Investigating mechanisms and biological functions of the bHLH/PAS transcription factors SIM1 and ARNT2

THE UNIVERSITY of ADELAIDE

Alexis Georgia Gerassimou

B.Sc. (Advanced), Honours (Biochemistry)

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

Department of Molecular and Biomedical Science
School of Biological Sciences
University of Adelaide, Australia
July 2021

Contents

Contents **2**
Abstract **5**
Declaration **6**
Acknowledgments **7**
Publications **8**
Additional publications 8
Conference and Symposia Poster Presentations **8**
Chapter 1: Introduction **9**
bHLH/PAS transcription factors **9**
ARNT and ARNT2 ..... **10**
Arnt and *Arnt2* Expression patterns 11
Arnt and *Arnt2* knockout mouse models 11
Overlapping and distinct functions of ARNT and ARNT2 12
Single Minded 1 ..... **13**
Project aims **31**
Chapter 2: Materials and Methods **32**
Abbreviations **32**
Materials **34**
Antibodies 34
Bacterial strains 35
Tissue culture cell lines 35
Solutions 35
Plasmids 38
shRNA sequences 39
Primers 39
Methods **40**
Bacterial cell culture 40
DNA and RNA analysis 40
Mammalian cell culture 41
Protein analysis 44
Mouse work **47**
Mouse colonies 47
Genotyping methods 48
Tissue sectioning and analysis 49
Primary cell culture 50
Fluorescence activated cell sorting (FACS) 51

Method for Analysing RNA following Intracellular Sorting (MARIS) 51
Chapter 3: Generation of shRNA knockdown mouse models to investigate unique and overlapping physiological functions of ARNT and ARNT2 52
Introduction 52
shRNA mouse models 52
Arnt and *Arnt2* shRNA mouse generation 54
Results 57
shRNA mouse genotyping 57
Knockdown of ARNT in E11.5 Mouse Embryonic Fibroblasts 57
Knockdown of ARNT2 in E16.5 and E18.5 cortical neurons 60
Knockdown of ARNT and ARNT2 in E16.5 embryos 63
Knockdown of ARNT and ARNT2 in adult tissues 66
Discussion 68
Efficiency of shRNA expression and knockdown in mouse tissues and primary cell culture 68
Other approaches to generate knockdown mouse models. 71
Chapter 4: Investigating the mechanism of reduced activity of the human obesity linked SIM1 mutant R171H 73
Introduction 73
SIM1 R171H mouse model 73
GFP mouse model 74
Results 75
Confirmation of Sim1-GFP expression in adult brain and kidney 75
Sim1 R171H mouse model 75
Effect of R171H on SIM1 function 78
Isolation of SIM1 expressing cells to investigate biochemical mechanisms of SIM1 78
Discussion 82
Mechanism of R171H dysfunction in obesity 82
Isolation of SIM1 positive cells 83
Chapter 5: Expression and purification of SIM1/ARNT2 heterodimer 84
Introduction 84
Previous expression systems 85
Expi293F expression system 87
Results 88
Designing tagged SIM1 and ARNT2 expression plasmids for production and purification of SIM1/ARNT2 bHLH/PAS heterodimers from Expi293F cells 88
Strep-Tactin XT Superflow and anti-Flag M2 affinity gel resins 88
Transient vs stable expression of SIM1.1-348.Flag/Strep and ARNT2.1-455.Myc in Expi293F cells 88

Mid scale culture and expression of SIM1.1-348.Flag/Strep and ARNT2.1-455.Myc 89
Large scale culture and expression of SIM1.1-348.Flag/Strep and ARNT2.1-455.Myc 92
Optimising conditions for Strep and Flag purifications of SIM1.1-348.Flag/Strep and ARNT2.1-455.Myc 95
Discussion 97
Optimisation of SIM1 and ARNT2 expression and purification 97
Future optimisation of purification of SIM1 and ARNT2 98
Chapter 6: Final discussion and future perspectives 100
Appendix 102
Identification of acetyl CoA Carboxylase 1 (ACC1) 102
References 103

Abstract

The basic Helix-Loop-Helix/PER-ARNT-SIM (bHLH/PAS) family of transcription factors are involved in gene regulation during both embryonic development and throughout adulthood to maintain cellular homeostasis. Members of this family must heterodimerise to activate or repress target gene expression.

ARNT and ARNT2 are homologues that are common dimerisation partners for many other bHLH/PAS transcription factors. Previous attempts to generate knockout mice of ARNT and ARNT2 transcription factors have resulted in embryonic or perinatal lethality due to the important developmental roles of both factors. These studies have also indicated unique as well as overlapping functions of ARNT and ARNT2 during embryonic development. Due to the lethality of homozygous null mice, there is limited knowledge of unique and overlapping functions of these two proteins in adult tissues. The development of a doxycycline (Dox) inducible short hairpin RNA (shRNA) system is an attractive approach to deplete gene expression without irreversibly damaging a gene and can allow normal embryonic development to occur prior to studying effects of reduced protein levels in adult tissues. In this study, we evaluated the efficiency of Dox inducible shRNA sequences to result in global knockdown of ARNT and ARNT2 in a mouse model. While we were able to show mosaic shRNA knockdown in embryonic shRNA mice, we did not observe any knockdown of either ARNT or ARNT2 in adult mice.

Genetic mouse models have revealed one key partner protein of ARNT2 to be Single-Minded 1 (SIM1), a bHLH/PAS factor predominately expressed in the paraventricular nucleus (PVN) and supraoptic nucleus (SON) of the hypothalamus. SIM1 has been established to be essential in the developing hypothalamus and has a physiological role in appetite control during adulthood. Haploinsufficiency of *Sim1* in mice results in increased weight gain due to hyperphagia and increased linear growth, and non-synonymous point mutants of *SIM1* can be key drivers of obesity in humans. Several studies indicate that SIM1 plays an important role in the appetite controlling Leptin-Melanocortin signalling pathway in the hypothalamus, however the target genes of SIM1 and its up- and downstream regulatory pathways have yet to be defined. To further our understanding of the role of SIM1 in appetite control, we generated a mouse model harbouring a weakly functioning mutant of SIM1 to demonstrate that a single amino acid change in SIM1 could underpin monogenetic obesity. Composite with a transgene expressing GFP under the control of the *Sim1* promoter, this mouse model, in combination with cell based assays, was used to investigate mechanisms of dysfunction that result from SIM1 mutations. We were able to show that the mutation did not decrease SIM1 protein stability or the ability of SIM1 to dimerise with interacting proteins, nor did it have an apparent effect on PVN development.

There is currently a lack of structural data for the bHLH/PAS family of transcription factors, including SIM1. Using the mammalian expression system Expi293F cells, we developed a system to express and purify a truncated SIM1/ARNT2 heterodimer for use in structural studies. A SIM1 structure will allow further understanding into how SIM1 functions and will aid development of targeted drug design.

Declaration

I certify that this work contains no material which has been accepted for the award of any other degree or diploma in my name, in any university or other tertiary institution 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. In addition, I certify that no part of this work will, in the future, be used in a submission in my name, for any other degree or diploma in any university or other tertiary institution without the prior approval of the University of Adelaide and where applicable, any partner institution responsible for the joint-award of this degree.

I 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 Search and also through web search engines, unless permission has been granted by the University to restrict access for a period of time.

I acknowledge the support I have received for my research through the provision of an Australian Government Research Training Program Scholarship.

Alexis Georgia Gerassimou

July 2021

Acknowledgments

First and foremost I would like to thank my supervisor Murray for accepting me into the Whitelaw lab and being the best supervisor anyone could ask for. Murray, I will always be thankful for your support, knowledge and guidance during the course of my PhD. Your encouragement and enthusiasm have helped me get through the last six years with my sanity intact, especially when things weren't going our way experimentally. I believe I'm a better scientist because of this.

I would also like to thank my co-supervisor Dan for all the insightful comments and suggestions, as well as all the encouragement you have given me over the years, especially when Murray was away on his long sabbaticals.

A huge thank you to all past and present members of the Whitelaw lab. In particular, my sincere thanks to Adrienne, Dave and Veronica for all your assistance and advice with experiments and your endless patience in answering what seemed to be my never ending questions. Thank you to Emily for making me feel welcome when I first started and our bonding chats over whether SIM1 and SIM2 actually exist – I value the friendship we have developed over the years. Joe and Tim, thank you for all the interesting conversations and sneaky lab lunches – the two of you could drive me crazy at times with all your banter but I wouldn't have had it any other way.

Also a big thank you to members of the Peet lab, especially Ice, Nat, Jay, Cam and Josh for all the hilarious chats, icecream runs, Eurovision conversations and general getting up to no good.

Thank you to my fellow students and staff members within the School of Biological Sciences, particularly the Beard and Thomas labs, for their support and friendship, especially Byron, who has been in the same boat as me for just as long. Thank you to Lynn and Velta for all the sage advice over the years and always being there for a chat – it was very much appreciated.

A very big thank you to my family and friends for their support throughout the last six years, for taking an interest in what I have been studying, for letting me vent frustration when experiments weren't working and always being there for me when needed.

Finally a huge thank you to Mum, Dad and Nicole, who have waited patiently (most of the time!) for thesis submission day to arrive – you have been my biggest supporters and I could not have survived this journey without you. It has been a long and challenging road to my PhD, however thanks to the support of all those mentioned I have thoroughly enjoyed the challenges and opportunities this journey has presented.

Publications

Gerassimou AG, Whitelaw ML, Bersten DC, *SIM1: the maestro of hunger circuits*. [Manuscript format]

Additional publications

Blackburn, PR, Sullivan, AE, **Gerassimou, AG**, Kleinendorst, L, Bersten, DC, Cooiman, M, Harris, KG, Wierenga, KJ, Klee, EW, van Gerpen, JA, Ross, OA, van Haelst, MM, Whitelaw, ML, Caulfield, TR, Atwal, PS. *Functional analysis of SIM1 variant p.G715V in patients with obesity*. J Clin Endocrinol Metab. 2020; 105(1):dgz192

Conference and Symposia Poster Presentations

Gerassimou AG, Sullivan AE, Bersten DC, Whitelaw ML. *Investigating mechanisms and biological functions of Single-Minded 1.* (2016) ANZSCDB meeting

Gerassimou AG, Sullivan AE, Bersten DC, Whitelaw ML. *Investigating mechanisms and biological functions of Single-Minded 1.* (2016) EMBL conference

Gerassimou AG, Sullivan AE, Bersten DC, Whitelaw ML. *Investigating mechanisms and biological functions of the obesity related transcription factor Single-Minded 1.* (2017) ComBio

Gerassimou AG, Sullivan AE, Bersten DC, Whitelaw ML. *Investigating mechanism and biological functions of the obesity related transcription factor Single-Minded 1.* (2017) ANZSCDB meeting

Gerassimou AG, Sullivan AE, Bersten DC, Whitelaw ML. *Mechanisms of regulation of the transcription factor Single-Minded 1 in obesity.* (2018) Lorne Genome conference

Chapter 1: Introduction

bHLH/PAS transcription factors

Transcriptions factors belonging to the basic helix-loop-helix/Per-ARNT-SIM (bHLH/PAS) family are actively involved in the regulation of many essential biological processes, including embryonic development, appetite control and responses to environmental stress such as low oxygen or infiltration of xenobiotics [1-3]. Mouse knockout studies have shown many of these factors to be embryonic lethal, establishing their roles as essential for life.

The factors within this family are characterised by a conserved N-terminal domain consisting of a bHLH motif followed by two adjacent repeats of the PAS domain. The basic region of the bHLH motif is involved in DNA binding and the HLH domain provides a dimerisation interface between family members. The two PAS domains, PAS A and PAS B, act as a secondary dimerisation interface that determines partner choice and influences the strength of partner interaction, while also responding to signals and interacting with cofactors to mediate gene expression [1, 4, 5]. Despite diverging at the amino acid level, the 3D structure of the PAS domain is highly conserved across the PAS family members [3]. The bHLH and PAS domains are both required to form functional DNA binding complexes to mediate target gene regulation. The C-terminal region contains transcriptional activation and repression domains for target gene regulation as well as cofactor binding domains and has low sequence conservation across family members [5-7] (Figure 1A).

bHLH/PAS transcription factors can be divided into two classes. Class I factors tend to have specific expression patterns and are tightly regulated, only forming functional transcription factors when dimerised with Class II factors. Class I factors include Single Minded 1 and 2 (SIM1 and SIM2), Hypoxia Inducible Factor 1/2 alpha (HIF1/2α), Aryl Hydrocarbon Receptor (AhR) and Neuronal PAS factor 4 (NPAS4). Class II factors, which include AhR Nuclear Translocator 1 and 2 (ARNT and ARNT2), are broadly expressed generic binding partners for Class I factors while also being capable of forming homodimers, although there are no known functional roles for homodimers. A bHLH/PAS heterodimer can activate or repress target gene promoters through DNA binding, typically through the binding to an asymmetric E-box-like element containing a conserved CGTG core [1, 8] (Figure 1B). Despite having high structural conservation and ability to recognise very similar DNA sequences, the family members are able to regulate distinct target genes and are functionally non-redundant. It is poorly understood how these factors are able to differentiate between a variety of cofactors as well as unique target genes.

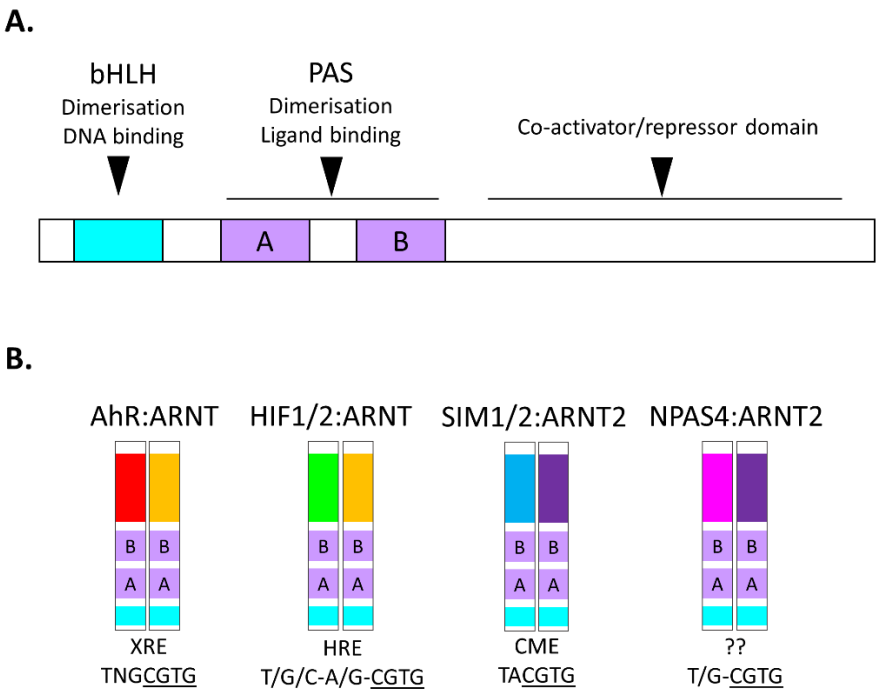

Figure 1. bHLH/PAS transcription factor family architecture and protein interactions.
A. Schematic of the general protein modules of bHLH/PAS family members. There is sequence similarity in the N-termini consisting of the bHLH and PAS domains, however there is low conservation in the C-terminal domain. B. Class I factors preferentially bind to a Class II factor in order to bind DNA and control gene transcription. Different heterodimers recognise variants of the E-box like sequence with a core CGTG to activate their target genes. XRE = xenobiotic response element, HRE = hypoxic response element, CME = central midline element.

ARNT and ARNT2

ARNT and its homologue ARNT2 are both Class II factors that are the generic binding partners for many Class I bHLH/PAS factors and as a result are involved in numerous essential biological pathways [1]. ARNT and ARNT2 have a high degree of sequence homology, sharing 63% identity across the whole protein, particularly in the bHLH/PAS domains which have 82% homology [9-12]. Both ARNT and ARNT2 have been shown to be equally capable of stably binding to the same set of Class I factors to activate response elements *in vitro*, however Class I factors seem to preferentially bind to either ARNT or ARNT2 to carry out their functional activities *in vivo*. For example, ARNT has been shown to be the preferred dimerisation partner for HIF1α and AhR in response to hypoxia and xenobiotics, respectively [13, 14]. ARNT2 on the other hand has been implicated to be the favoured partner for SIM1 in neuronal development and activity [15, 16]. This can be partially explained by the observed expression patterns of *Arnt* and *Arnt2*, with the expression of *Arnt* having a stronger overlap with *Hif1α* and *AhR* expression, while *Arnt2* and *Sim1* share similar expression patterns [17, 18]. Phenotypes observed for the knockout of *Arnt* or *Arnt2* which correlate with the knockout mice of other bHLH/PAS family members further supports Class I factors favouring either ARNT or ARNT2 *in vivo* [15, 16, 19-24].

Arnt and *Arnt2* Expression patterns

Mouse *Arnt* expression is ubiquitous throughout development and adulthood [9, 17, 20]. From as early as E9.0, stable *Arnt* mRNA has been predominately detected in tissues of mesoderm and endoderm origin with low levels of transcript being detected in ectodermal tissues. *Arnt* expression is notably weaker in the central nervous system (CNS), with expression decreasing in these areas during development as neurons reach their final differentiated states [9] .

Strong *Arnt2* mRNA expression has been noted to be predominately restricted to the CNS during embryonic development and is maintained throughout adulthood and is in a sense the reciprocal of *Arnt* expression [20]. Strong expression of *Arnt2* transcripts have also been detected in embryonic and adult kidney. At various stages of development, low expression of *Arnt2* has been detected in other tissues such as the retina, tongue, lung, stomach and intestines [9, 17].

Arnt and *Arnt2* knockout mouse models

Arnt knockout studies

Two independent research groups used resistance cassettes to disrupt the bHLH domain of the *Arnt* gene to generate *Arnt* homozygous knockout mice [21, 22]. Both studies found *Arnt* to be embryonic lethal with death occurring between E9.5 and E10.5 due to multiple development defects. These defects included defective angiogenesis of the yolk sac, stunted development, neural tube closure defects, forebrain hypoplasia, delayed rotation of the embryo and placental haemorrhaging [21, 22]. It has been postulated that the phenotype observed in these embryos is due to a decrease in HIF1α/ARNT heterodimer activity, with similar phenotypes seen in *Hif1α*$^{-/-}$ mice as well as mice that lack HIF1α/ARNT target genes [19, 24-27]. While *Arnt*$^{-/-}$ was embryonic lethal, *Arnt*$^{-/+}$ mice were found to be viable and didn't display any significant developmental defects.

An *Arnt* conditional knockout mouse using a Cre recombinase system targeting the bHLH domain has also been generated [28]. However, the knockout of *Arnt* in adult mice was observed to be incomplete with between 30-70% occurring in most tissues and more than 95% deletion only successful in the liver and thymus. Despite near complete deletion of ARNT in the liver resulting in a decrease in AhR and HIF1α target gene expression, there was still maximal target gene expression in other tissues that had high ARNT deletion.

Arnt2 knockout studies

As with *Arnt* knockout mice, a resistance cassette was used to disrupt the *Arnt2* gene in the bHLH domain [20], while a second research group inserted an inframe NLS-LacZ sequence into the middle of the basic domain [15]. Both studies reported that the heterozygous mice

were viable with no significant developmental defects, however the *Arnt2* null mice suffered perinatal lethality. Lethality of the null mice was attributed to impaired hypothalamic development of PVN and SON structures, which was characterised by the loss of hormone expression including corticotropin-releasing hormone (CRH), thyrotropin-releasing hormone (TRH), somatostatin (SS), vasopressin (AVP) and oxytocin (OXT). This phenotype is similar to that seen in *Sim1* knockout mouse studies in which these cell lineages were lost, indicating ARNT2 as the preferred binding partner of SIM1.

Compound Arnt and Arnt2 knockout studies

Compound *Arnt/Arnt2* knockout mice have been generated by the crossing of *Arnt* and *Arnt2* heterozygotes with the resultant $Arnt^{-/-}Arnt2^{-/-}$, $Arnt^{-/+}Arnt2^{-/-}$ and $Arnt^{-/-}Arnt2^{-/+}$ embryos not surviving beyond E8.5 [20]. This indicates that both proteins are capable of compensating for each other for any essential roles up until this developmental time point, but lethality beyond E8.5 suggests unique roles during later development. This indicates that only one allele of either *Arnt* or *Arnt2* is insufficient to maintain expression of essential genes beyond this developmental time point. However, embryos null for one of *Arnt* or *Arnt2*, while wild type for the other, survive after E8.5, suggesting either protein is capable of temporarily compensating for the other after this time point.

Overlapping and distinct functions of ARNT and ARNT2

Despite a high degree of sequence homology between ARNT and ARNT2, mRNA expression analysis and knockout studies in mice have indicated that both proteins have unique as well as overlapping functions.

$Arnt^{-/-}$ embryos have arrested development earlier and have a more severe phenotype than $Arnt2^{-/-}$ embryos. This indicates that ARNT2 is not capable of fully compensating for the loss of ARNT, implicating unique functions for the two proteins. This is also supported by the similar phenotypes observed between ARNT and ARNT2 with other bHLH/PAS protein knockout models. The $Arnt2^{-/-}$ phenotype reflects that seen in $Sim1^{-/-}$ mice, with both knockouts being perinatal lethal presumably due to hypocellular PVN and SON structures observed within the hypothalamus that are lacking essential cell lineages [15, 20, 23]. $Arnt^{-/-}$ mice display a similar phenotype to *Hif1α* knockout mice, with embryos from both knockouts having poor vascularisation and neural tube defects [19, 20, 24, 25]. However, it has been noted that a more severe phenotype resulting in earlier embryonic death is observed for the $Hif1\alpha^{-/-}$ mice compared to $Arnt^{-/-}$ mice. As ARNT2 is capable of forming a functional transcriptional factor with HIF1α, it has been reasoned that ARNT2 is able to compensate for the loss of ARNT to a limited degree in these mice, resulting in more advanced development compared to the $Hif1\alpha^{-/-}$ mice. Keith et al. [20] showed that ARNT/HIF1α and ARNT2/HIF1α heterodimers are both present and capable of inducing hypoxic response genes in cortical neurons. They found that cortical neurons derived from E14.5 $Arnt2^{-/-}$ embryos had slower, reduced induction of HIF1 target genes compared to wild type cells, indicating that the ARNT present is sufficient to maintain target gene expression to a limited degree in the absence of ARNT2.

Single Minded 1

Statement of Authorship

Title of Paper	SIM1: the maestro of hunger circuits
Publication Status	☐ Published ☐ Accepted for Publication ☐ Submitted for Publication ☑ Unpublished and Unsubmitted work written in manuscript style
Publication Details	N/A

Principal Author

Name of Principal Author (Candidate)	Alexis G. Gerassimou		
Contribution to the Paper	Wrote article. Helped with article concepts, structure and editing.		
Overall percentage (%)	70%		
Certification:	This paper reports on original research I conducted during the period of my Higher Degree by Research candidature and is not subject to any obligations or contractual agreements with a third party that would constrain its inclusion in this thesis. I am the primary author of this paper.		
Signature		Date	19/7/21

Co-Author Contributions

By signing the Statement of Authorship, each author certifies that:

i. the candidate's stated contribution to the publication is accurate (as detailed above);

ii. permission is granted for the candidate in include the publication in the thesis; and

iii. the sum of all co-author contributions is equal to 100% less the candidate's stated contribution.

Name of Co-Author	Murray L. Whitelaw		
Contribution to the Paper	Helped with article concepts, structure and editing.		
Signature		Date	19/07/21

Name of Co-Author	David C. Bersten		
Contribution to the Paper	Helped with article concepts, structure and editing.		
Signature		Date	24/07/21

SIM1: the maestro of hunger circuits

Alexis G. Gerassimou, Murray L. Whitelaw, David C. Bersten

Department of Molecular and Biomedical Science, University of Adelaide, Adelaide 5005, South Australia, Australia

Abstract

The hypothalamus is a highly evolved ancient brain region that regulates organismal homeostasis pathways essential for survival, including multiple signalling circuits that converge to regulate satiety and energy homeostasis. Recent studies have highlighted the growing number of specific cell populations and complexity of the neuronal networks that are involved in appetite regulation, where imbalances can result in obesity or starvation. Within the hypothalamus, the expression of the Single Minded 1 (SIM1) transcription factor has been shown to be essential for the correct development of the paraventricular nucleus (PVN), and consequently *Sim1* is a definitive marker of PNV neurons. In addition to its developmental role, *Sim1* plays a key role in postnatal PVN function and obesity, with *Sim1* PVN neurons being critically important in satiety circuitry. SIM1 variants have been identified in obese human patients, while mouse model studies have demonstrated that varying *Sim1* levels and/or activities can influence appetite regulation within the leptin-melanocortin signalling pathway. As such, SIM1 is emerging as an attractive target for obesity related therapeutics. This review will discuss recent advances in the study of SIM1 and *Sim1* neurons of the hypothalamus and their relevance to appetite regulation.

Introduction

The hypothalamus is central to maintaining several aspects of body homeostasis, with distinct sets of compartmentalised neurons controlling pathways that regulate circadian rhythm, fluid and electrolyte balance, thirst, energy stores and body temperature. Lesions or hormone imbalances in its neuronal networks can result in either hyperphagia induced obesity or starvation, establishing the hypothalamus as the central overseer of appetite [1-7]. Within the hypothalamus, the PVN integrates signals from other hypothalamic nuclei, in particular two molecularly distinct pro-opiomelanocortin (POMC) and agouti-related peptide (AgRP) neuronal populations in the arcuate nucleus (ARC), to regulate feeding behaviour to adapt to changing energy demands [8, 9].

Over the last few decades, there has been an increase in the number of studies in both humans and mice linking obesity with abnormal PVN development and function. Recent single cell sequencing experiments have revealed the complexity of the neuronal circuits that make up and control PVN function in satiety signalling and energy homeostasis [10-14], further improving our understanding of the genes involved in obesity. In particular there has

been an increase in the evidence outlining a key role for SIM1, a developmental transcription factor that specifies PVN neurons, in appetite control and obesity. Optogenetic studies have shown that *Sim1* neurons have the greatest effect on satiety while loss of function mutations in SIM1 have been associated with early onset obesity. Together, these make targeting *SIM1* neurons and/or the *SIM1* gene attractive therapeutic approaches for treating appetite disorders. In light of these recent advances, we will review the role of *Sim1* neurons, and the SIM1 protein, in the hypothalamus in regards to regulation of satiety signalling.

SIM1 and PVN circuitry in satiety signalling

Targeted deletion of *Sim1* in mice results in perinatal *Sim1* $^{-/-}$ lethality, attributed to severely hypocellular PVN and supraoptic (SON) structures within the hypothalamus, thus establishing an essential, non-redundant role for *Sim1* in the developing hypothalamus [15]. Multiple neurosecretory cell lineages were found to be missing in these structures, including those identified by the expression of Corticotrophin Releasing Hormone (CRH), Thyrotropin Releasing Hormone (TRH), Somatostatin (SS), Arginine Vasopressin (AVP), and Oxytocin (OXT). These studies revealed that *Sim1* is required for the proper migration and terminal differentiation of PVN and SON neural cell lineages, and as *Sim1* is expressed in almost all mature PVN neurons, it serves experimentally as the definitive PVN marker gene [15, 16]. The role of *Sim1* throughout the development of the hypothalamus has been reviewed elsewhere [17, 18].

The generation of independent mouse models that are haploinsufficient for *Sim1 (Sim1*$^{-/+}$*)* resulted in mice with severe weight gain due to hyperphagia, without any observed change in energy expenditure (Table 1). These mice also displayed increased linear growth and elevated blood insulin and leptin levels compared to wild type (WT) littermates [19-21]. While WT mice reduced their food intake when switched from a low fat to a high fat diet to compensate for the change in caloric intake, *Sim1*$^{-/+}$ mice did not change their food intake, resulting in more dramatic weight gain when on a high fat diet [20, 21]. Overall, the results of these studies indicated that the underlying cause of obesity in the *Sim1*$^{-/+}$ mice was increased eating without influencing energy expenditure.

The first reported case of *SIM1* haploinsufficiency causing severe obesity in humans was in a young girl with early onset obesity [28]. The girl was found to have a balanced chromosome translocation that resulted in a null *SIM1* allele. The haploinsufficiency of *SIM1* as a result of this translocation was hypothesised to be responsible for excessive food intake resulting in severe obesity as energy expenditure was found to be normal. The weight gain observed was found to be similar to young girls who had leptin, leptin receptor or melanocortin 4 receptor (Mc4r) mutations. Discovery of mutations in the genes for leptin and subsequently the leptin receptor were also originally made in severely obese mice (*ob/ob* and *db/db*, respectively) [3, 7], with administration of leptin to leptin deficient mice and humans rescuing the obesity phenotype [4-6, 29, 30]. This in turn led to the initial identification and characterisation of

Table 1. Summary of *Sim1* mouse models.

Mouse model	*Sim1*	Phenotype	Reference
Heterozygous	Disrupted *Sim1*	Obesity	[19]
Heterozygous	Disrupted *Sim1*	Obesity	[20, 21]
Modulation	Transgenic or viral mediated shRNA knockdown (k/d) or overexpression (o/e)	k/d = obesity o/e = rescue obesity	[22]
Conditional	Cre conditional knockout (Started either shortly after birth or P15)	Obesity	[23]
Conditional	Cre conditional knockout (Tamoxifen induced Cre in adult)	Obesity	[24]
Agouti yellow	Overexpression of human *SIM1*	Rescue obesity	[25]
M136K	Mutant identified in dominant screening of obese mice	Obesity	[26]
CRISPR activation	CRISPR activation of *Sim1* in heterozygous mice	Rescue obesity	[27]

the neuronal signalling pathways involved in energy homeostasis, in particular the leptin-melanocortin pathway which is a central energy homeostasis pathway initiated by leptin signalling in order to regulate hunger and satiety. In the leptin-melanocortin pathway, nutritional cues lead to changes in multiple neurotransmitter and hormone levels, including leptin, which act on POMC and AgRP neurons within the ARC nucleus [31-36]. Upon binding POMC neurons, leptin initiates the expression of POMC and its conversion to α-melanocortin stimulating hormone (α-MSH), which is transmitted to specific cells within the PVN, where it activates the G-coupled protein receptor Mc4r to ultimately reduce food intake. Conversely, binding of leptin to AgRP neurons suppresses expression and release of the hunger inducing peptide, AgRP. In the absence of leptin, AgRP neurons secrete AgRP into the PVN where it antagonises Mc4r, resulting in increased food intake [21, 37-41].

Loss of function mutations in leptin, leptin receptor, POMC and Mc4r have all been associated with hyperphagia induced obesity in both mice and humans [7, 42-51]. In contrast, deletion of the AgRP gene in mice causes starvation due to there no longer being AgRP present to oppose the action of α-MSH on the Mc4r in PVN neurons [52, 53]. Agouti yellow (A^y) mice that express high levels of agouti (an AgRP equivalent) have deficient Mc4r signalling, also resulting in obesity due to hyperphagia [25, 54]. Together, these results highlight that the leptin-melanocortin pathway is central in maintaining energy homeostasis and has been extensively reviewed in detail elsewhere [8, 9, 55, 56].
Over the years, a number of studies in mice have shown that optogenetic and chemogenetic stimulation of POMC and AgRP neurons causes a decrease or increase in food intake respectively, with stimulation of AgRP neuron terminals in the PVN also increasing food intake [57-61]. In particular, activation of *Sim1* positive PVN neurons, which make up the vast majority of PVN neurons that are downstream of the POMC and AgRP neurons, resulted in suppressed appetite, while inhibition of these *Sim1* neurons increased food intake [58].

Sim1 neuronal populations

Sim1 neurons of the PVN can be subdivided into populations demarcated by co-expression of distinct genes. For example, not all *Sim1* neurons express Mc4r, and a number of linages can be identified by specific neuropeptides (eg. SS, OXT, AVP, CRH and TRH). A major question is whether or not defined subsets of PVN ^{Sim1} neurons play crucial roles in appetite control.

Sim1 ^{Mc4r} *neurons*

A number of studies have re-expressed Mc4r specifically in *Sim1* positive neurons in Mc4r null mice [51, 62, 63]. Mc4r re-expression in *Sim1* positive neurons only dramatically rescued the obesity phenotype by abolishing hyperphagia. These results indicate that *Sim1*^{Mc4r} PVN neurons are sufficient for controlling feeding behaviour but not required for controlling energy expenditure. Further investigation using ChR2-assisted circuit mapping (CRACM) to identify neurons that were downstream of the PVN *Sim1*^{Mc4r} neurons showed that the *Sim1*^{Mc4r} neurons project to the lateral parabrachial nucleus (LPBN), a known feeding regulation area of the brain [63] (Figure 1). The role of Mc4r neurons in appetite regulation has been further supported with a recent study using a transgenic mouse line which uses a reporter expressed in Mc4r PVN neurons to detect Ca^{2+} levels, highlighting neuronal activity. Activation of Mc4r neurons, determined by an increase in Ca^{2+} levels, was observed in response to feeding, while Ca^{2+} levels and therefore activity of Mc4r neurons was decreased with fasting [64].

While current evidence indicates that SIM1 is a critical regulator in the leptin-melanocortin pathway downstream of Mc4r and that *Sim1*^{Mc4r} neurons are of major importance for regulating food intake, hyperphagia induced by the ablation of Mc4r only accounts for half of that seen compared to complete deletion of *Sim1*.

Sim1^{PDYN} *neurons*

A recent study has identified a *Sim1*:prodynorphin (PDYN) subpopulation of neurons, which lack *Mc4r*, as making up the other half of *Sim1* PVN neurons that are involved in regulating hunger and satiety [65]. *Sim1*^{PDYN} neurons were found to be anatomically distinct and functionally independently from *Sim1*^{Mc4r} neurons. Despite their functional independence, *Sim1*^{Mc4r} and *Sim1*^{PDYN} neurons have additive effects on satiety regulation. Chemogenetic inhibition of *Sim1*^{PDYN} neurons caused hyperphagia in mice to a similar degree as that seen when Mc4r neurons are inhibited, while activation of PDYN neurons resulted in decreased food intake in a similar manner to Mc4r neuronal activation. When both Mc4r and PDYN neurons were co-inhibited at the same time, the additive hyperphagia observed was comparable to that seen when all *Sim1* neurons are inactivated. Like *Sim1*^{Mc4r} neurons, *Sim1*^{PDYN} neurons were found to also receive regulatory input from AgRP neurons in the ARC

nucleus, which are themselves regulated by caloric state. As with *Sim1*^{Mc4r} neurons, tracing of *Sim1*^{PDYN} neuronal projections showed that they connect to the parabrachial complex (PB). However where *Sim1*^{Mc4r} neurons project to the LPBN, *Sim1*^{PDYN} neurons were found to predominately project to the pre-locus coeruleus (pLC), which in this study was identified as a novel satiety regulating node within the PB. Optogenetic stimulation of PDYN neuronal terminals in the PB was found to reduce food intake while inhibition in the same terminals promoted increased food intake, similar to what is seen in stimulation and inhibition of Mc4r neurons. These results provide evidence that in addition to *Sim1*^{Mc4r} neurons being crucial for regulating food intake, *Sim1*^{PDYN} neurons are also of major importance for the same outcomes. These two *Sim1* subpopulations work in distinct but parallel pathways from each other, having an additive effect in controlling appetite (Figure 1).

