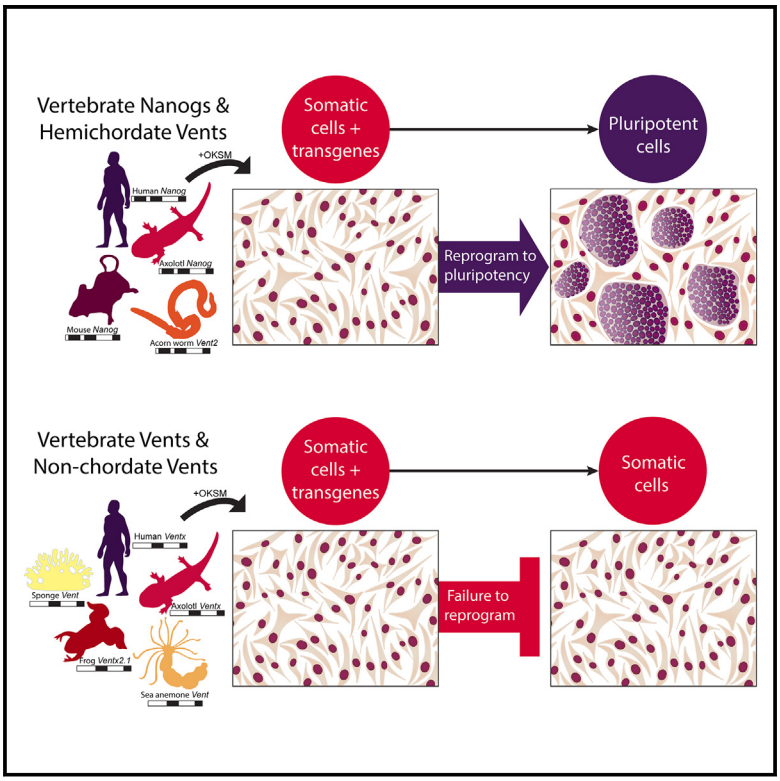
Cell Reports

Article

Programming of pluripotency and the germ line co-evolved from a Nanog ancestor

Graphical abstract



Highlights

- A *Vent* gene from an acorn worm possesses *Nanog*-like programming activity
- The *Vent gene* from a sea anemone supports self-renewal but fails to reprogram
- Vent and Nanog activities appear to sub-functionalize in vertebrates
- Nanog activity evolves coincidently with the evolution of mesoderm

Authors

Darren Crowley, Luke Simpson, Jodie Chatfield, ..., Matthew Loose, Ramiro Alberio, Andrew D. Johnson

Correspondence

darcrowley@gmail.com (D.C.), ramiro.alberio@nottingham.ac.uk (R.A.)

In brief

Crowley et al. investigated the evolutionary origins of the pluripotency gene *Nanog*. Their research suggests that a single *Vent* gene capable of programming both pluripotency and germline competency was present in the last common ancestor of deuterostomes. In vertebrates, these activities sub-functionalized into the *Vent* and *Nanog* gene families, respectively.

Crowley et al., 2025, Cell Reports *44*, 115396
March 25, 2025
https://doi.org/10.1016/j.celrep.2025.115396

CellPress

Cell Reports

CellPress
OPEN ACCESS

Article

Programming of pluripotency and the germ line co-evolved from a Nanog ancestor

Darren Crowley,[1,6,*] Luke Simpson,[1,6] Jodie Chatfield,[1] Teri Forey,[1] Cinzia Allegrucci,[1,7] Fei Sang,[1] Nadine Holmes,[1] Grigory Genikhovich,[2] Ulrich Technau,[2] Doreen Cunningham,[3] Elena Silva,[3] Nicholas Mullin,[4] James E. Dixon,[1,8] Matthew Loose,[1] Ramiro Alberio,[5,10,*] and Andrew D. Johnson[1,9]

[1]School of Life Sciences, University of Nottingham, Queens Medical Centre, Nottingham NG7 2UH, UK
[2]Department of Neurosciences and Developmental Biology, Faculty of Life Sciences, Vienna BioCenter, Djerassiplatz 1, 1030 Vienna, Austria
[3]Department of Biology, Georgetown University, Washington, D.C, USA
[4]Centre for Regenerative Medicine, University of Edinburgh, Edinburgh EH16 4UU, UK
[5]School of Biosciences, University of Nottingham, Sutton Bonington Campus, Loughborough LE12 5RD, UK
[6]These authors contributed equally
[7]Present address: School of Veterinary Medicine and Science, University of Nottingham, Sutton Bonington Campus, Loughborough LE12 5RD, UK
[8]Present address: School of Pharmacy, University of Nottingham, Nottingham NG7 2UH, UK
[9]Deceased
[10]Lead contact
*Correspondence: darcrowley@gmail.com (D.C.), ramiro.alberio@nottingham.ac.uk (R.A.)
https://doi.org/10.1016/j.celrep.2025.115396

SUMMARY

Francois Jacob proposed that evolutionary novelty arises through incremental tinkering with pre-existing genetic mechanisms. Vertebrate evolution was predicated on pluripotency, the ability of embryonic cells to form somatic germ layers and primordial germ cells (PGCs). The origins of pluripotency remain unclear, as key regulators, such as *Nanog*, are not conserved outside of vertebrates. Given NANOG's role in mammalian development, we hypothesized that NANOG activity might exist in ancestral invertebrate genes. Here, we find that *Vent* from the hemichordate *Saccoglossus kowalevskii* exhibits NANOG activity, programming pluripotency in *Nanog*$^{-/-}$ mouse pre-induced pluripotent stem cells (iPSCs) and NANOG-depleted axolotl embryos. *Vent* from the cnidarian *Nematostella vectensis* showed partial activity, whereas *Vent* from sponges and vertebrates had no activity. VENTX knockdown in axolotls revealed a role in germline-competent mesoderm, which *Saccoglossus Vent* could rescue but *Nematostella Vent* could not. This suggests that the last deuterostome ancestor had a *Vent* gene capable of programming pluripotency and germline competence.

INTRODUCTION

Pluripotency describes the ability of embryonic cells to give rise to both the soma and primordial germ cells (PGCs). The genetic and epigenetic regulation of these early embryonic cells determines the cellular complexity of the adult organism. Thus, the pluripotent state represents the raw material for phenotypic diversity. While pluripotency-related genes display a high degree of evolutionary conservation between eutherian mammals,[1,2] studying the gene regulatory network (GRN) of pluripotency outside of this clade is difficult due to poor genetic sequence conservation and gene duplications that often hamper ortholog identification.[2–7] Further, the evolution of clade-specific aspects of development, such as the formation of extra-embryonic lineages in mammals, makes direct comparisons of pluripotency gene knockout phenotypes difficult across different animal groups.

The evolution of the pluripotency gene *Nanog* is poorly understood; in mammals, *Nanog* appears to have a conserved expression profile demarcating the pluripotent cells of the inner cell mass (ICM) and pre-implantation epiblast.[8–12] NANOG is a master regulator of pluripotency *in vitro*, required for leukemia inhibitory factor (LIF)-independent self-renewal of naive mouse embryonic stem cells (mESCs)[8] and the reprogramming of somatic cells to a pluripotent state.[13] Critically, NANOG is fundamental in the establishment of pluripotency in pre-implantation embryos[9,11,14] and germline survival in mammals.[15]

Outside of mammals, the role of NANOG is less clear. Nanog is expressed in chicken epiblasts,[16,17] axolotl animal caps,[18] and zebrafish blastomeres.[19] Axolotl NANOG (AxNANOG) is required for gastrulation and lineage commitment[20] and is capable of maintaining LIF-independent pluripotency when dimerized.[18] While overall *Nanog* sequence conservation is low, the ability to facilitate the programming of pluripotency *in vitro* to pluripotency resides in its homeodomain (HD), which appears to be highly conserved across vertebrates.[2] The observation that zebrafish NANOG is capable of programming pluripotency in *Nanog*-null pre-induced pluripotent stem cells (piPSCs) suggests that the role of NANOG in governing the programming of pluripotency is conserved from the last common ancestor of vertebrates spanning some 429 million years.[2,21]

Despite the critical importance of NANOG in vertebrate development, no orthologs have been identified outside of vertebrates. Phylotypic analysis has suggested that *Nanog* may have evolved from the *Bsx*, *Nkx*,[2] or *Ventx*[22,23] gene family; however, no putative *Nanog* ancestors have ever been shown to be capable of functionally replacing a vertebrate *Nanog*. Given the importance of pluripotency and the central role of *Nanog* in vertebrate pluripotency, we searched for *Nanog* antecedents capable of programming pluripotency, focusing on animals with conditionally specified PGCs.[24–27]

Here, we show that pluripotency in *Nanog*$^{-/-}$ mouse piPSCs and NANOG-depleted axolotl embryos can be programmed with a *Vent* from the acorn worm *Saccoglossus kowalevskii* but not by *Vent* from the sea anemone *Nematostella vectensis.* The *Vent* gene from a more basal organism, the sponge *Ephydatia fluviatilis*, and vertebrate *Vent* genes showed no apparent role in the regulation of the pluripotent state. *VENTX* knockdown (KD) in axolotl embryos suggests a role of vertebrate *Vents* in establishing germline-competent mesoderm and that PGC development can be rescued by acorn worm, but not sea anemone, *Vent*. These findings suggest that a single precursor of *Vent* and *Nanog* genes capable of programming both germline-competent and pluripotent cell states was present in the last common ancestor of deuterostomes. In vertebrates, these activities have sub-functionalized within the *Vent* and *Nanog* gene families, respectively.

RESULTS

Ancient *Vent* genes possess *Nanog*-like roles

Nanog encodes an Nk-like (NKL) class HD transcription factor that programs pluripotency in vertebrate embryos.[2,8,9,13,28] Indeed, the *Nanog* HD is sufficient to program pluripotency in *Nanog*-null neural stem cell-derived piPSCs (NnNpiPSCs).[2] However, *Nanog* shows relatively low sequence conservation outside its HD and does not appear to be conserved outside of vertebrates.[2] We searched for *Nanog* ancestors focusing on HD conservation between the *Nanog*, *Vent*, and *Bsx* families rooted from human *Nkx2.5* (Figure S1). Phylogenetic information was used to estimate the evolutionary distance between each transcription factor from the observed differences in their sequences. In line with previous reports, invertebrate *Vent* genes were grouped with vertebrate *Nanog* and *Vent* orthologs, suggesting a homology of *Nanog* and *Vent*, rather than with the *Bsx* homeobox protein family (Figure S1).[22,23] Bayesian phylogeny also traced the *Vent* family to the metazoan origin (Figure S1); these results suggest that vertebrate *Nanog* and *Ventx* genes share a common *Vent*-like ancestor (herein referred to as *Vent*). We selected invertebrate *Vent* genes representing specific milestones in metazoan evolution, avoiding species known to contain germplasm or divergent modes of PGC specification,[7,24–27,29–33] and tested their capacity to maintain stemness by overexpression in mESCs followed by LIF withdrawal (see STAR Methods; Figures 1B and 1C). Given that NANOG proteins outside of mammals act as homodimers and that dimerization appears to be critical to the maintenance of the pluripotent state in mouse,[34,35] a tryptophan repeat domain (WR) was added to each *Vent* gene to drive homodimerization.[18,34,35] Positive control cell lines expressed vertebrate *Nanog* genes from mouse (*MNanog*) and axolotl (*AxNanog-WR*).[18] Cell lines expressing *Vent* genes from axolotl (*AxVentx-WR*), *Xenopus laevis* (*XVent2-WR*), and human (*HVentx-WR*) were established to test whether vertebrate *Vents* possess any ability to maintain pluripotency in the absence of LIF and were compared to those expressing *Vent* genes from *Saccoglossus kowalevskii* (*SkVent1-WR* and *SKVent2-WR*), *Nematostella vectensis* (*NvVent-WR*), and the freshwater sponge *Ephydatia fluviatilis* (*EfVent-WR*). Cells were then cultured in the presence or absence of LIF, and self-renewal was assessed by alkaline phosphatase (AP) staining (Figure 1C).[2] Colonies from cells expressing vertebrate *Nanog* genes stained positive for AP at least 13 passages after LIF withdrawal. Remarkably, *SkVent1-WR*, *SkVent2-WR*, and *NvVent-WR* supported the growth of AP+ colonies under the same conditions. Vertebrate *Vent* genes and *EfVent-WR*, in contrast, did not support self-renewal without LIF. These data suggest that the ability to maintain pluripotency evolved in *Vent* genes from primitive metazoans but was lost in vertebrate *Vent* after gene duplication, giving rise to *Nanog*.

