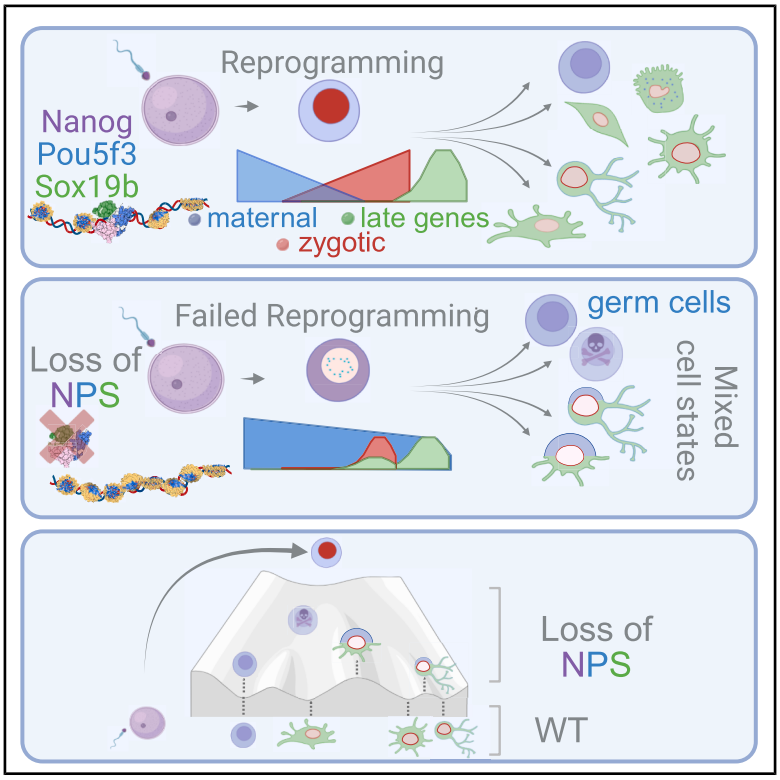
Cell Reports

Article

In vivo differentiation of embryonic cells devoid of key reprogramming factors

Graphical abstract



Highlights

- Cells lacking NPS can survive in a wild-type environment, albeit many cells die
- Cells lacking NPS are impaired in differentiation capacity
- NPS regulate cell identity through coordinating nuclear and cytoplasmic reprogramming
- Some NPS mutant cells can acquire states resembling germ and neural lineages

Authors

Scott E. Youlten, Liyun Miao, Caroline Hoppe, ..., Smita Krishnaswamy, Antonio J. Giraldez, Valerie A. Tornini

Correspondence

antonio.giraldez@yale.edu (A.J.G.), vtornini@ucla.edu (V.A.T.)

In brief

Youlten, Miao et al. use zebrafish transplantation to understand the fate of cells lacking pluripotency factors Nanog, Pou5f3, and Sox19b (NPS) in developmental reprogramming and cell identity establishment. They found that NPS are required for proper cell identity for most, but not all, lineages, likely through nuclear and cytoplasmic reprogramming.

Youlten et al., 2025, Cell Reports *44*, 116498
November 25, 2025
https://doi.org/10.1016/j.celrep.2025.116498

CellPress

Cell Reports

CellPress
OPEN ACCESS

Article

In vivo differentiation of embryonic cells devoid of key reprogramming factors

Scott E. Youlten,[1,8] Liyun Miao,[1,7,8] Caroline Hoppe,[1] Curtis W. Boswell,[1] Damir Musaev,[1] Mario Abdelmessih,[1] Damilola Olowookere,[1] Linnea A. Weiss,[1] Smita Krishnaswamy,[1,2] Antonio J. Giraldez,[1,3,4,*] and Valerie A. Tornini[1,5,6,9,*]

[1]Department of Genetics, Yale University School of Medicine, New Haven, CT 06510, USA
[2]Department of Computer Science, Yale University, New Haven, CT 06510, USA
[3]Yale Stem Cell Center, Yale University School of Medicine, New Haven, CT 06510, USA
[4]Yale Cancer Center, Yale University School of Medicine, New Haven, CT 06510, USA
[5]Department of Integrative Biology and Physiology, University of California, Los Angeles (UCLA), Los Angeles, CA 90095, USA
[6]Institute for Society and Genetics, University of California, Los Angeles (UCLA), Los Angeles, CA 90095, USA
[7]Present address: Department of Biological Science, Florida State University, Tallahassee, FL 32306, USA
[8]These authors contributed equally
[9]Lead contact
*Correspondence: antonio.giraldez@yale.edu (A.J.G.), vtornini@ucla.edu (V.A.T.)
https://doi.org/10.1016/j.celrep.2025.116498

SUMMARY

Embryonic cell differentiation depends on reprogramming of the oocyte and sperm nucleus into a transient totipotent state. In zebrafish, this coincides with genome activation, which is regulated by the pioneer factors Nanog, Pou5f3, and Sox19b (NPS). Here, we investigate the role of NPS in developmental reprogramming and differentiation by analyzing the fate of NPS mutant cells in a wild-type embryo using single-cell RNA-seq. We find that many cells fail to activate transcription or undergo cell death, while others acquire gene expression profiles that resemble germ cells, neural progenitors, and motoneuron states. These cells achieve intermediate transcriptional states, revealing the essential role of NPS in coordinating nuclear and cytoplasmic reprogramming and preventing the premature activation of lineage-specific differentiation programs. These results demonstrate that most developmental programs require developmental reprogramming by NPS, yet some cells can bypass transient totipotency to achieve intermediate developmental states resembling wild-type states *in vivo*.

INTRODUCTION

Fertilization initiates a large-scale remodeling of the nucleus and the cytoplasm during maternal-to-zygotic transition (MZT). This is a conserved developmental event in which maternal gene products deposited during oogenesis initiate the activation of the zygotic genome and trigger the clearance of maternal RNAs, enabling the onset of embryonic development.[1] During this transition, chromatin remodeling, transcriptional activation, and posttranscriptional regulation collaborate to shift control from maternal to zygotic programs. This transition resembles a reprogramming event, conceptually linked to the pioneering work demonstrating that the egg cytoplasm can reprogram a somatic nucleus back to a totipotent state.[2] These nuclear transfer experiments established that maternal factors are sufficient to erase and reset gene expression programs, highlighting the intrinsic reprogramming capacity of the oocyte. However, the specific factors and mechanisms by which this reprogramming is executed remain only partially understood. These observations also suggest that MZT provides the physiological context in which the nucleus and cytoplasm are reset to a transient totipotent state competent for differentiation. Still, it remains unclear whether this transient acquisition of developmental plasticity is strictly necessary for lineage specification. Can cells that fail to undergo nuclear reprogramming still differentiate, and can extrinsic signals compensate for the loss of this developmental reprogramming?

Somatic cells can be reprogrammed into IPS cells by the transcription factors (TFs) NANOG, POU5f1 (formerly Oct4), and SOX2, which drive epigenetic reprograming in development.[3,4] In zebrafish, Nanog, Pou5f3, and Sox19b (NPS) are pioneer factors that mediate chromatin opening and coordinate the first waves of gene expression during zygotic genome activation (ZGA).[5–8] They also activate the expression of the conserved microRNA *miR-430*, which represses and degrades many maternally deposited transcripts to facilitate the transition to the zygotic program.[1,9,10] To study their function during embryogenesis, we previously generated a triple maternal-zygotic *nanog*−/−; *pou5f3*−/−; *sox19b*−/− mutant zebrafish (MZ*nps*) devoid of all three pioneer factors.[11] MZ*nps* embryos display defects in chromatin accessibility, fail to activate a fraction of the zygotic genome, and fail to clear maternal mRNAs, resulting in developmental arrest before gastrulation at 4 h postfertilization (hpf).[7,11] It has been hypothesized that the molecular effects of chromatin

opening and maternal mRNA clearance, initiated by these pioneer factors, lead to a transient totipotent state required for subsequent cell differentiation following gastrulation.[1] However, because MZ*nps* embryos fail to gastrulate, it has not been possible to assess their differentiation capacity *in vivo*.

In this study, we ask whether embryonic cells that fail to reprogram via NPS factors retain any capacity to differentiate. Do they arrest development due to the failure of ZGA, or can they give rise to cell types resembling those in wild-type (WT) embryos? Alternatively, does the lack of NPS lead to the emergence of novel or aberrant cell states? To address these questions, we transplanted MZ*nps* cells into a WT embryonic environment to investigate their developmental potential. This system allowed us to test the nuclear and cytoplasmic reprogramming hypothesis *in vivo* and ask whether distinct cellular states can emerge despite disrupted reprogramming.

RESULTS

Single-cell transcriptome analysis of WT and MZ*nps* mutant cells

Embryos lacking Nanog, Pou5f3, and Sox19b (MZ*nps*) fail to properly activate a significant subset (37%) of zygotic genes, arrest at 4 hpf, and fail to gastrulate, precluding the analysis of their developmental potential beyond gastrulation.[7,11] To understand whether MZ*nps* mutant cells can survive and differentiate into different cell types, we transplanted mutant (MZ*nps*) or WT donor cells into GFP$^+$ WT hosts at the blastula stage (~3.3 hpf), when cells are pluripotent.[12] Donor cells were then allowed to develop and imaged at 12 hpf when the major cell lineages have been specified (illustrated in Figure 1A). We found that MZ*nps* donor cells integrated into WT hosts, albeit at a lower frequency than WT donor cells (Figure 1B). Indeed, consistent with the critical function of NPS in early development,[7,11,13,14] MZ*nps* mutant cells undergo cell death at a higher proportion (~80%) than WT donor cells (<2%; cleaved Caspase-3$^+$ or nuclear DAPI$^-$; Figure S1). Yet, remarkably, ~20% of MZ*nps* donor cells presented a normal morphology, stained positively for nuclear DAPI, and were not labeled by cleaved Caspase-3 (Figures 1C, S1C, and 1D). This indicates that MZ*nps* cells are capable of surviving beyond gastrulation if they develop in a WT embryonic environment, providing the unique opportunity to understand their developmental potential *in vivo* in the absence of NPS-mediated reprogramming.

To investigate the transcriptional state of MZ*nps* cells, we developed a method for rapid enrichment of donor cells relative to host cells for subsequent single-cell RNA sequencing (scRNA-seq; illustrated in Figure 2A).[15,16] Briefly, we expressed a truncated human cell-surface marker CD4 (hCD4) and DsRed in donor WT or MZ*nps* embryos by mRNA injection. At 3.3 hpf, we transplanted these labeled WT or MZ*nps* donor cells into similarly staged WT host embryos. Embryos were allowed to develop, and at the 3–6 somite stage (~11–12 hpf), 50 embryos were dissociated and pooled for each replicate (see methods). We enriched the dissociated cell suspensions for donor cells 2.26-fold with hCD4-binding MicroBeads and magnetic column retention (Figures S2A–S2C). We then performed scRNA-seq (10x Genomics) on donor-enriched cell suspensions (Figure 2A).

We obtained high-quality transcriptome data from 10,551 cells. WT and MZ*nps* donor cells were distinguished from WT hosts based on the enrichment of DsRed and hCD4 marker transcripts and the pattern of single-nucleotide variants (SNVs; see methods; Figures 2B and 2C). In total, we identified 810 MZ*nps* mutant donor cells across two replicates and 3,869 WT donor cells, with the remaining 5,872 cells derived from the WT host (Figures 2D–2F, S2D, and S2E). The lower recovery of MZ*nps* donor cells compared with WT donor cells was consistent with our image analysis and the increased cell death of MZ*nps* mutant cells (Figures 1C and S1). Among a total of 19,782 genes detected ($\geq$1 UMI in >5 cells), 17,775 (89.9%) were detected in both WT and MZ*nps* populations, including 770 zygotic genes (Figure S2F, S2G, and S2H). However, 1,991 genes were only detected in WT cells (Table S1; Figure S2F), consistent with an essential role for NPS in activating a fraction of the genome.[11,17] Together, these results suggest that we can distinguish MZ*nps* and WT donor-derived cells from WT host cells and profile their transcriptomes at single-cell resolution following development *in vivo*.

To compare the transcriptomes of MZ*nps* and WT donor cells, we analyzed the gene expression differences for 6,632 highly variable genes, selected the top 100 principal components, and embedded the data in 2D (see methods). WT donor cells aligned with WT host cells (Figure 2F), indicating their transcriptomes closely resembled those of host tissues. In contrast, MZ*nps* mutant cells largely segregated from both donor WT and host WT cells (Figure 2F). This indicates that the transcriptional state of MZ*nps* cells is distinct from that of WT cells. Profiling the transcriptome of individual MZ*nps* donor cells provided a unique opportunity to determine whether these cells can differentiate *in vivo* without NPS-dependent reprogramming, and, if so, what cell fates they achieve. We reasoned that MZ*nps* cells might follow one of three possible trajectories: (1) arrest their development entirely as occurs in MZ*nps* embryos; (2) collapse into a single, uniform developmental state; or (3) acquire diverse developmental states, indicative of developmental plasticity toward differentiation. Support for the latter would be evidenced by the emergence of reproducible subpopulations of mutant cells with distinct patterns of enriched genes and biological processes.

MZ*nps* cells differentiate into subpopulations with distinct, reproducible expression states

To investigate whether MZ*nps* cells can differentiate *in vivo* into different populations, we first constructed a graph of gene expression similarity among MZ*nps* cells that was used to cluster cells into subpopulations with distinct expression characteristics. We identified five subpopulations that were represented in both mutant replicates (KO 1–5; Figures 3A and S3A).

