Evolution of pluripotency: Hijacking of an ancient network

By Darren Jerome Crowley

BSc, MRes

Thesis submitted to the University of Nottingham for the degree of Doctor of Philosophy, December 2017

School of Life Sciences

University of Nottingham

United Kingdom

DECLARATION

I declare that this thesis is my own work and contains nothing which is the outcome of work done in collaboration with others, except as specified in the acknowledgments, and that has not previously been presented as part of any other degree.

ABSTRACT

Pluripotency is conserved in the major trunk of Vertebrate evolution, but how the gene regulatory network (GRN) that governs it evolved is poorly understood. A central component of this network is the Homeodomain containing transcription factor Nanog.

How Nanog evolved is not understood, as Nanog sequences have not been identified in invertebrate genomes. This study provides evidence of Nanog activity encoded in the homeodomain of the invertebrate Vent gene family. The Vent2 gene from Saccoglossus kowalevskii, a model hemichordate, successfully reprogrammed mammalian pre-iPS cells to pluripotency, as demonstrated by the activation of dormant pluripotency genes, and the ability to generate all three primary germ layers.

A second property of invertebrate Vents was also characterised in the Vent gene found in Nematostella vectensis, a sea anemone and model for cnidarian development, expression of which was insufficient to activate the endogenous pluripotency network of pre-iPS cells, though it could induce the cells to a XEN-like state that demonstrated up regulation of extra-embryonic markers, and subsequently gained dependence on ERK signalling.

A direct comparison between the Saccoglossus and Nematostella Vent homeodomains was used to provide insight into the step-wise changes that appear to have given rise to Nanog activity. Swapping the homeodomains from one CDS to another, to create hybrid molecules, I demonstrated that the respective reprogramming activities of these genes is conserved in the homeodomain. I then identified specific amino acid (AAs) differences in the homeodomains that conferred a Nanog-like capacity for reprogramming to the Nematostella gene.

Identification of a Nematostella EsrrB ortholog, which demonstrated reprogramming activity in mammalian pre-iPS cells, suggests wider conservation of pluripotency factors. I therefore propose that an ancient GRN for pluripotent mesoderm evolved in vertebrates to form part of the ground state network.

ACKNOWLEDGEMENTS

It’s been one heck of ride…

Firstly, to Professor Andrew Johnson, thank you for showing me how to pursue rigorous logic, creativity, the importance of real thinking, and what it really means to be a scientist. AJ leads his lab on the front line, and makes the rest of us feel like the knights at the round table- confididient that we’re all in it together no matter what, and fighting for something worthwhile.

To Ramiro Alberio, you’ve been both a fantastic teacher, and an amzing friend. I still try to adhere to all the good habits you thought me in tissue culture, and beyond, during my MRes degree. Knowing I can turn to you for advice has really helped me through the PhD.

To Jodie Chatfield, my life in Nottingham has benefitted so much from your help, support and friendship, both inside and out of the lab. You’re an incredible person, and thank you for everything. A special thanks to my pal Luke Simpson, who’s ‘Thomas Edison’ like creativity is an inspiration. Massive thanks to Matt Loose for sound advice in and out of work. A huge thank you to Gary Morgan who kindly shared both his expertise on cell biology, and Friday pints too! Thank you to Prof. Yoshi Seki, who thought me a lot, and is a truly lovely guy. A big shout out to Pratyush Deepth Roy, always had my back. Thank you to Tilo Kunath for providing the XEN cells for my comparison with my work, and thank you to Cinzia Allegrucci for always making herself available if I needed some help with work.

To my old housemate Áine Byrne-“Though she be little, she is fierce.” What an incredible friend you are. Your kindness, and insatiable drive and determination lit up the years we lived together. To my buddy Fionn Dunphy-Doherty, couldn’t have done it without you. Our coffee breaks and spontaneous pub trips got me through the PhD. Bags of Cans and all night sessions staying up watching boxing matches was a real highlight! Karolina, Ben and Izzy- thank you for you love and support. To Jen Biggs, you never cease to amaze me, and I couldn’t have done this without your support. For Tony, Cian and Shane, ye’re the best friends a man could have- a terrible influence, and bottomless reservoir of support, loyalty, fun, and terrible ideas! To the rest of the lads- Kieran, Barry, Dave, Emily, D Kelly, Symone, Chris, Áine H- You’re all amazing and I’m lucky to have you as my friends. A massive shout out to the St. Barnabas hurling boys as well- fantastic bunch of lads, and the championship will be ours next year!

To my family, your love and support was the driving force throughout the last few years. My aunts Bridget, Marg, Ann, and T- you guys are amazing, as is the whole extended family, lucky boy that I am!

To my brother Johnny, you’re a constant source of inspiration. You’ve always got my back, and always listen when I need you the most. Looking forward to a good puck around and some fishing next year! To my sister Ellen, I’m so proud of you, and you’ve always encouraged me to reach my best! To Michael, your kindness never ceases to amaze me. You’ve helped me so much, and I can’t wait tell you all about what I’ve learnt. To my dad John, thank you so much for always checking in and encouraging me to fight hard and never give up. The phone calls listening to hurling results on the radio with you always made me feel close to home. To my beautiful girlfriend Teri, I can’t imagine where I’d be without you. Even in the closing stages of my PhD you acted as my thesis-midwife. You’re an amazing friend, and I love you loads. To my mother Catherine, you’re the strongest person I know. I’ve reached my goals because of you, and if I turn out to be half the person you are, I’ll be very proud. Go raibh mile maith agaibh!!

CONTENTS

Declaration 2
Abstract 3
Acknowledgements 4
Contents 6
List of Figures 9
List of Tables 11
List of Abbreviations used 12
Chapter 1 15
1 Introduction 15
1.1 Literature Review 15
1.1.1 Pluripotency 15
1.1.2 The Core Pluripotency Network 17
1.1.3 Nanog as the gateway to ground-state pluripotency 18
1.1.4 Role of Nanog in Regulating Epigenetics 20
1.1.5 Cellular Reprogramming 23
1.1.6 Pluripotency is conserved 28
1.1.7 Reprogramming capacity of Nanog is conserved in the homeodomain 31
1.1.8 The origin of Nanog 33
1.2 Aims and objectives of this study 41
Chapter 2 42
2 General Materials and Methods 42
2.1 Cell lines 42
2.2 Cell Media 43
2.2.1 HEK 293T culture media 43
2.2.2 Basic ESC Serum Lif media 43
2.2.3 Ground state 2i + LIF media 43
2.2.4 Basic XEN culture media 44
2.3 Molecular cloning 44
2.3.1 Construction of plasmids 44
2.3.2 Plasmid mini preparation 45
2.4 Gene delivery systems 45
2.4.1 Lentiviral transduction 46
2.4.2 PiggyBac lines 47
2.5 Antibiotic selection 47

2.6 General Reprogramming methods 48
2.6.1 Neural Stem cell derived Nanog-null pre-iPS cells 48
2.6.2 TNGA MEF pre-iPS cells 48
2.7 Formation and differentiation of Embryoid bodies 48
2.8 AP staining 49
2.9 LacZ staining 49
2.10 Analysis of gene expression 50
2.10.1 RNA extraction 50
2.10.2 cDNA synthesis 50
2.10.3 qPCR analysis 51
Chapter Three 52
3 Results: Section One 52
3.1 Reprogramming Nanog-null NS derived pre-iPS cells 52
3.1.1 Lentiviral cell lines in serum + Lif conditions 52
3.1.2 2i + LIF reprogramming 57
3.1.3 iPS cells identified at day ten 61
3.1.4 Staining of iPS colonies 64
3.1.5 Efficiency of reprogramming 67
3.1.6 Saccoglossus iPS cells in culture 68
3.1.7 Monomer cell lines in serum + LIF 69
3.1.8 Saccoglossus Vent2 can generate iPS colonies in its native form 71
3.1.9 Saccoglossus Vent2 can activate endogenous Nanog as a monomer 74
3.1.10 Saccoglossus Vent2 iPSCs unstable in culture 76
3.1.11 Nematostella Vent drives an unusual phenotype 78
3.2 qPCR analysis of reprogrammed cells 81
3.3 Embryoid body differentiation 83
3.4 qPCR analysis of Embryoid body differentiation 84
3.5 MEF reprogramming 86
3.5.1 Creation of TNGA pre-iPS lines 86
3.5.2 Reprogramming of the TNGA lines 87
3.6 qPCR analysis of MEF derived iPSCs 92
3.7 qPCR analysis of Sea anemone Vent pre-iPSC lines 94
Chapter Four 97
4 Results: Section Two 97
4.1 Swapping the Homeodomains 97
4.2 Modified Sea anemone Vent gene 101

4.2.1 Amino Acid swap....101
4.2.2 Reprogramming with Modified Sea anemone Vent....103
4.3 An EsrrB orthologue in Sea anemone....108
Chapter Five....114
5 Discussion....114
5.1 Nanog activity is conserved....114
5.2 The origin of Nanog activity....119
5.3 Hypothesis....122
6 Appendices....125
6.1 Coding sequences of genes used in this study....125
7 Bibliography....132

LIST OF FIGURES

Figure 1: Phases of the reprogramming progress (Buganim et al., 2013). Pg. 25

Figure 2. : The structure of Nanog (Chambers and Tomlinson, 2009; Dixon et al., 2010) Pg. 30

Figure 3. : Phylogenetic analysis of Nanog paralogs in osteichthyans (Scerbo et al. 2014) Pg. 35

Figure 4: Evolutionary tree of relevant species for this study. Pg. 37

Figure 5: Phylogenetic tree of Nanog, Vent and BSX. Pg. 38

Figure 6: Alkaline Phosphatase staining of mESC lines in cytokine independent self-renewal assay. Pg. 40

Figure 7: NS derived pre-iPS lines in Serum + LIF expressing candidate genes with a WR domain three passages after transduction. Pg. 55

Figure 8: Colony formation at day three of reprogramming. Pg. 58

Figure 9: Partially reprogrammed pre-iPS Vent cultures struggling in 2i + LIF. Pg. 59

Figure 10: Saccoglossus Vent2 colonies emerge at day five as Nematostella Vent collapses. Pg. 61

Figure 11: Day 10 of reprogramming. Pg. 62

Figure 12: Extensive G418 treatment required for Xenopus Vent2-WR cultures. Pg. 64

Figure 13: LacZ staining at day 10 of reprogramming. Pg. 65

Figure 14: Alkaline phosphatase staining for pluripotent cells. Pg. 66

Figure 15: Passaged iPS lines growing in 2i + LIF. Pg. 69

Figure 16: Nematosella Vent induces a round phenotype. Pg. 71

Figure 17: Optimising seeding densities for reprogramming. Pg. 73

Figure 18: Saccoglossus Vent2 in its native form is sufficient to reprogram Nanog null pre-iPSCs. Pg. 74

Figure 19: Instability of the Saccoglossus Vent2 iPSCs. Pg.77

Figure 20: Testing the effects of 2i molecules on NvVent cells. Pg. 79

Figure 21: AP staining of Nematostella Vent line in different inhibitors Pg.80

Figure 22: qPCR analysis for expression of pluripotency markers in reprogrammed NS cells Pg.82

Figure 23: Silencing of exogenous pluripotency factors. Pg. 83

Figure 24: Differentiation of Embryoid bodies. Pg. 84

Figure 25: Differentiation of EBs to all three germ layers. Pg. 86

Figure 26: TNGA MEFs in Serum + LIF. Pg. 88

Figure 27: Nematostella Vent and Xenopus Vent TNGA MEFs. Pg. 89

Figure 28: GFP-positive MEF derived iPSC colonies. Pg. 90

Figure 29: Expansion of TNGA iPSC lines. Pg. 92

Figure 30: Activation of endogenous pluripotency genes in TNGA cells. Pg.93

Figure 31: Xen cell comparison with NvVent transfected pre-iPSCs. Pg. 94

Figure 32: Genes expressed in XEN cells are upregulated in NvVent pre-iPS cell lines. Pg. 95

Figure 33: Invertebrate Vent activity regulated specifically by the homeodomain. Pg. 98

Figure 34: SkVent2HD LacZ and AP staining. Pg. 99

Figure 35: The homeodomain of SKVent2 can activate endogenous pluripotency genes. Pg. 100

Figure 36: Alignment of the Homeodomain Amino Acid sequences of Nanogs and Vents. Pg. 102

Figure 37: Alignment displaying Homeodomain AAs which were swapped in NvVent to create the modified Nematostella Vent. Pg. 102

Figure 38: Nucleotide alterations in the homeodomain of Nematostella Vent to create the synthetic Modified NvVent. Pg. 103

Figure 39: The modified NvVent in culture, and emergence of iPSC colonies. Pg. 104

Figure 40: LacZ and AP staining of the ModNvVent iPSC colonies. Pg. 105

Figure 41: ModNvVent iPSCs maintained unusual flat and jagged phenotype. Pg.106

Figure 42: Activation of pluripotency factors by the Modified sea anemone Vent. Pg. 107

Figure 43: NvEsrrB colonies emerging during reprogramming in 2i + LIF. Pg. 109

Figure 44: NvEsrrB colonies 4 days and 16 days after passage. Pg. 110

Figure 45: NvEsrrB AP and LacZ 16 days after passage. Pg. 111

Figure 46: NvEsrrB is sufficient to activate endogenous pluripotency genes. Pg. 112

Figure 47: Section of Vent and EsrrB morphants compared with uninjected. Pg. 123

Figure 48: Illustration of my findings in a wider biological context. Pg. 124

List of Tables

Table 1: Candadate genes tested. Pg. 53

Table 2: Reprogramming efficiency of positively stained cell lines at day 10. Pg. 67

Table 3: Reprogramming efficiency of monomeric Saccoglossus Vent2. Pg. 75

Table 4: Efficiency of reprogramming TNGA MEFs. Pg. 91

Table 5: reprogramming efficiency of Saccoglossus Vent2 containing Nematostella flanking regions. Pg. 100

List of Abbreviations used

AP ……Alkaline Phosphatase

BME ……β-Mercaptoethanol

cDNA ……coding DNA

CCS ……caspase cleavage site

CDS ……coding sequence

DNA ……Deoxyribonucleic acid

EBs ……Embryoid Bodies

ESC ……Embryonic Stem Cells

EpiSC ……Epiblast-derived Stem cells

FBS …… Fetal bovine serum

GFP ……Green fluorescent protein

GRN ……Gene Regulatory Network

HD ……Homeodomain

H&E …… Haematoxylin and Eosin

H3K4me3 …… Histone 3 Lysine 4 trimethylation

H3K9me3 ……Histone 3 Lysine 9 trimethylation

H3K27ac …… Histone 3 Lysine 27 acetylation

ICM ……Inner Cell Mass

iPSC ……Induced Pluripotent Stem Cells

ITRs …… Inverted terminal repeat sequences

iXEN ……Induced XEN cells

LB …… Lysogeny broth

LIF …… Leukemia inhibitory factor

MEF ……Mouse embryonic fibroblasts

OSKM ……Oct4 Sox2 Klf4 cMyc (Yamanaka factors for reprogramming)

PBS …… Phosphate-buffered saline

PCH …… Pericentromeric heterochromatin

PCR ……Polymerase chain reaction

PGC ……Primordial Germ Cell

PURO ……Puromycin antibiotic

qPCR ……Quantitative (Real time) PCR

RNA ……Ribonucleic acid

RNAi ……RNA interference

TNGA ……GFP targeted to Nanog clone-A cell line

WR ……Tryptophan repeat domain

CHAPTER 1

1 INTRODUCTION

1.1 LITERATURE REVIEW

1.1.1 Pluripotency

Multicellular organisms originate from a fertilized egg with the developmental potential to form an adult consisting of different cell types and tissues. In most vertebrates a cell population is established in the early embryo which has the capacity to produce all three embryonic germ layers, ectoderm, mesoderm and endoderm, in addition to the primordial germ cells (PGCs). This capacity is described as “naïve pluripotency” (Nichols, J. & Smith, A., 2012). Although pluripotency is maintained only transiently in the developing embryo, it can be captured and maintained indefinitely *in vitro* through the derivation of cultured embryonic stem cells (ESCs). Evans and Kaufman (1981) first demonstrated the derivation of mouse ESCs from the inner cell mass (ICM) of mouse blastocysts, the ICM at this point consisting of approximately 20 cells (Tang et al., 2010). Blastocysts at this stage can be maintained in culture for about five days, after which point ICM-derived explants are disrupted into clusters of cells that proliferate and are passaged, establishing lines of immortalized ESCs.

During the transition from ICM to ESC in vitro, the normal developmental program of embryogenesis is replaced, as the progression of cellular development is arrested, and the ESCs can be propagated indefinitely in a self-renewing state in response to exogenously supplied cytokine signalling. These cells can also be induced to exit the self-renewing state and embark down defined paths of differentiation by the application of alternate signalling regimes. (Ying et al., 2008; Wray et al., 2010; Nichols and Smith, 2012; Jang et al., 2017). The propagation of mouse ESCs in an undifferentiated state can be maintained by leukemia inhibitory factor (LIF) which mediates self-renewal through the activation of STAT3 (Niwa et al., 1998), and either serum or bone morphogenetic protein (BMP) to induce inhibitor-of-differentiation proteins (Ying et al., 2003). In initial efforts to derive ESCs, stem cells were grown on a layer of mitotically inactivated fibroblasts known as 'feeders', until it was later discovered that the application of LIF was a sufficient substitute for the multifactor secretions provided by feeder cells (Williams et al., 1988).

Further elucidation of how pluripotency is regulated led to the development of culture conditions that more rigorously support the naïve self-renewing state. Most important of these are the so-called 2i conditions, which combines dual inhibition (2i) of mitogen-activated protein kinase (MAPK) signalling and glycogen synthase kinase-3 (GSK3) with the self-renewal cytokine LIF. Kunath et al. (2007) reported that Fgf4 stimulation of Erk1/2 is an autoinductive stimulus for naïve ES cells to exit the self-renewal programme, suggesting that the Erk cascade directs transition to a state that is primed for germ layer segregation. Removal of this FGF-Erk1/2 input arrests the progression of stem cells to either neural or mesodermal lineage commitment, and there are no significant alterations in the expression levels of key pluripotency markers Oct4, Nanog and Rex1 (Kunath et al., 2007). As neither LIF nor serum/BMP block the activation of ERK, Ying et al. (2008) proposed that the LIF and serum/BMP signals act downstream of phospho-ERK to block ES-cell commitment. They found that inhibition of the FGF4-MAPK cascade by addition of a MEK inhibitor to the culture media, in combination with LIF, replaces the requirement for serum/BMP and supports robust long-term ES-cell propagation. Inhibitors of the FGF4–MAPK signalling cascade reduce differentiation, though they are insufficient to support clonal propagation of ES cells, and apoptosis is relatively high (Ying et al., 2008). Relief of GSK3-mediated negative regulation of biosynthetic pathways was reported to enhance ES-cell propagation (Sato et al., 2004), and Ying et al. (2008) showed that the addition of GSK-3 inhibitor (CHIR99021) restores growth and viability of undifferentiated ES cells in the absence of ERK signalling (Ying et al., 2008).

When 2i conditions are combined with LIF (2i + LIF), the ground state is stabilized in 2i and does not intrinsically fluctuate, as demonstrated by the homogenously high levels of pluripotency factors such as Nanog and Rex1 (Wray et al., 2010). Cells maintained in 2i + LIF conditions are also epigenetically distinct from cells cultured in serum/LIF. They promote differential binding of core pluripotency transcription factors, which correlate with putative enhancer activity and precede changes in bivalent histone marks and DNA hypomethylation (Marks et al., 2012; Galonska et al., 2015; Blinka and Rao, 2017).

Pluripotency is characterised by a pluripotent gene regulatory network (GRN) which orchestrates the developmental potency of both the early embryo, as well as its manifestation in vitro in the form of ESCs, and it can be recapitulated through the reprogramming of somatic cells. In this network, specific transcription factors (TFs) positively regulate loci essential to maintaining pluripotency, repress differentiation programs and form a complex interconnected web of interactions that direct pluripotency, and when force expressed can activate the endogenous pluripotency network in somatic cells (Chen et al., 2008; Jaenisch and Young, 2008; Kim et al., 2008). Constitutive ectopic expression of pluripotent transcription factors in somatic cells can also induce the cells to be reprogrammed back to pluripotency (Takahashi and Yamanaka, 2006).

1.1.2 The Core Pluripotency Network

Transcriptional regulation is achieved through combinatorial interactions between regulatory elements in the genome and a vast range of factors that modulate the recruitment and activity of RNA polymerase. In an early mammalian embryo, pluripotency in the ICM is characterised by the expression of three transcription factors: Oct4, Sox2, and Nanog (Mitsui et al., 2003; Young, 2011; Li and Belmonte, 2017). Genome-wide studies have identified the co-localisation of these three factors in ESC chromatin, and their interactions mark the foundation of the transcriptional network that directs ESC identity (Boyer et al., 2005; Loh et al., 2006). For example, Boyer et al. (2005) documented that Nanog, Oct4, and Sox2 co-occupy approximately 350 target genes in human ESCs. In addition, each member of this triumvirate binds to its own regulatory sequences as well as those of other core members, leading to auto-regulatory and positive forward-feeding circuits. The precise manner in which these factors interact with DNA contributes to their function. Nanog is a homeobox-containing transcription factor which has a variant homeodomain that binds to DNA (Jauch et al., 2008). The level of identity between the Nanog homeodomain and other homeodomains is below that which permits assignment to a particular family (Kappen et al., 1993). The DNA sequence bound by Nanog has been disputed as it has affinity for multiple sequences (Mitsui et al., 2003; Loh et al., 2006; Jauch et al., 2008), but Oct4, which has a low-affinity POU-specific domain (POUS) adjacent to a higher affinity homeodomain (POUHD), and the Sox2 HMG domain, recognise a single consensus site each (Falkner and Zachau, 1984; Bowles et al., 2000).

Oct4 and Sox2 bind DNA co-operatively (Ambrosetti et al., 1997). Sox2-null ES cells have a propensity to differentiate to trophectoderm, which mimics the deletion of *Oct4* (Masui et al., 2007) though the level of Oct–Sox enhancer-dependent genes can be maintained in Sox2-null ES cells through forced expression of Oct4 (Masui et al., 2007). However Sox2 deletion in the zygote results in early embryonic lethality due to failed epiblast formation (Avilion et al., 2003). Ding et al. (2015) used a Sox2 protein interactome to identify Tex10 as a binding target. They reported that Tex10 regulates super enhancer (SE) assembly and activity in a Sox2 dependent manner through the recruitment of the histone acetyltransferase p300 and the DNA demethylase Tet1, regulating the transcription of SE-associated eRNA and mRNA expression. In ESCs, SEs are highly enriched for the core pluripotency factors Oct4, Sox2, and Nanog (Whyte et al., 2013). Sox2-dependent functions of Tex10 play a functionally significant role in ESC self-renewal, early embryo development, and reprogramming (Ding et al., 2015). Activation of the *Sox2* locus is also a critical step during the later stages of somatic cell reprogramming to pluripotency (Buganim et al., 2012).

Among the genes whose expression is driven by the co-operative interaction of Oct4/Sox2 is Nanog. Rodda et al. (2005) identified a Sox-Oct cis-regulatory element, which is necessary for the transcription of Nanog that is conserved within the Nanog proximal promoter. In a developing mouse embryo, it has been established that Nanog-null ICMs are blocked from maturing into pluripotent epiblasts, and either divert to trophectoderm, or they die (Silva et al., 2009). This explains why mutant embryos cannot give rise to ES cells (Mitsui et al., 2003), though an established ESC line remains viable when devoid of Nanog (Chambers et al., 2007).

1.1.3 Nanog as the gateway to ground-state pluripotency

Nanog is a divergent homeodomain-containing protein that is accorded a central role in the gene regulatory network that orchestrates pluripotency (Boyer et al., 2005;

Wang et al., 2006). It was first identified in 2003 by Chambers et al. when Nanog mRNA was isolated as a self-renewal determinant in mouse ESCs, acting in parallel with cytokine stimulation of Stat3. Coincidentally it was also reported by Mitsui et al. (2003) who used in silico differential display to identify several genes specifically expressed in mouse ES cells and preimplantation embryos. It was discovered that endogenous Nanog works in parallel with LIF to drive stem cell self-renewal, and that its forced exogenous expression could bypass the need for cytokine mediated stimulation of Stat3 when expanding ESC clones (Chambers et al., 2003). This property is dependent on Nanog dimerization which is mediated by a Tryptophan repeat domain (WR) in the carboxy terminal of Nanog (Mullin et al., 2008; Wang et al., 2008).

Transient down-regulation of Nanog will predispose ES cells to differentiation, but does not trigger differentiation, as reported by Chambers et al. (2007). They showed that after genetic deletion of Nanog by homologous recombination they could establish an ESC line capable of self-renewing in perpetuity, but with a greater propensity to differentiate than their wild-type counterparts. Although Nanog null ES cells can produce PGCs in in a chimeric mouse, the Nanog null PGCs fail to mature upon reaching the genital ridge and they are unable to commit to development beyond E11.5 (Chambers et al., 2007). Nanog is essential, however, for a mouse embryo's progression beyond implantation (Mitsui et al., 2003; Silva et al., 2009). Silva et al. (2009) contrasted the distribution of Nanog expression with that of Gata4 and Gata6, which mark the development of the hypoblast, an essential extraembryonic tissue from which primitive endoderm is derived. Nanog expression was discovered to be mutually exclusive of the GATA factors, and instead marked the epiblast, exclusively, in the mature blastocyst (Silva et al., 2009).

Reactivation of a silenced X-chromosome is also an epigenetic hallmark of the acquisition of full pluripotency (Mak et al., 2004), and during mouse development the expression of Nanog in the ICM at E4.5 marks the subset of cells where X chromosome silencing is erased (Silva et al., 2009). Concurrent with this, Nanog-null ICMs fail to reactivate the X chromosome, and fail to generate pluripotent cells (Silva et al., 2009).

The general consensus is that NANOG is first expressed in humans at the ICM stage (Cauffman et al., 2009), and at the morula stage in mouse (Mitsui et al., 2003; Hart *et al*., 2004). It has been shown to act as an antagonist of primitive endoderm specification during murine blastocyst formation, as the ICM cells differentiate into either epiblast, which is Nanog dependent, or primitive endoderm (PrE) which relies on Gata6 expression (Mitsui et al., 2003; Silva et al., 2009; Frankenberg et al., 2011). In vitro, ectopic Gata6 expression in ES cells is sufficient to transdifferentiate them into PrE cells (Fujikura et al., 2002); however, Nanog binds to Gata6 regulatory sequences to repress its expression (Singh et al., 2007). Early ICM progenitors co-express Nanog and Gata6 before acquiring distinct identities cell by cell as these factors acquire a mutually exclusive expression pattern (Hermitte and Chazaud, 2014). This is dependent on FGF4 signalling, which induces PrE cells at the expense of the Epi cells (Yamanaka et al., 2010). The discovery that blocking the Erk pathway inhibits PrE specification while promoting Epi identity lead to the development of 2i + LIF culture conditions to promote ground state pluripotency (Nichols et al., 2009).

Nanog is therefore regarded as being predominantly involved in the establishment of the pluripotent inner cell mass and germ cell states in vivo, and it plays less of a role in the general maintenance of pluripotency (Chambers et al., 2007; Silva et al., 2009). It is this indispensable role in the establishment of pluripotency in the embryo that is recapitulated *in vitro* through nuclear reprogramming of somatic cells to a naïve state.

1.1.4 Role of Nanog in Regulating Epigenetics

The genome of eukaryotic cells is organized into euchromatin, which is generally permissive for gene transcription and activation, and heterochromatin, which is highly condensed and largely gene-poor. This compartmentalization plays a crucial role in regulating and stabilizing the genome, and thereby impacting lineage commitment and contributing to cell identity (Fraser and Bickmore 2007; Misteli 2007).

The chromatin of mouse ESCs exhibits a widely dispersed and open architecture that is a defining property of naïve pluripotency in vivo and in vitro (Meshorer et al. 2006; Efroni et al. 2008; Ahmedet al. 2010; Boskovic et al. 2014). This includes constitutive heterochromatic domains such as pericentromeric satellite repeats (Meshorer et al.

2006; Efroni et al. 2008; Fussner et al. 2011; de Wit et al. 2013). A dispersed chromatin configuration characterizes pluripotent epiblast cells within the mouse blastocyst (Ahmed et al. 2010; Boskovic et al. 2014). Chromatin is compacted upon differentiation, creating a repressive environment (Meshorer et al. 2006; Efroni et al. 2008; Wen et al. 2009; Ahmed et al. 2010; Wijchers et al. 2015). But the open chromatin structure associated with pluripotency creates a transcriptionally permissive environment for regulating access of factors to the genome (Gaspar-Maia et al. 2011; Cavalli and Misteli 2013).

Reducing the expression of several epigenetic regulators like Chd1, or Padi4, in ESCs increases the amount of condensed heterochromatin, disrupts self-renewal, and alters differentiation potential (Meshorer et al. 2006; Gaspar-Maia et al. 2009; Lessard and Crabtree 2010; Christophorou et al. 2014). Pericentromeric heterochromatin (PCH) in somatic cells contains compact fibers and high levels of histone H3 Lys9 trimethylation (H3K9me3) (Peters et al. 2001; Lehnertz et al. 2003). However, in ESCs PCH appears open and decondensed, with relatively low levels of H3K9me3 (Meshorer et al. 2006; Efroni et al. 2008).

Novo et al. (2017) reported that Nanog is necessary and sufficient for PCH organization in mouse ESCs. Deletion of Nanog leads to compaction of heterochromatin domains, whereas forced expression of NANOG in epiblast stem cells (EpiSCs) is sufficient to decondense PCH organization and redistribute constitutive heterochromatin domains. It was found that NANOG associates with satellite repeats within PCH domains to promote the overall dispersed heterochromatin architecture of ESCs (Novo et al., 2017). NANOG actively regulates heterochromatin through its C-terminal transactivation domain which is recruited directly to the major satellite DNA within the PCH domains (Novo et al., 2017). Heterochromatin is readily compacted upon ESC differentiation and embryo development, implying that the epigenetic regulation of chromatin is coordinated with the cell’s differentiation status (Meshorer et al. 2006; Ahmed et al. 2010; Boskovic et al. 2014). Novo et al. (2017) suggest that down-regulation of Nanog early in ESC differentiation leads to heterochromatin compaction, and that its absence is a compromising factor in Nanog-null ESCs, which exhibit diminished colony formation and are more prone to spontaneous differentiation than wild-type ESCs (Mitsui et al. 2003; Chambers et al. 2007).

Nanog also has a repressive role, and can negatively regulate genes associated with differentiation. Liang et al. (2008) found that, together with Oct4, Nanog regulates pluripotency and cell fate by interacting with several repressive complexes such as NuRD, Sin3A and Pml. Nanog and Oct4 associate with NuRD to form the NODE (Nanog and Oct4 associated deacetylase) complex. RNAi knockdown of NODE subunits instigates an increased expression of developmentally regulated genes, and drives ES-cell differentiation (Liang et al., 2008).

In human pluripotent stem cells (hPSCs), Nanog directly associates with both Smad2/3 and DPY30 to recruit the complex of proteins associated with Set1 (COMPASS) in response to Nodal signalling (Bertero et al., 2015). COMPASS maintains a specific pattern of gene expression throughout development by facilitating the methylation of H3K4 residues in histones bound to Nodal-responsive genes. NANOG, which regulates its own expression, and that of other pluripotency factors, is part of a broad SMAD2/3-dependent transcriptional network characteristic of hPSCs (Mullen et al. 2011; Teo et al. 2011). The loss of H3K4me3 upon Activin/Nodal inhibition in hESCs (human embryonic stem cells) quickly switches off the expression of master pluripotency regulators, including NANOG, and subsequently leads to differentiation (Bertero et al., 2015). Wang *et al* (2006) had previously illustrated a Nanog interactome for mouse ESCs highlighting its physical interaction with the repressive Polycomb complex, associated with the deposition of H3K27me3. The combination of activating H3K4me3 and repressive H3K27me3 marks creates a state permissive to nodal signalling, which is necessary to establish pluripotency (Bertero *et al*, 2015). This concurs with earlier findings by Bernstein et al. (2006), which described bivalent domains consisting of large regions of H3K27me and smaller regions of H3K4me that are associated with low expression of developmental genes in pluripotent cells, while keeping them poised for activation upon differentiation.

1.1.5 Cellular Reprogramming

The first established method for nuclear reprogramming was through somatic cell nuclear transfer (SCNT). This was first reported by Briggs and King (1952) when they transferred nuclei from frog blastulae into enucleated eggs. The transplanted nuclei could competently direct early development, with up to 40% of the recipients giving rise to blastulae, before then aborting at the late blastula or early gastrula stages. John Gurdon (1958) then employed SCNT to successfully clone a frog using intact nuclei from the somatic cells of a Xenopus tadpole, demonstrating that mature cells can be reprogrammed to become pluripotent. This ability was shown not to be exclusive to amphibians when Wilmut et al. (1997) produced “Dolly” the sheep from the transplanted adult nucleus of a mammary epithelial cell.

In vitro reprogramming was first established when adult thymocytes were fused with ESCs (Tada et al., 2001). Rearrangement of the T cell receptor and immunoglobin genes demonstrated hybridization of the ESCs with the somatic cells. Reactivation of the X chromosome in female thymocytes, along with activation of the pluripotency specific transcription factor Oct4, signalled a transition to pluripotency. This was confirmed when the hybrid cells contributed to all three germ layers in chimeric embryos (Tada et al., 2001). Hou et al. (2013) demonstrated that exposure to a specific cocktail of small chemical compounds can also reprogram somatic lineages to a naïve state. However, it was the reprogramming of cells through the forced expression of defined transcription factors to produce iPSCs (Takahashi and Yamanaka, 2006) which established reprogramming of somatic cell nuclei as a model for studying development and disease, simultaneously opening up the potential for patient-specific cell-based therapy in regenerative medicine.

iPSCs were first obtained from mouse embryonic and adult fibroblasts, which were transduced with the reprogramming factors Oct4, Sox2, Klf4, and c-Myc (together referred to as OSKM) which are now commonly known as the "Yamanaka factors" (Takahashi and Yamanaka, 2006). Methods such as SCNT or cell fusion based reprogramming immediately expose a somatic nucleus to a pluripotent environment and this allows for a rapid reprogramming of the cell. By contrast, the process of transcription-factor-mediated reprogramming is a long and inefficient step-wise process, and the products vary greatly when profiled. Immature stages of reprogramming are progressively marked by intermediate cell types, where the changes induced by the exogenous cues are still on-going (Takahashi and Yamanaka, 2015). Yet the factors that initiate the process are known and can easily be modulated, allowing for a precise dissection of the different stages of the reprogramming process (Buganim et al., 2013).

Gene expression profiling and RNA interference first identified three separate phases of MEF reprogramming: initiation, maturation and stabilization (Li et al., 2010; Samavarchi-Tehrani et al., 2010; Golipour et al., 2012). But the reprogramming process is now more broadly divided into two phases: stochastic and deterministic (Fig.1; Buganim et al., 2013). Initially there is a long phase of stochastic gene activation induced by OSKM where the cells are directed to one of several possible fates, such as apoptosis, senescence, transformation, transdifferentiation or full reprogramming to pluripotency (Buganim et al., 2012; 2013). OSKM will open chromatin regions and allow the activation of genes that increase proliferation, initiate mesenchymal-to-epithelial transition (MET) and activate DNA repair and RNA processing (Li et al., 2010; Samavarchi-Tehrani et al., 2010; Buganim et al., 2012). There is then a rate-limiting step which contributes to the long latency of the reprogramming process (Buganim et al., 2013), before the instigation of a shorter, more hierarchal, deterministic second phase, which begins with the activation of the Sox2 locus (Buganim et al., 2012). In this late phase, the cells eventually stabilize into the pluripotent state, in which the transgenes are silenced, the cytoskeleton is remodelled to an embryonic stem cell (ESC)-like state, the epigenome is reset and the endogenous core pluripotency circuitry is activated (Golipour et al., 2012; Polo et al., 2012; Hansson et al., 2012; Buganim et al., 2012).

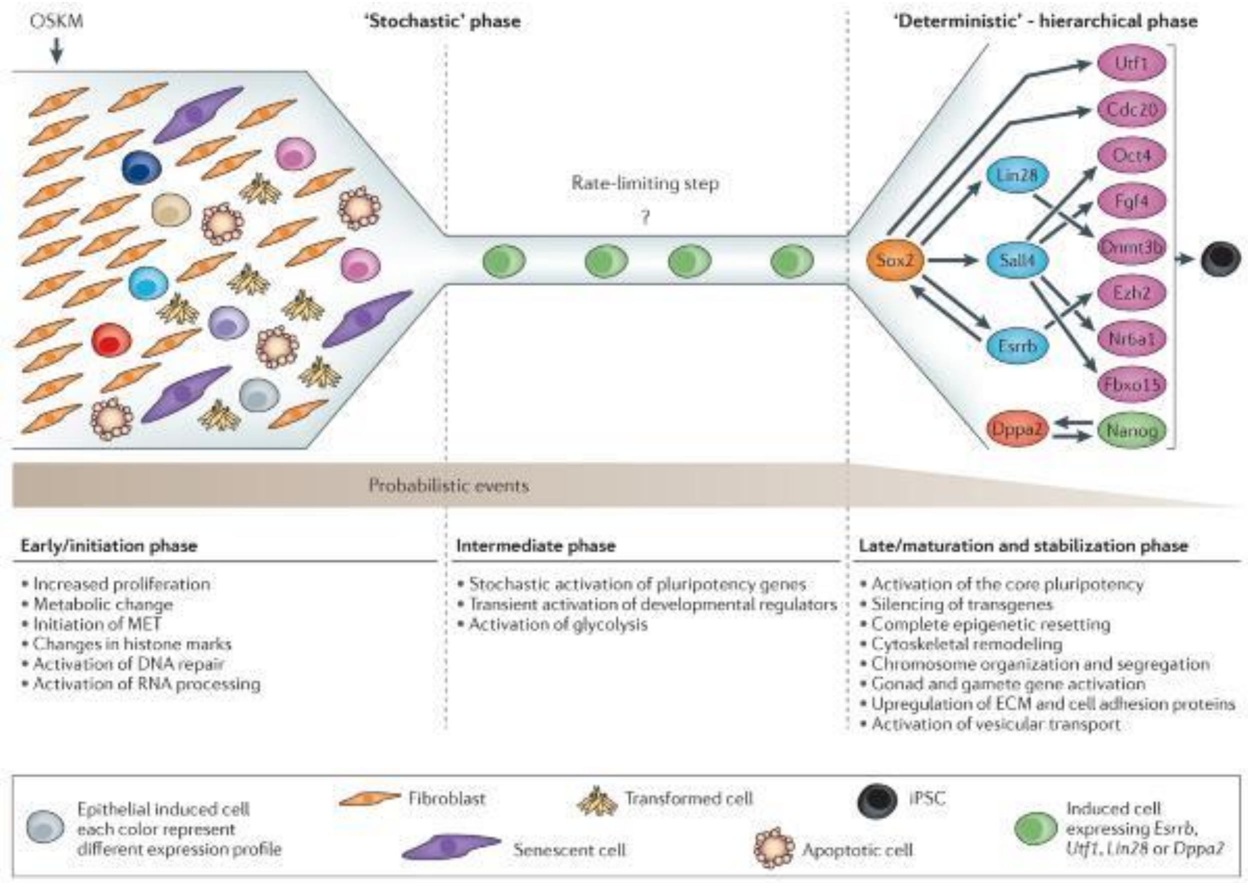


Figure 1: Phases of the reprogramming progress (Buganim et al., 2013).

Buganim et al. (2013) divide the reprogramming process into two phases: firstly, a long `stochastic' phase of gene activation; and secondly, a shorter hierarchical more `deterministic' phase of gene activation that begins with the activation of the *Sox2* locus.

From SCNT and fusion based experiments to transcription factor and small molecule mediated reprogramming, the efficiency of reprogramming varies greatly depending on which method is used. In the case of transduction with OSKM, most cells initiate the reprogramming process, but few successfully make it to ground-state pluripotency (Yamanaka, 2009). The binding sites of the individual OSKM factors were mapped to the human genome during the reprogramming of fibroblasts by Soufi et al. (2012), and these studies found that c-MYC promotes the binding of the other three factors to enhancer regions that promote reprogramming. However, repressive epigenetic regulation in the form of H3K9me^3 was present at many of these loci preventing the reprogramming factors from accessing the chromatin (Soufi et al., 2012). As a consequence of this, during the stochastic phase of reprogramming cells may be diverted to a fate other than reprogramming. Sometimes they can be directly reprogrammed to another somatic lineage such as cardiomyocytes (Efe et al., 2011), or neural progenitor cells (Kim et al., 2011). Other times the cells are induced to a metastable state. During chemical reprogramming of fibroblasts, cells pass through an extra-embryonic endoderm (XEN)-like state (Li et al., 2017). These XEN-like cells could be stably expanded in this state, but they also maintained the potential for other lineages, and could be induced to both hepatocytes and functional neurons (Li et al., 2017).

The level of expression of each pluripotency factor has also been shown to directly affect the efficiency of reprogramming. A higher level of both Oct4 and Klf4 (Tiemann et al., 2011), while simultaneously lowering the level of Sox2 (Yamaguchi et al., 2010), has been shown to enhance reprogramming. Also, each of the four factors in the OSKM reprogramming cocktail have important individual roles in development, and they regulate the transcription of targets unrelated to the reprogramming pathway (Payne et al., 1982; Nichols et al., 1998; Avilion et al., 2003; Jiang et al., 2008). Thus, an unbalanced stoichiometry may cause the OSKM factors to favour another pathway (Tiemann et al., 2011). During the immature phases of reprogramming, Oct4 and Sox2 have been shown to upregulate mesendodermal and neuroectodermal genes respectively (Shu et al., 2013), while Klf4 can promote epidermal gene expression (Kim et al., 2015). An unbalanced expression of OSKM can therefore divert partially reprogrammed cells away from the path to pluripotency, and prevent them from completing the transition (Tanabe et al., 2013). Furthermore, heterochromatin regions impede the establishment of pluripotency; its decompaction is a critical event that occurs at a late stage in reprogramming (Brambrink et al. 2008; Fussner et al. 2011; Schwarz et al. 2014). Therefore, reprogramming efficiency can be improved at an epigenetic level by forcing the heterochromatin to decondense with inhibitors of both DNA methyltransferases and histone deacetylases, or by genetic depletion of H3K9 methyltransferases (Huangfu et al. 2008; Mikkelsen et al. 2008; Soufi et al. 2012; Sridharan et al. 2013).

Among the earliest events in the reprogramming of fibroblasts is that the cells increase their division rates from a fibroblastic cell cycle of about 22 hours to a shorter ES-like cell cycle of approximately 11-12 hours after exogenous expression of the OSKM factors (Smith et al., 2010). The cells exhibit a change in nuclear to cytoplasmic ratio, as their morphology proceeds to a smaller, rounder, phenotype as the quickly dividing cells expand firstly as a monolayer, before the smaller cells begin to form compact colonies associated with a MET transition (Smith et al., 2010). At a more mature stage of reprogramming, cells begin to express pluripotency associated markers such as Alkaline Phosphatase (AP) and the surface marker SSEA1 (Stadtfeld et al., 2008). The expression of SSEA1 is maintained throughout the reprogramming process. Cells which stain positive for SSEA1 exhibit transient expression of miR-302a, which is associated with the post-implantation epiblast and EpiSCs, but is not present in ESCs (Polo et al., 2012). This suggests that the cells pass through a "primed" epiblast-like state during the transition to naïve pluripotency (Takahashi and Yamanaka, 2015).

In addition to established iPS cell colonies, transduction with OSKM can give rise to a partially reprogrammed cell type that can be stably maintained without ever making the transition to naïve pluripotency. These cells are called "pre-iPS" cell lines, and they appear ES-like in morphology, but they do not express endogenous pluripotency factors like Nanog, or associated downstream markers such as EsrrB and Rex1 (Papp and Plath, 2011).

Nanog is not part of the transcription factor cocktail required to initiate the reprogramming of mouse somatic cells to pluripotency as identified by Takahashi and Yamanaka (2006). However, its expression is required to progress reprogramming beyond the point of dedifferentiation and acquire the naïve pluripotent state (Silva et al., 2009). Therefore Nanog expression can be used as an indicator of fully reprogrammed iPSCs (Okita et al., 2007). Silva et al. (2009) demonstrated that Nanog-null pre-iPS cells do not reach ground state pluripotency; however, once Nanog expression is restored to the stalled pre-iPS cells they can then transition to the ground state, cooperatively promoted by transfer of the stable transfectants to the optimized 2i + LIF mESC ground state culture conditions. The inability of Nanog-null cells to progress beyond a pre-iPS state provides an experimental system to assay for Nanog activity in related genes, a system which might elucidate the origin of pluripotency factors (Theunissen et al., 2011).

1.1.6 Pluripotency is conserved

Pluripotency is well characterized in mammals, yet its evolutionary history is not well understood. Nieuwkoop demonstrated that the primitive ectoderm (animal cap) of embryos from axolotls can be induced to form PGCs and somatic lineages (Nieuwkoop, 1969; Sutasurya and Nieuwkoop, 1974), which are the defining characterisitics of pluripotent cells in mammals. This suggests that pluripotency is conserved between urodele amphibians and mammals (Johnson et al., 2001; Johnson et al., 2003a; Johnson et al., 2003b).

The existence and equivalent functions of orthologous pluripotency genes in non-mammalian vertebrates has been long debated. Functional assays were used to identify the zebrafish *pou2* gene as the *Oct4* orthologue (Burgess et al., 2002), but its primary role was in the specification of endoderm (Reim et al., 2004; Lunde et al., 2004), bring its assignment as an Oct ortholog into doubt. Three Pou factors with significant homology to Oct4 were identified in Xenopus laevis, and XlPou91 could compensate for Oct4 depletion in mESCs (Morrison and Brickman, 2006). Then Lavial et al. (2007) proposed the existence of an avian Oct4 homolog (cPouV) due to its expression in chicken ESC (cESC), which also express an ortholog of Nanog. Previous to this report, the relationship between *Oct4* and *Nanog* in conducting pluripotency had only been demonstrated in mammals. Yet syntenic comparison of the different Pou factors revealed a relationship between chicken PouV, xenopus Pou91 and zebrafish pou2 genes that is absent in mammalian species (Lavial et al., 2007; Frankenberg et al., 2009). It is now concluded that Oct4 is not conserved in Xenopus, chicken or zebrafish, the three historical non-mammalian models for vertebrate development (Frankenberg and Renfree, 2013; Frankenberg et al., 2014).

