Identification of downstream target genes and analysis of obesity-related variants of the bHLH/PAS transcription factor Single-minded 1

Anne Raimondo

B. Science (Molecular Biology) (Honours)

A thesis submitted in fulfilment of the requirements for the degree of
Doctor of Philosophy

Discipline of Biochemistry

School of Molecular and Biomedical Science

University of Adelaide, Australia

June 2011

CONTENTS

ABSTRACT

Single-minded 1 (SIM1) is a basic Helix-Loop-Helix/PER-ARNT-SIM (bHLH/PAS) transcription factor essential for survival in mice. The early post-natal lethality exhibited by *Sim1*$^{-/-}$ mice is believed to be the consequence of severely compromised hypothalamus development, although the contribution of reduced SIM1 signalling in the numerous other tissues in which it is expressed has never been formally investigated. The presence of a single *Sim1* allele is sufficient to avoid this perinatal lethality, and instead confers an early onset, hyperphagic obesity phenotype, potentially via disruption of critical intracellular signalling pathways that are activated in *Sim1*-expressing hypothalamic neurons in response to food intake. Similar correlations between reduced *SIM1* gene dosage and severe obesity have also been documented in humans. Alterations in *SIM1* expression and/or function therefore have important implications in health and disease, and warrant a detailed investigation into the downstream target genes and regulatory behaviours of this critical transcription factor, which are thus far almost entirely lacking in the literature.

The studies presented in this thesis describe a twofold approach to dissecting the gene regulatory properties of the SIM1 protein. Firstly, we optimised and performed a range of functional assays, including a cell-based luciferase reporter gene assay, a subcellular localisation assay, a co-immunoprecipitation assay, and an electrophoretic mobility shift assay, which were designed to assess the contribution of nineteen unique point mutations within the SIM1 protein sequence to altered SIM1 expression and behaviour. These nineteen mutations were identified in multiple cohorts of severely obese humans, and therefore represent potentially pathogenic alterations in the SIM1 sequence. Indeed, we observed a significant loss of function for many of these variants in luciferase reporter gene assays relative to wild type SIM1. The severe loss of function observed for one of these variants, SIM1 T292A, could be further attributed to altered subcellular localisation, thus impacting on its ability to form a stable heterodimer with ARNT2 in co-immunoprecipitation experiments. Secondly, we performed microarray studies on cultured kidney-derived cells inducibly overexpressing Myc-tagged SIM1 and its obligate partner factor ARNT2, and subsequently identified several genes that selectively responded to SIM/ARNT2 overexpression in this context. Further validation in hypothalamus-derived cultured cells highlighted *Myomesin 2* (*Myom2*) as a potentially

genuine downstream SIM1 target gene in both kidney and hypothalamus. We also present data that are the first to indicate *Somatostatin* (*Ss*) as a hypothalamic target gene regulated by SIM1 in a cell-autonomous manner.

These data are among the first to dissect the downstream target genes and regulatory properties of the SIM1 protein, and therefore make an important contribution to our understanding of the molecular basis to the hyperphagic obesity exhibited by *Sim1*$^{+/-}$ mice. They are also the first to link reduced activities of mis-sense mutations in the SIM1 coding sequence to increased weight gain in humans, and give further credence to the possibility that *SIM1* represents a novel genetic contributor to obesity disorders in the wider population. This knowledge may inform future attempts to develop therapies for obese phenotypes in humans, and broaden our understanding of the molecular events that underpin *Sim1*-mediated survival and maintenance of homeostasis.

DECLARATION

This work contains no material which has been accepted for the award of any other degree or diploma in any university or other tertiary institution to Anne Raimondo, and, to the best of my knowledge and belief, contains no material previously published or written by another person, except where due reference has been made in the text.

I give consent to this copy of my thesis, when deposited in the University Library, being made available for loan and photocopying, subject to the provisions of the Copyright Act 1968.

I also give permission for the digital version of my thesis to be made available on the web, via the University's digital research repository, the Library catalogue and also through web search engines, unless permission has been granted by the University to restrict access for a period of time.

Anne Raimondo

June 2011

ACKNOWLEDGEMENTS

I have benefited from the intelligence, insight, enthusiasm and friendship of countless people during my time at this University. I would particularly like to acknowledge the following people:

Associate Professor Murray Whitelaw – an unfailingly supportive, enthusiastic and encouraging supervisor who has given me an extraordinary amount of freedom to develop independently as a scientist. I am very appreciative of his co-operative and stimulating approach to scientific research and collaboration, and value his contribution to my scientific development over the last six years.

Dr. Anne Chapman-Smith – a source of continued inspiration, from whose fastidious, rational and methodical approach to science I have constantly benefited, and for whose contribution to my professional and personal development I am extremely grateful.

Dr. Simon Koblar and Dr. Dan Peet – official and unofficial co-supervisors respectively, who have been so generous in their contributions to my work, and who have always been available for advice and support.

Colleen Bindloss – a wonderful scientist and human being, without whose assistance this thesis would not have been completed.

Dr. Shwetha Ramachandrappa – a truly inspiring person whose friendship and support I value highly, and whose hard work and determination have made an invaluable contribution to this work. I hope you enjoy this, the companion volume to "Planes, Trains and Automobiles".

I am indebted to Mark Van der Hoek and Rosalie Kenyon at the Adelaide Microarray Facility, and Dr. Steven Pederson at the Women's and Children's Hospital, for their contributions to the microarray studies presented in this thesis, and to Associate Professor Gary Glonek for advice on statistical analysis.

I am grateful to our collaborators Dr. Sadaf Farooqi, Dr. Phillippe Froguel and Dr. Fanny Stutzmann, and all other scientists at the Metabolic Research Laboratories at the University of Cambridge and the Institut de Biologie de Lille at the Institut Pasteur de

Lille, for giving me the opportunity to contribute to their projects, and their patience in waiting for results.

To the many other members of the Whitelaw and Peet labs that I have known over the years, particularly Jodi, Susi, Jo, Fiona, Alix, Scott, Margo, Andrew, Dave, Matt, Adrienne, Veronica, Anthony, Cameron, Sarah, Bec, Karolina, Rachel, Sam, Teresa, Natalia and Sarah – I value your friendship, and the contributions you have made to my personal and professional development, very much.

To the professional staff who make a critical contribution to the daily maintenance of this building and its facilities. I am indebted particularly to John Mackrill for his skill in maintaining the tissue culture facility in the Department of Biochemistry, upon which I was absolutely reliant for my studies, and Serge Volgin and Shirley Coad for their dedication in running the Store.

To all other members of the Thomas and Jensen labs and the Departments of Biochemistry and Genetics, amongst whom I count many friends and admirable scientists.

I must also acknowledge the assistance of the Healthy Development Adelaide Research Cluster in financially supporting my studies.

Finally, deepest thanks must to go my parents, brothers and sister, extended family and friends for their unfailing support and encouragement, which I am not always very good at acknowledging but appreciate nonetheless.

CHAPTER 1: INTRODUCTION

1.1. The bHLH/PAS family of transcription factors

The basic Helix-Loop-Helix/Per-ARNT-SIM (bHLH/PAS) family of transcriptional regulatory proteins represent a class of transcription factors that possess diverse roles during embryonic development and the adaptive response to stress [1-5]. Many members of this family are essential for survival, and have been identified in a wide range of organisms including human, mouse, fruit fly, fish, frog, chicken, and dog. They typically act as heterodimers in order to alter expression of their direct target genes, and in this respect share a common domain architecture that influences their regulatory properties.

bHLH/PAS proteins display a high degree of sequence and structural conservation in their N-terminal regions, which have been shown to influence their DNA binding and dimerisation characteristics ([6-11] and **Figure 1.1.**). The bHLH domain consists of a short stretch of 8-10 residues rich in basic amino acids (aa) immediately followed by a three-dimensional protein fold consisting of two α-helices separated by a short loop [12]. This domain is critical for DNA binding and appears to be the principal determinant of dimer strength [6-9, 11, 13]. The contiguous PAS region, whilst divergent at the amino acid level between different bHLH/PAS factors, also folds into a characteristic three-dimensional structure, and has been shown to mediate the strength and specificity of partner choice in both invertebrates and mammals, as well as influence the stability and specificity of DNA binding, potentially through direct DNA contact ([8, 14-19] and A. Chapman-Smith, unpublished observation). This in turn impacts on downstream target gene expression, and in this way is believed to be at least partially responsible for the specificity of target gene regulation observed between family members [15, 20]. The C-terminal regions of bHLH/PAS proteins display low sequence conservation and are thought to possess transcriptional activation and/or repression and cofactor binding domains [6, 7, 9, 10, 13, 21].

1.2. The PAS domain

The PAS domain is a sensory module found across all three kingdoms of life. In a general sense, it is associated with proteins involved in signal reception and transduction in response to various stimuli, such as light, oxygen and small ligands. It is often found in combination with a range of other regulatory modules, and as such has been adapted across phylogeny to regulate an astonishing range of cellular functions [14].

The definition of a PAS domain has been modified over the years to incorporate shared sequence and structural elements, but refers primarily to a highly conserved three-dimensional structural fold, in part determined by regions of relatively low sequence homology amongst species [14, 22]. For this reason it is often depicted as a bipartite domain containing two PAS "repeats" of approximately 130aa each, designated PAS A and PAS B [14]. Structurally, it consists of an α-helical PAS core and helical connector on a β-scaffold [23]. In bHLH/PAS proteins, it is also a protein-protein interaction interface that mediates dimer formation, and in this way allows further subdivision of bHLH/PAS proteins into one of two classes based on their dimerisation properties ([19] and **Figure 1.1.**). Class I factors are unable to form homodimers, and function only as heterodimers with a class II factor. Class I includes the Aryl Hydrocarbon Receptor (AHR), which has a well-characterised role in the metabolism of environmental pollutants; the Hypoxia Inducible Factor-1 alpha and -2 alpha (HIF-1α and -2α) proteins, which are stabilised under conditions of low oxygen and subsequently perform critical roles in the adaptive response to cellular hypoxia; and the more recently characterised Neuronal PAS Domain Protein 4 (NPAS4) protein, which possesses roles in neuronal activity-dependent synapse formation [4, 5, 24]. Class II bHLH/PAS factors appear capable of forming homodimers *in vitro* as well as *in vivo*, although the functional significance of this observation is not well defined [25-27]. Signalling typically involves nuclear localisation of a class I factor in response to an exogenous signal, dimerisation with a class II factor such as AHR Nuclear Translocator (ARNT) or ARNT2, and DNA binding and regulation of gene expression via interactions with the transcriptional machinery and other cofactors [7, 9-11, 28, 29]. Generally, bHLH/PAS proteins bind asymmetric E-box elements containing a conserved 5' CGTG 3' core.

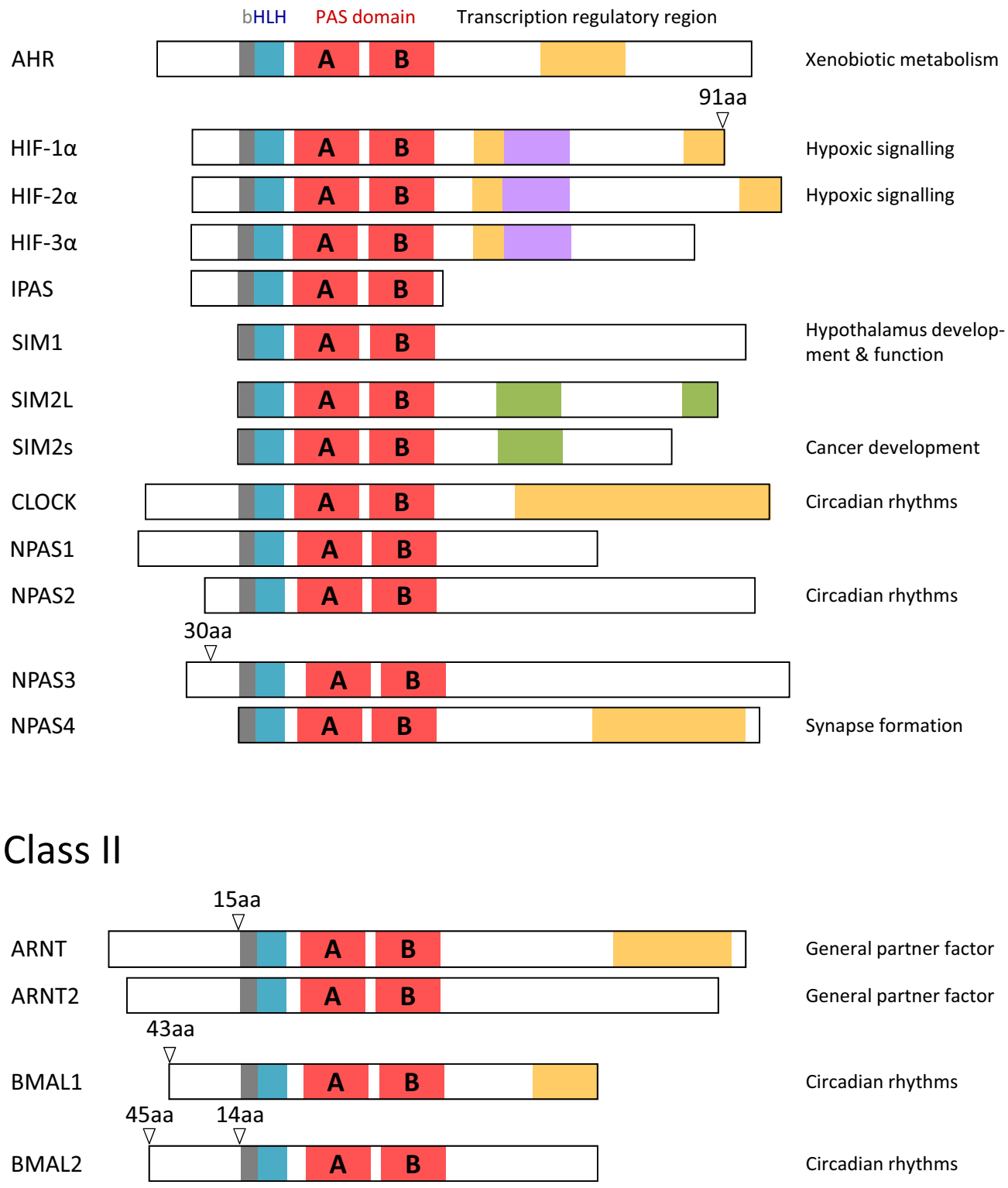


Figure 1.1. Schematic representation of the domain structure of the mammalian bHLH/PAS transcription factor family. Domains are colour coded as follows: **grey**, basic domain; **blue**, Helix-Loop-Helix domain; **red**, PAS domain containing PAS A and PAS B repeats; **orange**, transactivation domain; **green**, transrepression domain; **purple**, oxygen dependent degradation domain. A brief description of the known physiological role for each factor is included on the right.

The subfamily of bHLH/PAS factors involved in regulation of circadian rhythms has been reviewed in detail elsewhere and will not be discussed here [30].

1.3. Drosophila *single-minded* (*sim*)

The *sim* gene was first identified as part of a larger chromosomal region that conferred embryonic lethality in *Drosophila melanogaster* upon complete or partial deletion [31]. It spans 8 exons over 20 kilobases (kb) of DNA and encodes two transcripts of approximately 3.0kb and 3.5kb in size [32, 33]. The latest release of the *Drosophila melanogaster* genome indicates that two distinct proteins may be made from the *sim* gene, of 664aa (SIM B) and 688aa (SIM A), due to selection of alternative initiator methionines during translation (<http://flybase.org>). The significance of these two isoforms appears to be unexplored.

1.3.1. *sim* expression and function

sim is first expressed at the transcript level in the fly at the cellular blastoderm stage (within 3 hours post-fertilisation), in a narrow strip of cells running the anterior-posterior length of the embryo on the ventral side, at the border between the surface ectoderm and presumptive mesoderm ("mesectoderm") [32]. These cells later delaminate and send their nuclei internally, whilst maintaining cytoplasmic contact with the surface of the embryo. They eventually differentiate into the neural and specialised glial cells that comprise the central midline, a distinct morphological structure that influences commissural axon guidance, as well as differentiation of the adjacent mesoderm and lateral neurectoderm throughout the peak neurogenic period (3-12 hours post-fertilisation) [2]. In *sim* mutant embryos these cells are completely absent from the delaminating layer, resulting in collapse and subsequent fusion of the two, normally separate parallel longitudinal commissures, and adoption of a more lateral neuroectodermal cell fate by surrounding tissues [32-34].

It is not known whether the absence of midline cells in *sim* mutant embryos is due to cell death or a change of fate [32]. What is clear, however, is that the correct migration, differentiation, and potentially survival of these cells is dependent on *sim* expression [33]. A wealth of information is known about the genes involved in midline cell formation, and

expression of many of these is controlled, directly or indirectly, by SIM (**Section 1.6.2.2.1.**). Furthermore, ectopic overexpression of *sim* in lateral central nervous system (CNS) tissues is sufficient to convert them to a midline cell fate, adopting both the morphology and gene expression patterns characteristic of this region [35]. This defines SIM as a master regulator transcription factor, both necessary and sufficient for midline cell formation [33, 35].

1.3.2. Control of *sim* expression and activity

In contrast to other class I bHLH/PAS factors, SIM has yet to be ascribed an exogenous regulatory signal. Its highly restricted tissue expression pattern and constitutively nuclear localisation suggests that regulation of direct SIM targets may in fact occur in a spatial manner [32, 36, 37]. This is supported by the observation that expression of the class II partner factor Tango (the *Drosophila* orthologue of ARNT), which is usually cytoplasmic and presumably latent, co-localises with SIM in the nucleus in midline cells [37]. This phenomenon has also been observed in tracheal cells, development of which is dependent on the *Drosophila* class I bHLH/PAS factor Trachealess, which also interacts with Tango to form a functional heterodimer [37].

sim is one of many genes whose products control dorsal-ventral patterning throughout *Drosophila* embryogenesis. Its expression is therefore influenced by several genes that confer positional information along this axis, although direct *cis*-regulation of *sim* expression by most of these factors has not been conclusively shown. The *sim* gene possesses two promoters, denoted P_E and P_L, which influence *sim* expression in the early and late stages of development respectively [35, 38]. Initiation of *sim* expression from P_E and its restriction to mesectodermal cells may be directly regulated by the Dorsal, Twist, Snail, Daughterless and Scute transcription factors, for which multiple binding sites within P_E have been identified and tested [39, 40]. Notch signalling is also necessary to activate *sim* expression within mesectodermal cells in the early stages of *Drosophila* cell lineage specification, via alleviation of direct Suppressor of Hairless (Su(H))-mediated repression [41]. SIM also maintains its own expression in later stages of development by binding directly to P_L [35, 38].

1.4. The mammalian *Sim* genes: *Sim1* and *Sim2*

sim has two homologues in mammals. The very first identified mammalian *sim* gene was human *SIM2* (*hSIM2*), the result of exon trapping sequences from the Down Syndrome critical region on chromosome 21 [42-44]. These sequences were subsequently used to identify homologous full-length clones from both mouse and human cDNA libraries, yielding human *SIM1* (*hSIM1*), mouse *Sim1* (*mSim1*) and mouse *Sim2* (*mSim2*) [10, 45-47]. Concurrently, researchers also identified *mSim1* and *mSim2* on the basis of sequence homology to a *sim* probe [48, 49]. The *hSIM1* gene encodes a single transcript of ~9.5kb and a single protein of 766aa [45]. The *mSim1* gene encodes a single protein of 765aa [10]. The human, and presumably mouse, *SIM2* gene encodes multiple transcripts of varying lengths [45]. To date, two distinct SIM2 protein isoforms have been reported for both species, designated SIM2 long (SIM2L) and SIM2 short (SIM2s) respectively, the latter incorporating a unique C-terminus via alternative splicing ([45, 50] and **Figure 1.1.**).

Mouse and human SIM1, SIM2L and SIM2s display a high degree of sequence homology to SIM, and an exceptionally high degree of homology to each other, in their N-terminal regions (**Figure 1.2.**). Homology decreases markedly after the end of the PAS region, in a manner characteristic of the wider bHLH/PAS family. The respective mammalian orthologues, however, maintain their sequence conservation throughout the entire length of both proteins, indicating a high likelihood of conserved function between species. *Sim1* and *Sim2* homologues have been identified in organisms as diverse as frog, rat, chick, zebrafish, cow, horse, pig, cat, dog and amphipod [43, 51-56]. SIM1 and SIM2 function, however, has only been studied in detail in human and mouse. The information derived from these studies informs the experiments presented in this thesis, which focus mainly on SIM1.

1.4.1. *Sim1* and *Sim2* expression patterns

1.4.1.1. *Sim1*

Sim1 expression has been examined in detail in the developing mouse, and has been detected in the embryonic brain, kidney, and somites, as well as the neural tube, head

mesenchyme, foregut, muscle, dermis and genital eminence at various stages of embryogenesis [10, 49]. Its earliest expression is in the presomitic mesoderm at E8.0, and throughout somitogenesis becomes gradually more refined in this area, until it is expressed specifically in the lateral aspect of the dermatome that later forms the dermis [10, 49]. In the brain, *Sim1* is first detected at E9.0 in an area that extends from the basal portion of the caudal diencephalon into the rostral mesencephalon [49]. This later resolves into a longitudinal zone of expression that incorporates both the ventricular and mantle portions of a region extending through the basolateral mammillary nuclei to the anterior midbrain, including the zona limitans [49]. By E10.5 a small, separate area of expression also emerges in the alar plate of the secondary prosencephalon, which constitutes the prospective paraventricular and supraoptic nuclei (PVN and SON) and amygdala of the hypothalamus [49, 57] (**Section 1.6.1.** and **Figure 1.3.**). Its expression within the prospective anterior hypothalamus appears to be restricted to post-proliferative neurons that reside within the mantle layer. It is interesting to note that *Sim1* is not expressed in the floor plate cells of the ventral spinal cord – a structure believed to correspond most closely to the *Drosophila* central midline – although it can be detected in bilateral stripes flanking this area from E10.0 [49].

Although not explicitly examined after E16.5, *Sim1* expression within the murine central nervous system is probably maintained throughout the remainder of embryogenesis, and has been specifically detected in the post-natal PVN, SON, anterior periventricular nucleus (aPV), and basomedial amygdala, as well as scattered cells of the lateral hypothalamic area (LHA), by *in situ* hybridisation [58, 59]. *Sim1* expression has also been detected in the adult mouse kidney and skeletal muscle via Northern blot on whole tissue [10, 60] and the human adult kidney, skeletal muscle, bone marrow, breast, liver, ovary, pancreas, placenta, testis and tonsil via PCR [61].

1.4.1.2. *Sim2*

Like *Sim1*, *Sim2* is expressed in the embryonic kidney, muscle, and brain [10, 45, 49, 61, 62]. In the newborn hypothalamus, *Sim2* transcript is expressed in the dorsal preoptic area, as well as the aPV and the anterior and mid-PVN [62]. It is therefore expressed in a partially overlapping manner with *Sim1*, which is expressed throughout the entire aPV and PVN [49, 57]. *Sim2* transcript has also been detected in the facial cartilage, palate,

		SIM A		SIM B		hSIM1		mSIM1		hSIM2L		mSIM2L	
		bHLH/PAS	Full length	bHLH/PAS	Full length	bHLH/PAS	Full length	bHLH/PAS	Full length	bHLH/PAS	Full length	bHLH/PAS	Full length
SIM B	bHLH/PAS	100											
	Full length		100										
hSIM1	bHLH/PAS	75		75									
	Full length		42		44								
mSIM1	bHLH/PAS	74		74		99							
	Full length		43		45		96						
hSIM2L	bHLH/PAS	73		73		88		88					
	Full length		40		40		55		56				
mSIM2L	bHLH/PAS	73		73		89		89		97			
	Full length		42		42		58		58		91		
hSIM2s	bHLH/PAS	73		73		88		88		100		97	
	Full length		46		46		62		62		92		85

Figure 1.2. Percentage amino acid identity between *Drosophila* SIM and human and mouse SIM1, SIM2L and SIM2s. SIM A is 688aa in length and possesses an extra N-terminal 24aa relative to SIM B of 664aa (see main text). bHLH/PAS refers to the entire N-terminal sequence of each protein up to the end of the PAS domain, which is defined with reference to [22]. Genbank accession numbers are sourced from NCBI Protein <http://www.ncbi.nlm.nih.gov/protein> and are as follows: SIM A NP_524340.2; SIM B NP_731771.3; hSIM1 NP_005059.2; hSIM2L NP_005060.1; hSIM2s NP_033664.2; mSIM1 NP_035506.2; mSIM2L NP_035507.2. Full-length mSIM2s has yet to be lodged in the NCBI database.

Figure 1.3. Pictorial representation of the anatomical subdivisions of the hypothalamus (saggital view). The paraventricular and supraoptic nuclei are depicted in **blue**. The periventricular nucleus is depicted in **green**. Figure adapted from images sourced from <http://www.wiredtowinthemovie.com/mindtrip_xml.html> and <http://brainmind.com/LimbicOverview.html>.

ribs, vertebrae and lung of the embryonic mouse [10, 48, 49, 63]. It is expressed in similar tissues in the adult, and like *Sim1* has also been detected in the human testis and tonsil by PCR, as well as the lung, heart and prostate [60, 61]. *Sim2s* transcript is specifically upregulated in pancreatic, colon and prostate cancer tissues relative to corresponding benign tissues, and significant effort has been directed towards investigating its suitability as a possible solid tumour marker in this respect [61, 64, 65]. A more comprehensive summary of *Sim2* expression patterns in the embryo and adult can be found elsewhere [S. Woods, PhD thesis 2004; A. Farrall, PhD thesis 2009].

1.4.2. Control of *Sim1* and *Sim2* expression and activity

1.4.2.1. *Sim1*

Little is known of the upstream regulatory factors that control *Sim1* expression, and it is only recently that clues have begun to emerge concerning regulation of *Sim1* expression and activity *in vivo*. *Sim1* is frequently used as a marker of V3 interneurons of the ventral spinal cord in numerous species, and in this respect is sensitive to the correct spatio-temporal expression of several factors involved in ventral neural tissue specification, including the bHLH proteins OLIG1 and 2 [66], NGN3 [67, 68], and members of the NKX family of homeodomain transcription factors [66, 69, 70]. It is not known whether *Sim1* is a downstream target of any of these factors, although Treff *et al* reported two putative NGN3 binding sites in the *mSim1* promoter that have yet to be functionally tested [68].

The *mSim1* promoter has been cloned and the transcription start site located 1121 base pairs (bp) upstream of the initiator ATG [71]. It contains consensus TATA, GC and CAAT boxes as well as, interestingly, a consensus AHR binding site (xenobiotic response element or XRE, 5' TNGCGTG 3') that responds to the presence of AHR/ARNT and/or AHR/ARNT2 dimers in electrophoretic mobility shift assays (EMSAs) and luciferase reporter gene assays. Given the lack of phenotypic overlap between *Sim1*$^{-/-}$ and *Ahr*$^{-/-}$ mice, however, regulation of *Sim1* gene expression by the AHR *in vivo* seems unlikely [58, 72-74]. No further work that attempts to elucidate a possible role for the AHR in *Sim1* gene expression has been published.

SIM1 and SIM2, like SIM, display a constitutively nuclear subcellular localisation [75]. This initially suggested that their activity, like that of SIM, was more likely to be spatio-temporally controlled. However, recent work indicates that *Sim1* expression can indeed be modulated in response to physiological stimuli [76]. *Sim1* mRNA levels in the hypothalamus are increased in response to two signals associated with the neural response to increased food intake: leptin, a CNS-acting hormone released from adipose tissue, and melanotan II (MTII), a synthetic agonist of the Melanocortin 4 Receptor (MC4R), both of which are known to act in the hypothalamus, and thus have important implications in the ability of the body to maintain appropriate energy balance [76, 77]. This will be discussed in more detail below.

1.4.2.2. *Sim2*

Regulation of *Sim2* gene expression is poorly understood and will not be discussed here in detail. Briefly, animal studies and cell-based assays have implicated Notch and Sonic Hedgehog signalling in *Sim2* gene regulation, and consensus sites for the transcriptional regulators E2F, c-MYB and E47 have been identified in the *Sim2* promoter, although their functionality has not been explicitly tested [78-80]. There is no evidence to suggest that *Sim2* transcript levels are mediated in a signal-responsive manner.

1.4.3. Biochemical properties of SIM1 and SIM2

Our knowledge of the regulatory properties of the SIM1 protein is derived mostly from cell-based assays and other *in vitro* experiments. Yeast two-hybrid and co-immunoprecipitation (CoIP) studies have shown that SIM1 is capable of heterodimerising with both ARNT and ARNT2 [10, 60, 81]. *Sim1* is expressed in many of the same tissue types as *Arnt* and *Arnt2* so it is possible that it interacts with one or both of these proteins in a cell- or tissue-specific manner [82, 83]. However, the overlapping phenotypes between *Sim1*$^{-/-}$ and *Arnt2*$^{-/-}$ mice in the brain indicate that the likely *in vivo* dimerisation partner of SIM1 in this area is ARNT2 ([81] and **Section 1.5.3.2.**).

Preliminary experiments that expressed a GAL4-ARNT fusion protein with mSIM1 on a UAS promoter-driven reporter construct in adult monkey kidney COS-7 cells indicated that SIM1 may act as a transcriptional repressor [10]. However, SIM1 has subsequently

been shown to activate gene expression with both ARNT and ARNT2 in similar artificial reporter gene assays mediated by regulatory sequences derived from the *Drosophila toll* gene, a direct SIM target ([13] and **Section 1.6.2.2.1.**). This activity appears to be at least partially dependent on the presence of a C-terminal activation domain in the ARNT protein, since removal of this region abolishes SIM1/ARNT-mediated reporter gene activity [13, 50]. Similar delineation of the functional region(s) critical for SIM1/ARNT2-mediated gene expression has not been reported.

SIM2 possesses both activation and repression properties on endogenous target genes [84, 85]. This raises the possibility that SIM1 may behave similarly *in vivo*. Direct as well as indirect repression has been observed for several members of the bHLH/PAS family including SIM, the latter potentially occurring via sequestration of shared partner factors away from other bHLH/PAS protein complexes, or by activation of genes whose products behave as transcriptional repressors [11, 13, 75, 86, 87]. Furthermore, attempts to characterise a consensus DNA binding sequence for SIM have described a preferential 5' (G/A)(T/A)ACGTG 3' sequence, which overlaps with the response element bound by SIM2 and HIF-1α in mammals *in vivo* [26, 88]. This strengthens the possibility of direct competition for, and cross-regulation of, target genes amongst class I bHLH/PAS factors [75]. This phenomenon has already been described for SIM2 and HIF-1α [85].

Further details of SIM1-mediated gene expression are discussed in **Section 1.6.**

1.4.4. *Sim1* and *Sim2* gene knockout/transgenic mouse studies

1.4.4.1. *Sim1*

1.4.4.1.1. Germline *Sim1*$^{-/-}$ mice

Homozygous germline deletion of *Sim1* in mice was first published in 1996, and replaced 750bp encompassing the 5' region, initiator ATG, and basic domain of the *Sim1* gene with a pGK-neo cassette [58]. This construct effectively replaced endogenous *Sim1* transcript production with that of a non-functional, "mutant" transcript, which was expressed in a manner similar to endogenous transcript in the hypothalamus and enabled the authors to monitor the fate of *Sim1*-expressing cells. Michaud *et al* reported a perinatal lethal phenotype (uniform mortality within 24 hours of birth) and a complete loss of

hypothalamic PVN and SON cells in *Sim1*$^{-/-}$ mice, as inferred from the lack of expression of numerous peptides normally secreted by these nuclei (Corticotrophin Releasing Hormone (CRH), Thyrotropin Releasing Hormone (TRH), Somatostatin (SS), Arginine Vasopressin (AVP), and Oxytocin (OXT)) [58]. No changes in aPV development were described, although it was noted that this structure is morphologically more difficult to distinguish from the contiguous PVN (**Figure 1.3.**). The posterior pituitary was also smaller and its constituent cells more densely packed, presumably due to loss of axonal projections from the PVN/SON region.

The fact that *Sim1*$^{-/-}$ mice die perinatally indicates that, like other bHLH/PAS factors, it plays a critical, non-redundant role in some aspect of embryonic development [3]. Given the striking hypocellularity of the hypothalamic areas in which *Sim1* is normally expressed, it was proposed that the observed lethality was a consequence of this severe phenotype [58]. The fact that production of "mutant" *Sim1* transcripts was undetectable in newborns also led the authors to suggest that the cells expressing it must have changed fate or died. However, subsequent generation of *Sim1*$^{-/-}$ mice that replaced endogenous *Sim1* expression with that of a lacZ reporter gene led these same authors to report a migration defect, due to detection of an ectopic lacZ-positive region midway between the third ventricle (where the PVN is normally situated) and the surface of the brain [89]. This lateral accumulation of lacZ-expressing cells was evident from as early as E14.5 and, coupled with reduced staining in the more medial PVN/SON areas at later stages, indicated that *Sim1* may in fact be important for the correct positioning of PVN/SON cells, in addition to their terminal differentiation [57, 89]. These observations draw obvious parallels with *sim* function, in that both *sim* and *Sim1* possess roles in the correct migration, and possibly survival, of specific neural cell types [90].

Abnormal PVN/SON cell migration has also been observed in homozygous knockout mouse models for other hypothalamus-expressed transcription factors, including *Arnt2*, the Helix-Turn-Helix transcription factor *Otp*, and the POU domain transcription factor *Brn2* [81, 91, 92]. *Otp* appears to work in parallel with *Sim1* to mediate the correct positioning of PVN/SON cells and is also important for their proliferation [91, 93]. *Brn2*, whose expression is modified in both *Otp*$^{-/-}$ and *Sim1*$^{-/-}$ mice, may be a downstream effector of SIM1 in controlling PVN/SON formation ([58, 91] and **Section 1.6.2.**). Mice lacking the *Arnt2* gene have been published on three separate occasions and, like *Sim1*$^{-/-}$

mice, uniformly describe perinatal lethality and a lack of any distinguishable PVN/SON structures ([81, 94, 95] and **Section 1.5.3.2.**). However, ectopic expression of *Sim1* and *Brn2* in these mice indicates a similar accumulation of cells in regions lateral to the normal PVN/SON position, indicating that both SIM1 and ARNT2 are critical for the correct positioning of these neurons, potentially as a functional heterodimer [95].

1.4.4.1.2. Germline *Sim1*$^{+/-}$ mice

Germline *Sim1*$^{+/-}$ mice have been generated by two independent research teams [59, 96]. Unlike *Sim1*$^{-/-}$ mice, these mice survive to adulthood and display a severe, early onset obesity phenotype principally characterised by hyperphagia, increased linear growth, and elevated blood insulin and leptin levels, without any corresponding decrease in energy expenditure. The first reported *Sim1*$^{+/-}$ mouse displayed hypocellular PVN/SON formation at a gross anatomical level, with a reported 24% decrease in PVN cell counts without loss of any specific neurosecretory cell type [96]. It was proposed that the hyperphagia displayed by these mice was attributable to PVN/SON hypocellularity, which would presumably result in defective PVN/SON function. A related paper published eight years later by the same research group specifically described reductions in PVN/SON *Avp* and *Oxt* transcript and protein levels in these mice relative to wild type (WT) controls, and a corresponding 70% decrease in PVN projections to the dorsal vagal complex (DVC) as detected via retrograde labelling experiments [97]. The DVC and the median eminence (ME) are two areas to which PVN/SON projections involved in propagation of the neuroendocrine feeding response are known to extend (**Section 1.6.1.**). Duplan *et al* also noted that *Sim1*$^{+/-}$ mice displayed increased sensitivity to anorexia in response to chronic ingestion of a high salt solution relative to WT controls [97]. Independent researchers confirmed many of the characteristics conferred by heterozygous *Sim1* deletion three years after publication of the first *Sim1*$^{+/-}$ mouse model, and established that weight gain in these mice is due to a failure to modulate food intake in response to changes in dietary fat content [59]. However, the absence of any detectable PVN/SON hypocellularity in these mice led the authors to propose that dysfunction, not absence, of *Sim1*-expressing neurons contributed to the observed phenotypic traits [59].

The contribution of *Sim1* heterozygosity to PVN hypocellularity, and its subsequent impact on regulation of feeding in the hypothalamus, remains a disputed one. Successive

publications from the research teams responsible for each of the two original germline *Sim1*$^{+/-}$ mouse models have presented equally convincing data to support their original, and disparate, findings [59, 96-98]. The data do not appear to be misleading on a methodological basis, and no concrete suggestions have been made to explain what appear to be genuine differences between the two strains. It is possible that genetic background may be contributing to the observed phenotypic differences. Michaud *et al* generated their mice via electroporation of their *Sim1* "null" targeting construct (described above) into 129-derived R1 embryonic stem (ES) cells, followed by selection of antibiotic resistant clones and injection into C57BL/6 blastocysts to produce chimeric mice. Chimeras were then backcrossed to C57BL/6 females and maintained on a mixed C57BL/6-129/Sv background [58]. The targeting construct of Holder *et al* was designed to utilise Cre/loxP technology and contained a floxed pGK-neo cassette downstream of an additional loxP site inserted into exon 2 of the *Sim1* coding sequence (cds), which contains the initiator ATG [59]. They generated 129-derived RW-4 ES cell clones via antibiotic selection and used them to breed C57BL/6 heterozygotes, which were crossed with FVB mice ubiquitously expressing Cre recombinase to recombine out the pGK-neo selection marker as well as DNA encoding the first 17aa of the SIM1 protein. The resultant heterozygotes were backcrossed to inbred C57BL/6 mice and therefore possess a mixed C57BL/6-129/Sv-FVB background. There is a precedent for genetic background affecting mutant phenotypes in mice, presumably due to the presence of "modifier" alleles that affect the penetrance of a particular trait [99]. It is not known if this is the reason for the observed differences between the different *Sim1*$^{+/-}$ strains.

Any effect of germline *Sim1* deletion in the other tissues in which it is expressed has not been described.

1.4.4.1.3. Conditional *Sim1*$^{-/-}$ and *Sim1*$^{+/-}$ mice

Conditional *Sim1*$^{-/-}$ and *Sim1*$^{+/-}$ mice have recently been published by the Zinn group, utilising a similar gene targeting strategy to that employed for their germline *Sim1*$^{+/-}$ mouse model [59]. Expression of one or both *Sim1* alleles in post-natal neurons was abolished via CRE recombinase expression from an *α-calcium/calmodulin-dependent protein kinase II* promoter [98]. Conditional *Sim1*$^{-/-}$ and *Sim1*$^{+/-}$ mice are viable and display the phenotypic characteristics of germline *Sim1*$^{+/-}$ mice, including hyperphagia,

increased linear growth, and increased weight gain and food intake on both low and high fat diets relative to littermate controls [59, 96, 98]. Food intake and weight gain were greater in conditional *Sim1*$^{-/-}$ mice compared to *Sim1*$^{+/-}$ mice, indicating a dose-dependent effect of *Sim1* deletion on hyperphagic obesity. Stereological cell counting indicated that the PVN/SON region was histologically indistinguishable from WT in both *Sim1*$^{-/-}$ and *Sim1*$^{+/-}$ mice, as were their axonal projections to the DVC and ME [98]. These data strengthen the argument in favour of an active, post-natal signalling role for *Sim1* in the neural response to food intake, in a manner distinct from its obvious critical role in formation of hypothalamic structures during development. It is possible, of course, that reduced *Sim1* levels contribute to a more generalised defect in PVN/SON neuronal development and/or identity, a consequence of which is a failure to express those gene products that are the hallmarks of a terminally differentiated neuron, resulting in failure to respond to afferent PVN/SON signals [98]. Given the lack of specific markers for the PVN/SON region, this may be difficult to verify [89]. However, recent data describe increased food intake in mice unilaterally injected with adenoviral vectors producing short hairpin RNA (shRNA) sequences directed against the *Sim1* transcript, thus temporarily knocking down *Sim1* expression in the post-natal PVN [100]. Similar injection of *Sim1* overexpression vectors resulted in decreased food intake [100]. Cumulatively, these data suggest a dual role for *Sim1* in PVN/SON development and function, both of which have the potential to impinge upon hypothalamus-mediated regulation of food intake in the adult ([77, 101] and **Section 1.6.**).

1.4.4.1.4. *SIM1* overexpressing mice

Transgenic *SIM1*-overexpressing mice were published in 2006 and involved the generation of C57BL/6 mice constitutively overexpressing a bacterial artificial chromosome (BAC) encompassing 65kb of 5' flanking sequence, the *hSIM1* gene, and 58kb of 3' flanking sequence [102]. They do not display a lean phenotype, as may be expected from the severe obesity displayed upon loss of a single *Sim1* allele, but rather display resistance to weight gain on a high fat diet, without any corresponding change in basal metabolic rate, feeding efficiency or energy expenditure [59, 96, 102]. Feeding efficiency refers to the ease with which an organism gains weight in response to a fixed caloric intake, and can reflect changes in the ability of the organism to utilise and metabolise available energy. The resistance to diet-induced obesity (DIO) displayed by

SIM1 transgenic mice is due to reduced food intake relative to WT mice, indicating that *SIM1*-overexpressing mice are capable of modulating their food consumption in response to changes in dietary fat content in order to maintain isocaloric intake. *SIM1* is therefore more likely to play a role in the ability of these mice to sense the initial "desire to eat" in response to depleted energy levels, without affecting the subsequent metabolic processes that determine the ease with which they utilise and digest this energy from food. Furthermore, *SIM1*-overexpressing mice display normal PVN histology relative to WT controls, indicating that structural PVN malformations cannot explain the observed phenotypic traits.

1.4.4.2. *Sim2*

1.4.4.2.1. *Sim2*$^{-/-}$ and *Sim2*$^{+/-}$ mice

Sim2$^{-/-}$ mice have been generated by two independent research teams from the same ES cell clone [63, 103]. The targeting construct replaced 422bp of the 5'UTR and 144bp of exon 1 of the *Sim2* gene with a pGK-neo cassette [103]. Goshu *et al* also generated compound *Sim1/Sim2* knockout mice, via mating with the *Sim1* heterozygote bred by Michaud *et al* [58, 63]. *Sim2*$^{-/-}$ mice, like *Sim1*$^{-/-}$ mice, display perinatal lethality, however the cause of death according to Goshu *et al* appears to be multiple subtle malformations in the structural components necessary for adequate lung inflation, resulting in breathing failure. These include defects in rib and diaphragm formation, intercostal muscle attachment, and mesothelium formation [63]. Shamblott *et al* did not report similar malformations, but did describe an aerophage phenotype (accumulation of air within the gastrointestinal (GI) tract). They also described craniofacial abnormalities not observed by Goshu *et al*, including cleft secondary palate and malformations of the tongue and spheroid bone [103]. Embryonic or newborn compound *Sim1/Sim2*-deficient mice were analysed for gross PVN histology, and weight gain was also analysed over a seven month period following birth [63]. No contribution of reduced *Sim2* gene dosage to either of these traits in either a WT, *Sim1*$^{+/-}$ or *Sim1*$^{-/-}$ background was observed, although the perinatal lethality of both *Sim1*$^{-/-}$ and *Sim2*$^{-/-}$ mice precluded satisfactory analysis of weight gain in these contexts.

PVN formation and composition in *Sim2*$^{-/-}$ and *Sim2*$^{+/-}$ mice has since been studied more closely by Goshu *et al* [62]. Re-analysis of these mice, as well as compound *Sim1/Sim2* knockout mice, revealed reductions in *Trh* and *Ss* expression in the dorsal preoptic area, aPV, and anterior PVN relative to WT from as early as E12.5, in a subset of those cells that express *Sim2*. *Sim2* expression was also abolished in the *Sim1*$^{-/-}$ PVN. These data indicate partially overlapping roles for *Sim1* and *Sim2* in PVN formation and/or identity, which may indicate direct regulation of shared target genes between the two transcription factors. Further discussion of this concept, including a possible role for SIM1 in regulating *Sim2* gene expression, can be found in **Section 1.6.**

1.4.4.2.2. *Sim2* overexpressing mice

Transgenic *Sim2*-overexpressing mice have also been generated by two independent research teams, the first by ubiquitous *Sim2* overexpression from a *β-actin* promoter and the second by expression via its own regulatory elements from a BAC clone [104, 105]. Spatial *Sim2* expression in the latter was intended to mimic that of endogenous *Sim2*, with a view towards partially reproducing the gene dosage effects observed in Down Syndrome cases of trisomy 21 ([105] and **Section 1.4.**). Neither mouse model displayed any gross phenotypic differences relative to WT controls and exhibited normal brain histology. Behavioural tests indicated a moderate defect in fear conditioning, as determined by the tendency of the animal to "freeze" in anticipation of an imminent electric shock, and slightly reduced exploratory behaviour and resistance to pain [104, 105]. These behavioural characteristics may indicate a partial contribution of increased *SIM2* gene dosage to the etiology of Down Syndrome in humans.

1.5. The mammalian general partner factors ARNT and ARNT2

ARNT was first identified in mammals as an 87 kiloDalton (kDa) protein necessary for cytoplasmic-nuclear translocation of the AHR upon ligand binding and hence ligand-mediated AHR activity [106-108]. It has since been recognised as the mammalian orthologue of the *Drosophila* class II partner factor Tango [109]. In humans and mice the *ARNT* gene spans 22 exons over approximately 65kb of DNA, and it encodes multiple transcripts of varying lengths in several species due to alternative splicing [8, 108, 110-114]. The functions of these different ARNT isoforms have yet to be elucidated [108].

The mouse *Arnt2* (*mArnt2*) gene was identified on the basis of homology to the *hARNT* bHLH/PAS region and was soon after identified in rat and human [27, 112, 115, 116]. In mice, the *Arnt2* gene spans 21 exons over more than 120kb of DNA, with the initiator ATG residing in exon 1. Exons 2 and 3 each contain an in-frame stop codon that subsequently yield "truncated" transcripts. As a result, these two exons are skipped to yield "full-length" mARNT2 of 712aa [27, 95]. *mArnt2* also appears to possess an alternative initiator methionine in exon 4 that results in the absence of 11aa at the N-terminus of full-length mARNT2, upstream of the basic domain [95]. A similar phenomenon has not been described for ARNT2 from other species, nor is there any information to indicate what the function of these different isoforms might be.

ARNT and *ARNT2* homologues have since been identified in numerous species and, like other members of the wider bHLH/PAS family, display a high degree of homology to each other in their N-terminal regions [8, 27, 109, 112, 117-119]. The respective mammalian orthologues maintain their sequence homology throughout the entire length of both proteins. With regards to function they are best characterised in terms of their role as shared partner factors between the various class I bHLH/PAS proteins, and are involved in a wide variety of cellular functions as a result. The experiments presented in this thesis mainly utilise ARNT2, since evidence suggests it is the requisite dimerisation factor for SIM1 *in vivo*.

1.5.1. *Arnt* and *Arnt2* expression patterns

1.5.1.1. *Arnt*

Arnt expression in the mouse is essentially ubiquitous throughout development and into adulthood [27, 82, 83, 120]. At the transcript level it can be detected in tissues of mesodermal and endodermal origin from as early as E9.0, and is also expressed in ectodermal tissue at low levels [27, 82, 83, 120]. Throughout the CNS its expression is relatively low and it appears to be expressed most strongly in peripheral tissues [82, 83, 120]. Widespread *Arnt* expression in the rat brain has also been reported from more detailed studies in this area [121, 122].

1.5.1.2. *Arnt2*

Arnt2 expression is relatively restricted. In the mouse it is expressed most strongly in the embryonic and adult CNS, and has been detected in the neuroepithelium of the neural tube from as early as E9.0 [83]. It is, in this sense, the reciprocal of *Arnt* [27, 82, 83]. *Arnt2* also displays high levels of expression in the embryonic and adult kidney, and has been detected at low levels in the olfactory epithelium, retina, tongue, oesophagus, stomach, intestines, lung, and spleen at various stages of development [27, 82, 83]. In the CNS it is widely expressed at high levels, and is co-expressed with *Sim1* in the prospective PVN/SON [81, 122].

1.5.2. Biochemical properties of ARNT and ARNT2

The functional domains of ARNT have been mapped and suggest the presence of a strong C-terminal transactivation domain that is necessary, at least in part, for SIM1-mediated gene regulation [13, 50, 123]. ARNT also possesses a strong N-terminal nuclear localisation signal (NLS), and is a constitutively nuclear protein in mammals [124-126]. The functional domains of ARNT2 have not been mapped; however, it too is a constitutively nuclear protein (A. Raimondo, unpublished observation).

The high degree of N-terminal sequence homology between bHLH/PAS family members means that ARNT and ARNT2 are capable of forming stable heterodimers with numerous class I factors [1, 9-11]. Whilst able to compensate for each other to a limited degree, ARNT and ARNT2 also clearly possess unique gene regulatory properties, and can contribute to significantly different outcomes in concert with the same class I factor [127, 128]. They are thought to achieve this specificity of target gene regulation via class I factor-mediated protein-DNA interactions and recruitment of specific cofactors ([9-11, 28, 129-131], for a more detailed discussion see **Section 1.6.2.**). At the whole organism level, this concept is evidenced in the markedly different phenotypes exhibited by *Arnt* and *Arnt2* knockout mice [81, 94, 95, 132, 133].

1.5.3. *Arnt* and *Arnt2* gene knockout studies

1.5.3.1. *Arnt*$^{-/-}$ mice

Arnt$^{-/-}$ mice have been generated by two independent research teams [132, 133]. Both teams adopted a similar strategy of interrupting the ARNT bHLH domain with a selectable marker, thus abolishing its DNA binding and dimerisation abilities. *Arnt*$^{-/-}$ mice display embryonic lethality between E9.5 and E10.5, due to multiple defects including defective vascularisation of the yolk sac and branchial arches, stunted development, failure of the neural tube to close, forebrain hypoplasia, placental haemorrhaging, and general wasting of the embryo [132, 133]. Many of these defects highlight the importance of ARNT in facilitating HIF-1α-mediated gene expression programs in the earliest stages of development, since partially similar phenotypes have been reported in mice lacking expression of known HIF-1α/ARNT target genes [134, 135]. *Arnt*$^{-/-}$ ES cells also fail to activate the expression of genes that normally respond to decreased glucose levels and oxygen tension, reinforcing the importance of ARNT in mediating hypoglycaemic and hypoxic responses [132].

1.5.3.2. *Arnt2*$^{-/-}$ mice

The *Arnt2*$^{-/-}$ phenotype was first described in mice lacking a larger chromosomal region encompassing the entire *Arnt2* gene as well as the neighbouring *Fumarylacetoacetate hydrolase* (*Fah*) gene [81]. These mice displayed a perinatal lethal phenotype, previously attributed to liver dysfunction associated with decreased FAH activity [136]. Closer inspection of E18.5 embryos by Michaud *et al* confirmed that, like *Sim1*$^{-/-}$ mice, these mice also lack any distinguishable PVN/SON structures and do not produce detectable *Trh*, *Crh*, AVP and SS. *Sim1* expression in the hypothalamus was comparable to WT at E13.5, suggesting that the early stages of PVN/SON neurogenesis are unaffected by loss of *Arnt2*, however, *Sim1* expression in the PVN/SON region had ceased by E15.5 and was instead ectopically expressed along with *Brn2* in a region lateral to the normal PVN/SON position. This migratory defect is strikingly similar to that observed in *Sim1*$^{-/-}$ mice expressing lacZ under the control of the endogenous *Sim1* promoter, and indicates the importance of both SIM1 and ARNT2 for the correct positioning of PVN/SON neurons, probably as a functional heterodimer [81, 89].

Specific deletion of the *Arnt2* gene has since been performed by two independent research teams, one of which replaced most of the ARNT2 bHLH domain with a pGK-neo cassette, and the other that inserted a NLS-lacZ sequence in frame with the middle of the ARNT2 basic domain to enable tracing of *Arnt2*-expressing cells [94, 95]. Both teams describe a perinatal lethal phenotype and lack of PVN/SON structures characterised by a loss of *Crh, Trh, Ss, Avp* and *Oxt* expression. Hosoya *et al* also reported hypoplasia of the posterior pituitary, as observed in *Sim1*$^{-/-}$ mice, and reduced neurohormone staining in the posterior pituitary and ME, indicating a loss of afferent PVN/SON axonal projections to these areas [58, 95]. Hosoya *et al* also observed an ectopic *Brn2* expression domain lateral to the normal *Brn2*-positive PVN/SON region of E13.5 and E15.5 *Arnt2*$^{-/-}$ mice, similar to that observed by Michaud *et al* [81, 95].

Crossing of *Arnt* and *Arnt2* heterozygotes to generate compound *Arnt/Arnt2*-deficient mice has also been performed [94]. *Arnt*$^{-/-}$*Arnt2*$^{-/-}$, *Arnt*$^{+/-}$*Arnt2*$^{-/-}$ and *Arnt*$^{-/-}$*Arnt2*$^{+/-}$ embryos do not survive beyond E8.5, indicating non-overlapping, essential functions for these two genes beyond this developmental time point. *Arnt2*$^{+/-}$ mice are viable without any obvious morphological or developmental phenotypes, suggesting that a single allele of *Arnt2* is sufficient for normal development, or that *Arnt* is capable of compensating for loss of a single *Arnt2* allele during development [94, 95]. Cumulatively, these data suggest essential roles for both *Arnt* and *Arnt2* during embryonic development and post-natal survival, and more particularly for *Arnt2* in the correct migration, identity and possibly survival of PVN/SON cells.

1.6. The molecular basis of SIM1 function

Studies of the effects of *Sim1* deletion on hypothalamus development and function have naturally prompted a desire to understand the molecular basis of the observed defects [137]. Detailed physiological and pharmacological experiments in mice indicate a critical role for SIM1 in the regulation of energy balance, which is likely to be conserved in humans. It is not surprising, therefore, that SIM1 has recently emerged as a novel downstream effector of MC4R signalling in the PVN, and therefore a potential key mediator of hypothalamic signalling in response to increased food intake [76, 102].

1.6.1. Hypothalamic control of food intake and weight regulation

The hypothalamus has been acknowledged as a brain region central to mediation of the feeding response for several decades, and is regarded as the homeostatic regulatory region of the brain with the greatest influence on maintenance of appropriate energy balance [77, 101]. It responds to a variety of signals that influence energy intake and expenditure, as well as other homeostatic processes, and subsequently provides the organism with the means to adapt continually to changes in energy levels, as well as blood pressure, temperature, heart rate, and numerous other processes [138, 139]. At the physiological level, the hypothalamus integrates hormonal signals and metabolic fuels produced by the gut, pancreas, and adipose tissue, and converts them into downstream effector molecules that influence feelings of hunger and satiety [138-140]. Obesity can therefore result when there is an imbalance between energy intake and utilisation, which may stem from neural defects in the control of appetite [139].

1.6.1.1. The leptin-melanocortin signalling pathway

The signalling pathways activated in response to feeding are numerous and complex [138, 140, 141]. One of the most intensely studied of these is the leptin-melanocortin signalling pathway, the principal mediators of which are leptin, an adiposity signal best characterised in terms of its effects on leptin receptor-expressing cells of the hypothalamic arcuate nucleus (ARC), and alpha-melanocyte stimulating hormone (αMSH), a natural agonist of the MC4R and a downstream effector of leptin signalling in the hypothalamus ([142, 143] and **Figure 1.4.**). Leptin, which is secreted from adipose tissue in high amounts after a meal, enters the CNS and acts on independent populations of neurons in the ARC to stimulate expression of anorexigenic signals such as Pro-opiomelanocortin (POMC) and Cocaine and Amphetamine-Related Transcript (CART), and suppress expression of orexigenic signals such as Neuropeptide Y (NPY) and Agouti-Related Peptide (AGRP) [138, 144]. These so-called "first order" neurons project to "second order" neurons in several hypothalamic areas including the PVN, where αMSH, a cleavage product of POMC, is released. αMSH binds and increases MC4R activity, ultimately reducing food intake [77].

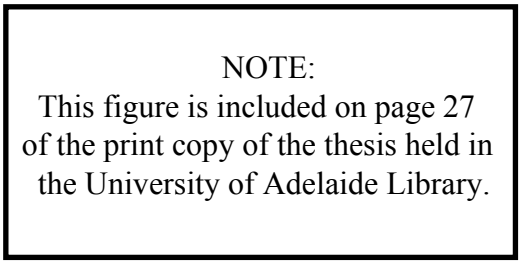


Figure 1.4. Hypothalamic influence on food intake and energy expenditure in response to peripheral signals, with an emphasis on leptin-melanocortin signalling. Leptin, which is released from adipose tissue, binds to its receptor on "first order" ARC neurons, and has an inhibitory effect on orexigenic AGRP and NPY production and a stimulatory effect on anorexigenic POMC and CART production. The high levels of leptin released after a meal therefore favour POMC and CART production. POMC is cleaved to produce αMSH, which binds to the MC4R on "second order" neurons, including PVN neurons, ultimately reducing food intake. Ghrelin, which is released from the stomach, and insulin, which is released from the pancreas, also act centrally. Adapted from [143].

The downstream effectors of MC4R signalling in PVN neurons are poorly understood. In any case, it appears that the effects of MC4R activity on food intake and energy expenditure are mediated by distinct populations of neurons, since reintroduction of *Mc4r* expression specifically in *Sim1*-expressing neurons of *Mc4r*-deficient mice influences food intake but has no effect on energy expenditure [145]. The PVN, its contiguous aPV and the SON cumulatively produce CRH, TRH, SS, AVP and OXT, several of which have anorectic effects when administered centrally [77]. A cleavage product of the SS pre-pro-peptide, Neuronostatin, has also been recently identified, which is expressed in an overlapping manner with SS and also reduces food intake when administered centrally [146]. The PVN/SON has efferent connections to several extrahypothalamic sites, including the DVC and ME, and the posterior pituitary via the ME [147, 148]. These connections are positive for CRH, TRH, SS, AVP and/or OXT and provide direct access for these hormones to the blood and cerebrospinal fluid [149-151]. It is possible that neurohormone activity at all these sites is responsible for propagation of PVN/SON-mediated signalling in response to feeding.

1.6.1.2. *Sim1* and the leptin-melanocortin signalling pathway

The majority of the work that informs our knowledge of a role for SIM1 in leptin-melanocortin signalling has been performed using the germline $Sim1^{+/-}$ mouse model generated by Holder *et al*, and the widely used $Agouti^{Yellow}$ (A^{Y}) mouse model, which ectopically overexpresses the Agouti protein, a MC4R antagonist [59, 76, 102, 152, 153]. A^{Y} mice present with a range of symptoms including hyperphagia, obesity, type II diabetes, increased linear growth, and a yellow coat colour, the latter due to an additional role for the Agouti protein in regulating hair pigmentation [154, 155]. Overexpression of the aforementioned *hSIM1* BAC in A^{Y} mice normalises their food intake and partially rescues their body weight, indicating that *SIM1* is capable of overriding the effect of decreased MC4R signalling in these mice, and may be acting downstream of this receptor to influence weight gain ([102] and **Section 1.4.4.2.2.**).

Sim1 mRNA levels in the hypothalamus of 6-8 week old WT mice are increased in response to intraperitoneal injection of both leptin and the synthetic MC4R agonist MTII [76]. However, MTII treatment does not attenuate food intake to the same extent in $Sim1^{+/-}$ mice as in WT controls due to reduced activation of PVN neurons [76]. The latter

phenomenon has also been observed upon MTII treatment of mice deficient for *Mc4r* expression [156]. Whether the *Mc4r*-positive cells activated in response to MTII treatment are the same cells that display elevated *Sim1* levels is yet to be conclusively proven, although this seems likely given that *Sim1* appears to be expressed throughout the entire PVN/SON, and at least some of the cells stimulated by MTII treatment are *Sim1*-positive [76, 145]. Interestingly, PVN *Mc4r* mRNA levels are reduced in conditional *Sim1*$^{-/-}$ and *Sim1*$^{+/-}$ mice relative to littermate controls, indicating the presence of a potential positive regulatory feedback loop within PVN neurons [98].

Despite their previously described hyperphagia and elevated blood insulin and leptin levels, *Sim1*$^{+/-}$ mice display elevated *Pomc* mRNA levels and decreased *Agrp* mRNA levels in response to leptin treatment, indicating a normal physiological response to leptin that these mice are clearly unable to propagate in the absence of *Sim1* [76]. This in turn suggests dysregulation of the signalling pathways further downstream of leptin receptor activation. Reduced levels of *Crh*, *Trh*, *Ss*, *Avp* and *Oxt* mRNA, as well as AVP and OXT protein, have all been detected in *Sim1*$^{+/-}$ hypothalamus relative to WT [97, 153]. OXT, an anorectic peptide whose expression is normally modulated in WT mice in response to fasting and feeding states, does not follow this pattern in *Sim1*$^{+/-}$ mice [153]. This suggests a direct role for SIM1 in regulation of *Oxt* expression in response to feeding. The MC4R agonist cyclo(β-Ala-His-D-Phe-Arg-Trp-Glu)-NH_2 induces c-Fos expression in PVN cells that co-stain for OXT [153], however, MTII treatment does not affect *Oxt* mRNA levels in WT mice, excluding the possibility of a direct link between MC4R activation, *Sim1* upregulation, and *Oxt* expression in the PVN [76]. However, intracerebroventricular (ICV) injection of an OXT receptor antagonist exacerbates the hyperphagia of *Sim1*$^{+/-}$ mice and ICV OXT injection normalises their food intake, indicating that SIM1-mediated *Oxt* signalling underlies the obesity of these mice in some way [153]. The effect of downregulated OXT production and/or reduced OXT receptor signalling on feeding in *Sim1*$^{+/-}$ mice has been suggested to occur via projections to the DVC [97].

1.6.2. Direct SIM1 target genes

SIM1 has yet to be ascribed a single *bona fide* direct target gene. Detailed animal studies in the last few years have increased our knowledge of the signalling pathways SIM1 is

likely to effect in mammals, but the precise role(s) of SIM1 in PVN/SON development and function, as well as the development and function of the other tissues in which it is expressed, remain unsatisfactorily understood. The identification of SIM1-regulated direct target genes will therefore be of assistance in understanding the molecular basis of severe obesity disorders, as well as perhaps providing novel insights into any other gene expression programs SIM1 may effect throughout embryonic development and post-natal life.

Whilst phenotypic and molecular data indicate an essential, non-redundant function for SIM1, there is also a precedent for functional cross-talk between class I bHLH/PAS factors that must be taken into account when attempting to identify or characterise direct SIM1 target genes. Such cross-talk can occur via competition for binding sites within the regulatory regions of shared target genes, or more indirectly via competition for ARNT or ARNT2 and/or other limiting cofactors [84, 85, 87, 157, 158]. This is perhaps not surprising, given the partial commonality between the various class I bHLH/PAS factors in terms of their expression patterns, DNA binding properties, dimerisation requirements, and N-terminal sequence homologies. The search for direct SIM1 target genes must therefore be informed not only by studies of *Sim1* function and activity *in vitro* and *in vivo*, but also by pioneering studies of the *sim* gene in *Drosophila*, as well as information derived from a small number of direct SIM2-regulated target genes that have only recently been identified and partially characterised.

1.6.2.1. Clues from studies of *Sim1* function and activity

Based on the observed data, SIM1 target genes may be expected to fall into one of three classes. The first includes those genes regulated in response to increased leptin and/or αMSH levels in the hypothalamus after feeding, via MC4R-mediated *Sim1* activation. The second includes those genes involved in establishment of PVN/SON structure and/or identity throughout embryogenesis and early post-natal life, and may include gene products critical for the correct migration and/or terminal differentiation of PVN/SON neurons. The third comprises any other genes involved in SIM1-mediated development and/or function of the other tissues in which it is expressed, such as the kidney. In all cases, SIM1 is presumed to bind DNA sequences containing a conserved 5' A<u>CGTG</u> 3' core as a heterodimer with ARNT and/or ARNT2 (see **Section 1.6.2.2.1.**).

With regards to the first class of direct SIM1 target genes, decreased levels of *Crh*, *Trh*, *Ss*, *Avp*, *Oxt* and *Mc4r* mRNA have all been detected in *Sim1*$^{+/-}$ hypothalamic tissues relative to WT [97, 98, 153]. Provided that this is not merely due to a decrease in PVN/SON cell survival resulting from reduced *Sim1* gene dosage, these data may provide the first glimpses of the signalling events downstream of *Sim1* expression in the hypothalamus (see **Section 1.6.1.**). However, none of these genes has been successfully characterised as a direct SIM1 target, nor is it known if increased leptin or αMSH/MTII activity is necessary to effect their expression by SIM1. Unlike *Sim1*, *Oxt* transcript levels do not increase in response to MTII treatment in mice, excluding the possibility of a direct link between MC4R activation, *Sim1* upregulation, and *Oxt* expression in the PVN [76]. Also, there are no conserved SIM1 binding sites within 5kb of the *Oxt* transcription start site, suggesting that this particular gene may be an indirect target ([153], see **Section 1.6.2.2.1.**).

Little is known about the direct target genes belonging to classes two and three. Broad-scale attempts to identify SIM1-regulated target genes have been published on only two occasions in more than ten years, and have not been pursued [57, 137]. Firstly, Liu *et al* attempted to identify genes differentially regulated in response to doxycycline-induced overexpression of a SIM1-VP16 fusion protein, and ARNT2, in mouse neuroblastoma Neuro-2a cells relative to mock-treated cells [137]. The potent VP16 activation domain replaced the entire SIM1 C-terminal region and was used to achieve maximal induction of any potential target genes upon doxycycline treatment. Of the 268 genes that displayed >1.7-fold upregulation in SIM1-VP16/ARNT2-overexpressing cells relative to control cells by microarray analysis, nine were verified by Northern blot. These included the tyrosine-protein kinase *Jak2*, the nuclear thyroid hormone receptor *Thrβ2*, and the *Melanocortin 3 Receptor* (*Mc3r*). MC3R and MC4R possess unique, non-redundant roles in regulation of body weight: whereas MC4R regulates food intake and energy expenditure via distinct populations of neurons, MC3R influences feeding efficiency and the partitioning of fuel stores into fat by an unknown mechanism [145, 159, 160]. *In situ* hybridisation analysis of *Jak2* and *Thrβ2* expression in *Sim1*$^{-/-}$ PVN sections indicated a subtle reduction in staining relative to WT levels, however this may simply reflect dual expression of these genes with *Sim1* in PVN/SON cells, which are known to be lost in *Sim1*$^{-/-}$ mice [58, 137]. Further analysis of any other potential targets has not been

performed. Caqueret *et al* also analysed gene expression changes in *Sim1*$^{-/-}$ hypothalamic tissue relative to WT by microarray analysis, in the hope of identifying genes important in patterning of the hypothalamic region during embryonic development [57]. They uncovered a small group of genes that did not overlap with those identified by Liu *et al*, including *Trh*, *Cart* and *Sim2*, expression of which were all reduced in *Sim1*$^{-/-}$ samples relative to WT and were subsequently verified by *in situ* hybridisation.

Reduced *Sim2* expression in *Sim1*$^{-/-}$ hypothalamic tissues has been observed on a few occasions [57, 63]. Whether this reflects genuine upstream regulation of *Sim2* expression by SIM1, and any subsequent effect of this event on PVN/SON formation and/or function, remains to be determined. SIM1 and SIM2 are both important for development of *Trh*- and *Ss*-positive cell populations in the PVN/SON region, which may reflect overlapping gene regulatory pathways between the two proteins [62]. *Brn2*, which is expressed in a partially overlapping manner with *Sim1* in the PVN, is important in the late period of *Crh*-, *Avp*- and *Oxt*-positive cell development and may be dependent on the expression and activity of SIM1 at earlier stages [57, 81, 92, 161].

1.6.2.2. Clues from studies of related bHLH/PAS factors

1.6.2.2.1. SIM

As previously discussed, SIM is a master regulator transcription factor both necessary and sufficient for midline cell formation during *Drosophila* embryogenesis (see **Section 1.3.**). It directly controls the expression of numerous genes involved in midline cell formation, and indirectly affects the expression of many more [33, 35, 38, 86, 88, 162-172]. Several of these genes have roles in such processes as midline cell differentiation and function, cell-cell interactions, axon guidance, and cell migration [173]. These are summarised in **Figure 1.5.**

SIM regulates expression of its direct target genes by interacting with a DNA sequence referred to as the Central Midline Element (CME), a consensus SIM response element originally derived from comparisons of multiple SIM-binding sequences in the regulatory regions of three direct SIM target genes: *slit*, *toll* and *sim* itself ([88] and **Figure 1.6.**). It has since been extrapolated to include the SIM-binding regulatory regions of seven

different genes and contains the core sequence 5' ACGTG 3', although this is sometimes expanded into 5' (G/A)(T/A)ACGTG 3' to represent nucleotide preferences 5' of the basic core sequence [35, 88, 163-165]. In all cases SIM behaves as a transcriptional activator, and only achieves downstream gene repression indirectly, by activating the expression of repressive factors [86].

More detailed knowledge of the molecular events surrounding SIM-mediated gene expression are known for only a few well-characterised genes, including those encoding the repulsive axon guidance molecule Slit and the cell-cell adhesion molecule Roughest [165, 174]. *slit* expression is regulated via binding of a ternary complex including SIM, Drifter, and the Sox HMG domain protein Fish-hook within a 1kb region of DNA located in intron 1 [174]. Fish-hook associates directly with the PAS domain of SIM and the POU domain of Drifter to activate *slit* expression in midline glial cells [174]. *roughest* expression is controlled via binding of the transcription factors SIM, Dmef2, Pointed and Su(H) within 10kb of the transcription start site [165]. However, a lack of detailed characterisation of other genes has precluded the formation of a general model for SIM-mediated gene regulation. In general, it is presumed that specificity of target gene activation occurs via the expression and/or recruitment of midline cell- (and/or gene-) specific cofactors to DNA, in a manner dependent on the presence of additional flanking *cis*-regulatory DNA elements [163, 165, 174]. To date, the possibility of SIM1-mediated regulation of SIM target gene orthologues in mammals has not been explored.

