# Comprehensive profiling of migratory primordial germ cells reveals niche-specific differences in non-canonical Wnt and Nodal-Lefty signaling in anterior vs posterior migrants

Rebecca G. Jaszczak*, Jay W. Zussman*, Daniel E. Wagner, and Diana J. Laird[#]

Eli and Edythe Broad Center for Regeneration Medicine and Stem Cell Research and Department of Obstetrics, Gynecology and Reproductive Science, UCSF, San Francisco, CA 94143 USA

* indicates equal contribution

[#] Correspondence to: diana.laird@ucsf.edu

## ABSTRACT

Mammalian primordial germ cells (PGCs) migrate asynchronously through the embryonic hindgut and dorsal mesentery to reach the gonads. We previously found that interaction with different somatic niches regulates PGC proliferation along the migration route. To characterize transcriptional heterogeneity of migrating PGCs and their niches, we performed single-cell RNA sequencing of 13,262 mouse PGCs and 7,868 surrounding somatic cells during migration (E9.5, E10.5, E11.5) and in anterior versus posterior locations to enrich for leading and lagging migrants. Analysis of PGCs by position revealed dynamic gene expression changes between faster or earlier migrants in the anterior and slower or later migrants in the posterior at E9.5; these differences include migration-associated actin polymerization machinery and epigenetic reprogramming-associated genes. We furthermore identified changes in signaling with various somatic niches, notably strengthened interactions with hindgut epithelium via non-canonical WNT (ncWNT) in posterior PGCs compared to anterior. Reanalysis of a previously published dataset suggests that ncWNT signaling from the hindgut epithelium to early migratory PGCs is conserved in humans. Trajectory inference methods identified putative differentiation trajectories linking cell states across timepoints and from posterior to anterior in our mouse dataset. At E9.5, we mainly observed differences in cell adhesion and actin cytoskeletal dynamics between E9.5 posterior and anterior migrants. At E10.5, we observed divergent gene expression patterns between putative differentiation trajectories from posterior to anterior including Nodal signaling response genes *Lefty1, Lefty2,* and *Pycr2* and reprogramming factors *Dnmt1, Prc1,* and *Tet1.* At E10.5, we experimentally validated anterior migrant-specific *Lefty1/2* upregulation via whole-mount immunofluorescence staining for LEFTY1/2 and phosphorylated SMAD2/3, suggesting that elevated autocrine Nodal signaling in migrating PGCs occurs as they near the gonadal ridges. Together, this positional and temporal atlas of mouse PGCs supports the idea that niche interactions along the migratory route elicit changes in proliferation, actin dynamics, pluripotency, and epigenetic reprogramming.

## INTRODUCTION

The mammalian germline is initially specified as a small pool of primordial germ cells[1] (PGCs) which will give rise to oocytes or spermatozoa and transmit genetic information across generations. A founding population of ~40 mouse PGCs is specified at E7.25[2,3,4] from the epiblast driven by bone morphogenetic protein (BMP) signaling from adjacent extraembryonic tissues[5,6]. Shortly after specification, PGCs begin migrating through the hindgut, moving anteriorly towards the forming gonadal ridges. Signaling cues such as Cxcr4 and KitL are known PGC chemoattractants, and contribute to this journey while also promoting PGC survival[7,8]. Studies observing this migratory process have characterized distinct behavioral patterns in the migratory population. Time lapse imaging across the period of mouse germ cell migration from E9.5-E11.5 showed that PGCs early in their migration are motile within the hindgut, moving both with the morphogenetic changes of the developing embryo and anteriorly in their own right[9]. At E9.5, a behavioral change occurs, as some PGCs begin directed migration out of the hindgut and into the dorsal mesentery tissues. This targeted migration continues into E10.5, by which point leading PGC migrants home in on the developing gonadal ridges[9]. At E10.5, PGCs are located within diverse tissue niches, with migratory leaders arriving in the gonadal ridges, actively migrating cells still in the dorsal mesentery, and migratory laggards remaining in the hindgut. By E11.5, most PGCs have reached the gonadal ridges; however, a population of PGCs remains medial to the gonad and is considered ectopic[10].

Though the spatiotemporal distribution of migrating PGCs is well characterized, molecular features distinguishing leading vs. lagging migrants remain poorly understood. Work from our lab demonstrated that early in migration, PGCs receive non-canonical WNT (ncWNT) signaling which suppresses their proliferation within the hindgut. As they leave the hindgut and progress into the mesentery and gonad, they display progressively greater canonical WNT signaling and proliferate more rapidly, reaching maximal proliferation within the gonad[11]. Epigenetic reprogramming occurs concurrently with PGC migration from E9.5-E11.5, during which global DNA demethylation erases genomic imprints and enables later upregulation of gametogenesis and meiosis-specific genes[12–14]. It remains unknown how shifting niche interactions throughout migration may alter PGC transcriptomes and epigenetic reprogramming, or how these influences might differentially affect earlier vs. later migrants.

Investigating transcriptional differences within the PGC population during migration has been challenging as the cell population at migratory timepoints starts out quite small with just ~200 cells present at E9.5[15], ~1,000 cells present at E10.5[16], and ~2600 cells present at E11.5[17]. Other groups have harnessed single-cell RNA sequencing to investigate germ cell development[18–25]; however, these studies have precluded in-depth study of migratory PGC behavior and transcriptional dynamics within single timepoints due to the relatively small numbers of migratory PGCs in their respective datasets.

By pooling many somite-matched embryos, we successfully profiled large populations of PGCs from E9.5, E10.5, and E11.5 embryos, separating anterior and posterior embryo halves in order to isolate leading from lagging migrants and facilitate direct transcriptional comparisons between these populations. We also profiled representative cells from PGC niches along the migratory path. By investigating cell-cell communication pathways between migrating PGCs and their somatic niches, we uncover interactions that influence PGC differentiation over the course of migration.

## RESULTS

### A comprehensive positional survey of migrating primordial germ cells and their niches

To ascertain the transcriptional dynamics within mouse PGCs across the migratory time period, we performed droplet-based single cell RNA sequencing to create libraries from E9.5, E10.5, and E11.5, to span the onset of directed migration [9] and homing of PGCs into the gonadal ridges (Fig 1A). Since PGCs are few in number at early migratory timepoints, we used the Oct4-ΔPE-eGFP reporter (MGI:3057158[26]) to purify PGCs from multiple age-matched embryos via FACS sorting (see Methods). We constructed sequencing libraries from bisected E9.5 and E10.5 embryos (see Methods) to facilitate comparisons between the transcriptomes of

leading (anterior) and lagging (posterior) migrants (Fig 1A, first two panels). At E11.5, 90-95% of PGCs have arrived at the forming gonadal ridge[17] with some remaining in the midline between the gonadal ridges (Fig 1A, third panel). We purified GFP+ PGCs from the aorta-gonad-mesonephros (AGM) to enrich both for successful migrants as well as any PGCs remaining in the midline and mesentery proximal to the gonads. To survey the supporting somatic cell niches at each migratory timepoint and anatomical position, we included GFP-negative somatic cells from the same sort.

In total, we profiled 21,205 cells (13,262 PGCs and 7,943 somatic cells) from all migratory timepoints (Fig 1B). Results were visualized by Uniform Manifold Approximation and Projection (UMAP) embeddings. We saw clear expression of early PGC markers *Pou5f1* and *Dppa3 (Stella)* among E9.5 and E10.5 PGCs and increasing expression of *Dazl* with PGC developmental stage (Fig 1C). Using automated cell type identification from previously published datasets (see Methods), we identified the following somatic populations: coelomic epithelium, bipotential early gonadal supporting cells, neural tube, hindgut epithelium, primitive erythroid lineage, endothelial cells, neural progenitors, epithelium other, schwann cell precursors, white blood cells, cholinergic neurons, myocytes, and sensory neurons (Fig 1D).

To validate the positional separation of the cells included in the anterior vs. posterior libraries, we examined genes with established roles in caudal patterning of the embryo. As expected, Wnt5a, Wnt5b, Hoxb1, Rarg, Lfng, and T [27] were highly enriched in somatic cells of our posterior E9.5 sample (Fig 1E). At E10.5, Hoxd10 expression is known to span somites 31-41 [28,29] and was higher in somatic cells of our posterior libraries, which contained tissue from somites greater than 21. Hoxb4 expression has been previously described from somite 6 to 41 [28], and was present in both anterior and posterior somatic libraries at E10.5, while expression of Hoxd10 was mostly restricted to the posterior sample (Fig 1E). This analysis supports that anterior and posterior tissues were successfully segregated using our embryo splitting strategy.

**Transcriptomic shifts over developmental time in migratory and post-migratory primordial germ cells**

All together, we examined 1,268 PGCs at E9.5, 1,664 PGCs at E10.5, and 10,330 PGCs at E11.5. This large number of PGCs permitted assessment of migratory PGC heterogeneity across developmental time. After pairwise differential expression testing among PGCs at each developmental timepoint, we conducted Gene Ontology (GO) or Gene Set Enrichment Analysis (GSEA) on the resultant lists of differentially expressed genes. E9.5 PGC migrants exhibited an elevated signature of regulation of canonical WNT signaling relative to E10.5 (Fig S1A). This ontology term result does not indicate the direction of such regulation, but may be consistent with prior findings that WNT signaling regulates proliferative capacity of migrating PGCs[11], with increased canonical WNT signaling promoting PGC proliferation later in migration. E9.5 migrants also differentially expressed genes associated with focal adhesion compared to E10.5 (Fig S1B), suggesting that a shift in cytoskeletal dynamics accompanies the behavior change from chiefly migratory to more proliferative over the course of migration. During PGC migration, progressive genome-wide DNA demethylation permits expression and potential transposition of transposable elements [30,31]. Consistent with this, we found piRNA processing was an upregulated GO term at E10.5 compared to E9.5 (Fig S1A), suggesting that expression of genes relevant for piRNA processing increases during PGC migration despite the fact that mature piRNAs are not detected until E13.5[32,33]. Additionally, compared to migrating PGCs at E9.5, post-migratory E11.5 germ cells express more genes associated with GO/GSEA terms related to chromatin remodeling (Fig S1C), consistent with the global demethylation present by E11.5[12], and oxidative phosphorylation (Fig S1D), which may reflect the increased proliferation previously observed within the gonadal niche[11].

Next, we leveraged trajectory inference techniques in CellRank after subsetting the PGC population to include 1,268 cells from each timepoint (Fig 2A); this subsampling avoided potential biases in embedding graph connectedness resulting solely from differences in PGC numbers from each timepoint and we further regressed out cell cycle phase based on S and G2M scores. As expected for this developmental window[16], we observed robust expression of *Pou5f1* across E9.5-E11.5 PGCs (Fig 2B) and successive upregulation of *Dazl* with increasing developmental time (Fig 2C). To better understand likely differentiation trajectories across these key migratory and early post-migratory timepoints, we employed CellRank's pseudotime kernel to build a transition matrix among PGCs. Simulating random walks of 100 randomly selected E9.5 germ cells for 100 steps revealed that E9.5 posterior PGCs primarily reached endpoints among E11.5 PGCs, with some

terminating among E10.5 and others remaining within a terminal state mostly occupied by E9.5 anterior PGCs (Fig S1E). When this was repeated using CellRank's realtime kernel only, which used only timepoint information and transcriptional relationships to build its transition matrix, all E9.5 PGCs reached final positions among E11.5 PGCs (Fig S1F). This discrepancy may reflect differences in differentiation potential of some E9.5 PGCs that end in a terminal state among anterior E9.5 PGCs, but could also result from technical batch effects generated during library preparation. These possible interpretations are further discussed in the Discussion section.

We detected initial and terminal states found in E9.5-11.5 PGC data based on diffusion pseudotime with a root cell within the posterior E9.5 sample and CellRank's pseudotime kernel. The predicted initial state was found among E9.5 PGCs (Fig 2D, green highlight 2) and was also identified as a putative terminal state, We also found one terminal state at E10.5 (red, 3) and one terminal state at E11.5 (orange, 1). Next, we predicted fate probabilities for each cell not assigned to a terminal state to reach each predicted terminal state (Fig 2E). Genes with expression patterns found to be highly correlated with specific terminal states and significantly up or downregulated during a specific transition from a predicted initial state to a terminal state were considered candidate driver genes of that terminal state.

Since terminal state 1 was found within the E11.5 PGCs in the dataset, we hypothesized that cells with high fate probabilities (Fig 2E) for terminal states 1 could represent successful migrants reaching the gonadal ridge and continuing germline development. This is supported by our discovery of various canonical germ cell development genes as genetic drivers for terminal state 1, including *Dazl, Dppa4, Dppa5a, Lhx1,* and *Rhox* genes [34–37] (Fig 2F). We also uncovered driver genes for state 1 relevant for critical germ cell functions at these timepoints, including *Smc1b* which is directly involved with epigenetic reprogramming[38–40] and *Mov10l1* and *Asz1,* which were previously implicated in genome defense against retrotransposons during reprogramming[41–45]. In this analysis, even the developmentally earliest PGCs in the dataset (E9.5 posterior) were assigned high fate probabilities of reaching state 1 (Fig 2E); this supports the possibility of an *in vivo* differentiation pathway linking initial state 3 with terminal state 1 during PGC migration.

To test our findings for conservation between mouse and human embryos, we took advantage of a publicly available dataset containing human (h) PGCs from 4-12 weeks (W), which corresponds developmentally to mouse PGCs from E9.5-E11.5[21] (Fig 2G). Like in their mouse counterparts, *POU5F1* marked hPGCs, and a gonocyte cluster was marked by *DAZL*. We identified somatic populations similar to those found in our mouse dataset through automated annotation (Fig 2H). Next, we explored the expression of genes identified as drivers for terminal state 1 in mouse PGCs. We found that *RHOX1* and *MOV10L1* were both expressed more in the gonocyte cluster compared to the PGC cluster (Fig 2I), corroborating these genes as putative markers of post-migratory PGCs across species.

**Receptor-ligand interactions between PGCs and their somatic niches across time and position**

We next sought to understand how migratory PGCs interact with their somatic niche cells across time and anatomical position throughout their migration. We employed CellChat[46] to assess candidate intercellular signaling networks linking cell types in anterior and posterior positions at E9.5 and E10.5. At E9.5, posterior PGCs remain exclusively within the hindgut, while the most advanced anterior migrants have left the hindgut and are located within mesentery tissues. At E10.5, lagging migrants remain within the hindgut, the majority of PGCs are in the midst of migration in the dorsal mesentery, and the most advanced PGCs are reaching the gonadal ridges[11]. Despite this diversity in tissue niches, at E9.5 and E10.5, all PGCs across anterior and posterior positions received Kit signaling from their adjacent somatic cells, including the hindgut epithelium, coelomic epithelium, neural tube, and endothelial cells (Fig S2A-D). Kit signaling information flow was not found to be significantly different between anterior and posterior migrants at E9.5 or E10.5 (Fig 2A, B). This is consistent with prior reports that PGCs are surrounded by KitL-expressing cells throughout their entire migration [47].

Across several signaling pathways, we identified dramatic cell-cell communication changes between developmental timepoints and posterior and anterior tissue niches. One major finding was significant ncWnt signaling unique to posterior PGC migrants; the source of this ncWNT signaling was the hindgut epithelium

and coelomic epithelium at E9.5 and primarily the coelomic epithelium at E10.5 (Fig 3C). This corroborates earlier experimental work demonstrating ncWNT signaling in the hindgut that suppresses PGC proliferation while in this early migratory niche[11]. The hindgut and coelomic epithelium mediated ncWNT signaling to PGCs was not present in the anterior migrants at E9.5 (data not shown) or E10.5 (Fig 3D, E), suggesting a transient and important[17] developmental signaling niche specific to posterior migrants. In further support, our reanalysis of signaling networks in human migratory PGCs[21] revealed that ncWNT signaling from the hindgut epithelium to early migratory PGCs is conserved in human embryos (Fig 3F).