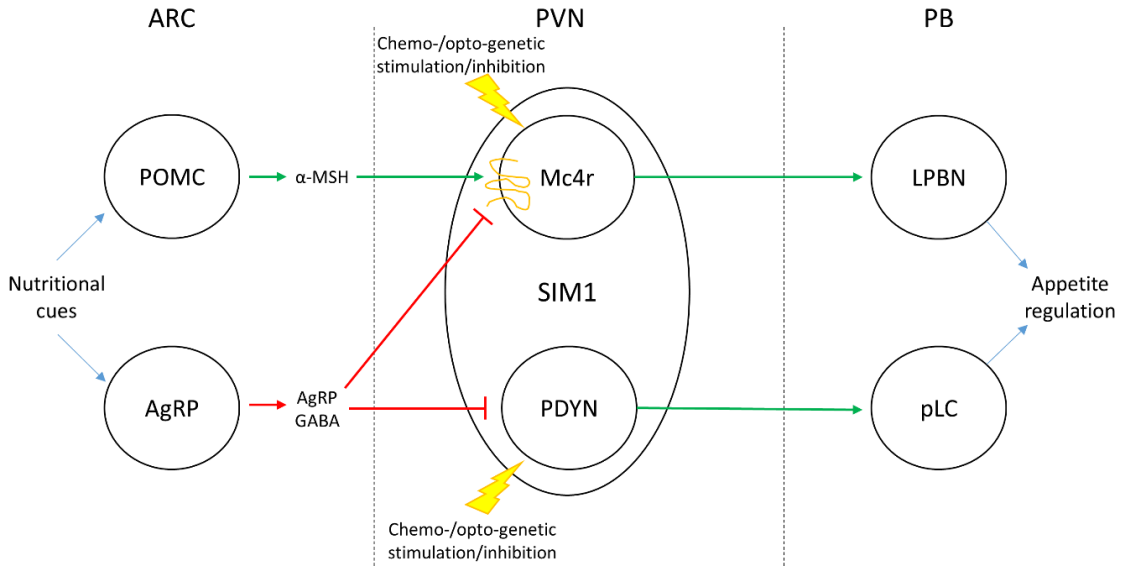

Figure 1. SIM1 neuronal pathway in feeding regulation. Nutritional cues regulate POMC and AgRP activity in the ARC, with activation of POMC neurons resulting in α-MSH being released into the PVN where it simulates Mc4r on *Sim1*^{Mc4r} neurons, leading to projections being made to the LPBN and ultimately reduced food intake. Activation of AgRP neurons antagonises both *Sim1*^{Mc4r} and *Sim1*^{PDYN} neurons in the PVN, reducing signalling to the PB and resulting in increased feeding. The Mc4r and PDYN *Sim1* neurons are distinct subpopulations that work together in parallel to regulate appetite. Opto- and chemogenetic stimulation of either Mc4r or PDYN *Sim1* neuronal subpopulations results in decreased food intake, while activation leads to increased food intake.

Sim1 neurons and Oxt

Sim1 has been implicated in the regulation *Oxt* as part of its function during appetite control. *Sim1* and *Oxt* are both predominately expressed in the PVN and SON, with *Oxt* shown to be co-localised in a subset of *Sim1* neurons [15, 66]. *Oxt* appears to have a role in appetite regulation, with feeding state modulating *Oxt* levels which in turn results in changes in food intake [66]. In addition, administration of *Oxt* decreased both food intake and weight gain in

Sim1$^{-/+}$ mice but not in WT mice, further supporting a connection between *Sim1* and *Oxt* in appetite regulation [66]. However, optogenetic studies from different groups have shown conflicting results in terms of a linked *Sim1* and *Oxt* pathway in the hypothalamus, with one group showing that PVN^{Oxt} neurons are post-synaptic targets of ARC^{AgRP} neurons [58], while another showed few very or no connections between the two neuronal populations [67].

In summary, results of the above studies establish that the ARC and PVN contain mixed populations of distinct neuron identities, highlighting the growing complexity of satiety signalling within the hypothalamus and illustrating that there is still much to uncover about the circuitry involved in appetite regulation.

Postnatal SIM1 activity is a key regulator of appetite

As *Sim1* is required for the correct migration and differentiation of neuronal populations within the PVN, dysregulation of food intake in *Sim1*$^{-/+}$ mice could be due to either a reduced number of *Sim1* neurons in the hypothalamus, or impaired postnatal physiological function of SIM1. This issue has been debated. One research group reported a decrease in the number of *Sim1* positive neurons in *Sim1*$^{-/+}$ mice compared to WT controls [15, 19, 68], while the other group claimed to see no significant changes in cell number [20, 21, 66]. The latter group favoured a model whereby hyperphagia seen in *Sim1* deleted mice resulted from decreased SIM1 regulation of target genes that modify food intake.

Modulation of Sim1 dosage

The hypothesis for SIM1 performing physiological functions in mature neurons was addressed by studies that modulated *Sim1* levels in adult mice post-development of the hypothalamus. Yang et al [22] manipulated levels of *Sim1* mRNA using adenoviral vectors to either deliver a shRNA to knockdown *Sim1* or a *Sim1* overexpression vector, in the PVN of adult WT mice. shRNA knockdown of *Sim1* resulted in increased food intake, while overexpression of *Sim1* reduced food intake. Due to limitations in viral delivery to the brain, a decrease in *Sim1* levels occurred in a limited number of PVN cells, however the limited reduction was sufficient to result in diet induced obesity.

As viral mediated modulation of *Sim1* levels was not complete throughout the PVN and homozygous knockout of *Sim1* is lethal due to a hypocellular hypothalamus, different PVN specific conditional postnatal knockouts of *Sim1* have been generated. These used different Cre systems to knockout *Sim1* soon after birth or in adult mice [23, 24] (Table 1). In these studies the conditional *Sim1*$^{-/-}$ and *Sim1*$^{-/+}$ mice phenocopied germline *Sim1*$^{-/+}$ mice, stimulating onset of hyperphagic obesity without any effect on energy expenditure, together with increasing linear growth compared to littermate controls. As with germline *Sim1*$^{-/+}$ mice, weight gain was greater when mice were placed on a high fat diet compared to a low fat diet. Informatively, the comparison of conditional *Sim1*$^{-/+}$ and *Sim1*$^{-/-}$ mice revealed a

dose-dependent effect of *Sim1* on food intake, with a higher increase of food intake and weight gain in *Sim1*$^{-/-}$ conditional knockouts compared to *Sim1*$^{-/+}$ mice.

As the mice used in these studies had mature adult hypothalamic structures, the loss of *Sim1* favours the hypothesis that SIM1 itself plays a distinct, postnatal role in appetite regulation and the obesity phenotype observed is not solely a consequence of poor hypothalamic development due to a lack of *Sim1* during development. Indeed, the more recent observation that *Sim1*$^{-/+}$ haploinsufficient mice induced hyperphagia and obesity can be reversed by postnatal restoration of *Sim1* mRNA levels by CRISPR activation (CRISPRa) in PVN neurons implies that signalling rather than loss of cells is responsible for obesity [27].

Sim1 and Mc4r

A number of lines of evidence have shown the important relationship between *Sim1* and Mc4r in the leptin-melanocortin pathway. Mc4r is a G-coupled protein receptor that can be regulated by either its agonist α-MSH or antagonist AgRP [37]. Recent structural studies have shown that Mc4r requires Ca^{2+} as a cofactor for activation by α-MSH while Ca^{2+} is not required for inhibition by AgRP, consistent with different conformational changes being observed upon binding of either the agonist or antagonist. In addition, Ca^{2+} levels appeared to have a dosage effect on Mc4r activity, with increased Ca^{2+} resulting in increased affinity and potency of α-MSH [69, 70]. Injection of either leptin or the synthetic analogue of α-MSH, melanoton-2 (MTII), are able to induce POMC expression and α-MSH production to activate Mc4r signalling, resulting in increased *Sim1* mRNA and reduced food intake in WT mice [21]. Mice heterozygous for *Sim1* were found to be resistant to both leptin and MTII signalling, with injection of either compound unable to supress feeding to the same extent as that seen in WT controls. Normal energy expenditure was observed in *Sim1* heterozygous mice in response to leptin or MTII injection, however the inability of the compounds to reduce feeding indicated that the signalling pathway for controlling food intake cannot be relayed past *Sim1* and that *Sim1* is not required for controlling energy expenditure [21]. In addition, the obesity phenotype observed in A^{y} mice, due to aberrant Mc4r signalling, was found to be partially rescued by reduced food intake upon the overexpression of human *Sim1* [25]. The ability of *Sim1* to override a lack of Mc4r signalling in the A^{y} mice indicates that *Sim1* is a downstream effector of Mc4r in the leptin-melanocortin pathway. This is further supported by chemogenetic suppression of *Sim1* neurons, which was found to increase food intake in a similar manner to activation of AgRP neurons [58]. In the same study, experiments that co-stimulated *Sim1* and AgRP neurons showed that activation of *Sim1* neurons was able to suppress feeding, overcoming AgRP induced feeding [58].

Sim1 and Oxt

There is conflicting data as to whether *Oxt* is a bona fide target of *Sim1* or is indirectly regulated by the functions of *Sim1*. *Oxt* levels are not affected by the injection of the Mc4r

agonist MTII, suggesting that *Oxt* is not downstream of activation of Mc4r and its resultant upregulation of *Sim1* [66]. However, studies have shown decreased levels of *Oxt* in *Sim1*$^{-/+}$ mice compared to WT controls [66, 68]. While one of these studies suggested that *Oxt* is somehow a mediator of the hyperphagic obesity seen in the *Sim1*$^{-/+}$ mice [66], the other study reasons that the change is due to a decrease in *Oxt* neurons as a result of an under developed PVN rather than direct transcriptional control by *Sim1* [68]. Conditional knockout of *Sim1* did not result in a change in the number of *Oxt* neurons within the hypothalamus, however a decrease in *Oxt* levels was also observed to be correlated to decreased levels of *Sim1* upon conditional knockout [23]. Although this result indicates that changes in a leptin-melanocortin-oxytocin pathway may be responsible for hyperphagic obesity, it was unclear if Mc4r, SIM1 and OXT are acting in the same linear pathway. Controversially, others reported that there is no change in the number of *Oxt* cells nor a change in *Oxt* levels in *Sim1*$^{-/+}$ mice [21], suggesting that SIM1 is not responsible for *Oxt* regulation in the PVN. This is further supported by the lack of conserved SIM1 binding sites near the transcription start site of the *Oxt* gene, reducing the possibility that SIM1 acts directly as a transcription factor on the *Oxt* promoter [66]. Despite the inconclusive evidence, there seems to be a missing link between *Sim1* and *Oxt* which appears to have an underlying role in appetite control, which needs to be thoroughly investigated by more in depth studies.

SIM1 as a monogenetic cause of obesity

A mechanistic role for SIM1 in the leptin-melanocortin pathway is consistent with the discovery of many non-synonymous mutants found in early onset obesity patients (Table 1, Figure 2A), where varying degrees of deficiency have been tracked with physicochemical properties and domain location of the amino acid changes. SIM1 is a member of the basic Helix-Loop-Helix/PER-ARNT-SIM (bHLH/PAS) transcription factor family. bHLH/PAS transcription factors are involved in many essential biological processes, including embryonic development and responses to various forms of environmental stress [71]. The family is characterised by an N-terminal architecture comprising a bHLH motif followed by two PAS domains, which are involved in dimerisation and DNA binding. The C-terminal halves of the family have low sequence conservation and contain transcriptional activation and repression domains [71-74]. Family members heterodimerise to regulate target gene expression through activation or repression and are divided into one of two classes. Class I factors tend to have specific expression patterns, are tightly regulated by environmental cues and only form a functional transcription factor when dimerised with a Class II factor, while Class II factors are generally broadly expressed and are generic binding partners for Class I factors.

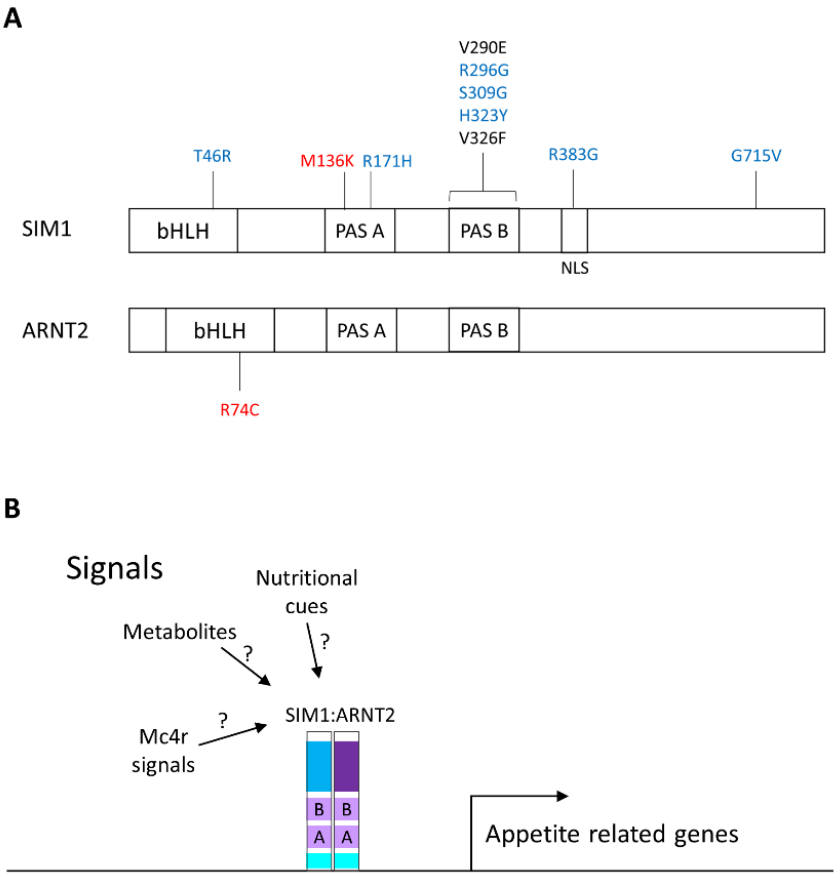

Figure 2. Position of mutations in SIM1 that result in decreased activity. A. Schematic of SIM1 and ARNT2, showing the position of mutations identified in humans (blue), mouse (red) or genome database (black). The majority of SIM1 mutations occur in the bHLH and PAS A and B repeats that form the dimerisation interface with the partner protein ARNT2. **B.** ARNT2 is the preferred heterodimerisation partner of SIM1 *in vivo*. Potential mechanisms of signal regulation of the SIM1/ARNT2 heterodimer to induce target gene activation including signals from the leptin-melanocortin pathway via Mc4r, circulating metabolites or nutritional cues.

SIM1 and ARNT2

In order for SIM1 to function, it must heterodimerise with a Class II factor. While SIM1 is capable of forming functional heterodimers with both the Class II factors aryl hydrocarbon nuclear receptor 1 and 2 (ARNT and ARNT2) *in vitro* [75, 76], studies have implicated ARNT2 as the preferential binding partner of SIM1 *in vivo* (Figure 2B). This has been indicated by *Sim1* and *Arnt2* having overlapping expression patterns within the central nervous system (CNS) [75] and homozygous knockout mouse studies of both factors have revealed similar phenotypes, namely perinatal lethality due to impaired hypothalamic development of the PVN and SON [15, 77, 78]. In contrast, *Arnt* expression is ubiquitous, with lower levels detected in cells of the CNS where *Sim1* is predominately expressed [79, 80]. The phenotype of knockout *Arnt* mice is also more severe than that for *Arnt2* and *Sim1* with multiple developmental defects causing embryonic death, reflecting what is seen for knockout of

another bHLH/PAS family member, *HIF1α* [81, 82]. This further indicates that ARNT2 is the preferred *in vivo* dimerisation partner for SIM1. A recent study has also shown an ARNT2 mutation, R74C, to cause obesity in mice, phenocopying *Sim1* mutant obesity mouse models and further supporting SIM1 and ARNT2 heterodimerisation preference *in vivo* [83].

Screening of patient cohorts with severe obesity has identified a number of *SIM1* variants (Table 2) that have been found to exhibit reduced function on reporter gene assays *in vitro* [84-87]. The majority of the variants tested in these studies had mutations in critical domains of the SIM1 protein, in particular in the dimerisation interface, with many variants having been shown to have impaired dimerisation with its partner protein ARNT2.

Table 2. Summary of human SIM1 variants that potentially contribute to obesity.

Variant	Number of known carriers	Human phenotype	Domain location of variant	Patient or genome database	Reference
Translocation	1	Obesity	Split between bHLH and PAS	Patient	[28]
T46R	7	Obesity	bHLH	Patient	[86]
S71R	1	Obesity		Patient	[85]
Q152E	2	Obesity	PAS A	Patient	[85]
R171H	3	Obesity	PAS A	Patient	[85, 86]
L238R	3	Obesity	PAS B	Patient	[85]
V290E	1	Unknown	PAS B	Genome	[86]
R296G	1	Obesity	PAS B	Patient	[88]
S309G	1	Obesity	PAS B	Patient	[88]
H323Y	2	Obesity	PAS B	Patient	[86]
V326F	1	Unknown	PAS B	Genome	[86]
R383G	1	Obesity	NLS	Patient	[85]
R550H	4	Obesity	C-terminal	Patient	[85]
P692L	4	Obesity	C-terminal	Patient	[85]
D707H	9	Obesity	C-terminal	Patient	[85]
T712I	2	Obesity	C-terminal	Patient	[85]
G715V	2	Obesity	C-terminal	Patient	[87]

In addition, a recent study performed dominate screening of mutagenic C57/BL6 mouse substrains for inheritance of obesity, identifying a mutation within the PAS A domain of SIM1, M136K [26] (Table). This M136K variant was unable to activate a reporter assay for SIM1, even in a dose dependent manner, indicating it represents a null allele of SIM1. Breeding of this mouse line resulted in inherited early onset obesity, with M136K mice heavier than WT mice at 6 weeks of age, due to greater food intake compared to WT mice, further supporting SIM1 haploinsufficiency as a driver of hyperphagic obesity [26]. Further, the R171H mutation, which was identified in humans and has been shown to have a 70-80% decrease in luciferase activity [85, 86], results in severe obesity when introduced into a mouse model (unpublished data).

Together with the results of the studies where *Sim1* levels are manipulated, there is a clear indication that additional to *Sim1* dosage, intrinsic protein activity influences resultant phenotypes, with amino acid variants potentially being mild or strong contributors of obesity. Strong variants, such as M136K and R171H, are monogenetic drivers of obesity, while patients with mild mutations may not necessarily be obese but could be predisposed to obesity under certain genetic and/or environmental conditions. Further supporting this concept of mild and strong variants is a recent study that highlights that not all Mc4r mutations are monogenetic drivers of obesity, rather different mutations effect the Mc4r protein activity levels to varying degrees, subsequently presenting a range of severity phenotypes [89].

The screening studies performed on both obese humans and mice has resulted in the identification of SIM1 variants, particularly loss of function mutations in the dimerisation interface. Biochemical, genetic and mouse model studies have provided convincing evidence to implicate non-functional SIM1 as a monogenetic cause of obesity. This highlights the importance of a yet to be defined role for SIM1 in adult PVN neurons to control appetite in response to changing energy demands. In order to develop effective therapeutics for the treatment of obesity, it is essential to understand the underlying molecular mechanisms involved in satiety signalling, including the role of SIM1 and *Sim1* neurons within the leptin-melanocortin pathway, particularly as *Sim1* is expressed throughout the PVN. Finally, the mechanisms underpinning loss of function of SIM1 non-synonymous variants, for which ARNT2 dimerisation remains intact, need to be defined.

Therapeutics targeting the PVN circuitry

There has been substantial evidence linking abnormal PVN development and function with obesity, thus making homeostatic pathways in this brain region prime therapeutic drug targets for the treatment of obesity. In the past, administration of leptin and compounds that activate Mc4r have been trialled as therapeutics for the treatment of obesity, but without much success due to the development of leptin resistance or side effects resultant of the multiple roles of Mc4r in energy homeostasis. Additionally, targeting Mc4r is unlikely to be useful where defective signalling is downstream of the Mc4r neurons. Therefore, SIM1 is starting to emerge as an enticing drug target for the treatment of obesity as it is downstream of both leptin and Mc4r and to the best of our knowledge, is only involved in the food intake component of energy homeostasis.

Studies in mice have shown that decreased levels of *Sim*1 results in obesity, while overexpression of *Sim1* can rescue obesity that results from dampened Mc4r activity [19-25, 68], indicating a *Sim1* dosage effect on feeding behaviour. Compounds that target SIM1 activity, in particular those that activate and upregulate SIM1 function, will make ideal drugs for the treatment of obesity. Given the strong association between Mc4r engagement and

SIM1 activity, it would be particularly critical to gain a solid understanding of the biochemical signalling between these two proteins as this link could provide insight to novel drug targets.

Alternatively to small molecule drugs, which may have off target effects and require consistent administration, advances in CRISPR technologies have promising potential as therapeutic tools to correct many genetic deficiencies without off target effects. Matharu *et al* [27] have used CRISPRa targeting the promoter or enhancer region of the endogenous *Sim1* gene to upregulate functional *Sim1* expression to normal levels in *Sim1*$^{-/+}$ mice. The study reported enduring reduced body weight due to a decrease in food intake compared to the *Sim1*$^{-/+}$ mice that were not treated with CRISPRa, with no apparent off target effects. It was also found in the study that upregulation of *Sim1* was tissue specific to where the *Sim1* promoter is active (ie. the brain and kidney), reducing the chances of any unwanted off target side effects. While CRISPRa is showing promising results as a therapeutic tool, it faces major challenges for translating therapy from a mouse to human setting, including the delivery methods used and ensuring physiological levels of expression are not exceeded.

Future directions

Although it has been established that *Sim1* dosage plays a role in appetite regulation downstream of Mc4r in the leptin-melanocortin pathway, and recent evidence shows that *Sim1*^{PDYN} neurons are also required for regulating food intake, there is limited knowledge of the molecular regulation and function of *Sim1*. Therefore, there are still a number of questions that still need to be answered in order to advance our knowledge of SIM1 regulation and function and how we can use this to our advantage in the generation of specific therapeutics for *Sim1* genetic obesity.

At the molecular level, this includes the identification of direct target genes of SIM1, which are currently unknown. Studies have attempted to identity novel target genes of SIM1 using different cell and mouse models, however there were no overlapping target genes identified between the studies [27, 90, 91]. Elucidating the signalling pathways that directly influence SIM1 expression and activity, such as post-translation modification or connecting with interacting co-factors, is urgently required to better understand SIM1 function in PVN neurons. Such knowledge will aid definition of the exact role(s) of SIM1 in different *Sim1* neuronal subpopulations, including in Mc4r and PDYN neurons, and how these independent circuits utilise SIM1 to achieve appetite homeostasis outcomes.

It would also be important to develop an understanding of the roles of SIM1 outside of appetite regulation. The leptin-melanocortin pathway and *Sim1* neurons have been implicated to play a role in sexual function in addition to its role in energy homeostasis. More investigation is still required to better understand the neuronal circuitry that is involved [92-94]. *Sim1* mRNA has also been shown to have strong expression within both the developing and adult kidney, however the specific cells types *Sim1* is expressed in within the

kidney, as well its function(s) therein, are currently unknown and requires detailed investigation.

References

1. A. W. Hetherington, S.W.R., *Hypothalamic lesions and adiposity in the rat.* The Anatomical Record, 1940. **78**: p. 149-172.
2. Anand, B.K. and J.R. Brobeck, *Hypothalamic control of food intake in rats and cats.* Yale J Biol Med, 1951. **24**(2): p. 123-40.
3. Zhang, Y., et al., *Positional cloning of the mouse obese gene and its human homologue.* Nature, 1994. **372**(6505): p. 425-32.
4. Campfield, L.A., et al., *Recombinant mouse OB protein: evidence for a peripheral signal linking adiposity and central neural networks.* Science, 1995. **269**(5223): p. 546-9.
5. Halaas, J.L., et al., *Weight-reducing effects of the plasma protein encoded by the obese gene.* Science, 1995. **269**(5223): p. 543-6.
6. Pelleymounter, M.A., et al., *Effects of the obese gene product on body weight regulation in ob/ob mice.* Science, 1995. **269**(5223): p. 540-3.
7. 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.
8. Sohn, J.W., *Network of hypothalamic neurons that control appetite.* BMB Rep, 2015. **48**(4): p. 229-33.
9. Andermann, M.L. and B.B. Lowell, *Toward a Wiring Diagram Understanding of Appetite Control.* Neuron, 2017. **95**(4): p. 757-778.
10. Chen, R., et al., *Single-Cell RNA-Seq Reveals Hypothalamic Cell Diversity.* Cell Rep, 2017. **18**(13): p. 3227-3241.
11. Kim, D.W., et al., *The cellular and molecular landscape of hypothalamic patterning and differentiation.* bioRxiv, 2020: p. 657148.
12. Campbell, J.N., et al., *A molecular census of arcuate hypothalamus and median eminence cell types.* Nat Neurosci, 2017. **20**(3): p. 484-496.
13. Romanov, R.A., et al., *Molecular design of hypothalamus development.* Nature, 2020. **582**(7811): p. 246-252.
14. Xu, S., et al., *Behavioral state coding by molecularly defined paraventricular hypothalamic cell type ensembles.* Science, 2020. **370**(6514).
15. Michaud, J.L., et al., *Development of neuroendocrine lineages requires the bHLH-PAS transcription factor SIM1.* Genes & Development, 1998. **12**(20): p. 3264-3275.
16. 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.
17. Biran, J., et al., *Role of developmental factors in hypothalamic function.* Frontiers in Neuroanatomy, 2015. **9**.
18. Alvarez-Bolado, G., *Development of neuroendocrine neurons in the mammalian hypothalamus.* Cell Tissue Res, 2019. **375**(1): p. 23-39.
19. Michaud, J.L., et al., *Sim1 haploinsufficiency causes hyperphagia, obesity and reduction of the paraventricular nucleus of the hypothalamus.* Human Molecular Genetics, 2001. **10**(14): p. 1465-1473.

20. Holder, J.L., Jr., et al., *Sim1 gene dosage modulates the homeostatic feeding response to increased dietary fat in mice.* American Journal of Physiology, 2004. **287**(1(1)): p. E105-E113.
21. Kublaoui, B.M., et al., *Sim1 haploinsufficiency impairs melanocortin-mediated anorexia and activation of paraventricular nucleus neurons.* Molecular Endocrinology, 2006. **20**(10): p. 2483-2492.
22. 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.
23. Tolson, K.P., et al., *Postnatal Sim1 Deficiency Causes Hyperphagic Obesity and Reduced Mc4r and Oxytocin Expression.* Journal of Neuroscience, 2010. **30**(10): p. 3803-3812.
24. Tolson, K.P., et al., *Inducible neuronal inactivation of Sim1 in adult mice causes hyperphagic obesity.* Endocrinology, 2014. **155**(7): p. 2436-44.
25. Kublaoui, B.M., et al., *SIM1 overexpression partially rescues agouti yellow and diet-induced obesity by normalizing food intake.* Endocrinology, 2006. **147**(10): p. 4542-4549.
26. Hossain, M.S., et al., *Identification of mutations through dominant screening for obesity using C57BL/6 substrains.* Sci Rep, 2016. **6**: p. 32453.
27. Matharu, N., et al., *CRISPR-mediated activation of a promoter or enhancer rescues obesity caused by haploinsufficiency.* Science, 2019. **363**(6424).
28. Holder, J.L., Jr., N.F. Butte, and A.R. Zinn, *Profound obesity associated with a balanced translocation that disrupts the SIM1 gene.* Human Molecular Genetics, 2000. **9**(1): p. 101-108.
29. Farooqi, I.S., et al., *Effects of recombinant leptin therapy in a child with congenital leptin deficiency.* N Engl J Med, 1999. **341**(12): p. 879-84.
30. Farooqi, I.S., et al., *Beneficial effects of leptin on obesity, T cell hyporesponsiveness, and neuroendocrine/metabolic dysfunction of human congenital leptin deficiency.* J Clin Invest, 2002. **110**(8): p. 1093-103.
31. Hahn, T.M., et al., *Coexpression of Agrp and NPY in fasting-activated hypothalamic neurons.* Nat Neurosci, 1998. **1**(4): p. 271-2.
32. Mizuno, T.M., et al., *Hypothalamic pro-opiomelanocortin mRNA is reduced by fasting and [corrected] in ob/ob and db/db mice, but is stimulated by leptin.* Diabetes, 1998. **47**(2): p. 294-7.
33. Morton, G.J., T.H. Meek, and M.W. Schwartz, *Neurobiology of food intake in health and disease.* Nat Rev Neurosci, 2014. **15**(6): p. 367-78.
34. Müller, T.D., et al., *Ghrelin.* Mol Metab, 2015. **4**(6): p. 437-60.
35. Flak, J.N. and M.G. Myers, Jr., *Minireview: CNS Mechanisms of Leptin Action.* Mol Endocrinol, 2016. **30**(1): p. 3-12.
36. Krashes, M.J., B.B. Lowell, and A.S. Garfield, *Melanocortin-4 receptor-regulated energy homeostasis.* Nature Neuroscience, 2016. **19**(2): p. 206-219.
37. Lu, D.S., et al., *Agouti protein is an antagonist of the melanocyte-stimulating-hormone receptor.* Nature, 1994. **371**(6500): p. 799-802.
38. Schwartz, M.W., et al., *Leptin increases hypothalamic pro-opiomelanocortin mRNA expression in the rostral arcuate nucleus.* Diabetes, 1997. **46**(12): p. 2119-2123.
39. Pritchard, L.E., A.V. Turnbull, and A. White, *Pro-opiomelanocortin processing in the hypothalamus: impact on melanocortin signalling and obesity.* Journal of Endocrinology, 2002. **172**(3): p. 411-421.
40. Jovanovic, Z. and G.S.H. Yeo, *Central leptin signalling: Beyond the arcuate nucleus.* Autonomic Neuroscience-Basic & Clinical, 2010. **156**(1-2): p. 8-14.

41. Schwartz, M.W., et al., *Central nervous system control of food intake.* Nature, 2000. **404**(6778): p. 661-71.
42. 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.
43. Farooqi, I.S. and S. O'Rahilly, *20 years of leptin: human disorders of leptin action.* J Endocrinol, 2014. **223**(1): p. T63-70.
44. Chua, S.C., Jr., et al., *Phenotypes of mouse diabetes and rat fatty due to mutations in the OB (leptin) receptor.* Science, 1996. **271**(5251): p. 994-6.
45. 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.
46. Huszar, D., et al., *Targeted disruption of the melanocortin-4 receptor results in obesity in mice.* Cell, 1997. **88**(1): p. 131-141.
47. Vaisse, C., et al., *A frameshift mutation in human MC4R is associated with a dominant form of obesity.* Nature Genetics, 1998. **20**(2): p. 113-114.
48. Yeo, G.S.H., et al., *A frameshift mutation in MC4R associated with dominantly inherited human obesity.* Nature Genetics, 1998. **20**(2): p. 111-112.
49. Farooqi, I.S., et al., *Dominant and recessive inheritance of morbid obesity associated with melanocortin 4 receptor deficiency.* Journal of Clinical Investigation, 2000. **106**(2): p. 271-279.
50. Vaisse, C., et al., *Melanocortin-4 receptor mutations are a frequent and heterogeneous cause of morbid obesity.* Journal of Clinical Investigation, 2000. **106**(2): p. 253-262.
51. Balthasar, N., et al., *Divergence of melanocortin pathways in the control of food intake and energy expenditure.* Cell, 2005. **123**(3): p. 493-505.
52. Gropp, E., et al., *Agouti-related peptide-expressing neurons are mandatory for feeding.* Nat Neurosci, 2005. **8**(10): p. 1289-91.
53. Luquet, S., et al., *NPY/AgRP neurons are essential for feeding in adult mice but can be ablated in neonates.* Science, 2005. **310**(5748): p. 683-5.
54. Bultman, S.J., E.J. Michaud, and R.P. Woychik, *Molecular characterization of the mouse agouti locus.* Cell, 1992. **71**(7): p. 1195-204.
55. Klaauw, A.A.v.d. and I.S. Farooqi, *The hunger genes: pathways to obesity.* Cell (Cambridge), 2015. **161**(1): p. 119-132.
56. Flier, J.S. and E. Maratos-Flier, *Leptin's Physiologic Role: Does the Emperor of Energy Balance Have No Clothes?* Cell Metab, 2017. **26**(1): p. 24-26.
57. Aponte, Y., D. Atasoy, and S.M. Sternson, *AGRP neurons are sufficient to orchestrate feeding behavior rapidly and without training.* Nat Neurosci, 2011. **14**(3): p. 351-5.
58. Atasoy, D., et al., *Deconstruction of a neural circuit for hunger.* Nature, 2012. **488**(7410): p. 172-7.
59. Fenselau, H., et al., *A rapidly acting glutamatergic ARC-->PVH satiety circuit postsynaptically regulated by alpha-MSH.* Nat Neurosci, 2017. **20**(1): p. 42-51.
60. Krashes, M.J., et al., *Rapid, reversible activation of AgRP neurons drives feeding behavior in mice.* J Clin Invest, 2011. **121**(4): p. 1424-8.
61. Zhan, C., et al., *Acute and long-term suppression of feeding behavior by POMC neurons in the brainstem and hypothalamus, respectively.* J Neurosci, 2013. **33**(8): p. 3624-32.
62. Xu, Y., et al., *Glutamate Mediates the Function of Melanocortin Receptor 4 on Sim1 Neurons in Body Weight Regulation.* Cell Metabolism, 2013. **18**(6): p. 860-870.
63. Shah, B.P., et al., *MC4R-expressing glutamatergic neurons in the paraventricular hypothalamus regulate feeding and are synaptically connected to the parabrachial nucleus.*

Proceedings of the National Academy of Sciences of the United States of America, 2014. **111**(36): p. 13193-13198.
64. Sweeney, P., et al., *Network dynamics of hypothalamic feeding neurons.* Proc Natl Acad Sci U S A, 2021. **118**(14).
65. Li, M.M., et al., *The Paraventricular Hypothalamus Regulates Satiety and Prevents Obesity via Two Genetically Distinct Circuits.* Neuron, 2019. **102**(3): p. 653-667.e6.
66. Kublaoui, B.M., et al., *Oxytocin deficiency mediates hyperphagic obesity of Sim1 haploinsufficient mice.* Mol Endocrinol, 2008. **22**(7): p. 1723-34.
67. Garfield, A.S., et al., *A neural basis for melanocortin-4 receptor-regulated appetite.* Nature Neuroscience, 2015. **18**(6): p. 863-U299.
68. 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.
69. Yu, J., et al., *Determination of the melanocortin-4 receptor structure identifies Ca(2+) as a cofactor for ligand binding.* Science, 2020. **368**(6489): p. 428-433.
70. Israeli, H., et al., *Structure reveals the activation mechanism of the MC4 receptor to initiate satiation signaling.* Science, 2021.
71. Kewley, R.J., M.L. Whitelaw, and A. Chapman-Smith, *The mammalian basic helix-loop-helix/PAS family of transcriptional regulators.* International Journal of Biochemistry & Cell Biology, 2004. **36**(2): p. 189-204.
72. Jain, S., et al., *Potent transactivation domains of the Ah receptor and the Ah receptor nuclear translocator map to their carboxyl termini.* Journal of Biological Chemistry, 1994. **269**(50): p. 31518-31524.
73. Jiang, B.H., et al., *Dimerization, DNA binding, and transactivation properties of hypoxia-inducible factor 1.* Journal of Biological Chemistry, 1996. **271**(30): p. 17771-17778.
74. Crews, S.T. and C.M. Fan, *Remembrance of things PAS: regulation of development by bHLH-PAS proteins.* Current Opinion in Genetics & Development, 1999. **9**(5): p. 580-587.
75. Michaud, J.L., et al., *ARNT2 acts as the dimerization partner of SIM1 for the development of the hypothalamus.* Mechanisms of Development, 2000. **90**(2): p. 253-261.
76. Ema, M., et al., *Two new members of the murine Sim gene family are transcriptional repressors and show different expression patterns during mouse embryogenesis.* Molecular and Cellular Biology, 1996. **16**(10): p. 5865-5875.
77. Hosoya, T., et al., *Defective development of secretory neurones in the hypothalamus of Arnt2-knockout mice.* Genes to Cells, 2001. **6**(4): p. 361-374.
78. 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.* Proceedings of the National Academy of Sciences of the United States of America, 2001. **98**(12): p. 6692-6697.
79. Jain, S., et al., *Expression of ARNT, ARNT2, HIF1 alpha, HIF2 alpha and Ah receptor mRNAs in the developing mouse.* Mechanisms of Development, 1998. **73**(1): p. 117-123.
80. Aitola, M.H. and M.T. Pelto-Huikko, *Expression of Arnt and Arnt2 mRNA in developing murine tissues.* Journal of Histochemistry & Cytochemistry, 2003. **51**(1): p. 41-54.
81. Kozak, K.R., B. Abbott, and O. Hankinson, *ARNT-deficient mice and placental differentiation.* Developmental Biology, 1997. **191**(2): p. 297-305.
82. Maltepe, E., et al., *Abnormal angiogenesis and responses to glucose and oxygen deprivation in mice lacking the protein ARNT.* Nature, 1997. **386**(6623): p. 403-407.