In zebrafish, a Nanog-like gene (zNanog-like) was isolated based on homology within the homeodomain, but no chromosomal synteny exists between zNanog-like and verified Nanog orthologues (Schuff et al., 2011; Xu et al., 2012; Theunissen et al., 2011; Wang et al., 2016). Xu et al. (2012) reported that zNanog-like regulates formation of the extraembryonic lineage which provides the Nodal signalling required to instigate endoderm specification through the Mxtx2-Nodal pathway. Schuff et al.

(2011) reported that zebrafish Nanog-like prevents murine embryoid body (EB) differentiation. However, contradictory to this, Xu et al. (2012) revealed that overexpression of zebrafish nanog-like is not sufficient to rescue LIF dependence in murine ES cells. Nanog is essential for establishing pluripotency in the mammalian embryo (Mitsui et al., 2003; Chambers et al., 2007; Silva et al., 2009), and as stated, this role appears conserved in axolotl (Johnson et al., unpublished).

In zebrafish embryos zNanog-like and pou2 activate transcription at the MBT (Lee et al. 2013), which is not conserved in axolotls (Johnson and colleagues, unpublished).

However, zNanog-like does maintain the ability to facilitate the transition of pre-iPS cells to naïve pluripotency, a fundamental characteristic of Nanog which is conserved in the homeodomain (Theunissen et al., 2011).

Activation of Nanog marks the acquisition of pluripotency in both development and reprogramming (Mitsui et al., 2003; Silva et al., 2009), so the discovery of a Nanog orthologue in axolotls (axNanog) by Dixon et al. (2010) (Fig. 2) suggests this function may be conserved. Prior to this, Oct4 was defined as a master regulator of pluripotency by Niwa et al. (2000), and was shown to be conserved in axolotls by Bachvarova et al. (2004). Together, these reports suggest that the transcription factors which regulate pluripotency are conserved from urodeles to mammals, which is consistent with functional demonstrations of pluripotency in axolotl animal caps (Boterenbrood and Nieuwkoop, 1973; Sutasurja and Nieuwkoop, 1974; Chatfield et al. 2014). The axNanog gene is encoded by four exons, with an intron/exon structure conserved in mammalian Nanog genes (Dixon et al., 2010)(Fig. 2). Homology of axNanog with other Nanog sequences is evident only in the homeodomain (HD) and caspase cleavage site (CCS) (Fujita et al., 2008) (Fig. 2), but comparative mapping demonstrated conserved synteny, and established orthology amongst the Nanog genes of axolotl, human and chicken (Smith et al., 2005; Dixon et al., 2010) (Fig. 2).

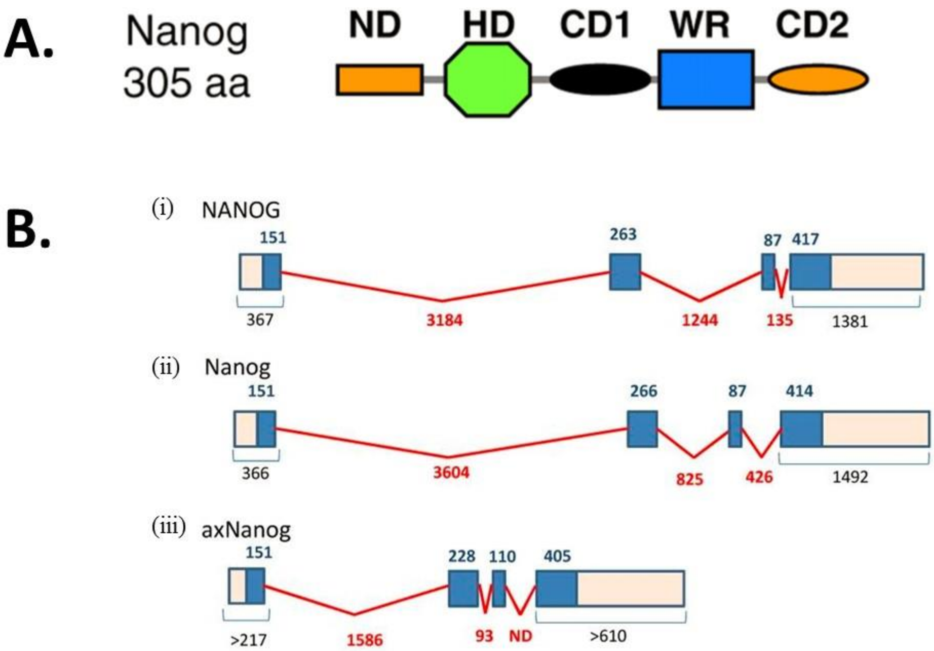


Figure 2: The structure of Nanog
(A) The mammalian Nanog structure (From paper: Chambers and Tomlinson, 2009). The Nanog protein can be divided into N-terminal and C-terminal halves. The N-terminal half contains a DNA-binding homeodomain (HD) and an N-terminal domain (ND). The C-terminal half contains a dimerisation domain (blue) referred to as the tryptophan repeat (WR). (B) Conservation of *Nanog* genomic structure (From paper: Dixon et al., 2010). Intron/exon structures of (i) Human NANOG (ii) Mouse Nanog (iii) Axolotl Nanog are aligned. Blue boxes denote protein coding regions. Numbers in blue represent amino acids. Numbers in red are intron lengths; ND, not determined. Black numbers indicate combined length of protein coding region with 5′ and 3′ untranslated regions.

Significantly, axNanog does not contain the WR domain, which is exclusive to the Nanogs of eutherian mammals (Fig. 2). Dixon et al. (2010) suggested that WR domain evolution is associated with the rapid expansion of the epiblast during development. The WR domain promotes cell proliferation in vitro which is integral to the isolation of stem cells from early eutherian embryos (Evans and Kaufman, 1981; Thomson et al., 1998; Schneider et al., 2007; Tesar et al., 2007; Brons et al., 2007; Buehr et al., 2008; Honda et al., 2009; Alberio et al., 2010). As the WR domain facilitates the homodimerization that allows mammalian Nanogs to promote ESC self-renewal in culture (Mullin et al., 2008; Wang et al., 2008), Dixon et al. (2010) found that addition of a mouse WR domain to axNanog confers this same ability to the gene. However, axNanog can regulate pluripotency as a monomer (Dixon et al., 2010). Overexpression of axNanog in ESCs enhanced the transfer of pluripotency to neural stem cells (NSCs) after fusion, an established attribute of mouse Nanog (Silva et al., 2006; Silva et al., 2009; Dixon et al., 2010). The efficiency of hybrid colony formation was unaffected by the capacity of axNanog to form homodimers (Dixon et al. 2010), which demonstrates that axNanog participates with factors in ESCs to program pluripotency in mammalian cells. This was also exhibited by the ability to suppress differentiation, with both the monomeric and dimeric states proving equally competent at inhibiting lineage specific markers in the embryoid bodies (EBs) formed from ESCs constitutively expressing axNanog (Dixon et al., 2010). More recently, a second Nanog gene was identified in Axolotl by Johnson and colleagues (Unpublished). This second axolotl Nanog orthologue displays a differential expression profile to axNanog and is maternally inherited as RNA, therefore it has been dubbed ‘Eggnog’.

1.1.7 Reprogramming capacity of Nanog is conserved in the homeodomain

Nanog shows relatively low sequence conservation across species in which it is found, so Theunissen et al. (2011) tested the functional capacity of Nanog orthologues to reprogram Nanog-null somatic cells to pluripotency. Theunissen et al. (2011) tested whether the transcriptional regulators of naive pluripotency are conserved. Most genes associated with pluripotency and reprogramming are highly conserved between eutherian mammals, yet mouse and human Nanog orthologues share only 54% sequence identity (Theunissen et al., 2011). Fully reprogrammed iPS cells can be isolated by screening for activation of the endogenous Nanog locus (Maherali et al., 2007; Okita et al., 2007; Wernig et al., 2007). Somatic cell reprogramming does not progress to full pluripotency without Nanog, even in the 2i + Lif culture conditions optimised to support naïve pluripotency (Silva et al., 2009). Thus, Nanog can be seen as a molecular switch that controls the establishment of pluripotency during reprogramming (Theunissen and Silva, 2011), and allowed Theunissen et al. (2011) to test Nanog orthologues for the capacity to reprogram N-/- mouse pre-iPS cells to naïve pluripotency.

They first established that mammalian orthologues of Nanog in Rat and Human were sufficient to restore the reprogramming capacity of Nanog $^{-/-}$ ($N^{-/-}$) pre-iPS cells, before then testing genes isolated from non-eutherian vertebrate species (Theunissen et al., 2011). Both chick Nanog and zNanog-like were capable of replacing mouse Nanog in reprogramming the cells to ground state efficiently, confirmed by the upregulation of endogenous pluripotency genes such as Rex1 and Klf2, and silencing of the exogenous factors (Theunissen et al., 2011). They also found that the non-eutherian orthologues could directly regulate mNanog targets including the activation of Oct4, and repression of Xist, downregulation of which is required for X chromosome reactivation (Silva et al., 2009).

Amino acid sequence conservation is confined to the homeodomain (HD) in Nanog orthologues, so Theunissen et al. (2011) tested related HD-containing transcription factors Msx1 and Nkx2.5 for the capacity to reprogram $N^{-/-}$ somatic cells, and neither factor could generate iPSCs (Theunissen et al., 2011).
Nanog is not encoded in the invertebrate genomes available, so to identify an invertebrate precursor, Theunissen et al. (2011) identified the Bsx gene in amphioxus, an invertebrate chordate (cephalochordate) (Bertrand and Escriva, 2011), as having the highest HD sequence identity to Nanog. They also identified the HD-containing Vent1 gene from amphioxus as being a possible candidate for Nanog activity. However neither factor could reprogram the $N^{-/-}$ somatic cells to pluripotency (Theunissen et al., 2011), suggesting that the capacity to do so arose within, and was exclusive to, the vertebrate orthologues of Nanog. As only the homeodomain shows any conservation of sequence between the Nanog orthologues, Theunissen et al. (2011) then reported that the reprogramming activity is conserved within the HD domain. In this regard, they found that transfection with the isolated homeodomain can substitute for the full-length Nanog protein, albeit at low efficiency, indicating that it is sufficient to complete the reprogramming of pre-iPSCs to ground state pluripotency.

1.1.8 The origin of Nanog

Pluripotency is a necessary and transient stage in development, but its evolution is unknown.
Nieuwkoop first suggested that pluripotency is conserved in basal vertebrates, when he demonstrated that the primitive ectoderm of an axolotl embryo could be induced to form both PGCs and somatic cells (Nieuwkoop, 1969; Sutasurya and Nieuwkoop, 1974). This was later confirmed by Chatfield et al. (2014), which demonstrated that pluripotent mesodermal precursors can give rise to multiple somatic lineages, as well as PGCs. This provides support to the hypothesis that pluripotency is conserved in the major trunk of vertebrate evolution (Johnson et al., 2001; Johnson et al., 2003a; Johnson et al., 2003b; Dixon et al. 2010; Theunissen et al., 2011). Thus the regulatory networks which direct early vertebrate development may hold the key to elucidating the origin of pluripotency, perhaps by tracing the evolution of candidate precursors in species where the pluripotency is a crucial step in embryological development. Within this study, I focused my search on species without germ plasm, and therefore without a derived form of embryogenesis (Johnson and Alberio, 2015).

Understanding the evolution of the GRN that governs pluripotency is a necessary step in understanding how pluripotency first emerged in vertebrates. Previous studies have not been able to demonstrate Nanog-like activity in homeobox orthologs identified in invertebrates (Theunissen et al., 2011), Nanog sequence are not encoded in prominent vertebrate models for development such as Xenopus (Koziol et al., 2007), so a question remains as to its origin. Due to its central role in orchestrating pluripotency, several studies have attempted to elucidate the evolutionary history of the gene, and its family, with the hope of better understanding pluripotency and the GRNs that governs it (Banerjee-Basu and Baxevanis, 2001; Booth and Holland, 2004; Theunissen et al., 2011; Scerbo et al., 2014; Pashaiasl et al., 2015).
Pashaiasl et al. (2015) used sequences of 47 Nanog genes from various species to computationally extract two datasets of sequence-specific features which could then be compared. The first dataset analysed 76 nucleotide attributes for each sequence, and the second dataset was based on repeated nucleotide sequences. Computational

algorithms such as decision tree models were applied on these datasets to find the evolutionary pathways of Nanog diversion (Pashaiasl et al., 2015). They reported that the results from their second database identified short sequence strings, such as ACTACT, TCCTGA, CCTGA, GAAGAC, and TATCCC as candidates to identify Nanog genes in various species.

Frankenberg and Renfree (2013) used phylogenetic analysis of POU5F1, POU2 and their flanking genes to elucidate the origin of POU5F1 (Oct4). Scerbo et al. (2014) used a similar approach to understand the evolutionary history of the Nanog. Scerbo et al. retrieved the sequences of known Nanog orthologues from public repositories, and identified exons two and three as the most highly conserved region, which incorporate the homeodomain, and then utilised tblastn searches against vertebrate genomes to identify Nanog paralogs. The searches spanned a broad taxonomic range of craniate genomes, but excluded any invertebrate data. The Novel paralogs they identified are found to exist in tandem duplicates, which they suggest is the result of at least four independent duplication events (Scerbo et al., 2014). Phylogenetic analyses were then conducted on a manually curated protein alignment to produce a strict consensus of the Maximum likelihood (ML) and Bayesian trees (Fig 3.), using the topology of the ML tree as a backbone. They rooted the three with actinopterygii (ray-finned fishes) sequences, and this accordingly lead to the actinopterygian spotted gar paralogues forming a well-supported monophyletic group at the base of the tree. The observed topology shows high conservation in teleosts, which encompasses zebrafish; however, their data demonstrates that the chromosomal synteny of the Nanog paralogs identified is not conserved between actinopterygii and sarcopterygii (Scerbo et al., 2014), the latter clade encompassing tetrapods, from which amniotes are descended (Bachvarova et al., 2009a; Bachvarova et al., 2009b).

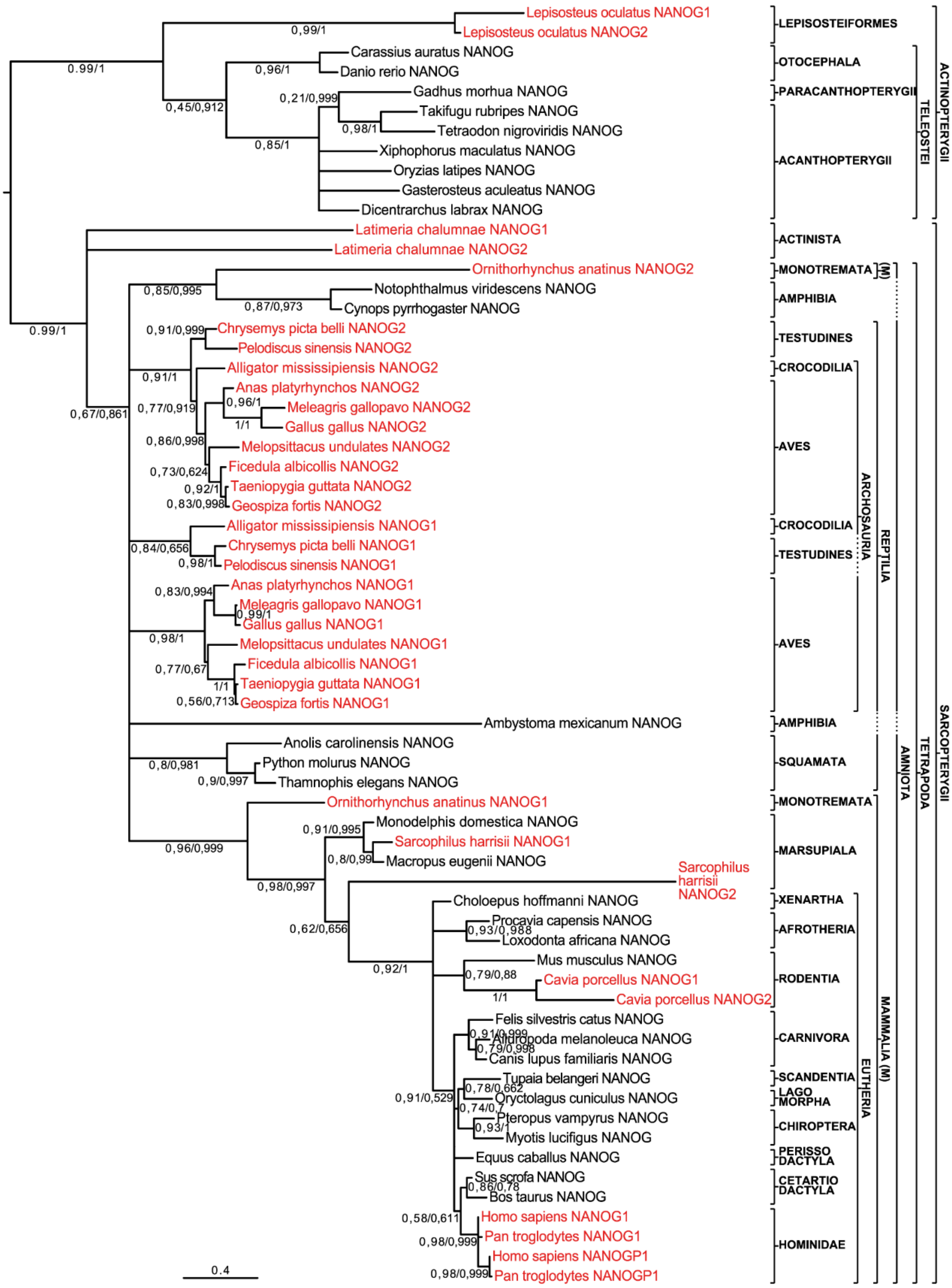


Figure 3. : Phylogenetic analysis of Nanog paralogs in osteichthyans (Scerbo et al. 2014)

ML and Bayesian trees using the translated sequences of exons two and three, to demonstrate the phylogeny of Nanog paralogs; the topology of the ML tree was used as a backbone, with aLRT and PP support values given for nodes that were recovered in both topologies. The actinopterygian sequences form the group used to root the tree. Duplicates are highlighted in red.

As the teleost sequences show high conservation with the actinopterygian sequences used to form the base of their phylogenetic analyses, they suggest that the role of orthologues such as zNanog-like, in controlling cell migration and embryonic morphogenesis (Xu et al., 2012; Zhou et al., 2013), may in fact be the ancient, and most widely conserved, function of Nanog (Scerbo et al., 2014). They further postulate that the role of Nanog in regulating pluripotency was acquired specifically by sarcopterygians (or tetrapods, or amniotes).

However, the embryos of teleost fish evolved from yolky amphibian-like embryos which undergo holoblastic cleavage (Virta and Cooper, 2011). These more slowly developing embryos were characteristic of the ancestor to tetrapods (Johnson et al., 2003b), and resemble those of ancestral actinopterigians, like sturgeons. The divergence of teleost Nanog sequences, and different function in development (Xu et al., 2012), may be due to the emergence of a deterministic embryology (Hashimoto et al., 2004), which may be related to the evolution of germ plasm, removing the requirement for ground state potential (Crother et al., 2007; Camp et al., 2009; Xu et al. 2012;). Conversely, these results raise the possibility that the role of Nanog in ground state pluripotency is conserved in more primitive fish, such as sturgeon, as predicted by their embryology.

The mechanisms that direct vertebrate development were most likely modified from pre-existing GRNs expressed in the embryos of lower chordates (Johnson et al., 2003a; Bachvarova et al., 2009a; Peter and Davidson, 2011). Embryological evidence indicates that the amphibian ancestor of amniotes was urodele-like (Bachvarova et al., 2009a; Bachvarova et al., 2009b). The development of the most widely studied lower chordates, ascidians, is largely determinative (Lemaire, 2011), and ascidian embryos do not manifest ground state pluripotency. Data from species with more regulative embryos, like amphioxus (Branchiostoma floridae) (Yu et al., 2007) and acorn worms (Saccoglossus kowaleskii) (Lowe et al., 2006), also do not reveal an obvious Nanog candidate. Johnson and colleagues (unpublished) suggested that Nanog is a vertebrate innovation, and reasoned that Nanog activity might be conserved in a related ancestral gene. Figure 4 shows the phylogenetic relationship of species from which candidate genes were chosen for this study and assayed for Nanog activity.

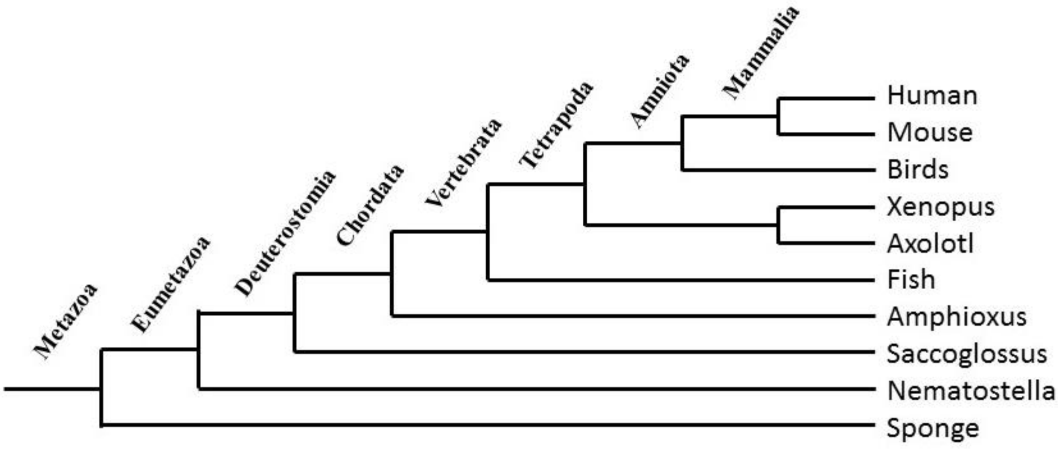


Figure 4: Evolutionary tree of relevant species for this study.

The phylogenetic relationships of the different species referenced in this study according to NCBI Taxonomy.

The Vent gene family was identified as a candidate due to homology with the Nanog homeodomain, and Vent genes were identified in the amphioxus and Saccoglossus genomes, representing cephalochordates, and hemichordates, respectively, the latter of which are positioned at the base of deuterostomes (Bourlat et al., 2006).

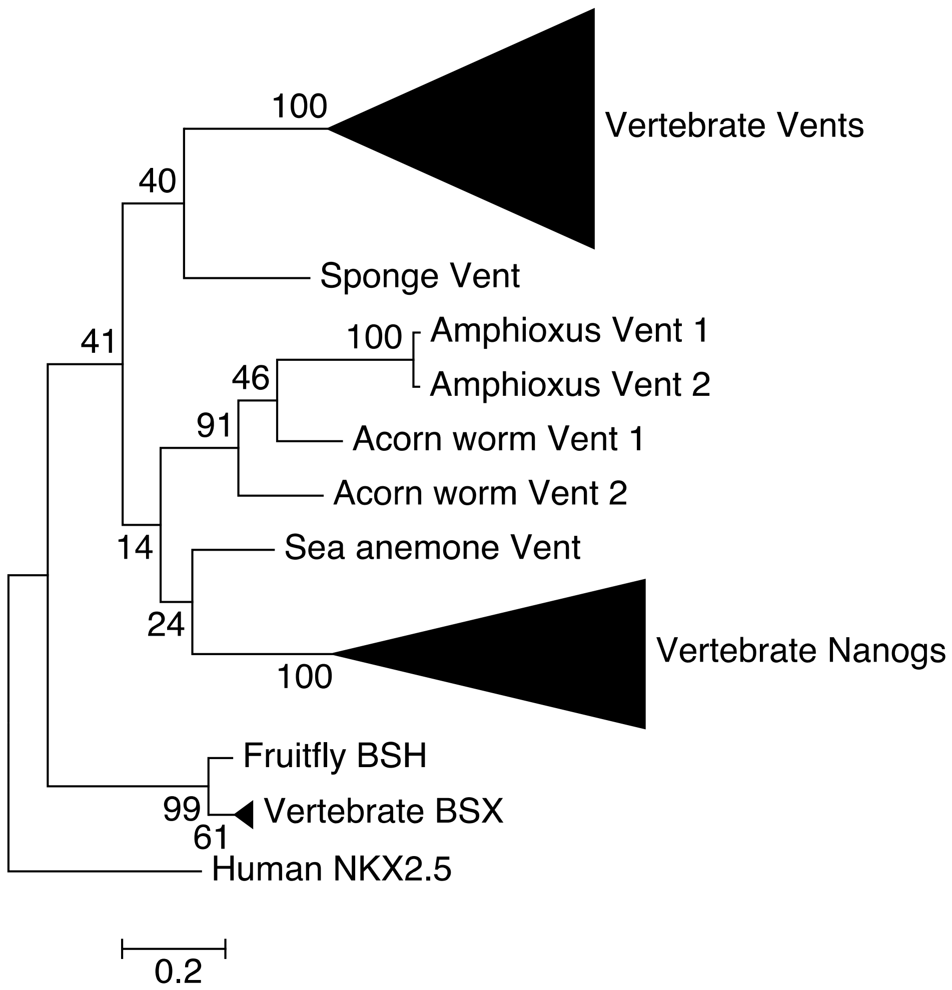


Figure 5: Phylogenetic tree of Nanog, Vent and BSX.

This shows Nanog and Vent grouped together, separate from BSX. Bayesian tree was created using the GTR+G model in Mr. Bayes based on the amino acid alignment of the homeoboxes. Posterior probability values are shown at each node. The tree was rooted on human NKX2.5 and vertebrate clades are compressed. Figure adapted from Evans 2015.

Phlogenetic analysis grouped invertebrate Vent genes with the Nanog and Vent orthologs found in vertebrate genomes. As seen in Figure 5, a Bayesian tree rooted by human NKX2.5 grouped invertebrate Vent genes with vertebrate Nanog and Vent orthologs, suggesting greater homology between Nanog and Vent than the BSX homeobox gene family. This homology raises the possibility that some Nanog-like activity may reside in the Vent gene family. The Vent transcription factor family was identified in Xenopus, where they regulate mesoderm patterning (Onichtchouk et al., 1996). XVent-2, a founding member of the family, has greater similarity to Nanog than any other sequence in the Xenopus genome, and it clusters with Vents from axolotls and other vertebrates. Scerbo et al. (2012) suggested that the VentX (Vent-1 and Vent-2) provides Nanog-like governance of developmental potential during Xenopus development. Their study reports that joint overexpression of Vent1 and Vent2 counteracts lineage commitment, whereas inversely, their downregulation facilitates premature differentiation. Scerbo et al. (2012) suggested that Nanog and Vent were interchangeable in a Xenopus embryo, and that expression of mouse Nanog could specifically rescue axis formation in *Ventx1/2* deficient embryos. However, this contradicts a study by Schuff et al. (2012) which compared Vent and Nanog genes in Xenopus and Zebrafish, respectively, and found that XVent1/2 and Nanog cannot substitute for each other. Depletion of Nanog in zebrafish cannot be rescued by ectopic expression of Xvent, and Xvent depletion in Xenopus cannot be overcome by ectopic expression of zebrafish Nanog.

Johnson and colleagues (Unpublished) investigated the evolutionary relationships between Nanog and Vent families including the two Vents genes from amphioxus (amphiVent1, 2) and Saccoglossus (SkVent1, 2). The phylogenetic tree groups Saccoglossus and amphioxus Vents together, falling outside the distinct vertebrate Nanog and Vent groups (Fig. 5). The identity of amphioxus genes as Vent family members is unambiguous, as they retain synteny with mammalian Vents (Luke et al., 2003). Addressing the possibility that the invertebrate Vent genes may also be related to Nanog, the expression of Saccoglossus Vent1 and Saccoglossus Vent2 was analyzed during development; Saccoglossus Vent1 RNA is maternally inherited and is found throughout the embryo, until its expression drops in mid neurula stages (Johnson et al., Unpublished). In contrast, Saccoglossus Vent2 is first expressed at midblastula stage, and is found in primitive ectoderm and endoderm (Johnson et al., Unpublished). Expression peaks at gastrulation, and is extinguished prior to neurulation, in a pattern very similar to the axNanog gene (Dixon et al., 2010). Thus, the expression pattern of Saccoglossus Vent2, in particular, is consistent with it being related to an invertebrate Nanog ancestor. Johnson and colleagues (Unpublished) conducted preliminary assays for Nanog activity in Vent genes, by testing if their constitutive expression in wildtype ESCs was suffienient to confer Nanog-like cytokine independence in cells cultured in the absence of LIF (Fig. 6).

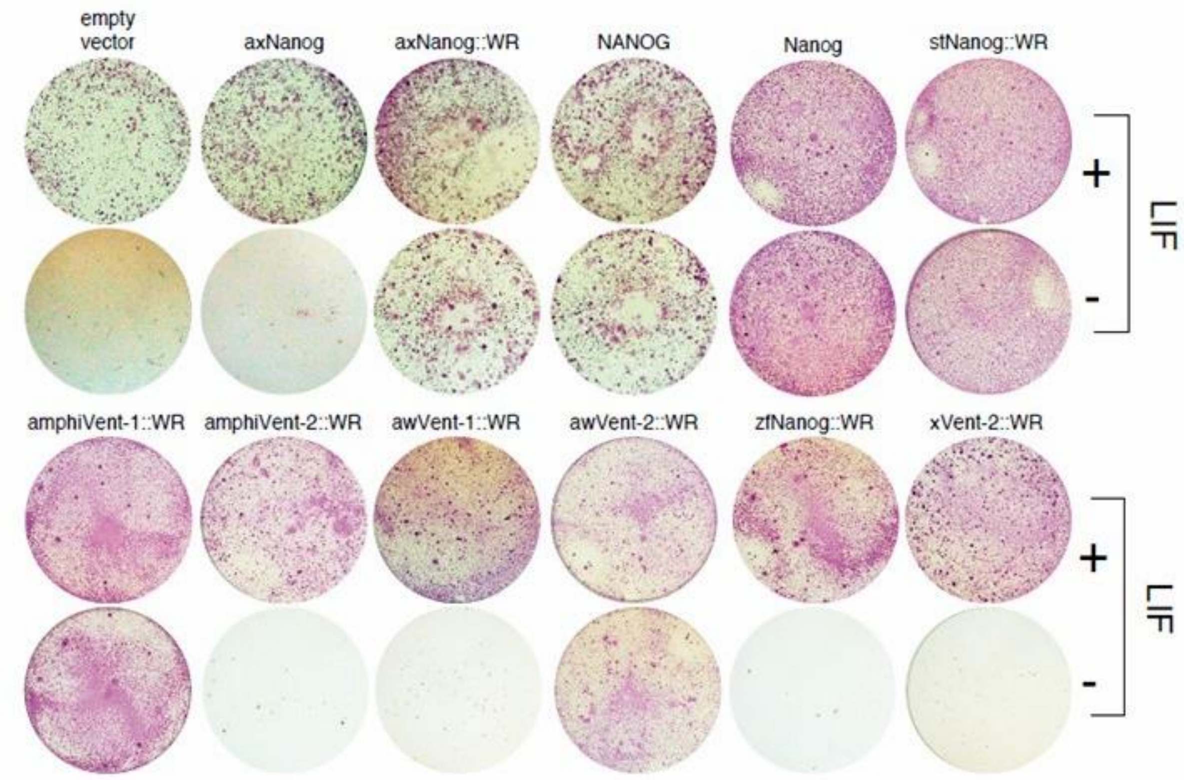


Figure 6: Alkaline Phosphatase staining of mESC lines in cytokine independent self-renewal assay.

AP staining indicates mESC lines expressing Nanog or Vent genes which maintained pluripotent colonies in the absence of LIF. (Johnson and colleagues, unpublished)

An individual Vent gene from amphioxus (Vent1) and Saccoglossus (Vent2), each tailored to contain a mouse WR domain on the carboxy terminus, demonstrated the ability to rescue the self-renewal of ESCs in the absence of LIF, a characteristic attribute of Nanog (Chambers et al., 2003). Xenopus Vent2 did not exhibit any Nanog-like activity, concurring with the findings of Schuff et al. (2012) that Xenopus Vents cannot substitute for Nanog. The relationship of Nanog with other homeoboxes has been debated (Theunissen et al., 2011; Schuff et al., 2012; Scerbo et al., 2012; Scerbo et al., 2014). BSX has been suggested to be the closest ortholog (Theunissen et al., 2011). However, the Vent gene family demonstrates conserved Nanog function (the rescue of LIF-independent self-renewal) (Fig. 6), and it occupies a phylogenetic position suggesting a direct relationship with Nanog (Fig. 5). Here I investigated this possibility more thoroughly.

1.2 AIMS AND OBJECTIVES OF THIS STUDY

My working hypothesis was that Nanog activity is conserved in invertebrate Vent genes, which are closely related to Nanog. My aim was to assay the Vent genes found in basal invertebrate genomes for Nanog reprogramming activity required by pre-iPS cell lines to transition to ground state pluripotency (Silva et al., 2009), which would provide an insight into the ancestry of Nanog.

CHAPTER 2

2 GENERAL MATERIALS AND METHODS

2.1 CELL LINES

All cell lines were cultured at 37°C in 5% CO_2 conditions. Unless otherwise stated, cells were passaged with 0.1% Trypsin-EDTA at 37°C, after washing with 1X PBS. Detached cells were collected in serum-containing media to deactivate the trypsin before pelleting by centrifugation for 5 minutes. Pelleted cells were resuspended in fresh media containing serum, counted using a haemocytometer, and plated. For cryopreservation of cell lines, pelleted cells were resuspended in a freeze medium containing 10% Dimethyl sulfoxide (DMSO) (SIGMA), 60% Foetal bovine serum (FBS) (SIGMA) and 30% complete medium. Cells were aliquoted into cryopreservation tubes, stored in a Mr Frosty cooling rack at -80°C, before transfer to liquid nitrogen tank for long term storage.

The following cell lines were used in this study:

- HEK 293T Human Embryonic Kidney cells
- CGR8 Mouse Embryonic Stem cells
- Neural Stem cell derived Nanog-null pre-iPS cells (Silva et al., 2009)
- TNGA MEF pre-iPS cells, derived from a TNGA mouse embryo (Chambers et al., 2007)
- Nanog-null mouse embryonic stem cells (Chambers et al., 2007)
- IMA81 E11 XEN cells (Kunath et al., 2005)

2.2 Cell Media

2.2.1 HEK 293T culture media

Cultured in high glucose Dulbecco's Modified Eagles Medium (DMEM) (SIGMA) supplemented with 10% FBS, 1% Penicillin/Streptomycin (SIGMA), 2mM L-Glutamine (SIGMA) and 2mM non-essential amino acids (Hyclone).

2.2.2 Basic ESC Serum Lif media

Wild-type mouse ES cells of the CGR8 (ECACC 07032901) line, TNGA MEF pre-iPSCs, and the Nanog null neural stem cell (NS) derived pre-iPS cells were all cultured on 0.1% gelatine. Cells were cultured in high glucose DMEM supplemented with 10% Foetal bovine serum (FBS), 1% Penicillin/Streptomycin, 2mM L-Glutamine, 0.1mM β-Mercaptoethanol (BME), 1mM Sodium pyruvate, and 2mM non-essential amino acids. Cytokine leukaemia inhibitory factor (LIF) was added fresh at 100ng/ml prior to use.

2.2.3 Ground state 2i + LIF media

2i + LIF media was used to culture ground state ESCs, and to mediate the reprogramming of pre-iPS cell lines to pluripotency. Media consisted of a 1:1 mix of DMEM/F12 and Neurobasal media, supplemented with 1% N2 supplement, 10% B27 supplement, 2mM non-essential amino acids, 1% Penicillin/Streptomycin and 0.1mM β-Mercaptoethanol (BME). LIF (100ng/ml) and small molecule inhibitors were added fresh to the media just prior to use. PD0325901 (Tocris) was used at a final concentration of 1uM, and CHIR99021 (Tocris) was used at a final concentration of 3uM.

2.2.4 Basic XEN culture media

For culture of XEN cells, I used complete Glasgow Minimal Essential Medium (GMEM) (SIGMA), supplemented with 10% FBS, 2mM non-essential amino acids, 2mM L-Glutamine, 1mM sodium pyruvate, and 0.1 mM β-mercaptoethanol.

2.3 MOLECULAR CLONING

2.3.1 Construction of plasmids

The coding sequence of each candidate gene was amplified by standard PCR using the Q5 High-Fidelity polymerase ready mix by New England Biolabs (NEB), and Sigma primers. The completed PCR reactions were run on a 1% agarose gel beside a NEB 1KB Molecular marker DNA ladder, and the appropriate band of PCR product was excised from the gel using a sterile scalpel. The product was column purified using the QIAquick Gel Extraction kit (Qiagen).

The following vectors were used in this study:

- pSIN-EF2-Puro
- PiggyBac-PGK-HYGRO
- pLenti-PGK-HYGRO

The appropriate vector, and the purified PCR product, were digested by restriction enzyme double-digest for one hour at 37^{o}C using NEB enzymes and buffers in aqueous conditions. These molecules were gel purified, and then ligated overnight using Promega T4 ligase and buffer. The ligation mix was then transformed under standard conditions into chemically-competent DH5 alpha cells and streaked on LB agar plates containing 100ug/ml ampicillin, to be incubated overnight at 37^{o}C. Colonies were picked by pipette tip, and added to mini-cultures of LB broth containing 100ug/ml ampicillin, and incubated on a shaker-rack for a maximum of 16 hours at 37^{o}C.

2.3.2 Plasmid mini preparation

After incubation, 1ml of the culture was pelleted and supernatant was removed. Pellet was resuspended in 250µl buffer P1. Cells were lysed using 250µl buffer P2 and incubated at room temperature for 3 minutes before adding 300µl neutralisation buffer P3. Lysed mini-cultures were centrifuged at 13,000rpm for 15 minutes to remove cellular debris. 700µl of the supernatant was transferred into a separate tube. Chloroform was added and the mixture vortexed. The sample was spun at 17,000g for 10 minutes to allow separation of the phases. Five hundred microliters of the aqueous phase was transferred to a separate tube and equal volume of isopropanol was added before briefly vortexing and centrifugation for 10 minutes at 13,000rpm. The supernatant was discarded and the pellet of plasmid DNA was washed with 70% ethanol. The pellet was air dried and resuspended in nuclease-free water. A restriction digest was performed, and subsequently run on 1% agarose gel, to screen clones for successful insertion of cDNA fragment into the ligated vector. Upon confirmation of successful plasmid assembly, the rest of the corresponding mini-culture was purified according to the standard protocol of a QIAGEN mini-prep kit, and quantified on the NanoDrop 1000, with a 260:280 value of 1.8 considered pure. Samples of each plasmid were then sequence verified by Source Bioscience Sequencing, Nottingham UK.

2.4 Gene delivery systems

To generate stable transgenic cell lines, Lipofectamine 3000 reagent (Thermofisher Scientific) was used to transfect mammalian cells with the appropriate expression constructs. Lines constitutively expressing genes of interest were then established through lentiviral transduction and PiggyBac transposase systems.

2.4.1 Lentiviral transduction

A lentivirus stock was produced for each plasmid.

Day 1:

293T cells were plated on 10cm (Sterilin) petri dishes at a density of $8x10^6$, six hours before transfection to allow the cells to evenly distribute prior to transfection.

Transfection reaction mix was prepared according to standard protocol of Lipofectamine 3000, with plasmid DNA added at the following concentrations:

- Backbone Vector DNA: 10µg
- psPAX2 packaging DNA: 7.5µg
- pMD2-G envelope DNA: 2.5µg

This mixture was incubated at room temperature for twenty minutes. Finally the transfection mixture was added to 6ml of 293T media, placed on previously plated cells and cultured overnight.

Day2:

The transfection media was replaced with fresh 293T media.

Day3:

Cells to be transduced were plated onto a six well plate for each line to be created.

Day4:

The virus media was collected from 293T cells, and filtered with a 45 micron filter. 4µg/ml Polybrene was added to virus media to increase transduction efficiency of the virus, and 1ml of neat virus media was added to each well of the cells to be transduced for 6 hours, then topped up with fresh cell media at a 1:1 ratio for overnight incubation.

Day5:

After overnight transduction, virus media is replaced with fresh cell media.

Day6:

Cells are passaged prior to antibiotic selection.

2.4.2 PiggyBac lines

To establish the PiggyBac (PB) lines, cells were co-transfected with a sequence verified PiggyBac contruct containing a gene of interest, and the PBase-expressing vector (pCAGPBase) at a concentration of 1ug:3ug, using the standard Lipofectamine 3000 protocol. During transposition, the transposase recognizes transposon-specific inverted terminal repeat sequences (ITRs) located on both ends of the PB vector, cuts the coding sequence from the original vector site and efficiently integrates them into TTAA chromosomal sites. Three days after transfection, cells underwent antibiotic selection to establish lines from stable transfectants.

2.5 ANTIBIOTIC SELECTION

The following antibiotics were used in this study:

- Puromycin (Gibco):
  A concentration of 1ug/ml was used for antibiotic selection of somatic cells for a minimum of 5 days. A reduced concentration of 0.5ug/ml was utilised for ESCs and iPSCs for two separate passages of 5 day selection periods.

- Hygromycin B (Thermo Fisher scientific):
  Selection was carried out for a minimum of 10 days using a final concentration of 150ug/ml.

- Geneticin Sulphate- G418 (Gibco):
  Selection was carried out at a final concentration of 400ug/ml, and cultures could be subsequently maintained in media containing 200ug/ml.

2.6 GENERAL REPROGRAMMING METHODS

2.6.1 Neural Stem cell derived Nanog-null pre-iPS cells

Cells are maintained in basic ESC serum + LIF media on 0.1% gelatin (prepared in house). They are seeded at low density under the same serum + LIF conditions for reprogramming, and allowed to attach and proliferate for 48 hours, before culture media is switched to 2i + LIF N2B27 neurobasal serum-free media to initiate reprogramming. Antibiotic selection for complete reprogramming is applied after six days in 2i + LIF, and colonies are isolated and passaged after ten days. In place of the endogenous Nanog alleles in the N-/- NS derived pre-iPS line, one allele contains the Neo gene which confers resistance to the antibiotic G418, the other allele contains a lacZ reporter gene which allows for fully reprogrammed cells to be detected by blue staining after addition of X-Gal.

2.6.2 TNGA MEF pre-iPS cells

TNGA pre-iPS cells are seeded in basic ESC serum + LIF media on 0.1% gelatin at a low density and maintained until emergence of microscopic GFP-negative colonies. Upon visualisation of GFP-negative pre-iPS colonies, media is switched to 2i + LIF N2B27 neurobasal serum-free media to mediate the transition to ground state. Colonies are subsequently checked under LED fluorescence for activation of the GFP-PURO reporter located in place of one of the endogenous Nanog alleles (Chambers et al., 2007: sup. Figure1; sup. Figure 2). GFP-positive colonies are selected by addition of puromycin to the culture media.

2.7 FORMATION AND DIFFERENTIATION OF EMBRYOID BODIES

The ‘Hanging drop’ method was used to create embryoid bodies from pluripotent cell lines. Cells are detached with 0.1% Trypsin-EDTA at 37°C, then collected and centrifuged in basic media. Cells are resuspended, and counted by haemocytometer, and an appropriate volume of the cell suspension calculated to contain 250K cells is then added to 10ml fresh media. An eight-channel multi-pipette was used to dispense lines of 40ul droplets onto the upturned lid of a petri dish. Each 40ul droplet from the 10ml suspension therefore contains 1000 cells. The lid was carefully replaced onto the base of the dish, which contained 1X PBS to maintain humidity. Embryoid bodies (EBs) formed over three days, before collection by p1000 pipette. The EBs were then maintained in suspension in basic media and allowed to spontaneously differentiate for up to ten days.

2.8 AP STAINING

The Alkaline Phosphatase kit from SIGMA was used to stain the cells as per the manufacturer's manual. Cells were fixed in a citrate-acetone-formaldehyde solution at room temperature for 30 seconds. They were then rinsed in deionised water. An alkaline-dye mix of sodium nitrite and FRV-Alkaline at a 1:1 ratio in solution was added to the cells, and they were incubated at room temperature for 15 minutes, protected from light. Cells were then counterstained for two minutes in hematoxylin solution, before being thoroughly rinsed in tap water, and air dried.

2.9 LACZ STAINING

The following buffers were used to stain the NS derived cell lines for activation of the LacZ gene.

- Fix buffer:
  Added to 1L of 0.1M PBS (pH 7.3) were: 50ml 0.1M EGTA (5mM [final]), 2ml 1M MgCl2 (2mM [final]), 2.86ml 70% glutaraldehyde (0.2% [final]).

- Wash buffer:
  2ml 1M mgCl2 (2mM [final]) was added to 1L of 0.1M PBS (pH 7.3).

- Staining buffer:
  Added to 1L of wash buffer were: 2.1g potassium ferrocyanide (5 mM [final]), 1.65g potassium ferricyanide (5mM [final]).

1 mg/ml X-gal (5-bromo-4-chloro-3-indolyl-β-D-galactosidase) was added fresh before staining.

Cells were fixed for 15 minutes. They were then washed twice, each wash was conducted for 5 minutes. The cells were then incubated in staining buffer for 12–16 hours at 37°C. Staining buffer was then removed, cells were washed twice, then maintained in 1X PBS at room temperature for storage and analysis.

2.10 ANALYSIS OF GENE EXPRESSION

2.10.1 RNA extraction

To extract RNA from cells in culture, each well of a six well plate to be processed was first washed with sterile PBS, and 600ul of TRI Reagent (SIGMA) was added directly into the well. The TRI Reagent was evenly dispersed, and allowed to stand for a minimum of 2 minutes. The suspension was collected to a sterile 15ml falcon tube, and either stored at -20^{o}C, or processed immediately. 200ul of chloroform (SIGMA) was added to the TRI Reagent suspension, and mixed by vortex until milky pink. The mixture was then transferred to a sterile 1.5ml Eppendorf tube and left to stand for 10 minutes at room temperature. The sample was centrifuged at 14,000rpm for 30 minutes at 4^{o}C to achieve phase separation. The upper aqueous layer was extracted and moved to a clean 1.5ml Eppendorf, and an equal volume plus 50ul of 100% ethanol was added, and the mixture was vortexed. The sample was then DNase treated (AMBION) and column purified as per the guide of the QIAGEN RNeasy Mini kit. RNA integrity was checked by running on an agarose gel.

2.10.2 cDNA synthesis

First-Strand cDNA synthesis was used to prepare samples for qPCR analysis. A reaction consisting of 1ul of Oligo$(dT)_{20}$ (50uM), 2ug of total RNA, and 1ul of 10mM dNTP mix, was topped up to 13ul using sterile distilled water. The reaction was heated to 65^{o}C for 5 minutes, and then incubated at 4^{o}C for 5 minutes. The tube was briefly centrifuged before the addition of 4ul 5X First-Strand Buffer, 1ul 0.1M DTT, 1ul RNaseOUT recombinant RNase inhibitor (40 units/ul), and 1ul SuperScript III RT (200 units/ul). All reagents from Thermo Fisher Scientific. The reaction mix was then heated to 25^{o}C for 5 minutes, 50^{o}C for 60 minutes, and 70^{o}C for 15 minutes, before being chilled to 4^{o}C. cDNA samples could then be analysed by qPCR.