Enriched genes within each subpopulation included developmental TFs typically expressed in distinct tissues during development (Figures 3C and 3D; Tables S2 and S3). KO 1, the most abundant subpopulation accounting for more than 60% of MZ*nps* cells (n = 550), was associated with the enriched expression of *meis1a*, *znf703*, *cdh1*, *bcar1*, and *anxa11a*. These genes are associated with epithelial morphogenesis, and their expression is enriched in somites and the notochord (Figures 3C, 3D,

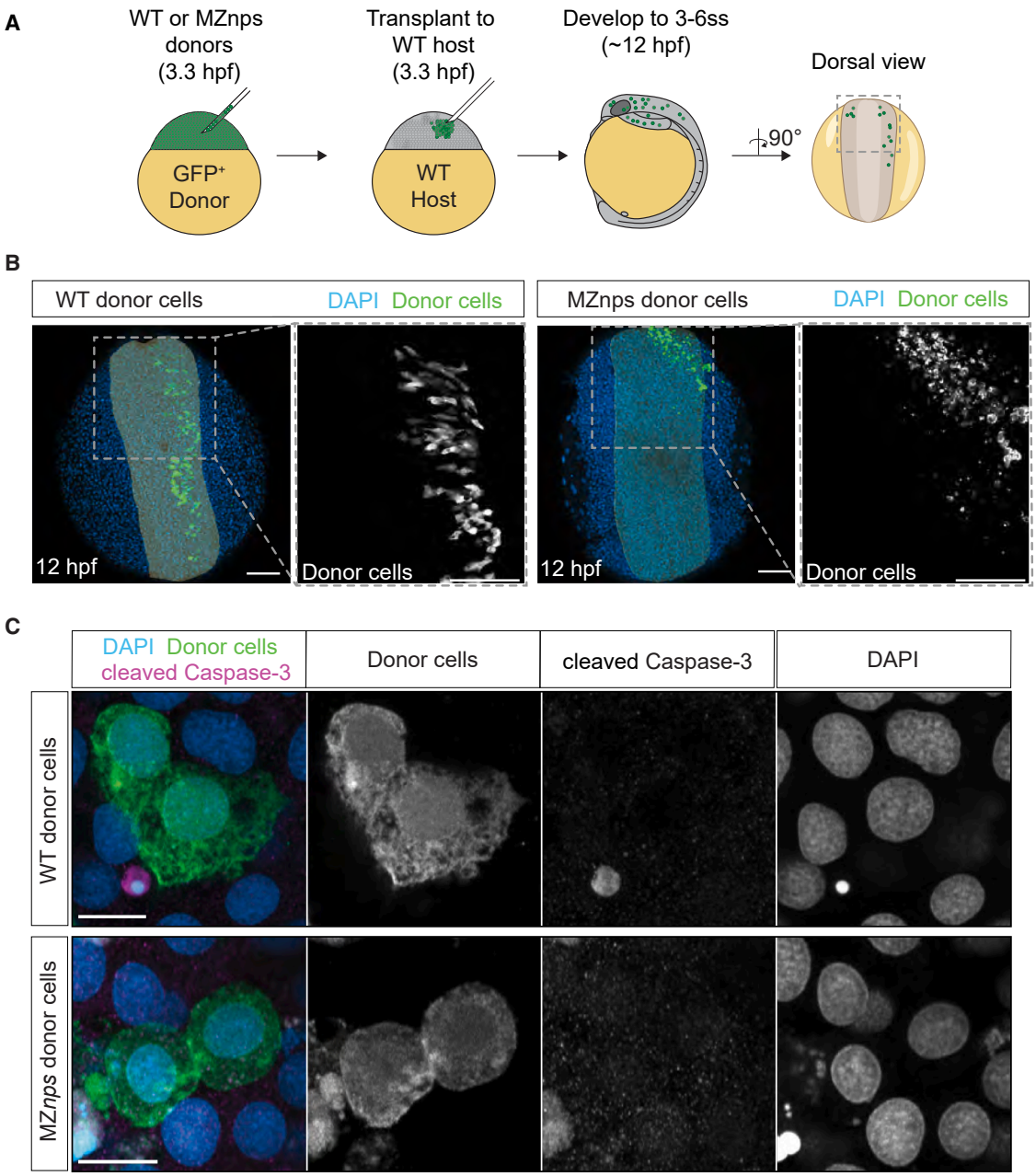


Figure 1. Embryonic cells lacking pioneer factors Nanog, Pou5f3, and Sox19b can survive in a wild-type embryo

(A) Illustration of zebrafish blastula transplants. Wild-type (WT) or MZ*nps* donor embryos were injected with GFP mRNA. At 3.3 hpf, WT or MZ*nps* GFP donor cells were transplanted into unlabeled WT hosts and imaged at the 3–6 somite stage (ss; ~12 hpf).

(B) Host embryos imaged at 3–6 ss containing GFP+ donor cells (green) from either WT (left) or MZ*nps* mutant embryos (right) as in)A). The region of analysis is indicated by the transparent shading. Nuclei labeled with DAPI (blue). Insets show GFP+ cells (gray). Scale bars, 100 μm.

(C) Representative examples of GFP+ donor cells (green) from WT (top) and MZ*nps* (bottom) showing "normal morphology." Nuclei of WT host and donor cells are labeled with DAPI (blue); cleaved Caspase-3 is shown in pink. Scale bars, 10 μm. See Figure S1.

and S3B).[18,19] These cells showed increased expression of apoptosis-associated genes relative to WT cells (95% above mean WT expression levels), suggesting that these cells may have reached a dead-end due to the lack of developmental reprogramming in the absence of NPS (Figures 3B, S3C, and S3D).

KO 2 (n = 132) was characterized by a major failure in transcriptional activation, with 2,597 genes downregulated relative to other subpopulations. This included 858 genes (33%) that are co-bound by NPS during ZGA.[11] Among the 88 upregulated genes, 56 belonged to the histone cluster 1 H2A family located

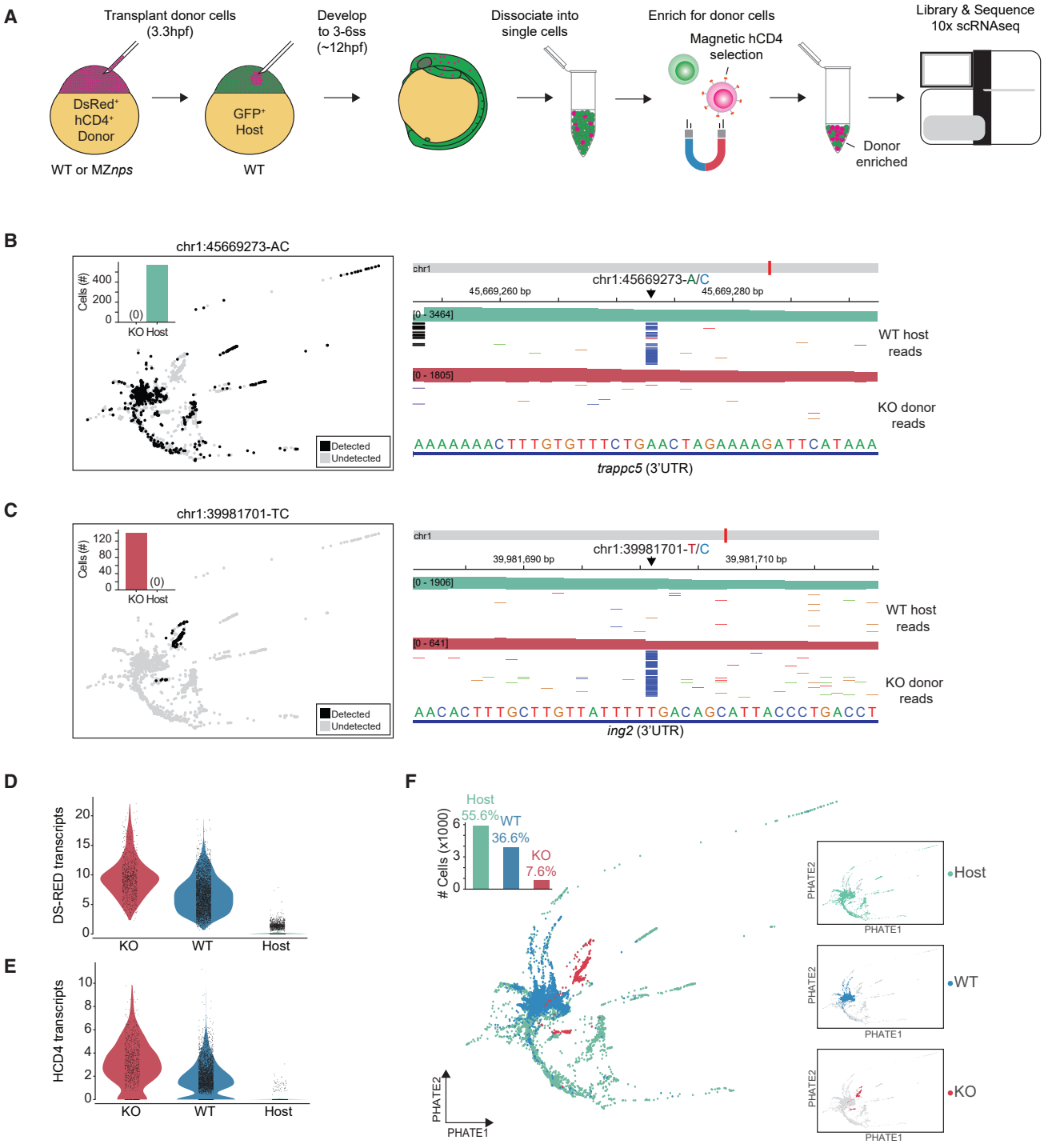


Figure 2. Enrichment and identification of MZ*nps* donor cells enables *in vivo* profiling of cell states

(A) Approach used to enrich cell suspensions for donor cells from dissociated host embryos (see methods). Single-cell sequencing was performed on the donor-cell enriched suspension.

(B). Detection per cell (left) and read-data alignment (right) for an A to C mutation in the 3′UTR of *trappc5* (chr1:45669273), present in reads from WT host cells (top) and not MZ*nps* donor cells (bottom). Bar plots, number of cells in which the SNV was detected per condition. Black, cells with the detected variant in a 2D PHATE embedding of all sequenced cells.

(C) Detection per cell (left) and read-data alignment (right) for a T to C mutation in the 3′UTR of *ing2* (chr1: 39981701), only detected in reads from MZ*nps* donor cells (bottom). Format of graph same as in (B).

(legend continued on next page)

on chromosomes 7 and 25 (Figures 3C, 3D, and S3B). These results are consistent with a significant defect in genome activation in this cell population.

KO 3 (n = 93 cells) was characterized by neuroectodermal genes and was associated with GO terms such as the generation of neurons (OR 26.7, FDR 7e-13) and nervous system development (OR 14.5, FDR 1e-07; Figures 3B–3D; Tables S2 and S3). Upregulated genes included multiple genes from the orthodenticle homeobox family (*otx1*, *otx2a*, and *otx2b*),[20,21] paired box genes (*pax3a* and *pax6a*), and SoxB1 orthologs to *sox19b* (*sox2*, *sox19a*, and *sox21a*) that are critical to establish and maintain neural progenitors.[22] Expression of SoxB1 orthologs might function as a compensatory response to Sox19b loss-of-function in NPS triple mutants.[23–25] Wnt signaling was the molecular pathway most significantly over-represented among KO 3 markers (Figure S3B).

KO 4 had GO term enrichment for neurogenesis-related terms, which include the genes ELAV-like neuron-specific RNA-binding protein *elavl3*, hairy-related 13 (*her13*), and scratch 2 (*scrt2*; Figures 3B–3D). We also observed expression of the ligands and receptors of the Notch signaling pathway (*dla*, *dlb*, *dlc*, *dld*, *and notch1a*), which is involved in motoneuron differentiation (Figures 3C and S3B; Tables S2 and S3)[26–28]

Finally, KO 5 (n = 11) was enriched for *nanos3*, *ddx4 (vasa)*, and *dnd1*, all genes that are maternally loaded and play critical roles in primordial germ cell (PGC) development (Figures 3B–3D).[29–31] Together, these data suggest that MZ*nps* cells in a WT environment are not a homogenous population; instead, they activate different gene expression programs leading to diverse developmental states.

MZ*nps* expression states resemble neural progenitors, neurons, and germ cells

We next asked whether MZ*nps* cells are directed toward WT cellular states or activate distinct gene expression programs and cellular states. To investigate these possibilities, we first analyzed whether MZ*nps* cells align transcriptionally with WT cells (Figure 4 and S4). Briefly, we used publicly available single-cell data to assemble a WT reference of ~120,000 cells encompassing 13 time points spanning 4–24 hpf (Figure 4C).[32–34] We then determined the Euclidean distance between cells using the top 100 principal components of highly variable genes for mutant cells, WT cells, and the reference dataset. We identified the cells in the reference that were most similar to WT or MZ*nps* cells in our dataset, known as "nearest neighbors." This approach allowed us to (1) annotate the cell types for WT host and donor cells (Figures 4A and 4B; see methods), (2) identify the cell types most similar to MZ*nps* subpopulations over embryonic development (Figures 4C–4F and S4C), and (3) define the chronological state of mutant cells based on the developmental stage of the nearest neighbors (Figures 4C, S4A, and S4B).

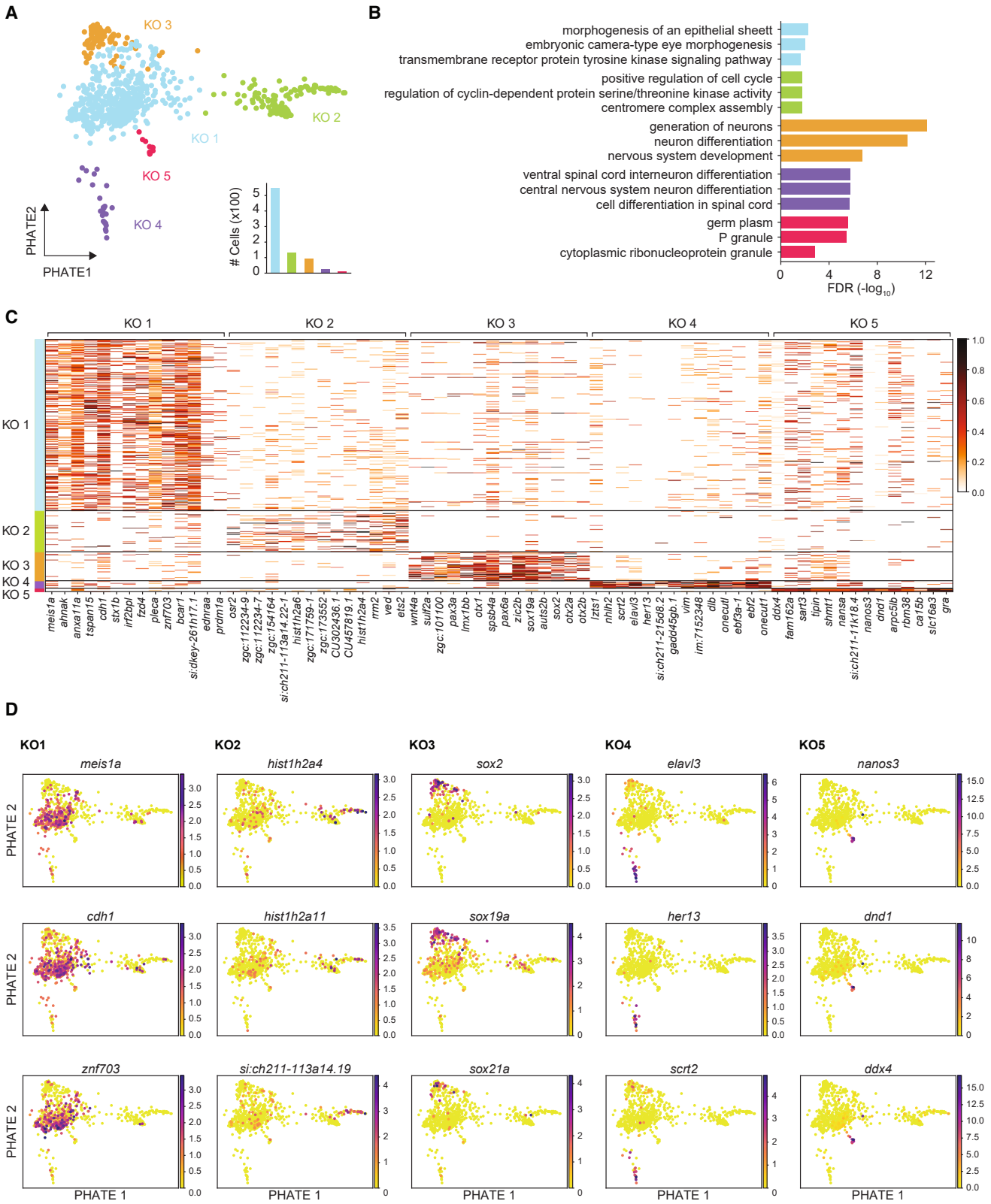
Because the developmental fate of cells is influenced by the local morphogen signals and their position within the embryo, we compared cell-type abundance from WT hosts vs. WT donors. We observed an enrichment for ectodermal lineages and a depletion of mesodermal lineages among WT donor cells (Figures 4B and S4A). Similarly, the nearest-neighbors to the MZ*nps* cells were also enriched in neuroectodermal lineages, suggesting that the transplantation method for WT and MZ*nps* donor cells was similar (Figures 4C and 4E). This analysis also indicates that the differentiation state of MZ*nps* cells was influenced by the host environment. Indeed, except for KO 2, MZ*nps* mutant cells expressed receptors for secreted signaling factors (BMP, FGF, Nodal, Hh, Wnt, and retinoic acid) and their target genes (Figure S4D). These results suggest that most MZ*nps* cells can respond to extrinsic developmental signals in a WT embryo.