83. Turer, E.E., et al., *A viable hypomorphic Arnt2 mutation causes hyperphagic obesity, diabetes and hepatic steatosis.* Dis Model Mech, 2018. **11**(12).
84. Bonnefond, A., et al., *Loss-of-function mutations in SIM1 contribute to obesity and Prader-Willi-like features.* J Clin Invest, 2013. **123**(7): p. 3037-41.
85. Ramachandrappa, S., et al., *Rare variants in single-minded 1 (SIM1) are associated with severe obesity.* Journal of Clinical Investigation, 2013. **123**(7): p. 3042-3050.
86. Sullivan, A.E., et al., *Characterization of human variants in obesity-related SIM1 protein identifies a hot-spot for dimerization with the partner protein ARNT2.* Biochemical Journal, 2014. **461**: p. 403-412.
87. Blackburn, P.R., et al., *Functional analysis of the SIM1 variant p.G715V in two patients with obesity.* J Clin Endocrinol Metab, 2019.
88. Montagne, L., et al., *Identification of two novel loss-of-function SIM1 mutations in two overweight children with developmental delay.* Obesity (Silver Spring), 2014. **22**(12): p. 2621-4.
89. Brouwers, B., et al., *Human MC4R variants affect endocytosis, trafficking and dimerization revealing multiple cellular mechanisms involved in weight regulation.* Cell Rep, 2021. **34**(12): p. 108862.
90. 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.
91. Caqueret, A., F. Boucher, and J.L. Michaud, *Laminar organization of the early developing anterior hypothalamus.* Dev Biol, 2006. **298**(1): p. 95-106.
92. Jorgenson, E., et al., *Genetic variation in the SIM1 locus is associated with erectile dysfunction.* Proc Natl Acad Sci U S A, 2018. **115**(43): p. 11018-11023.
93. Semple, E. and J.W. Hill, *Sim1 Neurons Are Sufficient for MC4R-Mediated Sexual Function in Male Mice.* Endocrinology, 2018. **159**(1): p. 439-449.
94. Semple, E., F. Shalabi, and J.W. Hill, *Oxytocin Neurons Enable Melanocortin Regulation of Male Sexual Function in Mice.* Mol Neurobiol, 2019. **56**(9): p. 6310-6323.

Project aims

The aims of this project was to further understand the function and regulation of SIM1, ARNT and ARNT2. More specifically:

- Ascertain the utility of mouse models, engineered to provide inducible and reversible depletion of ARNT and ARNT2, to investigate the unique and overlapping physiological functions of ARNT and ARNT2.
- Investigate reduced activity of the human obesity linked SIM1 mutant, R171H, in a recapitulated mouse model and in cultured cells.
- To optimise a mammalian cell protein expression system that will facilitate purification of a truncated SIM1/ARNT2 heterodimer for use in structural studies.

Chapter 2: Materials and Methods

Abbreviations

Ab	antibody
α-MSH	α-melanocyte stimulating hormone
APS	ammonium persulphate
ARC	arcuate nucleus
ARNT	aryl hydrocarbon nuclear translocator
bp	base pair
BSA	bovine serum albumin
°C	degrees Celsius
cDNA	coding deoxyribonucleic acid
CME	central midline element
DAPI	4’6-diamidino-2-phenylindole
DMEM	Dulbecco’s modified Eagle’s medium
DMSO	dimethylsulphoxide
DNA	deoxyribonucleic acid
dNTP	deoxyribonucleotide triphosphate
Dox	Doxycycline
DTT	dithiothreitol
EDTA	ethylene diamine tetraacetic acid
EtBr	ethidium bromide
EtOH	ethanol
FBS	fetal bovine serum
HBSS	Hanks’ Balanced Salt Solution
HEK	human embryonic kidney
HEPES	4-(2-hydroxyethyl)piperazine-1-ethanesulphonic acid
HIF	hypoxia inducible factor
hr	hour
HRP	horseradish peroxide
IF	immunofluorescence
Igepal	Igepal CA-630
IP	immunoprecipitation
IRES	internal ribosome entry site
kDa	kiloDalton
L	litre
LB	Luria broth
L-Glu	L-glutamine
M	molar
mA	milliampere
mAU	milli arbitrary units

Mc4r	melanocortin 4 receptor
MEF	mouse embryonic fibroblast
mES cell	mouse embryonic stem cell
MeOH	methanol
miR/miRNA	micro ribonucleic acid
mnucTomato	monomeric nuclear Tomato
µg	microgram
µl	microlitre
µm	micrometre
µM	micromolar
min	minute
ml	millilitre
mM	millimolar
mRNA	messenger ribonucleic acid
MTII	melanotan II
Neo	neomycin
ng	nanogram
NLS	nuclear localisation signal
O/N	overnight
PAGE	polyacrylamide gel electrophoresis
PBS	phosphate buffered saline
PCR	polymerase chain reaction
PFA	paraformaldehyde
PIC	protease inhibitor cocktail
POMC	pro-opiomelanocortin
P/S	penicillin/streptomycin
PVN	paraventricular nucleus
qPCR	real time PCR/quantitative PCR
RNA	ribonucleic acid
RNAi	RNA interference
RO	reverse osmosis
rpm	revolutions per minute
RT	room temperature
RT-PCR	reverse transcriptase PCR
SD	standard deviation
SDS	sodium dodecyl sulphate
sec	second
SEM	standard error of the mean
shRNA	short hairpin RNA
SIM	single minded
siRNA	small interfering RNA
S/N	supernatant

SON	supraoptic nucleus
TBE	tris-borate-EDTA
TE	tris-EDTA
TetR	tetracycline and doxycycline responsive repressor protein
TEMED	N,N,N^1,N^1-tetramethyl-ethylenediamine
TGE	tris-glycine-EDTA
Tris	tris (hydroxymethyl) aminomethane
Tween-20	polyoxyethylene-sorbitan monolaurate
U	unit
UV	ultraviolet
V	volts
v/v	volume per volume
WCE	whole cell extract
WT	wild type
w/v	weight per volume

Materials

Antibodies

Antibodies were diluted in 2% skim milk/PBST for western blots and in indicated blocking solution for IF.

Primary antibodies

Target	Species	Brand	Cat#/clone	WB dilution	IF dilution
ARNT	Rabbit	Proteintech	14105-1-AP	1:2000	-
ARNT2	Rabbit	Santa Cruz	M-165	1:1000	1:500
dsRed	Rabbit	Clontech	632496	1:1000	-
Flag	Mouse	Sigma	M2	1:10000 – 1:20000	-
GFP	Goat	Rockland	600-101-215	-	1:1000
HA	Rabbit	Cell signalling technologies	C29F4	1:2000	-
Myc	Mouse	Merck	4A6	1:2000	-
Myc	Rabbit	Abcam	ab9106	1:2000	-
OTP	Rabbit	GeneTex	GTX119601	1:2000	-
Strep	Mouse	IBA	2-1507-001	1:10000	-
Strep-HRP	Mouse	IBA	2-1502-001	1:10000	-
α-Tubulin	Rat	AbD Serotec/Bio-Rad	MCA78G	1:10000	-

Secondary antibodies

Antibody	Brand	Cat#/clone	WB dilution	IF dilution
Goat α-mouse HRP	Pierce	31430	1:10000	-
Goat α-rabbit HRP	Pierce	31460	1:10000	-
Rabbit α-rat HRP	Dako	P0450	1:10000	-
Donkey α-rabbit Alexa Fluor® 488	Invitrogen	A-21206	-	1:1000
Donkey α-rabbit Alexa Fluor® 594	Invitrogen	A-21207	-	1:1000
Donkey α-goat Alexa Fluor® 488	Abcam	Ab150129	-	1:1000

Bacterial strains

E.coli DH5α
Used for plasmid DNA propagation.

Tissue culture cell lines

Established immortalised cell lines

HEK293T
HEK293 cell lines constitutively expressing the SV40 large T antigen.

Expi293F™
HEK293 cell line variant that has been adapted to grow as a high density suspension culture in Expi293 media (Gibco).

Solutions

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

General laboratory solutions

Buffer W
100mM Tris pH 8, 150mM NaCl, 1mM EDTA

Buffer W (high salt and detergent)
100mM Tris pH 8, 500mM NaCl, 1% Triton X-100, 1mM EDTA

Coomassie Blue Stain
0.03% (w/v) Coomassie Blue R250 (Sigma), 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

Chemiluminescence detection solution
10ml 10mM Tris pH 8.5, 3µl 30% H_2O_2, 30µl 250mM luminol, 25µl 90mM p-coumaric acid

DNA load buffer, 6x
1mM EDTA pH 8.0, 0.25% (w/v) Bromophenol Blue, 50% (w/v) glycerol

Flag IP wash buffer
500mM NaCl, 1% Triton X-100, 10mM Tris pH 8, 1mM EDTA

Flag peptide buffer
100mM Tris pH 8, 200mM NaCl, 1mM EDTA

GST running buffer, 10x
14.4% (w/v) glycine, 3% (w/v) Tris, 1% (w/v) SDS, corrected to pH 8.3 with 1M HCl

Lower Buffer, 4x
1.5M Tris base, 0.4% (w/v) SDS corrected to pH 8.8 with 1M HCl and filtered 0.45µm

PBS, 1x
0.8% (w/v) NaCl, 0.02% (w/v) KCl, 0.4% (w/v) KH_2PO_4 (anhydrous), 2.3% Na_2HPO_4 (anhydrous)

PBST
0.1% (w/v) Tween-20 in PBS
Ponceau stain
0.5% (w/v) Ponceau S, 1% (w/v) glacial acetic acid

PIC, x100
10µl aprotinin (2mg/ml), 10µl leupeptin (5mg/ml), 10µl pepstatin (1mg/ml), 30µl EDTA (pH 7.5, 500mM), 40µl MQ

SDS loading buffer, 4x
40% (w/v) glycerol, 10% (v/v) 4x Upper Buffer, 5% (w/v) SDS, 0.01% (w/v) Bromophenol Blue, 0.08M DTT

Separation gel, 7.5%
1x Lower Buffer, 7.5% (v/v) 30-40% Acrylamide:Bis-Acrylamide 29:1 (Bio-Rad), 0.08% (w/v) APS, 0.08% (v/v) TEMED

Stacking gel, 4.5%
1x Upper Buffer, 4.5% (v/v) 30-40% Acrylamide:Bis-Acrylamide 29:1 (Bio-Rad), 0.08% (w/v) APS, 0.08% (v/v) TEMED

TBE buffer, 20x
0.1M EDTA, 1M Tris

TE, 100x
0.1M EDTA, 1M Tris

Upper Buffer, 4x
0.5M Tris base, 0.4% (w/v) SDS corrected to pH 6.8 with 1M HCl and filtered 0.45µm

Western blocking solution
10% (w/v) skim milk powder in PBST

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 PIC added fresh to a final concentration of 1mM and 1x respectively

WCE buffer – Expi293F cells
50mM Tris pH 8, 150mM NaCl, 0.5% NP-40, 1mM EDTA with 0.2mM PMSF, 0.5mM DTT and 1x PIC added fresh

Wet transfer buffer, 5x
50mM Tris, 400mM glycine

Wet transfer buffer, 1x
200ml 5x wet transfer buffer, 200ml methanol, 600ml RO water

Bacterial growth media

LB
2.5% (w/v) Luria Broth Base

SOB
20mg/ml Tryptone (Oxoid), 5mg/ml Yeast extract (Oxoid), 10mM NaCl, 2.5mM KCl, 10mM $MgCl_2$, 10mM $MgSO_4$

SOC
SOB + 2% glucose

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

Final concentrations of antibiotics for liquid and solid media
Ampicillin 100µg/ml
Kanamycin 50µg/ml

Tissue culture media

HEK293T, MEF and adult kidney media
DMEM with HEPES (Gibco, cat#12430) supplemented with 10% FBS, 1% P/S, 1% L-Glu

Expi293F™ media
Expi293F™ Expression Medium (Gibco, cat# A14351)

Neuronal cells
Neurobasal® Medium (Gibco, cat#21103) supplemented with x1 B27 (Gibco, cat#17504044), 1% P/S, 1% L-Glu

Plasmids

All plasmids were previously generated in the Whitelaw laboratory unless otherwise stated.

pEF-IRES-puro6-hSIM1-2HA-3FLAG
EF promoter drives constitutive expression of full length human SIM1 with 3xFlag and 2xHA tag. Contains a puromycin resistance cassette for selection. Parental vector pEF-IRES-puro6 described by Hobbs et al. [29].

pEF-IRES-puro6-hSIM1-R171H-2HA-3FLAG
EF promoter drives constitutive expression of full length human R171H mutant SIM1 with 3xFlag and 2xHA tag. Contains a puromycin resistance cassette for selection.

pcDNA5-FRT-TO-hOTP
CMV promoter drives constitutive expression of full length human OTP. Contains a hygromycin resistance cassette for selection. Parental vector pcDNA5-FRT-TO purchased from Invitrogen.

pEF-puro-hSIM1-1-348-3xFLAG-2xSTREP
EF promoter drives constitutive expression of truncated human SIM1 comprising of amino acids 1-348 with 3xFlag and 2xStrep tag. Contains a puromycin resistance cassette for selection.

pEF-neo-hARNT2-1-455-2xMYC
EF promoter drives constitutive expression of truncated human ARNT2 comprising of amino acids 1-455 with 2xMyc tag and contains a neomycin resistance cassette for selection [30].

shRNA sequences

***Arnt* shRNA 1498**
TGCTGTTGACAGTGAGCGCCAGATGAAATTGAGTATATTATAGTGAAGCCACAGATGTATAATATAC
TCAATTTCATCTGATGCCTACTGCCTCGGA

***Arnt2* shRNA 318**
TGCTGTTGACAGTGAGCGCGATGGTGAAGGTCCCAGTAAATAGTGAAGCCACAGATGTATTTACTG
GGACCTTCACCATCTTGCCTACTGCCTCGGA

Primers

Genotyping primers

Arnt2$^{-/+}$	
WT3	5’ AGTCCTTGTGACTGTGATGGTCCT 3’
WT5	5’ CAGATAAAACGTGTATGTGCCCTC 3’
TV neo	5’ TCAGAGCAGCCGATTGTCTGTTGTGCCCAGTCAT 3’
Arnt and *Arnt2* shRNA	
Col1a1 5' geno F2	5’ AATCATCCCAGGTGCACAGCATTGCGG 3’
Col1a1 5' geno R2	5’ CTTTGAGGGCTCATGAACCTCCCAGG 3’
Flp Inducer geno R1	5’ TGGACTACTGCGCCCTACAGATCTGC 3’
Sim1 R171H	
MW21	5’ AGGGGCATTGCACCATTACAG 3’
MW25	5’ GCTAAGGCTTTGGTTCTTAACTTCC 3’
qPCR 171 genotyping Fwd	5’ CACTTGTGTCGTCCAGAACTTT 3’
ACN homol recomb Rev	5’ GAAGTTATAAGCTTTCGCGAGCTCGAG 3’
Cre-pos Fwd	5’ CTAGGCCACAGAATTGAAAGATCT 3’
Cre-pos Rev	5’ GTAGGTGGAAATTCTAGCATCATCC 3’
Sim1-GFP	
mSim1 (-11) F	5’ TTGTAAGAAGAAAGGGAGCCCG 3’
Genstat GFP Rev	5’ TAGCGGCTGAAGCACTGCA 3’

qPCR

mrSim1 RT F	5’ ACTGGGAGAGATCCCTGGTG 3’
mrSim1 RT R	5’ GAGCTGACCACACTATCTTCATC 3’
d2.2GFP F	5’ AGAAGAACGGCATCAAGGTG 3’
d2.2GFP R	5’ GAACTCCAGCAGGACCATGT 3’
GABRE F	5’ TCAATGCGAAGAACACTTGG 3’
GABRE R	5’ AGAAGGAGACCCAGGAGAGC 3’
β-tubulin real time F	5’ CGACAATGAAGCCCTCTACGAC 3’
β-tubulin real time R	5’ ATGGTGGCAGACACAAGGTGGTTG 3’
β-actin qPCR F	5’ CTGCCTGACGGCCAGG 3’
β-actin qPCR R	5’ GATTCCATACCCAAGAAGGAAGG 3’

Methods

Bacterial cell culture

Heat shock transformation of chemically competent bacteria

Approximately 10ng plasmid DNA was added to 50µl competent DH5α E. coli cells followed by incubation on ice for 20min. Bacteria were then heat shocked: 42°C, 45sec or 2min; ice, 2min. 250µl SOC was added to cells, followed by shaking at 37°C for 20-40min. Bacteria were plated on agar plates with the appropriate antibiotic and grown O/N at 37°C.

Growth and maintenance of bacteria

Colonies picked from antibiotic agar plates were grown in 3ml, 5ml or 100ml LB cultures containing the appropriate antibiotic. Cultures were grown O/N at 37°C with agitation.

Freezing bacteria stocks

Glycerol was added to liquid culture to a final concentration of 16% in a screw top cryotube, incubated for 5min at RT and stored at -80°C.

DNA and RNA analysis

Preparation of plasmid DNA

Purified plasmid DNA was extracted from O/N bacterial cultures using QIAprep Spin Miniprep, QIAfilter Plasmid Midi, QIAGEN Plasmid Plus Midi, QIAGEN Plasmid Maxi or Giga kits. Purifications performed as per manufacturer’s instructions. Plasmid DNA was eluted in MQ water.

Isolation of total RNA

Media was aspirated and 200-800ul TRI Reagent (Trizol) was added directly to cells and incubated at RT for 1min. The homogenate was then transferred to an Eppendorf tube and either stored at -80°C or used immediately. TRI Reagent protocol was followed for RNA extraction, using chloroform in place of BCP, as per manufacturer's protocol. Pellets were resuspended in 1x TE buffer or MQ water.

cDNA synthesis from RNA

500ng-2µg RNA was incubated at 65°C with 1µl 500ng/µl Oligo dTs (15), 1 µl 50ng/µl dN6 and 2µl 5mM dNTPs and made up to 12 µl with MQ water followed by incubation on ice for 1min. Superscript III reverse transcriptase (Invitrogen) in 4µl 5x First Strand Buffer and 1µl 0.1M DTT was then added to the RNA-oligo mix to a final volume of 20µl made up with MQ water and incubated at 25°C, 50min; 50°C, 90min; 70°C, 30min. 30 µl 1x TE buffer or MQ water was added to the cDNA and stored at -20°C.

qPCR

Master mixes comprising of 22.5 µl SYBR Green Master Mix, 4.5 µl of each 5 µM primer and 2 µl cDNA were made up to 45 µl with MQ water. Master mixes were aliquoted into 15 µl triplicates into a 96 well plate. Samples were analysed on an Applied Biosystems Step One Plus thermocycler. Standard reaction conditions: 95°C, 10min; [95°C, 10sec; 60°C, 30sec] x40 cycles; 95°C, 15sec followed by melt curve analysis of 60°C, 1min, increase in sample temperature to 95°C with fluorescence measurements made every 0.5°C; 95°C, 15sec. Results were analysed using Step One software, QGene and Microsoft Excel.

Agarose gel electrophoresis

6x DNA loading buffer was added to DNA samples and separated by 1-2% (w/v) agarose gel containing EtBr. Electrophoresis was performed at 110-120V in 1x TBE buffer. 1Kb Plus DNA ladder (Invitrogen) was used as a size marker and DNA fragments were visualised by BioDoc-It Imaging System (UVP). Electrophoresis of RNA was performed as described for DNA. The equipment used for RNA electrophoresis was soaked in 10% SDS, 2M NaOH and RO water for at least 20min prior to use to remove RNases to prevent RNA degradation during gel electrophoresis.

Mammalian cell culture

Thawing frozen cell line stocks

HEK293T cells

Frozen cells were thawed in 37°C water bath and resuspended in 5-8ml appropriate media. Cells were centrifuged at 1000-1500rpm for 3-5min followed by removal of S/N. Cells were resuspended in an appropriate volume of fresh media for plating and incubated at 37°C, 5% CO_2.

Expi293F cells
Frozen Expi293F cells were thawed in 37°C water bath and resuspended in 5-8ml appropriate media. Cells were centrifuged at 1000-1500rpm for 3-5min followed by removal of S/N. Cells were resuspended in an appropriate volume of fresh media in a conical cell culture flask and incubated at 37°C, 8% CO_2 on an orbital shaker platform (125rpm).

Freezing cell line stocks

HEK293T cells
Cells were trypsinised in 0.05% or 0.25% trypsin/EDTA (Gibco) and resuspended in media before being centrifuged at 1000-1500rpm for 3-5min. S/N was removed and cell pellets were resuspended in freezing media (50% appropriate media, 40% FBS, 10% DMSO), aliquoted into Cryotube™ vials and stored in an isopropanol filled Cryo 1°C freezing container (NALGENE™) at -80°C.

Expi293F cells
Cells were centrifuged at 1000-1500rpm for 3-5min, S/N removed and cell pellets resuspended in Expi293F media (Gibco) with 10% DMSO. Cells were aliquoted into Cryotube™ vials and stored in an isopropanol filled Cryo 1°C freezing container (NALGENE™) at -80°C.

Routine maintenance of cell lines

HEK293T cells
Cells were passaged every 2-4 days or when near 90-100% confluent. Media was aspirated and cells washed once with PBS. 0.05% or 0.25% trypsin/EDTA (Gibco) was used to lift cells from culture dishes and media added to inactivate trypsin. Cells were typically diluted 1:10 or 1:20 into a new dish with fresh media. Cells were incubated at 37°C, 5% CO_2.

Expi293F cells
Cells were passaged every 3-4 days or when cell density reached $5x10^6$ cells/ml. An aliquot of cell suspension was transferred to a new conical cell culture flask and diluted in fresh Expi293F media. Cells were typically diluted 1:10 and were incubated at 37°C, 8% CO_2 on an orbital shaker platform (125rpm).

Cell counting

Cells were stained with trypan blue (Sigma) and counted using either a hemocytometer or an automated cell counter (TC20™ automated cell counter, Bio-Rad).

Transient transfections

HEK293T cells

Cells were typically transfected at 60-80% confluency. Plasmid DNA was diluted in serum free DMEM before addition of Polyethylenimine (PEI, Polysciences) transfection reagent, using PEI at a 3:1 ratio of PEI(µg):DNA(µg). PEI:DNA mix was incubated at RT for 15min then added dropwise to cells. Cells were incubated at 37°C, 5% CO_2.

Expi293F cells

Cells were seeded at $2x10^6$ cells/ml. For each ml of cell culture, 2µg of DNA was transfected using 4µl PEI in 108µl PBS. Ratio of SIM1:ARNT2 plasmid used was 10:1. DNA, PEI and PBS was premixed with vigorous vortexing and incubated at RT for 20min. The DNA/PEI/PBS mix was added to cells dropwise with swirling of the flask. Following transfection, cells were incubated on an orbital shaker (125rpm) at 37°C, 8% CO_2 until harvesting for protein extraction.

Immunofluorescence of monolayer cell cultures

Cells were seeded onto 0.2% gelatin (Sigma) or 2µg/ml Poly-D-Lysine (Sigma) coated glass coverslips in a 24 well tray. Cells were treated with Dox as appropriate. Media was removed and cells washed in PBS before fixation with 4% PFA for 15-20min at RT. Following fixation, PFA was removed and cells washed 3x in PBST with gentle shaking. Cells were then permeabilised with 0.2% Triton X-100 in PBS for 10min at RT followed by 2x PBST washes. Cells were incubated in blocking solution (10% NHS/PBST) for 2hr at RT followed by incubation in primary antibody O/N at 4°C in a humidified chamber with gentle shaking. Following primary antibody incubation, cells were washed 3x in PBST before incubation in secondary antibody for at least 2hr at RT with gentle shaking and covered in foil. Cells were washed with PBST before mounting coverslips onto microscope slides using ProLong Gold antifade reagent with DAPI (Invitrogen). Slides were allowed to dry at 4°C O/N before imaging on a fluorescence microscope.

Cell imaging and image analysis

All brightfield and fluorescent cell and embryo images were taken on Nikon Eclipse Ti microscope, Nikon Australia. Fiji (a distribution of ImageJ) software [31] was used for image analysis.

Protein analysis

Whole cell extract preparation

Adherent cultured cells
Cells were washed with 1ml ice cold 1x PBS and WCE buffer was added directly to cells. Cells were scraped and transferred to an Eppendorf tube followed by incubation on ice for 10-20min. Cells were pelleted at 14000rpm for 30min at 4°C. S/N containing the protein was transferred to a new Eppendorf tube and stored at -80°C.

Expi293F cells
Expi293F suspension cells were centrifuged at 1500rpm for 15min followed by washing of cell pellets in PBS. Cell pellets were resuspended in ice cold WCE buffer and incubated at 4°C for 20min with rotation. NaCl was added dropwise to a final concentration of 0.42M and mixed well. Cells were centrifuged at 22000-450000rpm at 4°C for 45-60min. S/N containing the protein was used for Co-IP protocols.

Mouse tissues
Mouse tissues were placed directly into WCE buffer and homogenised using a dounce homogeniser followed by incubation on ice for 30min. The homogenate was pelleted at 14000rpm for 30min at 4°C. S/N containing the protein was transferred to a new Eppendorf tube and stored at -80°C.

Protein quantification

Protein concentrations of WCE were measured using Bradford Protein Assay as per manufacturer's protocol and calculated from OD_{600} measurements with reference to a BSA standard curve.

Concentration estimates of purified protein were made by using Image Lab (Version 5.2.1, Bio-Rad) to analyse Coomassie stained gels using the included BSA standards.

Co-Immunoprecipitation

Flag Co-IP (SDS elution)
Anti-Flag M2 affinity gel resin (Sigma Aldrich) was pre-washed in IPP150 wash buffer and resin pelleted by centrifugation at 2000rpm for 3min. Resin was blocked with 0.05mg/ml BSA for 30min-1hr at 4°C with rocking, followed by pelleting and washed in 1ml IPP150 buffer. Resin was resuspended 50% by volume in wash buffer, added to WCE and incubated at 4°C for at least 3hr or O/N with rotation. WCE/resin was centrifuged at 2000rpm for 3min and washed in 1ml IPP150 wash buffer 3x 5min at 4°C. To elute the protein, the resin was incubated in 4xSDS LB for 2min at 95°C and centrifuged at 10000rpm for 1min.

Flag Co-IP (Flag peptide elution)

Anti-Flag M2 affinity gel resin (Sigma Aldrich) was pre-washed in wash buffer and resin pelleted by centrifugation at 2000rpm for 5min. WCE was added to resin and incubated at 4°C O/N with rotation. WCE/resin was pelleted and washed 3x for 10min at 4°C in wash buffer followed by washing in Flag peptide elution buffer without peptide. To elute the protein, the resin was rotated at 4°C for 1hr with 250ng/µl 3x Flag peptide diluted in Flag peptide elution buffer, centrifuged and S/N containing eluted protein removed. For small scale purification, incubation in Flag peptide was repeated 2x 5min while large scale was repeated at 2x 1hr. The anti-Flag M2 affinity gel resin was regenerated for subsequent uses by washing with 0.1M glycine followed by PBS until effluent was pH 7. Resin was stored in 50% glycerol and 0.02% sodium azide in PBS.

Flag affinity column chromatography

Anti-Flag M2 affinity gel resin (Sigma Aldrich) in a column was pre-washed with 5 column volumes of wash buffer. Protein solution was passed through the resin 3x followed by 5 column volumes wash buffer then 5 column volumes peptide elution buffer without peptide added. To elute protein, resin was incubated with 250ng/µl 3x Flag peptide diluted in Flag peptide elution buffer in capped column for 20min before collecting elutions. The anti-Flag M2 affinity gel resin was regenerated for subsequent uses by washing with 3 column volumes 0.1M glycine followed by PBS until effluent was pH 7. Resin was stored in 50% glycerol and 0.02% sodium azide in PBS.

Strep Co-IP (Biotin elution)

Strep-Tactin XT Superflow resin (IBA) was pre-washed in buffer W and resin pelleted by centrifugation at 2000rpm for 5min. WCE was added to resin and incubated at 4°C O/N with rotation. WCE/resin was pelleted and washed 4x with buffer W. To elute the protein, the resin was rotated at 4°C for 1hr with 50mM biotin in 1x BXT buffer (IBA), centrifuged and S/N containing eluted protein removed. Incubation in biotin peptide was repeated 2x 5min.

Strep affinity column chromatography

Strep-Tactin XT Superflow resin (IBA) in a column was pre-washed with 5 column volumes of WCE buffer. WCE was passed through the resin, filtrate collected and passed through resin a second time. Resin was washed with 10 column volumes of WCE buffer, followed by 5 column volumes buffer W. Where indicated, the resin was washed with an additional 5 column volumes of buffer W containing higher salt and detergent concentrations in between the WCE buffer and buffer W wash steps. Protein was eluted from the resin by incubating in 50mM biotin in 1x BXT buffer (IBA) in the plugged column for 5min before collecting elutions. The Strep-Tactin XT resin was regenerated for subsequent uses by washing with 2 column volumes of freshly made 10mM NaOH followed by 10 column volumes buffer W. Resin was stored in the column with buffer W and 0.02% NaN_3.

Denaturing SDS-PAGE

Denaturing SDS-PAGE was performed in 7.5% Bis-Acrylamide separation gels with 4.5% stacking Bis-Acrylamide stacking gel. Separation gel was set at RT for 20-30min with an overlay of MQ water. After setting, MQ water was removed and stacking gel was applied and allowed to set at RT for 20-30min. Gels were used straight away or stored at 4°C O/N. Alternatively, precast Bio-Rad Mini-PROTEAN III 4.5% stacking, 7.5% or 12% separation gels were used for separation of proteins.

10-100µg protein was suspended in 4x SDS loading buffer. Protein was loaded onto SDS-PAGE gel and separated at 120-150V in 1x GST running buffer. Dual colour precision plus protein standards (Bio-Rad) was used for both western blot analysis and Coomassie gels.

Western blot analysis

Following SDS-PAGE separation, proteins were transferred to a nitrocellulose membrane using the wet transfer method at 250mA for 80min at 4°C or using the Turbo Blot (Bio-Rad) apparatus, following manufacturer's protocol for Bio-Rad precast Mini-PROTEAN III gels. After transfer, membranes were blocked with 10% (w/v) skim milk/PBST at RT for at least 1hr. Membranes were then incubated O/N at 4°C or 2-4hr at RT with primary Ab. Membranes were washed 3x in PBST for 3min before incubation with secondary Ab at RT for 1-2hr. The membranes were washed 3x for 3min in PBST followed by incubation with chemiluminescence detection solution or Clarity Western ECL Substrate (Bio-Rad) as per manufacturer's protocol for 2-5min. Proteins was visualised using a Chemidoc MP Imaging System (Bio-Rad).

Coomassie staining

Following SDS-PAGE separation, gels were stained in Coomassie Blue Stain O/N at RT with shaking. Coomassie Blue Stain was then removed and the gel placed into Coomassie Destain I for 1-2hr at RT with shaking. Destain I was replaced with Destain II for 1-2hr at RT with shaking. Gel images were taken using Chemidoc MP Imaging System (Bio-Rad).

Analytical Size-Exclusion Chromatography (SEC)

Concentrated purified protein (~1mg/ml) was loaded onto a Superdex 200 10/300 GL analytical SEC (GE Healthcare), which was equilibrated in buffer W at a flow rate of 0.4 mL/min at room temperature. The size exclusion column was calibrated using protein standards (Sigma) containing bovine thyroglobulin (670kDa), bovine γ-globulin (158kDa), chicken ovalbumin (44kDa), horse myoglobin (17kDa) and bovine aprotinin (7kDa). SEC was performed by Dr Blagojce Jovcevski (Pukala Laboratory, University of Adelaide).

Mass spectrometry analysis

The band of interest was excised from the SDS-PAGE gel, destained and digested with trypsin. The unknown protein was analysed using MALDI TOF mass spectrometry (MS) and MS/MS using a Bruker Ultraflextreme MALDI TOF/TOF mass spectrometer (Bruker Daltonik GmbH) under the control of the flexControl software (Version 3.4, Bruker Daltonik GmbH). Protein identification was performed using MOWSE and probability scores calculated by the in-house Mascot database-searching engine (Version 2.2, Matrix Science). MS was performed by Chris Cursaro (University of Adelaide).

Mouse work

All mouse work was approved by the University of Adelaide Animal Ethics Committee and was performed in compliance with ethical standards.

Mice had access to normal chow diet and water ad libitum. Where stated, mice were fed either Doxycycline at 600mg/kg in normal chow (Specialty Feeds, SF08-026) or high fat diet (HFD) (Specialty Feeds, SF00-219).

Mouse colonies

All mice were maintained on a C57BL/6 background.

Arnt2$^{-/+}$

Mouse line was obtained from Yoishiaki Fujii-Kuriyama [15].

Arnt2$^{-/+}$ mice were generated by targeting *Arnt2* in ES cells and generating mice with germline transmission of the mutant allele. A NLS-LacZ cassette was inserted into the basic region, disrupting the bHLH domain and making ARNT2 non-functional. The NLS-LacZ was inserted so the functionally active LacZ protein is expressed, mimicking ARNT2 expression patterns [15].

Arnt sh1498 and *Arnt2* sh318

Generated during Honours Thesis research [32].

Dox inducible shRNA vectors against either *Arnt* or *Arnt2* were targeted for insertion into the *Col1a1* locus in KH2 VJ1 mES cells which were injected into Albino Swiss mouse embryos resulting in chimeric mice. Germline transmission of the targeted *Col1a1* allele was passed on through selected breeding to obtain *Arnt* and *Arnt2* shRNA colonies on a C57BL/6 background.

SIM1-GFP

The mouse strain used for this research project, STOCK Tg(Sim1-EGFP)AX55Gsat/Mmmh, RRID:MMRRC_000306-MU, was obtained from the Mutant Mouse Resource and Research Center (MMRRC) at University of Missouri, an NIH-funded strain repository, and was donated to the MMRRC by Nathaniel Heintz, Ph.D., The Rockefeller University, GENSAT.

A bacterial artificial chromosome system was sued to integrate GFP under the control of the *Sim1* promoter region into a mouse genome, generating Sim1-GFP mice [33]. The expression of the *GFP* matches the expression of *Sim1* reported in *in situ* mRNA studies of the PVN. The endogenous *Sim1* gene in not disrupted by the Sim1-GFP transgene.

***Sim1* R171H**

Generated by Whitelaw laboratory (unpublished mouse line).

A homologous recombination vector plasmid [34] containing the SIM1 mutation R171H in exon 5 of the *Sim1* gene and a LoxP flanked selection cassette in an intron was introduced to the genome of mES cells using homologous recombination. The removal of the selection cassette was achieved by crossing mice with a ubiquitous CMV-Cre expressing mouse line, followed by outbreeding to rid the mice of the CMV-Cre transgene (Whitelaw laboratory, unpublished mouse line).

Genotyping methods

Standard genotyping

Mouse ear clippings were obtained by using ear punches to create holes/notches in the ear. gDNA was extracted from the ear notches followed by genotyping PCR using either KAPA Mouse Genotyping kit (KAPA Biosystems) or MyTaq Extract-PCR kit (Bioline Australia) as per manufacturer's protocol with the appropriate genotyping primers. Standard reactions conditions: 94-95°C, 2-3min; [94-95°C, 15-45sec; 51-65°C, 15-45sec; 72°C, 30sec-1min] x35-40 cycles; 72°C, 10min. Reactions were separated by agarose gel electrophoresis and visualised by UV light.