2.10.3 qPCR analysis

Each qPCR reaction contained 5ul of SYBR-Green Jumpstart Taq readyMix (SIGMA), 1ul (10uM [final]) of each of the forward and reverse primers (SIGMA), 2µl of nuclease-free water, and 1µl of cDNA template. Each reaction was prepared in triplicate on an ABI FAST Systems 0.2ml 96-well PCR plate (STARLAB), and then sealed with an Optical Adhesive cover (Life Technologies).

The following qPCR conditions were utilized for each run on the QuantStudio 6 Flex (Life Technologies) qPCR instrument:
The plate was heated to 105^{o}C to activate the Jumpstart Taq before being held at 50^{o}C for 2 minutes. The initial denaturation step was at 94°C for 10 minutes, followed by 40 cycles of 94 °C for 15 seconds (denaturation) and 60 °C for
1 minute (annealing and extension).
The raw data was then extracted to be analysed by comparative CT (Cycle threshold). An endogenous control gene was chosen for each qPCR run to normalise the data in order to compare relative fold change of target genes amongst cDNA samples.
The double delta CT value (**ΔΔCT**) was calculated using the following formula:

ΔCT target- ΔCT reference= **ΔΔCT**

As all calculations are in logarithm base 2, the expression fold change of each target gene was calculated using **2^-ΔΔCt**.

CHAPTER THREE

3 RESULTS: SECTION ONE

3.1 REPROGRAMMING NANOG-NULL NS DERIVED PRE-IPS CELLS

3.1.1 Lentiviral cell lines in serum + Lif conditions

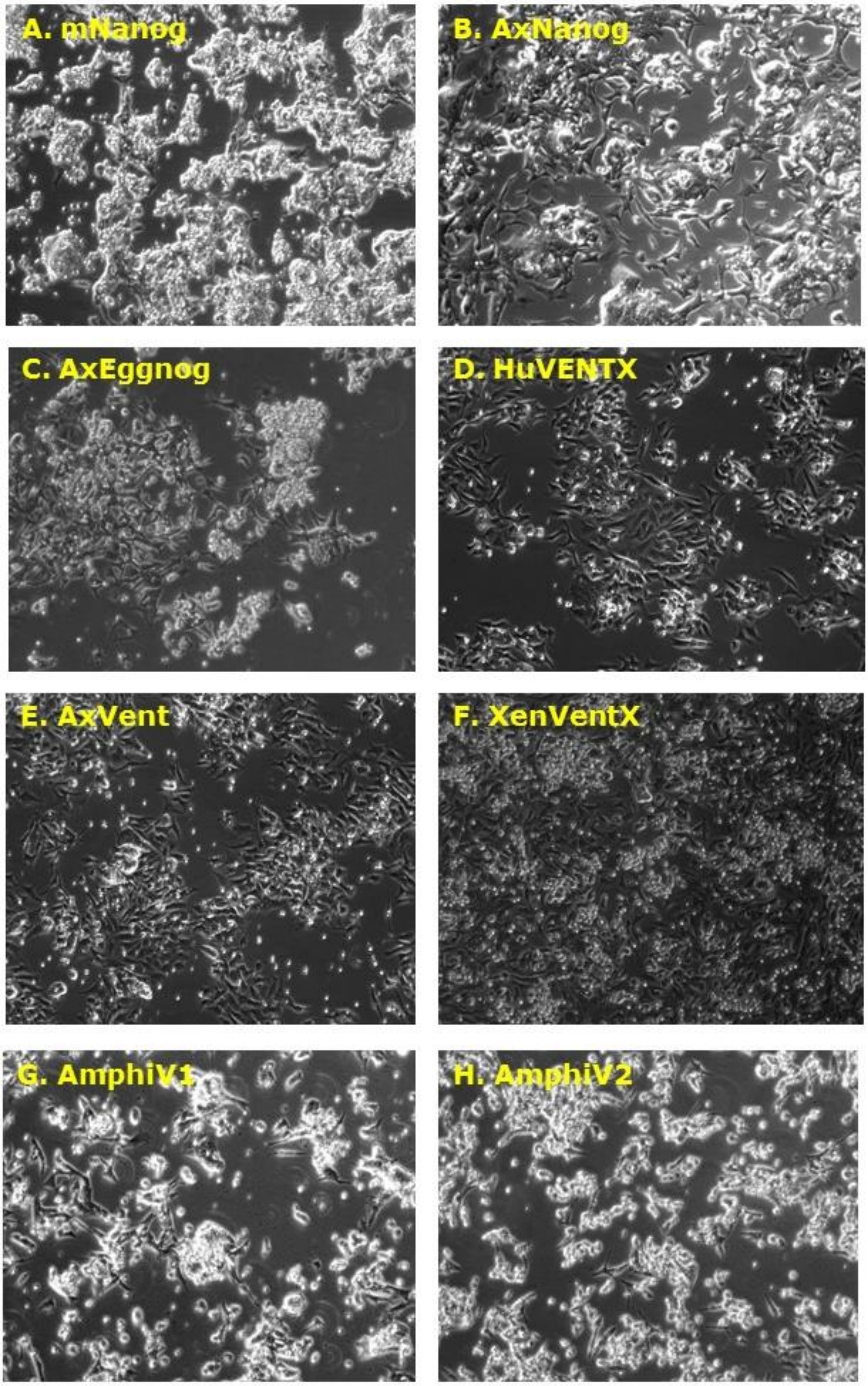
I tested the hypothesis that the invertebrate Vent genes may have Nanog activity, as they appear to be phylogenetically related (Fig. 3). I began by testing whether Vent genes could substitute for Nanog in reprogramming the Nanog-null neural stem cell derived pre-iPS cells (N-/- NS pre-iPSCs) to ground state pluripotency. Silva et al. (2009) demonstrated that Nanog null pre-iPSCs do not reprogram to pluripotency absent a functioning Nanog gene. To do this, I first established pre-iPSC lines constitutively expressing the candidate genes. The N-/- NS pre-iPSCs were tranduced with a different lentiviral media for each gene, made using the pSIN expression plasmid. The mouse Nanog was chosen as a positive control as it has been previously shown to efficiently reprogram these NS pre-iPSCs (Silva et al. 2009). Both axolotl Nanog genes were also tested for comparison. Pre-iPS cell lines expressing Xenopus laevis Vent2, axolotl Vent, and human VENTX were established by lentiviral transduction to test the reprogramming capacity of vertebrate Vents. Both Vent1 and Vent2 genes found in the invertebrates acorn worm Saccoglossus kowalevskii and amphioxus Branchiostoma floridae were tested, along with the single Vent gene found in each of the sea anemone Nematostella vectensis, and the freshwater sponge Ephydatia fluviatilis (Table 1).

Candidate Genes	
Mouse Nanog	Amphioxus Vent1
Axolotl Nanog	Amphioxus Vent2
Axolotl Eggnog	Saccoglossus Vent1
Human Vent	Saccoglossus Vent2
Axolotl Vent	Nematostella Vent
Xenopus Vent	Sponge Vent

Table 1: Candadate genes tested.

An empty pSIN plasmid was transfected as a negative control, and all cell lines were tested side by side with the untranduced pre-iPS cells. As the tryptophan repeat (WR) domain of mammalian Nanogs allows for the homodimerization that drives ESC self-renewal (Mullin et al., 2008; Wang et al., 2008), it was hypothesised that the addition of a WR domain to the carboxy end of Vent genes and orthologs of Nanog from lower vertebrates, would provide the best chance of establishing an iPS line that could successfully self-renew once reprogrammed. The pre-iPS cells were grown and transduced in serum + LIF conditions on gelatin treated six-well plates. The lines were maintained stably in culture after antibiotic selection for the transgene. The concentration of the antibiotic puromycin was sequentially increased from 0.5ug/ml to 1ug/ml to ensure selection of the cells with the highest level of the transgene, for more efficient reprogramming. Concentrations above 1ug/ml compromised the integrity of the cell line.

Transduction with each of the WR-containing constructs resulted in an accelerated proliferation rate, a characteristic specific to Nanog homodimers that was previously demonstrated by Dixon et al. (2010). Mouse Nanog (Fig. 7A), axolotl Nanog (Fig. 7B), and Eggnog (Fig. 7C) all displayed rapid growth. The cell lines began to demonstrate differential phenotypes in the serum + LIF conditions (See Fig. 7). The lines expressing Nanog genes (Fig. 7A, B, C) began to exhibit a smaller, rounded morphology associated with more intermediate, partially-reprogrammed cells, and the pre-iPS colony-like structures which exist in the NS pre-iPS parent cell line (Fig. 7N) became more prominent. Cells transfected with the Nematostella Vent gene exhibited a rounded morphology in the serum + LIF conditions (Fig. 7K), which appeared to be representative of more partially reprogrammed cells. The other invertebrate Vent lines demonstrated this morphology to a lesser extent (Fig. 7G, H, I, J, L). However, the vertebrate Vent lines more closely resembled the negative controls, and demonstrated no change in morphology (Fig. 7M).


A. mNanog
B. AxNanog
C. AxEggnog
D. HuVENTX
E. AxVent
F. XenVentX
G. AmphiV1
H. AmphiV2

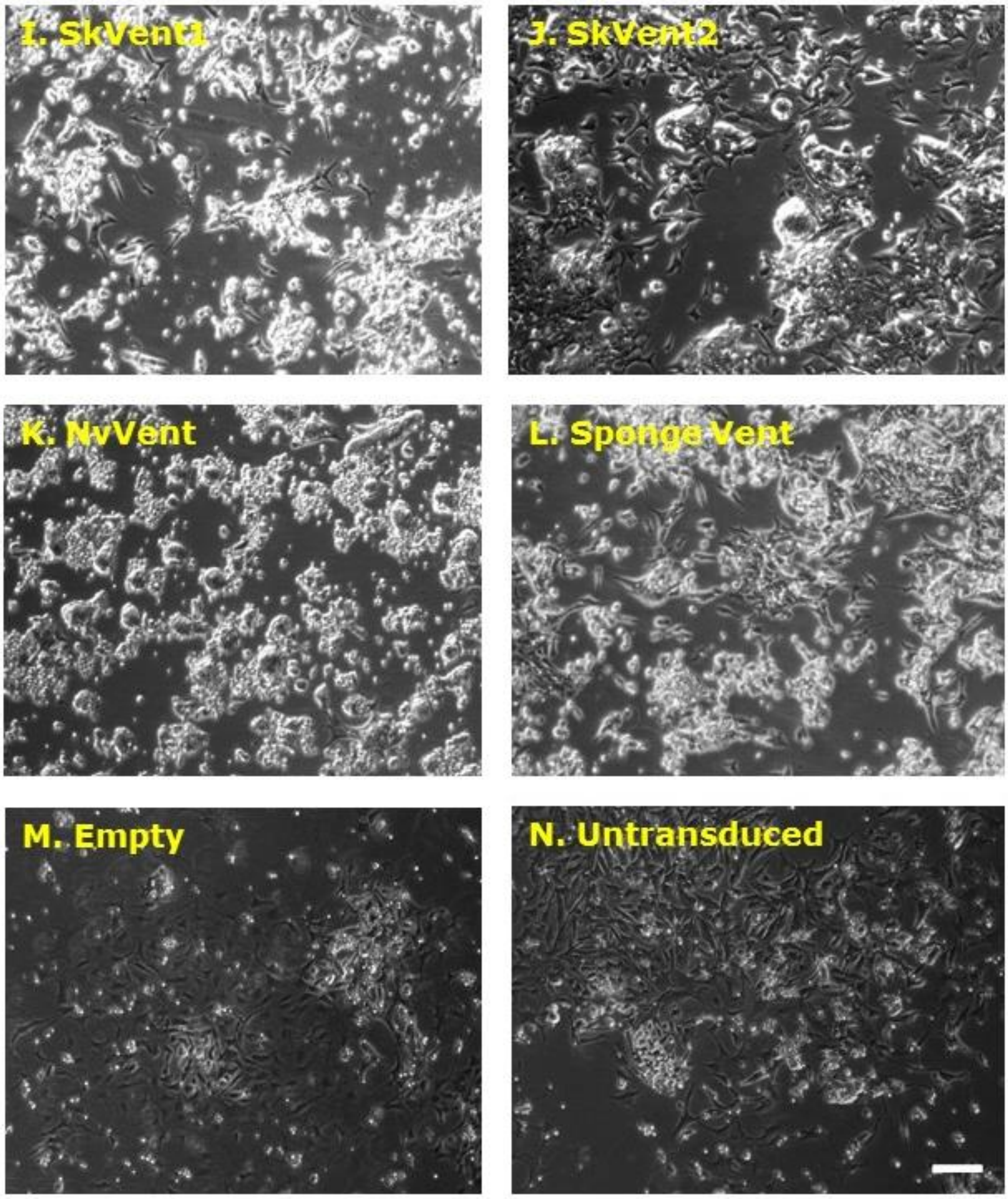


Figure 7: NS derived pre-iPS lines in Serum + LIF expressing candidate genes with a WR domain three passages after transduction.
(A) Mouse Nanog, (B) Axolotl Nanog, (C) Eggnog, (D) Human Vent, (E) Axolotl Vent, (F) Xenopus Vent2, (G) Amphioxus Vent1, (H) Amphioxus Vent2, (I) Saccoglossus Vent1, (J) Saccoglossus Vent2, (K) Nematostella Vent, (L) Sponge Vent, (M) Empty plasmid, (N) Untransduced cells. Scale bar represents 200um at 10x magnification.

In line with the established method for reprogramming the N-/- NS pre-iPS cells; the cells were initially seeded in serum + LIF conditions for two days, prior to transfer to a 2i + LIF culture. The initial seeding density of the WR-containing lines was 20K cells per well of a six-well plate; however this number was increased to 30k/well for the empty plasmid control and the untransduced cells. This was to accommodate the lower proliferation rate of these cells in the absence of the WR domain, ensuring a comparable confluency between cell lines at the point the reprogramming assay was initiated by switching to the 2i + LIF media.

3.1.2 2i + LIF reprogramming

The media was switched to 2i + LIF N2B27 serum free conditions two days after seeding, to initiate reprogramming. The switch to 2i + LIF media represents T0 of the reprogramming assay. By day three in 2i + LIF, small round iPS-like colonies were visible in the Nanog cultures (Fig. 8). The mouse Nanog colonies already displaying a defined edge associated with emerging iPSC colonies (Fig. 8A). Both axolotl Nanog and Eggnog cultures were also beginning to show early colony formation at a visibly lower frequency than the mouse Nanog cultures, but with large partially reprogrammed colonies (Fig. 8C, D).

At this point the Nematostella Vent cultures contained very few cells of the more elongated somatic morphology, the majority of cells instead showing a partially reprogrammed phenotype of more rounded cells, and appeared to be reprogramming at an efficient level (See Fig. 8B). There were no iPSC-like colonies that could be identified in the other Vent lines at day three.

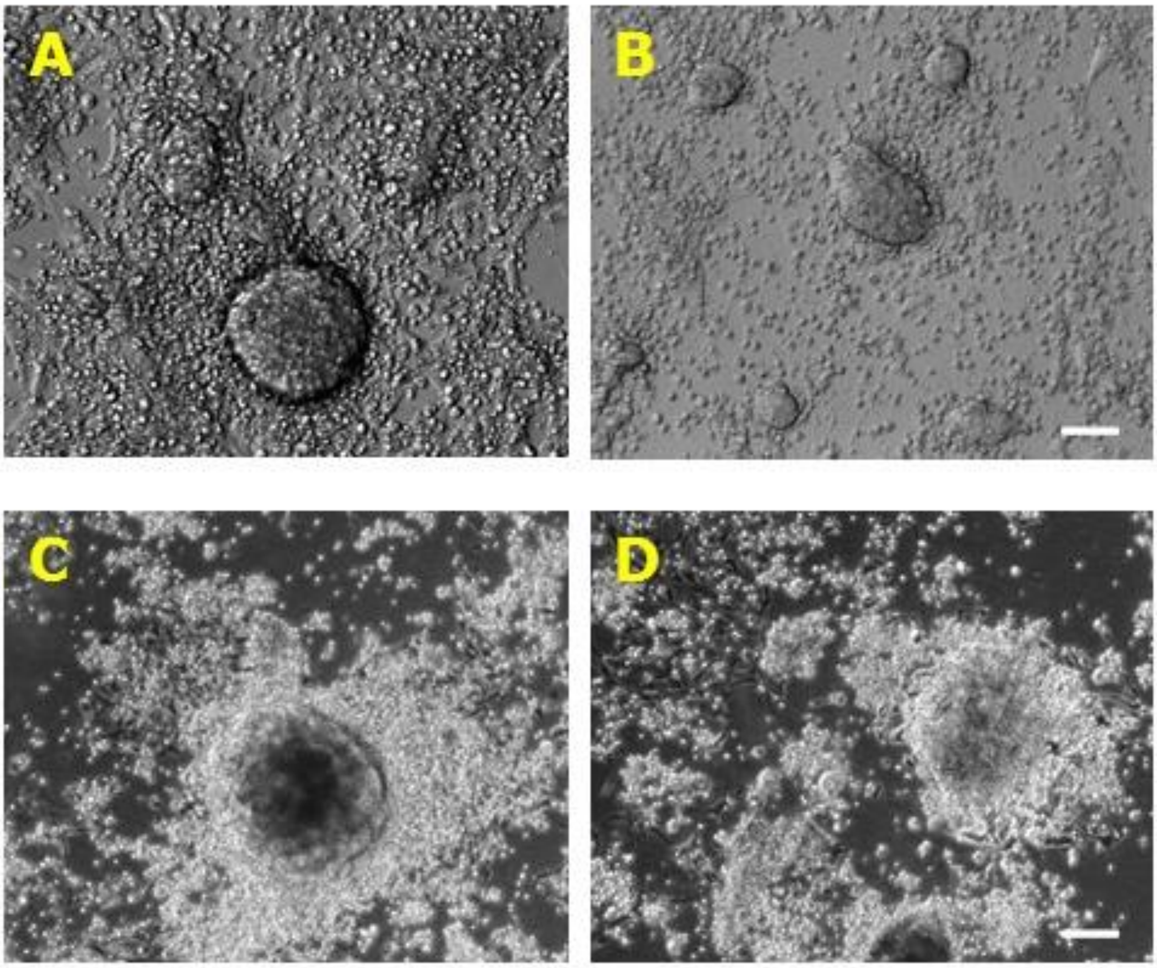
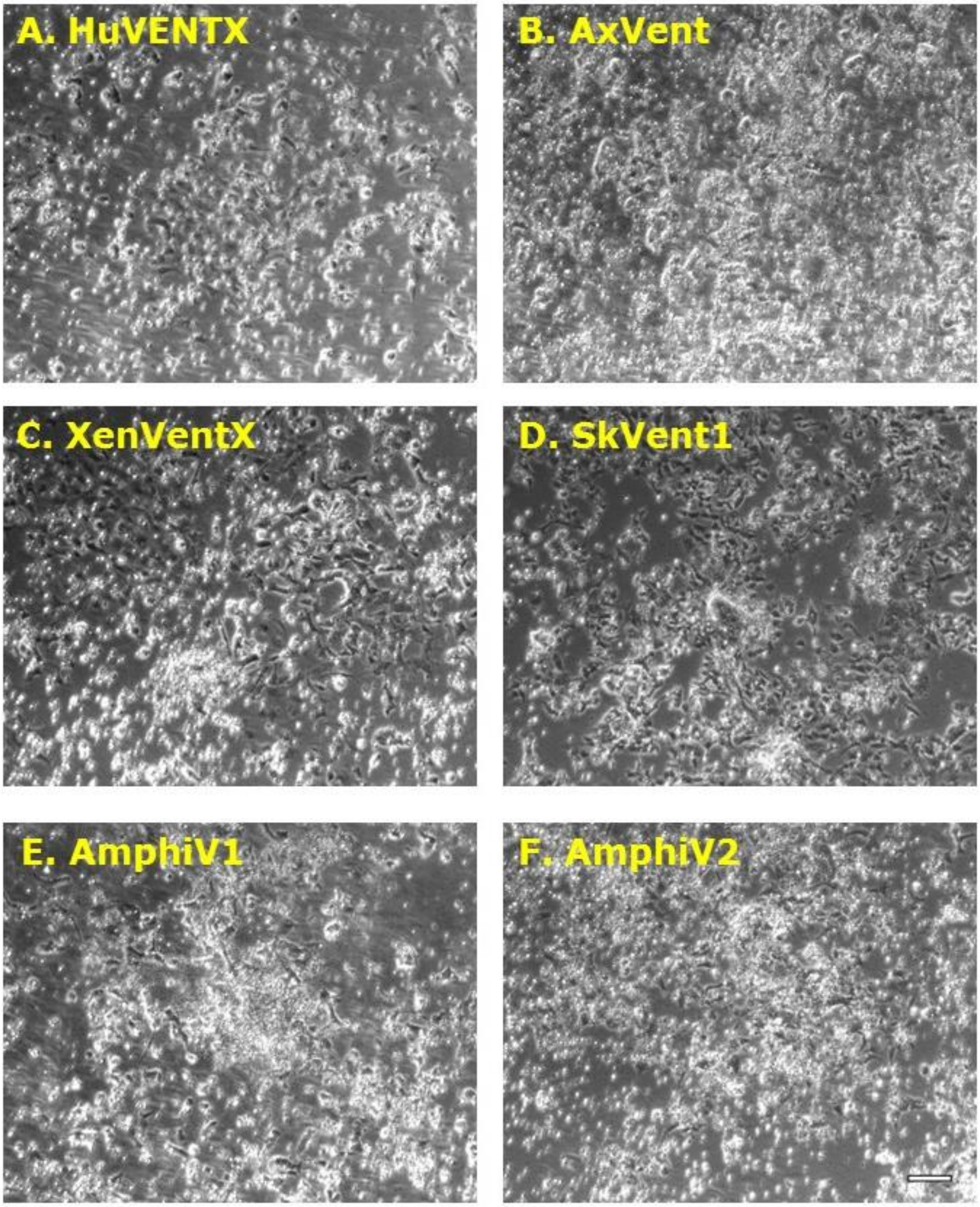


Figure 8: Colony formation at day three of reprogramming.
Both the (A) mouse Nanog and (B) Nematostella Vent cultures predominantly contained partially reprogrammed cells by day three, and microscopic iPS-like colonies with defined borders. Both axolotl Nanog (C) and Eggnog (D) showed large pre-iPS colony formation. Scale bar for images (A) and (B) represents 100um. Scale bar for images (C) and (D) represents 200um.

All of the Vent lines exhibited large amounts of cell death in the 2i + LIF conditions by day five (Fig. 9). The cells exhibiting a round phenotype representative of a more partially reprogrammed intermediate state, as seen in wells containing cells expressing human VENT (Fig. 9A), amphioxus Vent1 (Fig. 9E), and Sponge Vent (Fig. 9G), struggled to propagate in the 2i + LIF conditions, and began to die out.


A. HuVENTX
B. AxVent
C. XenVentX
D. SkVent1
E. AmphiV1
F. AmphiV2

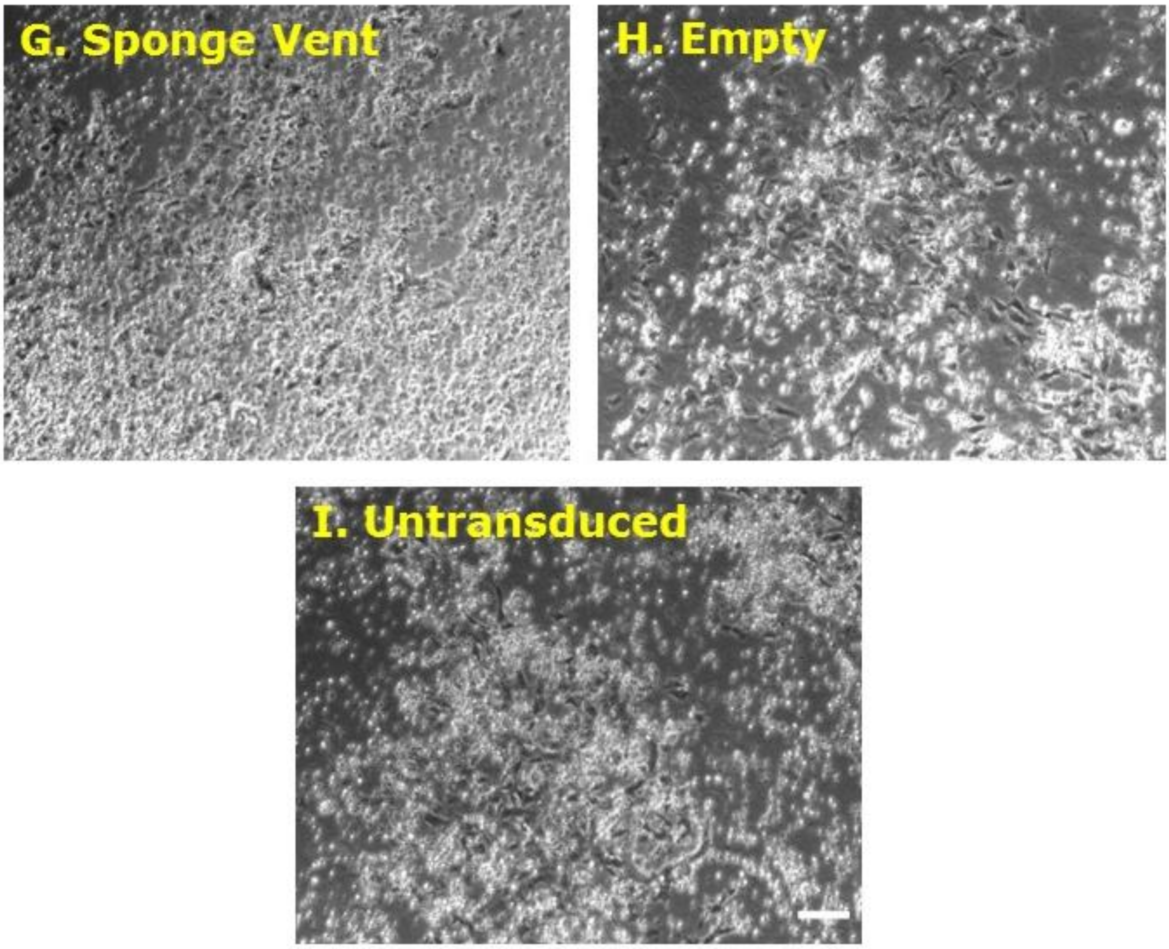


Figure 9: Partially reprogrammed pre-iPS Vent cultures struggling in 2i + LIF
Both the Vertebrate and Invertebrate Vent cultures showed large amounts of cell death by day 5 of reprogramming. Pre-iPS cells exhibiting a more intermediate state with a round partially reprogrammed phenotype struggled with prolonged exposure to 2i + LIF conditions.
(A) Human Vent, (B) Axolotl Vent, (C) Xenopus Vent2, (D) Saccoglossus Vent1,
(E) Amphioxus Vent1, (F) Amphioxus Vent2, (G) Sponge Vent, (H) Empty plasmid,
(I) Untransduced cells. Scale bar represents 200um at 10x magnification.

However, unlike these Vent cell lines which were showing signs of gradually dying out in 2i + LIF, the Nematostella Vent cultures looked to be rapidly dying (Fig. 10A), with very few cells remaining attached to the gelatinised surface of the plate. Only the Nematostella Vent cultures entirely failed to propagate beyond this stage.

At the same time point, day five, microscopic colonies could be identified in the Saccoglossus Vent2 cultures as rounded and compact, 3D structures with a defined edge (Fig. 10B) comparable to the colonies in the cultures expressing Nanog genes, such as axolotl Nanog (Fig. 10

C). No other Vent line demonstrated any signs of reprogramming activity at this point.

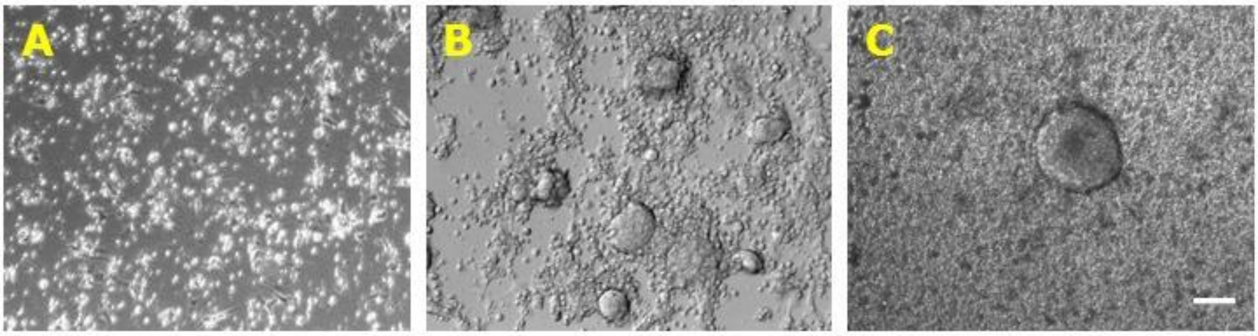


Figure 10: Saccoglossus Vent2 colonies emerge at day five as Nematostella Vent collapses
(A) Only the Nematostella Vent cultures completely died out in 2i + LIF conditions prior to antibiotic selection. (B) Microscopic colonies emerged in the Saccoglossus Vent2 cultures around day five.
(C) Axolotl Nanog iPSCs emerging.
Scale bar represents 200um

G418 antibiotic was added to the 2i + LIF media on day 6 at a concentration of 400ug/ml to select for expression of the Neo resistance allele embedded in the endogenous Nanog locus. By selecting for activation of the Nanog locus, I was able to isolate the cells which were fully reprogrammed to ground state pluripotency. Cultures were maintained in G418 for four days until assayed at the assigned end point of day ten.

3.1.3 iPS cells identified at day ten

By day ten of the reprogramming assay, the G418 had selected for the cells in which the Neo resistance gene had been activated. The iPS colonies could be more clearly identified by day ten (Fig. 11), as the majority of the cells which failed to activate endogenous Nanog locus died during selection.
All three Nanog genes, mouse Nanog, axolotl Nanog, and Eggnog, successfully reprogrammed the cells to form G418 resistant colonies. The mouse Nanog cultures, based on observation (Fig. 11A), appeared to have worked the most efficiently, forming large compact colonies comparable to naïve ESCs. Both of the axolotl Nanog genes had produced large round colonies with a defined edge (Fig. 11; B, C). Interestingly, the Saccoglossus Vent2 gene also successfully generated colonies; however, they were generally much smaller than those generated by the Nanog

cultures (Fig. 11; D), apart from a few colonies which expanded to a size comparable to the Nanog colonies. But all of the Saccoglossus Vent2 colonies displayed the round morphology and defined borders characteristic of emerging iPS colonies (See arrow in Fig. 11D).

Many of the colony-like structures that appeared in the Saccoglossus Vent2 cultures by mid-point of the reprogramming assay did not survive the G418 selection, and therefore appear to have failed to convert to a fully reprogrammed state. The other invertebrate Vent lines expressing Saccoglossus Vent1, both amphioxus Vents, and Sponge Vent failed to generate anything that resembled an iPS colony, and they died out completely in the antibiotic by day ten of the assay. The Nematostella vent line died prior to any selection. The three vertebrate Vent gene lines also failed to produce reprogrammed cells (Fig. 11E). Cultures of cells either transduced with the empty plasmid or negative controls containing cells that were not transduced died under selection by the end of reprogramming (Fig. 11F). By day ten of the assay, all Vent cultures apart from the Saccoglossus Vent2 line, and the negative controls, failed to produce G418 resistant colonies, and ceased to propagate in culture under antibiotic selection (Fig. 11E, F).

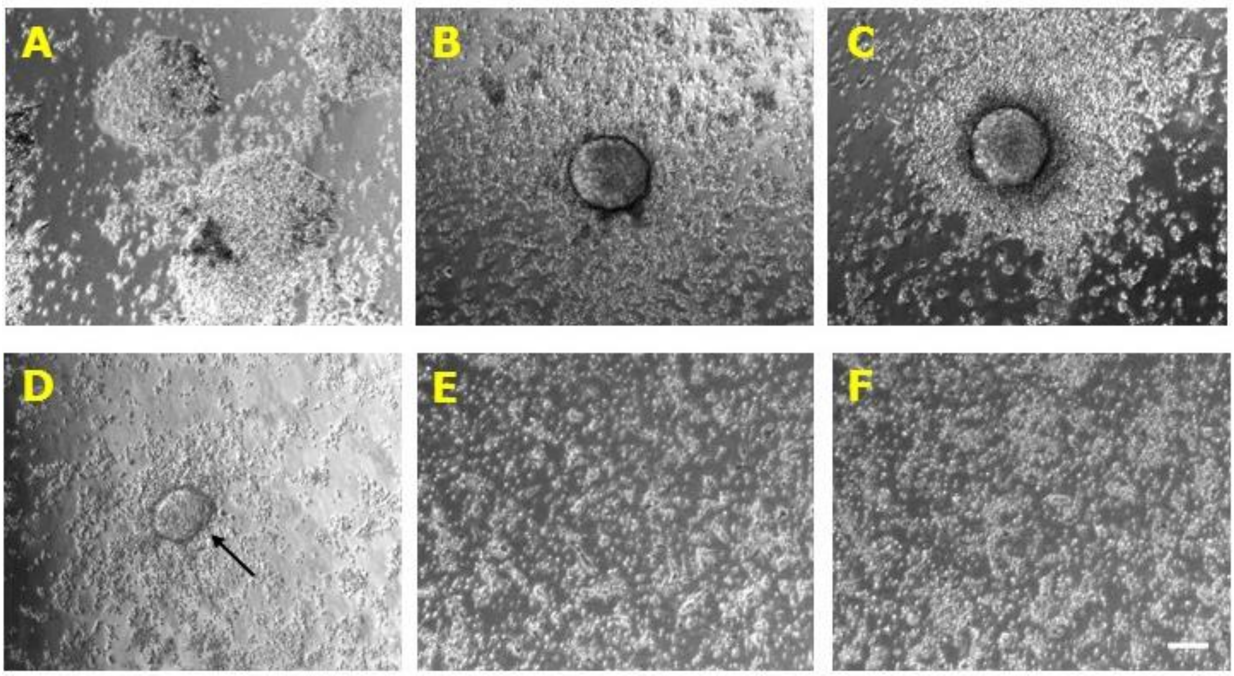


Figure 11: Day 10 of reprogramming.
(A) Large compact ESC-like colonies in mouse Nanog cultures. (B) G418 resistant colonies in Axolotl Nanog and (C) Eggnog (D) Smaller colonies in Saccoglossus Vent2 cultures. (E) Image of the human vent culture, which is representative of all the vertebrate Vent cultures at

this point of the assay. (F) Empty plasmid line did not survive the G418 selection. Scale bar represent 200um.

The cultures which had successfully generated colonies were analysed to determine if complete reprogramming was achieved at day ten. Plates containing each of the reprogrammed cell lines were passaged onto new six well plates with a gelatinised surface, and maintained in 2i + LIF N2B27 media conditions to establish lines of iPSCs. To passage the Saccoglossus Vent2 colonies which were sparse on the reprogramming plates, larger colonies were manually picked using a p200 pipette in order to limit enzymatic disruption with Trypsin, before the plates were then exposed to 1X Trypsin-EDTA solution to dislodge and collect the remaining cells, and then passaged as standard. The G418 antibiotic selection is a slow process, and by day ten of the reprogramming assay it was found that a few relatively intact cells with a somatic morphology were still attached to the plate, most noticeably in the Xenopus Vent2 cultures where no iPS colonies had formed. This is most likely due to the high confluency of the cultures at the onset of antibiotic selection, resulting from the accelerated proliferation rate attributed to the WR domain. However, these cells were also collected by exposure to Trypisn-EDTA, and passaged to fresh plates in 2i + LIF to rule out any possibility of Neo resistance having been activated within these cells. Very few cells from the Xenopus Vent2 cultures survived passaging, but those that attached were maintained in 2i + LIF conditions for ten days in the absence of antibiotic, and the cells grew from a low density to form large pre-iPS structures (Fig. 12A, B). G418 was then added to the media at the maintenance concentration of 200ug/ml, and over the course of one week, proceeded to eradicate all cells from the plate (Fig. 12C). This indicates that no cells were reprogrammed by Xenopus Vent2, and the cells which survived the initial antibiotic selection did so due to insufficient exposure to G418 selection.

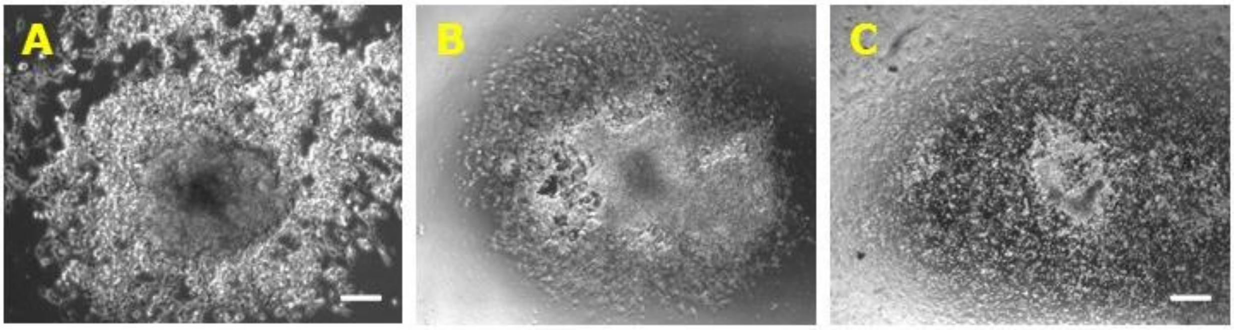


Figure 12: Extensive G418 treatment required for Xenopus Vent2-WR cultures
(A) Xenopus Vent2 cultures grew to form large pre-iPS structures in 2i + LIF conditions over the course of 10 days. Scale bar represents 200um. (B) A low number of cells successfully attached to the plate after passaging, and grew to form tightly packed clusters from sparely seeded cells. (C) G418 eliminated all cells from the culture after 7 days. Scale bar for images (B) and (C) represents 500um.

3.1.4 Staining of iPS colonies

When the reprogramming reached the assigned end point, day ten in 2i + LIF as per the methods of Theunissen et al. (2011), cultures which had generated G418 resistant colonies were stained for activation of the LacZ gene that was knocked-into one of the endogenous Nanog alleles. Blue staining was indicative of activation of the endogenous Nanog locus which indicates complete reprogramming to ground state potential. Cultures containing cells expressing Mouse Nanog, axolotl Nanog, Eggnog, and the Saccoglossus Vent2 all stained positive for LacZ (Figure 13).

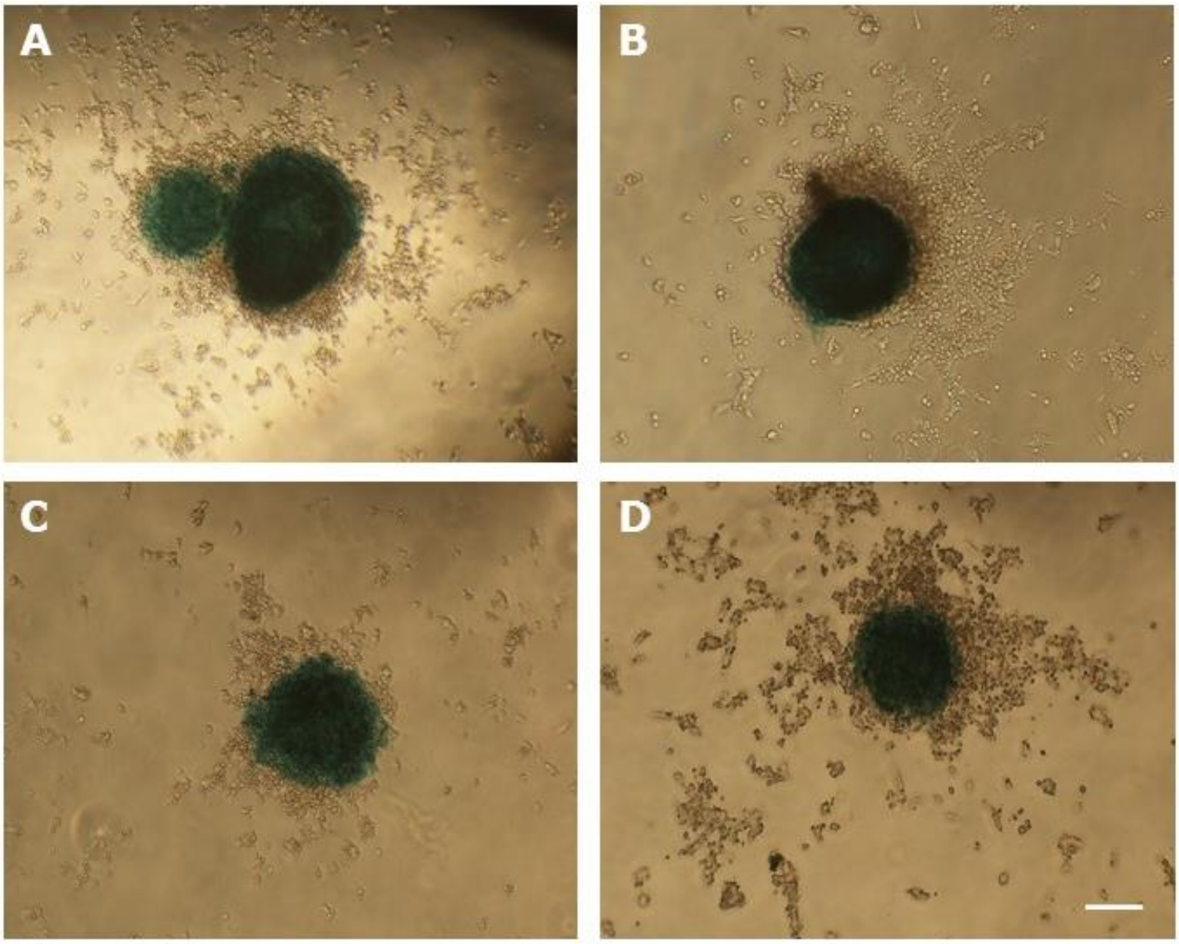


Figure 13: LacZ staining at day 10 of reprogramming
(A) Mouse Nanog, (B) Axolotl Nanog, (C) Eggnog, and (D) Saccoglossus Vent2, all stained positive for activation of the LacZ gene at the endogenous Nanog locus.
Scale bar represents 200um.

As a second test for reprogramming, cultures were also stained for expression of alkaline phosphatase (AP), which is present at elevated levels on the membrane of pluripotent cells, and is therefore used as a marker for pluripotency. Cultures of cells expressing mouse Nanog contained high numbers of AP positive colonies (Fig. 14A), demonstrating the high reprogramming efficiency of the gene. Cultures containing colonies generated by Axolotl Nanog (Fig. 14B), Eggnog (Fig. 14C), and Saccoglossus Vent2 (Fig. 14D) all stained positive for AP.

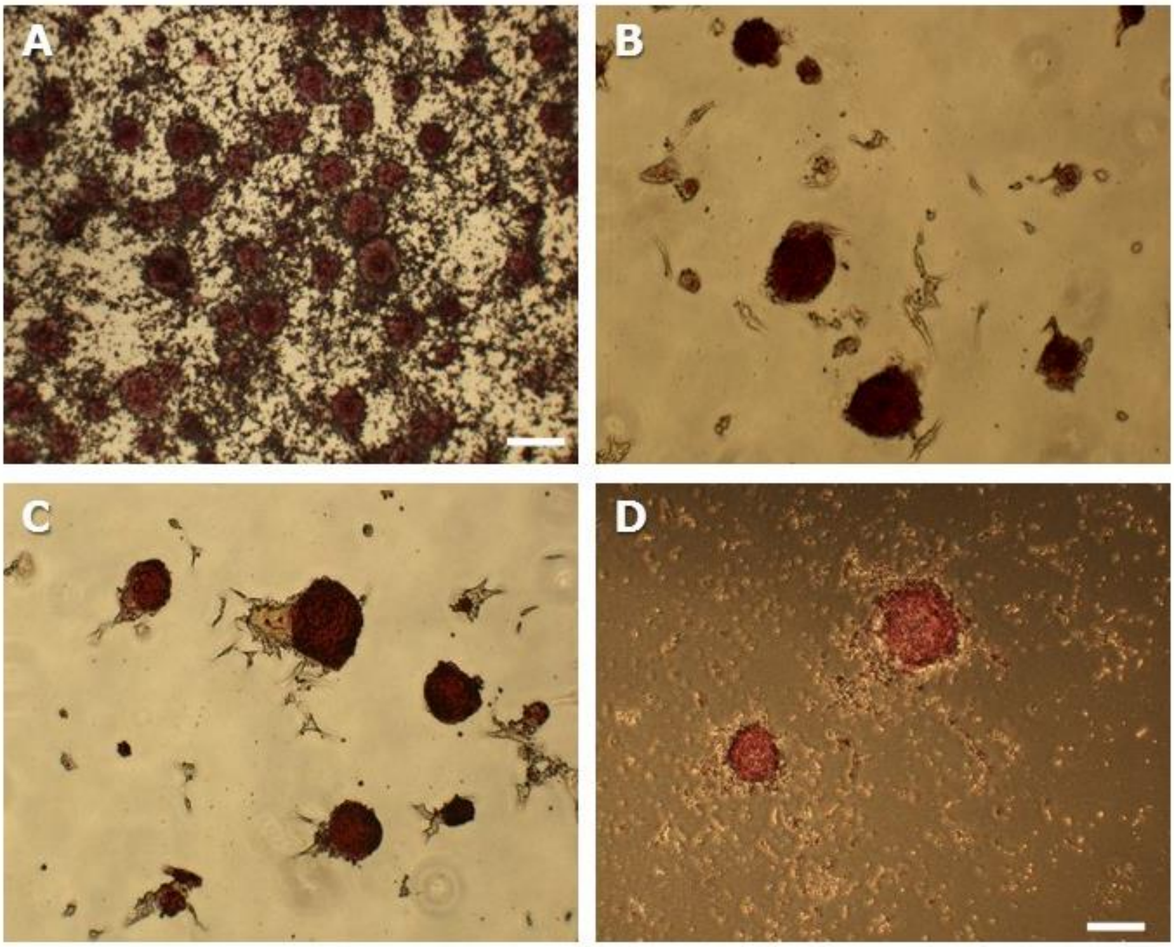


Figure 14: Alkaline phosphatase staining for pluripotent cells
Red cells indicate positive AP staining. (A) A dense population of AP positive mouse Nanog reprogrammed colonies, shown here at x40 magnification. Scale bar represents 500um. (B) Axolotl Nanog, (C) Eggnog, and (D) Saccoglossus Vent2, all stained positive for AP. Scale bar represents 200um.

