Title

Memory of Somatic Cell Identity Affects Specific Cell Differentiation Programs Upon Reprogramming *In Vivo*

Authors

Tomas Zikmund[1,6 #], Jonathan Fiorentino[1,2,3,6 †], Chris Penfold[5], Marco Stock[1,2,3,4], Gaurav Agarwal[5 $], Larissa Langfeld[1], Stephan Hamperl[1], Antonio Scialdone[1,2,3 *] and Eva Hoermanseder[1,7 *]

[1] Institute of Epigenetics and Stem Cells, Helmholtz Zentrum München, German Research Center for Environmental Health, Munich 81377, Germany
[2] Institute of Functional Epigenetics, Helmholtz Zentrum München, German Research Center for Environmental Health, Neuherberg 85764, Germany
[3] Institute of Computational Biology, Helmholtz Zentrum München, German Research Center for Environmental Health, Neuherberg 85764, Germany
[4]TUM School of Life Sciences Weihenstephan, Technical University of Munich, Freising 85354, Germany
[5] Wellcome Trust/Cancer Research, Gurdon Institute, University of Cambridge, UK
[6] Equal contribution
[7] Lead contact
[†] Current address: Center for Life Nano- & Neuro-Science, Fondazione Istituto Italiano di Tecnologia (IIT), Rome, Italy
[#] Current address: Institute of Pathological Physiology, First Faculty of Medicine, Charles University, Prague, Czech Republic.
[$] Current address: Division of Hematology/Oncology, Boston Children's Hospital, Harvard Medical School, Boston, MA, USA and Broad Institute of MIT and Harvard, Cambridge, MA, USA.
* Correspondence: antonio.scialdone@helmholtz-munich.de (A.S.), eva.hoermanseder@hemholtz-munich.de (E.H.)

Summary

Nuclear reprogramming can change cellular fates. Yet, reprogramming efficiency is low and the resulting cell types are often not functional. Here, we used nuclear transfer to eggs to follow single cells during reprogramming *in vivo*. We show that the differentiation success of reprogrammed cells varies across cell types and depends on the expression of genes specific to the previous cellular identity. We find subsets of reprogramming resistant cells that fail to form functional cell types, undergo cell death, or disrupt normal body patterning. Reducing expression levels of genes specific to the cell-type of origin leads to better reprogramming and improved differentiation trajectories. Thus, our work demonstrates that failing to reprogram *in vivo* is cell-type specific and emphasises the necessity of minimising aberrant transcripts of the previous somatic identity for improving reprogramming.

Introduction

Reprogramming somatic cells into alternative cell fates is a critical aspect of regenerative medicine and stem cell biology. This transformation can be achieved through somatic cell nuclear transfer (NT) to eggs or cloning[1], ectopic expression of specific transcription factors[2,3], or cell-cell fusion[4]. Despite their potential, the clinical use of reprogrammed cells has been hindered by the challenge of achieving an efficient conversion to functional cells fully integrated into tissues and able to replace faulty cells in organisms[5,6].

Successful cell fate reprogramming requires the erasure of the existing cell fate, followed by the establishment of a different cell fate. On a molecular level, epigenomic and transcriptomic changes are likely to be key for complete cell fate switches during nuclear reprogramming[6]. Since cloned embryos consist only of reprogrammed cells, it has been hypothesised that their viability requires complete reprogramming of the donor cell fate and correct cell differentiation to functioning tissues and organs. Conversely, when the development of NT-embryos fails, it has been proposed that cell differentiation is defective across cell types due to a failure in epigenome and transcriptome reprogramming of the donor cell fate[7]. Indeed, bulk transcriptome analyses have revealed that most NT-embryos show an overall persistence of gene expression patterns reminiscent of the somatic donor cell type from which they originated[8–12], termed transcriptional memory[6,13]. A distinct set of genes were identified as ON-memory genes, as they were expressed in the donor cell and their expression persisted at unusually high levels in NT-embryos when compared to *in vitro* fertilized (IVF) embryos[11]. Additionally, the expected upregulation of many essential genes during embryonic development in IVF-embryos was not observed in NT-embryos[10,11,14], which led to their categorization as OFF-memory genes[11]. The retention of high levels of ON- and OFF-memory gene expression at the totipotent stages of NT-embryos is indicative of a poor developmental outcome[12,15]. By interfering with chromatin marks associated with reprogramming-resistant ON-memory genes, a reduction of both ON- and OFF-memory gene expression is achieved in NT-embryos, correlating with increased developmental success[11,12]. Similarly, reducing chromatin marks linked to OFF-memory genes rescues both OFF- and ON-memory gene expression in cloned embryos and improves their developmental outcome[10,12,16–19]. Collectively, these findings suggest that transcriptional ON- and/or OFF- memory gene expression at the totipotent stage may hamper the successful development of cloned embryos.

Currently, however, the differentiation defects across cell types in developing NT-embryos are unknown. Specifically, it is uncertain whether all cell types generated by reprogramming undergo equally defective differentiation or if some cell lineages are more severely affected than others. Furthermore, the role of transcriptional memory in causing these defects, especially in the context of a full organism, is unclear.

In this study, we transplanted endoderm nuclei to enucleated eggs of the frog *Xenopus laevis* to generate NT-embryos and monitored the differentiation of the produced reprogrammed cells into epidermal cell types. We assessed reprogramming outcomes at the single-cell level, using transcriptomic analysis and imaging techniques. Our findings demonstrate that the reprogramming outcome and the success of establishing functional cells varies immensely across cell types in NT-embryos. While some cell types of the epidermis,

such as goblet cells, are formed correctly from reprogrammed cells, other cell lineages, such as basal stem cell-derived ones, show severe differentiation defects. We furthermore observe reprogramming resistant cells that retain an endoderm-like state, leading to aberrant body patterning in NT-embryos. These phenotypes specific to NT-embryos are accompanied by an increase in cell death. We then addressed whether aberrant donor-cell specific gene expression causes the observed defects. By mimicking transcriptional ON-memory in the epidermis of fertilised embryos, differentiation, and body patterning defects analogous to those observed in NT-embryos were induced. Conversely, reducing transcriptional ON-memory in NT-embryos rescued the observed epidermal defects. These results indicate that transcriptional memory is a key determinant of these phenotypes.

In summary, our study reveals substantial variability in reprogramming efficiency across cell types and identifies the inappropriate expression of lineage-determining genes previously active in the donor nucleus as a crucial obstacle for the generation of functional cell types during tissue development in cloned embryos.

Results

Differentiation success varies across epidermal cell types of cloned embryos.

It is generally assumed that developmental failures in cloned embryos are caused by widespread cell differentiation defects. We asked whether this is indeed the case, or if specific cell lineages are more susceptible to defects in cloned embryos. Thus, we examined the transcriptomes of individual cells within the developing epidermis. This was carried out in endoderm-derived nuclear transfer (NT) embryos and compared to their naturally conceived counterparts from *in vitro* fertilized (IVF) embryos of the vertebrate frog *Xenopus laevis.*

Specifically, endoderm nuclei from neurula stage embryos were transplanted to enucleated eggs to generate NT-embryos (Fig.1A). To assess cell-fate conversion success in NT-embryos, we chose the developing epidermis (ectoderm) as the reprogramming target tissue, as the cell types and differentiation pathways are well characterized in fertilized embryos, thus facilitating the analyses of developmental defects. Therefore, at the late gastrula stage (stage 12) when cells of the animal region of fertilized embryos are fully committed to an epidermal fate, the respective 2-cell thick layers from 10 IVF- and 10 NT-embryos (two biological replicates), were isolated. It has been previously shown that 70% of stage 12 NT-embryos fail in further development[20], therefore we expected that most NT-cells collected from these NT-embryos would show defects. Isolated NT- and IVF-tissues of each biological replicate were pooled separately, cells were dissociated, and single-cell RNA sequencing (scRNA-seq) was performed (Fig.1A). The single-cell transcriptome profiles from 3,405 cells (1841 IVF and 1564 NT cells) passing quality filters (Fig.S1 A,B) were generated. On average, 42439 unique transcripts and 6329 genes were detected in a typical cell. Unsupervised clustering of the cells' transcriptomes via the Louvain algorithm identified 10 clusters of good-quality cells present in the isolated cell layers, which we visualized using Uniform Manifold Approximation and Projection (UMAP)[21] (Fig.1B Fig.S1C-D). Both NT- and IVF-embryo cells contributed to all identified clusters (Fig.1C, Fig.S1E). To those clusters, 8 major cell types (Fig.1B) were assigned by manually matching the expression of cluster-

specific marker genes (Fig.S1F) to *Xenopus tropicalis* scRNAseq datasets[22–24] or *in situ* hybridization data deposited on Xenbase[25] (Fig.1B). Goblet cell states (clusters 1, 2) and Cement gland primordium cell states (CGP; cluster 4) corresponded to the outer cell layer (*krt* high; Fig.S1G). Cells in the remaining clusters expressed marker genes specific for the sensorial inner layer (*sox11* high; Fig.S1H). There, cell states were identified as non-neural ectoderm (cluster 10), multiciliated cell progenitors (MCPs; cluster 7), basal stem cells (BSCs; clusters 3 and 8) as well as chordal- and anterior neural plate border cells (CNP and ANP, clusters 6 and 5, respectively; Fig.1B and Fig.S1F). Interestingly, an additional cell cluster (cluster 9) was identified within the inner layer that, based on its marker genes (Fig.S1F), could not unambiguously be attributed to any cell type normally found in the developing epidermis of fertilised embryos. Cells of this cluster expressed transcripts characteristic for ionocytes (*foxi1*), as well as transcripts normally seen in endoderm cells *(sox17b*; Fig.S1K). Thus, this cluster 9 is referred to as 'mixed states'. Further sub-clustering of this cluster revealed two subpopulations- one subcluster with ionocyte characteristics (*foxi1* positive, subcluster 9.2, Fig.S1 I,J and K) comprised of both IVF- and NT-embryo cells, and one subcluster with endoderm-like characteristics (*sox17b* positive) consisting mainly of NT-embryo cells (subcluster 9.1, Fig.S1 I, J and K). However, cells of both subclusters express *sox11* (Fig. S1H), suggesting that all cells originate from the inner layer of the developing epidermis. Thus, subcluster 9.1 is likely composed of particularly reprogramming resistant NT-ectoderm cells displaying a cell state related to the donor cell type endoderm, which will be addressed in more depth later. Overall, our transcriptional analysis suggests that NT-embryos can, in principle, form all cell types of the developing epidermis as well as an additional endoderm-like cell type with both ecto- and endoderm characteristics, which has never been described before.

Next, we asked if cell states are formed at the same proportions in the developing epidermis of NT-embryos when compared to IVF-embryos (Fig. 1D). Cell-type composition analyses of the scRNA-seq data revealed that NT- and IVF-cells were represented at overall equal proportions in 5 out of 8 cell types, and, thus, these cell types are likely formed with normal efficiencies in NT-embryos. However, NT-cells were significantly overrepresented in cluster 9 with mixed ionocyte and endoderm-like characteristics. Interestingly, NT-cells were also significantly underrepresented in the basal stem cell clusters, suggesting that these cell types are formed at a reduced number or are delayed in NT-embryos. Hence, this indicates that not all cell states emerge under similar proportions in the developing epidermis of NT-embryos.

The observed changes in the fractions of NT-cells within one cell cluster could be due to differences in the proliferation rate of NT cells compared to control IVF-cells. Therefore, we computationally estimated the cell-cycle phase of the cells within each cluster using our scRNA-seq data[16]. The relative abundance of cells in each cell-cycle phase was comparable between clusters of NT- and IVF-embryos (Fig.S1L), suggesting that a difference in cell proliferation cannot explain the observed defects in cell-type composition.

To address if differentiation dynamics are altered in NT-embryos, and hence could underlie the differences in cell-type composition, we performed single-cell fate mapping on our scRNA-seq data using Cellrank[26], an algorithm that combines information from RNA velocity[27] and gene expression similarity to study cell state dynamics. Specifically, we computed cell-state transition matrices for IVF- and NT-cells separately and identified terminal states of

differentiation. These cell-fate probability maps identified, as terminal macrostates, 'goblet cell', 'anterior neural plate border', 'basal stem cell' and 'multiciliated cell progenitor' in IVF-embryos (Fig.1E). In NT-embryos, as in IVF, 'goblet cell' was identified within the outer epithelial layer as terminal state (Fig. 1F), suggesting that goblet cells are formed efficiently in both NT- and IVF-embryos. In the inner layer, however, we observed the appearance of a new terminal state, 'mixed states', and the disappearance of the terminal state 'basal stem cell' which agrees with the changes in cell-type composition observed before (Fig.1D). In addition, we also observed a loss of the 'neural plate border' terminal state and a shift in differentiation of inner layer cells towards the terminal state 'multiciliated cell progenitor' within our NT-embryo datasets. These findings indicate that the differentiation dynamics in NT-embryos are disrupted specifically in the cells of the inner cell layer, while in the outer cell layer, they remain largely unaltered compared to IVF-embryos.

In summary, we found evidence that some cell differentiation programs are more vulnerable than others upon reprogramming. Overall, this suggests that the unsuccessful development of NT-embryos is linked to a failure in producing specific cell types during development.

Inefficient transcriptome reprogramming is observed in epidermal cell types with differentiation defects in cloned embryos.

Considering that the reprogramming defects occur only in specific cell clusters, we next asked if transcriptional memory affects these cell types more than others in the epidermis in NT-embryos.

For this purpose, we first characterized transcriptional memory globally, across all clusters combined, in a pseudo bulk differential gene expression analysis. To this end, we first identified transcripts significantly up- or down-regulated (FDR<0.05) in all NT-epidermal cells when compared to all IVF-epidermal cells (DE up and DE down; Fig.1G). Then, we intersected the set of genes more than 2-fold upregulated in the NT-epidermis compared to the IVF-epidermis with the set of genes expressed in the somatic endoderm donor-cell type (Donor mean RPKM > 1; Donor vs. IVF: FDR<0.05, fold change >2). By doing this, we found ON-memory genes, i.e., genes that were expressed in the endoderm donor cell and that persisted at too high levels in the epidermis of NT-embryos compared to IVF-embryos (Fig.1G,H). To identify the global OFF-memory gene set, we selected instead genes that were more than 2-fold downregulated in NT-embryos cells and that were expressed at low levels in the somatic endoderm donor cell (Donor mean RPKM <20; IVF vs. Donor: FDR<0.05, fold change >2; Fig.1G,I). Next, we evaluated the average expression levels of these ON- and OFF-memory genes across the epidermal cell types. We found that ON-memory gene expression is highest in cells belonging to cluster 9 with mixed states (Fig.1H), suggesting that this cluster contains cells with high levels of transcriptional memory. Global OFF-memory was instead more evenly distributed across all cell types (Fig.1I). This suggests that transcriptome reprogramming efficiency is not uniform across cell types.