1.6.2.2.2. SIM2

A handful of direct SIM2 target genes have been uncovered in the last few years: *Myomesin 2* (*Myom2*), which encodes a muscle structural protein; *Bcl2/Adenovirus E1B 19kD Protein-Interacting Protein 3* (*Bnip3*); and the zinc finger transcription factor *Slug* [84, 85, 175]. The latter two have been implicated in SIM2s-mediated cancer progression in prostate and breast tissue [61, 175]. SIM2 regulates *Myom2* expression via binding to a single 5' GAACGTG 3' sequence in the upstream proximal promoter region, however the functional consequence of this observation remains undefined [84]. Interestingly, SIM2s represses *Bnip3* expression via binding to a palindromic E-box sequence (5' CACGTG 3') within 200bp of the transcription start site, a motif that has been previously identified as a direct binding site for HIF-1α/ARNT and that mediates upregulation of *Bnip3* expression

Gene name	Identity/function
abrupt	Zinc finger transcription factor.
breathless	Fibroblast growth factor receptor and receptor tyrosine kinase.
center divider	Serine/threonine kinase domain-containing protein.
DER	Epidermal growth factor receptor and receptor tyrosine kinase.
drifter	POU domain-containing transcription factor. *Brn2* homologue.
engrailed	Homeodomain-containing transcription factor. Axon guidance.
jing	Zinc finger transcription factor.
midline fasciclin	Membrane-associated protein. Cell-cell adhesion/axonogenesis.
orthodenticle/ocelliless	Homeodomain-containing transcription factor.
pointed	ETS (Helix-Turn-Helix) domain-containing transcription factor.
rhomboid/veinlet	Transmembrane serine protease. Processes Spitz into receptor-binding form.
roughest	Immunoglobulin domain-containing protein. Cell-cell adhesion.
slit	Ligand for Roundabout receptor. Axon guidance/repulsion.
spitz	Ligand for Epidermal growth factor receptor.
star	Spitz processing.
toll	Transmembrane leucine-rich repeat domain-containing protein. Dorsal-ventral patterning.
vnd	Homeodomain-containing transcription factor. *Nkx2.2* homologue.
wrapper	Membrane-bound Fibronectin and Immunoglobulin domain-containing protein. Neuron-glia adhesion.

Figure 1.5. List of genes whose expression is affected in *sim* mutant embryos and their known or putative function(s). Information on gene identity and function is derived from <http://flybase.org> and NCBI Entrez Gene <http://www.ncbi.nlm.nih.gov/gene>. Further references are in the main text. *vnd*, *ventral nervous system defective*; *DER*, *Drosophila epidermal growth factor receptor*.

wrapper	5′	TCAGAGAACGTGCTGTGTT 3′
roughest	5′	AGAGAAGACGTGCCTGTTG 3′
	5′	AAAGCAGACGTGCGCAGCC 3′
rhomboid/veinlet	5′	TGAGGGAACGTGGGAAAAA 3′
	5′	GCTGATGACGTGCCTGGTC 3′
breathless	5′	GTAGATTACGTGCCCGGCA 3′
	5′	ATCCGTTACGTGACTGATA 3′
	5′	GTGCAGTACGTGGCATGCG 3′
slit	5′	AACTCAAACGTGCCGCAGA 3′
toll	5′	CCTGAGTACGTGTTAATTT 3′
	5′	GAGTCGAACGTGTGTGATA 3′
	5′	AGTTTGTACGTGTGTGTGA 3′
	5′	AATTTGTACGTGCCACAGA 3′
single-minded	5′	ACTTCAGACGTGCATGTTG 3′
	5′	TTAAGATACGTGACCTAGG 3′
	5′	TGTATTTACGTGCGAATTC 3′
	5′	TTTTCGAACGTGATTTTGG 3′
	5′	AGCCAAAACGTGTGCCATG 3′
CONSENSUS	5′	AA GTACGTG 3′

Figure 1.6. Alignment of functional SIM binding motifs from several direct SIM target genes. Regulatory sequences from target genes for which individual CMEs have been mutated and functionally tested *in vivo* are indicated. In each case the core 5′ CGTG 3′ sequence is underlined. The corresponding target gene name is on the left. References are in the main text.

under hypoxic conditions [176]. SIM2s binding to the *Slug* promoter has been observed in breast cancer-derived MCF-7 cells, but more precise delineation of the SIM2s response element within this region has not been performed. It does, however, contain a palindromic E-box sequence, an XRE, and a single 5' ATACGTG 3' within 3kb of the transcription start site [175].

1.7. *SIM1*: a novel contributor to severe obesity in humans

A link between *SIM1* and obesity in humans was first established approximately ten years ago, prior to publication of the *Sim1*$^{+/-}$ mouse phenotype, in a girl possessing a balanced translocation between chromosomes 1p22.1 and 6q16.2 [177]. This first reported case contained a breakpoint in the first intron of SIM1 and resulted in separation of the 5' promoter and bHLH region from the 3' end [177]. There have been several subsequent publications that specifically link obesity with disruption of the chromosomal region encompassing the *SIM1* gene in humans [177-186]. They typically manifest in a severe, early onset obesity phenotype principally characterised by hyperphagia, increased linear growth, and elevated blood insulin and leptin levels, but without any corresponding change in energy expenditure. These features have also been observed in heterozygous *Sim1* mouse models ([59, 96] and **Section 1.4.4.1.**).

1.7.1. The genetic contribution to human obesity

Obesity can be generally described as an abnormal accumulation of adipose tissue. This may arise from a disorder in energy balance, due to excess energy intake and/or reduced expenditure; defects in fat storage and/or mobilisation, leading to an increase in adipocyte size or number; or behavioural defects in the neurological control of appetite [187]. Obesity is consequently defined as a multifactorial disease, and its penetrance can be determined by numerous dietary, lifestyle and genetic factors that combine to confer susceptibility to excess weight gain [188].

At the genetic level, obesity can be conferred by pathogenic mutations in a reasonably large pool of genes [187, 189]. Within the individual, it may be the result of a relatively rare but highly penetrant mutation in a single gene, or the combined effect of several unlinked, less detrimental gene variants. In this case, the exact "combination" of

contributing variants may vary between individuals [189]. The former phenomenon often manifests independently of dietary and lifestyle choices and may only be transmitted within families for a few generations [188]. The latter phenomenon may only manifest under certain environmental conditions (for example, in combination with a sedentary lifestyle in which the individual consumes large quantities of energy-dense foods), and the individual contributing genetic variants are thus more likely to be found at high frequency within the general population [190]. It is hoped that the identification and characterisation of highly penetrant disease-associated gene variants may provide insight into the underlying genetic "risk factors" for common obesity in the wider community [191, 192].

To date, the number of known contributors to severe monogenic obese phenotypes (*i.e.*, severe obesity resulting from a mutation in, or deficiency of, a single gene) in humans is small ([177, 193-208] and **Figure 1.7.**). Of these, many are involved in neurodevelopment and/or signalling pathways in the hypothalamus [188]. The most common contributor is *Mc4r* and mutations in this gene have been calculated to correspond to approximately 5% of all severe childhood obesity cases [209]. This figure, however, is extraordinarily high, and a much lower prevalence is observed for the remaining candidates. Of these, the majority were fortuitously discovered as the result of knockout or mutagenesis studies in rodents, or via linkage studies in those rare cases of severe familial obesity. Very few have been the result of association studies in larger cohorts, making it difficult to ascertain their contribution to obesity in the broader population (**Figure 1.7.**). Single nucleotide polymorphisms (SNPs) in the first intron of the *FTO* gene, for example, have been reproducibly associated with severe human obesity via association studies [210, 211]. However, none of the identified SNPs have yet been proven to affect *FTO* expression, nor is the biological contribution of the *FTO* gene product to obesity well understood [212, 213]. There are also many gene variants that contribute to obesity in rodents, for which corresponding human mutations have so far not been identified ([152, 214-218] and **Figure 1.7.**).

1.7.2. *SIM1* mutations in severely obese humans

The link between *SIM1* haploinsufficiency and severe, early onset obesity has made it an obvious candidate for further screening in human subjects [182, 219-223]. Given the

dramatic phenotype observed in mice upon loss of a single *Sim1* allele (and hence, presumably, 50% SIM1 activity), it is likely that the ability to regulate food intake is sensitive to relatively subtle changes in SIM1 behaviour. These changes may be conferred by deleterious variations in the *SIM1* cds or its regulatory regions, resulting in reduced SIM1 expression and/or activity. To date there are five published studies that have specifically looked for unique *SIM1* sequence variations in obese subjects relative to lean controls; obese subjects were typically required to display severe hyperphagic obesity (with reference to body mass index[1] (BMI) scores and age of onset) and were occasionally recruited on the basis of previous studies suggesting linkage of obesity to chromosome 6q [182, 219-223]. The majority failed to identify any *SIM1* variants that segregated with obesity, although Ahituv *et al* reported six non-synonymous variants within the *SIM1* gene of 379 obese subjects that were not present in 378 lean subjects, without revealing their exact identities [220]. Hung *et al* reported a nominal association of the SIM1 P352T/A371V haplotype with mildly increased BMI in adults, however this was not supported by a later study [221, 222]. Cumulatively these data suggest that, like other known contributors to severe monogenic obesity disorders, pathogenic *SIM1* sequence variants are unlikely to be found at high frequency within the severely obese [189]. However, the identification of any such variants may in turn inform our understanding of the factors that contribute to common obesity within the broader population, thus increasing our knowledge of the genetic basis to what is clearly a complex, multifactorial disease.

[1] Defined as weight in kilograms divided by the square of the height in metres.

Figure 1.7. List of known or suspected contributors to severe monogenic obesity phenotypes in humans and rodents. Further loci statistically linked to severe obesity phenotypes via genome-wide association studies have been identified but are not listed; *FTO* is included because it is the most statistically robust. See [189, 191]. Indicated references are not exhaustive.

Gene name and product	How identified	Name and/or type of mutation	Mutations in this gene identified in humans?
Asip (Agouti)	Mouse phenotype[152, 214]	A^{y} (ectopic Agouti expression)	No – mouse only
BDNF (Brain-Derived Neurotrophic Factor)	Single human subject[193, 194]	Sequence inversion upstream of *BDNF*	Originally identified in humans
Cpe (Carboxypeptidase E)	Mouse phenotype[215]	*fat* (serine→proline substitution)	No – mouse only
FTO (Fat Mass- and Obesity-Associated Gene)	Association study[210, 211]	Single nucleotide change in first intron	Originally identified in humans
Htr2c (5-Hydroxytryptamine Receptor 2C)	Mouse knockout[216]	*Htr2c*$^{-/-}$	No – mouse only
Lep (Leptin)	Mouse phenotype[199]	*obese* (premature stop codon)	Yes[200, 201]
Lepr (Leptin Receptor)	Mouse phenotype[202, 203]	*diabetes* (aberrant splicing)	Yes[204]
Mc4r (Melanocortin 4 Receptor)	Mouse knockout[205]	*Mc4r*$^{-/-}$	Yes[206-208]
Nhlh2 (Nescient Helix Loop Helix 2)	Mouse knockout[217]	*Nhlh2*$^{-/-}$	No – mouse only
NTRK2/TRKB (BDNF Receptor)	Single human subject[195]	tyrosine→cysteine substitution	Originally identified in humans
PCSK1 (Prohormone Convertase 1)	Single human subject[196]	glycine→arginine substitution	Originally identified in humans
POMC (Pro-opiomelanocortin, precursor of αMSH)	Two human subjects[197, 198]	Frameshift mutation Out-of-frame alternative start codon	Originally identified in humans
SIM1 (Single-minded 1)	Single human subject[177]	Chromosome breakpoint in first intron	Originally identified in humans
Tub (Tubby)	Mouse phenotype[218]	*tubby* (aberrant splicing)	No – mouse only

1.8. Project Aims and Approaches

1.8.1. Project Aims

The aims of this project are:

i) To elucidate a link between obesity-associated *SIM1* sequence variants and altered SIM1 activity and/or function *in vivo*, and

ii) To search for target genes of SIM1, given its importance in embryonic development and hypothalamic function.

1.8.2. Project Approaches

This will be examined by:

i) Developing functional assays (including a cell-based luciferase reporter gene assay, a subcellular localisation assay, a co-immunoprecipitation assay, and an electrophoretic mobility shift assay) to assess the impact of nineteen unique point mutations within the *SIM1* sequence on SIM1 expression and activity, which were identified in two separate studies via sequence analysis of the *SIM1* gene in individuals displaying severe, early onset obesity phenotypes,

ii) Studying transcripts that are differentially expressed in response to SIM1 and ARNT2 overexpression, as identified via microarray analysis of RNA samples isolated from human embryonic kidney (HEK) 293-derived cell lines inducibly overexpressing SIM1 and its obligate partner factor ARNT2, and confirming their differential expression in HEK 293T cell lines inducibly expressing shRNA constructs directed against the SIM1 transcript, and

iii) Studying the response of these genes, as well as some additional genes previously implicated in hypothalamic regulation of the feeding response, to SIM1 and/or SIM1 and ARNT2 overexpression and knockdown in embryonic hypothalamus-derived cell lines.

CHAPTER 2: MATERIALS AND METHODS

2.1. Abbreviations

aa	amino acid/s
AdML	Adenovirus major late
AHR	Aryl Hydrocarbon Receptor
αMSH	α-Melanocyte Stimulating Hormone
Amp	Ampicillin
AMP	Adenosine monophosphate
ANOVA	Analysis of variance
APS	Ammonium persulphate
aPV	Anterior periventricular nucleus
BSA	Bovine serum albumin
°C	degrees Celsius
cAMP	cyclic AMP
cDNA	coding deoxyribonucleic acid
cds	coding sequence
CME	Central Midline Element
cpm	counts per minute
ctrl	control
DAPI	4'6-diamidino-2-phenylindole
DMEM	Dulbecco's modified Eagle's medium
DMSO	Dimethylsulphoxide
DNA	Deoxyribonucleic acid
dNTP	Deoxyribonucleotide triphosphate
ds	double stranded
DTT	Dithiothreitol
DVC	Dorsal vagal complex
EDTA	Ethylene diamine tetraacetic acid
EMSA	Electrophoretic mobility shift assay
EtBr	Ethidium bromide
EtOH	Ethanol
FAM	6-carboxyfluorescein
FCS	Foetal calf serum
G418	Geneticin ®
GAPDH	Glyceraldehyde 3-Phosphate Dehydrogenase
GFAP	Glial Fibrillary Acidic Protein
H_6	6-histidine

HEK	Human embryonic kidney
HEPES	4-(2-hydroxyethyl)piperazine-1-ethanesulphonic acid
hr	hour/s
HRE	Hypoxic Response Element
HRP	Horseradish peroxidase
ICV	Intracerebroventricular
Ig	Immunoglobulin
Igepal	Igepal CA-630
IP	Immunoprecipitation
IPTG	Isopropyl-β-D-thiogalactopyranoside
IRES	Internal ribosomal entry site
Kan	Kanamycin
kb	kilobase/s
kDa	kiloDalton/s
L	litre/s
LB	Luria Broth
LHA	Lateral hypothalamic area
m	metre/s
M	molar
mA	milliampere/s
µL	microlitre/s
µm	micrometre/s
µM	micromolar
ME	Median eminence
MeOH	Methanol
min	minute/s
mL	millilitre/s
mm	millimetre/s
mM	millimolar
MOPS	3-(N-morpholino)propanesulphonic acid
MTII	Melanotan II
NES	Nuclear export signal
nM	nanomolar
NLS	Nuclear localisation signal
OD_{600}	Optical density at 600nm
ON	overnight
PAGE	Polyacrylamide gel electrophoresis
PBS	Phosphate buffered saline
PCR	Polymerase chain reaction
PFA	Paraformaldehyde

PGS	Protein G-Sepharose
Poly(dI-dC).poly(dI-dC)	Poly(deoxyinosinic-deoxycytidylic) acid sodium salt
Polybrene	1,5-dimethyl-1,5-diazaundecamethylene polymethobromide
PS	Penicillin/Streptomycin
PVN	Paraventricular nucleus
QPCR	Quantitative PCR
RNA	Ribonucleic acid
RO	Reverse osmosis
rpm	revolutions per minute
RT	Room temperature
RT-PCR	Reverse transcriptase-PCR
SAP	Shrimp alkaline phosphatase
SD	Standard deviation
SDS	Sodium dodecyl sulphate
sec	second/s
SEM	Standard error of the mean
shRNA	short hairpin RNA
SIM	Single-minded
SN	Supernatant
SNP	Single nucleotide polymorphism
SON	Supraoptic nucleus
SV40	Simian Virus 40
TBE	Tris-Borate-EDTA
TE	Tris-EDTA
TetR	Tetracycline- and Doxycycline-Responsive Repressor Protein
TGE	Tris-glycine-ETDA
TEMED	N,N,N^{1},N^{1}-tetramethyl-ethylenediamine
Tris	Tris(hydroxymethyl)aminomethane
Tween-20	Polyoxyethylene-sorbitan monolaurate
U	unit/s
UTR	Untranslated region
UV	Ultraviolet
V	Volt/s
VMN	Ventromedial nucleus
v/v	volume per volume
WCE	Whole cell extract
WT	Wild type
w/v	weight per volume
X-gal	5-Bromo-4-Chloro-3-Indolyl-β-D-Galactopyranoside

2.2. Materials

2.2.1. General Materials

Materials used and their main supplier are listed.

8 well chamber slides, Permanox	Nunc Lab-Tek
Acrodisc ® Syringe Filters, 0.2μm, 0.45μm, 0.8μm	Pall (Gelman Laboratory)
Cover glasses (coverslips), 22x60mm	Knittel Gläser
CryoTubes	Nunc
Kimwipes	Kimberly Clark/KimTech Science
Luer-Lok ® Syringes, 20mL and 50mL	BD Plastipak
Microamp ® Fast 96-well Reaction Plates	Applied Biosystems
Microseal 'B' adhesive seals	Bio-Rad
Mr. Frosty	Nalgene Labware
Multiplate 96-well unskirted PCR plates	Bio-Rad
Nitrocellulose	Pall Corporation
Phosphorimager (Storage Phosphor Screen)	Molecular Dynamics
Plugged (Aerosol Resistant) pipette tips	Molecular BioProducts
ProbeQuant G-50 Micro Columns	GE Healthcare
Surgical Blades, sterile	Swann-Morton
Tissue culture plasticware and disposable pipettes	Falcon, Corning Life Sciences
Whatman (3MM Chromatography) paper	Whatman
X-Ray film	AGFA

2.2.2. Chemicals and Reagents

All chemicals and reagents were molecular biology grade or higher, and were sourced from Sigma-Aldrich except those listed below.

1Kb Plus DNA Ladder	Invitrogen
AccuGel^{TM} 29:1 acrylamide:bis-acrylamide	National Diagnostics
Acetic acid	VWR International Ltd.
Agarose, DNA grade	AppliChem
α-^{32}P-dCTP (3000Ci/mmol, 10mCi/mL)	Perkin Elmer
Bacto-agar	Difco Labs
Blasticidin S HCl	Invitrogen
Bradford Protein Assay Reagent	Bio-Rad
Butanol (Butan-1-ol)	Merck

DAPI	Gibco (Invitrogen)
DMEM	Gibco (Invitrogen)
dNTPs	Finnzymes
DTT	BioVectra
Dual Colour Precision Plus Protein Standards	Bio-Rad
Easy Tag L-^{35}S-Met (>1000Ci/mmol)	Perkin Elmer
Ethanol	Merck
Fast SYBR Green Master Mix	Applied Biosystems
FCS	Bovogen Biologicals
First Strand Buffer, 5x	Invitrogen
Fluorescent Mounting Medium	Dako
Foetal Calf Serum	Gibco (Invitrogen)
Fugene 6	Roche
Glacial acetic acid	Merck
Glycerol	Merck
Glycine (powder)	Amresco
Glycogen	Boehringer Mannheim
High Fidelity Buffer, 5x	Finnzymes
Horse serum, heat inactivated	Gibco (Invitrogen)
Hygromycin B	Invitrogen
Immobilon Western Chemiluminescent HRP Substrate	Millipore
Isopropanol (Propan-2-ol)	Merck
Lipofectamine 2000	Invitrogen
Mark12 Unstained Standard	Invitrogen
Melanotan II [acetyl-(Nle^{4}, Asp^{5}, D-Phe^{7}, Lys^{10})-cyclo-α-MSH (4-10) amide acetate salt]	Bachem
Methanol	Merck
Oligo(dT)$_{15}$ Primer	Geneworks, Promega
Opti-MEM I	Gibco (Invitrogen)
Passive Lysis Buffer, 5x	Promega
Penicillin/Streptomycin [10,000U/mL Penicillin and 10,000µg/mL Streptomycin]	Invitrogen
Pfu Turbo polymerase buffer, 10x	Stratagene
Pyroneg (powder)	Suma
Random Hexamer Primer (RP-6)	Geneworks
Rapid Ligation Buffer, 2x	Promega
rec-Protein G-Sepharose 4B	Zymed Laboratories
Restriction enzyme reaction buffers	New England Biolabs
SDS (powder)	Bio-Rad
Shrimp Alkaline Phosphatase buffer, 10x	Promega

Skim milk powder	Diploma
Sodium hypochlorite, 12.5% (w/v) solution	Ajax Finechem
SUPERase-In RNAse Inhibitor	Ambion
T4 DNA Ligase Buffer, 10x	Geneworks
Thermo Pol Buffer, 10x	Geneworks
TRI Reagent	Ambion
Tris (powder)	Amresco
Trypsin-EDTA	Gibco (Invitrogen)
X-gal	BioVectra
Zeocin™	Invitrogen

2.2.3. Kits

Commercially available kits used and their supplier are listed.

Dual Luciferase Assay System	Promega
pGEM-T Easy Vector System	Promega
Plasmid Midi Kit	QIAGEN
QIAprep Spin Miniprep Kit	QIAGEN
QIAquick Gel Extraction Kit	QIAGEN
RNeasy Mini Kit	QIAGEN
TNT ® Quick Coupled Transcription/Translation System	Promega

2.2.4. Enzymes

Restriction endonucleases were supplied by New England Biolabs. Other enzymes were obtained from the following sources:

BIG DYE™ Version 3.1	Invitrogen
Gateway ® BP Clonase Enzyme Mix	Invitrogen
Gateway ® LR Clonase Enzyme Mix	Invitrogen
Klenow Large Fragment	New England Biolabs
Pfu Turbo polymerase	Stratagene
Phusion	Finnzymes
Shrimp Alkaline Phosphatase	Promega
SuperScript III Reverse Transcriptase	Invitrogen
T4 DNA Ligase	Geneworks
T4 Polynucleotide Kinase	Geneworks
Taq polymerase	New England Biolabs

2.2.5. Antibodies

2.2.5.1. Primary antibodies

4A6 αMyc	purified mouse monoclonal IgG_1 raised against aa 410-420 of hMYC (Upstate).
ab9106 αMyc	purified rabbit polyclonal IgG raised against the c-MYC epitope tag (Abcam).
M-165 αARNT2	purified rabbit polyclonal IgG raised against aa 457-622 of hARNT2 (Santa Cruz Biotechnology, Inc.).
MCA78G αalpha-tubulin	purified rat monoclonal IgG_{2a} raised against yeast Tubulin (AbD Serotec).
M2 αFlag	purified mouse monoclonal IgG_1 raised against the Flag epitope (Sigma-Aldrich).
222-6 αU2AF65	purified mouse monoclonal IgG_{2a} raised against a peptide derived from hU2AF65 (Zymed Laboratories).
MW6 αSIM1	purified rabbit polyclonal IgG raised against aa 550-649 of mSIM1 (Whitelaw lab).
H-300 αMEF-2	purified rabbit polyclonal IgG raised against aa 1-300 of hMEF-2 (Santa Cruz Biotechnology, Inc.). Isotype match antibody for M-165 αARNT2 antibody.
PG-M3 αPML	purified mouse monoclonal IgG_1 raised against aa 37-51 of hPML (Santa Cruz Biotechnology, Inc.). Isotype match antibody for 4A6 αMyc antibody.
Z0344 αGFAP	purified rabbit polyclonal IgG raised against GFAP isolated from cow spinal cord (Dako).
25/NESTIN αNestin	purified mouse monoclonal IgG_1 raised against aa 402-604 of rNestin (BD Transduction).
NS-1 αRIP	mouse monoclonal ascites fluid generated by immunising mice with rat olfactory bulb extracts (StemCell Technologies).

2.2.5.2. Secondary antibodies

Immunopure ® Goat Anti-Mouse IgG, (H+L), Peroxidase Conjugated	Thermo Scientific
Immunopure ® Goat Anti-Rabbit IgG, (H+L), Peroxidase Conjugated	Thermo Scientific
Polyclonal Rabbit Anti-Rat Immunoglobulins/HRP	Dako
Rabbit IgG SmartBlot-HRP	Vicgene Biotechnology
Alexa Fluor ® 594 Donkey Anti-Rabbit IgG (H+L)	Invitrogen
Alexa Fluor ® 594 Donkey Anti-Mouse IgG (H+L)	Invitrogen

2.2.6. Bacterial strains

The *E. coli* DH5α strain was used for DNA propagation and subcloning [Φ80*lac*ZΔM15, *rec*A1, *end*A1, *gyr*A96, *thi*-1, *hsd*R17(r_K^-, m_K^+), *sup*E44, *rel*A1, *deo*R, Δ(*lac*ZYA-*arg* F)U169].

2.2.7. Tissue culture cell lines

293 Flp-In™ T-Rex™

HEK cell line, containing a single integrated Flp Recombination Target (FRT) site from pFRT/*lac*Zeo, and multiple integrated copies of the TetR-expressing vector pcDNA6/TR ©. These enable tetracycline- and doxycycline-inducible expression of a gene of interest from a single genomic location via Flp recombinase-mediated integration of cDNA encoding the gene of interest into the FRT site. Purchased from Invitrogen.

N4, N7, N39

Immortalised murine embryonic neuronal cell lines, derived from E15.0, 17.0 or 18.0 hypothalamic primary cultures by retroviral transfer of the SV40 large T antigen and subsequent selection with G418 [224]. Purchased from CELLutions Biosystems Inc.

293T

HEK cell line constitutively expressing the SV40 large T antigen. Kindly donated by Dr. K. Jensen.

293A

HEK cell line constitutively expressing the SV40 large A antigen. Obtained from Biochemistry Department shared stocks.

2.2.8. Solutions

Unless otherwise stated, all solutions were made up in sterile MQ H_2O.

2.2.8.1. General lab solutions

Coomassie Blue stain	0.03% (w/v) Coomassie Blue R250, 50% (v/v) MeOH, 8.75% (v/v) acetic acid
Coomassie Destain I	50% (v/v) MeOH, 8.75% (v/v) acetic acid
Coomassie Destain II	5% (v/v) MeOH, 7% (v/v) acetic acid

Detergent-free hypotonic buffer	10mM HEPES pH 8.0, 1.5mM $MgCl_2$, 10mM KCl, 10% (v/v) glycerol with DTT and Protease Inhibitor Cocktail added fresh to final concentrations of 1mM and 1x respectively
DNA load buffer, 6x	1mM EDTA pH 8.0, 0.25% (w/v) Bromophenol Blue, 50% (v/v) glycerol
Fixing Solution	50% (v/v) MeOH, 10% (v/v) glacial acetic acid
Gel Drying Solution	5% (v/v) glycerol, 35% (v/v) crude EtOH
Gel Shift Buffer, 4x	40mM Tris pH 7.5, 12mM $MgCl_2$, 16mM spermidine, 2mM DTT, 40% (v/v) glycerol
GTS running buffer, 10x	14.4% (w/v) glycine, 3% (w/v) Tris, 1% (w/v) SDS, corrected to pH 8.3 with 1M HCl
Immunofluorescence Blocking Solution	10% (v/v) horse serum in PBST
IP buffer	250mM NaCl, 20mM HEPES pH 8.0, 0.1% Igepal, 1mM EDTA pH 8.0 with Protease Inhibitor Cocktail added fresh to a final concentration of 1x
Lower Buffer, 4x	1.5M Tris base, 0.4% (w/v) SDS corrected to pH 8.8 with 1M HCl and filtered 0.45µm
Nuclear extract buffer	20mM HEPES pH 8.0, 1.5mM $MgCl_2$, 0.5mM EDTA pH 8.0, 20% (v/v) glycerol, 0.42M KCl with DTT and Protease Inhibitor Cocktail added fresh to final concentrations of 1mM and 1x respectively
PBS, 1x	0.8% (w/v) NaCl, 0.02% (w/v) KCl, 0.4% (w/v) KH_2PO_4 (anhydrous), 2.3% (w/v) Na_2HPO_4 (anhydrous)
PBST	0.1% (v/v) Tween-20 in PBS
Ponceau stain	0.5% (w/v) Ponceau S, 1% (v/v) glacial acetic acid
Protein Storage buffer	50mM Tris pH 8.0, 0.1mM EDTA pH 8.0, 0.2mM DTT, 100mM NaCl, 5% (v/v) glycerol
Reaction Buffer, 10x	250mM Tris pH 9.0, 160mM $(NH_4)_2SO_4$
Sequencing Buffer, 5x	0.2M Tris pH 9.0, 5mM $MgCl_2$
STE buffer	10uL 4M NaCl added to 210uL 1x TE
TBE buffer, 20x	21.6% (w/v) Tris, 11% (w/v) boric acid, 1.86% (w/v) EDTA corrected to pH 8.3 with 1M HCl
TE, 100x	0.1M EDTA, 1M Tris
TEN buffer	40mM Tris pH 7.5, 10mM EDTA pH 8.0, 150mM NaCl
TFB1	30mM KAc, 100mM RbCl, 10mM $CaCl_2(2H_2O)$, 50mM $MnCl_2(4H_2O)$, 15% (v/v) glycerol corrected to pH 5.8 with 0.2M acetic acid
TFB2	10mM MOPS, 10mM RbCl, 75mM $CaCl_2(2H_2O)$, 15% glycerol corrected to pH 6.5 with 0.1M KOH
TGE buffer, 5x	125mM Tris, 0.95M glycine, 0.5mM EDTA pH 8.0
TM buffer, 10x	100mM Tris pH 7.5, 50mM $MgCl_2$

Tris-glycine-SDS load buffer, 4x	40% (v/v) glycerol, 10% (v/v) 4x Upper Buffer, 5% (w/v) SDS, 0.01% (w/v) Bromophenol Blue, 0.08M DTT
Tris-tricine-SDS load buffer, 2x	0.12M Tris pH 8.3, 0.06% (w/v) SDS, 45% (v/v) glycerol, 0.04% (w/v) Bromophenol Blue with DTT added fresh to a final concentration of 0.05M
Tris/tricine running buffer	1.21% (w/v) Tris, 1.79% (w/v) tricine, 0.1% (w/v) SDS
Upper Buffer, 4x	0.5M Tris base, 0.4% (w/v) SDS corrected to pH 6.8 with 1M HCl and filtered 0.45µm
WCE buffer	20mM HEPES pH 8.0, 0.42M NaCl, 0.5% (v/v) Igepal, 25% (v/v) glycerol, 0.2mM EDTA pH 8.0, 1.5mM $MgCl_2$ with DTT and Protease Inhibitor Cocktail added fresh to final concentrations of 1mM and 1x respectively
Western Blocking Solution	10% (w/v) skim milk powder in PBST
Wet Transfer Buffer	50mM Tris, 400mM glycine

2.2.8.2. Bacterial growth media

2.2.8.2.1. Basic media

Liquid:

LB	2.5% (w/v) Luria Broth Base
SOC	20mg/mL Tryptone (Oxoid), 5mg/mL Yeast Extract (Oxoid), 10mM NaCl, 2.5mM KCl, 10mM $MgCl_2$, 10mM $MgSO_4$, with 0.4% (v/v) glucose added after autoclaving

Solid:

LB + 1.5% (w/v) bacteriological- (bacto-) agar

2.2.8.2.2. Final concentrations of antibiotics and other additives

In liquid media:

Ampicillin: 200µg/mL

Carbenicillin: 100µg/mL

In solid media:

Ampicillin: 100µg/mL

Kanamycin: 50µg/mL

Chloramphenicol: 30µg/mL

Occasionally, LB+agar plates were spread with the following and left to dry prior to plating bacteria:

Ampicillin: 100µL of 25mg/mL stock

IPTG: 100µL of 100mM stock (pGEM-T Easy cloning)

X-gal: 20µL of 50mg/mL stock (pGEM-T Easy cloning)

2.2.8.3. Tissue culture solutions

2.2.8.3.1. Basic media

Complete media used for cell lines:

293T and 293 Flp-In™ T-Rex™	DMEM 12430 + 10% (v/v) FCS + 1% (v/v) PS
N4, N7, N39	DMEM 11995 + 10% (v/v) FCS + 1% (v/v) PS
Freezing Medium	10% (v/v) DMSO in FCS

2.2.8.3.2. Final concentrations of antibiotics

293 Flp-In™ T-Rex™ parent (untransformed) cell line: 100µg/mL Zeocin™, 15µg/mL blasticidin

293 Flp-In™ T-Rex™ stable cell lines: 200µg/mL hygromycin B, 15µg/mL blasticidin

N4 6-TR and N7 6-TR cell lines: 2µg/mL blasticidin

N4 6-TR/mSIM1-Flag cell lines: 2µg/mL blasticidin, 200µg/mL Zeocin™

N7 6-TR/mSIM1-Flag cell lines: 2µg/mL blasticidin, 100µg/mL Zeocin™

N39 empty/empty, empty/mARNT2 and mSIM1-Flag/mARNT2 cell lines: 2µg/mL blasticidin (selecting for LV501 vector only)

2.2.9. Plasmids

Unless explicitly labelled "mSIM1FL", all mSIM1 vectors express a truncated version of mSIM1 comprising aa 1-696 of mSIM1 only, followed by a contiguous 2xMyc tag. All hSIM1 vectors express full-length hSIM1.

2.2.9.1. Cloning vectors

pCIN4

aka. pIRES1neo [225]. Genbank accession number U89673.

pBSK-TH1

Kindly donated by Dr. S. Barry, Women's & Children's Health Research Institute, 72 King William Rd,

North Adelaide, SA 5006.

pDR2-hSIM1

Kindly donated by Dr. S. Antonarakis, Laboratory of Human Molecular Genetics, Department of Genetics and Microbiology, Geneva University Medical School, 1211 Geneva, Switzerland [45]. Expresses full-length hSIM1 (Genbank accession number U70212) cloned into BamHI/XbaI site of pDR2 (Genbank accession number U02428).

pBKS(+)-mSIM1

Kindly donated by Dr. C.M. Fan, Department of Embryology, Carnegie Institution of Washington, 3520 San Martin Drive, Baltimore, Maryland 21218, USA. Expresses full-length mSIM1 (Genbank accession number U40575) cloned into the EcoRI site of pBKS(+).

pGEM-T Easy

Purchased from Promega.

pGEM-T Easy hSIM1 (Myc)

Full-length hSIM1 was PCR amplified from pEF-hSIM1 no tag-IRES-puro with primers puro sense and hSIM1-rev, A-tailed and ligated into pGEM-T Easy.

pGEM-T Easy hSIM1 (Myc) T714A

The EcoRV/HpaI fragment was digested out of pGEM-T Easy hSIM1 T714A and ligated into similarly digested pGEM-T Easy hSIM1 (Myc).

pGEM-T Easy hSIM1 T714A

hSIM1 T714A was PCR amplified from pcDNA3-hSIM1 T714A using primers CMV forward and hSIM1-rev, A-tailed and ligated into pGEM-T Easy. This vector also contains at least one other amino acid substitution (A450V), probably due to errors incorporated during PCR.

pGEM-T Easy mSIM1 (CFlag)

Full-length mSIM1 was PCR amplified from pBSK(+)-mSIM1 using primers T7 promoter and mSIM1 full-lengthrev, A-tailed and ligated into pGEM-T Easy.

2.2.9.2. Expression vectors

pEF-IRES-puro

Hobbs *et al* [226]. Genbank accession number Z75185.1

pEF-IRES-neo

Hobbs *et al* [226]. Genbank accession number Y11035.1

pEF-mSIM1-Myc-IRES-puro

S. Woods. Does not express full-length mSIM1 (see above).

pEF-mSIM2L-Myc-IRES-puro

S. Woods

pEF-hARNT-Ha-IRES-neo

M. Kleman. Expresses hARNT transcript variant 3 (Genbank accession number NM_178427.2).

pEF-mARNT2-IRES-neo

J. White

pGEM7Zf(+)-hARNT

M. Whitelaw

pcDNA3-hSIM1 H323Y
pcDNA3-hSIM1 R581G
pcDNA3-hSIM1 T714A

Kindly donated by Dr. F. Stutzmann, Institut de Biologie de Lille, Institut Pasteur de Lille, 1 Rue du Professor Calmette, 59019 Lille, France.

pEF-hARNT2-IRES-neo

Full-length hARNT2 was PCR amplified from 293T cDNA using primers MluI-hARNT25'UTR and hARNT2stop-XbaI, digested with MluI/XbaI and ligated into similarly digested pEF-IRES-neo.

pEF-IRES-mSIM1FL-Myc-IRES-puro

Created via three-way ligation of the XbaI/XhoI fragment from pBKS(+)-mSIM1 (mSIM1 N-terminal fragment), the PCR fragment amplified from pBKS(+)-mSIM1 using primers mSIM1 sense and mSIM1 full-lengthrev and then digested with XhoI/HpaI (mSIM1 C-terminal fragment), and NheI/EcoRV digested pEF-mSIM2L-Myc-IRES-puro. This removed mSIM2L from the vector backbone and ligated mSIM1FL in frame with the existing 2xMyc tag.

pEF-CFlag-mSIM1FL-IRES-puro

Full-length mSIM1 was digested out of pGEM-T Easy mSIM1 (CFlag) with XbaI/HpaI and ligated into NheI/EcoRV digested pEF-His-Myc-hSIM2s-Flag-puro. This removed His-Myc-hSIM2s from the vector backbone and ligated mSIM1FL in frame with the existing C-terminal 3xFlag tag.

pGEM7Zf(+)-hSIM1-Myc

Full-length hSIM1-Myc was digested out of pEF-hSIM1-Myc-IRES-puro via sequential digestion with NotI, blunt-ending with Klenow, and digestion with NheI. It was then ligated into the recipient pGEM7Zf(+) backbone, which was generated from pGEM7Zf(+)-hARNT via sequential digestion with HindIII to remove the hARNT insert, blunt-ending with Klenow, and digestion with XbaI.

pEF-hSIM1-Myc-IRES-puro

Full-length hSIM1 was digested out of pGEM-T Easy hSIM1 (Myc) with ClaI/HpaI and ligated into ClaI/EcoRV digested pEF-mSIM2L-Myc-IRES-puro. This ligated hSIM1 in-frame with the existing 2xMyc tag.

pEF-hSIM1 no tag-IRES-puro

Full-length hSIM1 was digested out of pEF-hSIM1-IRES-puro with AflII, blunt-ended with Klenow and ligated into EcoRV digested pEF-IRES-puro. Primarily designed to decrease the length of the 5'UTR sequence between the promoter and initiator ATG cf. pEF-hSIM1-IRES-puro.

pEF-hSIM1-IRES-puro

Full-length hSIM1 was digested out of pDR2-hSIM1 using EcoRI/NdeI and ligated into similarly digested pEF-IRES-puro.

pEF-hSIM1 K51N-Myc-IRES-puro
pEF-hSIM1 S71R-Myc-IRES-puro
pEF-hSIM1 I128T-Myc-IRES-puro
pEF-hSIM1 Q152E-Myc-IRES-puro
pEF-hSIM1 R171H-Myc-IRES-puro
pEF-hSIM1 L238R-Myc-IRES-puro
pEF-hSIM1 T292A-Myc-IRES-puro
pEF-hSIM1 T361I-Myc-IRES-puro
pEF-hSIM1 R383G-Myc-IRES-puro
pEF-hSIM1 S541L-Myc-IRES-puro
pEF-hSIM1 P692L-Myc-IRES-puro
pEF-hSIM1 R703Q-Myc-IRES-puro
pEF-hSIM1 D707H-Myc-IRES-puro
pEF-hSIM1 T712I-Myc-IRES-puro

Full-length hSIM1-Myc containing one of each of the indicated mutations was digested out of its respective pcDNA5-FRT-TO-hSIM1-Myc mutant vector (e.g. pcDNA5-FRT-TO-hSIM1 K51N-Myc) with NheI/NotI and ligated into similarly digested pEF-hSIM1-Myc-IRES-puro.

pEF-hSIM1 P497R-Myc-IRES-puro
pEF-hSIM1 R550H-Myc-IRES-puro

Generated via overlap extension PCR. The plasmids used as templates for each of the two first round amplification reactions are indicated along with the corresponding primers used, followed by a brief explanation of the subsequent subcloning strategy.

<u>P497R:</u>

pEF-hSIM1-Myc-IRES-puro	SIM1-332F, hSIM1 P497R-R longer
pEF-hSIM1-Myc-IRES-puro	hSIM1 P497R-F, hSIM1-rev

Full-length second round PCR product was digested with BclI/PshAI and ligated into similarly digested pEF-hSIM1-Myc-IRES-puro.

<u>R550H:</u>

pEF-hSIM1-Myc-IRES-puro	SIM1-332F, hSIM1 R550H-R
pEF-hSIM1-Myc-IRES-puro	hSIM1 R550H-F, hSIM1-rev

Full-length second round PCR product was digested with EcoRV/PshAI and ligated into similarly digested pEF-hSIM1-Myc-IRES-puro.

pEF-hSIM1 R581G-Myc-IRES-puro

The EcoRV/PshAI fragment was digested out of pcDNA3-hSIM1 R581G and ligated into similarly digested pEF-hSIM1-Myc-IRES-puro.

pEF-hSIM1 T714A-Myc-IRES-puro

Full-length hSIM1 T714A was digested out of pGEM-T Easy hSIM1 (Myc) T714A using ClaI/HpaI and ligated into ClaI/EcoRV digested pEF-mSIM2L-Myc-IRES-puro. This ligated hSIM1 T714A in-frame with the existing 2xMyc tag.

2.2.9.2.1. Flp-In™ T-Rex™ vectors

pcDNA5-FRT-TO

Purchased from Invitrogen.

pOG44

Purchased from Invitrogen.

pcDNA5-FRT-TO-hSIM2s-Myc

A. Farrall

pcDNA5-FRT-TO-mARNT2

The mARNT2 cds was digested out of pEF-mARNT2-IRES-neo with XhoI/SalI and ligated into XhoI digested pcDNA5-FRT-TO.

pcDNA5-FRT-TO-mSIM1-Myc-mARNT2

mSIM1-Myc was digested out of pEF-mSIM1-Myc-IRES-puro with BamHI/NotI and ligated into similarly digested pcDNA5-FRT-TO-mARNT2. Does not express full-length mSIM1 (see above).

pcDNA5-FRT-TO-mSIM1-Myc-IRES-mARNT2

IRES sequence was digested out of pCIN4 with MscI/SmaI and ligated into NotI digested and Klenow treated pcDNA5-FRT-TO-mSIM1-Myc-mARNT2. Does not express full-length mSIM1 (see above).

pcDNA5-FRT-TO-mSIM1FL-Myc-mARNT2

mSIM1FL-Myc was digested out of pEF-IRES-mSIM1FL-Myc-IRES-puro with BamHI/NotI and ligated into similarly digested pcDNA5-FRT-TO-mARNT2.

pcDNA5-FRT-TO-mSIM1FL-Myc-IRES-mARNT2

IRES sequence was PCR amplified from pCIN4 using primers pCIN4 IRES for and pCIN4 IRES rev, digested with NotI and ligated into similarly digested pcDNA5-FRT-TO-mSIM1FL-Myc-mARNT2. Two ARNT2-like proteins of different molecular weights are expressed from this vector.

pcDNA5-FRT-TO-mSIM1FL-Myc-IRESfixed-mARNT2

The NotI/NarI fragment from pcDNA5-FRT-TO-mSIM1-Myc-IRES-mARNT2 was amplified via PCR using primers 5'flankIRESFRTS1T-F and 3'flankIRESFRTS1T-R, digested with NotI/NarI and ligated into similarly digested pcDNA5-FRT-TO-mSIM1FL-Myc-IRES-mARNT2. This effectively replaced the existing IRES and its flanking sequences with those from the source vector.

pcDNA5-FRT-TO-hSIM1-Myc

hSIM1-Myc was digested out of pEF-IRES-hSIM1-Myc-puro with NheI/NotI and ligated into similarly digested pcDNA5-FRT-TO-hSIM2s-Myc.

pcDNA5-FRT-TO-hSIM1 K51N-Myc
pcDNA5-FRT-TO-hSIM1 S71R-Myc
pcDNA5-FRT-TO-hSIM1 I128T-Myc
pcDNA5-FRT-TO-hSIM1 Q152E-Myc
pcDNA5-FRT-TO-hSIM1 R171H-Myc
pcDNA5-FRT-TO-hSIM1 L238R-Myc
pcDNA5-FRT-TO-hSIM1 T292A-Myc
pcDNA5-FRT-TO-hSIM1 T361I-Myc

pcDNA5-FRT-TO-hSIM1 R383G-Myc
pcDNA5-FRT-TO-hSIM1 S541L-Myc
pcDNA5-FRT-TO-hSIM1 P692L-Myc
pcDNA5-FRT-TO-hSIM1 R703Q-Myc
pcDNA5-FRT-TO-hSIM1 D707H-Myc
pcDNA5-FRT-TO-hSIM1 T712I-Myc

Generated via overlap extension PCR. The plasmids used as templates for each of the two first round amplification reactions are indicated along with the corresponding primers used, followed by a brief explanation of the subsequent subcloning strategy.

K51N:

pEF-hSIM1-Myc-IRES-puro	puro sense, hSIM1 K51N-R
pEF-hSIM1-Myc-IRES-puro	hSIM1 K51N-F, SIM1-297R

Full-length second round PCR product was digested with NheI/BssHII and ligated into similarly digested pcDNA5-FRT-TO-hSIM1-Myc.

S71R:

pEF-hSIM1-Myc-IRES-puro	puro sense, hSIM1 S71R-R
pEF-hSIM1-Myc-IRES-puro	hSIM1 S71R-F, SIM1-297R

Full-length second round PCR product was digested with NheI/BssHII and ligated into similarly digested pcDNA5-FRT-TO-hSIM1-Myc.

I128T:

pEF-hSIM1-Myc-IRES-puro	puro sense, hSIM1 I128T-R
pEF-hSIM1-Myc-IRES-puro	hSIM1 I128T-F, SIM1-334C

Full-length second round PCR product was digested with NheI/BssHII and ligated into similarly digested pcDNA5-FRT-TO-hSIM1-Myc.

Q152E:

pcDNA5-FRT-TO-hSIM1-Myc	CMV forward, hSIM1 Q152E-R longer
pcDNA5-FRT-TO-hSIM1-Myc	hSIM1 Q152E-F, SIM1-334C

Full-length second round PCR product was digested with NheI/BssHII and ligated into similarly digested pcDNA5-FRT-TO-hSIM1-Myc.

R171H:

pEF-hSIM1-Myc-IRES-puro	puro sense, hSIM1 R171H-R
pEF-hSIM1-Myc-IRES-puro	hSIM1 R171H-F, SIM1-334C

Full-length second round PCR product was digested with NheI/BssHII and ligated into similarly digested pcDNA5-FRT-TO-hSIM1-Myc.

L238R:

pEF-hSIM1-Myc-IRES-puro	puro sense, hSIM1 L238R-R
pEF-hSIM1-Myc-IRES-puro	hSIM1 L238R-F, SIM1-297R

Full-length second round PCR product was digested with NheI/BssHII and ligated into similarly digested pcDNA5-FRT-TO-hSIM1-Myc.

T292A:

pEF-hSIM1-Myc-IRES-puro	hSIM1-264F, hSIM1 T292A-R
pEF-hSIM1-Myc-IRES-puro	hSIM1 T292A-F, hSim1 antisense

Full-length second round PCR product was digested with BssHII/BamHI and ligated into similarly digested pcDNA5-FRT-TO-hSIM1-Myc.

T361I:

pEF-hSIM1-Myc-IRES-puro	hSIM1-264F, hSIM1 T361I-R
pEF-hSIM1-Myc-IRES-puro	hSIM1 T361I-F, hSim1 antisense

Full-length second round PCR product was digested with BssHII/BamHI and ligated into similarly digested pcDNA5-FRT-TO-hSIM1-Myc.

R383G:

pEF-hSIM1-Myc-IRES-puro	hSIM1-264F, hSIM1 R383G-R
pEF-hSIM1-Myc-IRES-puro	hSIM1 R383G-F, hSim1 antisense

Full-length second round PCR product was digested with BssHII/BamHI and ligated into similarly digested pcDNA5-FRT-TO-hSIM1-Myc.

S541L:

pEF-hSIM1-Myc-IRES-puro	SIM1-332F, hSIM1 S541L-R
pEF-hSIM1-Myc-IRES-puro	hSIM1 S541L-F, SIM1 QPCR 627R

Full-length second round PCR product was digested with BamHI/EcoRV and ligated into similarly digested pcDNA5-FRT-TO-hSIM1-Myc.

P692L:

pcDNA5-FRT-TO-hSIM1-Myc	QPCRSim1-F, hSIM1 P692L-R
pcDNA5-FRT-TO-hSIM1-Myc	hSIM1 P692L-F, BGH reverse

Full-length second round PCR product was digested with EcoRV/NotI and ligated into similarly digested pcDNA5-FRT-TO-hSIM1-Myc.

R703Q:

pcDNA5-FRT-TO-hSIM1-Myc	QPCRSim1-F, hSIM1 R703Q-R
pcDNA5-FRT-TO-hSIM1-Myc	hSIM1 R703Q-F, BGH reverse

Full-length second round PCR product was digested with EcoRV/NotI and ligated into similarly digested pcDNA5-FRT-TO-hSIM1-Myc.

D707H:

pcDNA5-FRT-TO-hSIM1-Myc	QPCRSim1-F, hSIM1 D707H-R
pcDNA5-FRT-TO-hSIM1-Myc	hSIM1 D707H-F, BGH reverse

Full-length second round PCR product was digested with EcoRV/NotI and ligated into similarly digested pcDNA5-FRT-TO-hSIM1-Myc.

T712I:

pEF-hSIM1-Myc-IRES-puro	QPCRSim1-F, hSIM1 T712I-R
pcDNA5-FRT-TO-hSIM1-Myc	hSIM1 T712I-F, BGH reverse

Full-length second round PCR product was digested with EcoRV/NotI and ligated into similarly digested pcDNA5-FRT-TO-hSIM1-Myc.

pcDNA5-FRT-TO-hSIM1 H323Y-Myc

The NheI/BamHI fragment was digested out of pcDNA3-hSIM1 H323Y and ligated into similarly digested pcDNA5-FRT-TO-hSIM1-Myc.

pcDNA5-FRT-TO-hSIM1 P497R-Myc
pcDNA5-FRT-TO-hSIM1 R550H-Myc
pcDNA5-FRT-TO-hSIM1 R581G-Myc
pcDNA5-FRT-TO-hSIM1 T714A-Myc

Full-length hSIM1-Myc containing one of each of the indicated mutations was digested out of its respective pEF-hSIM1-Myc-IRES-puro mutant vector (e.g. pEF-hSIM1 P497R-Myc-IRES-puro) with NheI/NotI and ligated into similarly digested pcDNA5-FRT-TO.

2.2.9.3. pSUPER vectors

pSUPER

Purchased from OligoEngine.

pSUPER scrambled

A. Farrall

pSUPER si1200

Oligos hSIM1si1200F and hSIM1si1200R were phosphorylated, annealed and ligated into BglII/HindIII digested pSUPER.

pSUPER si1381

Oligos hSIM1si1381F and hSIM1si1381R were phosphorylated, annealed and ligated into BglII/HindIII digested pSUPER.

pSUPER si2227

Oligos hSIM1si2227F and hSIM1si2227R were phosphorylated, annealed and ligated into BglII/HindIII digested pSUPER.

pSUPER si2378

Oligos hSIM1si2378F and hSIM1si2378R were phosphorylated, annealed and ligated into BglII/HindIII digested pSUPER.

2.2.9.4. Luciferase reporter vectors

pML-6CWT

Firefly luciferase plasmid that contains two consecutive 3xCME-containing sequences from the *toll* promoter upstream of a minimal AdML promoter [13]. Kindly donated by Dr. J. Pelletier, Department of Biochemistry, RM811, McIntyre Medical Sciences Building, 3655 Promenade Sir William Osler, McGill University, Montreal, Quebec, Canada H3G 1Y6.

phRL-CMV

Renilla luciferase internal control plasmid with CMV promoter (Promega).

2.2.9.5. Lentiviral vectors

2.2.9.5.1. Gateway ® vectors

2.2.9.5.1.1. Donor and entry vectors

pDONR201

Purchased from Invitrogen.

pDONR201 si1200
pDONR201 si1381
pDONR201 si2227

Generated via overlap extension PCR. Of the two first round reactions, only one required PCR amplification. The plasmid used as a template for this reaction is indicated along with the corresponding primers used. The second first round reaction required only annealing and end-filling of the two oligos indicated. The full-length second round PCR product was recombined into pDONR201 using BP Clonase (see **2.4.2.8.4.1.**).

si1200:

pBSK-TH1	AttB1F, LVsi1200no1R
(annealing and end-filling only)	LVsi1200no2F, AttB2R

si381:

pBSK-TH1	AttB1F, LVsi1381no1R
(annealing and end-filling only)	LVsi1381no2F, AttB2R

si2227:

pBSK-TH1	AttB1F, LVsi2227no1R
(annealing and end-filling only)	LVsi2227no2F, AttB2R

pENTR1A

Kindly donated by Dr. S. Gregory.

pENTR1A-CFlag-mSIM1FL

Full-length mSIM1-Flag was digested out of pEF-CFlag-mSIM1FL-IRES-puro with BamHI/SpeI, blunt-ended with Klenow and ligated into BamHI/NotI digested and Klenow treated pENTR1A.

pENTR1A-mARNT2

Full-length mARNT2 was digested out of pEF-mARNT2-IRES-neo with XhoI/SmaI and ligated into SalI/EcoRV digested pENTR1A.

2.2.9.5.1.2. Destination and expression vectors

pLV416neo

Kindly donated by Dr. S. Barry, Women's & Children's Health Research Institute, 72 King William Rd, North Adelaide, SA 5006.

pLV501

D. Bersten

pLV711G

Kindly donated by Dr. S. Barry, Women's & Children's Health Research Institute, 72 King William Rd, North Adelaide, SA 5006.

pLV416neo ΔccdB

N. Hao

pLV416neo-CFlag-mSIM1FL

Generated via Gateway recombination (see **2.4.2.8.4.**) between pLV416neo and pENTR1A-CFlag-mSIM1FL.

pLV501 ΔccdB

D. Bersten

pLV501-mARNT2

Generated via Gateway recombination (see **2.4.2.8.4.**) between pLV501 and pENTR1A-mARNT2.

pLV711G scrambled

D. Bersten

pLV711G si1200

Generated via Gateway recombination (see **2.4.2.8.4.**) between pLV711G and pDONR201 si1200.

pLV711G si1381

Generated via Gateway recombination (see **2.4.2.8.4.**) between pLV711G and pDONR201 si1381.

pLV711G si2227

Generated via Gateway recombination (see **2.4.2.8.4.**) between pLV711G and pDONR201 si2227.

pLenti6-TR

Purchased from Invitrogen.

pLenti4-CFlag-mSIM1FL

C. Bindloss

2.2.9.5.2. Lentivirus production vectors

pCMV-dvrp 8.2
pRSV-Rev
pMD2.G (VSV-G)
Kindly donated by Dr. S. Barry, Women's & Children's Health Research Institute, 72 King William Rd, North Adelaide, SA 5006.

2.2.10. Primers

Primers were synthesised by Geneworks, SA, Australia.

2.2.10.1. Sequencing/cloning primers

CMV forward	5' CGC AAA TGG GCG GTA GGC GTG 3'
BGH reverse	5' TAG AAG GCA CAG TCG AGG 3'
puro vector sense	5' AGT TCA ATT ACA GCT CTT 3'
pEF IRES puro rvs	5' GCG GAA TTG GGC TAG AGC 3'
T3 promoter	5' GCA ATT AAC CCT CAC TAA AGG 3'
T7 promoter	5' TAA TAC GAC TCA CTA TAG G 3'
SP6 promoter	5' GAT TTA GGT GAC ACT ATA G 3'
Luc27C	5' TAT GCA GTT GCT CTC CAG 3'
pCIN4 IRES for	5' GAT CGC GGC CGC CAC TTG AGT GAC AAT GAC 3'
pCIN4 IRES rev	5' TAT TGG GCG GCC GCA TTA TCC CGG GTT GTG 3'
5'flankIRESFRTS1T-F	5' GAT TTG TGA GCG GCC GCT CGA GCA TGC ATC 3'
3'flankIRESFRTS1T-R	5' GGC TTG TAG GCG CCG TCA GTC GAT TTG TTG 3'
MluI-hARNT25'UTR	5' TAA CGC GTC AAG CGG GCG CCT ATC 3'
hARNT2stop-XbaI	5' CGT CTA GAC TAC TCA GAA AAC GG 3'
mSIM1 sense	5' GCC TGC GAG GGA GGA CGG 3'
mSIM1 full-lengthrev	5' AAA ACA GTT AAC ACT TCC ATT GGT TAT 3'
SIM1-297R	5' GAA CCT GTA GTA CTT GGT GG 3'
SIM1-334C	5' GAA TTG CGG CCG CGT ATT CTG TGT CTG T 3'
hSIM1-264F	5' CTG TAC CAC CAT GTG CAC G 3'
hSIM1-rev	5' AAA ACA GTT AAC GCT TCC GTT GGT TAT 3'
QPCRSim1-F	5' CCT CTT ACC AAG GCC TCC 3'
ENTR1Aseq F	5' TTA GTT AGT TAC TTA AGC TCG GGC 3'
ENTR1Aseq R	5' GTA ACA TCA GAG ATT TTG AGA CAC 3'
AttB1F	5' GGG GAC AAG TTT GTA CAA AAA AGC AGG CTT AGA TCT GAA TTC AAG 3'
AttB2R	5' GGG GAC CAC TTT GTA CAA GAA AGC TGG GTA AAA A 3'
hSIM1si1200F	5' GAT CCC CCC TCA CAG ACA CAG AAT ACA ATT CAA GAG ATT GTA TTC TGT GTC TGT GAG GTT TTT GGA AA 3'
hSIM1si1200R	5' AGC TTT TCC AAA AAC CTC ACA GAC ACA GAA TAC AAT CTC TTG AAT TGT ATT CTG TGT CTG TGA GGG GG 3'
hSIM1si1381F	5' GAT CCC CGG ATT TCA CAC AGA AAG ATT TCA AGA GAA TCT TTC TGT GTG AAA TCC TTT TTG GGA A 3'
hSIM1si1381R	5' AGC TTT TCC AAA AAG GAT TTC ACA CAG AAA GAT TCT CTT GAA ATC TTT CTG TGT GAA ATC CGG G 3'

hSIM1si2227F	5' GAT CCC CGC TAA AGA CTA TCT GCA TTC GTT CAA GAG ACG AAT GCA GAT AGT CTT TAG CTT TTT GGA AA 3'
hSIM1si2227R	5' AGC TTT TCC AAA AAG CTA AAG ACT ATC TGC ATT CGT CTC TTG AAC GAA TGC AGA TAG TCT TTA GCG GG 3'
hSIM1si2378F	5' GAT CCC CGC GAA ACC ATT AGA AAC TAT TCA AGA GAT AGT TTC TAA TGG TTT CGC TTT TTG GAA A 3'
hSIM1si2378R	5' AGC TTT TCC AAA AAG CGA AAC CAT TAG AAA CTA TCT CTT GAA TAG TTT CTA ATG GTT TCG CGG G 3'
LVsi1200no1R	5' CAA TCT CTT GAA TTG TAT TCT GTG TCT GTG AGG GAT CTC TAT CAC 3'
LVsi1200no2F	5' CAA TTC AAG AGA TTG TAT TCT GTG TCT GTG AGG TTT TTA CCC AGC 3'
LVsi1381no1R	5' GAT TCT CTT GAA ATC TTT CTG TGT GAA ATC CGA TCT CTA TCA C 3'
LVsi1381no2F	5' GAT TTC AAG AGA ATC TTT CTG TGT GAA ATC CTT TTT ACC CAG C 3'
LVsi2227no1R	5' TCG TCT CTT GAA CGA ATG CAG ATA GTC TTT AGC GAT CTC TAT CAC 3'
LVsi2227no2F	5' TCG TTC AAG AGA CGA ATG CAG ATA GTC TTT AGC TTT TTA CCC AGC 3'
hSIM1 K51N-F	5' CCA GCT ATC TCA ATA TGA GAG TGG TG 3'
hSIM1 K51N-R	5' CAC CAC TCT CAT ATT GAG ATA GCT GG 5'
hSIM1 S71R-F	5' CAA GTC GGA CCA GGC CCC TGG ACA AC 3'
hSIM1 S71R-R	5' GTT GTC CAG GGG CCT GGT CCG ACT TG 3'
hSIM1 I128T-F	5' GAA TAC ACT CAC CCG 3'
hSIM1I128T-R	5' CGG GTG AGT GTA TTC 3'
hSIM1 Q152E-F	5' TTC GTG GAG GAG TAT 3'
hSIM1 Q152E-R longer	5' GAT CTC ATA CTC CTC CAC GAA GTG AGA 3'
hSIM1 R171H-F	5' GCC AAG CAT AAC GCC 3'
hSIM1 R171H-R	5' GGC GTT ATG CTT GGC 3'
hSIM1 L238R-F	5' GTT CCG CGC CAG CCG GGA CAT GAA GCT CAT C 3'
hSIM1 L238R-R	5' GAT GAG CTT CAT GTC CCG GCT GGC GCG GAA C 3'
hSIM1 T292A-F	5' GGA CAG GTG ACC GCC AAG TAC TAC AGG 3'
hSIM1 T292A-R	5' CCT GTA GT ACTT GGC GGT CAC CTG TCC 3'
hSIM1 T361I-F	5' CCA GCA GCT CCA TCC CCA CCA TGA CTG 3'
hSIM1 T361I-R	5' CAG TCA TGG TGG GGA TGG AGC TGC TGG 3'
hSIM1 R383G-F	5' CAA AGT CAA AAT CCG GG ACTT CCC CAT AC 3'
hSIM1 R383G-R	5' GTA TGG GGA AGT CCC GGA TTT TGA CTT TG 3'
hSIM1 P497R-F	5' GCC TTG CGC CTG ACA 3'
hSIM1 P497R-R longer	5' GGA GGC CTT TGT CAG GCG CAA GGC TGC GCG AGA G 3'

hSIM1 S541L-F	5' CCT GGG TTG GCC AGT 3'
hSIM1 S541L-R	5' ACT GGC CAA CCC AGG 3'
hSIM1 R550H-F	5' CGA TAT CAT ACT GAG CAG 3'
hSIM1 R550H-R	5' CTG CTC AGT ATG ATA TCG 3'
hSIM1 P692L-F	5' GCA GGA GAC CAC CTT ACT GTC TCT CC 3'
hSIM1 P692L-R	5' GGA GAG ACA GTA AGG TGG TCT CCT GC 3'
hSIM1 R703Q-F	5' GCT TTG GCT CTC ACC AGC AGT ATT TTG AC 3'
hSIM1 R703Q-R	5' GTC AAA ATA CTG CTG GTG AGA GCC AAA GC 3'
hSIM1 D707H-F	5' CGG CAG TAT TTT CAC AAG CAT GCT TAC 3'
hSIM1 D707H-R	5' GTA AGC ATG CTT GTG AAA ATA CTG CCG 3'
hSIM1 T712I-F	5' CAA GCA TGC TTA CAT ATT AAC TGG ATA TGC 3'
hSIM1 T712I-R	5' GCA TAT CCA GTT AAT ATG TAA GCA TGC TTG 3'

2.2.10.2. RT-PCR primers

SIM1-332F	5' CAC AGA CAC AGA ATA CAA AGG 3'
hSim1 sense aka. hSIM1-455N	5' GCT GGT GGA AGA GAG GC ATT 3'
hSim1 antisense aka. hSIM1-609C	5' TGG AGA ACT GAC CAC ACT AT 3'
mARNT2 sense	5' CCT TCA GCT CTT CCG TGG 3'
mARNT2 antisense	5' TCT GGG CAG TAG AAG CCT 3'
mARNT2-R2	5' CAT GGG TAG CAT GTC CTG G 3'
hArnt sense	5' TCT GTC ATG TTC CGG TTC CGG TCT 3'
hArntasense	5' TCA AGG GGC TTG CTG TGT TCT GGT 3'
GAPDH 113n	5' GGA GCC AAA AGG GTC ATC 3'
GAPDH 323c	5' ACC ACC CTG TTG CTG TAG 3'

2.2.10.3. QPCR primers

Primers used to amplify human transcripts:

hMYOM2_QPCR_F	5' AGA GAA GAA TCG TGG CAG GTT G 3'
hMYOM2_QPCR_R	5' GGA TGT CCT GGT CGT TCT TG 3'
hACTBL2_QPCR_F	5' ATC ACC ATT GGG AAT GAA CG 3'
hACTBL2_QPCR_R	5' GAT CCC ACT GGA CTC AAT GC 3'
hFAT4_QPCR_4187F	5' TGC AAA GAG GGA CTC ACT GG 3'
hFAT4_QPCR_4213R	5' TGA CTC ATG TGG TAG TCC AAG C 3'
hBMP3_QPCR_F	5' GAT ATT GGC TGG AGT GAA TGG 3'
hBMP3_QPCR_R	5' GGA TGG TAG CAT GAT TTG ATG G 3'
hODZ1_QPCR_F	5' CTC CTT CCT AAA CCA TTC AAC C 3'

hODZ1_QPCR_R	5' GGA AAC CAA ATA GCT CCA ACC 3'	
hNEFL_QPCR_F	5' CAA CGT GAA GAT GGC TTT GG 3'	
hNEFL_QPCR_R	5' TGA AAC TGA GTC GGG TCT CC 3'	
hTMOD1_QPCR_F	5' AAT TCG GCT ACC ACT TTA CCC 3'	
hTMOD1_QPCR_R	5' CCG CCA CAC ACT AGA CAC C 3'	
hPCDH20_QPCR_944-F	5' TGC ATA TGC GAG AGA GAA AGC 3'	
hPCDH20_QPCR_3'UTR-R	5' GTT GTC CTT CAT TGG AAC TGC 3'	
hDUB3_QPCR_444F	5' AAC GAA GCC TGA GTT CAA CG 3'	
hDUB3_QPCR_481R	5' TTT GCT GTT CAG GAT GAT GG 3'	
hPPFIA2_QPCR_F	5' TCG CAC ATA CTC ATG TTG ACC 3'	
hPPFIA2_QPCR_R	5' CAC TGC TCG TCT TCT TAG ATG G 3'	
hCHRM3_QPCR_F	5' TCT GGT CCA CTC CTC TGT CC 3'	
hCHRM3_QPCR_R	5' TGT GCA AGG TCA TTG TGA CTC 3'	
ARNT2_QPCR_F	5' CAG CAG ATC TAC TCC CAA GGA 3'	also used to amplify mARNT2
ARNT2_QPCR_R	5' TCA TTC ACT CCA GGC ACA TG 3'	also used to amplify mARNT2
SIM1_QPCR_587F	5' CAG ACC AAC TGG CTT CCA TT 3'	
SIM1_QPCR_627R	5' TGC TGG ATA TGG TCA CAT GG 3'	
hPOLR2A_qPCR_For	5' GAG AGT CCA GTT CGG AGT CCT 3'	
hPOLR2A_qPCR_Rev	5' CCC TCA GTC GTC TCT GGG TA 3'	

Primers used to amplify mouse transcripts:

mMYOM2 QPCR-1381F	5' GGC TTA CCA GAC GTG GTG AC 3'
mMYOM2 QPCR-1416R	5' TGA GCT CAA TGT CCT TGT CG 3'
mACTBL2 QPCR 3'UTR-F	5' AGG ATG TTG CTA TGG GAA GC 3'
mACTBL2 QPCR 3'UTR-R	5' CGA CAA CCA AAT CTC CCA AG 3'
mFAT4 QPCR-4353F	5' AAG CCC ATC GAG ACA CTC AG 3'
mFAT4 QPCR-4384R	5' AGC TGT GCT TCC CAC TGA AG 3'
mBMP3 QPCR-291F	5' ACA GCG GCA GCA GTA GAG TC 3'
mBMP3 QPCR-336R	5' TGT AAC CCC TTC GTT TGA GG 3'
mODZ1 QPCR-2448F	5' TGG ATT TCC CAA ACC AGA AC 3'
mODZ1 QPCR-2478R	5' CTG GAG CTC ACA CTG AAT GC 3'
mNEFL QPCR-323F	5' GGG TAT GAA CGA AGC TCT GG 3'
mNEFL QPCR-362R	5' TCG TGC TTC TCA GCT CAT TC 3'
mTMOD1 QPCR-376F	5' CCA TGA TGA GCA ACA ACG AC 3'
mTMOD1 QPCR-410R	5' CCT TCC CTT CGC ATG CTA GAC 3'
mPCDH20 QPCR-29F	5' CAT CTA CAT CGT CCC AGC AG 3'
mPCDH20 QPCR-66R	5' CGG TAG CAC GGC TGT AAC TC 3'
mPPFIA2 QPCR-1219F	5' GCT GGA TTC AGG TTG ACC AC 3'
mPPFIA2 QPCR-1253R	5' ATG TGC GAA CAG TGG AGT TG 3'
mCHRM3 QPCR 5'UTR-F	5' ACC AAG CAC GTG ACA TTC TG 3'

mCHRM3 QPCR 5'UTR-R	5' TCC AAG GAA GAA CAC ACC AAC 3'
mOxt QPCR-F	5' TTG GCT TACT GG CTC TGA CC 3'
mOxt QPCR-R	5' GGG AGA CAC TTG CGC ATA TC 3'
maVP QPCR-F	5' CTG CAG CGA CGA GAG CTG 3'
maVP QPCR-R	5' CTG TGT GGC GTT GCT TGG 3'
mCrh QPCR-F	5' CAC CTA CCA AGG GAG GAG AAG 3'
mCrh QPCR-R	5' AGG ACG ACA GAG CCA CCA G 3'
mTrh QPCR-F	5' GGT GCT GCC TTA GAT TCC TG 3'
mTrh QPCR-R	5' CTC TCT TCG GCT TCA ACG TC 3'
mSs QPCR-F	5' CAG ACT CCG TCA GTT TCT GC 3'
mSs QPCR-R	5' GGG CAT CAT TCT CTG TCT GG 3'
Nestin FWD	5' GCT TCT CTT GGC TTT CCT GA 3'
Nestin REV	5' AGA GAA GGA TGT TGG GCT GA 3'
mGFAP QPCR-246F	5' TGC AAG AGA CAG AGG AGT GG 3'
mGFAP QPCR-284R	5' CGG CGA TAG TCG TTA GCT TC 3'
mNeuroD qPCR-F	5' GCT CCA GGG TTA TGA GAT CG 3'
mNeuroD qPCR-R	5' CCG CTC TCG CTG TAT GAT TT 3'
Pax6 FWD	5' GCA CAT GCA AAC ACA CAT GA 3'
Pax6 REV	5' ACT TGG ACG GGA ACT GAC AC 3'
Mash1 qRT-PCR F	5' TAG CCC AGA GGA ACA AGA GC 3'
Mash1 qRT-PCR R	5' CTG CTT CCA AAG TCC ATT CC 3'
Rax F 97bp	5' TCT GGA ACC ATA CCT GGA CC 3'
Rax R 97bp	5' GCC TTC GAG AAG TCC CAC TA 3'
Six3 F 88bp	5' CCT CAC CCC CAC ACA AGT AG 3'
Six3 R 88bp	5' CTG ATG CTG GAG CCT GTT C 3'
mrArnt1_F	5' CTT GGC TCT GTG AAG GAA GG 3'
mrArnt1_R	5' TCA TCA TCT GGG AGG GAG AC 3'
mSIM2 QPCR-380F	5' GCT GAG AAC CAA CCC ATA TCC 3'
mSIM2 QPCR-409R	5' GTT CTC CAG TTT CCC ACC TG 3'
mouseMc4r.Fwd927	5' GGT CGG GAT CAT CAT AAG TTG TA 3'
mouseMc4r.Rev.1086	5' CAT CAG GAA CAT GTG GAC ATA GA 3'
QPCRSIM1-513F	5' GAG GCA GGC AGG TAC TTC C 3'
QPCRSIM1-551R	5' CTG ACC ACA CTA TCT TCA TCC 3'
mrPolr2a_qPCR1_F	5' GCA CCA TCA AGA GAG TGC AG 3'
mrPolr2a_qPCR1_R	5' GGG TAT TTG ATA CCA CCC TCT G 3'

2.2.10.4. EMSA oligos

WT MYOM2 probe:

F-S2-1	5' [FAM]AA TCA CGA TTC AAA ACT GGA ACG TGT CCT TTC TGA GTC CC 3'
F-S2-2	5' [FAM]GG GAC TCA GAA AGG ACA CGT TCC AGT TTT GAA TCG TG ATT 3'

MUT MYOM2 probe:

S2mut1/2UPP	5' [FAM]AA TCT TTA TTC AAA ACT GGA AAA AGT CCT TTC TGA GTC CC 3'
S2mut1/2LOW	5' [FAM]GG GAC TCA GAA AGG ACT TTT TCC AGT TTT GAA TAA AGA TT 3'

WT hEPO probe:

F-hEPOupper	5' [FAM]GA GGG GGC TGG GCC CTA CGT GCT GTC TCA CAC AGC CTG 3'
F-hEPOlower	5' [FAM]CA GGC TGT GTG AGA CAG CAC GTA GGG CCC AGC CCC CTC 3'

MUT hEPO probe:

F-hEPOmtUPP	5' [FAM]GA GGG GGC TGG GCC CTA AAA GCT GTC TCA CAC AGC CTG 3'
F-hEPOmtLOW	5' [FAM]CA GGC TGT GTG AGA CAG CTT TTA GGG CCC AGC CCC CTC 3'

WT hVEGF probe:

F-hVEGFupper	5' [FAM]CT CCA CAG TGC ATA CGT GGG CTC CAA CAG GTC CTC TG 3'
F-hVEGFlower	5' [FAM]CA GAG GAC CTG TTG GAG CCC ACG TAT GCA CTG TGG AG 3'

OL331/332 probe:

EMSA OL331 upper	5' GGG CCA TGG GAT GTG CGT GAC ATT TC 3'
EMSA OL332 lower	5' GGG AAA TGT CAC GCA CAT CCC ATG GC 3'

OL464/465 probe:

EMSA OL464 upper	5' GGG CCA TGG GAT GTA CGT GAC ATT TC 3'
EMSA OL465 lower	5' GGG AAA TGT CAC GTA CAT CCC ATG GC 3'

1xCME probe:

EMSA 1xCME upper	5' GGA ATT TGT ACG TGC CAC AGA 3'
EMSA 1xCME lower	5' GGT CTG TGG CAC GTA CAA ATT 3'

2xCME probe:

EMSA 2xCME upper	5' GGA ATT TGT ACG TGC CAC AGA CTA GAA ATT TGT ACG TGC CAC AGA 3'
EMSA 2xCME lower	5' GGT CTG TGG CAC GTA CAA ATT TCT AGT CTG TGG CAC GTA CAA ATT 3'

2.3. Electronic Resources

Pubmed <http://www.ncbi.nlm.nih.gov/pubmed/>

Source database for nucleotide and protein sequences and Entrez Gene.