Sim1-GFP:R171H genotyping

gDNA was extracted using the High Pure PCR Template Preparation kit (Roche), following manufacturer's protocol ('isolation of nucleic acid from mammalian tissue' followed by 'protocol for washing and elution'). gDNA was eluted in 40µl elution buffer and concentration measured by Nanodrop. Genotyping master mixes comprising of 22.5 µl SYBR Green Master Mix, 4.5 µl of each 5 µM R171H genotyping primers or Cre-pos control primers

and 50ng gDNA were made up to 45 µl with MQ water. Master mixes were aliquoted into 15 µl triplicates into a 96 well plate. Samples were analysed on an Applied Biosystems Step One Plus thermocycler. Reaction conditions: 95°C, 10min; [95°C, 15sec; 60°C, 1min] x40 cycles; 95°C, 15sec followed by melt curve analysis of 60°C, 1min, increase in sample temperature to 95°C with fluorescence measurements made every 0.5°C; 95°C, 15sec. Results were analysed using Step One software, QGene and Microsoft Excel.

Tissue sectioning and analysis

Transcardial Perfusions

Mice were administered a lethal dose of anaesthetic (Lethabarb, Virbac) of approximately 150 µl at 32.5mg/ml via intraperitoneal injection. When not responsive (established by squeezing foot with no response), PBS was injected into the heart and a cut made in the right atrium to allow returning blood to exit circulation. PBS was flushed through the mouse at 80mmHg until PBS flowed clear of blood, followed by flushing of 4% PFA at 120mmHg for approximately 5min. Following perfusion, desired tissues were dissected and incubated in 4% PFA O/N at 4°C with rotation.

Freezing tissues for sectioning

PFA fixed tissues were washed 3x in PBS followed by coating in optimal cutting temperature compound (O.C.T compound, Tissue-Tek). Tissues were placed into moulds containing O.C.T compound and frozen in isopentane cooled in an EtOH/dry ice bath. Tissue blocks were stored at -80°C.

Sectioning

Sectioning of PFA fixed tissues mounted in O.C.T compound was performed using Leica CM1900 cryostat. Sections were typically cut at 10-12µm thickness and mounted on Superfrost® Plus (Thermo Scientific) microscope slides. Mounted sections were stored between 4°C and -20°C.

Immunofluorescence of sections

O.C.T was washed off slides with 1x PBS. Tissue sections were permeabilised in 1% Triton X-100/PBS for 15min at RT followed by incubation in 10% NHS/0.3% Triton X-100/PBS blocking solution for 30min at RT. Sections were incubated with primary antibody O/N at 4°C in a humidified chamber. Following primary antibody incubation, sections were washed 3x with PBS before incubation with secondary antibody at RT for at least 2hr with gentle shaking and covered in foil. Sections were washed in PBS before mounting a coverslip onto the microscope slide using ProLong Gold antifade reagent with DAPI (Invitrogen). Slides were allowed to dry at 4°C O/N before imaging on a fluorescence microscope (Nikon Eclipse Ti microscope, Nikon Australia).

Primary cell culture

E11.5 primary mouse embryonic fibroblast (MEF) cultures

Pregnant mothers were sacrificed at day E11.5 by cervical dislocation and embryos were dissected in PBS. Embryo head and internal organs were removed and the trunk was washed with 1x PBS before incubating in 0.05% trypsin at 37°C for 5min. MEF media was added and centrifuged at 500rpm for 1min. Media was removed and the embryos were resuspended in fresh media followed by dissociation of cells with pipetting. Cells were typically plated at $1x10^5$ cells/ml on gelatin coated dishes or coverslips. MEFs were incubated at 37°C, 5% CO_2 and routinely passaged using 0.25% trypsin. Where indicated, cells were treated with 1µg/µl Dox.

E16.5 and E18.5 primary mouse cortical and hypothalamic neurons cultures

Pregnant mothers were sacrificed at either day E16.5 or E18.5 by cervical dislocation and embryos were dissected in 1x HBSS (Sigma) to remove the cortex and/or hypothalamus. The cortex/hypothalamus was digested in 0.25% trypsin for 10-15min before addition of 10mg/ml trypsin inhibitor (Sigma), followed by 3x washes in fresh Neurobasal media. The tissue was dissociated with pipetting in fresh Neurobasal media. Dissociated neurons were typically plated at $2.5x10^5$-$5x10^5$ cells/ml in 24 well trays or on coverslips that had been coated with 2µg/ml Poly-D-Lysine and washed 2x with water. Neurons were incubated at 37°C, 5% CO_2 and were maintained for indicated periods by replacing half of the culture media every 2 days. Where indicated, cells were treated with 1µg/µl Dox.

For KCl depolarisation, neurons were incubated with 55mM KCl for the indicated time. Where indicated, neurons were pre-treated O/N with 2µM Tetrodotoxin (TTX, Tocris) and 100µM D(-)-2-amino-5-phosphopentanoic acid (AP-5, Tocris).

Adult kidney cultures

Adult mice were sacrificed by cervical dislocation and the kidneys dissected. The fibrous capsule and adjacent medulla from cortical tissue, fat and connective tissues were removed. The kidneys were cut into small pieces and transferred to a sterile tube, followed by vigorous rinsing in ice cold HBSS (containing EGTA) until the solution was clear of blood. Kidney fragments were transferred to a petri dish and finely minced into approximately $1mm^3$ pieces. Minced fragments were resuspended in pre-warmed 5mg/ml dispase in HBSS (StemCell Technologies), x1 collagenase (Worthington) and 5mM $CaCl_2$ and incubated for 45min at 37°C with gentle stirring. The tissue was then dissociated using a glass pipette and passed through 40µm sieve. Cells were washed 3x in HBSS and resuspended in adult kidney media. Cells were typically seeded at $5x10^4$ cells/cm^2 on collagen or gelatin coated plates. Cells were incubated at 37°C, 5% CO_2. Media was changed 24hr after seeding and at 48hr

intervals thereafter and routinely passaged using 0.25% trypsin. Where indicated, cells were treated with 1µg/µl Dox.

Fluorescence activated cell sorting (FACS)

Pregnant mothers were sacrificed by cervical dislocation and embryos were dissected in 1x HBSS to remove the cortex. The cortex was digested in 0.25% trypsin for 10-15min followed by 3x 1min washes in trypsin inhibitor. Cells were dissociated with pipetting in cold PBS with 2% FBS and strained through 40µm cell strainer. Cells were transferred to a FACS tube for sorting on a BD FACSAria using Diva8 software. FACS was performed by Dr Iain Comerford (McColl laboratory, University of Adelaide).

Method for Analysing RNA following Intracellular Sorting (MARIS)

Primary kidney cells grown in culture were trypsinised and washed twice with PBS. Fixation and staining of cells was performed as previously described [35], using goat α-GFP primary antibody (1:1000) and alexa fluor 488 donkey α-goat secondary antibody (1:1000). Cell sorting was performed on BD FACSAria II using Diva8 software (performed by Detmold, University of Adelaide). RNA was extracted from isolated cells using TRI Reagent (Trizol) as per manufacturer's protocol, using chloroform in place of BCP. RNA was resuspended in 1xTE.

Chapter 3: Generation of shRNA knockdown mouse models to investigate unique and overlapping physiological functions of ARNT and ARNT2

Introduction

Gene targeting to generate ARNT or ARNT2 null mice resulted in embryonic or perinatal death, highlighting that both factors are essential for embryonic development. The generation of heterozygous mice, which are viable with no significant developmental defects, have limited use in studying global physiological roles of ARNT and ARNT2 when comparing to wild type animals. Although they have lower levels of ARNT or ARNT2 protein, they do not permit the study of the complete absence of the proteins at the whole organism level beyond embryonic development. Conditional knockout models in adult mice have also been limiting in researching the adult physiological roles of the proteins as the knockout has been predominately tissue specific with reports of difficulties in achieving complete abrogation at the whole organism level [28].

Adult mouse models in which there is no or very little expression of either ARNT or ARNT2 will allow further study into their physiological functions at the whole organism level, as well as their roles in disease progression. As both transcription factors have important roles during embryonic development, mouse models need to be generated in which the genes are not irreversibly depleted during the embryonic development phase, but where global loss of the protein can be invoked in the adult animal. RNA interference (RNAi) allows controlled, reversible knockdown of target proteins *in vivo* without the need to irreversibly modify the endogenous gene. This makes it a desirable system to use for investigating the effects of reduced ARNT and ARNT2 protein levels on normal physiology in an adult mouse model, in particular their functional redundancy, as well as their roles in disease progression.

shRNA mouse models

A tetracycline (Tet) response system, that was used to overexpress genes of interest, was adapted to express RNAi to control endogenous protein expression and was successfully introduced into mouse models [36]. The RNAi system employs an inducible promoter containing the Tet-Response Element (TRE), which is activated by the transcription factor TetOn, to transcribe a microRNA-30 (miR-30) embedded shRNA that targets the gene of interest. A conformational change in TetOn is induced by the addition of doxycycline (Dox), a stable analogue of tetracycline, allowing it to bind to the TRE to drive transcription of the shRNA. Endogenous microRNA (miRNA) cellular machinery is utilised to process the shRNA into a small interfering RNA (siRNA), which binds to the mRNA of interest preventing translation and ultimately resulting in the knockdown of the target protein (Figure 2). The removal of Dox deactivates TetOn which ceases shRNA production, allowing target protein expression to return to normal levels, effectively making the system reversible [36-38]. The reversibility of this system, especially in an *in vivo* situation, is highly advantageous

compared to the traditional methods previously employed to generate null *Arnt* and *Arnt2* mice where the genes are irreversibly destroyed. This will allow investigations into the effect of temporary changes in protein levels during development or disease.

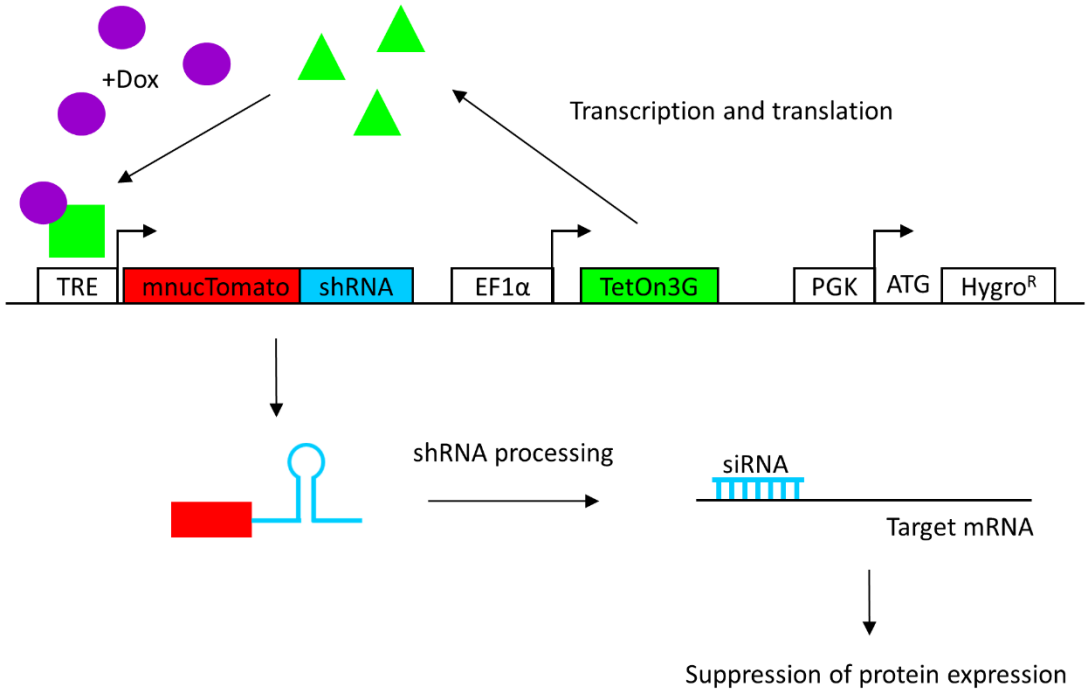

Figure 2. Schematic of doxycycline inducible shRNA system. TetOn3G transcription factor is constitutively expressed from the EF1α promoter. A conformational change is induced in TetOn3G by the addition of Dox, allowing it to bind to the TRE to drive shRNA and mnucTomato expression. The shRNA is processed by endogenous miRNA machinery to produce a siRNA that binds to the mRNA target and suppresses protein expression. The hygromycin resistance ($Hygro^R$) gene enables easy selection of cells harbouring the insert while mnucTomato expression allows the easy tracking of shRNA expressing cells. Removal of Dox prevents shRNA production, allowing protein expression to return to normal levels. System modified by Bersten et al [39].

Initial generation of RNAi mouse models involved pronuclear DNA injection of the shRNA constructs into embryonic stem cells or zygotes, which generally resulted in mice with variable knockdown [36, 40]. This method was then optimised to use FLP Recombination Mediated Cassette Exchange (RMCE) into the Collagen 1A1 (*Col1a1*) locus in previously engineered embryonic stem (ES) cells, thereby inserting single copies of the shRNA construct into the cell genome [37, 38, 41, 42]. These studies have either required breeding the transgenic RNAi mice with existing Tet transactivator mouse strains to generate the final mouse model [36, 37] or having the Tet transactivator engineered into the RMCE ES cells at the Rosa26 locus [38, 41, 42]. A number of these studies also introduced a fluorescent protein into the system that is co-expressed with the shRNA, enabling easy tracking and isolation of cells successfully expressing the shRNA in response to Dox treatment [37, 38, 42]. Work performed in the Whitelaw laboratory has further optimised this system by

generating a RMCE targeting vector that contains the shRNA, fluorescent protein and TetOn transactivator all on the one construct (Figure 3A) [39]. A TetOn variant that is more sensitive to Dox while having lower background expression, TetOn3G, was placed under the control of the EF1α promoter that produces constitutive, global expression and could alternatively be placed under tissue specific promoters in this optimised construct [39, 43]. The TRE3G promoter, a variant of the TRE promoter, was also introduced into the construct due to being optimised for lower background expression [39, 44].

We designed our experiments to use transgenic mice harbouring the modified version of the Dox inducible shRNA system which co-expresses with a monomeric nuclear red fluorescent protein (mnucTomato), targeting either *Arnt* or *Arnt2* for reversible knockdown *in vivo*.

Arnt and *Arnt2* shRNA mouse generation

Previously in the Whitelaw laboratory, shRNA sequences predicted to target different locations along the *Arnt* or *Arnt2* mRNA, were cloned into the optimised shRNA targeting vector and were stably integrated into the *Col1a1* locus in CIW mES cells [39]. During work prior to my PhD (Gerassimou, A. 2014 Honours thesis, University of Adelaide) [32], I used these modified ES cells to test the efficiency of the shRNA sequences to knockdown ARNT or ARNT2 protein at different concentrations of Dox and at different time points. Induced shRNA expression was tightly regulated in these cells, with mnucTomato expression only occurring in Dox treated cells. This confirmed that the Dox inducible system was not leaky and would not result in background expression of the shRNA, thus no significant background effects on ARNT or ARNT2 protein levels. Dox induction of mnucTomato was seen at concentrations as low as 5ng/ml, however efficient shRNA knockdown of ARNT and ARNT2 was only tested using 500ng/ml and 1µg/ml Dox. In an *in vivo* situation, these low Dox concentrations that can result in efficient knockdown of protein would be of an advantage as it should allow sufficient shRNA expression in tissues that have low delivery of Dox, such as the brain, as well as avoiding any potential toxicity effects that may result from higher doses. Efficient knockdown was seen with Dox treatments for 48 hours, however it was shown that mnucTomato expression was present 24 hours after Dox delivery [32].

From the results of this work, the shRNA sequences sh1498 for *Arnt* and sh318 for *Arnt2,* were selected to be introduced into mouse models as they were found to have potent knockdown of their respective protein at low Dox concentrations over a short time period. The selected shRNA vectors were targeted to the *Col1a1* site, which was engineered to contain two FRT sites flanking a Neo^R cassette, for integration into KH2 VJ1 mES cells (Figure 3B and 3C). Cells with successful integration of the shRNA vectors were selected with hygromycin, expanded and injected into Albino Swiss mouse embryos to generate shRNA chimeric mice (Figure 4). A total of four chimeric male pups resulted from the injections and were used for breeding mouse colonies onto a C57BL/6 background to use for this PhD project [32].

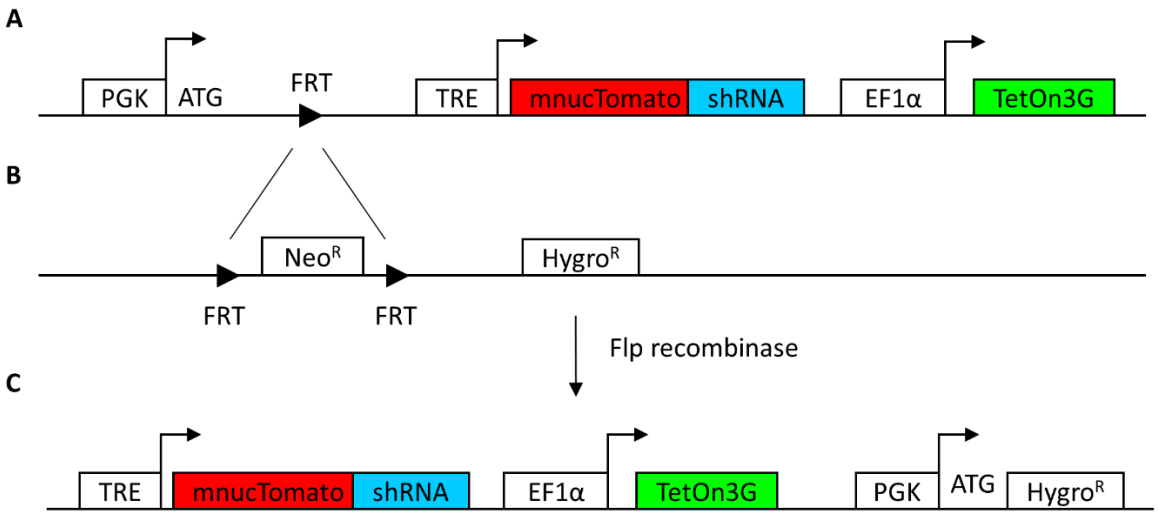

Figure 3. Integration of shRNA vector into ES cell genome. A. Schematic of shRNA vector comprising of mnucTomato and the shRNA under the control of the TRE, EF1α promoter driving TetOn3G, PGK promoter followed by an ATG codon and a FRT site. **B.** Engineered mES cells have a neomycin resistance (Neo^R) cassette flanked by two FRT sites with a $Hygro^R$ cassette downstream integrated into the *Col1a1* locus. Neomcyin is used to select cells that have the FRT sites. Due to a lack of promoter and ATG start codon, $Hygro^R$ is not possible. **C.** Flp recombinase is used to excise the Neo^R cassette and insert the shRNA vector in the FRT sites. $Hygro^R$ can be used as a selectable marker due to the presence of the PGK promoter and ATG codon introduced by the shRNA vector.

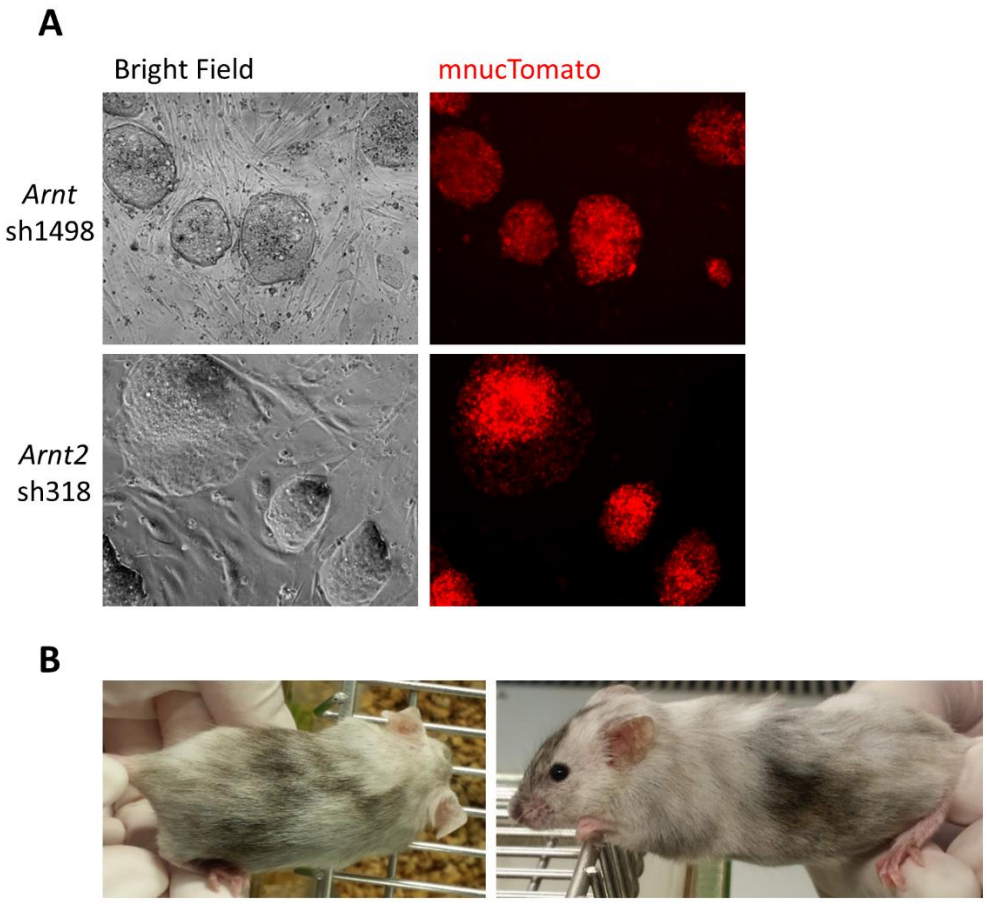

Figure 4. Generation of chimeric shRNA mice. A. KH2 VJ1 mES cells were electroporated with either *Arnt* or *Arnt2* shRNA vectors and selected with 140µg/ml hygromycin. Surviving colonies were expanded and treated with 1µg/ml Dox for 24hr to screen for mnucTomato expression. x10 magnification. **B.** Chimeric mice were generated from the injection of KH2 VJ1 mES cells harbouring the shRNA vector into Albino Swiss embryos. The chimeric mice were then breed to a C57/BL6 background [32].

Results

shRNA mouse genotyping

Confirmation of germline transmission of the shRNA construct was determined by PCR using three primers (Figure 5A). A common forward primer that sits in the *Col1a1* sequence was used with a reverse primer that also sits in *Col1a1* to amplify the WT *Col1a1* allele to produce an approximately 240bp PCR product. The common forward primer also amplified a PCR product of approximately 300bp when used with a reverse primer that sits in the shRNA construct sequence. The insertion of the shRNA construct into the *Col1a1* locus results in destruction of the reverse *Col1a1* primer site, therefore there is no amplification of WT sequence on the mutant allele. Using these three primers together in the same PCR reaction mix allowed easy detection of both WT and shRNA alleles. As seen in Figure 5B, both *Arnt* sh1498 and *Arnt2* sh318 was detected in both the heterozygous and homozygous state. Mice found to be homozygous for the shRNA alleles were breed to maintain shRNA homozygous colonies.

Compound *Arnt2*$^{-/+}$:sh318 mice were generated by the crossing of *Arnt2* sh318 homozygous mice with *Arnt2*$^{-/+}$ mice. *Arnt2*$^{-/+}$ genotyping PCR was performed using primers to simultaneously detect both WT and mutant *Arnt2*, as previously described [15]. As shown in Figure 5C, all mice from this cross breeding were heterozygous for *Arnt2* sh318, with a number of mice also positive for the *Arnt2* null allele. The compound heterozygous *Arnt2*$^{-/+}$:sh318 were then breed to obtain mice that were homozygous for the *Arnt2* sh318 allele and heterozygous for *Arnt2* mutant allele (Figure 5D).

Knockdown of ARNT in E11.5 Mouse Embryonic Fibroblasts

As ARNT is known to be expressed in mouse embryonic fibroblasts (MEFs) we decided to derive MEFs from *Arnt* sh1498 embryos to test knockdown of ARNT in primary embryonic cell culture. Primary MEFs were derived from E11.5 embryos and were treated with 1µg/ml Dox upon cell plating. mnucTomato expression was observed after 48hr of Dox treatment with no mnucTomato observed in untreated cells (Figure 6A) and knockdown of ARNT was seen by western blot (Figure 6B). Next, we maintained MEF cultures in Dox for up to five days. mnucTomato expression was stable over the timecourse (Figure 6C), however no knockdown was observed when western blot analysis was performed on the cell lysates (Figure 6D). The proportion of MEFs expressing mnucTomato at 48hr (Figure 6A) was higher than the number of cells expressing mnucTomato at two days in the extended time point experiment (Figure 6C), which would explain the difference in the level of ARNT seen in the western blots (Figure 6B and 6D). This was mostly likely due to the fact that the MEFs in Figure A and B were treated with Dox straight away upon plating when the culture was first made, while the cells used in the extended Dox time course (Figure C and D) had already been maintained in culture for a longer period of time before commencement of Dox treatment, reducing their viability.

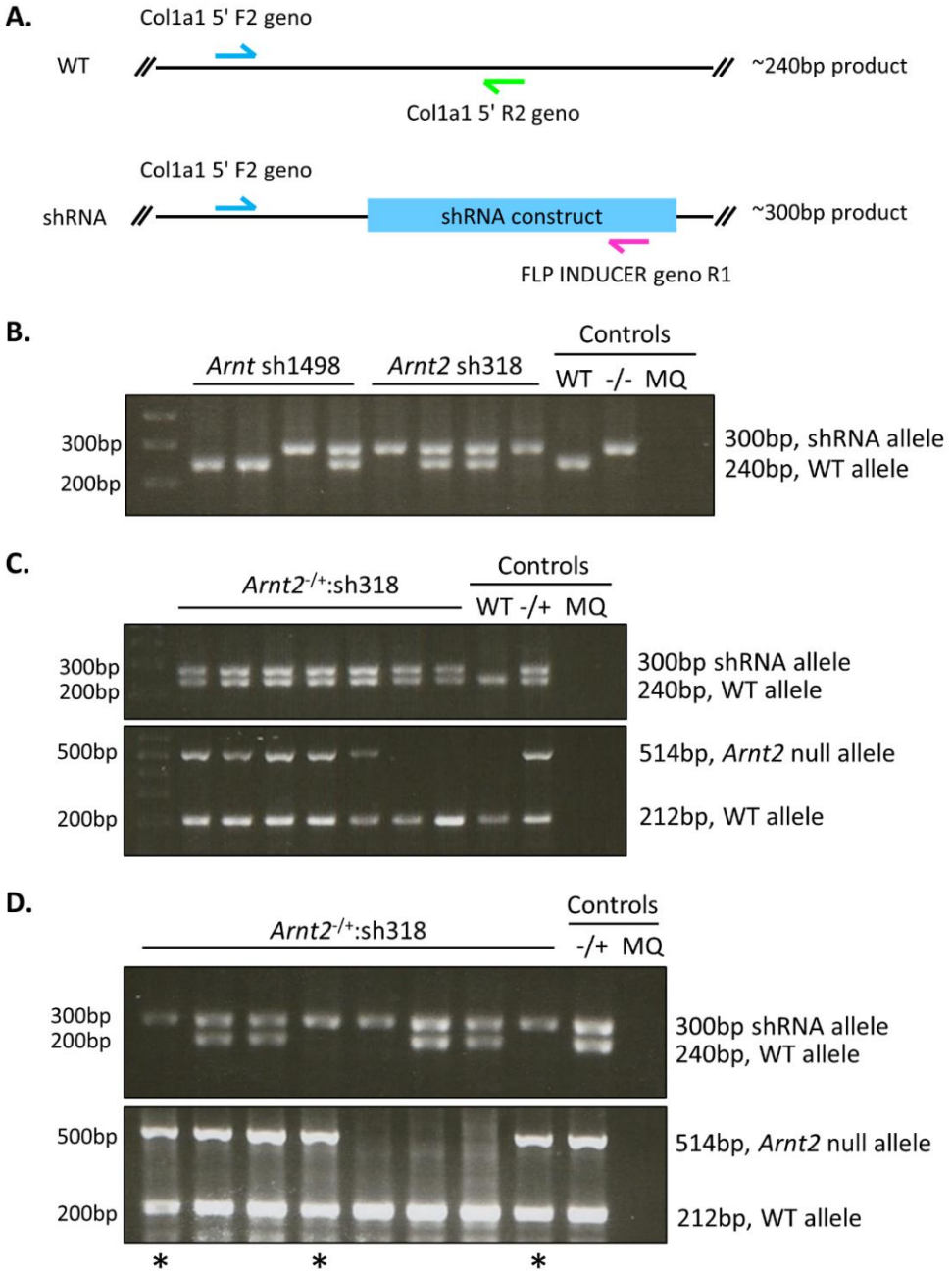

Figure 5. PCR of mouse DNA to determine genotype of *Arnt* sh1498, *Arnt2* sh318 and *Arnt2*$^{-/+}$:sh318 mice. Ear tissue taken from 3 week old pups was used for DNA extraction and PCR was performed to determine the genotype of mice. Genotyping controls included heterozygous (-/+) and homozygous (-/-) shRNA mouse DNA. **A.** Schematic of the *Col1a1* locus before (WT) and after (shRNA) insertion of the shRNA construct and the position of genotyping primers Col1a1 5' F2 geno, Col1a1 5' R2 geno and FLP INDUCER geno R1. The use of different reverse primers that bind to either WT or inserted shRNA construct sequence with a common forward primer allows differential PCR to be used for genotyping. **B.** Genotyping PCR gels showing mice that are either heterozygous or homozygous for the *Arnt* sh1498 or *Arnt2* sh318 constructs. **C.** Genotyping PCR gels for mice that are compound heterozygous for *Arnt2* sh318 and *Arnt2*$^{-/+}$. These mice were then breed to obtain mice that were compound homozygous for *Arnt2* sh318 and heterozygous for *Arnt2*$^{-/+}$. **D.** Genotyping PCR gels of pups generated from the crossing of compound heterozygous *Arnt2*$^{-/+}$:sh318 mice. Asterisks below the gel indicate compound *Arnt2*$^{-/+}$:sh318 homozygous mice used for knockdown analysis.

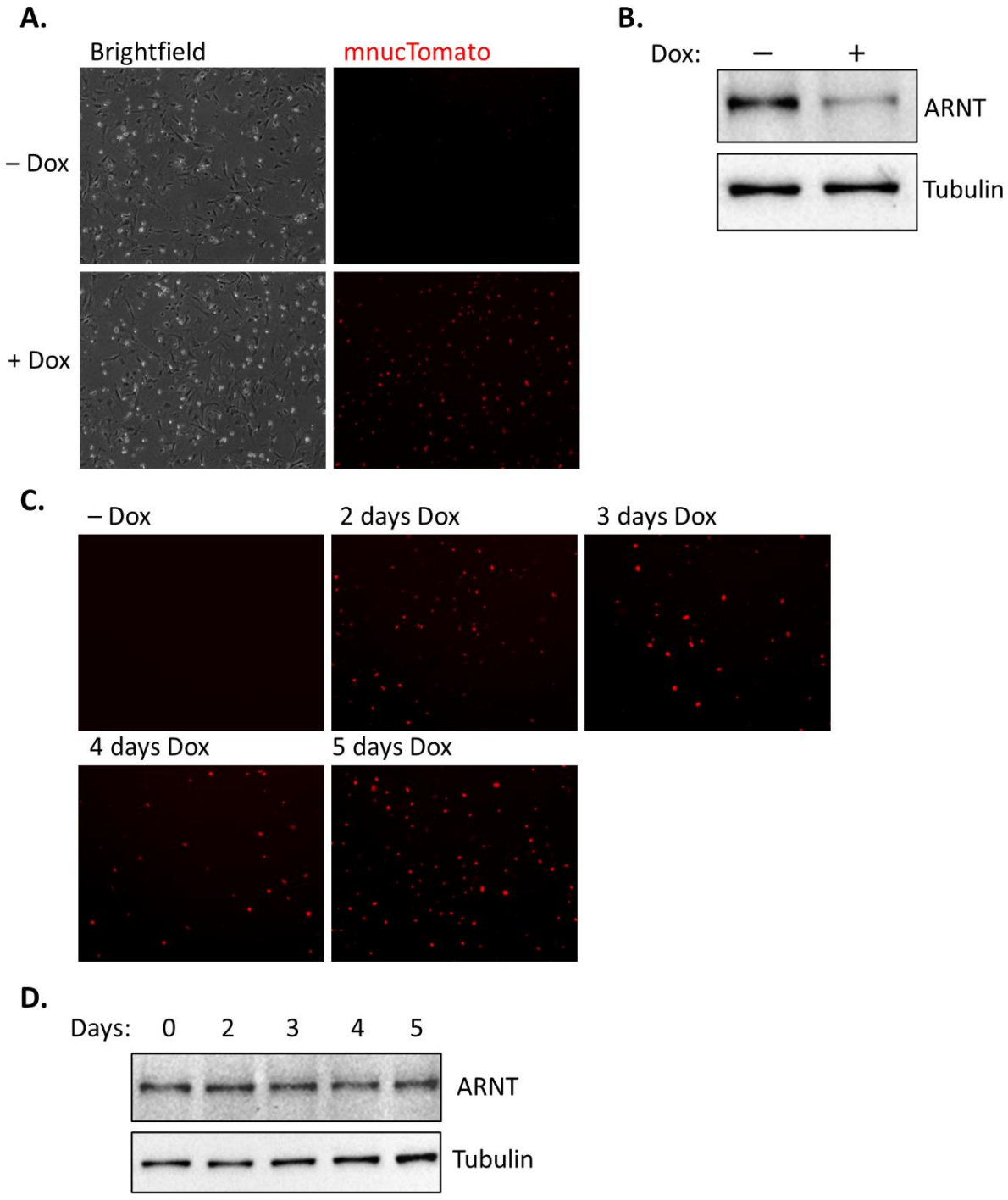

Figure 6. ARNT knockdown in primary E11.5 MEF cell culture derived from *Arnt* sh1498 embryos. A. Inducible mnucTomato expression in primary MEFs treated with 1µg/ml Dox for 48hr. x4 magnification. **B.** Western blot of shRNA knockdown of ARNT in primary MEFs treated with 1µg/ml Dox for 48hr. **C.** Inducible mnucTomato expression in primary MEFs treated with 1µg/ml Dox for 2-5 days. x4 magnification. **D.** No knockdown of ARNT was seen by western blot in MEFs treated with 1µg/ml Dox for 2-5 days.

Knockdown of ARNT2 in E16.5 and E18.5 cortical neurons

As ARNT2 is highly expressed in embryonic cortical neurons, we derived primary E16.5 cortical neurons from *Arnt2* sh318 mice and cultured them with 1µg/ml Dox. mnucTomato expression was only observed in Dox treated cells after 3 days treatment (Figure 7A), however there was no change in ARNT2 levels compared to the no Dox control when western blot analysis was performed on lysates from cells cultured for 4 days (Figure 7B) or for 8 days (Figure 7C). Since we expected a visible difference in the protein level of ARNT2 in the western blot due to the high mnucTomato expression observed, we next performed immunofluorescence (IF) on the cortical neurons using ARNT2 antibody. As can be seen in Figure 7D, cells that express mnucTomato do not show staining for ARNT2 as expected, however the majority of cells were not expressing mnucTomato and were positive for ARNT2. While this IF data does confirm that the knockdown of ARNT2 is occurring in neurons expressing the shRNA, it also explains why there is no significant reduction in overall ARNT2 levels observed in the western blot as most cells are not expressing the shRNA construct.

