Function and Mechanism of the Single-Minded 2 Gene

Emily Lyn Button

B.Sc. Honours (Biochemistry)

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

Department of Molecular and Biomedical Science

School of Biological Sciences

University of Adelaide, Australia

May 2023

Contents

Abstract

Single-Minded 2 (SIM2) is a member of the basic Helix-Loop-Helix PER-ARNT-SIM (bHLH/PAS) family of transcription factors which are known to play diverse roles in development, cellular homeostasis, and disease. The amino-terminal halves of bHLH/PAS transcription factors contain a basic DNA binding region followed by a helix-loop-helix dimerisation domain. This is then followed by a PAS domain consisting of two PAS repeats, PAS-A and PAS-B which also function in dimerisation and protein:protein interaction and ligand binding. Their carboxy-terminal halves contain transactivation or transrepression domains for target gene regulation.

In mice *Sim2* is essential for normal development. Mice lacking *Sim2* have been found to have multiple abnormal phenotypes, including abnormal skeletal structures and overgrowth of gut bacteria. *SIM2* has also been implicated in the progression of several cancers, with its function appearing to be highly context dependent. Upregulation of *SIM2* in prostate, pancreatic and colon cancers favours tumour progression, whereas downregulation of *SIM2* in breast cancer favours tumour progression. There is still much to discover regarding the function and target genes of SIM2 during development and in human disease. Therefore, the aim of this thesis was to further investigate the functions and mechanisms of action of *SIM2* both in developmental and disease contexts.

To investigate whether SIM2 may contribute to the pathogenesis of human developmental disorders, *SIM2* non-synonymous gene variants identified in patients with neurological phenotypes were functionally assessed to determine their impact on activity of SIM2 as a transcription factor. This study identified five variants that caused a significant reduction in the transcriptional activating potency of SIM2 and were further characterised to determine the mechanism associated with the deficiency. This work identified a set of residues that are important for the function of SIM2 as a transcription factor and may contribute to human pathology.

To investigate the function of SIM2 in breast cancer, the weak *SIM2* expressing MDA-MB-231 cell lines was modified to enable inducible upregulation of SIM2 and

subsequently subjected to RNA-sequencing. This study found a set of genes that were significantly differentially expressed upon upregulation of SIM2, culminating in a proposed mechanism whereby SIM2 crosstalks with other bHLH/PAS transcription factors to modulate their protumourigenic functions.

To investigate the function of *Sim2* during development, a conditional *Sim2* knockout mouse model was generated to selectively remove *Sim2* expression from the brain. These mice underwent feeding and behavioural studies to assess the impact of *Sim2* knockout in the brain, however the tests did not identify any significant differences between the conditional knockout animals and their normal litter mate controls. An epitope tagged SIM2 mouse line was generated that will provide a valuable tool for assessing endogenous SIM2 protein functions and interactions *in vivo*. In addition, attempts were made to generate a reporter mouse in which expression of *Sim2* would be replaced with a fluorescent protein. While various CRISPR based attempts were not successful in replacing a *Sim2* allele with fluorescent Tomato coding sequence in zygotes, cultured mouse embryonic stem cells were successfully targeted. These *Sim2-Tomato* ES cells can be used in future to generate the desired *Sim2-Tomato* reporter mouse line.

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.

The author acknowledges that copyright of published works contained within the thesis resides with the copyright holder(s) of those works.

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.

06/12/2022

Emily Lyn Button   Date

Acknowledgements

Firstly, I would like to thank my supervisor Murray for all your support and guidance throughout my studies. Thank you for accepting me as an Honours student and for encouraging me to continue to a PhD. Your invaluable advice, feedback, and wealth of knowledge has made me the scientist that I am today and for that I will be forever grateful. Thank you also to my co-supervisor Paul for your support and assistance, in particular with the mouse model aspects of my project.

Thank you to all past and present members of the Whitelaw Laboratory. Adrienne for all of your help and advice, especially when I was starting out as an Honours student. Thank you to Dave, Veronica, Alexis, Joe and Tim for your help, advice, enthusiasm, for being my sounding board and for your friendship over these years.

Thank you to the members of the Peet Laboratory for all your valuable advice, support, and feedback. A special thank you to Ice for being there with me through our Honours and PhD years. Your friendship and support certainly helped me get through the highs and lows. I would also like to give thanks to all the members of the Molecular and Biomedical Science Department I have crossed paths with, there are too many to name but you have all helped me along this journey.

I would also like to extend my gratitude to all those who have contributed to the work in this thesis. Thank you to Dr Emily Jahne for performing the behavioural studies, Sandy and Mel for the mouse embryology and transfers, and my co-authors on the publications presented in this thesis, Joe, Daniel, John, Dave, Dan, Jill, and Murray.

Most importantly I would like to thank my family and friends. I truly would not be where I am today without your support and guidance and for this, I can't thank you all enough. Thank you to my parents Lynette and Michael, and my brother Robert for always being there for me. And thank you especially to my husband Nathan, I could not have done this without you.

Publications

Button, E. L., Bersten, D. C., & Whitelaw, M. L. (2017). "HIF has Biff - Crosstalk between HIF1a and the family of bHLH/PAS proteins". *Exp Cell Res, 356*(2), 141-145. doi: 10.1016/j.yexcr.2017.03.055

Button, E. L., Rossi, J. J., McDougal, D. P., Bruning, J. B., Peet, D. J., Bersten, D. C., Rosenfeld, J.A., Whitelaw, M. L. (2022). Characterization of functionally deficient SIM2 variants found in patients with neurological phenotypes. Biochemical Journal, 479(13), 1441-1454. doi: 10.1042/bcj20220209

Conference and Symposia Presentations

Oral Presentations

Button E.L., Bersten D.C, Whitelaw M.L. (2017, October). *Analysis of transcription factor SIM2 variants present in patients with intellectual disabilities*. Poster teaser oral presentation at ComBio, Adelaide, SA.

Button E.L., Bersten D.C, Whitelaw M.L. (2017, November). *Analysis of transcription factor SIM2 variants present in patients with intellectual disabilities*. Oral presentation at ANZSCDB Meeting, Adelaide, SA.

Button E.L., Sullivan A.E., Bersten D.C, Whitelaw M.L. (2018, September). *Not so Single-Minded: SIM2 function and cross-talk in breast cancer*. Oral presentation at Oxygen et al Meeting, Flinders University, Adelaide, SA.

Button E.L., Sullivan A.E., Bersten D.C, Whitelaw M.L. (2018, November). *SIM2 function and cross-talk in breast cancer*. Oral presentation at ANZSCDB Adelaide Meeting, Adelaide, SA.

Poster Presentations

Button E. L., Lines T.E.P., Thomas P.Q., Whitelaw M.L. (2014, October). *Analysis of mouse models of the Single-Minded 2 gene*. Poster session presented at ComBio, Canberra, ACT.

Button E. L., Lines T.E.P., Thomas P.Q., Whitelaw M.L. (2015, July). *Analysis of mouse models of the Single-Minded 2 gene*. Poster session presented at the University of Adelaide Molecular and Biomedical Sciences Postgraduate Symposium, Adelaide, SA.

Button E., Piltz S., Thomas P., Whitelaw M. (2016, November). *Developing mouse models to investigate functions of the Single-Minded 2 gene*. Poster session presented at ANZSCDB Meeting, Adelaide, SA.

Button E.L., Bersten D.C, Whitelaw M.L. (2017, October). *Analysis of transcription factor SIM2 variants present in patients with intellectual disabilities*. Poster session presented at ComBio, Adelaide, SA.

Button E.L, Bersten D.C., Mokry J.A, Whitelaw M.L. (2018, February) *Analysis of functionally deficient variants of transcription factor SIM2 present in patients with intellectual disabilities*. Poster session presented at Lorne Genome Conference, Lorne, Victoria.

Chapter 1: Introduction

1.1 bHLH/PAS Transcription factors

1.1.1 Mechanism of action and molecular functions

The basic Helix-Loop-Helix/PER-ARNT-SIM (bHLH/PAS) proteins are a family of highly conserved transcription factors. Members of this family are known to function in diverse roles including oxygen homeostasis, stress response, development, and the circadian rhythm. These transcription factors are named so due to their conserved amino-terminal domains (Figure 1). The bHLH domain consists of the basic DNA binding regions, and a helix-loop-helix motif which forms the primary dimerisation interface. This is followed by a PAS domain consisting of two PAS repeats, termed PAS-A and PAS-B, which function as a secondary dimerisation interface and controls partner protein specificity (Crews & Fan, 1999; Kewley, Whitelaw, & Chapman-Smith, 2004). PAS-B has also been shown to function in signal regulation through small molecule and regulatory protein interactions, recruitment of cofactors and protein stability through ubiquitylation (Furness, Lees, & Whitelaw, 2007; Guo et al., 2013; N. Hao & Whitelaw, 2013; Huang et al., 2012; Okui et al., 2005; Partch & Gardner, 2011; To, Sedelnikova, Samons, Bonner, & Huang, 2006). The carboxy-terminal halves of these proteins, which share no conserved homology and contain transcriptional activation and repression domains for target gene regulation, have been shown to be regions for coactivator recruitment for some of the family members (Figure 1.1) (Beischlag et al., 2002; Bersten, Sullivan, Peet, & Whitelaw, 2013; Emily L. Button, Bersten, & Whitelaw, 2017; Crews & Fan, 1999; Kewley et al., 2004; F. Wang, Zhang, Wu, & Hankinson, 2010).

There are two classes of bHLH/PAS transcription factors, class I and class II, which form heterodimers in order to become functional transcription factor complexes. Class I factors are signal regulated and/or have spatiotemporally regulated expression and include the Single-Minded proteins (SIM1, SIM2), the Hypoxia Inducible Factors (HIF1α, HIF2α), Aryl Hydrocarbon Receptor (AHR), Neuronal PAS proteins (NPAS1,

NPAS3, NPAS4) and circadian rhythm factors (NPAS2, CLOCK). Class II factors are more ubiquitously expressed and include the Aryl Hydrocarbon Receptor Nuclear Translocator proteins (ARNT, ARNT2) and circadian rhythm factor Brain and Muscle ARNT-like 1 (BMAL1). Class II factors act as general partner factors and can heterodimerise with multiple class I factors. The heterodimers bind to atypical E-box like DNA sequences to regulate target gene expression (Figure 1.1) (Bersten et al., 2013; Emily L. Button et al., 2017; Crews & Fan, 1999; Kewley et al., 2004).

These transcription factors are essential for normal development, homeostasis and stress response, with knockout mouse models revealing lack of these proteins leads to lethality in many cases (K.-J. Chen, Lizaso, & Lee, 2014; Goshu et al., 2002; Hosoya et al., 2001; Keith, Adelman, & Simon, 2001; Kotch, Iyer, Laughner, & Semenza, 1999; Kozak, Abbott, & Hankinson, 1997; Maltepe, Schmidt, Baunoch, Bradfield, & Simon, 1997; J. L. Michaud, Rosenquist, May, & Fan, 1998; Shamblott, Bugg, Lawler, & Gearhart, 2002; Tian, Hammer, Matsumoto, Russell, & McKnight, 1998). However, when they are mutated or aberrantly regulated this can lead to human disease states, including cancer (where many of the factors have been implicated in having both pro- and anti-tumourigenic roles), hyperphagic obesity and developmental disorders (Bersten et al., 2013; Bonnefond et al., 2013; Ramachandrappa et al., 2013; Webb et al., 2013). Therefore, in depth study of these transcription factors is required to understand their molecular functions in both normal physiology and in relevant disease contexts.
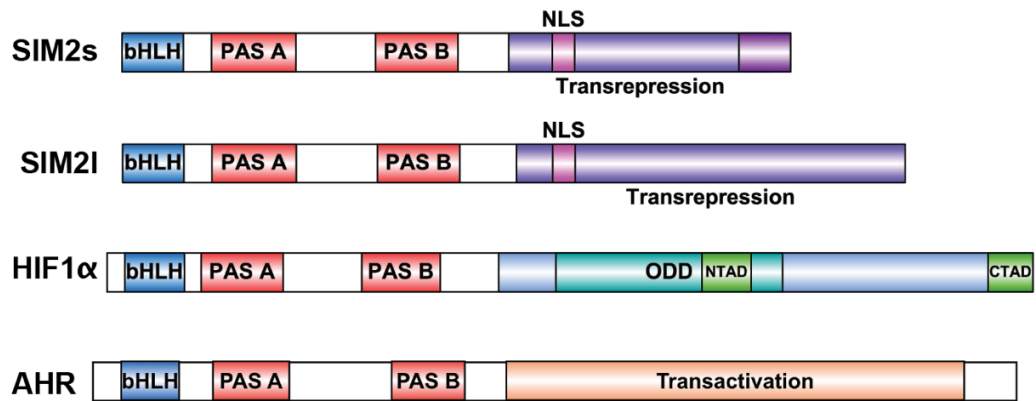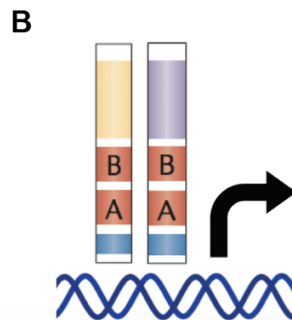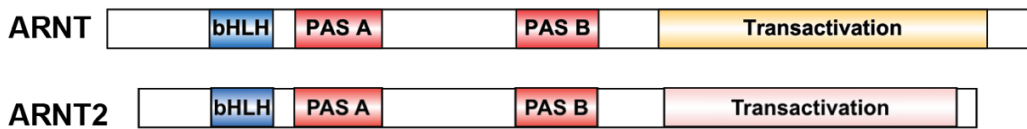
Figure 1.1: Mammalian bHLH/PAS transcription factors. **A)** Schematic of class I (SIM2s, SIM2l, HIF1α, AHR) and class II (ARNT, ARNT2) bHLH/PAS transcription factors relevant to this thesis. The N-terminal halves are architecturally conserved among family members. The bHLH domain functions in DNA binding and primary dimerisation. The PAS domain (PAS A and PAS B) also functions in dimerisation as well as mediating ligand binding and protein:protein interactions in some family members. The C-terminal halves contain transcriptional regulatory domains for control of target gene expression. **B)** Class I and class II factors must form heterodimers to become functional transcription factors that can bind to their DNA response elements and regulate target gene expression. bHLH: basic Helix-Loop-Helix, PAS: PER-ARNT-SIM, NLS: nuclear localisation signal, ODD: oxygen degradation domain, NTAD: N-terminal transactivation domain, CTAD: C-terminal transactivation domain.

This thesis will focus on the molecular functions and mechanisms of action of the class I bHLH/PAS transcription factor Single-Minded 2 (SIM2). Other factors relevant to this thesis include the class I factors Hypoxia Inducible Factors (HIF1α, HIF2α) and the Aryl Hydrocarbon Receptor (AHR), and the class II factors Aryl Hydrocarbon Receptor Nuclear Translocator (ARNT) and ARNT2.

HIF1α and HIF2α are known as the master regulators of the cellular response to hypoxia. Expression and activity of the HIFα proteins are tightly regulated by oxygen dependent hydroxylase enzymes. Under normal oxygen conditions (normoxia), HIFα subunits are inactivated and rapidly degraded through hydroxylation of conserved proline and asparagine residues in their C-terminal halves. Hydroxylation at two proline residues in the Oxygen Dependent Degradation (ODD) domain by the Prolyl Hydroxylase enzymes (PHDs) allows for recruitment of the Von Hippel Lindau protein (VHL), which leads to ubiquitylation and subsequent proteasomal degradation of the HIFα proteins (Ivan et al., 2001; Jaakkola et al., 2001; Gregg L. Semenza, 2012; G. L. Semenza, 2014; F. Yu, White, Zhao, & Lee, 2001). Transcriptional activity of HIFα is regulated by Factor Inhibiting HIF (FIH) which hydroxylates an asparagine residue in the C-terminal Transactivation Domain (CTAD), blocking recruitment of the CREB-binding protein (CBP)/p300 transcriptional coactivator complex (Lando, Peet, Gorman, et al., 2002; Lando, Peet, Whelan, Gorman, & Whitelaw, 2002). Under hypoxic conditions, these hydroxylase enzymes are inactive, leading to stable and transcriptionally active HIFα which can then translocate to the nucleus and become a functional transcription factor complex through dimerising with its partner protein ARNT (also known as HIF1β) or ARNT2. HIF heterodimers upregulate transcription of genes for cellular adaptation to hypoxia (e.g. *EPO*, *VEGF*, *BNIP3*, and *BNIP3L* and the glycolytic genes *PDK1*, *LDHA*,) through binding Hypoxic Response Elements (HRE) in promoters of target genes (Gregg L. Semenza, 2012; G. L. Semenza, 2014). The HIFs are also implicated in the pathology of many diseases, including cancer where they have been shown to have both pro- and anti-tumourigenic roles (Bersten et al., 2013; G. L. Semenza, 1999, 2003; Gregg L. Semenza, 2012; G. L. Semenza, 2014; G. L. Wang, Jiang, Rue, & Semenza, 1995).

AHR is a ligand activated bHLH/PAS transcription factor that functions in xenobiotic metabolism and immune cell development and response. It has been shown to respond to a number of xenobiotic ligands such as polycyclic and halogenated aromatic hydrocarbons (Denison & Nagy, 2003; N. Hao & Whitelaw, 2013; Waller & McKinney, 1995; Whelan, Hao, Furness, Whitelaw, & Chapman-Smith, 2010). In the absence of ligand, AHR is held inactive in the cytoplasm by chaperone proteins. Upon ligand binding, AHR translocates to the nucleus and dimerises with its partner factor ARNT to become a functional transcription factor. AHR/ARNT dimers bind to Xenobiotic Response Elements (XRE) and upregulate genes to initiate xenobiotic metabolism such as cytochrome P4501A1 (CYP1A1) and P4501B1 (CYP1B1). AHR is also implicated in cancer as having both pro- and anti-tumourigenic roles (Bersten et al., 2013; N. Hao & Whitelaw, 2013; Murray, Patterson, & Perdew, 2014; Safe, Cheng, & Jin, 2017).

ARNT and ARNT2 are class II bHLH/PAS transcription factors that act as general dimerisation partner factors for multiple class I factors including SIM2, HIF1α, HIF2α and AHR. ARNT is ubiquitously expressed with low expression in the brain. ARNT2 has reciprocal expression, being highly prevalent the brain and kidneys (Aitola & Pelto-Huikko, 2003; Nan Hao, Bhakti, Peet, & Whitelaw, 2013; Jain, Maltepe, Lu, Simon, & Bradfield, 1998; Sojka, Kern, & Pollenz, 2000). Hence it is thought that ARNT2 is the predominant neuronal dimerisation partner factor, whereas ARNT is the dimerisation factor in other tissues in the body. ARNT has been shown to contain a transactivation domain within the C-terminal half (Kewley et al., 2004; Whitelaw, Gustafsson, & Poellinger, 1994).

1.1.2   Crosstalk between bHLH/PAS family members

Given the high degree of conservation in the N-terminal halves of bHLH/PAS transcription factors, crosstalk between family members is likely, and has been shown to occur in a number of different contexts. Crosstalk can occur through competition between class I factors for dimerisation with their common class II factors. In situations where the class II factor is limiting, this can reduce the ability of class I factors to regulate expression of their individual target genes. In addition to this, the DNA response elements bound by these transcription factors are very similar, if not

identical, leading to the possibility that certain genes may be able to be regulated by multiple class I factors. This has been experimentally shown for a number of genes and is likely the case for many more in both normal physiological and disease contexts (Emily L. Button et al., 2017). For example, the pro-apoptotic gene *BNIP3* (BCL2/Adenovirus E1B 19 kDa-Interacting Protein 3) is known to be upregulated under hypoxic conditions by HIF1α. However SIM2 can also compete with HIF1α for the *BNIP3* HRE, downregulating the HIF mediated induction of this gene (Farrall & Whitelaw, 2009) (Whitelaw lab unpublished).

It is thought that crosstalk is, to an extent, avoided through regulation of the class I factors, which commonly have tissue restricted expression and/or are signal regulated. However, there are many tissues and cell types where multiple class I factors are expressed and can be active transcription factors. Therefore, a number of questions surrounding this idea still remain such as, how is crosstalk avoided, and are there different functional biological outcomes of crosstalk between family members?

Recent studies have shown that there is crosstalk between the circadian rhythm pathway, known to be controlled by CLOCK/BMAL, and oxygen sensing pathway, controlled by HIF1α. In certain contexts, it appears as though components of these pathways may be able to regulate one another and also potentially work together to regulate gene expression (Adamovich, Ladeuix, Golik, Koeners, & Asher, 2017; Emily L. Button et al., 2017; Eckle et al., 2012; Peek et al., 2017; Y. Wu et al., 2017). While this area of research is fairly new, it is likely there is much more to be discovered here and that these signalling pathways are more intertwined than initially thought. Similarly, in the immune system both AHR and HIF are known to play roles in the function and differentiation of a number of immune cell types. Emerging evidence suggests there may be a complex interplay between these two signalling pathways important to immunomodulation. These concepts are discussed in more detail in Chapter 2 (Emily L. Button et al., 2017).

In contexts where these factors are aberrantly regulated, such as cancer, there may be opportunity for crosstalk to occur between family members. Upregulation of a number of bHLH/PAS transcription factors, namely the HIFs, AHR and SIM2, has been

implicated in cancer progression. Studies have shown that reversing the increase in activity of these transcription factors through genetic interference and small molecule inhibition, growth and tumourigenesis of certain types of cancer cells both *in vitro* and *in vivo* can be decreased (Aleman et al., 2005; Bersten et al., 2013; DeYoung, Tress, & Narayanan, 2003a, 2003b; Kolluri, Jin, & Safe, 2017; Lu, Asara, Sanda, & Arredouani, 2011; G. L. Semenza, 2003, 2014; Xue, Fu, & Zhou, 2018; T. Yu, Tang, & Sun, 2017). In cases where these transcription factors are expressed above normal endogenous levels, it may create a situation where class II partner factors are limiting, leading to enhanced competition for partner factor and DNA response element binding. To complicate things further, both HIF1α and AHR have been shown to have both pro- and anti-tumourigenic roles. Additionally, SIM2 has been shown to be both up- and down-regulated in various cancers, implying that SIM2 can also have both pro- and anti-tumourigenic functions (Bersten et al., 2013). Therefore, the exact interplay between the crosstalk and regulation of these factors in a cancer context is likely to be highly complex and context specific. The implications of this for breast cancer progression is explored in Chapter 5, looking at potential crosstalk between SIM2, AHR and HIF1α in a breast cancer context.

1.2 Single-Minded 2

1.2.1   Transcriptional properties and target gene regulation

The Single-Minded (*sim*) gene was first identified in Drosophila, where it is known to regulate midline development of the central nervous system (CNS) (Crews, 1998; Nambu, Franks, Hu, & Crews, 1990; Nambu, Lewis, Wharton, & Crews, 1991). There have been two mammalian homologs identified, *SIM1* and *SIM2* that are encoded by two separate genes. The amino terminal halves of SIM1 and SIM2 show a high degree of homology through their bHLH and PAS regions, with 87% and 86% protein sequence identity for the human and mouse proteins, respectively. Their C-terminal halves, however, are highly divergent showing no significant homology (Chrast et al., 1997; Dahmane et al., 1995; Ema, Morita, et al., 1996; Ema, Suzuki, et al., 1996; Fan et al., 1996; J. Michaud & Fan, 1997; Moffett, Dayo, Reece, McCormick, & Pelletier, 1996; Yamaki et al., 1996). *SIM1* has been found to be essential for the development of

neuroendocrine lineages within the hypothalamus and plays a role in controlling appetite (J. L. Holder, Jr., Butte, & Zinn, 2000; J. L. Holder et al., 2004; Kublaoui, Gemelli, Tolson, Wang, & Zinn, 2008; Kublaoui, Holder, Gemelli, & Zinn, 2006; Kublaoui, Holder, Tolson, Gemelli, & Zinn, 2006; J. L. Michaud et al., 2001; Tolson et al., 2010).

The mouse *Sim2* gene has been mapped to the distal end of chromosome 16 and the human homologue, *SIM2*, to chromosome 21q22.2 (Chrast et al., 1997; Ema, Suzuki, et al., 1996; Fan et al., 1996). This lies within the Down syndrome Critical Region (DSCR), implicating *SIM2* in the pathogenesis of Down syndrome (H. Chen et al., 1995; Rachidi & Lopes, 2007). There are two isoforms of *SIM2*, SIM2-long (*SIM2-l*) and SIM2-short (*SIM2-s*), which arise due to alternative splicing. The long isoform transcript includes all 11 coding exons. The short isoform lacks exon 11 and contains a read-through from exon 10 into the following intron, creating a *SIM2-s* specific protein coding region (Chrast et al., 1997; Metz, Kwak, Gustafson, Laffin, & Porter, 2006). The functional differences between the two isoforms are currently unknown. SIM2 has been found to contain two transrepression domains, Pro/Ser and Pro/Ala rich regions, which are located within the C-terminal half of the protein. The short isoform lacks the Pro/Ala rich region (Moffett, Reece, & Pelletier, 1997). Both isoforms contain a Nuclear Localisation Signal (NLS) and have been shown to be predominantly localised to the nucleus within cells (Figure 1.1) (A. E. Sullivan, Peet, & Whitelaw, 2016; Adrienne E. Sullivan et al., 2014; Yamaki, Kudoh, Shimizu, & Shimizu, 2004).

SIM2 can form a functional heterodimer with both ARNT and ARNT2 that binds to the Central Midline Element (CME) to regulate gene expression (Ema, Morita, et al., 1996; Ema, Suzuki, et al., 1996; Moffett & Pelletier, 2000; Adrienne E. Sullivan et al., 2014). Both isoforms of SIM2 have been shown to function as either repressors or activators of gene expression in cell-based reporter assays and on verified SIM2 target genes (Figure 1.2). This activity is highly context dependent, for example, ectopic SIM2 expression increases *Myomesin2* (*MYOM2*) transcript levels in Human Embryonic Kidney 293a cells, suggesting SIM2 is acting as a transcriptional activator in this cell type. Whereas knockdown of SIM2 in an immortalised human myoblast cell line (LHCN-M2) cause an increase in *MYOM2* expression, suggesting SIM2 represses *MYOM2* in this context (Ema, Morita, et al., 1996; Ema, Suzuki, et al., 1996; Metz et al.,

2006; Moffett et al., 1997; Probst, Fan, TessierLavigne, & Hankinson, 1997; S. Woods, Farrall, Procko, & Whitelaw, 2008; S. L. Woods & Whitelaw, 2002). There are multiple mechanisms by which Sim2 can act as a transcriptional repressor. These include direct repression by binding to the promoter region of genes and utilising the C-terminal transrepression domain of SIM2, competition with other class I bHLH/PAS factors HIF1α and AHR for binding to ARNT and by competing with HIF dimers for binding directly to hypoxia response elements (HREs) (Figure 1.2) (Emily L. Button et al., 2017; Farrall & Whitelaw, 2009; S. L. Woods & Whitelaw, 2002). The action of SIM2 as a transcriptional activator is thought to be mediated through the transactivation domain found in the C-terminus of ARNT (Metz et al., 2006; S. Woods et al., 2008). Currently there are very few target genes of SIM2 known and these include *BNIP3*, the muscle related genes *MyoD* (Myogenic Differentiation 1) and *MYOM2*, and the antimicrobial genes *Cramp* (cathelicidin-related antimicrobial peptide) and *Tcf7l2* (transcription factor 7-like 2) (K.-J. Chen et al., 2014; Farrall & Whitelaw, 2009; Havis et al., 2012; S. Woods et al., 2008). Discovery of more genuine SIM2 target genes will help to further elucidate the function of this poorly understood transcription factor.
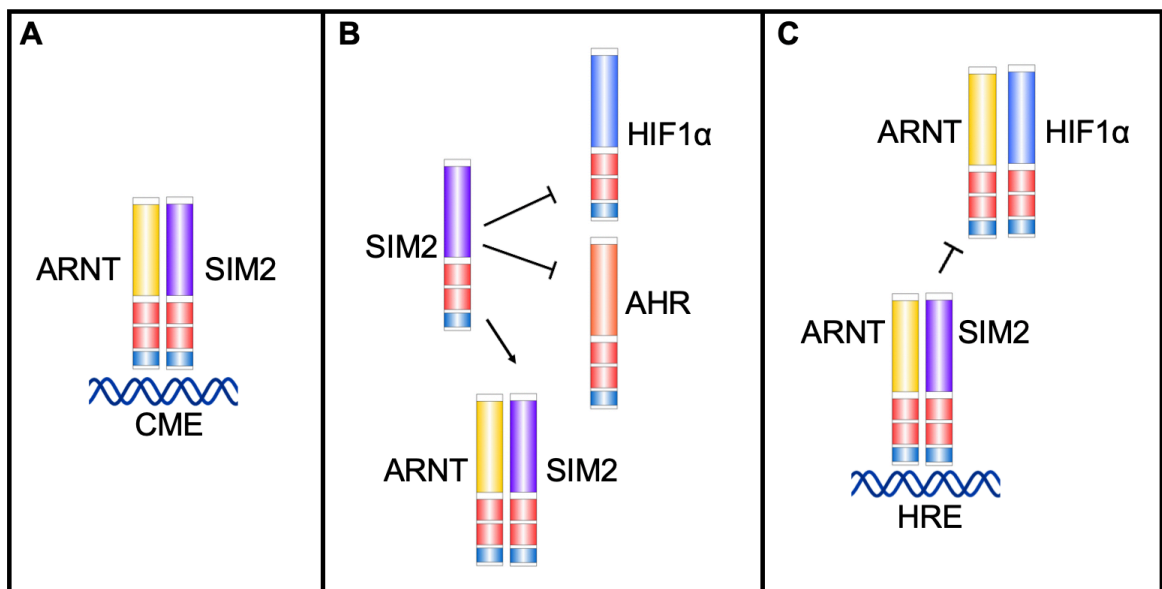
Figure 1.2: SIM2 mechanisms of transcriptional regulation. SIM2 can act as a transcriptional activator through dimerising with ARNT (or ARNT2) and binding to the CME DNA response element (**A**). SIM2 can also act as a transcriptional repressor through multiple mechanisms including actively repressing by binding at CME response elements (**A**), competing with other class I factors for dimerisation with

ARNT (**B**), or through binding to HRE response elements, blocking HIF dimers from accessing and upregulating gene expression (**C**).

1.2.2   Expression of *Sim2*

Current expression data is limited to mRNA expression due to lack of antibodies that will specifically detect endogenous SIM2. In the developing mouse embryo, *Sim2* expression is first seen at embryonic day 8.0 (E8.0) in a small band of cells at the midbrain to forebrain boundary. Throughout development *Sim2* expression is extended to other regions within the midbrain and forebrain including the paraventricular nucleus (PVN) and anterior periventricular (aPV) nuclei within the hypothalamus, cortex, olfactory bulb and the mammillary body. Outside the CNS *Sim2* is expressed in the cartilage and bone of the ribs, vertebrae, craniofacial structures, limbs and digits, skeletal muscle and kidney tubules (Coumailleau & Duprez, 2009; Ema, Morita, et al., 1996; Fan et al., 1996; Marion, Yang, Caqueret, Boucher, & Michaud, 2005). In the adult mouse, *Sim2* expression can be seen in most regions within the brain, intestines and mammary glands, and has also been found in skeletal muscle tissue and the kidneys, with varying levels observed between the short and long isoforms (Figure 1.3) ("Allen Developing Mouse Brain Atlas," ; K.-J. Chen et al., 2014; Laffin et al., 2008; Metz et al., 2006). A more detailed understanding of SIM2 protein expression both during development and postnatally will aid in the discovery of SIM2 target genes and elucidation of SIM2 function.
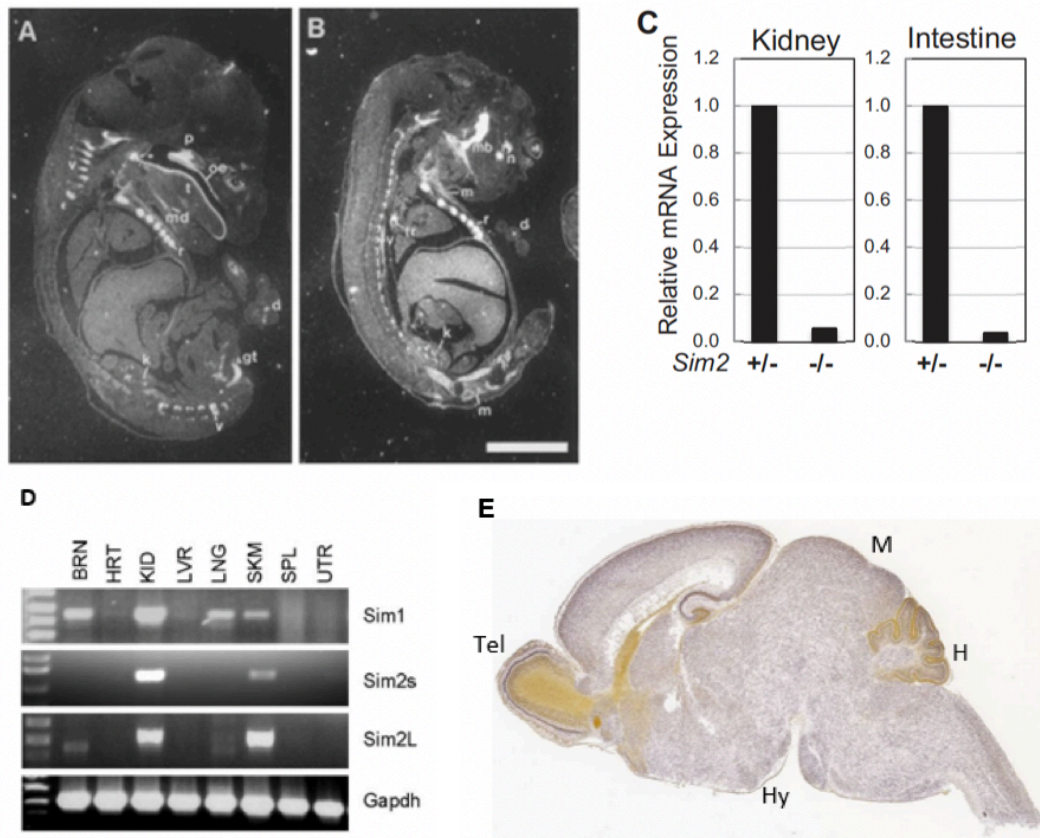
Figure 1.3: Expression of *Sim2.* RNA expression analysis of *Sim2* at E16.5 in a midsagittal **A)** and parasagittal **B)** E16.5 mouse section. *Sim2* expression was detected in vertebrae (v), ribs (r), kidney (k), oral epithelium (oe), mandible (md), mandibular bone (mb), palate (p), tongue (t), nasal pit (n), trachea (tr), muscles (m), and the digits (d) of the limb. Images reprinted from Molecular and Cellular Neuroscience, 7(1), Fan CM, Kuwana E, Bulfone A, Fletcher CF, Copeland NG, Jenkins NA, Crews S, Martinez S, Puelles L, Rubenstein JL, Tessier-Lavigne M, Expression patterns of two murine homologs of Drosophila single-minded suggest possible roles in embryonic patterning and in the pathogenesis of Down syndrome, pg 1-16, Copyright (1996), with permission from Elsevier. **C)** Quantitative PCR analysis showing expression of *Sim2* in the kidney and intestine of heterozygous *Sim2* animals (-/+), but not in homozygous KO animals (-/-). Image used with permission of American Physiological Society, from SIM2 maintains innate host defense of the small intestine. American Journal of Physiology-Gastrointestinal and Liver Physiology, Chen, K.-J., Lizaso, A., & Lee, Y.-H, 307(11), 2014; permission conveyed through Copyright Clearance Center, Inc.. **D)** RT-PCR analysis of the expression of *Sim1*, *Sim2-s* and *Sim2-l* in mouse tissues. *Sim1* is expressed in the brain, kidney, lung and skeletal muscle. *Sim2-s* is expressed at

relatively high levels in the kidney and to a lesser extent in skeletal muscle. *Sim2-l* shows relatively high expression in both the kidney and skeletal muscle, with lower levels of expression in the brain. Image reproduced under Attribution 4.0 International (CC BY 4.0) Creative Commons license, from Metz et al., 2006, doi: 10.1074/jbc.M508858200. **E)** *Sim2* RNA *in situ* hybridisation data from the Allen Developing Mouse Brain Atlas https://developingmouse.brain-map.org/, (Thompson et al., 2014). Postnatal day 4 (P4) mouse brain showing expression of *Sim2* in most brain regions as depicted by purple staining. Tel; telencephalic vesicle, Hy; hypothalamus, M; midbrain, H; hindbrain. Allen Developing Mouse Brain Atlas, https://developingmouse.brain-map.org/experiment/show/100092990.

1.2.3   Current mouse models of *Sim2* and roles during development

There are now a number of studies that have analysed the phenotype of *Sim2* knockout (KO) mice. Interestingly, these studies report varying phenotypes, presumably due to the differences in genetic background of the animals and the analysis that was performed by each of the groups. In general, these studies suggest that *Sim2* is an important factor for normal development and survival with homozygous *Sim2* KO typically presenting as lethal after birth (K.-J. Chen et al., 2014; Goshu et al., 2002; Shamblott et al., 2002) (Whitelaw Laboratory unpublished data).

Originally there were two papers published reporting phenotypes of *Sim2* knockout (KO) mice. Although these two studies analysed the same mouse line, the phenotypes of the *Sim2* KO mice are not totally in agreement. Both studies showed that *Sim2* is essential for normal growth and survival, with homozygous KO pups dying within a few days of birth due to breathing difficulties (Goshu et al., 2002; Shamblott et al., 2002). Aside from this common observation, the other reported phenotypes differ. Shamblott et al. found that the KO mice had craniofacial abnormalities, with either a fully or partially cleft palate, with these malformations likely resulting from premature accumulation of Hyaluronan (an extracellular matrix component) in the palate. These mice have accumulation of air in their gastrointestinal (GI) tract, which they attribute to the breathing difficulties caused by the craniofacial abnormalities (Figure 1.4 B). Conversely, Goshu et al. did not find any craniofacial abnormalities, but found other

skeletal defects including incompletely penetrant scoliosis and small protrusions on the ribs and vertebrae that formed aberrant connections to the surrounding intercostal muscles (Figure 1.4 A). It was proposed that the breathing difficulties seen in these mice were due to the compromised structures surrounding the pleural cavity, which ultimately lead to tearing of the pleural mesothelium. This KO mouse model has also been used to show that Sim2 is required for the development of a full set of Somatostatin (Ss) and Thyrotropin releasing hormone (Trh) expressing neurons within the hypothalamus (Goshu et al., 2004). *Sim2* may also play a role in the development of neurons of the mammillary body, however it appears as though *Sim1* can compensate for loss of *Sim2* expression (Marion et al., 2005). In combination, these studies infer a role for *Sim2* in skeletal and hypothalamic development, however the molecular mechanisms behind these phenotypes are still not fully understood, given the lack of known *Sim2* target genes.

More recently, a study was published that analysed the phenotype of an independent *Sim2* KO mouse line. A conditional *Sim2* KO allele was generated with exon 1 of Sim2 floxed to allow for removal by Cre recombinase. They found that global *Sim2* KO is postnatal lethal in 50% of cases and all KO mice have gas accumulation in their GI tract, to a varying extent (Figure 1.4 C). Further analysis identified that this was due to increased microbial growth in the GI tract of the KO mice. It was confirmed that this effect was due to loss of *Sim2* within intestinal cells through an intestinal-specific *Sim2* KO. It was found that *Sim2* directly or indirectly regulates expression of key antimicrobial peptides including *Cryptdin 1, Cryptdin 2, Cryptdin 6, MMP7* and *TCF7L2*, identifying a role for *Sim2* in innate immunity in the intestine (K.-J. Chen et al., 2014).
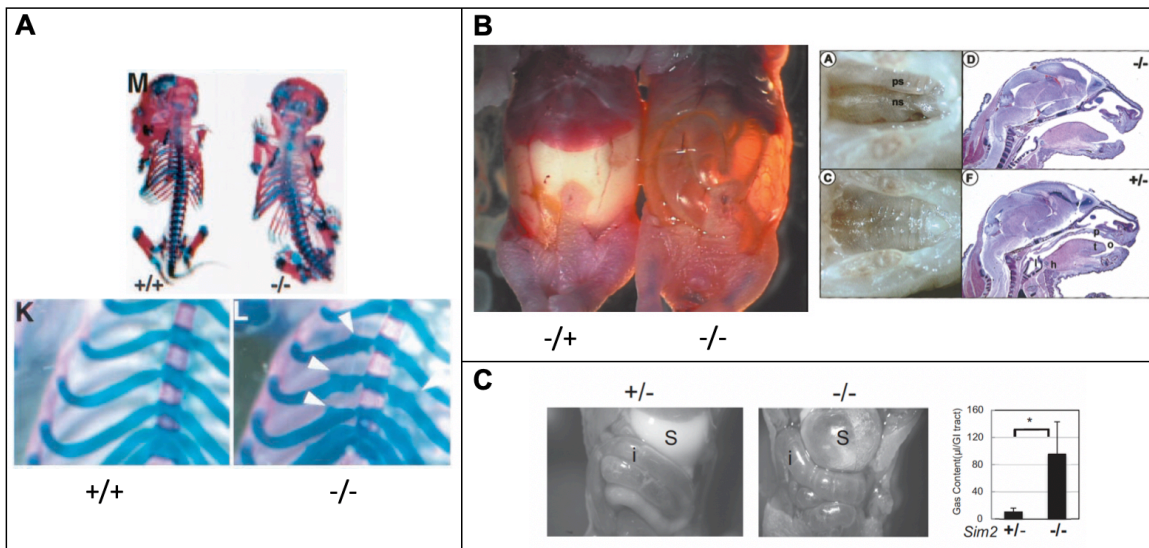
Figure 1.4: Phenotypes of the current *Sim2* knockout mouse models. **A)** Goshu et al (2002) found *Sim2* null mice developed scoliosis and rib protrusions not found in WT littermates. Image from; Sim2 mutants have developmental defects not overlapping with those of Sim1 mutants, Goshu, E., Jin, H., Fasnacht, R., Sepenski, M., Michaud, J. L., & Fan, C. M., Molecular and Cellular Biology, Copyright © 2002, American Society for Microbiology, reprinted by permission of Informa UK Limited, trading as Taylor & Taylor & Francis Group, http://www.tandfonline.com. **B)** Shamblott et al (2002) found gas accumulation in the gastrointestinal tract as well as fully or partially cleft palates in the *Sim2* KO mice but not their littermate controls. Image used with permission of John Wiley & Sons - Books, from Craniofacial abnormalities resulting from targeted disruption of the murine Sim2 gene, Shamblott, M.J., Bugg, E.M., Lawler, A.M., Gearhart, J.D., Developmental Dynamics, 224(4), 2002; permission conveyed through Copyright Clearance Center, Inc. **C)** Chen et al (2014) also saw gas accumulation in the GI tract of *Sim2* null mice, going on to show that this was due to increased microbial growth. Image used with permission of American Physiological Society, from SIM2 maintains innate host defense of the small intestine. American Journal of Physiology-Gastrointestinal and Liver Physiology, Chen, K.-J., Lizaso, A., & Lee, Y.-H, 307(11), 2014; permission conveyed through Copyright Clearance Center, Inc.

Transgenic mouse models overexpressing *Sim2* have been analysed in attempt to explore the possible function of *Sim2* in Down syndrome. These mice were found to have neuronal associated phenotypes of reduced sensitivity to pain, mild learning and memory impairment, and reduced anxiety related exploratory behaviour (Chrast et al.,

2000; Ema et al., 1999). Again, with the current lack of target gene information, these phenotypes and the potential role that *SIM2* may be playing in Down Syndrome is still not well understood.

Given the discrepancies between the established *Sim2* mouse models and the lack of understanding behind some of the reported phenotypes, there is a need for independent and tissue-specific knockout mouse models to further investigate the function of *Sim2*. In addition, mouse models that will allow us to further characterise the molecular function of *Sim2* through detailed expression profiling and target gene discovery will be essential tools for this process.

1.2.4 SIM2 and cancer

Aberrant regulation of SIM2 has been linked to the progression of a number of human cancers. Both isoforms of SIM2 have been shown to be commonly upregulated in prostate, pancreatic and colon tumours, with this upregulation appearing to promote tumour progression. In prostate cancer, SIM2 upregulation correlates with poorer patient prognosis and has been suggested as a potential biomarker and therapeutic target (Aleman et al., 2005; Arredouani et al., 2009; DeYoung et al., 2003a, 2003b; Ole Johan Halvorsen et al., 2007; Lu et al., 2011).

Conversely to this, in breast and esophageal cancers the short and long isoforms respectively have been shown to be downregulated, with data to suggest that this downregulation promotes tumour progression (Gustafson et al., 2009; Kwak et al., 2007; Laffin et al., 2008; S. J. Pearson et al., 2019; Scott J. Pearson et al., 2019; Scribner, Behbod, & Porter, 2013; Su et al., 2016; Tamaoki et al., 2018).

Overall, this highlights the action of SIM2 in cancer is highly context/type specific and further research is required to understand the molecular mechanisms relevant to distinct cancers. Presumably tissue-specific gene regulation by SIM2 as well as variable SIM2:cofactor interactions in different cell types determine overall pro- or anti-tumourigenic outcomes. The role SIM2 plays in cancer is the focus of Chapter 5 of this thesis.

1.3 Genome Editing

Traditional methods for generating mouse models are relatively inefficient and can take long periods of time to generate and establish founders. This typically involved modifying endogenous loci in mouse embryonic stem cells in culture through homologous recombination of a large plasmid targeting construct. Targeting constructs contained large homology arms flanking the genetic variation to be made along with selection markers. These constructs were electroporated into the ES cells, which were then cultured in selective media to enrich for those that were successfully electroporated and potentially contained the genetic modification of interest. ES cells would need to be screened in order to find one that contained the correct genetic modification. These cells would then be injected into a blastocyst and transferred to pseudo-pregnant female with the resulting pups being chimeric, derived from both the modified ES cells and those that were from the injected blastocyst. Pups with germline incorporation of the genetically modified ES cells would be able to pass the mutation on and from this next generation a heterozygote colony could be established (Hall, Limaye, & Kulkarni, 2009; Patrick D. Hsu, Lander, & Zhang, 2014).

The discovery and development of CRISPR/Cas9 technology has led to significant advances in the genome editing field. Clustered regularly interspaced short palindromic repeat DNA sequences (CRISPR) and the CRISPR-associated genes (Cas9) were first identified in bacteria where they have been shown to act as an immune system by providing defence against bacteriophage infection (Adli, 2018; Doudna & Charpentier, 2014; Patrick D. Hsu et al., 2014). Researchers have since adapted this system to allow for precise genome editing in mammalian cells.

The CRISPR/Cas9 system works through the RNA-guided Cas9 DNA endonuclease generating a double stranded break at a precise genomic location. The endogenous bacterial system uses two RNAs; a CRISPR RNA (crRNA) which acts as a guide, specifying the DNA target site through specific RNA:DNA base-pairing, and the trans-activating crRNA (tracrRNA) which binds the Cas9 enzyme. This has been engineered into the single guide RNA (sgRNA) which performs both of these functions for use in genome editing. In the case of the most commonly used Cas9 enzyme, SpCas9

(*Streptococcus pyogens* Cas9), the guide sequence that base-pairs with the target DNA is 20bp in length and requires an adjacent PAM (protospacer adjacent motif) NGG sequence that the Cas9 will recognise and bind to in the target DNA sequence (Figure 1.5) (Adli, 2018; Doudna & Charpentier, 2014). Plasmids that contain simple cloning sites for the sgRNA as well as the Cas9 gene have been generated for use in mammalian cell culture systems and generating sgRNA and Cas9 mRNA for zygote injections of these reagents (Cong et al., 2013).
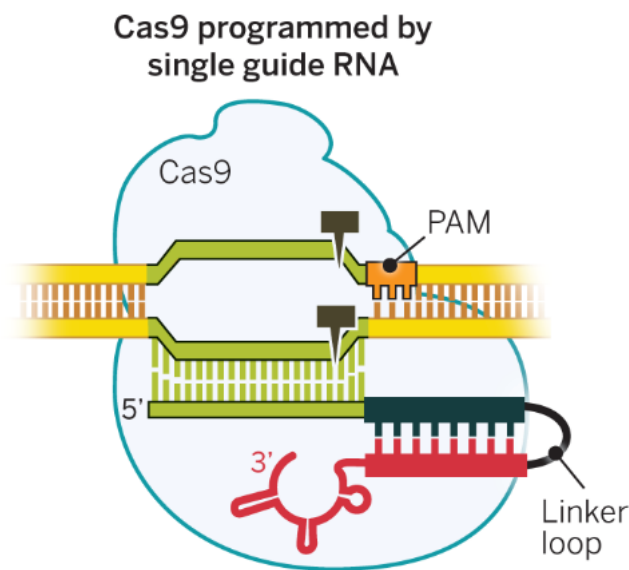
Figure 1.5: Structure of Cas9 endonuclease, sgRNA and target DNA complex. sgRNA directs the Cas9 endonuclease to a specific DNA target site through complementary RNA:DNA base pairing. The PAM sequence allows for Cas9 DNA binding with a double stranded break generated 3 nucleotides upstream of the PAM site. Image from Doudna, J. A., & Charpentier, E. (2014). The new frontier of genome engineering with CRISPR-Cas9. Science, 346(6213). doi: 10.1126/science.1258096. Reprinted with permission from AAAS.

Given the customisability of the guide RNA sequence, almost anywhere in the mammalian genome can be targeted. Once a double stranded break in the genome has been made, there are two main mechanisms the cell can use for repair; non-homologous end joining (NHEJ), and homology directed repair (HDR) (Figure 1.6). NHEJ is an error prone process that typically results in insertions or deletions at the CRISPR cut site (Patrick D. Hsu et al., 2014). This has been shown to be a very efficient

way of creating mutations at a target site, with reports of close to 100% cells/embryos containing some form of genetic lesion at the CRISPR cut site (Bell, Magor, Gillinder, & Perkins, 2014; X. Wang et al., 2013). Generation of double stranded breaks has been shown to increase the efficiency of homologous recombination events (Patrick D. Hsu et al., 2014). Therefore, through supplying a cell or embryo with a template for repair, such as a single stranded DNA donor oligonucleotide or double stranded donor plasmid, HDR can occur and be harnessed to make specific genetic modifications. This method has been shown to be relatively successful for generating genomic modifications such as knock-ins of epitope tags, fluorescent reporters and floxed alleles, as well as correcting disease mutations (Figure 1.6) (Doudna & Charpentier, 2014; Patrick D. Hsu et al., 2014; Y. Wu et al., 2013; Hui Yang et al., 2013; H. Yang, Wang, & Jaenisch, 2014).
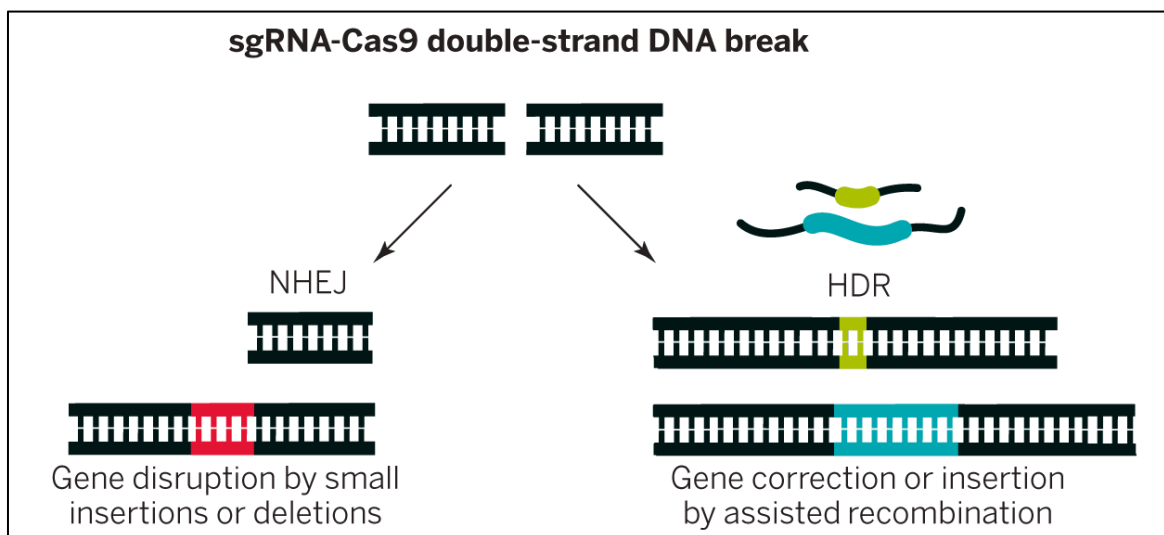
Figure 1.6: Repair pathways employed to generate targeted genetic modifications. Non-homologous end joining (NHEJ) results in insertions or deletions at the Cas9 cut site. Homology directed repair (HDR) allows for specific DNA modifications through the addition of a donor template for repair. Image from Doudna, J. A., & Charpentier, E. (2014). The new frontier of genome engineering with CRISPR-Cas9. Science, 346(6213). doi: 10.1126/science.1258096. Reprinted with permission from AAAS.

This technology has been used successfully in cell culture and in many organisms including drosophila, mice and even mammals as complex as monkeys (Bassett, Tibbit, Ponting, & Liu, 2013; Cho, Kim, Kim, & Kim, 2013; Cong et al., 2013; P. D. Hsu et al.,

2013; Niu et al., 2014; H. Y. Wang et al., 2013; Hui Yang et al., 2013). The use of this technology has both reduced the time taken and increased the efficiency of the generation of genetically modified mouse models, with the possibility of generating mouse models in as little as a few weeks. The process of generating genetically modified mouse models is achieved through injecting the CRISPR components (sgRNA, Cas9 mRNA or protein and an optional HDR template) into a mouse embryo that is then transferred to a pseudo pregnant female (Patrick D. Hsu et al., 2014; H. Yang et al., 2014). Injecting at the one cell stage allows for the genetic modification to be present in every cell in the mouse.

One disadvantage of this system is the high frequency of off-target mutations being generated at genomic locations other than the target sequence (Cho et al., 2014; Fu et al., 2013; Pattanayak et al., 2013; Zhang, Tee, Wang, Huang, & Yang, 2015). This is an outcome that needs to be considered and factored into experimental design. In the case of generating mouse models, there is the possibility of breeding out any off-target mutations through successive rounds of breeding and selecting for mice that only contain the genetic modification of interest. This can also be addressed in both animal models and cell culture through the use of two independent guides that target different regions of the same gene to ensure any phenotypes observed are not the result of off-target mutations.

1.4 Project Aims

The overall aim of this project is to investigate the molecular function and mechanisms of action of SIM2 in both a normal developmental and disease context. This aim was approached using both cell culture and mouse models. There are three main aims;

1. Functionally characterise SIM2 variants found in human patients with intellectual disabilities.
2. Investigate the function of SIM2 in breast cancer, focusing on gene expression analysis.
3. Generate, validate and analyse mouse models of SIM2 with the following subaims;
    i. Phenotypically analyse a conditional *Sim2* knockout mouse model.
    ii. Generate, validate, and characterise an epitope tagged SIM2 mouse model.
    iii. Generate, validate, and characterise a *Sim2* fluorescent reporter mouse model.

This work should contribute key knowledge in the understanding of how SIM2 functions as well as establish valuable tools in order to further investigate the molecular function of this transcription factor.

Chapter 2: HIF has Biff – Crosstalk between HIF1α and the family of bHLH/PAS proteins

2.1 Summary

Hypoxia Inducible Factor 1α (HIF1α) is a member of the bHLH/PAS family of transcription factors and is well studied due to its role in controlling the cellular response to hypoxia. While a lot is known about HIFs function, there is still much to discover. This chapter is a literature review focusing on what is known regarding the interactions and crosstalk between HIF1α and other members of the bHLH/PAS family of transcription factors. This includes crosstalk with SIM2 in a cancer context, AHR in immunomodulation, BMAL1 in the regulation of the circadian rhythm, and NPAS1 and NPAS3 in neuron development. This review has been published in a peer reviewed journal.

2.2 Publication 1: HIF has Biff – Crosstalk between HIF1α and the family of bHLH/PAS proteins: Article

Button, E. L., Bersten, D. C., & Whitelaw, M. L. (2017). "HIF has Biff - Crosstalk between HIF1a and the family of bHLH/PAS proteins". *Exp Cell Res, 356*(2), 141-145. doi: 10.1016/j.yexcr.2017.03.055

Statement of Authorship

Title of Paper	HIF has Biff - Crosstalk between HIF1a and the family of bHLH/PAS proteins
Publication Status	☒ Published ☐ Accepted for Publication ☐ Submitted for Publication ☐ Unpublished and Unsubmitted work written in manuscript style
Publication Details	Button, E. L., Bersten, D. C., & Whitelaw, M. L. (2017). HIF has Biff - Crosstalk between HIF1a and the family of bHLH/PAS proteins. Exp Cell Res, 356(2), 141-145. doi: 10.1016/j.yexcr.2017.03.055

Principal Author

Name of Principal Author (Candidate)	Emily Button		
Contribution to the Paper	Preparation of the manuscript unless otherwise stated.		
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	01/11/22

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	David Bersten		
Contribution to the Paper	Prepared section "HIF and the NPAS transcription factors" of the manuscript and provided feedback.		
Signature		Date	01/11/22

Name of Co-Author	Murray Whitelaw		
Contribution to the Paper	Prepared sections "abstract", and "conclusions and future directions" and provided feedback on the manuscript.		
Signature		Date	31-07-22

Please cut and paste additional co-author panels here as required.

34

HIF has Biff – Crosstalk between HIF1a and the family of bHLH/PAS proteins

Emily L. Button, David C. Bersten, Murray L. Whitelaw*

Dept Molecular and Cellular Biology, University of Adelaide, 5005 South Australia, Australia

A B S T R A C T

Two decades of research into functions of the ubiquitous transcription factor HIF have revealed pervasive roles in development, oxygen homeostasis, metabolism, cancer and responses to ischemia. Unsurprisingly, HIF activities impinge on many pathologies, for which underlying molecular mechanisms are actively sought. HIF is a member of the heterodimeric bHLH/PAS family of transcription factors, a set of proteins that commonly function in developmental pathways and adaptive responses to environmental or physiological stress. Similarities in the mechanisms that regulate gene targeting by these transcription factors create opportunities for extensive crosstalk between family members. Data supporting pathway interactions between HIF1a and other bHLH/PAS factors, both collaborative and antagonistic, is beginning to surface in the areas of cancer, circadian rhythm, and immune responses. This review summarises the status of HIF1a-bHLH/PAS protein crosstalk and is dedicated to the memory of Lorenz Poellinger, a pioneer investigator into the molecular mechanisms of HIF, AHR, and ARNT bHLH/PAS factors.

1. Introduction

The Hypoxia-Inducible Factor 1α (HIF1α) is an oxygen regulated member of the basic Helix-Loop-Helix/PER-ARNT-SIM (bHLH/PAS) family of transcription factors that plays key roles in embryonic and adult oxygen homeostasis. HIF1α is essential for development of embryonic vasculature and thereafter maintains roles in angiogenesis throughout life, particularly during times of oxygen deficiency, when HIF1α activity co-ordinately adjusts cellular metabolism and erythropoiesis to aid adaption and survival. To achieve these functions, HIF1α dimerises with a second bHLH/PAS protein, ARNT (also known as HIF1β), to regulate hundreds of genes across a range of tissues. HIF1α represents a Class I bHLH/PAS factor, proteins characterized by being either ubiquitously expressed and signal regulated, or alternatively, constitutively active and tissue specific. Class I factors bind DNA upon dimerision with a Class II partner factor, one of ARNT (Aryl Hydrocarbon Receptor Nuclear Translocator), ARNT2 or ARNT-Like (ARNTL, aka BMAL). Other well studied Class I factors include the Aryl hydrocarbon Receptor (AhR, aka Dioxin Receptor), which is critical for toxin metabolism and differentiation of distinct subsets of T and B immune cells [1]; Single Minded 1 (SIM1), essential for specification of neuroendocrine secreting cells of the hypothalamus and control of appetite [2,3]; Neural PAS factors (NPAS 1, 3 & 4), which function in neuron development and synaptic plasticity [4]; and CLOCK and NPAS2, which control circadian rhythms. Members of this family all share a conserved domain organisation. The amino-terminal bHLH domain is involved in DNA binding and serves as a primary dimerization interface. This is followed by a PAS domain consisting of two PAS repeats, PAS A and PAS B, that are important for secondary dimerization and in some cases, binding small organic ligands. The carboxy-terminal halves contain either transactivation or transrepression domains for target gene regulation. Cognate DNA sequences bound by the heterodimers are variations of the canonical E-box sequence, all containing a core NNCGTG [5,6].

Class I bHLH/PAS factors have structural similarities and commonly heterodimerise with ARNT to bind similar DNA response elements, features that invite direct crosstalk between the proteins. Additionally, an increasing number of studies are showing that there are intersections between the pathways bHLH/PAS factors control in biological functions such as immune cell regulation, the circadian rhythm and cancer progression. This review focuses on the various bHLH/PAS family members and how their functions and pathways are involved in crosstalk with the activities and functions of HIF.

2. HIF1α and SIM2 crosstalk in cancer

Single-Minded 2 (SIM2) is a Class I bHLH/PAS factor that has been identified as misregulated in a number of cancers, particularly prostate,

---

* Corresponding author.
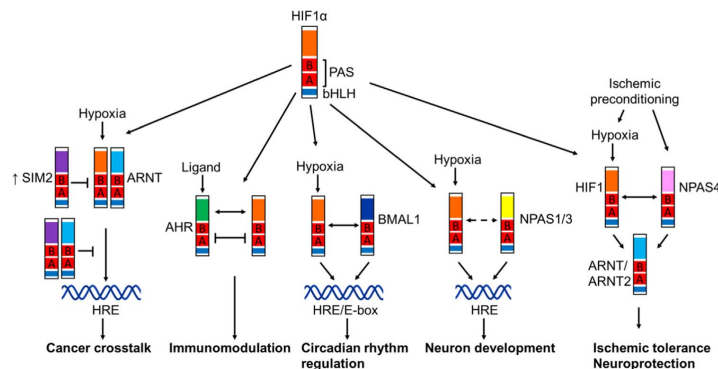*E-mail address:* murray.whitelaw@adelaide.edu.au (M.L. Whitelaw).

