University of Oxford

bHLH and bHLH-LZ factor exchange at promoters

DPhil Thesis

Supervisor: Prof. Colin Goding

Trinity Term 2014

Pakavarin Louphrasitthiphol
Ludwig Institute for Cancer Research
Nuffield Department of Clinical Medicine

Abstract

Mammalian promoters often contain DNA-elements that can be bound by a number of closely related transcription factors (TFs) that cannot bind to the same DNA-element simultaneously. It is possible that each TF responds to distinct cues, allowing the gene to be activated in response to multiple stimuli. An alternative possibility is that each TF binds sequentially, each contributing to the pre-initiation events leading up to transcription.

Here, we explore the exchange of basic-Helix-Loop-Helix-Leucine-Zipper (bHLH-LZ) factors, USF1, USF2 and MITF at the *TYROSINASE* promoter following induction by UVB-irradiation and methotrexate-administration. We demonstrate, for the first time in human melanoma, differential induction kinetics of *TYROSINASE* gene in response to an initial or re-induction, a phenomenon akin to “transcription memory” previously described in yeast.

We also show that USF2, specifically detected by two different antibodies targeting the N-terminal region, is largely cytoplasmic, at least in the cell lines we have investigated. We also showed that nucleo-cytoplasmic shuttling of these USF2 species is partly regulated by glucose. Using deletion mutants, we demonstrated the requirement of the amino-acids surrounding the USF-specific region and the basic domain in nuclear localisation of USF2, and that amino-acids $1 - 193$ appear to enhance dimerization of USF2 in addition to the classical HLH-LZ dimerization domain.

We will further investigate the role(s) played by MYC, MITF, HIF and USF exchange at common targets (which we identified through our ChIP-seq analysis) in gene activation and the effect on the (re)activation potential of these genes when DNA-binding by one or more of these factors are abolished, as well as when the promoter is monopolised by one of these factors through overexpression using cell lines expressing one of the bHLH-TFs under a tet-inducible promoter. In the long run, we aim to understand the potential differences in the role(s) of each bHLH-factors co-occupying E-box elements.

Table of contents

Abstract I
Abbreviations VII
1 Introduction 1
1.1 Classical transcription 1
1.2 Transcription factors (TFs) 2
Regulation of transcription by TFs 3
"TF binding kinetics and gene activation" and "bookmarking" 4
1.3 Adaptive transcription memory 5
Promoter education at *S. cerevisiae PHO5* 7
1.4 The bHLH and bHLH-LZ family 11
The classification of the bHLH-proteins 13
The HLH- and LZ-domain 13
bHLH-TFs to be studied and their known properties 14
1.5 MYC/MAX/MAD and related family of bHLH-LZ-TFs 15
MYC 16
MYC and cancer 16
MYC's biological function 17
Regulation of MYC's activity 18
MYC mediates transcription activation and chromatin remodelling 18
MYC as a repressor 20
MAX 20
1.6 Microphthalamia-associated transcription factor (MITF) 21
MITF mutants 21
MITF and development 22
Transcription activation by MITF 22
Regulation of MITF 23
Regulation by cAMP and CREB 23
Regulation by c-Kit 24
MITF and cancer 25
1.7 Hypoxia inducible factor (HIF) 25
Biological function of HIF 27
Transcription activation by HIF 28
Regulation of HIF-α proteins 29

HIF and cancer 30
Effects of loss of HIF-α in cancer 30
HIF's transcriptional targets and cancer 31
Cross-talk between HIF and MYC family proteins 31
1.8 Upstream stimulatory factors (USF) 33
USF and development 34
Regulation of USF 35
p38 and USF 35
Other regulation of USF 35
USF and transcription control 36
Transcription activation by USF 37
Cooperative binding of USF with cue-specific and tissue-specific TFs 37
Transcription repression by USF 38
USF may switch between an activator and a repressor to regulate gene expression 39
USF and cancer 39
1.9 Factor exchange and promoter education in mammals 41
Evidence and consequence of factor exchange in mammals 41
USF and promoter education 42
USF and promoter priming 43
USF factor exchange with inducible TFs through direct competition 44
1.10 The aims of the project 45
2. Material and Methods 47
2.1 Chemicals, buffers and reagents 47
2.2 Bacterial methods 48
2.3 Mammalian cell line methods 49
2.4 Nucleic acid methods 52
2.5 Protein techniques 56
2.6 Cell biology methods 61
2.7 Bioinformatics 62
2.8 Primer list 68
2.9 Antibody list 69
3. Results 70
3.1 ChIP-seq analysis of bHLH-TFs 72
3.1.1 MITF ChIP-seq 81

3.1.2 MYC ChIP-seq 84
3.1.3 USF ChIP-seq 87
3.1.4 HIF ChIP-seq 93
Distinction between HIF-1α and HIF-2α 94
3.1.5 Binding specificity 101
DNA-phosphate backbone contact and genome titration 101
Target E-box sequence preference 109
Core-E-box preference 109
Flanking sequence preference 113
3.1.6 Binding cooperativity 120
3.1.7 Co-occupancy 126
Motifs of the co-occupied peaks 132
Relative binding at co-occupied sites and comparability of ChIP-seq and qPCR 142
3.1.8 501mel specific peaks 144
3.2 RNA-seq 151
Basic RNA-seq analysis 153
Differential gene expression analysis 155
Assigning gene expression data to ChIP-seq peaks 158
GO-analysis of RNA-seq datasets 160
Correlating fold-changes obtained from RNA-seq and RT-qPCR 168
3.3 Factor exchange and promoter education 170
USF and histone modifications 170
Factor exchange at the human *Tyrosinase* (*TYR*) gene 172
Induction of the *TYR* gene by UVB 174
Induction of the *TYR* gene by an α-MSH analogue 179
Induction of MITF via Methotrexate 181
Other means to induce factor exchange 186
Approaches to interfere with bHLH-TFs' function 195
USF knock-out 195
A-USF (dominant-negative USF) 197
3.4 Regulation of USF2 subcellular localisation 204
USF2 is cytoplasmically localised 204
Nuclear-cytoplasmic shuttling of USF2 206
Over-expressed, heterologous USF2 is mainly nuclear localised 207

Induction of USF2 nuclear accumulation 209
Regulation of USF through nutrients 209
Glucose stimulation of USF2 212
Defining the regions in USF2 required for sub-cellular localisation 218
Intramolecular interaction of USF2 224
USF2 sub-cellular localisation 228
USF2 localisation to organelles/cellular structures. 231
USF2 and mitosis 231
4. Discussion 241
Aspects to consider while studying factor exchange in humans 241
The thought experiment 241
Regulation of factor exchange *in vivo* 245
Genome-wide analysis 248
Target specificity against the observed bound-targets 253
bHLH-TF exchange 257
Regulation of USF2 subcellular localisation 260
5. Future directions 264
6. Supplementary Figures and Tables 267
6.1 ChIP-seq analysis of bHLH-TFs 268
6.1.1 MITF ChIP-seq 273
6.1.2 MYC ChIP-seq 275
6.1.3 USF ChIP-seq 277
6.1.4 HIF ChIP-seq 281
6.1.5 Binding specificity 287
DNA-phosphate backbone contact and genome titration 287
Core-E-box and the flanking sequence preference 288
6.1.6 Binding cooperativity 295
6.1.7 Co-occupancy 298
6.3 Factor exchange and promoter education 301
Factor exchange at the human *Tyrosinase* gene (*TYR*) 301
Induction of *TYR* gene by UVB 301
Approaches to interfere with bHLH-TFs 304
6.4 Regulation of USF2 subcellular localisation 305
Induction of USF2 nuclear accumulation 305

Glucose stimulation of USF2 305
Defining the regions in USF2 required for sub-cellular localisation 306
Intramolecular interactions of USF2 308
USF2 sub-cellular localisation 310
USF2 localisation to organelles/cellular structures. 315
USF2 and mitosis 319
Bibliography 1

Abbreviations

0-9	
4-OHT	4-hydroxy tamoxifen
A	
α-MSH	Alpha melanocyte stimulating hormone
AdML	Adenovirus major late
AhR	Aryl hydrocarbon receptor
aka	Also known as
APS	Ammonium persulfate
ARNT	Aryl hydrocarbon receptor nuclear translocator (aka HIF1-β)
B	
b	Basic (domain)
BAF	Brm-associated factor
bHLH	Basic Helix-Loop-Helix
bHLH-LZ	Basic-Helix-Loop-Helix-Leucine-Zipper
BNIP3	Bcl-2/E1B 19 kDa-interacting protein 3
C	
CAD	C-terminal transcription activation domain
CAD gene	Carbamoyl-phosphate synthase/aspartate transferase/dihydroorotase
cAMP	3'-5'-cyclic adenosine monophosphate
CBP	CREB-binding protein
cDNA	Complementary DNA
ChIP-seq	Chromatin immunoprecipitation followed by high throughput sequencing
ChoRE	Carbohydrate responsive element
ChREBP	Carbohydrate responsive element-binding protein
CKI	Cyclin dependent kinase inhibitor
CKII	Casein kinase II
CoIP	Co-immunoprecipitation
CRE	cAMP response element
CREB	cAMP response element-binding protein
CT	Cholera Toxin
D	
DCT	Dopachrome tautomerase
DMEM	Dulbecco's modified eagle medium
dn	Dominant negative
dsDNA	Double stranded DNA
DTT	Dithiothreitol
E	
END1	Endothelin-1
EPAS-1	Endothelial PAS domain protein 1 (aka HIF-2α)
ER	Oestrogen-receptor
ER-A-USF	4-hydroxy tamoxifen inducible dominant negative USF
ESC	Embryonic stem cell

F	
FCHL	Familial combined hyperlipidaemia
FIH-1	Factor inhibiting HIF-1
FT	Flow-through
G	
gDNA	Genomic DNA
GO	Gene ontology
H	
HAT	Histone acetyl-transferase
HDAC	Histone deacetylase
HIF	Hypoxia inducible factor
HIF-1α	Hypoxia inducible factor-1α
HLH	Helix-Loop-Helix
HMT	Histone methyl-transferase
HNF	Hepatic nuclear factor
HRE	Hypoxia responsive elements
hRFC	Human reduced folate carrier
HS	Hypersensitive site
hSCF	Human stem cell factor
I	
ID	Inhibitory domain
Inr	Initiator
IP	Immunoprecipitation
IPAS	Inhibitory PAS domain protein (aka HIF-3α)
K-L	
LDHA	Lactate dehydrogenase A
L-PK	L-pyruvate kinase
LZ	Leucine-Zipper
M	
MAX	Myc-associated factor X
MB	MYC box
MC1r	Melanocortin 1 receptor
MEF	Mouse embryo fibroblast
MITF	Microphthalamia-associated transcription factor
MIZ-1	Myc-interacting Zn-finger protein-1
MLTF	Major late promoter transcription factor
MMP	Matrix metalloproteinase
MS	Mass-spectrometry
mSIN3	Mammalian homologue of the *Saccharomyces cerevisiae* global corepressor, SIN3
MTX	Methotrexate
MXI1	Max interactor 1
MYC	Myelocytomatosis
N	
NAD	N-terminal transcription activation domain
NDR	Nucleosome-depleted region

NES	Nuclear export signal
NFR	Nucleosome-free region
NLS	Nuclear localisation sequence
NPC	Nuclear pore complex
O	
ODC	Ornithine decarboxylase
ODD	Oxygen degradation domain
P-Q	
P/CAF	p300/CBP-associated factor
$p70^{S6K}$	p70 ribosomal protein S6 kinase
PAC	PAS-associated, C-terminal
PAI-1	Plasminogen activator inhibitor 1
PAS	Per-Arnt/AhR-Sim
PBS	Phosphate buffered saline
PBS-T	PBS + 0.5% Tween-20
PC	Principal component
PCA	Principal component analysis
PCR	Polymerase chain reaction
PDK1	Pyruvate dehydrogenase kinase
PHD	Prolyl-hydroxylase domain-containing protein
PI_3K	Phosphatidylinositol-4,5-bisphosphate 3-kinase
PIC	Protease inhibitor cocktail
PMSF	Phenylmethanesulfonylfluoride
P-TEFb	Positive transcription elongation factor b
PTM	Post-translational modification
qPCR	quatitative PCR
R	
RE	Restriction enzyme
REF	Rat embryo fibroblasts
RIN	RNA integrity number
RIPA	Radioimmunoprecipitation assay buffer
RNAPII	RNA-polymerase II
ROS	Reactive oxygen species
RPE	Retinal pigmented epithelia
RPMI-1640	Roswell park memorial institute medium-1640
RSK	Ribosomal S6 protein kinase
RT-qPCR	Reversed transcribed-quantitative PCR
S	
SCF	Stem cell factor
SDS	Sodium dodecyl sulphate
SDS-PAGE	Sodium dodecyl sulphate polyacrylamide gel electrophoresis
Si	Steel factor
SID	Sin3 interaction domain
SRE	Sterol regulatory element
SREBP	Sterol regulatory element-binding protein

T	
TA	Transcription activator
TAD	Transcription activation domain
TBS	Tris-buffered saline
TBS-T	TBS + 0.5% Tween-20
TEMED	Tetramethylethylenediamine
TERT	Telomerase reverse transcriptase
TF	Transcription factor
TFBS	Transcription factors binding site
TPA	Phorbol 12-myristate 13-acetate
TSS	Transcription-start site
TTS	Transcription-termination site
TYR	Tyrosinase
TYRP1	Tyrosinase-related protein 1
U	
UAS	Upstream activating sequence
UCSC	University of California at Santa Cruz
USF	Upstream stimulatory factor
USR	USF specific-region
UTR	Untranslated region
UV	Ultra-violet
V-W	
VEGF	Vascular endothelial growth factor
VHL	von Hippel-Lindau
WCE	Whole cell extract

1 Introduction

1.1 Classical transcription

The classical view of gene expression and chromosome dynamics has linked transcription initiation to binding of transcription activators (TAs) and associated co-factors (Gill, 2001; Ptashne and Gann, 1997). This leads to chromatin remodelling permissive for the recruitment of the basal transcription machinery necessary to fire transcription (Figure 1.1) (Krishnamurthy and Hampsey, 2009). Central to this is the target specificity conferred by transcription factors (TFs).

TFs play pivotal regulatory roles in gene activation; hence they must be regulated in order to prevent aberrant transcription initiation (Latchman, 1997). Many regulatory mechanisms exist including: subcellular localisation (Whiteside and Goodbourn, 1993), allostery (Kumar and McEwan, 2012), post-translational modifications (PTMs) (Bohmann, 1990; Lalli and Sassone-Corsi, 1994), expression levels, protein stability and degradation (Edmondson and Olson, 1993; Li et al., 2011).

Beyond the availability and ability of the TFs themselves to bind naked DNA, regulation also happens at the level of the accessibility and availability of TF binding sites (TFBSs) (Schulze and Wallrath, 2007). In higher eukaryotes, DNA is wrapped around core histone proteins to form nucleosomes (Kornberg, 1974). The phasing of nucleosomes and the rotational positioning of DNA around nucleosomes may occlude or expose different sets of TFBSs (Bai and Morozov, 2010; Jiang and Pugh, 2009; Rippe et al., 2007). This implies that nucleosome organisation is dynamic (Henikoff, 2008) and must be amenable for remodelling to expose occluded sites during gene activation (Avvakumov et al., 2011; Janicki et al., 2004). This initial process of “gene priming” is accomplished through a specific class of TFs, called “pioneer TFs” that are able to bind to their cognate site in condensed chromatin and alter the epigenetic landscape in order to make the locus accessible for

downstream TFs (Zaret and Carroll, 2011). Once, a TFBS is exposed, there may also be competition between factors binding to a shared or overlapping TFBS (Boyd and Farnham, 1999).

1.2 Transcription factors (TFs)

TFs act as *trans*-acting factors binding to specific nucleotide sequence on DNA-*cis*-regulatory elements to activate or repress defined sets of genes. The combination of TFs expressed in a given cell specifies the transcription profile which defines the character of a

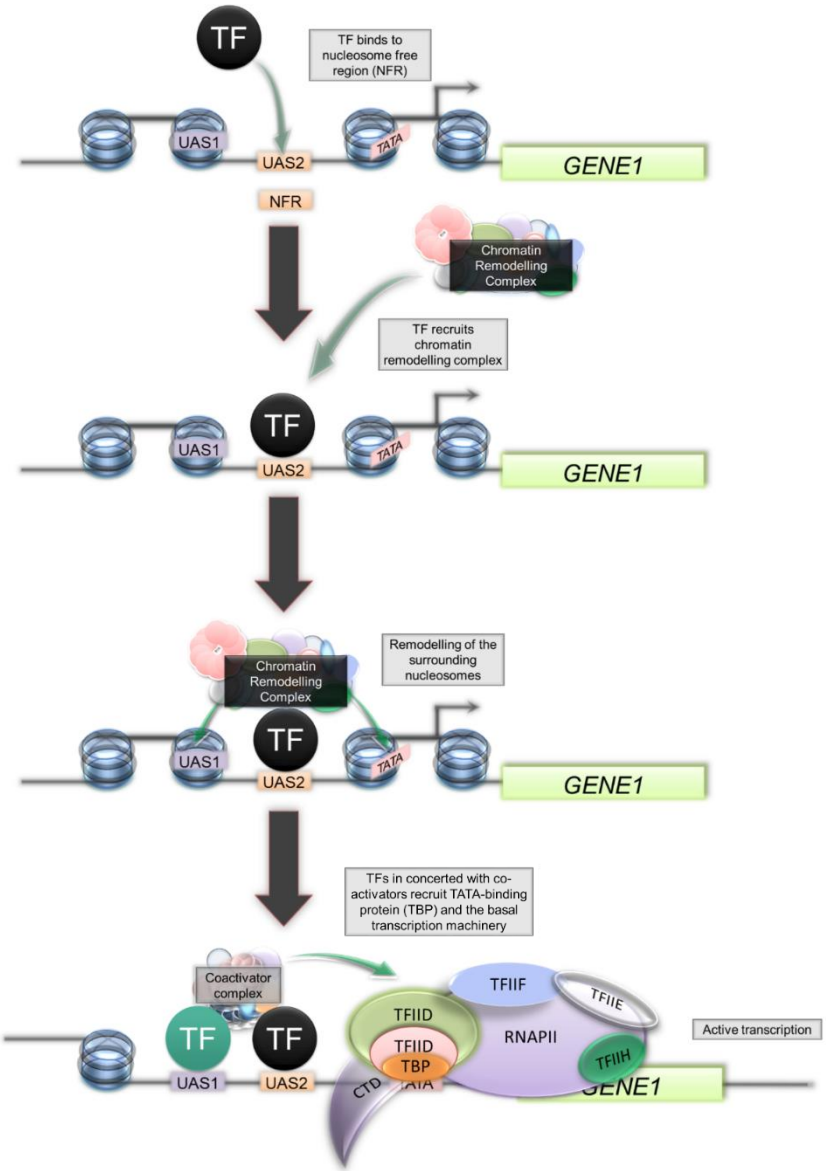


Figure 1.1 Schematic representation of classical transcription. Transcription factor (TF) binds to upstream activating sequence (UAS) in the nucleosome free region (NFR), resulting in recruitment of histone modifiers and chromatin remodelling complexes to establish an epigenetic landscape suitable for recruitment of the basal machinery necessary for transcription. The concerted action of TFs bound at the promoter and co-activators recruited result in establishment of transcription-competent RNA-polymerase II (RNAPII) and gene activation. TFs may also promote promoter clearance.

given cell type. As many TFs are ubiquitously expressed, yet display cell-type specific binding profiles, interactions between TFs and their cognate sequences is insufficient to explain the observed transcription output. Various factors including: chromatin structure, epigenetic modifications, and the presence of other TFs competing for DNA-*cis*-regulatory elements act in concert to define the expression profile at a given time.

Regulation of transcription by TFs

The importance of transcription regulation is highlighted by the observation that many steps leading to transcription activation can be interfered with by transcription repressors (Cowell, 1994). Although relevant, the understanding of the mechanism of transcription, how TFs recruit mediators and cofactors to drive changes in the epigenetic landscape, and the consequences of those changes is beyond the scope of this thesis (for review see Hahn and Young, 2011). Recent reviews on the mechanism of transcription by RNA-polymerase II (RNAPII) (Kaplan, 2013), the intricate regulations of RNAPII by cofactors at different stages from the formation of pre-initiation complex to elongation (Barrero and Malik, 2013) are available elsewhere. The role of epigenetic modifications [DNA methylation and histone modifications (Miller and Grant, 2013)], histone variants (Hubner et al., 2013; Law and Cheung, 2013)], chromatin structure and higher order genome organization in transcription regulation has also been reviewed elsewhere. For now, it suffices to be aware that binding of TFs often leads to changes in local chromatin properties through interaction and recruitment of chromatin modifying enzymes (Kooistra and Helin, 2012) and chromatin remodelling complexes (Aoyagi et al., 2003; Watanabe and Peterson, 2010).

TFs generally act as molecular switches that turn on gene expression. This is achieved through recruitment of histone modifiers, chromatin remodellers and other cofactors necessary for gene activation. Some TFs can stimulate transcription of active genes, a process termed “transcription amplification”. One such example is MYC that can act at the promoter clearance stage of active genes via the recruitment of positive transcription elongation factor b (P-TEFb). This promotes productive transcription elongation (Rahl and

Young, 2014), a rate limiting step of active transcription (Kwak and Lis, 2013). This is in line with the observation that many of MYC's positively-regulated genes have MYC's E-boxes located downstream of the promoter region (Bello-Fernandez et al., 1993; Eilers et al., 1991; Reisman et al., 1993; Wagner et al., 1993).

The distinct mode of gene activation by different TFs is illustrated through their ability to activate genes in a manner dependent on the position of their binding sites. For example, the TFBS (E-box) in the 5'UTR of the *carbamoyl-phosphate synthase/aspartate transferase/dihydroorotase* (*CAD*) gene and intron one of the *α-prothymosin* gene are bound by both MYC and USF (Boyd and Farnham, 1997). However, both genes are only activated by MYC (Boyd et al., 1998; Desbarats et al., 1996). Although USF proteins can function as TAs they are unable to activate these genes when the E-boxes are located downstream of the transcription start site (TSS).

"TF binding kinetics and gene activation" and "bookmarking"

Sequence specific DNA binding by a TF is a combination of generic binding to the DNA phosphate backbone and the sequence specific base contact. It is conceivable that to induce gene activation, a TF has to dwell at a promoter sufficiently long to recruit all the complexes necessary for gene activation (co-activator, chromatin remodeller, etc.). As this "recruitment" is likely diffusion driven, a low affinity binding site where the TF resides for a short period of time is less likely to complete the process of forming transcription competent, basal transcription machinery for gene activation. The proposed requirement for stable binding by a TF to activate a gene has been indirectly demonstrated for USF binding at the human reduced folate carrier (hRFC)-A1/A2 gene (Payton et al., 2005). USF activates *hRFC-A1/A2* gene when bound to E-box(b) (5'-CACGTG-3') but represses the gene when bound to E-box(a) (5'-CACTTG-3'). The former is a consensus USF binding site, which presumably allows USF to bind stably, while the latter is a weak USF binding site and is, perhaps, only bound transiently by USF. Furthermore, cooperative binding of USF with GATA to E-box(c) (5'-GTGGAC-3') which is also a USF low affinity site results in gene

activation, likely because cooperative binding with GATA help retains USF for sufficiently long to activate the gene.

The notion that TF binding kinetics is a critical determinant of gene activation helps explain why a TF does not invoke transcription wherever it binds. Binding of a TF to a TFBS with high k_{off} may not be able to complete the transcription pre-initiation steps necessary for transcription despite its ability to bind to that TFBS. This concept explains why highly related TFs that are potent activators and have overlapping target sequence specificity may be able to activate or repress different subsets of genes even when ChIP-assays demonstrate binding to all those genes.

The “epigenetic bookmarking” concept describes how stable binding of TFs lacking a transactivation domain (TAD), such as MAX, help retain an open chromatin structure permissive for binding of the “real” TA resulting in a more rapid induction kinetic. Epigenetic bookmarking may also be achieved through binding of TFs at TFBSs with high k_d (in other words, low dwell time). The brief occupancy of the TF helps retain an open chromatin structure for binding by cognate TA of those genes.

The subtle differences in the binding specificity of highly related TFs dictate the binding kinetic at a particular site and can influence transcription outcome based on their relative binding affinity. Thus, regulated exchange of TFs with overlapping DNA-binding specificities at a TFBS may represent a novel regulatory mechanism of gene expression.

1.3 Adaptive transcription memory

It is widely accepted that certain promoters can be poised for activation in a developmental stage-specific manner (Bernstein et al., 2006). That is, some genes are primed for rapid activation in response to specific developmental cues (Mercer et al., 2011). These genes may be characterised by the bivalent H3K4me3/H3K27me3 histone modifications as observed at the *Hox* cluster and genes that are required during early development (Cui et

al., 2009). Alternatively, some enhancers may be poised by pioneer factors which can bind to nucleosome-occluded TFBSs and induce early chromatin changes (H3K4me1) which mark those loci for activation during development by TFs that are unable to bind to nucleosome-occluded TFBSs (Zaret and Carroll, 2011). This sort of "gene priming" was first shown for the liver-specific enhancer of the *ALB1* gene in the precursor gut endoderm prior to its activation in nascent liver (Bossard and Zaret, 1998; Selimyan et al., 2013).

Upon differentiation, lineage specific TFs establish and maintain the repertoire of genes to be switched on by converting bivalent histone marks to active histone marks, or prime silenced genes for rapid activation. This is in contrast to "latent enhancers" which are not primed and have to acquire histone features (including H3K4me1, H3K9ac and H3K27ac) and appropriate chromatin remodelling to establish the epigenetic landscape necessary for gene activation following stimulation (Ostuni et al., 2013).

Some genes, however, are able to adjust their induction kinetics appropriate to the stimulus, and adapt their responses on repeated cycles of stimulation (Kundu et al., 2007; Light et al., 2010). This implies mechanisms exist to provide "memories" of previous signals at promoters in the absence of on-going gene expression. These changes in activation potential without changes in the underlying DNA-sequence likely involve epigenetic modifications that inform genes of their current activation potential in a way that is dependent on their previous transcription activity and can be termed "adaptive transcription memory". Adaptive transcription memory has been associated with a more rapid re-induction kinetic as has been shown for *HXK1*, *GAL1* and *INO1* (Tan-Wong et al., 2009; Brickner, 2009), but, concievably, may also be associated with a slower re-induction kinetic.

Any genes subject to adaptive transcription memory can broadly be said to exist in three qualitatively distinct states: a long term repressed state, an active state, and an epigenetic memory state, each of which may be associated with a distinct nuclear territory (Ahmed and Brickner, 2007; Andrulis et al., 1998; Brickner and Walter, 2004; de Wit and van

Steensel, 2009; Feuerbach et al., 2002; Tan-Wong et al., 2009). The last decade has seen a large number of publications reporting many inducible genes in yeast cycling between these three distinct states in response to cycles of repression, stimulation and reactivation. The prime example would be the sugar responsive *GAL* loci (Brickner, 2009; Kundu and Peterson, 2009).

Promoter education at *S. cerevisiae PHO5*

The phosphate regulated *PHO5* gene in *S. cerevisiae* represents an excellent example of an inducible gene that may possess adaptive transcription memory. *PHO5* is activated by intracellular phosphate depletion (Hirst et al., 1994), primarily during passage through S-phase, following the cascade depicted in Figure 1.2. Inactivation of the Pho80/Pho85, cyclin/cdk complex is permissive for nuclear translocation of the basic Helix-Loop-Helix (bHLH)-TF Pho4 (Jayaraman et al., 1994; O'Neill et al., 1996), where it binds to the

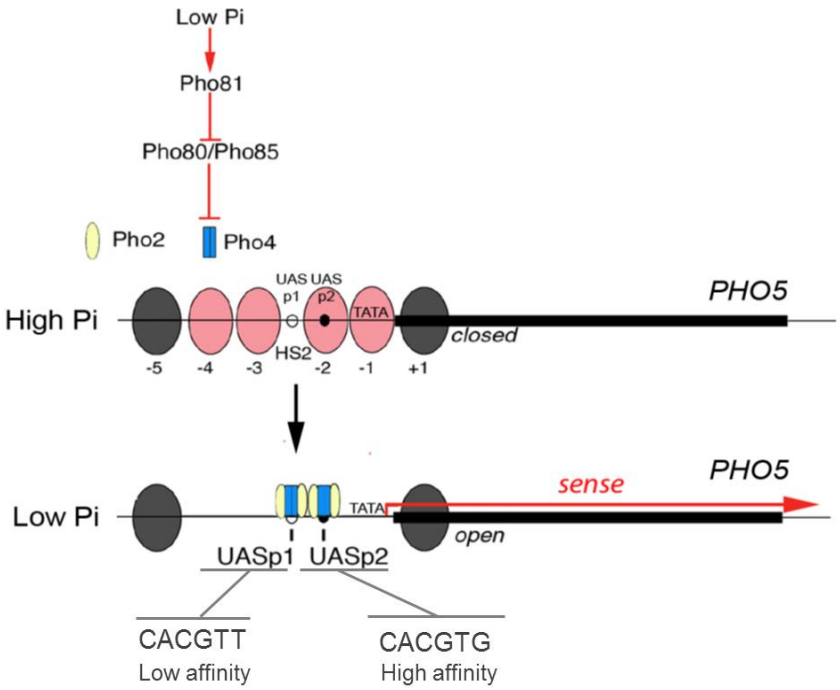


Figure 1.2 Gene activation at the *PHO5* of *S. cerevisiae*. Intracellular phosphate depletion activates Pho81 that inhibits the Pho80/Pho85 cyclin/cdk complex that is required for cytoplasmic retention of Pho4. Upon nuclear entry, Pho4 interacts with the homeodomain protein Pho2, forming a DNA-binding competent complex. Initial binding of Pho2/Pho4 occurs at a DNA hypersensitive site (HS2), containing a low affinity binding site, UASp1, located between nucleosome -2 and -3. This is followed by remodelling of surrounding, highly-positioned nucleosomes, revealing the high affinity UASp2. UASp2 is initially bound by Cbf1-Pho2 but is gradually displaced by Pho4-Pho2 following prolonged phosphate starvation. Binding of Pho4-Pho2 coincides with the recruitment of the basal transcription machinery and subsequent transcription of the sense mRNA. Abbreviation: Pi, phosphate; UASp, promoter Upstream Activating Sequence; HS, hypersensitive site.

upstream activating sequence (UAS) in cooperation with a homeodomain factor Pho2 to remodel four highly positioned nucleosomes (Barbaric et al., 1998; McAndrew et al., 1998; Reinke and Hörz, 2003). This unveils the TATA-box (Martinez-Campa et al., 2004) and permits the recruitment of the basal transcription machinery for transcription.

Recent evidence has, however, implicated the role of the bHLH-Leucine Zipper (bHLH-LZ) factor Cbf1 in regulation of *PHO5* and suggested that transcription activation and memory at *PHO5* requires a timely exchange of Pho4 and Cbf1 at the UAS (Goding lab unpublished data). The available evidence suggests that initially, Pho4 binds the low affinity UASp1 element located between nucleosomes -2 and -3. This leads to remodelling of nucleosome -2 to reveal the high affinity UASp2 (Venter et al., 1994). UASp2 is bound by Cbf1 in a complex with Pho2 that only occurs in high phosphate conditions. Binding of Cbf1 is required for subsequent stable binding of Pho4, and as phosphate levels drop Cbf1 binding to the UAS is replaced by Pho4-Pho2 that is favoured under low phosphate conditions. Once stably bound, Pho4 recruits the basal transcription machinery to fire transcription.

Non-productive turnover of TF binding was recently reported in a genome-wide study of *S.cerevisiae* Rap1 (Lickwar et al., 2012). Analogous to Pho4, stable binding of Rap1 correlates with transcription, while rapid turnover ("treadmilling"), although giving a similar ChIP profile, does not result in productive transcription. The results in this report suggested that competition for DNA-occupancy between nucleosomes and a TF leads to a "clutch-like mechanism" which can rapidly transform a treadmilling TF into a stable binding state permissive for productive transcription. In this light, Cbf1, appears to facilitate the nucelosome remodeling which permits stable binding of the treadmilling Pho4, most likely via the recruitment of ISWI/IOC3 chromatin-remodelling complex (Moreau et al., 2003) which destabilises the nucleosomes competing with Pho4 binding. In the absence of Pho4, binding of Cbf1 to the UAS does not lead to transcription activation, and unlike Pho4, recruitment of Cbf1 to DNA via a heterologous DNA-binding domain does not lead to

transcription activation (Thomas et al., 1992). Thus, Cbf1 is not a classical transcription-activator but is nonetheless, required for transcription activation by a classic bHLH-activator Pho4 (O'Neill et al., 1996). The ability of a factor to bind a promoter and actively establish a permissive epigenetic environment for the subsequent binding of a classical TA is a process we term 'promoter education' which is distinct from the "epigenetic bookmarking" concept (Rada-Iglesias et al., 2011; Zacharioudakis et al., 2007).

In addition to its role on initial activation of *PHO5*, Cbf1 in association with Pho2 is transiently re-recruited to the UAS on phosphate addition, a time when Pho4 is rapidly exported out of the nucleus. Cbf1 re-recruitment post-activation is required for association with components of the nuclear pore complex (NPC) and is critically dependent on prior activation of the *PHO5* promoter by Pho4. Since, NPC association has been previously associated with transcription memory (Brickner et al., 2007; Brickner, 2009; Tan-Wong et al., 2009), Cbf1 may be a *bona fide* memory factor and its role on transcription shut down may be to hold the promoter in a conformation compatible with rapid reactivation by Pho4 should phosphate levels decrease during Cbf1's 30' binding window.

The results from *PHO5* point to the existence of factors like Cbf1 that act as so-called 'promoter education factors' whose binding to target DNA sequences does not lead to productive transcription activation, but are nonetheless required for induction of gene expression by classical TAs. The role of Cbf1 post-activation at *PHO5*, suggests that promoter education is also implicated in shaping the epigenetic landscape post-gene activation ("post-activation promoter education"), leading up to transcription memory, either by recruitment of an epigenetic bookmarker or establishment of an epigenetic landscape that leads to a re-induction kinetic that is distinct from the initial activation. Refer to Figure 1.3 for a detailed comparison of classic transcription activation (Figure 1.3A), adaptive transcription memory (Figure 1.3B) and the promoter education involved at genes subjected to adaptive transcription memory cycling between the long-term repression, active transcription and epigenetic memory (Figure 1.3C).

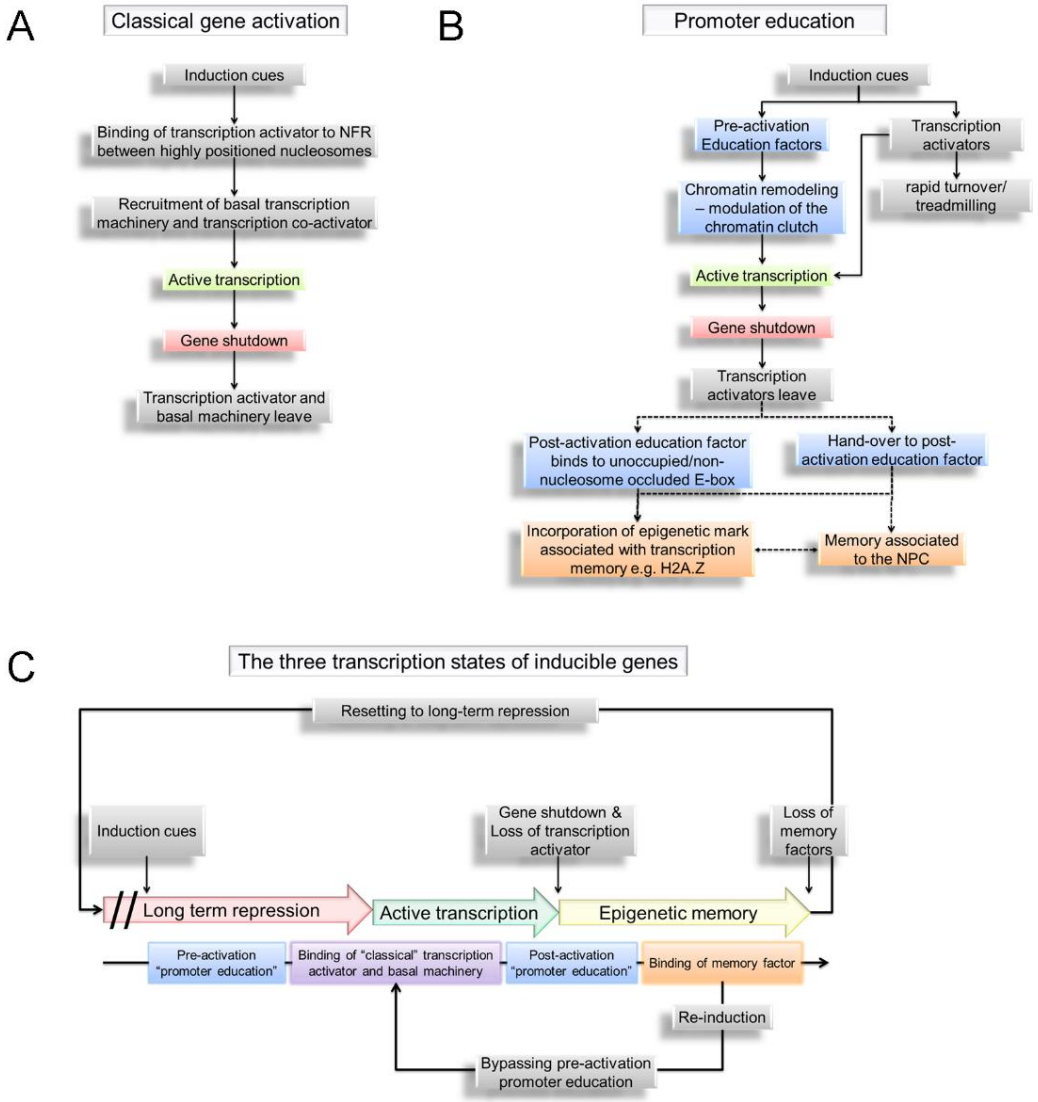


Figure 1.3 Schematic diagrams showing steps involved in classical gene activation and promoter education. **A)** In the classical view of gene activation, transcription of a gene is dependent on the binding of transcription activator (TA) that recruits the basal transcription machinery to fire transcription. Transcription ceases when the TA leaves as the TAs' roles are to recruit, stabilise and promote the transcription competent form of the basal transcription machinery. **B)** In some genes, however, the TA is unable to stably bind to the promoter in the absence of prior binding by the "pre-activation education factor" (aka pioneering factor) whose job is to establish an epigenetic landscape permissive for stable binding of TA and subsequent gene activation. In the absence of promoter education factor, TA is unable to stably bind due to competition with the surrounding nucleosome, effectively "treadmilling" at the DNA binding element. Upon gene shutdown, the TA either dissociates, leaving the nucleosome depleted promoter which can be bound by "post-activation education factor" or the TA may actively recruit a post-activation education factor, effectively handing over the promoter to post-activation education factor which leads to "adaptive transcription memory". This memory may be linked to association with the NPC that has been suggested to bypass initial recruitment of basal machinery and pre-mRNA processing and export machinery permitting rapid gene re-induction (Ahmed and Brickner, 2007; Dieppois and Stutz, 2010; Menon et al., 2005). Alternatively, transcription memory may be associated with changes in epigenetic modification or incorporation of histone variants such as H2A.Z (Brickner et al., 2007) which is associated with a more open chromatin conformation presumably lowering the kinetic barrier for the next round of gene activation. **C)** A schematic diagram correlating the three states of a gene, the events during promoter education and the exchanging of transcription factors. Upon receipt of inductive cues, pre-activation promoter education factors bind to the long-term repressed genes and establish epigenetic landscape suitable for stable binding of the classical TA whose binding correlates with transcription activation. After gene shutdown, post-activation education factors bind and establish epigenetic memory. During the memory period, the gene can be rapidly reactivate as the memory factor help retains the epigenetic landscape that is non-permissive for transcription but is distinct from that of the long-term repressed state. Beyond the memory period, in the absence of any re-activation of the gene, the gene returns to the long-term repressed state and memory is loss. Abbreviation: NFR; nucleosome free region, NPC, nuclear pore complex.

Here, we wish to explore the concept of promoter education, revealed by Cbf1 in *S. cerevisiae*, in the mammalian system where the identity of promoter education factors and the mechanisms underlying "transcription memory" are the key issues. Understanding this phenomenon may ultimately allow us to understand, for example, how cells may adapt to respond more rapidly to growth factor stimulation or how cancer cells dampen the apoptotic response following prolonged administration of DNA-damage-inducing anti-cancer compounds.

Crucial to the promoter education model at *PHO5* is the coordinate exchange of factors binding to shared DNA-elements which "educate" the gene during its activation and the formation of transcription memory which may be accompanied by gene re-localisation to the NPC. Since bHLH-families are well conserved, we suspect that they may play similar roles in promoter education of mammalian genes.

1.4 The bHLH and bHLH-LZ family

To understand how the bHLH- and bHLH-LZ-family of TFs may function in factor exchange and promoter education, it is important to understand the molecular mechanism underlying their actions and how these may be regulated. Of paramount importance is how target specificity is conferred and regulated as this determines which factors may exchange.

The bHLH-LZ-TFs have been implicated in a diverse range of biological processes including metabolism (Shimano, 2007; Wong and Sul, 2010), cell fate decisions

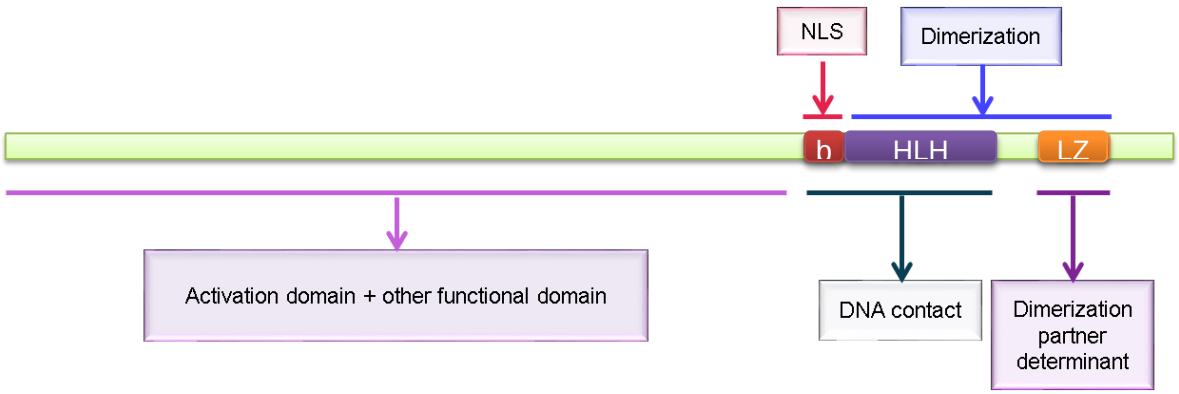


Figure 1.4 Diagrammatic representation of a basic helix-loop-helix leucine-zipper (b-HLH-LZ) transcription factors. NLS, nuclear localisation signal.

(Anantharaman et al., 2011), phenotype-switching (Hoek and Goding, 2010) and sex determination (Garrell, 1991). bHLH-LZ family proteins are characterised by; a region rich in arginine- and lysine- basic-amino-acid residues (b) responsible for DNA binding, sequence specificity and nuclear localisation; two amphipathic α-helical domains (HLH) required for dimerization followed by an amphipathic helix leucine-zipper (LZ) containing leucine heptad repeats required for dimerization and dimerization specificity (Dang et al., 1989; Ferre-D'Amare et al., 1993) (Figure 1.4) (for review see Pabo and Sauer, 1992). The bHLH-LZ-domain can be found in many transcription regulators such as the ubiquitous MYC, USF1, USF2 and TFII-I, the lineage specific MITF (melanocytes), ADD1 (adipocytes) and condition-specific sterol regulatory element-binding protein (SREBP, activated through low sterol level and insulin, exported from the ER and cleaved in the Golgi to release the soluble active form).

bHLH-LZ TFs function as dimers. The basic domain of each monomer binds to the major groove, contributing to the specificity of the half-site of the E-box motif (Figure 1.5) (Jones, 1990). Upon binding to the DNA, bHLH-LZ proteins undergo DNA-induced coil-to-helix folding transition (Anthony-Cahill et al., 1992; Ferre-D'Amare et al., 1994; Fisher et al., 1993a). The basic region and helix−1 of the HLH-domain form a contiguous α-helix that

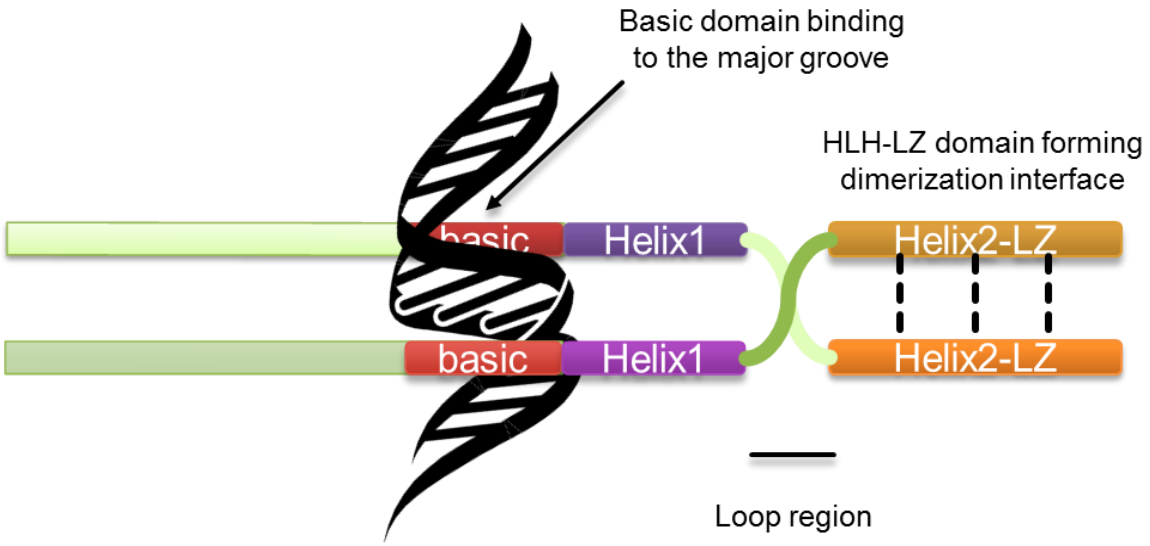


Figure 1.5 Graphical representation of a bHLH-LZ dimer bound to the DNA. The basic domain of each monomer confers the specificity for each half-site of the E-box motif through specific base contact at the major groove of the DNA.

connects to the second α-helix, helix−2, of the HLH-domain and the LZ (Brownlie et al., 1997; Ferre-D'Amare et al., 1993). Dimerization through the interaction of these four α-helices to form left-handed four-helix bundle allows precise positioning of the DNA-recognition basic domain into the major groove of the DNA (Figure 1.5). The bHLH-LZ-TFs bind a diverse set of related E-box motifs (CANNTG).

The classification of the bHLH-proteins

The bHLH-motif was first described for the murine TFs, E12 and E47 (Murre et al., 1989). Proteins containing bHLH-motifs were first classified based on their tissue distribution, target E-box specificities, and dimerization potential (Murre et al., 1994) resulting in 6 classes of bHLH-proteins (Massari and Murre, 2000). This was superseded by a phylogenetic based classification which took into account the E-box specificity, sequence conservation and the additional domain(s) (Atchley and Fitch, 1997). This resulted in six major phylogenetic groups where group B is probably the most ancestral of all (Table 1.1) (Ledent et al., 2002). Many of the phylogenetic group B proteins also possess a LZ-dimerization domain.

The HLH- and LZ-domain

The HLH-domain is sufficient for dimerization of many HLH-LZ proteins (Kerkhoff et al., 1991). The loop of the bHLH factors may contribute to overall DNA binding affinity as shown for MAX, USF, MITF and yeast Pho4 that contact the phosphate backbone outside the core E-box (Figure 1.6, marked by *****, the highly conserved lysine involved is highlighted in yellow) (Ferre-D'Amare et al., 1994; Shimizu et al., 1997).

The LZ functions independently of the bHLH-domain (Ma et al., 1994) and is responsible for dimerization specificity. For example, MITF preferentially forms homodimers (Pogenberg et al., 2012) whereas MYC must heterodimerize with MAX based on biophysical (Ayer et al., 1993; Brownlie et al., 1997; Muhle-Goll et al., 1994; Muhle-Goll et al., 1995) and *in vivo* studies (Grinberg et al., 2004).

Table 1.1

Classification of bHLH proteins

Phylogenetic group	Description	Murre et al. 1994 classification	Example
A	Bind to 5'-CA[G/C]CTG-3'	I, II	MyoD, Twist
B	Bind to 5'-CACGTG-3' or 5'-CATGTG-3'	III, IV	USF1/2, MYC, MITF
C	Bind to 5'-[A/G]CGTG-3'. Contain bHLH-PAS domain	N/A	HIF-1α, ARNT, Clock
D	Δbasic-domain, group A antagonists	V	ID
E	Bind to 5'-CACG[C/A]G-3' N-box. Contain orange domain and WRPW peptide	VI	Hairy
F	Contain COE dimerization and DNA binding domain	N/A	Coe

Evolutionary based classification of bHLH proteins. PAS domain is named after: *Drosophila* Period (Per), human aryl hydrocarbon receptor nuclear translocator (ARNT) and *Drosophila* Single-minded (Sim) (Zelzer et al., 1997). PAS domain consist of tandem ≈ 50 amino-acids repeats (aka PAS A and PAS B) separated by ≈ 150 amino-acids that are poorly conserved (Crews, 1998) and function as dimerization motifs (Kewley et al., 2004). Orange domain is a $30-$amino-acids domain C-terminal of bHLH, separated by a short, variable length sequence and function in transcription repression (Steidl et al., 2000) and dimerization (Davis and Turner, 2001).

bHLH-TFs to be studied and their known properties

To explore factor exchange and the promoter education hypothesis in humans we need to determine the appropriate factors to study. To this end we chose bHLH-TFs that may be

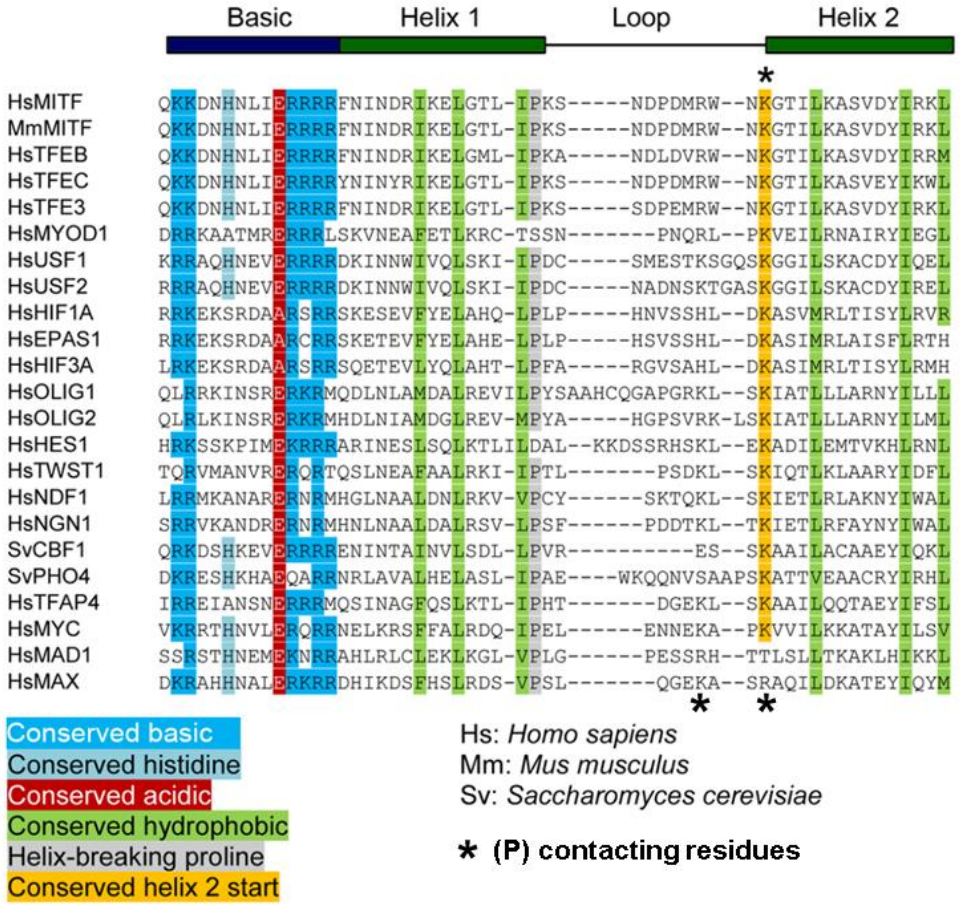


Figure 1.6 Sequence alignment of some bHLH-LZ-TFs showing the highly conserved basic and helix-looph-helix domain. ***** indicates additional phosphate contacting residues outside the basic domain. Figure adapted from our lab manuscript under preparation.

involved in a number of stages outlined in the promoter education scheme (Figure 1.3) and discuss some of their known biological properties. This knowledge will be instrumental in designing appropriate experiments. For example, optimal conditions to induce MITF, which promoters are likely to display factor exchange, and the time-scale of the experiment.

The following section will describe the known biology of the ubiquitous USF, the lineage specific MITF, and the extensively studied MYC. In addition, we will discuss HIF of the bHLH-PAS family that have been reported to intersect with MYC and USF.

1.5 MYC/MAX/MAD and related family of bHLH-LZ-TFs

The MYC/MAX/MAD family is involved in a diverse range of cellular processes including cell-cycle control (proliferation and apoptosis), cell fate decisions (differentiation, self-renewal and maintenance of pluripotency), signal transduction and metabolism (DePinho et al., 1991; Henriksson and Lüscher, 1996; Marcu et al., 1992).

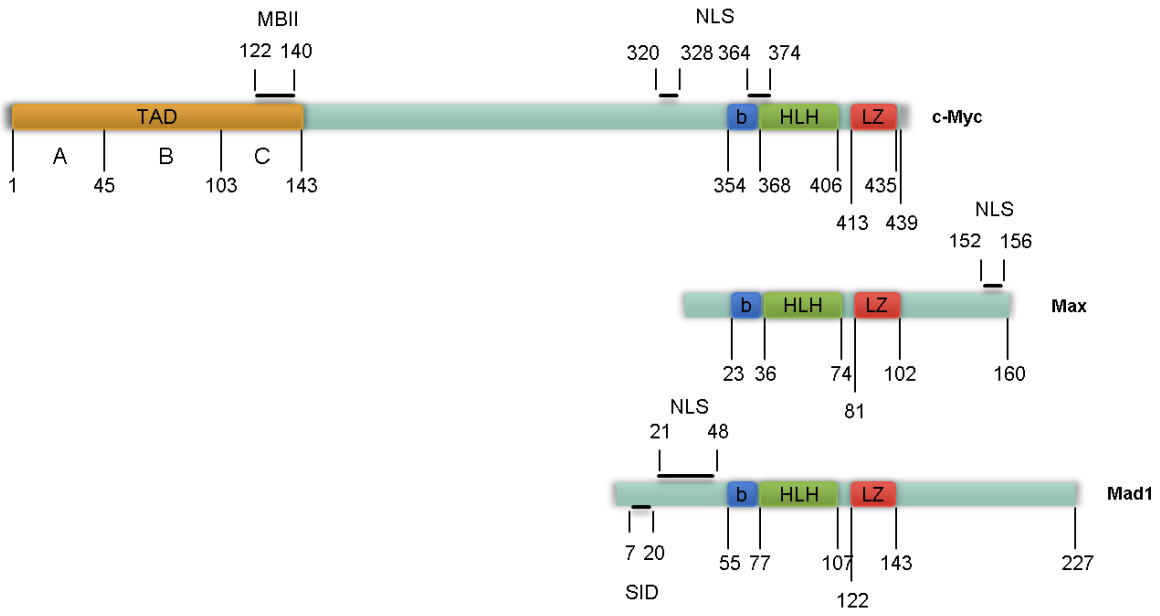


Figure 1.7 Diagrammatic representation of the MYC (represented here by c-Myc), MAX and MAD (represented here by Mad1) family. As with other bHLH-LZ-TFs, MYC/MAX/MAD family is characterised by the basic (b) DNA-binding domain adjacent to the helix-loop-helix (HLH) leucine-zipper (LZ) dimerization domain. **MYC** contains three N-terminal transactivation domains (TAD, denoted A, B and C), each of which is able to mediate gene activation. MYC Box II (MBII) is located at the end of the TAD and is required for MYC's ability to act as a repressor. **MAX** protein, on the other hand, lack a TAD and acts as a common, obligate binding partner for all MYC and MAD family members. **MAD** family members functionally antagonize MYC. Like MAX, MAD lack the N-terminal TAD. Instead MAD contains a Sin3 interaction domain (SID) that acts to recruit Sin3-HD1 repressive complex. Abbreviation: NLS, nuclear localisation sequence.

MYC/MAX/MAD family proteins are characterised by the highly conserved bHLH-LZ-domain (Figure 1.7). Although highly conserved within the subfamily, MYC and related-subfamily proteins contain relatively little homology outside the bHLH-LZ-domain.

MYC

MYC family proteins contain three N-terminal TADs (Kato et al., 1990), and a tripartite C-terminal domain containing: a basic domain (Blackwell et al., 1990), the HLH- and LZ-domain (Blackwood and Eisenman, 1991) (Figure 1.7).

The TAD interacts with retinoblastoma (pRb) and TATA-binding protein (Hateboer et al., 1993; Rustgi et al., 1991) and can activate transcription of a reporter gene $30-$ to $100-$fold when fused to the GAL4 DNA-binding-domain (Kato et al., 1990). Heterologously expressed MYC:MAX in yeast is able to activate transcription up to a $1000-$fold (Amati et al., 1992; Crouch et al., 1993), yet is only able to modestly activate reporter constructs ($2-$ to $10-$fold) in a transfection assay of a variety of cell lines (Hurlin et al., 1997 and reference therein). This implies that MYC transcription-activation is tightly limited in mammalian cells.

MYC and cancer

MYC was first identified in the late 1970s as the transforming sequence of an avian acute leukemia retrovirus (MC29), called v-Myc (for myelocytomatosis), which promotes a spectrum of malignancies such as carcinomas, myelocytomas and sarcomas (Roussel et al., 1979; Sheiness and Bishop, 1979). MYC is able to promote tumour formation via its ability to drive resting cells into the cell-cycle and to block cell differentiation (Grandori et al., 2000). It is probably through these abilities that MYC expression favours the establishment of immortalized cell lines (Barnett and Crouch, 1995).

MYC overexpression induces neoplastic transformation but also promotes apoptosis in cultured cell and transgenic animals (Askew et al., 1991; Coppola and Cole, 1986). Yet, MYC is found overexpressed in many human tumours (Adhikary and Eilers, 2005; Nilsson and Cleveland, 2003) and is the most commonly affected gene in human cancer. For

example, N-*myc* and L-*myc* amplification are found in neuroblastoma (Wenzel A., 1995) and small cell lung cancer respectively, and translocation of *MYC* onto either the heavy or the light chain immunoglobulin loci is observed in Burkitt's lymphoma (Xu et al., 2010). MYC's oncogenic potential generally requires cooperation with at least one additional oncogene such as *RAS* or *BCL-2* (Barnett et al., 1998). In fact, it has been long-recognised that MYC can act cooperatively to activate H-RAS (V12)-mediated transformation of rat embryo fibroblasts (REF) (Land et al., 1983). MYC induced cellular transformation is strongly correlated with MYC's ability to repress a number of promoters (Gartel and Shchors, 2003). Deregulation and amplification of MYC has also been associated with maintenance of cancer stem cells (Vita and Henriksson, 2006; Wang et al., 2008). Reviews on MYC and cancer can be found in Marcu et al., 1992; Meyer and Penn, 2008.

MYC's biological function

The MYC family comprising c-, N- and L-MYC represent classical nuclear proto-oncogenes. All MYC family members function through common mechanisms (Ma et al., 1993), although the specific aspect and the magnitude of the observed effects may differ. They bind to the same E-box targets as demonstrated by the ability of trans-activation incompetent mutants of any family member to dominantly supress the ability to transform REFs by all three MYC proteins (Mukherjee et al., 1992).

Homozygous deletion of either c-Myc or N-Myc is embryonically lethal (Stanton et al., 1992). Activation of MYC in resting fibroblasts is sufficient to drive cells into the cell cycle in the absence of serum, but also induces apoptosis unless supplied with specific survival cytokines (Evan and Littlewood, 1993; Marcu et al., 1992). MYC can also sensitize cells to apoptosis by various stimuli (Packham and Cleveland, 1995; Thompson, 1998), putting MYC at the centre of a life-and-death switch. Thus, MYC's activity is tightly regulated. One of the major regulatory mechanisms involves targeted degradation by the ubiquitin-proteasome system (for review see Farrell and Sears, 2014).

Regulation of MYC's activity

MYC is subject to various PTMs including phosphorylation and glycosylation that affect its transcription activation potential. For example, phosphorylation within Myc box I (MBI) by GSK-3 and MAPK (Henriksson et al., 1993; Pulverer et al., 1994) stabilize and modulate MYC's ability to interact with other proteins (Vervoorts et al., 2006).

Dimerization with MAX is obligatory for MYC's transcription-activation-activity (Amati et al., 1993a; Amati et al., 1992; Amati et al., 1993b; Blackwood and Eisenman, 1991; Gu et al., 1993; Kretzner et al., 1992). Thus, the transcription activity of MYC may be regulated through the availability of its dimerization partner, MAX that can also heterodimerise with MAD and MNT to repress genes. *In vitro* EMSAs demonstrate similar specificity and affinity to DNA by MYC:MAX, MAD:MAX and MNT:MAX (Sommer et al., 1998). Thus, MYC:MAX and MAD:MAX may act as molecular switches that activate and repress the same promoter, respectively (Ayer et al., 1993; Hurlin et al., 1996; Lüscher, 2001). The highly regulated expression of MYC and MAD family members (Ayer et al., 1993; Larsson et al., 1994; Wu, 1996; Zervos et al., 1993) allow for rapid factor exchange at promoters to regulate the transcription response.

MYC mediates transcription activation and chromatin remodelling

MYC regulates transcription by RNAP-I, -II and -III (Eilers and Eisenman, 2008) and is estimated to bind to $10-15\%$ of promoters (Dang, 2006). To activate genes, MYC recruits the SWI/SNF-like and TRRAP complexes which, in turn, recruit the HAT GCN5 (McMahon et al., 2000) or TIP60 (Frank et al., 2003) to hyper-acetylate histones H3 and H4. Acetylation neutralises the positive charge of the lysine side-chain is believed to weaken histone-DNA interaction and can also recruit bromo-domain containing proteins. Histone acetylation usually activates genes (Frank et al., 2001) but is not always sufficient (Fernandez et al., 2003). The requirement for histone acetylation in gene activation is in line with the observed decrease in histone acetylation at the telomerase reverse transcriptase (*TERT*) promoter (Xu et al., 2001) and lost of TRRAP binding and increased histone deacetylase (HDAC)-1

at the *CCND2* promoter and reduced *TERT* gene expression during differentiation of HL60 cells (Bouchard et al., 2001).

More recently, an increasing number of studies have implicated MYC in global regulation of transcription beyond its initial role as a classical TA acting in a gene specific manner. For example, MYC was proposed to be a universal amplifier of expressed genes as shown for two primary lymphocytes and embryonic stem cells (ESCs) of the mouse (Nie et al., 2012). This is achieved through stimulation of pause-released and facilitation of promoter re-loading which is believed to be non-gene specific based on the discordance in MYC's targets between different cell types. This mode of positive regulation of gene expression resulted in a non-linear amplification of gene expression where the more highly expressed genes are more strongly enhanced by MYC. Thus, MYC repressed targets may actually be activated albeit to a lesser extent than the average gene. The actual repression of MYC repressed targets may be mediated indirectly through MYC up-regulation of *EZH2* or miRNA (Bui and Mendell, 2010).

MYC may positively regulate gene expression globally through establishment and maintenance of positive histone marks. A genome-wide study showed that MYC is able to bind to a specific chromatin pattern defines by H3K4me3, H3K27me3 and hyper-acetylated H3, in an E-box independent manner (Guccione et al., 2006). Binding of MYC to pre-existing poised chromatin state is followed by histone hyper-acetylation/exchange reactions in various combinations depending on the cell lines (Martinato et al., 2008; Varlakhanova and Knoepfler, 2009). Depletion of MYC is associated with repressive chromatin state and disrupted nuclear localisation of HP1α that is phenocopied by overexpression of MAD. (For reviews on MYC regulation of global chromatin structure see Knoepfler, 2007; Varlakhanova and Knoepfler, 2009).

The direct involvement of MYC in global gene amplification through global histone modifications is not without dispute. Expression of *GCN5*, one of MYC positively regulated

genes, was shown to promote widespread histone-acetylation, epigenetic marks associated with gene activation, in MYC-null cells (Knoepfler et al., 2006).

MYC as a repressor

MYC may also act as a repressor in a manner independent of MYC's DNA-binding activity by directly binding to other TAs and prevent them from interacting with cofactors necessarily for their function as activators. For example, MYC interacts with Yin-Yang-1 (Shrivastava et al., 1993; Shrivastava et al., 1996), MIZ-1 (Myc-interacting Zn-finger protein-1) (Peukert et al., 1997) and TFII-1 (Roy et al., 1993) through its C-terminal domain to inhibit gene activation through initiator (Inr) sequences. While TFII-I also interacts with USF and MIZ-1, TFII-I binds specifically to the HLH-domain of MYC but not that of USF or MAX, suggesting that this DNA-independent repression might be specific to MYC.

MAX

MAX (Myc-associated factor X) was first identified through a screen of c-Myc against a cDNA expression library (Blackwood and Eisenman, 1991). Unlike the short-lived MYC and MAD (Blackwood et al., 1992), MAX is ubiquitously expressed throughout all stages of the cell cycle, is very stable, and is able to support MYC oncogenic activity (Berberich et al., 1992; Blackwood et al., 1992; Larsson et al., 1994). The predominant MAX isoforms are p21 and p22 (MAX and MAX9, respectively) (Arsura et al., 1995), expressed at equal levels. The latter has an extra nine amino-acids insertion (Blackwood and Eisenman, 1991). Both isoforms have an alternatively spliced isoform, ΔMAX, lacking the C-terminus containing an NLS and the putative regulatory domain but still able to bind an E-box in complex with MYC. Expression of ΔMAX enhances RAS-mediated transformation (Makela et al., 1992).

MAX may negatively-regulate MYC-responsive genes either through displacement of the MYC:MAX complex at the promoters with the transcriptionally inert MAX:MAX (Kato et al., 1992; Kretzner et al., 1992; Mukherjee et al., 1992), or with the transcription repressor complexes with MAD:MAX (Ayer et al., 1993; Hurlin et al., 1996; Prendergast et al., 1992;

Zervos et al., 1993) or MNT:MAX (Hurlin et al., 1997). The former is supported by the fact that the suppressive effect of MAX is dependent on its NLS (Makela et al., 1992).

1.6 Microphthalamia-associated transcription factor (MITF)

The *MITF* gene was identified as the locus associated with loss of pigmentation, reduced eye size, failure of secondary bone resorption, reduced number of mast cells, and early onset of deafness (Hodgkinson et al., 1993). Although, MITF was initially characterised as a bHLH-LZ-TF involved in melanocytic lineage specification and a potent TA of pigmentation genes (Hodgkinson et al., 1993), survival of pigmented cells and melanoma is also dependent on the expression of MITF (Widlund and Fisher, 2003). This is partly mediated through one of MITF's targets, the anti-apoptotic gene *BCL2* (McGill et al., 2002). *BCL2*$^{-/-}$ mice display postnatal loss of melanocytes (Veis et al., 1993).

MITF mutants

Loss of MITF function is associated with loss of viable melanocytes rather than defective pigmentation. Many mutant alleles of MITF-M, the melanocyte specific isoform, have been reported (Nakamura et al., 2002), several of which result in a gradual loss of pigmentation through impacts on survival and melanocytic-lineage-commitment during development (for review see Goding, 2000; Tachibana, 2000).

Mitf$^{-/-}$ mice exhibit loss of neural crest-derived melanocytes and failure of retinal pigmented epithelia (RPE) differentiation (Moore, 1995). Mutation of the *Mitf* gene in mice resulted in reduced melanocyte precursor cells (Hornyak et al., 2001) and failure of the RPE to produce pigment. This was accompanied by hyperproliferation and neuroretinal transdifferentiation (Nakayama et al., 1998), implicating the role of MITF in maintaining the terminal differentiation status of the RPE. A number of mutations effecting DNA-binding and dimerization of MITF which affect the coat colour of mice have also been reported (Hemesath et al., 1994).

MITF and development

Melanocytes are derived from the non-pigmented melanoblast precursors in the neural crest, corresponding to embryonic day 10.5 (E 10.5). By E 12, they undergo rapid proliferation and migration to the limb buds and enter the epidermis at the level of the lateral trunk by E13/E14 (Mayer, 1973). Initial expression of the *MITF* gene and commitment to melanocytes from neural crest cell is independent of MITF (Opdecamp et al., 1997). Once committed, however, MITF is critically important for the maintenance and survival of melanocytes.

In general, *MITF*-positive cells increase in number during development before decreasing and disappearing at birth. Melanocytes can also be found in the choroid layer of the eye (Tachibana, 2000), the heart (Hodgkinson et al., 1993) and in the stria vascularis of the cochlea (Steel and Barkway, 1989). In post-natal mice, *MITF*-positive cells are primarily found in the hair follicle (Tachibana, 2000). In humans, melanocytes are also distributed in the epidermis where they continue to divide and produce melanin during the tanning response (Jimbow et al., 1975).

Transcription activation by MITF

MITF activates genes mainly through the N-terminal TAD (AD1), located between residues $114 - 135$ within exon-4 (Sato et al., 1997). This region mediates interaction with p300/CREB-binding protein (CBP) and is homologous to the E1A, but distinct from Myb and the CREB-interaction region within p300/CBP (Yavuzer et al., 1995). MITF also harbours a second TAD (AD2) at the C-terminus. AD2 is highly conserved and bears some similarity to AD1. The biological significance of AD2 is, however, subject to controversy as *MITF* in mutant mice deleted in AD1 functions as a dominant negative TF (Hallsson et al., 2000).

MITF preferentially binds to 5'-CAC[G/A]TG-3' E-boxes (Bentley et al., 1994) as homo- or heterodimers with other members of the MiT-family of bHLH-LZ-TFs, namely TFE3, TFEB and TFEC (Hemesath et al., 1994). However, MITF most likely functions as a homodimer

during development as TFE3 and TFEB are absent from the neural crest at the times when the effect of MITF mutation becomes apparent (Nakayama et al., 1998).

MITF regulates a number of pigmentation-related genes through binding to the E-box and M-box (5'-AGT-CATGTG-3') (Aksan and Goding, 1998) presented at the promoters and/or enhancers of the genes encoding the rate-limiting enzyme in melanin production, *tyrosinase* (*TYR)* (Ganss et al., 1994) and many other key pigmentation enzymes (Lowings et al., 1992; Yasumoto et al., 1997).

The M-box is also recognised by other bHLH-TFs including USF1 and USF2. In fact, the M-box present in a heterologous promoter will drive gene expression in non-melanocytic lineage cells (Lowings et al., 1992). Melanocytic lineage specificity is believed to be partly conferred by the basal promoter region. For example, the TATA-region of the *tyrosinase-related protein 1* (*TYRP1*) gene confers some degree of tissue-specificity, possibly through a melanocyte-specific factor that derepresses the gene in melanocytes (Yavuzer and Goding, 1994).

Regulation of MITF

Stable commitment to the melanocytic lineage requires expression of *MITF* whose promoter can be induced by a number of TFs including PAX3 (Watanabe et al., 1998), SOX10 (Lee et al., 2000), LEF/TCF (Dorsky et al., 2000) and CREB (Bertolotto et al., 1998a). As these factors are widely expressed within the neural crest cells, yet very few cells are committed to the melanocytic lineage, it was proposed that melanocytic lineage commitment is a stochastic event that only happens when signalling from all the positive regulators are able to overcome the threshold barrier for *MITF* gene activation (Goding 2000).

Regulation by cAMP and CREB

3'-5'-cyclic adenosine monophosphate (cAMP) induces MITF target genes containing M-boxes such as *TYR*, *TRP1* and *TRP2* (Bertolotto et al., 1996; Bertolotto et al., 1998b) through increased MITF binding to M-boxes as a result of cAMP-induced *MITF* gene

expression (Bertolotto et al., 1998b). cAMP may be induced physiologically through MC1R receptor (García-Borrón et al., 2005; Rouzaud et al., 2005) which activates adenylyl cyclase through G-protein signalling, resulting in elevated cAMP levels (Bertolotto et al., 1998b; Chhajlani and Wikberg, 1992). Elevated cAMP activates the *MITF* promoter through PKA mediates phosphorylation CRE-binding protein (CREB) which enable it to bind to cAMP response element (CRE) presented at the *MITF* promoter (Bertolotto et al., 1998a). Activation of *MITF-M* through CREB is restricted to melanocytes (Bertolotto et al., 1998a). CREB is activated through phosphorylation on S133 and this can be mediated by a number of signalling cascades. For example, the growth factors EGF and NGF signalling, via MAPK, activate $p90^{RSK}$ that phosphorylates CREB (Xing et al., 1996; Xing et al., 1998).

Regulation by c-Kit

During melanocyte development, the Kit receptor tyrosine kinase may be activated by a cytokine, Steel factor [SI, mouse homologue of human stem cell factor (SCF)]. Mutation of Kit or Kit-ligand (SI) results in partial or complete loss of neural crest-derived melanocytes (Lev et al., 1994). Kit signalling through phosphatidylinositol-4,5-bisphosphate 3-kinase (PI_3K) results in activation of pyruvate dehydrogenase kinase (PDK1) and Akt kinases which have strong anti-apoptotic properties. Signalling through the PI_3K-AKT axis results in inhibition of GSK3β and nuclear accumulation of β-CATENIN that results in *MITF* gene activation. However, PI_3K-independent signalling from Kit is necessary for the melanocyte development (Nishikawa et al., 1991).

Kit signalling through the classical MAPK pathway resulted in ERK phosphorylation of MITF at S73 (Hemesath et al., 1998) and is required for recruitment of p300/CBP (Price et al., 1998a). The significance of p300/CBP is shown by the observation that the adenovirus E1A that binds and inactivates p300 (Eckner et al., 1994) is able to suppress expression of pigmentation-related genes (Yavuzer et al., 1995). The p300/CBP co-activator complex functions through recruitment of the RNAPII holo-enzyme (Nakajima et al., 1997), intrinsic HAT activity (Ogryzko et al., 1996) and recruitment of other histone acetyl transferases

(HATs) such as p300/CBP-associated factor (P/CAF) (Yang et al., 1996). ERK also activates ribosomal S6 protein kinase (RSK), $p90^{RSK}$ that phosphorylates MITF at $S409$, priming it for ubiquitination and proteasome-mediated degradation by ubiquitin/SUMO-conjugating enzyme UBC9 (Wu et al., 2000b). $S73$ −phosphorylation may also mediate MITF degradation, at least in the context of overexpressed UBC9 (Xu et al., 2000).

MITF and cancer

Melanoma is one of the most invasive types of cancer with poor prognosis owing to its propensity for vertical growth in the epidermis with deep invasion and early metastasis. Melanomas often carry inactivating mutations at the cyclin-dependent kinase inhibitor (CKI) *INK4a/ARF* locus that encodes p16 and p19. Loss of CKIs results in hyper-phosphorylated $p105$ −Rb and cell-cycle progression. A number of receptor tyrosine kinases not normally found in melanocytes may be found in melanoma (Chin et al., 1998). Stabilizing mutations of β-CATENIN are also found in $\leq 30\%$ of melanoma cell lines (Rubinfeld et al., 1997). As discuss above, signalling through MAPK and β-CATENIN positively regulates MITF expression and activity. MITF, in turn, regulates several downstream targets important for the survival of melanoma cells such as *BCL2* and *TBX2*. The expression of *BCL2* was found to correlate with *MITF* in many human primary melanoma (McGill et al., 2002). TBX2 is an anti-senescence factor which suppresses *CDKN2A/ARF* (Jacobs et al., 2000) and thus inhibits the p53 response (Sherr, 2001). Taken together the level of MITF plays a central role in fate, phenotype and survival of melanoma as discussed in Hoek and Goding, 2010.

1.7 Hypoxia inducible factor (HIF)

Adaptation to hypoxic stress during normal development as well as in pathologic states is mediated through hypoxia-inducible factors (HIF). HIF is a bHLH-PAS (Per-Arnt/AhR-Sim) TF that is regulated by cellular oxygen tension (Wang et al., 1995). The members of the HIF family include hypoxia inducible factor 1α (HIF-1α), endothelial PAS domain protein 1 (EPAS-1/HIF-2α), HIF-3α (aka IPAS, inhibitory PAS domain protein) and their common

heterodimeric partner aryl hydrocarbon receptor (AhR) nuclear translocator (ARNT/HIF1-β) that exists in three paralogues (ARNT1, ARNT2 and ARNT3) (Figure 1.8). HIF functions as a heterodimer of an O_2-labile-α-subunit (HIF-1α, HIF-2α and HIF-3α) and a stable β-subunit under low O_2 (hypoxic condition) and bind to the HIF-responsive element (HRE, 5'-TACGTG-3'). The α-subunits are rapidly degraded under normoxia (Semenza, 2001).

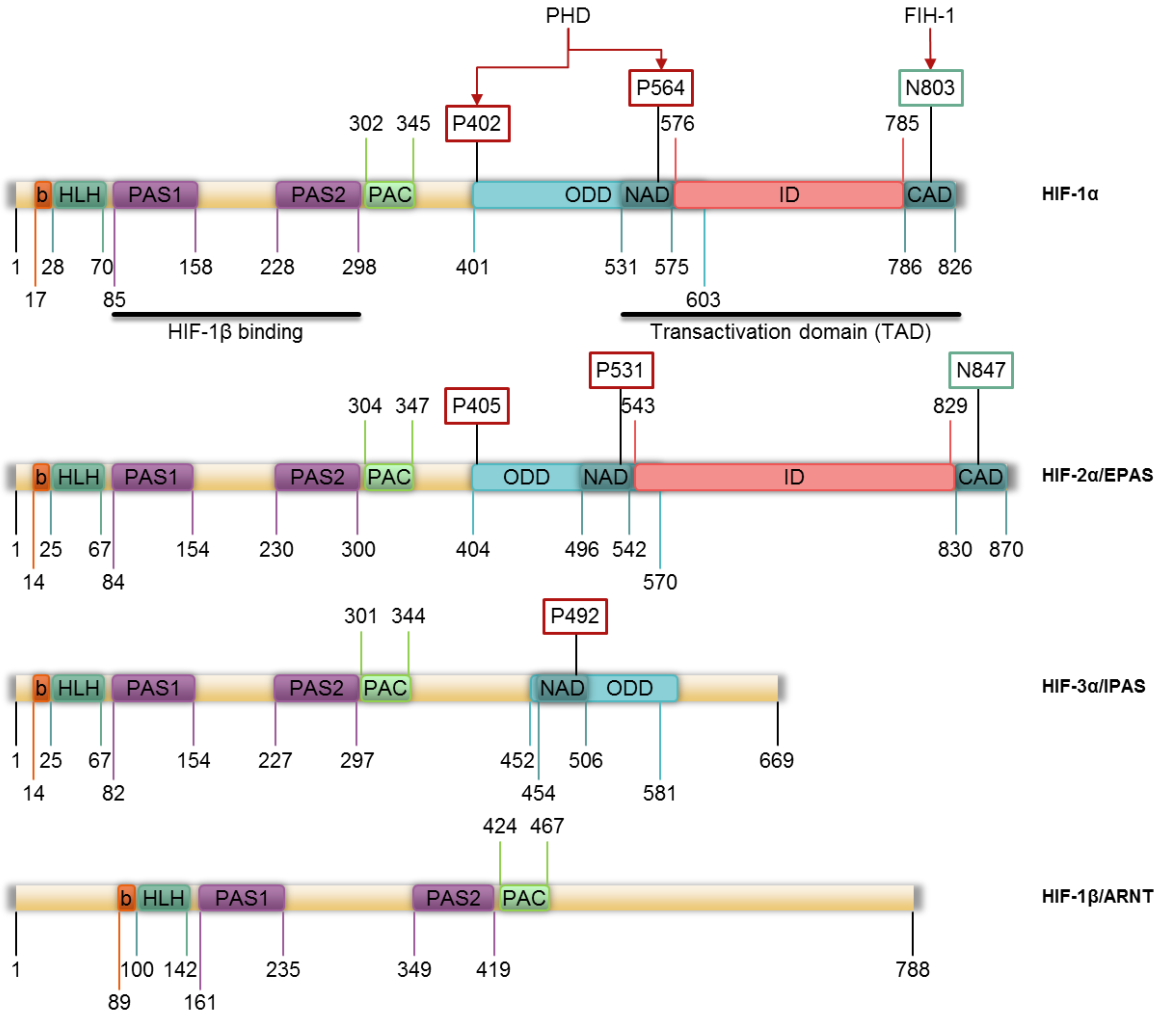


Figure 1.8 Diagrammatic representation of hypoxia inducible factor (HIF) family members. The α-subunits include HIF-1α, endothelial PAS domain protein 1 (EPAS-1/HIF-2α) and inhibitory PAS domain protein (IPAS/ HIF-3α). The three paralogues of the β-subunit is presented here as aryl hydrocarbon receptor (AhR) nuclear translocator (ARNT/HIF1-β). Amino-acid numbers at the border of each domain are labelled. To avoid ambiguity in the labelling of residue numbers, colour lines matching the colour of the domains are used. Common among all HIFs are the basic (b) DNA-binding domain and the HLH-PAS domain for dimerization followed by the PAC motif. HIF-α subunit also contains oxygen degradation domain where the two key proline residues, under high O_2 is hydroxylate by PHD, allowing it to be targeted by von Hippel-Lindau (VHL), the substrate recognition unit of E3 ubiquitin ligase for poly-ubiquitination and proteosomal degradation. The two transactivation domains (TADs) are separated by the inhibitory domain (ID) in HIF-1α and HIF-2α. The CAD also harbours another hydroxylation site mediated by FIH, resulting in inhibition of binding to the p300/CBP co-activator complex. The proline residues hydroxylated by prolyl hydroxylase domain-containing protein (PHD) are labelled with red boxes while asparagine residues hydroxylated by factor inhibiting HIF-1 (FIH-1) are labelled with green box. Abbreviations: b, basic; HLH, Helix-Loop-Helix; PAS, Per-Arnt/AhR-Sim; PAC, PAS-associated, C-terminal; ODD, oxygen degradation domain; NAD, N-terminal transactivation domain; ID, inhibitory domain; CAD, C-terminal transactivation domain.

HIF-1α and HIF-2α are structurally similar in the bHLH-region but differ in their TAD. The role of HIF-3α is less well understood, although it has been shown to act as a dominant negative of HIF-1α (Makino et al., 2001). The two TADs of HIF-1α and HIF-2α are located in the C-terminus and are referred to as the N-terminal TAD (NAD) and C-terminal TAD (CAD). Both NAD and CAD are required for optimal transcription activation of hypoxic genes (see Transcription activation by HIF below). The TADs, particularly NAD, were shown to be essential for target gene specificity by interacting with additional transcription cofactors (Dayan et al., 2006; Hu et al., 2007).

Biological function of HIF

The role of HIF is to orchestrate the cellular response to hypoxia through induction of glycolysis, erythropoiesis and angiogenesis (Gordan and Simon, 2007). This is achieved through binding to HREs found in genes involved in angiogenesis, metabolism, cell growth and survival, for examples, genes encoding glycolytic enzymes, *vascular endothelial growth factor* (*VEGF*) (Hu et al., 2003) and *CCND1* (Baba et al., 2003). HIF's role in vasculogenesis is demonstrated by embryonic lethality at E10.5 due to a complete failure in placental vascularization in *Arnt*$^{-/-}$ embryos (Adelman et al., 2000; Kozak et al., 1997).

The two most studied HIF-α subunits are not identical in their expression profile and preferred target genes (Hu et al., 2003; Wang et al., 2005). HIF-1α is expressed ubiquitously at a low level while HIF-2α is limited to bone marrow macrophage, endothelium, kidney epithelial cells, liver parenchyma, cardiac myocytes, lung mesenchymal cells, uterine decidual cells, pancreatic parenchymal cells and small intestine (Ema et al., 1997; Tian et al., 1997; Wiesener et al., 2003) although it has been observed in tumours of other tissues (Semenza, 2003). The more restricted expression of HIF-2α is supported by the observation that *Hif-1α*$^{-/-}$ ES and mouse embryo fibroblasts (MEFs) abrogate expression of all hypoxia inducible gene expression even though microarray analysis revealed a plethora of hypoxia-inducible genes regulated by HIF-2α (Hu et al., 2003).

HIF-1α can induce apoptosis and impair cell proliferation in response to hypoxia and hypoglycaemia in ESC through HIF-1α regulation of many genes regulating the cell-cycle including *p53*, *p21*, *p27* and *BCL-2*. HIF-1α mediates cell cycle arrest under mild hypoxia through p27-mediated inhibition of CDK2 (Gardner et al., 2001) leading to hypo-phosphorylated Rb (Goda et al., 2003) a key regulatory event of the G_1/S transition. However, cell-cycle arrest that occurs at very low oxygen tensions (0.01% O_2) is independent of HIF-1α-mediated induction of $p21^{Clp1}$ and $p27^{Klp1}$. p21 and p27, nonetheless, appear to be involved in preventing premature re-entry into the cell cycle following re-oxygenation in an additive manner (Green et al., 2001). Intriguingly, induction of p27 under severe hypoxia is independent of the HIF-1α-dependent region of the promoter (Gardner et al., 2001). Although hypoxia can induce cell-cycle arrest through induction of HIF-1α-independent genes, HIF-1α knockout tumours resulted in increased tumour growth due to decreased apoptosis and increased stress-induced proliferation (Carmeliet et al., 1998).

Transcription activation by HIF

The C-terminus of HIF-1α harbours two TADs located between amino-acids $531-826$, both of which interacts with a number of transcription coactivators in an oxygen dependent manner. Some of the reported ones include: p300/CBP and HAT SRC-1 and transcription intermediary factor 2. Initial binding of p300/CBP helps the recruitment of accessory coactivators that, together, stabilise the RNAPII pre-initiation complex and modify nucleosomes through HAT activity (Carrero et al., 2000; Ruas et al., 2005). At limiting concentrations, SRC-1 and CBP can synergise to activate target genes (Carrero et al., 2000). The minimum TADs were located to amino-acids $531-575$ and $786-826$ separated by inhibitory domain which is oxygen responsive (Jiang et al., 1997) (Figure 1.8). Similar domain structure and regulation was reported for HIF-2α. The two TADs are located at amino acids $517-682$ and $828-870$ of HIF-2α, separated by repressive sequences regulated by O_2 availability (Figure 1.8). The inducibility of HIF-2α was, however, reported to be less than that of HIF-1α (O'Rourke et al., 1999).

HIF-1α's targets have been categorized into PHD- (driven by NAD) and PHD/FIH-dependent cohorts (driven by NAD and CAD) based on the observation that both TADs are required for full hypoxic gene expression (Dayan et al., 2006). Further quantitative analysis revealed FIH roles in determining whether a given gene should be induced mildly or strongly under hypoxia (Dayan et al., 2009). An example of a PHD-dependent gene is *Bcl-2/E1B 19kDa interacting protein 3* (*BNIP3*) that becomes fully activated in response to a marginal drop in O_2 and is largely insensitive to FIH1 (Hu et al., 2011).

Regulation of HIF-α proteins

To enable instantaneous induction of HIF-α in response to hypoxia (Jewell et al., 2001), HIF-α is constitutively transcribed and translated in the cells but under normoxia, HIF-1α has a half-life of $< 5'$ (Huang et al., 1996).

O_2 mediates hydroxylation of the two prolines located in the oxygen-dependent degradation domain (ODD) by the highly conserved O_2 and Fe^{2+} dependent prolyl hydroxylase domain-containing protein (PHD) -1, -2 and -3 (Bruick and McKnight, 2001; Epstein et al., 2001). Proline hydroxylation results in recruitment of the E3 ubiquitin ligase-containing von Hippel-Lindau tumour-suppressor protein (VHL) (Ivan et al., 2001; Jaakkola et al., 2001) and rapid degradation by 26S proteasome (Cockman et al., 2000) such that almost no HIF- α protein accumulates (Figure 1.9A).

HIF is also regulated through asparagine hydroxylation in the CAD (Lando et al., 2002) by an Fe^{2+} and 2 −oxoglutarate-dependent dioxygenase (termed factor inhibiting HIF-1, FIH-1) that interferes with the ability of HIF-α to recruit p300/CBP necessary for its full transcription activity (Kung et al., 2000) (Figure 1.9B). The interactions between CAD and p300 (Gu et al., 2001)/CBP (Ema et al., 1999) also appears to be under additional regulation at the cysteine residue in a manner dependent on redox regulatory protein, Ref−1 and thioredoxin, at least for HIF-1α. Nuclear translocation of thioredoxin is dependent on reduced O_2 concentration (Ema et al., 1999).

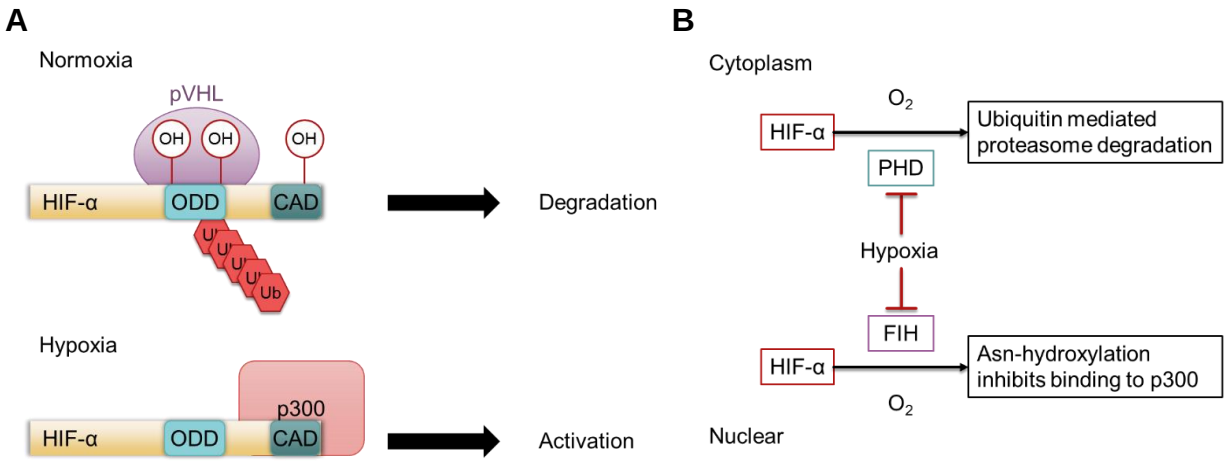


Figure 1.9 Regulation of HIF-α **A)** Schematic diagram of HIF-α regulation under normoxia and hypoxia. HIF-α is rapidly degraded in normoxia by von Hippel-Lindau (VHL) binding to the two hydroxylated proline within the oxygen degradation domain (ODD). Under hypoxia, the non hydroxylated arginine residue located in the C-terminal transactivation domain allow binding to p300 coactivator leading to gene activation. **B)** A flow **d**iagram showing the two level of regulation of HIF-α during high and low O_2 tension. HIF-α is rapidly degraded under high O_2 in the cytoplasm via PHD-mediated hydroxylation. An additional layer of regulation occurs through asparagine hydroxylation in the nucleus which serves to regulate HIF-α interaction with p300 coactivator. This interaction is necessary for activation of PHD/FIH-dependent genes cohort.

An increasing number of O_2-independent mechanisms of HIF-1α regulation have been reported (Koh et al., 2008; Lei et al., 2009; Liu et al., 2007; Luo et al., 2010) and represent potential mechanisms which might induce HIF exchange with other bHLH-TFs in response to different induction cues. However, we shall limit our study to regulation by O_2 as this is the most extensively studied condition with reported cross-talk with USF1.

HIF and cancer

The response to hypoxic condition, frequently found in solid tumours, has important effects on the natural history of the disease, partly through HIF-mediated stimulation of genes directly involved in tumour growth, angiogenesis, invasion and metastasis (Pugh and Ratcliffe, 2003; Semenza, 2003). HIF-α up-regulation is often found in glioma cells that harbour extensive hypoxic domains (Acker et al., 2005). HIF-α has been associated with poor prognosis in a broad range of sporadic cancers (Semenza, 2003).

Effects of loss of HIF-α in cancer

Despite the observed frequent loss of HIF-1α, loss of HIF-1α alone fails to confer consistent growth advantages in ESC-derived teratomas. Although HIF-1α inhibits *Vhl*$^{-/-}$ Ras-transformed MEF subcutaneous allograft growth (Mack et al., 2005), *Hif-1α* knock-outs

exhibit smaller tumour mass (Ryan et al., 1998), probably relating to HIF-1α's role in vascularisation.

HIF-2α is able to mediate neoplastic progression and enhanced MYC-mediated transformation of primary MEFs (Gordan et al., 2007a). The superior ability of HIF-2α to promote tumour growth was also demonstrated through targeted replacement of HIF-1α with HIF-2α in primary murine ESCs. The teratomas derived from HIF-2α knock-in ESCs were up to 4 –times the mass, had higher microvessel density and elevated expression of *VEGF*, *TGF-α* and *CCND1* than controls (Covello et al., 2005).

HIF's transcriptional targets and cancer

In addition to HIF regulation of erythropoiesis, angiogenesis and glycolytic metabolism, many other HIF's targets can also influence cancer progression. For instance, HIF-1α inhibits aerobic metabolism through induction of the *PDK1* gene. This was suggested to free the limited supply of oxygen for other cellular processes and minimize accumulation of reactive oxygen species (ROS) (Kim et al., 2006; Papandreou et al., 2006).

HIF also targets a number of genes important in metastasis. This ranges from the reported up-regulation of the chemokine receptor *CXCR4* (Staller et al., 2003), the *matrix metalloproteinases* (*MMPs*) *2* and *9* (Leufgen et al., 2005) which degrade extracellular matrix, and *lysyl oxidase* which alters extracellular matrix components such as elastin and collagen (Erler et al., 2006).

Cross-talk between HIF and MYC family proteins

The intricate concerted actions of HIF and MYC was suggested to fine tune the adaptive responses of cells growing in hypoxic environment (reviewed in Gordan et al., 2007b). Although both HIF and MYC share a number of common targets and a direct competition for E-boxes by HIF-1α and MYC has also been reported (Mazure et al., 2002), HIF generally regulates additional targets that can act to cap the detrimental effect of MYC regulated genes on overall cell viability under hypoxia. In solid tumours where O_2 levels are likely to

oscillate (Dewhirst 2007), transient induction of HIF-α inhibits anabolic pathways and suppress ROS production without long-term detrimental effect on mitochondrial mass, metabolic enzyme expression or translation machinery components.

MYC and HIF display antagonistic effects on carbon utilization and translation initiation (Gordan et al., 2007b). This is partly achieved through differential target gene regulation. While both positively regulate most of the glycolytic enzymes, only MYC is able to support growth and anabolic metabolism. This metabolic shift induced by HIF was suggested to be dominant over MYC, as MYC-up-regulated anabolic pathways are substrate-limited and should help tumour cells to survive till reoxygenation (Gordan et al., 2007b). HIF also acts epistatically to MYC's positive effects on translation through limiting availability of capped mRNA substrates for translation by the machinery produced by MYC (Coller et al., 2000).

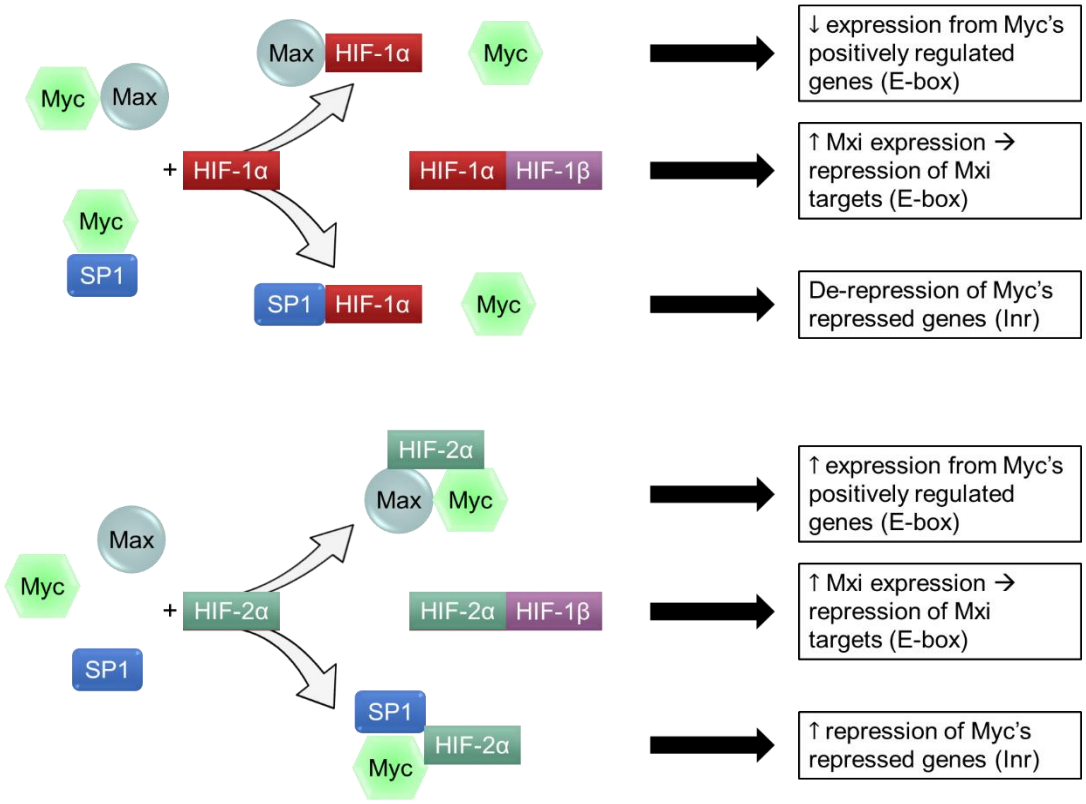


Figure 1.10 Diagrammatic representation of HIF-α cross-talk with the MYC family of transcription factor. HIF-1α and HIF-2α is stabilised following exposure to low O_2 tension. The stabilised HIF-α can form a functional heterodimer with HIF-1β to activate Mxi that specifically inhibits a subset of Myc activated target E-boxes. HIF-1α antagonise Myc function by binding to SP1 and Max, disrupting their interaction with Myc. This leads to decrease Myc-Max heterodimer, hence decrease expression from Myc's target E-box. Similarly disrupting Myc interaction with SP1 and Miz (not shown) results in derepression of Myc's repressed genes containing initiator element (Inr). The opposite situation occurs with HIF-2α as HIF-2α stabilised the interaction between Myc and Max and SP1.

HIF-1α, independent of HIF-1β, competes with MYC for SP1 complexes and MAX (Koshiji et al., 2005) at both active (To et al., 2006) and repressed genes (Koshiji et al., 2004; Mack et al., 2005; Gordan et al., 2007a), (Figure 1.10 Top). The reduction in DNA-binding-competent MYC allows factor switching with other bHLH-TFs. HIF-2α, on the other hand, augments MYC's activity through stabilisation of MYC complexes with SP1, MIZ1 and MAX (Gordan et al., 2007a) (Figure 1.10, Bottom) and promote tumour progression (Gordan et al., 2008).

HIF may also induce MYC:MAX exchange with MXI:MAX through direct induction of *MXI* gene expression which results in suppression of a specific subset of MYC's target genes (Figure 1.10) (Corn et al., 2005; Zhang et al., 2007). As MXI only binds to a subset of MYC's targets involved in energy-intensive process and not proliferation (O'Hagan et al., 2000), MYC:MAX factor exchange with MXI:MAX likely promotes tumour cell survival under hypoxia.

1.8 Upstream stimulatory factors (USF)

USF was first identified as major late promoter TF (MLTF) which bound cooperatively with $TF_{II}D$ to activate transcription from the adenovirus major late (*AdML*) promoter (Carthew et al., 1985; Kirschbaum et al., 1992; Sawadogo and Roeder, 1985). Efficient transcription from the *AdML* promoter requires USF dimers (Gregor et al., 1990).

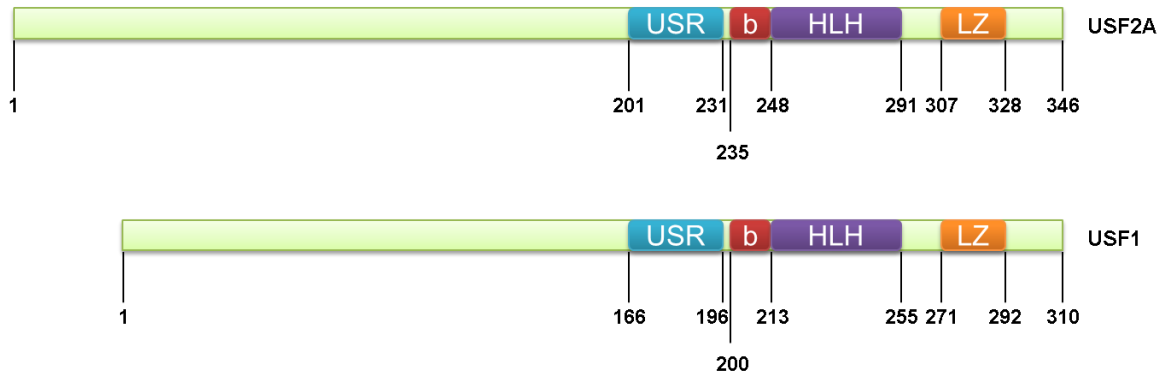


Figure 1.11 Diagrammatic representation of the major isoforms of USF2 and USF1. Amino-acid numbers at the border of each domain are labelled. Abbreviation: USR, USF specific-region; b, basic domain; HLH, Helix-loop-Helix; LZ, leucine-zipper

USF is transcribed from two distinct genes that encode the USF1 and USF2 proteins of $43-$ and $44-\mathrm{kDa}$, respectively (Figure 1.11). Both USFs exist in alternatively-spliced isoforms (Sawadogo, 1988; Sawadogo et al., 1988) and although ubiquitously expressed, different ratios of USF homo- and hetero-dimers are found in different cells with heterodimers being the most abundant form and USF2 homodimers being the most scarce (Sirito et al., 1994; Viollet et al., 1996). USF binds DNA as either homo- or hetero-dimers (Sirito et al., 1992). Homo- and hetero-dimerization requires the highly conserved C-terminus from the basic region onward (Sirito et al., 1992). While the LZ is not strictly necessary for USF1 DNA-binding activity (Gregor et al., 1990), it appears to be essential for USF2.

USF and development

USF1 and USF2, despite being highly ubiquitous postnatally, are differentially expressed during embryonic development, highlighting their differential and important roles during development (Fujimi and Aruga, 2008). In *Xenopus laevis,* expression of both genes initially comes from maternal transcripts. USF1 associates with $TF_{III}A$ and plays important roles in activation of late-stage oocyte genes (Kaulen et al., 1991).

USF1$^{-/-}$ or *USF2*$^{-/-}$ accompanied by heterozygous deletion of *USF2* or *USF1*, respectively, results in early embryonic lethality in mice (Sirito et al., 1998). Although they can partially compensate for each other, they also have non-redundant roles. While *USF1*$^{-/-}$ mice are viable and fertile with occasional epileptic seizures and enhanced barbering behaviour, *USF2*$^{-/-}$ mice are smaller at birth and display proportionate body features throughout postnatal development (Sirito et al., 1998).

The roles of USF during differentiation is supported by the reported induction of USF by thrombopoietin during platelet production (Kirito et al., 2003) and during differentiation of mast cells (Ma et al., 2008), osteoclasts (Zhang et al., 2008) and Sertoli cells (Wood and Walker, 2009). In murine erythroleukemia, Usf was subject to degradation by the intracellular non-lysosomal Ca^{2+}-dependent protease m-calpain in undifferentiated but not

differentiated cells (Lin et al., 2009). Cleavage by m-calpain leaves the dimerization domain intact which can act as a dominant negative protein (Watt and Molloy, 1993), suggesting that, at least in the context of erythroleukemia, USF's binding and transcription are required in the differentiated state.

Regulation of USF

p38 and USF

p38 SAPK may act as a hub which integrates a number of signals leading to USF activation. For example, endothelin-1 (Clerk et al., 1998) and Alpha-melanocyte stimulating hormone (α-MSH) (Smalley and Eisen, 2000) secreted from the surrounding keratinocytes to the melanocytes during the tanning response act via the p38 pathway. Glucose and hyper-osmolality also increase p38 kinase activity via PKCδ-dependent and PKC-independent pathways, respectively, in a time-dependent manner (Igarashi et al., 1999).

USF1 may be activated through phosphorylation at T153 by p38 in response to stress (Galibert et al., 2001). As p38 may be activated in response to oxidative stress, especially relevant in hypoxic tumours, USF1 acetylation at K199, that is dependent on T153-Ⓟ by p38 which inhibits DNA binding (Corre et al., 2009), may explain the observed inactivation of USF1 in some breast cancer cell lines (Ismail et al., 1999).

Other regulation of USF

USF1 is regulated in a redox-dependent manner at the two highly conserved cysteine residues located within the dimerization interface of the HLH-domain. Under non-reducing conditions, USF1 DNA-binding is significantly reduced (Pognonec et al., 1992).

USF1 is also subject to phosphorylation by cyclin A1- and cyclin B1-$p34^{cdc2}$ (Cheung et al., 1999), and PKA (Xiao et al., 2002), resulting in enhanced DNA-binding and transcription activation potential.

The consequence of regulable USF DNA binding and transcription activity is that USF can act at multiple stages and have multiple roles during a promoter education process in a context-specific manner. For example, while USF has been implicated in basal transcription at promoters where other TAs have been identified, USF can also serve as major TAs in response to, say p38-signalling. A number of direct targets of activated USF have been reported many of which are discussed in Corre and Galibert, 2005.

USF and transcription control

The TAD of USF1 has been mapped to the two sub-regions at the N-terminus, both of which are required for maximal gene activation (Kirschbaum et al., 1992). This N-terminal TAD, unlike the bHLH-domain, is not very well conserved evolutionarily. This raises the possibility that USF may not have initially evolved to be a strong TA. Instead, USF's major role may be to facilitate recruitment of co-activators and/or gene priming through appropriate chromatin remodelling. This is supported by the observation that USF1 does not activate transcription in a highly purified system (Kirschbaum et al., 1992) and that, although constantly bound to the *AdML* promoter, USF does not constitutively drive its expression (Leong et al., 1988). Moreover, it was noted that transcription activation by a USF1 homodimer is often poor and is affected by the distance to the TSS (Desbarats et al., 1996; Vallet et al., 1998). This led to a suggestion that USF1 may act as a transcription repressor by competing with the more potent TFs for the promoter (Lun et al., 1997; Takahashi et al., 1997). Other studies in a number of cell lines, however, demonstrated strong transcription activity by USF1 and a much more modest one by USF2 and vice versa in human mammary epithelial cells (Ismail et al., 1999). In any case, attenuation of transcription activation potential seems to happen at gene activation and not DNA binding, as fusion of USF2 to the activation domain of VP-16 restores transcription activation (Ismail et al., 1999). In fact, depending on the context, USF may display more potent transcription activation potential than MYC as has been shown for the human polymeric immunoglobulin receptor gene (Bruno et al., 2004).

Transcription activation by USF

Initially identified as MLTF, USF1 was shown to interact with the pyrimidine-rich initiator (Inr) element (consensus YYAYTCYY) of the *AdML* promoter and mutation of the Inr-element reduces transcription from the promoter (Du et al., 1993). In HeLa cells, the USR is necessary and sufficient for activation of promoters containing Inr-elements in addition to a TATA-box (Luo and Sawadogo, 1996b). However, the USR is not required for activation at other loci such as *IGF2R* (Szentirmay et al., 2003). This can be partially accounted for by the N-terminal TAD, located in Exon-5 which is required for activation of promoters lacking an Inr-sequence (Qyang et al., 1999). Interaction of USR with co-activators was suggested to induce a conformational change that unmasks the TAD located in Exon-5 (Luo and Sawadogo, 1996b).

USF also directly interacts with a number of HATs including P/CAF, SRC-1 and p300/CBP as well as histone methyltransferases (HMTs) including the SET1-complex containing H3K4 methyltransferase and the H4R3-specific methyltransferase PRMT1 (Huang et al., 2007). p300/CBP is required to potentiate transcription activation of USF at a number of promoters (Breen and Jordan, 2000; Goueli and Janknecht, 2003). However, this does not appear to be a generic mechanism as p300 is dispensable for expression from other promoters (Kingsley-Kallesen et al., 1999; Sayasith et al., 2004). This indicates that the involvement of p300 might be promoter-specific.

Cooperative binding of USF with cue-specific and tissue-specific TFs

USF, as a ubiquitous TF, may be involved in setting the basal transcription level. Cue-specific induction may, however, requires cooperative binding with cue-specific, inducible factors to confer maximal gene induction. For example, USF binding is required for basal expression of *FSHβ* (Ciccone et al., 2008; Harris and Mellon, 1998) in response to gonadotropin-releasing hormone. Induction, of *FSHβ*, however, requires CREB. The CRE half-site ($-73/-69$) overlaps with the USF E-box ($-70/-65$) and was suggested to bind cooperatively with USF (Ciccone et al., 2008).

USF also bind cooperatively with metabolite-induced TFs such as SREBP and ChREBP (Kaytor et al., 1997; Lefrancois-Martinez et al., 1995; Kaytor et al., 1997; Wang and Wollheim, 2002). Direct interaction between USF bHLH-domain and the SREBP C-terminal domain (Griffin et al., 2007), is necessary for regulation of many lipogenic promoters carrying tandem E-box and SRE in response to insulin (Sul et al., 2000). (For reviews on SREBP in metabolism see Osborne, 2000; Shao and Espenshade, 2012.)

Cooperative binding may also be important for the binding of USF to low affinity sites as shown for USF acting in concert with tissue specific-TFs like Ap2, Ikaros and GATA (Payton et al., 2005). Cooperative binding through DNA-looping was also proposed to mediate long-ranged interaction between USF and hepatic nuclear factor (HNF) 4 (Rada-Iglesias et al., 2005). USF binds cooperatively with HNF-4 to regulate some glucose responsive (Bergot et al., 1992) and plasma lipoprotein metabolism genes (Pastier et al., 2002; Ribeiro et al., 1999).

Transcription repression by USF

USF serves as a repressor for a number of genes, for example the *ABCA1* (Yang et al., 2002), *MyoD* (Lun et al., 1997), *GCLC* (Cheng et al., 2005) and *CYP19A1* (Takahashi and Kamataki, 2001; Takahashi et al., 1997) genes. USF was proposed to repress the human follicle-stimulating hormone receptor gene (*FSHR*) by recruiting a repressive complex through USF1 binding at the *Inr* and E-box 3 (5'-CACATG-3') and USF2 at E-box 3 (Putowski et al., 2004). It is also possible that USF binding at E-box 3 located proximal downstream of the TSS stearically blocks recruitment of the basal transcription machinery.

PTMs may also serve as switches that change USF from TAs to repressors. The polyphenol quercetin was shown to induce phosphorylation of USF2 and invoke USF2 binding and repression of a cardiovascular disease risk factor, *PAI-1* in endothelial cells (Olave et al., 2010). Thus, USF may repress genes through competition with other activators for E-boxes (Takahashi et al., 1997), or interaction with corepressor complexes (Cvekl et al., 1994).

USF may switch between an activator and a repressor to regulate gene expression

In the fasting state, USF represses lipogenic genes through recruitment of HDAC9, a class II HDAC transcription corepressor, to the LZ of USF1 that results in USF1 deacetylation and inactivation. Upon feeding, insulin activates the PI_3K-PKB/Akt cascade resulting in dephosphorylation and activation of DNA-PK by PP1 that, in turn, phosphorylates USF1 at $S262$. This enables USF1 to recruit P/CAF to the bHLH-domain for its acetylation at $K237$, induce double strands DNA (dsDNA) breaks and activate the *FAS* gene (Wong et al., 2009). A USF-mediated transient dsDNA break is required for local changes in chromatin architecture necessary for transcription of lipogenic genes (Ju et al., 2006).

Insulin, via atypical PKCζ/λ, phosphorylates Brg1/Brm-associated factor (BAF) 60c, a modulator component of the mammalian SWI/SNF chromatin remodelling complex. This results in nuclear translocation and association with acetylated USF1 through the bHLH-domain for chromatin remodelling to activate lipogenic genes (Wang et al., 2013). Increased chromatin accessibility around the first two nucleosomes from the TSS was observed within 10 minutes following insulin treatment (Wang et al., 2013). This is accompanied by increased RNAPII, H3K14Ac and H3S10P.

USF and cancer

Ectopic expression of USF2 can inhibit MYC + RAS transformation of REFs by almost 40 –fold and to a more modest 5 –fold by USF1 (Choe et al., 2005). This antagonism is dependent on USF's ability to activate genes (Choe et al., 2005) and not simply through competition for MYC's target genes (Luo and Sawadogo, 1996a). While the effect of USF1 is specific to MYC, USF2 was also able to inhibit E1A-mediated transformation among other oncogenes, and strongly suppress colony formation in HeLa cells, implicating USF2's ability to suppress growth in general (Luo and Sawadogo, 1996a). A more recent report, however, indicated that USF2 does not confer any growth disadvantage in MYC + RAS transformed REFs on plastic dishes, but instead, specifically attenuates the transformed phenotype such as anchorage-independent growth and tumourigenicity in nude mice (Choe et al., 2005).

It is interesting to note that USF derived from non-cancer cell lines, even when overexpressed are strong TAs (Ismail et al., 1999). In many cancer cell lines, both endogenous and exogenous USF fail to activate reporter constructs and supress proliferation even though they retain their DNA-binding ability (Ismail et al., 1999; Qyang et al., 1999). This supports the notion that USF target genes are important for USF's anti-proliferative properties. In thyroid follicular cells USF induces a delay in the G2/M phase of the cell cycle and thus inhibits cell proliferation. USF suppression of cell proliferation is accompanied by a decrease in cyclin B1-CDK-2 and an increase in p27 expression that overrides the thyroid stimulating hormone-induced cAMP response (Jung et al., 2007).

In osteosarcoma Saos-2 cells USF's anti-proliferative property was attributed to the inactivity of the USR-domain. Similarly, the USR is required for inhibition of colony formation in MYC + RAS REFs (Choe et al., 2005). This may imply the existence of a USR-specific co-activator absent in many cancer cell lines. Of note, the USR is sufficient for gene activation from Inr-element containing promoters (Luo and Sawadogo, 1996b). As the USR is generic to both USF1 and USF2, this cannot account for all USF inactivity because selective inactivation of USF2 over USF1 has been reported in some breast cancer cell lines (Ismail et al., 1999). This implies the existence of a USF2-specific anti-proliferative effect(s).

Inactivated USF2 presumably confers a growth advantage through competition and interfering with MYC's target involved in apoptosis (for review see Prendergast, 1999) or through down-regulation of tumour suppressor genes which respond specifically to USF, such as *IGF2R* (Souza et al., 1999; Szentirmay et al., 2003), *p53* (Reisman and Rotter, 1993) and *APC* (Jaiswal and Narayan, 2001), or DNA-repair genes such as *BRCA2* (Davis et al., 1999; Wu et al., 2000a).

1.9 Factor exchange and promoter education in mammals

Mammalian promoters often contain TFBSs that can be bound by a number of closely related TFs that cannot bind to the same DNA-elements simultaneously. It is possible that each TF responds to distinct cues, allowing the gene to be activated in response to multiple stimuli. An alternative possibility, as discussed in section 1.3 is that each TF binds sequentially, contributing to the pre-initiation events leading up to transcription.

Evidence and consequence of factor exchange in mammals

MAD:MAX complexes switching with the MYC:MAX complex during differentiation (Ayer and Eisenman, 1993; Hurlin et al., 1994) may represent a paradigm for TF exchange at a shared E-box through direct competition. In the case of MAD and MYC, this factor exchange either turns the gene off or on respectively. As MAX protein is in vast excess over MYC, it

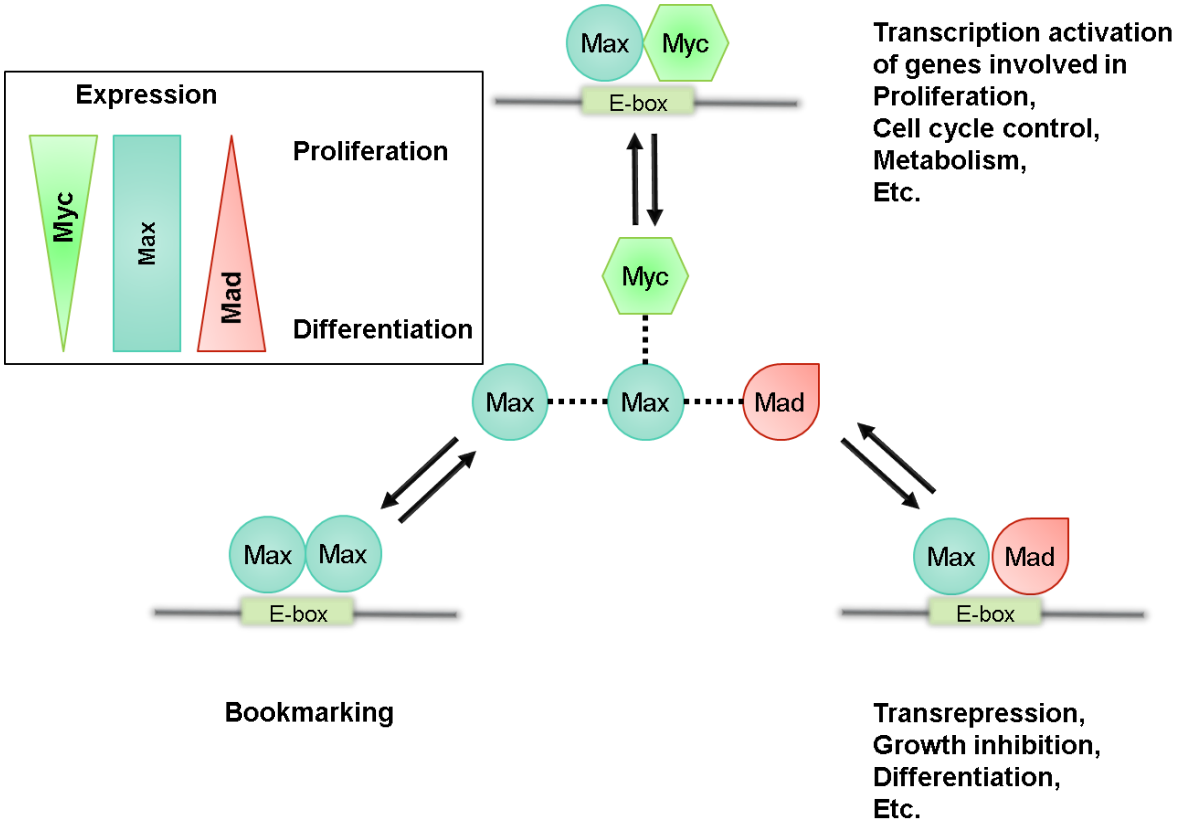


Figure 1.12 Factor exchange by MYC/MAX/MAD family of bHLH-LZ-TF. The protein level of MYC and MAD family members are highly dynamic while their dimerization partner, MAX is ubiquitous and stable. The expression of MYC and MAD are usually inversely correlated. MAX homodimer is unstable and hence, the preferred DNA binding species is in complex with MYC or MAD. The complex of MYC:MAX and MAD:MAX may exchange at promoters through direct competition for the binding sites with opposing effects on transcription output. As MAX:MAX homodimers are unstable and transcriptionally inert, promoter occupancy by a MAX:MAX homodimer is thought to either directly compete away other MAX heterodimers or to bookmark the gene for subsequent exchanges with other MAX heterodimer when they become avaliable.

was suggested that MYC:MAX complex switching with other MAX repressive complexes occurs through competition for target E-boxes (Figure 1.12) (Hurlin et al., 1997).

It is known that many tissues-specific TFs are also bHLH-TFs that demonstrate overlapping E-box specificity with MYC and MAX, for example MITF and MyoD. It is possible that MYC:MAX-heterodimers exchange with tissue specific bHLH-TFs may also occur as part of gene expression profile switching following differentiation.

A slightly different scenario, analogous to that explained in yeast, is factor exchange as part of the promoter education process. For instance, it was shown through *in vivo* cross-linking that both MYC and USF1 bind to the *CAD* gene (Boyd et al., 1998). As it is physically impossible for two bHLH (LZ) TFs to occupy the same E-box at the same instance, MYC must exchange with USF1. This exchange determines the transcriptional response of *CAD* gene (Boyd and Farnham, 1997). Why USF1 binds to the *CAD* gene if it is unable to activate the gene was not addressed. Based on the promoter education hypothesis, one possibility would be that USF1 acts as a bookmarker, keeping the chromatin open through its interaction with SET1A (Deng et al., 2013) for rapid *CAD* gene activation following binding by MYC. In the absence of USF1, the chromatin may be in the closed state, leading to a slower induction kinetic by MYC.

USF and promoter education

We hypothesised that USF might be involved in promoter education as a bookmarker and/or pre-initiation education factor because: USF lacks an evolutionarily conserved TAD yet its binding is required for chromatin remodelling and gene activation. For example, USF1 was reported to be essential for maintaining bivalent domains in ESCs as well as shifting these bivalent marks toward active H3K4me3 during differentiation (Deng et al., 2013). Moreover, USF1:USF2 heterodimer binding correlates with increase H3-acetylation and gene activation (Rada-Iglesias et al., 2008). H3-acetylation is promoted through PRMT1-induced asymmetric dimethylation of H3R3 (Li et al., 2010).

USF and promoter priming

Development-linked promoter education

USF is involved in establishing epigenetic landscape necessary for expression of β-globin locus during development. USF2 binds to the LCR and mediates recruitment of CBP, $TF_{II}B$ and RNAPII in the undifferentiated state to prime the gene for activation through USF interaction with the tissue-restricted activator NF-E2 (Zhou et al., 2010). During erythroid cell differentiation, USF1 expression is induced. The exchange of a USF2 homodimer with a USF1:USF2 heterodimer results in the recruitment of PRMT1 to LCR which induces histone acetylation (Li et al., 2010). This histone modification is thought to promote transfer of transcription activity from the LCR to the β-globin gene promoter.

USF and promoter priming at metabolic genes

SREBP binding and gene activation in response to insulin is dependent on USF binding to the adjacent E-box (Latasa et al., 2003) and insertion of SRE into a promoter containing a USF E-box enables that promoter to be induced in response to insulin (Wong et al., 2009). Importantly, SREBP-1c$^{-/-}$ mice fail to induce any lipogenic genes in response to insulin (Sheng et al., 1995), indicating USF's role in gene priming. This notion is further supported by the report that there are no accompanying changes in acetylation status of H3 and H4 following SREBP-1c binding even though p300/CBP is required for lipogenic gene activation (Ericsson and Edwards, 1998). Thus, it seems highly plausible that USF (as a promoter education factor)-mediated chromatin modification through recruitment of p300/CBP (Huang et al., 2007) prior to SREBP induction is sufficient to keep the open chromatin context permissive for rapid activation upon SREBP-1c binding.

The role of USF as a promoter education factor at insulin-responsive genes is further supported by the observed correlation between the highly dynamic SREBP-1c protein levels (Farese et al., 2005; Roth et al., 2000; Sundqvist et al., 2005) and the induction of insulin-responsive genes in response to fasting and refeeding (Horton et al., 1998). USF, although unable to mediate insulin-responsiveness, is nevertheless required for gene activation of insulin-responsive gene by SREBP.

USF also serves similar roles at glucose responsive genes where USF binding is not sufficient for induction of glucose-responsive genes in response to glucose. Carbohydrate-responsive element-binding protein (ChREBP), on the other hand, is necessary for full activation of glucose responsive genes in response to glucose. However, interfering with USF2 binding blocks normal induction by glucose (Kennedy et al., 1997).

USF factor exchange with inducible TFs through direct competition

A potential involvement of USF as a bookmarker has been hinted at in the literature where USF binds to the same E-box with other bHLH-TFs, such as MYC, but full activation requires MYC, for example, the *CAD* promoter (Boyd and Farnham, 1997, 1999). Similarly, USF competes with AhR and ARNT for binding to the xenobiotic-response element (XRE) but is unable to fully activate the genes (Takahashi and Kamataki, 2001). As USF expression, localisation and DNA-binding is relatively constant while AhR nuclear localisation can be induced, USF binding may keep the XRE accessible for rapid gene activation in response to AhR ligands.

USF and HIF factor exchange

A number of hypoxic response genes have been reported to contain both HRE and E-box that are bound by both HIF and USF. E-boxes that cross-talk with the HRE tend to be palindromic as this disfavours binding by HIF-1α, but favours USF binding. Examples of genes co-regulated by HIF-1α/USF include, *PAI-1* (Dimova and Kietzmann, 2006), *TERT* (Goueli and Janknecht, 2003; Yatabe et al., 2004) and *L-PK* (Krones et al., 2001). Binding of USF to E-box adjacent to HRE was shown to interfere with the HIF/HRE driven gene expression (Krones et al., 2001; Samoylenko et al., 2001).