The nearest neighbors for KO 1 and KO 2 in the reference dataset were enriched for embryonic stem cells (ESCs) and the ectoderm, which were predominantly observed in earlier developmental time points (Figures 4E, 4G, S4B, and S4C). Gene expression in KO 3 cells was enriched in nearest neighbors for neural tissues such as the developing mid-hindbrain boundary (MHB), forebrain, early neural progenitors and neural crest cells and was correlated with MHB cells (r = 0.60) (Figures 4E, 4G, S4B, and S4C). Gene expression in KO 4 was strongly correlated with both motoneurons (r = 0.74) and interneurons (r = 0.71), which, combined, represented ~86% of its nearest neighbors (Figures 4E, 4G, S4B, and S4C). Consistent with this correlation, the nearest neighbors to KO 4 were predominantly observed in later developmental stages (Figures 4D and S4B). Finally, all the nearest neighbors to KO 5 were PGCs and had a similar developmental stage (Figures 4E, 4G, S4B, and S4C). Gene expression in KO5 cells was highly similar to that of PGCs, with a correlation coefficient of 0.96—higher than the correlation observed between WT cells and their reference cell types (r = 0.92). Consistent with these observations, the genes most specific to subpopulations KO 3–5 were also enriched in neural tissues (KO 3), motoneurons/interneurons (KO 4), and PGCs (KO 5; Figure 4F). The tissue-specific marker genes for these tissues were also significantly over-represented in each KO subpopulation (Figure 4G). Together, these results reveal that in the absence of the reprogramming factors NPS, many cells stall in development, yet a subset of mutant cells can activate distinct transcriptional programs that resemble those of WT cells, including neural progenitors, interneurons/motoneurons, and PGCs.

Loss of NPS lead to partial WT-like states with hallmarks of defective nuclear and cytoplasmic reprogramming

We next examined which gene programs defined the similarities between KO subpopulations and specific WT states. KO 3, KO 4, and KO 5 each shared sets of genes with their closest WT neighbors—namely, neural progenitors, motoneurons/interneurons, and PGCs, respectively (Figure 5A). However, each KO group also lacked several key marker genes found in these WT states (Figures 5A–5D). We reasoned that the generation of mutant

(D) Violin plot of normalized DsRed mRNA expression in single cells from MZ*nps* (KO) donor cells (red) or WT donor cells (blue) and WT host cells (green).
(E) Same as (D) for normalized hCD4 mRNA expression.
(F) A 2D PHATE embedding of WT host (green), WT donor (blue), and MZ*nps* (red) cells overlaid and individually (inset). Bar plot, number and relative percentage of cells per condition in the total datasets. See Figure S2; Table S1.



(legend on next page)

cell states in the absence of the pioneer factors NPS could be derived from defective nuclear or cytoplasmic reprogramming after fertilization. Alternatively, mutant cells might be able to activate the gene expression modules (EMs) present in WT cells via mechanisms independent of NPS. To test these hypotheses, we first identified gene EMs based on the maternal or zygotic expression and their dependency on NPS in WT and MZ*nps* embryos between 2.5 and 6 hpf. This analysis revealed 9 EMs with distinct expression, chromatin accessibility, NPS binding, and *miR-430* site enrichment dynamics (Figures 6A–6F and S5; Table S5). These EMs represent distinct combinations of genome regulation dynamics, including NPS-dependent degradation (EMs 1 and 2), NPS-independent degradation (EM3), NPS-dependent activation (EMs 4–7), and NPS-dependent repression (EMs 8 and 9; Figures 6A–6F).

Next, we analyzed gene expression across the different KO cell clusters and identified four major patterns of gene regulation in the absence of NPS (Figures 5E, 6H, and S6). First, we observed a failure in cytoplasmic reprogramming, characterized by the upregulation of *miR-430* target genes due to impaired *miR-430*-mediated mRNA clearance (Figure 5E). This pattern was present across all KO subpopulations but was especially prominent in KO 1. Representative genes include *hoxb2a*, *hoxa2b*, *hif1a*, *meis1a*, *myocb1* (KO 1; Figure S6A); *sod1*, *nr2f1b*, and *gstp1* (KO 4; Figure 6H); and *hbp1*, *phax*, and *exrb1* (KO 5; Figure S6D; Table S5). These genes generally belong to embryonic module EM1 and exhibit strong maternal persistence, with some showing even higher expression in KO cells compared to WT cells at 12 hpf (Table S1). Their enhancer regions are typically depleted of NPS-binding sites, and the 3′UTRs are enriched for *miR-430* seed sequences (Figures 5E, 6B, and 6C). This suggests that another major source of transcriptional differences in MZnps cells is that they fail to clear many maternally provided transcripts. This pattern results from reduced *miR-430* expression, which depends on NPS for its zygotic activation, leading to maternal mRNA clearance and transcriptional reprogramming of the cytoplasm.

Second, we identified a failure in nuclear reprogramming, marked by the strong downregulation of genes in EMs 4–6, which are normally enriched for NPS binding and show reduced chromatin accessibility at their promoters and enhancers in KO cells (Figures 6E and 6F; Table S5). The genes that were downregulated in KO 3, KO 4, and KO 5 compared to WT showed features of NPS-dependent activation, including the enrichment of NPS-binding sites in their regulatory regions (Figure 6E; Table S4). For example, in KO 4, 65.5% (19/29) of the top downregulated genes were NPS bound (Figure 5C). Genes in these modules include *otx2a* and *foxd3* (KO 3; Figure S6C) and *ddl*, *cxcr4b*, and *elavl3* (KO 4; Figure 6H). Their enhancer regions include NPS binding (Figure 5E), and their reduced expression is consistent with NPS's role in the activation of the zygotic genome, which is necessary for transcriptional reprogramming of the nucleus. This suggests that one major source of transcriptional differences in MZnps cells is their failure to transcribe genes that are normally expressed immediately upon ZGA.

Third, we found a group of genes showing premature zygotic transcription in MZ*nps* mutants. Some of these genes, like *tet3* and *angptl4*, are normally repressed by NPS during nuclear reprogramming, while others—such as *mfap2* and *nat16*—are not bound by NPS and may be activated through compensatory mechanisms involving TFs that compete with NPS during genome activation (Figure 5C; Table S5). These genes are associated with either high (EM8) or low (EM9) NPS binding and show mild enrichment for *miR-430* seed sites, suggesting that their misregulation results from both failed nuclear and cytoplasmic reprogramming (Figure 5E). Notably, many of these genes (*neurod1*, *neurod4*, *nhlh2*, *scrt2*, and *isl1*) are neural regulators that are prematurely activated in MZ*nps* embryos as early as 5 hpf (Figure 6H). These findings suggest that NPS prevents premature expression of lineage-specific genes during early development, either through direct repression or by competing with other TFs for access to the transcriptional machinery, thus supporting a model in which NPS functions as a developmental brake.

Finally, the fourth pattern was specific to KO 5 and featured strong enrichment for PGC markers. Most of these genes (e.g., *ddx4*, *dnd1*, and *nanos3*) are maternally deposited in the germ plasm and showed consistent expression in both WT and MZ*nps* embryos (Figures 5A, 5D, and S5B). This expression pattern aligns with EM2, which is characterized by a lack of NPS binding, indicating that PGC specification occurs largely independently of NPS-dependent reprogramming.

Together, these analyses show that NPS governs three distinct developmental roles: cytoplasmic reprogramming through the clearance of maternal mRNAs, nuclear reprogramming through the direct activation of zygotic genes, and transcriptional competition that prevents premature lineage-specific transcription.

Cell states reflect reprogramming defects and compensatory gene activation

Next, we analyzed how early EM dynamics influence the developmental plasticity of MZ*nps* cells and their transcriptional states later in development at 12 hpf. To this end, we calculated the likelihood of different EMs contributing to KO clusters and WT cells (Figure 6G). Each KO group displayed unique combinations of EMs, some of which were absent in WT tissue states. For example, KO 1 states were represented by EMs 1, 2, 4, 5,

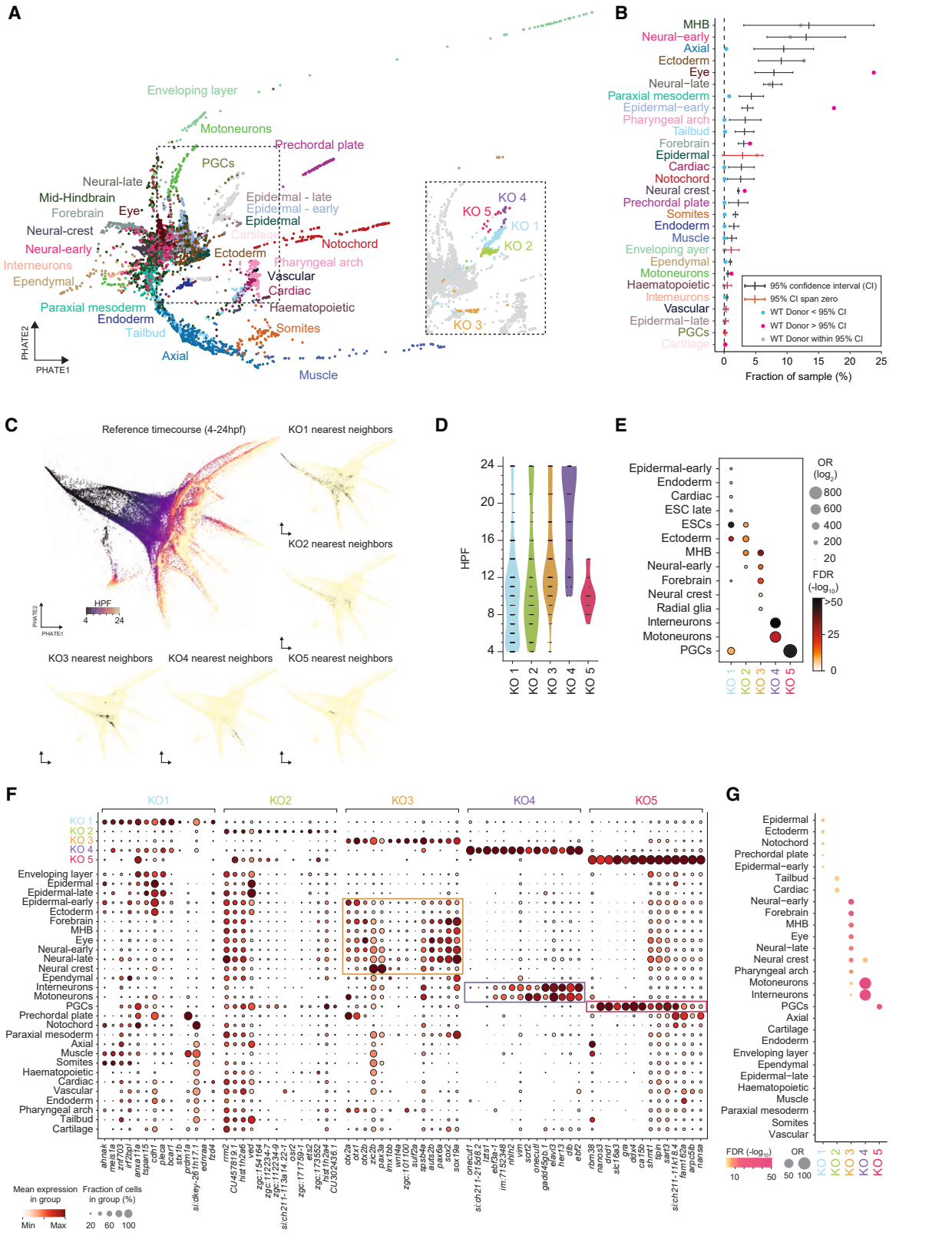
Figure 3. MZ*nps* cells differentiate into distinct expression states *in vivo*

(A) 2D PHATE embedding of MZ*nps* donor cells (810 total) colored by 5 subpopulations (KO 1–5). Bar plot, number of cells (x100) per subpopulation.

(B) The top 3 Gene Ontology terms significantly over-represented (FDR < 0.05) among top marker genes ($n = 100$) of each MZ*nps* subpopulation (indicated by bar color as in (A)).

(C) Heatmap showing the normalized expression in single cells for the top 14 marker genes in each MZ*nps* subpopulation. Rows, cells (sub populations indicated in the color bar as in (A)); columns, genes; color, expression (min-max scaled across cells).

(D) 2D PHATE embedding colored by smoothed expression of select transcription factors identified among markers of MZ*nps* subpopulations. See Figure S3; Tables S2 and S3.



(legend on next page)

8, and 9; this combination did not appear in any WT tissue states at 12 hpf, suggesting that these cell states do not exist in WT and derive from failed developmental reprogramming in MZ*nps* cells. In contrast, KO 4 cells were shaped exclusively by EMs 6 and 9, which also contributed to interneuron and motoneuron states, and KO 5 states were defined by EMs 2, 3, and 4, a combination that was uniquely associated with 12 hpf PGCs (Figure 6G). These results highlight the critical role of early embryonic EMs, shaped by developmental reprogramming, in determining the cell states reached later in development.

Finally, we analyzed how embryonic EMs were shaping differentiated KO cell state gene expression programs observed at 12 hpf (Figures 6H and S6). The strong resemblance between KO 5 and WT PGCs was consistent with the development of a viable germline derived from MZ*nps* PGCs when transplanted into PGC-depleted WT embryos (Miao et al., 2022), suggesting that PGC specification occurs independently of NPS. Thus, we focused on KO 4 due to their high resemblance to motoneurons and interneurons (Figures 4E and 5A), with the goal of understanding how gene regulatory patterns could dictate differentiation in these KO cells in the absence of reprogramming. EM6 (NPS-dependent ZGA; yellow) and EM9 (strong NPS-dependent repression; green) drove the KO 4 markers with the most disparate embryonic patterns (Figure 6H). By 12 hpf, a group of genes were still differentially expressed between KO4 and WT cells, accounting for the differences between the mutant and WT cells (Figure 5C). Upregulated genes in KO 4 were enriched in EM9, while downregulated genes were enriched in EM6 (Figure 6G), suggesting that defects in developmental reprogramming persist into later stages. However, a group of genes from EMs 6 and 9 showed similar levels of detection between KO 4 and WT motoneurons at 12 hpf (Figure 5C). For example, genes from EM6 that are co-bound by NPS during ZGA (*elavl3*, *cxcr4b*, *her13*, and *insm1b*) were not expressed in MZ*nps* embryos at 6 hpf (Figure 6H), but they were still induced in KO 4 cells by 12 hpf (Figure 5C), indicating that early failure in activation can be compensated by nonautonomous signals in the embryo that are independent of NPS reprogramming. On the other hand, genes from EM9 (*neurod4*, *scrt2*, *neurod1*, and *isl1*) were highly expressed in MZ*nps* embryos compared to WT embryos (Figure 6H), and their expression was maintained specifically in KO 4 cells relative to other KO cells (Figure 5A). These results are consistent with a developmental reprogramming role of NPS that activates early developmental programs and also prevents premature activation of later programs regulated by other TFs (Figure 6H). Interestingly, we identified additional genes (*elavl3*, *her13*, *insm1b*, and *cxcr4b*) that failed to activate in early MZ*nps* embryos but were later rescued in KO cells and became activated through NPS-independent mechanisms (Figures 5C and 6H). Altogether, these results suggest that the pioneer factors NPS establish a transient totipotent state by regulating nuclear and cytoplasmic reprogramming and preventing premature activation of differentiation programs. In some differentiation pathways, their absence can be partially compensated by nonautonomous signals from the WT environment.