At E9.5, we found that FGF signaling was significantly enriched in the anterior cells compared to the posterior (Fig 3A). In addition, the source of FGF shifted from the hindgut and neural lineages in the posterior to primarily the coelomic epithelium in the anterior (S2A, B, E). FGF signaling is involved in germ cell homing in zebrafish[48] and promotes survival and proliferation in migrating avian[49] and mouse PGCs[50]. The niche-specific shift in FGF signaling source and intensity may support a shift to greater proliferation following hindgut exit in anterior migrants [11].

Human PGCs have been shown to use nerve fibers as a migration substrate according to one report[51]. Although mouse germ cells have not been shown to colocalize with nerve fibers[52], we identified NCAM as an enriched signaling pathway (Fig 3A) from neural tube, neural progenitors, hindgut epithelium, and coelomic epithelium within the mesentery migratory niche. NCAM signaling to PGCs was only found in anterior migrants at E9.5 (Fig S2F) and was enriched in anterior migrants over posterior at E10.5 (Fig 3A, B).

We also observed Eph/Ephrin signaling as an enriched cell communication pathway in both the human and mouse datasets; Ephrins have established roles in coordination of adhesion and motility[53,54] and were recently identified as relevant for germ cell development in *Xenopus*[55] and mouse[24]. The spatial resolution in our mouse data reveals that posterior E9.5 PGCs exclusively send Ephrin type A signals to the hindgut epithelium (Fig S2G) while receiving Ephrin type B from the coelomic epithelium (Fig S2H); anterior E9.5 PGCs both send and receive Ephrins type A and B to and from the coelomic epithelium (Fig S2G, H). E10.5 posterior PGCs continue to send and receive Ephrin type A signaling to and from the coelomic epithelium, but shift to receive Ephrin type B from endothelial cells and neural progenitors in the mesentery (Fig S2I, J). Notably, E10.5 anterior PGCs located at the final migration station before entering the gonads do not participate in Ephrin A or B signaling as either senders or receivers (Fig S2I, J). Human migratory PGCs receive Ephrins type A and from the mesenchyme and endothelial cells, but also receive Ephrins type B from immune populations (Fig 3G, H); once they reach the gonad as gonocytes, these cells continue to receive ephrin B from the same sources as well as the gonadal soma, but instead send Ephrin type A to epithelial, endothelial, mesenchymal, and immune populations (Fig 3G). These data suggest a role for Ephrins in modulating distinct cell-cell interactions between PGCs and their diverse migratory niches.

**Transcriptomic shifts across anatomical position in E9.5 migratory PGCs**

The E9.5 anterior PGC population was comprised of leading migrants situated in the more developmentally advanced mesentery niche, as well as migrants in the anterior portions of the hindgut. Conversely, the E9.5 posterior PGC population encompassed migrants located mainly within the hindgut, occupying just the initial migratory niche. To identify transcriptomic differences by migratory position, we analyzed E9.5 PGCs alone. Initial clustering analysis revealed strong co-clustering of cells by cell cycle status (Fig S3A); thus, we regressed out cell cycle phase based on S and G2M scores to reveal additional sources of variation. The resultant cell embedding revealed that posterior cells predominantly clustered together, with a transition zone leading to two discrete zones of predominantly anterior cells (Fig 4A).

To probe transcriptional heterogeneity between anterior and posterior migrants, we conducted differential gene expression on pseudobulked anterior vs posterior populations with pseudoreplicates. We found that Thymosin-β4 (Tmsb4x) a globular G-actin-binding protein previously implicated in planar cell polarity[56], and Tpm4, a member of the tropomyosin family of actin binding proteins implicated in cell motility and adhesion[57 58], were both upregulated in posterior PGCs relative to anterior (Fig 4B). Using CellRank's pseudotime kernel on the E9.5 anterior and posterior samples, we identified one initial and four terminal states (Fig 4C) with similar probability (Fig. 4D). Transcripts that differ between anterior and posterior PGCs both decrease in expression

uniformly in inferred differentiation trajectories leading to terminal states among E9.5 anterior PGCs (Fig 4E). GSEA on differentially expressed genes across anatomical position revealed several terms related to cell-matrix interactions relevant for migration (Fig S3B). Glycosaminoglycans and proteoglycans have been previously shown to exhibit location-specific distributions along the PGC migratory path[59]; these data raise the possibility that distinct niche interactions modulate PGC transcriptomes through changes in glycosaminoglycan biosynthesis and degradation (Fig S3), which may in turn modulate migratory behavior across anatomical position.

We hypothesize that E9.5 anterior PGCs modulate expression of their actin polymerization machinery and cell surface-extracellular matrix interaction molecules as they transition from a chiefly migratory to more proliferative phenotype with distinct adhesion properties. Anterior PGCs at this timepoint upregulate Arpc1b, a member of the Arp2/3 complex[60–62], alongside Lgals1, known to promote or inhibit collective cell migration depending on the signaling context by upregulating downstream integrins[63,64] (Fig 4B). Interestingly, these genes exhibit diverging gene expression trends among identified terminal states (Fig 4E). Transcriptomic shifts between leading and lagging migrants extend beyond cytoskeletal regulation and extracellular matrix interactions. Mybl2, a potent pro-proliferative and pro-survival factor[65], is also upregulated in E9.5 anterior migrants, reflecting the shift to increased proliferation as PGCs encounter niches outside the hindgut. Finally, Kpna2, a gene encoding an importin previously shown to be required for spermatogenesis[66], was upregulated in anterior E9.5 PGCs (Fig 4B) and was found as a driver gene for one of four identified terminal states among anterior PGCs (Fig S3C). Kpna2 and Mybl2 also exhibit divergent trends in gene expression across terminal states, suggesting their potential involvement in alternative differentiation trajectories among early migrants.

**Transcriptomic shifts across anatomical position in E10.5 migratory PGCs**

After cell cycle phase regression (Fig S4A), anterior and posterior PGCs clustered mostly with cells of shared positional origin, with some cells appearing to occupy an intermediate between anterior and posterior states (Fig 5A). Differential gene expression analysis of anterior vs. posterior E10.5 PGCs revealed genes previously implicated in fertility relevant to spermatogenesis and oogenesis. Genes upregulated in anterior migrants included *Kpna2*, which remained upregulated at E10.5 to a similar degree as at E9.5, *Zfp42 (Rex1)*, an epigenetic regulator of genomic imprinting[40] and transcription factor necessary for proper differentiation of early and late spermatids[67], and *Mpc1*, one of two mitochondrial pyruvate transporters important for early ovarian folliculogenesis[68] (Fig 5B).

We found *Lefty1* and *Lefty2* among the top differentially expressed genes upregulated in the E10.5 anterior cells compared to posterior (Fig 5B). Lefty mRNA upregulation is a direct response to Nodal signaling[69], so we sought to identify candidate sources of Nodal ligand. Surveying Nodal expression across somatic and germ cells across timepoints revealed strong Nodal expression in PGCs from E10.5 (Fig S4B). Taken together, these results suggest differences in autocrine Nodal signaling across migratory spatial position at E10.5. A known target of Nodal signaling and mitochondrial gene involved in proline synthesis *Pycr2*[70], was also found to be significantly upregulated in anterior PGC migrants. Additionally, after conducting GO analysis on differentially expressed genes upregulated in anterior E10.5 PGCs relative to posterior, *Nodal* signaling terms were enriched (Fig S4C). Nodal signaling was previously implicated in collective cell migration, with increased cell protrusion and internalization movements correlated with high Nodal signaling activity[71] partly through induction of *Pitx2*[72], which was also upregulated in anterior migrants at E10.5 (Fig 5B). A recent preprint also found that Nodal influences cardiac progenitor cell migration by modulating F-actin activity[73]. *Fscn1*, which encodes an actin bundling protein, was significantly more highly expressed in anterior migrants than posterior migrants at E10.5 (Fig 5B). This gene is a direct target gene of Nodal signaling and is required to facilitate nuclear localization of phospho-Smad2 by shuttling internalized Nodal-Activin responsive TGF-beta Type I receptors to early endosomes[74]. In addition, *Nodal* upregulation in late migratory PGCs may be conserved in human embryos, as hPGCs express *NODAL* most highly from 9-10W (Fig S4D).

We next constructed and assessed CellRank trajectories to identify putative drivers of terminal states within the E10.5 timepoint alone. We specified the root cell to originate from the E10.5 posterior population, and found a single initial state composed mostly of posterior cells. We found two terminal states (orange and green, Fig 5F), primarily consisting of anterior cells and one terminal state (blue, Fig 5E) consisting of a mix of anterior

and posterior cells. Plotting terminal state probabilities by position revealed that many posterior PGCs were assigned high fate probabilities (Fig 5D) of reaching terminal state 2, suggesting a commonly-traversed differentiation pathway linking the initial state with terminal state 2. This was supported by the finding that genetic drivers of terminal state 2 included genes involved in PGC epigenetic reprogramming and survival: Tet1[12], Prc1[75], Chd4[39,76], Dnmt1[13], and Zcwpw1[77,78] (Fig S4E) We previously showed that later germ cells exhibiting inappropriately prolonged DNA methylation in the fetal testis are prone to later elimination by apoptosis during sex differentiation[79]; the high representation of reprogramming-related genes as drivers of terminal state 2 therefore suggests that this state is consistent with appropriate reprogramming and possibly future survival in the developing germline. In addition, we found *Mki67* as a driver of terminal state 2 (Fig S4E); as PGCs in more advanced migratory niches have increased proliferative capacity[11], this further supports the interpretation of terminal state 2 as a developmentally advanced cell state undergoing appropriate epigenetic reprogramming. To compare differentiation trajectories leading to the identified terminal states, we plotted trajectory-specific gene expression trends along pseudotime for key genes of interest, including genes involved in Nodal response, epigenetic reprogramming, and those with known fertility phenotypes. We found that Lefty1 and Lefty2 had the steepest upregulation trend within trajectory 2, but were upregulated to a lesser degree or stagnant for the other two states (Fig 5E). Lefty1 and Lefty2 may serve as markers of cells undergoing transcriptomic and epigenomic shifts consistent with differentiation into a putatively “appropriate” terminal state 2.

To test experimentally whether increased Nodal signaling is a hallmark of anterior migrating PGCs, we performed whole-mount immunofluorescence in E10.5 embryos (Fig 5F, H). We found anterior E10.5 PGCs exhibited significantly increased LEFTY1/2 and nuclear phosphorylated SMAD2/3 signal intensity compared to the posterior (Fig 5G, I), suggesting active signaling through TGF-beta type 1 receptors that bind Nodal. These results reveal a role for Nodal signaling during the terminal phases of PGC migration, with anterior E10.5 PGCs exhibiting distinct gene expression patterns and cell-cell signaling pathways compared to posterior PGCs. Our trajectory inference analysis and immunofluorescence staining suggest that enhanced Nodal signaling in anterior PGCs is associated with successful migration and epigenetic reprogramming, providing new insights into the molecular mechanisms driving PGC development.

## DISCUSSION

This study provides a comprehensive view of transcriptional heterogeneity among migratory mouse PGCs and revisits an existing dataset of migratory human PGCs at analogous timepoints. Key innovations of this study include the sheer number of high-quality PGC transcriptomes profiled and separation of PGC populations by their migratory position, enabling direct comparisons between PGCs present at the same developmental timepoints but within distinct tissue niches.

One of the most striking findings of this study is the identification of Nodal signaling upregulation specific to anterior migrants at E10.5, which we experimentally validated with whole-mount immunofluorescence. Lefty1 and Lefty2, both of which are Nodal-responsive inhibitors of Nodal, have been previously identified as part of the early PGC pluripotency network[14] and Nodal signaling has recently been recognized as a key regulator of hPGC-like cell specification *in vitro*[80]. Nodal and its co-receptor Cripto have also been identified as essential for maintaining testicular germ cell pluripotency from E12.5-E14.5 to prevent premature spermatogenic differentiation; conversely, too much Nodal signaling during this period promotes cell invasiveness and tumor formation[81]. At E13.5, *Lefty2* is a marker of putatively developmentally-defective apoptosis-poised testicular germ cells that have not undergone differentiation to prospermatogonia[79]. Importantly, we observed no obvious correlation between a cell’s *Lefty1* or *Lefty2* expression and its Y-chromosome transcript expression in our E10.5 PGC dataset (Fig S4F), suggesting that the niche-specific Nodal signaling signature at E10.5 is unrelated to regulating testis-specific germ cell differentiation.

An additional interesting feature of this finding is that Nodal expression is not differential between anterior and posterior PGCs at E10.5, but Nodal-responsive transcripts *Lefty1, Lefty2, Pitx2, Pycr2,* and *Fscn1* are all clearly upregulated in the anterior (Fig 5B). This pattern could result from a transcriptional response to Nodal

signaling initiated first in leading PGC migrants that persists after expression of the ligand has been downregulated. The Nodal-dependent upregulation of Fscn1 observed in anterior PGC migrants presents a tantalizing link between this signaling shift and changes in actin cytoskeletal dynamics associated with PGC migratory behavior[82–84] as E10.5 anterior migrants reach the nascent gonad. Moreover, *Fscn1's* role in potentiating TGF-beta signaling could reflect late migratory PGCs becoming primed to respond to BMP or Nodal cues once within the gonad in order to promote oogenic or spermatogenic fates, respectively[85].

Interestingly, the initial state identified in trajectory inference analysis across PGCs from all timepoints was also identified as a terminal state (Fig 2D, state 2). Since state 2 is identified among E9.5 PGCs, this may reflect that some E9.5 PGCs are less able to further differentiate to later terminal states; thus, their differentiation trajectories initiate and terminate within state 2. It is known that some mouse PGCs mis-migrate and fail to reach the gonad[86]; PGCs with high fate probabilities for terminal state 2 may include lagging and mis-migrants. In addition, despite the fact that initial and terminal states for this analysis were based on a diffusion pseudotime ordering with a root cell identified within the E9.5 posterior sample, initial state 3 was found among primarily E9.5 anterior cells using CellRank's cell-cell transition matrix. This likely reflects that transcriptional differences between these early migrating PGC populations are relatively subtle.

Several enriched GO terms for upregulated transcripts in posterior E9.5 PGCs relative to anterior were related to heparan sulfate proteoglycan and glycosaminoglycan synthesis and degradation (Fig 4B). The importance of this pathway during PGC migration was further supported by our cell-cell communication analysis. Interestingly, HSPG signaling from the hindgut epithelium and coelomic epithelium in the posterior was enriched relative to anterior at E9.5. By E10.5, HSPG signaling was instead enriched in the anterior, coming primarily from endothelial cells to PGCs (Fig S3D). In a manner similar to Kit signaling, HSPG signaling may shift to surround PGCs as they traverse their diverse migratory niches. HSPG signaling has been previously shown to modulate migratory behavior and survival in Zebrafish PGCs[87], but there has not been thorough investigation of this pathway in mouse or other mammalian systems[88]. Since posterior E9.5 PGCs are mostly located within the hindgut (Fig 1A), the reduction in HSPG signaling in E9.5 anterior migrants may reflect changes to the PGC cytoskeleton and extracellular matrix interactions that permit hindgut exit and enable a shift toward greater proliferation in more anterior migratory niches[11]. If HSPG signaling indeed facilitates directional migration, reestablishment of HSPG signaling in E10.5 anterior PGCs from the endothelium may guide final PGC homing to reach the gonadal ridges.