The interplay between USF and HIF can fine-tune gene expression in response to different levels of hypoxia. This is possible because USF and HIF bind to overlapping target DNA-elements with different affinity. For example, the *lactate dehydrogenase A* (*LDHA*) gene that contains three E-box/HRE element upstream of the TSS. E-box-I (5'-CACGTG-3') and E-box-II (5'-CACGTG-3') are strong USF- but weak HIF-binding sites, while E-Box-III (5'-

CACGTC-3') is a strong HIF binding site. Under hypoxia, HIF-1α might expand into E-box-I and E-box-II as HIF-1α expression is induced. Under high O_2, these E-boxes may be predominantly occupied by USF. This complementation was proposed to allow the expression from *LDHA* gene to be switched from HIF-1α to USF driven in response to increase O_2 and vice versa (Hu et al., 2011). If HIF-1α and USF are able to activate *LDHA* gene expression to a different degree, the transcription output from *LDHA* will be determined by the combination of HIF-1α and USF presented at the promoter.

Another interesting example is the exchange of USF and HIF-1α at the three tandem HRE/E-boxes separated by three nucleotides (5'-CACGCGccgCACGTGccaCACGCA-3') at the promoter of *BNIP3*. USF is not as strong an activator of *BNIP3* as HIF-1α and USF binding can interfere with HIF-1α-mediated *BNIP3* gene activation that is especially relevant at mild hypoxia where HIF-1α protein level is not as strongly induced. Although HIF-1α was shown to bind much more tightly than USF to the central E-box of *BNIP3* in an EMSA assay, USF binding persists under hypoxia *in vivo*. This persistent binding is dependent on the intact E-boxes flanking the central E-box. Thus, it appears that USF sidesteps on to the flanking E-boxes during hypoxia as HIF displaces USF from the central E-box (Hu et al., 2011). Thus, USF may act to cap the expression from *BNIP3* gene during mild hypoxia, allowing full expression only under strict hypoxia where elevated HIF-1α displaces the USF.

Thus, TF exchange between USF, AhR and HIF can be seen as a fine tuning mechanism which modulate the level of genes induction in response to varying degree of hypoxia, xenobiotic and physiological activator of USF such as p38 and glucose.

1.10 The aims of the project

We wish to explore the consequence of bHLH-TF exchange at promoters and some of the underlying mechanisms regulating bHLH-TFs exchange in humans. Factor exchange has been shown over the course of development by pioneer factors and lineage specific TFs, however, this does not represent factor exchange in the context of adaptive transcription

memory. Although of biological importance, development-linked factor exchange is often a non-dynamic, one-way process. Similarly, factor exchange in the context of MAX with MYC and MAD which resulted in gene activation and repression, respectively, although represent a good example of factor exchange is also binary and non-dynamic since MYC and MAD expression is generally inversely correlate. We are interested in factor exchange in the context of promoter education where factor exchange is dynamic. To this end, we will explore bHLH-TF exchanges at:

- Lineage specific, inducible genes such as the pigmentation-related *TYR* gene where binding of the ubiquitous USF1 and USF2 as well as the lineage-specific MITF to the E-box is required for *TYR* gene activation.
- Condition-specific, inducible genes in response to hypoxia. Conflicting reports in the literature implicated either HIF or USF in activation of the same genes. Others suggested HIF and USF bind to adjacent E-boxes and act in concert to activate hypoxic genes. While the latter notion is attractive, we wish to explore the dynamic and consequent of HIF and USF exchange following hypoxic induction, especially at genes which harbour only one HRE/E-box.
- Generic, housekeeping genes. As many genes are reported to be bound and activated by multiple bHLH-TFs. It is entirely possible that activation of these genes may be achieved by all bHLH-TFs. It is also possible, however, that exchange of generic TFs such as MYC and USF is involved in activation of these genes.

On top of factor exchange, we will explore the subtle differences in the target specificity and potential regulatory mechanism(s) that might affect DNA binding such as PTMs that influence DNA-binding affinity and regulation that influences local protein availability.

Lastly, we wish to explore how interfering with the function of one of the bHLH-TFs influences the outcome, in terms of the binding by other bHLH-TFs at common sites, and the consequences on transcription output.

2. Material and Methods

2.1 Chemicals, buffers and reagents

Table 2.1: List of buffers used. Chemicals and reagents were purchased from Sigma-Aldrich unless otherwise stated. 18MΩ water was used to make solutions. All DNA and RNA work were done using UltraPure™ DNase/RNase-Free Distilled Water (Invitrogen #10977 − 023)

Buffers	Composition
Band-shift buffer	25 mM HEPES (pH 7.4), 150 mM KCl, 5 µM DTT, 10% glycerol 200 µg/ml BSA, 1 µg Poly dI-dC (Fisher #11803865)
Bis-Tris MES running buffer	50 mM Tris base, 50 mM MES, 0.1% SDS,1 mM EDTA, pH 7.3
Bis-Tris transfer Buffer	25 mM Bis-Tris, 25 mM Bicine, 1 mM EDTA , pH 7.2
ChIP dilution buffer	16.7 mM Tris (pH 8.0), 167 mM NaCl, 1.2 mM EDTA, 1% Triton X−100, 0.01% SDS
ChIP elution buffer	100 mM $NaHCO_3$, 1% SDS
ChIP high salt wash buffer	20 mM Tris-HCl (pH 8.0), 500 mM NaCl, 2 mM EDTA, 1% Triton X−100, 0.1% SDS
ChIP low salt wash buffer	20 mM Tris-HCl (pH 8.0), 150 mM NaCl, 2 mM EDTA, 1% Triton X−100, 0.1% SDS
ChIP lysis buffer	50 mM Tris-HCl (pH 8.0), 10 mM EDTA, 10 mM sodium butyrate, 1% SDS, 0.5 mM PMSF, 4 × PIC
Laemmli sample buffer	62.5 mM Tris-HCl (pH 6.8), 10% glycerol, 2% SDS, 0.01% bromophenol blue
LDS sample buffer	0.2 M triethanolamine-Cl (pH 7.6), 10% glycerol, 1% LDS, , 1% Ficoll®−400, 0.025% phenol red, 0.025% coomassie G250, 0.5 mM EDTA
LiCl wash buffer	10 mM Tris-HCl (pH 8.0), 250 mM LiCl, 1 mM EDTA, 1% sodium deoxycholate,1% NP−40,
MITF band-shift buffer	10 mM Tris (pH 7.5), 100 mM NaCl, 2 mM $MgCl_2$, 1 mM EDTA (pH 8.0), 1 mM DTT, 4% glycerol, 10% FBS , 80 µg/ml sonicated salmon sperm DNA, 2 mM spermidine
Nuclear extraction buffer	20 mM HEPES (pH 7.9), 420 mM NaCl, 1.5 mM $MgCl_2$, 0.2 mM EDTA, 25% glycerol, 0.5 mM DTT, 10 mM sodium butyrate, 4 ×PIC
Nuclear isolation buffer	10 mM HEPES (pH 7.9), 10 mM KCl, 1.5 mM $MgCl_2$, 0.5 mM DTT, 10 mM sodium butyrate, 4 × PIC
PBS	137 mM NaCl, and 2.7 mM KCl. in 10 mM $-PO_4^{2-}$, pH 7.4
RIPA	25 mM Tris-HCl (pH 7.6), 150 mM NaCl, 5 mM EDTA, 1% NP−40, 0.1% SDS, 1% sodium deoxycholate, 4 × PIC, 1 × PhosSTOP
TBS	25 mM Tris base, 150 mM NaCl, 2 mM KCl, pH 7.4
Tris-Glycine running buffer	250 mM Tris, 192 mM Glycine, 1% SDS, pH 8.3
Tris-Glycine transfer buffer	25 mM Tris, 192 mM Glycine, 20% (v/v) methanol, pH 8.3

cOmplete, EDTA-free protease inhibitor cocktail (PIC) (Roche #05056489001)
Dithiothreitol (DTT)
Phenylmethanesulfonylfluoride (PMSF)
Lithium dodecyl sulfate (LDS)
Phosphate buffered saline (PBS)
PhosSTOP (Roche #04906837001)
Radioimmunoprecipitation assay buffer (RIPA)
Sodium dodecyl sulfate (LSDS)
Tris-buffered saline (TBS)

2.2 Bacterial methods

Growth conditions

For plasmid preparation, liquid bacterial cultures were grown in Luria-Bertani broth (Invitrogen #12780052), supplemented with appropriate antibiotics: ampicillin (100 µg/ml), kanamycin (30 µg/ml) or chloramphenicol (12.5 µg/ml) at 37°C, 225 rpm (Brunswick Scientific′s Innova® 44).

Luria-Bertani Petri dishes were prepared from LB-Agar (Lennox) (Invitrogen #22700025) supplemented with appropriate antibiotics.

Bacterial strains

For sub-cloning and PCR-based cloning, *E.coli* strain DH5α (Invitrogen #18265 − 017) (F^- Φ80*lac*ZΔM15 Δ[*lac*ZYA-*arg*F] U169 *rec*A1 *end*A1 *hsd*R17[r_k, m_k^+] *pho*A *sup*E44 *thi*-1 *gyr*A96 *rel*A1 λ^-) was used.

For TOPO cloning *E.coli* strain TOP10 (Invitrogen #C4040 − 03) (F- *mcrA* Δ[*mrr-hsd*RMS-*mcr*BC] Φ80*lac*ZΔM15 Δ *lac*X74 *rec*A1 *ara*D139 Δ[*araleu*]7697 *gal*U *gal*K *rps*L [StrR] *end*A1 *nup*G) was used.

For site-directed mutagenesis and deletion mutagenesis, *E.coli* strain XL-Gold (Agilent #200314) (Tetrdelta- (mcrA)183 delta- (mcrCB-hsdSMR-mrr)173 endA1 supE44 thi-1 recA1 gyrA96 relA1 lac Hte [F´ proAB lacIqZDM15 Tn10 (Tetr) Amy Camr]) was used.

Transformation of competent bacteria

All strains were transformed according to suppliers' protocols except, in all cases, recovery stage after transformations were carried out in S.O.C. media (Invitrogen #15544 − 034).

Frozen bacteria stock

Frozen bacteria stocks were prepared by mixing 1:1 ratio overnight-culture:autoclaved-glycerol then snap frozen on dry-ice-ethanol bath before storage at −80°C.

2.3 Mammalian cell line methods

Cell lines

Table 2.2: Mammalian cell lines used, culture conditions and origin. The HBL cell line was a kind gift from Jose Carlos García-Borrón. HERMES2C was obtained from St. George's Hospital cell bank. BT-549, HS-578T, T47D, HCT-116 and SW-620 cell lines were obtained from NCI-60. All other cell lines were obtained from the lab frozen stock.

Cell lines	Origin	Culture condition
501mel	Human melanoma	RPMI−1640
SkMel28	Human melanoma	RPMI−1640
HBL	Human melanoma (metastasis)	DMEM
IGR37	Human melanoma (metastasis)	RPMI−1640 and 15% FBS
IGR39	Human melanoma (primary)	RPMI−1640 and 15% FBS
B16	Mouse Melanoma	DMEM and 5% CO_2
Melan-A	Mouse melanocytes	RPMI−1640 and 200 nM TPA
Melan-a	Mouse melanocytes with defective MC1r	RPMI−1640 and 200 nM TPA
HERMES1	p16 null human melanocytes immortalized with hTERT	DMEM, 200 nM TPA, 200 pM CT, 10 ng/ml hSCF and 10 nM END−1
HERMES2C	Karyotypically normal subline of HERMES1	DMEM, 200 nM TPA, 200 pM CT, 10 ng/ml hSCF and 10 nM END−1
Phoenix (ampho)	Adenovirus E1A and SV40-T transformed HEK cells	DMEM
HEK293	Adenovirus E1A-transformed HEK cells	DMEM
HEK293FT	293 cell line carrying pFRT/lacZeo and pcDNA™6/TR (from the T-REx™ System)	DMEM
HeLa	Cervical adenocarcinoma	DMEM
MCF7	Breast adenocarcinoma (metastasis)	DMEM, 10 µg/ml Insulin, 1 mM Sodium Pyruvate
BT-549	Breast ductal carcinoma	RPMI−1640
HS-578T	Breast adenocarcinoma (metastasis)	RPMI−1640
T47D	Breast ductal carcinoma (metastasis)	RPMI−1640
HCT-116	Colorectal carcinoma	RPMI−1640
SW-620	Colorectal adenocarcinoma	RPMI−1640

Cells were grown at 37°C in humidified incubator with 10% CO_2 unless otherwise indicated
All media were supplemented with 10% FBS unless otherwise indicated
Dulbecco's Modified Eagle Medium (DMEM; Gibco #10566 − 016)
Roswell Park Memorial Institute Medium−1640 (RPMI; Gibco #61870 − 010)
Phorbol 12 −myristate 13 −acetate (TPA; Sigma #P1585)
Human stem cell factor (hSCF; Invitrogen #PHC2116), aka c-kit ligand
Endothelin-1 (END−1; Sigma #E7764)
Cholera Toxin (CT, Sigma #C8052)

Cell line maintenance

Monolayer cell cultures were regularly passaged at around 80 − 90% confluence, and used for no more than 25 passages, unless otherwise stated, to prevent replication-induced senescence or other problems associated with long-term culture. With the exception of

Phoenix cell lines which were directly dissociated in fresh media before diluting down, all cell lines were dissociated with Trypsin-EDTA $(\text{GIBCO } \#25200 - 056)$ at room temperature. Once the cells were dissociated, trypsin was inactivated in five volumes of complete medium, ready for either passage or experimental use.

Stable cell lines were kept in appropriate antibiotic at all times: Geneticin $(1\ \text{mg/ml})$ $(\text{GIBCO } \#10131 - 027)$, Puromycin $(5\ \mu\text{g/ml})$ $(\text{GIBCO } \#\text{A}11138 - 02)$ or Blasticidin S $(20\ \mu\text{g/ml})$ $(\text{GIBCO } \#\text{A}11139 - 03)$

For long-term storage, fresh complete media was replaced the day before frozen stocks were made. $\approx 30\%$ of T75 –flask at $80 - 90\%$ confluence was frozen per cryovial in FBS containing 10% DMSO and stored in liquid N_2.

Transient transfection of DNA

Transient transfections were carried out using a reverse-transfection protocol where exponentially growing cells were trypsinised and plated at $\approx 60\%$ confluent concurrent with transfection. The transfection complex was prepared with a $3:1$ ratio of FUGENE VI $(\text{Promega } \#\ \text{E}2692)$:DNA suspended in Opti-MEM $(\text{GIBCO } \#31985 - 062)$ according to the supplier's protocol.

Generation of stable cell lines

Stable cell lines were generated by transfecting expression plasmid containing either the Puromycin-, Neomycin- or Blasticidin-resistance genes or co-transfection of the expression vector with an empty vector containing the resistance genes using $10\ \mu\text{g}$ DNA per $10\ \text{cm}$ dish prepared as for transient transfection. In the case of co-transfection a $10:1$ ratio of expression vector to vector containing the resistance gene was used. After $24\ \text{h}$ post-transfection, complete media containing either $10\ \mu\text{g/ml}$ Puromycin, $5\ \text{mg/ml}$ Geneticin or $20\ \mu\text{g/ml}$ Blasticidin S was added into the medium. The media was changed daily for one week. Cloned cell lines were generated using a serial dilution method to obtain

single cell clones. Stable integration of ectopic genes was initially screened by Western-blot. Positive clones were then validated by immunofluorescence.

Transfection of siRNA

siRNA transfections were carried out using Lipofectamine RNAiMax (Invitrogen #13778 – 150), according to the manufacturer protocol.

Table 2.3 List of siRNA used

Target	siRNA identifier	Working concentration (nM)	Catalogue number	Supplier	Targeting sequence
USF1	SMARTpool: siUSF1#1	50	L-003614-00-0020	Dharmacon	GAGUAAAGGUGGGAUUCUA
	SMARTpool: siUSF1#2	50			CCAAAHCUUHUHAUUAUAU
	SMARTpool: siUSF1#3	50			AAACGCAGGGCUCAGCAUA
	SMARTpool: siUSF1#4	50			GUAACCACCGCUUGUCUGA
	siUSF1	50	s14719	Ambion	CUGCUGUUGUUACUACCCA
	siUSF1a	50	Custom	QIAGEN	GACCCAACCAGUGUGGCUA
	siUSF1b	50	Custom	QIAGEN	CUGGCGCCAUCAGUGGCU
USF2	SMARTpool: siUSF2#1	50	L-003618-00-0020	Dharmacon	GCAAGACGGGAGCGAGUAA
	SMARTpool: siUSF2#2	50			GGAGGGACAAGAUCAACAA
	SMARTpool: siUSF2#3	50			GAAGAGCCCAGCACAACGA
	SMARTpool: siUSF2#4	50			CAAAAUCCCUUCAGCAAUG
	siUSF2	50	s14720	Ambion	CCCUUACUCUCCAAAAAUU
	siUSF2a	50	Custom	QIAGEN	CCUCCACUUGGAAACGGUA
	siUSF2b	50	Custom	QIAGEN	AGGCCUGCGAUUACAUCCGG
MiTF	siMITF	83	Custom	QIAGEN	AGCAAGUACCUUUCUACCAC
siCTR	All-Star negative	83	1027281	QIAGEN	N/A

Treatments of cell lines

UVB induction

Cells were grown on 10 cm dishes and serum-starved (0.1% FBS) overnight prior to UVB-treatment. Growth medium was removed for UVB treatment. After treatment, the medium which was used to grow the cells prior to serum-starvation was added back to the cells (pre-conditioned medium) to avoid serum suppression of p38 SAPK.

MC1R activation

α-MSH (#M4135), NDP-α-MSH ([Nle^4,D-Phe^7]-α-MSH; #M8764) were obtained from SigmaAldrich while CUV9900 and afamelanotide (SCENESSE®) were obtained from Clinuvel Pharmaceuticals Limited (as part of collaboration with industry) and used at the indicated concentration.

2.4 Nucleic acid methods

Nucleic acid purification

Small- and large-scale plasmid purifications were done using QIAprep Spin Miniprep Kit (QIAGEN #27106) and QIAGEN Plasmid Maxi Kit (QIAGEN #12163) from 2 ml and 450 ml Luria-Bertani cultures respectively.

PCR, ChIP and ChIP-seq DNA were purified using QIAquick PCR Purification Kit (QIAGEN #28106). For cloning, PCR- and plasmid-digested products were electrophoresed on agarose, appropriate band(s) excised and purified using QIAquick Gel Extraction Kit (QIAGEN #28704).

Total RNA from cultured cells for RNA-seq, RT-qPCR and for cloning were purified using RNeasy Mini Kit (QIAGEN #74104). gDNA-contaminants were purified through on-column DNase (QIAGEN #79254) treatment for RNA-seq samples or gDNA Wipeout Buffer treatment (QIAGEN #205313) for RT-qPCR prior to reverse-transcription step.

All nucleic acid preps were carried out following the manufacturers' guidelines.

Nucleic acids quantification and assessment

Column-purified DNA- and RNA-concentrations were determined with a NanoDrop 2000c (Thermo). ChIP-seq purified DNA concentrations were measured using a Qubit® 2.0 Fluorometer (Invitrogen #Q32866) according to the manufacturer's protocol.

Assessment of ChIP-seq-DNA- and RNA-seq-RNA-sample quality was performed according to the manufacturer's protocol with a Bioanalyzer (Agilent #Q32866) using a

High Sensitivity DNA kit (Agilent #5067 – 4626) and RNA 6000 Nano Kit (Agilent #5067 – 1511), respectively. ChIP-seq samples where ≈≥ 85% of fragments lie between 50 – 300 bp were accepted. RNA-seq samples were only accepted when the RNA Integrity Number (RIN) values is ≥ 9.0.

Plasmids for transfection were tested for their integrity by electrophoresis on 0.8% agarose (Fisher #BP1356) gel. qPCR-, PCR-products and sonicated gDNA for ChIP applications were generally checked for specificity, product-size and fragment range, respectively, on 1% agarose gel. Samples were prepared in 1 × loading buffer (16% v/v formamide, 1% v/v 1 × TBE) and subjected to electrophoresis at 150 V, ≈ 25' in 1 × TBE (GIBCO #15581 – 044). Visualisation of DNA was done using SafeView Nucleic Acid Stain (NBS biologicals #NBS – SV1) under UVIdoc (UVItec).

Polymerase chain reaction (PCR)

All PCR reactions were carried out according to the suppliers' recommendations.

qPCR and RT-qPCR

RNA for gene expression analysis was extracted using RNeasy kit and cDNA generated using QuantiTect Rev. Transcription Kit (QIAGEN #205313) according to the supplier's protocol. To avoid sampling problems, we routinely use 50 ng starting RNA per quantitative real-time PCR (qPCR) reaction.

DNA for ChIP-qPCR was prepared as described in the section 2.5, ChIP. In general, one IP (8 µg antibodies) was used for no more than 40 –reactions to avoid sampling problems.

qPCR analysis was carried out in technical duplicate on the Rotor Gene Q (QIAGEN) using BRYT-master mix (Promega #A6002) and 66.7 nM primers with cycling parameters of 95°C, 5' enzyme activation and 60 cycles of 96°C, 8 s; 61°C, 10 s; 72°C, 5 s. The melting curve was determined by ramping temperature from 72°C to 96°C in steps of 1°C. Data was analysed using comparative C_t ($\Delta\Delta C_t$) method (Schmittgen and Livak 2008). For gene

expression, this meant each target is normalised against a housekeeping gene (*β-ACTIN*) and fold-enrichment is presented as a value relative to negative control. For ChIP-qPCR, this meant each target is normalised against a GAPDH primer pair targeting a region lacking any motif resembling an E-box and fold-enrichment is presented as a value relative to IgG.

All ChIP-qPCR primers used were subjected to *in silico* analysis for secondary structure and non-specific priming as well as further validation for amplification linearity against standard curves

Site-directed mutagenesis

Single point-mutations were introduced in existing plasmids using QuikChange Lightning Site-Directed Mutagenesis Kit (Agilent #210519). Multi-site mutations were introduced using QuikChange Lightning Multi Site-Directed Mutagenesis Kit (Agilent #210515). All steps were carried out according to the supplier's recommendations except: equimolar rather than equimass of primers was used (final concentration of 0.25 μM), 1 μl of DMSO was always added and Dpn1 enzyme treatment was performed for 4 h rather than 5'.

Deletion/Insertion mutagenesis

With the exception of the first batch of deletion mutants of USF2 which were cloned into the pCMV5 − 3 ×HA vector using EcoRI-digestion-cloning approach, all other deletion mutants were made using PCR coupled dsDNA-break (containing staggered nicks) rescue. For the PCR reaction the first half of the primer is specific to the end of the protein before the section to be deleted, while the other half of the primer is complementary to the beginning of the sequence right after the section to be deleted. PCR reactions were carried out using PfuUltra II (Agilent #600850) such that the polymerase copied the whole plasmid except the region to be deleted. The reactions were then treated with Dpn1 enzyme to remove the parental vectors and then transformed into XL-Gold chemically competent cells which would rescue the PCR products by stitching together the overlapping part of the primers (3'-end of both the forward- and the reversed-primers).

For insertion mutagenesis, if the region to be inserted was small ($\approx \leq 50$ bp), the sequence to be inserted was directly synthesised into the primers such that it resided between the two-half primers described above. PCR reactions and plasmid recues were performed as with deletion mutagenesis. Small insertion was used to introduce the second tag into the dual epitopes-tag constructs. For large insertion, the primers were designed such that the first half of the primer is complementary to the location on the plasmid which the insertion is to be inserted and the second half is complementary to the beginning of the sequence to be inserted. Initial PCR was carried out to generate mega-primers which were used for second PCR reaction as performed for deletion and small insertion above, except: initial denaturation was carried out for 5' and the first 5 cycles of the PCR was carried out at 55°C. This was used to generate all tet-inducible constructs as the parental plasmid lack appropriate restriction sites.

Cloning and sub-cloning

DNA-fragments for downstream restriction enzyme (RE) mediated-cloning were amplified from plasmids using Phusion High-Fidelity PCR Kit (Thermo #F − 553L). RE digestions were carried out using appropriate enzymes from Invitrogen or Promega. Digestions of PCR fragments for cloning and/or plasmid for sub-cloning were carried out overnight, 37°C. In the case of single RE digestion of plasmid backbones, 5U/reaction of CIAP (Invitrogen #18009 − 019) was also added into the reaction mixture to prevent re-ligation. All reactions were purified using QIAquick PCR Purification Kit prior to ligation with 5U/reaction of T4 DNA Ligase (Invitrogen #15224 − 041) using a 10: 1 ratio of insert:vector and left for 1 h at room-temperature prior to transformation into DH5α.

Cloning from RNA were performed from cDNA generated using AccuScript High-Fidelity 1st Strand cDNA Synthesis Kit (Agilent #200820) before subjected to PCR amplification using Phusion High-Fidelity PCR Kit and TOPO cloned into pcDNA™3.1 Directional TOPO® Expression Kit (Invitrogen #K4900 − 01) according to supplier's protocol except TOPO reactions were incubated for 30' rather than 5'.

Colony PCR technique was used to verify inserts size and orientation for all constructs and carried out using 1 µl of an ≈ 16 h-incubated colony dissolved in 100 µl water and MyTaq™ Red Mix $(\text{Bioline \#BIO} - 25044)$ in a 10 µl PCR reaction. Minipreps of positive clones were then sent for sequencing before scaling up for MaxiPrep.

2.5 Protein techniques

Whole cell extract

Whole cell extract (WCE) were prepared from, trypsinised, washed and centrifuged cell pellets dissolved in $4\times$ Laemmli sample buffer or $4\times$ NuPAGE® LDS Sample Buffer $(\text{Novex® \#B0008}) + 20\%$ β-mercaptoethanol, added fresh before use. Extracts were boiled at $95°\text{C}$ for 5'.

Subcellular fractionation

Cell pellets were re-suspended in ice-cold nuclear-isolation buffer (5 –volume of pellets) and rotated at $4°\text{C}$ for 10' before $\text{NP}-40$ was added to a final concentration of 0.5% and incubated for a further 5'. Plasma membrane breakdown was confirmed under the light microscope before centrifugation at $5{,}000 \times g$ for 10'. The resulting supernatant constituted the cytosolic fraction while the pellet was enriched in nuclei. The two fractions were reconstituted to the same final volume in $4\times$ Laemmmli- or $4\times$ LDS- buffer such that each µl of the nuclear and the cytosolic fraction correspond to the same number of cells.

SDS-PAGE

In order to better resolve PTM(s), we used $200{:}1$ acryamide:bisacrylamide $(\text{custom order, Severn Biotech})$ gels as opposed to the standard $37.5{:}1$ acryamide:bisacryamide. Resolving gels for Tris-glycine SDS polyacrylamide gel electrophoresis (SDS-PAGE) system contained 375 mM Tris (pH 8.9), 0.1% (v/v) SDS, and polymerised with 0.1% ammonium persulfate (APS) and Tetramethylethylenediamine (TEMED). Resolving gels were overlain with a stacking gel comprising 6% acrylamide $(37.5{:}1$ acryamide: bisacryamide), 146 mM Tris $(\text{pH } 6.8)$, 0.05% (v/v) SDS, and polymerised with 0.1% APS

and TEMED. 13.5% acrylamide gels were used to resolve MITF whereas 15% acrylamide gels were used to resolve USF bands. Gels were run at 170 V, 65'.

Coomassie and silver staining

Coomassie- (expedeon #ISB1L) and silver-staining (Biorad #161 − 0449) were used to guide the position of the band(s) to be excised for immunoprecipitation (IP)- and to assess the purity of co-immunoprecipitation (CoIP)-materials for Mass-spec (MS) applications.

Western-blotting

Gels were transferred onto Immun-Blot Low Fluorescence PVDF Membrane (Biorad) at 100 V , 60' Membranes were either probed with a single 1° antibody for blotting with 2° −HRP conjugated antibody and developed using Clarity™ Western ECL Substrate (Biorad #170 − 5061) or probed with three antibodies from different species and detected using Alexa Fluor−488, −546 and −647 2° antibodies and visualised using ChemiDoc.

Antibodies were diluted in 5% non-fat milk dissolved in PBS + 0.5% Tween−20 (PBS-T). 1° were incubated overnight, 4°C while 2° antibodies were incubated for 1 h , room-temperature. Phospho-specific and anti-acetyl-lysine antibodies were diluted in TBS + 5% non-fat milk dissolved in TBS + 0.5% Tween −20 (TBS-T). Washing was done before developing on the film or visualising on ChemiDoc for ≈ 30', changing fresh buffer every 5'.

Affinity purification

All steps involved in affinity purifications were carried out at 4°C . Either Axygen® 1.7 ml Microcentrifuge Tube (CORNING #MCT − 175 − C − S) or LoBind Tubes (Eppendorf #0030108116/0030108051) were used as these were certified free of DNAase, RNase and PCR clean. LoBind tube which reduces background was preferred for MS-proteomics.

Immunoprecipitation

Cells were trypsinised, washed in PBS and pelleted by centrifugation at 1200 × g, 4'. Cell pellets were either lysed in PBS + 1% Triton X−100 or RIPA supplemented with 4 × PIC for

CoIP or IP, respectively. The cell suspensions were briefly sonicated for $30\ s$ ($7.5\ s$ on/off cycle) using Bioruptor® Plus (Diagenode) then cleared by centrifugation at $13000 \times g, 10'$. The supernatants were transferred to fresh tube, keeping 10% as input and the remainder rotated overnight with either $0.5\ \mu g$ antibody for IP or $5\ \mu g$ antibody for CoIP (or $20\ \mu g$ antibody for MS-samples). An appropriate amount (10% on top of the theoretical binding capacity) of Dynabeads Protein A or G (Invitrogen #10002D/10004D) were washed in lysis buffer and blocked overnight with $0.5\ mg/ml$ BSA (Sigma #B8894) before adding to the samples and rotate for $1\ h$. The beads were washed 3 –times with lysis buffer, transferring to a new tube after the first wash and finally boiled in $4 \times$ Laemmli sample buffer.

GST-pull-down

GST-pull-down assays were performed as with IP but affinity pull-down was done with glutathione-sepharose beads ($GE\ \#17-0756-01$). $50\ \mu l$ of washed and resuspended beads ($50\%\ v/v$) were directly added to the lysate and rotated for $1\ h$ before washing 3 –times with lysis buffer, transferring to a new tube after the first wash and finally boiled in $4 \times$ Laemmli sample buffer.

Mass-spec

MS-proteomic analysis and MS-identification of PTM(s) and MS-peptide sequencing were performed in collaboration with Benedikt Kessler's lab where all the sample preps and validations were performed by us and the MS-run by Kessler's lab. For proteomic analysis of CoIP, whole eluates were directly subjected to trypsin digestion and MS-analysis. For PTMs-analysis and peptide sequencing, the samples were first run on Bis-Tris gel (as the neutral pH of the system should better preserve labile PTMs) to clean up the reaction. The band(s) were then excised and in-gel digestion was carried out before running on MS.

ChIP-seq

Chromatin immunoprecipitation followed by high throughput sequencing (ChIP-seq) were performed as described for ChIP (Nelson et al. 2006) with some adjustments. Three

80% confluence 15 cm dishes were trypsinised, collected into 50 ml falcon tube (Corning # 430828), centrifuged (800 × g, 4') and media aspirated. Cross-linking was done by adding 35 ml ice-cold PBS containing 0.4% paraformaldehyde and rotated at 4°C, 10' before quenching with glycine to a final concentration of 0.2 M for 10'. Samples were then washed and centrifuged (1500 × g, 10'). Lysis was done in 1 ml ChIP lysis buffer by passing cell suspension through a 25g needle (Terumo #NN − 2525R) until there were no visible clumps. A further 1 ml of ChIP dilution buffer was added before sonication for ≈ 12' in Covaris S220 (160 W peak incident power, 5% duty cycle and 200 cycles per burst) until 200 − 400 bp fragments were obtained (assessed by agarose gel electrophoresis).

The sonicated chromatin was cleared by centrifugation at 13,000 × g, 10'. Supernatant was diluted in 8 ml of ChIP dilution buffer (10 ml total) before 80 µg of respective antibodies were added and rotated in a 50 ml falcon tube overnight. At the same time 400 µl Dynabeads Protein A or G were washed and resuspended in ChIP dilution buffer and blocked in 0.5 mg/ml BSA overnight. IP was carried out using blocked-Dynabeads, rotated for 1 h and centrifuged (1500 × g, 10'). The beads were resuspended in 1 ml ChIP low salt wash buffer and transferred to a fresh microcentrifuge tube, washed, resuspended in a further 1 ml ChIP low salt wash buffer before transferring to a fresh microcentrifuge tube. Further washing, 2 × ChIP low salt wash buffer, 2 × ChIP high salt wash buffer and 2 × LiCl wash buffer was done in the same tube. The beads were eluted in 1.2 ml elution buffer. Reverse cross-linking of ChIPed-DNA was done at 65°C overnight with addition of 0.3 M NaCl (final concentration), 20 µg RNase A (Invitrogen #12091 − 039) and 20 µg Proteinase K (Fermentas #EO0492).

Recovery of ChIPed-DNA was done using QIAquick PCR Purification Kit. 6 ml PB buffer was added to the reversed cross-linked DNA before passing through four-columns as described per supplier's instruction. Final elution was done by passing 30 µl water sequentially through all the columns. This fraction was kept for ChIP-seq library preparation,

1 µl of which was assessed on a Bioanalyzer. A further 120 µl water were used to do a second elution from each of the four-columns (a total of 480 µl). This fraction was use in qPCR analysis at 5 µl per reaction.

Samples which passed the QC on the Bioanalyzer and showed enrichment at expected targets on qPCR were subjected to sequencing on HiSeq 2500 (Illumina) carried out using Wellcome Trust genomic service.

ChIP

ChIPs were carried out as described in the ChIP-seq protocol except: 1) each diluted, sonicated chromatin were used for 10 –ChIPs with different antibodies at 8 µg/ml in 1.5 ml tubes, 2) 50 µl Dynabeads were used per ChIP, 3) reverse cross-linking was done in 250 µl and 4) DNA was recovered with one PCR purification column per antibody and eluted in 2 × 120 µl water for use in qPCR analysis at 5 µl per reaction.

EMSA

Electrophoretic mobility shift assays (EMSA) were carried out using: total nuclear extract to examine endogenous proteins or transiently over-expressed epitope-tagged proteins expressed in Phoenix cells. Nuclei were isolated as described in Subcellular fractionation section before subjected to further extraction using nuclear extraction buffer (3 –volume nuclear pellet), 4°C, 10' then centrifuge at 13,000 × g, 10'. The resulting supernatants contained soluble nuclear protein extract which were aliquot and stored at −80°C until used.

To carry out band-shift experiment, a reaction mix containing 17 µl band-shift buffer, 1 – 2 µl nuclear extract, 200 pM fluorescently-labelled ds-oligonucleotide probe and water to a total of 20 µl was incubated on ice for 20' before loading on to a gel containing 6% sequencing grade acrylamide (29:1 acrylamide:bis – acrylamide; Severn Biotech #20 – 3500 – 05) and 0.5 × TBE, polymerised with addition of 0.09% APS and TEMED. Gels were run at 200 V, 90' and imaged on ChemiDoc.

For MITF, a reaction mixture containing $17\ \mu l$ MITF-band-shift buffer, $1-2\ \mu l$ nuclear extract, $200\ pM$ fluorescently-labelled ds-oligonucleotide probe and water to a total of $20\ \mu l$ was used.

The identity of the complex in the band-shift was verified by antibody super-shift. To do this, $1\ \mu g$ of antibody was added to the reaction mixture before the probe was added. In the case where the antibody blocked the DNA-binding interface of the protein, loss of band-shift would be seen. Alternatively, when the antibody binding did not interfere with DNA binding, a super-shift would be seen.

Relative binding affinity to different DNA-element was determined by titrating 10, 50 and $250\ ng$ of unlabelled-oligonucleotides carrying a specific DNA-element into the reaction mix containing a constant amount of the probe. Sonicated salmon sperm DNA was used as a non-specific competitor carrying various DNA-elements, most of which would not be the canonical binding sites of the TF of interested. Competition against sonicated salmon sperm DNA was performed to assess how amino-acid substitution at MITF-$K243$ equivalent residue that disrupt contact to the DNA-phosphate backbone would affect the overall DNA-binding affinity of the TF of interested. Salmon sperm DNA was titrated at a $2-$ or $4-$fold increment step maxing at $10\ mg/ml$.

2.6 Cell biology methods

Immunofluorescence

Cells were grown on $13\ mm\ Ø$ coverslips $(VWR\ \#\ 631-0149)$ overnight then fixed using 3.7% formaldehyde for $10'$, then permeablised for $10'$ in 0.5% TritonX-100 $-$PBS, washed twice in PBS, blocked $30'$ in 3% BSA-PBS before overnight incubation, $4°C$ with $1°$ antibodies diluted at $1:100$ in 3% BSA-PBS. $2°$ Alexa-fluor-conjugated antibodies were diluted at $1:250$ in 3% BSA-PBS, incubated for $1\ h$ at room-temperature, washed $3\times 10'$ then mounted on microscope slides using Vectashield-anti-fade medium containing DAPI $(Vector\ \#H-1200)$. All images were taken on LSM710 confocal microscope $(Zeiss)$.

Cell-cycle analysis

Cell-cycle analysis was performed using flow cytometry (FACS Canto, BD). Briefly, cells grown on 6 –well plates were trypsinised at $\approx 70-80\%$ confluent, and processed as for immunofluorescence samples up till washing, except in suspension (on rotating wheel) and incubation was done at room-temperature, $1\,\text{h}$. Finally, nuclei were stained using TO-PRO-3 $(\text{Invitrogen \#T3605})$ at $1\,\mu\text{M}$ for 30', room-temperature, washed twice in PBS and ready for measurement after adjusting for background and cross-talk between Alexa and TO-PRO-3 according to BD's recommendation.

Cell-cycle and proteins sub-cellular localisation were correlated by co-staining with anti-H3S10P antibody and visualised on LSM710.

2.7 Bioinformatics

ChIP-seq

Pre-processing, mapping and peak calling

All ChIP-seq experiments were carried out in two-biological replicates (except MITF which had three-biological replicates). The first replicate samples (first two for MITF) were carried out with 60-million pair-end reads while the second replicate (third for MITF) samples were carried out with 30-million single-end reads. Each biological replicate contained two technical replicates (same library sequence in two separate flow cells) that were stitched together using the concatenate command in UNIX to generate a single fastq file for each biological replicate. Raw fastq files were fastQCed to check the read quality and PCR duplication, processed and mapped to human genome build hg19 (GRCh37, February 2009) using Bowtie (Langmead et al. 2009) allowing for 2 mismatches and supressing all paired end data which mapped to more than one region. Mapped SAM-files were used to generate tag-directory (coordinate of mapped sequence) using Homer package (Heinz et al. 2010) separately for paired-end and single-end data before merging the tag-directory into one directory for each bHLH-TF.

Bedgraphs of mapped data were generated for visualisation in the University of California at Santa Cruz (UCSC) Genome Browser using the makeUCSCfile command of Homer. To aid direct visual comparison, all bedgraphs were normalised to 10^7 reads.

Peak calling was done using the findpeaks command of Homer. Background files used were generated by performing a parallel ChIP-seq experiment using HA-antibody against 501mel parental cell lines. Regions with ≥ 4 –fold tag-density compare to background files within ± 5 kb and has a cumulative Poisson p-value of ≤ 0.0001 were called as peaks. Peaks were reported as "peak score" which is a position-adjusted value and is correlated to the number of reads found in the peaks after normalisation of total mapped reads to 10-million.

Peaks were annotated by assignment to the nearest known gene or transcription start sites (TSS) or transcription termination site (TTS) using annotatedPeaks.pl command of Homer which use RefSeq genes as reference.

Gene and genome ontology

Gene ontology (GO) and genome ontology analysis were done using annotatePeaks.pl function of the Homer package. The GO analysis outputted from Homer covers most of the major databases including the KEGG and Gene Ontology Consortium database. For GO-analysis of the top 500 genes, we used AmiGo2.

The basic output of genome ontology categorised hits as falling under CpG islands, gaps, centromeres, promoters, 5'UTR, 3'UTR, exons, introns, coding, TTS, non-coding, pseudo, protein coding, ncRNA, rRNA, intergenic regions. This is essentially the same as using the Sitepro function of Cis-regulatory Element Annotation System (CEAS) (Shin et al. 2009) but for the whole dataset.

Motif analysis

De novo motif identification was carried out using findMotifsGenome.pl which is a differential motif finding algorithms. The sequence ± 30 bp from the peak summit were used

to identify enriched occurrence of certain strings of nucleotides against the background. Background is defined as sequences ± 50 kb of the TSS segmented into 200 bp, of which $\approx 50{,}000$ regions of similar GC content as the input sequence were used as background. As the algorithm looks for motifs of specified length and will report sequence as CATG of equal position weight score just to fill the gap at the ends, and in order to simplify the motifs reported, any terminal sequences which do not have at least one nucleotide with the log-odds probabilities of > 0.33 are manually trimmed. Motifs with P-value of ≥ -10 are considered as false positives while motifs with P-value of ≤ -70 were considered as significant. The motifs identified were screened against a database of known motifs and those with a similarity score of > 0.8 were reported as "best matches".

To correlate motifs with binding affinity, peaks were screened for specific motif(s) and a box-plot of peak score for each motif was plotted using R. To assess motif requirement for co-occupancy among different combinations of bHLH-TFs, co-occupying peaks were identified and sequences under the co-occupying peaks were used for motif analysis.

Co-occupy and unique targets

Identification of co-occupying targets among different bHLH-TFs was performed either by direct intersection of peak coordinates (from filtered-dataset where peak score is > 10) to give all incidences of co-occupancy, or by calling peaks using (an)other tag directory(s) as background to identified peaks preferentially bound by the TF in question relative to (an)other TF(s). Peaks were classified as being unique to one bHLH-TF when the tag-density from that bHLH-TF is $\geq 10-$fold enriched within ± 5 kb and has a cumulative Poisson p-value of ≤ 0.0001 over (an)other bHLH-TF(s).

Peaks were further classified according to the number of TFs co-occupying the peaks. All together, these analysis were used to identify potential sites where promoter education could potentially occur and also which bHLH-TFs were likely to be involved. The sequences

under the common peaks were also used for motif finding to identify sequence requirement for co-occupancy by a given set of bHLH-TFs.

501mel-specific peaks were identified through sequential intersection of peaks which are ≥ 10 −fold enriched within ± 5 kb and have a cumulative Poisson p-value of ≤ 0.0001 in 501mel over another cell line. Four cell lines were used as background for USF and MYC while the ccRCC cell line was used for HIF.

Histogram of ChIP-seq coverage

ChIP-seq mapped files were used to populate read-counts around selected peak summits (identified through peak-calling above) as a reference. For direct comparison, all mapped files were normalised to 10^7 reads. The read-counts for each mapped file around the peak summits were binned every 5 bp before plotting as a function of ChIP-fragment depth against distance from the peak summit. TFs whose binding overlap with the TF used to generate the reference peak will display a higher ChIP-fragment depth than those where binding does not extensively overlap with the reference.

Correlation analysis

Pairwise correlation analysis between different datasets against each reference peak file was performed to obtain the overall degree of similarity between different bHLH-TFs against any given peak file. Normalised, populated read-counts around the reference peak file generated as for the histogram, except no-binning, were used for linear regression analysis. Datasets which have higher R^2 values indicate a higher degree of similarity in terms of read-count around the reference peaks and likely indicate a more similar binding profile and, by extrapolation, binding specificity (two bHLH-TFs with a similar binding profile but different binding specificity will give a poor correlation coefficient.)

Heatmap of ChIP-seq coverage

As pairwise correlation analysis only gives an overall correlation between datasets, we also generated heatmaps of ChIP-seq coverage which retain the information at an individual

peak level. Normalised, populated read-counts around the reference peak file were centred average against the average read counts for each dataset to account for the difference in the background and ChIP efficiency between different datasets. A filter was applied to remove peaks whose normalised read-counts do not significantly differ among all the bHLH-TFs studied (that is, the peak is similarly bound by all bHLH-TFs in query) such that the number of peaks is limited to ≈ 2500 peaks. The remaining peaks were then centre normalised across all the bHLH-TFs at each peak and hierarchical clustering using average linkage carried out using Gene Cluster 3.0. The output matrix was used to generate heat-map for visualisation using TreeView.

RNA-seq

All RNA-seq samples were poly A-enriched and rRNA-depleted by the Wellcome Trust genomic service. Sequencing data was paired-end and strand-specific.

Pre-processing and mapping and visualisation

Raw fastq files were mapped to hg19 (GRCh37, February 2009) using TopHat2 only considering the refgene GTF file from UCSC. Bedgraph files for visualisation were generated from TopHat2-mapped files and normalised to 10^7 reads.

Transcript quantification

Raw-read count and RPKM-normalised data for each gene was calculated using analyzeRepeats.pl (Homer) and differential gene expression analysis was performed using EdgeR (Robinson et al. 2010). For simplicity, we merged all multi-copy genes to one and condensed all alternatively spliced genes into one to avoid multiple assignment of a gene to a single peak from the ChIP-seq datasets.

One of the major sources of variability between RNA-seq experiments is the efficiency of poly-A enrichment and rRNA-depletion steps. As rRNA represents the majority of the total-RNA extracted from the cells, we addressed this by limiting our normalisation to total mRNA when using the RPKM approach. For differential gene expression analysis using EdgeR,

which required a raw read count as input, we limited the total read count per position to three. By limiting the read count per position, we effectively limited the highly abundance RNA-species in each experiment to a roughly constant proportion (as these RNA-species will invariably exceed the threshold, so we essentially capped them to a fixed proportion per experiment). The one downside of this approach is that we would lose some of the dynamic-range of the very highly expressed genes.

Clustering and heat-maps

Gene Cluster 3.0 was used to cluster and normalise RPKM output from analyzeRNA.pl. We filter out any entry with ≤ 32 −mapped reads and differences among conditions were less than $8-$ or $16-$fold before being centre-normalised and hierarchical clustering using average linkage. The output matrix was used to generate a heat-map for visualisation using TreeView.

Principle component analysis

Using the RPKM of normalised mRNA-reads, we filtered out genes which have ≤ 32 normalised reads to avoid significant difference (but not statistical significance) arising from low coverage of those genes. The filtered matrix is used for principle component analysis (PCA), first considering $2-$ principal components (PCs), then repeated with $3-$PCs. PCA was performed in R.

Assigning gene expression data to ChIP-seq peaks

The differential gene expression matrix output from EdgeR was assigned to ChIP-seq peaks using annotatedPeaks.pl which matches the gene ID for both datasets.

2.8 Primer list

EMSA

Target	Forward primers (5'→3')	Reverse primers (5'→3')
***TYR* Enhancer M-box WT**	CTTGTGGAGATCATGTGATGACTTCCT GATTCC	GGAATCAGGAAGTCATCACATGATCTC CACAAG

Note: The E-boxes are highlighted in red. Only the wild-type sequence are shown. For TYR, refer to the result section for the mutation introduced.

ChIP

Target	Forward primers (5'→3')	Reverse primers (5'→3')	E-box
ASNA1 #1	CCTGTAGGTTCTTCGTATGGTG	GAGCATCTTCGAACTCCTCTG	T-CACGTG-A
ASNA1 #4	GGTGGTGTGGGCAAGAC	CCCTTCGCCCAGTGGTAT	C-CACCTG-C
BAX #1	CTTATCGGGAGATGCTCATTGG	TGAACGTGCGTCCTTCAC	T-CACGTG-A
BAX #3	TTGGACGGACGGCTGTT	GGAGGTTTGGTCCCGTCA	B-CACGTG-M
DVL2	CGTGACTTAAAGGCAGCCA	GCTGAAGGCAGCGATCAT	T-CACGTG-A
EGLN3	TGTCTGGGAAAGCCTCTACT	CCAGTTGTCCACCCAAATGA	Y-CACGTG-A
FAM21B #1	CCCAAACTCCAGTCCCAGTC	ACAGAACCCGCACGTGAT	T-CACGTG-M
FAM21B #3	CATGGCAGGTCGGATCAC	CGAAGCTGCCAGCACAC	T-CACGTG-A
GAPDH	GAAGGGCTTCGTAGACTGG	CTTAAGGCATGGCGCAACT	
HEXA #4	CACCCTGCCCTTTCACTTC	TACTGGGCCCGTGAACA	T-CACGTG-A
IGF2R	GGTTGGGAACAGTTGCTCTAA	TAGCCGGGCCTCAGTTT	T-CACGTG-A
ILF3 #1	CATCTCTAACCTTCTGTGCCTC	TGAGTACGTTAGCGGAAATGG	A-CACATG-C
ILF3 #3	TGGTAATGGCCGCTCAAA	TCCTTCGCCATCTGACAAC	T-CACGTG-A
ILF3 #4	TCACATGATGGAGCAACCTATC	TGCAAGCGGGAGTGTTT	Y-CACGTG-M
LAMP1 Enhancer	GATCCGTCTGCCCTTTCTCC	AGGTCCTGAAACGGGAAGTG	T-CACGTG-B
LAMP1 Promoter	CGCGGTGTCTTCTTCGTG	CAAGGGCTGGCGTTAGAGAC	T-CACGTG-A
LUC7L2	CTAACTACAACTCCCAGCATACG	ATTTTCGCCTTACCCTATCCC	T-CACGTG-G
NEURL2	GTCCGTAGGGACCAAAGAAG	CTGGAAGTCATGTGTACGAGTC	T-CACGTG-A
PLA2G1B #1	GCCGCCATTCACCAATAGA	CGCCACAGGATGTGATTGT	C-CACGTG-C
PLA2G1B #5	GATCGGGTTGGGCACAC	GCCGAAGCCAAGGAGAAA	T-CACGTG-A
PLD3	TGCGGCCATAACCATCAC	GCCGCTTGCTAGCTTCG	T-CACGTG-A
QTRT1	GCTCCAGAACGGATCGC	GCTCCCGCCATCTTGAC	T-CACGTG-A
RNASEK	CTAAGGCTAGTTCCTCGTCAC	CGTGAGTATTGGAAGAGCGA	T-CACGTG-A
RPUSD4 #4	CTACAATGAGGCGGGAGTG	TCACCGAGTCTCCTCCAAA	T-CACGTG-G
RPUSD4 #5	GCAACCTTGGCCGTTTC	CTCATTGTAGTCGCCCTTCT	C-CACGTG-G
SLC33A1	CGGTCATTGGGCACGAG	GAAACCTACACCCACGAAGAC	T-CACGTG-R
***TYR* Enhancer**	CAATATGGCTACAGC ATTGGAG	TCTCTCCCCTCTATCCTCTCTCT	T-CATGTG-A
***TYR* Promoter**	CCCACTGGTGGGATACGAGCCAA	CCACTCCCACCTCCAGCATCA	W-CATGTG-M
***TYR* 3'UTR**	AGCCTTCTTCCAACTCAGGT	ATGGGCTTAGGGGAAAATGT	
TNPO2	GAAACCCGGGCTGAGAC	GACCGGCTGGCTAGAGA	T-CACGTG-A
USF2 Inr	GCATCTCCTCCTCACATGAC	GGACGGAGGAGATACGGGAG	T-CACATG-A

Gene expression

Target	Forward primers (5'→3')	Reverse primers (5'→3')
18s	CCGGGGAGGTAGTGACGA	AAGGATTTAAAGTGGACTCATTCCA
β-ACTIN	AGAAAATCTGGCACCACACC	GGGGTGTTGAAGGTCTCAAA
ANF	CAGGATGGACAGGATTGGA	TGTCCTCCCTGGCTGTTATC
FASN	CAAGCTGAAGGACCTGTCTAG	CGGAGTGAATCTGGGTTGATG
E-CADHERIN	GGCCAGGAAATCACATCCTA	GGCAGTGTCTCTCCAAATCC
GAPDH	ACCCACTCCTCCACCTTTGA	CATACCAGGAAATGAGCTTGACAA
DCT	GGATGACCGTGAGCTGTGG	CCACAATTATAGCCGGCAAAGT
HIF-1α	CACCTCTGGACTTGCCTTTC	GGCTGCATCTCGAGACTTTT
HPRT	GCTTTCCTTGGTCAGGCAGTA	GGTCCTTTCACCAGCAAGCT

LAMP1	CGAGAAATGCAACACGTTACAG	TGATGTCAGTTATAGATTCCACAGTC
MC1R	GCTGCCAATTTCAGCTTTAGA	ATTGTGCATGCTGCTTTGAG
MITF	GGCCTCACCATCAGCAACTC	TCTCTTTGGCCAGTGCTCTTG
p16	GCCCAACGCACCGAATAG	CGCTGCCCATCATCATGA
p21	CTGGAGACTCTCAGGGTCG	CGGCGTTTGGAGTGGTAGA
p38α	TGAATGAAGACTCTGAGCTGAAGAT	TGGCCACGTAGCCTGTCA
p53	AGGCCTTGGAACTCAAGGAT	CCCTTTTTGGACTTCAGGTG
PGC1α	ACCAAACCCACAGAGAACAG	GGGTCAGAGGAAGAGATAAAGTTG
PKLR	GAGTCTTCCCCTTGCTTTACC	ACAATCACCAGGTCTCCAAC
POMC	CCTCACCACGGAAAGCAA	CCGGGAACATGGGAGTCT
RNAPII	TTGTGCAGGACACACTCACA	CAGGAGGTTCATCACTTCACC
TERT	AAGCTGTTTGCGGGGATT	CCAACAAGAAATCATCCACCA
TBX2	CTGACAAGCACGGGCTTCA	GTTGGCTCGCACTATGTGG
TBX3	AAAAATAGACAACAACCCTTTTGC	TGCAGGGTGAGCTGTTTTC
TERT	AAGCTGTTTGCGGGGATT	CCAACAAGAAATCATCCACCA
TFE3	ATCACTGTCAGCAACTCCTG	CTGTCGTTAATGTTGAATCGCC
TYR	CCTCTAGTCCTCACAAGGTCTGCA	GTGGGATACGAGCCAATTCGA
USF1	ACCCTTATTCCCCGAAGTCAGA	CGACGTCACTTCATTATGC
USF2	GTGACCCAGGTGGGTGTG	TGAAGGGATTTTGGATCACAG

2.9 Antibody list

Antibodies	Supplier	Catalogue numbers	Working dilution (WB)
β-ACTIN	Santa Cruz Biotechnologies	sc-47778 HRP	1:10,000
Acetylated-Lysine	Cell Signalling Techonology	9441	1:1,000
ERK-2	Santa Cruz Biotechnologies	sc-154	1:10,000
FLAG	SigmaAldrich	F1804	1:1,000
GAPDH	Santa Cruz Biotechnologies	sc-32233	1:10,000
HA	SigmaAldrich	H3663	1:1,000
HIF-1α	Home made	PM14	1:1,000
HIF-1β	Novu Biologicals	NB100-110	N/A
HIF-2α	Home made	PM9	N/A
LaminB	Santa Cruz Biotechnologies	sc-6216	1:1,000
MITF (AS2)	Home made	AS2	1:1,000
MITF (AS9)	Home made	AS9	1:1,000
MITF (C5)	Abcam	Ab12039	1:1,000
MITF (D5)	Dako	M3621	1:1,000
MITF (Rb)	Home made	Rb	1:1,000
MYC	SigmaAldrich	M4439	1:1,000
Normal Mouse IgG	Merck Millipore	12-371	N/A
Normal Rabbit IgG	Merck Millipore	12-370	N/A
RNAPII	Santa Cruz Biotechnologies	sc-899x	N/A
USF1	Santa Cruz Biotechnologies	sc-229	1:1,000
USF2 (A19)	Santa Cruz Biotechnologies	sc-30900	N/A
USF2 (Ab)	Abcam	ab60931	1:4,000
USF2 (C13)	Santa Cruz Biotechnologies	sc-30901	N/A
USF2 (C20)	Santa Cruz Biotechnologies	sc-862	1:4,000
USF2 (N18)	Santa Cruz Biotechnologies	sc-861	1:1,000

3. Results

The prerequisites to the “promoter education hypothesis”, as defined in section 1.3, include a TFBS bound by multiple TFs whose binding maybe regulated in a cue-specific manner. Depending on the properties of the TF occupying the TFBS, factor exchange at a promoter may result in a different gene activation status as outlined in Figure 1.3B & C. To study factor exchange and promoter education in humans, we need to identify the genomic loci where multiple bHLH-TFs bind. We chose human melanoma as a model because: 1) Melanoma contains a tissue specific bHLH-LZ-TF, MITF-M whose expression and activity can be induced (see section 1.6, Regulation of MITF). 2) The expression of a number of inducible, pigmentation-related genes have been reported to be dependent on both MITF and USF (Corre and Galibert, 2005), hinting to the involvement of MITF and USF exchange during gene activation. 3) MITF genome-wide ChIP-seq data, as well as our lab’s unpublished insight into mechanisms regulating MITF DNA-binding affinity and specificity, provide us with a good starting point for investigating factor exchange. In addition, MITF, USF1 and USF2 also share related sequence specificity with the MYC and HIF factors adding an additional layer of complexity to the system.

Key questions:

1) The literature indicates that each bHLH-TF exhibits subtle differences in the preferred nucleotide sequence of the core E-box and the flanking sequence. Do these preference pre-define the targets bound by a specific combination of bHLH-TFs? For example, do MYC and USF common sites lack the 5’-T and 3’-A nucleotides flanking the core E-box (5-’V-CACGTG-B-3’) apparently required by MITF? Are MITF and USF specific sites enriched in the 5’-T and 3’-A nucleotides flanking the core E-box (5-’T-CACGTG-A-3’) that prohibit MYC’s binding?
2) Are there any melanoma-specific targets for USF, MYC and HIF? If there are, are these sites also bound by MITF? If so, what are the effects of MITF knock-down on

gene expression and binding by the other bHLH-TFs co-occupying the sites? Perhaps, MITF dictates the melanoma-specific expression of these genes.

3) Can we demonstrate factor exchange and promoter education in the human-melanoma model?

To address the above questions we will:

1) Perform ChIP-seq for each bHLH-TF and the necessary bioinformatics analysis.
2) Analyse the effect of the variation in the core di-nucleotide (5-CANNTG- 3') and the nucleotide immediately flanking the core (5-'N-CACGTG-N-3') E-box element on relative ChIP-seq peak score and the ability to compete against the canonical E-box *in vitro* as well as the potential binding cooperativity of each bHLH-TF.
3) Identify TFBSs co-occupied by a number of bHLH-TFs and investigate any bias in the sequence of E-box found at the common TFBSs.
4) Investigate melanoma-specific binding by each of the bHLH-TF by subtracting 501mel peaks from peaks from other cell lines and extend this to co-occupied sites to explore potential melanoma-specific genes subject to promoter education.
5) Correlate ChIP-seq data with RNA-seq data: A) as a mean of validation for direct regulation of the genes bound by each bHLH-TF, B) to generate a database informing us of the transcription outcome on binding by each bHLH-TF at these genes for subsequent mechanistic analysis of promoter education. We already know that MYC and USF can act as repressors at some genes (see section 1.5, MYC as a repressor and section 1.8, Transcription repression by USF).
6) Explore tools to interfere with DNA-binding by one or more bHLH-TFs and investigate the consequence for promoter occupancy by the remaining factors and gene expression.
7) Investigate potential mechanism(s) regulating factor exchange, using regulation of sub-cellular localisation of USF2 as an example.

3.1 ChIP-seq analysis of bHLH-TFs

To identify common TFBSs we performed genome-wide ChIP-seq of the bHLH-LZ-TFs, MITF, USF1, USF2 and MYC. 501mel melanoma cell lines stably expressing either HA-epitope-tagged MITF, USF1 and USF2 (hereon, HA-MITF, HA-USF1 and HA-USF2, respectively) or FLAG-epitope-tagged MYC (hereon, FLAG-MYC) were generated to increase the efficiency as well as to reduce the background of the ChIP-seq experiments. We also performed ChIP-seq on the endogenous HIF as inducible bHLH-TFs that have been reported to co-regulate a number of genes with MYC (see section 1.7, Cross-talk between HIF and MYC family proteins) and USF (see section 1.9, USF and HIF factor exchange). 501mel cells were chosen because this melanoma cell line has been used both by our lab and our collaborators to perform ChIP-seq and RNA-seq on a number of other factors including RNA polymerase II (RNAPII) and components of the SWI-SNF and Mi-2/NuRD complexes (unpublished).

Stable cell lines were generated and tested as described in the Materials and Methods section. After clonal cell lines were initially validated by immunofluorescence and grown for a further few passages, the cell lines were re-tested for the expression of the epitope-tagged proteins by Western-blot (Figure 3.1.1) and immunofluorescence (Figure 3.1.2). The HA-

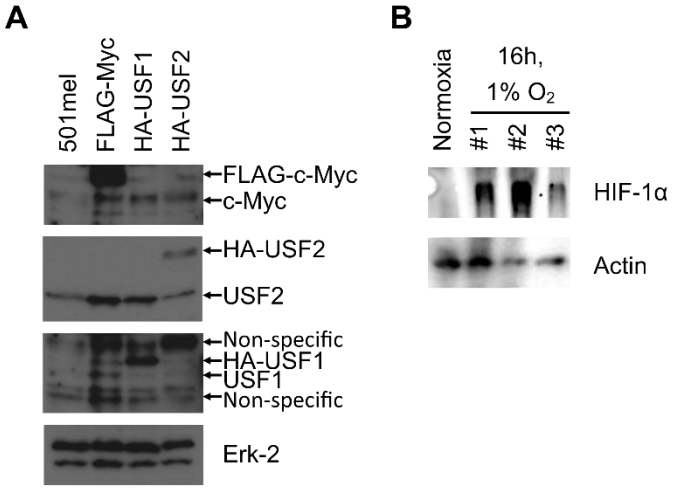


Figure 3.1.1 **A)** Western-blot of whole cell extract from 501mel cell lines stably expressing FLAG-MYC, HA-USF1 and HA-USF2 probed with MYC, USF2 (C20), USF1 (C20) and ERK-2 (loading control). Parental 501mel extract was used as a negative control. **B)** Western blot of 501mel cell line following **16 h**, **1**% O_2 induction, probing for HIF-1α and β-ACTIN (loading control). Treatment was done in triplicate (#1-#3).

MITF stable cell line was generated by Irwin Davidson's lab (Strub et al., 2011) and ChIP-seq experiments were performed in our laboratory (manuscript in preparation). Therefore, the stable cell lines described here will only include HA-USF1, HA-USF2 and FLAG-MYC. After the initial screen using anti-epitope antibody in immunofluorescence, Western-blotting was performed using antibodies raised against the endogenous proteins to enable an estimation of the relative levels of ectopic versus endogenous proteins.

MYC is short-lived and normally only accumulates in proliferating cells (Hurlin et al., 1996; Quéva et al., 1998). However, the FLAG-MYC cell line has a higher level of FLAG-MYC compared to the endogenous MYC (Figure 3.1.1A) when detected using an anti-MYC antibody, possibly reflecting the expression of FLAG-MYC from the highly active *CMV*-promoter. Similarly, the HA-USF1 also exceeds that of endogenous USF1 (Figure 3.1.1A). Intriguingly, both FLAG-MYC and HA-USF1 cell lines appear to have elevated levels of endogenous USF2 compared to the parental 501mel (Figure 3.1.1A), perhaps indicating that MYC and USF1 promote USF2 expression. The HA-USF2 stable cell line, on the other hand, displayed an approximately equal amount of endogenous and exogenous USF2 (Figure 3.1.1A). All stable cell lines express elevated level of MYC relative to 501mel.

It is worth pointing out that although we are primarily interested in identifying genomic loci that can be bound by MYC and USF, and the use of stable cell lines over-expressing MYC and USF1 may enable identification of sites normally only bound when MYC and USF proteins level are induced, the significantly overexpressed level of MYC and USF1 will also result in increased binding to low-affinity targets which may not be physiologically relevant. The latter point is of particular relevance as we use peaks called by these cell lines overexpressing MYC and USF1 for *de novo* motif identification and use these results for detail analysis of target sequence preference within the core E-box and the sequence immediately flanking the E-box. Notwithstanding all these caveats, if we are still able to identify differences in the E-box sequence specificity between different bHLH-TFs studied, the differences will most likely be significant.

Immunofluorescence using anti-epitope-specific antibodies revealed that all three epitope-tagged proteins are localised to the nucleus, (Figure 3.1.2). Note, however, that while MYC is expressed to a similar level in all cells, with the exception of those in mitosis, USF1 exhibits variable expression with cells containing larger nuclei appearing to express a greater amount of protein. This possibly indicates cell cycle related expression of USF1 at

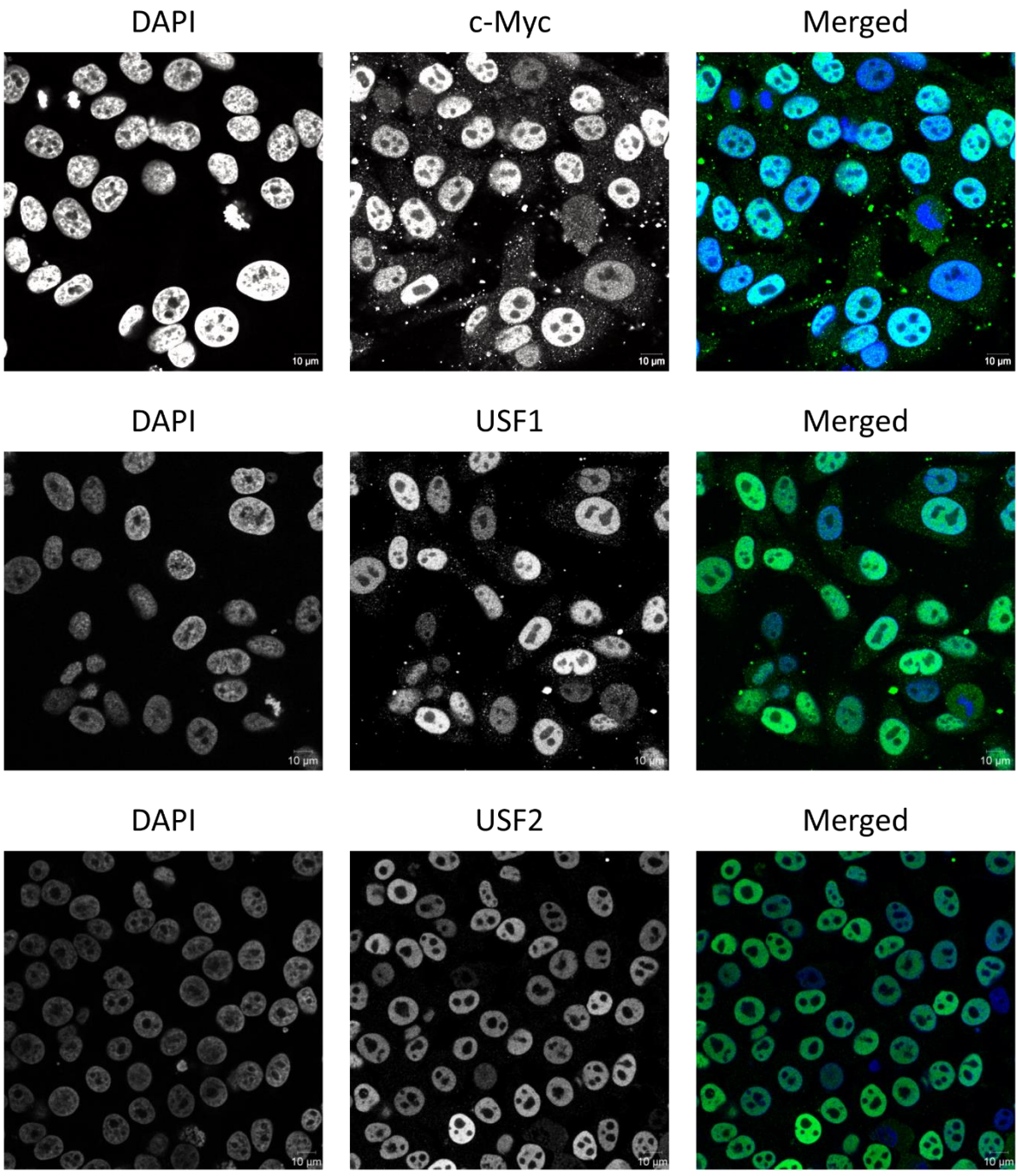


Figure 3.1.2 Immunofluorescence of stable cell lines derived from 501mel cells. 501mel cell lines stably expressing FLAG-MYC, HA-USF1 and HA-USF2 were stained using anti-FLAG- and anti-HA-antibodies after clonal expansion and grown for a further few passages to confirm the maintenance of exogenous proteins expression. Cells were also stained with DAPI to mark the nuclei. Scale bar is shown at the bottom right corner.

the protein level, since MYC and USF are both driven by the *CMV*-promoter. Any variable expression in USF2 expression was less easy to attribute to cell cycle–related events.

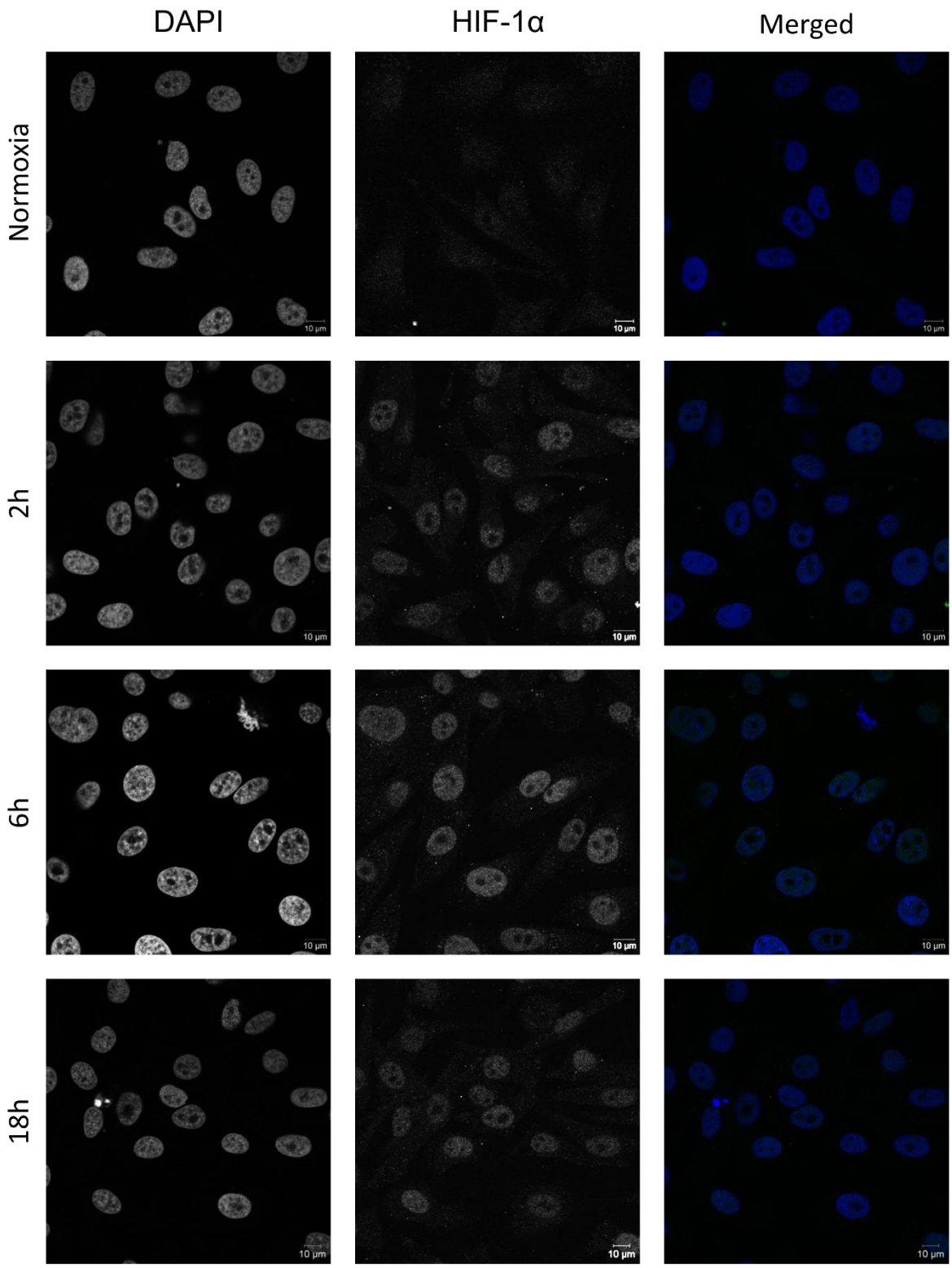


Figure 3.1.3 Immunofluorescence of 501mel cells following time course of **1**% O_2 and stained for HIF-1α. Cells were also stained with DAPI to mark the nuclei. Scaled bar is shown in white at the bottom right corner.

Cell lines stably expressing epitope-tagged HIF were not generated as HIF-α proteins are rapidly degraded under normoxic conditions (see section 1.7, Regulation of HIF-α proteins). As HIF-2α has a more restricted expression pattern but may be expressed in some tumour

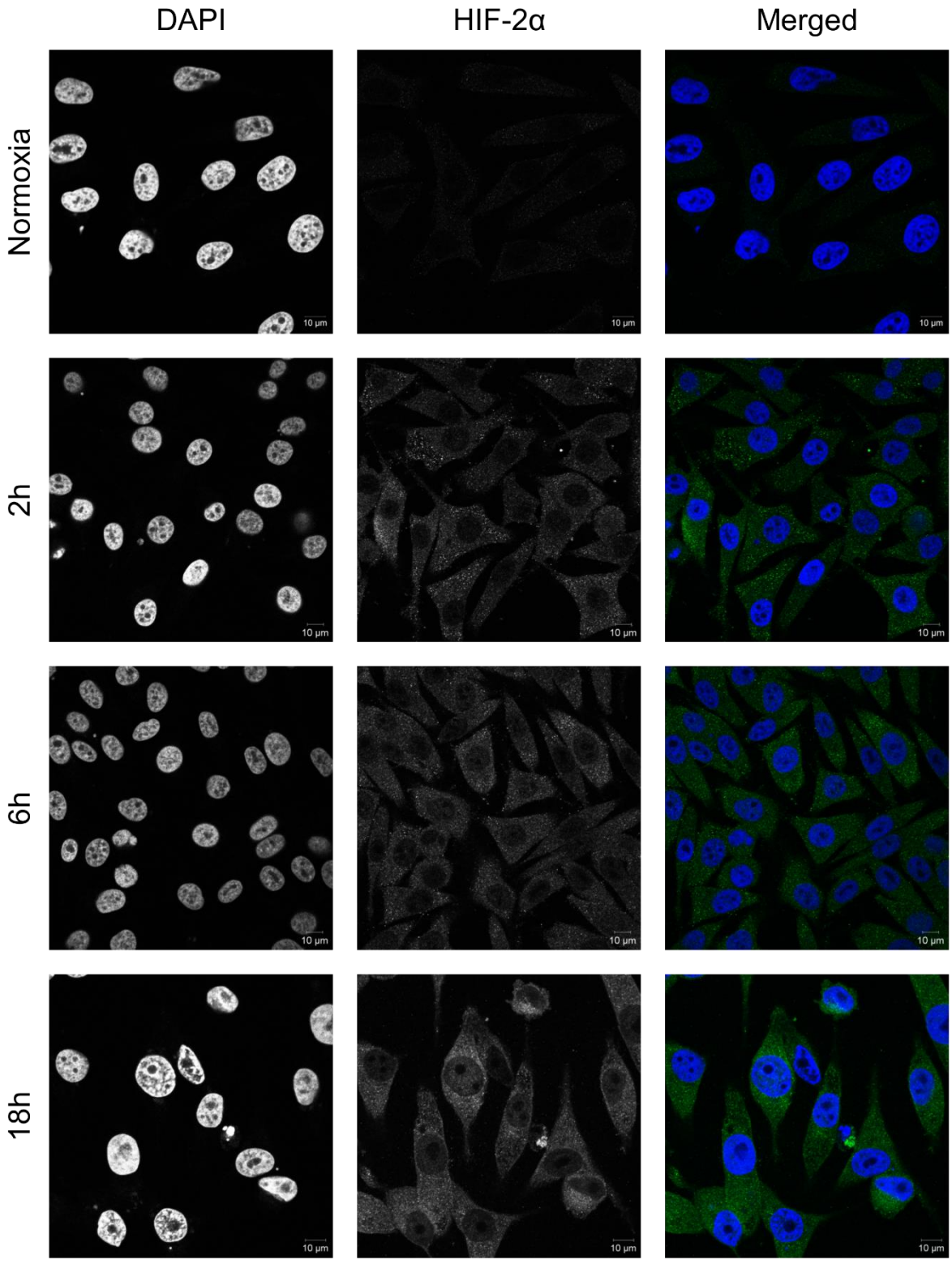


Figure 3.1.4 Immunofluorescence of 501mel cells following time course of **1**% O_2 and stained for HIF-2α. Cells were also stained with DAPI to mark the nuclei. Scaled bar is shown in white at the bottom right corner.

cells (Semenza, 2003), we tested for the inducibility of HIF-1α and HIF-2α in 501mel cell lines following a time course of 1% O_2 exposure (hypoxic treatment) (Figure 3.1.3 – 3.1.4). In 501mel cells, HIF-1α gradually accumulated over the time course of exposure to a hypoxic environment and becoming more prominent from 6 h onward. HIF-1α is nuclear localised at all times. HIF-2α, on the other hand, is visible within the first 2 h of hypoxic treatment and is mainly localised to the cytoplasm, though by 18 h a significant fraction is localised to the nucleus. Only the Western-blot of HIF-1α (Figure 3.1.1B) was carried out as no known reliable commercial antibodies were available that function in Western-blot for the other HIF at the time of experiment (Hammond lab and Radcliffe lab, personal communication).

Basic statistics of the ChIP-seq experiments

501mel cells stably expressing the epitope-tagged TFs were ChIPped and processed for high-throughput sequencing and sequencing data peak-called as described in the Materials and Methods. An example of a "peak" is visually demonstrated in Supplementary Figure 3.1.1A. The position-adjusted reads were reported as a "peak score" which is correlated to the number of reads found in the peaks after normalisation of total mapped reads to 10-million. In an ideal situation where peak score correlates with binding incidence of the ChIPped-protein, we can directly correlate peak score to binding affinity across different targets within each dataset. In reality, however, the amount of ChIPped protein at each locus is influenced by a number of factors including: accessibility of the target proteins to cross-linking agent as well as the epitope targeted by the antibody against the protein of interested. In the case where the protein of interest forms different functional complexes at different subsets of targets, cross-linking efficiency as well as epitope masking may influence the final amount of protein ChIPped at different loci. Further consideration needs to be made when comparing different ChIP-seq datasets including the relative amount of different bHLH-TFs being studied in different cell lines. As the peak score is also a function of the relative binding at each ChIPped locus, the relative amount of bHLH-TF of interested

versus other bHLH-TFs able to compete for those same sites will influence the final amount of target proteins ChIPed. In the case where these other bHLH-TFs have overlapping but not identical binding specificity, these other bHLH-TFs will influence binding of the bHLH-TF in question to different target E-boxes to different degrees.

As there is no easy and established standard for comparing different ChIP-seq datasets, we proceeded to analyse our datasets with the starting assumption that peak scores correlate to binding affinity (a measure of binding kinetics and thermodynamics), at least within a single dataset. In any case, the caveats of the ChIP-technique apply to both ChIP-seq and ChIP-qPCR. However, other factors such as bias in the library preparation and PCR amplification efficiency may also contribute to discrepancies between the two techniques. For the comparison of different ChIP-seq datasets, we attempted a simple centre normalisation approach which reports peak score for each locus as a value relative to the average peak score (Supplementary Table 3.1.1) of all the peaks used in each dataset. This normalisation should take into account the difference in background level for each experiment (by only considering the relative values of normalised read counts within the peaks) as well as the difference in the number of peaks-called between each dataset (due to the fact that we normalised total mapped reads to 10-million, assuming the same ChIP efficiency, datasets with lower number of peaks will have higher average peak score).

Visual inspection of the bedgraph uploaded onto the UCSC genome browser revealed that many peaks with score of < 10 tend to have a relatively broad profile and are sometimes difficult to distinguish from the surrounding region (exemplified in Supplementary Figure 3.1.1B). This may be explained by the fact that both background files used were prepared using a different protocol by a past lab member. This may result in different biases in the contaminant background.

To filter out some of the low affinity sites and false peaks, we filtered out peaks with peak score of ≤ 10 in most analyses. The use of a full- or filtered-dataset will be clearly indicated

(hereon, full-dataset and filtered-dataset respectively). In the case of the MITF ChIP-seq that has a low average peak score of 2.31, the top 6% of peaks were used to limit the unfiltered total of $99{,}380$ peaks down to $\approx 5{,}700$ peaks, comparable to that of the filtered-USF datasets (Supplementary Table 3.1.1). The low peak score for MITF can be explained by a combination of a much higher number of peaks called, and a lower average peak focus (which reports the tightness of the peaks).

A lower peak score most likely reflects binding to low affinity TFBSs, perhaps presented as tandem E-boxes as the average peak focus is higher when looking at the peaks from a filtered-dataset compare to a full-dataset (Supplementary Figure 3.1.2C & 3.1.2D). MITF is the only exception, where the average peak focus decreases for the top 6% of peaks ranked according to the peak score and may reflect cooperative binding to adjacent E-boxes giving rise to a high peak score.

Peaks called were stratified into different genome ontologies as either 5'-UTR, promoter, exon, intron, 3'-UTR, TTS, intergenic or non-coding (Supplementary Figure 3.1.3) . Interestingly, for the conditionally-induced HIF, a higher proportion of peaks were located within the promoter regions of genes when compared to the ubiquitously expressed USF and MYC ($\approx 40\%$ vs $\approx 30\%$). Another differences also lies in the $\approx 10\%$ higher binding of USF and MYC to intergenic regions compared to the HIF. One possible explanation is that the HIFs, as conditionally-induced TFs only need to bind to their target loci following induction before being rapidly degraded, whereas it may be advantageous for the ubiquitously expressed TFs like the USF to remain bound to the genome even when not anchored to their cognate TFBSs to facilitate rapid binding of USF to target E-boxes when those sites become available (based on a two-step binding model where TFs first scan the genome then anchor down at specific sites). This notion is supported by the genome ontology of MITF where $\approx 50\%$ of peaks were located within intergenic regions. Non-promoter binding in all cases may also reflect binding to the distal regulatory elements or cryptic promoters with no reported transcripts.

Based on the observed difference in the distribution of peaks into different genome ontologies by each bHLH-TF, we performed *de novo* motif identification on peaks stratified into different genome ontologies. To investigate potential differences in enriched motifs among peaks with high peak scores, we also compared motifs identified from full- and filtered-datasets. It is worth pointing out that although Homer uses sequences ± 50 kb from the TSS to establish the background file for motif identification, this should not biasedly affect motifs identified from different genome ontologies to a different degree. This is because the average gene size in humans is between $10 - 15$ kb and the average gene density is about one per $40 - 45$ kb. This is within the ± 50 kb from TSS used. Moreover, Homer selects region with a similar GC-content as the input sequence which is probably the major source of nucleotide bias around TFBS. Nonetheless, we should keep in mind that the background file used may give a biased result if a certain motif is only found at specific genome ontologies. This also applies to co-occurring motifs.

To unequivocally validate our ChIP-seq data without biasedly selecting for published targets we intersected MITF, USF and MYC narrow-peak files (called using MACS 2.0 against HA-background) and randomly selected subsets of genomic loci that are co-bound by one or more bHLH-LZ-TFs studied and have a different number of E-boxes for ChIP-qPCR analysis (Supplementary Table 3.1.2 & 3.1.3). It is important to point out that these targets were obtained from the MACS peak-called using the first biological replica ChIP-seq datasets [although the peak score reported in Supplementary Table 3.1.3 are values from merged tag-files of all biological replicate] while the ChIP-qPCR shown in this results section were performed using samples prepared for the second biological replicate of the ChIP-seq experiments, and all the genome-wide analyses were performed by merging all biological replicas before peak-calling as described in the Materials and Methods. As the two biological replicates do not give exactly the same peaks called using MACS 2.0 against the HA-background, with most irreproducibility occurring among peaks with a lower MACS "pile up" value (data not shown), the targets chosen for our ChIP-qPCR are not strictly

qualified as validation of ChIP-seq, and any differences observed may partially be attributed to the experimental variation between the two replicates.

Lastly, we performed gene ontology (GO) analysis of the top 500 genes, ranked by peak score using AmiGo and GO-terms derived from the whole dataset using Homer. The results are presented as summary tables grouping related GO-processes.

3.1.1 MITF ChIP-seq

With the exception of the more recently identified isoforms, isoform Mdel and isoform-12, all MITF isoforms (MITF-A, MITF-M, MITF-H and MITF-C) share common exon $2-8$ with alternative initial exon-1 usage and alternative splicing that adds or deletes six amino-acids ($+$ and $-$ isoforms, respectively) just before the basic domain (for review see Shibahara et al., 2001). We are interested in the melanocyte-lineage-specific isoform, isoform-M, whose expression is dependent on a melanocyte-restricted promoter (Price et al., 1998b). Ectopic expression of MITF in fibroblasts can induce a melanocytic phenotype (Tachibana et al., 1996) while ectopic expression of MITF in the neuroretina can induce transdifferentiation into melanocytes (Planque et al., 1999). Misexpression of MITF in zebrafish induces ectopic melanised cells (Lister et al., 1999). Collectively, the evidence implicates MITF as a key regulator of the melanocyte-lineage.

MITF-M on its own cannot induce oncogenic transformation in cell lines or tumours in nude mice (Tachibana, 1997). If anything, ectopic MITF expression via retroviral vector significantly increases apoptosis in cultured mast cells from mice (Tsujimura et al., 1997). Yet, MITF expression appears to be maintained in most malignant melanomas even when expression of many melanocytic markers is lost (King et al., 1999; Salti et al., 2000). This implies that MITF regulates a number of genes critical for the survival and progression of melanoma that may not necessarily have the canonical M-box, and that these E-boxes are likely to be of higher affinity than the canonical M-box as their expression is maintained when the M-box containing melanocytic gene expression is lost. These genes are likely to

also be regulated by other bHLH-TFs such as USF and MYC. To this end, I re-analysed the ChIP-seq data performed by Siddaway in our lab (manuscript in preparation) using the same parameters as the other ChIP-seq datasets for direct comparison.

The basic bioinformatics analysis of MITF ChIP-seq can be founded in Supplementary Figure 3.1.4 & 3.1.5. Briefly, the average peak score of MITF peaks do not differ significantly across different genome ontologies (Supplementary Figure 3.1.4A). The majority of the outliers with very high peak scores are located within intergenic regions and these may represent distal regulatory elements. Although the top 6% of the peaks display a slight enrichment in the binding to promoter regions (10%, Supplementary Figure 3.1.4B) when compared to the whole dataset (6%, Supplementary Figure 3.1.3A), the majority of binding still occurs within introns and intergenic regions.

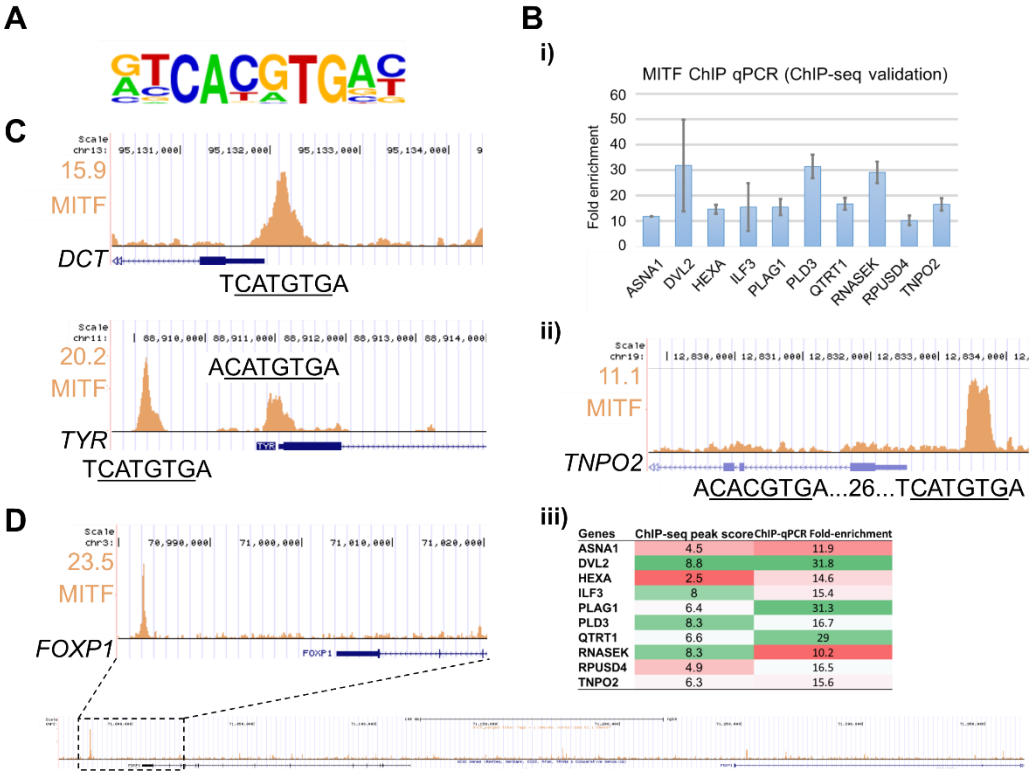


Figure 3.1.5 MITF's motif and targets **A)** The top motif discovered *de novo* using MITF ChIP-seq full-dataset. **B) i.** ChIP-qPCR at a subset of positive targets identified from the first biological replicate of ChIP-seq samples. Fold enrichment is calculated using $\Delta\Delta C_t$ approach as described in Materials and Methods. $\mathbf{N = 2}$ biological replicates each with **2** systematic replicates. **ii.** Representative snapshot of MITF bedgraph uploaded on to the UCSC genome browser at one randomly chosen target analysed by qPCR. **iii.** Table comparing ChIP-seq peak score against ChIP-qPCR fold-enrichment at indicated genes. To aid visual comparison the table are colour-coded such that the darkest green correspond to highest value while the darkest red correspond to lowest value in each column. **C)** Snapshot of an MITF bedgraph uploaded on to the UCSC genome browser at two known MITF targets, the *DCT* and *TYR* genes. The E-boxes underneath the peak are shown. The core nucleotides of the E-boxes are underlined. **D)** Same as **C** for a MITF specific target *FOXP1* gene (no E-box, not bound by MYC, USF or HIF). To illustrate the quality of the ChIP-seq dataset, a broader region around the peak is also shown.

Detailed analysis of the different motifs enriched across different genome ontologies reveals a clear preference for the 5'-CACGTG-3' E-box in promoter regions (Supplementary Figure 3.1.3C & 3.1.4). This supports the notion that MITF binds to a plethora of genes other than the M-box containing melanocytic genes. At other genome ontologies, however, there appeared to be a greater flexibility in the core di-nucleotide with the motif identified being 5'-CA[C/T][A/G]TG-3'. All the E-boxes identified contain either a 5'-T or 3'-A nucleotide flanking the core E-box, and these may be extended to 5'-GT-CACGTG-A-3' at some genome ontologies.

It may be tempting to interpret the motif bound by MITF shown in Figure 3.1.5A as either 5'-CACGTG-3' or 5'-CATATG-3' as MITF binds as a homodimer and hence ideally should prefer a symmetric E-box. However, this is certainly not the case as the canonical M-box is the asymmetric 5'-CATGTG-3'. This would imply that either MITF has a relaxed requirement for the core di-nucleotide of the core E-box or that MITF may be subjected to PTM(s) that modulate MITF binding specificity. In the latter scenario, modification of one of the MITF proteins within an MITF homodimer would result in a protein with specificity for an asymmetric E-box.

ChIP-qPCR at a number of target E-boxes was carried out to confirm the validity of ChIP-seq (Figure 3.1.5B, i). An example snapshot of an MITF bedgraph uploaded onto the UCSC genome browser at one of the targets analysed by ChIP-qPCR is shown in Figure 3.1.5B, ii. Although we were able to confirm binding at all targets chosen, ChIP-qPCR fold-enrichment did not always correlate well with the ChIP-seq peak score (Figure 3.1.5B, iii & Supplementary Table 3.1.3). We also cross-checked our ChIP-seq data with some of the published MITF targets, namely *dopachrome tautomerase* (*DCT*) and *TYR* (Figure 3.1.5C)

Table 3.1.1 Summary table of Gene-ontology (GO) analysis for MITF filtered dataset.

GO-terms with P < 0.05 derived from top 500 peaks using AmiGo	Total no. of term(s)	4
	No. of pigmentation related term(s)	1
Pigmentation related Go-terms derived from filtered dataset	No. of pigmentation related term(s)	23
	Average Coverage (%)	93.4
Top 50 GO-terms for filtered-dataset	No. of genes	11429
	No. of pigmentation related term(s)	0
	Average Coverage (%)	82.7

(Hodgkinson et al., 1993) and also identified *FOXP1* as a MITF-specific target which is not bound by MYC, HIF and USF in 501mel cells and lacks any apparent E-box (Figure 3.15D).

GO-analysis of the top 500 ChIP-seq peaks revealed the term "pigmentation" as fourth ranking. However, inspection of the top 50 GO-terms derived from the whole filtered-dataset did not reveal any term related to the pigmentation process (Table 3.1.1). The average coverage of genes in the top 50 terms is 82.7%, supporting the notion that MITF regulates a plethora of processes in addition to pigment production. Nonetheless, the role of MITF in pigmentation-related processes is affirmed by the high coverage of (93.4%) genes in pigmentation-related term (Table 3.1.1).

3.1.2 MYC ChIP-seq

MYC was proposed to function as a sensor of mitogen and nutrient availability (Dang, 1999) and is essential for the expression of many genes involved in cell cycle, proliferation and metabolism (Coller et al., 2000). MYC directly regulates a plethora of biosynthesis pathways that support growth (Dang et al., 2009; Sloan and Ayer, 2010). Reviews of MYC's role in DNA replication, transcription and splicing and nuclear structure can be found in Facchini and Penn, 1998; Luscher and Eisenman, 1990. MYC also regulates genes indirectly through microRNAs. For example, MYC represses miR-23a/b that suppresses glutaminase gene expression (Dang et al., 2009).

MYC is generally found in proliferating cells (Hurlin et al., 1996; Quéva et al., 1998) and is down-regulated during differentiation while other MAX dimerization partners, MAD and MXI are up-regulated (Ayer and Eisenman, 1993; Ayer et al., 1993; Larsson et al., 1994; Zimmerman et al., 1986). Overexpression of the MAD family (Chin et al., 1995; Hurlin et al., 1997; Hurlin et al., 1996; Koskinen et al., 1995; Lahoz et al., 1994) and MNT (Hurlin et al., 1997) can antagonize MYC-induced cellular transformation and cell cycle progression (Chen et al., 1995). This highlights the importance of a regulated temporal exchange of MYC and MAD (Spencer and Groudine, 1991) at target E-boxes (Figure 1.12).

The average peak score of the MYC ChIP-seq is highest at promoter regions and the outlier peaks with very high peak score spread between the promoters, introns and intergenic regions (Supplementary Figure 3.1.6A). A higher proportion of peaks are located within the promoter regions when comparing filtered dataset (41%, Supplementary Figure 3.1.6B) to the whole dataset (28%, Supplementary Figure 3.1.3B).

The consensus motif of MYC identified *de novo* from the whole dataset reveals a 5'-CA[C/T][G/A][T/C]G-3' core E-box specificity (Figure 3.1.6A). Closer inspection reveal a clear bias towards 5'-CACGTG-3' around the promoter regions (Supplementary Figure 3.1.6C & 3.1.7). The E-box motifs found in intron and intergenic regions are relatively degenerate while binding at other genome ontologies lack E-box motifs altogether. This is in contrast to what was observed for MITF (Supplementary Figure 3.1.4C), suggesting that

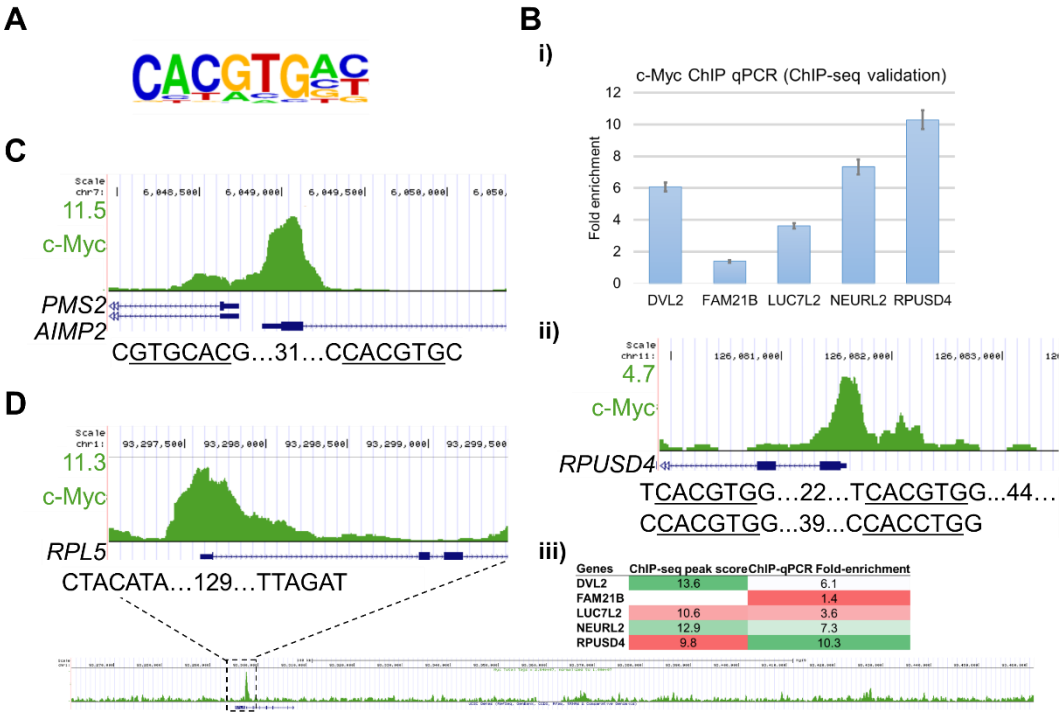


Figure 3.1.6 MYC's motif and targets **A)** The top motif discovered de novo using full MYC ChIP-seq dataset. **B) i.** ChIP-qPCR on a subset of positive targets identified from MYC ChIP-seq. Fold enrichment is calculated using $\Delta\Delta C_t$ approach as described in Materials and Methods. $\mathbf{N = 2}$ biological replicates each with **2** systematic replicates. **ii.** Representative snapshot of MITF bedgraph uploaded on to the UCSC genome browser at one of the target, randomly chosen, analysed by qPCR. The E-boxes underneath the peak are shown. The numbers between TFBSs indicate the distance between the TFBSs. The core nucleotides of the E-boxes are underlined. **iii.** Table comparing ChIP-seq peak score against ChIP-qPCR fold-enrichment at indicated genes. To aid visual comparison the table are colour-coded similar to heat map such that the darkest green correspond to highest value in each column and darkest red correspond to lowest value in each column. **C)** Snapshot of a MYC bedgraph uploaded on to the UCSC genome browser at a known MYC target, *AIMP2*. **D)** Same as **C)** for *RPL5*, a MYC unique target with no E-box. The initiator sequences under the peak are shown. To illustrate the quality of the ChIP-seq dataset, a broader region around the peak is also shown.

binding of MYC to regions other than promoters may either be E-box-independent or that MYC has the capability to bind to much more degenerate motifs than MITF.

In line with the reported interaction of MYC and SP1 at both active (To et al., 2006) and repressed (Koshiji et al., 2004) genes in competition with HIF-1α, the SP1 recognition motif was discovered as the motif with the second highest probability to the E-box around the promoter (Supplementary Figure 3.1.7).

With regard to the flanking sequences, it is apparent, even from the full-dataset (Figure 3.1.6A) that, unlike MITF, MYC's E-box is asymmetric. Closer inspection at peaks with score ≥ 10 reveal a strong 5'-GT bias, most likely reflecting the different target specificity confer by MYC and MAX.

ChIP-qPCR on a number of target E-boxes was carried out to confirm the validity of ChIP-seq (Figure 3.1.6B, i). An example snapshot of a MYC bedgraph uploaded onto the UCSC genome browser at one of the target analysed by ChIP-qPCR is shown in Figure 3.1.6B, ii . Although we were able to confirm enrichment at the expected targets, unlike the MITF dataset, the relative fold-enrichment does not correlate that well with the relative peak score (Figure 3.1.6B, iii & Supplementary Table 3.1.3).

We also cross-checked our ChIP-seq with some known MYC targets genes involved in protein synthesis. MYC induces components necessary for translation (tRNA and rRNA, RNAP-I (Arabi et al. 2005) and RNAP-III-dependent transcription (Gomez-Roman et al. 2003) as well as components necessary for initiation of cap-dependent transcription initiation (RNAPII-dependent transcription) (Coller et al. 2000). Shown in

Table 3.1.2 Summary table of Gene-ontology (GO) analysis for MYC filtered dataset.

GO-terms with P < 0.05 derived from top 500 peaks using AmiGo	Total no. of term(s)	35
	No. of metabolism related term(s)	16
	No. of biosynthetic related term(s)	3
Top 50 GO-terms for filtered-dataset	No of genes	4916
	No. of metabolism related term(s)	19
	No. of biosynthetic related term(s)	8
	No. of gene expression related term(s)	4
	Average Coverage (%)	38.4

Figure 3.1.6C & 3.1.6D are translation-related genes which contain an E-box or an initiator element respectively. Although none of the translation-related terms comes up when searched with AmiGo using the top 500 peaks with $P < 0.05$, a number of translation- and protein metabolism-related process terms come up a few times among the top 50 GO-term searched using the whole filtered-dataset. The most enriched terms are related to metabolism and biosynthesis (Table 3.1.2).

3.1.3 USF ChIP-seq

USF1 and USF2 are ubiquitous bHLH-LZ-TFs that bind DNA as either homo- or heterodimers (Sirito et al., 1992). USF represents the main MYC E-box binding proteins in nuclear extracts (Sommer et al., 1998). In a HeLa nucleus, the USF concentration was estimated to be $0.5\ \mu M$, corresponding to $\approx 20{,}000$ molecules per cells (Sawadogo et al., 1988). Although the USF genes are ubiquitously expressed and appear to function mostly in the context of differentiated cells, they are required for development and expression of many tissue-specific genes (Gao et al., 1997; Harris and Mellon, 1998; Bengtsson et al., 2002).

USF1 and USF2 share a highly similar sequence around the DNA-recognition basic-region. Therefore, one expects them to regulate similar sets of gene. However, there are reports of differential gene regulation by USF1 and USF2. For example, the *IGF2R* promoter is bound by USF *in vivo* in the non-tumorigenic mammary epithelial cell line MCF−10A, yet only USF2 transactivated the *IGF2R* gene (Szentirmay et al., 2003). We are therefore, interested in identifying USF1 and USF2 common and unique targets, both in terms of DNA-binding and gene expression within our melanoma system.

USF1

The average peak score of USF1 ChIP-seq is highest around the promoter regions and TTS (Supplementary Figure 3.1.8A). This may potentially reflect binding of USF1 to both promoters and terminators of the same genes (Supplementary Figure 3.1.10) either to drive

anti-sense transcription or perhaps forming a loop between the promoter and TTS. Outlier peaks with higher than average peak scores are quite well spread out among all genome-ontology terms. A higher proportion of peaks are located within the promoter regions when comparing filtered dataset (44%, Supplementary Figure 3.1.8B) to the whole dataset (31%, Supplementary Figure 3.1.3C).

The consensus motif of USF1 identified *de novo* from the whole dataset is 5'-[A/G]T-CACGTG-3' (Figure 3.1.7A). The preference for 5'-T is presented through most genome ontologies and appeared more prominent when including peaks with a lower peak score (compare Supplementary Figure 3.1.8C to Supplementary Figure 3.1.9). The preference for 5'-[A/G] at the -2-position flanking the core E-box is not present at the 5'UTR where the preference for 5'-T immediately flanking the core E-box is also weak. In general, the presence of 5'-T and/or 3'-A is found to be more abundance when including peaks with peak score ≤ 10. This may be interpreted as lower relative binding due to competition with MITF that favours 5'-T and 3'-A nucleotides flanking the core E-box. With regard to the core di-nucleotide, USF1 appears to bind solely to the 5'-CG-3'.

ChIP-qPCR on a number of target E-boxes was carried out to confirm the validity of the ChIP-seq including *ASNA1* which was not peak-called on the USF1 ChIP-seq but was chosen for ChIP-qPCR as USF2 was peak-called at the same locus. Closer inspection of the *ASNA1* bedgraph revealed a very broad USF1 peak compared to USF2 (Figure 3.1.7B). Detailed examination of the underlying sequence revealed two E-boxes 5'-T-CACGTG-A-3' and 5'-G-CACGTG-A3' separated by 570 bp. It is possible that USF2 only binds to one of the E-boxes, resulting in a normal, tight peak while USF1 may bind to both E-boxes, thus giving a broad peak profile that resulted in no *ASNA1* peak being called using our algorithm. Indeed, ChIP-qPCR confirms enrichment at all the targets chosen including *ASNA1* (Figure 3.1.7C, i), although the relative fold-enrichment did not all correlate with the normalised peak score (Figure 3.1.7C, ii).

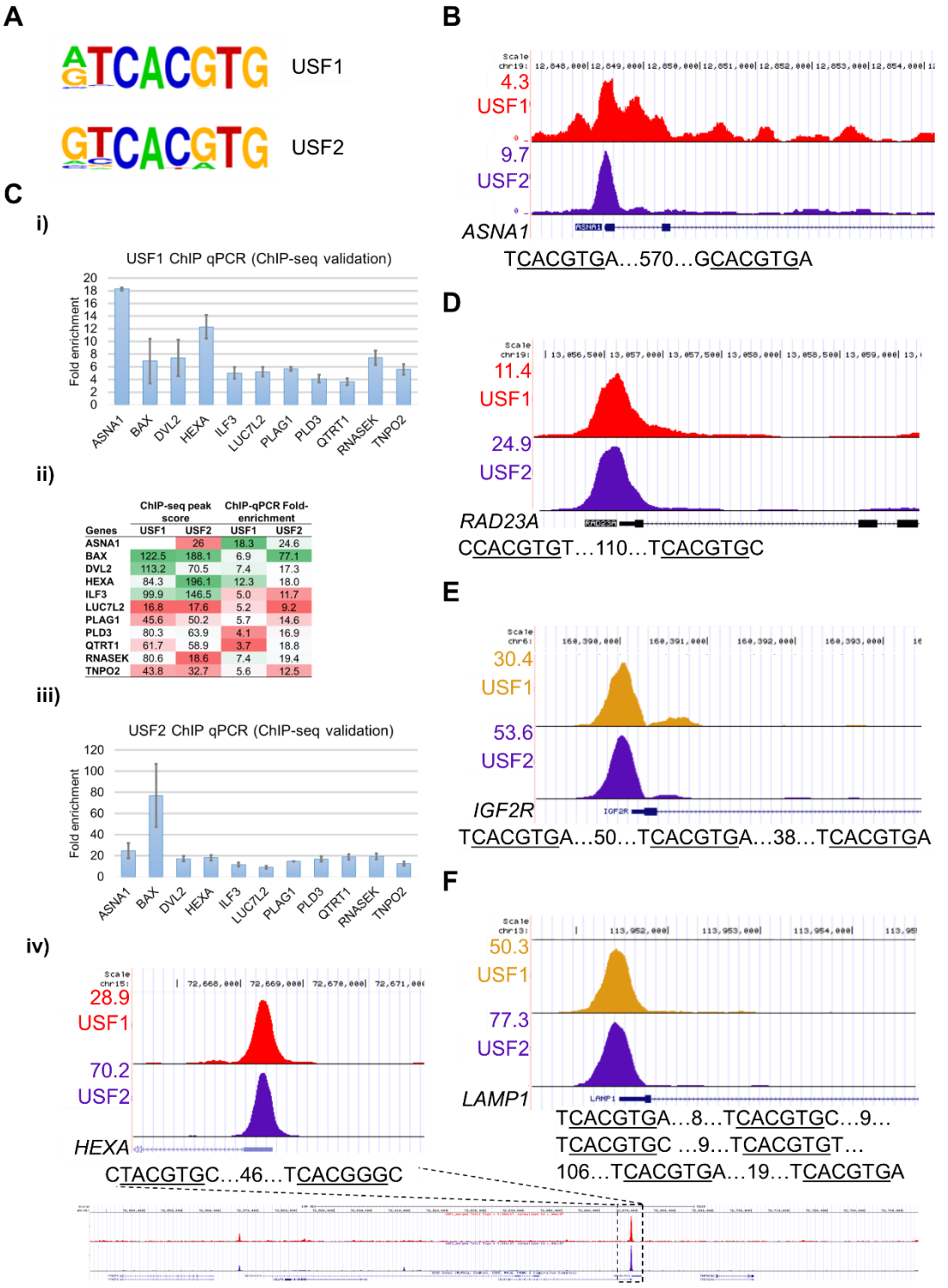


Figure 3.1.7 USF's motifs and targets **A)** The top motifs discovered *de novo* using the whole USF ChIP-seq dataset. **B)** Snapshot of USF1 and USF2 bedgraphs uploaded on to the UCSC genome browser at *ASNA1* gene. **C) i.** ChIP-qPCR on a subset of positive targets identified from the ChIP-seq of USF1. **ii.** Table comparing ChIP-seq peak score against ChIP-qPCR fold-enrichment at indicated genes. To aid visual comparison, the table are colour-coded such that the darkest green correspond to highest value in each column and darkest red correspond to lowest value in each column. **iii.** Same as **i.** for USF2. Fold enrichment is calculated using $\Delta\Delta\boldsymbol{C_t}$ approach as described in Materials and Methods. $\mathbf{N} = \mathbf{2}$ biological replicates each with **2** systematic replicates. **iv.** Representative snapshot of USF bedgraphs uploaded on to the UCSC genome browser at one of the target, randomly chosen, analysed by qPCR. **D)** Same as **B** at a known USF targets, *RAD23A*. **E)** Same as **B** for *IGF2R*, a known USF target with differential expression outcome when bound by USF1 or USF2. **F)** A novel USF target with numerous E-boxes underneath the peak, *LAMP1*. The E-boxes underneath the peak are shown. The numbers between the E-boxes indicate the distance between the E-boxes. The core nucleotides of the E-boxes are underlined.

Table 3.1.3 Summary table of Gene-ontology (GO) analysis for USF1 filtered dataset.

GO-terms with P < 0.05 derived from top 500 peaks using AmiGo	Total no. of term(s)	75
	No. of metabolism related term(s)	33
	No. of catabolism related term(s)	12
	No. of biosynthetic related term(s)	7
Top 50 GO-terms for filtered-dataset	No of genes	6675
	No. of metabolism related term(s)	20
	No. of biosynthetic related term(s)	5
	No. of sub-cellular localisation related term(s)	8
	Average Coverage (%)	51.9

We also cross-checked our ChIP-seq data against some published targets of USF (Figure 3.1.7D & 3.1.7E). For example, USF1 is critical for the maintenance of genome integrity in response to UV-induced DNA-photo-lesions through activation of two nucleotide-excision repair factors CSA and HR23A that are not targets of p53 (Baron et al., 2012). In addition to USF1, we also show binding of USF2 at *HR23A* (aka *RAD23A*) (Figure 3.1.7D). In addition to the known targets we also identify a novel USF target, *LAMP1* that encodes a protein involved in lysosome biogenesis (Figure 3.1.7F).

GO-analysis of USF1 top 500 genes revealed terms involved in DNA-repair. However, highest enrichment was found among metabolic- and catabolic-processes (Table 3.1.3). Whilst metabolic related terms are among the most enriched when considering the whole filtered-dataset, catabolism related terms are not. Instead, terms related to sub-cellular localisation are the second most abundant. The average coverage of the top 50 GO-terms is 51.9%, supporting the role of USF as a ubiquitous TF and not a master regulator of specific biological processes.

USF2

Four alternative spliced variants of USF2 are deposited in the Uniprot database (USF2A, USF2AΔ, USF2b and USF2c). USF2b (38 kDa) lacks the 67 –amino-acid fourth-exon and correlates with less transactivation potential at the *AdML* promoter (Viollet et al., 1996). USF2b can act as a dominant-negative regulator of USF-mediated activation of the major histocompatibility complex class I promoter (Howcroft et al., 1999). Within exon−4 resides a 27 –amino-acids region unique to mammalian USF2 (Fujimi and Aruga, 2008). As the

long USF2A (44 kDa) represents the major isoform in mammals, USF2's function as a TA is most likely recently acquired. Our analysis entails the long isoform, USF2A.

The average peak score of USF2 ChIP-seq is highest around promoters, 3'UTRs and TTSs (Supplementary Figure 3.1.11A) and may serve a role in DNA-looping as we hypothesised for USF1 (Supplementary Figure 3.1.10). The majority of the outliers with high peak scores are found distributed between promoters, introns and intergenic regions. Intronic and intergenic-binding may represent binding to distal enhancer regulatory elements. Alternatively, they may serve as a reservoir for USF2 when USF2 is not bound to promoters E-boxes, as hypothesised for MITF above. In support of this notion is the observation that the distribution of USF2 peaks among different genome-ontologies of USF2 shifts mainly from intergenic to promoter regions when comparing genome-ontologies derived from the peaks with score of ≥ 10 (25%, intergenic, 37%, promoter, Supplementary Figure 3.1.11B) to genome-ontologies derived from the whole dataset (32%, intergenic, 29%, promoter, Supplementary Figure 3.1.3D). To allow for rapid re-binding of USF2 to the promoter E-boxes, the intergenic target must have lower affinity for the shift to be thermodynamically favourable.

The consensus motif of USF2 identified *de novo* from the whole dataset is almost identical to that of USF1 with the 5'-T bias (Figure 3.1.7A) as expected based on the identical amino-acids in the basic domain (Figure 1.6). However, USF2 has a stronger bias for 5'-G at the -2-position from the core E-box (5'-GT-CACGTG-3') where USF1 has almost equal preference for 5'-G/A, indicating additional amino-acids influencing target specificity. In general, USF2 has a less stringent requirement for specific nucleotides flanking the core E-box across all genome ontology (Supplementary Figure 3.1.11C & 3.1.12) compared to MITF, MYC and USF1. USF2 appears to have greater flexibility in term of the core dinucleotide, similar to MYC and MITF. This suggests that USF1 and USF2 may have a different preference for the sites where USF exchange with MYC and MITF.

The increased occurrence of 5'-TA-3' core dinucleotide occurs mainly among the non-coding- and intronic-regions (Supplementary Figure 3.1.11C & 3.1.12). For the non-coding regions, this allowance increased for peaks with score ≥ 10 while at other regions, peaks with score ≤ 10 have higher occurance for the consensus 5'-CACGTG-3'. An interpretation of this is, at non-coding regions, binding of USF2 to non 5'-CACGTG-3' sites may be facilitated by other proteins in a cooperative manner, resulting in high affinity binding at lower affinity sites. This is supported by the report that USF can bind cooperatively with other TFs (refer to section 1.8 Cooperative binding of USF with cue-specific and tissue-specific TFs). For example, USF binds cooperatively with NF-Y to mediate expression at the *SVCT2* gene (Qiao and May, 2011). We also identified an NF-Y motif with a P-value second to the E-box around the promoter peaks (Supplementary Figure 3.1.12).

It is also worth noting that, unlike MITF and MYC, USF2 has an unequal weight for each nucleotide within the core-dinucleotides and this bias differs depending on the genome ontology. This indicates that USF2 allows for any combination of [C/G][T/A] at the core dinucleotide and the prevalence of each combination differs between different genome ontologies. This may simply reflect the GC-content bias, hence a clean 5'-CACGTG-3' motif at the promoter regions which has a higher GC content (Supplementary Figure 3.1.13). However, we cannot rule out the possibility of a bias in the occurrence of a specific type of E-boxes with another TFBS which influences USF2 binding at different genome ontologies.

ChIP-qPCR on a number of target E-boxes was carried out to confirm the validity of the ChIP-seq (Figure 3.1.7C, iii). We were able to confirm enrichment at all of the targets tested, however, the order of peak score and fold-enrichment across the targets tested do not

Table 3.1.4 Summary table of Gene-ontology (GO) analysis for USF2 filtered dataset.

GO-terms with P < 0.05 derived from top 500 peaks using AmiGo	Total no. of term(s)	41
	No. of metabolism related term(s)	26
	No. of catabolism related term(s)	5
Top 50 GO-terms for filtered-dataset	No of genes	5961
	No. of metabolism related term(s)	13
	No. of biosynthetic related term(s)	3
	No. of sub-cellular localisation related term(s)	4
	No. of catabolism related term(s)	5
	Average Coverage (%)	47.7

correlate perfectly (Figure 3.1.7C, ii , Supplementary Table 3.1.3). An example snapshot of a MYC bedgraph uploaded onto the UCSC genome browser at one of the target analysed by ChIP-qPCR is shown in Figure 3.1.6B, iv.

It is worth pointing out that USF binding is not detected at a number of reported target genes including many carbohydrate, lipid metabolism (refer to reference within Table 3.4.1) and DNA-damage response genes (Baron et al., 2012). As our HA-USF1 stable cell line substantially overexpresses HA-USF1 relative to endogenous USF1, yet no USF1 binding was detected at many of the reported USF-regulated genes, USF binding at these promoters is either conditionally induced or dependent on cooperative binding with (an)other conditionally induced factors such as SREBP and ChREBP.

GO-analysis of USF2 peaks reveals many similarities to USF1 with 14 –common terms among the top 50 –enriched terms derived from the whole ChIP-seq dataset and 22 – common terms among GO-term derived from USF's top 500 peaks with $P < 0.05$ using AmiGo. As we are only comparing the top terms from the USF1 peaks with the top terms from the USF2 peaks, it is possible that the unique terms identified for each USF may appear as a lower ranking term for the other USF. Nonetheless, this direct comparison indicates that, not all sites are identical in term of USF binding. USF2 peaks are primarily enriched in metabolism and catabolism related terms (Table 3.1.4). As with USF1 the average coverage of the top 50 GO-term is not complete (47.7%).

3.1.4 HIF ChIP-seq

HIF mediates the cellular response to hypoxic stress through binding to HREs [5'-RCGTG-3' (R: A/G)] that are enriched among hypoxia-responsive genes (Semenza et al., 1996). While HIF-1α is primarily, transiently induced under hypoxia, HIF-2α gradually accumulates under prolonged, low oxygen (5% O_2) conditions and activates genes like *VEGF* (Holmquist-Mengelbier et al., 2006). This suggests that while HIF-1α is responsible for mediating an acute hypoxic response, HIF-2α may be of more importance in tumour

progression where chronic hypoxia is often relevant. Indeed HIF-2α, but not HIF-1α, has been correlated with neuroblastoma progression (Holmquist-Mengelbier et al., 2006).

The distinct spatial-temporal expression of the two HIF-αs reflects their distinct physiological roles. Targeted disruption of HIF-1α (Carmeliet et al., 1998; Iyer et al., 1998; Ryan et al., 1998) and HIF-2α (Compernolle et al., 2002; Peng et al., 2000; Scortegagna et al., 2003; Tian et al., 1998) yields quite a different phenotype during development and the stage which the embryo dies. In the context of cancer, HIF-1α$^{+/+}$ tumours are preferentially found at neuronal tissues, whereas HIF-1α$^{-/-}$ tumours preferentially form from mesenchyme-derived tissue (Höpfl et al., 2002).

Distinction between HIF-1α and HIF-2α

HIF-1α and HIF-2α exhibit distinct effects on cell metabolism and proliferation (Gordan et al., 2007b). Although HIF-1α and HIF-2α share some common targets, a number of selective HIF-1α- and HIF-2α-responsive genes have been reported (Covello et al., 2006; Raval et al., 2005). For example, HIF-1α, but not HIF-2α, stimulates expression of glycolytic enzymes (Hu et al., 2003; Tian et al., 1997; Wang et al., 2005) while inhibiting anabolic pathways by limiting mitochondrial pyruvate uptake into the TCA-cycle whose metabolites are used as precursors for many anabolic pathways (Kim et al., 2006; Lum et al., 2007). HIF-2α preferentially activates *VEGF* (Ema et al., 1997), *TGF-α* (Gunaratnam et al., 2003), *lysly oxidase* (Erler et al., 2006), *Oct4* (Covello et al., 2006) and *CyclinD1* (Baba et al., 2003) (Figure 3.1.8).

A number of HIF-2α-specific coactivators have also been identified. These include: NF-κB, an essential modulator that enhances HIF-2α transcription activity in normoxia (Bracken et al., 2005), and the Ets1 and Elk-1 TFs (Aprelikova et al., 2006). Similarly, HIF-1α specifically interacts with HNF-4 to regulate some liver- and kidney-specific genes (Zhang et al., 1999).

The HIF ChIP-seq experiments were carried out from pooled material such that the same starting material was split into four fractions, one fraction for each of the HIF and a negative

control. This was done in an attempt to reduce batch-to-batch variations in the hope of improving comparability between each HIF dataset.

All of the HIF datasets have a very similar distribution of peak scores for each genome ontology. HIF-1α and HIF-2α have an almost identical average peak score while HIF-1β has nearly twice the average peak score despite having $\approx 2.5\times$ & $\approx 1.5\times$ the number of peaks called than HIF-1α and HIF-2α, respectively (Supplementary Table 3.1.1). This implies either a much more stringent binding by HIF-1β, or a significantly better performance of the HIF-1β antibody. The latter is much more likely for two reasons: firstly, HIF-α needs to dimerise with HIF-1β in order to bind DNA, thus target stringency should not be too dissimilar. Secondly, a brief inspection of motifs bound at different genome

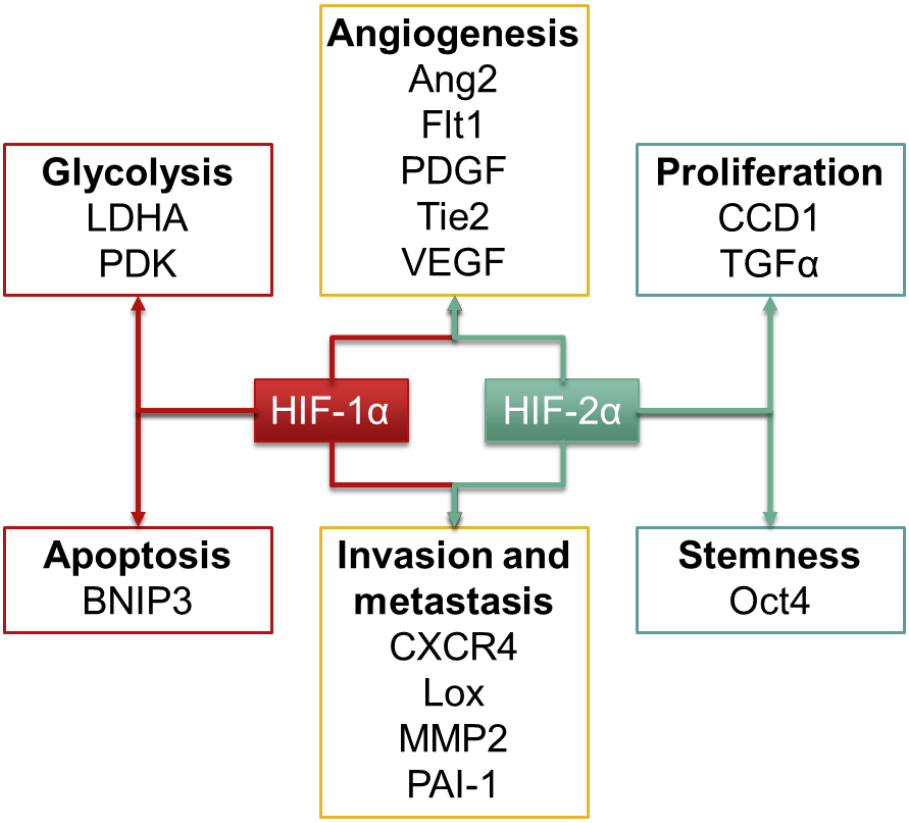


Figure 3.1.8 HIF-1α and HIF-2α share a number of common downstream targets including genes involved in the physiological process of angiogenesis and the associated chemotaxis (labelled as invasion and metastasis as these genes in the context of cancer are involved in invasion and metastasis). HIF-1α alone is involved in glycolysis and apoptosis whereas HIF-2α unique targets include stem cell factor *Oct4* and *Cyclin D1* (*CCD1*). HIF-2α also preferentially regulated *TGFα*. Abbreviation: *Ang2*, angiopoietin-2; *BNIP3*, Bc-2/E1B 19kDa interacting protein 3; CXCR4, C-X-C motif receptor 4; *CCD1*, *Cyclin D1*; *Flt1*, *fms-related tyrosine kinase 1*; *LDHA, lactate dehydrogenase A*; *Lox, lipoxygenase; MMP2*, *matrix metalloproteinase 2*; *Oct4*, *octamer-binding transcription factor 4*; *PAI-1*, *plasminogen activator inhibitor 1; PDK*, *pyruvate dehydrogenase kinase*; *PDGF, platelet-derived growth factor; TGFα*, *transforming growth factor α*; *Tie2, Tunica internal endothelial cell kinase 2; VEGF*, *Vascular endothelial growth factor*

ontologies reveal a much more stringent requirement for an HRE-like motif by HIF-αs over HIF-1β (Supplementary Figure 3.1.14C & 3.1.16C & 3.1.18C). Another possible explanation for the observed average peak score among the HIFs is that, the average peak score of HIF-1β represents the additive value of HIF-1β binding with both HIF-αs. As HIF-1α and HIF-2α have highly overlapping targets, HIF-1α and HIF-2α may be competing for the same sites, and dimerization with HIF-1β is obligate for their binding.

The rank of average peak score stratified by the genome ontology is almost identical with the highest score around the promoter and the TTS then the non-coding, 5'UTR, exon, intron, 3'UTR and intergenic regions in this order (Supplementary Figure 3.1.14A & 3.1.16A & 3.1.18A). The pattern of spreading of outlier peaks with high peak score is also very similar across all HIFs, with the majority of outlier peaks with high peak score at the promoter, intron and intergenic and depleted around the 3'UTR and TTS. HIF displays the highest proportion of binding around the promoter region ($\approx 50\%$) (Supplementary Figure 3.1.14B & 3.1.16B & 3.1.18B). The increases in the proportion of HIF peaks in the promoter when considering peaks with peak score of ≥ 10 are mainly balanced out by the decreased proportion of peaks bound to intergenic and intronic regions, indicating that HIF factor binding to these regions generally have lower peak scores than promoter targets (compare to Supplementary Figure 3.1.3E & F & G).