Fig. 1. Crosstalk of bHLH/PAS transcription factors with HIF1α. Other members of the bHLH/PAS transcription factor family are able to crosstalk with HIF1α in a number of ways. SIM2 competes with HIF1α both for binding to the partner factor ARNT and for binding HREs in HIF1α target genes. The AHR and HIF are able to compete for binding to ARNT and their pathways are both important for immunomodulation and can either promote the same or opposing outcomes. HIF1α crosstalk with the circadian clock occurs through both synergistic action of HIF1α/ARNT and CLOCK/BMAL dimers in regulating clock gene expression and through dimerising with the circadian factor BMAL to control gene expression. It is predicted that ischemic conditions will negatively influence NPAS activity.

where it has been suggested as a potential biomarker due to correlation of high levels with poor patient survival [7–9]. HIF1α can play multiple positive and negative roles in solid cancers. It primarily aids tumourigenesis and progression via adaption to hypoxia, but conversely, can also be anti-tumourigenic by activating pro-apoptotic genes. *In vitro* studies have shown that SIM2 is able to interfere with HIF1α activity in two important ways. Like HIF1α, SIM2 heterodimerises with ARNT, with seemingly similar affinity [10], setting a scenario for competition between HIF1α and SIM2 for binding to their common partner factor (see Fig. 1). Secondly, SIM2/ARNT heterodimers are able to compete with HIF1α/ARNT dimers for binding to hypoxic response elements (HREs). SIM2 can dampen HIF1α-mediated target gene activation through sequestering ARNT, with or without DNA binding. SIM2 can also actively repress gene transcription by utilising the transrepression domain within its C-terminus [11–14]. The latter appears to be important in a prostate cancer context, where SIM2 interferes with HIF1α mediated upregulation of BNIP3 (Bcl-2/adenovirus E1B 19 kDa interacting protein 3), a pro-apoptotic Bcl-2 family member [13]. We also found that SIM2 is able to repress HIF1α-mediated activation of NDRG1 (N-Myc Downstream Regulated 1, Sullivan et al., unpublished data), an established HIF1α target gene that suppresses metastasis in a number of cancer types [15,16]. NDRG1 is upregulated under hypoxia but downregulated when SIM2 is overexpressed in a number prostate cancer cell lines. This SIM2-mediated antagonism of the anti-tumorigenic roles of HIF1α in cancer cells may contribute significantly to the progression of late stage cancers, where SIM2 protein expression is most commonly upregulated (Sullivan and Whitelaw, unpublished data).

Natural functions of SIM2 have been little explored beyond proposed roles in the development of, and hormone expression in, certain types of neuroendocrine cells within the hypothalamus [17]. Whether SIM2 crosstalk with HIF1α has biological relevance in normal development or homeostasis remains to be determined.

3. HIF1α crosstalk with the circadian clock

BMAL1 (Brain and Muscle ARNT-Like, or simply ARNTL) is a Class II bHLH/PAS transcription factor that, together with its Class I partner factor CLOCK, forms one of the essential core components of the circadian clock that controls central and peripheral circadian rhythms. Other core proteins include the circadian repressors, PAS domain-containing PERIOD homologues 1, 2 and 3 (PER1 PER2, PER3), and

CRYPTOCHROME 1 and 2 (CRY1 and CRY2) [5]. It has long been known that HIF1α can heterodimerise with BMAL [18], however the biological relevance and implications for this are only now emerging. Current studies suggest that there is reciprocal regulation between the oxygen sensing and circadian clock pathways, which appears to occur primarily through the regulation of gene expression.

Recent work has reported that in certain cell types and under certain conditions, HIF1α is able to control the expression of some of the key circadian regulatory factors. Daily oxygen levels exhibit circadian like rhythms in the blood and tissue of rodents and mimicking these physiological oxygen rhythms in cultured Hepa-1c1c7 and NIH 3T3 cells synchronized circadian clocks [19]. Cells subjected to oxygen rhythms displayed rhythmic expression of classic clock genes including *Per1*, *Per2*, *Cry1*, *Cry2*, *Rev-erbα*, *Rora* and *Dbp*. In support of this phenomenon being HIF1α dependent, cells lacking HIF1α did not show synchronized rhythmic expression of clock genes [19]. Separate studies reported that both hypoxia mimetic treatments and genetic stabilization of HIF1α cause clock period lengthening and decreased amplitude in C2C12, U2OS cells and suprachiasmatic nucleus (SCN) *ex vivo* cultures. Additionally, chromatin Immunoprecipitation (ChIP) studies found HIF1α binding to the promoter regions of core clock genes including *PER1*, *PER2*, *Cry1*, *NR1D1*, *DEC1* and *DEC2*, and both hypoxia mimetic treatments and genetic stabilization of HIF1α induced expression of a number of core clock regulatory genes [20,21].

Conversely, core components of the circadian clock pathway can regulate HIF1α expression. In mouse brain and kidney tissues, nuclear HIF1α protein displays rhythmic expression [19]. Similarly, in the mouse liver and cardiac tissue, *Hif1α* mRNA levels cycle in a circadian manner [21,22]. Moreover, modulation of circadian activators and repressors effect hypoxic stabilization of HIF1α protein. For example, mouse fibroblasts lacking the circadian repressors *Cry1* and *Cry2* or *Per1* and *Per2* showed an increase in accumulation of HIF1α after hypoxia mimetic treatment. Conversely, knockout of the circadian transcriptional activator *Bmal1* in MEFs and C2C12 derived myotubes resulted in reduced HIF1α accumulation under hypoxic conditions. Consistent with this, *Bmal1* null myotubes revealed reduced cellular respiration and decreased anaerobic glycolysis [20,21]. Furthermore, analysis of ChIP-seq data showed association of clock proteins CLOCK, BMAL1 and CRY1 at an E-box motif within the promoter region of HIF1α [21] indicating that HIF1α is directly controlled at the genomic level by core clock genes. The consequence of circadian rhythm

impingement on the HIF1α dependent hypoxic response in whole animal physiology is remains to be determined.

Interestingly, there appears to be direct crosstalk between the oxygen sensing and circadian rhythm pathways in order to regulate the circadian clock. *In vitro* studies revealed that co-expression of HIF1α/BMAL1 was able to activate a HRE driven reporter and a *Per2*-luciferase reporter to a similar extent as canonical HIF1α/ARNT and CLOCK/BMAL dimers, respectively. The physiological relevance of the HIF1α/BMAL1 dimer, however, still remains to be determined [20]. Intriguingly, ChIP-seq analysis found that HIF1α and BMAL1 co-occupy 20–30% of BMAL1 binding sites, including genes that make up the core clock regulatory network such as *PER1*, *PER2*, *NR1D1*, *DEC1* and *DEC2*. In contrast to this being due to a HIF1α/BMAL1 dimer, it appears as though this crosstalk is a synergistic effect of HIF1α/ARNT and CLOCK/BMAL1 dimers acting through HRE and E-box motifs, respectively, within the same gene to regulate transcription (see Fig. 1) [21].

Taken together, these recent findings show that there is indeed crosstalk between the oxygen sensing and circadian rhythm pathways which is highly complex and may vary between different tissues. The core regulatory units of these two pathways appear to be able to control expression of one another and in the case of HIF1α and BMAL1, depending on the context, potentially work together to regulate gene expression. Clearly there is much more to be discovered in this area, including determination of other interacting factors that may be at play and elucidating how this newly discovered crosstalk between the two pathways might impinge on the phenomena of jet lag or chronic pathological conditions such as cardiovascular disease and stroke.

4. HIF and the NPAS transcription factors

Hypoxia plays multiple roles in nervous system development and function, via both HIF-dependent and independent pathways. Through HIF, hypoxia acts as a morphogen to organize normal developmental programs [23]. This is illustrated by conditional deletion of HIF1α in the CNS leading to hydrocephalus, failure of neuronal differential and increased neural apoptosis [24]. Not surprisingly, aberrant hypoxia can lead to disrupted development and pathological processes [23,25]. For example, genetic predisposing factors can interact with inappropriate gestational hypoxia to significantly increase the penetrance of congenital scoliosis via an FGF mediated pathway [26]. It has been proposed that the penetrance and/or severity of a number of common developmental defects result from combined genetic predisposition and disrupted oxygen supply, with potential roles for dysregulated HIF pathways being actively explored.

As has been found for HIF1α and CLOCK interacting with BMAL to regulate circadian rhythms, cross reactivity between the HIFs and other neuronal bHLH/PAS transcription factors is highly probable. The Neural PAS set of proteins, NPAS1, 2, 3 and 4 are expressed exclusively or predominantly in the brain. NPAS2 dimerises with BMAL to form another circadian rhythm regulating complex, which functions in brain regions outside the hypothalamic (suprachiasmatic nucleus) realm of CLOCK/BMAL. Biochemical experiments show NPAS1, 3 and 4 dimerise with ARNT and ARNT2. While limited data concerning the *in vivo* DNA binding and dimerization characteristics of NPAS1 and NPAS3 exist, it is striking that recent X-ray crystallography reveals structures of NPAS1/ARNT and NPAS3/ARNT to substantially overlap those of HIF1α/ARNT and HIF2α/ARNT. All four complexes bind with high affinity to the consensus hypoxia response element (HRE: TACGTG) [27]. The significant overlap between dimerization partners and response elements sets up a scenario where hypoxic induced HIF during neuronal development could alter NPAS1/3 mediated transcription factor function, affecting neuronal development [28]. Notably, early specification and proliferation of inhibitory neurons is controlled by NPAS1 and NPAS3 and disruption of either factor leads to defects in neuronal proliferation and animal behaviour [29,30].

Mutations in their genes have been linked to schizophrenia and autism spectrum disorder [29–31]. In support of cross coupling between NPAS factors and HIF1α, NPAS1 has been reported to bind the HRE of the EPO gene *in vivo* and repress a corresponding reporter gene in cell culture essays [28].

In addition to developmental cross talk, the HIFs may also interact with bHLH-PAS transcription factors following brain ischemic events or neuronal injury. Ischemic tolerance results from preconditioning by a sub-lethal ischemic event, which protects the tissue from subsequent ischemic challenge [25]. Preconditioning can protect neurons from cell death resultant from stroke, seizure or brain injury [25]. Both chemical and hypoxic induction of the HIFs has been shown to be neuroprotective to subsequent ischemic events [25] and conditional deletion of HIF1α in neurons increases brain injury following brain ischemia in mice [32,33]. NPAS4 is also upregulated following cerebral ischemia [34]. Experiments *in vitro* provide evidence that NPAS4 can contribute to neuroprotection and knockout of NPAS4 leads to progressive neuronal apoptosis and increased neurodegeneration in mice that have undergone cerebral ischemia [35,36]. Thus both HIF1α and NPAS4 are upregulated following cerebral ischemia and contribute to neuroprotection, however potential collaboration in this outcome remains unexplored.

In summary, the contribution of hypoxia to CNS development, ischemic tolerance and a variety of pathologies is well established. Determining the full contribution of HIF to these processes, including collaborative and/or cross interference mechanisms with other neuronal bHLH-PAS transcription factors, remains fertile ground for future research.

5. HIF and AHR function in immunomodulation

The Aryl Hydrocarbon Receptor (AHR) is a ligand dependent Class I bHLH/PAS factor important for regulating development and function of immune cells. In the absence of a ligand, latent AHR is located in the cytoplasm, bound to chaperone proteins. Upon binding of aromatic hydrocarbons, typified by xenobiotics such as dioxins and PCBs (polychlorinated biphenyls), dietary indoles or tryptophan metabolites, the AHR translocates to the nucleus where it dimerises with the Class II partner factor ARNT and activates target genes [1,6,37]. Biochemical and cell culture studies suggest that AHR and HIF activities can reciprocally inhibit each other by competing to bind to the common partner factor ARNT [38–41]. Situations where this crosstalk is physiologically relevant are beginning to emerge, best exemplified in the immune system, where complex interplay between HIF and AHR pathways features during the differentiation of a range of cell types.

An effective immune response is required to provide host defence against microorganisms. Mammals produce a complex network of specialised T cells that cause inflammation, clear the infection and regulate appropriate immune responses. Both AHR and HIF play important roles in the regulation and function of T cells (see Fig. 2 for overview). Naïve CD4+ cells have the ability to develop and differentiate into different types of regulatory (Treg) and helper (Th) T cells. This multifaceted process is controlled by several factors, including a group of cytokines called interleukins (IL), which each help stimulate differentiation of certain CD4+ T cell lineages. In general, helper T cells promote immune responses through releasing cytokines that act on other immune cells, whereas T regulatory cells act to suppress the immune responses of other cells and prevent autoimmunity. Studies have shown that depending on context, the AHR and HIF pathways can result in either the same or opposing outcomes for T cell regulation.

Treg cells inhibit the proliferation and activity of T helper cells. HIF1α is able to suppress the differentiation of Treg cells while conversely, AHR appears to, in some instances, promote Treg cell differentiation over T helper cell differentiation. The Tr1 cell population of regulatory T cells have anti-inflammatory roles, are abundant in
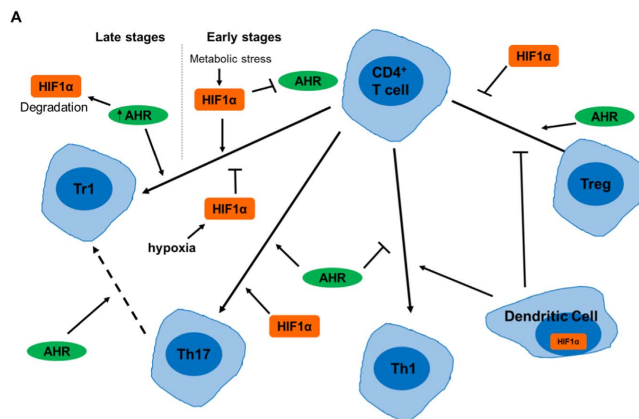
Fig. 2. Complex regulation of immune cells by HIF1α and AHR. Both HIF1α and AHR are able to influence the differentiation and activity of a number of immune T cells. Simplified overview for effects of these two factors is shown. HIF1α and AHR can either positively collaborate or have opposing effects on the differentiation and activity of certain types of T cells. The influence of HIF1α and AHR on differentiation of CD4+ T cells is summarised.

chronic infections and are characterized by their expression of IL-10. Under conditions where HIF1α is metabolically activated, HIF1α controls early stages of Tr1 cell development with AHR controlling the later stages. However, hypoxically activated HIF1α suppresses Tr1 cell differentiation [42]. As for helper T cell lineages, Th17 cells are proinflammatory T cells that differentiate from CD4+ T cells. Both HIF1α and AHR pathways can promote differentiation of the Th17 cell lineage. Th1 cells function to elicit cellular immunity against intracellular pathogens. Dendritic cells expressing HIF1α are able to favour differentiation of Th1 cells, whereas AHR has been shown to block Th1 cell differentiation.

Interestingly, accumulation of HIF1α protein in immune cells can occur in an oxygen-independent manner through a range of mechanisms, including bacterial infection, lipolysaccharide, ROS production by activated T cells and activation of nuclear factor-kappa B (NF-κB) [43]. This indicates that HIF may have important functions in immune cells outside of its role in modulating the response to oxygen levels.

The metabolic differentiation of type 1 regulatory T cells (Tr1) involves direct crosstalk between the HIF1α and AHR pathways [42]. The metabolic profile of T cells under Tr1-polarizing conditions suggests that a metabolic switch to aerobic glycolysis controlled by HIF1α may be necessary for the early stages of Tr1 cell development. Extracellular ATP (eATP) is increased through T cell activation and is able to increase HIF1α protein levels within T cells in Tr1-polarizing conditions. This leads to an increase in the abundance of the HIF1α/ARNT heterodimer with a concomitant reduction in the amount of the AHR/ARNT dimer within these T cells. Moreover, this increase in HIF1α protein causes degradation of AHR in a proteasome dependent manner. In later stages of Tr1 cell induction, AHR levels increase and promote HIF1α degradation, with AHR now controlling Tr1 cell differentiation. Under hypoxic conditions, however, stabilization of HIF1α prevents Tr1 cell differentiation as a result of persistent HIF1α protein and a decrease in the expression and binding of AHR at important target gene promoters [42]. Under conditions favouring Tr1 differentiation, IL-27 is able to induce AHR which then upregulates expression of IL-10 and IL-21, cytokines important for Tr1 cell activation. Additionally, in the presence of TGF-β, Th17 cells are able to *trans*-differentiate into Tr1 cells, with AHR activation favouring this process [37,44–47].

HIF1α is able to control the balance between differentiation of CD4+ T cells into Th17 and Treg cells both under normoxic and hypoxic conditions. HIF1α promotes the differentiation of Th17 cells through transcriptional activation of RORγt, an important Th17 cell transcriptional regulator, which results in production of IL-17, a key Th17 cell cytokine [48]. AHR is also able to promote differentiation of Th17 cells by decreasing expression and activity of IL-2, a known inhibitor of Th17 cell differentiation but stimulator of Treg cell differentiation [37,48–51]. Conversely, HIF1α suppresses Treg cell development through binding FoxP3, a key Treg cell transcriptional regulator, targeting it for degradation by the proteasome [48]. There is also evidence that dendritic cells expressing HIF1α are able to suppress Treg differentiation by favouring Th1 cell differentiation through increased expression of IL-12, a promoter of this process, and decreased TGF-β, a promoter of the alternative Treg cell differentiation [52]. The interplay with AHR is complex as there is evidence that AHR ligands can keep immune responses in check through favouring Treg cell differentiation and function over helper T cell differentiation. AHR can increase *FoxP3* expression by binding to its promoter and inducing transcription. AHR is also able to block differentiation of Th1 cells by inhibiting the phosphorylation of STAT1 [37,53,54].

In summary, there are clear examples of the HIF and AHR pathways leading to the same or opposing outcomes in T cell development and differentiation. Both transcription factors are subject to multiple levels of control, setting a significant challenge to decipher how variations in signalling mechanisms, emanating from intracellular and tissue microenvironment sources, combine with protein expression levels, presence of ligands, oxygen and metabolic status to coordinate immune responses. Whether there are more examples of direct crosstalk between these two pathways, similar to that seen in the differentiation of Tr1 cells, remains to be evaluated..

6. Conclusions and future directions

While the exploration of interplay between HIF1α and the signalling pathways of other bHLH/PAS transcription factors is in its infancy, some startling examples illustrate the potential for discoveries in this area to expand our understanding of how environmental and physiological perturbations are sensed and integrated at the molecular level. It will be interesting to see whether there are more instances of either direct competition or synergistic collaboration between HIF and other bHLH/PAS members in normal development, homeostasis or disease contexts. While this review has focused on HIF1α, as this has been the

most intensively investigated factor thus far, questions immediately arise as to cross coupling possibilities between bHLH/PAS family members and HIF2α and 3α. Further complexities are inherent when considering varied use of the three versions or ARNT, all of which are widely expressed. This research is particularly relevant to human pathology. Since HIF1 and 2α are currently being targeted in clinical trials to combat anemia and cancer, it is important to continue exploring the impact of HIF crosstalk in a range of tissues and cell types.

Acknowledgements

References

[1] N. Hao, M.L. Whitelaw, The emerging roles of AhR in physiology and immunity, Biochem. Pharmacol. 86 (5) (2013) 561–570.

[2] J.L. Michaud, et al., Development of neuroendocrine lineages requires the bHLH-PAS transcription factor SIM1, Genes Dev. 12 (20) (1998) 3264–3275.

[3] J.L. Michaud, et al., Sim1 haploinsufficiency causes hyperphagia, obesity and reduction of the paraventricular nucleus of the hypothalamus, Hum. Mol. Genet. 10 (14) (2001) 1465–1473.

[4] K. Ramamoorthi, et al., Npas4 regulates a transcriptional program in CA3 required for contextual memory formation, Science 334 (6063) (2011) 1669–1675.

[5] D.C. Bersten, et al., bHLH-PAS proteins in cancer, Nat. Rev. Cancer 13 (12) (2013) 827–841.

[6] R.J. Kewley, M.L. Whitelaw, A. Chapman-Smith, The mammalian basic helix-loop-helix/PAS family of transcriptional regulators, Int. J. Biochem. Cell Biol. 36 (2) (2004) 189–204.

[7] M.S. Arredouani, et al., Identification of the transcription factor single-minded homologue 2 as a potential biomarker and immunotherapy target in prostate cancer, Clin. Cancer Res. 15 (18) (2009) 5794–5802.

[8] O.J. Halvorsen, et al., Increased expression of SIM2-s protein is a novel marker of aggressive prostate cancer, Clin. Cancer Res. 13 (3) (2007) 892–897.

[9] Q. Long, et al., Protein-coding and microRNA biomarkers of recurrence of prostate cancer following radical prostatectomy, Am. J. Pathol. 179 (1) (2011) 46–54.

[10] M. Ema, et al., Two new members of the murine Sim gene family are transcriptional repressors and show different expression patterns during mouse embryogenesis, Mol. Cell Biol. 16 (10) (1996) 5865–5875.

[11] S.L. Woods, M.L. Whitelaw, Differential activities of murine single minded 1 (SIM1) and SIM2 on a hypoxic response element - Cross-talk between basic helix-loop-helix/Per-Arnt-Sim homology transcription factors, J. Biol. Chem. 277 (12) (2002) 10236–10243.

[12] P. Moffett, M. Reece, J. Pelletier, The murine Sim-2 gene product inhibits transcription by active repression and functional interference, Mol. Cell. Biol. 17 (9) (1997) 4933–4947.

[13] A.L. Farrall, M.L. Whitelaw, The HIF1 alpha-inducible pro-cell death gene BNIP3 is a novel target of SIM2s repression through cross-talk on the hypoxia response element, Oncogene 28 (41) (2009) 3671–3680.

[14] A.E. Sullivan, et al., Characterization of human variants in obesity-related SIM1 protein identifies a hot-spot for dimerization with the partner protein ARNT2, Biochem. J. 461 (2014) 403–412.

[15] Q. Wang, et al., HIF-1alpha up-regulates NDRG1 expression through binding to NDRG1 promoter, leading to proliferation of lung cancer A549 cells, Mol. Biol. Rep. 40 (5) (2013) 3723–3729.

[16] T.P. Ellen, et al., NDRG1, a growth and cancer related gene: regulation of gene expression and function in normal and disease states, Carcinogenesis 29 (1) (2008) 2–8.

[17] E. Goshu, et al., Sim2 contributes to neuroendocrine hormone gene expression in the anterior hypothalamus, Mol. Endocrinol. 18 (5) (2004) 1251–1262.

[18] J.B. Hogenesch, et al., The basic-helix-loop-helix-PAS orphan MOP3 forms transcriptionally active complexes with circadian and hypoxia factors, Proc. Natl. Acad. Sci. USA 95 (10) (1998) 5474–5479.

[19] Y. Adamovich, et al., Rhythmic oxygen levels reset circadian clocks through HIF1alpha, Cell Metab. 25 (1) (2017) 93–101.

[20] C.B. Peek, et al., Circadian clock interaction with hif1alpha mediates oxygenic metabolism and anaerobic glycolysis in skeletal muscle, Cell Metab. 25 (1) (2017) 86–92.

[21] Y. Wu, et al., Reciprocal regulation between the circadian clock and hypoxia signaling at the genome level in mammals, Cell Metab. 25 (1) (2017) 73–85.

[22] T. Eckle, et al., Adora2b-elicited Per2 stabilization promotes a HIF-dependent metabolic switch critical for myocardial adaptation to ischemia, Nat. Med. 18 (5) (2012) 774–782.

[23] S.L. Dunwoodie, The role of hypoxia in development of the Mammalian embryo, Dev. Cell 17 (6) (2009) 755–773.

[24] S. Tomita, et al., Defective brain development in mice lacking the Hif-1alpha gene in neural cells, Mol. Cell Biol. 23 (19) (2003) 6739–6749.

[25] F.R. Sharp, M. Bernaudin, HIF1 and oxygen sensing in the brain, Nat. Rev. Neurosci. 5 (6) (2004) 437–448.

[26] D.B. Sparrow, et al., A mechanism for gene-environment interaction in the etiology of congenital scoliosis, Cell 149 (2) (2012) 295–306.

[27] D. Wu, et al., NPAS1-ARNT and NPAS3-ARNT crystal structures implicate the bHLH-PAS family as multi-ligand binding transcription factors, Elife (2016) 5.

[28] S. Ohsawa, et al., Novel function of neuronal PAS domain protein 1 in erythropoietin expression in neuronal cells, J. Neurosci. Res. 79 (4) (2005) 451–458.

[29] A. Stanco, et al., NPAS1 represses the generation of specific subtypes of cortical interneurons, Neuron 84 (5) (2014) 940–953.

[30] C. Erbel-Sieler, et al., Behavioral and regulatory abnormalities in mice deficient in the NPAS1 and NPAS3 transcription factors, Proc. Natl. Acad. Sci. USA 101 (37) (2004) 13648–13653.

[31] G. Macintyre, et al., Association of NPAS3 exonic variation with schizophrenia, Schizophr. Res. 120 (1–3) (2010) 143–149.

[32] O. Baranova, et al., Neuron-specific inactivation of the hypoxia inducible factor 1 alpha increases brain injury in a mouse model of transient focal cerebral ischemia, J. Neurosci. 27 (23) (2007) 6320–6332.

[33] R. Helton, et al., Brain-specific knock-out of hypoxia-inducible factor-1alpha reduces rather than increases hypoxic-ischemic damage, J. Neurosci. 25 (16) (2005) 4099–4107.

[34] M. Shamloo, et al., Npas4, a novel helix-loop-helix PAS domain protein, is regulated in response to cerebral ischemia, Eur. J. Neurosci. 24 (10) (2006) 2705–2720.

[35] S.J. Zhang, et al., Nuclear calcium signaling controls expression of a large gene pool: identification of a gene program for acquired neuroprotection induced by synaptic activity, PLoS Genet 5 (8) (2009) e1000604.

[36] F.C. Choy, et al., Reduction of the neuroprotective transcription factor Npas4 results in increased neuronal necrosis, inflammation and brain lesion size following ischaemia, J. Cereb. Blood Flow Metab. 36 (8) (2016) 1449–1463.

[37] L. Zhou, AHR function in lymphocytes: emerging concepts, Trends Immunol. 37 (1) (2016) 17–31.

[38] K. Gradin, et al., Functional interference between hypoxia and dioxin signal transduction pathways: competition for recruitment of the Arnt transcription factor, Mol. Cell Biol. 16 (10) (1996) 5221–5231.

[39] W.K. Chan, et al., Cross-talk between the aryl hydrocarbon receptor and hypoxia inducible factor signaling pathways. Demonstration of competition and compensation, J. Biol. Chem. 274 (17) (1999) 12115–12123.

[40] M. Nie, A.L. Blankenship, J.P. Giesy, Interactions between aryl hydrocarbon receptor (AhR) and hypoxia signaling pathways, Environ. Toxicol. Pharmacol. 10 (1–2) (2001) 17–27.

[41] K.A. Lee, et al., Identification and characterization of genes susceptible to transcriptional cross-talk between the hypoxia and dioxin signaling cascades, Chem. Res. Toxicol. 19 (10) (2006) 1284–1293.

[42] I.D. Mascanfroni, et al., Metabolic control of type 1 regulatory T cell differentiation by AHR and HIF1-alpha, Nat. Med. 21 (6) (2015) 638–646.

[43] A. Palazon, et al., HIF transcription factors, inflammation, and immunity, Immunity 41 (4) (2014) 518–528.

[44] H. Zeng, et al., Type 1 regulatory T cells: a new mechanism of peripheral immune tolerance, Cell Mol. Immunol. 12 (5) (2015) 566–571.

[45] R. Gandhi, et al., Activation of the aryl hydrocarbon receptor induces human type 1 regulatory T cell-like and Foxp3(+) regulatory T cells, Nat. Immunol. 11 (9) (2010) 846–853.

[46] L. Apetoh, et al., The aryl hydrocarbon receptor interacts with c-Maf to promote the differentiation of type 1 regulatory T cells induced by IL-27, Nat. Immunol. 11 (9) (2010) 854–861.

[47] N. Gagliani, et al., Th17 cells transdifferentiate into regulatory T cells during resolution of inflammation, Nature 523 (7559) (2015) 221–225.

[48] E.V. Dang, et al., Control of TH17/Treg balance by hypoxia-inducible factor 1, Cell 146 (5) (2011) 772–784.

[49] A. Kimura, et al., Aryl hydrocarbon receptor regulates Stat1 activation and participates in the development of Th17 cells, Proc. Natl. Acad. Sci. USA 105 (28) (2008) 9721–9726.

[50] F.J. Quintana, et al., Aiolos promotes TH17 differentiation by directly silencing Il2 expression, Nat. Immunol. 13 (8) (2012) 770–777.

[51] M. Veldhoen, et al., Natural agonists for aryl hydrocarbon receptor in culture medium are essential for optimal differentiation of Th17 T cells, J. Exp. Med. 206 (1) (2009) 43–49.

[52] G. Liu, et al., Dendritic cell SIRT1-HIF1alpha axis programs the differentiation of CD4+ T cells through IL-12 and TGF-beta1, Proc. Natl. Acad. Sci. USA 112 (9) (2015) E957–E965.

[53] J.A. Goettel, et al., AHR activation is protective against colitis driven by T cells in humanized mice, Cell Rep. 17 (5) (2016) 1318–1329.

[54] F.J. Quintana, et al., Control of T(reg) and T(H)17 cell differentiation by the aryl hydrocarbon receptor, Nature 453 (7191) (2008) 65–71.

39

Chapter 3: Materials and Methods

This chapter details the materials and methods relating to chapters 5 and 6 of this thesis. Materials and methods for Chapter 4 are contained within the manuscript reprinted in this thesis.

3.1 Abbreviations

°C	degrees Celsius
bp	base pair
BSA	bovine serum albumin
DMSO	dimethyl sulfoxide
DNA	deoxyribonucleic acid
dNTPs	deoxyribonucleoside triphosphate
DTT	dithiothreitol
EDTA	ethylene diamine tetra-acetic acid
EtBr	ethidium Bromide
EtOH	ethanol
g	grams
hr(s)	hour(s)
kb	kilobase
kDa	kilodalton
L	litre
LB	luria broth
M	molar
μg	microgram
μL	microlitre
μM	micromolar
mL	millilitre
mM	millimolar
min(s)	minute(s)
MQ H_2O	milliQ Water
mRNA	messenger RNA

mV	millivolts
O/N	overnight
PAGE	polyacrylamide gel electrophoresis
PBS	phosphate buffered saline
PCR	polymerase chain reaction
PEG	polyethylene glycol
pmol	picomol
qPCR	quantitative PCR
RE	restriction endonuclease
RNA	ribonucleic acid
rpm	revolutions per minute
RT	room temperature
RT-PCR	reverse transcriptase PCR
sec(s)	second(s)
SDS	sodium dodecyl sulphate
SN	supernatant
Temp	temperature
Tfb1/2	Transformation buffer 1/2
Tween-20	Polyoxyethylene-sorbitan monolaurate
V	Volts
WCE	Whole cell extract
WT	Wild type

3.2 Materials

3.2.1 Chemicals and Reagents

2-mercaptoethanol	Sigma
Acrylamide	Sigma
Agarose	Sigma
Ammonium persulfate (APS)	Sigma
Bacto-agar	Difco Labs
Bacto-tryptone	Difco Labs

Boric acid	Sigma
Bradford Reagent	Sigma
Bromophenol Blue	Sigma
BSA Powder	Sigma
$CaCl_2$	Sigma
Chloroform	Emsure
Clarity ECL Reagent	Bio-Rad
Collegenase	Worthington
Dispase	Stem Cell Technologies
DMEM Media	Gibco
dNTPs	ThermoFisher
DMSO	Sigma
DTT	Sigma
EDTA	Sigma
EtBr	Sigma
EtOH	Sigma
F12 Media	Gibco
Fetal Calf Serum (FCS)	Gibco
Ficoll-400	Sigma
Gelatin	Sigma
GlutaMAX	Gibco
Glycerol	BDH Chemicals
Glycine	Sigma
Glycogen	Roche Diagnostics
HBSS (Hank's balanced salt solution)	Gibco
HEPES	Sigma
Isopropanol	Emsure
KCl	Sigma
LIF	Gibco
Methanol	Emsure
$MgCl_2$	Sigma
$MgSO_4$	Sigma
$MnCl_2$	Sigma

MOPS	Signa
Na$_2$HPO$_4$	Sigma
NaH$_2$PO$_4$	Sigma
NaCl	Emsure
NEAA (non-essential amino acids)	Gibco
NP-40 (Igepal)	Sigma
PEG-8000	Sigma
Penicillin/Streptomycin	Gibco
Phenol-Chloroform	Sigma
Potassium acetate	Sigma
Precision Plus Protein Standard	Bio-rad
Protease Inhibitors	Sigma
Protein G Sepharose	Invitrogen
RbCl$_2$	Sigma
RPMI-1640	Gibco
SDS	Fisher Scientific
Sodium deoxycholate	Sigma
SYBR Green	Roche and Applied Biosystems
TEMED	Sigma
Tris	Amresco
Triton X-100	ThermoFisher
TRIzol reagent	Invitrogen
Trypsin-EDTA	Gibco
Tryptone	Difco
Tween-20	Sigma
Yeast extract	Difco
1kb Plus DNA Ladder	Life Technologies

3.2.2 Oligonucleotides

Primer	Sequence (5'-3')	Description
Mouse Genotyping		
mSim2Ex1Fwd	CAAAGCGTCCATCATCCGACTC	Screening mouse Sim2 gene
mSim25'Ex1Fwd	GGAGACCAGAAGGAAGAGCGAG	
mSim2-PL451Rev	GCTTGGTTTTCCCGTAAGCTTG	
mSim2-genomicRev	CGAACTAGCCTGCCTGTGTC	
MW16	CCTGACTGATGAAGTTCCTATACTT	
MW14	GCGTTAAACCCTATGTATCTGC	
mSim2 5' Fwd	GCGCAGCGAGGTCTAATATGC	
NestinEn5'	CTTGGCTTTGTACTTTCTGTGACTG	Screening for the Nestin-Cre1 transgene
NestinEn3'	CCTCCATCCCAGACAAATACATTAC	
Cre-1	GCGGTCTGGCAGTAAAAACTATC	Screening for all Cre transgenes
Cre-2	GTGAAACAGCATTGCTGTCACTT	
Flp-1	CACTGATATTGTAAGTAGTTTGC	Screening for the Flp transgene
Flp-2	CTAGTGCGAAGTAGTGATCAGG	
Cre-Pos.Fwd	CTAGGCCACAGAATTGAAAGATCT	Positive control primers used in conjunction with the Flp and Cre primers
Cre-Pos.Rev	GTAGGTGGAAATTCTAGCATCATCC	
Sim2-Tomato		
Sim2TomatoRevATG	CGCGCCTCGGCTCTGGGCATATTAGACCTCGCTGCGCTCCTGGAGCACCCAACGTCTTCT	Cloning mSim2-dnucTomato-pWS-TK6
S2Tom.Ex1.Int1.Fwd	GCGCGCGGTCTTCCCGGAAGGTGAGGCTGCAGGTGGGCGTCGAAGTTATATTATGTACCTGACTGATG	
S2Tom 5'Vect Fwd	TTTGTGAACCGCCCGAAGC	
S2Tom 3'Vect Rev	GTCTCGTGCAGATGGACA	

NLS TdTom Fwd	GGAGCGCAGCGAGGTCTAATATGCC CAGAGCCGAGGCGCGATGACTGCTC CAAAGAAGAAGC	
NLS TdTom Rev	GTATGGCTGATTATGATCCTCTAGT ACTTCTCGACAAGCTTTACTTGTAC AGCTGTCCATG	
Sim2 CRISPR(1)	CACCGCAGGTGGGCGTCCGAGGCG	Cloning mSim2 Exon1 gRNA into pX330
Sim2 CRISPR(2)	AAACCGCCTCGGACGCCCACCTGC	
mSim2 Ex1 sgRNA Fwd	TAATACGACTCACTATAGGGCAGGT GGGCGTCCGAG	Generation of mSim2 Exon1 sgRNA from pX330
5'VectPCR.Seq.Re v	CTTGTTGATTCGAGGTGATG	
SV40 polyA Fwd	AGCTTGTCGAGAAGTACTAGAGG	mSim2-Tomato genotyping and sequencing primers
dnucTomato Rev Geno	CGCATGAACTCTTTGATGACC	
Neo Seq F	CGAGCTCTAGAGAATTGATCC	
mSim2 Seq 1 Fwd	GCAGTGGGAGAATGATTG	
mSim2 Seq 2 Fwd	CCCTGACACTCAACCTCG	
mSim2 5' Ex1 Fwd	GGAGACCAGAAGGAAGAGCGAG	
Heteroduplex Donor-Assisted Direct Integration (HD-ADI)		
S2Tom guide F	CACCGTCTAATATGCCCAGAGCCG	cloning S2Tom gRNA into pX330
S2Tom guide R	AAACCGGCTCTGGGCATATTAGAC	
S2Tom PF1	GGAGCGCAGCGAGGTCTAATATGCC CAGAGGGCGCGATGgtgagcaagggcga gga	PCR primers for generating HD-ADI donor DNA
S2Tom PF2	ATGgtgagcaagggcgagga	
S2Tom PR1	acgcccacctgcagcctc	
S2Tom PR2	TTTTTGGACTTCTCCTTCATCGCGCC TCGGacgcccacctgcagcctc	
mSim2 Tom sgRNA Fwd	TAATACGACTCACTATAGGGTCTAA TATGCCCAGAGCCG	Generation of S2Tom sgRNA from pX330 plasmid

Sim2-Tomato ssDNA oligo	AAGCGGGACTCCGCGGGCCTGGAGCGCAGCGAGGTCTAATATGCCCAGAGCCGAGGCGCGATGGTGAGCAAGGGCGAGGAGGTCATCAAAGAGTTCATGCGCTTCAAGGTGCGCATGGAGGGCTCCATGAACGGCCACGAGTTCGAGATCGAGGGCGAGGGCGAGGGCCGCCCCTACGAGGGCACCCAGACCGCCAAGCTGAAGGTGACCAAGGGCGGCCCCCTGCCCTTCGCCTGGGACATCCTGTCCCCCCAGTTCATGTACGGCTCCAAGGCGTACGTGAAGCACCCCGCCGACATCCCCGATTACAAGAAGCTGTCCTTCCCCGAGGGCTTCAAGTGGGAGCGCGTGATGAACTTCGAGGACGGCGGTCTGGTGACCGTGACCCAGGACTCCTCCCTGCAGGACGGCACGCTGATCTACAAGGTGAAGATGCGCGGCACCAACTTCCCCCCCGACGGCCCCGTAATGCAGAAGAAGACCATGGGCTGGGAGGCCTCCACCGAGCGCCTGTACCCCCGCGACGGCGTGCTGAAGGGCGAGATCCACCAGGCCCTGAAGCTGAAGGACGGCGGCCACTACCTGGTGGAGTTCAAGACCATCTACATGGCCAAGAAGCCCGTGCAACTGCCCGGCTACTACTACGTGGACACCAAGCTGGACATCACCTCCCACAACGAGGACTACACCATCGTGGAACAGTACGAGCGCTCCGAGGGCCGCCACCACCTGTTCCTGTACGGCATGGACGAGCTGTACAAGGCCGTGGATCCAAAAAAGAAGAGAAAGGTAGACCCTAAGAAAAAGAGGAAAGTCGATCCCAAAA	ssDNA oligo used to attempt to integrate dnucTomato coding sequence at the *mSim2* locus

	AGAAAAGAAAAGTGCACGGCTAAGC GGCCGCTTTCGAATCTAGAGCCTGC AGTCTCGACAAGCTTGTCGAGAAGT ACTAGAGGATCATAATCAGCCATAC CACATTTGTAGAGGTTTTACTTGCT TTAAAAAACCTCCCACACCTCCCCCT GAACCTGAAACATAAAATGAATGCA ATTGTTGTTGTTAACTTGTTTATTG CAGCTTATAATGGTTACAAATAAAG CAATAGCATCACAAATTTCACAAAT AAAGCATTTTTTTCACTGCATTCTA GTTGTGGTTTGTCCAAACTCATCAA TGAAGGAGAAGTCCAAAAATGCGGC	
Sim2-3xFLAG	ACACCCAGGCCCCGTCGCCGCCTCTG CGCCCGGCGCGCCCCGACCGCATTAC CTGGGCGCAAGTGTAATAATAACGA ATGGCAGGGACTACAAAGACCATGA CGGTGATTATAAAGATCATGACATC GATTACAAGGATGACGATGACAAGT GACCCGCCAGGGCCAGTCCCGCTCG GCAGGGCGGCCCTCAGGGAGAAGCC ATAGGCCAGGCC	ssDNA oligo used to insert 3xFLAG sequence at the 3' end of the coding sequence in exon 11 of *mSim2*
mSim2-3xFLAG		
mSim2 Ex11 Fwd	CGAGCTGCGGCCACTACC	genotyping FLAG mouse
mSim2 3'UTR Rev	GAGGTCACAGATTGACACCTGG	
mSim2 Tag CR1	CACGCCTCGGTCATCATCACAA	cloning tag gRNA into pX330
mSim2 Tag CR2	AAACTTGGTGATGATGACCGAGGC	
mSim2 Tag sgRNA Fwd	TAATACGACTACTATAGGGCCTCGG TCATCATCAC	Generation of mSim2 Tag sgRNA from pX330 plasmid
mSim2 Ex11 E LR	CTGGGCTGGAATGCACAAGG	

mSim2 3'UTR R LR	CAGGGCCATTCTCTAACATGCC	Genotyping and RT-PCR of *Sim2-3xFLAG*
mSim2 Ex10 Fwd	CCCGGCATGGTCTTGTGC	
mSim2 Tag Rev	TGGTCTTTGTAGTCCCTGCC	
mSim2 602 Fwd	CCTTCGTGCTGCTCAACTACC	
mSim2 Ex10-11 Fwd	AACTCTGAAGCGCCCTCCG	
mSim2 Ex11 Rev	GCCCAGGTAATGCGGTCG	
qPCR Primers		
EPB41L3 F	CGCTAAAGCTGTCCTGGAACAG	qPCR
EPB41L3 R	TACTCCTCCCGGTGAAACACTG	
VEGFA F	CCTTGCTGCTCTACCTCCAC	
VEGFA R	GCAGTAGCTGCGCTGATAGA	
IER3 F	CGGTCCTGAGATCTTCACCTTCG	
IER3 R	TGGTGAGCAGCAGAAAGAGAAG	
DDIT4 F	CTGGACAGCAGCAACAGTG	
DDIT4 R	TGGCACACAAGTGTTCATCC	
CYP1A1 F	CACATGCTGACCCTGGGAAAG	
CYP1A1 R	GGTGTGGAGCCAATTCGGATC	
IL24 F	CAGGGCCAAGAATTCCACTTTGG	
IL24 R	GGGCACTCGTGATGTTATCCTG	
NQO1 F	CCCTGCGAACTTTCAGTATCC	
NQO1 R	CTTTCAGAATGGCAGGGACTC	
POLR2A F	GAGAGTCCAGTTCGGAGTCCT	
POLR2A	CCCTCAGTCGTCTCTGGGTA	
Miscellaneous		
HumanU6.Seq Fwd	ACTATCATATGCTTACCGTAAC	Sequencing
P15Ori.Seq	GTATCACATATTCTGCTGACG	Sequencing
sgRNA Rev	AAAAGCACCGACTCGGTGCC	general reverse primer for generating sgRNAs

T7 promoter primer	TAATACGACTCACTATAGG	Sequencing pX330 CRISPR plasmid

3.2.3 Plasmids

3.2.3.1 Plasmids Described Elsewhere

Plasmid	Purpose
pX330 (addgene)	Cloning and expression of crRNA
pX330mSim2 (Emily Button, 2013)	CRISPR vector containing crRNA sequence that targets the mouse *Sim2* locus in Exon 1.
pNSEN-d2	Plasmid containing dnucTomato sequence
mSim2-LoxP.Frt-Neo-Frt.pWS-TK6 (cloned by Murray Whitelaw)	Plasmid for targeting Sim2 locus in mouse ES cells. Contains Sim2 Ex1 flanked by LoxP sites, neomycin resistance gene cassette flanked by Frt sites, 5' and 3' homology arms for recombination at the Sim2 locus.

3.2.3.2 Plasmids Cloned in This Thesis

Plasmid	Purpose
mSim2M.LoxP.Frt-Neo-Frt.pWS-TK6	Intermediate cloning plasmid. mSim2.LoxP.Frt-Neo-Frt.pWS-TK6 plasmid was digested with Mfe1 restriction endonuclease and ligated to remove 9 kb of Sim2 3' homology arm sequence.
mSim2M-dnucTomato-pWS-TK6	Intermediate cloning plasmid. Sim2 exon1 sequence in mSim2M.LoxP.Frt-Neo-Frt.pWS-TK6 plasmid replaced with dnucTomato coding sequence.
mSim2-dnucTomato-pWS-TK6	Targeting the *Sim2* locus in mouse ES cells and zygotes.

pX330mSim2 Tag	CRISPR vector containing crRNA sequence that targets the mouse *Sim2* locus for generating 3xFLAG tagged mice.
pX330 S2Tom	CRISPR vector containing crRNA sequence that targets the mouse *Sim2* locus in intron 1 for generating Sim2-Tomato mice by the HD-ADI method.

3.2.4 Enzymes

BigDye Terminator   ThermoFisher

DNase I   ThermoFisher

Phusion DNA Polymerase   NEB

Restriction endonucleases   NEB

SAP   NEB

T4 DNA Ligase   NEB

T4 PNK   NEB

T5 exonuclease   NEB

Taq DNA Polymerase   NEB

3.2.5 Antibodies

3.2.5.1 Primary antibodies

Anti-FLAG   M2   Sigma

Anti-SIM2   21069-1-AP   ProteinTech

Anti-α-tubulin   MCA78G   Bio-Rad

Anti-ARNT   14105-1-AP   ProteinTech

Anti-HA   C29F4   Cell Signalling

FLAG Resin   M2   Sigma

3.2.5.2 Secondary antibodies

Goat anti-mouse HRP   Pierce

Goat anti-rabbit HRP	Pierce
Rabbit anti-rat HRP	Dako
Donkey α-mouse Alexa Fluor® 594	Invitrogen

3.2.6 Buffers and Solutions

DNA/RNA loading buffer, 6x	50% (v/v) glycerol, 0.1 mM EDTA pH 8.0, 0.1% (w/v) bromophenol blue
GTS (glycine/tris/SDS) buffer	25 mM Tris, 192 mM glycine and 0.1% SDS
Hypotonic buffer	10 mM HEPES pH 7.9, 1.5 mM $MgCl_2$, 10 mM KCl, 0.4% NP-40 (Igepal), 10% Ficoll-400, 1x protease inhibitors, 1 mM DTT, 1 mM PMSF (optional)
IP buffer	20 mM HEPES pH 7.9, 150 mM NaCl, 1 mM EDTA, 0.02% NP-40 (Igepal), 1x protease inhibitors
Isothermal assembly reaction buffer, 5x	25% PEG-8000, 500 mM Tris/HCl pH 7.5, 50 mM $MgCl_2$, 50 mM DTT, 1 mM of each of the 4 dNTPs, 5 mM NAD
Isothermal assembly reaction, 1.33x	4 μL 5x isothermal assembly reaction buffer, 0.06 units T5 exonuclease, 60 units Taq DNA ligase, 0.375 units Phusion DNA polymerase
LB	1% (w/v) bacto-tryptone, 0.5% (w/v) yeast extract, 1% (w/v) NaCl, pH 7.0
LB agar	LB media with 1.5% bacto-agar
Nuclear extract buffer	20 mM HEPES pH 7.9, 1.5 mM $MgCl_2$, 0.42M KCl, 0.5 mM EDTA, 20% glycerol, 1x protease inhibitors, 1 mM DTT, 1 mM PMSF
PBS	130 mM NaCl, 2.5 mM KCl, 10 mM Na_2HPO_4, 30 mM NaH_2PO_4, pH 7.4
PBS-Tween	PBS with 0.1% (v/v) Tween-20
RIPA lysis buffer	150 mM NaCl, 50 mM Tris pH 8.0, 1% NP-40 (Igepal), 0.5% sodium deoxycholate, 0.1% SDS, 1 mM DTT, 1x protease inhibitors

Separation gel buffer, 4x	181.6 g tris, 40 mL 10% (w/v) SDS, HCl to pH 8.8, H_2O to 1L
Stacking gel buffer, 4x	60.5 g tris, 40 mL 10% (w/v) SDS, HCl to pH 8.8, H_2O to 1 L
SDS loading buffer, 4x	100 mM tris/HCl pH 6.8, 5% (w/v) SDS, 40% (v/v) glycerol, 0.01% (w/v) bromophenol blue, 100 mM DTT
SOC media	0.5% yeast extract, 2% tryptone, 10 mM NaCl, 2.5 mM KCl, 10 mM $MgCl_2$, 10 mM $MgSO_4$, 20 mM glucose
TBE, 20x	1.8 M tris-borate, 40 mM EDTA
TE	10 mM tris pH 8.0, 0.1 mM EDTA
TEN buffer	40 mM tris pH 8.0, 10 mM EDTA, 150 mM NaCl
Tfb1	1.18 g potassium acetate, 4.835 g $RbCl_2$, 0.59 g $CaCl_2$ (dihydrate), 3.96 g $MnCl_2$ (tetrahydrate), 75 mL 80% glycerol, up to 400 mL with H_2O, pH 5.8
Tfb2	0.1256 g MOPS, 0.0726 g $RbCl_2$, 0.6616 g $CaCl_2$ (dihydrate), 11.26 mL 80% glycerol, up to 60 mL with H_2O, pH 6.5
Western wet transfer buffer, 5x	30.25 g Tris, 142.5 g glycine, H_2O to 1 L
Western wet transfer buffer, 1x	400 mL Western wet transfer buffer 5 X, 400 mL methanol, 1.2 L H_2O
Whole cell extract lysis buffer	20 mM HEPES, 0.42 M NaCl, 0.5% NP-40 (Igepal), 25% glycerol, 0.2 mM EDTA, 1.5 mM $MgCl_2$, 12.5 mM DTT, 1mM PMSF, 1x protease inhibitors

3.2.7 Commercial Kits

HiScribe T7 High Yield RNA Synthesis Kit	NEB
KAPA Mouse Genotyping Kit	KAPA Biosystems
mirVana miRNA Isolation Kit	ThermoFisher
NucleoSpin Plasmid Mini Kit	Macherey-Nagel
NucleoBond Xtrra Midi Plus Kit	Macherey-Nagel
Plasmid Mini-prep Kit	QIAGEN

Plasmid Midi-prep Kit	QIAGEN
QIAmp DNA Blood Mini Kit	QIAGEN
QIAquick Gel Extraction Kit	QIAGEN
RNeasy Kit	QIAGEN
SuperScript™ III First-Strand Synthesis System	Invitrogen

3.3 Methods

3.3.1 DNA and RNA Manipulation Methods

3.3.1.1 Plasmid Miniprep and Midiprep

Plasmid purification bacterial cultures was performed using Qiagen Plasmid Mini- or Midi-prep Kits, or Macherey-Nagel NucleoSpin Plasmid Mini or NucleoBond Xtrra Midi Plus kits, according to manufacturer's instructions.

3.3.1.2 Genomic DNA Extraction

Genomic DNA extraction for cell line identification testing was performed using the Qaigen QIAmp DNA Blood Mini Kit according to manufacturer's instructions on cell pellets containing approximately $5x10^6$ cells.

Genomic DNA extraction from mouse embryonic stem cells was performed using the KAPA Express Extract Enzyme. After removal of cell culture media, 50 µl of the following mix was added to each well of a 96 well plate containing ES cell colonies; 5 µl 10x KAPA Express Extract Buffer, 44 µl H_2O, 1 µl KAPA Express Extract Enzyme. The plate was sealed and incubated for 10 mins at 75°C and the 5 mins at 95°C.

3.3.1.3 Polymerase Chain Reaction (PCR)

Routine PCR was performed using Taq DNA Polymerase with the following reaction conditions: 1x thermo pol buffer, 1.5 mM $MgCl_2$, 200 µM dNTPs, 400 nM primers and 1 unit of Taq DNA Polymerase, and the following thermal cycling conditions: 95°C for 3 mins; 24-35 cycles of 95°C for 15 sec, Tm (primer pair specific) for 15 sec and 72°C for 1 min/kb; hold at 11°C.

PCR reactions requiring high fidelity were performed using Phusion DNA Polymerase with the following reaction conditions: 1x GC or HF buffer, 2.5 mM $MgCl_2$, 400 μM dNTPs, 400 nM primers and 0.5-1 unit of Phusion DNA Polymerase, and the following thermal cycling conditions: 98°C for 30 sec; 35 cycles of 98°C for 10 sec, Tm (primer pair specific) for 30 sec and 72°C for 30 sec/kb; hold at 11°C.

3.3.1.4 DNA and RNA Gel Electrophoresis

6x DNA/RNA loading buffer was added to samples to a final 1x concentration, which were then separated by either a 1% (w/v) or 2% (w/v) agarose gel containing EtBr. 1 Kb Plus DNA Ladder was used as a size marker and electrophoresis was performed at 80-120 mV in 1X TBE buffer. DNA fragments were visualised by a BioDoc-It Imaging System (UVP).

3.3.1.5 DNA Gel Extraction

DNA gel extraction was performed using the QIAGEN QIAquick Gel Extraction Kit according to manufacturer's instructions.

3.3.1.6 Restriction Endonuclease (RE) Digest

All restriction endonucleases were purchased from NEB. Restriction endonuclease digestions were performed using 1-5 units of enzyme in the appropriate NEB buffer plus BSA if required and at the temperature specified by NEB. Digestion times were dependent on the amount of DNA to be digested, ranging from 1 hr to overnight.

3.3.1.7 Shrimp Alkaline Phosphatase (SAP) Treatment

DNA was mixed with 1 unit of SAP in 1X CutSmart Buffer and incubated at 37°C for at least 1 hr.

3.3.1.8 DNA Ligation

Ligation reactions were carried out using T4 DNA ligase in 1X T4 DNA Ligase Buffer. Reactions were incubated at room temperature for 1-3 hours or overnight at 16°C before transformation into chemically competent DH5α cells. Correct ligation was

confirmed by RE digest (separated by DNA gel electrophoresis) and/or DNA sequencing.

3.3.1.9 Gibson Isothermal Assembly

5X isothermal assembly reaction buffer (25% PEG-8000, 500 mM Tris/HCl pH 7.5, 50 mM $MgCl_2$, 50 mM DTT, 1 mM of each of the 4 dNTPs, 5 mM NAD). 1.33X isothermal assembly reaction (4 μL 5X isothermal assembly reaction buffer, 0.06 units T5 exonuclease, 60 units Taq DNA ligase, 0.375 units Phusion DNA polymerase). 5 μL of a DNA mix with equimolar ratios of overlapping DNA fragments was added to 15 μL of 1.33X isothermal assembly reaction mix and incubated at 50°C for 1 hr. Assembled products were then transformed into bacteria. For transformation into chemically competent cells, 10 μL of the assembly reaction was transformed into 100 μL of cells. For transformation into electro-competent cells, the reaction was either purified by eluting through a gel extraction column or diluted 1/3 into MQ H_2O with 1 μL of the diluted reaction used for transformation.

3.3.1.10 Sanger DNA Sequencing

Big Dye Terminator reactions were carried out with the following reaction conditions; 200-300 ng template DNA, 4 μL 5x Big Dye buffer, 1 μL Big Dye, 1 μL Primer (4 μM), MQ H_2O up to 20 μL and thermal cycling conditions; 96°C for 2 mins; 25 cycles of 96°C for 30 sec, 50°C for 15 sec, and 60°C for 4 mins; hold at 11°C. The reaction was then transferred to a 1.5 mL Eppendorf tube along with 1 μL of 20 mg/mL glycogen and 80 μL of 75% isopropanol. The reaction mix was then incubated at room temperature for 15 mins, followed by centrifugation at 14,000 rpm at 4°C for 20 mins. The supernatant was removed, and the pellet was washed with of 250 μL of 75% isopropanol, followed by centrifugation at 14,000 rpm at 4°C for 5 mins. The supernatant was removed and the pellet air dried before being sent to the SA Pathology Sequencing Centre for capillary electrophoresis.

Alternatively, 150-300 ng of purified DNA template along with 10 pmol primer was sent to the Australian Genomics Research Facility (AGRF) for Sanger DNA sequencing. Sequencing trace files were then aligned against the appropriate reference sequence using ApE Plasmid Editor (v2.0.53) to confirm the correct sequence.

3.3.1.11 Guide RNA (gRNA) Design

gRNA sequences were designed to fit the required criteria for the pX330 CRISPR plasmid. Potential gRNA sequences were identified in the target region that fit the following sequence criteria; target sequence (protospacer) 20 nucleotides in length beginning with a G, immediately followed by a protospacer adjacent motif (PAM) (NNG);

5' – GNNNNNNNNNNNNNNNNNNN NNG – 3'

Candidate sequences were then screened for any potential off-target effects in a similar method as previously described (H. Y. Wang et al., 2013). 12bp of the gRNA sequence immediately adjacent to the PAM, with the four possible PAM sequences (AGG, TGG, CGG, GGG) were entered into BLAST (NCBI) to search for homology to other genes. Any crRNA sequences (with adjacent PAM sequence) that had 100% homology to regions other than the intended target region were discarded.

3.3.1.12 CRISPR Vector Cloning

Approximately 1 μg of the pX330 vector was digested with the BbsI restriction endonuclease as described in section 3.3.1.6, followed by DNA gel electrophoresis to confirm digestion. The digested vector was then purified by DNA gel extraction as described in section 3.3.1.5.

4 μL of 100 μM of complementary oligonucleotides containing the gRNA sequence and appropriate 5' overhangs (CACC/AAAC) were each phosphorylated using T4 PNK in supplied buffer. The reaction was incubated at 37°C for 2 hrs before heat inactivation of PNK at 70°C for 10 mins. The reaction was diluted in 1X TE buffer and stored at -20°C until use. For annealing, equal amounts of phosphorylated oligos were mixed with 50 mM NaCl (final concentration) and incubated at 95°C for 5 mins before slowly cooling to room temperature. Annealed oligos were diluted 1:100 in H_2O and stored at -20°C until use.

5 μL of the phosphorylated and annealed oligonucleotide mix then ligated into the BbsI digested pX330 plasmid (50 ng) as described in section 3.3.1.8, with an incubation time

of at least 15 mins. 10 μL of the ligation reaction was then transformed into CaCl$_2$ competent DH5α E.coli cells as described in section 3.3.3.2.

3.3.1.13 sgRNA Synthesis

sgRNAs were prepared for injection into mouse embryos by amplification of the sgRNA sequence cloned into the pX330 plasmid by high fidelity PCR as described in section 3.3.1.3. The PCR product was then separated by DNA gel electrophoresis and purified as described in sections 3.3.1.4 and 3.3.1.5, respectively. *In vitro* transcription (IVT) of the purified template was performed using the HiScribe T7 High Yield RNA Synthesis Kit (NEB) according to manufacturer's instructions. The reaction was then DNase treated under the following reaction conditions; 38 μL IVT reaction, 63 μL H$_2$O, 4 μL DNase I; and incubated at 37°C for 15 mins. IVT sgRNA was then purified using the QIAGEN RNeasy Kit according to manufacturer's instructions. Successful IVT was confirmed by RNA quantitation by Nanodrop, and visualisation by RNA gel electrophoresis as described in section 3.3.1.4.

3.3.1.14 Heteroduplex Assay

The target region was amplified by PCR as described in section 3.3.1.3, followed by denaturation and annealing under the following conditions; 95°C for 5 mins, cool to 85°C at 2°C per sec, cool to 25 °C at 0.1°C per sec, hold at 11°C. Products were then separated by polyacrylamide gel electrophoresis. 12% polyacrylamide gel was prepared from the following mix; 3 mL 40% acrylamide, 6.5 mL H$_2$O, 0.5 mL 20X TBE, 120 μL 10% APS, 4.5 mL TEMED. The gel was then run in 1x TBE buffer at 150 V for at least 2 hrs (total time dependent on the size of the PCR product being run). The gel was then stained with EtBr and visualised by a BioDoc-It Imaging System (UVP).

3.3.1.15 RNA Extraction

RNA extraction from tissue or cells for downstream cDNA preparation was performed using either TRIzol or the QIAGEN RNeasy Kit according to manufacturer's instructions. For RNA sequencing, RNA was extracted from MDA-MB-231 cells using the mirVana miRNA Isolation Kit (Thermo Fisher) according to manufacturer's instructions. RNA quality was assessed by quantitation by Nanodrop and visualisation by RNA gel electrophoresis as described in section 3.3.1.4.

3.3.1.16 cDNA Synthesis

cDNA synthesis was performed using the SuperScript III First-Strand Synthesis System. Step 1 reaction was performed using the following reaction conditions; 1 µL 500 ng/µL Oligo dTs, 1 µL 25 µM random hexamers, 2 µL 5mM dNTPs, and 2 µg purified RNA (total reaction volume 20 µL) and incubated at 65°C for 5 mins. Step two reaction was performed by adding the following mix to the Step 1 reaction; 4 µL 5x First Strand Buffer, 1 µL 0.1 M DTT, 1 µL RNAse inhibitor, 0.5 µL Superscript III, and 1.5 µL of H_2O. The reaction was carried out under the following thermal cycling conditions; 25°C for 5 mins, 50°C for 1.5 hrs, 70°C for 15 mins, hold at 11°C. The reaction was then diluted with 30 µL of 1X TE buffer and stored at -20°C.

3.3.1.17 Quantitative PCR (qPCR)

qPCR reactions were performed using either Roche or Applied Biosystems SYBR Green reagent. A master mix for each primer pair was prepared with the following; 22.5 µL 2x SYBR Green Master Mix, 2.25 µL 5 µM each primer, 18µL H_2O. 2 µL cDNA was added to 45 µL of the prepared master mix and 15 µL was aliquoted into 3 wells of a 96 well qPCR reaction plate. Real Time (RT) PCR was carried out using an Applied Biosystems StepOne Real-Time PCR System using the Comparative Ct ($\Delta\Delta C_T$) method with the following conditions; 95°C for 10 mins, 40 cycles of 95°C for 10 sec and 60°C for 30 sec, followed by a melt curve with 0.5°C increments (Pfaffl, 2001). Expression of the target gene was normalised to an internal control gene. Mean normalised expression and standard deviation were calculated for each gene. Where three independent biological repeats were performed, the mean and standard error of the mean were calculated from the mean of each technical triplicate. Statistical analysis was performed on log-transformed data. Statistical significance was determined by One-way ANOVA followed by Dunnett's multiple comparisons test using GraphPad Prism (version 9.0 for MacOS, GraphPad Software, La Jolla California U.S.A., www.graphpad.com).

3.3.1.18 RNA Sequencing

1 µg of purified RNA was sent to the Australian Cancer Research Foundation (ACRF) Cancer Genomics Facility for Illumina while transcriptome sequencing (WTS, RNA

sequencing), with an output of 20-25 million paired-end reads per sample. Bioinformatics were performed by ACRF Cancer Genomics Facility. Briefly, adapters were trimmed from the reads, followed by alignment to the human reference genome (hg19) using the STAR aligner. Differential expression analysis performed using Bioconductor limma and edgeR packages.