Blat Search Genome <http://genome.ucsc.edu/cgi-bin/hgBlat?command=start>

Online genomic DNA database and alignment tool, usually used to look up the location of an input sequence in the genome or determine the exon structure of an mRNA.

ClustalW <http://align.genome.jp/>

Multiple sequence alignment tool.

Primer3 <http://frodo.wi.mit.edu/primer3/>

Online primer design tool used to pick QPCR primer pairs.

PsiPred <http://bioinf.cs.ucl.ac.uk/psipred/>

Protein secondary structure prediction tool.

Q-Gene [227]

Microsoft Excel based application used to analyse QPCR data.

SPSS 17.0-18.0

Software package used for all statistical analyses performed in this thesis.

TESS (Transcription Element Search System) <http://www.cbil.upenn.edu/cgibin/tess/tess?RQ=SEA-FR-QueryF>

Online predictive transcription factor binding site tool.

2.4. Methods

2.4.1. Bacterial culture

2.4.1.1. Preparation of chemically competent DH5α

Day 1: A single LB+agar plate was streaked from a frozen stock of DH5α bacteria and incubated ON/37°C.

Day 2: 2mL liquid LB was inoculated with a single colony picked from the ON growth plate and incubated ON/37°C with shaking.

Day 3: 1.1mL of the ON culture was subcultured into 25mL fresh LB pre-warmed to 37°C, and incubated at 37°C with shaking until the OD_{600} reached 0.6 (approximately 1.5hr). The entire culture was then subcultured into 500mL fresh LB pre-warmed to 37°C, and incubated at 37°C with shaking until the OD_{600} once again reached 0.6 (approximately 2hr). Cells were then pelleted in a pre-cooled centrifuge at 4700rpm/40min/4°C, the SN removed, and cells kept on ice. Cells were resuspended in 200mL cold TFB1 and re-centrifuged 4700rpm/20min/4°C. The SN was removed and cells resuspended immediately in 20mL cold TFB2, and kept on ice whilst aliquoting into pre-chilled microcentrifuge tubes and snap freezing in a dry ice/EtOH bath. Cells were stored at -80°C.

2.4.1.2. Heat shock transformation of chemically competent bacteria

With intact plasmid: 5-20ng DNA was gently mixed with 100µL chemically competent DH5α and incubated for 8min/RT followed by incubation for 2min/37°C. Cells were then plated onto LB+agar containing antibiotic (**2.2.8.2.2.**) and incubated ON/37°C.

With ligation mix: the ligation reaction (**2.4.2.8.**) was gently mixed with 100µL chemically competent DH5α and incubated on ice for 20min. Cells were then heat shocked for 2min/42°C and cooled on ice for 2min. 500µL SOC was added and cells incubated for 30min/37°C on a rotary wheel, followed by centrifugation at 6000rpm/2min/RT. SN was removed, leaving approximately 50µL with which to resuspend the pellet and plate onto an LB+agar plate containing antibiotic (**2.2.8.2.2.**). Plates were then incubated ON/37°C.

2.4.1.3. Growth and maintenance of bacteria

Transformed bacteria were propagated in liquid LB or on solid media, supplemented with antibiotic and/or other additives where appropriate, at 37°C. Frozen stocks were made from liquid cultures by mixing 1mL overnight culture with glycerol to a final

concentration of 16%, followed by incubation for 5min/RT and long-term storage at -80°C.

Final concentrations of antibiotics, colour selectors and other additives in liquid and solid media are listed in **2.2.8.2.2.**

2.4.2. DNA manipulation

2.4.2.1. Preparation of plasmid DNA

Purified plasmid DNA was extracted from ON liquid cultures using the QIAprep Spin Miniprep Kit for small scale purifications, and the Plasmid Midi Kit for large scale purifications, as per manufacturer's instructions.

2.4.2.2. PCR

Each set of PCR reaction conditions was optimised for the particular primer pair used. Reactions were performed in 0.2mL tubes in an MJ Research PTC-200 Peltier Thermal Cycler.

2.4.2.2.1. Using Taq polymerase

2.4.2.2.1.1. RT-PCR

25µL reactions were prepared on ice comprising 1x Reaction Buffer, 1.5mM $MgCl_2$, 0.25mM dNTPs, 1U Taq polymerase and 0.5µM each primer. 1µL cDNA or 1-5ng plasmid DNA was used as template. Standard reaction conditions: 95°C/2min, [95°C/30sec, 45-60°C/30sec, 72°C/30sec-3min]x25-40 cycles, 72°C/10min. Reactions were then separated by TBE/EtBr agarose gel electrophoresis (**2.4.5.1.**) and visualised by UV detection.

2.4.2.2.1.2. Colony PCR

15µL reactions were prepared on ice comprising 1x Reaction Buffer, 1.5mM $MgCl_2$, 0.25mM dNTPs, 1U Taq polymerase and 0.8µM each primer. 1µL overnight bacterial culture was used as template, or a single colony was picked from an ON transformation

plate with a toothpick and used to inoculate the reaction mix. Standard reaction conditions: 95°C/2min, [95°C/30sec, 45-60°C/30sec, 72°C/30sec-1min]x40 cycles, 72°C/10min. Reactions were then separated by TBE/EtBr agarose gel electrophoresis (**2.4.5.1.**) and visualised by UV detection.

2.4.2.2.2. Using Pfu Turbo

The high fidelity enzyme Pfu Turbo was used to generate PCR products ≤1kb ultimately used for protein expression. 50µL reactions were prepared on ice comprising 1x Pfu Turbo buffer, 0.25mM dNTPs, 2.5U Pfu Turbo polymerase and 0.5µM each primer. 1µL cDNA or 1-5ng plasmid DNA was used as template. Standard reaction conditions: 95°C/2min, [95°C/30sec, 45-60°C/30sec, 72°C/30sec-3min]x40 cycles, 72°C/10min. Reactions were then separated by TBE/EtBr agarose gel electrophoresis (**2.4.5.1.**) and visualised by UV detection.

2.4.2.2.3. Using Phusion

The high fidelity enzyme Phusion was used to generate PCR products >1kb ultimately used for protein expression. 50µL reactions were prepared on ice comprising 1x High Fidelity Buffer, 0.2mM dNTPs, 1U Phusion polymerase and 0.5µM each primer. 1µL cDNA or 1-5ng plasmid DNA was used as template. Standard reaction conditions: 98°C/30sec, [98°C/10sec, 45-60°C/30sec, 72°C/30sec-2min]x40 cycles, 72°C/10min. Reactions were then separated by TBE/EtBr agarose gel electrophoresis (**2.4.5.1.**) and visualised by UV detection.

2.4.2.2.4. Overlap Extension PCR

First round amplification: two 50µL reactions were prepared on ice comprising 1x Pfu Turbo buffer, 0.2mM dNTPs, 2.5U Pfu Turbo polymerase and 0.5µM each primer. 1-5ng plasmid DNA was used as template. Standard reaction conditions: 95°C/2min, [95°C/30sec, 45-60°C/30sec, 72°C/30sec-3min]x40 cycles, 72°C/10min. Reactions were then separated by TBE/EtBr agarose gel electrophoresis (**2.4.5.1.**) and visualised by UV detection. Recovery of DNA fragments from excised gel slices was performed using the QIAquick Gel Extraction Kit as per manufacturer's instructions, eluting in a final volume of 30µL.

End-filling: a single 20µL reaction was prepared on ice comprising 1x Reaction Buffer, 0.5mM dNTPs, 2.5U Taq polymerase, and 5µL of each of the two first round amplification products. Standard reaction conditions: 95°C/5min, 55°C/50sec, 72°C/15min.

Second round amplification: a single 25µL reaction was prepared on ice comprising 1x Pfu Turbo buffer, 0.4mM dNTPs, 1.25U Pfu Turbo polymerase and 0.5µM each primer, using the forward and reverse flanking primers used in the first round amplification. 1µL of the end-filling reaction was used as template. Standard reaction conditions: 95°C/2min, [95°C/30sec, 45-60°C/30sec, 72°C/30sec-3min]x40 cycles, 72°C/10min. Reactions were then separated by TBE/EtBr agarose gel electrophoresis (**2.4.5.1.**) and visualised by UV detection.

2.4.2.2.5. Sequencing

20µL reactions were prepared on ice comprising 0.875x Sequencing Buffer, 0.2µM or 100ng primer, 200ng plasmid DNA, and 1µL BIG DYE^{TM} reaction mix. Reaction conditions: 96°C/3min, [96°C/30sec, 50°C/15sec, 60°C/4min]x27 cycles. PCR products were precipitated from solution by addition of 1µg/µL glycogen and 60% (v/v) isopropanol, vortexed, and incubated for 15min/RT, followed by centrifugation in a pre-cooled centrifuge at 14,000rpm/20min/4°C. The SN was removed and the pellet washed in 250µL 75% (v/v) isopropanol, re-centrifuged at 14,000rpm/5min/4°C, the SN removed and the pellet dried for 1min/95°C. Samples were processed at the DNA sequencing facility, Institute for Medical and Veterinary Sciences (IMVS), Frome Rd, Adelaide, Australia.

2.4.2.2.6. QPCR

2.4.2.2.6.1. Experimental design

Triplicate 15uL reactions were prepared on ice comprising 1x Fast SYBR Green Master Mix, 1uL cDNA (**2.4.3.2.**), and 250nM each of forward and reverse primers, then aliquoted into 96 well trays and analysed on an Applied Biosystems Step One Plus thermo cycler. Standard reaction conditions: 95°C/20sec, [95°C/3sec, 60°C/15-30sec]x40

cycles, followed by melt curve analysis (60°C/1min, increase sample temperature to 95°C and measure fluorescence every 0.5°C, 95°C/15sec).

2.4.2.2.6.2. Data processing and statistical analysis

Raw data was analysed using a combination of Step One v2.1 and the Microsoft Excel based software application Q-Gene [227]. Threshold values were always adjusted to 0.2 in the Step One program with an arbitrary C_T value ≤ 30 cycles required for each gene. Primer efficiency values were adjusted using Q-Gene with reference to a previously determined value between 1.6 and 2.0.

Efficiency values for primer pairs:

hMYOM2_QPCR_F, hMYOM2_QPCR_R	1.88
hACTBL2_QPCR_F, hACTBL2_QPCR_R	2.0
hFAT4_QPCR_4187F, hFAT4_QPCR_4213R	2.0
hBMP3_QPCR_F, hBMP3_QPCR_R	2.0
hODZ1_QPCR_F, hODZ1_QPCR_R	1.97
hNEFL_QPCR_F, hNEFL_QPCR_R	1.93
hTMOD1_QPCR_F, hTMOD1_QPCR_R	2.0
hPCDH20_QPCR_944-F, hPCDH20_QPCR_3'UTR-R	1.97
hDUB3_QPCR_444F, hDUB3_QPCR_481R	1.90
hPPFIA2_QPCR_F, hPPFIA2_QPCR_R	2.0
hCHRM3_QPCR_F, hCHRM3_QPCR_R	2.0
ARNT2_QPCR_F, ARNT2_QPCR_R	2.0
SIM1_QPCR_587F, SIM1_QPCR_627R	2.0
hPOLR2A-qPCR-For, hPOLR2A-qPCR-Rev	1.90

mMYOM2 QPCR-1381F, mMYOM2 QPCR-1416R	1.96
mACTBL2 QPCR 3'UTR-F, mACTBL2 QPCR 3'UTR-R	2.0
mFAT4 QPCR-4353F, mFAT4 QPCR-4384R	1.94
mBMP3 QPCR-291F, mBMP3 QPCR-336R	2.0
mODZ1 QPCR-2448F, mODZ1 QPCR-2478R	1.94
mNEFL QPCR-323F, mNEFL QPCR-362R	2.0
mTMOD1 QPCR-376F, mTMOD1 QPCR-410R	2.0
mPCDH20 QPCR-29F, mPCDH20 QPCR-66R	2.0
mPPFIA2 QPCR-1219F, mPPFIA2 QPCR-1253R	1.86
mCHRM3 QPCR 5'UTR-F, mCHRM3 QPCR 5'UTR-R	2.0
mOxt QPCR-F, mOxt QPCR-R	1.84

maVP QPCR-F, maVP QPCR-R	2.0
mCrh QPCR-F, mCrh QPCR-R	2.0
mTrh QPCR-F, mTrh QPCR-R	1.90
mSs QPCR-F, mSs QPCR-R	1.82
mSIM2 QPCR-380F, mSIM2 QPCR-409R	1.68
mouseMc4r.Fwd927, mouseMc4r.Rev.1086	2.0
QPCRSIM1-513F, QPCRSIM1-551R	2.0
mrPolr2a_qPCR1_F, mrPolr2a_qPCR1_R	2.0

The fluorescence value for each target gene in each well was normalised to the corresponding fluorescence value for the reference gene, and the three normalised expression values for each triplicate were averaged. These figures were then converted to natural log values for the purposes of statistical analysis. Univariate ANOVA or two-tailed paired t test was performed using SPSS 18.0. Separate analyses were performed for each gene within an experiment.

Representative melt curves for each of the above primer pairs can be found in the **Appendix**.

2.4.2.3. Phosphorylation and annealing of dsDNA oligos for subcloning

Upper and lower oligos were phosphorylated in separate tubes by mixing 1x T4 DNA Ligase Buffer and 20U T4 Polynucleotide Kinase with 14µM oligo on ice in a final volume of 30uL and incubating for 30min/37°C, followed by heat inactivation for 10min/65°C. Each reaction was then diluted to 60µL with 1x TE. 50µL of each phosphorylated oligo were then pooled in a new microcentrifuge tube with 50mM NaCl, incubated for 5min in a 95°C heating block, then the block removed and allowed to cool to RT ON.

2.4.2.4. Generation of dsDNA probes for use in EMSAs

2.4.2.4.1. Annealing

100µM each upper and lower oligo were mixed in a single microcentrifuge tube with 1x TE and 50mM NaCl, incubated for 5min in a 95°C heating block, then the block removed

and allowed to cool to RT ON. The oligo stock was then diluted 1:10 with 1x TE to yield a 1µM working stock.

2.4.2.4.2. ^{32}P labelling

50nM annealed ds oligo was mixed with 1x TM buffer, 10mM DTT, 10U Klenow Large Fragment, and 40µCi α-^{32}P-dCTP in a final volume of 20µL. The reaction was incubated for 20min/30°C and then diluted to 50µL with STE buffer. The entire reaction was then purified using a ProbeQuant G-50 Micro Column as per manufacturer's instructions. 1µL of the eluted probe was counted in a scintillation counter to determine the labelling efficiency of the reaction. A minimum labelling efficiency of 40,000cpm/µL was required for further experiments.

2.4.2.5. Restriction enzyme digestion

Typically, 1-10U restriction enzyme per 1µg DNA was used in accordance with the guidelines laid out in the New Englad Biolabs catalogue and technical reference book. Reactions were incubated at the appropriate temperature for 3hr-ON.

2.4.2.6. Dephosphorylation of plasmid DNA

Linearised plasmid DNA was dephosphorylated by mixing with 1x SAP buffer and 1U SAP per 1µg DNA on ice in a final volume of 40-50µL. The reaction was then incubated for 15min-1hr/37°C before heat inactivation for 15min/65°C, or purification using the QIAquick Gel Extraction Kit as per manufacturer's instructions, eluting in a final volume of 30µL.

2.4.2.7. Generation of blunt-ended DNA using Klenow Large Fragment

Plasmid DNA previously digested with "sticky end" restriction enzyme/s was separated by TBE/EtBr agarose gel electrophoresis (**2.4.5.1.**) and visualised by UV detection. Recovery of DNA fragments from excised gel slices was performed using the QIAquick Gel Extraction Kit as per manufacturer's instructions, eluting in a final volume of 30µL. Eluted DNA was then mixed with 1x T4 DNA Ligase Buffer, 33µM dNTPs and 1U Klenow Large Fragment per 1µg DNA on ice in a final volume of 40µL. The reaction

was incubated at 25°C/15min then heat inactivated by addition of EDTA pH8.0 to a final concentration of 10mM and incubation for 20min/75°C.

2.4.2.8. Ligations

2.4.2.8.1. Ligation of purified DNA fragments

Recovery of DNA fragments from excised gel slices was routinely performed using the QIAquick Gel Extraction Kit as per manufacturer's instructions, eluting in a final volume of 30µL. Typically, a 3:1 or 4:1 molar ratio of insert:vector backbone was used for subsequent ligations as calculated by the following equation:

$$\frac{\text{ng of vector in 1µL} \times \text{size of insert in kb}}{\text{size of vector in kb}} \times \text{desired } \frac{\text{insert}}{\text{vector}} \text{ ratio} = \text{ng insert needed}$$

For three-way ligation of full-length mSIM1 into the pEF-Myc-IRES-puro backbone, all three DNA fragments (the two halves of the mSIM1FL insert and the vector backbone) were incubated simultaneously in a single ligation reaction. Equal moles of the two mSIM1 DNA fragments were added, at a 3:1 molar ratio of insert:vector backbone relative to the size of the larger insert fragment.

Ligation reactions were set up on ice and typically comprised 1x T4 DNA Ligase Buffer, 5U T4 DNA Ligase, and 10-50ng SAP treated vector DNA (**2.4.2.6.**) along with the appropriate amount of insert in a final reaction volume of 10-20µL. Occasionally, 1x Rapid Ligation Buffer was used in place of T4 DNA Ligase Buffer. Reactions were incubated for 3hr/RT for sticky-sticky ligations, and ON/4°C for sticky-blunt, blunt-blunt and 3-way ligations, before transformation into chemically competent DH5α (**2.4.1.2.**).

2.4.2.8.2. Ligation of phosphorylated and annealed dsDNA oligos

Phosphorylated and annealed dsDNA oligos (**2.4.2.3.**) were diluted 1:10,000 in 1xTE. Typically, a 3:1 volume ratio of diluted insert:vector backbone was used for ligations. The vector backbone was always SAP treated to prevent re-ligation (**2.4.2.6.**). Ligation reactions were set up on ice and incubated as above (**2.4.2.8.1.**), before transformation into chemically competent DH5α (**2.4.1.2.**).

2.4.2.8.3. pGEM-T Easy cloning

2.4.2.8.3.1. A-tailing

Purified DNA fragments generated by restriction enzyme digestion (**2.4.2.5.**) or PCR amplification (**2.4.2.2.2.-2.4.2.2.4.**) were recovered from excised gel slices using the QIAquick Gel Extraction Kit as per manufacturer's instructions, eluting in a final volume of 30µL. 5µL of this eluate was then added to a new tube with 1x Thermo Pol Buffer, 1.5mM $MgCl_2$, 17µM dATP and 1U Taq polymerase on ice in a total volume of 15µL and incubated for 30min/70°C.

2.4.2.8.3.2. Ligation into pGEM-T Easy

Reactions typically comprised 1x Rapid Ligation Buffer, 1µL linearised pGEM-T Easy vector, 5U T4 DNA ligase and 5.5µL A-tailed dsDNA (or MQ H_2O as a negative control) set up on ice in a final volume of 15µL. Reactions were incubated ON/4°C then transformed into chemically competent DH5α (**2.4.1.2.**) and plated on LB+amp plates previously spread with IPTG and X-gal and left to dry (**2.2.8.2.2.**).

2.4.2.8.4. Gateway ® cloning

2.4.2.8.4.1. Generation of entry vectors

dsDNA inserts with attB ends were recombined into pDONR201 as per manufacturer's instructions using BP Clonase™, in a final volume of 10µL. dsDNA inserts without attB ends were cloned into pENTR1A via standard ligation methods (**2.4.2.8.**). Reactions were incubated ON/4°C then transformed into chemically competent DH5α (**2.4.1.2.**) and plated on LB+kan plates.

2.4.2.8.4.2. Generation of expression vectors

Entry vectors were recombined with destination vectors as per manufacturer's instructions using LR Clonase™, in a final volume of 10µL. Occasionally, the ON LR reaction mix was also treated with a destination vector-unique restriction enzyme prior to Proteinase K digestion, to remove any destination vector not containing insert. This was done via

standard restriction enzyme digestion methods (**2.4.2.5.**). Reactions were then transformed into chemically competent DH5α (**2.4.1.2.**) and plated on LB+amp plates.

2.4.3. RNA manipulation

All RNA protocols were performed with RNAse-free solutions, gloves, plugged pipette tips and sterile plasticware to minimise contamination.

2.4.3.1. Isolation of total RNA

Total RNA was isolated from mammalian tissues and cultured cells using TRI Reagent. Briefly, cells were homogenised in TRI Reagent (2mL per confluent 10cm dish). Chloroform (200µL per 1mL lysate) was added and the tube shaken vigorously for 15sec before centrifugation at 14,000rpm/10min/4°C. The aqueous phase was removed to a new tube and an equal volume of 70% (v/v) ethanol was added and mixed by inversion. The entire volume was then loaded onto an RNeasy mini spin column and centrifuged at 14,000rpm/15sec/RT. The bound RNA was then washed and purified as per manufacturer's instructions, and eluted twice in a final volume of 30µL. Evaluation of RNA concentration and quality was performed by measurement of A_{260} and A_{280} with a spectrophotometer and visualisation of approximately 0.2-1µg RNA on a TBE/EtBr agarose gel (**2.4.5.1.**).

2.4.3.2. cDNA synthesis from RNA

For a single reaction, 2µg total RNA was mixed with 0.5µL of 500ng/µL Random Hexamer Primer and 1µL of 500ng/µL Oligo$(dT)_{15}$ Primer on ice in a total volume of 19µL, incubated for 10min/70°C and allowed to cool for 5min on ice. 11µL Reverse Transcriptase master mix was then added on ice, comprising 6µL 5x First Strand Buffer, 2µL 0.1M DTT, 1µL 25mM dNTPs, 1µL SUPERase-In RNAse Inhibitor and 1µL SuperScript III Reverse Transcriptase (or 1µL MQ H_2O as a no RT control) in a total volume of 30µL. The tube was then incubated for 1.5hr/50°C, followed by 15min/70°C before long-term storage at -20°C.

2.4.3.3. Microarray studies

2.4.3.3.1. Sample preparation and validation

293 Flp-In™ T-Rex™ parent (untransformed) cells and 293 T-Rex mSIM1-Myc/IRES/mARNT2 stable cells were seeded in triplicate in 10cm dishes and allowed to reach ~80% confluency ON (6 dishes total). Medium was then removed and replaced with fresh complete medium, with (mSIM1-Myc/IRES/mARNT2 stable cells) or without (parent cells) doxycycline to a final concentration of 1µg/mL (**2.4.4.3.1.**). Cells were harvested for RNA 12hr later (**2.4.3.1.**) and analysed for sample integrity by TBE/EtBr agarose gel electrophoresis (**2.4.5.1.**). 2µg each sample was also used to synthesise cDNA (**2.4.3.2.**), which was assayed via RT-PCR for the presence of *Sim1* and *Arnt2* with reference to the housekeeping gene *GAPDH* (**2.4.2.2.1.1.**). RNA quality and concentration was also verified by the Adelaide Microarray Centre using an RNA Nano LabChip inserted in an Agilent Bioanalyser.

2.4.3.3.2. Microarray and statistical analyses

300ng each sample was provided to the Adelaide Microarray Centre for analysis using the Affymetrix GeneChip Human Gene 1.0 ST Array platform. This single-label hybridisation array platform covers 28,869 genes, and each gene is represented by a set of approximately 26 different probes spread across the full length of the gene. Each chip is hybridised to a single sample only. Normalisation between chips is achieved with reference to a series of approximately 20,000 control probes, standard polyA controls and hybridisation controls represented on every chip. The analyses presented in this thesis therefore utilised six chips in total. Each of the six provided RNA samples underwent whole transcript amplification, fragmentation and biotin end labelling prior to hybridisation, as per standard protocols. After scanning, normalisation and background correction, a mean "summary" for each probe set on each chip was obtained, taking into account differences in GC content between probes within each set. All microarray analyses and data processing were performed by M. Van der Hoek. Processed data were then subjected to statistical analysis according to the method of Limma, with a false discovery rate adjusted p value threshold of <0.05, by Steve Pederson, Discipline of Paediatrics, University of Adelaide, Women's & Children's Health Research Institute, 72 King William Rd, North Adelaide, SA 5006. Further information, including standard

protocols, can be found at the Adelaide Microarray Centre website <http://www.microarray.adelaide.edu.au/>.

2.4.4. Mammalian cell culture and protein analysis

2.4.4.1. Thawing frozen cell line stocks

Complete medium was pre-warmed to 37°C and 10mL aliquoted into a 20mL tube. A frozen aliquot of cells was then removed from liquid nitrogen or -80°C storage and warmed to 37°C in a water bath. Once cells were almost entirely thawed the tube was sprayed with 70% (v/v) EtOH and the cells carefully added to the complete medium before transferring to the appropriately sized culture vessel for incubation at 37°C/5%CO_2.

2.4.4.2. Freezing cell line stocks

Cells in exponential growth phase were harvested with trypsin (**2.4.4.4.**) and resuspended in Freezing Medium before immediate transfer into CryoTubes, followed by ON storage in a Mr. Frosty and long term storage at -80°C.

2.4.4.3. Production of stable cell lines

2.4.4.3.1. 293 Flp-In™ T-Rex™ lines

Cell lines were produced as per manufacturer's instructions and maintained under antibiotic selection (see below and **2.2.8.3.2.**). Gene expression was routinely induced by removal of existing medium and replacement with fresh complete medium supplemented with doxycycline to a final concentration of 1µg/mL.

The following cell lines were generated and used for the studies presented in this thesis:

293 Flp-In™ T-Rex™ hyg p1

Cell line generated via transfection of the parent 293 Flp-In™ T-Rex™ cell line with the expression vector pcDNA5-FRT-TO. Hereafter referred to in this thesis as "empty vector" cell line.

293 Flp-In™ T-Rex™ mARNT2 m1 and m2

Two monoclonal cell lines inducibly overexpressing mARNT2, generated via transfection of the parent 293

Flp-In™ T-Rex™ cell line with the expression vector pcDNA5-FRT-TO-mARNT2. Two isolated antibiotic resistant colonies were then picked using cloning rings and expanded in separate dishes.

293 Flp-In™ T-Rex™ mSIM1FL-Myc/IRESfixed/mARNT2 p1 and p2

Two cell lines inducibly overexpressing Myc-tagged full-length mSIM1 and mARNT2, independently generated via separate transfections of the parent 293 Flp-In™ T-Rex™ cell line with the expression vector pcDNA5-FRT-TO-mSIM1FL-Myc-IRESfixed-mARNT2. **Used for microarray analyses**: hereafter referred to as 293 T-Rex mSIM1-Myc/IRES/mARNT2.

293 Flp-In™ T-Rex™ hSIM1-Myc p1 and p2

Two cell lines inducibly overexpressing Myc-tagged WT hSIM1, independently generated via separate transfections of the parent 293 Flp-In™ T-Rex™ cell line with the expression vector pcDNA5-FRT-TO-hSIM1-Myc.

293 Flp-In™ T-Rex™ hSIM1-Myc K51N, S71R, I128T, Q152E, R171H, L238R, T292A, H323Y, T361I, R383G, P497R, S541L, R550H, R581G, P692L, R703Q, D707H, T712I and T714A

Nineteen cell lines inducibly overexpressing a single Myc-tagged point mutant of hSIM1, generated via transfection of the parent 293 Flp-In™ T-Rex™ cell line with one of the expression vectors listed in **2.2.9.2.1.** (pcDNA5-FRT-TO-hSIM1 K51N-Myc, etc.).

2.4.4.3.2. Production of stable cell lines via lentiviral infection

The following protocol was routinely performed in a PC2 tissue culture laboratory. Contaminated plasticware was rinsed with 1.25% (v/v) hypochlorite solution prior to disposal.

Day 1: 293T cells (passaged fewer than 20 times) were trypsinised (**2.4.4.4.**) and seeded in serum-free medium at approximately 30% confluency in a T75cm^2 flask. Cells were then incubated ON at 37°C/5%CO_2.

Day 2: 293T cells were transfected with 3.75µg pMD2.G (VSV-G), 6.25µg pRSV-Rev, 7.5µg pCMV-dvrp 8.2 and 12.5µg expression vector, which had previously been mixed with 70µL Lipofectamine 2000 and 1mL serum-free medium in a sterile 20mL yellow cap tube and incubated for 20min/RT.

Day 3: 293T medium was removed and replaced with fresh complete medium. Cells to be infected were trypsinised, counted (**2.4.4.4.**) and seeded in complete medium in 12 well trays (1x10^5 cells per well for N4, N7 and N39 cells).

Day 5: 293T medium was harvested into a yellow cap tube and centrifuged in a swing-out rotor at 1500rpm/5min/RT. The SN was then filtered into a fresh, sterile yellow cap tube using a 0.45µm filter attached to a 20mL or 50mL Luer-Lok syringe. Cells (with existing medium removed) were routinely infected with 50% virus/50% fresh complete medium plus 8ug/mL polybrene and incubated ON at 37°C/5%CO_2.

Day 6: medium was removed and replaced with fresh complete medium containing antibiotic selection where appropriate (**2.2.8.3.2.**).

Infected cell lines were routinely passaged at least 3 times before transfer to PC1 conditions.

The following cell lines were generated and used for the studies presented in this thesis:

293T 711G scrambled

Cell line generated via lentiviral infection of 293T cells with the vector pLV711G scrambled.

293T 711G si1200

Cell line generated via lentiviral infection of 293T cells with the vector pLV711G si1200.

293T 711G si1381

Cell line generated via lentiviral infection of 293T cells with the vector pLV711G si1381.

293T 711G si2227

Cell line generated via lentiviral infection of 293T cells with the vector pLV711G si2227.

N4 6-TR and N7 6-TR

Cell lines generated via lentiviral infection of N4 or N7 cells with the expression vector pLenti6-TR. Generated by C. Bindloss.

N4 6-TR/CFlag-mSIM1FL p1 and p2

Two cell lines inducibly overexpressing Flag-tagged full-length mSIM1, independently generated via separate lentiviral infections of the N4 6-TR cell line with the expression vector pLenti4 CFlag-mSIM1FL. These two cell lines exhibit some basal overexpression of mSIM1-Flag in the untreated state relative to N4 6-TR cells. Generated by C. Bindloss.

N7 6-TR/CFlag-mSIM1FL p1 and p2

Two cell lines inducibly overexpressing Flag-tagged full-length mSIM1, independently generated via separate lentiviral infections of the N7 6-TR cell line with the expression vector pLenti4 CFlag-mSIM1FL. These two cell lines exhibit some basal overexpression of mSIM1-Flag in the untreated state relative to N7 6-TR cells. Generated by C. Bindloss.

N39 501 ΔccdB

Cell line generated via lentiviral infection of N39 cells with the expression vector pLV501 ΔccdB. Used solely as a source cell line for subsequent infection with LV416neo ΔccdB (see below).

N39 501 mARNT2

Cell line constitutively overexpressing mARNT2, generated via lentiviral infection of N39 cells with the expression vector pLV501 mARNT2. Used solely as a source cell line for subsequent infection with LV416neo ΔccdB or LV416neo CFlag-mSIM1FL (see below).

N39 416neo ΔccdB/501 ΔccdB

Cell line generated via lentiviral infection of the N39 501 ΔccdB cell line with the expression vector pLV416neo ΔccdB. Referred to in this thesis as N39 empty/empty. Generated by V. Bhakti.

N39 416neo ΔccdB/501 mARNT2

Cell line constitutively overexpressing mARNT2, generated via lentiviral infection of the N39 501 mARNT2 cell line with the expression vector pLV416neo ΔccdB. Referred to in this thesis as N39 empty/mARNT2. Generated by V. Bhakti.

N39 416neo CFlag-mSIM1FL/501 mARNT2 p1 and p2

Two cell lines constitutively overexpressing mARNT2 and Flag-tagged full-length mSIM1, independently generated via separate lentiviral infections of the N39 501 mARNT2 cell line with the expression vector pLV416neo CFlag-mSIM1FL. Referred to in this thesis as N39 mSIM1-Flag/mARNT2. Generated by V. Bhakti.

N39 416neo ΔccdB

Cell line generated via lentiviral infection of N39 cells with the expression vector pLV416neo ΔccdB.

N39 416neo CFlag-mSIM1FL p1 and p2

Two cell lines constitutively overexpressing Flag-tagged full-length mSIM1, independently generated via separate lentiviral infections of the N39 cell line with the expression vector pLV416neo CFlag-mSIM1FL.

N39 711G scrambled

Cell line generated via lentiviral infection of N39 cells with the vector pLV711G scrambled.

N39 711G si1200

Cell line generated via lentiviral infection of N39 cells with the vector pLV711G si1200.

shRNA expression was routinely induced in the appropriate cell lines by removal of existing medium and replacement with fresh complete medium supplemented with doxycycline to a final concentration of 5µg/mL.

2.4.4.4. Routine maintenance of cell lines

Cells were maintained in complete medium, supplemented with antibiotics where appropriate (**2.2.8.3.2.**), at 37°C/5%CO_2. Cells were routinely passaged by aspirating

medium off cells, washing with 1x PBS, incubating with trypsin (1mL per 10cm dish, 2mL per T175cm^2 flask), resuspension by gentle pipetting, and removal of 5-20% into fresh complete medium, with antibiotics where appropriate, in a new culture vessel. To count cells, resuspended cells were diluted 1:5-1:10. An aliquot of this diluted stock was further diluted 1:5-1:10 and counted using a haemocytometer with the following equation:

Cells/mL = number of cells per haemocytometer box x dilution factor x 10^4

Details of complete media used are listed in **2.2.8.3.1.**

2.4.4.5. Transient transfections

Typically, cell lines were transfected at 40-60% confluency using Fugene 6 at a 3:1 ratio of Fugene:DNA as per manufacturer's instructions. Cells were incubated at 37°C/5%CO_2 and harvested 8-24hr post-transfection.

2.4.4.6. Harvesting cells with TEN buffer

Medium was aspirated off cells, and cells were washed with 1x PBS. Ice-cold TEN buffer was then added directly (1mL per 10cm dish) and cells harvested into a microcentrifuge tube by scraping with a rubber policeman. Cells were centrifuged at 1400rpm/5min/4°C, the SN discarded and the cell pellet either stored at -80°C or kept on ice and used immediately.

2.4.4.7. Whole cell protein extract preparation and quantification

This method is adapted from Whitelaw *et al* [29]. Briefly, cells were harvested with TEN (2.4.4.6.) and the cell pellet resuspended in approximately 3 pellet volumes of ice-cold WCE buffer, followed by incubation on a nutator for 30-45min/4°C. The crude lysate was then centrifuged at 14,000rpm/30min/4°C and the SN transferred to a new microcentrifuge tube. Protein yields were determined by the Bradford Protein Assay as per manufacturer's instructions, calculated from OD_{600} measurements with reference to a BSA standard curve. Typically, 30-50µg WCE was used for subsequent denaturing Tris-glycine SDS-PAGE analysis (**2.4.5.3.**). Samples were stored long term at -80°C.

2.4.4.8. Detergent-free nuclear and cytoplasmic protein extract preparation and quantification

This method is adapted from Lees *et al* [228]. Briefly, cells were harvested with TEN (**2.4.4.6.**) and the cell pellet resuspended in approximately 3 pellet volumes of ice-cold detergent-free hypotonic buffer, followed by incubation for 5min on ice, or homogenisation with a pre-cooled Dounce homogeniser on ice (10 strokes). The crude lysate was then centrifuged at 14,000rpm/30min/4°C and the SN (cytoplasmic fraction) transferred to a new microcentrifuge tube. The remaining pellet was washed once in detergent-free hypotonic buffer, centrifuged at 14,000rpm/5min/4°C and the SN removed. The washed pellet was then resuspended in approximately 2 pellet volumes of nuclear extract buffer, followed by incubation on a nutator for 45min/4°C. The lysate was then centrifuged at 14,000rpm/30min/4°C and the SN (nuclear fraction) transferred to a new microcentrifuge tube. Protein yields were determined by the Bradford Protein Assay as per manufacturer's instructions, calculated from OD_{600} measurements with reference to a BSA standard curve. Typically, 20µg extract was used for subsequent denaturing Tris-glycine SDS-PAGE analysis (**2.4.5.3.**). Samples were stored long term at -80°C.

2.4.4.9. *In vitro* coupled transcription/translation of mammalian proteins using ^{35}S-Met

The TNT ® Quick Coupled Transcription/Translation System was used essentially as per manufacturer's instructions. For a single reaction, 10µL TNT master mix was added to 0.5µL Easy Tag L-^{35}S-Met along with 250µg plasmid DNA (pGEM7Zf(+)-hSIM1-Myc or pGEM7Zf(+)-hARNT) or MQ H_2O as a blank (negative) control in a final volume of 12.5µL. For multiple parallel reactions, a "master mix" of TNT master mix and Easy Tag L-^{35}S-Met was prepared and aliquoted into the required number of tubes, before adding the template DNA or MQ H_2O separately to each tube as required. Tubes were then incubated for 1.5hr/30°C. Efficiency of protein synthesis was determined by removal of 1µL from each reaction, addition of 20µL 1x SDS load buffer, boiling for 10min/70°C and separation by denaturing Tris-glycine SDS-PAGE (**2.4.5.3.**), followed by fixing the gel for 30min/RT in Fixing Solution, soaking the gel for 5min/RT in 10% glycerol, drying onto Whatman paper and exposure to a Phosphorimager. Phosphorimager plates were routinely scanned after approximately 40hr of exposure at 100 microns on an Amersham

Biosciences Typhoon Trio Variable Mode Imager. Samples were stored long term at -80°C.

2.4.4.10. EMSA

2.4.4.10.1. FAM labelling experiments

Reactions were prepared on ice in reduced light conditions, comprising 1x Gel Shift Buffer, 30µg BSA, 1.5µg poly(dI-dC).poly(dI-dC) and 10nM ds probe (**2.4.2.4.1.**). 20-45µg cytoplasmic or nuclear extract was then added (**2.4.4.8.**), or 5µg bacterially expressed and purified HIF-1α 1-241/H_6ARNT 1-362 in Protein Storage buffer as a positive control (kindly donated by A. Chapman-Smith). For supershifting experiments, the extract had first been mixed with 0.2µg or 0.5µg antibody (4A6 αMyc, H-300 αMEF-2 or M-165 αARNT2) and incubated on ice for 10min. Finally, NaCl was added to those reactions containing cytoplasmic extract, to adjust the final salt concentration to ~0.1M. The total reaction volume was 20µL. Reactions were incubated on ice for 30min before loading onto a pre-electrophoresed TGE gel (**2.4.5.2.**).

A free probe sample was always loaded for each of the different probe types used in each experiment, consisting of the required amount of ds probe with 1µL Bromophenol Blue and 5µL 80% glycerol added. Gels were run at 200V/1.5hr/4°C and then scanned at 200 microns on an Amersham Biosciences Typhoon Trio Variable Mode Imager.

2.4.4.10.2. ^{32}P labelling experiments

For initial experiments using increasing amounts of 293 T-Rex stable cell nuclear extract: Reactions were prepared on ice, comprising 1x Gel Shift Buffer and BSA(µg):extract(µg) and poly(dI-dC).poly(dI-dC)(µg):extract(µg) ratios of 4:1 and 1:7 respectively. 5-20µg nuclear extract was then added, along with 60,000cpm of labelled probe, in a total volume of 20µL (**2.4.4.8.** and **2.4.2.4.2.**). Reactions were incubated for 20min/RT before loading onto a pre-electrophoresed TGE gel (**2.4.5.2.**). Gels were run at 100V/1hr/4°C, then dried onto Whatman paper and exposed to a Phosphorimager.

For subsequent experiments using N39 and 293 T-Rex stable cell extracts:
Reactions were prepared on ice, comprising 1x Gel Shift Buffer, 10µg BSA, and 1.5µg poly(dI-dC).poly(dI-dC). 30µg cytoplasmic or nuclear extract was then added, along with 60,000cpm of labelled probe (**2.4.4.8.** and **2.4.2.4.2.**). Finally, NaCl was added to those reactions containing cytoplasmic extract, to adjust the final salt concentration to ~0.16-0.2M. The total reaction volume was 20µL or 25µL. Reactions were incubated for 20min/RT before loading onto a pre-electrophoresed TGE or TBE gel (**2.4.5.2.**). Gels were run at 100V/1-1.5hr/4°C, then dried onto Whatman paper and exposed to a Phosphorimager.

For subsequent experiments using *in vitro* transcribed and translated proteins:
Reactions were prepared on ice, comprising 1x Gel Shift Buffer, 0.1M KCl, and 10µg/mL poly(dI-dC).poly(dI-dC). 8µL blank lysate or 4µL each of *in vitro* transcribed and translated hSIM1-Myc and hARNT (that had been previously allowed to dimerise on ice for 1hr) was then added, along with 30,000cpm of labelled probe, in a total volume of 20µL (**2.4.4.9.** and **2.4.2.4.2.**). Samples were incubated for 30min on ice before loading onto a pre-electrophoresed TGE or TBE gel (**2.4.5.2.**). Gels were run at 100V/1-1.5hr/4°C, then dried onto Whatman paper and exposed to a Phosphorimager.

A free probe sample was always loaded for each of the different probe types used in each experiment. This sample consisted of the required amount of ds probe with 1µL Bromophenol Blue and 5µL 80% glycerol added. Dried gels were scanned after approximately 40hr of exposure at 200 microns on an Amersham Biosciences Typhoon Trio Variable Mode Imager.

2.4.4.11. Reporter gene studies in 24 well tray format

2.4.4.11.1. Transfection

293 T-Rex stable cell lines were counted (**2.4.4.4.**) and seeded into 24 well trays at $1.7x10^5$ cells/well. After ON incubation at 37°C/5%CO_2, transfections were performed in triplicate with Fugene 6 at a 3:1 ratio of Fugene:DNA as per manufacturer's instructions. For initial experiments transfecting cells with increasing amounts of hARNT and hARNT2 expression plasmid, 293 T-Rex hSIM1-Myc WT.1 cells were transfected with

0, 10, 20, 50, 100 or 200ng pEF-hARNT-Ha-IRES-neo or pEF-hARNT2-IRES-neo expression plasmid, made up to 200ng with pEF-IRES-neo where required, as well as 400ng pML-6CWT and 0.5ng phRL-CMV per well. Empty vector cells were transfected with 200ng pEF-hARNT-Ha-IRES-neo or pEF-hARNT2-IRES-neo expression plasmid, or 200ng pEF-IRES-neo, as well as 400ng pML-6CWT and 0.5ng phRL-CMV per well. Subsequent studies involving analysis of hSIM1 mutant cell lines involved transfection of 400ng pML-6CWT, 0.5ng phRL-CMV, and 20ng pEF-hARNT-IRES-neo or 50ng pEF-hARNT2-IRES-neo per well. Transfected cells were left for 6hr at 37°C/5%CO_2 before induction of hSIM1-Myc expression for a further 16hr with doxycycline (**2.4.4.3.1.**).

2.4.4.11.2. Dual luciferase activity assay

After 16hr doxycycline treatment, medium was aspirated off cells, and cells were lysed via addition of 100µL 1x Passive Lysis Buffer/well followed by vigorous shaking for 15min/RT. 10µL lysate/well was then analysed for Firefly and *Renilla* luciferase activities using the Dual Luciferase Assay System as per manufacturer's instructions, in 96 well tray format on a Promega Glomax Luminometer. 50µL each of LARII and Stop'n'Glo were injected per well, with a luminescence measurement taken prior to the first injection to control for background luminescence. This initial background value was subtracted from the subsequent Firefly luciferase activity value to give a corrected Firefly luciferase activity measurement. Corrected Firefly luciferase activity was then normalised to *Renilla* luciferase activity for each well, and the three normalised values for each triplicate were averaged. For those studies which analysed the activity of hSIM1 mutant cell lines relative to two independently derived WT hSIM1 cell lines, the corrected Firefly/*Renilla* values for both WT hSIM1 cell lines (6 wells) were combined into a single average value. These figures were then converted to log values for the purposes of statistical analysis. Empty vector, WT hSIM1 and mutant hSIM1 cell lines were assayed in triplicate in at least three independent experiments.

Lysates were also used for denaturing Tris-glycine SDS-PAGE (**2.4.5.3.**) and Western analysis (**2.4.4.12.**) to detect the presence of induced and/or transfected proteins. Usually, 10µL lysate/well was analysed. Lysates were stored long term at -80°C.

2.4.4.11.3. Statistical analysis

Univariate ANOVA was performed on the natural log values generated for the empty vector cell line and each mutant hSIM1 cell line with reference to WT hSIM1 using SPSS 17.0. Separate analyses were performed for hARNT- and hARNT2-transfected cells.

2.4.4.12. Western blotting

2.4.4.12.1. Protein transfer to nitrocellulose membrane

Following sample separation by denaturing SDS-PAGE (**2.4.5.3.**), proteins were transferred to nitrocellulose using Wet Transfer Buffer and a Bio-Rad Mini Trans-Blot Electrophoretic Transfer Cell as per manufacturer's instructions. Transfers were performed at 250mA/2hr/4°C for a single tank (500mA/2hr/4°C for two tanks from the same power pack).

2.4.4.12.2. Ponceau staining

Efficiency of transfer was always confirmed by incubating the nitrocellulose membrane with Ponceau stain for 2min/RT with rocking immediately after transfer. The membrane was then repeatedly washed with RO H_2O until background staining was reduced and protein bands clearly visible. Remaining stain was subsequently removed during incubation in Western Blocking Solution.

2.4.4.12.3. Blocking and antibody incubations

Membranes were incubated in Western Blocking Solution for 30min-1hr/RT with rocking, followed by incubation with primary antibody solution ON/4°C with rocking.

The following primary antibody solutions were used:

4A6 αMyc	1:1000-1:5000 of 1mg/mL antibody stock in 1% (w/v) skim milk powder in PBST
M-165 αARNT2	1:250 (endogenous)-1:5000 (overexpressed) of 200µg/mL antibody stock in 1% (w/v) skim milk powder in PBST
MCA78G αalpha tubulin	1:50,000 of 1mg/mL antibody stock in 1% (w/v) skim milk powder in PBST

M2 αFlag	1:50,000 of 0.5mg/mL antibody stock in 3% (w/v) skim milk powder in PBST
MW6 αSIM1	1:500 of neat antibody stock in 1% (w/v) skim milk powder in PBST
222-6 αU2AF65	1:1000 of 0.5mg/mL antibody stock in 2% (w/v) skim milk powder in PBST

Following primary antibody incubation the blot was washed 3x in PBST for 5min/RT with rocking, then incubated with the appropriate secondary antibody solution for 1hr/RT with rocking, followed by a further 3x washes in PBST for 5min/RT with rocking.

The following secondary antibody solutions were used:

Immunopure ® Goat Anti-Mouse IgG (Thermo Scientific)	1:20,000 dilution of 0.4mg/mL antibody stock in 1% (w/v) skim milk powder in PBST
Immunopure ® Goat Anti-Rabbit IgG (Thermo Scientific)	1:20,000 dilution of 0.4mg/mL antibody stock in 1% (w/v) skim milk powder in PBST
Polyclonal Rabbit Anti-Rat Immunoglobulins/HRP (Dako)	1:10,000 of 1.3mg/mL antibody stock in 1% (w/v) skim milk powder in PBST
Rabbit IgG SmartBlot-HRP (Vicgene)	1:1000 dilution of neat antibody stock in 5% (w/v) skim milk powder in PBST

2.4.4.12.4. Chemiluminescent detection of proteins

After the final washing step, the membrane was blotted dry between two pieces of Whatman paper and treated with the Immobilon Western Chemiluminescent HRP Substrate. Equal volumes of luminol and peroxide solutions were mixed and pipetted directly onto the membrane surface and allowed to incubate for 2min/RT. The membrane was then blotted dry between two pieces of Whatman paper and exposed to X-Ray film.

2.4.4.13. Immunohistochemistry of monolayer cultures

2.4.4.13.1. Fixing and permeabilisation

Cells were counted (**2.4.4.4.**) and seeded into 8 well chamber slides at $2x10^4$ cells/well (N4 and N7 cells), $0.5x10^4$ cells/well (N39 cells) or $3x10^4$ cells/well (293 T-Rex stable cell lines). After ON incubation at 37°C/5%CO_2, 293 T-Rex cells were treated with doxycycline for 16hr (**2.4.4.3.1.**). Otherwise, medium was removed and cells gently washed in PBST. Cells were then fixed in fresh 4% (w/v) PFA and incubated for 20min/RT, before careful removal of the PFA and 3x washes in PBST for 10min/RT with rocking. Cells were then permeabilised via addition of 0.2% (v/v) Triton X-100 in PBS, followed by incubation for 10min/RT and 4x washes in PBST for 5min/RT with rocking.

2.4.4.13.2. Blocking and antibody incubations

Cells were incubated in Immunofluorescence Blocking Solution for 1hr/RT with rocking, followed by incubation with primary antibody solution ON/4°C with rocking, in a 10cm dish wrapped in parafilm to prevent condensation.

The following primary antibody solutions were used:

4A6 αMyc	1:500 of 1mg/mL antibody stock in 10% (v/v) horse serum in PBST
Z0344 αGFAP	1:250 of 4.1mg/mL antibody stock in 10% (v/v) horse serum in PBST
25/NESTIN αNestin	1:250 of 250µg/mL antibody stock in 10% (v/v) horse serum in PBST
NS-1 αRIP	1:5 of neat ascites fluid in 10% (v/v) horse serum in PBST

Following primary antibody incubation cells were washed 3x in PBST for 5min/RT with rocking, then incubated with the appropriate secondary antibody solution for 2hr/RT with rocking, followed by a further 3x washes in PBST for 5min/RT with rocking. After addition of the secondary antibody solution, exposure to light was minimised by wrapping the chamber slides in aluminium foil.

The following secondary antibody solutions were used:

Alexa Fluor ® 594 Donkey Anti-Rabbit IgG (H+L)	1:1000 dilution of 2mg/mL antibody stock in 10% (v/v) horse serum in PBST

Alexa Fluor ® 594 Donkey Anti-Mouse IgG (H+L)	1:1000 dilution of 2mg/mL antibody stock in 10% (v/v) horse serum in PBST

2.4.4.13.3. Nuclear staining, mounting and photographing

DAPI was added to the final PBST wash to a final concentration of 2µg/mL. The chamber wells were then detached from the slide, and the slide allowed to air dry. One drop of Fluorescent Mounting Medium was then added on top of each "well", and the slide mounted with a coverslip and sealed with nail polish. The slide was allowed to dry fully before photographing with a TE-FM Epi-Fluorescence Attachment (Nikon) attached to an Eclipse T300 Inverted Microscope (Nikon).

2.4.4.14. Immunoprecipitation

293 T-Rex stable cell lines were treated for 16hr with doxycycline (**2.4.4.3.1.**) and WCEs prepared (**2.4.4.7.**). 250µg WCE was then mixed with 1µg ab9106 αMyc antibody and diluted to 100µL in ice-cold IP buffer. Lysates were incubated for 1hr/4°C with gentle rocking. Meanwhile, the required volume of rec-PGS 4B slurry was blocked in 0.5mg/mL BSA in IP buffer for 30min/4°C with gentle rocking. Beads were pelleted, SN removed and the pellet washed three times in ice-cold IP buffer before final resuspension of the rec-PGS 4B pellet in an equal volume of IP buffer. 30µL of this 50% rec-PGS 4B slurry was then added to each tube and samples incubated for a further 2hr/4°C with gentle rocking. Beads were pelleted, SN removed and the pellet washed three times in ice-cold IP buffer before elution of immune complexes in 25µL 2x SDS load buffer. Samples were boiled for 2min/95°C, separated by denaturing Tris-glycine SDS-PAGE (**2.4.5.3.**), and analysed by Western blot (**2.4.4.12.**). Proteins were detected using the M-165 αARNT2 polyclonal antibody and the 4A6 αMyc antibody. To minimise detection of denatured Ig molecules, the rabbit IgG SmartBlot secondary antibody was used to detect ARNT2 staining.

2.4.5. Gel electrophoresis

2.4.5.1. Agarose gel electrophoresis

Separation of nucleic acids was performed using 1-3% (w/v) agarose melted in 1x TBE gels, with EtBr added prior to setting to a final concentration of 0.75µg/mL. RNAse-free gels were similarly prepared, using RNAse-free stocks of TBE buffer, plugged pipette tips and sterile plasticware. Samples were prepared for electrophoresis by mixing with DNA load buffer to a final concentration of 1x. Gels were electrophoresed at 80-130V/45-90min/RT.

2.4.5.2. Non-denaturing PAGE

EMSA midi gels were poured using the ADELAB large OWL Electrophoresis apparatus. Mini gels were poured using the Bio-Rad Mini-PROTEAN II gel system. Prior to use, the glass plates, spacers and comb were cleaned by soaking in a 10% (w/v) Pyroneg solution for approximately 30min/RT, followed by thorough rinsing in RO H_2O and cleaning with 100% EtOH using Kimwipes and powder-free gloves.

Gel solutions were prepared in autoclaved glass beakers using autoclaved glass pipettes. For TGE gels, the gel solution comprised 7% (v/v) acrylamide, 5% (v/v) glycerol, 0.06% (w/v) APS and 0.1% (v/v) TEMED in 1x TGE buffer. For TBE gels, the gel solution comprised 7% (v/v) acrylamide, 5% (v/v) glycerol, 0.06% (w/v) APS and 0.1% (v/v) TEMED in 0.5x TBE buffer. Gels were pre-electrophoresed at 200V/40min/4°C (for midi gels) or 100V/40min/4°C (for mini gels), then samples (**2.4.4.10.**) loaded and electrophoresed at 200V/1.5-2hr/4°C (for midi gels) or 100V/1-1.5hr/4°C (for mini gels).

2.4.5.3. Denaturing SDS-PAGE

Denaturing SDS gels were used to separate proteins for Western analysis (**2.4.4.12.**) and were prepared using the Bio-Rad Mini-PROTEAN III gel system. Gel solutions were prepared in clean glass beakers using autoclaved glass pipettes.

For Tris-glycine gels, the separation gel comprised 1x Lower Buffer, 7.5% (v/v) acrylamide, 0.08% (w/v) APS and 0.08% (v/v) TEMED. After pouring it was carefully

overlaid with water-saturated butanol and allowed to set at RT. The butanol was then removed and the top of the gel rinsed with RO H_2O before addition of the stacking gel, which comprised 1x Upper Buffer, 4.5% (v/v) acrylamide, 0.08% (w/v) APS and 0.08% (v/v) TEMED. Once the gel had set, the gel running apparatus was assembled and the tank filled with fresh 1x GTS running buffer. Samples (**2.4.4.7.** and **2.4.4.8.**) were prepared for electrophoresis by mixing with Tris-glycine-SDS load buffer to a final concentration of 1x and boiling for 2min/95°C prior to loading. Gels were routinely electrophoresed at 110V/60-90min/RT.

For Tris-tricine gels, the separation gel comprised 1M Tris pH 8.3, 10% (v/v) acrylamide, 0.09% (w/v) APS and 0.09% (v/v) TEMED, and the stacking gel 1.26M Tris pH 8.3, 3% (v/v) acrylamide, 0.08% (w/v) APS and 0.08% (v/v) TEMED. Gels were poured as per Tris-glycine gels. Once the gel had set, the gel running apparatus was assembled and the tank filled with fresh 1x Tris/tricine running buffer. Samples (**2.4.4.7.** and **2.4.4.8.**) were prepared for electrophoresis by mixing with Tris-tricine-SDS load buffer to a final concentration of 1x and boiling for 2min/95°C prior to loading. Gels were routinely electrophoresed at 110V/60-90min/RT.

CHAPTER 3: RESULTS

Identification and characterisation of obesity-associated *SIM1* variants

3.1. Screening for disease-associated *SIM1* variants

As mentioned in **Chapter 1**, previous screens for *SIM1* variants statistically and reproducibly associated with severe obesity phenotypes in humans have proven difficult [182, 220-223]. This is despite substantial evidence to suggest that loss of *SIM1* expression and/or function in mammals is likely to have an effect on the ability of the organism to maintain adequate energy balance. This is perhaps unsurprising, given the statistical challenges associated with identifying comparatively rare monogenic disease-associated gene variants within a genetically diverse population that are also unlikely to be transmitted through families for more than a few generations [187, 189]. However, despite their relative rarity, monogenic contributors to severe obese phenotypes remain the most promising candidates for such screens, since they are more likely to survive the statistical rigours of separating natural sequence polymorphisms from genuine pathogenic mutations [187, 229]. For the two separate human genetic studies presented in this thesis, our collaborators adopted unique directed approaches, in which individuals were recruited based not only on the presence of phenotypic and physiological characteristics reminiscent of *Sim1* knockout mouse models, but also similarities between these models and other human syndromic models of severe obesity. The information provided by these studies was then utilised by us for further *in vitro* and cell based functional analyses.

3.1.1. The UK study: a population based screen for *SIM1* variants associated with severe, early onset obesity

The UK GOOS study was recently established to enable population based analyses of the genetic factors that contribute to severe, early onset obesity, and consists of approximately 4000 individuals recruited on the basis of two criteria: a BMI standard deviation score >3, and an established age of onset <10 years[2]. The nucleotide substitutions identified in the current study were obtained from sequence analysis of the

[2] In other words, eligible individuals must achieve a BMI score that deviates from the average for a person of that age by a factor of 3, prior to the age of 10 years.

SIM1 coding region in a subset (1776 Caucasian individuals) of this cohort relative to 1690 ethnically matched population based control individuals (The Isle of Ely Study, [230]). Any individuals possessing *MC4R* mutations or a congenital leptin deficiency were automatically excluded from the study, since the cause of their obesity could already be attributed to altered MC4R or leptin signalling ([200, 209] and S. Ramachandrappa, personal communication). In this way, the study was designed to favour the identification of novel genetic contributors to severe, early onset obesity phenotypes, of which *SIM1* was an initial candidate. This seemed plausible given the already established link between *SIM1* haploinsufficiency and severe weight gain in both mice and humans [59, 96, 177]. Ultimately, we were provided with sixteen different missense mutations to analyse that had been identified in twenty-five unrelated patients (**Figure 3.1.A.**). Three of these were also identified in a control cohort and therefore may represent non-pathogenic *SIM1* sequence polymorphisms (**Figure 3.1.A.**).

3.1.2. The French study: a familial screen for *SIM1* variants associated with a Prader-Willi like phenotype and severe obesity

Prader-Willi syndrome (PWS) is the most common form of syndromic obesity in humans, and is associated with numerous features including severe obesity, hypotonia, mental retardation, craniofacial abnormalities, and developmental delay [231]. It is believed to be the consequence of deletions or abnormalities in chromosome 15 and has also been ascribed to abnormal methylation patterns resulting from imprinting centre defects in this region [182]. Deletions and rearrangements of the long arm of chromosome 6, which harbours the *SIM1* gene, have been consistently associated with Prader-Willi like (PWL) phenotypes for approximately twenty years, suggesting that chromosome 6q abnormalities may also contribute to some of the features of PWS [179-184, 186, 232]. Indeed, analyses have revealed reduced PVN volume and OXT immunoreactivity in the hypothalamus of five PWS patients, indicating that compromised hypothalamic function may be a contributing factor [233]. This has prompted clinicians to encourage screening for chromosome 6q abnormalities, in addition to chromosome 15 abnormalities, when attempting to establish PWS or PWL syndrome in humans [179-182].

The mutations discovered in this study were identified via sequencing of the *SIM1* coding region in 330 French and American children with PWL features (defined as hyperphagic

A UK Genetics of Obesity study (GOOS)

OBESE COHORT

WT allele	Mutant allele	Peptide position	WT peptide	Mutant peptide	Frequency
A	T	51	K	N	1
C	G	71	S	R	1
T	C	128	I	T	4
C	G	152	Q	E	2
G	A	171	R	H	1
T	G	238	L	R	1
A	G	292	T	A	1
C	T	361	T	I	1
A	G	383	R	G	1
C	G	497	P	R	1
C	T	541	S	L	1
G	A	550	R	H	3
C	T	692	P	L	1
G	A	703	R	Q	1
G	C	707	D	H	4
C	T	712	T	I	1

CONTROL COHORT

WT allele	Mutant allele	Peptide position	WT peptide	Mutant peptide	Frequency
T	C	128	I	T	1*
C	G	152	Q	E	1*
G	C	707	D	H	1*

B French study

PWL COHORT

WT allele	Mutant allele	Peptide position	WT peptide	Mutant peptide	Frequency
T	C	128	I	T	1
C	G	152	Q	E	1
G	A	458	E	K	1
A	G	581	R	G	1
A	G	714	T	A	1

OBESE COHORT

WT allele	Mutant allele	Peptide position	WT peptide	Mutant peptide	Frequency
T	C	128	I	T	1*
C	T	323	H	Y	2

CONTROL COHORT

WT allele	Mutant allele	Peptide position	WT peptide	Mutant peptide	Frequency
T	C	128	I	T	1
G	A	458	E	K	1
C	A	515	H	Q	1
A	G	735	N	S	1

Figure 3.1. List of SIM1 point mutations identified by the UK Genetics of Obesity Study (A) and the French study (B). In each case the nucleotide change is indicated, as well as the resultant amino acid substitution in the SIM1 peptide sequence. **Peptide position** refers to the location of the amino acid within the SIM1 protein sequence. **Frequency** refers to the number of different patients within the cohort that possessed that particular nucleotide substitution. An asterisk (*) indicates that the number of patients that possess that substitution has not been fully determined; it has, however, definitely been identified in at least one patient. PWL, Prader-Willi-like.

obesity and at least one other feature indicative of PWL syndrome, including neonatal hypotonia, difficulty to thrive, mental retardation, developmental delay, behavioural problems, skin picking, facial dysmorphism, hypogenitalism or hypogonadism, and an absence of any detectable chromosome 15q or 6q abnormalities), plus an additional 1026 French and American individuals with severe obesity (defined as a history of weight >150% of ideal body weight for height or BMI $\geq 95^{th}$ percentile before the age of 13 years for American subjects, BMI ≥ 40 for French adults, and BMI $\geq 97^{th}$ percentile for age and gender and at least one obese first-degree relative for French children), relative to 1179 control individuals matched for age, gender and ethnicity (F. Stutzmann, personal communication). Patients were therefore selected for this study according to markedly different criteria, encompassing histories of transmitted obesity and/or the existence of physical characteristics reminiscent of syndromic obesity within individuals. Using this approach, five missense mutations were identified in the PWL cohort and two in the severely obese cohort, two of which were also identified in the control cohort (**Figure 3.1.B.**). At the time of writing the French study is ongoing; the variants presented here are therefore the result of initial *SIM1* sequence analyses in all three cohorts.