Reprogramming with invertebrate Vent genes

We next tested whether constitutive expression of invertebrate *Vent* genes could reprogram NnNpiPSCs (Figure S2), which are unable to survive the transition to serum-free 2i (small-molecule inhibitors PD0325901 and CHIR99021) + LIF conditions (2i+LIF) and establish ground-state pluripotency in the absence of NANOG activity.[13] Complete reprogramming in these cells is indicated by G418 resistance and expression of LacZ, demonstrating bi-allelic activation of the *Nanog* locus.[13] NnNpiPSCs transfected with each test construct and passaged in serum-LIF until piPSC colonies emerged were transitioned to 2i+LIF in the presence of G418 selection (Figures S2A and S2B), and surviving colonies were subjected to 5-bromo-4-chloro-3-indolyl-beta-D-galacto-pyranoside (X-GAL) and AP staining[2] (Figure S2C). Cells expressing either the mouse or axolotl *Nanog* construct formed colonies after transitioning to 2i+LIF, and these colonies showed AP and LacZ expression (Figure S2C). *SkVent2-WR* also mediated stable cell reprogramming, albeit at a lower efficiency than the *Nanog* genes (Figures S2B–S2E). Interestingly, *NvVent-WR* showed partial reprogramming activity, indicated by the appearance of small, round, refractile cells throughout the culture; however, these cells did not survive the transition to 2i+LIF. Cells from *SkVent1-WR* and *XVent2-WR* cultures, in contrast, resembled untransduced control cultures and died after G418 treatment (Figure S2B).

We next tested if *NvVent* and *SkVent2* could reprogram cells in their native forms without WR repeats (Figure 2). We created NnNpiPSC lines stably expressing these factors, using *MNanog* and *XVent2* as positive and negative controls, respectively. Cells expressing *MNanog* and *SkVent2* again formed colonies in serum-LIF, while cells expressing *XVent2* remained unchanged (Figure 2A). *NvVent* induced the formation of small, round, refractile cells in serum-LIF, similar to its dimerized form (Figure S2B). Cells expressing *SkVent2* and *MNanog* expanded in 2i+LIF, while the transition to 2i+LIF again resulted in cell death in *NvVent* cultures. However, these results showed a fraction of

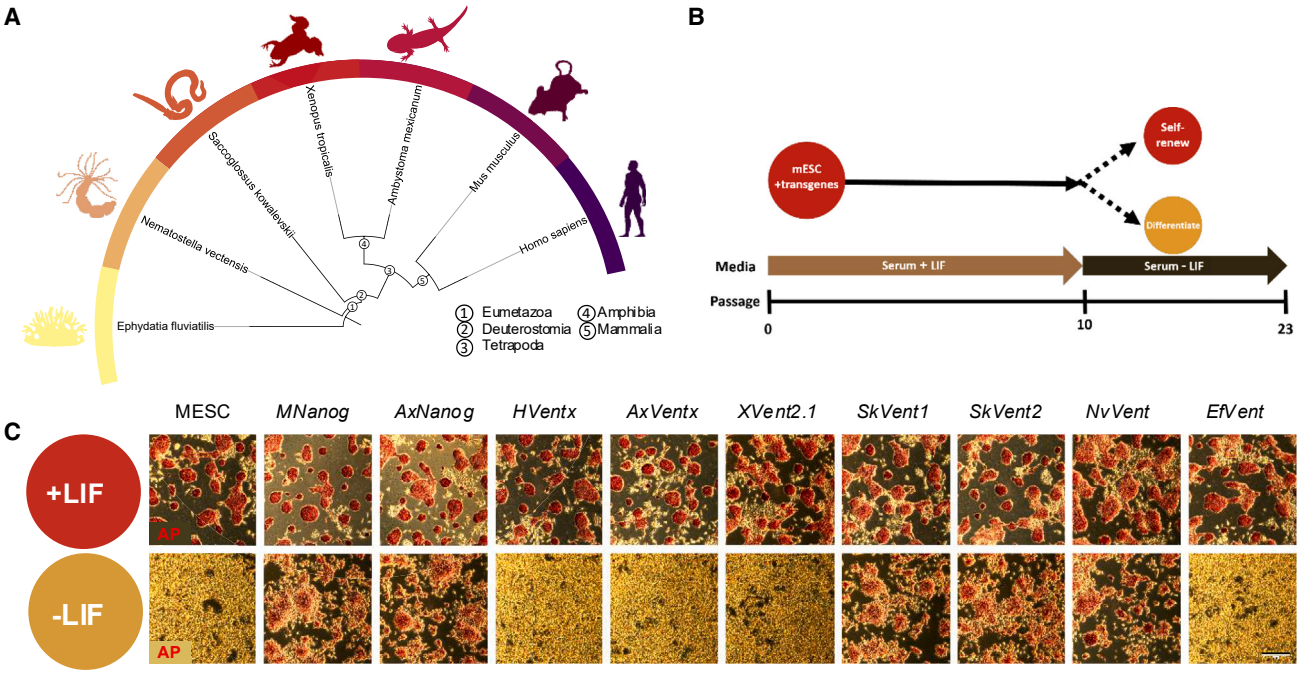


Figure 1. Ancient *Vent* genes convey *Nanog*-like cytokine independence

Ancient *Vents* can be traced back to the origin of metazoans.

(A) Phylogenetic tree with organisms selected for testing highlighted.

(B) Outline of the LIF withdrawal assay.

(C) mESC cultures overexpressing *Nanog* and *Vent* candidate genes, cultured in the presence or absence of LIF and stained for alkaline phosphatase activity after 13 passages. Red staining indicates the expression of alkaline phosphatase. Scale bar, 60 μm.

mESC, mouse embryonic stem cells; *MNanog*, mouse *Nanog*; *AxNanog*, axolotl *Nanog*; *HVentx*, human *Ventx*; *XVent2.1*, *Xenopus Vent2.1*; *SkVent1/2*, *Saccoglossus kowalevskii* (acorn worm) *Vent1/2*; *NvVent*, *Nematostella vectensis* (sea anemone) *Vent*; *EfVent*, *Ephydatia fluviatilis* (sponge) *Vent*.

the surviving cells staining AP+ (Figure S4B), suggesting some reprogramming activity. Cultures expressing *XVent2* did not survive G418 selection (Figure 2A).

Several colonies from cells expressing *SkVent2* survived the transition to 2i+LIF with G418 resistance while staining positive for both AP and LacZ (Figure 2B) along with *MNanog* cultures. Colony formation from *SkVent2* cultures occurred at a lower efficiency than *MNanog* (Figure 2C), but colonies from both cultures were stably maintained in 2i+LIF for several passages, enabling the establishment of iPSC lines (Figure 2A).

We next compared the expression levels of naive pluripotency markers in iPSC lines induced by *SkVent2* or *MNanog* wild-type mESCs and untransduced NnNpiPSCs (Figure 2D). All iPSCs showed activation of naive pluripotency factors *Pou5f1*, *Klf2*, *Rex1*, *Esrrb*, *Nrob1*, *Tbx3*, and *Fbx15* (herein referred to as naive pluripotency markers), indicating that the cells were reprogrammed to the ground state by both *MNanog* and *SkVent2* in their native forms. Notably, most genes in iPSCs demonstrated elevated expression levels relative to mESCs, suggesting their expression was upregulated by both the *MNanog* and *SkVent2* transgenes. qPCR also showed that the three retroviral Yamanaka factors expressed in the piPSCs were silenced in cells collected after ten or more passages (Figure 2E), verifying full reprogramming to the ground state.[36,37] We then tested iPSCs from both the *MNanog* and *SkVent2* lines for their spontaneous differentiation into the three germ layers using embryoid bodies (EBs) (Figure 2F). EBs from both the *MNanog* and *SkVent2* lines showed expression of markers representative of all three germ layers (Figure 2G). *SkVent2*-reprogrammed cells displayed a higher level of induction of *Sox17* relative to both *MNanog* iPSCs and wild-type mESCs. Mesodermal differentiation was confirmed in EBs from all three cell lines by *Myf5* and *T* expression, and *NeuroD* activation demonstrated the induction of neural genes.

The reprogramming capacity of *SkVent2* and *NvVent* was next tested on a piPSC mouse embryonic fibroblast (MEF) cell line containing a GFP-PURO reporter (TNG MEF)[38] to track the activation of the endogenous *Nanog* locus in real time (Figure S3). *SkVent2*-reprogrammed cells were identified by GFP+ colonies and activated naive pluripotency genes (Figures S3B and S3D). MEFs reprogrammed by *NvVent* again showed a propensity to produce a round, refractile cell type but did not activate GFP expression (data not shown). As a subset of the small, rounded cell type in piPSCs expressing *NvVent* demonstrated positive staining for AP (Figure S4B), we investigated this partial reprogramming in more detail.

The small, round, refractile phenotype induced by *NvVent* superficially resembles XEN cells, which are derived from the extra-embryonic endoderm of mouse blastocysts.[39] Moreover, XEN-like cells (often referred to as iXEN cells) arise in

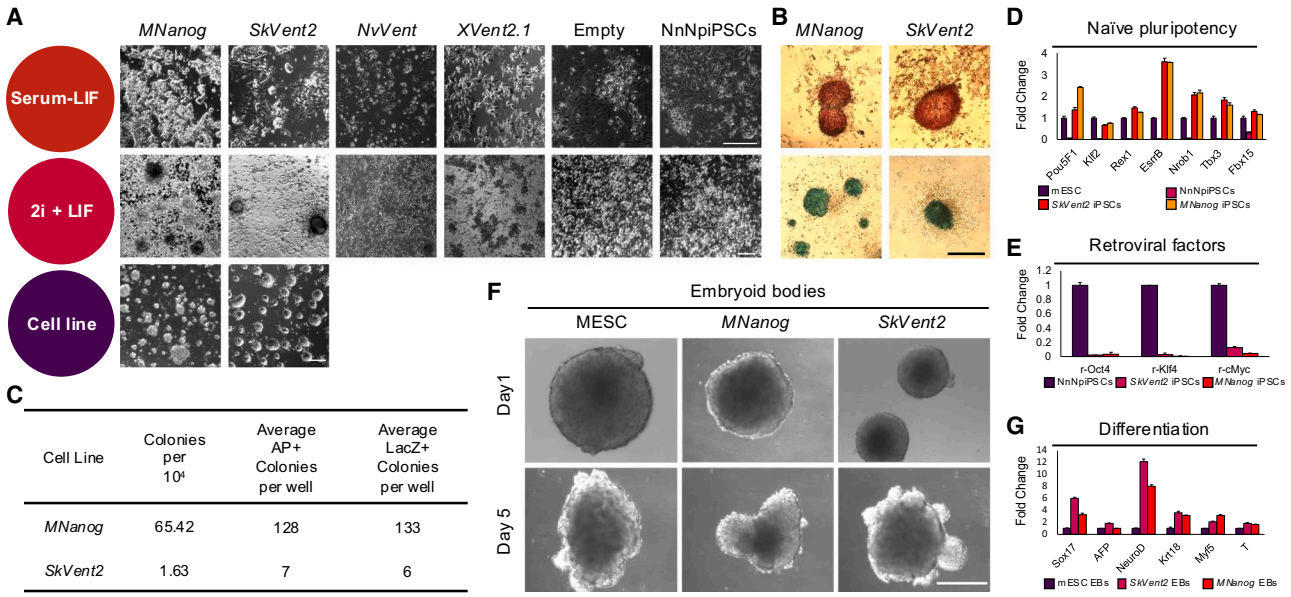


Figure 2. Ancient *Vent* monomers can program pluripotency

(A) piPSC lines expressing candidate genes in serum-LIF (top) and after the transition to 2i+LIF (middle) were passaged to establish fully reprogrammed iPSC lines (bottom). Scale bar: 100 μm.

(B) AP (top) and LacZ (bottom) staining of *MNanog*- and *SkVent2*-reprogrammed iPSCs. Scale bar: 100 μm.

(C) Reprogramming efficiencies. Values represent the averages across three independent experiments.

(D) RT-qPCR analysis showing expression of endogenous pluripotency markers in reprogrammed cells. Relative to *Actb*. Error bars indicate SD from three technical replicates.

(E) RT-qPCR analysis showing reprogrammed cells silence retroviral exogenous factors. Relative to *Actb*. Error bars indicate SD from three technical replicates.

(F) iPSCs form embryoid bodies. Scale bar: 200 μm.