3.3.2 Protein Methods

3.3.2.1 Protein Extraction

Protein extraction from mouse tissue was performed by homogenisation in RIPA lysis buffer followed by centrifugation at 14,000 rpm for 30 mins at 4°C.

Cultured cells were harvested for whole cell protein extraction by aspirating the media, adding 1 mL of cold 1x PBS to the dish and scraping cells to detach them from the plate. The cell suspension was then transferred to a 1.5 mL Eppendorf tube and centrifuged at 2,000 rpm for 2 mins at 4°C. Supernatant was discarded, and the cell pellet was resuspended in an appropriate volume of whole cell extract (WCE) buffer), incubated on ice for 10 mins and centrifuged at 14,000 rpm for 30 mins at 4°C. Protein extracts were stored at -80°C until further use.

Cultured cells were harvested for nuclear and cytoplasmic protein extraction by aspirating the media, adding 1.5 mL of TEN buffer to the dish to detach cells. Cells were then scraped, transferred to a 1.5 mL Eppendorf tube and centrifuged at 2,000 rpm for 5 mins at 4°C. Supernatant was removed, and the cell pellet was washed with 1 mL 1X PBS. The cell pellet was then resuspended in hypotonic buffer (approximately 2.5X volume of cell pellet), incubated on ice for 5 mins and centrifuged at 14,000 rpm for 30 mins at 4°C. The supernatant was then removed and stored as the cytosolic fraction at -80°C until further use. The remaining nuclear pellet was resuspended in nuclear extract buffer (approximately 2X volume of cell pellet), incubated with rocking at 4°C for 30-45 mins and centrifuged at 14,000 rpm for 30 mins at 4°C. The supernatant was removed and stored as the nuclear fraction at -80°C until further use.

Protein extracts were quantified by the Bradford protein assay according to manufacturer's instructions.

3.3.2.2 Co-Immunoprecipitation (Co-IP)

Co-IP reactions performed using the FLAG-M2 resin (Sigma-Aldrich) were performed by first washing the resin (20 µL per IP reaction) twice in 1 mL of IP Buffer followed by centrifugation at 2,000 rpm for 2 mins at 4°C. The resin was then blocked in IP buffer containing 0.05 µg/µL BSA by incubation at 4°C with rocking for 30 mins, pelleted, washed once with IP buffer, and resuspended at 50% in IP buffer. 20 µL of 50% blocked FLAG M2 resin was added to 500 µg of protein extract in a total volume of 300 µL in IP buffer and incubated with rocking at 4°C for at least 3 hrs (up to overnight). The reaction was then pelleted and washed three times with 1 mL IP buffer. Protein was eluted from the resin by adding 10 µL of 4x SDS loading buffer and stored at -20°C until further use.

Co-IP reactions performed using primary antibodies were performed by incubating 1 µg of primary antibody with 500 µg of protein extract in 300 µL of IP buffer for at least 3 hrs (up to overnight). Protein G Sepharose (PGS) (40 µL per IP reaction) was washed and blocked with BSA as described above. 40 µL of blocked PGS was then added to each IP reaction and incubated at 4°C with rocking for 1 hr. The IP reaction was then washed and eluted as described above.

3.3.2.3 Immunoblotting

Samples were separated by SDS-PAGE on acrylamide gels that were either prepared in house, or Mini-PROTEAN® TGX™ Precast Gels (Bio-Rad) for 1-2 hrs at 120 V. Protein was transferred to a nitrocellulose membrane using either the wet transfer method at 250 mV for 80mins, or using Trans-Blot® Turbo™ Transfer System (Bio-Rad). Transfer efficiency was assessed by ponceau staining and membranes were blocked in 10% skim milk for 1 hr at RT. Primary antibodies incubations were performed in 1% skim milk at 4°C O/N. Secondary antibody incubations were performed in 1-2% skim milk at RT for 1 hr. Three washes in PBS-Tween for 5 mins were performed after each antibody incubation. Membranes were developed using Clarity™ Western ECL Blotting Substrates and imaged with a Chemidoc MP Imaging System (Biorad).

3.3.2.4 Immunocytochemistry

Cells grown on a glass coverslip or tissue sections on glass slides were fixed with 4% PFA, permeabilised with 0.2% Triton X-100, washed three times with 0.1% PBS-Tween and blocked with 10% normal horse serum. Coverslips or slides were incubated with FLAG M2 primary antibody overnight at 4°C followed by donkey α-mouse Alexa Fluor® 594 secondary antibody at room temperature for 2 hrs. Coverslips or slides were washed three times with 0.1% PBS-Tween after each antibody incubation. Coverslips were mounted onto glass slides with ProLong Gold antifade reagent with DAPI (Invitrogen). Images were taken using a Nikon Eclipse Ti microscope and Nikon Digital Sight DS-Qi1 camera.

3.3.3 Bacterial Methods

3.3.3.1 Preparation of CaCl$_2$ Competent DH5α Cells

DH5α cells were streaked onto a fresh LB agar plate without antibiotics and incubated at 37°C overnight. A single colony was then isolated and cultured in LB overnight at 37°C. 1.1 mL of the overnight culture was subcultured into 25 mL of LB and grown at 37°C until OD$_{600}$ 0.6. The culture was then subcultured into 500 mL of pre-warmed LB and grown at 37°C until OD$_{600}$ 0.4-0.6. The culture was then centrifuged at 4,700 rpm for 40 mins at 4°C, resuspended in 200 mL of cold Tfb1, and centrifuged at 4,700 rpm for 20 mins at 4 °C. The cell pellet was then resuspended in 20 mL of cold Tfb2, aliquoted into 1.5 mL Eppendorf tubes and snap frozen in an EtOH dry ice slurry. Cells were stored at -80°C until use.

3.3.3.2 Transformation of CaCl2 Competent DH5α Cells

CaCl$_2$ competent DH5α cells were allowed to thaw on ice before being incubated with DNA for 20 mins on ice (plasmid, ligation reaction or isothermal assembly reaction). The mixture was then heat shocked at 42°C for 1 min and then placed on ice. For ligation or isothermal assembly reactions 250 μL of SOC media was added and the mixture was incubated at 30-37°C for 45 mins. For plasmid DNA 100 μL of LB was added. 150 μL of the suspension was plated onto LB agar plates with appropriate antibiotics and incubated at 30-37°C overnight.

3.3.3.3 Preparation of Electro-Competent DH5α Cells

DH5α cells were streaked onto a fresh LB agar plate without antibiotics and incubated at 37°C overnight. A single colony was then isolated and cultured in LB overnight at 37°C. The overnight culture was then subcultured 1:250 in LB and grown at 37°C until OD_{600} 0.4-0.6. Cells were pelleted and washed with the following; 1X culture volume MQ H_2O, 1X culture volume 10% glycerol, 0.025X culture volume 10% glycerol. Cells were then resuspended in 0.025X culture volume 10% glycerol, aliquoted into 1.5 mL Eppendorf tubes and stored at -80°C until use.

3.3.3.4 Transformation of Electro-Competent DH5α Cells

Electro-competent DH5α cells were allowed to thaw on ice before being mixed with the DNA for transformation (isothermal assembly reaction mix). The mixture was then transferred to an electroporation cuvette and electroporated using a BioRad MicroPulser. The mixture was immediately suspended in 1 mL of room temperature LB and incubated for 1 hr at 30-37°C with shaking. Cells were pelleted and resuspended in 100 μL of the supernatant which was plated on LB agar plates with appropriate antibiotics and incubated overnight at either 30-37°C.

3.3.4 Cell Culture Methods

3.3.4.1 Primary kidney cell culture

Protocol based off previously described methods (Valente et al., 2011). Kidneys were dissected out from 4-month-old female mice and cut into small pieces. The tissue was then washed 2-3 times in Hank's balanced salt solution (HBSS) to clear the blood, then finely minced. Tissue was digested with 5 mg/mL Dispase (Stem Cell Technologies) and 1mg/mL Collegenase (Worthington) in HBSS (+ 5 mM $CaCl_2$) for 20 mins at 37°C. Following digestion the tissue was passed through a 40 μM filter. Cells were centrifuged at 1000 rpm for 5 mins then resuspended in 2 mL HBSS. This step was repeated 2-3 times. Cells were then counted and seeded on to collagen or gelatin coated plates in DMEM/F-12 with 1% GlutaMAX and 10% FBS.

3.3.4.2 Mouse Embryonic Stem (mES) Cell Culture

mES cells were cultured in ES medium; DMEM (Gibco) with 15% FCS, 1% Penicillin/Streptomycin, 1% glutamine, 0.1% Beta-mercaptoethanol, 0.01 % LIF (ESGRO) and 1% NEAA on dishes coated with gelatin and a layer of irradiated mouse embryonic fibroblast (MEF) cells. Media was changed daily to maintain stem cell state.

3.3.4.3 Generation of Sim2-Tomato mES Cells

4 hours after media change cells were washed twice with PBS and harvested by trypsinisation with 0.25% Trypsin-EDTA (Invitrogen/Gibco) at 37°C for 4 mins. The reaction was stopped by addition of ES medium and cells were pelleted, washed twice with PBS and then resuspended in Ca Mg free PBS. Cells were mixed with 25 μg of pX330 containing mSim2 sgRNA and 25 μg of the mSim2-dnucTomato-pWS-TK6 plasmid, transferred to an electroporation cuvette and electroporated at 320 V for 375 microseconds. Cells were then plated in gelatin coated dishes with a layer of irradiated MEF cells. The next day the following selective agents were added to the media; 2 μM gancilovir and 100 μg/mL G418 (geneticin). Individual colonies were then isolated and grown in 96 well plates.

3.3.4.4 Culture of Human Cell Lines

MDA-MB-231 (human breast epithelial adenocarcinoma), MDA-MB-453 (human brest epithelial carcinoma; metastatic), MCF7 (human breast epithelial adenocarcinoma), T-47D (human breast epithelial carcinoma; ductal), LNCaP (human prostate epithelial carcinoma), PC3-AR+ (human prostate mesenchymal carcinoma) and DU145 (human prostate carcinaoma) cells were cultured in RPMI-1640 medium (Gibco) with 10% FCS, 1% GlutaMax and 1% Penicillin/Streptomycin.

MCF10A (human breast epithelial) cells were cultured in DMEM/F12 (Gibco) with 6 mM L-glutamine, 20 ng/ml EGF, 0.25 U/ml insulin, 0.5 μg/ml hydrocortisone, 100 ng/ml cholera toxin and 5% horse serum.

3.3.4.5 Lentiviral Transduction

Cells were seeded and allowed to settle for approximately 3 hrs before 0.5 mL of lentivirus was added to the cells. 48 hrs later media was replaced with fresh media

containing the appropriate selective agent (200 ng/mL puromycin or 10 µg/mL blasticidin). Successful transduction was confirmed by RNA extraction and qPCR as described in sections 3.3.1.15-17 and by protein extraction followed by immunoblotting as described in sections 3.3.2.1 and 3.3.2.3.

3.3.5 Animal model methods

Animal ethics for all animal based experiments was obtained through the University of Adelaide Animal Ethics Committee. The Nes-Cre1 mice were a kind gift from the McColl Laboratory (School of Biological Sciences, University of Adelaide).

3.3.5.1 Genotyping

Mice were genotyped using the KAPA Mouse Genotyping Kit (KAPA Biosystems). For standard genotyping reactions, ear tissue was taken from mice for genomic DNA extraction. For tissue specific genotyping reactions, a small amount of the desired tissue was taken. The tissue was incubated with 2 units of KAPA Express Extract Enzyme in the KAPA Express Extract Buffer (1X final concentration) at 75°C for 10 mins, 95°C for 5 mins, followed by centrifugation at 10,000 rpm for 1 min. Genomic DNA was stored at -20°C. 1 µL of the supernatant was used in each genotyping PCR reaction. Standard genotyping PCRs were carried out using the following conditions: 1X KAPA2G Fast Genotyping Mix, 400 nM primers and 5% DMSO, and the following thermal cycling conditions: 95°C for 3 mins; 40 cycles of 95°C for 15 sec, Tm (primer pair specific) for 15 sec and 72°C for 30 sec/kb; hold at 11°C.

3.3.5.2 Behavioural Studies

All behavioural studies were conducted by Dr Emily Jaehne (Baune Laboratory, University of Adelaide) and performed as described previously (Baune et al., 2008; Jaehne, Klarić, Koblar, Baune, & Lewis, 2015). The protocols are briefly described below.

Baseline locomotor and anxiety-like behaviour

Open Field: baseline locomotor activity and anxiety-like behaviour

Mice were placed into an open field arena and observed for time spent in each zone (inner and outer), and total distance travelled. Baseline locomotor activity (LMA) was measured as total distance travelled and time spent in each zone as a measure of anxiety-like behaviour, with higher time spent in the inner centre zone demonstrating less anxiety. Statistical significance was measured using an unpaired 2-tailed t-test.

Elevated Zero Maze: Anxiety-like behaviour

The elevated zero maze (EZM) consists of four quadrants, two open (without walls) and two enclosed. Mice were placed into one of the open quadrants and allowed to explore for 5 minutes with time in each quadrant measured. Time spent in the open arms minus the delay to enter the closed arms was calculated as a measure of anxiety-like behaviour. Statistical significance was measured using an unpaired 2-tailed t-test.

Cognition-like behaviour

Y-maze: Spatial Recognition Memory

The Y-maze has three arms, one start arm and two test arms with coloured markings on the inside walls. During the training phase, mice were placed in the maze with one test arm closed and observed for 10 mins. 30 mins later the mice were placed back into the maze with all three arms open and observed for 5 mins. Spatial recognition memory was measured by time spent in the novel test arm. The ratio of the amount of time spent in the novel test arm compared to the familiar test arm was calculated as the preference index and used to measure recognition memory. Statistical significance was measured by a 2-way Anova and Bonferroni post hoc test.

Depression-like behaviour

Forced Swim Test: Despair (depression-like) behaviour

Mice were placed into a circular container filled approximately halfway with water and monitored for 6 minutes. Time spent immobile was used as a measure for depression-like behaviour. More time spent immobile would indicate increased depression-like behaviour. Statistical significance was measured using an unpaired 2-tailed t-test.

Social Behaviour

Sociability test

A rectangular chamber with three enclosures was used to perform this test. Two 7 cm x 15 cm stranger cages were placed in the outer two chambers of the enclosure. In the first stage mice were first acclimatised to the apparatus for 5 mins. The second stage tests sociability. A novel mouse, designated as the stranger mouse, was added in one of the stranger cages. A preference index was calculated as the ratio of time interacting with the stranger mouse compared to the empty cage. The third stage tests preference for social novelty. A novel stranger mouse, designated as the novel mouse, mouse was put in the second stranger cage. Preference for social novelty was determined by a preference index which was calculated as the ratio of time interacting with the novel mouse over the familiar mouse. For each preference index, a value approaching 1 was considered normal behaviour. Statistical significance was measured by 2-way Anova and Bonferroni post hoc test for the preference for social novelty, and unpaired 2-tailed t-test for the preference index.

3.3.5.3 Feeding study

Mice were place on a high fat diet (21% Fat, Specialty Feeds) from age 4 weeks and weighed each week. Weights were averaged between each genotype group and statistical significance was measured by one-way Anova.

Chapter 4: Characterization of functionally deficient SIM2 variants found in patients with neurological phenotypes

4.1 Summary

While it is known that *Sim2* is essential for normal growth, development, and survival in mice, it is not known whether *SIM2* variants could underpin genetic pathologies in humans. This chapter addresses Aim 1 of this thesis; to functionally characterise SIM2 variants found in human patients with intellectual disabilities. *SIM2* non-synonymous variants identified by exome sequencing in patients with neurological phenotypes were assessed to determine the impact on the function of SIM2 as a transcription factor, and therefore whether these variants could be contributing to the clinical presentation of these individuals. This work has been published in a peer reviewed journal.

4.2 Publication 2: Characterization of functionally deficient SIM2 variants found in patients with neurological phenotypes

Button, E. L., Rossi, J. J., McDougal, D. P., Bruning, J. B., Peet, D. J., Bersten, D. C., Rosenfeld, J.A., Whitelaw, M. L. (2022). Characterization of functionally deficient SIM2 variants found in patients with neurological phenotypes. Biochemical Journal, 479(13), 1441-1454. doi: 10.1042/bcj20220209

Statement of Authorship

Title of Paper	Characterization of functionally deficient SIM2 variants found in patients with neurological phenotypes.
Publication Status	☒ Published ☐ Accepted for Publication ☐ Submitted for Publication ☐ Unpublished and Unsubmitted work written in manuscript style
Publication Details	Button, E. L., Rossi, J. J., McDougal, D. P., Bruning, J. B., Peet, D. J., Bersten, D. C., Rosenfeld, J.A., Whitelaw, M. L. (2022). Characterization of functionally deficient SIM2 variants found in patients with neurological phenotypes. Biochemical Journal, 479(13), 1441-1454. doi: 10.1042/bcj20220209

Principal Author

Name of Principal Author (Candidate)	Emily Button
Contribution to the Paper	Designed and performed all experiments (unless otherwise stated) and prepared the manuscript.
Overall percentage (%)	80%
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 06-12-2022

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	Joseph Rossi
Contribution to the Paper	Performed experiment 1c and provided feedback on the manuscript.
Signature	Date

Name of Co-Author	Daniel McDougal
Contribution to the Paper	Contributed the homology modelling.
Signature	Date

Name of Co-Author	John Bruning		
Contribution to the Paper	Contributed the homology modelling.		
Signature		Date	

Name of Co-Author	Daniel Peet		
Contribution to the Paper	Contributed to experiment design and analysis and provided feedback on the manuscript.		
Signature		Date	

Name of Co-Author	David Bersten		
Contribution to the Paper	Contributed to experiment design and analysis and provided feedback on the manuscript.		
Signature		Date	

Name of Co-Author	Jill Rosenfeld		
Contribution to the Paper	Provided patient variant details and feedback on the manuscript.		
Signature		Date	

Name of Co-Author	Murray Whitelaw		
Contribution to the Paper	Designed and analysed experiments and edited the manuscript.		
Signature		Date	06/12/22

Please cut and paste additional co-author panels here as required.

Statement of Authorship

Title of Paper	Characterization of functionally deficient SIM2 variants found in patients with neurological phenotypes.
Publication Status	☒ Published ☐ Accepted for Publication ☐ Submitted for Publication ☐ Unpublished and Unsubmitted work written in manuscript style
Publication Details	Button, E. L., Rossi, J. J., McDougal, D. P., Bruning, J. B., Peet, D. J., Bersten, D. C., Rosenfeld, J.A., Whitelaw, M. L. (2022). Characterization of functionally deficient SIM2 variants found in patients with neurological phenotypes. Biochemical Journal, 479(13), 1441-1454. doi: 10.1042/bcj20220209

Principal Author

Name of Principal Author (Candidate)	Emily Button		
Contribution to the Paper	Designed and performed all experiments (unless otherwise stated) and prepared the manuscript.		
Overall percentage (%)	80%		
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	

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	Joseph Rossi		
Contribution to the Paper	Performed experiment 1c and provided feedback on the manuscript.		
Signature		Date	21/10/22 .

Name of Co-Author	Daniel McDougal		
Contribution to the Paper	Contributed the homology modelling.		
Signature		Date	

70

Statement of Authorship

Title of Paper	Characterization of functionally deficient SIM2 variants found in patients with neurological phenotypes.
Publication Status	☒ Published ☐ Accepted for Publication ☐ Submitted for Publication ☐ Unpublished and Unsubmitted work written in manuscript style
Publication Details	Button, E. L., Rossi, J. J., McDougal, D. P., Bruning, J. B., Peet, D. J., Bersten, D. C., Rosenfeld, J.A., Whitelaw, M. L. (2022). Characterization of functionally deficient SIM2 variants found in patients with neurological phenotypes. Biochemical Journal, 479(13), 1441-1454. doi: 10.1042/bcj20220209

Principal Author

Name of Principal Author (Candidate)	Emily Button
Contribution to the Paper	Designed and performed all experiments (unless otherwise stated) and prepared the manuscript.
Overall percentage (%)	80%
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

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	Joseph Rossi
Contribution to the Paper	Performed experiment 1c and provided feedback on the manuscript.
Signature	Date

Name of Co-Author	Daniel McDougal
Contribution to the Paper	Contributed the homology modelling.
Signature	Date 27/07/2022

Name of Co-Author	John Bruning		
Contribution to the Paper	Contributed the homology modelling.		
Signature		Date	26/7/2020

Name of Co-Author	Daniel Peet		
Contribution to the Paper	Contributed to experiment design and analysis and provided feedback on the manuscript.		
Signature		Date	

Name of Co-Author	David Bersten		
Contribution to the Paper	Contributed to experiment design and analysis and provided feedback on the manuscript.		
Signature		Date	

Name of Co-Author	Jill Rosenfeld		
Contribution to the Paper	Provided patient variant details and feedback on the manuscript.		
Signature		Date	

Name of Co-Author	Murray Whitelaw		
Contribution to the Paper	Designed and analysed experiments and edited the manuscript.		
Signature		Date	

Please cut and paste additional co-author panels here as required.

Name of Co-Author	John Bruning		
Contribution to the Paper	Contributed the homology modelling.		
Signature		Date	

Name of Co-Author	Daniel Peet		
Contribution to the Paper	Contributed to experiment design and analysis and provided feedback on the manuscript.		
Signature		Date	19/07/2022

Name of Co-Author	David Bersten		
Contribution to the Paper	Contributed to experiment design and analysis and provided feedback on the manuscript.		
Signature		Date	

Name of Co-Author	Jill Rosenfeld		
Contribution to the Paper	Provided patient variant details and feedback on the manuscript.		
Signature		Date	

Name of Co-Author	Murray Whitelaw		
Contribution to the Paper	Designed and analysed experiments and edited the manuscript.		
Signature		Date	

Please cut and paste additional co-author panels here as required.

Name of Co-Author	John Bruning		
Contribution to the Paper	Contributed the homology modelling.		
Signature		Date	

Name of Co-Author	Daniel Peet		
Contribution to the Paper	Contributed to experiment design and analysis and provided feedback on the manuscript.		
Signature		Date	

Name of Co-Author	David Bersten		
Contribution to the Paper	Contributed to experiment design and analysis and provided feedback on the manuscript.		
Signature		Date	01/11/22

Name of Co-Author	Jill Rosenfeld		
Contribution to the Paper	Provided patient variant details and feedback on the manuscript.		
Signature		Date	

Name of Co-Author	Murray Whitelaw		
Contribution to the Paper	Designed and analysed experiments and edited the manuscript.		
Signature		Date	

Please cut and paste additional co-author panels here as required.

Name of Co-Author	John Bruning		
Contribution to the Paper	Contributed the homology modelling.		
Signature		Date	

Name of Co-Author	Daniel Peet		
Contribution to the Paper	Contributed to experiment design and analysis and provided feedback on the manuscript.		
Signature		Date	

Name of Co-Author	David Bersten		
Contribution to the Paper	Contributed to experiment design and analysis and provided feedback on the manuscript.		
Signature		Date	

Name of Co-Author	Jill Rosenfeld		
Contribution to the Paper	Provided patient variant details and feedback on the manuscript.		
Signature		Date	7/8/22

Name of Co-Author	Murray Whitelaw		
Contribution to the Paper	Designed and analysed experiments and edited the manuscript.		
Signature		Date	

Please cut and paste additional co-author panels here as required.

PORTLAND PRESS

Research Article

Characterization of functionally deficient SIM2 variants found in patients with neurological phenotypes

Emily L. Button[1], Joseph J. Rossi[1], Daniel P. McDougal[1], John B. Bruning[1,2], Daniel J. Peet[1], David C. Bersten[1], Jill A. Rosenfeld[3,4] and Murray L. Whitelaw[1]

[1]Department of Molecular and Biomedical Science, University of Adelaide, Adelaide, Australia; [2]Institute of Photonics and Advanced Sensing, School of Biological Sciences, University of Adelaide, Adelaide, Australia; [3]Department of Molecular and Human Genetics, Baylor College of Medicine, Houston, TX, U.S.A.; [4]Baylor Genetics Laboratories, Houston, TX, U.S.A.

Correspondence: Murray L. Whitelaw (murray.whitelaw@adelaide.edu.au) or Emily L. Button (emily.button@adelaide.edu.au)

OPEN ACCESS

Single-minded 2 (SIM2) is a neuron-enriched basic Helix–Loop–Helix/PER–ARNT–SIM (bHLH/PAS) transcription factor essential for mammalian survival. *SIM2* is located within the Down syndrome critical region (DSCR) of chromosome 21, and manipulation in mouse models suggests *Sim2* may play a role in brain development and function. During the screening of a clinical exome sequencing database, nine *SIM2* non-synonymous mutations were found which were subsequently investigated for impaired function using cell-based reporter gene assays. Many of these human variants attenuated abilities to activate transcription and were further characterized to determine the mechanisms under-pinning their deficiencies. These included impaired partner protein dimerization, reduced DNA binding, and reduced expression and nuclear localization. This study highlighted several *SIM2* variants found in patients with disabilities and validated a candidate set as potentially contributing to pathology.

Introduction

Single-minded 2 (SIM2) is a member of the basic Helix–Loop–Helix/PER–ARNT–SIM (bHLH/PAS) family of transcription factors, which are broadly known to play important roles in development, homeostasis and cellular stress responses. These factors dimerize with a general bHLH/PAS partner protein (e.g. aryl hydrocarbon nuclear receptor translocator (ARNT)) or the neuronally enriched paralog, ARNT2) to become DNA binding, functional transcription factors. This protein family is characterized by their N-terminal bHLH domain, required for DNA binding and primary dimeriza-tion, followed by a PAS domain consisting of two repeats, PAS-A and PAS-B, which function as sec-ondary dimerization interfaces and confer partner protein specificity. Their C-terminal halves contain transcription regulatory domains (Figure 1A) [1–3]. SIM2/ARNT2 heterodimers bind to the central midline element (CME) DNA binding site to regulate gene expression [4,5]. Two isoforms of SIM2, SIM2-long (SIM2l) and SIM2-short (SIM2s), arise due to alternative splicing of the *SIM2* gene, with each isoform containing a unique C-terminus [6,7]. Two transrepression domains are located in the C-terminal half of the long isoform, with the short isoform only containing one. Both SIM2l and SIM2s have shown the ability to activate or repress transcription, dependent on context [5,7–10]. Target genes and cofactors associated with SIM2 are still largely unknown.

In mice, *Sim2* mRNA is expressed within the brain both during embryonic development and post-natally, indicating that *Sim2* may have important neural functions [4,11,12]. The location of *SIM2* within the Down syndrome critical region (DSCR) on human chromosome 21 stimulated the proposal that *SIM2* may be one of the genes contributing to the complex aetiology of this condition [6,13].
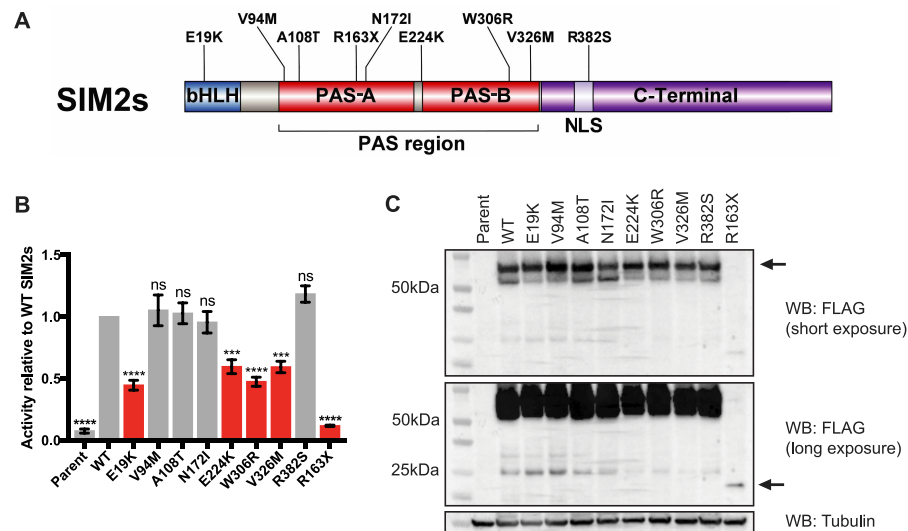
1441

Figure 1. SIM2 variants have reduced activity on a CME reporter.
(**A**) Schematic of SIM2s protein with domains and variants tested shown. N-terminal bHLH domain functions in DNA binding and primary dimerization. PAS domains are important for secondary dimerization and partner protein specificity. NLS directs SIM2 protein import into the nucleus. The C-terminal half of SIM2 contains transcriptional regulatory regions. Variants selected for functional testing are displayed above. (**B**) Dual-luciferase assays with a 6xCME reporter gene. Expression of SIM2s-HF and variants was induced with dox. Variants highlighted in red showed a significant reduction in reporter activity compared with WT SIM2s. Graph represents mean of $n = 3$ independent experiments, normalized to WT SIM2s control. Error bars represent SEM. Statistical significance determined by one-way ANOVA. *** $P \leq 0.001$, **** $P \leq 0.0001$, ns, not significant. (**C**) Western blot from whole-cell extracts. SIM2s-HF and variants were detected with FLAG antibody. Arrow indicates SIM2s variants, with the truncated *R163X* variant only detectable on long exposure.

Supporting this notion, studies characterizing phenotypes of transgenic mice overexpressing *Sim2* reported neuronal/behavioural anomalies of reduced sensitivity to pain and anxiety-related/reduced exploratory behaviour [14,15]. *Sim2* knockout mouse models revealed *Sim2* to be an essential gene, with null mice dying perinatally. Interestingly, newborn *Sim2* knockout pups display many structural defects, including facial dysmorphologies such as cleft palates and skeletal defects of incompletely penetrant congenital scoliosis and abnormal rib protrusions [16,17]. *Sim2* KO mice also have impaired development of somatostatin and thyrotropin-releasing hormone-expressing neurons within the hypothalamus, leading to a reduction in the number of these neurons [18]. The mechanisms behind these phenotypes are yet to be elucidated, so the exact functions that *Sim2* plays during normal development are unclear. Overall, *Sim2* mouse models have shown that aberrant expression of *Sim2* can deleteriously affect brain development and/or function.

Given the phenotypes of *Sim2* KO mice, it is possible that mutations in *SIM2* may be causing or contributing to human developmental disorders including intellectual disabilities, facial dysmorphologies and congenital scoliosis. Exome sequencing data from patients with intellectual disabilities revealed many non-synonymous single-nucleotide variants in the *SIM2* gene. In this study, we functionally tested and characterized *SIM2* variants in order to determine whether they are candidates for causing or contributing to disorders in the patients harbouring those variants. Many patient variants showed a significant reduction in the ability to activate transcription of a reporter gene. These variants were further characterized to determine the mechanism behind the deficiency in protein activity. This work has highlighted many *SIM2* variants as candidates for disability causing or contributing mutations and expanded the set of amino acids that are essential for SIM2 to maintain complete function as a transcription factor.

Results

SIM2 variants in patients with neurological phenotypes

Clinical exome sequencing data were screened to look for rare variants in the *SIM2* gene. Many heterozygous, non-synonymous variants were found, and nine were selected for functional analysis based on their predicted effect on SIM2 protein function (Figure 1A, Table 1, Supplementary Table S1). All variants selected are either not present or present at low frequencies in the Genome Aggregation Database (gnomAD) database (v2.1.1) [19]. Except for the *E19K* variant, which was apparently mosaic in the proband and absent from the parents by trio exome sequencing, the variants were apparently heterozygous and found via proband exome sequencing, so the inheritance of the variants is unknown. The *E19K* variant is within the first helix of the bHLH domain and thus might disrupt DNA binding or dimerization with ARNT2. The *V94M*, *A108T* and *N172I* variants are within the PAS-A repeat of the PAS domain. *E224K*, found in three patients from two families, lies at the N-terminal border of the PAS-B repeat. *W306R* and *V326M* variants lie near the C-terminal end of the PAS-B repeat, positioned in a region known to be important for activity [20]. As the PAS domain is crucial for dimerization with ARNT2, all of these variants were logical choices for activity screening. Additionally, these variants are predicted to be probably damaging to the protein using PolyPhen-2 (Polymorphism Phenotyping v2) predictive software [21]. The *R382S* variant is within the nuclear localization signal (NLS) in the C-terminus of the protein. The *R163X* mutation introduces a premature stop codon truncating the protein in the PAS-A repeat which, therefore, is predicted to disrupt dimerization (Figure 1A). All variants are in protein-coding regions common to both the short and long isoform of SIM2. Phenotypes in common between some patients harbouring functionally deficient *SIM2* alleles include intellectual disabilities, delayed speech, seizure disorders, hypotonia, dysmorphic features and scoliosis. It is worth noting that some of these patients also have mutations in other genes that may well contribute, albeit to unknown degrees, to their complex phenotypes (see Table 1, Supplementary Table S1). Expression cassettes for the short isoform of SIM2 (SIM2s) and each variant, incorporating a 2xHA-3xFLAG epitope tag (SIM2s-HF), were integrated into a set position in the genome of the T-REx293 cell line. This provided a set of stable cell lines with doxycycline-inducible expression for WT SIM2s-HF and each variant for functional characterization of each protein.

SIM2 variants show reduced transcriptional activity on a reporter gene

To test whether the amino acid variants change the activity of SIM2s as a transcription factor, reporter gene assays were performed using the T-REx293 stable cell lines with induced expression of WT SIM2s or patient-derived variants. SIM2s was previously shown to activate the expression of a CME-based reporter gene *in vitro* [5,20,22]. There was no significant change in the activity of the *V94M*, *A108T*, *N172I* and *R382S* variants compared with WT SIM2s, suggesting that these amino acids are not critical for SIM2s function as a transcription factor in this system. These variants exhibit comparable potencies to WT SIM2s so are unlikely to be causing or contributing to any human phenotypes, thus were not analyzed any further. The *E19K*, *E224K*, *W306R* and *V326M* variants all showed approximately 50% reduction in reporter gene activity when compared with the WT protein, whereas the *R163X* variant had near null activity (Figure 1B). Western blot showed that WT SIM2s-HAFLAG and the variants were all expressed as expected in this system after dox induction; however, the *R163X* variant exhibited a low level of protein expression compared with the WT protein (Figure 1C). This may be because the truncated protein is less stable than the WT protein.

Select SIM2 variants show reduced dimerization with partner protein ARNT2

To function as a transcription factor, SIM2 forms a heterodimer with ARNT2. Co-immunoprecipitation experiments were, therefore, performed to assess whether the activity-deficient variants had impaired dimerization with ARNT2. The *E19K*, *E224K* and *V326M* protein variants all appeared to co-immunoprecipitate ARNT2 to a similar extent to WT SIM2s, indicating that they are able to efficiently dimerize with ARNT2 (Figure 2A). In contrast, the *W306R* and *R163X* variants revealed weakened abilities to co-immunoprecipitate ARNT2 (Figure 2A,B). Due to the low level of expression of the R163X variant in the stable cell line, co-immunoprecipitations were repeated using extracts from 293T cells transiently transfected with WT or *R163X* expression plasmids (the latter in 5-fold excess to achieve comparable expression levels). Despite clear immunoprecipitation of *R163X*, a reproducible lack of ARNT2 co-immunoprecipitation was observed (Figure 2B). The weaker dimerization of *W306R* and *R163X* with ARNT2 is consistent with the decrease in

Table 1. SIM2 *(NM_005069.6) gene variants found by clinical exome sequencing shown to be functionally deficient.*

Sex	Age (years)	Phenotype	Variants thought to explain phenotypes	Nucleotide (amino acid) Polyphen-2	Reads	gnomAD Database (allele frequency)
F	19.4	Delayed motor milestones and speech, intellectual disability, hypotonia, seizure disorder, dysmorphic features, short stature, microcephaly, joint contractures, failure to thrive, cerebral palsy, scoliosis and sensitive skin.	Heterozygous c.116C>G (p.S39X) pathogenic variant in the SMC1A gene. SMC1A variants cause Cornelia de Lange syndrome 2 (CDLS2), an X-linked developmental disorder. Phenotypes include facial dysmorphisms, abnormal hands and feet, and growth and developmental delay [45].	**c.55G>A (p.E19K)** *probably damaging*	31/268	Not present
M	3.6	Dysmorphic features, reduced vision, mild hypotonia. Parents are consanguineous. Sibling of below patient.	Homozygous c.1340T>C (p.V447A) variant of unknown clinical significance (VUS) in the TUBGCP6. TUBGCP6 variants cause microcephaly and chorioretinopathy, autosomal recessive, 1 (MCCRP1). Phenotypes include microcephaly, cognitive and visual impairment [46].	**c.670G>A (p.E224K)** *probably damaging*	30/51	0.000007964
F	2.8	Dysmorphic features, reduced vision, unilateral reduced hearing. Parents are consanguineous. Sibling of above patient	Homozygous c.288T>A (p.C96X) pathogenic variant in the SPATA7 gene. SPATA7 variants cause Leber congenital amaurosis 3 (LCA3) and retinitis pigmentosa vision impairment disorders [47].	**c.670G>A (p.E224K)** *probably damaging*	26/47	
F	0.4	Elevated 3-methylglutaconic acid, hyperammonemia, and hypoglycaemia. Chromosomal microarray revealed large regions of absence of heterozygosity (AOH). Parents are consanguineous.	Homozygous c.442C>T (p.R148X) disease causing variant in the SERAC1 gene. SERAC1 variants cause 3-methylglutaconic aciduria with deafness, encephalopathy, and Leigh-like syndrome (MEGDEL), an autosomal recessive disorder. Phenotypes include delayed psychomotor development, hearing loss, movement disorder, and elevated 3-methylglutaconic acid [48].	**c.670G>A (p.E224K)** *probably damaging*	26/70	
M	8.8	Speech delay, autism, intellectual disability, spasticity, seizures, and joint contractures.		**c.916T>C (p.W306R)** *probably damaging*	17/37	0.000008107
F	17.5	Global developmental delay, failure to thrive, seizure disorders, hypertonia, nystagmus, microcephaly, mild scoliosis, decreased volume of grey and white matters and thinned corpus callosum of the brain.		**c.976G>A (p.V326M)** *probably damaging*	21/54	0.00002689
F	21.5	Autism, intellectual disability, hypotonia, hyperextensibility, tachycardia, cardiomyopathy, scoliosis, dysautonomia, and temperature intolerance.	Heterozygous c.490C>T (p.R164X) pathogenic variant in the HDAC8 gene previously reported to cause Cornelia de Lange syndrome 5 (CDLS5) (Deardorff et al 2012), an X-linked dominant developmental disorder. Phenotypes include facial dysmorphisms, abnormal hands and feet, and growth and developmental delay [49].	**c.487C>T (p.R163X)**	31/52	0.00003186

activating potency of this variant, while for *R163X*, the weak dimerization with confounding low expression culminates in little or no activity (Figure 1B).

Dimerization-deficient SIM2 variants show reduced competition with HIF1α

It has been shown previously that SIM2 is able to repress HIF1α reporter gene activity through competitive binding with their common partner factors ARNT/ARNT2 [10,20]. To investigate if any of the SIM2 variants
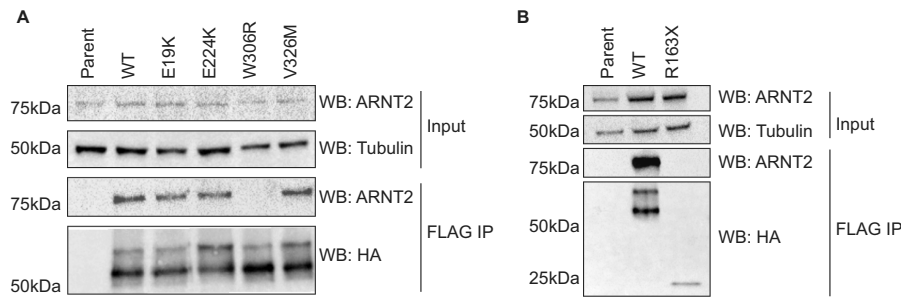
Figure 2. Select SIM2s variants show impaired dimerization with ARNT2.
Co-immunoprecipitation experiments performed to determine the ability of SIM2s-HF and variants to dimerize with ARNT2. FLAG immunoprecipitation followed by western blots for ARNT2, HA (SIM2s-HF) and Tubulin were performed. Blots are representative of 3 independent experiments. (**A**) Expression of SIM2s-HF and the *E19K*, *E224K*, *W306R* and *V326M* variants induced by dox. (**B**) SIM2s-HF and *R163X* variants expressed at a 1 : 5 ratio.

had a reduced ability to act as a transcription repressor in this way, we assayed a reporter gene driven by the HIF1α activated hypoxia response element (HRE) (Figure 3). When HIF1α expression was induced with the hypoxia mimetic dimethyloxalylglycine (DMOG), the activity of the reporter was increased as expected, confirming that HIF1α was dimerizing with ARNT/ARNT2 to form a functional transcription factor. When WT SIM2s expression was concomitantly induced with dox, the activity of the reporter was reduced to basal levels, indicating that SIM2s was competitively dimerizing with ARNT/ARNT2 to prevent HIF1α induction of the reporter. Agreeing with the immunoprecipitation data (Figure 2A,B), the *E19K*, *E224K* and *V326M* variants were all able to repress the HRE reporter gene to a similar extent as WT SIM2s. In contrast, the *W306R* and *R163X* variants were not able to repress HIF1α mediated reporter activity, consistent with attenuated SIM2s dimerization with ARNT/ARNT2.
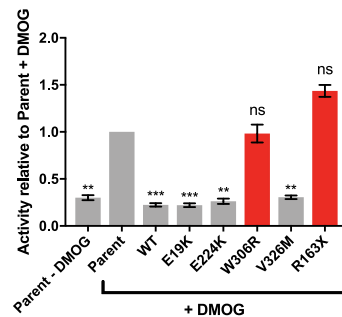
Figure 3. Dimerization-deficient variants lose the ability to repress HIF1α transcriptional activity.
Dual-luciferase reporter assay with a HIF1α responsive 4xHRE reporter. HIF1α expression was induced with the hypoxia mimetic DMOG; expression of SIM2s-HF and variants were induced with dox. Variants highlighted in red show no repression of HIF1α-mediated activation of the reporter. Graph represents mean of $n = 4$ independent experiments, normalized to parent control. Error bars represent SEM. Statistical significance determined by one-way ANOVA. ** $P \leq 0.01$, *** $P \leq 0.001$, ns not significant.

1445

SIM2 variants maintain nuclear localization

SIM2 is localized to the nucleus of cells, and any change of this localization may dampen the transcriptional outputs of SIM2 [20,22,23]. Immunofluorescence was performed to determine whether any of the variants disrupted the transport of SIM2s to the nucleus. As shown in Figure 4, WT SIM2s is predominantly nuclear in T-REx293 cells. The *E19K*, *E224K*, *W306R* and *V326M* variants all show predominant nuclear localization, comparable to WT SIM2s. Therefore, altered cellular localization would not be affecting the activity of these variants. The *R163X* variant is expressed at a much lower level compared with WT SIM2s and also has a distinctly different localization, being distributed throughout the entire cell. This is not unexpected, as this variant is truncated before the NLS. This change in localization of *R163X*, additional to poor expression and minimal
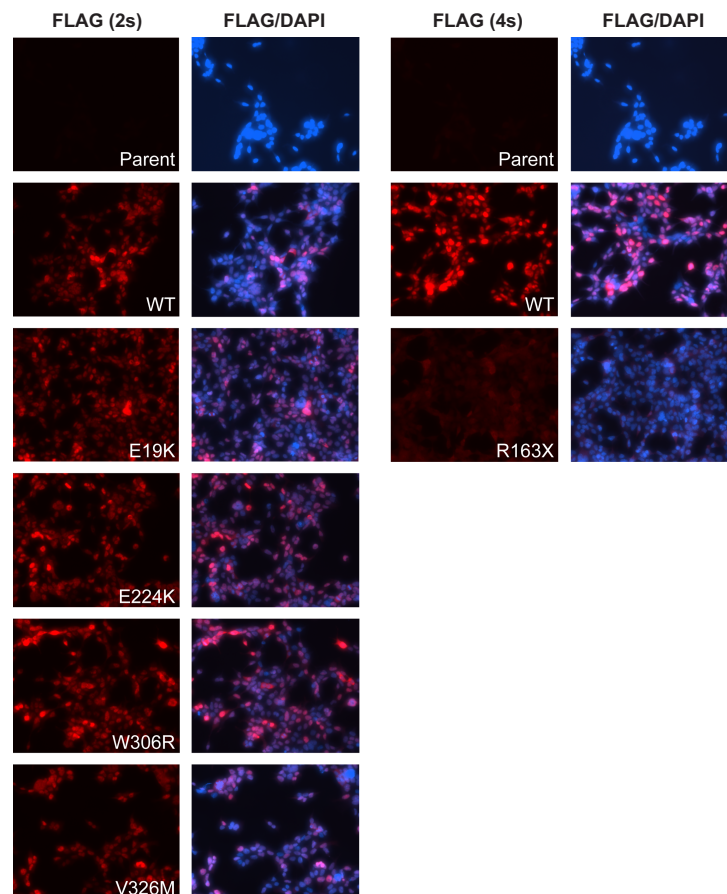

Figure 4. Cellular localization of SIM2 variants.
Immunofluorescence of fixed T-REx293 cells dox induced to express WT SIM2s-HF or mutant variants. FLAG antibody used to detect SIM2 and variants shows all variants are localized to the nucleus except the *R163X* variant, which is localized throughout the entire cell. 2s; 2 second exposure, 4s; 4 second exposure.

dimerization with ARNT2, explains the dramatically decreased activity when compared with the other attenuated SIM2s variants.

DNA binding deficiency of the SIM2 E19K variant

E19K is the sole variant located in the basic DNA binding region of the protein, highlighting this residue as potentially involved in SIM2s/ARNT2 dimer interaction with the DNA response element, CME. To investigate this, chromatin immunoprecipitation (ChIP) experiments were performed by inducing expression of SIM2s WT or the *E19K* variant and assessing enrichment of the 6xCME response element in the pML-6xCME reporter plasmid. Significant enrichment of the 6xCME response element was observed in the WT SIM2s cell line compared with the parent cell line (Figure 5), demonstrating that SIM2s directly binds to the 6xCME response element in the reporter plasmid to activate the expression of the reporter gene. Compared with the WT protein, the *E19K* variant showed approximately a 50% reduction in the enrichment of the 6xCME DNA binding region. Given that *E19K* is not deleterious for nuclear localization (Figure 4) or dimerization with ARNT2 (Figures 2A and 3), reduced DNA binding capability is the most likely cause for decreased transcriptional activity observed for this variant.

Modelling of the SIM2 : ARNT2 : DNA heterodimer structure

Given that no X-ray crystal structure is available for the SIM2 : ARNT2 : DNA complex, a homology model was created to better understand the atomic details regarding the variants analyzed in this study. The model was based on the mouse HIF2α : ARNT : DNA bHLH/PAS-A/PAS-B co-crystal structure (PDB 4ZPK), as HIF2α is the most closely related bHLH/PAS factor to SIM2 with a solved crystal structure [24]. We created a wildtype model that closely mimicked chain lengths, and also modelled mutations for *E19K, E224K, W306R* and *V326M* (Figure 6A). All mutations except *V326M* change the charge of the amino acid, indicating that these substitutions may alter the charge state within the local vicinity of the position, likely interfering with intramolecular interactions and/or the interaction energy between subunits of the dimer. The Trp to Arg mutation of SIM2 at position 306 is found at the SIM2 : ARNT2 dimer interface between the PAS-A and PAS-B domains, at an identical position to HIF2α *Trp318* in the HIF2α : ARNT crystal structure (Figure 6B,C). Within the SIM2 PAS-B domain, *Trp306* likely makes hydrophobic interactions (and potentially pi–pi stacking in a dynamic environment) with *Tyr294* and also with *Leu332*, which itself makes hydrophobic interactions with *Tyr294*. Together, these residues form a small hydrophobic cluster. This cluster is likely to form favourable inter-domain hydrophobic interactions *Ile238* and *Val279* of ARNT2 PAS-A. Mutation of *Trp306* to Arg would disrupt inter-domain hydrophobic interactions and likely perturb interactions between SIM2 *Tyr294* and *Leu332* by (1) altering charge at the interface and (2) sterically affecting interface conformation (Figure 6C,D). The *E19K* mutation disrupts charge complementarity at the dimer interface close to the DNA binding surface, abrogating
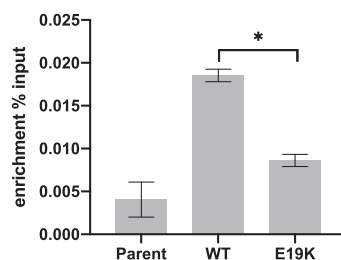
Figure 5. DNA binding of the *E19K* variant.
SIM2 Chromatin immunoprecipitation assays from cells with dox-induced expression of SIM2s-HF or *E19K* variant. Enrichment of the 6xCME response element was assessed by qPCR. Graph represents mean of $n = 3$ independent experiments presented as percent enrichment compared with input. Error bars represent SD. Statistical significance determined by one-way ANOVA. * $P \leq 0.05$ *** $P \leq 0.001$.
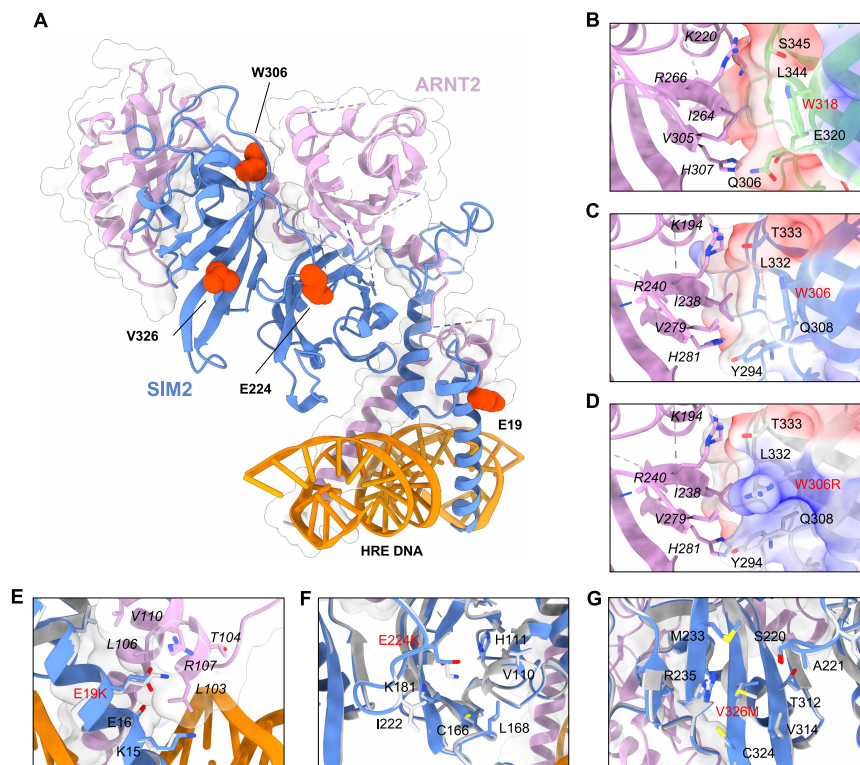
Figure 6. SIM2 : ARNT2 : DNA structural models.

(**A**) Homology model of the WT SIM2 : ARNT : DNA structure based on the mouse HIF2α : ARNT : DNA co-crystal structure (PDB:4ZPK). SIM2 is depicted in blue, ARNT2 in purple with transparent surface shown to highlight protein–protein interfaces. SIM2 missense variant residues are shown as red spheres. (**B**) PAS-A/PAS-B interface of the mouse HIF2α : ARNT : DNA co-crystal structure (PDB:4ZPK). Interface residues are shown as sticks and clearly labelled. Surface electrostatics of the HIF2α are shown. (**C**) PAS-A/PAS-B interface of the WT SIM2 : ARNT2 model and (**D**) mutant W306R SIM2 : ARNT2 model. Surface electrostatics are shown to demonstrate charge disruption introduced by W306R mutation. (**E–G**) Mutations E19K, E224K and V326M, respectively. ARNT1/2 residue labels are represented in italics for clear differentiation, and mutant residues are shown in red.

a potential salt bridge between *Glu19* of SIM2 and *Arg107* of ARNT2 and destabilization of helical secondary structure (Figure 6E). This provides an explanation for the observed weakened affinity for DNA and deficient activity. Mutation *E224K* located on the PAS-A/B loop also disrupts the charge adjacent to the interface (Figure 6F). However, it is difficult to make accurate predictions regarding the effect of *E224K*, as like the crystal structure, the model is missing many residues from the ARNT2 PAS-A domain. The *V326M* mutation is found on the PAS-B β-sheet directly opposing the PAS-A/B loop. The much larger side chain of Met at position 326 of SIM2 likely causes a steric clash with the tip of the loop containing *Ser220* of SIM2, possibly reorganizing the loop at the SIM2:ARNT2 interface and disrupting allostery between the two domains (Figure 6F).

PORTLAND
PRESS

Discussion

Although *SIM2* is an essential gene, its exact functions and molecular pathways are still not well understood. Mouse models have established that deletion of the *Sim2* gene produces a variety of developmental phenotypes; however, the extent to which *SIM2* gene variants cause or contribute to human pathologies remains undetermined. Recently a study proposed a homozygous SIM2 variant (p.Tyr154Cys) as the cause of the clinical presentation in a child with craniofacial abnormalities, developmental delay and intellectual disability [25]. Understanding the molecular basis of how non-synonymous nucleotide variants can alter the activity of SIM2 will not only provide a greater understanding of the way SIM2 functions as a transcription factor but could also highlight the possibility of these variants causing or contributing to human disease.

To study whether single amino acid SIM2 variants found in patients with intellectual disabilities cause functional changes in activity, we used an *in vitro* cell-based system. This consisted of stable cell lines harbouring site-specific inducible expression cassettes for WT SIM2s or mutant variants, which were used to compare relative outputs on a reporter gene. Out of the nine variants selected from the patient exome sequencing dataset, five of these showed a clear significant reduction in transcriptional activity compared with the WT protein (Figure 1). These variants lie across known important functional domains of the protein where there is strong potential for amino acid changes to disrupt protein function. While there are no known human variants linked with human disease as of yet, this finding highlights the possibility that human SIM2 variants can significantly change protein activity.

Further functional analysis of the activity-deficient variants was performed in order to elucidate the mechanism behind the reduced transcriptional activity. As SIM2 must dimerize with ARNT or ARNT2 to form a functional transcription factor, co-immunoprecipitation experiments were used to determine the capacity of these variants to form dimers with ARNT2. This showed that the *W306R* and *R163X* variants had little or no capacity to dimerize with ARNT2. The *W306R* variant lies within the previously defined 'hot spot' region for dimerization [20], confirming that residues within this region form a critical interaction surface. The *R163X* variant is truncated after the first PAS repeat of the PAS domain (PAS-A). It has been shown previously that the entire PAS domain (PAS-A and PAS-B) is important for SIM2 to dimerize with ARNT2 [20]. Therefore, our truncated variant lacking PAS-B confirmed the expected dimerization deficiency. Consistent with previously reported data for the closely related protein SIM1, where the variant SIM1 *V326F* did not influence dimerization [20], the identically located SIM2s *V326M* variant had no change in ARNT2 dimerization compared with the WT protein.

Complementary to the co-immunoprecipitation data, competition reporter gene assays showed that the *W306R* and *R163X* variants failed to repress the activity of the related bHLH/PAS factor, HIF1α, whereas WT SIM2 and the dimerization efficient variants afforded near full repression of HIF1α on a hypoxic response element. While the functional relevance of the competition between HIF1α and SIM2 *in vivo* during development is yet to be established, this loss of competition could have functional and phenotypic consequences. The interaction between genes and the environment has been implicated in the penetrance of scoliosis. Instances of hypoxia during embryogenesis can significantly increase the penetrance of scoliosis in mice that are heterozygous for genes associated with human scoliosis [26]. Many the patients with *SIM2* variants (*E19K*, *V326M*, *R163X*) have scoliosis, a phenotype that is seen in both the heterozygous and homozygous *Sim2* knockout mouse model with incomplete penetrance. Together these observations suggest that *SIM2* may be a susceptibility allele for scoliosis, increasing the incidence of this phenotype, likely in combination with environmental stresses (such as hypoxia) or other gene variants. It is possible that the competition between SIM2 and HIF1α may come into play here, which could be tested through generating mouse models of these variants and observing the penetrance of scoliosis with and without other environmental stresses, such as hypoxic events during embryogenesis, or genetic variants which have been shown to promote scoliosis [26].

It has previously been established that SIM2 is localized to the nucleus in cells [20,23]. Immunofluorescence analysis of the SIM2 variants showed that most are localized to the nucleus, with the exception of the *R163X*, which is localized throughout the cell. As this variant is truncated before the NLS and small in size, this result is to be expected. While such a change in localization would dampen transcriptional outputs of SIM2, we found this variant was also defective for ARNT2 dimerization and additionally has a lower level of expression compared with the WT protein. These properties in combination are the likely reason why this variant has essentially null activity on the CME reporter gene. It is also possible that the native *R163X* transcript would be subject to nonsense-mediated decay *in vivo* due to the introduction of a premature stop codon, resulting in

severely reduced protein being produced from this allele. As the truncated protein is being expressed from cDNA in our experimental cell lines, it was not possible for us to test if this is the case.

Homology modelling indicated that the *E19K* and *W306R* activity-deficient missense variants are likely located at SIM2:ARNT2 dimer interfaces, while *E224* and *V326M* lie adjacent to interfaces. However, by co-immunoprecipitation only one (*W306R*) showed a clear reduction in dimerization, suggesting that structural perturbations occurring at or near interfaces affect other protein properties important for activity, such as allostery. Analysis of the mouse HIF-2α : ARNT crystal structure (4ZPK) suggests a conserved role for HIF-2α *W318* (equivalent position to *W306* in SIM2) in promoting favourable interaction with the ARNT PAS-A domain (Figure 6B). Strikingly, crystal structures show this interaction is disrupted by the HIFa antagonist proflavine, which prevents HIFa/ARNT dimerization in biochemical assays [24,27]. Notably, Wu et al. [28] identified this Trp as being highly conserved in bHLH/PAS family members that dimerize with ARNT or ARNT2, suggesting its crucial role in dimerization that we have established for SIM2 may be general within the protein family. It is, therefore, also likely that mutation of Trp at equivalent positions in ARNT2 partner proteins would have deleterious effects on heterodimerization and ultimately function.

The *E19K* variant is close in proximity to the basic DNA binding region but does not appear to be in direct contact with the DNA response element. ChIP studies showed that the *E19K* variant has reduced DNA binding capabilities compared with the WT protein, so it is possible that this variant may be disrupting DNA binding through structural changes of the bHLH-motif caused by physicochemical properties of the E to K change. The *E224K* and *V326M* variants lie within regions encompassing the PAS-B domain, which is known to serve as a ligand-binding domain for other members of the bHLH/PAS protein family [29]. PAS domains are also known to mediate protein : protein interactions [30–34]. It is, therefore, possible that these variants are disrupting interactions with endogenous ligands or cofactors that are required for SIM2s function, or disrupt intramolecular allostery between the SIM2 PAS domains. As there are no known ligands for SIM2 and cofactors required for SIM2s transcriptional activity are yet to be discovered, these models cannot currently be tested.

While this study clearly shows that these SIM2 variants are transcriptionally deficient within the *in vitro* cell-based assays, these data do not show that these effects alone would translate to a clinical phenotype in patients. However, it is worth noting that while there were some commonalities in phenotypes for patients harbouring these variants, the complete clinical phenotypes were broadly diverse, and some patients also harbour gene variants that have either been linked to, or are potentially associated with, human pathologies (see Table 1). These observations suggest that rather than *SIM2* variant alleles being directly causative of human pathologies, it is more likely that the variants represent susceptibility alleles, where deficient SIM2 functions contribute to human pathologies in combination with other gene variants. In addition, the *SIM2* variants reported in this study when present in a homozygous or compound heterozygous state may result in developmental abnormalities in humans. Future animal studies that recapitulate the *SIM2* variants will be critical to assess their relative contribution(s) to the patient developmental phenotypes. In conclusion, this study has identified an additional set of four amino acids that are important for SIM2 to function as a transcription factor and provides a reference list that will be useful as more patients with intellectual disability and/or developmental abnormalities are discovered to harbour *SIM2* anomalies.

Materials and methods

Exome sequencing

A database of ∼8600 patients undergoing clinical exome sequencing at Baylor Genetics was searched for rare *SIM2* variants. A majority of patients (∼78%) were referred for phenotypes including neurologic symptoms. Most individuals (∼92%) had proband-only sequencing, so the inheritance of variants was not determined. Exome sequencing was conducted according to previously described methods [35].

Plasmid construction

Human SIM2s-2xHA-3xFLAG and variants were generated through PCR amplification and mutagenesis and cloned into pENTR1A by Gibson assembly [36]. These were then subcloned into pcDNA5-FRT/TO-Gateway through LR recombination according to the manufacturer's instructions (Invitrogen). Plasmids pEF-ARNT2-IRES-neo [37] and pGL3-4HRE luciferase [38] have been described previously. pML-6CME was a gift from Dr J Pelletier (Department of Biochemistry, McGill University, Montreal, Canada).

Generation and maintenance of cell lines

Cells were cultivated in DMEM (Dulbecco's modified Eagle's medium; Gibco) with 10% FBS (Gibco), 2 mM GlutaMAX (Gibco), 10 000 units/ml penicillin and 10 mg/ml Streptomycin (Invitrogen) at 37°C, 5% CO_2. Stable cell lines were generated using the Flp-in T-REx293 system according to the manufacturer's instructions (Invitrogen).

Dual-luciferase reporter assays

CME SIM2 activity assays: T-REx293 stable cell lines were seeded in a 24-well tray in triplicate. The following day cells were transfected with 0.1 ng pRL-CMV (Promega), 400 ng pML-6xCME and 50 ng of pEF-ARNT2-IRES-neo using PEI (polyethylenimine) (Polysciences, U.S.A.) transfection reagent [39]. Seven hours after transfection media was replaced with complete medium containing 1 µg/ml doxycycline (dox) (Sigma). Cells were harvested 20 h after dox induction with passive lysis buffer (Promega). Dual Luciferase assays (Promega) were carried out according to the manufacturer's protocol using Promega GloMax™ 96 Microplate Luminometer for at least three biological replicates. To calculate relative luciferase units (RLUs) Firefly luciferase units were normalized to Renilla luciferase units. RLUs were then used to calculate the means of the triplicate transfections for each biological replicate.

HIF1α competition assay: T-REx293 stable cell lines were seeded as above and transfected with 0.1 ng pCMV-RL (renilla luciferase) and 400 ng pGL3-4xHRE luciferase. Seven hours post-transfection media was replaced with complete medium containing 1 µg/ml dox and 1 mM DMOG (Sigma–Aldrich) to induce HIF1α expression. Twenty hours after dox and DMOG treatment cells were harvested and dual-luciferase assays were performed as above.

