scATAC-Seq reveals epigenetic heterogeneity associated with an EMT-like process in male germline stem cells and its regulation by G9a

Jinyue LIAO[1#], Hoi Ching SUEN[1#], Alfred Chun Shui LUK[1], Annie Wing Tung LEE[1], Judy Kin Wing NG[1], Ting Hei Thomas CHAN[1], Man Yee CHEUNG[1], David Yiu Leung CHAN[2], Tin Chiu LI[2], Huayu QI[3], Wai Yee CHAN[1], Robin M. HOBBS[4]*, Tin-Lap LEE[1]*

[1]Developmental and Regenerative Biology Program, School of Biomedical Sciences, Faculty of Medicine, The Chinese University of Hong Kong, Shatin, Hong Kong SAR, China
[2]Department of Obstetrics and GynaecologyFaculty of Medicine, The Chinese University of Hong Kong, Shatin, Hong Kong SAR, China
[3]Guangzhou Regenerative Medicine and Health Guangdong Laboratory, Guangzhou Institutes of Biomedicine and Health, Guangzhou, China
[4]Development and Stem Cells Program, Monash Biomedicine Discovery Institute and Department of Anatomy and Developmental Biology, Monash University, Melbourne, Victoria 3800, Australia

Author list footnotes

[#]JL and HCS contributed equally to the work.

*Correspondence should be addressed to RMH and TLL.

Robin M. HOBBS, PhD

Address: Australian Regenerative Medicine Institute, Level 1, 15 Innovation Walk, Monash University, Clayton, Melbourne, VIC 3800, Australia.

Email: robin.hobbs@monash.edu

Telephone: +61 3 9902 9611

Tin-Lap LEE, PhD

Address: Rm 622A, 6/F., Lo Kwee-Seong Integrated Biomedical Sciences Building, The Chinese University of Hong Kong, Shatin, Hong Kong SAR, China

Email: leetl@cuhk.edu.hk

Telephone: +852 3943 4436

Abstract

Epithelial-mesenchymal transition (EMT) is a phenomenon in which epithelial cells acquire mesenchymal traits. It contributes to organogenesis and tissue homeostasis, as well as stem cell differentiation. Emerging evidence indicates that heterogeneous expression of EMT gene markers presents in sub-populations of germline stem cells (GSCs). However, the functional implications of such heterogeneity are largely elusive. Here, we unravelled an EMT-like process in GSCs by *in vitro* extracellular matrix (ECM) model and single-cell genomics approaches. Through modulating ECM, we demonstrated that GSCs exist in interconvertible epithelial-like and mesenchymal-like cell states. GSCs gained higher migratory ability after transition to a mesenchymal-like cell state, which was largely mediated by the TGF-β signaling pathway. Dynamics of epigenetic regulation at the single-cell level was also found to align with the EMT-like process. Chromatin accessibility profiles generated by single-cell sequencing assay for transposase-accessible chromatin (scATAC-seq) clustered GSCs into epithelial-like and mesenchymal-like states, which were associated with differentiation status. The high-resolution data revealed novel regulators in the EMT-like process, including transcription factors *Zeb1* and *Hes1*. We further identified putative enhancer-promoter interactions and *cis*-co-accessibility networks at loci such as *Tgfb1*, *Notch1* and *Lin28a*. Importantly, we demonstrated that histone methyltransferase G9a promoted the GSC EMT-like process *in vitro* and contributed to neonatal germ cell migration *in vivo*. Our work thus provides the foundation for understanding the EMT-like process and a comprehensive resource for future investigation of epigenetic regulatory networks in GSCs.

Introduction

Stem cell migration underpins many developmental processes of metazoan animals and germ cell development is no exception. Similar to other stem cells, appropriate migration of spermatogonial stem cells (SSCs) is important for their survival, maintenance and differentiation. Shortly after birth, gonocytes, the precursors of SSCs, resume mitotic activity and begin to migrate to the basal lamina of testicular cords. During a broadly defined timeframe of postnatal days (PND) 0–6, they transition to postnatal spermatogonia, including SSCs[1–3]. In adult testis, SSC migration in seminiferous tubules has also been visualized by intravital live-imaging[4]. Since migrating cells are subjected to different physical microenvironments, spermatogonial migration may be linked to cell fate determination through external signals. A recent study suggested that migrating SSCs could tune their self□renewal and differentiation fate in response to fibroblast growth factor consumption on the basement membrane, suggesting a mechanistic link between SSC differentiation and migration[5]. However, at present, molecular regulatory mechanisms underlying SSC migration as well as its connection with stemness remain largely unknown.

Epithelial-to-mesenchymal transition (EMT) provides an attractive mechanism for explaining how SSCs regulate migration, stemness and the crosstalk between them. EMT is a crucial cellular program that enables polarized epithelial cells to transition toward a mesenchymal phenotype with increased cellular motility. In addition, EMT actively participates in stem cell differentiation during various developmental processes. For example, during gastrulation, epiblast cells undergo EMT to differentiate into three germ layers[6]. During the transition, a hybrid E/M state might exist to allow cells to migrate and respond to external signaling[7]. Classic EMT-transcription factors (EMT-TFs) also participate in developmental programs. Snail regulates neural crest delamination in ectoderm[8]. *Zeb2*, together with *Sip1*, modulates

the differentiation of embryonic hematopoietic stem cells to progenitor cells[9]. *Zeb2* is also involved in germ layer differentiation from ESCs[10].

Epithelial-mesenchymal (E/M) heterogeneity implies that cells within a population can exhibit epithelial, mesenchymal, or hybrid E/M phenotypes. There is evidence that E/M heterogeneity may exist in undifferentiated spermatogonia, a population of cells containing SSCs and differentiation-primed progenitors. SSCs have generally been considered as neither mesenchymal or epithelial due to expression of both epithelial (*Cdh1*) and mesenchymal (*Thy1*) markers. E-cadherin (CDH1) is a cell-cell adhesion molecule that plays essential roles in epithelial cell behavior and cancer suppression[11], while loss of E-cadherin is a hallmark of EMT[12]. In contrast, THY1, a cell surface glycoprotein, is an adhesion molecule of the immunoglobulin superfamily expressed in fibroblasts and mesenchymal stromal cells[13]. In addition, several SSC-related genes have also implicated roles in EMT processes. For instance, *Cxcr4* and *Foxc2* promote migration and invasiveness through EMT in different types of cancer[14–18]. CD87 (urokinase-type plasminogen activator (uPA) receptor/uPAR), a marker for neonatal undifferentiated spermatogonia, can be regulated by TGFβ and participates in tumor invasion and metastasis[19,20]. Recent whole transcriptome analysis also confirmed that THY1+ subpopulations of neonatal germ cells were mesenchymal-like due to higher levels of EMT genes (*Vim*, *Zeb2*), while the ID4+ population (SSC-enriched) expressed lower levels of EMT genes[21].

Whether an EMT or EMT-like process exists in undifferentiated spermatogonia and how it contributes to the regulation of SSC function remain largely unknown. In this study, we explored these questions by characterising an EMT-like process in germline stem cells (GSCs) cultured in vitro, which was derived from PND5.5 Oct4-GFP+/KIT- undifferentiated spermatogonia that contain transplantable SSC populations. We found that GSCs maintained on Matrigel displayed heterogeneity of gene expression related to EMT markers and ECM organization, which largely resembled germ cells in vivo. Through manipulating ECM in vitro, we revealed the presence of interconvertible epithelial-like and mesenchymal-like GSCs. To account for cellular heterogeneity in these cultures, we leveraged highly multiplexed single-cell ATAC-Seq (scATAC-Seq) to reconstruct the epigenetic regulatory program and revealed distinct cell states during this EMT-like process. Lastly, we identified Euchromatic histone-lysine N-methyltransferase 2 (EHMT2), also known as G9a, as a key player in GSC migration both in vivo and in vitro. These data and described strategies provide a useful resource for future epigenetic studies of GSC differentiation and migration.

Results

In vitro modelling of epithelial-mesenchymal heterogeneity in neonatal germ cells

Epithelial/mesenchymal (E/M) heterogeneity has been largely overlooked in current scRNA-Seq studies of spermatogonia, with most studies focused on the regulation of stemness and differentiation. As the first step to elucidate whether spermatogonia may display E/M or E/M-like heterogeneity, we re-analyzed published scRNA-Seq datasets of adult and neonatal spermatogonia[22,23]. In both sets of data, undifferentiated spermatogonia can be separated into stem cell-like (*Id4*-high) and progenitor cell-like (*Rarg*-high) population, and an additional differentiated-primed cell-like population (*Kit*-high) in neonatal spermatogonia (Supplementary Figure S1A and B). Interestingly, the progenitor cell-like and differentiation-

primed cell-like population showed higher expression of EMT-TFs (*Snai1*, *Zeb1*) and EMT marker genes (*Loxl2*, *Mmp2*), implying that a subpopulation of undifferentiated spermatogonia may undergo EMT or EMT-like process (Supplementary Figure S1A and B).