The consensus motifs of HIF-1α, HIF-1β and HIF-2α identified *de novo* from the whole datasets are reported in Figure 3.1.9A. HIF-1α appears to have quite a strong stringency for the 5'-CACGT-3' HRE which can be found in promoter, intronic and intergenic regions of both the full- and filtered-datasets, although this can be extended to TTS in the full-dataset (Supplementary Figure 3.1.14C & 3.1.15). The stringent requirement for the HRE is more apparent for sites with score of ≥ 10. This allowance for binding to more degenerate sites may be explained through cooperative binding with SP1 (Koshiji et al., 2005; Gordan et al., 2007a). Indeed, SP1 is identified as the second most prevalent motif around HIF-1α peaks located around the promoter regions (Supplementary Figure 3.1.15) and with a much

greater statistical significant than the HRE itself for the HIF-1β (Supplementary Figure 3.1.17) and HIF-2α (Supplementary Figure 3.1.19) datasets.

HIF-1β and HIF-2α display an overall binding motif more similar to the canonical 5'-CACGTG-3' E-box, albeit a rather degenerate one. For HIF-1β-only peaks at the promoter, intronic and intergenic regions harbour any motifs resembling an HRE-element for both full and filtered-datasets (Supplementary Figure 3.1.16C & 3.1.17). There is very little flexibility

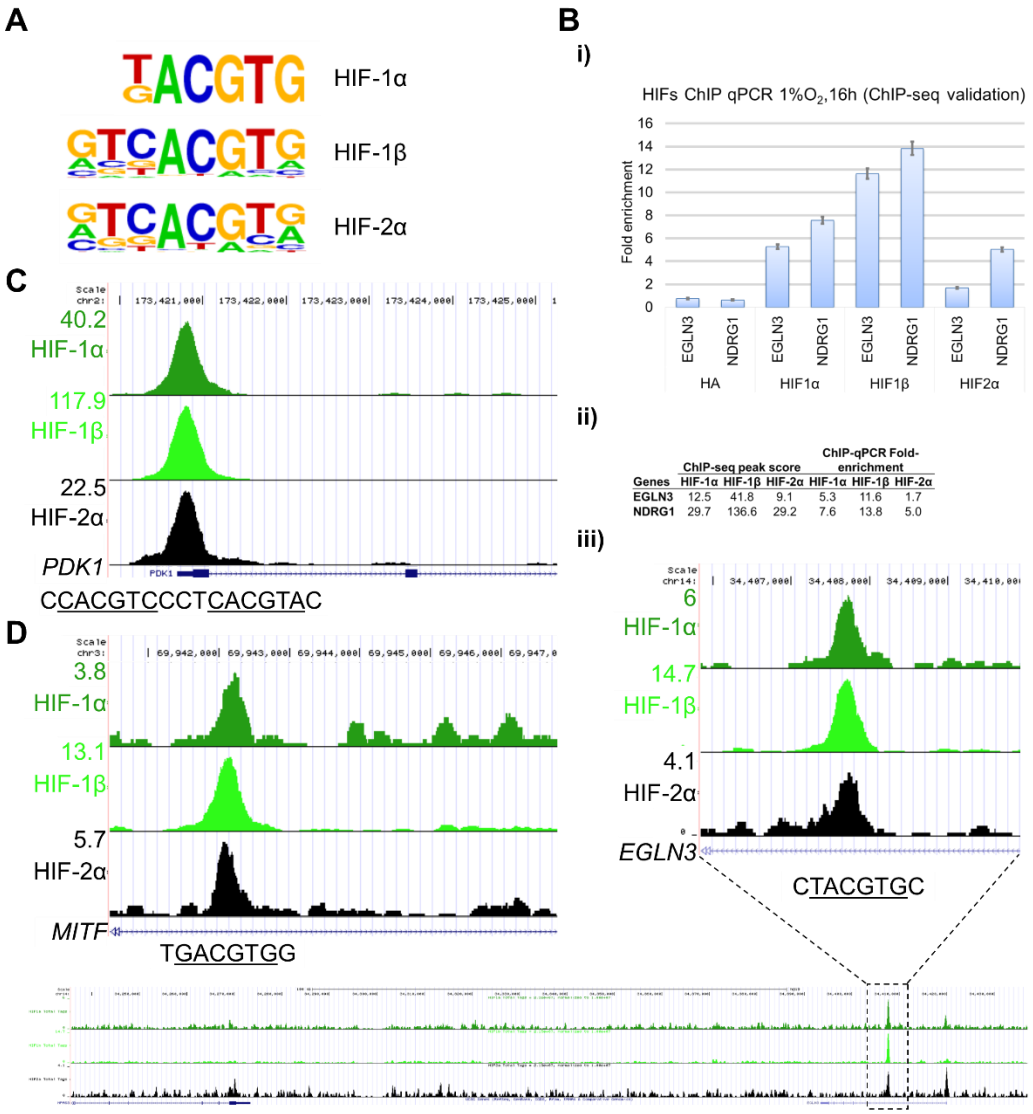


Figure 3.1.9 HIFs' motifs and targets **A)** Top motifs discovered *de novo* using the full HIF ChIP-seq dataset. **B) i.** ChIP-qPCR at known HIF targets, *EGLN3* and *NDRG1*. HA-ChIP was used as a negative control. Fold enrichment is calculated using $\Delta\Delta C_t$ approach as described in Materials and Methods. $\mathbf{N = 2}$ biological replicates each with 2 systematic replicates. **ii.** Table comparing ChIP-seq peak score against ChIP-qPCR fold-enrichment at indicated genes. **iii.** Representative snapshot of HIF bedgraphs uploaded on to the UCSC genome browser at *EGLN3* gene. To illustrate the quality of the ChIP-seq dataset, a broader region around the peak is also shown. **C)** Same as **B, iii** but at a reported HIF-1α specific-target, *PDK-1*. **D)** Same as **C)** for *MITF*, a novel HIF target we identified. The HRE/E-boxes underneath the peak are shown, the core E-box underlined.

with regard to the middle 5'-ACG-3' nucleotides. Additional invariable 5'-G (5'-G-CACGYW-3') can also be found around HIF-1β peaks with score of ≥ 10 at promoters.

HIF-2α peaks display very little signature of the HRE-motif except for peaks with score of ≥ 10 around promoters which have 5'-GT-CACGT-3' (Supplementary Figure 3.1.18C). This 5'-GT-CACGT-3' motif is lost when considering the full-dataset (Supplementary Figure 3.1.19), indicating that the HRE-like motif is enriched among HIF-2α high affinity sites. Interestingly, HIF-2α's intronic peaks are enriched in 5'-GT-CATGTG-AT3', a sequence which is also recognised by USF2.

ChIP-qPCR at *EGLN3* and *NDRG1* were carried out to confirm the success of HIF ChIP-seq samples (Figure 3.1.9B, i). [Note that, due to time constraints we did not perform any ChIP-qPCR validation of the HIF-samples as we only received the ChIP-seq data back less than a month before thesis submission, so ChIP-qPCR was carried out at known HIF's targets]. An HA-ChIP was used as an additional negative control, and the data was normalised to pre-immune rabbit serum. The relative fold-enrichment correlates well with the peak scores for each HIF, but not between HIFs, most likely reflecting a difference in the antibody quality.

We also compared our ChIP-seq data to the published targets of the HIF, especially those differentially regulated by HIF-1α and HIF-2α discussed above. However, we were not able to detect significant binding at any of the angiogenic genes, nor genes involved in invasion and metastasis (data not shown). However, we were able to detect binding at some of the more recently published targets of HIF including collagen-modifying enzymes such as *P4HA1* and *PLOD1* (Gilkes et al., 2014). For all genes reported to be HIF-1α-specific that we examined, we always detected a significant, albeit lower peak score from HIF-2α (an example is shown in Figure 3.1.9C). Thus, rather than being HIF-1α specific, these sites are probably better binding sites for HIF-1α over HIF-2α. At least for *PDK1*, a reported HIF-1α-specific target, the two E-boxes underneath the peak are closer to those of the HIF-1α over

the HIF-2α consensus motif. As for HIF-2α specific targets that have been reported for proliferation and stemness genes, we did not detect any binding (by HIF and by any of the bHLH-TF we are studying), either because these genes are not expressed in 501mel or they are not induced under the conditions we used to study hypoxic induction.

In addition to the known targets of HIF, we also identified *MITF* as a novel HIF target (Figure 3.1.9D). This binding site is unique to HIF, that is, it is not occupied by the other bHLH-TFs examined, and may represent a pigment-cell specific HIF target, as MITF is known as the master regulator of pigment cells.

Gene ontology analysis of HIF's top 500 genes reveal enrichment of hypoxic response and glycolytic pathway terms as expected (Table 3.1.5). HIF-1α also contains an additional term related to protein metabolism that only turns up in the top 50 GO-terms derived from the whole filtered-dataset. However, contrary to the literature, we found many terms relating to negative-regulation of biological processes such as cellular metabolism, gene expression and biosynthetic processes only in the HIF-2α top 50 GO-terms, but not in HIF-1α set (data not shown). Also contrary to the expectation is the average coverage of all of the GO-terms related to response to oxygen levels which are 47.9%, 68% and 52.3% for HIF-1α, HIF-1β and HIF-2α, respectively (Table 3.1.5). In other words, not a complete coverage. It is possible that other oxygen responsive genes may be involved in response to different levels of hypoxic stress. Interestingly, none of the hypoxic response and oxygen sensing terms comes up in the top 50 GO-term derived from the whole filtered-dataset (Table 3.1.5). This indicates that while hypoxic-responsive genes represent high affinity targets bound by HIF following exposure to low O_2 tension, HIF factors also regulate a plethora of other genes which may not contain HREs but rather have E-box motifs instead. Hence, the observed E-box like motifs identified *de novo*.

The top 50 GO-terms derived from the whole filtered-datasets contain 23 overlapping terms while 53 out of the combined 115 GO-terms derived from the top 500 peaks with $P < 0.05$ using AmiGo overlap among the HIF (data not shown).

Table 3.1.5 Summary table of Gene-ontology (GO) analysis for HIF-1α (**A**), HIF-1β (**B**) and HIF-2α (**C**) ChIP-seq filtered dataset.

A

HIF-1α	GO-terms with P < 0.05 derived from top 500 peaks using AmiGo	Total no. of term(s)	38
		No. of response to oxygen levels related term(s)	6
		No. of metabolic related term(s)	10
		No. of catabolism related term(s)	3
	Go-terms related to response to oxygen levels derived from filtered dataset	No. of response to oxygen levels related term(s)	16
		Average Coverage (%)	47.9
	Top 50 GO-terms for filtered-dataset	No of genes	3113
		No. of response to oxygen levels related term(s)	0
		No. of metabolic related term(s)	29
		No. of biosynthetic related term(s)	10
		Average Coverage (%)	25.6

B

HIF-1β	GO-terms with P < 0.05 derived from top 500 peaks using AmiGo	Total no. of term(s)	54
		No. of response to oxygen levels related term(s)	6
		No. of metabolic related term(s)	21
		No. of biosynthetic related term(s)	4
		No. of cellular component organization related term(s)	4
		No. of histone modification related term(s)	3
		No. of protein modification related term(s)	3
	Go-terms related to response to oxygen levels derived from filtered dataset	No. of response to oxygen levels related term(s)	11
		Average Coverage (%)	68
	Top 50 GO-terms for filtered-dataset	No of genes	7237
		No. of response to oxygen levels related term(s)	0
		No. of metabolic related term(s)	23
		No. of biosynthetic related term(s)	5
		No. of cellular component organization related term(s)	5
		No. of protein modification related term(s)	5
		Average Coverage (%)	57.9

C

HIF-2α	GO-terms with P < 0.05 derived from top 500 peaks using AmiGo	Total no. of term(s)	22
		No. of response to oxygen levels related term(s)	6
		No. of metabolic related term(s)	7
	Go-terms related to response to oxygen levels derived from filtered dataset	No. of response to oxygen levels related term(s)	17
		Average Coverage (%)	52.3
	Top 50 GO-terms for filtered-dataset	No of genes	4617
		No. of metabolic related term(s)	24
		No. of biosynthetic related term(s)	8
		No. of cellular component organization related term(s)	4
		No. of protein modification related term(s)	4
		Average Coverage (%)	37.4

3.1.5 Binding specificity

The highly conserved basic region explains the highly overlapping DNA motifs bound by bHLH-TFs (Figure 1.6). However, the observation that genes involved in specific biological processes tend to be regulated preferentially by a specific bHLH-TF suggested that subtle differences in the amino acids contacting the DNA, together with potential PTMs that affect DNA-contact, must be responsible for the observed differences in the hierarchical order of motifs preferred. Moreover, the requirement for this class of TF to bind as a dimer implies sequence specificity may also be regulated through choice of dimerization partner. For example, when each monomer has a different specificity for the half site, or one of the monomers carries an altered basic domain, this might lead to choice of an asymmetric target sequence (Garrell, 1991).

In this section, we investigated the difference in the E-box target specificity for MITF, MYC, USF and HIF in an attempt to address the questions of how subtle differences in the E-box sequence influence which subset of bHLH-TFs are likely to bind (see section 3.1.7), and the effect on the relative occupancy of those that are able to bind at the basal level.

DNA-phosphate backbone contact and genome titration

Anchoring of a TF to a specific DNA sequence depends on the combination of non-sequence-specific binding to the phosphate backbone and sequence-specific base contacts. TFs with high affinity for the DNA phosphate backbone will require less contribution from sequence-specific base contacts in order to be anchored to a specific locus. In the simplest sense, this means that a given bHLH-TF will be able to bind to a range of E-box motifs provided that the overall binding energy is sufficient to anchor the TF at that locus. This was demonstrated through MYC:MAX site selection studies and ChIP experiments which revealed a number of E-box variants including 5'-CATGTG-3', 5'-CACGCG-3', 5'-CATGCG-3' and 5'-C-AACGTG-3' (Blackwell et al., 1993; Grandori et al., 1996).

Regulation of non-sequence-specific DNA phosphate backbone contact has been demonstrated for MITF in our lab (unpublished data). Acetylation of MITF at K243 (K243ac), the last residue of the loop region (Figure 1.6) resulted in the disruption of the non-sequence-specific hydrogen-bond interaction between the amide side chain of K243 and the phosphate of the adenine at position 2 (Figure 3.1.10A). This K243ac, which is increased by MAPK signalling (Figure 3.1.10B), disrupts MITF-DNA-phosphate backbone interaction resulting in a reduced binding affinity to DNA-phosphate backbone. This is demonstrated through the reduced ability for the *TYR* enhancer probe containing MITF canonical E-box (5'-T-CACGTG-A-3') to be competed away by genomic salmon sperm DNA (containing a plethora of E-box variants, most of which would not correspond to MITF's canonical E-box) when using MITF-K243Q (an acetylation mimetic) but not MITF-K243R (a non-acetylable lysine-mimetic) relative to MITF-WT proteins (Figure 3.1.10C). This shifts in relative DNA-binding from high affinity sites (canonical binding sites) to a higher relative binding to low affinity sites (degenerate E-boxes presented across the genome) is a phenomena we termed "genome titration". Thus, the acetylation status of MITF at K243 influences the relative binding to different types of E-boxes by MITF.

As this last residue of the loop region (Figure 1.6) is highly conserved, we hypothesised that similar genome titration mechanism might be at work for other bHLH-TFs, although not necessarily regulated by the same lysine modifying enzymes. Genome titration in response to different induction cues would allow for global regulation of factor exchange in the case where each bHLH-TF is acetylated at the MITF-K243-equivalent residue in response to different induction cues. To this end, we set out to generate equivalent bacterial expression vectors for USF1 and USF2. However, we were not able to obtain satisfactory induction of heterologous protein expression in bacteria even when we considered secondary structure, codon usage and destabilizing sequences. Although, induction of heterologous protein was confirmed by Western-blot, we were unable to detect any protein by Coomassie-staining in some cases (data not shown). Since, we were not able to detect any acetylation of bHLH-

LZ-TFs transiently overexpressed in the Phoenix cell line, affinity purified and probed with pan-acetyl-lysine antibody (data not shown), we concluded that there is no significant lysine-acetylation in these overexpressed proteins and went ahead to perform all the subsequent band-shift assays using nuclear extracts from Phoenix cells transiently overexpressing the bHLH-LZ-TFs of interested. TFE3 and TFEB were cloned from cDNA obtained from MCF7 cells and MYC from 501mel cells as described in the Material and

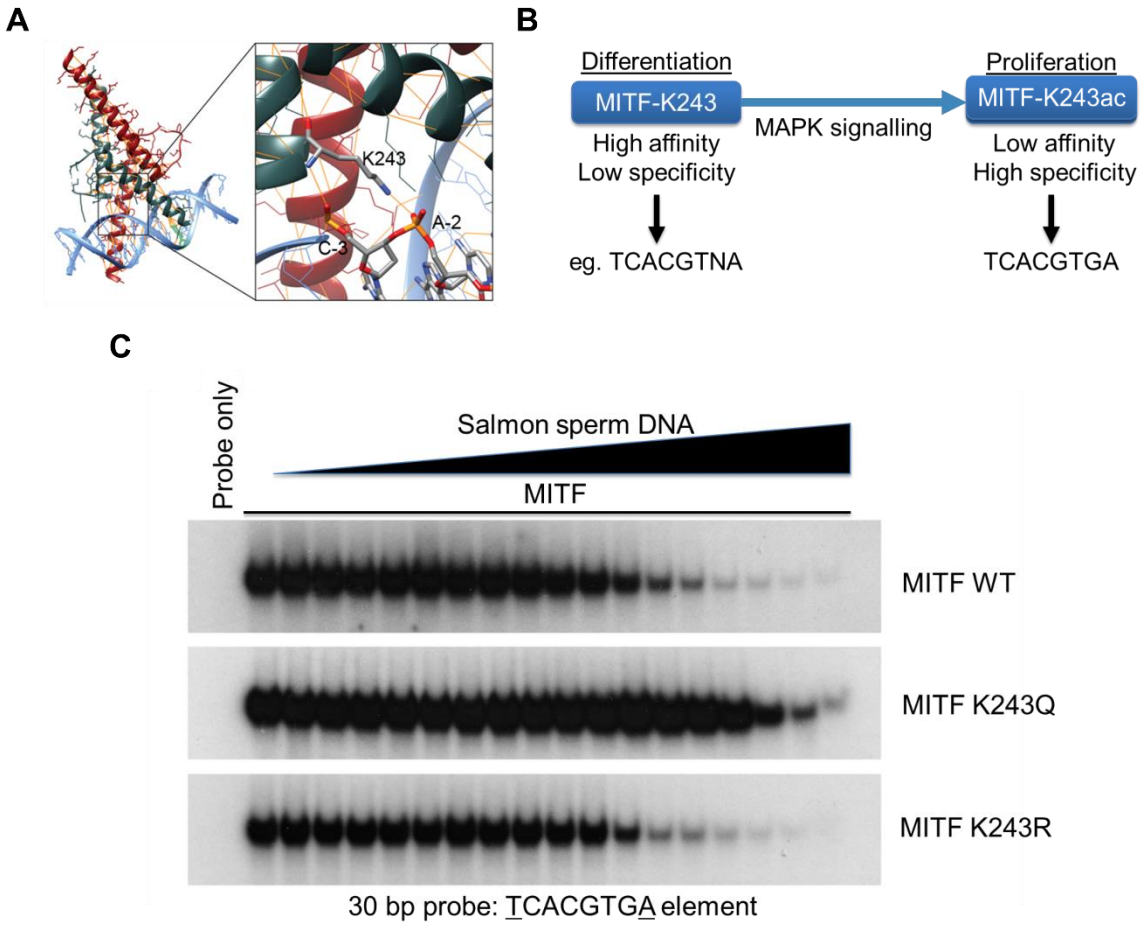


Figure 3.1.10 Acetylation of MITF at K243 decreases MITF's DNA-binding affinity. **A)** Graphical rendering of MITF bound to dsDNA derived from MITF crystal structure. The inset shows the **K243** whose amide side chain forms hydrogen bonds to the phosphate of the adenine at position **2**. **B)** Schematic diagram of the "genome titration" model where acetylation at MITF-**K243** results in the loss of the non-sequence specific protein-DNA−interaction, resulting in a decreased DNA-binding affinity. As a result, MITF-**K243ac** can only bind to the high affinity sites as these represent sites where binding energy from the sequence specific DNA-interaction is sufficient to compensate for the loss of non-sequence specific interaction made by **K243**. It was also proposed that the switch between acetylated and non-acetylated MITF can be induced by MAPK-signalling pathway, effectively determining which subset of genes (as pre-determined by the E-box presented) are to be activated during proliferation and differentiation. **C)** *In vitro* competition assay where bacterially purified MITF was first incubated with a *TYR* enhancer probe containing the consensus M-box, 5'-T-CACGTG-A-3' before being competed away with increasing amount of salmon sperm DNA (representing competitor harbouring various combination of E-box motifs, most of which would not have the canonical sequence) in **4** −fold increment steps maxing at **10 mg/ml** that corresponds to the lane furthest to the right of the band-shift presented. MITF-**K243R** which represents non-acetylable lysine mimetic displays similar behaviour to MITF-WT in term of its ability to bind to salmon-sperm DNA. MITF-**K243Q** which represents lysine acetylation mimetic displays a much reduced overall DNA-binding affinity (≥ **256**-fold), as inferred from the band-shift data (each lane represent a four-fold dilution of Salmon-sperm DNA). Data taken from our lab manuscript under preparation.

Methods. We mutagenize TFE3, TFEB, USF and MYC mammalian expression vectors at the MITF-K243-equivalent residues and checked for expression in the nuclear extract by Western-blotting (Supplementary Figure 3.1.20).

Competition assays analogous to that performed for MITF were carried out as described in Materials and Methods. We were able to obtain similar results for the other two members of the MiT-family, TFE3 and TFEB (Figure 3.1.11), notably that the Q mutants were competed less well than the WT or R mutant proteins. (We omitted TFEC as it is not expressed in most adult cell types including melanoma cell lines). For USF2, the K276Q-mutant displayed the weakest binding to non-specific DNA. However, the K276R-mutant also displayed decreased generic binding relative to wild-type, but less so than the K276Q-mutant. It is possible that differences in the amino-acids in the loop region of USF2 from the MiT-family result in a different effect of the arginine substitution. Although commonly used as a non-acetylable mimetic of lysine, the functional group of arginine (CH_2-CH_2-CH_2-NH-C-N_2H_3) and lysine (CH_2-CH_2-CH_2-CH_2-NH_2) are quite different and certainly of a different size. Thus, our observation indicated that the DNA-contacting interface of USF2 may not be identical to the MiT-family proteins and substitutions of K276 with either a Q or an R decrease binding to DNA-phosphate backbone.

The result from USF1 was more difficult to interpret as the overexpressed protein forms doublets on Western-blotting (Supplementary Figure3.1.20E), implicating the existence of at least two USF1 species that were also reflected in the band-shift, most apparently for the Q−mutant (Figure 3.1.11). Our initial experiment with 4 −fold dilution of salmon-sperm DNA

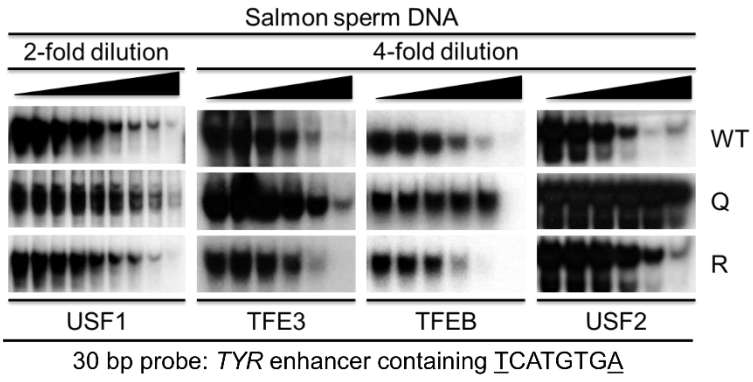


Figure 3.1.11 Competition assay against genomic DNA. Nuclear extracts of Phoenix cells transiently over-expressing the indicated proteins were incubated with the *TYR* enhancer probe and salmon sperm DNA as described in Materials and Methods. Only the bound DNA is shown, but in all cases the probe was in excess. A **2** – fold dilution of salmon sperm DNA was used for USF1 and a **4** –fold dilution gradient for TFE3, TFEB and USF2. The furthest lane to the right correspond to **10 mg/ml** salmon sperm DNA.

did not clearly resolve a difference between the wild-type and the K240Q/R mutant (data not shown). We therefore repeated the experiment using a 2 −fold dilution gradient which demonstrate a decreased ability for the K240Q mutant to bind to the salmon sperm DNA and a similar binding by the wild-type and the K208R mutant.

We also performed the same analysis on MYC. However, we did not observe significant binding even to a non 5'-T + 3'-A nucleotides flanking the core E-box probe (5'-V-CACGTG-B-3', data not shown). This may be explained by the strict requirement for heterodimerization of MYC with MAX for DNA-binding. In a transient overexpression experiment, the majority of MYC in the nuclear extract may be undimerised or homodimerised MYC which is unable to bind to DNA stably (Grinberg et al., 2004). For the little that does bind, presumably as MYC:MAX heterodimers, there was no noticeable difference between the wild-type and the mutants. As the MITF-K243-equivalent is not conserved in MAX (MAX has an arginine at this position), this residue is likely irrelevant for MAX. The role of MYC-K407-acetylation may, therefore, be irrelevant for MYC:MAX binding.

In conclusion, substitution of the conserved lysine-residue at the end of the loop region with an acetylation mimetic (glutamine) resulted in a reduced binding to DNA-phosphate backbone. As sequence-specific DNA-binding by a TF is a combination of sequence specific base contact and non-sequence specific DNA-phosphate backbone contact, a PTM that decreases non-sequence specific DNA-phosphate backbone contact such as MITF K − 243ac will affect the relative binding primarily to low affinity motifs as the overall binding energy at these sites may now be below the threshold for the TF to bind stably. The consequence of decreased non-sequence specific DNA-phosphate backbone contact is a general shift toward binding to high affinity motif(s) and decreased binding to low affinity targets. Since, MITF-K243 equivalent is highly conserved and substitution of this lysine with an acetylation mimetic resulted in a similar genome titration in *in vitro* competition assay, acetylation at this conserved lysine-residue may represent a common regulatory mechanism for regulating relative DNA-binding to different motifs.

We next looked for evidence of USF acetylation *in vivo.* Based on our lab's unpublished data on MITF acetylation by p300 in response to MAPK-signalling, we attempted to look for USF1 and USF2 acetylation when co-expressed with some HATs (p300/CBP and P/CAF) known to interact with USF (Huang et al., 2007). As USF was shown to be activated via MEK-ERK1/-2 in response to *H. pylori* infection to drive proliferation-associated gene expression (Juttner et al., 2003), MAPK-driven acetylation of USF that directs USF to proliferative genes, similar to MITF, was deemed possible. To investigate acetylation of USF/MITF, we co-transfected USF/MITF with HATs or constitutively active component of the MAPK-cascade (as MITF acetylation by p300 is dependent on MAPK pathway, Goding

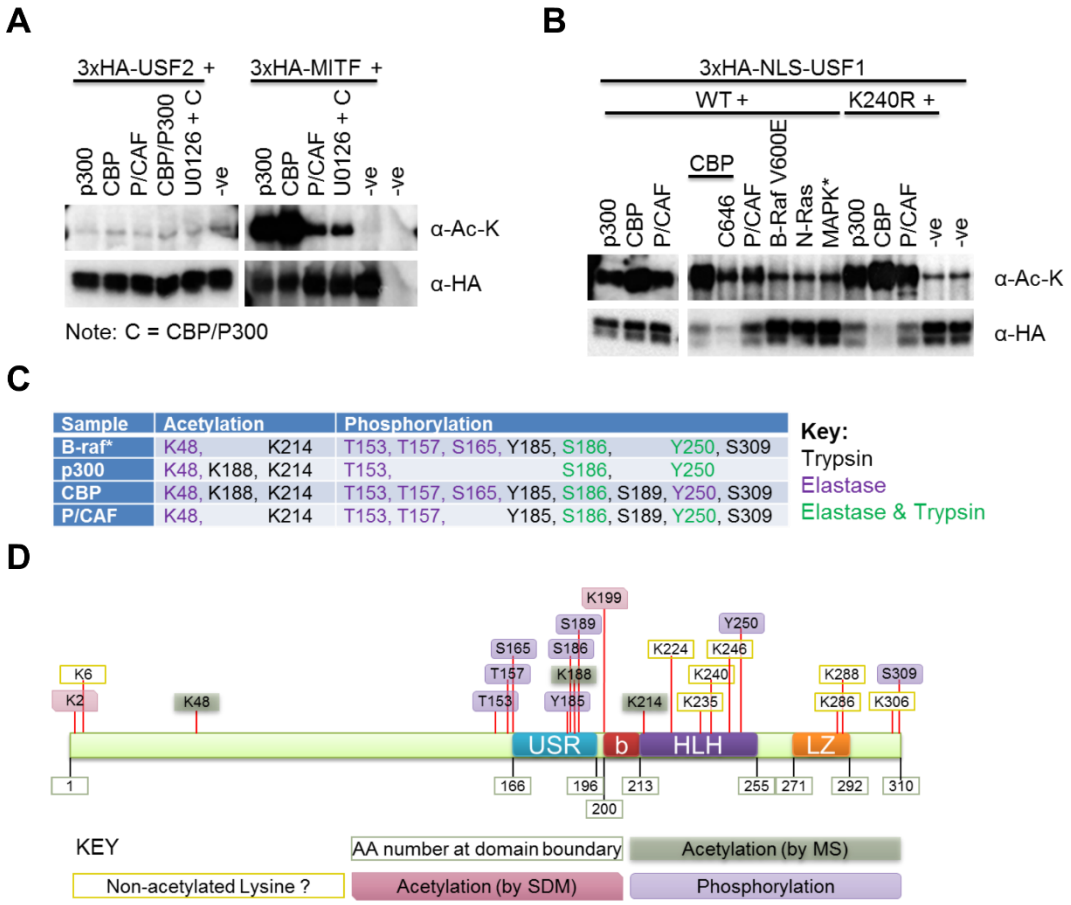


Figure 3.1.12 Acetylation of affinity purified USF and MITF. **A)** Western-blot of HA-USF2 and MITF co-transfected with the indicated expression vectors and probed with pan-acetyl-lysine antibody or HA-tag antibody. In the lane treated with U0126, the cells were also co-transfected with p300/CBP. **B)** Same as **A)** for HA-USF1. The 4th and 5th lane correspond to CBP without and with C646 inhibitor. Asterisk (*) denote constitutively active mutant. **C)** A table listing acetylation and phosphorylation sites identify by MS-approach. USF1 was co-transfected with either activating BRaf (B-raf*), p300, CBP or P/CAF. The sample were affinity purifies, SDS-PAGE, appropriate band excise and in-gel digested by either trypsin or elastase before running on reverse-phase chromatography and MS-analysis. Reported modifications only encompass those which appear in peptide with score ≥ **20**. Three biological replicas were analysed. **D)** Schematic diagram of USF1 as in Figure 1.11 but with all lysine indicated (red line). The sites of post-translational modification as identified by MS and inferred from the site-directed mutagenesis experiments are labelled as indicated.

lab unpublished observation) into Phoenix cells. HA-USF or MITF were then affinity purified with mouse-α-HA antibody and Western-blotted for acetylation using anti-pan-acetyl-lysine antibody or rabbit-α-HA to check for acetylation and the amount of protein purified, respectively (Figure 3.1.12A & B).

We were able to reproduce MITF acetylation by p300 and additionally demonstrated MITF acetylation by CBP and to a much lesser extent by P/CAF (Figure 3.1.12A 3xHA − MITF; p300, CBP or P/CAF). We were also able to show that MITF acetylation by p300 and CBP is reduced when cells were treated with U0126, a MEK inhibitor concomitantly with the transfection (Figure 3.1.12A H3xHA − MITF; U0126 + C). [The activity of CBP and p300 is dependent on phosphorylation by p42/p44 MAPK (ERK1/2) (Gusterson et al., 2002; Liu et al., 1999), hence a MEK inhibitor should reduce phosphorylation of MITF by p300 and CBP.]

We detected high levels of USF1 acetylation when co-expressed with p300, CBP and P/CAF, with CBP inducing the highest levels of acetylation (Figure 3.1.12B WT + p300, CBP or P/CAF). The involvement of CBP was confirmed by the reduced acetylation observed upon co-treatment with C646, a p300/CBP inhibitor (Figure 3.1.12B WT + CBP ± C646). None of the constitutively active components of MAPK-cascade resulted in changes in USF1 acetylation (Figure 3.1.12B WT + B − RafV600E, N − Ras or MAPK*).

The USF1-K240R-mutation did not abolish acetylation when co-expressed with p300, CBP or P/CAF, if anything, the acetylation level greatly increased when normalised to the protein level (Figure 3.1.12B K240R + p300, CBP or P/CAF). As with the wild-type USF1, co-transfection with CBP resulted in the highest level of acetylation. It is worth noting that co-expression of USF1 or MITF with either p300, CBP or P/CAF resulted in a greatly diminished USF1 and MITF purified protein level (Figure 3.1.12A & B α − HA).

As we were not able to detect acetylation of USF2 when co-expressed with p300, CBP or P/CAF (Figure 3.1.13A, USF2), and acetylation of MITF is being pursued by another lab member, we focused our investigation on USF1. The inability of K240R mutation to abolish

USF1-acetylation by p300, CBP and P/CAF does not exclude the possibility of K240 being an acetylation target by these HATs as other lysines may also be acetylated. We therefoere mutagenized every lysine residues in USF1 and investigated the effect on the overall acetylation of USF1 by p300, CBP and P/CAF. In all cases, there were no single lysine mutants that significantly affected acetylation by p300, CBP and P/CAF (data not shown).

MS-analysis of affinity-purified USF1 co-transfected with activated-BRaf or p300, CBP and P/CAF were subjected, initially to digestion by trypsin before reverse-phase chromatography and MS-peptide sequencing (Figure 3.1.12C). (A representative of sample validation by Western-blotting and Coomasie staining can be found in Supplementary Figure 3.1.21.). With the exception of K214ac, none of the other modified residues identified (K188ac, Y185Ⓟ, S189Ⓟ and S309Ⓟ) occur close to the DNA-binding region. As our site-directed mutagenesis, as well as random multi-site-directed mutagenesis to generate pools of multiply-mutated lysine residues clearly implicated additional residues other than USF1-K188 and K214 in acetylation by p300, CBP and P/CAF (data not shown), we repeated the MS-experiment with elastase to increase sequencing coverage. We were able to get almost a complete coverage with elastase and identified additional PTMs (K48ac, T153Ⓟ, T157Ⓟ and S165Ⓟ), however, none of these corresponds to USF1-K240 and none are close to the DNA-binding region (Figure 3.1.13C). The MS-result confirmed our lab's previously published data on T153-phosphorylation (Galibert et al., 2001) but we were unable to detect K199-acetylation (Corre et al., 2009). In any case, we will need to perform *in vitro* acetylation of USF1 using purified USF1 and each of p300, CBP and P/CAF (and all the necessary substrates) to unequivocally confirm direct acetylation by these HATs.

A summary diagram of USF1 and our identified PTM sites are indicated in (Figure 3.1.12D). As it is impossible to test every possible combination of mutants and candidate HATs, we have formed a collaboration with Philippe Lambert (Toronto) who will use our cell lines

stably expressing HA-USF1, HA-USF2 and FLAG-MYC to perform proteomic analysis to identify interaction partners as well as endogenous PTMs.

Target E-box sequence preference

To explore the effect of single nucleotide variations within and immediately flanking the core E-box on DNA-binding affinity by bHLH-TFs of interested, we first looked at the average and the spread of the peak score when the E-boxes are presented as either one, two or three E-boxes in tandem. One interesting observation was that we did not observe a single 5'-CACGTG-3' E-box present on its own in any of the ChIP-seq datasets we looked at. As we set the conditions such that the box-plots in Supplementary Figure 3.1.22 – 3.1.28 only include peaks which carry the indicated motif alone, in the absence of any other E-box screened, it would appear that 5'-CACGTG-3', at least in our datasets, is never present alone. Since our analysis in section 3.1.6 did not reveal any obvious cooperative binding to tandem E-boxes within a given peak, we shall limit our discussion to the most complete dataset in Supplementary Figure 3.1.22 – 3.1.28.

Core-E-box preference

MITF/TFE3

The genome-wide ChIP-seq data revealed 5'-CACGTG-3' as the most preferred target (highest average peak score), followed by 5'-CATGTG-3' (Top row, Supplementary Figure 3.1.22, 2 – E – boxes). Our *in vitro* band-shift data, however, suggested that 5'-T-CATGTG-A-3', at least in the context of the *TYR* enhancer, is MITF's highest affinity target as the 5'-T-CACGTG-A-3' cold competitor has a significantly decreased ability to compete with the 5'-T-CATGTG-A-3' probe for MITF (Figure 3.1.13A, CAcGTG) compared to the 5'-T-CATGTG-A-3' cold competitor (Figure 3.1.13A, T – Y – A where Y = CATGTG). Mutation of other nucleotides within the core M-box revealed that all bases are important for MITF binding.

We repeated the experiment using TFE3, another member of the MiT-family which is identical to MITF in the basic and the dimerization helices, although it differs at two-positions within the loop region (Figure 1.6). *In vitro* band-shift experiments of TFE3 demonstrated

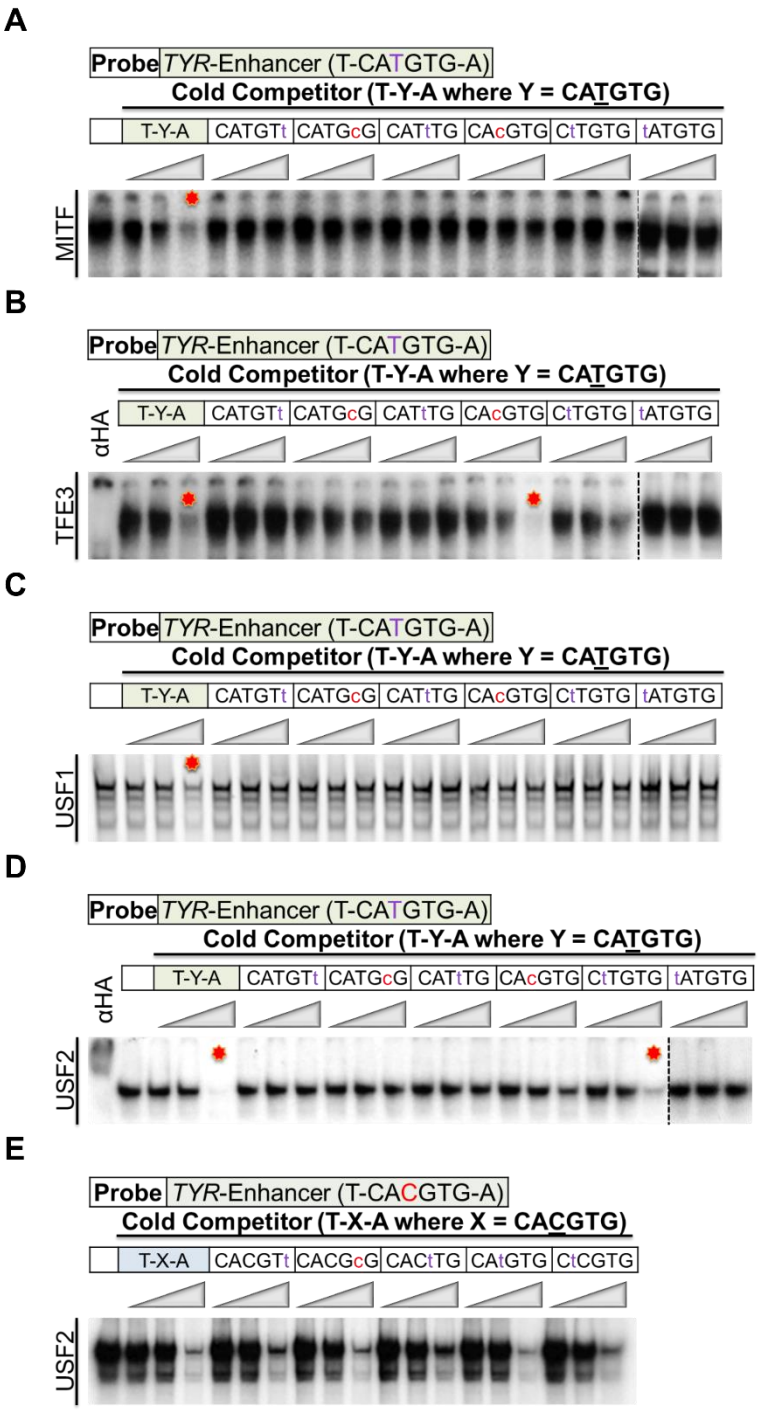


Figure 3.1.13 *In vitro* DNA-binding competition assay to determine the relative contribution of each nucleotide within the core E-box to the overall binding affinity. Nuclear extracts of Phoenix cell line transiently overexpressing MITF (**A**), TFE3 (**B**), USF1 (**C**) and USF2 (**D-E**) were incubated with the indicated cold-competitor carrying *TYR* enhancer sequence with the indicated mutations within the E-box prior to binding with the labelled probe. Dotted line indicated region cropped out from the film. The concentrations of competitor oligonucleotides were **10**, **50** and **250 ng/reaction** from left to right.

the ability of 5'-T-CACGTG-A-3' to compete against the *TYR* enhancer probe as well as, if not better than a 5'-T-CATGTG-A-3' cold competitor (Figure 3.1.13B). The contribution of the other nucleotides within the E-box, however, displayed the same pattern as for MITF, although the differences are not as strong as for MITF.

MYC

Analysis of the peak score against motifs bound by MYC reveals 5'-CACGTG-3' as the most preferred target followed by 5'-CATGTG-3' based on the average peak score (Top row, Supplementary Figure 3.1.23). With the exception of 5'-CACTTG-3' other variations within the core di-nucleotide of the E-box around [C/N][N/G] do not significantly affect the average peak score. This is also in a good agreement with motif analysis where, with the exception of 5'-CACGTG-3' enriched among peaks with score of ≥ 10, peaks with lower scores display an appreciable log-probability score for other combinations of the core-dinucleotide (Supplementary Figure 3.1.6 & 3.1.7).

USF1

Through site selection studies, it was reported that the USF1 full consensus binding site is 5'-GGT-CACGTG-ACC-3' (Bendall and Molloy, 1994). This is in agreement with our genome-wide data which demonstrated a strong preference for 5'-CACGTG-3' by USF1. The average peak score is almost twice that of the next preferred targets 5'-CATGTG-3' (Top row, Supplementary Figure $3.1.24, 2-\mathrm{E}-\mathrm{boxes}$).

USF1 and USF2 homo- and hetero-dimers were demonstrated to bind to all E-boxes with the core 5'-CAC[G/A]TG-3' in site-selection study (Bendall and Molloy, 1994). However, our genome-wide data show that other combinations of [C/N][N/G] at the core di-nucleotides, although not identical in terms of the hierarchy of the preference, do not differ significantly between each other relative to 5'-CACGTG-3'. This is in a striking contrast to our *in vitro* competition assay where only 5'-CATGTG-3' was able to compete with the labelled probe (Figure 3.1.13C).

USF2

Analysis of USF2 genome-wide data revealed a very similar pattern of E-box preference as that observed for USF1 with 5'-CACGTG-3' being the most preferred followed by 5'-CATGTG-3'. As with USF1, the difference between the average peak score of 5'-CATGTG-3' peaks and the rest did not differ significantly (Top row, Supplementary Figure 3.1.25, 2 – E – boxes). Analogous to the situation with USF1, our *in vitro* competition assay revealed 5'-CATGTG-3' as the most preferred target sequence. Although, in this case we were able to see minor competition by 5'-CACGTG-3' at $250\ \mathrm{ng/reaction}$. Interestingly, and in contrast to USF1, mutation of the A–to–T in 5'-CTTGTG-3' did not significantly affect the ability to compete with labelled probe (Figure 3.1.13D). The contribution of the Gs and the C (5'-T-CATGTG-A-3') on the binding of USF2 is in agreement with the reported dimethylsulphate *in vivo* foot-printing assay which identified three-protected guanine residues within the E-box of the haem oxygenase-1 promoter bound by USF (Figure 3.1.14) (Hock et al., 2004) and the report that the outer GC is indispensable for USF binding (Lennard and Egly, 1987).

It is possible that some of the observed discrepancies between the preferred motifs derived from *in vivo* (5'-CACGTG-3') and *in vitro* (5'-CATGTG-3') experiments might depend on the context in which the E-box is situated. To test this possibility and to confirm the band-shift result that 5'-CATGTG-3' represents the highest affinity site in the context of *TYR* enhancer, we repeated the same experiment with USF2 and the *TYR* enhancer probe carrying 5'-CACGTG-3' (Figure 3.1.13E). In this experiment, we were able to show that 5'-CACGTG-3'

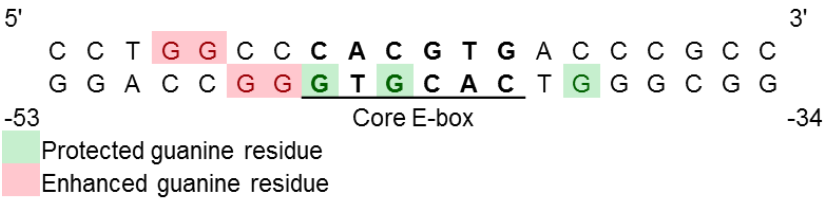


Figure 3.1.14 Schematic diagram representing the E-box and the immediately surrounding sequence at the haem oxygenase-1 promoter. The core E-box is marked in bold and underlined. The guanines protected from dimethylsulphate footprinting are highlighted in green while the enhanced guanine resides are highlighted in red. The most important contact being the 3'-G of the transcribed strand of the core E-box, mutation of which abolished all binding.

cold competitor was able to compete away 5'-CACGTG-3' probe at $250\ \mathrm{ng/reaction}$ as with 5'-CATGTG-3' cold competitor. This confirms that 5'-CATGTG-3' is the highest affinity motif. 5'-CACGTG-3' cold-competitor, although unable to efficiently compete with a 5'-CATGTG-3' probe when presented at $250\ \mathrm{ng/reaction}$ is able to compete with 5'-CACGTG-3' probe.

With regards to the other nucleotides, we also observed a decreased ability of the competitor oligos carrying a mutation in one of the Gs (5'-T-CATGTG-A-3') to resist the competition by the probe.

HIF

HIF-1α shows a clear preference for its consensus 5'-ACGTG-3' HRE and not much difference with regards to motifs carrying different variation of [C/N][N/G] (5'-A[C/N][N/G]TG-3') when presented as a single HRE E-box (Top row, Supplementary Figure $3.1.26, 1-\mathrm{E}-\mathrm{box}$). Although the consensus 5'-ACGTG-3' still displays the highest average peak score at peaks with two-HRE E-boxes, the difference between 5'-ACGTG-3' and other combinations of 5'-A[C/N][N/G]TG-3' is less prominent.

HIF-1β (Supplementary Figure 3.1.27) and HIF-2α (Supplementary Figure 3.1.28) which showed a consensus motif of 5'-BACGTR-3' also showed a clear preference for the consensus sequence at peaks with one HRE E-box. However, HIF-1β and HIF-2α appear to differentiate different variations in the core di-nucleotide better than HIF-1α. This is in a good agreement with the *de novo* motif identification performed in section 3.1.4 where HIF-1α displays a clean 5'-ACGTG-3' motif, while the motifs of HIF-1β and HIF-2α resemble a degenerate E-box (Figure 3.1.9A).

Flanking sequence preference

Another aspect which might differentiate one E-box from another is sequences flanking the core E-box. Binding site selection studies have demonstrated subtle differences in flanking sequence preference. For example, USF was reported to have a preference for the 5'-T and 3'-C nucleotide flanking the core E-box (Bendall and Molloy, 1994). This difference may

be determined by the variable residues at positions 3 and 4 of the basic region (Figure 1.6). Structural evidence suggested that amino-acid at position 4 of USF contacts the phosphate backbone (Ferre-D'Amare et al., 1994). Although this does not confer sequence specificity, the extra binding stability conferred may compensate for the loss of the preferred 3'-G bound by His at position 4 (as compared to MYC), potentially allowing USF to bind DNA irrespective of flanking sequences. However, without the structure of the full-length complex of various bHLH-TFs with DNA it is impossible to accurately assign the role of variable amino-acids and the potential regulatory contribution of the region outside the bHLH on DNA binding specificity.

The contribution of the flanking sequences may extend even further than the immediate flanking sequences. *In vitro* studies revealed an extended 12bp 5'-GAC-CACGTG-GTC-3' E-box as a MYC:MAX's high affinity site (Halazonetis and Kandil, 1991; Sommer et al., 1998). Binding site selection studies demonstrated preference for the first G at this high affinity site. ChIP using MYC and MAX-specific antibodies, on the other hand, indicated a preference for 5'-GC, 5'-CG or 5'-AG immediately preceding the core E-box (Grandori et al., 1996). These experiments demonstrated that: firstly, flanking sequences contribute to the observed differences in target specificity of bHLH-TFs and secondly, flanking sequence preference may be regulated. To this end we investigated the effect of the flanking sequences on the relative binding affinity of our bHLH-TFs of interest.

MITF/TFE3

As MITF is required for the development and survival of pigmented cells; yet proteins involved in pigmentation (containing 5-'T-CATGTG-3' M-box) (Lowings et al., 1992) are not, MITF must regulates other genes, most likely through E-boxes (Aksan and Goding, 1998).

Motifs identified *de novo* from our genome-wide data demonstrate 5'-GT flanking the E-box (5'-GT-CACGTG-3', Supplementary Figure 3.1.4C) to be most prevalent. The average peak score also demonstrated 5'-GT as the most preferred flanking sequence (Supplementary

Figure 3.1.22 third and second row, respectively). This is complemented by the observation made from our *in vitro* competition experiment which shows that the effect of a 5'-T on the binding affinity is influenced by the flanking sequence on the 3'-end. 5'-T + 3'-A represents the highest affinity binding sites (Figure 3.1.15A). However, a T-present at both the 5'- and 3'-end and 5'-C + 3'-C represented the weakest sites while 5'-T and 3'-C/G display similar affinity. The influence of 5'-T/3'-A on MITF binding is roughly symmetric. That is, the direction of the E-box, at least in the context of the *TYR* enhancer probe, does not affect the relative binding affinity by MITF as varying the 5'-nucleotide while fixing the 3'-A gave a similar pattern of binding affinity as that for 5'-T.

As MITF and TFE3 behave differently with regard to the specificity of the nucleotide around the core E-box, we also looked at the effect of flanking sequence on TFE3 binding (Figure 3.1.15B). Only a 5'-T + 3'-A competitor is able to effectively bind TFE3. Again, the observed difference highlights the potential involvement of the variable residues in the loop in determining sequence specificity.

MYC

5'-T or 3'-A bases flanking the core E-box was shown to disfavour MYC:MAX binding but not MAX:MAX (Fisher et al., 1993b; Solomon et al., 1993). Our genome-wide analysis of peak height against motifs, however, did not show specific discrimination of either 5'-T or 3'-A on its own (Second row, Supplementary Figure $3.1.23, 2-\mathrm{E}-\mathrm{boxes}$). In fact, our genome-wide data indicated 5'-A as the highest affinity site, at least in the context of a single E-box and that 5'-T represented the highest affinity site in the context of three tandem E-boxes. The latter is in agreement with the motifs identified *de novo* where 5'-T is enriched among peaks with score of ≥ 10 (Supplementary Figure 3.1.6C).

With regard to the second closest flanking sequence, there is a discordance between the preferred nucleotides when considering one, two or three-tandem E-boxes solely based on the mean peak score. However, 5'-G comes up as the most preferred target when

considering peak score of the upper quartile. The preference for 5'-GT is in agreement with the enrichment found among peaks with score of ≥ 10 (Supplementary Figure 3.1.6C).

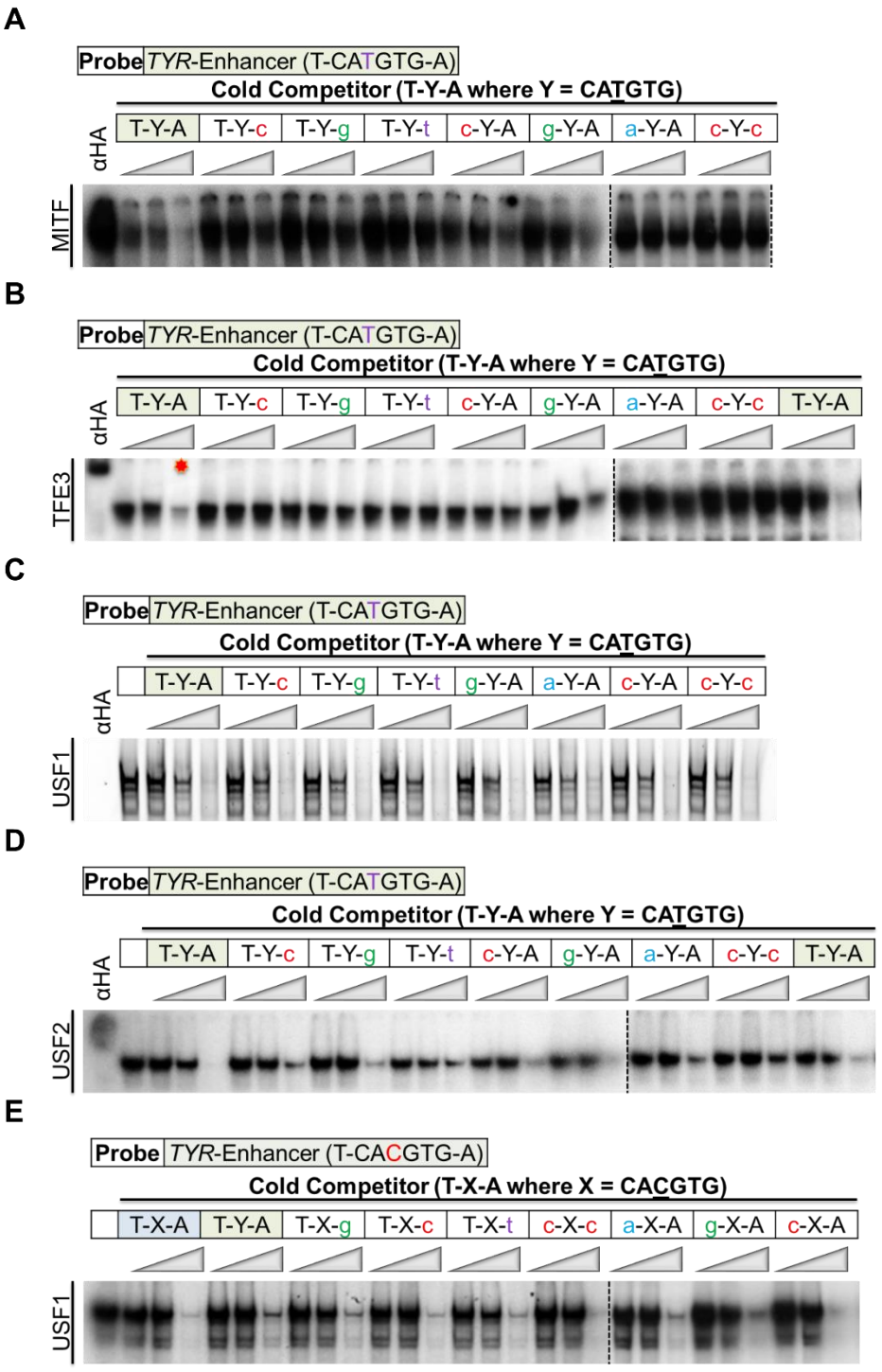


Figure 3.1.15 *In vitro* DNA-binding competition assay to determine the relative contribution of the nucleotides flanking the core E-box on the overall binding affinity. Nuclear extracts of Phoenix cells transiently overexpressing MITF (**A**), TFE3 (**B**), USF1 (**C, E**) and USF2 (**D**). The nuclear extracts were incubated with the indicated cold-competitor carrying *TYR* enhancer sequence with the indicated mutation of the nucleotide flanking the core E-box (denoted as X and Y) prior to binding with the labelled probe. Red asterisk indicates unique pattern to note. Dotted-line indicated region cropped out of the film. The concentrations of competitor oligonucleotides were **10**, **50** and **250ng/reaction** from left to right.

USF

As with the reported 5'-T and 3'-C flanking sequence preferred by USF from *in vitro* site selection assay (Bendall and Molloy, 1994), our *in vivo* data demonstrated preference for 5'-T/C with 5'-T being the most preferred for USF1 (Second row, Supplementary Figure 3.1.24). The most preferred second closest flanking nucleotide for USF1 is 5'-A for a single E-box and 5'-G for two and three-tandem E-boxes based on average peak score (Third row, Supplementary Figure 3.1.24). The observed preference for 5'-[G/A]T flanking sequence is reflected in 5'-[G/A]T appearing as part of the most frequently occurring motif of USF1 (Supplementary Figure 3.1.8C & 3.1.9). However, our *in vitro* competition assay demonstrated essentially no discrimination against any combination of 5'-T+ 3'-N and 5'-N+ 3'-A and 5'-C + 3'-C on binding by USF1 (Figure 3.1.15C).

For USF2, based on the average peak score alone, 5'-C represent the most preferred nucleotide immediately flanking the core E-box when presented as 2 –E-boxes and 5'-T as the most preferred nucleotide when considering 3 – tandem E-boxes (Second row, Supplementary Figure 3.1.25). 5'-G is the most preferred second nearest nucleotide flanking the core E-box (Third row, Supplementary Figure 3.1.25). 5'-GT nucleotides flanking sequence also appear in the most frequently occurring motif of USF2 (Supplementary Figure 3.1.11C & 3.1.12). In the context of the *TYR* enhancer probe, *in vitro*, USF2 displays a similar ability to bind to any combination of 5'-T-Y-N-3' and 5'-N-Y-A-3' but clearly shows a decreased ability to bind when both 5'-T & 3'-A are mutated to C (Figure 3.1.15D). The difference in the binding affinity of USF1 and USF2 to 5'-C-Y-C-3' suggest how 5'-C & 3'-C may confer specificity for USF1 homodimers.

As the *TYR* enhancer E-box is asymmetric, yet there is no major difference in the ability of 5'-T-Y-N-3' and 5'-N-Y-A-3' cold competitors to compete against the labelled *TYR* enhancer probe, the contribution of the nucleotides immediately adjacent to the core E-box on the overall binding affinity of USF2 is symmetric, equivalent and additive as the presence of

both 5'-T & 3'-A appear to compete marginally better than cold competitors which have either 5'-T or 3'-A.

As a validation of our competition assay with regard to the flanking sequence and to investigate whether the difference in the core E-box might influence how the flanking sequences affect the binding, we repeated the experiment for USF1 using the *TYR* enhancer probe carrying a 5'-CACGTG-3' core E-box in place of 5'-CATGTG-3' (Figure 3.1.15E). The result appears very similar, with USF1 having little discrimination against any nucleotides immediately flanking the core E-box.

HIF

HIF-1α demonstrates preference for 5'-G and 3'-C nucleotide immediately flanking the core HRE (5'-ACGTG-3') when considering $2-$HREs under a peak base on average peak score. For $3-$HREs the preference for 5'-nucleotide becomes a T while the 3'-nucleotide remain a G (Second row, Supplementary Figure 3.1.26). The preference for 5'-G/T is in agreement with HIF-1α's motif identified *de novo* (Figure 3.1.9A). With the exception of a 5'-A second closest nucleotide to the core HRE when presented as two HREs, there is no obvious preferred nucleotide for either 5' or 3' (Third row, Supplementary Figure 3.1.26, $2-$E$-$boxes).

HIF-1β demonstrated a slight bias for a 5'-G and a 3'-C nucleotide flanking the core HRE for peaks with $2-$HREs. The 3'-C preference is retained for peaks with $3-$E-boxes while the 5'-nucleotide preference becomes a C, based on average peak score alone (Second row, Supplementary Figure 3.1.27). This is also in good agreement with motifs identified *de novo*, especially when considering motifs identified for promoter peaks with a peak score ≥ 10 (Supplementary Figure 3.1.16C). With regard to the second closest nucleotide flanking the core E-box, 5'-C nucleotides have the highest average peak score. The 3' nucleotide preference, however is a C when presented as one or three E-boxes under the peak and a G when presented as two E-boxes under a peak (Third row, Supplementary Figure 3.1.27).

The preference for the nucleotide immediately flanking the core HRE for HIF-2α is less obvious. This is reflected in a rather degenerat motif identified *de novo* for HIF-2α (Figure 3.1.9A). 5'-T has a marginally higher average peak score for peaks with one and two-HREs while 5'-A has the highest average peak score for peaks with three-E-boxes (Second row, Supplementary Figure 3.1.28). 5'-G is the most preferred nucleotide second closest to the core E-box. This is in agreement with the motif identified *de novo* for promoter peaks with peak score ≥ 10 (Supplementary Figure 3.1.18C). 3'-C nucleotide second closest to the core E-box has marginally higher average peak scores when presented as one and three E-boxes under the peak while 3'-G has a highest average peak score when presented as two E-boxes under a peak (Third row, Supplementary Figure 3.1.28).

Amino-acids determining flanking sequence specificity.

Our genome-wide correlation of peak score against motifs and *in vitro* competition assay revealed differences in flanking sequence preferred by the bHLH-LZ-TFs studied. Based on the sequence conservation and the available crystal structure, we tried to correlate specific amino-acids with the flanking sequence preference.

Figure 3.1.16A shows the amino-acids around the beginning of the basic-domain (position -1 to $+4$). The sequence implicated in flanking sequence specificity for MITF (based on the crystal structure) and the equivalent position of MYC, CBF1 and USF are highlighted with a red box. MITF (Aksan and Goding, 1998) and CBF1 (Fisher and Goding 1992) have Asp at the third position and display a preference for the 5'-T and 3'-A flanking sequence. MYC, on the other hand, has an Arg at this third position and does not bind 5'-T and 3'-A flanking sequences (Fisher et al., 1993b). USF, which from our *in vitro* competition experiment demonstrated no clear discrimination against specific flanking sequences carry an alanine at the third position. One hypothesis would be alanine, which has no-side chain, neither forms sequence-specific base contacts nor contributes to stearic interference with specific bases, enabling USF to accommodate a broad range of flanking sequences. Therefore, we might hypothesise that substitution of USF's Ala at the third position of the

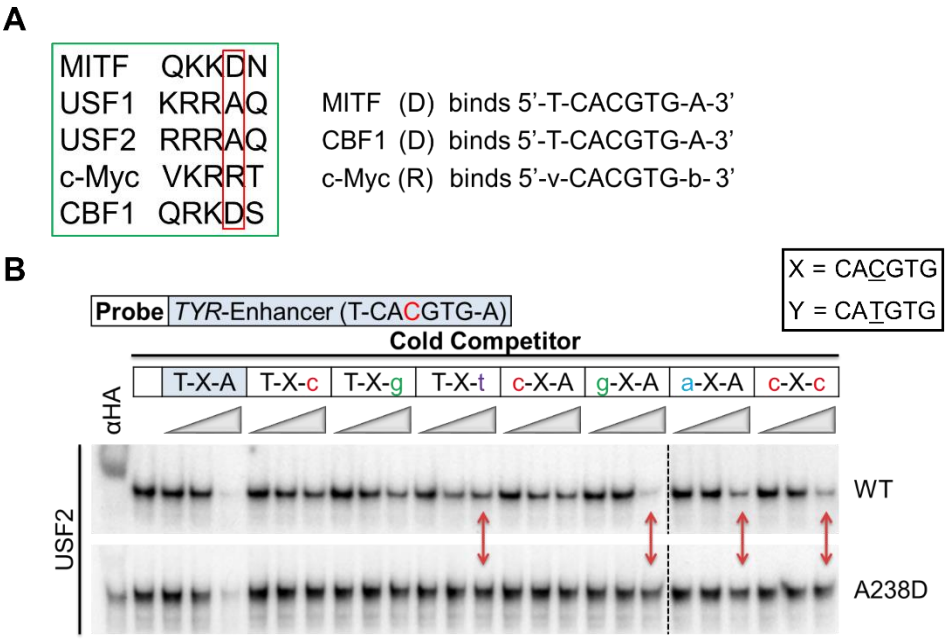


Figure 3.1.16 *In vitro* DNA-binding competition assay to determine the effect of mutating the amino-acids at the third position of the basic domain on the binding to the E-boxes with different nucleotides flanking the core E-box. **A)** Sequence alignment of the MITF, USF1, USF2, MYC and CBF1 at the sequence around the beginning of the basic domain (-1 to + 4). The residues proposed to be involved in flanking sequence specificity based on the crystal structure is highlighted in the red box. On the right side, the sequence bound by the indicated bHLH-TF and the residues thought to be involved in determining the flanking sequence specificity are shown. **B)** Nuclear extracts of Phoenix cell line transiently overexpressing either USF2 wild-type (WT) or USF2 carrying the **A238D** mutation (mimicking MiT-family) were incubated with the indicated cold-competitor and labelled probe as described in Materials and Methods. Red arrows indicate major differences between the wild-type and the USF2 $-$**A238D**. The concentrations of competitor oligonucleotides were **10**, **50** and **250 ng/reaction** from left to right. Dotted line indicated region of film cropped away.

basic domain with an Asp might confer MITF's flanking sequence preference onto USF. To address this, we performed a competition experiment using either wild-type USF2 or USF2-A238D mutant against various flanking sequences (Figure 3.1.16B). Gratifyingly, USF2-A238D displays almost an identical competition pattern as TFE3 and MITF with diminished binding to *TYR* enhancer oligonucleotides with any flanking sequence other than the 5'-T + 3'-A.

3.1.6 Binding cooperativity

Cooperative binding was postulated to facilitate binding to nearby low affinity sites by generating super-high affinity binding proteins by coupling multiple monovalent DNA-binding proteins into a multivalent DNA binding complex. Some bHLH-LZ-TFs have been shown to oligomerize in solution. This was suggested to play a role in DNA looping (Brownlie et al., 1997), allowing binding to two, otherwise low affinity binding sites. For

example, it was reported that Myc:Max and Max:Max but not USF bind cooperatively to E-box elements (Lüscher and Larsson, 1999). However, no cooperativity was observed using a transient COS-7 expression system in both competition and titration experiments using either oligonucleotide probes or reporter constructs (Vervoorts and Lüscher, 1999). Another conflicting report indicated that USF is able to form a bivalent homotetramer capable of binding simultaneously to two independent sites, possibly through DNA-looping (Ferre-D'Amare et al., 1994). This was thought to explain the observed rapid kinetic of transfer from spatially separated binding sites by passing through the double bound intermediate (Sawadogo, 1988). These different observations indicated that binding-cooperativity might be context specific.

We wished to explore potential binding cooperativity by MITF, MYC, USF and HIF as this will be useful while considering how factor might exchange following an induction cue. Consider a hypothetical situation where only MITF is able to bind cooperatively but not USF and that both USF and MITF have identical affinity for their shared E-boxes. In the case where a promoter only contains one E-box, MITF and USF will compete for the E-box equally. However, for a gene that contains two E-boxes, and assuming only MITF is able to bind cooperatively as a homotetramer, MITF will be able to outcompete USF at that promoter when both bHLH-LZ-TFs are presented at equimolar concentrations. The potential consequences of cooperative binding are: graded control of E-box-regulated gene expression when MITF is induced (promoters with multiple E-boxes will be the first to have USF displaced by MITF) and preferential retention of MITF at sites containing multiple E-boxes when MITF is depleted.

In this thesis, we explore the potential short-range cooperative binding to proximal sites, similar to what had been reported between USF and a number of other TFs such as NF-Y (Qiao and May, 2011) and SREBP (Griffin et al., 2007) but by each bHLH-TF studied. In an ideal ChIP-seq experiment where peak score correlates with bHLH-TFs binding. One hypothesis would be; cooperative binding will give rise to peaks with a peak score of $> 2 \times$

the added peak score when the motifs are present individually, in other words, when no cooperativity occurs. This hypothesis was fuelled by a report made by a past lab member who reported a highly elevated "pile up" value (generated by MACS 2.0, $\approx\equiv$ to un-normalised peak score used in this thesis) at the *LAMP1* promoter (contain $2\times$ 5'-T-CACGTG-A-3' E-box). There was $\approx 7-8$ fold greater peak height (value at the summit of the peak) when comparing *LAMP1* to the *BAX* promoter's wiggle files (generated by MACS 1.4, visualise on the UCSC genome browser) which contains only one 5'-T-CACGTG-A-3' E-box.

We started our analysis by hypothesising a situation where cooperative binding to an adjacent E-box, regardless of the type, is a default outcome. To this end, we stratified peaks according to the number of E-boxes present under the peaks and box-plotted the peak score against the number of E-boxes (Supplementary Figure 3.1.29). With the exception of MITF that shows a gradual but nonlinear increment of the average peak score up to peaks with $5\times$ E-boxes, we did not observe any trend for other bHLH-TFs we studied. It is entirely possible that cooperative binding only occurs at specific type of E-boxes, or at least those which carry a near or exact match to the consensus E-box or those with a specific spacing. To address some of these possibilities, we stratified peaks according to the number of motifs matching the consensus motif identified *de novo* using the sequence under the peaks from the full-datasets or filtered-datasets. We also stratified peaks according to specific motifs by screening all the possible E-box motifs generated in section 3.1 against the full-dataset and drew box-plots of peak score against motifs (data available upon request). As all bHLH-TFs looked at displayed similar pattern, we limit our discussion to MITF, which is exemplified in Figure 3.1.17A – C and the data which deviate from the general pattern described for MITF (Figure 3.1.17D).

In all cases we observed a gradual increase in the average peak score as the number of E-boxes increases. The increases in the peak score is not additive, that is, the average score of peaks harbouring two motifs is less than twice that of peaks containing only one

motif. This is true for both peaks screened using motif identified *de novo* from the full-dataset or filtered-dataset (Figure 3.1.17A & B).

Inspection of the peak score against an increasing number of a specific motifs explains the observed non-linear increment of peak score against the number of motifs under the peak (Figure 3.1.17A & B). For peaks containing a near match to the consensus E-box sequence we observed approximately double the average peak score when comparing peaks containing one E-box with those containing two (Figure 3.1.17C). Peaks containing motifs that deviated from the consensus motifs generally have the same average peak score regardless of the number of motifs present. An interpretation of this observation is that, none of the bHLH-TFs studied display cooperative binding to proximal E-boxes under a given peak. At high affinity motifs, the bHLH-TFs may be able to bind for sufficiently long that the multiple E-boxes present in close proximity may all be occupied at once, giving rise to a linear increment in the peak score with increasing number of motifs presence. At low affinity motifs, the bHLH-TFs may only bind transiently, effectively sliding between E-boxes presented in close proximity giving rise to an average binding of only a single site at a time.

It is worth pointing out that using the peak score of binding sites with no E-box as background to subtract from all the other average peak scores for each motif is not valid as MITF (and presumably all of the other bHLH-TFs) may also bind in an E-box independent manner. For example, passively as we observed in the recruitment of MITF and USF to DNA-damage sites (Goding lab unpublished observation). An appropriate background to use would be the average pile up across genome-wide that, in the case of MITF, correspond to 0.068 tags per bp, a negligible number.

HIF-1β and HIF-2α that did not have a clear consensus motif (the motif we identified *de novo* was similar to the consensus E-box, albeit a very relaxed one) did not possess any specific motifs where peak score correlated linearly with the number of the motifs present under the peak (data not shown).

The result for USF1 was different. While peaks carrying motifs that deviate from the USF1 consensus motif display a similar peak score regardless of the number of motifs, those which are very similar to the consensus motifs display a roughly similar peak score for peaks with one and two-motifs. However, peaks with > 2 motifs have a significantly higher average peak score than peaks with one or two motifs (Figure 3.1.17D).

Our genome-wide analysis did not reveal cooperative binding at a global level, both when the E-boxes are of the same type or of a different type. This does not, however, exclude

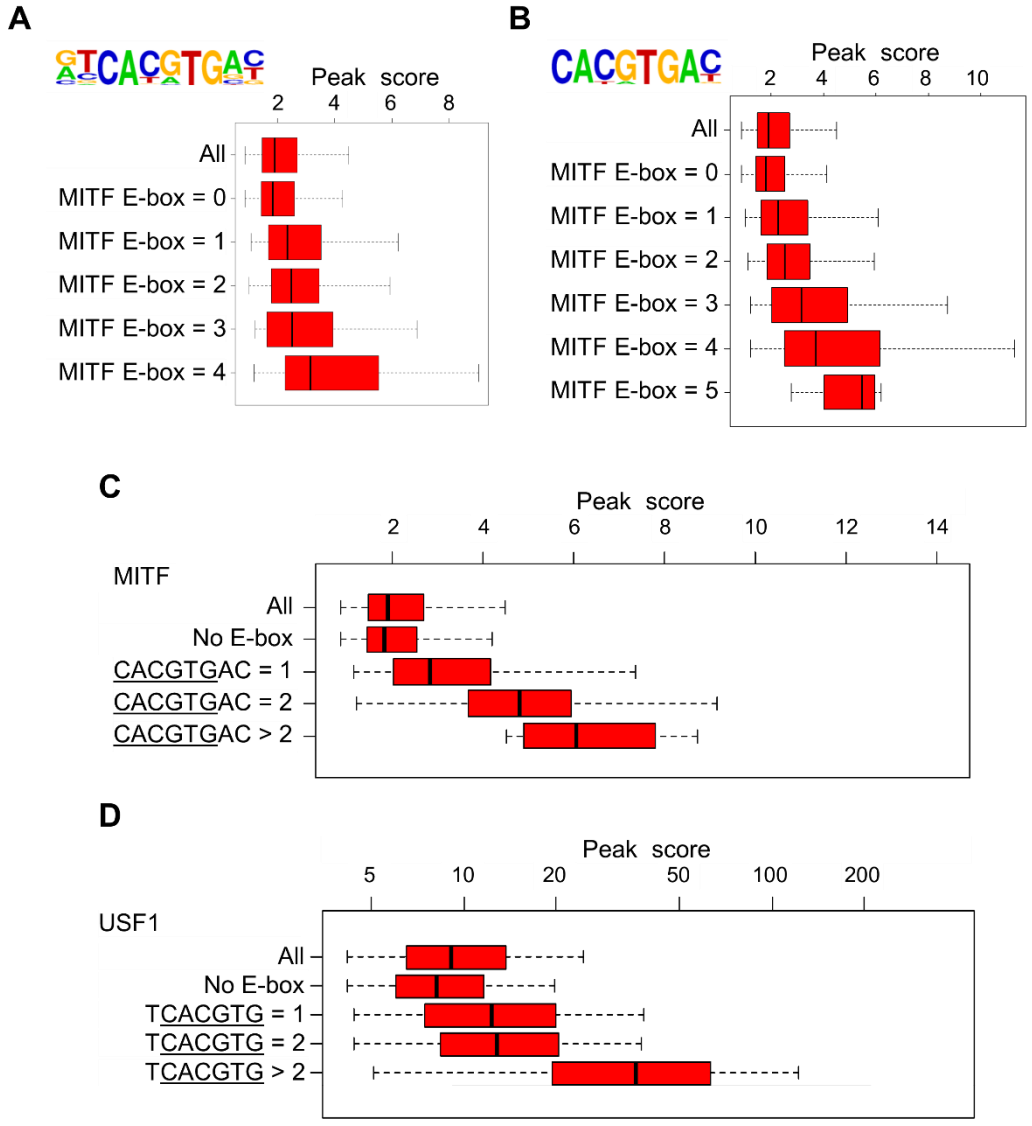


Figure 3.1.17 Peak score against number of motifs. **A)** Box-plots of MITF's peak score stratified into the number of motifs derived from MITF's full-dataset. The motif file used is displayed on the top right hand corner and displayed on the graph as "MITF E-box". "All" denotes distribution of the peak score form all the peaks with any number of motifs. **B)** Same as **A)** but with MITF's motif derived from MITF's top 6% of the peaks (filtered-dataset). **C)** Same as **A)** but using the motif files indicated on the left of each box to screen MITF full-dataset. Shown here are the only top score motifs when occur as a single, two tandem or multiple E-boxes. Full dataset involving all variation of the motifs identified in section 3.1 can be found in Supplementary Figure **3.1.30**. **D)** Same as **C)** for USF1 top score motif. Full dataset involving all variation of the motifs identified in section 3.1 can be found in Supplementary Figure **3.1.31**.

the possibility that cooperative binding requires a specific spacing of the E-boxes and thus may only be found at specific promoters. In fact, the whisker of the boxes for every motif demonstrated a very diverse range of peak score for a given motif and the number of E-boxes considered. If we compare the average of one E-box to the upper quartile (the whisker) of the same motif presented as two-E-boxes, the range between peak score would be very large. To address this possibility, we took a closer inspection at *LAMP1*- and *ARF-GEF2*- promoter (Figure 3.1.18A). The latter contains two tandem E-boxes with no spacer

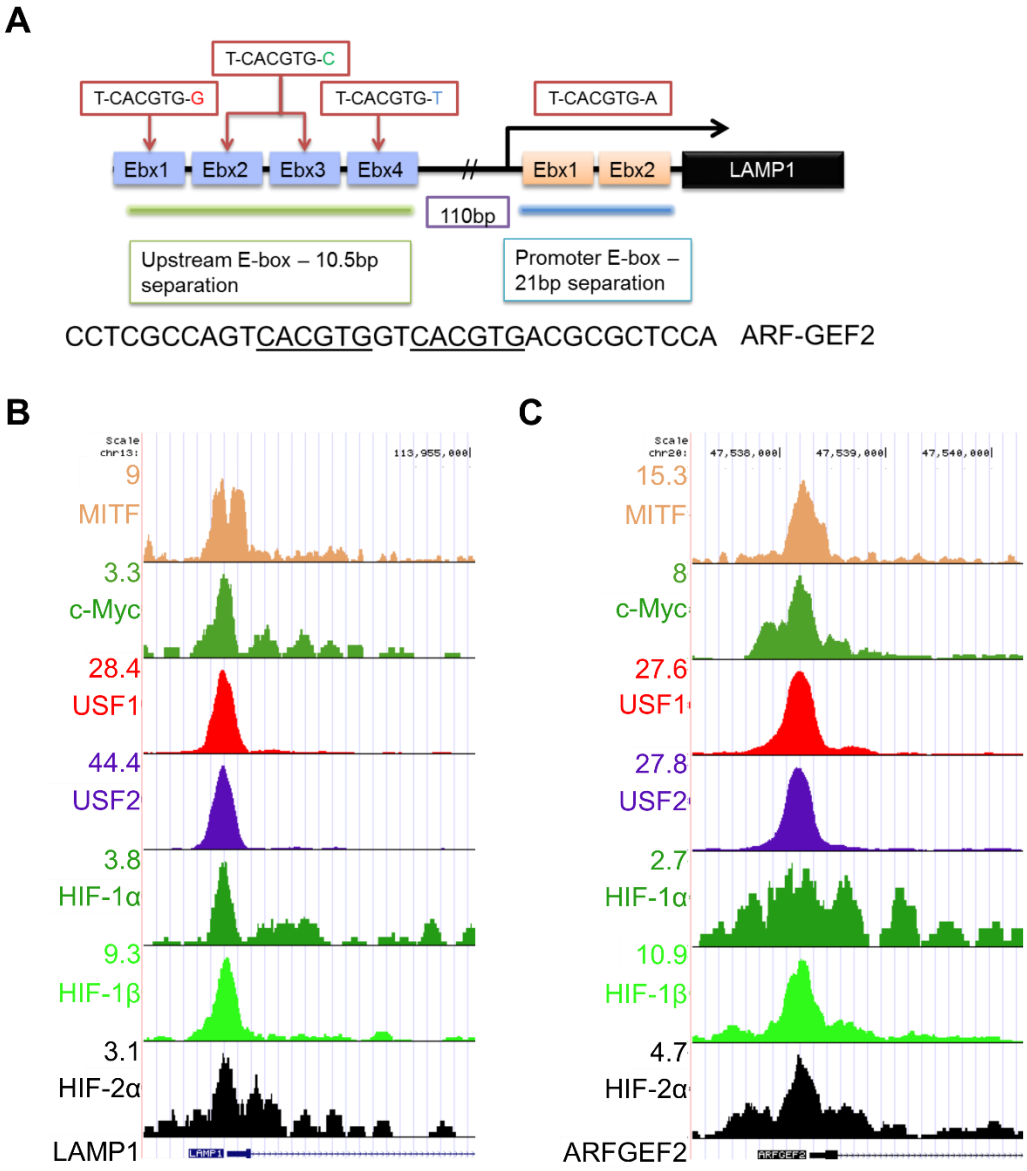


Figure 3.1.18 Study of potential cooperative binding based on E-boxes spacing against ChIP-seq peak score. **A)** Top, a graphical display of the *LAMP1* gene where the location of E-boxes found underneath the peak are indicated. The E-box sequences are shown in the red boxes and the flanking sequence highlighted in different colours. The spacing between the E-boxes are indicated beneath the peaks. Below, the sequence of part of *ARF-GEF2* promoter, an example of a gene positively peak-called that contains tandem E-boxes with no spacing. **B)** Snapshot of bedgraph files uploaded on to the UCSC genome browser for all the ChIP-seq experiments at the *LAMP1* promoter. **C)** Same as **B)** at *ARF-GEF2* promoter.

sequence. For the *LAMP1* promoter, we also noticed that there are additional four-E-boxes upstream of the proximal promoter two-E-boxes and that MITF's *LAMP1* peak actually has two summits (Figure 3.1.18B). Taken together, it appeared likely that the reported co-operative binding by MITF at the *LAMP1* promoter is actually an additive effect of binding to the weaker, non 5'-T + 3'-A flanking sequence at the upstream E-boxes together with the high affinity 5'-T + 3'-A, proximal promoter E-boxes.

As both the 5'-T-CACGTG-G-3' and 5'-T-CACGTG-A-3' type E-boxes present at the *ARF-GEF2* promoter are also found at the *LAMP1* promoter, it is interesting to see the different pattern of binding by our bHLH-TFs (Figure 3.1.18B & C). For example, USF2 shows a higher peak height at *LAMP1* than at the *ARF-GEF2* promoter, while MITF and MYC display greater peak height at *ARF-GEF2* promoter.

Based on the peak height of the bedgraph file at *ARF-GEF2* promoter alone, it seems more likely that cooperative binding by MITF or MYC would occur at tandem E-boxes. While USF1 and USF2 do not appear to exhibit cooperative binding at *LAMP1* and *ARF-GEF2* (comparing the peak score at *LAMP1* and *ARF-GEF2* to the average peak score for a single 5'-T-CACGTG-3' type E-box, data not shown), USF1 binding to closely spaced E-boxes (not tandem) may be more restricted than that of USF2 (comparing the difference in peak height on the bedgraph of USF1 and USF2 at the *LAMP1* and *ARF-GEF2* promoters).

3.1.7 Co-occupancy

Having investigated E-box specificity, we next investigated co-occupied sites. By "co-occupied sites" we mean, common targets that may be bound by the bHLH-TFs discussed. (It is impossible for multiple bHLH-TFs to occupy an E-box at the same time.) The aim is to generate a database of peaks and the corresponding list of bHLH-TFs co-occupied that we can back-reference when picking candidates to study factor exchange following induction of a specific bHLH-TFs or interfering with the binding of one or more bHLH-TF(s) co-occupying the site. Using the list of co-occupied targets, we also wished to address the

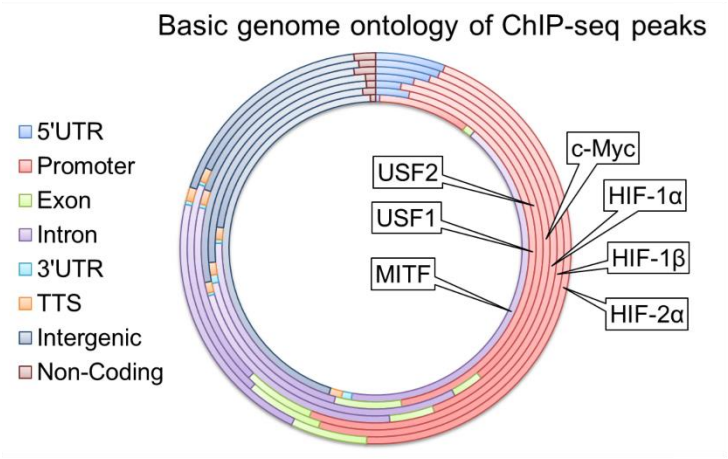
question of the bias in the sequence of the E-box when considering specific set of bHLH-TFs. For example, do USF and MYC co-occupied sites lack 5'-T or 3'-A nucleotide flanking the core E-box apparently prohibitive for the binding of MYC:MAX?

While MITF and USF peaks exhibit enrichment of 5'-CACGTG-3'-type E-boxes across all genome ontologies, 5'-CACGTG-3' is only found enriched at some genome ontologies for MYC (promoter and intergenic), HIF-1α, HIF-1β (promoter, intron and intergenic) and HIF-2α (intergenic). The fact that all bHLH-TFs studied could bind to 5'-CACGTG-3', but the motif was not found enriched across the genome, implicated the existence of mechanism(s) regulating the binding of a given bHLH-TF at a particular genome ontology, for example through cooperative binding with (an)other TF(s) whose TFBS is only enriched at a specific genome ontology. This means that co-occupancy of a given set of bHLH-TFs is influenced by the genomic loci in which the E-box is located. As the relative distribution of peaks into different genome ontologies by each bHLH-TF differs (Figure 3.1.19A), the extent of co-occupancy by a given set of bHLH-TFs is partly determined by the number of targets in each genome ontology.

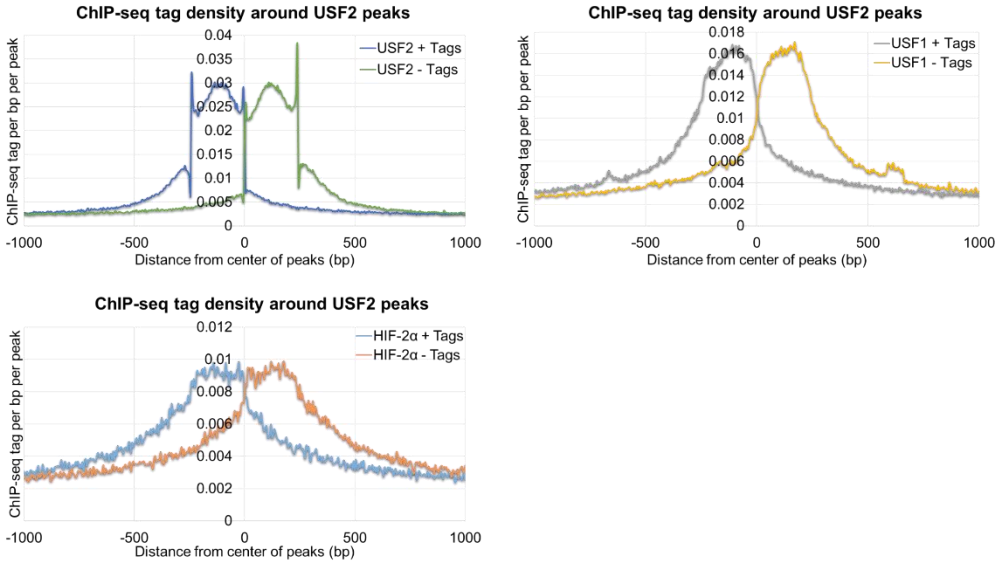
Through the analysis in section 3.1.1 – 3.1.5, we have established that the bHLH-TFs we looked at exhibit overlapping but not identical target specificity. Assuming direct correlation between read counts and binding affinity, bHLH-TFs that are more similar in E-box specificity will display a similar pattern of the relative read count across different targets. In the context of factor exchange, it is conceivable that bHLH-TFs that are more similar in binding specificity are more likely to exchange, unless the bHLH-TF with low binding affinity to a given co-occupied site can be massively induced or its target specificity altered. To assess this hypothesis we performed a pair-wise correlation coefficient analysis of read counts against different peak files as described in Materials and Methods. Histograms of average tag-intensity (≈≡ read-counts) as a function of distance from the peak summits of each reference peaks file were generated as a quick validation for all pair-wise analysis

(data not shown). To illustrate the point, histograms of USF2, USF1 and HIF-2α mapped reads against USF2 reference peak files is shown in Figure 3.1.19B.

A



B



C

Peak files	Centred correlation coefficient (whole dataset)						
	HIF-1α	HIF-1β	HIF-2α	c-Myc	MITF	USF1	USF2
HIF-1α		1.08	1.12	1.00	0.87	1.01	0.93
HIF-1β	1.85		1.63	0.95	0.38	0.54	0.66
HIF-2α	1.90	1.38		1.42	0.25	0.75	0.30
c-Myc	1.65	0.86	1.76		0.26	1.10	0.36
MITF	0.59	0.83	0.85	0.57		1.85	1.30
USF1	0.88	0.78	0.96	0.72	0.55		2.11
USF2	0.28	0.29	0.34	0.30	0.56	4.23	

Figure 3.1.19 Similarity in binding profile between ChIP-seq experiments. **A)** Summary graph of the relative distribution of ChIP-seq peaks stratified into different genome ontologies for filtered peaks (MITF top 6% of the peaks and peaks with score of > 10 for the rest). **B)** Tag-density of USF2, USF1 and HIF-2α around USF2 peaks. **C)** Colour coded table showing pair-wise correlation coefficient between each peak files against tag-density from other ChIP-seq experiments. For direct comparison across whole datasets, the correlation coefficients were centred to average correlation coefficient of all ChIP-seq experiments against each peak file such that the value in each row can be directly compared. Colour coding such that higher intensity of green reflects higher correlation while more intense red reflects lower correlation.

As the peak summit is defined as a position with ≈ equal number of 5' $(+)$ and 3' $(-)$ tags, a plot of ChIP-seq tag density of USF2 around USF2 peaks should give a mirror image of $+$ and $-$ tags with the axis of symmetry around the peaks' summit (Figure 3.1.19B, USF2). For USF1 whose incidence of co-occupancy with USF2 is high and the target specificity is also very similar to USF2, the average tag-density around USF2 peaks will be high (Figure 3.1.19B, USF1). The lower tag-density of HIF-2α around USF2 peaks indicates either: a low level of co-occupancy or a weak binding by HIF-2α to USF2 targets (Figure 3.1.19B, HIF $-$ 2α). As the average fragment size is not identical across all ChIP-seq experiments, and as each bHLH-TF may have different bias for different E-boxes presented in tandem, we do not expect the calculated peak summits to be identical across different experiments. Hence, we do not observe sharp drops in $+$ & $-$tags when tracing past the peak summit. This applied to all of our ChIP-seq datasets (data not shown).

Pairwise correlation coefficients for all ChIP-seq mapped files against each peak file were calculated before centre normalisation at each peak file for direct comparison (Figure 3.1.19C). As expected, USF1 and USF2 that can heterodimerise demonstrate the highest correlation coefficient. Although USF2 peaks are heavily biased towards co-occupancy by USF1 (almost 8 $-$fold over MITF, the second most correlated), the reverse is not true. The correlation coefficient of USF2 co-occupancy with USF1 is only ≈ 2 $-$fold over the second highest, HIF-2α.

In line with the reported role of USF1 as a mediator of stress-signalling and tanning in response to UVB (Galibert et al., 2001), MITF shows the highest correlation to USF1 and to a lesser extent USF2 around MITF peaks. MYC, on the other hand, displays the highest correlation coefficient with HIF-αs around MYC's peaks. This is also in line with the reported cross-interaction between MYC and HIF-α (To et al., 2006; Koshiji et al., 2004; Koshiji et al., 2005; Gordan et al., 2007a). USF1, too, displays a high correlation with MYC binding at MYC's peaks with $\approx 3-$ fold greater correlation than USF2 and $\approx 4-$ fold greater correlation than MITF. This is also in line with the reported regulation of genes such as

hTERT by USF1 and MYC (Chang et al., 2005; Goueli and Janknecht, 2003). Lastly, HIF displays a high correction coefficient among themselves as would be expected. HIF-1α and HIF-2α have a higher correlation coefficient at all HIF peak files than HIF-1β at HIF-α peak file. This may partly be explained by the fact that HIF-1β (aka ARNT) can also bind to different motifs as a heterodimer with AhR and that HIF-α can bind in a complex with SP1 independent of HIF-1β (Koshiji et al., 2005; Gordan et al., 2007a).

Having performed pairwise-correlation analysis of the whole dataset, we next looked to see if we could obtain a similar correlation between bHLH-TFs at each peak file when we only consider peaks which show differential binding by the bHLH-TFs studied. Using the peak files populated with tag-counts from different ChIP-seq tag-files generated above, we filtered out peaks with low differences among different bHLH-TFs such that each dataset contained only ≈ 2500 peaks before hierachical clustering and present them in the form of heat map which retains detail at an individual peak level (Figure 3.1.20). In this analysis, the correlation between each pair of bHLH-TFs at each peak file is largely in good agreement with the correlation coefficient obtained from the whole dataset. For example, HIFs always cluster together, and USF1 and USF2 displayed the greatest correlation around USF2 peaks. However, some differences exist. Unlike the overall correlation coefficient analysis, which include peaks that contains E-box motifs that are bound similarly by all of the bHLH-TFs studied, clustering analysis, which only keep peaks which are differentially bound by the bHLH-TFs studied, would be enriched for peaks which harbour E-box motifs that favour binding by a subset of bHLH-TFs. Whilst USF1 and USF2 show the highest correlation coefficient around USF1 peaks, clustering of differentially bound peaks put MITF closer to USF1 than USF2. This can be explained through motif preference where MITF (Supplementary Figure 3.1.4C) and USF1 (Supplementary Figure 3.1.8C) exhibit 5'-T flanking sequence preference while USF2 did not exhibit any preference for 5'-flanking nucleotide (Supplementary Figure 3.1.11C) at peaks with score of ≥ 10. Similarly, while HIF-2α display higher correlation coefficeint to HIF-1α around HIF-2α peaks, HIF-1β and HIF-

2α whose preferred motifs resemble a 5'-CACGTG-3'-type E-box cluster more closely than HIF-1α whose preferred motifs is a strong HRE at peaks with score of ≥ 10.

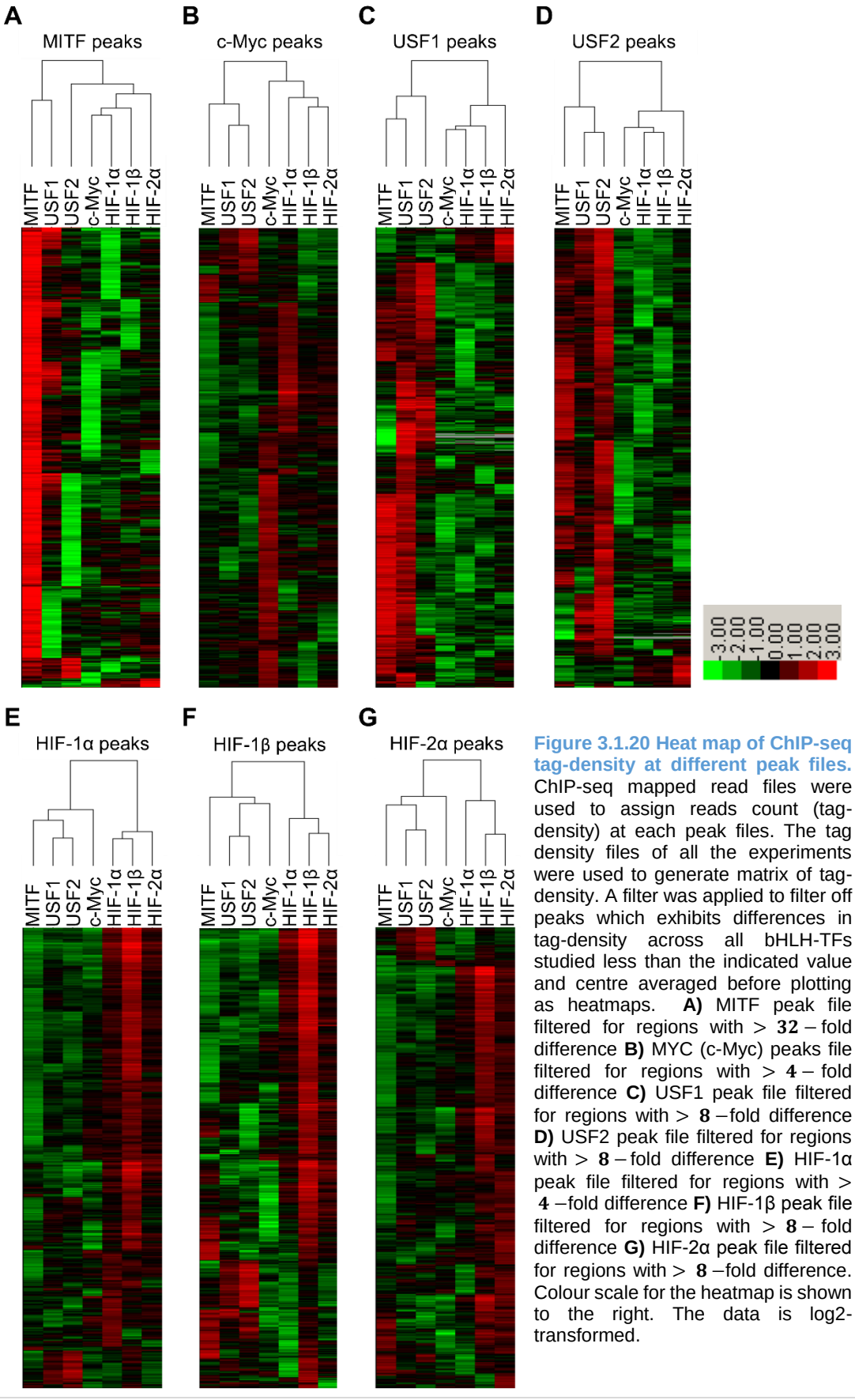


Figure 3.1.20 Heat map of ChIP-seq tag-density at different peak files. ChIP-seq mapped read files were used to assign reads count (tag-density) at each peak files. The tag density files of all the experiments were used to generate matrix of tag-density. A filter was applied to filter off peaks which exhibits differences in tag-density across all bHLH-TFs studied less than the indicated value and centre averaged before plotting as heatmaps. **A)** MITF peak file filtered for regions with $> 32-$ fold difference **B)** MYC (c-Myc) peaks file filtered for regions with $> 4-$ fold difference **C)** USF1 peak file filtered for regions with $> 8-$fold difference **D)** USF2 peak file filtered for regions with $> 8-$fold difference **E)** HIF-1α peak file filtered for regions with $> 4-$fold difference **F)** HIF-1β peak file filtered for regions with $> 8-$ fold difference **G)** HIF-2α peak file filtered for regions with $> 8-$fold difference. Colour scale for the heatmap is shown to the right. The data is log2-transformed.

Motifs of the co-occupied peaks

To investigate the bias in the motif preference at co-occupied peaks, we ran *de novo* motif identification using sequence under co-occupied peaks (Although uniquely occupied sites are fascinating in their insight to how additional regulatory mechanisms such as cooperative binding with (an)other TF(s) may confer unique specificity, this is beyond the scope of this thesis). We hypothesised that co-occupied sites are enriched in high affinity E-boxes, because: it is unlikely that multiple bHLH-TFs studied would adopt similar non-sequence specific DNA-binding, and that non-E-box targets with low peak score may either be non-biologically relevant or artefacts of ChIP-seq experiments. Therefore, sites where bHLH-TFs co-occupied should harbour a clear E-box. High affinity E-boxes also imply that read-counts at co-occupied sites are likely to be higher than the average of the whole dataset. Of course, there may also be super-high affinity, unique targets as a result of cooperative binding. However, these are likely to be a minority when considering the whole dataset.

Assuming direct proportionality between ChIP-seq coverage and the relative occupancy by each of the factors bound at co-occupied sites, we use ChIP-fragment depth at the peak summit to calculate the relative change in ChIP-fragment depth at co-bound peaks relative to the whole dataset for each bHLH-TF (Supplementary Table 3.1.4). In all cases, the ChIP-fragment depth of all bHLH-TFs at common targets have, on average, higher fragment depth. In other words, relative to the full-dataset, commonly bound sites appear to have a higher tag-density for all bHLH-TFs bound.

Based on the motif identified *de novo* (Summarised in Supplementary Figure 3.1.33 – 3.1.34) and correlation analysis (Figure 3.1.19C & Figure 3.1.20) we know that: USF1 and USF2, USF1 and MITF, USF1 and MYC, HIF-1α and USF1, and HIF-1α and HIF-2α are most similar in their targets, overall. As these co-occupied sites are not preferentially occupied by the remaining bHLH-TFs studied, we anticipated different signature motifs at common targets which dictate preferential binding by one and not another bHLH-TF. To address this hypothesis, we perform *de novo* motif identification using sequence from peaks

at co-occupied peaks (labelled as “Common peaks” column in Figure 3.1.21, 3.1.23 – 3.1.27). The common and unique sites are presented as Venn-diagrams for each set of bHLH-TFs (Part A of Figure 3.1.21, 3.1.23 – 3.1.27).

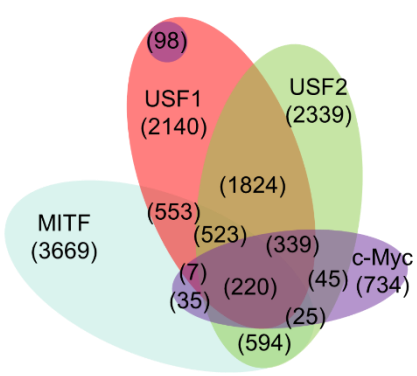
MITF, MYC, USF1 and USF2 common targets

For all combinations of MITF, MYC and USF a large fraction of co-occupied sites harbour the consensus motifs ($46 - 71\%$). The motif for USF1 + MYC (664 peaks) co-occupied sites is identical to USF1 (5’-[A/G]T-CACGTG-3’) suggesting that 5’-TA-3’ core-dinucleotide disfavours USF1 binding (Figure 3.1.21B, USF1 + c – Myc) (the MYC motif identified from the whole dataset allow 5’-TA-3’ while USF1 motif was a clean 5’-CG-3’ core-dinucleotide). As a 5’-TA-3’ core-dinucleotide is found at USF1 co-occupied site with MITF and USF2, but not with MYC, MYC likely exhibit little discrimination between 5’-TA-3’ and 5’-CG-3’ core-dinucleotide. Moreover, although MYC preferred a G at the -2 position of the 5’-flanking sequence (5’-GT-CACGTG), MYC binding is not prohibited by an A at the -2 position of the 5’-flanking sequence (5’-AT-CACGTG). As a 5’-TA-3’ core-dinucleotide can be found at USF1 + MITF (1283 peaks) and USF1 + USF2 (2906 peaks) co-occupied sites, despite being disfavoured by USF1, the relative preference of USF2 and MITF for 5’-TA-3’ over 5’-CG-3’ core-dinucleotide is greater than the unfavourability of 5’-TA-3’ over 5’-CG-3’ core-dinucleotide on USF1 binding.

Both MITF and USF1 have a similar flanking sequence preference, so their co-occupied sites also have 5’-T & 3’-A nucleotides flanking the core E-box. For USF2, that does not exhibit a strong flanking-sequence preference *in vivo* (Supplementary Figure 3.1.11C, All) or *in vitro* (Figure 3.1.15D), the flanking-sequence preference is dictated by USF1 at co-occupied sites. Conversely, the motif at MYC + MITF common sites (261 peaks) is dictated by the unfavourability of 5’-T + 3’-A nucleotides flanking the core E-box on MYC’s binding even though these flanking sequence represent MITF’s highest affinity targets.

Similar reasoning can be applied to USF2 + MITF (1362 peaks) and USFs + MITF (743 peaks). Interestingly, higher order co-occupancy, such as USFs + MYC (559 peaks) and USFs + MYC + MITF (220 peaks) have a more degenerate requirement for flanking sequences. One possible explanation is that although highly similar in their sequence preference, the slight differences in the preferred flanking sequences mean that to be able to accommodate all the bHLH-LZ-TFs queried, the sequence must be more degenerate.

A



B

Motif	P-value	Best matches	% of query	Common peaks
	$1e^{-172}$	USF1 (GM12878 Homer)	46	USF1 + c-Myc
	$1e^{-722}$	USF1 (GM12878 Homer)	57	USF1 + MITF
	$1e^{-1020}$	USF1 (GM12878 Homer)	53	USF1 + USF2
	$1e^{-221}$	USF1 (GM12878 Homer)	59	USF2 + c-Myc
	$1e^{-767}$	USF1 (GM12878 Homer)	78	USF2 + MITF
	$1e^{-61}$	USF1 (Jaspar)	55	c-Myc + MITF
	$1e^{-206}$	USF1 (GM12878 Homer)	49	USFs + c-Myc
	$1e^{-481}$	USF1 (GM12878 Homer)	71	USFs + MITF
	$1e^{-134}$	USF1 (GM12878 Homer)	63	USFs + c-Myc + MITF
	$1e^{-11}$	Max (Jaspar)	5	USFs + c-Myc + MITF

Figure 3.1.21 MITF, MYC, USF1 and USF2 co-occupied targets. **A)** Venn-diagram showing unique and overlapping peaks with peak score of > 10 for MYC, USF1 and USF2 and the top 6% of all peaks sorted by peak score for MITF. **B)** Motifs identified *de novo* from sequences under common peaks of indicated proteins (right hand column) along with percentage of occurrence within the queried sequence, best matches from the database and P-values.

Alternatively, co-occupied peaks may harbour multiple E-boxes of different type each of which are not necessarily bound equally by each of the bHLH-TFs co-occupying the site.

To gain insight into how exclusivity is achieved (that is, binding by some but not the other bHLH-TFs), as well as the observed more relaxed requirement for the E-box when considering peaks co-occupied by many bHLH-TFs, we inspected the E-box(es) under the peaks of various genes (Figure 3.1.22A – E). At the *QTRT1* promoter, an MITF-, MYC- and USF-specific target (no HIF), the E-boxes under the peak are both the 5'-T + 3'-A and 5'-C + 3'-G-type (Figure 3.1.22A). Whilst 5'-T + 3'-A disfavour MYC's binding, MYC is able to bind to the adjacent 5'-C + 3'-G E-box. When running *de novo* motif identification, both E-boxes will contribute to the motif identified, hence the more degenerate flanking sequence observed when considering sites co-occupied by many bHLH-TFs. We also looked at the *KIAA0556* gene, an MITF and USF common target (no MYC or HIF), where the three E-boxes under the peaks have different flanking sequences (Figure 3.1.22B). Why MYC does not bind at this locus is not clear. Moreover, there is no clear correlation between centre-normalised peak score and the E-box present. The same can be said for USF1 + USF2

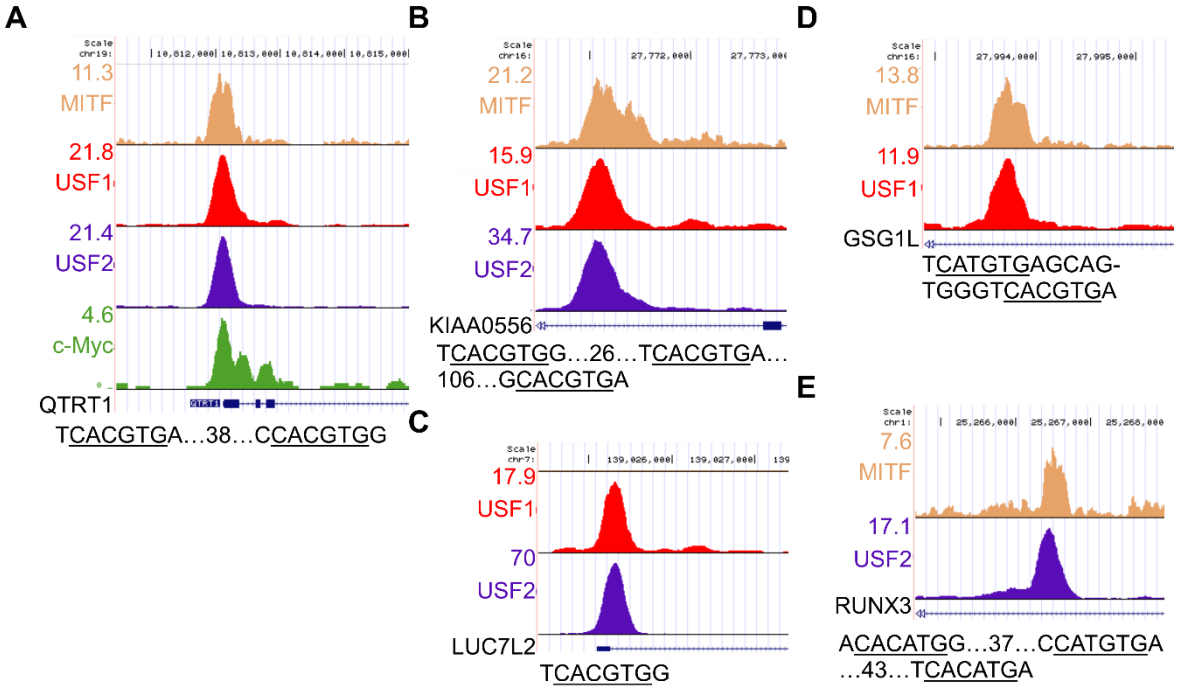


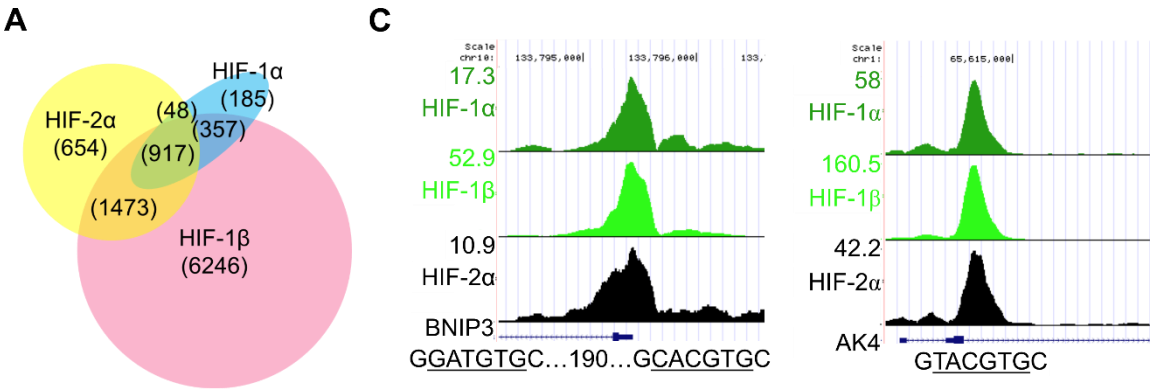
Figure 3.1.22 Snapshot of MITF, MYC, USF1 and USF2 bedgraphs at co-occupied sites. **A-E)** Snapshots of bedgraphs uploaded on to the UCSC genome browser at: the *QTRT1* gene, a common MITF, MYC and USF target (**A**), the *KIAA0556* gene, a common MITF and USF target (**B**), the *LUC7L2* gene, a common MITF, USF1 and USF2 target (**C**), the *GSG1L* gene, a common MITF and USF1 target (**D**), the *RUNX3* gene, a common MITF and USF2 target (**E**). The E-boxes under the peaks are shown under the graph where the core E-boxes are underlined.

(Figure 3.1.22C), MITF + USF1 (Figure 3.1.22D) and MITF+ USF2 (Figure 3.1.22E) unique peaks. In all cases, it was not clear how exclusivity is achieved based on the E-box sequence under the peak alone.

It is worth mentioning that we had to go through a large number of "unique targets" before finding peaks which we can confidently say are truly unique based on visual inspection of the bedgraphs uploaded onto the UCSC genome browser. This is because, we only use peaks with score of > 10 to minimize the false-positives due to the use of background files prepared using a different protocol, therefore many "unique targets" are, in fact, also occupied to a much lesser extent by other bHLH-TFs.

HIF common targets

The similarity in the motifs identified for HIF-1α + HIF-1β (1268 peaks) and HIF-2α + HIF-1β (2390 peaks) (Figure 3.1.23B) explains why we observed both HIF-1α + HIF-1β binding



B

Motif	Common peaks	% of query	Best matches	P-value
	HIF-1α + HIF-1β	34	HIF-1α (MCF7 Homer)	$1e^{-95}$
	HIF-1β + HIF-2α	26	bHLHE40 (HepG2 Homer)	$1e^{-85}$
	HIF-1α + HIF-2α	22	HIF-1α (MCF7 Homer)	$1e^{-62}$
	HIFs	13	E-box (Homer)	$1e^{-21}$

Figure 3.1.23 HIF1-α, HIF-1β and HIF-2α co-occupied targets. **A)** Venn-diagram showing unique and overlapping peaks with peak scores of > 10 for HIF. Snapshot of bedgraphs uploaded to the UCSC genome browser at: *BNIP3* and *AK4* genes (**B**). The hypoxia responsive elements (HREs) under the peaks are shown under the graph where the core HRE is underlined. **C)** Motifs identified *de novo* from sequences under common peaks of indicated proteins (right hand column) along with percentage of occurrence within the queried sequence, best matches from the database and P-values.

Table 3.1.6

Gene	Peak height normalised to average peak score			Motifs
	HIF-1α	HIF-1β	HIF-2α	
PDK1	3.67	6.38	2.09	C-CACGTC-C C T-CACGTA-C
MITF	0.35	0.71	0.52	T-GACGTG-G
BNIP3	1.58	2.87	1.01	G-GATGTG-C…190…G-CACGTG-C
AK4	5.29	8.70	3.91	G-TACGTG-C

Normalised peak height (to average peak score) for direct comparison of HIF-1α, HIF-1β and HIF-2α binding at *PDK1*, *MITF*, *BNIP3* and *AK4* gene. The E-box/HRE motifs identified under the peaks are indicated in the right hand column. The core E-box/HRE is underlined.

at all the reported HIF-1α or HIF-2α unique targets we looked at. Moreover, we identified a strong E-box motif at HIF-1α + HIF-2α (965 peaks) and HIF co-occupied targets which explain the observed binding of HIF at many MYC, MITF and USF targets.

In contrast to USF + MYC + MITF datasets, HIF co-occupied sites have a relatively low occurrence of E-boxes ($13 - 34\%$) at common targets raising a question of what are the recognition motifs for other co-occupying sites. Inspection of secondary motifs at co-occupied sites did not yield any E-boxes or HREs (data not shown). Since analysis of secondary motifs of each HIF individually (Supplementary Figure 3.1.15, 3.1.17 & 3.1.19) also did not identify any unique motif for each of the HIF, the requirement for binding at other co-occupied sites remain unsolved.

As is the case for MYC, MITF and USF common targets, there is no clear correlation between motifs and peak height or peak score (Figure 3.1.22C & Table 3.1.6). The one instance where HIF-2α has a higher occupancy relative to HIF-1α (at MITF), the motif is actually closer to a HIF-1α consensus motif.

Co-occupancy of HIF with MITF, MYC and USF

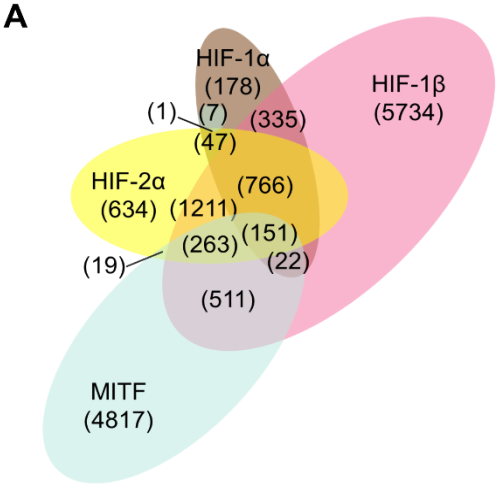
In this section we looked at motifs at sites co-occupied by HIF and each of MITF, MYC and USF (Figure 3.1.24 – 3.1.27A). Motifs identified *de novo* from MITF, MYC and USF peaks co-occupying with HIF are highly similar across each of the HIF co-bound sites.

In all instances of MITF and HIF-1α (181 peaks) and HIF-1β (947 peaks) co-occupied sites, the motifs are dictated by the more stringent requirement of MITF, 5'-RT-CACGTG-3'

(Figure 3.1.24B). The motifs for HIF-2α and MITF (434 peaks) is also similar to MITF's motif derived from the full-dataset (Supplementary Figure 3.1.5), albeit with a higher weight of T and A at the core-dinucleotide. In all cases, the 5'-C & 3'-G outermost of the core E-box is essentially invariable as these are critical for the binding of MITF (Figure 3.1.13A). The motifs of HIF-2α + HIF-1β + MITF (414 peaks), however, is more relaxed, with the core-E-box similar to that of HIF-2α, having higher weight of T and A at the core-dinucleotide, suggesting that the 5'-TA-3' core-dinucleotide, whilst being preferred by MITF and HIF-2α, disfavour binding by HIF-1α. This is supported by the absence of E-box motifs enriched among HIF + MITF common targets.

All of the motifs identified enriched at MYC and HIF co-occupied sites (980 peaks, Figure 3.1.25B) are very similar to those observed for sites co-occupied by MITF and each of the

A

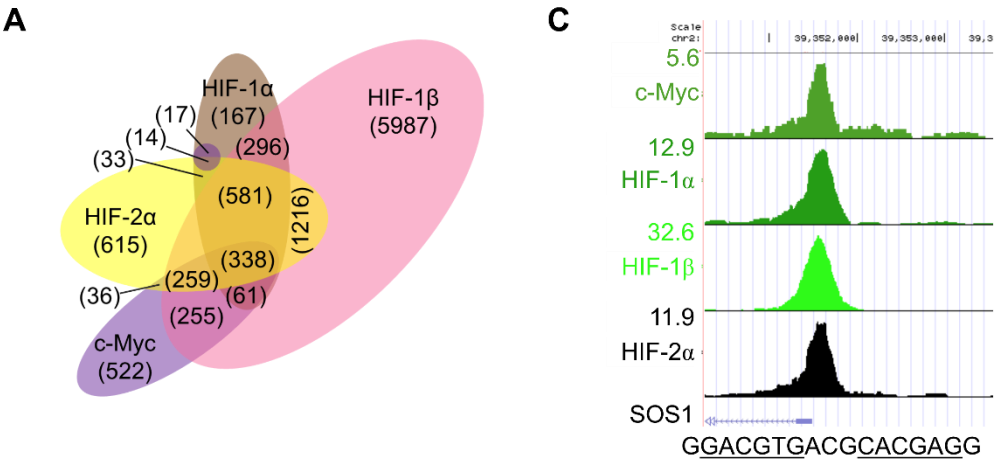


B

Motif	Common peaks	% of query	Best matches	P-value
	MITF + HIF-1α	32	USF2 (C2C12 Homer)	$1e^{-32}$
	MITF + HIF-1β	28	USF1 (GM12878 Homer)	$1e^{-180}$
	MITF + HIF-2α	35	USF2 (C2C12 Homer)	$1e^{-83}$
	HIF-2α + HIF-1β + MITF	31	USF1 (GM12878 Homer)	$1e^{-74}$

Figure 3.1.24 MITF, HIF1-α, HIF-1β and HIF-2α co-occupied targets. **A)** Venn-diagram showing unique and overlapping peaks with peak score of > 10 for HIF and the top 6% of all peaks sorted by peak score for MITF. **B)** Motifs identified *de novo* from sequences under common peaks of indicated proteins (right hand column) along with percentage of occurrence within the queried sequence, best matches from the database and P-values.

HIFs. The core E-box sequence at MYC and HIF common peaks are much more defined than each of the HIFs alone (Supplementary Figure 3.1.34). The flanking sequence preference is also more relaxed than that identified for MYC and are more similar to that identified for each of the HIFs. As 5'-CACGTG-3' is also a target for MITF and USF, MYC and HIF unique sites (no MITF and USF) may bear sequences dissimilar to 5'-CACGTG-3' as exemplified in (Figure 3.1.25C) where the E-box motifs are 5'-GACGTG-3' and 5'-CACGAG-3'. The relatively low percentage of E-box motif occurrence within HIF + MYC



B

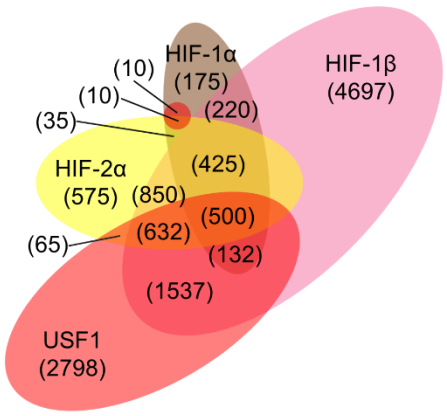
Motif	Common peaks	% of query	Best matches	P-value
	c-Myc + HIF-1α	13	bHLH-E40 (HepG2 Homer)	$1e^{-19}$
	c-Myc + HIF-1β	19	USF1 (GM12878 Homer)	$1e^{-59}$
	c-Myc + HIF-2α	18	USF1 (GM12878 Homer)	$1e^{-35}$
	HIF-1α + HIF-1β + c-Myc	25	CLOCK (Liver Homer)	$1e^{-20}$
	HIF-2α + HIF-1β + Myc	26	CLOCK (Liver Homer)	$1e^{-35}$
	HIFs + c-Myc	25	CLOCK (Liver Homer)	$1e^{-19}$

Figure 3.1.25 MYC, HIF1-α, HIF-1β and HIF-2α co-occupied targets **A)** Venn-diagram showing unique and overlapping peaks with peak score of $>$ **10** for MYC and HIF. Snapshot of bedgraph uploaded on to UCSC genome browser at *SOS1* gene (**B**). The E-boxes under the peaks are shown under the graph where the core E-boxes are underlined. **C)** Motifs identified *de novo* from sequence under common peaks of indicated proteins (right hand column) along with percentage of occurrence within the queried sequence, best matches from the database and P-values.

common peaks is a little puzzling and implicates the existence of additional factor(s) which bring MYC and HIF to other common targets in an E-box/HRE independent manner.

As demonstrated *in vivo* (Supplementary Figure 3.1.33 & 3.1.34) and *in vitro* (Figure 3.1.13C), USF1 has a requirement for the 5'-CG-3' core di-nucleotide, the motifs derived *de novo* for USF1 co-occupied sites with HIF (2886 peaks) are almost invariably 5'-CACGTG-3' (Figure 3.1.26B). The strong preference for 5'-GT nucleotide immediately flanking the core E-box at sites co-occupied by USF1 and each of the HIFs is not as strongly enriched among motifs identified from USF1, HIF-1β and HIF-2α individually and is absent at HIF-1α sites altogether. This is true for motifs derived from both the full- (Supplementary Figure 3.1.33) and filtered-datasets (Supplementary Figure 3.1.34). A 5'-T nucleotide immediately flanking

A

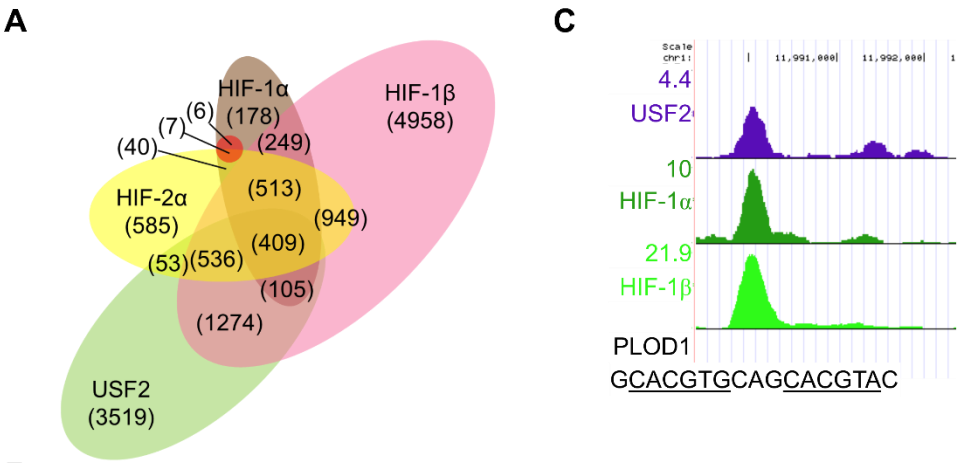


B

Motif	Common peaks	% of query	Best matches	P-value
	USF1 + HIF-1α	32	USF1 (GM12878 Homer)	$1e^{-67}$
	USF1 + HIF-1β	24	USF1 (GM12878 Homer)	$1e^{-256}$
	USF1 + HIF-2α	26	USF1 (GM12878 Homer)	$1e^{-120}$
	HIF-1α + HIF-1β + USF1	30	CLOCK (Liver Homer)	$1e^{-52}$
	HIF-2α + HIF-1β + USF1	23	USF1 (GM12878 Homer)	$1e^{-90}$
	HIFs + USF1	28	USF1 (GM12878 Homer)	$1e^{-41}$

Figure 3.1.26 USF1, HIF1-α, HIF-1β and HIF-2α co-occupied targets **A)** Venn-diagram showing unique and overlapping peaks with peak score of $>$ **10** for USF1 and HIF. **B)** Motifs identified *de novo* from sequences under common peaks of indicated proteins (right hand column) along with percentage of occurrence within the queried sequence, best matches from the database and P-values.

the core sequence is, however, found enriched at motifs derived from USF1. Moreover, the sequence of the core E-box enriched among USF1 and HIF co-occupied sites are strictly 5'-CACGTG-3', similar to that of USF1. This would suggest that USF1 and HIF co-occupied sites are strong USF1 targets and may represent secondary hypoxic target genes. Immediately after exposure to hypoxia, HIF binds first to high affinity sites, most probably cooperatively with additional factor(s) as motifs derived from peaks with score of > 10 are more degenerate (Supplementary Figure 3.1.34). As HIF-αs further accumulate, they would compete for sites bound by other bHLH-TFs such as the USF. At USF + HIF co-occupied



B

Motif	Common peaks	% of query	Best matches	P-value
	USF2 + HIF-1α	40	USF1 (GM12878 Homer)	$1e^{-99}$
	USF2 + HIF-1β	33	USF1 (GM12878 Homer)	$1e^{-396}$
	USF2 + HIF-2α	38	USF1 (GM12878 Homer)	$1e^{-193}$
	HIF-1α + HIF-1β + USF2	33	USF1 (GM12878 Homer)	$1e^{-68}$
	HIF-2α + HIF-1β + USF2	35	USF1 (GM12878 Homer)	$1e^{-133}$
	HIFs + USF2	32	USF2 (C2C12 Homer)	$1e^{-52}$

Figure 3.1.276 USF2, HIF1-α, HIF-1β and HIF-2α co-occupied targets **A)** Venn-diagram showing unique and overlapping peaks with peak score of > 10 for USF2 and HIF. Snapshot of bedgraphs uploaded on to UCSC genome browser at *PLOD1* (**B**). The E-boxes under the peaks are shown under the graph where the core E-boxes are underlined. **C)** Motifs identified *de novo* from sequences under common peaks of indicated proteins (right hand column) along with percentage of occurrence within the queried sequence, best matches from the database and P-values.

genes, USF may prevent premature activation while keeping these sites nucleosome-free for HIF binding.