Considering the differences observed in Ephrin signaling to PGCs between anterior and posterior niches and between mouse and human datasets (Fig 3G,H, S2G, H, I, J), Ephrins likely warrant further investigation for mediating precise control of PGC migratory behavior. Ephrin B signaling mediates germ layer separation in *Xenopus* embryos; a similar mechanism could mediate PGC repulsion from hindgut and mesentery tissues at the appropriate transitions during PGC migration. In our dataset, E10.5 anterior PGCs no longer exhibit enriched Eph/Ephrin signaling; reduced repulsive signaling between PGCs and their surrounding tissues could help PGCs bind and settle into the gonadal ridges as they complete their migration. Close analysis of the particular Eph and Ephrin ligands and receptors expressed in each somatic niche and relevant PGC subsets could help further clarify the relevance of a shifting Ephrin code in migratory PGC development.

Limitations of this study include batch effects between sequencing libraries stemming from separate library preparation procedures on cells from different timepoints and anatomical locations in our mouse dataset. When appropriate, we corrected for batch effects between libraries (see Methods - Batch Correction). Batch effects may have contributed some spurious differentially-expressed genes between timepoints and anatomical positions, leading to false detection of some differentially regulated GO/GSEA terms and cell-cell interaction pathways. Batch effects may explain the discovery of a putative terminal differentiation state within the bulk of E9.5 anterior PGCs, as it seems unlikely that such a large proportion of the PGCs profiled at this timepoint and anatomical position represent an aberrant state that is not conducive to further PGC migration and development.

Because ectopic PGCs have been shown to give rise to germ cell tumors[10], one goal of our analyses was to identify cells with clearly distinct transcriptomes consistent with mis-migration and/or aberrant differentiation.

Aside from the small population of PGCs consisting of mixed anterior and posterior migrants at E10.5 (Fig 5F state 0), we did not identify any minor clusters of obviously transcriptionally distinct PGCs suggesting ectopic localization. It is possible that this population is indeed composed of ectopic PGCs, but their small number precluded identification of specific markers for experimental validation. Moreover, it is possible that at the developmental stages profiled, ectopic PGCs have not yet acquired transcriptional differences reflecting their mislocalization in the embryo. Explicit profiling of ectopic PGCs separately from gonadal PGCs at E11.5 or later could be a fruitful way to establish a ground truth of expected transcriptional differences between successful and unsuccessful migrants to help identify these cells in future datasets and explore differentiation trajectories leading to the formation of germ cell tumors.

Finally, trajectory inference analyses like those in this study are more often performed on clearly divergent cell types within a differentiating lineage rather than within single cell types[89]; thus, the differences between putative differentiation trajectories we observed at E9.5 and E10.5 may be subtler than in other applications of these methods. Overall, the abundance of PGCs profiled in this study has enabled unprecedented comparisons among PGC subtypes within and across developmental timepoints.

## CONCLUSION

In summary, this work provides a transcriptional survey of migrating and early post-migratory mouse and human germ cells. We interrogated migrating PGCs in the context of their somatic niche, and identified known and novel cell-cell interactions which may be involved in regulation of migration. We identified transcriptional differences between PGCs in different phases along the migratory route, connecting their spatial heterogeneity to transcriptional heterogeneity. In the mouse, we assayed transcriptional differences between migratory leaders and laggards at E9.5 and E10.5, identifying increased Nodal signaling in anterior migrants at E10.5. In the human dataset, we characterized temporally specific gene expression patterns and identified key similarities and differences between migratory and post-migratory mouse and human PGCs.

## ACKNOWLEDGEMENTS

The authors would like to acknowledge Dan Bunis, David Wu, Aparna Bhadui, Bikem Soygur, Dan Nguyen, Lina Afonso, and Anjali Prabhu for their support of this work through helpful discussions and contributions to data collection and analysis. We thank the San Francisco Chan Zuckerberg Biohub for sequencing support.

## ARTICLE AND AUTHOR INFORMATION

### Funding

This work was funded by NIH grants 1DP2OD007420 (DJL), 1R01GM122902 (DJL), 1R01ES023297 (DJL), 1F31HD096840-01 (RGJ), as well as the San Francisco Chan-Zuckerberg Biohub (DJL, DEW) and Global Consortium for Reproductive Health through the Bia-Echo Foundation GCRLE-0123 (DJL).

## DISCLOSURES

D.L. is a scientific advisor at Vitra, Inc.

## FIGURE LEGENDS

Figure 1:
A: Confocal microscopy images of whole-mount mouse embryos at E9.5, E10.5, and E11.5 (left to right). PGCs are visualized via the Pou5f1-ΔPE-eGFP transgene (MGI:3057158) stained with anti-GFP antibody (ab13970). P indicates posterior of migratory stream; A indicates anterior. White arrows indicate the general direction of PGC migration by timepoint. B: UMAP plot of all cells profiled in this dataset (from pooled pre-sex specification XY and XX embryos) colored by timepoint arranged using STITCH by timepoint (see methods - Batch correction). C: UMAP plots of all cells colored by log-normalized expression of marker genes relevant in PGC development. D: UMAP plot of all cells profiled in this dataset colored by annotated celltype. E: Matrixplot of genes with anteroposterior domain-specific expression at E9.5 and E10.5.

Figure 2
A: UMAP plot of PGCs downsampled to 1268 cells/timepoint colored by position-time of sample. B: UMAP plot colored by *Pou5f1* gene expression. C: UMAP plot colored by *Dazl* gene expression. D All initial and terminal states predicted from CellRank transition matrix analysis based on diffusion pseudotime plotted on UMAP representation of downsampled PGCs across timepoints. E: Matrixplot of fate probability of reaching identified terminal states by position-time identity of PGCs. F: Putative driver genes with expression dynamics correlated with terminal state 1. Color bar at top of plot indicates position-time of cells most highly expressing driver genes below; see color key in 2A. G: UMAP plot of human PGCs and somatic cells from Li *et al.* 2017 colored by gestational week. H: UMAP plot of human PGCs and somatic cells colored by annotated celltypes. I: Violin plots of log-normalized expression of selected genes with conserved expression patterns between mouse and human.

Figure 3:
A: Comparison of intercellular communication networks enriched in E9.5 anterior cells vs. enriched in E9.5 posterior cells by normalized information flow computed with CellChat. B: Comparison of intercellular communication networks enriched in E10.5 anterior cells vs. enriched in E10.5 posterior cells computed with CellChat. In A and B, the color of the signaling pathway name corresponds to whether it is significantly enriched in either anterior or posterior. Black pathway names are not significantly differentially enriched. Colors correspond to legend. C-H: Cell populations in grey text do not participate significantly in plotted signaling pathway based on CellChat analysis. Arrows are colored according to cell type from which they originate. C: ncWNT signaling pathway network within mouse E9.5 posterior. D: ncWNT signaling pathway network within mouse E10.5 posterior. E: ncWNT signaling pathway network within mouse E10.5 anterior. F: Human ncWNT signaling pathway network. G. Human Ephrin A signaling pathway network. H. Human Ephrin B signaling pathway network.

Figure 4:
A: UMAP plot of E9.5 PGCs colored by anteroposterior position. B: Matrixplot of expression of selected differentially expressed genes between anterior and posterior E9.5 PGCs. C: All initial and terminal states predicted from CellRank transition matrix analysis computed with diffusion pseudotime plotted on UMAP representation of E9.5 PGCs. D: Matrixplot of fate probability of reaching identified terminal states by position identity of PGCs at E9.5. E: Gene expression trends between initial state 3 and the four identified terminal states for selected genes.

Figure 5
A: UMAP plot of E10.5 PGCs colored by anteroposterior position. B: Matrixplot of expression of selected differentially expressed genes between anterior and posterior E10.5 PGCs. C: All initial and terminal states predicted from CellRank transition matrix analysis computed with diffusion pseudotime plotted on UMAP representation of E10.5 PGCs. D. Matrixplot of fate probability of reaching identified terminal states by position identity of PGCs at E10.5. E: Gene expression trends between initial state 3 and the four identified terminal states for selected genes. Diamonds correspond to functional relevance of selected genes. F: Representative image of whole-mount immunofluorescence staining of Oct4-ΔPE-eGFP (Oct4GFP), and LEFTY1/2. Fluorescence from Oct4GFP reporter signal is amplified by anti-GFP antibody (ab13970) stain. For both the

whole image and zoom in portion, we selected slices for the maximum projection based on the presence of GFP signal. G: Quantification of LEFTY1/2 signal intensity of PGCs in n=6 embryos, one of which is shown in F. H: Representative image of whole-mount immunofluorescence staining of Oct4GFP and pSMAD2/3. Fluorescence from Oct4-ΔPE-eGFP reporter signal is amplified by anti-GFP antibody (ab13970) stain. For both the whole image and zoom in portion, we selected slices for the maximum projection based on the presence of GFP signal. I: Quantification of pSMAD2/3 signal intensity of PGCs in n=3 embryos, one of which is shown in F. P values for all quantifications were conducted using a paired Wilcoxon signed-rank test with BH adjustment in R.

Figure 6:
Schematic representation of PGC migration during migratory timepoints. PGCs migrate through the Hindgut, Dorsal Mesentery, and arrive in the Gonadal Ridge, encountering diverse signaling niches. Multiple signaling pathways, including Nodal/Lefty, ncWnt, and survival/homing factors such as Kit, regulate PGC proliferation and migration. Seminal literature references supporting key aspects of the migratory model are listed[90,91], including contributions from this study.

**SUPPLEMENTARY FIGURE LEGENDS**

Figure S1:
A-B: Barplot of enriched terms from gene ontology analysis of differentially expressed genes between E9.5 and E10.5 PGCs. C-D: Barplot of enriched terms from gene ontology analysis of differentially expressed genes between E9.5 and E11.5 PGCs. E: Visualization of random walks through CellRank transition matrix computed with diffusion pseudotime of 100 cells randomly selected from the E9.5 posterior sample stepping through 100 time steps. Black dots indicate starting positions, yellow dots indicate final positions, and lines between dots connect the positions of subsequent steps along the random walk, colored from violet (early) to orange (late). E-F: Visualization of random walks through CellRank transition matrix computed with experimental time of 100 cells randomly selected from the E9.5 timepoint stepping through 100 time steps. Black dots indicate starting positions, yellow dots indicate final positions, and lines between dots connect the positions of subsequent steps along the random walk, colored from violet (early) to orange (late). G. Violin plots of log-normalized expression levels of key genes of interest identified from mouse migratory PGC dataset in PGC, gonocyte, gonadal soma, and epithelial cell populations from Li et al. 2017 human dataset.

Figure S2:
A-D: Incoming signaling networks between somatic populations (x axis) and PGCs across developmental time and position. E) FGF signaling during E9.5 PGC migration is spatially distinct, with anterior cells receiving more FGF signal from the neural tube. F) NCAM signaling during E9.5 PGC migration is spatially distinct, with anterior cells receiving signal from most somatic niches. G) EPHA signaling during E9.5 PGC migration is spatially distinct, with anterior cells receiving signal from most somatic niches as well as cell intrinsic signal. H) EPHB signaling during E9.5 PGC migration is not spatially distinct, with anterior cells and posterior cells receiving signal from most somatic niches. I) EPHA signaling during E10.5 PGC migration is spatially distinct, with posterior cells exchanging signals most strongly with the coelomic epithelium. J) EPHB signaling during E10.5 PGC migration is spatially distinct, with posterior cells receiving signal from endothelial cells and neural progenitors.

Figure S3:
A: E9.5 mouse PGC samples before and after cell cycle phase regression labeled by predicted cell cycle phase. B: Barplot of enriched terms from gene ontology analysis of differentially expressed genes between E9.5 anterior and posterior migrating PGCs. C: Putative driver genes with expression dynamics correlated with terminal state 1.

Figure S4:
A: E10.5 mouse PGC samples before and after cell cycle phase regression labeled by predicted cell cycle phase. B: Nodal expression is strongest in E10.5 PGCs. Top, UMAP plot showing Nodal gene expression among E10.5 PGCs and not from other cells. Bottom, violin plot of log-normalized Nodal expression in PGCs

at each surveyed timepoint. C: Barplot of enriched terms from gene ontology analysis of differentially expressed genes between E10.5 anterior and posterior migrating PGCs. D. Expression of log-normalized NODAL mRNA across germ cells by timepoint in Li *et al.* 2017 human dataset. E: Putative driver genes with expression dynamics correlated with terminal state 2. F) Correlation plot of Lefty1/2 and Nodal genes with Y-chromosome genes, showing little correlation between Lefty/Nodal expression and Y-chromosome genes. G) Paired mean intensity of LEFTY1/2 and pSMAD2/3 expression in E10.5 anterior and posterior embryos.

## MATERIALS AND METHODS

### Animals

CD1 females were crossed to mixed background CD1/C57Bl6 males homozygous for Pou5f1-ΔPE-eGFP (MGI:3057158) for both tissue dissociations for single-cell RNA sequencing and whole-mount immunofluorescence validation experiments. Counting plug date as 0.5 days post conception, pregnant mice were euthanized and age matched embryos were dissected in 0.4% bovine serum albumin (BSA) in 1x phosphate-buffered saline (PBS) at E9.5, E10.5, and E11.5. Embryos were further staged by somite number, to acquire a tighter, more standardized range of developmental time for included embryos. Embryos of both sexes were pooled without genotyping, as the timepoints analyzed were prior to sex specification. All animal work was performed under strict adherence to the guidelines and protocols set forth by the University of California San Francisco's Institutional Animal Care and Use Committee (IACUC), and all experiments were performed in an animal facility approved by the Association for the Assessment and Accreditation of Laboratory Animal Care International (AAALAC). All procedures performed were approved by the UCSF IACUC. All mice were maintained in a temperature-controlled animal facility with 12 hr light dark cycles, and were given access to food and water.

### Tissue dissociation

*E9.5 dissociation*

At E9.5, 5 litters were dissected and embryos with total somite counts of 20-25 were kept. To split PGCs and somatic cells into Anterior and Posterior Bins (containing migratory leaders and laggards, respectively), embryos were bisected at somite 15, corresponding to where the PGCs move from the hindgut into the dorsal mesentery. (Fig1A, dotted line). All tissue rostral to the forelimb was removed and the pooled anterior and posterior segments were bathed in 200uL of pre-warmed 0.25% Trypsin/ETDA, and incubated at 37°C. After 10 minutes, the tissue was triturated with a pipet tip enlarged by cutting with a razor blade. If after 10 minutes, the digest still looked clumpy, we incubated another 10 minutes in the 200uL 0.25% Trypsin/ETDA. Next, 50uL of 1mg/ml DNAse was added, samples triturated gently, and incubated for 5 minutes at 37C. After 5 minutes, samples were checked for consistency. If they were thin, watery, and homogenous, digests were quenched with a 1:1 volume of FBS. If samples were still viscous, 50uL more DNAse was added and samples were incubated for 3 more minutes at 37C, after which samples were quenched with a 1:1 volume of FBS and placed on ice.