To avoid having a mixed culture of neurons that express and do not express the shRNA construct, we next performed FACS on embryonic cortical neurons from Dox treated pregnant females. Females were placed on Dox diet three days before being sacrificed at E16.5. Cortical neuron suspensions were made from the collected embryos and FACS sorted for mnucTomato expression before plating and being maintained in 1µg/ml Dox in culture. FACS confirmed that there were no mnucTomato positive cortical neurons collected from embryos that were not on Dox (Figure 7E), however there were two peaks seen for mnucTomato positive neurons from Dox treated embryos (Figure 7F). Upon plating, all neurons sorted as mnucTomato low were found to be dead, presumably fluorescing red due to either cell death autofluorescence or residual mnucTomato which has a long half-life. While neurons sorted as mnucTomato+ were alive at the time of plating and were all still showing red fluorescence, they did not survive prolonged culturing and were not used for further experiments (Figure 7G). FACS was not used again on primary neurons as this experiment showed that they were too delicate to be placed under the conditions required for FACS.

Primary cortical neuron cultures were also derived from E18.5 *Arnt2* sh318 embryos collected from pregnant females that were on Dox feed for three days. Upon plating, neurons were cultured in the presence or absence of 1µg/ml Dox over eight days. As seen in Figure 8A, mnucTomato expression is still maintained after 24hr and four days of Dox withdrawal, with complete loss of mnucTomato expression occurring by eight days, supporting that the system is reversible. Western blot analysis however showed that there was no knockdown of ARNT2 in lysates either made directly from the embryos on the day of collection or from extracts made from cells cultured in Dox for eight days (Figure 8B). We once again performed IF on the cortical neurons cultured in Dox for eight days, which showed the same result as the E16.5 neurons, with cells expressing mnucTomato not having ARNT2 staining while the majority of cells did have ARNT2 protein and no mnucTomato expression (Figure 8C).

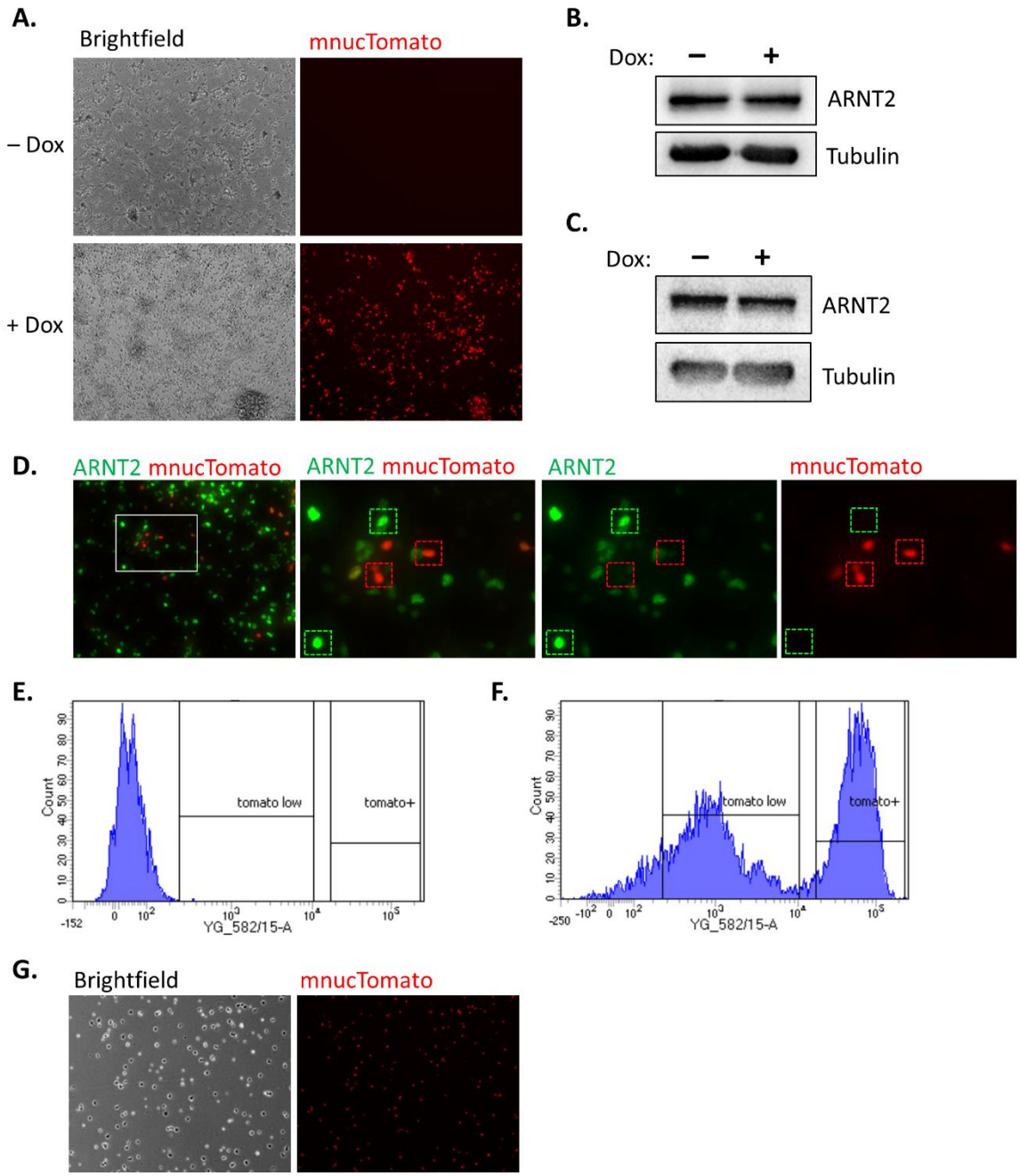

Figure 7. Dox inducible *Arnt2* sh318 expression in primary E16.5 cortical neuron cell culture. Primary cortical neurons were derived from E16.5 *Arnt2* sh318 embryos. 1µg/ml Dox was added to neurons upon plating. **A.** Induced mnucTomato expression in primary cortical neurons that were treated with Dox for 3 days. x10 magnification. **B and C.** Western blot of lysates of cortical neurons cultured with Dox for 4 days (**B**) and 8 days (**C**). **D.** mnucTomato expression and IF staining of ARNT2 in cortical neurons treated with Dox for 5 days. Cells were either positive for ARNT2 or induced mnucTomato. First panel = x10 magnification, other panels = enlargement of white boxed area. Green box = cells that only have ARNT2, red box = cells that only have mnucTomato. **E and F.** A pregnant female was fed Dox for 3 days prior to taking embryos at E16.5 for FACS analysis. FACS plots show no mnucTomato expressing cells from no Dox control embryos (**E**) and cells sorted for either low or high mnucTomato expression in Dox treated embryos (**F**). **G.** mnucTomato expression in cortical neurons 24hr after FACS sorting for mnucTomato. Cells were maintained in 1µg/ml Dox when plated after FACS. x10 magnification.

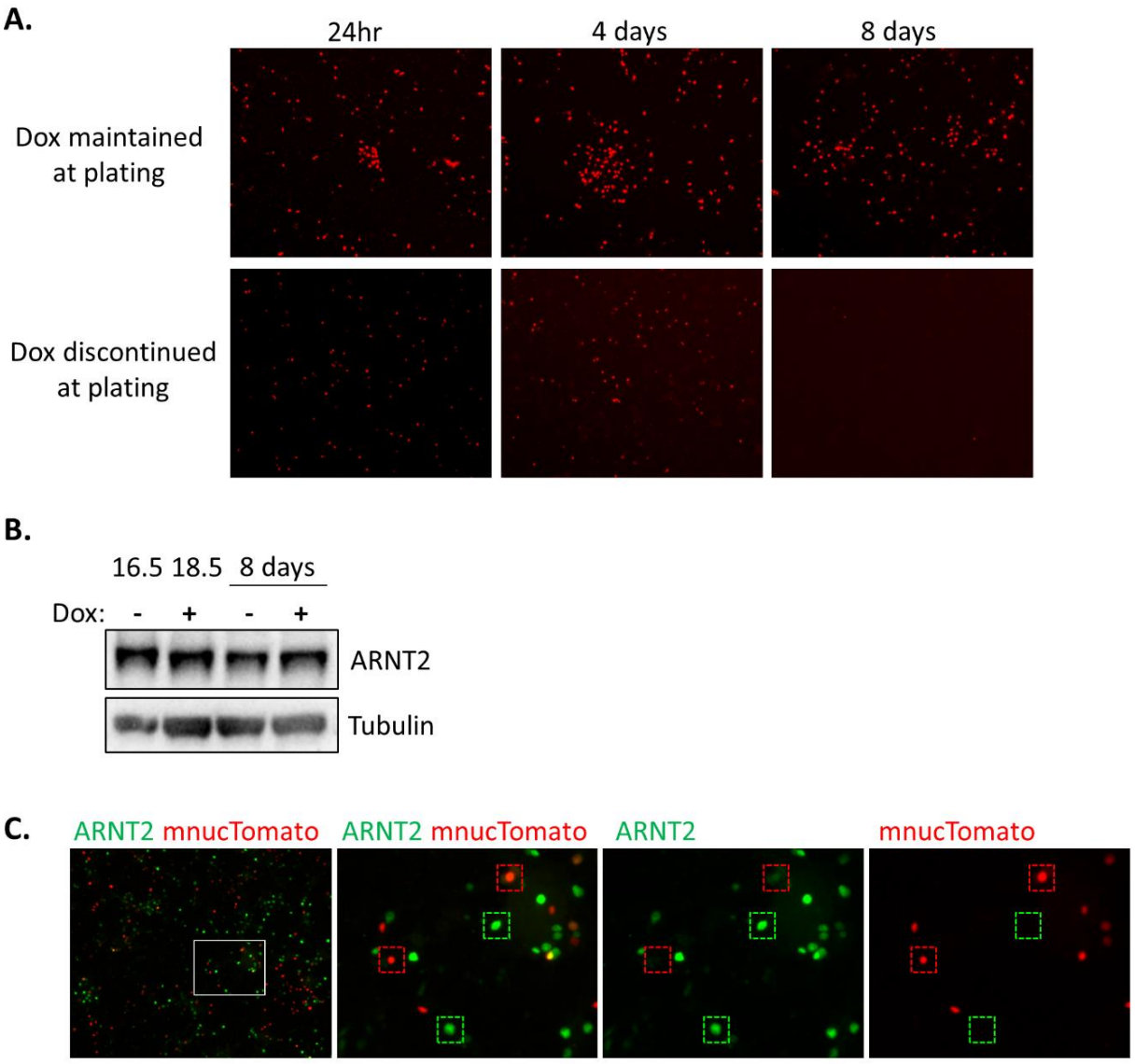

Figure 8. Dox inducible *Arnt2* sh318 expression in primary E18.5 cortical neuron cell culture. Female *Arnt2* sh318 mouse was fed Dox feed for 3 days before taking E18.5 embryos to make primary cortical neuron cultures. Cultures were maintained with or without 1µg/ml Dox upon plating. **A.** Loss of dTomato expression over 8 days of Dox withdrawal in culture. x10 magnification. **B.** Western blot showing no knockdown of ARNT2 in cortical neurons taken directly from embryos on day of collection (E18.5) or after a further 8 days of *in vitro* Dox treatment. **C.** mnucTomato expression and IF of ARNT2 in cortical neurons treated with 1µg/ml Dox for 8 days. Cells that express mnucTomato do not express ARNT2, with the majority of cells being negative for mnucTomato expression. First panel = x10 magnification, other panels = enlargement of white boxed area. Green box = cells that only have ARNT2, red box = cells that only have mnucTomato.

Knockdown of ARNT and ARNT2 in E16.5 embryos

In addition to using mice with WT endogenous ARNT2, we crossed the *Arnt2* sh318 mice with *Arnt2*$^{-/+}$ mice. As the *Arnt2*$^{-/+}$ mice only have one functional allele, they have reduced levels of ARNT2 protein compared to WT mice. By crossing our *Arnt2* sh318 mice with the *Arnt2*$^{-/+}$ mice, there will theoretically be half the amount of starting ARNT2 for the shRNA to knockdown, increasing our chances of successfully causing complete knockdown of ARNT2.

We decided to look at the knockdown of both ARNT and ARNT2 in individual embryos collected from pregnant females that were placed on Dox feed three days prior to embryo collection. Strong mnucTomato expression was seen throughout the embryos collected from both shRNA strains, including the brain, kidney, paws and head (Figure 9A and 9B).

Western blots were used to assess the level of knockdown of ARNT and ARNT2 in individual embryos. As the shRNA sequences were specific to either *Arnt* or *Arnt2* and therefore could not knockdown the other homologue, the *Arnt2* shRNA mouse line could be used as a control for ARNT knockdown, and vice versa. In the cortex, ARNT was visibly reduced in *Arnt* sh1498 embryos (Figure 9C), while ARNT2 was not convincingly decreased in *Arnt2*$^{+/+}$:sh318, nor *Arnt2*$^{-/+}$:sh318 embryos (Figure 9D). As expected, mice homozygous for the *Arnt2* knockout allele generated by the crossing of the *Arnt2*$^{-/+}$:sh318 compound mice did not have any detectable levels of ARNT2 (Figure 9D). Quantification of the westerns indicate approximately 20-55% knockdown of ARNT, while ARNT2 was knocked down by 50% at best with some embryos showing no knockdown occurring at all (Figure 9E and 9F). The higher levels of knockdown observed in the *Arnt2*$^{-/+}$:sh318 embryos were in those that were found to be heterozygous null for *Arnt2* after genotyping, therefore the reduced levels of ARNT2 could be from only having one functional allele of *Arnt2* rather than knockdown by the shRNA. ARNT expression was visibly reduced in *Arnt* sh1498 embryonic kidneys, with quantification analysis indicating a 10-80% decrease across embryos (Figure 9G and 9I). ARNT2 in the kidney had obvious knockdown seen by western blot with between 10% and 95% reduction in protein levels also occurring across embryos (Figure 9H and 9J).

As knockdown of both ARNT and ARNT2 were variable across individual embryos from within the same litters, we decided to extend the time period the pregnant females were given Dox feed from three days to six days before collecting embryos. As the western blot in Figure 9K shows, there was no significant change in the level of knockdown of ARNT occurring between three and six days in *Arnt* sh1498 embryonic kidneys. Quantification showed knockdown was 25-60% across embryos at 6 days (Figure 9L), which is within the 10-80% seen in embryos on Dox for three days.

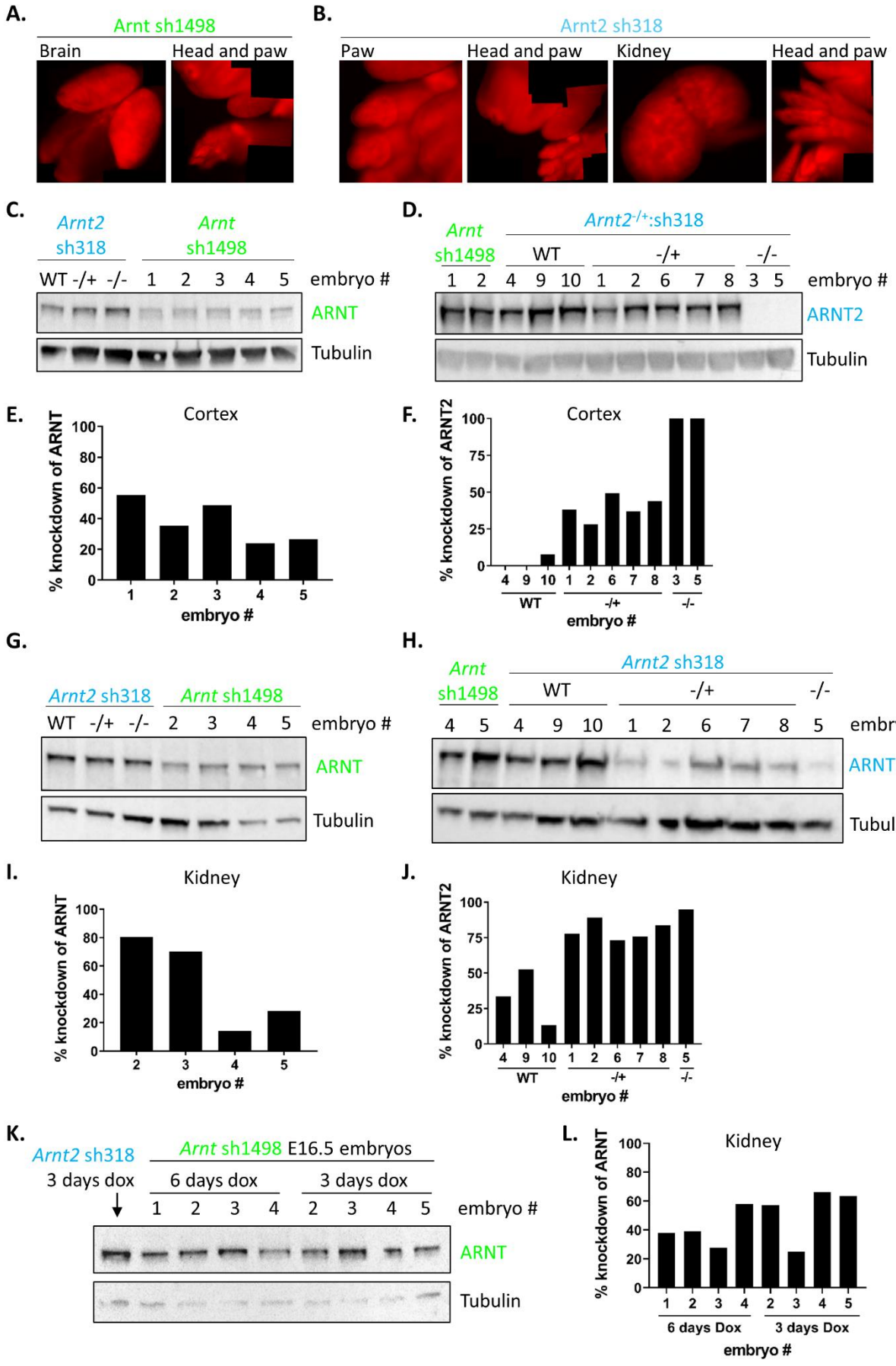
A.
Arnt sh1498
Brain
Head and paw
B.
Arnt2 sh318
Paw
Head and paw
Kidney
Head and paw
C.
Arnt2 sh318
Arnt sh1498
WT -/+ -/- 1 2 3 4 5 embryo #
ARNT
Tubulin
D.
Arnt sh1498
Arnt2-/+:sh318
WT -/+ -/-
1 2 4 9 10 1 2 6 7 8 3 5 embryo #
ARNT2
Tubulin
E.
Cortex
% knockdown of ARNT
100 80 60 40 20 0
1 2 3 4 5
embryo #
F.
Cortex
% knockdown of ARNT2
100 75 50 25 0
4 9 10 1 2 6 7 8 3 5
WT -/+ -/-
embryo #
G.
Arnt2 sh318
Arnt sh1498
WT -/+ -/- 2 3 4 5 embryo #
ARNT
Tubulin
H.
Arnt sh1498
Arnt2 sh318
WT -/+ -/-
4 5 4 9 10 1 2 6 7 8 5 embryo #
ARNT2
Tubulin
I.
Kidney
% knockdown of ARNT
100 80 60 40 20 0
2 3 4 5
embryo #
J.
Kidney
% knockdown of ARNT2
100 75 50 25 0
4 9 10 1 2 6 7 8 5
WT -/+ -/-
embryo #
K.
Arnt2 sh318
Arnt sh1498 E16.5 embryos
3 days dox
6 days dox
3 days dox
1 2 3 4 2 3 4 5 embryo #
ARNT
Tubulin
L.
Kidney
% knockdown of ARNT
100 80 60 40 20 0
1 2 3 4 2 3 4 5
6 days Dox
3 days Dox
embryo #

Figure 9. ARNT and ARNT2 knockdown in E16.5 embryonic tissues. Females were fed Dox feed for 3 days before taking embryos at E16.5. Protein lysates were made from individual embryos within the same litter. *Arnt* sh1498 was used as the no knockdown control for *Arnt2* sh318, and vice versa. **A and B**. mnucTomato expression in brain, kidney, head and paws of *Arnt* sh1498 embryos (**A**) and *Arnt2* sh318 embryos (**B**). x4 magnification. **C.** Western blot showing knockdown of ARNT in cortical neurons from *Arnt* sh1498 embryos. ARNT levels were not affected by the induction of *Arnt2* shRNA. **D.** ARNT2 knockdown was not obvious by western blot in compound *Arnt2*$^{-/+}$:sh318 embryonic cortical neurons. No ARNT2 was detected in embryos homozygous for the ARNT2 null allele. ARNT2 levels were not affected by Dox induction of *Arnt* shRNA. **E and F.** Percentage knockdown of ARNT (**E**) and ARNT2 (**F**) seen in western blots of embryonic cortical neurons. **G.** Western blot of *Arnt* sh1498 kidneys showing knockdown of ARNT. ARNT levels were not affected by the induction of *Arnt2* shRNA. **H.** Western showing a decrease in ARNT2 levels in *Arnt2*$^{-/+}$:sh318 embryonic kidneys, with no effect on ARNT2 expression when *Arnt* shRNA is expressed. No ARNT2 was detected in *Arnt2* homozygous null kidneys. **I and J.** Percentage knockdown of ARNT (**I**) and ARNT2 (**J**) seen in western blots of embryonic cortical kidneys. **K.** Western blot of reduced ARNT levels in kidneys taken from embryos collected from pregnant females that were on Dox fed for either 3 or 6 days before being sacrificed. **L.** Percentage knockdown of ARNT in 3 or 6 day Dox treated embryos.

Knockdown of ARNT and ARNT2 in adult tissues

Brain and kidney tissues were taken from adult *Arnt* sh1498 that had been on a Dox diet for two weeks. Western blot analysis of these tissue samples did not show any reduction in ARNT protein compared to tissues taken from mice on normal chow diet (Figure 10A). We then looked at Dox feeding over one, two and four weeks for both *Arnt* and *Arnt2* shRNA mice. Western blots show that the duration the mice were on the diet did not result in any knockdown of either ARNT or ARNT2 in adult hypothalamic tissues (Figure 10B).

As large pieces of adult tissue give high levels of background fluorescence when viewed under the microscope, we were unable to confirm mnucTomato expression before making the tissue lysates for western blot. To confirm the expression of mnucTomato in our Dox fed adult mice, we performed western blot using dsRed antibody to probe for mnucTomato and used lysates made from shRNA embryos that had visible mnucTomato expression under the microscope as positive controls. As seen in Figure 10C, there was no expression of mnucTomato in the cortex from either *Arnt* or *Arnt2* shRNA adult mice. Very low mnucTomato expression in adult tissues was then confirmed using primary cell culture. We derived primary kidney cultures from both *Arnt* sh1498 and *Arnt2* sh318 adult mice and maintained the cells in 1µg/ml Dox for four days. As seen in Figure 10D, very few cells were expressing mnucTomato in cultured primary kidney cells for both shRNA strains.

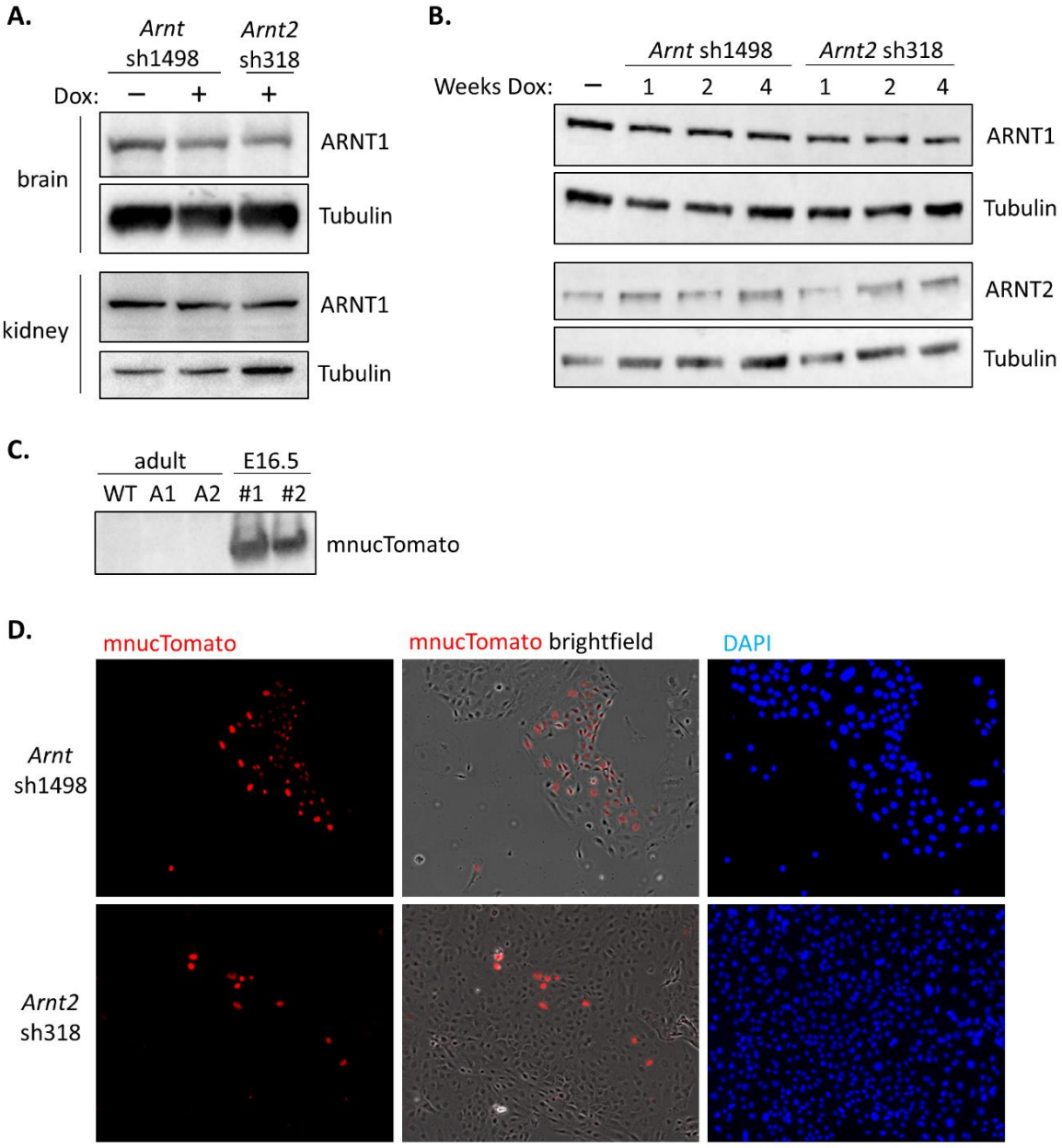

Figure 10. shRNA induction in *Arnt* sh1498 and *Arnt2* sh318 in adult mouse tissues. A. Western blot of ARNT in brain and kidney tissue extract from mice fed Dox for 2 weeks. There was no shRNA knockdown of ARNT in Dox fed *Arnt* sh1498 mice. ARNT levels were not affected by the expression of shRNA targeting *Arnt2*. **B.** Western blot showing no knockdown of ARNT or ARNT2 in hypothalamic tissue extract from either *Arnt* sh1498 or *Arnt2* sh318 adult mice fed Dox for 1, 2 or 4 weeks. **C.** Western blot showing a lack of mnucTomato expression in the cortex of adult *Arnt* and *Arnt2* shRNA mice fed Dox for 1 week. Lysates made from embryos with visible expression of mnucTomato were used as a control for dsRed antibody staining of mnucTomato. **D.** Low expression of mnucTomato in primary kidney cell culture derived from *Arnt* sh1498 and *Arnt2* sh318 adult mice and cultured in 1µg/ml Dox for 4 days. Nuclei are stained with DAPI. x10 magnification.

Discussion

Generation of ARNT and ARNT2 null mice have been lethal, establishing the importance of both proteins being required during embryonic development. While heterozygous mice have been useful to a limited degree for the study of reduced ARNT and ARNT2 levels, there is a need for a mouse model to study the effect of the complete absence of either protein on adult physiology and disease progression. This models needs to keep the *Arnt* and *Arnt2* genes intact to allow for normal development of the embryo. In this study we have attempted to generate inducible shRNA mouse models that do not destroy the *Arnt* or *Arnt2* genes, therefore allowing normal development of embryos with controlled knockdown of ARNT or ARNT2 occurring upon Dox treatment. An advantage of such models are that they will allow temporal control of the knockdown that can easily be reversed by withdrawing Dox treatment, making effective tools for experiments looking at temporary fluctuations in protein levels.

Previously, the Whitelaw laboratory was able to show inducible co-expression of mnucTomato and shRNA sequences targeting *Arnt* and *Arnt2* that resulted in the knockdown of the target proteins in mES cell culture. The TetOn variant used, TetOn3G, is more sensitive to Dox compared to other variants, meaning lower Dox concentrations should still result in efficient knockdown, making it ideal for use in animal models where there will be tissues with low Dox delivery. We then took the shRNA sequences that gave the most potent knockdown for the two proteins and generated shRNA mouse models.

Efficiency of shRNA expression and knockdown in mouse tissues and primary cell culture

The Dox induced expression of the shRNAs were tightly regulated by Dox treatment as determined by the expression of mnucTomato. In both primary cell cultures and whole embryos, mnucTomato expression was only seen in the presence of Dox. The absence of mnucTomato in non-treated cells indicated no background expression was present and therefore no significant background effects on the expression of ARNT or ARNT2. The inducible system was also confirmed to have reversible mnucTomato expression, with withdrawal of Dox treatment from primary cell cultures resulting in a decrease in mnucTomato to non-detectable levels.

Dox delivered to pregnant mice via food pellets alone was enough to induce strong mnucTomato expression throughout the whole E16.5 or E18.5 embryo after just three days, with consistent expression seen across embryos from all litters collected for analysis. Despite this, there was high variability in knockdown of both ARNT and ARNT2 between embryos from within the same litter (Table 1). The consistent mnucTomato expression across individual embryos indicated they were receiving the same levels of Dox, therefore we would presume that the embryos should have been expressing shRNA to the same level which should have been reflected in the western blots. Extending Dox treatment to six days

did not have any effect on the levels of inducible knockdown compared to three days Dox in our embryos. Crossing our *Arnt2* sh318 mice with a mouse line heterozygous for *Arnt2*, which already have reduced ARNT2 compared to WT mice, also did not have any knockdown effect on the levels of total ARNT2 seen. We expected that this time frame of 3-6 days would have been sufficient to induce strong knockdown of ARNT and ARNT2 due to the results of Premsrirut's group [38]. In their study, they introduced a shRNA targeting luciferase, which was co-expressed with GFP, into mice that had global expression of luciferase. Dox treated embryos were seen to have global GFP expression with a corresponding decrease in overall luciferase levels within 2-6 days of Dox treatment.

Table 1. Summary of percentage of protein knockdown in cortical and kidney tissues from E16.5 *Arnt* and *Arnt2* shRNA embryos.

Tissue	ARNT or ARNT2	% knockdown	# days on Dox
Cortex	ARNT	20-55	3
Cortex	ARNT2	0-50	3
Kidney	ARNT	10-80	3
Kidney	ARNT	25-60	6
Kidney	ARNT2	10-95	3

While embryonic knockdown would be useful for investigation into ARNT and ARNT2 during development, the main reason for generating our mouse models was to investigate the effects of the loss of ARNT and ARNT2 in adult physiology. These mice would be particularly useful to help answer questions about the role of ARNT with HIF and SIM2 in cancer or with HIF and AhR in inflammatory bowel disease and the role of ARNT2 with NPAS4 in learning and memory or appetite control with SIM1. Hence, we wanted to see convincing knockdown of the two proteins in adult tissues. Unfortunately, we did not see any knockdown of ARNT or ARNT2 in either adult brain or kidney tissues with up to four weeks of Dox treatment delivered through food, nor did we see mnucTomato expression in adult tissues by western blot. In the study performed by Premsrirut et al [38], they observed GFP expression with clear knockdown of luciferase after just four days of Dox treatment in almost all adult mouse tissues. They then went on to show that GFP expression and luciferase knockdown was completely reversed within 12 days of Dox withdrawal. Another two independent studies also showed up to 80% reduction of their target proteins with corresponding GFP expression in most tissues in adult mice with 1-3 weeks of Dox feed [45, 46]. While we did not test Dox penetration in various tissues in our mice, these previous pieces of data indicate that the Dox delivery to most tissues should have been successful throughout our mice, with a corresponding knockdown of target protein, within the timeframes used for our Dox treatments in both embryos and adults.

In addition to testing knockdown efficiency directly from Dox treated embryonic and adult tissues, we also tested primary cultures derived from our shRNA mice. Primary embryonic MEF and cortical neurons and adult kidney cells were cultured in a monolayer, thus all cells received a more equal Dox delivery compared to cells *in vivo*. In spite of this, not all cells treated *in vitro* exhibited mnucTomato expression, with western blot analysis of ARNT and ARNT2 showing no change in levels. IF analysis of embryonic derived cultured cells showed that the majority of the cells were not expressing mnucTomato and consequently were still expressing either ARNT or ARNT2. The number of mnucTomato negative cells significantly outnumbered the mnucTomato positive cells upon Dox treatment for both embryonic and adult cells, thus no overall significant change in protein levels could be detected in our analysis. Increasing the time the cultured cells were treated had no impact on the levels of mnucTomato and ARNT or ARNT2 seen. Initial testing of the shRNA sequences in an established mES cell line to identify those that gave potent knockdown were performed using both 500ng/ml or 1µg/ml of Dox [32]. Dox was used at 1µg/ml for all primary cell cultures performed in this study, therefore we would expect to see effective knockdown of both ARNT and ARNT2 with very little protein being detected at most. Additionally, previous studies that derived primary MEFs from shRNA transgenic mice were able to show near complete knockdown of their target protein with similar *in vitro* Dox treatments as those performed in this study [36, 38].

The results of this study show that mnucTomato and shRNA inducible expression was mosaic despite the mice being homozygous for the shRNA construct and genotyping confirming the transgene was still present in the genome of the mice. Dox was used at known effective concentrations and while Dox delivery to all tissues may have been a limiting factor *in vivo*, this is not the case for primary cell culture where cells are grown in monolayer. Initial testing of the shRNA sequences showed almost complete knockdown of both ARNT and ARNT2, therefore the lack of knockdown seen in mice and primary cells was not due to using unreliable shRNA sequences. Given all of this, we therefore suspect that either silencing of the transgene or shutdown of mnucTomato/shRNA processing through unknown mechanisms is responsible for the lack of widespread expression and activity of the system in the presence of Dox in our mouse models.

Previous studies utilising viral delivery vectors to introduce shRNA constructs into cell genomes have been shown to be susceptible to epigenetic gene silencing and promoter interference, resulting in heterogeneous expression of shRNA constructs and consequently variable knockdown [47]. While our mice were not generated using a viral delivery system, similar mechanisms of gene silencing could potentially be at play in our mouse models. Alternatively, degradation of the mnucTomato/shRNA transcript or limited availability of miRNA cellular machinery could result in significantly decreased number of processed transcripts that are not capable of meeting the thresholds required for knockdown.
Since the initial generation of shRNA germline mouse models, there have been very limited reports of further mouse models being generated. The models reported have resulted in

80% knockdown at best in selected tissues [45, 46]. The lack of studies reported using this system may indicate that other research groups have encountered similar issues in inducing knockdown when using germline shRNA mouse models. Consequently there are no published investigations into why shRNAs that work extremely well in cell culture no longer work in germline animal models.

As our mouse models were clearly not knocking down ARNT and ARNT2 to levels that would make downstream experiments viable, we did not investigate the mechanisms for loss of mnucTomato or shRNA expression and activity any further.

Other approaches to generate knockdown mouse models.

While shRNA knockdown does not result in the complete elimination of target protein as knockout models would [38], an shRNA mouse system would be ideal to study the effects of near complete reduction of ARNT and ARNT2 in adult physiology and disease progression. The reversibility of the system would also allow investigation into the effects of drug treatments that are designed to inactivate either of these proteins.

Unfortunately, as we were not able to observe any knockdown of either ARNT or ARNT2 in adult mice and knockdown seen in embryos was too variable, it was concluded that experiments using these models would not be reliable. As the only knockdown we saw was variable, this would likely have generated questionable, non-reproducible results.