Immunoblotting

Samples were separated by SDS–PAGE using 7.5%, 12% or 4–20% Mini-PROTEAN® TGX™ Precast Gels (Bio-Rad) before being transferred to nitrocellulose membranes using Trans-Blot® Turbo™ Transfer System (Bio-Rad) and probed with the following primary antibodies; anti-FLAG (M2, Sigma), anti-ARNT2 (M165, Santa Cruz Biotechnology), anti-HA (C29F4, Cell Signalling) and anti-α-tubulin (MCA78G, Bio-Rad). Blots were probed with the following secondary antibodies; goat anti-mouse HRP (Pierce), goat anti-rabbit HRP (Pierce) and rabbit anti-rat HRP (Dako) and developed using Clarity™ Western ECL Blotting Substrates (Bio-Rad). Blots were imaged with a Bio-Rad ChemiDoc MP Imaging System.

Immunoprecipitation

For the SIM2s WT, *E19K*, *E224K*, *W306R* and *V326M* variant immunoprecipitations (IP); T-REx293 cells were treated with 1 µg/µl dox for 6 h before lysis and protein extraction. For the *R163X* variant IP; HEK 293T cells were transfected with 1 µg of pEF-ARNT2-IRES-neo and either 1 µg of pcDNA5-FRT/TO-SIM2s-HA-FLAG plasmid or 5 µg pcDNA5-FRT/TO-R163X-SIM2s-HA-FLAG plasmid. Proteins were immunoprecipitated with FLAG M2 resin (Sigma–Aldrich) overnight at 4°C. Following washes, proteins were eluted from the resin by boiling in 4× SDS load buffer (20% glycerol, 2.5% SDS, 200 mM DTT) and analyzed by immunoblotting.

Immunocytochemistry

Cells were seeded on a glass coverslip and allowed to settle overnight. After treatment for 24 h with dox, cells were fixed with 4% PFA (Sigma), permeabilized with 0.2% Triton X-100 (Sigma) and blocked with 10% normal horse serum. Coverslips were incubated with anti-FLAG antibody (M2, Sigma) overnight at 4°C followed by donkey α-mouse Alexa Fluor® 594 (Invitrogen) at room temperature for 2 h. Coverslips were then mounted onto glass slides with ProLong Gold antifade reagent with DAPI (Invitrogen). Images were taken using a Nikon Eclipse Ti microscope and Nikon Digital Sight DS-Qi1 camera.

Chromatin immunoprecipitation (ChIP)

Cells were seeded in a T175 flask and allowed to settle overnight before being transfected with 20 µg pML-6xCME using PEI transfection reagent. Seven hours after transfection media was replaced with complete medium containing 1 µg/ml doxycycline (dox) (Sigma). After 24 h dox induction cells were cross-linked with 1% formaldehyde for 10 min at room temperature. Chromatin extracts were prepared and immunoprecipitation performed as described previously [40]. Immunoprecipitation was performed with 5 µg of anti-SIM2 antibody

(21069-1-AP, Proteintech), or control anti-rabbit antibody (cyclin D2, M-20, Santa Cruz Biotechnology). ChIP samples were purified through NucleoSpin DNA columns (Macherey-Nagel) using QIAGEN PCR Purification kit buffers according to the manufacturer's instructions and analyzed by qPCR using primers designed to amplify the 6xCME region of the pML-6xCME reporter plasmid.

Statistics

Data from the dual-luciferase assays are presented normalized to the control of that group (WT SIM2s for CME activity assays or parent T-REx293 line + DMOG for HIF1α competition assays) as mean; error bars represent the standard error of the mean (±SEM). Data from the ChIP assays are presented as percent enrichment of the 6xCME response element compared with input; error bars represent standard deviation (±SD). Statistical analysis was performed on log-transformed data. One-way ANOVA followed by Dunnett's multiple comparisons test was performed using GraphPad Prism version 9.0 for MacOS, GraphPad Software, La Jolla California U.S.A., www.graphpad.com.

Homology modelling

All simulations were carried out in ICM Pro software suite (Molsoft) [41]. Homology models were carried out in the Homology Model module of ICM Pro [42,43]. SIM2 (Uniprot: Q14190) and ARNT2 (Uniprot: Q9HBZ2) homology models were first created separately and docked together, and on DNA, using the HIF2α : ARNT : DNA co-crystal structure (PDB:4ZPK) as a reference [24]. Further regularization, minimization and model refinement were then carried out to ensure the integrity of the dimer interface and interactions of the individual mutations. Figures were created using UCSF ChimeraX [44].

Data Availability

All data and reagents are available upon request.

Competing Interests

The authors declare that there are no competing interests associated with the manuscript.

Open Access

Open access for this article was enabled by the participation of University of Adelaide in an all-inclusive *Read & Publish* agreement with Portland Press and the Biochemical Society under a transformative agreement with CAUL.

CRediT Author Contribution

Emily Lyn Button: Conceptualization, Formal analysis, Investigation, Visualization, Writing — original draft. **Joseph J. Rossi**: Investigation, Writing — review and editing. **Daniel P. McDougal**: Investigation, Visualization. **John B. Bruning**: Investigation, Visualization. **Daniel Peet**: Conceptualization, Writing — review and editing. **David Bersten**: Conceptualization, Writing— review and editing. **Jill Rosenfeld**: Resources, Writing — review and editing. **Murray Whitelaw**: Conceptualization, Supervision, Funding acquisition, Writing — review and editing.

Abbreviations

ChIP, chromatin immunoprecipitation; CME, central midline element; DSCR, Down syndrome critical region; HRE, hypoxia response element; IP, immunoprecipitations; NLS, nuclear localization signal; PEI, polyethylenimine; SIM2, single-minded 2; SIM2l, SIM2-long; SIM2s, SIM2-short.

References

1   Bersten, D.C., Sullivan, A.E., Peet, D.J. and Whitelaw, M.L. (2013) bHLH-PAS proteins in cancer. *Nat. Rev. Cancer* **13**, 827–841 https://doi.org/10.1038/nrc3621

2   Button, E.L., Bersten, D.C. and Whitelaw, M.L. (2017) HIF has biff – crosstalk between HIF1a and the family of bHLH/PAS proteins. *Exp. Cell Res.* **356**, 141–145 https://doi.org/10.1016/j.yexcr.2017.03.055

PORTLAND
PRESS

3   Kewley, R.J., Whitelaw, M.L. and Chapman-Smith, A. (2004) The mammalian basic helix-loop-helix/PAS family of transcriptional regulators.
     Int. J. Biochem. Cell Biol. **36**, 189–204 https://doi.org/10.1016/S1357-2725(03)00211-5

4   Ema, M., Suzuki, M., Morita, M., Hirose, K., Sogawa, K., Matsuda, Y. et al. (1996) cDNA cloning of a murine homologue of drosophila single-minded,
     its mRNA expression in mouse development, and chromosome localization. *Biochem. Biophys. Res. Commun.* **218**, 588–594 https://doi.org/10.1006/
     bbrc.1996.0104

5   Moffett, P. and Pelletier, J. (2000) Different transcriptional properties of mSim-1 and mSim-2. *FEBS Lett.* **466**, 80–86 https://doi.org/10.1016/
     S0014-5793(99)01750-0

6   Chrast, R., Scott, H.S., Chen, H.M., Kudoh, J., Rossier, C., Minoshima, S. et al. (1997) Cloning of two human homologs of the drosophila single-minded
     gene SIM1 on chromosome 6q and SIM2 on 21q within the down syndrome chromosomal region. *Genome Res.* **7**, 615–624 https://doi.org/10.1101/
     gr.7.6.615

7   Metz, R.P., Kwak, H.I., Gustafson, T., Laffin, B. and Porter, W.W. (2006) Differential transcriptional regulation by mouse single-minded 2s. *J. Biol. Chem.*
     281, 10839–10848 https://doi.org/10.1074/jbc.M508858200

8   Moffett, P., Reece, M. and Pelletier, J. (1997) The murine Sim-2 gene product inhibits transcription by active repression and functional interference.
     Mol. Cell. Biol. **17**, 4933–4947 https://doi.org/10.1128/MCB.17.9.4933

9   Woods, S., Farrall, A., Procko, C. and Whitelaw, M.L. (2008) The bHLH/Per-Arnt-Sim transcription factor SIM2 regulates muscle transcript myomesin2
     via a novel, non-canonical E-box sequence. *Nucleic Acids Res.* **36**, 3716–3727 https://doi.org/10.1093/nar/gkn247

10   Woods, S.L. and Whitelaw, M.L. (2002) Differential activities of murine single minded 1 (SIM1) and SIM2 on a hypoxic response element - cross-talk
     between basic helix-loop-helix/Per-Arnt-Sim homology transcription factors. *J. Biol. Chem.* **277**, 10236–10243 https://doi.org/10.1074/jbc.
     M110752200

11   Ema, M., Morita, M., Ikawa, S., Tanaka, M., Matsuda, Y., Gotoh, O. et al. (1996) Two new members of the murine Sim gene family are transcriptional
     repressors and show different expression patterns during mouse embryogenesis. *Mol. Cell. Biol.* **16**, 5865–5875 https://doi.org/10.1128/MCB.16.10.
     5865

12   Fan, C.M., Kuwana, E., Bulfone, A., Fletcher, C.F., Copeland, N.G., Jenkins, N.A. et al. (1996) Expression patterns of two murine homologs of
     Drosophila single-minded suggest possible roles in embryonic patterning and in the pathogenesis of down syndrome. *Mol. Cell. Neurosci.* **7**, 1–16
     https://doi.org/10.1006/mcne.1996.0001

13   Michaud, J. and Fan, C.M. (1997) Single-minded - two genes, three chromosomes. *Genome Res.* **7**, 569–571 https://doi.org/10.1101/gr.7.6.569

14   Chrast, R., Scott, H.S., Madani, R., Huber, L., Wolfer, D.P., Prinz, M. et al. (2000) Mice trisomic for a bacterial artificial chromosome with the
     single-minded 2 gene (Sim2) show phenotypes similar to some of those present in the partial trisomy 16 mouse models of down syndrome.
     Hum. Mol. Genet. **9**, 1853–1864 https://doi.org/10.1093/hmg/9.12.1853

15   Ema, M., Ikegami, S., Hosoya, T., Mimura, J., Ohtani, H., Nakao, K. et al. (1999) Mild impairment of learning and memory in mice overexpressing the
     mSim2 gene located on chromosome 16: an animal model of down's syndrome. *Hum. Mol. Genet.* **8**, 1409–1415 https://doi.org/10.1093/hmg/8.8.
     1409

16   Goshu, E., Jin, H., Fasnacht, R., Sepenski, M., Michaud, J.L. and Fan, C.M. (2002) Sim2 mutants have developmental defects not overlapping with
     those of Sim1 mutants. *Mol. Cell. Biol.* **22**, 4147–4157 https://doi.org/10.1128/MCB.22.12.4147-4157.2002

17   Shamblott, M.J., Bugg, E.M., Lawler, A.M. and Gearhart, J.D. (2002) Craniofacial abnormalities resulting from targeted disruption of the murine Sim2
     gene. *Dev. Dyn.* **224**, 373–380 https://doi.org/10.1002/dvdy.10116

18   Goshu, E., Jin, H., Lovejoy, J., Marion, J.F., Michaud, J.L. and Fan, C.M. (2004) Sim2 contributes to neuroendocrine hormone gene expression in the
     anterior hypothalamus. *Mol. Endocrinol.* **18**, 1251–1262 https://doi.org/10.1210/me.2003-0372

19   Karczewski, K.J., Francioli, L.C., Tiao, G., Cummings, B.B., Alföldi, J., Wang, Q. et al. (2020) The mutational constraint spectrum quantified from
     variation in 141,456 humans. *Nature* **581**, 434–443 https://doi.org/10.1038/s41586-020-2308-7

20   Sullivan, A.E., Raimondo, A., Schwab, T.A., Bruning, J.B., Froguel, P., Farooqi, I.S. et al. (2014) Characterization of human variants in obesity-related
     SIM1 protein identifies a hot-spot for dimerization with the partner protein ARNT2. *Biochem. J.* **461**, 403–412 https://doi.org/10.1042/BJ20131618

21   Adzhubei, I.A., Schmidt, S., Peshkin, L., Ramensky, V.E., Gerasimova, A., Bork, P. et al. (2010) A method and server for predicting damaging missense
     mutations. *Nat. Methods* **7**, 248–249 https://doi.org/10.1038/nmeth0410-248

22   Sullivan, A.E., Peet, D.J. and Whitelaw, M.L. (2016) MAGED1 is a novel regulator of a select subset of bHLH PAS transcription factors. *FEBS J.* **283**,
     3488–3502 https://doi.org/10.1111/febs.13824

23   Yamaki, A., Kudoh, J., Shimizu, N. and Shimizu, Y. (2004) A novel nuclear localization signal in the human single-minded proteins SIM1 and SIM2.
     Biochem. Biophys. Res. Commun. **313**, 482–488 https://doi.org/10.1016/j.bbrc.2003.11.168

24   Wu, D., Potluri, N., Lu, J., Kim, Y. and Rastinejad, F. (2015) Structural integration in hypoxia-inducible factors. *Nature* **524**, 303–308 https://doi.org/10.
     1038/nature14883

25   Al-Kurbi, A.A., Da'as, S.I., Aamer, W., Krishnamoorthy, N., Poggiolini, I., Abdelrahman, D. et al. (2022) A recessive variant in SIM2 in a child with
     complex craniofacial anomalies and global developmental delay. *Eur. J. Med. Genet.* **65**, 104455 https://doi.org/10.1016/j.ejmg.2022.104455

26   Sparrow, D.B., Chapman, G., Smith, A.J., Mattar, M.Z., Major, J.A., O'Reilly, V.C. et al. (2012) A mechanism for gene-environment interaction in the
     etiology of congenital scoliosis. *Cell* **149**, 295–306 https://doi.org/10.1016/j.cell.2012.02.054

27   Lee, K., Zhang, H., Qian, D.Z., Rey, S., Liu, J.O. and Semenza, G.L. (2009) Acriflavine inhibits HIF-1 dimerization, tumor growth, and vascularization.
     Proc. Natl Acad. Sci. U.S.A. **106**, 17910–17915 https://doi.org/10.1073/pnas.0909353106

28   Wu, D., Su, X., Potluri, N., Kim, Y. and Rastinejad, F. (2016) NPAS1-ARNT and NPAS3-ARNT crystal structures implicate the bHLH-PAS family as
     multi-ligand binding transcription factors. *Elife* **5**, e18790 https://doi.org/10.7554/eLife.18790

29   Hao, N. and Whitelaw, M.L. (2013) The emerging roles of AhR in physiology and immunity. *Biochem. Pharmacol.* **86**, 561–570 https://doi.org/10.1016/
     j.bcp.2013.07.004

30   Furness, S.G.B., Lees, M.J. and Whitelaw, M.L. (2007) The dioxin (aryl hydrocarbon) receptor as a model for adaptive responses of bHLH/PAS
     transcription factors. *FEBS Lett.* **581**, 3616–3625 https://doi.org/10.1016/j.febslet.2007.04.011

31   Guo, Y., Partch, C.L., Key, J., Card, P.B., Pashkov, V., Patel, A. et al. (2013) Regulating the ARNT/TACC3 axis: multiple approaches to manipulating
     protein/Protein interactions with small molecules. *ACS Chem. Biol.* **8**, 626–635 https://doi.org/10.1021/cb300604u

1453

32 Huang, N., Chelliah, Y., Shan, Y., Taylor, C.A., Yoo, S.-H., Partch, C. et al. (2012) Crystal structure of the heterodimeric CLOCK:BMAL1 transcriptional activator complex. *Science* **337**, 189–194 https://doi.org/10.1126/science.1222804

33 Partch, C.L. and Gardner, K.H. (2011) Coactivators necessary for transcriptional output of the hypoxia inducible factor, HIF, are directly recruited by ARNT PAS-B. *Proc. Natl Acad. Sci. U.S.A.* **108**, 7739–7744 https://doi.org/10.1073/pnas.1101357108

34 To, K.K.W., Sedelnikova, O.A., Samons, M., Bonner, W.M. and Huang, L.E. (2006) The phosphorylation status of PAS-B distinguishes HIF-1alpha from HIF-2alpha in NBS1 repression. *EMBO J.* **25**, 4784–4794 https://doi.org/10.1038/sj.emboj.7601369

35 Yang, Y., Muzny, D.M., Xia, F., Niu, Z., Person, R., Ding, Y. et al. (2014) Molecular findings among patients referred for clinical whole-exome sequencing. *J. Am. Med. Assoc.* **312**, 1870–1879 https://doi.org/10.1001/jama.2014.14601

36 Gibson, D.G., Young, L., Chuang, R.-Y., Venter, J.C., Hutchison CA, I.I.I. and Smith, H.O. (2009) Enzymatic assembly of DNA molecules up to several hundred kilobases. *Nat. Methods* **6**, 343–345 https://doi.org/10.1038/nmeth.1318

37 Ramachandrappa, S., Raimondo, A., Cali, A.M.G., Keogh, J.M., Henning, E., Saeed, S. et al. (2013) Rare variants in single-minded 1 (SIM1) are associated with severe obesity. *J. Clin. Investig.* **123**, 3042–3050 https://doi.org/10.1172/JCI68016

38 Ema, M., Taya, S., Yokotani, N., Sogawa, K., Matsuda, Y. and Fujii-Kuriyama, Y. (1997) A novel bHLH-PAS factor with close sequence similarity to hypoxia-inducible factor 1α regulates the *VEGF* expression and is potentially involved in lung and vascular development. *Proc. Natl Acad. Sci. U.S.A.* **94**, 4273–4278 https://doi.org/10.1073/pnas.94.9.4273

39 Longo, P.A., Kavran, J.M., Kim, M.-S. and Leahy, D.J. (2013) Transient mammalian cell transfection with polyethylenimine (PEI). *Methods Enzymol.* **529**, 227–240 https://doi.org/10.1016/B978-0-12-418687-3.00018-5

40 Farrall, A.L. and Whitelaw, M.L. (2009) The HIF1 alpha-inducible pro-cell death gene BNIP3 is a novel target of SIM2s repression through cross-talk on the hypoxia response element. *Oncogene* **28**, 3671–3680 https://doi.org/10.1038/onc.2009.228

41 Abagyan, R., Totrov, M. and Kuznetsov, D. (1994) ICM—a new method for protein modeling and design: applications to docking and structure prediction from the distorted native conformation. *J. Comput. Chem.* **15**, 488–506 https://doi.org/10.1002/jcc.540150503

42 Abagyan, R., Batalov, S., Cardozo, T., Totrov, M., Webber, J. and Zhou, Y. (1997) Homology modeling with internal coordinate mechanics: deformation zone mapping and improvements of models via conformational search. *Proteins: Struct. Funct. Bioinform.* **29**, 29–37 https://doi.org/10.1002/(SICI)1097-0134(1997)1+<29::AID-PROT5>3.0.CO;2-J

43 Cardozo, T., Totrov, M. and Abagyan, R. (1995) Homology modeling by the ICM method. *Proteins* **23**, 403–414 https://doi.org/10.1002/prot.340230314

44 Eric, F.P., Thomas, D.G., Conrad, C.H., Elaine, C.M., Gregory, S.C., Tristan, I.C. et al. (2021) UCSF ChimeraX: Structure visualization for researchers, educators, and developers. *Protein Sci.* **30**, 70–82 https://doi.org/10.1002/pro.3943

45 Musio, A., Selicorni, A., Focarelli, M.L., Gervasini, C., Milani, D., Russo, S. et al. (2006) X-linked cornelia de lange syndrome owing to SMC1L1 mutations. *Nat. Genet.* **38**, 528–530 https://doi.org/10.1038/ng1779

46 Martin, C.-A., Ahmad, I., Klingseisen, A., Hussain, M.S., Bicknell, L.S., Leitch, A. et al. (2014) Mutations in PLK4, encoding a master regulator of centriole biogenesis, cause microcephaly, growth failure and retinopathy. *Nat. Genet.* **46**, 1283 https://doi.org/10.1038/ng.3122

47 Wang, H., den Hollander, A.I., Moayedi, Y., Abulimiti, A., Li, Y., Collin, R.W.J. et al. (2009) Mutations in SPATA7 cause leber congenital amaurosis and juvenile retinitis pigmentosa. *Am. J. Hum. Genet.* **84**, 380–387 https://doi.org/10.1016/j.ajhg.2009.02.005

48 Wortmann, S.B., Vaz, F.M., Gardeitchik, T., Vissers, L.E., Renkema, G.H., Schuurs-Hoeijmakers, J.H. et al. (2012) Mutations in the phospholipid remodeling gene SERAC1 impair mitochondrial function and intracellular cholesterol trafficking and cause dystonia and deafness. *Nat. Genet.* **44**, 797–802 https://doi.org/10.1038/ng.2325

49 Kaiser, F.J., Ansari, M., Braunholz, D., Concepción Gil-Rodríguez, M., Decroos, C., Wilde, J.J. et al. (2014) Loss-of-function HDAC8 mutations cause a phenotypic spectrum of Cornelia de Lange syndrome-like features, ocular hypertelorism, large fontanelle and X-linked inheritance. *Hum. Mol. Genet.* **23**, 2888–2900 https://doi.org/10.1093/hmg/ddu002

Supplementary Table 1: Additional SIM2 (NM_005069.6) gene variants found by clinical exome sequencing assayed in this study.

Sex	Age (years)	Phenotype	Variants thought to explain phenotypes	Nucleotide (Amino Acid) Polyphen-2	Reads	gnomAD Database (allele frequency)
F	5.8	Acute renal and liver dysfunction, enteropathy, heart disease, global developmental delay, mild hypotonia, dysmorphic features, short statue, failure to thrive, mild structural brain abnormality, and an abnormal N-glycan and transferrin test. Unusual movements possibly related to seizures associated with Streptococcus pneumonia bacteremia		**c.280G>A (p.V94M)** *probably damaging*	31/72	Not present
F	20.7	Postural orthostatic tachycardia syndrome (POTS), possible seizures, abnormal movements (random flinching), gastrointestinal symptoms, slight scoliosis and skin anomalies.		**c.322G>A (p.A108T)** *probably damaging*	46/85	0.00001063
M	66.0	Bilateral hearing loss, multifactorial gait difficulty, balance issues, weakness, peripheral neuropathy, cerebellar atrophy, short term memory loss, vitamin B12 deficiency and history of concussion.	Heterozygous c.35delG (p.G12fs) variant in the *GJB2* gene, a common pathogenic variant associated with autosomal recessive deafness (1). A novel hemizygous c.197C>A (p.A66D) variant of unknown significance (VUS) in the *BCAP31* gene. *BCAP31* variants cause deafness, dystonia, and cerebral hypomyelination (DDCH), an X-linked recessive disorder with phenotypes including lack of psychomotor development, dysmorphic facial features, deafness, dystonia and cerebral hypomyelination(2).	**c.515A>T (p.N172I)** *probably damaging*	21/53	0.000003977
F	7.9	Prematurity, delayed speech, refractory epilepsy, structural brain abnormalities (right frontal developmental venous anomaly), chronic otitis media, and persistent asthma.		**c.1146A>C (p.R382S)** *probably damaging*	83/184	0.000007964

90

Supplementary References

1.   Kenneson A, Van Naarden Braun K, Boyle C. GJB2 (connexin 26) variants and nonsyndromic sensorineural hearing loss: A HuGE review. Genetics in Medicine. 2002;4(4):258-74.
2.   Cacciagli P, Sutera-Sardo J, Borges-Correia A, Roux J-C, Dorboz I, Desvignes J-P, et al. Mutations in BCAP31 Cause a Severe X-Linked Phenotype with Deafness, Dystonia, and Central Hypomyelination and Disorganize the Golgi Apparatus. The American Journal of Human Genetics. 2013;93(3):579-86.

Chapter 5: Molecular functions of SIM2 in cancer

5.1 Introduction

All types combined, cancer is one of the highest causes of mortality globally each year (WHO, 2022). Therefore, it is important to continue adding to our understanding of the molecular characteristics in tumours that result in enhanced growth and tumour progression. We have now entered the era of precision medicine, where cancer therapeutics can be tailored to subsets of patients whose cancer presents with specific molecular or cellular features, or biomarkers. These commonly include changes in gene or protein expression, and genomic sequence or architecture aberrations. Biomarkers are identified by a variety of clinical tests, including genomic analysis (whole genome sequencing and targeted mutation testing), immunohistochemistry and fluorescence in situ hybridisation. There are already many approved biomarkers targeted by cancer therapeutics, for example the estrogen/progesterone receptors (ER/PR) for breast cancer and *BCR/ABL* for chronic myeloid leukaemia (Malone, Oliva, Sabatini, Stockley, & Siu, 2020; Oliveira et al., 2020). Given these successes, it is now important to understand the molecular functions of recently proposed cancer biomarkers, to determine whether they will be suitable cancer therapeutic targets.

bHLH/PAS transcription factors have diverse roles in normal development and cellular homeostasis, and their molecular functions in cancer are also highly varied and context dependent. Many bHLH/PAS transcription factors have been shown to have both pro- and anti-tumourigenic functions, including the hypoxia inducible factors (HIF1α and HIF2α), AHR, and SIM2 (Bersten et al., 2013; Emily L. Button et al., 2017; G. L. Semenza, 2003).

The molecular functions of SIM2 in cancer has been shown to be highly context dependent and having apparent opposing effects in different cancer types. Expression of both isoforms of SIM2 have been demonstrated to be upregulated in prostate, pancreatic and colon tumours, compared to the normal surrounding tissue (Aleman et al., 2005; Arredouani et al., 2009; DeYoung et al., 2003a, 2003b; Ole J. Halvorsen et al.,

2005; Ole Johan Halvorsen et al., 2007; Lu et al., 2011). Further studies indicate that SIM2 may be an important factor for the progression of these cancers and a number of studies have suggested SIM2 as a potential biomarker and therapeutic target for prostate, pancreatic and colon tumours (Arredouani et al., 2009; DeYoung et al., 2003a, 2003b; Ole Johan Halvorsen et al., 2007; Souza et al., 2017).

Knockdown of SIM2 expression both *in vitro* and in mouse xenograft models reduces cancer cell growth and tumourigenic properties, suggesting a role for SIM2 in the progression of pancreatic and colon cancers (Aleman et al., 2005; DeYoung et al., 2003a, 2003b). Treatment of both CAPAN-1 pancreatic cells and RKO colon carcinoma cells caused growth inhibition and induction of apoptosis *in vitro* and in the case of the RKO cells, in a mouse xenograft model (DeYoung et al., 2003a, 2003b).

Upregulation of both isoforms of SIM2 in prostate cancer has been shown to correlate with a number of adverse clinical indicators, including increased serum prostate specific antigen (PSA), increased levels of tumour growth and proliferation, and a high histological grade. In addition, increased expression of SIM2 in prostate cancer correlates with a poorer patient prognosis, with the estimated 10- and 13-year survival rate in patients without and with SIM2s expression being 98.1% and 98.1% versus 79.5% and 72.8%, respectively (Ole Johan Halvorsen et al., 2007). SIM2 has been proposed as a potential immunotherapy target for prostate cancer patients, as it has been identified that some patients with upregulation of SIM2 display an immunological response, with autoantibodies to SIM2 detected in patient serum. In addition, in mice it has been shown that a cytotoxic T-cell response can be triggered in response to SIM2 epitopes (Arredouani et al., 2009; Kissick, Sanda, Dunn, & Arredouani, 2014). Microarray analysis of PC3 human prostate cancer cells with *SIM2* knockdown identified potential *SIM2* target genes in this cancer type. In addition, a number of metabolic pathways were dysregulated with *SIM2* knockdown, including purine and pyrimidine metabolism, and glycolysis or glycogenesis (Lu et al., 2011).

Interestingly, and opposing to the above, in breast and esophageal cancers the short and long SIM2 isoforms, respectively, have been observed to be downregulated, and studies suggest that downregulation promotes tumour progression (Gustafson et al.,

2009; Kwak et al., 2007; Laffin et al., 2008; S. J. Pearson et al., 2019; Scott J. Pearson et al., 2019; Scribner et al., 2013; Su et al., 2016; Tamaoki et al., 2018). In Harvey-*Ras* (Ha-*Ras*) transformed MCF10A normal human breast cells downregulation of *SIM2s* expression appears to occur through the NOTCH signalling pathway (Gustafson et al. 2009). In mouse tumour models containing SIM2 low expressing breast cancer cells, the tumours formed have a more de-differentiated state and increased lung metastasis (Scribner et al., 2013). Loss of SIM2 also leads to epithelial to mesenchymal transition (EMT) in tumour cells and SIM2 has been shown to regulate the expression of key EMT factors SLUG and Matrix Metalloprotease 1, 2, 3 and 9 (MMP1, MMP2, MMP3, MMP9) (Kwak et al., 2007; Laffin et al., 2008; Scott J. Pearson et al., 2019; Scribner et al., 2013). Manipulation of SIM2 expression in breast cancer cells can cause changes in tumourigenic properties both *in vitro* and in mouse xenograft models. Knockdown of SIM2 increases tumour cell growth, whereas overexpression of SIM2 reduces tumour cell growth and invasiveness (Kwak et al., 2007; Laffin et al., 2008; Scribner et al., 2013). This highlights a role for SIM2 in repression of EMT and mammary cell development and differentiation (Scott J. Pearson et al., 2019; Scribner et al., 2013; Wellberg, Metz, Parker, & Porter, 2010).

Furthermore, it has been shown SIM2 regulates genomic stability through promoting homologous recombination, and loss of SIM2 expression results in decreased genomic DNA stability. Experiments in human breast cancer cell line DCIS.com cells showed that SIM2s is stabilised in response to ionising radiation induced DNA damage through phosphorylation by ATM (ataxia-telangiectasia mutated) (Scott J. Pearson et al., 2019). Once stabilised, SIM2s interacts with BRCA1 to allow for RAD51 recruitment at sites of DNA damage, leading to repair through homologous recombination. Loss of SIM2s stabilisation through mutation of the predicted ATM phosphorylation residue (S115A) leads to genomic instability, resulting in EMT of DCIS.com cells both *in vitro* and *in vivo* in a mouse xenograft model (Scott J. Pearson et al., 2019). Furthermore, genomic instability caused by loss of SIM2s results in replication fork collapse due to lack of RAD51 recruitment, leading to aneuploidy and chromatin fragmentation through abnormal sister chromatid segregation during mitosis (S. J. Pearson et al., 2019).

SIM2 has also been implicated in the regulation of the NFκB signalling pathway, which is known to promote breast cancer progression and metastasis. SIM2 is able to repress NFκB signalling through regulation of its critical factors IKKα, IKKβ, phosphorylated-p65 and p65 protein, which are all downregulated with SIM2 overexpression, and conversely upregulated with SIM2 knockdown in breast cancer cells. SIM2 expression is reciprocally repressed through NFκB signalling (Wyatt et al., 2019).

SIM2 has been proposed as a biomarker for uterine cervical squamous cell carcinoma (CvSCC), with high levels of SIM2 expression correlating with a better survival rate in CvSCC patients. Knockdown of SIM2 in CvSCC cell lines suggests that this may be through interference with the bHLH/PAS transcription factor HIF1α, due to an observed increase in *HIF1A* expression and its target genes, including *VEGFA*. Knockdown of SIM2 in CvSCC cells also leads to increased cell growth, and increased resistance to oxidative and radiation stress (Nakamura et al., 2017).

As detailed above, it is apparent that the function of SIM2 in cancer is highly variable and context dependent, with both pro- and anti-tumourigenic roles for SIM2 demonstrated in various cancer types. Given that SIM2 has been suggested as potential biomarker and therapeutic target for cancers where its expression is aberrantly upregulated, it is important to further understand both the pro- and anti-tumourigenic functions of *SIM2* to determine if it indeed would be a suitable therapeutic target. The work in this chapter aims to add further knowledge to the molecular functions of SIM2 in a cancer context to aid in determining how SIM2 can have apparently opposing actions in different cancer types.

5.2 Aims

The overall aim of this chapter is to investigate the molecular function of *SIM2* in a human cancer context, focusing on gene regulation. This is broken up into the following sub aims:

1. To generate and characterise human breast and prostate cancer cell lines with inducible overexpression of *SIM2*.
2. To identify *SIM2* target genes in a breast cancer context.

5.3 Experimental Systems and Approaches

5.3.1   Cell culture systems

This chapter utilises cell culture systems to investigate the function of *SIM2* in a cancer context. Several human breast and prostate cancer cell lines were used in the work presented in this study, as detailed in Table 5.1.

Table 5.1: Human cell lines used in this study.

Cell Line	Origin
MDA-MB-231	Human breast epithelial adenocarcinoma
MDA-MB-453	Human breast epithelial carcinoma; metastatic
MCF7	Human breast epithelial adenocarcinoma
T-47D	Human breast epithelial carcinoma; ductal
MCF10A	Human breast epithelial
LNCaP	Human prostate epithelial carcinoma
PC3-AR+	Human prostate mesenchymal carcinoma
DU145	Human prostate carcinoma

The MDA-MB-231, MCF10A and T-47D cell lines were a kind gift from the Tilley Laboratory (Dame Roma Mitchell Cancer Research Laboratories, University of Adelaide).

In addition, manipulated versions of these cell lines were generated and used to investigate the aims of this chapter. These cell lines are described in Table 5.2. MDA-MB-231 and LNCaP cells with doxycycline inducible expression of both a low, and high level of SIM2s were generated as part of this project. SIM2s was epitope tagged with a HA-FLAG tag to allow for detection of protein expression. These cells lines will be referred to as MDA-MB-231 SIM2-low, MDA-MB-231 SIM2s-high, LNCaP SIM2s-low and LNCaP SIM2s-high.

These cell lines were generated using the LVTPTIP lentiviral mediated method (Figure 5.1) (Bersten et al., 2015). This method makes use of the Tet-On 3G system for

doxycycline (dox) inducible expression of the gene of interest, without significant background gene expression in the absence of dox. The gene of interest is cloned via Gateway cloning into the lentiviral vector downstream of the TRE3G promoter. The constitutive Phosphoglycerate Kinase-1 promoter (PGKp) drives expression of the TET-ON 3G protein which when in the presence of dox will activate the TRE3G promoter and drive expression of the gene of interest. The PGKp also drives expression of an antibiotic resistance cassette for selection of cells with vector integration (Bersten et al., 2015; Loew, Heinz, Hampf, Bujard, & Gossen, 2010).

The SIM2s-high cell lines were generated using lentivirus produced with the LVTPTIP vector containing a puromycin resistance selection cassette, which had been observed to result in a high level of dox inducible expression of the gene of interest. The SIM2s-low cell lines were generated using lentivirus produced with a modified LVTPTIP vector that had a blastocydin resistance selection cassette, which had been observed to result in a low level of dox inducible expression of the gene of interest (Bersten et al. 2015, Whitelaw Laboratory, unpublished data). The expression differences observed between these two antibiotic resistance cassettes presumably reflects the higher stringency of puromycin selection. The lentiviral vectors and lentivirus were made by Dr David Bersten and Dr Adrienne Sullivan (Whitelaw Laboratory, University of Adelaide).
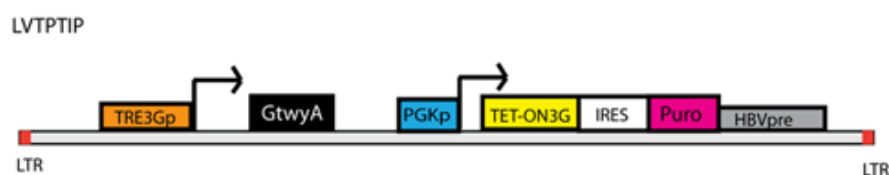
Figure 5.1: Vector used to generate Lentiviral destination plasmids for doxycycline inducible expression of the gene of interest, which is incorporated into the plasmid via Gateway cloning. TRE3G promoter (TRE3Gp) is upstream of the Gateway (GtwyA) cloning site where the gene of interest is inserted. The PGK promoter (PGKp) drives expression of TET-ON3G protein, which when in the presence of dox activates expression from the TRE3Gp. This is followed by a selection gene (in this case Puromycin resistance (Puro) with an internal ribosomal entry site (IRES) to allow for expression. Image reproduced under Attribution 4.0 International (CC BY 4.0) Creative Commons license, from Bersten et al., 2015, doi: 10.1371/journal.pone.0116373.

All other cell lines described in Table 5.2 were generated by Dr Adrienne Sullivan (Whitelaw Laboratory, University of Adelaide), as indicated by asterisks.

Table 5.2: Manipulated cell lines used in this study.

Cell Line	Description
MDA-MB-231 SIM2s-HF-low	Doxycycline inducible expression of SIM2s tagged with HA-FLAG at a relatively low level. Referred to as MDA-MB-231 SIM2s-low.
MDA-MB-231 SIM2s-HF-high	Doxycycline inducible expression of SIM2s tagged with HA-FLAG at a relatively high level. Referred to as MDA-MB-231 SIM2s-high.
*MDA-MB-453 SIM2s	Doxycycline inducible expression of SIM2s.
*MDA-MB-453 SIM2l	Doxycycline inducible expression of SIM2l.
*MCF7 SIM2s	Doxycycline inducible expression of SIM2s.
*MCF7 SIM2l	Doxycycline inducible expression of SIM2l.
LNCaP SIM2s-HF-low	Doxycycline inducible expression of SIM2s tagged with HA-FLAG at a relatively low level. Referred to as MDA-MB-231 SIM2s-low.
LNCaP SIM2s-HF-high	Doxycycline inducible expression of SIM2s tagged with HA-FLAG at a relatively high level. Referred to as MDA-MB-231 SIM2s-high.
*PC3-AR+ SIM2s	Doxycycline inducible expression of SIM2s.
*PC3-AR+ SIM2l	Doxycycline inducible expression of SIM2l.
*DU145 SIM2s	Doxycycline inducible expression of SIM2s.
*DU145 SIM2l	Doxycycline inducible expression of SIM2l.

*Generated by Dr Adrienne Sullivan, Whitelaw Laboratory.

5.3.2   RNA sequencing

RNA sequencing is commonly used to assess gene expression and differential gene expression across sample groups. Standard laboratory processing steps include RNA extraction, mRNA enrichment or ribosomal RNA depletion, cDNA synthesis and library

preparation through the addition of adaptors and high-throughput next generation sequencing. Sequencing data is then processed through bioinformatic pipelines to perform an alignment of sequencing reads to the reference transcriptome, quantification, filtering and normalisation of reads, and statistical analysis of to determine differentially expressed genes and transcripts (DEG) between the test groups (Conesa et al., 2016; Stark, Grzelak, & Hadfield, 2019). A significant advantage of RNA sequencing over microarray is the ability to quantify expression levels of the entire transcriptome.

For the RNA sequencing detailed in this chapter extracted RNA samples were sent to the ACRF Cancer Genomics Facility (Centre for Cancer Biology, SA Pathology and University of South Australia) for RNA sequencing and bioinformatic analysis to determine differentially expressed genes between the test groups.

5.4 Results

5.4.1 Expression of SIM2 in human breast and prostate cell lines

In order to determine the appropriate cell lines for *SIM2* overexpression, the expression level of *SIM2* was analysed in the normal human breast cells MCF10A, MDA-MB-231 breast cancer cells, and prostate cancer cells LNCaP and PC3-AR+. This was assessed using three sets of primers, specific for the short isoform (*SIM2s*), long isoform (*SIM2l*) and total *SIM2*.

Total *SIM2* expression is significantly lower in the MDA-MB-231 human breast carcinoma cell line compared to the MCF10A normal human breast cells (Figure 5.2A), consistent with previously reported literature, and with reports that have demonstrated downregulation of *SIM2* in breast cancer (Gustafson et al., 2009; Kwak et al., 2007; Laffin et al., 2008; Scribner et al., 2013). This is due to a reduction in the level of *SIM2l* expression, with no significant difference seen in the expression of *SIM2s* (Figure 5.2 B,C). For the prostate cancer cells, the LNCaP cells have statistically significant less total *SIM2* expression compared to PC3-AR+ cells (Figure 5.2A), also consistent with previous reports (Lu et al., 2011). This was due to a significant

reduction in the expression of both the short and long isoforms of *SIM2* (Figure 5.2 B,C). This indicates that LNCaP cells may not have upregulation of *SIM2* expression, at least to the same extent as what has been observed in human prostate tumours and other prostate cancer cell lines. Both prostate cancer cell lines have a significantly higher level of total *SIM2* expression compared to the breast cells lines, which reflected in expression of both the short and long isoform, with PC3-AR+ cells having the highest level of expression (Figure 5.2 A-C).

Despite the fact that MDA-MB-231 cells have a reduction in the long isoform of *SIM2*, MDA-MB-231 cells were chosen to generate a *SIM2s* overexpression cell line, as the short isoform has been reported in the literature to be downregulated in human breast tumours compared to the surrounding tissue (Gustafson et al., 2009; Kwak et al., 2007; Laffin et al., 2008; Scribner et al., 2013). Given that *SIM2* expression is reported to be upregulated in prostate cancer (Arredouani et al., 2009; DeYoung et al., 2003b; Ole Johan Halvorsen et al., 2007; Lu et al., 2011) the LNCaP cell line was chosen as the prostate cancer cell line to generate a *SIM2* overexpression cell line. It was expected this this may result in *SIM2* expression levels comparable to that of the PC3-AR+ cells, mimicking *SIM2* upregulation in prostate cancer. Both isoforms have been reported to be upregulated in prostate cancer, so the short isoform was also chosen for overexpression in prostate cancer.
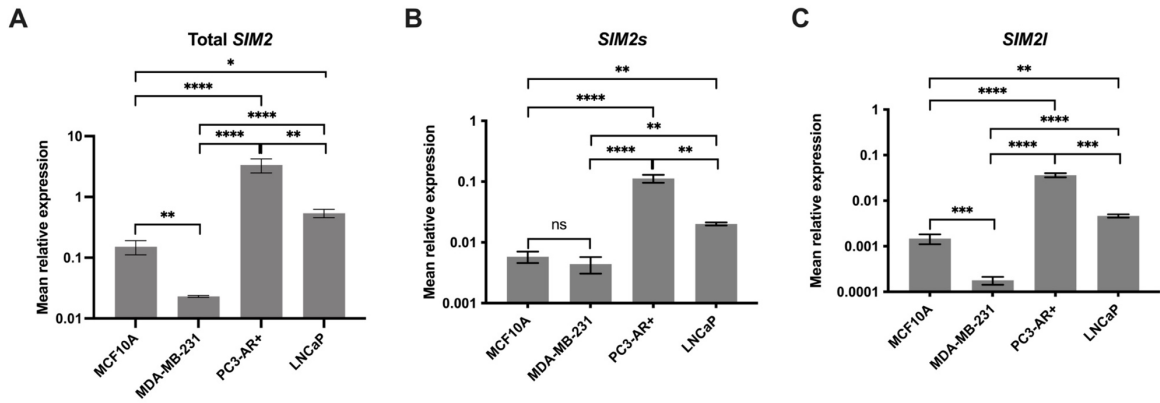
Figure 5.2: Expression of *SIM2* in normal human breast MCF10A, breast cancer cell line MDA-MB-231, and prostate cancer cell lines PC3AR+ and LNCaP. **A)** Total *SIM2* expression, **B)** *SIM2s* expression, **C)** *SIM2l* expression. Data presented as mean ± SEM for 3 biological replicates, relative to *POLR2A* (housekeeping gene) expression as an endogenous control. Y-axis presented as a log scale. Statistical significance determined by one-way ANOVA. * $P \leq 0.05$, ** $P \leq 0.01$, *** $P \leq 0.001$, **** $P \leq 0.0001$, ns, not significant.

5.4.2   Generation and characterisation of SIM2s overexpressing cell lines

MDA-MB-231 and LNCaP cells first underwent identification testing to confirm cell identity. Testing was performed by CellBank Australia using the Promega PP16HS kit. Comparison of 9 loci to the MDA-MB-231 reference sample, 88% (22/25) alleles were identical to MDA-MB-231 (ATCC: HTB-26). Greater than 80% being identical is considered a match, therefore the identity of these cells was confirmed as MDA-MB-231 (report generated by CellBank, 18/07/2017). Comparison of 9 loci to the LNCaP reference sample, 100% (32/32) alleles were identical to LNCaP clone FGC (ATCC: CRL-1740), confirming identity of these cells as LNCaP (report generated by CellBank, 18/07/2017).

MDA-MB-231 and LNCaP cells with dox inducible expression of SIM2s at both a low and high level were generated through lentiviral transduction and either blastocidin or puromycin selection using the LVTPTIP system as described in section 5.3.1 (Bersten et al., 2015). There was no observable difference in the growth or appearance of these cells compared to the parent lines.

A co-immunoprecipitation (co-IP) was performed from nuclear extracts of the MDA-MB-231 SIM2s-low, MDA-MB-231 SIM2s-high, LNCaP SIM2s-low and LNCaP SIM2s-high cell lines to confirm expression levels of SIM2s and dimerisation with ARNT (Figure 5.3). As can be seen in the 10% input, there is a much lower level of SIM2s expression in the SIM2s-low cells for both the MDA-MB-231 and LNCaP lines compared to the SIM2s- high cells, with SIM2s expression not detectable by western blot in the whole cell extract (Figure 5.3, left panel). Expression of endogenous ARNT can be seen at a similar level in both the low and high expressing MDA-MB-231 and LNCaP lines. The FLAG IP again highlights the significant difference in SIM2s expression in these lines, with only a small amount of SIM2s visible in the low expressing lines IP compared to the high expressing lines (Figure 5.3, right panel). ARNT can be seen in the FLAG IP in both the low and high expressing lines, with significantly more ARNT observed in the high expressing lines. By estimation, the majority of the ARNT in the cells appears to co-immunoprecipitate with SIM2s in the high expressing line, and a very low proportion of ARNT co-immunoprecipitated with SIM2s in the low expressing lines (Figure 5.3, ARNT band only detectable on high exposure).
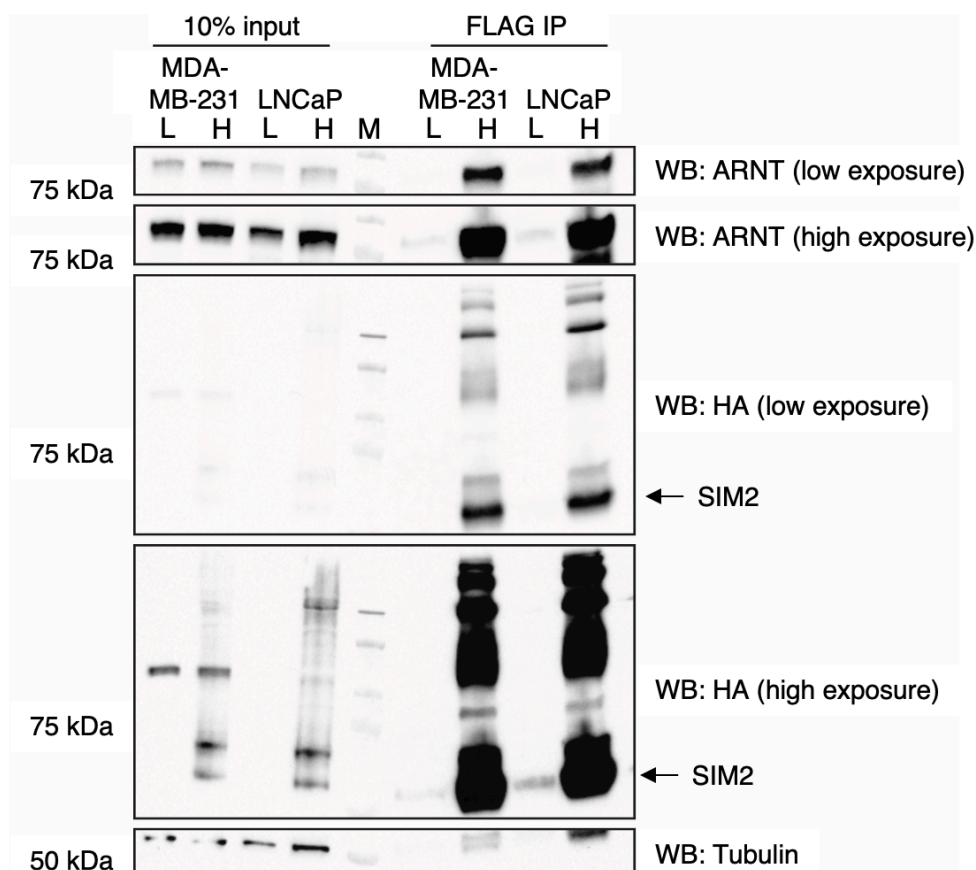
Figure 5.3: Co-immunoprecipitation to determine SIM2s levels of expression and dimerisation with ARNT in MDA-MB-231 and LNCaP SIM2s-HF- low and -high expressing cell lines. M: marker, L: SIM2s-HF-low, H: SIM2s-HF-high.

During titration of dox to determine the optimal concentration of dox required for expression of MDA-MB-231 SIM2s-HF-high, cells were grown and treated with dox for 24 hours, ranging from a final concentration of 0 μg/mL to 1 μg/mL (Figure 5.4). As shown in Figure 5.4A, no observable SIM2s-HF expression can be seen in the absence of dox induction, indicating that there is no, or minimal expression of SIM2s-HF without the addition of dox. SIM2s-HF expression can be seen after treatment with as little as 0.1 μg/mL dox, albeit at a lower level compared to treatment with higher concentrations of dox. Maximum SIM2s-HF expression is observed with 0.6 μg/mL dox treatment. This concentration was chosen to proceed with further experiments.

To determine relative levels of SIM2s-HF protein expression after dox induction, MDA-MB-231 SIM2s-HF-high cells were treated with 0.6 μg/uL dox and harvested after 0, 4, 6, 8, 16 and 20 hours. Expression of SIM2s-HF was assessed through western blot

(Figure 5.4B). There was no observable SIM2s-HF expression without dox treatment (Figure 5.4A, 0 hours). Expression of SIM2s-HF can be seen after 4 hours dox treatment, indicating that the onset of protein expression is rapid after the addition of dox. Increasing expression of SIM2s-HF can be seen up to the 16 hour time point, at which maximum expression appears to have been reached. The 6 hour time point can be used for a lower level and shorter time point of SIM2s-HF expression. Any time point after and including 16 hours would represent maximum SIM2s-HF expression post dox induction.
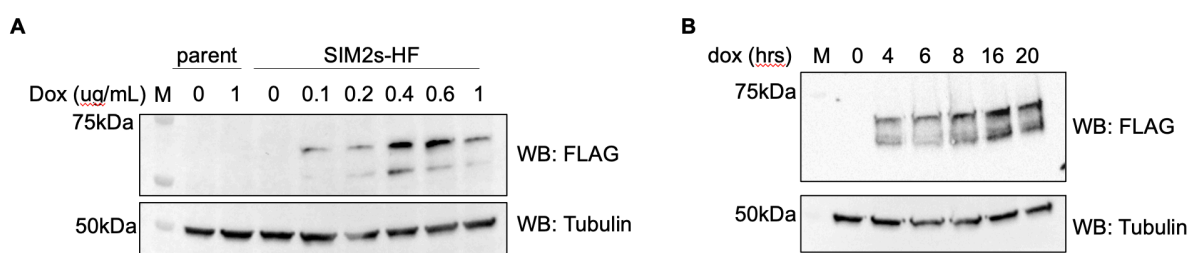

Figure 5.4: Optimisation of dox induction of SIM2s in MDA-MB-231 cells. **A)** Western blot of dox concentration optimisation in MDA-MB-231 SIM2s-HF-high expressing cells. Dox concentration shown above blot. **B)** Dox induction time point optimisation. Time of dox induction (0.6 µg/mL) shown above blot. M: marker.

To characterise these cell lines further, *SIM2* transcript expression was analysed by RT-PCR and compared to a number of cell lines; normal breast cell line MCF10A, breast cancer cell lines T-47D, MCF7 and parent MDA-MB-231, and prostate cancer cell lines PC3-AR+, DU145, and LNCaP parent (Figure 5.5). RNA extracted from two LNCaP cell lines that had inducible expression of SIM2s and SIM2l were used as controls known to have an increased level of SIM2 expression (cDNA provided by Dr Adrienne Sullivan, Whitelaw Laboratory). The primers used will detect both the long and short isoform of *SIM2*. The normal breast MCF10A and breast cancer T-47D and MCF7 cell lines all appear to have comparable *SIM2* expression, whereas the MDA-MB-231 parent cell line appears to have a lower level of *SIM2* expression. Both the MDA-MB-231 SIM2s-low and -high cell lines have significantly higher level of expression compared to the other breast cell lines. The PC3-AR+ cell line has a comparable level of *SIM2* expression to both the MDA-MB-231 SIM2s-low and -high cell lines, and to all of the LNCaP *SIM2*

overexpressing cell lines, whereas the prostate cancer DU145 and parent LNCaP cell lines have a lower level of expression.

SIM2 levels in cancer lines

RT-qPCR

Log scale

n=3
Error bars = SEM

to the low cells (Figure 5.6B). The level of *SIM2l* transcript was comparable across all three lines (Figure 5.6C), however there was a statistically significant reduction of *SIM2l* in MDA-MB-231 SIM2s-HF-high cells compared with the parent MDA-MB-231 cells. This is likely an artefact due to the manipulation of *SIM2s* expression in the MDA-MB-231 SIM2s-HF-high cell line. The difference in total *SIM2* expression between the MDA-MB-231 SIM2s-HF-low and -high cells was not found to be statistically significant

(Figure 5.6A), however there is a demonstrated a clear difference in SIM2 protein expression between the two lines as demonstrated by western blot (Figure5.3). Coupled with the statistically significant difference in *SIM2s* but not *SIM2l* expression (Figure 5.6B,C), we have demonstrated that these two cell lines have a low, and high level of *SIM2s* overexpression, respectively.
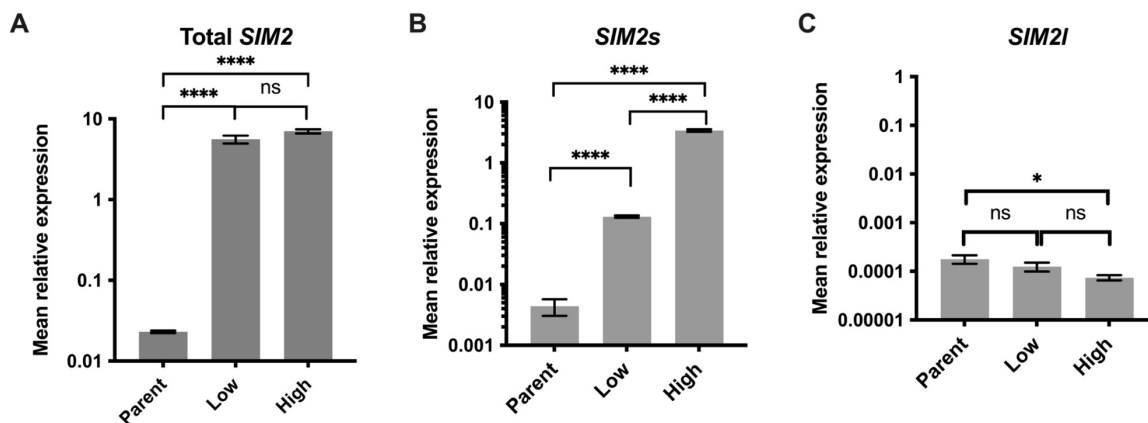
Figure 5.6: RT-qPCR analysis of *SIM2* expression in parent, SIM2s low and SIM2s high overexpressing MDA-MB-231 cells treated with 0.6 µg/mL dox for 24 hours. **A)** Total *SIM2* expression, **B)** *SIM2s* expression, **C)** *SIM2l* expression. Data presented as mean ± SEM for 3 biological replicates, relative to *POLR2A* expression. Statistical significance determined by one-way ANOVA. * P ≤ 0.05, **** P ≤ 0.0001, ns, not significant.

RT-qPCR analysis was also performed to assess *SIM2* expression in MDA-MB-231 overexpressing cells compared to the MCF10A and PC3-AR+ (Figure 5.7). This was performed using primers that will detect total *SIM2*, *SIM2s* and *SIM2l*. Both the MDA-MB-231-low and -high overexpressing cells have significantly higher *SIM2* expression compared to the normal human breast MCF10A cells (Figure 5.7A). This is due to an increase in *SIM2s* expression (Figure 5.7B). Although the observed level of *SIM2* expression in both the MDA-MB-231 SIM2s-low and -high overexpressing cells is significantly higher than that seen in the MCF10A normal breast cell line, it is not significantly different to that seen in the PC3-AR+ human prostate cancer cell line, therefore both still represent a biologically relevant level of *SIM2* expression in a cancer context (Figure 5.7A).
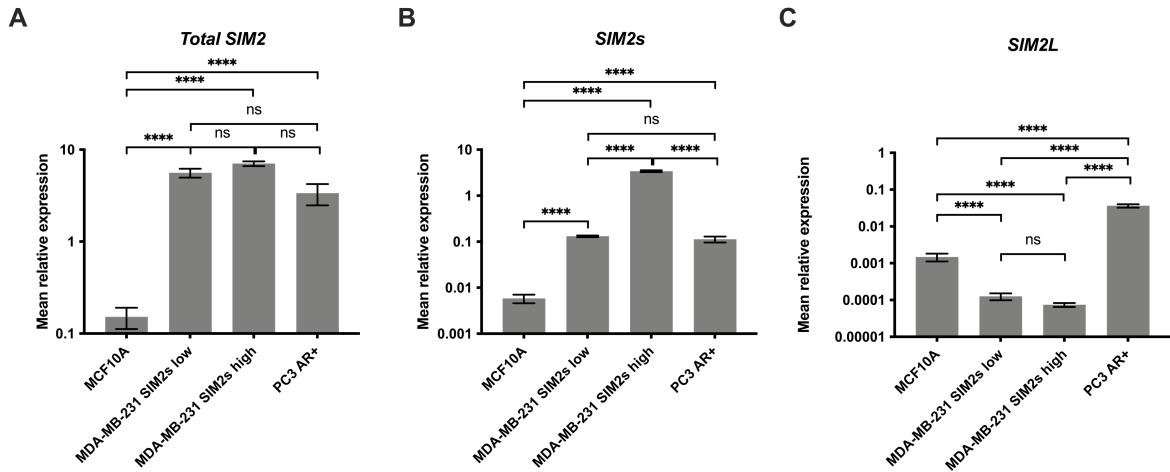
Figure 5.7: RT-qPCR analysis of *SIM2* expression in MCF10A, MDA-MB-231 SIM2s-low – and high overexpressing cells, and PC3-AR+ cells. **A)** Total *SIM2* expression, **B)** *SIM2s* expression, **C)** *SIM2l* expression. Data presented as mean ± SEM for 3 biological replicates, relative to *POLR2A* expression. Statistical significance determined by one-way ANOVA. **** $P \leq 0.0001$, ns, not significant.

Characterisation of the MDA-MB-231 SIM2s-low and -high overexpressing cells demonstrated successful generation of the cell lines with both a low and high level of overexpression of SIM2, respectively. SIM2 is able to form a heterodimer with its partner protein ARNT, therefore should be able to function as a transcription factor and regulate target gene expression. It is known that one of the mechanisms by which SIM2 acts as a transcriptional repressor is to compete with other bHLH/PAS transcription factors for dimerisation with their common partner proteins ARNT or ARNT2 (Emily L. Button et al., 2017; Farrall & Whitelaw, 2009; S. L. Woods & Whitelaw, 2002). In the *SIM2s*-high overexpression cells there would be expected to be significant competition for ARNT dimerisation, as SIM2 appears to dimerise with majority of ARNT in the cells (Figure 5.3). Competition would not be expected to be significant in the *SIM2s*-low overexpressing cells, as SIM2 is only dimerising with a very small proportion of ARNT protein in the cells (Figure 5.3), therefore there should be ample ARNT remaining for other Class I bHLH/PAS transcription factors to dimerise with. The MDA-MB-231 SIM2s-HF-high overexpressing line was chosen to perform RNA-sequencing experiments in order to give the best chance of identifying differentially expressed genes. The MDA-MB-231 SIM2s-HF-low overexpressing line can then be used to further investigate the differentially expressed genes found.

5.4.3 Discovery of SIM2s target genes by RNA sequencing

Two biological repeats of the parent MDA-MB-231 cells treated with 0.6 µg/mL dox for 24 hours, MDA-MB-231 SIM2s-high cells treated with 0.6 µg/mL dox for 6 hours and MDA-MB-231 SIM2s-high cells treated with 0.6µg/mL dox for 24 hours underwent RNA sequencing. The 6-hour time point was chosen to investigate early gene expression changes after SIM2s induction. The 24-hour time point was chosen to investigate long-term gene expression changes after induction of SIM2s expression, after maximum expression has been reached (approximately at the 16-hour time point, see Figure 5.4). The two time points of SIM2s induction were also compared in order to assess whether there are any early vs late gene expression changes, and to help highlight direct SIM2s target genes, as SIM2s upregulated/downregulated genes may influence later expression of additional genes.   Genes differentially expressed >2 fold, and <0.5 fold, and defined as being statistically significant (based on adjusted p-value for multiple comparisons), were selected for further analysis. This criteria was chosen as it would most likely represent biologically relevant and statistically significant gene expression changes.