Both axolotl Nanog (Fig. 14B) and Eggnog (Fig. 14C) appeared to have generated AP positive colonies to a similar extent, but were both observed to be sparser than the mouse Nanog cultures (Fig. 14A). Based on observation, Saccoglossus Vent2 cultures had generated very few AP (Fig. 14D), and LacZ (Fig. 13D), positively stained colonies. However, positive staining in the Saccoglossus Vent2 cultures suggests that the Nanog activity required to reprogram the NS pre-iPS cells to pluripotency is conserved in Hemichordates.

3.1.5 Efficiency of reprogramming

Colonies which stained positive for both LacZ and AP were counted at the end of the reprogramming assay, to assess the reprogramming efficiency by comparing the number of colonies generated per 10^4 cells seeded for each cell line.

Cell Line	LacZ+			AP+			Total Colonies	Average per well	Colonies per 10^4 seeded
Mouse Nanog	173	193	139	175	191	181	1052	175.3	**87.6**
Axolotl Nanog	35	43	36	48	41	38	241	40.17	**20.85**
Eggnog	27	32	28	31	29	33	180	30	**15**
Saccoglossus Vent2	8	7	12	9	8	8	52	8.67	**4.34**

Table 2: Reprogramming efficiency of positively stained cell lines at day 10
Three reprogramming cultures were established for each cell line on a six well plate. For each plate one well was stained for LacZ, another for AP. Positive staining colonies were counted for each well. The total number of colonies generated for each gene was calculated as the sum of both LacZ and AP positive colonies. The final efficiency of colonies generated per 10^4 cells seeded, was based on the initial seeding density of 20K cells per well for each line.

Mouse Nanog was the most efficient gene for generating iPSC colonies from the somatic pre-iPS cells (Table 2). Both axolotl Nanog genes demonstrated a similar efficiency which was however, much lower than the mouse gene. The Saccoglossus Vent2 gene was approximately 5% as efficient as Mouse Nanog in generating colonies. However, activation of the endogenous Nanog locus (Fig. 13), and subsequent generation of iPS colonies indicate that the Saccoglossus Vent2 gene can successfully substitute for Nanog in reprogramming, and that hemichordate Vent genes have the capacity reprogram cells to pluripotency. To further characterise the activity of Saccoglossus Vent2, the iPS line was expanded and analysed. As the mouse Nanog gene was the most efficient at reprogramming, and is itself the native transcription factor that was deleted from the parental line, it was chosen as the positive control for experiments here after. Both iPS lines were expanded side-by-side for comparative purposes.

3.1.6 Saccoglossus iPS cells in culture

The iPSCs reprogrammed with Saccoglossus Vent2 were continuously maintained on gelatinised plates in 2i + LIF N2B27 serum-free conditions. The cells proliferated at a rate above that which would be normal for wild type ESCs, presumably due to the expression of the WR domain in the lentiviral contruct. The line expressing mouse Nanog showed very limited differentiation in 2i + LIF conditions, and maintained homogeneity within the culture for several passages at a time absent G418 (Fig. 15A). In contrast, the Saccoglossus cells had a noticeable propensity to differentiate. Relative to the mouse Nanog line, the Saccoglossus Vent2 cells continuously displayed disproportionately high levels of differentiation on the periphery of the iPS colonies (Fig 15B). Interconnecting webs of differentiated cells (Indicated by yellow arrows in Fig. 15C) formed across the culture in the absence of G418. The central colonies themselves otherwise exhibited a compact 3D structure characteristic of ground state pluripotent colonies. Yet to retain homogeneity of the cultures, they were periodically maintained in 200ug/ml G418 antibiotic to remove differentiation (Fig. 15D). The central colonies remained intact, yet prolonged absence of any antibiotic selection for pluripotent cells lead to the formation of interconnecting webs of differentiated cells, present universally throughout the culture. The mouse Nanog line, in comparison, behaved more akin to wild type ESCs in 2i + LIF.

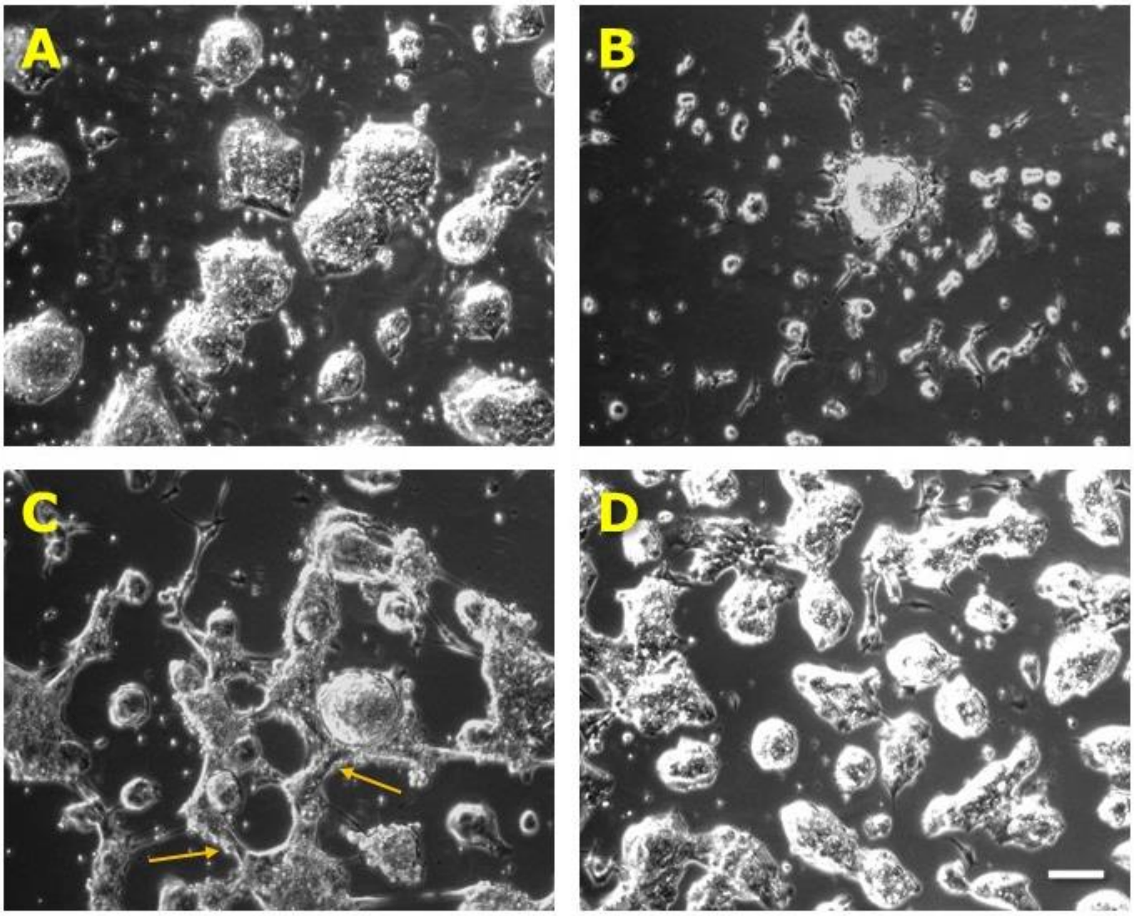


Figure 15: Passaged iPS lines growing in 2i + LIF
(A) Mouse Nanog iPS cells maintained a homogeneous ESC-like morphology in 2i + LIF conditions. (B) Saccoglossus Vent2 cells had a propensity to differentiate. (C) Saccoglossus Vent2 colonies cultures absent G418. (D) iPS colony homogeneity restored with G418. Scale bar represents 200um.

Both iPS lines were expanded further. As this reprogramming experiment was conducted with the mammalian WR domain attached to the carboxy terminus of the Vent genes, we decided to test if the Saccoglossus Vent2 gene could reprogram the N-/- somatic pre-iPS cells acting as a monomer, and successfully establish an iPS line in its native state.

3.1.7 Monomer cell lines in serum + LIF

To test the reprogramming capacity of the invertebrate Vent genes further, we used the PiggyBac vector system to stably transfect N-/- pre-iPS cells with Saccoglossus

Vent2 and Nematostella Vent. These constructs lacked the WR domain. The mouse Nanog gene, was used as a positive control. However, the mouse Nanog CDS naturally encodes the WR domain, and was therefore capable of homodimer formation. As the Xenopus Vent2 gene had been the most extensively tested in the previous assay, and was conclusively shown to have no reprogramming activity, it was chosen as a negative control for vertebrate Vent activity in this reprogramming assay, and was expressed in its natural monomeric state.
The PiggyBac vector system was chosen for several reasons. Unlike lentiviral vectors, the PiggyBac vector is not silenced by epigenetic reprogramming in pluripotent cells (Chen et al., 2015), and therefore its expression is maintained at a constant level in an established iPS line. In addition, lentiviral vectors randomly insert into the genome, whereas the PiggyBac vectors recognise a specific sequence for insertion, which provides a greater level of control over the copy number delivered. PiggyBac is also reversible, the transgene inserted can be excised from established cell lines.

As previously, the N-/- pre-iPS cells were grown in serum + LIF conditions, and were stably transfected with each transgene. Selection for the transgenes was achieved by addition of the antibiotic hygromycin B to the culture media at a concentration of 100ug/ml for a minimum of ten days to select for transformants. Stable cell lines were then seeded for reprogramming. The pre-iPS morphology of the mouse Nanog and Saccoglossus Vent2 lines showed a slightly increased propensity to form pre-iPS colony-like structures in serum LIF conditions, but cells expressing Xenopus Vent2 remained unchanged. Cells expressing the Nematostella Vent gene, however, again began to demonstrate a disposition to form a more rounded and refractile cell type throughout the culture (Fig. 16A), and its propensity to do so increased after selection and became more prominent after the cultures were passaged (Fig. 16B). The smaller cells at a more intermediate phase of reprogramming in the starting cell line have a tendency to attach to the larger cells that display a more elongated somatic morphology, and the smaller cells form 3D aggregates similar to early colony formation. However, the round Nematostella cells could comfortably propagate as a monolayer attached to the gelatin treated surface of the six well plate (Fig. 16C).

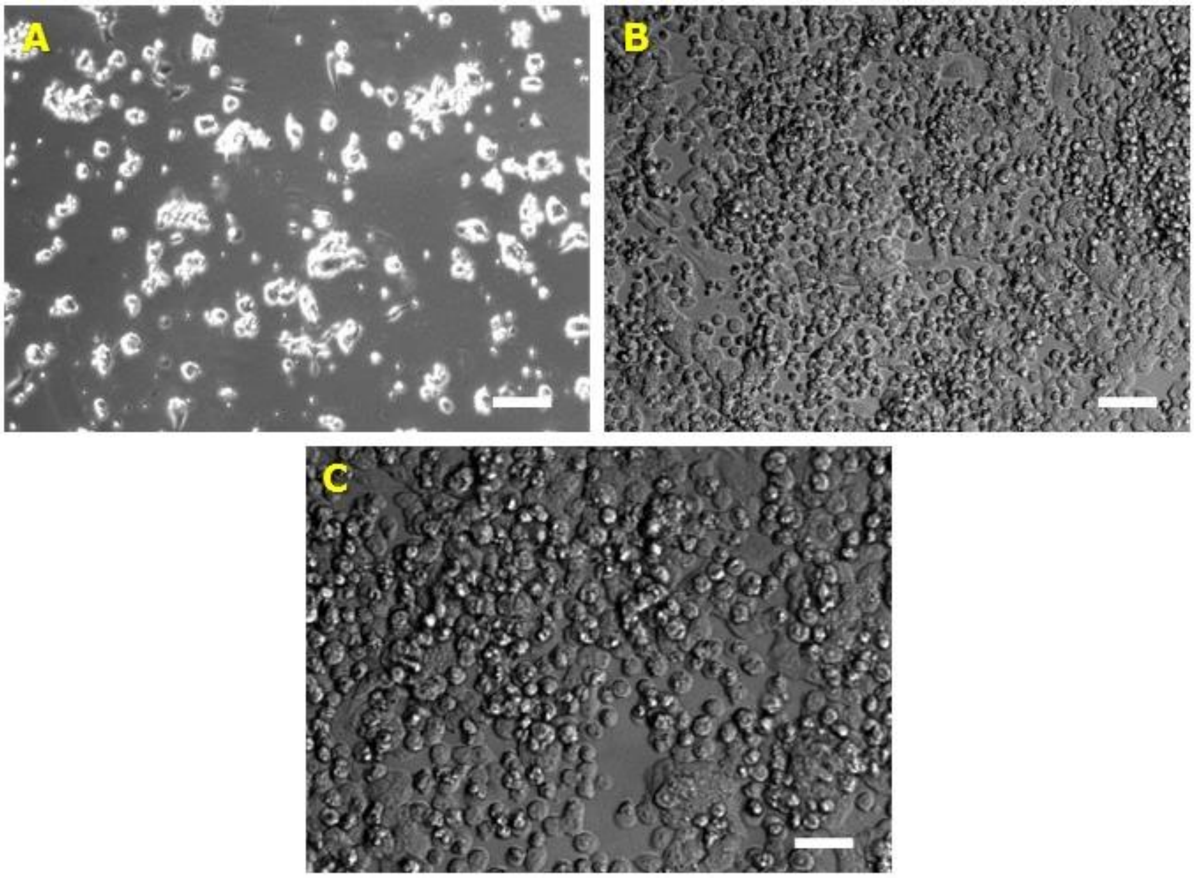


Figure 16: Nematosella Vent induces a round phenotype
(A) Pre-iPS cells transfected with Nematostella Vent begin to show an increasing proportion of partially reprogrammed cells. Scale bar represents 200um.
(B) Cell which have a round morphology become more prominent with increasing passage number. Scale bar represents 100um.
(C) Round cells attach to the gelatin treated surface in a monolayer, and can proliferate comfortably. Higher magnification; scale bar represents 50um.

All four pre-iPS lines were then seeded in serum + LIF on gelatinised six well plates, before media was switched to 2i + LIF after 48 hours to initiate reprogramming. The switch to 2i + LIF media represents T0 of the reprogramming assay.

3.1.8 Saccoglossus Vent2 can generate iPS colonies in its native form

By the time the media was switched to 2i + LIF, the Saccoglossus Vent2, Nematostella Vent, and xenopus Vent cultures all remained sparse. The monomer cell cultures had grown to a noticeably lower extent than their counterparts expressing the WR domain. The absence of a WR domain, which had conferred accelerated growth on the lentiviral lines, combined with the removal of serum from the culture conditions at this point to inhibit the cell proliferation. Cells expressing Saccoglossus Vent2 failed to produce any colony-like structures by day 6, and died in the subsequent four day selection for reprogrammed cells with G418 (Fig. 17A). The unhealthy appearance of the cultures suggested the initial seeding density was inadequate to facilitate the testing of the genes for reprogramming activity. Therefore the seeding numbers were doubled for all lines. The mouse Nanog positive control showed early colony formation as expected (Fig. 17B), however this gene in its native form conferred an accelerated growth rate on the cells, and due to the increased seeding density was already fully confluent by the midpoint of the assay. The G418 selection could not remove the over-growth of cells, and the culture died out by day 10 (Fig. 17C).

The Mouse Nanog line was seeded at 20K cells per well, and the other three lines at 40K per well. This ensured adequate cell numbers for colony formation (Fig. 17D), and removed the possibility of over-growth which would compromise the cultures. The Nematostella Vent cultures exhibited massive cell death when switched to 2i + LIF serum-free conditions. Concurrent with the lentiviral line, the Nematostella Vent cultures could not be maintained to the point of antibiotic selection (Fig. 17; E). The Xenopus Vent2 negative control cultures displayed no reprogramming activity, and died off in response to antibiotic selection (Fig. 17F).

Cultures which generated colonies were stained with Xgal for LacZ activation in the endogenous Nanog locus, and also AP staining with a Sigma Alkaline phosphatase kit. The number of positively stained colonies was counted.

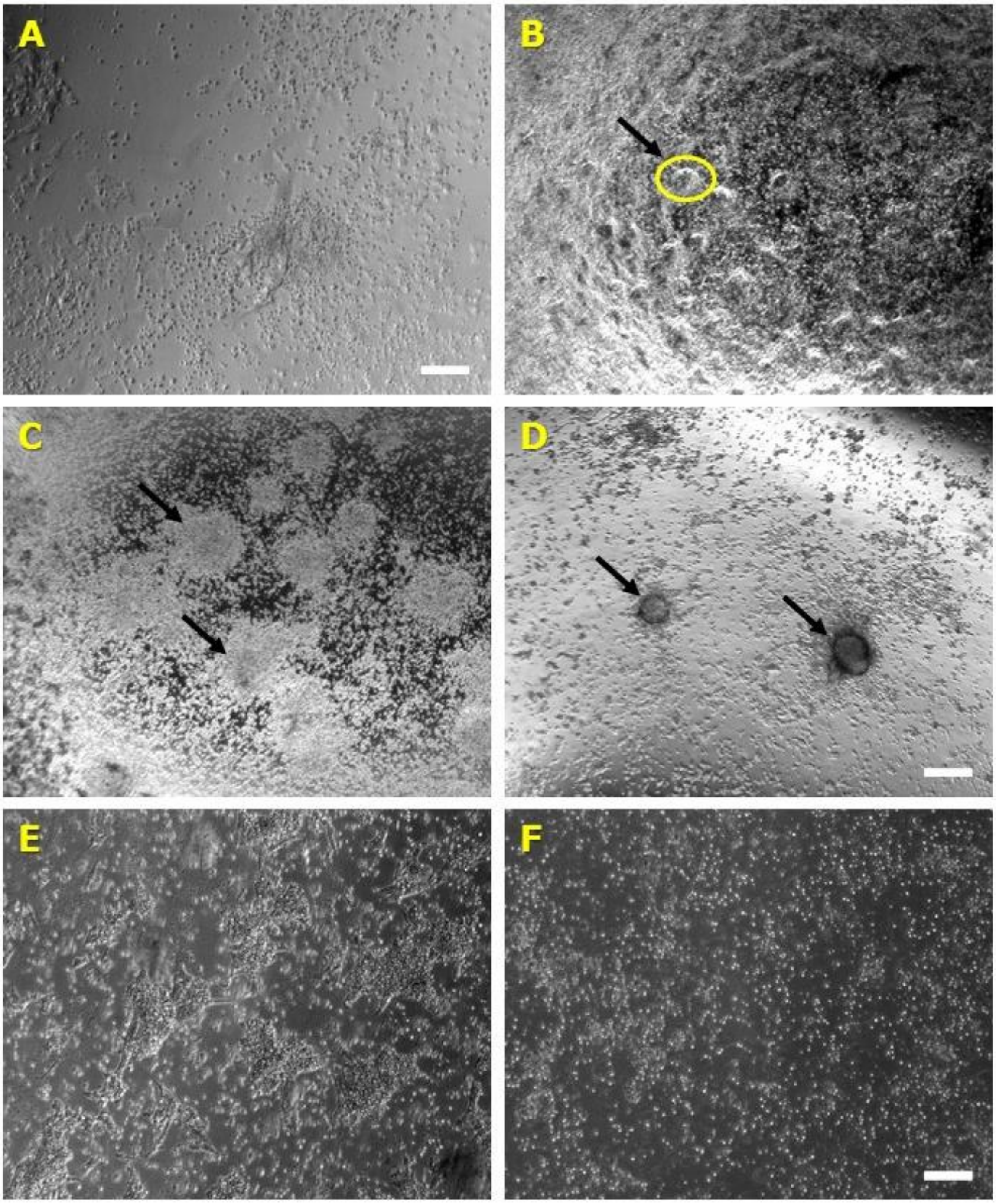


Figure 17: Optimising seeding densities for reprogramming
(A) Saccoglossus Vent2 by day 8 in 2i + LIF when seeded at 20K per well. (B) Increased seeding caused mouse Nanog to quickly reach maximum confluency. Colonies indicated by Arrows. (C) Mouse Nanog at day 10; colonies had collapsed. (D) Saccoglossus Vent2 cells seeded at 40K cell/well facilitated colony formation. (E) The Nematostella Vent line dying prior to selection. (F) Xenopus Vent negative control dead by day 10. Scale bar for A, E, and F represents 200um. Scale bar for B, C, and D represents 500um.

3.1.9 Saccoglossus Vent2 can activate endogenous Nanog as a monomer

Colonies formed from pre-iPS cells expressing the Saccoglossus Vent2 gene in its native monomeric form were analysed to determine if the endogenous Nanog locus was transcriptionally active by staining for lacZ activity. The culture generated on average 6.5 colonies per well (Table 3), indicating that Saccoglossus Vent2 protein in a monomeric form is sufficient to fully reprogram the pre-iPS cells (Fig. 18B). The Saccoglossus Vent2 colonies were comparable in size to those of the mouse Nanog culture (Fig. 18A), however there were much fewer colonies (Table 3).

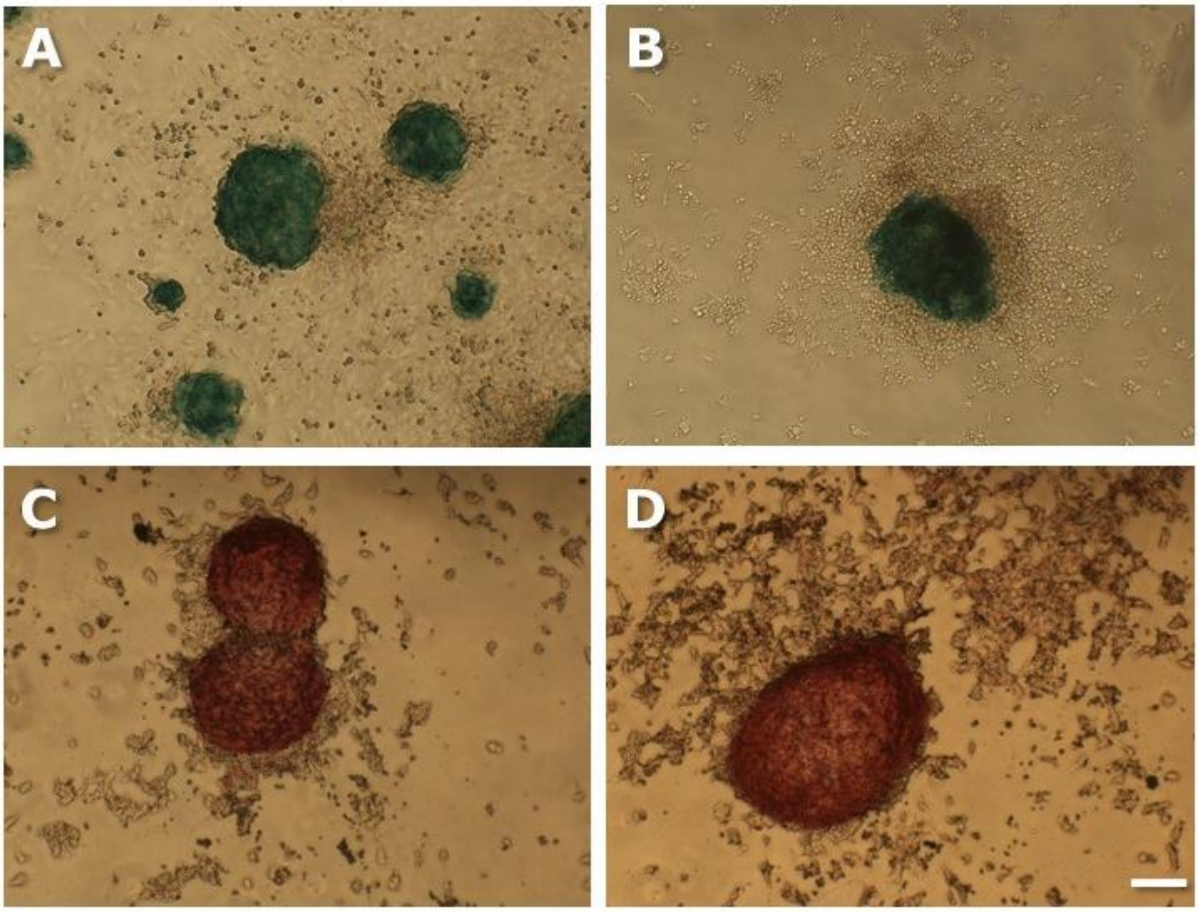


Figure 18: Saccoglossus Vent2 in its native form is sufficient to reprogram Nanog null pre-iPSCs
Positive staining for LacZ indicates that the cells have been reprogrammed to the point where the endogenous Nanog locus has been activated in both (A) the mouse Nanog positive control cultures and (B) the line expressing Saccoglossus Vent2 in its native monomeric form. (C) Mouse Nanog stained positive for AP. (D) Red staining indicates that the Saccoglossus Vent2 line has been reprogrammed to the point where it now displays ESC like levels of the AP marker for pluripotency. Scale bar represents 100um.

The cultures expressing lacZ were also stained for expression of alkaline phosphatase, as an indicator of pluripotency. AP staining gave concurrent results to that of the LacZ stain, for both the mouse Nanog and Saccoglossus Vent2 cultures (Fig. 18; C, D). The positively stained colonies reaffirm the capacity of the Saccoglossus Vent2 gene to reprogram somatic cells to pluripotency.

Positively stained colonies were counted and the efficiency of reprogramming was calculated (Table 3).

Cell Line	**LacZ+**			**AP+**			**Total Colonies**	**Average per well**	**Colonies per 10^4 seeded**
Mouse Nanog	123	146	131	125	111	149	785	130.83	**65.42**
Saccoglossus Vent2	6	4	7	6	8	8	39	6.5	**1.63**

Table 3: Reprogramming efficiency of monomeric Saccoglossus Vent2
Positive staining colonies were counted for each well. The total number of colonies generated for each gene was calculated as the sum of both LacZ and AP positive colonies. The final efficiency of colonies generated per 10^4 cells seeded, was based on the initial seeding density of 20K cells per well for mouse Nanog, and 40K per well for Saccoglossus Vent2.

The culture of cells expressing Saccoglossus Vent2 generated only 2.5% as many colonies as the mouse Nanog gene, producing on average 1.63 colonies per well for every 10k cells seeded for the assay. Interestingly, both the mouse Nanog and Saccoglossus Vent2 cultures were less efficiently reprogrammed by the PiggyBac vector system (Table 3), than the Lentiviral transduction system of gene delivery (Table 2). The mouse Nanog cultures on average produced 25.3% less colonies when reprogrammed using PiggyBac. The Saccoglossus Vent2 gene in PiggyBac showed a larger reduction in the number of colonies produced, as on average it generated 62.4% less colonies per 10K cells seeded. This may be due to a higher level of expression required to initiate reprogramming. As stated, the mouse Nanog CDS used in reprogramming naturally contains the WR domain which facilitates the formation of Nanog homodimers. Future work testing the reprogramming activity of a mouse Nanog molecule minus the WR domain, would provide a more direct comparison with the monomeric Saccoglossus Vent2 used in this assay. It is also possible that the reduced proliferation rate displayed by the monomeric Saccoglossus line generates less colonies than the cells expressing the WR domain as a result of attaining a lower overall confluency during the assay. Cells from both sets of cultures were passaged in 2i + LIF to establish an iPSC line.

3.1.10 Saccoglossus Vent2 iPSCs unstable in culture

Reprogrammed cells were passaged on to fresh plates and maintained in serum-free conditions. The Saccoglossus Vent2 line displayed a lower proliferation rate than that of the Nanog reprogrammed cells, presumably due to the forced expression of the WR domain in the mouse Nanog iPSC line. However, similar to cells reprogrammed with lentiviral directed expression of Saccoglossus Vent2-WR fusion, cells reprogrammed with the monomer construct displayed a propensity to differentiate. The instability of these cells was specific to the Saccoglossus Vent2 line, as cells expressing Nanog remained as relatively homogeneous ESC-like cells in culture. Maintenance of ESC-like colonies from cells reprogrammed with Vent2 required regular addition of G418 to the 2i + LIF culture media at a concentration of 200ug/ml (Fig. 19B). In the absence of antibiotic selection, an interconnecting web of differentiated cells surrounded the iPS colonies throughout all Vent2 cultures (Fig. 19A). The culture retains ESC-like cells throughout this process, as the compact iPSC colonies remain G418 resistant, which demonstrates that the endogenous Nanog locus remains active in these cells (Fig. 19B). Homogeneous cultures of iPSCs were collected for RNA isolation, and later quantitative and qualitative analysis of gene expression.

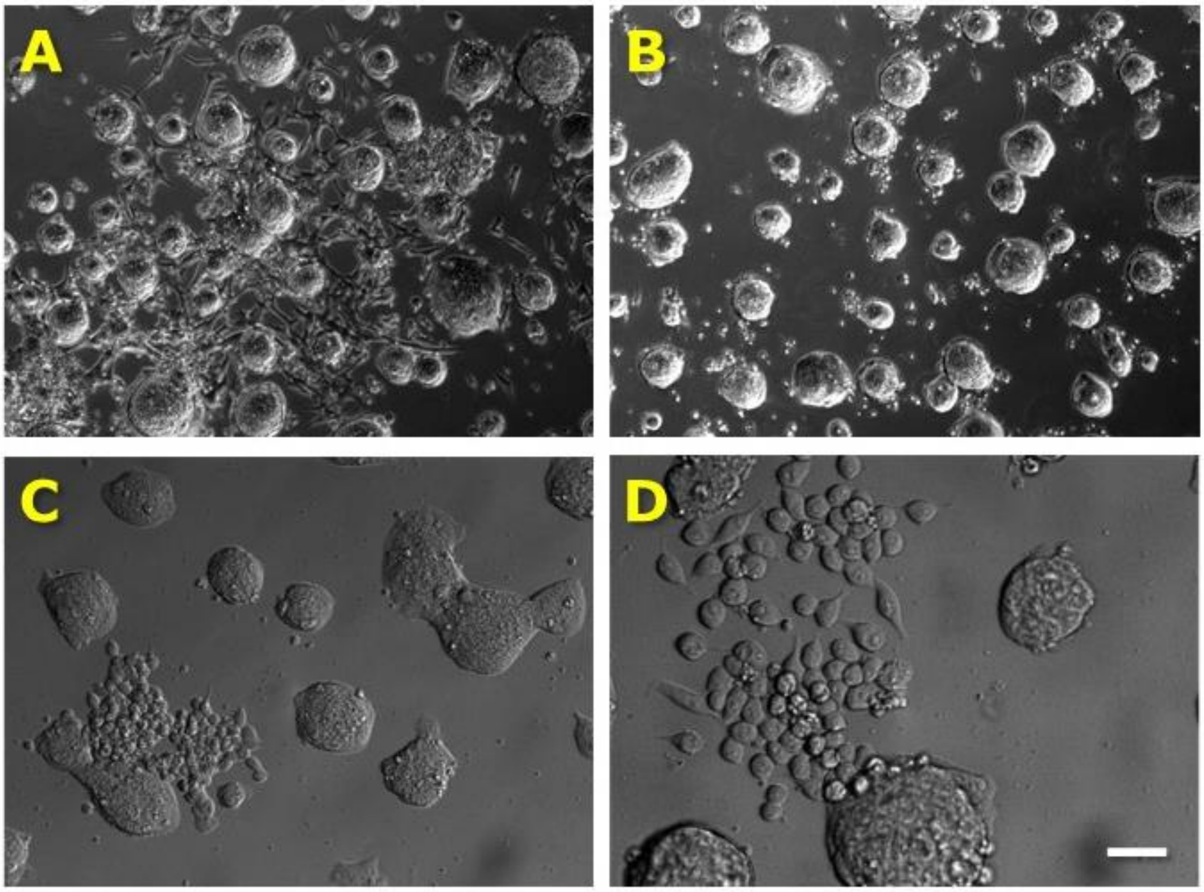


Figure 19: Instability of the Saccoglossus Vent2 iPSCs
(A) Saccoglossus Vent2 line absent G418. (B) G418 restores homogeneity, central iPSC colonies remain. (C) Cells also differentiate to large rounded cells. (D) Differentiated Saccoglossus Vent2 cells display both a prominent round morphology and an epithelial-like morphology, similar to Nematostella Vent pre-iPS line. Scale bar represents: 200um for A, B. 100um for C. 50um for D.

Elongated neural-like cells most commonly differentiate within these cultures (Fig. 19A). These cells form consistently in the absence of the antibiotic. However, a second type of differentiated cell also arises less frequently, as patches of more rounded cells which do not expand well in the culture, and appear only in small clusters briefly in the culture (Fig. 19C).

These cells closely resemble the rounded cells induced in Nematostella as part of heterogeneous pre-iPS cell cultures (Fig. 16). In the Saccoglossus Vent2 iPSC cultures, the rounded cells are a consequence of differentiation which makes them easier to identify due to their morphology contrasting to that of the large compact iPSC colonies. It can be seen that this population of large differentiated cells consists of a round cell type and cells of a more epithelial morphology (Fig. 19D). The round cells resemble those induced by Nematostella Vent, suggesting their induction is reflects a specific property of invertebrate Vent genes. We decided, therefore, to examine the Nematostella Vent (NvVent) pre-iPS cell line.

3.1.11 Nematostella Vent drives an unusual phenotype

As seen in figure 16, Nanog -/- pre-iPS cells transfected with the Nematostella Vent gene divert to a rounded cell type, the number of which increases exponentially to form the majority of the cell population within those cultures. These pre-iPS cells were seeded for reprogramming uniformly with lines expressing other candidate genes; however only the sea anemone Vent line could not survive in 2i + LIF conditions. To gain a closer look at the morphology of these cells, we isolated a culture which displayed a high ratio of the rounded cells within its population, and applied a haematoxylin and eosin (H&E) stain. The high proportion of blue hemalum, which specifically colours the nucleus, demonstrates a high nuclear to cytoplasmic ratio. Prominent nucleoli indicate that the cells are transcriptionally active (Fig. 20B). The fact that these cells do not survive 2i neurobasal conditions suggested that the cells have developed dependency on a signalling pathway with which the small molecule inhibitors interfere. The inhibitors used are PD which inhibits MEK signalling, and the Chiron inhibitor of GSK-3 which causes an upregulation of Wnt signalling. One, or both, of these two pathways were subsequently identified as crucial to the expansion of the Nematostella Vent pre-iPS cells as they do not survive 2i + LIF conditions (Fig 10A, Fig. 17E).

To identify the inhibitor that disrupts the cultures, I tested the effects of the media components separately. The GSK-3 inhibitor at a concentration of 3uM did not alter the phenotype of the cells when added to the media (Fig. 20C). Cells under these conditions maintained the sub population of round cells at levels similar to basic serum + LIF conditions (Fig. 20A). However, addition of the MEK inhibitor at a concentration of 1uM began to eliminate the round cells from the culture over the course of 24 hours, demonstrating that these cells are ERK dependent (Fig. 20D).

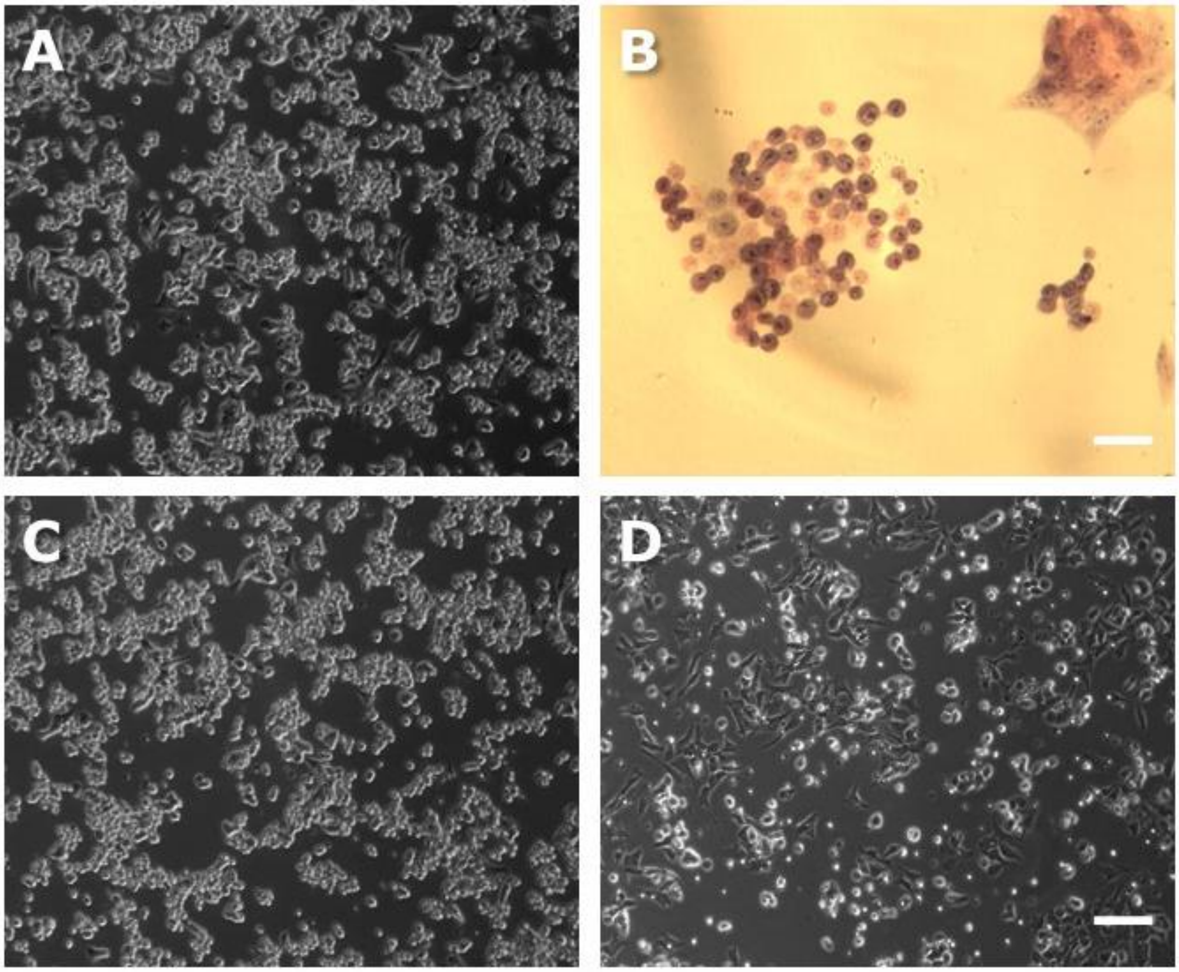


Figure 20: Testing the effects of 2i molecules on NvVent cells
(A) The NvVent pre-iPS cells grown in serum-LIF show a rounded phenotype. (B) H&E staining demonstrates high nuclear to cytoplasmic ratio, and pronounced nucleoli demonstrate that the cells are transcriptionally active. Scale bar represents 50um. (C) The cells grown in serum + LIF + GSK-3 inhibitor. (D) The cells grown in MEK inhibitor for 24 hours; high cell death, and the rounded cells become less prominent in the culture. Scale bar represents 100um.

Interestingly, a subset of the Nematostella Vent culture, growing in a monolayer of rounded cells, demonstrated positive staining for AP, a pluripotency marker (Fig. 21A). These were reduced in number after the addition of MEK inhibitor (Fig. 21C). The proportion of cells staining positive for AP was unaffected by the addition of GSK-3 (Fig. 21B).

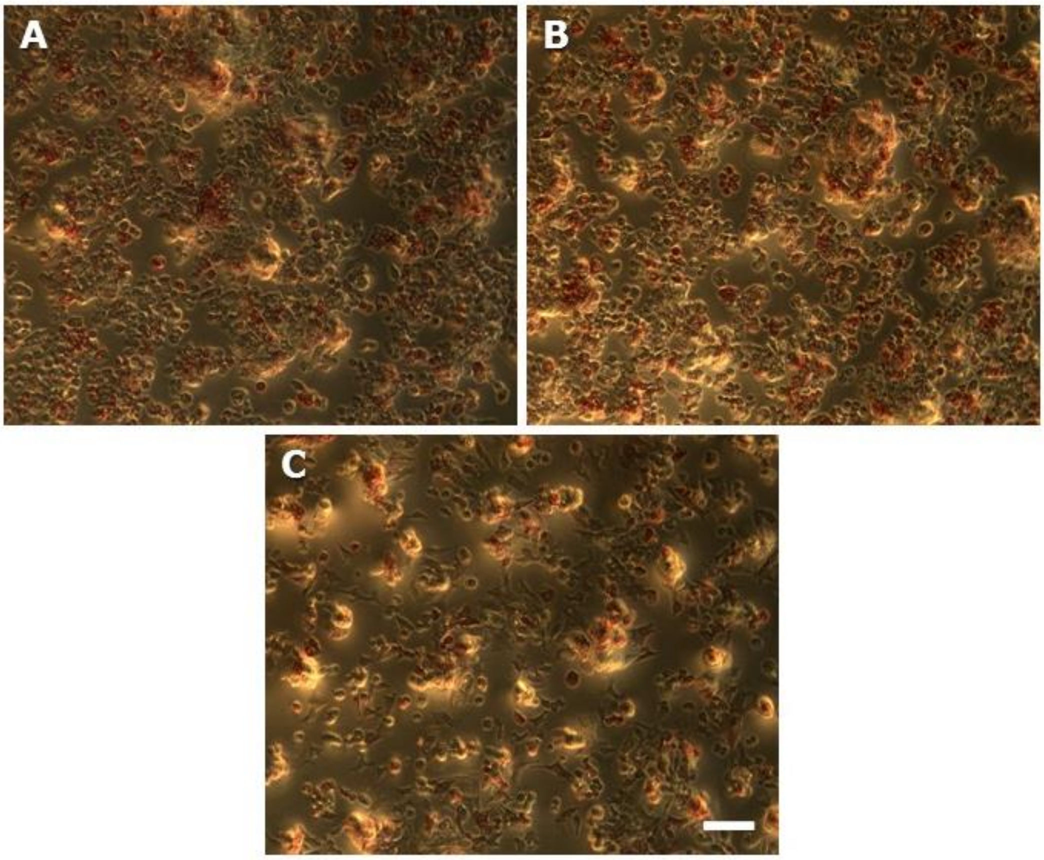


Figure 21: AP staining of Nematostella Vent line in different inhibitors
(A) A small portion of the NvVent pre-iPS cells stain positive for AP in serum LIF. (B) These cells are unaffected by the addition of GSK-3 inhibitor. (C) MEK inhibitor reduces the number of round cells present in the culture, coinciding with a drop in the level of positively stained cells. Scale bar represents 100um.

Transfection of the pre-iPSC line with Nematostella Vent appears to induce a more rounded and refractile cell type throughout the culture, and the cells acquired dependence on ERK signalling. These attributes are characteristic of XEN cells, which are representative of the extra-embryonic endoderm of the mouse blastocyst (Kunath et al. 2005). RNA samples were collected from the cultures displaying the highest proportion of XEN-like cells, to be analysed at a later point.

3.2 QPCR ANALYSIS OF REPROGRAMMED CELLS

To determine if NS pre-iPS cells were completely reprogrammed by the Nanog gene from mouse (mNanog) and the Vent2 gene from the acorn worm Saccoglossus kowalevskii (SkVent2), RNA samples were collected and used to make First strand cDNA. mNanog and SkVent2 cells reprogrammed using both the lentiviral constructs and the PiggyBac gene delivery system, were analysed for expression of pluripotency markers relative to wildtype mouse ESCs and the untransfected Nanog-null NS derived pre-iPSCs (Fig. 22). Cells reprogrammed with a lentiviral SkVent2 construct contained a tryptophan repeat domain (WR) added to the carboxy terminus of the gene, and the lentiviral reprogrammed cells are denoted in figure 21 by the addition of –WR to the sample name. Primer sequences from Silva et al. (2011) were used to specifically target the endogenous Oct4 gene, and differentiate its expression from that of the retroviral Oct4 that is constitutively present in the pre-iPS cells. All iPSC samples showed activation of the endogenous Oct4, with cells reprogrammed with mouse Nanog showing a relatively higher level than the cells reprogrammed with the invertebrate Vent. Krüppel-like factor 2 (Klf2) is essential for ground state pluripotency (Yeo et al., 2014), and was also activated in all iPSC samples, but was present at a lower level than in mESCs. Rex1 is a direct target of Nanog (Shi et al., 2006), and was activated in all iPSCs, but was most highly upregulated in mouse Nanog iPSCs. Estrogen-related receptor b (EsrrB) is also a direct target of Nanog (Festuccia et al., 2012), and was successfully activated in all the iPSC lines. Interestingly, it was similarly upregulated in both the mouse Nanog and Saccoglossus Vent lines, almost four-fold higher relative to wildtype ESCs, suggesting that it may have been directly targeted by the invertebrate Vent gene as well as Nanog. Nrob1 is a marker of naïve pluripotency (Nichols and Smith, 2009), and was similarly activated and upregulated by both mouse Nanog and Saccoglossus Vent2. Tbx3 is a pluripotency factor that is directly bound by Nanog, and can be used as a reprogramming factor itself to increase germ-line competency of iPSCs (Han et al., 2010), and Fig. 22 shows it can be activated by both Nanog and invertebrate Vent genes during reprogramming. Fbx15, expression of which is specific to undifferentiated ESCs (Tokuzawa et al. 2003), was also activated in all iPSC lines. Activation of these endogenous pluripotency factors indicates that the cells were fully reprogrammed to ground state pluripotency by both mouse Nanog and Saccoglossus

Vent2. I observed that the majority of the genes targeted by the qPCR analysis demonstrated expression levels which were elevated relative to wildtype ESCs, suggesting that their expression was being directly upregulated by both the mouse Nanog and Saccoglossus Vent transgenes.