(G) RT-qPCR analysis showing gene expression analysis of three somatic germ layers in embryoid bodies. Relative to *Actb*. Error bars indicate SD from three technical replicates.

mESC, mouse embryonic stem cells; *MNanog*, mouse *Nanog*; *AxNanog*, axolotl *Nanog*; *HVentx*, human *Ventx*; *XVent2.1*, *Xenopus Vent2.1*; *SkVent1/2*, *Saccoglossus kowalevskii* (acorn worm) *Vent1/2*; *NvVent*, *Nematostella vectensis* (sea anemone) *Vent*; *EfVent*, *Ephydatia fluviatilis* (sponge) *Vent*; EBs, embryoid bodies.

parallel to iPSCs as a result of partial OSKM (*Oct4*, *Sox2*, *Klf4*, *c-Myc*) reprogramming,[40] suggesting that our XEN-like cells could be products of partial reprogramming by *NvVent*. In line with previous reports,[40] survival of the XEN-like cells was extracellular signal-regulated kinase (ERK) dependent (Figure S4A), thus explaining why they could not survive 2i+LIF. We then compared gene expression in cells programmed with *NvVent* to XEN cells and untransduced NnNpiPSCs (Figure S4E). Endoderm markers *Sox7*, *Sox17*, *Gata6*, and *Foxa2* are specifically expressed in XEN cells but not in untransduced piPSCs.[40] Gene expression analysis revealed the distinct upregulation of *Sox17*, *Gata6*, and *Foxa2*, but not *Sox7*, in NnNpiPSC lines transfected with *NvVent* relative to untransduced piPSCs, albeit expression levels were lower compared to pure XEN cell cultures; nonetheless, cells expressing *NvVent* displayed a similar expression pattern, with *Sox17* being the most highly expressed marker. Together, these data suggest that *NvVent* can drive the partial reprogramming of the NnNpiPSCs but that these cells fail to establish naive pluripotency and instead divert to a XEN-like state. This activity stands in marked contrast to the effects observed with acorn worm, sponge, and vertebrate genes.

Characterizing Vent expression in invertebrates

Given the reprogramming capability of *SkVent2* and the partial reprogramming activity of *NvVent*, we next investigated the expression profiles of ancient *Vent* genes in their native organisms (Figure 3). In acorn worm, *in situ* hybridization (ISH) revealed distinct expression patterns of *SkVent1* and *SkVent2* during early development. At early gastrulation (18 h post-fertilization [hpf]), both *Vent* genes were expressed in the cells of the presumptive ectoderm, which will give rise to ectodermal, mesodermal, and endodermal cells (Figure 3A). By mid-gastrulation (22 hpf), both genes showed a speckled expression pattern across the ectoderm. However, by early neurulation (38 hpf), neither gene was expressed. Semi-qPCR further revealed their temporal dynamics (Figure 3B). *SkVent1* was maternally inherited, with expression detectable prior to fertilization and throughout gastrulation but extinguished shortly after, in line with ISH data. *SkVent2*, on the other hand, was expressed zygotically around 12 hpf during the blastula stages. Notably, the expression profile of *SkVent2*—which is capable of reprogramming—parallels that of vertebrate *Nanog*s, such as axolotl.[18,20]

To explore *NvVent* expression in sea anemone, we utilized a recent single-cell RNA sequencing (RNA-seq) dataset that

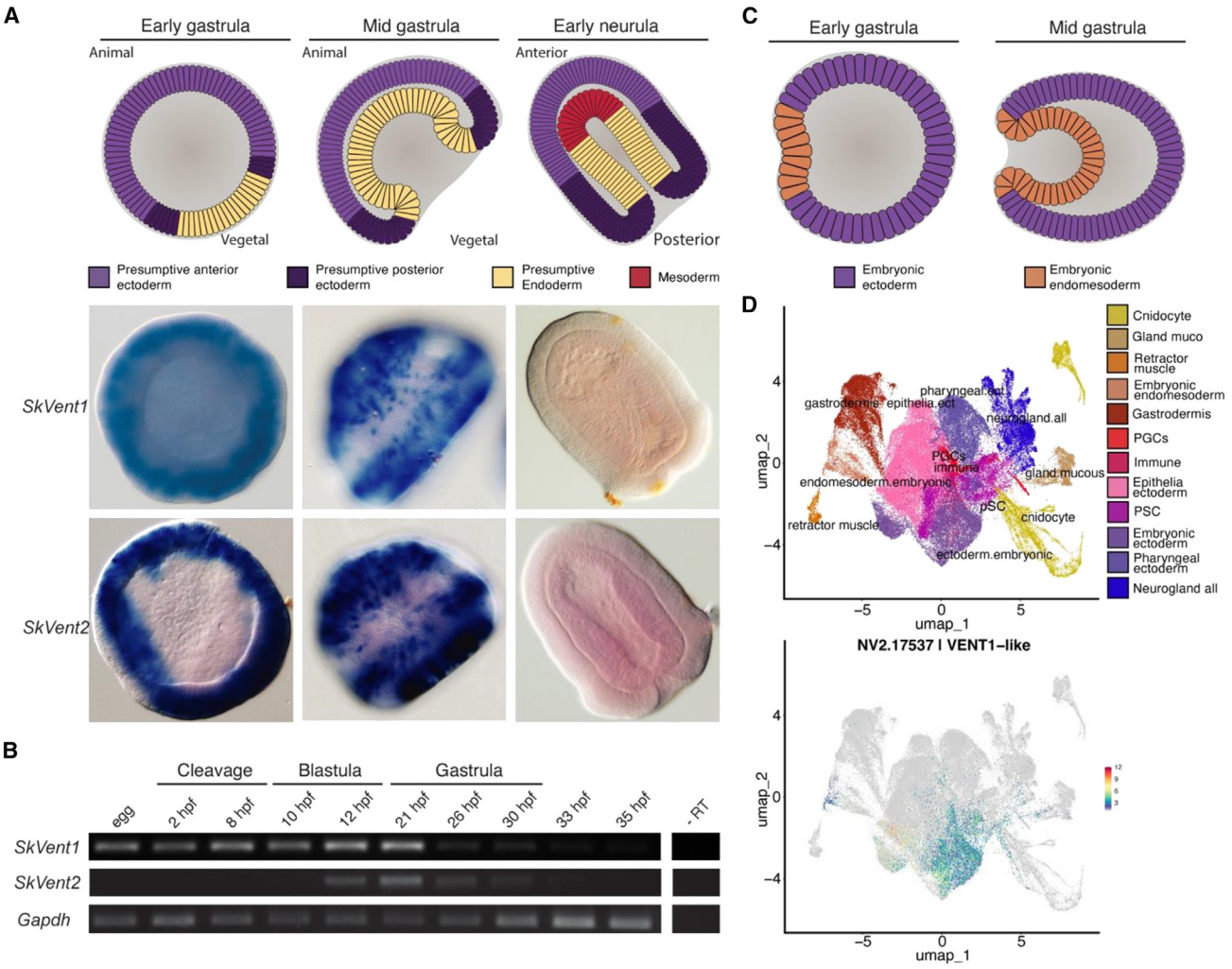


Figure 3. Characterization of ancient *Vent* expression profiles

(A) Schematic showing early and mid-gastrulation as well as neurulation stages in *Saccoglossus kowalevskii* embryos (top). *In situ* hybridizations showing SkVent1 and SkVent2 expression in early (18 hpf), mid (22 hpf), and early neurula (38 hpf) stage *Saccoglossus kowalevskii* embryos (bottom).

(B) Expression of SkVent1, SkVent2, and GAPDH across *Saccoglossus* early developmental stages.

(C) Schematic showing early and mid-gastrulation in *Nematostella vectensis* embryos.

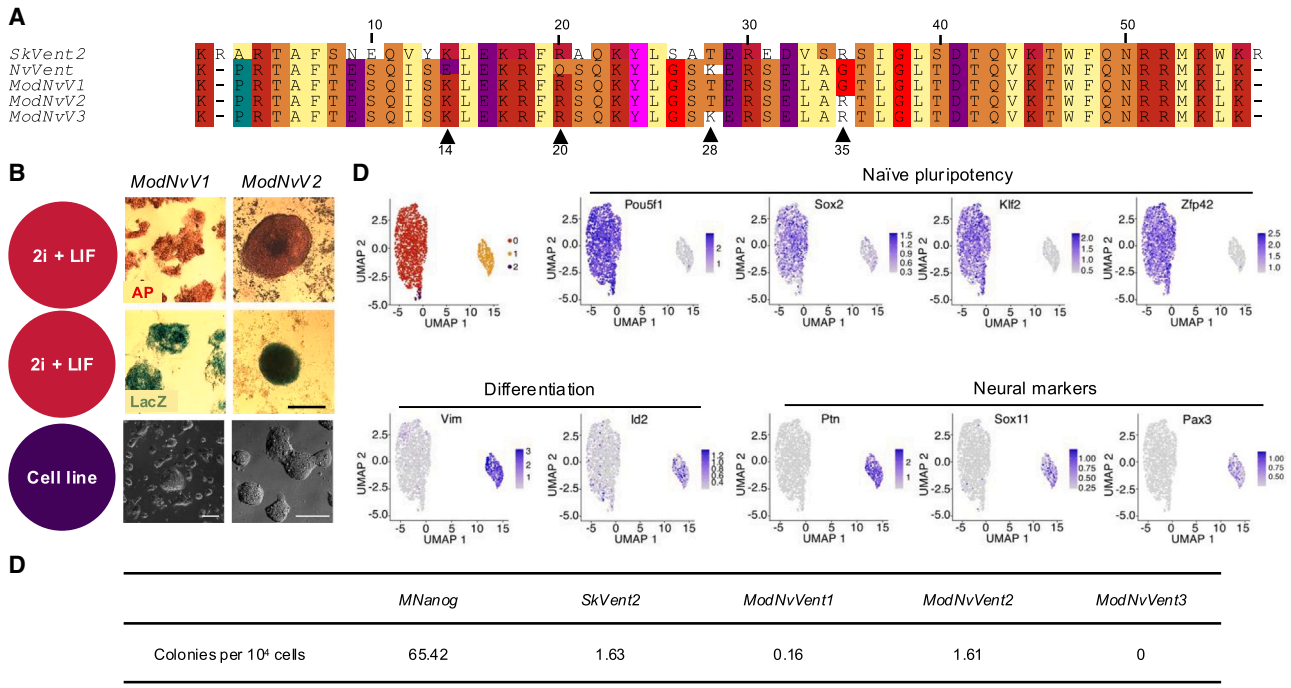
(D) Uniform manifold approximations and projections (UMAPs) showing single-cell data of *Nematostella vectensis* embryos from 18 hpf through to 16 days from Cole et al.[41] (top), as well as the expression of NvVent (bottom).

SkVent1/2, *Saccoglossus kowalevskii* (acorn worm) Vent1/2; NvVent, *Nematostella vectensis* (sea anemone) Vent; hpf, h post-fertilization.

contains embryos from the earliest gastrula stages (18 hpf) through to polyp stages.[41] In contrast to the expression profile of acorn worm vents, which showed expression exclusively within multipotent cells prior to and during gastrulation, *NvVent* was expressed broadly in both embryonic cell layers (Figures 3C and 3D): in the aboral and the oral embryonic ectoderm, including pharyngeal ectoderm, as well as in the endomesoderm. It is also expressed in the yet unidentified "immune" cell populations and in a subset of cells in the neural progenitor cluster.

Programming Nanog-like activity into sea anemone Vent

NANOG's reprogramming capacity is encoded in the HD[2]; we, therefore, swapped the HDs of *SkVent2* and *NvVent* to test if they accounted for the differential activities (Figure S5A). First, we transduced NnNpiPSCs with hybrid molecules, one containing the *SkVent2HD* fused to the N and C termini of *NvVent* (*Sk2HD-NvV*) and another consisting of the *NvVentHD* fused to the flanking regions of *SkVent2* (*NvVHD-SkV*). *NvVHD-SkV* induced the same iXEN-like cells observed with *NvVent* and died after G418 selection. In contrast, the *SkV2HD-NvVent* line transitioned to ground-state pluripotency following 2i+LIF conditions and G418 selection shown by LacZ and AP staining as well as activation of endogenous pluripotency genes (Figures S5B–S5E). Moreover, the efficiency of *SkVent2HD-NvVent* was comparable to that of the full-length *SkVent2* coding sequence (CDS), indicating that the NANOG activity exhibited by *SkVent2* is indeed encoded in the HD.[2]



D

	MNanog	*SkVent2*	*ModNvVent1*	*ModNvVent2*	*ModNvVent3*
Colonies per 10^4 cells	65.42	1.63	0.16	1.61	0

Figure 4. Conferring NANOG activity to *NvVent*

(A) Amino acid alignments of *SkVent2*, *NvVent*, *ModNvV1*, *ModNvV2*, and *ModNvV3* homeodomains. Arrows indicate where amino acids were changed between *NvVent*, *ModNvV1*, *ModNvV2*, and *ModNvV3*.

(B) Reprogramming with modified *NvVents*. piPSCs expressing *ModNvV1* or *ModNvV2* in serum + LIF (top), and in 2i conditions (middle) (*ModNvV3* cultures missing), fully reprogrammed iPSCs were then passaged to establish cell lines (bottom). Scale bar: 100 μm. AP (top) and LacZ (bottom) staining of *ModNvV1* and *ModNvV2* iPSCs. Scale bar: 100 μm.