As a complementary strategy to assess differences in transcriptional ON- and OFF-memory across cell types, we determined significantly differentially expressed genes between NT- and IVF-cells (FDR<0.05) for each cluster separately. We then intersected the set of upregulated genes in each cluster with the genes expressed in the endoderm (see definition

above) to identify cell-type specific ON-memory genes. Conversely, to reveal cell-type-specific OFF-memory genes, we intersected them with genes lowly expressed in the endoderm (see definition above). With this analysis, we found that the “mixed states” cluster 9 possesses the highest number of ON- and OFF-memory genes, corroborating our conclusion that this cluster has relatively high levels of transcriptional memory (Fig.1J). Other cell types with ~50 or more ON-memory genes include cement gland primordium, multiciliated cell progenitor and basal stem cells. Interestingly, we found evidence of differentiation defects for all these cell types (Fig.1D,E and F). Together, this indicates that transcriptional memory varies across cell types and is highest in the “mixed state” cluster 9, which likely includes cells in NT-embryos that resist reprogramming and maintain an endoderm-like state.

Finally, we evaluated the overlap of ON- and OFF-memory genes across cell types. We found that most OFF- and ON-memory genes are cell-type-specific. For instance, 84 ON-memory genes and 46 OFF-memory genes are specific to the mixed states cluster (Fig.1K,L) and another 15 ON- and 3 OFF- memory genes are specific to basal stem cell clusters (FIG.1 K,L). The largest number of overlapping ON-memory genes is between the mixed states and the cement gland (13; Fig. 1K), while the highest overlap of OFF-memory genes occurs between mixed states and chordal neural plate border (5; Fig. 1L). Together, these suggest that each cell type has a specific set of memory genes, and only very few memory genes are shared across multiple cell types.

In summary, we observed cell-type-specific defects in NT-embryos associated with high levels of ON- and OFF-memory gene expression, which in some instances culminated in cell fate transformations of epidermal cells into an endoderm-like state.

Endoderm gene expression domains expand into ectoderm regions in cloned embryos

The mixed states cluster 9 contains NT-cells that continue to express genes typical of the endoderm donor cell used to generate them (ON-memory genes) as well as genes indicative of an inner layer epidermal fate (ectoderm, *sox11*; Fig.S1H). Such endoderm-like cells displayed, based on marker gene expression, a double identity that has not been described so far. Hence, we further characterised their occurrence and localization in intact NT-embryos.

To identify the localisation of the endoderm-like cells in the NT-embryo, we first selected candidate endoderm ON-memory genes that can serve as markers, including the key endoderm transcription factors *sox17b* and *foxa4*, as well as the endoderm genes *march8* and *cdx1.* These genes are expressed in NT-cells with the endoderm-like state within cluster 9 in our scRNA-seq data set (Fig.2A,D,G and J). We then generated probes against transcripts of these marker genes and used them in fluorescence *in situ* hybridization (FISH, for lowly abundant transcripts) or chromogenic whole mount *in situ* hybridization (WISH, for highly abundant transcripts) assays of endoderm derived NT-embryos at embryonic stage 12. IVF-embryos served as wild-type control for normal gene expression patterns. *In vivo* FISH analyses revealed *sox17b* expression foci in a large proportion of cells, up to 30%, within the developing epidermis of NT-embryos, but not of IVF-embryos (Fig.2B and C). WISH against *foxa4* showed that the mRNA was also expressed in cells across the epidermis in NT-embryos, but not in IVF-embryos (Fig.2E). Furthermore, we could observe that the wild-type

expression domain of *foxa4* in the endoderm, which is visible in intact gastrula IVF-embryos as a ring around the blastopore, was broadened and extended in most gastrula NT-embryos until the animal pole and, thus, into the ectoderm (Fig.2E,F). We could also observe the same aberrant and expanded expression domains in NT-embryos for *march8* and *cdx1* (Fig.2H,I and Fig.2K,L respectively).

Together, these experiments confirm, in NT-embryos, the presence of endoderm-like cells in aberrant positions that normally correspond to ectoderm. Moreover, this data suggests that such endoderm-like cells probably originate from multiple locations. Notably, the observed expanded expression patterns of these key endoderm genes indicate a disruption of normal embryonic body patterning in NT-embryos.

Basal stem cell numbers are reduced, while epidermal progenitor cells emerge at normal rates in cloned embryos.

We then further investigated the defects in epidermal cell type composition, as indicated by our computational analyses of scRNA-seq data, in intact gastrula stage NT-embryos.

We analysed *in vivo* cell-type composition of the developing epidermis (Fig.3A) in stage 12 NT- and IVF-gastrula embryos and compared them to published studies in wild-type IVF-embryos[28]. At this time in development, we can observe the expected specification of multiciliated cell progenitors marked by *foxj1* expression (Fig.3A,B), of ionocyte progenitors marked by *foxi1* expression (Fig.3A,C) and of goblet cells marked by *otogl* expression (Fig.3A,D) in our scRNA-seq dataset in IVF- and NT-embryos. Using probes against these specific marker genes in WISH analyses, we observed that multiciliated cell progenitors *(foxj1)*, ionocyte progenitors (*foxi1*) and goblet cells (*otogl2*) emerge at similar rates in IVF- and in NT- gastrula embryos (Fig.3F,H). We could not detect the emergence of small secretory cells marked by *foxa1* in our analyses (from both the scRNA-seq data and *in vivo*), which is thus likely to occur at a later developmental stage (data not shown).

Basal stem cells are marked by the expression of *tp63*[24,29] and importantly, their numbers are significantly reduced in the epidermis of NT-embryos, when compared to IVF-embryos of the same developmental stage (Fig.3,G,H), confirming our findings of basal stem cell differentiation defects in the scRNA-seq data analyses (Fig.3E).

Together, this indicates that during early epidermal differentiation of cloned embryos, the first defect that can be observed *in vivo* is a reduction in basal stem cell numbers. However, cell types arising from basal stem cells, like multiciliated cell progenitors and ionocyte progenitors, emerge at a normal rate at this point in development.

Basal stem cell loss and defective mature epidermis coincide with increased cell death in cloned embryos.

The observed lower numbers of basal stem cells are sufficient to support the generation of normal numbers of epidermal progenitor cells in cloned embryos at early developmental gastrula stages (Fig.3). However, basal stem cells could become limiting as growth and differentiation progress and mature epidermal tissues (Fig.4A) are formed. Thus,

defects in epidermal cell type differentiation could become more apparent at later developmental stages in cloned embryos.

To test this, we generated endoderm-derived NT-embryos, as well as control IVF-embryos, and collected them at the tailbud stage, a time in development when the epidermis has fully matured and differentiated cell types have been formed[28] (Fig.4A). We first tested if basal stem cell numbers continue to be reduced in NT-tailbuds when compared to IVF-tailbuds using immunohistochemical staining of embryos against the basal stem cell marker Tp63. In 43% of NT-embryos (N=30) we observed areas of the epidermis depleted of basal stem cells, a phenotype never observed in IVF-embryos (N=26; Fig.4 B,C). Basal stem cells (BSCs) are located in the inner layer of the epidermis, and the mature cell types that differentiate from them intercalate into the outer layer of the epidermis in wild-type embryos (Fig.4A). We observed that multiciliated cells (MCCs), ionocytes (ISCs) and small secretory cells (SCCs) are intercalated between goblet cells in the outer layer of the epidermis at expected frequencies in IVF-embryos (Fig.4 D,E, IVF) and in the NT-embryos with normal numbers of Tp63 positive basal stem cells (Fig.4D,E Tp63 positive area). However, in NT-embryos that show a depletion of BSCs in the inner layer, the BSC-derived cell types MCCs, ISCs and SSCs are missing or reduced in the outer layer and primarily goblet cells can be found (Fig.4D,E Tp63 negative areas). Together, this reveals that cloned embryos continue to show basal stem cell reduction in the mature epidermis, which is accompanied by a loss of the mature basal stem cell-derived cell types multiciliated cells, ionocytes and secretory cells.

Previously, it has been reported that defective differentiation of BSC-derived cell types results in the depletion of the stem cell pool and increased cell death in the skin of mouse embryos[30]. Since similar defects were observed here in the epidermis of NT-derived embryos, we next addressed if increased cell death can also be detected in our system. Therefore, we evaluated the number of apoptotic cells by immunostaining against activated Caspase 3 in endoderm-derived NT-embryos, as well as in IVF-embryos as controls, at the tailbud stage. Indeed, most NT-embryos showed increased numbers of Caspase 3 positive cells in the epidermis, when compared to IVF-embryos (Fig.4F and G), which is indicative of increased cell death in the epidermis of cloned embryos.

Together, this suggests that differentiation of basal stem cells to epidermal cell types is defective in cloned embryos, as we observe a reduction of the stem cell pool, loss of the terminal cell types (multiciliated cells, ionocytes and small secretory cells) and increased cell death in the epidermis of cloned embryos.

Cell differentiation defects are recapitulated by the ectopic expression of ON-memory genes in fertilised embryos

Our single-cell transcriptome analyses of the developing epidermis in NT-embryos revealed that inner layer cells show transcriptional memory of the somatic endoderm cell used to generate the cloned embryo. Sox17 and Foxa4 are key endoderm-determining transcription factors[22,31] that were aberrantly expressed as ON-memory genes in NT epithelial cells. Based on this, we hypothesised that their expression could contribute to the aberrant basal stem cell differentiation occurring in NT-embryos.

Therefore, we tested if overexpression of Sox17b and Foxa4 in the epithelium of embryos generated by fertilisation could phenocopy defects observed in NT-embryos. To achieve this, we injected mRNA encoding FLAG tagged versions of these transcription factors individually into the ventral blastomeres of 8-cell embryos, which will give rise to the epidermis (Fig.5A). As controls, we injected mRNA encoding a FLAG tagged DNA binding protein without known TF activity (Kdm5b^{ci}, a catalytically dead histone demethylase) as well as uninjected fertilised embryos. Staining against the basal stem cell marker Tp63 at the tailbud stage revealed that most embryos expressing endoderm transcription factors Sox17b or Foxa4 in the epidermis showed basal stem cell-depleted regions in the inner layer (Fig.5B,C). As in NT embryos, the outer cell layer located above these depleted areas contained goblet cells but showed a reduction of basal stem cell-derived multiciliated cells, ionocytes and small secretory cells (Fig.5D,E). Furthermore, we found an increase in the number of cells positive for the cell death marker Caspase 3 in the epidermis of embryos ectopically expressing Sox17b and Foxa4, but not in embryos expressing the control protein Kdm5b^{ci} (Fig.5F,G)^{ci}.

Together, these data suggest that the ectopic expression of endoderm ON-memory genes can induce the same epithelial defects we identified in NT-embryos.

Reducing expression of the key endoderm ON-memory gene Sox17b rescues epidermal defects in endoderm derived NT-embryos

As our experiments above indicated, the expression of ON-memory genes, particularly of endoderm lineage-inducing transcription factors like Sox17b, are likely to be responsible for aberrant transcriptomes and differentiation defects of basal stem cells in NT-embryos. Thus, we reasoned that reducing the aberrant expression of Sox17b in epithelial cells of NT-embryos may rescue the observed epithelial phenotypes in NT-embryos.

To test this, we inhibited Sox17b translation in the developing epidermis by injecting *sox17b* antisense morpholinos (asMOs) into the ventral blastomeres of 8-cell embryos. Embryos were generated by NT of endoderm cells to enucleated eggs or by *in vitro* fertilization (Fig.6A). To trace asMO targeted areas in NT-embryos, we co-injected fluorescently labelled dextran. We collected embryos at the tailbud stage and identified basal stem cells using Tp63 immunostainings. In NT-embryos with successful targeting of *sox17b* asMOs to the epidermis, as indicated by dextran-positive cells, we observed a wild-type representation of epidermal basal stem cells in the epidermis in 8 out of 9 embryos (n=3; Fig.6B and C) indistinguishable from IVF embryos (N=15 embryos, n=3; Fig.6B and C). Instead, in untreated NT-embryos, basal stem cell-depleted regions of the epidermis were observed in 4 out of 9 embryos (n=3; Fig.6B and C), as observed before (Fig.4B and C). This suggests that reducing Sox17b ON-memory gene expression in the epidermis of endoderm derived NT- embryos restores basal stem cell numbers.

Next, we examined the number of apoptotic cells in these embryos and found that inhibiting *sox17b* ON-memory gene expression in the epidermis by asMO injection reduced the number of caspase-positive cells per embryo to levels similar to those observed in IVF embryos (n=2; Fig 6D,E). Instead, 5 out of 6 untreated NT-embryos (n=2; Fig.6D,E) showed elevated numbers of apoptotic cells, as observed before (see Fig.4F,G).

Together, the data suggest that a reduction in the expression of the endoderm ON-memory gene Sox17b rescues epidermal defects in endoderm-derived NT-embryos. This, in turn, indicates that ON-memory gene expression is a major contributor to the abnormalities observed in NT-embryos.

Discussion

Our study reveals that reprogramming success *in vivo* is cell-type-specific and uncovers memory of active transcriptional states as both an indicator and an underlying cause for the observed developmental problems in NT-embryos. Using single-cell transcriptome analyses and *in vivo* studies, we uncovered a diverse landscape of cell behaviours within the epidermis of NT-embryos, identifying distinct reprogramming outcomes ranging from normal developmental programs for some cell types to abrogated cell differentiation and cell death for others. We show that the differentiation defects of specific cell types in NT-embryos are associated with transcriptional memory. Moreover, mimicking ON-memory by exogenous expression of the endoderm lineage transcription factors Sox17b and Foxa4 in the developing epidermis of fertilized embryos produced defects resembling those observed in NT-embryos. Conversely, the reduction of *sox17b* ON-memory gene expression in the developing epidermis of NT-embryos improved the efficient generation of functional, fully differentiated epidermal tissues. This suggests that the expression of key ON-memory genes is necessary and sufficient for the differentiation defects of reprogrammed cells and the developmental phenotypes observed in nuclear transfer embryos.