*E10.5 dissociation*

At E10.5, 3 litters were dissected and embryos with total somite counts of 34-38 were kept. To further split our E10.5 embryos into an Anterior Bin and Posterior Bin (containing migratory leaders and laggards, respectively), we did the split right below where the PGCs have begun to bifurcate to settle into the left or right gonadal ridge, between somite 19 and 20 (Fig1A, dotted line). Dissected embryos were cut in half just below the heart, and the posterior half of the embryo was dissociated to prepare for FACS germ cell concentration (GFP- heads were used as GFP- control tube). Embryos posterior halves were split into groups of 10, and placed in a 37C water bath with 400uL of pre-warmed 0.25% Trypsin/ETDA. An additional 200uL of pre-warmed 0.25% Trypsin/ETDA was added if the solution appeared too thick. Embyos were digested at 37C for 15 more minutes, and digest progress aided by flicking and assessing solution viscosity every 5 minutes. 200uL of 1mg/ml DNAse was added, samples triturated with a cut pipette (see E9.5 for description), and then incubated for 5mins in 37C water bath. After a 5 minute digest, samples were assessed for remaining clumps; if clumps remained, an additional 5mins or additional 200uL of 1mg/ml DNAse was added. Once the solution pipetted homogenously and water-like, it was quenched with a volume matched amount of FBS and placed on ice.

*E11.5 AGM dissociation*

At 11.5, 4 litters were dissected; AGMs of age matched embryos (45-49 somite range) were microdissected out from posterior embryos. AGMs were grouped in sets of 6, and placed in a 37C water bath. GFP- heads were also processed as a control. 500uL of 0.25% Trypsin/EDTA was added to embryos. Tubes were incubated for 20 minutes, triturating with a cut pipet tip (see E9.5 for description) at every 5 minute interval to gently mix the solution. After 20 minutes, 100uL of 1mg/ml DNase was added to each tube, and tubes were

mixed by flicking. After 5 minutes of incubation with DNAse added, solutions were checked for homogeneity. Once the solution was water-like, it was quenched with 1:1 volume of FBS and placed on ice. For the first biological replicate, 23 AGMs were dissociated. For the second replicate, 12 AGMs were dissociated.

**PGC enrichment**

Once on ice, each sample was filtered through a 35 micron strainer. Tip and tube and filter were rinsed with a small amount of 0.1% BSA to wash all possible GFP+ PGCs into the cell suspension. Sytox Blue live-dead indicator was added at the appropriate 1:1000 concentration to each digest. Samples were sorted on either a BD FACS Aria2, Aria3, or Aria Fusion, following a gating pattern as follows based on established methods using the Pou5f1-ΔPE-eGFP mouse strain[92] : 1) side scatter area vs forward scatter area to select for cells; 2) side scatter width vs side scatter height to select for cell singlets; 3) forward scatter width vs forward scatter height to select again for cell singlets; 4) Sytox Blue vs forward scatter area to select for Sytox Blue negative live cells; and finally 5) GFP vs forward scatter area. The range of GFP negative vs positive gates was calibrated each experiment based on the GFP- digest sample, to put GFP- cells in the $10^3$ range and GFP+ cells in the $10^4$ to $10^5$ range. GFP+ cells were collected in 500uL of 0.2% BSA, and GFP- somatic cells were also collected in 500uL of 0.2% BSA. For each GPF+ library, we spiked back in somatic cells at a 1:1 ratio. We concentrated cells into a 300-1500 cell/uL range, resuspending in 0.1% BSA.

**Single-cell RNA sequencing library construction**

We pooled sorted GFP+ germ cells and an equal amount of GFP- somatic cells from each timepoint and anatomical location. We loaded counted cells onto a 10X V2 chip to create a single cell emulsion. For E9.5 Anterior, we loaded 3400 cells, and for E9.5 Posterior, 2800 cells. For E10.5 Anterior, we loaded two technical replicates totaling 9300 cells. For E10.5 Posterior, we loaded two technical replicates totaling 11200 cells. For the two E11.5 biological replicates, we loaded 15,000 and 10,000 cells, respectively. Library creation was done in house following 10X ChromiumTM Single Cell 3' Reagent Kits v2 User Guide Rev A. Briefly, we used Single-Cell 3′ Reagent Version 2 Kit (10X Genomics) and used fluidics to create individual cells in gel bead-in-emulsions (GEMS). Samples were prepared directly according to the protocol, with the following steps tailored to maximize library yield: at the cDNA amplification step, we used 14 cycles, and at the sample index PCR step, we used 14 cycles. Samples were tested for quality on an Agilent Bioanalyzer High Sensitivity chip before sequencing. E9.5 and E11.5 libraries were sequenced on HiSeq 4000 (Illumina) with paired-end sequencing parameters: Read1, 98 cycles; Index1, 14 cycles; Index2, 8 cycles; and Read2, 10 cycles. E10.5 libraries were sequenced on Novaseq (Illumina) with paired-end sequencing parameters: Read1, 150 cycles; i7 Index, 8 cycles; i5 index, 0 cycles; and Read2, 150 cycles.

**Whole-mount immunofluorescence**

Embryos from CD1 dams crossed to sires homozygous for Pou5f1-ΔPE-eGFP (MGI:3057158) were euthanized at E9.5 (20-25 somites) or E10.5 (34-38 somites) and then immunostained according to the iDISCO protocol[93] with minor modifications. Briefly, embryos were fixed in 4% PFA overnight at 4°C overnight, blocked in 0.2% Gelatin, 0.5% Triton X-100 in 1XPBS overnight at room temperature, incubated with primary antibodies prepared in the blocking solution for 10 nights at 4°C, and incubated with secondary antibodies in the blocking solution for 2 nights at 4°C. Samples were washed in 0.2% Gelatin, 0.5% Triton X-100 at least 6 times for 30 minutes each at room temperature between incubation steps. Then samples were cleared with 50% Tetrahydrofuran (THF):dH2O overnight at room temperature, 80% THF:dH2O and 100% THF for 1.5 hours each at room temperature, Dichloromethane (DCM) for 30 minutes at room temperature, and finally Dibenzyl ether (DBE) for a minimum of 3 nights at room temperature until ready for imaging. For whole-mount immunofluorescence staining with primary antibody for pSMAD2/3, antigen retrieval was performed with 100% acetone for 1 hour at -20°C immediately before blocking. The following antibodies were used at the specified dilutions: Chicken anti-GFP (cat ab13970) 1:200, Goat anti-LEFTY1/2 (cat AF746) 1:200, Rabbit pSMAD2/3 (cat D27F4) 1:200, Donkey anti-Rat AlexaFluor 405 (cat A48268)1:200, Donkey anti-Chicken AlexaFluor 488 (cat 703-546-155) 1:200, Donkey anti-Rabbit AlexaFluor 555 (cat A31572) 1:200, Donkey anti-Goat AlexaFluor 647 (cat A21447)1:200.

**Imaging**

All imaging was performed on the Leica TCS SP8 inverted scanning confocal microscope with 2 micron Z-steps.

**Image Analysis**
Image analysis to measure fluorescence intensity of LEFTY1/2 and pSMAD2/3 in E10.5 embryos was performed in Imaris software by masking on GFP signal intensity to identify PGCs expressing the Pou5f1-ΔPE-eGFP reporter and using automatic spot detection to call a 10-micron diameter spot for each PGC checked with manual correction. Then, identified PGC spots were characterized by their location relative to a line extending ventrally from the junction of somites 19 and 20; those anterior to this location were called anterior and posterior were called posterior. Then, the max signal intensities for LEFTY1/2 or pSMAD2/3 signal were recorded, plotted, and compared with an unequal variances t-test (Welch's t-test) in Prism.

**Data processing**

**SoupX Decontamination**
When aberrant hemoglobin gene expression in non-red blood cells was discovered, data was processed through SoupX (version 1.5.2) [94] to remove ambient RNA contamination. Contamination parameters were either set by the function autoEstCont to automatically estimate ambient RNA contamination, or set by slowly increasing the threshold of filtering until Hbb* genes ceased to inappropriately be found as markers for non-red blood cell genes, as follows: E9.5 Anterior, 12%; E9.5 Posterior, autoEstCont; E10.5 Anterior, 19%, E10.5 Posterior, autoEstCont.

Initial processing of the v2 10x libraries was done through CellRanger v3.1.1. Libraries for all timepoints were aligned to the mouse genome (10× Genomics prebuilt mm10 version 3.0.0 reference genome) with all cellranger default parameters for demultiplexing and aligning. Our libraries were sequenced at a depth of 125,100 mean reads per cell for the E9.5 Anterior sample and 111,258 mean reads per cell for the E9.5 posterior sample. At E10.5, our anterior technical replicates 1 and 2 were 493,803 and 399,673, and our posterior replicates 3 and 4 492,512 and 556,014. At E11.5, our first biologic replicate was 15,542 mean reads per cell, and the second biologic replicate 20,967. The resulting gene by cell matrices were analyzed with the python package ScanPy v1.9.3 [95]. Cells were filtered to retain only high quality cells with a minimum of 2000 transcripts per cell barcode, a maximum of 10% mitochondrial transcripts, and a doublet score of 0.15 computed with Scrublet v0.2.3 [96]. Further downstream, any remaining cell populations lacking unique marker genes or with a high density of doublet scores near the 0.15 doublet score cutoff were also removed.

To reanalyze the human datasets, 4-12W count files were downloaded from GSE86146[21]. Individual timepoints contained too few cells for Seurat v3.2.3[97] to integrate effectively, therefore, male and female matrices from individual timepoints were concatenated, and the two sex datasets merged via Seurat's integration anchors based approach. Briefly, after running NormalizeData and running FindVariableFeatures to select 2000 variable genes, FindIntegrationAnchors was run (dims 1:30). ScaleData, RunPCA, and RunUMAP were then run with all default flags.

**Batch correction**
To faithfully annotate cell types across time points, anatomical locations and sequencing libraries, libraries were first integrated using harmonypy v0.0.9, generating a manifold that most closely co-clusters cells with transcriptomic similarity across batches to yield consistent cell type identification across sequencing libraries. To integrate cells within timepoints but allow for visual comparison of similarities and differences between primordial germ cells on the full cell manifold (Fig 1C), we used STITCH[98], which joins the k-nearest neighbor graphs from multiple timepoints by identifying neighbors in adjacent timepoints and projecting cells from a timepoint $t_i$ into the next timepoint $t_{i+1}$. To handle batch variability across libraries for different timepoints, we also provided the library_id as a batch_key in order to prioritize highly variable genes identified across batches rather than batch-specific highly variable genes for use in UMAP graph construction. This light batch correction using highly variable genes was also used to build the PGC-only manifold in Figure 2.

**Cell type identification**

Of the cells we sequenced, 1,268 E9.5 PGCs, 1,664 E10.5 PGCs, and 10,330 E11.5 PGCs survived quality control filtering steps. We acquired an average of 3,724 genes per cell at E9.5, 2,076 genes per cell at E10.5, and 4,115 genes per cell at E11.5. We next began to cluster the data and identify the germ cell and somatic cell components present. Automated cell type label transfer was performed in Celltypist (v1.6.2) using published datasets[25,99] with good coverage of gonadal and extragonadal cell types at equivalent developmental stages (E9.5-E11.5). Automated cell type predictions were verified through manual gene set scoring. Clustering resolution was varied to best capture the distribution of predicted cell types in the dataset. While the limited cell number of PGCs present in the tissue at E9.5 did not permit us to construct multiple libraries from multiple biologic replicates, we prepared the E10.5 anterior and posterior libraries as well as the E11.5 library in biological duplicate. Automated cell type identification was performed on the Li *et al.* 2017 human dataset using Celltypist's Developing_Human_Gonads model based on Garcia-Alonso *et al.* 2022.

**Differential Expression Testing**

Differential expression testing was performed using pyDESeq2 (v0.4.4) and pseudobulking cells by sequencing library of origin. In cases with fewer than three libraries per timepoint or anatomical position, 3 pseudoreplicates were generated and cells were randomly allocated to one of these three pseudoreplicates, then gene expression data were pseudobulked by pseudoreplicate. We compared gene expression using the Wald test and corrected for multiple comparisons via the Benjamini-Hochberg procedure.

**Gene Set Enrichment Analysis**

Gene set enrichment analysis was performed with GSEApy (v1.1.2) on differentially expressed genes from pyDESeq2 (see Differential Expression Testing) filtered to include genes with a log2 fold change ≥0.5 or ≤-0.5 and an adjusted p value of ≤0.01. Gene sets used for enrichment analysis were acquired from the Molecular Signatures Database (MSigDB).

**Trajectory Inference Methods**

We used CellRank (v 2.0.2)[89] to analyze differentiation trajectories across and within timepoints. We began by using CellRank's realtime kernel with germ cells from E9.5-E11.5 to generate a cell-cell transition matrix incorporating transcriptional similarities among germ cells as well as their developmental stage, subsampling the data to contain the same number of germ cells from each timepoint to reduce biases from cell number. Assigning E9.5 Posterior to contain the starting state, we plotted random walks through the transition matrix to identify the most likely endpoints of differentiation. Then, using diffusion pseudotime computed in ScanPy for cells across all timepoints in CellRank's pseudotime kernel, we identified putative initial and terminal macrostates within migratory PGCs. For all trajectory analyses, the number of macrostates used was the lowest possible such that the CellRank algorithm identified an initial state containing the root cell used for pseudotemporal ordering. Each cell not assigned to a macrostate was scored on its likelihood of reaching a predicted terminal state using Markov chain absorption probabilities, which approximate the results of infinite random walks through the state manifold. For intra-timepoint trajectory analyses at E9.5 and E10.5, we used CellRank's pseudotime kernel with diffusion pseudotime computed in ScanPy to build cell-cell transition matrices. Intra-timepoint initial and terminal states as well as random walks were also computed within CellRank's pseudotime kernel based on diffusion pseudotime.

**Cell-cell Communication Analysis**

All cell-cell communication network analysis was performed in Cellchat (v2.1.2)[46]. Enriched cell-cell communication networks were initially inferred from somatic and germ cells derived from each timepoint and anatomical position separately (E9.5 posterior, E9.5 anterior, E10.5 posterior, and E10.5 anterior), and then CellChat objects were compared via cross-dataset analysis. Since initial cell type annotations were performed on the integrated cell manifold across all cells from E9.5-E11.5, subsetted embeddings for each position and timepoint were checked for annotation accuracy and cluster annotations were slightly modified as needed prior to CellChat analysis.

**Data Availability**

Raw and processed transcriptomic data generated in this study have been deposited in the Gene Expression Omnibus (GEO) under the accession code GSE274603. Code used to analyze the data is located at github.com/LairdLabUCSF/Migratory-PGC-Atlas. Additionally, an R Shiny app to facilitate quick exploration of the mouse single cell data is located at: rebeccagj.shinyapps.io/MousePGCAtlas.