A series of experiments were performed by Fellmann *et al* [48] to optimise the shRNA backbone, resulting in a modified backbone that allows more efficient processing into siRNA, therefore more efficient knockdown of target genes. This optimised backbone was termed miR-E. They were able to show a significant increase in the efficiency of knockdown of target proteins when using the miR-E backbone compared to miR-30. Previous work done in the Whitelaw laboratory has also shown an improvement in knockdown of HIF1α in mES cells using the miR-E backbone compared to miR-30, however the improvements were not enough to result in complete knockdown of HIF1α [32]. While generating shRNA mouse models using miR-E may improve the efficiency of induced shRNA processing and therefore should result in improved knockdown of ARNT and ARNT2, the mice would most likely experience the same limitations as those generated in this study using the miR-30 backbone.

In the past 15 years, there have been very few studies published using germline shRNA mouse models. These studies have shown variable levels of incomplete knockdown across different tissues [36, 38, 45, 46]. Rather than generating germline shRNA mouse models, most studies employing shRNA technologies have predominately used various viral delivery systems to induce transient knockdown of target genes. Consequently, the generation of new mouse model systems that allow finely controlled expression of target genes over long periods of time are required. The development of new CRISPR technologies over the past

decade, such as CRISPR interference (CRISPRi), would be more likely to result in the quick generation of more reliable knockdown mouse models. Recent work has used deactivated Cas9 (dCas9), which has lost its DNA nuclease ability, fused to the transcription inhibitor KRAB to downregulate target gene expression without irreversibly modifying the gene of interest. Single guide RNA (sgRNA) guides the fused dCas9-KRAB protein to the target gene of interest, where it blocks transcription at the transcription start site [49-51]. Multiplex sgRNA systems are also being developed to allow the downregulation of multiple genes to investigate more complex pathways [50, 52]. Results of studies using CRISPRi have indicated that the knockdown of target proteins is more efficient compared to established RNAi technologies such as the shRNA system employed in this study [50, 51]. This could be due to CRISPRi preventing transcription of the target gene while the shRNA initiates degradation of an already expressed mRNA. Expression of either the dCas9-KRAB or the sgRNA can be placed under the control of chemical inducible promoters, such as Dox, to allow temporal control of gene inactivation, or more recently shown, under the control of light sensitive systems for even tighter regulation of gene expression [49]. Combining CRISPRi with different inducible expression systems as well as cell specific promoters will vastly enhance the temporal and spatial control of target gene expression. Not only will using a system such as CRISPRi for the knockdown of ARNT and ARNT2 allow easier investigation into the role of these proteins in adult physiology and disease progression, but will also allow further investigation into more complex pathways by the knockdown of multiple components.

Chapter 4: Investigating the mechanism of reduced activity of the human obesity linked SIM1 mutant R171H

Introduction

The generation of SIM1 knockout mouse models, which results in pup death within 24hr of birth, has established the importance of the role of SIM1 during development, in particular its role in the correct development of the hypothalamus [53, 54]. Post-development, SIM1 has been shown to have a role in the hypothalamic regulatory pathway that controls appetite, with deficient SIM1 resulting in hyperphagic obesity, while increases in SIM1 rescues genetically induced obesity phenotypes [55-62]. Loss of function mutations in SIM1 have been established as a monogenic cause of obesity in both humans and mice. However, the target genes and both up- and downstream regulatory pathways of SIM1 are currently unknown. As such, further biochemical investigation is required to elucidate novel target genes, regulation signals and interacting co-factors of SIM1, as well as how mutations in SIM1 affect its function. Understanding these mechanisms underlying SIM1 function will allow the development of effective therapeutics for the treatment of obesity.

Screening studies of both human patients with early onset obesity, as well as obese mice, have resulted in the identification of SIM1 variants which have been further analysed for their activities on a SIM1 reporter gene [30, 63]. A number of these mutations, which were located in the dimerisation interface of SIM1, were found to be loss of function mutations. One of the variants identified in these studies was the mutation R171H, which sits in the PAS A domain of SIM1, and was used in this study for further investigation into SIM1 function.

SIM1 R171H mouse model

The SIM1 mutation R171H was identified in a screen of early onset obesity patients and is known to be present in at least three carriers [30]. Previous studies performed in the Whitelaw laboratory showed a significant decrease in SIM1 R171H activity compared to WT SIM1 on a SIM1 responsive reporter gene [30, 63], with an approximately 70-80% decreased in activity. Hence, R171H is likely a monogenetic driver of obesity.

As the R171H mutation causes severe deficiency in function, it was selected to be introduced into a mouse model to aid investigation into the mechanisms of function of SIM1. Using homologous recombination, a plasmid containing the R171H mutation in exon 5 of the *Sim1* gene and a LoxP flanked Neo^R selection cassette in the following intron was introduced into the genome of mES cells. Removal of the Neo^R selection cassette was achieved by crossing mice with a ubiquitous CMV-Cre expressing mouse line (Figure 11), after which the CMV-Cre transgene was subsequently bred out of the final C57/BL6 SIM1 R171H mouse model (Whitelaw laboratory, unpublished mouse line).

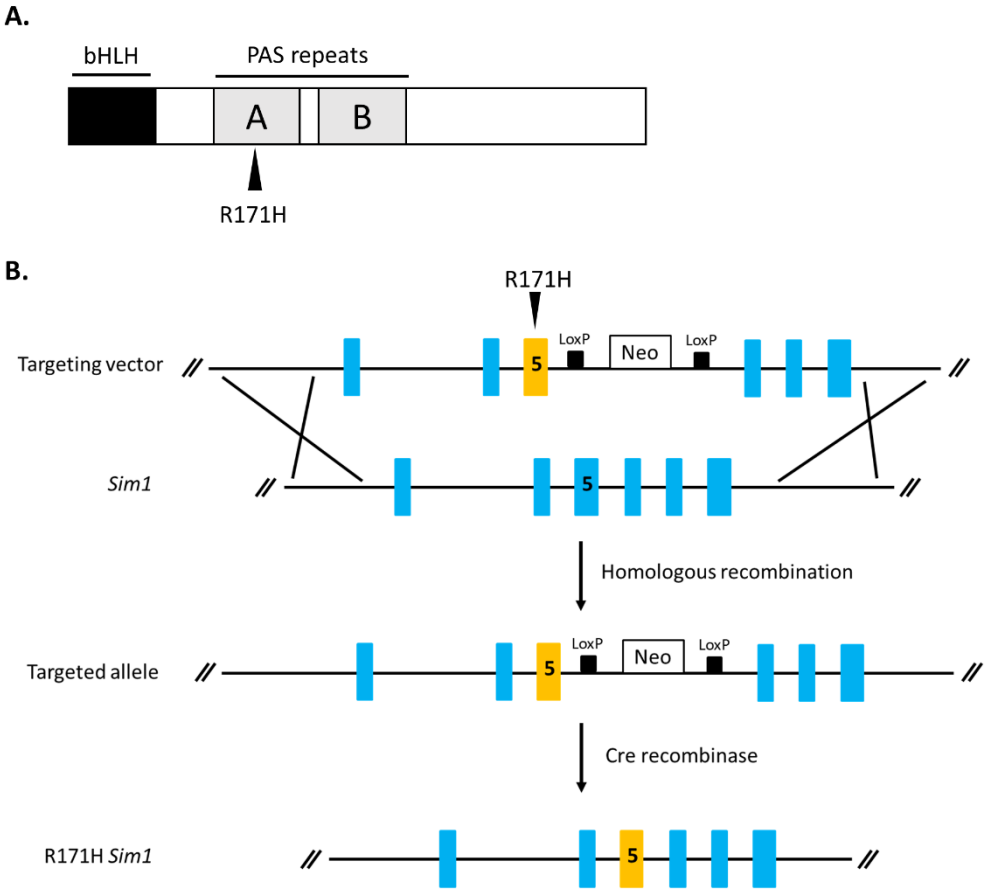

Figure 11. Schematic of generation of *Sim1* R171H mice. A. Position of the R171H mutation along SIM1. **B.** *Sim1* targeting vector has the R171H mutation in exon 5 (orange box) of the *Sim1* gene followed by a LoxP flanked Neo^R selection cassette in the following intron. Mice were crossed with a ubiquitous Cre-recombinase expressing mouse line to remove the Neo^R selection cassette (Whitelaw laboratory, unpublished mouse line).

GFP mouse model

There are currently no antibodies available that are capable of detecting endogenous SIM1. In order to be able to identify cells that express SIM1 protein, we obtained a Sim1-GFP mouse line from GENSAT. To generate this mouse line, a bacterial artificial chromosome system was used to integrate GFP under the control of the *Sim1* promoter region into the mouse genome [33]. The *Sim1* promoter controlling GFP expression allows the easy identification of SIM1 expressing cells by detecting GFP expression. An advantage of using Sim1-GFP mice over using a tagged SIM1 mouse line is that the endogenous *Sim1* gene itself is not disrupted by the Sim1-GFP transgene. This allowed us to cross the Sim1-GFP mice with our R171H mutant mice in order to generate Sim1-GFP:$Sim1^{R171H}$ mice to further investigate the effect of the R171H mutation on the function of SIM1, as well using GFP labelled cells to attempt identification of novel SIM1 target genes.

Results

Confirmation of Sim1-GFP expression in adult brain and kidney

It has been established that *Sim1* mRNA is expressed in embryonic and adult brain and kidney, however there is currently limited expression analysis of the SIM1 protein in adult tissues, partly due to the lack of a SIM1 antibody that can detect endogenous protein. In order to visualise and isolate SIM1 expressing cells, we decided to employ the Sim1-GFP mouse model.

To confirm expression of GFP was under the control of the *Sim1* promoter in the Sim1-GFP mice, we perfused and harvested the brain and kidney of adult mice for tissue sectioning and subsequent immunostaining for GFP. As seen in Figure 12A, there is clear GFP expression in the PVN and nucleus of lateral olfactory tract (NLOT) within the brain, matching *in situ* RNA data from the Allen Mouse Brain Atlas (http://mouse.brain-map.org/) (Figure 12B). However, due to damage in the sections where the SON is located, we were unable to definitively confirm expression of SIM1 in this subset of neurons.

Sim1 RNA has been detected in both embryonic and adult kidney tissues by northern blot analysis, however expression within the kidney had only been narrowed down to the developing tubules using *in situ* analysis of E16.5 embryonic mice [5, 18]. Using immunofluorescence (IF), we were able to show Sim1-GFP expression in the tubules of the adult kidney (Figure 12C), indicating that SIM1 expression is maintained in the tubules throughout kidney development and into adulthood.

ARNT2 has been established to be the preferred dimerisation partner of SIM1 within the brain to carry out its function. Therefore, in order to further confirm expression of GFP in the correct cells, we performed IF on Sim1-GFP mouse brain and co-stained with GFP and ARNT2 antibodies. As seen in Figure 12D, there is clear co-expression of GFP and ARNT2 within the same cells of the PVN, further confirming the correct expression pattern of the Sim1-GFP construct.

Sim1 R171H mouse model

Confirmation of germline transmission of the *Sim1* R171H mutation was determined by PCR, using two primers that give different product sizes depending on the absence or presence of a LoxP scar (Figure 13A). The LoxP scar is a result of the Cre-recombination that occurs to remove the Neo^R selection cassette during the generation of the mouse line. In WT mice, the primers give a product of approximately 680bp while the LoxP scar increases the distance between the two primers giving a product of approximately 780bp in mice harbouring the R171H mutation (Figure 13A and 13B). This genotyping strategy allows easy detection of the R171H mutation in both the heterozygous and homozygous state.

As R171H was identified in a study of obesity patients and proven to weaken activity of SIM1 on a reporter gene, we wanted to confirm that the mutation was functionally deleterious enough to drive obesity when introduced into a mouse model. We placed 12 week old adult female R171H homozygous and WT mice on a HFD. As seen in Figure 13C, the mice homozygous for R171H had significant weight gain compared to WT mice, whose weight increased marginally after four weeks on a HFD. Figure 13D shows the physical difference in size between the WT and R171H homozygous mice after 4.5 weeks on a HFD.
This SIM1 R171 mouse model establishes that non-synonymous point mutations in the SIM1 gene have the ability to underpin monogenic forms of obesity.

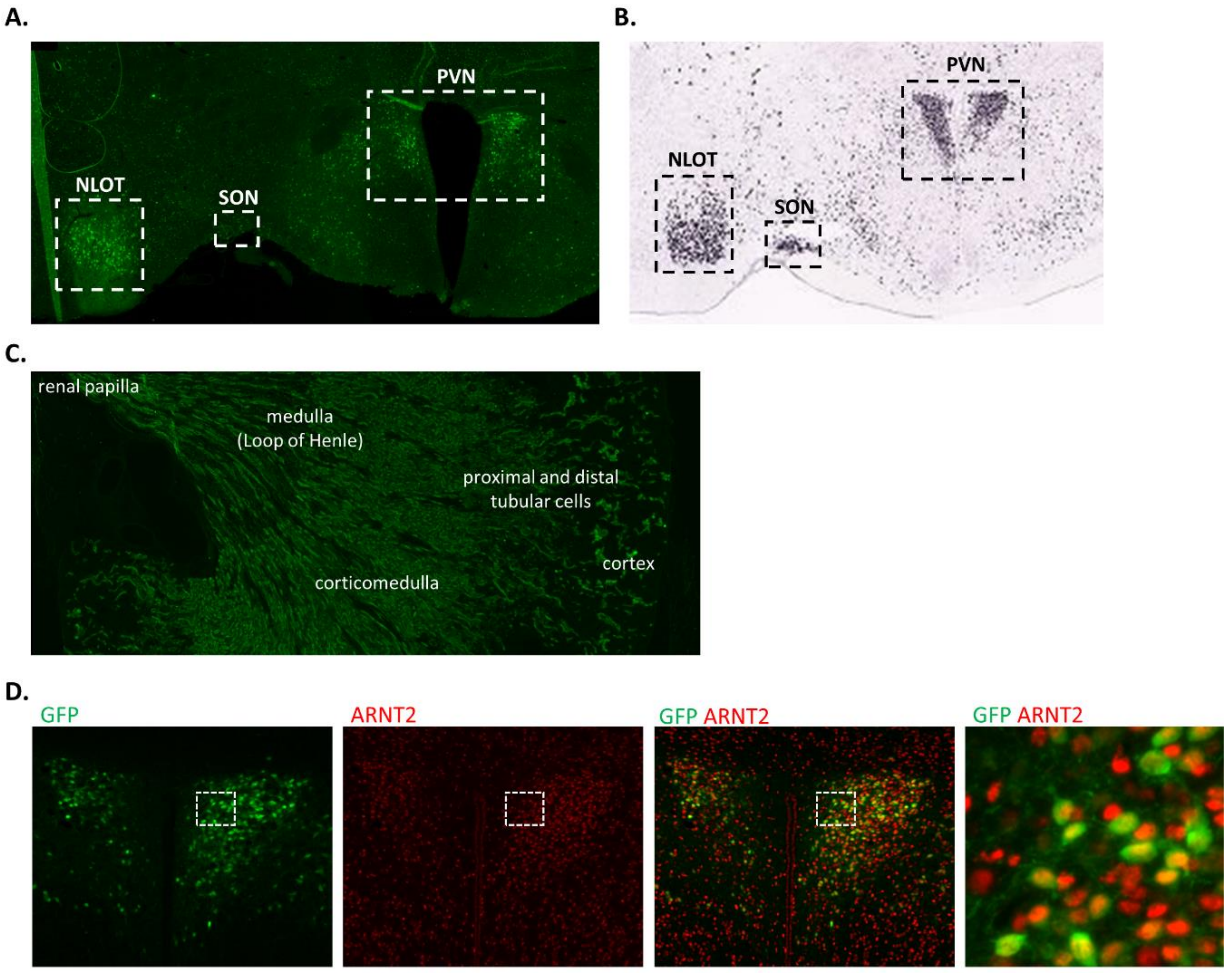

Figure 12. Sim1-GFP expression patterns in adult mouse brain and kidney. A. Sim1-GFP adult mouse brain immunostained for GFP. x4 magnification. PVN: paraventricular nucleus, SON: supraoptic nucleus, NLOT: nucleus of lateral olfactory tract. **B.** *In situ* hybridisation of *Sim1* mRNA in the adult brain of *Sim1*-Cre dtTomato mice (Allen brain atlas, http://mouse.brain-map.org/). **C.** Sim1-GFP adult kidney immunostained for GFP. x4 magnification. **D.** Co-expression of Sim1-GFP and ARNT2 in the PVN of adult Sim1-GFP mice. x10 magnification, last panel = enlargement of white boxed area. Images representative of three mice.

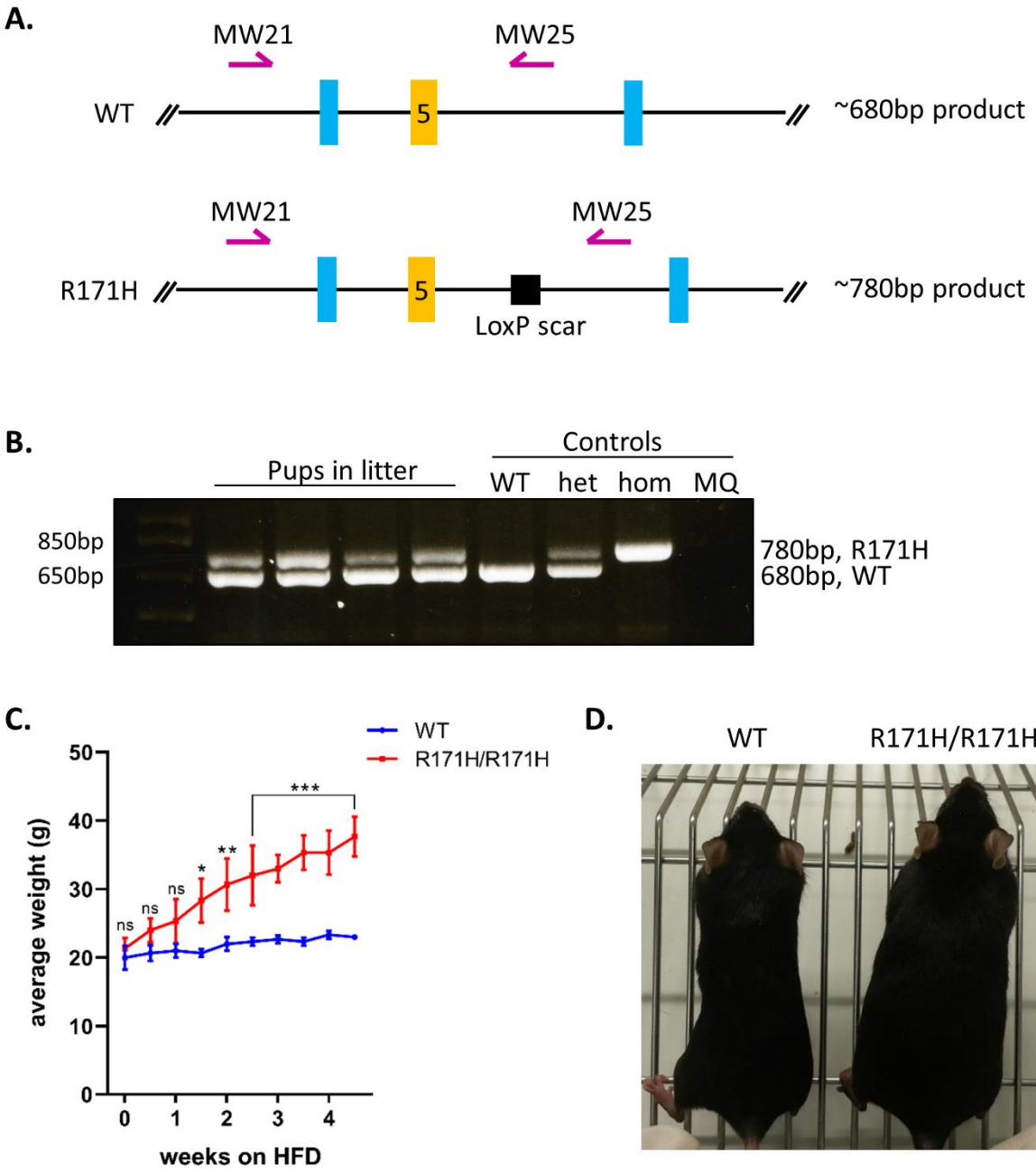

Figure 13. *Sim1* R171H genotyping and weight gain on HFD. **A.** Schematic of the *Sim1* locus before (WT) and after (R171H) Cre-Lox recombination and the position of genotyping primers MW21 and MW25. Cre-Lox recombination used to delete the intronic Neo selection cassette from the R171H mutation locus (exon 5, orange box) results in a LoxP scar, increasing the distance between the genotyping primers and allowing differential PCR to be used for genotyping. **B.** Ear tissue taken from 3 week old pups was used for DNA extraction and PCR was performed to determine the genotype of mice. Genotyping PCR gels showing mice that are either WT, heterozygous or homozygous for *Sim1* R171H. **C.** Average weight of 12 week old female WT and R171H/R171H mice on a HFD. WT n = 3, R171H/R171H n = 3. ns = not significant, * P <0.001, ** P <0.0002, *** P <0.0001. **D.** Photo of WT and R171H/R171H mice after 4.5 weeks of being fed a HFD.

Effect of R171H on SIM1 function

As the R171H mutation causes significant weight gain in mice (Figure 13) and reduced activity on reporter gene assays [30, 63], we decided to look into the mechanism of loss of function. As the R171H mutation occurs in the PAS A domain of SIM1, which is involved in dimerisation with its partner protein ARNT2, a co-immunoprecipitation (Co-IP) was performed to assess the ability of the mutant SIM1 to dimerise with ARNT2. As seen in Figure 14A, R171H SIM1 resulted in the Co-IP of ARNT2 levels that are comparable to that of WT SIM1, indicating that there is no impairment in the ability of SIM1 R171H to dimerise with ARNT2. We next performed Co-IP to look at the ability of SIM1 R171H to interact with orthopedia homeobox (OTP). OTP is a second hypothalamic, obesity related protein [64] which has been identified to interact with SIM1 in experiments performed in the Whitelaw laboratory (unpublished data). Western blot of the Co-IP in Figure 14B shows that there is no difference in the abilities of the SIM1 WT or SIM1 R171H to bind to OTP.

SIM1 is established to play a role in the correct development of the PVN, including the correct differentiation of a number of cell lineages. Therefore, we next looked at the hypothalamus of adult mice to look at differences in PVN structure and cell number to elucidate whether improper development of the PVN was responsible for the increase in weight gain seen in the *Sim1* R171H mice. As seen in the hypothalamic sections, there was no change in the structure of the PVN (Figure 14C) or the number of Sim1-GFP positive cells (Figure 14D), indicating the low levels of SIM1 R171H function does not impact development of the PVN.

These experiments reveal that the obesity driving mechanism(s) of the R171H mutation lie beyond simple deficiencies of PVN development, or connection with known interacting proteins, and will require further research to decipher.

Isolation of SIM1 expressing cells to investigate biochemical mechanisms of SIM1

Satiety signalling from PVN neurons results from their activation (depolarisation) in response to α-melanocortin stimulating hormone (α-MSH) binding to the melanocortin 4 receptor (Mc4r). This process is reported to increase expression of SIM1 [57]. To aid our investigation into SIM1 biochemistry using the Sim1-GFP mouse line, we first tested whether expression of *GFP* and endogenous *Sim1* mRNA would respond to KCl depolarisation. Embryonic brains were harvested from E16.5 Sim1-GFP embryos and GFP expression was used to guide dissection of the hypothalamus (Figure 15A and 15B). Figure 15C shows qPCR analysis of the expression of *Sim1* and *GFP* in the hypothalamic neurons with low to no expression in control cortical neurons. The hypothalamic marker gamma-aminobutyric acid receptor subunit epsilon (*GABRE*) was also tested and found to be expressed in the hypothalamus with virtually no expression in the cortical neurons. This confirms we can use GFP expression to guide dissection of the embryonic brain to obtain a SIM1 expressing population of cells.

We next depolarised the hypothalamic cultures with KCl, harvested RNA and performed a pilot qPCR analysis for *Sim1* and *GFP*. As expected, *Sim1* expression increased upon depolarisation and *GFP* followed suit, confirming that *GFP* expression was regulated in the same manner as *Sim1* due to GFP being driven by the *Sim1* promoter (Figure 15D).

We therefore proceeded to investigate if *Sim1* and *GFP* expression were also induced *in vivo* in the hypothalamus of adult mice. As hypothalamic *Sim1* has previously been shown to increase following administration of two satiety signalling molecules, leptin and MTII (a synthetic MSH mimetic) [57], we decided to perform a fast and refeed study to stimulate hunger and satiety signalling, and assess associated changes in *Sim1* and *GFP* expression.

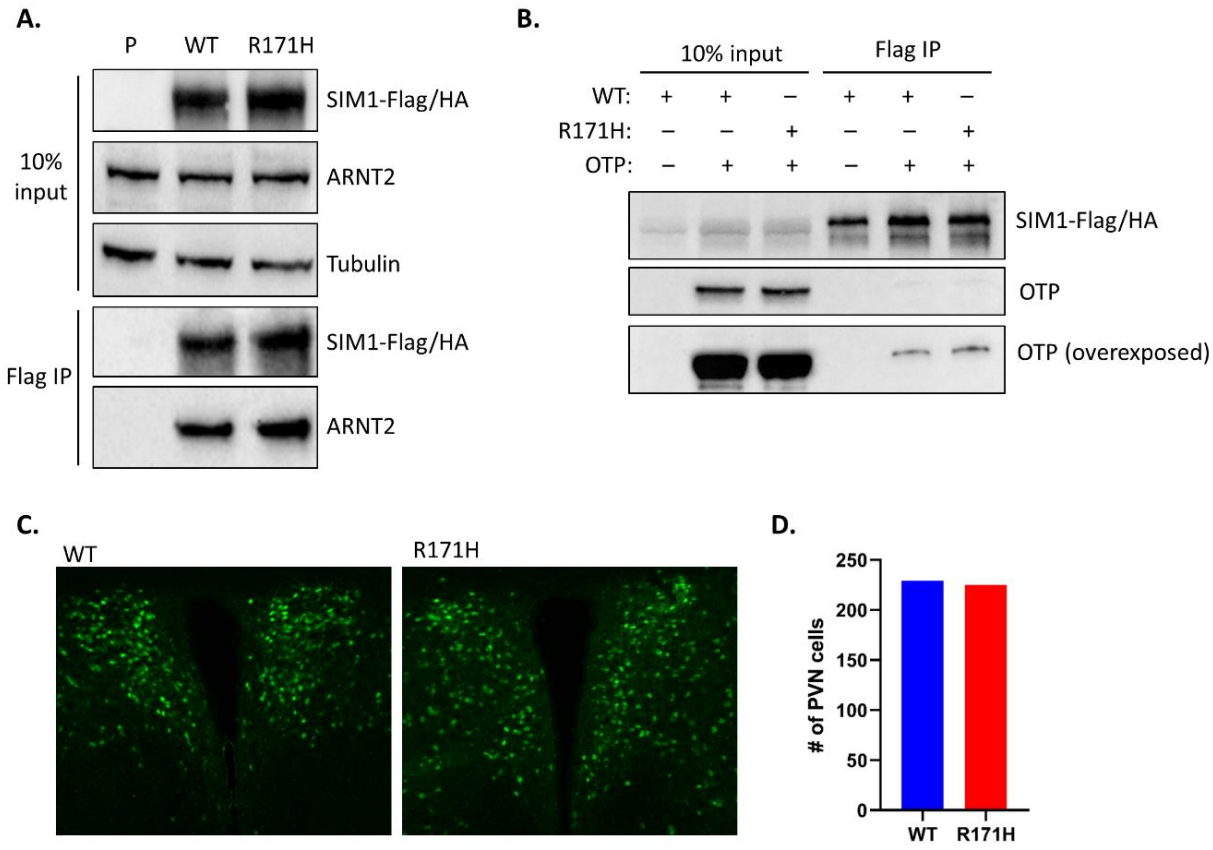

Figure 14. Effect of R171H mutation on protein-protein interactions and structure of PVN. A. HEK293T cells were transiently transfected with either WT or R171H SIM1-Flag and Co-IP performed using anti-Flag resin. Both WT and R171H SIM1 were able to Co-IP with equal amounts of endogenous ARNT2. P = parent cell line. **B.** HEK293T cells were transiently transfected with OTP and either WT or R171H SIM1-Flag. Co-IP was performed using anti-Flag resin with both WT and R171H SIM1 being able to Co-IP with equal amounts of OTP. **C.** PFA fixed brains from Sim1-GFP:WT and Sim1-GFP:R171H mice were sectioned and immunostained for GFP. x10 magnification. Images representative of at least four mice for each genotype. **D.** Number of Sim1-GFP positive cells detected in the PVN from part C.

Mice were fasted for 40hr, then half of the mice were fed for 4hr before harvesting the hypothalamus and extracting RNA for use in qPCR analysis. As seen in Figure 15E, we did not see an increase in either *Sim1* or *GFP* expression upon refeeding in either SIM1 WT or R171H mice. Unfortunately, we were unable to detect GFP expression using fluorescent microscopy to guide our dissection of the adult hypothalamus to obtain the PVN, therefore it was possible we were also harvesting other SIM1 expressing populations (ie. the SON and NLOT). *Sim1* expression within these cells could have skewed the results, making this experimental approach to assess *SIM1* levels to not be feasible due to the technical limitations of obtaining the desired cells types.

As we could not isolate the PVN from adult hypothalamus using fluorescent microscopy directly on tissue from Sim1-GFP mice, we attempted to sort GFP cells using FACS, as it should allow more accurate isolation of GFP positive cells. To initially explore this technique, we decided to use the more accessible kidney cells derived from the Sim1-GFP mice as we had shown that the kidney cells visibly express the Sim1-GFP as well as having the advantage of more cells to work with compared to using dissected hypothalamus. FACS sorting for GFP without the use of GFP antibodies was unsuccessful, with no cells being detected as GFP positive (data not shown), therefore we went on to attempt isolating GFP cells using MARIS (method for analysis RNA following intracellular sorting) [35]. This technique has been reported to isolate high quality RNA from specific cell types that have been fixed, stained with desired antibody and FACS sorted. Using this technique resulted in approximately 1% of cells put through FACS being detected and sorted as GFP positive and was unfortunately not enough to obtain RNA quantities required for downstream applications (data not shown). This approach was therefore abandoned.

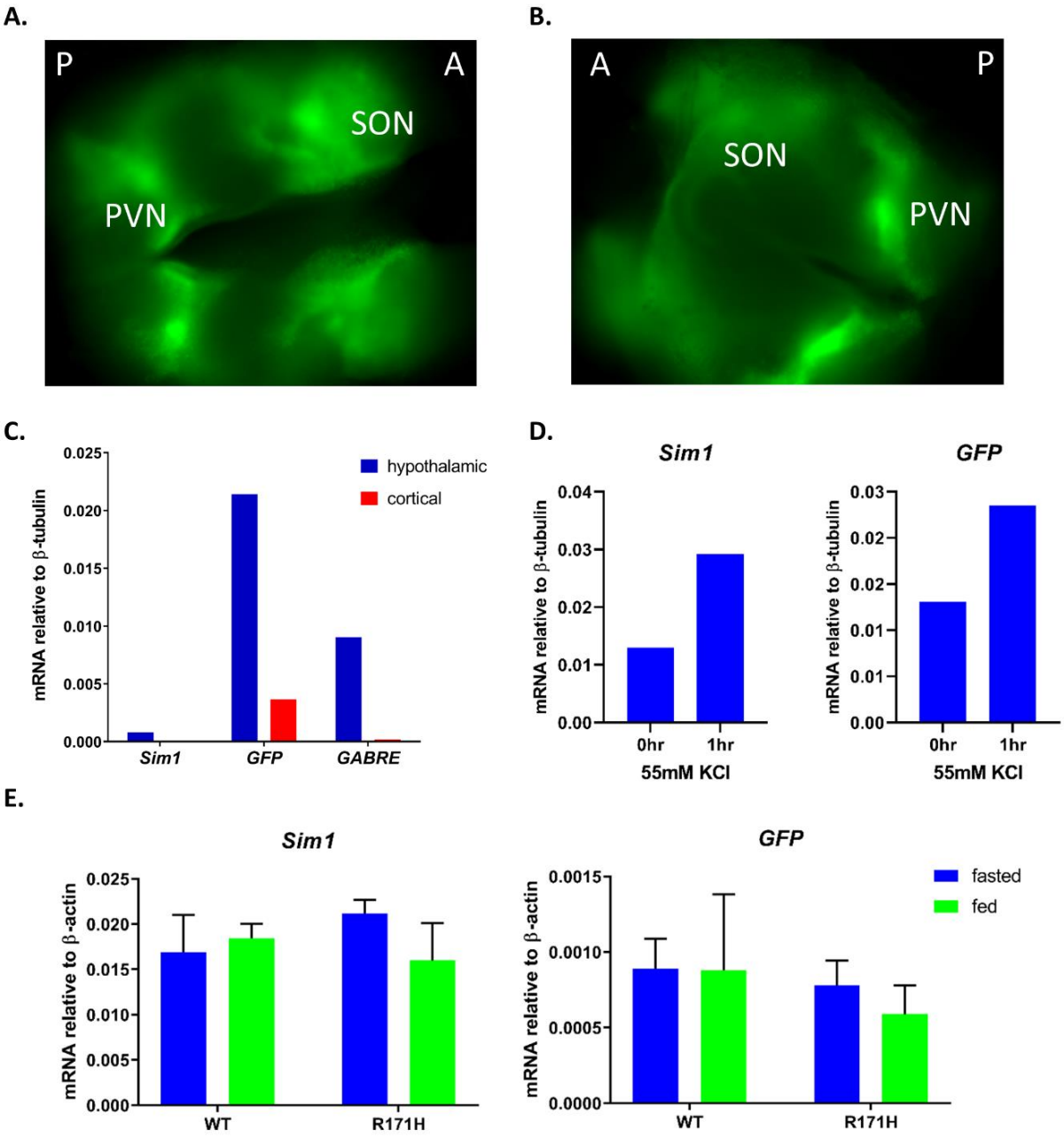

Figure 15. Isolation of Sim1-GFP cells. A and B. E16.5 embryonic hypothalamus dissected based on GFP expression. A = anterior, P = posterior. x4 magnification. **A.** Superior view. **B.** Inferior view. **C.** qPCR of gene markers of hypothalamic and cortical cultures. **D.** KCl depolarisation of cultured hypothalamic neurons results in a change in *GFP* expression that reflects changes in *Sim1* expression. **E.** RNA from the hypothalamus of adult mice that were either fasted or re-fed was harvested and qPCR of *Sim1* and *GFP* performed. WT fed n = 3, WT fasted n =2, R171H fed n = 3, R171H fasted n = 3.

Discussion

SIM1 has been established as being essential for embryonic development with a continued role in energy homeostasis during adulthood. Despite point mutations such as R171H being monogenetic drivers of obesity, there is currently still an insufficient understanding of the mechanisms of SIM1 function, including but not limited to how SIM1 itself is regulated and the target genes regulated by SIM1. With the aid of a mouse model that has a weakly functioning SIM1 mutation, R171H, we sought to develop methods towards investigating the mechanism(s) behind the functional deficiency of the mutation as well as identification of highly desired target genes of SIM1.

Mechanism of R171H dysfunction in obesity

The R171H mutation, identified in a screen of obese patients, had previously been shown to have significantly decreased activity on a reporter assay [30, 63]. Introduction of this mutation into a mouse model resulted in significantly increased weight gain compared to control mice, reflecting the human phenotype and confirming that the R171H is a SIM1 mutation able to drive obesity.

We therefore wanted to understand the mechanism behind the dysfunction of the R171H mutation. As the R171H mutation sits within the PAS A domain, which is involved in dimerisation with partner proteins as well as influencing the strength of partner interactions, we tested the ability of the mutated SIM1 to bind to its dimerisation partner ARNT2 by co-expressing the proteins and performing Co-IP. The mutated SIM1 was equally capable of dimerising with ARNT2 as the WT SIM1, indicating that the mutation is not affecting the structure in a way that inhibits dimerisation with its essential partner proteins.