DISCUSSION

Here, we investigated whether developmental reprogramming by the maternal pioneer factors NPS is required for cell differentiation. Our results reveal three major findings. First, MZ*nps* cells fail to fully undergo nuclear and cytoplasmic reprogramming. Second, despite these defects, some cells adopt transcriptional profiles resembling WT germ cells or neural progenitors, while others enter a dead-end or a mixed state, suggesting limited differentiation potential. Third, we observe premature activation of later developmental gene programs, indicating that reprogramming also serves to constrain the timing of transcriptional activation. Together, these results reveal that NPS are required to mediate nuclear and cytoplasmic reprogramming of the transcriptome and to establish the temporal coordination of gene expression for proper cell differentiation after MZT.

NPS-mediated reprogramming and developmental specification

Cell differentiation is driven by lineage-specific TFs that activate gene expression programs, which are further refined post transcriptionally by microRNAs and RNA-binding proteins.

Figure 4. MZ*nps* achieve expression states resembling germ cells, neural progenitors, and neurons

(A) 2D PHATE embedding of total dataset, colored by WT cell-type annotations (see methods; MZ*nps* cells colored gray). Inset, magnified view of MZ*nps* cells colored by KO subpopulations (WT cells colored gray).

(B) Proportional representation of WT donor cell types compared to host cell types. Host indicated by 95% confidence intervals (CIS) of mean cell-type proportions. Percentage of WT donor cells for each cell type indicated by colored dots. Blue, below 95% CI of host; red, above 95% CI; gray, within 95% CI.

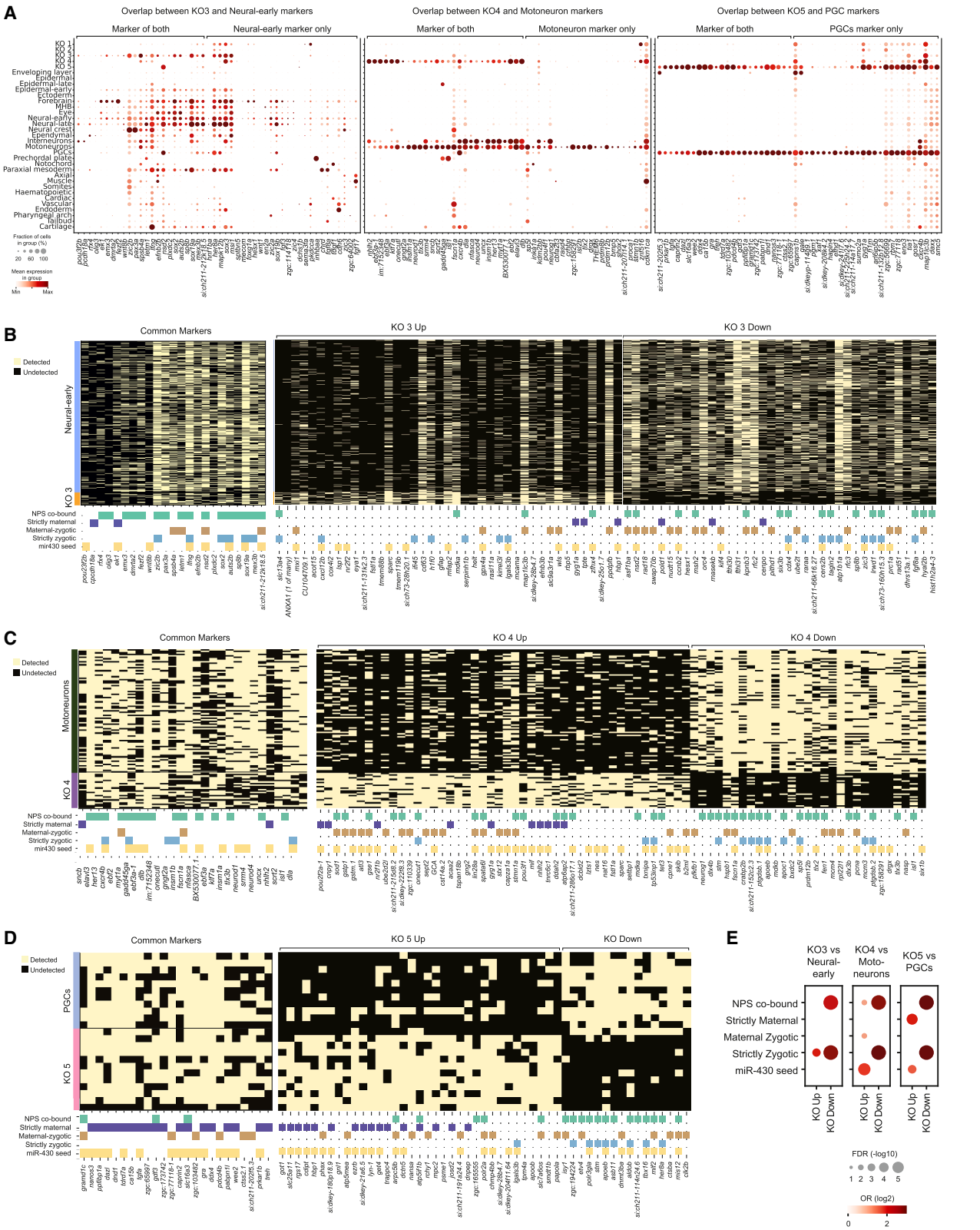
(C) 2D PHATE embedding of WT reference dataset (see methods) spanning 4–24 hpf with cells colored according to time point of collection. Insets indicate in black the WT reference cells that are nearest neighbors (k = 5) to MZ*nps* cells for each subpopulation.

(D) Violin plots showing the hpf distribution of WT reference cells that are nearest neighbors (k = 5) to MZ*nps* cells for each subpopulation as shown in (C). Cells shown in black.

(E) Bubble plot showing significant enrichment (FDR < 0.01) of cell types among WT reference cells that are nearest neighbors (k = 5) to MZ*nps* cells for each subpopulation. The size of the bubble reflects the odds ratio (OR), and color reflects the false discovery rate (FDR; -log10). Over-representation determined by one-sided Fisher's Exact test and corrected for multiple hypothesis testing.

(F) Bubble plot showing the mean expression of (color) and fraction (size) of cells expressing the top marker genes (n = 14) identified for each MZ*nps* subpopulation. Expression is summarized for both MZ*nps* KO subpopulations and WT cell types. Rows correspond to cell types, columns correspond to genes, AND color is scaled by column. Markers for KO 3 are most strongly expressed in neuroectodermal progenitors (yellow box); for KO 4, in interneurons and motoneurons (blue box); and for KO 5, markers in PGCs (red box).

(G) Dot plot showing significant enrichment (FDR < 0.05) of MZ*nps* subpopulation markers among markers of WT cell types (see methods). The size of the dots, odds ratio (OR). Color, false discovery rate (FDR; -log10). Over-representation determined by one-sided Fisher's Exact test and corrected for multiple hypothesis testing. See Figure S4.



(legend on next page)

According to the developmental reprogramming hypothesis, both nuclear and cytoplasmic reprogramming are required during MZT to enable proper cell differentiation.[1,10,37] The extensive reprogramming that occurs during the MZT is thought to influence the landscape of future cell-fate potential. We find that NPS mutant cells can adopt transcriptional states partially resembling WT fates. A subset of MZ*nps* cells display gene expression profiles consistent with early developmental arrest, reflecting a failure to initiate genome activation or the induction of apoptosis. In contrast, a subset of mutant cells acquire gene expression patterns that resemble peripheral and central neural lineages, supporting the idea that in the absence of reprogramming, cells may default toward neuroectodermal fates.[38] While these neural-like cells share a core set of fate-specifying genes with their WT counterparts, they also exhibit misexpressed genes. Upregulated transcripts are enriched in *miR-430* target sites, consistent with failed maternal mRNA clearance and cytoplasmic reprogramming. Interestingly, several upregulated genes—including *neurod1*, *neurod4*, *and islet1*—are normally activated at later stages of development but appear to be prematurely transcribed in mutant cells. This is consistent with a function of NPS in preventing the activation of later expressed genes that are normally driven by other TFs and thus preserving temporal fidelity by means of transcriptional competition.[39–42] Conversely, downregulated genes are enriched in NPS-binding sites, indicating that loss of NPS early on has detrimental effects on the subsequent activation of these lineage genes. Together, these results suggest that in the absence of developmental reprogramming by NPS, cells adopt intermediate states that lack proper temporal and identity resolution.

We propose that pioneer factors such as NPS sculpt epigenetic trajectories that canalize cells toward stable fates. In their absence, cells deviate into hybrid transcriptional states, lacking proper nuclear and cytoplasmic reprogramming. These findings align with models implicating maternal pioneer factors in both the activation of the zygotic genome and the repression of the previous developmental program.[7,9,11,43–45] Notably, comparison of MZ*nps* cells with canonical signaling mutants, such as Nodal receptor-deficient embryos, reveals an important distinction: while signaling mutants retain the flexibility to adopt alternative WT fates, MZ*nps* cells occupy transcriptional territories that are distinct from any WT lineage.[32] We propose that the loss of NPS disrupts the activation of key lineage programs. As a result, failed reprogramming does not simply cause arrest but can redirect cells toward aberrant transcriptional states.

PGC formation independent of NPS activity

We found that PGC states do not require *nanog*, *pou5f3*, and *sox19b* for lineage commitment or developmental reprogramming. Although we attempted germline transplantation[46] from MZ*nps* mutants into WT embryos, we were unable to determine whether these cells could mature into functional germ cells and reconstitute the germ line. This inability may stem from the substantial reduction in *cxcr4a/b* expression in MZ*nps* mutant embryos (Miao et al. 2022), which likely disrupts PGC migration. Indeed, Nanog knockdown reduces *cxcr4* expression and leads to PGC migration defects in medaka.[47] Nonetheless, multiple lines of evidence indicate that PGC specification is not significantly compromised by NPS loss of function. Following early developmental rescue, MZ*nps* mutants are fertile,[11,17] suggesting that NPS are dispensable for germline maintenance during later stages. In support of this conclusion, Nanog knockdown results in an increased number of PGCs in zebrafish,[48] and ectopic PGCs can be induced via a cocktail of nine maternal factors[49] or through germplasm transplantation, as observed in Drosophila.[50] Indeed, one of the earlier functions for *miR-430* in somatic cells is to clear and repress maternal PGC-specific genes in somatic cells.[45] Although KO 5 cells showed some differences from PGCs, their identity is primarily defined by maternal inheritance.[49,51] Our findings suggest that reprogramming by NPS is not required to establish the majority of the PGC gene expression program or specify germ-cell fate.

Developmental context as a modulator of cell fate

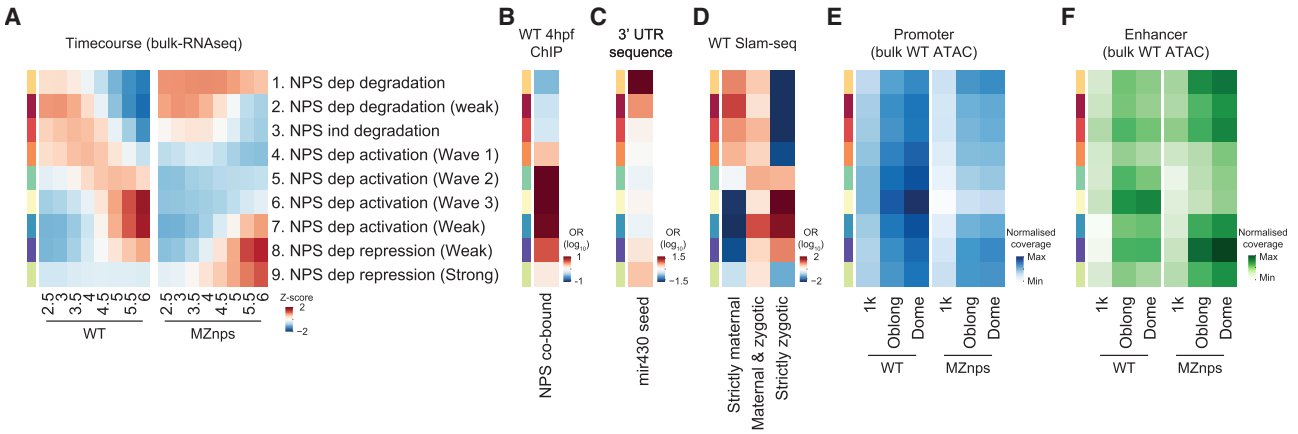
Despite intrinsic deficits, MZ*nps* cells that survived in the WT environment displayed transcriptional plasticity. Multiple genes expressed by mutant cells are reminiscent of developmental programs corresponding to neural progenitors, interneurons, motoneurons, and PGCs, albeit with transcriptional signatures distinct from canonical WT fates. For cell types underrepresented in the WT control, their absence in mutants could reflect either a true differentiation failure or transplantation bias. In neural progenitor cells, we observed transcriptional compensation by SoxB1 family members, reminiscent of neuroectodermal lineages.[23] The expression of SoxB1 family members in these cells may represent a developmental buffer that allows cells to access ectodermally-derived states. Other neural-like cells showed premature and incomplete activation of differentiation genes such as *elavl3* and *neurod4*. Although genes categorized as dependent upon NPS for activation were not turned on at 6 hpf, they were identified as marker genes for KO 4 at 12 hpf, suggesting that developmental signals in a WT environment may play a

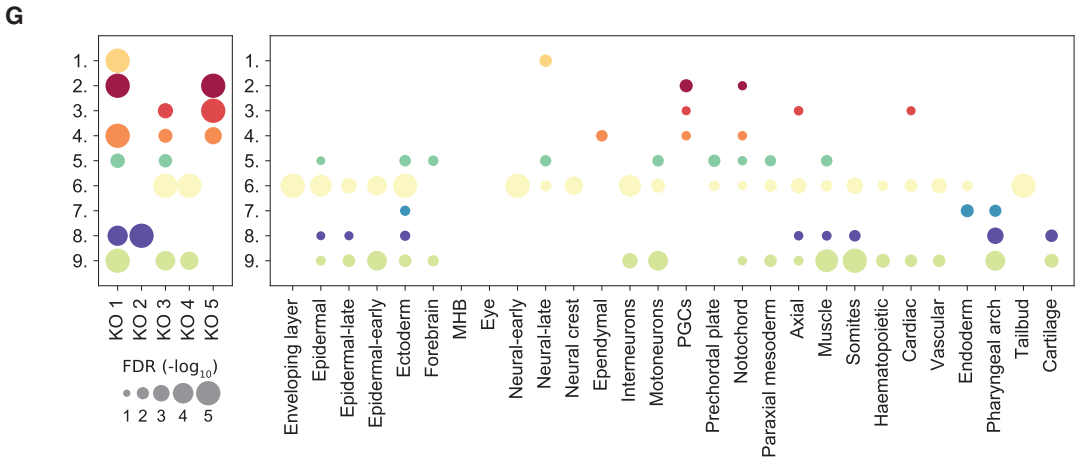
Figure 5. MZ*nps* expression states combine cell type-specific marker expression with altered expression of NPS and *miR-430* targets

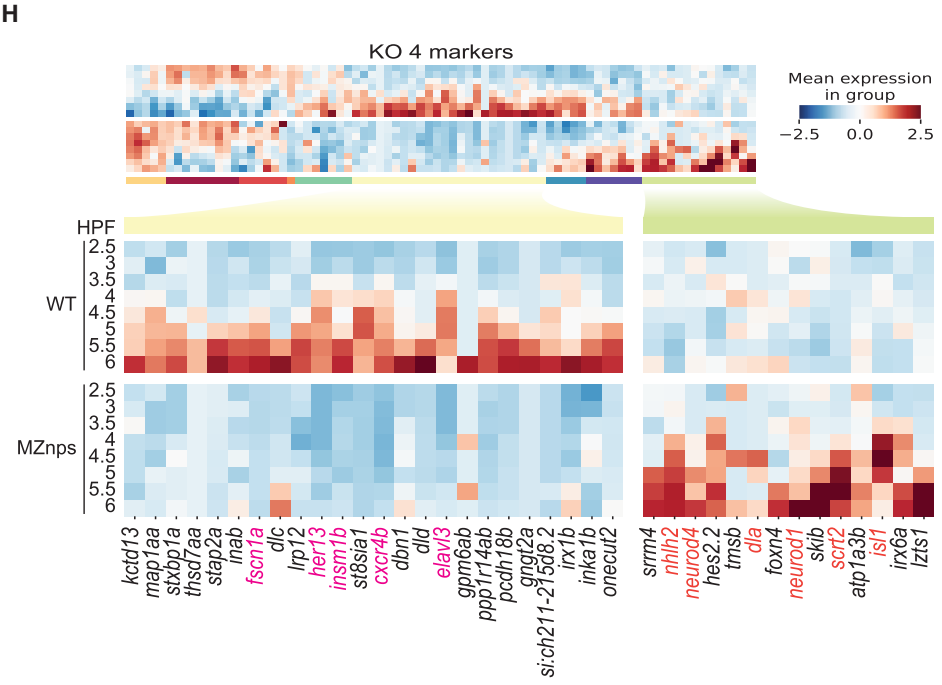
(A) Dot plot showing the overlap of expression of subpopulation-specific marker genes between MZ*nps* subpopulations (KO 3, 4, and 5) and their most closely related tissue types (neural-early, motoneuron, and PGC, respectively). "Marker of both," genes grouped according to MZ*nps* subpopulation-specific marker genes that are also expressed in the related tissue type. "Cell-type marker only," genes for the related tissue type not specific for the relevant MZ*nps* subpopulation. Color, mean expression represented; size, fraction of cells expressing each gene.