**Bibliography**


1. Lawson, K. A. & Hage, W. J. Clonal analysis of the origin of primordial germ cells in the mouse. *Ciba Found. Symp.* **182**, 68–84; discussion 84 (1994).
2. Anderson, R., Copeland, T. K., Schöler, H., Heasman, J. & Wylie, C. The onset of germ cell migration in the mouse embryo. *Mech. Dev.* **91**, 61–68 (2000).
3. Canovas, S., Campos, R., Aguilar, E. & Cibelli, J. B. Progress towards human primordial germ cell specification in vitro. *Mol. Hum. Reprod.* **23**, 4–15 (2017).
4. Ginsburg, M., Snow, M. H. & McLaren, A. Primordial germ cells in the mouse embryo during gastrulation. *Development* **110**, 521–528 (1990).
5. Lawson, K. A. *et al.* Bmp4 is required for the generation of primordial germ cells in the mouse embryo. *Genes Dev.* **13**, 424–436 (1999).
6. Ying, Y., Liu, X. M., Marble, A., Lawson, K. A. & Zhao, G. Q. Requirement of Bmp8b for the generation of primordial germ cells in the mouse. *Mol. Endocrinol.* **14**, 1053–1063 (2000).
7. Pesce, M., Farrace, M. G., Piacentini, M., Dolci, S. & De Felici, M. Stem cell factor and leukemia inhibitory factor promote primordial germ cell survival by suppressing programmed cell death (apoptosis). *Development* **118**, 1089–1094 (1993).
8. Farini, D., La Sala, G., Tedesco, M. & De Felici, M. Chemoattractant action and molecular signaling pathways of Kit ligand on mouse primordial germ cells. *Dev. Biol.* **306**, 572–583 (2007).
9. Molyneaux, K. A., Stallock, J., Schaible, K. & Wylie, C. Time-lapse analysis of living mouse germ cell migration. *Dev. Biol.* **240**, 488–498 (2001).
10. Runyan, C., Gu, Y., Shoemaker, A., Looijenga, L. & Wylie, C. The distribution and behavior of extragonadal primordial germ cells in Bax mutant mice suggest a novel origin for sacrococcygeal germ cell tumors. *Int. J. Dev. Biol.* **52**, 333–344 (2008).
11. Cantú, A. V., Altshuler-Keylin, S. & Laird, D. J. Discrete somatic niches coordinate proliferation and migration of primordial germ cells via Wnt signaling. *J. Cell Biol.* **214**, 215–229 (2016).
12. Hill, P. W. S. *et al.* Epigenetic reprogramming enables the transition from primordial germ cell to gonocyte. *Nature* **555**, 392–396 (2018).

13. Hargan-Calvopina, J. *et al.* Stage-Specific Demethylation in Primordial Germ Cells Safeguards against Precocious Differentiation. *Dev. Cell* **39**, 75–86 (2016).
14. Seisenberger, S. *et al.* The dynamics of genome-wide DNA methylation reprogramming in mouse primordial germ cells. *Mol. Cell* **48**, 849–862 (2012).
15. Durcova-Hills, G., Wianny, F., Merriman, J., Zernicka-Goetz, M. & McLaren, A. Developmental fate of embryonic germ cells (EGCs), in vivo and in vitro. *Differentiation.* **71**, 135–141 (2003).
16. Saiti, D. & Lacham-Kaplan, O. Mouse Germ Cell Development in-vivo and in-vitro. *Biomark Insights* **2**, 241–252 (2007).
17. Laird, D. J., Altshuler-Keylin, S., Kissner, M. D., Zhou, X. & Anderson, K. V. Ror2 enhances polarity and directional migration of primordial germ cells. *PLoS Genet.* **7**, e1002428 (2011).
18. Stévant, I. *et al.* Dissecting Cell Lineage Specification and Sex Fate Determination in Gonadal Somatic Cells Using Single-Cell Transcriptomics. *Cell Rep.* **26**, 3272-3283.e3 (2019).
19. Hermann, B. P. *et al.* The Mammalian Spermatogenesis Single-Cell Transcriptome, from Spermatogonial Stem Cells to Spermatids. *Cell Rep.* **25**, 1650-1667.e8 (2018).
20. Wang, J.-J. *et al.* Single-cell transcriptome landscape of ovarian cells during primordial follicle assembly in mice. *PLoS Biol.* **18**, e3001025 (2020).
21. Li, L. *et al.* Single-Cell RNA-Seq Analysis Maps Development of Human Germline Cells and Gonadal Niche Interactions. *Cell Stem Cell* **20**, 858-873.e4 (2017).
22. Mayère, C. *et al.* Single-cell transcriptomics reveals temporal dynamics of critical regulators of germ cell fate during mouse sex determination. *BioRxiv* (2019) doi:10.1101/747279.
23. Zhao, Z.-H. *et al.* Single-cell RNA sequencing reveals the landscape of early female germ cell development. *FASEB J.* **34**, 12634–12645 (2020).
24. Alexander, A. K. *et al.* Single-nucleus multiomics reveals the gene-regulatory networks underlying sex determination of murine primordial germ cells. (2024) doi:10.7554/eLife.96591.1.
25. Garcia-Alonso, L. *et al.* Single-cell roadmap of human gonadal development. *Nature* **607**, 540–547 (2022).
26. Szabó, P. E., Hübner, K., Schöler, H. & Mann, J. R. Allele-specific expression of imprinted genes in mouse migratory primordial germ cells. *Mech. Dev.* **115**, 157–160 (2002).

27. Gofflot, F., Hall, M. & Morriss-Kay, G. M. Genetic patterning of the developing mouse tail at the time of posterior neuropore closure. *Dev. Dyn.* **210**, 431–445 (1997).

28. Burke, A. C., Nelson, C. E., Morgan, B. A. & Tabin, C. Hox genes and the evolution of vertebrate axial morphology. *Development* **121**, 333–346 (1995).

29. Wellik, D. M. Hox patterning of the vertebrate axial skeleton. *Dev. Dyn.* **236**, 2454–2463 (2007).

30. Brennecke, J. *et al.* Discrete small RNA-generating loci as master regulators of transposon activity in Drosophila. *Cell* **128**, 1089–1103 (2007).

31. Teixeira, F. K. *et al.* piRNA-mediated regulation of transposon alternative splicing in the soma and germ line. *Nature* **552**, 268–272 (2017).

32. Ernst, C., Odom, D. T. & Kutter, C. The emergence of piRNAs against transposon invasion to preserve mammalian genome integrity. *Nat. Commun.* **8**, 1411 (2017).

33. Ramakrishna, N. B., Battistoni, G., Surani, M. A., Hannon, G. J. & Miska, E. A. Mouse primordial germ-cell-like cells lack piRNAs. *Dev. Cell* **57**, 2661-2668.e5 (2022).

34. Lin, Y. & Page, D. C. Dazl deficiency leads to embryonic arrest of germ cell development in XY C57BL/6 mice. *Dev. Biol.* **288**, 309–316 (2005).

35. Madan, B. *et al.* The pluripotency-associated gene Dppa4 is dispensable for embryonic stem cell identity and germ cell development but essential for embryogenesis. *Mol. Cell. Biol.* **29**, 3186–3203 (2009).

36. Tanaka, S. S. *et al.* Loss of Lhx1 activity impacts on the localization of primordial germ cells in the mouse. *Dev. Dyn.* **239**, 2851–2859 (2010).

37. Tan, K. *et al.* The Rhox gene cluster suppresses germline LINE1 transposition. *Proc Natl Acad Sci USA* **118**, (2021).

38. Rengaraj, D. *et al.* Finer resolution analysis of transcriptional programming during the active migration of chicken primordial germ cells. *Comput. Struct. Biotechnol. J.* **20**, 5911–5924 (2022).

39. Zhao, H. *et al.* The chromatin remodeler Chd4 maintains embryonic stem cell identity by controlling pluripotency- and differentiation-associated genes. *J. Biol. Chem.* **292**, 8507–8519 (2017).

40. Kim, J. D. *et al.* Rex1/Zfp42 as an epigenetic regulator for genomic imprinting. *Hum. Mol. Genet.* **20**, 1353–1362 (2011).

41. Rolland, A. D., Lehmann, K. P., Johnson, K. J., Gaido, K. W. & Koopman, P. Uncovering gene regulatory networks during mouse fetal germ cell development. *Biol. Reprod.* **84**, 790–800 (2011).

42. Ikeda, S. *et al.* Disruption of piRNA machinery by deletion of ASZ1/GASZ results in the expression of aberrant chimeric transcripts in gonocytes. *J. Reprod. Dev.* **68**, 125–136 (2022).

43. Ollinger, R. *et al.* Deletion of the pluripotency-associated Tex19.1 gene causes activation of endogenous retroviruses and defective spermatogenesis in mice. *PLoS Genet.* **4**, e1000199 (2008).

44. Reichmann, J. *et al.* Tex19.1 inhibits the N-end rule pathway and maintains acetylated SMC3 cohesin and sister chromatid cohesion in oocytes. *J. Cell Biol.* **219**, (2020).

45. Vourekas, A. *et al.* The RNA helicase MOV10L1 binds piRNA precursors to initiate piRNA processing. *Genes Dev.* **29**, 617–629 (2015).

46. Jin, S. *et al.* Inference and analysis of cell-cell communication using CellChat. *Nat. Commun.* **12**, 1088 (2021).

47. Gu, Y., Runyan, C., Shoemaker, A., Surani, A. & Wylie, C. Steel factor controls primordial germ cell survival and motility from the time of their specification in the allantois, and provides a continuous niche throughout their migration. *Development* **136**, 1295–1303 (2009).

48. Chang, C.-T., Lee, Y.-H., HuangFu, W.-C. & Liu, I.-H. Cell-intrinsic Fgf signaling contributes to primordial germ cell homing in zebrafish. *Theriogenology* **158**, 424–431 (2020).

49. Whyte, J. *et al.* FGF, Insulin, and SMAD Signaling Cooperate for Avian Primordial Germ Cell Self-Renewal. *Stem Cell Reports* **5**, 1171–1182 (2015).

50. Takeuchi, Y., Molyneaux, K., Runyan, C., Schaible, K. & Wylie, C. The roles of FGF signaling in germ cell migration in the mouse. *Development* **132**, 5399–5409 (2005).

51. Møllgård, K. *et al.* Human primordial germ cells migrate along nerve fibers and Schwann cells from the dorsal hind gut mesentery to the gonadal ridge. *Mol. Hum. Reprod.* **16**, 621–631 (2010).

52. Wolff, E., Suplicki, M. M. & Behr, R. Primordial germ cells do not migrate along nerve fibres in marmoset monkey and mouse embryos. *Reproduction* **157**, 101–109 (2019).

53. Santiago, A. & Erickson, C. A. Ephrin-B ligands play a dual role in the control of neural crest cell migration. *Development* **129**, 3621–3632 (2002).

54. Singh, A., Winterbottom, E. & Daar, I. O. Eph/ephrin signaling in cell-cell and cell-substrate adhesion. *Front Biosci (Landmark Ed)* **17**, 473–497 (2012).

55. Owens, D. A. *et al.* High-throughput analysis reveals novel maternal germline RNAs crucial for primordial germ cell preservation and proper migration. *Development* **144**, 292–304 (2017).

56. Padmanabhan, K. *et al.* Thymosin β4 is essential for adherens junction stability and epidermal planar cell polarity. *Development* **147**, (2020).

57. Jeong, S., Lim, S., Schevzov, G., Gunning, P. W. & Helfman, D. M. Loss of Tpm4.1 leads to disruption of cell-cell adhesions and invasive behavior in breast epithelial cells via increased Rac1 signaling. *Oncotarget* **8**, 33544–33559 (2017).

58. Zhao, X., Jiang, M. & Wang, Z. TPM4 promotes cell migration by modulating F-actin formation in lung cancer. *Onco Targets Ther* **12**, 4055–4063 (2019).

59. Soto-Suazo, M., San Martin, S., Ferro, E. S. & Zorn, T. M. T. Differential expression of glycosaminoglycans and proteoglycans in the migratory pathway of the primordial germ cells of the mouse. *Histochem. Cell Biol.* **118**, 69–78 (2002).

60. Laurila, E., Savinainen, K., Kuuselo, R., Karhu, R. & Kallioniemi, A. Characterization of the 7q21-q22 amplicon identifies ARPC1A, a subunit of the Arp2/3 complex, as a regulator of cell migration and invasion in pancreatic cancer. *Genes Chromosomes Cancer* **48**, 330–339 (2009).

61. Gamallat, Y. *et al.* ARPC1B Is Associated with Lethal Prostate Cancer and Its Inhibition Decreases Cell Invasion and Migration In Vitro. *Int. J. Mol. Sci.* **23**, (2022).

62. Rauhala, H. E., Teppo, S., Niemelä, S. & Kallioniemi, A. Silencing of the ARP2/3 complex disturbs pancreatic cancer cell migration. *Anticancer Res.* **33**, 45–52 (2013).

63. Auvynet, C. *et al.* Galectin-1 promotes human neutrophil migration. *Glycobiology* **23**, 32–42 (2013).

64. Rizqiawan, A. *et al.* Autocrine galectin-1 promotes collective cell migration of squamous cell carcinoma cells through up-regulation of distinct integrins. *Biochem. Biophys. Res. Commun.* **441**, 904–910 (2013).

65. Musa, J., Aynaud, M.-M., Mirabeau, O., Delattre, O. & Grünewald, T. G. MYBL2 (B-Myb): a central regulator of cell proliferation, cell survival and differentiation involved in tumorigenesis. *Cell Death Dis.* **8**, e2895 (2017).

66. Navarrete-López, P. *et al.* Loss of the importin Kpna2 causes infertility in male mice by disrupting the translocation of testis-specific transcription factors. *iScience* **26**, 107134 (2023).

67. Rezende, N. C. *et al.* Rex1 (Zfp42) null mice show impaired testicular function, abnormal testis morphology, and aberrant gene expression. *Dev. Biol.* **356**, 370–382 (2011).

68. Tanaka, K. *et al.* Abnormal early folliculogenesis due to impeded pyruvate metabolism in mouse oocytes†. *Biol. Reprod.* **105**, 64–75 (2021).

69. Shen, M. M. Nodal signaling: developmental roles and regulation. *Development* **134**, 1023–1034 (2007).

70. Lee, K. L. *et al.* Graded Nodal/Activin signaling titrates conversion of quantitative phospho-Smad2 levels into qualitative embryonic stem cell fate decisions. *PLoS Genet.* **7**, e1002130 (2011).

71. Pinheiro, D., Kardos, R., Hannezo, É. & Heisenberg, C.-P. Morphogen gradient orchestrates pattern-preserving tissue morphogenesis via motility-driven unjamming. *Nat. Phys.* (2022) doi:10.1038/s41567-022-01787-6.

72. Collins, M. M. *et al.* Pitx2c orchestrates embryonic axis extension via mesendodermal cell migration. *eLife* **7**, (2018).

73. Gonzalez, V. *et al.* Cooperation between Nodal and FGF signals regulates zebrafish cardiac cell migration and heart morphogenesis. *BioRxiv* (2024) doi:10.1101/2024.01.05.574380.

74. Liu, Z. *et al.* Fscn1 is required for the trafficking of TGF-β family type I receptors during endoderm formation. *Nat. Commun.* **7**, 12603 (2016).

75. Mochizuki, K. *et al.* Repression of germline genes by PRC1.6 and SETDB1 in the early embryo precedes DNA methylation-mediated silencing. *Nat. Commun.* **12**, 7020 (2021).

76. de Castro, R. O. *et al.* Mouse Chd4-NURD is required for neonatal spermatogonia survival and normal gonad development. *Epigenetics Chromatin* **15**, 16 (2022).

77. Yuan, S. *et al.* The histone modification reader ZCWPW1 promotes double-strand break repair by regulating cross-talk of histone modifications and chromatin accessibility at meiotic hotspots. *Genome Biol.* **23**, 187 (2022).

78. Wells, D. *et al.* ZCWPW1 is recruited to recombination hotspots by PRDM9 and is essential for meiotic double strand break repair. *eLife* **9**, (2020).

79. Nguyen, D. H. *et al.* Apoptosis in the fetal testis eliminates developmentally defective germ cell clones. *Nat. Cell Biol.* **22**, 1423–1435 (2020).

80. Jo, K. *et al.* Efficient differentiation of human primordial germ cells through geometric control reveals a key role for Nodal signaling. *eLife* **11**, (2022).

81. Spiller, C. M. *et al.* Endogenous Nodal signaling regulates germ cell potency during mammalian testis development. *Development* **139**, 4123–4132 (2012).

82. Yamashiro, S., Yamakita, Y., Ono, S. & Matsumura, F. Fascin, an actin-bundling protein, induces membrane protrusions and increases cell motility of epithelial cells. *Mol. Biol. Cell* **9**, 993–1006 (1998).