We next looked at its interaction with a protein outside of the bHLH/PAS family of transcription factors. OTP mutations can drive obesity and OTP has previously been identified by the Whitelaw laboratory to interact with SIM1 (unpublished data). We once again saw the same result of the R171H mutation not having an effect on the interaction of SIM1 and OTP, with both WT and R171H SIM1 binding equal amounts of OTP. In addition to these experiments showing that the R171H mutation does not weaken dimerisation, the western blots show equal expression of WT and R171H SIM1 in the input sample indicating that the R171H mutation is not decreasing the overall stability of the protein.

SIM1 plays an essential role in the development of the PVN, being responsible for the differentiation of specific neuronal cell lineages. With the significantly reduced activity of SIM1 resulting from the R171H mutation, we looked at the effect of the mutation on the development of the PVN in mice. Using the compound Sim1-GFP:Sim1^{R171H} mice, we were able to show that there is no difference in the number of SIM1 positive cells nor is there any apparent differences in the general structure of the PVN compared to Sim1-GFP:Sim1^{WT} mice. This indicates that there is sufficient SIM1 function during embryonic development to ensure correct differentiation of the

required cell types and that a lack of PVN neurons is not responsible for hyperphagic obesity in R171H mice.

The results presented in this study indicates that the R171H mutation causes SIM1 dysfunction through a currently unknown mechanism. This gives rise to the question whether the R171H mutation causes a deficiency in DNA sequence recognition or binding yet to be discovered co-factors, and will required further investigation to elucidate. Previous work in the Whitelaw laboratory has developed an EMSA method using the central midline element (CME) which SIM1 is known to bind, however preliminary data using baculoviral expressed bHLH/PAS termini of the SIM1/ARNT2 heterodimer revealed no difference in EMSA between WT SIM1 and SIM1 R171H (D. Bersten, unpublished data). A more robust method to assess DNA binding, which may eventuate once genomic sites recognised by SIM1 in neurons have been defined, may be needed to determine differences between WT SIM1 and SIM1 R171H.

Isolation of SIM1 positive cells

With an ultimate aim of identifying target genes of SIM1, we initially attempted isolation of SIM1 positive cells that had either WT or the functional weak R171H mutant of SIM1. From the results in this study, we were able to show that with the Sim1-GFP mouse line we could use GFP expression to guide dissection of the embryonic brain to isolate embryonic SIM1 positive cells. In addition, we obtained evidence that the *GFP* expression could be induced using depolarisation in the same manner as *Sim1.* However, GFP expression could not be detected in the adult brain using microscopy. Consequently, we dissected the whole hypothalamus in order to test if *Sim1* expression in response to feeding might still be analysed despite being unable to use the GFP as a guide to dissect the PVN. Unfortunately we were unable to detect alterations in *Sim1* levels upon feeding, perhaps due to also collecting other nearby SIM1 positive cell populations (ie. SON and NLOT) that could have been avoided if we were able to carry out the dissection with Sim1-GFP guidance. As we could not detect GFP in adult mouse brain by microscopy, we explored whether FACS and MARIS might be used to isolate SIM1 positive cells, using a simpler kidney dissection and cell disruption protocols. Unfortunately both methods were unable to isolate the Sim1-GFP cells.

Ideally we would like to retrieve populations of Sim1-GFP cells with either endogenous WT or R171H SIM1 isolated from the PVN for use in RNA sequencing (RNAseq). Comparison of the gene expression profiles from RNAseq between the WT and R171H populations would give insight into the target genes of SIM1, opening the door for further investigation into SIM1 function in appetite control and how functional mutants of SIM1 result in obesity. Future work will involve developing an effective method to isolate SIM1 positive neurons from the PVN. This could include using laser capture microdissection to dissect the PVN from the hypothalamus using brain landmarks for guidance, followed by single cell RNAseq. Sequencing of single cells rather than a heterogeneous population of cell types will allow exclusion of *Sim1* negative cells by identifying those that express *Sim1.* Additionally, single cell analysis will allow expression profiles of SIM1 neuron subpopulations to be examined.

Chapter 5: Expression and purification of SIM1/ARNT2 heterodimer

Introduction

The N-terminal architecture of bHLH/PAS factors is highly conserved across the family, with the bHLH and PAS domains being involved in DNA binding and partner specificity and dimerisation. The PAS domains are also involved in signal regulation and interactions with co-factors. The C-termini vary in sequence and length between family members and contain the transcriptional activation and repression domains that bind to co-factors to regulate gene expression.

There is currently incomplete, partial structural data for the bHLH/PAS family, consisting of either isolated PAS A or B domains, or the N-terminal region containing the bHLH motif and both PAS domains, being solved using nuclear magnetic resonance (NMR) or X-ray crystallography (Table 2). A limited number of these structures have been reported bound to DNA, small molecules or co-factors. The reported structures have indicated that the 3D fold of the PAS repeats are highly conserved across the family despite low sequence homology at the amino acid level, while the overall quaternary arrangements can differ across heterodimers, particularly in the interacting surfaces between different family members [65]. The C-termini on the other hand lack any conserved domain homology with previously described structural motifs and are believed to be structurally flexible. Consequently, the C-terminal halves of bHLH/PAS proteins have yet to be structurally characterised.

While the structural data available for bHLH/PAS proteins indicates that the general propensity for heterodimer formation is conserved, it does not necessarily translate to the same surfaces being involved in heterodimerisation, DNA binding and interactions with co-factors and ligands. For example, the HIF2α/ARNT structure has six binding interactions within the heterodimer compared to the CLOCK/BMAL structure which has only four interfaces [66, 67]. While the same domains appear to be responsible for dimerisation in both cases, with some conserved residues being involved, it has been noted that there are some slight differences in the orientation of the folds at other contact points. These observed differences may contribute to the variation in the overall protein structure between heterodimers, which could consequently be responsible for the specificity and strength of partner dimerisation and target gene selection.

Due to the limited structural data available for bHLH/PAS proteins, there is still much to learn about their protein-protein interactions and how these may dictate specificity in target gene selection. Elucidating the full complement of protein structures of this family would be particularly important in deciphering how subtle differences in heterodimer structure could lead to DNA binding sequence selectively, especially given the similar recognition sequences between different heterodimers.

There is currently no structural data for SIM1, with only partial computer generated models based on the HIF2α/ARNT and CLOCK/BMAL structures available [30, 63]. While these computational models may give some insight into the function of SIM1 and help explain the reason for dysfunction of SIM1 point mutants, these SIM1 models remain too rudimentary for use in further structural based studies such as in silico drug design. Having an empirical SIM1 structure will allow further understanding of how SIM1 functions, including DNA binding, Class II partner selection and dimerisation, co-factor interactions and small molecule binding.

Importantly, mapping SIM1 mutations known to cause obesity would help explain the impaired function of the protein. While mutation hotspot modelling based on HIF2α/ARNT and CLOCK/BMAL structures has been performed in the past, the studies were insufficient to properly characterise SIM1/ARNT2 interaction as the modelling resulted in the mutations mapping to different PAS domain surfaces [30, 63, 68]. For example, modelling based on the earlier CLOCK/BMAL PAS B structure has the R171H mutation sitting on a dimerisation interface, while the model based on HIF2α/ARNT suggests that R171H sits on a loop that makes contacts with DNA. These two models give rise to different explanations for the mode of R171H dysfunction: 1) If R171H lays within the dimerisation interface, it could be weakening the dimerisation with its partner proteins and decreasing the overall stability of the SIM1 protein and 2) the variant may cause a deficiency in either DNA sequence recognition and/or binding, as well as negatively affect the stability of the transcription factor when bound to DNA. While previous work done in the Whitelaw laboratory does not favour the first scenario, the latter scenario of the R171H mutations affecting DNA binding remains a possibility as we have yet to define a native SIM1 bound cis element on a SIM1 target gene. Obtaining a SIM1 structure with this mutation would be of major benefit to further understand SIM1 function.

Previous expression systems

As there is a pressing need for a SIM1 structure, there have been multiple attempts to express and purify enough SIM1 protein in order to perform structural analysis. Previously in the Whitelaw laboratory, there have been attempts to express mammalian SIM1 protein in both bacteria and yeast. Both attempts failed, presumably due to the lack of mammalian chaperones and systems required for the correct folding of SIM1. Expression in insect cells using baculovirus has also been attempted with limited success. Electrophoretic mobility shift assays indicated that functional SIM1 protein capable of binding DNA was expressed (Whitelaw laboratory, unpublished data). However, low protein expression yields and the potential lack of cell machinery required for complete proper mammalian protein folding did not make insect cells a viable choice for large scale protein expression and purification.

Table 2. Current bHLH/PAS protein structures determined by NMR or X-ray crystallography.

Protein	Domain	Method	DNA/Ligand/Co-activator	Human/mouse	Reference
HIF1α	PAS B	X-ray	-	Human	[69]
HIF2α	PAS B	NMR	-	Human	[70]
ARNT	PAS B	NMR/X-ray	-	Human	[71], PDB: 2B02
AhR	PAS A	X-ray	-	Mouse	[72]
HIF1α/ARNT	PAS B	X-ray	Co-activator	Human/mouse	[73]
HIF1α/ARNT	bHLH, PAS A, PAS B	X-ray	DNA (HRE)	Mouse	[67]
HIF2α/ARNT	PAS B	NMR/X-ray	-	Human	[67, 71, 74]
HIF2α/ARNT	PAS B	X-ray	ligand	Human	[67, 74]
HIF2α/ARNT	bHLH, PAS A, PAS B	X-ray	-/+ DNA (HRE)	Mouse	[67]
AhR/ARNT	bHLH, PAS A	X-ray	-/+ DNA (XRE)	Human/mouse	[75, 76]
NPAS1/ARNT	bHLH, PAS A, PAS B	X-ray	-	Mouse	[77]
NPAS3/ARNT	bHLH, PAS A, PAS B	X-ray	DNA (HRE)	Mouse	[77]
CLOCK/BMAL	bHLH	X-ray	-/+ DNA (E box)	Human	[78]
CLOCK/BMAL	bHLH, PAS A, PAS B	X-ray	-	Mouse	[66]

Expi293F expression system

Expi293F cells are a modified HEK293 cell line that has been adapted to be grown as a high density suspension culture for easy scalability and high transfection efficiency, allowing increased yields of protein to be expressed when compared to adherent cell lines.

As Expi293F cells are mammalian, they have an advantage over bacterial, yeast and insect cells in that they have the necessary mechanisms to ensure the correct folding of mammalian proteins. Potential PTMs made to the protein of interest are more likely to be maintained in a mammalian system, as well as the presence of chaperones to ensure proteins are kept in their native form. In this study we use the suspension Expi293F cell protein expression system to express and purify a truncated human SIM1/ARNT2 heterodimer. This cell line should allow increased yields of correctly folded SIM1/ARNT2 protein that can be used for either cryogenic electron microscopy (cryo-EM) or crystallisation trials for structural analysis.

Results

Designing tagged SIM1 and ARNT2 expression plasmids for production and purification of SIM1/ARNT2 bHLH/PAS heterodimers from Expi293F cells

Due to difficulties in past attempts to express and purify full length bHLH/PAS transcription factors for structural analysis, we decided to attempt expressing and purifying human SIM1 truncated to the end of the PAS B domain (hSim1.1-348). SIM1 was co-expressed with human ARNT2, also truncated to the end of the PAS B domain (hArnt2.1-455), as mouse genetic studies indicate it to be the preferred dimerisation partner of SIM1 [15, 16]. Co-expression of SIM1 with ARNT2 has also been shown to result in more stable SIM1 expression (Whitelaw laboratory, unpublished data). To aid purification and detection of the expressed protein heterodimer, SIM1 was tagged with both 2xFlag and 3xTwin-Strep tags, while ARNT2 had a 2xMyc tag. Resistance for puromycin was included on the SIM1 plasmid and neomycin for ARNT2 to use for selection in the generation of stable cell lines. Both proteins were expressed from the pEF promoter. For simplicity, these truncated proteins will be referred to as SIM1.1-348.Flag/Strep and ARNT2.1-455.Myc (Figure 16A).

Strep-Tactin XT Superflow and anti-Flag M2 affinity gel resins

In this study, two different resins were used for the purification of the SIM1.1-348.Flag/Strep and ARNT2.1-455.Myc heterodimer. The first resin is the Strep-Tactin XT Superflow resin (IBA), which consists of a modified streptavidin bound to resin that can bind to twin-Strep tag with high affinity, even under physiological buffer conditions. Due to biotin's high affinity for Strep-Tactin XT, it is used to competitively elute the twin-Strep tagged protein of interest from the resin. The second resin used in these experiments is the anti-Flag M2 affinity gel resin (Sigma Aldrich) which has a Flag antibody bound to agarose beads. The Flag antibody recognises and binds to the Flag tag epitope that is conjugated to the protein of interest. The 3xFlag peptide which competitively binds to the Flag antibody on the resin is used to elute the Flag tagged protein of interest. An advantage of both these resins is that they can be regenerated by removing the elution reagents, allowing the resins to be reused for additional Co-IPs. For the experiments presented here, both the twin-Strep and Flag tags were conjugated to the truncated SIM1 protein to enable easy single or two-step purification as well as detection using Strep or Flag tag specific antibodies.

Transient vs stable expression of SIM1.1-348.Flag/Strep and ARNT2.1-455.Myc in Expi293F cells

Confirmation of protein co-expression from our truncated SIM1.1-348.Flag/Strep and ARNT2.1-455.Myc plasmids in Expi293F cells was performed using both transient and stable small scale transfections and purified using Co-IP with either Flag or Strep resin. This was to confirm that the correct size protein was being expressed from each of the plasmids and to compare which transfection method gave the greater yield of protein. As seen in Figure 16B,

SIM1.1-348.Flag/Strep and ARNT2.1-455.Myc were expressed at the correct sizes of 45kDa and 55kDa, respectively. While transient co-transfection of both plasmids resulted in protein that could be detected by Coomassie stain after either Flag or Strep Co-IP, Expi293F cells selected to stably express ARNT2.1-455.Myc with transient SIM1.1-348.Flag/Strep expression did not produce enough protein to be detected on a Coomassie gel (Figure 16B). Western blotting for SIM1.1-348.Flag/Strep and ARNT2.1-455.Myc confirmed the stable cell line produced the correct proteins, while also confirming that transient transfections resulted in a greater yield of SIM1.1-348.Flag/Strep and ARNT2.1-455.Myc (Figure 16C). A significant amount of SIM1.1-348.Flag/Strep and ARNT2.1-455.Myc was observed in the flow through for both the Flag and Strep Co-IPs from the transiently transfected cells (Figure 16C). This was most likely due to the large amounts of tagged SIM1 protein being transiently expressed saturating the capacity of the small volumes of resin used in these Co-IPs.

Based on these results, we decided to upscale protein expression using transient transfection to ensure we obtained the greatest yield of SIM1.1-348.Flag/Strep and ARNT2.1-455.Myc possible. This small scale purification experiment was performed using batch Co-IP due the limitation of small volumes not flowing through resin using gravity alone. It was decided to perform subsequent Co-IP purifications on up-scaled volumes using the Strep resins in a column format to eliminate resin loss, which results in decreased yields, that occurs when using batch format. In addition, the column format allows easier handling of larger volumes compared to batch purification as there is no need for multiple centrifugation steps.

Mid scale culture and expression of SIM1.1-348.Flag/Strep and ARNT2.1-455.Myc

We next up-scaled our culture volume from 20ml to 200ml of Expi293F cells and performed affinity column chromatography using the Strep-Tactin XT resin, using biotin to elute the proteins. Initially, we started with a cell concentration of $1x10^6$ cells/ml for transfection, however at the time of harvest the cell viability was 73%, well below the recommended viability of 90% to ensure greater protein yields. While purification of SIM1.1-348.Flag/Strep and ARNT2.1-455.Myc was still successful from this culture (Table 3), this result indicated that a low starting cell concentration would result in lower cell numbers and consequently lower protein yield. Therefore, for subsequent cultures the starting cell concentration was increased to $2x10^6$ cells/ml to ensure greater cell numbers and viability at the time of harvesting to improve protein yields.

As seen in the representative coomassie in Figure 17A, approximately equal amounts of both SIM1.1-348.Flag/Strep and ARNT2.1-455.Myc protein were eluted from the Strep-Tactin XT resin. Estimates based on a BSA standard included in the coomassie gels indicated approximately 170µg of SIM1.1-348.Flag/Strep and ARNT2.1-455.Myc was obtained in the fractions eluted from the Strep column purification. However, there were also contaminants eluted from the column, with the two most predominate background bands at

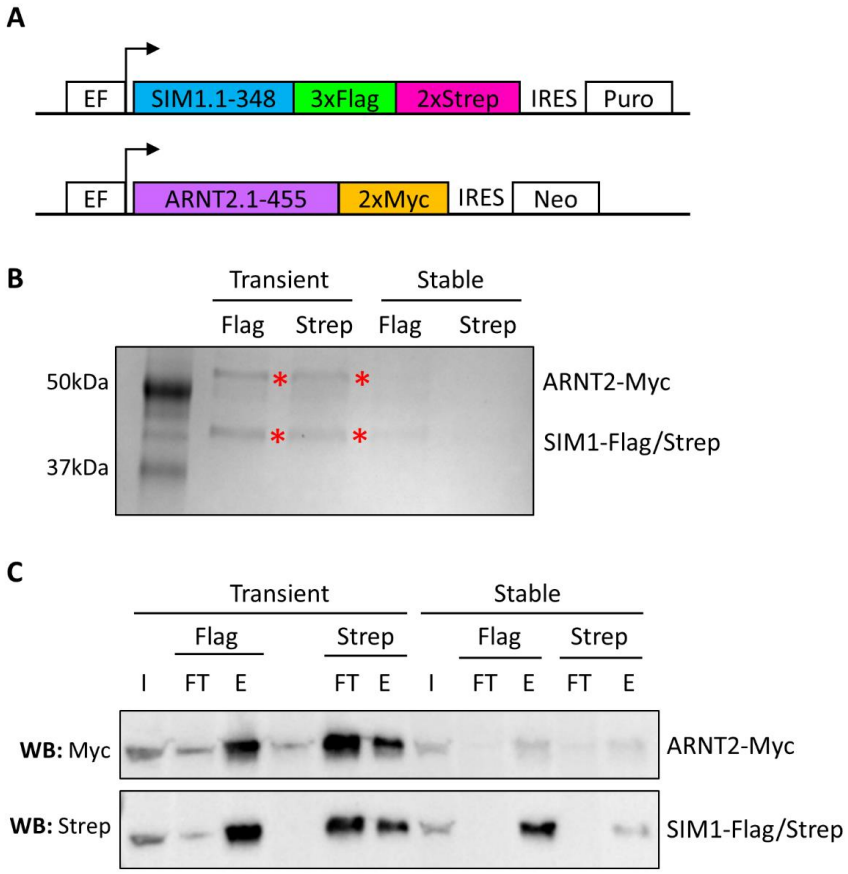

Figure 16. Small scale expression of transient and stably transfected truncated SIM1.1-348.Flag/Strep and ARNT2.1-455.Myc in Expi293F cells. A. Schematic of expression plasmids used to express truncated SIM1.1-348.Flag/Strep and ARNT2.1-455.Myc. Tagged proteins and either puromycin (Puro) or neomycin (Neo) resistance cassettes are expressed from the pEF promoter. IRES = internal ribosome entry site. **B.** Coomassie gel of transiently transfected SIM1.1-348.Flag/Strep with either transiently or stably expressed ARNT2.1-488.Myc. Co-IP was performed using Flag or Strep-Tactin XT resin purification methods. 3mg WCE was used for Co-IP and protein eluted in 150µl of appropriate elution buffer. 10% of elution loaded onto coomassie gel. **C.** Western blot of purified SIM1.1-348.Flag/Strep and ARNT2.1-455.Myc from Flag and Strep-Tactin XT resin purifications. I = 1% Co-IP input, FT = 1% flow through, E = 10% elution.

approximately 70kDa and 250kDa. As the 70kDa band has previously been identified in other attempted bHLH/PAS purifications performed in the Whitelaw laboratory, it was concluded this band represented heat shock protein 70 (Hsp70). Mass spectrometry performed on the gel purified protein from the unknown 250kDa band revealed it to be acetyl CoA Carboxylase 1 (ACC1) (appendix). ACC1 is a biotinylated protein, therefore is able to bind to the Strep-Tactin XT resin.

We next concentrated the elution fractions using a concentrating spin column. As seen in Figure 17B, the SIM1.1-348.Flag/Strep and ARNT2.1-455.Myc heterodimer was successfully concentrated. As the truncated SIM1.1-348.Flag/Strep and ARNT2.1-455.Myc heterodimer

Table 3. Summary of attempts to express and purify SIM1.1-348.Flag/Strep and ARNT2.1-455.Myc heterodimer from Expi293F cells. All protein concentrations estimates were based on analysis of coomassie gels using the included BSA standards. All concentration and volume values are approximate. Conc^{n} = concentration.

Batch	Starting		Harvest			Pellet volume (ml)	Strep or Flag Co-IP	Column or batch	Buffer salt and detergent conc^{n}	Concentrated		Total volume after Flag Co-IP (µl)	Total purified SIM1/ARNT2 (µg)	Estimate after conc^{n}
	Cell conc^{n} (cells/ml)	Volume (ml)	Cell conc^{n} (cells/ml)	Viability (%)	Volume (ml)					Protein (µg/µl)	Volume (µl)			
#1	$1x10^{6}$	200	$1.1x10^{6}$	73	220	1	Strep	Column	150mM NaCl 0.5% NP-40	0.8	50	-	40	Yes
#2	$2x10^{6}$	200	$2.95x10^{6}$	92	220	5	Strep	Column	150mM NaCl 0.5% NP-40	2	50	-	100	Yes
#3	$2x10^{6}$	200	$2.5x10^{6}$	86	220	5	Strep	Column	150mM NaCl 0.5% NP-40	1.5	100	-	150	Yes
#4-6	$2x10^{6}$	1400	$5.42x10^{6}$	96	1100	8 x 5ml (purified in batches #4-6)								
			$2.47x10^{6}$	93	220									
			$2.7x10^{6}$	93	220									
#4						20	Strep	Column	150mM NaCl 0.5% NP-40	1	200	-	200	Yes
#5						5	Strep + Flag	Column	500mM NaCl 1% Triton X-100	-	-	500	Strep = 150 Flag = 63	No
#6						5	Flag	Batch	500mM NaCl 1% Triton X-100	-	-	900	119	No

is approximately 100kDa, a concentration column with a size cut off of 30kDa was used. This meant that both Hsp70 and ACC1 were also concentrated during the concentration step. Analysis of the coomassie gels using the BSA standards indicated approximately 100-150µg of SIM1.1-348.Flag/Strep and ARNT2.1-455.Myc were purified and concentrated. This indicates an approximate 10-40% loss in protein during the concentration step. While there was significant protein loss, SIM1.1-348.Flag/Strep and ARNT2.1-455.Myc levels remained stoichiometric, indicating that the two proteins form a strong heterodimer.

The western blot in Figure 17C shows good expression of both SIM1.1-348.Flag/Strep and ARNT2.1-455.Myc in the input used for the Co-IP, with only low amounts of the proteins being detected in the initial flow through over the column and the final regeneration of the resin. This indicates that almost all of the SIM1.1-348.Flag/Strep is binding to the column and being eluted, validating the use of the Strep-Tactin XT resin for up-scaled protein purification.

Large scale culture and expression of SIM1.1-348.Flag/Strep and ARNT2.1-455.Myc

We next up-scaled the culture from 200ml to 1400ml, using the same starting cell concentration of $2x10^6$ cells/ml and kept the same ratio of plasmid DNA. Due to the limitations of the size of the flasks, the culture was split over one 2L and two 500ml flasks. Cell viability was maintained in the up-scaled culture, with the 2L flask having a greater increase in cell concentration at the time of harvest compared to the smaller flasks (Table 3). Instead of using all of the culture for a single purification attempt, we decided to start with only half of the culture, which would allow us to use the one culture for altering conditions to obtain the best purification results.

As seen in Figure 18A, there was a visible increase in SIM1.1-348.Flag/Strep and ARNT2.1-455.Myc yield compared to the 200ml culture purifications (Figure 17A) as expected. However, there was also an increase in the amount of contaminants remaining bound to the resin during the wash steps and eluting with the desired proteins. We next concentrated the elutions (Figure 18B) and performed western blot using the concentrated protein (Figure 18C). As can been seen in Figure 18C, SIM1.1-348.Flag/Strep and ARNT2.1-455.Myc were also detected in both the flow through and NaOH resin regeneration wash. As there was protein detected in the flow through, there could have been some saturation of the resin during the procedure with the excess protein flowing through. The presence of the SIM1.1-348.Flag/Strep and ARNT2.1-455.Myc in the NaOH wash flow through indicates that the heterodimer was once again not completely eluted off the resin and would require more elution fractions to completely elute the proteins to obtain the greatest yield possible.

The concentrated protein was then put through a size exclusion column to try to remove contaminants and obtain a more pure SIM1.1-348.Flag/Strep:ARNT2.1-455.Myc heterodimer. As seen in Figure 18D, there was a peak at roughly 100kDa, the expected size of our truncated SIM1.1-348.Flag/Strep and ARNT2.1-455.Myc heterodimer. There was also

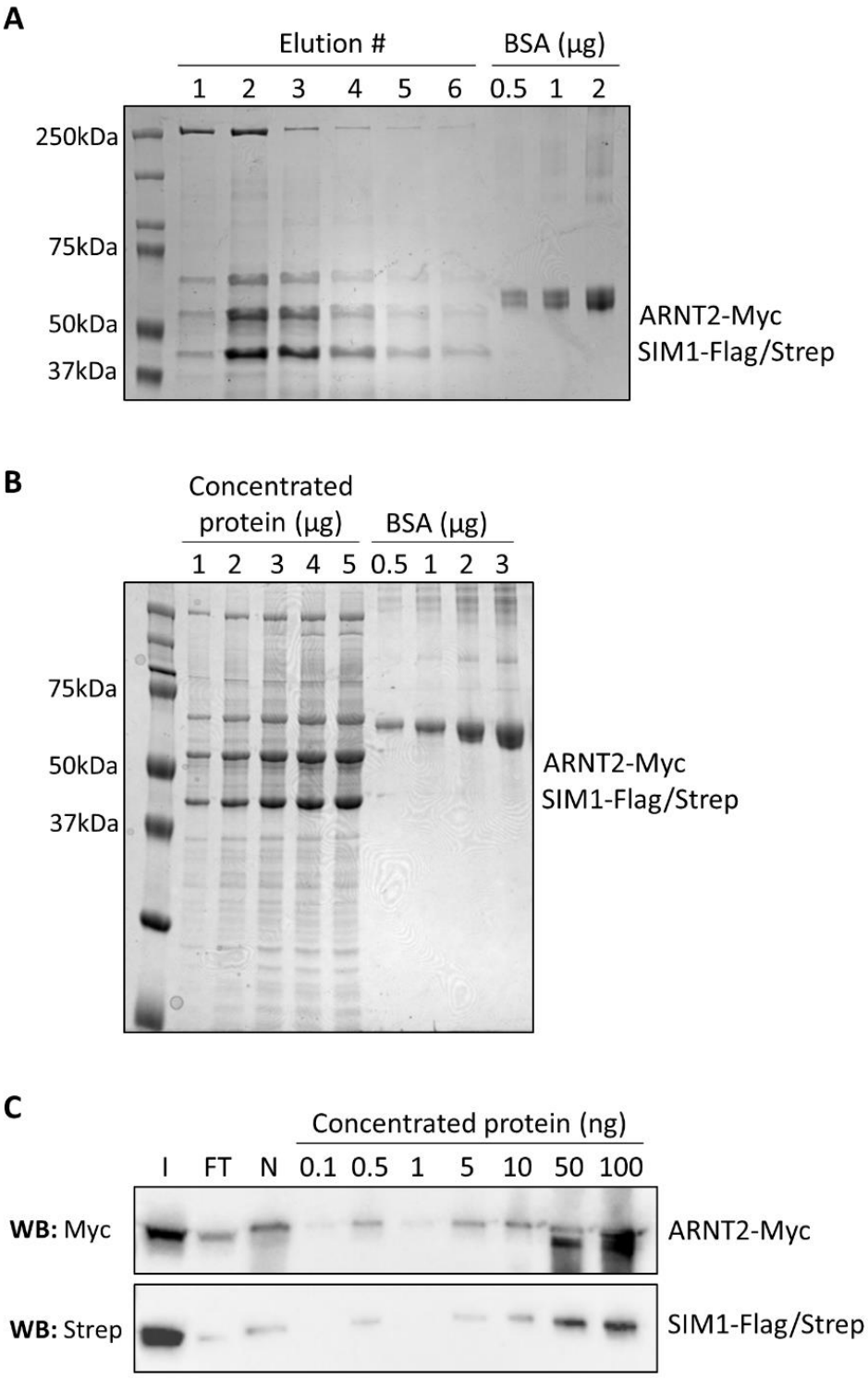

Figure 17. Strep resin purification of transiently transfected truncated SIM1.1-348.Flag/Strep and ARNT2.1-455.Myc in Expi293F cells. BSA standards are included in coomassie gels for estimating purified protein concentrations. **A.** Coomassie of SIM1.1-348.Flag/Strep and ARNT2.1-455.Myc eluted from a Strep resin column using Biotin. Elutions #1-6 = 35ul of 1ml elution. **B.** Coomassie of the concentrated protein from elutions #2-4 from the Strep resin column. Amount of protein loaded onto coomassie gel calculated from estimates based on part A. **C.** Western blot of proteins after concentration of elutions #2-4. Amount of protein loaded onto western blot calculated from estimates based on part B. I = 25µg Co-IP input, FT = 25µg flow through, N = 10µl of 2ml NaOH wash.

another peak at the size that corresponds to the ARNT2.1-455.Myc monomer. Western blotting of the elution volumes corresponding to these peaks (Figure 18E) confirmed the presence of both the heterodimer and monomer. We next concentrated volumes 9-12 ml from the SEC as these gave the most intense SIM1.1-348.Flag/Strep and ARNT2.1-455.Myc bands on the western blot. Coomassie gel of the concentrated protein from the SEC (Figure 18F) showed that it was not able to remove all of the contaminants seen after the initial Strep purification, with a visible loss of SIM1.1-348.Flag/Strep and ARNT2.1-455.Myc protein yield also seen.

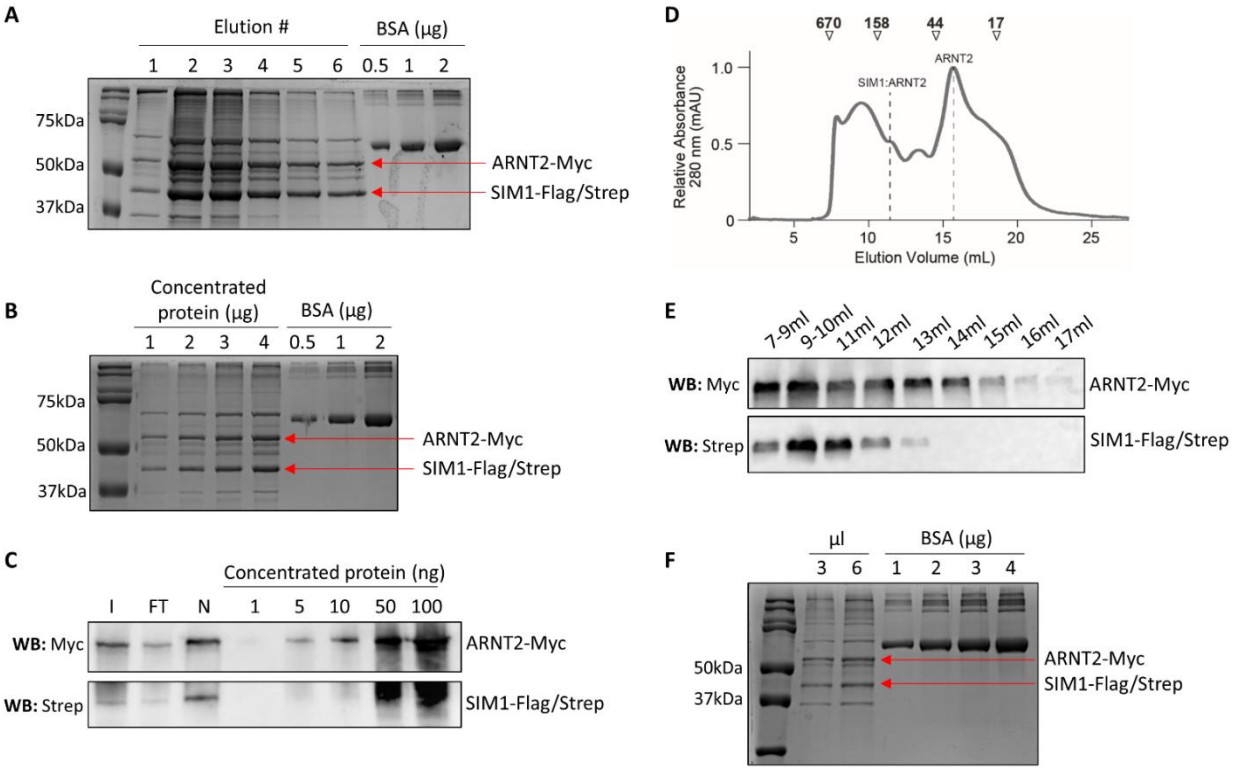

Figure 18. Large scale expression of transiently transfected truncated SIM1.1-348.Flag/Strep and ARNT2.1-455.Myc in Expi293F cells. 1400ml Expi293F cell culture was transiently transfected with SIM1.1-348.Flag/Strep and ARNT2.1-455.Myc. Half of the whole cell extract was used for Strep resin column purification. BSA standards are included in coomassie gels for estimating purified protein concentrations. **A.** Coomassie of SIM1.1-348.Flag/Strep and ARNT2.1-455.Myc eluted from a Strep resin column using Biotin. Elutions #1-6 = 35ul of 1ml elution. **B.** Coomassie of the concentrated protein from elutions #2-6. Amount of protein loaded onto coomassie gel calculated from estimates based on part A. **C.** Western blot of the concentrated protein from elutions #2-6. Amount of protein loaded onto western blot calculated from estimates based on part B. I = 25µg Co-IP input, FT = 25µg flow through, N = 10µl of 4ml NaOH wash. **D.** Concentrated protein was put through a size exclusion column, resulting in peaks at volumes in which SIM1.1-348.Flag/Strep:ARNT2.1-455.Myc heterodimer and ARNT2.1-455.Myc monomer were eluted. mAU = milli arbitury units. **E.** Western blot of SIM1.1-348.Flag/Strep and ARNT2.1-455.Myc from the elution volumes obtained from the SEC. Loaded 10µl from each 500µl elution volume. **F.** Coomassie of concentrated SIM1.1-348.Flag/Strep and ARNT2.1-455.Myc from the 9-12ml fractions from the SEC.

Optimising conditions for Strep and Flag purifications of SIM1.1-348.Flag/Strep and ARNT2.1-455.Myc

As nonspecific proteins were still being pulled down along with SIM1.1-348.Flag/Strep and ARNT2.1-455.Myc on the Strep resin and were not completely removed using size exclusion, we decided to try changing the buffer conditions (Table 3) as well as use Flag Co-IP as a second purification step to help reduced contaminants in the elutions. Using part of the large scale culture, we performed affinity chromatography using the Strep column using more stringent wash buffers that had increased salt and detergent concentrations to help prevent non-specific binding of proteins to the resin. As seen in Figure 19A, the use of the more stringent buffers resulted in generally less non-specific protein being eluted from the column compared to the initial buffer conditions (Figure 18A). To further purify the sample, the Strep-Tactin elutions were then passed through a Flag resin column and eluted using 3xFlag peptide. Comparison of Figure 19A and 19B shows that the Flag Co-IP was able to further reduce the levels of background bands in the eluted fractions, however this clean-up was not perfect and there was still some non-specific binding to the Flag resin.