Essentially identical motifs were identified for USF2 sites co-occupied with HIF (2390 peaks, Figure 3.1.276B) with the only exception being a more relaxed requirement for the −2 nucleotide flanking sequence in all cases (primarily, 5'-RT-CACGTG-3') and additionally for the −1 flanking sequence for HIF-2α (5'-NN-CACGTG-3'). The motif identified for HIF-2α also displayed an allowance for 5'-TA-3' in the core-di-nucleotide (5'-CA[C/T][G/A]TG), reflecting the ability of USF2 and HIF-2α to bind to these E-boxes.

An example of USF2 + HIF-1α + HIF-1β (409 peaks) unique sites is shown in Figure 3.1.27C. Based on our data from section 3.1.5, we anticipated such unique sites would not have 5'-T flanking sequence (as these are preferred by MITF and USF1), and the core-dinucleotide to be 5'-TA-3' (as these would disfavour USF1). However, the E-boxes presented at the *PLOD1* gene contradict our expectations with regard to the core-dinucleotide (5'GCACGTGCAGCACGTAC-3'). Whilst HIF-1α binds as expected based on the presence of an HRE (in the second of the two E-boxes), why USF1 is excluded from this site is not clear. This, and the lack of clear correlation between peak height and the motif under the peak (further exemplified in Supplementary Figure 3.1.35), as discussed earlier, can undermine the use of signature motifs to predict relative occupancy by bHLH-TFs of interest.

Relative binding at co-occupied sites and comparability of ChIP-seq and qPCR

Information obtained from our ChIP-seq datasets are very useful for identifying co-occupied targets which represent potential candidates for factor exchange. However, it would be even better if we could directly translate normalised peak scores to relative occupancy by a given set of bHLH-TFs. (For example, a two-fold higher normalised peak score of MITF over USF1 indicated that the incidence of MITF being found at that site is twice that of USF1.) Numerous sources of variability in ChIP-seq experiments may influence the

comparability of different ChIP-seq experiments. For examples, ChIP-efficiency, fragment length, variable background reads, bias during amplification and library preparation, and most importantly each of our ChIP-seq dataset (except for the HIFs) were generated using different cell lines; although they were all derivatives of 501mel cells, clonal section during stable cell line generation may result in cells with slightly different phenotypes (some clones were visibly more pigmented than others, data not shown).

As it is generally accepted that comparative studies within a given ChIP-seq data set is acceptable while comparison across different dataset is not because there is no good way of normalising ChIP-seq data like ChIP-qPCR, we adopted a simple centre normalisation of peak score against the average peak score of the whole dataset before direct comparison to ChIP-qPCR fold-enrichment (Figure 3.1.28A & 3.1.287B). Centre normalisation was carried out to account for sample-to-sample variation and aimed to present each data point as a value relative to the whole dataset. Centre normalised values should take into account all common sources of variability such as cross-linking efficiency, antibody specificity and background noise.

We performed ChIP-qPCR on 11 –targets that do not overlap with repetitive elements and multiple, closely spaced E-boxes that may not be resolved by ChIP-qPCR and ChIP-seq. We centre normalised the ChIP-qPCR-fold-enrichment (Figure 3.1.28A) and peak scores (Figure 3.1.28B) before plotting them as a stack bar graph presenting the centre-normalised peak score as the percentage of total centre-normalised peak score by MITF, USF2 and USF1 at each target. At first glance the relative binding by each bHLH-LZ-TFs appears to correlate quite well between the two-techniques. Closer inspection, however, reveals that only 3 of the 11 targets are identical with regard to the relative binding by each bHLH-TF as inferred from ChIP-seq and ChIP-qPCR data. Thus, although giving us good insight into which bHLH-TF is bound to a given target, the ChIP-seq peak score cannot be used directly to determine the relative occupancy by each of the bHLH-TFs.

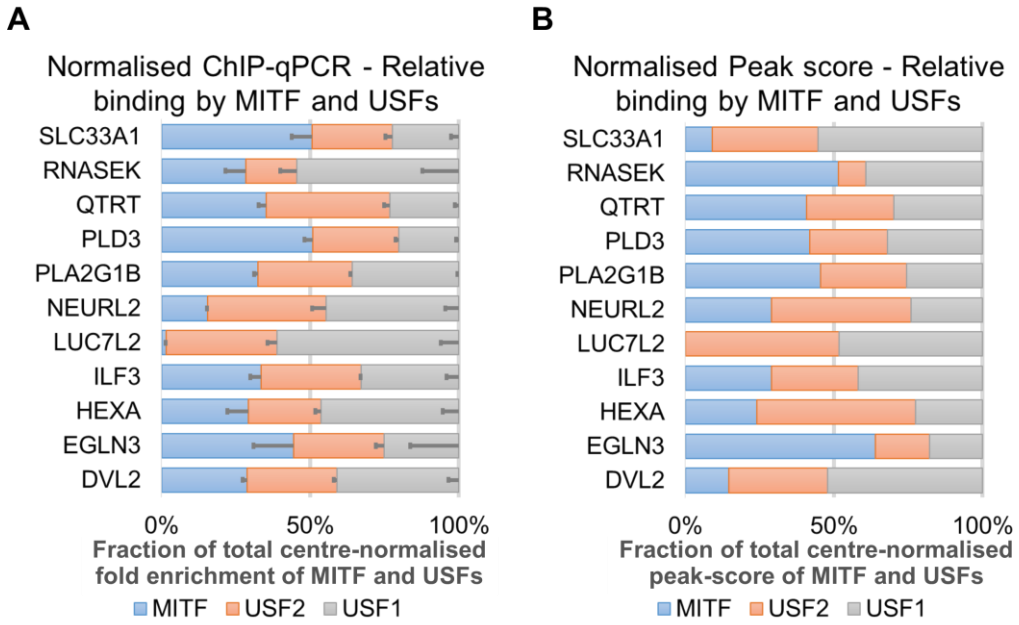


Figure 3.1.28 Relative occupancy of MITF, USF1 and USF2 at some targets. **A)** Stack bar graph of centre normalised ChIP-qPCR fold enrichment for MITF, USF1 and USF2 at indicated genes (on the left) performed in 501mel stable cell lines used to generate the ChIP-seq datasets. ChIP-qPCR of all the genes listed were run using the same ChIPped sample. To account for the sample-to-sample variation (that is between cell lines) the fold enrichment at each gene are centre normalised against the average fold-enrichment of all the genes looked at for each proteins. The centre normalised values were used to plot the stack bar graph, presenting the centre-normalised value as percentage of total centre-normalised fold enrichment by MITF, USF2 and USF1. $\mathbf{N} = \mathbf{2}$ biological replicates each with **2** systematic replicates. **B)** Stack bar graph of centre normalised ChIP-seq peak score for MITF, USF1 and USF2 at indicated genes (on the left). Centre normalisation of peak score was carried out to account for the variability between ChIP-seq samples. Centre normalised peak scores for each proteins were used to plot the stack bar graph, presenting the centre-normalised peak score as percentage of total centre-normalised peak score by MITF, USF2 and USF1.

3.1.8 501mel specific peaks

One of the questions we set out to investigate was whether 501mel-specific targets are also targets of MITF. In other words, are 501mel-specific targets unique to 501mel cells because 501mel cells express MITF that is not present in any of the cell lines used to generate the publicly available ChIP-seq data? To address this question we re-analysed a large number of publicly available datasets for USF, MYC and HIF. We performed sequential filtering for 501mel specific peaks against four cell lines for MYC (A549, K562, HeLa-S3 and NB4), USF1 (A549, K562, GM12878 and HepG2) and USF2 (HeLa-S3, K562, GM12878 and HepG2) and against ccRCC for HIF. Filtering resulted in 186 501mel-specific MYC's peak, 377 501mel-specific USF1's peak. 356 501mel-specific USF2's peak, 1633 501mel-specific HIF-1α's peak, 2629 501mel-specific HIF-1β's peak and 1892 501mel-specific HIF-2α's peak.

Table 3.1.7 Average Peak score of 501mel-specific peaks

ChIP	Average Peak score (501mel-specific peaks)	Average Peak score (Full dataset)
MYC	11.21	8.36
USF1	22.75	12.20
USF2	27.56	15.66
HIF-1α	11.35	10.96
HIF-1β	15.14	18.46
HIF-2α	12.57	12.57

The average peak score of 501mel specific peaks (Table 3.1.7) is higher than the average peak score of the full-ChIP-seq datasets (MYC, $8.36 \rightarrow 11.21$; USF1 $12.20 \rightarrow 22.75$; USF2 $15.66 \rightarrow 27.56$; HIF-1α $10.96 \rightarrow 11.35$ and HIF-2α, $10.76 \rightarrow 12.57$) with the exception of HIF-1β ($18.46 \rightarrow 15.14$) indicating that 501mel-specific peaks may be biased towards higher affinity targets.

We performed GO-analysis on 501mel-specific peaks and looked for enrichment of related GO-terms as well as melanocytic-lineage-related terms (Data not shown but summarized in Table 3.1.8). With the exception of 501mel-specific USF2 peaks, which have one

Table 3.1.8 Summary table for analysis of top 50 GO-terms for 501mel specific-peak

MYC	No. of metabolic related term(s)	4
	No. of biosynthetic related term(s)	3
	Average Coverage (%)	11.4
USF1	No. of immune response related term(s)	34
	No. of differentiation related term(s)	2
	Average Coverage (%)	62.1
USF2	No. of differentiation related term(s)	6
	No. of metabolism related term(s)	3
	No. of catabolism related term(s)	7
	No. of development related term(s)	9
	No. of growth factor signalling related term(s)	7
	No. of pigmentation related term(s)	1
	Average Coverage (%)	30.7
HIF-1α	No. of metabolism related term(s)	10
	No. of response to oxygen levels related term(s)	1
	No. of development related term(s)	7
	No. of differentiation related term(s)	5
	No. of protein modification related term(s)	5
	Average Coverage (%)	10.9
HIF-1β	No. of development related term(s)	22
	No. of differentiation related term(s)	9
	Average Coverage (%)	14.7
HIF-2α	No. of metabolism related term(s)	10
	No. of catabolism related term(s)	3
	No. of development related term(s)	6
	No. of differentiation related term(s)	4
	Average Coverage (%)	15.6

pigmentation-related term (7% coverage) within the top 50 GO-terms, none of the other datasets contain melanoma/melanocyte specific terms. There are some biased toward enrichment of specific terms with some datasets, for example, HIF-α peaks are enriched in metabolic processes, while USF1 peaks are enriched in immune-response related processes.

As expression of pigmentation-related genes are specific to melanocyte lineage, yet we did not observed general enrichment of pigmentation-related GO-term, we specifically looked for melanoma/melanocyte-specific genes manually. These genes are listed in Table 3.1.9. It is worth noting that the gene lists do not overlap, reflecting differential occupancy of melanoma/melanocyte-related genes by bHLH-TFs under investigation. The very low coverage, however, indicated that the E-boxes of many melanoma/melanocyte-related genes are occupied by MYC, USF and HIF in non-melanoma cell lines. As binding does not necessarily equate to gene expression, it is possible that these targets are occupied in a "treadmilling" mode (Lickwar et al., 2012) which does not result in gene activation. We will have to perform gene-expression analysis to address this.

To find 501mel-specific, co-occupied sites, we performed a series of peak file intersections of 501mel specific peaks. The final result of the 501mel-specific targets co-occupied by MYC, MITF and USF are HEAT repeat-containing protein 2, and ciliary rootlet coiled-coil, rootletin, both of which are related to cilia. The former contains five 5'-A-CACGTG-G-3' + three 5'-A-CACGTG-A-3' E-boxes while the latter contains a single 5'-C-CATGTG-A-3' E-box. As these proteins are non-melanoma specific, also bound by HIF and the E-boxes involved are also generic we did not pursue these genes further.

Brief summary of Chapter 3.1

We performed ChIP-seq analysis of seven highly related bHLH-TFs, in part, to generate database of co-occupying peaks for studying factor exchange following induction of a specific bHLH-TFs or interfering with the binding of one or more bHLH-TF(s) co-occupying

the site. Our data demonstrated high degree of overlapping targets that is reflective of their highly similar motifs enriched under the peaks. Despite similar target specificity, these factors demonstrate a rather different spread of binding across different genome ontologies with most HIF binding occurring around the promoter while most of MITF binding occurs within the intergenic region. There is relatively little change in the distribution of peaks into different genome ontologies when considering full- or filtered-datasets indicating that peak score is largely independent of genomic location and that non-promoter peaks are likely functional, for example, as distal enhancer or non-protein coding promoters.

GO analysis of MITF binding sites revealed almost a complete coverage in pigmentation-related process genes. HIF, on the other hand, demonstrated incomplete coverage of GO-terms related to response to oxygen indicating that many oxygen responsive genes may only be activated under specific conditions. Moreover, we did not detect HIF binding at a number of reported angiogenic and invasion and metastatic genes.

Motifs with the highest average peak scores, generally, are also the most prevalent motif. Whilst the most prevalent motif *in vivo* for MITF and USF is 5'-CACGTG-3', 5-CATGTG-3' is the highest affinity target in *in vitro* competition assay, at least when presented in the context of the *TYR* enhancer. With regard to the nucleotides flanking the core E-box, 5'-T + 3'-A compete significantly better than other combinations of flanking sequence in *in vitro* competition assay for MITF and TFE3. USF, on the other hand, is able to bind to any combination of flanking sequence *in vitro*, despite demonstrating 5'-T flanking nucleotide in the motif identified from the ChIP-seq experiment, while USF2, when considering only peaks with peak score ≥ 10, did not have any preferred flanking sequence. The difference in the flanking sequence specificity appears to reside within the third amino-acid of the basic domain. Mutation of USF's Ala to MITF's Asp result in USF that exhibits a clear preference for 5'-T + 3'-A nucleotide flanking the core E-box similar to MITF.

Our genome-wide analysis did not reveal cooperative binding by each of the bHLH-TF to the adjacent E-box under the same peak at a global level, both when the E-boxes are of the same type or of a different type, based on the average peak score alone. This does not, however, exclude the possibility that cooperative binding requires a specific spacing of the E-boxes and thus may only be found at specific promoters. For peaks containing a near match to the consensus E-box sequence we observed approximately double the average peak score when comparing peaks containing one E-box with those containing two while peaks containing non-consensus motifs generally have the same average peak score regardless of the number of motifs present. This argues against short-range binding cooperativity, at the global level. Interestingly, USF1 was the only factor where the average peak score of peak harbouring one or two near-consensus motif were roughly equal. However, the average peak score was significantly higher for peak with > 2 E-box.

We investigated bias in the sequence of the E-box at co-occupied sites. We also performed pairwise correlation coefficient which revealed the expected correlation, for example, USF1 and USF2 that can heterodimerise demonstrated the highest correlation coefficient at USF peaks. Similarly, HIF also displayed high level of correlation supportive of our observed lack of HIF-1α or HIF-2α specific targets in 501mel cells.

The motifs at co-occupied sites are dictated by the factor with the more stringent requirement in the E-box sequence it can bind. For example, USF1 + MYC common targets do not have 5'-T + 3'-A flanking sequence which is known to inhibit MYC's binding but retain 5'T- as presented in the motif identified for both factors individually. However, the motifs identified *de novo* from sequence under co-occupied peaks appear more degenerat as we increased the number of co-bound factors into consideration. This is most probably because to be able to accommodate all the bHLH-LZ-TFs queried, the sequence must be more degenerate as each bHLH-TF will have a slight difference in the preferred flanking sequence. Alternatively, co-occupied peaks may represent peaks with multiple E-boxes of different types under the peak.

Through sequential filtering of 501mel ChIP-seq peaks with peaks-called form publically available ChIP-seq datasets, we identified a number of 501mel specific peaks. With the exception of one GO-term related to pigmentation, we did not identify any pigmentation-related GO-term among 501mel specific peaks. Interestingly USF1 peaks are enriched in immune-response related processes. At an individual gene level, however, we were able to identify a number of melanoma specific genes, among 501mel specific USF and HIF peaks (Table 3.1.9).

Table 3.1.9 501mel-specific, melanoma/melanocyte related-genes

Gene Symbol	Genes	Nearest PromoterID	Peak files	Peak score		Genome ontology
				Raw	Norm	
DDT	D-dopachrome tautomerase	NM_001084392	USF1	21.4	0.94	promoter-TSS
MAGEC1	melanoma antigen family C, 1	NM_005462	USF1	14.2	0.62	Intergenic
MAGEC2	melanoma antigen family C, 2	NM_016249	USF1	11.2	0.49	Intergenic
MAGEA2	melanoma antigen family A, 2	NM_175742	USF1	7.68	0.34	intron (intron 3 of 5)
MAGEH1	melanoma antigen family H, 1	NM_014061	USF1	6.05	0.27	Intergenic
MAGEA6	melanoma antigen family A, 6	NM_175868	USF1	5.82	0.26	promoter-TSS
MAGEA9	melanoma antigen family A, 9	NM_005365	USF1	5.59	0.25	Intergenic
MAGEA6	melanoma antigen family A, 6	NM_005363	USF1	5.36	0.24	intron (intron 1 of 2)
MAGEA1	melanoma antigen family A, 1	NM_004988	USF1	5.12	0.23	Intergenic (-32034)
MAGEA1	melanoma antigen family A, 1	NM_004988	USF1	5.12	0.23	Intergenic (+73890)
MAGEA2B	melanoma antigen family A, 2B	NM_153488	USF1	5.12	0.23	Intergenic
DDT	D-dopachrome tautomerase	NM_001084392	USF2	32.4	1.18	promoter-TSS
DDTL	D-dopachrome tautomerase-like	NM_001084393	USF2	31.6	1.15	promoter-TSS
MAGEA2	melanoma antigen family A, 2	NM_175742	USF2	6.82	0.25	promoter-TSS (NM_001282505)
MAGEA1	melanoma antigen family A, 1	NM_004988	USF2	5.37	0.19	Intergenic
PRAME	preferentially expressed antigen in melanoma	NM_206955	HIF-1α	9.47	0.83	promoter-TSS (NR_027426)
TYRP1	tyrosinase-related protein 1	NM_000550	HIF-1β	19.5	1.29	Intergenic
MITF	microphthalmia-associated transcription factor	NM_198177	HIF-1β	38.6	2.55	intron (intron 2 of 9)
MITF	microphthalmia-associated transcription factor	NM_198178	HIF-1β	11.2	0.74	exon (NM_006722, exon 3 of 10)
TYR	tyrosinase	NM_000372	HIF-1β	18.6	1.23	Intergenic
TYR	tyrosinase	NM_000372	HIF-1β	9.76	0.64	promoter-TSS (NM_000372)
MLANA	melan-A	NM_005511	HIF-1β	15.8	1.04	intron (intron 2 of 4)
MC1R	melanocortin 1 receptor (alpha melanocyte stimulating hormone receptor)	NM_002386	HIF-1β	15.8	1.04	intron (NM_014972, intron 15 of 17)
MLPH	melanophilin	NM_001281473	HIF-1β	12.5	0.83	promoter-TSS (NR_104019)
MAGEF1	melanoma antigen family F, 1	NM_022149	HIF-1β	8.83	0.58	Intergenic
BAGE	B melanoma antigen	NM_001187	HIF-1β	20.4	1.35	Intergenic
MCHR1	melanin-concentrating hormone receptor 1	NM_005297	HIF-1β	19.1	1.26	Intergenic
PRAME	preferentially expressed antigen in melanoma	NM_206955	HIF-2α	10.5	0.84	promoter-TSS (NR_027426)
MCHR1	melanin-concentrating hormone receptor 1	NM_005297	HIF-2α	8.68	0.69	Intergenic
POMC	proopiomelanocortin	NM_000939	HIF-2α	8.68	0.69	exon (exon 3 of 3)

3.2 RNA-seq

RNA-seq was carried out to study transcription outcome at genes bound by each bHLH-TF characterised in our ChIP-seq analysis. Knowing the transcription outcome of over-expressing USF and MYC, hypoxia, and knocking-down/interfering with USF can provide us with clues as to the role played by each factor at selected genes for subsequent mechanistic analysis of promoter education. Although we have performed the popular transcript assembly and differential gene expression analysis using the Cufflink and Cuffdiff packages, these data are not of relevance to the theme of this thesis. Therefore, we will limit our discussion to correlations between differential gene expression and assignment to ChIP-seq peaks and principal component analysis (PCA) that informs us of the similarity between each dataset. We also limit our discussion to global analyses as it would be meaningless to discuss specific genes before we define specific targets of interest. Specific data for individual genes can be referred to in the database generated.

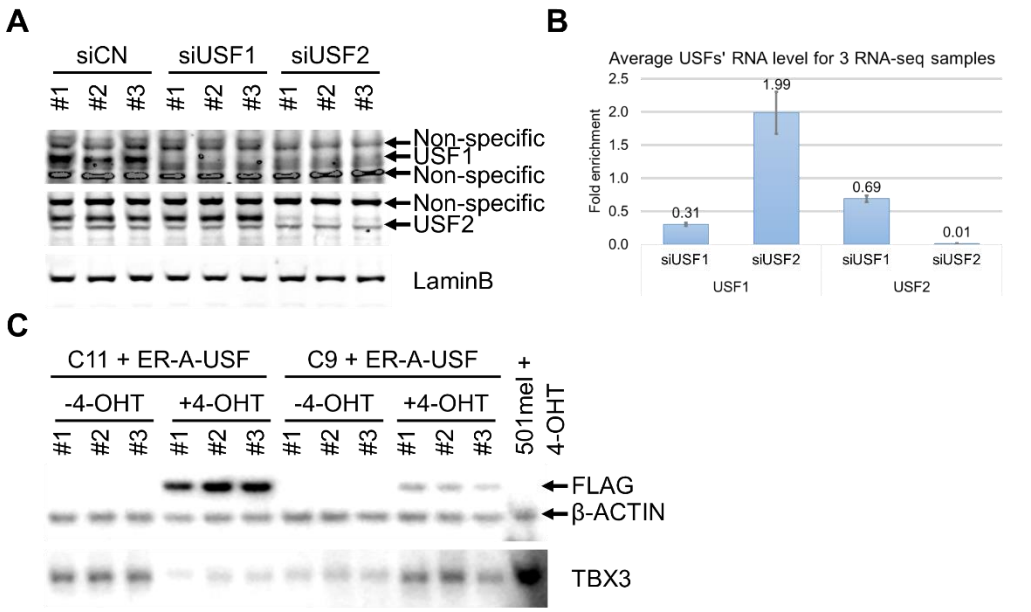


Figure 3.2.1 Validation of siRNA and ER-A-USF samples for RNA-seq. **A)** Western-blot of extract from 501mel cells transfected with the indicated siRNA and samples collected after 72 h siRNA treatment. LaminB was used as a loading control. USF1 and USF2 bands as well as non-specific bands are indicated by arrows. **B)** Gene expression analysis (RT-qPCR) of the same RNA-seq samples as in **A**. The graph is presented as fold enrichment relative to scramble siRNA (siCN), normalised using $\Delta\Delta C_t$ approach as described in Materials and Methods. $\mathbf{N = 3}$ biological replicates each with $\mathbf{2}$ systematic replicates. **C)** Western-blot of extract from two-clones of 501mel (C11 and C9) stably expressing $\mathbf{4}$ –hydroxy tamoxifen ($\mathbf{4-OHT}$) inducible dominant negative USF (A-USF) before and after $\mathbf{72\ h}$ induction with $\mathbf{4-OHT}$ ($\mathbf{600\ ng/ml}$). FLAG bands indicate FLAG-ER-A-USF. β-ACTIN was used as a loading control. TBX3 was used as controls for functionality of ER-A-USF. TBX3 was a novel USF target we identified.

We performed RNA-seq on parental 501mel, 501mel treated under $1\%\ O_2, 16\ h$ for hypoxic treatment, 501mel stably expressing epitope-tagged MYC, USF1, USF2 (hereon referred to as MYC, USF1 and USF2) and 4-hydroxy-tamoxifen ($4-$OHT) inducible dominant-negative USF (hereon referred to as ER-A-USF), and 501mel siRNA knock-down (siUSF1 and siUSF2) or scrambled-siRNA (siCN). Dataset from hypoxic treatment was used to annotate the HIF ChIP-seq datasets. MYC, USF1 and USF2 datasets were used to annotate MYC, USF1 and USF2 ChIP-seq datasets. As these cell lines, especially MYC and USF1 express greatly elevated levels of bHLH-LZ-TFs of interest, their respective RNA-seq datasets can be used to investigate differential gene expression at each gene in response to elevated level of MYC and USF. siRNA and ER-A-USF samples were prepared for the opposite study, the effect of interfering with USF binding on gene expression.

For siRNA samples, we verified our samples prior to submission by Western-blotting with USF1 and USF2 antibody, using LaminB as a loading control (Figure 3.2.1A) and reverse-transcribed qPCR (RT-qPCR) (Figure 3.2.1B). At the protein level, all three-replicates of siUSF1 samples demonstrated a clean knock-down of USF1-specific bands while maintaining USF2-protein level when compared to the siCN. siUSF2 samples also show a highly efficient knock-down of the USF2-specific band. However, in these samples, we also saw a knock-down of USF1 equally efficient as in the siUSF1 samples, perhaps indicating that USF2 regulates USF1 expression. There are reports of asymmetric cross-regulation between USF that is system specific. For example, studies in mouse liver revealed no functional loss nor any cross-regulation between USF1 and USF2 (Vallet et al., 1997) but asymmetric regulation between the USFs was observed in foetal fibroblasts where USF2 knock-out resulted in reduced USF1 protein, and USF1 knock-out led to increased USF2 protein (Sirito et al., 1998). In the 501mel cell line, USF2 positively regulates USF1, but not vice-versa at the protein level, but not at the transcript level. siUSF1 gave $\approx 70\%$ knock-down of USF1-transcript (Figure 3.2.1B) but appears to enhance USF2 transcript level by $\approx 2-$fold, potentially reflecting a compensatory mechanism. siUSF2, on the other hand,

gave ≈ 99% knock-down of USF2-transcript while moderately effecting USF1 transcript (≈ 30% knock-down).

It is important to bear in mind that there is no way of achieving 100% knock-down using siRNA, in fact, our RT-qPCR demonstrates residual USF mRNA. The remaining USF protein will be able to bind to the highest affinity/most accessible sites either alone or cooperatively with (an)other TF(s). These sites conceivably represent targets that contribute to a significant proportion of USF biological function. Ideally we would like to perform RNA-seq on USF1 and USF2 knock-out cell lines. However, we experienced difficulties generating USF knock-out cell lines using both TALEN and Cas9-CRISPR technology (refer to Chapter 3.3 USF knock-out). Therefore, we also carried out RNA-seq experiments using alternative approach to interfere with USF.

As an alternative approach to interfere with USF, we generated 501mel cell lines stably expressing ER-A-USF. A-USF is a version of USF1 that lacks a basic region but retains the dimerization domain. As such, A-USF can dimerize with both USF1 and USF2 and act as a dominant-negative by blocking DNA binding of the endogenous proteins. We 4 −OHT-treated the A-USF-expressing cells for 72 h prior to RNA-extraction (Figure 3.2.1C). We submitted clone−11 of ER-A-USF as this was the one that gave a high induction of ER-A-USF (Figure 3.2.1C, FLAG) and essentially no detectable protein in an un-induced state. To validate the functionality of our ER-A-USF, we also checked for the expression of TBX3, an anti-senescence factor that we demonstrated to be a USF target (data not shown) and which is upregulated by 4 −OHT (Figure 3.3.12E). In protein samples from cells used for RNA-seq analysis, we saw a reduction in TBX3 on Western-blot in clone−11 but not clone−9 following 72 h 4 −OHT induction (Figure 3.2.1C).

Basic RNA-seq analysis

As detailed presentation of our RNA-seq analysis is beyond the scope of this thesis, we will only be discussing the overall pipeline along with some examples.

The example of PCA, given in Figure 3.2.2, were generated using all RNA-seq datasets. All biological replicates demonstrated good clustering, indicating a high reproducibility with the exception of one hypoxia and one 501mel sample (Figure 3.2.2A). In good agreement with the clustering done for the ChIP-seq peaks that demonstrated HIF targets to be most different from the rest (Figure 3.1.20), the 2 −components PCA revealed hypoxia samples as being the most different from other datasets (Figure 3.2.2A).

USF1 and USF2 clustered very closely with all 3- principal components (PCs) considered, most likely reflecting the highly overlapping target genes regulated. USF differs from MYC when considering a third PC indicating many overlapping but also unique targets (Figure 3.2.2B).

siRNA as well as ER-A-USF + 4 −OHT samples clustered quite well together as would be expected. However, they are also relatively similar to 501mel (parental, untreated) and ER-A-USF − 4 −OHT, potentially indicating that interfering with USF, at least within the time scale of 72 h does not affect gene expression profiles as much as hypoxic treatment. It is possible that, at least within 72 h other bHLH-TFs such as MITF might be able to partially compensate for the loss of USF functionality. Alternatively, the residual functional USF may

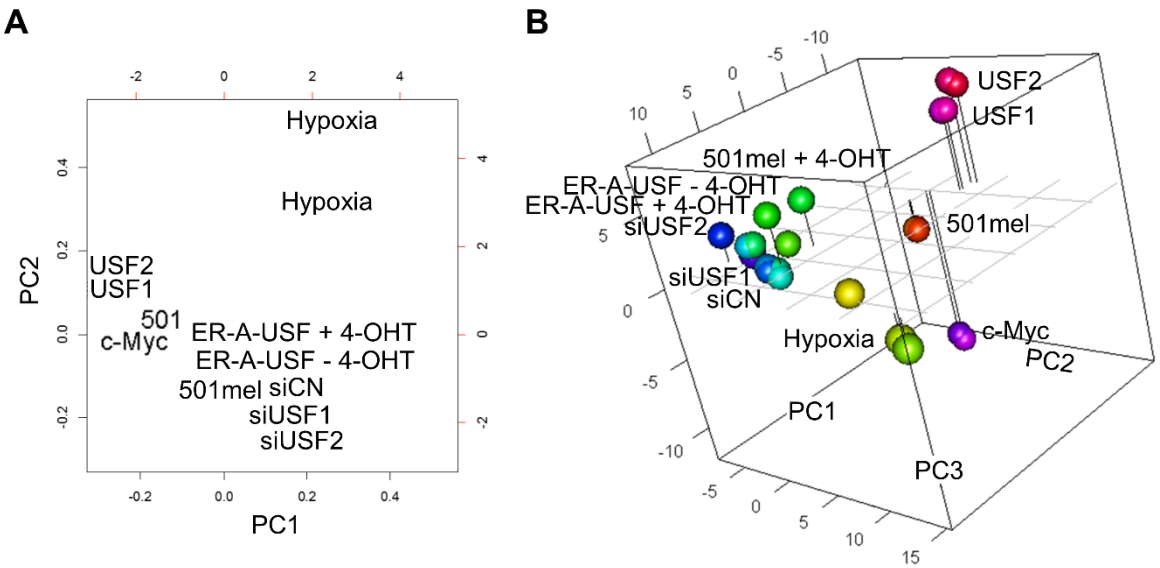


Figure 3.2.2 Principal component (PC)-analysis (PCA) of all RNA-seq datasets. **A)** 2 −D PCA for 501mel (501mel and 501), 501mel subjected to $1\%\ O_2$, 16 h hypoxia, 501mel cell lines stably expressing USF1, USF2, MYC (c-Myc) and 4 −hydroxy Tamoxifen inducible dominant negative USF (ER-A-USF), 501mel siRNA treated for USF1 and USF2. **B)** 3 −D PCA of the same dataset as A).

be sufficient to mediate gene expression from major USF targets that mediate the majority of USF's biological functions. In any case, siUSF samples and the 501mel cell lines stably expressing USF are roughly mirrored along the diagonal axis of the first two PCs (Figure 3.2.2A) indicating that they are rather different along the first two PCs.

We also performed pairwise correlation analysis as well as hierarchical clustering analogous to those performed for ChIP-seq dataset, as these data confirm and essentially give the same conclusion as the PCA analysis, we shall not report them.

Differential gene expression analysis

EdgeR was used to generate a pair-wise differential gene expression matrix for correlation analysis between RNA-seq samples. The differential gene expression output is also used to annotate ChIP-seq peaks. Representatives of such analyses can be seen in Figure 3.2.3. The differential gene expression profile of 501mel cell line stably expressing HA-USF1 (relative to 501mel) and siUSF1 (relative to siCN) show a high degree of discordance (Figure 3.2.3A). In a simple situation where genes which are either up-regulated or repressed in the presence of USF1 and repressed or activate in its absence, we would expect to see an inverse correlation. That is, a straight line with a gradient of -1 that intersects the co-ordinate $(0,0)$. The star-shape pattern of the scatter-plot is typical of uncorrelated samples. That is, up-regulated genes in one sample corresponded to either no change or down-regulation in another sample and vice versa. The observed pattern can be interpreted as mutual exclusivity. That is, genes which demonstrated a high difference in the sample dataset when compared to the control, display little difference in the second sample-control pair. This is unlikely to be due to low sequence coverage giving rise to a statistically-insignificant significance-difference as we filtered out genes with low read count. Another, more plausible explanation is that genes which are highly affected by USF1 overexpression (relative to the 501mel-control) may give rise to relatively little difference in the siUSF1 sample (relative to siCN) as genes whose expression are highly dependent on USF1 are not necessarily repressed in the absence of USF1. More probably, however, is

that they simply would not be expressed. The converse may be applied to USF1-repressed targets where the absence of USF1 may simply result in little difference in gene expression relative to control unless the activator of that gene is also induced. The same pattern can be seen for the USF2 and siUSF2 pair (Figure 3.2.3B). Essentially the same observation was made for USF1 and USF2 against ER-A-USF (data not shown).

The only datasets that show a good correlation were the siRNA samples for the siUSF1 and siUSF2 pair (Figure 3.2.3C) where genes repressed/activated by USF are activated /repressed by knocking down either USF1 or USF2. As USFs are mainly found as heterodimers in cells and our siUSF sample was performed in the 501mel parental cell lines which do not carry an elevated level of either USF through heterologous promoter, knocking-down one is likely to affect the gene expression of the targets of both. Moreover, we already demonstrated a reduced USF1 protein level following siUSF2 (Figure 3.2.1A). In which case, all genes affected by siUSF1, even if they are USF1-specific targets would also be affected by siUSF2. Interestingly, when the same plot was made using differential gene expression data with a P-value of ≤ 0.05 we get a cross-shape pattern consisting of positive correlation and inverse-correlation (Figure 3.2.3D). The inverse-correlation represents genes that are inversely regulated by USF and USF1 alone (knock-down of USF2 also abolishes USF1 in 501mel), or genes that are better activated/repressed by either USF or USF1. When both USFs are present, they compete for the promoter, or bind as a heterodimer to these targets. However, when USF1 is knocked-down, USF2 can exert its effect on gene expression alone. As the P-value is a function of the read-counts, it is worth pointing out that genes that are exclusively regulated by USF1 and/or USF2 alone would be excluded as these genes will have a very low read-count in the siUSF samples and gave low statistical significance during the analysis. Similarly, low-expressed genes with low read-counts in both the sample and control would be excluded and can be seen as a white line across the x-axis (Figure 3.2.3D). For the data that remain, these must have a relatively high read count in both sample and control and therefore are likely to be the

more highly expressed genes which may be regulated by multiple TFs and thus may not be heavily affected by knocking-down USF. Hence, we lose many of the data-points with high differential gene expression.

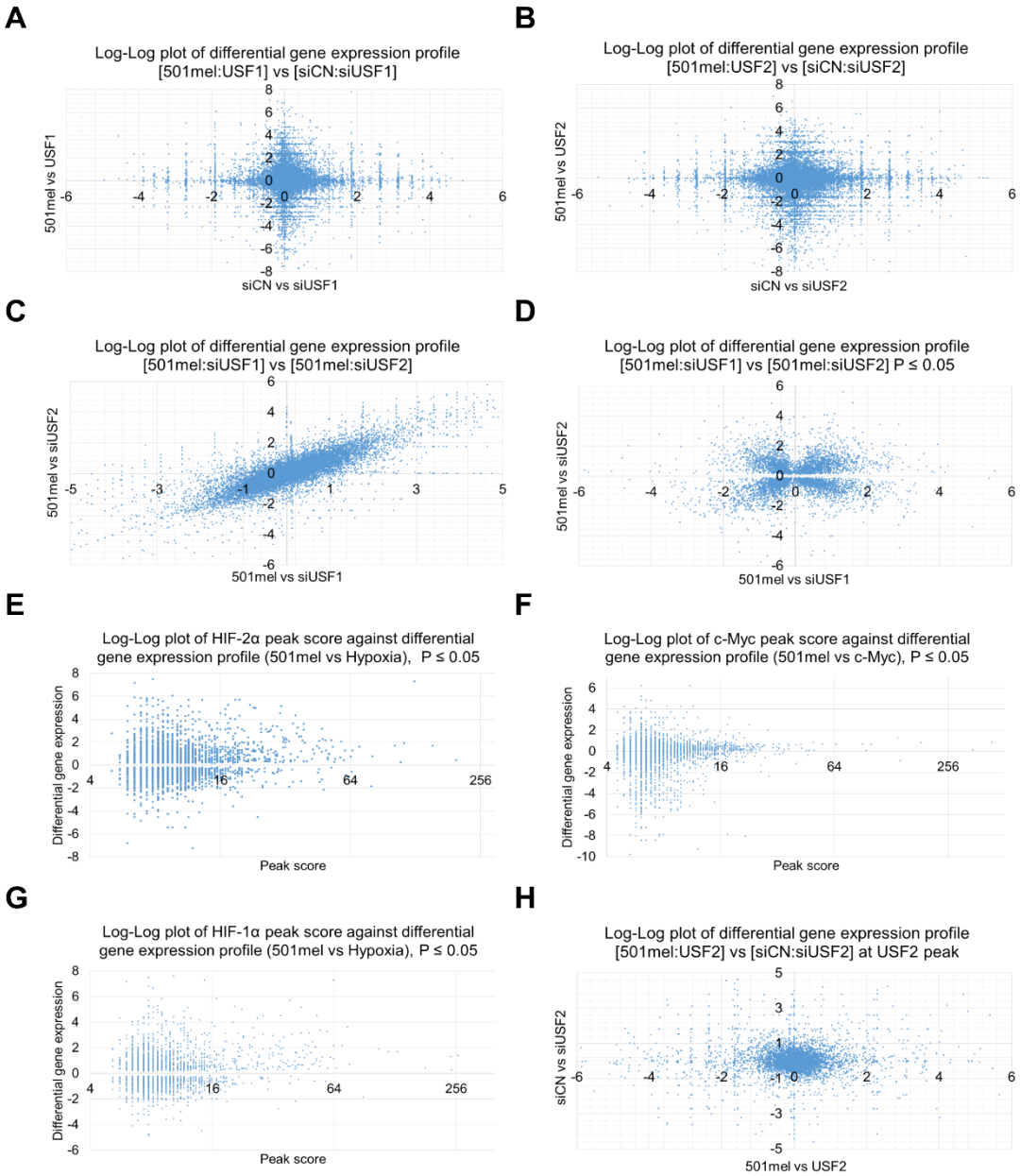


Figure 3.2.3 Example of correlation analysis among RNA-seq datasets and RNA-seq and ChIP-seq datasets. **A)** Log-log plot of differential gene expression profile of 501mel treated with siUSF1 (**50 pM, 72 h**, relative to siCN) against 501mel stably expressing over-expressed HA-USF1 (USF1, relative to 501mel) for the full RNA-seq datasets. **B)** Log-log plot of differential gene expression profile of 501mel treated with siUSF2 (**50 pM, 72 h**, relative to siCN) against 501mel stably expressing HA-USF2 (USF2, relative to 501mel) for the full-RNA-seq datasets. **C)** Log-log plot of differential gene expression profile of 501mel treated with siUSF1 (**50 pM, 72 h**, relative to 501mel) against 501mel treated with siUSF2 (**50 pM, 72 h**, relative to 501mel) for the full-RNA-seq datasets. **D)** Same as **C)** for differential gene expression with P-value of $\leq \mathbf{0.05}$. **E)** Log-log plot of peak score against differential gene expression between 501mel and 501mel subjected to hypoxic treatment (**1**% O_2,**16h**) at HIF-2α peaks for differential gene expression with P-value of $\leq \mathbf{0.05}$. **F)** Log-log plot of peak score against differential gene expression between 501mel and 501mel stably expressing over-expressed MYC at MYC peaks for differential gene expression with P-value of $\leq \mathbf{0.05}$. **G)** Log-log plot of peak score against differential gene expression between 501mel and 501mel subjected to hypoxic treatment (**1**% O_2,**16h**) at HIF-1α peaks for differential gene expression with P-value of $\leq \mathbf{0.05}$. **H)** Log-log plot of differential gene expression of HA-USF2 (relative to 501mel) against siUSF2 (relative to siCN) at USF2 peaks.

Assigning gene expression data to ChIP-seq peaks

Based on the "treadmilling" model (Lickwar et al., 2012), we hypothesised that bHLH-TF binding to peaks with a low peak score, which presumably are bound more transiently by the TF in question, might negatively influence gene expression through competition with other bHLH-TFs which are more capable of activating those genes. Conversely, peaks with a high peak score, which presumably are bound by the TF in question with a longer residence time, might correlate with positive gene expression. To address this hypothesis, we assigned an RNA-seq differential gene expression matrix to appropriate ChIP-seq peaks (exemplified in Figure 3.2.3E – G). From our analysis, there is no relationship between peak score and whether the genes are activated or repressed, although the number of repressed and activated genes against peak score are somewhat different for each bHLH-TF studied. For example, more of MYC's peaks are associated with down-regulated genes (Figure 3.2.3F). HIF-1α's peaks, on the other hand, appear to have a slight bias for up-regulated genes (exemplified in Figure 3.2.3G).

The lack of correlation does not mean there is no correlation between residency time and gene expression, as peak score does not distinguish between sites with one or multiple E-boxes. In fact most of the genes that show the largest differential gene expression (both up- and down-regulated are concentrated towards lower peak score. One interpretation of this is that these actually represent sites with a single E-box of high affinity, whereas higher peak scores, that are associated with relatively small differential gene expression, are sites with multiple-E-boxes of low affinity. Base on the "TF binding kinetic and gene activation" hypothesis (section 1.2), binding to E-boxes with a dwell time below the threshold to activate the gene will not result in gene activation. It is also possible that the context in which the E-boxes are situated and other factors co-recruited may play a more prominent role. The location to which the bHLH-TFs are bound relative to the TSS may also dictate the mode which the bHLH-TF in question regulates the gene. Alternatively, all the caveats associated with translating peak-score to binding affinity may obscure any correlation that may exist.

Lastly we analysed differential gene expression profiles of two RNA-seq datasets for each peak file as exemplified in Figure 3.2.3H where we looked at differential gene expression of 501mel cells stably expressing USF2 (HA-USF2, relative to parental 501mel) against 501mel cells treated with siUSF2 (relative to siCN), selecting only RNA-seq data that has corresponding USF2 peaks. Unlike for the whole dataset where we saw some mutual exclusivity between the two datasets (Figure 3.2.3B) the most extremely discordant are lost when considering only genes where USF2 binds, indicating that there may be many indirect effects of USF2 on global gene expression. [A similar pattern was observed when comparing the same differential gene expression data against combined USF1 + USF2 peaks, (data not shown). This was taken into consideration as siUSF2 also affects USF1 protein level.] For the dataset with corresponding USF2 peaks, there is a relatively higher ratio of genes which show an inverse correlation between siUSF2 and HA-USF2 ([+, −] and [−, +] quadrant) as would be expected of samples expressing USF2 or USF2 knock-down. The remainder of the genes show a positive correlation between USF2 and siUSF2, and may represent genes that are weakly regulated by USF2. As bHLH-TFs have overlapping target specificity, a given TFBS-motif may be bound and activated by USF2 in competition with other bHLH-TFs. In a situation where USF2 is not a strong activator or repressor of the gene, increasing USF2 levels would only result in a modest activation/repression of the gene. Conversely, knocking down USF2 would also only result in a modest repression/activation of the gene. In both cases, the presence or absence of USF2 would only affect competition with other bHLH-TFs.

For USF2 bound genes that are up- or down-regulated in the HA-USF2 dataset, but exhibit no change in the siUSF2 dataset, we may argue that this is due to incomplete knock-down of USF2. Residual amount of USF2 may be able to bind to and activate genes, especially those with high affinity sites that may represent major USF2 targets.

GO-analysis of RNA-seq datasets

GO-analysis of ChIP-seq datasets demonstrates enrichment of some related GO-terms. Therefore, we sought to identify enrichment of some related gene-ontologies in each RNA-seq dataset. To this end, we used our differential gene-expression data and selected for genes whose log-fold changes have the P-value of < 0.05 and separated these into up-regulated ($FC > 1.1$) and down-regulated ($FC < 0.9$) datasets for gene ontology analysis.

Most RNA-seq datasets have an approximately equal distribution of genes up-regulated and down-regulated (Figure 3.2.4A). Only the ER-A-USF± 4 $-$OHT pair and the siCN vs siUSF1 pair display $> 10\ \%$ difference between number of genes up- and down-regulated (15 & $20\ \%$, respectively).

To eliminate indirect effects associated with the treatment used to prepare some of the RNA-seq datasets, we also looked at the distribution of differential gene expression which has corresponding peaks (Figure 3.2.4B). Under hypoxia, there is a slight enrichment (average $\approx 10\%$) of up-regulated genes associated with HIF-peaks. There are $\approx 5\%$ more up- than down-regulated genes in HA-USF1, HA-USF2 and FLAG-MYC cell lines with corresponding USF1, USF2 and MYC peaks, respectively. Conversely, there are $\approx 5\%$ more down- than up-regulated genes in siUSF2 samples with corresponding USF1 and USF2 peaks. Interestingly, there are $\approx 18\%$ more up- than down-regulated genes in siUSF1 samples with corresponding USF1 and USF2 peaks. This is similar to what was seen when considering the whole RNA-seq dataset with P-value of < 0.05.

ER-A-USF± 4 $-$OHT pair, on the other hand, display no major differences between up- and down-regulated genes even though an $\approx 15\%$ difference was seen when considering the whole RNA-seq dataset with P-value of < 0.05. This indicates that 4 $-$OHT treatment has a greater influence on gene expression than the induction of dominant negative USF.

We performed a number of pairwise analyses. To simplify the presentation of data, we grouped together terms which belong to related processes and presented them here as

tables with the number of GO-terms out of the top 50 GO-terms under the indicated processes. We first looked at hypoxia and the effects of heterologous expression of MYC and USF on gene expression relative to the 501mel parental cell line (Table 3.2.1).

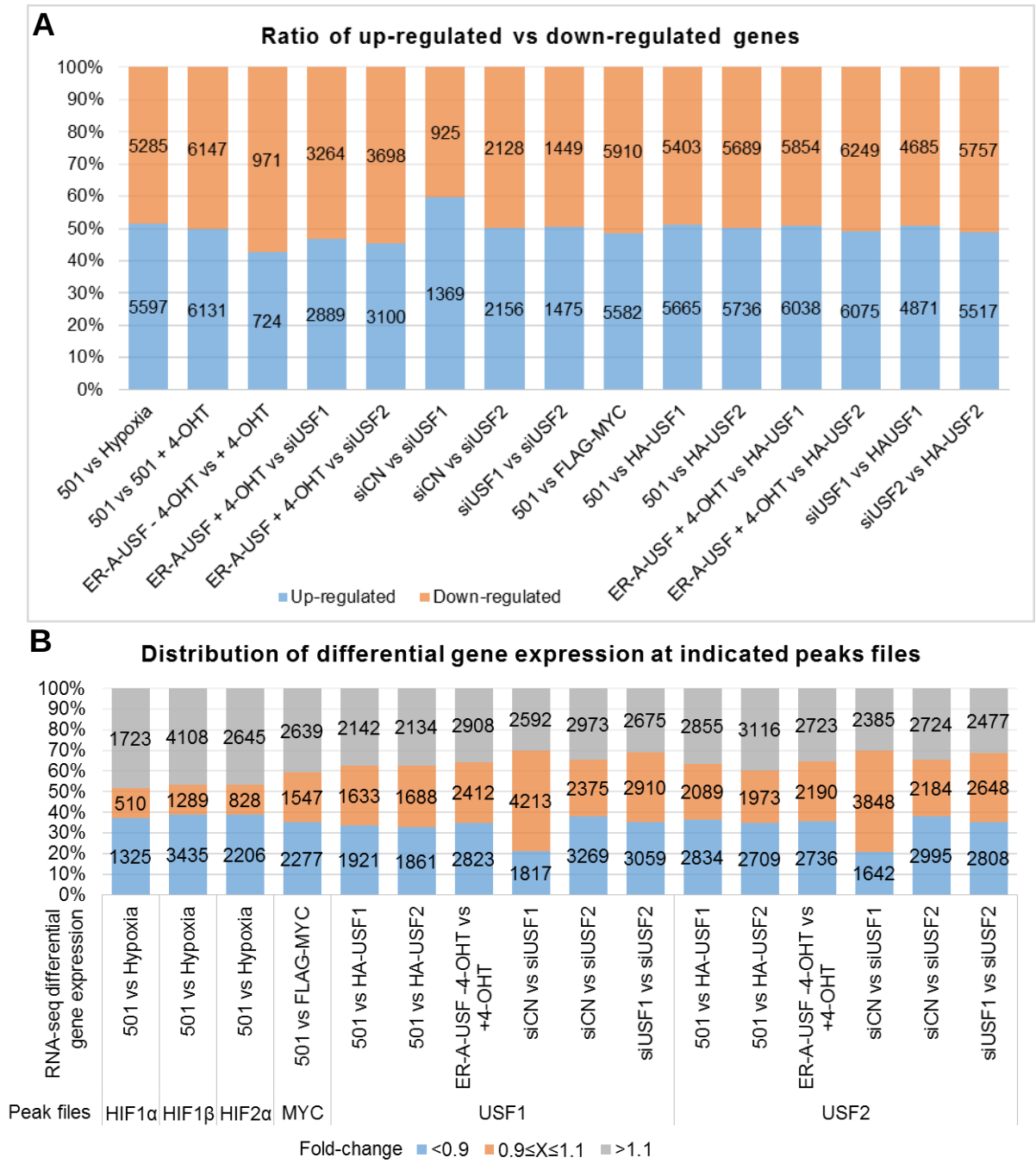


Figure 3.2.4 Stack-bar graphs showing up-regulated and down-regulated genes in the pairwise-comparison indicated. **A)** Differential gene-expression data from RNA-seq calculated using EdgeR was filtered for genes whose log-fold changes have the P-value of $< \mathbf{0.05}$ and separated into up-regulated and down-regulated dataset and plotted as stack-bars. **B)** Differential gene-expression data from RNA-seq calculated using EdgeR was fitted to the indicated ChIP-seq peak files before stratified in to genes with fold change of $< \mathbf{0.9}$, $\mathbf{0.9} \leq \boldsymbol{x} \leq \mathbf{1.1}$ or $\geq \mathbf{1.1}$ and plotted as stack-bars. The numbers on the bar indicated the number of genes in each dataset. Labelling are as followed: 501, parental 501mel cells. $\mathbf{4-OHT}$, $\mathbf{4}-$Hydroxy-tamoxifen treated; ER-A-USF, 501mel cells stably expressing $\mathbf{4-OHT}$ inducible dominant negative USF; siUSF1, 501mel treated with siUSF1 at $\mathbf{50\ pM, 72\ h}$; siUSF2, 501mel treated with siUSF2 at $\mathbf{50\ pM, 72\ h}$; siCN, 501mel treated with scramble siRNA at $\mathbf{50\ pM, 72\ h}$; FLAG-MYC, 501mel cells stably expressing FLAG-MYC; HA-USF1, 501mel cells stably expressing HA-USF1; and HA-USF2, 501mel cells stably expressing HA-USF2.

Table 3.2.1 Clustered GO-terms derived from differential gene expression analysis of 501mel treated with $\mathbf{16\ h, 0.1\%}$ O_2 or 501mel cell lines stably expressing either FLAG-MYC, HA-USF1 or HA-USF2 relative to 501mel parental cells. Green box in the first column indicated up-regulated genes while red box indicated down-regulated genes. Differential expression analysis is always second dataset relative to the first. The number of terms on the right indicates the number of GO-terms under the indicated process out of the top **50** GO-terms looked at.

Samples	Process	No. of terms
501mel vs 16 h 0.1% O_2	Development and differentiation	17
	Metabolic process	11
501mel vs 16 h 0.1% O_2	Cell cycle related process	17
	Metabolic process	17
	Biosynthetic/Biogenesis process	5
501mel vs FLAG-MYC	Metabolic process	21
	Biosynthetic/Biogenesis process	14
501mel vs FLAG-MYC	Metabolic process	12
	Negative regulation of metabolic process	4
	Cell cycle related process	6
	Biosynthetic/Biogenesis process	6
501mel vs HA-USF1	Development and differentiation	18
	Metabolic process	6
501mel vs HA-USF1	Metabolic process	19
	Cell cycle related process	9
	Biosynthetic/Biogenesis process	9
501mel vs HA-USF2	Metabolic process	26
	Cell cycle related process	8
501mel vs HA-USF2	Biosynthetic/Biogenesis process	32
	Metabolic process	6

Exposure to hypoxia resulted in up-regulation of development and differentiation processes ($17/50$) and down-regulation of cell-cycle-related processes ($17/50$). The former, presumably is a reflection of the inhibitory effect of hypoxia on proliferation as expression of differentiation genes are associated with exit from the cell-cycle. A number of metabolism-related terms are found up- and down-regulated presumably reflecting shifts in the metabolic processes from aerobic to anaerobic energy production.

The FLAG-MYC cell line that has elevated levels of MYC, displays up-regulation of many metabolic- and biosynthetic-processes ($21/50$ & $14/50$), presumably reflecting the effect of MYC in driving cell proliferation. This is supported by the observed enrichment of down-regulated terms relating to negative regulation of metabolic processes and cell-cycle-related processes ($4/50$ & $6/50$).

The HA-USF1 cell line displays increased expression of genes related to the development and differentiation terms ($18/50$), and decreased expression from cell-cycle-related processes ($9/50$). This is in good agreement with the reported role of USF in differentiation (see section 1.8, USF and development) and the reported ability of USF1 to inhibit growth induced by MYC+ RAS (see section 1.8, USF and cancer). Somewhat puzzling is the higher number of metabolic processes being supressed in the presence of HA-USF1, as USF1 was reported to activate many metabolic genes (refer to relevant reference in Table 3.4.1). The HA-USF2 cell line, on the other hand, demonstrates up-regulation of a large number of metabolic processes as would be expected since USF is involved in regulating metabolic genes. However, contrary to the HA-USF1 cell line, in the presence of HA-USF2 a number of cell-cycle-related processes are up-regulated, while many biosynthetic processes are down-regulated ($8/50$ & $32/50$). The enrichment in biosynthetic and biogenesis process down-regulated is in agreement with the reported ability of USF2 to more efficiently inhibit proliferation (see section 1.8, USF and cancer).

Table 3.2.2 Clustered GO-terms derived from differential gene expression analysis for 501mel treated with 72 h, 50 pM scrambled siRNA (siCN) or siUSF1 or siUSF2. Green box in the first column indicates up-regulated genes while the red box indicates down-regulated genes. Differential expression analysis is always the second dataset relative to the first. The number of terms on the right indicates the number of GO-terms under the indicated process out of the top 50 GO-terms looked at. FC, fold-change

Samples	Process	No. of terms
siCN vs siUSF1	Metabolic process	11
	Development and differentiation	9
	Biosynthetic/Biogenesis process	2
siCN vs siUSF1	Cell cycle related process	32
	Biosynthetic/Biogenesis process	2
siCN vs siUSF2	Metabolic process	12
	Development and differentiation	10
	Biosynthetic/Biogenesis process	6
	Growth Factor signalling	4
siCN vs siUSF2	Cell cycle related process	38
siUSF1 vs siUSF2	Development and differentiation	12
	Metabolic process	12
	Biosynthetic/Biogenesis process	2
siUSF1 vs siUSF2	Metabolic process	12
	Development and differentiation	10
	Biosynthetic/Biogenesis process	6
	Growth Factor signalling	4
siUSF1 vs siUSF2 (FC $0.8 < x < 1.2$)	Metabolic process	23
	Biosynthetic/Biogenesis process	13

We proceeded to analyse the effect of siUSF on gene expression (Table 3.2.2). Knocking down USF1 has the effect of increasing gene expression related to metabolic- and development-and-differentiation-processes ($11/50$ & $9/50$) and a large decrease in cell-cycle-related processes ($32/50$). A similar result was observed for the siUSF2 samples, with increased metabolic- and development-and-differentiation-processes ($12/50$ & $10/50$) and a large decrease in cell-cycle-related processes ($38/50$). The similarity may be explained by the fact that siUSF2 also knocks-down USF1 in 501mel cells. Therefore, any effect seen in the siUSF1 samples will also be translated to the siUSF2 samples. To address this we performed differential gene expression analysis between siUSF1 and siUSF2. In this case, we observed enrichment of metabolic-, development-and-differentiation- and biosynthetic-processes among both up- and down-regulated genes. This indicated that either knocking down USF1 and USF2 has an additive effect on USF regulated genes, or that there are genes differentially regulated by either USF1 or USF2 alone. In any case, we are able to identify many metabolic- and development-and-differentiation-processes ($23/50$ & $13/50$) among terms which siUSF1 and siUSF2 have similar effects on their gene expression (defined as genes whose fold-change is between $0.8 - 1.2$ among the two samples). These include genes which are similarly affected by USF1 and USF1 + USF2 knock down (hence, no significant difference between siUSF1 and siUSF2 samples) or non-USF regulated genes whose expression is not significantly affected by siUSF. As genes implicated in both metabolic- and biosynthetic/biogenesis are among those regulated by USF, their enrichment among genes similarly affected by siUSF1 and siUSF2 most likely reflect genes similarly regulated by USF1 and USF2.

The ER-A-USF cell line was generated as an alternative, complementary approach to interfere with USF. However, we also noted the potential problem with the use of $4-$OHT that also regulates some endogenous genes other than the ER-A-USF. Presented here is a pairwise comparison of gene expression in the ER-A-USF cell line against siUSF (Table 3.2.3). 501mel cells treated with $4-$OHT display up-regulation of cell-cycle related

Table 3.2.3 Clustered GO-terms derived from differential gene expression analysis for 501mel cells treated with: $72\,\mathrm{h}\ 600\mathrm{ng/ml}, 4-$hydroxy-tamoxifen $(4-\mathrm{OHT})$ or $72\,\mathrm{h}, 50\ \mathrm{pM}$ scramble siRNA (siCN) or siUSF1 or siUSF2. The green box in the first column indicates up-regulated genes while the red box indicates down-regulated genes. Differential expression analysis is always the second dataset relative to the first. The number of terms on the right indicates the number of GO-terms under the indicated process out of the top 50 GO-terms looked at. FC, fold-change

Samples	Process	No. of terms
501mel $-$ vs $+4-$OHT	Metabolic process	29
	Cell cycle related process	7
501mel $-$ vs $+4-$OHT	Metabolic process	21
	Biosynthetic/Biogenesis process	14
ER-A-USF $-4-$OHT vs $+4-$OHT	Cell cycle related process	18
	Development and differentiation	3
	Metabolic process	2
ER-A-USF $-4-$OHT vs $+4-$OHT	Development and differentiation	37
ER-A-USF $+4-$OHT vs siUSF1	Biosynthetic/Biogenesis process	18
	Metabolic process	15
	Mitochondrial/Energy related process	14
ER-A-USF $+4-$OHT vs siUSF1	Development and differentiation	24
	Cell cycle related process	3
ER-A-USF $+4-$OHT vs siUSF1 (FC $0.5 < x < 1.5$)	Metabolic process	21
	Biosynthetic/Biogenesis process	11
	Cell cycle related process	4
ER-A-USF $+4-$OHT vs siUSF2	Mitochondrial/Energy related process	14
	Biosynthetic/Biogenesis process	13
	Metabolic process	9
ER-A-USF $+4-$OHT vs siUSF2	Cell cycle related process	11
	Development and differentiation	10
	Metabolic process	3
	Biosynthetic/Biogenesis process	3
ER-A-USF $+4-$OHT vs siUSF2 (FC $0.5 < x < 1.5$)	Metabolic process	23
	Biosynthetic/Biogenesis process	14
	Cell cycle related process	2

processes $(7/50)$. Terms related to metabolic processes are found enriched among both up-regulated and down-regulated genes $(29/50\ \&\ 21/50)$, while biosynthetic terms are enriched among the down-regulated genes $(14/50)$. ER-A-USF cell line treated with $4-$OHT display an enrichment in cell-cycle-related process among up-regulated genes $(18/50)$ and enrichment in development and differentiation process among down-regulate genes $(37/50)$. As proliferative and differentiation genes are often inversely expressed, this result indicates that the effect of $4-$OHT on activation of proliferative genes (most likely through ER) and repression of differentiation genes is confounding the effect of A-USF. Nevertheless, we are able to pick out a number of processes affected by siUSF as well as

ER-A-USF (by considering genes whose fold-change between ER-A-USF and siUSF lie between $0.5 - 1.5$). Interestingly, a large number of processes related to biosynthesis, metabolism and mitochondria-related processes are up-regulated in the siUSF samples relative to ER-A-USF $+ 4$ −OHT. This indicates that ER-A-USF $+ 4$ −OHT exerts a greater effect in repressing these genes related to these terms when compared to siUSF. The converse, where some genes are more down-regulated in siUSF samples relative to ER-A-USF $+ 4$ −OHT, also exists. These mainly include cell-cycle-related, development and differentiation processes. As A-USF interferes with both USF1 and USF2 functions, this may explain some of the differences observed when comparing ER-A-USF $+ 4$ −OHT and siUSF1. Since, siUSF2 affects the protein level of both USF1 and USF2 we should expect to see a similar effect on gene expression. As this is clearly not the case, we must be cautious when using the ER-A-USF cell line as a tool to study factor exchange. Although siRNA may also exert some confounding effects, especially though eliciting a stress response, we did not see stress-response terms among the genes up-regulated by siRNA. Therefore, the confounding effect of $+ 4$ −OHT is likely to be more significant.

Lastly, we performed pair-wise differential gene expression analysis on the HA-USF1 and HA-USF2 stable cell lines against ER-A-USF $+ 4$ −OHT and siUSF (Table 3.2.4). For ER-A-USF $+ 4$ −OHT vs HA-USF1, we could see enrichment of cell-cycle-related processes ($8/50$), metabolic processes ($20/50$) and mitochondria-related processes ($5/50$). Most of the genes repressed in HA-USF1 relative to ER-A-USF $+ 4$ −OHT belong to metabolic- and biosynthetic-processes ($22/50$ & $17/50$). For HA-USF2 relative to ER-A-USF $+ 4$ −OHT cell-cycle-related- and metabolic-processes are enriched among the up-regulated genes ($17/50$ & $9/50$), while biosynthetic, metabolic and development-and-differentiation processes are enriched among down-regulated genes ($15/50$ & $9/50$ & $8/50$).

Cell-cycle- and metabolic- related-processes are enriched among the up-regulated genes ($24/50$ & $9/50$) in HA-USF1 relative to siUSF1 while metabolic- and biosynthetic-processes ($20/50$ & $17/50$) are among genes repressed. The repressed processes are similar to HA-

USF1 relative to ER-A-USF + 4 −OHT. The result for HA-USF2 relative to siUSF2 is somewhat opposite to what was observed relative to ER-A-USF + 4 − OHT with biosynthetic-, metabolic and development and differentiation-processes among genes up-regulated (16/50 & 11/50 & 6/50) and enriched cell-cycle-related- and metabolic processes among genes down-regulated (21/50 & 14/50).

As already pointed out, when comparing ER-A-USF + 4 −OHT and siUSF, the different confounding effects of the two approaches used to interfere with USF function results in some discordance when directly comparing the effect of ER-A-USF + 4 −OHT and siUSF against HA-USF1 and HA-USF2 stable cell lines. Nevertheless, there are a number of GO-terms which are similarly enriched among up- and down-regulated genes with both the ER-A-USF + 4 −OHT and siUSF approaches.

Table 3.2.4 Clustered GO-terms derived from differential gene expression analysis for 501mel treated with: 72 h 600 ng/ml, 4 −hydroxy-tamoxifen (4 − OHT) or 72 h, 50 pM scramble siRNA (siCN) or siUSF1 or siUSF2 and or 501mel cell lines stably expressing either HA-USF1 or HA-USF2. The green box in the first column indicates up-regulated genes while the red box indicates down-regulated genes. Differential expression analysis is always second dataset relative to the first. The number of terms on the right indicates the number of GO-terms under the indicated process out of the top 50 −GO terms looked at. FC, fold-change

Samples	Process	No. of terms
ER-A-USF + 4 −OHT vs USF1	Metabolic process	20
	Cell cycle related process	8
	Mitochondrial/Energy related process	5
ER-A-USF + 4 −OHT vs USF1	Metabolic process	22
	Biosynthetic/Biogenesis process	17
ER-A-USF + 4 −OHT vs USF2	Metabolic process	17
	Cell cycle related process	9
	Biosynthetic/Biogenesis process	3
ER-A-USF + 4 −OHT vs USF2	Biosynthetic/Biogenesis process	15
	Metabolic process	9
	Development and differentiation	8
siUSF1 vs USF1	Cell cycle related process	24
	Metabolic process	9
	Biosynthetic/Biogenesis process	3
siUSF1 vs USF1	Metabolic process	20
	Biosynthetic/Biogenesis process	17
siUSF2 vs USF2	Biosynthetic/Biogenesis process	16
	Metabolic process	11
	Development and differentiation	6
siUSF2 vs USF2	Cell cycle related process	21
	Metabolic process	14

Correlating fold-changes obtained from RNA-seq and RT-qPCR

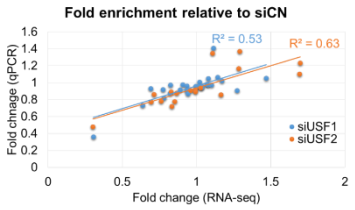
As RT-qPCR is generally accepted as the gold-standard for differential gene expression analysis (because each sample is internally controlled) while normalisation of RNA-seq still remains a challenge in differential gene expression analysis, we carried out RT-qPCR on 28 genes to validate our RNA-seq data.

Of the 28 –genes analysed, six-genes show no overlap in fold-change between RT-qPCR and RNA-seq data for USF1 and nine-genes for USF2 even when considering the 95% confidence interval of ChIP-qPCR data (Figure 3.2.5A). For the remaining genes the fold-change as determined from the RNA-seq were plotted against fold-change as determined from RT-qPCR, and R^2-value of 0.53 and 0.63 were obtained for siUSF1 and siUSF2 samples, respectively (Figure 3.2.5B). In general, we can say that the two approaches gave a similar trend in the differential gene expression analysis, although the absolute values differ.

A

	siUSF1		siUSF2	
	qPCR	RNA-seq	qPCR	RNA-seq
ANF	1.47	1.05		
E-Cad	0.77	0.92		
LAMP1	1.69	1.18	1.64	1.07
RNAPII			1.70	1.23
p21CIP1			1.15	1.62
p53			2.36	1.38
p66α	1.27	0.91	2.73	0.89
PGC1α			0.30	0.48
POMC	1.43	0.15	2.07	0.15
USF1			1.78	0.87
USF2	0.69	0.93	0.01	0.26

B



C

	siUSF1		siUSF2	
	qPCR	RNA-seq	qPCR	RNA-seq
β-ACTIN	0.99	0.96	0.71	0.86
GAPDH	1.11	1.41	1.11	1.35
HPRT	0.89	0.92	0.83	0.72
RNAPII	1.02	1.01		
ANF			1.69	1.10
Bcl2	0.64	0.73	0.76	0.78
DCT	1.10	0.97	1.53	0.92
Dia1	0.78	0.82	0.96	0.91
E-CAD			1.29	1.37
HIF1α	0.93	0.91	0.85	0.78
MC1R	1.03	0.94	1.16	0.86
MITF	1.07	0.98	1.03	0.94
P16	1.08	1.05	0.83	0.89
P21	1.01	0.96		
P38α	0.94	0.96	0.99	0.92
P53	1.14	1.07		
PGC1α	0.82	0.88		
SILVER	0.91	0.98	0.99	0.89
TBX2	1.08	0.97	1.28	1.17
TBX3	1.17	1.02	0.69	0.78
TERT	0.83	0.98	0.50	0.61
TFE3	0.94	0.87	1.27	1.04
TYR	1.03	0.96	0.87	0.88
USF1	0.31	0.36		

Figure 3.2.5 Correlation analysis of RNA-seq and RT-qPCR. **A)** A table showing some of the tested genes whose fold-change in siUSF1 or siUSF2 samples relative to siCN samples by RT-qPCR (qPCR, using $\Delta\Delta C_t$ method as described in Materials and Methods) is significantly different from fold-change as determined from RNA-seq (analysed using EdgeR). The value from qPCR and RNA-seq is considered to be the same if the two values overlap when considering 95% confidence interval of the qPCR data. Red highlight indicates very different fold change obtained using the two RNA-quantification methods. **B)** Scatter plot of fold-change as determined from RNA-seq against fold-change as determined from RT-qPCR. 28 –genes (listed in **C**) were used for comparison and genes highlighted in red in **A)** are omitted from the plot. The R^2 value for siUSF1 is shown in blue while the R^2 value for siUSF2 is shown in orange. For RNA-seq $\mathrm{N} = 2$ biological replicates, for RT-qPCR $\mathrm{N} = 3$ biological replicates each with 2 systematic replicates. **C)** A table showing fold-change in gene expression of the indicated genes in siUSF1 or siUSF2 samples relative to siCN samples as determined by RT-qPCR (qPCR, using $\Delta\Delta C_t$ method as described in Materials and Methods) and RNA-seq (analysed using EdgeR). These data were used to plot the scatter graph in **B**.

Brief summary of Chapter 3.2

We have performed and analysed RNA-seq experiments on siUSF, ER-A-USF, hypoxia treated cells, FLAG-MYC, HA-USF and the negative control. Using these data, we performed differential gene expression analysis and assigned them to ChIP-seq peaks. Correlation analysis of the whole RNA-seq datasets indicated that: genes highly affected by overexpression of USF may give rise to relatively little difference in the siUSF or ER-A-USF+ 4 −OHT samples, probably because they are not regulated in the absence of USF.

With the exception of ER-A-USF± 4 −OHT and the siCN vs siUSF1 pair, all RNA-seq dataset have $\leq 10\%$ difference between up- and down-regulated genes. However, we were able to see, on average, $\approx 10\%$ more HIF's target genes upregulated in response to hypoxia and $\approx 5\%$ more genes up-regulated than down-regulated in HA-USF and FLAG-MYC RNA-seq samples with corresponding USF and MYC's peaks. On translating from pair-wise analysis of the whole RNA-seq to genes with corresponding ChIP-seq peaks, we find that there are probably many indirect effects of siRNA and 4 −OHT treatment on gene expression and that they must be interpreted with caution. These discrepancies are unlikely to be due to the simplified gene-expression analysis approach used in this thesis as the fold-changes determined from RNA-seq correlate rather well with fold-enrichment obtained from RT-qPCR.

GO-analysis of up- and down-regulated genes revealed enrichment in terms largely in line with expectation. For example, the FLAG-MYC cell line that has elevated levels of MYC, displays up-regulation of many metabolic- and biosynthetic-processes, reflective of MYC's ability to drive cell proliferation. Also, the enrichment in biosynthetic and biogenesis process down-regulated in HA-USF2 samples which is in agreement with the reported ability of USF2 to more efficiently inhibit proliferation.

3.3 Factor exchange and promoter education

USF and histone modifications

In the hematopoietic stem/progenitor cell, maintenance of the bivalent marks at haematopoietic genes by USF is necessary for subsequent haematopoietic lineage specification by HoxB4 and TAL1 that are also activated by USF/NF-Y (Zhu et al., 2003). During differentiation, USF1 recruits the hSET1A H3K4 HMT that shifts the bivalent marks toward H3K4me3 resulting in the recruitment of TAF3 and RNAPII and gene expression. USF1:USF2 heterodimer activates hematopoietic expression of the *HoxB4* gene in response to cytokine-mediated self-renewal and expansion of haematopoietic stem cells (Giannola et al., 2000).

USF1 also has a more general role in maintaining chromatin boundaries by binding to insulators and maintaining histone modifications associated with active transcription (West et al., 2004). This is achieved through direct interaction with the PRMT1, P/CAF and SRC-1 (Huang et al., 2007). In hematopoietic cells, USF1 interaction with SET1 and NURF is necessary to maintain chromatin domains (Li et al., 2011).

As USF has no strong transcription-activation domain (see section 1.8, USF and transcription control), it is possible that USF may have a global role in setting up the epigenetic landscape for gene activation. Although USF has also been reported as a strong TA, it is entirely possible that the reported gene activation by USF is a consequence of chromatin modifications that allow the gene to be better activated by other TFs. To address this, we turned to the publicly available data and identified a hepatic-carcinoma (HepG2)-cell line as having the most comprehensive histone ChIP-seq datasets and searched for enrichment of specific histone marks at USF occupied targets. If USF is involved in promoter education globally and not limited to developmentally-linked genes (differentiation genes), we may detect an enrichment of epigenetic marks associated with poised-promoter and bivalent marks even in the differentiated HepG2-cell line.

Table 3.3.1 Histone marks and the associated consequence on gene status.

Histone	What it means?
H3K4me1	Associated with enhancers, and downstream of transcription starts.
H3K9ac	Associated with transcriptional initiation and open chromatin structure.
H3K27ac	Associated with transcriptional initiation and open chromatin structure.
H3K36me3	Marks regions of RNAPII elongation, including coding and non-coding transcripts.
H4K20me1	Associated with active and accessible regions
H3K79me1	Associated with transcription elongation
H3K4me2	Marks promoters and enhancers. Most CpG islands are marked by H3K4me2 in primary cells. May also be associated with poised promoters
H3K4me3	Marks promoters that are active or poised to be activated
H3K79me2	Mark of the transcriptional transition region - the region between the initiation marks (K4me3, etc) and the elongation marks (K36me3).
H3K27me3	Marks promoters that are silenced by Polycomb proteins in a given lineage and large domains found at inactive developmental loci.

In HepG2 USF binding correlates with enrichment of histone marks characteristic of active genes (see Table 3.3.1 for a brief summary of histone marks and the associated consequence), especially H3K4me3, H3K27ac and H3K4me2 in the order of prevalence, and to a lesser extent H3K4me1 and H3K9ac (Figure 3.3.1A & B for USF1 and USF2 respectively). For H3K4me1 and H3K9ac, this might simply reflect the lesser prevalence of poised enhancers in HepG2 cells. H4K20me1 is the only active histone mark in the dataset that is not globally enriched around the USF summit (Figure3.3.1C & D for USF1 and USF2 respectively).

The repressive chromatin mark H3K27me3 is at background level around USF peaks (Figure3.3.1C & D). H3K79me2, a mark of TSS (between promoter and gene body), is found

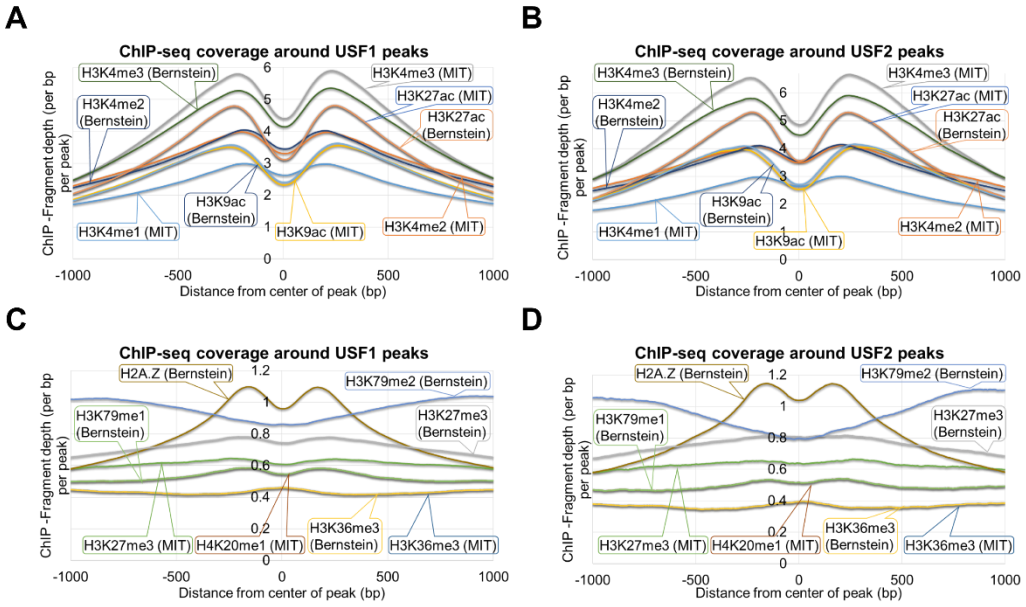


Figure 3.3.1 Histone ChIP-seq at USF targets. Histones mapped file were used to populate read-counts at USF1 (**A** and **C**) and USF2 (**B** and **D**) peak summits with a bin size of 5 bp before plotting as a function of ChIP-fragment depth against distance from peak summit. To aid visualisation, the data were split into two sets of high coverage (**A** and **B**) and low coverage (**B** and **D**).

to be globally depleted at USF1 and USF2 peak summits, indicating that USF factors generally do not bind very close to the TSS. Lastly, in line with the reported role of H3K79me1 and H3K36me3 in elongation but not transcription initiation, H3K79me1 and H3K36me3 are depleted around USF peaks (Figure3.3.1C & D) (Karlić et al., 2010).

H2A.Z, a histone variant that is characteristically positioned around TSSs and was shown to stimulate ISWI chromatin-remodelling activity through the extended acidic patch on H2A.Z (Goldman et al., 2010) necessary for promoter remodelling and H2A.Z exchange prior to RNAPII loading (Hardy et al., 2009) is strongly enriched around USF peaks (Figure3.3.1C & D).

Factor exchange at the human *Tyrosinase* (*TYR*) gene

Although there are many similarities between transcription at *PHO5* and mammalian genes, a key question is the applicability of the "promoter education" model at mammalian genes. To study promoter education in a mammalian context we need an inducible gene that is targeted by multiple TFs binding to (a) common DNA element(s). One such example would be the pigmentation-related gene, *TYR*. The *TYR* promoter contains three M-boxes (Figure 3.3.2A) (Bentley et al., 1994; Shibata et al., 1992) and is induced as part of the pigmentation process in response to α-MSH and ultraviolet (UV)-B irradiation (Abdel-Malek et al., 1999).

The *TYR* promoter (Hodgkinson et al., 1993) and enhancer M-boxes (Shibata et al., 1992) are targets of MITF and *TYR* gene expression is dependent on MITF (Krylov et al., 1997). UV responsiveness, but not basal *TYR* gene expression, is dependent on USF1 phosphorylation by p38 (Galibert et al., 2001), as expression of dominant negative USF that blocks binding of both USF1 and USF2 failed to significantly affect basal promoter activity.

To study promoter education at the the *TYR* gene, we first looked at basal promoter occupancy in 501mel cells. Based on our ChIP-seq results, only MITF was peak-called at the *TYR* promoter as reflected on the bedgraph visualised using UCSC genome browser.

(Figure 3.3.2B). Our ChIP-qPCR of HA-MITF and HA-USF at *TYR* enhancer M-box (Figure 3.3.2C), however, confirmed the reported binding of USF and MITF (Gaggioli et al., 2003) and also demonstrated the highest enrichment of MITF at both the promoter and enhancer compared to USF1 (by $\approx 2.3-$ fold) and USF2 (by $\approx 1.8-$ fold). In a complementary approach we performed *in vitro* EMSA of *TYR* enhancer M-box probe against nuclear extract derived from Phoenix cells transiently overexpressing USF1, USF2 or MITF (Figure 3.3.2D). The specificity of the protein binding was confirmed by using antibodies to super-shift the DNA-bound proteins. Lastly, we affirmed the requirement of USF and MITF in basal expression of the *TYR* gene through RT-qPCR of 501mel treated

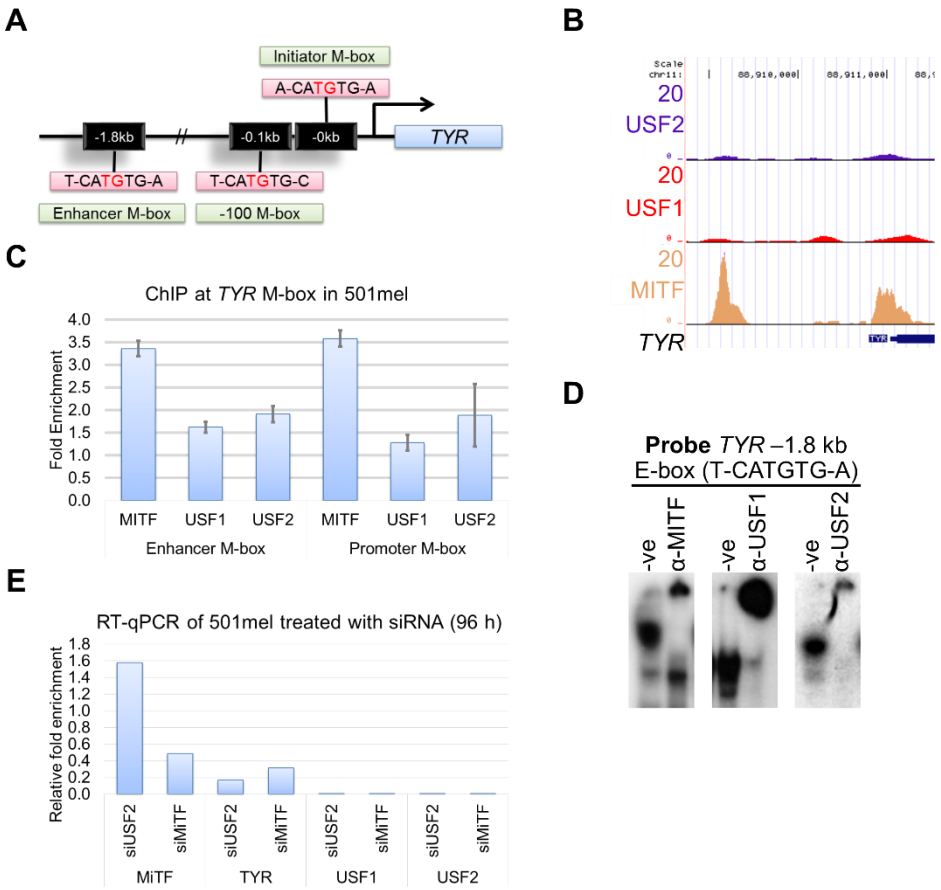


Figure 3.3.2 Binding of USF and MITF to the *TYR* enhancer. **A)** Schematic view of the *tyrosinase* (*TYR*) promoter. The three E/M-boxes upstream of the transcription start-site are indicated. **B)** Snapshot of MITF and USF bedgrapgh files at *TYR* promoter uploaded on to the UCSC genome browser. **C)** ChIP-qPCR of USF1, USF2 and MITF at the *TYR* enhancer M-box and promoter proximal M-box (qPCR primer spanning both -100 M-box and the initiator M-box). Fold enrichment is calculated using $\Delta\Delta C_t$ approach as described in Materials and Methods. $N = 2$ biological replicates each with 2 systematic replicates. **D)** EMSA showing binding of Phoenix nuclear extract transiently overexpressing either USF1, USF2 or MITF to *TYR* enhancer oligonucleotide probe. The identity of the proteins bound is confirmed using antibodies super-shifted as indicated. **E)** Gene expression analysis of 501mel cell subjected to siMITF $(83\ \mathrm{nM})$ or siUSF2 $(50\ \mathrm{nM})$ for $96\ \mathrm{h}$ before the samples were harvested and RNA-extracted for RT-qPCR analysis of *MITF*, *TYR*, *USF1* and *USF2* mRNA. Fold enrichment reported as value relative to siCN at $t = 96\ \mathrm{h}$, normalised to *β–ACTIN*. RNA samples from **3** biological replicates were pooled before subjected RT-qPCR analysis.

with siRNA against USF2 and MITF for $96\,\mathrm{h}$. At $96\,\mathrm{h}$, silencing either USF2 or MITF strongly suppressed USF1 and USF2 mRNA levels (Figure 3.3.2E). Although siMITF, resulted in $\approx 50\,\%$ reduction in *MITF* mRNA level, siUSF2 resulted in ≈ 1.6 –fold increase in *MITF* mRNA level. siMITF, which decreased USF and MITF, resulted in $\approx 65\,\%$ reduction in *TYR* mRNA level while siUSF2 which abolished both USF1 and USF2 but increased MITF mRNA levels, resulted in a slightly larger decrease in *TYR* mRNA level ($\approx 80\,\%$ reduction). The better reduction in *TYR* mRNA in the absence of USF may indicate that exchange between MITF and USF might result in a larger magnitude of *TYR* gene activation than MITF alone.

To study promoter education we needed to establish conditions where we could efficiently induce *TYR* gene expression. Several stimuli may be used to induce the *TYR* promoter. For example, UVB, the peptide hormone α-MSH and its synthetic analogues such as NDP-α-MSH (FitzGerald et al., 2006), and small molecules such as forskolin that induce MITF expression.

Induction of the *TYR* gene by UVB

In humans, MITF plays a central role in the tanning response through coordinated-regulation of melanocytic enzyme expression (Liu and Fisher, 2010). Following exposure to UV-irradiation, human epidermal melanocytes increase melanin synthesis in the compartmentalized organelles called melanosomes where all the melanogenic enzymes are situated. Melanosomes are transferred to the surrounding keratinocytes to increase skin pigmentation. This “tanning response” is specific to humans and was suggested to play a role in protection against UV-induced DNA-damage (Hill, 1992).

UVB-irradiation is among the various stimuli of p38 SAPK (Iordanov et al., 1997; Tan et al., 1996; Xing et al., 1998) that leads to phosphorylation-induced activation of CREB via MAPKAP-K2. CREB, in turn, activates the *MITF-M* promoter (Bertolotto et al., 1998a). Although MITF is a master regulator of all pigmentation-related genes, USF is required for

expression of at least some UV-responsive genes (Corre et al., 2004). For example, the expression and UV-responsiveness of the human Mimecan/osteoglycin, a corneal keratin sulphate proteoglycan, is entirely dependent on the E-box element in the promoter (Tasheva and Conrad, 2003). At our model gene, *TYR*, binding by USF and MITF are required for full activation (Bentley et al., 1994; Gaggioli et al., 2003; Galibert et al., 2001). Based on our promoter education hypothesis, different levels of USF and MITF proteins may result in different induction kinetics of the *TYR* gene. For example, cell lines with high basal level of MITF and USF1 proteins may be able to activate the *TYR* gene in response to UVB more rapidly and have a higher gene expression level than cell lines that only express a basal level of MITF. To address this possibility, we checked for the expression of USF and MITF in a number of melanoma/melanocyte cell lines by Western-blotting prior to their use. Refer to Supplementary Figure 3.3.1 for a more detailed discussion regarding the relative expression of USF and MITF across cell lines and the rationale behind choosing each cell line.

UVB irradiation activates a stress signalling cascade mediated by p38 that directly activates MITF by phosphorylating $S307$ and USF1 by phosphorylating $\mathrm{T153}$ (Corre et al., 2009; Galibert et al., 2001), resulting in activation of pigmentation related genes (Figure 3.3.3A). Whether p38 activation leads to exchange of MITF and USF at the *TYR* promoter has not been shown. To test this, we performed ChIP-qPCR of 501mel cells treated with $200\ \mathrm{J/m^2}$ of UVB after overnight serum-starvation (a pre-treatment to overcome p38 suppression by serum factors). We focussed our study on the *TYR* enhancer M-box as our ChIP-seq show the highest occupancy by MITF at the enhancer M-box (Figure 3.3.2B), and we were not able to identify qPCR primer pairs which worked well and able to separate the -100 E-box and the initiator E-box.

In 501mel cells, at $\mathrm{t} = 0$ we saw a relatively low M-box occupancy with USF2 being the highest (≈ 3.5 –fold, Figure 3.3.3B) relative to control. At $\mathrm{t} = 30'$, we saw maximal MITF (≈ 17 –fold), USF1 (≈ 8 –fold), USF2 (≈ 3 –fold) along with RNAPII (≈ 8 –fold relative to $\mathrm{t} =$

0) accumulation at the *TYR* enhancer M-box (Figure 3.3.3D, 30′). The accumulation of MITF and RNAPII rapidly subsided at $t = 3\ h$ (down to $\approx 3.5\ \&\ 0.2 - fold$ relative to $t = 0$, Figure 3.3.3B, 3 h). The kinetic of RNAPII accumulation that coincided with that of MITF, suggested that MITF might recruit RNAPII to the *TYR* enhancer M-box while USF might be involved in the release of RNAPII as USF signals remained after MITF and RNAPII leave. Alternatively, USF may be involved in transcription memory and their binding may help to retain the chromatin in an open state for rapid re-induction following the next round of UVB-induction. USF1, on the other hand, showed a much smaller reduction in ChIP signal (down to $\approx 6 - fold$ relative to $t = 0$) while USF2 almost returned to the signal at $t = 0$ and remained low when the cells were harvested after $120\ h$ (Figure 3.3.3B, 120 h). Interestingly, at this time point, we also observed a re-recruitment of MITF, USF1 and RNAPII. In the

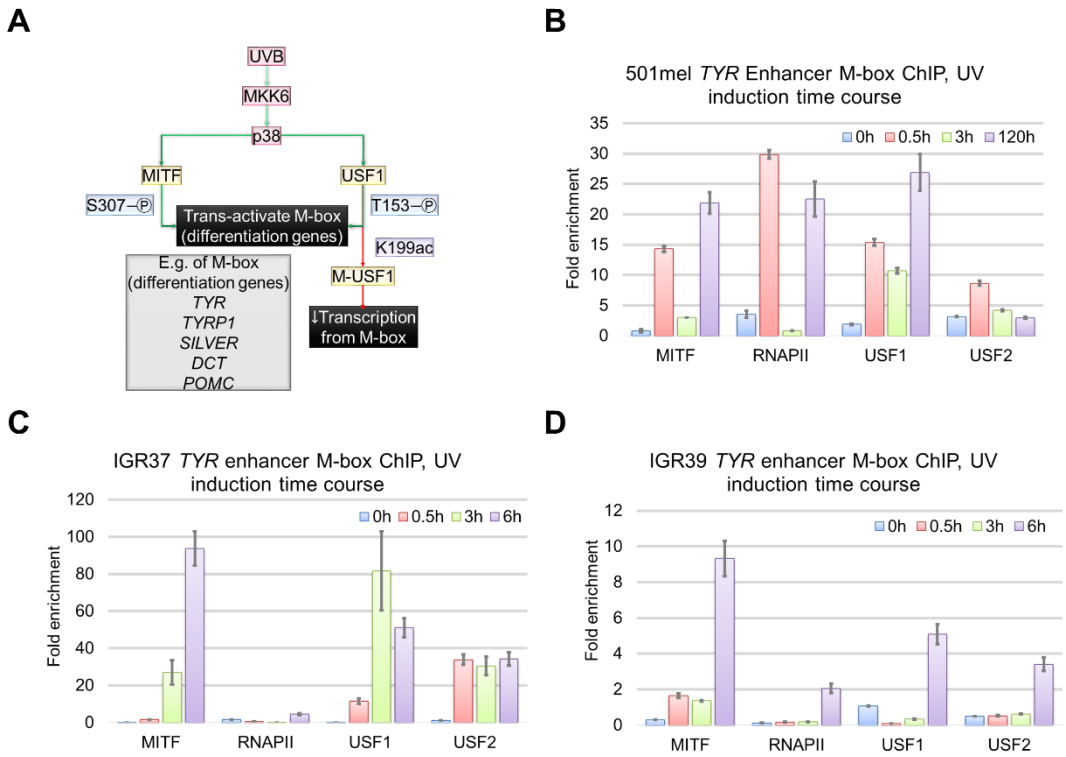


Figure 3.3.3 UVB irradiation results in changes in the ChIP profile at the *TYR* enhancer. **A)** Schematic representation of the signalling cascade induced in melanocytes following UVB irradiation. UVB irradiation triggers activation of MKK6 (stress–specific MAPKK) that in turn activates p38 stress-activated protein kinase. In the context of pigment cells, p38 mediates activation of MITF by phosphorylating S307. This phosphorylation gears MITF towards activation of differentiation genes, in other words, genes involved in pigment formation. p38 also phosphorylates USF1 at T153, resulting in USF1 with high transcription activation potential. However, prolonged signalling through p38 also leads to acetylation of USF1 at K199 creating a modified USF1 (M-USF1) that inhibits USF1's transcription activation potential. **B-D**) Time course of ChIP-qPCR at the *TYR* enhancer M-box following $200\ J/m^2$ UVB irradiation. The ChIPed proteins (MITF, RNAPII, USF1 and USF2) are indicated below each set of ChIP experiments for 501mel (**B**), IGR37 (**C**) and IGR39 (**D**). Fold enrichment is calculated using $\Delta\Delta C_t$ approach as described in Materials and Methods. $N = 3$ biological replicates each with 2 systematic replicates.

case of MITF and USF1, the enrichment was greater than at $t = 30'$ (≈ 1.5 & 1.7 –fold relative to $t = 0$, respectively). As UVB-treatment resulted in a high level of apoptosis, it is possible that cells which have survived and proliferate till $t = 120$ h are those which have especially high basal level of USF1 and MITF, allowing them to better cope with UVB-exposure. Hence, the observed high ChIP-signal in the absent of extra UVB-induction.

We also performed the induction experiments in IGR37 (Figure 3.3.3C) and IGR39 (Figure 3.3.3D) primary and metastatic melanoma cell lines, respectively. Their basal MITF level is much lower than 501mel cells (Supplementary Figure 3.3.1) but it can be induced (our lab unpublished observation). Without going into great detail, both cell lines demonstrated a much slower kinetic of MITF and RNAPII recruitment to the *TYR* enhancer M-box. This slower kinetic most likely reflected the time needed to induce MITF protein level when compared to 501mel cells. In IGR37 cells, recruitment of USF to the *TYR* enhancer occurred as early as $t = 30'$, which was similar to 501mel cells, possibly because both cell lines contained similar USF protein level (Supplementary Figure 3.3.1) . However, accumulation of USF1 only peaked at $t = 3$ h before decreasing at $t = 6$ h. USF2, on the other hand, remained bound to the *TYR* enhancer M-box throughout the duration of our experiments. In IGR39 cells, there was a slight decrease in USF1 occupancy at $t = 30'$ and USF did not really accumulate till $t = 6$ h. This delay in USF recruitment may be due to a much lower level of USF2 in IGR39 cells (Supplementary Figure 3.3.1). As USF1:USF2 heterodimers are the most common species in many cell lines, it is possible that the initial USF1 recruitment to the M-box following p38 activation requires dimerization with USF2. However, subsequent binding of USF1 at a much later time appeared to be independent of USF2 (compare Figure 3.3.3B, $30'$ to Figure 3.3.3B, 120 h).

Although the relative fold-enrichment displayed by each of the factors were not identical and there were some differences in the predominant factors occupying the enhancer E-box following the time course of UV induction; it was interesting to see that in all instances, increased M-box binding could be observed by all factors. This suggested that the observed

enrichments at each time point may reflect the average ChIPed signals of factors rapidly exchanging (fast on/off rate) or the average of the different population of cells with either USF or MITF bound to the *TYR* M-box, in the case where activation of *TYR* can be achieved by any of the USF and MITF. As it is physically impossible for an M-box to be occupied by USF and MITF simultaneously, USF and MITF must exchange following UVB induction. It remains to be seen whether USF and MITF coordinately exchange to activate *TYR* gene or they exchange passively and the increased ChIPped-signal following UVB induction simply reflected a more permissive chromatin structure for binding by USF and MITF.

The very different ChIP-qPCR fold-enrichment between the three cell lines following induction may be due to the different basal levels as well as the inducibility of USF and MITF proteins between the three cell lines. 501mel which have a high basal expression of MITF and USF, although enables it to response more rapidly to UV-induction, also means the dynamic range following induction will be lower than the IGR. The difference between the two IGR cell lines can be explained by the fact that IGR39 is dedifferentiated while IGR37 still retain its pigmented-cell phenotype. Based on the MITF-induced phenotype-switching model, it is conceivable that only a fraction of IGR39 switch back to express MITF.

Complications with the use of UVB induction

UV exposure activates a number of TFs including p53 (Prives and Hall, 1999), AP-1 (activator protein 1) (Minden and Karin, 1997), NF-κB (nuclear factor κB) (Li and Karin, 2000), Oct-1 (octamer binding factor) (Zhao et al., 2000), SRF (serum-response factor) (Chai and Tarnawski, 2002) and USF1 (Galibert et al., 2001). The binding sites of these factors are often present at the promoters of UV responsive genes. This could potentially complicate our analysis of promoter education, especially in the light of changes in transcription response when USF or MITF's functions are interfered with, if the gene being studied is also regulated by other TFs. Equally important, is the complication with the survivability of the cells. The reason that the time-course experiments carried out above for the IGR cells were only up to 6 h was due to high level of apoptosis, leaving too few cells to

be used for the full ChIP-experiments, rendering it very difficult to perform studies at time points beyond $6\ \mathrm{h}$. We only included in this thesis the results obtained up to $\mathrm{t} = 3\ \mathrm{h}$ and skipped all the time points until $\mathrm{t} = 120\ \mathrm{h}$ for 501mel because we only had sufficient material to perform ChIP for USF at $\mathrm{t} = 6\ \mathrm{h}$ (whose fold-enrichment returned to non-induced state, data not shown) and only until $\mathrm{t} = 120\ \mathrm{h}$ did the cells recover in sufficient quantity to be used for the ChIP experiment. Moreover, different activation pathways that result in a different profile and activation kinetic of downstream TFs potentially result in different initial induction kinetics and different re-activation potentials. To this end, we explored various alternative routes to induce factor exchange.

Induction of the *TYR* gene by an α-MSH analogue

The tanning response in human skin following UVB irradiation involves keratinocyte-secreted α-MSH binding to the MC1R presented on the surface of melanocytes (García-Borrón et al., 2005; Rouzaud et al., 2005). Activation of MC1R results in activation of p38 SAPK and activation of adenylyl cyclase through G-protein signalling resulting in elevated cAMP levels (Figure 3.3.4A) (Bertolotto et al., 1998b; Chhajlani and Wikberg, 1992). Activation of the p38 pathway leads to activation of USF1 and MITF as described above. Activation of cAMP, however, results in elevated *MITF* transcription through CREB (BuscÀ and Ballotti, 2000). Our working hypothesis for factor exchange following α-MSH addition involves initial binding of USF1 in competition with MITF as a result of p38 activation. MITF protein accumulation through increased *MITF* transcription via the cAMP pathway coincides with gradual accumulation of the transcription incompetent acetylated M-USF1 in a manner parallel to the length of p38 activation (Corre et al., 2009). This leads to the displacement of USF1 by MITF and probably corresponds to maximal *TYR* transcription. Thus, crucial to the induction of *TYR* expression via MC1R is the accumulation of MITF protein. Therefore, we monitored the accumulation of MITF as a read-out of induction via MC1R.

We carried out a time course induction using various α-MSH analogues in a number of cell lines. Shown here are the results for afa-melanotide in HERMES and 501mel cells,

CUV9900 in HBL cells and NDP α-MSH in HERMES cells (Figure 3.3.4B – E). In HERMES, MITF accumulates from the 30' time-point and decreases by 24 h post afa-melanotide (Figure 3.3.4B) and by 48 h post NDP α-MSH (Figure 3.3.4C) treatment. MITF accumulation at the 30' time-point was probably through p38 regulation of MITF protein stability, as has been shown for activated *MITF* in response to lipid-stimulated activation of p38 (Saha et al., 2006), while MITF accumulation at later time points most likely mediated through changes in *MITF* mRNA levels.

In line with the observed accumulation of MITF, RT-qPCR of HBL melanoma cells treated with CUV9900 showed increased gene expression from MITF-downstream differentiation genes such as *TYR* and *Silver*. The *MITF* gene was also induced at $t = 30'$ compare to $t = 10'$ while USF1 decreased following CUV9900 treatment (Figure 3.3.4D).

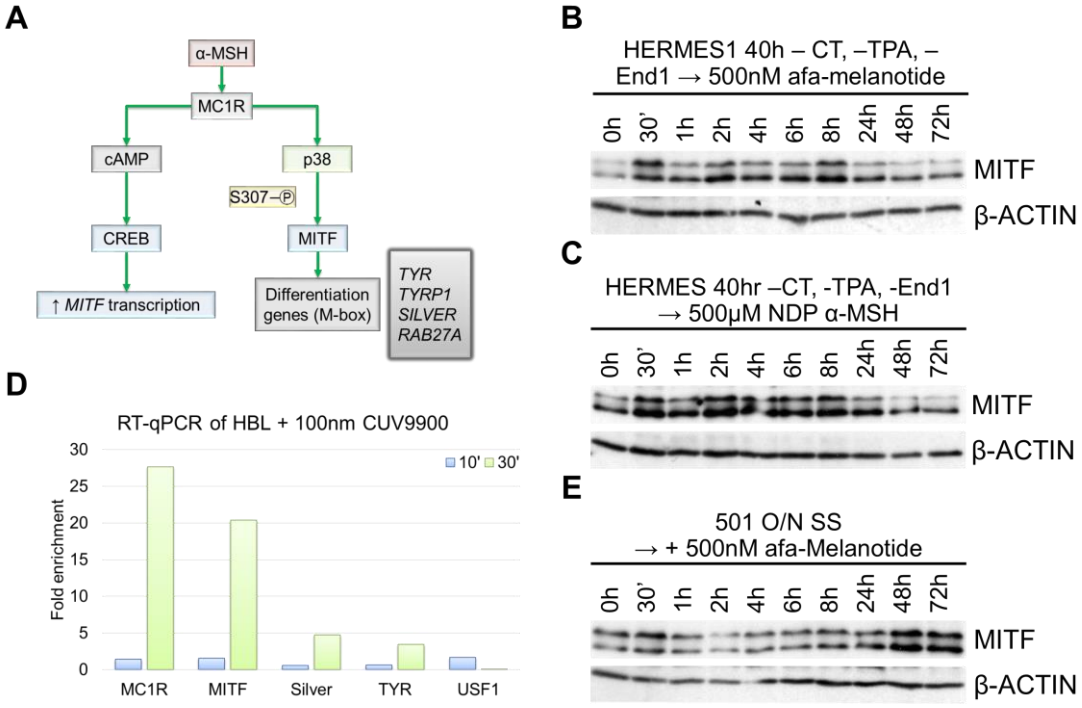


Figure 3.3.4 Induction of MITF by α-MSH analogue. **A)** Schematic pathway depicting the two downstream effect of α-MSH on MITF activation. The binding of the peptide hormone α-melanocortin stimulating hormone (α-MSH) to its cognate receptor, melanocortin receptor (MC1R) results in activation of cyclic-AMP (cAMP) and p38 signalling pathway. cAMP acts on CREB to increase *MITF* transcription while p38 phosphorylate MITF, gearing it towards differentiation genes. Western-blot of whole cell lysate from HERMES treated with 500 nM afa-melanotide (α-MSH analogue) (**B**) or 500 μM NDP α-MSH (**C**). HERMES cells were subjected to 40h serum-, cholera toxin (CT)-, TPA- and endothelin-1 (End1)- starvation to suppress cAMP response prior to induction. The lysates were collected at the indicated time point following induction. β-ACTIN was used as a loading control. **D)** Gene expression analysis by RT-qPCR of the HBL melanoma cell line treated with 100 nM CUV9900. Fold enrichment is calculated using $\Delta\Delta C_t$ approach as described in Materials and Methods and presented as values relative to normalised values at $t = 0$. **E)** Same as **B** for 501mel melanoma. 501mel cells were serum starved overnight prior to afa-melanotide induction.

We performed the induction in 501mel cells to assess the effect on MITF levels following MC1R activation independent of CREB; 501mel cells have a defective cAMP response (José Carlos Garcia Barron, personal communication), so changes in MITF levels are most likely through p38 activation. In 501mel, there was a transient and minor increase of MITF protein at the 30' time-point. Significant MITF protein accumulation was only evident at around 48 h post-induction which was unlikely to be a result of protein stabilisation via p38 (Figure 3.3.4E) and may be mediated via a pathway not investigated in our study. For example, p38 can also activate CREB via MAPKAP-K2 that would result in increased transcription from *MITF* promoter via CRE. In any case, the lack of significant changes in MITF accumulation was also accompanied by no major changes in promoter occupancy following administration of α-MSH analogues (data not shown).

Induction of MITF via Methotrexate

Our lab has recently published a directed phenotype-switching in melanoma through induction of *MITF* gene by methotrexate (MTX)(Saez-Ayala et al., 2013). Although the exact

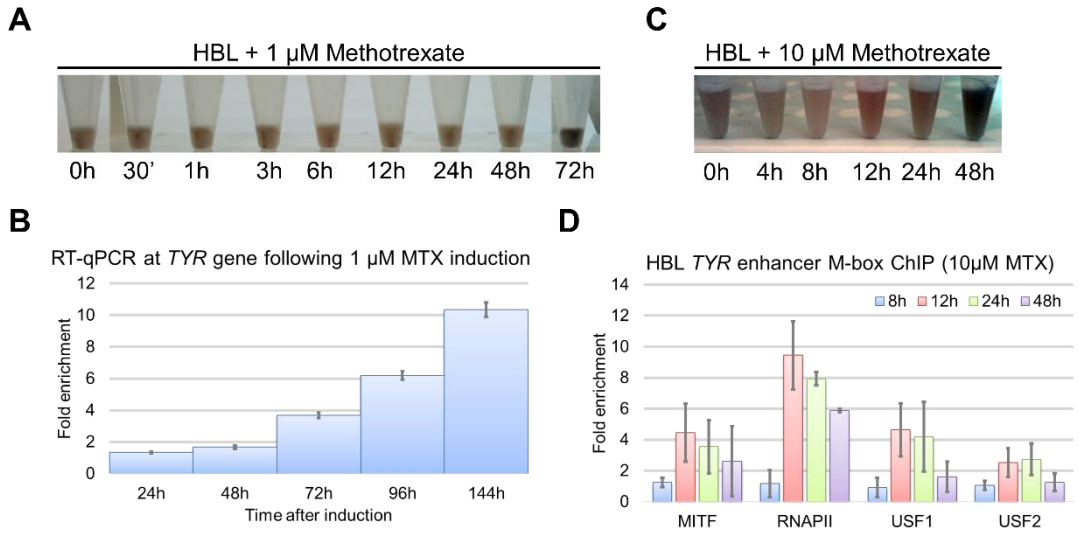


Figure 3.3.5 Methotrexate induces pigmentation via induction of *TYR* gene expression in HBL cells. **A)** Photograph of HBL melanoma cell line treated with 1 μM methotrexate (MTX). Cells were trypsinised at the indicated time points, washed in PBS, centrifuged and photographed. **B)** HBL cells treated with 1 μM MTX and samples collected at the indicated time points for gene expression analysis of *TYR* gene by RT-qPCR. Fold-enrichment is reported as normalised values (to *β-ACTIN*) relative to $t = 0$. $N = 3$ biological replicates each with 2 systematic replicates. **C)** Same as **A** but HBL cells were treated with 10 μM MTX and resuspended in 200 μl PBS before being photographed. **D)** HBL cells were treated with 10 μM MTX and samples collected at the indicated time points for ChIP-qPCR analysis at the *TYR* enhancer M-box using the indicated antibodies (MITF, RNAPII, USF1 and USF2). Fold enrichment is calculated using $\Delta\Delta C_t$ approach as described in Materials and Methods and presented as values relative to normalised values at $t = 0$. $N = 3$ biological replicates each with 2 systematic replicates.

mechanism which MTX induces *MITF* expression is yet to be elucidated, it promised to be an attractive and clean alternative to study factor exchange as it robustly activated MITF in a number of cell lines even in the dedifferentiated, non-pigmented IGR39 cells. To demonstrate the ability of MTX to induce *TYR* gene and pigmentation, we tested a number of cell lines simply by following the level of pigmentation within the cells. Among the panel of cell lines we tested (data not shown), HBL demonstrated the best inducibility by basic colorimetric inspection and at the *TYR* mRNA level following MTX induction. HBL, was the only cell line we were able to visually see pigment accumulation by $72\ \mathrm{h}$ following $1\ \mu\mathrm{M}$ MTX (Figure 3.3.5A) induction. This is also reflected in the progressive accumulation of *TYR* mRNA level from $\approx 2-$ fold at $\mathrm{t} = 24\ \mathrm{h}$ upto $\approx 10-$ fold at $\mathrm{t} = 144\ \mathrm{h}$ relative to $\mathrm{t} = 0\ \mathrm{h}$ (Figure 3.3.5B). At $10\ \mu\mathrm{M}$ (the concentration used for other cell lines), accumulation of pigment within the HBL cells can be seen as early as $12\ \mathrm{h}$ (Figure 3.3.5C). This accumulation coincided with the maximal, MITF, USF and RNAPII ChIPed-signal at the *TYR* enhancer that gradually decreased from that point (Figure 3.3.5D). More specifically, the MITF ChIP-signal changed from $\approx 1.2 \rightarrow 4 \rightarrow 3.8 \rightarrow 2.5-$ fold at $\mathrm{t} = 8\ \mathrm{h} \rightarrow 12\ \mathrm{h} \rightarrow 24\ \mathrm{h} \rightarrow 48\ \mathrm{h}$ relative to normalised-value at $\mathrm{t} = 0$. A similar trend was observed for USF1, peaking at $\mathrm{t} = 12\ \mathrm{h}, 4.5-$fold relative to at $\mathrm{t} = 0$ before decreasing. The enrichment of USF2 was more modest, peaking at $\mathrm{t} = 24\ \mathrm{h}, 2.6-$fold relative to $\mathrm{t} = 0$ before decreasing. RNAPII also peaked at $\mathrm{t} = 12\ \mathrm{h}, 9-$ fold relative to normalised-value at $\mathrm{t} = 0$ before decreasing.

Having demonstrated *TYR* gene activation and changes in ChIP profile at the enhancer M-box following MTX-induction with no immediate detrimental effect on the cell viability, we asked if we could use MTX to study transcription memory at the *TYR* gene. To this end, we explored changes in the ChIP-profile and the associated gene expression profile at the *TYR* gene following re-induction with MTX. However, MTX induces accumulation of cells in the S-phase of the cell-cycle, most likely through depletion of dTTP pools and hence interferes with DNA-replication. To avoid complications associated with the indirect effect of cell-cycle

arrest on the study of promoter education at the *TYR* gene as well as to ensure that the cells have rid themselves of MTX from the initial induction prior to our re-induction study we performed cell-cycle analysis of 501mel cell treated with $10\ \mu M$ MTX then chased with fresh-RPMI (no MTX) as depicted in (Figure 3.3.6A). As MTX induced cell-cycle arrest through induction of MITF, we made the assumption that MITF levels must have returned to normal when the cell-cycle profile return to the non-induced state and re-induction should only be carried out after this time point.

Example of stack bar-graphs of the cell cycle distribution of 501mel cells treated with MTX and RPMI-chased, and the associated flow-cytometry profiles of nuclear staining in the PE-channel are shown in Figure 3.3.6B & C, respectively. For this analysis we could confirm that MTX induced an accumulation in S-phase of the cell cycle which could be seen to increase from $\approx 18\ \%$ in the non-treated samples to $\approx 27\ \%$ at $52\ h$ MTX and continued to increase to $\approx 36\%$ at $50\ h$ RPMI-chased. The cell-cycle profile only returned to the pre-induced state after an $\approx 96\ h$ RPMI-chase (60% G1, $\approx 18\ \%$ S and $\approx 20\ \%$ G2). Therefore, we established a $96\ h$ RPMI-chase to be the time gap before re-induction.

From visual inspection, the kinetic of pigment accumulation in 501mel cells following initial round of MTX-induction is much slower than that of HBL cells. While pigment accumulated as early as $12\ h$ and cells appear black by $\approx 48\ h$ in HBL cells (Figure 3.3.5C), it took $\approx 120\ h$ for 501mel cell suspension to appear black (Figure 3.3.6D). This was in good agreement with *TYR* mRNA level which peaked at $96\ h$ corresponding to $\approx 11-$ fold enrichment relative to $t = 0$ (Figure 3.3.6E, Initial induction). The highest level of pigmentation presumably occurred with a slight time lag following the maximal *TYR* mRNA level. Interestingly, re-induction of 501mel cells with $10\ \mu M$ MTX after $96\ h$ RPMI-chased did not result in induction of the *TYR* gene (Figure 3.3.6E, Re-induction). We performed ChIP-qPCR at the *TYR* enhancer to address this observation. For simplicity fold-enrichments are presented as values relative to $t = 0$.

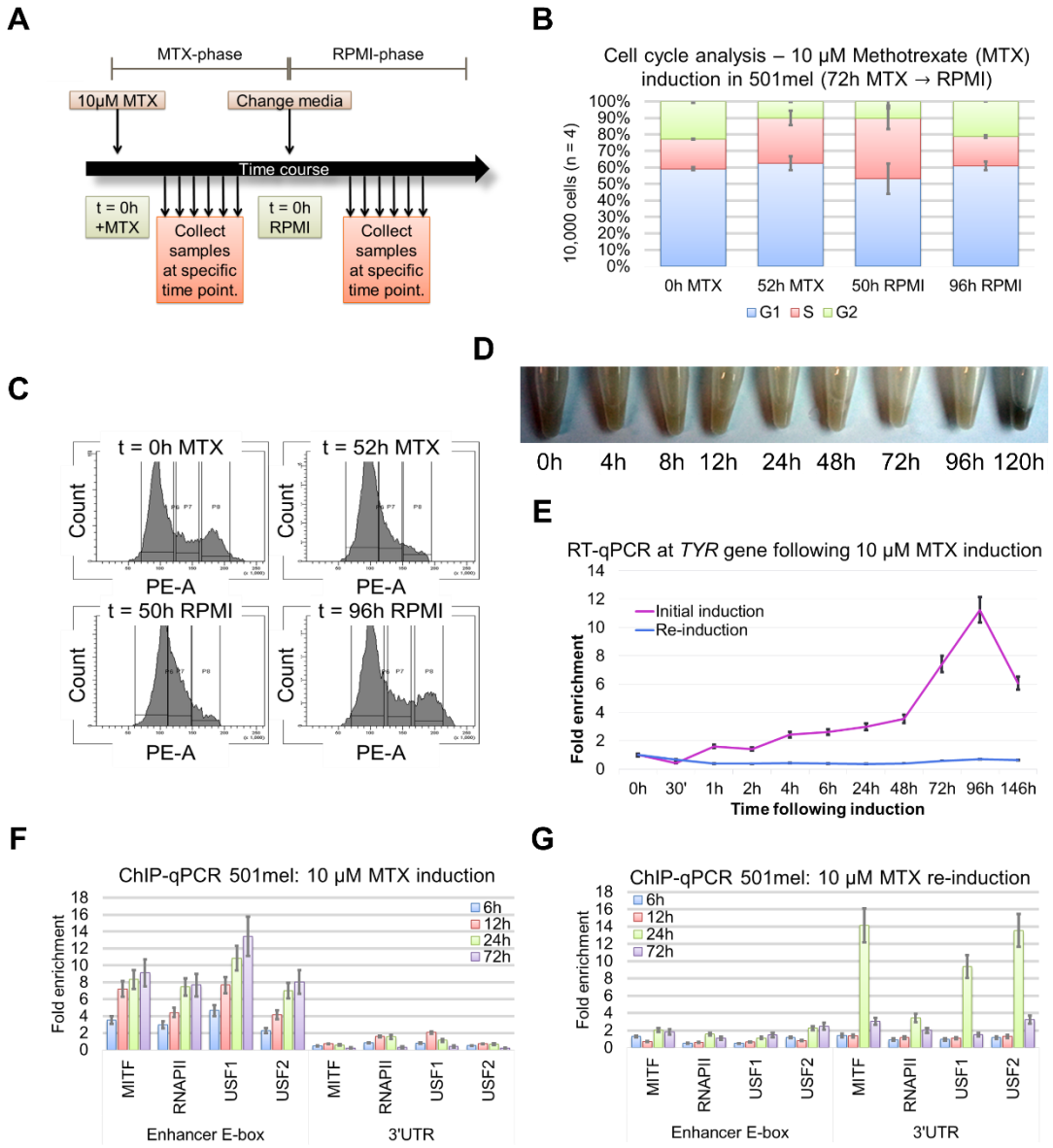


Figure 3.3.6 Transcription memory at the *TYR* gene in 501mel cells following induction by methotrexate. **A)** Schematic diagram of cell-cycle analysis following induction by methotrexate (MTX) in 501mel. MTX was shown to induce cell-cycle arrest through MITF up-regulation. To ensure that subsequent studies of transcription memory at the *TYR* gene following MITF induction via MTX is not complicated by secondary effects of cell-cycle arrest, we performed a time-course experiment of MTX-induction followed by RPMI-phase and samples were collected till the time point where the cells have returned into cycle. **B)** Stacked bar-graph demonstrating cell cycle distribution in to G1, S and G2 at the indicated time-points following MTX induction and RPMI-phase**.** $N = 3$ biological replicates. **C)** Examples of cell cycle profiles and the gating used to obtain the values used to plot the graphs in **B**. **D)** Photograph of 501mel cells treated with $10\ \mu M$ MTX. Cells were trypsinised at the indicated time point, washed in PBS, centrifuged and resuspended in $150\ \mu l$ PBS before being photographed. **E)** Gene expression analysis by RT-qPCR of 501mel cells treated with $10\ \mu M$ MTX and samples collected at the indicated time points for analysis. The purple line corresponds to 501mel cells treated with MTX for the first time ("initial induction") while blue line correspond to 501mel treated with MTX for $72\ h$ then MTX-chased by recovering the cells in RPMI for $96\ h$ then re-induced with $10\ \mu M$ MTX ("re-induction"). Fold enrichment is calculated using $\Delta\Delta C_t$ approach as described in Materials and Methods and presented as values relative to normalised values at $t = 0$. $N = 3$ biological replicates each with 2 systematic replicates. **F-G)** ChIP-qPCR analysis at the *TYR* enhancer M-box in 501mel cells treated with $10\ \mu M$ MTX and samples collected at the indicated time point and ChIPed with the indicated antibody. Fold enrichment is calculated using $\Delta\Delta C_t$ approach as described in Materials and Methods and presented as values relative to normalised values at t=0 for the "initial induction" (**F**) and "re-induction" phase (**G**). $N = 3$ biological replicates each with 2 systematic replicates.

Following the initial induction with $10\ \mu M$ MTX, we could see a gradual enrichment of MITF, USF and RNAPII at the *TYR* enhancer but not at the 3'UTR all the way to $t = 72\ h$, the final time point we measured (Figure 3.3.6F). The increased binding by all factors studied following induction is similar to what was observed for *TYR* gene induction following UVB-irradiation (Figure 3.3.3B – D). More specifically, MITF was observed to peak at ≈ 9.1 –fold while USF1 was at ≈ 13.5 –fold and USF2 at ≈ 8 –fold at $t = 72\ h$ relative to $t = 0$. RNAPII demonstrated similar trend of enrichment, albeit plateau out after $t = 24\ h$ corresponding to ≈ 7.5 – fold enrichment relative to $t = 0$. We also looked at MITF, USF and RNAPII enrichment along the *TYR* gene but only shown here the 3'UTR as this was the only region which demonstrated some differences following the induction and re-induction. For the initial-induction, there is a small gradual increased in ChIP-signal up to $t = 24\ h$ before decreasing to the level slightly below that of $t = 6\ h$ by $t = 72\ h$, except for USF1 which peaked at $t = 12\ h$ before decreasing.

We also performed an analogous ChIP of re-induced 501mel cells following the initial $72\ h$ induction and $96\ h$ RPMI-chase, then re-induced with $10\ \mu M$ MTX (Figure 3.3.6G). In line with the observed lack of induction at the mRNA level (Figure 3.3.6E), there was essentially no MITF, USF and RNAPII ChIP-signal at the *TYR* enhancer E-box. We could see a very minor increase following the induction time course, but these signals were mostly below 2 –fold. Intriguingly, however, was the observed peak of MITF and USF ChIP-signal at the 3'-UTR of the *TYR* gene at $t = 24\ h$ corresponding to ≈ 14 –fold enrichment of MITF, ≈ 9 –fold enrichment of USF1 and ≈ 13.5 –fold enrichment for USF2 relative to $t = 0\ h$. RNAPII, on the other hand, only gave an enriched signal of ≈ 3.4 –fold at the same time point. This was in a striking contrast to an approximately equal fold-enrichment of RNAPII compared to that of MITF at the enhancer M-box at $t = 24\ h$. The lack of RNAPII signal argued against transcription of any anti-sense transcript and transcription interference being responsible for suppression of the *TYR* sense-transcript. Although, we were not able to determine the nature of the signal at $t = 24\ h$, we could conclude that *TYR*

gene expression was refractory to further induction even after $96\ \mathrm{h}$ time gap when the effect of MTX on MITF-induced cell cycle arrest has already subsided. Therefore, the *TYR* gene, at least in response to MTX, can be educated such that it becomes refractive to the second round of induction, at least under the time-course studied. It is possible that this dulled induction of the *TYR* gene may be due to a decreased induction of MITF following the second round of MTX-induction. This was not addressed at the time of the experiment but will have to be investigated in order to determine whether the *TYR* or *MITF* gene is a *bona fide* example of gene that retains transcription memory in response to MTX-induction. Alternatively, the promoter activity of the *TYR* gene may not have returned to the pre-induced state preventing further induction in response to the second round of MTX-induced MITF accumulation. We may assess these by using two independent means to consecutively induce MITF, for example, initial induction by MTX followed by doxycycline induction of MITF in a cell line stably expressing doxycycline inducible MITF.

Other means to induce factor exchange

So far, we have explored means to induce factor exchange at our model gene (*TYR*) through MITF protein accumulation. In this section we explored the converse, decreasing MITF protein level. One such condition that our lab is exploring is amino-acid starvation. It was noted that melanoma cells grown in MEM lacking three-amino-acids (Q, S and G) relative to RPMI displayed depleted MITF protein. The underlying mechanism(s) of metabolic regulation of MITF is an active, on-going research topic in our lab and is beyond the scope of this thesis. We present here the kinetic of MITF demise following amino-acid starvation as well as USF2 ChIP at some of the genes in the HA-USF2 stable cell line upon amino-acid starvation and re-feeding.

Western-blotting of 501mel cells grown in MEM media showed the initial shift of MITF from the faster-migrating band to the slower migrating band, indicating increased phosphorylation at $S73$ which this upper band correspond to (Figure 3.3.7A). This shift was transient and was only seen at $\mathrm{t} = 1\ \mathrm{h}$. By $\mathrm{t} = 24\ \mathrm{h}$ we could see a significant decrease in

MITF levels ($> 3-$fold). This decrease may reflect increased protein degradation or reduced translation from MITF mRNA and not transcription as a similar reduction was observed for HA-MITF that is expressed from the *CMV*-promoter in the HA-MITF stable cell line. Concomitant with the decrease in MITF, we saw an increase in USF1 at $t = 24$ h which was not reproduced in extracts from the HA-MITF cell line. USF2 protein levels, on the other hand, were constant throughout the 24 h time-course in both 501mel and HA-MITF cells regardless of the antibody used. (We will discuss more about different USF2 antibodies in section 3.4.) GAPDH was used as a loading control.