We reasoned that an in vitro culture system would facilitate the investigation of germ cell EMT-like process. Conventional methods to culture germ cells rely on mouse embryonic fibroblasts (MEF) as feeder cells. However, MEF's ECM properties and its effect on EMT regulation are largely unknown. The presence of a MEF feeder layer also complicates pharmacological manipulation in an in vitro assay. Therefore, we adopted a Matrigel-based feeder-free culture system as Matrigel can mimic the mechanical and biochemical properties of ECM and support long-term maintenance of GSCs[24]. Cultured undifferentiated spermatogonia (GSC lines) from isolated Oct4+/KIT- germ cells of PND5.5 mice were derived and stabilized in MEF-dependent conditions, then transferred to Matrigel-coated plates for subsequent long-term culture (Supplementary Figure S2A). To address whether cultured cells still maintain a self-renew gene expression program, cells were examined during 9 days of the MEF-to-Matrigel transition period (Supplementary Figure S2B). We confirmed GSCs in Matrigel-based culture expressed comparable levels of a panel of stem (*Gfra1*, *Etv5*) and progenitor-associated genes (*Upp1*, *Rarg*) as MEF-dependent cultures throughout the transition period (Supplementary Figure S2C).

We previously showed that neonatal differentiation-primed germ cells (Oct4-GFP-high) expressed higher levels of mesenchymal markers (e.g. *Vim*, *Dab2* and *Mmp3*) along with genes significantly enriched in ECM organization and cell migration (Supplementary Figure S2D and E) [19]. To examine whether such E/M-like heterogeneity is maintained in vitro, we separated GSC subpopulations from culture based on Oct4-GFP intensity by FACS and performed RNA-Seq (Figure 1A, Supplementary Figure S2F). Consistent with expression patterns in vivo, stemness gene expression was higher in the Oct4-GFP-low subpopulation, while progenitor gene expression was higher in Oct4-GFP-high subpopulation (Supplementary Figure S2D). These observations are consistent with a previous study that analysed spermatogonial cultures derived from adult Oct4-GFP mice [22]. We also observed that the Oct4-GFP-high subpopulation in both datasets exhibited upregulation of well-known mesenchymal markers, such as *Vim*, *Mmp2*, *Col1a1* and *Loxl2*, although there was no significant difference in EMT-TF expression (*Zeb*, *Snail*) (Supplementary Figure S2E). Of the 589 differentially expressed genes (DEGs) (FDR < 0.05 and fold change ≥ 2) identified, 371 DEGs were upregulated in Oct4-GFP-high GSCs while 218 were upregulated in Oct4-GFP-low GSCs. Mfuzz clustering analysis reveals 8 clusters of different expression trends (Supplementary Figure 2G). The GO terms enriched in DEGs in Cluster 1 (n=103) and Cluster 5 (n=49) with positive correlation to Oct4-GFP expression were associated with Wnt signaling and ECM (Figure 1B). For example, Cluster 1 DEGs include metalloprotease genes *Adamts1* and *Mmp17*, which are known to participate in cancer progression[25,26]. Cluster 3 showed a similar trend and enriched GO terms were related to growth factor binding. Among them, FGFR2 mediates mouse SSC self-renewal through MAP2K1 signaling[27]. On the other hand, the DEGs in Cluster 2 (n=145) with negative correlation to Oct4-GFP levels are enriched for immunity (Figure 1D). *Tlr2*, a Toll-like receptor gene in Cluster 2, has been shown to promote EMT in cancer[28]. Taken together, higher Oct4-GFP levels were associated with a more mesenchymal-like property both in vitro and in vivo. Our culture system thus successfully recapitulated E/M-like heterogeneity of germ cells.

Modulation of epithelial-mesenchymal plasticity of GSCs through controlling ECM properties

Upregulation of genes related to ECM and ECM organization in progenitor or differentiating cells may lead to increased ECM components, which can promote EMT induction[29]. To test this possibility, we modulated ECM components using different concentrations of Matrigel coating. Strikingly, we observed the formation of two morphologically distinct GSC colonies with similar cell proliferation rates: low (0.1-1 μg/cm^2) and high (5-20 μg/cm^2) concentrations of Matrigel promoted the formation of domed-shaped colonies and flat-shaped colonies respectively (Figure 1C, Supplementary Figure S3A). Domed GSCs displayed a more epithelial-like phenotype and were packed into a round colony, possibly due to higher cell-cell adhesion. In contrast, flat GSCs exhibited a more mesenchymal-like phenotype, marked by more cell protrusions. RT-PCR analysis indicated that GSCs cultured on 0.1 μg/cm^2 of Matrigel expressed higher levels of epithelial marker *Cdh1* (encoding E-cadherin), while GSCs cultured on 20 μg/cm^2 of Matrigel expressed higher levels of the mesenchymal marker *Vim* (encoding vimentin) (Supplementary Figure S3B).

Domed GSC colonies were less firmly attached to the matrigel and would detach after gentle pipetting, which allowed easy separation of domed and flat GSCs for further characterization. The loss of cell-cell adhesion molecule CDH1 is considered a hallmark of EMT[12]. Therefore, we further investigated the CDH1 protein level. Flow cytometry analysis indicated that the CDH1+ proportion of domed GSCs is 93.2%, which is significantly higher than that of flat GSCs (71.1%) ($n=3$, $p<0.05$), while the CDH1+ proportion of GSC on MEF was at an intermediate level (82.1%). (Figure 1D, Supplementary Figure S3C). Consistent with this, immunofluorescence staining revealed that flat GSCs expressed higher levels of vimentin, while domed GSCs showed higher E-cadherin levels (Figure 1E). Interestingly, we found that flat colonies expressed higher levels of CD87 (uPAR), which has been suggested to be involved in germ cell migration (Figure 1E) [19]. Indeed, transwell invasion assay showed that flat GSCs had higher invasive capacity (Figure 1F, Supplementary Figure S3D). Based on their characteristics, we applied the term epithelial-like GSCs (EL-GSC) to refer to domed GSCs and the term mesenchymal-like GSCs (ML-GSC) to refer to flat SSCs.

To assess the ability of GSC to transition between the epithelial-like and mesenchymal-like states, we cultured flat GSCs on low concentration Matrigel and monitored them for formation of domed colonies. Following re-plating of cell suspensions, domed-shaped colonies became apparent over a period of 14 days. Likewise, cells derived from domed colonies on low matrigel concentrations formed flat colonies on high concentration Matrigel over this time frame (Supplementary Figure S3E).

Taken together, these results demonstrated that GSCs undergo EMT- or MET-like processes in vitro, which is dependent on ECM property.

TGF-β signaling is involved in the activation of EMT-like program in GSC

To investigate the underlying regulatory pathways during GSC EMT/MET-like process, we performed bulk RNA-Seq analysis on EL-GSCs and ML-GSCs and identified 381 DEGs with FDR < 0.05 and fold change ≥ 2. Surprisingly, most of the DEGs were upregulated in ML-GSCs (Figure 1G). Of these, many were key players of the EMT pathway. We then examined the top upstream regulators in ML-GSCs predicted by Ingenuity Pathway Analysis (IPA) (Figure 1H). TGFB1 caught our attention as many of its interacting partners were also upregulated in ML-GSC, such as CDH2, LOXL2 and MMP2 (Supplementary Figure S4A). We also used IPA to determine the top significant networks upregulated in ML-GSC. The

key nodes in the highest scored network are collagens and ERK1 (Supplementary Figure S4B). The activation of ERK1/2 pathway promotes TGFβ-induced EMT and cancer metastasis[30]. The network with second highest score infers NF-kB pathway, which is another pathway essential for both the induction and maintenance of EMT (Supplementary Figure S4C) (Huber et al. 2004).