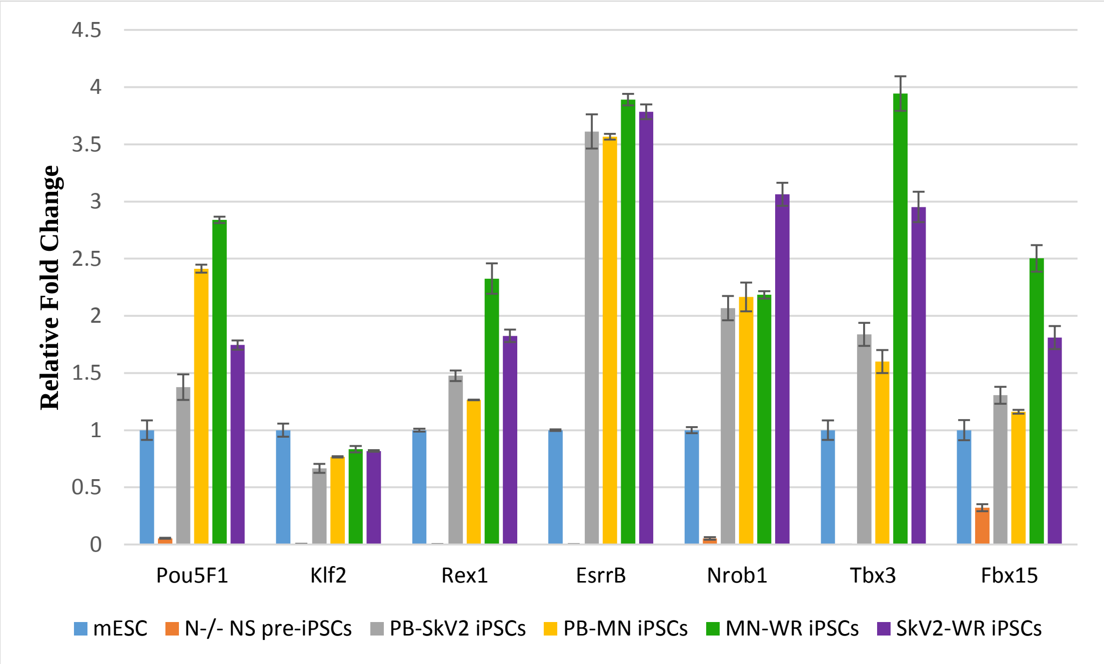


Figure 22: qPCR analysis for expression of pluripotency markers in reprogrammed NS cells

qPCR analysis of iPSCs derived from reprogramming of Nanog-null NS pre-iPSCs. Data was standardised against ground state CGR8 wildtype mouse ESCs, using β-Actin as an endogenous control. PB-MN and PB-SkV2 represent the PiggyBac reprogrammed mouse Nanog and acorn worm Vent2 iPSCs respectively. MN-WR and SkV2-WR indicates the lentiviral constructs, which added the tryptophan repeat domain to the SkV2 gene.

As the PiggyBac reprogrammed cells utilised a gene delivery system that inserts constructs at specific sites, resulting in a constant controlled level of expression in the resulting iPSCs, these lines were chosen for further analysis. These iPSCs also permitted analysis of cells reprogrammed by the invertebrate Vent gene in its native monomeric form; without the WR domain. Silencing of the exogenous retroviral factors marks the acquisition of a completely reprogrammed state (Hotta and Ellis, 2008; Buganim et al, 2012). Cells were collected after ten or more passages and analysed to determine if the retroviral Yamanaka factors expressed in the pre-iPS cells were silenced in the iPSC lines (Fig. 23). The NS pre-iPS cells constitutively express exogenous Oct4, Klf4 and cMyc which were transduced by retrovirus (Silva et al., 2009), but maintain high levels of endogenous Sox2 due to the neural origin of the pre-iPSC line.

All three retroviral Yamanaka factors are downregulated in the established iPSC lines (Fig. 23). This verifies complete reprogramming of the cells to ground state pluripotency.

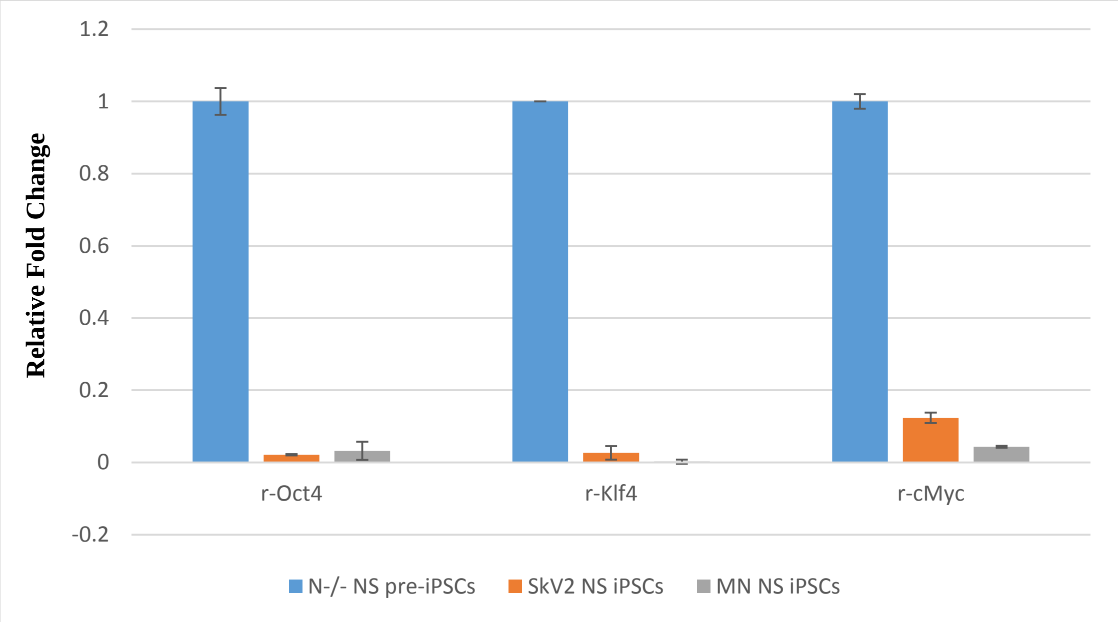


Figure 23: Silencing of exogenous pluripotency factors.

Retroviral Oct4, Klf4, and cMyc were downregulated in all reprogrammed iPSCs, relative to the untransfected pre-iPSCs. β-Actin was used as an endogenous control.

3.3 EMBRYOID BODY DIFFERENTIATION

At the same late passage used for RNA extraction, iPSCs were also collected from both the mouse and acorn worm to make embryoid bodies (EBs). This was to test the capacity of cells reprogrammed by an invertebrate Vent gene to differentiate, relative to EBs made from wildtype mouse ESCs. Hanging drops were formed as previously described (Section 2.7) using mNanog iPSCs, SkVent2 iPSCs, and CGR8 wildtype mESCs. The EBs formed over three days, were then collected (Fig. 24A), and allowed to differentiate and grow in suspension cultures for ten days (Fig. 24B). The suspension cultures were maintained in basic serum containing media minus any inhibitors to promote spontaneous differentiation. After ten days the EBs were collected, and RNA was extracted for qPCR analysis.

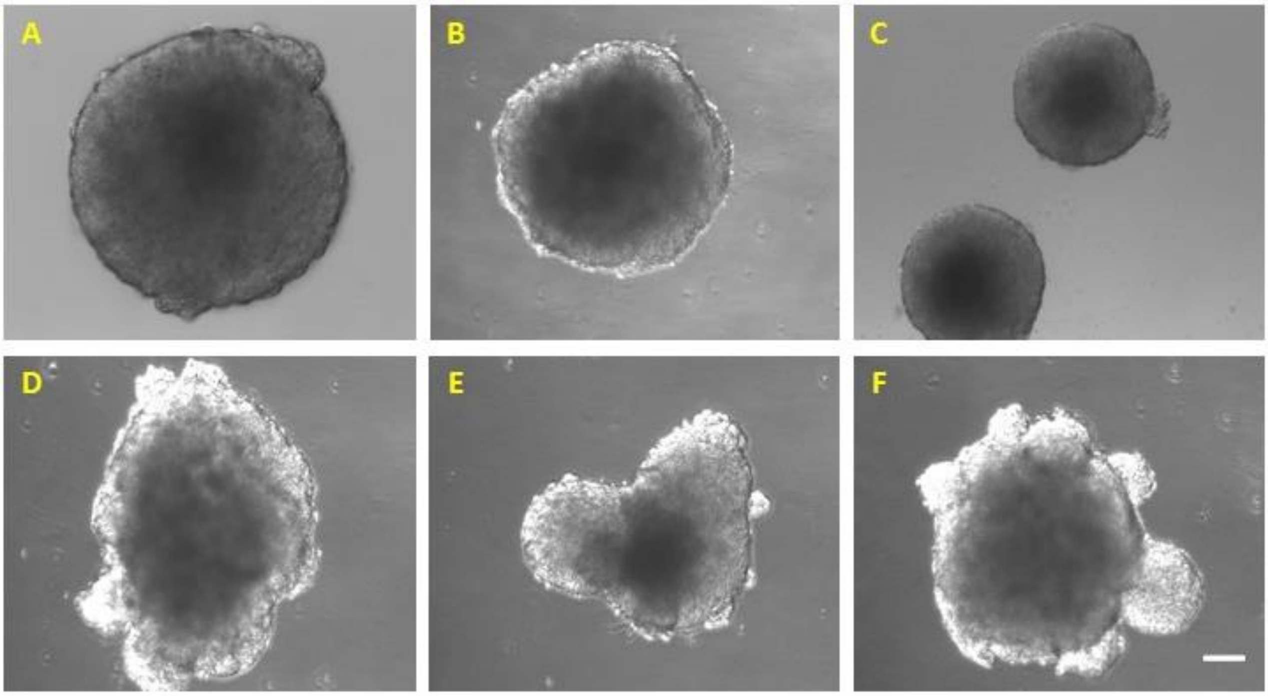


Figure 24: Differentiation of Embryoid bodies

(A) Embryoid bodies formed from mESCs in suspension culture at day 1. (B) Embryoid bodies formed from mNanog iPSCs in suspension culture at day 1 (C) Embryoid bodies formed from SkVent2 iPSCs in suspension culture at day 1. The mESC (D), mouse Nanog (E) and SkVent2 (F) embryoid bodies beginning to show spontaneous differentiation 5 days after collection from hanging drops. Scale bar represents 100um for A, B, D, E, F. Scale bar represents 200um for C.

3.4 qPCR analysis of Embryoid body differentiation

When differentiating in the absence of signalling inhibitors, EBs recapitulate many aspects of cell differentiation during early mammalian embryogenesis and give rise to mature cells of all three primary germ layers (Desbaillets et al., 2000). To determine if the iPSCs were capable of generating cells representative of all three germ layers, two markers were chosen for each lineage. Sox17 and alpha fetoprotein (AFP) were selected for endoderm, NeuroD and Cytokeratin (Krt18) for ectoderm, and Myf5 and Brachyury (T) were chosen to assay for mesodermal differentiation (Fig. 25).

Both the mNanog and Saccoglossus Vent2 EBs showed high upregulation of markers representative of all three germ layers, indicating differentiation, and verifying that cells had been reprogrammed to the ground state (Fig. 25). The Saccoglossus Vent2 cells displayed a greater induction of endodermal markers Sox17 and AFP relative to either wildtype ESCs or mouse Nanog iPSCs. Mesodermal differentiation identified by Myf5 and T was further verified by spontaneous differentiation to cardiomyocytes in the cultures, identified by the characteristic pulsation of the EBs. The ectodermal marker NeuroD, associated with neural differentiation was the gene most highly upregulated in both iPSC lines. This may be influenced by the neural origin of the pre-iPS cells. The harbouring of residual DNA methylation signatures characteristic of the somatic tissue from which the pre-iPSCs originated, has been previously demonstrated to direct differentiation in favour of lineages related to the donor cell (Kim et al. 2010). However, the cells successfully activated genes associated with all three germ layers, indicating that they are fully reprogrammed.

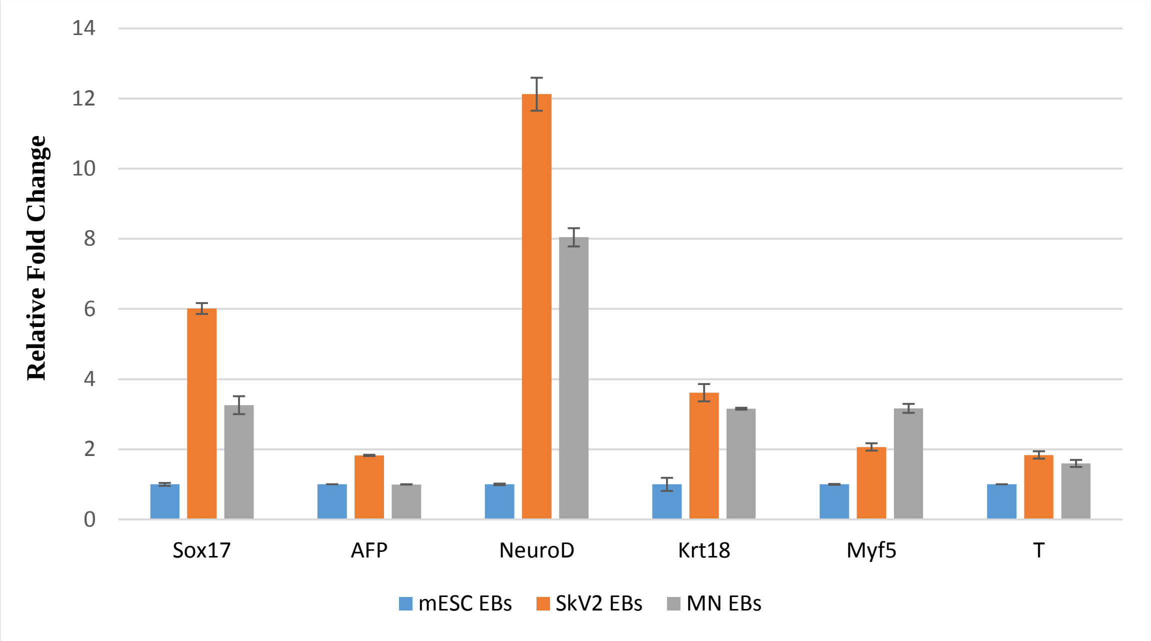


Figure 25: Differentiation of EBs to all three germ layers

EBs formed from ESCs and both mouse Nanog and Saccoglossus Vent iPSC lines show activation of markers specific to all three germ layers. Data was standardised to mESC EBs. β-actin was used as an endogenous control.

3.5 MEF REPROGRAMMING

3.5.1 Creation of TNGA pre-iPS lines

The MEF pre-iPS cell line used was derived from a chimeric mouse where one of the endogenous Nanog alleles had been replaced by the GFP targeted to Nanog clone-A (TNGA) through homologous recombination (Chambers et al. 2007). To test the reprogramming capacity of the Saccoglossus Vent2 gene further, we decided to transfect it into the TNGA MEF pre-iPS cell line, and attempt to reprogram primary cells to the ground state. The GFP-PURO reporter located at the endogenous Nanog locus in one allele in the TNGA cells (Chambers et al., 2007), would allow us to track activation of endogenous Nanog in real-time during reprogramming. Unlike the N-/- neural stem cell derived pre-iPS line used previously, the TNGA MEFs have one functional endogenous Nanog allele remaining intact. Therefore, production of endogenous Nanog protein could stabilise an established iPS line, and potentially counteract the differentiation observed in the reprogrammed neural pre-iPS lines described above (Fig. 19).

The PiggyBac gene delivery system was used to establish TNGA pre-iPS lines constitutively expressing candidate genes. The reprogramming activity of both Saccoglossus Vent2 and Nematostella Vent were compared. Mouse Nanog and Xenopus Vent2 were used as positive and negative controls respectively. The TNGA pre-iPS cells were transfected with PiggyBac constructs in serum + LIF conditions. Stable transfectants were expanded after standard selection with HygromycinB, then seeded for reprogramming. The cells were maintained at a confluency above 50% throughout the establishment and expansion of the lines, and demonstrated no change in their morphology relative to the untransfected MEFs (Fig. 26; A).

3.5.2 Reprogramming of the TNGA lines

Each TNGA MEF line was seeded for reprogramming in serum + LIF conditions on gelatin. All lines were seeded at a low density to facilitate the cells remaining on a single culture vessel for an extended period of time without reaching maximum confluency. This was to facilitate the formation of GFP-negative partially reprogrammed colonies in serum + LIF conditions, prior to a switch to 2i + LIF media which mediates the reprogramming to ground state pluripotency. All Vent lines were seeded at 10K cells/well, but the mNanog line was seeded at 5k/well to account for an accelerated growth rate conferred on it by the presence of the WR domain. The cells showed a greater propensity to form GFP-negative micro-colonies when grown from a lower confluency. This was identified by the elongated fibroblast cells growing in tightly packed clusters from clonal density. The density of the clusters increased as a monolayer as the cells wrapped around one-another. Individual cells then became difficult to distinguish in pre-iPS colonies which were GFP-negative, as indicated by the arrow in Figure 24C. Complete reprogramming was then mediated by a switch to 2i + LIF conditions. Reprogramming of the TNGA lines does not entail defined time points for media change like those that were applied when reprogramming the Nanog-null NS pre-iPSCs. The timing of the switch to 2i + LIF media for each line was dictated by observation of GFP-negative colony formation within the cultures (Fig.

26B, C). The mNanog cells proliferated at a higher rate than the other pre-iPSC lines. However both the mNanog (Fig. 22B) and the SkVent2 (Fig. 22C) cultures began to show pre-iPSC colony formation. The mouse Nanog line developed larger colonies at a higher frequency than the Saccoglossus Vent2; however all cultures remained negative for GFP in serum + LIF indicating the endogenous Nanog locus had not been activated.

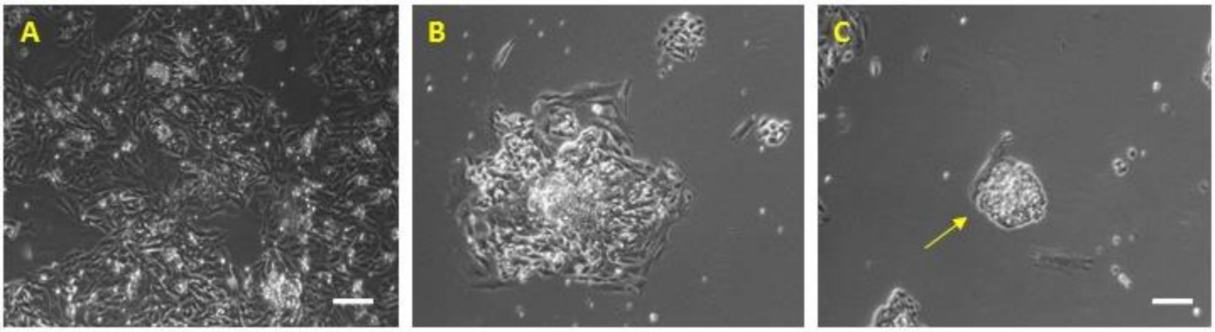


Figure 26: TNGA MEFs in Serum + LIF

(A) Untransfected TNGA MEF pre-iPSCs. Scale bar represents 200um at 100x magnification (B) The mouse Nanog line beginning to form GFP-negative colonies 4 days after being seeded. (C) The Saccoglossus Vent2 line showed micro-colony formation 4 days after being seeded for reprogramming. Scale bar for B and C represents 100um at 200x magnification

The line expressing Xenopus Vent2 failed to generate pre-iPSC GFP-negative structures similar to the mNanog or SkVent2 lines in serum + LIF conditions; individual cells remained visible in a monolayer (Fig. 27A). There was no activation of the GFP allele in 2i + LIF conditions, and the cells which morphologically displayed a more partially-reprogrammed phenotype died in 2i conditions (Fig. 27B).

The Nematostella Vent line showed a distinct morphology, different to the other lines in serum + LIF conditions, as they unexpectedly displayed a small rounded morphology (Fig. 27C), similar to the Nanog-null NS derived pre-iPSCs when transfected with the Nematostella gene (Fig. 16A). In accordance with this phenotype, the cells did not survive prolonged culture in 2i + LIF conditions, and after four days in 2i + LIF the culture died out (Fig. 27D).

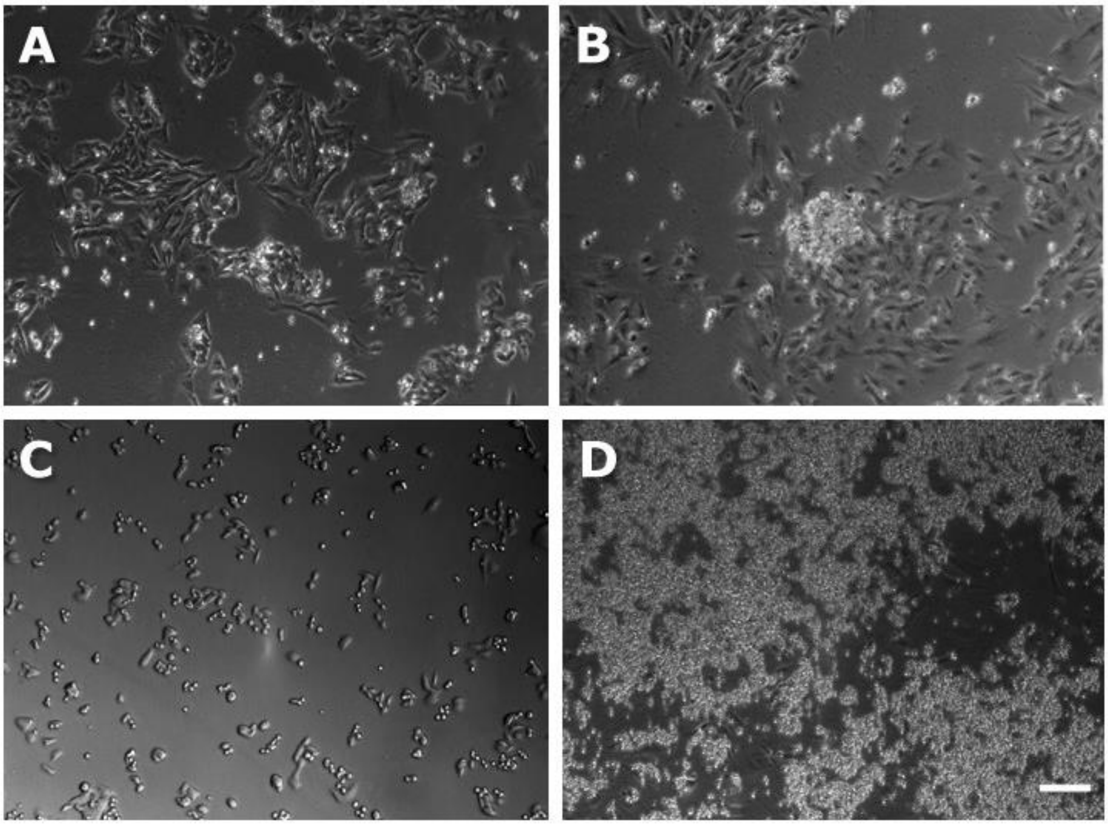


Figure 27: Nematostella Vent and Xenopus Vent TNGA MEFs

(A) Xenopus Vent line in serum + LIF (B) Xenopus Vent in 2i + LIF. (C) Sea anemone Vent in serum + LIF, (D) Sea anemone Vent dies out in 2i + LIF. Scale bar represents 100um.

Once micro-colonies were observed in the serum + LIF conditions, the media was swapped to 2i + LIF. The cells were subsequently checked for activation of the GFP reporter in the endogenous Nanog locus. GFP-positive colonies in both the mNanog (Fig.28 B) and SkVent2 (Fig. 28 E) cultures confirmed activation of the Nanog gene, suggesting fully reprogrammed cells. GFP expression is linked through an IRES to puromycin resistance encoded by the *pac* gene (Chambers et al., 2007), therefore GFP positive-colonies of fully reprogrammed cells were then selected for by adding puromycin to the culture media at a final concentration of 1ug/ml.

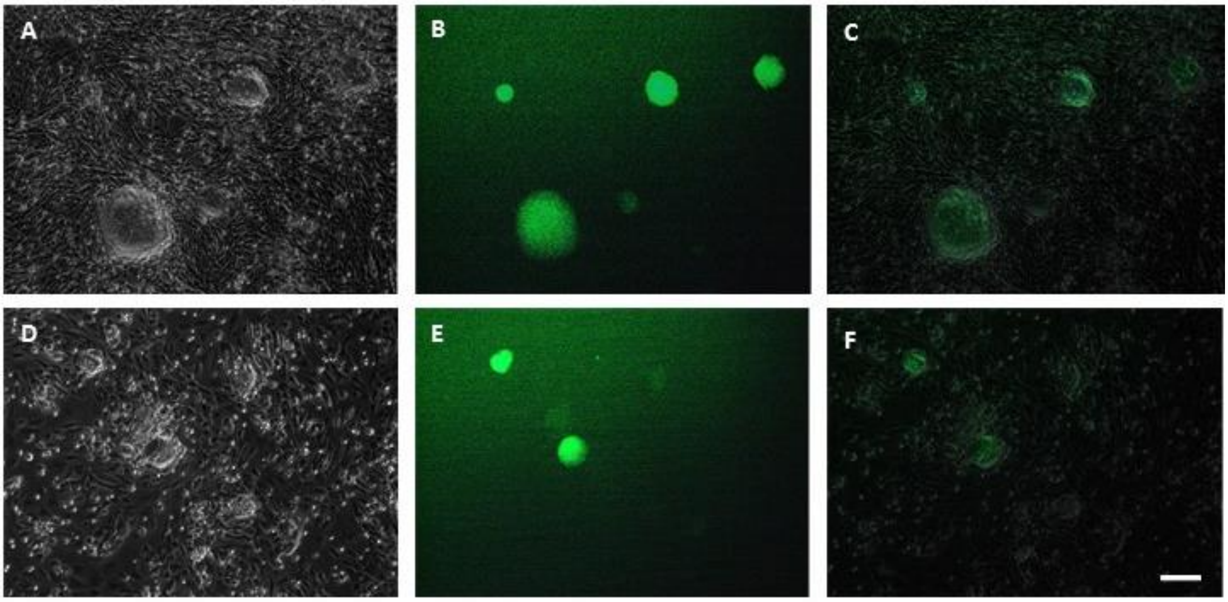


Figure 28: GFP-positive MEF derived iPSC colonies

(A) Mouse Nanog: Brightfield view of Mouse Nanog culture after 10days in 2i + LIF media. (B) GFP activation in the mouse Nanog colonies. (C) Merged image show that the GFP signal is localised to the colonies.

(D) Acorn worm Vent2: Small iPSC colony formation in the acorn worm Vent2 cultures. (E) GFP activation in acorn worm Vent colonies, (F) Merged image shows GFP signal in emerging acorn worm Vent2 iPSC colonies. Scale bar represents 100um.

GFP-positive colonies were counted in six separate wells of a six-well plate for each line, and the average number of colonies generated per 10K cells seeded was counted (Table 4). Mouse Nanog reprogrammed the MEFs more efficiently than the acorn worm Vent2 gene, generating a larger number of colonies from a lower seeding density. Mouse Nanog generated 166 colonies from 30K cells seeded on the plate in total, compared to 26 colonies generated from 60K acorn worm Vent2 cells seeded (Table 4).

Cell Line	GFP Positive colonies						Total Colonies	Average per well	Colonies per 10^4 seeded
Mouse Nanog	29	28	36	24	25	24	166	27.6	**55.3**
Acorn worm Vent2	4	6	6	5	2	3	26	4.3	**4.3**

Table 4: Efficiency of reprogramming TNGA MEFs
GFP-positive colonies were counted for each well. Each line was tested on a six well-plate. Efficiency was calculated per 10K cells seeded.

The cells were selected with Puromycin for 48 hours, before iPSCs were passaged with 1X trypsin EDTA onto a fresh gelatinised plate, and maintained in 2i + LIF. Puromycin was added at a maintenance concentration of 0.5ug/ml to clear the well of differentiated cells. Both the mNanog iPSC cultures (Fig. 28A), and the SkVent2 iPSCs (Fig. 28B) looked similar; however the mNanog line showed a higher proliferation rate than SkVent2. Both lines were seeded at 400K cells per well on a six-well plate at each passage, however, mNanog reached confluency after two days, compared with four days for SkVent2. Based on observation, the MEFs reprogrammed with Saccoglossus Vent2 displayed less of a propensity to differentiate than the NS cells reprogrammed with the same gene (Fig. 19). However, TNGA MEFs reprogrammed with Saccoglossus Vent2 (Fig. 29B) were less stable in culture than the mouse Nanog line (Fig. 29A), with more differentiation appearing in the Saccoglossus cultures over multiple passages. Homogeneity could be re-established by the addition of 0.5ug/ml puromycin to the media.

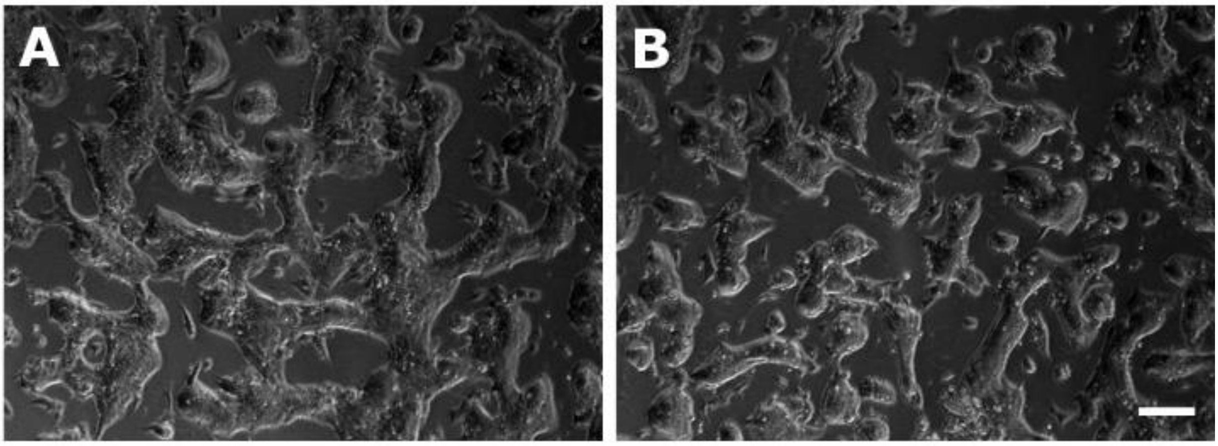


Figure 29: Expansion of TNGA iPSC lines

(A) Mouse Nanog iPSCs derived from TNGA line expanded in 2i + LIF.

(B) Acorn worm Vent2 iPSCs in 0.5ug/ml puromycin media after passaging. Scale bar represents 100um.

Both iPSC lines were expanded for several passages, before RNA samples were collected for later qPCR analysis.

3.6 QPCR ANALYSIS OF MEF DERIVED IPSCS

Both TNGA iPSC lines were characterised by assaying for activation of endogenous pluripotency markers. The same markers were used as in the assay of the reprogrammed NS cells.

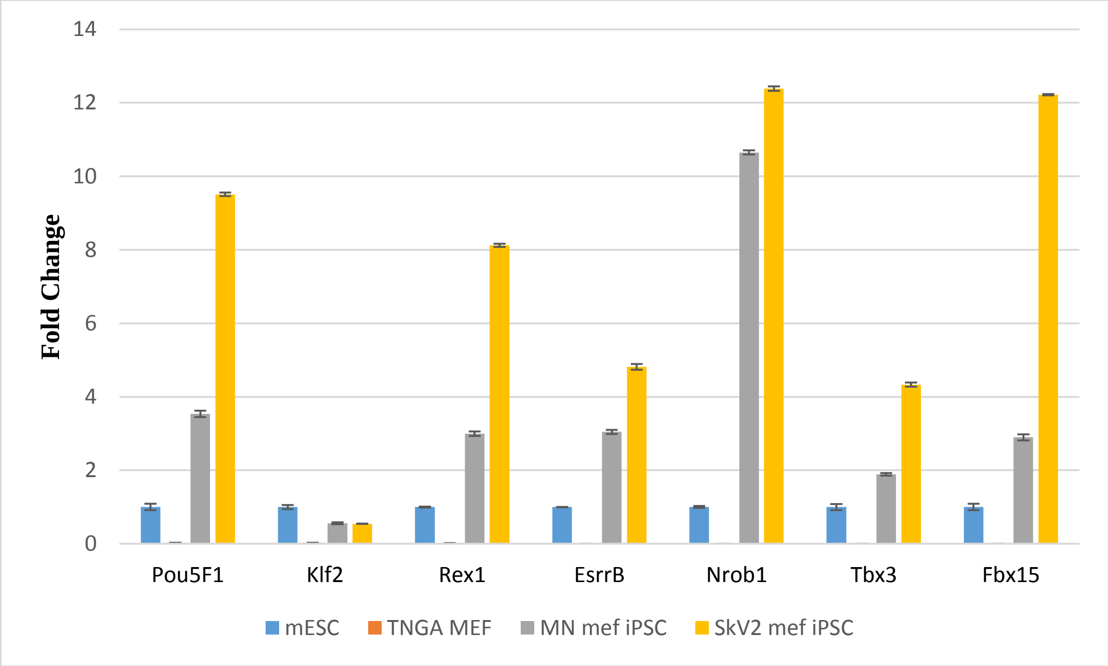


Figure 30: Activation of endogenous pluripotency genes in TNGA cells

qPCR analysis of iPSCs derived from reprogramming of Nanog-null NS pre-iPSCs. Data was standardised against CGR8 wildtype mouse ESCs, using β-Actin as an endogenous control.

Both mNanog and SkVent2 induced activation of endogenous pluripotency factors. Interestingly, I observed that the levels of expression were consistently higher in the cells reprogrammed with SkVent2 than either the wild types ESCs or the iPSCs reprogrammed with mNanog (Fig. 30). This included the relative levels of expression of Oct4 and Rex1, both direct binding targets of Nanog. Activation of endogenous markers concurred with the positive GFP signal emitted from the colonies, and suggests that Nanog activity sufficient to reprogram primary somatic cells to pluripotency is conserved in the Vent gene family at the base of deuterstome evolution.

In contrast to this, repeated attempts to reprogram with a Vent gene from the starlet sea anemone Nematostella vectensis were unsuccessful. However, the TNGA MEFs transfected with Nematostella Vent (NvVent) showed induction of a more rounded and refractile cell type throughout the culture, similar to the phenotype seen in the Nanog null pre-iPSCs transfected with the same gene. The cells also died early in 2i + LIF conditions during the reprogramming assay (Fig. 27D), suggesting they may have gained a dependence on ERK similar to the NS cells transfected with NvVent. To test this I compared the cells directly by means of qPCR analysis.

3.7 qPCR analysis of Sea anemone Vent pre-iPSC lines

A rounded and highly refractile phenotype was observed in both types of pre-IPS cells when transfected with NvVent. The cells also seem to have acquired a dependence on ERK signalling. As mentioned, these attributes resemble XEN cells, which are derived from the extra-embryonic endoderm of the mouse blastocyst (Kunath et al. 2005). IMA81 E11 Xen cells characterised by Kunath and colleagues (Fig. 31A), were cultured in basic Xen cell culture conditions, and contain a heterogeneous culture of cells representative of both visceral and parietal endoderm (Kunath et al., 2005). The XEN cells (Fig. 31A) display a refractile and rounded cell type similar to that of the pre-iPSCs transfected with NvVent (Fig. 31 B).

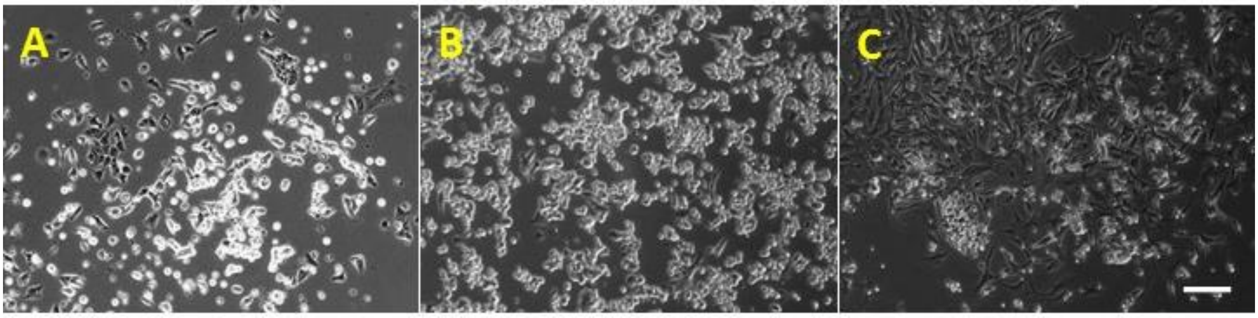


Figure 31: Xen cell comparison with NvVent transfected pre-iPSCs

(A) IMA81 E11 Xen cells in culture. (B) Nanog-null NS pre-iPS cells transfected with NvVent. (C) Untransfected NS pre-iPS cells. Scale bar represents 200um.

A sample of the XEN cells was collected and processed for RNA, and cDNA was synthesised for qPCR analysis (Fig. 32).

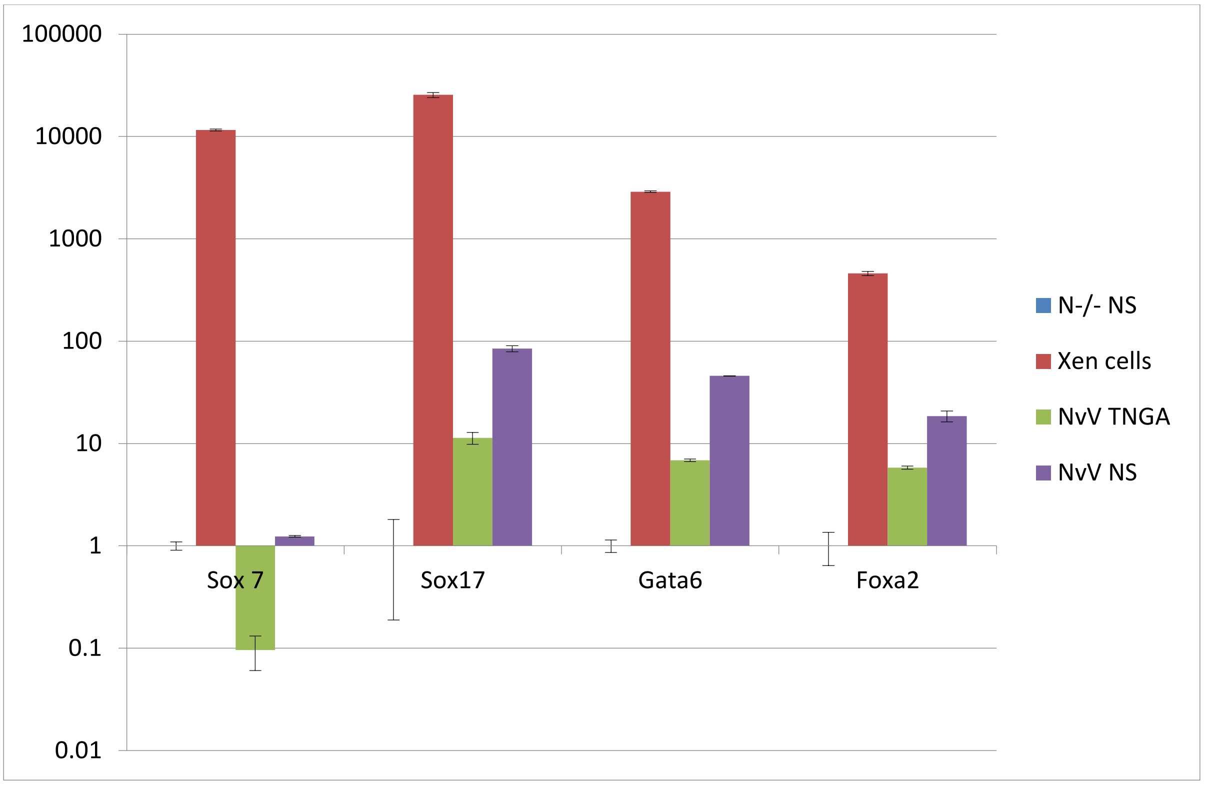


Figure 32: Genes expressed in XEN cells are upregulated in NvVent pre-iPS cell lines

Data from qPCR analysis in a logarithmic scale of 10 shows that extra-embryonic endoderm markers which are expressed in XEN cells, are upregulated in pre-iPS cells transfected with NvVent, relative to the untransfected pre-iPS cells. β-Actin was used as an endogenous control.

Sox7, Sox17, Gata6 and Foxa2 are expressed in XEN cell lines. As seen in figure 32, these genes are not expressed in the untransfected pre-iPS cells, therefore we analysed in the cells transfected with NvVent by qPCR to determine if those pre-iPS cells have acquired a gene expression profile complimentary to their phenotype which is concurrently XEN-like (Fig. 31). There is distinct upregulation of XEN cell markers in both pre-iPS cell lines transfected with NvVent (Fig. 32), relative to the untransfected pre-iPS cells. The relative fold change is relatively low when compared with the XEN cell culture, however cells expressing NV Vent display a similar pattern of expression, with Sox17 the most highly expressed marker. This XEN-like transcription profile is more prominent in the NS pre-iPS cell than in the TNGA MEFs. A recent study by Li et al. (2017) has demonstrated that iPSCs established by chemical induction pass through an extra-embryonic endoderm (XEN)-like state, and previous research by Parenti et al. (2016) described how the OSKM factors induce expression of endodermal genes, leading to formation of induced XEN (iXEN) cells, which display similar properties of blastocyst-derived XEN cells, including morphology, transcription profile, self-renewal, and multipotency. Parenti et al. (2016) show that iXEN cells arise in parallel to induced pluripotent stem cells, indicating that OSKM drive cells to two distinct cell fates during reprogramming. That the XEN-like cells are induced to such a prominent level in the NvVent cultures relative to the untransfected pre-iPS cells suggests that the sea anemone Vent gene is reprogramming the cells to an extent where they are then diverted to a XEN-like state. If so, this may be due to the inability of NvVent to activate Nanog and complete the transition to reprogramming. The XEN-like transcription profile is more prominent in the NS derived cells relative to the TNGA cells, this may be due to the more partially reprogrammed state of the NS pre-iPS cells, making them more likely to divert to an alternative lineage.

CHAPTER FOUR

4 RESULTS: SECTION TWO

My results suggest that Nanog-like reprogramming activity is found in the *Vent2* gene of the acorn worm; Saccoglossus kowalevskii, representative of basal deuterostomes (Cameron et al., 2000). The *Vent* gene found in the sea anemone Nematostella vectensis does not exhibit the same activity, suggesting that Nanog activity arose after the evolution of Cnidaria. However, the sea anemone Vent gene does encode some reprogramming activity, as its expression appears to divert pre-iPS cells to a XEN-like state. I therefore further investigated a functional comparison of the two sequences with the aim of gaining insight into how Nanog activity evolved.

4.1 SWAPPING THE HOMEODOMAINS

Theunissen et al. (2011) demonstrated that the reprogramming capacity of Nanog is conserved in the DNA-binding homeodomain. I therefore decided to test the invertebrate Vent genes to determine if the reprogramming activity of SaccoglossusVent2 and the XEN-like inducing ability of Nematostella Vent is conserved in the respective homeodomains. To assess whether the activity is consigned to the DNA-binding domain, I swapped the homeodomains from one CDS to another, and created hybrid plasmids containing the Saccoglossus Vent2 homeodomain with the N and C termini of the Nematostella Vent (SkVent2HD), and a second molecule consisting of the Nematostella Vent homeodomain plus the flanking regions of Saccoglossus Vent2 (NvVentHD). These constructs were cloned into a pLenti vector, and subsequently transduced into the nanog-null NS derived pre-iPS cells. After standard selection with HygromycinB, the cells were seeded for reprogramming.

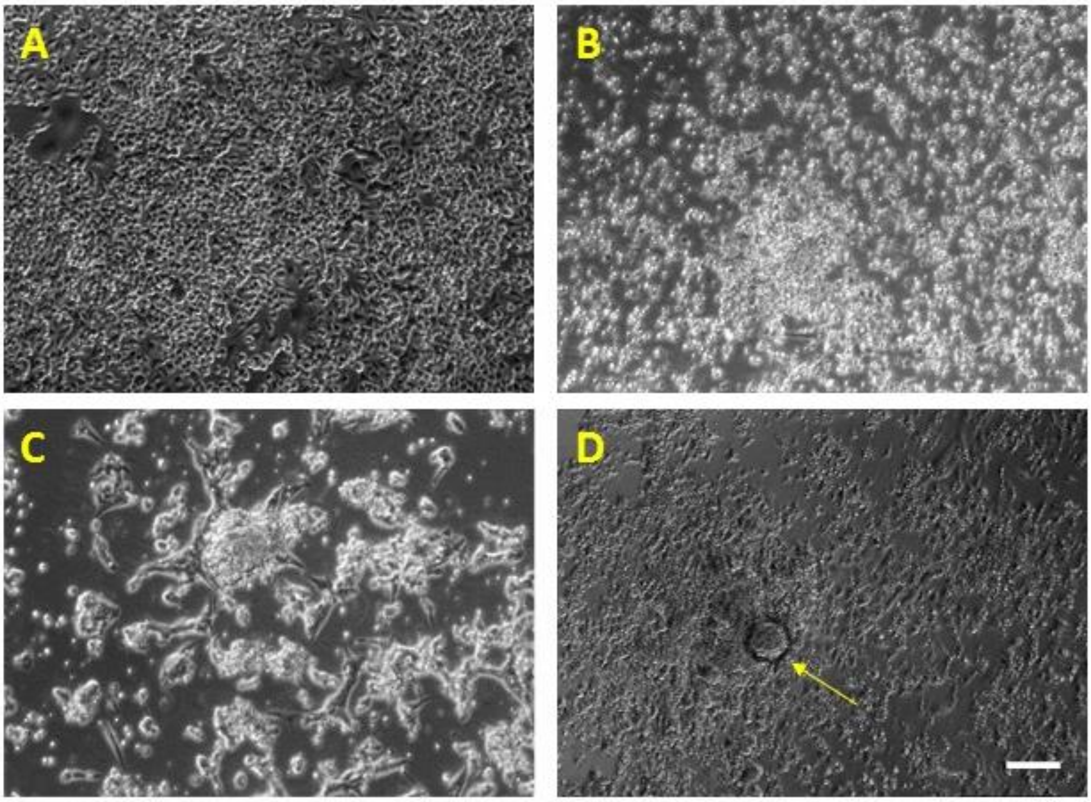


Figure 33: Invertebrate Vent activity regulated specifically by the homeodomain

(A) NvVentHD produces XEN-like cells in serum + LIF (B) NvVentHD cells after 6 days in 2i + LIF.

(C) SkVent2HD in serum LIF. (D) Colonies emerge in SkVent2HD cultures after 6 days in 2i + LIF. Scale bar represents 200um for A, B, D, and 100um for C.

The cells were seeded at 40K per well as previously optimised for the lines expressing the natural Saccoglossus Vent2. The NvVentHD cultures did not maintain a healthy appearance in the 2i + LIF conditions during the initial reprogramming experiment, but the less partially reprogrammed elongated pre-iPS cells persisted until the addition of G418 antibiotic, at which point the subsequently died. After several passages in serum + LIF conditions, the NvVentHD cultures began to display the same rounded cell type as previously observed in cultures expressing Nematostella Vent. This phenotype became more prominent as the line was expanded (Fig. 33; A), and switching the media to 2i + LIF now rapidly eradicated the NvVentHD cultures, which expired by the fourth day in 2i conditions (Fig. 33; B).