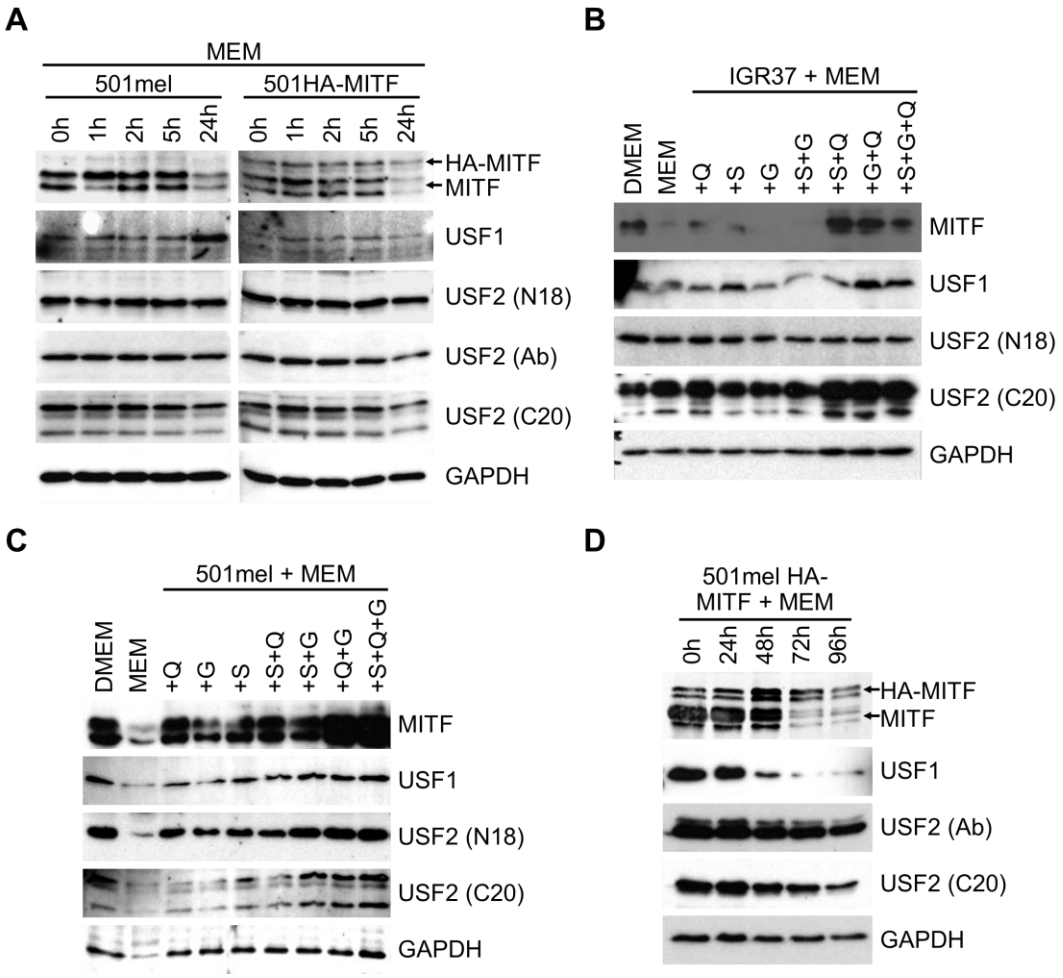


Figure 3.3.7 Regulation of MITF through –Q, –S and –G amino-acid starvation. **A)** Western-blot of whole cell extract obtained from 501mel and 501mel stably expressing HA-MITF grown in MEM instead of the standard RPMI. The samples were collected at the indicated time point and probed with the indicated antibodies. **B)** Western-blot of whole cell extract obtained from IGR37 cells grown in the indicated condition for 24 h. DMEM was used as a control while MEM was used as a control for the growth in MEM + one or more amino-acids supplement to re-constitute to the level presented in the DMEM. **C)** Same as **B)** for 501mel cells. **D)** Same as **A)** for 501mel-HA-MITF under longer time scale. GAPDH was used as a loading control.

As MEM differs from RPMI in the three-amino-acids (Q, S and G), we went on to explore the amino-acid(s) responsible for the reduction in MITF protein level by supplementing MEM with one or more of the missing amino-acids to the level found in DMEM. We performed the initial experiment in IGR37 cells, as this cell line demonstrated the most striking response to MEM treatment with regard to MITF reduction (Figure 3.3.7B). From the Western-blot we could infer that MITF protein can be restored to the level of DMEM upon the addition of Q and either S or G. Note that, under all conditions tested, the level of USF1 and USF2 remain largely unchanged after $24\ \mathrm{h}$. Although when compared to the GAPDH loading control, USF2 as detected by N-18 may be slightly decreased in the double and triple amino-acid supplemented cells except in samples supplemented with S and G.

As we are most interested in the response in 501mel, we repeated the same amino-acid reconstitution in the 501mel cell line (Figure 3.3.7C). In 501mel cells addition of Q, alone was sufficient to restore the level of MITF protein to that of the DMEM while G and S partially restored MITF levels but did not appear to have an additive effect on MITF protein levels when presented together. USF1 appeared to remain constant under all conditions tested while USF2 levels appeared to be decreased by ≈ 2 –fold in MEM, MEM supplemented with Q, G, S and S + Q when detected with USF2 (N18) & (C20).

The use of amino-acid starvation was not without complications, however. Although we consistently observed a reduction in MITF protein levels, the exact time point at which this happened is somewhat variable between experiments. Moreover, USF levels are also affected at later time point. For example, when we repeated the experiment in the HA-MITF cell line in Figure 3.3.7D, we observed a greatly reduced level of USF1 by $48\ \mathrm{h}$ in MEM while a significant reduction of endogenous MITF was not observed until $\mathrm{t} = 72\ \mathrm{h}$. The level of HA-MITF (expressed under the *CMV*-promoter), on the other hand, was not affected until $\mathrm{t} = 96\ \mathrm{h}$. USF2 remained largely unchanged until $\mathrm{t} = 96\ \mathrm{h}$ where it decreased by ≈ 2 –fold with the C20-antibody and $\approx 30\ \%$ with the Ab-antibody when taking into account the GAPDH loading control.

This inconsistency led us to experiment with different amino-acid starvation conditions. While Q, G and S can be synthesised in humans, L and K absent in SILAC media are essential amino acids. That is, they cannot be synthesised in human cells. To this end we carried out a time-course starvation experiment where 501mel cells grown in DMEM media were transferred to SILAC media and whole cell extracts analysed by Western-blot (Figure 3.3.8A). In the SILAC media, we could see decreased MITF protein within the first hour of amino-acid deprivation. The level remained consistently low throughout the $24\ \mathrm{h}$ time-course. At $\mathrm{t} = 24\ \mathrm{h}$ MITF existed predominantly in the middle migrating isoform. (Depending on the running conditions, sometimes we are able to resolve MITF to three-bands. Work done in our lab had identified the middle band to be dependent on phosphorylation by ERK.) USF1 levels remained relatively constant, while USF2 increased slightly $\approx 1.5-$fold between $4\ \mathrm{h}\ \&\ 6\ \mathrm{h}$, taking into account GAPDH loading control. This increase in USF2 was consistently observed using all USF2 antibodies.

It is worth noting, however that although 501mel cells were able to survive growth in MEM without significant cell death up to $\approx 48\ \mathrm{h}$, 501mel cells grown in SILAC media do not survive well over $24\ \mathrm{h}$. We attempted to alleviate this problem by supplementing SILAC

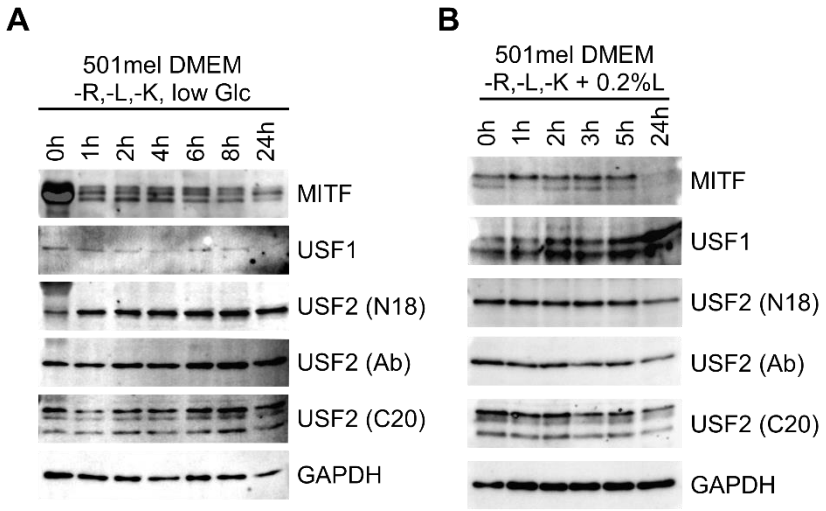


Figure 3.3.8 Regulation of MITF through –R, –L and –K amino-acids and low glucose starvation. **A)** Western-blot of 501mel whole cell extract grown in SILAC media, lacking R, L and K and samples collected at the indicated time points and probed with the indicated antibodies. **B)** Same as **A)** but the SILAC media was supplemented with $0.2\ \%$ L (relative to the level in DMEM). GAPDH was used as a loading control. SILAC media contains $1\mathrm{mg/ml}$ glucose compare to $4.5\mathrm{mg/ml}$ glucose in DMEM.

media with $0.2\ \%$ L as it was shown by another member of the lab that IGR37 cells grown in SILAC media supplemented with $0.2\ \%$ L had a much improved viability at longer incubation times. In the 501mel cells we were able to improve the survivability of the cells with the $0.2\ \%$ L supplement. However, our Western-blot of MITF indicated that this improved survivability was due to the much later time point at which MITF levels reduce (Figure 3.3.8B). MITF protein level only diminished at the $24\ \mathrm{h}$ time-point in SILAC media supplemented with $0.2\ \%$ L. USF1 remained largely unchanged while USF2 showed a minor decrease (≈ 1.5 –fold) in the protein level at the $24\ \mathrm{h}$ time-point by all antibodies.

Changes in E-box occupancy following starvation

We proceeded to investigate changes in the promoter occupancy when MITF protein is reduced following amino-acid starvation. As we no longer have the antibody clone which we used for our initial MITF ChIP at the *TYR* gene and the new batch of our MITF antibody (in-house) did not perform well for either ChIP or IF, we attempted the experiment in our HA-USF2 cell line, looking at USF2 signals. We made the assumption that for any sites that USF2 co-occupied with MITF and/or MYC, starvation, treatment that reduced MITF and/or MYC protein level would result in an increased occupancy by USF2 as there would be less competition with MITF and/or MYC in these cells. As MEM-treatment also resulted in an increase in G1-cell cycle arrested population (Supplementary Figure 3.3.2), we also assumed a reduction in MYC that is expressed in proliferating cells.

Whether we see factor exchange (increased USF2 binding when MITF and/or MYC level decreases) will depend on USF2's affinity for a given locus relative to MITF and MYC. One could imagine that very weak USF2 targets relative to MITF and/or MYC would not yield a very noticeable change in USF2 binding following MITF depletion. At the other extreme, a very strong USF2 target that is also a very weak MITF and/or MYC target will also not yield any major changes in USF2 occupancy following MITF and/or MYC removal as these loci will already be predominantly occupied by USF2. By extrapolation, we anticipate that sites

which are bound strongly both by MITF and/or MYC and USF2 to represent loci with highest magnitude of factor exchange.

To address this hypothesis we looked at a number of genes co-occupied by at least one of the USFs and MITF and/or MYC. We performed ChIP-qPCR analysis of HA-USF2 cell line grown in either RPMI or MEM for 72 h (Figure 3.3.9A) or MEM for 72 h then grown in RPMI for a further 24 h (as a recovery phase for MITF, Figure 3.3.9B). The genes in the graph are ranked according to the ratio of USF2 peak score over the combined MITF + MYC peak score. From our previous analysis in section 3.1.7 we know that the normalised peak scores do not correlate perfectly with ChIP-qPCR fold-enrichment and should only be used as a guideline. Nevertheless, we used these values to calculate the ratio of the combined (USF1 + USF2) peak score over the (MITF + MYC) combined peak score, as well as the ratio of USF2 peak score over the combined (MITF + MYC) peak score (Figure 3.3.9C). The former value was calculated as USF tends to bind as a heterodimer although we would focus on the ratio of USF2 peak score over combined (MITF + MYC) peak score as our Western-blots indicated that USF1 is down-regulated at the protein level following growth in MEM (Figure 3.3.7D). Based on our hypothesis, genes which show a ratio of USF2/(MITF + MYC) peak score close to one and have a high USF2 peak score are expected to demonstrated high enrichment of USF2 in MEM relative to RPMI whereas genes with a very low and very high USF2/(MITF + MYC) ratio are expected to demonstrated little enrichment of USF2 under MEM growth media relative to RPMI.

In support of our hypothesis, genes which are weak USF2 targets including *EGLN3*, *TNPO2*, *ASNA1* and *QTRT1* displayed only minor (between $2-3-$fold) differences in USF2 binding when comparing cells grown in MEM to those grown in RPMI (Figure 3.3.9A). At the *LUC7L2* and *PLAG1* genes where the USF2/(MITF + MYC) peak scores are around one, we saw a larger enrichment of USF2 in MEM relative to RPMI ($\approx 10.5-$ & $6-$fold, respectively). As *LUC7L2* has a very low normalised USF2 peak score, this suggests that

for MYC and USF2 co-occupied sites, the relative ratio of USF/MYC peak score might be a better predictor of whether we can expect to see a larger magnitude of factor exchange. However, this argument would contradict what is observed for *QTRT1* and *SLC33A1* where the ratio of USF2/(MITF + MYC) peak score is closer to one but did not display a large enrichment of USF2 when cells were grown in MEM media. For *PLAG1*, it is possible that starvation by MEM has a larger effect on MYC protein level than MITF. As *LUC7L2* is a MYC target but not an MITF target, we saw a larger enrichment of USF2 binding in MEM-media, while *PLAG1* and *QTRT1*, which are targets of MITF but not MYC displayed much smaller enrichment of USF2 under MEM relative to RPMI media. In any case, we did not assess the relative changes in MYC and MITF protein level following starvation in the samples used in this specific experiment, and therefore we will have to address this before any firm conclusion can be drawn.

Another point to consider is the relative bias for USF1 homodimer binding over USF2 homo- and heterodimers. *LUC7L2* displays a high enrichment of USF2 in MEM over RPMI growth conditions and has a close to $1:1$ ratio of USF1 and USF2 normalised peak scores. This was the same for the *PLAG1* E-box. The *QTRT1* and *SLC33A1* E-boxes, on the other hand, displayed a bias towards USF1 over USF2 ($\approx 1.34-$fold & $2.08-$fold). Therefore, a reduction in MITF and MYC, although freeing up the E-boxes, may favour USF1 homodimer binding over USF2, hence the smaller differences. The fact that *LUC7L2* also displays a bias of USF1 over USF2 ($\approx 1.27-$fold), which is similar to *QTRT1* yet displays a large difference in USF2 enrichment, supports our notion that MEM treatment has a larger effect on MYC than MITF. Moreover, although *SLC33A1* has a larger bias toward USF1 over USF2 compared to *QTRT1*, *SLC33A1* displays a larger enrichment of USF2 under MEM relative to RPMI growth conditions compare to *QTRT1* that is an MITF but not a MYC target.

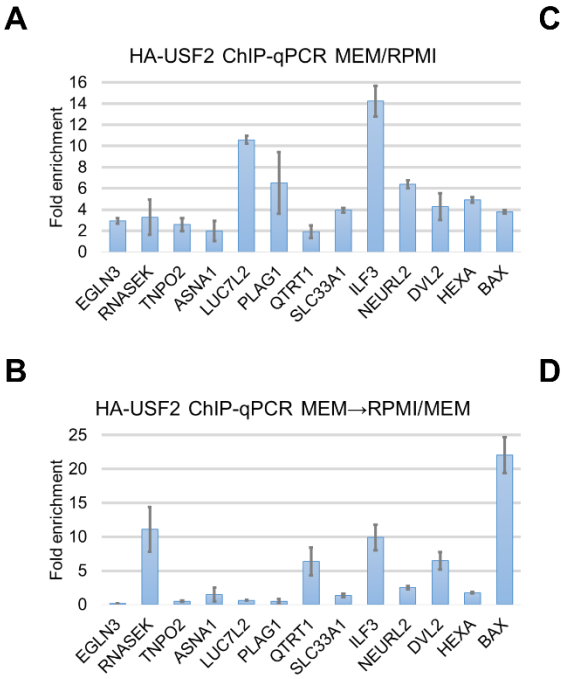
In support of our hypothesis, *ILF3* that displayed the largest fold-enrichment ($\approx 14-$fold) of USF2 binding under MEM relative to RPMI was a relatively good target for MITF, MYC and USF. Also in agreement with our hypothesis is that a high ratio of USF2/(MITF + MYC)

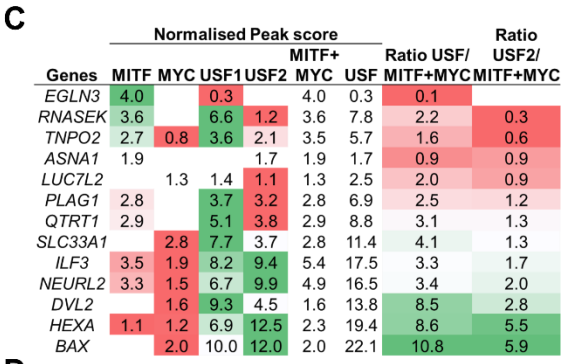
peak score correlates with smaller USF2 fold-enrichment in MEM over RPMI. These sites represent strong USF2 targets but weak MITF & MYC sites and therefore are mainly occupied by USF even under RPMI growth conditions.

As part of the promoter education hypothesis we also wished to explore the effect on target occupancy when the bHLH-TF that was taken out of the equation is reintroduced into the system. To address this we looked at the relative changes in USF2 binding at E-boxes in cells grown for a further $24\,\mathrm{h}$ in RPMI following $72\,\mathrm{h}$ growth in MEM (RPMI add-back, Figure 3.3.9B). In this scenario we would expect genes that display relatively small changes in USF2 occupancy upon MEM starvation to also display relatively small changes upon RPMI add-back. At genes which became highly enriched in USF2 following MEM treatment, depending on the relative affinity of USF2 to the E-box compared to MITF and/or MYC, we can expect to see a different kinetic of USF2 displacement by MITF and/or MYC. The time-point of $24\,\mathrm{h}$ after PRMI add-back was arbitrarily chosen as our lab's unpublished observations indicated that MITF protein levels are restored by this time point (data not shown).

Our ChIP-qPCR results however, did not display the anticipated outcome, except that weak USF2 targets with a relatively low ratio of USF2/(MITF + MYC) peak score did not change much when comparing MEM to RPMI and also did not change much following RPMI add-back. *EGLN3*, *TNPO2*, *ASNA1*, *LUC7L2*, *PLAG1* and *SLC33A1* (relatively low USF2 normalised peak score) as well as *NEURL2* and *HEXA* (relatively high USF2 normalised peak score) all display lower than 2.5 –fold differences between RPMI add-back and the sample under MEM treatment. Genes which showed > 6 –fold enrichment of USF2 ChIP signal under RPMI add-back conditions relative to MEM are *RNASEK*, *QTRT1*, *ILF3*, *DVL2* and *BAX*. There appears to be no clear correlation between the genes where the USF2 ChIP signal under RPMI add-back was enriched relative to MEM and whether these E-boxes are MYC and/or MITF targets.

As MYC is known to be highly responsive to growth factors in the serum supplement, we also performed a ChIP-qPCR analysis of HA-USF1, HA-USF2 and FLAG-MYC grown in either RPMI + 10 % FBS or RPMI + 0.1 % FBS (serum-starved) for 24 h (Figure 3.3.8D). In all cases, we observed an enrichment in USF1 and USF2 and a decrease in MYC at MYC and USF co-occupied sites when comparing to the normalised fold-enrichment of serum-starved cells against cells grown in RPMI + 10 % FBS. As the experiment was performed using different cell lines, as well as our observation that FLAG-MYC and HA-USF1 stable cell lines display elevated level of USF2 protein (Figure 3.1.1A), we do not intend to interpret





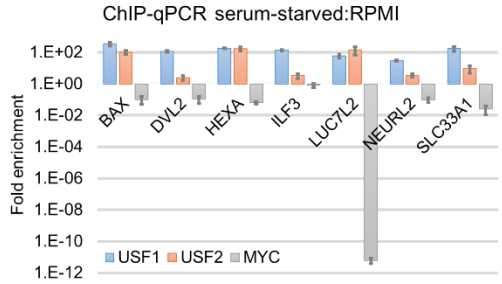


Figure 3.3.9 ChIP-qPCR of 501mel cells under starvation. **A)** 501mel cell line stably expressing HA-USF2 was subjected to growth in RPMI or MEM for 72 h before samples were collected for ChIP of HA-USF2 and qPCR at the indicated targets. The genes are ranked in the order according to the ratio of USF2 over combined MITF + MYC peak score calculated in **C**. The fold-enrichment is presented as the ratio of normalised $\Delta\Delta C_t$ of cells grown in MEM against cells grown in RPMI for 72 h. **B)** Same as **A** but showing the ratio of ChIP-qPCR signal following growth in MEM for 72 h before grown for a further 24 h in RPMI relative to cells grown for 72 h in MEM. **C)** Table showing the normalised peak scores at the loci studied by qPCR in **A)** and **B)**. As we were trying to look for a shift from MITF and/or MYC to USF, we combined the peak score of MITF + MYC and USF1 + USF2 then used these combined values to calculate the ratio of USF1 + USF2 combined peak score over combined MITF + MYC peak score. As amino-acid starvation may affect USF1 protein levels and the fact that we were ChIPping for USF2, we also calculated the ratio of USF2 peak score over combined MITF + MYC peak score and use this to sorted the order of genes presented in **A** and **B**. The table is colour-coded in two ways. First, each row of peak score from each bHLH-LZ-TF against each gene are colour-coded such that bHLH-TF with the highest score are highlighted in dark green while lowest score are highlight in dark red. Second, the last two columns are colour-coded such that genes with the lowest ratio of USF over MITF + MYC combined peak score are in darkest red while the one with highest ratio are in darkest green. **D)** ChIP-qPCR of 501mel cells stably expressing either HA-USF1, HA-USF2 or FLAG-MYC subjected to serum starvation (0.1 %) for 24 h before samples collected for ChIP using either HA- or FLAG-antibodies followed by qPCR analysis at the indicated targets. The fold-enrichments are presented as the ratio of normalised $\Delta\Delta C_t$ for serum-starved cells relative to cells grown in RPMI.

these result beyond the observation that at MYC and USF co-occupied sites, we observe an increase in USF and a decreased in MYC ChIP-signal.

Approaches to interfere with bHLH-TFs' function

We explored various means to study different stages of promoter education by USF1, USF2 and MITF. Since, the exact role of each factor exchanging at the promoter remained elusive, we attempted various methods including: 4-hydroxy-tamoxifen (4 – OHT)-inducible hormone binding domain of the oestrogen receptor (ER) fused to dominant-negative (dn) proteins, adenovirus containing dn-protein, siRNA, sh-RNA, and epigenetic silencing using dsRNA (Monika Gullerova, personal communication) and genome editing. Since we had very little success with sh-RNA and epigenetic silencing, we focussed on the other approaches in this report.

USF knock-out

We previously discussed the different phenotype of USF1 and USF2 knock-out mice reported in the literature (see section 1.8, USF and development). These reports indicated that knocking-out USF is a viable option for studying the role of USF in factor exchange and promoter education at selected genes. To this end, we adopted the recently emerging genome-editing tools, TALEN and the Cas9-CRISPR system (for review see Gaj et al., 2013). The ability of the TALEN and Cas9-CRISPR to edit the target locus is affirmed using the Surveyor assay as described in Supplementary Figure 3.3.3 before proceeding to generate knock-out cell lines.

TALEN technology utilises two sequence-specific DNA-binding proteins (TAL) fused to the Fok1 nuclease. DNA-excision and double-strand excision requires dimerization of the Fok1 nuclease, meaning that the spacing between the two TALENs is critical for DNA-cleavage. Thus, target specificity is conferred through the sequence specificity of each arm of the TALEN pair and the spacing between the two TALEN arms. We obtained TALEN from GeneArt (LifeTech) who also performed *in vitro* screening to minimize off-target editing.

Using these TALENs, we attempted to generate 501mel cell lines carrying a USF2 knock-out. As the efficiency of genome editing at all alleles of a gene by TALEN in a given cell is widely accepted to be quite low, we screened a large number of clones following selection of cells positive for integration of TALEN (emGFP-tagged). The whole cell extract from these clones were Western-blotted and probed with two-USF2 antibodies that detect different species of USF2. In a number of clones, we were able to greatly reduce the USF2 species detected by the N18-antibody and in some clones totally abolish it (exemplified by clone #67 in Figure 3.3.10A), but not the faster migrating species detected by C20-antibody.

In a complimentary approach, we also designed a number of guide-RNAs for use with the Cas9-CRISPR genome editing system against USF2. Only one of the guide-RNAs (directed against USF2 exon 2) successfully demonstrated some knock-out of USF2 as detected by Western-blot (Figure 3.3.10B). We tried this guide-RNA in both Phoenix (Φ) and 501mel cell lines and in both cases were able to see consistent knock-out of the faster migrating band of USF2 as detected by the C20-antibody, but not the species detected by the Ab-antibody. Since, we were able to confirm the functionality of our TALEN and Cas9-CRISPR *in vitro*

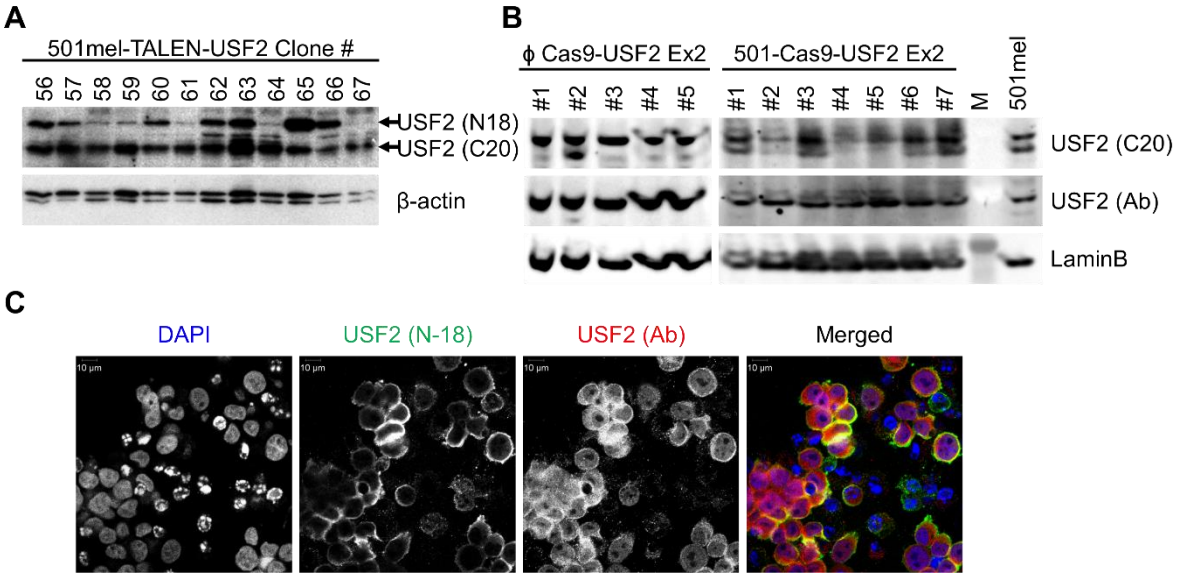


Figure 3.3.10 Genome editing of USF2 by the TALEN and Cas9-CRISPR approaches. **A)** Western-blot of whole cell extract of clonally isolated 501mel cell lines selected for stable expression of TALEN directed against USF2. The numbers above each lane indicated the clone number of the isolated lines. The blot was dually probed with two USF2 antibodies (N18 and C20) which detect different species of USF2. β–ACTIN was used as a loading control. **B)** Western-blot of whole cell extract of clonally isolated Phoenix (Φ) or 501mel cell lines selected for stable expression of Cas9-CRISPR directed against USF2 exon-2 (Ex2). The numbers above each lane indicated the clone number of the isolated lines. The blot was probed with two USF2 antibodies (C20 and Ab) which detect different species of USF2. LaminB was used as a loading control. **C)** Immunofluorescence of 501mel cell line positively selected for TALEN directed against USF2, stained with DAPI, USF2 (N18 and Ab).

using the Surveyor assay, but consistently failed to generate cell lines in which knock-out of all USF2-species was achieved, we suspected that knocking-out USF2 may be inviable in the cell lines we used. Moreover, during the process of USF2 knock-out cell line generation using TALEN, we saw a large proportion of emGFP-positive cells rounded up and detached. This observation persisted even after the cells had been grown under selection for a number of weeks. As genome-editing by TALEN is inefficient, we suspected that the detached cells were cells which were homozygous USF2 knock-outs. To further investigate this, we performed immunofluorescence of the polyclonal population of emGFP-positive cells and stained for USF2. Remarkably, cells which were double negative for USF2 staining by both N18- and Ab-antibodies were cells which rounded up and had nuclear foci characteristic of apoptotic cells (Figure 3.3.10C). This indicated that knock-out of USF2 is inviable in 501mel cells, hence we never observed clones which completely abolished all USF2-species. This most likely is due to the asymmetric regulation of USF2 over USF1 as knocking-down USF2 (using siRNA) also knocks-down USF1 at the protein level. We have tried a number of siRNAs to ensure that the observed USF1 knock-down is not an off-target effect of the siUSF2 used (data not shown). [Although based on mRNA-levels (Figure 3.2.1B), we already know that this could not have been the case]. This asymmetric regulation was observed in all melanoma cell lines we tested but not in the Phoenix cell line (data not shown). Interestingly, contrary to the published report, our collaborator found that USF2 knock-out mice die at an early stage and present a high level of iron in the skin (Marie-Dominique Galibert, personal communication). Since iron interferes with the stress response, even if we are able to establish a USF2 knock-out cell line from the embryos of these mice, it would still complicate our study of factor exchange as USF is the effector of a number of stress response pathways (Corre and Galibert, 2005).

A-USF (dominant-negative USF)

In the light of the involvement of USF in activation of a diverse range of genes, and the difficulty in getting a satisfactory level of knock-down using sh-RNA, even after applying

selection for cells carrying the short hairpin constucts, and the reduced viability of USF2 knock-out cells we concluded that constitutive interference of USF function may confer growth disadvantages. Therefore, we adopted an inducible dominant-negative system. To this end, we established stable cell lines expressing a 4 −OHT inducible dominant-negative USF (ER-A-USF). The A-USF contains the HLH-LZ domain of USF1 that allows dimerization with USF1 and USF2 but the basic region is replaced with a 12 −amino-acid acidic poly-glutamic acid stretch (A-USF) (Figure 3.3.11A). Inducibility is achieved through coupling of A-USF to the 4 −OHT responsive oestrogen receptor ligand-binding domain (ER). This results in increased dimer stability and abolished DNA-binding (Krylov et al., 1997). A-USF is able to form a more stable dimer with the endogenous USF than the native USFs themselves (Krylov et al., 1997), thus A-USF can efficiently titrate USF away from their targets (Qyang et al., 1999). The major disadvantage of this approach is that A-USF targets both USF1 and USF2.

The ER of ER-A-USF covers the ligand-binding domain of the oestrogen receptor containing the $\mathrm{G525R}$ mutation that increases its responsiveness to $4-$OHT. In the absence of 4 −OHT, the protein is rapidly degraded, but when ER-A-USF binds to 4 −OHT, it re-localises to the nucleus (Figure 3.3.11B). This protects it from the cytosolic proteasome and results in nuclear accumulation. This mean of induction also ensures that under non-induced conditions there should not be any basal-effect of A-USF as it should be cytosolic. As a proof of principle, shown here in Figure 3.3.11C are 501mel cells transiently transfected with HA-A-USF for $24\ \mathrm{h}$ prior to staining. In the HA-A-USF positive cells, we see depletion of nuclear USF2 (as detected by the N18-antibody). As the majority of HA-A-USF are cytosolic, this indicates dimerization of HA-A-USF with USF2. These positive cells also demonstrated decreased level of TBX3 protein, a novel USF downstream target we identified (data not shown). To validate the inducibility of ER-A-USF, we transiently transfected 501mel cells with FLAG-ER-A-USF $\pm 4-$ OHT. ER-A-USF is largely cytoplasmic and is only present as background staining without 4 −OHT (Figure 3.3.11D).

Following $16\,\mathrm{h}$ $4-\mathrm{OHT}$ induction, FLAG-ER-A-USF accumulates in the nucleus as expected.

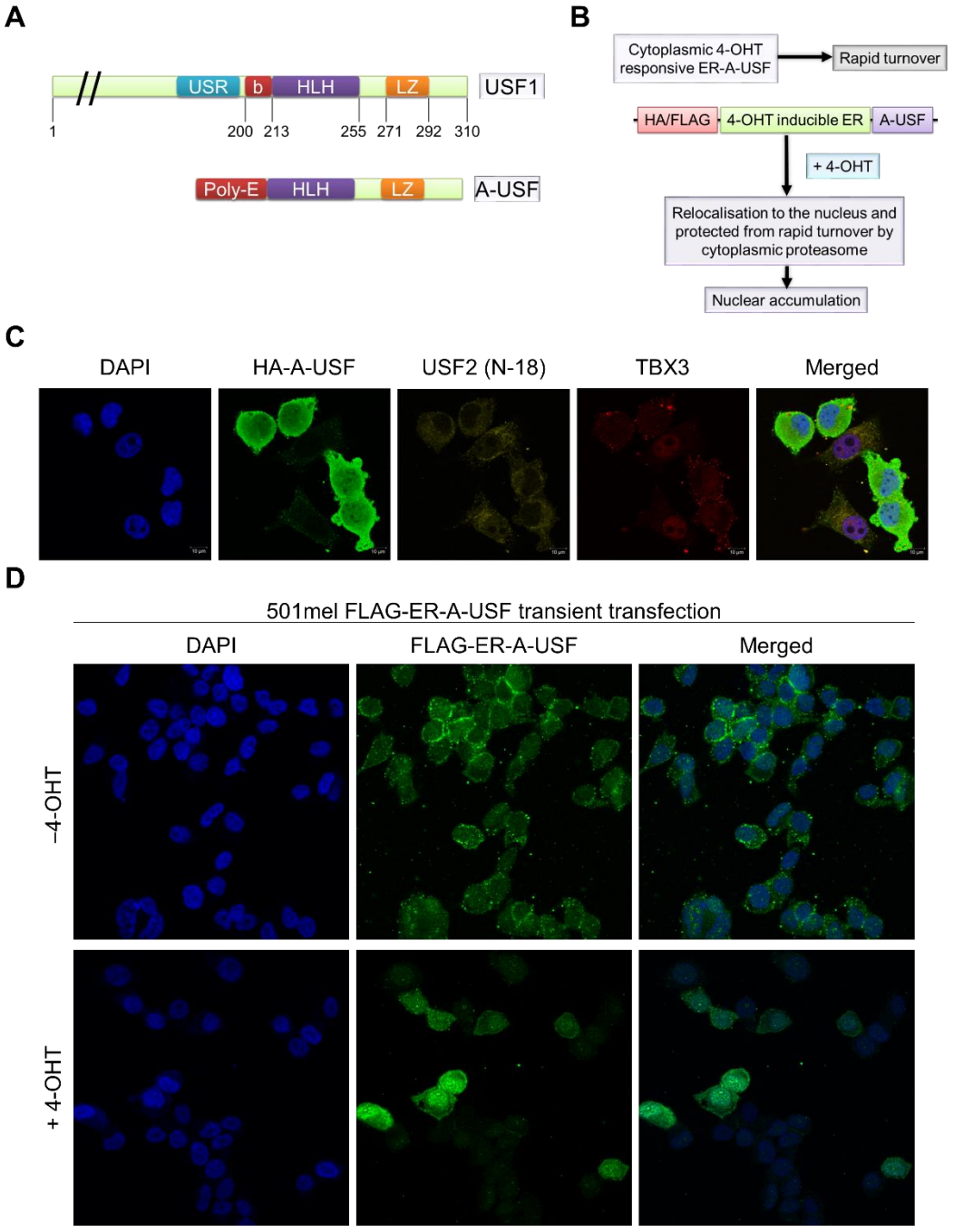


Figure 3.3.11 Generation of inducible dominant negative USF (A-USF). **A)** Schematic diagram of USF1 and dominant negative USF (A-USF). The domain of USF1 are indicated and the amino-acid residue number at the boundary of the domain are indicated under the diagram. USR (USF specific region); b (basic); HLH (Helix-loop-Helix) and LZ (leucine zipper). A-USF is generated by fusing USF1 HLH-to-C-terminal domain to the acidic stretch, poly-glutamic acid (Poly-E) in place of the basic domain. **B)** Schematic diagram of how the inducibility of $4-\mathrm{OHT}$ responsive, truncated oestrogen-receptor fusion of A-USF work. In the non-induced state, ER-A-USF is largely cytoplasmic and therefore, are rapidly degraded by cytoplasmic proteasome. Induction by 4 –hydroxyl–tamoxifen $(4-\mathrm{OHT})$ result in nuclear relocalisation of ER-A-USF to the nuclear which protected it from proteasome mediated degration. **C)** Immunofluorescence of 501mel cells transiently transfected with HA-tagged A-USF and stained with the indicated antibody. **D)** IF of 501mel cells transiently transfected with FLAG-tagged-ER-A-USF $\pm\,300\,\mathrm{ng/ml}$ $4-\mathrm{OHT}$ and stain for DAPI and FLAG.

Having validated the functionality and inducibility of ER-A-USF, we went on to create stable cell lines containing HA/FLAG-ER-A-USF in 501mel and performed time course induction to demonstrate that overnight induction $(16\ \mathrm{h})$ resulted in high level of accumulation of A-USF for the selected clone (Figure 3.3.12A). We then tested the functionality of our inducible system by looking at the effect on TBX3. Following $16\ \mathrm{h}$ of ER-A-USF, we see by immunofluorescence a noticeable drop in TBX3 expression, while there are no significant changes in USF2 levels (Figure 3.3.12B). This observation matches well with our results on Western-blot (Figure 3.3.12C) showing a decrease in TBX3 levels as the total amount of A-USF accumulates. (although we did not include $\mathrm{t} = 0$ in the blot, this should not influence our interpretation as we are directly comparing $+\ \mathrm{vs} - 4$ −OHT.) However, we observed a small but noticeable accumulation of TBX3 in the parental 501mel treated with 4 −OHT after $4\ \mathrm{h}$ treatment (Figure 3.3.12D). Moreover, as we extended our observation to other cell lines, we were struck by the observation that TBX3 was induced by overnight 4 −OHT treatment (Figure 3.3.12E) in primary melanoma (IGR37), metastatic melanoma (IGR39) as well as non-pigmented (Phoenix) cell lines. This was a very important observation with diverse medical implications. TBX3 is a well characterized anti-senescence (Carlson et al., 2001; Hoogaars et al., 2008) and pro-metastatic factor (Rodriguez et al., 2008). Thus, TBX3 induction by 4 −OHT must be taken into account when formulating a medical prognosis for the use of tamoxifen in ER+ breast cancers. Tamoxifen may compromise the effect of other combined treatments by enhancing senescence bypass and potentially inducing metastasis of breast tumours. Equally important is the potential effect of TBX3 induction by 4 −OHT in research using ER-inducible systems such as: ER-N-Ras in oncogenic studies (Talotta et al., 2008; Tarutani et al., 2003) and ER-Myc in maintenance and anti-senescence of murine neural crest stem cells JoMa1.3 cell line (Maurer et al., 2007). An important question is, how much of the reported effects are attributed to TBX3 induced anti-senescence? These are key questions which are beyond the scope of this thesis but do need to be addressed.

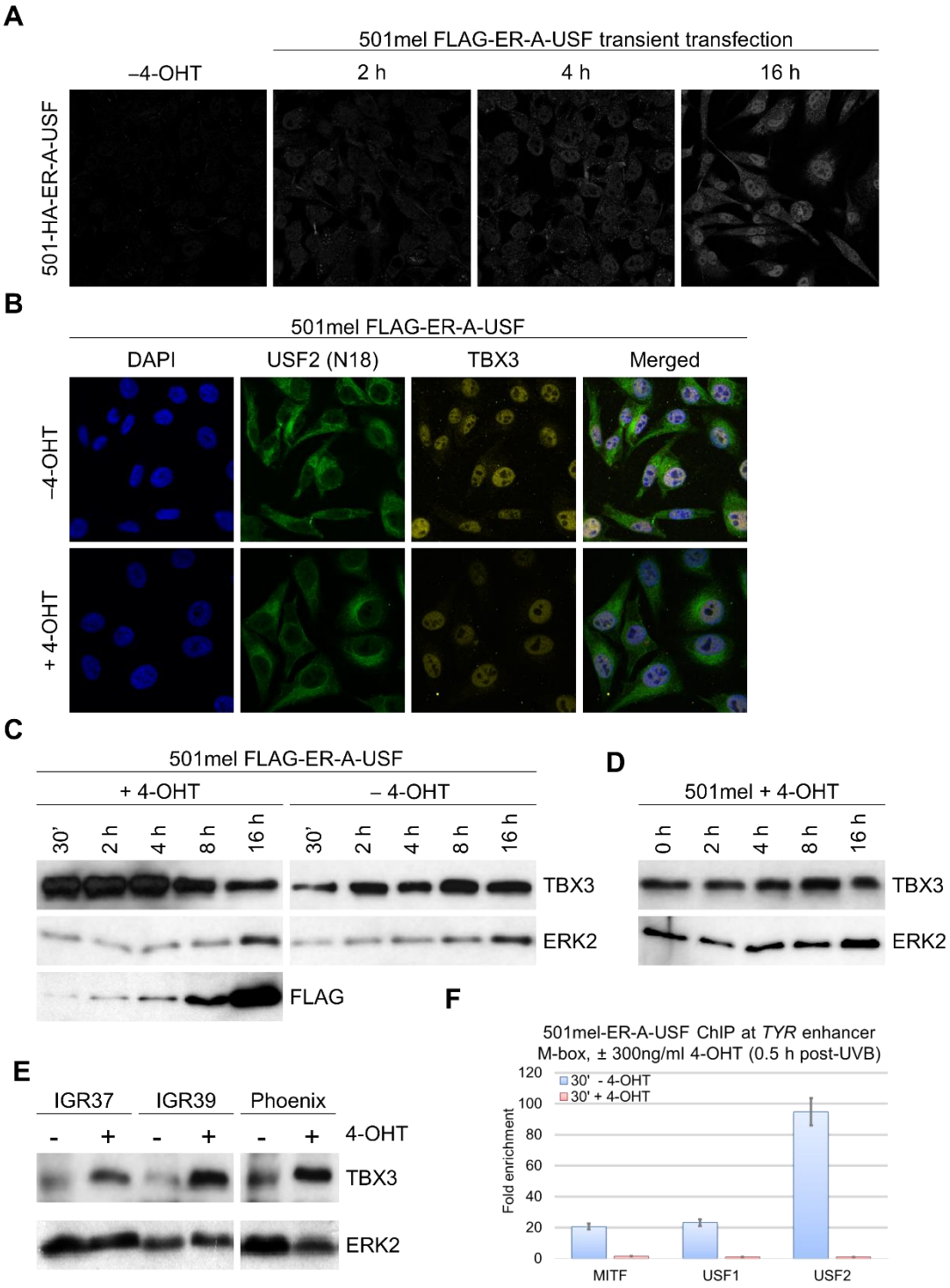


Figure 3.3.12 Cell line stably expressing ER-A-USF and the effect of Tamoxifen on TBX3 expression. 501mel clonal cell lines stably expressing 4 –hydroxyl–tamoxifen ($4-\mathrm{OHT}$)-inducible dominant negative USF (ER-A-USF) was generated as described in the Materials and Methods and inducibility and functionality was tested. **A)** Immunofluorescence of one of the 501mel cell lines stably expressing HA-ER-A -USF following time course of $4-\mathrm{OHT}$ addition. The samples were probed with anti-HA antibody. **B)** IF of one of the 501mel stable cell lines expressing FLAG-ER-A -USF $\pm\, 4-\mathrm{OHT}\ 16\ \mathrm{h}$. The samples were stained using the indicated antibodies. **C)** Western-blots following time course of $4-\mathrm{OHT}$ induction, showing the accumulation of FLAG-ER-A-USF as detected by anti-FLAG antibody. The treated cells were probed for the changes in TBX3 using ERK-2 as a loading control. **D)** Same as **C** for the parental 501mel cell. **E)** Western-blots showing the effect of $\pm\, 4{-}\mathrm{OHT}$ overnight on the level of TBX3 in IGR37, IGR39 and Phoenix using ERK-2 as a loading control. **F)** ChIP-qPCR of 501mel cell line stable expressing ER-A-USF $\pm\, 4-\mathrm{OHT}$ overnight prior to treatment with $200\ \mathrm{J/m^2}$ UVB and sample collected $30'$post UVB-treatment and subjected to ChIP-qPCR at *TYR* enhancer M-box using the indicated antibodies. The fold-enrichments are presented as the ratio of normalised $\Delta\Delta C_t$ of cells treated with UVB at $\mathrm{t} = 30'$ relative to $\mathrm{t} = 0$.

Using an ER-A-USF stable cell line, we performed ChIP at the *TYR* enhancer after $200\ \mathrm{J/m^2}$ UVB following overnight incubation $\pm 300\ \mathrm{ng/ml}\ 4-$ OHT (Figure 3.3.12F). The fold enrichment is presented relative to $\mathrm{t} = 0$. Addition of $4-$OHT decreases the MITF ChIP signal following $30'$ UVB-treatment by $\approx 14-$fold, USF1 by $\approx 23-$fold and USF2 by $\approx 100-$fold relative to $-\ 4-$OHT. The accompanying reduction in MITF ChIP-signal following UVB-induction in the presence of A-USF supports our promoter education hypothesis where coordinate exchange of USF and MITF might be required for stable binding of MITF. This notion is further supported by our ChIP in the parental 501mel following UVB- (Figure 3.3.3D) and MTX-induction (Figure 3.3.5F) where concomitant increase of MITF and USF at the *TYR* enhancer was observed following induction. In the ER-A-USF cell line $+\ 4-$OHT, interfering with USF binding also interferes with MITF binding to the *TYR* enhancer following UVB-induction. The differences in the observed ratio of MITF, USF1 and USF2 ChIP fold-enrichment following UVB induction is probably due to the slight difference in the phenotype of the clonal cell when compare to the parental heterogeneous population used in previous *TYR* induction experiment.

Brief summary of Chapter 3.3

Central to the promoter education hypothesis is the coordinate exchange of bHLH-TFs at shared E-box elements. USF, which has no strong transcription activation domain and was reported to be involved in maintenance of bivalent marks as well as establishment of active histone modifications, was suspected to play the role of the so called “promoter education factor”. Analysis of the publicly available data in HepG2 correlates USF binding with active histone marks and H2A.Z globally, but not histone modifications associated with poised enhancers.

Through ChIP-qPCR we demonstrated increased M-box binding by USF and MITF at our model gene, *TYR* following induction by UVB and MTX. The kinetic of recruitment appeared to be cell line specific that may be dependent on the basal expression level of USF and MITF as well as the inducibility of MITF in each cell line. As it is physically impossible for

an M-box to be occupied by USF and MITF simultaneously, USF and MITF must exchange following induction. It remains to be seen whether the observed increase USF and MITF ChIP-signals reflect USF and MITF co-ordinately exchanging to activate the *TYR* gene or the average of different populations of cells with either USF or MITF bound to the *TYR* M-box, in the case where activation of *TYR* can be achieved by either USF or MITF. Co-ordinate exchange of MITF and USF is more likely, at least at the *TYR* enhancer because, induction of ER-A-USF prior to UVB induction significantly reduced the USF and MITF ChIP-signal even though A-USF only interferes with DNA-binding by USF.

We also investigated factor exchange through conditions which decrease MITF (and by extrapolation MYC) protein levels, namely amino-acid and serum starvation. At all targets co-occupied by USF and MYC, serum-starvation resulted in an increase in USF occupancy at the expense of MYC. ChIP-qPCR at a number of USF2, MYC and MITF co-occupied site demonstrated the highest increase in the USF2 ChIP-signal following starvation at sites at which USF2, MYC and MITF all display a high normalised peak score. Sites which are weak USF2 targets relative to MITF and/or MYC, however, did not demonstrate a significant increase in the USF2 ChIP-signal following starvation. At the other extreme, sites that are strong USF2 but weak MITF and MYC targets did not show a significant increase in the USF2 ChIP-signal following amino-acid starvation. Sites that have a relatively higher normalised peak score for USF1 over USF2 also demonstrate a relatively small increase in the USF2 ChIP-signal following starvation. This, most likely, is because, reduction in MITF and MYC, although freeing-up the E-boxes, may favour USF1 homodimer binding over USF2. Lastly, we found that sites with higher normalised peak scores for MYC than MITF demonstrate greater enrichment of USF2 following starvation.

3.4 Regulation of USF2 subcellular localisation

USF2 is cytoplasmically localised

Sub-cellular localization has been shown to regulate the nuclear activity of many TFs. For example, the bHLH-TF, Mash-2, is cytoplasmically localized in cytotrophoblasts (Jiang and Mendelson, 2003). USF1 and USF2, on the other hand, are regarded as nuclear TFs. However, we noticed a significant amount of cytoplasmic USF2 staining with the N18-antibody when we switched to the staining protocol (described in Materials and Methods) compatible with phalloidin staining, while the C20-antibody consistently gave nuclear staining in both protocols as did the USF1-antibody (Figure 3.4.1A) in 501mel cells. Phalloidin, a phallotoxin that binds at the interface between F-actin monomers, is commonly used to stain the actin-cytoskeletal. Many of the staining protocols that focus on nuclear proteins tend to include a cytoskeletal extraction step which removes most of the cytoskeletal components to significantly reduce background staining.

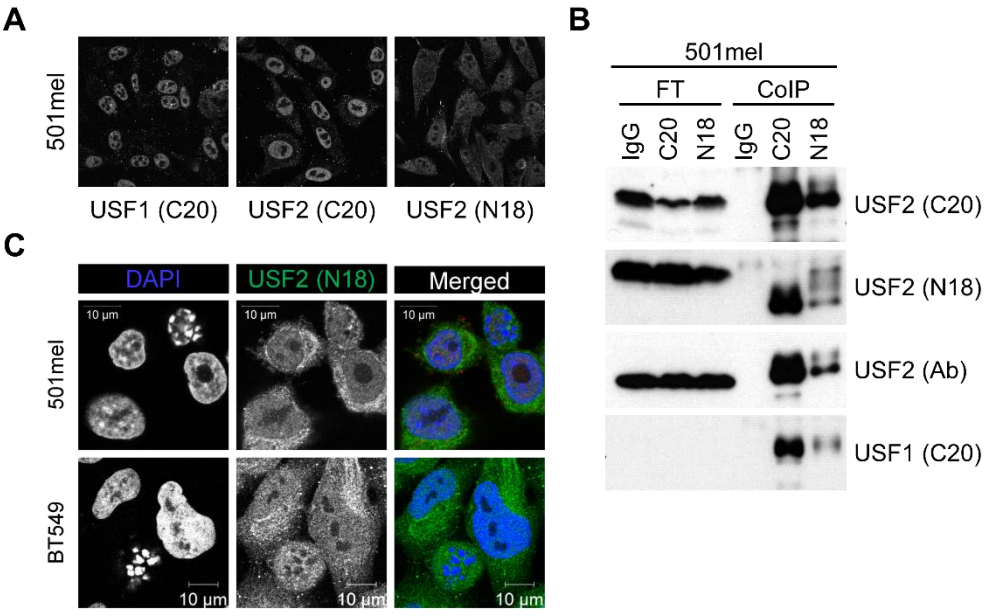


Figure 3.4.1 Staining and CoIP using USF2 antibodies. **A)** Immunofluorescence of 501mel cells using the indicated antibodies. **B)** Co-immunoprecipitation (CoIP) of 501mel cell extract with either IgG, USF2 (C20) or USF2 (N18) antibodies, Western-blotted and probed with the indicated antibodies. FT (flow-through) indicated the lysate after the antibodies have been purified with DynaBeads. CoIP indicates the antibody used for the immunoprecipitation. **C)** Immunofluorescence of 501mel and BT549 cells stained with USF2 (N18) antibody and DAPI.

The most straightforward explanation for the observed cytoplasmic localisation of USF2 would be that the N18-antibody is detecting cytoplasmic proteins other than USF2. This argument seemed logical in the light of the observed different migration of the USF2 band detected by the C20- and the N18-antibodies. The N18-antibody targets the N-terminal 18 amino-acids of USF2 while the C20-antibody targets the 20-amino-acid at the C-terminus. The C20-antibody gave an ≈ 44 kDa band while the N18 detects an ≈ 55 kDa band (Figure 3.3.10A). Since USF2A isoform predicted size is ≈ 44 kDa, this may argue against the specificity of the N18-antibody. However, a large number of our observations argued against the N18-antibody being non-specific. Firstly, we performed all ChIP-qPCR with both the C20- and N18-antibodies, though in all cases we only report the N18-data as these consistently gave a higher fold-enrichment (section 3.3). Secondly, CoIP of endogenous USF2 with either the C20- or N18-antibodies immunoprecipitated USF2 species that were detected by other antibodies (Figure 3.4.1B). Moreover, both the C20- and N18-antibodies immunoprecipiated USF1 as well as USF2, as detected using a third antibody (Ab). What was puzzling, however, was the different migration pattern of the USF2 species. The N18-antibody detects the faster migrating species of USF2 (probably corresponding to the C20-specific species) in the samples CoIPed with C20-antibody but not vice-versa. A possible explanation is the involvement of labile PTMs that interfere with the binding by one of the USF2-antibodies but not the other, such that they are mutually exclusive but the PTMs may be lost during the CoIP steps. Alternatively, novel alternative splicing may be involved which generates USF2 species that can interact but lack part of USF2 recognised by each antibody. To address this, we looked at the transcript assembly from our RNA-seq data and performed RT-PCR using a number of primer pairs targeting the 5'- and 3'-ends of the USF2 reference mRNA but were unable to identify any novel alternative first and last exon usage by USF2 (data not shown). Lastly, (in collaboration with the Kessler lab, TDI, Oxford) we performed Mass-spec analysis of the N18 and C20-specific bands from the IP samples. In both cases, we were able to confirm that the C20-recognised band is USF2 in both N18- and C20-immunoprecipitates (data not shown). The amino-acid sequence of the N18-

recognised band, however, did not match the canonical USF2A. However, at the time of the experiment, the Kessler lab lacked the algorithms for *de novo* peptide sequencing (without reference database) therefore, we were not able to identify the exact amino-acids sequence of the protein.

The most compelling evidence is that we were able to knock-out the N18-specific band through a genome editing approach (Figure 3.3.10A & C). Immunofluorescence in various cell lines using N18-antibody also support the argument we made in section 3.3, USF knock-out that apoptosis may be a consequence rather than a causal of knocking-out all USF2 species using TALEN as USF2, as detected by the N18-antibody, persists in apoptotic cells as exemplified in 501mel and BT549 cells with an apoptotic phenotype (Figure 3.4.1C).

Nuclear-cytoplasmic shuttling of USF2

As alluded to above, factor exchange at a shared DNA-binding element may also be regulated at the level of availability and ability of each factor to bind. Our data suggested that USF2 can be cytoplasmic localised. This is not entirely surprising as some TFs such as STATs (Kerr et al., 2003; Williams et al., 2004) (re-localised from the plasma membrane to the nucleus following recruitment from the cytoplasm and activation by JAK) and p53 (re-localised from the cytoplasm to the nucleus) (Liang and Clarke, 1999) and transcription co-factors such as β-CATENIN (relocalised from the plasma membrane to the nucleus following WNT signalling) (Piedra et al., 2001) are sequestered away from their target genes and only enter the nucleus in response to specific cues. We, therefore, wished to explore the regulatory mechanism involved in USF2 subcellular re-localisation. The mechanistic insight of USF2 subcellular re-localisation will be invaluable to the understanding of the regulatory mechanisms involved in controlling the availability of USF2 during transcription activation and memory. Moreover, it will provide us with a means to explore the effect of inducing USF2 on factor exchange and promoter education.

Over-expressed, heterologous USF2 is mainly nuclear localised

It came to our attention that the majority of over-expressed USF2 is nuclear. We verified that the N18-antibody recognises HA-epitope-tagged-USF2A (HA-USF2) expressed in 501mel cells. A Western-blot of the extract was dually probed with anti-HA and the N18-antibody and imaged on an Odyssey infra-red imager (data not shown). The HA-USF2 band picked up by the anti-HA antibody was also recognised by the N18-antibody, confirming that N18-antibody detects USF2A. To eliminate the potential cross-reactivity of N18-antibody with other cytoplasmic proteins displaying sequence related to USF2 epitope, we verified by BLAST that the region of USF2 targeted by the N18-antibody does not exhibit any sequence homology to known proteins in the database (as of July, 2014).

We further performed immunofluorescence on 501mel cells transiently ectopically expressing HA-USF2-FLAG dual-epitopes-tagged protein and probed with anti-HA, anti-USF2 (N18) and anti-FLAG antibodies (Figure 3.4.2A). The result indicated that ectopic USF2 recognised by the three different antibodies has the same nuclear profile. This nuclear staining was not observed with endogenous USF2 in 501mel (Figure 3.4.1A) but was, nevertheless, detected by the N18-antibody. Therefore, the N18-antibody recognises USF2 protein encoded by the cDNA. As a final confirmation, we generated USF2 deletion mutants lacking either the N-terminal 20 −amino-acids or the C-terminal 20 −amino-acids and performed transient transfection assays in retinal-pigmented epithelial (RPE) cells stably expressing GFP-actin and probed with HA- and either C20- or N18-antibody to check the epitope specificity of the two antibodies (Figure 3.4.2B). The C20-antibody did not pick up the HA-USF2ΔC20 while the N18-antibody did not pick up the HA-USF2ΔN20, respectively. RPE-GFP-actin was used as we were interested to see whether these USF2-mutants co-localised with actin.

In a complementary approach to confirm the cytoplasmic localisation of endogenous USF2, we sub-fractionated 501mel cells into nuclear and cytoplasmic fractions and probed for MITF, USF1 and USF2 (Figure 3.4.2C). This approach should eliminate any potential

artefacts, such as fixation conditions, generated during slide preparation for immunofluorescence. This sub-fractionation also revealed the majority of USF2, as detected by the N-18 antibody, to be cytosolic, while MITF and USF1 are almost completely nuclear as expected. LaminB was used as a nuclear loading control to show that there is no contamination from the nuclear fraction in the cytoplasmic fraction. This increased our confidence that USF2 may be regulated at the level of nuclear-cytoplasmic shuttling.

The observed nuclear localisation of exogenous, over-expressed USF2 suggested that endogenous USF2 may be actively sequestered in the cytoplasm by (an)other protein(s) or

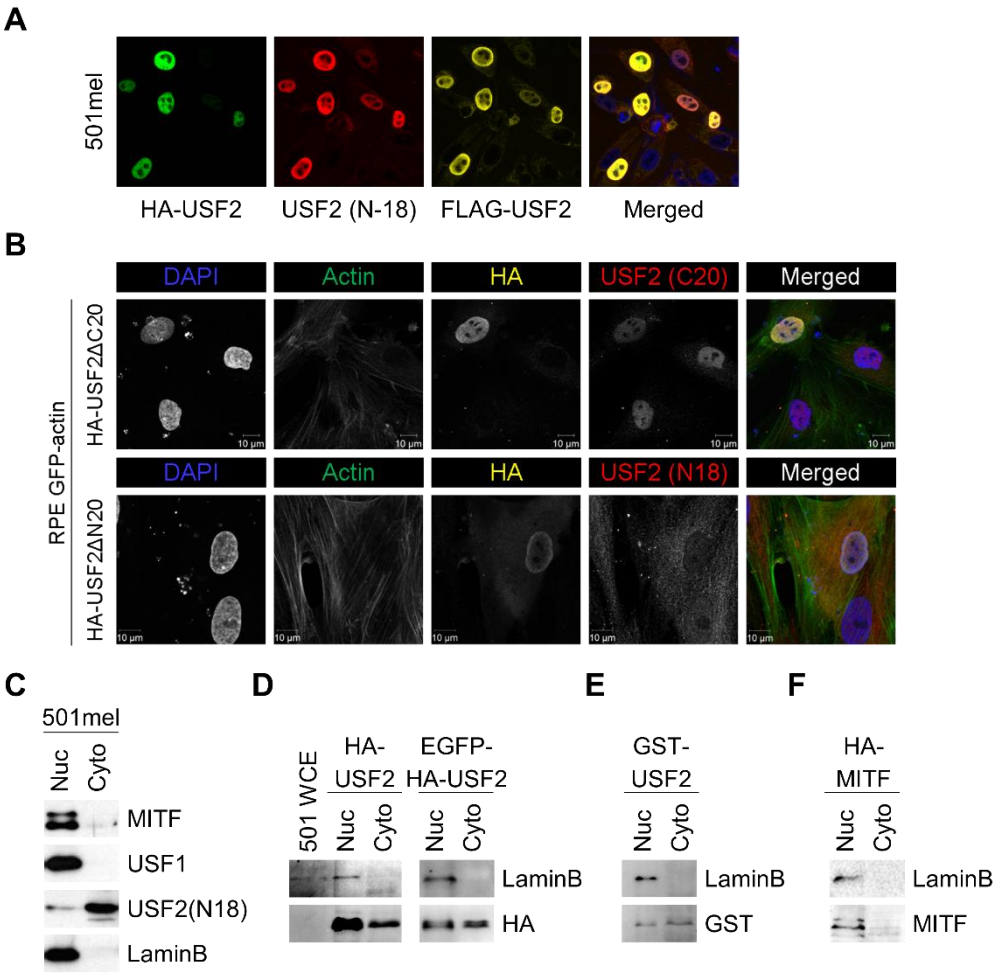


Figure 3.4.2 Subcellular-localisation of ectopic and endogenous USF2 by immunofluorescence or subcellular fractionation. **A)** Immunofluorescence of 501mel cells transiently transfected with HA-USF2-FLAG dual epitope-tags and stained with either anti-HA-, USF2 (N18)- or FLAG-antibodies as indicated. **B)** Immunofluorescence of retinal pigmented epithelia (RPE) stably expressing GFP-actin transfected with vectors expressing USF with either a 20-amino-acid deletion at the C-terminus or the N-terminus (either HA-USF2ΔC20 or HA-USF2ΔN20, respectively) and probed with either anti-HA or anti-USF2 (C20 or N18) antibodies. **C)** Subcellular fractionation of 501mel cells into nuclear and cytoplasmic fractions (Nuc and Cyto, respectively). **D-F)** Western blot of sub-cellular fractionation of 501mel cell lines stably expressing HA-USF2 and EGFP-USF2 (**D**), GST-USF2 (**E**) and HA-MITF (**F**).

efficiently exported out of the nucleus. In transient over-expression, we may saturate the protein(s) required for cytoplasmic localisation, resulting in more nuclear USF2.

To test this hypothesis, we generated stable cell lines expressing HA-USF2 or EGFP-HA-USF2 under the control of the *CMV*-promoter (Figure 3.4.2D) and GST-USF2 under the control of 5'LTR of Moloney murine leukemia virus (*MMLV*) (Figure 3.4.2E). The latter is a relatively weak promoter compared to *CMV*, allowing a stable, low-level expression. We utilised three different tags to reduce the chance of fusion-induced conformational changes on USF2 influencing the sub-cellular localisation of ectopic USF2. In stable cell lines, the expression level of USF2 is much lower than transient over-expression and may not saturate the mechanism(s) involved in cytoplasmic retention of USF2. Indeed, sub-cellular fractionation of these cell lines reveals a substantial amount of cytoplasmic ectopic USF2 (Figure 3.4.2D – E). Moreover, HA-USF2 cells (Figure 3.4.2D) which express a much higher level of USF2 have a higher ratio of ectopic nuclear to cytoplasmic USF2 when compared to other stable cell lines, further supporting the system saturation hypothesis. It is worth noting that this saturation phenomenon is not observed with another bHLH-LZ-TF, MITF as demonstrated in 501mel HA-MITF (Figure 3.4.2F), where endogenous and ectopic MITF are nuclear. This suggests that retention to a specific sub-cellular compartment is not a phenomenon common to all bHLH-LZ-TFs, at least in 501mel cells.

Induction of USF2 nuclear accumulation

We explored a diverse range of induction cues as well as activating kinases in an attempt to promote a change in subcellular localisation of USF2. However with the exception of glucose stimulation, nothing shifted USF2 (as detected by the N18-antibody) into the nucleus. Therefore, we shall only list them here without supplying the results (Table 3.4.1).

Regulation of USF through nutrients

Ample evidence implicates USF1 and USF2 in glucose (Vallet et al., 1998; Vallet et al., 1997) and fatty acid metabolism (Casado et al., 1999; Pajukanta et al., 2004; Wang and

Sul, 1995), and that both DNA binding and protein levels of USF1 and USF2 are regulated in response to these metabolites (Decaux et al., 1991; Shi et al., 2008; Weigert et al., 2004). Although, there had been no report relating metabolic regulation of USF2 to the sub-cellular localisation of USF2, we deemed this highly likely, because USF2 must re-localise to the nucleus to function as a TF, be it at glucose-responsive or stress-responsive genes.

Table 3.4.1 Table listing experiments performed in an attempted to induce USF2 (as detected by N18-antibody) nuclear localisation

Compounds/Cues/ Activating kinase	**What it does?**	**Rationale**
CDK5 & p35	p35 is activated by Calpain during aberrant cell cycle activity and DNA-damage and function together with CDK5	USF has been reported to be a target of calpain in undifferentiated murine erythroleukemia (Lin et al., 2009), leaving behind dimerization domain which acts as dominant-negative (Watt and Molloy, 1993).
LATS1	Negative regulator of YAP1. Important in organ size control and suppress tumourigenesis through restrict proliferation and promotes apoptosis.	USFs act as regulators of cell cycle and proliferative genes (Corre and Galibert, 2005). USF suppresses tumour formation (Luo and Sawadogo, 1996a; Choe et al., 2005). *USF2*$^{-/-}$ mice are smaller at birth and display proportionate body features throughout postnatal development (Sirito et al., 1998).
AMPK	AMPK plays pivotal role in energy homeostasis, stimulates catabolism	USF1 regulates hepatic gluconeogenesis through activation of an inhibitor of gluconeogenesis genes, small heterodimer partner, in response to AMPK and DNA-PK (Chanda et al., 2009).
p38	Stress activated MAPK	USFs are known to be a versatile stress-responsive TF (Corre and Galibert, 2005).
α-MSH analogue	Activates the cAMP response, induces p38 and pigmentation	USF is required for expression of many pigmentation related genes (Corre and Galibert, 2005).
UVB	Activates MKK6 and p38	Refer to section 1.8, p38 and USF
Tunicamycin	ER-stress inducer	SREBP which cooperate with USF to regulate a number of metabolic gene, is exported from the ER before being cleaved in the Golgi and released into the nucleus (Osborne, 2000; Shao and Espenshade, 2012; Shimomura et al., 1998; Sul et al., 2000; Wong and Sul, 2010). Tunicamycin has been used to induce SREBP activation. As USF2 sometimes appears to localise to the ER/Golgi like organelles, we tried this.
Oleic acid/ CD-lipid (LifeTech #11905)/ Cholesterol (LifeTech #12531 – 018)	Lipid/Choresterol supplement	USF is linked to the familial combined hyperlipidaemia (FCHL) (Pajukanta et al., 1998; Pajukanta et al., 2004) and dyslipidaemia (Naukkarinen et al., 2005; Hoffstedt et al., 2005). USF regulates some 40 genes involved in lipid metabolism (van Deursen et al., 2009; Casado et al., 1999; Naukkarinen et al., 2006; Vallet et al., 1998)

		USF was shown to be the major hepatic nuclear component that bind to the ChoRE (Vaulont et al., 1989). Thus, addition of cholesterol or fatty acid that induce/inhibit, ChREBP (Kawaguchi et al., 2002) may result in shifted of USF localisation as USF might exchange with ChREBP.
Essential amino-acid /Serum starvation ± Pyruvate/Lactate	Nutrient starvation	Essential amino-acids starvation results in MITF depletion while serum starvation result in MYC depletion. If USF exchange with MYC and MITF at share promoters, conditions which decrease MITF and MYC may induce USF2 nuclear localisation. Pyruvate and lactate are key regulator of TCA and as USF is involved in regulation of a number of metabolic genes, conditions which induce switching in metabolic profile may induce changes in USF2 localisation.
Nicotinamide	Inhibits 3',5'-cyclic-AMP phosphodiesterase. PARP inhibitor, HDAC inhibitor	Sirtuin has been implicated in a wide range of biological process including transcription, cell fate, DNA-repair, inflammation and bioenergetics as it requires NAD for lysine deacetylation. USF regulates a number of these processes (refer to section 1.8). During fasting, glucagon activates cAMP response that inhibits USF mediated glucose responsive gene expression (Bergot et al., 1992).
Torin1	Inhibits mTOR, a key positive regulator of biosynthetic pathway and inhibitor of autophagy pathway.	Torin1 increases nuclear accumulation of TFEB, a highly related bHLH-LZ-TF. Moreover, USF is involved in a number of biosynthetic pathways including the anabolic pathway towards fatty-acids catabolism (Casado et al., 1999; Latasa et al., 2003; Wang and Sul, 1997).
Chloroquine	Inhibits protein degradation through lysosomal pathway resulting in suppression of autophagy.	links to the Torin1 experiment. If USF exchanges with TFEB at autophagy genes between starvation and feeding, we may see nuclear-cytoplasmic shuttling of USF2 following the starve-feed cycle.
Metformin	Anti-diabetic drug. Inhibits gluconeogenesis, at least in part through activation of LKB1-AMPK pathway	USF has been linked to metabolic syndrome (Langefeld et al., 2004; Shoulders and Naoumova, 2004) and type-2 diabetes and glucose intolerance in various populations (Elbein et al., 1999; Hanson et al., 1998; Ng et al., 2005; Putt et al., 2004; Wiltshire et al., 2001).
Insulin	Promotes glucose metabolism	Insulin negatively regulated USF heterodimer DNA binding (Nowak et al., 2005)
GSK3 inhibitor	GSK3 is a key regulator of glycogen metabolism and mediates signalling from a number of pathway including WNT and insulin.	Our lab (unpublished data) has demonstrated increased nuclear accumulation of MITF following GSK3-inhibitor treatment.
1% O_2, 16 h or DMOG	Hypoxic induction/ Inhibits PHD containing proteins	USF directly interact with the HIF-signalling pathway to regulate a number of hypoxic genes (Refer to section 1.9, USF and HIF factor exchange)

Glucose stimulation of USF2

High extracellular glucose enhances expression of USF1 and/or USF2 (Weigert et al., 2004) and was associated with increased DNA-binding activity and transactivation potential in a manner dependent on PKC, ERK and p38 (Bidder et al., 2002; Wang et al., 2004; Weigert et al., 2004). At least in HepG2 cells, there appeared to be more USF under high-glucose conditions (van Deursen et al., 2008). Some studies, however, reported differential regulation of USF1 (Bidder et al., 2002) or USF2 (Wang et al., 2004) alone by glucose. This difference may reflect cell-type specific responses to glucose or alteration in specific signalling cascade in some cell lines. Indeed supplementing RPMI media to $4.5\mathrm{g/l}$ glucose (hyperglycemia) did not result in any changes in USF protein level in a number of melanoma cell lines (data not shown).

Glucose induction of USF1 and USF2 protein levels can be further augmented by high fatty acids (van Deursen et al., 2008). This induction, DNA-binding and subsequent transcription activation is apparently cell type and species-dependent. For example, in HepG2 cells, glucose increases predominantly USF1 but not USF2, and activation of downstream hepatic lipase is only diminished by silencing *USF1* and not *USF2*. Similarly, USF1 has a very low basal protein level and could be induced to high levels in response to glucose, at least in rat liver while the level of USF2 remained unchanged (Wang and Sul, 1995). As with glucose induction, we did not see any accumulation of USF1 and USF2 in conjunction with chemically-defined lipids, oleic acid and/or cholesterol (data not shown).

In primary rat mesangial cells, glucose induces USF2 expression through CREB (Shi et al., 2008) in a time and dose-dependent manner. This induction was unaffected by cyclohexamide or actinomycin D, indicating an unchanged rate of translation and mRNA stability. In mouse liver, on the other hand, glucose induces both USF1 and USF2 that predominantly act as a heterodimer (Vallet et al., 1998). In the case of asymmetric induction, changes in the ratio of USF1 to USF2 can induce a switch from USF1:USF2 heterodimers to USF homodimers.

Considering the large number of observations implicating glucose in the induction of USF, we reasoned our lack of observed changes in USF protein level to be melanoma specific. Unlike liver, melanocytes are not a primary glucose-responsive cell-type, and it is entirely possible that pigmented cells are not as responsive to glucose induction. In the light of our observed cytoplasmic localisation of USF2, we hypothesised that glucose might induce nuclear accumulation of USF2 and not the overall USF protein level.

We carried out a time course experiment to analyse USF2 localisation following glucose-starvation, or on addition of glucose following glucose starvation (glucose add-back) (Figure 3.4.3). To eliminate any potential confounding effects dependent on cell lines, we first analysed the sub-cellular localization of USF2 across three melanoma cell lines (IGR39, 501mel, IGR37) following the time course of $2\ g/l$ glucose add-back after $24\ h$ glucose starvation as outlined in the scheme in Figure 3.4.3A, i. Cell pellets were divided into nuclear and cytoplasmic fractions and analysed by Western-blotting. β-ACTIN was used as a loading control for the cytoplasmic fraction Figure 3.4.3A, ii and LaminB was used as a loading control for the nuclear fraction Figure 3.4.3A, iii. LaminB and β-ACTIN were also used to assess the quality of fractionation. As we did not see any bands on Western-blot in the cytoplasmic and nuclear fraction, respectively, these data were omitted.

In IGR39 cells, USF2 was exclusively localised to the cytoplasm following overnight glucose starvation. Following glucose add-back, we observed an increase in nuclear USF2 at $2\ h$ which persisted throughout the time-course we studied (Figure 3.4.3A, iii). This increase coincided with a decrease in the faster migrating band of cytoplasmic USF2 at $2\ h$ (Figure 3.4.3A, ii). At $4\ h$ post glucose add-back, there was an accumulation of the faster migrating form of USF2 in the cytoplasmic fraction (Figure 3.4.3A, ii). This was neither accompanied by a decrease in nuclear USF2 (Figure 3.4.3A, iii) nor a decrease in the slower migrating form of USF2 in the cytoplasm, indicating that the total level of USF2 may have increased. There appeared to be a slight reduction in total level of USF2 in both nuclear and cytoplasmic fractions at $6\ h$ when compare to $4\ h$.

A similar pattern of changes in the level of nuclear and cytoplasmic USF2 was observed in both 501mel and IGR37 cells, potentially reflecting a common regulatory mechanism that diverts USF2 away from its target metabolic genes by cytoplasmic sequestration. There were, however, some differences that probably reflected cell line-specific characteristics. For example, in 501mel at 6 h post-glucose add-back the total level of cytoplasmic USF2 also decreased, as in IGR39, albeit much more dramatically (Figure 3.4.3A, ii). Moreover, unlike the IGR39 samples where either the slower and the faster migrating forms of USF2 were not well resolved or nuclear USF2 in IGR39 was exclusively in the slower migrating isoform (Figure 3.4.3A, ii), we could clearly resolve the two forms of USF2 in 501mel and IGR37. In general, glucose add-back following 24 h glucose-starvation initially led to a

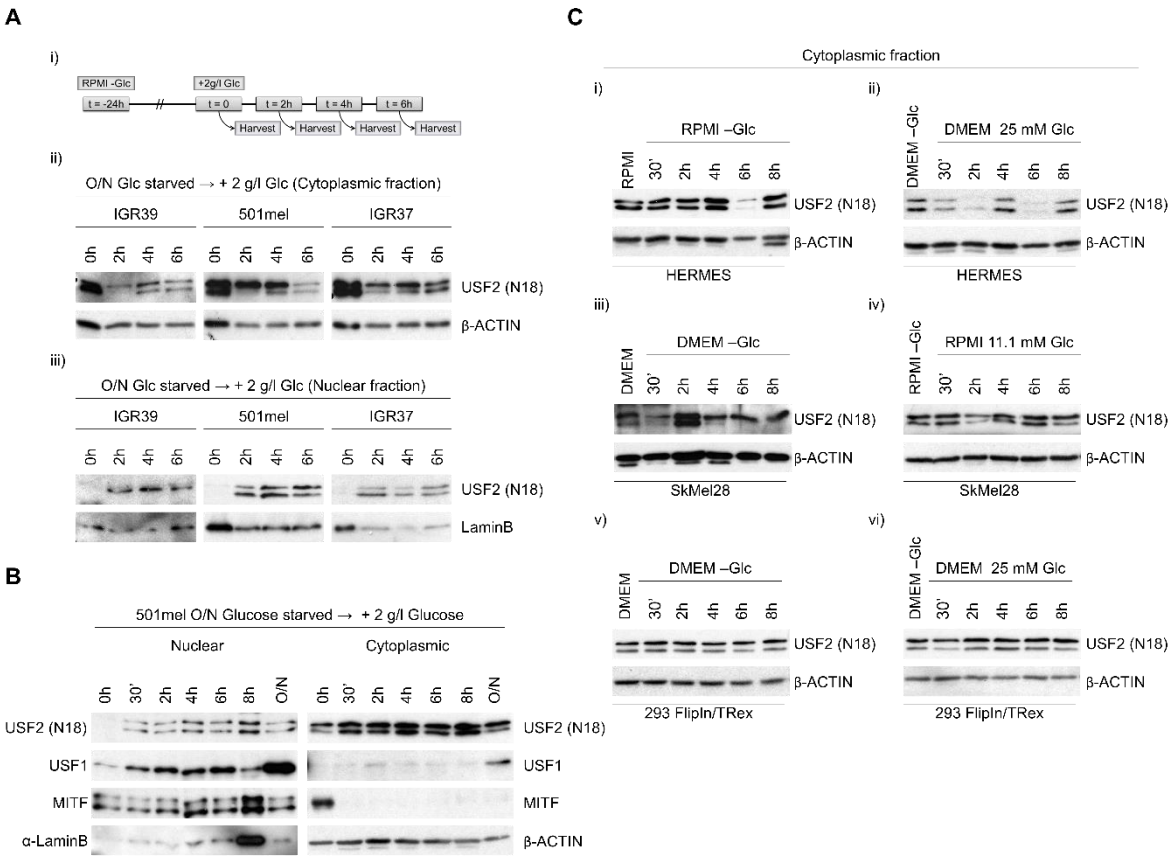


Figure 3.4.3 Glucose regulation of USF2. **A) i**) Schematic representation of the time course experiment carried out for glucose add-back experiment. Three melanoma cell lines IGR37, 501mel and IGR39 were subjected to 24 h overnight glucose starvation (O/N Glc starved) before 2 g/l glucose is added back. Cells were harvested at the time points indicated and subject to sub-cellular fractionation into cytoplasmic (**ii**)) and nucleoplasmic (**iii**)) fractions and the western blot was probed with the indicated antibodies. **B)** Time course glucose add-back experiment in 501mel following 12 h overnight glucose starvation. Cells were harvested at the indicated time point, fractionated and the Western-blot was probed with the indicated antibodies as in **A**. **C)** Western-blot of cytoplasmic fraction of melanocytes (HERMES), melanoma (SkMel28) and a non-pigmented cell line (293 FlipIn/TRex) subjected to either glucose starvation (**i**), **iii**) and **v**) respectively) or glucose add-back following 24 h overnight glucose starvation (**ii**), **iv**) and **vi**) respectively). O/N, overnight

decrease in the faster migrating form of cytoplasmic USF2 which coincided with the increase in nuclear USF2. At 6 h post-glucose add-back, the ratio of faster-to-slower migrating forms of cytoplasmic USF2 approached one. In the nucleus, there was a minor but noticeable trend of increasing the ratio of slower-to-faster migrating form of USF2.

One question arising from the observations in Figure 3.4.3A is that if we extend the time points far enough, does USF2 leave the nucleus? To this end, we extended our time course, focusing on 501mel, to 24 h post-glucose add-back (Figure 3.4.3B). We observed similar accumulation of nuclear USF2 from a depleted state following 12 h glucose starvation to a steady level of USF2. However, we did not see exclusion of USF2 from the nucleus over the time course of the experiment (Figure 3.4.3B, O/N), indicating that glucose starvation may be an effective means for inducing cytoplasmic retention of USF2. The ratio of the faster-to-slower migrating form of cytoplasmic USF2 varies slightly over time (Figure 3.4.3A, ii). The more dramatic effect observed with the first experiment (Figure 3.4.3A) may reflect the more synchronous population of cells as they were starved for 24 h as opposed to 12 h (Figure 3.4.3B).

We note that there appeared to be an increase in nuclear USF2 at 8 h (Figure 3.4.3B, 8 h) that coincided with a decrease in USF1 and an increase in MITF. The decrease in USF1 appeared to disrupt the trend of gradual USF1 accumulation and the largely unchanged MITF protein level except from at 8 h where there is a distinct increase. The increases in USF2 and MITF were not due to unequal loading (based on LaminB as a loading control), since the ratio of between USF1, USF2 and MITF changed. Moreover, if we examine the level of β-ACTIN in the cytoplasmic fraction at 8 h, it is clear that the significant changes in the level of LaminB at 8 h cannot be attributed to sample preparation as they were prepared such that the ratio between the nuclear and cytoplasmic fractions loaded on to the gel was always equal to one (as described in Materials and Methods). That is, the amount of sample loaded for the nuclear and cytoplasmic fractions corresponds to the same number of cells. Since there is no corresponding increase in the β-ACTIN at 8 h for the cytosolic fraction,

unequal loading cannot fully account for the observed differences in LaminB. Whilst we are unable to explain this observation, it appears that there may be periodic changes in the protein level of USF1, USF2 and MITF that might be enhanced by glucose add-back following starvation. Moreover, there may be glucose-enhanced changes in nuclear LaminB structure, either due to rapid turnover or changes in PTMs that affect the detection by the LaminB antibody we used.

It is worth noting that glucose starvation also regulates MITF. Following overnight glucose starvation, there is a differentially modified form of MITF in the cytoplasmic fraction that migrates on SDS-PAGE in between the two bands observed in the nucleus (Figure 3.4.3B). Our results also indicate that there is a shift in MITF migration pattern from the usual 75 kDa band to the 45kDa and 35 kDa bands following time course of glucose starvation, most strikingly in SkMel28 cells (Supplementary Figure 3.4.1). The exact identity of the 45 kDa and 35 kDa species are yet to be elucidated, but is unlikely to reflect non-specific binding by the anti-MITF antibody as the same bands were observed independently with other anti-MITF antibodies targeting different regions of MITF. Based on the observed detection by the two antibodies, we can infer that this novel form of nuclear MITF still contains the transcription activation domain and DNA binding domain. These observations suggest that MITF may be either proteolytically cleaved or differentially translated (internal translation initiation) in response to glucose that may influence MITF target gene specificity or transcription activation capability.

To further explore the generality of USF2 regulation by glucose, we performed similar glucose add-back, as well as glucose starvation, experiments on a melanocyte cell line (HERMES) (Figure 3.4.3C, i & ii, respectively), another melanoma cell line (SkMel28) (Figure 3.4.3C, iii & iv, respectively) and a non-melanocyte lineage cell line (293 FlipIn/TRex) (Figure 3.4.3C, v & vi, respectively). Bearing in mind the apparent regulation of LaminB, we only present here the results for the cytoplasmic fraction. Based on our previous experiments (Figure 3.4.3A & B) which suggested that glucose may shift the

equilibrium toward increasing nuclear USF2, while overnight glucose starvation resulted in almost exclusively cytoplasmic USF2, we predicted that the opposite experiment, transfer into no glucose conditions should yield an increasing cytoplasmic USF2. In HERMES there appears to be a gradual accumulation of cytoplasmic USF2, until 6 h post-glucose starvation where there is a decrease in USF2 level (even after taking uneven loading into account) before increasing again at 8 h (Figure 3.4.3C, i). Without an appropriate loading control for the nuclear fraction we cannot conclusively distinguish between degradation of USF2 at 6 h and its transfer to the nuclear compartment, although base on the observed lack of changes in total USF2 protein level in 501mel, this most likely reflects transfer of USF2 to the nuclear compartment.

In the reverse experiment where we added glucose back to HERMES following 24 h overnight glucose starvation we again saw an oscillating pattern of cytosolic USF2. Oscillation occurs with a 2 h periodicity (Figure 3.4.3C, ii), starting with high cytosolic USF2 following overnight starvation. This is similar to what we observed for previous experiments (Figure 3.4.3A, ii and Figure 3.4.3B). We will need to check whether this oscillation correlates with changes in nuclear USF2 or is reflective of changes in total USF2 levels.

In agreement with our prediction that USF2 is exported from the nucleus following glucose starvation as a means to exclude USF2 binding to glucose-responsive genes, the level of cytosolic USF2 in SkMel28 cells increases up to the 2 h time-point following glucose starvation (Figure 3.4.3C, iii), then reduces to a more stable level from 4 h onwards. It is worth noting that only the slower migrating form of USF2 is retained in the cytoplasm after the 4 h time-point. The pattern of changes in cytosolic USF2 in response to glucose add-back in SkMel28 (Figure 3.4.3C, iv) is similar to what was seen earlier (Figure 3.4.3A, ii), albeit with a slightly different kinetic and less dramatic. In this case, there is only a minor decrease in the faster migrating form of USF2 between the 2 h and 4 h time-points in the cytoplasm. We believe that this is a reflection of the different cell lines used as well as the period of starvation prior to glucose add-back.

We were struck by the fact that USF2 in the 293 FlipIn/TRex cell line is unaffected by either glucose starvation (Figure 3.4.3C, v) or glucose add-back (Figure 3.4.3C, vi), at least within the time-points studied. This may reflect pigment cell-specific, glucose-regulation of USF2 or that 293 FlipIn/TRex contains mutations or oncogenes (for examples, SV40 T-antigen and adenovirus E1A) that override the glucose effect on USF2. In any cases the observed differences imply that cytoplasmic retention may act as a mechanism which regulates USF2 DNA-binding and hence factor exchange.

Defining the regions in USF2 required for sub-cellular localisation

Having established that USF2 can be cytosolic we next asked which regions of USF2 are involved in nuclear and cytoplasmic localisation. To narrow down the region involved in regulating USF2 sub-cellular localization we also made a series of N- and C-terminal deletion mutants of HA-epitope-tagged USF2 based on exon structures and functional domains and assessed the sub-cellular localisation when transiently transfected in to 501mel (Figure 3.4.4A). Extracts of transfected cells were divided into nuclear and cytoplasmic fractions, and analysed by Western-blotting and probed using anti-HA antibody. LaminB was used to assess the quality of fractionation and that there is no nuclear contamination in the cytoplasmic fraction (only one representative is shown in Figure 3.4.4A). Systematic deletion from the C-terminus reveals that USF2 containing the amino-acid sequence from the N-terminus to the end of HLH-domain are almost exclusively nuclear (Figure 3.4.4A, FL, $1-328$ & $1-291$). Mutants lacking amino-acids after the basic region and the USR are equally distributed between the nuclear and cytoplasmic fractions (Figure 3.4.4A, $1-248$ & $1-234$). This is in contrast to the published observation that USR and the basic domain are necessary and sufficient for the nuclear localization of USF2 (Luo and Sawadogo, 1996b).

The $1-192$ mutant which contains only the N-terminus up to exon -5 and lacks the reported nuclear localisation signal (NLS) located in the basic region and the USR of USF2 was found to be almost entirely cytoplasmic as expected (Figure 3.4.4A, $1-192$) based on

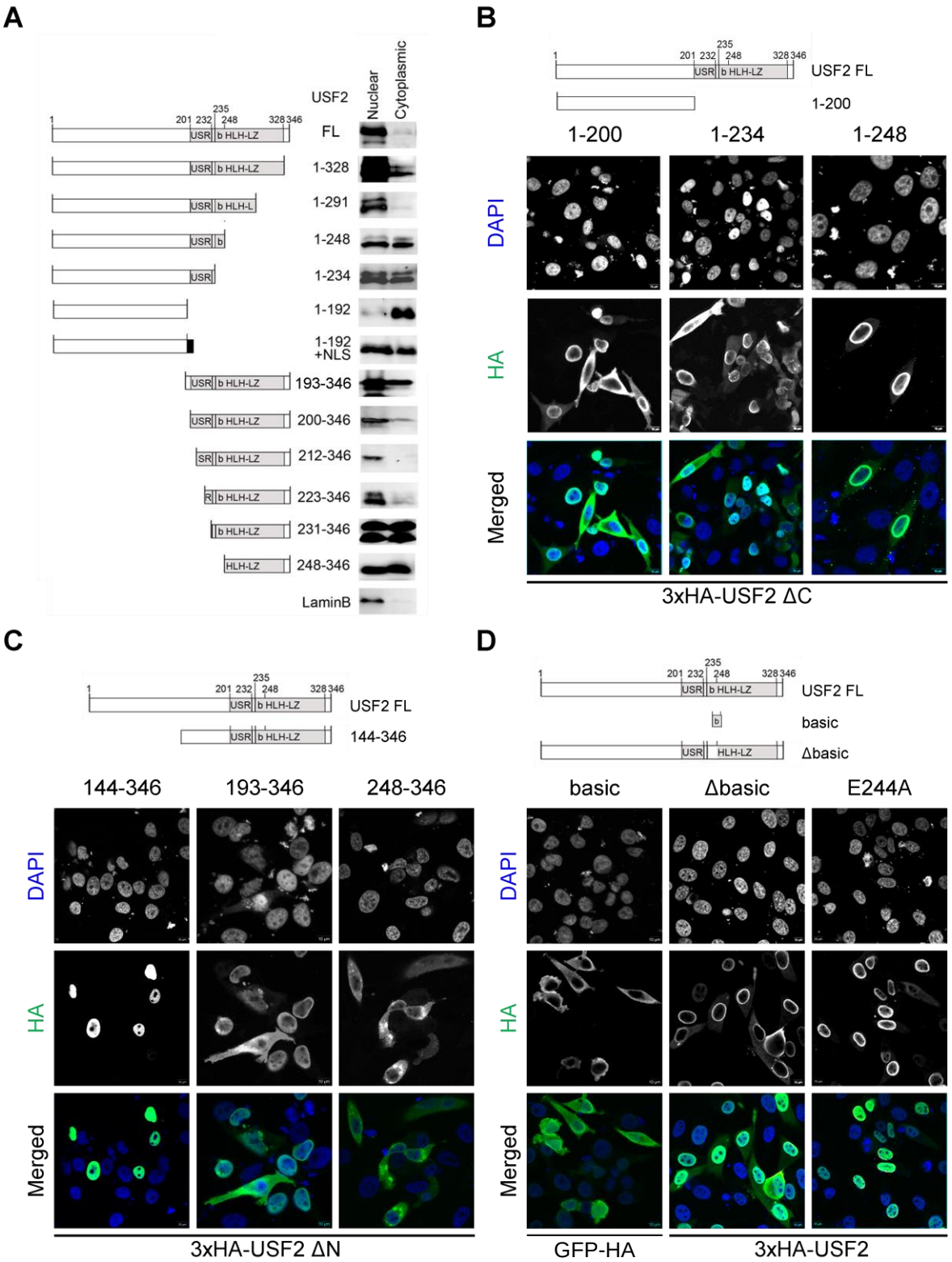


Figure 3.4.4 Nuclear-cytoplasmic localisation of USF2 deletion mutants. **A)** Schematic diagram of USF2 deletion mutant used to transfect 501mel cells before cell pellets were collected and subjected to fractionation into nuclear and cytoplasmic fractions. All the constructed were HA-affinity tagged. To the right is the Western-blot of fractionated 501mel cell lysate into nuclear and cytoplasmic fractions and probed with HA-antibody. The quality of the fractionation was assessed using LaminB to ensure no contamination of nuclear fraction into the cytoplasmic fraction. **B)** Representative immunofluorescence of 501mel cells transfected with the indicated construct from the C-terminal deletion series (refer to Supplementary Figure 3.4.2 for more). **C)** Same as B for N-terminal deletion series (refer to Supplementary Figure 3.4.3 for more). **D)** Same as **B** for USF2 basic region, USF2 lacking the basic region (Δbasic) and DNA-binding mutant form of USF2 (E244A). Note that the basic region is fused the GFP as the basic domain alone is too small to be an efficient open reading frame.

the absence of the reported NLS. In the light of a recent study reporting the NPC cut-off size of $40 - 60$ kDa (Ma et al., 2012) or about 5 nm across [although the exact size will depend on the hydrophobicity of the molecule (Kapon et al., 2010)], we can infer that the $1 - 192$ mutant is below the cut-off size and is expected to freely diffuse into the nucleus. The absence of $1 - 192$ mutant in the nucleus suggest that this mutant is actively exported from the nucleus and/or that it is efficiently held in the cytoplasm.

To further address the role of the N-terminus of USF2 in its sub-cellular localisation, we also sub-cloned the $1 - 192$ mutant into a vector containing two NLSs upstream of the HA-epitope-tag ($1 - 192 +$ NLS). The $1 - 192 +$ NLS mutant appears in both nuclear and cytoplasmic fractions (Figure 3.4.4A, $1 - 192 +$ NLS). This suggests that amino-acids $1 - 192$ of USF2 contains a nuclear export signal (NES) or is cytoplasmic retained. In this case, the observed localisation of the $1 - 192 +$ NLS mutant would simply reflect an equilibrium between the two NLSs provided by the plasmid backbone and the NES or cytoplasmic retention signal on the $1 - 192 +$ NLS mutant.

There is a significant amount of $193 - 346$ mutant in the cytoplasmic fraction even though this mutant contains both of the putative NLSs (Figure 3.4.4A, $193 - 346$). Further deletion of eight amino-acids N-terminal of the USR result in a mutant that was almost exclusively nuclear as expected (Figure 3.4.4A, $200 - 346$). This suggested that amino-acids $193 - 200$ may be implicated in negative regulation of USF2 nuclear localisation.

The $212 - 346$ and the $223 - 346$ mutants that include part of the USR to the C-terminus of USF2 were nuclear, showing that intact USR is not required for nuclear localisation of USF2 (Figure 3.4.4A, $212 - 346$ & $223 - 346$). The $231 - 346$ mutant lacking the USR was found to be equally partitioned between the nuclear and cytoplasmic fractions (Figure 3.4.4A, $231 - 346$), indicating that in 501mel cells, the basic-region alone is insufficient for nuclear localisation.

The $248 - 346$ mutant containing the region from the HLH-domain to the C-terminus also lacking the reported NLSs within the basic region and the USR of USF2, however, was equally partitioned into the nuclear and the cytoplasmic compartments (Figure 3.4.4A, $248 - 346$). This may reflect indirect nuclear localisation through dimerization of the $248 - 346$ mutant with the endogenous USF, interaction with other factors or passive diffusion through the NPC.

In a complementary approach, we also performed immunofluorescence of the USF2 C- (Supplementary Figure 3.4.2) and N-terminal deletion mutants (Supplementary Figure 3.4.3). Shown in Figure 3.4.4B & C are some of the examples. The immunofluorescence data are in good agreement with the subcellular fractionation although the exact ratio between the two compartments may differ. For example the USF2 mutant containing amino-acids up to the basic region was found to be equally distributed between nuclear and cytoplasmic fraction by subcellular fractionation but appeared mainly nuclear by immunofluorescence (Figure 3.4.4B, $1 - 248$). Immunofluorescence data for the N-terminal deletion mutant reveal nuclear localisation of the $144 - 346$ mutant, which contains both the USR- and the basic-domain and 59 amino-acids N-terminal of the USR (Figure 3.4.4C, $144 - 346$) while the $193 - 346$ mutant and the $248 - 346$ mutant distribute between both the nuclear and the cytoplasm (Figure 3.4.4C, $193 - 346$ & $248 - 346$), in agreement with subcellular-fractionation. Taken together, it appears that the 59 amino-acids N-terminal of the USR (corresponding to exon-5) act to negate the negative effect of the 8 amino-acids N-terminal of the USR on efficient nuclear accumulation of USF2.

As our fractionation and immunofluorescence data argue against the previous observation that the basic region is sufficient for USF2 localisation in most contexts while the USR and basic are sufficient for nuclear localisation of USF2 (Luo and Sawadogo, 1996b), we went on to investigate the role of the basic domain in USF2 nuclear localisation. We generated GFP-HA-tag to USF2 basic region (GFP-HA-tagged because HA-basic-domain would be too small to be an effective open-reading frame) and assessed the sub-cellular localisation

by immunofluorescence (Figure 3.4.4D, basic). In agreement with our USF2 deletion mutants, the basic domain on its own is insufficient for nuclear localisation. To further assess the importance of the basic domain in USF2 nuclear localisation, we generated a USF2 mutant lacking the basic domain and found the majority of the transfected USF2 to be nuclear localised (Figure 3.4.4D, Δbasic). As the basic-domain is involved in DNA-binding, we reason that the reported role of the basic-domain in USF2 nuclear localisation maybe related to its DNA-binding properties. To address this we generated a DNA-binding mutant of USF2 and confirmed its inability to bind to the DNA by EMSA using *TYR* enhancer probe (data not shown) before assessing its sub-cellular localisation by immunofluorescence (Figure 3.4.4D, E244A). The E244A mutant is almost exclusively nuclear, separating DNA-binding from nuclear localisation of USF2.

The data from our initial N- and C-terminal deletion mutant indicates that nuclear localisation of USF2 in 501mel is dependent on a region beyond the USR- and the basic-domain, more specifically the region encoded by exon−5 and the HLH-domain may play some role. To test this, we generate further deletion mutants, lacking both the N- and the C-terminal regions surrounding the USR and the basic domain as illustrated in Figure 3.4.5A. Using these constructs we are able to demonstrate that USR plays a more significant contribution to USF2 nuclear localisation (Figure 3.4.5B, 193 – 232) than the basic-domain (Figure 3.4.4D, basic) in 501mel cells. Although the USR- and basic-domain appear to act cooperatively to induce USF2 nuclear localisation (Figure 3.4.5B, 193 – 248), extending the construct to cover the HLH-domain result in significant cytoplasmic localisation (Figure 3.4.5B, 193 – 291). However, a protein that also includes the LZ-domain appears almost exclusively nuclear (Figure 3.4.5B, 193 – 328).

While 144 – 346 mutant appears exclusively nuclear (Figure 3.4.4C, 144 – 346), the 144 – 248 mutant has a significant amount of cytoplasmic USF2 in the highly expressing cells (Figure 3.4.5C, 144 – 248). This cannot be fully explained through passive diffusion of the small USF2 mutant out of the nucleus as the 193 – 248 mutant was largely nuclear

(Figure 3.4.5B, 193 – 248). Interestingly, extending the mutant to cover the first 13 –amino-acids of the HLH-domain significantly increased the ratio of nuclear to cytoplasmic localisation (Figure 3.4.5C, 144 – 261). This changes to exclusive nuclear localisation when the whole HLH-domain is included (Figure 3.4.5C, 144 – 291).

Taken together, nuclear localisation of USF2 requires the USR- and basic-domain. The presence of the HLH-LZ-domain increases the proportion of the proteins in the nucleus. It may be tempting to interpret these as the HLH-LZ domain enable dimerization and DNA-binding, thereby enhancing its nuclear retention; without these regions, even in the

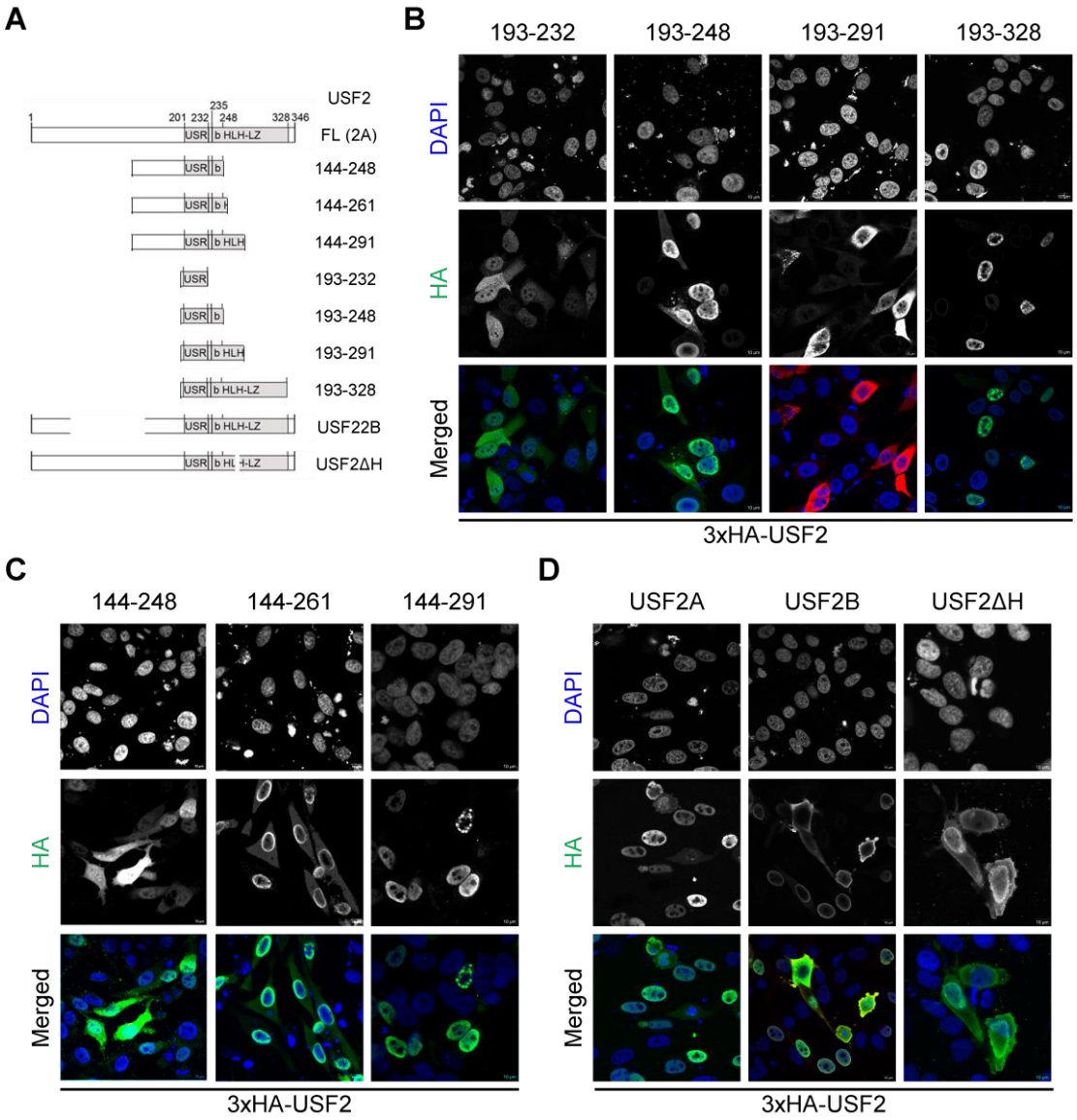


Figure 3.4.5 Immunofluorescence of USF2 deletion mutants and USF2 isoforms. **A)** Schematic diagram of USF2 deletion mutants and USF2 alternatively-spliced isoforms used to transfect 501mel cells for immunofluorescence in **B-D**. **B)** Immunofluorescence of 3 × HA-tagged to the indicated amino-acids of USF2 (starting from the USR-domain) and stained with HA-antibody. **C)** Same as **B** or USF2 mutants starting from exon−5. **D)** Immunofluorescence of 3 × HA-tagged USF2 alternatively spliced isoforms.

presence of the basic region NLS, USF2 will continually cycle between nuclear and cytoplasmic fractions, with amino-acids $1-192$ contributing to the cytoplasmic localization. However, the DNA-binding mutation would argue against nuclear retention through DNA-binding. As nuclear accumulation is enhanced by the addition of only the first $13-$amino-acids of the HLH-domain which would be insufficient for stable dimerization, it is possible that the function of the USR- and the basic-domain function as an NLS is dependent on the adjacent amino-acids which place the USR and the basic domain in an appropriate context to function as an NLS. The more complex observations with regard to exon-5 where its inclusion together with the HLH results in nuclear localisation (Figure 3.4.4C, $144-346$) while in the absence of HLH a sizeable portion of the proteins localised to the cytoplasm (Figure 3.4.5C, $144-248$), indicating potential intra-molecular masking which regulates the effect of the amino acids encoded by exon-5.

Up to this point we had been focussing on the canonical USF2A isoform. We also looked at the two other isoforms of USF2, USF2B which lacks amino-acids $77-143$ (Figure 3.4.5A, USF2B) and USF2AΔH (Figure 3.4.5A, USF2ΔH) that lacks amino-acids $275-282$ (a fourth isoform has now been reported, USF2c, lacking amino-acids $76-206$). As USF2B and USF2AΔH retain both the N- and C-terminal epitopes recognised by the USF2 N18- and C20-antibodies they cannot account for the observed differences in subcellular localisation of the USF2 species detected by the two antibodies. Nonetheless, we checked the subcellular localisation of these two isoforms out of curiosity and were surprised by the observed increased cytoplasmic localisation, especially for USF2AΔH (Figure 3.4.5D, USF2B & USF2ΔH). At least for USF2AΔH, this supports our observation that an intact HLH is required for efficient nuclear localisation of USF2.

Intramolecular interaction of USF2

As discussed above, intra-molecular interactions may be involved in regulating USF2 subcellular localisation. Intra-molecular interactions may mask sites involved in PTMs

and/or protein-protein interaction, directly influencing USF2 subcellular localisation. In support of this is the observation that while transient over-expression of HA- or FLAG-tagged USF2 are almost exclusively nuclear localised, N-terminal-EGFP- and C-terminal-mCherry-tagged USF2 contain a large proportion of cytoplasmic USF2 (Figure 3.4.6A). Large protein fusions may interfere with the ability of USF2 protein to fold on to itself, hence affecting the overall conformation. Conformation-induced changes in subcellular localisation of bHLH-LZ-TFs may not be unique to USF2 and the 501mel cells as our preliminary data also demonstrated similar observations for USF1 and MITF and to a lesser extent TFE3 in both fibroblasts and RPE cells (Supplementary Figure 3.4.4 & 3.4.5).

To study intramolecular interaction we performed GST-affinity-pull-down using either FLAG-GST-USF2 or GST-HA-USF2 as bait against the library of USF2 deletion mutants (Figure 3.4.6B). Initial pull-down experiment was performed on 3xHA-USF2 with deletion in the USR-domain and progressive deletion by exon. In some instance, we also include another construct carrying an additional V5-C-terminal tag. The pull-down demonstrated the interaction between USF2A (full-length) and all the construct with intact HLH-LZ-domain (Figure 3.4.6C, ΔUSRV5, ΔUSR, $34-346$, $144-346$V5, $144-346$, $193-346$, $223-346$, $231-346$ & $200-346$)). The only exception is the $77-346$ mutant. This may partially be attributed to the poor extractability (compare insoluble and input fraction of $77-346$ mutant with other N-terminal deletion mutant of smaller size) and relatively low efficiency of GST-pull-down (Figure 3.4.6C, $77-346$). $1-76$ and $1-143$ mutants which extend from the N-terminal up to exon-3 and exon-4, respectively but lacking the HLH-LZ do not interact with the full length GST-HA-USF2 as expected (Figure 3.4.6C, $1-76$ & $1-143$). However, extending the N-terminal region up to exon-5 results in interaction with GST-HA-USF2 (Figure 3.4.6C, $1-193$). This interaction is enhanced by adding the NLS- to the $1-193$ mutant (Figure 3.4.6C, NLS $-$ $1-193$). This may be explained by the observation that heterologous overexpressed USF2 full-length have high ratio of nuclear localisation and NLS $-1-193$ are expressed at a higher level, potentially reflecting protection from

proteasome degradation when localised to the nucleus. Importantly, NLS place NLS−1 – 193 mutant in the same sub-cellular compartment as the bait for them to interact. To narrow down the N-terminal region which interacts with full-length USF2 and to eliminate potential exclusion due to different subcellular localisation of the deletion mutants (cytoplasmic) and the full-length protein (nuclear), we repeated the experiment using USF2 mutants carrying NLSs to drive nuclear localisation (Figure 3.4.6D). GST-pull-downs using FLAG-GST-USF2 as bait against the 3xHA-NLS-proteins revealed a weak, background level of interaction with the $1 - 76$ mutant that is enhanced with the addition of exon−4 (Figure 3.4.6D, $1 - 76$ & $1 - 143$). Addition of exon−5 resulted in a decreased pull-down efficiency that can be partially explained by the reduced extraction efficiency of the sample (Figure 3.4.6D, FLAG, $1 - 193$).

We also performed the experiment on a mutant lacking the region beyond the LZ-domain (Figure 3.4.6D, $1 - 326$) and a mutant carrying only the HLH-LZ-dimerization domain (Figure 3.4.6D, $241 - 346$) as well as the full-length USF1 and USF2 (Figure 3.4.6D, FLAG + USF1 FL & USF2 FL) proteins to assess the relative interaction efficiency of the HLH-LZ dimerization domain lacking either the N-terminal or the C-terminal region of the HLH-LZ. By comparing the amount of pull-down material against the input, we can say that the $1 - 328$ mutant interacts with FLAG-GST-USF2 almost as efficiently as the full-length USF1 and USF2 (Figure 3.4.6D, $1 - 328$ vs USF1 FL & USF2 FL), indicating that the amino acids beyond the LZ-domain are not critical for the interaction. The $241 - 346$ mutant, on the other hand, displayed a greatly reduced interaction with FLAG-GST-USF2 when comparing to USF1- and USF2-full-length (Figure 3.4.6D, $241 - 346$ vs USF1 FL & USF2 FL). This suggest that amino-acids N-terminal of the HLH-LZ region play some roles in the interaction between FLAG-GST-USF2 and deletion mutants. In the light of the observed interaction between FLAG-GST-USF2 and the N-terminal region of USF2, we can conclude that amino-acids $1 - 193$, with the core at $77 - 143$, of USF2 play some roles in intrameolecular interaction within USF2.

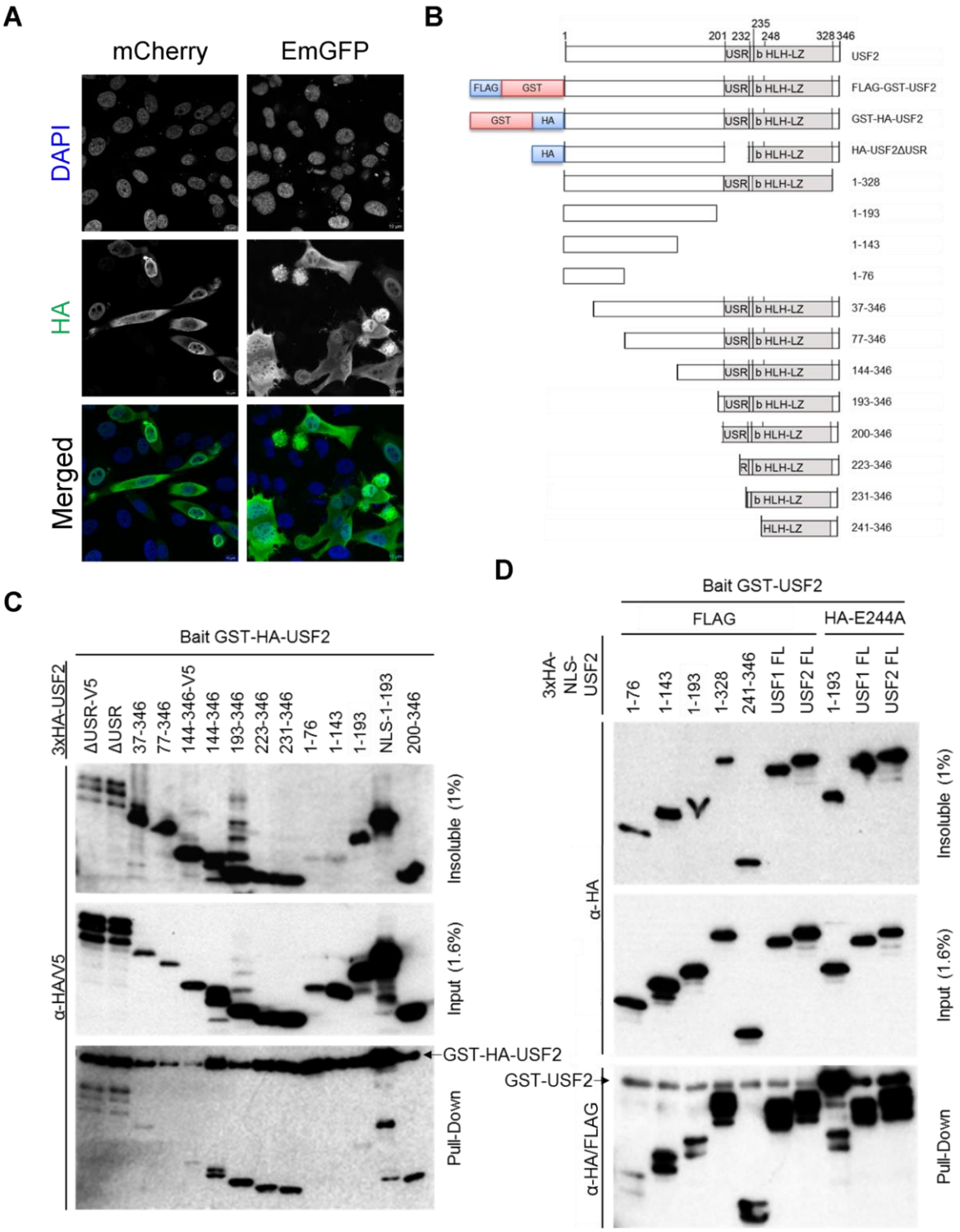


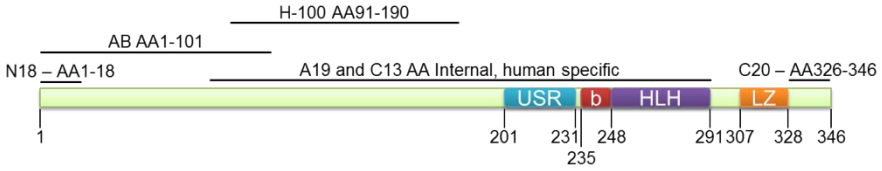
Figure 3.4.6 Intramolecular interaction and potential conformation induced changes in subcellular localisation of USF2. **A)** Immunofluorescence of C-terminal HA-mCherry and N-terminal EmGFP-HA tagged USF2A probed with HA-antibody. **B)** Schematic diagram of USF2 deletion mutants used in the GST-pull-down experiment in **C** and **D**. **C)** GST-HA-USF2 bait was used to screen against USF2 fragments, N-terminal 3 × HA-tagged or with C-terminal V5-tag. NLS, denote 3 × HA vector carrying two-nuclear localisation signals. **D)** Same as **C** but the baits were either FLAG-GST-USF2 or the DNA-binding mutant GST-HA-USF2-E244A. The target USF2 deletion mutants for the study of intramolecular interaction are 3 × HA-NLS tagged to force all the N-terminal fragment into the nucleus where the full length USF2 are mainly found.

To assess whether the interaction between amino-acids $1-193$ and the C-terminal region is dependent on DNA-binding or not, we performed GST-pull-downs using a DNA-binding mutant of USF2, GST-HA-USF2-$E244A$ as bait against the $1-193$ mutant using full-length USF1 and USF2 as a control (Figure 3.4.6D, $HA-E244A+1-193$, USF1 FL & USF2 FL). In this case, we were able to confirm that the interaction between amino-acids $1-193$ and the C-terminal region is independent of DNA-binding.

USF2 sub-cellular localisation

To test whether cytoplasmic localisation of USF2 is melanoma specific, we performed immunofluorescence on cell lines listed in Table 2.3. Based on the observed difference in the staining pattern with the C20- and N18-antibodies, we tested a number of commercially available anti-USF2 antibodies targeting different regions of USF2 to try to narrow down the region responsible for the observed difference. The summary diagram entailing antigen specificities of the antibodies used, the observed band size and migration pattern on Western-blot, the subcellular localisation between nuclear and cytoplasm as observed on Western-blot and immunofluorescence in 501mel cells can be found in Figure 3.4.7A. All the antibodies were validated by their ability to detect heterologously expressed USF2A in transient transfection assay by immunofluorescence. A19-, C13- and H100-antibodies did not detect heterologous USF2 on Western-blots. The full immunofluorescence datasets can be found in Supplementary Figure $3.4.6-10$. Here we will only briefly discuss some of the examples just to illustrate that there are some cell line specific differences in the observed pattern of USF2 localisation. In Figure 3.4.7B USF2 (N18) staining appears as filamentous structures within the cytoplasm, somewhat similar to TUBULIN and similar to ACTIN stress fibres around the edge of the cells in the breast ductal carcinoma cell line, BT549 (Figure 3.4.7B, BT549). In BT549, nuclear staining appears more uniform. In the metastatic breast adenocarcinoma, HS578T, similar uniform nuclear and filamentous cytoplasmic staining can be observed. However, USF2 can also be found localised to the leading edge of the migrating cells (Figure 3.4.7B, HS578T, red arrows) suggesting an intriguing possibi-

A



Antibodies	Bands on Western-blot	Nuclear/Cytoplasmic on Western-blot	Nuclear/Cytoplasmic on immunofluorescence
N18	≈ 55 kDa, Doublets	Cytoplasmic	Mainly Cytoplasmic
C20	≈ 44 kDa, Triplets	Nuclear	Nuclear
AB	≈ 55 kDa, Singlet	Both	Both
A19			Mainly Nuclear
C13			Mainly Nuclear
H100			Mainly Nuclear

B

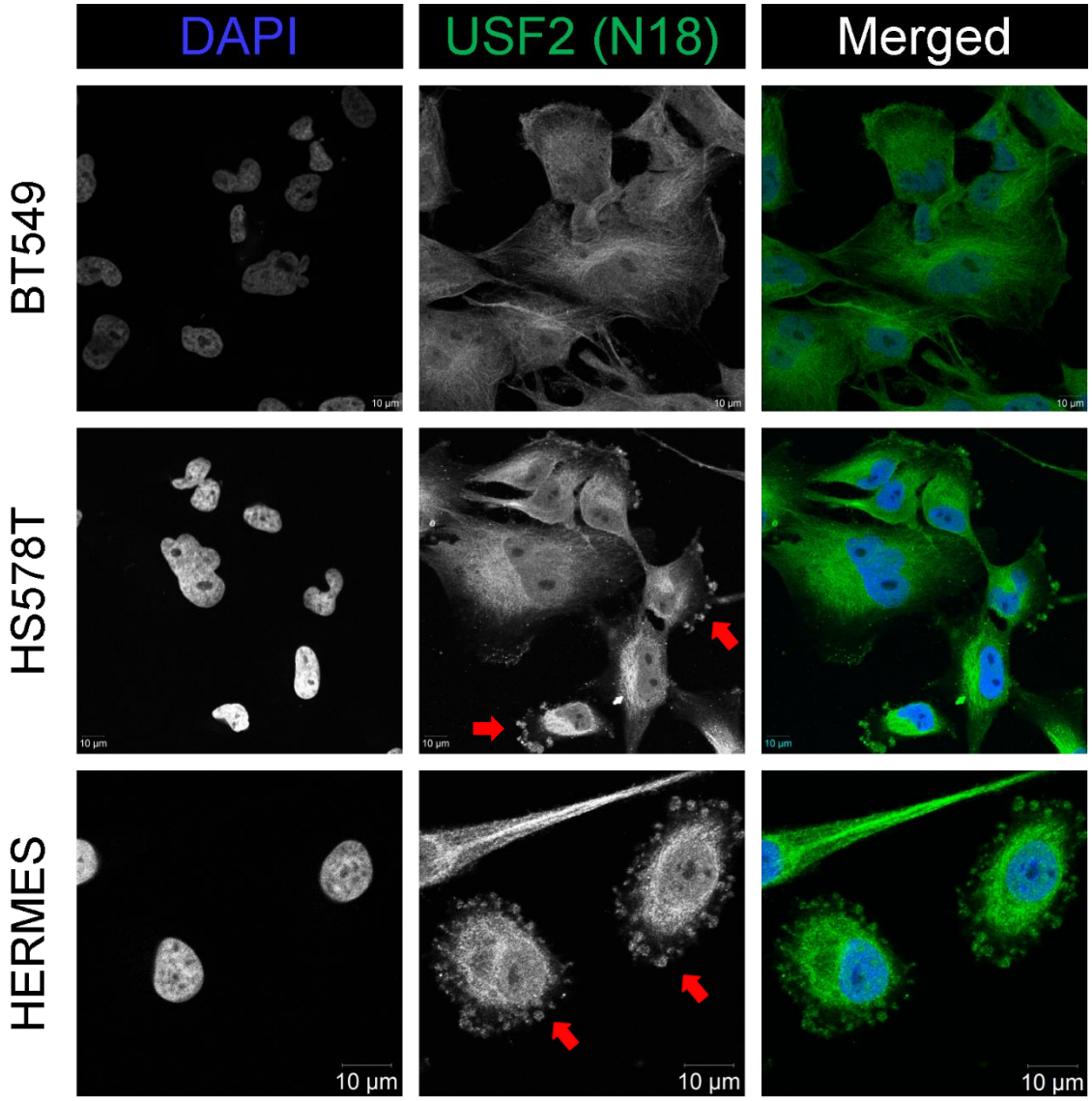


Figure 3.4.7 Various USF2 antibodies, and its target specificity. **A)** Schematic diagram of epitopes on USF2 detected by various antibodies used in this thesis and a summary table underneath outlining the approximate size and the nature of the bands observed on Western-blot, the localisation to the nuclear or cytoplasmic fraction as detected by Western-blot and immunofluorescence in 501mel melanoma cell lines. **B)** Examples of immunofluorescence probing for USF2 using the N18 antibody on various cell line. Full dataset can be found in Supplementary Figure 3.4.6 – 10. Red arrows highlight region of interest mentioned in the main text.

lity that USF2 may directly associate with membrane proteins and shuttle into the nucleus following receipt of specific cues. In the melanocyte, HERMES cell line, similar nuclear and cytoplasmic staining is observed. However, USF2 can be seen localised to the vacuole structure around the membrane (Figure 3.4.7B, HERMES, red arrow). The specific staining pattern of nuclear USF2 and USF2 localised to the leading edge of the membrane is similarly observed with the AB-antibody that has overlapping antigen specificity as the N18-antibody (Figure 3.4.8, USF2(Ab), red arrow). Indicating that USF2 species containing the N-terminal region can be found in the nucleus, cytoplasm and at the leading edge of the membrane. Note that the similar staining pattern can also be seen using a C20-antibody, however, as observed in 501mel cells, C20 staining is biased towards nuclear staining in all cell lines studied (Supplementary Figure 3.4.8 – 9). Interestingly the A19- and C13-antibodies which target the middle region of human USF2 (specific amino-acids not specified by the supplier) staining is biased towards nuclear staining as with C20-antibody Figure 3.4.8, USF2 (A19) and Supplementary Figure 3.4.7 – 10). Also as with the C20-antibody, a different ratio of nuclear and cytoplasmic staining was observed for the A19-

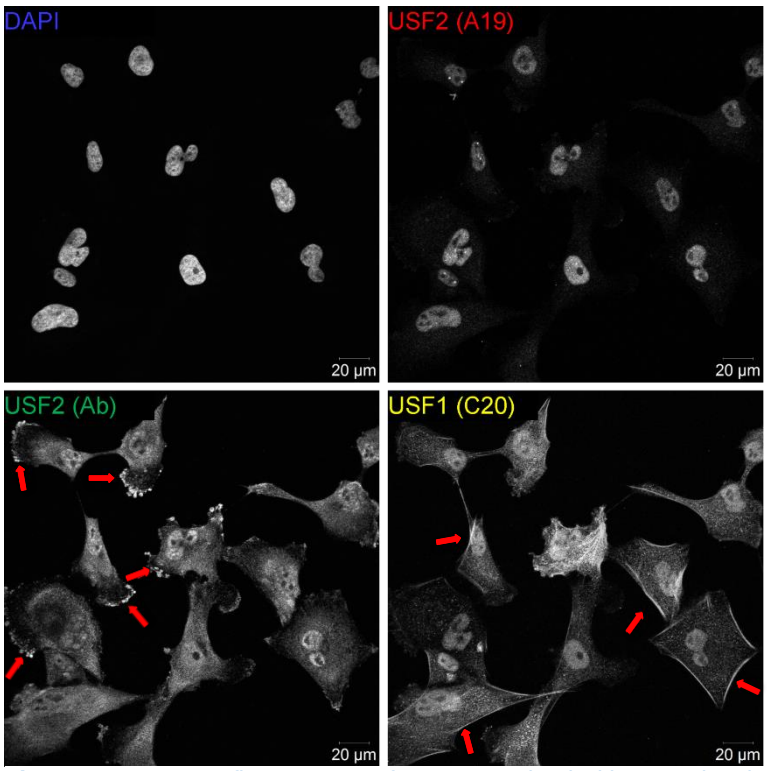


Figure 3.4.8 Immunofluorescence of HS578T stained with USF2 (A19), USF2 (Ab) and USF1 (C20) in HS578T. Red arrows denote region of interest mentioned in the main text.

and C13-antibodies. Taken together, we can conclude that USF2 displays cell line-specific localisation patterns between the nuclear and cytoplasmic compartment. Although not discussed here, it is worth noting that USF1 also displays cell line-specific patterns of distribution (Supplementary Figure 3.4.9), most striking is the localisation to the plasma membrane observed in HS578T (Figure 3.4.8).

USF2 localisation to organelles/cellular structures.

Having observed cell-line-specific organelle/cellular structure-like staining patterns, we ask whether we could co-localise USF2 to specific cellular compartments. We co-stained for USF2 and various organelles and cellular structures including lysosomes (LAMP1 and Lyso-tracker), melanosomes (TRP-2 and HMB-45), mitochondria (JC-1), actin cytoskeleton (Phalloidin) and β-TUBULIN (Supplementary Figure 3.4.11 – 14). As we did not observe any significant co-localisation with the above markers we concluded that USF2 may be localised to a number of organelles or bound to structures not tested in our study.

USF2 and mitosis

One way to regulate factor exchange as briefly discussed in the introduction is through cell cycle. During mitosis, the genome undergoes massive condensation during prophase in preparation for mitotic chromosome alignment during metaphase followed by sister chromatid separation during anaphase and cytokinesis during telophase where the cleavage furrow from the plasma membrane separates the cell into two daughter cells. DNA compaction can physically evict TFs from their binding sites by secluding away their TFBS. TFs whose levels fluctuate during the cell cycle, for example those that are degraded during mitosis, would leave the TFBS open for occupancy by TFs with shared DNA-binding specificity but whose levels remain constant throughout mitosis. Our lab's unpublished data on MITF reveal a fraction of MITF remains bound on to the mitotic chromosome. This implies that MITF can bookmark specific genes for rapid reactivation following cytokinesis. In the case where other bHLH-TFs with shared DNA-binding specificity with MITF get degraded during the cell cycle, one can envisage that MITF will be able to occupy these

sites until other factors return to their G1 level. To study the level of USF2 at different stages of mitosis, we co-stained USF2 with H3S10-Ⓟ, a marker of condensed chromatin frequently used for this type of analysis. We performed experiments on a number of cell lines but only include the results from 501mel and another cell line in this thesis.