The GO terms significantly enriched in ML-GSC included "collagen-containing extracellular matrix", "extracellular structure organization" and "biological adhesion" (Figure 1I). Genes enriched in GO term "biological adhesion", including *Fn1* (encoding fibronectin) and *Itga5* (encoding integrin β4), have also been implicated in cell invasion and migration during EMT [31,32]. Gene Set Enrichment Analysis (GSEA) validated that the most enriched curated gene sets were associated with ECM proteins and regulators, such as "Core Matrisome" and "Extracellular Matrix Organization" (Figure 1J, Supplementary Figure S4D). Furthermore, "Epithelial-Mesenchymal Transition" hallmark gene set was enriched in ML-GSC, indicating ML-GSCs were potentially GSCs that had undergone an EMT-like transition (Figure 2E). TGF-β signaling pathway is widely known to induce EMT in many cell types[33,34]. Thus, we took a closer look at genes that are involved in TGF-β-induced EMT pathway, such as *Tgfb1*, *Zeb2* and *Snai2* (Figure 1K). In accordance with IPA results, we found that most of the genes involved in TGF-β-induced EMT were also upregulated in ML-GSCs (FDR < 0.05) (Figure 2F).

In conclusion, these results indicated that the TGF-β signaling pathway is involved in the GSC EMT-like process in the condition of high concentration Matrigel.

Single cell chromatin states define EMT spectrum in GSC

EMT is dynamically orchestrated by transcription factors (TFs) in both developmental and tumor settings[35]. ATAC-Seq captures the chromatin accessibility and allows elucidation of TF activity. However, bulk ATAC-Seq may miss important biological signals carried by only a subset of cells in heterogeneous populations of GSC. Therefore, we used scATAC-Seq to investigate both cellular heterogeneity and dynamic TF regulation during the GSC EMT/MET-like transition (Figure 2A). We performed scATAC-Seq from a pool of cells collected from 3 time points (1, 7 and 14 days after they were transferred to high concentration Matrigel coating) to capture cells from different stages of the whole EMT-like process. After filtering out low-quality nuclei and doublets, we obtained 2,263 high-quality single nuclei for downstream analysis (Figure 2B).

We adopted SnapATAC to cluster the cells based on the similarity of the chromatin accessibility landscape, which is one of the preferred computational pipelines in recent benchmarking studies[36,37]. SnapATAC portioned these nuclei profiles into six clusters (Figure 2C) and the clustering results were not affected by sequencing depth (Supplementary Figure S5A).

First, we exploited gene scores to explore cellular properties. A gene score is essentially a prediction of how highly expressed a gene will be based on the accessibility of regulatory elements in the vicinity of the gene, including the gene body and putative distal regulatory elements. SSC-1 is likely to be SSC enriched, as most cells in this subset expressed previously recognized mouse SSC marker genes, including *Gfra1* and *Eomes* (Figure 2D). Cells in SSC-2 also expressed *Gfra1* and *Eomes* but at a lower level. Prog-1 cells selectively express progenitor markers *Neurog3*, *Sox3*, *Upp1* and *Nanos3* (Figure 2D). Prog-2

and Prog-3 share a similar gene expression profile, with the expression of progenitor genes *Ddit4* and *Upp1*. Lastly, we found increased gene scores of differentiation markers *Kit* and *Sohlh1* in the Diff-SPG subset (Figure 2D). We then explored the existence of an EMT-like spectrum among the 6 spermatogonia subsets. A subset of cells from SSC-1 and Prog-1 located at the upper region of the UMAP embedding shared higher levels of epithelial markers *Cdh1* and *Ocln*, which we referred to as epithelial-like state (EL-state). Consistently, this subset of cells displayed lowest levels of mesenchymal markers *Vim* and EMT-TFs such as *Zeb1*, *Zeb2*, *Snai1* and *Snai2* (Figure 2D, Supplementary Figure S5B). In contrast, Diff-SPG subset was predicted to express lower levels of epithelial markers *Cdh1* and *Ocln* and increased levels of EMT-TFs (Figure 2D), which we referred to as mesenchymal-like state (ML-state). Thus, cell clustering based on chromatin accessibility revealed cell states spanning GSC differentiation and EMT-like transition.

Second, we tried to associate the chromatin state and phenotypic information. We reasoned that chromatin accessibility signatures associated with bulk populations of EL- and ML-GSCs should preferentially be enriched at either end of the EMT-like spectrum in single cell data. To test this, we performed bulk ATAC-Seq on EL- and ML-GSCs. We first confirmed that EL-GSCs showed higher chromatin accessibility in epithelial marker genes and lower chromatin accessibility in mesenchymal markers and EMT regulators, compared to ML-GSCs (Supplementary Figure S6). Using an analytical framework similar to chromVAR to evaluate enrichment of differentially accessible regions (DARs) between these two phenotypes, we found that EL-GSC and ML-GSC signatures were clearly enriched in EL- and ML-state, respectively (Figure 2E).

Third, we inferred TF motif activity (or in short, TF activity) from scATAC-Seq data. ChromVAR calculates a TF z-score, which infers the binding of TF to open chromatin regions based on the presence of the TF motifs within these regions[38]. We found that the TF z-score of core EMT-TFs *Zeb1*, *Twist2*, *Snai1* and *Snai2* are higher in Diff-SPG, supporting that Diff-SPG are more mesenchymal-like (Supplementary Figure S5C). In addition, chromVAR analysis showed that the most variable TF motifs across all the cells represented EMT regulators *Smarcc1*, *Bach1* and *c-Fos* (Figure 2F and G). BACH1 has been reported to promote EMT by repressing the expression of a set of epithelial genes (Sato et al. 2020). SMARCC1 interacts with NKX6.1 to activate EMT in HeLa cells[39]. FOS forms a multimeric SMAD3/SMAD4/JUN/FOS complex to regulate TGF-β-mediated signaling upon TGF-β activation[40]. Notably, these factors display higher activity in Diff-SPG, suggesting their possible role at the later stage of EMT. In addition, a decreased DMRT1 TF z-score was found in Diff-SPG, which is consistent with the previous finding that loss of DMRT1 promoted spermatogonial differentiation through increasing RA responsiveness and activating *Sohlh1* transcription[41].

Taken together, single-cell chromatin accessibility profiling successfully deconvoluted cellular epigenetic states associated with both differentiation and EMT-like process in GSC. Our results suggested that chromatin accessibility dynamics associated with GSC differentiation are coupled with E/M heterogeneity: stem cells are more epithelial-like while differentiating cells are more mesenchymal-like, which is reminiscent of the finding from in vivo scRNA-seq as described above.

Dynamic changes in accessibility of regulatory elements and TF activity during the EMT-like process

While scRNA-seq has revealed the continuous nature of spermatogonial differentiation through transcriptome analysis, we hypothesized that chromatin accessibility could highlight an additional and distinct layer of transcription factor dynamics. We adopted a semi-supervised pseudotime approach to reconstruct the EMT-like trajectory of GSCs. We reasoned that stem cells at the start of the trajectory would transverse through intermediate states before generating differentiating cells. Given the close association between GSC differentiation and EMT-like process, this path should also chart a trajectory from the EL- to ML- state. 3D UMAP embedding suggested two possible trajectories as cell fates appeared to diverge via two distinct branches from SSC-1 (Figure 3A, Supplementary Figure S5E). The state sequence in Trajectory 1 corresponds to SSC-1 (EL-state) differentiating into progenitors (Prog-1/2/3) and later into Diff-SPG (ML-state), while in Trajectory 2, SSC-1/2 bypassed the intermediate progenitor-state and directly reached the differentiating state (Diff-SPG).

We generated an ordering of single cells (referred to as 'pseudotime') along these trajectories (Figure 3B). As expected, gene score of the mesenchymal marker *Vim* varied in a reciprocal gradient to the epithelial marker *Cdh1* across the embedded cells in both trajectories (Figure 3C). We then determined which genes drive the transition in these trajectories through identifying stage-specific factors that co-vary over pseudotime (Supplementary Figure S7). We found prominence of *Hist1* locus genes at the early stage and olfactory receptors at the late stage of pseudotime. Other candidates in the intermediate stage include several genes with known roles in EMT, including *Macrod2*, *Dock1*, *Ank2* and *Cdh13*. For instance, overexpression of HIST1H1A could upregulate several EMT regulators[42] and MACROD2 overexpression could reverse EMT[43].