So far, the study of reprogrammed cell differentiation relied largely on cell culture systems. Typically, induced pluripotent stem cells (iPSCs) are generated and directed to adopt new identities, or alternatively, one cell type is directly transformed into another using specific lineage transcription factors[32]. These methods have significantly enhanced our understanding of cell lineage decisions during reprogramming, yet they have limitations. In particular, the *in vitro* nature of cell differentiation entails, for example, the lack of key signalling molecule gradients or tissue mechanical properties that are present *in vivo* and are required for accurate cell specification. Here, by leveraging NT in the frog model system, we followed the development of reprogrammed cells within an organismal context and described their *in vivo* differentiation pathways. Since NT-embryos are built entirely from reprogrammed cells, we could make comprehensive comparisons of their differentiation success into various cell types.

We uncover a previously unappreciated heterogeneity of differentiation success across cell-types in NT-embryos. Being able to compare the single-cell transcriptomes of reprogrammed cells in NT-embryos to their exact *in vivo* counterparts allowed in-depth computational analyses of the data generated here, including determining the frequency of NT- and IVF-cells within the cell clusters and cell differentiation dynamics. These analyses reveal that, surprisingly, many cell states differentiated normally in NT-embryo. However, the relative abundance of basal stem cells was reduced in NT-embryos compared to IVF-embryos, and their ability to differentiate was disrupted in our *in silico* analyses. Moreover, we found a new cell state formed exclusively in NT-embryos, which co-expressed epidermis and endoderm markers and altered the tissue's differentiation dynamics. Together, these analyses of the single-cell transcriptome data singled out differentiation defects in specific types of

reprogrammed cells, which we confirmed *in vivo* in NT-embryos and that are likely to result in developmental problems. This suggests that embryonic death observed in cloned organisms is not due to an overall failure in differentiation of all cell types, but rather the result of defective differentiation of specific cell lineages only.

The aberrant expression of endodermal genes in the developing epidermis of NT-embryos has been reported before in bulk RNA-seq studies[11]. Such genes were called ON-memory genes since they represented a transcriptional memory of the donor cell type. Nevertheless, it was unclear whether the ON-memory genes are expressed throughout all cells of the tissue or only in specific cell types, and if their presence is detrimental to the generation of functional cell types during reprogramming. Thanks to the single-cell resolution of our study, we discovered that the transcriptional ON-memory is not equally present everywhere but is especially prominent in basal stem cells, cement gland primordium cells, multiciliated cell progenitors, and an endoderm-like cell state with a mixed identity only found in endoderm derived NT-embryos (Fig. 1J, cluster 9).

We speculated that NT-cells with mixed endoderm/ectoderm identity in an epidermal environment could result in aberrant body patterning of the developing embryos and thus addressed the localization of these cells *in vivo*. The observed expansion of endoderm gene expression domains far into normally ectoderm regions confirmed this *in vivo* and suggests that this could result in defective ectoderm cell specification and differentiation.

Furthermore, we hypothesized that the defects we found in basal stem cells might be driven by the high degree of transcriptional memory. We thus performed experiments that manipulated the expression levels of ON-memory genes in the developing epidermis of IVF- and NT-embryos, while leaving the other tissues unperturbed. These experiments confirmed our hypothesis that the expression of specific ON-memory genes is an underlying cause for at least some of the developmental phenotypes observed in NT-embryos.

Importantly, we then showed for the first time that reducing the expression level of one key lineage determining transcription factor showing ON-memory, Sox17b, can increase the differentiation success of reprogrammed cells. We propose a functional hierarchy amongst the genes showing transcriptional memory in cells of NT-embryos. Lineage determining transcription factors, such as Sox17b, are upstream regulators of the endoderm transcription network and that can induce the expression of downstream endoderm genes during development of IVF-embryos[31,33]. We previously found that sox17b is a gene especially resistant to reprogramming due to the stabilization of its active epigenetic state by chromatin modifications such as H3K4me3[11]. When erroneously expressed in epidermal cells of NT-embryos, due to epigenetic ON-memory, we speculate that Sox17b induces expression of other endodermal genes and disrupts the normal activation of epidermal gene expression networks, such as the one of basal stem cells and their progeny, thereby preventing normal differentiation. It is possible that other cell lineages, such as the one of goblet cells, are less sensitive to ON-memory or can more efficiently reprogram the endoderm fate to a goblet cell fate. What determines the ability of some differentiation programs of reprogrammed cells to overcome epigenetic memory of the donor cell type in NT-embryos remains unknown.

We observed that NT-embryos also show an increase in caspase-positive cells, suggesting an increase in cell death. This raises the question of which cells undergo cell death and how this is induced: Are these basal stem cells that fail in differentiation? Does inefficient

transcriptome reprogramming induce cell death? Our discovery that the expression levels of Sox17b in epidermal tissues of IVF- and NT-embryos correlate with cell death indicates a link between transcriptional ON-memory and cell death. Interestingly, in zebrafish, cells with a transcriptional signature that does not fit the surrounding tissue become apoptotic via cell competition[34]. Furthermore, phenotypes similar to those we observed in frog NT-embryos are present in the developing epidermis of mouse embryos when cell competition is induced[30]. Hence, it is tempting to speculate that the observed poorly reprogrammed cells with high degrees of ON memory are eliminated by the more efficiently reprogrammed cells in NT embryos through cell competition.

Overall, our results provide new insights into the outcomes of cellular reprogramming *in vivo* and its molecular drivers, which are crucial for regenerative medicine. Reducing the expression of specific ON-memory genes emerges as a novel strategy for efficiently producing functional cells and tissues capable of replacing irreversibly damaged tissues through cell reprogramming.

Acknowledgements

T.Z. and E.H. were supported by CRC1064 Chromatin Dynamics Project-ID 213249687 and HO 6864/2-1 project grant, both DFG; Recognition Award, Helmholtz Association; Marie Sklodowska-Curie Postdoctoral Fellowship; Project grant MR/P00479/, MRC. S.H. by ERC starting grant 852798. Work in the labs of E.H., A.S. and S.H. was funded by the Helmholtz Association. J.F. and M.S. were supported by the Joachim Herz Stiftung Add-on Fellowship for Interdisciplinary Life Science. M.S. was supported by the Helmholtz Association under the joint research school “Munich School for Data Science—MUDS”.

We would like to thank John Gurdon, Jerome Jullien, Charles Bradford and the Gurdon; Ralph Rupp and his team, Peter Walentek, Maria-Elena Torres-Padilla and members of our teams and institutes for reagents, their critical input and support throughout the project.

Author contributions

T.Z., J.F., C.A.P, G.A. and E.H. conducted the experiments and analyses; S.H. provided critical support; T.Z., J.F., E.H. and A.S. designed the analyses and experiments; E.H, A.S, J.F. and T.Z. wrote the paper.

Declaration of interests

The authors declare no competing interest.

Data and code availability

The data have been deposited at GEO under the accession number GSE269252. All original code can be accessed at https://github.com/ScialdoneLab/scXen . Any additional information required to reanalyse the data reported in this paper is available from the lead contact upon request.

Supplemental information

Figure S1 related to figure 1.

Figure legends

Figure 1. Differentiation defects vary across epidermal cell types and are associated with incomplete transcriptome reprogramming in cloned embryos

(A) Experimental design of scRNAseq experiment.
(B,C) NT-embryos can form all cell types of the developing epidermis. (B) UMAP plot of scRNA-seq data from IVF- and NT- cells. Cells are coloured according to the cluster or (C) according to condition IVF or NT.
(D) Not all cell states emerge at similar proportions in NT- and IVF-embryos. *In silico* cell type compositional analyses of scRNA-seq data where each bar represents the mean fraction of NT- over IVF-cells in each cell cluster. p-value: for each cell type computed using a Wald test on the regression coefficient and corrected with the FDR method; * FDR < 10^{-5}; over: overrepresented in NT-embryos; under: underrepresented in NT-embryos.
(E,F) NT-cells of the inner layer show altered differentiation dynamics. (E) UMAP showing IVF-cells coloured by terminal state of differentiation computed using CellRank. Colours match cell clusters shown in B. Brighter or darker colours indicate a lower or higher probability of assignment to a terminal state, respectively. (F) NT-cells analysed and visualized as described in (E).
(G) MA plot comparing gene expression between epidermal cells of NT- and IVF- embryos. The average log2 FC in NT-cells over IVF-cells is plotted on the y-axis and the mean log2 (RPKM+1) gene expression in the donor endoderm tissue (bulk RNA-seq data) on the x-axis. Red: upregulated transcripts in NT; dark red: ON-memory genes; blue: downregulated transcripts in NT; dark blue: OFF-memory genes.
(H) UMAP plot of ON-memory gene expression in individual NT- and IVF-cells. For each cell, the average of the scaled and centred expression levels of the ON-memory genes is shown.
(I) UMAP plot of OFF-memory gene expression in individual NT- and IVF-cells. For each cell, the average of the scaled and centred expression levels of the OFF-memory genes is shown.
(J) Bar plot displaying counts of ON- and OFF-memory genes in individual epidermal cell types of NT-embryos.
(K) UpSet plot showing the overlap of ON-memory genes across cell types in NT-embryos. The bottom left horizontal bar plot shows the total number of ON-memory genes detected in each cell type. The number of intersected genes between individual cell states is depicted in the upper vertical bar plot. Dots connected with vertical lines represent cell states sharing the same ON-memory genes. NPB: neural plate border.
(L) UpSet plot showing the overlap of OFF-memory genes across cell types in NT-embryos as described in (K).

Figure 2. Endoderm gene expression domains expand into ectoderm regions in cloned embryos

(A, D, G, J) Expression levels of *sox17b*, *foxa4*, *march8* and *cdx1* in NT- and IVF-cells shown in UMAP plots.

(B) Representative image of epithelia of IVF- and NT-embyros at stage 12 stained by fluorescent *in situ* hybridization (FISH) with antisense RNA probe against *sox17b* transcript; Two channels merged: Cyan – nuclei (DAPI); Red – *sox17b* probe (Cy5)
(C) Quantification of (B). Dot plot of relative abundance of *sox17b* positive cells in NT- and IVF- epithelium.
(E, H, K) Representative images of IVF- and NT- stage 12 embryos stained by whole mount *in situ* hybridization (WISH) with *foxa4*, *march8* and *cdx1* antisense RNA probes.
(F,I, L) Dot plot showing the quantification of *foxa4*, *march8* and *cdx1* expression patterns in NT- and IVF-embryos. For each embryo, the ratio of the signal length from dorsal blastopore to animal-vegetal axis diameter was calculated; numbers of embryos for *foxa4* staining: IVF n = 5, NT n = 9; for *march8*: IVF n = 9, NT n = 9; for *cdx1*: IVF n = 10, NT n =9;

Figure 3. Basal stem cell numbers are reduced , while epidermal progenitor cells emerge at normal rates in cloned embryos

(A) Schematic overview of cell type early specification in mucociliary epidermis in *Xenopus* embryos.
(B-E) Expression levels of *foxj1*, *foxi1*, *otog* and *tp63* in NT- and IVF-cells shown in UMAP plots.
(F) Representative images of IVF- and NT- stage 12 embryos stained by whole mount *in situ* hybridization with *foxj1*, *foxi1* and *otogl2* antisense RNA probes.
(G) Representative images of IVF- and NT- epithelia stained by FISH antisense RNA probe against *tp63* transcript in stage 12 embryos; Two channels merged: Cyan – nuclei (DAPI); Red – *sox17b* probe (Cy5); scale bar = 20 μm
(H) Quantification of (F,G). Dot plots of relative abundance of *foxj1*, *foxi1*, *otogl2* and *tp63* positive cells in NT- and IVF- epithelia; for *foxj1*: IVF n = 5, NT n = 5; for *otogl2*: IVF n = 8, NT n = 6; for *foxi1*: IVF n = 10 NT n = 7; for *tp63*: IVF n = 10, NT n = 7; Statistical test: Student's unpaired t-test.

Figure 4. Basal stem cell loss and defective mature epidermis coincide with increased cell death in cloned embryos

(A) Schematic representation of cell type-specific morphogenesis and cell maturation in mucociliary epidermis of *Xenopus* embryos.
(B) Representative images of IVF- and NT-embryos at tailbud stage stained by whole mount immuno-histochemistry using monoclonal antibody anti-Tp63 protein; scale bar = 1 mm.
(C) Bar plot showing proportions of NT- and IVF-embryos that experienced loss of Tp63 positive basal stem cells (BSC) in their epidermis.
(D) Immunofluorescence staining of Tp63 positive basal stem cells and α-ac-tubulin positive multiciliated cells in epidermis of tailbud stage NT- and IVF-embryos. Antibody anti-ZO-1 (tight junction protein) was used for the detection of cell borders.
(E) Quantified proportions of epidermal cell types in IVF- and NT-embryos at stage 32. The middle chart represents proportions in regions lacking Tp63+ cells. Data represents mean values from IVF (n = 7), NT Tp63 negative area (n = 5) and NT Tp63 positive area (n = 5).
(F) Representative images from immunofluorescence of representative NT- and IVF- embryos at tailbud stage 32 using anti-cleaved Caspase-3 (Asp175) antibody. In the merged image: nuclei in cyan (DAPI); cleaved Casp3 in red.

(G) Quantification of cells positive for activated (cleaved) Caspase-3 per lateral side of stage 32 embryo. NT (n = 10) and IVF (n = 15).*p < 0.01, statistical test: Student's unpaired t-test.