(C) Single-cell sequencing of *ModNvV1* iPSCs and selected marker expression. Values represent relative counts on a log2 scale, scaled by a factor of 10,000.

(D) Reprogramming efficiencies. Values represent the averages across three independent experiments.

SkVent2, *Saccoglossus kowalevskii* (acorn worm) *Vent2*; *NvVent*, *Nematostella vectensis* (sea anemone) *Vent*; *ModNvV1*, modified *Nematostella vectensis Vent, NvVent*$^{14E>K/20Q>R/28K>T}$; *ModNvV2*, modified *ModNvV1, ModNvV-1*$^{35G>R}$; *ModNvV3*, modified *ModNvV2, ModNvV-2*$^{14T>K}$.

We then sought to find the minimum number of amino acid (aa) changes required to confer full reprogramming activity on *NvVent*. We changed 3 aa near the N terminus of the *NvVent* HD (*NvVent*$^{14E>K/20Q>R/28K>T}$), which we reasoned, based on differential charge, might contribute to the respective biochemical activities of *NvVent* and *SkVent2*, and expressed this modified factor (*ModNvV-1*) in NnNpiPSCs (Figure 4A). Following the switch to 2i+LIF, small numbers of compact colonies appeared that stained positive for LacZ and AP (Figure 4B). Notably, these colonies exhibited the flat, irregular morphology reminiscent of *Nanog*$^{-/-}$ mESCs[38] and were expanded into cell lines. Gene expression analysis showed that the cells activated endogenous pluripotency genes (Figure S6A). To determine the extent of reprogramming, we performed single-cell sequencing on the *ModNvV-1* line (Figures 4C and S6B–S6D). We identified three unique cell clusters. Most cells in cluster 0 showed high expression of naive pluripotency genes; cluster 2, in contrast, showed lower expression of *Sox2*, *Klf4*, *Esrrb*, and *Nr0b1*. Cluster 1 was grouped separately from 0 and 2 and exhibited very little to no naive pluripotency gene expression, instead expressing higher levels of neural stem cell markers such as *Ptn*, *Pax3*, *En1*, and *Sox11*. Together, these data suggest that cluster 0 represents fully reprogrammed cells, while cluster 2 may represent less stable cells prior to differentiation, and cluster 1 represents a separate, partially differentiated neural cell type, possibly reflecting the neural stem cell origin of the piPSCs.[2,13] We also found few cells co-expressing iXEN markers *Foxa2*, *Gata6*, and *Sox17*, indicating that these cells are eliminated after continuous culture (Figure S6D). Importantly, this demonstrates that despite a relatively low reprogramming efficiency, the 3 aa changes in *ModNvV-1* are sufficient to fully activate the endogenous mouse pluripotency network to form *bona fide* iPSCs.

We then asked if the reprogramming efficiency of *ModNvV-1* could be enhanced by changing a fourth aa closer to the C terminus of the HD to introduce a charge present in *SkVent2* (*ModNvV-1*$^{35G>R}$). When this new construct (*ModNvV-2*) was transduced into NnNpiPSCs, it induced colony formation with about a 10-fold enhancement over *ModNvV-1* (Figures 4B and 4D), and, importantly, the colonies induced by *ModNvV-2* were dome shaped, resembling the colonies formed by wild-type ESCs in 2i+LIF (Figure 4B). As an additional change, we reverted aa 28 of *ModNvV-2* to the wild type (*ModNvV-2*$^{14T>K}$). When this molecule, *ModNvV-3*, was expressed in NnNpiPSCs, it again induced the small, refractive cell type; however, the cells in these cultures did not produce colonies after the transition to 2i+LIF (Figure 4D), indicating that the change of aa 14 is essential for

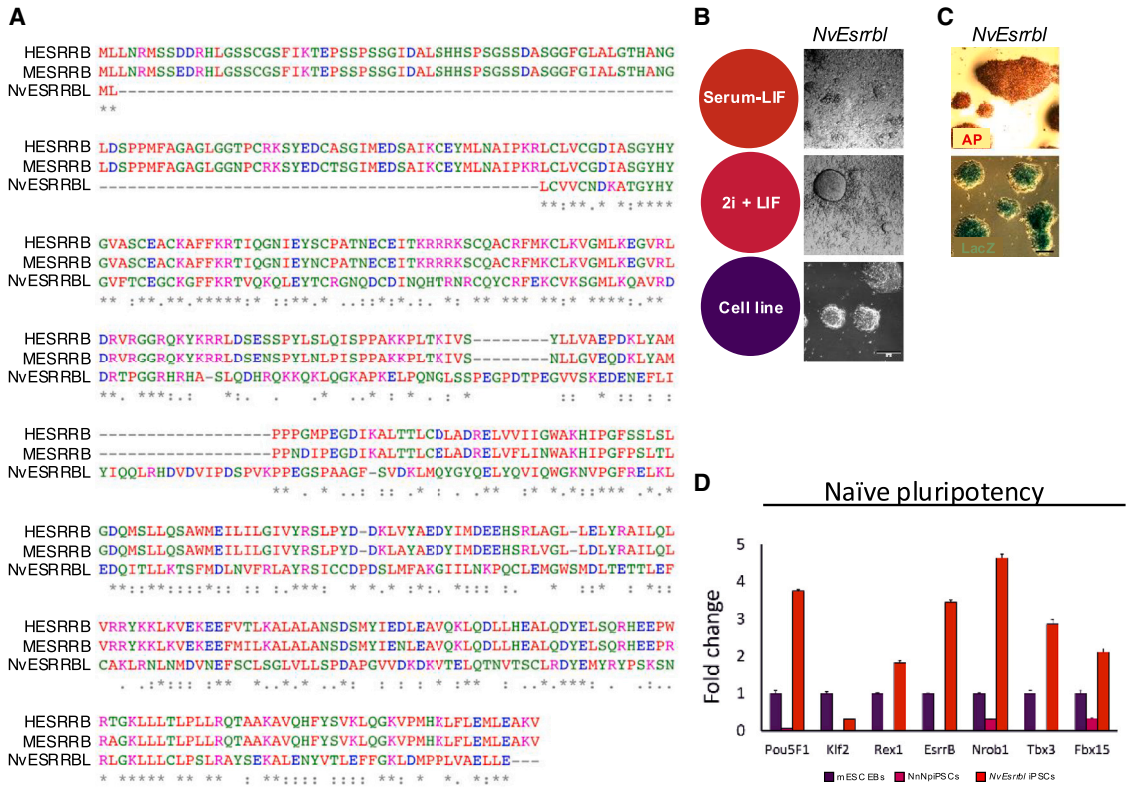


Figure 5. Reprogramming with *NvEsrrbl*

(A) Amino acid sequence alignments of HESRRB, MESRRB, and NvESRRBL.

(B) *NvEsrrbl* piPSCs in serum-LIF conditions (top) and after transitioning to 2i (middle); fully reprogrammed iPSCs were then passaged to establish cell lines (bottom). Scale bar: 200 μm.

(C) AP (top, red) and LacZ (bottom) staining of reprogrammed iPSCs. *NvEsrrbl* iPSCs stained positive for AP and LacZ. Scale bar: 60 μm.

(D) RT-qPCR analysis showing *NvEsrrbl*-reprogrammed iPSCs activate endogenous naive pluripotency genes. Relative to *Actb*. Error bars indicate SD from three technical replicates.

NvEsrrbl, *Nematostella vectensis* (sea anemone) *Esrrb-like gene*; *HESRRB*, human *Esrrb*; *MESRRB*, mouse *Esrrb*; mESC, mouse embryonic stem cells.

reprogramming. These data suggest a tri-phasic model for the evolution of NANOG activity wherein accumulated mutations to aa 14, 20, and 28 of an ancestral *Vent* gene HD conferred an ability to induce pluripotency, and an additional mutation at aa 35 enhanced this effect to result in the NANOG activity observed in *SkVent2*, which stabilizes the acquired pluripotent state in vertebrates. Subsequently, aa 42 (tyrosine) and 43 (lysine) within the vertebrate *Nanog* HD evolved, leading to much greater reprogramming efficiency.[2]

To investigate this model, we subjected the array of *Vent* variants to *in silico* structural analysis in comparison with *Nanog* (Figures S7A and S7B). This revealed that the aa changes in *ModNvV-1/2* are outside of the DNA-binding region, suggesting that these changes do not directly alter DNA binding and may act by recruiting other pluripotency factors.

Based on our model of incremental *Nanog* evolution, we then asked if direct NANOG targets might also be conserved in invertebrates. *Esrrb* is a direct target of NANOG and can substitute for *Nanog* in reprogramming assays.[42] Homology analysis identified the closest gene to *Esrrb* in the *Nematostella* genome (annotated as *Rxra*); however, the overall sequence conservation was relatively low (Figure 5A). We then tested this *Esrrb*-like gene (*NvEsrrbl*) in a reprogramming assay (Figures 5B and 5C). Remarkably, *NvEsrrbl* successfully reprogrammed the NnNpiPSCs evidenced by positive AP and LacZ staining as well as activation of endogenous pluripotency genes (Figure 5D). *NvEsrrbl* iPSCs displayed a slow growth rate and irregular, small, flat, jagged colonies, similar to the *ModNvV-1* iPSCs (Figure S4). Together, our data suggest that core elements of the pluripotency gene network are ancestral components of metazoan embryos and that NANOG activity evolved in the *Vent* gene family, which coincided with the evolution of mesoderm in Bilateria. We sought to understand how the emergence of NANOG activity may have contributed to vertebrate evolution, so we tested *Vent* function in axolotls, which express the conserved pluripotency network and model development in the tetrapod ancestor.

SkVent2 can specify vertebrate germline-competent mesoderm

In axolotls, NANOG KD results in the complete arrest of embryogenesis before gastrulation commences (Figure 6A).[20] It has been suggested that VENT1/2 carries out NANOG function in

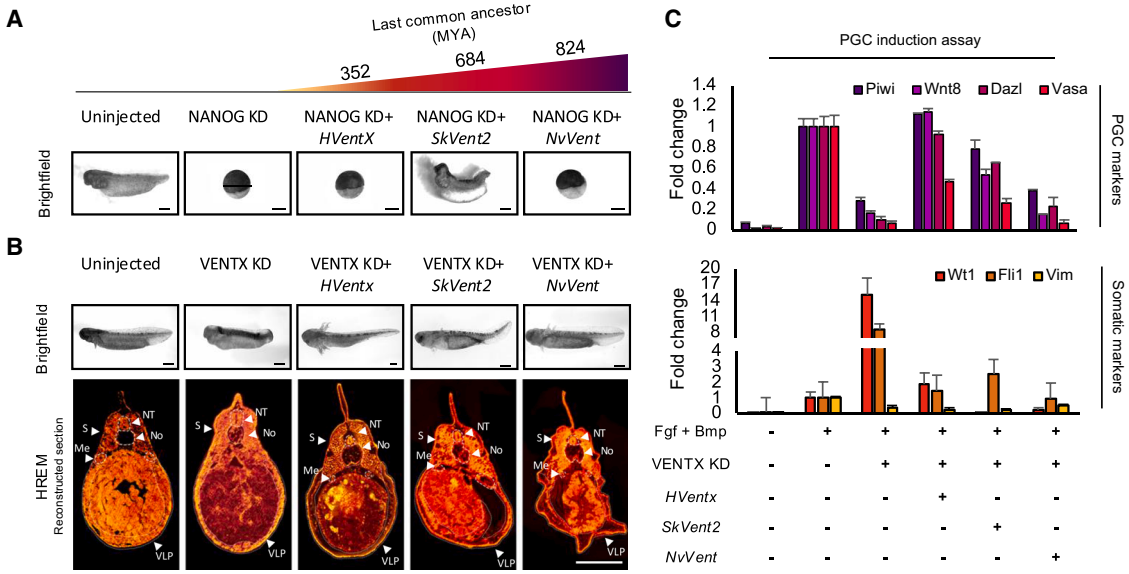


Figure 6. Vertebrate *Vent* genes are required for germline competence

(A) NANOG KD in axolotl arrests development pre-gastrula (stage 9). Rescue by co-injection of *SkVent2* mRNA, but not *NvVent* or *HVentx*, is shown. Scale bar: 1 mm.

(B) VENTX KD in axolotl can be rescued by co-injection of *SkVent2* and *HVentx* mRNA. Scale bar: 1 mm. Bright-field images and high-resolution electron microscopy (HREM) section reconstructions of uninjected and VENTX depletion with or without co-injection of *SkVent2*, *NvVent*, or *HVentx* mRNA are shown. Dashed lines highlight somites (S), neural tube (NT), notochord (No), and mesonephric ducts (Me). Dotted line delimits the ventral lateral plate (VLP). Scale bar, 1 mm.

(C) Gene expression (RT-qPCR) of uninjected and PGC-induced induced animal caps with and without VENTX-KD animal caps. Relative to *Odc1*. Error bars indicate SD from three biologically independent pools of 8–12 animal caps.