(B–D). Heatmaps showing common markers (left, from (A)) and differentially expressed genes (right, columns) between MZ*nps* subpopulations KO 3, KO 4, and KO 5, and their respective closest WT tissue type (see methods). Each row is a cell. Genes organized as statistically "up"- or "down"-regulated across cells in the MZ*nps* subpopulation relative to the closest tissue type; detected genes in yellow (≥1UMI). Genes are annotated (color bar below) for function at MZT (4 hpf): NPS co-bound, strictly maternal, maternal zygotic, strictly zygotic, and *miR-430* seed (see methods).

(E) Bubble plot of MZT annotation over-representation among differentially expressed genes in B-D (FDR <0.05). Dot size, false discovery rate (FDR; -log10); color, odds ratio (OR; log2). Over-representation determined by one-sided Fisher's exact test and corrected for multiple hypothesis testing. See Table S4.







(legend on next page)

role in the differentiation of these cell states. Thus, the activation of genes characteristic of central and peripheral neural fates in these cells may originate from a cell's intrinsic competence to receive extracellular signals that induce neuronal expression programs during development. However, due to the absence of other factors required for further lineage commitment, including NPS developmental reprogramming, these cells may then be pushed toward a hybrid expression state.

In summary, our findings show that NPS function during early development to (1) reprogram the cytoplasm by eliminating maternal transcripts, (2) reprogram the nucleus by activating the zygotic genome, and (3) inhibit premature transcriptional programs required later in development. Through this integrated activity, NPS ensures the coordination of gene expression programs across different cell-fate trajectories.

Limitations of the study

While the transplantation analysis allows us to assess different cell fates, given the limitation of the transplant, we have not been able to explore all potential cell types. Future studies will be needed to understand how initial chromatin remodeling following genome activation creates a permissive chromatin environment for downstream developmental TFs to activate specific gene expression programs during embryogenesis. Finally, while some MZ*nps* cells (KO 5) resemble PGCs, their transcriptional states are not identical. Further studies are required to determine whether MZ*nps* mutant PGCs are functional.

RESOURCE AVAILABILITY

Lead contact

Requests for reagents and information should be directed to, and will be fulfilled by, the lead contact, Valerie Tornini (vtornini@ucla.edu).

Materials availability

All constructs generated in this study are available from the lead contact upon request.

Data and code availability

Raw sequencing data have been deposited under SRA BioProject ID PRJNA1126192. Processed sequencing data are deposited under GEO accession ID GSE309740. Original code is available at https://github.com/GiraldezLab/scMZnps.

ACKNOWLEDGMENTS

We thank reviewers for their constructive comments. We thank Guilin Wang from the Yale Center for Genome Analysis for sequencing support; Hiba Codore and Nitya Khatri for technical help; Stefania Nicoli for sharing their lab space; and Sarah Dube and Tim Gerson for excellent animal care. This work is supported by funding from the US NIH (R00HD105001) and the UCLA Division of Life Sciences startup funds to V.A.T; the Surdna Foundation and the Yale Genetics Venture Fund to L.M.; Human Frontiers Postdoctoral Fellowship LT0073/2022-L and EMBO Long-Term Postdoctoral Fellowship ALTF #794-2021 to C.H.; a Canadian Institutes of Health Research (CIHR) postdoctoral fellowship to C.W.B.; and R01HD100035 (NIH) to S.K. and A.J.G. and R35GM122580 (NIH) to A.J.G. The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health or any funding sources.

AUTHOR CONTRIBUTIONS

L.M., V.A.T., and A.J.G. conceived this project. L.M., V.A.T., and A.J.G. designed and performed experiments. S.E.Y. designed and performed data analyses, with L.M., V.A.T., and A.J.G. contributing to guiding analyses and performing *in vivo* validations. C.H. performed confocal image acquisition and analyses. C.W.B. generated a custom annotation of cell types in previously published wild-type single-cell datasets (from 32). D.M. and M.A. identified an endogenous stabilizing 3′UTR. D.O. contributed to the experiments. L.A.W., S.K., V.A.T., and A.J.G. contributed to funding acquisition and project supervision. V.A.T. and S.E.Y. wrote the original draft, with L.M., L.A.W., and A.J.G. contributing to editing and revisions. All authors revised this manuscript. S.E.Y. and L.M. contributed equally to this work.

DECLARATION OF INTERESTS

A.J.G. is a founder of, and has an equity interest in, RESA Therapeutics, Inc. S.E.Y. is a founder and the director of Compbiosphere, a computational biology analytics and consulting company.

STAR★METHODS

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

- KEY RESOURCES TABLE
- EXPERIMENTAL MODEL AND STUDY PARTICIPANT DETAILS
  - Zebrafish husbandry and maintenance
  - Embryo injections, transplants, and dissociations
- METHOD DETAILS
  - hCD4-mediated enrichment of transplanted cells
  - The truncated hCD4 sequence was
  - Embryo staining and image acquisition

Figure 6. Expression module analysis of cell states reflecting reprogramming defects and compensatory gene activation

(A) Heatmap showing the median z-scores of genes belonging to nine expression modules (EMs) over 2.5–6 hpf in temporal bulk transcriptome data from WT or MZ*nps* embryos. Z scores clipped at −2 and 2. Transcriptome data from.[17]

(B) Heatmap of enrichment of genes from each EM cobound by Nanog (N), Pou5f3 (P), and Sox19b (S) at accessible enhancer regions at 4 hpf (from (A)). Color, odds ratio of enrichment (log2). Data from.[11,35]

(C–F) Heatmap of enrichment for genes with *miR-430* seed sequences within their 3′UTRs among genes from each EM from (A); (C), for genes that are maternal only, maternal zygotic, and zygotic only among genes from each EM; (D), of chromatin accessibility at promoters (E) and enhancers (F) of genes in each EM in WT and MZ*nps* cells at the 1k (3 hpf), oblong (3.7 hpf), and dome (4.3 hpf) stages, increased blue/green intensity, more open; data from.[17] Color, odds ratio of enrichment (log2). Z, M, and MZ annotations from.[36]

(G) Bubble plot of over-representation of marker genes for MZ*nps* subpopulations (left) and WT host cell types (right) among EMs (from (A); FDR < 0.05). Size denotes FDR (-log10). Over-representation determined by one-sided Fisher's exact test and corrected for multiple hypothesis testing.

(H) Heatmap of KO 4 marker expression (top) in temporal bulk transcriptome data from WT or MZ*nps* embryos. Color bar, EM membership (from A) highlighting (zoom below) those corresponding to NPS dependent activation (yellow) and NPS-dependent strong repression (green). Data from.[17] The mean of scaled (*Z* score) expression for each gene at each time point in each condition is shown. Gene names in magenta/orange, motoneurons markers. See Figures S5 and S6; Table S5.

- Imaging quantification for single transplants
- Data alignment and preprocessing
- Data dimension reduction, graphing and embedding
- Donor cell identification
- Characterization of MZ*nps* subpopulations
- WT cell-type annotation and donor/host composition
- Nearest neighbor developmental age and cell-type
- Marker enrichment in MZ*nps* subpopulations
- Differentially expressed genes analyses
- EM regulation in WT and MZ*nps* mutant embryos

- QUANTIFICATION AND STATISTICAL ANALYSIS

SUPPLEMENTAL INFORMATION

Supplemental information can be found online at https://doi.org/10.1016/j.celrep.2025.116498.

Received: June 17, 2024
Revised: August 19, 2025
Accepted: October 10, 2025

REFERENCES

1. Kojima, M.L., Hoppe, C., and Giraldez, A.J. (2025). The maternal-to-zygotic transition: reprogramming of the cytoplasm and nucleus. Nat. Rev. Genet. *26*, 245–267. https://doi.org/10.1038/s41576-024-00792-0.
2. 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.
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. Yu, J., Vodyanik, M.A., Smuga-Otto, K., Antosiewicz-Bourget, J., Frane, J. L., Tian, S., Nie, J., Jonsdottir, G.A., Ruotti, V., Stewart, R., et al. (2007). Induced pluripotent stem cell lines derived from human somatic cells. Science *318*, 1917–1920. https://doi.org/10.1126/science.1151526.
5. Chambers, I., and Tomlinson, S.R. (2009). The transcriptional foundation of pluripotency. Development *136*, 2311–2322. https://doi.org/10.1242/dev.024398.
6. Onichtchouk, D. (2012). Pou5f1/oct4 in pluripotency control: Insights from zebrafish. genesis *50*, 75–85. https://doi.org/10.1002/dvg.20800.
7. Lee, M.T., Bonneau, A.R., Takacs, C.M., Bazzini, A.A., DiVito, K.R., Fleming, E.S., and Giraldez, A.J. (2013). Nanog, Pou5f1 and SoxB1 activate zygotic gene expression during the maternal-to-zygotic transition. Nature *503*, 360–364. https://doi.org/10.1038/nature12632.
8. 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.
9. Giraldez, A.J., Mishima, Y., Rihel, J., Grocock, R.J., Van Dongen, S., Inoue, K., Enright, A.J., and Schier, A.F. (2006). Zebrafish MiR-430 promotes deadenylation and clearance of maternal mRNAs. Science *312*, 75–79. https://doi.org/10.1126/science.1122689.
10. Yartseva, V., and Giraldez, A.J. (2015). The Maternal-to-Zygotic Transition During Vertebrate Development: A Model for Reprogramming. Curr. Top. Dev. Biol. *113*, 191–232. https://doi.org/10.1016/bs.ctdb.2015.07.020.
11. Miao, L., Tang, Y., Bonneau, A.R., Chan, S.H., Kojima, M.L., Pownall, M.E., Vejnar, C.E., Gao, F., Krishnaswamy, S., Hendry, C.E., and Giraldez, A.J. (2022). The landscape of pioneer factor activity reveals the mechanisms of chromatin reprogramming and genome activation. Mol. Cell *82*, 986–1002.e9. https://doi.org/10.1016/j.molcel.2022.01.024.
12. Kimmel, C.B., and Law, R.D. (1985). Cell lineage of zebrafish blastomeres: III. Clonal analyses of the blastula and gastrula stages. Dev. Biol. *108*, 94–101. https://doi.org/10.1016/0012-1606(85)90012-0.
13. Veil, M., Schaechtle, M.A., Gao, M., Kirner, V., Buryanova, L., Grethen, R., and Onichtchouk, D. (2018). Maternal Nanog is required for zebrafish embryo architecture and for cell viability during gastrulation. Dev. Camb. Engl. *145*, dev155366. https://doi.org/10.1242/dev.155366.
14. Lunde, K., Belting, H.-G., and Driever, W. (2004). Zebrafish *pou5f1/pou2*, Homolog of Mammalian *Oct4*, Functions in the Endoderm Specification Cascade. Curr. Biol. *14*, 48–55. https://doi.org/10.1016/j.cub.2003.11.022.
15. Miltenyi, S., Müller, W., Weichel, W., and Radbruch, A. (1990). High gradient magnetic cell separation with MACS. Cytometry *11*, 231–238. https://doi.org/10.1002/cyto.990110203.
16. Wattrus, S.J., and Zon, L.I. (2020). A Transgenic System for Rapid Magnetic Enrichment of Rare Embryonic Cells. Zebrafish *17*, 354–357. https://doi.org/10.1089/zeb.2020.1904.
17. Riesle, A.J., Gao, M., Rosenblatt, M., Hermes, J., Hass, H., Gebhard, A., Veil, M., Grüning, B., Timmer, J., and Onichtchouk, D. (2023). Activator-blocker model of transcriptional regulation by pioneer-like factors. Nat. Commun. *14*, 5677. https://doi.org/10.1038/s41467-023-41507-z.
18. Babb, S.G., and Marrs, J.A. (2004). E-cadherin regulates cell movements and tissue formation in early zebrafish embryos. Dev. Dyn. *230*, 263–277. https://doi.org/10.1002/dvdy.20057.
19. Thisse, B., Pflumio, S., Fürthauer, M., Loppin, B., Heyer, V., Degrave, A., Woehl, R., Lux, A., Steffan, T., Charbonnier, X.Q., et al. (2001). Expression of the zebrafish genome during embryogenesis.
20. Martinez-Morales, J.R., Signore, M., Acampora, D., Simeone, A., and Bovolenta, P. (2001). Otx genes are required for tissue specification in the developing eye. Dev. Camb. Engl. *128*, 2019–2030. https://doi.org/10.1242/dev.128.11.2019.
21. Lane, B.M., and Lister, J.A. (2012). Otx but Not Mitf Transcription Factors Are Required for Zebrafish Retinal Pigment Epithelium Development. PLoS One *7*, e49357. https://doi.org/10.1371/journal.pone.0049357.
22. Okuda, Y., Ogura, E., Kondoh, H., and Kamachi, Y. (2010). B1 SOX Coordinate Cell Specification with Patterning and Morphogenesis in the Early Zebrafish Embryo. PLoS Genet. *6*, e1000936. https://doi.org/10.1371/journal.pgen.1000936.
23. Oosterveen, T., Kurdija, S., Ensterö, M., Uhde, C.W., Bergsland, M., Sandberg, M., Sandberg, R., Muhr, J., and Ericson, J. (2013). SoxB1-driven transcriptional network underlies neural-specific interpretation of morphogen signals. Proc. Natl. Acad. Sci. *110*, 7330–7335. https://doi.org/10.1073/pnas.1220010110.
24. Graham, V., Khudyakov, J., Ellis, P., and Pevny, L. (2003). SOX2 functions to maintain neural progenitor identity. Neuron *39*, 749–765. https://doi.org/10.1016/s0896-6273(03)00497-5.
25. Archer, T.C., Jin, J., and Casey, E.S. (2011). Interaction of Sox1, Sox2, Sox3 and Oct4 during primary neurogenesis. Dev. Biol. *350*, 429–440. https://doi.org/10.1016/j.ydbio.2010.12.013.
26. Tan, G.C., Mazzoni, E.O., and Wichterle, H. (2016). Iterative Role of Notch Signaling in Spinal Motor Neuron Diversification. Cell Rep. *16*, 907–916. https://doi.org/10.1016/j.celrep.2016.06.067.
27. Wutikeli, H., Xie, T., Xiong, W., and Shen, Y. (2025). ELAV/Hu RNA-binding protein family: key regulators in neurological disorders, cancer, and other diseases. RNA Biol. *22*, 1–11. https://doi.org/10.1080/15476286.2025.2471133.
28. England, S., Hilinski, W., de Jager, S., Andrzejczuk, L., Campbell, P., Chowdhury, T., Demby, C., Fancher, W., Gong, Y., Lin, C., et al. (2014). Identifying Transcription Factors Expressed by Ventral Spinal Cord Interneurons. ZFIN Direct Data Submission. http://zfin.org.
29. Köprunner, M., Thisse, C., Thisse, B., and Raz, E. (2001). A zebrafish nanos-related gene is essential for the development of primordial germ cells. Genes Dev. *15*, 2877–2885. https://doi.org/10.1101/gad.212401.