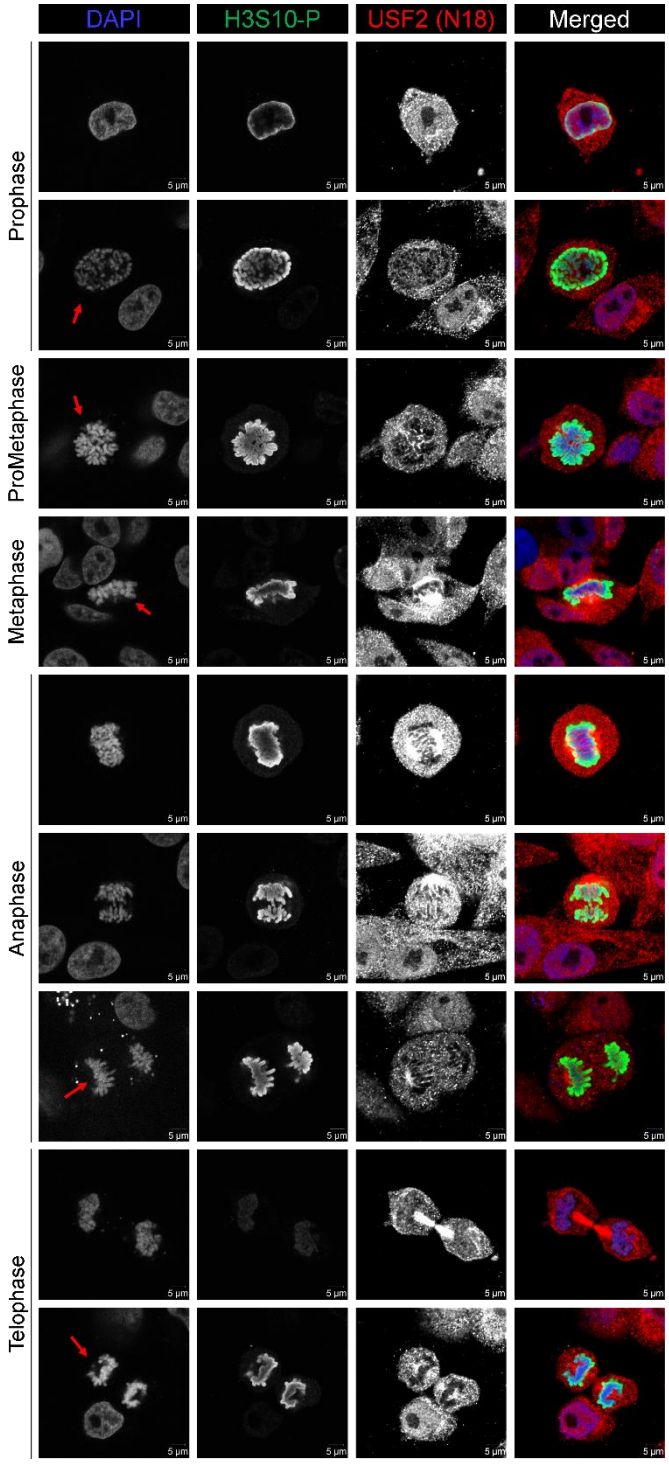


Figure 3.4.9 Staining of USF2 as detected by N18-antibody at different stages of the cell cycle in 501mel cell line. Red arrows indicate cells undergoing mitosis.

In 501mel cells, USF2 as detected by the N18-antibody, is distributed in the nucleus and cytoplasm during the prophase and pro-metaphase (Figure 3.4.9, Prophase & ProMetaphase) but clearly excluded from the condensed chromatin throughout mitosis. Interestingly, USF2 as detected by the N18-antibody co-localised with the microtubule spindle like structure and can be seen to permeate from the microtubule organising centre to the kinetochore but is excluded from the chromatids (Figure 3.4.9, Metaphase & Anaphase). During telophase, USF2 is concentrated at the non-kinetochore microtubules that push the two sets of chromosome apart (Figure 3.4.9, Telophase). Similar observations were made in other cell lines and is exemplified in IGR37 (Supplementary Figure 3.4.15).

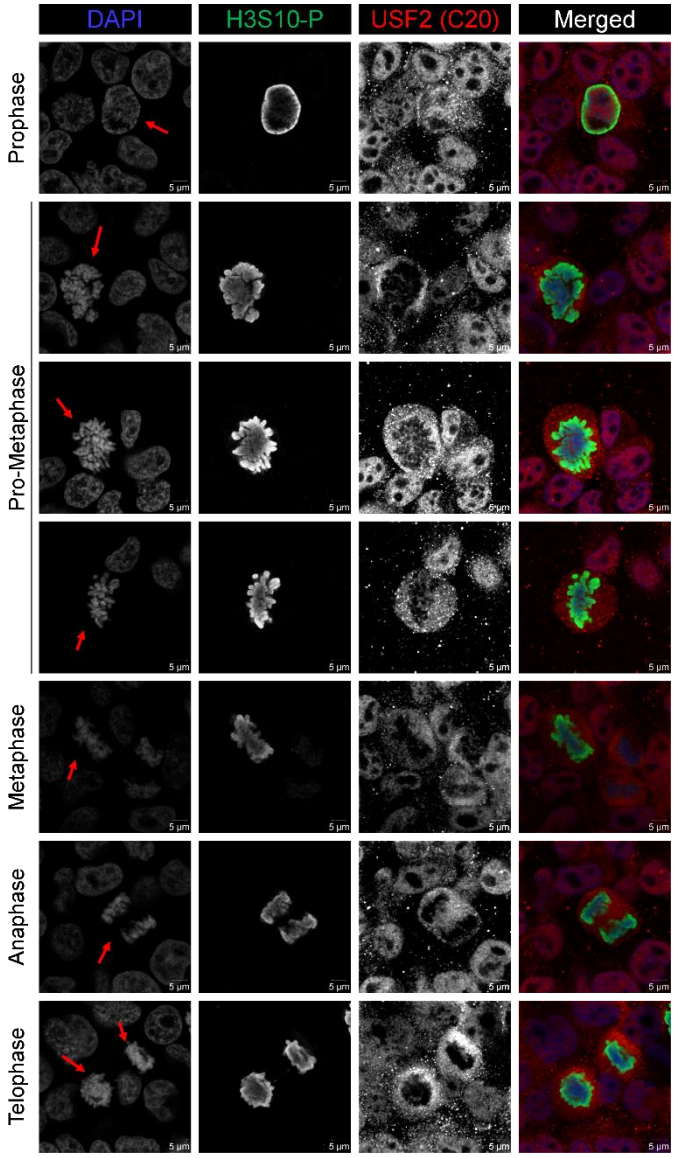


Figure 3.4.10 Staining of USF2 as detected by C20-antibody at different stages of the cell cycle in 501mel cell line. Red arrows indicate cells undergoing mitosis.

USF2 as detected by the C20-antibody, which displayed predominantly nuclear staining during the G1-phase of the cell cycle, appeared largely unaffected during mitosis at the protein level in 501mel cells (Figure 3.4.10). As with N18-antibody staining, C20-staining is excluded from the condense chromatids but unlike N18-staining, C20-staining does not co-

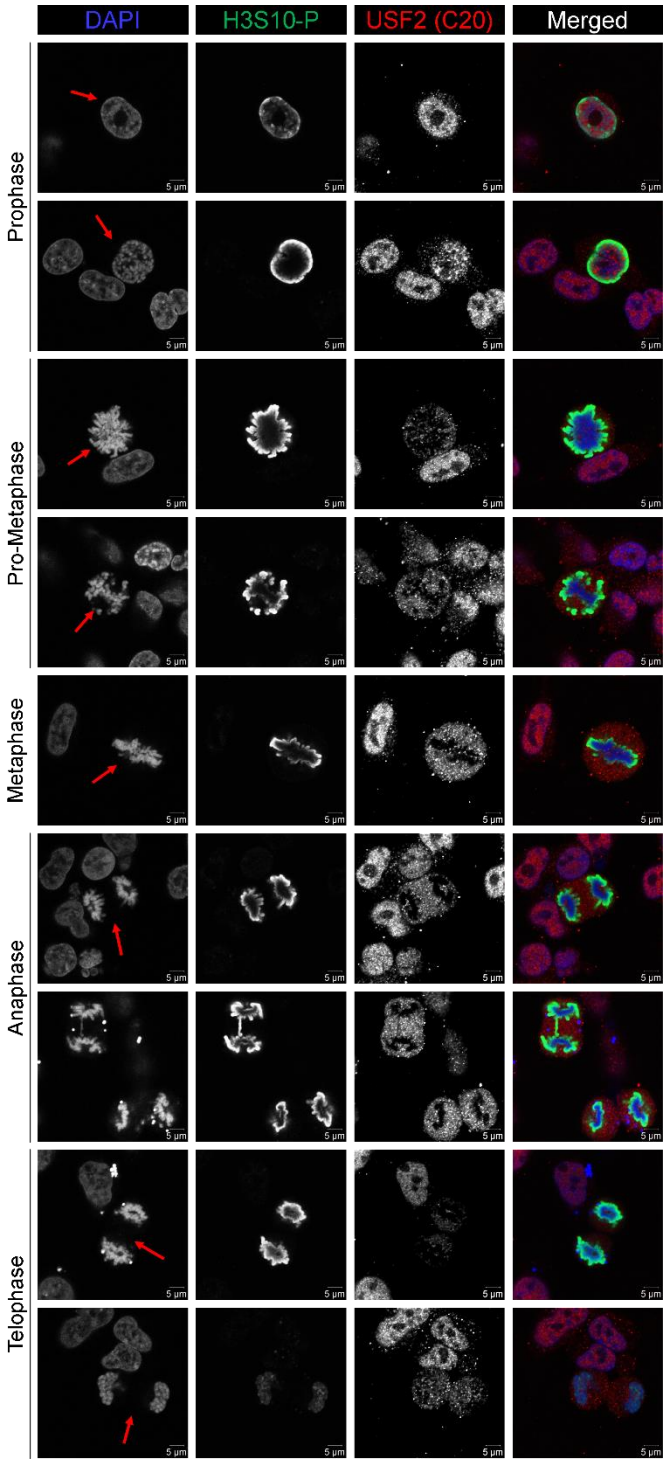


Figure 3.4.11 Staining of USF2 as detected by C20-antibody at different stages of the cell cycle in the HBL cell line. Red arrows indicate cells undergoing mitosis.

localise to the mitotic microtubules (Figure 3.4.10, Metaphase, Anaphase & Telophase). Similar observations were made for all other cell lines except for HBL melanoma cells where the overall C20-staining level decreased following entry into mitosis and remained low throughout (Figure 3.4.11). The staining of A19-antibody demonstrated clear changes at

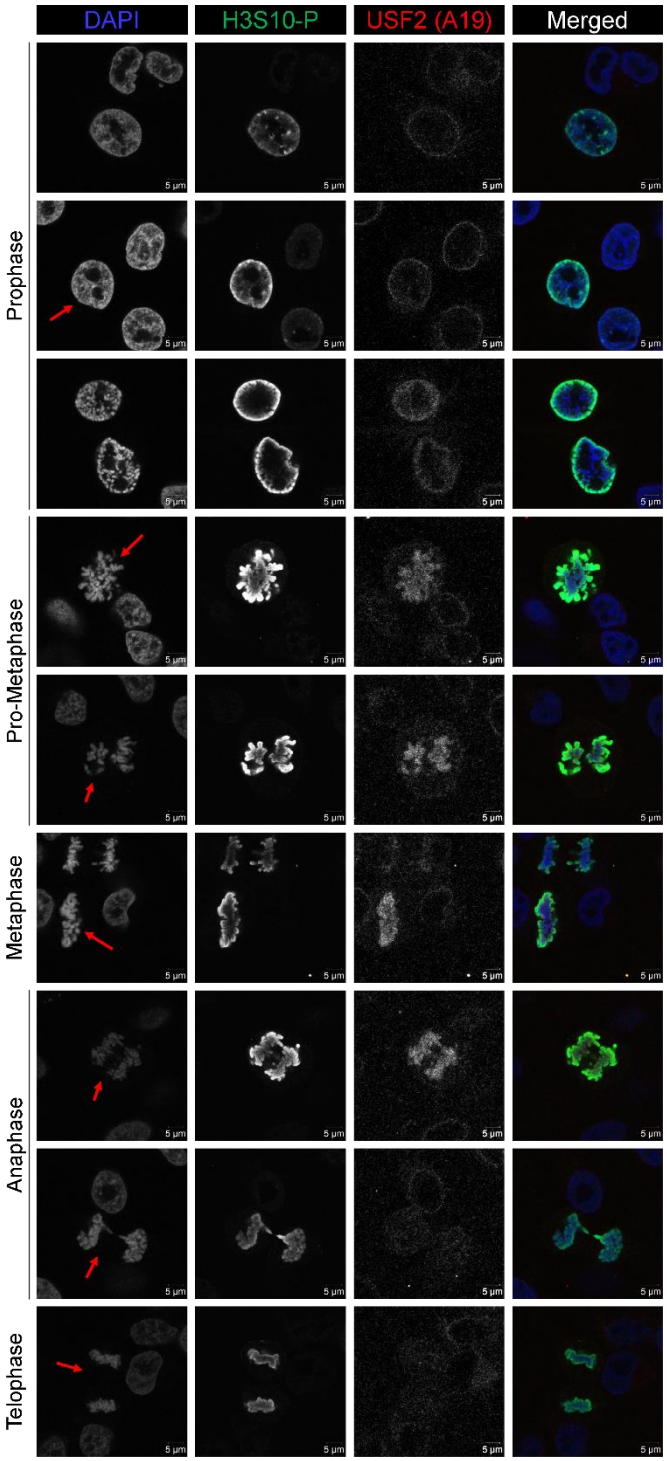


Figure 3.4.12 Staining of USF2 as detected by A19-antibody at different stages of the cell cycle in 501mel cell line. Red arrows indicate cells undergoing mitosis.

different stages of mitosis. In this experiment, USF2 as detected by the A19-antibody is nuclear localised especially around the nuclear envelope in non-dividing cells in 501mel cells (Figure 3.4.12). Nuclear USF2 (A19) accumulated in cells during late prophase (Figure 3.4.12, Prophase, third row) and peaked after chromatids become visible (Figure

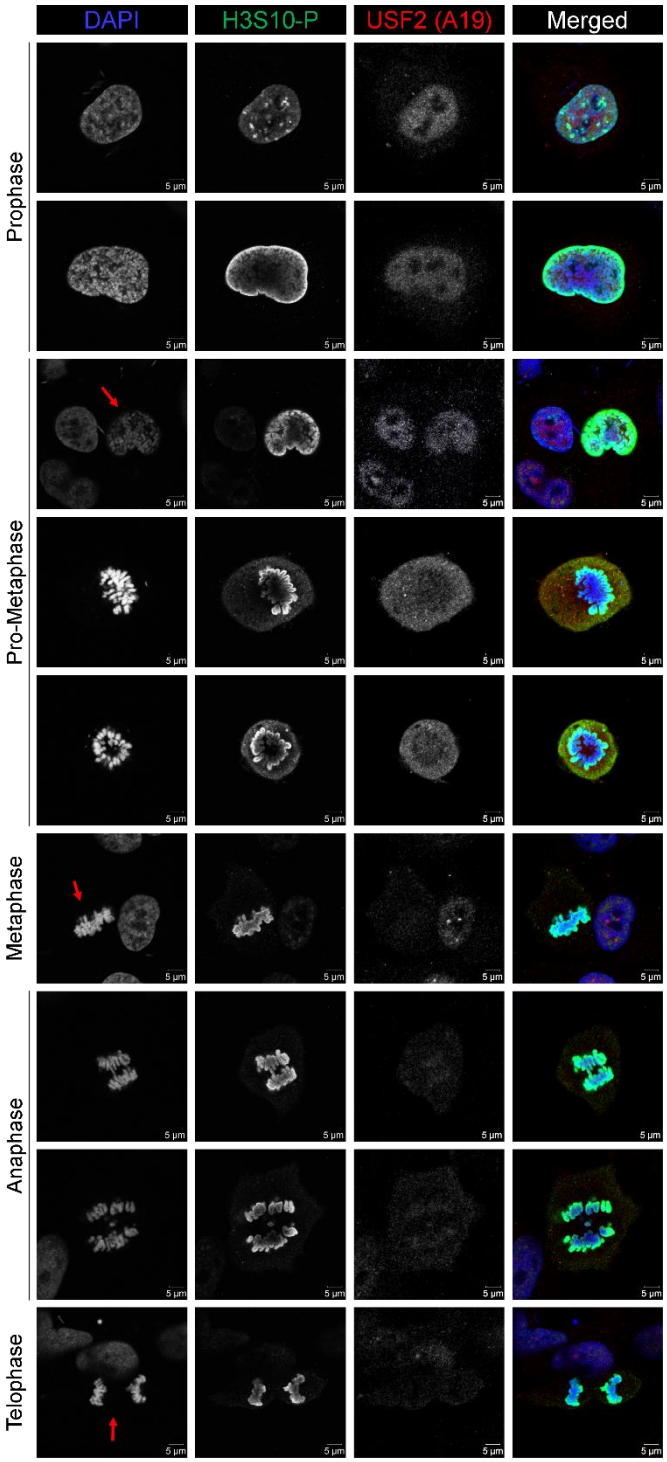


Figure 3.4.13 Staining of USF2 as detected by A19-antibody at different stages of mitosis in the HS578T cell line. Red arrows indicate cells undergoing mitosis.

3.4.12, Pro – Metaphase) and remained high until late anaphase (Figure 3.4.12, Metaphase & Anaphase) where A19-staining returned to the dispersed state in the nucleus (Figure 3.4.12, Anaphase, second row) and remained so through the telophase (Figure 3.4.12, Telophase). This pattern of increased staining as chromatids appear might be specific to

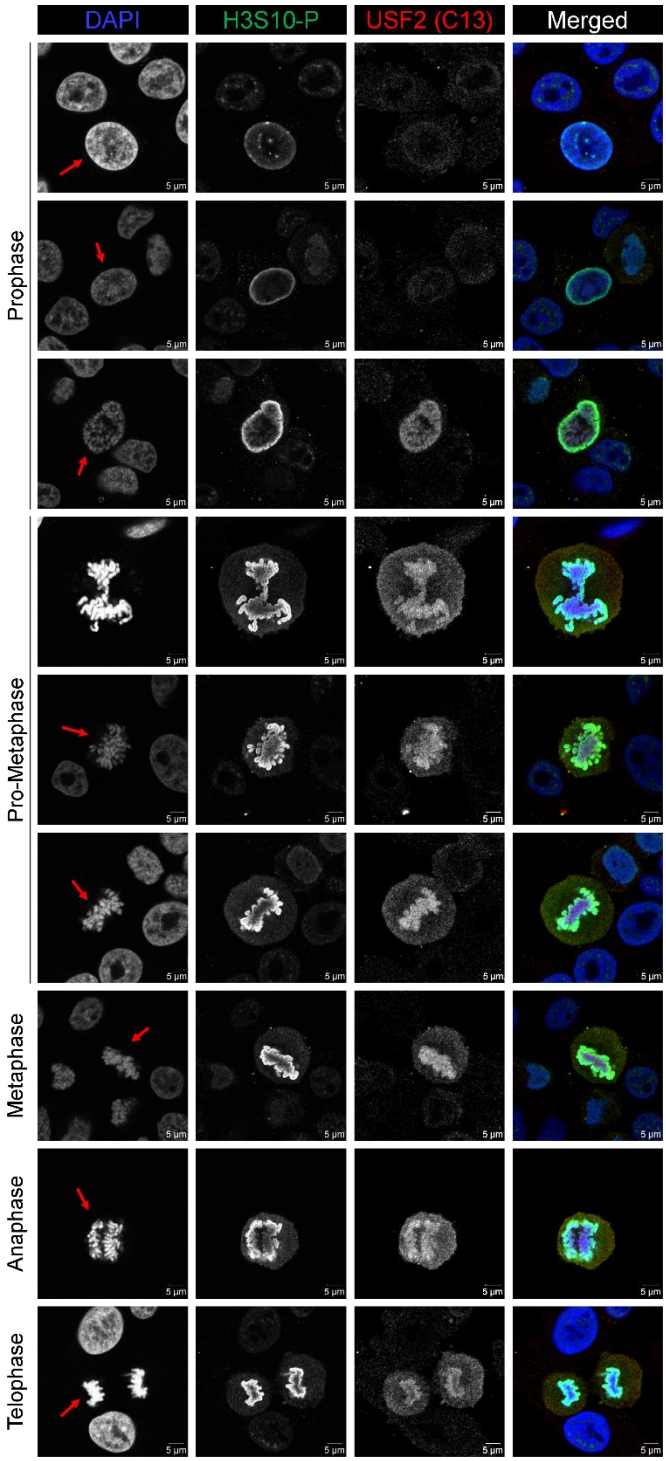


Figure 3.4.14 Staining of USF2 as detected by C13-antibody at different stages of the cell cycle in 501mel cell line. Red arrows indicate cells undergoing mitosis.

501mel cells as we observed a general decrease in A19-staining in other cell lines we looked at as exemplified by HS578T (Figure 3.4.13) and HELA cells (Supplementary Figure 3.4.16). In HS578T, A19-staining persisted until metaphase where the level drastically reduces relative to non-dividing cells.

Lastly, we looked at the staining pattern using the C13-antibody during mitosis. In this case, similar to what was observed for A19-antibody in 501mel, C13-staining was predominantly nuclear in non-dividing cells and began to accumulate in the nucleus of cells during late prophase/early-prometaphasae (Figure 3.4.14, Prophase, third rows in 501mel and Supplementary Figure 3.4.17, Pro – Metaphase, first row for BT549). Also as with A19-staining, C13-staining could be seen on the chromatids. However, unlike the A19-staining that disappeared from the chromatids from anaphase onwards (Figure 3.4.12, Anaphase, second row), C13-staining persisted through telophase (Figure 3.4.14, Telophase).

Brief summary of Chapter 3.4

For the first time, we demonstrated that USF2 species, as detected by two antibodies targeting the N-terminus of USF2, can be found in the nuclear, cytoplasm and at the leading edge of the membrane. The specific detail and ratio appeared to be cell-line specific. However, ectopic USF2A are primarily nuclear localised in many melanoma cell lines tested. The ratio of nuclear:cytoplasmic USF2 is proportional to the amount of USF2 expressed, hinting for a potential involvement of cytoplasmic retention mechanism which may be saturated in overexpressed system.

We showed that USF2, as detected by N18-antibody were excluded from the mitotic chromosome but co-localised with the microtubule spindle like structure and could be seen to permeate from the microtubule organising centre as well as localised at the non-kinetochore microtubules which push the two sets of chromosome apart. The C20-specific species are localised to nucleoplasm throughout mitosis. USF2 species as detected by two antibodies targeting the internal region of USF2 are predominantly nuclear and

accumulates as cells enter mitosis and can be seen concentrated on the chromatids, at least in 501mel. There appeared to be some minor cell line specific differences in the staining pattern.

Contrary to the literature which reported USF induction by glucose and fatty acids at the protein level, we did not observed any changes in the total USF protein level in any of the melanoma/melanocyte cell lines looked at. We have, however, demonstrated a shift in USF2 (N18-species) from cytoplasm to the nuclear compartment in a glucose add-back experiment following overnight glucose starvation. This changes was not seen in 293 FlipIn/TRex cell line. Additionally, we identified novels form of nuclear MITF which still contains the transcription activation domain and DNA binding domain following glucose starvation and a differentially modified form of MITF in the cytoplasmic fraction that migrates on SDS-PAGE in between the two bands observed in the nucleus in the glucose add-back following overnight glucose starvation.

To characterise region involved in regulating USF2 sub-cellular localization we generated deletion mutants based on exon structures and functional domains and assessed the sub-cellular localisation when transiently transfected in to 501mel. The summary of the result can be found in Table 3.4.2. We demonstrated that amino-acids $1-192$ may harbour NES or involved in cytoplasmic retention and that the basic and USR are not sufficient for nuclear localisation. In fact, USF2 mutant lacking the basic domain is nuclear localised. It appears that the USR- and the basic-domain require amino-acids immediately adjacent to function as NLS. Inclusion of exon-5 to the USR-b-domain promotes cytoplasmic retention which is overridden by the presence of HLH-dimerization domain. This suggests that intra-molecular masking might play a role in regulating USF2 subcellular localisation. Moreover, the N- and C-terminal fluorescent protein-tagged USF2 have a significant proportion of USF2 localised to the cytoplasm, indicating a potential conformation induced changes in sub-cellular localisation.

Based on the relative ratio of USF2 deletion mutants pulled-down by GST-tagged full-length USF2 against that of full-length USF1 and USF2 prey (normalised to input) amino-acids $1-193$, with the core at $77-143$, of USF2 probably plays some roles in intramolecular interactions within USF2.

Table 3.4.2 Table summarizing subcellular localisation of various USF2 construct as determine from subcellular fractionation or immunofluorescence.

			Localisation	
	Amino acids	**Domains**	**Fractionation**	**Immunofluorescence**
C-terminal deletion	1-328	N-LZ	N	N
	1-291	N-HLH	N	
	1-248	N-b	N/C	N
	1-234	N-USR + 2AA	N/C	N/C
	1-200	N-Ex5 + 8AA		65% N/C, 35% N
	1-192	N-Ex5	C	N/C
	1-143	N-Ex4		N/C
	1-76	N-Ex3		N/C
N-terminal deletion	37-346	Ex3 – C-ter		N
	77-346	Ex4 – C-ter		N
	144-346	Ex5 – C-ter		N
	193-346	8AA + USR-C-ter	75%N, 25%C	25% N/C, 75% N
	200-346	USR-C-ter	N	N/C
	212-346	24AA + b-C-ter	N	
	223-346	13AA + b-C-ter	N	40% N/C, 60% N
	231-346	5AA + b-C-ter	N/C	15% N/C, 85% N
	248-346	HLH-C-ter	N/C	N/C
ΔNΔC-terminal deletion	144-248	Ex4-b		N/C
	144-261	Ex4-b + 13AA		N/C
	144-291	Ex4-b + 43AA		N
	193-232	8AA + USR		N/C
	193-248	8AA + USR-b		20% N/C, 80% N
	193-291	8AA + USR-HLH		C (with some in N)
	193-328	8AA + USR-LZ		N
	235-248	b		C
Others	Δ235-248	Δb		N (with trace in C)
	FL E244A	FL		N
	USF2 (2B)	Isoform 2B		85% N, 15% C
	USF2 (ΔH)	Isoform ΔH		N/C
	USF2A	Isoform 2A	N	N

% shown for immunofluorescence reflect the % of cells that display: nuclear (N), cytoplasmic (C) or both nuclear and cytoplasmic staining (N/C) and not the relative distribution between the nuclear and cytoplasmic compartment. $N \geq 100$ transfected cells. Abbreviation: LZ, leucine-zipper; HLH, helix-loop-helix; b, basic; USR, USF-specific region; AA, amino-acid; Ex, exon; C-ter(minal), carboxyl-terminal ; N-ter(minal), amino-terminal ; FL, full-length

4. Discussion

Brief summary

To study the role of bHLH-TFs exchange at E-boxes we needed to identify co-occupied targets (by means of ChIP-seq). To dissect the role(s) played by each factor, we also needed to establish approaches to induce as well as intervene or knock-out the functions of specific factor(s) [by means of siRNA and inducible, dominant-negative ER-A-USF (also ER-A-MAX and ER-A-MITF, data not shown) as well as serum and amino-acid starvation).

We demonstrated changes in USF and MITF ChIPped-signal at the *TYR* enhancer M-box in response to UVB and MTX. Using MTX we demonstrated differences in the ChIP-signal and gene-induction at the *TYR* gene following the initial round of induction and re-induction, potentially demonstrating transcription memory.

We also showed, for the first time, that different species of USF2 can be localised to the cytoplasm in different ratios and this ratio is cell-type specific. We also demonstrated nucleo-cytoplasmic shuttling of USF2 by glucose, and characterised regions of USF2 involved in nuclear-cytoplasmic localisation of USF2. Some species of USF2 can be found localised to the condensed chromatids during mitosis.

Aspects to consider while studying factor exchange in humans

The thought experiment

Using a bottom-up approach, one may start by considering a two-step binding model where a bHLH-TF scans the genome through non-sequence specific phosphate backbone contacts and anchor down where the non-specific- plus sequence-specific-binding energy is sufficient to retain the bHLH-TF at that locus. The dwell time at each TFBS will depend on the overall binding energy. A given bHLH-TF is likely to bind a range of target E-boxes (denoted as Ebx in Figure 4.1), albeit with different affinity. The dwell time at each motif is proportional to binding affinity (Figure 4.1A, left).

Next, consider introducing a second bHLH-TF (hereon, TF2) with an identical binding property to the first bHLH-TF (hereon, TF1). When the E-boxes are in excess, the relative binding to each motif by TF1 and TF2 will be identical to that observed with TF1 alone. If the number of TF exceed the number of sites for each motif, one can expect to see a proportional decrease in the binding at high affinity sites by each TF, as the system now contains twice as many TFs competing for those same number of sites, and a proportional increase in binding to lower affinity sites as those TFs unable to bind to higher affinity sites are now available to bind to the lower affinity sites (Figure 4.1A, right).

We, next, consider epigenetic and the higher order nuclear organisation (de Wit and van Steensel, 2009; Hakim et al., 2010; Kind and van Steensel, 2010). Changes in DNA topology as well as wrapping of DNA around nucleosome may leave certain E-boxes facing inwards toward the nucleosome, effectively occluding the binding site. Similarly, higher order chromatin compaction can seclude some E-boxes away in the heterochromatin. This effectively pre-selects the range of genomic loci available for binding by bHLH-TFs.

Now take into account DNA-modifications, such as cytosine-methylation, and differentiate TF1 and TF2 by attributing TF2 with the ability to bind to modified DNA as well as nucleosome-bound E-boxes and reconsider the binding profile (Figure 4.1B). Again, there will be no change in the hierarchy of motif bound; however, the distribution of bound loci will now look very different. In a nucleus where the amount of the TF is limiting compared to the number of available E-boxes, one can expect to see a shift of binding by TF2 towards sites prohibitive to TF1 binding as these sites are always available when not bound by TF2. In a situation where TF2 acts as a pioneer factor to remodel the nucleosome, TF1 can only compete with TF2 for the TFBS after TF2 has opened up the chromatin.

Next, imagine introducing a protein that can inhibit binding of TF1 but not TF2. This may be achieved through protein-protein interaction that interferes with DNA-binding and/or dimerization of TF1 or sequestering TF1 away from the DNA. For examples, Yin-Yang-1

inhibit dimerization and DNA-binding by MYC (Austen et al., 1998), and sequestration of SREBP in the cytoplasm. Interfering with the binding of TF1 obviously favours binding by TF2 to common targets through decreased competition.

Next, consider introducing a third bHLH-TF (hereon, TF3) with overlapping but not identical binding specificity. In this system we should see most binding by each factor at the canonical E-box and competition at sites with motif permissive for binding by both factors (Figure 4.1C, left, note the reduced binding to Ebx2-Ebx8 by TF1 as compared to Figure 4.1A due to direct competition with TF3). The relative occupancy by each TF will depend on the relative binding affinity to each motif. This relative occupancy (but not the dwell time of each binding event) can be altered by altering the protein level of the TF. As each binding event is diffusion-driven, increasing the amount of one TF will increase the incidence of binding (but not the dwell time which depend on binding energy).

Lastly, consider introducing PTMs that may affect phosphate-backbone contacts or sequence-specific base contacts of, say TF3. Altering the phosphate-backbone contact will not affect the hierarchical profile of preferred motif, but will affect relative binding to each motifs. If a PTM increases the binding affinity to the phosphate-backbone of TF3, TF3 will now be able to bind more stably to the lower affinity motifs (non-canonical motif). Assuming the overall binding energy required for anchoring remain unchanged one should see a broader profile of binding toward the lower affinity motifs (Figure 4.1D, red line). If a PTM decreases binding to phosphate backbone, one would expect to see the relative binding across the motifs to be concentrated around high affinity sites (Figure 4.1D, green line). In the former situation TF3 will be able to compete with other TFs across a broader range of motifs as TF3 can now better bind to non-canonical motifs. In the latter situation, TF3's low affinity sites becomes available for binding by other TFs while TF3's high affinity sites demonstrate increased occupancy by TF3 even though the overall binding affinity decreases because there are more TF3 available to compete for these sites.

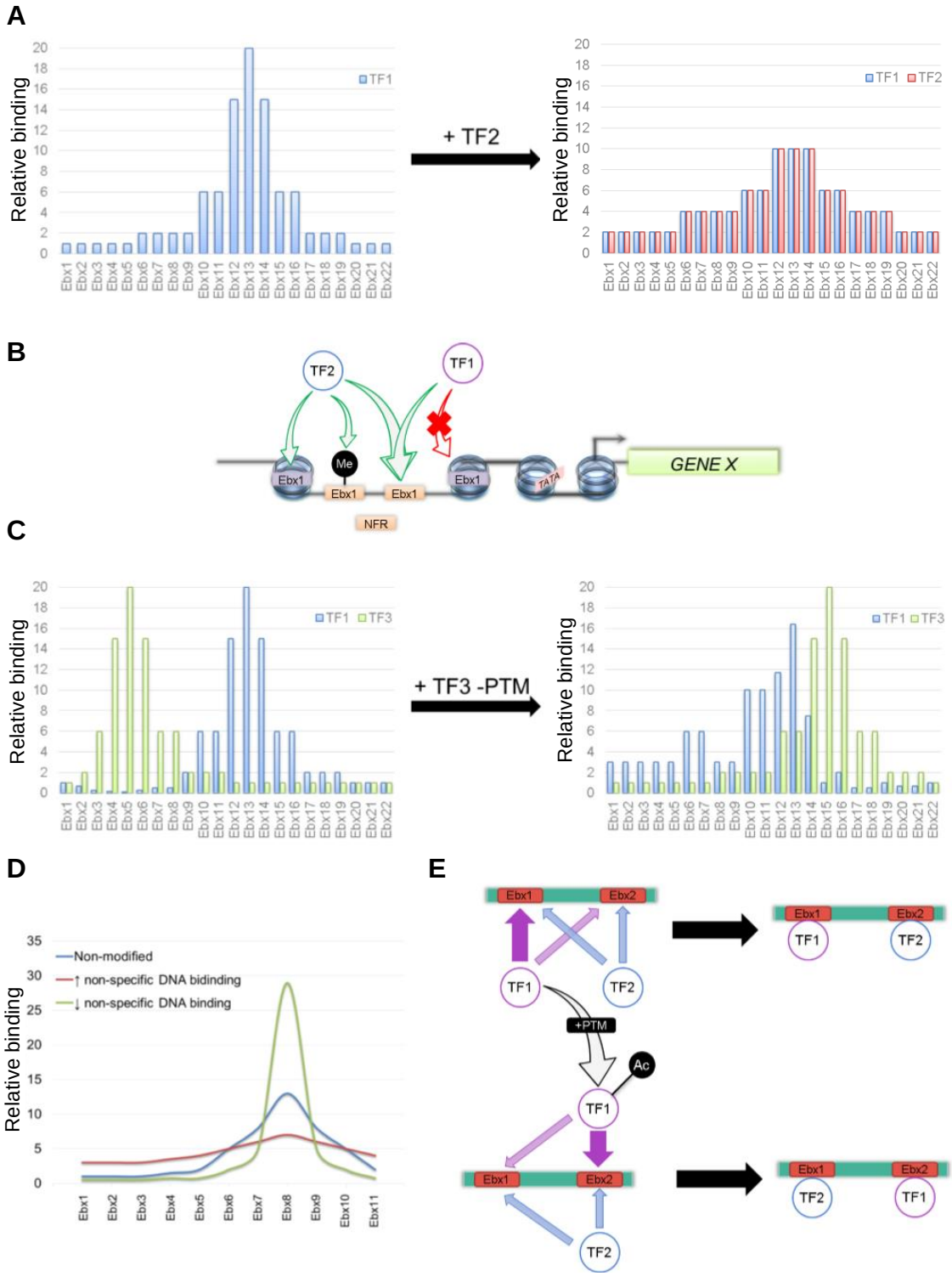


Figure 4.1 **A)** Hypothetical distribution profile of relative binding by a bHLH-LZ-TF (TF1) to different E-box motifs (Ebx) across the genome. TF1 binds to a range of Ebx with highest affinity for Ebx13 and graded affinity for Ebx with sequences that deviate from the canonical Ebx13 (adjacent Ebx are more similar than ones further apart). To the right is the hypothetical profile of relative binding when TF2 is introduced with identical DNA-binding property to TF1, assuming the total number of Ebx is limiting. **B)** Diagram depicting a situation where TF1 is only able to bind to unmodified, Ebx1 in a nucleosome free region (NFR) context while TF2 can bind to modified as well as nucleosome occluded Ebx1. **C)** Hypothetical distribution profile of relative binding by TF1 and TF3 that have overlapping, but very different canonical binding sites. To the right is the hypothetical distribution profile of relative binding by TF1 and TF3 when a post-translational modification (PTM) alters the specificity of TF3 from Ebx5 to Ebx15 and related sequences. **D)** Hypothetical distribution profile for a given TF and the effect of PTM that alter the phosphate-back-bone contact. **E)** Schematic diagram depicting the most favourable outcome given the relative affinity to each Ebx and the effect of a PTM that changes the specificity of the Ebx. (The filled, thick arrow represents high affinity binding relative to the transparent, thin arrow.)

If a PTM alters sequence-specific base contact (exemplified as acetylation in Figure 4.1E, Ac) or alters the conformation of the TF, such that the amino-acids contacting the bases change, one can expect a change in hierarchical profile of preferred motifs. In Figure 4.1E, factor exchange occurs as TF1, which previously preferred Ebx1 now changes its preference to Ebx2 while TF2 retains its low affinity for both E-boxes. Thus, TF2 exchanges (gets competed out) with TF1 (from Ebx1 by TF1 and from Ebx2 by TF1-Ac) as it has lower affinity for the E-boxes than TF1. In a situation where each TF in question has a preference for certain, largely non-overlapping E-box motifs over others as depicted in Figure 4.1C, left, a PTM which alters target specificity such that the two TFs in questioned are now largely overlapping will result in competition for overlapping sites and a drastic change in the binding profiles at those sites (Figure 4.1C, right).

Regulation of factor exchange *in vivo*

Central to the promoter education hypothesis is the timely, coordinated exchange of TFs at co-occupied sites. The underlying mechanism which orchestrated this factor exchange is of paramount importance. Hence, we consider various aspects that may regulated factors exchange.

Local protein availability at the E-boxes

The local concentration of a given bHLH-TF can influence the dynamic of factor exchange at the E-box through direct competition. Local protein concentration may be regulated at the level of sub-cellular localisation (Grinberg et al., 2004) and sequestration (Osborne, 2000), that often affects their functions (Fabbro and Henderson, 2003; Xiao et al., 2003), or at the protein synthesis and degradation level (Blackwood et al., 1992).

As dimerization is obligatory for DNA binding by bHLH-TFs the relative stability of the dimmer can influence factor exchange. Dimerization stability may be regulated by PTM(s) as shown for reduced MYC:MAX complex formation and DNA binding in response to IFN-γ (Bahram et al., 1999). A possible scenario is exchange between the ubiquitous, unstable

dimer with the inducible, more stable dimer. Upon induction, equilibrium will favour the occupancy by the more stable dimer. This way, the less stable dimer with low dwel time on the promoter serves as a bookmarker, maintaining NDR necessary for H2A.Z deposition (Hartley and Madhani, 2009), allowing the gene to be rapidly activated when the inducible, stable dimer binds.

To study factor exchanges through direct competition by induction at the protein level, we looked at hypoxic induction of HIF to identify E-box co-occupied with other bHLH-LZ-TFs studied. The next important question that will need to be investigated is: how HIF induction influences the relative binding of other bHLH-TFs at the co-occupied sites. From the aspect of sub-cellular localisation we investigated cues which might induce nuclear localisation of USF2. Although we identified glucose as a cue which promoted nuclear accumulation of USF2, the effect was rather modest and is unlikely to result in significant changes in USF2 ChIP-seq profile. However, we cannot rule out the potential effect of glucose induction on USF2 to be specific to glucose responsive genes. We will address this by looking at changes in gene expression of glucose responsive genes which are USF2 targets and cherry-pick candidate genes for further ChIP-qPCR studied to address whether glucose induces USF2 factor exchange with other bHLH-TFs co-occupying the site.

PTM(s) may influence DNA binding affinity and specificity

PTM(s) may influence specific and/or generic DNA-protein interaction. PTMs which decrease overall DNA binding affinity will allow displacement by other TFs which can bind more stably to the same motif, provided that the overall thermodynamic (taking in to consideration local protein concentration) is favourable. PTM(s) that alter specific base contact (either directly or through conformational changes) can change target preferences. These PTM(s) may be induced by specific signalling pathway, allowing transcriptional output to be responsive to specific cues.

Evidence for PTM(s) influencing DNA binding affinity stemmed from the observation that bacterially expressed MAX binds DNA much better than *in vitro* transcribed one, indicating that MAX DNA-binding activity can be regulated (Berberich and Cole, 1992; Kato et al., 1992). The DNA binding activity of MAX homodimer is inhibited by casein kinase II (CKII) phosphorylation (Berberich and Cole, 1992; Bousset et al., 1993; Prendergast et al., 1992). MYC is also phosphorylated by CKII at the equivalent site (Luscher et al., 1989). However, this phosphorylation does not affect MYC DNA binding (Berberich and Cole, 1992). This CKII phosphorylation site appears to be conserved among bHLH-proteins including JUN and FOS and has been implicated in regulation of JUN binding *in vivo* (Boyle et al., 1991).

CKII can also phosphorylate the C-terminus of MAX. Mutation of this phosphorylation site promotes DNA binding by MAX-MAX but suppress binding by MYC-MAX. There is no apparent change in dimerization efficiency or the nuclear localisation by the C-terminal phosphorylation mutant (Prendergast et al., 1992).

Cell cycle-dependent PTM(s)-mediated changes in DNA binding specificity has also been reported for MYC. Hyper-phosphorylation of MYC during mitosis correlated with reduced non-specific DNA binding (Luscher and Eisenman, 1992).

Based on the ongoing work in our lab on acetylation of MITF at K243 (which contacts DNA-phosphate back bone), we investigated the effect of acetylation-mimetic at the MITF K243 equivalent residues on TFE and USF on their ability to bind to salmon sperm DNA *in vitro*. As with MITF K243Ac, TFE and USF mutants with reduced DNA binding affinity also show reduced ability to bind to salmon sperm DNA containing various non canonical E-boxes.

Chromatin and DNA modifications

Chromatin structure may also impede binding by secluding away TFBSs, as often the case for silenced genes. It has been shown that Max-Max and Myc-Max are able to bind to nucleosome bound E-boxes with similar affinity to free E-boxes (Wechsler et al., 1994) while USF cannot (Lüscher and Larsson, 1999). In this capacity, MYC-MAX exchange with

USF at nucleosome bound E-boxes can only occurred after MYC-MAX mediated chromatin opening. In this thesis, we demonstrated that chromatin compaction during mitosis lead to USF being excluded from the chromatin.

DNA methylation at the CpG dinucleotide within the E-box was shown to inhibit binding by MYC and MAX but not USF (Prendergast and Ziff, 1991). As this methylation has been shown to be cell-type specific in some instances, for examples, *CASP8* and *EGFR*, DNA methylation can directly regulated factor exchanged, for example, following differentiation and hence, MYC-dependent transcription (Perini et al., 2005). We have not addressed this in this thesis, however, future work will address the effect of CpG methylation on DNA binding (see Chapter 5, CpG methylation and binding of MITF, MYC, HIF and USF).

In short, binding of a sequence-specific TF is influenced by:

1. Number of available binding sites
2. Number of binding-competent TFs
3. Availability of the binding sites (accessibility and modification state)
4. Competition with related TFs for common binding sites and competition with TFs whose binding sites overlap the binding sites of the TF in question
5. Presence of proteins which regulate the affinity, specificity and the ability to bind by the TF in question through PTM(s), allostery and/or stearic inhibition

Genome-wide analysis

Although a large volume of MYC and USF ChIP-seq data exist, there is no published USF, MYC or HIF ChIP-seq in melanoma cell lines. To the best of our knowledge, we are the first to carry out a comprehensive comparative study of as many as seven, highly related bHLH-TFs, covering a lineage specific (MITF), four- generic (USF, HIF-1β and MYC) and two conditionally induced (HIF-1α and HIF-2α) factors in a human melanoma model.

There is no strong correlation between peak score and the motif under the peaks.

Based on thermodynamic argument, one may expect most binding by a bHLH-TF at the highest affinity E-box and a gradual decrease in relative binding levels as we go down the list of E-boxes with progressively lower affinity. However, our ChIP-seq analysis revealed a sizable portion of binding at sites lacking any apparent E-box motifs, even among co-occupied sites. With the exception of canonical motifs, we did not see any correlation between peak score and number of motifs under the peak, globally. Moreover, we were able to identify binding to motifs, which in the band-shift experiments were not able to compete against the canonical E-box, in our ChIP-seq peaks. How can these be explained?

Binding at non canonical E-box may be explained through: modulation of target specificity and/or non-sequence specific DNA-binding affinity, competition with other bHLH-TFs which may titrate away high affinity sites as well as artefact of performing the ChIP-seq experiments with over-expressed proteins. Another intriguing possibility is that bHLH-TFs bind cooperatively to adjacent E-boxes (Walhout et al., 1997) through heterotetramer formation as observed for Myc-Max X-ray structure where two heterodimers join through the head-to-tail assembly of the LZ to form an antiparallel four-helix bundle (Dang et al., 1989). This heterotetramer formation is specific to Myc-Max and not Mad-Max (Nair and Burley, 2003). This was proposed to increase the affinity of Myc-Max to non-canonical E-boxes in a flanking sequence, sequence between the two E-boxes and the distance between the two E-boxes-dependent manners (Walhout et al., 1998). Cooperative binding may also explain some of the observed binding at sites with no apparent E-box motifs through passive binding with (a)other sequence specific TF(s).

MYC, MITF, USF and HIF peaks are dissimilar in their distribution across the genome.

Earlier studies of MYC demonstrated selective enrichment of Myc's E-boxes at the 5' regulatory elements and not just any E-box in the genome. (Fernandez et al., 2003). This indicated that not all E-boxes are equally accessible to Myc and/or that Myc was unable to bind to every E-box with the consensus sequence. The former may involve prohibitive

chromatin structure or occlusion of the E-boxes by other proteins. The latter would suggest that: Myc required additional proteins to bind that are not present at non-promoter E-boxes and/or DNA-modification/other proteins inhibit Myc's binding at certain E-boxes. This has led us to study the distribution of peaks across different genome ontologies (Figure 3.1.19A).

In 501mel cells, we did not observe specific enrichment of MYC at the promoter when considering all peaks-called, or peaks with score of ≥ 10. The relative distribution into each genome ontology differed for each bHLH-TF with HIF displaying the highest proportion of peaks located in the promoter region, followed by USF1 and MYC. MITF is the only bHLH-TF we studied that uniquely displayed essentially no binding to the 5'UTR, a very low proportion of peaks bound to the promoter ($\approx 10\,\%$ for top 6% peaks) and a very high enrichment of peaks within introns and intergenic regions ($\approx 41\,\% \,\&\, 45\,\%$ for the top 6% peaks). It is important to be aware that the distribution of peaks changes very little (slight decreases in the intergenic and intronic region in favour of the promoter region) when comparing full- and filtered-dataset of all bHLH-TFs studied, indicating that the peak score is largely unaffected by the genomic location which peaks appeared. Going by the study which reported that productive transcription is related to the dwell time at the promoter (Lickwar et al., 2012) and the assumption that peak score correlates with residency time, non-promoter peaks most likely also have regulatory functions, perhaps regulating transcription from anti-sense promoters and non-coding transcripts or represent binding to distal regulatory element. The difference in the relative distribution of peaks into different genome ontologies by each bHLH-TF studied is indicative of their different biological targets.

There were no HIF-1α or HIF-2α specific targets in 501mel.

The lack of exclusivity in HIF-1α and HIF-2α binding at reported HIF-1α or HIF-2α specific targets, even at genes which harbour a consensus HIF-1α HRE and not the E-box preferred by HIF-2α, in our ChIP-seq datasets is unlikely to be due to the limitation of comparing different ChIP-seq samples because: firstly, all HIF-ChIP-seqs used the same pool of starting material split into four fractions for each HIF antibody and pre-immune serum

(negative control) which were processed in parallel, and secondly, the normalised peak height of HIF-1β is approximately double that of HIF-1α + HIF-2α. The latter reflects a direct competition between HIF-1α and HIF-2α, both of which require HIF-1β dimerization partners for binding to their target binding sites. By a process of elimination, we suspect that the difference may be attributed to the relative HIF-αs protein levels in 501mel cells at the time point of sampling. This is not something we can directly assess with the available tools and reagents. Once, we have access to antibodies that work reliably for Western-blotting, we could attempt to quantify the amount of HIF by doing a titration curve of purified HIF of known concentration (and hence, the number of molecules) against the HIF level in 501mel cells at the time point which we sampled our ChIP-seq experiment. According to the literature (Holmquist-Mengelbier et al., 2006), we may hypothesise that the relatively small differences between HIF-1α and HIF-2α occupancy at HIF targets observed in our datasets is due to high ratio of HIF-2α over HIF-1α after prolong hypoxic treatment. This may potentially allow invasion of HIF-2α in to targets preferentially bound by HIF-1α.

MYC, MITF, USF and HIF demonstrate enrichment of highly similar motif and co-occupy a large number of sites but also regulate many unique targets.

Our ChIP-seq analysis indicates that; provided that the sequence of the E-box is permissive for binding by the bHLH-TFs, there is very little exclusivity for a given E-box to binding by a specific bHLH-TF. The lack of enrichment of a specific bHLH-TF at a given site may be a consequence of performing ChIP-seq using cells with different background. For example, the MYC ChIP-seq is prepared from cell lines expressing elevated level of MYC and USF2 relative to the parental 501mel cells. Elevated level of MYC allows MYC to bind to non-canonical sites which may be more favourably bound by other bHLH-TFs. In the parental 501mel, where MYC-protein levels are low, thermodynamics will favour MYC binding to MYC's cognate sites, leaving very little MYC to compete with other bHLH-TFs for binding to MYC's low affinity sites. This notion is supported by previously published observations that overexpressed MYC can bind non-specifically (for review see Sabò and Amati, 2014)

while MYC's highest affinity sites are constitutively bound regardless of MYC's protein level (Fernandez et al., 2003).

The E-box motifs enriched in the bHLH-TFs studied (with the exception of the motif identified from the HIF-1α full-dataset) are remarkable similar with subtle differences in the preferred core dinucleotide and the flanking nucleotides. The similarity in the preferred motif of the bHLH-TFs studied explain the high degree of co-occupied targets. Therefore, we argue that the difference in the sequence immediately flanking the E-box and the context which the E-box is presented (for example, in conjunction with TFBS of another TF that can bind cooperatively with the bHLH-TF in question), result in the observed biases for regulation of a given gene by the bHLH-TF which can bind sufficiently long to activate the gene.

The observed high degree of co-occupancy and the argument based on dwell time of the bHLH-TFs may explain the conflicting reports that try to attribute activation of some genes to a specific bHLH-TF. For example, the conflicting reports attributing either HIF, USF or MYC as the positive regulator of *hTERT* (Chang et al., 2005; Goueli and Janknecht, 2003; Yatabe et al., 2004), and the relative contribution of HIF and USF at many of the metabolic genes such as *L-PK* (Krones et al., 2001) and *LDHA* (Hu et al., 2011). As transcription activation potential of each bHLH-TF at each gene may be cell-type and context specific [encompassing PTMs (that modulate affinity and specificity) and epigenetic] as had been shown for USF2 (Ismail et al., 1999), the differential regulation observed may depend on the system studied.

Cooperative binding with (an)other TF(s) may also confer target specificity to each of the highly related bHLH-TF and may explain the absence of any obvious correlation between specific E-box motif and biased occupancy toward one of the bHLH-TFs, even though all bHLH-TFs demonstrate unique sites. The biological relevance of unique targets is supported by the lack of correlation in the pairwise comparison of the RNA-seq from cell

lines overexpressing MYC and USF. Despite having similar motif preference, MYC and USF do not regulate identical set of genes.

Target specificity against the observed bound-targets

With respect to sequence conservation within the basic domain of subclass B TFs, from a theoretical standing, one would infer that their binding specificity and affinity should be very similar. This is supported by many *in vitro* experiments involving MYC, MAX, MAD and USF family members (Sommer et al., 1998). As some genes are preferentially bound by certain bHLH-TFs while others are bound by many. How, then, is the target specificity achieved? The former observation implies mechanism(s) exist to limit the range of bHLH-TFs able to bind, while the latter raises a number of conundrums. As it is physically impossible for multiple factors to bind to the same locus at the same time, either each bHLH-TF binds in response to distinct cues, and/or PTMs and other proteins are involved in determining factor exchange by each bHLH-TF (see section 1.3).

The potential involvement of PTMs is supported by the observation that the source of protein can influence the differences observed in binding site selection experiments. For examples, the inhibitory effect 5'-T or 3'-A residues flanking the core E-box on MYC-MAX binding (Fisher et al., 1993b) is less pronounced when using transient overexpressed protein from COS-7 cells (Sommer et al., 1998). It is possible that PTMs are required for the observed discrimination against flanking T/A bases. In an overexpression system, the system may be saturated and a large fraction of MYC remains unmodified, hence no preference for the flanking sequence. However, the difference in the flanking sequence preference between MYC:MAX and USF (Bendall and Molloy, 1994; Blackwell et al., 1993) was also shown to be negligible when comparing COS-7-derived native proteins complexes (Sommer et al., 1998), suggesting a cell-type specific differences.

Modulation of target specificity though PTM is largely unexplored. PTM may affect residue contacting the phosphate backbone or directly contacting a base, that may alter binding

affinity and specificity respectively, as we showed for MITF (Goding lab unpublished observations). Unacetylated MITF at $\mathrm{K243}$ (involved in DNA-phosphate backbone contact), which has a higher non-sequence-specific DNA binding affinity, is able to bind to more degenerate E-boxes, effectively being titrated away from its canonical E-boxes to non-canonical sites until specific induction cues lead to acetylation of MITF, reduced non-sequence specific DNA-binding affinity and re-directed MITF back to its canonical E-box containing targets. The consequence of $\mathrm{K}-243\mathrm{Ac}$ is that; in response to MAPK-signalling, there is more MITF bound to the consensus M-/E-boxes allowing increased binding/competition to these sites against other bHLH-TFs with overlapping E-box specificities such as USF and MYC. Since, modulation of generic DNA-binding affinity does not allow MITF to discriminate proliferative genes (5'-T-CACGTG-A-3') and differentiation genes (5'-T-CATGTG-A-3') an additional regulatory mechanism must exist to allow discrimination of these targets.

The highest affinity motifs identified *in vitro* differ from the most preferred motif identified from *in vivo* ChIP-seq.

de novo motifs identification informs us of the most prevalent motifs bound by each TF which may not necessarily be the highest affinity sites due to ChIP-seq artefacts as well as nucleotide bias found in different genome ontologies. To investigate this difference, we compliment peak score-motif correlation analysis with *in vitro* band-shift experiments.

The preferred E-box identified from ChIP-seq (5'-CACGTG-3') and band-shift (5'-CATGTG-3') experiments differed for MITF and USF1. It is possible that, globally, 5'-CACGTG-3'-targets are more frequently presented in conjunction with other TFBSs whose cognate TFs can bind cooperatively with MITF and USF1 resulting in high affinity binding and hence, thermodynamically more preferred. Alternatively, 5'-CACGTG-3' may be more prevalent among housekeeping genes while 5'-CATGTG-3' may be more prevalent among inducible genes. As ChIP-seq peak score is a population average value, housekeeping genes which are constitutively express will have a higher proportion of cells with MITF bound to the

promoters than inducible genes. In both scenarios, MITF is permitted to have 5'-CATGTG-3' as its highest affinity target yet shows 5'-CACGTG-3' as motif with highest peak score.

The difference in the motifs preferred derived from ChIP-seq and *in vitro* competition assay may also be accounted for by the limitations inherent to each technique. For ChIP-seq this may include the accessibility of MITF to cross-linking agent and the antibody. 5'-CACGTG-3', which is more prevalence among MITF target around the promoter region (Supplementary Figure 3.1.4, C), may be associated with a more open chromatin structure, allowing efficient cross-linking of MITF. Equally possible is that 5'-CATGTG-3' may be more widespread among targets where MITF binds in conjunction with other protein-complexes. This may result in a decreased cross-linking efficiency or epitope masking, both of which negatively influence the amount of MITF ChIPped. Whilst these issues are absent in *in vitro* competition assay, we have to bear in mind that we used overexpressed protein from Phoenix cells' nuclear extract. Although we took precaution by measuring the acetylation of our overexpressed proteins using pan-acetyl-lysine antibody against purified protein, we have not ruled out the possibility of other PTMs which may influence the binding specificity. Conversely, if modification which influence DNA-binding is the default state as has been shown for MAX (Berberich and Cole, 1992; Kato et al., 1992), overexpressed system may saturated the internal mechanism responsible for the modification, resulting in protein which behave dissimilarly from *in vivo*. If the TF of interested can exist in multiple modified form with different overall binding affinity to the DNA, extractability can also become an issue. Lastly, these competition assay, whilst attempt to reconstitute physiological pH and salt concentration presented in the nucleus, are not identical. It is conceivable that the differences in the binding condition may influence DNA-binding specificity.

The motifs at the co-occupied sites are dictated by the sequence preference of the factors co-occupying those sites.

The motifs at co-occupied sites are dictated by the factor with the more stringent requirement in the E-box sequence it can bind. For example, USF1 + MYC common targets

do not have 5'-T + 3'-A flanking sequence which is known to inhibit MYC's binding (Fisher et al., 1993b) but retain 5'-T as presented in the motif identified for both factors individually. However, the motifs identified *de novo* from sequence under co-occupied peaks appear more degenerated as we increase the number of co-bound factors into consideration. For example, USF + MYC + MITF common targets demonstrate a degenerated requirement for the flanking sequence. This is most probably because to be able to accommodate all the bHLH-LZ-TFs queried, the sequence must be more degenerated as each bHLH-TF have a slightly difference in the preferred flanking sequence. Alternatively, co-occupied peaks may represent peaks with multiple E-boxes of different types under the peak.

501mel specific peaks are not enriched in pigmentation related genes.

Based on the reported role of MITF as a master regulator of pigmentation genes, we reasoned that some of 501mel-specific sites co-bound by MITF should be enriched in the differentiation genes associated with pigmentation processes. To our surprise, however, we did not observe enrichment of any GO-terms related to pigmentation among 501mel-specific-peaks for individual bHLH-TF as well as peaks co-occupied with MITF, except for one GO-term for 501mel-specific-USF2 peak. Closer inspection of 501mel-specific peaks revealed some melanoma antigen proteins and some pigmentation-related genes. Interestingly, 501mel-specific-peaks for each bHLH-TF do not overlap with each other well, indicative of differential occupancy by each factor and that the promoters of the pigmentation-related genes may be differentially occupied in non-pigmented cells used as background to identify 501mel-specific peaks. It is possible that occupancy at pigmentation-related genes in non-pigmented cell lines are stochastic and is independent of MITF. Since we used four cell lines as background for MYC, USF1 and USF2, we ended up with very few pigmentation-related terms. This is supported by the higher incidence of pigmentation-genes among 501mel specific HIF peaks which only have one publically available data for use as a background file. Keeping the stochastic binding argument in mind, it is fascinating to see enrichment of specific processes. For example, 501mel-specific HIF-α peaks are

enriched in metabolic processes, while 501mel-specific USF1 peaks are enriched in immune-response related processes. As these GO-terms are not melanoma-specific, it may be worth investigating these processes in more detail as some of them may be uniquely important in the progression and survival of melanoma cells.

bHLH-TF exchange

To address the question pertaining factor exchange at co-occupied sites, we performed ChIP-qPCR at the *TYR* enhancer M-box following induction cues which specifically enhance the level of MITF or induce PTM that modulate the ability of MITF and USF1 to bind to the M-box. Using dominant-negative USF which interferes with the ability of USF to bind, amino-acid starvation which diminishes the level of MITF, and serum-starvation which is supposed to reduce the level of MYC, we demonstrated how modulating the availability of specific bHLH-LZ-TFs can affect relative binding of the remaining bHLH-LZ-TFs.

There is no correlation in differential gene expression profile among stable cell lines overexpressing USF1 or MYC relative to 501mel, arguing against co-ordinate exchange of bHLH-TFs at co-occupied sites being required for maximal gene expression.

If coordinate exchange of MYC, MITF and/or USF at co-occupied sites is required for maximal gene expression, biasing the promoter to being occupied by either MYC or USF ought to similarly negatively influence gene expression. As there is a large number of overlapping targets bound by any combination of MYC, MITF and USF, yet differential gene expression analysis comparing stable cell lines overexpressing USF1 or MYC did not yield any correlation, at a glance, this may argue against cycling exchange of bHLH-TF at co-occupied site being required for maximal gene expression.

Given that: our stable cell lines overexpress MYC and USF to a different degree and overexpression of MYC and USF1 also influence the level of USF2, our stable cell lines do not recapitulate a situation where the co-occupied site is monopolised by one bHLH-TF. Moreover, depending on the sequence of the E-box, each gene is likely to have a different level of dependency on each of the bHLH-TF studied. As we only include correlation

analysis of the whole RNA-seq dataset and not dataset sub-divided into different E-boxes, we are unlikely to see any correlation. However, it is also unlikely that all of the co-occupied genes can independently be activated by any one of the bHLH-TFs co-bound. Had this been the case, we should observe a positive correlation in differential gene expression analysis when we did pairwise-comparisons of HA-USF and FLAG-MYC cell lines.

Continuous exchange of USF and MITF may be required for expression from pigmentation-related genes.

The fact that 501mel-specific peaks were not enriched by pigmentation genes and that many peaks associated with pigmentation genes are also found in other cell types implies that closed chromatin structure is not required to prevent expression from pigmentation-related genes in non-pigmented cells. This notion is in contrast to some nucleosome mapping experiments that identified nucleosomes mapped over the TSSs of silenced genes (Bargaje et al., 2012). Instead, continuous exchange of USF and MITF might be required for expression of pigmentation-related gene as we showed for *TYR* gene where increased binding by USF1, USF2 and MITF are required for expression of *TYR* following induction by MTX (Figure 3.3.5D & 3.3.6F – G). In a cell lineage lacking MITF, USF may still be able to compete with the promoter proximal H3.3/H2A.Z double variant nucleosome core particles (Jin and Felsenfeld, 2007; Jin et al., 2009) presented at many cis-regulatory element (Deal et al., 2010; Mito et al., 2007), resulting in USF ChIPped-signal. However, analogous to the situation at *PHO5* gene (see section 1.3), productive transcription elongation of pigmentation genes may require MITF. In the absence of MITF, there may be no recruitment of the basal transcription machinery just as we have shown in Figure 3.3.3B that a maximal RNAPII ChIPed-signal coincides with a peak MITF ChIPped-signal while the USF ChIPped-signal persists after the RNAPII ChIPped-signal decreases.

Induction of the *TYR* gene with MTX is refractive to re-induction by MTX.

For the first time, we demonstrated the phenomenon akin to transcription memory in a human system. The *TYR* gene was chosen as a human specific example of inducible gene

where we suspected promoter education may lead to transcription memory (tanning response is specific to humans). Following an initial round of induction by MTX, we saw increased binding by MITF, USF1, USF2 and RNAPII at the *TYR* enhancer. As it is physically impossible for MITF and USF to occupy the *TYR* enhancer M-box simultaneously and unlike the situation observed at the *S.cerevisiae PHO5* gene where we saw sequential binding by each factor following induction, we reasoned that MITF and USF might exchange at the *TYR* promoter following induction. This continual exchange prior to each round of transcription burst is likely required as interfering with USF binding through A-USF also abolish binding by MITF following UVB induction. As *TYR* mRNA increases with a slight time lag to MITF and USF binding, either coordinate exchange of MITF and USF is required for each round of transcription or both MITF and USF can drive *TYR* gene expression. In the latter case, the increased MITF and USF signal may simply reflect a more permissive chromatin structure for MITF and USF binding and that both can drive *TYR* gene expression. As ChIP is a population average technique, and single cell analysis technique is not immediately available, we will address this question by using a tet-inducible system where we massively overexpress either MITF, USF1 or USF2 and investigate *TYR* M-box occupancy and gene expression. If continual exchange is required for each round of transcription, we should see a decrease expression following doxycycline induction of either USF or MITF.

Interestingly, re-induction of the *TYR* gene by MTX, even after allowing a $96\ \mathrm{h}$ recovery period, did not result in increased binding at the *TYR* enhancer nor did we see any increase in *TYR* gene expression. Although we observed a spike in the MITF and USF ChIP signal at the 3'UTR corresponding to $24\ \mathrm{h}$ post re-induction with MTX, the absence of *TYR* transcript accumulation is unlikely to be due to transcription interference as there was no corresponding RNAPII ChIP-signal at the 3'UTR at this time point. We suspect that this may be due to dulled induction of MITF in response to the second round of induction by MTX, although this still needs to be assessed.

Regulation of USF2 subcellular localisation

USF2 may be cytoplasmically localised in a manner dependent on the N-terminal domain.

While MITF, MYC and USF1 are almost exclusively nuclear in 501mel; USF2, as detected by the N18-antibody, is cytoplasmic localised. As far as we are aware, there is no report of different sub-cellular localisation of USF2 in a manner dependent on the N- and the C-terminal domains, let alone the observation that USF2 can be primarily localised to the cytoplasm. More specifically, we demonstrated that USF2 species, as detected by two different antibodies targeting the N-terminal region of USF2 are primarily localised to the cytoplasm while four different antibodies targeting the C-terminus and the central region of USF2 are primarily localised to the nucleus in 501mel cells. This cytoplasmic localisation of USF2 is most likely regulated as we observed different ratio of nuclear:cytoplasmic localisation across different cell lines.

USF2 species detected by the N-terminal specific antibodies (N18 & Ab) display slower migration on SDS-PAGE than the species detected by the C-terminal antibody (C20), indicative of cytoplasmic species being more modified relative to the nuclear species. Remarkably, however, although both the N- and C-terminal antibodies are able to detect heterologously expressed full-length canonical form of USF2, the bands detected by the N- and the C-terminal antibodies on Western-blots are mutually exclusive when probing for the endogenous USF2. Puzzlingly, however, the N18-antibody is able to detect the C20-specific specie of USF2 in CoIP experiments using both N18- and C20-antibodies. A possible explanation is that the bands detected by the N-terminal antibodies are modified at the C-terminus such that it can no longer be detected by the C20-antibody and vice versa.

Whilst it is tempting to speculate the existence of a novel USF2 isoform such that the C20-specific isoform lacks the first 18 –amino-acids, this is unlikely as the N18-antibody is able to detect the C20-specific isoform in the CoIP samples but not the endogenous one. A plausible explanation is that the C20-specific isoform harbours a labile PTM at the N-

terminus that interfere(s) with the recognition by the two N-terminal specific antibodies. This PTM may be lost during the preparation of CoIP samples.

The N18-specific species may correspond to a novel USF2 isoform lacking the C-terminal $20-$amino-acids and/or harbour a stable, large PTM (as the N18-specific band is $\approx 10\ \mathrm{kDa}$ larger than C20-specific species) that interferes with detection by C20-antibody.

Cytoplasmic USF2 shuttles into the nucleus in response to glucose.

USF1 and/or USF2 protein levels can be induced in response to high glucose, although the specific details appear to be cell line specific (Bidder et al., 2002; Shi et al., 2008; van Deursen et al., 2008; Wang et al., 2004). We have shown, for the first time, specific regulation of USF2 by glucose at the level of nuclear-cytoplasmic shuttling. Following $24\ \mathrm{h}$ glucose starvation there is an increased in the cytoplasmic localisation of USF2 N18-species and a corresponding decreased in the nuclear fraction. Glucose add-back following starvation results in a shift from the cytoplasm to the nucleus, the kinetics of which is dependent on cell line and the length of starvation as well as the amount of glucose added-back. Interestingly, we did not observed this shuttling in the 293 FlipIn/TRex cell line. Thus, USF2 nuclear-cytoplasmic shuttling in response to glucose may be pigment cell-specific.

The putative nuclear localisation signals in the basic and the USR domain are insufficient for complete nuclear localisation of USF2.

In contrast to the previously published observations (Luo and Sawadogo, 1996b), we found that both the USR and the basic domain are insufficient for nuclear localisation. USF2 mutant lacking the basic-domain is almost exclusively nuclear. The amino-acids surrounding the USR- and basic-domain positively and negatively influence the nuclear-cytoplasmic localisation of USF2 in a manner that some regions can act dominantly over the other (Figure 4.2). In general, the USR- and the basic-domain with amino-acids $143-193$ or amino-acids $248-261$ in combination with amino-acid N- and C-terminal immediately adjacent to the USR- and the basic-domain is required for total nuclear localisation.

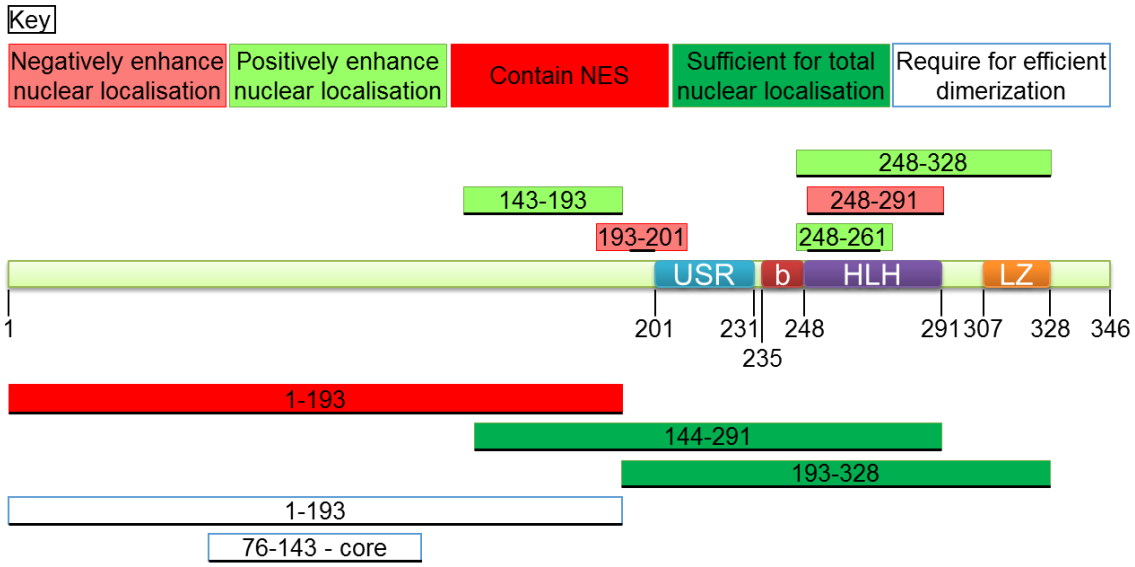


Figure 4.2 Schematic diagram summarizing regions influencing the nuclear and cytoplasmic localisation and dimerization of USF2. Regions surrounding the USR and the basic-domain positively (light green) and negatively (pink) influence the relative distribution between nuclear and cytoplasmic compartment of USF2. These regions are presented on the diagram such that the boxes which are higher up exert dominant effect on the region indicated in the box lower down. For example, amino-acids $248 - 328$ when presented with amino-acids $193 - 201$ resulted in a predominantly nuclear localised USF2. Red-box indicates region which appears to harbour nuclear export signal (NES). Green box indicates minimum amino-acids required for total nuclear localisation. White box with blue outlined indicates region requires for efficient dimerization. The minimum region required for this interaction is labelled underneath as the "core" sequence.

Exon $1 - 5$ of USF2 may be required for efficient dimerization.

In addition to the HLH-LZ domain, we showed that dimerization of USF2 may be complemented by amino-acids $1 - 193$ N-terminal of the USR. The amount of USF2 proteins pulled-down with full-length USF2 is approximately double when the HLH-LZ is presented in conjunction with the N-terminal 193 amino-acids. The minimum region required for this interaction lies between amino-acids $76 - 143$ (Figure 4.2).

Concluding remark

Through comparative analysis of seven ChIP-seq datasets we demonstrated a high degree of overlapping targets specificity by MITF, USF, MYC and HIF. It remains to be seen whether continuous exchange of these factors at a co-occupied target E-box is necessary for maximal gene expression. Although we have demonstrated differential promoter occupancy and induction kinetic at the *TYR* gene following initial and re-induction by MTX, the underlying mechanisms for the differences remain to be elucidated. Similarly, the effect of interfering with one or more of the factors at different stages of promoter education during

the initial round and re-induction stage has not been fully addressed in this thesis. We have, however, developed tools to address these questions both for induction of one specific factor (in contrast to multiple bHLH-TFs in this thesis) as well as co-induction with the dn-TF(s). These tools will enable us to dissect the role of each of the factors during promoter education as well as identifying genes that retain transcription memory.

5. Future directions

Study of factor exchange at genes through induction of/interfering with specific bHLH-TFs

We have demonstrated factor exchange at *TYR* gene using physiological cues such as α-MSH and UVB-irradiation. However, these cues result in activation of multiple bHLH-TFs and elicit a number of biological responses such as DNA-damage and p38 responses that may complicate the interpretation of the results.

During the preparation of this thesis, we developed a tetracycline (Tet)-inducible system that can be used as an alternative to the 4 −OHT inducible system, used in this thesis. With the Tet-inducible system, we are able to specifically induce chosen bHLH-TFs to a variable level and use it to investigate the consequence on factor exchange and gene expression. We are currently generating cell lines which co-expressed bHLH-TFs of interest with one or more of the dn-TFs (A-USF, A-MAX and A-MITF) to investigate the effect of interfering with one or more bHLH-TF(s) on gene activation following specific induction of other bHLH-TFs.

Using the data we have generated from our ChIP-seq we are able to pull out genes co-occupied by any combination of MYC, MITF, HIF and USF in 501mel cells for study with the Tet-inducible system. We can also use our differential gene expression data from the RNA-seq datasets as reference to investigate the effect of interfering or inducing one or more of the bHLH-TFs with the Tet-inducible system.

Ultimately we wish to address the following questions:

1) **What are the consequences of monopolising a promoter which is co-occupied by multiple bHLH-TFs with one bHLH-TF?** We have shown that in all instances of induced factor exchange at *TYR* enhancer M-box, there were concomitant increased occupancy by USF and MITF. This suggested that exchange of USF and MITF may be required for each round of *TYR* gene activation. In this case, monopolising the *TYR* promoter with, say MITF alone should interfere with *TYR* gene expression.

2) **What are the consequences of interfering with one or more of the bHLH-TF(s) co-occupying a promoter?** Interfering with any of the co-occupying bHLH-TFs may prevent efficient gene activation in the case where continuous exchange is required for maximal gene expression. If each bHLH-TF co-occupying the promoter has (a) different role(s) as outlined in the promoter education and transcription memory hypothesis, interfering with the exchange of one or more factors may affect the induction and re-induction kinetic. Similarly, interfering with a memory/bookmarker factor may result in a slower re-induction kinetic. One hypothetical situation would be the exchange between USF and MITF at pigmentation-related genes. USF which is involved in the maintenance of active and primed epigenetic marks may act as a promoter education factor while MITF with a potent transcription activation domain may act as a real activator that requires USF to prime the promoter for MITF binding. In the absence of USF, the promoter may become occluded by nucleosomes preventing binding by MITF when it is induced.
3) **Is continuous exchange of bHLH-TFs required for expression of housekeeping genes?** Our ChIP-seq data revealed a large number of co-occupied sites. At least in the case of co-occupancy with HIF, we may attribute factor exchange with HIF to be illustrative of gene regulation in response to different induction cues. Sites co-occupied by generic TFs such as MYC and USF, however, may represent sites that require continuous exchange of MYC and USF for maximal gene expression. Analysis of E-box occupancy and the corresponding expression of some MYC and USF co-occupied targets under conditions where we induce A-USF or A-MAX should address this.

CpG methylation and binding of MITF, MYC, HIF and USF

CpG methylation has been implicated in regulation of many cell-specific gene expression during development by blocking binding of many TFs (Ehrlich, 2003; Shiota, 2004) especially relevant for ubiquitously expressed ones such as c-Myb (Klempnauer, 1993), MYC (Prendergast et al., 1991), MAX (Prendergast and Ziff, 1991) and E2F (Campanero

et al., 2000). Moreover, CpG methylation can lead to binding by MeCP1 and MeCP2 which result in histone deacetylation and chromatin compaction (Bird and Wolffe, 1999).

Evidence for the effect of CpG methylation on DNA binding by USF is conflicting. One report indicated that CpG methylation at the E-box does not inhibit USF binding, but it inhibits USF-mediated expression at the *SVCT2* gene (Qiao and May, 2011). On the other hand, EMSA using oligonucleotides containing selectively methylated CpG within the E-box' demonstrates decreased binding affinity by USF1:USF2, at least in the context of the *FSHR* promoter (Griswold and Kim, 2001). If CpG methylation affects USFs' binding such that the dwell time decreases but not the steady-state levels, ChIP would still reveal USF binding. However, reducing dwell time on a promoter may interfere with USF's ability to recruit co-activators resulting in reduced gene expression as observed at the *SVCT2* gene.

To address the effect of CpG methylation on binding kinetics we are preparing samples for ChIP-seq experiments of USF and HIF in DNMT-null and the corresponding wild-type cells in collaboration with Skirmantas Kriaucionis (LICR, Oxford). We will first address the changes in relative peak score across different target E-boxes and correlate this with CpG methylation status and differential gene expression between the two cell lines. We anticipate the existence of a critical threshold dwell time of factor binding for gene activation.

Regulation of USF2 subcellular localisation

Although we identify glucose as a compound that increases USF2 nuclear accumulation, we did not observe complete shuttling of USF2 into the nucleus. This implies other conditions exist to induce USF2 nuclear accumulation and some of these may act in an additive manner. To investigate this, we plan to use a high-throughput, immunofluorescence base screening approach against a library of small molecules. Alternatively, we may adopt some of the bar-coded Cas9-library to screen for genes that affect USF2 subcellular localisation when knocked-out. We will also use MS-based peptide sequencing to obtain full peptide sequence of the USF2 species detected by the C20-, N18- and Ab-antibodies.

6. Supplementary Figures and Tables

Box-plots

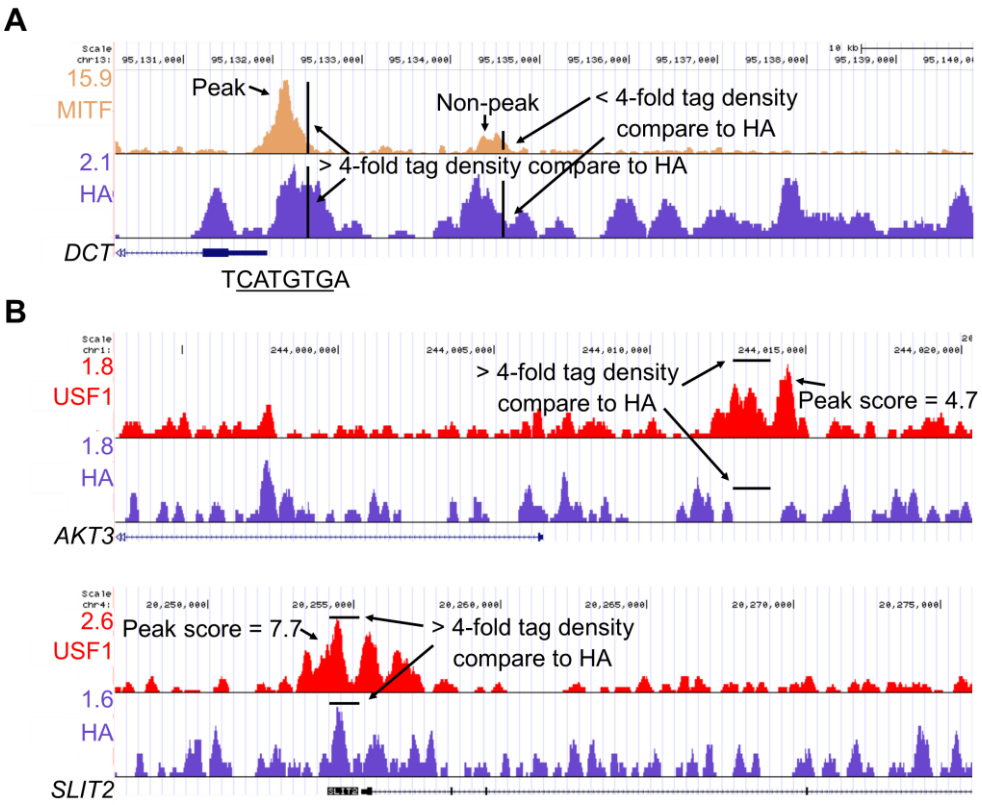
All box-plots are presented as standard “box and whisker plots” such that the lines on the box represent data from the second quartile ($25-50\%$), median and third quartile ($50-75\%$), respectively. The whisker to the left of the box demonstrates the spread of the lowest quartile (first 25%) while the whisker to the right of the box demonstrates the spread of the highest quartile (top 25%). Dots outside the whisker represent the outliers.

Motif finding

Table of Motifs found are presented as a motif logo, source of sequence used, best-matched motif from the database [presented as TF (Cell lines where the motif was derived from, database which the matched motif is taken from)] and P-value. In the case where multiple motifs have P-value of ≤ -70, only motifs with P values of ≤ -70 are reported.

In the case of *de novo* motif identification from peaks stratified by genome ontology, at least one motif will be reported, unless the top motif has a P-value of ≥ -10. For the analysis of motifs at the unique and co-occupied sites, percentage of input sequence containing the motif are also given.

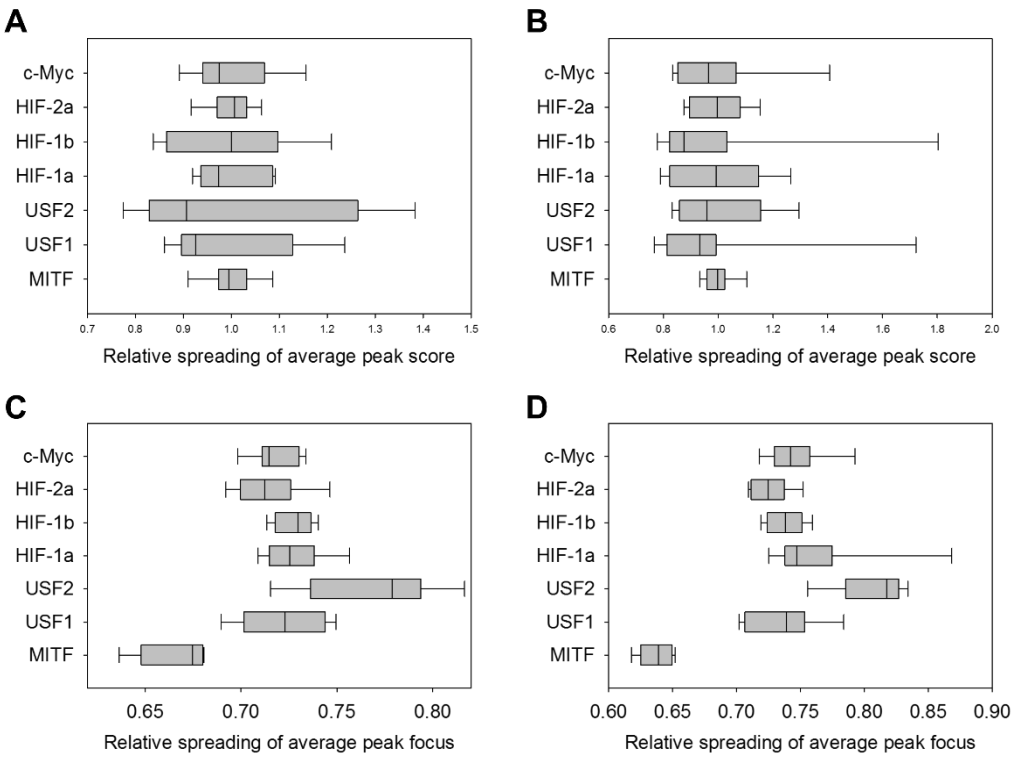
6.1 ChIP-seq analysis of bHLH-TFs



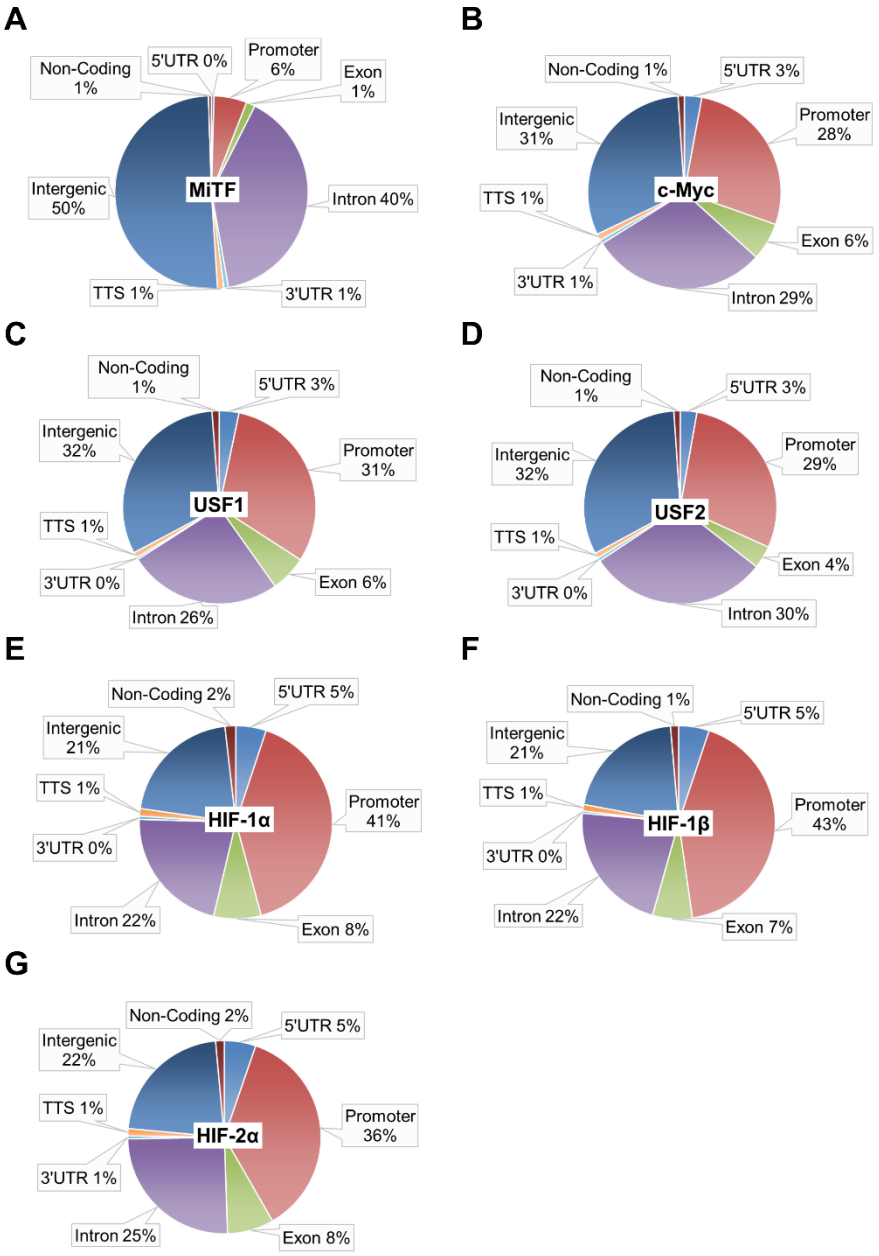
Supplementary Figure 3.1.1 Examples of peak and non-peak regions. Snapshots of bedgraphs generated from mapped ChIP-seq files and uploaded on to the UCSC genome browser. **A**) Snapshot of MITF ChIP-seq at *DCT* promoter demonstrating regions called as peak and non-peak based on regions with ≥ 4 –fold tag-density compare to background (HA-ChIP-seq) files within ± 5 kb and has a cumulative Poisson p-value of ≤ 0.0001. For simplicity of illustration, straight vertical lines were used to label a region called peak against HA-background as well as a region that appear like a peak but not peak-called. **B)** As HA-background ChIP-seq experiments were performed using a different protocol by Siddaway in our lab and submitted at a different time (same as MITF ChIP-seq datasets), we noticed that experimental-noise as well as different sample preparation resulted in different biased in the background files generated. Based on visual inspection of bedgraph files uploaded on to the UCSC genome browser, we arbitrarily place a filter to only keep peaks with score of ≥ 10 to remove peaks which maybe false positive for some of the subsequent analysis. Exemplified here are regions peak-called at *AKT3* (peak score $=$ 4.7) and *SLIT2* (peak score $=$ 7.7) genes. Horizontal black bars indicate regions called as peak in the sample (exemplified with USF1) and the equivalent region in the background (HA).

Supplementary Table 3.1.1 Summary of peaks called and peak focus of ChIP-seq experiments.

ChIP-seq	No. of peaks ≥4-fold enriched	Average peaks score	Average peak focus	No. of peaks with peak score ≥10	% of peaks with peak score ≥10
HIF-1α	5,095	10.96	0.73	1,507	29.6
HIF-1β	12,659	18.46	0.73	8,993	71.0
HIF-2α	8,108	10.78	0.71	3,094	38.1
c-Myc	9,980	8.36	0.72	1,503	15.1
MITF	99,380	2.31	0.68	237	0.2
USF1	13,073	12.20	0.73	5,685	43.5
USF2	11,780	15.66	0.78	5,911	50.1



Supplementary Figure 3.1.2 Box plots of peak score and peak focus. Average peak score for different genome ontology were centred normalised and box-plotted to demonstrate the relative spreading of peak score across different genome ontology for full dataset (**A**) and peaks with score of ≥ 10 (**B**). **C)** Box plot of average peak focus for full dataset and peaks with score of ≥ 10 (**D**). Note that the top 6% peaks for MITF were used for both **B)** and **D)** as only $0.2\ \%$ of peaks has score of ≥ 10.



Supplementary Figure 3.1.3 Basic genome ontology of ChIP-seq peaks generated from full ChIP-seq datasets for MITF (**A**), MYC (c-Myc) (**B**), USF1 (**C**), USF2 (**D**), HIF-1α (**E**), HIF-1β (**F**) and HIF-2α (**G**). The basic genome ontology classified peaks into either: 5'-Untranslated region (5'-UTR), promoter, exon, intron, 3'-UTR, transcription termination sites (TTS), intergenic or non-coding.

Supplementary Table 3.1.2

Genes	Protein name	Brief Description	Link
ASNA1	ATPase ASNA1	ATPase required for the delivery of tail-anchored (TA) proteins to the ER	http://www.uniprot.org/uniprot/O43681
BAX	Bax protein	Activation of cysteine-endopeptidase activity involved in apoptotic signalling pathway, cellular response to UV	http://www.uniprot.org/uniprot/Q5ZPJ1
DVL2	dishevelled homolog DVL-2	Plays a role both in canonical and non-canonical Wnt signalling	http://www.uniprot.org/uniprot/O14641
FAM21B	WASH complex subunit FAM21B	Component of the WASH complex, presented at the surface of endosomes that recruits and activates the Arp2/3 complex to induce actin polymerization.	http://www.uniprot.org/uniprot/Q5SNT6
HEXA	Beta-hexosaminidase subunit alpha	Responsible for the degradation of GM2 gangliosides, and a variety of other molecules containing terminal N-acetyl hexosamines.	http://www.uniprot.org/uniprot/P06865
ILF3	Interleukin enhancer-binding factor 3	May facilitate dsRNA-regulated gene expression post-transcriptionally. Can act as a translation initiation inhibitor. Can regulate protein arginine N-methyltransferase 1 activity. Can promote the formation of stable DNA-dependent protein kinase holoenzyme complexes on DNA.	http://www.uniprot.org/uniprot/Q12906
LUC7L2	Putative RNA-binding protein Luc7-like 2	mRNA splice site selection	http://www.uniprot.org/uniprot/Q9Y383
NEURL2	Neuralized-like protein 2	Plays an important role in myofiber differentiation and maturation. Probable substrate-recognition component of a SCF-like ECS E3 ubiquitin-protein ligase.	http://www.uniprot.org/uniprot/Q9BR09
PLAG1	Zinc finger protein PLAG1	Transcription factor whose overexpression results in up-regulation of target genes, such as IGFII, leading to uncontrolled cell proliferation. Overexpression is frequently observed in hepatoblastoma.	http://www.uniprot.org/uniprot/Q6DJT9
PLD3	Phospholipase D3	Phosphatidylcholine-hydrolyzing enzyme	http://www.uniprot.org/uniprot/Q8IV08
QTRT1	Queuine tRNA-ribosyltransferase	Interacts with QTRTD1 to form an active queuine tRNA-ribosyltransferase. This enzyme exchanges queuine for the guanine at the wobble position of tRNAs with GUN anticodons (tRNA-Asp, -Asn, -His and -Tyr), thereby forming the hypermodified nucleoside queuosine (Q) (7-(((4,5-cis-dihydroxy-2-cyclopenten-1-yl)amino)methyl)-7-deazaguanosine)	http://www.uniprot.org/uniprot/Q9BXR0
RNASEK	Protein RNASEK-C17orf49	Inferred from electronic annotation as DNA-binding protien	http://www.uniprot.org/uniprot/H0YIS7
RPUSD4	RNA pseudouridylate synthase domain-containing protein 4	Inferred from electronic annotation as pseudouridine synthesis	http://www.uniprot.org/uniprot/Q96CM3
TNPO2	TNPO2 protein	Intracellular protein transport	http://www.uniprot.org/uniprot/Q05D48

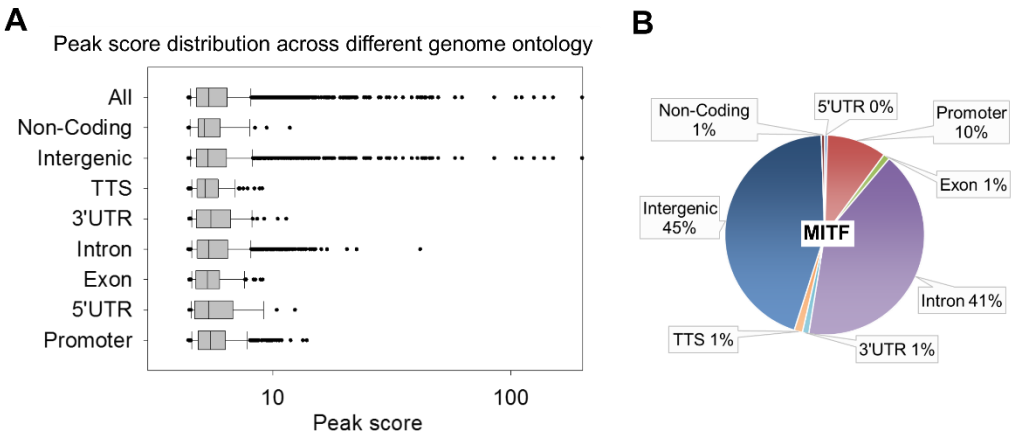
List of target genes selected for ChIP-qPCR to validate ChIP-seq result.

Supplementary Table 3.1.3

	ChIP-seq peak score				Normalised score				
Genes	**MITF**	**MYC**	**USF1**	**USF2**	**MITF**	**MYC**	**USF1**	**USF2**	**E-box(es) within the peak**
ASNA1	4.5			26.0	1.9			1.7	TCACGTGA 570 GCACGTGA
BAX		17.0	122.5	188.1		2.0	10.0	12.0	TCACGTGA 27 GCAGGTGCGGTCACGTGA 39 CCACGTGA 10 TCACGTGA
DVL2	8.8	13.6	113.2	70.5	3.8	1.6	9.3	4.5	TCACGTGA 39 TCACGTGA 11 TCACGTGA 26 CCACATGA
FAM21B									TCACGTGC 8 TCACGTGA
HEXA	2.5	9.8	84.3	196.1	1.1	1.2	6.9	12.5	CTACGTGC 46 TCACGGGC
ILF3	8.0	15.9	99.9	146.5	3.5	1.9	8.2	9.4	TCACGTGG 48 TCACGTGA 77 TCACATGA 39 TCACGTGA
LUC7L2		10.6	16.8	17.6		1.3	1.4	1.1	TCACGTGG
NEURL2	7.7	12.9	81.5	154.4	3.3	1.5	6.7	9.9	TCACGTGGCGCCGGAGTTCACGTGACTCGTACACATGA
PLAG1	6.4		45.6	50.2	2.8		3.7	3.2	CCACGTGC 85 TCACGTGA 19 CCACGTGG
PLD3	8.3		80.3	63.9	3.6		6.6	4.1	TCACGTGA 37 TCACGTGA
QTRT1	6.6		61.7	58.9	2.9		5.1	3.8	TCACGTGA 38 CCACGTGG
RNASEK	8.3		80.6	18.6	3.6		6.6	1.2	TCACGTGA 9 TCACGTGA
RPUSD4	4.9	9.8		14.1	2.1	1.2		0.9	TCACGTGG 22 TCACGTGG 44 CCACGTGG 39 CCACCTGG
TNPO2	6.3	6.4	43.8	32.7	2.7	0.8	3.6	2.1	ACACGTGC 26 TCACGTGA

Table listing raw peak score, score normalised to average peak score for direct comparison and the E-box(es) motifs found under the peaks of genes used for ChIP-qPCR to validate ChIP-seq results. The core E-box is underlined. The number between each E-box indicates the distance from the first E-box to the next.

6.1.1 MITF ChIP-seq



C

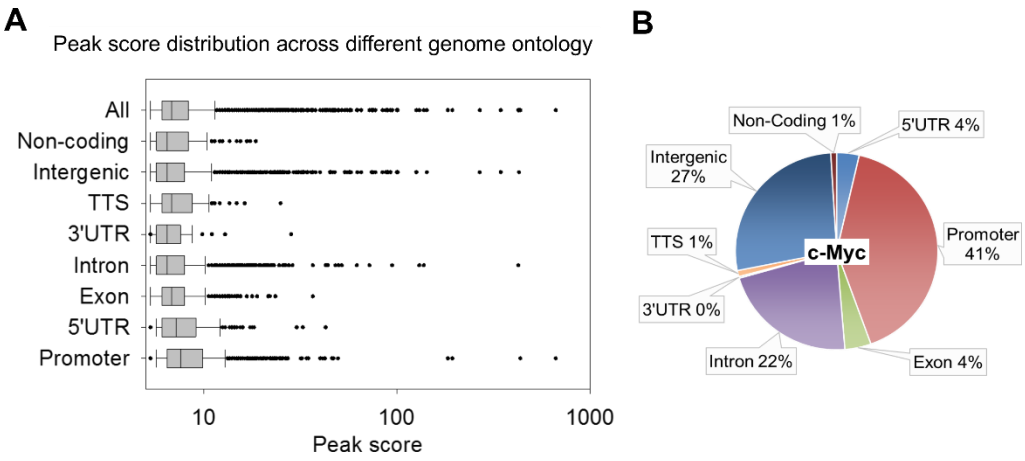
Motif	Peak location	Best matches	P-value
	All	USF2 (C2C12 Homer)	$1e^{-1855}$
	Promoter	E-box (Homer)	$1e^{-418}$
	5'UTR	USF2 (C2C12 Homer)	$1e^{-22}$
	Exon	USF2 (C2C12 Homer)	$1e^{-33}$
	Intron	USF2 (C2C12 Homer)	$1e^{-794}$
	3' UTR	USF2 (C2C12 Homer)	$1e^{-30}$
	TTS	E-box (Homer)	$1e^{-30}$
	Intergenic	USF2 (C2C12 Homer)	$1e^{-633}$
	Non-coding	USF1 (GM12878 Homer)	$1e^{-16}$

Supplementary Figure 3.1.4 Basic bioinformatics of MITF ChIP-seq. **A)** A box-plot of peak score across full-dataset stratified into different genome ontologies. **B)** Pie chart showing the relative distribution of top 6 % of peaks (by peak score) across different genome ontologies. **C)** E-box motifs identified de novo from top 6 % of peaks stratified into different genome ontologies

Motif	Peaks location	Best matches	P-value
	All	USF1 (GM12878 Homer)	$1e^{-7694}$
	Promoter	USF1 (GM12878 Homer)	$1e^{-932}$
	Promoter	Sp1 (Homer)	$1e^{-101}$
	Promoter	Elk1 (HeLa Homer)	$1e^{-71}$
	Promoter	NRF1 (MCF7 Homer)	$1e^{-71}$
	5'UTR	USF1 (GM12878 Homer)	$1e^{-47}$
	Exon	USF2 (C2C12 Homer)	$1e^{-245}$
	Intron	USF2 (C2C12 Homer)	$1e^{-6066}$
	Intron	Sox2 mES	$1e^{-95}$
	Intron	Sox10 (Jaspar)	$1e^{-88}$
	Intron	JUN:FOS (Jaspar)	$1e^{-85}$
	3' UTR	USF2 (C2C12 Homer)	$1e^{-176}$
	TTS	USF1 (GM12878 Homer)	$1e^{-189}$
	Intergenic	USF1 (GM12878 Homer)	$1e^{-7212}$
	Intergenic	AP1 ThioMac-PU1	$1e^{-85}$
	Non-coding	USF1 (GM12878 Homer)	$1e^{-95}$

Supplementary Figure 3.1.5 Motifs identified de novo from MITF full-dataset stratified into different genome ontologies.

6.1.2 MYC ChIP-seq



C

Motif	Peak location	Best matches	P-value
	All	USF1 (GM12878 Homer)	$1e^{-83}$
	Promoter	USF1 (GM12878 Homer)	$1e^{-57}$
	5'UTR		$1e^{-5}$
	Exon	GFY (Jaspar)	$1e^{-11}$
	Intron	USF1 (GM12878 Homer)	$1e^{-17}$
	3' UTR	HIF-2α (785-RCC Homer)	$1e^{-5}$
	TTS		$1e^{-10}$
	Intergenic	USF1 (GM12878 Homer)	$1e^{-28}$
	Non-coding	MA0081.1_SPIB (Jaspar)	$1e^{-5}$

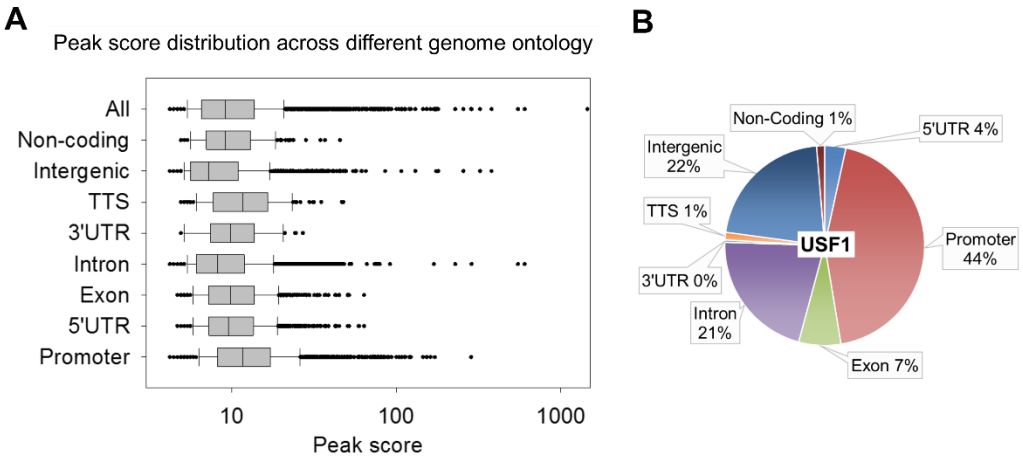
Supplementary Figure 3.1.6 Basic bioinformatics of MYC ChIP-seq. **A)** Box-plot of peak score across full dataset stratified into different genome ontologies. **B**) Pie chart showing the relative distribution of peaks with score of ≥ 10 stratified into different genome ontologies. **C)** E-box motifs identified de novo from peaks with score of ≥ 10 stratified into different genome ontologies.

Motif	Peaks location	Best matches	P-value
	All	USF1 (GM12878 Homer)	$1e^{-137}$
	Promoter	USF1 (GM12878 Homer)	$1e^{-123}$
	Promoter	Sp1 (Homer)	$1e^{-112}$
	Promoter	NRF1 (MCF7 Homer)	$1e^{-84}$
	Promoter	Elk4 (HeLa Homer)	$1e^{-84}$
	5'UTR		$1e^{-12}$
	Exon	N-Myc (mES, GSE11431)	$1e^{-24}$
	Intron	USF2 (C2C12 Homer)	$1e^{-36}$
	3' UTR	N-Myc (mES, GSE11431)	$1e^{-10}$
	TTS		$1e^{-11}$
	Intergenic	USF1 (GM12878 Homer)	$1e^{-28}$
	Non-coding		$1e^{-8}$

Supplementary Figure 3.1.7 Motifs identified de novo from the MYC full-dataset stratified into different genome ontologies.

6.1.3 USF ChIP-seq

USF1



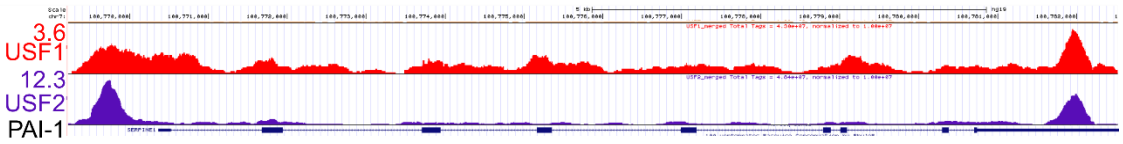
C

Motif	Peak location	Best matches	P-value
	All	USF1 (GM12878 Homer)	$1e^{-926}$
	Promoter	USF1 (GM12878 Homer)	$1e^{-461}$
	5'UTR	CLOCK (GSE 39860 Homer)	$1e^{-21}$
	Exon		$1e^{-23}$
	Intron	USF2 (C2C12 Homer)	$1e^{-220}$
	3' UTR	USF1 (GM12878 Homer)	$1e^{-11}$
	TTS	USF1 (GM12878 Homer)	$1e^{-20}$
	Intergenic	PB0007.1 (Bhlh2_1 Jaspar)	$1e^{-290}$
	Non-coding	USF1 (Jaspar)	$1e^{-9}$

Supplementary Figure 3.1.8 Basic bioinformatics of USF1 ChIP-seq. **A)** Box-plot of peak score across USF1 full-dataset stratified into different genome ontologies. **B)** Pie chart showing the relative distribution of peaks with score of ≥ 10 stratified into different genome ontologies. **C)** E-box motifs identified de novo from peaks with score of ≥ 10 stratified into different genome ontologies

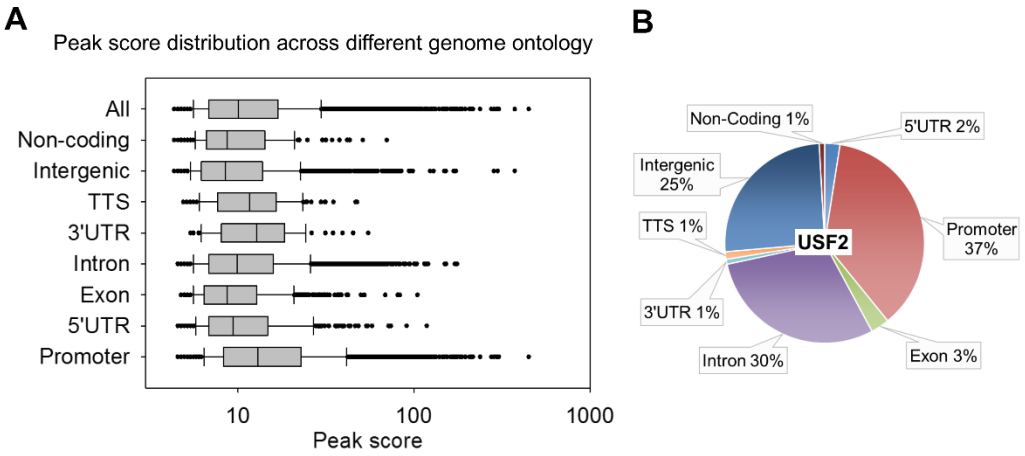
Motif	Peaks location	Best matches	P-value
	All	USF1 (GM12878 Homer)	$1e^{-2452}$
	Promoter	USF1 (GM12878 Homer)	$1e^{-497}$
	5'UTR	Max (Jaspar)	$1e^{-44}$
	Exon	USF1 (GM12878 Homer)	$1e^{-18}$
	Intron	USF2 (C2C12 Homer)	$1e^{-307}$
	3' UTR	USF1 (GM12878 Homer)	$1e^{-17}$
	TTS	USF1 (GM12878 Homer)	$1e^{-18}$
	Intergenic	PB0007.1 Bhl-h2_1	$1e^{-414}$
	Non-coding	USF1 (GM12878 Homer)	$1e^{-11}$

Supplementary Figure 3.1.9 Motifs identified de novo from the USF1 full-dataset stratified into different genome ontologies.



Supplementary Figure 3.1.10 Snapshot of UCSC file at the PAI-1 locus showing USF1 and USF2 peaks at the promoter and terminator region.

USF2



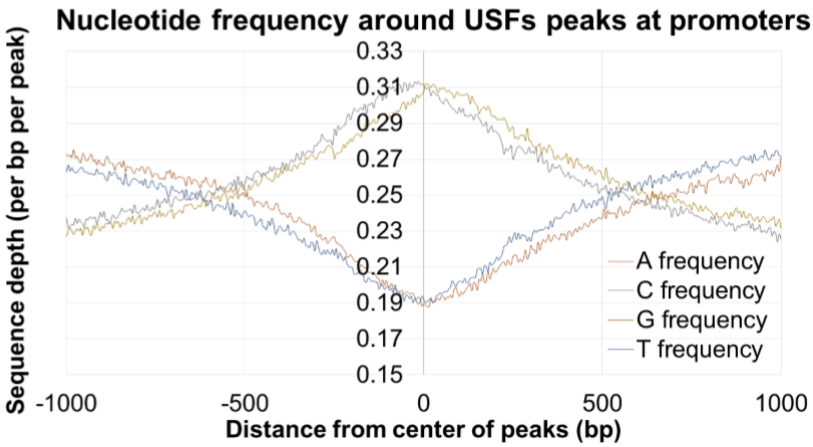
C

Motif	Peak location	Best matches	P-value
	All	USF1 (GM12878 Homer)	$1e^{-2277}$
	Promoter	USF1 (GM12878 Homer)	$1e^{-799}$
	5'UTR	BMAL1 (Liver, Homer)	$1e^{-40}$
	Exon	N-Myc (mES, GSE11431)	$1e^{-42}$
	Intron	USF2 (C2C12 Homer)	$1e^{-877}$
	3' UTR	USF2 (C2C12 Homer)	$1e^{-36}$
	TTS	c-Myc (LNCAP, Homer)	$1e^{-30}$
	Intergenic	USF2 (C2C12 Homer)	$1e^{-621}$
	Non-coding	N-Myc (mES, GSE11431)	$1e^{-18}$

Supplementary Figure 3.1.11 Basic bioinformatics of USF2 ChIP-seq. **A)** Box-plot of peak score across USF2 full-dataset stratified into different genome ontologies. **B**) Pie chart showing the relative distribution of peaks with score of ≥ 10 stratified into different genome ontologies. **C)** E-box motifs identified de novo from peaks with score of ≥ 10 stratified into different genome ontologies.

Motif	Peaks location	Best matches	P-value
	All	USF2 (C2C12 Homer)	$1e^{-999}$
	Promoter	USF1 (GM12878 Homer)	$1e^{-715}$
	Promoter	NFY (Homer)	$1e^{-215}$
	Promoter	Sp1 (Homer)	$1e^{-186}$
	Promoter	Elk4 (Hela Homer)	$1e^{-87}$
	Promoter	NRF1 (MCF7 Homer)	$1e^{-69}$
	5'UTR	USF1 (GM12878 Homer)	$1e^{-32}$
	Exon	c-Myc (LNCAP, Homer)	$1e^{-43}$
	Intron	USF2 (C2C12 Homer)	$1e^{-1133}$
	3' UTR	BMAL1 (Liver, Homer)	$1e^{-41}$
	TTS	c-Myc (mES Homer)	$1e^{-37}$
	Intergenic	USF1 (GM12878 Homer)	$1e^{-856}$
	Non-coding	USF2 (C2C12 Homer)	$1e^{-20}$

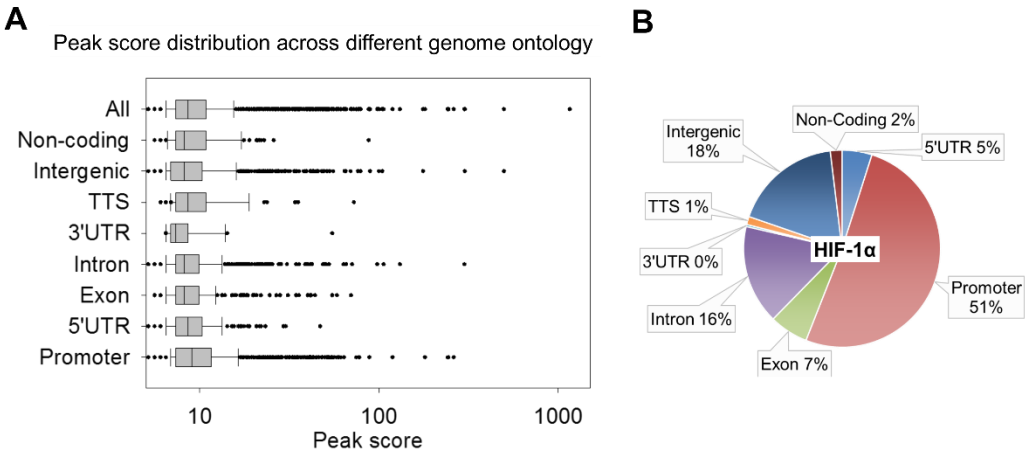
Supplementary Figure 3.1.12 Motifs identified de novo from USF2 full-dataset stratified into different genome ontologies.



Supplementary Figure 3.1.13 Nucleotide distribution around USF promoter peaks. The frequency of each nucleotide found around the USF peaks at the promoter region are binned every 5 bp and plotted as a function of distance from the centre of USFs peaks

6.1.4 HIF ChIP-seq

HIF-1α



C

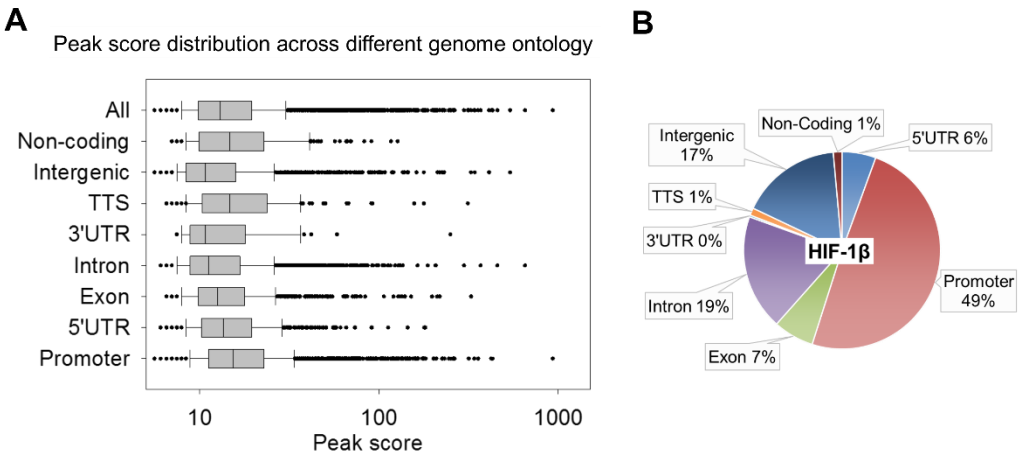
Motif	Peak location	Best matches	P-value
	All	HIF-1α (MCF7 Homer)	$1e^{-61}$
	Promoter	HIF-1α (MCF7 Homer)	$1e^{-42}$
	5'UTR		$1e^{-10}$
	Exon		$1e^{-10}$
	Intron	HIF-1α (MCF7 Homer)	$1e^{-13}$
	TTS	HIF-1α (MCF7 Homer)	$1e^{-6}$
	Intergenic	HIF-2α (785-RCC Homer)	$1e^{-14}$
	Non-coding		$1e^{-9}$

Supplementary Figure 3.1.14 Basic bioinformatics of HIF-1α ChIP-seq. **A)** Box-plot of peak score across full-dataset stratified into different genome ontologies. **B)** Pie chart showing the relative distribution of peaks with score of ≥ 10 stratified into different genome ontologies. **C)** E-box motifs identified de novo from peaks with score of ≥ 10 stratified into different genome ontologies.

Motif	Peaks location	Best matches	P-value
	All	HIF-2α (785-RCC Homer)	$1e^{-97}$
	Promoter	HIF-1α (MCF7 Homer)	$1e^{-80}$
	Promoter	Sp1 (Homer)	$1e^{-79}$
	5'UTR		$1e^{-11}$
	Exon		$1e^{-21}$
	Exon	HIF-2α (785-RCC Homer)	$1e^{-18}$
	Intron	HIF-2α (785-RCC Homer)	$1e^{-30}$
	3' UTR	HIF-2α (785-RCC Homer)	$1e^{-8}$
	TTS	HIF-1α (MCF7 Homer)	$1e^{-8}$
	Intergenic	HIF-1α (MCF7 Homer)	$1e^{-16}$
	Non-coding		$1e^{-9}$

Supplementary Figure 3.1.15 Motifs identified de novo from HIF-1α full-dataset stratified into different genome ontologies.

HIF-1β



C

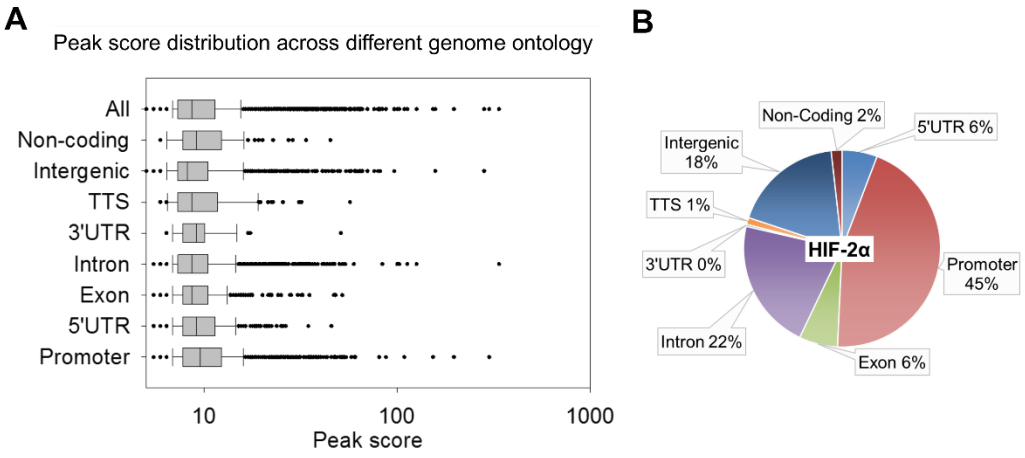
Motif	Peak location	Best matches	P-value
	All	CLOCK (Liver Homer)	$1e^{-173}$
	Promoter	Sp1 (Homer)	$1e^{-139}$
	Promoter	HIF-1α (MCF7 Homer)	$1e^{-40}$
	5'UTR		$1e^{-12}$
	Exon	HIF-1α (MCF7 Homer)	$1e^{-23}$
	Intron	HIF-2α (785-RCC Homer)	$1e^{-54}$
	3' UTR	HIF-1α (MCF7 Homer)	$1e^{-7}$
	TTS		$1e^{-11}$
	Intergenic	CLOCK (Liver Homer)	$1e^{-29}$
	Non-coding		$1e^{-8}$

Supplementary Figure 3.1.16 Basic bioinformatics of HIF-1β ChIP-seq. **A)** Box-plot of peak score across full-dataset stratified into different genome ontologies. **B)** Pie chart showing the relative distribution of peaks with score of ≥ 10 stratified into different genome ontologies. **C)** E-box motifs identified de novo from peaks with score of ≥ 10 stratified into different genome ontologies.

Motif	Peaks location	Best matches	P-value
	All	USF1 (GM12878 Homer)	$1e^{-245}$
	Promoter	Sp1 (Homer)	$1e^{-232}$
	Promoter	Elk1 (HeLa Homer)	$1e^{-157}$
	Promoter	USF1 (GM12878 Homer)	$1e^{-139}$
	Promoter	NFY (Homer)	$1e^{-116}$
	Promoter	NRF1 (MCF7)	$1e^{-97}$
	Promoter	YY1 (Homer)	$1e^{-61}$
	5'UTR	MyoD (Jaspar)	$1e^{-18}$
	Exon	Myc/Max (Jaspar)	$1e^{-43}$
	Exon	HIF-1α (MCF7 Homer)	$1e^{-31}$
	Intron	CLOCK (Liver Homer)	$1e^{-86}$
	Intron	ARNT:Ahr (Jaspar)	$1e^{-22}$
	3' UTR		$1e^{-7}$
	TTS	ATF4 (MEF, Homer)	$1e^{-13}$
	TTS	HIF-2α (785-RCC Homer)	$1e^{-10}$
	Intergenic	CLOCK (Liver Homer)	$1e^{-56}$
	Non-coding		$1e^{-10}$

Supplementary Figure 3.1.17 Motifs identified de novo from HIF-1β full-dataset stratified into different genome ontologies.

HIF-2α



C

Motif	Peak location	Best matches	P-value
	All	CLOCK (Liver Homer)	$1e^{-60}$
	Promoter		$1e^{-36}$
	5'UTR		$1e^{-10}$
	Exon	HIF-2α (785-RCC Homer)	$1e^{-10}$
	Intron	USF2 (Jaspar)	$1e^{-17}$
	3' UTR		$1e^{-5}$
	TTS		$1e^{-8}$
	Intergenic	USF2 (C2C12 Homer)	$1e^{-14}$
	Non-coding		$1e^{-9}$

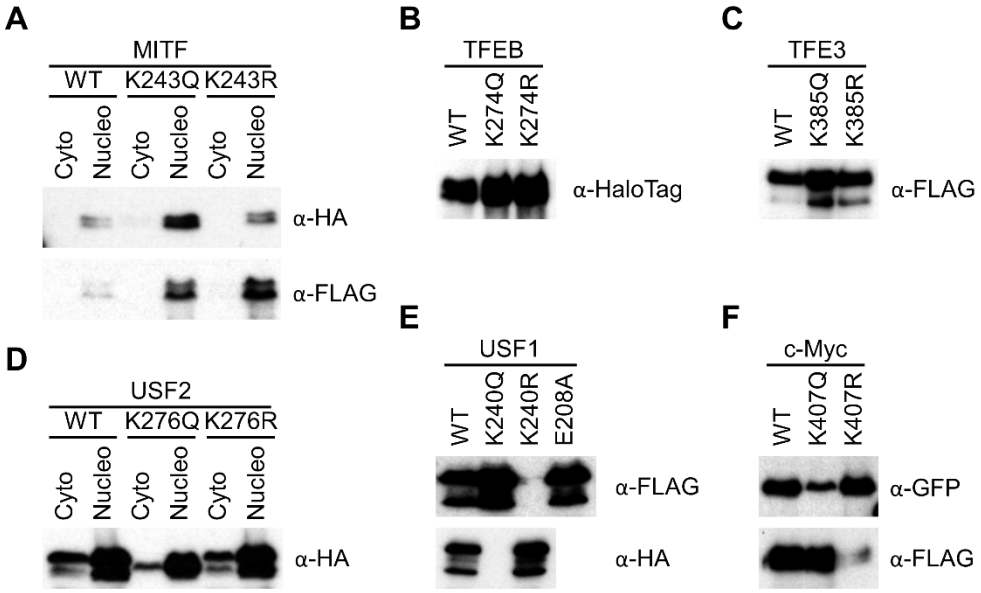
Supplementary Figure 3.1.18 Basic bioinformatics of HIF-2α ChIP-seq. **A)** Box-plot of peak score across full-dataset stratified into different genome ontologies. **B)** Pie chart showing the relative distribution of peaks with score of ≥ 10 stratified into different genome ontologies. **C)** E-box motifs identified de novo from peaks with score of ≥ 10 stratified into different genome ontologies.