The duplicates for the three different sample groups clustered together, and distinctly from the other sample groups as shown in the multidimensional scaling (MDS) plot in Figure 5.8A. Plots summarising the statistically significant and potentially biologically relevant differentially expressed genes is shown in Figure 5.8 B-D for the three sample group comparisons. A relatively small number of differentially expressed genes are observed as either being upregulated (green) or downregulated (red).
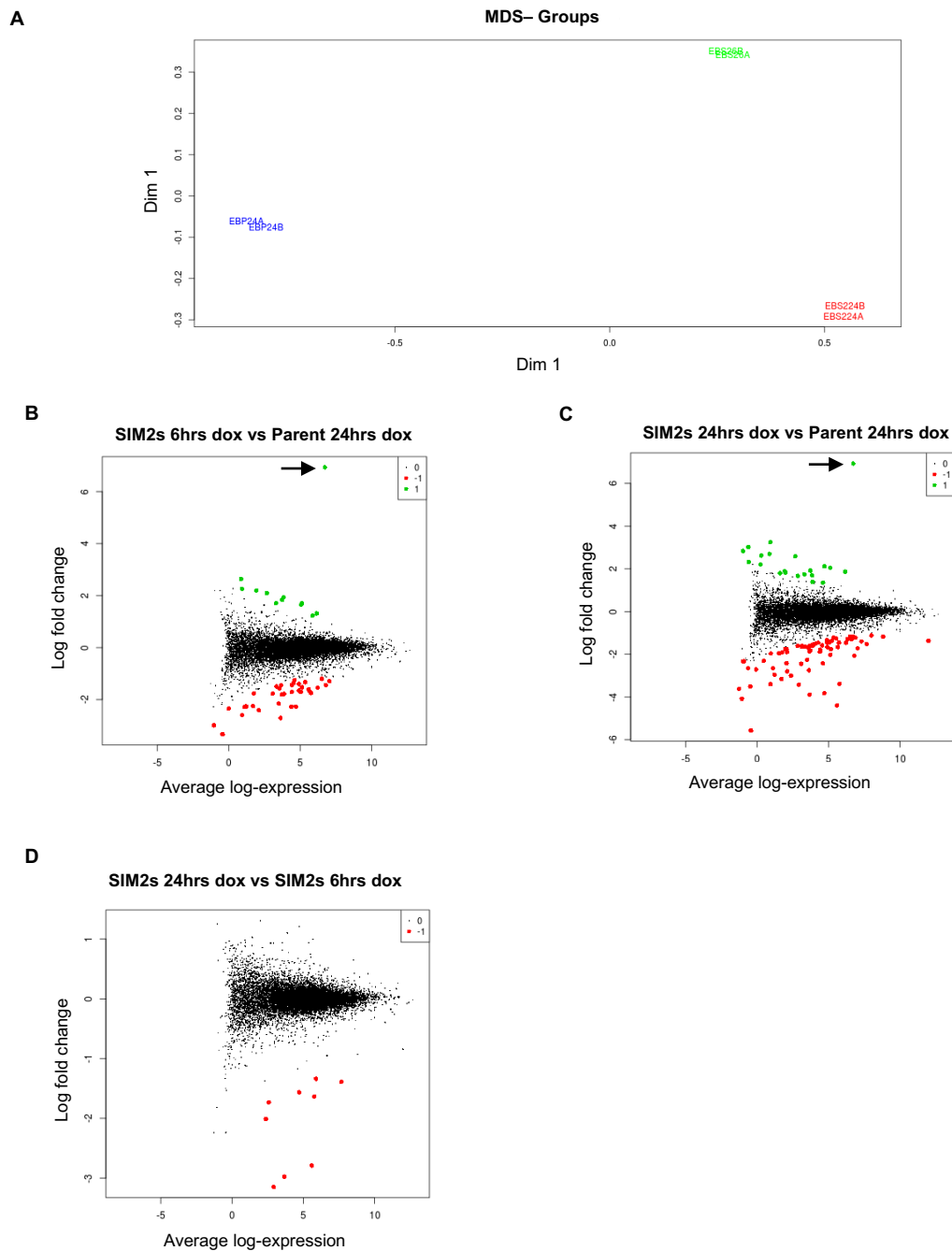
Figure 5.8: Analysis of RNA sequencing data. **A)** MDS plot showing the relationship between the samples; MDA-MB-231 parent 24 hrs dox (EBP24 replicate A and B) in blue, MDA-MB-231 SIM2s 24 hrs dox in red, MDA-MB-231 SIM2s 6 hrs dox in green. **B-D)** Plots comparing log fold change to the average log expression across the replicates for the three sample comparisons; SIM2s 6 hrs dox vs Parent 24 hrs dox **(B),** SIM2s 24 hrs dox vs parent 24 hrs dox **(C)**, and SIM2s 6 hrs dox vs SIM2s 24 hrs dox **(D)**. Significantly differentially expressed genes are shown in green (upregulated) and red (downregulated). Arrow in **B,C** points to *SIM2*. (Graphs generated as part of the initial analysis of data performed by ACRF, Dr Andreas Schreiber).

A total of 102 genes were found to be differentially expressed in the MDA-MB-231 SIM2s-high overexpression line compared to the parent MDA-MB-231 cells across both the 6 hr and 24 hr dox induction time points (Figure 5.9A). A heat map listing all the differentially expressed genes can be seen in Figure 5.10. The majority of differentially expressed genes were downregulated, with a total of 79 genes being downregulated compared to 23 genes being upregulated, indicating that SIM2s is mostly acting as a transcriptional repressor in this context. There was a total of 47 differentially expressed genes in the SIM2s-high overexpressing cells after 6hours dox induction compared to the parent cells, with 35 being downregulated and 12 being upregulated (Table 5.3). The regulation of this set of genes occurs early after SIM2s expression is induced. After 24 hours dox induction, there were a total of 94 differentially expressed genes compared to the parent cells, with 73 genes being downregulated and 21 upregulated (Table 5.4). The regulation of this set of genes would occur both early and later after SIM2s induction. There were only 9 genes that were differentially expressed after 24 hours SIM2s overexpression compared to 6 hours, all of which were downregulated (Table 5.5). The regulation of this set of genes is likely to occur later after induction of SIM2s, as their expression is significantly downregulated between these two time points of SIM2s induction.

A comparison of the number of differentially expressed genes across all conditions is shown in Figure 5.9B, with the genes and log2 fold changes listed in Tables 5.6-5.11. 6 genes were significantly differentially expressed after 6 hours of SIM2s upregulation only, 4 downregulated and 2 upregulated (Table 5.6). This set would represent the genes that are SIM2s targets immediately after expression is upregulated, with their expression beginning to return to normal between 6 and 24 hours post SIM2s induction. These genes would likely not contribute to the tumour suppressive function of SIM2, given that their expression is only transiently affected by upregulation of SIM2s. 49 genes were significantly differentially expressed at 24 hours of SIM2s upregulation only, 38 downregulated and 11 upregulated (Table 5.7). This set of genes would likely represent SIM2s target genes that are not in the immediate response but may be either early or chronically regulated genes. Only 2 genes were differentially expressed only after 24 hours of SIM2s induction compared to 6 hours of SIM2s induction, both of which were downregulated (Table 5.8). 38 genes were differentially

expressed both after 6 and 24 hours of SIM2s upregulation compared to the parent MDA-MB-231 cells, 28 downregulated and 10 upregulated (Table 5.9). This set of genes would likely be part of the early SIM2 regulated genes, and the change in their expression persisted long term. There were 4 differentially expressed genes after 24 hours of SIM2s upregulation which were also differentially expressed after 24 hours compared to 6 hours of SIM2s upregulation, all of which were downregulated (Table 5.10). These genes are likely only regulated at a later time point after SIM2s upregulation. As would be expected, there were no differentially expressed genes after 6 hours SIM2s induction commonly differentially expressed also only after 24 hours of SIM2s induction compared to the 6hr induction time point. 3 genes were differentially expressed in all three group comparisons (SIM2s 6hours vs parent, SIM2s 24 hours vs parent and SIM2s 24 hours vs SIM2s 6 hours), all of which were downregulated (Table 5.11). These genes likely form part of the early and chronic SIM2s response after upregulation.
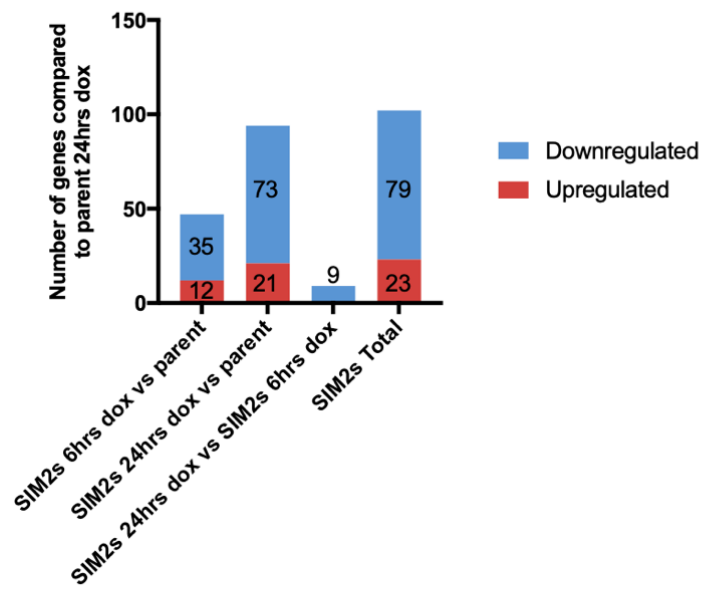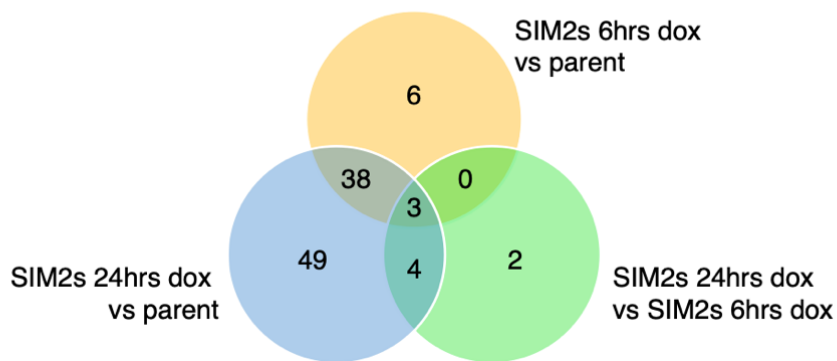
Figure 5.9: Differentially expressed genes across each comparison (SIM2s 6 hrs dox vs parent, SIM2s 24 hrs dox vs parent and SIM2s 6 hrs dox vs SIM2s 24 hrs dox). **A)** Graph depicting the number of upregulated and downregulated genes across each comparison and in total across all comparisons. **B)** Venn diagram of the number of differentially expressed genes common and unique to each of the three comparisons.
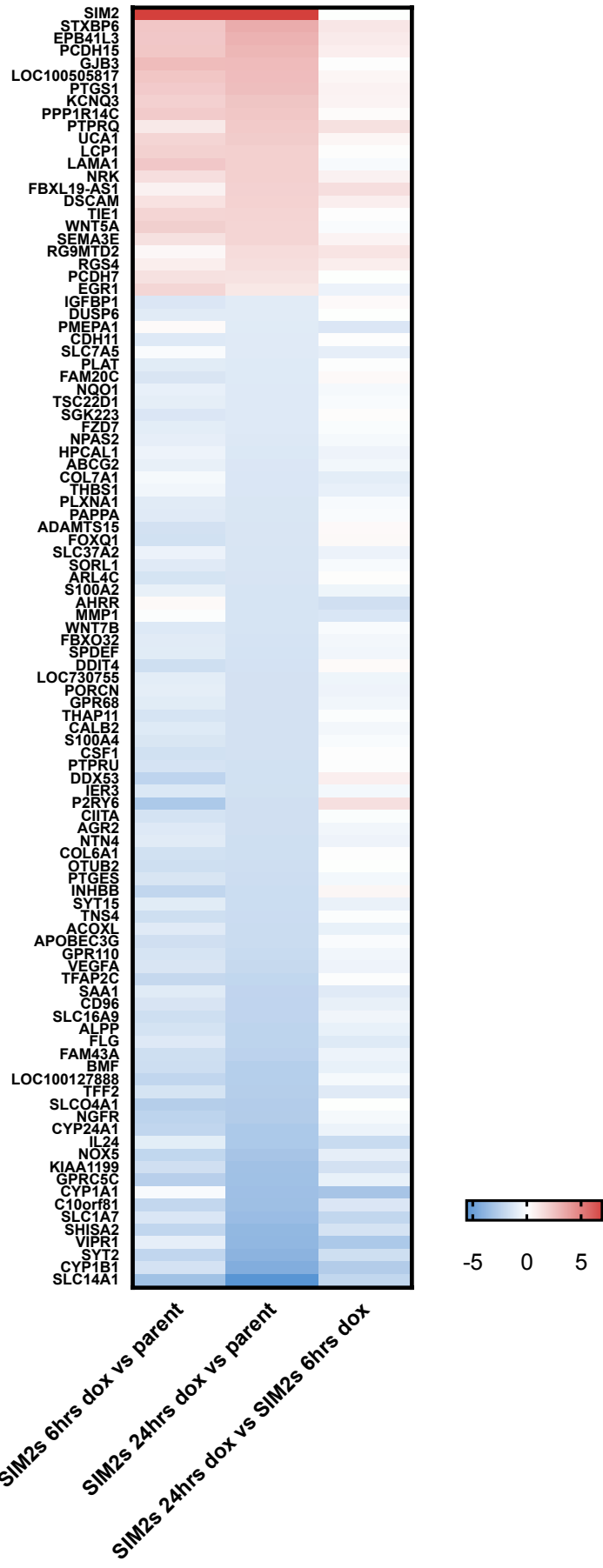
Figure 5.10: Heatmap of differentially expressed genes identified by RNA sequencing of MDA-MB-231 parent and SIM2s overexpressing lines. Comparisons performed between the parent MDA-MB-231 cells and the MDA-MB-231 SIM2s-HF-high cells after 6 (left column) and 24 hours of dox induction (middle column), and between the SIM2s-HF-high cells at the two induction time points (right column). Note that not all genes are significantly differentially expressed in all comparisons. Scale represents Log2 fold change in expression.

Approximately twice as many differentially expressed genes were observed after 24hours dox induction compared to 6hours dox induction. In many cases, the genes that were differentially expressed at 24hours and not at 6hours appeared to be changing in expression at the 6-hour time point, however this difference had not yet reached the threshold to become a statistically significant and biologically relevant gene expression change (Figure 5.10). Figure 5.11 shows examples of genes significantly differentially expressed after 24 hours of SIM2s upregulation, but not after 6 hours of SIM2s upregulation. In this example set, expression of *CYP1A1*, *IL24*, *MMP1*, *NPAS2* and *NQO1* are all decreased and expression of *EPB4IL3* is increased after 6 hours of SIM2s upregulation, however the change in expression for all genes is not significant until after 24 hours of *SIM2s* upregulation. *SIM2* was not differentially expressed between the 6-hours and 24-hours dox induction time points in the SIM2s-HF-high cells. This indicates that *SIM2s* transcript is at maximum expression after only 6 hours of dox induction, whereas SIM2 protein does not reach maximum expression until approximately 16 hours post initiation of dox induction (see Figure 5.4B).
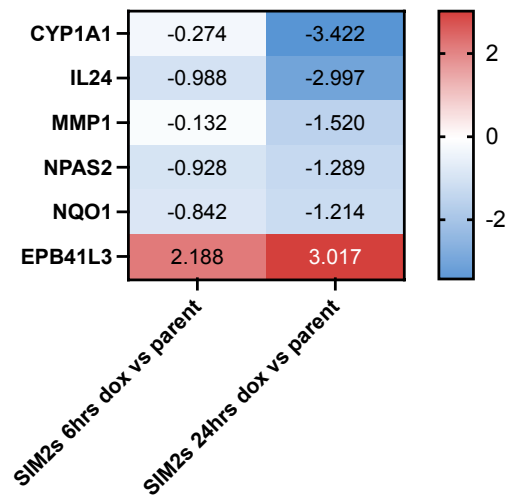
Figure 5.11: Heat map depicting example set of genes with significant differential expression after 24 hrs of *SIM2s* upregulation, but not after 6 hrs of *SIM2s* upregulation. Values and scale depict the Log2 fold change in expression for each of the genes. All *SIM2s* 6 hrs dox vs parent gene expression changes are not statistically significant. All *SIM2s* 24 hrs dox vs parent gene expression changes are statistically significant.

Overall, there were a relatively small number of differentially expressed genes identified after upregulation of SIM2s expression, particularly after 6-hours of dox induction (47 genes). This is the first analysis of transcriptome changes in response to upregulation of SIM2 expression in breast cancer cells and will provide valuable insight into the varied function of SIM2 in cancer.

5.4.4   Analysis of SIM2s target genes in MDA-MB-231 cells

A selection of differentially expressed genes were chosen for RT-qPCR confirmation, and to determine whether the genes were also differentially expressed in the MDA-MB-231 SIM2s-low overexpressing cell line. Two biological replicates were performed from cells treated with dox for 24 hours. Consistent trends in target gene expression were observed between the replicates, which were also in agreement with the RNA-sequencing data (Figure 5.12).

Erythrocyte membrane protein band 4.1 like 3 (EPB41L3, also known as 4.1B, DAL-1) is a membrane skeletal adapter protein which functions in connecting the plasma

membrane and the cytoskeleton and is known to have roles in cell motility and adhesion. *EPB41L3* is reported to have anti-tumourigenic functions, with downregulation of expression observed in multiple cancer types (including breast cancer) associated with enhanced tumourigenesis (Z. Wang et al., 2014). Expression of *EPB41L3* was significantly upregulated after 24 hours of SIM2s overexpression (3.017 log2 fold change, Table 5.4). Upregulation of *EPB41L3* by SIM2s would be expected to be anti-tumourigenic. Consistent with the RNA sequencing data, *EBP41L3* expression was found to be upregulated in both the MDA-MB-231 SIM2s-low and -high overexpressing cells, compared to the parent MDA-MB-231 cells (Figure 5.12 A,B).

Vascular endothelial-derived growth factor A (VEGFA or VEGF) is a member of the platelet-derived growth factor (PDGF) family that has well defined roles in the formation of new blood vessels, known as angiogenesis. *VEGFA* expression is known to be upregulated in most human tumours and plays a key role in tumour growth and metastasis through the promotion of angiogenesis to maintain a supply of oxygen. Upregulation of *VEGFA* in human cancers is correlated with increased tumour invasiveness and metastasis, and poorer prognosis. Expression of *VEGFA* is regulated by factors including fibroblast growth factor (FGF), epidermal growth factor (EGF), tumour necrosis factor (TNF), and under hypoxic conditions is upregulated by the bHLH/PAS transcription factor HIF1α (Apte, Chen, & Ferrara, 2019; Carmeliet, 2005; Claesson-Welsh & Welsh, 2013; G. L. Semenza, 2003). Expression of *VEGFA* was downregulated after both 6 and 24 hours of SIM2s overexpression (-1.44 and -2.064 log2 fold change, respectively) in MDA-MB-231 cells (Table 5.3, Table 5.4). Downregulation of *VEGFA* by SIM2s would be expected to be anti-tumourigenic by inhibiting tumour vascularisation, limiting oxygen supply. Consistent with RNA sequencing data, expression of *VEGFA* was downregulated in both the MDA-MB-231 SIM2s-low and -high overexpressing cells, compared to the parent MDA-MB-231 cells (Figure 5.12 C,D).

Immediate early response 3 (IER3, also known as IEX-1) is a cellular stress induced gene with functions in the regulation of apoptosis, cell proliferation, differentiation and metabolism. IER3 function in cancer is complex and reported to have both pro-and anti-tumorigenic actions, being able to both promote apoptosis and protect cells

against apoptosis. Increased expression of *IER3* is correlated with a good prognosis in some cancer types, including pancreatic and ovarian cancer, and a poor prognosis in other cancer types, including breast cancer, myeloma and acute myeloid leukaemia (Arlt & Schäfer, 2011; M. X. Wu, 2003; M. X. Wu, Ustyugova, Han, & Akilov, 2013). Expression of *IER3* was found to be downregulated after both 6 and 24 hours of SIM2s overexpression in MDA-MB-231 cells (-1.293 and -1.716 log2 fold change, respectively) (Table 5.3, Table 5.4). Downregulation of *IER3* by SIM2s in breast cancer would be expected to be anti-tumourigenic, through the loss of protection from apoptosis. Consistent with RNA sequencing data, expression of *IER3* was downregulated in both the MDA-MB-231 SIM2s-low and -high overexpressing cells, compared to the parent MDA-MB-231 cells (Figure 5.12 E,F).

DNA-damage-inducible transcript 4 (*DDIT4*, also known as *REDD1* or *RTP801*) is a stress response protein that is induced under cellular stress conditions including hypoxia energy deficiency, DNA damage and stress hormones. It's primary reported function is to inhibit mTOR through the stabilisation of the TSC1-TSC2 inhibitory complex, however it has also been shown to have roles in promoting cell proliferation and inhibition of apoptosis (Pinto et al., 2017; Sofer, Lei, Johannessen, & Ellisen, 2005; Zhidkova et al., 2022). *DDIT4* has been suggested as a prognostic biomarker for a number of cancer types, with high expression of *DDIT4* correlating with unfavourable prognostic outcomes in cancers including breast cancer, AML, glioblastoma and ovarian cancer (Pinto et al., 2017). *DDIT4* was also found to be a potential AHR target gene in breast cancer, with expression of *DDIT4* downregulated after *AhR* knockdown in MDA-MB-231 cells, implicating *DDIT4* in both the xenobiotic and hypoxic stress responses (Goode, Pratap, & Eltom, 2014). Expression of *DDIT4* was found to be downregulated after both 6 and 24 hours of SIM2s overexpression in MDA-MB-231 cells (-1.784 and -1.604 log2 fold change, respectively) (Table 5.3, Table 5.4). Downregulation of *DDIT4* by SIM2s in breast cancer would be expected to be anti-tumourigenic, given that high expression of *DDIT4* correlates with a poorer prognosis in breast cancer (Pinto et al., 2017). Presumably this would be through preventing the role of *DDIT4* in cell proliferation and inhibition of apoptosis. Consistent with RNA sequencing data, expression of *DDIT4* was downregulated in both the MDA-MB-231

SIM2s-low and -high overexpressing cells, compared to the parent MDA-MB-231 cells (Figure 5.12 G,H).

Cytochrome P450 family 1 subfamily A member 1 (*CYP1A1*) is a member of the cytochrome P450 family of enzymes known to function in the metabolism of xenobiotics and estrogen. Importantly, *CYP1A1* is one of the classic AHR target genes induced for phase I metabolism of xenobiotics (Bersten et al., 2013; N. Hao & Whitelaw, 2013; Tsuchiya, Nakajima, & Yokoi, 2005). Expression of *CYP1A1* has been shown to be upregulated in breast cancer with high expression correlating with tumour grade (Sneha et al., 2021; Tsuchiya et al., 2005). CYP1A1 functions in breast cancer cell proliferation and survival through the suppression of AMPK signalling. Knockdown of *CYP1A1* in MCF7 and MDA-MB-231 breast cancer cells reduces cell proliferation and survival (Rodriguez & Potter, 2013). In addition, *CYP1A1* has been shown to be overexpressed in anti-estrogen treatment resistant breast cancers (Sneha et al., 2021). Expression of *CYP1A1* was significantly downregulated after 24 hours of SIM2s overexpression in MDA-MB-231 cells (-3.422 log2 fold change), and it was also significantly downregulated after 24 hours of SIM2s overexpression compared to 6 hours (-3.148 log2 fold change) indicating that it may be a late regulated gene (Table 5.4, Table 5.5). Downregulation of *CYP1A1* by SIM2s in breast cancer would support the anti-tumourigenic function of SIM2s through preventing the cell proliferation and survival roles of *CYP1A1*. Consistent with RNA sequencing data, expression of *CYP1A1* was downregulated in both the MDA-MB-231 SIM2s-low and -high overexpressing cells, compared to the parent MDA-MB-231 cells (Figure 5.12 I, J).

Interleukin 24 (IL24) is a member of the IL-10 cytokine family and is reported to have anti-tumorigenic functions in many cancer types including ovarian, pancreatic, colon, prostate, and breast cancer, with downregulation of *IL24* expression in breast cancer correlating with poor prognosis (Menezes et al., 2014; Whitaker, Filippov, & Duerksen-Hughes, 2012). Expression of *IL24* was significantly downregulated after 24 hours of SIM2s overexpression in MDA-MB-231 cells (-2.997 log2 fold change), and it was also significantly downregulated after 24 hours of SIM2s overexpression compared to 6 hours (-2.009 log2 fold change) indicating that it may be a late regulated gene (Table 5.4, Table 5.5). Conversely to the above described genes, downregulation of *IL24* would

be expected to be pro-tumourigenic. Consistent with RNA sequencing data, expression of *IL24* was downregulated in both the MDA-MB-231 SIM2s-low and -high overexpressing cells, compared to the parent MDA-MB-231 cells (Figure 5.12 K,L).

All genes tested were similarly regulated in both the MDA-MB-231 SIM2s-low and -high overexpressing cells, indicating that regulation of these genes by SIM2 is not an artefact of the high level of overexpression. Therefore, it would be expected that most, if not all target genes found to be differentially expressed with overexpression of *SIM2s* at a high level would also be differentially expressed with a low level of *SIM2*s overexpression.
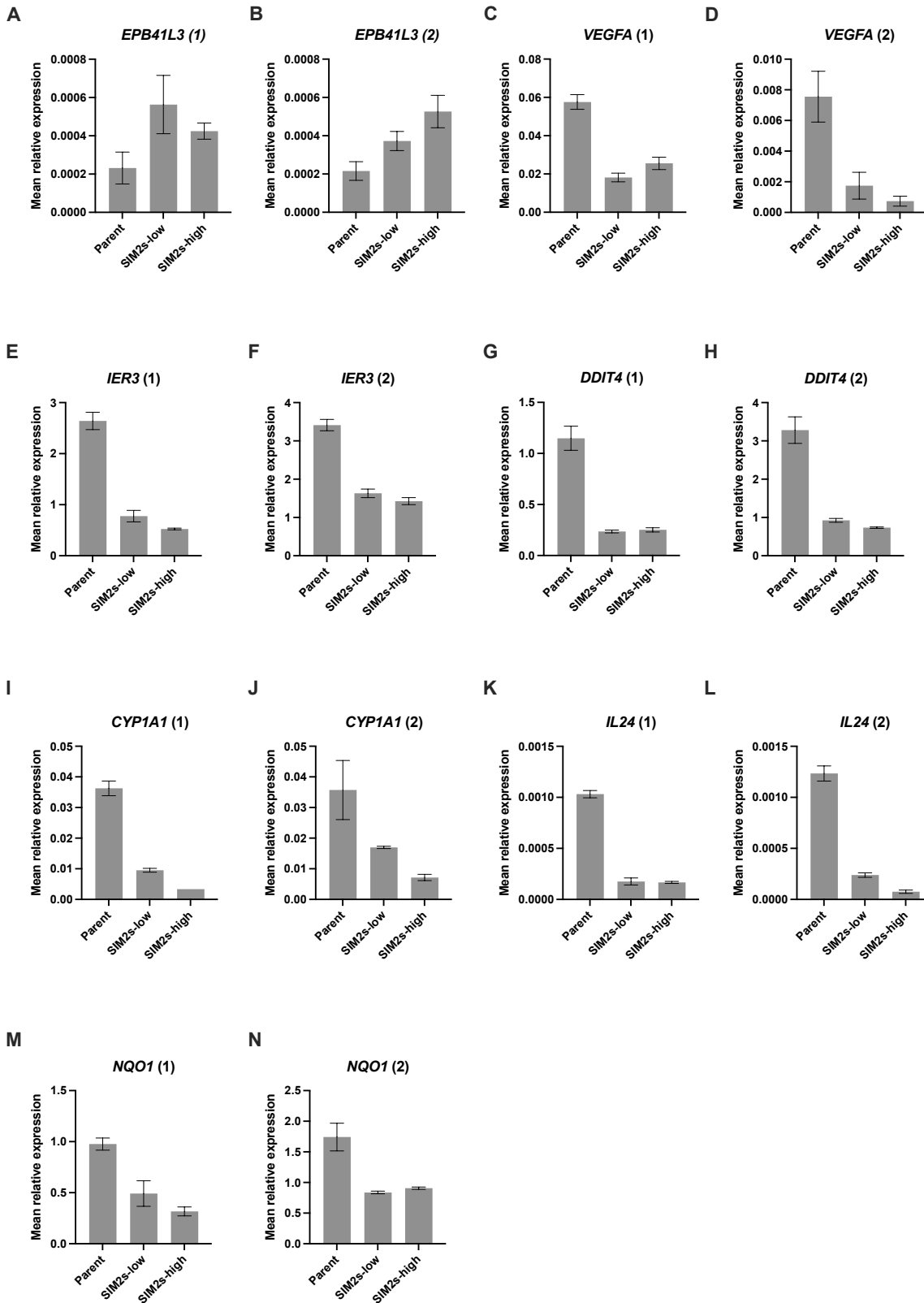
Figure 5.12 Target gene confirmation was performed by RT-qPCR comparing MDA-MB-231 parent, SIM2s-low and -high overexpressing cells treated with dox for 24 hours. Two biological replicates for each gene were performed and are presented individually. **A-B)** *EBP41L3*. **C-D)** *VEGFA*. **E-F)** *IER3*. **G-H)** *DDIT4*. **I-J** *CYP1A1.* **K-L)** *IL24*.

M-N) *NQO1*. Data presented as mean ± SD for 3 technical replicates, relative to *POLR2A* expression.

NAD(P)H quinone dehydrogenase 1 (NQO1) is a member of the quinone family of enzymes and is an FAD (flavin adenine dinucleotide) binding protein that functions in reducing quinones to hydroquinones (Oh & Park, 2015; Preethi, Arthiga, Patil, Spandana, & Jain, 2022). *NQO1* has known associations with the bHLH/PAS family of transcription factors, being a classic target gene of AHR, and has been shown to increase the stability of HIF1α in cancer cells (N. Hao et al., 2012; N. Hao & Whitelaw, 2013; Oh & Park, 2015; Okey et al., 2005). High expression of *NQO1* is reported to correlate with poor prognosis in a number of different cancer types, including breast, colon, lung, cervical and pancreatic cancers (Oh & Park, 2015; Preethi et al., 2022; Y. Yang et al., 2014; Y. Yang et al., 2019). Overexpression of NQO1 increases tumourigenic properties both *in vitro* and in an *in vivo* mouse xenograft model (Yang et al 2019). In addition, *NQO1* overexpression promotes increased glycolysis through maintaining NADPH homeostasis(Y. Yang et al., 2019). Increased glycolytic metabolism (known as the Warburg effect) is important for tumour progression. Expression of *NQO1* was significantly downregulated after 24 hours of SIM2s overexpression (-1.214 log2 fold change, Table 5.4). Downregulation of *NQO1* by SIM2s in breast cancer would be expected to be anti-tumourigenic, and interestingly, would result in interference with both the ARH and HIF1α signalling pathways. Consistent with RNA sequencing data, expression of *NQO1* was downregulated in both the MDA-MB-231 SIM2s-low and -high overexpressing cells, compared to the parent MDA-MB-231 cells (Figure 5.12 M,N).

Expression of *NQO1* in various human cancer cell lines with and without *SIM2* upregulation was further investigated, to determine if modulation of *NQO1* expression could be a mechanism by which downregulation of *SIM2* in breast cancer is favourable for tumour progression. The cell lines available for analysis that were already generated by the Whitelaw laboratory included the MDA-MB-231 SIM2s-low and -high overexpressing cells, and LNCaP SIM2s-low and -high overexpressing cells (generated as part of this project). In addition, the following cell lines had previously been generated by the Whitelaw laboratory; human breast cancer cell lines MCF7 and MDA-MB-453 cells, and human prostate cancer cell lines PC3-AR+ and DU145 cells with

inducible expression of *SIM2s* and *SIM2l* (Table 5.2) (Dr Adrienne Sullivan, University of Adelaide) (A. E. Sullivan, 2016).

RT-qPCR for *NQO1* was performed on cDNA from the above described cell lines. All three of the breast cancer cell lines showed a reduction of *NQO1* expression with overexpression of *SIM2s* or *SIM2l* (MCF7 and MDA-MB-453 cells only) compared to the respective parent cells (Figure 5.13 A, B, C), further supporting the finding that *NQO1* is a target gene of *SIM2* in a breast cancer context. *NQO1* also does not appear to be an isoform specific SIM2 target gene as overexpression of both the short and long isoform in MCF7 and MDA-MB-453 cells results in a reduction in *NQO1* expression (Figure 5.13 B,C). DU145 cells were the only prostate cancer cell line that showed a reduction in *NQO1* expression with overexpression of *SIM2s* or *SIM2l* compared to the parent cells (Figure 5.13F). The LNCaP cells overexpressing *SIM2s* and the PC3-AR+ cells overexpressing *SIM2s* or *SIM2l* did not appear to have a change in the level of *NQO1* expression compared to the respective parent cells (Figure 5.13 D, E), indicating that *NQO1* may be more likely to be a breast cancer specific SIM2 target gene. LNCaP and DU145 cells have a relatively similar level of *SIM2* expression (Figure 5.5), therefore it is unlikely the SIM2 protein already present in LNCaP cells is sufficient to downregulate *NQO1* expression to an extent that overexpression of SIM2 would not have any further effect on *NQO1* expression levels. There may be specific cofactors or expression of other genes present in DU145 cells that allows SIM2 to regulate *NQO1* expression.

The implication that *NQO1* is a SIM2 target gene in breast cancer, but not commonly in prostate cancer supports the proposal that downregulation of *NQO1* by SIM2 in breast cancer could be a mechanism by which SIM2 is able to reduce tumourigeneis. Therefore, downregulation of *SIM2* expression in breast cancer allows for high expression of *NQO1*, promoting enhanced tumourigenesis.
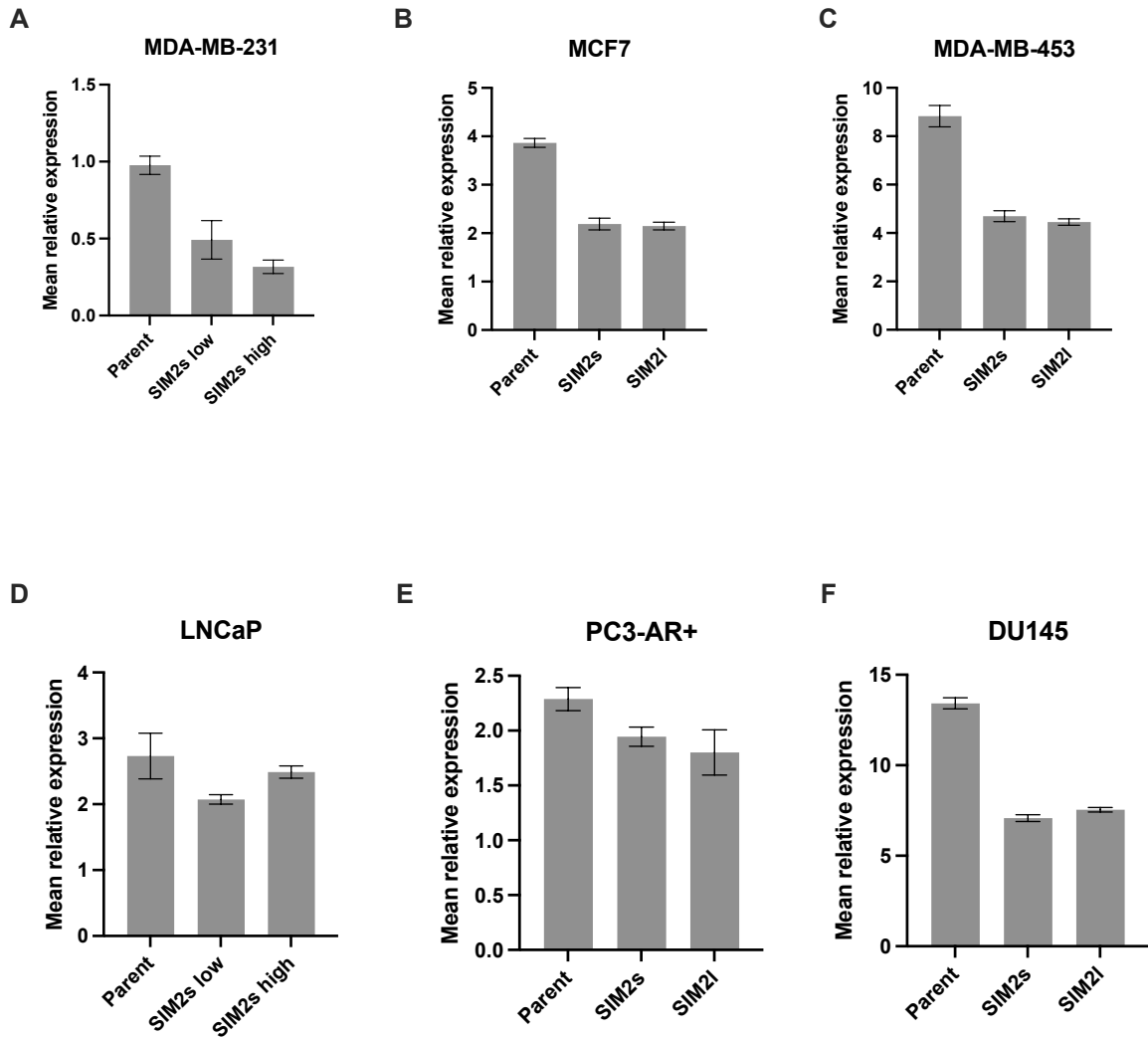
Figure 5.13: RT-qPCR analysis of *NQO1* expression in human breast and prostate cancer cell lines with overexpression of SIM2. **A)** MDA-MB-MDA-MB-231 parent, SIM2s-low and -high overexpressing cells. **B)** MCF7 parent, SIM2s and SIM2l overexpressing cells. **C)** MDA-MB-453 parent, SIM2s and SIM2l overexpressing cells. **D)** LNCaP parent, SIM2s-low and -high overexpressing cells. **E)** PC3-AR+ parent, SIM2s and SIM2l overexpressing cells. **F)** DU145 parent, SIM2s and SIM2l overexpressing cells. Data presented as mean ± SD for 3 technical replicates, relative to *POLR2A* expression.

5.4.5   SIM2 target gene dataset comparisons

Analysis of gene expression changes resulting from manipulation of *SIM2* expression has been performed previously in other cancer types. To further investigate how *SIM2*

has opposing actions in different cancer types, a comparison was performed between the MDA-MB-231 SIM2s overexpression RNA sequencing dataset, and the following available microarray datasets; overexpression of SIM2s and SIM2l in HT-1080 human fibrosarcoma cells and LNCaP prostate cancer cells (unpublished data, Whitelaw Laboratory) (A. E. Sullivan, 2016), and knockdown of *SIM2* in PC3 prostate cancer cells (Lu et al., 2011).

The HT1080 and LNCaP SIM2 microarray included overexpression of SIM2s and SIM2l at two different time points (4 hours and 24 hours). Combining all differentially expressed genes across the HT1080 dataset and for the MDA-MB-231 SIM2s overexpression at both the 6 and 24 hour time points compared to the parent MDA-MB-231 cells, there are a total of 17 commonly differentially expressed genes (Figure 5.14A). 3 genes are commonly upregulated (one of these being *SIM2*), 12 genes are commonly downregulated, and 2 genes are oppositely regulated (upregulated in one cell type and downregulated in the other) (Figure 5.14 A,B). Comparison between the LNCaP and MDA-MB-231 differentially expressed genes identified that there were only three common genes. One gene was commonly downregulated and two commonly upregulated (one of these being *SIM2*) (Figure 5.14 A,B). Comparison between the MDA-MB-231 SIM2s overexpression differentially expressed genes and PC3 knockdown differentially expressed genes did not find any genes common to the two datasets (data not shown). However, there was a similar proportion of genes that are downregulated by SIM2 expression across both cell types, with approximately 78% of genes downregulated with SIM2s overexpression in MDA-MB-231 cells and approximately 82% of genes upregulated with *SIM2* knockdown in PC3 cells (Lu et al., 2011).

This analysis provides further support for SIM2 function being highly context dependent, with a strikingly low number of commonly regulated genes observed across the dataset comparisons. The highest number of commonly regulated genes was found between the MDA-MB-231 and HT1080 datasets. However, there are two genes, *WNT5A* and *SAA1* that were differentially regulated, with *WNT5A* being upregulated in MDA-MB-231 cells and downregulated HT1080 cells, and *SAA1* being downregulated in MDA-MB-231 cells and upregulated in HT1080 cells, highlighting that the

mechanisms of SIM2 target gene regulation may be highly variable between the two cell types.

Only two genes (not including *SIM2*) are commonly regulated by SIM2 in MDA-MB-231 and LNCaP cells, *EBP41L3* which was upregulated, and *IER3* which was downregulated. Of note, these two genes are also regulated commonly regulated in the HT1080 cells, therefore these genes may be common SIM2 target genes across many tissue or cell types. The lack of commonly regulated genes between the breast cancer MDA-MB-231 cells and prostate cancer (LNCaP and PC3) cells indicates that tissue specific SIM2 target genes is likely the mechanism by which downregulation of SIM2 in breast cancer and upregulation of SIM2 in prostate cancer is reported to have pro-tumourigenic effects.
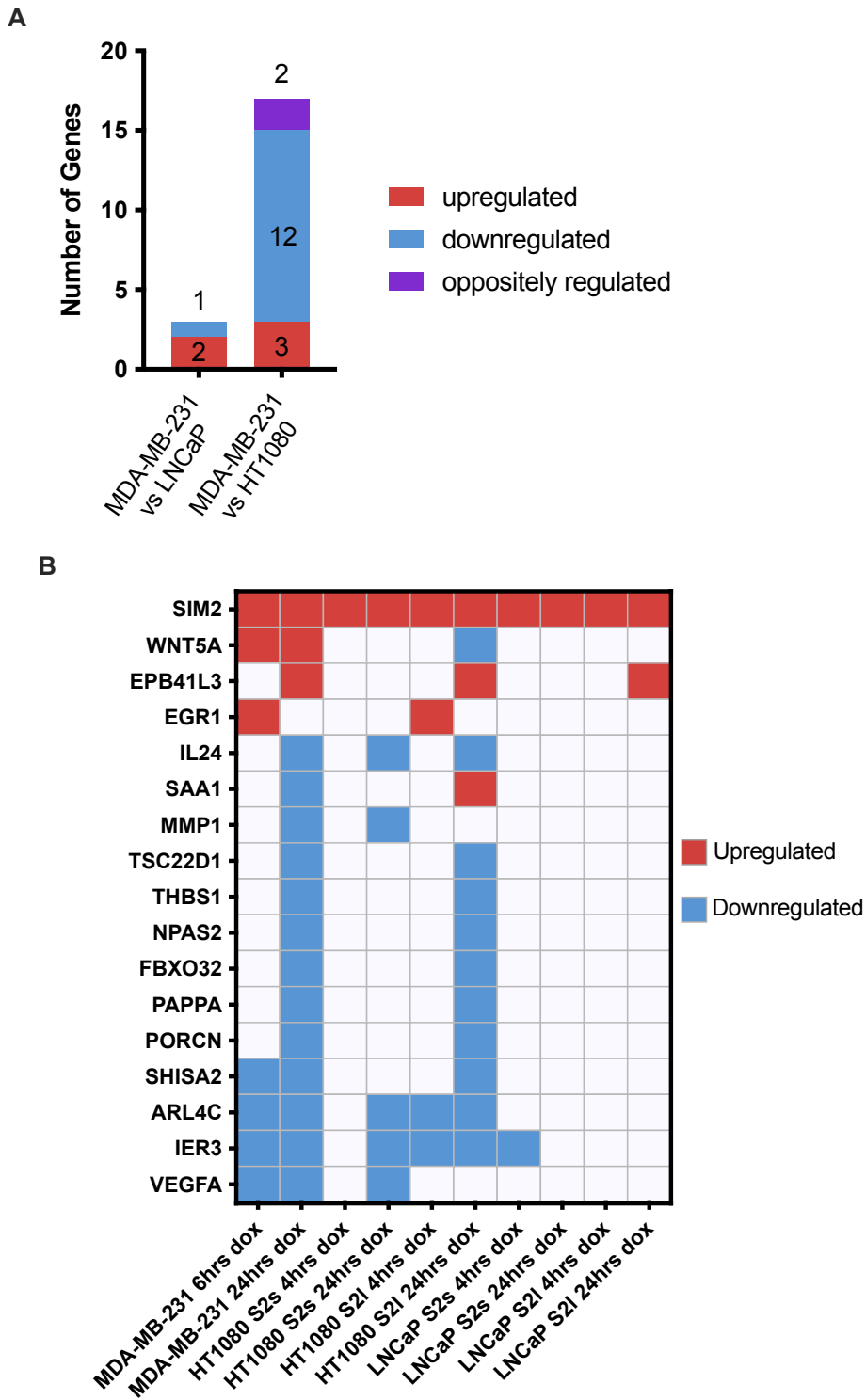
Figure 5.14: Comparison of SIM2 regulated genes identified across different datasets; MDA-MB-231 SIM2s overexpression, and HT-1080 and LNCaP SIM2s (S2s) and SIM2l (S2l) overexpression. **A)** Graph displaying the number of differentially expressed genes in common with the MDA-MB-231 dataset and the LNCaP and HT-1080 datasets and whether they were commonly upregulated, downregulated or oppositely regulated. **B)** Graph of differentially expressed in common between the MDA-MB-231 dataset and

the LNCaP and HT-1080 datasets, displaying which genes were up- or downregulated under each condition.

5.4.6 Crosstalk between SIM2s and AHR in breast cancer cells

AHR is known to have both pro- and anti-tumourigenic functions in human cancer. There are two main mechanisms through which AHR function promotes tumourigenesis. The first is through upregulation of enzymes that function in xenobiotic metabolism (including the cytochrome P450 enzymes). These enzymes produce carcinogenic side products, resulting in enhanced tumourigenesis. The second is through AHR regulation of genes that function in cellular proliferation and the immune system. Consistent with the pro-tumourigenic role of AHR, transgenic mice that express constitutively active AHR develop stomach and liver tumours (Moennikes et al., 2004). Conversely to this, *AhR* knockout mice develop gastrointestinal tumours, demonstrating a tumour suppressor role for AHR.

Even within the same cancer type AHR has been demonstrated to have both pro- and anti-tumourigenic functions, including in breast cancer. *AHR* has been reported to have increased expression, and constitutive activation in breast cancer and cancer cell lines, with high expression correlating with a high tumour malignancy grade (Powell, Goode, & Eltom, 2013; Schlezinger et al., 2006). Knockdown of *AHR* in MDA-MB-231 breast cancer cells was shown to result in decreased tumourigenic properties both *in vitro* and *in vivo*. Cells with *AHR* knockdown displayed reduced proliferation, growth and migration, and increased apoptosis, while these cells had reduced tumour growth and lung metastasis in a mouse xenograft model (Goode et al., 2013). Gene expression analysis of *AHR* knockdown MDA-MB-231 cells showed differential expression in genes known to function in cell growth, survival, migration and invasion, supporting a pro-tumourigenic role for AHR in breast cancer (Goode et al., 2014).

As discussed, (see section 5.4.4) the classic AHR target genes *CYP1A1* and *NQO1* were significantly downregulated with SIM2s overexpression. In addition, the classic AHR target genes *CYP1B1* and *AHRR* were also found to be significantly downregulated by SIM2s (Tables 5.3, 5.4 and 5.5) (Bersten et al., 2013). *CYP1B1* was significantly

downregulated after both 6 and 24 hours of SIM2s overexpression (-1.604 and -4.394 log2 fold change, respectively), being the second most highly downregulated gene at 24 hours. Expression of *CYP1B1* was even further downregulated between 6 and 24 hours (-2.17 log2 fold change), indicating that *CYP1B1* is strongly downregulated by overexpression of SIM2s both initially, and expression continued to be repressed over time. Expression of *AHRR* was significantly downregulated after 24 hours of SIM2s overexpression compared to 6 hours (-1.731 log2 fold change), indicating that expression changes later after SIM2s induction.

To investigate whether SIM2s is interfering with regulation of AHR target genes in MDA-MB-231 cells, a comparison was performed between the MDA-MB-231 SIM2s overexpression differentially expressed genes identified by RNA sequencing and a published dataset of differentially expressed genes identified by microarray in MDA-MB-231 cells with knockdown of *AHR* expression (Goode et al., 2014). If there is crosstalk between SIM2 and AHR resulting in an impact on AHR target gene regulation, it would be expected that AHR target genes would be differentially regulated after AHR knockdown and with SIM2s overexpression.

This analysis found that there were 14 differentially expressed genes common to both datasets, supporting the proposal that SIM2 is interfering with AHR target gene regulation in breast cancer cells. (Figure 5.15A). 3 genes (*PCDH15*, *LCP1*, and *RGS4*) were commonly upregulated, 9 genes (*SORL1, DDIT4, GPR68, S100A4, CSF1, PTPRU, GPR110, VEGFA,* and *SLCO4A1*) were commonly downregulated, and 2 genes (*MMP1* and *CYP24A1*) were oppositely regulated, being downregulated by SIM2s overexpression and upregulated by AHR knockdown (Figure 5.15 B,C)
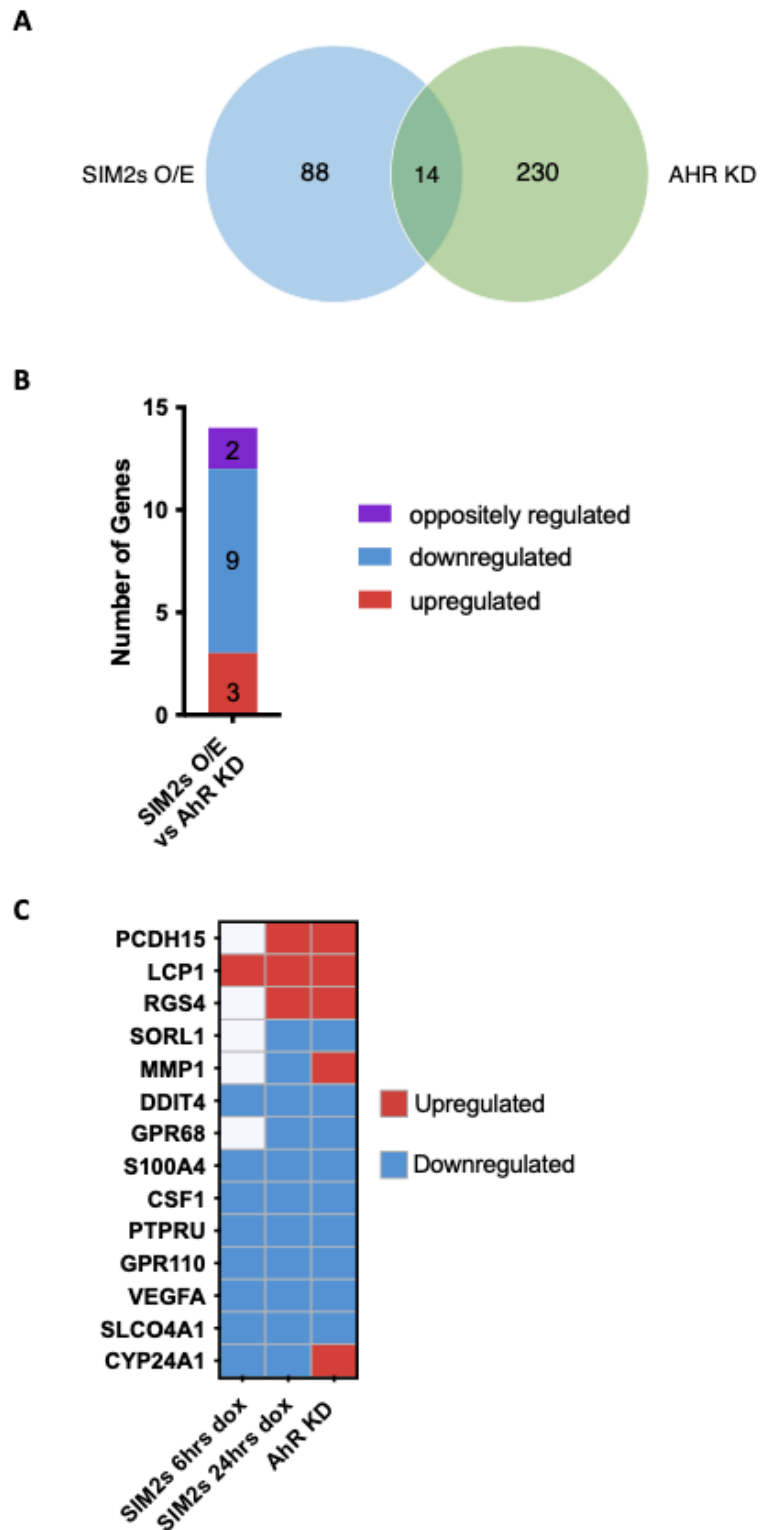
Figure 5.15: Comparison of MDA-MB-231 *SIM2s* overexpression RNA-seq data with AhR KD microarray data. **A)** Venn diagram comparing differentially expressed genes across the MDA-MB-231 SIM2s overexpression RNA sequencing experiment and AhR knockdown in MDA-MB-231 cells. **B)** Graph displaying the number of differentially expressed genes in common with the SIM2s overexpression and AHR KD datasets and

whether they were commonly upregulated, downregulated or oppositely regulated. **C)** Graph of differentially expressed in common between the SIM2s overexpression and AHR KD datasets, displaying which genes were up- or downregulated under each condition.

This analysis has identified the first demonstration of crosstalk between SIM2s and the bHLH/PAS transcription factor AHR in breast cancer cells. A relatively large proportion of the SIM2s target genes in MDA-MB-231 cells overlap with AHR target genes in the same cell type (approximately 14%). SIM2s interference with AHR signalling in breast cancer would have the potential to be tumour suppressive, when AHR is functioning in a pro-tumourigenic manner.

5.5 Discussion and Future Directions

5.5.1 Expression of SIM2 in cancer cell lines and generation of cell line models

The overarching aim of this chapter was to further investigate the molecular functions of SIM2 in human cancer. In order to explore this, cell lines with doxycycline inducible expression of SIM2s at both a high and low level were generated with tight control of SIM2s under the TET-ON 3G system, which was confirmed to have no detectable expression of SIM2s in the absence of dox (Figure 5.3). These cell line models will serve as valuable tools for future investigations stemming from the findings of this chapter.

While the characterisation of these cell line models was limited to gene expression changes, additional assessment could be performed to interrogate the effect of overexpression of SIM2s in these cells. The growth and invasiveness of these cells could be studies in xenograft mouse experiments by assessing tumour growth and metastasis. It would be expected that the MDA-MB-MDA-MB-231 breast cancer cells overexpressing SIM2s would have reduced tumour burden, with the converse being true in LNCaP prostate cancer cells overexpressing SIM2s. It would also be interesting to assess the metabolic functions of these cells with ectopic SIM2s expression, as it has been previously found that knockdown of SIM2 expression in prostate cancer leads to dysregulation of a number of genes in metabolic pathways (Lu et al., 2011). These

studies would provide a useful baseline reading for the impact of SIM2s overexpression in MDA-MB-231 and LNCaP cells. From here, these cell line models would be ideal for assessing the potential of any small molecule drugs that could modulate activity of SIM2. A drug that could upregulate SIM2 expression or stabilise SIM2 protein could offer a potential therapeutic option for breast cancers where SIM2 expression is downregulated, while a small molecule inhibitor of SIM2 function could serve as a therapeutic for cancers where SIM2 expression is aberrantly upregulated.

5.5.2 RNA sequencing identified SIM2s target genes in MDA-MB-231 cells

While SIM2 has been shown to function as both a transcriptional activator and repressor it is clear from the differentially regulated genes identified by RNA sequencing that SIM2s acts primarily as a transcriptional repressor in this context, with approximately 80% differentially expressed genes being downregulated (Ema, Morita, et al., 1996; Ema, Suzuki, et al., 1996; Metz et al., 2006; Moffett et al., 1997; Probst et al., 1997; S. Woods et al., 2008; S. L. Woods & Whitelaw, 2002). This is consistent with SIM2s action in prostate cancer cells. After knockdown of SIM2 expression in PC3 prostate cancer cells, approximately 80% of differentially expressed genes were upregulated, indicating that the expression of these genes was repressed in the presence of high levels of *SIM2* expression (Lu et al., 2011). Even though the set of genes regulated by SIM2 in these two cell types were mutually exclusive, the nature of SIM2 as a transcription factor was the same.

Overall, there was a relatively small number of SIM2s target genes identified in MDA-MB-231 cells, with a total of 102 differentially expressed genes found across all comparisons performed (excluding *SIM2*) (Figure 5.9). Even though *SIM2* is expressed in normal breast tissue and cells (Figure 5.2) (Gustafson et al., 2009; Kwak et al., 2007; Laffin et al., 2008; Scribner et al., 2013), the functions of SIM2 may be limited in this cell type, therefore only a discrete set of genes are perturbed by up- or downregulation of SIM2. In addition, the effects on the transcriptome from upregulation of SIM2s was only assessed up to 24 hours post induction of SIM2s expression, which does not reach a maximum until approximately 16 hours (Figure 5.4B). It is likely that if this time was extended additional differentially expressed genes would be identified. They may

however no longer be genes that are directly regulated by SIM2s but are genes that are as a flow on effect of the up- or downregulation of genes by SIM2s that influence their expression. Additionally, there may be genes that are regulated by SIM2s which had not yet reached the defined threshold to be called as differentially expressed. For example, a gene repressed by SIM2s that has an mRNA with a long half-life may not reach the defined threshold of expression being reduced by less than a half. These genes while not identified as SIM2s target genes by this study may still be important genes in the function of SIM2s in breast cancer.

It is unknown from this study which of the identified target genes identified by RNA sequencing in MDA-MB-231 cells are specific to the short isoform and what genes would be similarly regulated by the long isoform. This could be determined through generation of MDA-MB-231 cell lines overexpressing *SIM2l* at both a low and high level and interrogation of expression of the key genes identified in this study. Since *NQO1* was found to be downregulated by both the short and long isoform of SIM2 in two other breast cancer cell lines, MCF7 and MDA-MB-453 (Figure 5.13), it would be expected that there would be other genes among this set similarly regulated by both SIM2 isoforms. In addition, this analysis would help to determine whether there are any key isoform specific difference in the mechanism of action of SIM2 in breast cancer, and therefore why the short isoform has been reported to be downregulated in breast cancer, not the long isoform (Gustafson et al., 2009; Kwak et al., 2007; Laffin et al., 2008; Scribner et al., 2013).

Strikingly, the set of genes regulated by SIM2 across different cell types is highly variable. These investigations did not identify a core set of genes regulated by SIM2 across all cell types. This is unlike what has been found for other bHLH/PAS transcription factors with HIF1α and AHR both known to each have a core set of genes which are regulated in response to their respective signals (hypoxia and ligand binding). The action of SIM2 as a transcription factor has already been shown to be highly context dependent. This may be due to cell-type specific co-factors required to specify SIM2 function. So far only one protein, MAGED1 (melanoma-associated antigen D1), has been shown to regulate the transcriptional activity of SIM2 (A. E. Sullivan et al., 2016). However, it is highly possible that there are other proteins that interact with

SIM2 and modulate its transcriptional activities that are yet to be discovered. One of the known functions of the PAS domain is protein-protein interaction, therefore it is likely that cofactor proteins would interact with SIM2 via the PAS domains (Furness et al., 2007; Guo et al., 2013; N. Hao & Whitelaw, 2013; Huang et al., 2012; Okui et al., 2005; Partch & Gardner, 2011). The presence and absence of SIM2 cofactors could lead to differential regulation of SIM2 target genes in a cell and tissue specific manner.

Throughout development *Sim2* expression is seen to be restricted to certain tissues (see section 1.1.2 for an overview of *Sim2* expression). Given that SIM2 is aberrantly upregulated in certain cancer types including prostate cancer, it is possible that the set of genes regulated by SIM2 in these cells does not form part of a potential core set of SIM2 target genes as the genes that would be regulated by SIM2 during development are no longer accessible and have been silenced. While SIM2 was initially thought to only be a tissue-restricted bHLH/PAS transcription factors whereby regulation of its spatiotemporal expression is the mechanism through which SIM2 function is controlled, recently it has also been discovered that SIM2 is also signal regulated. Pearson, Sarkar, et al (2019) identified that in a breast cancer context phosphorylation of SIM2 by ATM stabilised the protein in response to DNA damage (Scott J. Pearson et al., 2019). There may be other signals that lead to phosphorylation and stabilisation of SIM2, modulating its function in different cell and tissue types.

From here it will be important to determine which in the set of differentials expressed genes are direct target genes of SIM2s and which are indirectly controlled by modulation of SIM2s expression. This could be achieved by performing a chromatin immunoprecipitation (ChIP) using a SIM2 antibody, similar to previously described methods (Emily L. Button et al., 2022) and then either performing next generation sequencing to identify all regions where SIM2s is bound throughout the genome, or targeted qPCR at the promoter regions of the genes identified by RNA sequencing.

5.5.3 Mechanism of action of SIM2s in breast cancer

Two mechanisms of action of SIM2s in breast cancer can be observed from the RNA sequencing generated in this study. The first is SIM2s regulation of target genes that

will directly influence tumourigenic properties. Among the set of differentially expressed genes identified in MDA-MB-231 cells with overexpression of SIM2s a number of genes have been identified that would contribute to the antitumourigenic function of SIM2s in breast cancer. This is a combination of repression of genes that would favour tumour progression and upregulation of genes that would inhibit tumour progression. This dataset supports regulation of pro- and anti-tumourigenic genes by SIM2s as a mechanism for why downregulation of SIM2s in breast cancer results in increased tumour growth and progression. Figure 5.16 details SIM2s regulation of genes that will result in inhibition of tumour progression.
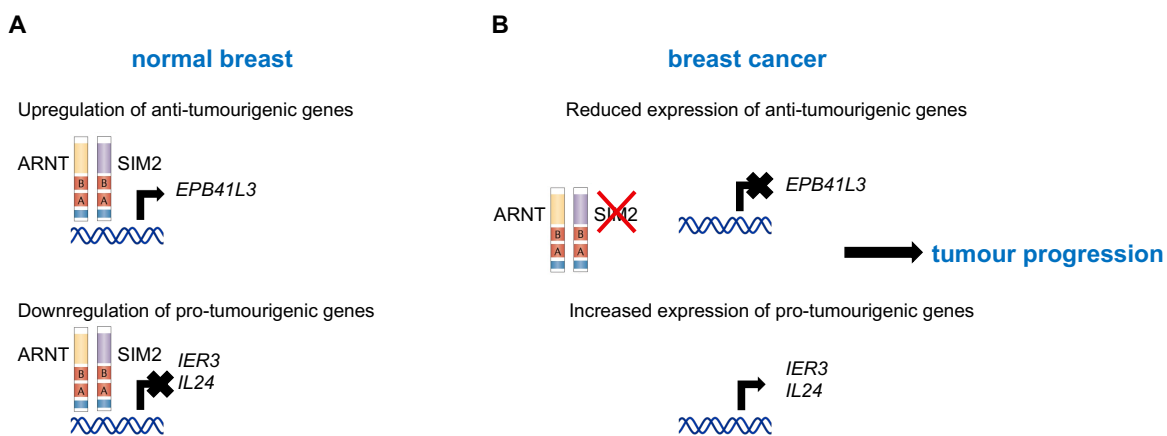
Figure 5.16: Proposed mechanism of action of SIM2 regulating target gene expression in normal and breast cancer tissue. **A)** In normal breast tissue SIM2 regulates expression of genes maintaining an overall antitumourigenic function through the upregulation of anti-tumourigenic and suppression of protumourigenic genes. **B)** Breast cancers where *SIM2* expression is downregulated results in reduced expression of antitumourigenic and increased expression of protumourigenic genes, resulting in enhanced tumour growth and progression.