3.2. Physiological data

Initial metabolic studies in twenty-three of the twenty-five *SIM1* mutation carriers identified in the UK GOOS cohort indicated that, as expected, these individuals cumulatively displayed increased energy intake relative to control (lean) individuals, confirming that they are indeed hyperphagic (**Figure 3.2.A.**). Energy intake was similar to that observed for *MC4R*-deficient patients, although not as severe as that exhibited by *LEPTIN*-deficient patients (**Figure 3.2.A.**). The same cohort of *SIM1* mutation carriers did not display any deviation in their basal metabolic rate relative to predicted values (**Figure 3.2.B.**). These phenomena correspond to those already observed for *Sim1*$^{+/-}$ mice, which also display severe weight gain due to increased food intake without any corresponding change in energy expenditure [59, 96].

At the time of writing, similar analyses of the French cohort(s) have not been completed.

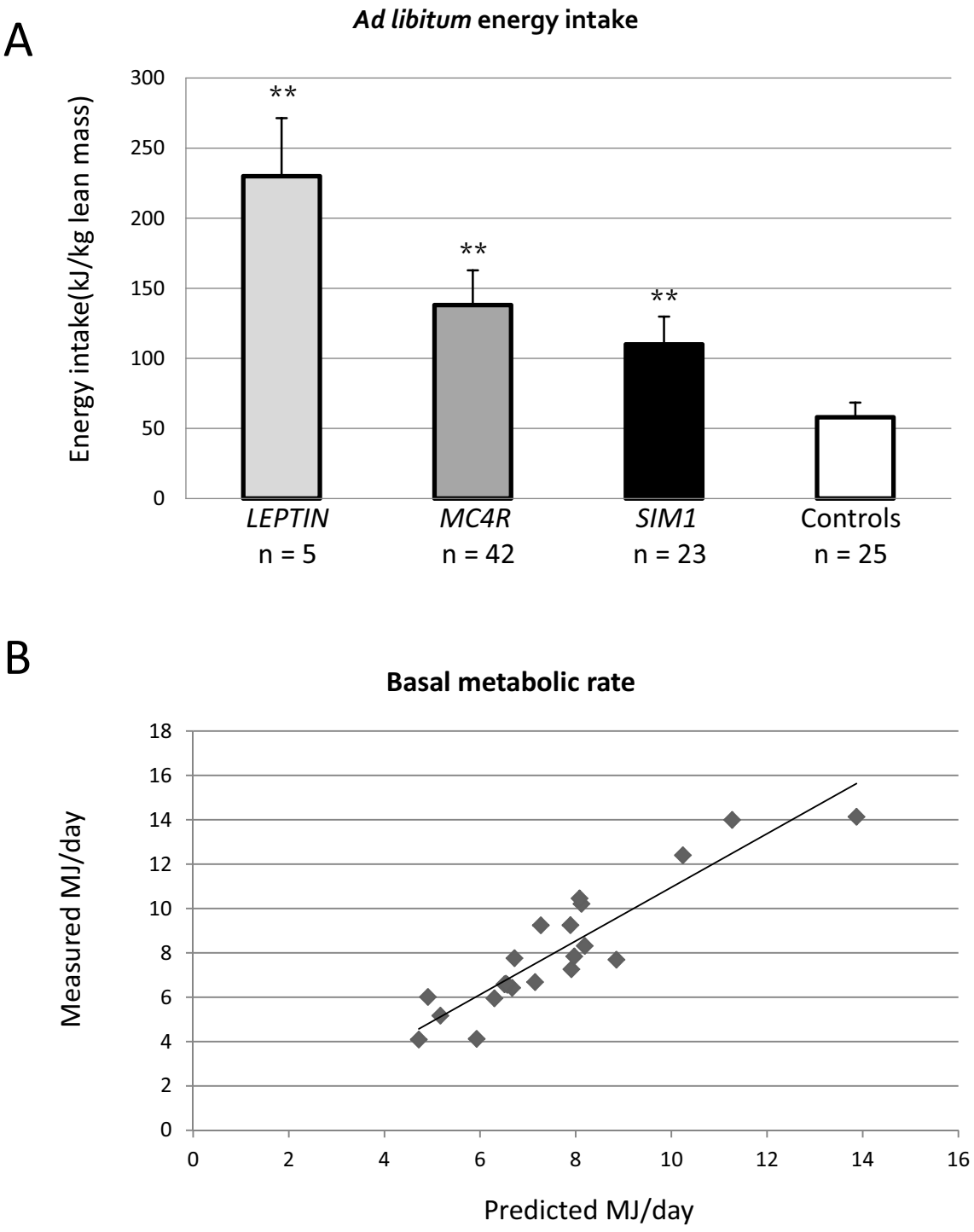
3.3. Functional characterisation of disease-associated *SIM1* variants

Of the sixteen *SIM1* point mutations identified in the UK obese cohort and the six *SIM1* point mutations identified in the French PWL and obese cohorts, only the I128T and Q152E variants are common to both (**Figure 3.1.**). Together, the twenty different point mutations are spread throughout the length of the SIM1 protein and do not cluster within any observable "hot spots", except for the P692L, R703Q, D707H, T712I and T714A mutations, which are located within a linear 23aa region close to the C-terminus (**Figure 3.3.**). All twenty WT residues are conserved between human and mouse SIM1, and are also highly conserved between SIM1 homologues from other species (**Figures 3.3.** and **3.4.**).

In light of what is already known about the domain structure of the SIM1 protein, it can be supposed that pathogenicity may therefore be conferred via disruption of numerous domain-mediated processes, including DNA binding, dimerisation with ARNT and/or ARNT2, and recruitment of specific coregulatory proteins, as well as protein or transcript expression, stability and subcellular localisation (**Figure 3.3.** and **Chapter 1**). Prior to this study, assays for most of these processes had not been reported for SIM1. It was therefore necessary to develop, in the first instance, a functional assay capable of ascertaining how many of the twenty identified mutations could be reproducibly associated with altered SIM1 activity, to which end we successfully developed and optimised a cell-based dual luciferase reporter gene assay system. In addition, we successfully developed a CoIP protocol to assess the relative affinities of some mutants for ARNT and ARNT2, performed immunohistochemical staining on fixed cells to determine their subcellular localisation, and attempted to optimise an EMSA protocol to assess their relative DNA binding affinities. In this way, our studies focused on the molecular biology of the SIM1 protein, and aimed to elucidate how selective disruption of one or more of its biochemical properties could impact on its activity *in vivo*.

3.3.1. Several variants display altered activity in a CME-driven reporter gene assay system

A lack of any well characterised endogenous SIM1 target genes in the last 15 years has necessarily restricted studies of SIM1 function and activity to artificial reporter gene



Data: S. Ramachandrappa

Figure 3.2. *SIM1* mutation carriers display increased *ad libitum* energy intake without any change in basal metabolic rate. A. *Ad libitum* energy intake for *LEPTIN*-, *MC4R*- and *SIM1*-deficient patients and control patients was assessed using a breakfast meal of known macronutrient content after an overnight fast and is expressed as kilojoules (kJ) of energy consumed per kilogram of lean body mass as measured by dual-emission X-ray absorptiometry, to allow comparison between individuals of different body weights and compositions. Results shown are the average energy intake +SEM for each group. P values derived from two-tailed paired t test indicating likelihood of differential energy intake relative to controls: *** <0.001, ** <0.01, * <0.05, ns not significant. **B.** Basal metabolic rate was determined by indirect calorimetry on n= 23 patients after an overnight fast using an open circuit, ventilated, canopy measurement system (Europa Gas Exchange Monitor, NutrEn Technology Ltd.). After adjustment for body composition, basal metabolic rate as measured in megajoules (MJ) per day (diamonds) was plotted against predicted metabolic rate based on age- and sex-specific equations, yielding a line of best fit.

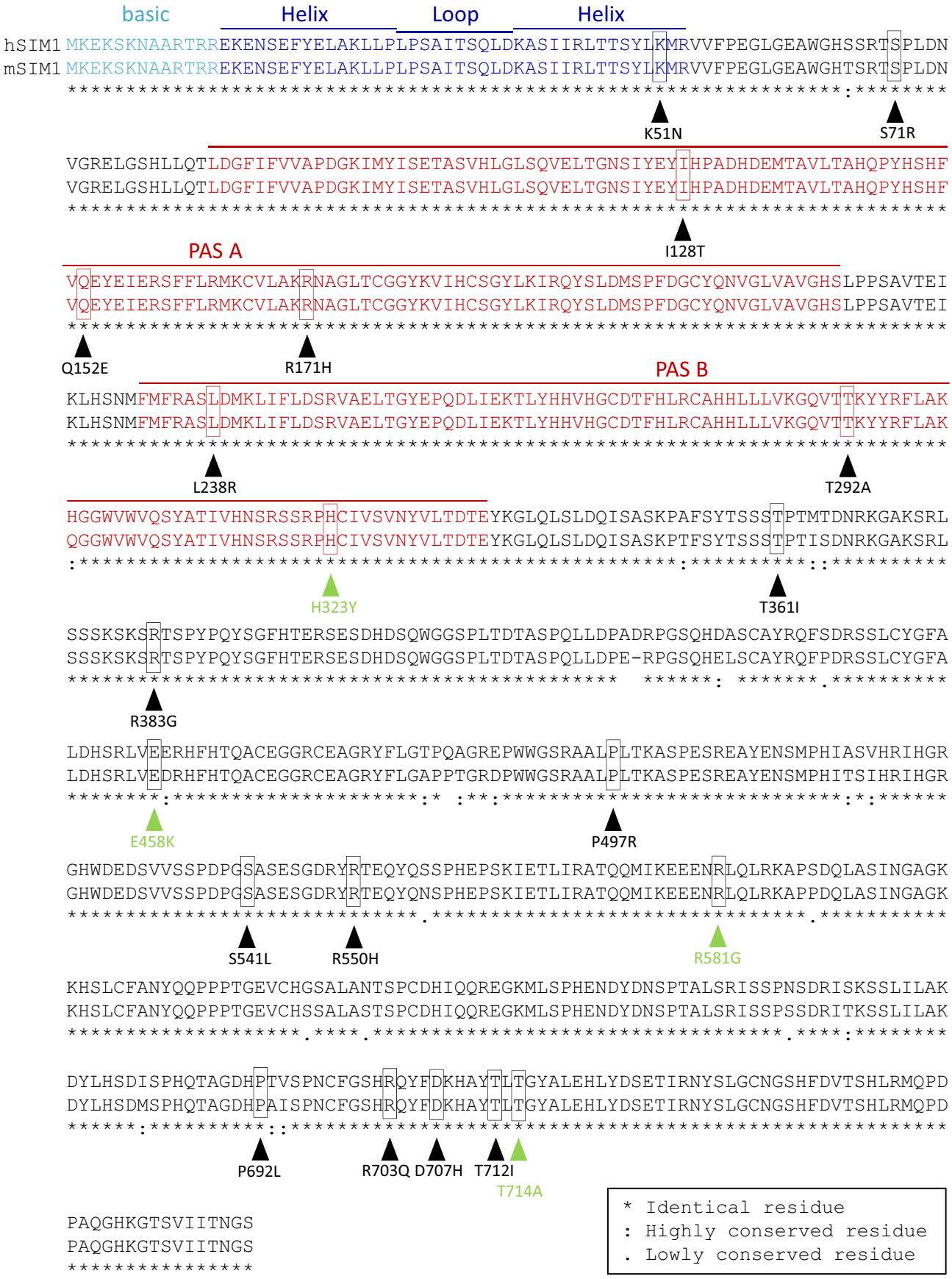
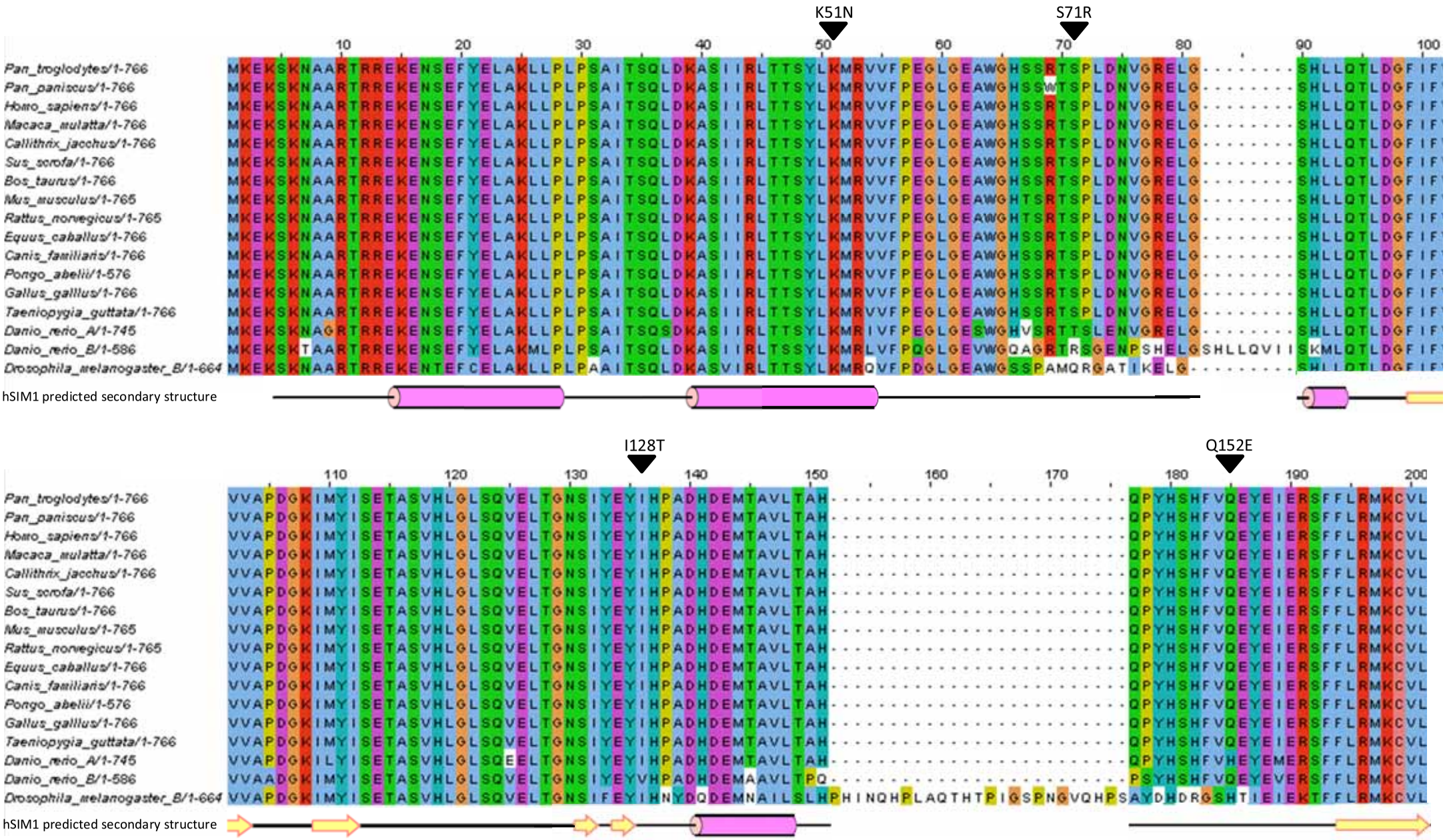
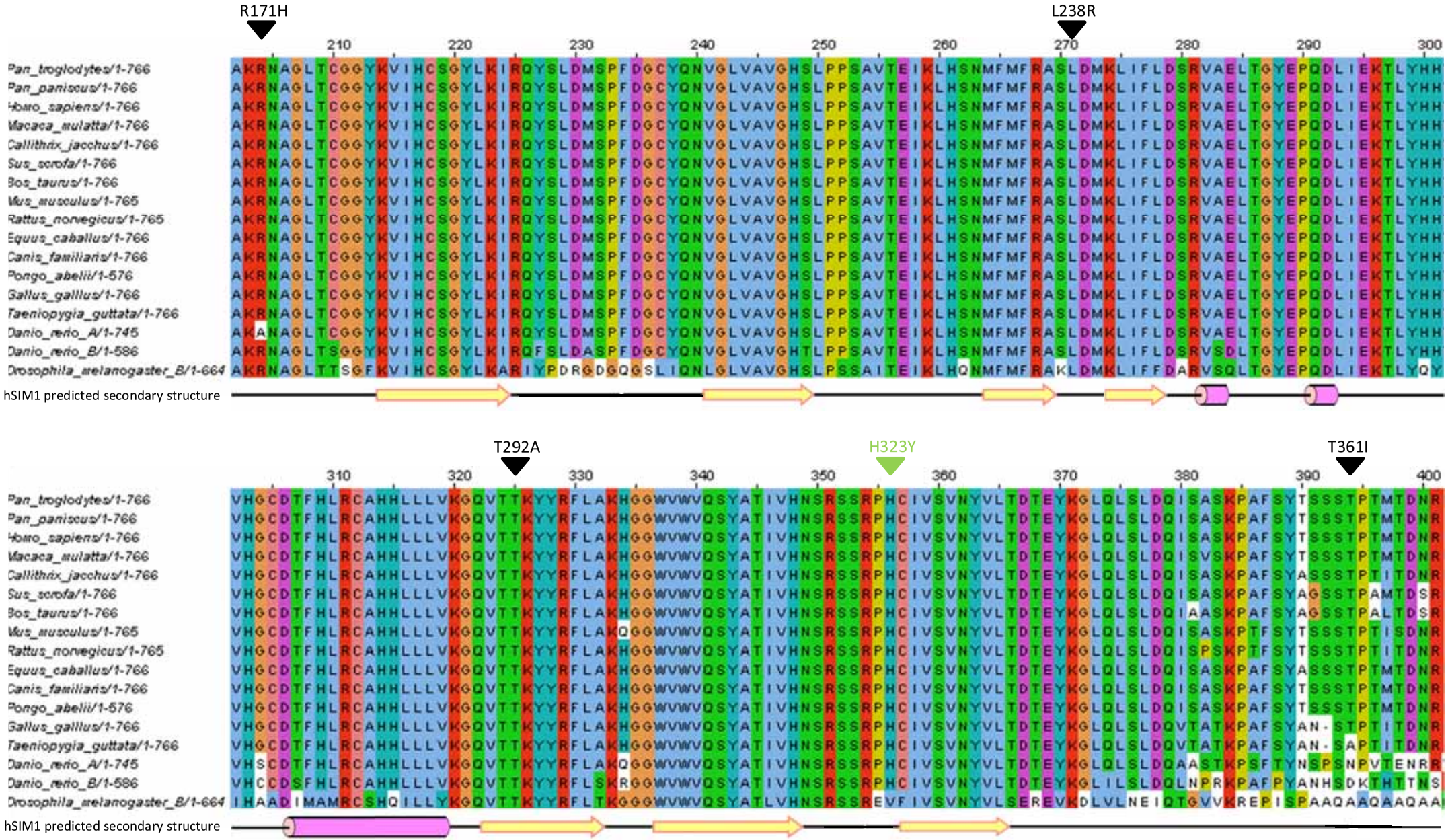
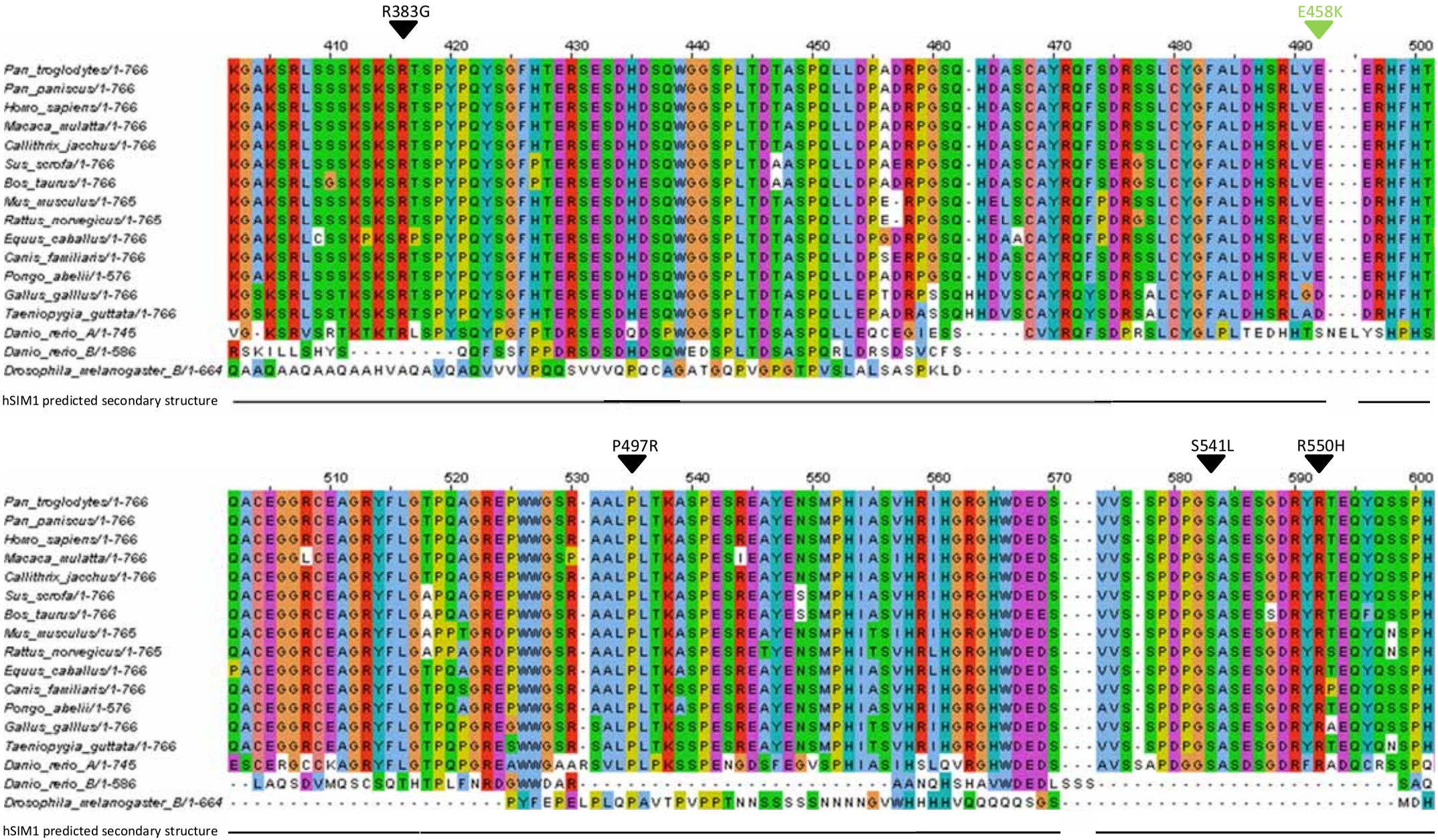
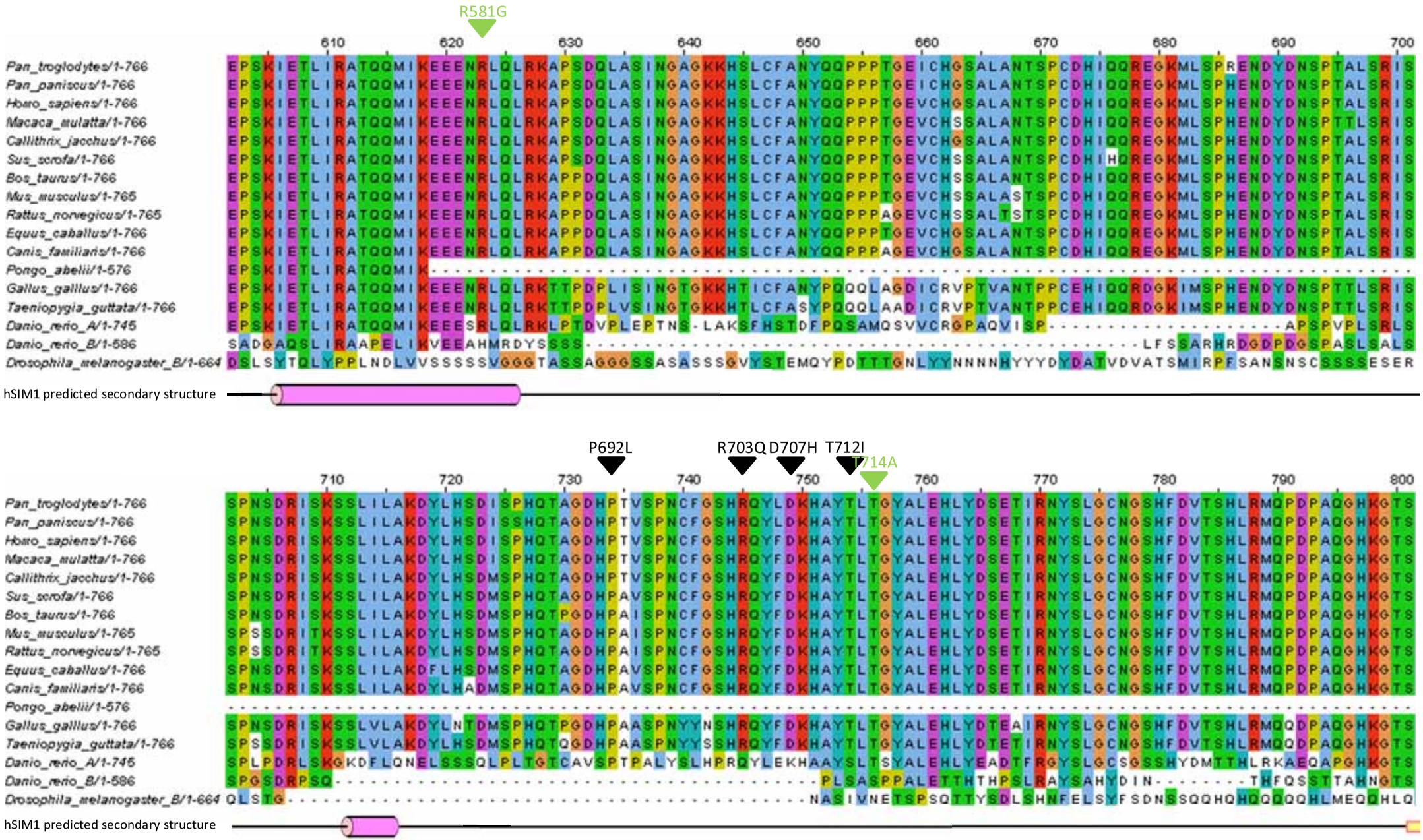


Figure 3.3. Alignment of human and mouse SIM1 protein sequences. Shared amino acid identity at each position is indicated underneath. Domain boundaries are defined with reference to [22]. Protein domains are colour coded as follows: **light blue**, basic domain; **dark blue**, Helix-Loop-Helix domain; **red**, PAS domain; black, inter-domain loop regions and C-terminus. The sixteen SIM1 variants identified by the UK GOOS study are indicated in **black**. Four additional unique variants identified by the French study are indicated in **green**. Protein alignment was performed using ClustalW with Genbank accession numbers NM_005059.2 (hSIM1) and NM_035506.2 (mSIM1).


K51N
S71R
10
20
30
40
50
60
70
80
90
100
Pan_troglodytes/1-766
Pan_paniscus/1-766
Homo_sapiens/1-766
Macaca_mulatta/1-766
Callithrix_jacchus/1-766
Sus_scrofa/1-766
Bos_taurus/1-766
Mus_musculus/1-765
Rattus_norvegicus/1-765
Equus_caballus/1-766
Canis_familiaris/1-766
Pongo_abelii/1-576
Gallus_gallus/1-766
Taeniopygia_guttata/1-766
Danio_rerio_A/1-745
Danio_rerio_B/1-586
Drosophila_melanogaster_B/1-664
hSIM1 predicted secondary structure
I128T
Q152E
110
120
130
140
150
160
170
180
190
200
Pan_troglodytes/1-766
Pan_paniscus/1-766
Homo_sapiens/1-766
Macaca_mulatta/1-766
Callithrix_jacchus/1-766
Sus_scrofa/1-766
Bos_taurus/1-766
Mus_musculus/1-765
Rattus_norvegicus/1-765
Equus_caballus/1-766
Canis_familiaris/1-766
Pongo_abelii/1-576
Gallus_gallus/1-766
Taeniopygia_guttata/1-766
Danio_rerio_A/1-745
Danio_rerio_B/1-586
Drosophila_melanogaster_B/1-664
hSIM1 predicted secondary structure


R171H
L238R
210
220
230
240
250
260
270
280
290
300
Pan_troglodytes/1-766
Pan_paniscus/1-766
Homo_sapiens/1-766
Macaca_mulatta/1-766
Callithrix_jacchus/1-766
Sus_scrofa/1-766
Bos_taurus/1-766
Mus_musculus/1-765
Rattus_norvegicus/1-765
Equus_caballus/1-766
Canis_familiaris/1-766
Pongo_abelii/1-576
Gallus_gallus/1-766
Taeniopygia_guttata/1-766
Danio_rerio_A/1-745
Danio_rerio_B/1-586
Drosophila_melanogaster_B/1-664
hSIM1 predicted secondary structure
T292A
H323Y
T361I
310
320
330
340
350
360
370
380
390
400
Pan_troglodytes/1-766
Pan_paniscus/1-766
Homo_sapiens/1-766
Macaca_mulatta/1-766
Callithrix_jacchus/1-766
Sus_scrofa/1-766
Bos_taurus/1-766
Mus_musculus/1-765
Rattus_norvegicus/1-765
Equus_caballus/1-766
Canis_familiaris/1-766
Pongo_abelii/1-576
Gallus_gallus/1-766
Taeniopygia_guttata/1-766
Danio_rerio_A/1-745
Danio_rerio_B/1-586
Drosophila_melanogaster_B/1-664
hSIM1 predicted secondary structure


R383G
E458K
410
420
430
440
450
460
470
480
490
500
Pan_troglodytes/1-766
Pan_paniscus/1-766
Homo_sapiens/1-766
Macaca_mulatta/1-766
Callithrix_jacchus/1-766
Sus_scrofa/1-766
Bos_taurus/1-766
Mus_musculus/1-765
Rattus_norvegicus/1-765
Equus_caballus/1-766
Canis_familiaris/1-766
Pongo_abelii/1-576
Gallus_gallus/1-766
Taeniopygia_guttata/1-766
Danio_rerio_A/1-745
Danio_rerio_B/1-586
Drosophila_melanogaster_B/1-664
hSIM1 predicted secondary structure
P497R
S541L
R550H
510
520
530
540
550
560
570
580
590
600
Pan_troglodytes/1-766
Pan_paniscus/1-766
Homo_sapiens/1-766
Macaca_mulatta/1-766
Callithrix_jacchus/1-766
Sus_scrofa/1-766
Bos_taurus/1-766
Mus_musculus/1-765
Rattus_norvegicus/1-765
Equus_caballus/1-766
Canis_familiaris/1-766
Pongo_abelii/1-576
Gallus_gallus/1-766
Taeniopygia_guttata/1-766
Danio_rerio_A/1-745
Danio_rerio_B/1-586
Drosophila_melanogaster_B/1-664
hSIM1 predicted secondary structure


R581G
610
620
630
640
650
660
670
680
690
700
Pan_troglodytes/1-766
Pan_paniscus/1-766
Homo_sapiens/1-766
Macaca_mulatta/1-766
Callithrix_jacchus/1-766
Sus_scrofa/1-766
Bos_taurus/1-766
Mus_musculus/1-765
Rattus_norvegicus/1-765
Equus_caballus/1-766
Canis_familiaris/1-766
Pongo_abelii/1-576
Gallus_gallus/1-766
Taeniopygia_guttata/1-766
Danio_rerio_A/1-745
Danio_rerio_B/1-586
Drosophila_melanogaster_B/1-664
hSIM1 predicted secondary structure
P692L
R703Q
D707H
T712I
T714A
710
720
730
740
750
760
770
780
790
800
Pan_troglodytes/1-766
Pan_paniscus/1-766
Homo_sapiens/1-766
Macaca_mulatta/1-766
Callithrix_jacchus/1-766
Sus_scrofa/1-766
Bos_taurus/1-766
Mus_musculus/1-765
Rattus_norvegicus/1-765
Equus_caballus/1-766
Canis_familiaris/1-766
Pongo_abelii/1-576
Gallus_gallus/1-766
Taeniopygia_guttata/1-766
Danio_rerio_A/1-745
Danio_rerio_B/1-586
Drosophila_melanogaster_B/1-664
hSIM1 predicted secondary structure

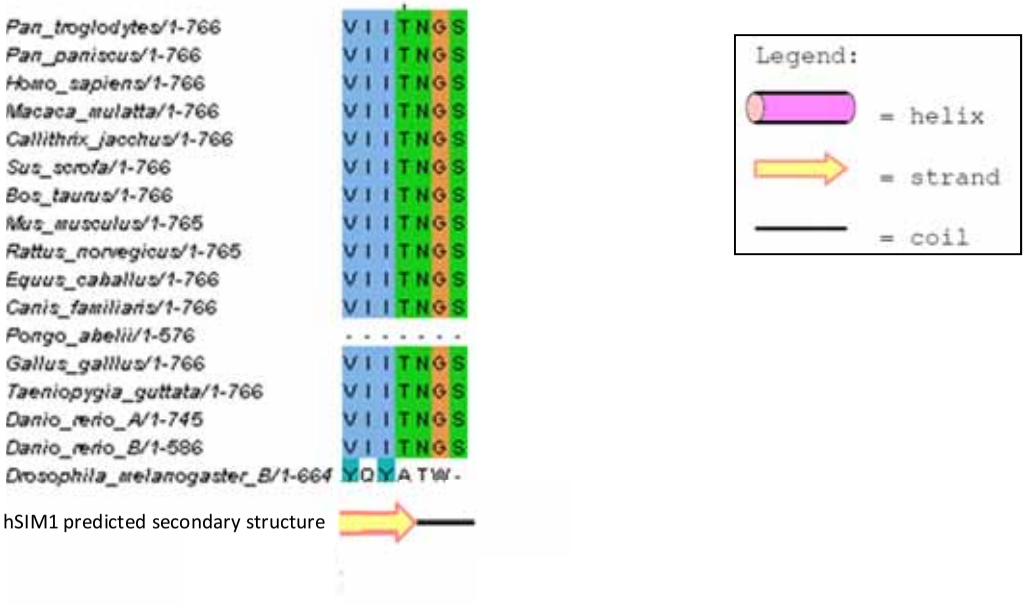


Figure 3.4. Alignment of SIM1 protein sequences from different species. Amino acid identity at each position is indicated using the single letter amino acid code, each of which is assigned a colour. Predicted secondary structure for hSIM1 is shown underneath (see legend). The twenty hSIM1 variants are indicated with arrows along the top of the alignment (UK variants in **black** and French variants in green). Protein alignment was performed using ClustalW with Genbank accession numbers NP_001074957.1 (P. troglodytes SIM1), A1YFY6.1 (P. paniscus SIM1), NP_005059.2 (H. sapiens SIM1), XP_001086054.1 (M. mulatta SIM1), XP_002746908.1 (C. jacchus SIM1), NP_001166056.2 (S. scrofa SIM1), NP_001179901 (B. taurus SIM1), NM_035506.2 (M. musculus SIM1), NP_001101111.1 (R. norvegicus SIM1), XP_001503954.1 (E.caballus SIM1), XP_539058.2 (C. familiaris SIM1), CAH91969.1 (P. abelli SIM1), XP_419817.2 (G. gallus SIM1), XP_002196673.1 (T. guttata SIM1), NP_835740.1 (D.rerio SIM1-A), NP_001153057.1 (D. rerio SIM1-B), and NP_731771.3 (D.melanogaster SIM B), which was then used to generate this image using Jalview. Predicted secondary structure of hSIM1 was generated using PsiPred and adjusted with reference to [22].

assays. At the commencement of this study, the only reported SIM1-responsive construct consisted of a repeated regulatory sequence derived from the direct SIM target gene *toll* (herein referred to as 6xCME), in which context SIM1 behaves as a transcriptional activator with both ARNT and ARNT2 [13]. Using this construct, we optimised a dual luciferase assay protocol that enabled us to distinguish any alteration in activity for nineteen of the twenty SIM1 variants relative to WT SIM1. We chose not to analyse the E458K variant as it had been identified in both the French PWL and control cohorts; since the I128T variant had similarly been identified in obese and control cohorts in both the UK and French studies and therefore did not co-segregate with obesity, we decide that it would suffice as a "control" variant for both the UK and French studies in our analyses.

Initial experiments in which WT and mutant SIM1 proteins were overexpressed with ARNT or ARNT2 in HEK 293T cells, human neuroblastoma SK-N-BE-2C cells and mouse neuroblastoma Neuro-2a cells via transient transfection proved highly inconsistent, due to variability in transfection efficiency and protein expression both within and between experiments (A. Raimondo and S. Ramachandrappa, data not shown). We therefore decided to adopt an alternative experimental approach that gave us greater control of SIM1 expression, by utilising the HEK 293 Flp-In™ T-Rex™ (hereafter referred to as 293 T-Rex) cell system recently developed by Invitrogen. This cell line was designed to enable doxycycline-inducible expression of a protein of interest from a single, pre-defined locus via Flp recombinase mediated genome integration (<http://products.invitrogen.com/ivgn/product/R78007>). Ideally for our purposes, it therefore offered the possibility of homogeneous protein expression between multiple independently generated cell lines, since gene expression in each case would occur from the same, genomically active locus, under the control of the same regulatory elements. The 293 T-Rex cell line endogenously expresses *SIM1*, *ARNT* and *ARNT2* and is therefore a relevant cell line within which to study SIM1 function and activity (**Figure 3.5.A.**). We therefore generated nineteen separate 293 T-Rex stable cell lines inducibly overexpressing one of each of the nineteen SIM1 variants of interest, as well as two independently derived WT SIM1 stable cell lines and a single empty vector cell line, and tested the reproducibility of the 293 T-Rex system via Western blot and reporter gene analyses.

Western analysis of the two independently derived WT SIM1 stable cell lines and nineteen different mutant SIM1 cell lines treated with doxycycline showed that SIM1 protein expression was successfully induced in each line relative to an empty vector cell line, and that protein expression was indeed highly comparable between the two WT SIM1 lines (**Figure 3.5.B.** and **C.**). Protein expression between WT and mutant SIM1 cell lines was also highly similar, indicating that none of the nineteen mutations is likely to have a considerable effect on SIM1 protein stability (**Figure 3.5.B.** and **C.**). Transient transfection of increasing amounts of ARNT or ARNT2 expression plasmid into one of the WT SIM1 cell lines indicated a dose-responsive increase in luciferase reporter gene activity relative to empty vector cells upon doxycycline treatment (**Figure 3.6.A.**). This was confirmed by Western analysis (**Figure 3.6.B.**). Since SIM1/ARNT-mediated gene expression was higher than SIM1/ARNT2-mediated gene expression in this system, subsequent experiments used 20ng transfected ARNT expression plasmid or 50ng ARNT2 expression plasmid per well, in order to achieve a sufficient degree of reporter gene induction to discern differences between WT and mutant SIM1 proteins, whilst remaining within the quantitative range for each heterodimer.

Subsequent analysis of each of the sixteen SIM1 variants identified in the UK obese cohort under these conditions indicated statistically significant and highly reproducible differences in reporter gene activity for nine variants relative to WT in the presence of ARNT, and thirteen variants in the presence of ARNT2 (**Figure 3.7.**). Activity was highly comparable between the two independently generated WT SIM1 cell lines, indicating that these differences are likely to be genuine (**Figure 3.7., inset**). Similar analysis of five SIM1 variants from the French PWL and obese cohorts identified a further two for which activity was also significantly reduced relative to WT, in the presence of both ARNT and ARNT2 (**Figure 3.8.**). Surprisingly, the UK R383G variant displayed a statistically significant increase in activity in the presence of ARNT2, which may be a reflection of its slightly increased expression at the protein level by Western analysis relative to WT SIM1 (**Figure 3.5.C.**). The I128T "control" variant showed no significant difference in activity relative to WT in the presence of ARNT and a comparatively modest decrease in activity in the presence of ARNT2, indicating that the I→T substitution at this position may have a subtle effect on SIM1/ARNT2 transcriptional activity *in vivo* (**Figures 3.7.** and **3.8.**).

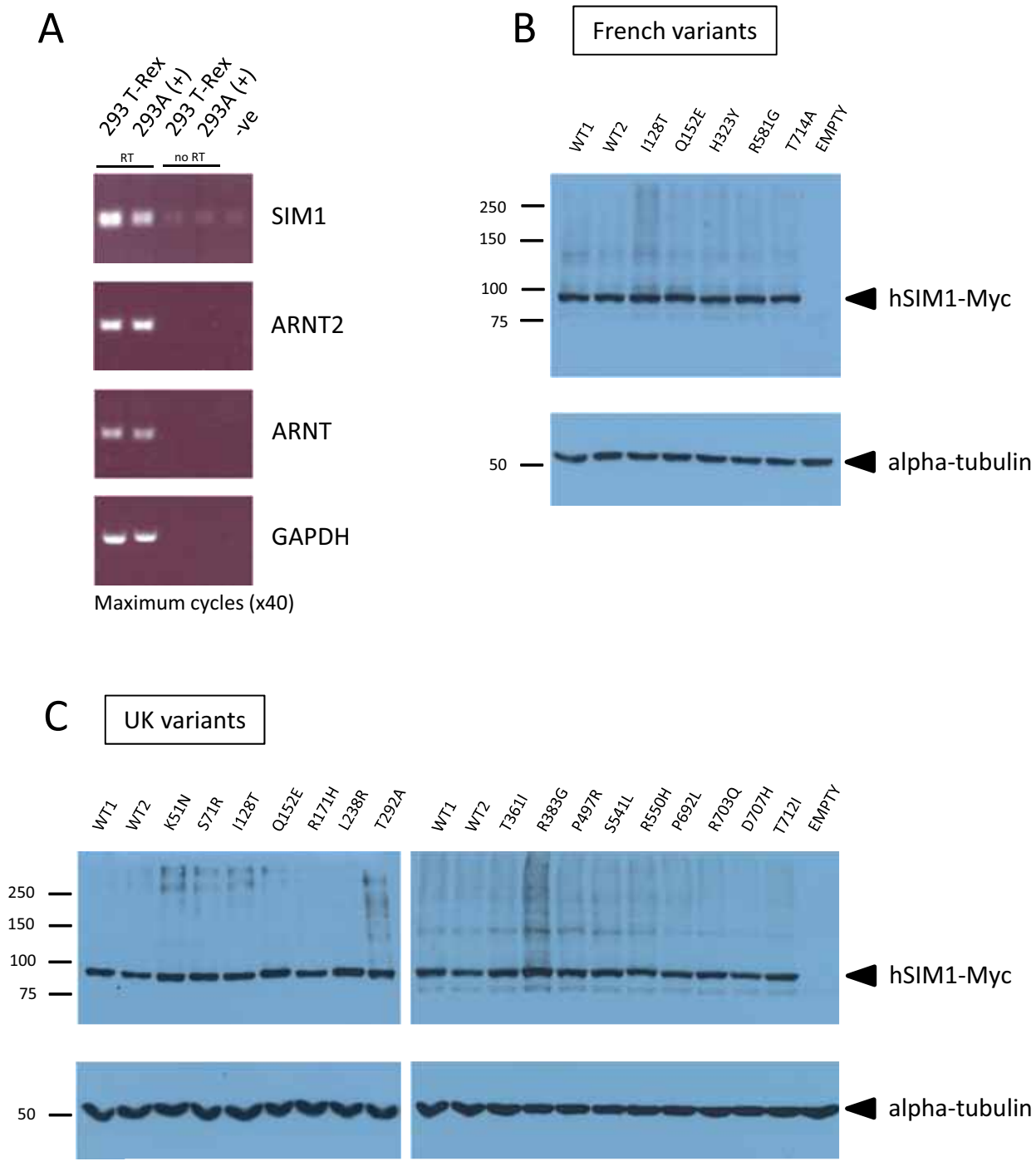


Figure 3.5. Characterisation of SIM1, ARNT2 and ARNT expression and activity in HEK 293 T-Rex parent and stable cell lines. A. cDNA prepared from parent 293 T-Rex cells was assayed via RT-PCR for the presence of endogenous *SIM1*, *ARNT2* and *ARNT*, with reference to the housekeeping gene *GAPDH*. 293A cDNA was used as a positive control and water (-ve) as a negative control. **B.** Western analysis of SIM1 variants identified in the French cohort. Two independently derived wild type (WT) SIM1 293 T-Rex stable cell lines, five different SIM1 mutant cell lines, and one empty vector cell line were treated with doxycycline to a final concentration of 1µg/mL for 16hr. 30µg WCE was subsequently separated by SDS-PAGE, transferred to nitrocellulose, and probed with 4A6 anti-Myc monoclonal antibody and MCA78G anti-alpha-tubulin monoclonal antibody. **C.** Western analysis of SIM1 variants identified in the UK GOOS cohort. Two independently derived WT SIM1 293 T-Rex stable cell lines, sixteen different SIM1 mutant cell lines, and one empty vector cell line were stimulated with doxycycline to a final concentration of 1µg/mL for 16hr. 30µg WCE was subsequently separated by SDS-PAGE, transferred to nitrocellulose, and probed with 4A6 anti-Myc monoclonal antibody and MCA78G anti-alpha-tubulin monoclonal antibody. The two WT cell lines are included as a reference on each blot.

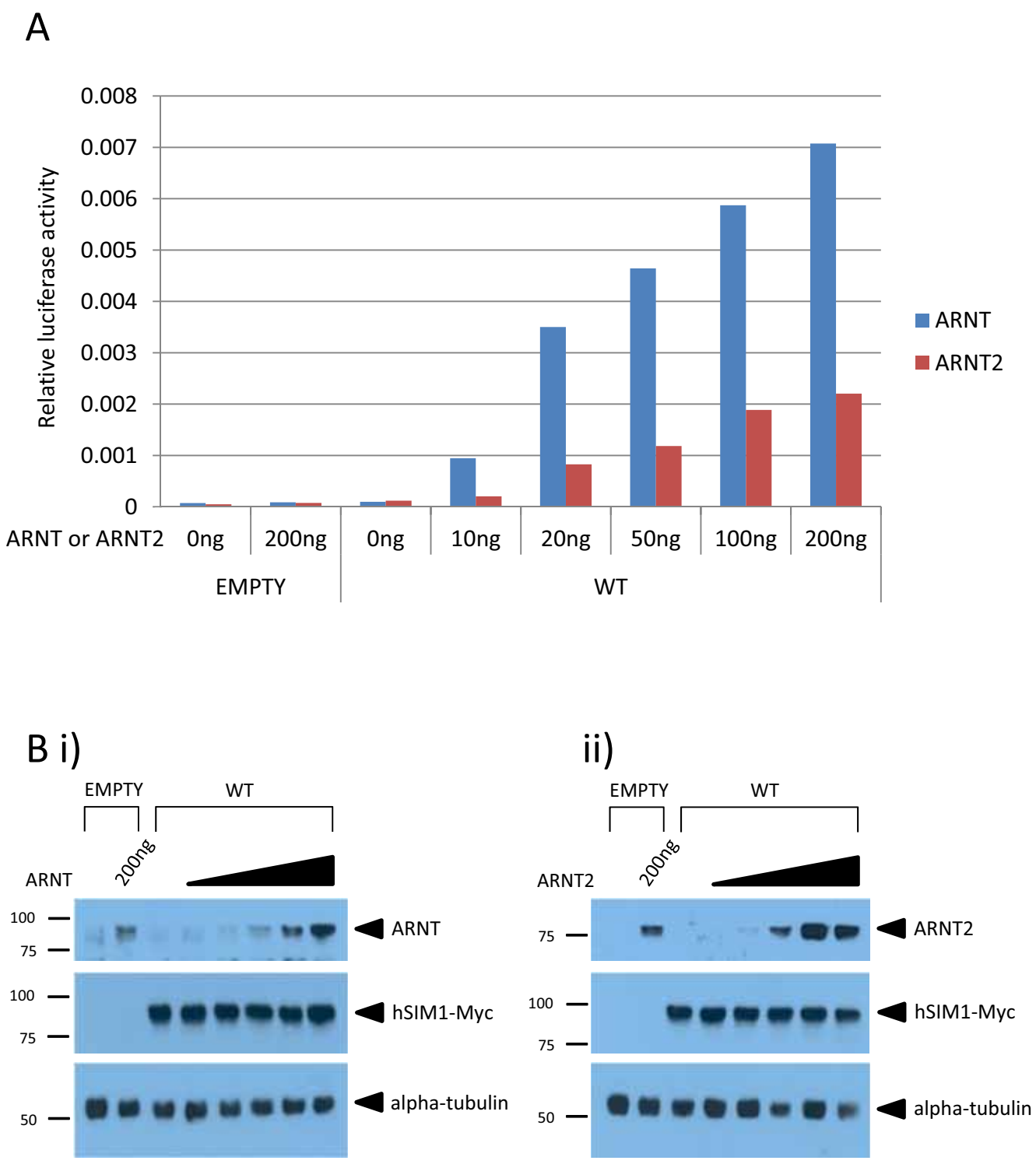


Figure 3.6. Titration of *ARNT* and *ARNT2* expression plasmid to define the quantitative range in wild type SIM1 293 T-Rex stable cells. A single wild type (WT) SIM1 293 T-Rex stable cell line and an empty vector (EMPTY) cell line were transiently transfected with the 6xCME Firefly luciferase reporter plasmid, increasing amounts of *ARNT* or *ARNT2* expression plasmid (0, 10, 20, 50, 100, or 200ng), and a *Renilla* control plasmid. Cells were then treated with doxycycline to a final concentration of 1µg/mL for 16hr before harvesting. **A.** Analysis of luciferase activity. Firefly luciferase activity was normalised to *Renilla* luciferase activity for each well. Results shown are from a single experiment performed in triplicate. **B.** Western blot analysis of Myc-tagged SIM1 and ARNT **(i)** or Myc-tagged SIM1 and ARNT2 **(ii)** overexpression in reporter assay lysates. 10µL lysate from one well of each triplicate was separated by SDS-PAGE, transferred to nitrocellulose, and probed with 4A6 anti-Myc monoclonal antibody, MA1-515 anti-ARNT monoclonal antibody or M-165 anti-ARNT2 polyclonal antibody, and MCA78G anti-alpha-tubulin monoclonal antibody.

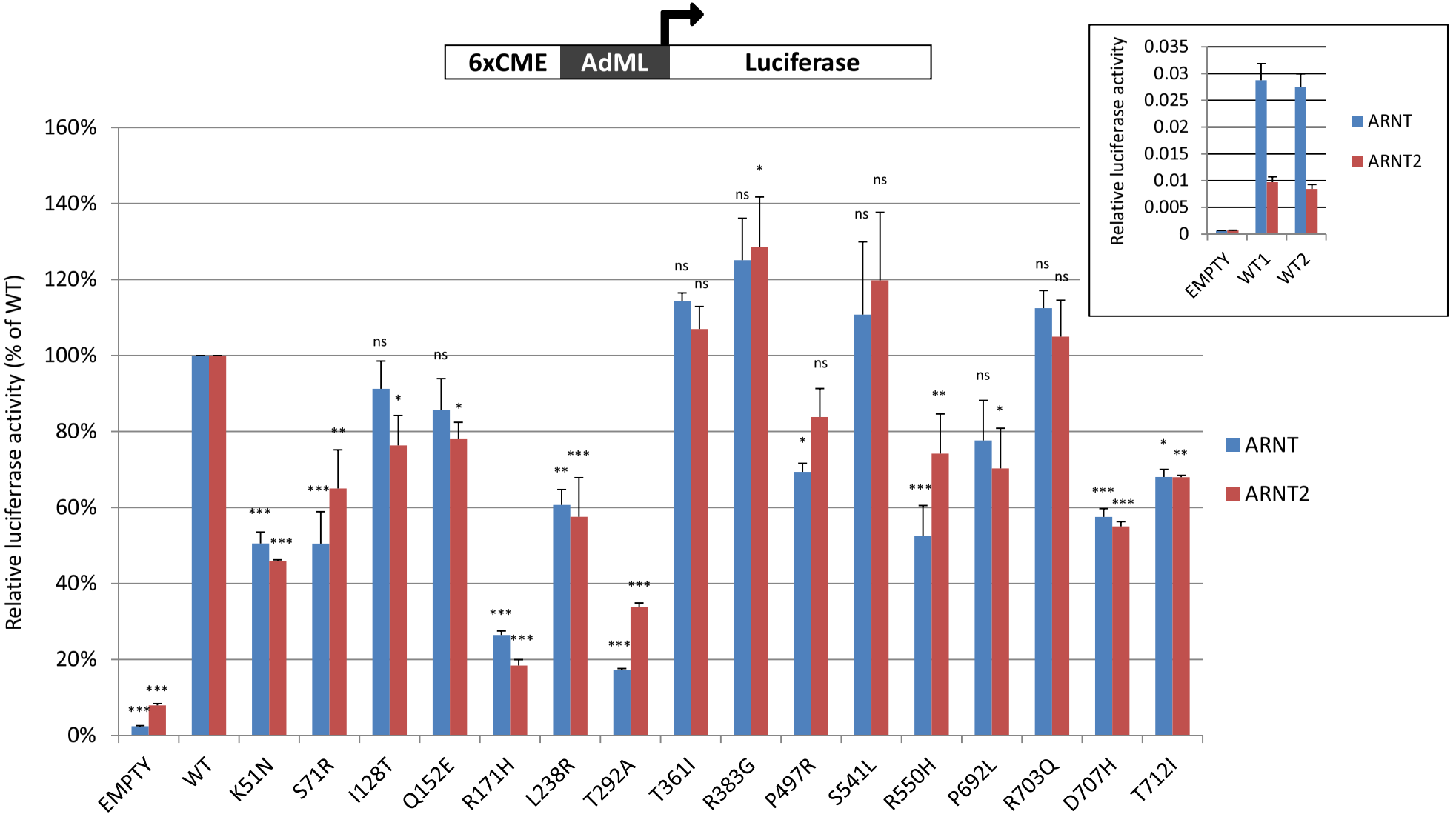


Figure 3.7. Effect of SIM1 point mutations on the transcriptional activity of SIM1 (UK variants). Two independently derived wild type (WT) SIM1 293 T-Rex stable cell lines, sixteen different SIM1 mutant cell lines, and one empty vector cell line were transiently transfected with the 6xCME Firefly luciferase reporter plasmid, *ARNT* or *ARNT2* expression plasmid, and a *Renilla* luciferase control plasmid. Cells were then treated with doxycycline to a final concentration of 1µg/mL for 16hr before harvesting and analysing for luciferase activity. Firefly luciferase activity was normalised to *Renilla* luciferase activity for each well. Results shown are the average of at least three experiments performed in triplicate +SEM expressed relative to the two WT lines, which have been combined into a single average value and normalised to 100% (Inset: results for the two WT lines, prior to normalisation, indicating their activities relative to the empty vector line). P values derived from univariate ANOVA indicating likelihood of differential activity relative to WT: *** <0.001, ** <0.01, * <0.05, ns not significant.

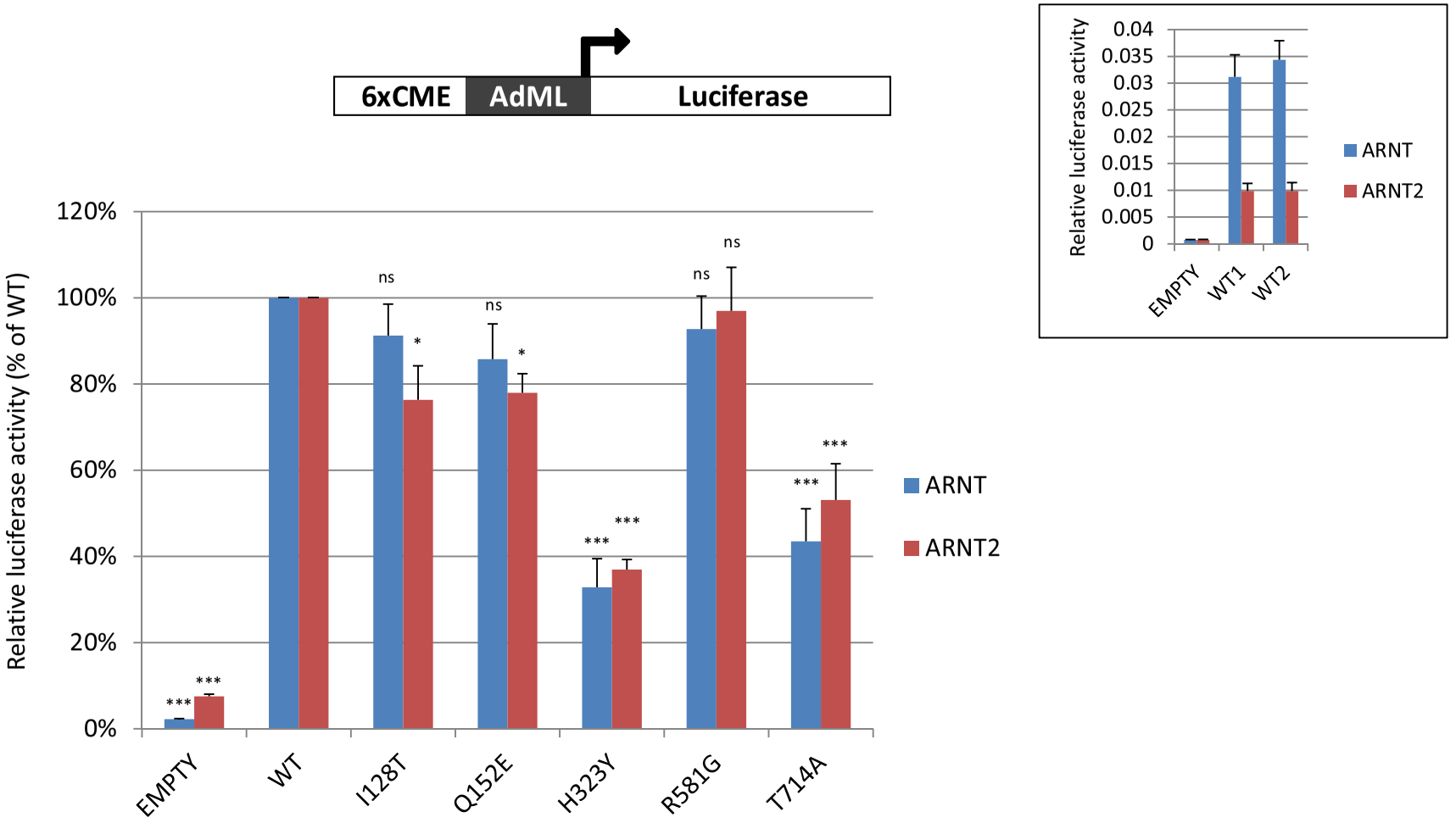


Figure 3.8. Effect of SIM1 point mutations on the transcriptional activity of SIM1 (French variants). Two independently derived wild type (WT) SIM1 293 T-Rex stable cell lines, five different SIM1 mutant cell lines, and one empty vector cell line were transiently transfected with the 6xCME Firefly luciferase reporter plasmid, *ARNT* or *ARNT2* expression plasmid, and a *Renilla* control plasmid. Cells were then treated with doxycycline to a final concentration of 1µg/mL for 16hr before harvesting and analysing for luciferase activity. Firefly luciferase activity was normalised to *Renilla* luciferase activity for each well. Results shown are the average of at least three experiments performed in triplicate +SEM expressed relative to the two WT lines, which have been combined into a single average value and normalised to 100% (Inset: results for the two WT lines, prior to normalisation, indicating their activities relative to the empty vector line). P values derived from univariate ANOVA indicating likelihood of differential activity relative to WT: *** <0.001, ** <0.01, * <0.05, ns not significant.

The T361I, S541L, R703Q and R581G variants were the only ones to display no significant differences in activity relative to WT in the presence of either ARNT or ARNT2 (**Figures 3.7.** and **3.8.**). This result may indicate that they represent non-pathogenic sequence variants present at low frequency in their respective cohorts, or alternatively, that they are genuine mutations of potentially mild effect that may only manifest under certain environmental or genetic conditions that we were unable to account for in our assay (see **Section 1.7.1.**). This may also be the case for the D707H variant, which displayed an approximate 40% reduction in activity in the presence of both ARNT and ARNT2 relative to WT SIM1, yet was identified in both the UK obese and control cohorts (**Figures 3.1.** and **3.7.**). Our reporter assays were not designed to account for additional unknown factors that may influence SIM1 activity, and hence expression of its downstream target genes, *in vivo*. On the other hand, as a quantitative assay designed to assess any changes in activity conferred by the identified amino acid substitutions, our results do indeed support the hypothesis that deleterious *SIM1* sequence variations are associated with obesity in humans, and establish the suitability of the 293 Flp-In™ T-Rex™ cell system for investigating the effect of these amino acid substitutions on SIM1-mediated gene expression. The veracity of our reporter assay results was further substantiated by IHC and CoIP analysis of one of the most inactive variants: T292A.

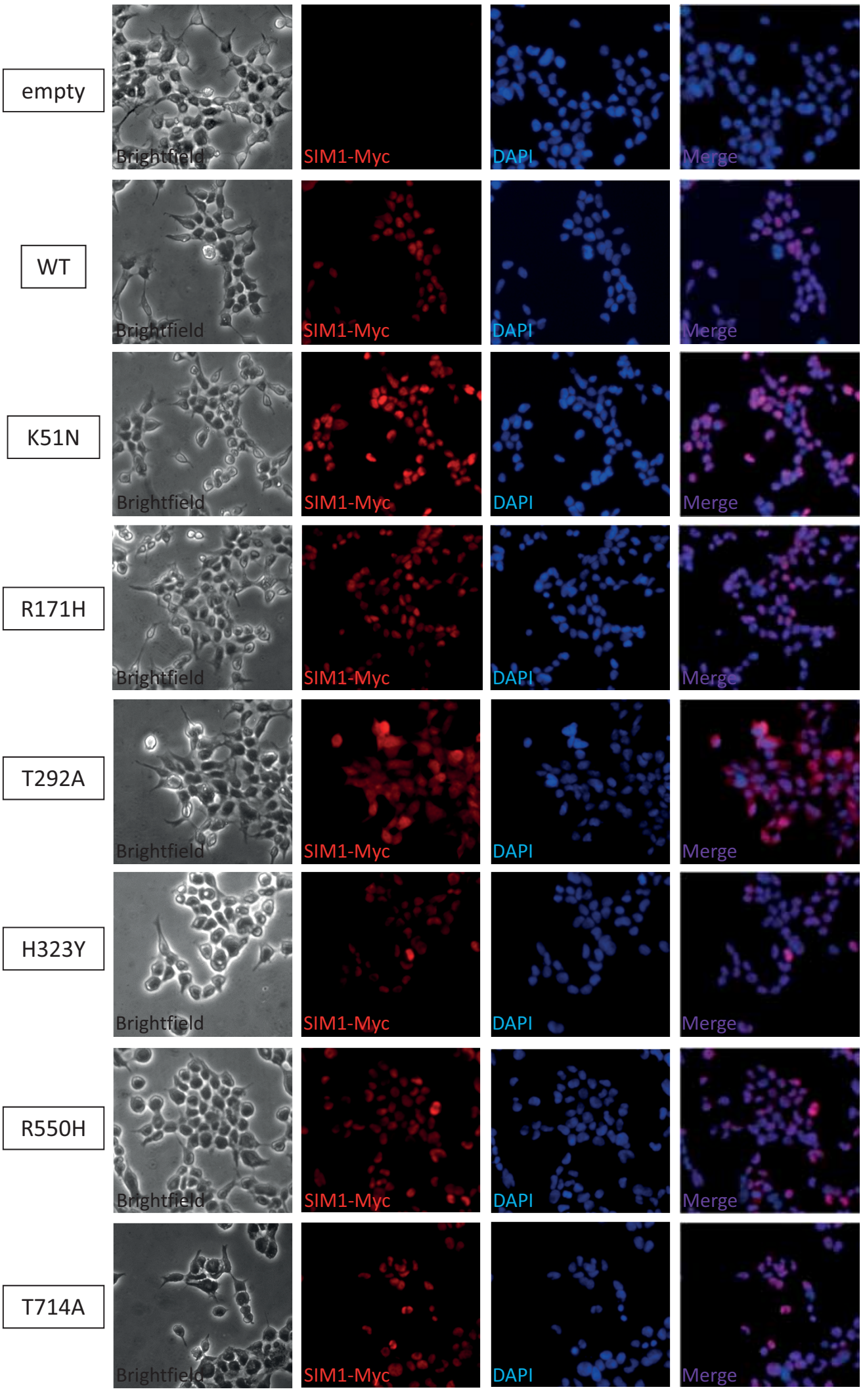
3.3.2. SIM1 T292A displays an altered subcellular localisation pattern

Using our reporter assay data as a guide, we performed immunohistochemical staining on some of the most severely affected variants to determine whether they retained the constitutively nuclear localisation characteristic of WT SIM1 [75]. Interestingly, staining of six different doxycycline-induced mutant SIM1 293 T-Rex stable cell lines indicated an altered localisation pattern for the T292A variant (**Figure 3.9.**). SIM1 T292A displayed a more diffuse expression pattern relative to WT SIM1, being noticeably present in the cytoplasm as well as the nucleus, including the cytoplasmic processes characteristic of HEK 293 cells (**Figure 3.9.**). The K51N, R171H, H323Y, R550H and T714A variants all displayed a nuclear accumulation similar to WT SIM1 (**Figure 3.9.**).

3.3.3. SIM1 T292A is reduced in its ability to dimerise with ARNT and ARNT2

Further investigation indicated that the T292A variant was severely deficient in its ability to CoIP both ARNT2 and ARNT (**Figure 3.10.** and data not shown). We were unable to

Figure 3.9. SIM1 T292A displays an altered subcellular localisation pattern. Empty vector, wild type (WT) SIM1, or mutant SIM1 293 T-Rex stable cell lines were seeded onto coverslips and treated with doxycycline to a final concentration of 1µg/mL for 16hr. Cells were then fixed with 4% paraformaldehyde and stained with 4A6 anti-Myc monoclonal antibody and Alexa Fluor 594-conjugated anti-mouse secondary antibody (in **red**). Nuclei were visualised using DAPI (in **blue**). An overlaid image of these two layers is included on the right (**merge**). A brightfield image of each cell line is included on the left. All images at 20x magnification.


empty
Brightfield
SIM1-Myc
DAPI
Merge
WT
Brightfield
SIM1-Myc
DAPI
Merge
K51N
Brightfield
SIM1-Myc
DAPI
Merge
R171H
Brightfield
SIM1-Myc
DAPI
Merge
T292A
Brightfield
SIM1-Myc
DAPI
Merge
H323Y
Brightfield
SIM1-Myc
DAPI
Merge
R550H
Brightfield
SIM1-Myc
DAPI
Merge
T714A
Brightfield
SIM1-Myc
DAPI
Merge

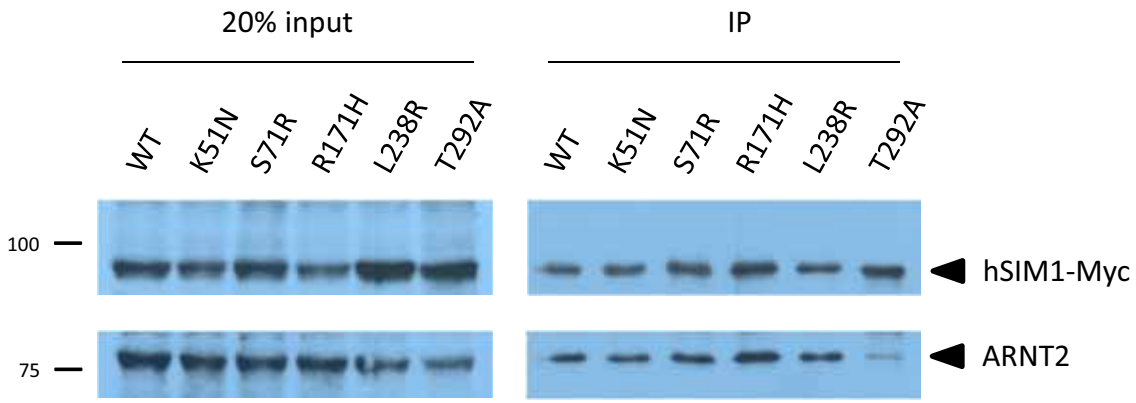


Figure 3.10. SIM1 T292A has reduced affinity for ARNT2. Coimmunoprecipitation analysis of SIM1/ARNT2 dimerisation in whole cell extracts. 293 T-Rex stable cell lines were treated with doxycycline to a final concentration of 1µg/mL for 16hr. 250µg WCE from each cell line was then immunoprecipitated with the ab9106 anti-Myc antibody conjugated to Protein G sepharose beads. The resultant immunoprecipitates (IP) were separated by SDS-PAGE alongside 20% starting material as input, transferred to nitrocellulose, and probed with 4A6 anti-Myc monoclonal antibody and M-165 anti-ARNT2 polyclonal antibody.

detect any reduction in dimerisation for the PAS domain K51N, S71R, R171H or L238R variants, despite their substantially reduced activity in our luciferase reporter assays (**Figures 3.7.** and **3.10.**). The H323Y variant, which is also localised within the SIM1 PAS domain, likewise did not show any reduction in dimerisation ability under these conditions (data not shown). The remaining variants analysed in our reporter assays do not lie within those domains known to confer dimerisation potential on the SIM1 protein and were therefore not assayed using this method (see **Chapter 1**).