Figure 5. Cell differentiation defects are recapitulated by the ectopic expression of ON-memory genes in fertilized embryos

(A) Schematic of the microinjection experiment. The injected ventral blastomeres of 8-cell stage embryos are coloured in black.
(B) Images of representative embryos at stage 32 stained with anti-Tp63 immuno-histochemistry. Uninj.: uninjected, OE: injected embryos overexpressing Sox17b, Foxa4 and Kdm5b^{ci}. Scale bar = 0.2 mm.
(C) Proportions of embryos displaying an area with a depletion of Tp63+ basal stem cell (BSC) (white bar) across the different conditions.
(D) Immunofluorescence staining of Tp63+ basal stem cells, α-ac-tubulin+ multiciliated cells and ZO-1+ cell borders in epidermis of uninjected embryos or embryos overexpressing Sox17b, Foxa4 and control proteins. In merged image: Tp63 in magenta, ZO-1 in green, α-ac-tubulin in blue.
(E) Quantified proportions of epidermal cell types at stage 32 in each condition. The second chart from the top shows the proportions in regions lacking Tp63+ cells, and the third chart from the top shows epidermal cell proportions in regions with normal Tp63+ cell distribution in Sox17b OE embryos. Data represents mean values from Uninj, Sox17b OE Tp63 neg (n = 5), Sox17b OE Tp63 pos (n = 5), ctrl OE (n = 6).
(F) Representative images of immunofluorescent staining for cleaved Caspase-3 and FLAG-Sox17b or FLAG-Foxa4 overexpressed proteins. Merged Image: Caspase 3 in magenta, DNA in green, FLAG tagged protein in blue.
(G) Quantification of cells positive for activated (cleaved) Caspase-3 per lateral side of stage 32 embryo. Uninj (n = 19), Sox17b OE (n = 10), Foxa4 OE (n = 16), ctrl OE (n = 7).).*p < 0.01, statistical test: Student's unpaired t-test.

Figure 6. Reducing expression of key endoderm ON-memory gene Sox17b rescues epidermal defects observed in endoderm derived NT-embryos

(A) Schematic of the rescue experiment where NT-embryos were injected with antisense morpholinos (MO) into ventral blastomeres (in black) at 8-cell stage.
(B) Immunohistochemistry for Tp63 protein at tailbud stage embryos in different conditions indicated at the top. Scale bar = 0.1 mm. Ctrl - control, MO - antisense morpholino.
(C) Proportions of embryos showing perturbations in the composition of Tp63+ cells in the epidermis in the three conditions indicated on the vertical axis.
(D) Immunofluorescence for activated Caspase-3 (Magenta), DAPI (Cyan) and fluorescent dextran (Green) in IVF, NT and NT embryos injected with *sox17b* morpholino.
(E) Quantification of cells positive for activated (cleaved) Caspase-3 per lateral side of stage 32 embryo. IVF, NT Ctrl, NT *sox17b* MO.).*p < 0.01, statistical test: Student's unpaired t-test.

Supplementary Figure 1.

(A) Violin plots showing the number of unique molecular identifiers (UMIs) identified per cell for each cluster. Red and green colours represent IVF and NT conditions, respectively.

(B) Violin plot showing the number of genes identified per cell for each cluster. Red and green colors represent IVF and NT conditions, respectively.
(C) Variation of Information (VI) between cell clusterings, averaged over M=50 random sub-samplings of the highly variable genes, versus the resolution parameter of the Louvain algorithm. Colours indicate different values of the number of neighbours used to compute the Shared Nearest Neighbour (SNN) graph of the cells.
(D) UMAP plot of the scRNA-seq data coloured according to the cell clusters identified via the Louvain algorithm.
(E) Stacked barplot showing the composition of each cell cluster in terms of sequencing batches. Cell percentages are normalised by the total number of cells in each experiment (IVF1 and NT1 vs IVF2 and NT2). IVF1 and NT1 belong to the first technical replicate. IVF2 and NT2 belong to the second one.
(F) Heatmap showing the scaled expression level of selected marker genes (columns) of the identified cell clusters (rows), used for cell type annotation. Cells are ordered by cluster assignment and, within each cluster, by condition (IVF or NT), as indicated by the two coloured bars. Expression levels are zero-centred and scaled by the standard deviation for each gene.
(G-H) Feature plots showing the expression levels of *krt* and *sox11* transcripts in epidermal cells at stage 12.
(I) UMAP plot of cells in cluster 9. Colours correspond to IVF and NT conditions (see panel K).
(J) UMAP plot of cluster 9 (Mixed states) sub-clustering. Colours correspond to two different subclusters identified in cluster 9.
(K) Detailed view of NT- and IVF- cells of Cluster 9 on the UMAP plot depicted in panel D. Middle and bottom UMAP plots show expression levels of *sox17b* or *foxi1* in cells of cluster 9.
(L) Cell-cycle analysis of NT and IVF epidermal cells per cluster using the cyclone algorithm. Each bar represents the proportion of cells in G1, G2/M and S-phase of the cell cycle.
(M) Box plot of the RNA velocity confidence values per cell cluster, for IVF- and NT-cells separately.

Methods

Xenopus Laevis culture

Adult *Xenopus Laevis* animals were obtained from Nasco (901 Janesville Avenue, P.O. Box 901, Fort Atkinson, WI 53538-0901, USA) and Xenopus1 (Xenopus 1, Corp. 5654 Merkel Rd. Dexter, MI. 48130) All frog maintenance and care were conducted according to German Animal Welfare Act. Research animals were used in accordance with guidelines approved and licensed by ROB-55.2-2532.Vet_02-23-126.

Embryo handling

Xenopus eggs were collected*, in vitro* fertilized and handled as described[35]. Developmental stage of embryos was determined according to Nieuwkoop and Faber[36].

Endoderm donor cell preparation

Donor cells for nuclear transfer were isolated from endoderm tissue of neurula stage embryos (stage 21). Before procedure, the jelly coat was removed from embryos using 2% cysteine

solution in 0.1x MMR (pH7.8). Embryos were then washed three times with 0.1xMMR and transferred into 2% agarose coated Petri dish containing calcium- and magnesium-free Modified Barth's Saline (1xMBS, 88 mM NaCl, 1 mM KCl, 2.5 mM NaHCO3, 10 mM HEPES-NaOH pH7.4) supplemented with 1 mM EDTA and 0.1% BSA. Using two Dumont Nr. 5 tweezers, the inner endoderm mass was isolated from embryos and kept for 20 min in the same MBS buffer until endoderm cells were completely dissociated and a single cell suspension was achieved.

Nuclear transfer

Freshly collected eggs were placed on small piece of paper and transferred into UV crosslinker (Jena Analytic, # CL-1000) equipped with UV-C G8T5/SW lamps to be exposed 3000J/cm2 of 254 nm light. Within 1 min after the exposure, donor cell nuclei were transplanted to these enucleated and partially dejellied eggs using polished glass micropipettes. After NT, embryos were placed into Marc's Modified Ringers buffer (1xMMR, 100mM NaCl, 2mM KCl, 1mM MgSO4, 2mM CaCl2, 0.1mM EDTA, 5mM HEPES pH 7.8). Once the NT embryos reached stage 4, buffer was exchanged to 0.1xMMR and embryos were further cultured at 14°C. As control for egg and embryo quality, *in vitro* fertilized (IVF) embryos were prepared from the same egg batch.

Single cell preparation from epidermal explants

Epidermal tissues were isolated from IVF- and NT-embryos of stage 12 and shortly washed in Newport 2.0 dissociation buffer (100 mM sodium isethionate, 20 mM sodium pyrophosphate, 10 mM CAPS, 20 mM glucose, pH10.5 adjusted by NaOH) to remove calcium ions. Epidermal tissues were transferred into fresh dissociation buffer in BSA coated low bind microcentrifuge tubes and incubated for 30min at 18°C with agitation. During the period of incubation, tubes were inverted several times. Resulting single cell suspensions were resuspended in PBS-BSA solution and filtered through 30 μm cell strainers. Cells were washed 3-times with 1 ml PBS-BSA, counted and analysed for viability. 2500 cells with viability more than 95% and not detectable RNA in supernatant were submitted for library preparation.

Single cell capturing, barcoding and library preparation

In each experiment, single cell suspensions from pools of 5 IVF- or 5 NT- epidermal tissue samples were processed using Chromium 10x Genomics platform to generate single cell libraries. Libraries from all samples were pooled and sequenced in two lanes using Illumina HiSeq 2500 to generate paired-end 100-bp data.

Single cell RNA sequencing data pre-processing

Single cell libraries were processed using the 10x Genomics Cell Ranger pipeline (v2.2.0 and v3.0.2 for 10x chemistry 2 and 3 respectively), with reads aligned to *Xenopus* laevis genome (v9.1) using STAR[37]. The estimated number of cells was calculated based on the distribution of barcode counts and data realigned using --force-cells option. Quality control of individual libraries was based on the mean number of detected genes per cell (typically in the range of 3k-6k) and mean UMIs per cell (typical values 80k-500k). Mapping efficiency was high, with >75% reads mapped to the genome, with 40-60% mapped to the transcriptome. After quality control, 3405 cells with more than 6000 UMIs per cell and more than 2000 detected genes per

cell were obtained: 566 for SIGAA2 (IVF1), 514 for SIGAB2 (NT1), 1275 for SIGAH5 (IVF2) and 1050 for SIGAH12 (NT2), for a total of 1841 IVF and 1564 NT good quality cells.

Data normalization, identification of highly variable genes and batch integration

R package Seurat v3[38] was used for data normalization, selection of highly variable genes for each batch and batch integration. The data for each batch was normalized using the function "NormalizeData", and the top 2000 highly variable genes were computed using the function "FindVariableGenes", with the "vst" selection method. The data for the four batches were then integrated using the functions "FindIntegrationAnchors", using the first 20 dimensions from the Canonical Correlation Analysis (CCA), and "IntegrateData". The integrated data was scaled using the function "ScaleData".

Dimensionality reduction and clustering of cells

Using the matrix of the integrated and scaled data, a Principal Component Analysis (PCA) using the function "RunPCA" was performed, and a UMAP using the function "RunUMAP" was generated. Then, a k-nearest neighbour (KNN) graph of the cells using the function "FindNeighbors" was computed, and cell clustering was performed on the matrix of the integrated and scaled data with the function "FindClusters". The number of neighbors ("k.param" parameter in "FindNeighbors" function) and the resolution for clustering ("resolution" parameter in "FindClusters" function) were chosen using the information theoretic criterion defined in[39], whose rationale is to choose the set of parameters (in our case "k.param" and "resolution") that minimizes the variation of information between the clusters obtained on the full dataset and those obtained after performing M=50 random sub-samplings of 50% of the highly variable genes, while maximizing the resolution. Following this criterion, we selected "resolution"=0.5 and "k.param=20" (Fig.S1C), and 11 cell clusters were initially identified. Cells from one cluster with notably lower UMI counts and a low number of genes detected compared to the other clusters (cluster 0 in Fig. S1A-B) were removed from downstream analyses. Moreover, a subset of 14 cells from cluster 6 were identified as outliers using the outlier detection algorithm based on the distance from the k-nearest neighbours for each cell (Python package "PyOD"[40] and removed from downstream analyses.
Cells from cluster 9 were subclustered using the batch integration pipeline from Seurat v3 as applied as described above, but setting "k.filter"=10, "k.score"=10 and retaining the first 10 dimensions in the CCA in the function "FindIntegrationAnchors", and setting "k.weight"=10 in the function "IntegrateData", given the smaller number of cells in this case. Then, cell clustering with default parameters was performed, obtaining two clusters.

Cell type annotation

Cell clusters were annotated based on the expression of known markers and on information available at *Xenopus* Bioinformatics Database (Xenbase)[25] and *Xenopus* scRNA-seq datasets[22–24].

Cell cycle annotation

The cell cycle phase of cells was annotated using the "cyclone" function of the "scran" R package in Lun et al.[41]

In silico cell type composition analysis

Changes in cell type proportions between IVF and NT were tested following the method used in Haber et al. (2017)[42]. The detected number of cells of each cell type in each batch is modelled as a random count variable using a Poisson process. The total number of cells in each batch is provided as an offset variable, representing the rate of detection, while the condition (IVF or NT) represents a covariate. The model was fitted using the function "glm" from the "stats" package. The p-value for each cell type was computed using a Wald test on the regression coefficient, and corrected with the FDR method using the function "p.adjust".

CellRank analyses

We conducted a fate mapping analysis on the scRNA-seq data using the CellRank python package (v1.1.0)[26]. Since there is not a well-established method to take into account batch effects in RNA velocity analysis, we considered only the SIGAH5 (IVF) and the SIGAH12 (NT) samples due to their larger number of cells compared to the other samples and because the mixed states cluster is mostly made of cells from these two samples.
First, we computed the matrices of spliced and unspliced counts using velocyto[27]. Next, we used the "scvelo" package[43] to compute the RNA velocities of single cells, separately for the IVF and NT samples. We filtered out genes expressed in fewer than 10 cells (function "scvelo.pp.filter_genes"), with fewer than 20 or 10 spliced or unspliced counts, respectively (function "scvelo.pp.filter_and_normalize"), we normalized the count matrices and selected the top 2000 highly variable genes. Next, we computed the moments for velocity estimation using 30 neighbours and the top 30 Principal Components (PCs) (function "scvelo.pp.moments"). Then, we estimated the kinetic parameters of the RNA velocity model with the function "scvelo.tl.recover_dynamics" and we computed the RNA velocities using the dynamical model with the function "scvelo.tl.velocity". We estimated the RNA velocity confidence with the function "scvelo.tl.velocity_confidence" and we represented the distribution of the velocity confidence values in each cell cluster, separately for IVF and NT cells, using the "seaborn" python package ("boxplot" function). The confidence scores were relatively high for all clusters in the data (Fig. S1M), which supported the use of RNA velocity in the analysis of differentiation dynamics that we performed with CellRank, as described below.
We used the CellRank function "cellrank.tl.terminal_states" (with default parameters) to infer the terminal states of cellular dynamics based on the velocity kernel and on the connectivity kernel, which is based on gene expression similarities. The plots of the terminal states for IVF and NT cells are obtained using the function "cellrank.pl.terminal_states". In those plots, the colors indicate the most likely terminal state estimated for each cell, with darker colors corresponding to higher-confidence estimations.