30. Knaut, H., Pelegri, F., Bohmann, K., Schwarz, H., and Nüsslein-Volhard, C. (2000). Zebrafish vasa RNA but Not Its Protein Is a Component of the Germ Plasm and Segregates Asymmetrically before Germline Specification. J. Cell Biol. *149*, 875–888. https://doi.org/10.1083/jcb.149.4.875.

31. Weidinger, G., Stebler, J., Slanchev, K., Dumstrei, K., Wise, C., Lovell-Badge, R., Thisse, C., Thisse, B., and Raz, E. (2003). dead end, a Novel Vertebrate Germ Plasm Component, Is Required for Zebrafish Primordial Germ Cell Migration and Survival. Curr. Biol. *13*, 1429–1434. https://doi.org/10.1016/S0960-9822(03)00537-2.

32. Farrell, J.A., Wang, Y., Riesenfeld, S.J., Shekhar, K., Regev, A., and Schier, A.F. (2018). Single-cell reconstruction of developmental trajectories during zebrafish embryogenesis. Science *360*, eaar3131. https://doi.org/10.1126/science.aar3131.

33. Lange, M., Granados, A., VijayKumar, S., Bragantini, J., Ancheta, S., Santhosh, S., Borja, M., Kobayashi, H., McGeever, E., Solak, A.C., et al. (2023). Zebrahub – Multimodal Zebrafish Developmental Atlas Reveals the State-Transition Dynamics of Late-Vertebrate Pluripotent Axial Progenitors. Preprint at bioRxiv. https://doi.org/10.1101/2023.03.06.531398.

34. Sur, A., Wang, Y., Capar, P., Margolin, G., Prochaska, M.K., and Farrell, J. A. (2023). Single-cell analysis of shared signatures and transcriptional diversity during zebrafish development. Dev. Cell *58*, 3028–3047.e12. https://doi.org/10.1016/j.devcel.2023.11.001.

35. Xu, C., Fan, Z.P., Müller, P., Fogley, R., DiBiase, A., Trompouki, E., Unternaehrer, J., Xiong, F., Torregroza, I., Evans, T., et al. (2012). Nanog-like regulates endoderm formation through the Mxtx2-Nodal pathway. Dev. Cell *22*, 625–638. https://doi.org/10.1016/j.devcel.2012.01.003.

36. Baia Amaral, D., Egidy, R., Perera, A., and Bazzini, A.A. (2024). miR-430 regulates zygotic mRNA during zebrafish embryogenesis. Genome Biol. *25*, 74. https://doi.org/10.1186/s13059-024-03197-8.

37. Giraldez, A.J. (2010). microRNAs, the cell's Nepenthe: clearing the past during the maternal-to-zygotic transition and cellular reprogramming. Curr. Opin. Genet. Dev. *20*, 369–375. https://doi.org/10.1016/j.gde.2010.04.003.

38. Muñoz-Sanjuán, I., and Brivanlou, A.H. (2002). Neural induction, the default model and embryonic stem cells. Nat. Rev. Neurosci. *3*, 271–280. https://doi.org/10.1038/nrn786.

39. Judson, R.L., Babiarz, J.E., Venere, M., and Blelloch, R. (2009). Embryonic stem cell–specific microRNAs promote induced pluripotency. Nat. Biotechnol. *27*, 459–461. https://doi.org/10.1038/nbt.1535.

40. Anokye-Danso, F., Trivedi, C.M., Juhr, D., Gupta, M., Cui, Z., Tian, Y., Zhang, Y., Yang, W., Gruber, P.J., Epstein, J.A., and Morrisey, E.E. (2011). Highly efficient miRNA-mediated reprogramming of mouse and human somatic cells to pluripotency. Cell Stem Cell *8*, 376–388. https://doi.org/10.1016/j.stem.2011.03.001.

41. Pasque, V., Jullien, J., Miyamoto, K., Halley-Stott, R.P., and Gurdon, J.B. (2011). Epigenetic factors influencing resistance to nuclear reprogramming. Trends Genet. *27*, 516–525. https://doi.org/10.1016/j.tig.2011.08.002.

42. Ugolini, M., Kerlin, M.A., Kuznetsova, K., Oda, H., Kimura, H., and Vastenhouw, N.L. (2024). Transcription bodies regulate gene expression by sequestering CDK9. Nat. Cell Biol. *26*, 604–612. https://doi.org/10.1038/s41556-024-01389-9.

43. Giraldez, A.J., Cinalli, R.M., Glasner, M.E., Enright, A.J., Thomson, J.M., Baskerville, S., Hammond, S.M., Bartel, D.P., and Schier, A.F. (2005). MicroRNAs regulate brain morphogenesis in zebrafish. Science *308*, 833–838. https://doi.org/10.1126/science.1109020.

44. Lund, E., Liu, M., Hartley, R.S., Sheets, M.D., and Dahlberg, J.E. (2009). Deadenylation of maternal mRNAs mediated by miR-427 in Xenopus laevis embryos. RNA *15*, 2351–2363. https://doi.org/10.1261/rna.1882009.

45. Mishima, Y., Giraldez, A.J., Takeda, Y., Fujiwara, T., Sakamoto, H., Schier, A.F., and Inoue, K. (2006). Differential Regulation of Germline mRNAs in Soma and Germ Cells by Zebrafish miR-430. Curr. Biol. *16*, 2135–2142. https://doi.org/10.1016/j.cub.2006.08.086.

46. Ciruna, B., Weidinger, G., Knaut, H., Thisse, B., Thisse, C., Raz, E., and Schier, A.F. (2002). Production of maternal-zygotic mutant zebrafish by germ-line replacement. Proc. Natl. Acad. Sci. USA *99*, 14919–14924. https://doi.org/10.1073/pnas.222459999.

47. Sánchez-Sánchez, A.V., Camp, E., Leal-Tassias, A., Atkinson, S.P., Armstrong, L., Díaz-Llopis, M., and Mullor, J.L. (2010). Nanog regulates primordial germ cell migration through Cxcr4b. Stem Cells Dayt. Ohio *28*, 1457–1464. https://doi.org/10.1002/stem.469.

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

49. Wang, X., Zhu, J., Wang, H., Deng, W., Jiao, S., Wang, Y., He, M., Zhang, F., Liu, T., Hao, Y., et al. (2023). Induced formation of primordial germ cells from zebrafish blastomeres by germplasm factors. Nat. Commun. *14*, 7918. https://doi.org/10.1038/s41467-023-43587-3.

50. Illmensee, K., and Mahowald, A.P. (1974). Transplantation of posterior polar plasm in Drosophila. Induction of germ cells at the anterior pole of the egg. Proc. Natl. Acad. Sci. USA *71*, 1016–1020. https://doi.org/10.1073/pnas.71.4.1016.

51. Raz, E. (2003). Primordial germ-cell development: the zebrafish perspective. Nat. Rev. Genet. *4*, 690–700. https://doi.org/10.1038/nrg1154.

52. Lewis, B.P., Burge, C.B., and Bartel, D.P. (2005). Conserved seed pairing, often flanked by adenosines, indicates that thousands of human genes are microRNA targets. Cell *120*, 15–20. https://doi.org/10.1016/j.cell.2004.12.035.

53. Cunningham, F., Achuthan, P., Akanni, W., Allen, J., Amode, M.R., Armean, I.M., Bennett, R., Bhai, J., Billis, K., Boddu, S., et al. (2019). Ensembl 2019. Nucleic Acids Res. *47*, D745–D751. https://doi.org/10.1093/nar/gky1113.

54. Abraham, A., Pedregosa, F., Eickenberg, M., Gervais, P., Mueller, A., Kossaifi, J., Gramfort, A., Thirion, B., and Varoquaux, G. (2014). Machine learning for neuroimaging with scikit-learn. Front. Neuroinform. *8*, 14. https://doi.org/10.3389/fninf.2014.00014.

55. Howe, D.G., Bradford, Y.M., Eagle, A., Fashena, D., Frazer, K., Kalita, P., Mani, P., Martin, R., Moxon, S.T., Paddock, H., et al. (2017). The Zebrafish Model Organism Database: new support for human disease models, mutation details, gene expression phenotypes and searching. Nucleic Acids Res. *45*, D758–D768. https://doi.org/10.1093/nar/gkw1116.

56. Westerfield, M. (2000). The Zebrafish Book. A Guide for the Laboratory Use of Zebrafish (Danio rerio), 4th Edition (University of Oregon Press).

57. Kemp, H.A., Carmany-Rampey, A., and Moens, C. (2009). Generating Chimeric Zebrafish Embryos by Transplantation. J. Vis. Exp. *1394*, 1394. https://doi.org/10.3791/1394.

58. Kimmel, C.B., Warga, R.M., and Schilling, T.F. (1990). Origin and organization of the zebrafish fate map. Development *108*, 581–594. https://doi.org/10.1242/dev.108.4.581.

59. Boswell, C.W., Hoppe, C., Sherrard, A., Miao, L., Kojima, M.L., Martino, P., Zhao, N., Stasevich, T.J., Nicoli, S., and Giraldez, A.J. (2025). Genetically encoded affinity reagents are a toolkit for visualizing and manipulating endogenous protein function in vivo. Nat. Commun. *16*, 5503. https://doi.org/10.1038/s41467-025-61003-w.

60. Schneider, C.A., Rasband, W.S., and Eliceiri, K.W. (2012). NIH Image to ImageJ: 25 years of image analysis. Nat. Methods *9*, 671–675. https://doi.org/10.1038/nmeth.2089.

61. Virshup, I., Bredikhin, D., Heumos, L., Palla, G., Sturm, G., Gayoso, A., Kats, I., Koutrouli, M., and Scverse Community; and Berger, B. (2023). The scverse project provides a computational ecosystem for single-cell omics data analysis. Nat. Biotechnol. *41*, 604–606. https://doi.org/10.1038/s41587-023-01733-8.

62. Wolf, F.A., Angerer, P., and Theis, F.J. (2018). SCANPY: large-scale single-cell gene expression data analysis. Genome Biol. *19*, 15. https://doi.org/10.1186/s13059-017-1382-0.

63. Wolock, S.L., Lopez, R., and Klein, A.M. (2019). Scrublet: Computational Identification of Cell Doublets in Single-Cell Transcriptomic Data. Cell Syst. *8*, 281–291.e9. https://doi.org/10.1016/j.cels.2018.11.005.

64. Moon, K.R., van Dijk, D., Wang, Z., Gigante, S., Burkhardt, D.B., Chen, W. S., Yim, K., Elzen, A.v.d., Hirn, M.J., Coifman, R.R., et al. (2019). Visualizing structure and transitions in high-dimensional biological data. Nat. Biotechnol. *37*, 1482–1492. https://doi.org/10.1038/s41587-019-0336-3.

65. van Dijk, D., Sharma, R., Nainys, J., Yim, K., Kathail, P., Carr, A.J., Burdziak, C., Moon, K.R., Chaffer, C.L., Pattabiraman, D., et al. (2018). Recovering Gene Interactions from Single-Cell Data Using Data Diffusion. Cell *174*, 716–729.e27. https://doi.org/10.1016/j.cell.2018.05.061.

66. Heaton, H., Talman, A.M., Knights, A., Imaz, M., Gaffney, D.J., Durbin, R., Hemberg, M., and Lawniczak, M.K.N. (2020). Souporcell: robust clustering of single-cell RNA-seq data by genotype without reference genotypes. Nat. Methods *17*, 615–620. https://doi.org/10.1038/s41592-020-0820-1.

67. Robinson, J.T., Thorvaldsdóttir, H., Winckler, W., Guttman, M., Lander, E. S., Getz, G., and Mesirov, J.P. (2011). Integrative genomics viewer. Nat. Biotechnol. *29*, 24–26. https://doi.org/10.1038/nbt.1754.

68. Traag, V.A., Waltman, L., and van Eck, N.J. (2019). From Louvain to Leiden: guaranteeing well-connected communities. Sci. Rep. *9*, 5233. https://doi.org/10.1038/s41598-019-41695-z.

69. Harris, M.A., Clark, J., Ireland, A., Lomax, J., Ashburner, M., Foulger, R., Eilbeck, K., Lewis, S., Marshall, B., Mungall, C., et al. (2004). The Gene Ontology (GO) database and informatics resource. Nucleic Acids Res. *32*, D258–D261. https://doi.org/10.1093/nar/gkh036.

70. Kanehisa, M., and Goto, S. (2000). KEGG: kyoto encyclopedia of genes and genomes. Nucleic Acids Res. *28*, 27–30. https://doi.org/10.1093/nar/28.1.27.

71. Agrawal, A., Balcı, H., Hanspers, K., Coort, S.L., Martens, M., Slenter, D. N., Ehrhart, F., Digles, D., Waagmeester, A., Wassink, I., et al. (2024). WikiPathways 2024: next generation pathway database. Nucleic Acids Res. *52*, D679–D689. https://doi.org/10.1093/nar/gkad960.

72. Fang, Z., Liu, X., and Peltz, G. (2023). GSEApy: a comprehensive package for performing gene set enrichment analysis in Python. Bioinforma. Oxf. Engl. *39*, btac757. https://doi.org/10.1093/bioinformatics/btac757.

73. Waskom, M. (2021). seaborn: statistical data visualization. J. Open Source Softw. *6*, 3021. https://doi.org/10.21105/joss.03021.

74. Clough, E., and Barrett, T. (2016). The Gene Expression Omnibus Database. Methods Mol. Biol. *1418*, 93–110. https://doi.org/10.1007/978-1-4939-3578-9_5.