Xenopus,[22,23] which lacks a *Nanog* gene.[18] We therefore tested the ability of *Vent* variants to rescue this phenotype. Although *HVentX* could not rescue NANOG depletion in axolotls, NANOG activity was rescued by injecting mRNA coding for human *Nanog* (*HNanog*). Strikingly, RNA for *SkVent2* rescued gastrulation and tissue differentiation in NANOG-depleted embryos, but *NvVent* had no effect. We depleted axolotl embryos of VENTX, which diminished the posterior ventral region superficially (Figure 6B). High-resolution episcopic microscopy[43] reconstruction revealed that VENTX morphants lacked mesonephric ducts. Morphants appeared to show reduced posterior ventral lateral plate mesoderm displaying a thickened epidermal layer (Figure 6B). Co-injection of RNA encoding *HVentX*, *SkVent2*, or *NvVent* rescued the development of these structures, though aberrant internal morphology was still observed. In particular, morphants rescued with *NvVent* showed severe distortions of abdominal and lateral plate development. These data indicate a clear distinction between NANOG and VENT activities as well as the conservation of VENT function. Given that VENTX depletion diminished both posterior ventral lateral plate and intermediate mesoderm and that, in axolotls, germline-competent mesoderm[44,45] develops within the posterior ventral lateral plate,[46] we reasoned that VENTX might play a role in germline-competent mesoderm specification. To test this idea, we carried out a PGC induction assay[44] in axolotl animal caps (Figure 6C). Depletion of VENTX caused increased expression of somatic mesoderm markers, but crucially, the expression of PGC markers *Piwi*, *Dazl*, and *Vasa*, as well as *Wnt8*, a germline-competent mesoderm marker,[46,47] was diminished (Figure 6C). Together, these data suggest that VENTX is required for germline-competent mesoderm specification. Remarkably, RNA for *HVentX* and *SkVent2* rescued the expression of PGC markers as well as *Wnt8* and somatic mesoderm markers. *NvVent* partially rescued the expression of somatic genes and did not rescue the expression of *Wnt8* and PGC markers. These results align the ability to specify germline competence with the acquisition of NANOG activity in *SkVent2*.

DISCUSSION

Pluripotency, as characterized by the capacity of early embryonic cells to differentiate into both somatic and germinal lineages, is a fundamental cornerstone of multicellular organismal development. In mammals, this capability is underpinned by a GRN centered on NANOG, along with OCT-4 and SOX2.[48–50] However, the extent to which the mechanisms of mammalian pluripotency can be tracked back in non-mammalian vertebrates and invertebrates is poorly understood. For instance, in mammals, totipotent cells must first commit to trophectoderm (TE) or ICM fates. The ICM cells are then faced between epiblast and visceral endoderm/hypoblast fates. Given that both TE and visceral endoderm/hypoblast represent mammal-specific cell types, it is unclear why OCT-4 and NANOG, which prevent TE specification and hypoblast specification, respectively,[51,52] would be conserved in non-mammals, and yet both are conserved across vertebrates.[2–4,7,18] Indeed, both *in vivo* and

in vitro investigations have demonstrated that *Nanog* is required for the programming of pluripotency in mouse cells.[2,13] NANOG is also indispensable for gastrulation and germ layer differentiation in axolotl embryos.[20] These findings suggest that NANOG's conserved role may be to establish embryonic lineage competence in early embryonic cells rather than repress species-specific lineages. Surprisingly, despite the conservation of NANOG activity in vertebrates, no ortholog has been identified in invertebrates. While phylogenetic analyses have identified candidate *Nanog* orthologs, to date, none have been shown to possess the ability to facilitate the reprogramming of mammalian cells. Previous phylotypic analysis suggested that *Nanog* may have evolved from members of the *Bsx* gene family in Cephalochordates; however, the amphioxus *Bsx* gene was unable to functionally replace NANOG in reprogramming.[2] It has also been suggested that the *Nanog* gene arose from the *Vent* gene family and that latent NANOG activity in *XVent2.1* has allowed for the loss of *Nanog* in frogs.[22,23,53] Given the importance of pluripotency and the central role of NANOG, here we have searched for *Nanog* antecedents capable of programming mammalian pluripotency and provided evidence that the ability to program pluripotency is conserved in the invertebrate *Vent* family of transcription factors. Therefore, the ability to program pluripotency appears to have emerged in the bilaterian ancestor. This is evidenced by the ability of *SkVent2* to reprogram somatic cells to a pluripotent state. We have also demonstrated that the *Vent* gene from a deuterostome *Nematostella* possesses partial reprogramming activity. We found that *NvVent* reprogramming induces the formation of XEN-like cells, an extra-embryonic endodermal cell type. The induction of XEN cells has been reported as a by-product of incomplete reprogramming[40,54,55]; we, therefore, postulate that *NvVent* is able to initiate the reprogramming process but fails to establish ground-state pluripotency. In line with this hypothesis, *NvVent* was able to be endowed with *bona fide* NANOG activity with relatively few changes to the aa sequence. Interestingly, an *Esrrb*-like gene in sea anemone in its native form was able to reprogram NnNpiPSCs with no alterations, much like mammalian *Esrrb* genes.[42] Given that *SkVent2*, *ModNvV-1*, and *ModNvV-2* were able to also activate endogenous *Esrrb* and other downstream pluripotency factors, we propose that the aa changes and subsequent acquisition of NANOG activity during the evolution of the *Vent* gene family may have facilitated the hijacking of a pre-existing pluripotency GRN (pGRN), where factors such as *Esrrb* were already present.

Notably, the emergence of NANOG activity coincides with the emergence of mesoderm. While it is unclear to what extent these two events are linked, questions regarding the development of mesoderm and its evolutionary origins have been central in biology for over a hundred years.[56] Mesoderm represents a fundamental innovation of complex animals, giving rise to important organ systems such as skeletal elements, musculature, heart, kidneys, somites, and notochord. Interestingly, *SkVent2* was also able to rescue the development of NANOG-KD axolotl embryos, which fail to gastrulate or form mesodermal cell types, including somites, notochord, cardiogenic mesoderm, or intermediate mesoderm.[20] Together, this suggests that the evolution of NANOG activity may underpin the competence of early embryonic cells to form mesodermal and other chordate-specific cell types. It is tempting to speculate that the elaboration of the pGRN may itself have facilitated the increase of cellular diversity seen in bilaterians and, later, chordates.

Limitations of the study

While our study provides significant insights into the evolutionary origins of NANOG-like activity and its role in pluripotency, it has several limitations that warrant consideration.

Firstly, we have, in this and another recent study,[20] observed significant differences in the phenotypic outcomes following VENT2.1 depletion in *Xenopus* versus NANOG depletion in axolotl, while axolotl VENTX depletion does bare some resemblance to the VENT2.1-KD phenotype. Further, although we found no evidence supporting the ability of the Xenopus *Vent2.1* to maintain or program mammalian pluripotency, we cannot definitively exclude the possibility of a divergent regulatory mechanism in Anuran amphibians. Therefore, further comparative investigations in both axolotl and *Xenopus* species are essential to fully understand whether axolotl *Nanog* and *Xenopus Vent2.1* overlap in their functional roles.

Additionally, while we have uncovered a novel role of *Vent* genes in the regulation of germline-competent mesoderm in a vertebrate, the precise role of *Vent* genes within their native organisms remains undefined, necessitating further research. To date, most studies investigating the role of *Vent* genes in development have suggested roles in early cell potency,[22,57] hematopoiesis,[58–60] and regulation of neural crest cells.[61] While further study is required to elucidate whether the requirement for VENTX in PGC formation is conserved, it is noteworthy that *Ventx* expression is detected in human PGCs both *in vitro* and *in vivo*.[62,63] Further, the underlying mechanism by which it influences germline formation remains unclear. Given the divergence of vertebrate *Vent* genes from pluripotency regulation, the necessity for both *Vent* and *Nanog* genes in germ cell specification is also unclear.

Although we identified and tested *Vent* genes from diverse metazoan lineages, given the estimated 1.5 million metazoan species, our sample is clearly limited. A broader analysis of *Nanog/Vent* genes, encompassing a wider range of invertebrate species with different modes of PGC specification, would strengthen the understanding of the evolutionary relationship between *Nanog/Vent* genes, pluripotency, and germline development. Additionally, while the reprogramming capabilities of invertebrate *Vent* genes were investigated in mammalian cells and ISH was used to determine expression patterns, additional functional studies in native organisms are needed to fully clarify the roles of these genes during development. Furthermore, our study of gene expression in *Nematostella vectensis* was limited by the availability of single-cell sequencing data from earlier periods of development. While we identified *NvVent* expression in more specialized cell populations at this stage, earlier developmental time points (pre-gastrulation) may reveal whether *NvVent* is also expressed in unspecialized cells, providing additional insights into its role. Similarly, the identification of an *Esrrb*-like gene in *Nematostella vectensis* was preliminary, and while its ability to reprogram is suggestive of a pre-existing pGRN, direct experimental validation is also required to confirm functional conservation in the context of endogenous pluripotency networks.

Lastly, the reprogramming assays used to evaluate partial versus full pluripotency induction could not definitively determine the underlying mechanisms by which specific aa substitutions in the HD altered function. Although structural modeling provided clues, direct biochemical and structural validation, such as DNA binding assays and co-factor interaction studies, were not performed and require further assessment.

These limitations underscore the need for future studies employing broader taxonomic sampling, native developmental models, and mechanistic biochemical investigations to refine our understanding of the emergence of pluripotency and its impact on metazoan evolution. Despite these challenges, our findings lay the groundwork for further exploration into the evolution of pluripotency and its role in metazoan development.

RESOURCE AVAILABILITY

Lead contact

Requests for further information and resources should be directed to and will be fulfilled by the lead contact, Ramiro Alberio (ramiro.alberio@nottingham.ac.uk).

Materials availability

This study did not generate new unique reagents.

Data and code availability

- This paper did not create any original code.
- All data generated or analyzed during this study are included in this published article and its supplemental information files. The source data underlying the single-cell transcriptomes of ModNvVent1 reprogrammed NnNpiPSCs are deposited in the European Nucleotide Archive under the project number EMBL-EBI: PRJEB51885.
- Any additional information required to reanalyze the data reported in this paper is available from the lead contact upon request

ACKNOWLEDGMENTS

This project was funded by a grant from UK MRC (MR/N020979/1) (to A.D.J.). We would also like to thank John Gerhart and Michael Wu for making available the cloned vectors containing *Saccoglossus kowalevskii* Vents 1 and 2.

AUTHOR CONTRIBUTIONS

D.C., J.E.D., R.A., C.A., and A.D.J. conceived the project. D.C., L.S., R.A., C.A., J.E.D., M.L., and A.D.J. designed the experiments and analyzed the data. D.C. and L.S. led the experimental work, assisted by J.C. and N.H. F.S., G.G., U.T., and L.S. performed the bioinformatic analysis. T.F. and M.L. performed the phylogenetic analysis. L.S., D.C., and E.S. performed the *in vivo* work. G.G., U.T., D.C., and E.S. provided additional reagents and materials. N.M. performed the structural analysis. D.C., L.S., G.G., U.T., C.A., M.L., R.A., and A.D.J. prepared the manuscript.

DECLARATION OF INTERESTS

The authors declare no competing interests.

STAR★METHODS

Detailed methods are provided in the online version of this paper and include the following:

SUPPLEMENTAL INFORMATION

Supplemental information can be found online at https://doi.org/10.1016/j.celrep.2025.115396.