In addition to SIM2s directly regulating pro- and antitumourigenic genes, crosstalk with other bHLH/PAS transcription factors appears to be occurring in breast cancer. This is evidenced by the overlap between AHR and SIM2 target genes in MDA-MB-231 cells, and the identification of *VEGFA* as a downregulated gene in SIM2s overexpressing cells (Figure 5.10). Even though AHR can have both a pro- and antitumourigenic functions in breast cancer, in MDA-MB-231 cells AHR has been reported to promote tumourigenic properties both *in vitro* and *in vivo* (Goode et al., 2013; Goode et al.,

2014). RNA-sequencing of the MDA-MB-231 cells overexpressing SIM2s found that SIM2s can repress expression of protumourigenic AHR target genes. In this context SIM2s crosstalk with AHR would reduce the protumourigenic function of AHR. AHR has previously been reported to be constitutively active in MDA-MB-231 cells, therefore crosstalk between SIM2 and AHR has been observed in an AHR activating ligand free setting (Goode et al., 2013).

HIF1α has been found to be overexpressed in human breast cancers, and overexpression is associated with increased mortality (Bos et al., 2003; Schindl et al., 2002; G. L. Semenza, 2003). There are many mechanisms by which HIF1α promotes tumourigenesis including angiogenesis, cell survival and proliferation and alterations to metabolism (Bersten et al., 2013). The RNA-sequencing of MDA-MB-231 cells overexpressing SIM2s was performed under normal oxygen conditions, therefore HIF1α would not be expected to be stabilised and regulating target gene expression. However, expression of *VEGFA* was repressed in these cells, indicating that SIM2s is also interfering in the HIF1α signalling pathway. Given that SIM2 has been reported to be able to repress expression of another HIF1α target gene, *BNIP3*, it would be expected that under conditions of hypoxia where HIF1α is stabilised and transcriptionally active SIM2 would also be able to repress expression of additional HIF1α target genes (Farrall & Whitelaw, 2009).

Target genes identified by these gene expression studies suggest a mechanism whereby expression of SIM2s in normal breast is required to compete and crosstalk with AHR and HIF1α to potentially dampen or modulate the pro-tumourigenic functions of these transcription factors. In breast cancers where SIM2 expression is downregulated, this favours tumour progression through removing competition between these transcription factors, thereby increasing the pro-tumourigenic activity of AHR and HIF1α. Overall resulting in increased expression of AHR and HIF1α target genes that favour tumour progression. An additional layer of crosstalk exists through SIM2s downregulation of the AHR target gene *NQO1*, which has been previously reported to increase the stability of HIF1α protein in cancer cells (N. Hao et al., 2012; N. Hao & Whitelaw, 2013; Oh & Park, 2015; Okey et al., 2005). Therefore, SIM2 also appears to be able crosstalk with HIF1α signalling through potentially influencing

factors that affect the stability of HIF1α protein. Figure 5.17 depicts the proposed model of crosstalk between SIM2, AHR and HIF1α bHLH/PAS transcription factors in breast cancer.
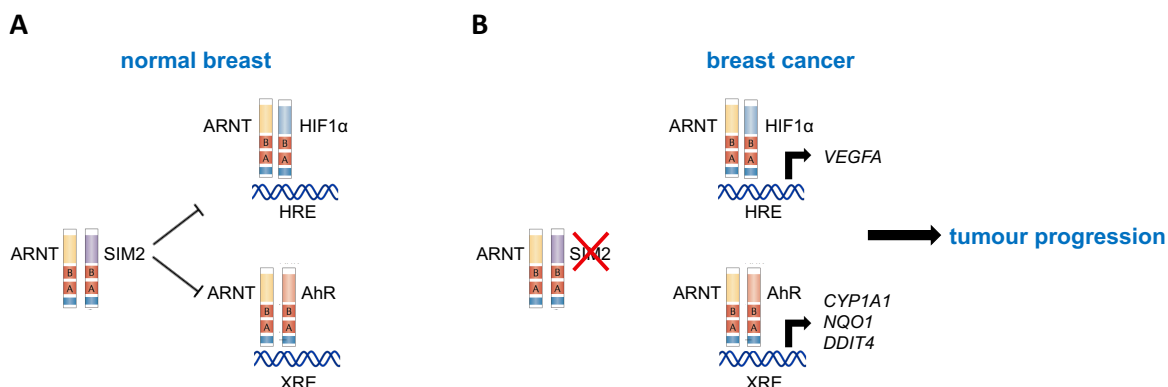
Figure 5.17: Proposed mechanism of SIM2 crosstalk in breast cancer progression. **A)** In normal breast tissue SIM2 is expressed and competes with HIF1α and AhR for dimerisation with their common partner factor ARNT. This leads to repression of HIF1α and AhR protumourigenic functions. In breast cancer cells where SIM2 expression is downregulated, this competition is removed, allowing HIF1α and AhR upregulate expression of target genes that favour tumour progression.

The proposed mechanism of crosstalk between SIM2s and the two other bHLH/PAS transcription factors AHR and HIF1α in breast cancer cells could form the basis of further exploration into this mechanism of action. It was known that SIM2 can act as a transcriptional repressor through competition with other bHLH/PAS transcription factors for dimerising with their common partner factors ARNT or ARNT2 and for these dimers binding to DNA response elements (Emily L. Button et al., 2017; Farrall & Whitelaw, 2009; S. L. Woods & Whitelaw, 2002). Co-immunoprecipitation experiments could be performed to determine whether SIM2s is outcompeting AHR and HIF1α for dimerisation with ARNT. While the MDA-MB-231 cells with a high level of overexpression of SIM2s were found to dimerise with the majority of ARNT protein, the low-level overexpression cells were only dimerising with a very small proportion of ARNT (Figure 5.3), therefore it would be unlikely that SIM2s would be outcompeting AHR and HIF1α in these cells. Chromatin immunoprecipitation followed by high throughput sequencing or qPCR experiments could also be performed to determine if

SIM2s-ARNT heterodimers are occupying AHR and HIF1α DNA response elements, preventing AHR-ARNT or HIF1α-ARNT dimers from binding and regulating target gene expression. Furthermore, additional RNA-sequencing experiments could be performed under conditions of hypoxia and with an AHR ligand to determine the level of crosstalk that occurs between these proteins when they are signal activated. It would be expected that additional genes normally upregulated by HIF1α or AHR under signal activation conditions would be found to be repressed by upregulation of SIM2s.

Also of interest is the downregulation of the bHLH/PAS transcription factor Neuronal PAS Domain Protein 2 (*NPAS2*) by SIM2s (Figure 5.10). NPAS2 functions in the circadian rhythm and forms a heterodimer with BMAL1 (or ARNT-like, ARNTL) to regulate expression of core circadian rhythm components including *PER1/2/3* and *CRY1/2*. NPAS2 is also implemented in cancer progression, acting as both a tumour suppressor and oncogene, dependent on cancer type. In breast cancer downregulation of *NPAS2* has been shown to promote tumour progression, whereas high expression is correlated with increased disease free survival (Bersten et al., 2013; Peng, Bai, & Pang, 2021; Sancar & Van Gelder, 2021). While downregulation of *NPAS2* by SIM2s would likely have a protumourigenic effect on cancer cells, this finding does add another level of complexity to the ways in which SIM2s can crosstalk with other members of the bHLH/PAS transcription factor family. It is still to be determined whether *NPAS2* is a direct or indirect target of SIM2s, however this is the first instance of SIM2s being found to repress expression of another bHLH/PAS transcription factor.

Overall this study builds on the current knowledge of the mechanism of action of SIM2 in breast cancer by adding a putative set of SIM2s target genes and defining crosstalk as a key anti-tumourigenic function of SIM2s.

5.5 RNA sequencing differentially expressed genes tables

Table 5.3: Differentially expressed genes in MDA-MB-231 cells after 6 hours of SIM2s induction compared to parent MDA-MB-231 cells.

Gene	Log2 fold change	p.value
SLC14A1	-3.334	1.00E-06
P2RY6	-2.991	0.00022
SLCO4A1	-2.707	0
GPRC5C	-2.596	2.00E-06
NGFR	-2.406	0
DDX53	-2.345	0.0277
LOC100127888	-2.28	1.00E-04
INHBB	-2.275	0
SHISA2	-2.266	0
CYP24A1	-2.261	6.70E-05
NOX5	-2.249	6.00E-06
TFAP2C	-2.152	0
OTUB2	-1.807	0
DDIT4	-1.784	0
FAM43A	-1.77	1.10E-05
TNS4	-1.76	0.01588
KIAA1199	-1.745	0
COL6A1	-1.707	0
FOXQ1	-1.7	0
CSF1	-1.679	0
ADAMTS15	-1.653	0
CYP1B1	-1.604	0
PTPRU	-1.561	0.00025
ARL4C	-1.544	0
GPR110	-1.527	1.00E-06
THAP11	-1.494	0.00894
PTGES	-1.448	0.00558
VEGFA	-1.44	0
FAM20C	-1.439	0.00038
S100A4	-1.396	0.0004
IGFBP1	-1.35	0.00394
SGK223	-1.329	0.00022
IER3	-1.293	0
FLG	-1.25	0.04791
CDH11	-1.203	0.00355
PCDH7	1.233	0.01717

Gene		
NRK	1.313	0.00708
EGR1	1.659	0
TIE1	1.704	0.00183
UCA1	1.705	0
LCP1	1.833	0.00025
WNT5A	1.937	0
PTGS1	2.097	0.00049
LAMA1	2.197	0.00056
STXBP6	2.263	0.03812
GJB3	2.639	0.00039
SIM2	6.932	0

Table 5.4: Differentially expressed genes in MDA-MB-231 cells after 24 hours of SIM2s induction compared to parent MDA-MB-231 cells.

Gene	Log2 fold change	p.value
SLC14A1	-5.571	0
CYP1B1	-4.394	0
SYT2	-4.08	0
VIPR1	-3.887	0
SHISA2	-3.828	0
SLC1A7	-3.614	0.006271
C10orf81	-3.504	3.00E-06
CYP1A1	-3.422	0
GPRC5C	-3.4	0
KIAA1199	-3.38	0
NOX5	-3.158	0
IL24	-2.997	0
CYP24A1	-2.957	0
NGFR	-2.793	0
SLCO4A1	-2.744	0
TFF2	-2.708	0.00068
LOC100127888	-2.652	2.00E-06
BMF	-2.651	0.001713
FAM43A	-2.441	0
FLG	-2.42	0
ALPP	-2.411	3.00E-06
SLC16A9	-2.329	0.042175
CD96	-2.323	0.026608
SAA1	-2.304	0.00438
TFAP2C	-2.251	0
VEGFA	-2.064	0

GPR110	-2.022	0
APOBEC3G	-1.965	0.035985
ACOXL	-1.959	0.010808
TNS4	-1.905	0.001497
SYT15	-1.892	0.001355
INHBB	-1.885	0
PTGES	-1.861	0
OTUB2	-1.817	0
COL6A1	-1.782	0
NTN4	-1.759	0
AGR2	-1.747	0.016623
CIITA	-1.746	0.005447
IER3	-1.716	0
PTPRU	-1.666	1.30E-05
CSF1	-1.66	0
S100A4	-1.659	0
CALB2	-1.641	0.001899
THAP11	-1.636	0.000285
GPR68	-1.616	7.10E-05
PORCN	-1.614	0.002403
LOC730755	-1.609	0.006529
DDIT4	-1.604	8.00E-06
SPDEF	-1.582	1.00E-06
FBXO32	-1.551	0.000122
WNT7B	-1.524	8.80E-05
MMP1	-1.52	0
S100A2	-1.481	0.00316
ARL4C	-1.463	0
SORL1	-1.458	0.000305
SLC37A2	-1.454	0
FOXQ1	-1.441	1.00E-06
ADAMTS15	-1.434	1.20E-05
PAPPA	-1.398	0.000291
PLXNA1	-1.393	0
THBS1	-1.374	0
COL7A1	-1.367	0.000843
ABCG2	-1.338	0.001208
HPCAL1	-1.298	1.90E-05
NPAS2	-1.289	2.60E-05
FZD7	-1.258	0.002716
SGK223	-1.227	0.013882
TSC22D1	-1.225	0.000154

Gene	Log2 fold change	p.value
NQO1	-1.214	0.003816
PLAT	-1.18	0.029492
SLC7A5	-1.168	0.000324
CDH11	-1.163	0.026302
DUSP6	-1.12	0.042949
RGS4	1.359	0.014833
RG9MTD2	1.395	0.009788
SEMA3E	1.676	0.005447
WNT5A	1.691	0.000106
TIE1	1.75	0.000409
DSCAM	1.796	0.032566
FBXL19-AS1	1.83	0.005447
NRK	1.871	0
LAMA1	1.891	0.020217
LCP1	1.915	2.70E-05
UCA1	2.047	0
PTPRQ	2.117	0
PPP1R14C	2.205	0.045571
KCNQ3	2.322	0.032566
PTGS1	2.588	0
LOC100505817	2.619	0.001497
GJB3	2.695	0.000133
PCDH15	2.84	0.000759
EPB41L3	3.017	0.000152
STXBP6	3.244	2.00E-06
SIM2	6.928	0

Table 5.5: Differentially expressed genes in MDA-MB-231 cells after 24 hours of SIM2s induction compared to 6 hours of SIM2s induction.

Gene	Log2 fold change	p.value
CYP1A1	-3.148	0
VIPR1	-2.974	0
CYP1B1	-2.79	0
IL24	-2.009	0.042476
AHRR	-1.731	0.042476
KIAA1199	-1.635	0
SHISA2	-1.562	0.010046
MMP1	-1.387	0
PMEPA1	-1.338	0.000959

Table 5.6: Differentially expressed genes in MDA-MB-231 cells exclusively after 6 hours of SIM2s induction compared to parent MDA-MB-231 cells.

Gene	Log2 fold change
P2RY6	-2.991
DDX53	-2.345
FAM20C	-1.439
IGFBP1	-1.35
EGR1	1.659
PCDH7	1.233

Table 5.7: Differentially expressed genes in MDA-MB-231 cells exclusively after 24 hours of SIM2s induction compared to parent MDA-MB-231 cells.

Gene	Log2 fold change
SYT2	-4.08
SLC1A7	-3.614
C10orf81	-3.504
TFF2	-2.708
BMF	-2.651
ALPP	-2.411
SLC16A9	-2.329
CD96	-2.323
SAA1	-2.304
APOBEC3G	-1.965
ACOXL	-1.959
SYT15	-1.892
NTN4	-1.759
AGR2	-1.747
CIITA	-1.746
CALB2	-1.641
GPR68	-1.616
PORCN	-1.614
LOC730755	-1.609
SPDEF	-1.582
FBXO32	-1.551
WNT7B	-1.524
S100A2	-1.481
SORL1	-1.458
SLC37A2	-1.454
PAPPA	-1.398

PLXNA1	-1.393
THBS1	-1.374
COL7A1	-1.367
ABCG2	-1.338
HPCAL1	-1.298
NPAS2	-1.289
FZD7	-1.258
TSC22D1	-1.225
NQO1	-1.214
PLAT	-1.18
SLC7A5	-1.168
DUSP6	-1.12
RGS4	1.359
RG9MTD2	1.395
SEMA3E	1.676
DSCAM	1.796
FBXL19-AS1	1.83
PTPRQ	2.117
PPP1R14C	2.205
KCNQ3	2.322
LOC100505817	2.619
PCDH15	2.84
EPB41L3	3.017

Table 5.8: Differentially expressed genes in MDA-MB-231 cells exclusively after 24 hours of SIM2s induction compared to 6 hours of SIM2s induction.

Gene	Log2 fold change
AHRR	-1.731
PMEPA1	-1.338

Table 5.9: Commonly differentially expressed genes in MDA-MB-231 cells after both 6 and 24 hours of SIM2s induction compared to parent MDA-MB-231 cells.

Gene	Log2 fold change SIM2s 6 hrs vs parent	Log2 fold change SIM2s 24 hrs vs parent
SLC14A1	-3.334	-5.571
GPRC5C	-2.596	-3.4
NOX5	-2.249	-3.158
CYP24A1	-2.261	-2.957
NGFR	-2.406	-2.793
SLCO4A1	-2.707	-2.744
LOC100127888	-2.28	-2.652
FAM43A	-1.77	-2.441
FLG	-1.25	-2.42
TFAP2C	-2.152	-2.251
VEGFA	-1.44	-2.064
GPR110	-1.527	-2.022
TNS4	-1.76	-1.905
INHBB	-2.275	-1.885
PTGES	-1.448	-1.861
OTUB2	-1.807	-1.817
COL6A1	-1.707	-1.782
IER3	-1.293	-1.716
PTPRU	-1.561	-1.666
CSF1	-1.679	-1.66
S100A4	-1.396	-1.659
THAP11	-1.494	-1.636
DDIT4	-1.784	-1.604
ARL4C	-1.544	-1.463
FOXQ1	-1.7	-1.441
ADAMTS15	-1.653	-1.434
SGK223	-1.329	-1.227
CDH11	-1.203	-1.163
WNT5A	1.937	1.691
TIE1	1.704	1.75
NRK	1.313	1.871
LAMA1	2.197	1.891
LCP1	1.833	1.915
UCA1	1.705	2.047
PTGS1	2.097	2.588
GJB3	2.639	2.695
STXBP6	2.263	3.244
SIM2	6.932	6.928

Table 5.10: Commonly differentially expressed genes in MDA-MB-231 cells after 24 hours of SIM2s induction and after 24 hours of SIM2s induction compared to 6 hours of SIM2s induction.

Gene	Log2 fold change SIM2s 24 hrs vs parent	Log2 fold change SIM2s 24 hrs vs 6 hrs
VIPR1	-3.887	-2.974
CYP1A1	-3.422	-3.148
IL24	-2.997	-2.009
MMP1	-1.52	-1.387

Table 5.11: Differentially expressed genes in MDA-MB-231 cells after both 6 and 24 hours of SIM2s induction, and after 24 hours of SIM2s induction compared to 6 hours of SIM2s induction.

Gene	Log2 fold change SIM2s 6 hrs vs parent	Log2 fold change SIM2s 24 hrs vs parent	Log2 fold change SIM2s 24 hrs vs 6 hrs
CYP1B1	-1.604	-4.394	-2.79
SHISA2	-2.266	-3.828	-1.562
KIAA1199	-1.745	-3.38	-1.635

Chapter 6: Generation and analysis of mouse models of SIM2

6.1 Introduction

6.1.1 Current understanding of *Sim2* function through mouse models

Previously reported mouse models of *Sim2* provide a valuable insight into function during development. These mouse models have demonstrated a role for *Sim2* in neuronal development, skeletal development and innate immunity in the intestines (K.-J. Chen et al., 2014; Chrast et al., 2000; Ema et al., 1999; Goshu et al., 2002; Goshu et al., 2004; Marion et al., 2005; Shamblott et al., 2002). However, the molecular mechanisms behind most of these phenotypes remains to be determined.

A common phenotype across two of the reported *Sim2* KO mouse models is accumulation of air or gas in the GI tract. While one study suggested this was due to craniofacial abnormalities resulting in breathing difficulties, causing accumulation of air (Shamblott et al., 2002), this proposal is unlikely given the other two studies did not find similar craniofacial abnormalities (K.-J. Chen et al., 2014; Goshu et al., 2002). Chen et al (2014) clearly demonstrated the mechanism behind this phenotype as due to the role of *Sim2* in innate immunity in the intestines. Loss of *Sim2* was shown to cause increased microbial growth in the GI tract as a result of lower expression of key antimicrobial peptides Cryptdin 1, Cryptdin 2, Cryptdin 6, CRAMP, MMP7 and TCF7L2. An *in vitro* secreted alkaline phosphatase (SEAP) reporter system was used to demonstrate SIM2 binding to promotor region of *Cramp*, *RegIIIγ*, *Pla IIa* (phospholipase A2) and *Tcf7l2*, suggesting direct regulation of these genes (K.-J. Chen et al., 2014).

Breathing difficulty was another phenotype common to two of the reported mouse models, however the proposed explanations for this observation differ between the two reports. Shamblott et al (2002) proposed the breathing difficulties were due to the craniofacial abnormalities, namely a fully or partially cleft palate seen in the *Sim2* KO mice, but not their heterozygous or WT littermates. It is worth noting that this

phenotype was not observed in the other two studies (K.-J. Chen et al., 2014; Goshu et al., 2002). The group further explored this phenotype and showed that this defect is likely due to hypocellularity and an excess of extracellular matrix, specifically hyaluronan, in the developing palate of KO animals. This suggests that *Sim2* may somehow be a regulator of hyaluronan accumulation in the palate, however the mechanism behind this was not determined. Alternatively, Goshu et al (2002) proposed the breathing difficulties to   result from other skeletal defects. Both heterozygous and KO *Sim2* mice had small protrusions from the ribs and vertebrae which made aberrant connections to the surrounding intercostal muscles. These were likely due to increased proliferation, as shown by increased BrdU (Bromodeoxyuridine) positive cells in and around the protrusions compared to other areas within the developing ribs. They proposed the breathing difficulties were caused by compromised structure surrounding the pleural cavity, including the rib protrusions as well as a thin diaphragm which ultimately led to tearing of the pleural mesothelium and lung atelectasis. Both the heterozygous and homozygous *Sim2* KO mice also displayed incompletely penetrant scoliosis of the spine, suggesting that *Sim2* may be a potential cause of this condition (Goshu et al., 2002). Since this phenotype was not seen in all KO animals, this suggests that loss of *Sim2* function alone may not be enough to cause scoliosis but may be a contributing factor in combination with other genetic or environmental factors. Again, the mechanism behind why *Sim2* KO mice display these phenotypes has not yet been explained.

Interestingly, the only phenotype that was reported in all of the *Sim2* KO mouse studies was perinatal lethality. The first two studies where the majority of *Sim2* null mice did not survive to adulthood attributed this to the observed breathing difficulties, albeit with differing causes of this phenotype (Goshu et al., 2002; Shamblott et al., 2002). In the most recent study, around half of the *Sim2* null mice survived to adulthood. It was noted that these were the mice with less severe gas accumulation suggesting that the increased microbial growth is the cause of lethality in this mouse line (K.-J. Chen et al., 2014).

The varying phenotypes in these mouse models may be explained at least in part by differences in genetic background. Chen et al (2014) stated that the mouse line was

backcrossed into a C57BL/6 background. Shamblott et al (2002) generated the mouse line through gene targeting in J1 embryonic stem (ES) cells, which were then injected into C57BL/6J mouse blastocysts. Goshu et al (2002) were provided with heterozygous female from Shamblott et al which was then crossed with a BL6 male and backcrossed for 5 generations. Given that there are many sub-strains of C57BL/6 mice it is likely all three of these mouse lines are on differing genetic backgrounds. This suggests that *Sim2* KO phenotypes may be highly variable and other genetic and environmental factors may influence the penetrance of these phenotypes.

6.1.2 *Sim2* and *Sim1* have overlapping and distinct functions

SIM1 and SIM2 share a high degree of homology through their N-terminal halves. Consequently, it has been shown that both SIM1 and SIM2 heterodimerise with the same partner proteins, ARNT and ARNT2, as well as bind to the same DNA motif, the Central Midline Element (CME). Given this, it is possible that there may be crosstalk between these two proteins through competition for both partner protein and DNA binding. Their C-terminal halves are highly divergent showing no significant homology, which likely allows the two transcription factors to perform distinct functions (Chrast et al., 1997; Ema, Morita, et al., 1996; Ema, Suzuki, et al., 1996; Fan et al., 1996; Moffett & Pelletier, 2000; S. L. Woods & Whitelaw, 2002).

The expression pattern of *Sim1* during mouse embryonic development is similar to that seen for *Sim2*. *Sim1* expression within the brain has a slightly later onset than that of *Sim2*, beginning within the caudal diencephalon at embryonic day 9.0 (E9.0). Expression within the brain is found later in regions derived from this area, most notably within the paraventricular (PVN), supraoptic (SON) and anterior periventricular (aPV) nuclei within the hypothalamus. Outside the central nervous system (CNS), *Sim1* expression is also seen in the kidney tubules and somites (Fan et al., 1996). In adult mouse tissues *Sim1* is expressed in the brain, kidneys, skeletal muscle and lungs (Metz et al., 2006).

In *Sim1* knockout mice the paraventricular nucleus (PVN) and supraoptic nucleus (SON) within the hypothalamus are hypocellular, resulting in lack of expression of the

neuropeptides Somatostatin (Ss), Thyrotropin releasing hormone (Trh), Oxytocin (OT), arginine vasopressin (AVP) and corticotropin releasing hormone (Crh) (J. L. Holder et al., 2004; Kublaoui et al., 2008; J. L. Michaud et al., 2001; J. L. Michaud et al., 1998; Tolson et al., 2010). Interestingly *Sim2* has also been shown to play a role in the development of cells within the hypothalamus, with both *Sim2* heterozygous and KO mice having a significant decrease in the number of Somatostatin (SS) and Thyrotropin Releasing Hormone (TRH) expressing neurons within the Paraventricular Nucleus (PVN) (Goshu et al., 2004). Heterozygous *Sim1* KO on a *Sim2* KO background further exacerbates this phenotype, suggesting *Sim2* and *Sim1* may genetically interact in pathways that regulate the development of these cell types, while still performing distinct functions evidenced by the fact that both genes are required for full development and neuropeptide expression of these cell types. Furthermore, it appears as though *Sim2* may be acting downstream of *Sim1*, supported by the fact that *Sim1* expression in the hypothalamus is not affected by loss of *Sim2*, whereas *Sim2* expression is lost in *Sim1* KO mice within the developing dorsal preoptic area (dP) and PVN (Goshu et al., 2004).

SIM1 has been shown to be a monogenetic cause of obesity in both mice and humans (J. L. Holder, Jr. et al., 2000; J. L. Holder et al., 2004; Kublaoui et al., 2008; Kublaoui, Holder, Gemelli, et al., 2006; Kublaoui, Holder, Tolson, et al., 2006; J. L. Michaud et al., 2001; Tolson et al., 2010). *Sim1* heterozygous KO mice and conditional KOs show increased weight gain compared to WT littermates (J. L. Michaud et al., 2001; Tolson et al., 2010). Multiple *SIM1* gene variants found in humans with hyperphagic obesity exhibit decreased transcriptional activity (Bonnefond et al., 2013; Ramachandrappa et al., 2013; Adrienne E. Sullivan et al., 2014). *Sim1* appears to act in the leptin-melanocortin appetite control pathway, with *Sim1* heterozygous and conditional KO mice having reduced hypothalamic expression of *Mc4r* (Tolson et al., 2010). Given the overlapping expression and high level of N-terminal conservation between *Sim1* and *Sim2*, it may be possible that *Sim2* also plays a role in appetite control. While Goshu et al (2002) showed that *Sim2* heterozygous mice did not display increased weight gain compared to WT mice, it is possible that remaining *Sim2* is sufficient for function(Goshu et al., 2002).

Sim1 and *Sim2* target genes in tissues and cell types where they are independently and co-expressed are largely unknown. Determining these will be a critical step towards elucidating their overlapping and distinct functions.

6.1.3 Challenges for in vivo analysis of *Sim2* function

In order to elucidate and fully understand the unexplained phenotypes of the *Sim2* KO mouse models a more detailed map of SIM2 RNA and protein expression, both pre- and postnatally, as well as defining tissue specific SIM2 target genes, is essential. In situ hybridisation to map *Sim2* mRNA expression during mouse embryonic development has proven to be problematic due to low levels of *Sim2* mRNA expression (C.M Fan, personal communication).

Currently, there is no antibody that will robustly detect endogenous SIM2 protein. Additionally, effectively determining SIM2 regulated genes requires an efficient way to select for specific cells in which SIM2 is expressed. There is a desperate need for new tools to rigorously define *Sim2* expression and isolate SIM2 expressing, and KO counterpart, cells. The lack of sensitive SIM2 antibody could be solved by introducing an epitope tag into the endogenous *Sim2* locus in a mouse model. Antibodies to the epitope tag could be used to investigate SIM2 protein expression during both embryonic and postnatal development and through to adulthood. In order to clearly identify and isolate cells in which SIM2 is expressed, together with KO counterpart cells, a knock-in mouse model containing a fluorescent reporter inserted into the *Sim2* locus, replacing endogenous *Sim2* expression, would be highly beneficial.

The function of *Sim2* in the brain in postnatal development has not been investigated due to the perinatal lethality of global *Sim2* KO. A conditional *Sim2* KO mouse model in which *Sim2* expression is specifically removed within the brain would be ideal to overcome this. The lethality of *Sim2* KO mice does not appear to be due to any neurological phenotypes, therefore these mice should survive to adulthood and allow for interrogation of any potential neurological phenotypes associated with loss of *Sim2* function in the brain. *Sim2* KO mice would be expected to manifest a neurological

phenotype given that *Sim2* is expressed within the brain both during embryonic development and postnatally, and that there is a reduction of SS and TRH expressing neurons within the hypothalamus of *Sim2* KO embryos ("Allen Developing Mouse Brain Atlas," ; Coumailleau & Duprez, 2009; Ema, Morita, et al., 1996; Fan et al., 1996; Goshu et al., 2004; Marion et al., 2005).

6.2 Chapter Aims

This chapter presents work towards a central aim of generating, validating and analysing new mouse models of SIM2. This aim was broken down into the following three sub-aims:

1) Phenotypically analyse a conditional neuronal *Sim2* knockout mouse model.
2) Generate, validate, and characterise an epitope tagged SIM2 mouse model.
3) Generate, validate, and characterise a *Sim2* fluorescent reporter mouse model.

This work aimed to develop the tools necessary to form a comprehensive understanding of the cells and tissues in which SIM2 is expressed, and the genes which are specifically regulated by SIM2. All three sub-aims are interconnected, with each likely to provide support to the other two.

6.3 Experimental systems and approaches

6.3.1 Conditional *Sim2* KO mouse line

A conditional *Sim2* knockout mouse model had been previously established by the Whitelaw laboratory, with part of this work presented in my Honours thesis (Whitelaw laboratory unpublished data), (E. L. Button, 2013). This was achieved through targeting the *Sim2* locus in mouse ES cells through homologous recombination. A targeting vector was designed to insert LoxP sites either side of Exon 1 of *Sim2*, creating a floxed allele, in addition to a neomycin resistance selection cassette flanked by Frt sites for selection of successfully targeted ES cells. The LoxP sites allow for selective deletion of a portion of Exon 1 containing the start codon, through cre-mediated recombination, to generate a *Sim2* KO allele (Figure 6.2A). Additionally, the Frt sites

allowed for removal of the selection cassette by flp-mediated recombination, in order to minimise the amount of exogenous DNA sequence present in the mouse line. The mouse line was generated through injection of successfully modified ES cells into a blastocyst, which was then transferred to a pseudo-pregnant female to generate chimeric mice (Sandra Piltz, SA Genome Editing Facility). Once established, this mouse line was crossed with a ubiquitously expressed Flp-recombinase mouse to remove the selection cassette (Myself & Murray Whitelaw as part of my honours project). This mouse line was designated as the *Sim2*-floxed mouse line with heterozygous mice of genotype *Sim2*$^{fl/+}$ and homozygous *Sim2*$^{fl/fl}$.

The floxed mouse line was crossed into a Nestin-Cre (Nes-Cre1) mouse line to generate a neuronal *Sim2* KO mouse. The Nes-Cre1 mouse line is used to generate neuronal gene knockouts and contains a modified Cre with a nuclear localisation sequence (NLS) and actin polyadenylation (polyA) signal (Dubois, Hofmann, Kaloulis, Bishop, & Trumpp, 2006). The Nestin-cre1 transgene consists of 5.8kb of Nestin promotor sequence upstream of the modified cre. This is then followed by the entire *Nestin* intron 2 sequence, which has been shown to contain an enhancer sufficient to direct *Nestin* expression to the CNS (Zimmerman et al., 1994). Nes-cre1-mediated recombination of floxed alleles has been shown to be detected as early as E7.5-8.0 in cells within the developing mesoderm, with close to complete recombination of floxed alleles in cells within the developing CNS by E15.5. In addition, recombination was also seen at this stage in somite-derived tissues including the muscle, dermis, cartilage and bone. In adult tissues, recombination of floxed alleles was seen in greater than 99% of cells within the kidney and brain and partial recombination was seen in the heart, lung and liver. The Nes-cre1 transgene is also active in germline cells, so floxed alleles will be inherited as recombined alleles in the next generation (Buchholz, Refaeli, Trumpp, & Bishop, 2000; Dubois et al., 2006; Trumpp, Depew, Rubenstein, Bishop, & Martin, 1999). The Nes-cre1 mouse line was a gift from the McColl laboratory (University of Adelaide).

In order to generate the *Sim2*-Nestin KO mouse line (from here on referred to as *Sim2*-Nes-cre1) the *Sim2*-floxed mouse line was crossed with the Nes-cre1 mouse line. Given that the Nes-cre1 transgene is expressed in germline cells, the breeding strategy to

obtain the desired genotypes had to be designed around this. See Figure 6.1 for outline of the breeding strategy. Two rounds of breeding were required to generate the desired *Sim2* neuronal KO genotype. First a mouse of genotype $Sim2^{fl/+}$ was crossed with a Nes-cre1 mouse, producing offspring with the somatic cell genotype $Sim2^{fl/+;Nes\text{-}cre1}$ and germline genotype $Sim2^{+/-;Nes\text{-}cre1}$. These mice were then crossed with $Sim2^{fl/fl}$ mice in order to produce mice with the following genotypes; $Sim2^{fl/+}$, $Sim2^{fl/-}$, $Sim2^{fl/+;Nes\text{-}cre1}$, $Sim2^{fl/-;Nes\text{-}cre1}$. The $Sim2^{fl/+}$ animals will have an overall WT *Sim2* genotype. The $Sim2^{fl/-}$ and $Sim2^{fl/+;Nes\text{-}cre1}$ mice will have an overall heterozygous *Sim2* genotype, differing only by the presence of the Nes-cre1 transgene in the latter. Mice with genotype $Sim2^{fl/-;Nes\text{-}cre1}$ are the neuronal *Sim2* KOs, with recombination of the floxed allele having occurred in all tissues where the Nes-cre1 is expressed. Note that the germline cells for these mice will have a genotype of $Sim2^{-/-}$.

Sim2^{fl/+}   ×   *Nes-cre1*

Sim2^{fl/+;Nes-cre1}   ×   *Sim2*^{fl/fl}

Sim2^{fl/+}   *Sim2*^{fl/-}   *Sim2*^{fl/+;Nes-cre1}   *Sim2*^{fl/-;Nes-cre1}
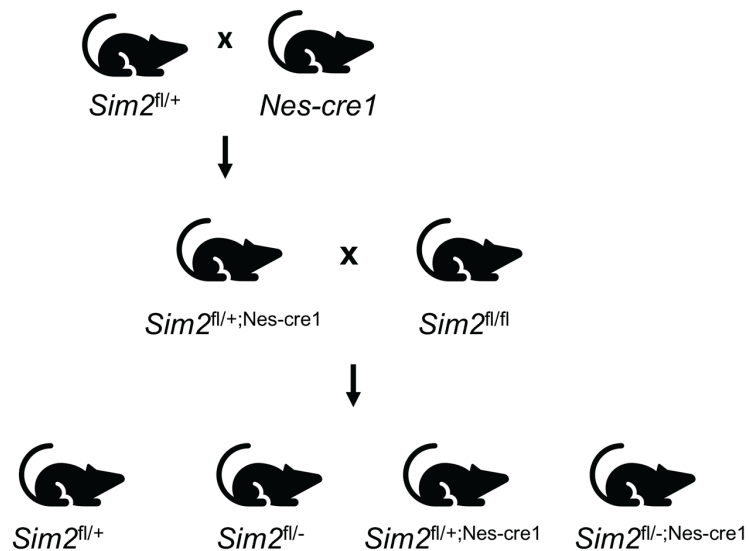
Figure 6.1: *Sim2*-Nestin-cre1 breeding strategy. Two rounds of breeding were required to generate the desired genotypes. First a heterozygous *Sim2* floxed mouse was crossed with the Nes-cre1 mouse line to generate $Sim2^{fl/+;Nes\text{-}cre1}$ mice. These mice were then crossed with homozygous *Sim2* floxed mice to generate littermates with four different genotypes; $Sim2^{fl/+}$, $Sim2^{fl/-}$, $Sim2^{fl/+;Nes\text{-}cre1}$, $Sim2^{fl/-;Nes\text{-}cre1}$. Genotypes depicted represent the somatic cell genotype.

Since the Nes-Cre transgene is expressed in germline cells, the genotype of the Nestin KO mice (*Sim2*-Nestin KO) will be $Sim2^{-/fl;\,Nes\text{-}Cre}$, with conditional deletion of the floxed *Sim2* allele in tissues where the Nes-Cre transgene is expressed. These mice are viable

so should allow for investigation into any neuronal phenotypes resulting from conditional deletion of *Sim2* from the brain.

6.3.2 Genome editing for generation of the epitope tagged and fluorescent reporter knock-in mouse lines

In order to generate the epitope tagged and fluorescent reporter mouse models, CRISPR/Cas9 technology will be used. It is possible to rapidly generate genetically modified mice with epitope tagged proteins using the CRISPR/Cas9 system (Hui Yang et al., 2013; H. Yang et al., 2014). In this system, a guide RNA (gRNA) is used to target the Cas9 nuclease to generate a double stranded break (DSB) at a specific location within the genome which has been shown to increase the efficiency of homologous recombination (Hui Yang et al., 2013). Exogenous sequences can then be inserted into that location through homologous recombination by the addition of donor DNA with homology regions in the form of a ssDNA oligonucleotide or plasmid. Genome editing can be achieved in mouse zygotes by co-injection of Cas9 mRNA or protein, a gRNA and donor DNA. This work aimed to generate two knock-in mouse models. The first being an epitope tagged model with a 3xFLAG tag sequence inserted at the 3' end of the *Sim2* coding sequence, immediately before the endogenous stop codon. The second being a knock-in mouse line where a portion of the endogenous *Sim2* exon 1 sequence is replaced with the coding sequence for the Tomato fluorescent protein, creating a reporter mouse line in the context of a *Sim2* KO allele. Both mouse lines were to be generated with the use of CRISPR-Cas9 technology to increase the efficiency of homologous recombination to insert the exogenous tag and fluorescent reporter sequences into the endogenous *Sim2* locus.

6.4 Results

6.4.1 Analysis of a conditional *Sim2* knockout mouse model

6.4.1.1 Characterisation of Sim2-NesCre conditional KO

In order to establish a neuronal *Sim2* KO mouse line, the floxed *Sim2* mouse line was crossed with the Nestin-Cre mouse line (Dubois et al., 2006). To determine the genotype of mice from matings described previously (see figure 6.1), two independent genotyping PCRs were performed. Figure 6.2B shows an example of these genotyping PCRs. One PCR determined whether the Nes-Cre1 transgene was present. As can be seen in Figure 6.2B (bottom panel), a 276bp band is amplified when the transgene is present. The other PCR determined the somatic cell status of the *Sim2* alleles. Figure 6.2A shows a schematic of the WT (+), floxed (fl) and recombined (-) *Sim2* alleles. Genotyping primers were designed either side of the region that is excised by the cre. If the WT (+), floxed (fl) or recombined(-) *Sim2* allele is present, 720bp, 816bp or 419bp products will be amplified, respectively. The result from these two genotyping PCRs was used to determine the overall genotype of each mouse (Figure 6.2C).
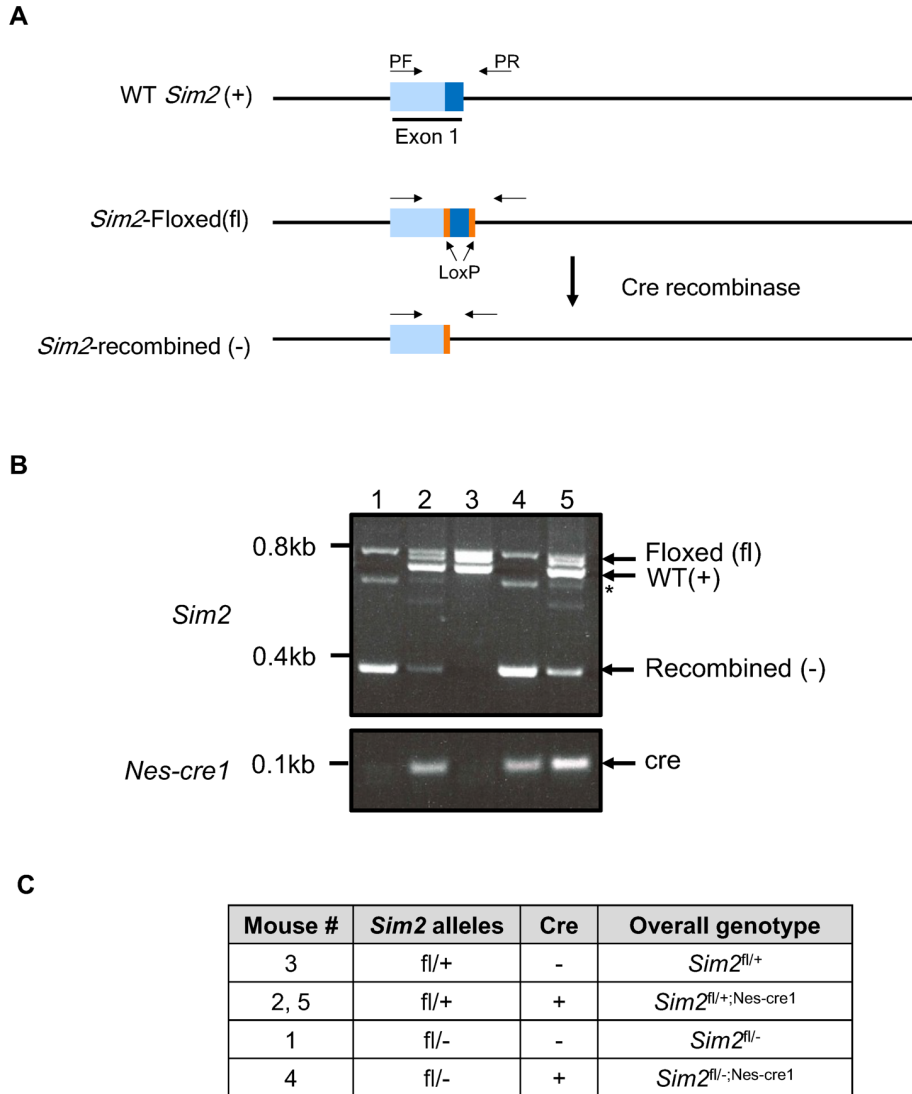
Figure 6.2: *Sim2*-Nes-cre1 genotyping. **A)** Schematic representation of *Sim2* WT (+), floxed (fl) and recombined KO (-) alleles with genotyping primer sequences shown (PF, PR). **B)** Example genotyping PCR for the *Sim2* alleles (top panel) and presence of cre transgene (bottom panel). * Non-specific product. **C)** Table displaying results of the genotyping PCRs in **B.**

A genotyping PCR was performed on genomic DNA extracts from brain, liver, kidney, lung and skeletal muscle tissues of *Sim2*^fl/fl and *Sim2*^fl/-;Nes-cre1 mice. Consistent with literature, near complete recombination of the floxed *Sim2* allele can be seen in whole brain, kidney and skeletal muscle tissues, with partial recombination seen in the liver and lung (Figure 6.3) (Dubois et al., 2006). Even though the Nestin-cre expression and allele recombination is not specific to neurons, given that kidney and muscle two other

tissues commonly reported to express *Sim2* the Nestin-cre should be advantageous to perform initial studies that might shed light on SIM2 functions. Despite the extensive recombination of the *Sim2* floxed allele in many tissues where *Sim2* is known to be expressed, these mice are viable and survive to adulthood. This suggests that the lethality phenotype observed in *Sim2* global KO mice is not due to any function *Sim2* has in tissues and at time points after the Nes-cre1 is expressed. Notably, mice with the Nes-cre1 transgene appeared to be smaller than mice that do not contain the transgene.
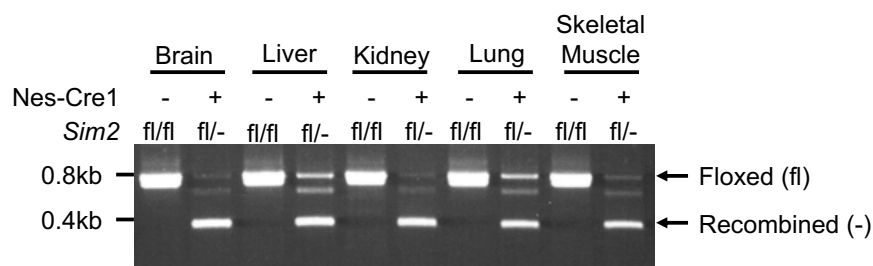
Figure 6.3: Nestin-cre1-mediated recombination of the floxed *Sim2* allele in mouse tissues. Presence (+) or absence (-) of the Nes-cre1 transgene and *Sim2* genotype (homozygous floxed: fl/fl, compound heterozygous floxed/KO: fl/-) shown above each lane. Arrows show expected band size for corresponding floxed (fl) and recombined (-) allele PCRs.

6.4.1.2   Behavioural testing

In order to investigate the neuronal function of *Sim2*, *Sim2*-Nestin KO mice were subjected to behavioural testing to determine whether there are any behavioural differences in mice that lack *Sim2* in the brain. This work was conducted as a collaboration with Professor Bernhard Baune and Dr. Emily Jaehne. A total of 20 mice, 10 *Sim2*-Nestin KO (KO) and 10 littermate controls (WT and heterozygous) were subjected to a group of behavioural tests. *Sim2* heterozygous mice were included in the WT cohort as phenotypically they do not show any major neurological defects (Goshu et al., 2004). The genotype of the KO mice was *Sim2*[flox/-; Nes-cre1]. The controls were littermates with the following genotypes; *Sim2*[flox/flox] (WT), *Sim2*[flox/WT] (WT) and *Sim2*[flox/WT; Nes-cre1] (het). There were both male and female mice in each group with ages ranging between 2.5 to 9 months (see Table 1 for details). The battery of behavioural

tests assayed baseline locomotor activity, anxiety-like behaviour, cognition-like behaviour, depression-like behaviour and social behaviour.

Table 1: Mice subjected to behavioural study.

Mouse #	Sex	*Sim2* alleles	Genotype (overall)	Age (months)
1	F	fl/-; Nes-cre1	KO	5
2	F	fl/-; Nes-cre1	KO	4
3	F	fl/-; Nes-cre1	KO	4
4	F	fl/-; Nes-cre1	KO	4
5	F	fl/-; Nes-cre1	KO	4
6	F	fl/-; Nes-cre1	KO	4
7	F	fl/-; Nes-cre1	KO	4
8	M	fl/-; Nes-cre1	KO	5
9	M	fl/-; Nes-cre1	KO	7
10	M	fl/-; Nes-cre1	KO	4
11	F	fl/fl	control (WT)	2.5
12	F	fl/WT	control (WT)	4
13	M	fl/WT	control (WT)	5
14	M	fl/WT	control (WT)	7
15	M	fl/WT	control (WT)	7
16	F	fl/WT; Nes-cre1	control (het)	5
17	F	fl/WT; Nes-cre1	control (het)	5
18	F	fl/WT; Nes-cre1	control (het)	9
19	F	fl/WT; Nes-cre1	control (het)	7
20	F	fl/WT; Nes-cre1	control (het)	2.5

The open field test was used to test for both baseline locomotor activity and anxiety related behaviour. In this study mice were observed in an open field divided into an outer and inner (centre) zone. Baseline locomotor activity was measured by distance travelled and anxiety was measured as time spent in the centre of the field, with more time spent in the centre indicating less anxiety. There was no significant difference seen between the WT and *Sim2*-Nestin KO mice for either the baseline locomotor activity or time spent in the centre of the open field (Figure 6.4A). The elevated zero

maze was also used as a measurement of anxiety like behaviour. The maze consists of a circular platform of four quadrants, two open (without walls) and two enclosed quadrants. Anxiety related behaviour was measure as the time spent in the open arms, with more time spent in open arm indicating less anxiety. There was no significant difference in time spent in the open arms between the WT and *Sim2*-Nestin KO mice, indicating that the *Sim2*-Nestin KO mice do not display a difference in anxiety-related behaviour compared to WT animals (Figure 6.4B).

The forced swim test was performed to assess depression-like behaviour. This test involved observing mice in a container filled with water and measuring time spent immobile, where greater time spent immobile would indicate depression-like behaviour. There appears to be a trend towards the *Sim2*-Nestin KO mice spending more time immobile, however this difference was not statistically significant due to high variability within the KO group. This result indicates that there may be an increase in depression-like behaviour for the *Sim2*-Nestin KO mice, however further study would be required to determine if this is truly the case.

To test spatial recognition memory the Y-maze test was performed. Mice were place in the Y-maze which consists of three arms, one start and two test arms (novel and familiar) with different coloured tape on the walls. One of the test arms (novel arm) was closed off in the first (training) and mice were observed for 10 minutes. 30 minutes later the mice were returned to the apparatus for the test phase with all three arms open for 5 minutes and time spent in each arm was measured. Both the WT and KO mice spent more time in the novel arm which was statistically significant. However there was no difference in the time spent in each arm between the WT and KO mice (Figure 6.4Dii). A preference index measurement used to assess recognition memory was calculated as the ratio of time in the novel arm to time in the novel and familiar test arms, with a ratio approaching 1 representing successful recognition memory. As can be seen in Figure 6.4Dii, there was no difference in preference index between the WT and KO mice, with both being approximately 0.6, indicating that there is no difference in spatial recognition memory.

The sociability test was performed to look for any difference in social behaviour. For the first stage mice were placed in an apparatus consisting of three chambers (two with empty cages) to acclimatise. A second mouse, designated as the stranger mouse, was then added into one of the empty cages for the second stage (sociability). Time spent interacting with each of the cages was measured and a social preference index was calculated. As would be expected, both the WT and KO mice spent a significantly larger amount of time interacting with the stranger mouse over the empty cage, however there was no difference between the two genotypes in social preference index (Figure 6.4Ei, ii). For the third stage (preference for social novelty) an additional mouse, designated as the novel mouse, was placed in the empty cage with the original stranger mouse now referred to as the familiar mouse. Interaction with the familiar mouse and novel mouse was measured and a preference index calculated as an indication of preference for social novelty. Both the WT and KO mice spent more time interacting with the novel mouse compared to the familiar mouse, with no significant difference in preference index seen between the two genotypes (Figure 6.4Eiii, iv). Overall this test did not reveal any difference in social behaviour between the WT and KO mice.

In summary, the behavioural testing of the *Sim2*-Nestin-Cre1 mouse line showed that there is no significant difference in baseline locomotor activity, anxiety, cognition and sociability in the *Sim2*-Nestin-cre1 KO mouse line. There did however appear to be a trend towards an increase in depression like behaviour, with greater time spent immobile for the KO mice in the forced swim test. This was not statistically significant due to large variation within the KO group. A larger sample size, or less variation of other factors within the group, such as age and gender, may help to determine whether this is truly a phenotype of the KO mice. It would also be interesting to include behavioural tests to assess pain sensitivity and further investigation into anxiety-related behaviour, as reduced pain sensitivity and anxiety related/reduced exploratory behaviour has been observed in a *Sim2* transgenic overexpression mouse model (Chrast et al., 2000).

A Open Field

i Baseline locomotor activity

Distance travelled (m)

15
10
5
0

Sim2 control   *Sim2* KO

ii Time spent in centre (anxiety)

Time in centre (s)

60
40
20
0

Sim2 control   *Sim2* KO

B Elevated Zero Maze

Time spent in open arms (anxiety)

Time in open arms (s)

40
30
20
10
0

Sim2 control   *Sim2* KO

C Forced Swim Test

Immobility (depression like behaviour)

Immobility time (s)

80
60
40
20
0

Sim2 control   *Sim2* KO

D Y-maze

i Spatial recognition memory

Time in arms (s)

150
100
50
0

■ Familiar Arm
■ Novel Arm

*   *

Sim2 control   *Sim2* KO

ii Spatial recognition memory

Preference index

0.8
0.6
0.4
0.2
0.0

Sim2 control   *Sim2* KO

E Sociability Test

i Sociability

Interaction time (s)

80
60
40
20
0

■ Empty Cage
■ Stranger Mouse

****   ****

Sim2 control   *Sim2* KO

ii Social preference index

Preference index

1.0
0.8
0.6
0.4
0.2
0.0

Sim2 control   *Sim2* KO

iii Preference for social novelty

Interaction time (s)

60
40
20
0

■ Familiar Mouse
■ Novel Mouse

*   **

Sim2 control   *Sim2* KO

iv Preference for social novelty index

Preference index

0.8
0.6
0.4
0.2
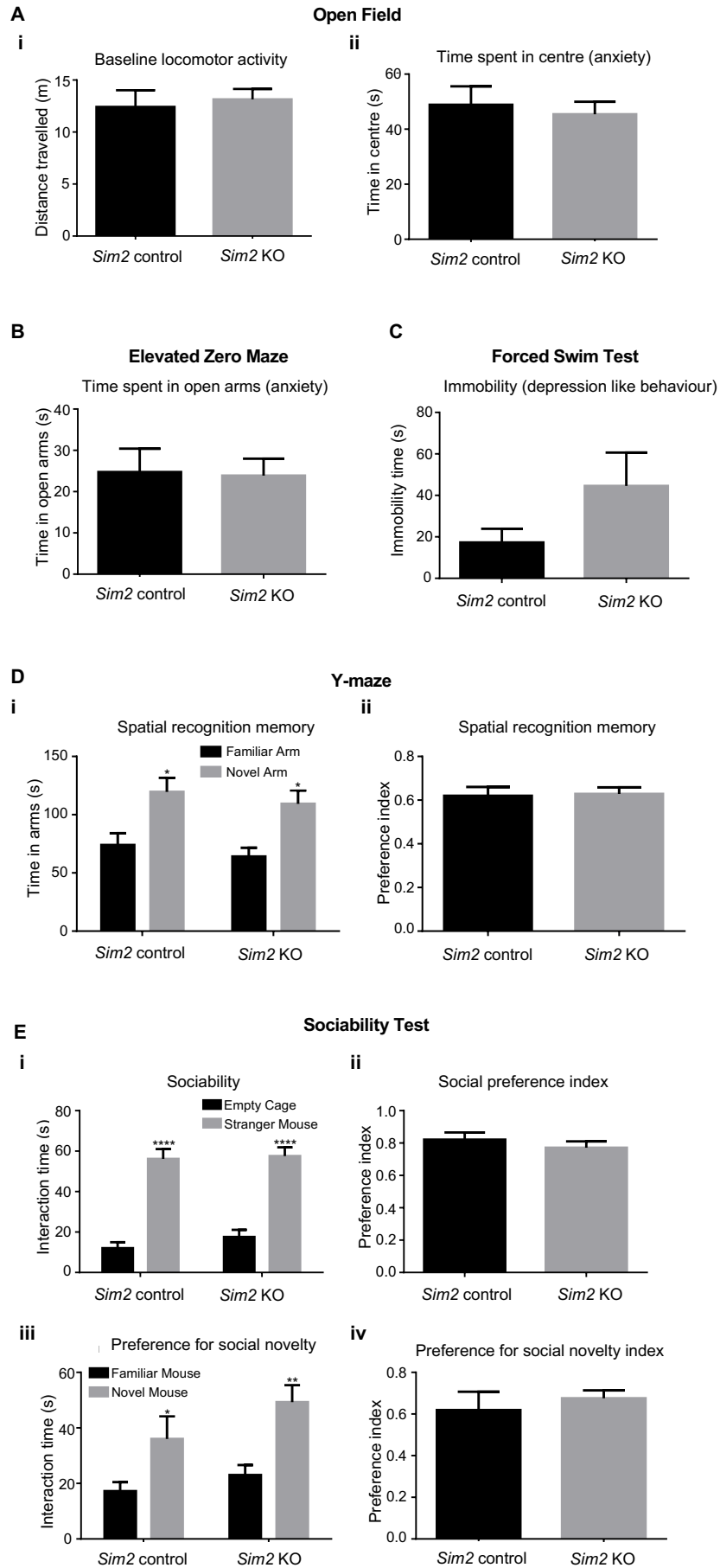0.0

Sim2 control   *Sim2* KO

161

Figure 6.4: Behavioural testing used to assess any potential neurological phenotypes in the *Sim2*- Nestin-cre1 KO mice. **A)** Baseline locomotor activity (i) and anxiety-like behaviour (ii) were assessed in the open field test. **B)** Anxiety-like behaviour was also measured by the elevated zero maze. **C)** Depression-like behaviour was measured by the forced swim test. **D)** Spatial recognition memory was assessed by the Y-maze (i, ii). **E)** Sociability (i, ii) and preference for social novelty (iii, iv) was measured by the sociability test. Data is presented as the mean and SD of n=10 animals in each group. * = p <0.05, ** = p <0.005, **** = p <0.0001.

6.4.1.3   Feeding study

Given that loss of function of the closely related paralog SIM1 has been shown to be a monogenic cause of obesity, we conducted a feeding study in order to assess whether *Sim2* was also involved in the satiety pathway. *Sim2*$^{fl/+}$ mice were used as a WT control. *Sim2*$^{fl/-}$ and *Sim2*$^{fl/+;Nes-cre1}$ represented two different heterozygous controls, with the *Sim2*$^{fl/-}$ mice having only one functional copy of *Sim2* in all tissues through all stages of development and postnatally, and the *Sim2*$^{fl/+;Nes-cre1}$ mice will only be heterozygous for *Sim2* in tissues and at developmental stages post Nestin-cre1-mediated recombination of the floxed allele (see section 6.3.1 for detailed description of Nestin-cre1 expression, Dubois et al., 2006). *Sim2*$^{fl/-;Nes-cre1}$ mice were used as the neuronal *Sim2* KO group. There were 6 mice in the WT and heterozygous groups and 5 mice in the KO group. A pilot study was conducted with additional animals to be added to the groups if the initial results were suggestive of a phenotype in the *Sim2*-Nestin KO mice. The mice were fed on a high fat diet from the age of 4 to 13 weeks and weighed weekly to measure weight gain and assess if there were any differences between the different genotypes (Figure 6.5).

Figure 6.5A shows a graph of the average recorded weights for all genotypes across the duration on the study. Overall there was a trend for the mice with the Nestin-cre1 transgene to weigh less than the mice without the transgene. The average starting weight at 4 weeks of age of the *Sim2*$^{fl/+;Nes-cre1}$ (8.0g) and *Sim2*$^{fl/-;Nes-cre1}$ (8.0g) mice was significantly less than the *Sim2*$^{fl/+}$ (12.3g) mice (Figure 6.5B). There was a trend for this to be true when comparing the *Sim2*$^{fl/+;Nes-cre1}$ and *Sim2*$^{fl/-;Nes-cre1}$ to the *Sim2*$^{fl/-}$ mice,

however this was not significant (Figure 6.5B). This overall trend continued through the duration of the study, with the final average weight at 13 weeks of age of the $Sim2^{fl/+;Nes-cre1}$ (19.8g) and $Sim2^{fl/-;Nes-cre1}$ (19.8g) mice being significantly less than the $Sim2^{fl/-}$ (24.5g) mice (Figure 6.5C). There was also a trend for this to be true when comparing the $Sim2^{fl/+;Nes-cre1}$ and $Sim2^{fl/-;Nes-cre1}$ to the $Sim2^{fl/-}$ mice, however this was not significant (Figure 6.5C). The $Sim2^{fl/-}$ mice had a significant increase in overall weight gain (13.8g) compared to the $Sim2^{fl/+}$ (10.0g), however there was no significant difference in overall weight gain when comparing any of the other genotypes, including the WT group with either the $Sim2^{fl/+;Nes-cre1}$ heterozygous group, or the $Sim2^{fl/-;Nes-cre1}$ KO group (Figure 6.5D). Given this, it is likely that the increase in overall weight gain in this group is due to other factors, not the loss of a single copy of $Sim2$ in this group.

This is the first study investigating the effect of a high fat diet in a conditional $Sim2$ KO mouse model. Consistent with previously reported findings in heterozygous $Sim2$ KO mice (Goshu et al., 2002), these results confirm that $Sim2$ does not appear to play a role in satiety signalling, unlike its closely related homolog $Sim1$.
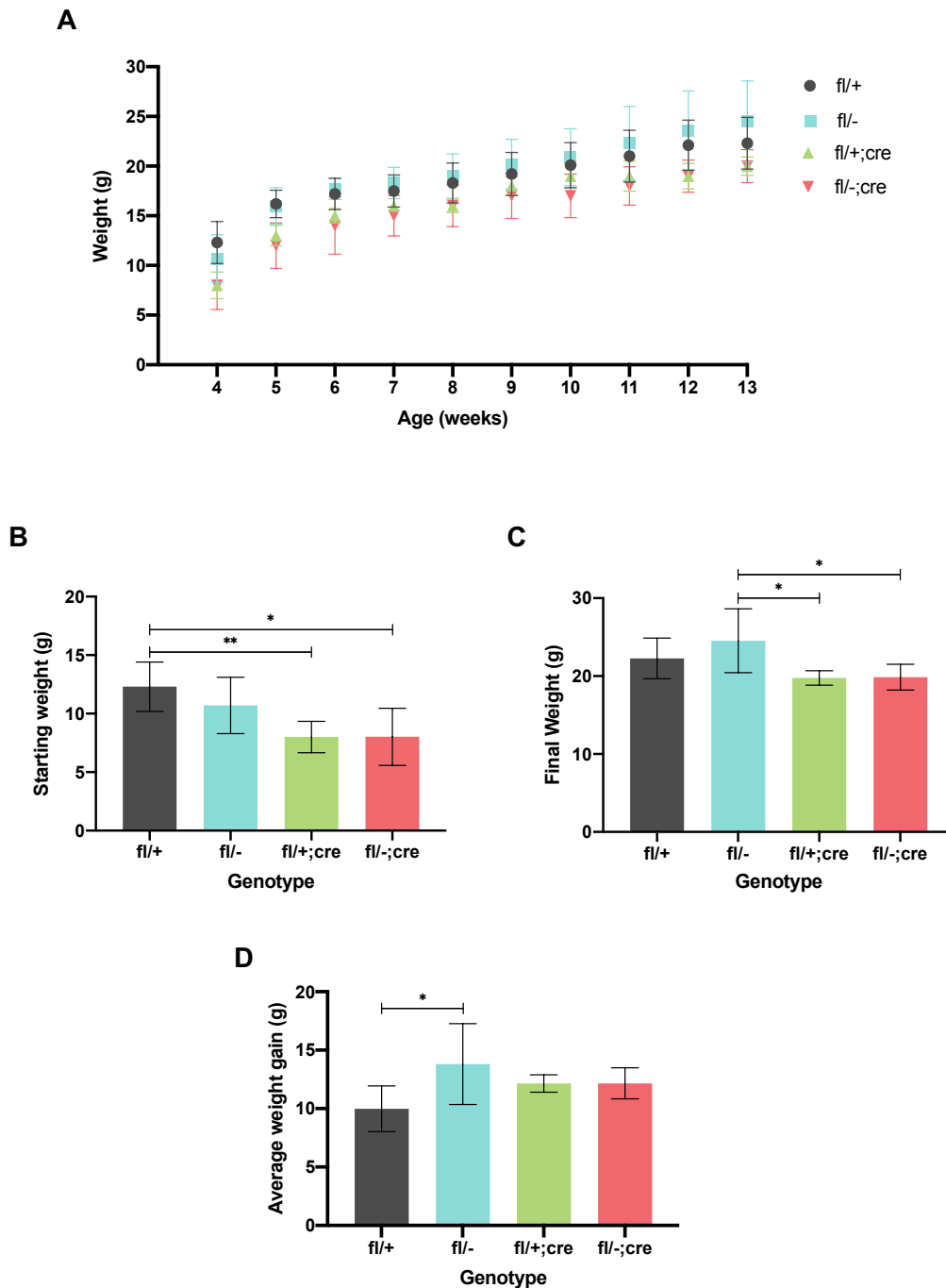
Figure 6.5: *Sim2*-Nestin KO feeding study. **A)** Weekly weight measurements of mice on a high fat diet. Data points represent the mean and standard deviation for each genotype group. The average starting weight at 4 weeks of age (**B**), final body weight at 13 weeks of age (**C**) and weight gain across the study (**D**) was calculated. Data represented as the mean and standard deviation for each genotype group. Statistical significance was measured by one-way Annova. * $p < 0.05$, ** $p < 0.01$.