STAR★METHODS

KEY RESOURCES TABLE

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Antibodies		
Anti-hCD4-FITC	Miltenyi Biotec	Cat# 130-114-531; RRID:AB_2726690
Mouse monoclonal anti-GFP	Thermo Fisher Scientific	Cat# A11120; RRID:AB_221568
Rabbit polyclonal anti-Cleaved Caspase-3 (Asp175)	Cell Signaling Technology	Cat# 9661S; RRID:AB_2341188
Goat anti-Mouse IgG (H + L) Cross-Adsorbed Secondary Antibody, Alexa Fluor™ 488	Thermo Fisher Scientific	Cat# A32723; RRID:AB_2633275
Goat anti-Rabbit IgG (H + L) Cross-Adsorbed Secondary Antibody, Alexa Fluor™ 546	Thermo Fisher Scientific	Cat# A11010; RRID:AB_2534077
Experimental models: organisms/strains		
Zebrafish: MZ*nps* (maternal-zygotic *nanog*$^{-/-}$; *pou5f3*$^{-/-}$;*sox19b*$^{-/-}$) triple mutant	Miao et al.[1]	N/A
Zebrafish: wildtype TU-AB and TL strains	N/A	N/A
Zebrafish: *Tg(His-GFP)*	Pauls et al.[52]	*kca66*
Chemicals, peptides, and recombinant proteins		
16% Paraformaldehyde	Electron Microscopy Sciences	Cat# 50-980-487
DAPI	Thermo Fisher Scientific	Cat# D1306
Protease from Streptomyces griseus (Pronase)	Sigma Aldrich	Cat# P5147, CAS: 9036-06-0
Tween 20	Sigma Aldrich	Cat# X100; CAS: 9036-19-5
Normal Goat Serum (NGS)	Thermo Fisher Scientific	Cat# 50062Z
Agarose GPG/LMP™	AmericanBio	Cat# AB00981; CAS: 9012-36-6
HEPES	Sigma Aldrich	Cat# H4034; CAS: 7365-45-9
Bovine Serum Albumin (BSA)	Sigma Aldrich	Cat# A9647
EDTA	Sigma Aldrich	Cat# ED3SS; CAS: 85715-60-2
Trypan Blue	Thermo Fisher Scientific	Cat# 15250061; CAS: 72-57-1
Proteinase K	Thermo Fisher Scientific	Cat# EO0492
Critical commercial assays		
hCD4 MicroBeads	Miltenyi Biotec	Cat#130-097-048
MiniMACS Separator	Miltenyi Biotec	Cat#130-090-312
mMESSAGE mMACHINE™ SP6 Transcription Kit	Thermo Fisher Scientific	Cat#AM1340
In-Fusion® HD Cloning Plus	TaKaRa	Cat#639650
10 x scRNA 3′ kit v3	10x Genomics	PN-1000075
Deposited data		
Zebrafish genome GRCz11	Ensembl	https://www.ensembl.org/Danio_rerio; RRID:SCR_002344
ATAC-seq datasets from WT and MZ*nps* mutant embryos	Gene Expression Omnibus Database; Clough et al.[2]	GSE215956
Single cell RNA-seq of transplanted WT and MZ*nps* cells	This paper	PRJNA1126192 and GSE309740
Embryonic zebrafish scRNA-seq timecourses	Farrell et al.[32]; Lange et al.[33]; and Sur et al.[34]	GSE106587; SRA BioProject ID PRJNA940501; and GSE223922
Datasets for Omni-ATAC seq, ChIP-seq, RNA-seq, CUT&TAG in WT and MZ*nps* embryos	Miao et al.[11]	SRA: SRP355652
Omni-ATAC-seq datasets in zebrafish embryos	Riesle et al.[17]	GSE215956
RNA-seq time series for zebrafish embryos	Riesle et al.[17]	GSE162415
ChIP-seq for Nanog	Xu et al.[35]	GSE34683

(*Continued on next page*)

Continued

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Recombinant DNA		
pCS2+dsRed	This paper	N/A
pCS2+EGFP-CAAX	This paper	N/A
pCS2+nanog	Lee et al.[3]	N/A
pCS2+pou5f3	Lee et al.[3]	N/A
pCS2-hCD4	This paper	N/A
Software and algorithms		
IMARIS; v10.0	Bitplane, Oxford Instruments, Concord MA	RRID:SCR_007370
CellRanger pipeline; v3.0.2	10x Genomics	https://www.10xgenomics.com/support/software/cell-ranger/latest; RRID:SCR_023221
Python; (Version 3.11.5)	The Python Software Foundation	https://www.python.org; RRID:SCR_008394
R; (Version 4.4)	The R foundation	https://www.r-project.org; RRID:SCR_001905
Souporcell; v2.5	Heaton et al.[4]	RRID:SCR_027462
anndata; v0.9.1	Virshup et al.[5]	RRID:SCR_018209
scanpy; v1.9.3	Wolf et al.[6]	RRID:SCR_018139
scrublet; v0.2.3	Wolock et al.[7]	RRID:SCR_018098
Integrative Genomics Viewer; v2.18.4	Robinson et al.[8]	http://www.broadinstitute.org/igv/; RRID:SCR_011793
leidenalg; v0.10.0	Traag et al.[9]	N/A
GSEApy; v1.0.6	Fang et al.[10]	RRID:SCR_025803
MAGIC; v3.0.0	Van Dijk et al.[11]	https://github.com/KrishnaswamyLab/MAGIC; RRID:SCR_022371
graphtools; v1.5.3	N/A	https://github.com/KrishnaswamyLab/graphtools
PHATE; v1.0.11	Moon et al.[12]	https://github.com/KrishnaswamyLab/PHATE; RRID:SCR_027119
Ensembl; Danio rerio v95	Cunningham et al.[53]	https://ftp.ensembl.org/pub/release-95/gtf/danio_rerio/Danio_rerio.GRCz11.95.gtf.gz; RRID:SCR_002344
Gene Ontology (GO); v2018	Harris et al.[13]	RRID:SCR_002811
KEGG	Kanehisa et al.[14]	RRID:SCR_012773
Wiki-pathways	Agrawal et al.[15]	RRID:SCR_002134
ggplot2; v3.5.1	R package	https://cran.r-project.org/web/packages/ggplot2/citation.html; RRID:SCR_014601
Scikit-learn; v1.3.0	Abraham et al.[54]	RRID:SCR_002577
seaborn	Python tool	https://seaborn.pydata.org/; RRID:SCR_018132
GraphPad Prism (Version 10)	GraphPad Software	RRID:SCR_002798
FIJI/ImageJ	Schneider et al.[16]	https://imagej.nih.gov/ij/; RRID:SCR_002285; RRID:SCR_003070
Other		
Coverslips No. 1.5; 24 × 60mm	Thermo Fisher Scientific	Cat# 22-050-246
Zeiss LSM 980 upright confocal microscope with Airyscan 2 detector	Zeiss	RRID:SCR_025048
10x Genomics Chromium Chip	10x Genomics	1000074

EXPERIMENTAL MODEL AND STUDY PARTICIPANT DETAILS

Zebrafish husbandry and maintenance

The husbandry of zebrafish was conducted as previously documented.[55,56] The maintenance of fish lines was carried out in accordance with the AAALAC research guidelines following a protocol approved by the Yale University IACUC.

The maternal-zygotic (MZ) *nanog*$^{-/-}$;*pou5f3*$^{-/-}$;*sox19b*$^{-/-}$ triple homozygous mutants (MZ*nps*) were maintained as described.[11] Briefly, MZ*nps* or MZ*nanog*$^{-/-}$;*pou5f3*$^{+/-}$;*sox19b*$^{-/-}$ females were crossed with MZ*nps* males, and the resulting embryos were rescued with mRNA injection (25 pg *nanog* mRNA +30 pg *pou5f3* mRNA) at the 1-cell stage. Alternatively, MZ*nanog*$^{+/-}$;*pou5f3*$^{+/-}$;*sox19b*$^{+/-}$ females were crossed with MZ*nps* males, and the resulting embryos were rescued by injecting 30 pg *pou5f3* mRNA at the 1-cell stage. Progenies from MZ*nanog*$^{-/-}$;*pou5f3*$^{+/-}$;*sox19b*$^{-/-}$ or MZ*nanog*$^{+/-}$;*pou5f3*$^{+/-}$;*sox19b*$^{+/-}$ females were genotyped as previously described.[11]

Embryo injections, transplants, and dissociations

Embryos were dechorionated at the 1-cell stage. 100 pg hCD4 mRNA (described below) and 50 pg dsRed capped mRNA were injected into donor embryos (both WT and MZ*nps* embryos, pooled from 2 to 8 pairs for each replicate) at the 1-cell stage. 75 pg EGFP-CAAX capped mRNA was injected into host wild-type embryos at the 1-cell stage (pooled from 4 to 8 pairs). At 3–4 hpf, about 50 cells were taken from donor embryos and transplanted into host WT embryos (one embryo serves as a donor for about five different hosts). Transplantation was carried out as described[57] by three independent operators for each transplantation. For each replicate, we transplanted 200–300 embryos and kept them at 28°C until dissociation. At ~3-6ss stage (corresponding to ~11–12 hpf), about 60 successfully transplanted and healthy-looking hosts for each replicate (across ~250 hosts) were selected for further procedures. Embryos were rapidly deyolked manually with forceps in fresh, ice-cold Ringer's solution (per 50 mL: 5.8mL 1M NaCl; 73μL 2M KCl; 250μL 1M HEPES in 43.5 mL RNase-free water). 50 perfectly deyolked embryos were rinsed and gently mechanically dissociated using polished glass pipettes in fresh ice-cold Ringer's solution, centrifuged at 4°C at 300 rpm for 3 min, supernatant removed, and the remaining cells were resuspended in fresh ice-cold PBE solution (1X RNase-free PBS, 0.5% BSA, 2mM EDTA). Cells were then incubated with a 1:100 hCD4 Microbeads solution (in PBS) (Miltenyi Biotec), incubated for 10 min on ice, and washed 2X with ice-cold PBE solution. Cells were resuspended in 1mL PBE and placed through a MACS magnetic column (MiniMACS Separator, Miltenyi Biotec). Cells were collected into a clean RNase-free Lo-Bind tube, washed 2X with ice-cold PBE, and immediately prepared for loading onto a 10X Genomics chip (v3). The protocol was optimized and enrichment efficiencies calculated using an anti-hCD4-FITC antibody (to detect expression in zebrafish embryos), and hCD4 mRNA-injected *Tg(H2B-GFP)* embryos, mixed 1:9 with WT embryos, and cells counted using a hemocytometer and Trypan Blue to assess cell viability. From single cell RNA sequencing experiments, we noted an over-representation of cell types of ectodermal origin in WT donor cells relative to WT host cell populations, although all tissue types were represented, indicating a potential technical bias introduced at the site of transplantation (as shown in Figures 2E and 4A).[58]

METHOD DETAILS

hCD4-mediated enrichment of transplanted cells

A truncated human CD4 sequence (corresponding to the sequence identified by hCD4 MicroBeads, Miltenyi Biotec) was cloned into a pCS2+ vector containing an SP6 promoter, a stabilizing 3′UTR derived from the zebrafish *prrg2* endogenous gene,[59] and an SV40 pA.

The truncated hCD4 sequence was

5′-ATGAACCGGGGAGTCCCTTTTAGGCACTTGCTTCTGGTGCTGCAACTGGCGCTCCTCCCAGCAGCCACTCAGGGAAAGAAAGTGGTGCTGGG
CAAAAAAGGGGATACAGTGGAACTGACCTGTACAGCTTCCCAGAAGAAGAGCATACAATTCCACTGGAAAAACTCCAACCAGATAAAGATTCTGG
GAAATCAGGGCTCCTTCTTAACTAAAGGTCCATCCAAGCTGAATGATCGCGCTGACTCAAGAAGAAGCCTTTGGGACCAAGGAAACTTCCCCCTG
ATCATCAAGAATCTTAAGATAGAAGACTCAGATACTTACATCTGTGAAGTGGAGGACCAGAAGGAGGAGGTGCAATTGCTAGTGTTCGGATTGAC
TGCCAACTCTGACACCCACCTGCTTCAGGGGCAGAGCCTGACCCTGACCTTGGAGAGCCCCCCTGGTAGTAGCCCCTCAGTGCAATGTAGGAGTC
CAAGGGGTAAAAACATACAGGGGGGGAAGACCCTCTCCGTGTCTCAGCTGGAGCTCCAGGATAGTGGCACCTGGACATGCACTGTCTTGCAGAAC
CAGAAGAAGGTGGAGTTCAAAATAGACATCGTGGTGCTAGCTTTCCAGAAGGCCTCCAGCATAGTCTATAAGAAAGAGGGGGAACAGGTGGAGTT
CTCCTTCCCACTCGCCTTTACAGTTGAAAAGCTGACGGGCAGTGGCGAGCTGTGGTGGCAGGCGGAGAGGGCTTCCTCCTCCAAGTCTTGGATCA
CCTTTGACCTGAAGAACAAGGAAGTGTCTGTAAAACGGGTTACCCAGGACCCTAAGCTCCAGATGGGCAAGAAGCTCCCGCTCCACCTCACCCTG
CCCCAGGCCTTGCCTCAGTATGCTGGCTCTGGAAACCTCACCCTGGCCCTTGAAGCGAAAACAGGAAAGTTGCATCAGGAAGTGAACCTGGTGGT
GATGAGAGCCACTCAGCTCCAGAAAAATTTGACCTGTGAGGTGTGGGGACCCACCTCCCCTAAGCTGATGCTGAGCTTGAAACTGGAGAACAAGG
AGGCAAAGGTCTCGAAGCGGGAGAAGGCGGTGTGGGTGCTGAACCCTGAGGCGGGGATGTGGCAGTGTCTGCTGAGTGACTCGGGACAGGTCCTG
CTGGAATCCAACATCAAGGTTCTGCCCACATGGTCGACCCCGGTGCAGCCAATGGCCCTGATTGTGCTGGGGGGCGTCGCCGGCCTCCTGCTTTT
CATTGGGCTAGGCATCTTCTTCTGTGTCAGGTGCCGGCAC – 3′

Embryo staining and image acquisition

Embryos were collected at ~12 hpf, fixed overnight in cold 4% paraformaldehyde (PFA) in 1x PBS overnight, washed with 1X PBS-Tw (0.1% Tween 20) three times, permeabilized with 10μg/ml Proteinase K for 1 min, washed with PBS-Tw three times, re-fixed with 4% PFA/PBS-Tw, and washed again with PBS-Tw three times. Embryos were blocked with 10% normal goat serum (Thermo Fisher Scientific, 50062Z) for one hour rotating at room temperature, stained with primary antibodies against GFP Tag Monoclonal Antibody (3E6) (1:500, mouse, Thermo Fisher Scientific, A11120) and cleaved Caspase-3 (rabbit, 1:500, CST.

9661S) rotating overnight at 4°C, washed with PBS-Tw three times at room temperature, then stained with Alexa Fluor 488 anti-mouse, Alexa Fluor 546 anti-rabbit, and DAPI for one hour rotating at room temperature. Embryos were washed with PBS-Tw three times. Embryos were mounted in 0.8% low melt agarose (GPG/LMP AmericanBio, AB00981-00050) in water against a no. 1.5 coverslip. Whole embryo images at 12 hpf were acquired using an upright Zeiss LSM 980 confocal microscope with an Airyscan 2 detector and an air EC Plan-Neofluar 10x/0.3 M27 objective. In detail, images were obtained at 16 Bit with 2x line averaging, bidirectional scanning, LSM scan speed 6 (pixel dwell time 0.51 μs), 1.1x optical zoom, GaAsP-PMT detectors, an image size of 4084 × 4084 pixels corresponding to 769.13 μm × 769.13 μm and three-dimensional optical stacks were acquired at 2 μm spacing. Samples were illuminated with the 405nm laser line (DAPI) at 0.2–1% laser power, 820V (gain) and 0 (offset); 488nm laser line (Alexa Fluor 488) at 4% laser power, 850V (gain) and 0 (offset); 561nm laser line (Alexa Fluor 546) at 1.1% laser power, 850V (gain) and 0 (offset). Raw images were deconvolved using the Airyscan software and images shown in the result section are maximum projected, with brightness and contrast adjusted separately between experiments.