Received: May 13, 2024
Revised: December 21, 2024
Accepted: February 13, 2025

REFERENCES

1. Endo, Y., Kamei, K.I., and Inoue-Murayama, M. (2020). Genetic Signatures of Evolution of the Pluripotency Gene Regulating Network across Mammals. Genome Biol. Evol. *12*, 1806–1818. https://doi.org/10.1093/gbe/evaa169.
2. Theunissen, T.W., Costa, Y., Radzisheuskaya, A., van Oosten, A.L., Lavial, F., Pain, B., Castro, L.F.C., and Silva, J.C.R. (2011). Reprogramming capacity of Nanog is functionally conserved in vertebrates and resides in a unique homeodomain. Development *138*, 4853–4865. https://doi.org/10.1242/dev.068775.
3. Frankenberg, S.R., Frank, D., Harland, R., Johnson, A.D., Nichols, J., Niwa, H., Schöler, H.R., Tanaka, E., Wylie, C., and Brickman, J.M. (2014). The POU-er of gene nomenclature. Development *141*, 2921–2923. https://doi.org/10.1242/dev.108407.
4. Frankenberg, S., and Renfree, M.B. (2013). On the origin of POU5F1. BMC Biol. *11*, 56. https://doi.org/10.1186/1741-7007-11-56.
5. Niwa, H., Nakamura, A., Urata, M., Shirae-Kurabayashi, M., Kuraku, S., Russell, S., and Ohtsuka, S. (2016). The evolutionally-conserved function of group B1 Sox family members confers the unique role of Sox2 in mouse ES cells. BMC Evol. Biol. *16*, 173. https://doi.org/10.1186/s12862-016-0755-4.
6. Niwa, H., Sekita, Y., Tsend-Ayush, E., and Grützner, F. (2008). Platypus Pou5f1 reveals the first steps in the evolution of trophectoderm differentiation and pluripotency in mammals. Evol. Dev. *10*, 671–682. https://doi.org/10.1111/j.1525-142X.2008.00280.x.
7. Sukparangsi, W., Morganti, E., Lowndes, M., Mayeur, H., Weisser, M., Hammachi, F., Peradziryi, H., Roske, F., Hölzenspies, J., Livigni, A., et al. (2022). Evolutionary origin of vertebrate OCT4/POU5 functions in supporting pluripotency. Nat. Commun. *13*, 5537. https://doi.org/10.1038/s41467-022-32481-z.
8. 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. https://doi.org/10.1016/s0092-8674(03)00392-1.
9. Mitsui, K., Tokuzawa, Y., Itoh, H., Segawa, K., Murakami, M., Takahashi, K., Maruyama, M., Maeda, M., and Yamanaka, S. (2003). The

homeoprotein Nanog is required for maintenance of pluripotency in mouse epiblast and ES cells. Cell *113*, 631–642. https://doi.org/10.1016/s0092-8674(03)00393-3.

10. Ramos-Ibeas, P., Sang, F., Zhu, Q., Tang, W.W.C., Withey, S., Klisch, D., Wood, L., Loose, M., Surani, M.A., and Alberio, R. (2019). Pluripotency and X chromosome dynamics revealed in pig pre-gastrulating embryos by single cell analysis. Nat. Commun. *10*, 500. https://doi.org/10.1038/s41467-019-08387-8.
11. Ortega, M.S., Kelleher, A.M., O'Neil, E., Benne, J., Cecil, R., and Spencer, T.E. (2020). NANOG is required to form the epiblast and maintain pluripotency in the bovine embryo. Mol. Reprod. Dev. *87*, 152–160. https://doi.org/10.1002/mrd.23304.
12. Cauffman, G., De Rycke, M., Sermon, K., Liebaers, I., and Van de Velde, H. (2009). Markers that define stemness in ESC are unable to identify the totipotent cells in human preimplantation embryos. Hum. Reprod. *24*, 63–70. https://doi.org/10.1093/humrep/den351.
13. Silva, J., Nichols, J., Theunissen, T.W., Guo, G., van Oosten, A.L., Barrandon, O., Wray, J., Yamanaka, S., Chambers, I., and Smith, A. (2009). Nanog is the gateway to the pluripotent ground state. Cell *138*, 722–737. https://doi.org/10.1016/j.cell.2009.07.039.
14. Allègre, N., Chauveau, S., Dennis, C., Renaud, Y., Meistermann, D., Estrella, L.V., Pouchin, P., Cohen-Tannoudji, M., David, L., and Chazaud, C. (2022). NANOG initiates epiblast fate through the coordination of pluripotency genes expression. Nat. Commun. *13*, 3550. https://doi.org/10.1038/s41467-022-30858-8.
15. Yamaguchi, S., Kurimoto, K., Yabuta, Y., Sasaki, H., Nakatsuji, N., Saitou, M., and Tada, T. (2009). Conditional knockdown of Nanog induces apoptotic cell death in mouse migrating primordial germ cells. Development *136*, 4011–4020. https://doi.org/10.1242/dev.041160.
16. Nakanoh, S., Fuse, N., Takahashi, Y., and Agata, K. (2015). Verification of chicken Nanog as an epiblast marker and identification of chicken PouV as Pou5f3 by newly raised antibodies. Dev. Growth Differ. *57*, 251–263. https://doi.org/10.1111/dgd.12205.
17. Lavial, F., Acloque, H., Bertocchini, F., Macleod, D.J., Boast, S., Bachelard, E., Montillet, G., Thenot, S., Sang, H.M., Stern, C.D., et al. (2007). The Oct4 homologue PouV and Nanog regulate pluripotency in chicken embryonic stem cells. Development *134*, 3549–3563. https://doi.org/10.1242/dev.006569.
18. Dixon, J.E., Allegrucci, C., Redwood, C., Kump, K., Bian, Y., Chatfield, J., Chen, Y.H., Sottile, V., Voss, S.R., Alberio, R., and Johnson, A.D. (2010). Axolotl Nanog activity in mouse embryonic stem cells demonstrates that ground state pluripotency is conserved from urodele amphibians to mammals. Development *137*, 2973–2980. https://doi.org/10.1242/dev.049262.
19. Gagnon, J.A., Obbad, K., and Schier, A.F. (2018). The primary role of zebrafish nanog is in extra-embryonic tissue. Development *145*, dev147793. https://doi.org/10.1242/dev.147793.
20. Simpson, L.A., Crowley, D., Forey, T., Acosta, H., Ferjentsik, Z., Chatfield, J., Payne, A., Simpson, B.S., Redwood, C., Dixon, J.E., et al. (2023). NANOG is required to establish the competence for germ-layer differentiation in the basal tetrapod axolotl. PLoS Biol. *21*, e3002121. https://doi.org/10.1371/journal.pbio.3002121.
21. Kumar, S., Suleski, M., Craig, J.M., Kasprowicz, A.E., Sanderford, M., Li, M., Stecher, G., and Hedges, S.B. (2022). TimeTree 5: An Expanded Resource for Species Divergence Times. Mol. Biol. Evol. *39*, msac174. https://doi.org/10.1093/molbev/msac174.
22. Scerbo, P., Girardot, F., Vivien, C., Markov, G.V., Luxardi, G., Demeneix, B., Kodjabachian, L., and Coen, L. (2012). Ventx factors function as Nanog-like guardians of developmental potential in Xenopus. PLoS One *7*, e36855. https://doi.org/10.1371/journal.pone.0036855.
23. Scerbo, P., Markov, G.V., Vivien, C., Kodjabachian, L., Demeneix, B., Coen, L., and Girardot, F. (2014). On the origin and evolutionary history of NANOG. PLoS One *9*, e85104. https://doi.org/10.1371/journal.pone.0085104.
24. Johnson, A.D., Drum, M., Bachvarova, R.F., Masi, T., White, M.E., and Crother, B.I. (2003). Evolution of predetermined germ cells in vertebrate embryos: implications for macroevolution. Evol. Dev. *5*, 414–431. https://doi.org/10.1046/j.1525-142x.2003.03048.x.
25. Johnson, A.D., Crother, B., White, M.E., Patient, R., Bachvarova, R.F., Drum, M., and Masi, T. (2003). 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. https://doi.org/10.1098/rstb.2003.1331.
26. Johnson, A.D., and Alberio, R. (2015). Primordial germ cells: the first cell lineage or the last cells standing? Development *142*, 2730–2739. https://doi.org/10.1242/dev.113993.
27. Evans, T., Wade, C.M., Chapman, F.A., Johnson, A.D., and Loose, M. (2014). Acquisition of germ plasm accelerates vertebrate evolution. Science *344*, 200–203. https://doi.org/10.1126/science.1249325.
28. Chambers, I., and Tomlinson, S.R. (2009). The transcriptional foundation of pluripotency. Development *136*, 2311–2322. https://doi.org/10.1242/dev.024398.
29. Extavour, C.G., and Akam, M. (2003). Mechanisms of germ cell specification across the metazoans: epigenesis and preformation. Development *130*, 5869–5884. https://doi.org/10.1242/dev.00804.
30. Bateson, W. (1885). The Later Stages in the Development of Balanoglossus Kowalewskii. Q. J. Microsc. Sci. *25*.
31. Bateson, W. (1886). Continued Account of the Later Stages in the Development of Balanoglossus Kowalev-skii, and of the Morphology of the Enteropneusta. J. Cell Sci. *S2–26*, 511–534.
32. Molè, M.A., Coorens, T.H.H., Shahbazi, M.N., Weberling, A., Weatherbee, B.A.T., Gantner, C.W., Sancho-Serra, C., Richardson, L., Drinkwater, A., Syed, N., et al. (2021). A single cell characterisation of human embryogenesis identifies pluripotency transitions and putative anterior hypoblast centre. Nat. Commun. *12*, 3679. https://doi.org/10.1038/s41467-021-23758-w.
33. Xiong, Y., Scerbo, M.J., Seelig, A., Volta, F., O'Brien, N., Dicker, A., Padula, D., Lickert, H., Gerdes, J.M., and Berggren, P.O. (2020). Islet vascularization is regulated by primary endothelial cilia via VEGF-A-dependent signaling. Elife *9*, e56914. https://doi.org/10.7554/eLife.56914.
34. 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. https://doi.org/10.1042/BJ20080134.
35. 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. https://doi.org/10.1073/pnas.0802288105.
36. 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*, 940–948. https://doi.org/10.1002/jcb.21912.
37. Buganim, Y., Faddah, D.A., Cheng, A.W., Itskovich, E., Markoulaki, S., Ganz, K., Klemm, S.L., van Oudenaarden, A., and Jaenisch, R. (2012). Single-cell expression analyses during cellular reprogramming reveal an early stochastic and a late hierarchic phase. Cell *150*, 1209–1222. https://doi.org/10.1016/j.cell.2012.08.023.
38. Chambers, I., Silva, J., Colby, D., Nichols, J., Nijmeijer, B., Robertson, M., Vrana, J., Jones, K., Grotewold, L., and Smith, A. (2007). Nanog safeguards pluripotency and mediates germline development. Nature *450*, 1230–1234. https://doi.org/10.1038/nature06403.
39. Kunath, T., Arnaud, D., Uy, G.D., Okamoto, I., Chureau, C., Yamanaka, Y., Heard, E., Gardner, R.L., Avner, P., and Rossant, J. (2005). Imprinted X-inactivation in extra-embryonic endoderm cell lines from mouse blastocysts. Development *132*, 1649–1661. https://doi.org/10.1242/dev.01715.
40. Parenti, A., Halbisen, M.A., Wang, K., Latham, K., and Ralston, A. (2016). OSKM Induce Extraembryonic Endoderm Stem Cells in Parallel to Induced Pluripotent Stem Cells. Stem Cell Rep. *6*, 447–455. https://doi.org/10.1016/j.stemcr.2016.02.003.

41. Cole, A.G., Steger, J., Hagauer, J., Denner, A., Ferrer Murguia, P., Knabl, P., Narayanaswamy, S., Wick, B., Montenegro, J.D., and Technau, U. (2024). Updated single cell reference atlas for the starlet anemone Nematostella vectensis. Front. Zool. *21*, 8. https://doi.org/10.1186/s12983-024-00529-z.

42. Festuccia, N., Osorno, R., Halbritter, F., Karwacki-Neisius, V., Navarro, P., Colby, D., Wong, F., Yates, A., Tomlinson, S.R., and Chambers, I. (2012). Esrrb is a direct Nanog target gene that can substitute for Nanog function in pluripotent cells. Cell Stem Cell *11*, 477–490. https://doi.org/10.1016/j.stem.2012.08.002.

43. Mohun, T.J., and Weninger, W.J. (2012). Embedding embryos for high-resolution episcopic microscopy (HREM). Cold Spring Harb. Protoc. *2012*, 678–680. https://doi.org/10.1101/pdb.prot069583.

44. Chatfield, J., O'Reilly, M.A., Bachvarova, R.F., Ferjentsik, Z., Redwood, C., Walmsley, M., Patient, R., Loose, M., and Johnson, A.D. (2014). Stochastic specification of primordial germ cells from mesoderm precursors in axolotl embryos. Development *141*, 2429–2440. https://doi.org/10.1242/dev.105346.

45. Savage, A.M., Alberio, R., and Johnson, A.D. (2021). Germline competent mesoderm: the substrate for vertebrate germline and somatic stem cells? Biol. Open *10*, bio058890. https://doi.org/10.1242/bio.058890.

46. 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. https://doi.org/10.1006/dbio.2001.0264.

47. Foster, S., Oulhen, N., Fresques, T., Zaki, H., and Wessel, G. (2022). Distinct mechanisms of germ cell factor regulation for an inductive germ cell fate. Preprint at bioRxiv *2022*, xxx. https://doi.org/10.1101/2022.02.04.479164.

48. Yurko, Y.Y., Scerbo, M.W., Prabhu, A.S., Acker, C.E., and Stefanidis, D. (2010). Higher mental workload is associated with poorer laparoscopic performance as measured by the NASA-TLX tool. Simulat. Healthc. J. Soc. Med. Simulat. *5*, 267–271. https://doi.org/10.1097/SIH.0b013e3181e3f329.