Motif	Peaks location	Best matches	P-value
	All	USF1 (GM12878 Homer)	$1e^{-103}$
	Promoter	Sp1 (Homer)	$1e^{-81}$
	Promoter	USF1 (GM12878 Homer)	$1e^{-69}$
	5'UTR		$1e^{-11}$
	Exon	MyoD (GSE 21614 Homer)	$1e^{-22}$
	Intron	EBF (pro B-cell Homer)	$1e^{-29}$
	Intron	USF2 (C2C12 Homer)	$1e^{-22}$
	3' UTR		$1e^{-9}$
	TTS		$1e^{-10}$
	Intergenic	USF2 (C2C12 Homer)	$1e^{-26}$
	Non-coding		$1e^{-10}$

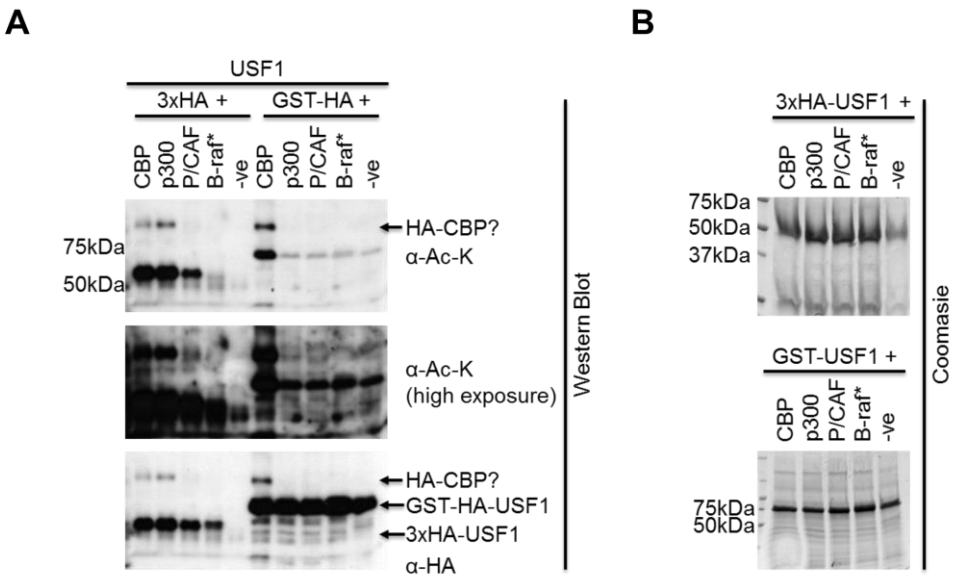
Supplementary Figure 3.1.19 Motifs identified de novo from HIF-2α full-dataset stratified into different genome ontologies.

6.1.5 Binding specificity

DNA-phosphate backbone contact and genome titration



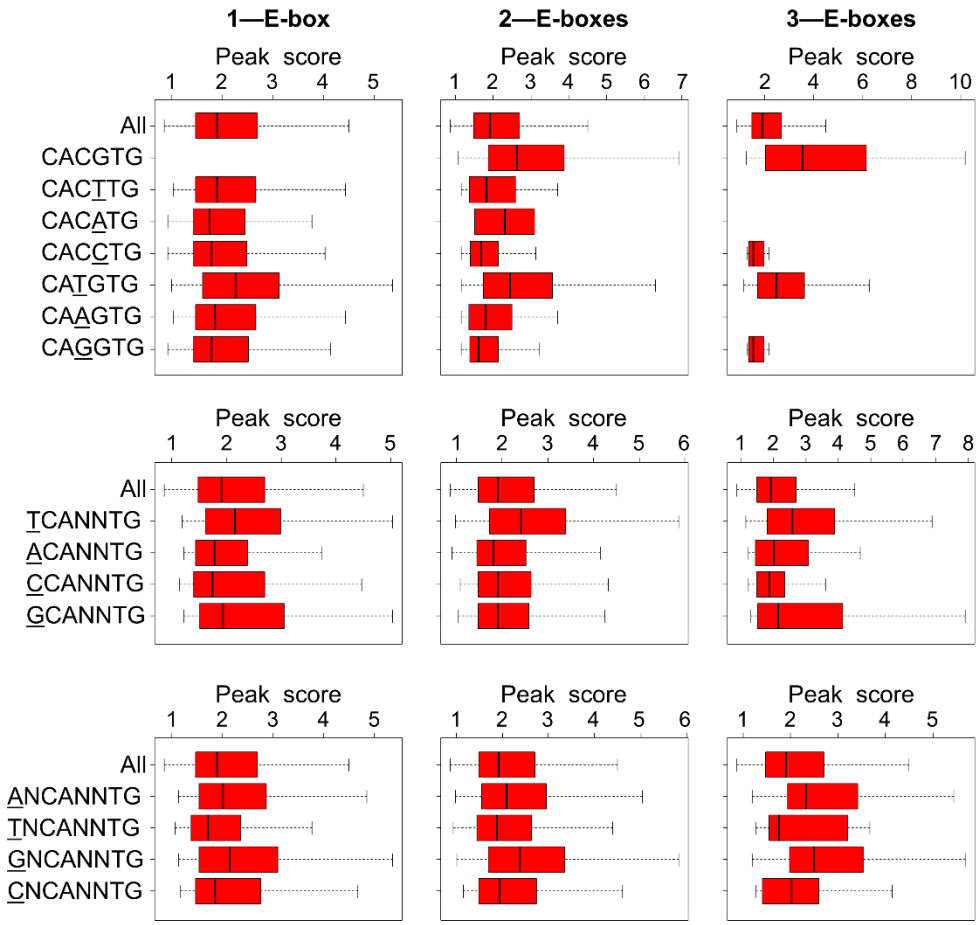
Supplementary Figure 3.1.20 Western-blotting of Phoenix cells transiently over-expressing the indicated proteins and probed with the indicated antibodies. **A)** Phoenix cell s transfected with HA-MITF and FLAG-MITF were fractionated into nucleoplasmic (nucleo) and cytoplasmic (cyto) fractions before Western-blotting. **B)** Phoenix cells transfected with Halotag-TFEB mutants and probed with TMR-dye. **C)** Phoenix cells transfected with FLAG-TFE3 mutants. **D**) Same as **A)** for 3xHA-USF2 mutants. **E)** Phoenix cells transfected with FLAG-USF1 or HA-USF1 mutants. **F)** Phoenix cells transfected with FLAG-EmGFP-c-Myc or FLAG-c-Myc.



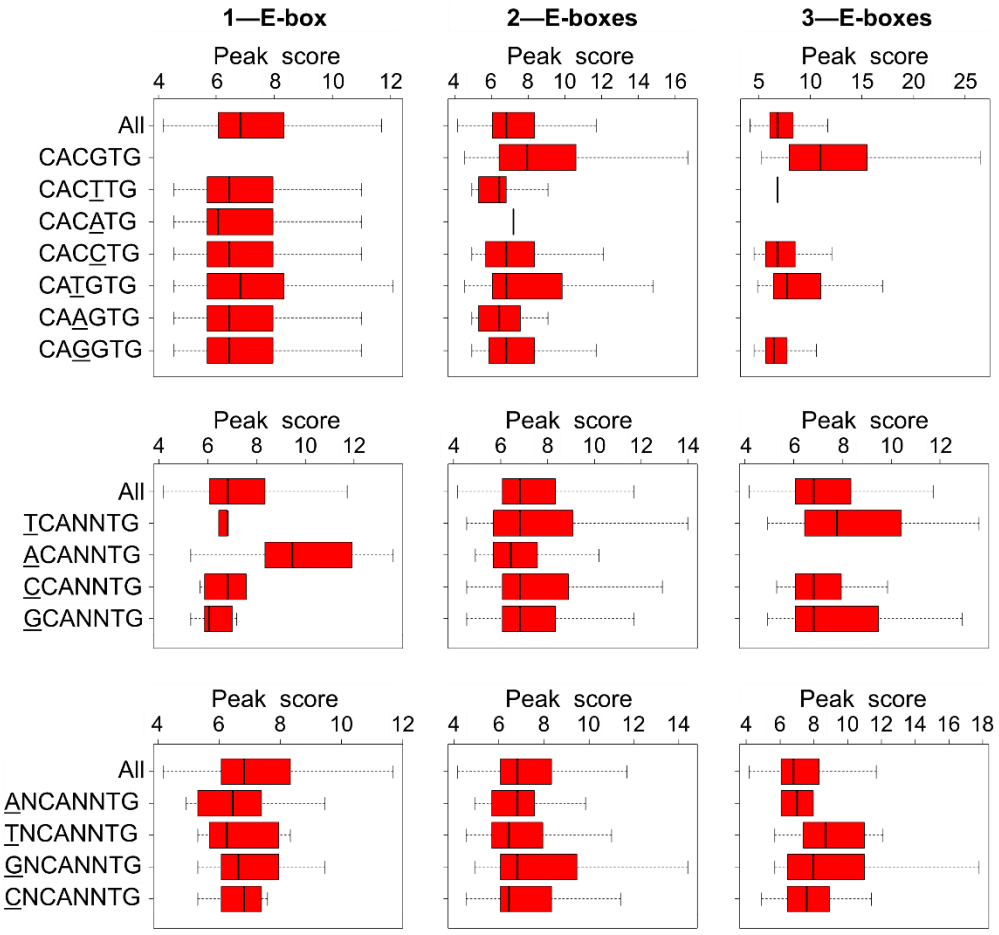
Supplementary Figure 3.1.21 Validation of MS-sample. **A)** Western-blotting of HA-USF1 and GST-HA-USF1 co-transfected with the indicated expression vectors and affinity purified before probing with either pan-acetyl-lysine antibody or HA-antibody to confirm the success of purification, the integrity of the purified protein and the acetylation level. **B)** Same sample as **A)** stained with Coomassie to check the protein level and the purity of the purified protein.

Core-E-box and the flanking sequence preference

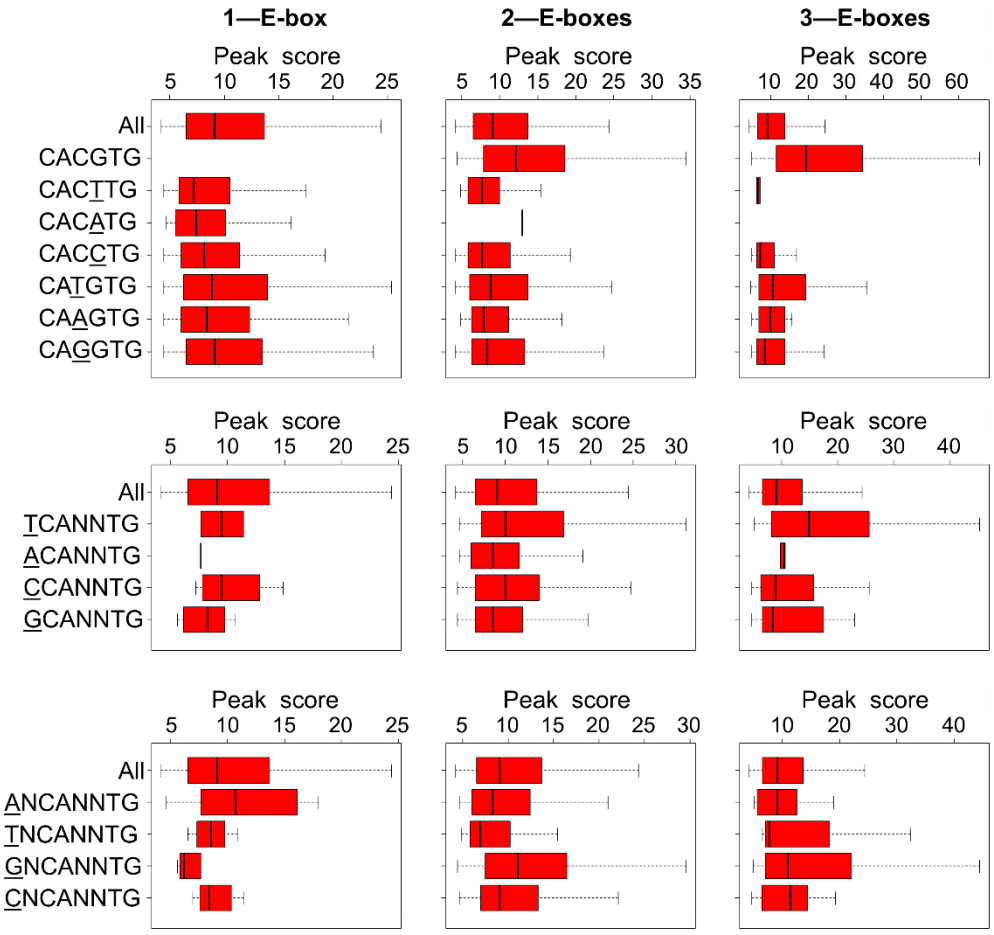
Note on the box-plots in this and the next sections: All the box-plots were plotted in R supplied with Tab-delimited data files where all the peaks were screened for the indicated E-box and conditions set such that only peaks with the specified E-box and nothing else are used for box-plots. No centering, nor any assumptions were made to adjust the peaks. All peaks were treated at the given size (in other words, the size which a region was defined as a peak when the peak was called). This is in contrast to the original analysis performed by Siddaway in our lab for MITF flanking sequences where MACs narrow peak files were intersected with genome coordinates of a specific E-box motifs using bedtools. The difference in analytical approach means that our analysis is less likely to miss out any motifs for non-centred peaks. However, the downside is that the peaks used will be of variable size and for samples with low peak focus, it may include non-biologically important motifs at the tail end of the peaks.



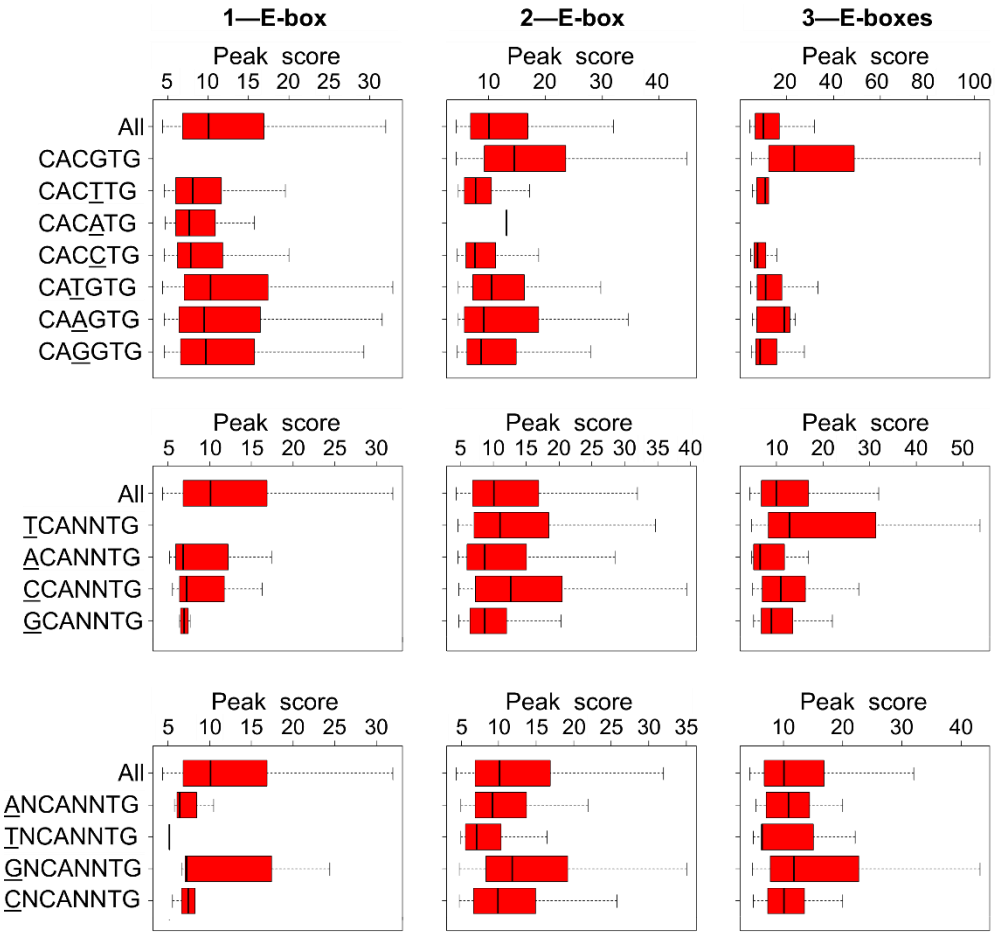
Supplementary Figure 3.1.22 Box-plot of peak score against different E-box motifs (based on the consensus motif identified de novo) for MITF. The first column represents data from peaks with one E-box, the second with two and the third with three E-boxes of the indicated type only. The first row looked at the effect of core di-nucleotides on binding affinity (strand specific). The second row looked at the effect of 5'- and 3'- nucleotides immediately adjacent to the core E-box. The third row looked at the effects of the second nearest 5'- and 3'- flanking nucleotides.



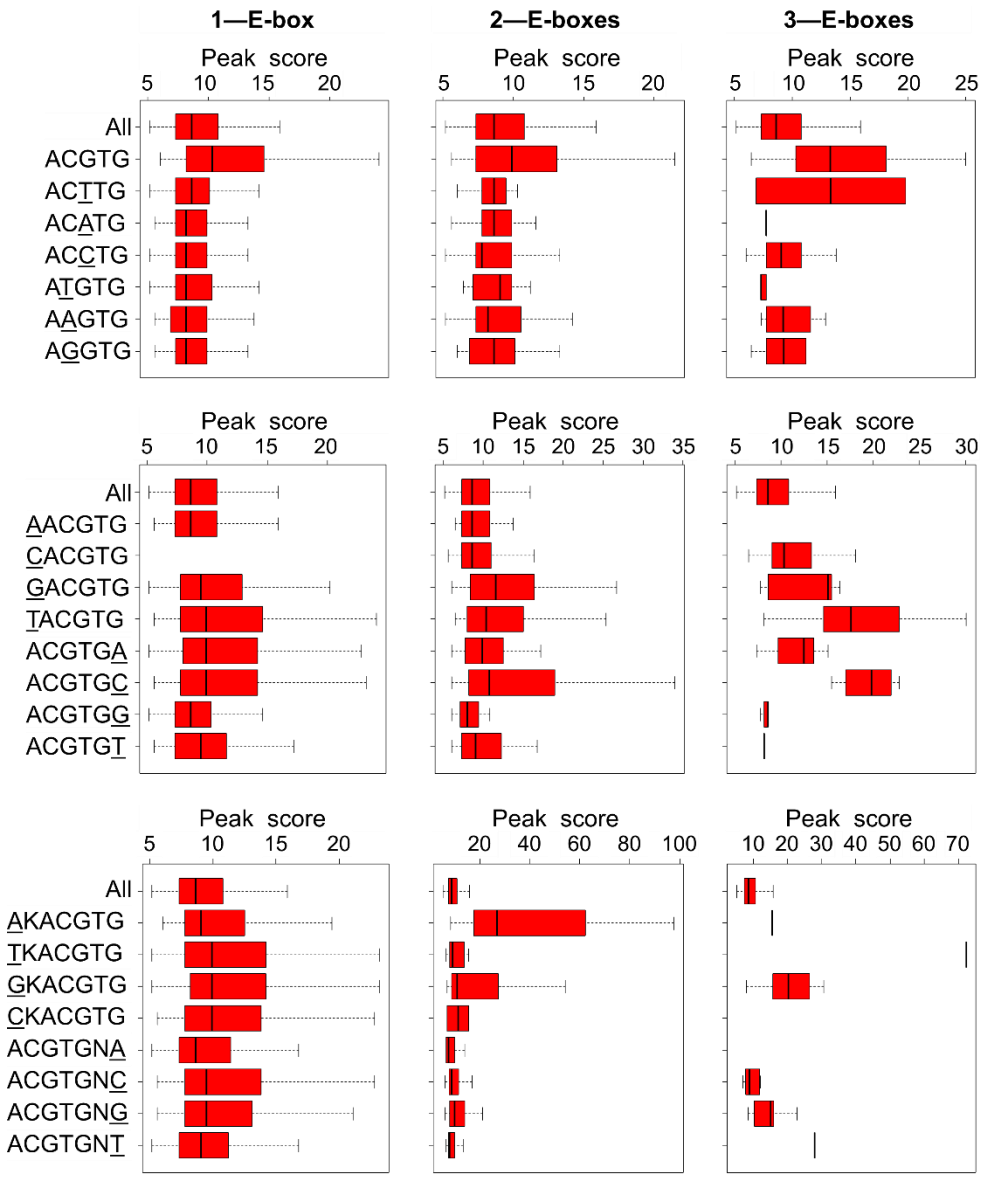
Supplementary Figure 3.1.23 Box-plot of peak score against different E-box motifs for (based on the consensus motif identified de novo) MYC. The first column represents data from peaks with one E-box, the second with two and the third with three E-boxes of the indicated type only. The first row looked at the effect of core di-nucleotides on binding affinity (strand specific). The second row looked at the effect of 5'- and 3'- nucleotides immediately adjacent to the core E-box. The third row looked at the effects of the second nearest 5'- and 3'- flanking nucleotides.



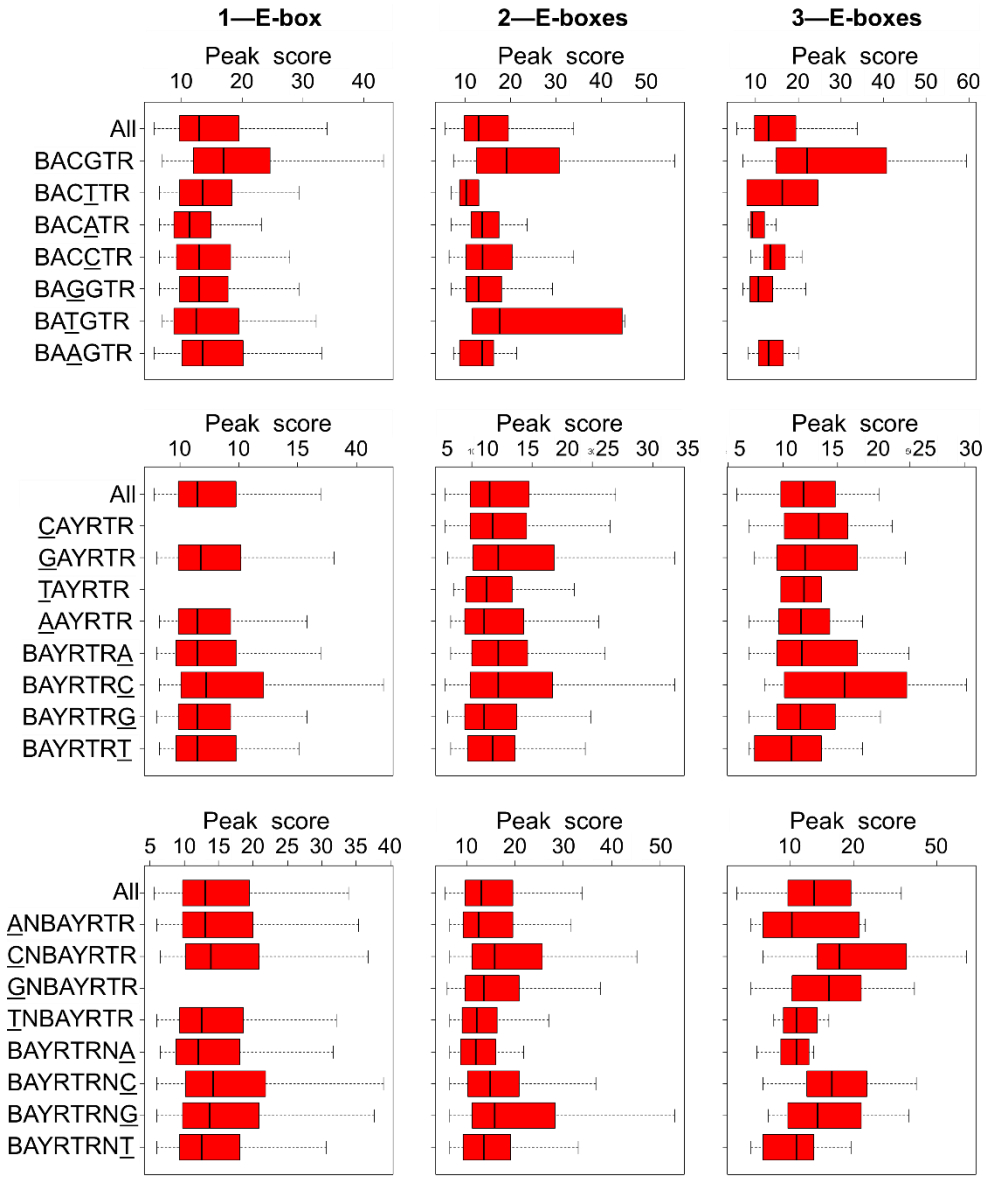
Supplementary Figure 3.1.24 Box-plot of peak score against different E-box motifs (based on the consensus motif identified de novo) for USF1. The first column represents data from peaks with one E-box, the second with two and the third with three E-boxes of the indicated type only. The first row looked at the effect of core di-nucleotides on binding affinity (strand specific). The second row looked at the effect of 5'- and 3'- nucleotides immediately adjacent to the core E-box. The third row looked at the effects of the second nearest 5'- and 3'- flanking nucleotides.



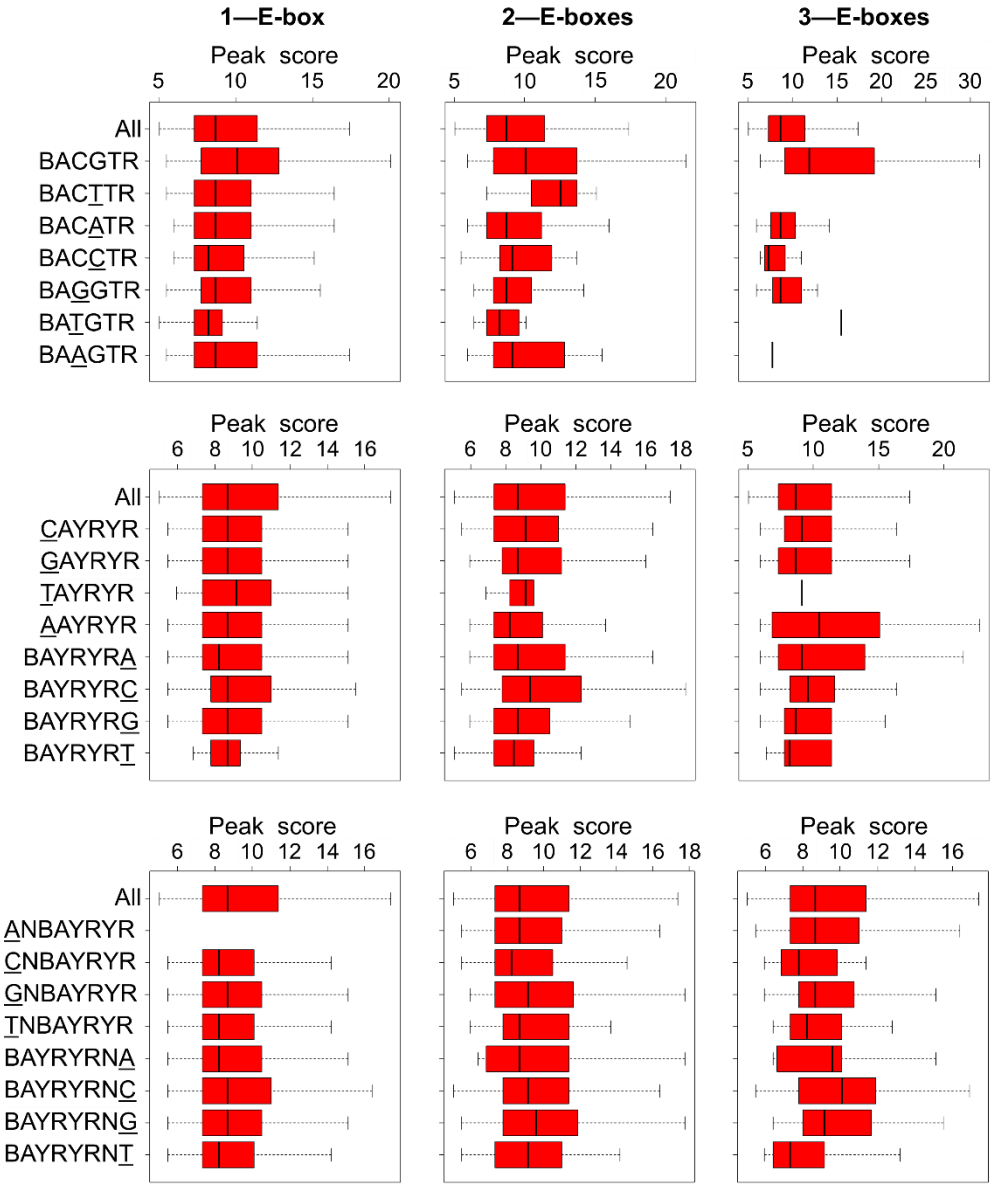
Supplementary Figure 3.1.25 Box-plot of peak score against different E-box motifs (based on the consensus motif identified de novo) for USF2. The first column represents data from peaks with one E-box, the second with two and the third with three E-boxes of the indicated type only. The first row looked at the effect of core di-nucleotides on binding affinity (strand specific). The second row looked at the effect of 5'- and 3'- nucleotides immediately adjacent to the core E-box. The third row looked at the effects of the second nearest 5'- and 3'- flanking nucleotides.



Supplementary Figure 3.1.26 Box-plot of peak score against different E-box motifs (based on the consensus motif identified de novo) for HIF-1α. The first column represents data from peaks with one E-box, the second with two and the third with three E-boxes of the indicated type only. The first row looked at the effect of core di-nucleotides on binding affinity (strand specific). The second row looked at the effect of 5'- and 3'- nucleotides immediately adjacent to the core E-box. The third row looked at the effects of the second nearest 5'- and 3'- flanking nucleotides.

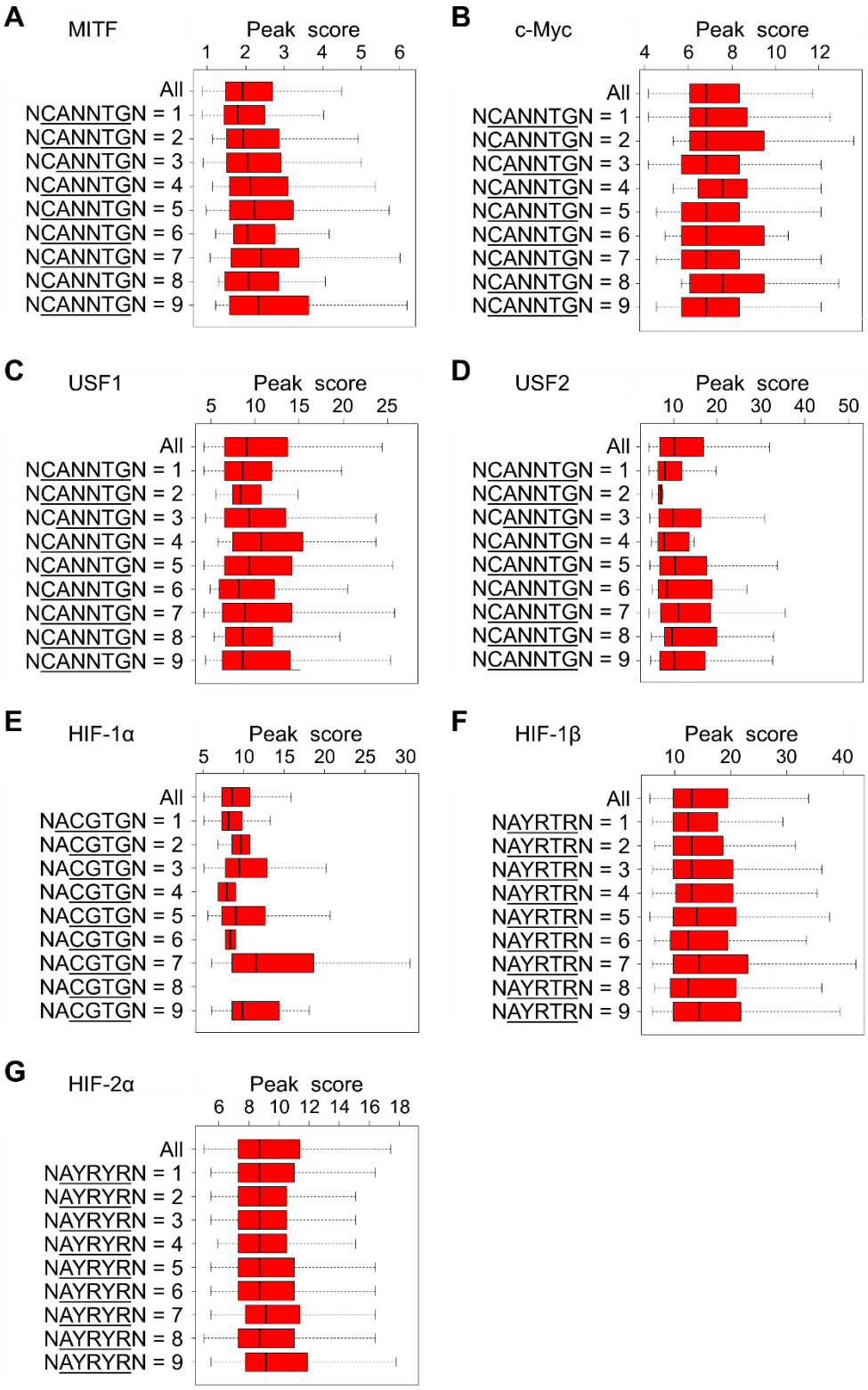


Supplementary Figure 3.1.27 Box-plot of peak score against different E-box motifs (based on the consensus motif identified de novo) for HIF-1β. The first column represents data from peaks with one E-box, the second with two and the third with three E-boxes of the indicated type only. The first row looked at the effect of core di-nucleotides on binding affinity (strand specific). The second row looked at the effect of 5'- and 3'- nucleotides immediately adjacent to the core E-box. The third row looked at the effects of the second nearest 5'- and 3'- flanking nucleotides.

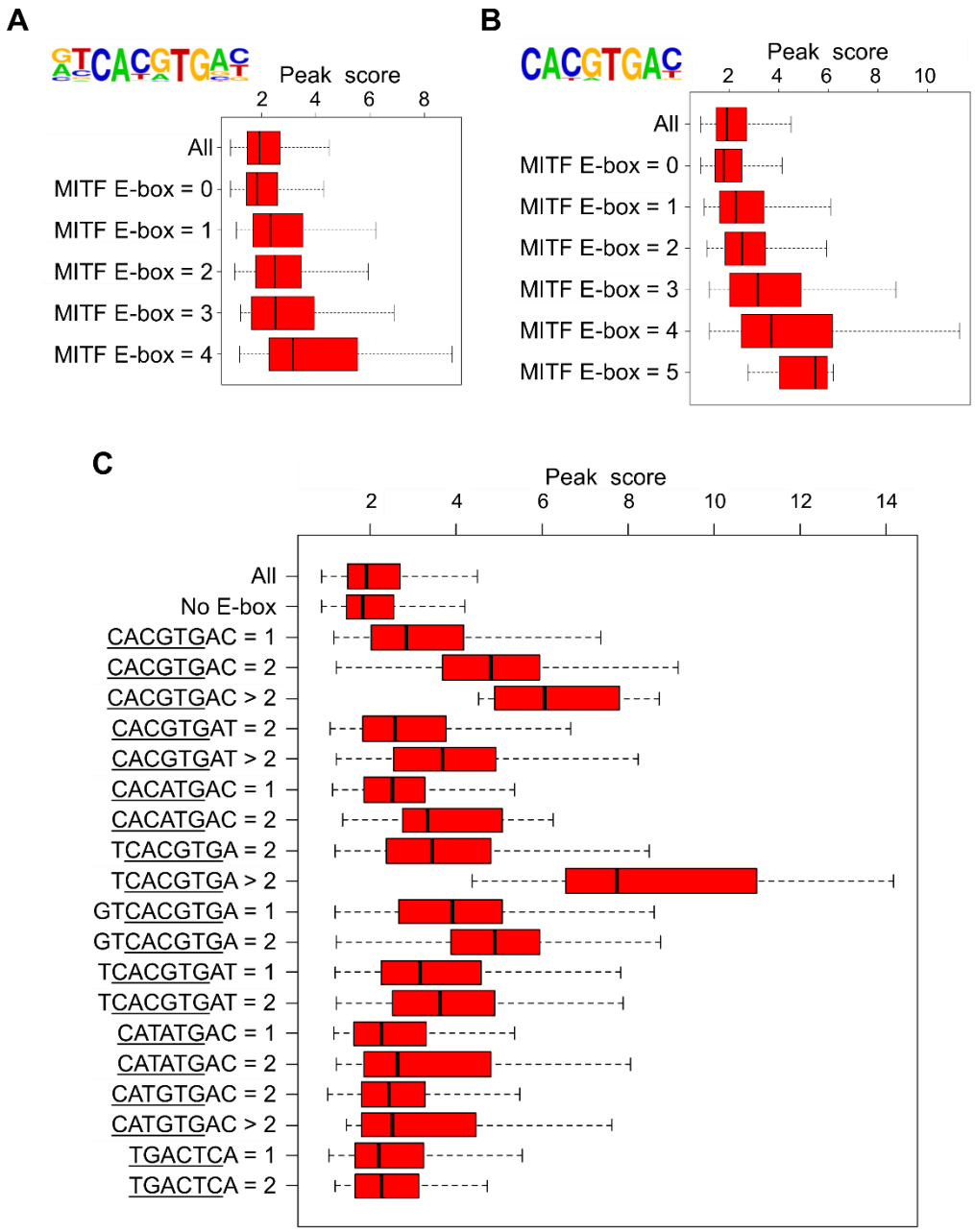


Supplementary Figure 3.1.28 Box-plot of peak score against different E-box motifs (based on the consensus motif identified de novo) for HIF-2α. The first column represents data from peaks with one E-box, the second with two and the third with three E-boxes of the indicated type only. The first row looked at the effect of core di-nucleotides on binding affinity (strand specific). The second row looked at the effect of 5'- and 3'- nucleotides immediately adjacent to the core E-box. The third row looked at the effects of the second nearest 5'- and 3'- flanking nucleotides.

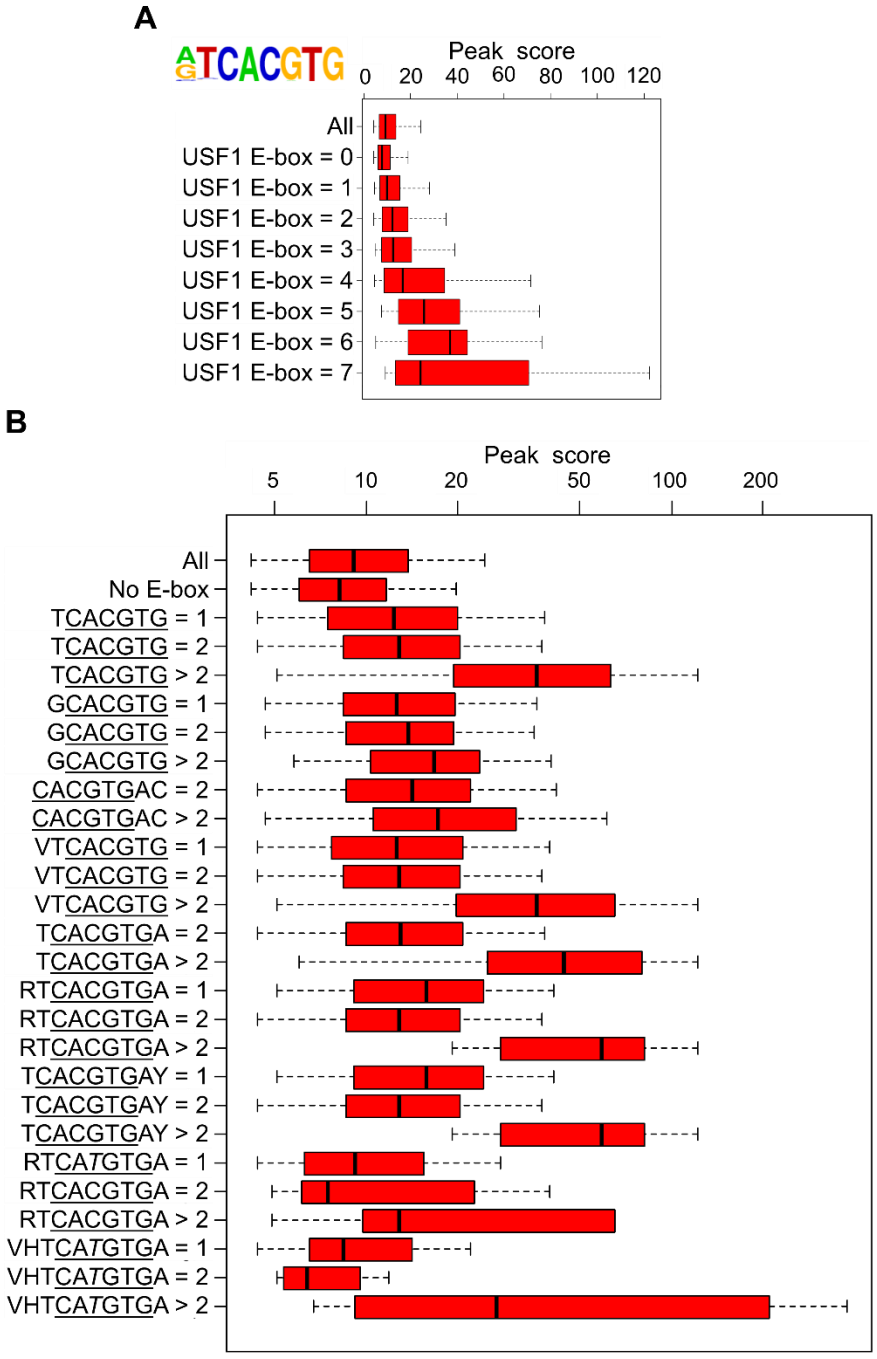
6.1.6 Binding cooperativity



Supplementary Figure 3.1.29 Box-plots of peak score against number of motifs. In order to investigate cooperative binding within a given peak disregarding the type of E-box underneath, ChIP-seq peak files were screened against the indicated degenerated E-box and box-plots of peak score against the number of motifs under a given peak were made for MITF (**A**), MYC (c-Myc) (**B**), USF1 (**C**), USF2 (**D**), HIF-1α (**E**), HIF-1β (**F**) and HIF-2α (**G**). The core E-box/HRE are underlined.



Supplementary Figure 3.1.30 Box-plots of peak score against different number of motifs identified de novo for MITF. **A)** Box-plots of MITF's peak score stratified into the number of motifs derived from MITF's ChIP-seq full-dataset. The motif file used is displayed on the top right hand corner and displayed on the graph as "MITF E-box". "All" denotes distribution of the peak score form all the peaks with any number of motifs. **B)** Same as **A)** but with MITF's motif derived from MITF's top 6% of the peaks. **C)** Same as A) but using the motif files indicated on the left of each box. The number next to the motif indicates the number of E-box occurred within the peak. Motifs with no data and motifs which demonstrated no difference in the peak score between one, two and > 2 E-box are omitted. The core E-boxes are underlined.



Supplementary Figure 3.1.31 Box-plots of peak score against different number of motifs identified de novo for USF1. **A)** Box-plots of USF1's peak score stratified into the number of motifs derived from USF1's full-dataset. The motif file used is displayed on the top right hand corner and displayed on the graph as "USF1 E-box". "All" denotes distribution of the peak score form all the peaks with any number of motifs. Note that the motif derived from peaks with score of ≥ 10 is almost identical to that derived from the full dataset and hence not shown. **B)** Same as A) but using the motif files indicated on the left of each box. The number next to the motif indicates the number of occurrence within the peak. Motifs with no data and motifs which demonstrated no difference in the peak score between one, two and > 2 E-box are omitted. The core E-boxes are underlined. Sequence within the core E-box that deviated from the consensus sequence are italicized.

6.1.7 Co-occupancy

Supplementary Table 3.1.4

Fold enrichment of ChIP-seq coverage for the indicated "Tag files" (mapped reads) in the commonly bound peaks (co-bound by all the bHLH-TFs in "Other peaks" column) relative to the average of ChIP-seq coverage of the same set of bHLH-TFs derived from whole dataset. The enrichment ratio is colour-coded such that most intense red corresponds to lowest enrichment ratio while most intense green corresponds to the most highly enriched. Colour-coding is stratified to peaks co-bound with HIF-1α, HIF-1β, HIF-2α or MITF, MYC and USF.

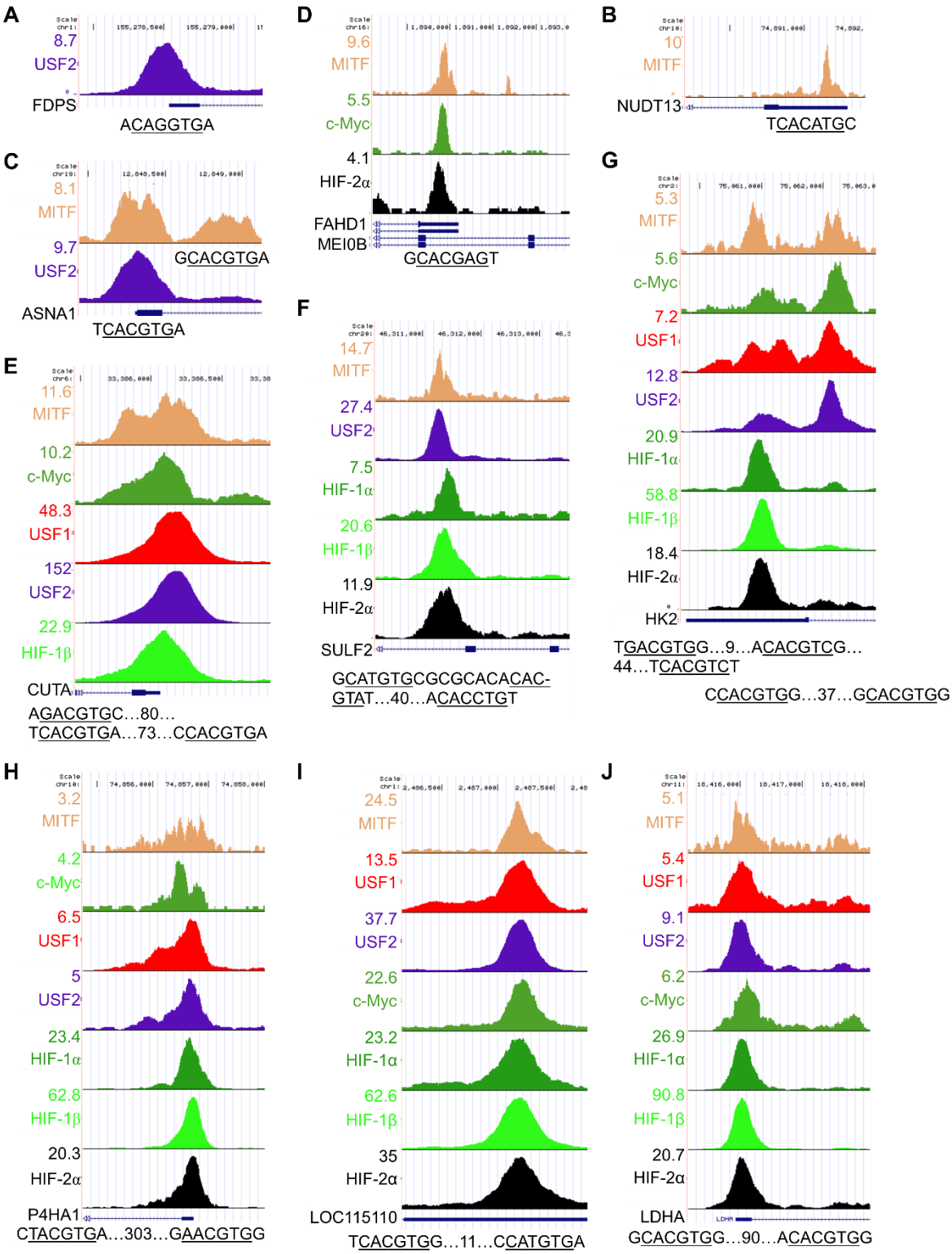
Peak1	Other peaks/ Tag files	Ratio enriched in co-bound peaks
HIF-1α	HIF-1α	1.5
	HIF-1β	3.4
HIF-1α	HIF-1α	2.2
	HIF-2α	2.2
HIF-1α	HIF-1α	4.7
	MITF	3.4
HIF-1α	HIF-1α	2.7
	MYC	3.2
HIF-1α	HIF-1α	2.2
	USF1	2.2
HIF-1α	HIF-1α	2.0
	USF2	3.0
HIF-1α	HIF-1α	2.2
	HIF-1β	4.6
	MYC	2.4
HIF-1α	HIF-1α	1.9
	HIF-1β	4.4
	USF1	1.8
HIF-1α	HIF-1α	1.8
	HIF-1β	4.2
	USF2	2.7
HIF-1α	HIF-1α	1.9
	HIF-1β	4.6
	HIF-2α	1.8
HIF-1α	HIF-1α	2.5
	HIF-1β	5.2
	HIF-2α	2.6
	MYC	2.8
HIF-1α	HIF-1α	2.2
	HIF-1β	5.3
	HIF-2α	2.2
	USF1	2.0
HIF-1α	HIF-1α	2.3
	HIF-1β	5.4
	HIF-2α	2.3
	USF2	2.9
HIF-1β	HIF-1β	2.6
	HIF-2α	1.3
HIF-1β	HIF-1β	2.9
	MITF	2.5
HIF-1β	HIF-1β	3.1
	MYC	1.9
HIF-1β	HIF-1β	2.1
	USF1	1.6
HIF-1β	HIF-1β	1.9
	USF2	2.1
HIF-1β	HIF-1β	4.3
	HIF-2α	2.5
	MITF	2.7
HIF-1β	HIF-1β	3.9
	HIF-2α	2.1
	MYC	2.3
HIF-1β	HIF-1β	3.4
	HIF-2α	1.7
	USF1	1.8
HIF-1β	HIF-1β	3.3
	HIF-2α	1.6
	USF2	2.6
HIF-2α	HIF-2α	3.1
	MITF	3.1
HIF-2α	HIF-2α	2.5
	MYC	2.7
HIF-2α	HIF-2α	1.9
	USF1	2.1
HIF-2α	HIF-2α	1.7
	USF2	2.9
MITF	MITF	3.7
	MYC	4.3
MITF	MITF	3.0
	USF1	2.9
MITF	MITF	2.7
	USF2	3.3
MITF	MITF	3.1
	USF1	3.0
	USF2	4.7
MITF	MITF	3.8
	USF1	4.7
	USF2	7.6
	MYC	4.3
MYC	MYC	2.5
	USF1	2.9
MYC	c-Myc	2.2
	USF2	4.0
MYC	MYC	2.5
	USF1	2.9
	USF2	4.6
USF1	USF1	1.8
	USF2	2.6

Motif	bHLH-TF	Best matches	P-value
	MITF	USF1 (GM12878 Homer)	$1e^{-7694}$
	c-Myc	USF1 (GM12878 Homer)	$1e^{-137}$
	USF1	USF1 (GM12878 Homer)	$1e^{-2452}$
	USF2	USF2 (C2C12 Homer)	$1e^{-999}$
	HIF-1α	HIF-2α (785-RCC Homer)	$1e^{-97}$
	HIF-1β	USF1 (GM12878 Homer)	$1e^{-245}$
	HIF-2α	USF1 (GM12878 Homer)	$1e^{-103}$

Supplementary Figure 3.1.33 Summary list of motifs derived from the sequence under the peaks of ChIP-seq full-dataset. The motifs are centred around the core E-box 5'-CACGTG-3'

Motif	bHLH-TF	Best matches	P-value
	MITF	USF2 (C2C12 Homer)	$1e^{-1855}$
	c-Myc	USF1 (GM12878 Homer)	$1e^{-83}$
	USF1	USF1 (GM12878 Homer)	$1e^{-926}$
	USF2	USF1 (GM12878 Homer)	$1e^{-2277}$
	HIF-1α	HIF-1α (MCF7 Homer)	$1e^{-61}$
	HIF-1β	CLOCK (Liver Homer)	$1e^{-173}$
	HIF-2α	CLOCK (Liver Homer)	$1e^{-60}$

Supplementary Figure 3.1.34 Summary list of motifs derived from the sequence under the peaks with peak score of ≥ 10 for MYC (c-Myc), USF1, USF2, HIF1-α, HIF-1β and HIF-2α. For MITF, the top 6% of peaks rank by peaks score was used. The motifs are centre around the core E-box 5'-CACGTG-3'



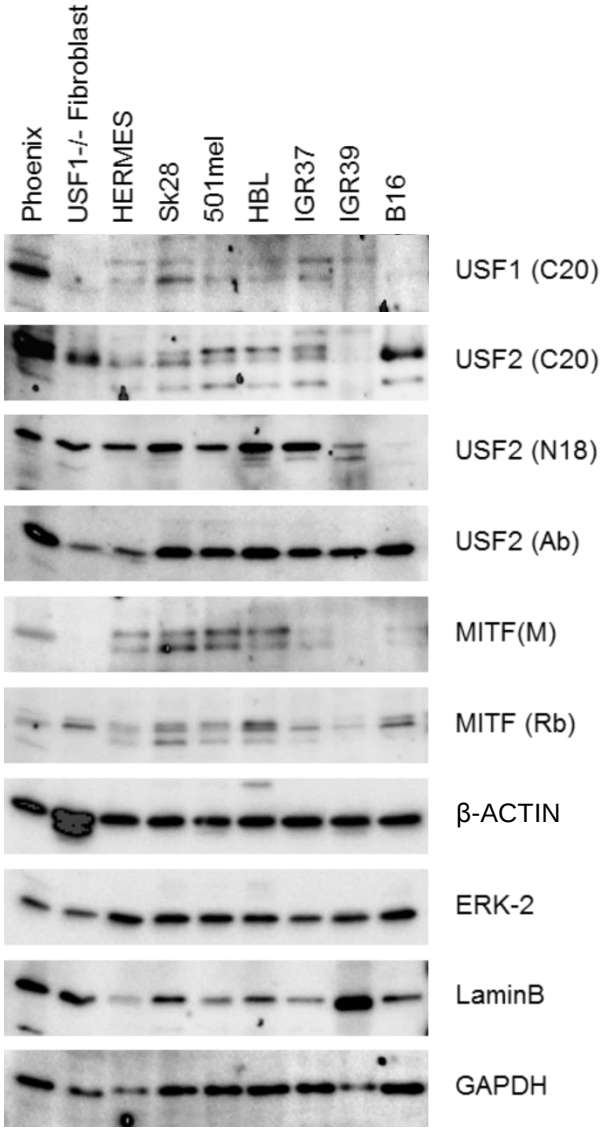
Supplementary Figure 3.1.35 Snapshot of bedgraphs uploaded on to the UCSC genome browser at various genes. In each case, only bHLH-TFs that show binding to a given gene are shown. E-box sequences are given beneath the peaks where the core E-boxes are underline.

6.3 Factor exchange and promoter education

Factor exchange at the human *Tyrosinase* gene (*TYR*)

Induction of *TYR* gene by UVB

As USF and MITF are required for maximal expression from the *TYR* gene, it is possible that the ratio of these factors in different cell lines may determine, partly, the basal activation potential of the *TYR* gene. Western-blot analysis of various cell lines was carried out to assess the level of USF and MITF in various cell lines. A number of USF2 antibodies were



Supplementary Figure 3.3.1 Western-blot of whole cell extracts from Phoenix (human embryonic kidney cell), Fibroblasts with heterozygous USF1 knock-out, HERMES (human melanocyte cell lines) and human melanoma cell lines (Skmel28, 501mel, HBL, IGR37 and IGR39) and B16 (mouse melanoma cell line) probed for USF1, USF2 and MITF. β-ACTIN, ERK-2, LaminB and GAPDH were used as loading controls.

used, as discussed in section 3.4 where we validated the specificity of these antibodies and believe that the observed differences in the specific bands detected on Western-blot reflects different antibodies picking up different species/modified forms of USF2. The two MITF antibodies used (denoted as M and Rb) are homemade antibodies targeting different region of MITF and were used to detect differently modified forms of MITF. A number of loading controls were used, but it is worth pointing out that, the amount of sample loaded for all the gels used to make this figure were the same. We performed initial normalisation to β-ACTIN, but included here all the loading controls used in this thesis. The poor correlation between different loading-control indicated that while these proteins are useful as loading controls for experiments within a given cell line, they may not be appropriate controls when comparing different cell lines. In any case, we are more interested in the ratio of USF1, USF2 and MITF in each cell line rather than the absolute difference between cell lines. To address the latter question we would need to count cells and load equal number of cells into each lane.

Shown here is the Western-blot of whole cell extract from two non-pigmented cell lines (Phoenix and Fibroblasts), one human melanocyte cell line (HERMES), five melanoma cell lines (SkMel28, 501mel, HBL, IGR37 and IGR39) and one mouse melanoma cell line (B16). Phoenix was used as a non-pigmented cell line control. Reflective of MITF-M being specific to the melanocyte lineage (Yavuzer and Goding, 1994), the MITF band detected in Phoenix cells has a different migration pattern from melanoma and melanocytes, possibly reflecting different isoforms of MITF expressed in Phoenix cell line or cross-reactivity of our antibodies with other MiT-family member such as TFEB. The ubiquitous bHLH-TF, USF1 and USF2 are expressed at a high level. Interestingly, the Phoenix cell line has a relatively high level of the faster migrating form of USF1 than other cell lines. Based on the β-ACTIN loading control, Phoenix cells express a much higher level of USF1 and USF2 (as detected by C20-antibody) than other cell lines. The level of USF2 as detected by N18- and Ab-antibodies are comparable between the SkMel28, 501mel, HBL and IGR37 melanoma cell lines.

USF1$^{-/-}$ fibroblasts was used as another non-pigmented cell line control and also to affirm the specificity of the USF1 antibody which did not pick up the two bands detected in other cell lines. Relative to Phoenix cells, the USF2 species detected by the C20- and Ab-antibodies is much lower than the N18-specific specie. Unlike Phoenix cells, no MITF band as detected by the MITF(M)-antibody could be seen, confirming the notion that the bands detected in Phoenix cells correspond to different isoforms of MITF.

HERMES is a human melanocyte cell line that was used as a control for non-transformed pigment cells. In addition to expressing USF1 and USF2, HERMES also express abundant MITF, reflecting its identity as a melanocyte. While expressing similar levels of USF1 to other pigmented cell lines, HERMES expresses fewer USF2 species detected by the Ab-antibody.

SkMel28 (Sk28) is a melanoma cell line, chosen as it expresses a different ratio of modified MITF compared to 501mel. This cell line was suspected to carry elevate levels of a sumoylated form of MITF (Miller et al., 2005), but lower overall levels of MITF than 501mel. However, in our blot, SkMel28 cells express higher levels of MITF as detected by the Rb-antibody and have a higher ratio of hypo-phosphorylated form of MITF (lower band of the two as detected by MITF(M)-antibody). SkMel28 is the only pigmented cell line that expresses a higher ratio of faster migrating form of USF1, similar to Phoenix cells. SkMel28

is also the only melanoma cell line that expresses higher level of the slower migrating form of the USF2 as detected by the C20-antibody (the middle band of the three on the blot). The level of USF2 as detected by N18-antibody is similar to that of HBL and IGR37 but higher than all the other cell lines.

501mel was chosen because many of the previous studies, including that presenting the MITF-ChIP-seq, were performed in this cell line. Moreover, it has been reported to express abundant MITF and USF. However, in our blot it appears to express lower levels of USF1 than the Phoenix, SkMel28 and IGR37 cell lines and although they have similar levels of USF2 as detected by the Ab-antibody, 501mel cells have a lower level of USF2 as detected by the N18-antibody compared to other melanoma cell lines except IGR39. 501mel cells have predominantly the slow migrating form of USF2 as detected by the C20-antibody. 501mel expresses a similar ratio of the hypo- and hyper-phosphorylated forms of MITF.

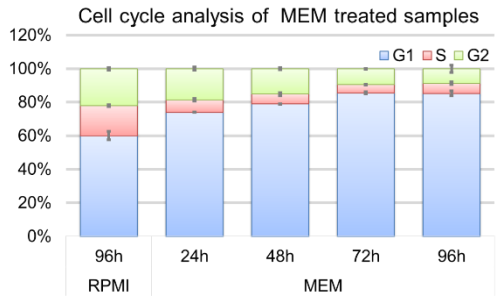
HBL is a human melanoma cell line chosen as it express a wild-type form of MC1R and would be an ideal cell line for studying induction by α-MSH and its analogues. This cell line also expresses abundant MITF which appears mainly as the hyper-phosphorylated form (the upper band in MITF(M) and the upper two bands on MITF(Rb). HBL expresses comparable level of USF1 and USF2 as detected by the C20- and Ab-antibodies compared to 501mel cells but a higher level of USF2 as detected by the N18-antibody than 501mel cells.

IGR37 was derived from a primary melanoma and expresses comparable levels of USF2 to the SkMel28 cell line but higher levels of both USF1 bands than other melanoma cell lines except SkMel28. The MITF level is relatively low compared to other melanoma cell lines with an equal distribution between the hypo- and hyper-phosphorylated form of MITF as detected by the anti-MITF(M) antibody. The lower level of MITF is reproduced with MITF(Rb)-antibody, although in this case, it's the middle band which is enriched.

IGR39 was derived from a metastatic melanoma from the same patient from which IGR37 was established. IGR39 has a depleted level of MITF when compare to other melanoma cell lines, reflective of its highly dedifferentiated phenotype. Unlike other cell lines, however, IGR39 expresses a very low level of USF1 and USF2 as detected by the C20- and N18-antibodies although it does express comparable levels of USF2, as detected by the Ab-antibody, to IGR37. IGR39 is the only human cell line which expresses comparable levels of the two forms of USF2 as detected by the N18-antibody whereas all the other cell lines have predominantly the upper, presumably more modified form.

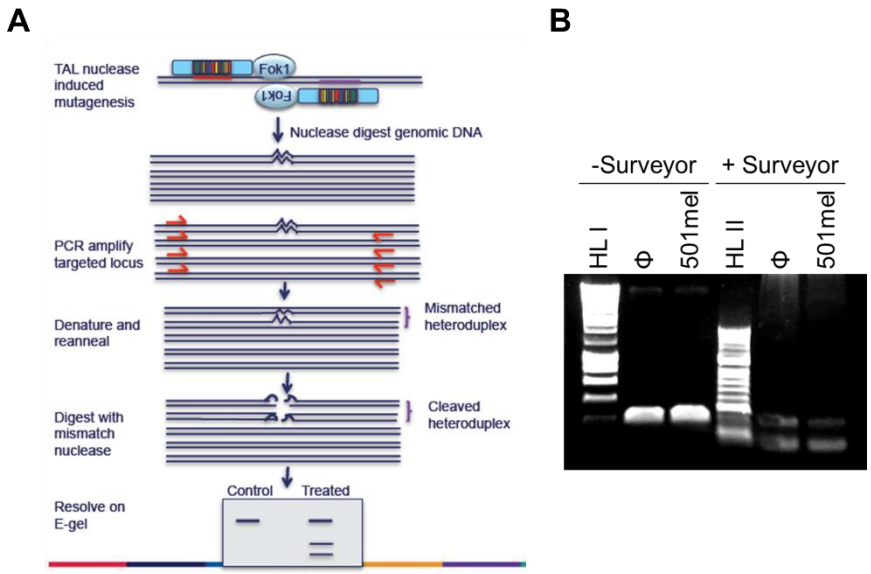
B16 is a mouse melanoma cell line chosen as a representative of non-human melanoma cell line. B16 expresses a comparable level of USF2 as detected by the Ab-antibody and an elevated level of USF2 as detected by the C20-antibody, but surprisingly has a very low level of USF2 as detected by the N18-antibody. There is very little USF1 compared to other human cell lines and the level of MITF is only comparable to that of the IGR37, that is, much lower than all of the human melanoma cell lines. The predominant MITF form as detected by MITF(Rb) antibody is the highly modified form (top two bands).

Other means to induce factor exchange



Supplementary Figure 3.3.2 Cell cycle analysis of 501mel cells grown in MEM. Samples were collected at the indicated time points and nuclear stained with DAPI. RPMI was used as a non-treated control.

Approaches to interfere with bHLH-TFs

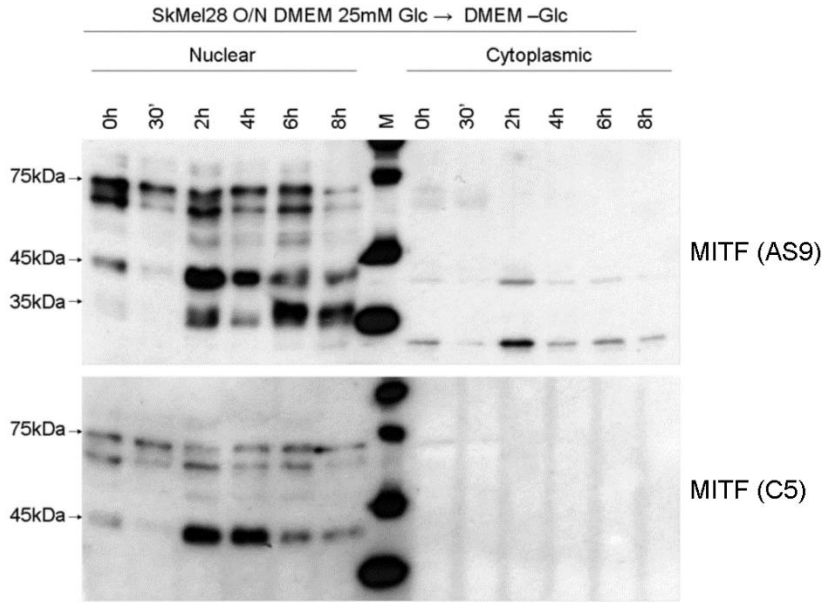


Supplementary Figure 3.3.3 Surveyor assay for validating genome editing by TALEN and Cas9-CRISPR. **A)** Schematic diagram explaining the underlying concept of the surveyor assay originally developed to detect SNPs. Briefly, target-specific binding of TALEN or Cas9-CRISP leads to DNA excision and double strand breaks which, in S-phase of the cell cycle triggers non-homologous end joining (NHEJ) that invariably introduces or excises some nucleotides at the break point. Within a population of cells transfected with TALEN/CAS9-CRISPR, some cells' genomes would be edited while other would not. As each excision is unlikely to result in an identical NHEJ product, this results in many allelic variations at the site of excision. During PCR-amplification of genomic DNA derived from the population of alleles, a large number of mismatched heteroduplexes will form following the renaturation step. These imperfect amplicons can be detected and cleaved by the "Surveyor" enzyme resulting in cleaved PCR products. Diagram taken from Invitrogen. **B)** An example of DNA-agarose gel electrophoresis of PCR product spanning 200 bp region around loci which TALEN cut. The genomic DNA is derived from poly-clonal population of Phoenix (Φ)- or 501mel-cells selected for TALEN directed against USF2. The undigested PCR product (-Surveyor) is shown on the left while the two bands on the right-half of the gel (+ Surveyor) demonstrate that the TALEN had worked and the Surveyor enzyme cleaves the PCR amplicon in two.

6.4 Regulation of USF2 subcellular localisation

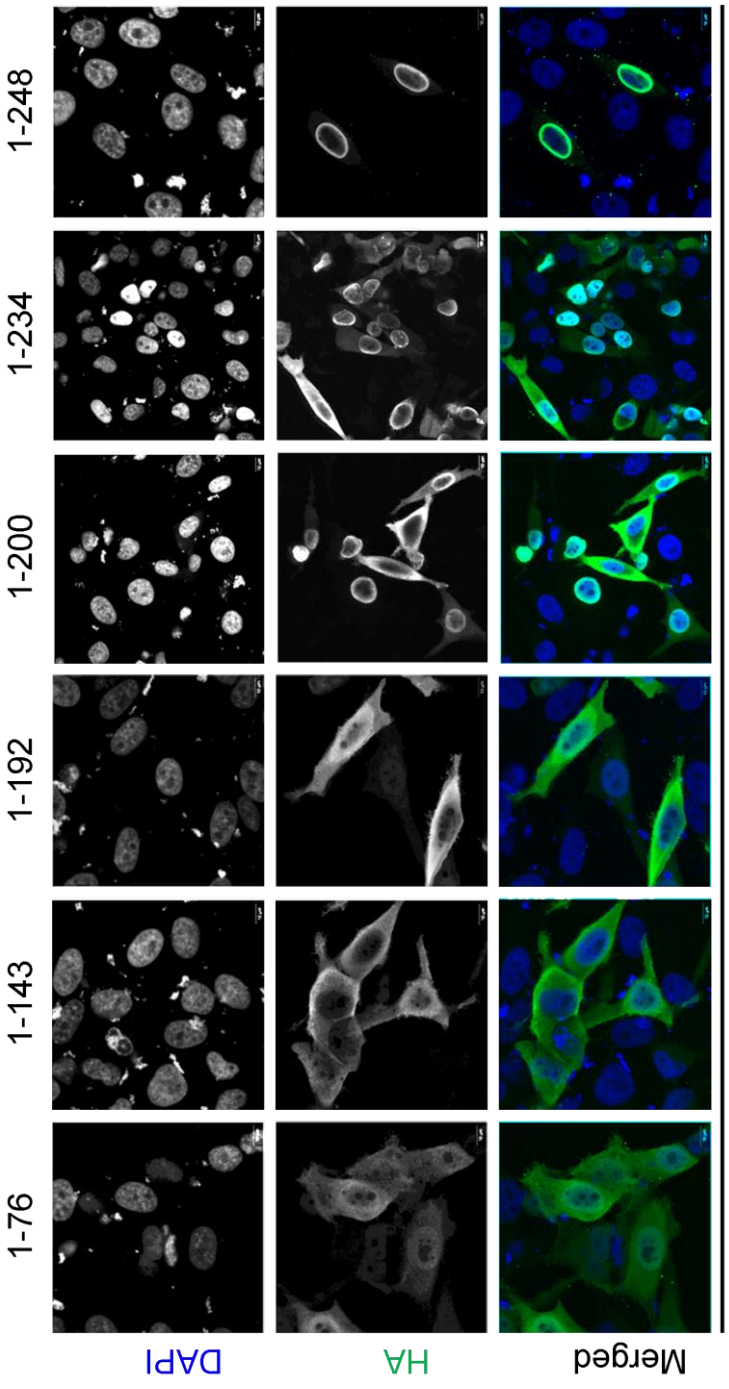
Induction of USF2 nuclear accumulation

Glucose stimulation of USF2

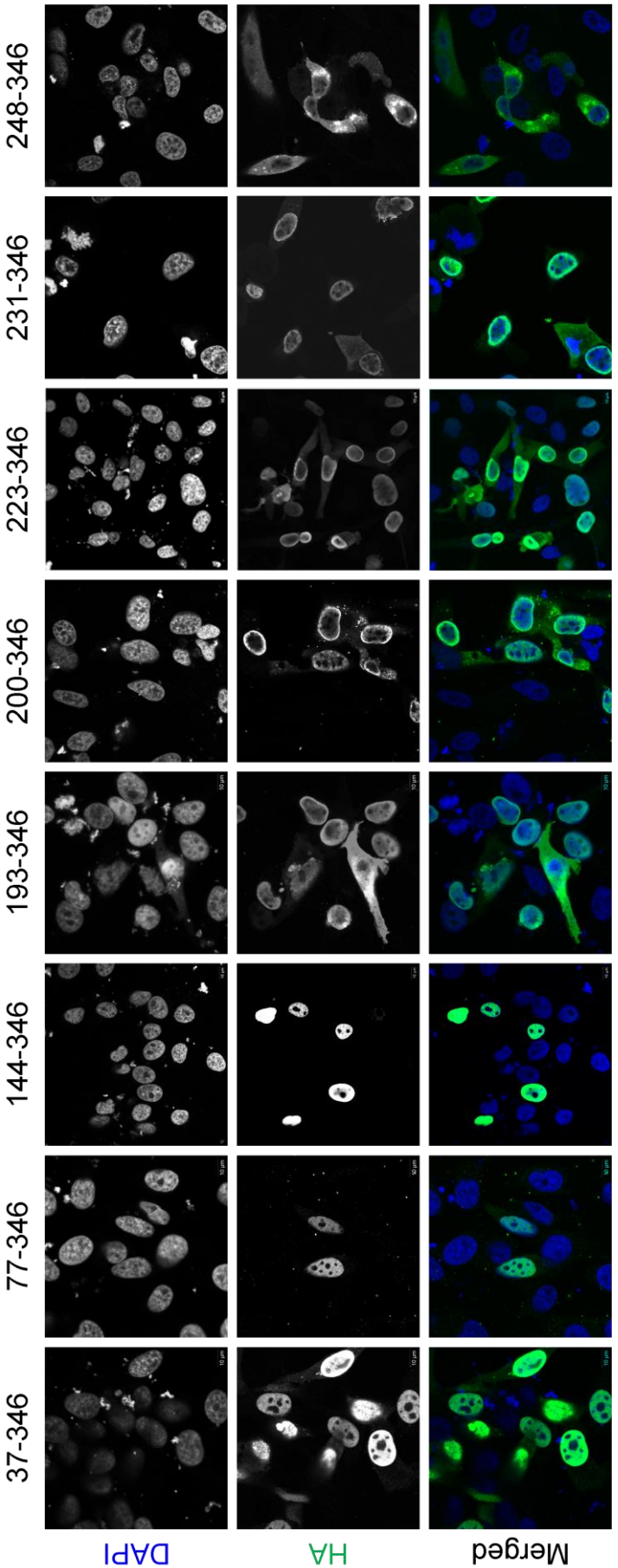


Supplementary Figure 3.4.1 Western-blot of SkMel28 subjected to glucose starvation following overnight DMEM (25 mM glucose), probed with different anti-MITF antibodies. Anti-MITF (AS9) is a monoclonal antibody targeting the C-terminal half of MITF. Anti-MITF (C5) is a commercial monoclonal antibody targeting the N-terminal half of MITF. Both antibodies show a similar trend for changes in the ratio of the two ≈ 75 kDa bands over time course of glucose deprivation. Both antibodies also reveal similar changes in the level of the 45 kDa form of MITF. However, only anti-MITF (AS9) detects the 35 kDa form of MITF, similar to what is seen in Supplementary Figure 3.4.1 using another monoclonal anti-MITF antibody, also targeting the C-terminal half of MITF.

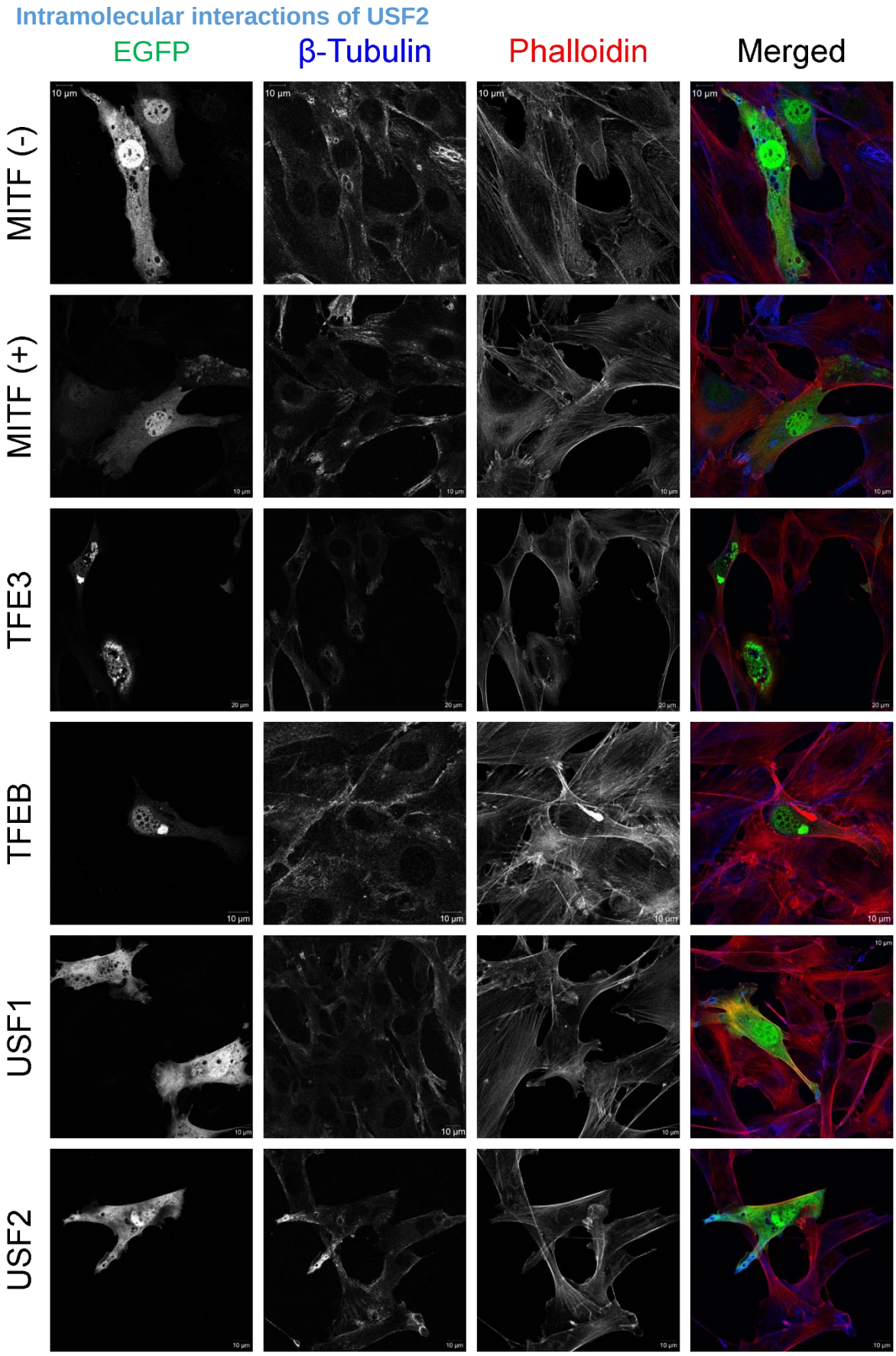
Defining the regions in USF2 required for sub-cellular localisation



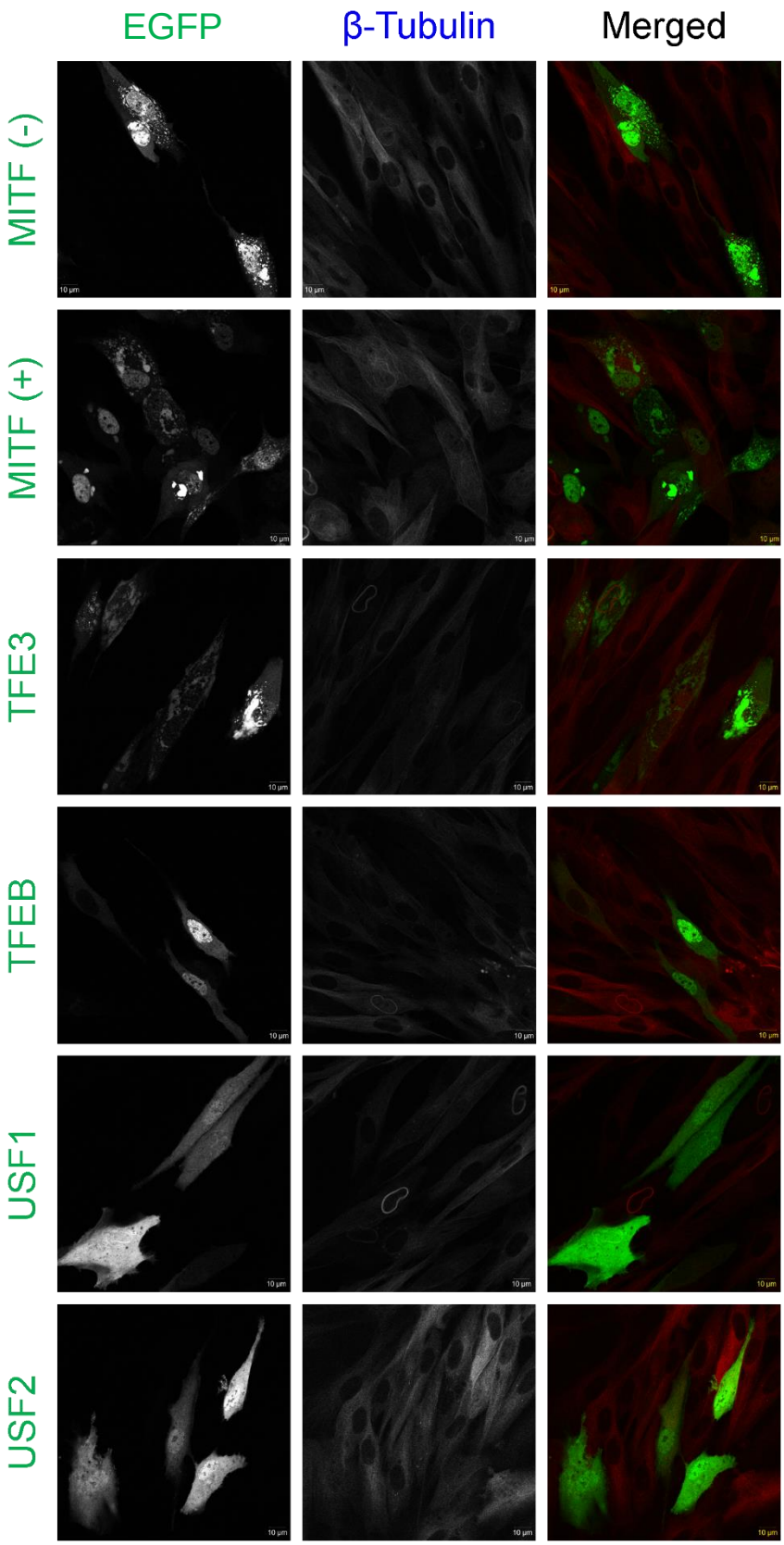
Supplementary Figure 3.4.2 Immunofluorescence of USF2 C-terminal deletion mutants. 501mel cells were transiently transfected with 3xHA-USF2 construct carrying truncation mutants from the C-terminus, leaving only the indicated amino-acids residues (above each column) before being fixed and stained with HA-antibody. Scale bars are shown at the corner of the images.



Supplementary Figure 3.4.3 Immunofluorescence of USF2 N-terminal deletion mutants. 501mel cells were transiently transfected with 3xHA-USF2 construct carrying truncation mutants from the N-terminus, leaving only the indicated amino-acids residues (above each column) before fixed and stained with HA-antibody. Scale bars are shown at the corner of the images.

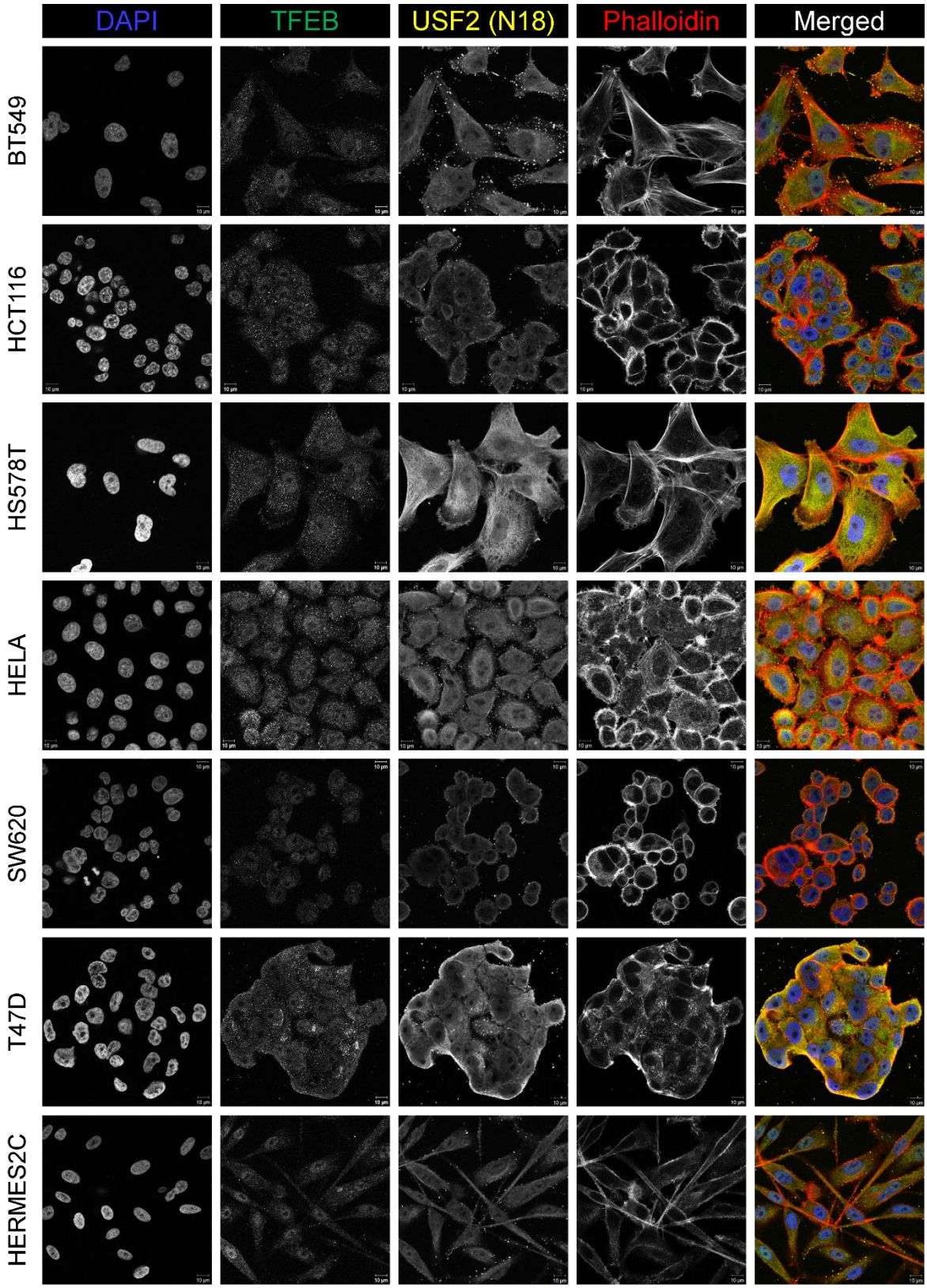


Supplementary Figure 3.4.4 Immunofluorescence of Fibroblast cells transfected with EGFP-tagged bHLH-LZ-TFs and stained with β-tubulin and Phallodin. Scale bars are shown at the corner of the images.

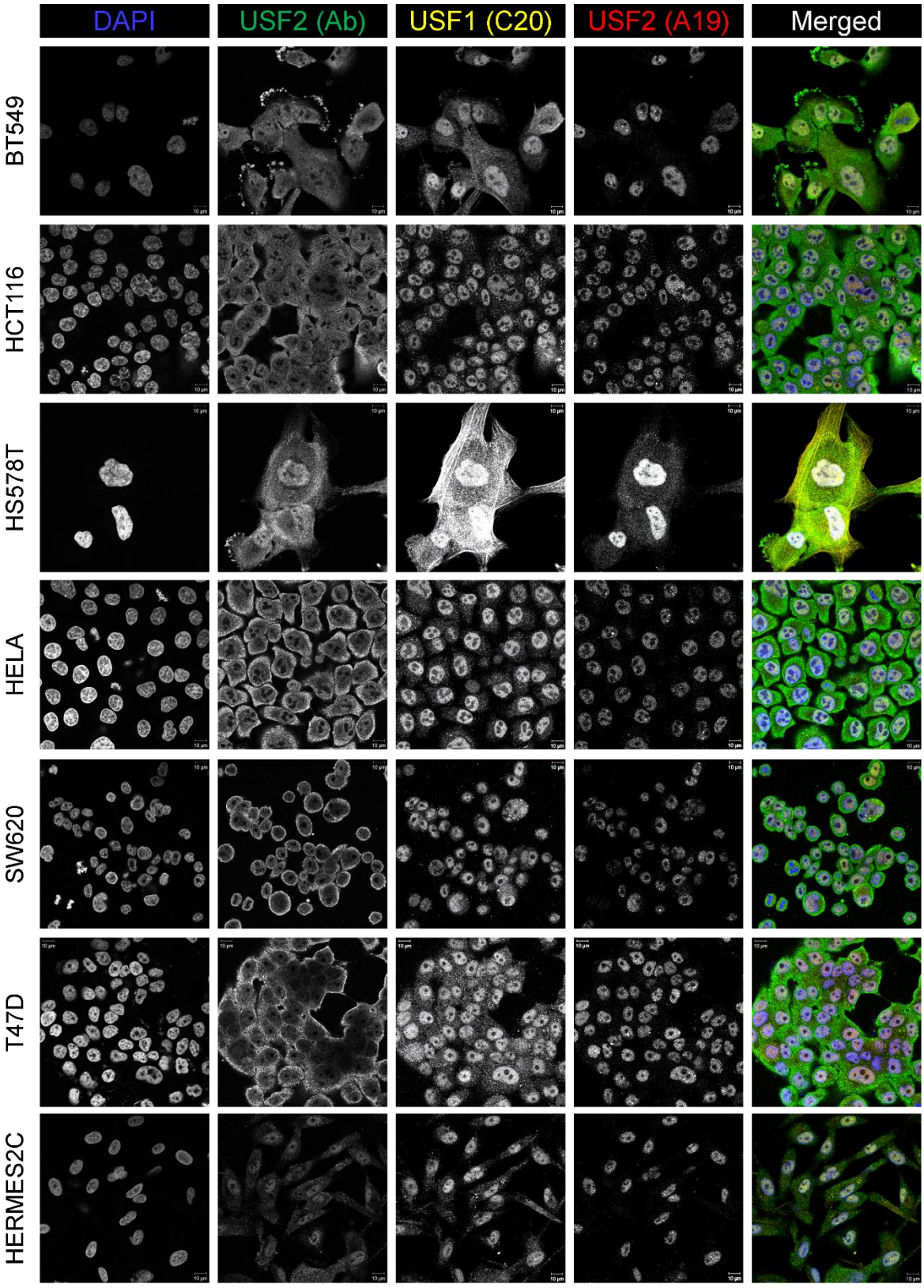


Supplementary Figure 3.4.5 Immunofluorescence of retinal pigmented epithelia cells (RPE) stably expressing mCherry β-tubulin transfected with EGFP-tagged bHLH-LZ-TFs. Scale bars are shown at the corner of the images.

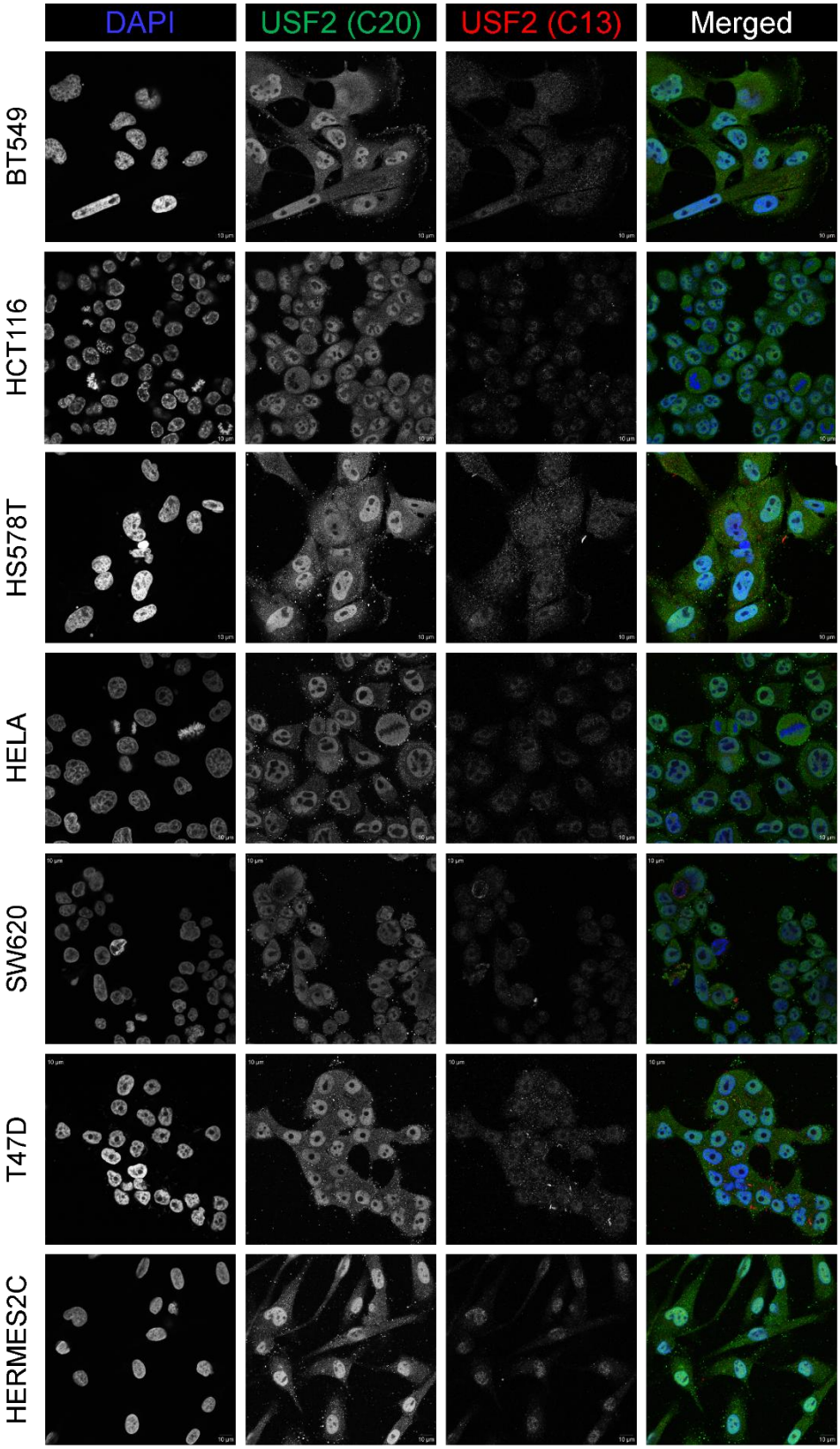
USF2 sub-cellular localisation



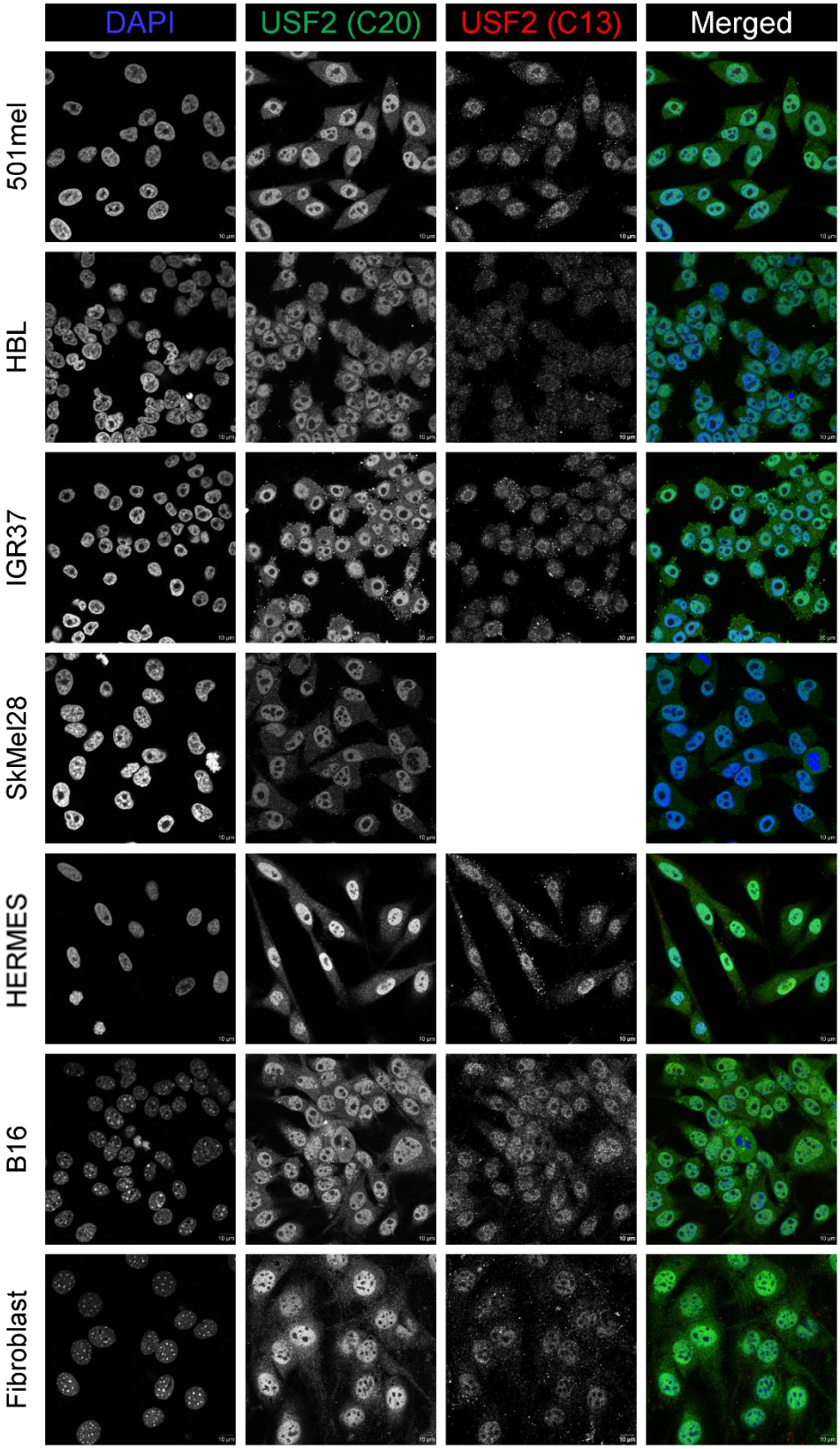
Supplementary Figure 3.4.6 Immunofluorescence of various cell lines stained with TFEB and USF2 (N18) antibodies. Scale bars are shown at the corner of the images.



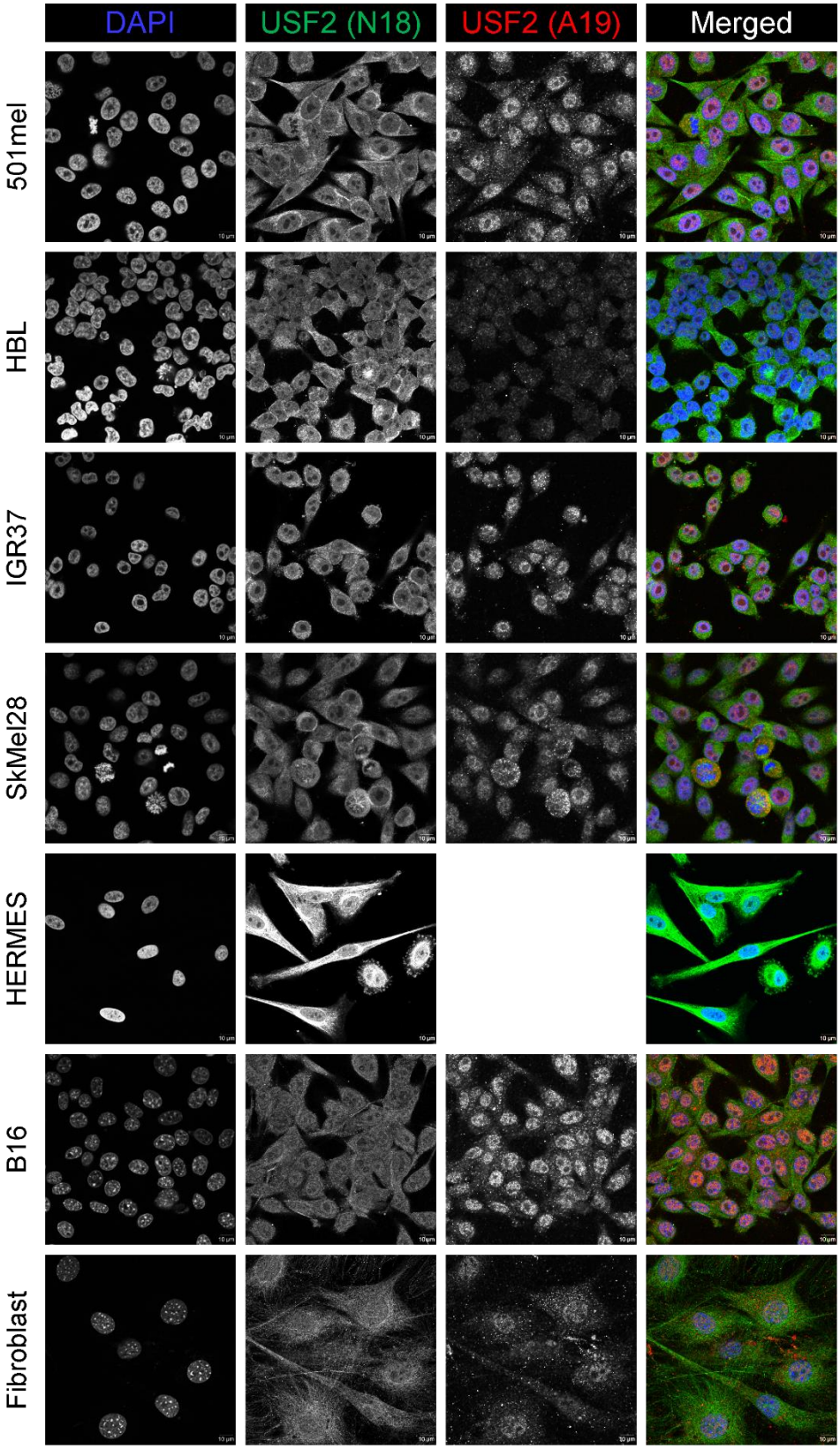
Supplementary Figure 3.4.7 Immunofluorescence of various cell lines stained with USF2 (Ab), USF1 (C20) and USF2 (A19) antibodies. Scale bars are shown at the corner of the images.



Supplementary Figure 3.4.8 Immunofluorescence of various cell lines stained with USF2 (C20) and USF2 (C13) antibodies. Scale bars are shown at the corner of the images.

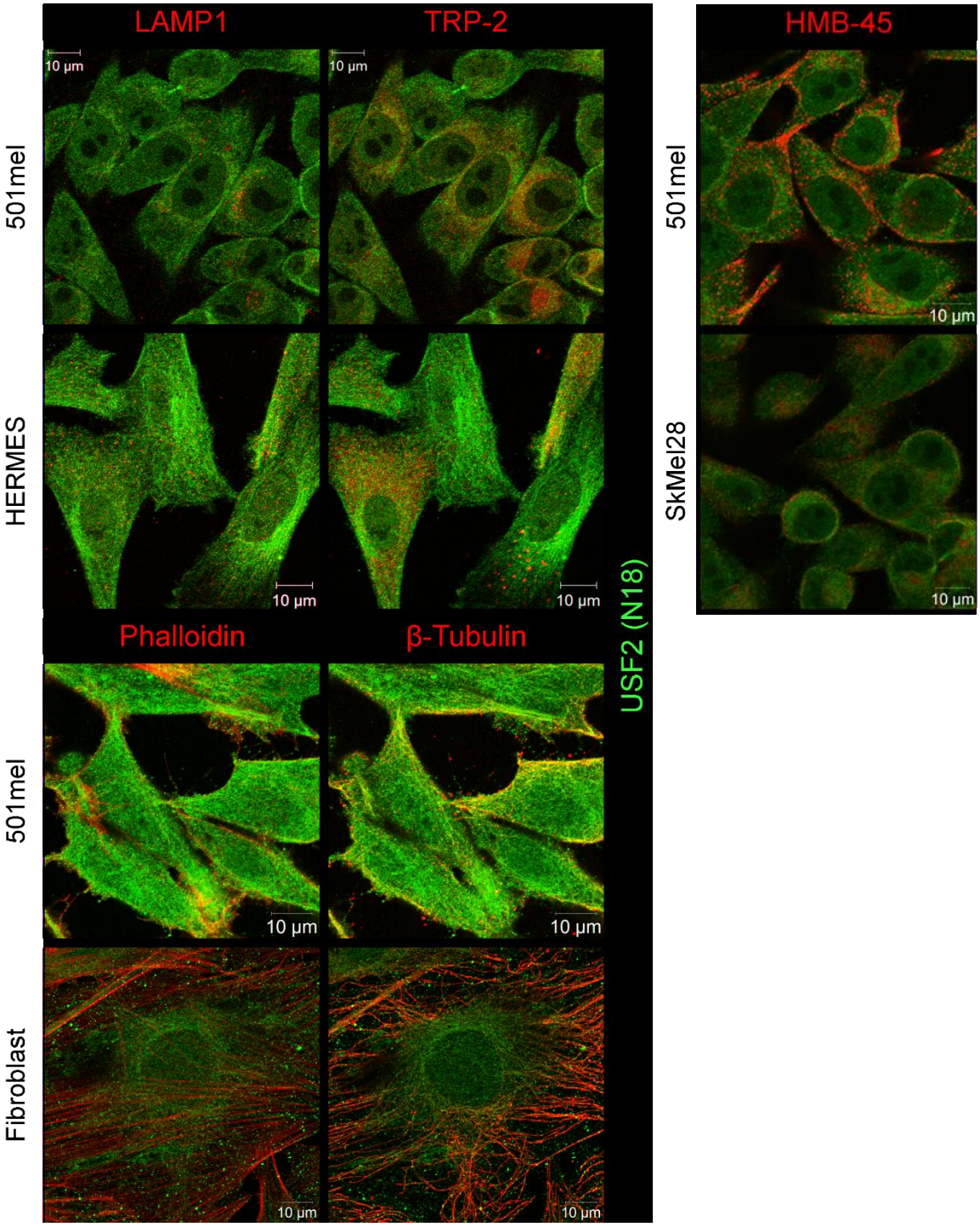


Supplementary Figure 3.4.9 Immunofluorescence of various cell lines stained with USF2 (C20) and USF2 (C13) antibodies. Scale bars are shown at the corner of the images.

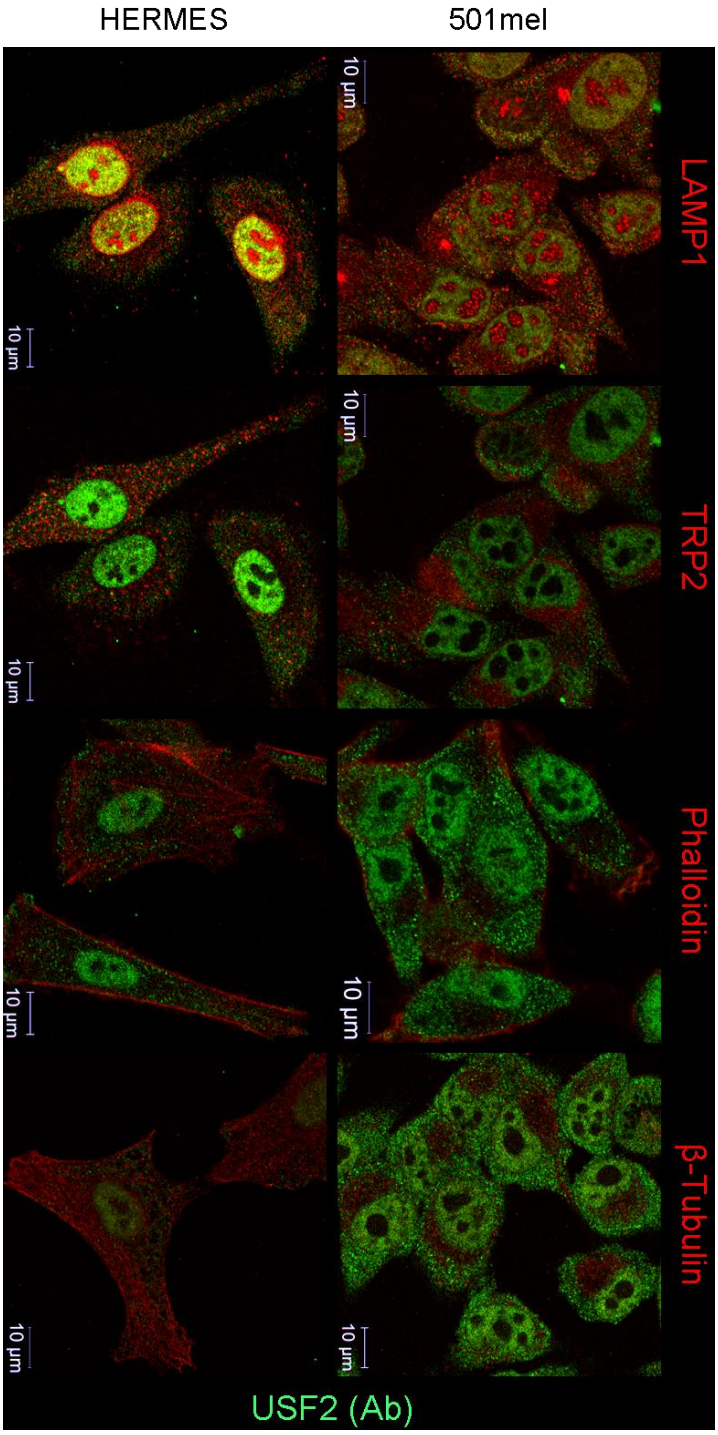


Supplementary Figure 3.4.10 Immunofluorescence of various cell lines stained with USF2 (N18) and USF2 (A19) antibodies. Scale bars are shown at the corner of the images.

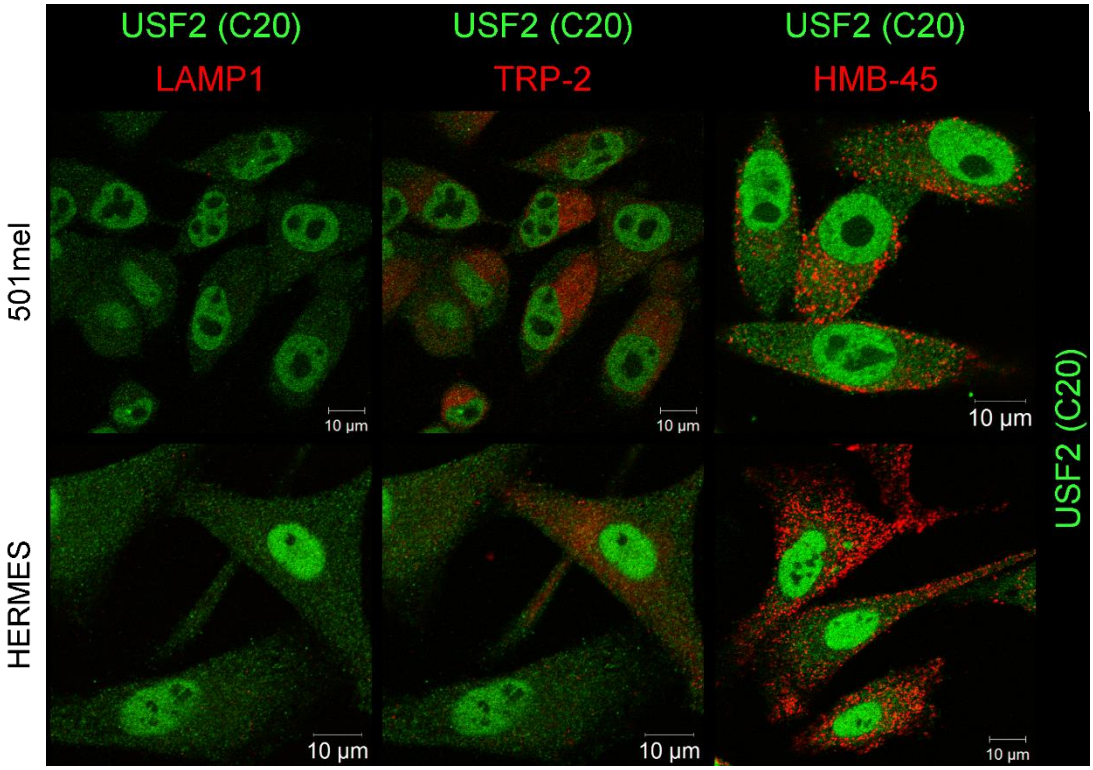
USF2 localisation to organelles/cellular structures.



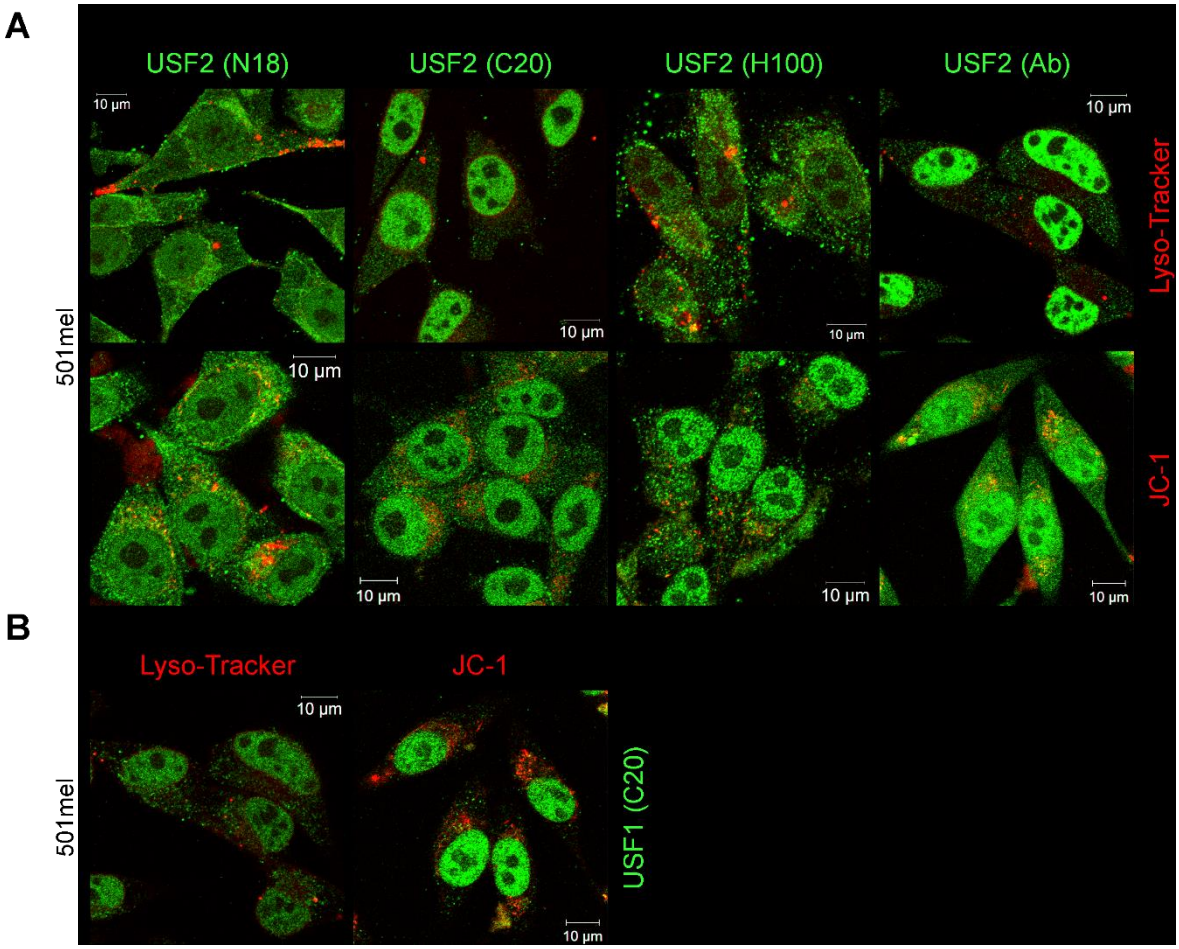
Supplementary Figure 3.4.11 Immunofluorescence staining with USF2 (N18) antibody for colocalisation with different cellular structure/organelles. Cells were co-stained with USF2 (N18) antibody (green) and one other antibody specific for organelles or cellular structure (red). Scale bars are shown at the corner of the images.



Supplementary Figure 3.4.12 Immunofluorescence staining with USF2 (Ab) antibody for colocalisation with different cellular structure/organelles. Cells were co-stained with USF2 (Ab) antibody (green) and one other antibody specific for organelles or cellular structure (red). Scale bars are shown at the corner of the images.

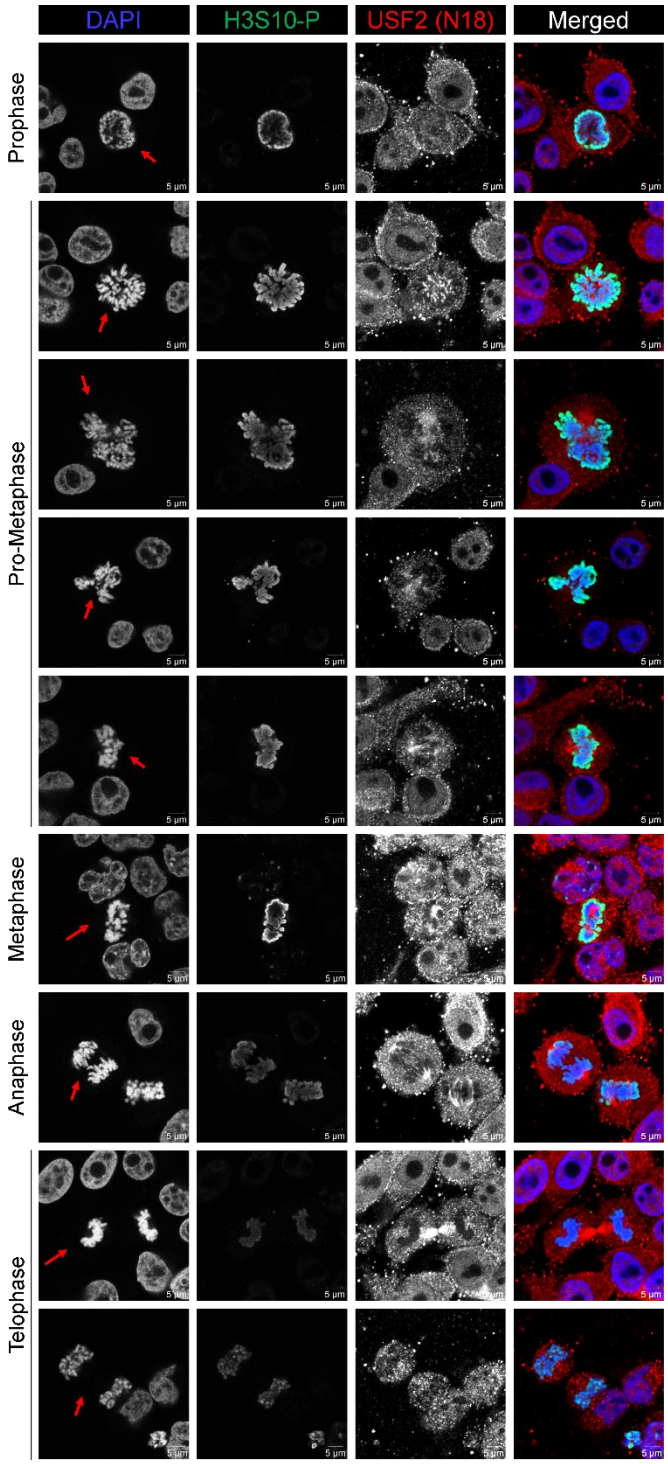


Supplementary Figure 3.4.13 Immunofluorescence staining with USF2 (C20) antibody for co-localisation with different cellular structure/organelles. Cells were co-stained with USF2 (C20) antibody (green) and one other antibody specific for organelles or cellular structure (red). Scale bars are shown at the corner of the images.

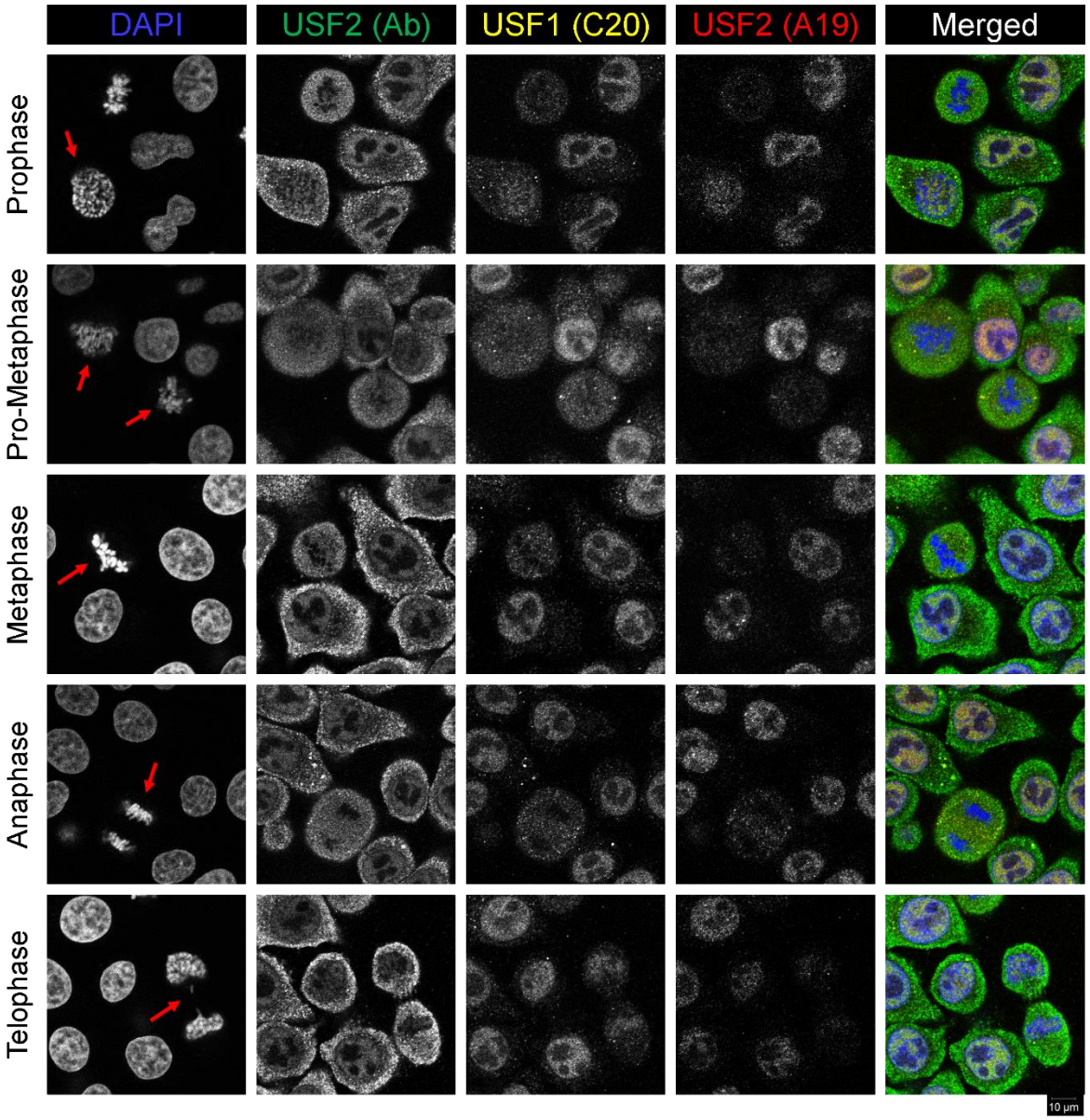


Supplementary Figure 3.4.14 Immunofluorescence staining with USF antibodies for colocalisation with mitochondria or lysosome compartments. 501mel cells were incubated with dye specific for the lysosome compartment (Lyso-Tracker, Invitrogen) or mitochondria (JC-1, Invitrogen) as per supplier recommendation before subjected to fixation and staining as described in Materials and Methods with either USF2 antibodies (A) or USF1 antibody (B). Scale bars are shown at the corner of the images.

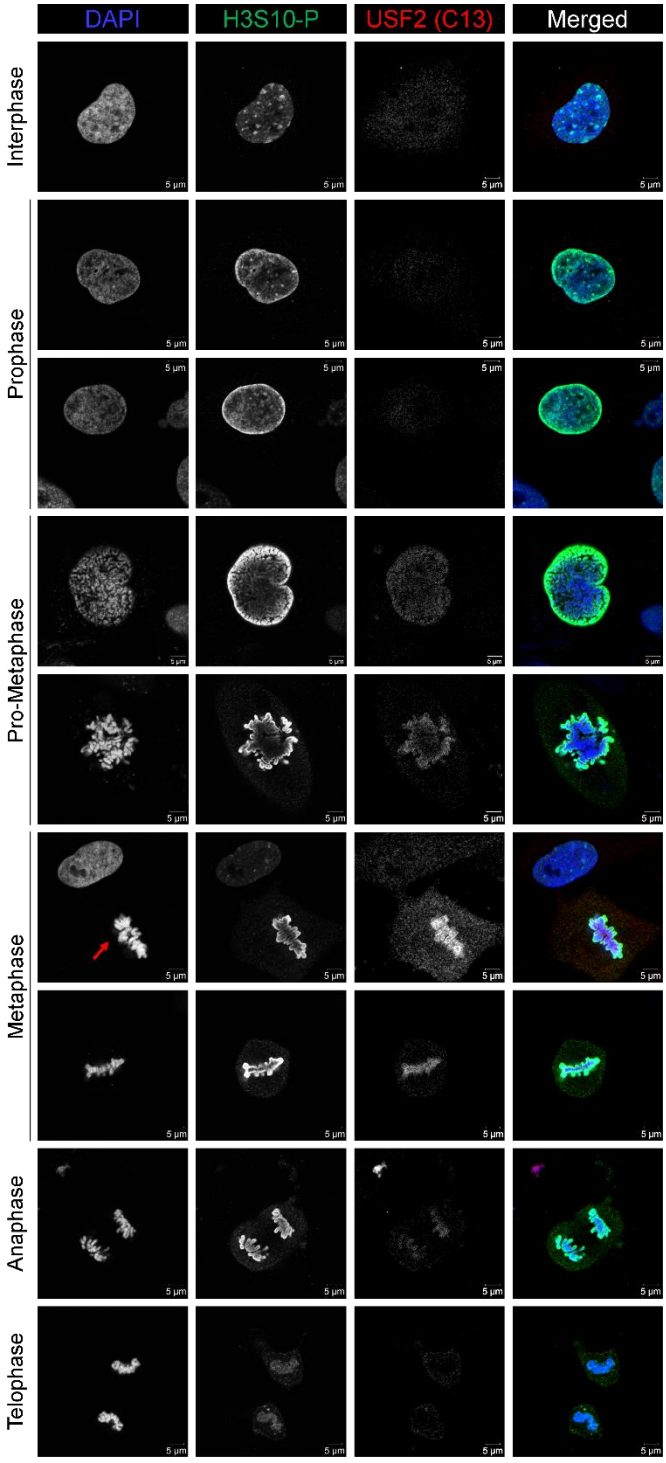
USF2 and mitosis



Supplementary Figure 3.4.15 Staining of USF2 as detected by N18-antibody at different stages of mitosis in IGR37 cell line. Red arrows indicate cells undergoing mitosis. Scale bars are shown at the corner of the images.