We then tried to identify critical TFs involved in the EMT-like process. chromVAR was inherently based on TF motifs and could not provide a one-to-one correspondence between binding motifs and particular TF proteins, because TFs of the same family often share a similar motif. We reasoned that increasing the abundance of an activator TF will increase its TF activity, or its effect on the state of chromatin, which can be reflected by positive correlation between its expression (gene score) and TF activity (chromVAR z-score). Thus, we correlated gene scores of a TF across pseudotime to its corresponding TF z-score to uncover high confidence regulators. This revealed TFs enriched at the EL-state or ML-state of each trajectory (Figure 3D to F). The EL-state featured ESRRG, EGR4 and SP/KLF family TFs. EGR4 regulates spermatogenesis via several testicular cell types including germ cells [44] and loss of EGR4 leads to premature germ cell death and infertility [45]. ESRRG is a member of a nuclear receptor (NR) superfamily of TFs, which has been proposed as a candidate gene for involvement in human male fertility[46]. Conversely, known EMT regulators are enriched in the ML-state. For instance, GLI2 has been reported to cooperate with ZEB1 for transcriptional repression of CDH1 expression in melanoma cells, while FOXE3 regulates EMT in lens epithelial cells[47].

This analysis also revealed differential TF binding dynamics in each branch, suggesting that a distinct set of TFs cooperatively regulates GSC differentiation and EMT-like transition in each trajectory. Several TFs showed a decreasing trend in Trajectory 1 but showing a transient increase in Trajectory 2, with some of them having known roles in both EMT and male reproduction (Figure 3F). For example, NFYA is overexpressed in some cancers and its motif is related to meiosis-specific gene regulation[48–50]. *Hes1* is expressed in prepubertal spermatogonia in mouse and SSEA+ human SSCs [51,52]. It can also promote EMT via the

AKT pathway in cancers[53]. Lastly, FEV, a solid tumor oncogene, is also upregulated in active-phase testes for cell differentiation[54,55].

Taken together, our results have provided comprehensive insights into chromatin accessibility dynamics and the distinct TF regulatory programs in GSC undergoing an EMT-like process.

Epithelial-like and mesenchymal-like-state-specific regulatory topics and enhancers in GSC

Next, we sought to explore gene enhancer usage in GSC. First, we identified the top 500 differentially accessible regions (DARs) within each cluster. Then, we located the enhancer regions by considering chromatin state revealed by chromHMM using recent ChIP-seq profiling of Id4-GFP spermatogonia[56]. Following this, we uncovered potential stage-specific enhancers of each cluster (Supplementary Figure S8). For instance, the enhancer located in *Zeb1* was the most accessible in Diff-SPG, in accordance with its ML-state. Other potential enhancer regions resided in *Egfr*, *Foxc1*, *Dnmt3a*, *Chd7* and *Cd9* loci, which all have known functions in spermatogonial regulation or spermatogenesis.

The EL-state contained subsets of cells from both SSC-1 and Prog-1, which could not be definitively resolved in the above cluster-oriented analysis. Therefore, we used cisTopic [57] to explore substructure within cell populations associated with a continuous EMT-like process, which is one of the best methods for dynamic populations [37]. cisTopic identifies co□accessible regulatory regions (cis-regulatory topics) for each cell by clustering all 364,232 scATAC-Seq peaks (Chen et al. 2019). Louvain clustering using the regulatory topics identified (Figure 4A) captured similar separation of cell states as shown in SnapATAC results (Figure 2C). Among the regulatory topics, Topic 10,12,14 and 16 were most prominent in Prog-2, while Topic 1, 8, 15 and 19 were less enriched in Prog-2 (Figure 4B, Supplementary Figure S9A and B). To gain insight into the function of genes associated with distinct regulatory topics, we performed functional enrichment analysis using GREAT analysis. Topic 10 and 16 were related to stem cell maintenance and division, while Topic 8 and 15 showed direct association with EMT (Supplementary Figure S9C and D). Thus, we then identified topics enriched in EL-state (SSC-1 and Prog-1) and ML-state (Diff-SPG). Three topics (Topic 7, 18 and 22) contributed to the segregation of EL-state, and three topics (Topic 6, 11 and 17) contributed more in the ML-state (Figure 4B and C). GREAT analysis revealed the GO terms enriched in mesenchymal-like topics are associated with NF-kB pathway and response to retinoic acid, which is consistent with the IPA analysis of ML-GSC enriched genes and differentiating spermatogonia, respectively (Figure 4D). Epithelial-like topics were associated with stem cell maintenance and Wnt signaling pathway (Figure 4D). Interestingly, more mesenchymal-like topic regions were located in active enhancer regions (4.9-5.0%), compared with epithelial-like topic regions (10.3-13.4%) (Figure 4E). We then determined specific enhancers in EL- and ML-states (Figure 4F). For example, epithelial-like topic regions included active enhancers located in *Fgfr1* and *Usf1*, which is related to undifferentiated spermatogonia and GSC maintenance[58,59]. The enhancer region located in *Cdh13* was also more accessible in EL-state, which is consistent with pseudotime analysis. Mesenchymal-like topic regions contained enhancer regions located in *Notch1* and *Mmp9*, which are associated with EMT.

Many enhancer elements are not located directly adjacent to the gene whose expression it regulates but reside far from the transcription start sites. Besides, relationships between

enhancers and target genes are not always one to one. Therefore, it is still challenging to connect enhancer elements to their target genes[60]. Leveraging a recently described strategy, Cicero[61], we linked distal regulatory elements to their prospective genes via patterns of co-accessibility (Figure 5A). First, we were able to interconnect 26,927 scATAC-Seq peak pairs at a co-accessibility score cut-off of 0.25, in which 5,082 peak pairs linked a distal regulatory element to a promoter. Of these, 1,037 links included distal elements that fell within active enhancer regions, which might represent putative enhancer-promoter interactions. GO analysis of the target genes revealed significant enrichment of EMT regulation (Figure 5B). For example, regions marked by both H3K4me1 and H3K27ac were connected to the promoters of *Notch1* and *Bmp7*, suggesting putative enhancers regulating their expression during the EMT-like process (Figure 5C). Other examples included enhancers for *Spry4* and *Lin28a*, which are both highly expressed in undifferentiated spermatogonia, including SSC (Figure 5C)[62,63].

Further, we identified cis-co-accessibility networks (CCANs) using Cicero, which can inform us about modules of coregulated chromatin in the genome. We focused on the 181 regions with an exceptionally large number of significant co-accessible peak links (high-density CCANs, greater than 10 links), which incorporated 3626 sites. GREAT analysis associated the peaks contained in these high-density CCANs to 1181 genes and revealed significant enrichment in EMT regulation (Binom FDR Q-Value, 2.80e-79). We further examined the relationship of high-density CCANs and super-enhancers, which are clusters of enhancers highly enriched in histone H3K27 acetylation (Adam et al., 2015). We found high-density CCANs significantly overlapped the 1529 super-enhancers identified by ROSE based on SSC H3K27ac mark ($p < 2.489e-61$, hypergeometric test) (Figure 5D and E). Accordingly, genes within high-density CCANs were strongly enriched in super-enhancer associated genes ($p < 8.339e-09$, hypergeometric test, representation factor = 1.5). Focusing on individual loci, we observed key EMT and spermatogonia related genes that exhibited patterns of coordinated regulation, which tended to occupy the same topologically associated domain (TAD) in spermatogonia[64]. We found super-enhancer regions located at *Nanos2* and *Tgfb1* loci. In accordance with the role of *Nanos2* in SSC maintenance, several peaks in the super-enhancer region were more accessible in SSC-2 subset (Figure 5F). Consistent with the role of TGF-β signaling in the EMT-like process, we found that the peaks at *Tgfb1* locus were less active in GSCs at EL-state (SSC-1 and Prog-1) (Figure 5G). We also identified super-enhancers at *Dab2ip* and histone cluster 1 loci. *Dab2ip* facilitates spermatogonial differentiation and overexpression of certain histones promotes EMT [65 42] (Supplementary Figure S10).

Taken together, scATAC-Seq allowed us to identify distal enhancers and enhancer-to-gene relationships associated with EMT-like process.

G9a promotes EMT-like process in GSC in vitro and is required for gonocyte migration in vivo

EMT is shown to be regulated by histone modifications and respective enzymes[66,67]. Among them, the Euchromatic histone-lysine N-methyltransferase 2 (EHMT2), also known as G9a, caught our attention. H3K9 methyltransferase G9a is reportedly induced by TGF-β and has an important role in the development of EMT in cancer cells[68]. We have previously shown that G9a is one of the top genes preferentially expressed in differentiation-primed spermatogonia (Oct4-GFP-high cells) in neonatal testis[19]. To consolidate this finding, we examined the scRNA-Seq datasets to reassess the *G9a* expression profile in spermatogonia (Supplementary Figure S1A-B). We confirmed that *G9a* mRNA levels were consistently higher in the

progenitor-like population (*Rarg*-high) compared to stem cell-like population (*Id4*-high). Given the higher EMT-related gene expression in progenitor-like/differentiation-primed populations and the established role of G9a in EMT in cancer cells, we hypothesized that G9a may also play a role in the GSC EMT-like process.