In contrast with this, the SKVent2HD line displayed no alteration in its pre-iPS phenotype when cultured in serum + LIF media, instead there were prominent pre-iPS colonies observed in the cultures (Fig. 33C). Switching to 2i + LIF conditions to mediate reprogramming subsequently generated small iPSC like colonies after approximately three to four days, which remained visible within the culture to the point of selection with G418 at day six. By day ten, the SkVent2HD colonies had survived, suggesting that the NEO gene in the endogenous Nanog locus had been activated. The iPSC colonies were stained for both LacZ and AP as previously described to verify the acquisition of pluripotency (Fig. 34).

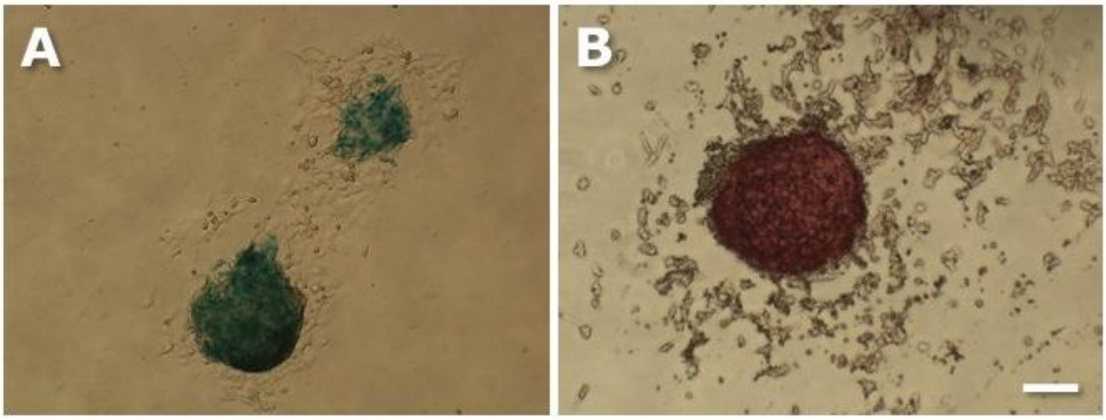


Figure 34: SkVent2HD LacZ and AP staining

(A) LacZ staining of cell reprogrammed with SkVent2HD. (B) AP staining of SkVent2 iPSCs. Scale bar represents 100 um.

Positive staining for LacZ (Fig. 34A) concurred with the observed resistance to G418, and confirmed activation of the endogenous Nanog locus. Positive AP staining (Fig. 34B) verified the acquisition of a pluripotent state. This indicates that the Nanog activity exhibited by Saccoglossus Vent2 in reprogramming cells to pluripotency is conserved in the homeodomain of the more basal Vent gene. Three wells of a six-well plate were stained for either LacZ or AP, and the positively staining colonies were counted to calculate the efficiency of reprogramming (Table 5).

Cell Line	LacZ+			AP+			Total Colonies	Average per well	Average Colonies per 10^4 seeded
SKVent2HD	4	3	5	6	5	4	27	4.5	**1.125**

Table 5: reprogramming efficiency of Saccoglossus Vent2 containing Nematostella flanking regions.

The capacity the Saccoglossus Vent to reprogram the NS pre-iPS cells has been shown to be inefficient when compared with mouse Nanog (Table 3). However, the average of 1.125 colonies generated by the SkV2HD gene per 10K cells seeded is comparable to the average 1.63 obtained through reprogramming with the complete CDS. This suggests that the capacity to reprogram is not affected by the presence of the sea anemone flanking regions. The SkVent2HD iPSCs were expanded for several passages, and RNA was extracted for qPCR analysis (Fig. 35).

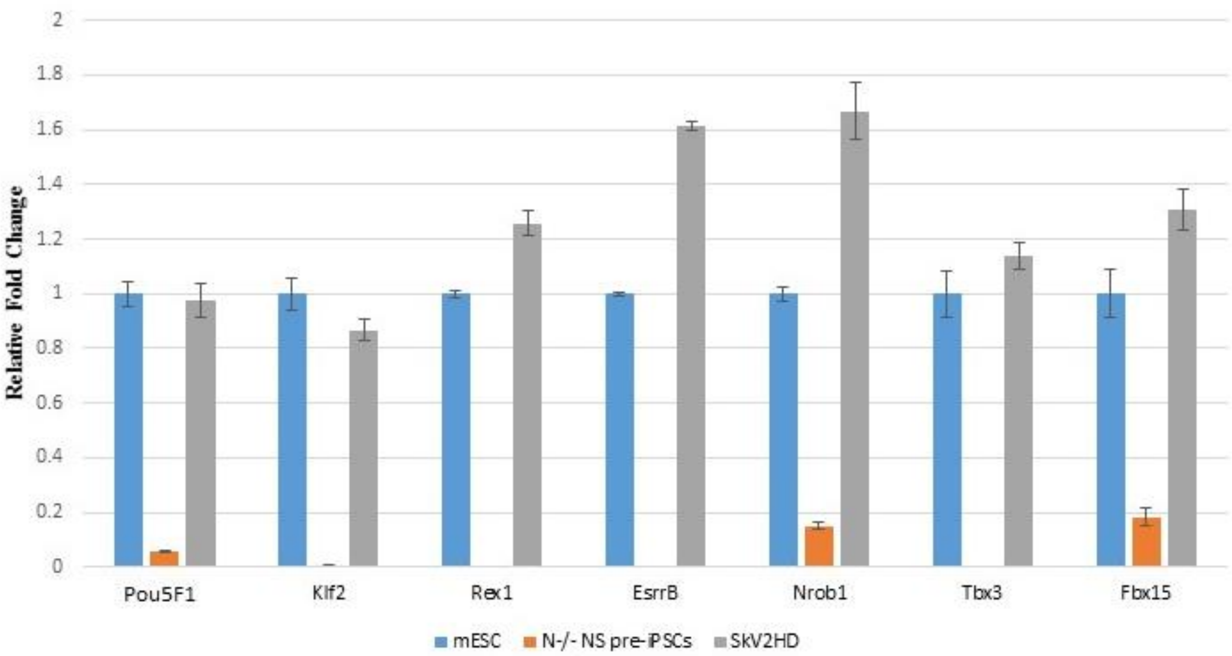


Figure 35: The homeodomain of SKVent2 can activate endogenous pluripotency genes

qPCR analysis demonstrated complete reprogramming of the NS pre-iPS cells by the SKVent2HD hybrid. Data is standardised to mESCs, B-Actin was used as the endogenous control.

The SKVent2HD iPSCs were analysed for activation of endogenous pluripotency genes to confirm acquisition of naïve pluripotency. The data was again normalised to wildtype mESCs, and shows that SkVent2HD reprogrammed cells exhibited activation of all qPCR targets. Interestingly, the levels of direct Nanog targets EsrrB and Rex1 were observably higher in the SkVent2HD iPSCs relative to the ESCs, as consistently observed previously with the natural Saccoglossus Vent2 sequence (Fig. 22, Fig. 30).

4.2 Modified Sea anemone Vent gene

The reprogramming capacity of Nanog appears to be conserved in the Vent homeodomains at the base of deuterostome evolution. The Vent gene found in the freshwater sponge Ephydatia fluviatilis demonstrated no activity in mammalian cells, but the sea anemone Vent repeatedly exhibited an ability to induce XEN-like cells from pre-iPS lines. However, neither gene could facilitate the transition to pluripotency, although this activity is found in an acorn worm Vent genes from, which is a hemichordate. These results are consistent with a model in which the emergence of Nanog activity coincides with the evolution of mesoderm. I therefore compared the amino acid sequences of the Vent homeodomains in an attempt to identify specific changes that may have facilitated the acquisition of this Nanog activity.

4.2.1 Amino Acid swap

The amino acid (AA) sequence of different vertebrate Nanog and Vent homeodomains (HDs), were aligned with the invertebrate Vents (Figure 36).

Figure 36: Alignment of the Homeodomain Amino Acid sequences of Nanogs and Vents

Homeodomains were aligned to identify Amino Acids that differ between the Nematostella and Saccoglossus Vent genes.

As the evolution of Nanog activity occurred between Nematostella and Saccoglossus, I identified three AAs whose structure would be most likely to change the binding activity of each HD region (Fig. 37). They were identified based on differential charge or polarity of the amino acids in question. The amino acids identified in the Nematostella HD were then changed to the corresponding AAs in the Saccoglossus Vent2 gene.

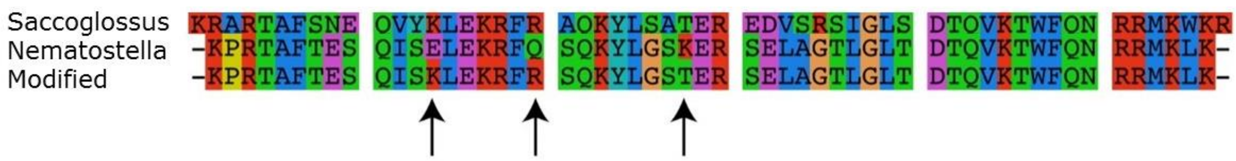


Figure 37: Alignment displaying Homeodomain AAs which were swapped in NvVent to create the modified Nematostella Vent

Three Nematostella Vent AAs were swapped to their corresponding AAs in Saccoglossus Vent.

From the amino terminus, the first change was Glutamic acid (E) to Lysine (K), which changed the charge from negative to a positive. The second change was from Glutamine (Q) to Argenine (R) which swapped a neutral polar AA for a positively charged AA. The third change was from Lysine (K) to threonine (T) which replaced the positive charge of Lysine with the uncharged polar group of Threonine (Fig. 37, Fig. 38).

```
           301
           |
NvVent:    TCCGAACTCG AGAAACGCTT CCAGTCACAG AAATACCTCG GGAGCAAGGA ACGGAGTGA
Modified:  TCCAAACTCG AGAAACGCTT CAGATCACAG AAATACCTCG GGAGCACAGA ACGGAGTGA
```

Figure 38: Nucleotide alterations in the homeodomain of Nematostella Vent to create the synthetic Modified NvVent.

Altered nucleotides are highlighted in red.

Once the sequence had been designed (Fig. 38), a synthetic molecule was order from Integrated DNA Technologies (IDT), and cloned into the pLenti plasmid containing a PGK promoter and Hygro resistance. The gene will subsequently be referred to as the modified Nematostella Vent (ModNvVent).

4.2.2 Reprogramming with Modified Sea anemone Vent

A lentiviral media was produced containing the ModNvVent construct and it was transduced in to the Nanog-null pre-iPS cells, and selected for with HygromycinB under standard conditions. Several passages after transduction the cultures began to display the rounded and refractile phenotype associated with the pre-iPS cell lines expressing the natural sea anemone Vent gene. However, the ModNvVent exhibited this to a greatly reduced extent (Fig. 39A). MEK inhibitor at a final concentration of 1uM was added to some of the serum + LIF cultures (Fig. 39B), and over the course of 48hours it removed the majority of the rounded XEN-like phenotype, similar to earlier results (Fig. 21D). However, unlike the cultures expressing the natural sea anemone Vent, the ModNvVent culture retained a larger portion of its population due to the reduced number of cells displaying either a XEN-like, or partially reprogrammed, morphology (Fig. 39; B). Cultures of the ModNvVent pre-iPS cells were maintained in serum + LIF for several passages before seeding for reprogramming. The initial seeding density was selected at 40K cells per well of a 6 well plate as this had proven optimum for the acorn worm Vent. After two days in serum + LIF, the media was switched to 2i + LIF to initiate reprogramming. However, after just three days in 2i + LIF, the cultures began to exhibit signs of stress, and a large portion of the cells had quickly died off. I assume this was due to the inhibition of ERK signalling, as described previously. Therefore an initial seeding density of

80K cells per well was deemed more appropriate to account for the large initial die-off of cells at the switch to 2i + LIF.

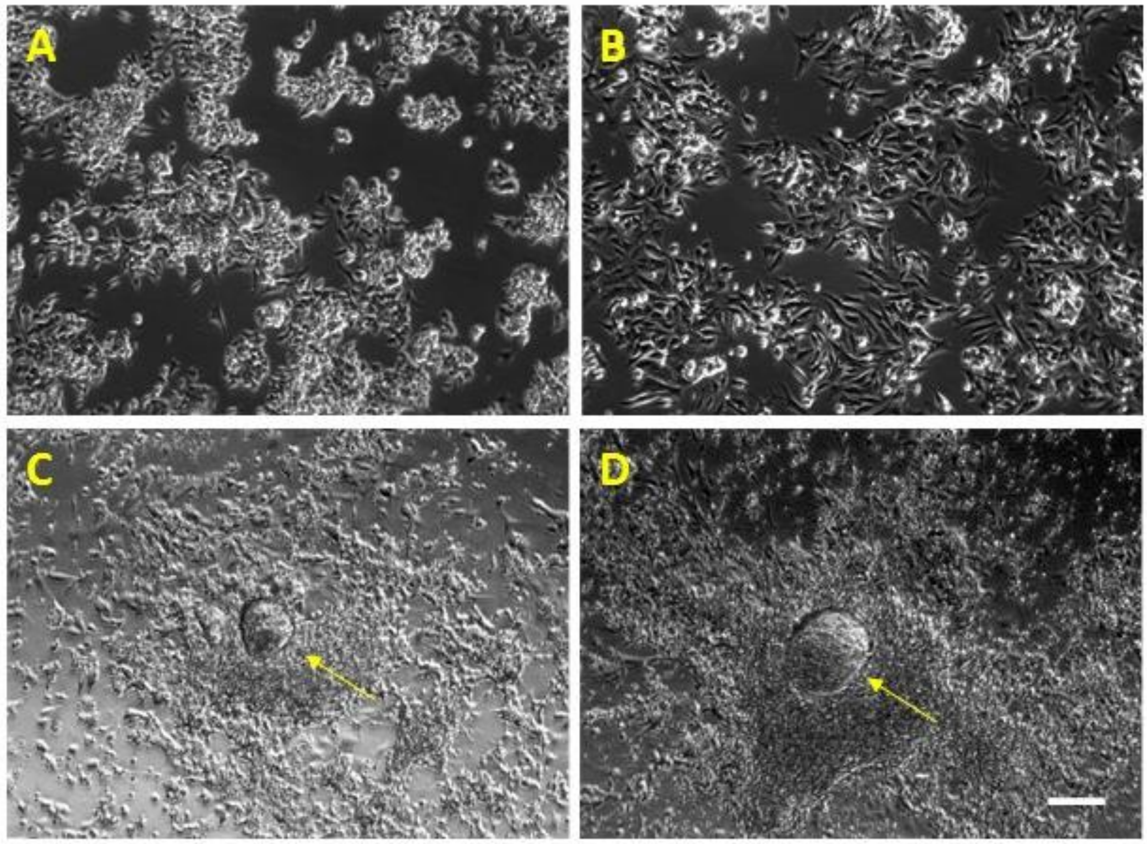


Figure 39: The modified NvVent in culture, and emergence of iPSC colonies

(A) The ModNvVent pre-iPS cells display a reduced number of rounded cells relative to NvVent (Seen in Fig. 19). (B) Addition of 1uM MEK inhibitor to the media removes some, but not all cells.

(C) Microscopic colonies emerge after 4 days in 2i + LIF. (D) Colonies exhibit ESC like appearance after 7 days in 2i + LIF. Scale bar represents 200um.

After four days in 2i + LIF, microscopic colonies could be observed in the cultures (Fig. 39C). However, this was at a minimum efficiency of one colony per well for most cultures. The colonies were maintained through the addition of G418 to the media to select for fully reprogrammed cells. By day seven in 2i + LIF the colonies had acquired a defined and compact ESC-like structure associated with true iPSCs (Fig. 39D).The antibiotic killed all but the reprogrammed cells, and by day ten the small iPSC colonies remained. However, due to the low efficiency of the reprogramming, the colonies were manually picked and collected with a pipette tip using a microscope under sterile conditions. Colionies were disassociated with trypsin-EDTA, and transferred to a new six-well plate and expanded on gelatin. After two passages, the cells were stained for LacZ and AP to confirm activation of the endogenous Nanog and acquisition of pluripotency (Fig. 40).

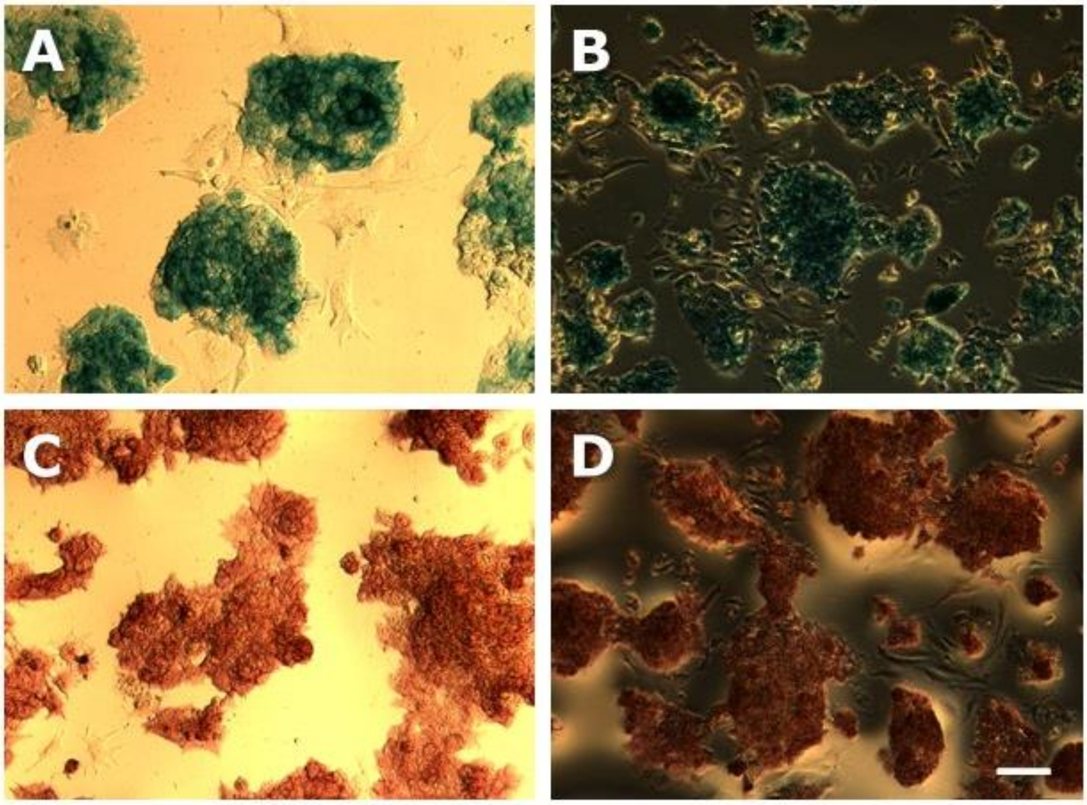


Figure 40: LacZ and AP staining of the ModNvVent iPSC colonies

(A) Positive staining for LacZ indicates that the endogenous Nanog was activated. (B) LacZ staining shown to be isolated to the centre of colonies displaying jagged morphology.

(C) Positive staining for AP indicates that the cells are pluripotent. (D) AP is seen to be more homogenously expressed than the LacZ (B). Scale bar represents 100um for A and C. Scale bar represents 200um for B and D.

The ModNvVent iPSCs stained positively for both LacZ (Fig. 40A, B), and AP (Fig. 40C, D), demonstrating that the cells were reprogrammed. However, the colonies displayed an irregular phenotype in that they maintained a jagged edge around the periphery and appeared more flat than standard ESC or iPSC colonies (Fig. 41B). This contrasts with the colonies observed in iPSC cultures reprogrammed by mNanog or Skvent2. Smaller colonies were prone to differentiation (Fig 41A), but homogeneity could be maintained through the addition of G418 to the culture media (Fig. 41B).

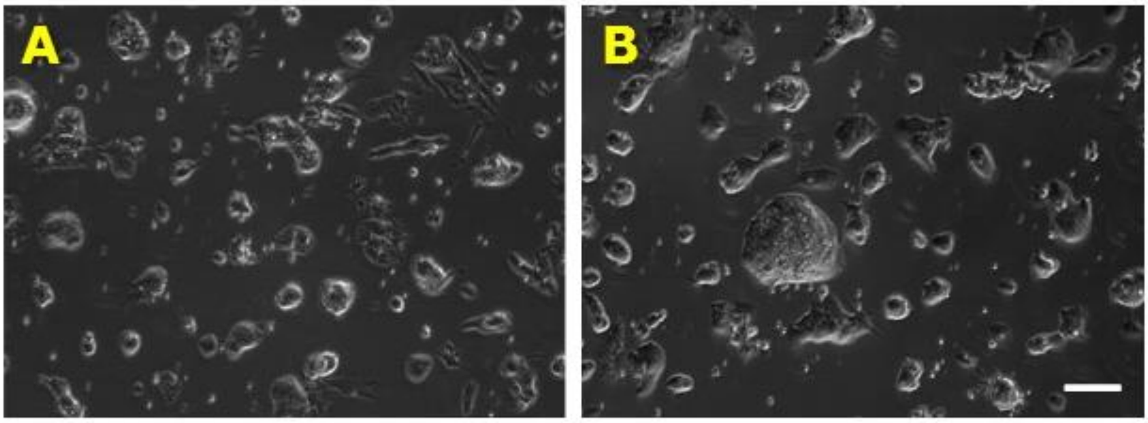


Figure 41: ModNvVent iPSCs maintained unusual flat and jagged phenotype.

(A) The Smaller colonies of ModNvVent iPSCs were prone to differentiation. (B) The line maintained an unusual jagged phenotype, but the culture could be maintained as a homogenous population of pluripotent cells with addition of G418. Scale bar represents 200um.

The cells propagated at a reduced rate relative to wildtype ESCs, but were successfully expanded, and a line was established. After six passages, a sample was collected and processed for RNA. CDNA was synthesised and the sample was analysed by qPCR for activation of a set of endogenous pluripotency factors (Fig. 42).

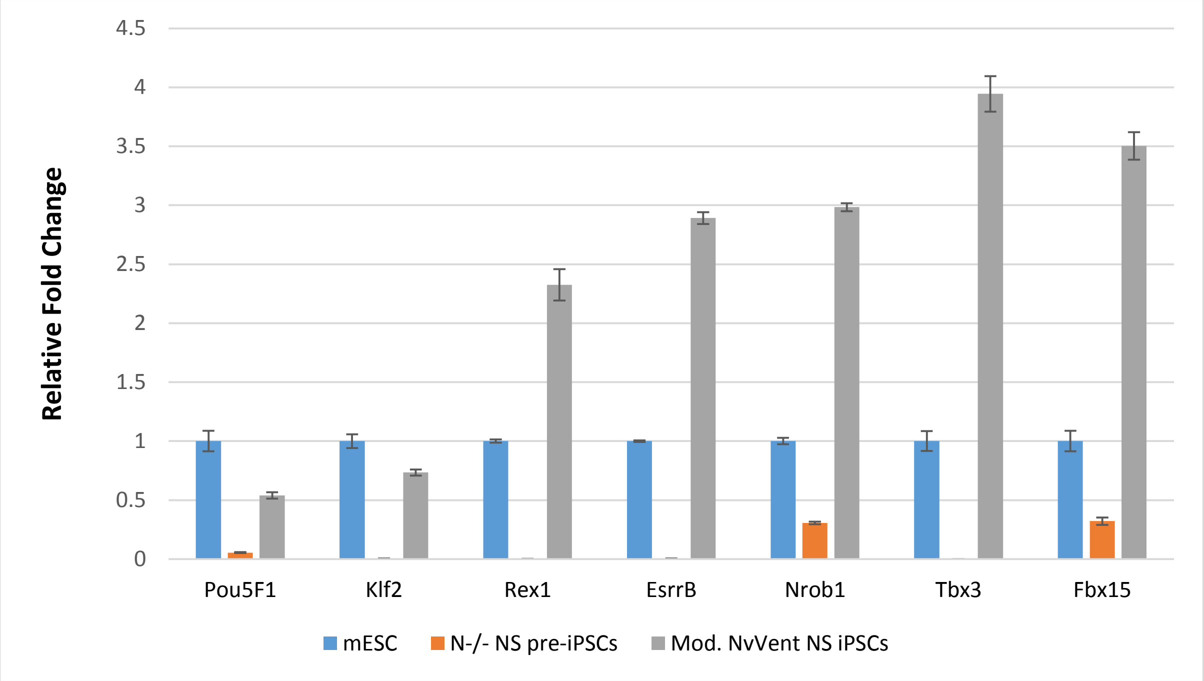


Figure 42: Activation of pluripotency factors by the Modified sea anemone Vent

The ModNvVent iPSCs were assayed for expression of endogenous factors relative to wildtype ESCs. B-Actin was used as an endogenous control.

The synthetic modified sea anemone Vent gene was observed to activate the native pluripotency factors targeted by the qPCR. The native Nanog targets EsrrB and Tbx3 were again upregulated to a large extent as seen previously with cells reprogrammed with acorn worm Vent. The naïve marker Nrob1, and the stem cell specific Fbx15 were both highly upregulated. This indicates that the three amino acids swapped into the sea anemone Vent2 were sufficient to alter its binding capacity and confer a reprogramming activity resembling with Nanog. These data suggest that Nanog activity had evolved in ancient Vent genes that precede the emergence of vertebrates. I postulated that they may form components of an ancestral gene regulatory network (GRN) that existed prior to the evolution of vertebrates.

4.3 An EsrrB orthologue in Sea anemone

Nanog is not present in the genomes of invertebrates; however Nanog activity appears to be conserved in the Vent gene family of basal deuterostomes. Vent genes which were identified in more primitive organisms, such as Sea anemone and Sponge, are unable to facilitate reprogramming to pluripotency. I searched for orthologues of other known pluripotency factors, and identified a gene related to Estrogen-related receptor B (EsrrB) from the sea anemone Nematostella vectensis based on homology with the vertebrate EsrrB family. EsrrB is a direct target of Nanog in vertebrates, and has been shown to be capable of substituting for Nanog activity in pluripotent cells (Festuccia et al., 2012). Nanog has been shown to directly bind to EsrrB in mouse ESCs, and subsequently stimulate EsrrB transcription. Festuccia et al. (2012) also demonstrated that forced EsrrB expression could confer cytokine-independent self-renewal on mESCs, and reprogram Nanog-null EpiSCs to the naïve pluripotency. Feng et al. (2009) also reported that EsrrB could promote the reprogramming of MEFs to pluripotency.

More strikingly, Festuccia et al. (2012) demonstrated that induction of an exogenous EsrrB transgene by doxycycline resulted in rapid activation of the endogenous Nanog locus in the Nanog-null NS derived pre-iPS cell line, and that the resulting colonies could be picked and expanded. This proved that the downstream Nanog target EsrrB could substitute for Nanog in the acquisition of pluripotency.

I therefore decided to test the sea anemone EsrrB (NvEsrrB) gene in the same capacity. A synthetic molecule containing the NvEsrrB was ordered from IDT, and cloned into the pLenti expression vector. It was then transduced into the Nanog-null pre-iPS cells, and seeded for reprogramming (Fig. 43).

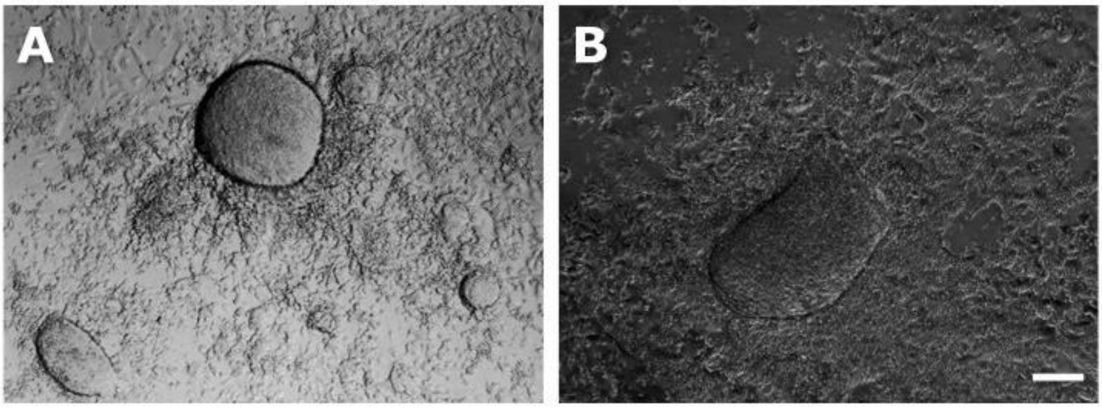


Figure 43: NvEsrrB colonies emerging during reprogramming in 2i + LIF

(A) Colonies by day 4. iPSCs emerge in the NvEsrrB pre-iPS cell cultures. (B) Few colonies were generated, but they appeared early, and subsequently grew to a large extent.

The NvEsrrB pre-iPS cells were seeded at 40K cells/well on a six well plate in serum + LIF conditions on gelatin, and media was switched to 2i + LIf after two days. Interestingly, large colonies displaying a defined edge and circular morphology appeared in the cultures early in reprogramming (Fig. 43A). However, only 7 colonies appeared on the plate prior to addition of G418 at day six in 2i + LIF. The colonies were manually picked by pipette tip, dissociated with Trypsin-EDTA, and transferred to a single well of a twelve-well plate to maintain a higher confluency.

The cells struggled to expand from a low density (Fig. 44A), and displayed an irregular phenotype of flat, jagged colonies, similar to the ModNvVent iPSCs. However, unlike the iPSCs generated by the modified NvVent, the majority of colonies within the NvEsrrB cultures failed to expand completely after a few days in culture. The resulting micro-colonies differentiated or exhibited signs of apoptosis, such as the appearance of large vacuoles. Not all colonies expired however, and small colonies maintained in culture as they were, absent any sign of proliferation (Fig. 44B, D).

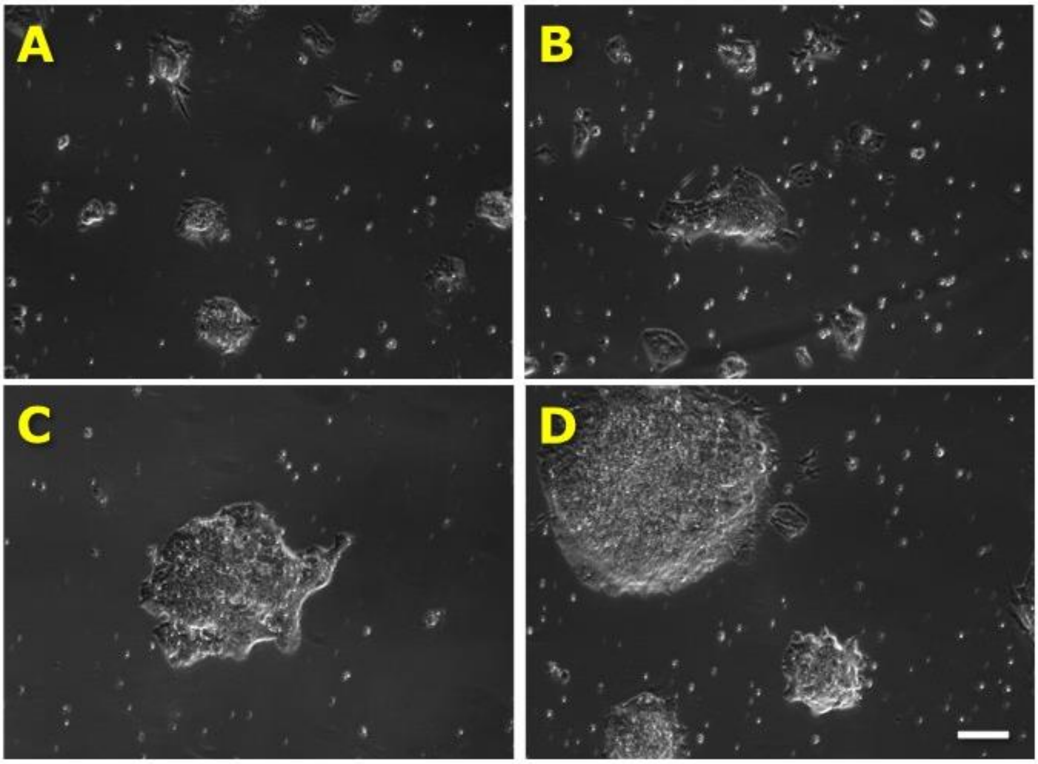


Figure 44: NvEsrrB colonies 4 days and 16 days after passage

(A) NvEsrrB iPSCs 4 days after passaging were expanded on gelatin. (B) Colonies displayed irregular shapes and expanded slowly. (C) NvEsrrB iPSCs after 16 days in culture. The colonies maintained irregular morphology upon expansion. (D) NvEsrrB iPSCs exhibited very low proliferation rate with some smaller colonies failing to expand relative to others. Scale bar represents 200um.

Not all colonies failed to propagate, and some NvEsrrB colonies expanded to form large irregular structures (Fig. 44; C). Both the surviving smaller colonies and larger expanding colonies displayed a uniform centre which more closely resembled ESC colonies of cells.

The proliferation rate of the cells remained consistently low, and they were passaged approximately every twenty days. After 16 days in culture after passaging, the cells were stained for LacZ and AP to confirm acquisition of pluripotency (Fig. 45). Interestingly, the both the smaller stalled colonies, and larger expanding structures similarly stained positive for both AP (Fig. 45A) and LacZ (Fig. 45C), indicating activation of the Nanog locus.

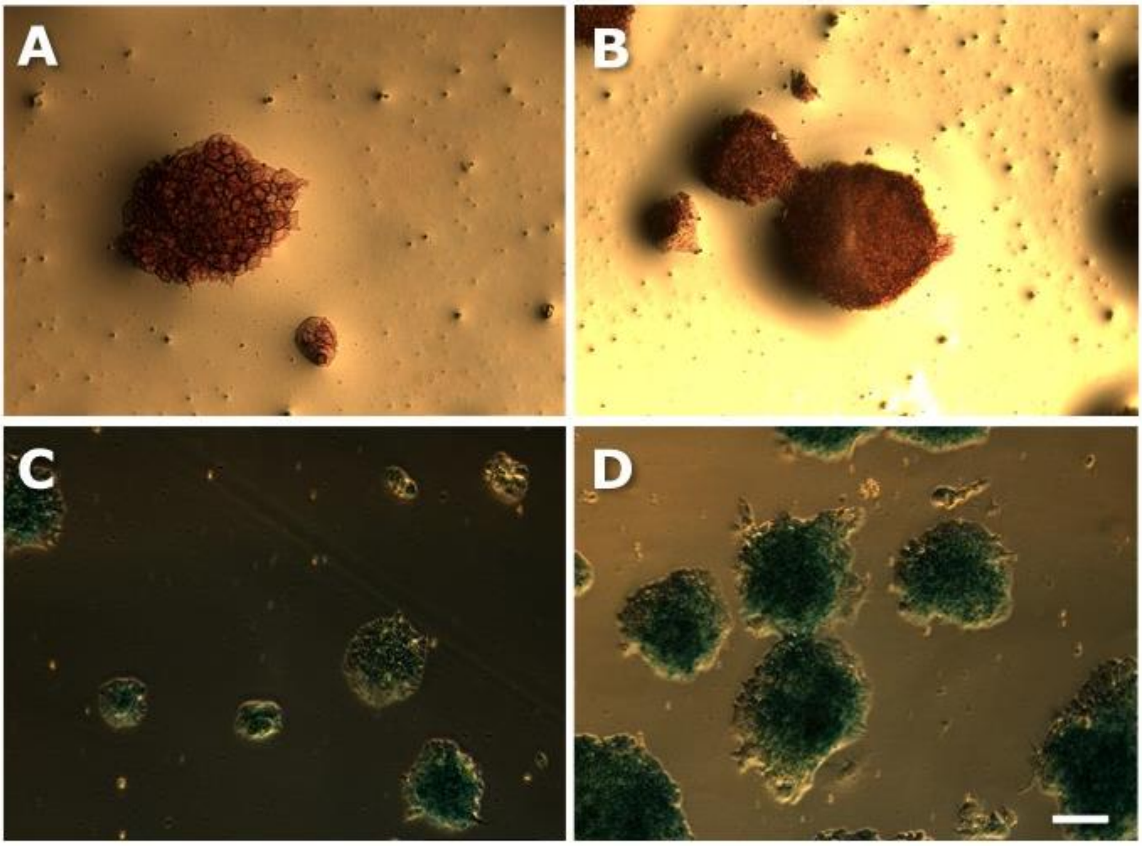


Figure 45: NvEsrrB AP and LacZ 16 days after passage

(A) NvEsrrB iPSCs stained positive for AP. (B) AP staining was homogenous across colonies. (C) Small colonies which failed to proliferate stained positive for LacZ. (D) LacZ staining was exclusive to the centre of the iPSC colonies. Scale bar represents 200um.

Similar to the ModNvVent iPSCs that stained positive for AP and LacZ (Fig. 40), the NvEsrrB colonies displayed a homogenous staining for AP (Fig. 45; B), but blue LacZ staining was exclusive to the centre of the colonies (Fig. 45; D), indicative of the endogenous Nanog locus failing to remain active in the peripheral cells. This was also evident by the propensity of differentiated cells appearing at the edges of colonies to expand when maintained in culture for multiple passages. However, homogeneity could be restored by addition of G418 at the maintenance concentration of 200ug/ml. A culture homogenous for undifferentiated cells was maintained to high confluency, and collected for processing for RNA extraction and qPCR analysis (Fig. 46).

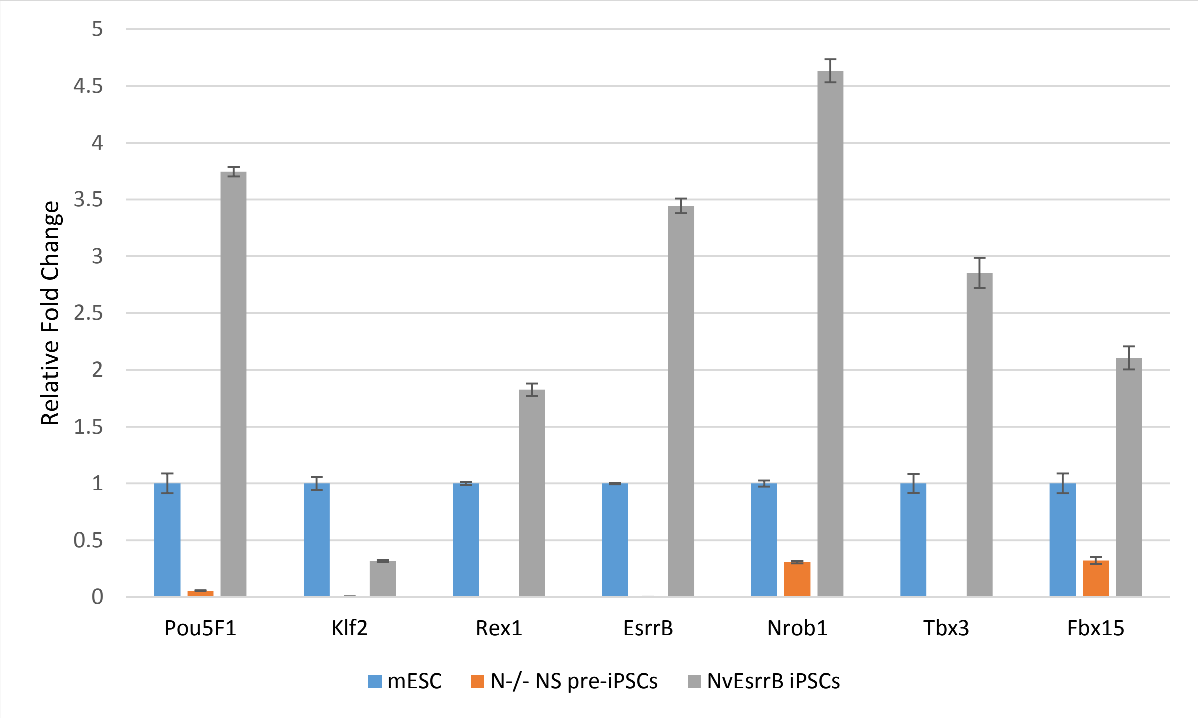


Figure 46: NvEsrrB is sufficient to activate endogenous pluripotency genes.

NvEsrrB iPSCs showed activation of endogenous pluripotency factors at a high level relative to mESCs. B-Actin was used as an endogenous control.

The sea anemone EsrrB-like gene was sufficient to activate endogenous pluripotency genes to a relatively high level, including the endogenous mouse EsrrB gene. Interestingly, the endogenous Oct4 was expressed approximately 3.75 fold higher in NvEsrrB iPSCs than in wildtype mouse ESCs. Van den Berg et al. (2008) reported that Esrrb binds directly to Oct4 and the *Nanog* proximal promoter, and it is possible that NvEsrrB may be fulfilling a similar role in these cells. That Nrob1, Tbx3 and Fbx15 are all highly upregulated indicates that these cells have been completely reprogrammed to pluripotency. For both the Modified NvVent and NvEsrrB lines, the nature of their states and the extent to which they are reprogrammed could be further elucidated by more functional assays.

That the sea anemone Nematostella Vectensis contains a Vent gene which can be modified with three amino acids to acquire Nanog activity, and an EsrrB orthologue which can substitute for Nanog in mammalian cells similar to a vertebrate EsrrB, suggests that the GRN which governs pluripotency arose from a pre-existing GRN that preceded the evolution of both vertebrates and basal deuterostomes. Johnson and colleagues have shown that Nanog, Vent and EsrrB are all members of the GRN which governs germ cell specification in an axolotl (Unpublished). Ground state pluripotency is conserved along the major trunk of vertebrate evolution (Johnson et al. 2011), suggesting that it is possible that the GRN which governs pluripotency evolved from a pre-existing GRN which governs germ cell specification.

CHAPTER FIVE

5 DISCUSSION

5.1 NANOG ACTIVITY IS CONSERVED

The results of this study support the hypothesis that Nanog activity first emerged within in the invertebrate Vent gene family. Initial reprogramming experiments with lentiviral constructs, which contained a WR domain, demonstrated that this was the most efficient vehicle for introducing candidate genes to reprogram Nanog-null pre-iPS cells (Table 2), and so this system was used in all preliminary studies. This system provided evidence that the Saccoglossus Vent2 gene encodes reprogramming activity related to that of vertebrate orthologs of Nanog, but to a lesser final efficiency in terms of colonies generated. The second axolotl Nonog ortholog, called Eggnog, gene exhibited a similar level of activity to axNanog (Dixon et al. 2010) in converting the NS pre-iPS cells to the ground state; however, their efficiency was much less than that of mouse Nanog. All vertebrate orthologs of Vent failed to generate any iPSC colonies, suggesting in vertebrates reprogramming activity is exclusive to Nanog genes.

Of the invertebrate Vent genes assayed, only expression of Saccoglossus Vent2 facilitated the transition to pluripotency. Orthologs found in Branchiostoma floridae (Amphioxus), Nematostella vectensis (sea anemone) and Ephydatia fluviatilis (Sponge) all failed to generate colonies, or reprogram the cells to a point where the endogenous Nanog gene was activated. Interestingly, the other otholog found in Saccoglossus, the Vent gene1, exhibits no reprogramming activity. This aligns with previous unpublished data generated by the Johnson lab, where Saccoglossus Vent2, but not Vent1, conferred cytokine independent self-renewal on mouse ESCs when combined with the mouse WR domain (Fig. 6). This suggests a divergence in function of the Vent orthologs following a genome duplication event, which occurred after the deuterostome–protostome split (Friedman and Hughes, 2001).

Saccoglossus Vent2 activated the endogenous Nanog locus (Fig.13), and subsequently conferred resistance to G418 on the reprogrammed cells, which facilitated selection and the establishment of an iPSC line. Alkaline phosphatase (AP) staining, and subsequent qPCR analysis confirmed that the Saccoglossus Vent2 gene was sufficient to activate endogenous pluripotency markers, providing further evidence that the ability to orchestrate the higher mammalian pluripotency network is conserved in a Vent gene from hemichordates. Activation of the endogenous Pou5f1, and markers of naïve pluripotency such as Klf2 and Nrob1, verify the acquisition of ground state potential. Activation of direct Nanog targets such as Rex1 and EsrrB, the latter of which was highly upregulated relative to wildtype ESCs, suggests that the activity of Saccoglossus Vent2 is similar to that observed with mouse Nanog. Contrasting with this however, cells reprogrammed by Vent2 exhibited a different phenotype than iPSCs induced by mouse Nanog. They had a greater propensity to differentiate than the cell line reprogrammed by mouse Nanog. They consistently gave rise to interconnected webs of elongated cells, which expanded exponentially throughout the culture (Fig. 15C), while simultaneously maintaining a compact population of ESC-like cells in central colonies. This property was exclusive to iPSCs reprogramed with the Saccoglossus gene, making them dependent on G418 for maintenance of homogenous ESC-like cultures (Fig. 15D). This level of differentiation suggests that Saccoglossus Vent2 exerts a more loosely maintained governance of the native pluripotency network in the cells than vertebrate Nanogs, and may allude to a difference in binding targets within the wider network. Nanog protein acts as a hub in pluripotency, mediating protein-protein interactions by interacting with multiple molecules simultaneously (Mullin and Chambers, 2011). It is possible that Saccoglossus Vent2 forms a different hub, which mediates interactions with additional factors outside of the pluripotency GRN, which promote differentiation.

Further analysis of the reprogramming activity of Saccoglossus Vent2 demonstrated that its native monomeric form is also sufficient to transition the NS pre-iPS cells to pluripotency. Colonies staining positive for both LacZ and AP (Fig. 18) were generated by transposase mediated insertion of a PiggyBac vector expressing Vent2. The lack of any reprogramming activity exhibited by the Xenopus Vent2 gene verified the hypothesis that Nanog activity is not conserved in the vertebrate orthologs of Vent. The assay utilising PiggyBac constructs displayed an all-round lower efficiency of reprogramming than the lentiviral constructs, suggesting that the higher level of transgene expression associated with lentiviral transduction initiated reprogramming in a larger population of the seeded cells. However, a further reduction in the efficiency of Saccoglossus Vent2 relative to mouse Nanog may directly result from the lower confluency attained by the culture over the ten day assay in 2i + LIF conditions. Mouse Nanog in its native form naturally contains a WR domain, which has previously been shown to increase cell proliferation (Dixon et al. 2010). The reduced proliferation rate of the monomeric line compared with the Saccoglossus Vent2 cells, which were transduced with the WR domain, resulted in cultures which were observed to be less confluent at the later stages of reprogramming, and may have contributed to a lower overall efficiency. To further elucidate if the WR domain plays a significant role in the efficiency of reprogramming with Saccoglossus Vent2, it will be compared to an altered monomeric mouse Nanog molecule in future work. However, the Saccoglossus Vent2 gene sufficiently activated all the native pluripotency markers whose expression was assayed by qPCR. The level to which the resident pluripotency factors were expressed was reduced in the cells transfected with PiggyBac relative to lentiviral reprogrammed cells, which may also reflect a reduced activity due to the lower level of expression of the transgenes. This may be responsible for the reduced efficiency of reprogramming the pre-iPS cells to naïve pluripotency. The IPSCs generated by the monomeric Saccoglossus Vent2 gene displayed the same propensity to differentiate as those previously reprogrammed with the vector expressing the WR domain. G418 antibiotic was added to the culture in alternating passages, and homogeneous ESC-like cultures were maintained for analysis. The qPCR data also indicated that the exogenous retroviral Yamanaka factors were silenced (Fig. 23), indicating that Saccoglossus Vent2 has fully reprogrammed the cells, as complete transition to a naïve ESC-like state is required to silence retroviral transgenes through epigenetic remodelling (Hotta and Ellis, 2008; Buganim et al, 2012).