49. Li, M., and Belmonte, J.C.I. (2017). Ground rules of the pluripotency gene regulatory network. Nat. Rev. Genet. *18*, 180–191. https://doi.org/10.1038/nrg.2016.156.

50. Shanak, S., and Helms, V. (2020). DNA methylation and the core pluripotency network. Dev. Biol. *464*, 145–160. https://doi.org/10.1016/j.ydbio.2020.06.001.

51. Lee, M., Oh, J.N., Choe, G.C., Kim, S.H., Choi, K.H., Lee, D.K., Jeong, J., and Lee, C.K. (2022). Changes in OCT4 expression play a crucial role in the lineage specification and proliferation of preimplantation porcine blastocysts. Cell Prolif. *55*, e13313. https://doi.org/10.1111/cpr.13313.

52. Wu, G., and Schöler, H.R. (2014). Role of Oct4 in the early embryo development. Cell Regen. *3*, 7. https://doi.org/10.1186/2045-9769-3-7.

53. Hellsten, U., Harland, R.M., Gilchrist, M.J., Hendrix, D., Jurka, J., Kapitonov, V., Ovcharenko, I., Putnam, N.H., Shu, S., Taher, L., et al. (2010). The genome of the Western clawed frog Xenopus tropicalis. Science *328*, 633–636. https://doi.org/10.1126/science.1183670.

54. Beyret, E., and Izpisua Belmonte, J.C. (2016). The XEN of reprogramming. Cell Res. *26*, 147–148. https://doi.org/10.1038/cr.2016.9.

55. Zhao, Y., Zhao, T., Guan, J., Zhang, X., Fu, Y., Ye, J., Zhu, J., Meng, G., Ge, J., Yang, S., et al. (2015). A XEN-like State Bridges Somatic Cells to Pluripotency during Chemical Reprogramming. Cell *163*, 1678–1691. https://doi.org/10.1016/j.cell.2015.11.017.

56. Rieger, R.M., and Ladurner, P. (2003). The significance of muscle cells for the origin of mesoderm in bilateria. Integr. Comp. Biol. *43*, 47–54. https://doi.org/10.1093/icb/43.1.47.

57. Chennen, K., Scerbo, M.J., Dollfus, H., Poch, O., and Marion, V. (2014). [Bardet-Biedl syndrome: cilia and obesity - from genes to integrative approaches]. Med. Sci. *30*, 1034–1039. https://doi.org/10.1051/medsci/20143011018.

58. Gentner, E., Vegi, N.M., Mulaw, M.A., Mandal, T., Bamezai, S., Claus, R., Tasdogan, A., Quintanilla-Martinez, L., Grunenberg, A., Döhner, K., et al. (2016). VENTX induces expansion of primitive erythroid cells and contributes to the development of acute myeloid leukemia in mice. Oncotarget *7*, 86889–86901. https://doi.org/10.18632/oncotarget.13563.

59. Rawat, V.P.S., Arseni, N., Ahmed, F., Mulaw, M.A., Thoene, S., Heilmeier, B., Sadlon, T., D'Andrea, R.J., Hiddemann, W., Bohlander, S.K., et al. (2010). The vent-like homeobox gene VENTX promotes human myeloid differentiation and is highly expressed in acute myeloid leukemia. Proc. Natl. Acad. Sci. USA *107*, 16946–16951. https://doi.org/10.1073/pnas.1001878107.

60. Leitoguinho, A.R., Ng, E., Stanley, E., and Elefanty, A. (2019). The role of ventx homeobox gene during human haematopoietic development. Exp. Hematol. *76*, S73.

61. Scerbo, P., and Monsoro-Burq, A.H. (2020). The vertebrate-specific VENTX/NANOG gene empowers neural crest with ectomesenchyme potential. Sci. Adv. *6*, eaaz1469. https://doi.org/10.1126/sciadv.aaz1469.

62. Castillo-Venzor, A., Penfold, C.A., Morgan, M.D., Tang, W.W., Kobayashi, T., Wong, F.C., Bergmann, S., Slatery, E., Boroviak, T.E., Marioni, J.C., and Surani, M.A. (2023). Origin and segregation of the human germline. Life Sci. Alliance *6*, e202201706. https://doi.org/10.26508/lsa.202201706.

63. Tyser, R.C.V., Mahammadov, E., Nakanoh, S., Vallier, L., Scialdone, A., and Srinivas, S. (2021). Single-cell transcriptomic characterization of a gastrulating human embryo. Nature *600*, 285–289. https://doi.org/10.1038/s41586-021-04158-y.

64. Swiers, G., Chen, Y.H., Johnson, A.D., and Loose, M. (2010). A conserved mechanism for vertebrate mesoderm specification in urodele amphibians and mammals. Dev. Biol. *343*, 138–152. https://doi.org/10.1016/j.ydbio.2010.04.002.

65. Colwin, A.L., and Colwin, L.H. (1950). The developmental capacities of separated early blastomeres of an enteropneust, Saccoglossus kowalevskii. J. Exp. Zool. *115*, 263–295.

66. Lowe, C.J., Tagawa, K., Humphreys, T., Kirschner, M., and Gerhart, J. (2004). Hemichordate embryos: procurement, culture, and basic methods. Methods Cell Biol. *74*, 171–194. https://doi.org/10.1016/s0091-679x(04)74008-x.

67. Lowe, C.J., Wu, M., Salic, A., Evans, L., Lander, E., Stange-Thomann, N., Gruber, C.E., Gerhart, J., and Kirschner, M. (2003). Anteroposterior patterning in hemichordates and the origins of the chordate nervous system. Cell *113*, 853–865. https://doi.org/10.1016/s0092-8674(03)00469-0.

68. Kunath, T., Saba-El-Leil, M.K., Almousailleakh, M., Wray, J., Meloche, S., and 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*, 2895–2902. https://doi.org/10.1242/dev.02880.

69. Takashima, Y., Guo, G., Loos, R., Nichols, J., Ficz, G., Krueger, F., Oxley, D., Santos, F., Clarke, J., Mansfield, W., et al. (2014). Resetting transcription factor control circuitry toward ground-state pluripotency in human. Cell *158*, 1254–1269. https://doi.org/10.1016/j.cell.2014.08.029.

70. Evans, T., Johnson, A.D., and Loose, M. (2018). Virtual Genome Walking across the 32 Gb Ambystoma mexicanum genome; assembling gene models and intronic sequence. Sci. Rep. *8*, 618. https://doi.org/10.1038/s41598-017-19128-6.

71. Nowoshilow, S., Schloissnig, S., Fei, J.F., Dahl, A., Pang, A.W.C., Pippel, M., Winkler, S., Hastie, A.R., Young, G., Roscito, J.G., et al. (2018). The axolotl genome and the evolution of key tissue formation regulators. Nature *554*, 50–55. https://doi.org/10.1038/nature25458.

72. Waterhouse, A., Bertoni, M., Bienert, S., Studer, G., Tauriello, G., Gumienny, R., Heer, F.T., de Beer, T.A.P., Rempfer, C., Bordoli, L., et al. (2018). SWISS-MODEL: homology modelling of protein structures and complexes. Nucleic Acids Res. *46*, W296–W303. https://doi.org/10.1093/nar/gky427.

STAR★METHODS

KEY RESOURCES TABLE

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Bacterial and virus strains		
Chemically-competent DH5 alpha cells	Made in-house (University of Nottingham)	N/A
psPAX2 packaging DNA	Addgene	CatNo #12260
pMD2-G envelope DNA	Addgene	CatNo # 12259
Chemicals, peptides, and recombinant proteins		
LIF (Leukemia inhibitory factor)	Made in-house (University of Nottingham)	N/A
PD0325901	Tocris	CatNo #4192
CHIR99021	Tocris	CatNo #4423
β-Mercaptoethanol	Thermo Fisher Scientific	CatNo #21985023
Sodium pyruvate	Thermo Fisher Scientific	CatNo #11360070
Polybrene	Sigma-Aldrich	CatNo #TR-1003-G
Geneticin Sulfate (G418)	Gibco	CatNo #A1720
Hygromycin B	Thermo Fisher Scientific	CatNo #10687010
Puromycin	Thermo Fisher Scientific	CatNo #A1113803
Ficoll	Sigma-Aldrich	CatNo #F5415
Ampicillin	Sigma-Aldrich	CatNo #A0166-5G
Fungizone	Sigma-Aldrich	CatNo #A2411
JB4 embedding kit	Sigma-Aldrich	CatNo #EM0100-1KT
Acridine orange	Sigma-Aldrich	CatNo #A8097
TrypLE Express	Gibco	CatNo #12604013
Lipofectamine 3000	Thermo Fisher Scientific	CatNo #L3000001
Critical commercial assays		
QIAquick Gel Extraction Kit	Qiagen	CatNo #28704
QIAGEN RNeasy Mini Kit	Qiagen	CatNo #74104
Alkaline Phosphatase Kit	Sigma-Aldrich	CatNo #MAK447-1KT
Chromium Single Cell 3′ Library and Gel Bead Kit v2	10X Genomics	PN-120267
Chromium Single Cell A Chip Kit	10X Genomics	PN-1000009
Chromium Multiplex Kit, 96 rxns	10X Genomics	PN-120262
Deposited data		
Single-cell transcriptomes of ModNvVent1 reprogrammed NnNpiPSCs	European Nucleotide Archive	PRJEB51885
Experimental models: Cell lines		
NS-derived pre-iPS cells	Kindly provided by J Silva (Cambridge Stem Cell Institute)	N/A
TNG MEF pre-iPSCs	Kindly provided by J Silva (Cambridge Stem Cell Institute)	N/A
Mouse embryonic stem cells (Nanog-null)	Kindly provided by J Silva (Cambridge Stem Cell Institute)	N/A
HEK 293T	Sigma-Aldrich	CatNo #12022001
CGR8 mouse ES cells ECACC 07032901	Derived in-house (University of Nottingham)	N/A
IMA81 E11 XEN cells	Kindly provided by Tilo Kunath (Univ of Edinburgh)	N/A

(*Continued on next page*)

Continued

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Experimental models: Organisms/strains		
Ambystoma mexicanum embryos	University of Nottingham	N/A
Nematostella vectensis embryos	Vienna BioCenter	N/A
Saccoglossus kowalevskii embryos	Georgetown University	N/A
Oligonucleotides		
QPCR Primers, see Table S1	Sigma-Aldrich	N/A
In situ probes, see Table S1	Sigma-Aldrich	N/A
Morpholino oligonucleotides, see method details	GeneTools, LLC	N/A
Recombinant DNA		
PSIN-EF2-Puro vector	Kindly provided by J Silva (Cambridge Stem Cell Institute)	N/A
PiggyBac-PGK-HYGRO vector	Kindly provided by J Silva (Cambridge Stem Cell Institute)	N/A
Software and algorithms		
QuantStudio 6 Flex	Life Technologies	N/A
Osirix MD	Pixmeo	N/A
Seurat R package	Satija Lab	N/A
Mr Bayes		N/A
MUSCLE	EBI	N/A
Swiss Model		N/A

EXPERIMENTAL MODEL AND STUDY PARTICIPANT DETAILS

Axolotl embryos and explants

Procedures involving animals have been approved by the Animal Welfare Ethics Committee Review Board, University of Nottingham. Embryos were collected following mating as described previously.[25,46,64] For all experiments, both male and female axolotl were used. The age of embryos used in this study generally ranged from 1 to 12 days old, although temperature can impact this. For consistency, embryos were staged according to the morphology of the WT embryos as previously described.[25,44,46,64] For microinjection, embryos were manually de-jellied and cultured in 1× modified Barth's solution (MBS) with 4% Ficoll (Sigma). From stage 7 onwards, embryos were maintained in 0.2× MBS, and dissected explants were maintained in 0.7×Marc's modified ringers solution (MMR). Culture solutions were supplemented with antibiotics (50 μg/mL penicillin and streptomycin, and 50 μg/mL kanamycin), 100 μg/mL Ampicillin and 50 μg/mL fungizone).

Collection and culture of acorn worm eggs and embryos

Gravid adult S. kowalevskii were collected in May and September from Waquoit Bay, in Massachusetts and maintained at the Marine Biological Laboratory (MBL; Woods Hole, MA) in sea tables with ambient running sea water. They are maintained in plastic dishes filled with sand and 3 animals to each dish, originally described by Colwin and Colwin[65] and modified by Lowe et al. 2004. Spawning, *in vitro* fertilizations, and fixations were performed as described previously.[66,67] Briefly, collected gravid females were spawned using a temperature shift of 29°C for 6–8 h. Eggs were collected over the course of the day. Sperm is collected by ripping a small piece of testes from a gravid male and collecting in a microfuge tube with sea water and kept on ice until needed. Fertilization was achieved by incubating a small drop of concentrated sperm with the eggs and monitoring for fertilization under a dissecting microscope. Embryos were maintained is filtered sea water and incubating the embyros at 20°C.