We first examined the levels of G9a protein and G9a-mediated H3K9me2 modification in our in vitro Matrigel-based GSC culture. The levels of G9a and H3K9me2 in ML-GSC (black arrows, cells with more protrusions) is higher than that of EL-GSC, which supports the idea that G9a might promote EMT-like process in GSC (Figure 6A). Next, we used UNC0638, a substrate-competitive inhibitor that selectively binds enzymatically active G9a, to corroborate this hypothesis[69]. We found that UNC0638 treatment in MEF-dependent GSC culture for 48 hours resulted in a dose-dependent increase of epithelial-like cells as revealed by CDH1 FACS analysis (Figure 6B).

To directly assess the impact of UNC0638 on EMT-like progression, we devised an in vitro assay in which cells from a domed-shaped colony (EL-GSC) migrate into unoccupied margins of the well (Figure 6C). We seeded GSCs on low concentration Matrigel to induce domed-shaped colonies. We then transferred the colonies to high concentration Matrigel after which cells at the colony border could spontaneously undergo EMT-like process and migrate outwards. Cells at the periphery of the patch acquired a spindle-like morphology and formed leading and protruding edges consistent with the acquisition of a mesenchymal phenotype (Supplementary Figure S11A). To demonstrate the robustness of this assay, we examined the effect of several kinase inhibitors. Consistent with the up-regulation of the TGF-β pathway in ML-GSC, spontaneous EMT-like progression depends on TGF-β, as TGF-β type 1 receptor inhibitor RepSox could significantly inhibit outward migration (Figure 6D). We also validated the action of other potential kinases (*Pdgfrb*, *Fgfr1*, *Flt4*, *Egfr*), whose gene expression was upregulated in ML-GSCs. These kinases are also reported to have roles in EMT and cancer metastasis[70–73]. Indeed, their respective inhibitors suppressed GSC migration (Supplementary Figure S11B and C). More importantly, we observed similar suppression of EMT-like progression when domed-shaped EL-GSC colonies were treated with UNC0638 (Figure 6D).

To elucidate the mechanisms underlying EMT suppression upon G9a inhibition, we investigated the transcriptomic changes upon different dosage of UNC0638 treatment (Supplementary Figure S11D). Mfuzz clustering of DEGs reveals 4 clusters, in which Cluster 1 (n=79) and Cluster 2 (n=50) showed upregulation while Cluster 3 (n=143) and Cluster 4 (n=61) showed downregulation (Figure 6E). GSEA and GO analysis showed that genes downregulated after UNC0638 treatment are significantly associated with EMT, such as extracellular matrix constituent, cell migration, cell adhesion (Figure 6F and G, Supplementary Figure S11E). For example, we observed downregulation of genes related to cell adhesion (*Itgav*, *Cdh11*), cell motility (*Gja1*, *Cd248*) and cell migration (*Acta2*, *Emilin1*). We also found several genes related to Wnt signaling pathway and P53 pathway were upregulated after UNC0638 treatment, such as *Axin2*, *Fzd7*, *Pten*, *Siah1*, *Perp* and *Ep300* (Supplementary Figure S11F).

We further validated the link between G9a, EMT and migration in GSC. UNC0638-inhibited DEGs significantly overlapped with DEGs upregulated in ML-GSC ($p < 3.715e-178$, Fisher's exact test), consisting of a prominent set of EMT genes associated with the activation of TGF-β signaling such as *Vim*, *Fn1* and *Twist2* (Figure 6H, Supplementary Figure S11G). GSEA verified that high expression gene set in ML-GSC were significantly enriched in the control group when compared to G9a treatment group ($p < 0.0001$) (Supplementary Figure

S12A). We further examined the gene set preferentially expressed in the recently discovered migratory pro-spermatogonia (I-ProSG), which presumably regulates gonocyte migration in vivo[74]. GSEA analysis revealed enrichment of this gene set in ML-GSC and downregulation of this gene set after G9a inhibition, with core-enriched genes related to cell adhesion (Supplementary Figure S12B to D). These results indicated that G9a might be involved in germ cell migration in vivo through EMT-related mechanisms.

To test this idea, we administered UNC0638 daily by PO to neonatal male mice from PND3 for three constitutive days and observed its effect (Figure 6I). On PND6, the testes were collected and proportion of Oct4-GFP+ nascent spermatogonia attached to the basement membrane were quantified. Most of the Oct4-GFP+ cells (92.9%, n=3) finished migration on PND6 in the control group. However, in UNC0638-treatment group, significantly fewer Oct4-GFP+ (74.7%, n=6, $p<0.001$) cells attached to the basement membrane and a large portion of cells remained at the centre of the tubule lumFigure 6J and K). This demonstrated that G9a is required for proper migration of GSC.

Potential target of G9a regulation in GSC

To assess the role of G9a in epigenetic control of the GSC EMT-like progression, we performed scATAC-Seq on UNC0638-treated GSCs. First, we generated synthetic pseudo-bulk datasets by merging data from cells in control and UNC0638-treated groups. Consistent with RNA-Seq results, chromatin accessibility of signature EMT genes (*Twist1*, *Foxc2*, *Lox*) were downregulated after G9a inhibition (Figure 7A). Accessibility at regulatory sites is driven by TF binding. Therefore, a comparison of the TF binding activity based on differential footprinting will allow us to reveal molecular mechanisms of G9a-mediated regulation of gene expression. Surprisingly, we only observed TFs with significant loss of activity after UNC0638 treatment with scHINT analysis[75] (Figure 7B and C). The most prominent factors are members of the Hes/Hey family. Among them, Hes1 was also identified to be associated with EMT-like trajectory in pseudotime analysis.

At single cell level, we compared the control and treatment groups using SnapATAC, because it clusters cells using the cell-by-bin matrix which makes it easy to combine multiple samples and perform comparative analysis. In accordance with the EMT suppression, the proportion of mesenchymal-like cells (Diff-SPG) decreased after UNC0638 treatment (Figure 7D). Interestingly, we observed an increased proportion of cells mainly in Prog-2. This raised the possibility that G9a might have a more profound effect on Trajectory 1, which is consistent with a higher activity of HES1 motif in Prog-2 along Trajectory 1 (Figure 7E). Concordantly, HES1 motif activity in Prog-2 substantially decreased after G9a inhibition. We further examined HES1 protein level and found that it drastically decreased upon G9a inhibition over time and was nearly depleted after 64 hours (Figure 7F). *Hes1* mRNA level also showed a mild decrease trend after UNC0638 treatment (Supplementary Figure S11H). These results suggested that G9a might regulate GSC EMT-like process through HES1.

We next used cisTopic to compare the control and UNC0638 treatment group (Figure 7G). cisTopic identified three “loss of accessibility topics”, Topic 2, 7 and 14 (Figure 7H). Topic 2 is highly enriched for genes associated with Wnt pathway; Topic 7 is related to actin-filament regulation while Topic 14 is related to Notch signaling pathway (Figure 7I). These pathways have been suggested to be necessary for EMT[76–78]. On the other hand, we identified two topics that represent an increase in accessibility (Figure 7H). Topic 3 contains regions

significantly associated with cell-cell adhesion while Topic 15 is associated with general transcriptional regulation (RNA splicing) (Figure 7I).

Collectively, our transcriptomic and epigenetic analysis allow the identification of the underlying mechanism of G9a in the regulation of EMT-like process and its potential role in germ cell migration through HES1 (Figure 7J).

Discussion

Analysis of several gene expression profiling studies have indicated that E/M heterogeneity exists in spermatogonia in vivo. However, the biological relevance of this phenomenon is yet to be determined. In this study, we demonstrated the epithelial-mesenchymal plasticity of GSCs, as we identified two interconvertible subtypes of GSCs with completely different cellular and molecular properties when cultured on Matrigel. We further showed that GSC migration is regulated by EMT-related mechanisms.