To further characterise the iPSCs generated by Saccoglossus Vent2, and verify that they had successfully attained ground state potential, I produced EBs which were allowed to spontaneously differentiate to lineages of all three primary germ layers. These were compared directly with EBs from mESCs and iPSCs reprogrammed with mNanog. Both the mouse Nanog and the Saccoglossus EBs showed high upregulation of markers representative of all three germ layers, indicating differentiation and verifying that cells had indeed been reprogrammed to pluripotency (Fig. 25). This provides support to the reprogramming data, as well as evidence that the Vent2 gene found in the hemichordate Saccoglossus can transition pre-iPS cells to a naïve state with developmental potential comparable to both iPSCs reprogrammed with mouse Nanog or wildtype ESCs.

The Nanog-null neural derived pre-iPS cells exist in an artificially induced, partially reprogrammed state, already teetering on the edge of conversion to pluripotency, with only the absence of Nanog, or a gene with equivalent function, preventing their transition to ground state. To elucidate the full extent of the reprogramming activity encoded by Saccoglossus Vent2, I transfected it into the TNGA MEF pre-iPSC line, which are representative of a primary somatic cell lineage. This allowed me to simultaneously test the effects that activation of a functional endogenous Nanog gene would have on the iPSCs produced, which might potentially counteract the instability exhibited by lines generated with the invertebrate Vent gene. Transfection of mouse Nanog or Saccoglossus Vent2 into the TNGA cells instigated microscopic colony formation from a low seeding density in their respective cultures (Fig. 26). Cultures expressing mouse Nanog exhibited more reprogramming, with a larger number of GFP-negative micro-colonies emerging, and these pre-iPS cell structures were larger than those induced by Saccoglossus Vent2 expression (Fig. 26 B, C). None of the cultures expressing Xenopus Vent2 generated any pre-iPS cell colonies with the same morphology (Fig. 27). Activation of the GFP-PURO reporter in the primary cells by both mouse Nanog and Saccoglossus Vent2, provided more conclusive evidence that Saccoglossus Vent2 encodes Nanog-like reprogramming activity. Cells expressing Saccoglossus Vent2 were seeded at twice the initial density of mouse Nanog, yet generated only 15.7% as many GFP-positive colonies. Interestingly, the TNGA cells reprogrammed by Saccoglossus Vent2 do not exhibit the same propensity to differentiate as Nanog-null cells (Fig. 29). This suggests that the translation of endogenous Nanog protein counteracts the instability demonstrated by Nanog-null iPSCs. However, although the cultures appear more stable than cells which are null for Nanog, they do exhibit a low level of differentiation at the periphery of IPSC colonies in 2i + LIF, which was not observed in the mouse Nanog cultures. This suggests that the pluripotent state is unstable in a sub-population of these cells, compromised by either the expression of Saccoglossus Vent2, or a low level of endogenous Nanog translated from the single allele remaining intact. The possibility that expression of Saccoglossus Vent2 actively drives the cells to differentiate would further distinguish it from Nanog, as the dual ability to reprogram cells and direct them to a somatic lineage suggests targeting of a network other than the pluripotency GRN. Further evidence of the reprogramming activity encoded by Saccoglossus Vent2 is the high level that endogenous pluripotency factors were upregulated relative to both ESCs and mouse Nanog iPSCs. Both endogenous Oct4 and Fbx15, and notably the direct Nanog target EsrrB, were all upregulated to more than twice the relative expression level in Saccoglossus iPSCs than in cells reprogrammed by mouse Nanog. This suggests that Saccoglossus Vent2 is directly targeting, and upregulating, the downstream targets of Nanog in mammalian cells. The resulting iPSC phenotype actively differentiates, and may be due to Vent2 also interacting with targets that reside outside of the Nanog interactome.

The Vent2 gene of Saccoglossus was the only invertebrate Vent gene to facilitate induction of pluripotency in mammalian pre-iPS cells. The single Vent gene from Nematostella Vectensis, a cnidarian, was unable to activate the endogenous Nanog locus or establish pluripotency in the pre-iPS cells tested. It did, however, induce a different XEN-like phenotype in all the pre-iPS cell lines expressing Nematostella Vent (Fig. 20, 27). These rounded cells were also briefly observed, albeit in lower proportion, in the Saccoglossus Vent2 iPSC cultures (Fig. 19C, D), suggesting that the activity is conserved in Vent, but diminished through the evolution of deuterostomes. The data from qPCR analysis (Fig. 32), showed that the same genes that are highly expressed in extraembryonic endoderm, of which XEN cells are representative (Kunath et al., 2005), were activated in the lines expressing Nematostella Vent. The relative fold change in the level at which these factors are expressed compared to the XEN cells suggests that the cells expressing the Nematostella gene are not directly programmed to a true XEN cell fate, but instead are induced to a XEN-like state in a heterogeneous culture that acquires a dependence on ERK signalling, but appear to lack the capacity for self-renewal which is characteristic of XEN cells (Kunath et al., 2005). Parenti et al. (2016) demonstrated that the OSKM factors can directly induce a XEN-like state in parallel with cells being reprogrammed to pluripotency; however the prominence of the XEN-like cells in all cultures expressing Nematostella Vent, relative to other pre-iPS cell lines that also express OSKM, suggests that the phenotype is induced by the cnidarian Vent gene. This may suggest that the Nematostella Vent gene encodes some reprogramming activity, but lacks activity sufficient to activate Nanog in mammalian cells and facilitate the final conversion to a pluripotent state. In the absence of this activity, the cells instead divert to the alternative XEN-like fate as described by Parenti et al. (2016).

5.2 THE ORIGIN OF NANOG ACTIVITY

Saccoglossus kowalevskii is a hemichordate, representative of basal deuterstomes, the embryo of which derives all three of the primary germ layers found in vertebrates during their development (Kaul-Strehlow and Stach, 2013). However cnidarians, such as Nematostella, are a much more basal organism that develops from an outer ectodermal and inner endodermal germ layer (Byrum and Martindale, 2004), without mesoderm. Davidson et al. (1995) gave a minimal estimation that the evolutionary split that diverged cnidarians from bilaterians occurred well over 600 million years ago. Because cnidarians do not develop the mesodermal layer possessed by bilaterians, which forms the tissues between the ecto- and endodermal derivatives, it is possible that the Nanog-like activity present in the Vent gene family from hemichordates evolved as part of a network permissive for mesoderm.

Another aim of this study was to gain an insight into how Nanog activity evolved. The activity required to attain ground state potential is conserved in the Nanog homeodomain (Theunissen et al., 2011), which inextricably links elucidating its origin to understanding the evolution of pluripotency itself. A more in depth comparison of the Saccoglossus and Nematostella Vent DNA-binding domains provided evidence that the activity encoded by each gene is also conserved in the homeodomain. The hybrid molecules containing swapped homeodomains retained the reprogramming properties of the complete CDS associated with each homeodomain (Fig. 33). The ability of the Saccoglossus Vent2 gene to activate the endogenous pluripotency markers (Fig. 35) was unaffected by the Nematostella flanking regions, just as the ability to induce a XEN-like state was retained by the Nematostella homeodomain hybrid.

The data from this study suggests that the ability to reprogram somatic cells and to orchestrate the pluripotency network of higher mammals evolved within the homeodomain of invertebrate Vent genes. My comparative analysis of the amino acid sequence of each homeodomain, allowed me to identify individual residues which sufficiently altered the binding capacity of Nematostella Vent to a configuration that facilitates the conversion of pre-iPS cells to naïve pluripotency (Fig. 40, Fig. 42). Future work may provide more in depth insight into the importance of these particular amino acids, and it is possible that alteration of these residues in Saccoglossus Vent to the Nematostella AAs could compromise the reprogramming ability of Vent2.

The morphology of the iPSCs reprogrammed with the modified NvVent differed from those established by Nanog or Saccoglossus Vent2 (Fig. 41). The ModNvVent iPSC line instead more closely resembled Nanog-null ESCs in culture, exhibiting the same phenotype as the Nanog null ESCs in serum + LIF that was described in Chambers et al. (2007), with colonies displaying a flatter shape with more prominent cytoplasmic protrusions, typical of embryonic stem cells on the periphery of colonies. The ModNvVent iPSCs also expand more slowly than wild-type ESCs, or mouse Nanog and Saccoglossus Vent expressing iPSC lines created in this study, which also resembles the properties of Nanog-null ESCs described by Chambers et al. (2007). Although expression of the modified NvVent induced activation of all endogenous pluripotency markers analysed by qPCR (Fig. 42), it was the upregulation of direct Nanog targets Rex1 and EssrB which were expressed at a noticeable higher level than in wildtype ESCs that was most interesting. These data suggest that the replacement of three amino acids in Nematostella Vent with the corresponding AAs from Saccoglossus, conferred a reprogramming capacity on the Cnidarian Vent gene by installing an ability to activate Nanog targets, and also the locus of the endogenous Nanog gene itself (Fig. 40). It is possible that the changes made to the Nematostella Vent gene conveyed on it hub activity, similar to Nanog (Mullin and Chambers, 2011), which was then capable of facilitating the interactions of native factors required in the transition to pluripotency. These data concur with the hypothesis that Nanog activity had evolved within an ancient Vent gene family, which preceded the emergence of vertebrates. And, that Vent was part of an ancestral gene regulatory network (GRN) that existed prior to the evolution of vertebrates.

The identification of an EsrrB gene in Nematostella vectensis which shared homology with vertebrate orthologs provided further support to the hypothesis that pluripotency evolved from an ancient gene regulatory network that existed prior to bilaterian evolution. EsrrB is a direct target of Nanog in vertebrates, and has been shown to be capable of substituting for Nanog activity in pluripotent cells (Festuccia et al., 2012). It is also capable of activating the endogenous Nanog locus in NS pre-iPS cells, thereby directly substituting for Nanog function in reprogramming (Festuccia et al., 2012).

That Nematostella EsrrB can replicate this activity, closely aligning it with the vertebrate orthologs of EsrrB (Fig. 45, Fig. 46). This provides striking evidence to support the hypothesis that pluripotency evolved from a GRN which preceded basal deuterostomes, such as the hemichordates. Expression of Nematostella EsrrB did not induce any XEN-like cells in pre-iPS line, which demonstrates a different activity to both the Nematostella Vent and modified Vent lines, indicating that they are not interchangeable. However, the iPSCs established by reprogramming with Nematostella EsrrB form a line which is unstable, and impractical for sustained culture. The reprogramming efficiency was considerably lower than that of all other genes assayed, as a total of seven colonies were generated from a plate seeded with 240K pre-iPS cells. Of the colonies that were passaged, a high proportion of individual cells underwent apoptosis (Fig 44A, B). The larger remaining cell population exhibited a stunted proliferation rate that was less than that of either the ModNvV cells or Nanog-null ESCs. The cells which expanded to form larger iPSC colonies over the course of a month, displayed a similar morphology to the ModNvVent reprogrammed cells consisting of a flatter shape with the same prominent cytoplasmic protrusions. In both the cultures containing Nematostella EsrrB and ModNvVent reprogrammed cells, the positive LacZ staining was exclusive to the centre of the ESC-like colonies, indicative of the endogenous Nanog locus being deactivated in the peripheral cells. This may be due to a lack of Nanog-like regulation of the level of endogenous pluripotency factors. In the cells reprogrammed with NvEsrrB, the endogenous Oct4 is upregulated to a noticeably higher level relative compared to its expression level in wildtype ESCs. A highly elevated level of Oct4 has been shown to be incompatible with the self-renewal of both ESCs and iPSCs (Niwa et al., 2000; Radzisheuskaya et al., 2013), which may account in part for the observations made here.

However, activation of the endogenous Nanog locus, and other naïve markers such as NroB1 and Klf2, demonstrably prove that Nematostella EsrrB has reprogrammed the pre-iPS cells to a pluripotent state. A future comparison with vertebrate orthologs and analysis of the Nematostella Vent expressing cells' capacity to differentiate will further elucidate the activity of Nematostella EsrrB.

5.3 HYPOTHESIS

I have proposed that Nanog-like activity evolved in the Vent homeodomain at the base of deuterstome evolution, coinciding with the advent of mesoderm. The presence of Vent and EsrrB orthologs in a cnidarian such as Nematostella, suggests that the genes participated in in a network which preceded the evolution of reprogramming activity in bilaterian members of the homeobox gene family. One such network, which appears to be conserved in the trunk of vertebrate evolution, is the GRN that governs mesoderm development upstream of PGC specification (Johnson et al., 2011; Johnson and Alberio, 2015). Previous research in our lab has demonstrated the stochastic specification of PGCs in axolotl embryos from mesodermal precursors capable of generating multiple somatic lineages, in addition to the PGCs, which are hypothesised to be "the last cells standing", or a default state of specification (Chatfield et al., 2014; Johnson and Alberio, 2015). This pluripotent mesodermal state is maintained until late irreversible commitment to the germ line around the mid-tailbud stage of development (Chatfield et al., 2014). We believe that development of this pluripotent mesodermal domain is reliant on a GRN which includes Nanog, Vent and EsrrB (Fig. 47; Johnson and colleagues, unpublished.).

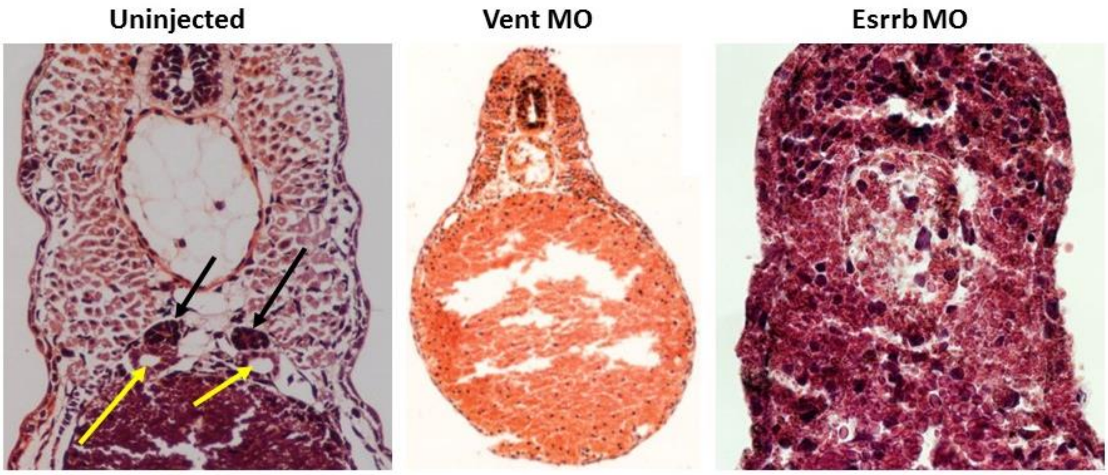


Figure 47: Section of Vent and EsrrB morphants compared with uninjected.

Knockdown of Vent or EsrrB disrupts development of the AGM.

Yellow arrows indicate mesonephric ducts. Black arrows indicate PGCs. (Johnson and colleagues, unpublished).

Knockdown of Vent expression in axolotl embryos, using antisense morpholinos, aborts development of the aorta-gonad-mesonephros (AGM) region from which the PGCs are derived in axolotls (Fig. 47) (Chatfield et al., 2014), and in mammals (Calvopina et al., 2015).

In vitro studies, using induced axolotl animal caps, show that mesoderm specified to produce PGCs is instead diverted to endoderm, in the absence of expression of VentX, the axolotl Vent ortholog. This also phenocopies the knock down of Nanog by injection of Nanog morpholino, which arrests normal development of the embryo and instead causes everything to divert to endoderm (Johnson and colleagues, unpublished). Nanog and Vent both form part of a GRN required for PGC induction in an axolotl. Interestingly, knockdown of EsrrB in an axolotl also disrupts development of the posterior lateral mesoderm, extinguishing pluripotency in mesodermal precursors, and preventing the specification of the PGCs (Fig. 47). EsrrB joins both Nanog and Vent as part of the axolotl GRN that regulates specification of the AGM region.

Stochastic specification of PGCs, which appears conserved, depends on pluripotency being maintained within a pool of mesodermal cells in the axolotl (Chatfield et al.

2014). Recently it was shown that human PGCs could be induced from hESCs via a competent mesendodermal intermediate state (Kobayashi et al., 2017).

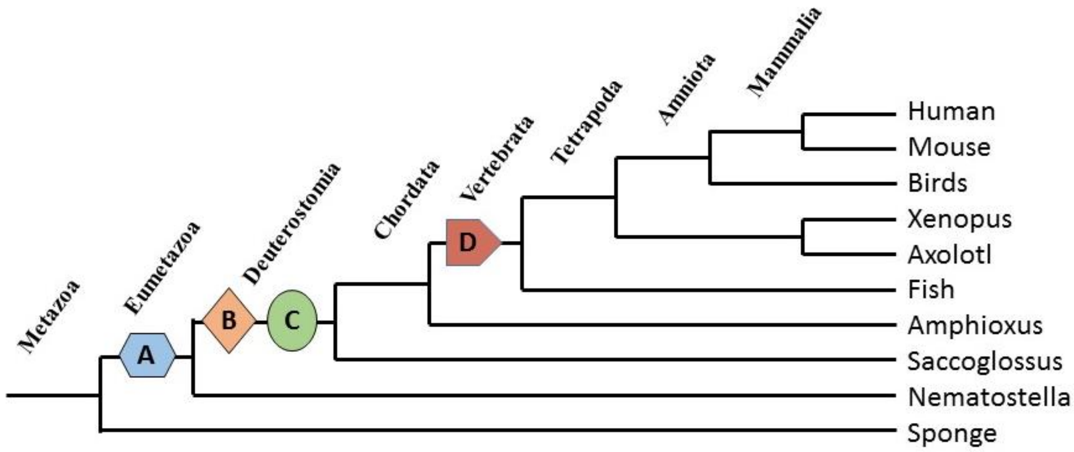


Figure 48: Illustration of my findings in a wider biological context.
(A) An ancient GRN containing Vent and EsrrB orthologues is conserved in cnidarians.

(B) Evolution of an ancient pluripotent mesoderm at the base of deuterostomes.

(C) Pluripotent mesoderm gave rise to Nanog-like activity.

(D) The ancient GRN for pluripotent mesoderm evolved at the base of vertebrates to form part of the ground state network.

Both Vent and Nanog are crucial regulators of mesodermal pluripotency in axolotl, and it is possible that the evolution of Nanog-like activity in the Vent homeodomain of basal deuterstomes, for which Saccoglossus kowalevskii is a model, facilitated the evolution of an ancient pluripotent mesoderm. Both the Nematostella Vent gene, and an EsrrB gene which encodes the pluripotent potential exhibited by vertebrate orthologs, provide evidence suggesting that an ancient GRN conserved in cnidarians (Fig. 48A) was configured to regulate formation of the pluripotent mesoderm in deuterstomes (Fig. 48B), which over time gave rise to Nanog activity (Fig. 48C). I propose that an ancient GRN for pluripotent mesoderm, containing the ancestors to both Nanog and EsrrB, evolved in vertebrates to form part of the ground state network (Fig. 48D).

6 APPENDICES

6.1 CODING SEQUENCES OF GENES USED IN THIS STUDY

Mouse Nanog:

ATGagtgtgggtcttcctggtccccacagtttgcctagttctgaggaagcatcgaattctgggaacgcctcatcaatgcct
gcagtttttcatcccgagaactattcttgcttacaagggtctgctactgagatgctctgcacagaggctgcctctcctcgccctt
cctctgaagacctgcctcttcaaggcagccctgattcttctaccagtcccaaacaaaagctctcaagtcctgaggctgacaa
gggccctgaggaggaggagaacaaggtccttgccaggaagcagaagatgcggactgtgttctctcaggcccagctgtgt
gcactcaaggacaggtttcagaagcagaagtacctcagcctccagcagatgcaagaactctcctccattctgaacctgagc
tataagcaggttaagacctggtttcaaaaccaaaggatgaagtgcaagcggtggcagaaaaaccagtggttgaagactag
caatggtctgattcagaagggctcagcaccagtggagtatcccagcatccattgcagctatccccagggctatctggtgaac
gcatctggaagcctttccatgtggggcagccagacttggaccaacccaacttggagcagccagacctggaccaacccaa
cttggaacaaccagacctggaccaacccaacttggagcagccaggcctggaccgctcagtcctggaacggccagccttg
gaatgctgctccgctccataacttcggggaggactttctgcagccttacgtacagttgcagcaaaacttctctgccagtgattt
ggaggtgaatttggaagccactagggaaagccatgcgcattttagcaccccacaagccttggaattattcctgaactactct
gtgactccaccaggtgaaatatga

Axolotl Nanog:

ATGCCCGCCCACTGCATGACCCCGCAAATGAGGGTGCCAGGATACCAGGC
CTACCCGGAACTGAACCAACCGTGTGCTGTGGCGGCGAACTACGCGCCAG
CATTCCAGGGGAACCTGCAGGCCGCAGAAGAGGGGCGCCCGGTCCCGGC
AGTCCTGCCGTCTTCCCCCGACTCAGCCACCAGCCCAAAGGTGGATCCGTT
CGCCCAGTGTGCCCCGGATGTAGCGGTGGGAGGTGAGACGAAGGCCCAG
GCGAGAAAGCGCACGTGCTTCAGTCAGGAACAGCTCGTGGCCCTGCACCG
AATGTTCCAGAAGCAGCATTACATGAACCCAATGCAGGCGCAGCAGCTGG
CGGCGGATCTCAACCTCACATACAAGCAGGTGAAGAACTGGTTCCAGAAC
CGAAGGATGAAGCACAAACTGTCCCTGAAGGACTCAGTGTGGCTGGACA
AGAGGTGCTGGCAACCACAGGCGAGCTCAATCCTAACACCGGCTCAGCCG
CAGTCCACGGGGTGCCCGGAAAGCTCCTCCCACTTGCCTCAGAGATACAC
AGTGCACCATAGTGCACTTGCCCAGCGAGTCACCAGTCACCCCTACCAGA
AGTACAGTGGGATTCAGAACCCTCACCAGAAAGTGCTGTCAGAAGACGCA
GCAACAGTGCAGCACAGGGAGGCCGCCCCGCAGTGCATGGGGCCACAGC
AGTACATGAACAGGCACCAAAATTACCCAACCATTGAGTACGCCGGAGCA
AGACCAGTAGAAGGGTACAACCTTAAGACTCCTCTGCAGTACCCTTCGAT
GGCACCATACCCCAACTACTACTACCAGCAGCCACCGCCCTACATTCATC
AGCAGGGCCGGCCCGATATCAGATTTCAGAGCACGCAGGGCATGTGA

Eggnog:

ATGagcagccccgcttcacctctgacccccggccaggggtccaaacccgcggtgaactctcagctccccggccccac
cttctcgccgagcagcggcatgatccacttccgctggaacgcgcaggctcaagcggagggctccgactgtgtcctagtgt
acgccgagaagggcgccccggctccagactggtcatctacatctccggactctgcaattgacttatctccaacggaccccg
tctgctctaagtttgagtttccctccctgcctgatgtggacaaggtgaaagtggctgagaagaagccaaagttgcgcgcttgc
ttctcaacaaagcagctccttgccatgcaggagcgctttgagcagcagaagtatcttagtcccatgcagatccaggaattga
gcacatctttgggcctcacctataaacaggtgaagacttggtttcagaatcgtagaattaagaacaagaaagctgagaagta
cacagctgcggctgggaagaacagcacaatgcagtttaactctcagcctgtgccgtgctgtgcgatgccaccaggatttcc
ctctggttaccccatccatgccggagaaaatatcattggacctgtggctgctccattgcatcatccaaaccctacactcatgct
tgggcctgaagaacagccaatgaggtggaacggggatgtcttcccactcgcccctgttcctcccttcaagcaacactcagc
agggcttgcgaagcaggacttggatttctggccgaacatcatggtgatggaaggcatgcggcagggtgactgctgcagttt
tcctccccccttggacctgtcaggttcctgctggattttatcatggctacgagcaggggccgaactgccttcctttgcagaatta
cccacctcagtgtgtgatgggcccgcctcagtgggggaatgcagagcag
Ctttga

Axolotl Vent:

ATGGGCAAGGCCAGCTTCTCCATAGATTATCTGGTACAGAGTCAACAAGG
ACAGCGCAGCCTCCACTGCACAGAGGGGCTCCTAAGCCTGCCGGCTAGGC
CGCATGTGCCGTGCCAGGCCCCACCAACGCCCCCGACCTCTCTCTACAAG
GTGACTCGCTTATCCCGGCCCAAGAAAGAATCCACACCGGGCTGCACTCC
TAGGAGCCGAGAGCTGGCACCGGAAAGCAGGCGCCAGGAGCCATCTTGT
CAGATCAGCCAGCCCCCAGAGAGCAGACACGAGGCCTCCGAGGGCAGGG
TCCACGCGAACAGGGAACAGGCAGCAGGGGTTCGATATCCGGAGCCGGT
GTGTGAGCGCATGCGGCAGGTCTCGCAGTACAGCGAGCAGGCGTCATTCT
GTGGGAGCAGGCAGCGGTCCCCAGATAACAGCCACGTTGAGCCCGGAAC
CAGTGAGCAGGGCCGCGAGTTCAAGGGTTCGGAGCCGGCGTGTGTGAGC
AGACAGCTGGAGCCTTCCTTCACGAGCAGAGAGCCACCCTTCGGGAGCAC
GTTCCAGTCGCCATCCGGCACACCGGACCCGTTCGATGAGGAGGAGAGAA
GCTGGAGTGCGGACGAGACATCTTGCCCATCCCTGAGCGACGAAGGTGAA
GCTCTGCGGCTCGCGGACTGTGAATCCCCGCGTCTGCTCGAGGAAGTGGG
GTCCCCGAAAGGCGACAGCCAGCGCAGAGCGAGGACCGCCTTCAGTAGC
CAGCAGCTCTCCAGGCTGGAGCAGACGTTCGCGAAGCAGCAGTACCTGCC
GGCCCAGGAGCGGAGGAAGCTGGCCACCGCCCTCCAGATCTCCGAGATGC
AGATCAAAACTTGGTTCCAGAATCGCAGGACAAAGATGAAAAGACAGAC
CCTGGACCAGCAGCACATTGTGATGCCTTTTGGAAACCCACATGTCCCCTC
TGGCTTTCCCTATGGATCCAAGCAGCTAGAGTTGCCCCCTTATTACTTGAA
CTGTGGGAGCCAGTACAACACTTCAACACCCCTCAACGGACCTTATCCCTT
TTATGGACCTGACGTGTCATACTACCAACAGACCCCATACCACCAAGCCC
CCTTTCATCACCAGCCATACCAGCAACTCTCCTATCACCAGCCCAACTATC
AGCAACTCCCATTTTACCAGCACACTTTCCACCGACAGCCATTTCACCCTG
ACACGATGAGTTGGGAAGCTCAGGAACAATCCCCACATCCACTGTACCCC
TTGACTGGCACCTTGAAGTGCTGA

Amphioxus Vent1:

ATGACTGCCTGCAAGCTCCCCTTCGCTCTCTCCGTGGAAAACTTGATCGGC
GCTCGAACTGACCTTGTTCCTGACATCAGAGCTTCCAGCCCCGAATGTCTT
CCACCGAAGAAGCGTCACCTCCTGTTCACCGCCTCGCCGGGCCCGGACGC

AAACACTCGGACCGCCGCTCAAGTCGTGCCGCCCGCCGCTAACGCTCTGC
GGAGTGACGGAGAAACGTCTGACGAGGAGGAGGAGATCGACGTCGTTTC
CATTGGGGACGCAGCCTCACCTGCTTCCAGCACGAGTTCAGACCGCTCCA
GCCCAGCACCTTCCTGCGACGACAGCCCGTCTCCCTCCCCGGCAGGCCGC
AAGGCACGCACGGCCTTCACCACCGAGCAGGTCATGGCCCTGGAGGAGC
GTTTCCGCCTGCAGAAGTACCTGTCCGCCGCCGACCGGGAAACTCTGGCC
AAGGCCACGGGACTGACCGACGAGCAGGTCAAGACGTGGTTCCAGAACC
GCCGTATGAAGCTGAAGCGCCAGCAGCAAGACTTCGCCACCTTCCCGCTG
CACGCCTCCGTCACCGTGCCCGGGATGCCCGGCTACTCCCCACAGCCCAA
CCCCTGGTACGACTCGGCGCACTTCGTCCCGTCCCTCGGCACCGCCGTCTG
TGGTCTGCGGCCTCACATGCTGCCTCAGCCCGCCAACACCGTGACCACCA
GCCGTGGGAACAGATACTCACCGTACCCGTCTGCCCGCGTGTCTCCGTCC
GGCCGCAGCCACCCCCAGATGACCAACATGTACATGTCTCCGTACCAGAA
CAACGTGGCAGGGGCGCACTACAACAACGTACCCAGCATGAACTTCTAA

Amphioxus Vent2:

ATGGCCCTGGAGGAGCGTTTCCGCCTGCAGAAGTACCTGTCCGCCGCCGA
CCGGGAAACTCTGGCCAAGGCCACGGGACTGACCGACGAGCAGGTCAAG
ACGTGGTTCCAGAACCGCCGCATGAAGCTGAAGCGCCAGCAGCAAGACTT
CGCCACCTTCCCGCTGCACGCCTCCGTCACCGTGCCCGGGATGCCCGGCTA
CTCCCCACAGCCCAACCCCTGGTACGACTCCGCGCACTTCGTCCCGTCCCT
CGGCACCGCCGTCTGTGGTCTGCGGCCTCACATGCTGCCCCAGCCCGCCA
ACACCGTGACCACCAGCCGTGGGAGCAGATACTCACCGTACCCGTCTGCC
CGCGTGTCTCCGTCCGGCCGCAGCCACCACCAGATGACCAACATGTACAT
GTCTCCGTACCAGAACACGTGGCAGGGGCGCACTACCACAACGTACCCGG
TGTTTACATGTAA

Saccoglossus kowalevskii Vent1:

ATGGCGGCTTCTAGGTCGGAGCAAGAGCAATCAGTTCTCTCTAATTCCCCT
GCGTTCTCCAGTCAATCAACAACACATGCACAGAATGAAGACTCTCTTTC
ACCATCAGTAACAACAACGACTGAATTACTGTGTTTAGTATGTTCCGACA
AGGGGTCTGGATACCATTATTCGGTATTTTCGTGCGAGGGTTGTAAGGGTT
TTTTCAAACGAACCGTTCAAAAACAACTGAAATACAGCTGTAAAGAAAGC
AGGAACTGCGAGATTACAAAGTACAGTAGGAACAGTTGCCAGTCATGCCG
CTTTCAGAAATGTGTCAACATGGGCATGAAAATAGACGCTGTACGAGAAG
ACAGAACACCGGGTGGTAAACAGAAAATAAAGCGCTCACGATCACATTC
ATCAAGTCTGCCCGTGAACCAAAGTACAAAAAAAACATTCCCTAATGTGA
GTGATACAAACAATGATACAGCATCCTCGTCATCGTTAACCGATACAAAT
TCCCCTCAACCTGAGAAACGTTTAAAAGAAGACAGTGACGTTTCACAGGA
GGATCAAAGTAAAGCGTATGAGACAATGGCCGAGTTACTAACCGCTGCAC
GTCCTGACATAGTACCTGAAGCTCAGTCATTCAAGGATATGAGTAATTTTG
GCATAGATGATTTGATGCAGTATGGATACAAAGAATTGCAGTTAATTATA
CAGTGGGCTCGGCATGTGCCGGGATTTCGAGGTCTACTGATAGAAGATCA
AATGGCACTTTTAAAAGCATCATTCATGGAACTAAATGTACTCAGGTTAG
CATATAGATCTAGAAGTGAAAATGGTTTGGTTAGATTTGGCGCTAACATG

AAGCTCACCAAAGATCAGGCCATGGAGATTGGCTGGGGACAAGAACTGA
TACAATCTACTTTAAATTATGTTTCAAGGATACAAGAAGTTAACGTTGATC
ACGTAGAGTTCTGTTTTCTTAACGCCATTGTACTTACTTATCCAGATGCCG
CAGGACTACAAGACAAAACCAGTGTAACAACTCTTCAAGCCAAAATAGTT
GATGCACTTCGACACTACATCAAAACACACTATCCAAAAGATGCCAAGAG
ATATGCAAAGGTGTTACTAAGGTTACCAACTTTACGAACTGTTGCTTCTAA
AGCAACTGAGAGATTTTTATCAATGACATTGGATGGCACAATACAAATCA
ATGAATTAGTATCAGAGATGATGGGATAA

Saccoglossus kowalevskii Vent2:

ATGGTTCAACATACAGCAACACCAACCGCCATGCATATTGGGGTTGGCGG
TAGCAGACCAGATACGAATGACTACGAGATTGCCATGATATTGGCGAGGA
ACATGGCACAACAAACTTCAAATGACCCATCTCCTATAGCTACATCTACG
CCCGAATCACCTGATAGCGGTTGCGATGGTAACAACGAACACGGACCAAA
ACGCAAGAGAGCTCGCACTGCCTTCAGTAACGAACAGGTTTATAAATTAG
AGAAGAGATTCCGAGCACAGAAATATTTATCCGCCACTGAAAGGGAGGA
TGTTTCTCGTTCTATTGGACTATCTGACACACAGGTAAAAACATGGTTTCA
AAACCGCCGCATGAAGTGGAAGCGTGAACGAAAAGACGACGTCCCCGAT
ATGTCGATGTCACACGCTGGAATGTACGCGCACACCACTCACAATTATAT
GCAATCCTATCCGATCAATGGAGCATATCCGTCACATCCTTGCATGTCGAG
CACCGCCTCATCCATCGGGTCCGTTCCCCAGCAACCCCAACCGATGACCA
CCATGTCGCTTGCCAGTACTCCGATGATGCCATCGTCGTATTCAGTACCGT
TTCCGTCGACAGGGACATGTCCCACTACCTACTACCCGACATATCATAGA
ATGAGGACGACGACCTCTCATTTTGCCATCGGGCATTACAACTACAACAT
GCCATTTGCGCACCGCCCCATGTATTAA

Nematostella vectensis Vent:

ATGGAAGACCAACAAAACAAGGACTTTGACAACCAGGCATCATTTGTCAC
CAGTAAAGGCGCCGAAAGAAAGAAGCCTACTTTCCTAATCAAATACATTA
TGGGTGATACACTACTGGGCGGCCAAGCTTTAAGACGTGCCGATGTGAAA
ACCTCAAAGGTCTTGAGCGCACAGGCAAATACGATTCCTCGGGATACGCC
TACAGAGGCCTTCACCGACAAGGCCTCTTCACCAGTCGGCGCCAGTCCAT
ACGAGCTACTAAAAAAAAAGCCGCGCACGGCATTCACAGAATCTCAAAT
ATCCGAACTCGAGAAACGCTTCCAGTCACAGAAATACCTCGGGAGCAAGG
AACGGAGTGAACTGGCCGGCACACTTGGACTCACAGACACTCAAGTGAA
GACGTGGTTTCAAAACAGACGAATGAAGCTGAAGCGACAGCGCCAAGAG
GACACAGATCGCCAGGCTAGATATTCACTTGTGAGCCCCTACAGGCTCGC
CCCTTATGCTAGTCACGTGGTGCCATCATACCCAGTGACGTCATACACCGG
TTACTATGGCAACTTCCCAATGCCTCGAGTTTTATGTGATGACCTTGTCGC
CTCTCCCGCATCGTCATCAGTGTCGCCAACAGGGACGACCGGAGAGTTCC
ACGTCACGACAGATCAAAGATACAGCTACTTGGCCTCCTATCCCTCCTCG
GGATACCTGTGTCCCGTAGCGCCATTCCATGTTTTT

Ephidatia Fluviatilis:

ATGGACAACTGCAGAGGCGAGAAGAAGCCCCTGCTGAGCACCAACCAGC
AGAGCTTCCGGATCGACAACCTGCTGACCCGGAAAGTGATCGAACAACAG
CAGCAGCCCGACCACTACACCATCTACAGCCCCAGCAAGATCAGCGACGG
CACCAAGATCTACGAGCACGGCAGAGAGAGCACCAGCAGCACCTCCGGC
AACGACGCCGACGATGATCTGCTGACCAGACGGAAGAAGGCCCGGACCG
CCTTCAGCAGAGAACAGGTGGCCGAGCTGGAAAAGAAGTTCCAGAACAA
GAAGTACCTGAGCAGCACCGAGCGGGGCGAGCTGGCCGAGAAGCTGAAG
CTGAGCGACATGCAAGTGAAAACCTGGTTCCAGAATCGGCGGATGAAGTA
CAAGCGGCAGAGCGAAGAGACAGAGATGGAAATGAAGTCCTCTAAGTAC
GGCAGCTTCGTGTACCGGGGCGTGGCCCACAGCATGTTCAACGCCGTGCT
CGAGTGA

Xenopus Laevis Vent2:

ATGAACACAAGGACTACTACAGGAAAGATGACCAAAGCTTTCTCTTCAGT
AGAATGGCTTGCCCAAAGCAGCCGGAGATCTCACAAAGAGCAGCCCAGC
AAAGGGGATCAGCGATATTCACCGTACCCAAGCCCATCCCTGCCTTCCTG
GAACAGTGATGTGTCCCCTTCTTCATGAACAGCCAGCTGTCTCCAGTTGCA
GGCAGTGCCCAAGTCTCACCGTGCCCTGGGAGTGCACAATATTCCTCAGA
CAGTGAATTATCTCTCTATTCCAATGAAGATGAAGACTTATTCTGTGAAAA
GGACCTTAATACCCCTTCTACCCCTGGGGACAATGGATTGCTACACAGGG
ACACAGCCACACATGAGTACTCTGGAATGGTTTCAGTGCCAGCAAATACT
CCCAGAACCACTTCCAATGAAGATGCTGCTAAGTCTGGCTACAGCACTAG
CACTGACTCAGGCTATGAAAGTGAAGCAAGTCGCTCCAGCTCTACAGCCC
CTGAAGGAGATGCCACCGTATCCCTGAGCCCCAATGATACTTCAGATGAA
GAGGGCAAGTTGGGCCGAAGGTTGCGAACAGCTTTCACCAGTGATCAGAT
CTCCACTCTGGAGAAGACTTTTCAGAAACACAGATACCTTGGGGCTTCTG
AGAGACGGAAACTCGCAGCCAAACTCCAGCTTTCTGAAGTCCAGATTAAA
ACCTGGTTCCAGAACCGCAGGATGAAATACAAACGGGAAATCCAAGATG
GCAGACCTGACTCATACCACCCAGCCCAGTTCTTTGGTGTGTACGGCTATT
CACAGCAGCCCACTCCTGTATTCCAACATGCAGTGCAACAACCCTACCCA
GGTTATAGCCCACTAATGGAAACCCTGCCTGGCACCATGCCCTATGCCAT
GCACCCCCCTGCCATGGACTCTCTGAATCATTTCAACTCTCCACCTTTTCA
AATGTTCTACATGCCCCAACAGCACCTTGGGCAGCCTCTGGCCTACTAA

Modified Nematostella Vent:

ATGGAAGACCAACAAAACAAGGACTTTGACAACCAGGCATCATTTGTCAC
CAGTAAAGGCGCCGAAAGAAAGAAGCCTACTTTCCTAATCAAATACATTA
TGGGTGATACACTACTGGGCGGCCAAGCTTTAAGACGTGCCGATGTGAAA
ACCTCAAAGGTCTTGAGCGCACAGGCAAATACGATTCCTCGGGATACGCC
TACAGAGGCCTTCACCGACAAGGCCTCTTCACCAGTCGGCGCCAGTCCAT
ACGAGCTACTAAAAAAAaagccgcgcacggcattcacagaatctcaaatatccaaactcgagaaacgctt
cagatcacagaaatacctcgggagcacagaacggagtgaactggccggcacacttggactcacagacactcaagtgaag
acgtggtttcaaaacagacgaatgaagctgaagCGACAGCGCCAAGAGGACACAGATCGCCA
GGCTAGATATTCACTTGTGAGCCCCTACAGGCTCGCCCCTTATGCTAGTCA
CGTGGTGCCATCATACCCAGTGACGTCATACACCGGTTACTATGGCAACTT

CCCAATGCCTCGAGTTTTATGTGATGACCTTGTCGCCTCTCCCGCATCGTC
ATCAGTGTCGCCAACAGGGACGACCGGAGAGTTCCACGTCACGACAGATC
AAAGATACAGCTACTTGGCCTCCTATCCCTCCTCGGGATACCTGTGTCCCG
TAGCGCCATTCCATGTTTTT

Nematostella EsrrB:

ATGTTGTTATGTGTTGTATGCAACGATAAGGCAACAGGGTACCATTATGGT
GTGTTCACATGTGAGGGATGCAAGGGGTTCTTCAAAAGAACCGTACAGAA
ACAACTTGAGTACACATGTCGTGGAAATCAAGACTGTGACATCAATCAGC
ACACACGTAACCGCTGCCAGTACTGTCGTTTCGAGAAATGTGTCAAAAGT
GGAATGCTTAAGCAAGCTGTTCGTGATGATAGAACTCCTGGAGGTCGCCA
CAGACATGCTTCCCTTCAGGATCACAGACAGAAGAAACAAAAATTACAAG
GCAAAGCTCCCAAAGAGCTTCCACAAAATGGCCTGTCATCTCCGGAGGGT
CCGGATACTCCAGAGGGGGTGGTATCGAAAGAGGACGAGAATGAATTCC
TGATTTATATACAGCAATTACGACATGATGTGGATGTTATCCCTGACAGCC
CTGTCAAGCCGCCCGAAGGCTCCCCGGCGGCAGGCTTTAGTGTCGACAAG
CTCATGCAATATGGCTACCAAGAACTATATCAGGTAATTCAGTGGGGTAA
AAACGTCCCAGGCTTCCGTGAGCTGAAACTAGAGGATCAAATAACACTCC
TGAAGACCTCTTTCATGGACCTGAATGTTTTTCGCCTGGCCTATCGCTCTA
TCTGTTGTGACCCGGATTCCCTAATGTTCGCCAAGGGCATTATACTGAACA
AGCCTCAGTGTCTGGAGATGGGATGGAGTATGGATCTAACGGAGACCACG
CTTGAGTTCTGCGCCAAGCTCAGAAATTTGAACATGGATGTGAATGAGTTT
TCATGTCTAAGTGGTTTGGTGCTACTAAGTCCTGATGCACCAGGTGTGGTC
GACAAAGACAAAGTGACAGAGCTCCAGACAAACGTGACATCGTGCTTGC
GTGACTATGAGATGTACCGTTACCCAAGCAAAAGCAACCGTCTTGGTAAA
CTCCTCCTATGCCTCCCGTCACTGCGTGCATACAGTGAGAAGGCCTTGGAA
AACTATGTCACACTTGAATTCTTTGGCAAACTTGACATGCCACCTCTTGTA
GCAGAGCTATTGGAGTAA

NvVent N-domain-Saccoglossus VentHD-NvVent C-Domain:

ATGGAAGACCAACAAAACAAGGACTTTGACAACCAGGCATCATTTGTCAC
CAGTAAAGGCGCCGAAAGAAAGAAGCCTACTTTCCTAATCAAATACATTA
TGGGTGATACACTACTGGGCGGCCAAGCTTTAAGACGTGCCGATGTGAAA
ACCTCAAAGGTCTTGAGCGCACAGGCAAATACGATTCCTCGGGATACGCC
TACAGAGGCCTTCACCGACAAGGCCTCTTCACCAGTCGGCGCCAGTCCAT
ACGAGCTACTAAAAAAAagagaattcAAGAGAGCTCGCACTGCCTTCAGTAAC
GAACAGGTTTATAAATTAGAGAAGAGATTCCGAGCACAGAAATATTTATC
CGCCACTGAAAGGGAGGATGTTTCTCGTTCTATTGGACTATCTGACACAC
AGGTAAAAACATGGTTTCAAAACCGCCGCATGAAGTGGAAGCGTcgcaagctt
CGACAGCGCCAAGAGGACACAGATCGCCAGGCTAGATATTCACTTGTGAG
CCCCTACAGGCTCGCCCCTTATGCTAGTCACGTGGTGCCATCATACCCAGT
GACGTCATACACCGGTTACTATGGCAACTTCCCAATGCCTCGAGTTTTATG

TGATGACCTTGTCGCCTCTCCCGCATCGTCATCAGTGTCGCCAACAGGGAC
GACCGGAGAGTTCCACGTCACGACAGATCAAAGATACAGCTACTTGGCCT
CCTATCCCTCCTCGGGATACCTGTGTCCCGTAGCGCCATTCCATGTTTTT

SkVent N-domain-Nematostella VentHD-SkVent C-Domain:

ATGGTTCAACATACAGCAACACCAACCGCCATGCATATTGGGGTTGGCGG
TAGCAGACCAGATACGAATGACTACGAGATTGCCATGATATTGGCGAGGA
ACATGGCACAACAAACTTCAAATGACCCATCTCCTATAGCTACATCTACG
CCCGAATCACCTGATAGCGGTTGCGATGGTAACAACGAACACGGACCAAA
ACGCagagaattcaagccgcgcacggcattcacagaatctcaaatatccgaactcgagaaacgcttccagtcacagaa
atacctcgggagcaaggaacggagtgaactggccggcacacttggactcacagacactcaagtgaagacgtggtttcaa
aacagacgaatgaagctgaagcgcaagcttGAACGAAAAGACGACGTCCCCGATATGTCGA
TGTCACACGCTGGAATGTACGCGCACACCACTCACAATTATATGCAATCC
TATCCGATCAATGGAGCATATCCGTCACATCCTTGCATGTCGAGCACCGCC
TCATCCATCGGGTCCGTTCCCCAGCAACCCCAACCGATGACCACCATGTC
GCTTGCCAGTACTCCGATGATGCCATCGTCGTATTCAGTACCGTTTCCGTC
GACAGGGACATGTCCCACTACCTACTACCCGACATATCATAGAATGAGGA
CGACGACCTCTCATTTTGCCATCGGGCATTACAACTACAACATGCCATTTG
CGCACCGCCCCATGTAT

Mouse Nanog Tryptophan repeat (WR) region:

TGGGGCAGCCAGACTTGGACCAACCCAACTTGGAGCAGCCAGACCTGGAC
CAACCCAACTTGGAACAACCAGACCTGGACCAACCCAACTTGGAGCAGCC
AGGCCTGGACCGCTCAGTCCTGGAACGGCCAGCCTTGGAATGCTGCTCCG
CTCCATAACTTCGGGGAGGACTTTCTGCAGCCTTACGTACAGTTGCAGCA
AAACTTCTCTGCCAGTGATTTGGAGGTGAATTTGGAAGCCACTAGGGAAA
GCCATGCGCATTTTAGCACCCCACAAGCCTTGGAATTATT

7 BIBLIOGRAPHY

Alberio, R., Croxall, N. and Allegrucci, C. (2010). Pig epiblast stem cells depend on activin/nodal signaling for pluripotency and self renewal. Stem Cells Dev. 2010 Oct; 19(10): 1627–1636.