83. Yamakita, Y. *et al.* Fascin1 promotes cell migration of mature dendritic cells. *J. Immunol.* **186**, 2850–2859 (2011).

84. Lamb, M. C. & Tootle, T. L. Fascin in cell migration: more than an actin bundling protein. *Biology (Basel)* **9**, (2020).

85. Lundgaard Riis, M. & Jørgensen, A. Deciphering Sex-Specific Differentiation of Human Fetal Gonads: Insight From Experimental Models. *Front. Cell Dev. Biol.* **10**, 902082 (2022).

86. Richardson, B. E. & Lehmann, R. Mechanisms guiding primordial germ cell migration: strategies from different organisms. *Nat. Rev. Mol. Cell Biol.* **11**, 37–49 (2010).

87. Wei, K.-H. & Liu, I.-H. Heparan sulfate glycosaminoglycans modulate migration and survival in zebrafish primordial germ cells. *Theriogenology* **81**, 1275–85.e1 (2014).

88. García-Castro, M. I., Anderson, R., Heasman, J. & Wylie, C. Interactions between germ cells and extracellular matrix glycoproteins during migration and gonad assembly in the mouse embryo. *J. Cell Biol.* **138**, 471–480 (1997).

89. Lange, M. *et al.* CellRank for directed single-cell fate mapping. *Nat. Methods* **19**, 159–170 (2022).

90. Haston, K. M., Tung, J. Y. & Reijo Pera, R. A. Dazl functions in maintenance of pluripotency and genetic and epigenetic programs of differentiation in mouse primordial germ cells in vivo and in vitro. *PLoS ONE* **4**, e5654 (2009).

91. Gill, M. E., Hu, Y.-C., Lin, Y. & Page, D. C. Licensing of gametogenesis, dependent on RNA binding protein DAZL, as a gateway to sexual differentiation of fetal germ cells. *Proc Natl Acad Sci USA* **108**,

7443–7448 (2011).

92. Spiller, C. M., Burnet, G. & Bowles, J. Mouse fetal germ cell isolation and culture techniques. *Methods Mol. Biol.* **1463**, 173–183 (2017).

93. Renier, N. *et al.* iDISCO: a simple, rapid method to immunolabel large tissue samples for volume imaging. *Cell* **159**, 896–910 (2014).

94. Young, M. D. & Behjati, S. SoupX removes ambient RNA contamination from droplet-based single-cell RNA sequencing data. *Gigascience* **9**, giaa151. (2020).

95. Wolf, F. A., Angerer, P. & Theis, F. J. SCANPY: large-scale single-cell gene expression data analysis. *Genome Biol.* **19**, 15 (2018).

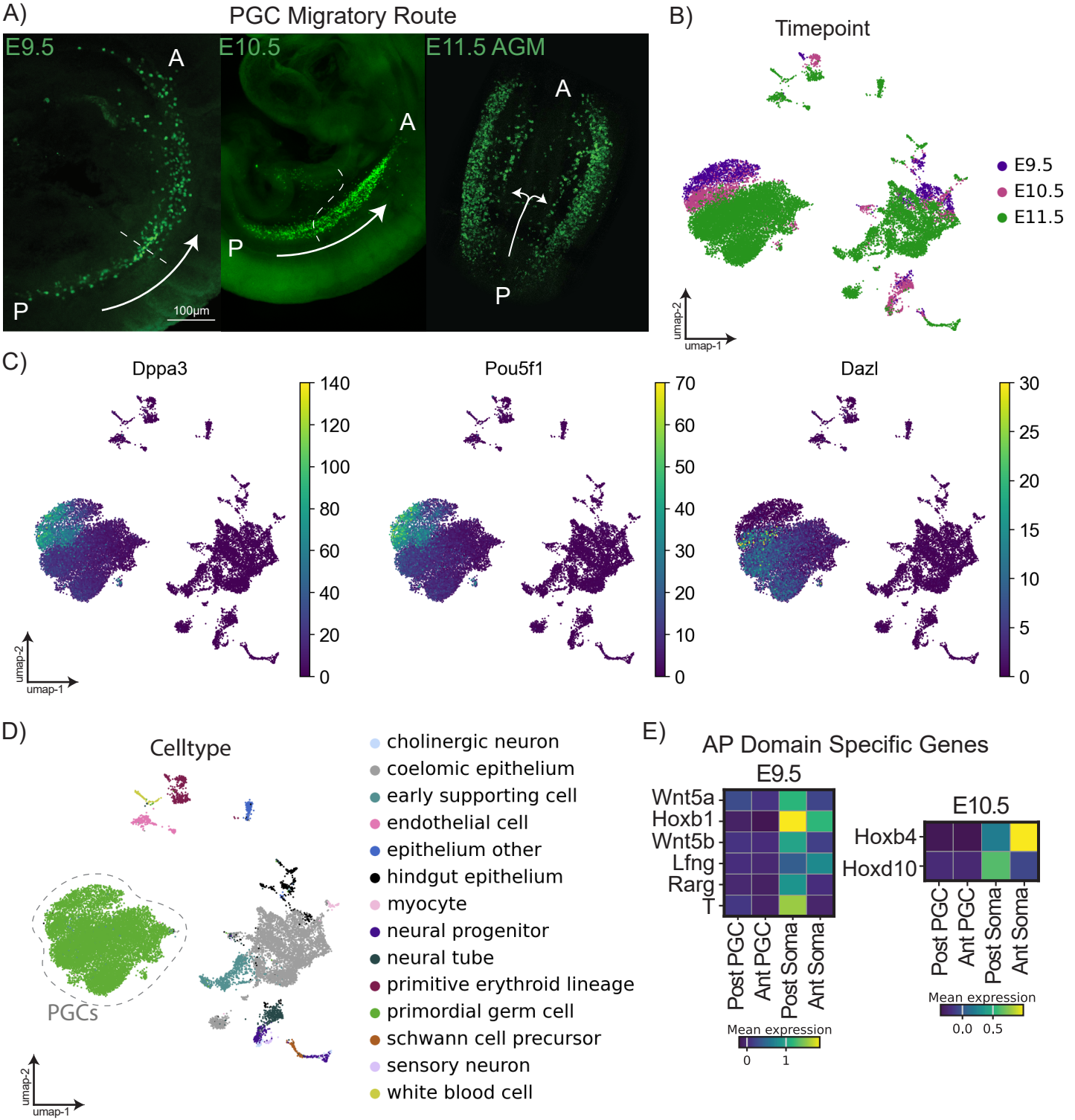
96. Wolock, S. L., Lopez, R. & Klein, A. M. Scrublet: Computational Identification of Cell Doublets in Single-Cell Transcriptomic Data. *Cell Syst.* **8**, 281-291.e9 (2019).

97. Satija, R., Farrell, J. A., Gennert, D., Schier, A. F. & Regev, A. Spatial reconstruction of single-cell gene expression data. *Nat. Biotechnol.* **33**, 495–502 (2015).

98. Wagner, D. E. *et al.* Single-cell mapping of gene expression landscapes and lineage in the zebrafish embryo. *Science* **360**, 981–987 (2018).

99. Cao, J. *et al.* The single-cell transcriptional landscape of mammalian organogenesis. *Nature* **566**, 496–502 (2019).

**Figure 1**


A)
PGC Migratory Route
E9.5
E10.5
E11.5 AGM
A
P
100µm
B)
Timepoint
E9.5
E10.5
E11.5
umap-2
umap-1
C)
Dppa3
Pou5f1
Dazl
umap-2
umap-1
D)
Celltype
PGCs
cholinergic neuron
coelomic epithelium
early supporting cell
endothelial cell
epithelium other
hindgut epithelium
myocyte
neural progenitor
neural tube
primitive erythroid lineage
primordial germ cell
schwann cell precursor
sensory neuron
white blood cell
umap-2
umap-1
E)
AP Domain Specific Genes
E9.5
Wnt5a
Hoxb1
Wnt5b
Lfng
Rarg
T
Post PGC
Ant PGC
Post Soma
Ant Soma
Mean expression
0
1
E10.5
Hoxb4
Hoxd10
Post PGC
Ant PGC
Post Soma
Ant Soma
Mean expression
0.0
0.5

# Figure 2

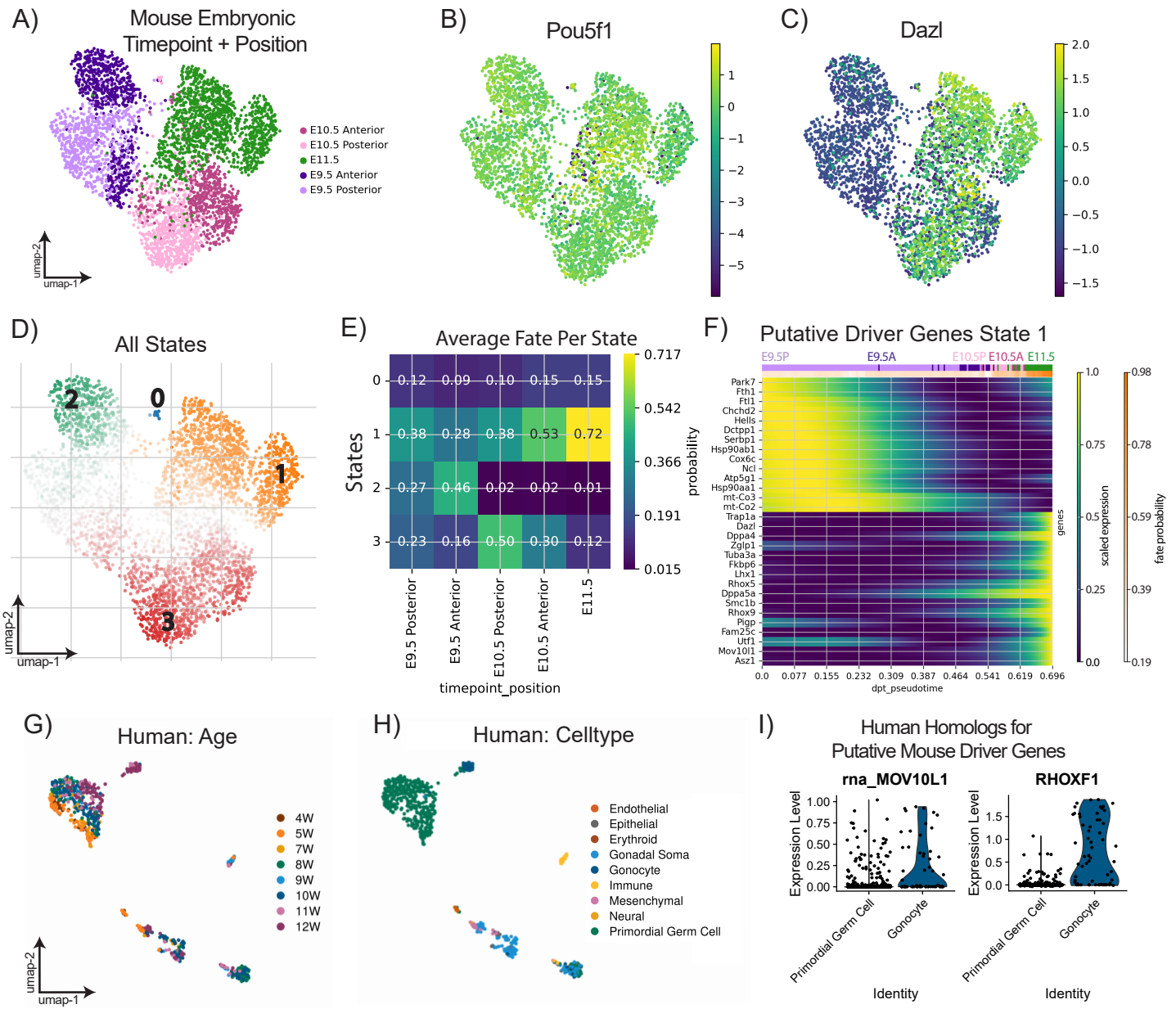
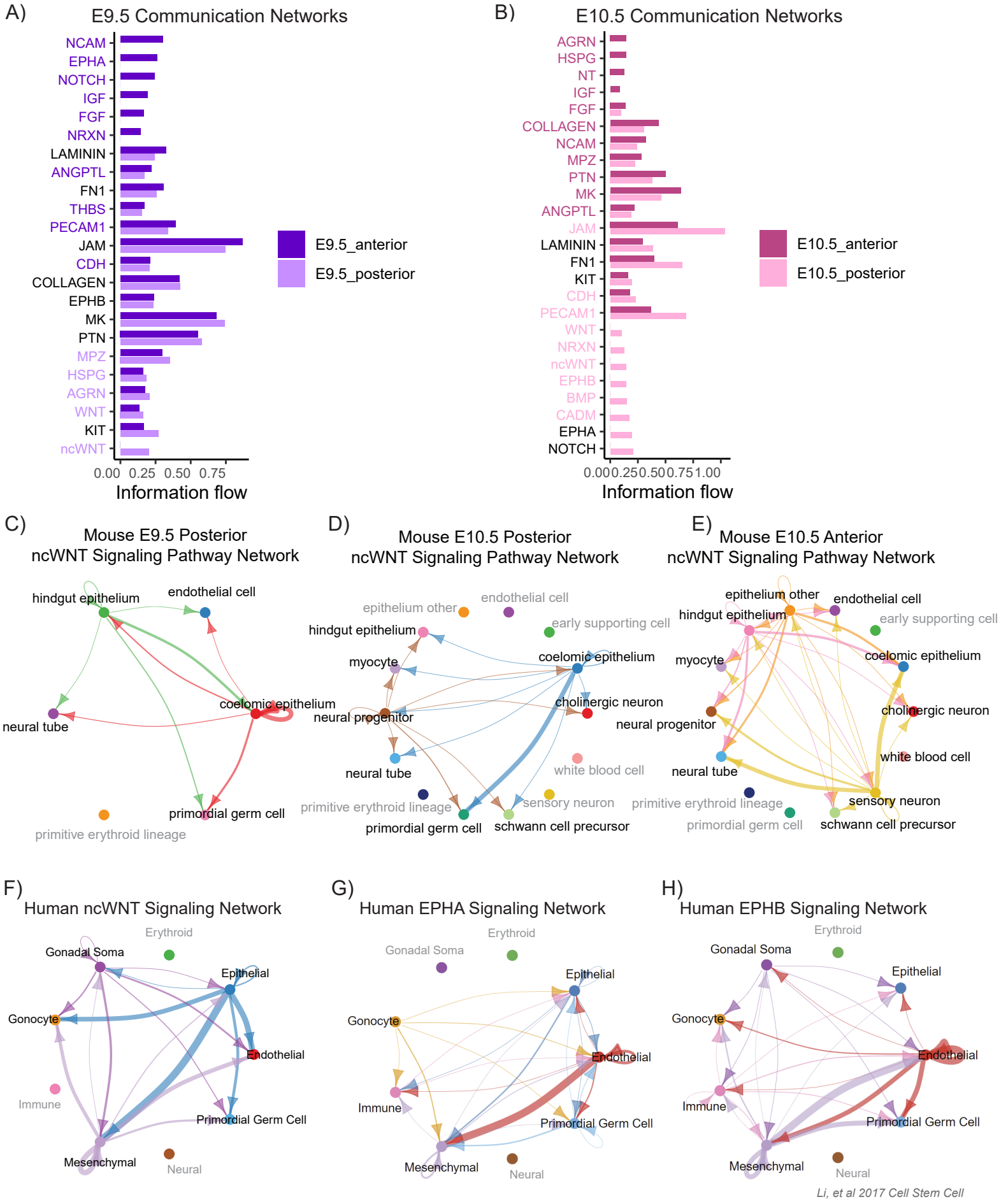