First, we recapitulated the EMT-like process in GSC through Matrigel-based culture. Matrigel is a reconstituted cell culture matrix able to control stem cell fate by influencing growth, differentiation and migration[79]. Our results revealed that high concentrations of Matrigel coating induced an EMT-like process as it promoted the formation of flat-shaped GSC colonies and mesenchymal-like cells. This phenotypic change was mainly mediated by TGF-β signaling as evident by RNA-seq analysis. In the SSC niche, ECM proteins contributing to the basement membrane include type IV collagens, laminin, heparan sulfate proteoglycan (HSPG) and entactin, many of which are contained in Matrigel[80]. Among them, HSPGs have been demonstrated to regulate several major signaling pathways, including TGF-β, fibroblast growth factor (FGF), Wnt and Hedgehog (Hh) pathways. It has been shown that FGFs can be immobilized and concentrated on the HSPG-rich basement membrane in testis[5] and FGFs can also augment TGF-β1-induced EMT in mammary epithelial cells[81]. Therefore, it is possible that increased deposition of HSPGs ensures an enriched pool of bioactive growth factors (e.g. TGF-β1 and FGFs) in the accumulating matrix to promote the EMT-like process in GSC.

To our knowledge, our study is the first to dissect the complex epigenetic regulatory networks of GSCs at single cell resolution. Relatively little is known about chromatin accessibility dynamics during SSC differentiation. Previous ATAC-Seq analysis of spermatogenesis revealed that open chromatin in differentiated spermatogonia (KIT+ cells) is similar to that in undifferentiated spermatogonia (THY1+ cells) and differentiation only involves chromatin closure without de novo formation of accessible chromatin[82]. However, bulk ATAC-Seq may mask true biological signals due to the proportions of the different cell states. We also provide evidence at the single cell level that GSC differentiation can proceed by different routes before converging on the differentiated state. Instead of a linear path, the differentiation pathway of our data resembles a “doughnut” shape, which mirrors the findings from scRNA-seq on human germ cells [83].

Regarding the relationship between GSC differentiation and the EMT-like process, our scATAC-Seq suggested that SSC-1 and Diff-SPG clusters coincide with EL- and ML-like states, respectively. We found decreased DMRT1 TF activity in the mesenchymal-like subset Diff-SPG. DMRT1 is one of the top motifs enriched in human SSC-specific ATAC-seq peaks and loss of DMRT1 causes spermatogonia to enter meiosis[41,51]. This suggested that SSCs initiate differentiation when entering the mesenchymal-like state. Supporting this, decreased

DMRT1 TF activity was accompanied by increased gene scores of differentiation genes. This observation resembles the findings in embryo development, in which EMT has been depicted as a process of differentiation, with gradual loss of potency[84]. Therefore, we speculate that the differentiation program is coupled with an EMT-like process in SSCs. Our results provide a clue to the link between EMT and SSC differentiation. However, the functional implications of our observation requires further validation using transplantation assay in the future.

scATAC-Seq analysis also brings the possibility to learn epigenetic regulation related to E/M plasticity. First, we observed gene scores and TF activity of several canonical EMT regulators displaying distinct patterns across the cell states, the most significant of which was the TF ZEB1. Multiple analyses suggested it increased in activity during the mesenchymal-like state, suggesting a role in the GSC EMT-like transition. This is in line with a recent study showing *Zeb1* as a potential MET suppressor in mouse SSC[85]. We also identified gene signatures (e.g. *Smarcc1*, *Bach1* and *c-Fos*) that were not typically considered as canonical EMT regulators but have been previously implicated in EMT. Second, based on predicted activity of these canonical EMT genes, cells can be regarded as progressively transitioned throughout the E/M spectrum. When assessing dynamics of chromatin accessibility throughout the pseudotemporal trajectories, we further identified novel regulators of the GSC EMT-like process, such as *Hes1*. Third, scATAC-Seq also reveals cis-regulatory elements associated with E/M plasticity. Using cisTopic, we were able to identify epithelial- or mesenchymal- associated regulatory topics. Annotating these putative regulatory elements suggested signals from multiple cues orchestrate the proper execution of EMT/MET cycles in GSC. For example, the Wnt signaling pathway showed activation, whereas NF-kB and TGFβ activity decreased in epithelial-like clusters. These regulatory elements represent interesting candidates for future studies.

Using our culture model, we have shown that ATAC-Seq performed on heterogeneous samples at the single-cell level can reveal important dynamic regulatory regions across the EMT-like process. Moreover, combined with the analysis of histone modifications, our data allowed us to identify putative enhancer regions more accessible in different cell states. scATAC-Seq also provides a unique opportunity to identify tightly coordinated co-accessible regulatory elements. We highlighted several potential enhancer-promoter interactions and super-enhancer regions associated with key EMT regulators and SSC-related genes. Our dataset, which represents an initial characterization of genome-wide chromatin accessibility and candidate distal regulatory sites in GSC EMT-like process, should be a valuable resource for investigators interested in male germ cell biology, as well as EMT in general.

Although genes related to germ cell migration such as the TF *Rhox10* and actin disassembly factor *Aip1* have been reported[65,86], mechanisms that regulate germ cell migration are largely unknown due to the lack of an effective cellular model. Our culture system provides a valuable platform to study the dynamics of germ cell migration in vitro. We showed that GSC migration is enhanced in the mesenchymal-like population. Further, we found that the gene expression signatures of ML-GSC are concordant with a recently discovered migratory pro-spermatogonia (ProSG) subset, I-ProSG[74] , both of which showed upregulation of migratory genes. This suggested a shared mechanism of GSC migration in vivo and in vitro. Following this idea, we demonstrated that G9a inhibition largely downregulated this migration-associated gene expression program, leading to inhibition of both the EMT-like process in vitro and gonocyte migration to the basement membrane in vivo.

Despite the known role of G9a as an essential epigenetic regulator during spermatogenesis, such studies have been predominantly limited in scope to male germ cell meiotic progression [87]. Existing data indicated that G9a exerts a transcriptional repression function dependent on its HMTase activity that mediates H3K9 methylation[88]. G9a might cause repression of regulators which in turn controls the EMT-like process in GSC. A plausible candidate is *Cdh1*. We showed that G9a inhibition by UNC0638 led to an increase in CDH1, possibly due to the removal of H3K9me2 on its gene regulatory elements. In line with this, G9a could cause transcriptional repression of *CDH1* in cancers[89,90] . Loss of CDH1 results in the liberation of β-catenin that in turn induces expression of EMT-TFs [91], which might contribute to the downregulation of EMT genes after G9a inhibition. Conversely, G9a inhibition downregulated HES1 in both protein level and TF activity in GSC, which has been reported to promote EMT and cell migration [92]. The underlying regulation can be explained by the function of G9a as a coactivator, since it has been reported that G9a activates the Notch signaling pathway by regulating *Notch*1 and *Hes1* in cancer progression of melanoma[93]. Another possibility is that G9a might promote AKT activity which in turn upregulates HES1[53,94]. Interestingly, we observed increased chromatin accessibility of *Notch1* and *Akt* in mesenchymal-like regulatory topic regions. implying their potential role in the EMT-like process. Yet, further investigation is needed to uncover the mechanisms of G9a-dependent regulation of the EMT-like process in GSC.

Of note, our study focused on chromatin accessibility and did not include scRNA-Seq analysis of GSCs. Recently, multimodal single cell profiling ("multi-omics") has greatly improved our ability to detect unique cell types and states. Joint data of scATAC-Seq and scRNA-Seq from the same cell showed that they are comparable for determining cell identity, suggesting cells coordinate chromatin structure with transcription. Nevertheless, some cell states may not reflect equally in both profiles. Future studies would benefit from joint profiling to provide additional information that further elucidates E/M heterogeneity. Despite the limitations, our scATAC-Seq analysis redefines cellular heterogeneity of the undifferentiated population driven by specific EMT-TFs. Our study thus takes a significant step forward in revealing the complex picture of male germ cell development. The concept that neonatal spermatogonial development may also involve an EMT-like process can provide a valuable model to shape future research.

Materials and Methods

Animals

All the animal experiments were performed according to the protocols approved by the Animal Experiment Ethics Committee (AEEC) of The Chinese University of Hong Kong (CUHK) and followed the Animals (Control of Experiments) Ordinance (Cap. 340) licensed from the Department of Health, the Government of Hong Kong Special Administrative Region. All the mice were housed under a cycle of 12-hour light/dark and kept in ad libitum feeding and controlled the temperature of 22-24°C. Oct4-EGFP transgenic mice (B6; CBA-Tg(Pou5f1-EGFP)2Mnn/J, Stock no.: 004654) were acquired from The Jackson Laboratory and maintained in CUHK Laboratory Animal Services Centre[95].