Cumulatively, these data indicate that the substantial reduction in luciferase reporter gene activity observed for the SIM1 T292A variant can be attributed to a combination of reduced heterodimerisation with its requisite partner factor and its aberrant localisation both to the cytoplasm and nucleus (**Figures 3.9.** and **3.10.**). There are no data to suggest that aa 292 lies within an NLS sequence, although given that SIM1 is a constitutively nuclear protein it would be reasonably expected to possess such a motif [234]. The fact that the T292A variant is not completely excluded from the nucleus suggests that NLS function is partially disrupted by this amino acid substitution (**Figure 3.9.**). The molecular basis of the loss of function observed for the remaining variants has yet to be established, but does not appear to be altered subcellular localisation or reduced dimerisation potential (**Figures 3.9.** and **3.10.**).

3.3.4. Attempts to develop a SIM1 DNA binding assay

3.3.4.1. Experimental approach

Previously published attempts to characterise SIM1 DNA binding affinity via EMSA are limited, despite a wealth of information regarding similar experiments for other class I bHLH/PAS factors such as the AHR [60]. We therefore believed a more thorough effort to optimise an *in vitro* EMSA protocol for SIM1 could be attempted, in a manner designed to enable analysis of DNA binding potential for any of the nineteen SIM1 variants. Full-length SIM1, ARNT and ARNT2 protein for this series of experiments was obtained either from mammalian cell extracts or *in vitro* transcription/translation reactions (**Figure 3.11.A., B.,** and **D.**). Interestingly, SIM1 can be detected in both cytoplasmic and nuclear protein isolates after subcellular fractionation of cells overexpressing Myc-tagged SIM1, as confirmed by its co-fractionation with the cytoplasmic marker alpha-tubulin and

the nuclear marker U2AF65 (**Figure 3.11.A.**). We consider this to be a consequence of our cytoplasmic and nuclear protein extract purification procedure, since SIM1 is clearly constitutively nuclear in fixed cells ([75]; S. Woods, PhD thesis 2004; and **Figure 3.9.**). A similar phenomenon has also been previously observed for SIM2 and ARNT, which are also constitutively nuclear proteins, during extract preparation ([108]; S. Woods, PhD thesis 2004; A. Farrall, PhD thesis 2009). Extracts from both HEK 293 T-Rex cells and mouse embryonic hypothalamus N39 cells constitutively overexpressing WT SIM1 and/or ARNT2 were used, to account for any cell type-specific cofactors that may be required for optimal DNA binding (for further explanation and characterisation of the N39 cell line, see **Chapter 4**). We also tested several different probe sequences (**Figure 3.11.C.**), and varied our binding and electrophoresis conditions.

3.3.4.2. Attempts to optimise an EMSA protocol using probe sequences derived from the regulatory regions of several SIM2 and HIF-1α direct target genes

Initially, we performed several assays using three different ~40bp FAM-labelled dsDNA probes already available in the laboratory, which contained regulatory sequences (encompassing a single active HRE or CME) from the SIM2 direct target gene *MYOM2* and the two HIF-1α direct target genes *ERYTHROPOIETIN* (*EPO*) and *VASCULAR ENDOTHELIAL GROWTH FACTOR* (*VEGF*) (**Figure 3.11.C.**). The HRE contains the same 5' ACGTG 3' core sequence bound by SIM and SIM2 *in vivo* and may therefore also bind SIM1 (see **Chapter 1**). We observed non-specific shifts for all three of these probes in the presence of cytoplasmic and nuclear extracts from 293 T-Rex mSIM1-Myc/IRES/mARNT2 cells treated with or without doxycycline (**Figure 3.12.A.**) Mutation of the core HRE or CME sequence within the *EPO* and *MYOM2* probes did not abolish binding, nor was it affected by supershift experiments using SIM1- and ARNT2-specific antibodies (**Figure 3.12.B.-C.**) A bacterially expressed and purified N-terminal fragment of the HIF-1α/ARNT heterodimer (kindly provided by A. Chapman-Smith) served as a positive control, and strong binding to each of the three probes was also observed (**Figure 3.12.A.**). Together, these results suggested that the observed bands were most likely attributable to non-specific binding of high affinity DNA binding factors within the cell extracts, and that further optimisation of our experimental approach was required.

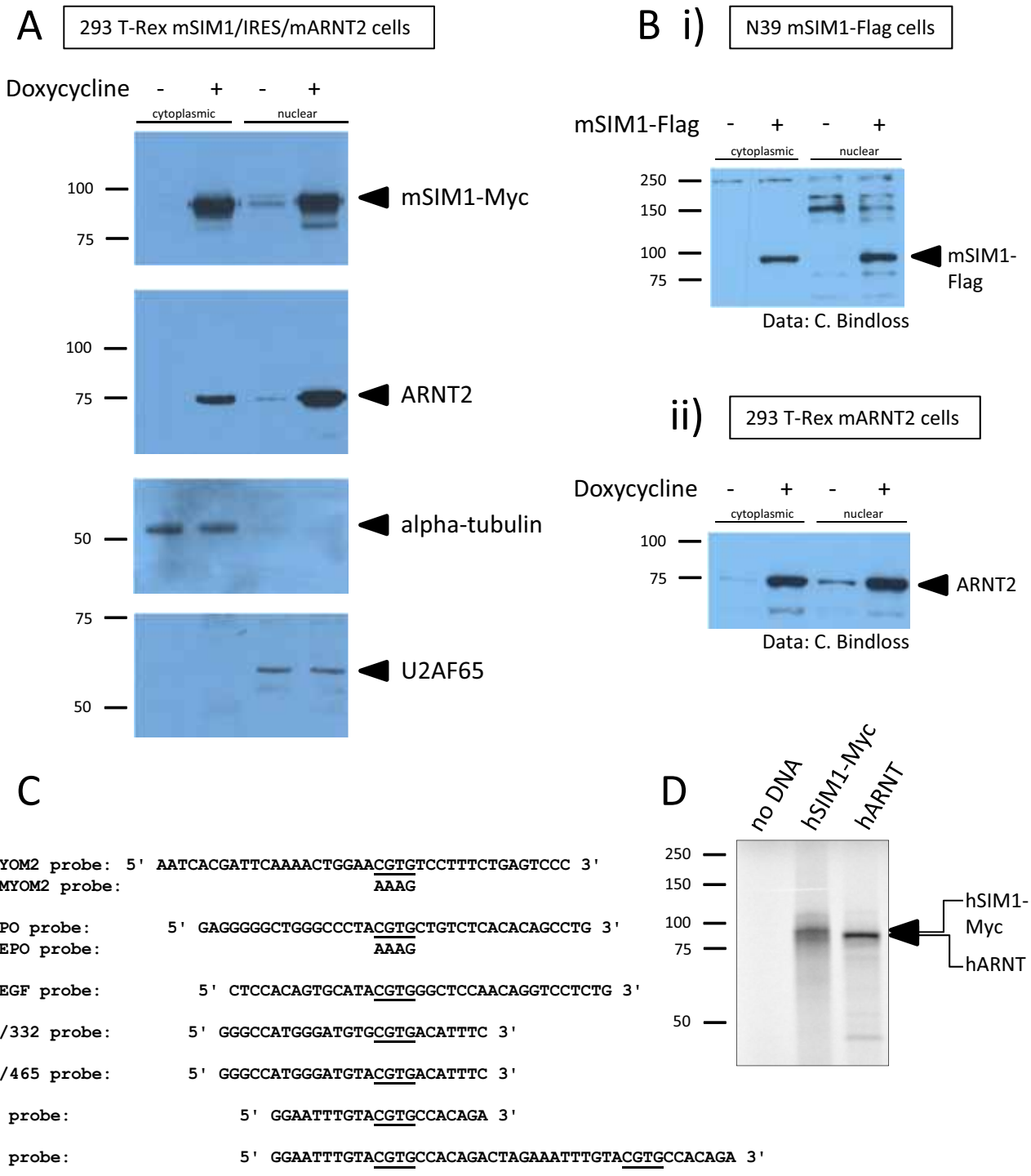


Figure 3.11. Generation of SIM1, ARNT2 and ARNT protein extracts for use in EMSAs, and summary of probe sequences used. A. 293 T-Rex mSIM1-Myc/IRES/mARNT2 cells were treated with (+) or without (-) doxycycline for 24hr and then harvested for cytoplasmic and nuclear extracts. 10µg extract was subsequently separated by SDS-PAGE, transferred to nitrocellulose, and probed with 4A6 anti-Myc monoclonal antibody, M165 anti-ARNT2 polyclonal antibody, 222-6 anti-U2AF65 monoclonal antibody and MCA78G anti-alpha-tubulin monoclonal antibody. **B.** N39 empty vector (-) or mSIM1-Flag (+) overexpressing cells **(i)** or 293 T-Rex mARNT2 cells treated with (+) or without (-) doxycycline for 24hr **(ii)** were harvested for cytoplasmic and nuclear extracts. 20µg extract was subsequently separated by SDS-PAGE, transferred to nitrocellulose, and probed with MW6 anti-SIM1 polyclonal antibody or M165 anti-ARNT2 polyclonal antibody. **C.** The sequence of the upper strand of each dsDNA probe used is indicated, with the core 5' CGTG 3' sequence underlined. The mutated (MUT) MYOM2 and EPO probes contain the same flanking sequences but have the core 5' CGTG 3' sequence mutated to 5' AAAG 3'. The 2xCME probe contains two core 5' CGTG 3' sequences. **D.** hSIM1-Myc and hARNT were transcribed and translated *in vitro* using ^{35}S-Met. 0.08% of the total reaction volume was subsequently separated by SDS-PAGE, dried onto Whatman paper, and exposed to a Phosphorimager.

A

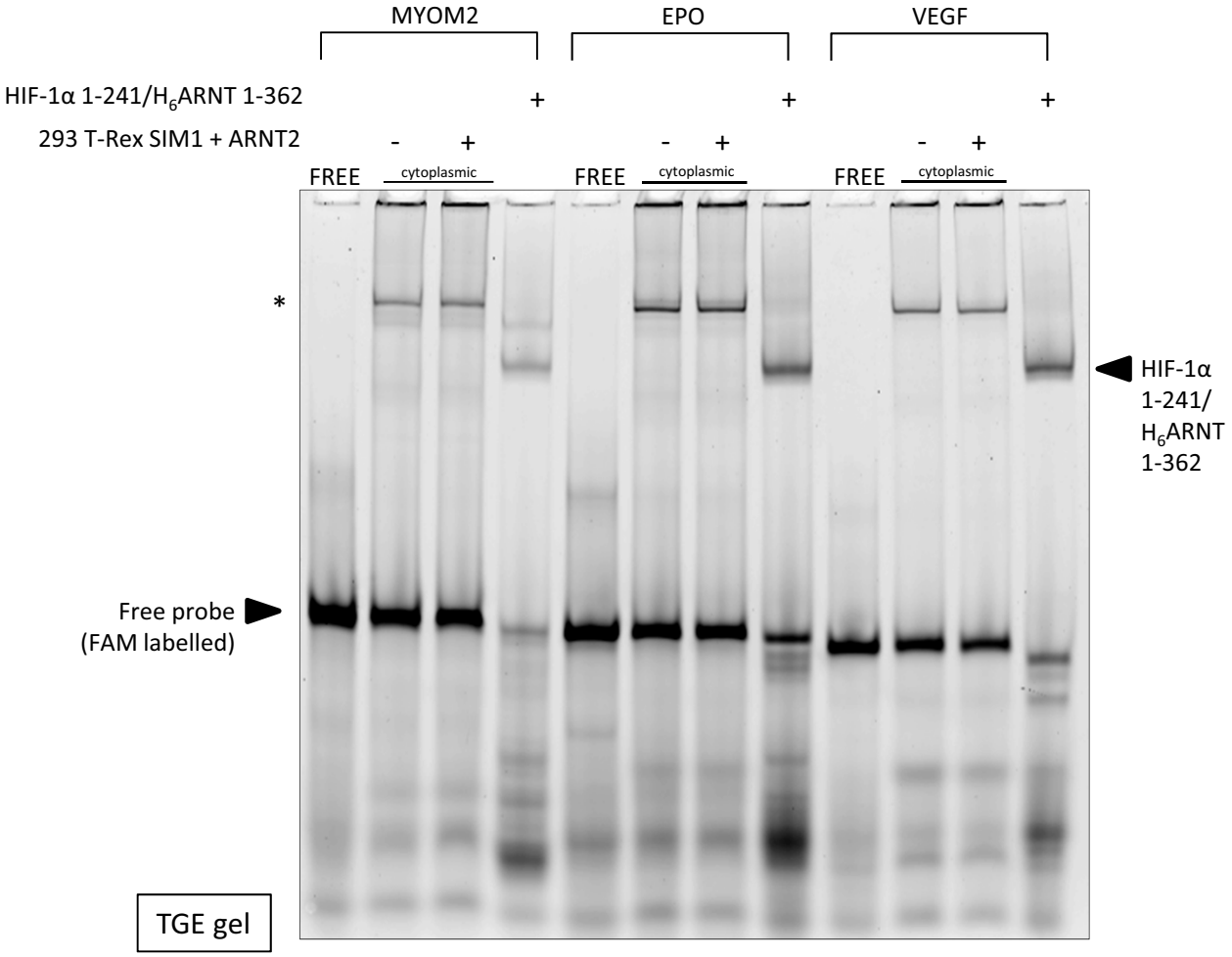
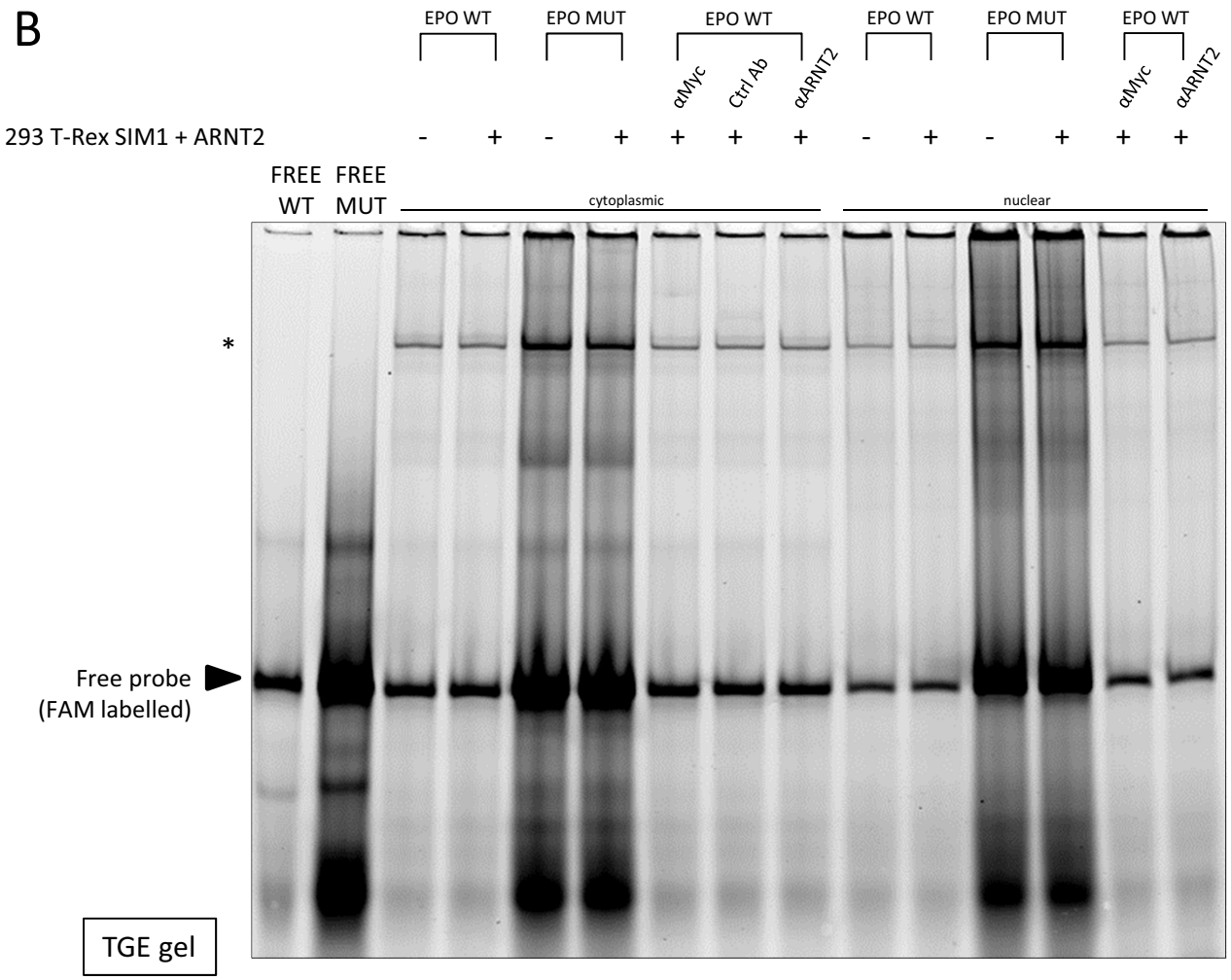
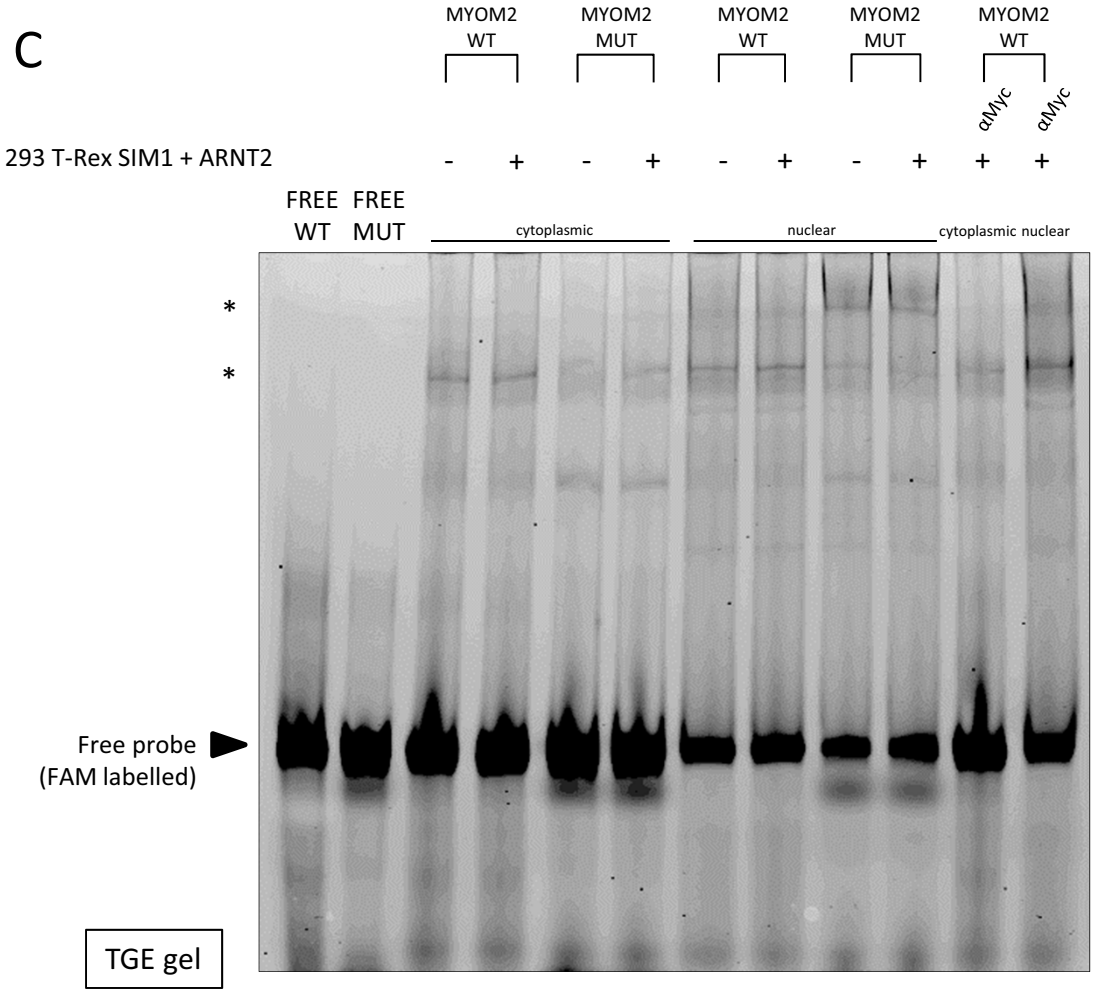


Figure 3.12. Attempted optimisation of SIM1 binding to HRE- and CME-containing probe sequences via EMSA. A. 45µg of cytoplasmic extract from induced (+) or uninduced (-) 293 T-Rex mSIM1-Myc/IRES/mARNT2 cells was bound to FAM-labelled *MYOM2*, *EPO* or *VEGF* probes and subjected to non-denaturing Tris-Glycine PAGE. Bacterially co-expressed and purified HIF-1α 1-241/H_6ARNT 1-362 was used as a positive control. **B.** 20µg of cytoplasmic or nuclear extract from induced (+) or uninduced (-) 293 T-Rex mSIM1-Myc/IRES/mARNT2 cells was bound to FAM-labelled wild type (WT) or mutant (MUT) *EPO* probes and subjected to non-denaturing Tris-Glycine PAGE. The presence of any SIM1/ARNT2 dimers in shifted bands was tested via incubation with αMyc or αARNT2 antibodies relative to an isotype match control antibody. **C.** 20µg of cytoplasmic or nuclear extract from induced (+) or uninduced (-) 293 T-Rex mSIM1-Myc/IRES/mARNT2 cells was bound to FAM-labelled wild type (WT) or mutant (MUT) *MYOM2* probes and subjected to non-denaturing Tris-Glycine PAGE. The presence of any SIM1/ARNT2 dimers in shifted bands was tested via incubation with αMyc antibodies. FREE, free probe. Non-specific shifts are indicated with a *.


B
EPO WT
EPO MUT
EPO WT
EPO WT
EPO MUT
EPO WT
αMyc
Ctrl Ab
αARNT2
αMyc
αARNT2
293 T-Rex SIM1 + ARNT2
- + - + + + + - + - + + +
FREE WT
FREE MUT
cytoplasmic
nuclear
*
Free probe
(FAM labelled)
TGE gel


C
MYOM2 WT
MYOM2 MUT
MYOM2 WT
MYOM2 MUT
MYOM2 WT
αMyc
αMyc
293 T-Rex SIM1 + ARNT2
- + - + - + - + + +
FREE WT
FREE MUT
cytoplasmic
nuclear
cytoplasmic nuclear
*
*
Free probe
(FAM labelled)
TGE gel

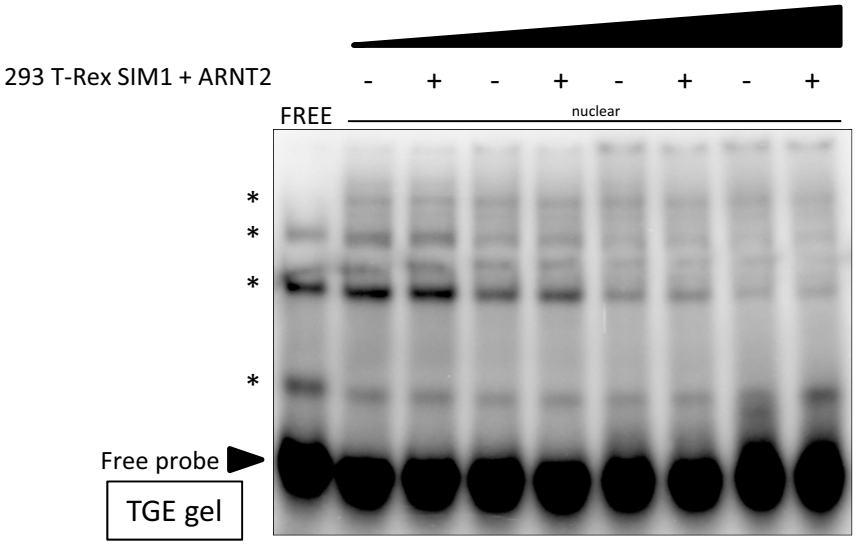
3.3.4.3. Further optimisation of probe sequences and binding conditions

PCR site selection and EMSA experiments by Swanson *et al* represent the only systematic published attempt to identify a preferential *in vitro* CME binding sequence [26]. They describe weak binding of full-length SIM and hARNT to a 5' GNNNNGTGCGTGANNNTCC 3' probe sequence via EMSA, mentioning that they deliberately used oligomers with a fixed 5' GCGTG 3' core to prevent ARNT homodimer formation [26]. Taking into account the 5' (G/A)(T/A)ACGTG 3' SIM binding sequence derived from bioinformatic analysis of several endogenous SIM target genes that had been published just months previously, Swanson *et al* further optimised this probe sequence by varying the nucleotides immediately flanking the core 5' CGTG 3' sequence ([26, 88] and **Chapter 1**). Their two most promising dsDNA probe sequences (designated OL331/332 and OL464/465) were adopted by us for the studies presented in this thesis. We also utilised two more CME-containing dsDNA probes identical in sequence to those bound by SIM1/ARNT and SIM1/ARNT2 in our luciferase reporter gene analyses (**Figures 3.6.-3.8.**), which contained one (1xCME) or two (2xCME) repeats of a 25bp sequence derived from the proximal promoter of the *toll* gene, a direct SIM target [13].

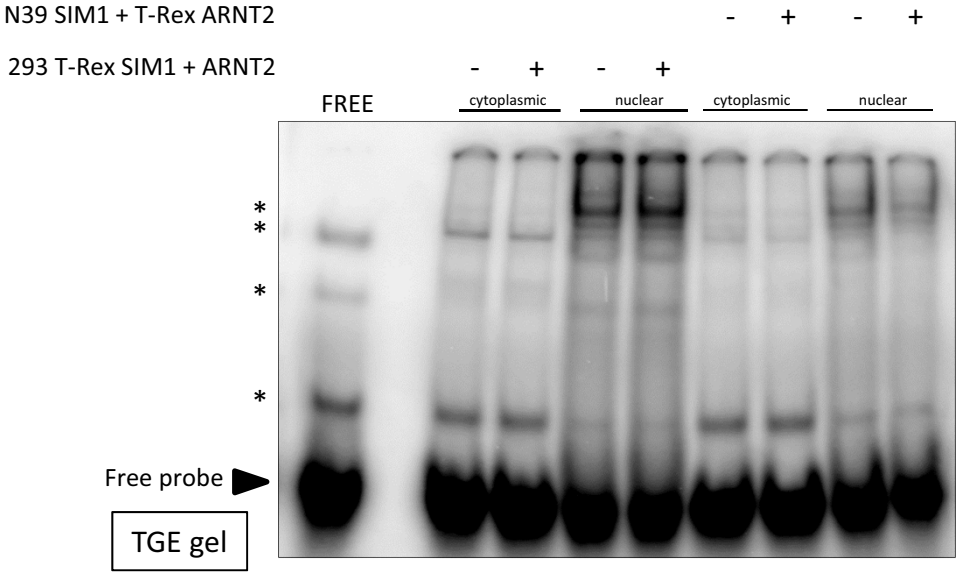
Figures 3.13.-3.16. illustrate our systematic attempts to achieve an observable SIM1-specific EMSA shift with each of the four probe sequences described above, using overexpressed SIM1, ARNT and/or ARNT2 proteins from various sources, different binding conditions, and two different gel chemistries to increase our chances of success. Of primary interest was to vary the conditions of each sequential assay such that non-specific binding might be discouraged without compromising specific, SIM1-mediated DNA binding. We also used ^{32}P-labelled probes to increase the sensitivity of our assays. We initially attempted a dose response curve using increasing amounts nuclear extract from 293 T-Rex mSIM1-Myc/IRES/mARNT2 cells treated with or without doxycycline, and varied the amount of non-specific competitors (BSA and polydI-dC.polydI-dC) proportionately. No specific shift was seen for any of the four probes in treated samples relative to untreated (**Figures 3.13.A., 3.14.A., 3.15.A.** and **3.16.A.**). We then attempted using the maximum amount of cytoplasmic or nuclear extract from doxycycline treated or untreated 293 T-Rex mSIM1-Myc/IRES/mARNT2 cells, as well as N39 mSIM1-Flag or empty vector control cells and doxycycline treated or untreated 293 T-Rex mARNT2 cells, under similar binding conditions. Once again, no specific shift was seen for the

Figure 3.13. Attempted optimisation of SIM1 binding to the CME-containing OL331/332 probe sequence via EMSA. A. Increasing amounts (5, 10, 15 and 20µg) of nuclear extract from induced (+) and uninduced (-) 293 T-Rex mSIM1-Myc/IRES/mARNT2 cells were bound to ^{32}P-labelled OL331/332 and subjected to non-denaturing Tris-Glycine PAGE. **B.** Equal amounts of cytoplasmic or nuclear extract from N39 empty vector and 293 T-Rex mARNT2 uninduced cells (-) or N39 mSIM1-Flag and 293 T-Rex mARNT2 induced cells (+), or 30ug cytoplasmic or nuclear extract from induced (+) and uninduced (-) 293 T-Rex mSIM1-Myc/IRES/mARNT2 cells were bound to ^{32}P-labelled OL331/332 and subjected to Tris-Glycine non-denaturing PAGE. **C.** Equal amounts of *in vitro* transcribed and translated hSIM1-Myc and hARNT (+) or blank lysate (-) were bound to ^{32}P-labelled OL331/332 and subjected to Tris-Glycine non-denaturing PAGE. **D.** Equal amounts of cytoplasmic or nuclear extract from N39 empty vector and 293 T-Rex mARNT2 uninduced cells (-) or N39 mSIM1-Flag and 293 T-Rex mARNT2 induced cells (+), or 30ug cytoplasmic or nuclear extract from induced (+) and uninduced (-) 293 T-Rex mSIM1-Myc/IRES/mARNT2 cells were bound to ^{32}P-labelled OL331/332 and subjected to Tris-Borate-EDTA non-denaturing PAGE. **E.** Equal amounts of *in vitro* transcribed and translated hSIM1-Myc and hARNT (+) or blank lysate (-) were bound to ^{32}P-labelled OL331/332 and subjected to Tris-Borate-EDTA non-denaturing PAGE. FREE, free probe. Non-specific shifts are indicated with a *.

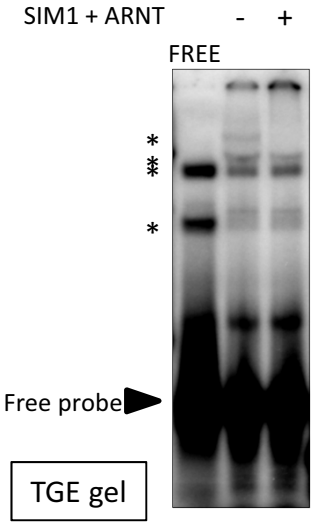
A


293 T-Rex SIM1 + ARNT2
- + - + - + - +
FREE
nuclear
*
*
*
*
Free probe
TGE gel

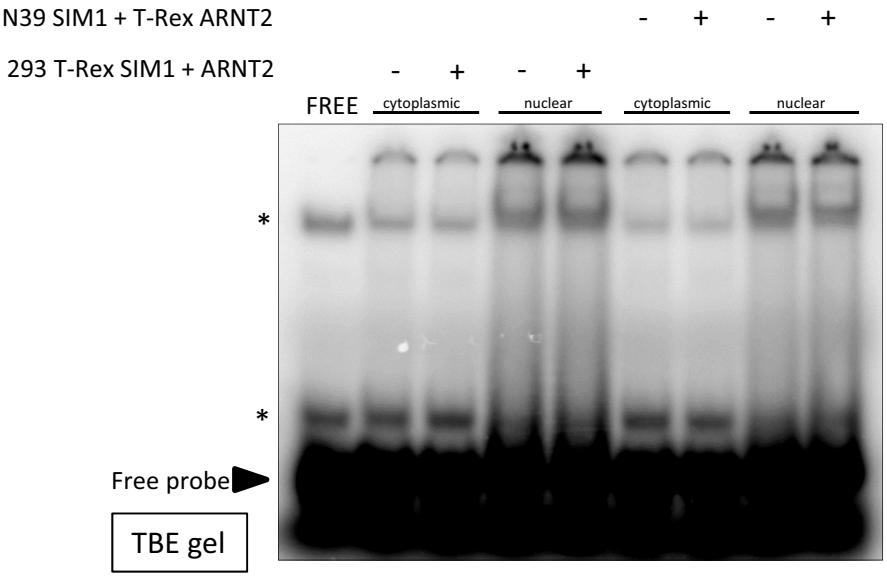
B


N39 SIM1 + T-Rex ARNT2
- + - +
293 T-Rex SIM1 + ARNT2
- + - +
FREE
cytoplasmic
nuclear
cytoplasmic
nuclear
*
*
*
*
Free probe
TGE gel

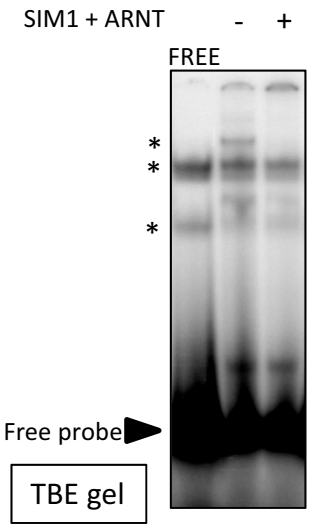
C


SIM1 + ARNT
- +
FREE
*
*
*
Free probe
TGE gel

D


N39 SIM1 + T-Rex ARNT2
- + - +
293 T-Rex SIM1 + ARNT2
- + - +
FREE
cytoplasmic
nuclear
cytoplasmic
nuclear
*
*
Free probe
TBE gel

E


SIM1 + ARNT
- +
FREE
*
*
*
Free probe
TBE gel

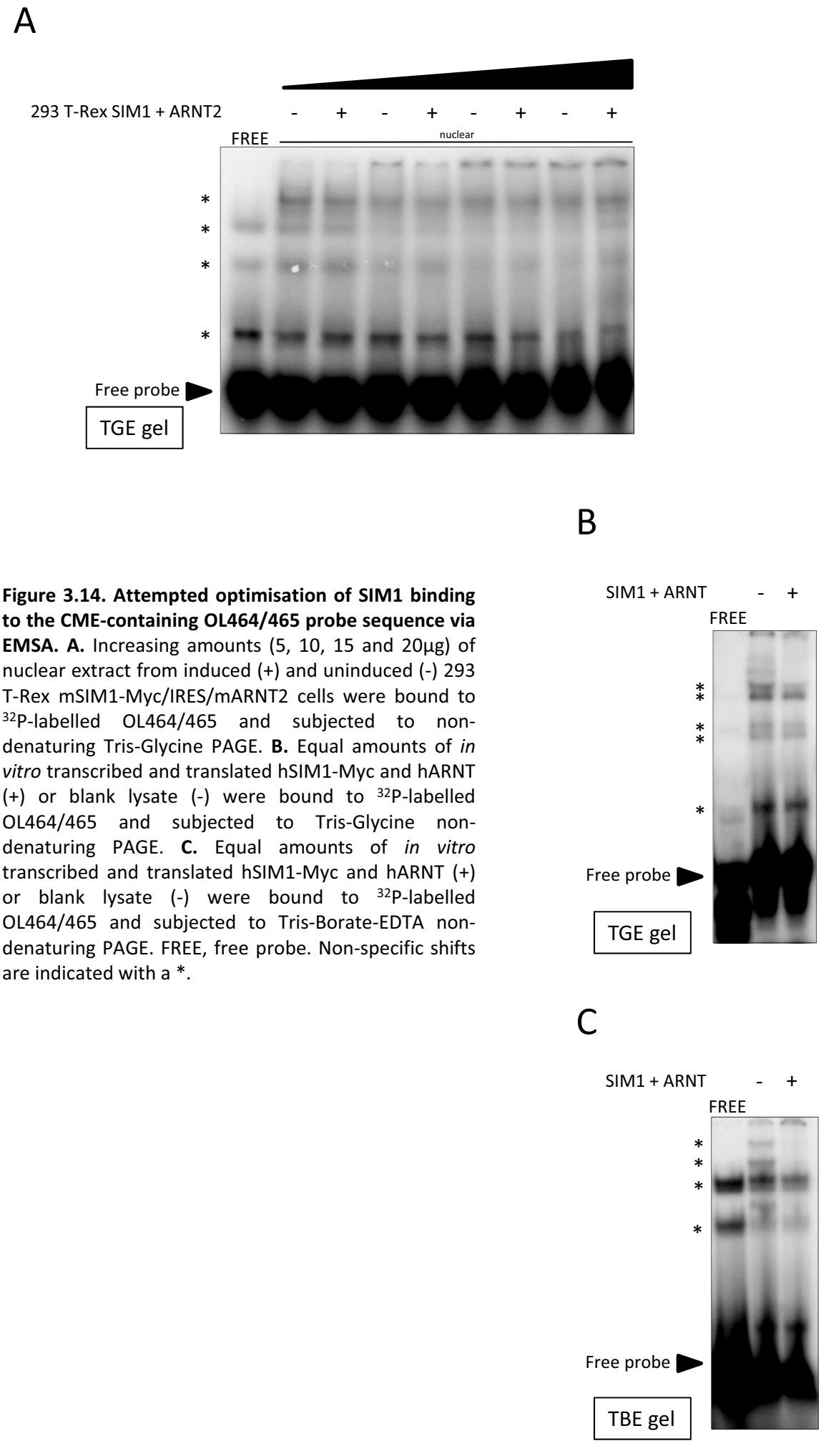


Figure 3.14. Attempted optimisation of SIM1 binding to the CME-containing OL464/465 probe sequence via EMSA. A. Increasing amounts (5, 10, 15 and 20µg) of nuclear extract from induced (+) and uninduced (-) 293 T-Rex mSIM1-Myc/IRES/mARNT2 cells were bound to ^{32}P-labelled OL464/465 and subjected to non-denaturing Tris-Glycine PAGE. **B.** Equal amounts of *in vitro* transcribed and translated hSIM1-Myc and hARNT (+) or blank lysate (-) were bound to ^{32}P-labelled OL464/465 and subjected to Tris-Glycine non-denaturing PAGE. **C.** Equal amounts of *in vitro* transcribed and translated hSIM1-Myc and hARNT (+) or blank lysate (-) were bound to ^{32}P-labelled OL464/465 and subjected to Tris-Borate-EDTA non-denaturing PAGE. FREE, free probe. Non-specific shifts are indicated with a *.

Figure 3.15. Attempted optimisation of SIM1 binding to the CME-containing 1xCME probe sequence via EMSA. A. Increasing amounts (5, 10, 15 and 20µg) of nuclear extract from induced (+) and uninduced (-) 293 T-Rex mSIM1-Myc/IRES/mARNT2 cells were bound to ^{32}P-labelled 1xCME and subjected to non-denaturing Tris-Glycine PAGE. **B.** Equal amounts of cytoplasmic or nuclear extract from N39 empty vector and 293 T-Rex mARNT2 uninduced cells (-) or N39 mSIM1-Flag and 293 T-Rex mARNT2 induced cells (+), or 30ug cytoplasmic or nuclear extract from induced (+) and uninduced (-) 293 T-Rex mSIM1-Myc/IRES/mARNT2 cells were bound to ^{32}P-labelled 1xCME and subjected to Tris-Glycine non-denaturing PAGE. **C.** Equal amounts of *in vitro* transcribed and translated hSIM1-Myc and hARNT (+) or blank lysate (-) were bound to ^{32}P-labelled 1xCME and subjected to Tris-Glycine non-denaturing PAGE. **D.** Equal amounts of cytoplasmic or nuclear extract from N39 empty vector and 293 T-Rex mARNT2 uninduced cells (-) or N39 mSIM1-Flag and 293 T-Rex mARNT2 induced cells (+), or 30ug cytoplasmic or nuclear extract from induced (+) and uninduced (-) 293 T-Rex mSIM1-Myc/IRES/mARNT2 cells were bound to ^{32}P-labelled 1xCME and subjected to Tris-Borate-EDTA non-denaturing PAGE. **E.** Equal amounts of *in vitro* transcribed and translated hSIM1-Myc and hARNT (+) or blank lysate (-) were bound to ^{32}P-labelled 1xCME and subjected to Tris-Borate-EDTA non-denaturing PAGE. FREE, free probe. Non-specific shifts are indicated with a *.

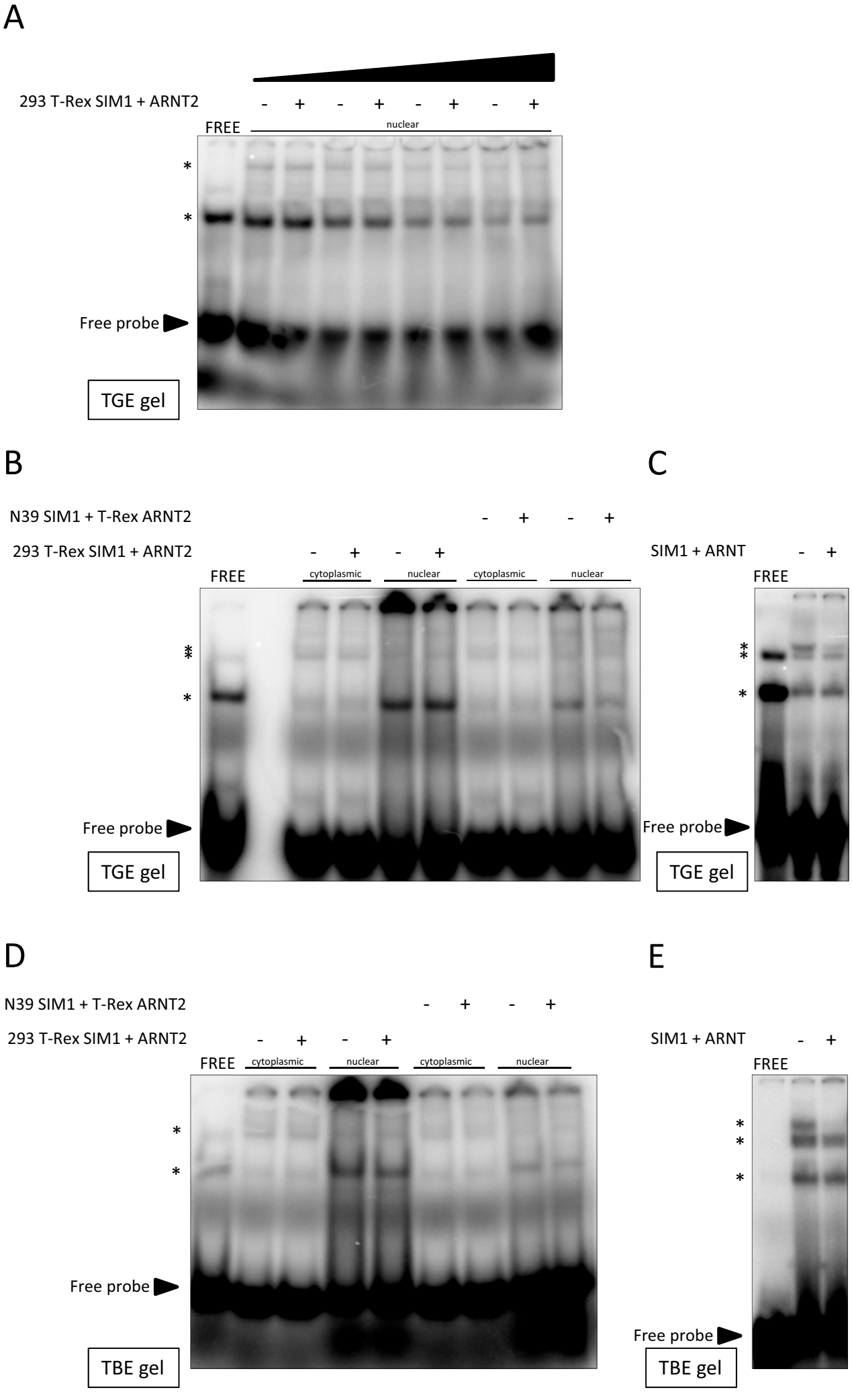
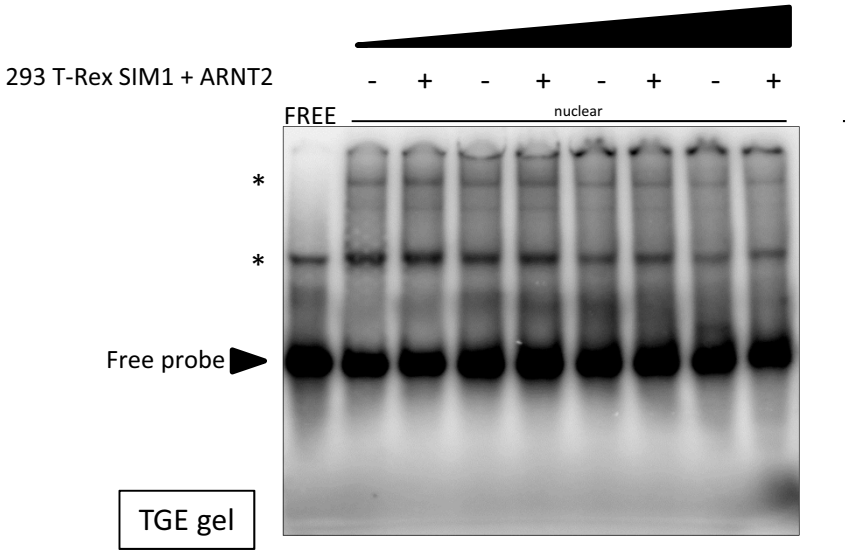
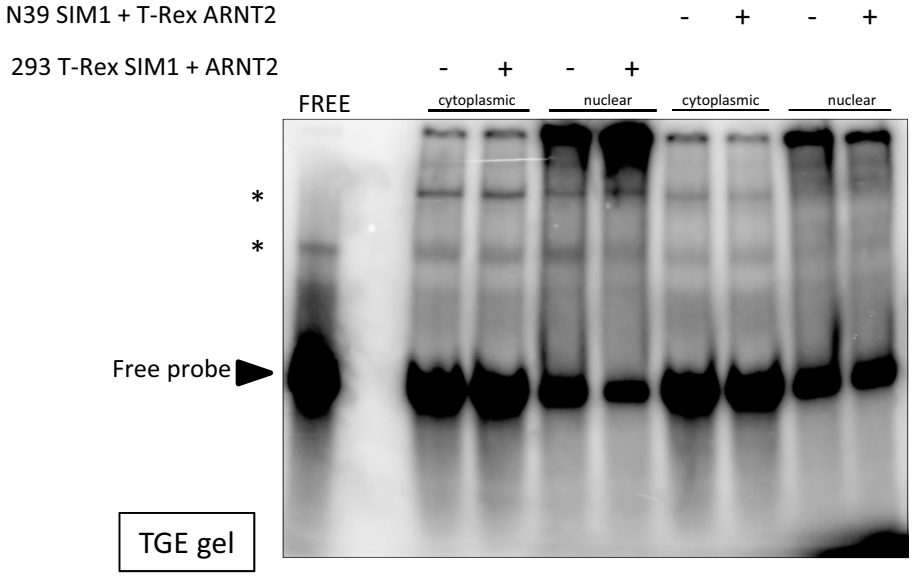

A
293 T-Rex SIM1 + ARNT2
- + - + - + - +
FREE
nuclear
*
*
Free probe
TGE gel
B
N39 SIM1 + T-Rex ARNT2
- + - +
293 T-Rex SIM1 + ARNT2
- + - +
FREE
cytoplasmic
nuclear
cytoplasmic
nuclear
**
*
Free probe
TGE gel
C
SIM1 + ARNT
- +
FREE
**
*
Free probe
TGE gel
D
N39 SIM1 + T-Rex ARNT2
- + - +
293 T-Rex SIM1 + ARNT2
- + - +
FREE
cytoplasmic
nuclear
cytoplasmic
nuclear
*
*
Free probe
TBE gel
E
SIM1 + ARNT
- +
FREE
*
*
*
Free probe
TBE gel

Figure 3.16. Attempted optimisation of SIM1 binding to the CME-containing 2xCME probe sequence via EMSA. A. Increasing amounts (5, 10, 15 and 20μg) of nuclear extract from induced (+) and uninduced (-) 293 T-Rex mSIM1-Myc/IRES/mARNT2 cells were bound to ^{32}P-labelled 2xCME and subjected to non-denaturing Tris-Glycine PAGE. **B.** Equal amounts of cytoplasmic or nuclear extract from N39 empty vector and 293 T-Rex mARNT2 uninduced cells (-) or N39 mSIM1-Flag and 293 T-Rex mARNT2 induced cells (+), or 30ug cytoplasmic or nuclear extract from induced (+) and uninduced (-) 293 T-Rex mSIM1-Myc/IRES/mARNT2 cells were bound to ^{32}P-labelled 2xCME and subjected to Tris-Glycine non-denaturing PAGE. **C.** Equal amounts of *in vitro* transcribed and translated hSIM1-Myc and hARNT (+) or blank lysate (-) were bound to ^{32}P-labelled 2xCME and subjected to Tris-Glycine non-denaturing PAGE. **D.** Equal amounts of cytoplasmic or nuclear extract from N39 empty vector and 293 T-Rex mARNT2 uninduced cells (-) or N39 mSIM1-Flag and 293 T-Rex mARNT2 induced cells (+), or 30ug cytoplasmic or nuclear extract from induced (+) and uninduced (-) 293 T-Rex mSIM1-Myc/IRES/mARNT2 cells were bound to ^{32}P-labelled 2xCME and subjected to Tris-Borate-EDTA non-denaturing PAGE. **E.** Equal amounts of *in vitro* transcribed and translated hSIM1-Myc and hARNT (+) or blank lysate (-) were bound to ^{32}P-labelled 2xCME and subjected to Tris-Borate-EDTA non-denaturing PAGE. FREE, free probe. Non-specific shifts are indicated with a *.

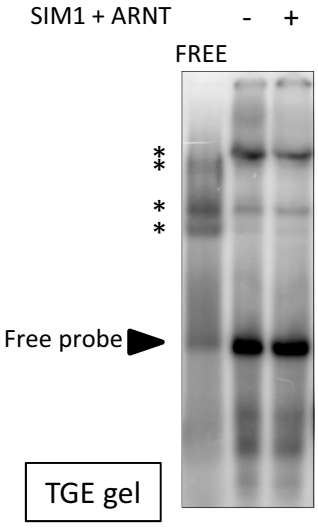
A


293 T-Rex SIM1 + ARNT2
- + - + - + - +
FREE
nuclear
*
*
Free probe
TGE gel

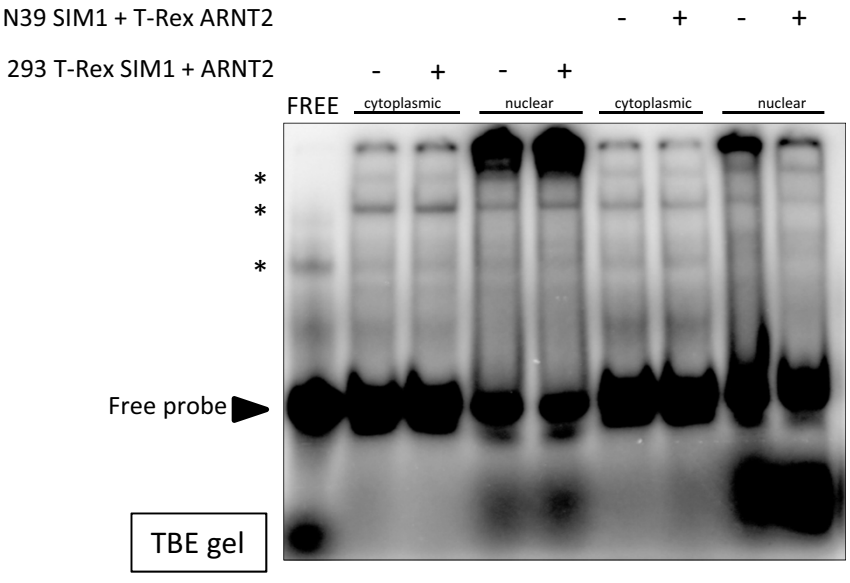
B


N39 SIM1 + T-Rex ARNT2
- + - +
293 T-Rex SIM1 + ARNT2
- + - +
FREE
cytoplasmic
nuclear
cytoplasmic
nuclear
*
*
Free probe
TGE gel

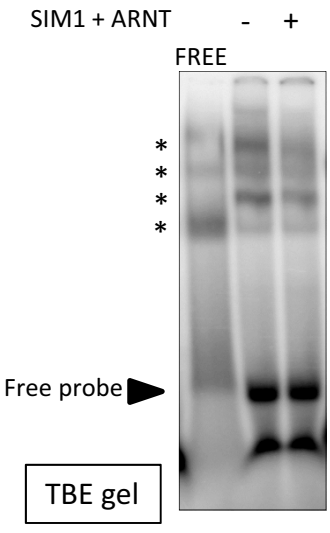
C


SIM1 + ARNT
- +
FREE
**
**
Free probe
TGE gel

D


N39 SIM1 + T-Rex ARNT2
- + - +
293 T-Rex SIM1 + ARNT2
- + - +
FREE
cytoplasmic
nuclear
cytoplasmic
nuclear
*
*
*
Free probe
TBE gel

E


SIM1 + ARNT
- +
FREE
*
*
*
*
Free probe
TBE gel

three probes we tested, despite the additional use of the less stringent Tris-Borate-EDTA (TBE) gel electrophoresis system for sample separation, in addition to the previously used Tris-Glycine-EDTA (TGE) gel electrophoresis system (**Figures 3.13.B.** and **D., 3.15.B.** and **D.** and **3.16.B.** and **D.**). Finally, we were also unable to detect a specific shift for any of the four probes using *in vitro* transcribed and translated hSIM1-Myc and hARNT using either TBE or TGE gel electrophoresis (**Figures 3.13.C.** and **E., 3.14.B.** and **C., 3.15.C.** and **E.** and **3.16.C.** and **E.**). At the time of writing we have abandoned this experimental approach, and are in the process of optimising a chromatin immunoprecipitation (ChIP) protocol for overexpressed hSIM1-Myc and the 6xCME reporter construct in 293 T-Rex hSIM1-Myc cells, in the hope of isolating DNA-bound SIM1 complexes under the same conditions used for our luciferase reporter assays (**Figures 3.7.** and **3.8.**).

In summary, we have successfully developed and optimised a dual luciferase reporter gene assay protocol capable of assessing the consequence of nineteen different point mutations within the SIM1 sequence on its transcriptional activity in HEK derived cells, and confirmed that the majority of them display significantly reduced transcriptional activity relative to WT SIM1 in the presence of both ARNT and ARNT2. These nineteen SIM1 variants were identified by two independent research groups via human genetic association studies in severely obese patients, and suggest a link between reduced SIM1 transcriptional activity and weight gain in humans. Thus far, the majority of variants that display altered transcriptional activity do so via an unknown mechanism. It cannot be ascribed to any substantial alteration in protein stability, since all nineteen variants express at a level comparable to WT SIM1 by Western analysis. However, the severe reduction in luciferase reporter gene activity observed for SIM1 T292A can be attributed to a combination of reduced dimerisation ability and a failure to localise exclusively to the nucleus like WT SIM1. This may indicate the importance of aa 292 for nuclear accumulation of the SIM1 protein. Finally, we also attempted to develop an *in vitro* DNA binding assay for full-length WT SIM1, using both cytoplasmic and nuclear extracts from SIM1 and/or ARNT2 overexpressing HEK- and embryonic hypothalamus-derived cells and *in vitro* transcribed and translated SIM1 and ARNT proteins, and seven different probe sequences derived from previously characterised SIM2, HIF-1α or SIM target genes. Despite these comprehensive analyses, we were unable to detect a SIM1-specific shift under any of the conditions we tested. The question of whether any of the identified SIM1 point mutations contribute to reduced DNA binding therefore remains unaddressed.

CHAPTER 4: RESULTS

Identification of novel downstream SIM1 target genes

4.1. Choice of experimental strategy

In light of the increasingly substantial link between altered *SIM1* gene dosage and severe obesity in both humans and mice, it is becoming ever more important to understand the molecular processes by which SIM1 influences weight regulation in mammals [59, 96, 98, 102, 177, 181, 183]. Furthermore, its obvious necessity for survival in mice suggests an additional critical role for SIM1 in embryonic development [58]. There is a dearth of molecular detail concerning the function of this essential transcription factor in both of these processes, the elucidation of which would aid our understanding of its critical roles in both survival and homeostasis [57, 137]. We therefore decided to adopt an unbiased microarray approach towards identifying SIM1 downstream transcriptional targets.

Ideally, we wished to adopt an experimental strategy that would favour the identification of genes whose products are important for SIM1-mediated signalling in the hypothalamus, since the severe hyperphagic obesity and PVN/SON malformations described for *Sim1* deficient mice were the only data indicative of a specific *in vivo* role for *Sim1* at the commencement of this study [58, 59, 96]. However, we were unable to obtain *Sim1*$^{+/-}$ mice for our studies, and an appropriate hypothalamus-derived cell line within which to manipulate SIM1 expression levels had yet to be described. We therefore decided to search for downstream SIM1 target genes in an alternative cellular context.

Extensive PCR analyses in our laboratory have indicated that *SIM1* is expressed at relatively high levels in cultured kidney cell lines such as HEK 293T (A. Raimondo, Honours thesis 2005). Given the reproducibility of the 293 T-Rex system for the studies presented in **Chapter 3** and the consistency in protein overexpression we observed between multiple independently generated stable cell lines, we reasoned that this system offered the means to analyse SIM1-mediated changes in downstream target gene expression in a highly imitable context. The relatively high levels of *SIM1* expression in 293T cells also raised the possibility of additional verification experiments in this cell line

via *SIM1* knockdown. Given the absence of a single reported *bona fide* direct SIM1 target gene in any cell or tissue type, this strategy offered a tractable and highly relevant experimental approach with which to uncover a wealth of information regarding SIM1-mediated changes in gene expression. We anticipated that it had the potential to uncover not only kidney-specific SIM1 target genes, but also more "generic" target genes that may prove equally relevant to SIM1 function in the other tissues in which it is expressed, including the hypothalamus. The identification of any such genes, and subsequent analysis of their regulatory regions, could in turn inform the identification of novel SIM1 target genes using bioinformatic techniques.

Ultimately, we chose to perform microarray analysis on 293 T-Rex cells inducibly expressing both mSIM1-Myc and its obligate partner factor mARNT2. We expressed the SIM1/ARNT2 heterodimer (as opposed to the SIM1/ARNT heterodimer) since gene knockout studies indicate that the dimerisation partner for SIM1 *in vivo* is likely to be ARNT2 [81]. The human and mouse orthologues of these two proteins are almost identical at the amino acid level (96% between hSIM1 and mSIM1 and 97% between hARNT2 and mARNT2, see also **Figure 1.2.**). We therefore anticipated that the mouse SIM1/ARNT2 heterodimer would be sufficiently capable of regulating downstream target gene expression in a human cell line. More importantly, their relative divergence at the nucleotide level allowed us to differentiate between endogenous *SIM1* and *ARNT2* and ectopic *Sim1* and *Arnt2* transcripts in subsequent QPCR analyses. We also performed QPCR analysis on a select number of candidate target genes, initially identified by microarray, upon *SIM1* knockdown in 293T cells. During the course of these studies, Belsham *et al* described the generation of a host of immortalised neuronal cell lines derived from the hypothalamus of E15.0-18.0 mice [224]. This offered a unique opportunity to test for conservation of target gene regulation by SIM1 between kidney- and hypothalamus-derived cells. We therefore purchased three of these cell lines and performed RT-PCR and immunohistochemical analyses to determine whether they might be derived from the PVN/SON/aPV region. We subsequently generated stable cell lines and tested a number of candidate target genes for their ability to respond to SIM1 overexpression and knockdown in these cells. Finally, we performed bioinformatic analysis on the upstream regulatory regions of two of our most promising candidates to identify CME sequences that may be bound by SIM1 *in vivo*.

4.2. 293 T-Rex cells can be successfully induced to express SIM1 and ARNT2

First, we performed Western blot and CoIP analyses on two independently derived stable cell lines inducibly overexpressing Myc-tagged mSIM1 and mARNT2 (**Figure 4.1.A.i** and **B.**). Western analysis of cells treated with doxycycline for time periods ranging from 6-24hr indicated that mSIM1-Myc and mARNT2 expression was strongly induced in both cell lines within 6hr (**Figure 4.1.A.i**). Overexpressed mSIM1-Myc and mARNT2 also formed a stable complex in CoIP experiments (**Figure 4.1.B.**). The 293 T-Rex system did not display complete repression of mSIM1-Myc and mARNT2 expression in the uninduced state, since RT-PCR analysis indicated considerable levels of basal *Sim1* and *Arnt2* transcript expression in the absence of doxycycline after only 25 cycles of amplification relative to the untransformed (parent) 293 T-Rex cell line, using primers capable of amplifying both human and mouse *Sim1* and *Arnt2* (**Figure 4.1.A.ii**). A similar phenomenon could also be observed by Western analysis after extended exposure (data not shown). Dual luciferase reporter assay experiments presented in **Chapter 3** confirmed that independently derived SIM1-expressing 293 T-Rex cell lines display a highly comparable degree of artificial reporter gene induction, suggesting that SIM1 is comparably active in these cell lines. Together, these data indicate that the 293 T-Rex cell line is a highly reproducible system within which to study SIM1-mediated changes in gene expression. We therefore decided to adopt this system for subsequent microarray experiments.

4.3. Microarray experiments

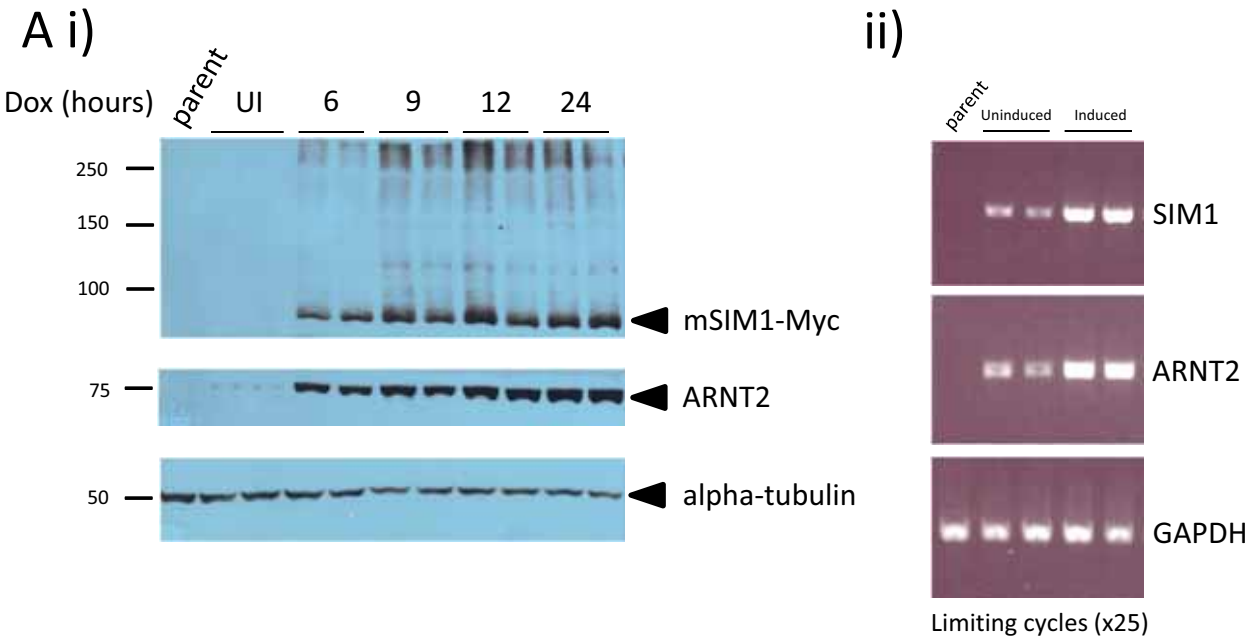
4.3.1. Experimental design

Initially, we attempted comparative microarray analysis of total RNA samples prepared from 293 T-Rex mSIM1-Myc/IRES/mARNT2 cells treated with or without doxycycline for 12hr. However, statistical analysis indicated a lack of significant differences in transcript levels between untreated and treated samples using this method (data not shown). We anticipated that this may be due to "leakiness" of the 293 T-Rex system, producing basal levels of mSIM1-Myc and mARNT2 expression in untreated cells, which may be maximally altering the expression of any downstream target genes in the absence of doxycycline (**Figure 4.1.A.ii**). We therefore decided to repeat our analyses using

triplicate total RNA samples prepared from parent 293 T-Rex cells, and 293 T-Rex mSIM1-Myc/IRES/mARNT2 cells treated with doxycycline for 12hr (**Figure 4.2.A.**). In this way, we hoped to reduce "background" levels of downstream target gene induction or repression and therefore favour the identification of any genes responding to mSIM1-Myc and mARNT2 induction in doxycycline treated 293 T-Rex mSIM1-Myc/IRES/mARNT2 cells. The absence of any statistically significant changes in transcript levels between 293 T-Rex mSIM1-Myc/IRES/mARNT2 cells treated with or without doxycycline in the first instance suggested that doxycycline itself was incapable of substantially affecting transcript levels in this system. We therefore did not perform "mock" induction of the parent 293 T-Rex cell line. RT-PCR analysis of cDNA synthesised from the array samples confirmed a high degree of *Sim1* and *Arnt2* induction in treated cells relative to parent cells after only 25 cycles of amplification (**Figure 4.2.B.**). A schematic representation of our final microarray experiment design is depicted in **Figure 4.2.C.**

4.3.2. Data processing, analysis and initial verification

All microarray analyses, data processing and statistical analyses were performed by the Adelaide Microarray Facility in collaboration with S. Pederson at the Women's and Children's Health Research Institute. 237 differentially expressed genes were identified using a false discovery rate-adjusted p value < 0.05 using the Limma data analysis package, all of which were upregulated in doxycycline treated 293 T-Rex mSIM1-Myc/IRES/mARNT2 cells relative to parent 293 T-Rex cells (**Figure 4.3.**). *Sim1* and *Arnt2* transcripts were only 2.12 and 2.45 fold upregulated respectively in doxycycline-treated 293 T-Rex mSIM1-Myc/IRES/mARNT2 cells relative to parent 293 T-Rex cells, despite evidence of strong induction of both genes in these samples by RT-PCR (**Figure 4.2.B.**). This is most likely due to the overexpressed *mSim1* and *mArnt2* transcripts having only partial homology to the *hSIM1* and *hARNT2* probes present on the array slides (data not shown). The remainder of the top twenty-five clones (ranked according to fold change) also did not display changes much greater than 2-3 fold (**Figure 4.3.**). As an initial step, therefore, we attempted to verify eleven of the top twenty-five differentially expressed clones by QPCR in the same RNA samples used for our microarray experiments.



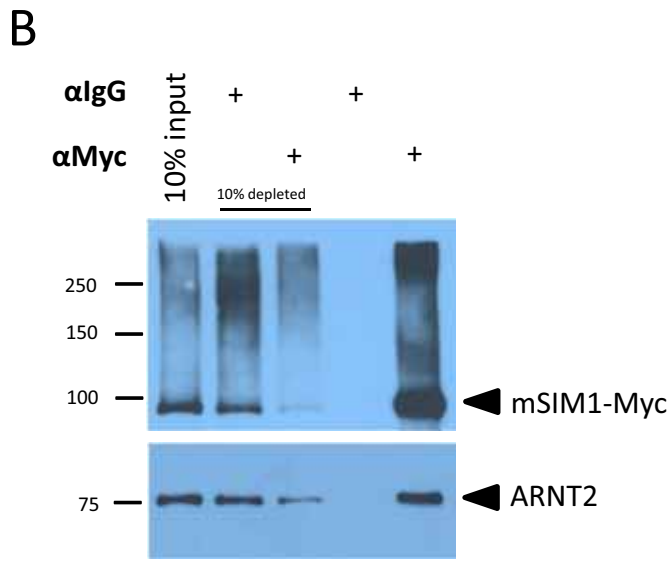


Figure 4.1. Characterisation of SIM1 and ARNT2 expression and activity in HEK 293 T-Rex parent and stable cell lines. A. Western blot **(i)** and RT-PCR **(ii)** analysis of Myc-tagged SIM1 and ARNT2 overexpression in 293 T-Rex stable cell lines. **i)** WCEs were prepared from parent 293 T-Rex cells and two independently derived 293 T-Rex mSIM1-Myc/IRES/mARNT2 stable cell lines treated with doxycycline to a final concentration of 1µg/mL for the time periods indicated. 30µg WCE was subsequently separated by SDS-PAGE, transferred to nitrocellulose, and probed with 4A6 anti-Myc monoclonal antibody, M-165 anti-ARNT2 polyclonal antibody and MCA78G anti-alpha-tubulin monoclonal antibody. UI, uninduced (untreated) cell extract. **ii)** cDNA was prepared from parent 293 T-Rex cells, and two independently derived 293 T-Rex mSIM1-Myc/IRES/mARNT2 stable cell lines treated with (induced) or without (uninduced) doxycycline to a final concentration of 1µg/mL for 24hr, and assayed for the presence of overexpressed *Sim1* and *Arnt2* by RT-PCR with reference to the housekeeping gene *GAPDH*. **B.** Co-immunoprecipitation analysis of mSIM1-Myc/mARNT2 dimerisation in WCEs. 293 T-Rex mSIM1-Myc/IRES/mARNT2 cells were treated with doxycycline to a final concentration of 1µg/mL for 24hr and WCEs prepared. 100µg extract was then immunoprecipitated with the 4A6 anti-Myc antibody or an isotype match (anti-IgG) antibody conjugated to Protein G sepharose beads. The resultant immunoprecipitates were separated by SDS-PAGE alongside 10% of the starting material (input) and 10% of the unbound (depleted) supernatant, transferred to nitrocellulose, and probed with ab9106 anti-Myc monoclonal antibody and M-165 anti-ARNT2 polyclonal antibody.