6.4.2 Generation and characterisation of a SIM2-3xFLAG mouse model

6.4.2.1 Genome editing to insert a 3xFLAG tag at the endogenous *Sim2* locus

Based on previous experiments conducted in the Whitelaw laboratory, there are no antibodies that will reliably detect endogenous SIM2 protein. This poses a problem for analysis of the function and expression of the endogenous protein *in vivo*. In order to attempt to overcome this problem, we decided to generate a knock-in mouse model with the endogenous SIM2 protein tagged with an epitope tag. A triple FLAG (3xFLAG) epitope tag was chosen as it was successfully used to tag endogenous neural transcription factors by our collaborators (Professor Paul Thomas, personal communication). Given that the basic DNA binding region is located at the N-terminus of the SIM2 protein and the FLAG epitope contains a number of acidic residues, the C-terminus was chosen to be tagged to ensure endogenous protein function was not disrupted. There are two isoforms of SIM2 (SIM2-s, SIM2-l), however as the current literature is focused mostly on the long isoform of mouse SIM2, only the long isoform was chosen to be tagged. Additionally, given that the short isoform is generated through a non-canonical splicing event, it was predicted that introducing the tag sequence at the C-terminus of the short isoform may disrupt alternative splicing of the *Sim2* transcript.

An overview of the strategy used to generate this mouse model can be seen in Figure 6.6. The guide RNA was designed using the format of GN_{19}-NGG that was standard for the *S. pyogenes* Cas9 (spCas9) endonuclease (Adli, 2018; H. Yang et al., 2014). The guide was designed to target the *Sim2* gene as close as possible to the endogenous stop codon within these design constraints, with the Cas9 nuclease cleavage site being 10 nucleotides upstream of the stop codon. The ssDNA oligonucleotide repair template was designed to have 60bp of homology to either side of the Cas9 cleavage site flanking the sequence for the 3xFLAG epitope. The 3xFLAG sequence was inserted directly before the endogenous stop codon to avoid loss of any upstream amino acids. Nine silent mutations were introduced into the gRNA recognition site and PAM sequence in order to prevent re-cleavage by the Cas9 endonuclease at sites where the 3xFLAG sequence had been inserted. *In silico* translation of the sequence resulted in the

predicted amino acid sequence and stop codon location for the targeted allele. The guide RNA sequence was cloned into the pX330 vector according to standard protocols and in vitro transcribed to generate gRNA for injection. The ssDNA donor oligonucleotide was commercially manufactured (Sigma Aldrich).

A microinjection mix was prepared that contained the ssDNA oligo (100 ng/ul), sgRNA (50 ng/ul) and Cas9 mRNA (100 ng/ul) (Figure 6.6D). Microinjection of this mix into mouse zygotes, and transfer of the injected zygotes to a pseudo-pregnant C57BL/6 female was performed by Sandra Piltz (SA Genome Editing Facility) (H. Yang et al., 2014). From this a total of 9 mice were born.

PCR genotyping was first performed in order to determine if there were any mice that potentially contained the modified 3xFLAG allele (Figure 6.6B). Primers were designed to specifically anneal outside of the targeted region and would produce a PCR product of 649bp for a WT allele and 715bp for a correctly targeted allele. Figure 6.6B shows the genotyping PCR results for the 9 mice born. The WT control gave a band at approximately the expected size as well as a larger faint non-specific band. Three of the 9 mice born showed a PCR product of approximately the predicted size for the 3xFLAG targeted allele (Figure 6.6B, mice 4, 7 and 9). The PCR genotyping analysis also showed 7/9 mice to harbour some form of mutation (3xFLAG integration or other) around the target site, shown by either the lack of a WT band, or additional bands not corresponding to the WT allele (Figure 6.6B, D). At least three of the mice appeared to not have a PCR product corresponding to the WT *Sim2* allele suggesting deletion encompassing at least one of the primer binding sites. Four also appeared to have PCR products larger than expected for the WT or 3xFLAG allele. This could be due to an insertion at the CRISPR target site.

PCR products predicted to correspond to the 3xFLAG allele from the 3 mice were sequenced (Figure 6.6C), confirming that all three had in-frame insertion of the 3xFLAG sequence. Founders 7 and 9 were then crossed with WT C57BL/6 mice in order to establish a colony and breed out the non-targeted allele, which would likely have in/dels due to targeting by CRISPR/Cas9. Mice from these matings heterozygous for the *Sim2*-3xFLAG allele were crossed in order to generate homozygous *Sim2*-3xFLAG

mice. Mouse 4 was shown by sequencing to have a non-synonymous mutation within the FLAG sequence so was not used for colony establishment.

Overall this was an efficient strategy for generation of the *Sim2*-3xFLAG allele with only one round of zygote injection and transfers required to generate founder mice with the correctly inserted sequence. Consistent with reported success rate of inserting short DNA sequences using a single stranded oligonucleotide (Hui Yang et al., 2013), approximately 30% of mice born (3 out of 9) harboured the integrated donor DNA sequence, with approximately 20% (2 out of 9) containing the the correctly targeted allele (Figure 6.6).
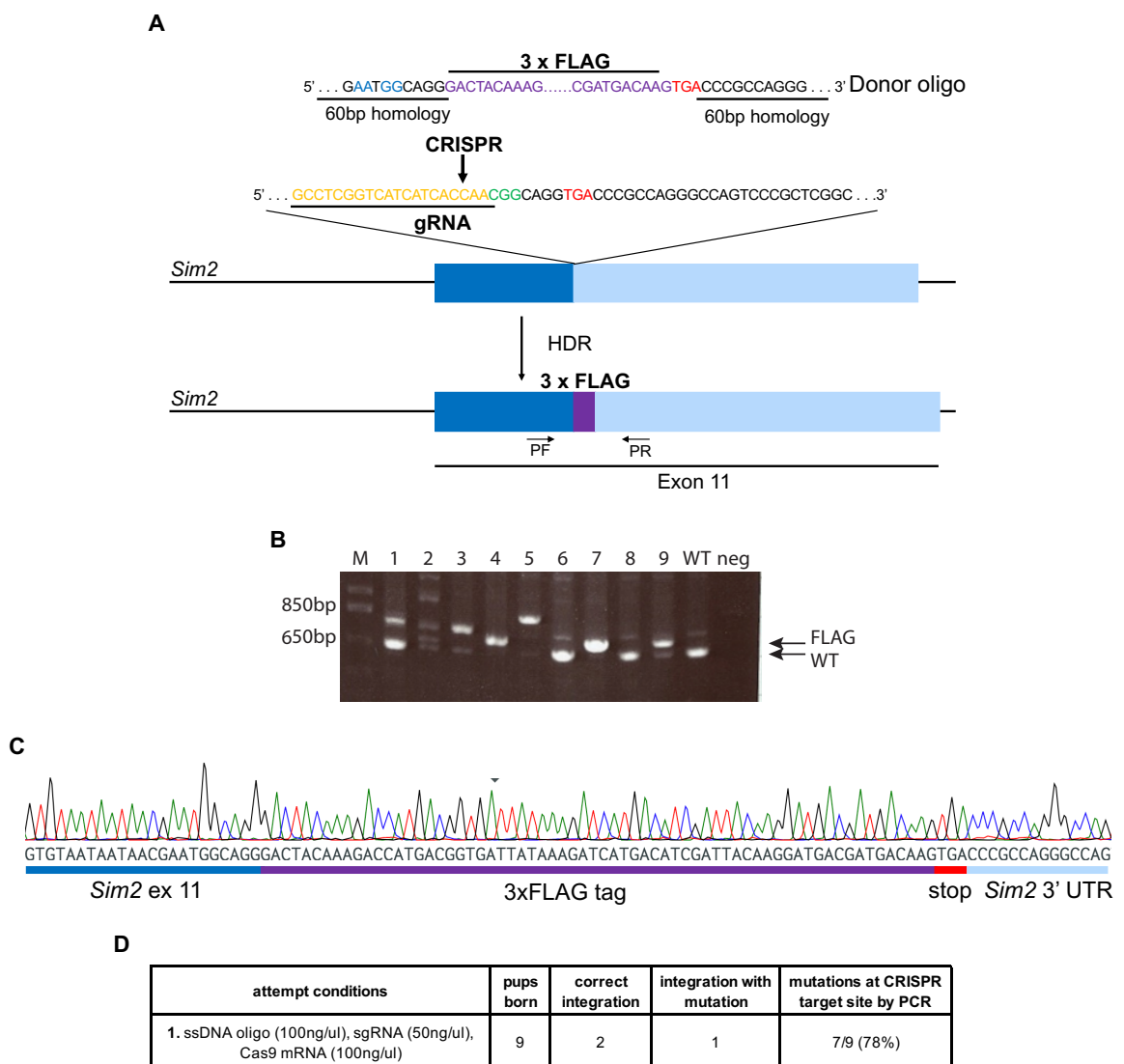

Figure 6.6: Generation of the SIM2-3xFLAG mouse model. **A)** Strategy for generating SIM2-3xFLAG mouse line. Genotyping primers (PF, PR) shown on targeted exon schematic.

6.4.2.2 Characterisation of the SIM2-3xFLAG mouse model

A number of methods were employed to attempt to validate this mouse line as a tool for investigating the expression and function of SIM2 in an *in vivo* context. Initially, detection of SIM2-3xFLAG protein was attempted through multiple immunoblotting based assays from tissue extracts of both heterozygous (*Sim2*^{FLAG/+}) and homozygous (*Sim2*^{FLAG/FLAG}) Sim2-3xFLAG newborn mouse kidney and brain tissue extracts. These included an α-FLAG western blot from newborn *Sim2*^{FLAG/+} mouse brain and kidney extracts, an α-FLAG immunoprecipitation followed by α-FLAG western blot from newborn *Sim2*^{FLAG/+} mouse brain and kidney extracts, an α-FLAG immunoprecipitation followed by α-SIM2 western blot and α-SIM2 immunoprecipitation followed by α-FLAG western blot on *Sim2*^{FLAG/FLAG} newborn mouse kidney extracts (the α-SIM2 antibody (21069-1-AP, Proteintech) had previously been successfully used for SIM2 IP experiments in the Whitelaw Laboratory (Dr. Adrienne Sullivan, unpublished data)). Very disappointingly, there was no detectable SIM2-3xFLAG protein identified in any of these immunoblotting experiments (data not shown).

In a second approach, primary kidney cell cultures were derived from four-month-old *Sim2*^{FLAG/+} and *Sim2*^{+/+} mice in order to assess SIM2-3xFLAG expression and their viability as an *in vitro* system for investigating SIM2 function. The establishment of these cells was based on a published protocol (Valente et al., 2011). Cells were grown on coverslips and immunofluorescence was performed to determine if SIM2-3xFLAG expression could be detected in these cells. As can be seen in Figure 6.7, a high level of non-specific staining can be seen in both the *Sim2*^{FLAG/+} and *Sim2*^{+/+} cells stained with α-FLAG antibody. There was little to no background seen for the cells stained with secondary antibody only, indicating that the background signal was due to the mouse α-FLAG primary antibody.
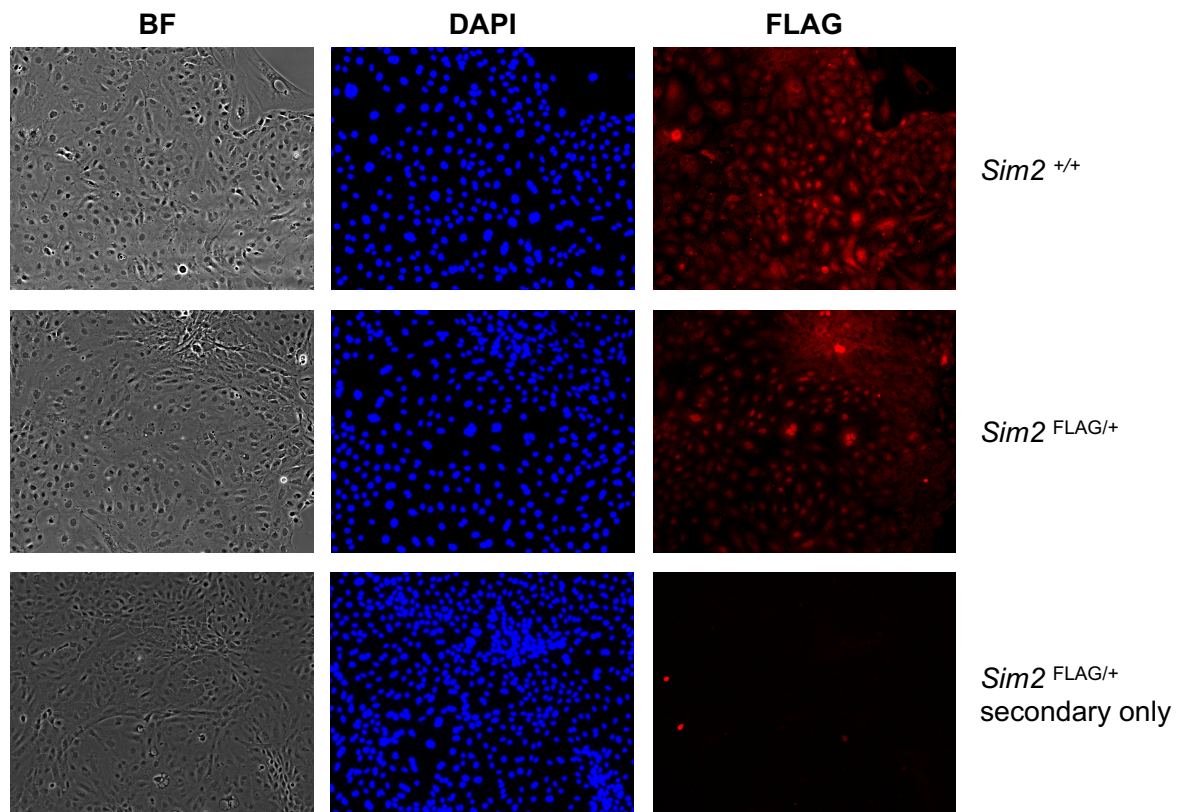
Figure 6.7: Immunofluorescence of primary kidney cell cultures derived from Sim2[+/+] or Sim2[FLAG/+] mice. Left panel are brightfield (BF) images of primary kidney cells. Middle pane is DAPI nuclear staining and right panel is α-FLAG.

One of the planned experiments for the SIM2-3xFLAG mouse line was to perform a comprehensive SIM2 protein expression pattern analysis through immunofluorescence. Given the high level of background staining seen in the IF experiments on the primary kidney cell cultures, mice were first perfused with PBS to remove blood before fixation with 4% formaldehyde. This process would be expected to reduce the amount of cross-reactive immunological material that would cause high levels of non-specific staining in IF as a result of using an α-mouse secondary antibody. A *Sim2*[+/+] and *Sim2*[FLAG/FLAG] mouse was perfused before dissecting and sectioning the kidneys. Kidney sections were stained with α-FLAG ab to look for SIM2-3xFLAG expression. Figure 6.8 shows representative images from the IF experiment. Some non-specific staining can be observed around the outer layer of the kidney tissue and a low level of background non-specific staining seen in both the *Sim2*[+/+] and *Sim2*[FLAG/FLAG] kidney sections. However there does not appear to be any specific staining of SIM2-3xFLAG seen in the kidney sections.
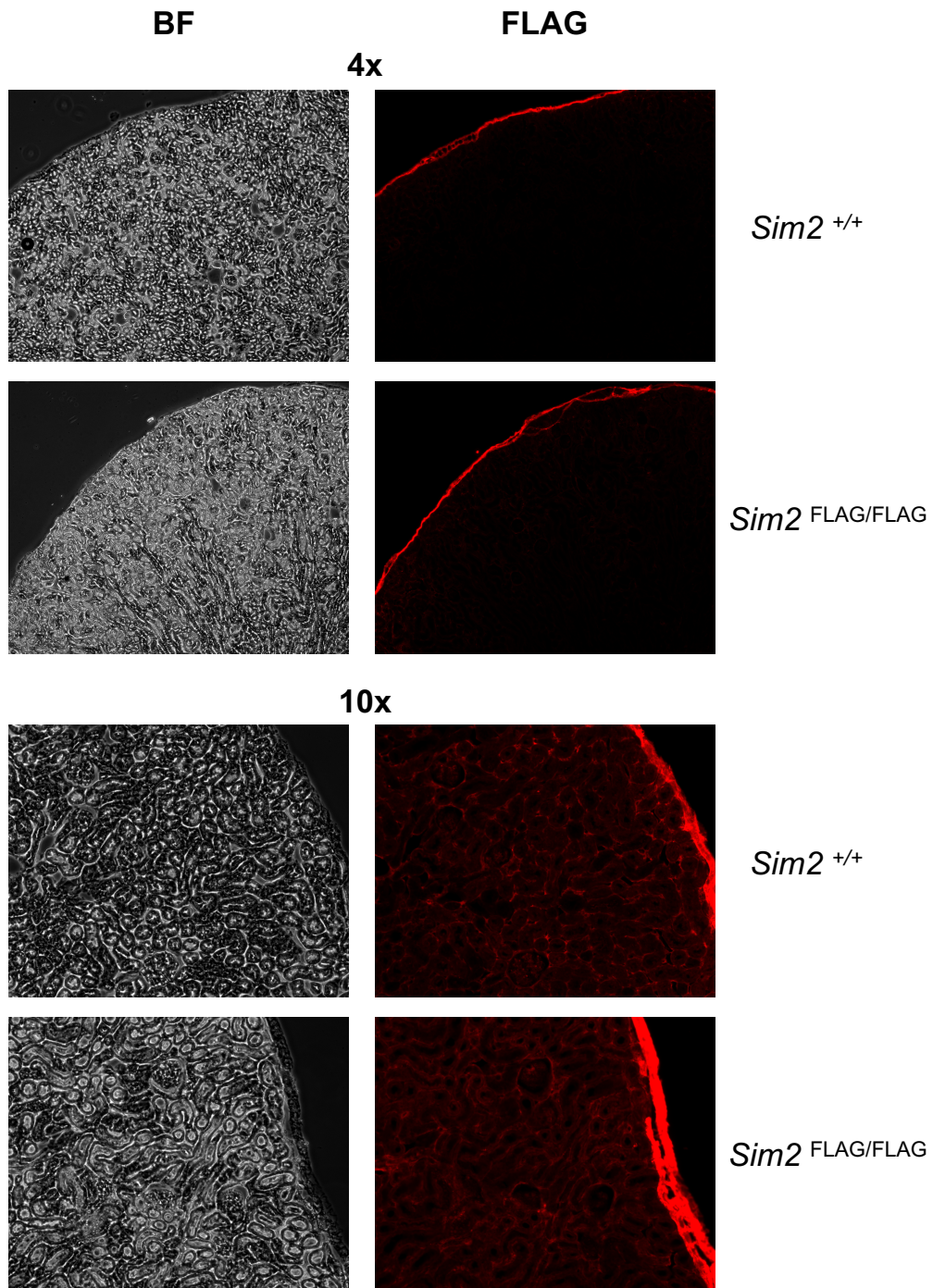
Figure 6.8: Kidney section IF. *Sim2*[+/+] and *Sim2*[FLAG/FLAG] kidney sections probed with α-FLAG antibody and viewed under 4x or 10x magnification.

6.4.2.3   Analysis of *Sim2* transcripts

Given that detection of the SIM2-3xFLAG protein was unsuccessful across a number of alternative methods, detection of the *Sim2-3xFLAG* transcript was attempted. Multiple RT-PCR experiments with different primers both specific to the tag sequence and common to both the tag and WT sequence were designed and attempted, however detection of the *Sim2-3xFLAG* transcript by this method was unsuccessful in kidney extracts from *Sim2*$^{FLAG/+}$ mice (data not shown). To investigate further, RT-PCR analysis was performed on RNA extracted from *Sim2*$^{FLAG/+}$ and *Sim2*$^{+/+}$ kidney tissue in order to determine whether *Sim2* is being expressed in this tissue, and the relative expression of the two isoforms. RT-PCR for both the short and long isoforms of *Sim2* was performed using primers previously been shown to detect these two isoforms (Kwak et al., 2007; Metz et al., 2006). As can be seen in Figure 6.9, both isoforms of *Sim2* appear to be expressed at comparable levels in both *Sim2*$^{FLAG/+}$ and *Sim2*$^{+/+}$ kidney extracts. Sequencing of *Sim2s* transcript aligns to the predicted sequence, with read through from exon 10 into intron 10, consistent with reported literature and what is seen in the human *SIM2s* transcript (Figure 6.10) (Chrast et al., 1997; Metz et al., 2006).

Given this, the failure to detect the 3xFLAG transcript was likely due to suboptimal primers as detection of both the long and short isoform of *Sim2* was successful using published primer sequences (Kwak et al., 2007; Metz et al., 2006). Levels of the expressed long and short isoform mRNA do not appear to differ when comparing WT, heterozygous and homozygous *Sim2*-3xFLAG mice, suggesting that the insertion of the tag sequence does not affect alternative splicing of *Sim2* mRNA, and that the modified long isoform is not subject to nonsense mediated decay or some other form of increased mRNA turnover (Figure 6.9).
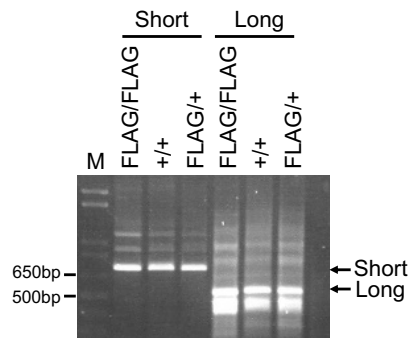
Figure 6.9: RT-PCR analysis of *Sim2s* and *Sim2l* transcripts in mouse kidney tissue from *Sim2*FLAG/FLAG, *Sim2*FLAG/+ and *Sim2*+/+ mice. Arrows depict PCR bands corresponding to the respective transcripts.

2102
AGTGGCACTGTGCCAACTCTCTAGTGCCCAGCAGCTCATCGCCAGCTAAAAACCTTTCTGAGCCTTCTCCTGTGAATGCTGCCCGGCATGGTCTTGTGCC

Exon10
template sequence mSim2s cDNA (predicted)

AGTGGCACTGTGCCAACTCTCTAGTGCCCAGCAGCTCATCGCCAGCTAAAAACCTTTCTGAGCCTTCTCCTGTGAATGCTGCCCGGCATGGTCTTGTGCC
aligned sequence Sim2s cDNA fwd seq

AGTGGCACTGTGCCAACTCTCTAGTGCCCAGCAGCTCATCGCCAGCTAAAAACCTTTCTGAGCCTTCTCCTGTGAATGCTGCCCGGCATGGTCTTGTGCC
aligned sequence Sim2s cNDA rev seq

Exon-Intron boundary

2202
AAACTATGAAGGTAGGTCCCATTTACAGGAGGGGAAGGGCAGGGGTTCAGCCTCAGTAGCAGTGGGTGGTGGATGGAAAAAGAATCCTCATGCTTTGGGC

Intron 10
Exon10 | Ex10 extention (short isoform)
template sequence mSim2s cDNA (predicted)

AAACTATGAAGGTAGGTCCCATTTACAGGAGGGGAAGGGCAGGGGTTCAGCCTCAGTAGCAGTGGGTGGTGGATGGAAAAAGAATCCTCATGCTTTGGGC
aligned sequence Sim2s cDNA fwd seq

AAACTATGAAGGTAGGTCCCATTTACAGGAGGGGAAGGGCAGGGGTTCAGCCTCAGTAGCAGTGGGTGGTGGATGGAAAAAGAATCCTCATGCTTTGGGC
aligned sequence Sim2s cNDA rev seq

Sim2s stop codon

2302
AAACTGGCCTCTTCTGCTTCTAAGCTGGGATTTCTGTGCTTCCTACTGCCCATCATGGTTTCATGGGCGTGAGTTACATATAACAAGAACACCGGAGATG
stop

Intron 10
Ex10 extention (short isoform)
template sequence mSim2s cDNA (predicted)

AAACTGGCCTCTTCTGCTTCTAAGCTGGGATTTCTGTGCTTCCTACTGCCCATCATGGTTTCATGGGCGTGAGTTACATATAACAAGAACACCGGAGATG
aligned sequence Sim2s cDNA fwd seq

AAACTGGCCTCTTCTGCTTCTAAGCTGGGATTTCTGTGCTTCCTACTGCCCATCATGGTTTCATGGGCGTGAGTTACATATAACAAGAACACCGGAGATG
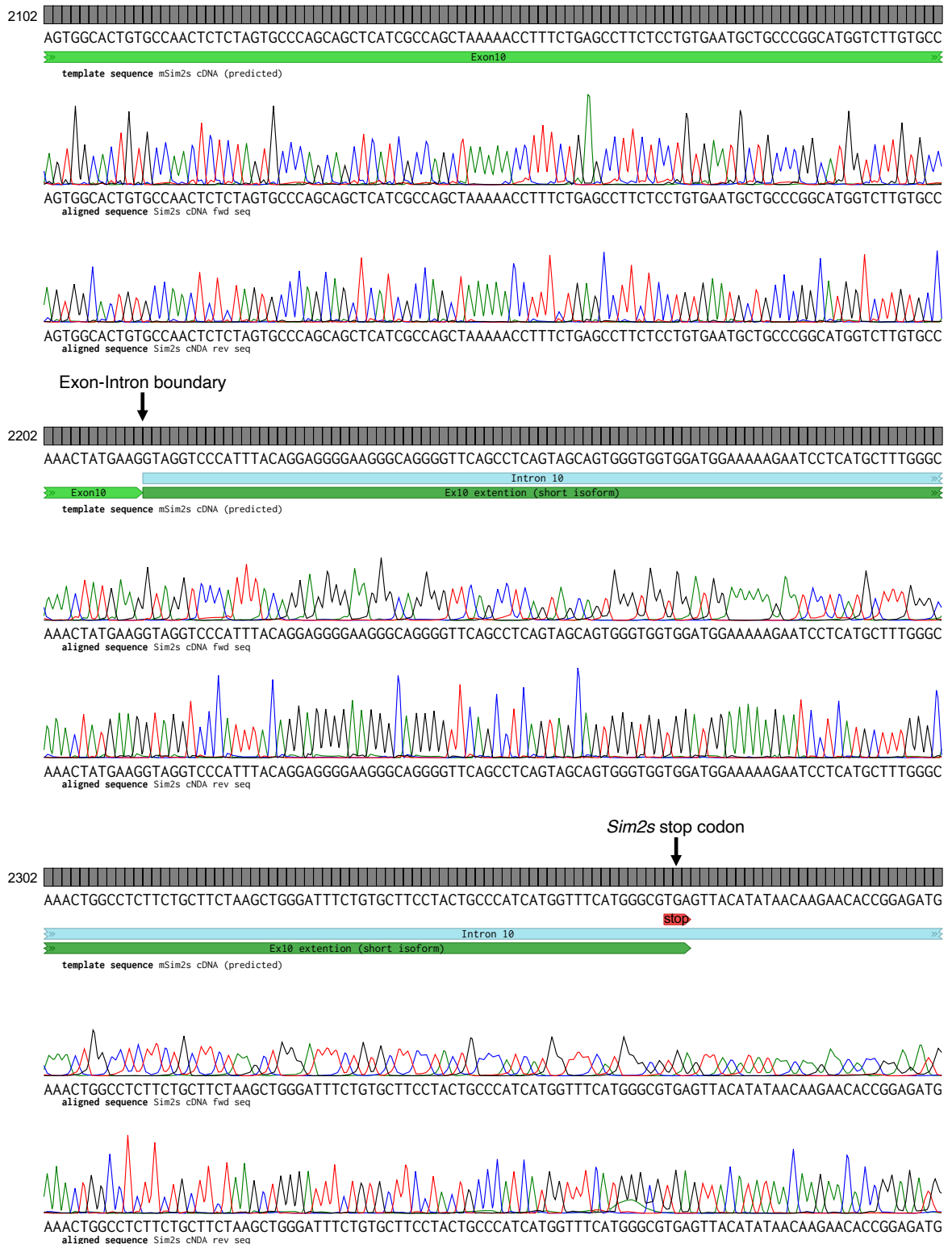**aligned sequence** Sim2s cNDA rev seq

Figure 6.10: Sequencing in both the forward (top sequencing trace) and reverse (bottom sequencing trace) of the *Sim2s* cDNA aligned to the m*Sim2* genomic DNA as a reference sequence. Exon 10, intron 10, the exon 10 extension specific to the *Sim2s* isoform and the *Sim2s* predicted stop codon are depicted below the reference

sequence. Reverse sequencing trace visualised as complement by alignment program (Benchling [Biology Software]. (2022). Retrieved from https://benchling.com)

Despite successful in frame integration of the 3xFLAG epitope tag at the endogenous *Sim2* locus, detection of the SIM2-3xFLAG protein was not successful. Multiple experimental detection methods were used in attempt to visualise SIM2-3xFLAG protein expression including western blot, immunoprecipitation and immunofluorescence. There are a number of possible explanations for the inability to detect expression of the SIM2-3xFLAG protein. Given SIM2 is a lowly expressed protein which is very difficult to detect in cells or tissues by commercial antibodies, the 3xFLAG epitope tag in combination with the chosen antibody may likewise not provide enough sensitivity to detect endogenous SIM2 protein using the explored methods. Additionally, SIM2 is a labile protein when ectopically expressed in cell lines (S. Woods et al., 2008), and therefore endogenous SIM2 may not survive the processing protocols used in the detection methods.

6.4.3 Attempts at generating a SIM2-Tomato reporter mouse model

Multiple attempts were made at generating a *Sim2*-Tomato fluorescent reporter mouse line. These attempts all used a donor DNA sequence that included homology regions specific to the *Sim2* gene flanking the coding sequence for the Tomato fluorescent protein. The aim was to replace expression of *Sim2* with the Tomato fluorescent protein by removing the coding sequence of *Sim2* exon1 and replacing it with the coding sequence for the Tomato protein, followed by a nuclear localisation sequence (NLS) and poly adenylation (poly A) signal. The NLS was included in order to concentrate the fluorescent signal in the nucleus which would hopefully assist with visualisation given that *Sim2* is a lowly expressed transcript (C.M. Fan, personal communication) (Fan et al., 1996).

6.4.3.1 Targeting vector donor DNA cloning

A number of the attempts at generating the *Sim2*-Tomato mouse line used a plasmid vector as the donor DNA for homologous recombination. This vector was cloned

through the modification of the mSim2-flox-pWS-TK6 targeting plasmid which was used to generate the *Sim2*-floxed mouse line (cloned by Murray Whitelaw). Multiple different strategies were designed and attempted; the successful strategy only will be detailed.

This vector was modified through restriction endonuclease digestion, ligation, PCR amplification and Gibson isothermal assembly to remove the *Sim2* exon1 coding sequence and replace it with the Tomato coding sequence. A schematic diagram of this vector can be seen in Figure 6.11A. The cloned vector (mSim2-Tomato-pWS-TK6) included two large *Sim2* homology regions flanking the coding sequence for the Tomato fluorescent protein followed by a NLS sequence and SV40 polyA. Immediately following the Tomato sequence was the Neomycin resistance selection marker flanked by FRT sites to allow for removal through flp-mediated recombination. The *Sim2* sequence is conserved until the endogenous start codon, at which point the native sequence is replaced with the nuclear Tomato construct. The targeting plasmid also contains thymidine kinase (TK) and ampicillin resistance (Amp^R) selection markers as well as a p15A origin of replication (p15A Ori) for propagation in bacteria (Figure 6.11A). The entire targeting vector is approximately 20kb in size.

As the CRISPR system was to be used to facilitate homologous recombination, a gRNA was designed to target *Sim2* intron 1, as close as possible to the region that will be replaced with the Tomato coding sequence (Figure 6.11B). This guide was cloned into the pX330 plasmid (pX330-mSim2).

6.4.3.2 Targeting the *Sim2* locus in embryonic stem cells

Generation of the *Sim2*-Tomato allele was first attempted in mouse embryonic stem (mES) cells using the cloned targeting vector to assess whether this strategy could be successful. mES cells were electroporated with 25ug of the CRISPR expression plasmid px330-mSim2 and 25ug of the mSim2-dnucTomato-pWS-TK6 targeting vector and then seeded into plates containing MEF feeder cells layer. Following treatment with 2uM ganciclovir for TK selection and 100uM G418 (geneticin) for positive neomycin selection individual colonies were isolated and transferred to 96 well tray for

monoclonal colony growth. A genotyping PCR was performed using a forward primer specific to the endogenous *Sim2* gene and a reverse primer specific to the Tomato sequence on DNA extracted from individual clones to determine which clones had been successfully targeted and contained the *Sim2*-Tomato allele (Figure 6.11B P1 + P3). As can be seen in Figure 6.12, 9 of the 12 screened colonies contained a PCR product at the expected size of 2.3kb for the *Sim2*-Tomato allele. The Tomato sequence and surrounding regions were sequenced to confirm correct integration (data not shown).

Generation of the *Sim2*-Tomato allele in ES cells using a large targeting vector and CRISPR/Cas9 to induce a double stranded DNA break at the integration site was an efficient strategy for generating the *Sim2*-Tomato allele with 75% of clones containing the target construct. However, the process to generate a founder mouse from this point is time consuming and would require multiple rounds of breeding to obtain a mouse line with pure genetic background. Given the success of our recombination strategy in ES cells, a more direct CRISPR method was therefore employed in attempts to generate the *Sim2*-Tomato mouse line.
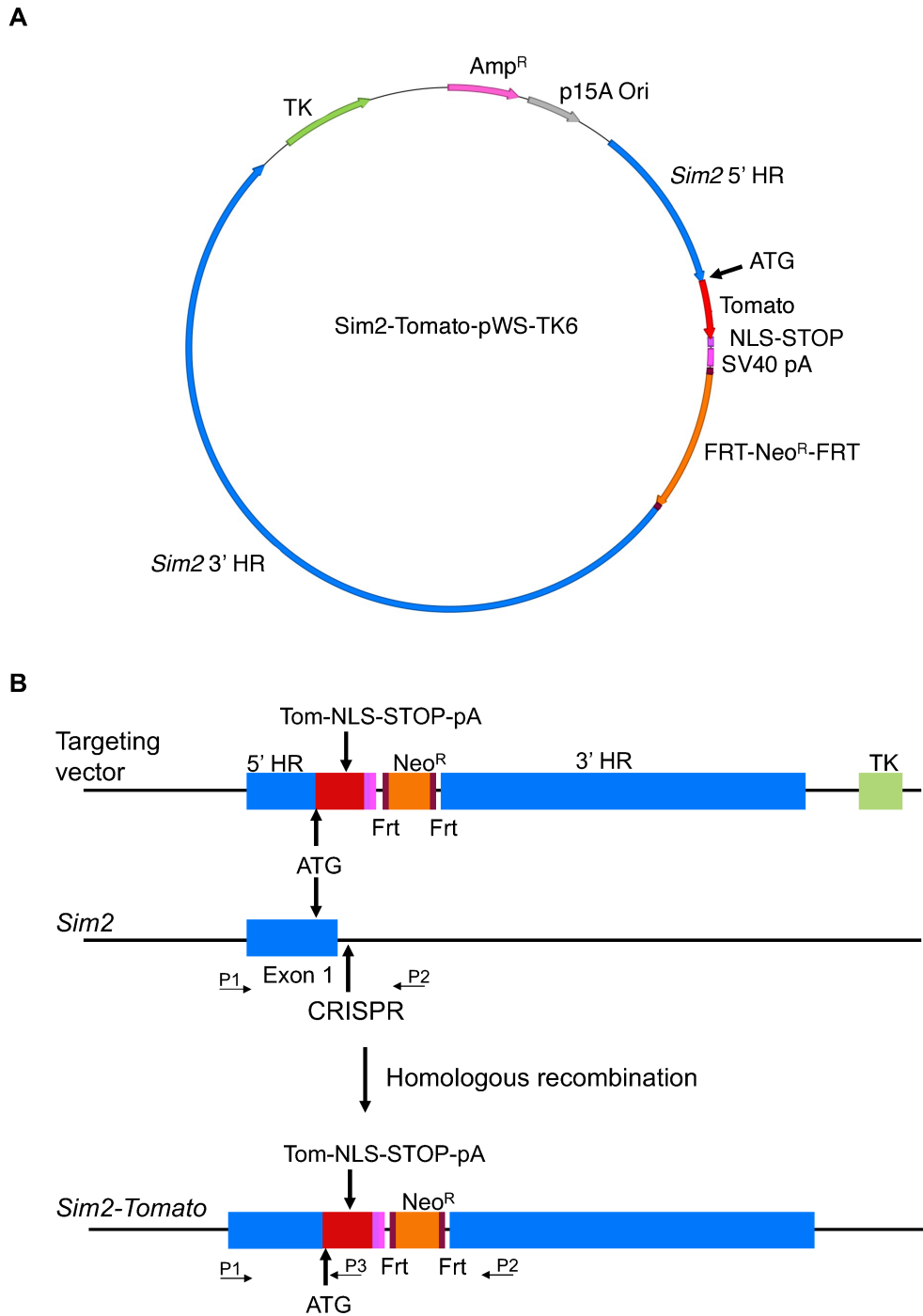
Figure 6.11: Large targeting vector strategy for generation of *Sim2-Tomato* mouse line. **A)** Schematic of the Sim2-Tomato-pWS-TK6 vector that was cloned. **B)** Strategy for generating the *Sim2*-Tomato allele using the large targeting vector. P1, P2, P3 are the approximate locations of primers used for genotyping. TK: thymidine kinase, Amp[R]: ampicillin resistance, p15Ori: p15A origin of replication, HR: homology region, NLS: nuclear localisation signal, pA: poly adenylation signal, FRT: Flp recombination site, Neo[R]: neomycin resistance, ATG: start codon.
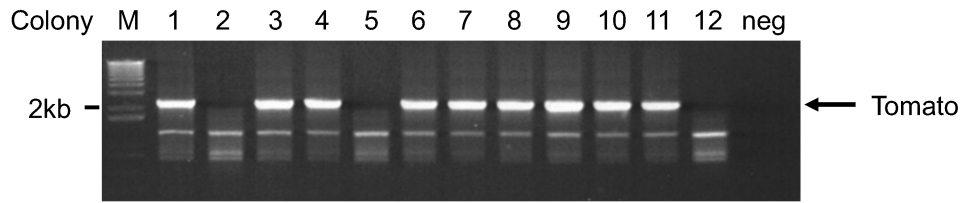
Figure 6.12: Genotyping potential *Sim2*-Tomato ES cell colonies. The presence of the successfully targeted *Sim2*-Toamato allele results in amplification of a 2.3kb PCR product. M: marker.

6.4.3.3 Attempts at targeting in mouse zygotes

6.4.3.3.1 Large targeting vector

Given that generating the *Sim2*-Tomato targeted allele was highly successful in mES cells, this method was attempted in mouse zygotes. Cytoplasmic injection into mouse zygotes of a mix containing the sgRNA at 50ng/μL, 20kb targeting vector donor DNA at 200ng/μl and spCas9mRNA at 100ng/μl was performed by Sandra Piltz (SA Genome Editing Facility). A total of 20 mice were born across two separate injection rounds.

Two genotyping PCRs were used to screen for any mice that may have correct integration of the targeting vector. The first PCR used primers either side of the targeted region that would give a 0.7kb product for the WT and 3.2kb product for *Sim2*-Tomato allele (Figure 6.11B P1 + P2). The second used primers that were specific to the exogenous Tomato sequence that would amplify a 1.2kb product if the Tomato sequence was present (Figure 6.11B P1 + P3). The results of both genotyping PCRs showed that there were zero mice that had the correctly integrated *Sim2*-Tomato allele, with the first PCR only amplifying the WT allele, and no products seen for the second (data not shown). Therefore the attempt to generate the *Sim2*-Tomato mouse line using donor plasmid was unsuccessful. Based on published data using this method, a success rate of approximately 10-30% would have been expected (Hui Yang et al., 2013).

A heteroduplex assay was used to test for the efficiency of the CRISPR guide. The PCR first genotyping PCR which produced a 0.7 kb product for the WT allele was denatured and then re-annealed, followed by separation by polyacrylamide gel electrophoresis.

An example of this experiment can be seen in Figure 6.13E. The PCR from a WT mouse with no mutations at the CRISPR target site contains only one species, which is seen as one band on the polyacrylamide gel. A PCR from mice where small in/del mutations have been generated at the CRISPR cut site will generate multiple species, some of which will have imperfect bp annealing, causing retardation of migration of this species through the polyacrylamide gel, which can be seen as multiple bands on the gel. From this strategy, 15 out of the 20 (75%) mice born had mutations at the CRISPR target site as indicated by the presence of multiple PCR species in the heteroduplex assay (Figure 6.13.E, data not shown). Therefore the gRNA used in this method was efficient at targeting double stranded DNA breaks at this location, indicating that the reason for failure to generate the mouse line was not due to an ineffective guide RNA.

6.4.3.3.2 Heteroduplex Donor-Assisted Direct Integration HD-ADI

Given that the initial attempts at generating the *Sim2*-Tomato mouse line in zygotes through the use of a large targeting vector were not successful, a new strategy was attempted. This strategy, Heteroduplex Donor-Assisted Direct Integration (HA-ADI) was a method for generating Knock-In (KI) mouse lines hypothesised by a colleague Dr Fatwa Adikusuma (University of Adelaide). This work was undertaken as a collaboration to collect data to determine if this would be a viable strategy for generating KI mouse models and attempt to generate the *Sim2*-Tomato mouse line. For this strategy the gRNA is designed to cut exactly at the modification site, with insertion of the donor DNA destroying the CRISPR target site. The donor DNA is a dsDNA PCR with short 30nt ssDNA tails complementary to the genomic DNA at the insertion site. This system takes advantage of the cellular DNA break repair mechanism microhomology-mediated end joining (MMEJ) (H. Wang & Xu, 2017). This process involves 5'-3' resection of DNA at the break site, introducing microhomology regions which then anneal, and gaps are filled in through DNA synthesis and ligation to reform dsDNA. In this strategy, regions of microhomology will be provided by the donor DNA, which will anneal to the complementary gDNA at the resected sites, introducing the exogenous Tomato DNA sequence in the process (Figure 6.13B).

The strategy for generating the *Sim2*-Tomato mouse line used a CRISPR guide that would generate a double stranded DNA break in exon 1, 10nt upstream of the ATG start

codon. The donor was generated through PCR amplification of the *Sim2* Kozak sequence followed by the Tomato coding sequence with 30nt tails complementary to the gDNA site of insertion either side. There were two alternate methods for generating the donor DNA. For method 1; the Kozak-Tomato sequence was amplified by two different PCRs, one that added the 30bp complementary tail to the 5' end of the PCR product, the other added the 30bp complementary tail to the 3' end of the PCR product. These two PCR products were then mixed, denatured and re-annealed to generate species that will have single stranded tails either side of the Kozak-Tomato sequence, one at the 5' and one at the 3' end on opposite strands. Note that in this mix the original PCR product species will also be present, so there will be 4 alternative species in total, two species with single stranded tails and two dsDNA species. For method 2; the Kozak-Tomato sequence was again amplified by two different PCRs, one that added the 30bp complementary tails on both the 5' and 3' end of the product, the other was only the Kozak-Tomato sequence with no sequence complementary to the site of insertion. These two PCR products were then mixed, denatured and re-annealed to generate species that will have single stranded complementary tails on both the 5' and 3' end on either the top or bottom strand. Note that in this mix the original PCR product species will also be present, resulting in 4 alternative species, two of which will have single stranded complementary tails, two will be double stranded.

A mix containing the sgRNA at 50ng/µl, PCR template mix at 100ng/µl and spCas9 mRNA at 100ng/µl was injected into the cytoplasm of mouse zygotes, followed by transfer of the injected zygotes to a pseudo-pregnant C57BL/6 female (Sandra Piltz, SA Genome Editing Facility). From this strategy, a total of 28 mice were born.

Two genotyping PCRs were used to screen for the presence of the *Sim2*-Tomato allele. The first used primers either side of the targeted region that would give a 0.7kb product for the WT and 1.8kb product for *Sim2*-Tomato allele (Figure 6.13B, P1 + P2). The second used a forward primer upstream of the target site and a reverse primer specific to the exogenous Tomato sequence that would give a 0.9kb product only if the Tomato sequence was present (Figure 6.13.B, P1 + P3). The results of both of these genotyping PCRs showed that there were zero mice with the *Sim2*-Tomato allele present, with PCRs for all mice generating only a product for the WT allele, and no

product for the Tomato allele. An example of the first PCR can be seen in Figure 6.13D. The PCR from mice 1-9 as well as the WT control only contained a 0.7kb PCR product for the WT allele. The positive control plasmid (mSim2-dnucTomato-pWS-TK6) gave the expected 3.2kb product. A heteroduplex assay was performed using the first genotyping PCR to determine the efficiency of the gRNA. Only 4 out of 28 (14%) of mice had mutations at the CRISPR cut site as shown by multiple species seen in the heteroduplex assay, indicating that this gRNA is not efficient (data not shown).

Given that the HD-ADI method has not been previously reported or shown to be an effective strategy at generating knock-in mouse models, it is unclear whether this method did not work due to the specifics of this experiment, or if it is just not a viable method for generating knock-in alleles. The gRNA did not appear to be optimal, given that only 14% of mice born appeared to contain mutations at the target site. Alternative gRNA sequences and target sites were not an option given the requirement for the guide to target the exact location of sequence insertion (Figure 6.13A). This may have been a contributing factor to the lack of success of this method. However given the uncertainty of this method, it was not explored any further as an option for generating the *Sim2*-Tomato mouse line.

6.4.3.3.3   Long ssDNA oligo

A third and final strategy was employed in an attempt to generate the *Sim2*-Tomato mouse line. This strategy was based on a method which uses a long ssDNA oligonucleotide as the donor DNA template for HDR and CRISPR to generate a double stranded DNA break at the point of insertion to increase the efficiency of this repair (Figure 6.13C) (Li et al., 2017; Quadros et al., 2016).

A guide RNA was designed and cloned that would generate a double stranded DNA break at the endogenous *Sim2* ATG start codon. The donor DNA was designed to include 60bp homology sequences flanking the Tomato coding sequence including a nuclear localisation signal and SV40 polyA sequence. Pronuclear injection of a mix containing spCas9 protein at 20 ng/µl, ssDNA donor oligo at 10 ng/µl, the sgRNA at 20 ng/µl was performed by Sandra Piltz (SA Genome Editing Facility). From this attempt 7 mice were born. Two genotyping PCRs were used to screen for insertion of the

Tomato sequence. The first used primers either side of the insertion site that would give a 0.7 kb product for the WT allele and a 1.8 kb product for the *Sim2*-Tomato allele (Figure 6.13C P1 + P2). The second used a forward primer upstream of the target site and a reverse primer specific to the exogenous Tomato sequence that would give a 0.9kb product only if the Tomato sequence was present (Figure 6.13C, P1 + P3). The results of both of these genotyping PCRs showed that there were zero mice with the *Sim2*-Tomato allele present, with PCRs for all mice generating only a product for the WT allele, and no product for the Tomato allele (data not shown). A heteroduplex assay was used to determine the efficiency of the gRNA. 2 out of the 7 (29%) mice had mutations at the CRISPR cut site as seen by multiple bands in the heteroduplex assay indicating that this is a relatively inefficient gRNA (data not shown).

The long ssDNA oligonucleotide method was only attempted once, with a limited number of mice born. The gRNA used in this method was again suboptimal, which may have contributed to the lack of success. However, the data is limited given that there were only 7 mice born and analysed from a single experiment.

A

| attempt	conditions	pups born	correct integration	mutations at CRISPR target site	guide
Large targeting vector	large targeting vector (200ng/ul), gRNA (50ng/ul), Cas9 mRNA (100ng/ul), cytoplasmic injection	20	0	15 (75%)	*Sim2* intron 1 (pX330-mSim2)
HD-ADI	sgRNA (50ng/ul), template (100ng/ul), cas9 mRNA (100ng/uL)	28	0	4 (14%)	*Sim2* 5'UTR (pX330-Sim2-Tom)
ssDNA oligo	ssDNA oligo (10ng/ul), gRNA (20ng/ul), Cas9 protein (20ng/ul), pronuclear injuction	7	0	2 (29%)	*Sim2* ATG (pX330-Sim2-Tomato-Ex1)

B   **HD-ADI**
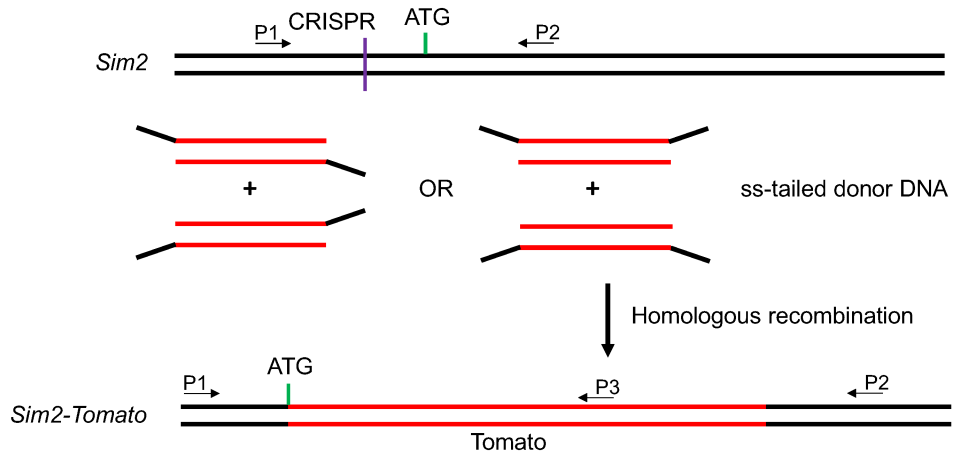
C   **ssDNA oligo**
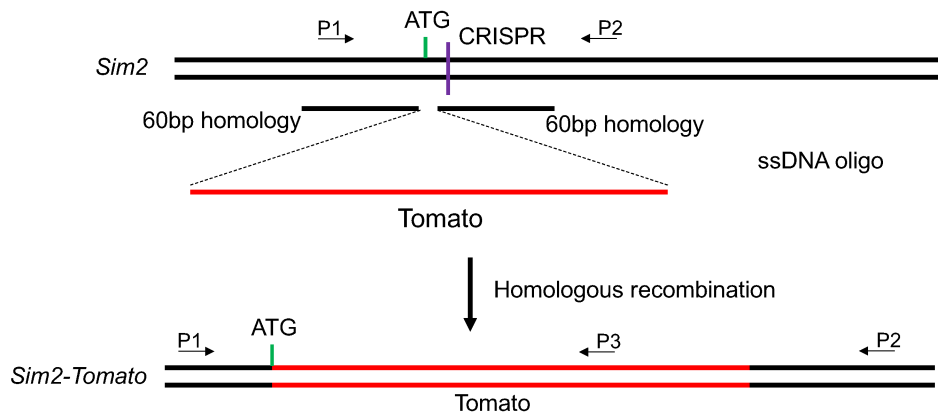
D

E

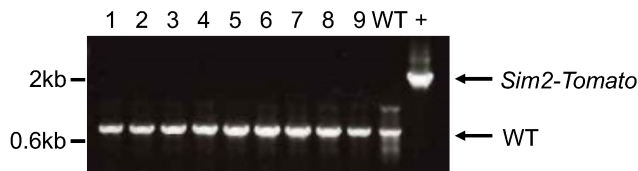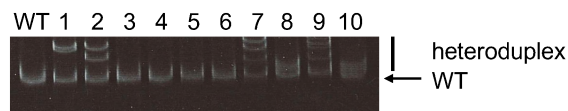**Figure 6.13:** Overview of attempts at generating the *Sim2*-Tomato mouse line. **A)** Summary of attempts. **B)** Schematic of HD-ADI strategy. **C)** Schematic of ssDNA oligo strategy. **D)** Example of genotyping PCR performed to determine whether the *Sim2*-Tomato allele was present. **E)** Example of heteroduplex assay performed to determine whether the gRNA was successfully generating double stranded breaks and mutations at the CRISPR cut site. P1, P2, P3 show the approximate locations of genotyping primers used.

Despite multiple attempts and various strategies employed, generation of the *Sim2*-Tomato reporter mouse line was unsuccessful. A total of 55 mice were born across the three alternative strategies. Genotyping confirmed lack of correct targeting in all 55 mice, with varying degrees of efficiencies for the gRNAs used. Without this mouse line, in addition to the inability to detect SIM2-3xFLAG protein expression in the *Sim2*-3xFLAG mouse line, in depth characterisation of *Sim2* expression throughout development and through to adulthood was not able to be performed. Therefore further *Sim2* mouse model based experiments were not pursued.

6.5 Discussion

6.5.1 *Sim2* conditional knockout mouse model

Given the current literature on *Sim2* it was hypothesised that *Sim2* has neuronal functions that would manifest as a phenotype in a knockout context. As the global *Sim2* knockout mouse line is perinatal lethal (K.-J. Chen et al., 2014; Goshu et al., 2002; Shamblott et al., 2002), we generated a conditional *Sim2* knockout mouse line using the Nestin-cre1 line to remove *Sim2* expression from the brain. The purpose of these initial studies was to determine any substantial phenotypes revealed in later development through to adulthood that could then be interrogated further by more discrete targeted KO mouse lines and experiments. Therefore the Nestin-Cre line provided a useful model as it removed *Sim2* expression in a number of tissues, while allowing the mice to survive to adulthood.

It has been shown that Nestin-cre mice have mild hypopituitarism, which results in a reduction in growth hormone levels and body weight, as well as reduced contextual- and cued-conditioned fear (Declercq et al., 2015; Galichet, Lovell-Badge, & Rizzoti, 2010; Giusti et al., 2014). Importantly, the study of Nestin-Cre1 mice by Giusti et al (2014) did not find any differences in locomotor activity, exploratory behaviour, learning and memory or sociability, a number of behaviours assessed in the battery of behavioural tests the *Sim2*-Nestin-Cre1 mice were subjected to.

Declercq et al (2015) found that Nestin-cre mice weighed on average approximately 5g less that mice without the cre transgene (Declercq et al., 2015). Consistent with this, the genotypes of the *Sim2* mice with the Nestin-cre transgene on average weighed between 2.5-4.5g less than the genotypes not containing the cre (*Sim2*$^{fl/+;Nes-cre1}$ and *Sim2*$^{fl/-;Nes-cre1}$ compared to the *Sim2*$^{fl/-}$ and *Sim2*$^{fl/+}$ mice) (Figure 6.5). Despite this, the results from this study have shown that *Sim2* is not involved in the feeding response, at least to the same extent or in the same pathway as *Sim1*. *Sim1* heterozygous mice have been reported to have weight differences up to and in excess of 10g compared to WT animals in similar high fat diet feeding studies (J. L. Holder et al., 2004). If *Sim2* was playing a similar role to *Sim1*, it would be likely that the increase in weight gained on a high fat diet would still be observed despite the lowered body weight of the Nestin-cre mice. In addition, these results are consistent with our later observations that human patients that have *SIM2* variants do not have obesity phenotypes (see Chapter 3) (Emily L. Button et al., 2022) and there have not been any reported *SIM2* variants in patients with hyperphagic obesity to date. Therefore this study suggests that *Sim2* is not involved in satiety signalling so this appears to be a unique role of *Sim1*.

Given the required strategy to generate neuronal *Sim2* KO mice included cre expression in the germline, *Sim2*$^{+/+; Nes-cre1}$ mice, the perfect littermate controls that might reveal subtle phenotypes, were not produced. Future experiments may benefit from the use more specific neuronal cre crosses, that allow for *Sim2* to remain expressed in germline cells. There are other neuronal-specific cre mouse lines available that may provide suitable neuronal specific KO of *Sim2* expression. The CaMKIIα cre mouse line in which the mouse calcium/calmodulin-dependent protein kinase II alpha (Camk2a) promoter drives expression of the Cre recombinase has been

shown to be specific to neurons (Rios et al., 2001; Tolson et al., 2010; Tsien et al., 1996). Characterisation of a *Sim2*-CaMIIα-cre mouse line may assist in elucidating any subtle neuronal phenotypes as a result of a neuronal *Sim2* KO. In addition, the Sim1-cre mouse line (Balthasar et al., 2005) would provide a useful tool for investigating the function of *Sim2* compared to *Sim1*. By generating a *Sim2-Sim1*-cre mouse line we would be able to observe the unique functions that *Sim2* performs in neurons that co-express these two related genes. Gene expression studies in this context would help to tease out the distinct and overlapping functions of *Sim2* and *Sim1*.

6.5.2   SIM2-3xFLAG mouse model

Even though successful generation of the *Sim2*-3xFLAG mouse model was shown at the DNA level, detection of endogenously tagged SIM2 protein was not achieved. Since homozygous FLAG mice survive and do not have any phenotypes seen in the global *Sim2* KO mouse lines, it can be assumed that total SIM2 protein is being expressed at a sufficient level for function. However it is known that there are two isoforms of *Sim2* (SIM2-s and SIM2-l), with little known regarding the functional differences and redundancy between the two. It could be that the short isoform is the predominantly expressed protein isoform in the tissues and at the time points where protein expression was investigated. This could be determined through generation of a mouse line where the short isoform is tagged.

The C-terminal end of the gene was chosen to insert the tag sequence due to the fact that the basic DNA binding region is at the very N-terminal end of the protein, so it was predicted that adding the acidic FLAG tag sequence may disrupt DNA binding. Consequently the position of the tag sequence insertion would only result in the long isoform being tagged, with the short isoform remaining unchanged. In order to tag both isoforms, it would be necessary to tag either the N-terminal end or tag the C-terminus of both isoforms independently. An advantage of tagging the C-terminus of each isoform independently would be that each isoform could have a unique tag, affording the opportunity to investigate their functions independently. However, the possibility of this causing disrupted splicing of the long isoform would need to be considered due

to the fact that this would require insertion of exogenous sequence into intron 10 of the *Sim2* gene.

Tagging the N-terminus of the protein would result in both expressed isoforms being tagged. It would be possible to determine which isoform is present by western blot through the expected size difference of the two isoforms. The two isoforms however would be indistinguishable through immunofluorescence methods. It would be essential to test whether tagging the N-terminus would disrupt the DNA binding function of SIM2. This could be done by expressing N-terminally tagged SIM2-s protein in culture and assessing transcriptional activity through a reporter gene assay, as the short isoform is identical at the N-terminal end and has been shown to activate reporter gene expression *in vitro* (Moffett & Pelletier, 2000; A. E. Sullivan et al., 2016; Adrienne E. Sullivan et al., 2014).

The most likely explanation for the inability to detect SIM2-3xFLAG protein is that endogenous SIM2 is a labile and perhaps signal regulated protein, and that the conditions (sample processing, signal presence, tissues and developmental time points) chosen to look for protein expression did not overlap with conditions where SIM2 protein is expressed to a detectable level.

A recent study looking at the role of SIM2 in breast cancer has shown that SIM2 protein is stabilised through phosphorylation in response to DNA damage induced by radiation treatment (Scott J. Pearson et al., 2019). Given that other members of the bHLH/PAS transcription factor family are also signal regulated (HIF, AhR), this is not unexpected and provides a rationale to suggest this being the reason SIM2-3xFLAG protein was not detected. In order to determine if this is the case, primary cell cultures from the tagged mice could be grown and treated with radiation to induce SIM2 expression. Western blot and immunofluorescence experiments could then be used to detect any SIM2-3xFLAG protein expression.

6.5.3 SIM2-Tomato reporter mouse

The aim of generating a *Sim2*-Tomato reporter mouse model was not achieved in the work described in this thesis. Generation of the *Sim2*-Tomato allele in mES cells was successful but was not taken further in attempts to generate the mouse line. However, given that other methods used to attempt to generate this mouse line were not successful, this would be a possible alternative. An advantage to this method would be that it would be possible to sequence the ES cells to find a monoclonal line with one successfully targeted allele and one unmodified allele. This would ensure that any mice born do not contain cells that are effectively *Sim2* KO to ensure survival of these mice.

While it is unclear why the attempts at targeting *Sim2* in mouse zygotes was unsuccessful, future attempts may benefit from some modifications to the strategies. A notable difference between the large targeting vector strategy employed in this study and previously reported successful experiments was the size of the donor plasmids used, with the *Sim2*-Tomato targeting vector being approximately twice the size of the plasmids used to generate the published reporter mice (Lengner et al., 2007; Hui Yang et al., 2013). The size and sequence of the *Sim2*-Tomato targeting vector was limited by the ability to modify the starting plasmid. An alternative approach would be to have plasmid synthesised, rather than cloned in house, or clone the targeting vector from scratch, without the use of the existing plasmid as a starting point. A plasmid could be designed and synthesised or cloned which only includes the necessary sequences, reducing the overall size of the targeting plasmid significantly.

The long ssDNA oligo method has been reported to have success rates for generating KI alleles between 8.5 to 100% at various loci (Quadros et al., 2017). Notably, Quadros et al (2017) successfully generated knock-in mouse lines with the use of pre-assembled Cas9 ribonucleoprotein, consisting of the crRNA, tracrRNA and Cas9 ribonucleoprotein (ctRNP). Based on reported literature, this appears to be the most efficient method for generating knock-in alleles, therefore future attempts at generating the *Sim2*-Tomato mouse line should be based on this method and use the ctRNP in order to increase efficiencies. In addition, the inefficiency of the guide RNA may be an advantage in this

case as it is more likely that there will be an unmodified allele, ensuring survival of any mice with successful insertion of the Tomato reporter sequence.

6.5.4 General discussion and future directions

There are still many areas to be explored regarding the developmental function of *Sim2*. Once these mouse models are generated this will open up multiple options for exploring endogenous *Sim2* functions. Both the *Sim2*-Tomato and FLAG mouse lines could be used to generate comprehensive *Sim2* gene and protein expression data. This would allow further studies to be focused on tissues and cell types at developmental time points of interest. A key use of the *Sim2*-Tomato mouse line would be to isolate *Sim2* expressing cells through fluorescence-activated cell sorting (FACS). RNA-sequencing studies can be performed on cells collected from WT (*Sim2*$^{Tomato/+}$) and KO (*Sim2*$^{Tomato/Tomato}$) mice (Brunskill et al., 2014). Identification of key and novel areas of SIM2 expression and the genes regulated will provide valuable insight into its function during development through to adulthood and help to resolve the unexplained phenotypes of the *Sim2* KO mouse models.