A)
Mouse Embryonic Timepoint + Position
E10.5 Anterior
E10.5 Posterior
E11.5
E9.5 Anterior
E9.5 Posterior
umap-2
umap-1
B)
Pou5f1
C)
Dazl
D)
All States
E)
Average Fate Per State
States
timepoint_position
E9.5 Posterior
E9.5 Anterior
E10.5 Posterior
E10.5 Anterior
E11.5
F)
Putative Driver Genes State 1
E9.5P
E9.5A
E10.5P
E10.5A
E11.5
probability
genes
scaled expression
fate probability
dpt_pseudotime
G)
Human: Age
4W
5W
7W
8W
9W
10W
11W
12W
H)
Human: Celltype
Endothelial
Epithelial
Erythroid
Gonadal Soma
Gonocyte
Immune
Mesenchymal
Neural
Primordial Germ Cell
I)
Human Homologs for Putative Mouse Driver Genes
rna_MOV10L1
RHOXF1
Expression Level
Primordial Germ Cell
Gonocyte
Identity

**Figure 3**




*Li, et al 2017 Cell Stem Cell*

**Figure 4**

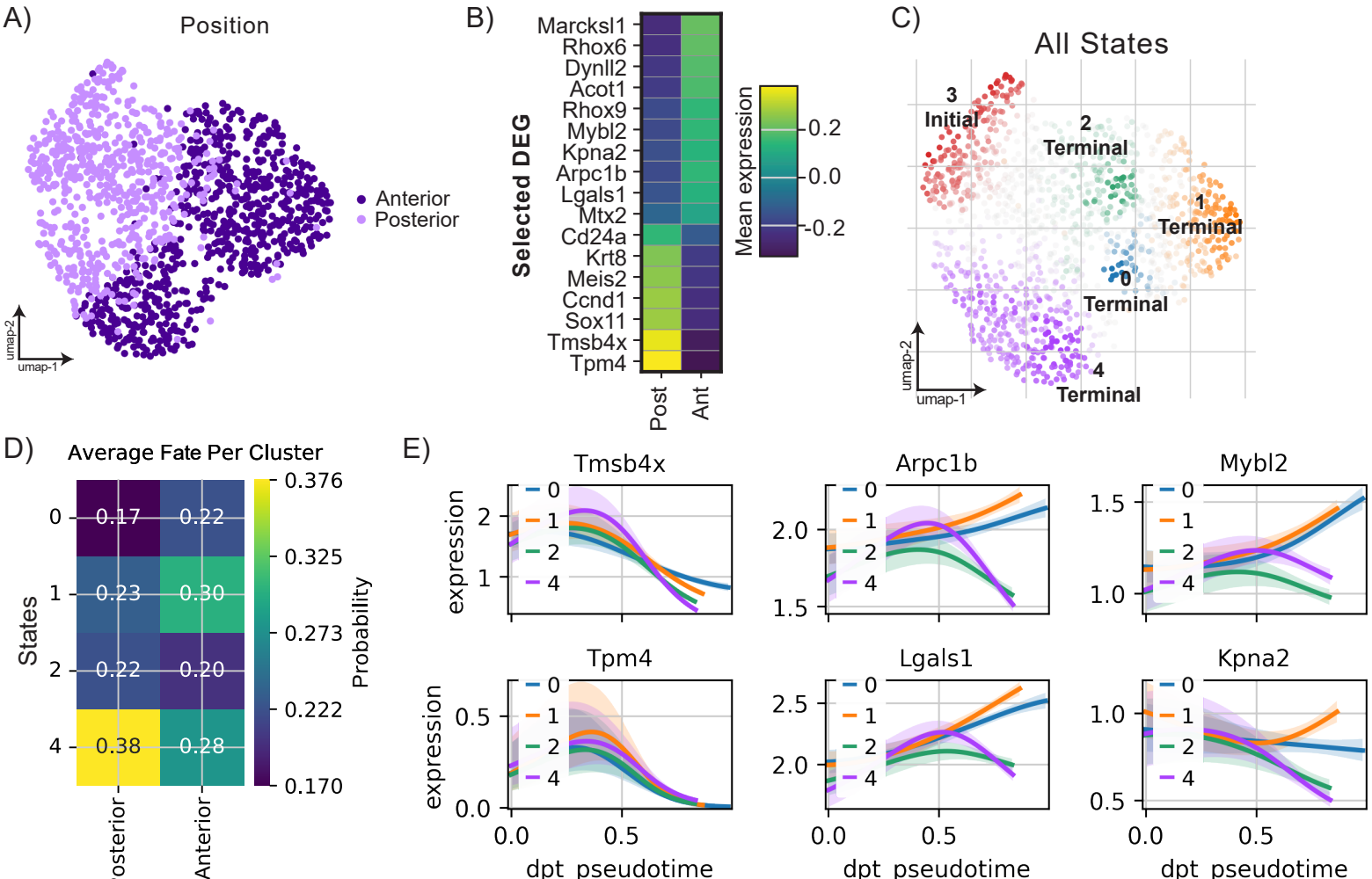
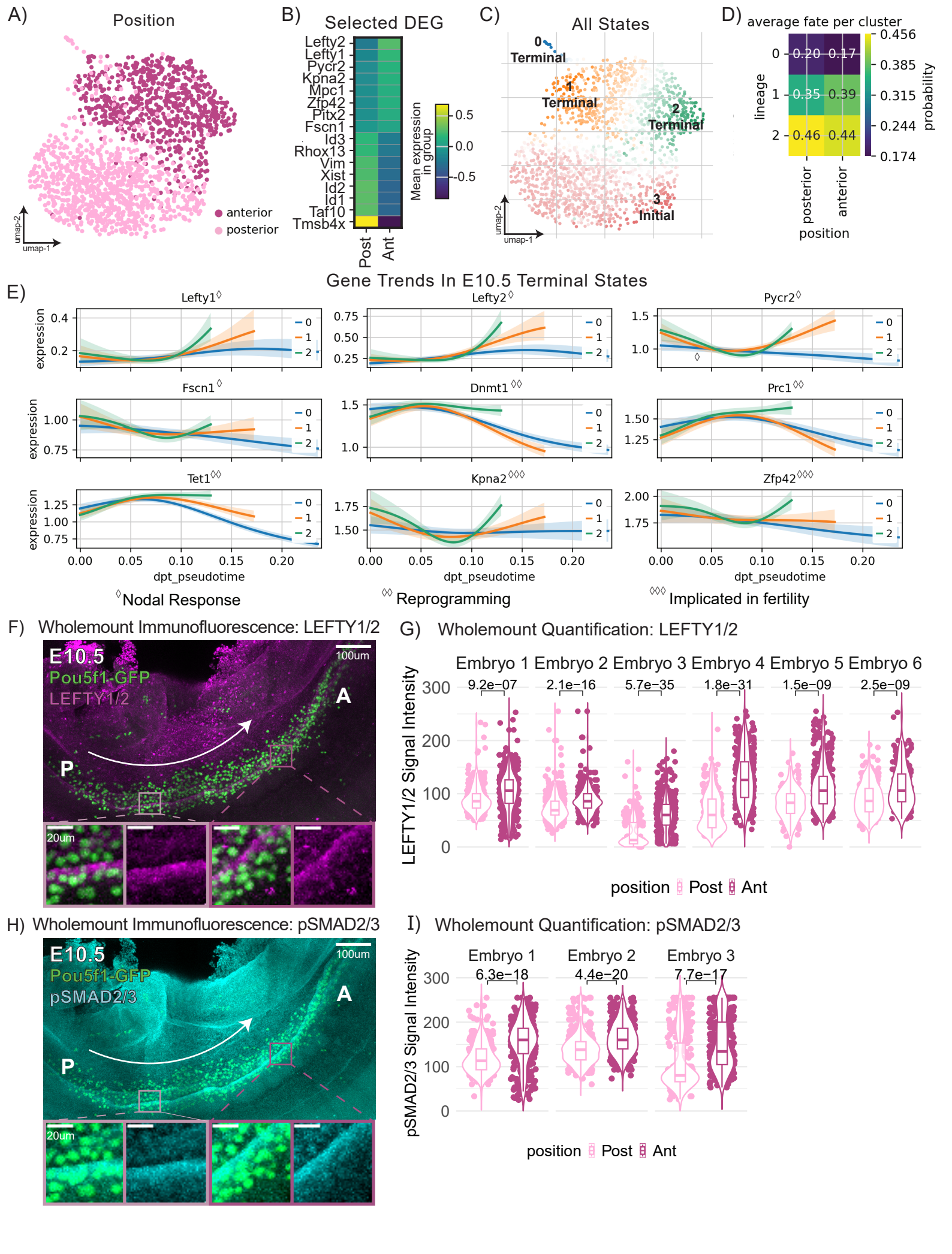
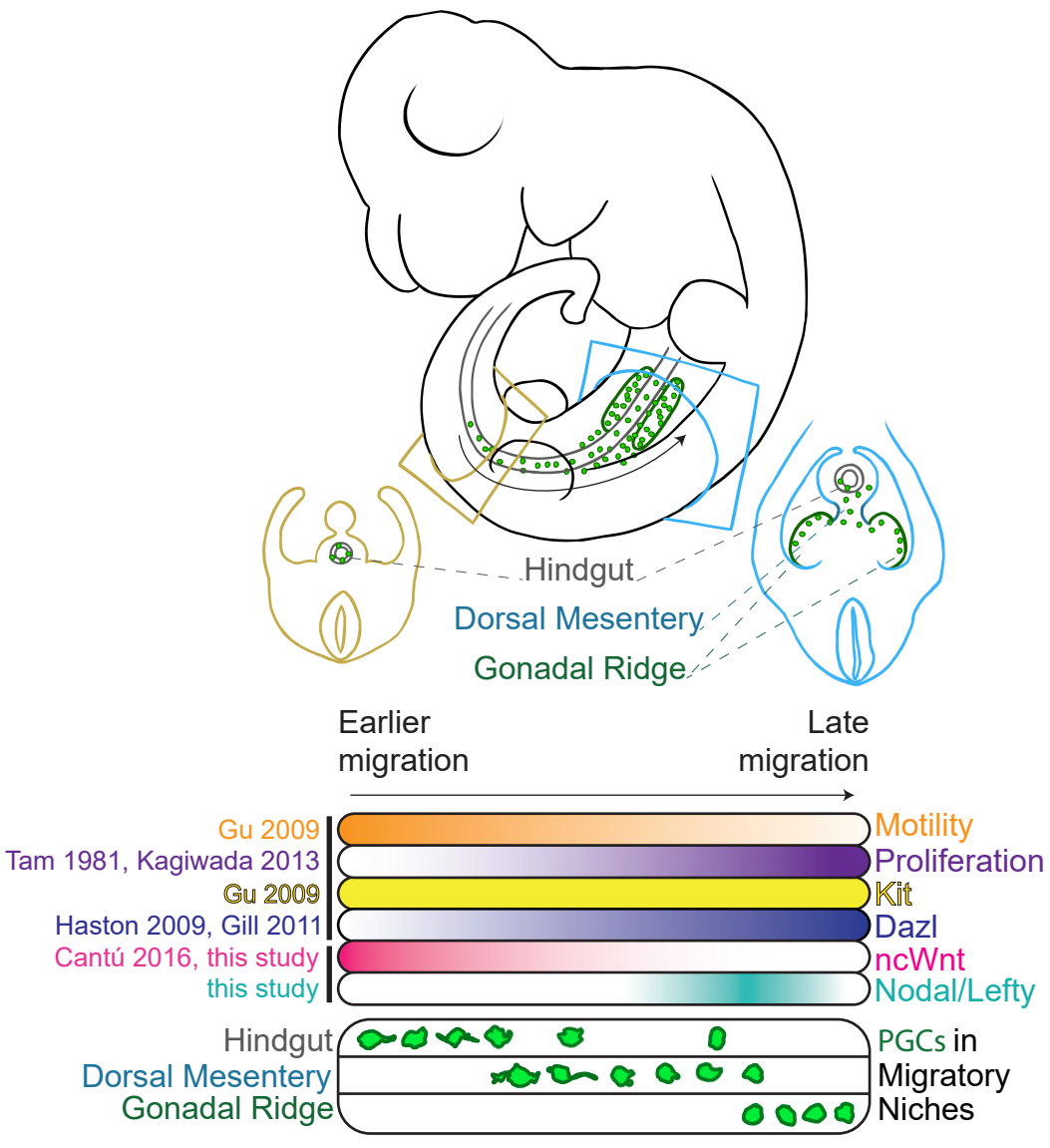
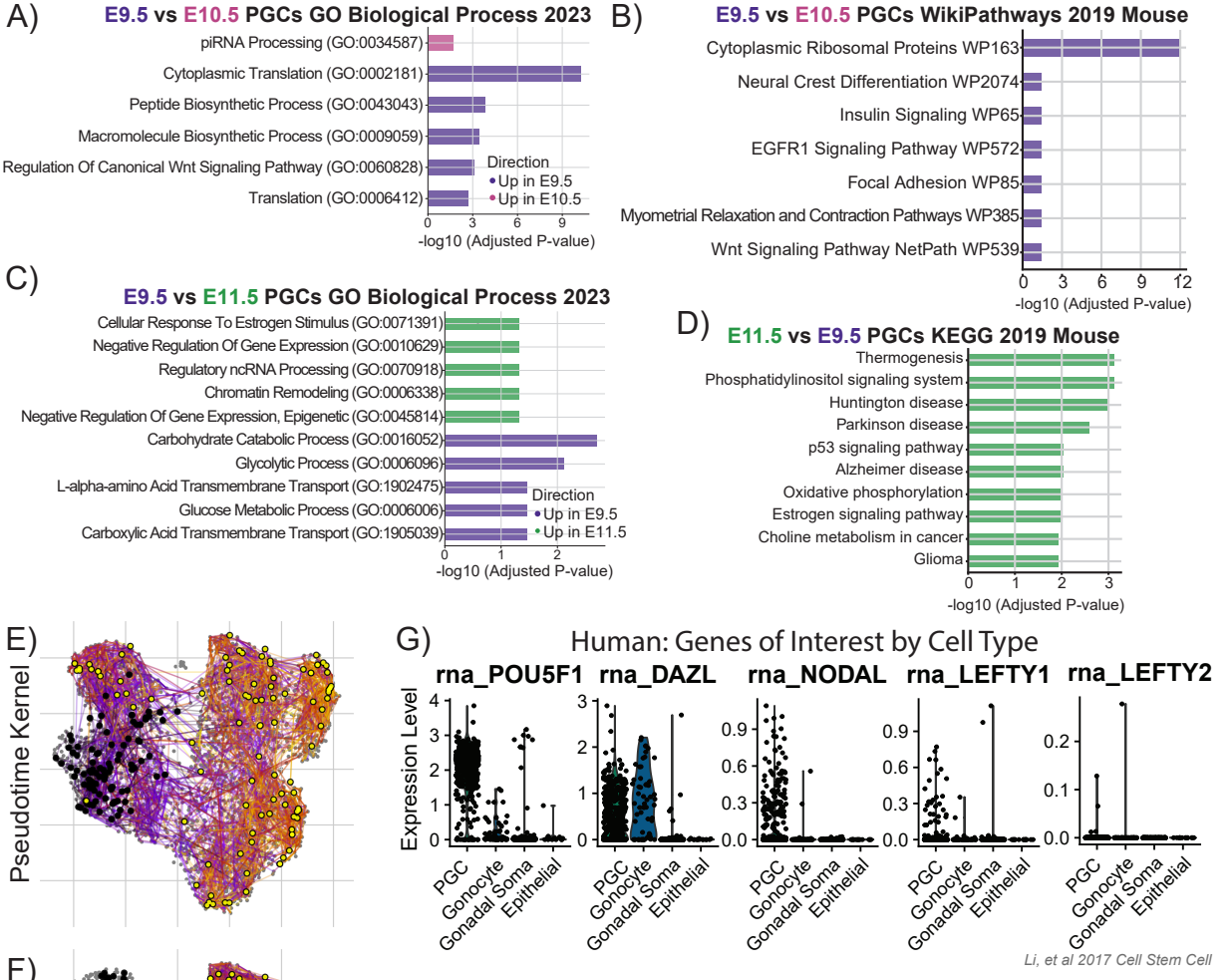