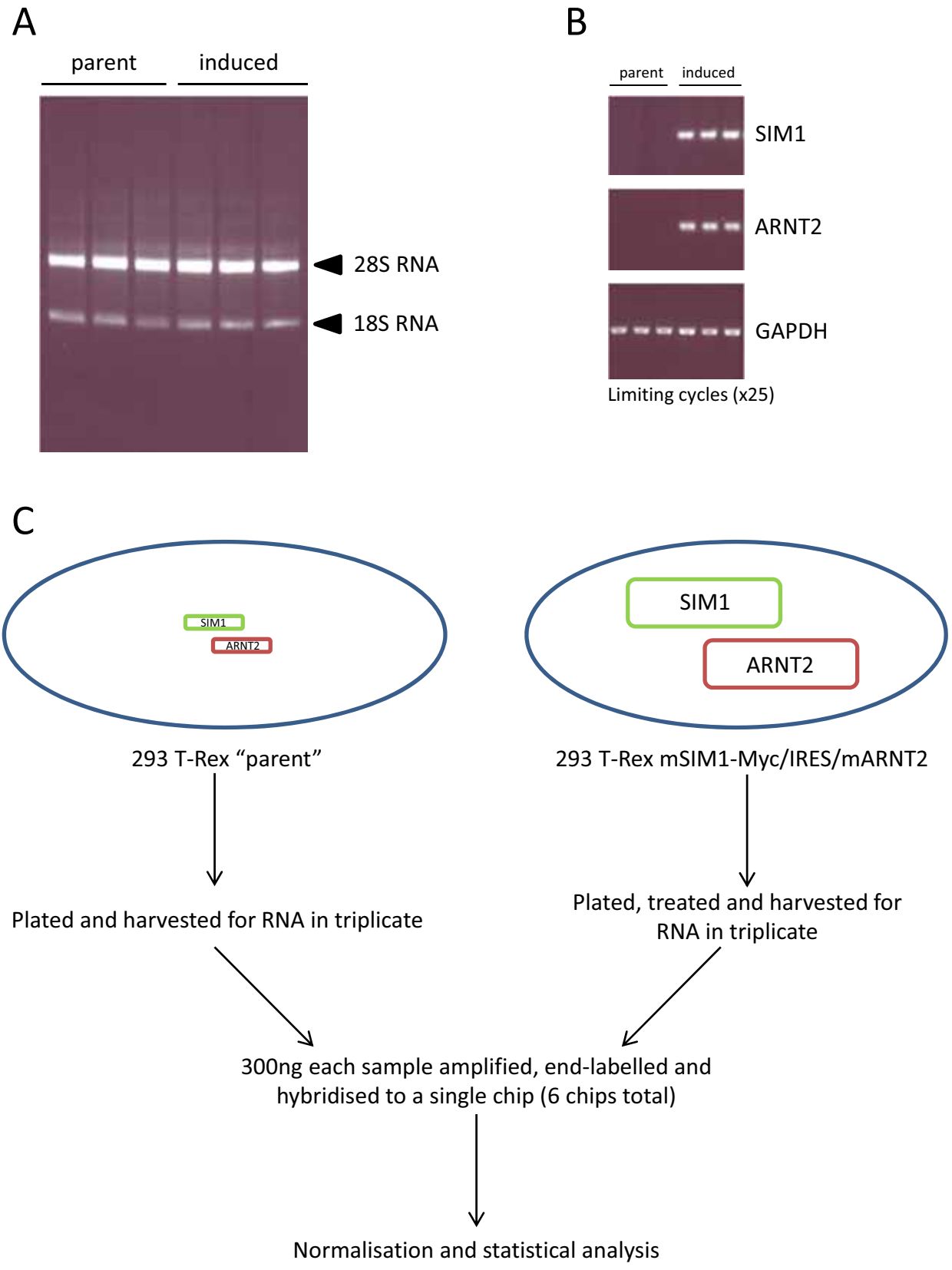


Figure 4.2. Validation of microarray RNA samples and schematic representation of experimental approach. A. Total RNA was prepared from 293 T-Rex parent cells and 293 T-Rex mSIM1-Myc/IRES/mARNT2 cells plated, treated (with or without doxycycline for 12hr) and harvested in triplicate. 1µg was then separated on a TBE/EtBr agarose gel and visualised by UV detection. **B**. 2µg each RNA sample was used to make cDNA, which was then assayed via RT-PCR for the presence of *Sim1* and *Arnt2*, with reference to the housekeeping gene *GAPDH*. **C**. Schematic representation of microarray analysis approach. Total RNA was prepared from 293 T-Rex parent cells and 293 T-Rex mSIM1-Myc/IRES/mARNT2 cells treated with doxycycline for 12hr in triplicate. 300ng of each sample was then analysed via microarray analysis using the Affymetrix single-label GeneChip Human Gene 1.0 ST Array platform.

Probeset ID	Gene Assignment	Gene Symbol	RefSeq	Fold Change	Description	P value	t statistic	B statistic
8144322	Myomesin (M-protein) 2, 165kDa // 8p23.3	*MYOM2*	NM_003970	3.122673136	upregulated	0.000125647	28.5945092	8.700408745
8112198	DKFZp686D0972 // similar to RIKEN cDNA 4732495G21 gene // 5q11.2	*ACTBL2*	NM_001017992	2.627064838	upregulated	0.000820243	18.6024963	7.223704982
8097288	FAT tumor suppressor homolog 4 (Drosophila) // 4q28.1	*FAT4*	NM_024582	2.480613005	upregulated	0.000228512	24.1336984	8.204073344
7985285	**Aryl hydrocarbon receptor nuclear translocator 2 // 15q24**	***ARNT2***	NM_014862	2.456607418	upregulated	0.001391368	16.0280478	6.541850393
8175531	Cerebellar degeneration-related protein 1, 34kDa // Xq27.1	*CDR1*	NM_004065	2.24183567	upregulated	0.004254645	11.72413	4.855463605
8128459	**Single-minded homolog 1 (Drosophila) // 6q16.3-q21**	***SIM1***	NM_005068	2.120645044	upregulated	0.001714488	14.8505663	6.160229667
8096070	Bone morphogenetic protein 3 (osteogenic) // 4q21	*BMP3*	NM_001201	1.992219107	upregulated	0.001391368	15.8109528	6.475199931
8174937	Odz, odd Oz/ten-m homolog 1(Drosophila) // Xq25	*ODZ1*	NM_014253	1.88976948	upregulated	0.001029125	17.2845648	6.898061733
8149835	Neurofilament, light polypeptide 68kDa // 8p21	*NEFL*	NM_006158	1.814376419	upregulated	0.000820243	19.1023664	7.335975071
8156706	Tropomodulin 1 // 9q22.3	*TMOD1*	NM_003275	1.797663589	upregulated	0.001029125	17.1035199	6.849671354
8166184	Carbonic anhydrase VB, mitochondrial // Xp21.1	*CA5B*	NM_007220	1.719819983	upregulated	0.001891322	14.3714614	5.989865461
7965166	Protein tyrosine phosphatase, receptor type, f polypeptide, interacting protein (liprin), alpha 2	*PPFIA2*	NM_003625	1.669402293	upregulated	0.0207078	7.74642397	2.265037604
8018902	Dynein, axonemal, heavy chain 17 // 17q25.3	*DNAH17*	NM_173628	1.664559671	upregulated	0.004242618	11.8687501	4.927160664
8149356	Deubiquitinating enzyme 3 // 8p23.1	*DUB3*	NM_201402	1.633678286	upregulated	0.001714488	15.1305563	6.255585374
8100578	EPH receptor A5 // 4q13.1	*EPHA5*	BX537946	1.621121953	upregulated	0.004042723	12.3452309	5.154637217
8030823	LOC402665 // hCG1651476 // 19q13.33	*LOC402665*	NM_001101372	1.626994573	upregulated	0.024376656	7.48600377	2.043179355
7971905	Protocadherin 20 // 13q21	*PCDH20*	NM_002843	1.61427807	upregulated	0.00664777	10.5558439	4.226684987
8171791	Small muscle protein, X-linked // Xp22.1	*SMPX*	NM_014332	1.609802278	upregulated	0.01056196	9.60476654	3.6416307
7918558	Potassium voltage-gated channel, Shal-related subfamily, member 3 // 1p13.3-p13.2	*KCND3*	NM_004980	1.603334858	upregulated	0.012280908	9.19845637	3.368912513
8111677	Leukemia inhibitory factor receptor alpha // 5p13-p12	*LIFR*	NM_002310	1.597965126	upregulated	0.004673012	11.5104222	4.747157665
8148059	DEP domain containing 6 // 8q24.12	*DEPDC6*	NM_022783	1.562574036	upregulated	0.016253049	8.43459159	2.814961041
7912145	Tumor necrosis factor receptor superfamily, member 9 // 1p36	*TNFRSF9*	NM_001561	1.556301435	upregulated	0.003509441	12.9499542	5.425140645
8089714	Limbic system-associated membrane protein // 3q13.2-q21	*LSAMP*	NM_002338	1.547065059	upregulated	0.015134405	8.73910464	3.042544461
7910915	Cholinergic receptor, muscarinic 3 // 1q43	*CHRM3*	AK056349	1.54254948	upregulated	0.020388255	7.77578793	2.289563319
8152522	Ectonucleotide pyrophosphatase/phosphodiesterase 2 // 8q24.1	*ENPP2*	NM_006209	1.534580494	upregulated	0.002395146	13.7727853	5.763383735

Figure 4.3. Top twenty-five clones from microarray experiments as determined by the method of Limma, ranked according to fold change. The gene name (**Gene Assignment**), gene symbol (**Gene Symbol**), and Genbank accession number (**RefSeq**) for each clone is indicated, as well as the corresponding Affymetrix probe set identification number (**Probeset ID**). **Fold change** refers to the ratio of the average signal intensity in the three "induced" samples over the three "parent" samples, and the **Description** indicates that expression of each gene is increased (upregulated) in the induced samples relative to the parent samples. All of the clones listed retained an adjusted **p value** < 0.05 after multiple test correction. For each clone, the **t statistic**, indicating the ratio of the log2 fold change to its standard error, and the **B statistic**, indicating the log odds that the gene is differentially expressed, are also listed. *SIM1* and *ARNT2* are highlighted in **green** and **red** respectively.

The genes we analysed (*MYOM2*, *ACTBL2*, *FAT4*, *BMP3*, *ODZ1*, *NEFL*, *TMOD1*, *PCDH20*, *DUB3*, *PPFIA2* and *CHRM3*) were chosen primarily based on their place in the top twenty-five clones differentially regulated in response to mSIM1-Myc and mARNT2 expression relative to parent cells (**Figure 4.3.**). We did endeavour, however, to choose genes that possessed links to *Sim1* expression and/or function *in vivo*. *MYOM2* has already been confirmed as a direct SIM2 target by our laboratory, and its expression is regulated via SIM2 binding to a core 5' AACGTG 3' motif within 1.3kb of the transcription start site ([84] and **Figure 4.14.**). Its expression within sarcomeres – the structural units of skeletal muscle fibres or myofibrils – suggests that, potentially, both SIM1 and SIM2 may influence its expression in muscle, although co-expression of *Sim1*, *Sim2* and *Myom2* in this organ at the cellular level has yet to be established [10, 49, 60, 84, 235-238]. *Myom2* transcripts have also been detected in other tissues including brain, kidney and heart, suggesting that it may possess additional unknown functions elsewhere in the body ([236-238] and the UniGene reference database <http://www.ncbi.nlm.nih.gov/unigene>). *TMOD1* is another gene best characterised in terms of its role in muscle, where it acts as an actin-capping protein during myofibril assembly [239-241]. Together with *MYOM2*, this may indicate a muscle-specific role for SIM1 in myofibril assembly, structure and function. *NEFL* is a well-characterised cytoskeletal component, important not only for maintaining cellular structure, but also facilitating the cytoskeletal rearrangements that occur in response to cell motility and axonal transport [242, 243]. The less well-characterised *ACTBL2* has greater than 90% amino acid homology to β-actin, which also possesses roles in these processes ([244-246] and data not shown). *FAT4* and *PCDH20* are both members of the cadherin superfamily of transmembrane proteins important for cell-cell and cell-matrix adhesion in multiple organs including brain and kidney [247, 248]. *FAT4* has also been previously implicated in intracellular signalling pathways important for kidney formation, suggesting a developmental role for SIM1 in this organ [249-252]. *PPFIA2*, a receptor tyrosine phosphatase (PTPR) interacting protein, is a member of a larger family of PTPR interacting proteins (liprins) that have been extensively studied in terms of their influence on axon guidance, synapse formation and function, neurotransmitter release and cell adhesion [253-256]. *ODZ1* is a member of the tenascin family of extracellular matrix glycoproteins that influences neuronal migration and pathfinding, and may also be involved in signal transduction via cleavage of its intracellular domain [257-260]. Interestingly, it has also been recently identified by Caqueret *et al* as a gene whose expression is reduced in *Sim1*$^{-/-}$ hypothalamic tissues

relative to WT tissues, highlighting its potential as a genuine SIM1-regulated gene [57]. *BMP3* is a member of the TGF-β superfamily of secreted extracellular messenger ligands that influence cell proliferation via cell surface serine/threonine kinase receptors. Several members of the *Bmp* family have kidney phenotypes in the absence of both alleles in mice, including defects in ureteric branch growth, mesenchymal survival, nephrogenesis and glomerulogenesis [261-263]. *DUB3* is a member of the ubiquitin processing protease subfamily of deubiquitinating enzymes, and is expressed at the transcript level in heart, skeletal muscle, colon, liver, kidney and brain [264]. Finally, *CHRM3* is a transmembrane G protein-coupled receptor (GPCR) protein that has been implicated in leptin-melanocortin signalling in response to food intake, making it a subject of particular interest in terms of SIM1 function in the hypothalamus [265].

Fortunately, ten of the eleven chosen clones displayed a greater degree of differential expression in treated 293 T-Rex mSIM1-Myc/IRES/mARNT2 cells relative to parent 293 T-Rex cells by QPCR than what was observed by microarray analysis, including a 125 fold increase in *Sim1* expression (**Figure 4.4.**). We concluded that our microarray analyses had yielded sufficiently reliable data to warrant further analysis, and therefore decided to pursue these ten candidates as potential downstream SIM1 target genes in additional studies.

4.4. Validation of differential expression in independent samples by QPCR

We designed the next set of experiments to verify that the genes we had chosen were indeed being upregulated specifically in response to SIM1 overexpression. We analysed transcript levels for each of the ten candidate genes by QPCR in two independently derived 293 T-Rex mSIM1-Myc/IRES/mARNT2 stable cell lines, two independently derived 293 T-Rex mARNT2 stable cell lines, one 293 T-Rex empty vector stable cell line, and the parent 293 T-Rex cell line (**Figure 4.5.A.**). Unlike the parent 293 T-Rex cell line, the 293 T-Rex empty vector stable cell line had undergone Flp recombinase-mediated integration with an "empty" expression vector, and subsequent antibiotic selection, in a manner similar to that experienced by the 293 T-Rex mSIM1-Myc/IRES/mARNT2 and 293 T-Rex mARNT2 stable cell lines. It therefore served as a control for any gene expression changes that may be occurring as a consequence of the

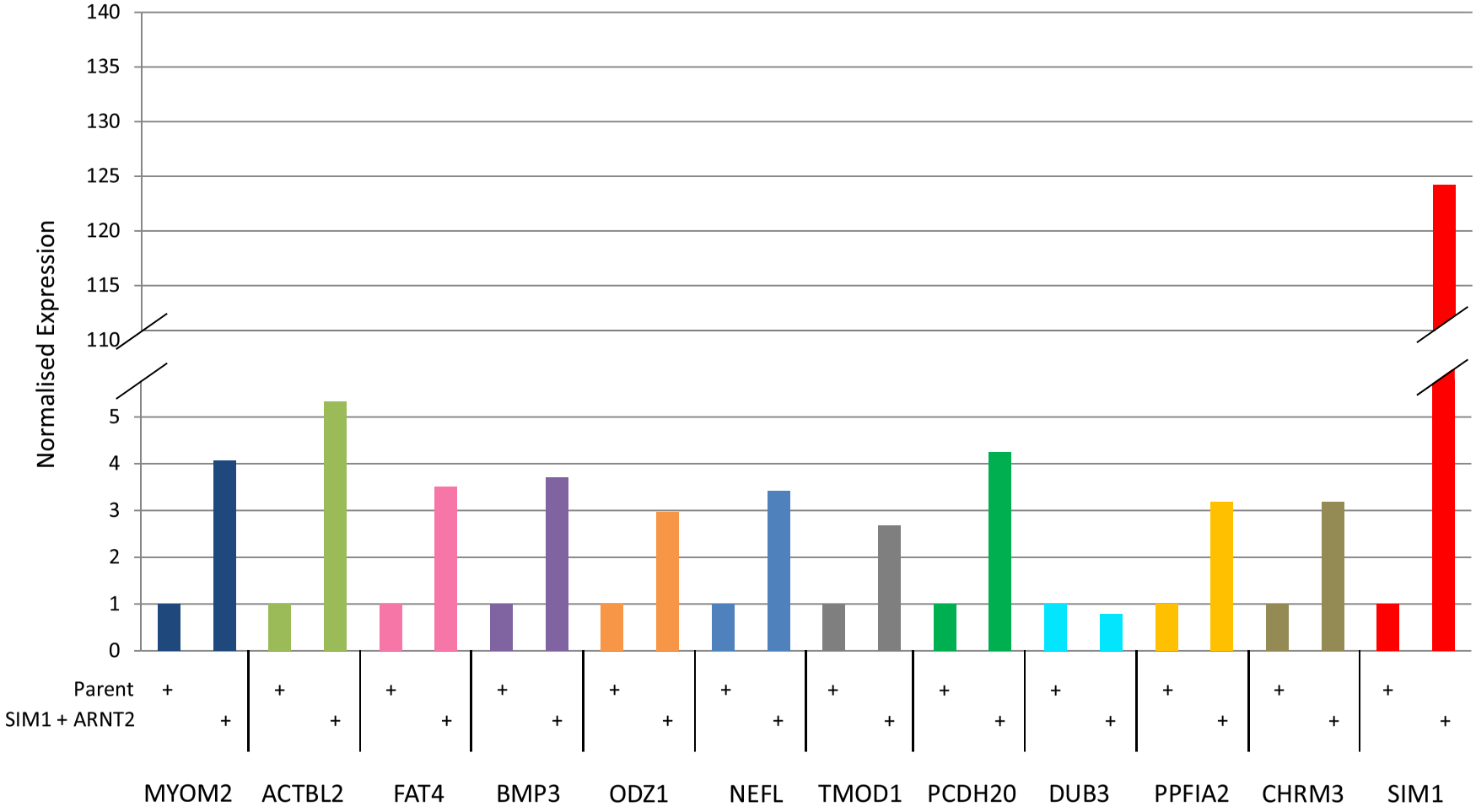


Figure 4.4. Validation of putative downstream SIM1/ARNT2 target genes in microarray samples. cDNA was synthesised from an aliquot of the RNA samples subjected to microarray analysis and subjected to QPCR analysis using gene-specific primers. Results are from a single experiment performed in triplicate normalised to the reference gene *POLR2A*. Transcript expression is depicted as a fold change relative to the parent cell line, which has been normalised to 1 for each gene.

A

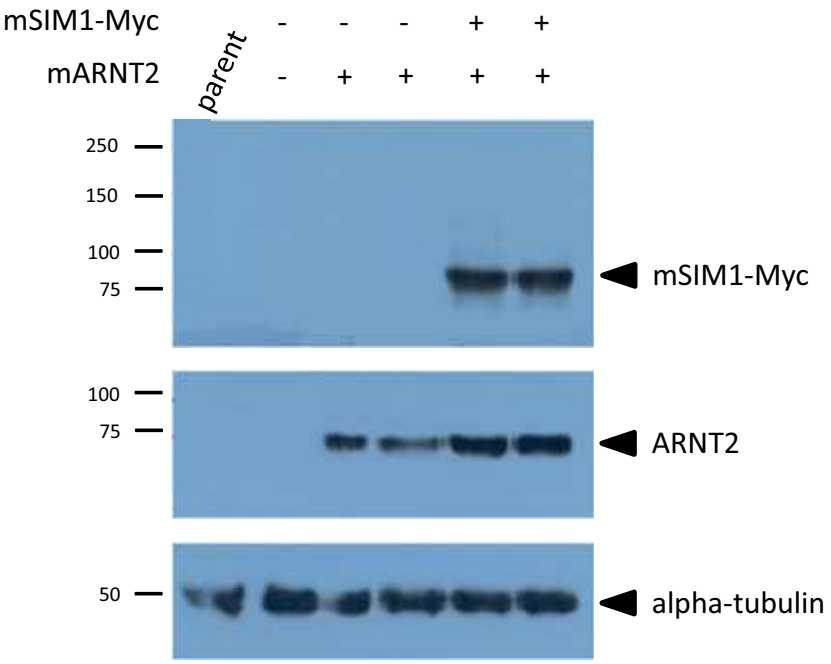
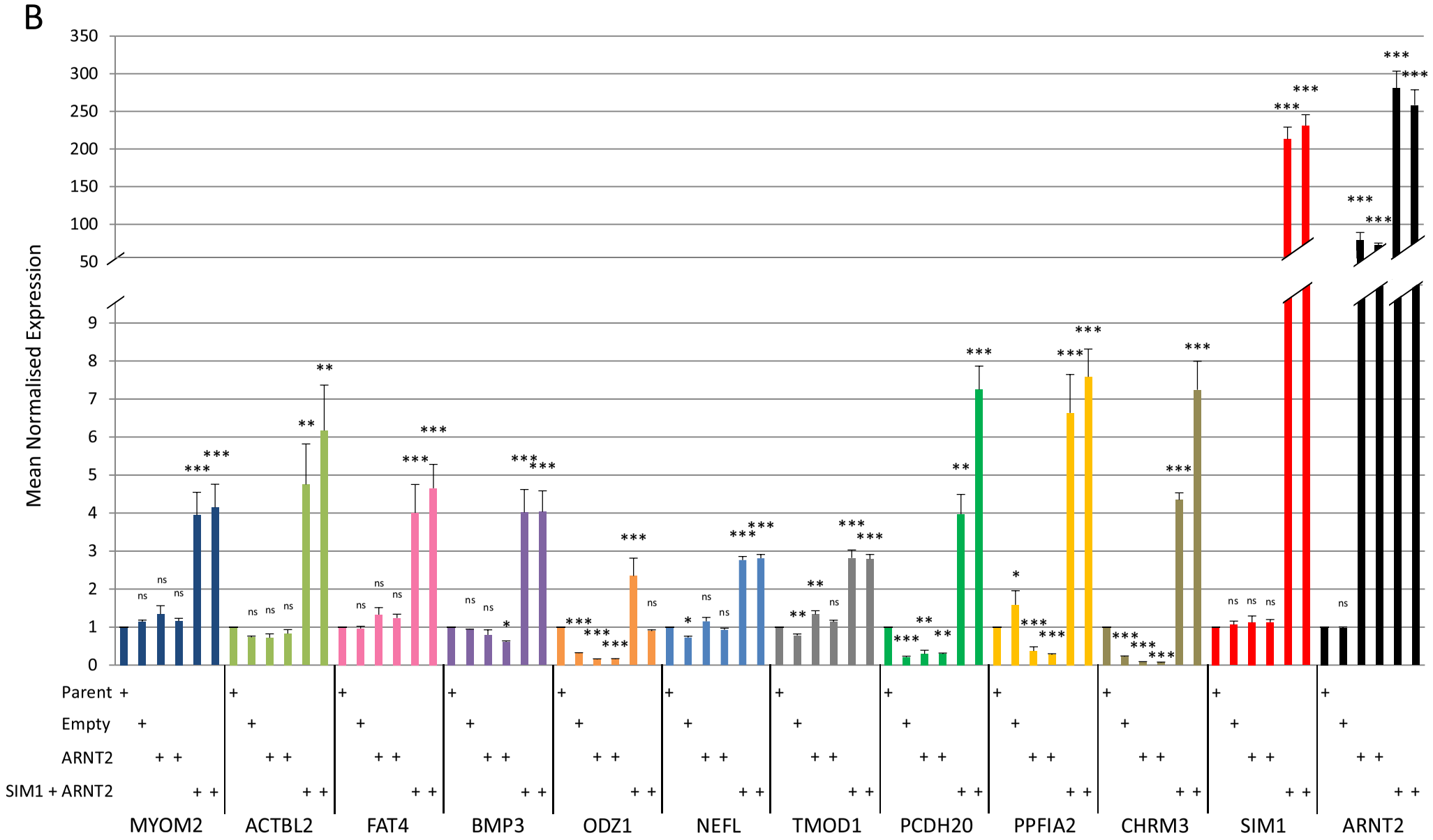


Figure 4.5. Validation of putative downstream SIM1/ARNT2 target genes in independently derived 293 T-Rex samples. Parent 293 T-Rex cells, a single empty vector 293 T-Rex stable cell line, two independently derived mARNT2 293 T-Rex stable cell lines, and two independently derived mSIM1/IRES/mARNT2 293 T-Rex stable cell lines were treated with doxycycline to a final concentration of 1µg/mL for 12hr. **A**. Western blot analysis of mSIM1-Myc and ARNT2 overexpression in WCEs. 30µg WCE from each cell line was separated by SDS-PAGE, transferred to nitrocellulose, and probed with 4A6 anti-Myc monoclonal antibody, M-165 anti-ARNT2 polyclonal antibody and MCA78G anti-alpha-tubulin monoclonal antibody. **B**. QPCR analysis of differential target gene transcript levels. RNA was prepared from each cell line, and cDNA subsequently synthesised and subjected to QPCR analysis using gene-specific primers. Results are the average of three experiments performed in triplicate +SEM normalised to the reference gene *POLR2A*. Transcript expression is depicted as a fold change relative to the parent cell line, which has been normalised to 1 for each gene. P values derived from univariate ANOVA indicating likelihood of differential expression relative to the parent cell line: *** <0.001, ** <0.01, * <0.05, ns not significant.


B
Mean Normalised Expression
350
300
250
200
150
100
50
9
8
7
6
5
4
3
2
1
0
Parent +
Empty +
ARNT2 + +
SIM1 + ARNT2 + +
MYOM2
ACTBL2
FAT4
BMP3
ODZ1
NEFL
TMOD1
PCDH20
PPFIA2
CHRM3
SIM1
ARNT2

recombination and antibiotic selection process. Three of the ten genes analysed (*MYOM2*, *ACTBL2* and *FAT4*) showed selective and highly reproducible increases in transcript levels only upon doxycycline induction of both mSIM1-Myc and mARNT2 (**Figure 4.5.B.**). Of the remaining genes, three (*BMP3*, *NEFL* and *TMOD1*) showed subtle but statistically significant changes in transcript levels between empty vector- and/or mARNT2-expressing cells relative to parent cells, as well as robust induction in response to mSIM1-Myc and mARNT2 overexpression; three (*ODZ1*, *PPFIA2* and *CHRM3*) displayed significantly reduced transcript levels in response to mARNT2 induction relative to both parent and empty vector cells, as well as robust induction in response to mSIM1-Myc and mARNT2 overexpression; and finally, *PCDH20* displayed equivalent reductions in transcript levels in empty vector- and mARNT2-overexpressing cells relative to parent 293 T-Rex cells, as well as strong induction upon doxycycline induction of both mSIM1-Myc and mARNT2 (**Figure 4.5.B.**). It is worth noting that the fold changes observed for each gene in doxycycline treated 293 T-Rex mSIM1-Myc/IRES/mARNT2 cells above the parent 293 T-Rex cell line are in good agreement with those observed in initial QPCR analyses of our microarray RNA samples (**Figure 4.4.**). We therefore concluded that, despite some evidence of SIM1-independent changes in candidate target gene transcript levels, our microarray analyses had yielded a number of genuine downstream SIM1 target genes. We therefore decided to pursue them in additional experiments.

4.5. Analysis of differential expression in independent cell systems

4.5.1. Analysis of differential expression upon endogenous *SIM1* knockdown in 293T cells

Next, we decided to examine the transcript levels of each candidate target gene in response to shRNA-mediated knockdown of endogenous *SIM1* in 293T cells. We designed four shRNA constructs against the *SIM1* transcript, and first tested their ability to reduce ectopic SIM1 expression upon transient transfection in 293T cells. All four shRNA sequences showed strong repression of hSIM1-Myc by Western analysis relative to a control (scrambled) shRNA (**Figure 4.6.A.**). We then subcloned three of the four shRNA sequences into a single vector, doxycycline-inducible shRNA expression plasmid and generated stable 293T cell lines via lentiviral infection. In the absence of an anti-

SIM1 antibody capable of detecting endogenous SIM1 protein we instead tested the ability of each sequence to knock down endogenous *SIM1* transcript by QPCR. All three sequences reduced *SIM1* transcript levels by 40-50% upon doxycycline induction relative to untreated cells (**Figure 4.6.B.**).

Unfortunately, subsequent QPCR analysis of candidate target gene expression in one of these cell lines did not show any substantial, *SIM1* knockdown-specific differences in transcript levels upon doxycycline treatment, although transcript levels of *BMP3* were slightly more reduced in *SIM1* knockdown cells relative to control (scrambled) cells (**Figure 4.7.**). Transcript levels of *FAT4* and *PPFIA2* were slightly increased in doxycycline treated *SIM1* knockdown cells relative to control cells, which contradicts evidence of their substantial induction in 293 T-Rex cells upon mSIM1-Myc and mARNT2 overexpression (**Figure 4.5.B.**). We concluded that either the level of *SIM1* knockdown was insufficient to reduce expression of its downstream targets, or that there are compensatory factors present in 293T cells that are capable of effecting downstream target gene expression despite a reduction in endogenous SIM1 levels. In a mechanistic sense, however, it is also possible that these genes only react to induced or upregulated SIM1. This may occur *in vivo* in response to some kind of physiological stimulus, which can be mimicked in cell culture by overexpressing SIM1 (and ARNT2) from cDNA, and would preclude any "basal" regulation of these genes under conditions of normal or reduced SIM1 levels. Such a model is reflective of the events that may occur in PVN/SON/aPV cells in response to αMSH signalling. We therefore adopted an alternative approach towards assessing the veracity of our ten chosen candidates as genuine downstream SIM1 target genes, by testing their responsiveness to SIM1 and/or ARNT2 expression in a newly characterised hypothalamus derived cell culture model.

4.5.2. Analysis of differential expression in hypothalamus derived cell lines

The hypothalamus is a remarkably heterogeneous collection of unique cell populations (nuclei), each possessing a distinct function and anatomical position *in vivo* (**Figure 1.3.**, for review see [266, 267]). It can therefore be argued that relevant analyses of SIM1 function in the hypothalamus should ideally occur within *Sim1*-expressing, PVN/SON/aPV-derived cells. As aforementioned, our inability to obtain *Sim1*$^{+/-}$ mice and the lack of a suitable hypothalamus derived cell line within which to manipulate *Sim1*

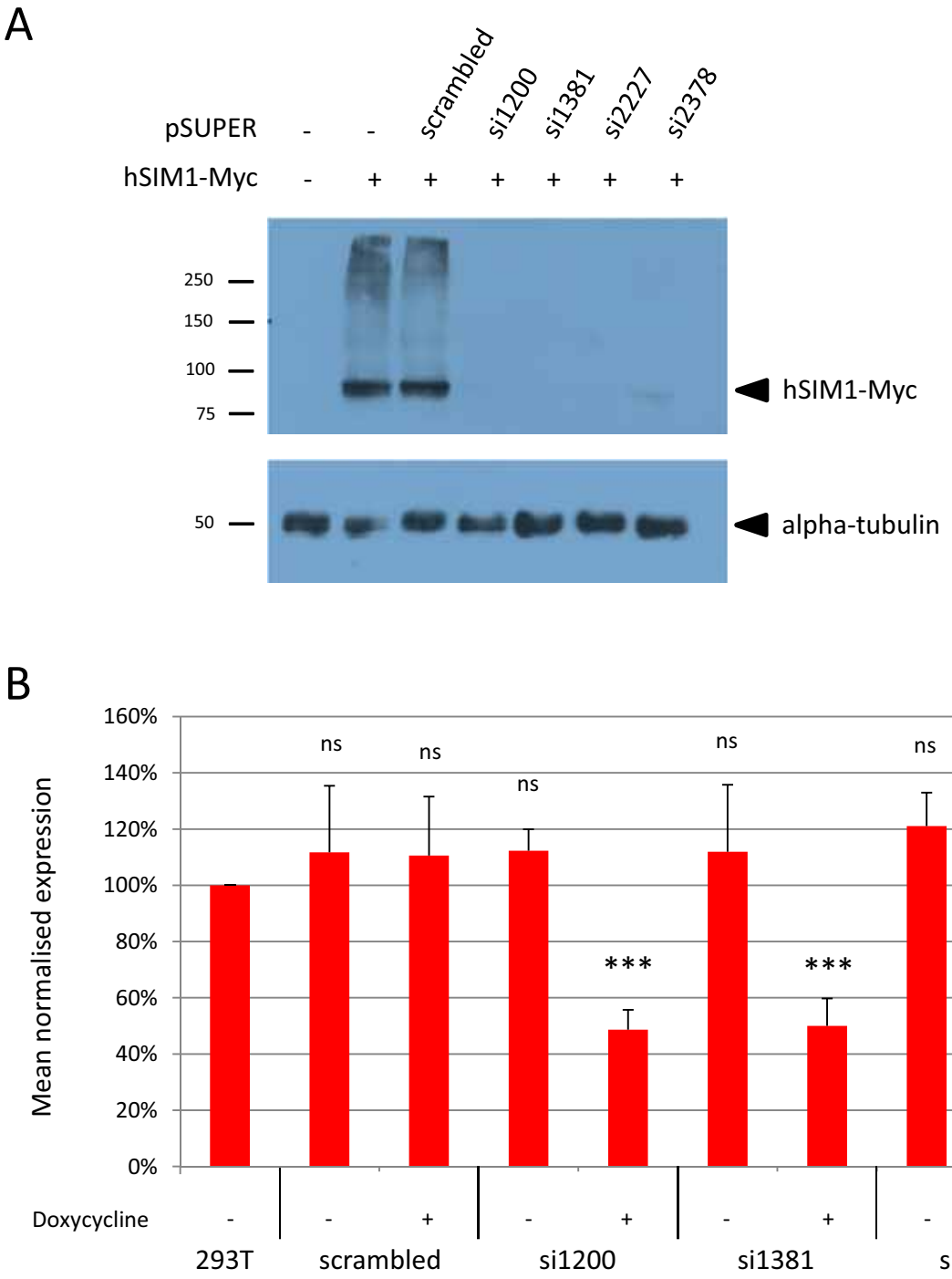


Figure 4.6. **Validation of SIM1 knockdown in 293T cells. A**. Western blot analysis of hSIM1-Myc knockdown via transient transfection. 293T cells were transiently transfected with 0.5µg hSIM1-Myc expression plasmid (+) or empty vector (-), as well as 5µg empty pSUPER vector (-) or pSUPER vector expressing either a scrambled shRNA sequence or one of four *SIM1* shRNA sequences (si1200, si1381, si2227 or si2378) and incubated for 24hr before harvesting. 30µg WCE was then separated by SDS-PAGE, transferred to nitrocellulose, and probed with 4A6 anti-Myc monoclonal antibody and MCA78G anti-alpha-tubulin monoclonal antibody. **B**. Analysis of endogenous *SIM1* knockdown in 293T stable cell lines via QPCR. 293T stable cell lines inducibly overexpressing a control (scrambled) shRNA sequence or one of three SIM1 shRNA sequences were treated with (+) or without (-) doxycycline to a final concentration of 5µg/mL for 48hr before harvesting for RNA. cDNA was subsequently synthesised and subjected to QPCR analysis using gene-specific primers. Results shown are the average of three experiments performed in triplicate +SEM normalised to the reference gene *POLR2A*, expressed relative to untreated 293T cells that have been normalised to 100%. P values derived from univariate ANOVA indicating likelihood of differential expression relative to untreated 293T cells: *** <0.001, ** <0.01, * <0.05, ns not significant.

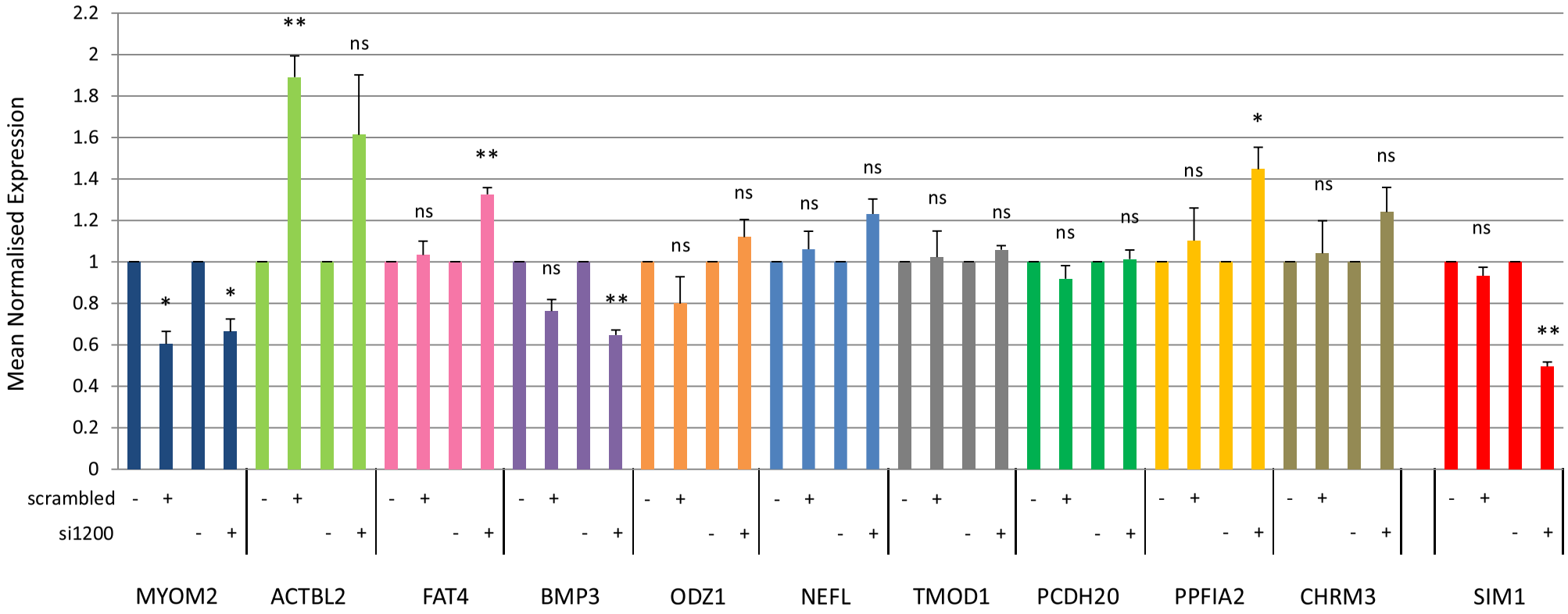


Figure 4.7. Analysis of putative downstream SIM1/ARNT2 target genes upon *SIM1* knockdown in 293T cells. 293T stable cell lines inducibly overexpressing a control shRNA sequence (scrambled) or a *SIM1* shRNA sequence (si1200) were treated with (+) or without (-) doxycycline to a final concentration of 5µg/mL for 48hr before harvesting for RNA. cDNA was subsequently synthesised and subjected to QPCR analysis using gene-specific primers. Results are the average of three experiments performed in triplicate +SEM normalised to the reference gene *POLR2A*. Transcript expression for each gene in each cell line (scrambled or si1200) is depicted as a fold change relative to untreated cells, which has been normalised to 1. P values derived from two-tailed paired t test indicating likelihood of differential expression relative to untreated cells for each gene: *** <0.001, ** <0.01, * <0.05, ns not significant.

expression levels had restricted our initial studies of SIM1 activity and function to HEK-derived cells (**Section 4.1.**). However, during the course of our microarray validation studies, Belsham *et al* described the generation of a host of immortalised neuronal cell lines derived from the hypothalamus of E15.0-18.0 mice [224]. These cell lines were generated via retroviral infection of primary hypothalamic neurons with the SV40 large T antigen and a neomycin resistance gene, which enabled subsequent selection with G418 and expansion of antibiotic resistant colonies [224]. Morphological and gene expression data indicated that some of these cell lines might be derived from the PVN/SON/aPV region, and therefore offered the opportunity to study hypothalamic regulation of downstream SIM1 target gene expression in a suitably relevant cell culture model ([224] and <http://www.cellutionsbiosystems.com/index.php?option=com_frontpage&Itemid=1>). We therefore purchased three of these cell lines (N4, N7 and N39) and attempted to characterise them via IHC and RT-PCR analysis.

4.5.2.1. Marker gene analysis of immortalised, embryonic hypothalamus derived cell lines N4, N7 and N39

First, we performed IHC analysis of our cell lines using antibodies directed against the neuronal precursor marker Nestin, the oligodendrocytic marker RIP, and the astrocytic marker Glial Fibrillary Acidic Protein (GFAP). All three cell lines were strongly positive for Nestin and did not express RIP, indicating that they were of the neuronal lineage (**Figure 4.8.**). To our surprise, however, all three cell lines also stained positive for GFAP (**Figure 4.8.**). RT-PCR analysis of *Nestin* and *Gfap* transcript levels in these cell lines confirmed their abundant expression of *Nestin*, although only weak levels of *Gfap* could be detected (**Figure 4.9.iii**). *Nestin* and *Gfap* are known to be co-expressed in neural cells, prior to their absolute commitment to either the neuronal or astrocytic lineage [268]. We therefore anticipate that these cells represent a type of undifferentiated neuronal precursor cell that has yet to commit fully to the neuronal lineage. This inference is supported by further RT-PCR analyses showing an absence of detectable *Neuro D* expression (a marker of more differentiated neurons), as well as *Crh*, *Trh*, *Ss*, *Avp* and *Oxt*, which are expressed in terminally differentiated neurons of the PVN, SON and aPV (**Figure 4.9.iii**). Excitingly, N39 cells were strongly positive for *Sim1* expression, and all three cell lines express the general partner factors *Arnt* and *Arnt2*. None of the three cell lines expressed detectable *Sim2* or *Mc4r* (**Figure 4.9.i**).

We also attempted limited characterisation of each cell line according to the expression of four genes expressed along the anterior-posterior and dorsal-ventral axes of the hypothalamus (**Figure 4.9.ii**). N39 cells (and, to a lesser extent, N4 and N7 cells) were positive for *Six3*, which displays a limited degree of overlap with *Sim1* in the LHA *in vivo* ([93] and **Figure 4.9.ii**). As expected, all three cell lines were negative for *Rax*, which is expressed within the earliest hypothalamic progenitor cells immediately lining the ventricular zone ([269] and **Figure 4.9.ii**). Additionally, none was positive for the ventral hypothalamic marker *Mash1* or the dorsal hypothalamic marker *Pax6*, which is expressed more anteriorly to the prospective PVN/SON region *in vivo* ([270-272] and **Figure 4.9.ii**).

Precise characterisation of hypothalamic cell lineage on the basis of marker gene analysis is difficult given the lack of region-specific markers, and it is also possible that the immortalisation and antibiotic selection process has caused these cell lines to drift somewhat from their initial identity [57, 269]. From these results, we concluded that N39 cells offered the most promising context within which to study hypothalamic regulation of downstream target genes by SIM1. We therefore decided to test the response of our candidate target genes to SIM1 and ARNT2 overexpression in N39-derived stable cell lines.

4.5.2.2. Analysis of differential expression in N39 cells stably overexpressing SIM1 and ARNT2

We initially generated N39 stable cell lines constitutively overexpressing empty vector, mARNT2, or mSIM1-Flag and mARNT2 via lentiviral infection (**Figure 4.10.A.**), and analysed the transcript levels of each of the ten candidate target genes by QPCR. Six of the ten genes were not expressed at sufficiently high levels to allow for accurate gene expression analysis (data not shown). Of the remaining four, none showed any reproducible difference in transcript levels upon mSIM1-Flag and mARNT2 overexpression relative to empty vector cells or mARNT2 overexpressing cells (**Figure 4.11.A.**).

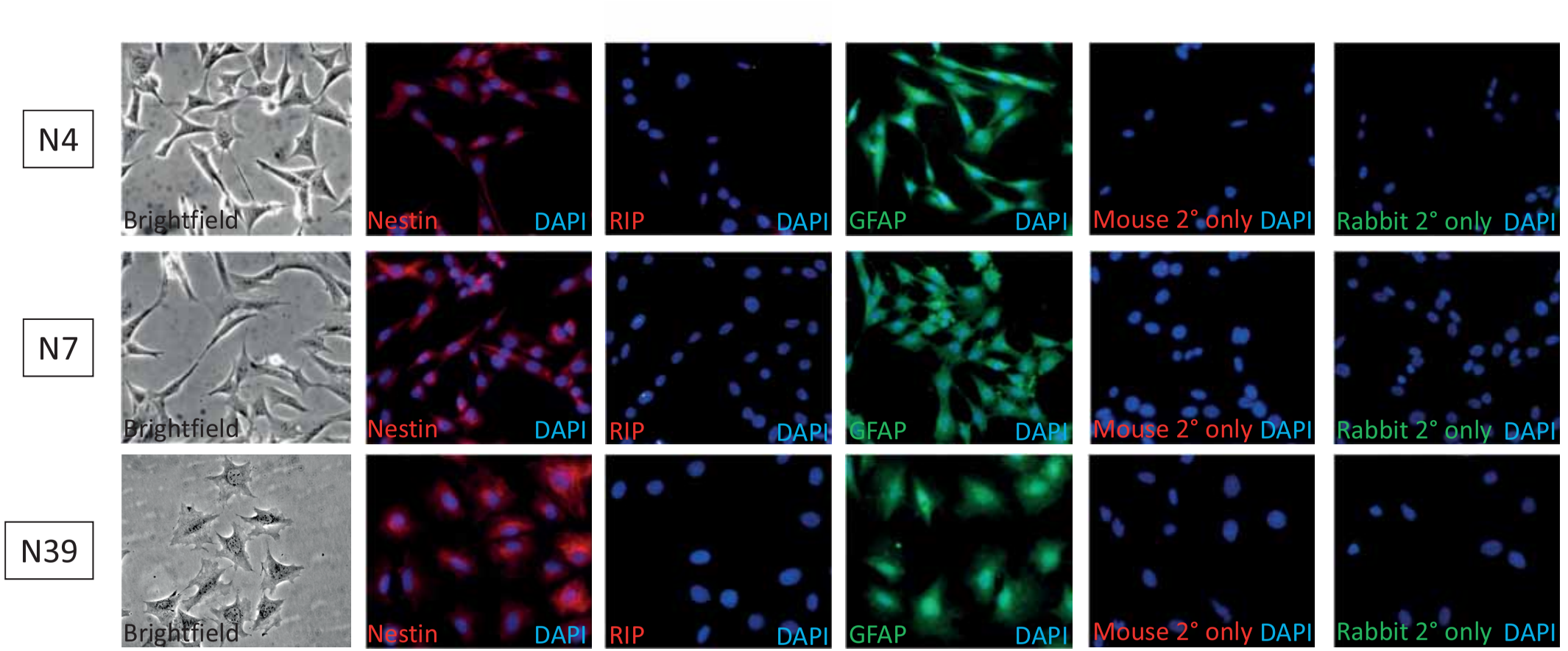


Figure 4.8. The embryonic hypothalamus-derived cell lines N4, N7 and N39 display expression patterns characteristic of neuronal precursor cells. Cells were seeded onto coverslips, fixed with 4% paraformaldehyde, and stained with antibodies against the early neuronal marker Nestin, the oligodendrytic marker RIP and the astrocytic marker GFAP, as well as Alexa Fluor 594-conjugated anti-mouse (Nestin, RIP, in **red**) and anti-rabbit (GFAP, in **green**) secondary antibodies. Nuclei were visualised using DAPI (in **blue**). A brightfield image of each cell line is included on the left. All images at 20x magnification.

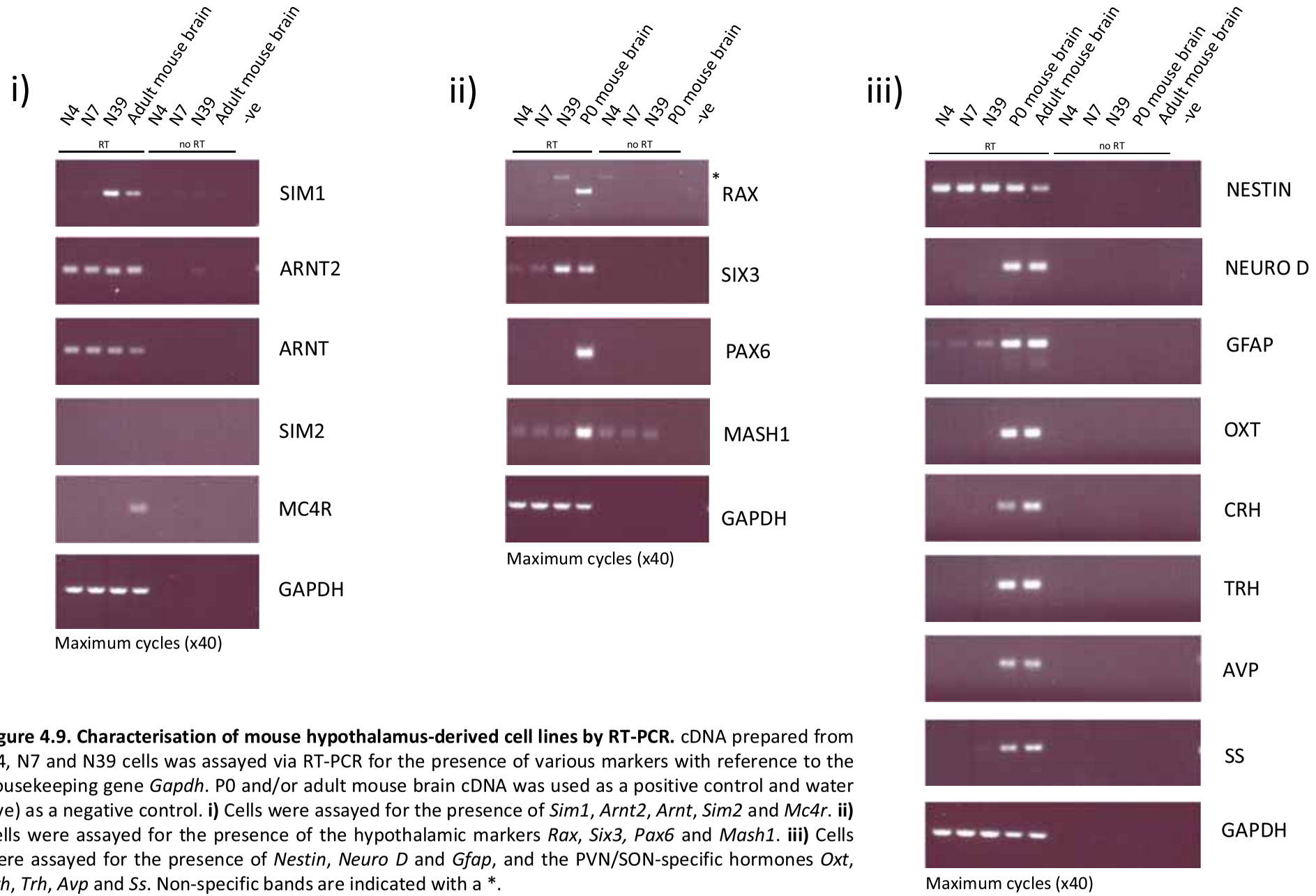


Figure 4.9. Characterisation of mouse hypothalamus-derived cell lines by RT-PCR. cDNA prepared from N4, N7 and N39 cells was assayed via RT-PCR for the presence of various markers with reference to the housekeeping gene *Gapdh*. P0 and/or adult mouse brain cDNA was used as a positive control and water (-ve) as a negative control. **i)** Cells were assayed for the presence of *Sim1*, *Arnt2*, *Arnt*, *Sim2* and *Mc4r*. **ii)** Cells were assayed for the presence of the hypothalamic markers *Rax*, *Six3, Pax6* and *Mash1*. **iii)** Cells were assayed for the presence of *Nestin*, *Neuro D* and *Gfap*, and the PVN/SON-specific hormones *Oxt*, *Crh*, *Trh*, *Avp* and *Ss*. Non-specific bands are indicated with a *.

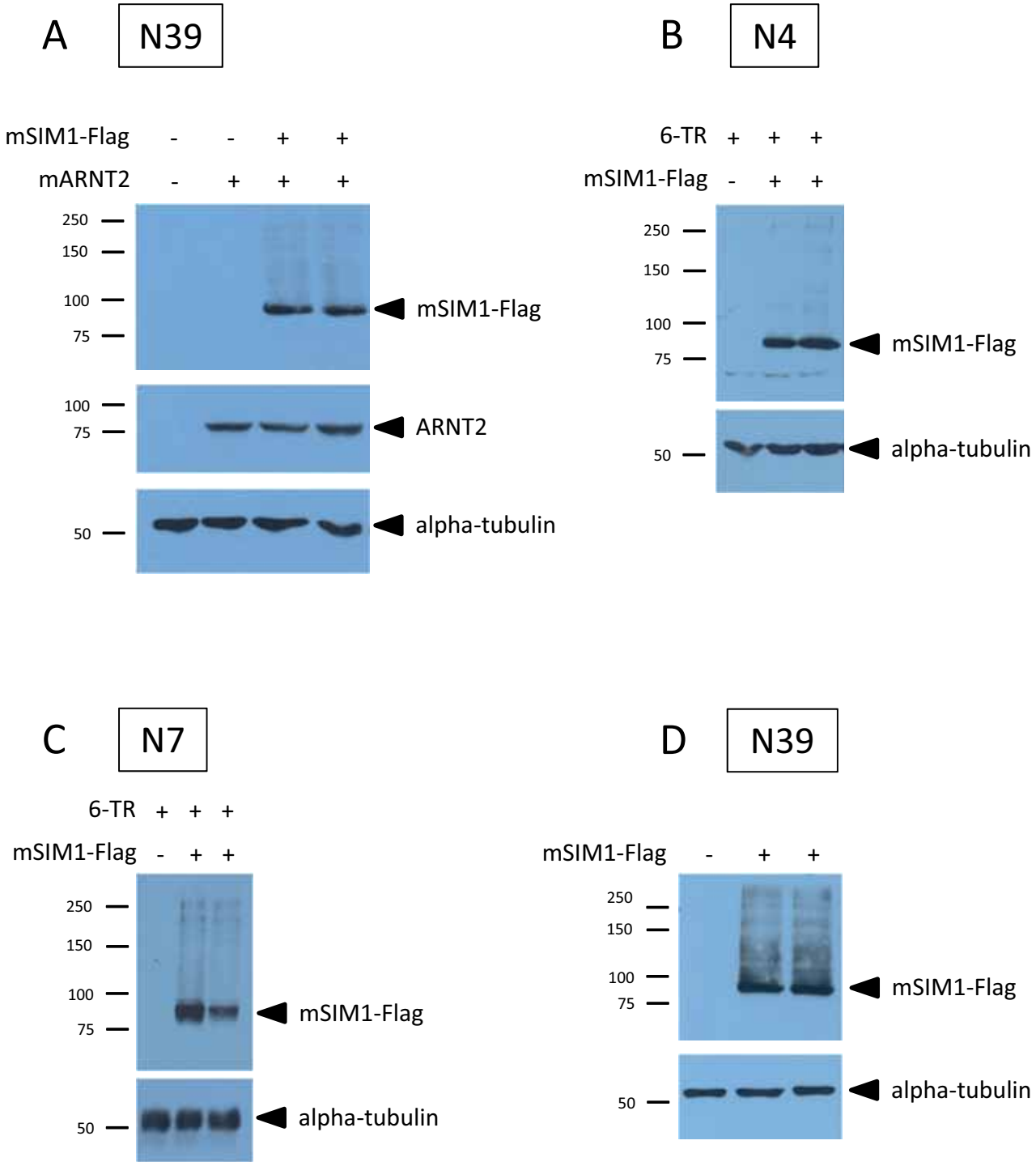
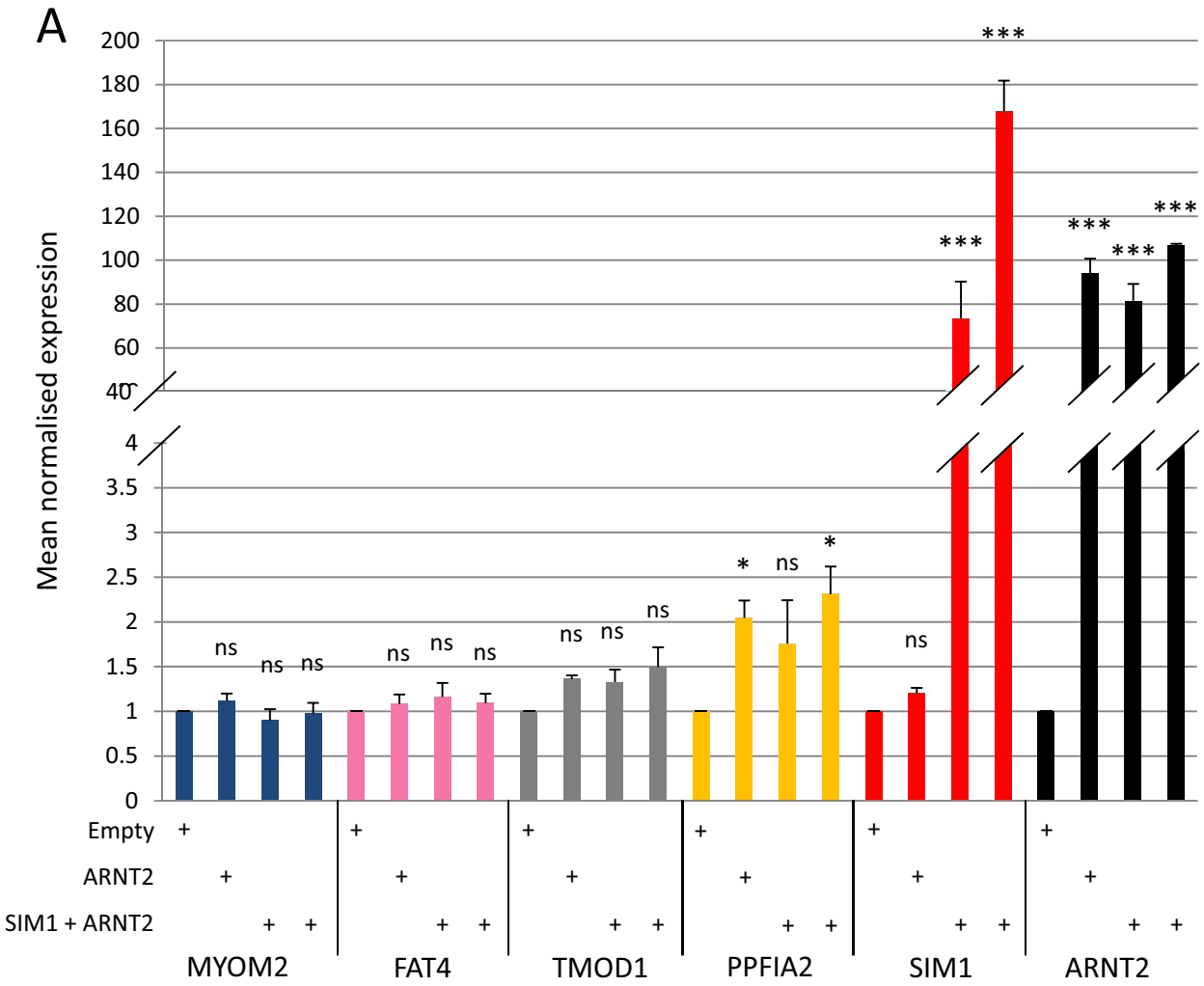
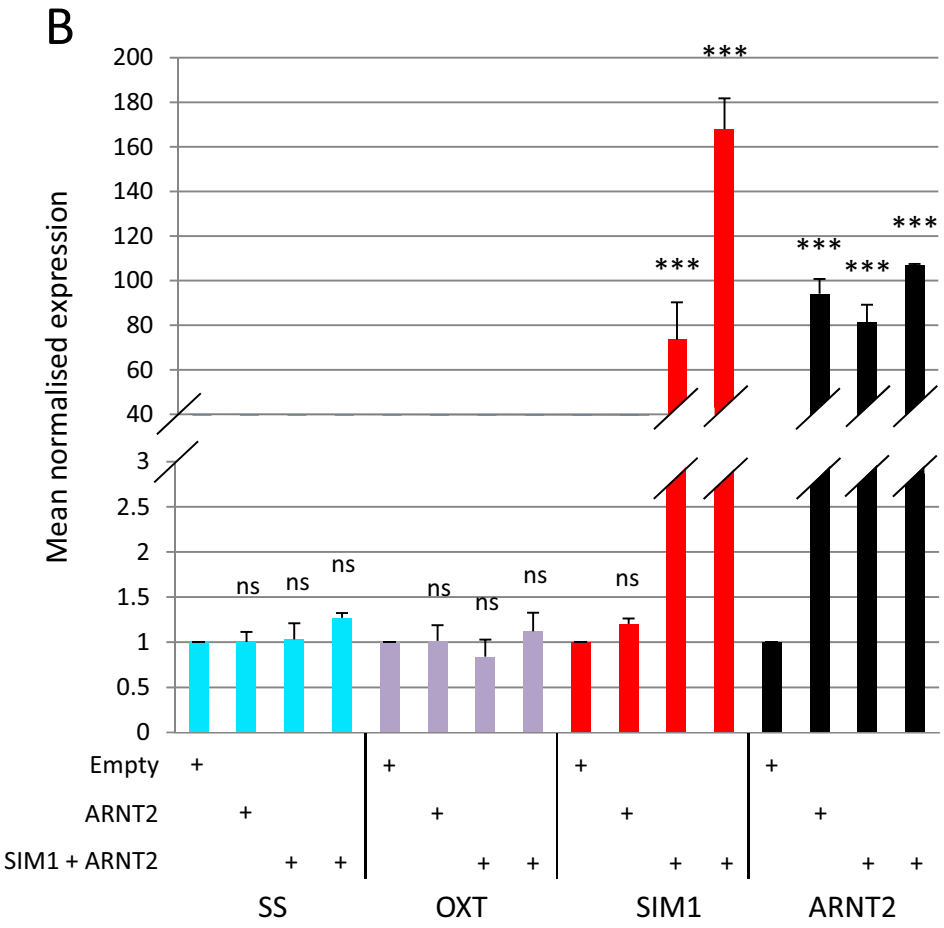


Figure 4.10. Validation of SIM1 and/or ARNT2 protein overexpression in hypothalamic stable cell lines. 30µg WCE from each cell line was separated by SDS-PAGE, transferred to nitrocellulose, and probed with M2 anti-Flag monoclonal antibody, M-165 anti-ARNT2 polyclonal antibody and/or MCA78G anti-alpha-tubulin monoclonal antibody. **A.** Western blot analysis of Flag-tagged mSIM1 and mARNT2 overexpression in N39 cells expressing empty vector, mARNT2 expression vector, or mSIM1-Flag and mARNT2 expression vectors via lentiviral infection. **B.** Western blot analysis of Flag-tagged SIM1 overexpression in N4 cells stably expressing a TetR (6-TR) expression vector, or a TetR expression vector and a mSIM1-Flag expression vector, via lentiviral infection. **C.** Western blot analysis of Flag-tagged SIM1 overexpression in N7 cells stably expressing a TetR (6-TR) expression vector, or a TetR expression vector and a mSIM1-Flag expression vector, via lentiviral infection. **D.** Western blot analysis of Flag-tagged SIM1 overexpression in N39 cells expressing empty vector (-) or a mSIM1-Flag expression vector (+) via lentiviral infection.

Figure 4.11. Analysis of putative downstream SIM1/ARNT2 target genes in hypothalamus-derived N39 cells overexpressing SIM1 and ARNT2. A-B. One empty vector N39 stable cell line (Empty), one mARNT2 N39 stable cell line (ARNT2), and two independently derived mSIM1/mARNT2 N39 stable cell lines (SIM1 + ARNT2) were harvested for RNA. cDNA was subsequently synthesised and subjected to QPCR analysis using gene-specific primers. Results are the average of three experiments performed in triplicate +SEM normalised to the reference gene *Polr2a*. Transcript expression for each expressed gene is depicted as a fold change relative to the empty vector cell line, which has been normalised to 1. P values derived from univariate ANOVA indicating likelihood of differential expression relative to empty vector cells: *** <0.001, ** <0.01, * <0.05, ns not significant.


A
Mean normalised expression
200
180
160
140
120
100
80
60
40
4
3.5
3
2.5
2
1.5
1
0.5
0
***
***
***
***
***
ns
ns
ns
ns
ns
ns
ns
ns
ns
ns
*
ns
*
ns
Empty
ARNT2
SIM1 + ARNT2
MYOM2
FAT4
TMOD1
PPFIA2
SIM1
ARNT2


B
Mean normalised expression
200
180
160
140
120
100
80
60
40
3
2.5
2
1.5
1
0.5
0
***
***
***
***
***
ns
ns
ns
ns
ns
ns
ns
ns
Empty
ARNT2
SIM1 + ARNT2
SS
OXT
SIM1
ARNT2

4.5.2.3. Additional analyses of candidate hypothalamic downstream target genes

As previously discussed, PVN *Mc4r* mRNA levels are reduced in conditional *Sim1*$^{-/-}$ and *Sim1*$^{+/-}$ mice relative to littermate controls [98]. Furthermore, reduced levels of *Crh*, *Trh*, *Ss*, *Avp* and *Oxt* mRNA have all been detected in *Sim1*$^{+/-}$ hypothalamus relative to WT [97, 153]. Reduced *Sim2* expression in *Sim1*$^{-/-}$ hypothalamic tissues has also been observed on a few occasions [57, 63]. These data suggest the possibility of direct regulation of these genes by SIM1 in the hypothalamus, although in the case of *Oxt* this has already been suggested to occur in an indirect manner [153]. We were therefore interested in testing the ability of *Crh*, *Trh*, *Ss*, *Avp* and *Oxt*, as well as *Mc4r* and *Sim2*, to respond to mSIM1-Flag and mARNT2 overexpression in N39 cells. As can be seen from **Figure 4.9.iii**, none of these genes is expressed at a level able to be detected by ethidium bromide staining in N4, N7 or N39 cells. QPCR analysis, however, was able to detect low levels of *Oxt* and *Ss* expression in N39 empty vector cells. Unfortunately, neither gene responded to mSIM1-Flag and mARNT2 overexpression (**Figure 4.11.B.**).

4.5.2.4. Analysis of differential expression in N4, N7 and N39 cells stably overexpressing SIM1

In a final attempt to discern whether any of our genes of interest responded to SIM1 expression in a hypothalamic context, we tested the expression levels of each gene in N4 and N7 cells, as well as N39 cells, constitutively expressing mSIM1-Flag only (**Figure 4.10.B-D.**). The N4 and N7 stable cell lines utilised for these experiments were doxycycline-inducible cell lines that displayed significant "leaky" expression of mSIM1-Flag in the untreated state (**Figure 4.10.B-C.**). We therefore performed our analyses on untreated N4 or N7 6-TR/mSIM1-Flag cells relative to parent N4 or N7 6-TR cells.

We detected a slight decrease in *Fat4* transcript levels, and a 3 fold increase in *Sim2* transcript levels, in response to mSIM1-Flag expression in N4 cells via QPCR (**Figure 4.12.A-B.**). The increase in *Sim2* transcript levels was not statistically significant but we felt it warranted further investigation. Indeed, we were able to detect increased *Sim2* transcript levels in an additional, independently derived mSIM1-Flag expressing N4 cell line (**Figure 4.13.A.**). The slight decrease in *Fat4* transcript levels in response to mSIM1-Flag overexpression could not be similarly confirmed in an additional, independently

Figure 4.12. Analysis of putative downstream SIM1/ARNT2 target genes in hypothalamus derived N4, N7 and N39 cells overexpressing SIM1. A-B. N4 cells stably expressing a TetR expression vector (-) or a TetR expression vector and a mSIM1-Flag expression vector (+) were harvested for RNA. cDNA was subsequently synthesised and subjected to QPCR analysis using gene-specific primers. Results are the average of three experiments performed in triplicate +SEM normalised to the reference gene *Polr2a*. Transcript expression for each gene is depicted as a fold change relative to the 6-TR cell line, which has been normalised to 1. **C-D.** N7 cells stably expressing a TetR expression vector (-) or a TetR expression vector and a mSIM1-Flag expression vector (+) were harvested for RNA. cDNA was subsequently synthesised and subjected to QPCR analysis using gene-specific primers. Results are the average of three experiments performed in triplicate +SEM normalised to the reference gene *Polr2a*. Transcript expression for each gene is depicted as a fold change relative to the 6-TR cell line, which has been normalised to 1. **E-F.** N39 cells stably expressing empty vector (-) or a mSIM1-Flag expression vector (+) were harvested for RNA. cDNA was subsequently synthesised and subjected to QPCR analysis using gene-specific primers. Results are the average of three experiments performed in triplicate +SEM normalised to the reference gene *Polr2a*. Transcript expression for each gene is depicted as a fold change relative to the empty vector cell line, which has been normalised to 1. Genes not expressed in each cell line are left blank. P values derived from two-tailed paired t test indicating likelihood of differential expression relative to empty vector cells: *** <0.001, ** <0.01, * <0.05, ns not significant.