ON and OFF memory genes identification and data representation

We performed a differential expression analysis between IVF- and NT- cells, across all cell clusters combined, using the R package "DESeq2", measuring the effect of the condition (NT vs IVF) while controlling for the effect of the experiment (SIGAA2-SIGAB2 vs SIGAH5-SIGAH12). We considered as significantly up- or down-regulated genes those having

FDR<0.05. Then, relying on the bulk RNA-seq data from[11], we defined global ON- and OFF-memory genes using the following criteria:

- ON-memory genes: mean($RPKM_{Donor}$)> 1 (from bulk RNA-seq); $FDR_{Donor\ vs\ IVF} < 0.05$ (from bulk RNA-seq); $log2(FC)_{NT/IVF} > 1$ (from scRNA-seq);
- OFF-memory genes: mean($RPKM_{Donor}$)< 20 (from bulk RNA-seq); $FDR_{Donor\ vs\ IVF} < 0.05$ (from bulk RNA-seq); $log2(FC)_{NT/IVF} < -1$ (from scRNA-seq).

Next, we repeated the same type of differential expression analysis between IVF and NT cells considering each cell cluster separately, and we defined cluster-specific ON- and OFF-memory genes using the same criteria described above. The results on ON- and OFF-memory genes analyses were visualized using R packages "UpSetR"and "ggplot2" using default settings.

Whole mount RNA *in situ* hybridization

For RNA *in situ* hybridization analysis, embryos were anesthetized in 0.05% Tricaine, fixed 1h at RT in MEMFA (0.1 M MOPS pH7.4, 2 mM EGTA, 1 mM MgSO4, 3.7% Formaldehyde), washed with PBS and stored in ethanol at -20°C. Each developmental stage is represented by at least 3 biological replicates. Embryos of each biological replicate is derived from a different mating pair. Plasmids for preparation of digoxigenin labelled antisense RNA probes were pBluescriptIISK-otogl2.L, pCMV-Sport6-*foxa1.S*[42] or cloned from cDNA stage 12 embryos (*tp63.S*, *foxi1.S*, *foxj1.S*, *hoxd1.S*, *cdx1.S*, *march8.L*). Primers used for amplification of selected genes are listed in "List of used primers" below. All amplicons were subcloned into pCS2+ vector via FseI/AscI restriction sites or using NEBuilder HiFi DNA Assembly Master Mix (New England Biolabs, #E2621S). The plasmids containing *tp63.S* and *hoxd1.S* were linearized using NotI; plasmids with *foxi1.S*, *foxj1.S*, *cdx1.S* genes were linearized using HindIII. The 4 μg of linearized vectors were transcribed 5h using RiboMAX™ Large Scale T7 RNA polymerase kit (Promega, #P1300) with 5.6 mM Ribo m7G Cap Analog (Promega, #P1711) in reaction and resulting antisense probes purified with RNeasy Mini-Kit (Qiagen, #74106). Fluorescent RNA *in situ* hybridization used tyramide amplification after addition of probes and incubation with horseradish peroxidase antibody conjugate (Sigma-Aldrich, #11207733910) as described in Lea et al. (2012)[43]. Chromogenic RNA *in situ* hybridization was performed as described in Sive et al. (2000)[32]. After the AP staining (BM purple, Sigma-Aldrich, #11442074001) the embryos were dehydrated with 75%EtOH/PBS and bleached 1-2h in bleaching solution (1% H2O2, 5% Formamide, 0.5xSSC). After refixation in MEMFA, the embryos were photographed using a Leica M205FA stereomicroscope. All images are presented as a compound z-stack projection.

Immunohistochemistry for Tp63 protein

The experiment was conducted similarly as in Lu et al.(2001)[44] and Tseng et al. (2007)[45]. In short, the embryos at stage 32-33 were fixed overnight at 4°C in MEMFA, rinsed with 1xPBS and incubated at least 24 hrs in 100% ethanol at -20°C. Embryos were then rehydrated with stepwise washes using 75%, 50% and 25% ethanol in PBST (1xPBS+0.1% Tween20). Endogenous alkaline phosphatase was inactivated by washing embryos once with 50% formamide/PBS solution and incubating them in the same solution at 65°C for 2 hrs. Embryos

were then permeabilized twice 10 min in 1xPBS+0.2% TritonX-100, blocked in antibody buffer (1%BSA in 1xPBS+0.02% Tween20) for 1 h at RT and stained overnight (o/n) with primary antibody anti-Tp63 (Abcam, clone 4A4, #ab735) diluted 1:1000 at 4°C. Next day six washes with PBST (1 h each wash), 1 h of blocking with antibody buffer and o/n incubation with anti-mouse AP conjugated antibody (Abcam, #ab97262) in dilution 1:1000 at 4°C. Finally, the embryos were again washed 6-7 times with PBST (each wash for 1h, the last one was o/n), incubated 20 min in AP buffer (100mM Tris/HCl pH=9.5, 100mM NaCl, 50mM MgCl2) and stained in dark with NBT (33.8 μg/ml)/BCIP(17.5 μg/ml) in AP buffer for 1.5 h at 4°C. After the signal was developed, embryos were fixed 30 min in 4% PFA/in 1xPBS solution, dehydrated with 30 min with 75% ethanol in 1xPBS and bleached for 2-4 h with bleaching solution until Tp63 positive cells were clearly visible.

Immunofluorescence

The detection of epidermal cells was performed as described in Walentek et al. (2018)[25] with some minor changes. If ZO-1 staining took place on embryos immunohistochemically stained for Tp63 (see above), samples were pre-treated with 2ug/ml proteinase K solution for 8min at 37°C. If MEMFA-fixed embryos were used, they were first rehydrated. Embryos were blocked 1 hr in antibody buffer at RT and incubated depending on the experiment with following primary antibodies: Anti-Acetylated Tubulin (1:1000, Sigma-Aldrich, #T6793); anti ZO-1 (1:200, Invitrogen, #61-7300) or β-catenin (1:5, rat, hybridoma supernatant gifted from prof. Ralph Rupp); anti-cleaved Caspase-3 (Asp175) (1:400, CST, #9661). After o/n incubation at 4°C, samples were washed 5x30min with TBST (1xTBS+0.02% Tween20), blocked with TBS antibody buffer (1%BSA in TBST) for 1h at RT and stained o/n with antibody fluorescent conjugates (1:500, anti-mouse-AF488, anti-rabbit-AF555 or 647, anti-rat-AF546, all from Thermo-Fisher Scientific), 1mg/ml DAPI solution 1:500 and PNA-lectin AF594 1:1000 (1mg/ml, Invitrogen, #L32459). Before mounting into Vectashield (Vector labs, H-1000), embryos were washed 3x30 min in TBST. Imaging was conducted on Leica SP8 system using oil 40x objective, Navigator of the LASX software, z-step 0.7μm, 1024x1024 per image, 0.75x zoom.

Microinjections and embryo treatments

mRNA for *sox17b.1.S*, *foxa4.L* and mouse *Kdm5b*^{ci} transcription factors was generated from pCS2+ vectors, where coding sequence of all genes was cloned from cDNA and fused with *Xenopus* globin 5' and 3' UTRs for increasing the translation potential[46]. Upon cloning into pCS2+ vector each gene was tagged on its N-teminus with triplet of FLAG sequence (DYKDDDDK). RiboMAX™ Large Scale SP6 RNA polymerase kit (Promega, #P1280) was used for *in vitro* transcription on XbaI linearized pCS2+ vectors. 250 pg of mRNA in 4.6nl was injected with 100 pg of fluorescent dextran AF488 (Invitrogen, #D22910) into both ventral blastomeres of 8-cell stage embryos. At late tailbud stage, embryos were fixed and stained in a similar way a as described in sections "Immunohistochemistry for Tp63 protein" and "immunofluorescence". Primary antibodies used for immunofluorescence were: anti-cleaved Caspase-3 (Asp175) (1:400, rabbit, CST, #9661) and anti-FLAG (1:400, mouse, Sigma-Aldrich, #F3165) for detection of overexpressing proteins. Secondary antibodies anti-mouse-AF555, anti-rabbit-AF647 (all from Thermo-Fisher Scientific)

were used in dilution 1:500. Morpholino oligo *sox17β* was used the same as in Clements et al. (2003)[47]. and injected in concentration of 5 ng per 8-cell stage ventral blastomere of NT or IVF embryos.

List of used Primers:

Primer_name	**Primer_sequence_for_cloning_into_pCS2+_vector**[a]
*cdx1.S*_fwd	TAAGCAGGCCGGCCTatgtacgtgggttatcttttgg
*cdx1.S*_rev	AGCTGAGGCGCGCCttacgaaagatattcttccttgatagg
*march8.L*_fwd	TAAGCAGGCCGGCCTatgaagcttcagaatgagaaaac
*march8.L*_rev	AGCTGAGGCGCGCCttaaacctgaaggatcgctg
*foxj1.S*_fwd	TAAGCAGGCCGGCCTatgtttgacctgcccag
*foxj1.S*_rev	AGCTGAGGCGCGCCttatatataggaacccaaggacg
*foxi1.S*_fwd	TAAGCAGGCCGGCCTatgagtgcctttgatccac
*foxi1.S*_rev	AGCTGAGGCGCGCCttatacttctgtgccctctc
*tp63.S*_Gib_fwd	gtggaggcggccgcggccggcctatgttgtatctggaaaacag
*tp63.S*_Gib_rev	atacgactcactatagggggcgcgccttatggatacattgaatggc
*hoxd1.S*_Gib_fwd	gtggaggcggccgcggccggcctatgaattcctacctagaatac
*hoxd1.S*_Gib_rev	atacgactcactatagggggcgcgccttagttttggctggcttg
*foxa4.L*_Fwd	TAAGCAGGCCGGCCTatgctaaatagagtcaaattgg
*foxa4.L*_Rev	AGCTGAGGCGCGCCttaaagggagctgaggatag
*sox17b.1.S*_Fwd	TAAGCAGGCCGGCCTATGAGCAGCCCGGACTGC
*sox17b.1.S*_Rew	AGCTGAGGCGCGCCTTATACGCCACAATAGTCATAGTAG

[a] This table contains sequences of primers used for amplification of coding sequences from cDNA and cloning it via FseI/AscI into pCS2+ vector

References

1. Gurdon, J.B., Elsdale, T.R., and Fischberg, M. (1958). Sexually Mature Individuals of Xenopus laevis from the Transplantation of Single Somatic Nuclei. Nature *182*, 64–65. https://doi.org/10.1038/182064a0.

2. Weintraub, H., Tapscott, S.J., Davis, R.L., Thayer, M.J., Adam, M.A., Lassar, A.B., and Miller, A.D. (1989). Activation of muscle-specific genes in pigment, nerve, fat, liver, and fibroblast cell lines by forced expression of MyoD. Proc Natl Acad Sci U S A *86*, 5434–5438. https://doi.org/10.1073/pnas.86.14.5434.

3. Takahashi, K., and Yamanaka, S. (2006). Induction of pluripotent stem cells from mouse embryonic and adult fibroblast cultures by defined factors. Cell *126*, 663–676. https://doi.org/10.1016/j.cell.2006.07.024.

4. Blau, H.M., Chiu, C.P., and Webster, C. (1983). Cytoplasmic activation of human nuclear genes in stable heterocaryons. Cell *32*, 1171–1180. https://doi.org/10.1016/0092-8674(83)90300-8.

5. Matoba, S., and Zhang, Y. (2018). Somatic Cell Nuclear Transfer Reprogramming: Mechanisms and Applications. Cell Stem Cell *23*, 471–485. https://doi.org/10.1016/j.stem.2018.06.018.

6. Hörmanseder, E. (2021). Epigenetic memory in reprogramming. Current Opinion in Genetics & Development *70*, 24–31. https://doi.org/10.1016/j.gde.2021.04.007.

7. Jullien, J., Pasque, V., Halley-Stott, R.P., Miyamoto, K., and Gurdon, J.B. (2011). Mechanisms of nuclear reprogramming by eggs and oocytes: a deterministic process? Nat Rev Mol Cell Biol *12*, 453–459. https://doi.org/10.1038/nrm3140.

8. Gao, S., Chung, Y.G., Williams, J.W., Riley, J., Moley, K., and Latham, K.E. (2003). Somatic Cell-Like Features of Cloned Mouse Embryos Prepared with Cultured Myoblast Nuclei1. Biology of Reproduction *69*, 48–56. https://doi.org/10.1095/biolreprod.102.014522.

9. Hirasawa, R., Matoba, S., Inoue, K., and Ogura, A. (2013). Somatic Donor Cell Type Correlates with Embryonic, but Not Extra-Embryonic, Gene Expression in Postimplantation Cloned Embryos. PLOS ONE *8*, e76422. https://doi.org/10.1371/journal.pone.0076422.

10. Matoba, S., Liu, Y., Lu, F., Iwabuchi, K.A., Shen, L., Inoue, A., and Zhang, Y. (2014). Embryonic development following somatic cell nuclear transfer impeded by persisting histone methylation. Cell *159*, 884–895. https://doi.org/10.1016/j.cell.2014.09.055.

11. Hörmanseder, E., Simeone, A., Allen, G.E., Bradshaw, C.R., Figlmüller, M., Gurdon, J., and Jullien, J. (2017). H3K4 Methylation-Dependent Memory of Somatic Cell Identity Inhibits Reprogramming and Development of Nuclear Transfer Embryos. Cell Stem Cell *21*, 135-143.e6. https://doi.org/10.1016/j.stem.2017.03.003.

12. Liu, W., Liu, X., Wang, C., Gao, Y., Gao, R., Kou, X., Zhao, Y., Li, J., Wu, Y., Xiu, W., et al. (2016). Identification of key factors conquering developmental arrest of somatic cell cloned embryos by combining embryo biopsy and single-cell sequencing. Cell Discov *2*, 1–15. https://doi.org/10.1038/celldisc.2016.10.

13. Firas, J., Liu, X., and Polo, J.M. (2014). Epigenetic memory in somatic cell nuclear transfer and induced pluripotency: Evidence and implications. Differentiation *88*, 29–32. https://doi.org/10.1016/j.diff.2014.09.001.

14. Liu, Y., Wu, F., Zhang, L., Wu, X., Li, D., Xin, J., Xie, J., Kong, F., Wang, W., Wu, Q., et al. (2018). Transcriptional defects and reprogramming barriers in somatic cell nuclear reprogramming as revealed by single-embryo RNA sequencing. BMC Genomics *19*, 734. https://doi.org/10.1186/s12864-018-5091-1.