As the ACC1 is consistently pulled out when using the Strep resin, we decided to perform purification using only the Flag resin and the more stringent buffers (Table 3). As seen in Figure 19C, strong SIM1.1-348.Flag/Strep and ARNT2.1-455.Myc bands are detected in the coomassie gel. No ACC1 was pulled out using the Flag resin. However there was still a strong Hsp70 band and another contaminant just below 50kDa still being pulled out in large amounts, with faint bands for multiple other proteins. In addition, there was both SIM1.1-348.Flag/Strep and ARNT2.1-455.Myc detected in the glycine wash of the resin which is used to regenerate the Flag resin for reuse. As with the NaOH wash of the Strep resin, this shows that not all of the desired proteins are being eluted off with the resin with the 3xFlag peptide. This indicates that extra elution fractions or a longer elution incubation time is required for maximal elution of SIM1.1-348.Flag/Strep and ARNT2.1-455.Myc from the Flag resin.

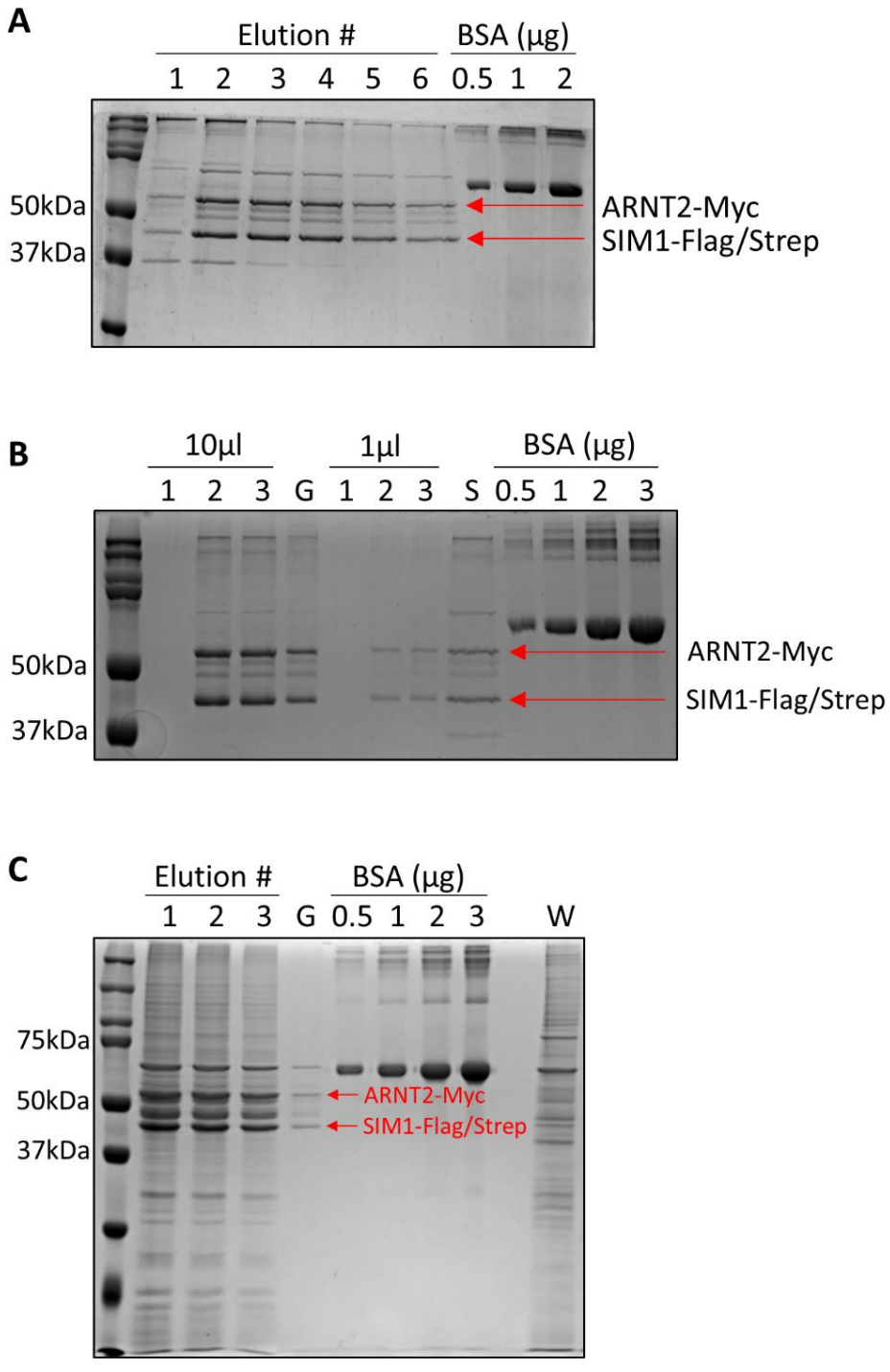

Figure 19. Strep and Flag purification of transiently expressed truncated SIM1.1-348.Flag/Strep ARNT2.1-455.Myc in Expi293F cells. 1400ml Expi293F cell culture was transiently transfected with SIM1.1-348.Flag/Strep and ARNT2.1-455.Myc. Parts of the culture were used to optimise conditions for purification. BSA standards are included in coomassie gels for estimating purified protein concentrations. **A and B.** Coomassie of SIM1.1-348.Flag/Strep and ARNT2.1-455.Myc that was first purified using Strep resin column Co-IP **(A)** followed by column Flag Co-IP **(B)**. **A.** Elution #1-6 = 35µl of 1ml fraction. **B.** 1 or 10ul of elutions #1-3, G = 10µl of 750µl glycine wash, S = 20µl of combined elutions from Strep IP used as input for Flag IP. **C.** Coomassie of SIM1.1-348.Flag/Strep and ARNT2.1-455.Myc purified using batch Flag Co-IP only. G = 10µl of 750µl glycine wash, W = 10µl of 1ml wash flow through.

Discussion

Currently available structural data for the bHLH/PAS family indicates a conserved general overall structure, with a number of key structural features within the heterodimerisation interface being common across family members. While there are similarities between family members, there have also been a number of differences observed, with these differences most likely being responsible for partner specificity and target gene selection. Given the ability of different family members to bind to the same partner proteins, as well as recognise very similar DNA binding sequences, it is important to decipher what subtle structural differences are responsible for the displayed level of specificity.

There is currently no structural data available for SIM1. Using previously determined HIF2α/ARNT and CLOCK/BMAL structures to model SIM1 structure has resulted in different models being predicted, highlighting the importance of obtaining structural data for SIM1 itself. This will allow us to gain a better understanding of SIM1 function, particularly for how mutations affect structure and function. Previous attempts to express and purify SIM1 for structural analysis have not been successful. Attempts to express SIM1 in bacteria and yeast have failed with limited success in insect cells, which was not viable for large scale protein expression required for structural analysis. In this study, we decided to use a mammalian cell system, Expi293F cells, to express truncated SIM1 and ARNT2, with the aim of optimising expression and purification conditions to reach the amounts and purity appropriate for structural analysis by either cryo-EM or crystallisation.

Optimisation of SIM1 and ARNT2 expression and purification

Transient transfection of SIM1.1-348.Flag/Strep and ARNT2.1-455.Myc plasmids into Expi293F cells resulted in higher expression of both truncated proteins compared to expression from a stable cell line. As we want to obtain the greatest yield of protein as possible for use in structural analysis, transient expression was therefore used for downstream optimisation. The results also showed that transient transfection was found to be easily scaled up as culture volumes were increased, to increase the yield of SIM1.1-348.Flag/Strep and ARNT2.1-455.Myc protein. Higher starting cell concentrations were shown to be beneficial, resulting in higher cell numbers and viability at harvest, thus higher protein yield. Expression of the correct proteins from the plasmids was confirmed by western blot using antibodies against each of the tags. As these proteins were expressed in mammalian cells, correct folding of the proteins should be maintained, as well as any PTMs, chaperones and co-factors. This is an advantage over previously used bacteria, yeast and insect expression systems that have been used in the past as they seemingly do not have the cell components required to maintain high levels of SIM1/ARNT2 mammalian proteins.

As seen in the results, Strep Co-IP of SIM1.1-348.Flag/Strep and ARNT2.1-455.Myc was successful, however a number of contaminants were also pulled out with the proteins of

interest. In particular, Hsp70 and ACC1 were pulled out in quantities comparable to the overexpressed SIM1.1-348.Flag/Strep and ARNT2.1-455.Myc. Hsp70 is a protein chaperone involved in a number of cellular activities, including ensuring proteins are correctly folded and targeting misfolded proteins for degradation [79]. This ability of binding to multiple proteins, including members of the bHLH/PAS family, often results in Hsp70 being pulled down as a contaminant during protein purification. Both the twin-Strep tag and biotin bind to the same binding pocket of the Strep-Tactin XT on the resin with high affinity, therefore biotin is used to elute the protein of interest from the resin as it competes for the same binding pocket. ACC1 uses endogenous biotin as a co-factor in order to function, so consequently the biotin bound ACC1 can also bind to the Strep-Tactin XT resin during the Co-IP protocol. This would explain why so much ACC1 is eluted with the proteins of interest.

As using a SEC did not appear to improve the purity of the SIM1.1-348.Flag/Strep and ARNT2.1-455.Myc solution, using more stringent wash buffers and including a secondary Co-IP step using Flag resin was attempted to remove the contaminants from the elutions. While secondary purification over the Flag resin considerably improved the purity of the elution, it did not completely remove all of the contaminating proteins, including ACC1. It was therefore decided to attempt purification using only the Flag resin with the stringent wash buffers, as it was expected that the ACC1 at least would not bind non-specifically to the Flag antibody. While no ACC1 was seen in the coomassie after Flag Co-IP, there was still a significant amount of Hsp70, as well as the presence of other protein contaminants in the elution. These results indicate a two-step purification method results in a cleaner purification product than using only one resin strategy for purification.

Future optimisation of purification of SIM1 and ARNT2

In addition to the optimisation methods presented here, optimised buffers, further steps or alternative purification methods will be required to remove the remaining contaminants to reach at least 95% purity which is required for structural analysis. This level of purity is required as high levels of contaminants will interfere with crystal formation or will distort image processing when used for cryo-EM.

The purity of the elutions can be further improved through pre-treating the lysates before performing Co-IP. For Strep purification, binding of biotin to the resin can be blocked by the addition of avidin, a biotin binding protein. Treatment with avidin should eliminate the binding of biotin bound ACC1 to the Strep resin. Lysates for Flag purification can be pre-cleared using either mouse IgG-agarose or unconjugated resin to bind non-specific proteins to reduce the amount of contaminating proteins that are eluted with SIM1.1-348.Flag/Strep and ARNT2.1-455.Myc. The use of more stringent wash buffers compared to what is used in the work presented here, could result in even more pure SIM1.1-348.Flag/Strep and ARNT2.1-455.Myc elutions. Both resins are compatible with higher salt and detergent concentrations than what was used in this work, therefore it would be feasible to increase

these concentrations in an attempt to help reduce non-specific binding of contaminating proteins.

An alternative approach could be to employ heparin affinity chromatography as an initial purification step, followed by either Strep or Flag Co-IP. Heparin affinity chromatography columns are widely used in purification methods due to the ability of heparin to bind to a wide range of proteins without the need for a tagged protein of interest. Heparin is particular good at binding to DNA binding proteins as its structure mimics that of nucleic acid. As SIM1 and ARNT2 are both transcription factors that bind to DNA in order to function, they would be able to be purified from cell lysates using a heparin column. The elutions from this column can then be used for either Strep or Flag Co-IP to obtain a highly pure solution of SIM1.1-348.Flag/Strep and ARNT2.1-455.Myc proteins.

The yield of SIM1.1-348.Flag/Strep and ARNT2.1-455.Myc purified protein (Table 3) is not yet at the levels to make structural studies feasible. Crystallisation studies require at least 1mg of starting protein for successful structural analysis. While the yields from these purifications do not indicate a sufficient amount will be obtained from using the Expi293F system for crystallisation, the resultant yields look promising for use in cryo-EM as it requires a lot less starting material. As the elution volumes acquired from the different purification methods are quite large, a concentration step is required. The results indicated a significant loss of protein during the concentrating steps using a centrifuge filter device, which would obviously impact the final yield of purified proteins for use in structural analysis. This could be due to the SIM1.1-348.Flag/Strep and ARNT2.1-455.Myc binding to the membrane of the filter device. Potential ways to avoid this would be to pre-treat the membrane overnight with buffer containing detergent to help avoid protein sticking to the membrane, as well as performing short spins and pipetting close to the membrane to prevent a concentration gradient forming with high protein concentration near the membrane. Alternatively, testing other concentration filter devices that use different membrane chemistry should be explored.

While further optimisation is necessary to obtain the yields and purity of SIM1.1-348.Flag/Strep and ARNT2.1-455.Myc protein required for structural analysis, the work presented here shows promise of being an appropriate system to achieve this goal. Even though determining a full length structure would be the ideal case, obtaining a truncated SIM1/ARNT2 heterodimer structure will still be a major leap forward in understanding how SIM1 structure and function are related, including DNA binding, Class II partner selection and co-factor interactions. Obtaining structures of SIM1 that contain obesity causing mutations, such as the R171H mutation, will also contribute to developing this understanding of function. Elucidating the SIM1 structure will also be beneficial in investigating how slight structural differences between bHLH/PAS family members plays a role in their specificity in choosing dimerisation partners and selecting their specific target genes despite binding to similar DNA recognition sequences.

Chapter 6: Final discussion and future perspectives

Many essential developmental and homeostatic pathways are regulated in part by members of the bHLH/PAS family of transcription factors. bHLH/PAS factors are essential for development, evident by the lethality of knockout mouse models, as well as playing important roles in response to changing environmental conditions, such as changes in nutritional cues and fluctuations in oxygen levels. As such, it is critical to understand the temporal and spatial function of these factors, in particular their contribution to disease states.

The aim of the work presented in this thesis was to further our understanding of the mechanisms and biological functions of the bHLH/PAS partner proteins SIM1 and ARNT2. This was divided into three separate aims: 1. Investigate the unique and overlapping functions of ARNT and ARNT2 by generating inducible shRNA mouse models; 2. Investigate the role of SIM1 in appetite regulation using a mutant SIM1 mouse model that recapitulates the human phenotype; 3. Optimise conditions for expression and purification of truncated SIM1/ARNT2 heterodimer for use in structural studies.

Prenatal and perinatal lethality of *Arnt* and *Arnt2* homozygous knockout mice have inhibited research into the unique and overlapping functions of these two proteins. Using a mouse model in which we can study the effects of reduced ARNT or ARNT2 beyond embryonic development will advance our understanding of the interplay of these factors with Class I factors, such as SIM1 and HIF1α, particularly during disease progression. It will also give insight into potential side effects of targeting the interaction of ARNT or ARNT2 with other family members as a therapeutic strategy to treat conditions such as SIM1 related obesity or HIF1α regulation in cancer. To overcome this hurdle of homozygous lethality, we generated Dox inducible, reversible shRNA mouse models for both *Arnt* and *Arnt2* that would allow study of reduced protein levels in adult mice. While this system showed robust knockdown in cell culture, the work presented in this thesis shows that the knockdown of ARNT and ARNT2 was unfortunately not translated into the mouse model, with mosaic knockdown only observed in Dox treated embryos with no observable change in protein levels in adult tissues. A review of the literature which utilises shRNA systems results in very few studies that successfully generate germline shRNA mouse models, all of which resulted in inconsistent knockdown of the target protein throughout the whole animal [45, 46]. After this thesis work began, research in the field of CRISPR technologies advanced in great leaps and bounds. Specifically, the development of CRISPRi and various methods that can be employed to tightly control gene expression [49-52] could be exploited in future experiments investigating the differential activities of ARNT and ARNT2 in adult physiology. Importantly, CRISPRi systems can be multiplexed to investigate complex pathways that ARNT or ARNT2 are involved in, for example targeting different components of the leptin-melanocortin signalling pathway which SIM1, presumably dimerised with ARNT2 [15, 16], plays a key role.

A key partner protein of ARNT2 is SIM1, a critical factor for the correct development of the hypothalamus and a key regulator of appetite. SIM1 deficiency results in hyperphagia induced obesity in both humans and mice. This thesis presents a mouse model in which the SIM1 weakly functioning mutation R171H, first identified in a human patient, is introduced into the mouse genome. For the first time, we can show that a single variant identified in a human patient can also result in hyperphagic obesity when introduced in a mouse model. A goal of the work presented here was to use this mouse model, which also expresses GFP under the control of the *Sim1* promoter, to isolate SIM1 expressing neuronal cells to use for analysis of SIM1 pathways and target genes by comparing WT and R171H SIM1 expressing cells. However, due to technical limitations of the experiments performed, we were unable to isolate appropriate numbers of SIM1 positive cells for use in downstream applications, such as RNAseq. Since undertaking this thesis work, there have been increasing studies published using circuit mapping and single cell RNAseq that have revealed a complex network of neuronal cell populations involved in appetite regulation [80-84], including a number of SIM1 expressing subpopulations within the hypothalamus [85, 86]. This highlights the requirement of future work in developing effective cell isolation methods in order to identify novel target genes of SIM1 and to further dissect the role of SIM1 in these complex pathways that regulate appetite.

Having structural data for a protein can help in understanding how it functions and aid in the development of targeted drug design. There is currently a lack of protein structures for bHLH/PAS proteins, including SIM1. The final goal of this thesis was to optimise a mammalian cell expression system and the purification conditions to yield truncated SIM1/ARNT2 heterodimer for use in structural studies. While the work presented here shows promise for obtaining purified SIM1/ARNT2 heterodimer, there is still a need for further optimisation to reach the purity and yield required for structural studies. Obtaining a SIM1 structure will be critical in helping investigations into multiple aspects of SIM1 function, such as protein-protein interactions, how the SIM1/ARNT2 heterodimer binds DNA, and importantly how mutations, such as R171H, affects structure and function resulting in a hyperphagic obesity phenotype.

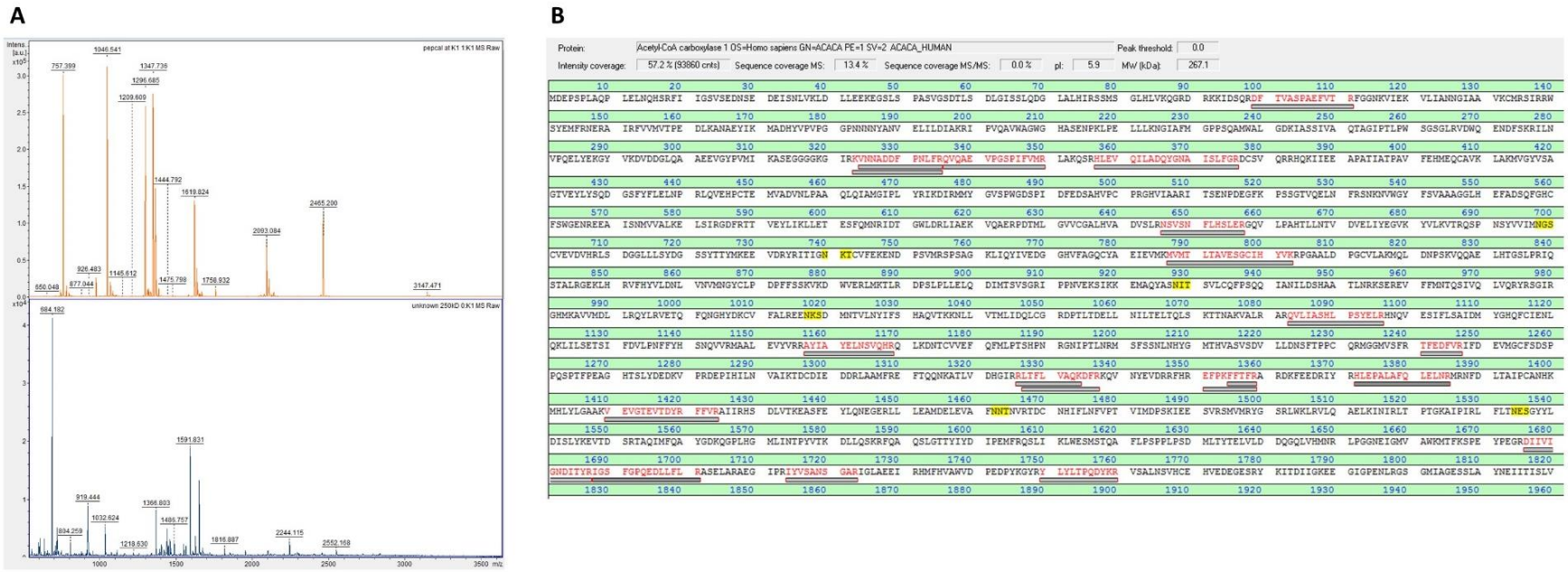

Figure 20. Identification of acetyl CoA Carboxylase 1 using mass spectrometry. The unknown 250kDa band of interest was excised from coomassie gel, destained and digested with trypsin. Peptides were analysed using MALDI TOF mass spectrometry (MS) and MS/MS to identify the unknown protein. **A.** Spectrum of peaks corresponding to fragments identified by MS. Pepcal standards (top) and unknown protein of interest (bottom). **B.** Alignment of fragments identified by MS to acetyl CoA carboxylase 1 protein sequence.

References

1. Kewley, R.J., M.L. Whitelaw, and A. Chapman-Smith, *The mammalian basic helix-loop-helix/PAS family of transcriptional regulators.* International Journal of Biochemistry & Cell Biology, 2004. **36**(2): p. 189-204.
2. Partch, C.L. and K.H. Gardner, *Coactivator Recruitment: A New Role for PAS Domains in Transcriptional Regulation by the bHLH-PAS Family.* Journal of Cellular Physiology, 2010. **223**(3): p. 553-557.
3. Bersten, D.C., et al., *bHLH-PAS proteins in cancer.* Nature Reviews Cancer, 2013. **13**(12): p. 827-841.
4. Crews, S.T. and C.M. Fan, *Remembrance of things PAS: regulation of development by bHLH-PAS proteins.* Current Opinion in Genetics & Development, 1999. **9**(5): p. 580-587.
5. Ema, M., et al., *Two new members of the murine Sim gene family are transcriptional repressors and show different expression patterns during mouse embryogenesis.* Molecular and Cellular Biology, 1996. **16**(10): p. 5865-5875.
6. Jain, S., et al., *Potent transactivation domains of the Ah receptor and the Ah receptor nuclear translocator map to their carboxyl termini.* Journal of Biological Chemistry, 1994. **269**(50): p. 31518-31524.
7. Jiang, B.H., et al., *Dimerization, DNA binding, and transactivation properties of hypoxia-inducible factor 1.* Journal of Biological Chemistry, 1996. **271**(30): p. 17771-17778.
8. Crews, S.T., *Control of cell lineage-specific development and transcription by bHLH-PAS proteins.* Genes & Development, 1998. **12**(5): p. 607-620.
9. Aitola, M.H. and M.T. Pelto-Huikko, *Expression of Arnt and Arnt2 mRNA in developing murine tissues.* Journal of Histochemistry & Cytochemistry, 2003. **51**(1): p. 41-54.
10. 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.
11. 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).* Molecular and Cellular Biology, 1996. **16**(4): p. 1706-1713.
12. 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.
13. Wang, G.L., et al., *Hypoxia-inducible factor-1 is a basic-helix-loop-helix-PAS heterodimer regulated by cellular O-2 tension.* Proceedings of the National Academy of Sciences of the United States of America, 1995. **92**(12): p. 5510-5514.
14. 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.
15. Hosoya, T., et al., *Defective development of secretory neurones in the hypothalamus of Arnt2-knockout mice.* Genes to Cells, 2001. **6**(4): p. 361-374.
16. Michaud, J.L., et al., *ARNT2 acts as the dimerization partner of SIM1 for the development of the hypothalamus.* Mechanisms of Development, 2000. **90**(2): p. 253-261.
17. Jain, S., et al., *Expression of ARNT, ARNT2, HIF1 alpha, HIF2 alpha and Ah receptor mRNAs in the developing mouse.* Mechanisms of Development, 1998. **73**(1): p. 117-123.
18. 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.* Molecular and Cellular Neuroscience, 1996. **7**(1): p. 1-16.
19. Iyer, N.V., et al., *Cellular and developmental control of O-2 homeostasis by hypoxia-inducible factor 1 alpha.* Genes & Development, 1998. **12**(2): p. 149-162.
20. 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.*

Proceedings of the National Academy of Sciences of the United States of America, 2001. **98**(12): p. 6692-6697.
21. Kozak, K.R., B. Abbott, and O. Hankinson, *ARNT-deficient mice and placental differentiation.* Developmental Biology, 1997. **191**(2): p. 297-305.
22. Maltepe, E., et al., *Abnormal angiogenesis and responses to glucose and oxygen deprivation in mice lacking the protein ARNT.* Nature, 1997. **386**(6623): p. 403-407.
23. Michaud, J.L., et al., *Development of neuroendocrine lineages requires the bHLH-PAS transcription factor SIM1.* Genes & Development, 1998. **12**(20): p. 3264-3275.
24. Ryan, H.E., J. Lo, and R.S. Johnson, *HIF-1 alpha is required for solid tumor formation and embryonic vascularization.* Embo Journal, 1998. **17**(11): p. 3005-3015.
25. Kotch, L.E., et al., *Defective vascularization of HIF-1 alpha-null embryos is not associated with VEGF deficiency but with mesenchymal cell death.* Developmental Biology, 1999. **209**(2): p. 254-267.
26. Carmeliet, P., et al., *Abnormal blood vessel development and lethality in embryos lacking a single VEGF allele.* Nature, 1996. **380**(6573): p. 435-9.
27. Ferrara, N., et al., *Heterozygous embryonic lethality induced by targeted inactivation of the VEGF gene.* Nature, 1996. **380**(6573): p. 439-42.
28. Tomita, S., et al., *Conditional disruption of the aryl hydrocarbon receptor nuclear translocator (Arnt) gene leads to loss of target gene induction by the aryl hydrocarbon receptor and hypoxia-inducible factor 1 alpha.* Molecular Endocrinology, 2000. **14**(10): p. 1674-1681.
29. 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.
30. Ramachandrappa, S., et al., *Rare variants in single-minded 1 (SIM1) are associated with severe obesity.* Journal of Clinical Investigation, 2013. **123**(7): p. 3042-3050.
31. Schindelin, J., et al., *Fiji: an open-source platform for biological-image analysis.* Nat Methods, 2012. **9**(7): p. 676-82.
32. Gerassimou, A.G., *Investigating biological roles of bHLH/PAS transcription factors*. 2014, University of Adelaide.
33. Gong, S.C., et al., *A gene expression atlas of the central nervous system based on bacterial artificial chromosomes.* Nature, 2003. **425**(6961): p. 917-925.
34. Wu, S., et al., *A protocol for constructing gene targeting vectors: generating knockout mice for the cadherin family and beyond.* Nat Protoc, 2008. **3**(6): p. 1056-76.
35. Hrvatin, S., et al., *MARIS: Method for Analyzing RNA following Intracellular Sorting.* Plos One, 2014. **9**(3).
36. Dickins, R.A., et al., *Tissue-specific and reversible RNA interference in transgenic mice.* Nature Genetics, 2007. **39**(7): p. 914-921.
37. Dow, L.E., et al., *A pipeline for the generation of shRNA transgenic mice.* Nat Protoc, 2012. **7**(2): p. 374-93.
38. Premsrirut, P.K., et al., *A Rapid and Scalable System for Studying Gene Function in Mice Using Conditional RNA Interference.* Cell, 2011. **145**(1): p. 145-158.
39. Bersten, D.C., et al., *Inducible and Reversible Lentiviral and Recombination Mediated Cassette Exchange (RMCE) Systems for Controlling Gene Expression.* Plos One, 2015. **10**(3).
40. Rubinson, D.A., et al., *A lentivirus-based system to functionally silence genes in primary mammalian cells, stem cells and transgenic mice by RNA interference.* Nature Genetics, 2003. **33**(3): p. 401-406.
41. Seibler, J., et al., *Reversible gene knockdown in mice using a tight, inducible shRNA expression system.* Nucleic Acids Res, 2007. **35**(7): p. e54.
42. Vidigal, J.A., et al., *An inducible RNA interference system for the functional dissection of mouse embryogenesis.* Nucleic Acids Res, 2010. **38**(11): p. e122.

43. Zhou, X., et al., *Optimization of the Tet-On system for regulated gene expression through viral evolution.* Gene Ther, 2006. **13**(19): p. 1382-90.
44. Loew, R., et al., *Improved Tet-responsive promoters with minimized background expression.* BMC Biotechnol, 2010. **10**: p. 81.
45. Alorro, M.G., et al., *Generation of an inducible mouse model to reversibly silence Stat3.* Genesis, 2017. **55**(4).
46. Cassidy, L.D., et al., *A novel Atg5-shRNA mouse model enables temporal control of Autophagy in vivo.* Autophagy, 2018. **14**(7): p. 1256-1266.
47. Hofmann, A., et al., *Epigenetic regulation of lentiviral transgene vectors in a large animal model.* Mol Ther, 2006. **13**(1): p. 59-66.
48. Fellmann, C., et al., *An Optimized microRNA Backbone for Effective Single-Copy RNAi.* Cell Reports, 2013. **5**(6): p. 1704-1713.
49. Navabpour, S., J.L. Kwapis, and T.J. Jarome, *A neuroscientist's guide to transgenic mice and other genetic tools.* Neurosci Biobehav Rev, 2020. **108**: p. 732-748.
50. Zheng, Y., et al., *CRISPR interference-based specific and efficient gene inactivation in the brain.* Nat Neurosci, 2018. **21**(3): p. 447-454.
51. MacLeod, R.S., et al., *Effective CRISPR interference of an endogenous gene via a single transgene in mice.* Sci Rep, 2019. **9**(1): p. 17312.
52. Qi, L.S., et al., *Repurposing CRISPR as an RNA-guided platform for sequence-specific control of gene expression.* Cell, 2013. **152**(5): p. 1173-83.
53. Alvarez-Bolado, G., *Development of neuroendocrine neurons in the mammalian hypothalamus.* Cell Tissue Res, 2019. **375**(1): p. 23-39.
54. Biran, J., et al., *Role of developmental factors in hypothalamic function.* Frontiers in Neuroanatomy, 2015. **9**.
55. Michaud, J.L., et al., *Sim1 haploinsufficiency causes hyperphagia, obesity and reduction of the paraventricular nucleus of the hypothalamus.* Human Molecular Genetics, 2001. **10**(14): p. 1465-1473.
56. Holder, J.L., Jr., et al., *Sim1 gene dosage modulates the homeostatic feeding response to increased dietary fat in mice.* American Journal of Physiology, 2004. **287**(1(1)): p. E105-E113.
57. Kublaoui, B.M., et al., *Sim1 haploinsufficiency impairs melanocortin-mediated anorexia and activation of paraventricular nucleus neurons.* Molecular Endocrinology, 2006. **20**(10): p. 2483-2492.
58. Kublaoui, B.M., et al., *SIM1 overexpression partially rescues agouti yellow and diet-induced obesity by normalizing food intake.* Endocrinology, 2006. **147**(10): p. 4542-4549.
59. 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.
60. Tolson, K.P., et al., *Postnatal Sim1 Deficiency Causes Hyperphagic Obesity and Reduced Mc4r and Oxytocin Expression.* Journal of Neuroscience, 2010. **30**(10): p. 3803-3812.
61. Tolson, K.P., et al., *Inducible neuronal inactivation of Sim1 in adult mice causes hyperphagic obesity.* Endocrinology, 2014. **155**(7): p. 2436-44.
62. Hossain, M.S., et al., *Identification of mutations through dominant screening for obesity using C57BL/6 substrains.* Sci Rep, 2016. **6**: p. 32453.
63. Sullivan, A.E., et al., *Characterization of human variants in obesity-related SIM1 protein identifies a hot-spot for dimerization with the partner protein ARNT2.* Biochemical Journal, 2014. **461**: p. 403-412.
64. Moir, L., et al., *Disruption of the homeodomain transcription factor orthopedia homeobox (Otp) is associated with obesity and anxiety.* Mol Metab, 2017. **6**(11): p. 1419-1428.
65. Wu, D. and F. Rastinejad, *Structural characterization of mammalian bHLH-PAS transcription factors.* Curr Opin Struct Biol, 2017. **43**: p. 1-9.
66. Huang, N., et al., *Crystal structure of the heterodimeric CLOCK:BMAL1 transcriptional activator complex.* Science (New York, N.Y.), 2012. **337**(6091): p. 189.

67. Wu, D., et al., *Structural integration in hypoxia-inducible factors.* Nature, 2015. **524**(7565): p. 303-8.
68. Bonnefond, A., et al., *Loss-of-function mutations in SIM1 contribute to obesity and Prader-Willi-like features.* J Clin Invest, 2013. **123**(7): p. 3037-41.
69. Cardoso, R., et al., *Identification of Cys255 in HIF-1α as a novel site for development of covalent inhibitors of HIF-1α/ARNT PasB domain protein-protein interaction.* Protein Sci, 2012. **21**(12): p. 1885-96.
70. Erbel, P.J., et al., *Structural basis for PAS domain heterodimerization in the basic helix--loop--helix-PAS transcription factor hypoxia-inducible factor.* Proc Natl Acad Sci U S A, 2003. **100**(26): p. 15504-9.
71. Card, P.B., P.J.A. Erbel, and K.H. Gardner, *Structural basis of ARNT PAS-B dimerization: Use of a common beta-sheet interface for hetero- and homodimerization.* Journal of Molecular Biology, 2005. **353**(3): p. 664-677.
72. Wu, D., et al., *Structure and dimerization properties of the aryl hydrocarbon receptor PAS-A domain.* Mol Cell Biol, 2013. **33**(21): p. 4346-56.
73. Guo, Y., et al., *Coiled-coil coactivators play a structural role mediating interactions in hypoxia-inducible factor heterodimerization.* J Biol Chem, 2015. **290**(12): p. 7707-21.
74. Scheuermann, T.H., et al., *Artificial ligand binding within the HIF2alpha PAS-B domain of the HIF2 transcription factor.* Proc Natl Acad Sci U S A, 2009. **106**(2): p. 450-5.
75. Schulte, K.W., et al., *Structural Basis for Aryl Hydrocarbon Receptor-Mediated Gene Activation.* Structure, 2017. **25**(7): p. 1025-1033.e3.
76. Seok, S.H., et al., *Structural hierarchy controlling dimerization and target DNA recognition in the AHR transcriptional complex.* Proc Natl Acad Sci U S A, 2017. **114**(21): p. 5431-5436.
77. Wu, D., et al., *NPAS1-ARNT and NPAS3-ARNT crystal structures implicate the bHLH-PAS family as multi-ligand binding transcription factors.* Elife, 2016. **5**.
78. Wang, Z., et al., *Intermolecular recognition revealed by the complex structure of human CLOCK-BMAL1 basic helix-loop-helix domains with E-box DNA.* Cell Res, 2013. **23**(2): p. 213-24.
79. Rosenzweig, R., et al., *The Hsp70 chaperone network.* Nat Rev Mol Cell Biol, 2019. **20**(11): p. 665-680.
80. Campbell, J.N., et al., *A molecular census of arcuate hypothalamus and median eminence cell types.* Nat Neurosci, 2017. **20**(3): p. 484-496.
81. Chen, R., et al., *Single-Cell RNA-Seq Reveals Hypothalamic Cell Diversity.* Cell Rep, 2017. **18**(13): p. 3227-3241.
82. Kim, D.W., et al., *The cellular and molecular landscape of hypothalamic patterning and differentiation.* bioRxiv, 2020: p. 657148.
83. Romanov, R.A., et al., *Molecular design of hypothalamus development.* Nature, 2020. **582**(7811): p. 246-252.
84. Xu, S., et al., *Behavioral state coding by molecularly defined paraventricular hypothalamic cell type ensembles.* Science, 2020. **370**(6514).
85. Li, M.M., et al., *The Paraventricular Hypothalamus Regulates Satiety and Prevents Obesity via Two Genetically Distinct Circuits.* Neuron, 2019. **102**(3): p. 653-667.e6.
86. Shah, B.P., et al., *MC4R-expressing glutamatergic neurons in the paraventricular hypothalamus regulate feeding and are synaptically connected to the parabrachial nucleus.* Proceedings of the National Academy of Sciences of the United States of America, 2014. **111**(36): p. 13193-13198.