These mouse lines could also be used to investigate the functions of *Sim2* compared to *Sim1*. Crossing the *Sim2*-Tomato mouse line with the SIM1-GFP mouse line would allow for direct comparison of *Sim2* vs *Sim1* expression profiles *in vivo* to identify unique and overlapping expression regions (Gong et al., 2003). Identifying regions that only express one of these transcription factors may allow for investigation into uniquely regulated genes and give a greater understanding of the functions specific to each transcription factor.

In addition to the suggested neuronal cre mouse lines, the *Sim2*-floxed mouse line could also be crossed with additional cre mouse lines to investigate the function of *Sim2* in other tissues. For example, *Sim2* is expressed in the developing and adult kidneys however there has not yet been a function described for *Sim2* in this tissue (Ema, Morita, et al., 1996; Fan et al., 1996; Metz et al., 2006). Crossing with one of the many kidney-cre mouse lines that have been successfully used to target proteins in the kidney could be used to investigate the function of *Sim2* in this tissue (Kohan, 2008).

There has already been an established function determined for *Sim2* in the gut, with *Sim2* responsible for the regulation of antimicrobial peptides (K.-J. Chen et al., 2014). With the recent discovery that SIM2 is stabilised through phosphorylation in breast cancer cells, in addition to the fact that other Class I bHLH/PAS transcription factors are signal regulated (N. Hao & Whitelaw, 2013; Scott J. Pearson et al., 2019; Gregg L. Semenza, 2012; G. L. Semenza, 2014), it may be possible that SIM2 is signal regulated in the gut, potentially by a ligand produced by microbes in the gut. Crossing the *Sim2*-floxed mouse line with an intestinal specific cre could be used to investigate this possibility through isolation and sequencing of microbes present in the gut of *Sim2* KO mice.

Chapter 7: Final Discussion

The overall aim of this thesis was to investigate the molecular function and mechanisms of action of SIM2 in both a normal developmental and disease context, and therefore contribute to the understanding of the biological functions of this transcription factor. While this study has certainly contributed significant additional findings regarding the function and mechanisms of SIM2, there are still many avenues for future investigations that can stem from the findings and tools developed as part of this thesis.

During development *Sim2* is expressed in regions of the developing brain including the hypothalamus, cortex, olfactory bulb and the mammillary body, and outside of the central nervous system in the cartilage and bone of the ribs, vertebrae, craniofacial structures, limbs and digits, skeletal muscle and kidney tubules (Coumailleau & Duprez, 2009; Ema, Morita, et al., 1996; Fan et al., 1996; Marion et al., 2005). Knockout Sim*2* mouse models show abnormal phenotypes in these tissues, including loss of hormone expressing neurons in the hypothalamus and abnormal skeletal structures (Goshu et al., 2002; Goshu et al., 2004; Shamblott et al., 2002). Therefore it is likely that loss of function *SIM2* variants could contribute to human developmental pathologies. While it remains to be determined whether heterozygous *SIM2* variants can cause human neurological phenotypes, a recent study has proposed clinical features in a child including craniofacial abnormalities, developmental delay, and intellectual disability to be caused by a homozygous *SIM2* variant (Al-Kurbi et al., 2022). As exome and whole genome sequencing are becoming more widely used and accessible diagnostic tools for determining the cause of genetic conditions it will remain to be seen whether additional *SIM2* variants will be identified as candidate disease causing variants. The set of experiments we presented in Chapter 4 will be useful for predicting and classifying the impact of additional identified SIM2 variants (Emily L. Button et al., 2022).

Aberrant expression of *SIM2* is linked with the progression of a number of human cancers, where upregulation of *SIM2* in prostate, pancreatic and colon cancers favours tumour progression, and downregulation in breast and esophageal cancers favours

tumour progression (Aleman et al., 2005; Arredouani et al., 2009; DeYoung et al., 2003a, 2003b; Gustafson et al., 2009; Ole Johan Halvorsen et al., 2007; Kwak et al., 2007; Laffin et al., 2008; S. J. Pearson et al., 2019; Scott J. Pearson et al., 2019; Scribner et al., 2013; Tamaoki et al., 2018). It is still largely unknown how SIM2 can have opposing actions in different cancer types. The SIM2s target genes in breast cancer we have identified are relatively exclusive to this cancer type as comparisons with other gene expression datasets in cells with manipulated *SIM2* expression found minimal overlap in identified differentially expressed genes (Chapter 5) (Goode et al., 2014; A. E. Sullivan, 2016). SIM2 has been proposed as a potential biomarker and cancer therapeutic where expression is aberrantly upregulated (Arredouani et al., 2009; DeYoung et al., 2003a, 2003b; Ole Johan Halvorsen et al., 2007). It will be important to determine the mechanisms behind the cell type specific gene regulation to ensure off-target effects of modulation of SIM2 function in other tissue types can be predicted and assessed for any detrimental effects.

Crosstalk between bHLH/PAS transcription factors is an area where further research is required to completely understand this occurrence, particularly as it is apparent that this is a biologically relevant function of these proteins in development, normal cellular homeostasis, and disease (reviewed in Chapter 2) (Emily L. Button et al., 2017). With overlap in dimerisation partners and DNA response elements, there is high potential for crosstalk to occur in cells and tissues where bHLH/PAS transcription factors are co-expressed. The finding that SIM2 can crosstalk with HIF1α, AHR and NPAS2 in breast cancer is intriguing and raises the question of whether crosstalk with other bHLH/PAS transcription factors is an important function of SIM2 during development. Whether SIM2 crosstalk with NPAS2 is a biologically relevant interaction or an artefact of mis-regulation of genes in cancer cells remains to be determined. Further investigations are required to advance our understanding of the overlapping and competing functions of this transcription factor family.

Target genes of SIM2 during development are largely still unknown and remain to be identified. However the epitope tagged mouse line we have generated may provide a model to enrich for endogenous SIM2 target genes once further specificity in the spatiotemporal expression of *Sim2* is characterised. Cell and tissue specific knockout

models targeted to key areas of *Sim2* expression throughout development should also reveal any yet to be discovered roles of SIM2. In breast cancer it has been shown that SIM2 is stabilised through phosphorylation by ATM following ionising radiation induced DNA damage (Scott J. Pearson et al., 2019). Whether SIM2 is similarly signal regulated during development remains to be determined, however these mouse models may aid in the identification of any additional SIM2 ligands or regulatory mechanisms.

References

Adamovich, Y., Ladeuix, B., Golik, M., Koeners, M. P., & Asher, G. (2017). Rhythmic Oxygen Levels Reset Circadian Clocks through HIF1alpha. *Cell Metab, 25*(1), 93-101. doi: 10.1016/j.cmet.2016.09.014

Adli, M. (2018). The CRISPR tool kit for genome editing and beyond. *Nature Communications, 9*(1), 1911. doi: 10.1038/s41467-018-04252-2

Aitola, M. H., & Pelto-Huikko, M. T. (2003). Expression of Arnt and Arnt2 mRNA in developing murine tissues. *Journal of Histochemistry & Cytochemistry, 51*(1), 41-54.

Al-Kurbi, A. A., Da'as, S. I., Aamer, W., Krishnamoorthy, N., Poggiolini, I., Abdelrahman, D., . . . Fakhro, K. A. (2022). A recessive variant in SIM2 in a child with complex craniofacial anomalies and global developmental delay. *Eur J Med Genet, 65*(4), 104455. doi: 10.1016/j.ejmg.2022.104455

Aleman, M. J., DeYoung, M. P., Tress, M., Keating, P., Perry, G. W., & Narayanan, R. (2005). Inhibition of Single Minded 2 gene expression mediates tumor-selective apoptosis and differentiation in human colon cancer cells. *Proceedings of the National Academy of Sciences of the United States of America, 102*(36), 12765-12770. doi: 10.1073/pnas.0505484102

Allen Developing Mouse Brain Atlas. Retrieved from http://developingmouse.brain-map.org/

Apte, R. S., Chen, D. S., & Ferrara, N. (2019). VEGF in Signaling and Disease: Beyond Discovery and Development. *Cell, 176*(6), 1248-1264. doi: 10.1016/j.cell.2019.01.021

Arlt, A., & Schäfer, H. (2011). Role of the immediate early response 3 (IER3) gene in cellular stress response, inflammation and tumorigenesis. *Eur J Cell Biol, 90*(6-7), 545-552. doi: 10.1016/j.ejcb.2010.10.002

Arredouani, M. S., Lu, B., Bhasin, M., Eljanne, M., Yue, W., Mosquera, J.-M., . . . Sanda, M. G. (2009). Identification of the Transcription Factor Single-Minded Homologue 2 as a Potential Biomarker and Immunotherapy Target in Prostate Cancer. *Clinical Cancer Research, 15*(18), 5794-5802. doi: 10.1158/1078-0432.ccr-09-0911

Balthasar, N., Dalgaard, L. T., Lee, C. E., Yu, J., Funahashi, H., Williams, T., . . . Lowell, B. B. (2005). Divergence of melanocortin pathways in the control of food intake and energy expenditure. *Cell, 123*(3), 493-505. doi: 10.1016/j.cell.2005.08.035

Bassett, Andrew R., Tibbit, C., Ponting, Chris P., & Liu, J.-L. (2013). Highly Efficient Targeted Mutagenesis of Drosophila with the CRISPR/Cas9 System. *Cell Rep, 4*(1), 220-228. doi: 10.1016/j.celrep.2013.06.020

Baune, B. T., Wiede, F., Braun, A., Golledge, J., Arolt, V., & Koerner, H. (2008). Cognitive dysfunction in mice deficient for TNF- and its receptors. *Am J Med Genet B Neuropsychiatr Genet, 147b*(7), 1056-1064. doi: 10.1002/ajmg.b.30712

Beischlag, T. V., Wang, S., Rose, D. W., Torchia, J., Reisz-Porszasz, S., Muhammad, K., . . . Hankinson, O. (2002). Recruitment of the NCoA/SRC-1/p160 family of transcriptional coactivators by the aryl hydrocarbon receptor/aryl hydrocarbon receptor nuclear translocator complex. *Molecular and Cellular Biology, 22*(12), 4319-4333. doi: 10.1128/mcb.22.12.4319-4333.2002

Bell, C. C., Magor, G. W., Gillinder, K. R., & Perkins, A. C. (2014). A high-throughput screening strategy for detecting CRISPR-Cas9 induced mutations using next-

generation sequencing. *BMC Genomics, 15*, 1002. doi: 10.1186/1471-2164-15-1002

Bersten, D. C., Sullivan, A. E., Li, D., Bhakti, V., Bent, S. J., & Whitelaw, M. L. (2015). Inducible and reversible lentiviral and Recombination Mediated Cassette Exchange (RMCE) systems for controlling gene expression. *Plos One, 10*(3), e0116373. doi: 10.1371/journal.pone.0116373

Bersten, D. C., Sullivan, A. E., Peet, D. J., & Whitelaw, M. L. (2013). bHLH-PAS proteins in cancer. *Nat Rev Cancer, 13*(12), 827-841. doi: 10.1038/nrc3621

Bonnefond, A., Raimondo, A., Stutzmann, F., Ghoussaini, M., Ramachandrappa, S., Bersten, D. C., . . . Froguel, P. (2013). Loss-of-function mutations in SIM1 contribute to obesity and Prader-Willi-like features. *Journal of Clinical Investigation, 123*(7), 3037-3041. doi: 10.1172/jci68035

Bos, R., van der Groep, P., Greijer, A. E., Shvarts, A., Meijer, S., Pinedo, H. M., . . . van der Wall, E. (2003). Levels of hypoxia-inducible factor-1alpha independently predict prognosis in patients with lymph node negative breast carcinoma. *Cancer, 97*(6), 1573-1581. doi: 10.1002/cncr.11246

Brunskill, E. W., Park, J. S., Chung, E., Chen, F., Magella, B., & Potter, S. S. (2014). Single cell dissection of early kidney development: multilineage priming. [Article]. *Development, 141*(15), 3093-3101. doi: 10.1242/dev.110601

Buchholz, F., Refaeli, Y., Trumpp, A., & Bishop, J. M. (2000). Inducible chromosomal translocation of AML1 and ETO genes through Cre/loxP-mediated recombination in the mouse. *Embo Reports, 1*(2), 133-139. doi: 10.1093/embo-reports/kvd027

Button, E. L. (2013). *Disrupting function of the bHLH/PAS transcription factor Single Minded 2.* (Honours Thesis), University of Adelaide.

Button, E. L., Bersten, D. C., & Whitelaw, M. L. (2017). HIF has Biff – Crosstalk between HIF1a and the family of bHLH/PAS proteins. *Experimental Cell Research, 356*(2), 141-145. doi: https://doi.org/10.1016/j.yexcr.2017.03.055

Button, E. L., Rossi, J. J., McDougal, D. P., Bruning, J. B., Peet, D. J., Bersten, D. C., . . . Whitelaw, M. L. (2022). Characterization of functionally deficient SIM2 variants found in patients with neurological phenotypes. *Biochemical Journal, 479*(13), 1441-1454. doi: 10.1042/bcj20220209

Carmeliet, P. (2005). VEGF as a key mediator of angiogenesis in cancer. *Oncology, 69 Suppl 3*, 4-10. doi: 10.1159/000088478

Chen, H., Chrast, R., Rossier, C., Gos, A., Antonarakis, S. E., Kudoh, J., . . . Shimizu, N. (1995). SINGLE-MINDED AND DOWN-SYNDROME. *Nature Genetics, 10*(1), 9-10.

Chen, K.-J., Lizaso, A., & Lee, Y.-H. (2014). SIM2 maintains innate host defense of the small intestine. *American Journal of Physiology-Gastrointestinal and Liver Physiology, 307*(11), G1044-G1056. doi: 10.1152/ajpgi.00241.2014

Cho, S. W., Kim, S., Kim, J. M., & Kim, J.-S. (2013). Targeted genome engineering in human cells with the Cas9 RNA-guided endonuclease. *Nature Biotechnology, 31*(3), 230-232.

Cho, S. W., Kim, S., Kim, Y., Kweon, J., Kim, H. S., Bae, S., & Kim, J.-S. (2014). Analysis of off-target effects of CRISPR/Cas-derived RNA-guided endonucleases and nickases. *Genome Research, 24*(1), 132-141. doi: 10.1101/gr.162339.113

Chrast, R., Scott, H. S., Chen, H. M., Kudoh, J., Rossier, C., Minoshima, S., . . . Antonarakis, S. E. (1997). Cloning of two human homologs of the Drosophila single-minded gene SIM1 on chromosome 6q and SIM2 on 21q within the down syndrome chromosomal region. *Genome Research, 7*(6), 615-624.

Chrast, R., Scott, H. S., Madani, R., Huber, L., Wolfer, D. P., Prinz, M., . . . Antonarakis, S. E. (2000). Mice trisomic for a bacterial artificial chromosome with the single-minded 2 gene (Sim2) show phenotypes similar to some of those present in the partial trisomy 16 mouse models of Down syndrome. *Human Molecular Genetics, 9*(12), 1853-1864. doi: 10.1093/hmg/9.12.1853

Claesson-Welsh, L., & Welsh, M. (2013). VEGFA and tumour angiogenesis. *J Intern Med, 273*(2), 114-127. doi: 10.1111/joim.12019

Conesa, A., Madrigal, P., Tarazona, S., Gomez-Cabrero, D., Cervera, A., McPherson, A., . . . Mortazavi, A. (2016). A survey of best practices for RNA-seq data analysis. *Genome Biology, 17*(1), 13. doi: 10.1186/s13059-016-0881-8

Cong, L., Ran, F. A., Cox, D., Lin, S., Barretto, R., Habib, N., . . . Zhang, F. (2013). Multiplex Genome Engineering Using CRISPR/Cas Systems. *Science, 339*(6121), 819-823.

Coumailleau, P., & Duprez, D. (2009). Sim1 and Sim2 expression during chick and mouse limb development. *International Journal of Developmental Biology, 53*(1), 149-157.

Crews, S. T. (1998). Control of cell lineage-specific development and transcription by bHLH-PAS proteins. *Genes & Development, 12*(5), 607-620. doi: 10.1101/gad.12.5.607

Crews, S. T., & Fan, C. M. (1999). Remembrance of things PAS: regulation of development by bHLH-PAS proteins. *Current Opinion in Genetics & Development, 9*(5), 580-587.

Dahmane, N., Charron, G., Lopes, C., Yaspo, M. L., Maunoury, C., Decorte, L., . . . Delabar, J. M. (1995). DOWN-SYNDROME CRITICAL REGION CONTAINS A GENE HOMOLOGOUS TO DROSOPHILA SIM EXPRESSED DURING RAT AND HUMAN CENTRAL-NERVOUS-SYSTEM DEVELOPMENT. *Proceedings of the National Academy of Sciences of the United States of America, 92*(20), 9191-9195. doi: 10.1073/pnas.92.20.9191

Declercq, J., Brouwers, B., Pruniau, V., Stijnen, P., Faudeur, G., Tuand, K., . . . Creemers, J. (2015). Metabolic and Behavioural Phenotypes in Nestin-Cre Mice Are Caused by Hypothalamic Expression of Human Growth Hormone. *Plos One, 10*, e0135502. doi: 10.1371/journal.pone.0135502

Denison, M. S., & Nagy, S. R. (2003). Activation of the aryl hydrocarbon receptor by structurally diverse exogenous and endogenous chemicals. *Annu Rev Pharmacol Toxicol, 43*, 309-334. doi: 10.1146/annurev.pharmtox.43.100901.135828

DeYoung, M. P., Tress, M., & Narayanan, R. (2003a). Down's syndrome-associated Single Minded 2 gene as a pancreatic cancer drug therapy target. *Cancer Letters, 200*(1), 25-31. doi: 10.1016/s0304-3835(03)00409-9

DeYoung, M. P., Tress, M., & Narayanan, R. (2003b). Identification of Down's syndrome critical locus gene SIM2-s as a drug therapy target for solid tumors. *Proceedings of the National Academy of Sciences of the United States of America, 100*(8), 4760-4765. doi: 10.1073/pnas.0831000100

Doudna, J. A., & Charpentier, E. (2014). The new frontier of genome engineering with CRISPR-Cas9. *Science, 346*(6213). doi: 10.1126/science.1258096

Dubois, N. C., Hofmann, D., Kaloulis, K., Bishop, J. M., & Trumpp, A. (2006). Nestin-Cre transgenic mouse line Nes-Cre1 mediates highly efficient Cre/loxP mediated recombination in the nervous system, kidney, and somite-derived tissues. *Genesis, 44*(8), 355-360.

Eckle, T., Hartmann, K., Bonney, S., Reithel, S., Mittelbronn, M., Walker, L. A., . . . Eltzschig, H. K. (2012). Adora2b-elicited Per2 stabilization promotes a HIF-dependent

metabolic switch critical for myocardial adaptation to ischemia. *Nat Med, 18*(5), 774-782.

Ema, M., Ikegami, S., Hosoya, T., Mimura, J., Ohtani, H., Nakao, K., . . . Fujii-Kuriyama, Y. (1999). Mild impairment of learning and memory in mice overexpressing the mSim2 gene located on chromosome 16: an animal model of Down's syndrome. *Human Molecular Genetics, 8*(8), 1409-1415. doi: 10.1093/hmg/8.8.1409

Ema, M., Morita, M., Ikawa, S., Tanaka, M., Matsuda, Y., Gotoh, O., . . . Fujii-Kuriyama, Y. (1996). Two new members of the murine Sim gene family are transcriptional repressors and show different expression patterns during mouse embryogenesis. *Mol Cell Biol, 16*(10), 5865-5875.

Ema, M., Suzuki, M., Morita, M., Hirose, K., Sogawa, K., Matsuda, Y., . . . FujiiKuriyama, Y. (1996). cDNA cloning of a murine homologue of Drosophila single-minded, its mRNA expression in mouse development, and chromosome localization. *Biochemical and Biophysical Research Communications, 218*(2), 588-594.

Fan, C. M., Kuwana, E., Bulfone, A., Fletcher, C. F., Copeland, N. G., Jenkins, N. A., . . . TessierLavigne, M. (1996). 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, 7*(1), 1-16. doi: 10.1006/mcne.1996.0001

Farrall, A. L., & Whitelaw, M. L. (2009). The HIF1 alpha-inducible pro-cell death gene BNIP3 is a novel target of SIM2s repression through cross-talk on the hypoxia response element. *Oncogene, 28*(41), 3671-3680.

Fu, Y., Foden, J. A., Khayter, C., Maeder, M. L., Reyon, D., Joung, J. K., & Sander, J. D. (2013). High-frequency off-target mutagenesis induced by CRISPR-Cas nucleases in human cells. *Nature Biotechnology, 31*(9), 822-+. doi: 10.1038/nbt.2623

Furness, S. G. B., Lees, M. J., & Whitelaw, M. L. (2007). The dioxin (aryl hydrocarbon) receptor as a model for adaptive responses of bHLH/PAS transcription factors. *Febs Letters, 581*(19), 3616-3625. doi: 10.1016/j.febslet.2007.04.011

Galichet, C., Lovell-Badge, R., & Rizzoti, K. (2010). Nestin-Cre mice are affected by hypopituitarism, which is not due to significant activity of the transgene in the pituitary gland. *Plos One, 5*(7), e11443. doi: 10.1371/journal.pone.0011443

Giusti, S. A., Vercelli, C. A., Vogl, A. M., Kolarz, A. W., Pino, N. S., Deussing, J. M., & Refojo, D. (2014). Behavioral phenotyping of Nestin-Cre mice: Implications for genetic mouse models of psychiatric disorders. *Journal of Psychiatric Research, 55*, 87-95. doi: https://doi.org/10.1016/j.jpsychires.2014.04.002

Gong, S. C., Zheng, C., Doughty, M. L., Losos, K., Didkovsky, N., Schambra, U. B., . . . Heintz, N. (2003). A gene expression atlas of the central nervous system based on bacterial artificial chromosomes. *Nature, 425*(6961), 917-925. doi: 10.1038/nature02033

Goode, G. D., Ballard, B. R., Manning, H. C., Freeman, M. L., Kang, Y., & Eltom, S. E. (2013). Knockdown of aberrantly upregulated aryl hydrocarbon receptor reduces tumor growth and metastasis of MDA-MB-231 human breast cancer cell line. *Int J Cancer, 133*(12), 2769-2780. doi: 10.1002/ijc.28297

Goode, G. D., Pratap, S., & Eltom, S. E. (2014). Depletion of the aryl hydrocarbon receptor in MDA-MB-231 human breast cancer cells altered the expression of genes in key regulatory pathways of cancer. *Plos One, 9*(6), e100103. doi: 10.1371/journal.pone.0100103

Goshu, E., Jin, H., Fasnacht, R., Sepenski, M., Michaud, J. L., & Fan, C. M. (2002). Sim2 mutants have developmental defects not overlapping with those of Sim1 mutants. *Molecular and Cellular Biology, 22*(12), 4147-4157.

Goshu, E., Jin, H., Lovejoy, J., Marion, J. F., Michaud, J. L., & Fan, C. M. (2004). Sim2 contributes to neuroendocrine hormone gene expression in the anterior hypothalamus. *Molecular Endocrinology, 18*(5), 1251-1262.

Guo, Y. R., Partch, C. L., Key, J., Card, P. B., Pashkov, V., Patel, A., . . . Gardner, K. H. (2013). Regulating the ARNT/TACC3 Axis: Multiple Approaches to Manipulating Protein/Protein Interactions with Small Molecules. *Acs Chemical Biology, 8*(3), 626-635. doi: 10.1021/cb300604u

Gustafson, T. L., Wellberg, E., Laffin, B., Schilling, L., Metz, R. P., Zahnow, C. A., & Porter, W. W. (2009). Ha-Ras transformation of MCF10A cells leads to repression of Singleminded-2s through NOTCH and C/EBP beta. *Oncogene, 28*(12), 1561-1568.

Hall, B., Limaye, A., & Kulkarni, A. B. (2009). Overview: generation of gene knockout mice. *Current protocols in cell biology, Chapter 19*, Unit-19.12.17. doi: 10.1002/0471143030.cb1912s44

Halvorsen, O. J., Rostad, K., Oyan, A. M., Bo, T. H., Stordrange, L., Olsen, S., . . . Akslen, L. A. (2005). Single-minded 2 homolog gene (SIM2) is overexpressed in prostate cancer. *Virchows Archiv, 447*(2), 452-452.

Halvorsen, O. J., Rostad, K., Oyan, A. M., Puntervoll, H., Bo, T. H., Stordrange, L., . . . Akslen, L. A. (2007). Increased expression of SIM2-s protein is a novel marker of aggressive prostate cancer. *Clinical Cancer Research, 13*(3), 892-897. doi: 10.1158/1078-0432.ccr-06-1207

Hao, N., Bhakti, V. L. D., Peet, D. J., & Whitelaw, M. L. (2013). Reciprocal regulation of the basic helix-loop-helix/Per-Arnt-Sim partner proteins, Arnt and Arnt2, during neuronal differentiation. *Nucleic Acids Research, 41*(11), 5626-5638. doi: 10.1093/nar/gkt206

Hao, N., Lee, K. L., Furness, S. G., Bosdotter, C., Poellinger, L., & Whitelaw, M. L. (2012). Xenobiotics and loss of cell adhesion drive distinct transcriptional outcomes by aryl hydrocarbon receptor signaling. *Mol Pharmacol, 82*(6), 1082-1093. doi: 10.1124/mol.112.078873

Hao, N., & Whitelaw, M. L. (2013). The emerging roles of AhR in physiology and immunity. *Biochem Pharmacol, 86*(5), 561-570. doi: 10.1016/j.bcp.2013.07.004

Havis, E., Coumailleau, P., Bonnet, A., Bismuth, K., Bonnin, M.-A., Johnson, R., . . . Duprez, D. (2012). Sim2 prevents entry into the myogenic program by repressing MyoD transcription during limb embryonic myogenesis. *Development, 139*(11), 1910-1920.

Holder, J. L., Jr., Butte, N. F., & Zinn, A. R. (2000). Profound obesity associated with a balanced translocation that disrupts the SIM1 gene. *Human Molecular Genetics, 9*(1), 101-108.

Holder, J. L., Zhang, L., Kublaoui, B. M., DiLeone, R. J., Oz, O. K., Bair, C. H., . . . Zinn, A. R. (2004). Sim1 gene dosage modulates the homeostatic feeding response to increased dietary fat in mice. *American Journal of Physiology-Endocrinology and Metabolism, 287*(1), E105-E113. doi: 10.1152/ajpendo.00446.2003

Hosoya, T., Oda, Y., Takahashi, S., Morita, M., Kawauchi, S., Ema, M., . . . Fujii-Kuriyama, Y. (2001). Defective development of secretory neurones in the hypothalamus of Arnt2-knockout mice. *Genes to Cells, 6*(4), 361-374. doi: 10.1046/j.1365-2443.2001.00421.x

Hsu, Patrick D., Lander, Eric S., & Zhang, F. (2014). Development and Applications of CRISPR-Cas9 for Genome Engineering. *Cell, 157*(6), 1262-1278. doi: 10.1016/j.cell.2014.05.010

Hsu, P. D., Scott, D. A., Weinstein, J. A., Ran, F. A., Konermann, S., Agarwala, V., . . . Zhang, F. (2013). DNA targeting specificity of RNA-guided Cas9 nucleases. *Nature Biotechnology, 31*(9), 827-+. doi: 10.1038/nbt.2647

Huang, N., Chelliah, Y., Shan, Y., Taylor, C. A., Yoo, S.-H., Partch, C., . . . Takahashi, J. S. (2012). Crystal Structure of the Heterodimeric CLOCK:BMAL1 Transcriptional Activator Complex. *Science, 337*(6091), 189-194. doi: 10.1126/science.1222804

Ivan, M., Kondo, K., Yang, H., Kim, W., Valiando, J., Ohh, M., . . . Kaelin, W. G., Jr. (2001). HIFalpha targeted for VHL-mediated destruction by proline hydroxylation: implications for O2 sensing. *Science, 292*(5516), 464-468. doi: 10.1126/science.1059817

Jaakkola, P., Mole, D. R., Tian, Y. M., Wilson, M. I., Gielbert, J., Gaskell, S. J., . . . Ratcliffe, P. J. (2001). Targeting of HIF-alpha to the von Hippel-Lindau ubiquitylation complex by O2-regulated prolyl hydroxylation. *Science, 292*(5516), 468-472. doi: 10.1126/science.1059796

Jaehne, E. J., Klarić, T. S., Koblar, S. A., Baune, B. T., & Lewis, M. D. (2015). Effects of Npas4 deficiency on anxiety, depression-like, cognition and sociability behaviour. *Behav Brain Res, 281*, 276-282. doi: 10.1016/j.bbr.2014.12.044

Jain, S., Maltepe, E., Lu, M. M., Simon, C., & Bradfield, C. A. (1998). Expression of ARNT, ARNT2, HIF1 alpha, HIF2 alpha and Ah receptor mRNAs in the developing mouse. *Mechanisms of Development, 73*(1), 117-123. doi: 10.1016/s0925-4773(98)00038-0

Keith, B., Adelman, D. M., & Simon, M. C. (2001). 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, 98*(12), 6692-6697. doi: 10.1073/pnas.121494298

Kewley, R. J., Whitelaw, M. L., & Chapman-Smith, A. (2004). The mammalian basic helix-loop-helix/PAS family of transcriptional regulators. *International Journal of Biochemistry & Cell Biology, 36*(2), 189-204.

Kissick, H. T., Sanda, M. G., Dunn, L. K., & Arredouani, M. S. (2014). Immunization with a peptide containing MHC class I and II epitopes derived from the tumor antigen SIM2 induces an effective CD4 and CD8 T-cell response. *Plos One, 9*(4), e93231. doi: 10.1371/journal.pone.0093231

Kohan, D. E. (2008). Progress in gene targeting: using mutant mice to study renal function and disease. *Kidney International, 74*(4), 427-437. doi: 10.1038/ki.2008.146

Kolluri, S. K., Jin, U. H., & Safe, S. (2017). Role of the aryl hydrocarbon receptor in carcinogenesis and potential as an anti-cancer drug target. *Arch Toxicol, 91*(7), 2497-2513. doi: 10.1007/s00204-017-1981-2

Kotch, L. E., Iyer, N. V., Laughner, E., & Semenza, G. L. (1999). Defective vascularization of HIF-1 alpha-null embryos is not associated with VEGF deficiency but with mesenchymal cell death. *Developmental Biology, 209*(2), 254-267. doi: 10.1006/dbio.1999.9253

Kozak, K. R., Abbott, B., & Hankinson, O. (1997). ARNT-deficient mice and placental differentiation. *Developmental Biology, 191*(2), 297-305. doi: 10.1006/dbio.1997.8758

Kublaoui, B. M., Gemelli, T., Tolson, K. P., Wang, Y., & Zinn, A. R. (2008). Oxytocin deficiency mediates hyperphagic obesity of Sim1 haploinsufficient mice. *Molecular Endocrinology, 22*(7), 1723-1734. doi: 10.1210/me.2008-0067

Kublaoui, B. M., Holder, J. L., Jr., Gemelli, T., & Zinn, A. R. (2006). Sim1 haploinsufficiency impairs melanocortin-mediated anorexia and activation of paraventricular nucleus neurons. *Molecular Endocrinology, 20*(10), 2483-2492. doi: 10.1210/me.2005-0483

Kublaoui, B. M., Holder, J. L., Jr., Tolson, K. P., Gemelli, T., & Zinn, A. R. (2006). SIM1 overexpression partially rescues agouti yellow and diet-induced obesity by normalizing food intake. *Endocrinology, 147*(10), 4542-4549. doi: 10.1210/en.2006-0453

Kwak, H.-I., Gustafson, T., Metz, R. P., Laffin, B., Schedin, P., & Porter, W. W. (2007). Inhibition of breast cancer growth and invasion by single-minded 2s. *Carcinogenesis, 28*(2), 259-266. doi: 10.1093/carcin/bgl122

Laffin, B., Wellberg, E., Kwak, H.-I., Burghardt, R. C., Metz, R. P., Gustafson, T., . . . Porter, W. W. (2008). Loss of singleminded-2s in the mouse mammary gland induces an epithelial-mesenchymal transition associated with up-regulation of slug and matrix metalloprotease 2. *Molecular and Cellular Biology, 28*(6), 1936-1946. doi: 10.1128/mcb.01701-07

Lando, D., Peet, D. J., Gorman, J. J., Whelan, D. A., Whitelaw, M. L., & Bruick, R. K. (2002). FIH-1 is an asparaginyl hydroxylase enzyme that regulates the transcriptional activity of hypoxia-inducible factor. *Genes & Development, 16*(12), 1466-1471. doi: 10.1101/gad.991402

Lando, D., Peet, D. J., Whelan, D. A., Gorman, J. J., & Whitelaw, M. L. (2002). Asparagine hydroxylation of the HIF transactivation domain: A hypoxic switch. *Science, 295*(5556), 858-861. doi: 10.1126/science.1068592

Lengner, C. J., Camargo, F. D., Hochedlinger, K., Welstead, G. G., Zaidi, S., Gokhale, S., . . . Jaenisch, R. (2007). Oct4 expression is not required for mouse somatic stem cell self-renewal. *Cell Stem Cell, 1*(4), 403-415. doi: 10.1016/j.stem.2007.07.020

Li, H., Beckman, K. A., Pessino, V., Huang, B., Weissman, J. S., & Leonetti, M. D. (2017). Design and specificity of long ssDNA donors for CRISPR-based knock-in. *bioRxiv*. doi: 10.1101/178905

Loew, R., Heinz, N., Hampf, M., Bujard, H., & Gossen, M. (2010). Improved Tet-responsive promoters with minimized background expression. *BMC Biotechnology, 10*(1), 81. doi: 10.1186/1472-6750-10-81

Lu, B., Asara, J. M., Sanda, M. G., & Arredouani, M. S. (2011). The Role of the Transcription Factor SIM2 in Prostate Cancer. *Plos One, 6*(12). doi: 10.1371/journal.pone.0028837

Malone, E. R., Oliva, M., Sabatini, P. J. B., Stockley, T. L., & Siu, L. L. (2020). Molecular profiling for precision cancer therapies. *Genome Med, 12*(1), 8. doi: 10.1186/s13073-019-0703-1

Maltepe, E., Schmidt, J. V., Baunoch, D., Bradfield, C. A., & Simon, M. C. (1997). Abnormal angiogenesis and responses to glucose and oxygen deprivation in mice lacking the protein ARNT. *Nature, 386*(6623), 403-407. doi: 10.1038/386403a0

Marion, J. F., Yang, C., Caqueret, A., Boucher, F., & Michaud, J. L. (2005). Sim1 and Sim2 are required for the correct targeting of mammillary body axons. *Development, 132*(24), 5527-5537.

Menezes, M. E., Bhatia, S., Bhoopathi, P., Das, S. K., Emdad, L., Dasgupta, S., . . . Fisher, P. B. (2014). MDA-7/IL-24: multifunctional cancer killing cytokine. *Adv Exp Med Biol, 818*, 127-153. doi: 10.1007/978-1-4471-6458-6_6

Metz, R. P., Kwak, H. I., Gustafson, T., Laffin, B., & Porter, W. W. (2006). Differential transcriptional regulation by mouse single-minded 2s. *Journal of Biological Chemistry, 281*(16), 10839-10848. doi: 10.1074/jbc.M508858200

Michaud, J., & Fan, C. M. (1997). Single-minded - Two genes, three chromosomes. *Genome Research, 7*(6), 569-571.

Michaud, J. L., Boucher, F., Melnyk, A., Gauthier, F., Goshu, E., Levy, E., . . . Fan, C. M. (2001). Sim1 haploinsufficiency causes hyperphagia, obesity and reduction of the paraventricular nucleus of the hypothalamus. *Human Molecular Genetics, 10*(14), 1465-1473. doi: 10.1093/hmg/10.14.1465

Michaud, J. L., Rosenquist, T., May, N. R., & Fan, C. M. (1998). Development of neuroendocrine lineages requires the bHLH-PAS transcription factor SIM1. *Genes & Development, 12*(20), 3264-3275. doi: 10.1101/gad.12.20.3264

Moennikes, O., Loeppen, S., Buchmann, A., Andersson, P., Ittrich, C., Poellinger, L., & Schwarz, M. (2004). A constitutively active dioxin/aryl hydrocarbon receptor promotes hepatocarcinogenesis in mice. *Cancer Res, 64*(14), 4707-4710. doi: 10.1158/0008-5472.Can-03-0875

Moffett, P., Dayo, M., Reece, M., McCormick, M. K., & Pelletier, J. (1996). Characterization of msim, a murine homologue of the Drosophila sim transcription factor. *Genomics, 35*(1), 144-155. doi: 10.1006/geno.1996.0333

Moffett, P., & Pelletier, J. (2000). Different transcriptional properties of mSim-1 and mSim-2. *Febs Letters, 466*(1), 80-86.

Moffett, P., Reece, M., & Pelletier, J. (1997). The murine Sim-2 gene product inhibits transcription by active repression and functional interference. *Molecular and Cellular Biology, 17*(9), 4933-4947.

Murray, I. A., Patterson, A. D., & Perdew, G. H. (2014). Aryl hydrocarbon receptor ligands in cancer: friend and foe. *Nature Reviews Cancer, 14*(12), 801-814. doi: 10.1038/nrc3846

Nakamura, K., Komatsu, M., Chiwaki, F., Takeda, T., Kobayashi, Y., Banno, K., . . . Sasaki, H. (2017). SIM2l attenuates resistance to hypoxia and tumor growth by transcriptional suppression of HIF1A in uterine cervical squamous cell carcinoma. *Sci Rep, 7*(1), 14574. doi: 10.1038/s41598-017-15261-4

Nambu, J. R., Franks, R. G., Hu, S., & Crews, S. T. (1990). THE SINGLE-MINDED GENE OF DROSOPHILA IS REQUIRED FOR THE EXPRESSION OF GENES IMPORTANT FOR THE DEVELOPMENT OF CNS MIDLINE CELLS. *Cell, 63*(1), 63-75.

Nambu, J. R., Lewis, J. O., Wharton, K. A., & Crews, S. T. (1991). THE DROSOPHILA SINGLE-MINDED GENE ENCODES A HELIX-LOOP-HELIX PROTEIN THAT ACTS AS A MASTER REGULATOR OF CNS MIDLINE DEVELOPMENT. *Cell, 67*(6), 1157-1167.

Niu, Y., Shen, B., Cui, Y., Chen, Y., Wang, J., Wang, L., . . . Sha, J. (2014). Generation of Gene-Modified Cynomolgus Monkey via Cas9/RNA-Mediated Gene Targeting in One-Cell Embryos. *Cell, 156*(4), 836-843. doi: 10.1016/j.cell.2014.01.027

Oh, E. T., & Park, H. J. (2015). Implications of NQO1 in cancer therapy. *BMB Rep, 48*(11), 609-617. doi: 10.5483/bmbrep.2015.48.11.190

Okey, A. B., Franc, M. A., Moffat, I. D., Tijet, N., Boutros, P. C., Korkalainen, M., . . . Pohjanvirta, R. (2005). Toxicological implications of polymorphisms in

receptors for xenobiotic chemicals: the case of the aryl hydrocarbon receptor. *Toxicol Appl Pharmacol, 207*(2 Suppl), 43-51. doi: 10.1016/j.taap.2004.12.028

Okui, M., Yamaki, A., Takayanagi, A., Kudoh, J., Shimizu, N., & Shimizu, Y. (2005). Transcription factor single-minded 2 (SIM2) is ubiquitinated by the RING-IBR-RING-type E3 ubiquitin ligases. *Experimental Cell Research, 309*(1), 220-228. doi: 10.1016/j.yexcr.2005.05.018

Oliveira, K. C. S., Ramos, I. B., Silva, J. M. C., Barra, W. F., Riggins, G. J., Palande, V., . . . Calcagno, D. Q. (2020). Current Perspectives on Circulating Tumor DNA, Precision Medicine, and Personalized Clinical Management of Cancer. *Mol Cancer Res, 18*(4), 517-528. doi: 10.1158/1541-7786.Mcr-19-0768

Partch, C. L., & Gardner, K. H. (2011). Coactivators necessary for transcriptional output of the hypoxia inducible factor, HIF, are directly recruited by ARNT PAS-B. *Proceedings of the National Academy of Sciences of the United States of America, 108*(19), 7739-7744. doi: 10.1073/pnas.1101357108

Pattanayak, V., Lin, S., Guilinger, J. P., Ma, E., Doudna, J. A., & Liu, D. R. (2013). High-throughput profiling of off-target DNA cleavage reveals RNA-programmed Cas9 nuclease specificity. *Nature Biotechnology, 31*(9), 839-+. doi: 10.1038/nbt.2673

Pearson, S. J., Elswood, J., Barhoumi, R., Ming-Whitfield, B., Rijnkels, M., & Porter, W. W. (2019). Loss of SIM2s inhibits RAD51 binding and leads to unresolved replication stress. *Breast Cancer Res, 21*(1), 125. doi: 10.1186/s13058-019-1207-z

Pearson, S. J., Roy Sarkar, T., McQueen, C. M., Elswood, J., Schmitt, E. E., Wall, S. W., . . . Porter, W. W. (2019). ATM-dependent activation of SIM2s regulates homologous recombination and epithelial–mesenchymal transition. *Oncogene, 38*(14), 2611-2626. doi: 10.1038/s41388-018-0622-4

Peek, C. B., Levine, D. C., Cedernaes, J., Taguchi, A., Kobayashi, Y., Tsai, S. J., . . . Bass, J. (2017). Circadian Clock Interaction with HIF1alpha Mediates Oxygenic Metabolism and Anaerobic Glycolysis in Skeletal Muscle. *Cell Metabolism, 25*(1), 86-92. doi: 10.1016/j.cmet.2016.09.010

Peng, L. U., Bai, G., & Pang, Y. (2021). Roles of NPAS2 in circadian rhythm and disease. *Acta Biochim Biophys Sin (Shanghai), 53*(10), 1257-1265. doi: 10.1093/abbs/gmab105

Pfaffl, M. W. (2001). A new mathematical model for relative quantification in real-time RT-PCR. *Nucleic Acids Res, 29*(9), e45. doi: 10.1093/nar/29.9.e45

Pinto, J. A., Rolfo, C., Raez, L. E., Prado, A., Araujo, J. M., Bravo, L., . . . Gomez, H. L. (2017). In silico evaluation of DNA Damage Inducible Transcript 4 gene (DDIT4) as prognostic biomarker in several malignancies. *Sci Rep, 7*(1), 1526. doi: 10.1038/s41598-017-01207-3

Powell, J. B., Goode, G. D., & Eltom, S. E. (2013). The Aryl Hydrocarbon Receptor: A Target for Breast Cancer Therapy. *Journal of cancer therapy, 4*(7), 1177-1186. doi: 10.4236/jct.2013.47137

Preethi, S., Arthiga, K., Patil, A. B., Spandana, A., & Jain, V. (2022). Review on NAD(P)H dehydrogenase quinone 1 (NQO1) pathway. *Mol Biol Rep*. doi: 10.1007/s11033-022-07369-2

Probst, M. R., Fan, C. M., TessierLavigne, M., & Hankinson, O. (1997). Two murine homologs of the Drosophila single-minded protein that interact with the mouse aryl hydrocarbon receptor nuclear translocator protein. *Journal of Biological Chemistry, 272*(7), 4451-4457.

Quadros, R. M., Miura, H., Harms, D. W., Akatsuka, H., Sato, T., Aida, T., . . . Gurumurthy, C. B. (2017). Easi-CRISPR: a robust method for one-step generation of mice carrying conditional and insertion alleles using long ssDNA donors and CRISPR ribonucleoproteins. *Genome Biology, 18*(1), 92. doi: 10.1186/s13059-017-1220-4

Quadros, R. M., Ohtsuka, M., Harms, D. W., Aida, T., Redder, R., Miura, H., . . . Gurumurthy, C. B. (2016). Easi-CRISPR: Efficient germline modification with long ssDNA donors. *bioRxiv*, 069963. doi: 10.1101/069963

Rachidi, M., & Lopes, C. (2007). Mental retardation in Down syndrome: From gene dosage imbalance to molecular and cellular mechanisms. *Neuroscience Research, 59*(4), 349-369. doi: 10.1016/j.neures.2007.08.007

Ramachandrappa, S., Raimondo, A., Cali, A. M. G., Keogh, J. M., Henning, E., Saeed, S., . . . Farooqi, I. S. (2013). Rare variants in single-minded 1 (SIM1) are associated with severe obesity. *Journal of Clinical Investigation, 123*(7), 3042-3050. doi: 10.1172/jci68016

Rios, M., Fan, G. P., Fekete, C., Kelly, J., Bates, B., Kuehn, R., . . . Jaenisch, R. (2001). Conditional deletion of brain-derived neurotrophic factor in the postnatal brain leads to obesity and hyperactivity. *Molecular Endocrinology, 15*(10), 1748-1757. doi: 10.1210/me.15.10.1748

Rodriguez, M., & Potter, D. A. (2013). CYP1A1 regulates breast cancer proliferation and survival. *Mol Cancer Res, 11*(7), 780-792. doi: 10.1158/1541-7786.Mcr-12-0675

Safe, S., Cheng, Y., & Jin, U.-H. (2017). The aryl hydrocarbon receptor (AhR) as a drug target for cancer chemotherapy. *Current Opinion in Toxicology, 2*, 24-29. doi: https://doi.org/10.1016/j.cotox.2017.01.012

Sancar, A., & Van Gelder, R. N. (2021). Clocks, cancer, and chronochemotherapy. *Science, 371*(6524). doi: 10.1126/science.abb0738

Schindl, M., Schoppmann, S. F., Samonigg, H., Hausmaninger, H., Kwasny, W., Gnant, M., . . . Oberhuber, G. (2002). Overexpression of hypoxia-inducible factor 1alpha is associated with an unfavorable prognosis in lymph node-positive breast cancer. *Clin Cancer Res, 8*(6), 1831-1837.

Schlezinger, J. J., Liu, D., Farago, M., Seldin, D. C., Belguise, K., Sonenshein, G. E., & Sherr, D. H. (2006). A role for the aryl hydrocarbon receptor in mammary gland tumorigenesis. *Biological Chemistry, 387*(9), 1175-1187. doi: 10.1515/bc.2006.145

Scribner, K. C., Behbod, F., & Porter, W. W. (2013). Regulation of DCIS to invasive breast cancer progression by Singleminded-2s (SIM2s). *Oncogene, 32*(21), 2631-2639. doi: 10.1038/onc.2012.286

Semenza, G. L. (1999). Regulation of mammalian O-2 homeostasis by hypoxia-inducible factor 1. *Annual Review of Cell and Developmental Biology, 15*, 551-578. doi: 10.1146/annurev.cellbio.15.1.551

Semenza, G. L. (2003). Targeting HIF-1 for cancer therapy. *Nature Reviews Cancer, 3*(10), 721-732. doi: 10.1038/nrc1187

Semenza, G. L. (2012). Hypoxia-Inducible Factors in Physiology and Medicine. *Cell, 148*(3), 399-408. doi: 10.1016/j.cell.2012.01.021

Semenza, G. L. (2014). Oxygen Sensing, Hypoxia-Inducible Factors, and Disease Pathophysiology. In A. K. Abbas, S. J. Galli & P. M. Howley (Eds.), *Annual Review of Pathology: Mechanisms of Disease, Vol 9* (Vol. 9, pp. 47-71).

Shamblott, M. J., Bugg, E. M., Lawler, A. M., & Gearhart, J. D. (2002). Craniofacial abnormalities resulting from targeted disruption of the murine Sim2 gene. *Developmental Dynamics, 224*(4), 373-380. doi: 10.1002/dvdy.10116

Sneha, S., Baker, S. C., Green, A., Storr, S., Aiyappa, R., Martin, S., & Pors, K. (2021). Intratumoural Cytochrome P450 Expression in Breast Cancer: Impact on Standard of Care Treatment and New Efforts to Develop Tumour-Selective Therapies. *Biomedicines, 9*(3). doi: 10.3390/biomedicines9030290

Sofer, A., Lei, K., Johannessen, C. M., & Ellisen, L. W. (2005). Regulation of mTOR and cell growth in response to energy stress by REDD1. *Mol Cell Biol, 25*(14), 5834-5845. doi: 10.1128/mcb.25.14.5834-5845.2005

Sojka, K. M., Kern, C. B., & Pollenz, R. S. (2000). Expression and subcellular localization of the aryl hydrocarbon receptor nuclear translocator (ARNT) protein in mouse and chicken over developmental time. *Anatomical Record, 260*(4), 327-334. doi: 10.1002/1097-0185(200012)260:4<326::aid-ar10>3.0.co;2-u

Souza, M. F., Kuasne, H., Barros-Filho, M. C., Cilião, H. L., Marchi, F. A., Fuganti, P. E., . . . Cólus, I. M. S. (2017). Circulating mRNAs and miRNAs as candidate markers for the diagnosis and prognosis of prostate cancer. *Plos One, 12*(9), e0184094. doi: 10.1371/journal.pone.0184094

Stark, R., Grzelak, M., & Hadfield, J. (2019). RNA sequencing: the teenage years. *Nature Reviews Genetics, 20*(11), 631-656. doi: 10.1038/s41576-019-0150-2

Su, P., Wen, S. W., Zhang, Y. F., Li, Y., Xu, Y. Z., Zhu, Y. G., . . . Tian, Z. Q. (2016). Identification of the Key Genes and Pathways in Esophageal Carcinoma. *Gastroenterology Research and Practice*. doi: 10.1155/2016/2968106

Sullivan, A. E. (2016). *Function and Regulation of Single-minded 1 and Single-minded 2 Transcription Factors in Disease.* University of Adelaide.

Sullivan, A. E., Peet, D. J., & Whitelaw, M. L. (2016). MAGED1 is a novel regulator of a select subset of bHLH PAS transcription factors. *FEBS J, 283*(18), 3488-3502. doi: 10.1111/febs.13824

Sullivan, A. E., Raimondo, A., Schwab, T. A., Bruning, J. B., Froguel, P., Farooqi, I. S., . . . Whitelaw, M. L. (2014). Characterization of human variants in obesity-related SIM1 protein identifies a hot-spot for dimerization with the partner protein ARNT2. *Biochemical Journal, 461*, 403-412. doi: 10.1042/bj20131618

Tamaoki, M., Komatsuzaki, R., Komatsu, M., Minashi, K., Aoyagi, K., Nishimura, T., . . . Sasaki, H. (2018). Multiple roles of single-minded 2 in esophageal squamous cell carcinoma and its clinical implications. *Cancer science, 109*(4), 1121-1134. doi: 10.1111/cas.13531

Thompson, C. L., Ng, L., Menon, V., Martinez, S., Lee, C. K., Glattfelder, K., . . . Jones, A. R. (2014). A high-resolution spatiotemporal atlas of gene expression of the developing mouse brain. *Neuron, 83*(2), 309-323. doi: 10.1016/j.neuron.2014.05.033

Tian, H., Hammer, R. E., Matsumoto, A. M., Russell, D. W., & McKnight, S. L. (1998). The hypoxia-responsive transcription factor EPAS1 is essential for catecholamine homeostasis and protection against heart failure during embryonic development. *Genes & Development, 12*(21), 3320-3324. doi: 10.1101/gad.12.21.3320

To, K. K. W., Sedelnikova, O. A., Samons, M., Bonner, W. M., & Huang, L. E. (2006). The phosphorylation status of PAS-B distinguishes HIF-1alpha from HIF-2alpha in NBS1 repression. *EMBO J, 25*(20), 4784-4794. doi: 10.1038/sj.emboj.7601369

Tolson, K. P., Gemelli, T., Gautron, L., Elmquist, J. K., Zinn, A. R., & Kublaoui, B. M. (2010). Postnatal Sim1 Deficiency Causes Hyperphagic Obesity and Reduced Mc4r and Oxytocin Expression. *Journal of Neuroscience, 30*(10), 3803-3812. doi: 10.1523/jneurosci.5444-09.2010

Trumpp, A., Depew, M. J., Rubenstein, J. L. R., Bishop, J. M., & Martin, G. R. (1999). Cre-mediated gene inactivation demonstrates that FGF8 is required for cell survival and patterning of the first branchial arch. *Genes & Development, 13*(23), 3136-3148. doi: 10.1101/gad.13.23.3136

Tsien, J. Z., Chen, D. F., Gerber, D., Tom, C., Mercer, E. H., Anderson, D. J., . . . Tonegawa, S. (1996). Subregion- and cell type-restricted gene knockout in mouse brain. *Cell, 87*(7), 1317-1326. doi: 10.1016/s0092-8674(00)81826-7

Tsuchiya, Y., Nakajima, M., & Yokoi, T. (2005). Cytochrome P450-mediated metabolism of estrogens and its regulation in human. *Cancer Lett, 227*(2), 115-124. doi: 10.1016/j.canlet.2004.10.007

Valente, M. J., Henrique, R., Costa, V. L., Jerónimo, C., Carvalho, F., Bastos, M. L., . . . Carvalho, M. (2011). A Rapid and Simple Procedure for the Establishment of Human Normal and Cancer Renal Primary Cell Cultures from Surgical Specimens. *Plos One, 6*(5), e19337. doi: 10.1371/journal.pone.0019337

Waller, C. L., & McKinney, J. D. (1995). Three-dimensional quantitative structure-activity relationships of dioxins and dioxin-like compounds: model validation and Ah receptor characterization. *Chem Res Toxicol, 8*(6), 847-858. doi: 10.1021/tx00048a005

Wang, F., Zhang, R. X., Wu, X. M., & Hankinson, O. (2010). Roles of Coactivators in Hypoxic Induction of the Erythropoietin Gene. *Plos One, 5*(3). doi: 10.1371/journal.pone.0010002

Wang, G. L., Jiang, B. H., Rue, E. A., & Semenza, G. L. (1995). 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, 92*(12), 5510-5514. doi: 10.1073/pnas.92.12.5510

Wang, H., & Xu, X. (2017). Microhomology-mediated end joining: new players join the team. *Cell & bioscience, 7*, 6-6. doi: 10.1186/s13578-017-0136-8

Wang, H. Y., Yang, H., Shivalila, C. S., Dawlaty, M. M., Cheng, A. W., Zhang, F., & Jaenisch, R. (2013). One-Step Generation of Mice Carrying Mutations in Multiple Genes by CRISPR/Cas-Mediated Genome Engineering. *Cell, 153*(4), 910-918.

Wang, X., Song, Y., Chen, L., Zhuang, G., Zhang, J., Li, M., & Meng, X.-F. (2013). Contribution of single-minded 2 to hyperglycaemia-induced neurotoxicity. *Neurotoxicology, 35*, 106-112. doi: 10.1016/j.neuro.2013.01.003

Wang, Z., Zhang, J., Ye, M., Zhu, M., Zhang, B., Roy, M., . . . An, X. (2014). Tumor suppressor role of protein 4.1B/DAL-1. *Cell Mol Life Sci, 71*(24), 4815-4830. doi: 10.1007/s00018-014-1707-z

Webb, E. A., AlMutair, A., Kelberman, D., Bacchelli, C., Chanudet, E., Lescai, F., . . . Dattani, M. T. (2013). ARNT2 mutation causes hypopituitarism, post-natal microcephaly, visual and renal anomalies. *Brain, 136*, 3096-3105. doi: 10.1093/brain/awt218

Wellberg, E., Metz, R. P., Parker, C., & Porter, W. W. (2010). The bHLH/PAS transcription factor singleminded 2s promotes mammary gland lactogenic differentiation. *Development, 137*(6), 945-952.

Whelan, F., Hao, N., Furness, S. G., Whitelaw, M. L., & Chapman-Smith, A. (2010). Amino acid substitutions in the aryl hydrocarbon receptor ligand binding domain

reveal YH439 as an atypical AhR activator. *Mol Pharmacol, 77*(6), 1037-1046. doi: 10.1124/mol.109.062927

Whitaker, E. L., Filippov, V. A., & Duerksen-Hughes, P. J. (2012). Interleukin 24: mechanisms and therapeutic potential of an anti-cancer gene. *Cytokine Growth Factor Rev, 23*(6), 323-331. doi: 10.1016/j.cytogfr.2012.08.004

Whitelaw, M. L., Gustafsson, J. A., & Poellinger, L. (1994). IDENTIFICATION OF TRANSACTIVATION AND REPRESSION FUNCTIONS OF THE DIOXIN RECEPTOR AND ITS BASIC HELIX-LOOP-HELIX/PAS PARTNER FACTOR ARNT - INDUCIBLE VERSUS CONSTITUTIVE MODES OF REGULATION. *Molecular and Cellular Biology, 14*(12), 8343-8355.

WHO. (2022, 03/02/2022). Cancer. Retrieved 24/04/2022, from https://www.who.int/news-room/fact-sheets/detail/cancer

Woods, S., Farrall, A., Procko, C., & Whitelaw, M. L. (2008). The bHLH/Per-Arnt-Sim transcription factor SIM2 regulates muscle transcript myomesin2 via a novel, non-canonical E-box sequence. *Nucleic Acids Research, 36*(11), 3716-3727. doi: 10.1093/nar/gkn247

Woods, S. L., & Whitelaw, M. L. (2002). Differential activities of murine single minded 1 (SIM1) and SIM2 on a hypoxic response element - Cross-talk between basic helix-loop-helix/Per-Arnt-Sim homology transcription factors. *Journal of Biological Chemistry, 277*(12), 10236-10243. doi: 10.1074/jbc.M110752200

Wu, M. X. (2003). Roles of the stress-induced gene IEX-1 in regulation of cell death and oncogenesis. *Apoptosis, 8*(1), 11-18. doi: 10.1023/a:1021688600370

Wu, M. X., Ustyugova, I. V., Han, L., & Akilov, O. E. (2013). Immediate early response gene X-1, a potential prognostic biomarker in cancers. *Expert Opin Ther Targets, 17*(5), 593-606. doi: 10.1517/14728222.2013.768234

Wu, Y., Liang, D., Wang, Y., Bai, M., Tang, W., Bao, S., . . . Li, J. (2013). Correction of a Genetic Disease in Mouse via Use of CRISPR-Cas9. *Cell Stem Cell, 13*(6), 659-662. doi: 10.1016/j.stem.2013.10.016

Wu, Y., Tang, D., Liu, N., Xiong, W., Huang, H., Li, Y., . . . Zhang, E. E. (2017). Reciprocal Regulation between the Circadian Clock and Hypoxia Signaling at the Genome Level in Mammals. *Cell Metabolism, 25*(1), 73-85. doi: 10.1016/j.cmet.2016.09.009

Wyatt, G. L., Crump, L. S., Young, C. M., Wessells, V. M., McQueen, C. M., Wall, S. W., . . . Lyons, T. R. (2019). Cross-talk between SIM2s and NFkappaB regulates cyclooxygenase 2 expression in breast cancer. *Breast Cancer Res, 21*(1), 131. doi: 10.1186/s13058-019-1224-y

Xue, P., Fu, J., & Zhou, Y. (2018). The Aryl Hydrocarbon Receptor and Tumor Immunity. *Frontiers in Immunology, 9*, 286-286. doi: 10.3389/fimmu.2018.00286

Yamaki, A., Kudoh, J., Shimizu, N., & Shimizu, Y. (2004). A novel nuclear localization signal in the human single-minded proteins SIM1 and SIM2. *Biochemical and Biophysical Research Communications, 313*(3), 482-488. doi: 10.1016/j.bbrc.2003.11.168

Yamaki, A., Noda, S., Kudoh, J., Shindoh, N., Maeda, H., Minoshima, S., . . . Shimizu, N. (1996). The mammalian single-minded (SIM) gene: Mouse cDNA structure and diencephalic expression indicate a candidate gene for Down syndrome. *Genomics, 35*(1), 136-143. doi: 10.1006/geno.1996.0332

Yang, H., Wang, H., Shivalila, C. S., Cheng, A. W., Shi, L., & Jaenisch, R. (2013). One-Step Generation of Mice Carrying Reporter and Conditional Alleles by CRISPR/Cas-Mediated Genome Engineering. *Cell, 154*(6), 1370-1379.

Yang, H., Wang, H. Y., & Jaenisch, R. (2014). Generating genetically modified mice using CRISPR/Cas-mediated genome engineering. [Article]. *Nature Protocols, 9*(8), 1956-1968. doi: 10.1038/nprot.2014.134

Yang, Y., Zhang, Y., Wu, Q., Cui, X., Lin, Z., Liu, S., & Chen, L. (2014). Clinical implications of high NQO1 expression in breast cancers. *J Exp Clin Cancer Res, 33*(1), 14. doi: 10.1186/1756-9966-33-14

Yang, Y., Zhu, G., Dong, B., Piao, J., Chen, L., & Lin, Z. (2019). The NQO1/PKLR axis promotes lymph node metastasis and breast cancer progression by modulating glycolytic reprogramming. *Cancer Letters, 453*, 170-183. doi: 10.1016/j.canlet.2019.03.054

Yu, F., White, S. B., Zhao, Q., & Lee, F. S. (2001). HIF-1alpha binding to VHL is regulated by stimulus-sensitive proline hydroxylation. *Proc Natl Acad Sci U S A, 98*(17), 9630-9635. doi: 10.1073/pnas.181341498

Yu, T., Tang, B., & Sun, X. (2017). Development of Inhibitors Targeting Hypoxia-Inducible Factor 1 and 2 for Cancer Therapy. *Yonsei medical journal, 58*(3), 489-496. doi: 10.3349/ymj.2017.58.3.489

Zhang, X.-H., Tee, L. Y., Wang, X.-G., Huang, Q.-S., & Yang, S.-H. (2015). Off-target Effects in CRISPR/Cas9-mediated Genome Engineering. *Molecular Therapy - Nucleic Acids, 4*, e264. doi: https://doi.org/10.1038/mtna.2015.37

Zhidkova, E. M., Lylova, E. S., Grigoreva, D. D., Kirsanov, K. I., Osipova, A. V., Kulikov, E. P., . . . Lesovaya, E. A. (2022). Nutritional Sensor REDD1 in Cancer and Inflammation: Friend or Foe? *Int J Mol Sci, 23*(17). doi: 10.3390/ijms23179686

Zimmerman, L., Lendahl, U., Cunningham, M., McKay, R., Parr, B., Gavin, B., . . . McMahon, A. (1994). INDEPENDENT REGULATORY ELEMENTS IN THE NESTIN GENE DIRECT TRANSGENE EXPRESSION TO NEURAL STEM-CELLS OR MUSCLE PRECURSORS. *Neuron, 12*(1), 11-24. doi: 10.1016/0896-6273(94)90148-1