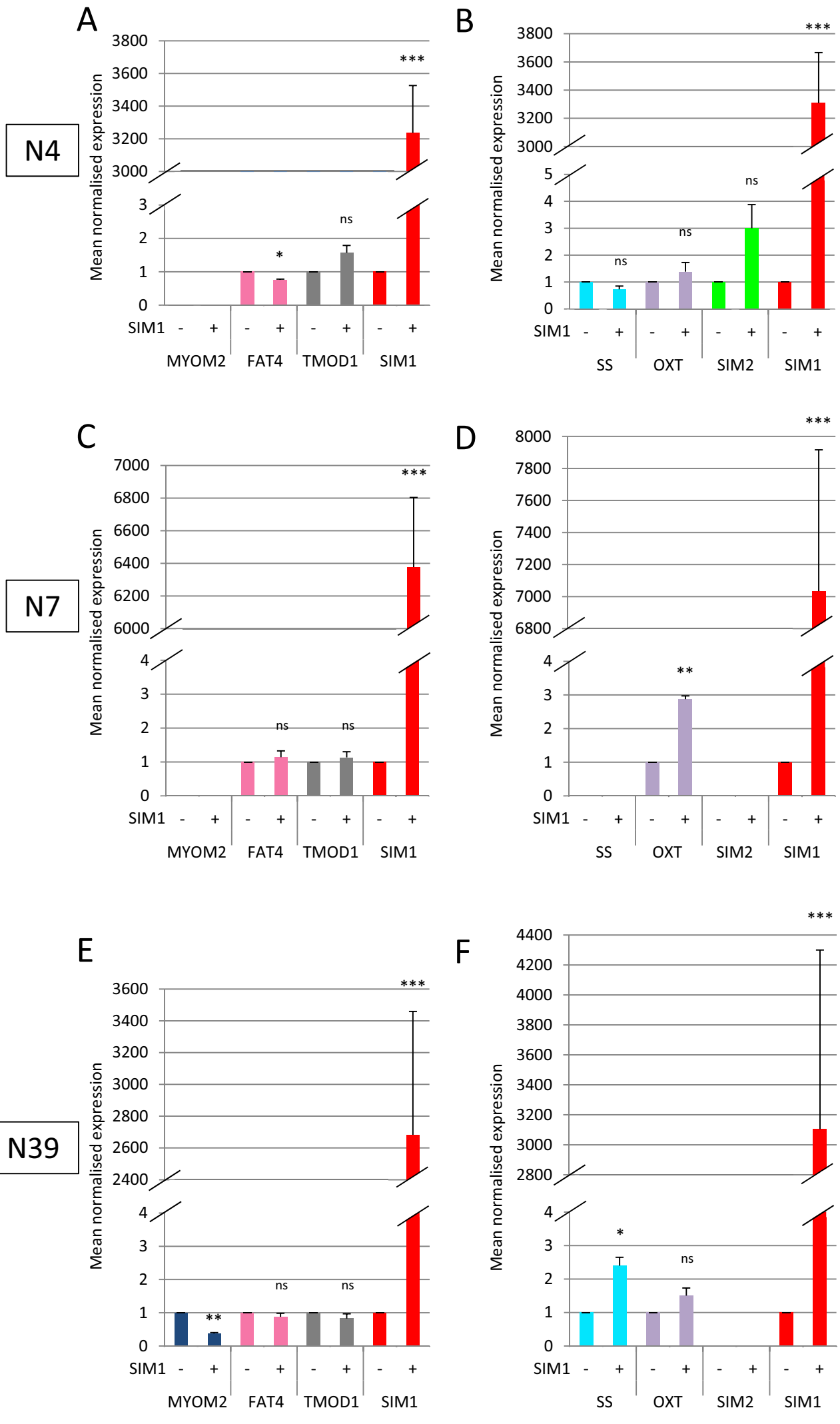
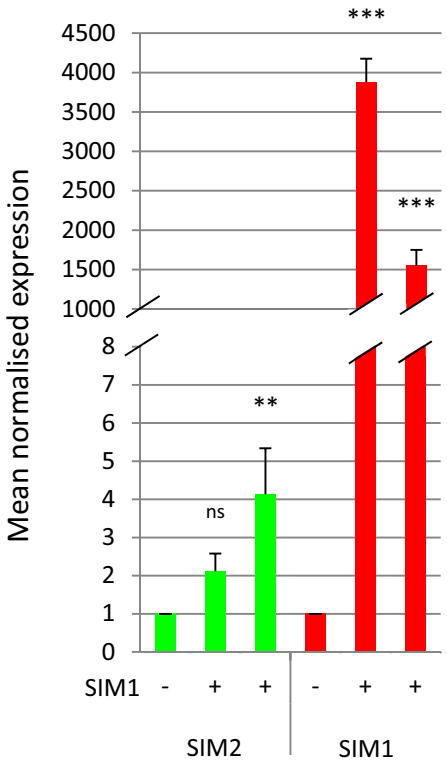
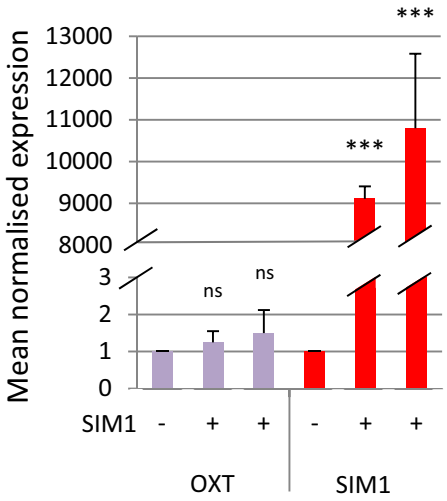

A
B
C
D
E
F
N4
N7
N39
Mean normalised expression
SIM1
MYOM2
FAT4
TMOD1
SIM1
SS
OXT
SIM2
SIM1
***
**
*
ns

Figure 4.13. Verification of putative downstream SIM1/ARNT2 target genes in independently derived SIM1-overexpressing or *Sim1* knockdown stable cell lines. A-B. One N4 cell line (**A**) or N7 cell line (**B**) stably expressing a TetR expression vector (-), or two independently derived cell lines stably expressing a TetR expression vector and a mSIM1-Flag expression vector (+) were harvested for RNA. cDNA was subsequently synthesised and subjected to QPCR analysis using gene-specific primers. Results are the average of three experiments performed in triplicate +SEM normalised to the reference gene *Polr2a*. Transcript expression for each gene is depicted as a fold change relative to the 6-TR cell line, which has been normalised to 1. **C.** One N39 cell line stably expressing empty vector (-) or two independently derived N39 cell lines stably expressing a mSIM1-Flag expression vector (+) were harvested for RNA. cDNA was subsequently synthesised and subjected to QPCR analysis using gene-specific primers. Results are the average of three experiments performed in triplicate +SEM normalised to the reference gene *Polr2a*. Transcript expression for each gene is depicted as a fold change relative to the empty vector cell line, which has been normalised to 1. **D.** N39 stable cell lines inducibly overexpressing a control shRNA sequence (scrambled) or a *SIM1* shRNA sequence (si1200) were treated with (+) or without (-) doxycycline to a final concentration of 5µg/mL for 48hr before harvesting for RNA. cDNA was subsequently synthesised and subjected to QPCR analysis using gene-specific primers. Results are the average of three experiments performed in triplicate +SEM normalised to the reference gene *Polr2a*. Transcript expression for each gene in each cell line (scrambled or si1200) is depicted as a fold change relative to untreated cells, which has been normalised to 1. P values derived from univariate ANOVA (**A-C**) or two-tailed paired t test (**D**) indicating likelihood of differential expression relative empty vector cells: *** <0.001, ** <0.01, * <0.05, ns not significant.

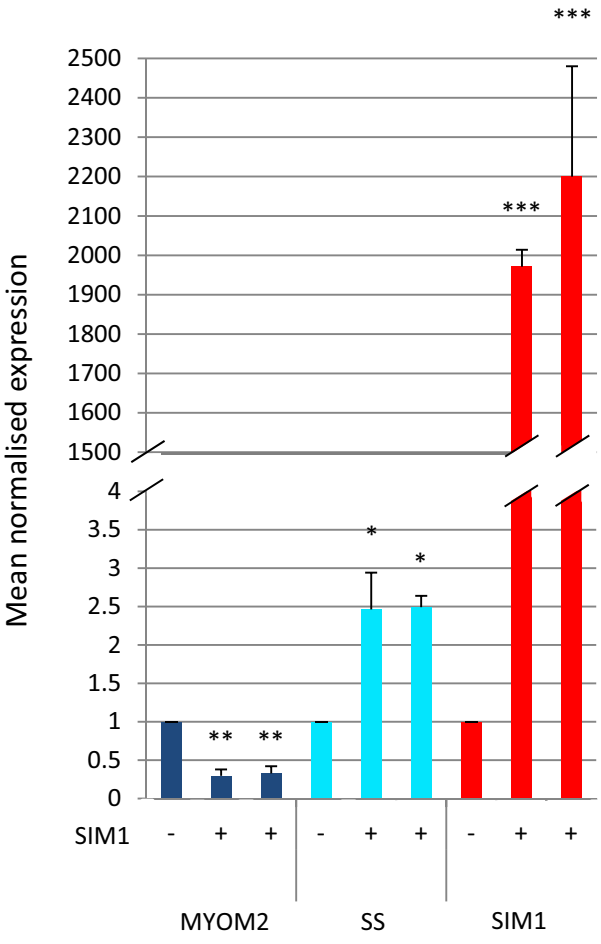
A N4



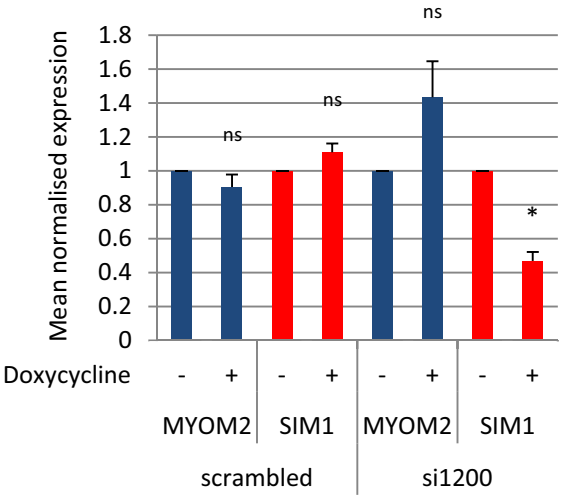
B N7



C N39



D N39



derived mSIM1-Flag expressing N4 cell line (data not shown). *Oxt* transcript levels appeared to respond to mSIM1-Flag expression in N7 cells, but we were also unable to replicate this result in an additional, independently derived mSIM1-Flag expressing N7 cell line (**Figures 4.12.D.** and **4.13.B.**). No further changes in candidate target gene transcript levels were observed in mSIM1-Flag-expressing N7 cells (**Figure 4.12.C-D**). Finally, we detected a reproducible 3 fold decrease in *Myom2* expression, and a reproducible 2.5 fold increase in *Ss* expression, in N39 cells expressing mSIM1-Flag relative to empty vector cells. Both of these results could also be replicated in an independently derived mSIM1-Flag-expressing N39 cell line (**Figures 4.12.E-F.** and **4.13.C.**). This result was surprising given that we observed no change for either of these genes in N39 cells expressing both mSIM1-Flag and mARNT2 (**Figure 4.11.**). This in turn suggests that ARNT2 is dispensable for SIM1-mediated regulation of these two genes in N39 cells, and may in fact have a stimulatory effect on the regulation of *Myom2* expression by SIM1 and an inhibitory effect on the regulation of *Ss* expression by SIM1 in this context. Both events could potentially occur via inhibition of SIM1/ARNT regulation of these genes. These results also indicate that SIM1 is capable of influencing both activation and repression of different genes within the same cell type, which may be indicative of divergent transcriptional regulatory properties upon its downstream target genes.

4.5.2.5. Analysis of *Myom2* transcript levels upon *Sim1* knockdown in N39 cells

Finally, we assessed the ability of *Myom2* to respond to endogenous *Sim1* knockdown in N39 cells. QPCR analysis of *Myom2* transcript levels in a stable cell line inducibly overexpressing a *Sim1*-specific shRNA sequence showed a modest increase in *Myom2* transcript levels upon doxycycline treatment relative to a control (scrambled) cell line, but it was not statistically significant, perhaps due to insufficient *Sim1* knockdown (**Figure 4.13.D.**). We were unable to perform similar experiments to test for changes in *Ss* transcript levels in response to *Sim1* knockdown, since it is not endogenously expressed at sufficient levels in N39 cells to allow for reliable QPCR analysis (data not shown). These data indicate a trend towards regulation of *Myom2* transcript levels by endogenous SIM1 protein in N39 cells that requires further analysis.

4.6. Bioinformatic analysis of the upstream regulatory regions of the *Myom2* and *Ss* genes

Myom2 has already been characterised as a direct SIM2 target, and contains many core 5' ACGTG 3' sequences within 10kb upstream of the initiator ATG, some of which are conserved across species ([84] and **Figure 4.14.B.**). We also searched for conserved 5' ACGTG 3' core sequences within 10kb upstream of the transcription start site of the *Ss* gene in human, mouse and rat. We identified a single 5' ACGTG 3' sequence in this region, which is conserved across species (**Figure 4.14.A.**). It is possible, therefore, that SIM1 controls *Ss* gene expression via this sequence, although its considerable distance from the immediate proximal promoter region suggests that primary control of *Ss* expression is mediated by other factors. There is, however, a precedent for SIM proteins being able to control the expression of downstream target genes via a single binding site, suggesting that direct control of *Ss* gene expression may be influenced by SIM1 *in vivo* [88, 168].

In summary, we have attempted to uncover novel downstream SIM1 target genes by performing microarray analysis on HEK-derived 293 T-Rex cells inducibly overexpressing both SIM1 and its obligate partner factor ARNT2. Western blot and RT-PCR analyses indicated robust induction of both genes in this cell line in response to doxycycline treatment. However, we also detected substantial basal expression of *Sim1* and *Arnt2* in the absence of doxycycline, indicating incomplete repression of *Sim1* and *Arnt2* gene expression in the untreated state. We therefore decided to perform microarray analysis on doxycycline-treated 293 T-Rex mSIM1-Myc/IRES/mARNT2 cells and untreated 293 T-Rex parent cells. 237 significantly upregulated genes were subsequently identified, ten of which were tested and verified for their response to mSIM1-Myc and mARNT2 expression in subsequent experiments via QPCR analysis. However, whilst these ten genes displayed robust induction upon overexpression of mSIM1-Myc and mARNT2 in independent experiments, none was markedly and specifically reduced upon *SIM1* knockdown in 293T cells. We therefore decided to investigate SIM1-mediated regulation of these same candidate target genes in a newly characterised embryonic hypothalamus derived cell culture model. Simultaneously, we examined the response of several other candidate hypothalamic target genes, including *Sim2*, *Oxt* and *Ss*, to SIM1 overexpression and knockdown in these cells. RT-PCR analysis indicated that each of the

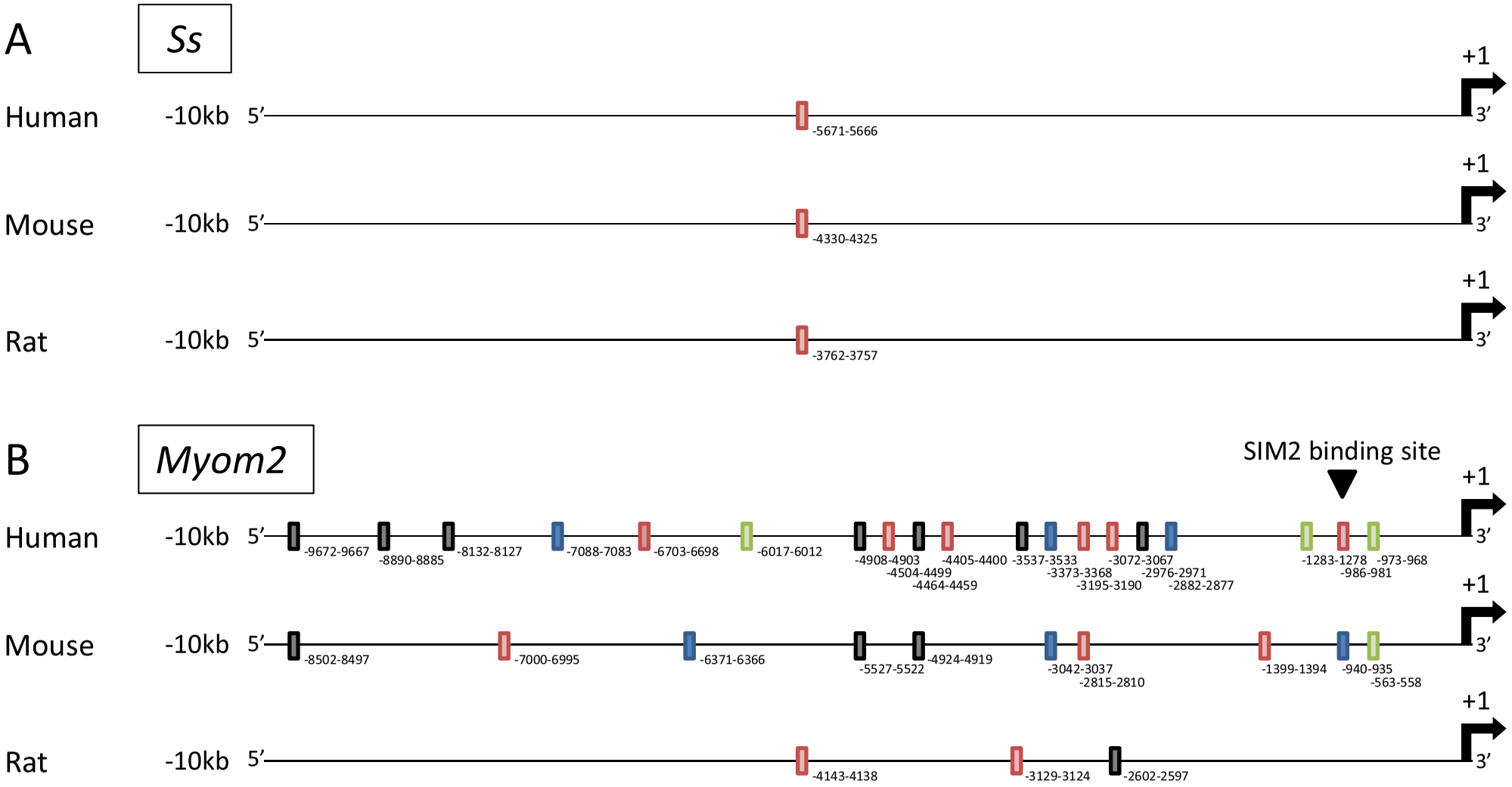


Figure 4.14. Schematic representation of consensus core CME sequences (containing a 5' ACGTG 3' core) in the upstream regulatory regions of the *Ss* and *Myom2* genes, as identified via bioinformatic analysis using TESS. Numbers refer to the position of each core sequence relative to the initiator ATG, which is labelled +1.

three cell lines we utilised for these experiments (N4, N7 and N39) were positive for both *Nestin* and *Gfap* expression, indicating that they most likely represent a type of undifferentiated hypothalamic neuronal precursor cell. Only one of the three cell lines, N39, endogenously expresses *Sim1*, although all three were positive for *Arnt* and *Arnt2* expression. We therefore performed preliminary QPCR analysis of candidate target gene expression in N39 cells stably expressing mSIM1-Flag and mARNT2, but were unable to detect reproducible changes in transcript levels for any of the genes we tested. Further analysis, however, revealed a reproducible and statistically significant decrease in *Myom2* transcript levels, and a reproducible and statistically significant increase in *Ss* transcript levels, in N39 cells overexpressing mSIM1-Flag relative to empty vector cells. This indicated that these two genes may indeed be regulated by SIM1 in these cells in an ARNT2-independent manner. We also detected an increase in *Sim2* transcript levels in N4 cells overexpressing mSIM1-Flag relative to empty vector cells, indicating that *Sim2* is able to respond to changes in SIM1 levels in a cell type-specific manner. Finally, knockdown of endogenous *Sim1* in N39 cells caused a slight increase in *Myom2* transcript levels that was not statistically significant. Bioinformatic analysis of the upstream regulatory regions of the *Myom2* and *Ss* genes indicates the presence of core 5' A<u>CGTG</u> 3' sequences, some of which are conserved across species, which may be important for SIM1-mediated regulation of these genes. *Myom2*, *Ss* and possibly *Sim2*, therefore, remain promising, probably tissue-specific SIM1 candidate target genes worthy of further investigation. Several other genes we identified by microarray analysis (*ACTBL2*, *FAT4*, *BMP3*, *ODZ1*, *NEFL*, *TMOD1*, *PCDH20*, *PPFIA2* and *CHRM3*) also warrant further investigation as tissue-specific SIM1 target genes, potentially in the development and/or function of organs such as the kidney.

CHAPTER 5: DISCUSSION

SIM1 is a transcription factor essential for survival and homeostasis in mice [58, 59, 96, 98]. Homozygous deletion of the *Sim1* gene results in perinatal lethality, presumably due to impaired development of critical neuroendocrine secretory cells of the hypothalamus, and heterozygous deletion confers early onset, hyperphagic obesity due to severely compromised function of these same cells [58, 59, 96]. Loss of a single *SIM1* allele has also been correlated with severe hyperphagic obesity in humans [177, 179, 181-183]. The past few years have seen an increase in our knowledge of the molecular properties of the SIM1 protein and how perturbation of its expression and behaviour influences development and disease. We now know that it is a tissue-specific transcription factor that responds to the MC4R agonist αMSH in order to influence the propagation of intrahypothalamic signals in response to increased food intake in the PVN/SON/aPV [76, 98, 102, 153]. SIM1 may also have additional, uncharacterised roles in peripheral tissues. Detailed characterisation of the biochemical properties and downstream effector molecules of this protein thus have important implications for human health and disease. The studies presented in this thesis endeavoured to substantiate a link between altered SIM1 function and severe obesity phenotypes in humans via a range of functional assays designed to detect any alteration in SIM1 expression, subcellular localisation, or activity in response to potentially deleterious amino acid substitutions. We also endeavoured to dissect the downstream target genes of SIM1 by performing microarray analyses on stable cell lines inducibly expressing SIM1 and its obligate partner factor ARNT2.

SIM1 as novel monogenic obesity gene.

The rising prevalence of obesity in the general population has intensified efforts to understand the molecular processes that underpin this complex disease. It is becoming increasingly evident that the trend towards obesity is a highly complex phenomenon, clearly influenced by environment, with an underlying genetic basis that can be contributed by functional alterations in a diverse number of genes [187, 189]. Current efforts to characterise these genes have focused on the identification of mutations or deletions that are sufficient to effect obesity in an autonomous manner, *i.e.* obese phentoypes associated with defects in a single gene [191, 192]. The phenotypic

observations in *Sim1*$^{+/-}$ mice, and a small number of documented cases that link severe obesity to haploinsufficiency of the *SIM1* gene in humans, has prompted efforts to consolidate the link between reduced SIM1 activity and obesity [59, 96, 177, 181-183, 219-223]. To this end, we participated in two collaborative projects that endeavoured to isolate and characterise *SIM1* variants associated with severe obesity phenotypes in humans, and performed numerous functional assays to compare the activity and behaviour of these variants to WT SIM1.

Identification of *SIM1* sequence variants statistically associated with obesity via human genetic studies.

Efforts to identify potentially deleterious variations in the *SIM1* cds by our collaborators in the UK and France identified a total of sixteen different point mutations that were unique to severely obese subjects relative to controls (**Figure 3.1.**). An additional four were identified in both obese and control subjects (**Figure 3.1.**). The twenty variants are located along the entire length of the SIM1 sequence, and subsequently reside within numerous functional domains, including those that confer DNA binding and dimerisation ability upon bHLH/PAS proteins (**Figures 3.3.** and **3.4.** and **Chapter 1**). Due to a lack of detailed functional analysis of the SIM1 protein those variants located beyond aa 336 cannot be ascribed to any specific functional domain, although the C termini of bHLH/PAS proteins are generally thought to possess transactivating properties as well as potential protein-protein interaction domains [6, 7, 9, 10, 13, 21]. We therefore concluded that disruption of any one of these functions has the potential to impact on SIM1 activity. Only two of the twenty variants (I128T and Q152E) are shared between the UK and French studies, which is probably reflective of both the relative infrequency of most variants within each cohort – the majority were identified in a single patient only – as well as the differing criteria employed to identify them (**Section 3.1.** and **Figure 3.1.**). The fact that these two were also identified in control subjects further implies that they may be low frequency SNPs, which may have the potential to manifest in obesity only in combination with additional genetic or environmental factors and could therefore occur in both obese and non-obese subjects. We were primarily interested in the effect of these amino acid substitutions on the molecular properties of the SIM1 protein, and subsequently performed numerous assays designed to identify any changes in SIM1 behaviour and activity for each variant in cell based and *in vitro* assays.

The HEK 293 T-Rex cell system is a robust and highly reproducible system within which to study SIM1 expression, activity and behaviour.

First, we sought to establish the extent to which the twenty identified SIM1 sequence variants contributed to altered SIM1 expression and/or activity. To this end, we adopted the HEK 293 T-Rex system for functional analysis of each *SIM1* variant in Western blot and luciferase reporter gene assays (see **Chapter 3**). While this cell line did not provide a hypothalamic cell context within which to study each variant, preliminary experiments indicated that we required some temporal control over the degree of SIM1 expression in our assays that we could not achieve via transient transfection studies (data not shown). Given that protein expression in the 293 T-Rex system occurs from a single genomic locus, this experimental approach enabled us to achieve highly comparable levels of SIM1 expression between multiple independently generated cell lines for short periods using doxycycline, thus allowing for comparisons of the intrinsic activity of each variant with WT SIM1 in subsequent reporter assays (**Figures 3.5.-3.8.**).

Of the nineteen variants analysed in luciferase reporter gene assays, eleven showed statistically significant differences in activity relative to WT SIM1 in the presence of ARNT, and fifteen showed statistically significant differences in activity relative to WT SIM1 in the presence of ARNT2 (**Figures 3.7.** and **3.8.**). Four variants (T361I, S541L, R703Q and R581G) showed no statistically significant differences in activity in the presence of either ARNT or ARNT2, indicating that they may represent SIM1 sequence variants of low or no pathogenicity. This in itself may be reflective of a need for additional, unknown environmental or genetic factors to influence the activity of these variants that we did not account for in our assay. Activity for most variants ranged from approximately 40-80% of WT, with two displaying activities from 20-40% (**Figures 3.7.** and **3.8.**). Given that deletion of a single *Sim1* allele (corresponding to a probable 50% decrease in SIM1 activity) is sufficient to confer severe obesity in mice, reductions of this magnitude may indeed be sufficient to impact on weight gain in humans. Interestingly, most variants did not distinguish between ARNT and ARNT2 with regards to their percentage activity relative to WT SIM1 (**Figures 3.7.** and **3.8.**). This is despite a three-fold greater induction for SIM1/ARNT on the 6xCME reporter construct relative to SIM1/ARNT2 (**Figure 3.7.** and **3.8., inset**). This may simply be reflective of the fact that

the regulatory sequences driving transcription from the 6xCME construct are derived from a *Drosophila* SIM/Tango target gene, and therefore have a greater affinity for SIM1/ARNT than SIM1/ARNT2 [13]. Alternatively, ARNT2 may have a less potent transactivation domain than ARNT, and could therefore provide a means to modulate the expression of SIM1 target genes *in vivo* via selective partner choice.

Characterisation of SIM1 T292A as a novel loss-of-function mutation.

Our reporter gene assay studies highlighted SIM1 T292A as one of the most severely compromised variants (**Figure 3.7.**). Further investigation showed that this loss-of-function could be attributed to a combination of altered subcellular localisation as well as reduced affinity for ARNT2 (**Figures 3.9.** and **3.10.**). It is possible that the T→A substitution at this position disrupts a dimerisation interface, since homology mapping places aa 292 in close proximity to residues involved in dimer formation in other PAS domain containing proteins ([22] and **Figure 3.4.**). This hypothesis assumes that nuclear retention of SIM1 is dependent on its ability to form a stable complex with ARNT or ARNT2, disruption of which would result in SIM1 translocation to the cytoplasm. However, ectopic WT SIM1 is exclusively localised to the nucleus even when it is almost certainly in excess of endogenous ARNT/ARNT2 ([75] and **Figure 3.9.**). This implies that SIM1 itself must possess an NLS, the function of which has somehow been compromised due to the T→A substitution at aa 292.

There has been only one reported study of a putative NLS in the SIM1 protein, which was pinpointed to aa 368-388 based on the exclusively nuclear accumulation observed for this short peptide when transfected into 293T cells [234]. Interestingly, although the T292A variant does not reside within this region, the R383G variant does. We did not, however, test the R383G variant for altered subcellular localisation since it did not display any statistically significant reduction in activity in our luciferase reporter gene assays (**Figure 3.7.**). It is possible that the T→A substitution at aa 292 has reduced the efficiency of a critical NLS, or alternatively, created or enhanced the function of a novel nuclear export signal (NES). Since NLSs are typically short sequences rich in the basic amino acids lysine and arginine, the effect of the T292A substitution may be indirect, either via loss of a critical phosphorylation site required for efficient nuclear localisation, or replacement of

a hydrophilic threonine residue with a hydrophobic alanine residue that masks or buries an atypical NLS [273-276]. Alternatively, since NES sequences are typically rich in hydrophobic residues, the addition of an alanine at this position may assist in creating a novel, relatively weak NES. This NES could compete with the NLS for trafficking of the SIM1 protein, the efficiency of which would be dependent on the relative affinities of nuclear import/export receptors for each sequence. The two possibilities (compromised nuclear import or enhanced nuclear export) could be easily distinguished via treatment with leptomycin B (LMB). LMB inhibits active nuclear export by blocking the interaction of NES sequences with the nuclear export receptor protein CRM1 [277]. If active CRM1-mediated export is necessary for the partially cytoplasmic localisation of SIM1 T292A, treatment with LMB would result in exclusively nuclear accumulation. No change in subcellular localisation upon LMB treatment would instead indicate that SIM1 T292A is retained in the cytoplasm due to compromised NLS function.

^{263}KTLYHHVHGCDTFHLRCAHHLLLVKGQVT[T]KYYRFLAKHGGWVWVQSYATIVHNSRSSRPH^{323}

Sequence of the hSIM1 protein from aa 263-323. Basic amino acids that may be involved in NLS function are indicated in **green**. Hydrophobic amino acids that may be involved in NES function are indicated in **red**. Amino acid 292 is boxed.

Unexpectedly, five other variants within the bHLH/PAS region that displayed substantially reduced activity in our reporter assays displayed no difference in dimerisation potential relative to WT SIM1 (**Figures 3.3., 3.7., 3.10.** and data not shown). It is possible that DNA binding is compromised for the remaining variants that lie within this region. Data from our laboratory indicates that, along with the bHLH domain, the PAS domain itself forms direct contacts with DNA, thus contributing to the binding affinity of bHLH/PAS heterodimers (A. Chapman-Smith, unpublished observation). We spent considerable time attempting to optimise an *in vitro* DNA binding assay to address this possibility, without success (**Figures 3.11.-3.16.**). Taking into account previously published, only moderately successful attempts to observe SIM1-specific shifts via EMSA, it appears that SIM1 may be a genuinely weak DNA binder *in vitro* under standard EMSA conditions [60]. This in turn implies that additional or alternative reaction conditions should be utilised in further attempts to optimise an EMSA protocol for full-length SIM1. Such conditions could include the addition of RNA species to the binding reaction, since recent work indicates a role for long non-coding RNA molecules in DNA-

bound gene regulatory complexes [278-280]. Alternatively, development of a ChIP protocol that enriches for DNA-bound SIM1 complexes, either via immunoprecipitation of ectopic SIM1 from the 6xCME reporter construct or native SIM1 from the regulatory regions of its endogenous target genes, may prove to be a more relevant means of studying SIM1 DNA binding affinity in a cellular context. The latter approach awaits the development of antibodies capable of immunoprecipitating endogenous SIM1.

Identification of transcripts that are differentially regulated in response to SIM1 and ARNT2 expression via microarray.

Years of research have thus far provided only limited insight into the downstream target genes of SIM1, in particular those involved in mediation of adiposity signals in the hypothalamus in response to feeding [57, 137]. This can be attributed, at least in part, to the highly selective expression pattern for SIM1 in the hypothalamus, which makes relevant cell-based analyses of its transcriptional activities particularly challenging [58]. With the tools available to us at the commencement of this study we decided to investigate downstream SIM1 target genes in HEK cells, in the hope that our ability to perform both SIM1 overexpression and knockdown studies in these cells would uncover "generic" SIM1 target genes that were also relevant to SIM1 function in the other tissues in which it is expressed, including the hypothalamus. To this end, we generated 293 T-Rex stable cell lines inducibly expressing Myc-tagged mSIM1 and mARNT2, and characterised their suitability for microarray studies by Western analysis and RT-PCR.

Application of the HEK 293 T-Rex cell system to study SIM1-mediated changes in downstream target gene expression.

Having previously established the suitability of the 293 T-Rex system for analysis of protein expression and activity between multiple independently generated cell lines, we decided to adopt this cell line for our microarray analyses. However, we detected a substantial degree of basal transcription in our mSIM1-Myc/mARNT2-overexpressing stable cell lines in the absence of doxycycline, which prompted us to perform microarray analysis on doxycycline treated 293 T-Rex mSIM1-Myc/IRES/mARNT2 cells relative to the parent (untransformed) 293 T-Rex cell line (**Figures 4.1.A.ii** and **4.2.**). Using this method, we identified several candidate downstream target genes whose transcript levels

were altered in response to mSIM1-Myc and mARNT2 overexpression (**Figures 4.3.** and **4.4.**).

Extensive QPCR analyses in additional 293 T-Rex stable cell lines indicated robust and statistically significant increases in transcript levels for nine of ten candidate genes in response to mSIM1-Myc and mARNT2 induction relative to parent cells (**Figure 4.5.B.**). A single gene, *ODZ1*, showed a statistically significant increase in transcript levels in only one of two independently derived 293 T-Rex mSIM1-Myc/IRES/mARNT2 stable cell lines relative to parent cells (**Figure 4.5.B.**). However, this gene showed some evidence of SIM1-independent changes in transcript levels between cell lines, since reduced levels were observed in empty vector cells relative to parent cells. A similar phenomenon was also observed for *NEFL*, *TMOD1*, *PCDH20*, *PPFIA2* and *CHRM3* (**Figure 4.5.B.**). This suggested that the transfection and antibiotic selection process for generating stable 293 T-Rex cell lines had impacted on the expression of these genes, and subsequently made the empty vector cell line a more suitable control cell line with which to compare candidate target gene transcript levels in 293 T-Rex mSIM1-Myc/IRES/mARNT2 cells. When normalised to empty vector cells, transcript levels for all ten genes were significantly increased in both 293 T-Rex mSIM1-Myc/IRES/mARNT2 stable cell lines (**Figure 4.5.B.** and data not shown). Furthermore, several genes (*ODZ1*, *NEFL*, *TMOD1*, *PPFIA2* and *CHRM3*) consistently displayed moderate changes in transcript levels in response to mARNT2 expression alone, suggesting that the relative levels of both SIM1 and ARNT2 may impact on the expression of these genes (**Figure 4.5.B.** and data not shown).

Transcript levels of the majority of genes do not alter in response to endogenous *SIM1* knockdown in 293T cells.

Despite the promising results obtained from our analyses in 293 T-Rex stable cell lines, subsequent QPCR analysis of each of the ten candidate genes in 293T cells engineered to express a control (scrambled) or *SIM1*-specific shRNA sequence in response to doxycycline induction did not show any *SIM1* knockdown-specific differences in transcript levels for any of the genes we tested, except perhaps *BMP3* (**Figure 4.7.**). Although *FAT4* and *PPFIA2* showed slight but statistically significant increases upon *SIM1* knockdown, these observations run contrary to their potent induction upon mSIM1-

Myc/mARNT2 overexpression in 293 T-Rex cells and are unlikely to be genuine (**Figures 4.5.B.** and **4.7.**). In the absence of an antibody capable of detecting endogenous SIM1 protein we were unable to assess the efficiency of SIM1 knockdown in this system, although QPCR analysis indicated that *SIM1* transcripts were reduced by at least 50% upon doxycycline treatment (**Figures 4.6.** and **4.7.**). This may be insufficient to achieve dramatic reductions in downstream target gene transcript levels. There is also a precedent for shared gene regulation between bHLH/PAS factors, and it is possible that SIM2 in particular may be compensating for loss of SIM1 in this cell line [75, 85, 87, 157, 158]. This is especially the case for *MYOM2*, which has already been characterised as a direct SIM2 target in 293T cells [84]. The remaining genes are largely uncharacterised in terms of other factors that control their expression, but these factors may also be able to maintain sufficient expression of each gene in the event of reduced SIM1 levels.

Analysing SIM1 activity in a hypothalamic context.

Given that one of the original aims of this study was to uncover SIM1 target genes involved in hypothalamic function, we were excited when, during the course of our analyses in HEK cells, Belsham *et al* described the development of numerous immortalised neuronal cell lines derived from E15.0-18.0 hypothalamus [224]. We were of course immediately interested in establishing whether any of the described cell lines were likely to be derived from the PVN/SON/aPV region, and therefore an appropriate means by which to study SIM1 function in hypothalamus. PVN/SON/aPV cells are born and generated between E10.0-18.0, with the peak period of neuronal differentiation and migration occurring between E15.0-18.0. Differentiation is not completed until final settling and synaptic integration from E18.0 onwards [281-283]. This fits reasonably well with expression of the neuronal precursor marker *Nestin* in the three cell lines we purchased (**Figures 4.8.** and **4.9.iii**). Additionally, RT-PCR analysis indicated that N39 cells were positive for *Sim1* (**Figure 4.9.i**). In the hypothalamus, *Sim1* is expressed exclusively in the PVN, its contiguous aPV, and SON, as well as scattered cells of the LHA in adult mice, implying that N39 cells may well be derived from one of these regions ([49, 57-59] and **Figure 1.3.**). This inference was also supported by a lack of *Rax*, *Pax6* and *Mash1* expression, since these genes are expressed in various zones surrounding the PVN/SON area ([93, 269-272, 284-286] and **Figure 4.9.ii**). However, N39 cells did not express *Mc4r*, which was unexpected given that *Sim1* and *Mc4r* are probably co-

expressed *in vivo* ([76, 145] and **Figure 4.9.i**). Ultimately, we decided that N39 cells offered a promising system within which to pursue our target gene analyses, and subsequently generated stable N39 cell lines constitutively overexpressing mSIM1-Flag and/or mARNT2 (**Figure 4.10.**).

***Myom2* and *Ss* may be novel downstream SIM1 targets in hypothalamus.**

Initial QPCR analysis of each of the ten candidate target genes we identified in the 293 T-Rex system indicated that the majority of them were not expressed in N39 cells (data not shown). This was difficult to anticipate, since detailed expression analysis for most of these genes, particularly in the hypothalamus, is not available. Those genes that were expressed endogenously (*Myom2*, *Fat4*, *Tmod1* and *Ppfia2*) did not display any SIM1-mediated changes in transcript levels relative to empty vector cells (**Figure 4.11.A.**). However, analysis of transcript levels for these same genes in N39 cells constitutively expressing only mSIM1-Flag showed a reproducible 3 fold decrease in *Myom2* transcript levels relative to empty vector cells that we were able to replicate in an additional, independently derived cell line (**Figures 4.12.E.** and **4.13.C.**). This contrasted strikingly with its potent upregulation upon mSIM1-Myc/mARNT2 induction in 293 T-Rex cells (**Figures 4.4.** and **4.5.B.**). Furthermore, we observed a reproducible 2.5 fold increase in *Ss* transcript levels relative to empty vector cells that we were also able to replicate in an additional, independently derived cell line (**Figures 4.12.F.** and **4.13.C.**). Our reasons for analysing *Ss* transcript levels are discussed in more detail below.

***Myom2* as a novel SIM1-regulated target gene.**

The phenomenon of a single transcription factor possessing both activator and repressor properties on its downstream target genes has been previously documented within the bHLH/PAS family [84, 85, 87, 287, 288]. It may be dependent on multiple factors including promoter architecture, the availability of specific coregulatory proteins, and the relative affinity of competing transcription factors for critical cofactors, resulting in target gene regulatory events that can be highly context-dependent [289-292]. Indeed, SIM2 has been shown to possess activator and repressor properties on the *MYOM2* gene in different cell types, suggesting that the same may also be true for SIM1 on this gene [84]. SIM2/ARNT activates a *MYOM2*-mediated luciferase reporter gene in 293T cells,

however, *Sim2* knockdown in human myoblast LHCN-M2 cells causes increased *MYOM2* transcript levels, indirectly inferring that SIM2 normally represses *MYOM2* in this cell line [84]. This may be a reflection of the fact that LHCN-M2 cells are an undifferentiated muscle progenitor cell type, in which differentiated muscle-specific factors such as *MYOM2* are not yet required and therefore need to be repressed. An additional level of *MYOM2* regulation may therefore be supplied in any cell or tissue type that co-expresses both *Sim1* and *Sim2*, particularly if, as suggested by **Figure 4.13.A.**, SIM1 is capable of affecting *Sim2* transcript levels in a cell type-specific manner (see below).

The method of *Myom2* repression by SIM1 in N39 cells is unclear. QPCR analysis of *Myom2* transcript levels upon shRNA-mediated *Sim1* knockdown showed a modest increase in *Myom2* transcript levels relative to a scrambled shRNA sequence, suggesting that endogenous SIM1 may act to repress *Myom2* (**Figure 4.13.D.**). It is unlikely that SIM2 is compensating for loss of SIM1 and maintaining repression of *Myom2* in this instance, since *Sim2* is not expressed endogenously in N39 cells ([84] and **Figure 4.9.i**). SIM1 could be effecting repression of *Myom2* via binding directly to the *Myom2* promoter in a repressive complex, and its success in doing so may also be reflective of a lack of essential factors, such as MEF2, that are necessary to maintain *Myom2* expression in a non-muscle context ([238] and **Figure 4.14.**). Alternatively, SIM1 may be having an indirect repressive effect via sequestration of critical factors away from other activating complexes. Interestingly, data presented in **Figure 4.11.A.** indicates that ectopic ARNT2 may in fact relieve SIM1-mediated repression of *Myom2* in N39 cells, suggesting that ARNT2 is a limiting factor in this cell line. A dynamic interplay between the relative levels of SIM1 and ARNT2 may therefore affect expression of this gene in the hypothalamus [84]. This may also be true of its role in HEK cells, or any other cell type in which multiple bHLH/PAS factors are expressed ([75] and **Figures 4.4.** and **4.5.B.**).

_Ss_ as a novel SIM1-regulated target gene.

Data published within the last three to five years has dramatically increased our understanding of SIM1 function within the adult hypothalamus in response to food intake [76, 98, 153]. In particular, its positioning along the leptin-melanocortin axis has highlighted a number of genes whose expression may be dependent on SIM1, potentially in a cell autonomous manner. These genes include *Crh*, *Trh*, *Ss*, *Avp* and *Oxt*, as well as

Sim2 and *Mc4r* [57, 63, 97, 98, 153]. Although we could not detect expression of any of these genes in N39 cells by RT-PCR analysis and ethidium bromide staining, QPCR analysis indicated that *Oxt* and *Ss* were expressed endogenously at low but sufficient levels to allow for reliable analysis by this method (**Figures 4.9.i** and **iii** and data not shown). We subsequently observed a reproducible 2.5 fold increase in *Ss* transcript levels in mSIM1-Flag overexpressing cells relative to empty vector cells that we were able to replicate in an additional, independently derived cell line (**Figures 4.12.F.** and **4.13.C.**). Once again, this phenomenon appears to be ARNT2-independent, and data presented in **Figure 4.11.B.** indicates that ectopic ARNT2 may in fact have an inhibitory effect on SIM1-mediated induction of *Ss*. *Oxt* has already been proposed to be an indirect SIM1 target, and its lack of response in N39 cells may reflect a deficiency of critical downstream factors necessary to effect its expression [153]. We also observed increased transcript levels for *Sim2* in N4 cells in response to mSIM1-Flag overexpression, as well as changes in *Fat4* and *Oxt* levels in N4 and N7 cells respectively that were not reproducible (**Figure 4.13.A., B.** and **D.**).

Ss expression, function and physiology have been well documented over a period of approximately forty years. It was initially isolated and characterised as a hypothalamic polypeptide that inhibited growth hormone release in rats, but has subsequently been proven to effect the release of numerous hormones and small molecules from several sites of expression including the anterior pituitary and the GI tract [293]. SS is also capable of regulating renal cell proliferation in an autocrine fashion [294, 295]. The SS pre-pro-peptide can be cleaved to produce three different products (SS-14, SS-28 and the more recently discovered Neuronostatin) that differ in terms of their expression patterns, relative affinities for the five SS receptors (SSTR1-5), and physiological effects [146, 293, 296, 297]. The extent of sequence overlap between SS-14 and SS-28 has made selective analysis of their expression *in vivo* difficult, although the pre-pro-peptide itself is expressed in the PVN and aPV of the hypothalamus and appears to regulate growth hormone secretion from the anterior pituitary primarily via SSTR1, SSTR2 and SSTR5 [58, 62, 93, 298-301]. Neuronostatin is co-expressed with SS-14 in the hypothalamus and has been demonstrated to have anorectic effects upon ICV injection in rats [146, 302]. The data presented in this thesis are the first to indicate that levels of *Ss* transcript (and hence SS function) may be dependent on SIM1 in a cell autonomous manner. At the physiological level, our results are supported by the increased linear growth observed for

Sim1$^{+/-}$ mice relative to WT mice, since reduced *Sim1* dosage would contribute to reduced *Ss* expression in these animals and hence attenuate the degree to which growth hormone release is inhibited by SS [59, 96, 98]. Reduced production of Neuronostatin from the PVN/aPV could also contribute to the hyperphagia displayed by *Sim1*$^{+/-}$ mice [59, 96, 98]. Interestingly, irregular growth hormone secretion patterns have been observed in severely MC4R-deficient obese patients relative to control (non-MC4R-deficient) obese patients, further substantiating the link between intact MC4R signalling – which may be mediated by SIM1 – and SS-mediated regulation of growth hormone release in humans [76, 102, 303]. *Ss* levels are also decreased in A^Y mice, which constitutively overexpress an MC4R antagonist [304].

The *Ss* promoter contains only one 5' ACGTG 3' motif in 10kb of sequence upstream of the initiator ATG, which is conserved across species (**Figure 4.14.**). This may indicate immediate control of *Ss* expression by SIM1 by binding directly to DNA, probably in concert with other transcriptional activators. Alternatively, SIM1 may regulate *Ss* expression indirectly, by controlling the availability of transcriptional regulators that impact on *Ss* expression. Surprisingly, the *Ss* promoter itself is largely uncharacterised, although SS secretion from primary cells can be modulated in response to a wide range of small molecules and secreted factors [293, 305]. There is a single cAMP response element (CRE) within 100bp of the transcription start site that has been shown to modulate *Ss* transcript production via binding of phosphorylated CRE Binding Protein (CREB) [306, 307]. This is intriguing in light of the fact that MC4R is a GPCR of the $G_{\alpha s}$ subtype, activation of which results in increased levels of cAMP [308]. The same response element is also bound and activated by CREB in response to BDNF signalling in cultured cortical neurons [309]. It would be interesting to see if *Ss* levels respond to αMSH and/or MTII treatment in WT mice, and whether this upregulation is attenuated in *Mc4r*- and *Sim1*- deficient mice. *Ss* transcript levels are known to be reduced in *Sim1*$^{+/-}$ hypothalamic tissues, and the levels of hypothalamic SS-14 and Neuronostatin in these mice would therefore also be of interest [153]. Finally, control of *Ss* expression by SIM1 on the conserved 5' ACGTG 3' motif could be tested via ChIP of ectopic SIM1 protein from N39 cells using anti-Flag antibodies, since antibodies capable of detecting endogenous SIM1 protein have yet to be developed (**Figure 4.14.**).

Further dissection of the physiological role/s and downstream target genes of SIM1.

The gene expression studies presented here have provided some of the first glimpses into the signalling pathways SIM1 may be effecting in multiple organs throughout development and in the adult. What these studies have highlighted is that these activities are likely to be cell type- or tissue-specific, perhaps via recruitment of specific cofactors necessary for SIM1 activity. This hypothesis has been frequently proposed but never investigated, and awaits the identification of SIM1 interacting proteins that may mediate tissue-specific gene regulatory events. Such studies may also provide a molecular basis to the observed lack of overlap between our gene regulatory studies in kidney- and hypothalamus-derived cells. Despite these challenges, we have identified *Myom2* and *Ss* as novel SIM1 target genes worthy of further investigation. Their relevance to SIM1 function *in vivo* awaits further analyses in mice.

A lack of any phenotypic characterisation of *Sim1*$^{-/-}$ mice outside the hypothalamus makes predictions of a role for SIM1 in peripheral tissues problematic [58]. A more comprehensive analysis of *Sim1* expression throughout embryogenesis, particularly in kidney and muscle, is therefore critical for further studies of its downstream target genes in these areas. Expressed sequence tags (ESTs) for all of the abovementioned candidates except *CHRM3* have been cloned from kidney tissue, indicating that they may well possess functions in this organ. Many of them have also been similarly cloned from muscle <http://www.ncbi.nlm.nih.gov/unigene>. On the basis of what is already known about each gene, it seems likely that SIM1 may be important for kidney formation throughout embryogenesis, although a role in kidney function in the adult is also possible. It may also be an important mediator of signalling pathways that control muscle structure and function (see above). Tissue-specific knockout of *Sim1* in mice would enable analysis of any phenotype that manifests beyond birth, and would also provide the means to compare the expression of any potential candidate downstream target genes, including those discussed above, via *in situ* hybridisation on *Sim1*$^{-/-}$ kidney or muscle tissue sections relative to WT. Dual knockout of one or both alleles of *Sim1* and *Sim2* also has the potential to highlight any shared regulatory pathways between these two genes. Interestingly, *Arnt2* is expressed more strongly in the kidney than any other organ in the body except the brain, and has been detected in embryonic kidney tissue from as early as E13.0 [83, 112]. It is also expressed in muscle from E11.0, although *Arnt* appears to be

expressed more strongly than *Arnt2* in this area [83]. Phenotypes in these two organs have not been reported for either *Arnt*$^{-/-}$ or *Arnt2*$^{-/-}$ mice, however like *Sim1*$^{-/-}$ mice they may not manifest until after birth and would therefore not be observed due to the embryonic lethality conferred by defective HIF-1α signalling in *Arnt*$^{-/-}$ mice, or perinatal lethality presumably conferred by compromised hypothalamic development in *Arnt2*$^{-/-}$ mice ([81, 94, 95, 132, 133] and **Chapter 1**).

Further dissection of a role for SIM1 in the penetrance and severity of obese phenotypes.

The data presented in this thesis represent the first functional evidence that loss of SIM1 activity is associated with severe obesity in humans. This strengthens the possibility that deleterious *SIM1* mutations may contribute to obese phenotypes in the wider population, and has important implications for future efforts to develop therapies for metabolic disorders. We are in the process of generating knock-in mice for several of the most severely affected variants via homologous recombination, in the hope that they recapitulate the obesity phenotype of *Sim1*$^{+/-}$ mice. This would provide definitive physiological evidence that these variants are sufficient to confer obesity in mice, and would also provide us with the means to dissect more thoroughly the SIM1-mediated signalling pathways that are activated in response to food intake.

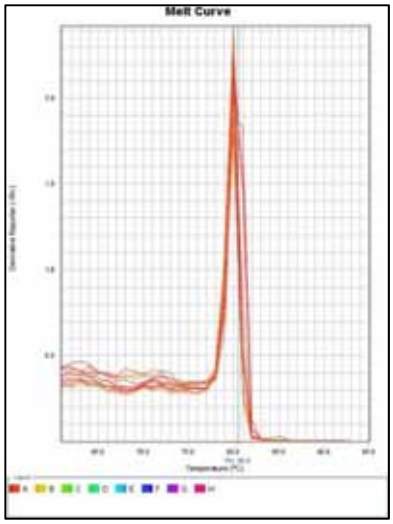
Data recently published by Tolson *et al* are the first to show that the signalling pathways SIM1 effects throughout development are distinct from those it effects in response to increased food intake in the adult [98]. This in turn suggests that informative studies into SIM1 function and activity must be tissue- and developmental stage-specific and should ideally be performed using *Sim1*-deficient mice. These could include analyses of hypothalamus histology and candidate target gene expression in *Sim1*$^{mut/mut}$ and *Sim1*$^{+/mut}$ mice relative to WT mice, both with and without αMSH and/or MTII treatment. Such analyses could include assessment of *Ss* transcript levels in the hypothalamus, as a means of dissecting more thoroughly its contribution to obesity in *Sim1*-deficient animals. Microarray or deep sequencing analysis could also be employed to uncover novel genes that are differentially regulated in response to altered SIM1 activity. Simultaneously, it would also be of interest to analyse more thoroughly the morphology and gene expression profiles of peripheral tissues in *Sim1*$^{mut/mut}$ and *Sim1*$^{+/mut}$ mice, as a means of assessing the

contribution of SIM1 in these tissues to the severe obesity observed upon global *Sim1* deficiency.

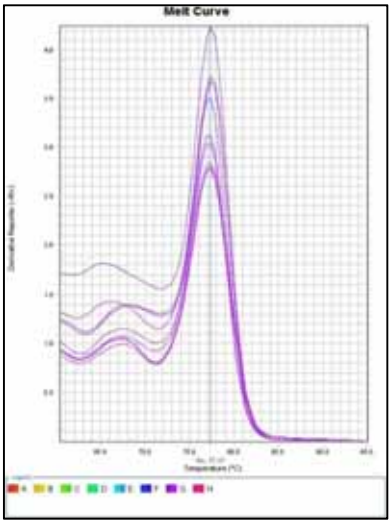
Addendum

Following completion of the studies presented in this thesis, it was brought to our attention that the Wellcome Trust Sanger Institute, Hinxton, which was responsible for sequencing and verification of the DNA samples derived from the UK GOOS and control cohorts, was unable to confirm the veracity of three of the sixteen SIM1 variants provided to us for functional analysis, namely the K51N, T292A and T361I variants. Unfortunately, this precludes their definition as naturally occurring pathogenic SIM1 sequence variants that may contribute to obesity. However, our results indicate that T292A in particular serves as a near-null SIM1 sequence variant that can be utilised in future studies as a benchmark for severe loss-of-function. Validation of the authenticity of these three variants awaits sample recovery and further re-sequencing by the Sanger Institute.

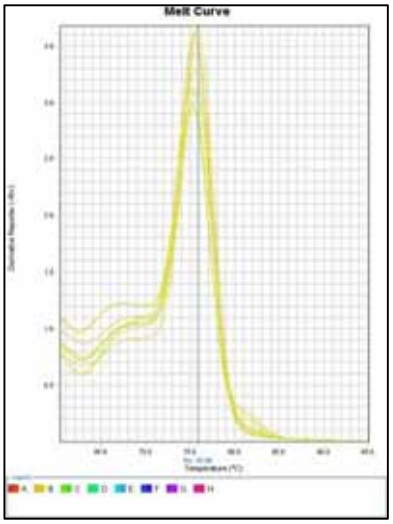
CHAPTER 6: APPENDIX



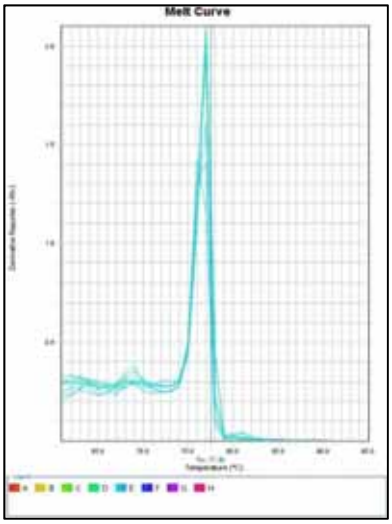
hMYOM2



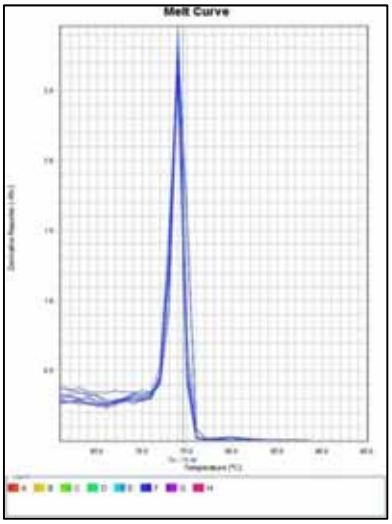
hACTBL2



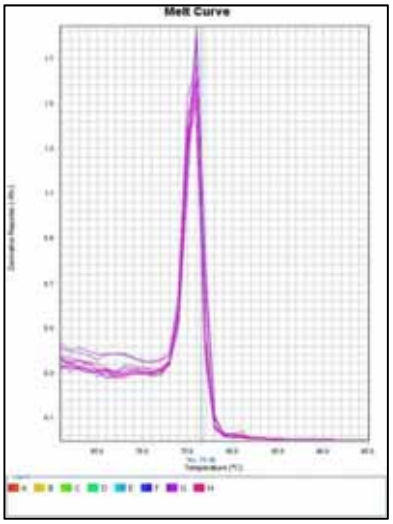
hFAT4



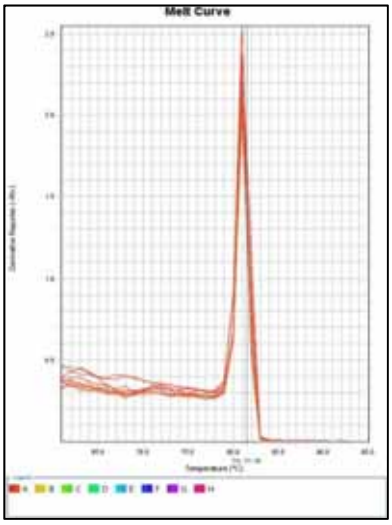
hBMP3



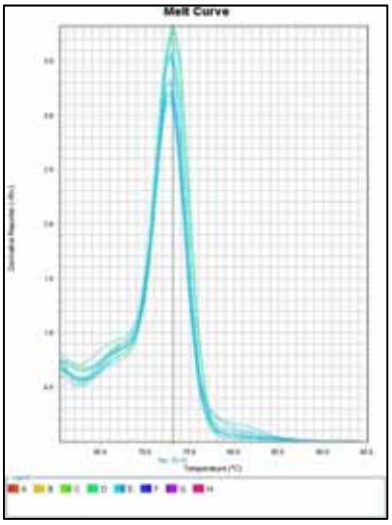
hODZ1



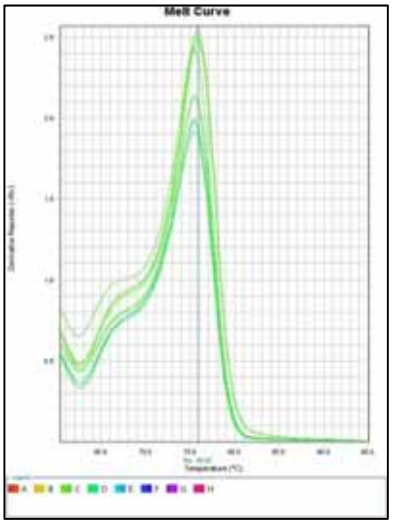
hNEFL



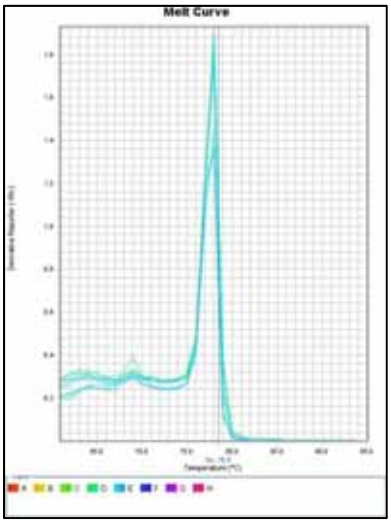
hTMOD1



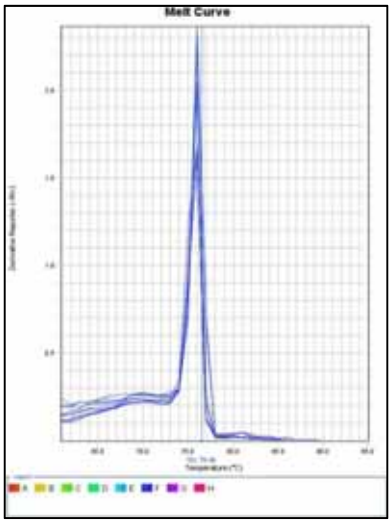
hPCDH20



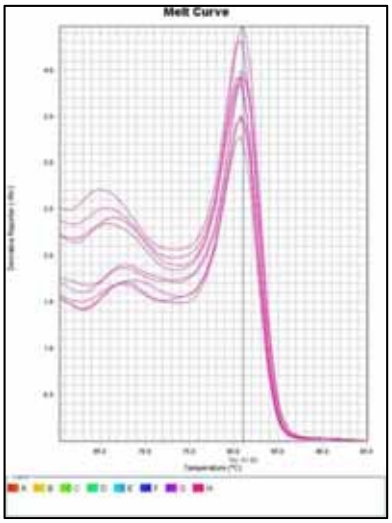
hDUB3



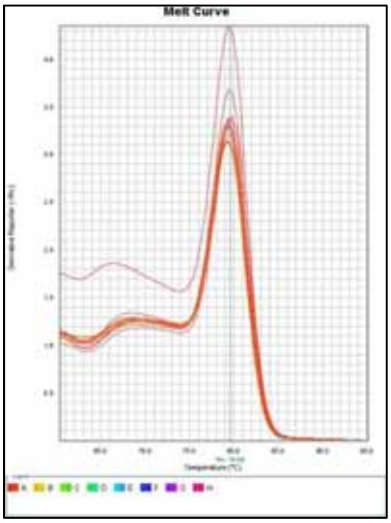
hPPFIA2



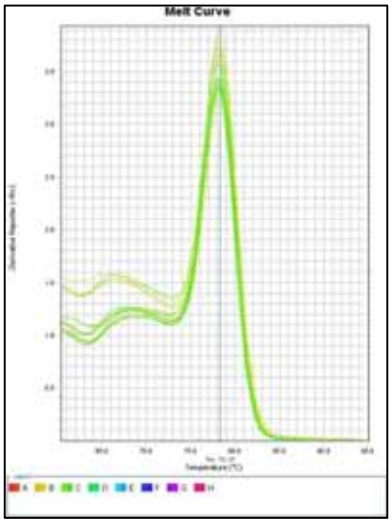
hCHRM3



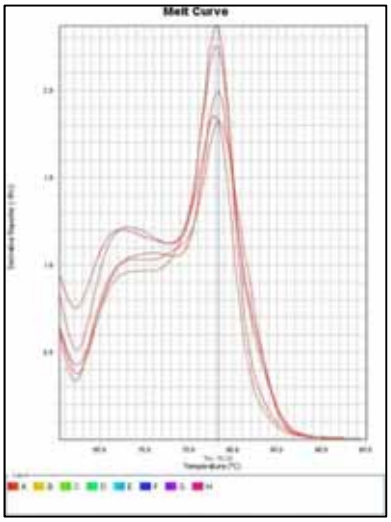
ARNT2



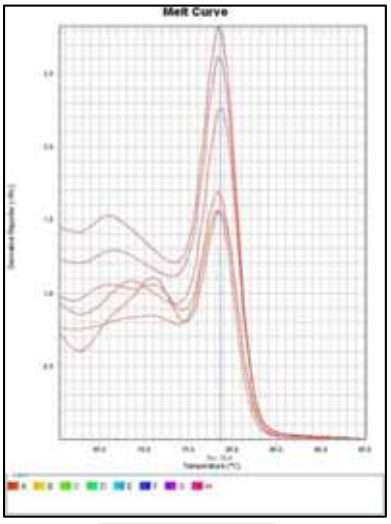
hSIM1
(587F, 627R)



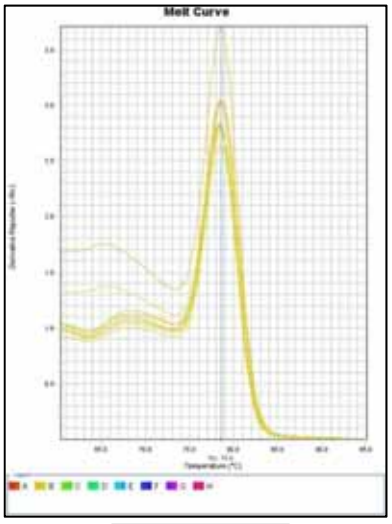
hPOLR2A



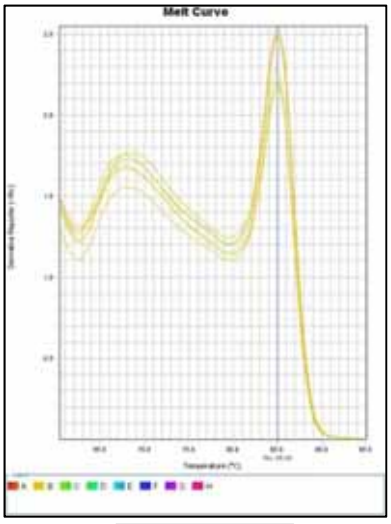
mMYOM2



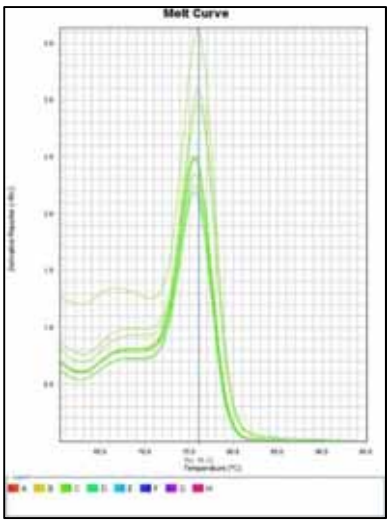
mACTBL2



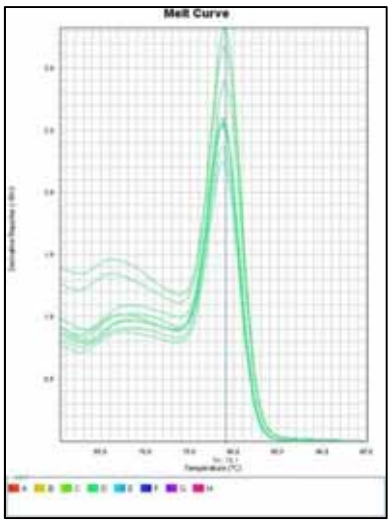
mFAT4



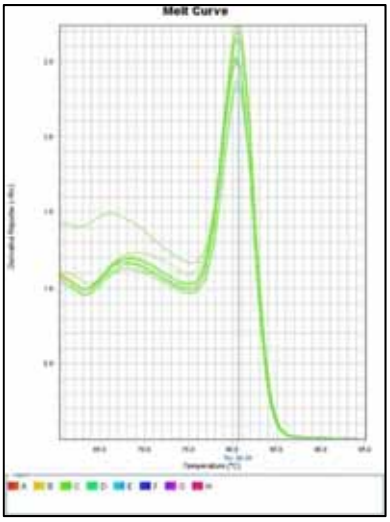
mBMP3



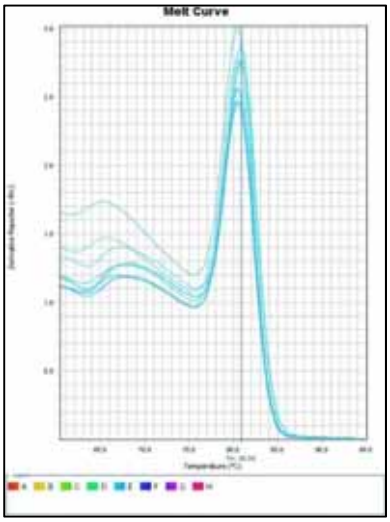
mODZ1



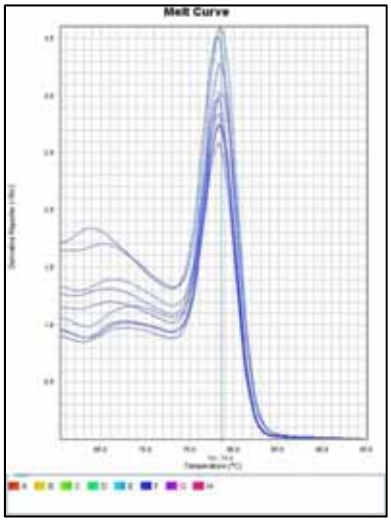
mNEFL



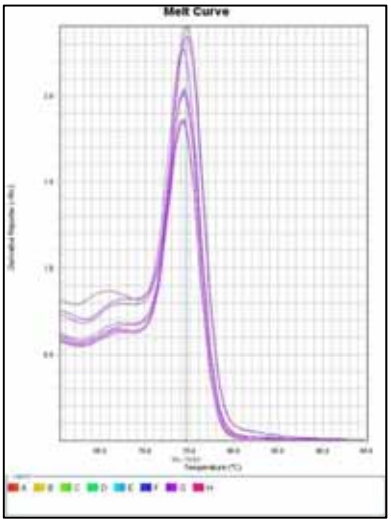
mTMOD1



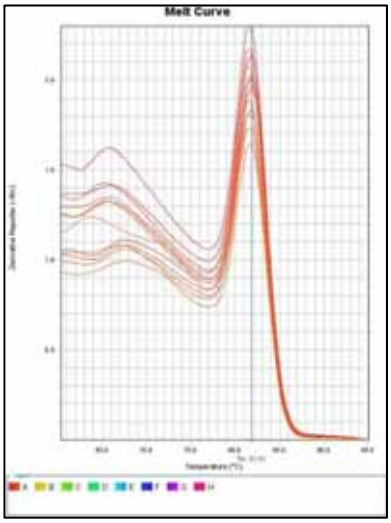
mPCDH20



mPPFIA2



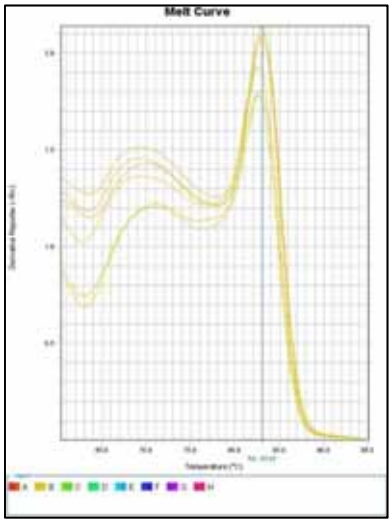
mCHRM3



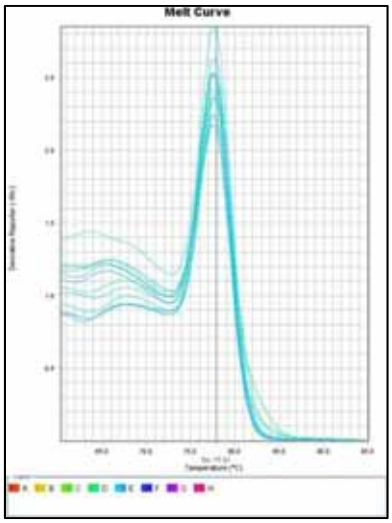
mOXT

Melt curve for this primer pair unable to be generated due to very low transcript expression.