15. Boiani, M., Eckardt, S., Schöler, H.R., and McLaughlin, K.J. (2002). Oct4 distribution and level in mouse clones: consequences for pluripotency. Genes Dev. *16*, 1209–1219. https://doi.org/10.1101/gad.966002.

16. Chung, Y.G., Matoba, S., Liu, Y., Eum, J.H., Lu, F., Jiang, W., Lee, J.E., Sepilian, V., Cha, K.Y., Lee, D.R., et al. (2015). Histone Demethylase Expression Enhances Human Somatic Cell Nuclear Transfer Efficiency and Promotes Derivation of Pluripotent Stem Cells. Cell Stem Cell *17*, 758–766. https://doi.org/10.1016/j.stem.2015.10.001.

17. Blelloch, R., Wang, Z., Meissner, A., Pollard, S., Smith, A., and Jaenisch, R. (2006). Reprogramming Efficiency Following Somatic Cell Nuclear Transfer Is Influenced by the Differentiation and Methylation State of the Donor Nucleus. STEM CELLS *24*, 2007–2013. https://doi.org/10.1634/stemcells.2006-0050.

18. Enright, B.P., Kubota, C., Yang, X., and Tian, X.C. (2003). Epigenetic Characteristics and Development of Embryos Cloned from Donor Cells Treated by Trichostatin A or 5-aza-2′-deoxycytidine1. Biology of Reproduction *69*, 896–901. https://doi.org/10.1095/biolreprod.103.017954.

19. Kishigami, S., Mizutani, E., Ohta, H., Hikichi, T., Thuan, N.V., Wakayama, S., Bui, H.-T., and Wakayama, T. (2006). Significant improvement of mouse cloning technique by treatment with trichostatin A after somatic nuclear transfer. Biochem Biophys Res Commun *340*, 183–189. https://doi.org/10.1016/j.bbrc.2005.11.164.

20. Gurdon, J.B. (1960). The developmental capacity of nuclei taken from differentiating endoderm cells of Xenopus laevis. J Embryol Exp Morphol *8*, 505–526.

21. Becht, E., McInnes, L., Healy, J., Dutertre, C.-A., Kwok, I.W.H., Ng, L.G., Ginhoux, F., and Newell, E.W. (2018). Dimensionality reduction for visualizing single-cell data using UMAP. Nat Biotechnol. https://doi.org/10.1038/nbt.4314.

22. Jansen, C., Paraiso, K.D., Zhou, J.J., Blitz, I.L., Fish, M.B., Charney, R.M., Cho, J.S., Yasuoka, Y., Sudou, N., Bright, A.R., et al. (2022). Uncovering the mesendoderm gene regulatory network through multi-omic data integration. Cell Rep *38*, 110364. https://doi.org/10.1016/j.celrep.2022.110364.

23. Briggs, J.A., Weinreb, C., Wagner, D.E., Megason, S., Peshkin, L., Kirschner, M.W., and Klein, A.M. (2018). The dynamics of gene expression in vertebrate embryogenesis at single-cell resolution. Science *360*, eaar5780. https://doi.org/10.1126/science.aar5780.

24. Lee, J., Møller, A.F., Chae, S., Bussek, A., Park, T.J., Kim, Y., Lee, H.-S., Pers, T.H., Kwon, T., Sedzinski, J., et al. (2023). A single-cell, time-resolved profiling of Xenopus mucociliary epithelium reveals nonhierarchical model of development. Science Advances *9*, eadd5745. https://doi.org/10.1126/sciadv.add5745.

25. Fisher, M., James-Zorn, C., Ponferrada, V., Bell, A.J., Sundararaj, N., Segerdell, E., Chaturvedi, P., Bayyari, N., Chu, S., Pells, T., et al. (2023). Xenbase: key features and resources of the Xenopus model organism knowledgebase. Genetics *224*, iyad018. https://doi.org/10.1093/genetics/iyad018.

26. Lange, M., Bergen, V., Klein, M., Setty, M., Reuter, B., Bakhti, M., Lickert, H., Ansari, M., Schniering, J., Schiller, H.B., et al. (2022). CellRank for directed single-cell fate mapping. Nat Methods *19*, 159–170. https://doi.org/10.1038/s41592-021-01346-6.

27. La Manno, G., Soldatov, R., Zeisel, A., Braun, E., Hochgerner, H., Petukhov, V., Lidschreiber, K., Kastriti, M.E., Lönnerberg, P., Furlan, A., et al. (2018). RNA velocity of single cells. Nature *560*, 494–498. https://doi.org/10.1038/s41586-018-0414-6.

28. Walentek, P. (2018). Manipulating and Analyzing Cell Type Composition of the Xenopus Mucociliary Epidermis. Methods Mol Biol *1865*, 251–263. https://doi.org/10.1007/978-1-4939-8784-9_18.

29. Haas, M., Vázquez, J.L.G., Sun, D.I., Tran, H.T., Brislinger, M., Tasca, A., Shomroni, O., Vleminckx, K., and Walentek, P. (2019). ΔN-Tp63 Mediates Wnt/β-Catenin-Induced Inhibition of Differentiation in Basal Stem Cells of Mucociliary Epithelia. Cell Rep *28*, 3338-3352.e6. https://doi.org/10.1016/j.celrep.2019.08.063.

30. Ellis, S.J., Gomez, N.C., Levorse, J., Mertz, A.F., Ge, Y., and Fuchs, E. (2019). Distinct modes of cell competition shape mammalian tissue morphogenesis. Nature *569*, 497–502. https://doi.org/10.1038/s41586-019-1199-y.

31. Sinner, D., Kirilenko, P., Rankin, S., Wei, E., Howard, L., Kofron, M., Heasman, J., Woodland, H.R., and Zorn, A.M. (2006). Global analysis of the transcriptional network controlling Xenopus endoderm formation. Development *133*, 1955–1966. https://doi.org/10.1242/dev.02358.

32. Cieślar-Pobuda, A., Knoflach, V., Ringh, M.V., Stark, J., Likus, W., Siemianowicz, K., Ghavami, S., Hudecki, A., Green, J.L., and Łos, M.J. (2017). Transdifferentiation and reprogramming: Overview of the processes, their similarities and differences. Biochimica et Biophysica Acta (BBA) - Molecular Cell Research *1864*, 1359–1369. https://doi.org/10.1016/j.bbamcr.2017.04.017.

33. Charney, R.M., Paraiso, K.D., Blitz, I.L., and Cho, K.W.Y. (2017). A gene regulatory program controlling early Xenopus mesendoderm formation: network conservation and motifs. Semin Cell Dev Biol *66*, 12–24. https://doi.org/10.1016/j.semcdb.2017.03.003.

34. Akieda, Y., Ogamino, S., Furuie, H., Ishitani, S., Akiyoshi, R., Nogami, J., Masuda, T., Shimizu, N., Ohkawa, Y., and Ishitani, T. (2019). Cell competition corrects noisy Wnt morphogen gradients to achieve robust patterning in the zebrafish embryo. Nat Commun *10*, 4710. https://doi.org/10.1038/s41467-019-12609-4.

35. Sive, H.L., Grainger, R.M., and Harland, R.M. (2000). Early development of Xenopus laevis: a laboratory manual (Cold Spring Harbor Laboratory Press).

36. Nieuwkoop, J.F., P.D. ed. (2020). Normal Table of Xenopus Laevis (Daudin): A Systematical & Chronological Survey of the Development from the Fertilized Egg till the End of Metamorphosis (Garland Science) https://doi.org/10.1201/9781003064565.

37. Dobin, A., Davis, C.A., Schlesinger, F., Drenkow, J., Zaleski, C., Jha, S., Batut, P., Chaisson, M., and Gingeras, T.R. (2013). STAR: ultrafast universal RNA-seq aligner. Bioinformatics *29*, 15–21. https://doi.org/10.1093/bioinformatics/bts635.

38. Stuart, T., Butler, A., Hoffman, P., Hafemeister, C., Papalexi, E., Mauck, W.M., Hao, Y., Stoeckius, M., Smibert, P., and Satija, R. (2019). Comprehensive Integration of Single-Cell Data. Cell *177*, 1888-1902.e21. https://doi.org/10.1016/j.cell.2019.05.031.

39. Thowfeequ, S., Fiorentino, J., Hu, D., Solovey, M., Ruane, S., Whitehead, M., Vanhaesebroeck, B., Scialdone, A., and Srinivas, S. (2021). Characterisation of the transcriptional dynamics underpinning the function, fate, and migration of the mouse Anterior Visceral Endoderm. Preprint at bioRxiv, https://doi.org/10.1101/2021.06.25.449902 https://doi.org/10.1101/2021.06.25.449902.

40. Zhao, Y., Nasrullah, Z., and Li, Z. PyOD: A Python Toolbox for Scalable Outlier Detection. Journal of machine learning research *20.96: 1-7.*

41. Lun, A.T.L., McCarthy, D.J., and Marioni, J.C. (2016). A step-by-step workflow for low-level analysis of single-cell RNA-seq data with Bioconductor. Preprint at F1000Research, https://doi.org/10.12688/f1000research.9501.2 https://doi.org/10.12688/f1000research.9501.2.

42. Haber, A.L., Biton, M., Rogel, N., Herbst, R.H., Shekhar, K., Smillie, C., Burgin, G., Delorey, T.M., Howitt, M.R., Katz, Y., et al. (2017). A single-cell survey of the small intestinal epithelium. Nature *551*, 333–339. https://doi.org/10.1038/nature24489.

43. Bergen, V., Lange, M., Peidli, S., Wolf, F.A., and Theis, F.J. (2020). Generalizing RNA velocity to transient cell states through dynamical modeling. Nat Biotechnol *38*, 1408–1414. https://doi.org/10.1038/s41587-020-0591-3.

44. Lu, P., Barad, M., and Vize, P.D. (2001). Xenopus p63 expression in early ectoderm and neurectoderm. Mech Dev *102*, 275–278. https://doi.org/10.1016/s0925-4773(01)00315-x.

45. Tseng, A.-S., Adams, D.S., Qiu, D., Koustubhan, P., and Levin, M. (2007). Apoptosis is required during early stages of tail regeneration in Xenopus laevis. Dev Biol *301*, 62–69. https://doi.org/10.1016/j.ydbio.2006.10.048.