A)
Position
Anterior
Posterior
umap-2
umap-1
B)
Marcksl1
Rhox6
Dynll2
Acot1
Rhox9
Mybl2
Kpna2
Arpc1b
Lgals1
Mtx2
Cd24a
Krt8
Meis2
Ccnd1
Sox11
Tmsb4x
Tpm4
Selected DEG
Post
Ant
Mean expression
0.2
0.0
-0.2
C)
All States
3
Initial
2
Terminal
1
Terminal
0
Terminal
4
Terminal
umap-2
umap-1
D)
Average Fate Per Cluster
States
0
1
2
4
0.17
0.22
0.23
0.30
0.22
0.20
0.38
0.28
Posterior
Anterior
Probability
0.376
0.325
0.273
0.222
0.170
E)
Tmsb4x
Arpc1b
Mybl2
Tpm4
Lgals1
Kpna2
expression
dpt_pseudotime
0
1
2
4

**Figure 5**

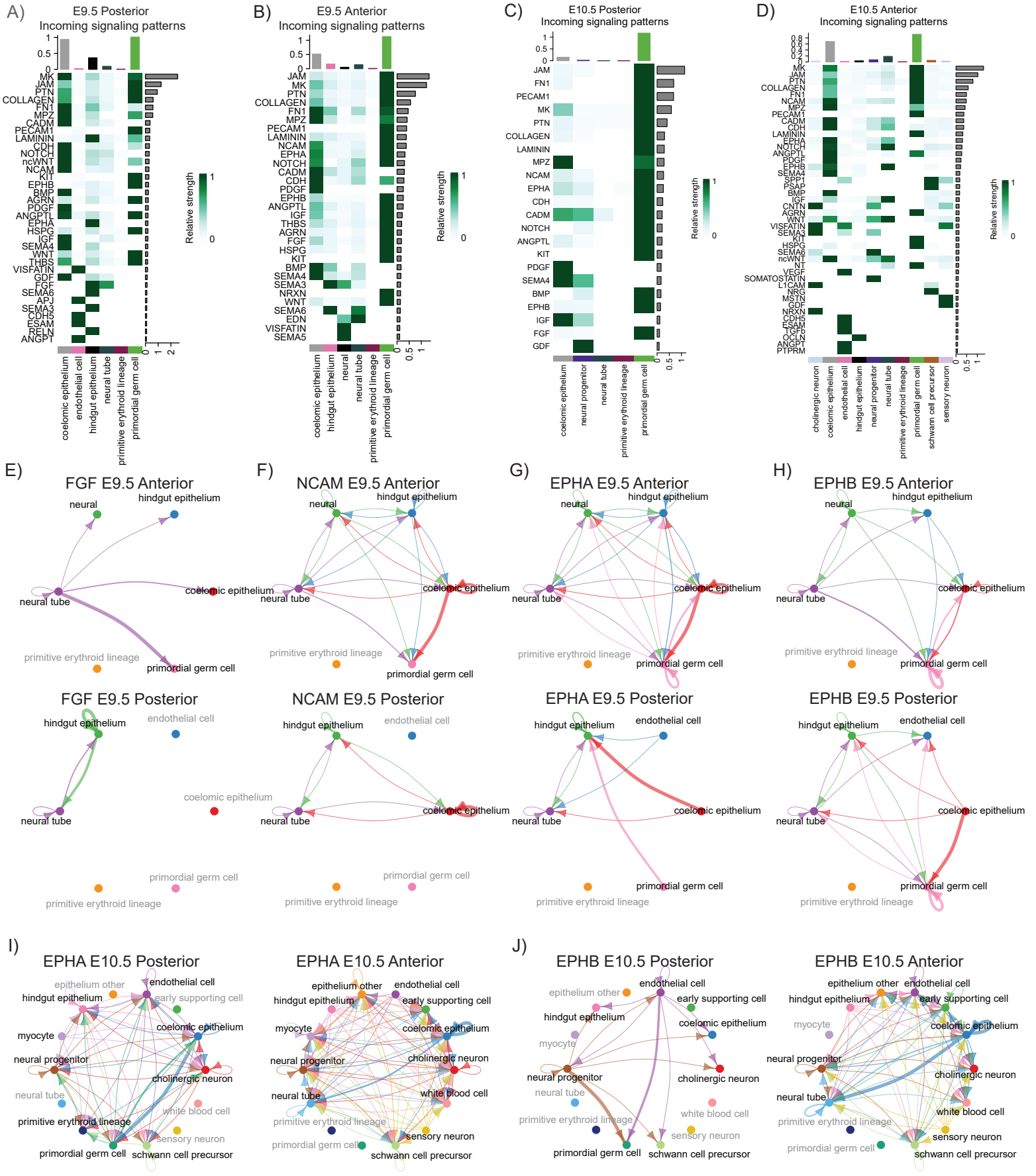

A)
Position
anterior
posterior
umap-1
umap-2
B)
Selected DEG
Lefty2
Lefty1
Pycr2
Kpna2
Mpc1
Zfp42
Pitx2
Fscn1
Id3
Rhox13
Vim
Xist
Id2
Id1
Taf10
Tmsb4x
Post
Ant
Mean expression in group
0.5
0.0
-0.5
C)
All States
0 Terminal
1 Terminal
2 Terminal
3 Initial
D)
average fate per cluster
lineage
0 0.20 0.17
1 0.35 0.39
2 0.46 0.44
posterior
anterior
position
probability
0.456
0.385
0.315
0.244
0.174
E)
Gene Trends In E10.5 Terminal States
Lefty1◊
Lefty2◊
Pycr2◊
Fscn1◊
Dnmt1◊◊
Prc1◊◊
Tet1◊◊
Kpna2◊◊◊
Zfp42◊◊◊
expression
dpt_pseudotime
◊Nodal Response
◊◊ Reprogramming
◊◊◊ Implicated in fertility
F) Wholemount Immunofluorescence: LEFTY1/2
E10.5
Pou5f1-GFP
LEFTY1/2
100um
A
P
20um
G) Wholemount Quantification: LEFTY1/2
Embryo 1 Embryo 2 Embryo 3 Embryo 4 Embryo 5 Embryo 6
9.2e−07
2.1e−16
5.7e−35
1.8e−31
1.5e−09
2.5e−09
LEFTY1/2 Signal Intensity
position Post Ant
H) Wholemount Immunofluorescence: pSMAD2/3
E10.5
Pou5f1-GFP
pSMAD2/3
100um
A
P
20um
I) Wholemount Quantification: pSMAD2/3
Embryo 1
Embryo 2
Embryo 3
6.3e−18
4.4e−20
7.7e−17
pSMAD2/3 Signal Intensity
position Post Ant


Hindgut
Dorsal Mesentery
Gonadal Ridge
Earlier migration
Late migration
Gu 2009
Tam 1981, Kagiwada 2013
Gu 2009
Haston 2009, Gill 2011
Cantú 2016, this study
this study
Motility
Proliferation
Kit
Dazl
ncWnt
Nodal/Lefty
Hindgut
Dorsal Mesentery
Gonadal Ridge
PGCs in Migratory Niches

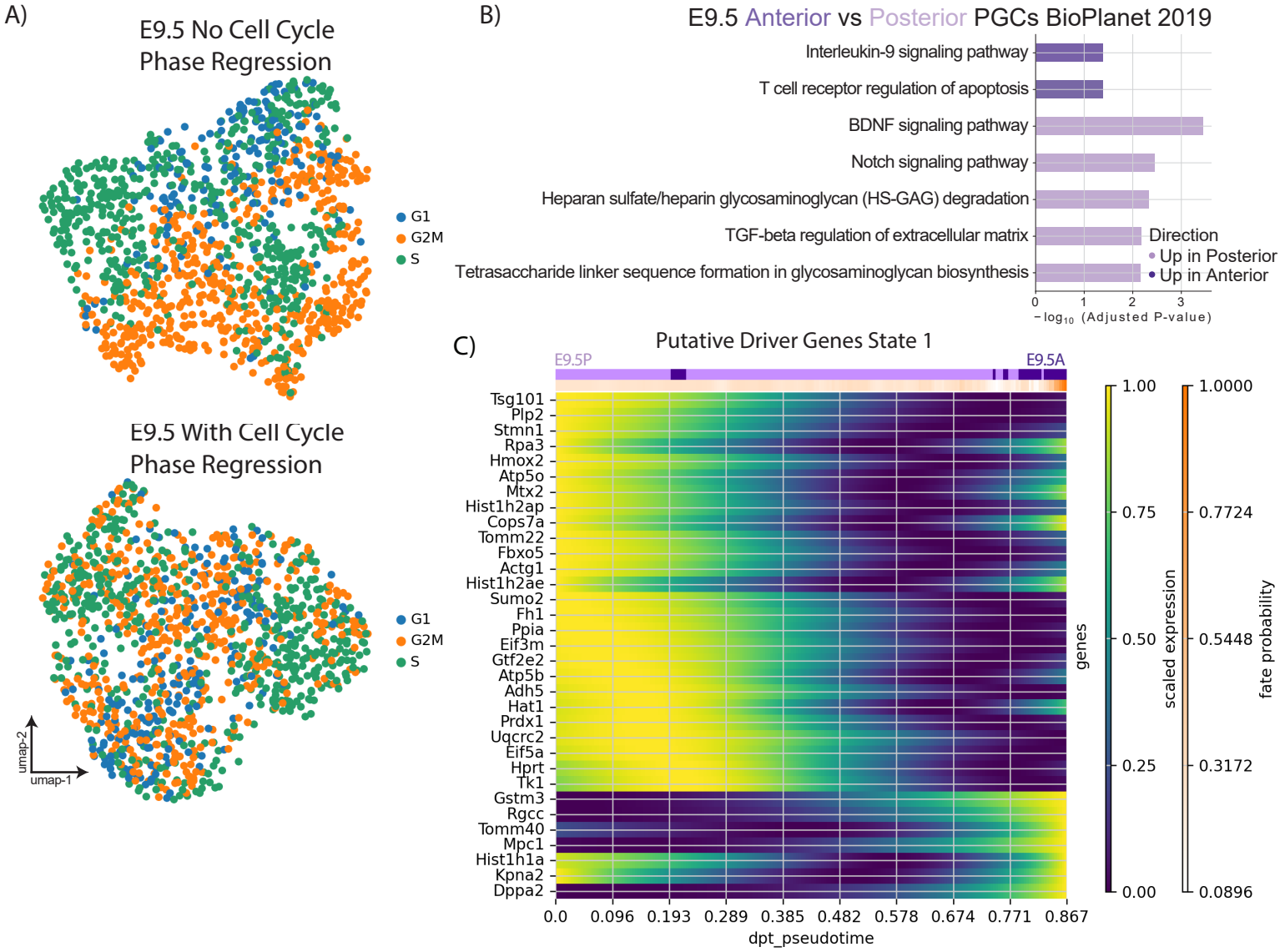
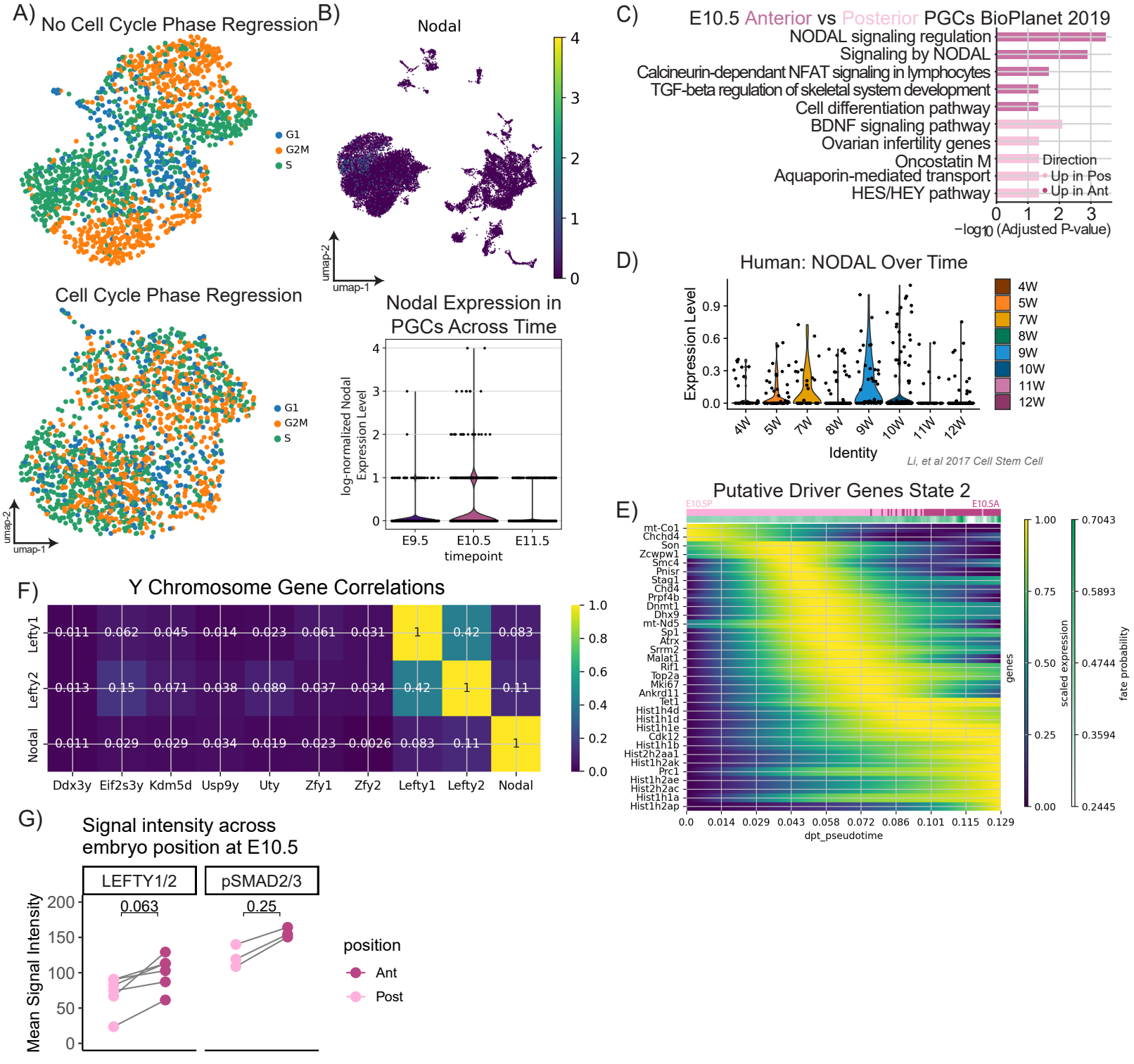
# Supplementary Figure 1



# Supplementary Figure 2



## Supplementary Figure 3



# Supplementary Figure 4