SSC isolation and derivation of long-term culture

Spermatogonial stem cells from testes of Oct4-GFP transgenic mice at PND5.5 were purified using the method described previously[96]. Isolated seminiferous tubules were digested with 1 mg/ml type 4 collagenase (Gibco), 1 mg/ml hyaluronidase (Sigma-Aldrich) and 5 µg/ml DNase I (Sigma-Aldrich) at 37°C for 20 min with occasional shaking. The suspension was passed through a 40- µm strainer cap (BD Falcon) to yield a uniform single cell suspension. After incubation in staining buffer (PBS supplemented with 1% FBS, HEPES, glucose, pyruvate and penicillin-streptomycin) with APC anti mouse CD117 ckit antibody (553356, BD Biosciences) at 4°C for 30 min, Oct4-GFP+/KIT- cells were collected with a BD FACSAria Fusion Flow Cytometer (BD Biosciences). Primary SSCs were then cultured on mouse embryonic fibroblast layer (150,000 cells per well of a 24 well tissue culture plate). GSC medium (Supplementary Table S1) was used for GSC maintenance. GSCs were passaged by treating with 0.25% trypsin (Invitrogen) for two minutes and adding fresh medium to stop trypsinization. After centrifugation at 1000 rpm for 5 minutes, the cell pellet was resuspended in medium with a density of 100,000 cells per well of a 24 well tissue culture plate, either on MEF layer or Matrigel-coated well.

Western blot

GSC cultured on MEF feeder were treated with 2 mM UNC0638 for 26, 50 and 64 hours. Cells were harvested and FACS sorted to remove MEF feeder cells. Cell lysate was separated by SDS/PAGE, transferred to Immobilon-P membrane (Millipore), and incubated with primary antibodies and HRP-conjugated secondary antibodies. Immunoreactive bands were detected using Western Lightning Plus-ECL (Perkin Elmer). The antibodies used are listed in Supplementary Table S2.

In vivo G9a inhibitor UNC0638 treatment

Oct4-EGFP transgenic mice were used. Treatment with G9a inhibitor was performed by oral administration of UNC0638 (Abcam) diluted in 1% tragacanth (G1128, Sigma) with a dosage of 3 mg/kg daily on PND3-5. DMSO diluted in 1% tragacanth was used as a control. For each litter, at least one male was used for drug treatment and at least one male was used as control. On PND6, the mice were euthanized and their testes were collected. After being

fixed in 4% (v/v) paraformaldehyde for 2 hours, 10 µm cryosections of neonatal mouse testes were prepared as described[97].

Bulk RNA-Seq

Low input RNA-Seq library preparation was conducted according to the modified protocol of Smart-seq2[98]. In brief, the mixture of 1 µl of 10 mM oligo-dT30VN primer (5'-AAGCAGTGGTATCAACGCAGAGTACT30VN-3'), 1 µl of 10 mM dNTP mix and 10 ng extracted RNA in 5 µl reaction volume was incubated at 72°C for 3 min. Then, 5 µl of the reverse transcription mixture (200U SuperScript II reverse transcriptase, 20U RNase inhibitor, 1x Superscript II first-strand buffer, 5 mM DTT, 1 M Betaine, 6 mM MgCl2, 1 µM TSO (5'- AAGCAGTGGTATCAACGCAGAGTACATrGrG+G-3') was added into the tube for RNA reverse transcription. The tubes containing a total 10 µl mixture were placed in ProFlex PCR System (Thermo Fisher Scientific) and the reaction condition was as follows: 42°C for 90 min, followed by ten cycles of 50°C for 2 min and 42°C for 2 min. The final enzyme inactivation was performed at 70°C for 15 min. After finishing the first-strand reaction, the 2x PCR preamplification mix (1x KAPA HiFi HotStart ReadyMix, 0.1 µM IS PCR primers (5'- AAGCAGTGGTATCAACGCAGAGT-3')) was added. PCR preamplification reaction was performed in ProFlex PCR System by using the following program: 98°C for 3 min, 12 cycles of 98°C for 20 s, 67°C for 15 s and 72°C for 6 min, with the final extension cycle of 72°C for 5 min. Amplified cDNA was purified by using Ampure XP beads (1:1 ratio) according to the manufacturer's manual and was eluted in 15 µl of ddH2O. The yield of amplified cDNA was measured by Qubit3.0 fluorometer. Amplified cDNA with good quality were used for library construction with TruePrep DNA Library Prep Kit V2 for Illumina (Vazyme) following the instruction. The quality of all constructed libraries was assessed by Agilent Bioanalyzer 2100 with High Sensitivity DNA Kit (Agilent Technologies). Qualified libraries were sequenced on Illumina HiSeq2000 to generate 150bp paired-end reads. Raw reads were pre-processed with trim-galore for trimming Illumina adaptor sequences (http://www.bioinformatics.babraham.ac.uk/projects/trim_galore/). Counts were assigned to genes defined by the Ensembl (release 67) annotation using RSEM[99]. Differentially expressed genes were identified with edgeR package[100]. Soft clustering using differentially expressed genes was obtained using the Mfuzz package implemented in R[101]. GO enrichment and pathway analysis was performed using ToppGene Suite[102], Ingenuity Pathway Analysis (IPA) and GSEA[103]. A hypergeometric test with the Benjamini and Hochberg false discovery rate (FDR) was performed using the default parameters to adjust p-value.

Bulk ATAC-Seq

ATAC-Seq was performed with ML- and EL-like GSCs according to Omni-ATAC protocol using TruePrep DNA Library Prep Kit V2 for Illumina (TD501) with modifications[104]. Briefly, around 100,000 cells were harvested and resuspended in 25 µL ATAC-Resuspension Buffer (RSB) containing 0.1% NP40, 0.1% Tween-20, and 0.01% Digitonin and incubated on ice for 3 min. The cell lysis was washed out with 0.5mL cold ATAC-RSB containing 0.1% Tween-20. After centrifugation at 500g for 10 mins, pellet nuclei were resuspended in 10ul PBS. Nuclei were counted and 50,000 nuclei were mixed in 50 µL of transposition mixture (10 µL 2x TTBL buffer, 5 µL transposase (TTE mix), 16.5 µL PBS, 0.5 µL 1% digitonin, 0.5 µL 10% Tween-20) by pipetting up and down. Tagmentation were performed at 37°C for 30 min in a thermomixer with 1000 RPM mixing. The transposed DNA was purified with Zymo

DNA clean and concentrator-5 kit. Subsequently, purified transposed DNA was then amplified by PCR for 7 cycles with TruePrep DNA Library Prep Kit. The quality of all constructed libraries was assessed by Agilent Bioanalyzer 2100 with High Sensitivity DNA Kit (Agilent Technologies). Qualified libraries were sequenced on Illumina HiSeq2000 to generate 150bp paired-end reads. Data were analyzed withATAC-Seq pipeline from nf-core with default setting[105]. Peaks identified by MACS2[106] were used for differential analysis with MAnorm[107].

scRNA-Seq analysis

Previously published scRNA-Seq datasets were downloaded from GEO (GSE112880 and GSE109049) [22,23]. UMI count tables of each cellular barcode were analyzed using Seurat (Butler et al. 2018). Seurat objects were built by loading UMI count tables using the "Read10X" function. Each sample was filtered and cells with greater than 1000 genes expressed and less than 20% of reads mapped to the mitochondrial genome were retained. After normalization, the top 2000 highly variable genes were selected using the "FindVariableFeatures" function of the Seurat package. Dimensionality of data was reduced by principal component analysis (PCA) (30 components) and visualized with UMAP. Clustering was performed using Louvain algorithm on 30 principal components (resolution = 0.5). After initial clustering, cluster-specific markers were identified with the "FindAllMarkers" function using default parameters and clusters were assigned based on known markers. Only Kit- clusters were included in subsequent analysis. The generated Seurat object was loaded to BioTuring for marker visualization in UMAP plots.

scATAC-Seq analysis

Sample collection

For feeder-free culture group, cells were cultured on Matrigel-coated plates (10 μg/cm2) for 1, 7 and 14 days respectively after being transferred from MEF-dependent culture. Samples at different time points are pooled together in a ratio of 1:2:2 (Day1:Day7:Day14). For G9a inhibition group, cells cultured on MEF feeder were treated with 2 mM UNC0638 for 26, 50 and 64 hours respectively. Samples at different time points are pooled together in equal numbers. For the control group, cells were cultured on the MEF feeder. Cells from each time point were purified using FACS to remove MEF feeder, cell debris and cell aggregate, and pooled (total 120,000 cells for each sample) as mentioned.