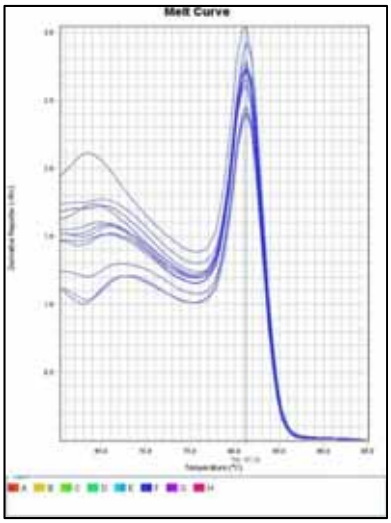
mAVP



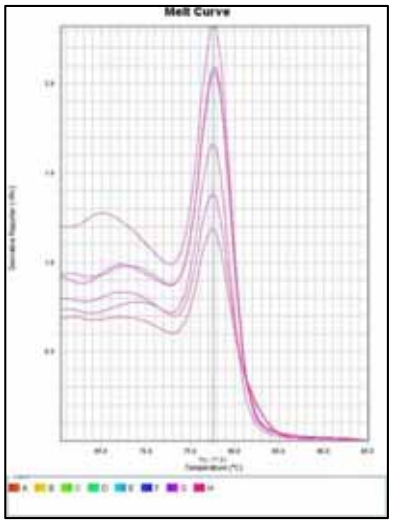
mCRH



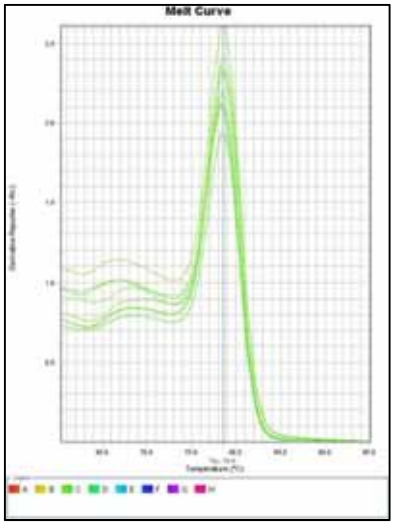
mTRH



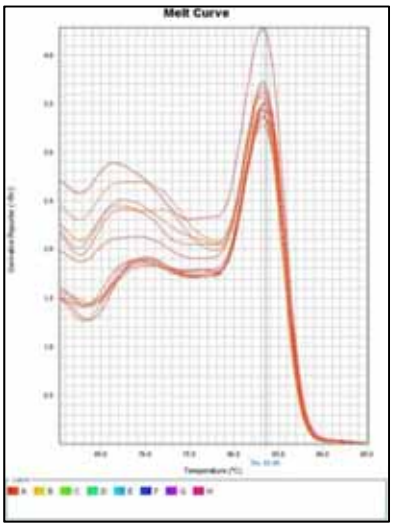
mSS



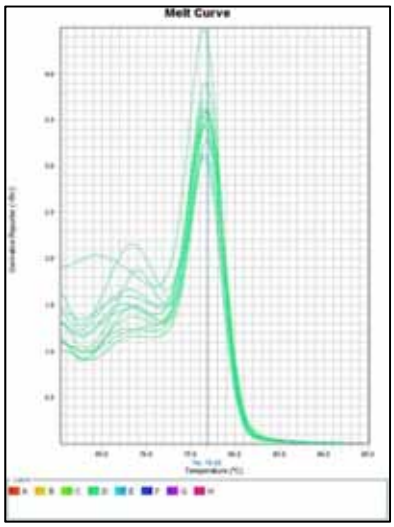
mSIM2



mMC4R



mSIM1
(513F, 551R)



mrPOLR2A

Appendix. Representative melt curves for QPCR primer pairs. Melt curve analysis was typically performed on triplicate serial 1:5 dilutions of 293T cDNA (for human primer pairs) or mouse brain cDNA (for mouse primer pairs). Primer efficiency values (listed in Materials and Methods) were calculcated using the Step One Plus software.

REFERENCES

1. Kewley, R.J., M.L. Whitelaw, and A. Chapman-Smith, *The mammalian basic helix-loop-helix/PAS family of transcriptional regulators.* Int J Biochem Cell Biol, 2004. **36**(2): p. 189-204.
2. Crews, S.T., *Control of cell lineage-specific development and transcription by bHLH-PAS proteins.* Genes Dev, 1998. **12**(5): p. 607-20.
3. Crews, S.T. and C.M. Fan, *Remembrance of things PAS: regulation of development by bHLH-PAS proteins.* Curr Opin Genet Dev, 1999. **9**(5): p. 580-7.
4. Bracken, C.P., M.L. Whitelaw, and D.J. Peet, *The hypoxia-inducible factors: key transcriptional regulators of hypoxic responses.* Cell Mol Life Sci, 2003. **60**(7): p. 1376-93.
5. Furness, S.G., M.J. Lees, and M.L. Whitelaw, *The dioxin (aryl hydrocarbon) receptor as a model for adaptive responses of bHLH/PAS transcription factors.* FEBS Lett, 2007. **581**(19): p. 3616-25.
6. Dolwick, K.M., H.I. Swanson, and C.A. Bradfield, *In vitro analysis of Ah receptor domains involved in ligand-activated DNA recognition.* Proc Natl Acad Sci U S A, 1993. **90**(18): p. 8566-70.
7. Fukunaga, B.N., et al., *Identification of functional domains of the aryl hydrocarbon receptor.* J Biol Chem, 1995. **270**(49): p. 29270-8.
8. Reisz-Porszasz, S., et al., *Identification of functional domains of the aryl hydrocarbon receptor nuclear translocator protein (ARNT).* Mol Cell Biol, 1994. **14**(9): p. 6075-86.
9. Jiang, B.H., et al., *Dimerization, DNA binding, and transactivation properties of hypoxia-inducible factor 1.* J Biol Chem, 1996. **271**(30): p. 17771-8.
10. Ema, M., et al., *Two new members of the murine Sim gene family are transcriptional repressors and show different expression patterns during mouse embryogenesis.* Mol Cell Biol, 1996. **16**(10): p. 5865-75.
11. Ooe, N., et al., *Identification of a novel basic helix-loop-helix-PAS factor, NXF, reveals a Sim2 competitive, positive regulatory role in dendritic-cytoskeleton modulator drebrin gene expression.* Mol Cell Biol, 2004. **24**(2): p. 608-16.
12. Ma, P.C., et al., *Crystal structure of MyoD bHLH domain-DNA complex: perspectives on DNA recognition and implications for transcriptional activation.* Cell, 1994. **77**(3): p. 451-9.
13. Moffett, P. and J. Pelletier, *Different transcriptional properties of mSim-1 and mSim-2.* FEBS Lett, 2000. **466**(1): p. 80-6.
14. Taylor, B.L. and I.B. Zhulin, *PAS domains: internal sensors of oxygen, redox potential, and light.* Microbiol Mol Biol Rev, 1999. **63**(2): p. 479-506.
15. Zelzer, E., P. Wappner, and B.Z. Shilo, *The PAS domain confers target gene specificity of Drosophila bHLH/PAS proteins.* Genes Dev, 1997. **11**(16): p. 2079-89.
16. Chapman-Smith, A., J.K. Lutwyche, and M.L. Whitelaw, *Contribution of the Per/Arnt/Sim (PAS) domains to DNA binding by the basic helix-loop-helix PAS transcriptional regulators.* J Biol Chem, 2004. **279**(7): p. 5353-62.
17. Chapman-Smith, A. and M.L. Whitelaw, *Novel DNA binding by a basic helix-loop-helix protein. The role of the dioxin receptor PAS domain.* J Biol Chem, 2006. **281**(18): p. 12535-45.

18. Pongratz, I., et al., *Role of the PAS domain in regulation of dimerization and DNA binding specificity of the dioxin receptor.* Mol Cell Biol, 1998. **18**(7): p. 4079-88.
19. Huang, Z.J., I. Edery, and M. Rosbash, *PAS is a dimerization domain common to Drosophila period and several transcription factors.* Nature, 1993. **364**(6434): p. 259-62.
20. Lindebro, M.C., L. Poellinger, and M.L. Whitelaw, *Protein-protein interaction via PAS domains: role of the PAS domain in positive and negative regulation of the bHLH/PAS dioxin receptor-Arnt transcription factor complex.* Embo J, 1995. **14**(14): p. 3528-39.
21. Franks, R.G. and S.T. Crews, *Transcriptional activation domains of the single-minded bHLH protein are required for CNS midline cell development.* Mech Dev, 1994. **45**(3): p. 269-77.
22. Yildiz, O., et al., *Crystal structure and interactions of the PAS repeat region of the Drosophila clock protein PERIOD.* Mol Cell, 2005. **17**(1): p. 69-82.
23. Vreede, J., et al., *PAS domains. Common structure and common flexibility.* J Biol Chem, 2003. **278**(20): p. 18434-9.
24. Lin, Y., et al., *Activity-dependent regulation of inhibitory synapse development by Npas4.* Nature, 2008. **455**(7217): p. 1198-204.
25. Sogawa, K., et al., *Possible function of Ah receptor nuclear translocator (Arnt) homodimer in transcriptional regulation.* Proc Natl Acad Sci U S A, 1995. **92**(6): p. 1936-40.
26. Swanson, H.I., W.K. Chan, and C.A. Bradfield, *DNA binding specificities and pairing rules of the Ah receptor, ARNT, and SIM proteins.* J Biol Chem, 1995. **270**(44): p. 26292-302.
27. Hirose, K., et al., *cDNA cloning and tissue-specific expression of a novel basic helix-loop-helix/PAS factor (Arnt2) with close sequence similarity to the aryl hydrocarbon receptor nuclear translocator (Arnt).* Mol Cell Biol, 1996. **16**(4): p. 1706-13.
28. Kobayashi, A., et al., *CBP/p300 functions as a possible transcriptional coactivator of Ah receptor nuclear translocator (Arnt).* J Biochem, 1997. **122**(4): p. 703-10.
29. Whitelaw, M., et al., *Ligand-dependent recruitment of the Arnt coregulator determines DNA recognition by the dioxin receptor.* Mol Cell Biol, 1993. **13**(4): p. 2504-14.
30. Panda, S., J.B. Hogenesch, and S.A. Kay, *Circadian rhythms from flies to human.* Nature, 2002. **417**(6886): p. 329-35.
31. Hilliker, A.J., et al., *Cytogenetic analysis of the chromosomal region immediately adjacent to the rosy locus in Drosophila melanogaster.* Genetics, 1980. **95**(1): p. 95-110.
32. Thomas, J.B., S.T. Crews, and C.S. Goodman, *Molecular genetics of the single-minded locus: a gene involved in the development of the Drosophila nervous system.* Cell, 1988. **52**(1): p. 133-41.
33. Nambu, J.R., et al., *The single-minded gene of Drosophila is required for the expression of genes important for the development of CNS midline cells.* Cell, 1990. **63**(1): p. 63-75.
34. Kim, I.O., et al., *CNS midline cells contribute to maintenance of the initial dorsoventral patterning of the Drosophila ventral neuroectoderm.* J Neurobiol, 2005. **62**(4): p. 397-405.

35. Nambu, J.R., et al., *The Drosophila single-minded gene encodes a helix-loop-helix protein that acts as a master regulator of CNS midline development.* Cell, 1991. **67**(6): p. 1157-67.
36. Crews, S.T., J.B. Thomas, and C.S. Goodman, *The Drosophila single-minded gene encodes a nuclear protein with sequence similarity to the per gene product.* Cell, 1988. **52**(1): p. 143-51.
37. Ward, M.P., J.T. Mosher, and S.T. Crews, *Regulation of bHLH-PAS protein subcellular localization during Drosophila embryogenesis.* Development, 1998. **125**(9): p. 1599-608.
38. Muralidhar, M.G., C.A. Callahan, and J.B. Thomas, *Single-minded regulation of genes in the embryonic midline of the Drosophila central nervous system.* Mech Dev, 1993. **41**(2-3): p. 129-38.
39. Kasai, Y., S. Stahl, and S. Crews, *Specification of the Drosophila CNS midline cell lineage: direct control of single-minded transcription by dorsal/ventral patterning genes.* Gene Expr, 1998. **7**(3): p. 171-89.
40. Kasai, Y., et al., *Dorsal-ventral patterning in Drosophila: DNA binding of snail protein to the single-minded gene.* Proc Natl Acad Sci U S A, 1992. **89**(8): p. 3414-8.
41. Morel, V. and F. Schweisguth, *Repression by suppressor of hairless and activation by Notch are required to define a single row of single-minded expressing cells in the Drosophila embryo.* Genes Dev, 2000. **14**(3): p. 377-88.
42. Chen, H., et al., *Single-minded and Down syndrome?* Nat Genet, 1995. **10**(1): p. 9-10.
43. Dahmane, N., et al., *Down syndrome-critical region contains a gene homologous to Drosophila sim expressed during rat and human central nervous system development.* Proc Natl Acad Sci U S A, 1995. **92**(20): p. 9191-5.
44. Lucente, D., et al., *Localization of 102 exons to a 2.5 Mb region involved in Down syndrome.* Hum Mol Genet, 1995. **4**(8): p. 1305-11.
45. Chrast, R., et al., *Cloning of two human homologs of the Drosophila single-minded gene SIM1 on chromosome 6q and SIM2 on 21q within the Down syndrome chromosomal region.* Genome Res, 1997. **7**(6): p. 615-24.
46. Moffett, P., et al., *Characterization of msim, a murine homologue of the Drosophila sim transcription factor.* Genomics, 1996. **35**(1): p. 144-55.
47. Yamaki, A., et al., *The mammalian single-minded (SIM) gene: mouse cDNA structure and diencephalic expression indicate a candidate gene for Down syndrome.* Genomics, 1996. **35**(1): p. 136-43.
48. Ema, M., et al., *cDNA cloning of a murine homologue of Drosophila single-minded, its mRNA expression in mouse development, and chromosome localization.* Biochem Biophys Res Commun, 1996. **218**(2): p. 588-94.
49. Fan, C.M., et al., *Expression patterns of two murine homologs of Drosophila single-minded suggest possible roles in embryonic patterning and in the pathogenesis of Down syndrome.* Mol Cell Neurosci, 1996. **7**(1): p. 1-16.
50. Metz, R.P., et al., *Differential transcriptional regulation by mouse single-minded 2s.* J Biol Chem, 2006. **281**(16): p. 10839-48.
51. Coumailleau, P., et al., *Characterization and developmental expression of xSim, a Xenopus bHLH/PAS gene related to the Drosophila neurogenic master gene single-minded.* Mech Dev, 2000. **99**(1-2): p. 163-6.
52. Coumailleau, P. and D. Duprez, *Sim1 and Sim2 expression during chick and mouse limb development.* Int J Dev Biol, 2009. **53**(1): p. 149-57.

53. Serluca, F.C. and M.C. Fishman, *Pre-pattern in the pronephric kidney field of zebrafish.* Development, 2001. **128**(12): p. 2233-41.
54. Haegeman, A., et al., *Mapping and SNP analysis of bovine candidate genes for meat and carcass quality.* Anim Genet, 2003. **34**(5): p. 349-53.
55. Frazer, K.A., et al., *Noncoding sequences conserved in a limited number of mammals in the SIM2 interval are frequently functional.* Genome Res, 2004. **14**(3): p. 367-72.
56. Vargas-Vila, M.A., et al., *A prominent requirement for single-minded and the ventral midline in patterning the dorsoventral axis of the crustacean Parhyale hawaiensis.* Development, 2010.
57. Caqueret, A., F. Boucher, and J.L. Michaud, *Laminar organization of the early developing anterior hypothalamus.* Dev Biol, 2006. **298**(1): p. 95-106.
58. Michaud, J.L., et al., *Development of neuroendocrine lineages requires the bHLH-PAS transcription factor SIM1.* Genes Dev, 1998. **12**(20): p. 3264-75.
59. Holder, J.L., Jr., et al., *Sim1 gene dosage modulates the homeostatic feeding response to increased dietary fat in mice.* Am J Physiol Endocrinol Metab, 2004. **287**(1): p. E105-13.
60. Probst, M.R., et al., *Two murine homologs of the Drosophila single-minded protein that interact with the mouse aryl hydrocarbon receptor nuclear translocator protein.* J Biol Chem, 1997. **272**(7): p. 4451-7.
61. DeYoung, M.P., M. Tress, and R. Narayanan, *Down's syndrome-associated Single Minded 2 gene as a pancreatic cancer drug therapy target.* Cancer Lett, 2003. **200**(1): p. 25-31.
62. Goshu, E., et al., *Sim2 contributes to neuroendocrine hormone gene expression in the anterior hypothalamus.* Mol Endocrinol, 2004. **18**(5): p. 1251-62.
63. Goshu, E., et al., *Sim2 mutants have developmental defects not overlapping with those of Sim1 mutants.* Mol Cell Biol, 2002. **22**(12): p. 4147-57.
64. DeYoung, M.P., M. Tress, and R. Narayanan, *Identification of Down's syndrome critical locus gene SIM2-s as a drug therapy target for solid tumors.* Proc Natl Acad Sci U S A, 2003. **100**(8): p. 4760-5.
65. Kwak, H.I., et al., *Inhibition of breast cancer growth and invasion by single-minded 2s.* Carcinogenesis, 2007. **28**(2): p. 259-66.
66. Sun, T., et al., *Olig bHLH proteins interact with homeodomain proteins to regulate cell fate acquisition in progenitors of the ventral neural tube.* Curr Biol, 2001. **11**(18): p. 1413-20.
67. Lee, J., et al., *Neurogenin3 participates in gliogenesis in the developing vertebrate spinal cord.* Dev Biol, 2003. **253**(1): p. 84-98.
68. Treff, N.R., et al., *Differentiation of embryonic stem cells conditionally expressing neurogenin 3.* Stem Cells, 2006. **24**(11): p. 2529-37.
69. Briscoe, J., et al., *Homeobox gene Nkx2.2 and specification of neuronal identity by graded Sonic hedgehog signalling.* Nature, 1999. **398**(6728): p. 622-7.
70. Yang, L., S. Rastegar, and U. Strahle, *Regulatory interactions specifying Kolmer-Agduhr interneurons.* Development. **137**(16): p. 2713-22.
71. Yang, C., et al., *Regulatory interaction between arylhydrocarbon receptor and SIM1, two basic helix-loop-helix PAS proteins involved in the control of food intake.* J Biol Chem, 2004. **279**(10): p. 9306-12.
72. Fernandez-Salguero, P., et al., *Immune system impairment and hepatic fibrosis in mice lacking the dioxin-binding Ah receptor.* Science, 1995. **268**(5211): p. 722-6.

73. Schmidt, J.V., et al., *Characterization of a murine Ahr null allele: involvement of the Ah receptor in hepatic growth and development.* Proc Natl Acad Sci U S A, 1996. **93**(13): p. 6731-6.
74. Mimura, J., et al., *Loss of teratogenic response to 2,3,7,8-tetrachlorodibenzo-p-dioxin (TCDD) in mice lacking the Ah (dioxin) receptor.* Genes Cells, 1997. **2**(10): p. 645-54.
75. Woods, S.L. and M.L. Whitelaw, *Differential activities of murine single minded 1 (SIM1) and SIM2 on a hypoxic response element. Cross-talk between basic helix-loop-helix/per-Arnt-Sim homology transcription factors.* J Biol Chem, 2002. **277**(12): p. 10236-43.
76. Kublaoui, B.M., et al., *Sim1 haploinsufficiency impairs melanocortin-mediated anorexia and activation of paraventricular nucleus neurons.* Mol Endocrinol, 2006. **20**(10): p. 2483-92.
77. Schwartz, M.W., et al., *Central nervous system control of food intake.* Nature, 2000. **404**(6778): p. 661-71.
78. Gustafson, T.L., et al., *Ha-Ras transformation of MCF10A cells leads to repression of Singleminded-2s through NOTCH and C/EBPbeta.* Oncogene, 2009. **28**(12): p. 1561-8.
79. Yamaki, A., et al., *Molecular mechanisms of human single-minded 2 (SIM2) gene expression: identification of a promoter site in the SIM2 genomic sequence.* Gene, 2001. **270**(1-2): p. 265-75.
80. Epstein, D.J., et al., *Members of the bHLH-PAS family regulate Shh transcription in forebrain regions of the mouse CNS.* Development, 2000. **127**(21): p. 4701-9.
81. Michaud, J.L., et al., *ARNT2 acts as the dimerization partner of SIM1 for the development of the hypothalamus.* Mech Dev, 2000. **90**(2): p. 253-61.
82. Jain, S., et al., *Expression of ARNT, ARNT2, HIF1 alpha, HIF2 alpha and Ah receptor mRNAs in the developing mouse.* Mech Dev, 1998. **73**(1): p. 117-23.
83. Aitola, M.H. and M.T. Pelto-Huikko, *Expression of Arnt and Arnt2 mRNA in developing murine tissues.* J Histochem Cytochem, 2003. **51**(1): p. 41-54.
84. Woods, S., et al., *The bHLH/Per-Arnt-Sim transcription factor SIM2 regulates muscle transcript myomesin2 via a novel, non-canonical E-box sequence.* Nucleic Acids Res, 2008. **36**(11): p. 3716-27.
85. Farrall, A.L. and M.L. Whitelaw, *The HIF1alpha-inducible pro-cell death gene BNIP3 is a novel target of SIM2s repression through cross-talk on the hypoxia response element.* Oncogene, 2009. **28**(41): p. 3671-80.
86. Estes, P., J. Mosher, and S.T. Crews, *Drosophila single-minded represses gene transcription by activating the expression of repressive factors.* Dev Biol, 2001. **232**(1): p. 157-75.
87. Moffett, P., M. Reece, and J. Pelletier, *The murine Sim-2 gene product inhibits transcription by active repression and functional interference.* Mol Cell Biol, 1997. **17**(9): p. 4933-47.
88. Wharton, K.A., Jr., et al., *Control of CNS midline transcription by asymmetric E-box-like elements: similarity to xenobiotic responsive regulation.* Development, 1994. **120**(12): p. 3563-9.
89. Xu, C. and C.M. Fan, *Allocation of paraventricular and supraoptic neurons requires Sim1 function: a role for a Sim1 downstream gene PlexinC1.* Mol Endocrinol, 2007. **21**(5): p. 1234-45.
90. Marion, J.F., et al., *Sim1 and Sim2 are required for the correct targeting of mammillary body axons.* Development, 2005. **132**(24): p. 5527-37.

91. Acampora, D., et al., *Progressive impairment of developing neuroendocrine cell lineages in the hypothalamus of mice lacking the Orthopedia gene.* Genes Dev, 1999. **13**(21): p. 2787-800.
92. Nakai, S., et al., *The POU domain transcription factor Brn-2 is required for the determination of specific neuronal lineages in the hypothalamus of the mouse.* Genes Dev, 1995. **9**(24): p. 3109-21.
93. Wang, W. and T. Lufkin, *The murine Otp homeobox gene plays an essential role in the specification of neuronal cell lineages in the developing hypothalamus.* Dev Biol, 2000. **227**(2): p. 432-49.
94. Keith, B., D.M. Adelman, and M.C. Simon, *Targeted mutation of the murine arylhydrocarbon receptor nuclear translocator 2 (Arnt2) gene reveals partial redundancy with Arnt.* Proc Natl Acad Sci U S A, 2001. **98**(12): p. 6692-7.
95. Hosoya, T., et al., *Defective development of secretory neurones in the hypothalamus of Arnt2-knockout mice.* Genes Cells, 2001. **6**(4): p. 361-74.
96. Michaud, J.L., et al., *Sim1 haploinsufficiency causes hyperphagia, obesity and reduction of the paraventricular nucleus of the hypothalamus.* Hum Mol Genet, 2001. **10**(14): p. 1465-73.
97. Duplan, S.M., et al., *Impact of Sim1 gene dosage on the development of the paraventricular and supraoptic nuclei of the hypothalamus.* Eur J Neurosci, 2009. **30**(12): p. 2239-49.
98. Tolson, K.P., et al., *Postnatal Sim1 deficiency causes hyperphagic obesity and reduced Mc4r and oxytocin expression.* J Neurosci. **30**(10): p. 3803-12.
99. Doetschman, T., *Influence of genetic background on genetically engineered mouse phenotypes.* Methods Mol Biol, 2009. **530**: p. 423-33.
100. Yang, C., et al., *Adenoviral-mediated modulation of Sim1 expression in the paraventricular nucleus affects food intake.* J Neurosci, 2006. **26**(26): p. 7116-20.
101. Sahu, A., *Minireview: A hypothalamic role in energy balance with special emphasis on leptin.* Endocrinology, 2004. **145**(6): p. 2613-20.
102. Kublaoui, B.M., et al., *SIM1 overexpression partially rescues agouti yellow and diet-induced obesity by normalizing food intake.* Endocrinology, 2006. **147**(10): p. 4542-9.
103. Shamblott, M.J., et al., *Craniofacial abnormalities resulting from targeted disruption of the murine Sim2 gene.* Dev Dyn, 2002. **224**(4): p. 373-80.
104. Ema, M., et al., *Mild impairment of learning and memory in mice overexpressing the mSim2 gene located on chromosome 16: an animal model of Down's syndrome.* Hum Mol Genet, 1999. **8**(8): p. 1409-15.
105. Chrast, R., et al., *Mice trisomic for a bacterial artificial chromosome with the single-minded 2 gene (Sim2) show phenotypes similar to some of those present in the partial trisomy 16 mouse models of Down syndrome.* Hum Mol Genet, 2000. **9**(12): p. 1853-64.
106. Hankinson, O., *Single-step selection of clones of a mouse hepatoma line deficient in aryl hydrocarbon hydroxylase.* Proc Natl Acad Sci U S A, 1979. **76**(1): p. 373-6.
107. Hankinson, O., *Evidence that Benzo(a) pyrene-resistant, aryl hydrocarbon hydroxylase-deficient variants of mouse hepatoma line, Hepa-1, are mutational in origin.* Somatic Cell Genet, 1981. **7**(4): p. 373-88.
108. Hoffman, E.C., et al., *Cloning of a factor required for activity of the Ah (dioxin) receptor.* Science, 1991. **252**(5008): p. 954-8.

109. Sonnenfeld, M., et al., *The Drosophila tango gene encodes a bHLH-PAS protein that is orthologous to mammalian Arnt and controls CNS midline and tracheal development.* Development, 1997. **124**(22): p. 4571-82.
110. Scheel, J. and D. Schrenk, *Genomic structure of the human Ah receptor nuclear translocator gene (hARNT).* Hum Genet, 2000. **107**(4): p. 397-9.
111. Korkalainen, M., J. Tuomisto, and R. Pohjanvirta, *Identification of novel splice variants of ARNT and ARNT2 in the rat.* Biochem Biophys Res Commun, 2003. **303**(4): p. 1095-100.
112. Drutel, G., et al., *Cloning and selective expression in brain and kidney of ARNT2 homologous to the Ah receptor nuclear translocator (ARNT).* Biochem Biophys Res Commun, 1996. **225**(2): p. 333-9.
113. Pollenz, R.S., et al., *Isolation and expression of cDNAs from rainbow trout (Oncorhynchus mykiss) that encode two novel basic helix-loop-Helix/PER-ARNT-SIM (bHLH/PAS) proteins with distinct functions in the presence of the aryl hydrocarbon receptor. Evidence for alternative mRNA splicing and dominant negative activity in the bHLH/PAS family.* J Biol Chem, 1996. **271**(48): p. 30886-96.
114. Wang, F., et al., *Structure and expression of the mouse AhR nuclear translocator (mArnt) gene.* J Biol Chem, 1998. **273**(38): p. 24867-73.
115. Nagase, T., et al., *Prediction of the coding sequences of unidentified human genes. VII. The complete sequences of 100 new cDNA clones from brain which can code for large proteins in vitro.* DNA Res, 1997. **4**(2): p. 141-50.
116. Barrow, L.L., et al., *Aryl hydrocarbon receptor nuclear translocator 2 (ARNT2): structure, gene mapping, polymorphisms, and candidate evaluation for human orofacial clefts.* Teratology, 2002. **66**(2): p. 85-90.
117. Tanguay, R.L., et al., *Identification and expression of alternatively spliced aryl hydrocarbon nuclear translocator 2 (ARNT2) cDNAs from zebrafish with distinct functions.* Biochim Biophys Acta, 2000. **1494**(1-2): p. 117-28.
118. Powell-Coffman, J.A., C.A. Bradfield, and W.B. Wood, *Caenorhabditis elegans orthologs of the aryl hydrocarbon receptor and its heterodimerization partner the aryl hydrocarbon receptor nuclear translocator.* Proc Natl Acad Sci U S A, 1998. **95**(6): p. 2844-9.
119. Powell, W.H., et al., *Functional diversity of vertebrate ARNT proteins: identification of ARNT2 as the predominant form of ARNT in the marine teleost, Fundulus heteroclitus.* Arch Biochem Biophys, 1999. **361**(1): p. 156-63.
120. Abbott, B.D. and M.R. Probst, *Developmental expression of two members of a new class of transcription factors: II. Expression of aryl hydrocarbon receptor nuclear translocator in the C57BL/6N mouse embryo.* Dev Dyn, 1995. **204**(2): p. 144-55.
121. Kainu, T., J.A. Gustafsson, and M. Pelto-Huikko, *The dioxin receptor and its nuclear translocator (Arnt) in the rat brain.* Neuroreport, 1995. **6**(18): p. 2557-60.
122. Petersen, S.L., et al., *Distribution of mRNAs encoding the arylhydrocarbon receptor, arylhydrocarbon receptor nuclear translocator, and arylhydrocarbon receptor nuclear translocator-2 in the rat brain and brainstem.* J Comp Neurol, 2000. **427**(3): p. 428-39.
123. Sonnenfeld, M.J., C. Delvecchio, and X. Sun, *Analysis of the transcriptional activation domain of the Drosophila tango bHLH-PAS transcription factor.* Dev Genes Evol, 2005. **215**(5): p. 221-9.
124. Eguchi, H., et al., *A nuclear localization signal of human aryl hydrocarbon receptor nuclear translocator/hypoxia-inducible factor 1beta is a novel bipartite*

type recognized by the two components of nuclear pore-targeting complex. J Biol Chem, 1997. **272**(28): p. 17640-7.
125. Sojka, K.M., C.B. Kern, and R.S. Pollenz, *Expression and subcellular localization of the aryl hydrocarbon receptor nuclear translocator (ARNT) protein in mouse and chicken over developmental time.* Anat Rec, 2000. **260**(4): p. 327-34.
126. Holmes, J.L. and R.S. Pollenz, *Determination of aryl hydrocarbon receptor nuclear translocator protein concentration and subcellular localization in hepatic and nonhepatic cell culture lines: development of quantitative Western blotting protocols for calculation of aryl hydrocarbon receptor and aryl hydrocarbon receptor nuclear translocator protein in total cell lysates.* Mol Pharmacol, 1997. **52**(2): p. 202-11.
127. Sekine, H., et al., *Unique and overlapping transcriptional roles of arylhydrocarbon receptor nuclear translocator (Arnt) and Arnt2 in xenobiotic and hypoxic responses.* J Biol Chem, 2006. **281**(49): p. 37507-16.
128. Dougherty, E.J. and R.S. Pollenz, *Analysis of Ah receptor-ARNT and Ah receptor-ARNT2 complexes in vitro and in cell culture.* Toxicol Sci, 2008. **103**(1): p. 191-206.
129. Kobayashi, A., K. Sogawa, and Y. Fujii-Kuriyama, *Cooperative interaction between AhR.Arnt and Sp1 for the drug-inducible expression of CYP1A1 gene.* J Biol Chem, 1996. **271**(21): p. 12310-6.
130. Antonsson, C., et al., *Distinct roles of the molecular chaperone hsp90 in modulating dioxin receptor function via the basic helix-loop-helix and PAS domains.* Mol Cell Biol, 1995. **15**(2): p. 756-65.
131. Kazlauskas, A., L. Poellinger, and I. Pongratz, *Evidence that the co-chaperone p23 regulates ligand responsiveness of the dioxin (Aryl hydrocarbon) receptor.* J Biol Chem, 1999. **274**(19): p. 13519-24.
132. Maltepe, E., et al., *Abnormal angiogenesis and responses to glucose and oxygen deprivation in mice lacking the protein ARNT.* Nature, 1997. **386**(6623): p. 403-7.
133. Kozak, K.R., B. Abbott, and O. Hankinson, *ARNT-deficient mice and placental differentiation.* Dev Biol, 1997. **191**(2): p. 297-305.
134. Carmeliet, P., et al., *Abnormal blood vessel development and lethality in embryos lacking a single VEGF allele.* Nature, 1996. **380**(6573): p. 435-9.
135. Ferrara, N., et al., *Heterozygous embryonic lethality induced by targeted inactivation of the VEGF gene.* Nature, 1996. **380**(6573): p. 439-42.
136. Grompe, M., et al., *Loss of fumarylacetoacetate hydrolase is responsible for the neonatal hepatic dysfunction phenotype of lethal albino mice.* Genes Dev, 1993. **7**(12A): p. 2298-307.
137. Liu, C., et al., *Identification of the downstream targets of SIM1 and ARNT2, a pair of transcription factors essential for neuroendocrine cell differentiation.* J Biol Chem, 2003. **278**(45): p. 44857-67.
138. Coll, A.P., I.S. Farooqi, and S. O'Rahilly, *The hormonal control of food intake.* Cell, 2007. **129**(2): p. 251-62.
139. Kalra, S.P., et al., *Interacting appetite-regulating pathways in the hypothalamic regulation of body weight.* Endocr Rev, 1999. **20**(1): p. 68-100.
140. Valassi, E., M. Scacchi, and F. Cavagnini, *Neuroendocrine control of food intake.* Nutr Metab Cardiovasc Dis, 2008. **18**(2): p. 158-68.
141. Murphy, K.G. and S.R. Bloom, *Gut hormones and the regulation of energy homeostasis.* Nature, 2006. **444**(7121): p. 854-9.
142. Beckers, S., et al., *The role of the leptin-melanocortin signalling pathway in the control of food intake.* Crit Rev Eukaryot Gene Expr, 2009. **19**(4): p. 267-87.

143. Barsh, G.S. and M.W. Schwartz, *Genetic approaches to studying energy balance: perception and integration.* Nat Rev Genet, 2002. **3**(8): p. 589-600.
144. Sahu, A., *Leptin signaling in the hypothalamus: emphasis on energy homeostasis and leptin resistance.* Front Neuroendocrinol, 2003. **24**(4): p. 225-53.
145. Balthasar, N., et al., *Divergence of melanocortin pathways in the control of food intake and energy expenditure.* Cell, 2005. **123**(3): p. 493-505.
146. Samson, W.K., et al., *Neuronostatin encoded by the somatostatin gene regulates neuronal, cardiovascular, and metabolic functions.* J Biol Chem, 2008. **283**(46): p. 31949-59.
147. Swanson, L.W. and H.G. Kuypers, *The paraventricular nucleus of the hypothalamus: cytoarchitectonic subdivisions and organization of projections to the pituitary, dorsal vagal complex, and spinal cord as demonstrated by retrograde fluorescence double-labeling methods.* J Comp Neurol, 1980. **194**(3): p. 555-70.
148. Lechan, R.M. and R. Toni (2004) *Functional Anatomy of the Hypothalamus and Pituitary*. Chapter 3B - The Hypothalamus and Pituitary. Free online full text from Endotext.com [http://www.endotext.org/index.htm].
149. Sawchenko, P.E. and L.W. Swanson, *Immunohistochemical identification of neurons in the paraventricular nucleus of the hypothalamus that project to the medulla or to the spinal cord in the rat.* J Comp Neurol, 1982. **205**(3): p. 260-72.
150. Rinaman, L., *Ontogeny of hypothalamic-hindbrain feeding control circuits.* Dev Psychobiol, 2006. **48**(5): p. 389-96.
151. Swanson, L.W. and P.E. Sawchenko, *Hypothalamic integration: organization of the paraventricular and supraoptic nuclei.* Annu Rev Neurosci, 1983. **6**: p. 269-324.
152. Bultman, S.J., E.J. Michaud, and R.P. Woychik, *Molecular characterization of the mouse agouti locus.* Cell, 1992. **71**(7): p. 1195-204.
153. Kublaoui, B.M., et al., *Oxytocin deficiency mediates hyperphagic obesity of Sim1 haploinsufficient mice.* Mol Endocrinol, 2008. **22**(7): p. 1723-34.
154. Lu, D., et al., *Agouti protein is an antagonist of the melanocyte-stimulating-hormone receptor.* Nature, 1994. **371**(6500): p. 799-802.
155. Millar, S.E., et al., *Expression and transgenic studies of the mouse agouti gene provide insight into the mechanisms by which mammalian coat color patterns are generated.* Development, 1995. **121**(10): p. 3223-32.
156. Marsh, D.J., et al., *Response of melanocortin-4 receptor-deficient mice to anorectic and orexigenic peptides.* Nat Genet, 1999. **21**(1): p. 119-22.
157. Chan, W.K., et al., *Cross-talk between the aryl hydrocarbon receptor and hypoxia inducible factor signaling pathways. Demonstration of competition and compensation.* J Biol Chem, 1999. **274**(17): p. 12115-23.
158. Gradin, K., et al., *Functional interference between hypoxia and dioxin signal transduction pathways: competition for recruitment of the Arnt transcription factor.* Mol Cell Biol, 1996. **16**(10): p. 5221-31.
159. Chen, A.S., et al., *Inactivation of the mouse melanocortin-3 receptor results in increased fat mass and reduced lean body mass.* Nat Genet, 2000. **26**(1): p. 97-102.
160. Cummings, D.E. and M.W. Schwartz, *Melanocortins and body weight: a tale of two receptors.* Nat Genet, 2000. **26**(1): p. 8-9.
161. Schonemann, M.D., et al., *Development and survival of the endocrine hypothalamus and posterior pituitary gland requires the neuronal POU domain factor Brn-2.* Genes Dev, 1995. **9**(24): p. 3122-35.

162. Wharton, K.A., Jr. and S.T. Crews, *CNS midline enhancers of the Drosophila slit and Toll genes.* Mech Dev, 1993. **40**(3): p. 141-54.
163. Ohshiro, T. and K. Saigo, *Transcriptional regulation of breathless FGF receptor gene by binding of TRACHEALESS/dARNT heterodimers to three central midline elements in Drosophila developing trachea.* Development, 1997. **124**(20): p. 3975-86.
164. Zelzer, E. and B.Z. Shilo, *Interaction between the bHLH-PAS protein Trachealess and the POU-domain protein Drifter, specifies tracheal cell fates.* Mech Dev, 2000. **91**(1-2): p. 163-73.
165. Apitz, H., et al., *Single-minded, Dmef2, Pointed, and Su(H) act on identified regulatory sequences of the roughest gene in Drosophila melanogaster.* Dev Genes Evol, 2005. **215**(9): p. 460-69.
166. Sedaghat, Y., W.F. Miranda, and M.J. Sonnenfeld, *The jing Zn-finger transcription factor is a mediator of cellular differentiation in the Drosophila CNS midline and trachea.* Development, 2002. **129**(11): p. 2591-606.
167. Chang, J., S.H. Jeon, and S.H. Kim, *The hierarchical relationship among the spitz/Egfr signaling genes in cell fate determination in the Drosophila ventral neuroectoderm.* Mol Cells, 2003. **15**(2): p. 186-93.
168. Estes, P., E. Fulkerson, and Y. Zhang, *Identification of motifs that are conserved in 12 Drosophila species and regulate midline glia vs. neuron expression.* Genetics, 2008. **178**(2): p. 787-99.
169. Sonnenfeld, M.J. and J.R. Jacobs, *Mesectodermal cell fate analysis in Drosophila midline mutants.* Mech Dev, 1994. **46**(1): p. 3-13.
170. Kim, S.H. and S.T. Crews, *Influence of Drosophila ventral epidermal development by the CNS midline cells and spitz class genes.* Development, 1993. **118**(3): p. 893-901.
171. Klambt, C., J.R. Jacobs, and C.S. Goodman, *The midline of the Drosophila central nervous system: a model for the genetic analysis of cell fate, cell migration, and growth cone guidance.* Cell, 1991. **64**(4): p. 801-15.
172. Kearney, J.B., et al., *Gene expression profiling of the developing Drosophila CNS midline cells.* Dev Biol, 2004. **275**(2): p. 473-92.
173. Crews, S., et al., *Drosophila single-minded gene and the molecular genetics of CNS midline development.* J Exp Zool, 1992. **261**(3): p. 234-44.
174. Camacho, A. and I. McDougall, *Intracratonic, strikeslip partitioned transpression and the formation of eclogite facies rocks: An example from the Musgrave Block, central Australia.* Tectonics, 2000. **19**: p. 978-996.
175. Laffin, B., et al., *Loss of singleminded-2s in the mouse mammary gland induces an epithelial-mesenchymal transition associated with up-regulation of slug and matrix metalloprotease 2.* Mol Cell Biol, 2008. **28**(6): p. 1936-46.
176. Bruick, R.K., *Expression of the gene encoding the proapoptotic Nip3 protein is induced by hypoxia.* Proc Natl Acad Sci U S A, 2000. **97**(16): p. 9082-7.
177. Holder, J.L., Jr., N.F. Butte, and A.R. Zinn, *Profound obesity associated with a balanced translocation that disrupts the SIM1 gene.* Hum Mol Genet, 2000. **9**(1): p. 101-8.
178. Klein, O.D., et al., *Interstitial deletions of chromosome 6q: genotype-phenotype correlation utilizing array CGH.* Clin Genet, 2007. **71**(3): p. 260-6.
179. Villa, A., et al., *De novo interstitial deletion q16.2q21 on chromosome 6.* Am J Med Genet, 1995. **55**(3): p. 379-83.
180. Gilhuis, H.J., et al., *Interstitial 6q deletion with a Prader-Willi-like phenotype: a new case and review of the literature.* Eur J Paediatr Neurol, 2000. **4**(1): p. 39-43.

181. Faivre, L., et al., *Deletion of the SIM1 gene (6q16.2) in a patient with a Prader-Willi-like phenotype.* J Med Genet, 2002. **39**(8): p. 594-6.
182. Varela, M.C., et al., *A new case of interstitial 6q16.2 deletion in a patient with Prader-Willi-like phenotype and investigation of SIM1 gene deletion in 87 patients with syndromic obesity.* Eur J Med Genet, 2006. **49**(4): p. 298-305.
183. Wang, J.C., et al., *A 5-Mb microdeletion at 6q16.1-q16.3 with SIM gene deletion and obesity.* Am J Med Genet A, 2008. **146A**(22): p. 2975-8.
184. Stein, C.K., et al., *Interstitial 6q deletion and Prader-Willi-like phenotype.* Clin Genet, 1996. **49**(6): p. 306-10.
185. Le Caignec, C., et al., *Interstitial 6q deletion: clinical and array CGH characterisation of a new patient.* Eur J Med Genet, 2005. **48**(3): p. 339-45.
186. Bonaglia, M.C., et al., *Detailed phenotype-genotype study in five patients with chromosome 6q16 deletion: narrowing the critical region for Prader-Willi-like phenotype.* Eur J Hum Genet, 2008. **16**(12): p. 1443-9.
187. Walley, A.J., J.E. Asher, and P. Froguel, *The genetic contribution to non-syndromic human obesity.* Nat Rev Genet, 2009. **10**(7): p. 431-42.
188. Ranadive, S.A. and C. Vaisse, *Lessons from extreme human obesity: monogenic disorders.* Endocrinol Metab Clin North Am, 2008. **37**(3): p. 733-51, x.
189. Barsh, G.S., I.S. Farooqi, and S. O'Rahilly, *Genetics of body-weight regulation.* Nature, 2000. **404**(6778): p. 644-51.
190. O'Rahilly, S. and I.S. Farooqi, *Human obesity: a heritable neurobehavioral disorder that is highly sensitive to environmental conditions.* Diabetes, 2008. **57**(11): p. 2905-10.
191. O'Rahilly, S., *Human genetics illuminates the paths to metabolic disease.* Nature, 2009. **462**(7271): p. 307-14.
192. Comuzzie, A.G. and D.B. Allison, *The search for human obesity genes.* Science, 1998. **280**(5368): p. 1374-7.
193. Gray, J., et al., *Hyperphagia, severe obesity, impaired cognitive function, and hyperactivity associated with functional loss of one copy of the brain-derived neurotrophic factor (BDNF) gene.* Diabetes, 2006. **55**(12): p. 3366-71.
194. Han, J.C., et al., *Brain-derived neurotrophic factor and obesity in the WAGR syndrome.* N Engl J Med, 2008. **359**(9): p. 918-27.
195. Yeo, G.S., et al., *A de novo mutation affecting human TrkB associated with severe obesity and developmental delay.* Nat Neurosci, 2004. **7**(11): p. 1187-9.
196. Jackson, R.S., et al., *Obesity and impaired prohormone processing associated with mutations in the human prohormone convertase 1 gene.* Nat Genet, 1997. **16**(3): p. 303-6.
197. Yaswen, L., et al., *Obesity in the mouse model of pro-opiomelanocortin deficiency responds to peripheral melanocortin.* Nat Med, 1999. **5**(9): p. 1066-70.
198. Krude, H., et al., *Severe early-onset obesity, adrenal insufficiency and red hair pigmentation caused by POMC mutations in humans.* Nat Genet, 1998. **19**(2): p. 155-7.
199. Zhang, Y., et al., *Positional cloning of the mouse obese gene and its human homologue.* Nature, 1994. **372**(6505): p. 425-32.
200. Montague, C.T., et al., *Congenital leptin deficiency is associated with severe early-onset obesity in humans.* Nature, 1997. **387**(6636): p. 903-8.
201. Strobel, A., et al., *A leptin missense mutation associated with hypogonadism and morbid obesity.* Nat Genet, 1998. **18**(3): p. 213-5.

202. Chen, H., et al., *Evidence that the diabetes gene encodes the leptin receptor: identification of a mutation in the leptin receptor gene in db/db mice.* Cell, 1996. **84**(3): p. 491-5.
203. Lee, G.H., et al., *Abnormal splicing of the leptin receptor in diabetic mice.* Nature, 1996. **379**(6566): p. 632-5.
204. Clement, K., et al., *A mutation in the human leptin receptor gene causes obesity and pituitary dysfunction.* Nature, 1998. **392**(6674): p. 398-401.
205. Huszar, D., et al., *Targeted disruption of the melanocortin-4 receptor results in obesity in mice.* Cell, 1997. **88**(1): p. 131-41.
206. Yeo, G.S., et al., *A frameshift mutation in MC4R associated with dominantly inherited human obesity.* Nat Genet, 1998. **20**(2): p. 111-2.
207. Hinney, A., et al., *Several mutations in the melanocortin-4 receptor gene including a nonsense and a frameshift mutation associated with dominantly inherited obesity in humans.* J Clin Endocrinol Metab, 1999. **84**(4): p. 1483-6.
208. Vaisse, C., et al., *A frameshift mutation in human MC4R is associated with a dominant form of obesity.* Nat Genet, 1998. **20**(2): p. 113-4.
209. Farooqi, I.S., et al., *Clinical spectrum of obesity and mutations in the melanocortin 4 receptor gene.* N Engl J Med, 2003. **348**(12): p. 1085-95.
210. Frayling, T.M., et al., *A common variant in the FTO gene is associated with body mass index and predisposes to childhood and adult obesity.* Science, 2007. **316**(5826): p. 889-94.
211. Dina, C., et al., *Variation in FTO contributes to childhood obesity and severe adult obesity.* Nat Genet, 2007. **39**(6): p. 724-6.
212. Cecil, J.E., et al., *An obesity-associated FTO gene variant and increased energy intake in children.* N Engl J Med, 2008. **359**(24): p. 2558-66.
213. Gerken, T., et al., *The obesity-associated FTO gene encodes a 2-oxoglutarate-dependent nucleic acid demethylase.* Science, 2007. **318**(5855): p. 1469-72.
214. Miller, M.W., et al., *Cloning of the mouse agouti gene predicts a secreted protein ubiquitously expressed in mice carrying the lethal yellow mutation.* Genes Dev, 1993. **7**(3): p. 454-67.
215. Naggert, J.K., et al., *Hyperproinsulinaemia in obese fat/fat mice associated with a carboxypeptidase E mutation which reduces enzyme activity.* Nat Genet, 1995. **10**(2): p. 135-42.
216. Tecott, L.H., et al., *Eating disorder and epilepsy in mice lacking 5-HT2c serotonin receptors.* Nature, 1995. **374**(6522): p. 542-6.
217. Good, D.J., et al., *Hypogonadism and obesity in mice with a targeted deletion of the Nhlh2 gene.* Nat Genet, 1997. **15**(4): p. 397-401.
218. Noben-Trauth, K., et al., *A candidate gene for the mouse mutation tubby.* Nature, 1996. **380**(6574): p. 534-8.
219. Meyre, D., et al., *A genome-wide scan for childhood obesity-associated traits in French families shows significant linkage on chromosome 6q22.31-q23.2.* Diabetes, 2004. **53**(3): p. 803-11.
220. Ahituv, N., et al., *Medical sequencing at the extremes of human body mass.* Am J Hum Genet, 2007. **80**(4): p. 779-91.
221. Hung, C.C., et al., *Studies of the SIM1 gene in relation to human obesity and obesity-related traits.* Int J Obes (Lond), 2007. **31**(3): p. 429-34.
222. Ghoussaini, M., et al., *Analysis of the SIM1 contribution to polygenic obesity in the French population.* Obesity (Silver Spring), 2010. **18**(8): p. 1670-5.
223. Traurig, M., et al., *Common variation in SIM1 is reproducibly associated with BMI in Pima Indians.* Diabetes, 2009. **58**(7): p. 1682-9.

224. Belsham, D.D., et al., *Generation of a phenotypic array of hypothalamic neuronal cell models to study complex neuroendocrine disorders.* Endocrinology, 2004. **145**(1): p. 393-400.
225. Rees, S., et al., *Bicistronic vector for the creation of stable mammalian cell lines that predisposes all antibiotic-resistant cells to express recombinant protein.* Biotechniques, 1996. **20**(1): p. 102-4, 106, 108-10.
226. Hobbs, S., S. Jitrapakdee, and J.C. Wallace, *Development of a bicistronic vector driven by the human polypeptide chain elongation factor 1alpha promoter for creation of stable mammalian cell lines that express very high levels of recombinant proteins.* Biochem Biophys Res Commun, 1998. **252**(2): p. 368-72.
227. Muller, P.Y., et al., *Processing of gene expression data generated by quantitative real-time RT-PCR.* Biotechniques, 2002. **32**(6): p. 1372-4, 1376, 1378-9.
228. Lees, M.J. and M.L. Whitelaw, *Multiple roles of ligand in transforming the dioxin receptor to an active basic helix-loop-helix/PAS transcription factor complex with the nuclear protein Arnt.* Mol Cell Biol, 1999. **19**(8): p. 5811-22.
229. Allison, D.B. and N.J. Schork, *Selected methodological issues in meiotic mapping of obesity genes in humans: issues of power and efficiency.* Behav Genet, 1997. **27**(4): p. 401-21.
230. Williams, D.R., et al., *Undiagnosed glucose intolerance in the community: the Isle of Ely Diabetes Project.* Diabet Med, 1995. **12**(1): p. 30-5.
231. Holm, V.A., et al., *Prader-Willi syndrome: consensus diagnostic criteria.* Pediatrics, 1993. **91**(2): p. 398-402.
232. Smeets, D.F., et al., *Prader-Willi syndrome and Angelman syndrome in cousins from a family with a translocation between chromosomes 6 and 15.* N Engl J Med, 1992. **326**(12): p. 807-11.
233. Swaab, D.F., J.S. Purba, and M.A. Hofman, *Alterations in the hypothalamic paraventricular nucleus and its oxytocin neurons (putative satiety cells) in Prader-Willi syndrome: a study of five cases.* J Clin Endocrinol Metab, 1995. **80**(2): p. 573-9.
234. Yamaki, A., et al., *A novel nuclear localization signal in the human single-minded proteins SIM1 and SIM2.* Biochem Biophys Res Commun, 2004. **313**(3): p. 482-8.
235. Vinkemeier, U., et al., *The globular head domain of titin extends into the center of the sarcomeric M band. cDNA cloning, epitope mapping and immunoelectron microscopy of two titin-associated proteins.* J Cell Sci, 1993. **106 (Pt 1)**: p. 319-30.
236. Carlsson, E., et al., *Myofibrillar M-band proteins in rat skeletal muscles during development.* Histochemistry, 1990. **95**(1): p. 27-35.
237. Grove, B.K., et al., *Myomesin and M-protein: expression of two M-band proteins in pectoral muscle and heart during development.* J Cell Biol, 1985. **101**(4): p. 1413-21.
238. Steiner, F., K. Weber, and D.O. Furst, *Structure and expression of the gene encoding murine M-protein, a sarcomere-specific member of the immunoglobulin superfamily.* Genomics, 1998. **49**(1): p. 83-95.
239. Fritz-Six, K.L., et al., *Aberrant myofibril assembly in tropomodulin1 null mice leads to aborted heart development and embryonic lethality.* J Cell Biol, 2003. **163**(5): p. 1033-44.
240. Gokhin, D.S., et al., *Tropomodulin isoforms regulate thin filament pointed-end capping and skeletal muscle physiology.* J Cell Biol, 2010. **189**(1): p. 95-109.

241. Sussman, M.A., et al., *Tropomodulin is highly concentrated at the postsynaptic domain of human and rat neuromuscular junctions.* Exp Cell Res, 1993. **209**(2): p. 388-91.
242. Millecamps, S., et al., *Conditional NF-L transgene expression in mice for in vivo analysis of turnover and transport rate of neurofilaments.* J Neurosci, 2007. **27**(18): p. 4947-56.
243. Moisse, K., et al., *Cytosolic TDP-43 expression following axotomy is associated with caspase 3 activation in NFL-/- mice: support for a role for TDP-43 in the physiological response to neuronal injury.* Brain Res, 2009. **1296**: p. 176-86.
244. Huttelmaier, S., et al., *Spatial regulation of beta-actin translation by Src-dependent phosphorylation of ZBP1.* Nature, 2005. **438**(7067): p. 512-5.
245. Karakozova, M., et al., *Arginylation of beta-actin regulates actin cytoskeleton and cell motility.* Science, 2006. **313**(5784): p. 192-6.
246. Shestakova, E.A., et al., *Correlation of beta-actin messenger RNA localization with metastatic potential in rat adenocarcinoma cell lines.* Cancer Res, 1999. **59**(6): p. 1202-5.
247. Kim, S.Y., et al., *Spatiotemporal expression pattern of non-clustered protocadherin family members in the developing rat brain.* Neuroscience, 2007. **147**(4): p. 996-1021.
248. Yagi, T. and M. Takeichi, *Cadherin superfamily genes: functions, genomic organization, and neurologic diversity.* Genes Dev, 2000. **14**(10): p. 1169-80.
249. Baena-Lopez, L.A., A. Baonza, and A. Garcia-Bellido, *The orientation of cell divisions determines the shape of Drosophila organs.* Curr Biol, 2005. **15**(18): p. 1640-4.
250. Mao, Y., et al., *Characterization of a Dchs1 mutant mouse reveals requirements for Dchs1-Fat4 signaling during mammalian development.* Development, 2011. **138**(5): p. 947-57.
251. Rock, R., S. Schrauth, and M. Gessler, *Expression of mouse dchs1, fjx1, and fat-j suggests conservation of the planar cell polarity pathway identified in Drosophila.* Dev Dyn, 2005. **234**(3): p. 747-55.
252. Saburi, S., et al., *Loss of Fat4 disrupts PCP signaling and oriented cell division and leads to cystic kidney disease.* Nat Genet, 2008. **40**(8): p. 1010-5.
253. Asperti, C., et al., *Liprin-alpha1 promotes cell spreading on the extracellular matrix by affecting the distribution of activated integrins.* J Cell Sci, 2009. **122**(Pt 18): p. 3225-32.
254. Asperti, C., E. Pettinato, and I. de Curtis, *Liprin-alpha1 affects the distribution of low-affinity beta1 integrins and stabilizes their permanence at the cell surface.* Exp Cell Res, 2010. **316**(6): p. 915-26.
255. Serra-Pages, C., et al., *Liprins, a family of LAR transmembrane protein-tyrosine phosphatase-interacting proteins.* J Biol Chem, 1998. **273**(25): p. 15611-20.
256. Spangler, S.A. and C.C. Hoogenraad, *Liprin-alpha proteins: scaffold molecules for synapse maturation.* Biochem Soc Trans, 2007. **35**(Pt 5): p. 1278-82.
257. Ben-Zur, T., et al., *The mammalian Odz gene family: homologs of a Drosophila pair-rule gene with expression implying distinct yet overlapping developmental roles.* Dev Biol, 2000. **217**(1): p. 107-20.
258. Drabikowski, K., A. Trzebiatowska, and R. Chiquet-Ehrismann, *ten-1, an essential gene for germ cell development, epidermal morphogenesis, gonad migration, and neuronal pathfinding in Caenorhabditis elegans.* Dev Biol, 2005. **282**(1): p. 27-38.

259. Minet, A.D., et al., *Teneurin-1, a vertebrate homologue of the Drosophila pair-rule gene ten-m, is a neuronal protein with a novel type of heparin-binding domain.* J Cell Sci, 1999. **112 (Pt 12)**: p. 2019-32.
260. Oohashi, T., et al., *Mouse ten-m/Odz is a new family of dimeric type II transmembrane proteins expressed in many tissues.* J Cell Biol, 1999. **145**(3): p. 563-77.
261. Cain, J.E., et al., *Bone morphogenetic protein signaling in the developing kidney: present and future.* Differentiation, 2008. **76**(8): p. 831-42.
262. Stewart, A., H. Guan, and K. Yang, *BMP-3 promotes mesenchymal stem cell proliferation through the TGF-beta/activin signaling pathway.* J Cell Physiol, 2010. **223**(3): p. 658-66.
263. Takao, M., et al., *Identification of rat bone morphogenetic protein-3b (BMP-3b), a new member of BMP-3.* Biochem Biophys Res Commun, 1996. **219**(2): p. 656-62.
264. Burrows, J.F., et al., *DUB-3, a cytokine-inducible deubiquitinating enzyme that blocks proliferation.* J Biol Chem, 2004. **279**(14): p. 13993-4000.
265. Yamada, M., et al., *Mice lacking the M3 muscarinic acetylcholine receptor are hypophagic and lean.* Nature, 2001. **410**(6825): p. 207-12.
266. Caqueret, A., et al., *Looking for trouble: a search for developmental defects of the hypothalamus.* Horm Res, 2005. **64**(5): p. 222-30.
267. Szarek, E., et al., *Molecular genetics of the developing neuroendocrine hypothalamus.* Mol Cell Endocrinol, 2010. **323**(1): p. 115-23.
268. von Bohlen Und Halbach, O., *Immunohistological markers for staging neurogenesis in adult hippocampus.* Cell Tissue Res, 2007. **329**(3): p. 409-20.
269. Shimogori, T., et al., *A genomic atlas of mouse hypothalamic development.* Nat Neurosci, 2010. **13**(6): p. 767-75.
270. Valverde, F., et al., *Development of the mammillothalamic tract in normal and Pax-6 mutant mice.* J Comp Neurol, 2000. **419**(4): p. 485-504.
271. Tsuchiya, R., et al., *Aberrant axonal projections from mammillary bodies in Pax6 mutant mice: possible roles of Netrin-1 and Slit 2 in mammillary projections.* J Neurosci Res, 2009. **87**(7): p. 1620-33.
272. Osorio, J., et al., *Phylotypic expression of the bHLH genes Neurogenin2, Neurod, and Mash1 in the mouse embryonic forebrain.* J Comp Neurol, 2010. **518**(6): p. 851-71.
273. Nigg, E.A., *Nucleocytoplasmic transport: signals, mechanisms and regulation.* Nature, 1997. **386**(6627): p. 779-87.
274. Jans, D.A. and S. Hubner, *Regulation of protein transport to the nucleus: central role of phosphorylation.* Physiol Rev, 1996. **76**(3): p. 651-85.
275. Holmberg, C.I., et al., *Multisite phosphorylation provides sophisticated regulation of transcription factors.* Trends Biochem Sci, 2002. **27**(12): p. 619-27.
276. Wen, W., et al., *Identification of a signal for rapid export of proteins from the nucleus.* Cell, 1995. **82**(3): p. 463-73.
277. Kudo, N., et al., *Leptomycin B inhibition of signal-mediated nuclear export by direct binding to CRM1.* Exp Cell Res, 1998. **242**(2): p. 540-7.
278. Lipovich, L., R. Johnson, and C.Y. Lin, *MacroRNA underdogs in a microRNA world: evolutionary, regulatory, and biomedical significance of mammalian long non-protein-coding RNA.* Biochim Biophys Acta, 2010. **1799**(9): p. 597-615.
279. Khalil, A.M., et al., *Many human large intergenic noncoding RNAs associate with chromatin-modifying complexes and affect gene expression.* Proc Natl Acad Sci U S A, 2009. **106**(28): p. 11667-72.

280. Hung, T. and H.Y. Chang, *Long noncoding RNA in genome regulation: Prospects and mechanisms.* RNA Biol, 2010. **7**(5): p. 582-5.
281. Alvarez-Bolado, G., M.G. Rosenfeld, and L.W. Swanson, *Model of forebrain regionalization based on spatiotemporal patterns of POU-III homeobox gene expression, birthdates, and morphological features.* J Comp Neurol, 1995. **355**(2): p. 237-95.
282. Karim, M.A. and J.C. Sloper, *Histogenesis of the supraoptic and paraventricular neurosecretory cells of the mouse hypothalamus.* J Anat, 1980. **130**(Pt 2): p. 341-7.
283. Okamura, H., et al., *Time of vasopressin neuron origin in the mouse hypothalamus: examination by combined technique of immunocytochemistry and [3H]thymidine autoradiography.* Brain Res, 1983. **285**(2): p. 223-6.
284. Wataya, T., et al., *Minimization of exogenous signals in ES cell culture induces rostral hypothalamic differentiation.* Proc Natl Acad Sci U S A, 2008. **105**(33): p. 11796-801.
285. Stoykova, A. and P. Gruss, *Roles of Pax-genes in developing and adult brain as suggested by expression patterns.* J Neurosci, 1994. **14**(3 Pt 2): p. 1395-412.
286. Oliver, G., et al., *Six3, a murine homologue of the sine oculis gene, demarcates the most anterior border of the developing neural plate and is expressed during eye development.* Development, 1995. **121**(12): p. 4045-55.
287. Chen, K.F., et al., *Transcriptional repression of human cad gene by hypoxia inducible factor-1alpha.* Nucleic Acids Res, 2005. **33**(16): p. 5190-8.
288. To, K.K., et al., *The phosphorylation status of PAS-B distinguishes HIF-1alpha from HIF-2alpha in NBS1 repression.* Embo J, 2006. **25**(20): p. 4784-94.
289. Leung, T.H., A. Hoffmann, and D. Baltimore, *One nucleotide in a kappaB site can determine cofactor specificity for NF-kappaB dimers.* Cell, 2004. **118**(4): p. 453-64.
290. Geserick, C., H.A. Meyer, and B. Haendler, *The role of DNA response elements as allosteric modulators of steroid receptor function.* Mol Cell Endocrinol, 2005. **236**(1-2): p. 1-7.
291. Gaston, K. and P.S. Jayaraman, *Transcriptional repression in eukaryotes: repressors and repression mechanisms.* Cell Mol Life Sci, 2003. **60**(4): p. 721-41.
292. Lu, J., et al., *Regulation of skeletal myogenesis by association of the MEF2 transcription factor with class II histone deacetylases.* Mol Cell, 2000. **6**(2): p. 233-44.
293. Patel, Y.C., *Somatostatin and its receptor family.* Front Neuroendocrinol, 1999. **20**(3): p. 157-98.
294. Ruiz-Torres, P., et al., *A dual effect of somatostatin on the proliferation of cultured rat mesangial cells.* Biochem Biophys Res Commun, 1993. **195**(2): p. 1057-62.
295. Hatzoglou, A., et al., *Identification and characterization of opioid and somatostatin binding sites in the opossum kidney (OK) cell line and their effect on growth.* J Cell Biochem, 1996. **63**(4): p. 410-21.
296. Yamada, Y., et al., *Cloning and functional characterization of a family of human and mouse somatostatin receptors expressed in brain, gastrointestinal tract, and kidney.* Proc Natl Acad Sci U S A, 1992. **89**(1): p. 251-5.
297. Pradayrol, L., et al., *N-terminally extended somatostatin: the primary structure of somatostatin-28.* FEBS Lett, 1980. **109**(1): p. 55-8.

298. Kreienkamp, H.J., et al., *Somatostatin receptor subtype 1 modulates basal inhibition of growth hormone release in somatotrophs.* FEBS Lett, 1999. **462**(3): p. 464-6.
299. Zheng, H., et al., *Somatostatin receptor subtype 2 knockout mice are refractory to growth hormone-negative feedback on arcuate neurons.* Mol Endocrinol, 1997. **11**(11): p. 1709-17.
300. Rohrer, S.P., et al., *Rapid identification of subtype-selective agonists of the somatostatin receptor through combinatorial chemistry.* Science, 1998. **282**(5389): p. 737-40.
301. Hokfelt, T., et al., *Cellular localization of somatostatin in endocrine-like cells and neurons of the rat with special references to the A1-cells of the pancreatic islets and to the hypothalamus.* Acta Endocrinol Suppl (Copenh), 1975. **200**: p. 5-41.
302. Dun, S.L., et al., *Neuronostatin is co-expressed with somatostatin and mobilizes calcium in cultured rat hypothalamic neurons.* Neuroscience, 2010. **166**(2): p. 455-63.
303. Martinelli, C.E., et al., *Obesity due to melanocortin 4 receptor (MC4R) deficiency is associated with increased linear growth and final height, fasting hyperinsulinemia, and incompletely suppressed growth hormone secretion.* J Clin Endocrinol Metab, 2011. **96**(1): p. E181-8.
304. Martin, N.M., et al., *Abnormalities of the somatotrophic axis in the obese agouti mouse.* Int J Obes (Lond), 2006. **30**(3): p. 430-8.
305. Epelbaum, J., et al., *The neurobiology of somatostatin.* Crit Rev Neurobiol, 1994. **8**(1-2): p. 25-44.
306. Montminy, M.R., et al., *Identification of a cyclic-AMP-responsive element within the rat somatostatin gene.* Proc Natl Acad Sci U S A, 1986. **83**(18): p. 6682-6.
307. Montminy, M.R. and L.M. Bilezikjian, *Binding of a nuclear protein to the cyclic-AMP response element of the somatostatin gene.* Nature, 1987. **328**(6126): p. 175-8.
308. Tao, Y.X., *The melanocortin-4 receptor: physiology, pharmacology, and pathophysiology.* Endocr Rev, 2010. **31**(4): p. 506-43.
309. Sanchez-Munoz, I., et al., *Regulation of somatostatin gene expression by brain derived neurotrophic factor in fetal rat cerebrocortical cells.* Brain Res, 2011. **1375**: p. 28-40.