Ambrosetti DC, Basilico C, Dailey L. (1997). Synergistic activation of the fibroblast growth factor 4 enhancer by Sox2 and Oct-3 depends on protein-protein interactions facilitated by a specific spatial arrangement of factor binding sites. Mol Cell Biol (11):6321-9.

Ahmed K, Dehghani H, Rugg-Gunn P, Fussner E, Rossant J, Bazett-Jones DP. (2010). Global chromatin architecture reflects pluripotency and lineage commitment in the early mouse embryo. PLoS One 5: e10531.

Avilion, A. A., Nicolis, S. K., Pevny, L. H., Perez, L., Vivian, N. and Lovell-Badge, R. (2003). Multipotent cell lineages in early mouse development depend on SOX2 function. Genes Dev. 17, 126-140.

Bachvarova R. F., Crother B. I., Johnson A. D. (2009a). Evolution of germ cell development in tetrapods: comparison of urodeles and amniotes. Evol. Dev. 11, 603-609.

Bachvarova R. F., Crother B. I., Manova K., Chatfield J., Shoemaker C. M., Crews D. P., Johnson A. D. (2009b). Expression of Dazl and Vasa in turtle embryos and ovaries: evidence for inductive specification of germ cells. Evol. Dev. 11, 525-534.

Banerjee-Basu S., and Baxevanis A. D. (2001). Molecular evolution of the homeodomain family of transcription factors. Nucleic Acids Res; 29(15): 3258–3269

Bernstein BE, Mikkelsen TS, Xie X, Kamal M, Huebert DJ, Cuff J, Fry B, Meissner A, Wernig M, Plath K, Jaenisch R, Wagschal A, Feil R, Schreiber SL, Lander ES. (2006). A bivalent chromatin structure marks key developmental genes in embryonic stem cells. Cell;125:315–326.

Bertrand S., Escriva H. (2011). Evolutionary crossroads in developmental biology: amphioxus. Development 138: 4819-4830.

Bessonnard S, De Mot L, Gonze D, Barriol M, Dennis C, Goldbeter A, Dupont G, Chazaud C. (2014) Gata6, Nanog and Erk signaling control cell fate in the inner cell mass through a tristable regulatory network. Development; 141(19):3637-48.

Blinka S. and Sridhar Rao S. (2017) Nanog Expression in Embryonic Stem Cells – An Ideal Model System to Dissect Enhancer Function. Bioessays.

Booth HA, Holland PW. (2004). Eleven daughters of NANOG. Genomics; 84(2):229-38.

Bourlat SJ, Juliusdottir T, Lowe CJ, Freeman R, Aronowicz J, Kirschner M, Lander ES, Thorndyke M, Nakano H, Kohn AB, Heyland A, Moroz LL, Copley RR, Telford MJ. (2006). Deuterostome phylogeny reveals monophyletic chordates and the new phylum Xenoturbellida. Nature;444(7115):85-8. Epub 2006 Oct 18.

Boskovic A, Eid A, Pontabry J, Ishiuchi T, Spiegelhalter C, Raghu Ram EV, Meshorer E, Torres-Padilla ME. (2014). Higher chromatin mobility supports totipotency and precedes pluripotency in vivo. Genes Dev 28: 1042–1047.

Bowles J, Schepers G, Koopman P. (2000). Phylogeny of the SOX family of developmental transcription factors based on sequence and structural indicators. Dev Biol.; 227(2):239-55.

Boyer LA, Lee TI, Cole MF, Johnstone SE, Levine SS, Zucker JP, Guenther MG, Kumar RM, Murray HL, Jenner RG, Gifford DK, Melton DA, Jaenisch R, Young RA. (2005). Core transcriptional regulatory circuitry in human embryonic stem cells. Cell; 122(6):947-56.

Brambrink T, Foreman R, Welstead GG, Lengner CJ, Wernig M, Suh H, Jaenisch R. 2008. Sequential expression of pluripotency markers during direct reprogramming of mouse somatic cells. Cell Stem Cell 2: 151–159.

Brons, I. G., Smithers, L. E., Trotter, M. W., Rugg-Gunn, P., Sun, B., Chuva de Sousa Lopes, S. M., Howlett, S. K., Clarkson, A., Ahrlund-Richter, L., Pedersen, R. A. et al. (2007). Derivation of pluripotent epiblast stem cells from mammalian embryos. Nature448, 191-195.

Buehr, M., Meek, S., Blair, K., Yang, J., Ure, J., Silva, J., McLay, R., Hall, J., Ying, Q. L. and Smith, A. (2008). Capture of authentic embryonic stem cells from rat blastocysts. Cell 135, 1287-1298.

Buganim Y, Faddah DA, Cheng AW, Itskovich E, Markoulaki S, Ganz K, Klemm SL, van Oudenaarden A, Jaenisch R. (2012). Single-cell expression analyses during cellular reprogramming reveal an early stochastic and a late hierarchic phase. Cell; 150(6):1209-22.

Buganim Y., Faddah D. A., Jaenisch R. (2013). Mechanisms and models of somatic cell reprogramming. Nature Reviews Genetics 14, 427–439

Burgess S, Reim G, Chen W, Hopkins N, Brand M. (2002). The zebrafish spiel-ohne-grenzen (spg) gene encodes the POU domain protein Pou2 related to mammalian Oct4 and is essential for formation of the midbrain and hindbrain, and for pre-gastrula morphogenesis. Development;129:5–916.

Byrum, C. A. & Martindale, M. Q. (2004). in Gastrulation: from Cells to Embryo (ed Stern, C. D.) Ch. 3, 33–50 (Cold Spring Harbor Laboratory Press, New York).

Calvopina J., Cook H., Vincent JJ, Nee K., T. Clark A.T. (2015). The Aorta-Gonad-Mesonephros Organ Culture Recapitulates 5hmC Reorganization and Replication-Dependent and Independent Loss of DNA Methylation in the Germline. Stem Cells Dev. 2015 Jul 1; 24(13): 1536–1545.

Cameron, C. B., Garey, J. R. & Swalla, B. J. (2000). Evolution of the chordate body plan: new insights from phylogenetic analyses of deuterostome phyla. Proc. Natl Acad. Sci. USA 97, 4469–4474.

Camp E, Sánchez-Sánchez AV, García-España A, Desalle R, Odqvist L, Enrique O'Connor J, Mullor JL. (2009). Nanog regulates proliferation during early fish development. Stem Cells. (9):2081-91.

Cauffman G, De RyckeM, Sermon K, Liebaers I, Van de Velde H. (2009). Markers that define stemness in ESC are unable to identify the totipotent cells in human preimplantation embryos. Hum Repro.;24:63– 70.

Cavalli G, Misteli T. 2013. Functional implications of genome topology. Nat Struct Mol Biol 20: 290–299.

Chambers, I., Colby, D., Robertson, M., Nichols, J., Lee, S., Tweedie, S. and Smith, A. (2003). Functional expression cloning of Nanog, a pluripotency sustaining factor in embryonic stem cells. *Cell* 113,643 -655.

Chambers, I. and Tomlinson, S. R. (2009). The transcriptional foundation of pluripotency. Development 136, 2311-2322.

Chazaud C, Yamanaka Y, Pawson T, Rossant J. (2006). Early lineage segregation between epiblast and primitive endoderm in mouse blastocysts through the Grb2-MAPK pathway. Dev Cell. (5):615-24.

Chen, X., Xu, H., Yuan, P., Fang, F., Huss, M., Vega, V.B., Wong, E., Orlov, Y.L., Zhang, W., Jiang, J., et al. (2008). Integration of external signaling pathways with the core transcriptional network in embryonic stem cells. Cell 133, 1106–1117.

Chen X, Cui J, Yan Z, et al. (2015). Sustained high level transgene expression in mammalian cells mediated by the optimized piggyBac transposon system. Genes Dis. (1):96–105

Crother, B. I., White, M. E., and Johnson, A. D. (2007). Inferring developmental constraint and constraint release: primordial germ cell determination mechanisms as examples. J. Theor. Biol. 248: 322–330.

Davidson E. H., Peterson K. J., Cameron R. A. (1995). Origin of Bilaterian Body Plans: Evolution of Developmental Regulatory Mechanisms Science 24.

Desbaillets, I., Ziegler, U., Groscurth, P., and Gassmann, M. (2000). Embryoid bodies: an in vitro model of mouse embryogenesis. Exp. Physiol. 85, 645–651.

de Wit E, Bouwman BA, Zhu Y, Klous P, Splinter E, Verstegen MJ, Krijger PH, Festuccia N, Nora EP, Welling M, et al. 2013. The pluripotent genome in three dimensions is shaped around pluripotency factors. Nature 501: 227–231.

Ding J, Huang X, Shao N, Zhou H, Lee DF, Faiola F, Fidalgo M, Guallar D, Saunders A, Shliaha PV, Wang H, Waghray A, Papatsenko D, Sánchez-Priego C, Li D, Yuan Y, Lemischka IR, Shen L, Kelley K, Deng H, Shen X, Wang J. (2015). Tex10 coordinates epigenetic control of super-enhancer activity in pluripotency and reprogramming. Cell Stem Cell, 16 pp. 653–668.

Dixon, J. E., Allegrucci, C., Redwood, C., Kump, K., Bian, Y., Chatfield, J., Chen, Y. H., Sottile, V., Voss, S. R., Alberio, R. et al. (2010). Axolotl Nanog activity in mouse embryonic stem cells demonstrates that ground state pluripotency is conserved from urodele amphibians to mammals. Development137, 2973-2980.

Dunn S.J., Martello G., Yordanov B., Emmott S., Smith A. G. (2014). Defining an essential transcription factor program for naïve pluripotency. Science Vol. 344, Issue 6188, pp. 1156-1160.

Efe JA, Hilcove S, Kim J, Zhou H, Ouyang K, Wang G, Chen J, Ding S. (2011). Conversion of mouse fibroblasts into cardiomyocytes using a direct reprogramming strategy. Nature Cell Biol. 13, 215–222.

Efroni S, Duttagupta R, Cheng J, Dehghani H, Hoeppner DJ, Dash C, Bazett-Jones DP, Le Grice S, McKay RD, Buetow KH, et al. (2008). Global transcription in pluripotent embryonic stem cells. Cell Stem Cell 2: 437–447.

Evans M.J, Kaufman M.H. (1981). Establishment in culture of pluripotential cells from mouse embryos. Nature, 292, pp. 154–156.

Evans, T. (2015). Investigating the association between germ line specification and sequence evolution in vertebrates. PhD thesis, University of Nottingham.

Falkner FG, Zachau HG. (1984). Correct transcription of an immunoglobulin kappa gene requires an upstream fragment containing conserved sequence elements. Nature; 310(5972):71-4.

Fernandez-Tresguerres B., Cañon S., Rayon T., Pernaute B., Crespo M., Torroja C., Manzanaresa M. (2010). Evolution of the mammalian embryonic pluripotency gene regulatory network. Proc Natl Acad Sci U S A.; 107(46): 19955–19960.

Festuccia N, Osorno R, Halbritter F, Karwacki-Neisius V, Navarro P, Colby D, Wong F, Yates A, Tomlinson SR, Chambers I. (2012). Esrrb is a direct Nanog target gene that can substitute for Nanog function in pluripotent cells. Cell Stem Cell;11(4):477-90.

Frankenberg S, Pask A, Renfree MB.The evolution of class V POU domain transcription factors in vertebrates and their characterisation in a marsupial. Dev. Biol., 337 (2009), pp. 162–170.

Frankenberg S, Renfree MB. (2013). On the origin of POU5F1. BMC Biol.; 11:56.

Frankenberg SR, Frank D, Harland R, Johnson AD, Nichols J, Niwa H, Schöler HR, Tanaka E, Wylie C, Brickman JM. (2014). The POU-er of gene nomenclature. Development; 141 (15):2921-3.

Fraser P, Bickmore W. (2007). Nuclear organization of the genome and the potential for gene regulation. Nature 447: 413–417.

Friedman R., Hughes A. L. (2001). Pattern and Timing of Gene Duplication in Animal Genomes. Genome Res.; 11(11): 1842–1847.

Fujikura J, Yamato E, Yonemura S, Hosoda K, Masui S, Nakao K, Miyazaki Ji J, Niwa H. (2002). Differentiation of embryonic stem cells is induced by GATA factors. Genes Dev. 2002 Apr 1; 16(7):784-9.

Fujita, J., Crane, A. M., Souza, M. K., Dejosez, M., Kyba, M., Flavell, R. A., Thomson, J. A. and Zwaka, T. P. (2008). Caspase activity mediates the differentiation of embryonic stem cells. Cell Stem Cell 2, 595-601.

Fussner E, Djuric U, Strauss M, Hotta A, Perez-Iratxeta C, Lanner F, Dilworth FJ, Ellis J, Bazett-Jones DP. (2011). Constitutive heterochromatin reorganization during somatic cell reprogramming. EMBO J 30: 1778–1789.

Galonska C., Ziller M. J., Karnik R., Meissner A. (2015). Ground State Conditions Induce Rapid Reorganization of Core Pluripotency Factor Binding before Global Epigenetic Reprogramming. Cell Stem Cell, 17, 462.

Gaspar-Maia A, Alajem A, Polesso F, Sridharan R, Mason MJ, Heidersbach A, Ramalho-Santos J, McManus MT, Plath K, Meshorer E, et al. 2009. Chd1 regulates open chromatin and pluripotency of embryonic stem cells. Nature 460: 863–868.

Gaspar-Maia A, Alajem A, Meshorer E, Ramalho-Santos M. 2011. Open chromatin in pluripotency and reprogramming. Nat Rev Mol Cell Biol 12: 36–47.

Golipour A, David L, Liu Y, Jayakumaran G, Hirsch CL, Trcka D, Wrana JL. (2012). A late transition in somatic cell reprogramming requires regulators distinct from the pluripotency network. Cell Stem Cell 11, 769–782.

Greber B., Wu G., Bernemann C., Joo J. Y., Han D. W., Ko K., Tapia N., Sabour D., Sterneckert J., Tesar P., et al. (2010). Conserved and divergent roles of FGF signaling in mouse epiblast stem cells and human embryonic stem cells. Cell Stem Cell 6, 215–226

Guo G., Yang J., Nichols J., Hall J. S., Eyres I., Mansfield W., Smith A. (2009). Klf4 reverts developmentally programmed restriction of ground state pluripotency. Development 136, 1063–1069

Hajkova P, Erhardt S, Lane N, Haaf T, El-Maarri O, Reik W, Walter J, Surani MA.. (2002). Epigenetic reprogramming in mouse primordial germ cells. Mech Dev.;117(1–2):15–23.

Han J, Yuan P, Yang H, Zhang J, Soh BS, Li P, Lim SL, Cao S, Tay J, Orlov YL, Lufkin T, Ng HH, Tam WL, Lim B: Tbx3 improves the germ-line competency of induced pluripotent stem cells. Nature. 463: 1096-1100.

Hanna J., Markoulaki S., Mitalipova M., Cheng A. W., Cassady J. P., Staerk J., Carey B. W., Lengner C. J., Foreman R., Love J., et al. (2009a). Metastable pluripotent states in NOD-mouse-derived ESCs. Cell Stem Cell 4, 513–524

Hanna J., Saha K., Pando B., van Zon J., Lengner C. J., Creyghton M. P., van Oudenaarden A., Jaenisch R. (2009b). Direct cell reprogramming is a stochastic process amenable to acceleration. Nature 462, 595–60

Hanna J., Cheng A. W., Saha K., Kim J., Lengner C. J., Soldner F., Cassady J. P., Muffat J., Carey B. W., Jaenisch R. (2010). Human embryonic stem cells with biological and epigenetic characteristics similar to those of mouse ESCs. Proc. Natl. Acad. Sci. USA 107, 9222–9227

Hansson J, Rafiee MR, Reiland S, Polo JM, Gehring J, Okawa S, Huber W, Hochedlinger K, Krijgsveld J. (2012). Highly coordinated proteome dynamics during reprogramming of somatic cells to pluripotency. Cell Rep. 2, 1579–1592.

Hansen C. H., van Oudenaarden A. (2013). Allele-specific detection of single mRNA molecules in situ. Nat. Meth., 10, 869.

Hart A. H., Hartley L., Ibrahim M., Robb L. (2004) Identification, cloning and expression analysis of the pluripotency promoting Nanog genes in mouse and human. Dev Dyn.;230(1):187-98.

Hermitte S., Chazaud C. (2014). Primitive endoderm differentiation: from specification to epithelium formation. Philos Trans R Soc Lond B Biol Sci.; 369(1657).

Honda, A., Hirose, M. and Ogura, A. (2009). Basic FGF and Activin/Nodal but not LIF signaling sustain undifferentiated status of rabbit embryonic stem cells. Exp. Cell Res. 315, 2033-2042.

Hotta A., and Ellis J. (2008). Retroviral vector silencing during iPS cell induction: an epigenetic beacon that signals distinct pluripotent states. J Cell Biochem.;105(4):940-8.

Hou, P., Li, Y., Zhang, X., Liu, C., Guan, J., Li, H., Zhao, T., Ye, J., Yang, W., Liu, K. et al. (2013). Pluripotent stem cells induced from mouse somatic cells by small-molecule compounds. Science 341, 651-654.

Huangfu D, Maehr R, Guo W, Eijkelenboom A, Snitow M, Chen AE, Melton DA. 2008. Induction of pluripotent stem cells by defined factors is greatly improved by small-molecule compounds. Nat Biotechnol 26: 795–797.

Jaenisch, R., and Young, R. (2008). Stem cells, the molecular circuitry of pluripotency and nuclear reprogramming. Cell 132, 567–582.

Jang S, Choubey S, Furchtgott L, Zou LN, Doyle A, Menon V, Loew EB, Krostag AR, Martinez RA, Madisen L, Levi BP. Ramanathan S. (2017). Dynamics of embryonic stem cell differentiation inferred from single-cell transcriptomics show a series of transitions through discrete cell states. Elife;6. pii: e20487.

Jauch R, Ng CK, Saikatendu KS, Stevens RC, Kolatkar PR. (2008). Crystal structure and DNA binding of the homeodomain of the stem cell transcription factor Nanog. J Mol Biol.; 376(3):758-70.

Jiang, J., Chan, Y.-S., Loh, Y.-H., Cai, J., Tong, G.-Q., Lim, C.-A., Robson, P., Zhong, S. and Ng, H.-H. (2008). A core Klf circuitry regulates self-renewal of embryonic stem cells. Nat. Cell Biol. 10, 353-360.

Johnson, A. D., Bachvarova, R. F., Drum, M. and Masi, T.

(2001). Expression of axolotl DAZL RNA, a marker of germ plasm: widespread maternal RNA and onset of expression in germ cells approaching the gonad.
Dev. Biol. 234, 402-415.

Johnson, A. D., Crother, B., White, M. E., Patient, R., Bachvarova, R. F., Drum, M. and Masi, T. (2003a). Regulative germ cell specification in axolotl embryos: a primitive trait conserved in the mammalian lineage. Philos. Trans. R.Soc. Lond. B Biol. Sci.358, 1371-1379.

Johnson, A. D., Drum, M., Bachvarova, R. F., Masi, T., White, M. E. and Crother, B. I. (2003b). Evolution of predetermined germ cells in vertebrate embryos: implications for macroevolution. Evol. Dev.5, 414-431.

Johnson AD, Alberio R. (2015) Primordial germ cells: The first cell lineage or the last cells standing? Development; 142(16):2730–2739.

Kallas A., Pook M., Trei A., Maimets T. (2014). SOX2 Is Regulated Differently from NANOG and OCT4 in Human Embryonic Stem Cells during Early Differentiation Initiated with Sodium Butyrate. Stem Cells Int.: 298163.

Kappen C, Schughart K, Ruddle FH. (1993). Early evolutionary origin of major homeodomain sequence classes. Genomics. 18(1):54-70.

Kaul-Strehlow S., Stach T. (2013) A detailed description of the development of the hemichordate Saccoglossus kowalevskii using SEM, TEM, Histology and 3D-reconstructions. Frontiers in Zoology 10:53.

Khan A, Nakamoto A, Okamoto S, Tai M, Nakayama Y, Kobayashi K, Kawamura A, Takeda H, Yamasu K. (2012). Pou2, a class V POU-type transcription factor in zebrafish, regulates dorsoventral patterning and convergent extension movement at different blastula stages. Mech Dev.;129(9-12):219-35.

Kim, J., Chu, J., Shen, X., Wang, J., and Orkin, S.H. (2008a). An extended transcriptional network for pluripotency of embryonic stem cells. Cell 132, 1049–1061.

Kim K, Doi A, Wen B., Ng K., Zhao R, Cahan P., Kim J., Aryee MJ., H Ji, L Ehrlich, A Yabuuchi, Takeuchi A, Cunniff KC, Hongguang H, Mckinney-Freeman S., Naveiras O., Yoon TJ, Irizarry RA., Jung N, Seita J., Hanna J, Murakami P, Jaenisch R., Weissleder R., Orkin SH., Weissman Il., Feinberg AP, Daley G. Q. (2010) Epigenetic memory in induced pluripotent stem cells. Nature; 467(7313): 285–290.

Kim J, Efe JA, Zhu S, Talantova M, Yuan X, Wang S, Lipton SA, Zhang K, Ding S. (2011). Direct reprogramming of mouse fibroblasts to neural progenitors. Proc. Natl Acad. Sci. USA 108, 7838–7843.

Kim, S.I., OcegueraYanez, F., Hirohata, R., Linker, S., Okita, K., Yamada, Y., Yamamoto, T., Yamanaka, S. and Woltjen, K. (2015). KLF4 N-terminal variance modulates induced reprogramming to pluripotency. Stem Cell Rep. 4, 727-743.

Kobayashi, T., Zhang, H., Tang, W. W. C., Irie, N., Withey, S., Klisch, D., Sybirna A., Dietmann S., Contreras D. A, Webb R., Allegrucci C., Alberio R., Surani, M. A. (2017). Principles of early human development and germ cell program from conserved model systems. Nature, 546(7658), 416–420.

Koziol MJ, Garrett N, Gurdon JB (2007) Tpt1 activates transcription of oct4 and nanog in transplanted somatic nuclei. Curr Biol 17: 801–807.

Kunath T., Arnaud D., Uy G. D., Okamoto I., Chureau C, Yamanaka Y., Heard E., Gardner R. L., Avner P., Rossant J. (2005). Imprinted X-inactivation in extra-embryonic endoderm cell lines from mouse blastocysts. Development (2005). 132: 1649-1661.

Kunath T., Saba-El-Leil M.K., Almousailleakh M., Wray J., Meloche S., Smith A. (2007). FGF stimulation of the Erk1/2 signalling cascade triggers transition of pluripotent embryonic stem cells from self-renewal to lineage commitment. Development, 134, pp. 2895–2902.

Lavial F, Acloque H., Bertocchini F, MacLeod D. J, Boast S, Bachelard E., Montillet G., Thenot S., Sang H. M., Stern C. D., Samarut J., Pain B. (2007). The Oct4 homologue PouV and Nanog regulate pluripotency in chicken embryonic stem cells. Development 134: 3549-3563; doi: 10.1242/dev.006569

Lee MT, Bonneau AR, Takacs CM, Bazzini AA, DiVito KR, Fleming ES, Giraldez AJ. (2013). Nanog, Pou5f1 and SoxB1 activate zygotic gene expression during the maternal-to-zygotic transition. Nature, 503 (7476), pp. 360–364 (PMCID 3925760).

Lehnertz B, Ueda Y, Derijck AA, Braunschweig U, Perez-Burgos L, Kubicek S, Chen T, Li E, Jenuwein T, Peters AH. (2003). Suv39h-mediated histone H3 lysine 9 methylation directs DNA methylation to major satellite repeats at pericentric heterochromatin. Curr Biol 13: 1192–1200.

Lemaire P. (2011). Evolutionary crossroads in developmental biology: the tunicates. Development. ;138(11):2143-52. doi: 10.1242/dev.048975.

Levine, M., and Tjian, R. (2003). Transcription regulation and animal diversity. Nature 424, 147–151.

Li R, Liang J, Ni S, Zhou T, Qing X, Li H, He W, Chen J, Li F, Zhuang Q, Qin B, Xu J, Li W, Yang J, Gan Y, Qin D, Feng S, Song H, Yang D, Zhang B, Zeng L, Lai L, Esteban MA, Pei D. (2010). A mesenchymal-to-epithelial transition initiates and is required for the nuclear reprogramming of mouse fibroblasts. Cell Stem Cell 7, 51–63.

Li M. and Belmonte J. C. I. (2017). Ground rules of the pluripotency gene regulatory network. Nature Reviews Genetics 180–191.

Li X, Liu D, Ma Y, Du X, Jing J, Wang L, Xie B, Sun D, Sun S, Jin X, Zhang X, Zhao T, Guan J, Yi Z, Lai W, Zheng P, Huang Z, Chang Y, Chai Z, Xu J, Deng H. (2017). Direct Reprogramming of Fibroblasts via a Chemically Induced XEN-like State. Cell Stem Cell, 21 (2), pp. 264-273.

Liang J, Wan M, Zhang Y, Gu P, Xin H, Jung SY, Qin J, Wong J, Cooney AJ,Liu D, Songyang Z. (2008). Nanog and Oct4 associate with unique transcriptional repression complexes in embryonic stem cells. Nat Cell Biol 10:731–739.

Loh YH, Wu Q, Chew JL, Vega VB, Zhang W, Chen X, Bourque G, George J, Leong B, Liu J, Wong KY, Sung KW, Lee CW, Zhao XD, Chiu KP, Lipovich L, Kuznetsov VA, Robson P, Stanton LW, Wei CL, Ruan Y, Lim B, Ng HH (2006). The Oct4 and Nanog transcription network regulates pluripotency in mouse embryonic stem cells. Nat Genet.; 38(4):431-40.

Luke GN, Castro LF, McLay K, Bird C, Coulson A, Holland PW. (2003). Dispersal of NK homeobox gene clusters in amphioxus and humans. Proc Natl Acad Sci U S A.;100(9):5292-5..

Lunde K, Belting HG, Driever W. (2004). Zebrafish PouVf1/pou2, homolog of mammalian Oct4, functions in the endoderm specification cascade. Curr. Biol. 14,48 -55.

Maherali N., Sridharan R., Xie W., Utikal J., Eminli S., Arnold K., Stadtfeld M., Yachechko R., Tchieu J., Jaenisch R., et al. (2007). Directly reprogrammed fibroblasts show global epigenetic remodeling and widespread tissue contribution. Cell Stem Cell 1, 55–70

Mak, W., Nesterova, T.B., de Napoles, M., Appanah, R., Yamanaka, S., Otte, A.P., Brockdorff, N. (2004). Reactivation of the paternal X chromosome in early mouse embryos. Science 303, 666–669.

Marks H., Kalkan T., Menafra R., Denissov S., Jones K., Hofemeister H., Nichols J., Kranz A., Stewart A. F., Smith A., Stunnenberg H. G. (2012). Cell, 149, 590.

Masui S, Nakatake Y, Toyooka Y, Shimosato D, Yagi R, Takahashi K, Okochi H, Okuda A, Matoba R, Sharov AA, Ko MS, Niwa H. (2007). Pluripotency governed by Sox2 via regulation of Oct3/4 expression in mouse embryonic stem cells. Nat Cell Biol. (6):625-35.

Meshorer E, Yellajoshula D, George E, Scambler PJ, Brown DT, Misteli T. (2006). Hyperdynamic plasticity of chromatin proteins in pluripotent embryonic stem cells. Dev Cell 10: 105–116.

Mikkelsen TS, Hanna J, Zhang X, Ku M, Wernig M, Schorderet P, Bernstein BE, Jaenisch R, Lander ES, Meissner A. 2008. Dissecting direct reprogramming through integrative genomic analysis. Nature 454: 49–55.

Misteli T. 2007. Beyond the sequence: cellular organization of genome function. Cell 128: 787–800.

Mitsui K, Tokuzawa Y, Itoh H, Segawa K, Murakami M, Takahashi K, Maruyama M, Maeda M, Yamanaka S. (2003). The homeoprotein Nanog is required for maintenance of pluripotency in mouse epiblast and ES cells. Cell; 113(5):631-42.

Miyanari Y., Torres-Padilla ME. (2012). Control of ground-state pluripotency by allelic regulation of Nanog. Nature, 483, 470.

Morrison, G. M. and Brickman, J. M. (2006). Conserved roles for Oct4 homologues in maintaining multipotency during early vertebrate development. Development 133,2011 -2022.

Mullin, N. P., Yates, A., Rowe, A. J., Nijmeijer, B., Colby, D., Barlow, P. N., Walkinshaw, M. D. and Chambers, I. (2008). The pluripotency rheostat Nanog functions as a dimer. Biochem. J.411, 227-231.

Mullin N., Chambers I. (2011). The Function of Nanog in Pluripotency. In: Ainscough J., Yamanaka S., Tada T. (eds) Nuclear Reprogramming and Stem Cells. Stem Cell Biology and Regenerative Medicine. Humana Press, Totowa, NJ

Nichols, J., Zevnik, B., Anastassiadis, K., Niwa, H., Klewe-Nebenius, D., Chambers, I., Schöler, H. and Smith, A. (1998). Formation of pluripotent stem cells in the mammalian embryo depends on the POU transcription factor Oct4. Cell 95, 379-391.

Nichols J, Silva J, Roode M, Smith A. (2009). Suppression of Erk signalling promotes ground state pluripotency in the mouse embryo. Development; 136(19):3215-22.

Nichols, J. and Smith, A. (2012). Pluripotency in the embryo and in culture. Cold Spring Harb. Perspect. Biol. 4, a008128.

Nieuwkoop, P. D. (1969). The formation of the mesoderm in urodele amphibians. Induction by the endoderm. Wilhelm Roux Arch.162, 341-373.

Niwa H., Burdon T., Chambers I., Smith A. (1998). Self-renewal of pluripotent embryonic stem cells is mediated via activation of STAT3. Genes Dev. 1998 Jul 1; 12(13): 2048–2060.

Niwa H., Miyazaki J., Smith A.G. (2000) Quantitative expression of Oct-3/4 defines differentiation, dedifferentiation or self-renewal of ES cells. Nat Genet.;24:372–376.

Novo C.L., Tang C., Ahmed K., Djuric U. , Fussner E. , Mullin N.P. , et al.(2016). The pluripotency factor Nanog regulates pericentromeric heterochromatin organization in mouse embryonic stem cells Genes Dev., 30, pp. 1101-1115

Okita, K., Ichisaka, T., Yamanaka, S. (2007). Generation of germline-competent induced pluripotent stem cells. Nature 448, 313–317.

Onichtchouk D1, Gawantka V, Dosch R, Delius H, Hirschfeld K, Blumenstock C, Niehrs C. (1996). The Xvent-2 homeobox gene is part of the BMP-4 signalling pathway controlling [correction of controling] dorsoventral patterning of Xenopus mesoderm. Development.;122(10):3045-53.

Orkin SH, Wang J, Kim J, Chu J, Rao S, Theunissen TW, Shen X, Levasseur DN. (2008). The transcriptional network controlling pluripotency in ES cells. Cold Spring Harb Symp Quant Biol.;73:195-202.

Papp B. and Plath K. (2011). Reprogramming to pluripotency stepwise resetting of the epigenetic landscape. Cell Res.; 21(3): 486–501.

Pashaiasl M, Khodadadi K, Kayvanjoo AH, Pashaei-asl R, Ebrahimie E, Ebrahimi M. Unravelling evolution of Nanog, the key transcription factor involved in self-renewal of undifferentiated embryonic stem cells, by pattern recognition in nucleotide and tandem repeats characteristics. Gene. 2016;578(2):194–204.

Payne, G. S., Bishop, J. M. and Varmus, H. E. (1982). Multiple arrangements of viral DNA and an activated host oncogene in bursal lymphomas. Nature 295, 209-214.

Peter I. S., Davidson E. H., (2011). Evolution of Gene Regulatory Networks that Control Embryonic Development of the Body Plan. Cell; 144(6): 970–985.

Peters AH, O'Carroll D, Scherthan H, Mechtler K, Sauer S, Schofer C, Weipoltshammer K, Pagani M, Lachner M, Kohlmaier A, et al. (2001). Loss of the Suv39h histone methyltransferases impairs mammalian heterochromatin and genome stability. Cell 107: 323–337.

Phillips, T. and Hoopes, L. (2008). Transcription factors and transcriptional control in eukaryotic cells. Nature Education 1(1):119

Polo JM, Anderssen E, Walsh RM, Schwarz BA, Nefzger CM, Lim SM, Borkent M, Apostolou E, Alaei S, Cloutier J, Bar-Nur O, Cheloufi S, Stadtfeld M, Figueroa ME, Robinton D, Natesan S, Melnick A, Zhu J, Ramaswamy S, Hochedlinger K. (2012). A molecular roadmap of reprogramming somatic cells into iPS cells. Cell 151, 1617-1632.

Radzisheuskaya A., Le Bin Chia G., Dos Santos R.L., Theunissen T.W., Castro L.F., Nichols J., Silva J.C. (2013) A defined Oct4 level governs cell state transitions of pluripotency entry and differentiation into all embryonic lineages. Nat Cell Biol.;15:579–590.

Reim G, Mizoguchi T, Stainier DY, Kikuchi Y, Brand M. (2004). The POU domain protein spg (pou2/Oct4) is essential for endoderm formation in cooperation with the HMG domain protein casanova. Developmental Cell;6:91–101.

Rodda D. J., Chew JL., Lim LH., Loh YH.,Bei Wang B., Ng HH., Robson P. (2005). Transcriptional Regulation of Nanog by OCT4 and SOX2. The Journal of Biological Chemistry280, 24731-24737.

Rodaway A., Patient R. (2001). Mesendoderm: An Ancient Germ Layer? Volume 105, Issue 2, Pages 169-172.

Samavarchi-Tehrani P, Golipour A, David L, Sung HK, Beyer TA, Datti A, Woltjen K, Nagy A, Wrana JL. (2010). Functional genomics reveals a BMP-driven mesenchymal-to-epithelial transition in the initiation of somatic cell reprogramming. Cell Stem Cell 7, 64–77.

Sato N, Meijer L, Skaltsounis L, Greengard P, Brivanlou AH. (2004). Maintenance of pluripotency in human and mouse embryonic stem cells through activation of Wnt signaling by a pharmacological GSK-3-specific inhibitor. Nat Med. 2004 Jan; 10(1):55-63.

Scerbo P, Girardot F, Vivien C, Markov GV, Luxardi G, et al. (2012) Ventx factors function as Nanog-like guardians of developmental potential in Xenopus. PLoS One 7: e36855.

Scerbo P, Markov GV, Vivien C, Kodjabachian L, Demeneix B, Coen L, et al. (2014). On the Origin and Evolutionary History of NANOG. PLoS ONE 9(1): e85104.

Schneider, M. R., Adler, H., Braun, J., Kienzle, B., Wolf, E. and Kolb, H. J. (2007). Canine embryo-derived stem cells-toward clinically relevant animal models for evaluating efficacy and safety of cell therapies.
Stem Cells25, 1850-1851.

Schuff, M., Siegel, D., Philipp, M., Bundschu, K., Heymann, N., Donow, C. & Knöchel, W. (2011). Characterization of Danio rerio Nanog and functional comparison to Xenopus Vents. Stem Cells Dev. 21, 1225–1238.

Schwarz BA, Bar-Nur O, Silva JC, Hochedlinger K. (2014). Nanog is dispensable for the generation of induced pluripotent stem cells. Curr Biol 24: 347–350.
Shu, J., Wu, C., Wu, Y., Li, Z., Shao, S., Zhao, W., Tang, X., Yang, H., Shen, L., Zuo, X. et al. (2013). Induction of pluripotency in mouse somatic cells with lineage specifiers. Cell 153, 963-975.

Shi W, Wang H, Pan G, Geng Y, Guo Y, Pei D. Regulation of the pluripotency marker Rex-1 by Nanog and SOX2. J Biol Chem. 2006 Aug 18;281(33):23319-25.

Silva, J., Barrandon, O., Nichols, J., Kawaguchi, J., Theunissen, T. W. and Smith, A. (2008). Promotion of reprogramming to ground state pluripotency by signal inhibition. PLoS Biol.6, e253.

Silva J, Nichols J, Theunissen TW, Guo G, van Oosten AL, Barrandon O, Wray J, Yamanaka S, Chambers I, Smith A. (2009). Nanog is the gateway to the pluripotent ground state. Cell ;138(4):722-37.

Singh AM, Hamazaki T, Hankowski KE, Terada N. (2007). A heterogeneous expression pattern for Nanog in embryonic stem cells. Stem Cells; 25(10):2534-42.

Smith, J. J., Kump, D. K., Walker, J. A., Parichy, D. M. and Voss, S. R. (2005). A comprehensive expressed sequence tag linkage map for tiger salamander and Mexican axolotl: enabling gene mapping and comparative genomics in Ambystoma. Genetics 171, 1161-1171.

Smith ZD, Nachman I, Regev A, Meissner A. (2010). Dynamic single-cell imaging of direct reprogramming reveals an early specifying event. Nat Biotechnol; 28(5):521-6.

Soufi, A., Donahue, G. and Zaret, K. S. (2012). Facilitators and impediments of the pluripotency reprogramming factors' initial engagement with the genome. Cell 151, 994-1004.

Sridharan R, Gonzales-Cope M, Chronis C, Bonora G, McKee R, Huang C, Patel S, Lopez D, Mishra N, Pellegrini M, et al. (2013). Proteomic and genomic approaches reveal critical functions of H3K9 methylation and heterochromatin protein-1γ in reprogramming to pluripotency. Nat Cell Biol 15: 872–882.

Stadtfeld M, Maherali N, Breault DT, Hochedlinger K. (2008). Defining molecular cornerstones during fibroblast to iPS cell reprogramming in mouse. Cell Stem Cell; 2(3):230-40.

Ström F. S., Pu Zhang, Stavreus-Evers A. (2012). Co-localization of NANOG and OCT4 in human pre-implantation embryos and in human embryonic stem cells. J Assist Reprod Genet. (10): 1021–1028.

Sutasurya, L. A. and Nieuwkoop, P. D. (1974). The induction of the primordial germ cells in the urodeles. Wilhelm Roux Arch. 175, 199-220.

Tada, M., Takahama, Y., Abe, K., Nakatsuji, N. and Tada, T. (2001). Nuclear reprogramming of somatic cells by in vitro hybridization with ES cells. Curr. Biol. 11, 1553-1558.

Takahashi K., and Yamanaka S. (2006). Induction of Pluripotent Stem Cells from Mouse Embryonic and Adult Fibroblast Cultures by Defined Factors Cell Volume 126, Issue 4, Pages 663–676.

Takahashi K, Okita K, Nakagawa M, Yamanaka S (2007) Induction of pluripotent stem cells from fibroblast cultures. Nat Protoc 2: 3081–3089.

Takahashi, K., Tanabe, K., Ohnuki, M., Narita, M., Sasaki, A., Yamamoto, M., Nakamura, M., Sutou, K., Osafune, K. and Yamanaka, S. (2014). Induction of pluripotency in human somatic cells via a transient state resembling primitive streak-like mesendoderm. Nat. Commun. 5, 3678.

Takahashi K., and Yamanaka S. (2015). A developmental framework for induced pluripotency. Development 142: 3274-3285.

Tanabe, K., Nakamura, M., Narita, M., Takahashi, K. and Yamanaka, S. (2013). Maturation, not initiation, is the major roadblock during reprogramming toward pluripotency from human fibroblasts. Proc. Natl. Acad. Sci. USA 110,12172-12179.

Tang F., Barbacioru C., Bao S., Lee C., Nordman E., Wang X, Lao K., Surani A. (2010). Tracing the Derivation of Embryonic Stem Cells from the Inner Cell Mass by Single-Cell RNA-Seq Analysis. Cell stem Cell, Vol6, Issue 5, Pages 468–478.

Tesar, P. J., Chenoweth, J. G., Brook, F. A., Davies, T. J., Evans, E. P., Mack, D. L., Gardner, R. L. and McKay, R. D. (2007). New cell lines from mouse epiblast share defining features with human embryonic stem cells. Nature448,196-199.

Tiemann, U., Sgodda, M., Warlich, E., Ballmaier, M., Schöler, H. R., Schambach, A. and Cantz, T. (2011). Optimal reprogramming factor stoichiometry increases colony numbers and affects molecular characteristics of murine induced pluripotent stem cells. Cytometry A 79A, 426-435.

Theunissen T. W., Costa Y., Radzisheuskaya A., van Oosten A. L., Lavial F., Pain B., Castro L. F. and Silva J. C. (2011). Reprogramming capacity of Nanog is functionally conserved in vertebrates and resides in a unique homeodomain. Development 138, 4853-4865. 10.1242/dev.06877

Theunissen T. W., Silva J. C. R. (2011). Switching on pluripotency: a perspective on the biological requirement of Nanog. Philos. Trans. Royal Soc. London Ser. B Biol. Sci. 366, 2222–2229.

Theunissen T. W., van Oosten A. L., Castelo-Branco G., Hall J., Smith A., Silva J. C. (2011). Nanog overcomes reprogramming barriers and induces pluripotency in minimal conditions. Curr. Biol. 21, 65–7.

Thomson, J. A., Itskovitz-Eldor, J., Shapiro, S. S., Waknitz, M. A., Swiergiel,

J. J., Marshall, V. S. and Jones, J. M. (1998). Embryonic stem cell lines derived from human blastocysts. Science 282, 1145-1147.

Tokuzawa Y, Kaiho E, Maruyama M, Takahashi K, Mitsui K, Maeda M, Niwa H, Yamanaka S. (2003). Fbx15 Is a Novel Target of Oct3/4 but Is Dispensable for Embryonic Stem Cell Self-Renewal and Mouse Development. Mol Cell Biol.; 23(8): 2699–2708.

Virta VC, Cooper MS. (2011). Structural components and morphogenetic mechanics of the zebrafish yolk extension, a developmental module. J Exp Zool B Mol Dev Evol.;316(1):76-92.

Wang, J., Rao, S., Chu, J., Shen, X., Levasseur, D.N., Theunissen, T.W., and Orkin, S.H. (2006). A protein interaction network for pluripotency of embryonic stem cells. Nature 444, 364–368.

Wang, J., Levasseur, D. N. and Orkin, S. H. (2008). Requirement of Nanog dimerization for stem cell self-renewal and pluripotency. Proc. Natl. Acad. Sci. USA 105, 6326-6331.

Wang, H., Liu, Y., Ye, D., Li, J., Liu, J. and Deng, F. (2016), Knockdown of zebrafish Nanog increases primordial germ cells during early embryonic development. Develop. Growth Differ., 58: 355–366. doi:10.1111/dgd.12279

Wernig M., Meissner A., Foreman R., Brambrink T., Ku M., Hochedlinger K., Bernstein B. E., Jaenisch R. (2007). In vitro reprogramming of fibroblasts into a pluripotent ES-cell-like state. Nature 448, 318–324.

Wijchers PJ, Geeven G, Eyres M, Bergsma AJ, Janssen M, Verstegen M, Zhu Y, Schell Y, Vermeulen C, de Wit E, et al. 2015. Characterization and dynamics of pericentromere-associated domains in mice. Genome Res 25: 958–969.

Williams R.L., Hilton D.J., Pease S., Willson T.A., Stewart C.L., Gearing D.P., Wagner E.F., Metcalf D., NicolaN.A., Gough N.M. (1988). Myeloid leukaemia inhibitory factor maintains the developmental potential of embryonic stem cells. Nature 336:684–687.

Whyte, W.A., Orlando, D.A., Hnisz, D., Abraham, B.J., Lin, C.Y., Kagey, M.H., Rahl, P.B., Lee, T.I., and Young, R.A. (2013). Master transcription factors and mediator establish super-enhancers at key cell identity genes. Cell 153, 307–319.

Wray J., Kalkan T., Smith A. G. (2010). The ground state of pluripotency Biochemical Society Transactions, 38(4)1027-1032.

Xu C. Fan Z. P., Muller P., Fogley R., DiBiase A., Trompouki E., Unternaehrer J., Xiong F., Torregroza I., Evans T., Megason S. G. (2012). Nanog-like Regulates Endoderm Formation through the Mxtx2-Nodal Pathway. Developmental cell Volume 22, Issue 3, Pages 625–638.

Yamanaka, S. (2009). Elite and stochastic models for induced pluripotent stem cell generation. Nature 460, 49-52.

Yamanaka Y, Lanner F, Rossant J. (2010). FGF signal-dependent segregation of primitive endoderm and epiblast in the mouse blastocyst. Development; 137(5):715-24.

Yamaguchi, S., Hirano, K., Nagata, S. and Tada, T. (2010). Sox2 expression effects on direct reprogramming efficiency as determined by alternative somatic cell fate. Stem Cell Res. 6, 177-186.

Yeo JC, Jiang J, Tan ZY, Yim GR, Ng JH, Göke J, Kraus P, Liang H, Gonzales KA, Chong HC, Tan CP, Lim YS, Tan NS, Lufkin T, Ng HH. (2014). Klf2 is an essential factor that sustains ground state pluripotency. Cell Stem Cell.;14(6):864-72

Ying QL, Nichols J, Chambers I, Smith A. (2003). BMP induction of Id proteins suppresses differentiation and sustains embryonic stem cell self-renewal in collaboration with STAT3. Cell; 115(3):281-92.

Ying Q.L , Wray J., Nichols J., Batlle-Morera L., Doble B., Woodgett J., Cohen P., Smith A. (2008). The ground state of embryonic stem cell self-renewal. Nature, 453, pp. 519–523.

Young R. (2011). Control of the Embryonic Stem Cell State. Cell Volume 144, Issue 6, Pages 940–954.

Zhou Y, Li S, Huang Q, Xie L, Zhu X. (2013). Nanog suppresses cell migration by downregulating thymosin beta4 and Rnd3. J Mol Cell Biol.