Regions of interest were imaged using the above confocal scope with an LD LCI Plan-Apochromat 40x/1.2 Imm Corr DIC M2 objective with water immersion. Settings that differ from the above whole embryo imaging were as follows: LSM scan speed 7 (pixel dwell time 0.35 μs) and three-dimensional optical stacks were acquired at 0.3 μm spacing. Samples were illuminated with the 405nm laser line (DAPI) at 0.2% laser power, 820V (gain) and 0 (offset); 488nm laser line (Alexa Fluor 488) at 1% laser power, 850V (gain) and 0 (offset); 561nm laser line (Alexa Fluor 546) at 1.1% laser power, 850V (gain and 0 (offset). Images were acquired at 4084 × 4084 or 3016 × 2160-3016 pixels corresponding to 176.26 μm × 176.26 μm and 124.29 μm x 89.01–124.29 μm, respectively. The optical zoom (1.2 or 1.7) and therefore image sizes were varied to optimally image specific tissue regions optimally. Raw images were deconvolved using the Airyscan software and then processed with ImageJ software.[60] Images shown in the result section show maximum projected sub-stacks with brightness and contrast adjusted separately between images to optimize visualization of GFP+ donor cells in their tissue environment.

Imaging quantification for single transplants

3D image analysis was performed in the IMARIS software (Bitplane, Oxford Instruments, Concord MA; Version: 10.0). An analysis region was defined for each image to include the embryo tissue and exclude the yolk area so that GFP+ donor cells outside this region were excluded from quantification. GFP+ cells were identified using the 'spot' function and a constant size of 8 μm spheres, to exclude GFP+ cell remnants. Cells were categorized based on their DAPI and cleaved Caspase-3 fluorescence intensity and then manually checked.

Data alignment and preprocessing

Single cell data was aligned to the zebrafish genome (GRCz11) with the addition of sequences of GFP, dsRed and hCD4. Unique molecular identifiers (UMIs) for each gene (Ensembl *Danio rerio* transcriptome v95[53] plus GFP, dsRed and hCD4) quantified per cell using the CellRanger pipeline (v3.0.2, 10x Genomics). The cells by genes counts matrix was loaded into python (v3.11.5) as an anndata object (v0.9.1,[61]) and preprocessed using scanpy (v1.9.3,[62]). Cells with a high likelihood of being doublets were identified with scrublet (doublet score >0.4, v0.2.3,[63]) and removed. Cells with <1000 UMIs or >25% mitochondrial content were removed. Genes detected (UMI>0) in fewer than 5 cells, as well as GFP, dsRed and hCD4 marker genes, and genes on the mitochondrial genome were also removed. The final single-cell dataset contained expression measurements for 19,803 genes in 10,551 cells (donor and host), comprised of 1,488 and 1,226 cells from each MZ*nps* transplant replicate, respectively, and 7,837 cells from the WT transplanted sample.

Data dimension reduction, graphing and embedding

Count data per cell were normalized by library size (UMIs per 10,000) to account for difference in sequencing depth between cells and transformed (square-root) to account for the heteroscedasticity between genes with intrinsically different expression levels. Highly variable genes (HVGs) were identified within each sample as the top 10% of genes with the highest mean to variance ratio, and the union of HVGs in all samples defined 6,632 HVGs in the dataset. The dimensionality of HVG expression was further reduced to the top 100 principal components (PCs) (explaining >99.9% of variance) which were used to calculate a cell x cell distance matrix (Euclidean) and generate a k-nearest neighbor-graph (k = 5) with graphtools (v1.5.3, https://github.com/KrishnaswamyLab/graphtools). For visualization, this graph was embedded in 2-dimensions using PHATE (v1.0.11,[64]) and scatterplots generated using sc.pl.embedding. This graph was also used to smooth gene expression values and correct for drop out using MAGIC (v3.0.0,[65]). Smoothed values were not used for differential expression testing.

Donor cell identification

MZ*nps* knockout (KO) donor cells were distinguished from WT host cells based on the patterns of single-nucleotide-variants (SNVs) in single-cell reads per cell using software package Souporcell (v2.5).[66] In both mutant-donor samples, this package identified SNVs present in genes on chromosome 1 and then clustered cells in each sample into distinct genotypes based on the count of variant alleles. In each sample, the cluster pertaining to MZ*nps* donor cells was identified based on enrichment of dsRed and hCD4 marker expression. Finally, strain-specific SNVs we validated using genome sequencing data from MZ*nps* and WT embryos where the genotype is known.[11] As the genetic background of the WT donor and WT host cells was identical, WT donor cells were defined based

on dsRed and hCD4 expression alone. Cells in the WT-donor sample were annotated as donor cells if both dsRed and hCD4 were detected (UMI >0), or either dsRed/hCD4 were highly expressed (≥4 UMI). Through this approach, 810 KO donor cells, 3,869 WT donor cells and 5,872 WT host cells were identified. SNV alignment tracks were visualized in IGV v2.18.4.[67]

Venn diagrams of genes detected in MZ*nps* and/or WT donor cells (UMI >0) were constructed using the venneuler R-package (v1.1-4). Genes were annotated as strictly zygotic as per.[36] Violin plots were generated using scanpy.pl.violin.

Characterization of MZ*nps* subpopulations

To identify MZ*nps* subpopulations, the total dataset was subset to the 810 KO-cells before data dimensionality reduction, graph construction and embedding was performed, as above. Leiden clustering[68] (leidenalg v0.10.0) was performed on the cell x cell graph (resolution = 0.4) which identified five cell clusters. Markers for each subpopulation were identified as protein-coding genes that were differentially expressed (scanpy.tl.rank_genes_groups, method = 't test_overestim_var', FDR <0.05) or differentially detected (two-sided Fisher's exact test, FDR <0.05) in each subpopulation relative to cells not in the subpopulation. Additionally, markers were filtered to those detected in fewer than 20% cells not in the subpopulation and fewer than 60% of cells overall (to exclude housekeeping genes). The top markers were identified based on the combined (sum) ranking of both differential expression and detection. Significant over-representation (one-sided Fisher's exact test, FDR <0.05) of Gene Ontology[69] (v2018), KEGG,[70] and Wiki-pathways[71] genesets among marker genes was tested using gseapy (v1.0.6).[72] Heatmaps of marker gene expression were generated using scanpy.pl.heatmap(standard_scale = 'var').

To score apoptosis gene expression in cells from each MZ*nps* sub-population and WT cell type, we first curated a geneset of apoptosis related genes from those annotated in the gene ontology genesets apoptotic process (GO:0006915) or positive regulation of apoptotic process (GO:0043065). The expression of these genes was smoothed using MAGIC (v3.0.0,[65]) and scaled to zero mean and unit variance (*Z* score). Negative z-scores were clipped at zero (to make sure genes with low expression do not have a negative contribution) and positive z-scores clipped at three (to ensure the score is not dominated by the high expression of a single gene). The apoptosis geneset score was calculated per cell as the mean of clipped z-scores in the geneset divided by the max clipped *Z* score (i.e., three). Violin plots of apoptosis geneset scores were generated using scanpy.pl.violin. Heatmaps of apoptosis geneset expression were generated using sns.clustermap(standard_scale = 1) from seaborn package (v0.12.2,[73]).

WT cell-type annotation and donor/host composition

To annotate host and donor WT cells in our dataset (query), we first curated a reference WT dataset spanning 4–24 hpf, combining cells from published datasets.[32–34] HVGs were detected as described above and expression scaled within each dataset before datasets were concatenated. The combined HVG expression was reduced to the top 100 principal components (PCs) which were used to create a joint mutual nearest neighbor (MNN) graph of cells from all three datasets with graphtools (v1.5.3, https://github.com/KrishnaswamyLab/graphtools). Edges between cells from non-adjacent timepoints were masked (set to zero) in the graph kernel unless their edge weight was >0.8. The masked graph was embedded in 2 dimensions using PHATE for visualization (v1.0.11,[64]). Leiden clustering (resolution = 2) was performed on the masked graph and identified 37 clusters of cells. We identified marker genes between these clusters as described above and annotated cell types in the reference based on their expression of lineage defining genes and their timepoint of collection (consistent with the known cell-types present at a given developmental stage). This annotated reference was then subset to 3.3–12 hpf timepoints and used to annotate WT cells in our dataset.

To transfer the cell type annotation from the reference dataset to the WT cells in our dataset, HVGs in each dataset were identified as above, and HVG expression scaled within each dataset before concatenation. The dimensionality of the combined dataset was reduced to the top 93 principal components (PCs) (explaining >99.9% of the variance). Based on these PCs, the Euclidean distance between cells in the reference and query dataset was calculated and used to train a k-nearest neighbor classifier (sklearn.neighbors.KNeighborsClassifier, neighbors = 5, v1.3.0,[54]) to annotate the cell types of WT cells in the query dataset.

To compare the cell-type composition between WT donor cells and WT host cells from each sample ($n = 3$), we calculated a 95% confidence-interval around the mean proportion of each cell-type in host. We then calculated the proportion of cells from each cell-type in the WT donor and compare it this reference. Cell-type composition reference and donor proportion was visualized using ggplot2 (v3.5.1) in R (v4.4).

Nearest neighbor developmental age and cell-type

To analyze the developmental age and cell-types of the WT cells most similar to MZnps cells, we identified the nearest-neighbors to the cells from each sub-population in the 4–24 reference time course. As above, HVGs in each dataset (query and reference) were identified and HVG expression scaled within each dataset before the dimensionality of the combined dataset was reduced to the top 100 PCs (explaining >99.9% of the variance). Based on these PCs, the Euclidean distance between query and reference cells was calculated and the 5 nearest neighbors to query cells identified in the WT reference. The distribution of developmental timepoints among the reference nearest neighbors was analyzed for each MZ*nps* sub-populations and WT cell-type in the query dataset and visualized using scanpy.pl.violin. The cell type enrichment among reference nearest neighbors to each MZ*nps* sub-population relative to the overall cell-type abundance in the reference dataset was determined (one-sided Fisher's exact test, FDR <0.05) using the gseapy package (gseapy.enrich(), v1.0.6).[72] Enrichment bubble plots were created using scanpy.pl.dotplot. Reference dataset

PHATE embeddings were visualized using scanpy.pl.embedding(). Correlation between query MZ*nps* sub-populations and WT tissues with reference WT-tissues at 12hpf was calculated based on the mean of PCs in cells from each group.

Marker enrichment in MZ*nps* subpopulations

To determine the enrichment of cell-type specific marker genes among the marker genes of MZ*nps* subpopulations, we first defined a database of gene enriched in each WT cell-type in the dataset using the same marker-gene identification strategy described for the MZ*nps* subpopulations above, filtering housekeeping genes expressed in more than 50% of cells in the total dataset. The markers of each MZ*nps* subpopulation were then tested for significant over representation (one-sided Fisher's exact test, FDR <0.05) among the top 50 markers of each WT cell-type using the gseapy package (gseapy.enrich(), v1.0.6).[72] Enrichment bubble plots were created in R (v4.4) using the ggplot2 package (v3.5.1). Dot plots of MZ*nps* marker, WT cell type marker, and morphogen receptor expression in MZ*nps* subpopulations and WT cell types were created using scanpy.pl.dotplot.

Differentially expressed genes analyses

To calculate differentially expressed genes (DEGs) between MZ*nps* subpopulations and their most similar WT cell-types (i.e., KO3 vs. early-neurons, KO4 vs. motoneurons, KO5 vs. PGCs) the dataset was subset to just cells in either group and genes filtered to those expressed in at least one group (UMI >0). We then identified genes that were both differentially expressed (scanpy.tl.rank_genes_groups, method = 't test_overestim_var', FDR <0.01) and differentially detected (two-sided Fisher's exact test, FDR <0.01) between groups, with the exception of the KO 5 vs. PGCs comparison. As this comparison was underpowered (12 vs. 11 cells in each group), uncorrected *p*-values were used. Heatmaps of gene detection were generated using scanpy.pl.heatmap and gene annotations visualized using the python package seaborn (v0.12.2, seaborn.heatmap).

We analyzed differentially expressed genes for the overrepresentation of NPS-bound and *miR-430* target genes, as well as maternally-deposited vs. zygotically-expressed genes at genome activation. NPS-binding at enhancer regions (TSS +/− 5Kb) was determined based on the intersection between the Nanog, Pou5f3 and Sox19b ChIP binding peak-regions in WT embryos as well as ATAC accessible regions in WT or mutant embryos at ~4 hpf, as defined in Miao et al.[11] Genes with *miR-430* target sites were annotated according to the presence of the seed 6-mer recognition sequence (GCACTT,[9,52]) in annotated 3′ untranslated regions (3′UTRs). Strictly maternal, maternal zygotic and strictly zygotic genes were annotated as per.[36] The statistically significant over-representation of NPS-binding and *miR-430* target sites was tested with gseapy (v1.0.6, one-sided Fisher's exact test). Bubble plots were created using scanpy.pl.dotplot.

EM regulation in WT and MZ*nps* mutant embryos

To analyze the expression of genes associated with MZ*nps* sub-populations at genome activation we used a publicly available time-course measuring gene expression in WT and MZ*nps* mutant embryos between 2.5-6hpf (sampled every 30 min, n = 2–6 per condition per timepoint, 70 samples total) (GEO ascension GSE162415).[74] Expression per sample was library size normalized (TMM) and square-root transformed. Genes were subset to those that were highly variable in our single cell dataset (the top 10% of genes with the highest mean to variance ratio), either overall or within MZ*nps* cells (8035 genes). The mean expression of each gene was calculated across replicates for each condition and timepoint and gene x gene correlation matrix calculated. The correlation matrix was used for hierarchical clustering (scipy.cluster.hierarchy.linkage(method = 'ward')) and the classification of genes into nine gene expression modules (EMs) (scipy.cluster.hierarchy.cut_tree(n_clusters = 9)). The median of scaled expression values for genes in each EM was visualized using seaborn.clustermap. EMs were analyzed for the significant overrepresentation (1-sided Fisher's exact test, FDR <0.05) of NPS-bound and *miR-430* target genes, as well as maternally-deposited vs. zygotically-expressed genes at genome activation (as annotated above) using the gseapy package (gseapy.enrich).

The accessibility of promoters (canonical TSS ±500, as annotated in Miao et al.[11]) and enhancers (defined as ±200 bp peak within +/−5Kb) regions for genes in each EM were analyzed using ATAC data collected from WT and MZ*nps* mutant embryos at 1k, dome and oblong stages of embryo development publicly available accessed through GEO[74] (GSE215956, preprocessed.bw files). The mean of library depth-normalized and scaled (sklearn.preprocessing.MinMaxScaler) read counts at promoters and enhancers (within ±200bp of summit) were calculated for genes within each EM. Heatmaps of enrichment statistics and promoter/enhancer coverage were generated using seaborn.heatmap.

The overrepresentation of MZ*nps* subpopulation and WT cell-type markers among EMs was tested using the gseapy package (gseapy.enrich, one-sided Fisher's exact test, FDR <0.05) and visualized using scanpy.pl.dotplot. Heatmaps of MZ*nps* marker temporal expression in each condition were generated using scanpy.pl.matrixplot.

QUANTIFICATION AND STATISTICAL ANALYSIS

A modified Student's t-test that overestimates variance was performed to identify differentially expressed markers for each subpopulation and genes that were differentially expressed between MZ*nps* subpopulations and their most similar WT cell-types. One-sided Fisher's exact tests were performed to test for significant over-representation of gene sets and marker genes for MZ*nps* subpopulations; for reference cell type enrichment of nearest neighbors to each MZ*nps* sub-population; for significant over-representation of

NPS-binding and *miR-430* target sites; for significant overrepresentation of NPS-bound, *miR-430* target, and maternally-deposited vs. zygotically-expressed at ZGA genes among EMs; and for over-representation of MZ*nps* subpopulation and WT cell-type markers among EMs. Two-sided Fisher's exact tests were performed to test for differentially detected markers for each subpopulation and genes that were differentially detected between MZ*nps* subpopulations and their most similar WT cell-types. The Pearson correlation coefficients were used to compare gene expression in MZ*nps* sub-populations and WT tissues with reference WT cell types.