Figure 1

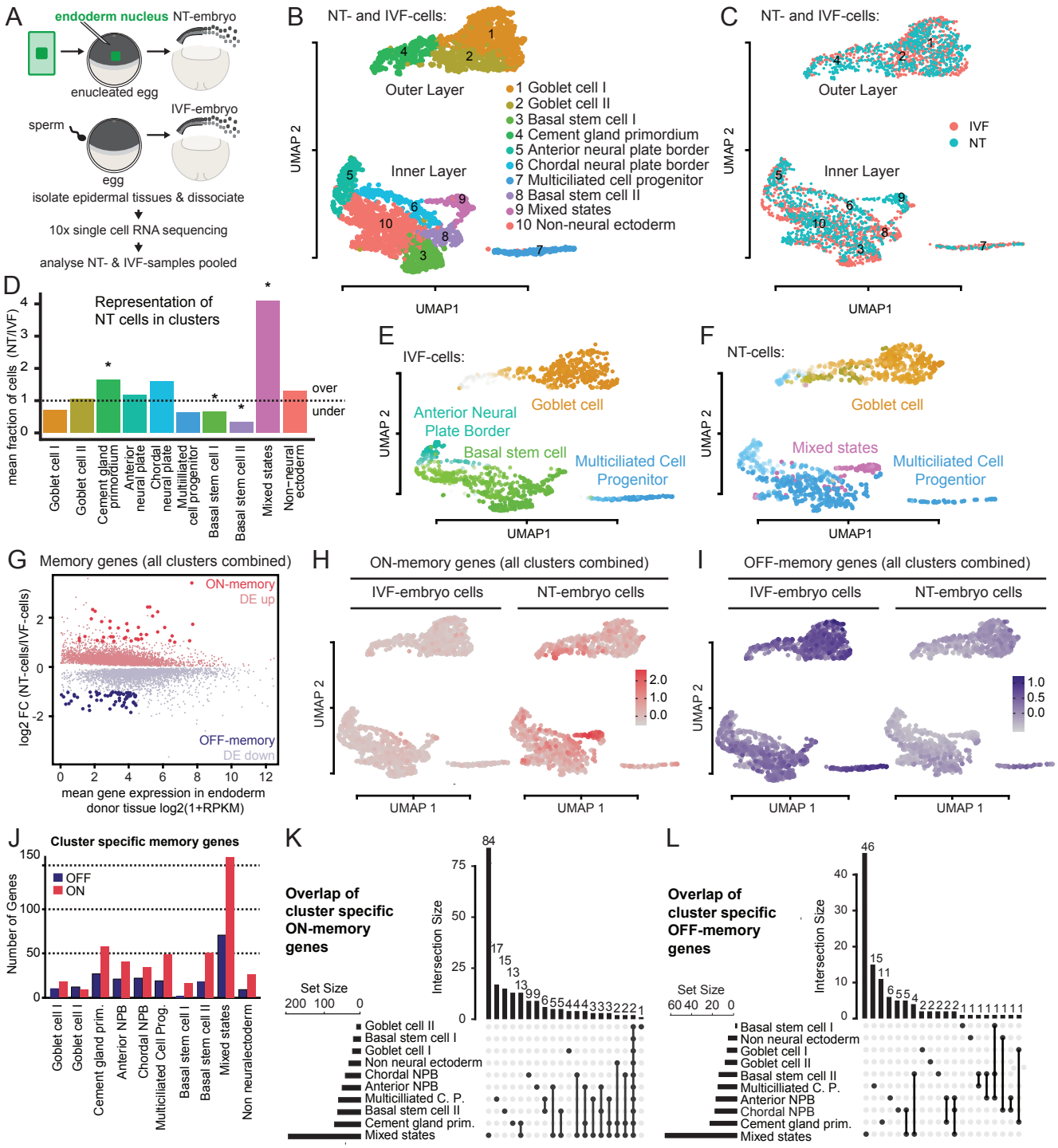

A
endoderm nucleus
NT-embryo
enucleated egg
sperm
IVF-embryo
egg
isolate epidermal tissues & dissociate
10x single cell RNA sequencing
analyse NT- & IVF-samples pooled
B
NT- and IVF-cells:
Outer Layer
Inner Layer
UMAP 2
UMAP1
1 Goblet cell I
2 Goblet cell II
3 Basal stem cell I
4 Cement gland primordium
5 Anterior neural plate border
6 Chordal neural plate border
7 Multiciliated cell progenitor
8 Basal stem cell II
9 Mixed states
10 Non-neural ectoderm
C
NT- and IVF-cells:
Outer Layer
Inner Layer
IVF
NT
UMAP 2
UMAP1
D
Representation of NT cells in clusters
mean fraction of cells (NT/IVF)
over
under
Goblet cell I
Goblet cell II
Cement gland primordium
Anterior neural plate
Chordal neural plate
Multiciliated cell progenitor
Basal stem cell I
Basal stem cell II
Mixed states
Non-neural ectoderm
E
IVF-cells:
Anterior Neural Plate Border
Goblet cell
Basal stem cell
Multiciliated Cell Progenitor
UMAP 2
UMAP1
F
NT-cells:
Goblet cell
Mixed states
Multiciliated Cell Progentior
UMAP 2
UMAP1
G
Memory genes (all clusters combined)
ON-memory
DE up
OFF-memory
DE down
log2 FC (NT-cells/IVF-cells)
mean gene expression in endoderm donor tissue log2(1+RPKM)
H
ON-memory genes (all clusters combined)
IVF-embryo cells
NT-embryo cells
UMAP 2
UMAP 1
I
OFF-memory genes (all clusters combined)
IVF-embryo cells
NT-embryo cells
UMAP 2
UMAP 1
J
Cluster specific memory genes
OFF
ON
Number of Genes
Goblet cell I
Goblet cell I
Cement gland prim.
Anterior NPB
Chordal NPB
Multiciliated Cell Prog.
Basal stem cell I
Basal stem cell II
Mixed states
Non neuralectoderm
K
Overlap of cluster specific ON-memory genes
Set Size
Intersection Size
Goblet cell II
Basal stem cell I
Goblet cell I
Non neural ectoderm
Chordal NPB
Anterior NPB
Multicilliated C. P.
Basal stem cell II
Cement gland prim.
Mixed states
L
Overlap of cluster specific OFF-memory genes
Set Size
Intersection Size
Basal stem cell I
Non neural ectoderm
Goblet cell I
Goblet cell II
Basal stem cell II
Multicilliated C. P.
Anterior NPB
Chordal NPB
Cement gland prim.
Mixed states

Figure 2

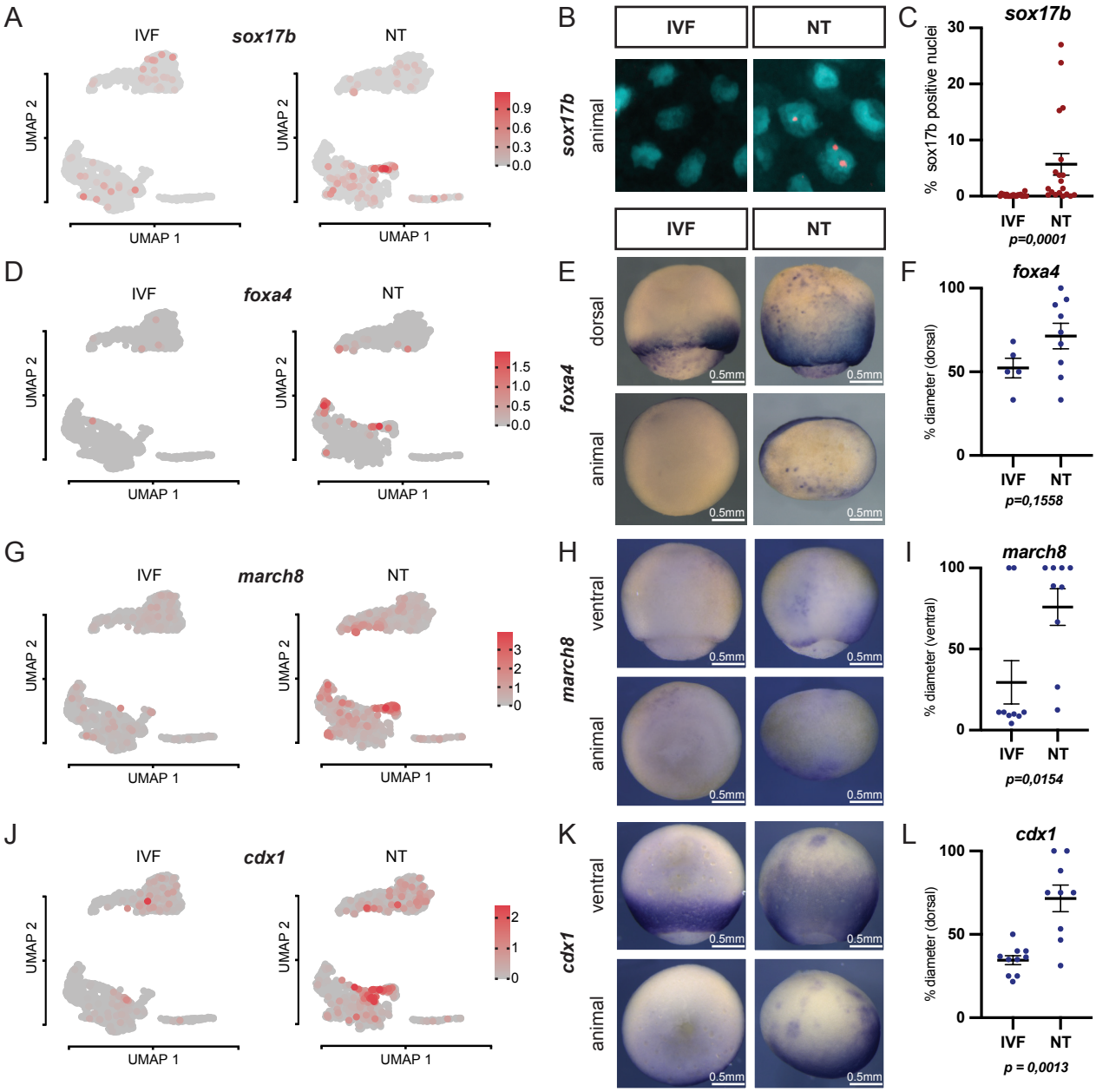

A
IVF
sox17b
NT
UMAP 2
UMAP 1
0.9
0.6
0.3
0.0
B
IVF
NT
sox17b
animal
C
sox17b
% sox17b positive nuclei
30
20
10
0
IVF
NT
p=0,0001
D
IVF
foxa4
NT
1.5
1.0
0.5
0.0
E
dorsal
animal
0.5mm
F
foxa4
% diameter (dorsal)
100
50
0
p=0,1558
G
IVF
march8
NT
3
2
1
0
H
ventral
animal
I
march8
% diameter (ventral)
p=0,0154
J
IVF
cdx1
NT
2
1
0
K
L
cdx1
% diameter (dorsal)
p = 0,0013

Figure 3

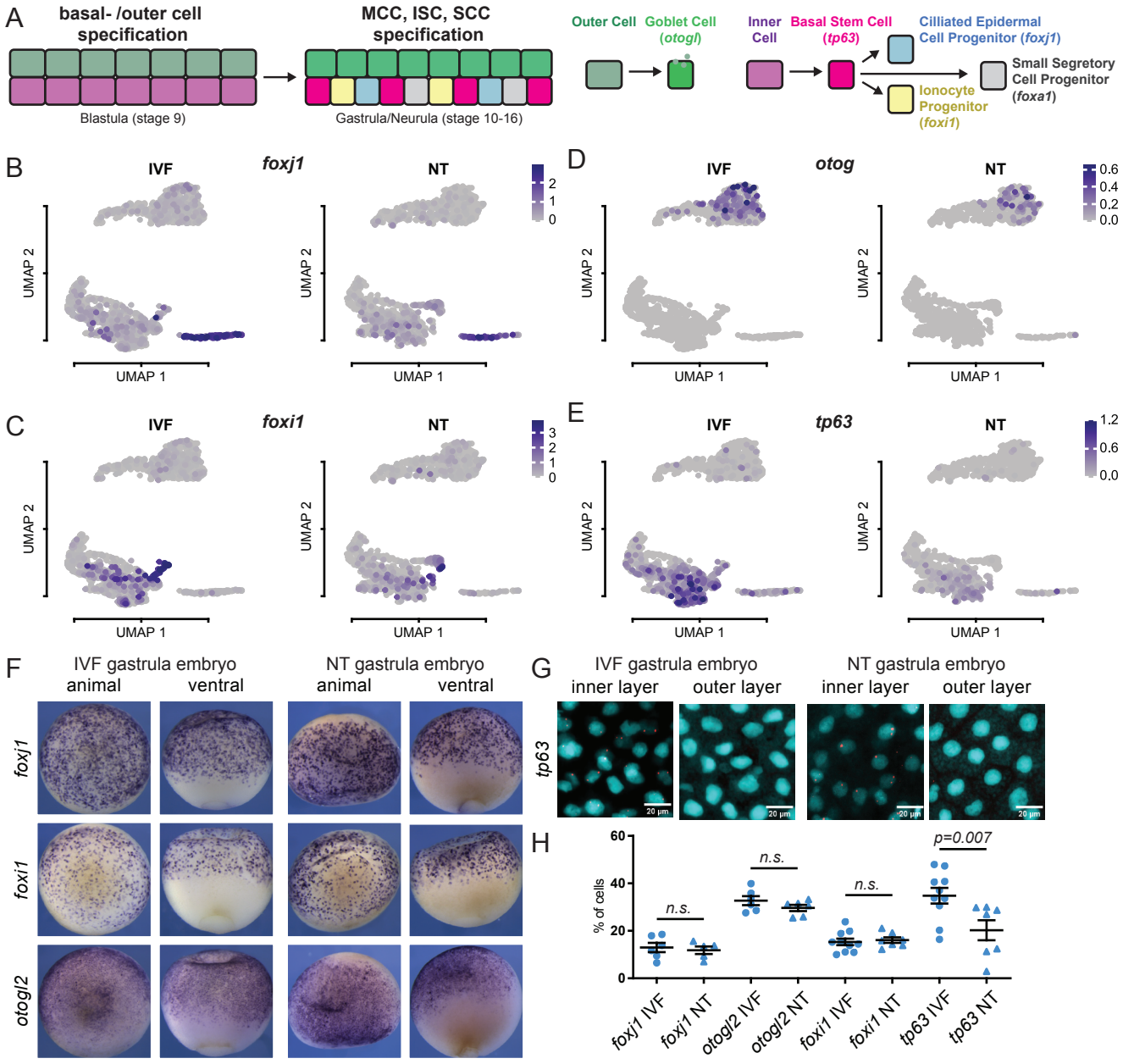

A
basal- /outer cell specification
MCC, ISC, SCC specification
Blastula (stage 9)
Gastrula/Neurula (stage 10-16)
Outer Cell
Goblet Cell (otogl)
Inner Cell
Basal Stem Cell (tp63)
Cilliated Epidermal Cell Progenitor (foxj1)
Ionocyte Progenitor (foxi1)
Small Segretory Cell Progenitor (foxa1)
B
IVF
foxj1
NT
UMAP 2
UMAP 1
C
foxi1
D
otog
E
tp63
F
IVF gastrula embryo
NT gastrula embryo
animal
ventral
foxj1
foxi1
otogl2
G
inner layer
outer layer
H
% of cells
n.s.
p=0.007
foxj1 IVF
foxj1 NT
otogl2 IVF
otogl2 NT
foxi1 IVF
foxi1 NT
tp63 IVF
tp63 NT