Cell culture

The study utilized various cell lines that have been previously genotyped, including male NS-derived pre-iPS cells and TNG MEF pre-iPSCs, as well as female mouse embryonic stem cells (both wild-type and Nanog-null), HEK 293T cells, and IMA81 E11 XEN cells. Further details of all cell lines can be found in the key resources table. 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 before pelleting by centrifugation for 5 min. Pelleted cells were resuspended in fresh media containing serum, counted using a haemocytometer, and plated.

Human Embryonic Kidney (HEK 293T; ECACC 12022001) cells were 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). Wild-type mouse ES cells of the CGR8 (ECACC 07032901) line, TNG MEF pre-iPSCs,[38] Nanog-null neural stem cell (NS) derived pre-iPS cells,[13] and the Nanog-null mouse embryonic stem cells[38] were all cultured on 0.1% gelatine. Cells were cultured in high glucose DMEM supplemented with 10% Fetal bovine serum (FBS), 1% Penicillin/Streptomycin, 2mM L-Glutamine, 0.1mM β-Mercaptoethanol (BME), 1mM Sodium pyruvate, and 2mM non-essential amino acids. Leukemia inhibitory factor (LIF) was added fresh at 100 ng/ml prior to use. Ground state 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, 2% B27 supplement, 2mM non-essential amino acids, 1% Penicillin/Streptomycin and 0.1mM β-Mercaptoethanol (BME). LIF (100 ng/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. IMA81 E11 XEN cells,[68] were maintained in 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. All cells were free of mycoplasma contamination, and were passaged and maintained as previously described[13,38,68,69] or according to manufacturer's protocols.

METHOD DETAILS

Morpholino and RNA microinjections

Morpholino oligonucleotides (GeneTools, LLC,OR) were designed to block translation of target proteins. Intron/exon boundaries were predicted by homology. Sequences were obtained from our axolotl genomic resource,[70] or via the axolotl genome website.[71] The morpholino sequences used were as follows: Translation MO: *Nanog*, 5′- GGTCAATCCAAAAGCTCCTCCTAAG-3′; Translation MO: *VentX* 5′-GAGAGGCCCCCATTTATACGGAGCC-3'. A nonspecific morpholino was injected in each experiment at equivalent levels to the specific splice morpholino combinations: MO: Control, 5′-GGATTTCAAGGTTGTTTACCTGCCG-3′. *In vitro* transcription and microinjection mRNAs for microinjection were synthesised using mMessage mMachine (Ambion) from plasmids encoding; Xenopus *FGF4*, Xenopus *BMP4*, Human *VentX,* Nematostella *Vent* and Saccoglossus *Vent*. PGC induction assay, preparation and *ex vivo* culture of animal cap explants was performed as outlined previously.[44]

Acorn worm embryo photography

Embryos to be cleared were dehydrated in methanol, washed using BB:BA (2:1 benzyl benzoate: benzyl alcohol) and either mounted on slides or washed in permount (Fisher Scientific) and mounted on slides for imaging. Cleared embryos were imaged under differential interference contrast optics (DIC). Chromogenic cleared embryos were photographed using a Jenoptik ProgRes C14 camera on a Zeiss Axioplan 2 Imager, a Zeiss Axiocam on an Axioimager or Q imaging micropublisher 3.3 camera on a Zeiss Axiophot.

In situ hybridization

Colorimetric whole mount *in situ* hybridization on acorn worm embryos was carried out using an established lab protocol. Embryos were kept in 2 mL glass vials for all steps until the colorimetric reaction performed in 6-well tissue culture plates. Samples were fixed in 4% paraformaldehyde (PFA). Proteinase K treatment was carried out at 10 μg/mL in PBST (0.1% Triton X- in 1x PBS) for 15 min at room temperature (RT). Acetic anhydride treatment at 250 μM for 5 min at RT followed by a 500-μM treatment for 5 min at RT. In situs for experimental embryos were stained using 1.6 μL:2.7μL ratio of 5-Bromo-4-chloro-3-indolyl phosphate: nitro-blue tetrazolium chloride and stopped with 3 × 5-min rinses at RT in 1XMAB. Some samples were further cleared by rinsing 2 × 5 min in MeOH and cleared with a 2:1 ratio of benzyl benzoate:benzyl alcohol (BBBA) before imaging on a Zeiss Axioimager. Probes used in this study are listed in Table S1.

Vector construction and gene delivery

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 (all primers are listed in key resources table). 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). PSIN-EF2-Puro, PiggyBac-PGK-HYGRO and pLenti-PGK-HYGRO vectors as well as the purified PCR product, were digested by restriction enzyme double-digest for 1 h at 37°C using NEB enzymes and buffers in aqueous conditions. These molecules were gel purified, and then ligated overnight using T4 ligase and buffer (Promega). 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°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 h at 37°C. After incubation, 1mL of the culture was taken and plasmid DNA was purified using the QIAprep Spin Miniprep Kit (Qiagen) according to manufacturer's protocols.

A restriction digestion was performed on the purified plasmid DNA, and subsequently run on 1% agarose gel, to screen clones for successful insertion of cDNA fragment into the ligated vector. Samples of each plasmid were then sequence verified.

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.

To produce lentiviral stocks, 293T cells were plated on 10cm petri dishes at a density of $8x10^6$, 6 h 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 and pMD2-G envelope DNA: 2.5μg. This mixture was incubated at room temperature for 20 min. Finally, the transfection mixture was added to 6mL of 293T media, placed on previously plated cells and cultured overnight. On Day 2, the transfection media was replaced with fresh 293T media. Cells to be transduced were plated onto a six well plate on day 3. The following day virus media was collected from 293T cells, and filtered with a 45-μm filter. Polybrene (4μg/ml) 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 h, then topped up with fresh cell media at a 1:1 ratio for overnight incubation. After overnight transduction, virus media is replaced with fresh cell media. At day 6, cells were passaged prior to antibiotic selection, which began on day 7.

To establish the PiggyBac (PB) lines, cells were co-transfected with a sequence verified PiggyBac construct 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. Puromycin (Gibco) at 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.

Cell reprogramming

Cells were maintained in basic ESC serum + LIF media on 0.1% gelatin (prepared in house), seeded at low density under the same serum + LIF conditions for reprogramming. Cells were allowed to attach and proliferate for 48 h, before culture media was switched to 2i + LIF N2B27 neurobasal serum-free media to initiate reprogramming. Antibiotic selection for complete reprogramming was carried out after six days in 2i + LIF. In place of the endogenous Nanog alleles in the $N^{-/-}$ NS derived pre-iPS line, one allele contained a Neomycin resistance gene which conferred resistance to the antibiotic Geneticin Sulphate- G418 (Gibco), the other allele contained a lacZ reporter gene which allows for fully reprogrammed cells to be detected by blue staining after addition of X-Gal. G418 selection was used at a final concentration of 400ug/ml. Colonies were isolated and passaged after ten days and cultures could be subsequently maintained in 2I + LIF media, however 200ug/ml G418 was required periodically for the maintenance of the SKVent2, ModNvV1 and ModNvV2 lines which were prone to differentiation.

TNG MEF pre-iPSCs were 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-iPSC colonies, media was switched to 2i + LIF N2B27 neurobasal serum-free media to mediate the transition to ground state. Colonies were subsequently checked under LED fluorescence for activation of the GFP-PURO reporter located in place of one of the endogenous Nanog alleles.[11] GFP-positive colonies were selected by addition of puromycin to the culture media.

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 were 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.

The Alkaline Phosphatase kit from SIGMA was used to stain the cells as per the manufacturer's manual.

To stain the NS derived cell lines for activation of the LacZ gene, cells were fixed for 15 min in fix buffer (5mM EGTA, 2mM MgCl2 0.2% glutaraldehyde in 0.1M PBS pH 7.3). cells were then washed for 5 min twice in wash buffer (2mM mgCl2 in 0.1M PBS pH 7.3). The cells were then incubated in staining buffer (5mM potassium ferrocyanide, 5mM potassium ferricyanide, 1 mg/mL 5-bromo-4-chloro-3-indolyl-β-D-galactosidase) for 12–16 h at 37°C. Staining buffer was then removed, cells were washed twice with wash buffer, then maintained in 1X PBS at room temperature prior to analysis.

Semi-quantitative reverse transcription quantitative PCR (RT-PCR)

Total RNA was extracted from cells in culture according to standard procedures. In brief, 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 min. The suspension was collected into a sterile 15mL falcon tube, and either stored at −20°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 min at room temperature. The sample was centrifuged at 14,000rpm for 30 min at 4°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.

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°C for 5 min, and then incubated at 4°C for 5 min. 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°C for 5 min, 50°C for 60 min, and 70°C for 15 min, before being chilled to 4°C.

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°C to activate the Jumpstart Taq before being held at 50°C for 2 min. The initial denaturation step was at 94°C for 10 min, followed by 40 cycles of 94°C for 15 s (denaturation) and 60°C for 1 min (annealing and extension). The raw data was then extracted to be analyzed by comparative CT (Cycle threshold). An endogenous control gene was chosen for each qPCR run to normalise the data and 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. Probes and primer information are listed in Table S1.

Semi-quantitative PCR

RT-PCR Twenty embryos were flash frozen at desired stages. Total RNA was isolated using PureLink RNA Micro Kit (Life Technologies), and cDNA was made using Tetro cDNA synthesis kit (Bioline) following the manufacturer's instructions. Semi-quantitative PCR was performed 28 cycles annealing at 60°C. Primer sequences can be found in Table S1.

High-resolution electron microscopy (HREM)

Samples were fixed overnight at 4°C in 4% PFA in PBS before being dehydrated overnight in increasing concentrations of methanol (50%, 70%, 80%, 90% and 100%) before mounted in JB4 medium with acridine orange (SIGMA), and processed for HREM as described by Mohun and Weninger.[43] Section TIFF images were trimmed using the image processor function on photoshop before being imported into Osirix MD to produce embryo 3D reconstructions using the magma shader.

Single-cell isolation, cDNA preparation and sequencing

Single-cell dissociation was performed by incubation in TrypLE Express (GIBCO) for 5 min at 37°C and ran through a cell strainer to remove clumped cells. Individual cells were subsequently transferred to DMEM +20% FCS to block TrypLE Express and washed in PBS-PVP. The 10X Genomics Chromium was used to generate a Single Cell 3′ Whole Transcriptome sequencing library using the Chromium Single Cell 3′ Library and Gel Bead Kit v2, the Chromium Single Cell A Chip Kit and the Chromium Multiplex Kit, 96 rxns (10X Genomics; PN-120267, PN-1000009 and PN-120262). The library was sequenced on the Illumina NextSeq 500 using a NextSeq 500 High Output v2.5 150 cycle kit (Illumina; 20024907), to generate >50,000 raw reads per cell.

Bioinformatic analysis

Sequenced sc-libraries were processed according to the 10X Cell Ranger pipeline. RNA-seq reads were mapped to the mouse reference genome (ensembl mm10).

Subsequent analysis was performed using the Seurat package in R (Available at https://satijalab.org/seurat/). Once raw data was imported, data was normalised with the NormalizeData function. 2000 Variable features were found using the FindVariableFeatures function with vst as the selection method. Prior to dimensionality reduction data was rescaled using the ScaleData function using all genes. Principal component analysis was performed using the RunPCA function using 30 npcs. K-nearest neighbors were calculated using the FindNeighbors function using a dimensionality setting of 1:10. Clusters were identified with the FindClusters function with a resolution of 0.1. UMAP dimensionality reduction was ran using default parameters and a dimensionality setting of 1:10.

Phylogenetic analysis

The 56 amino acid sequences of each of the homeodomains from the Nanog, Vent, and BSX families were input into Mr Bayes and a Bayesian tree was created rooting the tree on human NKX2.5. The program was run with the GTR+G model during the calculation of likelihood of the tree.

Sequence alignments

For comparison of homeodomains and full-length sequences the ebi resource MUSCLE was used using default parameters (available at https://www.ebi.ac.uk/Tools/msa/muscle/)

Structural modeling

The Vent homeodomains were modeled based on the structure of mouse Nanog homeodomain (pdb: 2vi6) using Swiss Model.[72] The resulting models were superposed onto that of human Nanog in complex with DNA (pdb: 4rb0) in order to visualize likely DNA orientation.

QUANTIFICATION AND STATISTICAL ANALYSIS

All quantitative analysis was carried out using R in R Studio or Microsoft Excel. Numbers of biological or technical replicates are described in the figure legends. Data are represented as mean plus SD. Normalisation and processing of single-cell RNA sequencing and qPCR is described above. For cell counting, cells were counted using a haemocytometer manually. The counting of AP-positive colonies was also done manually. All reprogramming experiments were repeated at least 3 times independently using different cell passage numbers.