Supplementary Figure 3.4.16 Staining of USF2 as detected by Ab- and A19-antibody along with USF1 at different stages of the mitosis in HeLa cell line. Red arrows indicate cells undergoing mitosis. Scale bars are shown at the corner of the images.



Supplementary Figure 3.4.17 Staining of USF2 as detected by C13-antibody at different stages of mitosis in BT549 cell line. Red arrow indicates cell undergoing mitosis. Scale bars are shown at the corner of the images.

Bibliography

Abdel-Malek, Z., Suzuki, I., Tada, A., Im, S., and Akcali, C. (1999). The melanocortin-1 receptor and human pigmentation. Ann N Y Acad Sci *885*, 117-133.

Acker, T., Diez-Juan, A., Aragones, 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-2α. Cancer Cell *8*, 131-141.

Adelman, D.M., Gertsenstein, M., Nagy, A., Simon, M.C., and Maltepe, E. (2000). Placental cell fates are regulated in vivo by HIF-mediated hypoxia responses. Genes Dev *14*, 3191-3203.

Adhikary, S., and Eilers, M. (2005). Transcriptional regulation and transformation by Myc proteins. Nature Reviews Mol Cell Biol *6*, 635-645.

Ahmed, S., and Brickner, J.H. (2007). Regulation and epigenetic control of transcription at the nuclear periphery. Trends in Genetics *23*, 396-402.

Aksan, I., and Goding, C.R. (1998). Targeting the microphthalmia basic helix-loop-helix-leucine zipper transcription factor to a subset of E-box elements in vitro and in vivo. Mol Cell Biol *18*, 6930-6938.

Amati, B., Brooks, M.W., Levy, N., Littlewood, T.D., Evan, G.I., and Land, H. (1993a). Oncogenic activity of the c-Myc protein requires dimerization with Max. Cell *72*, 233-245.

Amati, B., Dalton, S., Brooks, M.W., Littlewood, T.D., Evan, G.I., and Land, H. (1992). Transcriptional activation by the human c-Myc oncoprotein in yeast requires interaction with Max. Nature *359*, 423-426.

Amati, B., Littlewood, T.D., Evan, G.I., and Land, H. (1993b). The c-Myc protein induces cell cycle progression and apoptosis through dimerization with Max. EMBO J *12*, 5083-5087.

Anantharaman, A., Lin, I.-J., Barrow, J., Liang, S.Y., Masannat, J., Strouboulis, J., Huang, S., and Bungert, J. (2011). Role of Helix-Loop-Helix Proteins during Differentiation of Erythroid Cells. Mol Cell Biol *31*, 1332-1343.

Andrulis, E.D., Neiman, A.M., Zappulla, D.C., and Sternglanz, R. (1998). Perinuclear localization of chromatin facilitates transcriptional silencing. Nature *394*, 592-595.

Anthony-Cahill, S.J., Benfield, P.A., Fairman, R., Wasserman, Z.R., Brenner, S.L., Stafford, W.F., 3rd, Altenbach, C., Hubbell, W.L., and DeGrado, W.F. (1992). Molecular characterization of helix-loop-helix peptides. Science *255*, 979-983.

Aoyagi, S., Wade, P.A., and Hayes, J.J. (2003). Nucleosome sliding induced by the xMi-2 complex does not occur exclusively via a simple twist-diffusion mechanism. J Biol Chem *278*, 30562-30568.

Aprelikova, O., Wood, M., Tackett, S., Chandramouli, G.V., and Barrett, J.C. (2006). Role of ETS transcription factors in the hypoxia-inducible factor-2 target gene selection. Cancer Res *66*, 5641-5647.

Arsura, M., Deshpande, A., Hann, S.R., and Sonenshein, G.E. (1995). Variant Max protein, derived by alternative splicing, associates with c-Myc in vivo and inhibits transactivation. Mol Cell Biol *15*, 6702-6709.

Askew, D.S., Ashmun, R.A., Simmons, B.C., and Cleveland, J.L. (1991). Constitutive c-myc expression in an IL-3-dependent myeloid cell line suppresses cell cycle arrest and accelerates apoptosis. Oncogene *6*, 1915-1922.

Atchley, W.R., and Fitch, W.M. (1997). A natural classification of the basic helix-loop-helix class of transcription factors. Proc Natl Acad Sci USA *94*, 5172-5176.

Austen, M., Cerni, C., Luscher-Firzlaff, J.M., and Luscher, B. (1998). YY1 can inhibit c-Myc function through a mechanism requiring DNA binding of YY1 but neither its transactivation domain nor direct interaction with c-Myc. Oncogene *17*, 511-520.

Avvakumov, N., Nourani, A., and Côté, J. (2011). Histone Chaperones: Modulators of Chromatin Marks. Mol Cell *41*, 502-514.

Ayer, D.E., and Eisenman, R.N. (1993). Inhibition of cell proliferation by the Mad1 transcriptional repressor. Genes Dev *7*, 2110.

Ayer, D.E., Kretzner, L., and Eisenman, R.N. (1993). Mad: A heterodimeric partner for Max that antagonizes Myc transcriptional activity. Cell *72*, 211-222.

Baba, M., Hirai, S., Yamada-Okabe, H., Kenji, H., Tabuchi, H., Kobayashi, K., Kondo, K., Yoshida, M., Yamashita, A., Kishida, T.*, et al.* (2003). Loss of von Hippel-Lindau protein causes cell density dependent deregulation of CyclinD1 expression through Hypoxia-inducible factor. Oncogene *22*, 2728-2738.

Bahram, F., Wu, S., Oberg, F., Luscher, B., and Larsson, L.G. (1999). Posttranslational regulation of Myc function in response to phorbol ester/interferon-gamma-induced differentiation of v-Myc-transformed U-937 monoblasts. Blood *93*, 3900-3912.

Bai, L., and Morozov, A.V. (2010). Gene regulation by nucleosome positioning. Trends in Genetics *26*, 476-483.

Barbaric, S., Münsterkötter, M., Goding, C., and Hörz, W. (1998). Cooperative Pho2-Pho4 interactions at the *PHO5* promoter are critical for binding of Pho4 to UASp1 and for efficient transactivation by Pho4 at UASp2. Mol Cell Biol *18*, 2629-2639.

Bargaje, R., Alam, M.P., Patowary, A., Sarkar, M., Ali, T., Gupta, S., Garg, M., Singh, M., Purkanti, R., Scaria, V.*, et al.* (2012). Proximity of H2A.Z containing nucleosome to the transcription start site influences gene expression levels in the mammalian liver and brain. Nucleic Acids Research *40*, 8965-8978.

Barnett, S.C., and Crouch, D.H. (1995). The effect of oncogenes on the growth and differentiation of oligodendrocyte type 2 astrocyte progenitor cells. Cell Growth and Differentiation *6*, 69-80.

Barnett, S.C., Robertson, L., Graham, D., Allan, D., and Rampling, R. (1998). Oligodendrocyte-type-2 astrocyte (O-2A) progenitor cells transformed with c-myc and H-ras form high-grade glioma after stereotactic injection into the rat brain. Carcinogenesis *19*, 1529-1537.

Baron, Y., Corre, S., Mouchet, N., Vaulont, S., Prince, S., and Galibert, M.D. (2012). USF-1 is critical for maintaining genome integrity in response to UV-induced DNA photolesions. PLoS Genetics *8*, e1002470.

Barrero, M.J., and Malik, S. (2013). The RNA polymerase II transcriptional machinery and its epigenetic context. Sub-cellular Biochemistry *61*, 237-259.

Bello-Fernandez, C., Packham, G., and Cleveland, J.L. (1993). The ornithine decarboxylase gene is a transcriptional target of c-Myc. Proc Natl Acad Sci USA *90*, 7804-7808.

Bendall, A.J., and Molloy, P.L. (1994). Base preferences for DNA binding by the bHLH-Zip protein USF: effects of MgCl2 on specificity and comparison with binding of Myc family members. Nucleic Acids Research *22*, 2801-2810.

Bengtsson, S.H., Madeyski-Bengtson, K., Nilsson, J., and Bjursell, G. (2002). Transcriptional regulation of the human carboxyl ester lipase gene in THP-1 monocytes: an E-box required for activation binds upstream stimulatory factors 1 and 2. The Biochemical Journal *365*, 481-488.

Bentley, N.J., Eisen, T., and Goding, C.R. (1994). Melanocyte-specific expression of the human tyrosinase promoter: activation by the microphthalmia gene product and role of the initiator. Mol Cell Biol *14*, 7996-8006.

Berberich, S., Hyde-DeRuyscher, N., Espenshade, P., and Cole, M. (1992). max encodes a sequence-specific DNA-binding protein and is not regulated by serum growth factors. Oncogene *7*, 775-779.

Berberich, S.J., and Cole, M.D. (1992). Casein kinase II inhibits the DNA-binding activity of Max homodimers but not Myc/Max heterodimers. Genes Dev *6*, 166-176.

Bergot, M.O., Diaz-Guerra, M.J., Puzenat, N., Raymondjean, M., and Kahn, A. (1992). Cis-regulation of the L-type pyruvate kinase gene promoter by glucose, insulin and cyclic AMP. Nucleic Acids Research *20*, 1871-1877.

Bernstein, B., Mikkelsen, T., Xie, X., Kamal, M., Huebert, D., Cuff, J., Fry, B., Meissner, A., Wernig, M., Plath, K.*, et al.* (2006). A bivalent chromatin structure marks key developmental genes in embryonic stem cells. Cell *125*, 315 - 326.

Bertolotto, C., Abbe, P., Hemesath, T.J., Bille, K., Fisher, D.E., Ortonne, J.P., and Ballotti, R. (1998a). Microphthalmia gene product as a signal transducer in cAMP-induced differentiation of melanocytes. Journal of Cell Biology *142*, 827-835.

Bertolotto, C., Bille, K., Ortonne, J.P., and Ballotti, R. (1996). Regulation of tyrosinase gene expression by cAMP in B16 melanoma cells involves two CATGTG motifs surrounding the TATA box: implication of the microphthalmia gene product. Journal of Cell Biology *134*, 747-755.

Bertolotto, C., Busca, R., Abbe, P., Bille, K., Aberdam, E., Ortonne, J.P., and Ballotti, R. (1998b). Different cis-acting elements are involved in the regulation of TRP1 and TRP2 promoter activities by cyclic AMP: pivotal role of M boxes (GTCATGTGCT) and of microphthalmia. Mol Cell Biol *18*, 694-702.

Bidder, M., Shao, J.S., Charlton-Kachigian, N., Loewy, A.P., Semenkovich, C.F., and Towler, D.A. (2002). Osteopontin transcription in aortic vascular smooth muscle cells is controlled by glucose-regulated upstream stimulatory factor and activator protein-1 activities. J Biol Chem *277*, 44485-44496.

Bird, A.P., and Wolffe, A.P. (1999). Methylation-induced repression--belts, braces, and chromatin. Cell *99*, 451-454.

Blackwell, T.K., Huang, J., Ma, A., Kretzner, L., Alt, F.W., Eisenman, R.N., and Weintraub, H. (1993). Binding of myc proteins to canonical and noncanonical DNA sequences. Mol Cell Biol *13*, 5216-5224.

Blackwell, T.K., Kretzner, L., Blackwood, E.M., Eisenman, R.N., and Weintraub, H. (1990). Sequence-specific DNA binding by the c-Myc protein. Science *250*, 1149-1151.

Blackwood, E.M., and Eisenman, R.N. (1991). Max: A helix-loop-helix zipper protein that forms a sequence-specific DNA-binding complex with Myc. Science *251*, 1211-1217.

Blackwood, E.M., Lüscher, B., and Eisenman, R.N. (1992). Myc and Max associate in vivo. Genes Dev *6*, 71-80.

Bohmann, D. (1990). Transcription factor phosphorylation: A link between signal transduction and the regulation of gene expression. Cancer Cell *2*, 337-344.

Bossard, P., and Zaret, K.S. (1998). GATA transcription factors as potentiators of gut endoderm differentiation. Development *125*, 4909-4917.

Bouchard, C., Dittrich, O., Kiermaier, A., Dohmann, K., Menkel, A., Eilers, M., and Lüscher, B. (2001). Regulation of cyclin D2 gene expression by the Myc/Max/Mad network: Myc-dependent TRRAP recruitment and histone acetylation at the cyclin D2 promoter. Genes Dev *15*, 2042-2047.

Bousset, K., Henriksson, M., Luscher-Firzlaff, J.M., Litchfield, D.W., and Luscher, B. (1993). Identification of casein kinase II phosphorylation sites in Max: effects on DNA-binding kinetics of Max homo- and Myc/Max heterodimers. Oncogene *8*, 3211-3220.

Boyd, K.E., and Farnham, P.J. (1997). Myc versus USF: discrimination at the cad gene is determined by core promoter elements. Mol Cell Biol *17*, 2529-2537.

Boyd, K.E., and Farnham, P.J. (1999). Coexamination of site-specific transcription factor binding and promoter activity in living cells. Mol Cell Biol *19*, 8393-8399.

Boyd, K.E., Wells, J., Gutman, J., Bartley, S.M., and Farnham, P.J. (1998). c-Myc target gene specificity is determined by a post-DNAbinding mechanism. Proc Natl Acad Sci USA *95*, 13887-13892.

Boyle, W.J., Smeal, T., Defize, L.H., Angel, P., Woodgett, J.R., Karin, M., and Hunter, T. (1991). Activation of protein kinase C decreases phosphorylation of c-Jun at sites that negatively regulate its DNA-binding activity. Cell *64*, 573-584.

Bracken, C.P., Whitelaw, M.L., and Peet, D.J. (2005). Activity of hypoxia-inducible factor 2alpha is regulated by association with the NF-kappaB essential modulator. J Biol Chem *280*, 14240-14251.

Breen, G.A., and Jordan, E.M. (2000). Upstream stimulatory factor 2 stimulates transcription through an initiator element in the mouse cytochrome c oxidase subunit Vb promoter. Biochimica et biophysica acta *1517*, 119-127.

Brickner, D.G., Cajigas, I., Fondufe-Mittendorf, Y., Ahmed, S., Lee, P.-C., Widom, J., and Brickner, J.H. (2007). H2A.Z-Mediated Localization of Genes at the Nuclear Periphery Confers Epigenetic Memory of Previous Transcriptional State. PLoS biology *5*, e81.

Brickner, J.H. (2009). Transcriptional memory at the nuclear periphery. Current Opinion in Cell Biology *21*, 127-133.

Brickner, J.H., and Walter, P. (2004). Gene recruitment of the activated INO1 locus to the nuclear membrane. PLoS Biol *2*, e342.

Brownlie P, Ceska T, Lamers M, Romier C, Stier G, Teo H, and D., S. (1997). The crystal structure of an intact human Max-DNA complex: new insights into mechanisms of transcriptional control. Structure, 509-520.

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

Bruno, M.E., West, R.B., Schneeman, T.A., Bresnick, E.H., and Kaetzel, C.S. (2004). Upstream stimulatory factor but not c-Myc enhances transcription of the human polymeric immunoglobulin receptor gene. Mol Immunol *40*, 695-708.

Bui, T.V., and Mendell, J.T. (2010). Myc: Maestro of MicroRNAs. Genes Cancer *1*, 568-575.

BuscÀ, R., and Ballotti, R. (2000). Cyclic AMP a Key Messenger in the Regulation of Skin Pigmentation. Pigment Cell Res *13*, 60-69.

Campanero, M.R., Armstrong, M.I., and Flemington, E.K. (2000). CpG methylation as a mechanism for the regulation of E2F activity. Proc Natl Acad Sci USA *97*, 6481-6486.

Carlson, H., Ota, S., Campbell, C.E., and Hurlin, P.J. (2001). A dominant repression domain in Tbx3 mediates transcriptional repression and cell immortalization: relevance to mutations in Tbx3 that cause ulnar-mammary syndrome. Human Molecular Genetics *10*, 2403-2413.

Carmeliet, P., Dor, Y., Herber, J.M., Fukumura, D., Brusselmans, K., Dewerchin, M., Neeman, M., Bono, F., Abramovitch, R., Maxwell, P., *et al.* (1998). Role of HIF-1α in hypoxiamediated apoptosis, cell proliferation and tumour angiogenesis. Nature *394*, 485-490.

Carrero, P., Okamoto, K., Coumailleau, P., O'Brien, S., Tanaka, H., and Poellinger, L. (2000). Redox-regulated recruitment of the transcriptional coactivators CREB-binding protein and SRC-1 to hypoxia-inducible factor 1alpha. Mol Cell Biol *20*, 402-415.

Carthew, R.W., Chodosh, L.A., and Sharp, P.A. (1985). An RNA polymerase II transcription factor binds to an upstream element in the adenovirus major late promoter. Cell *43*, 439-448.

Casado, M., Vallet, V.S., Kahn, A., and Vaulont, S. (1999). Essential Role in Vivo of Upstream Stimulatory Factors for a Normal Dietary Response of the Fatty Acid Synthase Gene in the Liver. J Biol Chem *274*, 2009-2013.

Chai, J., and Tarnawski, A.S. (2002). Serum response factor: discovery, biochemistry, biological roles and implications for tissue injury healing. Journal of physiology and pharmacology : an official journal of the Polish Physiological Society *53*, 147-157.

Chanda, D., Li, T., Song, K.H., Kim, Y.H., Sim, J., Lee, C.H., Chiang, J.Y., and Choi, H.S. (2009). Hepatocyte growth factor family negatively regulates hepatic gluconeogenesis via induction of orphan nuclear receptor small heterodimer partner in primary hepatocytes. J Biol Chem *284*, 28510-28521.

Chang, J.T., Yang, H.T., Wang, T.C., and Cheng, A.J. (2005). Upstream stimulatory factor (USF) as a transcriptional suppressor of human telomerase reverse transcriptase (hTERT) in oral cancer cells. Molecular Carcinogenesis *44*, 183-192.

Chen, J., Willingham, T., Margraf, L.R., Schreiber-Agus, N., DePinho, R.A., and Nisen, P.D. (1995). Effects of the MYC oncogene antagonist, MAD, on proliferation, cell cycling and the malignant phenotype of human brain tumour cells. Nature Medicine *1*, 638-643.

Cheng, L.L., Li, B., Luo, J.D., Tu, H.B., Liu, Q.C., and Ran, P. (2005). Identification of an E-box motif as a transcriptional repressor element in the proximal promoter region of the GCLC gene in rat lung epithelial L2 cells. Free Radical Biology & Medicine *39*, 1030-1040.

Cheung, E., Mayr, P., Coda-Zabetta, F., Woodman, P.G., and Boam, D.S. (1999). DNA-binding activity of the transcription factor upstream stimulatory factor 1 (USF-1) is regulated by cyclin-dependent phosphorylation. Biochemical Journal *344 Pt 1*, 145-152.

Chhajlani, V., and Wikberg, J.E.S. (1992). Molecular cloning and expression of the human melanocyte stimulating hormone receptor cDNA. FEBS Letters *309*, 417-420.

Chin, L., Merlino, G., and DePinho, R.A. (1998). Malignant melanoma: modern black plague and genetic black box. Genes Dev *12*, 3467-3481.

Chin, L., Schreiber-Agus, N., Pellicer, I., Chen, K., Lee, H.W., Dudast, M., Cordon-Cardo, C., and DePinho, R.A. (1995). Contrasting roles for Myc and Mad proteins in cellular growth and differentiation. Proc Natl Acad Sci USA *92*, 8488-8492.

Choe, C., Chen, N., and Sawadogo, M. (2005). Decreased tumorigenicity of c-Myc-transformed fibroblasts expressing active USF2. Exp Cell Res *302*, 1-10.

Ciccone, N.A., Lacza, C.T., Hou, M.Y., Gregory, S.J., Kam, K.Y., Xu, S., and Kaiser, U.B. (2008). A composite element that binds basic helix loop helix and basic leucine zipper transcription factors is important for gonadotropin-releasing hormone regulation of the follicle-stimulating hormone beta gene. Molecular endocrinology (Baltimore, Md) *22*, 1908-1923.

Clerk, A., Michael, A., and Sugden, P.H. (1998). Stimulation of the p38 mitogen-activated protein kinase pathway in neonatal rat ventricular myocytes by the G protein-coupled receptor agonists, endothelin-1 and phenylephrine: a role in cardiac myocyte hypertrophy? Journal of Cell Biology *142*, 523-535.

Cockman, M.E., Masson, N., Mole, D.R., Jaakkola, P., Chang, G.W., Clifford, S.C., Maher, E.R., Pugh, C.W., Ratcliffe, P.J., and Maxwell, P.H. (2000). Hypoxia inducible factor-alpha binding and ubiquitylation by the von Hippel-Lindau tumor suppressor protein. J Biol Chem *275*, 25733-25741.

Coller, H.A., Grandori, C., Tamayo, P., Colbert, T., Lander, E.S., Eisenman, R.N., and Golub, T.R. (2000). Expression analysis with oligonucleotide microarrays reveals that MYC regulates genes involved in growth, cell cycle, signaling, and adhesion. Proc Natl Acad Sci USA *97*, 3260-3265.

Compernolle, 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-2α and inhibition of VEGF impair fetal lung maturation, whereas treatment with VEGF prevents fatal respiratory distress in premature mice. Nature Medicine *8*, 702-710.

Coppola, J.A., and Cole, M.D. (1986). Constitutive c-myc oncogene expression blocks mouse erythroleukaemia cell differentiation but not commitment. Nature *320*, 760-763.

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 Biology & Therapy *4*, 1285-1294.

Corre, S., and Galibert, M.D. (2005). Upstream stimulating factors: Highly versatile stress-responsive transcription factors. Pigment Cell Res *18*, 337-348.

Corre, S., Primot, A., Baron, Y., Le Seyec, J., Goding, C., and Galibert, M.-D. (2009). Target Gene Specificity of USF-1 Is Directed via p38-mediated Phosphorylation-dependent Acetylation. J Biol Chem *284*, 18851-18862.

Corre, S., Primot, A., Sviderskaya, E., Bennett, D.C., Vaulont, S., Goding, C.R., and Galibert, M.-D. (2004). UV-induced Expression of Key Component of the Tanning Process, the POMC and MC1R Genes, Is Dependent on the p-38-activated Upstream Stimulating Factor-1 (USF-1). J Biol Chem *279*, 51226-51233.

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-2alpha regulates Oct-4: effects of hypoxia on stem cell function, embryonic development, and tumor growth. Genes Dev *20*, 557-570.

Covello, K.L., Simon, M.C., and Keith, B. (2005). Targeted Replacement of Hypoxia-Inducible Factor-1α by a Hypoxia-Inducible Factor-2α Knock-in Allele Promotes Tumor Growth. Cancer Research *65*, 2277-2286.

Cowell, I.G. (1994). Repression versus activation in the control of gene transcription. Trends in Biochemical Sciences *19*, 38-42.

Crews, S.T. (1998). Control of cell lineage-specific development and transcription by bHLH-PAS proteins. Genes Dev *12*, 607-620.

Crouch, D.H., Fisher, F., Clark, W., Jayaraman, P.S., Goding, C.R., and Gillespie, D.A. (1993). Gene-regulatory properties of Myc helix-loop-helix/leucine zipper mutants: Max-dependent DNA binding and transcriptional activation in yeast correlates with transforming capacity. Oncogene *8*, 1849-1855.

Cui, K., Zang, C., Roh, T.Y., Schones, D.E., Childs, R.W., Peng, W., and Zhao, K. (2009). Chromatin signatures in multipotent human hematopoietic stem cells indicate the fate of bivalent genes during differentiation. Cell Stem Cell *4*, 80-93.

Cvekl, A., Sax, C.M., Bresnick, E.H., and Piatigorsky, J. (1994). A complex array of positive and negative elements regulates the chicken alpha A-crystallin gene: involvement of Pax-6, USF, CREB and/or CREM, and AP-1 proteins. Mol Cell Biol *14*, 7363-7376.

Dang, C.V. (1999). c-Myc target genes involved in cell growth, apoptosis, and metabolism. Mol Cell Biol *19*, 1-11.

Dang, C.V. (2006). The c-Myc target gene network. Semin Cancer Biol *16*, 253-264.

Dang, C.V., Le, A., and Gao, P. (2009). MYC-Induced Cancer Cell Energy Metabolism and Therapeutic Opportunities. Clinical Cancer Research *15*, 6479-6483.

Dang, C.V., McGuire, M., Buckmire, M., and Lee, W.M.F. (1989). Involvement of the 'leucine zipper' region in the oligomerization and transforming activity of human c-myc protein. Nature *337*, 664-666.

Davis, P.L., Miron, A., Andersen, L.M., Iglehart, J.D., and Marks, J.R. (1999). Isolation and initial characterization of the BRCA2 promoter. Oncogene *18*, 6000-6012.

Davis, R.L., and Turner, D.L. (2001). Vertebrate hairy and Enhancer of split related proteins: transcriptional repressors regulating cellular differentiation and embryonic patterning. Oncogene *20*, 8342-8357.

Dayan, F., Monticelli, M., Pouyssegur, J., and Pecou, E. (2009). Gene regulation in response to graded hypoxia: the non-redundant roles of the oxygen sensors PHD and FIH in the HIF pathway. Journal of theoretical biology *259*, 304-316.

Dayan, F., Roux, D., Brahimi-Horn, M.C., Pouyssegur, J., and Mazure, N.M. (2006). The oxygen sensor factor-inhibiting hypoxia-inducible factor-1 controls expression of distinct genes through the bifunctional transcriptional character of hypoxia-inducible factor-1alpha. Cancer Res *66*, 3688-3698.

de Wit, E., and van Steensel, B. (2009). Chromatin domains in higher eukaryotes: Insights from genome-wide mapping studies. Chromosoma *118*, 25-36.

Deal, R.B., Henikoff, J.G., and Henikoff, S. (2010). Genome-wide kinetics of nucleosome turnover determined by metabolic labeling of histones. Science *328*, 1161-1164.

Decaux, J.F., Marcillat, O., Pichard, A.L., Henry, J., and Kahn, A. (1991). Glucose-dependent and -independent effect of insulin on gene expression. J Biol Chem *266*, 3432-3438.

Deng, C., Li, Y., Liang, S., Cui, K., Salz, T., Yang, H., Tang, Z., Gallagher, P.G., Qiu, Y., Roeder, R.*, et al.* (2013). USF1 and hSET1A mediated epigenetic modifications regulate lineage differentiation and HoxB4 transcription. PLoS genetics *9*, e1003524.

DePinho, R.A., Schreiber-Agus, N., and Alt, F.W. (1991). myc Family Oncogenes in the Development of Normal and Neoplastic Cells. In Advances in Cancer Research, F.V.W. George, and K. George, eds. (Academic Press), pp. 1-46.

Desbarats, L., Gaubatz, S., and Eilers, M. (1996). Discrimination between different E-box-binding proteins at an endogenous target gene of c-myc. Genes Dev *10*, 447-460.

Dieppois, G., and Stutz, F. (2010). Connecting the transcription site to the nuclear pore: a multi-tether process that regulates gene expression. Journal of cell science *123*, 1989-1999.

Dimova, E.Y., and Kietzmann, T. (2006). Cell type-dependent regulation of the hypoxia-responsive plasminogen activator inhibitor-1 gene by upstream stimulatory factor-2. J Biol Chem *281*, 2999-3005.

Dorsky, R.I., Raible, D.W., and Moon, R.T. (2000). Direct regulation of nacre, a zebrafish MITF homolog required for pigment cell formation, by the Wnt pathway. Genes Dev *14*, 158-162.

Du, H., Roy, A.L., and Roeder, R.G. (1993). Human transcription factor USF stimulates transcription through the initiator elements of the HIV-1 and the Ad-ML promoters. EMBO J *12*, 501-511.

Eckner, R., Ewen, M.E., Newsome, D., Gerdes, M., DeCaprio, J.A., Lawrence, J.B., and Livingston, D.M. (1994). Molecular cloning and functional analysis of the adenovirus E1A-associated 300-kD protein (p300) reveals a protein with properties of a transcriptional adaptor. Genes Dev *8*, 869-884.

Edmondson, D.G., and Olson, E.N. (1993). Helix-loop-helix proteins as regulators of muscle-specific transcription. J Biol Chem *268*, 755-758.

Ehrlich, M. (2003). Expression of various genes is controlled by DNA methylation during mammalian development. Journal of Cellular Biochemistry *88*, 899-910.

Eilers, M., and Eisenman, R.N. (2008). Myc's broad reach. Genes Dev *22*, 2755-2766.

Eilers, M., Schirm, S., and Bishop, J.M. (1991). The MYC protein activates transcription of the alpha-prothymosin gene. EMBO J *10*, 133-141.

Elbein, S.C., Hoffman, M.D., Teng, K., Leppert, M.F., and Hasstedt, S.J. (1999). A genome-wide search for type 2 diabetes susceptibility genes in Utah Caucasians. Diabetes *48*, 1175-1182.

Ema, M., Hirota, K., Mimura, J., Abe, H., Yodoi, J., Sogawa, K., Poellinger, L., and Fujii-Kuriyama, Y. (1999). Molecular mechanisms of transcription activation by HLF and HIF1alpha in response to hypoxia: their stabilization and redox signal-induced interaction with CBP/p300. EMBO J *18*, 1905-1914.

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 1α regulates the VEGF expression and is potentially involved in lung and vascular development. Proc Natl Acad Sci USA *94*, 4273-4278.

Epstein, A.C., Gleadle, J.M., McNeill, L.A., Hewitson, K.S., O'Rourke, J., Mole, D.R., Mukherji, M., Metzen, E., Wilson, M.I., Dhanda, A.*, et al.* (2001). C. elegans EGL-9 and mammalian homologs define a family of dioxygenases that regulate HIF by prolyl hydroxylation. Cell *107*, 43-54.

Ericsson, J., and Edwards, P.A. (1998). CBP is required for sterol-regulated and sterol regulatory element-binding protein-regulated transcription. J Biol Chem *273*, 17865-17870.

Erler, J.T., Bennewith, K.L., Nicolau, M., Dornhofer, N., Kong, C., Le, Q.T., Chi, J.T., Jeffrey, S.S., and Giaccia, A.J. (2006). Lysyl oxidase is essential for hypoxia-induced metastasis. Nature *440*, 1222-1226.

Evan, G.I., and Littlewood, T.D. (1993). The role of c-myc in cell growth. Current Opinion in Genetics & Development *3*, 44-49.

Fabbro, M., and Henderson, B.R. (2003). Regulation of tumor suppressors by nuclear-cytoplasmic shuttling. Experimental Cell Research *282*, 59-69.

Facchini, L.M., and Penn, L.Z. (1998). The molecular role of Myc in growth and transformation: recent discoveries lead to new insights. FASEB *12*, 633-651.

Farese, R.V., Sajan, M.P., and Standaert, M.L. (2005). Insulin-sensitive protein kinases (atypical protein kinase C and protein kinase B/Akt): actions and defects in obesity and type II diabetes. Experimental Biology and Medicine *230*, 593-605.

Farrell, A.S., and Sears, R.C. (2014). MYC Degradation. Cold Spring Harbor Perspectives in Medicine *4*.

Fernandez, P.C., Frank, S.R., Wang, L., Schroeder, M., Liu, S., Greene, J., Cocito, A., and Amati, B. (2003). Genomic targets of the human c-Myc protein. Genes Dev *17*, 1115-1129.

Ferre-D'Amare, A.R., Pognonec, P., Roeder, R.G., and Burley, S.K. (1994). Structure and function of the b/HLH/Z domain of USF. EMBO J *13*, 180-189.

Ferre-D'Amare, A.R., Prendergast, G.C., Ziff, E.B., and Burley, S.K. (1993). Recognition by Max of its cognate DNA through a dimeric b/HLH/Z domain. Nature *363*, 38-45.

Feuerbach, F., Galy, V., Trelles-Sticken, E., Fromont-Racine, M., Jacquier, A., Gilson, E., Olivo-Marin, J.C., Scherthan, H., and Nehrbass, U. (2002). Nuclear architecture and spatial positioning help establish transcriptional states of telomeres in yeast. Nature Cell Biology *4*, 214-221.

Fisher, D.E., Parent, L.A., and Sharp, P.A. (1993a). High affinity DNA-binding Myc analogs: recognition by an alpha helix. Cell *72*, 467-476.

Fisher, F., Crouch, D.H., Jayaraman, P.S., Clark, W., Gillespie, D.A., and Goding, C.R. (1993b). Transcription activation by Myc and Max: flanking sequences target activation to a subset of CACGTG motifs in vivo. EMBO J *12*, 5075-5082.

FitzGerald, L.M., Fryer, J.L., Dwyer, T., and Humphrey, S.M. (2006). Effect of MELANOTAN®, [Nle4, D-Phe7]-[alpha]-MSH, on melanin synthesis in humans with MC1R variant alleles. Peptides *27*, 388-394.

Frank, S.R., Parisi, T., Taubert, S., Fernandez, P., Fuchs, M., Chan, H.M., Livingston, D.M., and Amati, B. (2003). MYC recruits the TIP60 histone acetyltransferase complex to chromatin. EMBO reports *4*, 575-580.

Frank, S.R., Schroeder, M., Fernandez, P., Taubert, S., and Amati, B. (2001). Binding of c-Myc to chromatin mediates mitogen-induced acetylation of histone H4 and gene activation. Genes Dev *15*, 2069-2082.

Fujimi, T.J., and Aruga, J. (2008). Upstream stimulatory factors, USF1 and USF2 are differentially expressed during Xenopus embryonic development. GEP *8*, 376-381.

Gaggioli, C., Busca, R., Abbe, P., Ortonne, J.P., and Ballotti, R. (2003). Microphthalmia-Associated Transcription Factor (MITF) is Required but is Not Sufficient to Induce the Expression of Melanogenic Genes. Pigment Cell Res *16*, 374-382.

Gaj, T., Gersbach, C.A., and Barbas, C.F., 3rd (2013). ZFN, TALEN, and CRISPR/Cas-based methods for genome engineering. Trends in Biotechnology *31*, 397-405.

Galibert, M.-D., Carreira, S., and Goding, C.R. (2001a). The Usf-1 transcription factor is a novel target for the stress-responsive p38 kinase and mediates UV-induced Tyrosinase expression. EMBO J *20*, 5022-5031.

Ganss, R., Schutz, G., and Beermann, F. (1994). The mouse tyrosinase gene. Promoter modulation by positive and negative regulatory elements. J Biol Chem *269*, 29808-29816.

Gao, E., Wang, Y., Alcorn, J.L., and Mendelson, C.R. (1997). The basic helix-loop-helix-zipper transcription factor USF1 regulates expression of the surfactant protein-A gene. J Biol Chem *272*, 23398-23406.

García-Borrón, J.C., Sánchez-Laorden, B.L., and Jiménez-Cervantes, C. (2005). Melanocortin-1 receptor structure and functional regulation. Pigment Cell Res *18*, 393-410.

Gardner, L.B., Li, Q., Park, M.S., Flanagan, W.M., Semenza, G.L., and Dang, C.V. (2001). Hypoxia Inhibits G1/S Transition through Regulation of p27 Expression. J Biol Chem *276*, 7919-7926.

Garrell J., C.S. (1991). The helix-loop-helix domain: a common motif for bristles, muscles and sex. Bioessays *13(10)*, 493-498.

Gartel, A.L., and Shchors, K. (2003). Mechanisms of c-myc-mediated transcriptional repression of growth arrest genes. Exp Cell Res *283*, 17-21.

Giannola, D.M., Shlomchik, W.D., Jegathesan, M., Liebowitz, D., Abrams, C.S., Kadesch, T., Dancis, A., and Emerson, S.G. (2000). Hematopoietic expression of HOXB4 is regulated in normal and leukemic stem cells through transcriptional activation of the HOXB4 promoter by upstream stimulating factor (USF)-1 and USF-2. Journal of Experimental Medicine *192*, 1479-1490.

Gilkes, D.M., Semenza, G.L., and Wirtz, D. (2014). Hypoxia and the extracellular matrix: drivers of tumour metastasis. Nat Rev Cancer *14*, 430-439.

Gill, G. (2001). Regulation of the initiation of eukaryotic transcription. Essays in Biochemistry *37*, 33-43.

Goda, N., Ryan, H.E., Khadivi, B., McNulty, W., Rickert, R.C., and Johnson, R.S. (2003). Hypoxia-inducible factor 1α is essential for cell cycle arrest during hypoxia. Mol Cell Biol *23*, 359-369.

Goding, C.R. (2000). Mitf from neural crest to melanoma: signal transduction and transcription in the melanocyte lineage. Genes Dev *14*, 1712-1728.

Goldman, J.A., Garlick, J.D., and Kingston, R.E. (2010). Chromatin remodeling by imitation switch (ISWI) class ATP-dependent remodelers is stimulated by histone variant H2A.Z. J Biol Chem *285*, 4645-4651.

Gomez-Roman, N., Felton-Edkins, Z.A., Kenneth, N.S., Goodfellow, S.J., Athineos, D., Zhang, J., Ramsbottom, B.A., Innes, F., Kantidakis, T., Kerr, E.R., *et al.* (2006). Activation by c-Myc of transcription by RNA polymerases I, II and III. Biochemical Society symposium, 141-154.

Gordan, J.D., Bertout, J.A., Hu, C.-J., Diehl, J.A., and Simon, M.C. (2007a). HIF-2α Promotes Hypoxic Cell Proliferation by Enhancing c-Myc Transcriptional Activity. Cancer Cell *11*, 335-347.

Gordan, J.D., Lal, P., Dondeti, V.R., Letrero, R., Parekh, K.N., Oquendo, C.E., Greenberg, R.A., Flaherty, K.T., Rathmell, W.K., Keith, B., *et al.* (2008). HIF-alpha effects on c-Myc distinguish two subtypes of sporadic VHL-deficient clear cell renal carcinoma. Cancer Cell *14*, 435-446.

Gordan, J.D., and Simon, M.C. (2007). Hypoxia-inducible factors: central regulators of the tumor phenotype. Curr Opin Genet Dev *17*, 71-77.

Gordan, J.D., Thompson, C.B., and Simon, M.C. (2007b). HIF and c-Myc: sibling rivals for control of cancer cell metabolism and proliferation. Cancer Cell *12*, 108-113.

Goueli, B.S., and Janknecht, R. (2003). Regulation of telomerase reverse transcriptase gene activity by upstream stimulatory factor. Oncogene *22*, 8042-8047.

Grandori, C., Cowley, S.M., James, L.P., and Eisenman, R.N. (2000). The myc/max/mad network and the transcriptional control of cell behavior. Annual Review of Cell and Developmental Biology *16*, 653-699.

Grandori, C., Mac, J., Siebelt, F., Ayer, D.E., and Eisenman, R.N. (1996). Myc-Max heterodimers activate a DEAD box gene and interact with multiple E box-related sites in vivo. EMBO J *15*, 4344-4357.

Green, S.L., Freiberg, R.A., and Giaccia, A.J. (2001). p21Cip1 and p27Kip1 regulate cell cycle reentry after hypoxic stress but are not necessary for hypoxia-induced arrest. Mol Cell Biol *21*, 1196-1206.

Gregor, P.D., Sawadogo, M., and Roeder, R.G. (1990). The adenovirus major late transcription factor USF is a member of the helix-loop-helix group of regulatory proteins and binds to DNA as a dimer. Genes Dev *4*, 1730-1740.

Griffin, M.J., Wong, R.H., Pandya, N., and Sul, H.S. (2007). Direct interaction between USF and SREBP-1c mediates synergistic activation of the fatty-acid synthase promoter. J Biol Chem *282*, 5453-5467.

Grinberg, A.V., Hu, C.-D., and Kerppola, T.K. (2004). Visualization of Myc/Max/Mad Family Dimers and the Competition for Dimerization in Living Cells. Mol Cell Biol *24*, 4294-4308.

Griswold, M.D., and Kim, J.S. (2001). Site-specific methylation of the promoter alters deoxyribonucleic acid-protein interactions and prevents follicle-stimulating hormone receptor gene transcription. Biology of Reproduction *64*, 602-610.

Gu, J., Milligan, J., and Huang, L.E. (2001). Molecular mechanism of hypoxia-inducible factor 1alpha -p300 interaction. A leucine-rich interface regulated by a single cysteine. J Biol Chem *276*, 3550-3554.

Gu, W., Cechova, K., Tassi, V., and Dalla-Favera, R. (1993). Opposite regulation of gene transcription and cell proliferation by c-Myc and Max. Proc Natl Acad Sci USA *90*, 2935-2939.

Guccione, E., Martinato, F., Finocchiaro, G., Luzi, L., Tizzoni, L., Dall' Olio, V., Zardo, G., Nervi, C., Bernard, L., and Amati, B. (2006). Myc-binding-site recognition in the human genome is determined by chromatin context. Nat Cell Biol *8*, 764-770.

Gunaratnam, L., Morley, M., Franovic, A., de Paulsen, N., Mekhail, K., Parolin, D.A.E., Nakamura, E., Lorimer, I.A.J., and Lee, S. (2003). Hypoxia Inducible Factor Activates the Transforming Growth Factor-α/Epidermal Growth Factor Receptor Growth Stimulatory Pathway in $VHL^{-/-}$ Renal Cell Carcinoma Cells. J Biol Chem *278*, 44966-44974.

Gusterson, R., Brar, B., Faulkes, D., Giordano, A., Chrivia, J., and Latchman, D. (2002). The transcriptional co-activators CBP and p300 are activated via phenylephrine through the p42/p44 MAPK cascade. J Biol Chem *277*, 2517-2524.

Hahn, S., and Young, E.T. (2011). Transcriptional regulation in Saccharomyces cerevisiae: transcription factor regulation and function, mechanisms of initiation, and roles of activators and coactivators. Genetics *189*, 705-736.

Hakim, O., Sung, M.H., and Hager, G.L. (2010). 3D shortcuts to gene regulation. Current Opinion in Cell Biology *22*, 305-313.

Halazonetis, T.D., and Kandil, A.N. (1991). Determination of the c-MYC DNA-binding site. Proc Natl Acad Sci USA *88*, 6162-6166.

Hallsson, J.H., Favor, J., Hodgkinson, C., Glaser, T., Lamoreux, M.L., Magnusdottir, R., Gunnarsson, G.J., Sweet, H.O., Copeland, N.G., Jenkins, N.A., *et al.* (2000). Genomic, transcriptional and mutational analysis of the mouse microphthalmia locus. Genetics *155*, 291-300.

Hanson, R.L., Ehm, M.G., Pettitt, D.J., Prochazka, M., Thompson, D.B., Timberlake, D., Foroud, T., Kobes, S., Baier, L., Burns, D.K., *et al.* (1998). An autosomal genomic scan for loci linked to type II diabetes mellitus and body-mass index in Pima Indians. American Journal of Human Genetics *63*, 1130-1138.

Hardy, S., Jacques, P.E., Gevry, N., Forest, A., Fortin, M.E., Laflamme, L., Gaudreau, L., and Robert, F. (2009). The euchromatic and heterochromatic landscapes are shaped by antagonizing effects of transcription on H2A.Z deposition. PLoS genetics *5*, e1000687.

Harris, A.N., and Mellon, P.L. (1998). The basic helix-loop-helix, leucine zipper transcription factor, USF (upstream stimulatory factor), is a key regulator of SF-1 (steroidogenic factor-1) gene expression in pituitary gonadotrope and steroidogenic cells. Molecular endocrinology *12*, 714-726.

Hartley, P.D., and Madhani, H.D. (2009). Mechanisms that specify promoter nucleosome location and identity. Cell *137*, 445-458.

Hateboer, G., Timmers, H.T.M., Rustgi, A.K., Billaud, M., Van 'T Veer, L.J., and Bernards, R. (1993). TATA-binding protein and the retinoblastoma gene product bind to overlapping epitopes on c-Myc and adenovirus E1A protein. Proc Natl Acad Sci USA *90*, 8489-8493.

Heinz, S., Benner, C., Spann, N., Bertolino, E., Lin, Y.C., Laslo, P., Cheng, J.X., Murre, C., Singh, H., and Glass, C.K. (2010). Simple combinations of lineage-determining transcription factors prime cis-regulatory elements required for macrophage and B cell identities. Mol Cell *38*, 576-589.

Hemesath, T.J., Price, E.R., Takemoto, C., Badalian, T., and Fisher, D.E. (1998). MAP kinase links the transcription factor Microphthalmia to c-Kit signalling in melanocytes. Nature *391*, 298-301.

Hemesath, T.J., Steingrimsson, E., McGill, G., Hansen, M.J., Vaught, J., Hodgkinson, C.A., Arnheiter, H., Copeland, N.G., Jenkins, N.A., and Fisher, D.E. (1994). microphthalmia, a critical factor in melanocyte development, defines a discrete transcription factor family. Genes Dev *8*, 2770-2780.

Henikoff, S. (2008). Nucleosome destabilization in the epigenetic regulation of gene expression. Nat Rev Genet *9*, 15-26.

Henriksson, M., Bakardjiev, A., Klein, G., and Luscher, B. (1993). Phosphorylation sites mapping in the N-terminal domain of c-myc modulate its transforming potential. Oncogene *8*, 3199-3209.

Henriksson, M., and Lüscher, B. (1996). Proteins of the Myc Network: Essential Regulators of Cell Growth and Differentiation. In Advances in Cancer Research, F.V.W. George, and K. George, eds. (Academic Press), pp. 109-182.

Hill, H.Z. (1992). The function of melanin or six blind people examine an elephant. Bioessays *14*, 49-56.

Hirst, K., Fisher, F., McAndrew, P.C., and Goding, C.R. (1994). The transcription factor, the Cdk, its cyclin and their regulator: directing the transcriptional response to a nutritional signal. EMBO J *13*, 5410-5420.

Hock, T.D., Nick, H.S., and Agarwal, A. (2004). Upstream stimulatory factors, USF1 and USF2, bind to the human haem oxygenase-1 proximal promoter in vivo and regulate its transcription. Biochemical Journal *383*, 209-218.

Hodgkinson, C.A., Moore, K.J., Nakayama, A., Steingrimsson, E., Copeland, N.G., Jenkins, N.A., and Arnheiter, H. (1993). Mutations at the mouse microphthalmia locus are associated with defects in a gene encoding a novel basic-helix-loop-helix-zipper protein. Cell *74*, 395-404.

Hoek, K.S., and Goding, C.R. (2010). Cancer stem cells versus phenotype-switching in melanoma. Pigment Cell & Melanoma Res *23*, 746-759.

Hoffstedt, J., Ryden, M., Wahrenberg, H., van Harmelen, V., and Arner, P. (2005). Upstream transcription factor-1 gene polymorphism is associated with increased adipocyte lipolysis. Journal of Clinical Endocrinology and Metabolism *90*, 5356-5360.

Holmquist-Mengelbier, L., Fredlund, E., Löfstedt, T., Noguera, R., Navarro, S., Nilsson, H., Pietras, A., Vallon-Christersson, J., Borg, Å., Gradin, K., *et al.* (2006). Recruitment of HIF-1α and HIF-2α to common target genes is differentially regulated in neuroblastoma: HIF-2α promotes an aggressive phenotype. Cancer Cell *10*, 413-423.

Hoogaars, W.M.H., Barnett, P., Rodriguez, M., Clout, D.E., Moorman, A.F.M., Goding, C.R., and Christoffels, V.M. (2008). TBX3 and its splice variant TBX3 + exon 2a are functionally similar. Pigment Cell & Melanoma Res *21*, 379-387.

Höpfl, G., Wenger, R.H., Ziegler, U., Stallmach, T., Gardelle, O., Achermann, R., Wergin, M., Kaser-Hotz, B., Saunders, H.M., Williams, K.J., *et al.* (2002). Rescue of hypoxia-inducible factor-1α-deficient tumor growth by wild-type cells is independent of vascular endothelial growth factor. Cancer Res *62*, 2962-2970.

Hornyak, T.J., Hayes, D.J., Chiu, L.Y., and Ziff, E.B. (2001). Transcription factors in melanocyte development: distinct roles for Pax-3 and Mitf. Mech Dev *101*, 47-59.

Horton, J.D., Bashmakov, Y., Shimomura, I., and Shimano, H. (1998). Regulation of sterol regulatory element binding proteins in livers of fasted and refed mice. Proc Natl Acad Sci USA *95*, 5987-5992.

Howcroft, T.K., Murphy, C., Weissman, J.D., Huber, S.J., Sawadogo, M., and Singer, D.S. (1999). Upstream stimulatory factor regulates major histocompatibility complex class I gene expression: the U2DeltaE4 splice variant abrogates E-box activity. Mol Cell Biol *19*, 4788-4797.

Hu, C.-J., Wang, L.-Y., Chodosh, L.A., Keith, B., and Simon, M.C. (2003). Differential Roles of Hypoxia-Inducible Factor 1α (HIF-1α) and HIF-2α in Hypoxic Gene Regulation. Mol Cell Biol *23*, 9361-9374.

Hu, C.J., Sataur, A., Wang, L., Chen, H., and Simon, M.C. (2007). The N-terminal transactivation domain confers target gene specificity of hypoxia-inducible factors HIF-1alpha and HIF-2alpha. Molecular Biology of the Cell *18*, 4528-4542.

Hu, J., Stiehl, D.P., Setzer, C., Wichmann, D., Shinde, D.A., Rehrauer, H., Hradecky, P., Gassmann, M., and Gorr, T.A. (2011). Interaction of HIF and USF signaling pathways in human genes flanked by hypoxia-response elements and E-box palindromes. MCR *9*, 1520-1536.

Huang, L.E., Arany, Z., Livingston, D.M., and Bunn, H.F. (1996). Activation of hypoxia-inducible transcription factor depends primarily upon redox-sensitive stabilization of its alpha subunit. J Biol Chem *271*, 32253-32259.

Huang, S., Li, X., Yusufzai, T.M., Qiu, Y., and Felsenfeld, G. (2007). USF1 recruits histone modification complexes and is critical for maintenance of a chromatin barrier. Mol Cell Biol *27*, 7991-8002.

Hubner, M.R., Eckersley-Maslin, M.A., and Spector, D.L. (2013). Chromatin organization and transcriptional regulation. Curr Opin Genet Dev *23*, 89-95.

Hurlin PJ, Ayer DE, Grandori C, and RN, E. (1994). The Max transcription factor network: involvement of Mad in differentiation and an approach to identification of target genes. Cold Spring Harb Symp Quant Biol, pp. 109-116.

Hurlin, P.J., Quéva, C., and Eisenman, R.N. (1997). Mnt, a novel Max-interacting protein is coexpressed with Myc in proliferating cells and mediates repression at Myc binding sites. Genes Dev *11*, 44-58.

Hurlin, P.J., Quéva, C., Koskinen, P.J., Steingrímsson, E., Ayer, D.E., Copeland, N.G., Jenkins, N.A., and Eisenman, R.N. (1996). Mad3 and Mad4: novel Max-interacting transcriptional repressors that suppress c-myc dependent transformation and are expressed during neural and epidermal differentiation. EMBO J *15*, 2030.

Igarashi, M., Wakasaki, H., Takahara, N., Ishii, H., Jiang, Z.Y., Yamauchi, T., Kuboki, K., Meier, M., Rhodes, C.J., and King, G.L. (1999). Glucose or diabetes activates p38 mitogen-activated protein kinase via different pathways. Journal of Clinical Investigation *103*, 185-195.

Iordanov, M., Bender, K., Ade, T., Schmid, W., Sachsenmaier, C., Engel, K., Gaestel, M., Rahmsdorf, H.J., and Herrlich, P. (1997). CREB is activated by UVC through a p38/HOG-1-dependent protein kinase. EMBO J *16*, 1009-1022.

Ismail, P.M., Lu, T., and Sawadogo, M. (1999). Loss of USF transcriptional activity in breast cancer cell lines. Oncogene *18*, 5582-5591.

Ivan, M., Kondo, K., Yang, H., Kim, W., Valiando, J., Ohh, M., Salic, A., Asara, J.M., Lane, W.S., and Kaelin, W.G., Jr. (2001). HIFalpha targeted for VHL-mediated destruction by proline hydroxylation: implications for O2 sensing. Science *292*, 464-468.

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 1α. Genes Dev *12*, 149-162.

Jaakkola, P., Mole, D.R., Tian, Y.M., Wilson, M.I., Gielbert, J., Gaskell, S.J., von Kriegsheim, A., Hebestreit, H.F., Mukherji, M., Schofield, C.J.*, et al.* (2001). Targeting of HIF-alpha to the von Hippel-Lindau ubiquitylation complex by O2-regulated prolyl hydroxylation. Science *292*, 468-472.

Jacobs, J.J., Keblusek, P., Robanus-Maandag, E., Kristel, P., Lingbeek, M., Nederlof, P.M., van Welsem, T., van de Vijver, M.J., Koh, E.Y., Daley, G.Q.*, et al.* (2000). Senescence bypass screen identifies TBX2, which represses Cdkn2a (p19(ARF)) and is amplified in a subset of human breast cancers. Nat Genet *26*, 291-299.

Jaiswal, A.S., and Narayan, S. (2001). Upstream stimulating factor-1 (USF1) and USF2 bind to and activate the promoter of the adenomatous polyposis coli (APC) tumor suppressor gene. Journal of Cellular Biochemistry *81*, 262-277.

Janicki, S.M., Tsukamoto, T., Salghetti, S.E., Tansey, W.P., Sachidanandam, R., Prasanth, K.V., Ried, T., Shav-Tal, Y., Bertrand, E., Singer, R.H., *et al.* (2004). From Silencing to Gene Expression: Real-Time Analysis in Single Cells. Cell *116*, 683-698.

Jayaraman, P.S., Hirst, K., and Goding, C.R. (1994). The activation domain of a basic helix-loop-helix protein is masked by repressor interaction with domains distinct from that required for transcription regulation. EMBO J *13*, 2192-2199.

Jewell, U.R., Kvietikova, I., Scheid, A., Bauer, C., Wenger, R.H., and Gassmann, M. (2001). Induction of HIF-1alpha in response to hypoxia is instantaneous. FASEB *15*, 1312-1314.

Jiang, B., and Mendelson, C.R. (2003). USF1 and USF2 mediate inhibition of human trophoblast differentiation and CYP19 gene expression by Mash-2 and hypoxia. Mol Cell Biol *23*, 6117-6128.

Jiang, B.H., Zheng, J.Z., Leung, S.W., Roe, R., and Semenza, G.L. (1997). Transactivation and inhibitory domains of hypoxia-inducible factor 1alpha. Modulation of transcriptional activity by oxygen tension. J Biol Chem *272*, 19253-19260.

Jiang, C., and Pugh, B.F. (2009). Nucleosome positioning and gene regulation: advances through genomics. Nat Rev Genet *10*, 161-172.

Jimbow, K., Roth, S.I., Fitzpatrick, T.B., and Szabo, G. (1975). Mitotic activity in non-neoplastic melanocytes in vivo as determined by histochemical, autoradiographic, and electron microscope studies. Journal of Cell Biology *66*, 663-670.

Jin, C., and Felsenfeld, G. (2007). Nucleosome stability mediated by histone variants H3.3 and H2A.Z. Genes Dev *21*, 1519-1529.

Jin, C., Zang, C., Wei, G., Cui, K., Peng, W., Zhao, K., and Felsenfeld, G. (2009). H3.3/H2A.Z double variant-containing nucleosomes mark 'nucleosome-free regions' of active promoters and other regulatory regions. Nat Genet *41*, 941-945.

Jones, N. (1990). Transcriptional regulation by dimerization: two sides to an incestuous relationship. Cell *61*, 9-11.

Ju, B.G., Lunyak, V.V., Perissi, V., Garcia-Bassets, I., Rose, D.W., Glass, C.K., and Rosenfeld, M.G. (2006). A topoisomerase IIbeta-mediated dsDNA break required for regulated transcription. Science *312*, 1798-1802.

Jung, H.S., Kim, K.S., Chung, Y.J., Chung, H.K., Min, Y.K., Lee, M.S., Lee, M.K., Kim, K.W., and Chung, J.H. (2007). USF inhibits cell proliferation through delay in G2/M phase in FRTL-5 cells. Endocrine Journal *54*, 275-285.

Juttner, S., Cramer, T., Wessler, S., Walduck, A., Gao, F., Schmitz, F., Wunder, C., Weber, M., Fischer, S.M., Schmidt, W.E., *et al.* (2003). Helicobacter pylori stimulates host cyclooxygenase-2 gene transcription: critical importance of MEK/ERK-dependent activation of USF1/-2 and CREB transcription factors. Cellular Microbiology *5*, 821-834.

Kaplan, C.D. (2013). Basic mechanisms of RNA polymerase II activity and alteration of gene expression in Saccharomyces cerevisiae. Biochimica et biophysica acta *1829*, 39-54.

Kapon, R., Naim, B., Zbaida, D., Nevo, R., Tsabari, O., and Reich, Z. (2010). Permeating the nuclear pore complex. Nucleus *1*, 475-480.

Kato, G.J., Barrett, J., Villa-Garcia, M., and Dang, C.V. (1990). An amino-terminal c-Myc domain required for neoplastic transformation activates transcription Mol Cell Biol *10*, 5914-5920.

Kato, G.J., Lee, W.M., Chen, L.L., and Dang, C.V. (1992). Max: functional domains and interaction with c-Myc. Genes Dev *6*, 81-92.

Kaulen, H., Pognonec, P., Gregor, P.D., and Roeder, R.G. (1991). The Xenopus B1 factor is closely related to the mammalian activator USF and is implicated in the developmental regulation of TFIIIA gene expression. Mol Cell Biol *11*, 412-424.

Kawaguchi, T., Osatomi, K., Yamashita, H., Kabashima, T., and Uyeda, K. (2002). Mechanism for fatty acid "sparing" effect on glucose-induced transcription: regulation of carbohydrate-responsive element-binding protein by AMP-activated protein kinase. J Biol Chem *277*, 3829-3835.

Kaytor, E.N., Shih, H., and Towle, H.C. (1997). Carbohydrate regulation of hepatic gene expression. Evidence against a role for the upstream stimulatory factor. J Biol Chem *272*, 7525-7531.

Kennedy, H.J., Viollet, B., Rafiq, I., Kahn, A., and Rutter, G.A. (1997). Upstream stimulatory factor-2 (USF2) activity is required for glucose stimulation of L-pyruvate kinase promoter activity in single living islet beta-cells. J Biol Chem *272*, 20636-20640.

Kerkhoff, E., Bister, K., and Klempnauer, K.H. (1991). Sequence-specific DNA binding by Myc proteins. Proc Natl Acad Sci USA *88*, 4323-4327.

Kerr, I.M., Costa-Pereira, A.P., Lillemeier, B.F., and Strobl, B. (2003). Of JAKs, STATs, blind watchmakers, jeeps and trains. FEBS Letters *546*, 1-5.

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 Metabolism *3*, 177-185.

Kind, J., and van Steensel, B. (2010). Genome-nuclear lamina interactions and gene regulation. Current Opinion in Cell Biology *22*, 320-325.

King, R., Weilbaecher, K.N., McGill, G., Cooley, E., Mihm, M., and Fisher, D.E. (1999). Microphthalmia transcription factor. A sensitive and specific melanocyte marker for MelanomaDiagnosis. American Journal of Pathology *155*, 731-738.

Kingsley-Kallesen, M.L., Kelly, D., and Rizzino, A. (1999). Transcriptional regulation of the transforming growth factor-beta2 promoter by cAMP-responsive element-binding protein (CREB) and activating transcription factor-1 (ATF-1) is modulated by protein kinases and the coactivators p300 and CREB-binding protein. J Biol Chem *274*, 34020-34028.

Kirito, K., Fox, N., and Kaushansky, K. (2003). Thrombopoietin stimulates Hoxb4 expression: an explanation for the favorable effects of TPO on hematopoietic stem cells. Blood *102*, 3172-3178.

Kirschbaum, B.J., Pognonec, P., and Roeder, R.G. (1992). Definition of the transcriptional activation domain of recombinant 43-kilodalton USF. Mol Cell Biol *12*, 5094-5101.

Klempnauer, K.H. (1993). Methylation-sensitive DNA binding by v-myb and c-myb proteins. Oncogene *8*, 111-115.

Knoepfler, P.S. (2007). Myc goes global: new tricks for an old oncogene. Cancer Res *67*, 5061-5063.

Knoepfler, P.S., Zhang, X.Y., Cheng, P.F., Gafken, P.R., McMahon, S.B., and Eisenman, R.N. (2006). Myc influences global chromatin structure. EMBO J *25*, 2723-2734.

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.

Kooistra, S.M., and Helin, K. (2012). Molecular mechanisms and potential functions of histone demethylases. Nature Reviews Molecular Cell Biology *13*, 297-311.

Kornberg, R.D. (1974). Chromatin structure: a repeating unit of histones and DNA. Science *184*, 868-871.

Koshiji, M., Kageyama, Y., Pete, E.A., Horikawa, I., Barrett, J.C., and Huang, L.E. (2004). HIF-1α induces cell cycle arrest by functionally counteracting Myc. EMBO J *23*, 1949-1956.

Koshiji, M., To, K.K.W., Hammer, S., Kumamoto, K., Harris, A.L., Modrich, P., and Huang, L.E. (2005). HIF-1α Induces Genetic Instability by Transcriptionally Downregulating MutSα Expression. Mol Cell *17*, 793-803.

Koskinen, P., Ayer, D., and Eisenman, R. (1995). Repression of Myc-Ras cotransformation by Mad is mediated by multiple protein-protein interactions. Cell Growth Differ *6*, 623-629.

Kozak, K.R., Abbott, B., and Hankinson, O. (1997). ARNT-deficient mice and placental differentiation. Dev Biol *191*, 297-305.

Kretzner, L., Blackwood, E.M., and Eisenman, R.N. (1992). Myc and Max proteins possess distinct transcriptional activities. Nature *359*, 426-429.

Krishnamurthy, S., and Hampsey, M. (2009). Eukaryotic transcription initiation. Current Biology *19*, 153-156.

Krones, A., Jungermann, K., and Kietzmann, T. (2001). Cross-talk between the signals hypoxia and glucose at the glucose response element of the L-type pyruvate kinase gene. Endocrinology *142*, 2707-2718.

Krylov, D., Kasai, K., Echlin, D.R., Taparowsky, E.J., Arnheiter, H., and Vinson, C. (1997). A general method to design dominant negatives to B-HLHZip proteins that abolish DNA binding. Proc Natl Acad Sci USA *94*, 12274-12279.

Kumar, R., and McEwan, I.J. (2012). Allosteric Modulators of Steroid Hormone Receptors: Structural Dynamics and Gene Regulation. Endocrine Reviews *33*, 271-299.

Kundu, S., Horn, P.J., and Peterson, C.L. (2007). SWI/SNF is required for transcriptional memory at the yeast GAL gene cluster. Genes Dev *21*, 997-1004.

Kundu, S., and Peterson, C.L. (2009). Role of chromatin states in transcriptional memory. Biochimica et Biophysica Acta *1790*, 445-455.

Kung, A.L., Wang, S., Klco, J.M., Kaelin, W.G., and Livingston, D.M. (2000). Suppression of tumor growth through disruption of hypoxia-inducible transcription. Nature medicine *6*, 1335-1340.

Kwak, H., and Lis, J.T. (2013). Control of transcriptional elongation. Annual Review of Genetics *47*, 483-508.

Lahoz, E.G., Xu, L., Schreiber-Agus, N., and DePinho, R.A. (1994). Suppression of Myc, but not E1a, transformation activity by Max-associated proteins, Mad and Mxi1. Proc Natl Acad Sci USA *91*, 5503-5507.

Lalli, E., and Sassone-Corsi, P. (1994). Signal transduction and gene regulation: the nuclear response to cAMP. J Biol Chem *269*, 17359-17362.

Land, H., Parada, L.F., and Weinberg, R.A. (1983). Tumorigenic conversion of primary embryo fibroblasts requires at least two cooperating oncogenes. Nature *304*, 596-602.

Lando, D., Peet, D.J., Whelan, D.A., Gorman, J.J., and Whitelaw, M.L. (2002). Asparagine hydroxylation of the HIF transactivation domain a hypoxic switch. Science *295*, 858-861.

Langefeld, C.D., Wagenknecht, L.E., Rotter, J.I., Williams, A.H., Hokanson, J.E., Saad, M.F., Bowden, D.W., Haffner, S., Norris, J.M., Rich, S.S*., et al.* (2004). Linkage of the metabolic syndrome to 1q23-q31 in Hispanic families: the Insulin Resistance Atherosclerosis Study Family Study. Diabetes *53*, 1170-1174.

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

Larsson, L.G., Pettersson, M., Öberg, F., Nilsson, K., and Luscher, B. (1994). Expression of mad, mxi1, max and c-myc during induced differentiation of hematopoietic cells: opposite regulation of mad and c-myc. Oncogene *9(4):*, 1247-1252.

Latasa, M.J., Griffin, M.J., Moon, Y.S., Kang, C., and Sul, H.S. (2003). Occupancy and function of the -150 sterol regulatory element and -65 E-box in nutritional regulation of the fatty acid synthase gene in living animals. Mol Cell Biol *23*, 5896-5907.

Latchman, D.S. (1997). Transcription factors: An overview. Int J Biochem Cell Biol *29*, 1305-1312.

Law, C., and Cheung, P. (2013). Histone variants and transcription regulation. Sub-cellular biochemistry *61*, 319-341.

Ledent, V., Paquet, O., and Vervoort, M. (2002). Phylogenetic analysis of the human basic helix-loop-helix proteins. Genome Biol *3*, 0030.1–0030.18.

Lee, M., Goodall, J., Verastegui, C., Ballotti, R., and Goding, C.R. (2000). Direct regulation of the Microphthalmia promoter by Sox10 links Waardenburg-Shah syndrome (WS4)-associated hypopigmentation and deafness to WS2. J Biol Chem *275*, 37978-37983.

Lefrancois-Martinez, A.M., Martinez, A., Antoine, B., Raymondjean, M., and Kahn, A. (1995). Upstream stimulatory factor proteins are major components of the glucose response complex of the L-type pyruvate kinase gene promoter. J Biol Chem *270*, 2640-2643.

Lei, Z., Li, B., Yang, Z., Fang, H., Zhang, G.M., Feng, Z.H., and Huang, B. (2009). Regulation of HIF-1alpha and VEGF by miR-20b tunes tumor cells to adapt to the alteration of oxygen concentration. PLoS One *4*, e7629.

Lennard, A.C., and Egly, J.M. (1987). The bidirectional upstream element of the adenovirus-2 major late promoter binds a single monomeric molecule of the upstream factor. EMBO J *6*, 3027-3034.

Leong, K., Brunet, L., and Berk, A.J. (1988). Factors responsible for the higher transcriptional activity of extracts of adenovirus-infected cells fractionate with the TATA box transcription factor. Mol Cell Biol *8*, 1765-1774.

Leufgen, H., Bihl, M.P., Rudiger, J.J., Gambazzi, J., Perruchoud, A.P., Tamm, M., and Roth, M. (2005). Collagenase expression and activity is modulated by the interaction of collagen types, hypoxia, and nutrition in human lung cells. J Cell Physiol *204*, 146-154.

Lev, S., Blechman, J.M., Givol, D., and Yarden, Y. (1994). Steel factor and c-kit protooncogene: genetic lessons in signal transduction. Critical reviews in oncogenesis *5*, 141-168.

Li, N., and Karin, M. (2000). Signaling pathways leading to nuclear factor-kappa B activation. Methods in Enzymology *319*, 273-279.

Li, S., Ogawa, W., Emi, A., Hayashi, K., Senga, Y., Nomura, K., Hara, K., Yu, D., and Kasuga, M. (2011a). Role of S6K1 in regulation of SREBP1c expression in the liver. Biochemical and Biophysical Research Communications *412*, 197-202.

Li, X., Hu, X., Patel, B., Zhou, Z., Liang, S., Ybarra, R., Qiu, Y., Felsenfeld, G., Bungert, J., and Huang, S. (2010). H4R3 methylation facilitates beta-globin transcription by regulating histone acetyltransferase binding and H3 acetylation. Blood *115*, 2028-2037.

Li, X., Wang, S., Li, Y., Deng, C., Steiner, L.A., Xiao, H., Wu, C., Bungert, J., Gallagher, P.G., Felsenfeld, G*., et al.* (2011b). Chromatin boundaries require functional collaboration between the hSET1 and NURF complexes. Blood *118*, 1386-1394.

Liang, S.-H., and Clarke, M.F. (1999). A Bipartite Nuclear Localization Signal Is Required for p53 Nuclear Import Regulated by a Carboxyl-terminal Domain. J Biol Chem *274*, 32699-32703.

Lickwar, C.R., Mueller, F., Hanlon, S.E., McNally, J.G., and Lieb, J.D. (2012). Genome-wide protein-DNA binding dynamics suggest a molecular clutch for transcription factor function. Nature *484*, 251-255.

Light, W.H., Brickner, D.G., Brand, V.R., and Brickner, J.H. (2010). Interaction of a DNA zip code with the nuclear pore complex promotes H2A.Z incorporation and INO1 transcriptional memory. Mol Cell *40*, 112-125.

Lin, I.J., Zhou, Z., Crusselle-Davis, V.J., Moghimi, B., Gandhi, K., Anantharaman, A., Pantic, D., Huang, S., Jayandharan, G., Zhong, L.*, et al.* (2009). Calpeptin increases the activity of upstream stimulatory factor and induces high level globin gene expression in erythroid cells. J Biol Chem *284*, 20130-20135.

Lister, J.A., Robertson, C.P., Lepage, T., Johnson, S.L., and Raible, D.W. (1999). nacre encodes a zebrafish microphthalmia-related protein that regulates neural-crest-derived pigment cell fate. Development *126*, 3757-3767.

Liu, J.J., and Fisher, D.E. (2010). Lighting a path to pigmentation: mechanisms of MITF induction by UV. Pigment Cell Melanoma Res *23*, 741-745.

Liu, Y.V., Baek, J.H., Zhang, H., Diez, R., Cole, R.N., and Semenza, G.L. (2007). RACK1 competes with HSP90 for binding to HIF-1alpha and is required for O(2)-independent and HSP90 inhibitor-induced degradation of HIF-1alpha. Mol Cell *25*, 207-217.

Liu, Y.Z., Thomas, N.S., and Latchman, D.S. (1999). CBP associates with the p42/p44 MAPK enzymes and is phosphorylated following NGF treatment. Neuroreport *10*, 1239-1243.

Lowings, P., Yavuzer, U., and Goding, C.R. (1992). Positive and negative elements regulate a melanocyte-specific promoter. Mol Cell Biol *12*, 3653-3662.

Lum, J.J., Bui, T., Gruber, M., Gordan, J.D., DeBerardinis, R.J., Covello, K.L., Simon, M.C., and Thompson, C.B. (2007). The transcription factor HIF-1alpha plays a critical role in the growth factor-dependent regulation of both aerobic and anaerobic glycolysis. Genes Dev *21*, 1037-1049.

Lun, Y., Sawadogo, M., and Perry, M. (1997). Autoactivation of Xenopus MyoD transcription and its inhibition by USF. Cell Growth Differ *8*, 275-282.

Luo, W., Zhong, J., Chang, R., Hu, H., Pandey, A., and Semenza, G.L. (2010). Hsp70 and CHIP selectively mediate ubiquitination and degradation of hypoxia-inducible factor (HIF)-1alpha but Not HIF-2alpha. J Biol Chem *285*, 3651-3663.

Luo, X., and Sawadogo, M. (1996a). Antiproliferative properties of the USF family of helix-loop-helix transcription factors. Proc Natl Acad Sci USA *93*, 1308-1313.

Luo, X., and Sawadogo, M. (1996b). Functional domains of the transcription factor USF2: atypical nuclear localization signals and context-dependent transcriptional activation domains. Mol Cell Biol *16*, 1367-1375.

Lüscher, B. (2001). Function and regulation of the transcription factors of the Myc/Max/Mad network. Gene *277*, 1-14.

Luscher, B., and Eisenman, R.N. (1990). New light on Myc and Myb. Part I. Myc. Genes Dev *4*, 2025-2035.

Luscher, B., and Eisenman, R.N. (1992). Mitosis-specific phosphorylation of the nuclear oncoproteins Myc and Myb. Journal of Cell Biology *118*, 775-784.

Luscher, B., Kuenzel, E.A., Krebs, E.G., and Eisenman, R.N. (1989). Myc oncoproteins are phosphorylated by casein kinase II. EMBO J *8*, 1111-1119.

Lüscher, B., and Larsson, L.G. (1999). The basic region/helix-loop-helix/leucine zipper domain of Myc proto-oncoproteins: Function and regulation. Oncogene *18*, 2955-2966.

Ma, A., Moroy, T., Collum, R., Weintraub, H., Alt, F.W., and Blackwell, T.K. (1993). DNA binding by N- and L-Myc proteins. Oncogene *8*, 1093-1098.

Ma, J., Goryaynov, A., Sarma, A., and Yang, W. (2012). Self-regulated viscous channel in the nuclear pore complex. Proc Natl Acad Sci USA *109*, 7326-31.

Ma, P.C., Rould, M.A., Weintraub, H., and Pabo, C.O. (1994). Crystal structure of MyoD bHLH domain-DNA complex: perspectives on DNA recognition and implications for transcriptional activation. Cell *77*, 451-459.

Ma, Z., Jhun, B., Jung, S.Y., and Oh, C.K. (2008). Binding of upstream stimulatory factor 1 to the E-box regulates the 4G/5G polymorphism-dependent plasminogen activator inhibitor 1 expression in mast cells. The Journal of allergy and clinical immunology *121*, 1006-1012.e1002.

Mack, F.A., Patel, J.H., Biju, M.P., Haase, V.H., and Simon, M.C. (2005). Decreased growth of Vhl-/- fibrosarcomas is associated with elevated levels of cyclin kinase inhibitors p21 and p27. Mol Cell Biol *25*, 4565-4578.

Makela, T.P., Koskinen, P.J., Vastrik, I., and Alitalo, K. (1992). Alternative forms of Max as enhancers or suppressors of Myc-Ras cotransformation. Science *256*, 373-377.

Makino, Y., Cao, R., Svensson, K., Bertilsson, G., Asman, M., Tanaka, H., Cao, Y., Berkenstam, A., and Poellinger, L. (2001). Inhibitory PAS domain protein is a negative regulator of hypoxia-inducible gene expression. Nature *414*, 550-554.

Marcu, K.B., Bossone, S.A., and Patel, A.J. (1992). Myc Function and Regulation. Annual Review of Biochemistry *61*, 809-858.

Martinato, F., Cesaroni, M., Amati, B., and Guccione, E. (2008). Analysis of Myc-Induced Histone Modifications on Target Chromatin. PLoS ONE *3*, e3650.

Martinez-Campa, C., Politis, P., Moreau, J.L., Kent, N., Goodall, J., Mellor, J., and Goding, C.R. (2004). Precise nucleosome positioning and the TATA box dictate requirements for the histone H4 tail and the bromodomain factor Bdf1. Mol Cell *15*, 69-81.

Massari, M., and Murre, C. (2000). Helix-loop-helix proteins: regulators of transcription in eucaryotic organisms. Mol Cell Biol *20*, 429 - 440.

Maurer, J., Fuchs, S., Jäger, R., Kurz, B., Sommer, L., and Schorle, H. (2007). Establishment and controlled differentiation of neural crest stem cell lines using conditional transgenesis. Differentiation *75*, 580-591.

Mayer, T.C. (1973). The migratory pathway of neural crest cells into the skin of mouse embryos. Dev Biol *34*, 39-46.

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

McAndrew, P.C., Svaren, J., Martin, S.R., Hörz, W., and Goding, C.R. (1998). Requirements for Chromatin Modulation and Transcription Activation by the Pho4 Acidic Activation Domain. Mol Cell Biol *18*, 5818-5827.

McGill, G.G., Horstmann, M., Widlund, H.R., Du, J., Motyckova, G., Nishimura, E.K., Lin, Y.L., Ramaswamy, S., Avery, W., Ding, H.F.*, et al.* (2002). Bcl2 regulation by the melanocyte master regulator Mitf modulates lineage survival and melanoma cell viability. Cell *109*, 707-718.

McMahon, S.B., Wood, M.A., and Cole, M.D. (2000). The Essential Cofactor TRRAP Recruits the Histone Acetyltransferase hGCN5 to c-Myc. Mol Cell Biol *20*, 556-562.

Menon, B.B., Sarma, N.J., Pasula, S., Deminoff, S.J., Willis, K.A., Barbara, K.E., Andrews, B., and Santangelo, G.M. (2005). Reverse recruitment: The Nup84 nuclear pore subcomplex mediates Rap1/Gcr1/Gcr2 transcriptional activation. Proc Natl Acad Sci USA *102*, 5749-5754.

Mercer, E.M., Lin, Y.C., Benner, C., Jhunjhunwala, S., Dutkowski, J., Flores, M., Sigvardsson, M., Ideker, T., Glass, C.K., and Murre, C. (2011). Multilineage priming of enhancer repertoires precedes commitment to the B and myeloid cell lineages in hematopoietic progenitors. Immunity *35*, 413-425.

Meyer, N., and Penn, L.Z. (2008). Reflecting on 25 years with MYC. Nat Rev Cancer *8*, 976-990.

Miller, J.L., and Grant, P.A. (2013). The role of DNA methylation and histone modifications in transcriptional regulation in humans. Sub-cellular Biochemistry *61*, 289-317.

Minden, A., and Karin, M. (1997). Regulation and function of the JNK subgroup of MAP kinases. Biochimica et Biophysica Acta *1333*, 85-104.

Mito, Y., Henikoff, J.G., and Henikoff, S. (2007). Histone replacement marks the boundaries of cis-regulatory domains. Science *315*, 1408-1411.

Moore, K.J. (1995). Insight into the microphthalmia gene. Trends in Genetics *11*, 442-448.

Moreau, J.L., Lee, M., Mahachi, N., Vary, J., Mellor, J., Tsukiyama, T., and Goding, C.R. (2003). Regulated displacement of TBP from the PHO8 promoter in vivo requires Cbf1 and the Isw1 chromatin remodeling complex. Mol Cell *11*, 1609-1620.

Muhle-Goll, C., Gibson, T., Schuck, P., Schubert, D., Nalis, D., Nilges, M., and Pastore, A. (1994). The dimerization stability of the HLH-LZ transcription protein family is modulated by the leucine zippers: a CD and NMR study of TFEB and c-Myc. Biochemistry *33*, 11296-11306.

Muhle-Goll, C., Nilges, M., and Pastore, A. (1995). The leucine zippers of the HLH-LZ proteins Max and c-Myc preferentially form heterodimers. Biochemistry *34*, 13554-13564.

Mukherjee, B., Morgenbesser, S.D., and DePinho, R.A. (1992). Myc family oncoproteins function through a common pathway to transform normal cells in culture: cross-interference by Max and trans-acting dominant mutants. Genes Dev *6*, 1480-1492.

Murre, C., Bain, G., Vandijk, M., Engel, I., Furnari, B., Massari, M., Matthews, J., Quong, M., Rivera, R., and Stuiver, M. (1994). Structure and function of helix-loop-helix proteins. Biochimica et Biophysica Acta *1218*, 129 - 135.

Murre, C., McCaw, P.S., and Baltimore, D. (1989). A new DNA binding and dimerization motif in immunoglobulin enhancer binding, daughterless, MyoD, and myc proteins. Cell *56*, 777-783.

Nair, S.K., and Burley, S.K. (2003). X-Ray Structures of Myc-Max and Mad-Max Recognizing DNA: Molecular Bases of Regulation by Proto-Oncogenic Transcription Factors. Cell *112*, 193-205.

Nakajima, T., Uchida, C., Anderson, S.F., Parvin, J.D., and Montminy, M. (1997). Analysis of a cAMP-responsive activator reveals a two-component mechanism for transcriptional induction via signal-dependent factors. Genes Dev *11*, 738-747.

Nakamura, M., Tobin, D.J., Richards-Smith, B., Sundberg, J.P., and Paus, R. (2002). Mutant laboratory mice with abnormalities in pigmentation: annotated tables. J Dermatol Sci *28*, 1-33.

Nakayama, A., Nguyen, M.T., Chen, C.C., Opdecamp, K., Hodgkinson, C.A., and Arnheiter, H. (1998). Mutations in microphthalmia, the mouse homolog of the human deafness gene MITF, affect neuroepithelial and neural crest-derived melanocytes differently. Mech Dev *70*, 155-166.

Naukkarinen, J., Ehnholm, C., and Peltonen, L. (2006). Genetics of familial combined hyperlipidemia. Current Opinion in Lipidology *17*, 285-290.

Naukkarinen, J., Gentile, M., Soro-Paavonen, A., Saarela, J., Koistinen, H.A., Pajukanta, P., Taskinen, M.R., and Peltonen, L. (2005). USF1 and dyslipidemias: converging evidence for a functional intronic variant. Human Molecular Genetics *14*, 2595-2605.

Nelson, J.D., Denisenko, O., and Bomsztyk, K. (2006). Protocol for the fast chromatin immunoprecipitation (ChIP) method. Nat Protocols *1*, 179-185.

Ng, M.C., Miyake, K., So, W.Y., Poon, E.W., Lam, V.K., Li, J.K., Cox, N.J., Bell, G.I., and Chan, J.C. (2005). The linkage and association of the gene encoding upstream stimulatory factor 1 with type 2 diabetes and metabolic syndrome in the Chinese population. Diabetologia *48*, 2018-2024.

Nie, Z., Hu, G., Wei, G., Cui, K., Yamane, A., Resch, W., Wang, R., Green, D.R., Tessarollo, L., Casellas, R.*, et al.* (2012). c-Myc is a universal amplifier of expressed genes in lymphocytes and embryonic stem cells. Cell *151*, 68-79.

Nilsson, J.A., and Cleveland, J.L. (2003). Myc pathways provoking cell suicide and cancer. Oncogene *22*, 9007-9021.

Nishikawa, S., Kusakabe, M., Yoshinaga, K., Ogawa, M., Hayashi, S., Kunisada, T., Era, T., Sakakura, T., and Nishikawa, S. (1991). In utero manipulation of coat color formation by a monoclonal anti-c-kit antibody: two distinct waves of c-kit-dependency during melanocyte development. EMBO J *10*, 2111-2118.

Nowak, M., Helleboid-Chapman, A., Jakel, H., Martin, G., Duran-Sandoval, D., Staels, B., Rubin, E.M., Pennacchio, L.A., Taskinen, M.R., Fruchart-Najib, J.*, et al.* (2005). Insulin-mediated down-regulation of apolipoprotein A5 gene expression through the phosphatidylinositol 3-kinase pathway: role of upstream stimulatory factor. Mol Cell Biol *25*, 1537-1548.

O'Hagan, R.C., Schreiber-Agus, N., Chen, K., David, G., Engelman, J.A., Schwab, R., Alland, L., Thomson, C., Ronning, D.R., Sacchettini, J.C.*, et al.* (2000). Gene-target recognition among members of the Myc superfamily and implications for oncogenesis. Nat Genet *24*, 113-119.

O'Neill, E.M., Kaffman, A., Jolly, E.R., and O'Shea, E.K. (1996). Regulation of PHO4 nuclear localization by the PHO80-PHO85 cyclin-CDK complex. Science *271*, 209-212.

O'Rourke, J.F., Tian, Y.M., Ratcliffe, P.J., and Pugh, C.W. (1999). Oxygen-regulated and transactivating domains in endothelial PAS protein 1: comparison with hypoxia-inducible factor-1alpha. J Biol Chem *274*, 2060-2071.

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.

Olave, N.C., Grenett, M.H., Cadeiras, M., Grenett, H.E., and Higgins, P.J. (2010). Upstream stimulatory factor-2 mediates quercetin-induced suppression of PAI-1 gene expression in human endothelial cells. Journal of Cellular Biochemistry *111*, 720-726.

Opdecamp, K., Nakayama, A., Nguyen, M.T., Hodgkinson, C.A., Pavan, W.J., and Arnheiter, H. (1997). Melanocyte development in vivo and in neural crest cell cultures: crucial dependence on the Mitf basic-helix-loop-helix-zipper transcription factor. Development *124*, 2377-2386.

Osborne, T.F. (2000). Sterol regulatory element-binding proteins (SREBPs): key regulators of nutritional homeostasis and insulin action. J Biol Chem *275*, 32379-32382.

Ostuni, R., Piccolo, V., Barozzi, I., Polletti, S., Termanini, A., Bonifacio, S., Curina, A., Prosperini, E., Ghisletti, S., and Natoli, G. (2013). Latent enhancers activated by stimulation in differentiated cells. Cell *152*, 157-171.

Pabo, C.O., and Sauer, R.T. (1992). Transcription Factors: Structural Families and Principles of DNA Recognition. Annual Review of Biochemistry *61*, 1053-1095.

Packham, G., and Cleveland, J.L. (1995). c-Myc and apoptosis. Biochimica et Biophysica Acta *1242*, 11-28.

Pajukanta, P., Lilja, H.E., Sinsheimer, J.S., Cantor, R.M., Lusis, A.J., Gentile, M., Duan, X.J., Soro-Paavonen, A., Naukkarinen, J., Saarela, J.*, et al.* (2004a). Familial combined hyperlipidemia is associated with upstream transcription factor 1 (USF1). Nat Genet *36*, 371-376.

Pajukanta, P., Lilja, H.E., Sinsheimer, J.S., Cantor, R.M., Lusis, A.J., Gentile, M., Duan, X.J., Soro-Paavonen, A., Naukkarinen, J., Saarela, J.*, et al.* (2004b). Familial combined hyperlipidemia is associated with upstream transcription factor 1 (USF1). Nat Genet *36*, 371-376.

Pajukanta, P., Nuotio, I., Terwilliger, J.D., Porkka, K.V., Ylitalo, K., Pihlajamaki, J., Suomalainen, A.J., Syvanen, A.C., Lehtimaki, T., Viikari, J.S.*, et al.* (1998). Linkage of familial combined hyperlipidaemia to chromosome 1q21-q23. Nat Genet *18*, 369-373.

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 Metabolism *3*, 187-197.

Pastier, D., Lacorte, J.M., Chambaz, J., Cardot, P., and Ribeiro, A. (2002). Two initiator-like elements are required for the combined activation of the human apolipoprotein C-III promoter by upstream stimulatory factor and hepatic nuclear factor-4. J Biol Chem *277*, 15199-15206.

Payton, S.G., Liu, M., Ge, Y., and Matherly, L.H. (2005). Transcriptional regulation of the human reduced folate carrier A1/A2 promoter: Identification of critical roles for the USF and GATA families of transcription factors. Biochimica et Biophysica Acta *1731*, 115-124.

Peng, J., Zhang, L., Drysdale, L., and Fong, G.H. (2000). The transcription factor EPAS-1/hypoxia-inducible factor 2α plays an important role in vascular remodeling. Proc Natl Acad Sci USA *97*, 8386-8391.

Perini, G., Diolaiti, D., Porro, A., and Della Valle, G. (2005). In vivo transcriptional regulation of N-Myc target genes is controlled by E-box methylation. Proc Natl Acad Sci USA *102*, 12117-12122.

Peukert, K., Staller, P., Schneider, A., Carmichael, G., Hanel, F., and Eilers, M. (1997). An alternative pathway for gene regulation by Myc. EMBO J *16*, 5672-5686.

Piedra, J., Martínez, D., Castaño, J., Miravet, S., Duñach, M., and de Herreros, A.G.a. (2001). Regulation of β-Catenin Structure and Activity by Tyrosine Phosphorylation. J Biol Chem *276*, 20436-20443.

Planque, N., Turque, N., Opdecamp, K., Bailly, M., Martin, P., and Saule, S. (1999). Expression of the microphthalmia-associated basic helix-loop-helix leucine zipper transcription factor Mi in avian neuroretina cells induces a pigmented phenotype. Cell Growth Differ *10*, 525-536.

Pogenberg, V., Ögmundsdóttir, M.H., Bergsteinsdóttir, K., Schepsky, A., Phung, B., Deineko, V., Milewski, M., Steingrímsson, E., and Wilmanns, M. (2012). Restricted leucine zipper dimerization and specificity of DNA recognition of the melanocyte master regulator MITF. Genes Dev *26*, 2647-2658.

Pognonec, P., Kato, H., and Roeder, R.G. (1992). The helix-loop-helix/leucine repeat transcription factor USF can be functionally regulated in a redox-dependent manner. J Biol Chem *267*, 24563-24567.

Prendergast, G.C. (1999). Mechanisms of apoptosis by c-Myc. Oncogene *18*, 2967-2987.

Prendergast, G.C., Hopewell, R., Gorham, B.J., and Ziff, E.B. (1992). Biphasic effect of Max on Myc cotransformation activity and dependence on amino- and carboxy-terminal Max functions. Genes Dev *6*, 2429-2439.

Prendergast, G.C., Lawe, D., and Ziff, E.B. (1991). Association of Myn, the murine homolog of max, with c-Myc stimulates methylation-sensitive DNA binding and ras cotransformation. Cell *65*, 395-407.

Prendergast, G.C., and Ziff, E.B. (1991). Methylation-sensitive sequence-specific DNA binding by the c-Myc basic region. Science *251*, 186-189.

Price, E.R., Ding, H.F., Badalian, T., Bhattacharya, S., Takemoto, C., Yao, T.P., Hemesath, T.J., and Fisher, D.E. (1998a). Lineage-specific signaling in melanocytes. C-kit stimulation recruits p300/CBP to microphthalmia. J Biol Chem *273*, 17983-17986.

Price, E.R., Horstmann, M.A., Wells, A.G., Weilbaecher, K.N., Takemoto, C.M., Landis, M.W., and Fisher, D.E. (1998b). alpha-Melanocyte-stimulating hormone signaling regulates expression of microphthalmia, a gene deficient in Waardenburg syndrome. J Biol Chem *273*, 33042-33047.

Prives, C., and Hall, P.A. (1999). The p53 pathway. The Journal of pathology *187*, 112-126.

Ptashne, M., and Gann, A. (1997). Transcriptional activation by recruitment. Nature *386*, 569-577.

Pugh, C.W., and Ratcliffe, P.J. (2003). Regulation of angiogenesis by hypoxia: Role of the HIF system. Nature Medicine *9*, 677-684.

Pulverer, B.J., Fisher, C., Vousden, K., Littlewood, T., Evan, G., and Woodgett, J.R. (1994). Site-specific modulation of c-Myc cotransformation by residues phosphorylated in vivo. Oncogene *9*, 58-70.

Putowski, L.T., Schillings, W.J., Lee, C.M., Reddy, E.P., and Jakowicki, J.A. (2004). Human follicle-stimulating hormone receptor (FSH-R) promoter/enhancer activity is inhibited by transcriptional factors, from the upstream stimulating factors family, via E-box and newly identified initiator element (Inr) in FSH-R non-expressing cells. Gynecological Endocrinology *19*, 9-17.

Putt, W., Palmen, J., Nicaud, V., Tregouet, D.A., Tahri-Daizadeh, N., Flavell, D.M., Humphries, S.E., and Talmud, P.J. (2004). Variation in USF1 shows haplotype effects, gene : gene and gene : environment associations with glucose and lipid parameters in the European Atherosclerosis Research Study II. Human Molecular Genetics *13*, 1587-1597.

Qiao, H., and May, J.M. (2011). CpG methylation at the USF-binding site mediates cell-specific transcription of human ascorbate transporter SVCT2 exon 1a. Biochemical Journal *440*, 73-84.

Quéva, C., Hurlin, P.J., Foley, K.P., and Eisenman, R.N. (1998). Sequential expression of the MAD family of transcriptional repressors during differentiation and development. Oncogene *16*, 967-977.

Qyang, Y., Luo, X., Lu, T., Ismail, P.M., Krylov, D., Vinson, C., and Sawadogo, M. (1999). Cell-type-dependent activity of the ubiquitous transcription factor USF in cellular proliferation and transcriptional activation. Mol Cell Biol *19*, 1508-1517.

Rada-Iglesias, A., Ameur, A., Kapranov, P., Enroth, S., Komorowski, J., Gingeras, T.R., and Wadelius, C. (2008). Whole-genome maps of USF1 and USF2 binding and histone H3 acetylation reveal new aspects of promoter structure and candidate genes for common human disorders. Genome Res *18*, 380-392.

Rada-Iglesias, A., Bajpai, R., Swigut, T., Brugmann, S.A., Flynn, R.A., and Wysocka, J. (2011). A unique chromatin signature uncovers early developmental enhancers in humans. Nature *470*, 279-283.

Rada-Iglesias, A., Wallerman, O., Koch, C., Ameur, A., Enroth, S., Clelland, G., Wester, K., Wilcox, S., Dovey, O.M., Ellis, P.D.*, et al.* (2005). Binding sites for metabolic disease related transcription factors inferred at base pair resolution by chromatin immunoprecipitation and genomic microarrays. Human Molecular Genetics *14*, 3435-3447.

Rahl, P.B., and Young, R.A. (2014). MYC and Transcription Elongation. Cold Spring Harbor Perspectives in Medicine *4*.

Raval, R.R., Lau, K.W., Tran, M.G.B., Sowter, H.M., Mandriota, S.J., Li, J.L., Pugh, C.W., Maxwell, P.H., Harris, A.L., and Ratcliffe, P.J. (2005). Contrasting properties of hypoxia-inducible factor 1 (HIF-1) and HIF-2 in von Hippel-Lindau-associated renal cell carcinoma. Mol Cell Biol *25*, 5675-5686.

Reinke, H., and Hörz, W. (2003). Histones are first hyperacetylated and then lose contact with the activated PHO5 promoter. Mol Cell *11*, 1599-1607.

Reisman, D., Elkind, N.B., Roy, B., Beamon, J., and Rotter, V. (1993). c-Myc trans-activates the p53 promoter through a required downstream CACGTG motif. Cell Growth & Differentiation *4*, 57-65.

Reisman, D., and Rotter, V. (1993). The helix-loop-helix containing transcription factor USF binds to and transactivates the promoter of the p53 tumor suppressor gene. Nucleic acids research *21*, 345-350.

Ribeiro, A., Pastier, D., Kardassis, D., Chambaz, J., and Cardot, P. (1999). Cooperative binding of upstream stimulatory factor and hepatic nuclear factor 4 drives the transcription of the human apolipoprotein A-II gene. J Biol Chem *274*, 1216-1225.

Rippe, K., Schrader, A., Riede, P., Strohner, R., Lehmann, E., and Längst, G. (2007). DNA sequence- and conformation-directed positioning of nucleosomes by chromatin-remodeling complexes. Proc Natl Acad Sci USA *104*, 15635-15640.

Robinson, M.D., McCarthy, D.J., and Smyth, G.K. (2010). edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. Bioinformatics *26*, 139-140.

Rodriguez, M., Aladowicz, E., Lanfrancone, L., and Goding, C.R. (2008). Tbx3 Represses E-Cadherin Expression and Enhances Melanoma Invasiveness. Cancer Res *68*, 7872-7881.

Roth, G., Kotzka, J., Kremer, L., Lehr, S., Lohaus, C., Meyer, H.E., Krone, W., and Muller-Wieland, D. (2000). MAP kinases Erk1/2 phosphorylate sterol regulatory element-binding protein (SREBP)-1a at serine 117 in vitro. J Biol Chem *275*, 33302-33307.

Roussel, M., Saule, S., Lagrou, C., Rommens, C., Beug, H., Graf, T., and Stehelin, D. (1979). Three new types of viral oncogene of cellular origin specific for haematopoietic cell transformation. Nature *281*, 452-455.

Rouzaud, F., Kadekaro, A.L., Abdel-Malek, Z.A., and Hearing, V.J. (2005). MC1R and the response of melanocytes to ultraviolet radiation. Mutation Research/Fundamental and Molecular Mechanisms of Mutagenesis *571*, 133-152.

Roy, A.L., Carruthers, C., Gutjahr, T., and Roeder, R.G. (1993). Direct role for Myc in transcription initiation mediated by interactions with TFII-I. Nature *365*, 359-361.

Ruas, J.L., Poellinger, L., and Pereira, T. (2005). Role of CBP in regulating HIF-1-mediated activation of transcription. Journal of Cell Science *118*, 301-311.

Rubinfeld, B., Robbins, P., El-Gamil, M., Albert, I., Porfiri, E., and Polakis, P. (1997). Stabilization of beta-catenin by genetic defects in melanoma cell lines. Science *275*, 1790-1792.

Rustgi, A.K., Dyson, N., and Bernards, R. (1991). Amino-terminal domains of c-myc and N-myc proteins mediate binding to the retinoblastoma gene product. Nature *352*, 541-544.

Ryan, H.E., Lo, J., and Johnson, R.S. (1998). HIF-1α is required for solid tumor formation and embryonic vascularization. EMBO J *17*, 3005-3015.

Sabò, A., and Amati, B. (2014). Genome Recognition by MYC. Cold Spring Harbor Perspectives in Medicine *4*.

Saez-Ayala, M., Montenegro, M.F., Sanchez-Del-Campo, L., Fernandez-Perez, M.P., Chazarra, S., Freter, R., Middleton, M., Pinero-Madrona, A., Cabezas-Herrera, J., Goding, C.R.*, et al.* (2013). Directed phenotype switching as an effective antimelanoma strategy. Cancer Cell *24*, 105-119.

Saha, B., Singh, S.K., Sarkar, C., Bera, R., Ratha, J., Tobin, D.J., and Bhadra, R. (2006). Activation of the Mitf promoter by lipid-stimulated activation of p38-stress signalling to CREB. Pigment Cell Res *19*, 595-605.

Salti, G.I., Manougian, T., Farolan, M., Shilkaitis, A., Majumdar, D., and Das Gupta, T.K. (2000). Micropthalmia transcription factor: a new prognostic marker in intermediate-thickness cutaneous malignant melanoma. Cancer Res *60*, 5012-5016.

Samoylenko, A., Roth, U., Jungermann, K., and Kietzmann, T. (2001). The upstream stimulatory factor-2a inhibits plasminogen activator inhibitor-1 gene expression by binding to a promoter element adjacent to the hypoxia-inducible factor-1 binding site. Blood *97*, 2657-2666.

Sato, S., Roberts, K., Gambino, G., Cook, A., Kouzarides, T., and Goding, C.R. (1997). CBP/p300 as a co-factor for the Microphthalmia transcription factor. Oncogene *14*, 3083-3092.

Sawadogo, M. (1988). Multiple forms of the human gene-specific transcription factor USF. II. DNA binding properties and transcriptional activity of the purified HeLa USF. J Biol Chem *263*, 11994-12001.

Sawadogo, M., and Roeder, R.G. (1985). Interaction of a gene-specific transcription factor with the adenovirus major late promoter upstream of the TATA box region. Cell *43*, 165-175.

Sawadogo, M., Van Dyke, M.W., Gregor, P.D., and Roeder, R.G. (1988). Multiple forms of the human gene-specific transcription factor USF. I. Complete purification and identification of USF from HeLa cell nuclei. J Biol Chem *263*, 11985-11993.

Sayasith, K., Bouchard, N., Sawadogo, M., Lussier, J.G., and Sirois, J. (2004). Molecular characterization and role of bovine upstream stimulatory factor 1 and 2 in the regulation of the prostaglandin G/H synthase-2 promoter in granulosa cells. J Biol Chem *279*, 6327-6336.

Schmittgen, T.D., and Livak, K.J. (2008). Analyzing real-time PCR data by the comparative CT method. Nature Protocols *3*, 1101-1108.

Schulze, S.R., and Wallrath, L.L. (2007). Gene Regulation by Chromatin Structure: Paradigms Established in Drosophila melanogaster. Annual Review of Entomology *52*, 171-192.

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

Selimyan, R., Gerstein, R.M., Ivanova, I., Precht, P., Subrahmanyam, R., Perlot, T., Alt, F.W., and Sen, R. (2013). Localized DNA demethylation at recombination intermediates during immunoglobulin heavy chain gene assembly. PLoS biology *11*, e1001475.

Semenza, G.L. (2001). HIF-1, O(2), and the 3 PHDs: how animal cells signal hypoxia to the nucleus. Cell *107*, 1-3.

Semenza, G.L. (2003). Targeting HIF-1 for cancer therapy. Nat Rev Cancer *3*, 721-732.

Semenza, G.L., Jiang, B.H., Leung, S.W., Passantino, R., Concordet, J.P., 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.

Shao, W., and Espenshade, P.J. (2012). Expanding roles for SREBP in metabolism. Cell Metabolism *16*, 414-419.

Sheiness, D., and Bishop, J.M. (1979). DNA and RNA from Uninfected Vertebrate Cells Contain Nucleotide Sequences Related to the Putative Transforming Gene of Avian Myelocytomatosis Virus. Journal of Virology *31*, 514-521.

Sheng, Z., Otani, H., Brown, M.S., and Goldstein, J.L. (1995). Independent regulation of sterol regulatory element-binding proteins 1 and 2 in hamster liver. Proc Natl Acad Sci USA *92*, 935-938.

Sherr, C.J. (2001). The INK4a/ARF network in tumour suppression. Nature Reviews Molecular Cell Biology *2*, 731-737.

Shi, L., Liu, S., Nikolic, D., and Wang, S. (2008). High glucose levels upregulate upstream stimulatory factor 2 gene transcription in mesangial cells. Journal of cellular biochemistry *103*, 1952-1961.

Shibahara, S., Takeda, K., Yasumoto, K., Udono, T., Watanabe, K., Saito, H., and Takahashi, K. (2001). Microphthalmia-associated transcription factor (MITF): multiplicity in structure, function, and regulation. Journal of Investigative Dermatology Symposium Proceedings *6*, 99-104.

Shibata, K., Muraosa, Y., Tomita, Y., Tagami, H., and Shibahara, S. (1992). Identification of a cis-acting element that enhances the pigment cell-specific expression of the human tyrosinase gene. J Biol Chem *267*, 20584-20588.

Shimano, H. (2007). SREBP-1c and TFE3, energy transcription factors that regulate hepatic insulin signaling. Journal of Molecular Medicine (Berlin, Germany) *85*, 437-444.

Shimizu, T., Toumoto, A., Ihara, K., Shimizu, M., Kyogoku, Y., Ogawa, N., Oshima, Y., and Hakoshima, T. (1997). Crystal structure of PHO4 bHLH domain–DNA complex: flanking base recognition. EMBO J *16*, 4689-4697.

Shimomura, I., Shimano, H., Korn, B.S., Bashmakov, Y., and Horton, J.D. (1998). Nuclear sterol regulatory element-binding proteins activate genes responsible for the entire program of unsaturated fatty acid biosynthesis in transgenic mouse liver. J Biol Chem *273*, 35299-35306.

Shin, H., Liu, T., Manrai, A.K., and Liu, X.S. (2009). CEAS: cis-regulatory element annotation system. Bioinformatics *25*, 2605-2606.

Shiota, K. (2004). DNA methylation profiles of CpG islands for cellular differentiation and development in mammals. Cytogenetic and Genome Research *105*, 325-334.

Shoulders, C.C., and Naoumova, R.P. (2004). USF1 implicated in the aetiology of familial combined hyperlipidaemia and the metabolic syndrome. Trends Mol Med *10*, 362-365.

Shrivastava, A., Yu, J., Artandi, S., and Calame, K. (1996). YY1 and c-Myc associate in vivo in a manner that depends on c-Myc levels. Proc Natl Acad Sci USA *93*, 10638-10641.

Sirito, M., Lin, Q., Deng, J.M., Behringer, R.R., and Sawadogo, M. (1998). Overlapping roles and asymmetrical cross-regulation of the USF proteins in mice. Proc Natl Acad Sci USA *95*, 3758-3763.

Sirito, M., Lin, Q., Maity, T., and Sawadogo, M. (1994). Ubiquitous expression of the 43- and 44-kDa forms of transcription factor USF in mammalian cells. Nucleic Acids Research *22*, 427-433.

Sirito, M., Walker, S., Lin, Q., Kozlowski, M.T., Klein, W.H., and Sawadogo, M. (1992). Members of the USF family of helix-loop-helix proteins bind DNA as homo- as well as heterodimers. Gene Expression *2*, 231-240.

Sloan, E.J., and Ayer, D.E. (2010). Myc, Mondo, and Metabolism. Genes & Cancer *1*, 587-596.

Smalley, K., and Eisen, T. (2000). The involvement of p38 mitogen-activated protein kinase in the alpha-melanocyte stimulating hormone (alpha-MSH)-induced melanogenic and anti-proliferative effects in B16 murine melanoma cells. FEBS Lett *476*, 198-202.

Solomon, D.L.C., Amati, N., and Land, H. (1993). Distinct DNA binding preferences for the c-Myc/Max and Max/Max dimers. Nucleic Acids Research *21*, 5372-5376.

Sommer, A., Bousset, K., Kremmer, E., Austen, M., and Lüscher, B. (1998). Identification and Characterization of Specific DNA-binding Complexes Containing Members of the Myc/Max/Mad Network of Transcriptional Regulators. J Biol Chem *273*, 6632-6642.

Souza, R.F., Wang, S., Thakar, M., Smolinski, K.N., Yin, J., Zou, T.T., Kong, D., Abraham, J.M., Toretsky, J.A., and Meltzer, S.J. (1999). Expression of the wild-type insulin-like growth factor II receptor gene suppresses growth and causes death in colorectal carcinoma cells. Oncogene *18*, 4063-4068.

Spencer, C.A., and Groudine, M. (1991). Control of c-myc regulation in normal and neoplastic cells. Adv Cancer Res *56*, 1-48.

Staller, P., Sulitkova, J., Lisztwan, J., Moch, H., Oakeley, E.J., and Krek, W. (2003). Chemokine receptor CXCR4 downregulated by von Hippel-Lindau tumour suppressor pVHL. Nature *425*, 307-311.

Stanton, B.R., Perkins, A.S., Tessarollo, L., Sassoon, D.A., and Parada, L.F. (1992). Loss of N-myc function results in embryonic lethality and failure of the epithelial component of the embryo to develop. Genes Dev *6*, 2235-2247.

Steel, K.P., and Barkway, C. (1989). Another role for melanocytes: their importance for normal stria vascularis development in the mammalian inner ear. Development *107*, 453-463.

Steidl, C., Leimeister, C., Klamt, B., Maier, M., Nanda, I., Dixon, M., Clarke, R., Schmid, M., and Gessler, M. (2000). Characterization of the human and mouse HEY1, HEY2, and HEYL genes: cloning, mapping, and mutation screening of a new bHLH gene family. Genomics *66*, 195-203.

Strub, T., Giuliano, S., Ye, T., Bonet, C., Keime, C., Kobi, D., Le Gras, S., Cormont, M., Ballotti, R., Bertolotto, C*., et al.* (2011). Essential role of microphthalmia transcription factor for DNA replication, mitosis and genomic stability in melanoma. Oncogene *30*, 2319-2332.

Sul, H.S., Latasa, M.J., Moon, Y., and Kim, K.H. (2000). Regulation of the fatty acid synthase promoter by insulin. The Journal of Nutrition *130*, 315s-320s.

Sundqvist, A., Bengoechea-Alonso, M.T., Ye, X., Lukiyanchuk, V., Jin, J., Harper, J.W., and Ericsson, J. (2005). Control of lipid metabolism by phosphorylation-dependent degradation of the SREBP family of transcription factors by SCF(Fbw7). Cell Metabolism *1*, 379-391.

Szentirmay, M.N., Yang, H.X., Pawar, S.A., Vinson, C., and Sawadogo, M. (2003). The IGF2 receptor is a USF2-specific target in nontumorigenic mammary epithelial cells but not in breast cancer cells. J Biol Chem *278*, 37231-37240.

Tachibana, M. (1997). Evidence to suggest that expression of MITF induces melanocyte differentiation and haploinsufficiency of MITF causes Waardenburg syndrome type 2A. Pigment Cell Res *10*, 25-33.

Tachibana, M. (2000). MITF: a stream flowing for pigment cells. Pigment Cell Res *13*, 230-240.

Tachibana, M., Takeda, K., Nobukuni, Y., Urabe, K., Long, J.E., Meyers, K.A., Aaronson, S.A., and Miki, T. (1996). Ectopic expression of MITF, a gene for Waardenburg syndrome type 2, converts fibroblasts to cells with melanocyte characteristics. Nat Genet *14*, 50-54.

Takahashi, Y., and Kamataki, T. (2001). Competitive inhibition of the transcription of rabbit CYP1A1 gene by upstream stimulatory factor 1 (USF1). Drug Metabolism Reviews *33*, 37-47.

Takahashi, Y., Nakayama, K., Itoh, S., Fujii-Kuriyama, Y., and Kamataki, T. (1997). Inhibition of the transcription of CYP1A1 gene by the upstream stimulatory factor 1 in rabbits. Competitive binding of USF1 with AhR.Arnt complex. J Biol Chem *272*, 30025-30031.

Talotta, F., Cimmino, A., Matarazzo, M.R., Casalino, L., De Vita, G., D'Esposito, M., Di Lauro, R., and Verde, P. (2008). An autoregulatory loop mediated by miR-21 and PDCD4 controls the AP-1 activity in RAS transformation. Oncogene *28*, 73-84.

Tan-Wong, S.M., Wijayatilake, H.D., and Proudfoot, N.J. (2009). Gene loops function to maintain transcriptional memory through interaction with the nuclear pore complex. Genes Dev *23*, 2610-2624.

Tan, Y., Rouse, J., Zhang, A., Cariati, S., Cohen, P., and Comb, M.J. (1996). FGF and stress regulate CREB and ATF-1 via a pathway involving p38 MAP kinase and MAPKAP kinase-2. EMBO J *15*, 4629-4642.

Tarutani, M., Cai, T., Dajee, M., and Khavari, P.A. (2003). Inducible Activation of Ras and Raf in Adult Epidermis. Cancer Res *63*, 319-323.

Tasheva, E.S., and Conrad, G.W. (2003). The UV responsive elements in the human mimecan promoter: a functional characterization. Molecular Vision *9*, 1-9.

Thomas, D., Jacquemin, I., and Surdin-Kerjan, Y. (1992). MET4, a leucine zipper protein, and centromere-binding factor 1 are both required for transcriptional activation of sulfur metabolism in Saccharomyces cerevisiae. Mol Cell Biol *12*, 1719-1727.

Thompson, E.B. (1998). The many roles of c-Myc in apoptosis. Annual review of Physiology *60*, 575-600.

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, 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.

To, K.K., Sedelnikova, O.A., Samons, M., Bonner, W.M., and Huang, L.E. (2006). The phosphorylation status of PAS-B distinguishes HIF-1alpha from HIF-2alpha in NBS1 repression. EMBO J *25*, 4784-4794.

Tsujimura, T., Hashimoto, K., Morii, E., Tunio, G.M., Tsujino, K., Kondo, T., Kanakura, Y., and Kitamura, Y. (1997). Involvement of transcription factor encoded by the mouse mi locus (MITF) in apoptosis of cultured mast cells induced by removal of interleukin-3. The American Journal of Pathology *151*, 1043-1051.

Vallet, V.S., Casado, M., Henrion, A.A., Bucchini, D., Raymondjean, M., Kahn, A., and Vaulont, S. (1998). Differential roles of upstream stimulatory factors 1 and 2 in the transcriptional response of liver genes to glucose. J Biol Chem *273*, 20175-20179.

Vallet, V.S., Henrion, A.A., Bucchini, D., Casado, M., Raymondjean, M., Kahn, A., and Vaulont, S. (1997). Glucose-dependent Liver Gene Expression in Upstream Stimulatory Factor 2 −/− Mice. J Biol Chem *272*, 21944-21949.

van Deursen, D., Jansen, H., and Verhoeven, A.J. (2008). Glucose increases hepatic lipase expression in HepG2 liver cells through upregulation of upstream stimulatory factors 1 and 2. Diabetologia *51*, 2078-2087.

van Deursen, D., van Leeuwen, M., Vaulont, S., Jansen, H., and Verhoeven, A.J. (2009). Upstream Stimulatory Factors 1 and 2 activate the human hepatic lipase promoter via E-box dependent and independent mechanisms. Biochimica et biophysica acta *1791*, 229-237.

Varlakhanova, N.V., and Knoepfler, P.S. (2009). Acting locally and globally: Myc's ever-expanding roles on chromatin. Cancer Res *69*, 7487-7490.

Vaulont, S., Puzenat, N., Levrat, F., Cognet, M., Kahn, A., and Raymondjean, M. (1989). Proteins binding to the liver-specific pyruvate kinase gene promoter. A unique combination of known factors. J Mol Biol *209*, 205-219.

Veis, D.J., Sorenson, C.M., Shutter, J.R., and Korsmeyer, S.J. (1993). Bcl-2-deficient mice demonstrate fulminant lymphoid apoptosis, polycystic kidneys, and hypopigmented hair. Cell *75*, 229-240.

Venter, U., Svaren, J., Schmitz, J., Schmid, A., and Hörz, W. (1994). A nucleosome precludes binding of the transcription factor Pho4 in vivo to a critical target site in the PHO5 promoter. EMBO J *13*, 4848-4855.

Vervoorts, J., Luscher-Firzlaff, J., and Luscher, B. (2006). The ins and outs of MYC regulation by posttranslational mechanisms. J Biol Chem *281*, 34725-34729.

Vervoorts, J., and Lüscher, B. (1999). DNA Binding of Myc/Max/Mad Network Complexes to Oligonucleotides Containing Two E Box Elements: c-Myc/Max Heterodimers Do Not Bind DNA Cooperatively. Biological Chemistry *380*,1121-6.

Viollet, B., Lefrancois-Martinez, A.M., Henrion, A., Kahn, A., Raymondjean, M., and Martinez, A. (1996). Immunochemical characterization and transacting properties of upstream stimulatory factor isoforms. J Biol Chem *271*, 1405-1415.

Vita, M., and Henriksson, M. (2006). The Myc oncoprotein as a therapeutic target for human cancer. Semin Cancer Biol *16*, 318-330.

Wagner, A.J., Meyers, C., Laimins, L.A., and Hay, N. (1993). c-Myc induces the expression and activity of ornithine decarboxylase. Cell Growth Differ *4*, 879-883.

Walhout, A.J., Gubbels, J.M., Bernards, R., van der Vliet, P.C., and Timmers, H.T. (1997). c-Myc/Max heterodimers bind cooperatively to the E-box sequences located in the first intron of the rat ornithine decarboxylase (ODC) gene. Nucleic Acids Research *25*, 1493-1501.

Walhout, A.J., van der Vliet, P.C., and Timmers, H.T. (1998). Sequences flanking the E-box contribute to cooperative binding by c-Myc/Max heterodimers to adjacent binding sites. Biochimica et Biophysica Acta *1397*, 189-201.

Wang, D., and Sul, H.S. (1995). Upstream Stimulatory Factors Bind to Insulin Response Sequence of the Fatty Acid Synthase Promoter USF1 is regulated. J Biol Chem *270*, 28716-28722.

Wang, D., and Sul, H.S. (1997). Upstream stimulatory factor binding to the E-box at -65 is required for insulin regulation of the fatty acid synthase promoter. J Biol Chem *272*, 26367-26374.

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 USA *92*, 5510-5514.

Wang, H., and Wollheim, C.B. (2002). ChREBP rather than USF2 regulates glucose stimulation of endogenous L-pyruvate kinase expression in insulin-secreting cells. J Biol Chem *277*, 32746-32752.

Wang, J., Wang, H., Li, Z., Wu, Q., Lathia, J.D., McLendon, R.E., Hjelmeland, A.B., and Rich, J.N. (2008). c-Myc Is Required for Maintenance of Glioma Cancer Stem Cells. PLoS ONE *3*, e3769.

Wang, S., Skorczewski, J., Feng, X., Mei, L., and Murphy-Ullrich, J.E. (2004). Glucose up-regulates thrombospondin 1 gene transcription and transforming growth factor-beta activity through antagonism of cGMP-dependent protein kinase repression via upstream stimulatory factor 2. J Biol Chem *279*, 34311-34322.

Wang, V., Davis, D.A., Haque, M., Huang, L.E., and Yarchoan, R. (2005). Differential gene up-regulation by hypoxia-inducible factor-1α and hypoxia-inducible factor-2α in HEK293T cells. Cancer Res *65*, 3299-3306.

Wang, Y., Wong, R.H., Tang, T., Hudak, C.S., Yang, D., Duncan, R.E., and Sul, H.S. (2013). Phosphorylation and recruitment of BAF60c in chromatin remodeling for lipogenesis in response to insulin. Mol Cell *49*, 283-297.

Watanabe, A., Takeda, K., Ploplis, B., and Tachibana, M. (1998). Epistatic relationship between Waardenburg syndrome genes MITF and PAX3. Nat Genet *18*, 283-286.

Watanabe, S., and Peterson, C.L. (2010). The INO80 family of chromatin-remodeling enzymes: regulators of histone variant dynamics. Cold Spring Harbor Symposia on Quantitative Biology *75*, 35-42.

Watt, F., and Molloy, P.L. (1993). Specific cleavage of transcription factors by the thiol protease, m-calpain. Nucleic Acids Research *21*, 5092-5100.

Wechsler, D.S., Papoulas, O., Dang, C.V., and Kingston, R.E. (1994). Differential binding of c-Myc and Max to nucleosomal DNA. Mol Cell Biol *14*, 4097-4107.

Weigert, C., Brodbeck, K., Sawadogo, M., Häring, H.U., and Schleicher, E.D. (2004). Upstream Stimulatory Factor (USF) Proteins Induce Human TGF-β1 Gene Activation via the Glucose-response Element–1013/–1002 in Mesangial Cells. J Biol Chem *279*, 15908-15915.

Wenzel A., S.M. (1995). The mycN/max protein complex in neuroblastoma. Short review. Eur J Cancer *31A(4)*, 516-519.

West, A.G., Huang, S., Gaszner, M., Litt, M.D., and Felsenfeld, G. (2004). Recruitment of histone modifications by USF proteins at a vertebrate barrier element. Mol Cell *16*, 453-463.

Whiteside, S.T., and Goodbourn, S. (1993). Signal transduction and nuclear targeting: regulation of transcription factor activity by subcellular localisation. Journal of Cell Science *104*, 949-955.

Widlund, H.R., and Fisher, D.E. (2003). Microphthalamia-associated transcription factor: a critical regulator of pigment cell development and survival. Oncogene *22*, 3035-3041.

Wiesener, M.S., Jürgensen, J.S., Rosenberger, C., Scholze, C.K., Hörstrup, J.H., Warnecke, C., Mandriota, S., Bechmann, I., Frei, U.A., Pugh, C.W.*, et al.* (2003). Widespread hypoxia-inducible expression of HIF-2alpha in distinct cell populations of different organs. FASEB *17*, 271-273.

Williams, C.C., Allison, J.G., Vidal, G.A., Burow, M.E., Beckman, B.S., Marrero, L., and Jones, F.E. (2004). The ERBB4/HER4 receptor tyrosine kinase regulates gene expression by functioning as a STAT5A nuclear chaperone. Journal of Cell Biology *167*, 469-478.

Wiltshire, S., Hattersley, A.T., Hitman, G.A., Walker, M., Levy, J.C., Sampson, M., O'Rahilly, S., Frayling, T.M., Bell, J.I., Lathrop, G.M.*, et al.* (2001). A genomewide scan for loci predisposing to type 2 diabetes in a U.K. population (the Diabetes UK Warren 2 Repository): analysis of 573 pedigrees provides independent replication of a susceptibility locus on chromosome 1q. American Journal of Human Genetics *69*, 553-569.

Wong, R.H., Chang, I., Hudak, C.S., Hyun, S., Kwan, H.Y., and Sul, H.S. (2009). A role of DNA-PK for the metabolic gene regulation in response to insulin. Cell *136*, 1056-1072.

Wong, R.H., and Sul, H.S. (2010). Insulin signaling in fatty acid and fat synthesis: a transcriptional perspective. Current Opinion in Pharmacology *10*, 684-691.

Wood, M.A., and Walker, W.H. (2009). USF1/2 transcription factor DNA-binding activity is induced during rat Sertoli cell differentiation. Biology of Reproduction *80*, 24-33.

Wu, K., Jiang, S.W., Thangaraju, M., Wu, G., and Couch, F.J. (2000a). Induction of the BRCA2 promoter by nuclear factor-kappa B. J Biol Chem *275*, 35548-35556.

Wu, M., Hemesath, T.J., Takemoto, C.M., Horstmann, M.A., Wells, A.G., Price, E.R., Fisher, D.Z., and Fisher, D.E. (2000b). c-Kit triggers dual phosphorylations, which couple activation and degradation of the essential melanocyte factor Mi. Genes Dev *14*, 301-312.

Wu S, P.A., Korcz A, Soprano DR, Soprano KJ. (1996). Overexpression of Mxi1 inhibits the induction of the human ornithine decarboxylase gene by the Myc/Max protein complex. Oncogene *12*, 621-629.

Xiao, Q., Kenessey, A., and Ojamaa, K. (2002). Role of USF1 phosphorylation on cardiac alpha-myosin heavy chain promoter activity. American journal of physiology Heart and circulatory physiology *283*, 213-219.

Xiao, Z., Latek, R., and Lodish, H.F. (2003). An extended bipartite nuclear localization signal in Smad4 is required for its nuclear import and transcriptional activity. Oncogene *22*, 1057-1069.

Xing, J., Ginty, D.D., and Greenberg, M.E. (1996). Coupling of the RAS-MAPK pathway to gene activation by RSK2, a growth factor-regulated CREB kinase. Science *273*, 959-963.

Xing, J., Kornhauser, J.M., Xia, Z., Thiele, E.A., and Greenberg, M.E. (1998). Nerve growth factor activates extracellular signal-regulated kinase and p38 mitogen-activated protein kinase pathways to stimulate CREB serine 133 phosphorylation. Mol Cell Biol *18*, 1946-1955.

Xu, D., Popov, N., Hou, M., Wang, Q., Björkholm, M., Gruber, A., Menkel, A.R., and Henriksson, M. (2001). Switch from Myc/Max to Mad1/Max binding and decrease in histone acetylation at the telomerase reverse transcriptase promoter during differentiation of HL60 cells. Proc Natl Acad Sci USA *98*, 3826-3831.

Xu, J., Chen, Y., and Olopade, O.I. (2010). MYC and Breast Cancer. Genes & Cancer *1*, 629-640.

Xu, W., Gong, L., Haddad, M.M., Bischof, O., Campisi, J., Yeh, E.T., and Medrano, E.E. (2000). Regulation of microphthalmia-associated transcription factor MITF protein levels by association with the ubiquitin-conjugating enzyme hUBC9. Exp Cell Res *255*, 135-143.

Yang, X.J., Ogryzko, V.V., Nishikawa, J., Howard, B.H., and Nakatani, Y. (1996). A p300/CBP-associated factor that competes with the adenoviral oncoprotein E1A. Nature *382*, 319-324.

Yang, X.P., Freeman, L.A., Knapper, C.L., Amar, M.J., Remaley, A., Brewer, H.B., Jr., and Santamarina-Fojo, S. (2002). The E-box motif in the proximal ABCA1 promoter mediates transcriptional repression of the ABCA1 gene. Journal of Lipid Research *43*, 297-306.

Yasumoto, K., Yokoyama, K., Takahashi, K., Tomita, Y., and Shibahara, S. (1997). Functional analysis of microphthalmia-associated transcription factor in pigment cell-specific transcription of the human tyrosinase family genes. J Biol Chem *272*, 503-509.

Yatabe, N., Kyo, S., Maida, Y., Nishi, H., Nakamura, M., Kanaya, T., Tanaka, M., Isaka, K., Ogawa, S., and Inoue, M. (2004). HIF-1-mediated activation of telomerase in cervical cancer cells. Oncogene *23*, 3708-3715.

Yavuzer, U., and Goding, C.R. (1994). Melanocyte-specific gene expression: role of repression and identification of a melanocyte-specific factor, MSF. Mol Cell Biol *14*, 3494-3503.

Yavuzer, U., Keenan, E., Lowings, P., Vachtenheim, J., Currie, G., and Goding, C.R. (1995). The Microphthalmia gene product interacts with the retinoblastoma protein in vitro and is a target for deregulation of melanocyte-specific transcription. Oncogene *10*, 123-134.

Zacharioudakis, I., Gligoris, T., and Tzamarias, D. (2007). A Yeast Catabolic Enzyme Controls Transcriptional Memory. Current Biology *17*, 2041-2046.

Zaret, K.S., and Carroll, J.S. (2011). Pioneer transcription factors: establishing competence for gene expression. Genes Dev *25*, 2227-2241.

Zelzer, E., Wappner, P., and Shilo, B.Z. (1997). The PAS domain confers target gene specificity of Drosophila bHLH/PAS proteins. Genes Dev *11*, 2079-2089.

Zervos, A.S., Gyuris, J., and Brent, R. (1993). Mxi1, a protein that specifically interacts with Max to bind Myc-Max recognition sites. Cell *72*, 223-232.

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, W., Tsuchiya, T., and Yasukochi, Y. (1999). Transitional change in interaction between HIF-1 and HNF-4 in response to hypoxia. Journal of Human Genetics *44*, 293-299.

Zhang, Y., Hassan, M.Q., Li, Z.Y., Stein, J.L., Lian, J.B., van Wijnen, A.J., and Stein, G.S. (2008). Intricate gene regulatory networks of helix-loop-helix (HLH) proteins support regulation of bone-tissue related genes during osteoblast differentiation. Journal of Cellular Biochemistry *105*, 487-496.

Zhao, H., Jin, S., Fan, F., Fan, W., Tong, T., and Zhan, Q. (2000). Activation of the transcription factor Oct-1 in response to DNA damage. Cancer Res *60*, 6276-6280.

Zhou, Z., Li, X., Deng, C., Ney, P.A., Huang, S., and Bungert, J. (2010). USF and NF-E2 cooperate to regulate the recruitment and activity of RNA polymerase II in the beta-globin gene locus. J Biol Chem *285*, 15894-15905.

Zhu, J., Giannola, D.M., Zhang, Y., Rivera, A.J., and Emerson, S.G. (2003). NF-Y cooperates with USF1/2 to induce the hematopoietic expression of HOXB4. Blood *102*, 2420-2427.

Zimmerman, K.A., Yancopoulos, G.D., Collum, R.G., Smith, R.K., Kohl, N.E., Denis, K.A., Nau, M.M., Witte, O.N., Toran-Allerand, D., Gee, C.E.*, et al.* (1986). Differential expression of myc family genes during murine development. Nature *319*, 780-783.