Figure 4

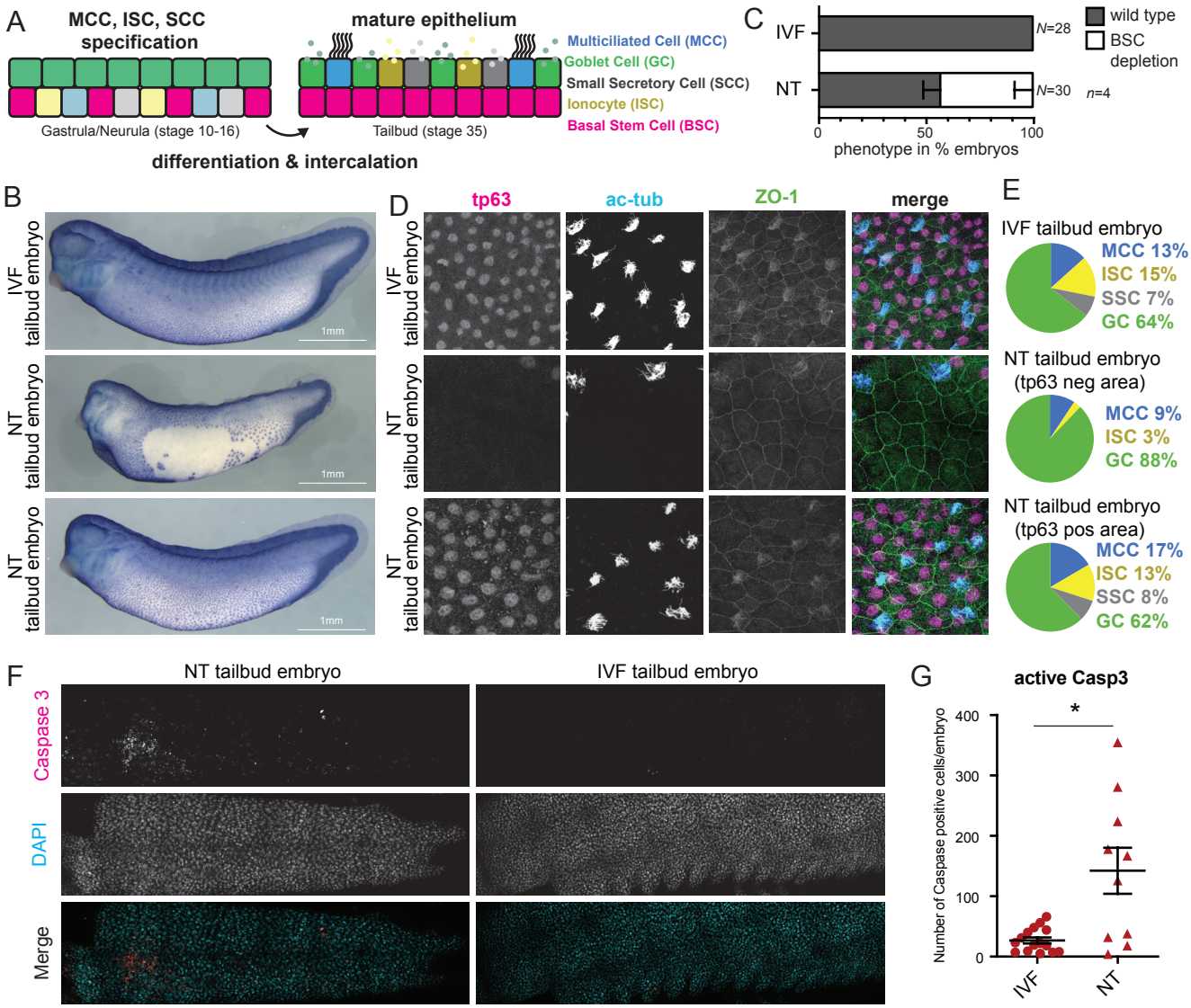

A
MCC, ISC, SCC specification
mature epithelium
Multiciliated Cell (MCC)
Goblet Cell (GC)
Small Secretory Cell (SCC)
Ionocyte (ISC)
Basal Stem Cell (BSC)
Gastrula/Neurula (stage 10-16)
Tailbud (stage 35)
differentiation & intercalation
C
IVF
NT
N=28
N=30
n=4
wild type
BSC depletion
0
50
100
phenotype in % embryos
B
IVF tailbud embryo
NT tailbud embryo
NT tailbud embryo
1mm
D
tp63
ac-tub
ZO-1
merge
E
IVF tailbud embryo
MCC 13%
ISC 15%
SSC 7%
GC 64%
NT tailbud embryo (tp63 neg area)
MCC 9%
ISC 3%
GC 88%
NT tailbud embryo (tp63 pos area)
MCC 17%
ISC 13%
SSC 8%
GC 62%
F
NT tailbud embryo
IVF tailbud embryo
Caspase 3
DAPI
Merge
G
active Casp3
Number of Caspase positive cells/embryo
400
300
200
100
0
*
IVF
NT

Figure 5

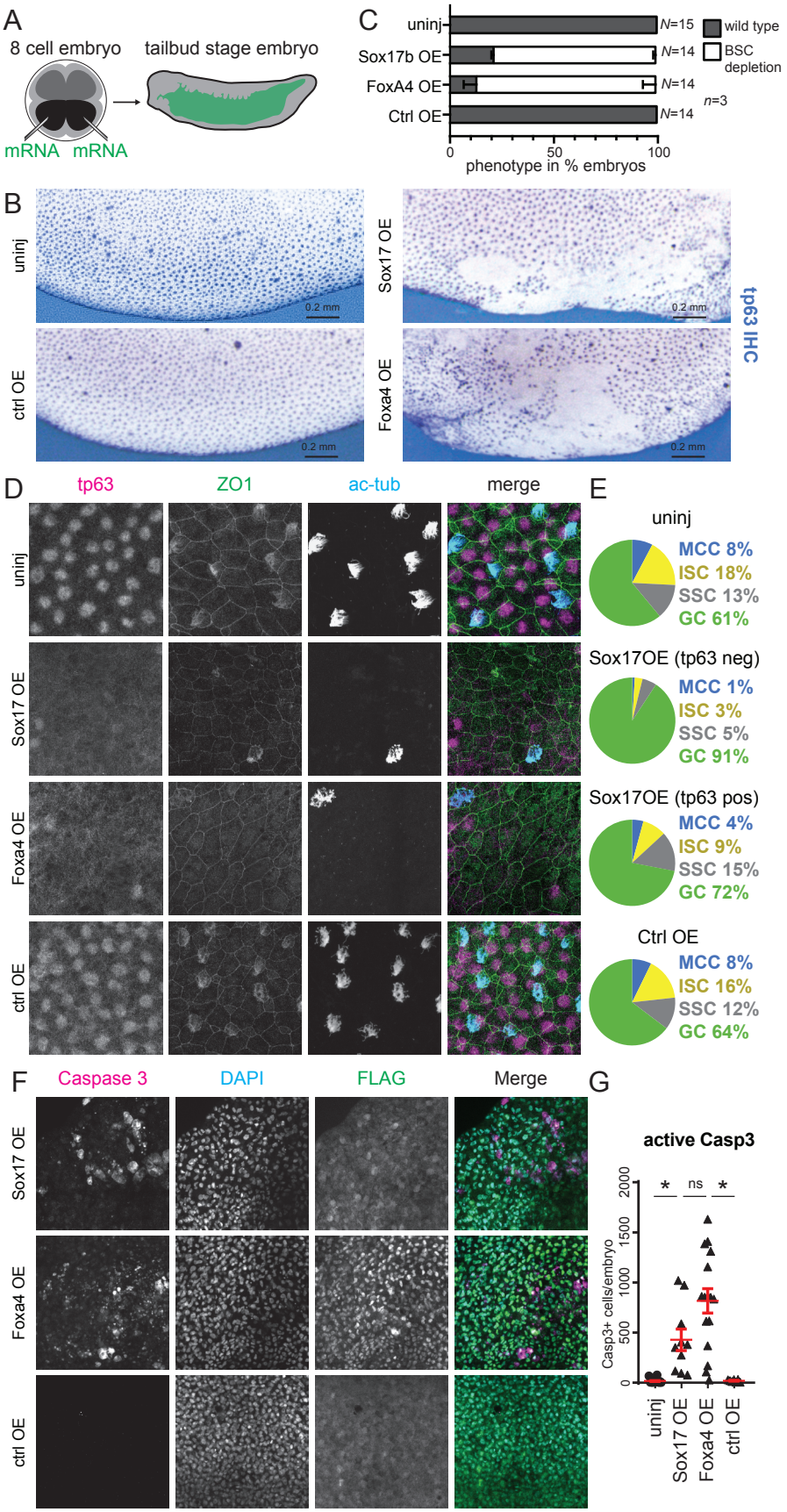
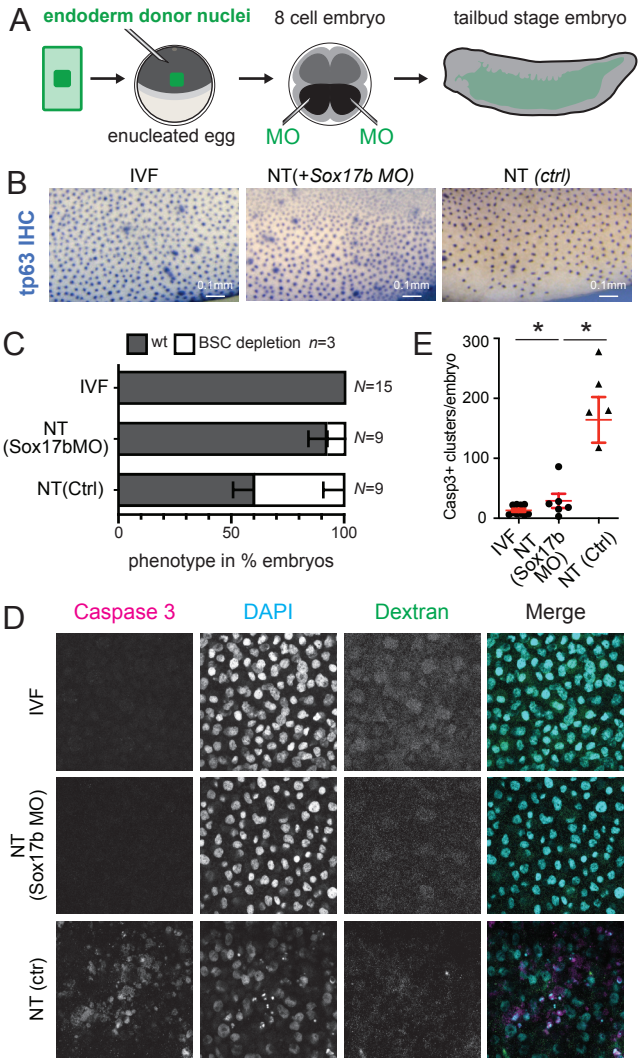
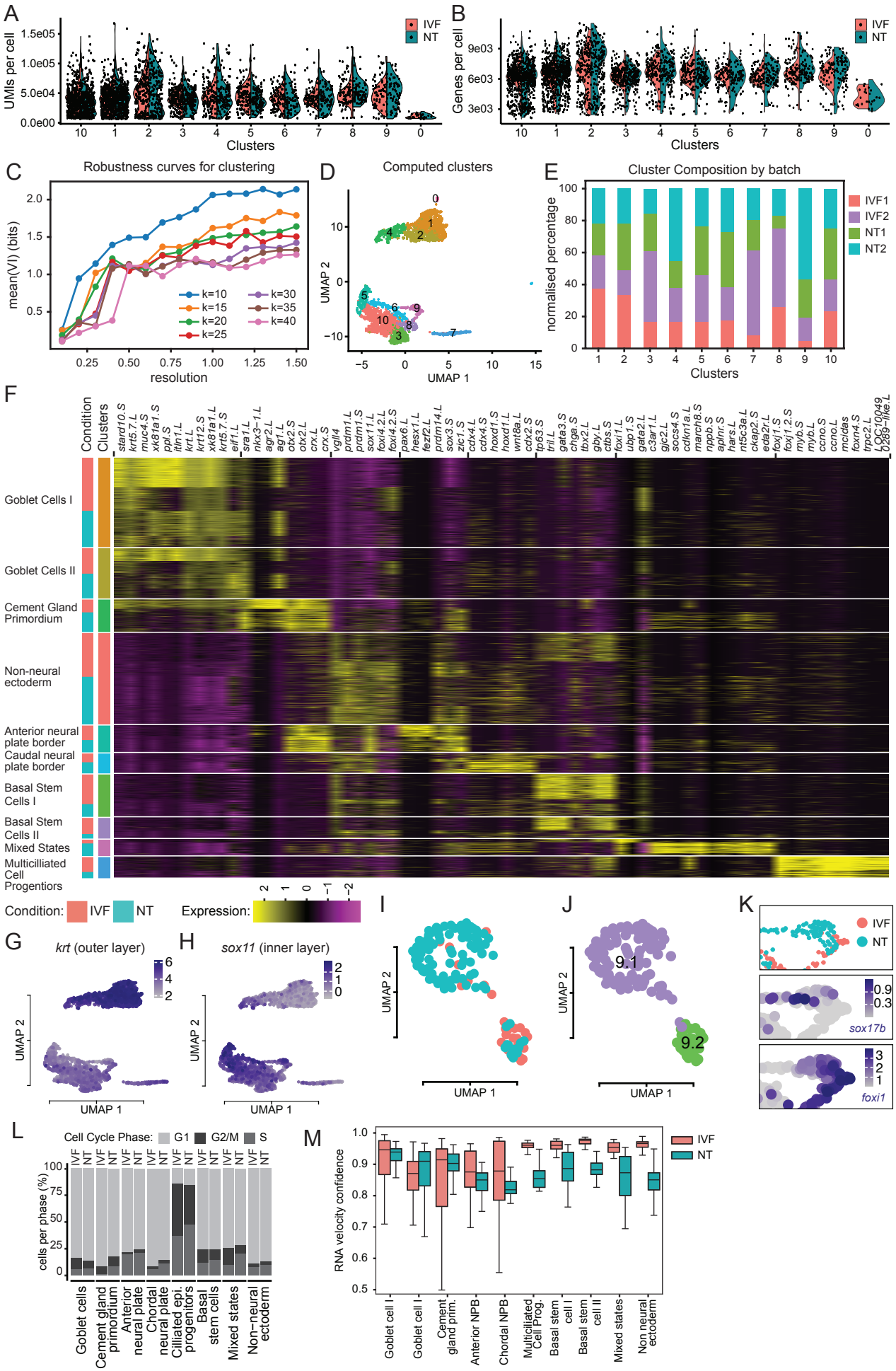

A
8 cell embryo
tailbud stage embryo
mRNA
mRNA
C
uninj
Sox17b OE
FoxA4 OE
Ctrl OE
N=15
N=14
N=14
N=14
wild type
BSC depletion
n=3
0
50
100
phenotype in % embryos
B
uninj
Sox17 OE
ctrl OE
Foxa4 OE
0.2 mm
tp63 IHC
D
tp63
ZO1
ac-tub
merge
uninj
Sox17 OE
Foxa4 OE
ctrl OE
E
uninj
MCC 8%
ISC 18%
SSC 13%
GC 61%
Sox17OE (tp63 neg)
MCC 1%
ISC 3%
SSC 5%
GC 91%
Sox17OE (tp63 pos)
MCC 4%
ISC 9%
SSC 15%
GC 72%
Ctrl OE
MCC 8%
ISC 16%
SSC 12%
GC 64%
F
Caspase 3
DAPI
FLAG
Merge
Sox17 OE
Foxa4 OE
ctrl OE
G
active Casp3
Casp3+ cells/embryo
0
500
1000
1500
2000
*
ns
*
uninj
Sox17 OE
Foxa4 OE
ctrl OE

Figure 6


A
endoderm donor nuclei
8 cell embryo
tailbud stage embryo
enucleated egg
MO
MO
B
IVF
NT(+Sox17b MO)
NT (ctrl)
tp63 IHC
0.1mm
C
wt
BSC depletion n=3
IVF
NT (Sox17bMO)
NT(Ctrl)
N=15
N=9
N=9
0
50
100
phenotype in % embryos
E
Casp3+ clusters/embryo
300
200
100
0
IVF
NT (Sox17b MO)
NT (Ctrl)
D
Caspase 3
DAPI
Dextran
Merge
IVF
NT (Sox17b MO)
NT (ctr)

Supplementary Figure 1