Cell lysis and tagmentation

Cell tagmentation was performed according to SureCell ATAC-Seq Library Prep Kit (17004620, Bio-Rad) User Guide (10000106678, Bio-Rad) and the protocol based on Omni-ATAC was followed (Corces et al. 2017). In brief, washed and pelleted cells were lysed with the Omni-ATAC lysis buffer containing 0.1% NP-40, 0.1% Tween-20, 0.01% digitonin, 10 mM NaCl, 3 mM MgCl2 and 10 mM Tris-HCl pH 7.4 for 3 min on ice. The lysis buffer was diluted with ATAC-Tween buffer that contains 0.1% Tween-20 as a detergent. Nuclei were counted and examined under microscope to ensure successful isolation. Same number of nuclei were subjected to tagmentation with equal ratio of cells/Tn5 transposase to minimize potential batch effect. Nuclei were resuspended in tagmentation mix, buffered with 1×PBS

supplemented with 0.1% BSA and agitated on a ThermoMixer for 30 min at 37 °C. Tagmented nuclei were kept on ice before encapsulation.

scATAC-Seq library preparation and sequencing

Tagmented nuclei were loaded onto a ddSEQ Single-Cell Isolator (Bio-Rad). scATAC-Seq libraries were prepared using the SureCell ATAC-Seq Library Prep Kit (17004620, Bio-Rad) and SureCell ATAC-Seq Index Kit (12009360, Bio-Rad). Bead barcoding and sample indexing were performed with PCR amplification as follows: 37 °C for 30 min, 85 °C for 10 min, 72 °C for 5 min, 98 °C for 30 s, eight cycles of 98 °C for 10 s, 55 °C for 30 s and 72 °C for 60 s, and a single 72 °C extension for 5 min to finish. Emulsions were broken and products were cleaned up using Ampure XP beads. Barcoded amplicons were further amplified for 8 cycles. PCR products were purified using Ampure XP beads and quantified on an Agilent Bioanalyzer (G2939BA, Agilent) using the High-Sensitivity DNA kit (5067-4626, Agilent). Libraries were sequenced on HiSeq 2000 with 150 bp paired-end reads.

Sequencing reads preprocessing

Sequencing data were processed using the Bio-Rad ATAC-Seq Analysis Toolkit. This Toolkit is a streamlined computational pipeline, including tools for FASTQ debarcoding, read trimming, alignment, bead filtration, bead deconvolution, cell filtration and peak calling. The reference index was built upon the mouse genome mm10. For generation of the fragments file, which contain the start and end genomic coordinates of all aligned sequenced fragments, sorted bam files were further process with "bap-frag" module of BAP (https://github.com/caleblareau/bap). Downstream analysis was performed with SnapATAC [36], cisTopic[57] and ArchR[108] . Resulted sorted bam file of each sample was structured into hdf5 snapfiles (single-nucleus accessibility profiles) using snaptools version 1.4.1. Cell-by-bin matrices were created with bin size of 5,000bp. SnapATAC package was used to merge snapfiles based on the same bin coordinates. Peak coordinates obtained from Bio-Rad ATAC-Seq Analysis Toolkit were loaded on snapfiles to generate cell-by-peak matrix. Cell-by-peak matrix exported from SnapATAC was imported to create cisTopic object. Fragment files were used to create the Arrow file in the ArchR package.

Clustering and gene score/transcription factor activity analysis

We filtered out low-quality nuclei with stringent selection criteria, including read depth per cell (>2,000) and TSS enrichment score (>20%). Potential doubles were further removed based on the ArchR method. Bin regions were cleaned by eliminating bins overlapping with ENCODE Blacklist regions, mitochondrial DNA as well as the top 5% of invariant features (house-keeping gene promoters). Dimensionality reduction was performed using a nonlinear diffusion map algorithm available in the SnapATAC 1.0.0 package to produce 50 eigen-vectors of which the first 10 were selected in order to generate K Nearest Neighbor graph with K = 15. The clustering was performed using Leiden Algorithm available in the package leidenalg version 0.7.0 with a resolution of 0.4 and UMAP embedding were generated using umap-learn. ArchR was used to estimate gene expression for genes and TF motif activity from single cell chromatin accessibility data. Gene scores were calculated using the addGeneScoreMatrix() function with gene score models implemented in ArchR. addDeviationsMatrix() function was used to compute enrichment of TF activity on a per-cell basis across all motif annotations based on Chromvar.

cisTopic analysis

Topic modelling was performed using 10 to 35 (1 by 1), 40 and 500 iterations with cisTopics. The model was selected based on the highest log-likelihood. The cell-topic UMAP representation was obtained by using UMAP on the normalized topic-cell matrix (by Probability). Cell clustering result (6 clusters) from SnapATAC was used for visualization. Heatmap was then generated based on the cell-cisTopic distributions to identify topics associated with each cluster. To select a representative set of regions of the topic, Region-topic distributions were binarized with a probability threshold of 0.99 to identify the top contributing regions in each topic. For identifying enriched GO terms per topic, the binarized topics (i.e. sets of top regions per topic) was analyzed over GREAT or clusterProfiler[109].

Trajectory analysis

Trajectory analysis was performed in ArchR. addTrajectory() function in ArchR was used to construct trajectory on cisTopic UMAP embedding. To perform integrative analyses for identification of positive TF regulators by integration of gene scores with motif accessibility across pseudo-time, we used the correlateTrajectories() function which takes two SummarizedExperiment objects retrieved from the getTrajectories() function.

Enhancer analysis

To connect distal regulatory elements to promoters, we first used the addCoAccessibility() function in ArchR to calculate co-accessibility. To select links connecting distal regulatory element and promoter, two peaks in each pair were annotated by ChIPseeker separately and the links with only one peak falling in the promoter region were retained. To identify enhancers, datasets of H3K4me3, H3K27me3, H3K4me1, H3K4me2 and H3K27ac modifications in Id4-GFP-bright SSC were downloaded from GSE131656[56]. ChromHMM was used to train a 12 state model on all histone marks assayed. Chromatin states of DARs among different clusters, topic regions and distal element-promoter links were annotated using annotatr package [110], and candidate regions overlap "Active enhancer" states were considered as putative enhancers. ROSE was used for the identification of super-enhancers with H3K27ac dataset (GSE131656)[56]. Overlapping regions of the resulting sets of super-enhancers and CCANs were evaluated using makeVennDiagram in ChIPpeakAnno. CCANs and super-enhancers were associated with nearby genes using GREAT analysis with the default setting[111].

Footprinting analysis

Differential transcription factor footprints between control and UNC0638 treatment group were identified using the Regulatory Genomics Toolbox application HINT[75]. Aligned BAM files from control or UNC0638 treatment group were treated as pseudo-bulk ATAC-seq profiles and then subjected to rgt-hint analysis. Based on MACS calling peaks, we used HINT-ATAC to predict footprints with the “rgt-hint footprinting” command. We then identified all binding sites of a particular TF overlapping with footprints by using its motif from JASPAR with “rgt-motifanalysis matching” command. Differential motif occupancy was identified with “rgt-hint differential” command and ”–bc” was specified to use the bias-corrected signal.

Real-time PCR

Total RNA was extracted from cells by AllPrep DNA/RNA Micro Kit (QIAGEN). RNA was then converted to cDNA by reverse transcription using PrimeScript RT Master Mix (Takara). Real-time PCR was performed using Power SYBR® Green PCR Master Mix (Life Technology) following manufacturer's instructions. Primers used: Cdh1 forward (5'-3') CAGGTCTCCTCATGGCTTTGC, reverse (5'-3') CTTCCGAAAAGAAGGCTGTCC; Vim forward (5'-3') CCTTTACTGCAGTTTTTCAGG, reverse (5'-3') GTATTCTAGCACAAGATTTC.

Cell proliferation analysis

For each well of 24-well plate, 10^5 cells were seeded and cultured in complete GSC medium. Each sample contains at least two technical replicates to minimize errors. The number of cells was counted on day 7 and day 14 of the experiment.

Cell invasion assay

Invasion assay was performed according to manufacturer's instructions (Corning). In brief, a layer of Matrigel (Corning) was coated on the transwell insert containing a polycarbonate membrane with 8 μm pores (Corning Costar). A single cell suspension containing 105 cells was seeded on the transwell membrane and the transwell insert was placed into a 24-well plate that contained complete medium. After incubation for 48 hours, cells in the upper chamber were removed and the transwell insert was fixed in 4% PFA for 30 minutes. Cells that migrated to the transwell membrane were stained with 0.1% crystal violet for 10 minutes and then observed under a light microscope. At least 6 fields were counted for each sample. The migration ratio was calculated by the area of cells divided by the total area of the visual field using the ImageJ software (NIH).

Spontaneous EMT assay and small molecule inhibitor treatment

Matrigel (Corning) was diluted with DMEM-F12 (Invitrogen) to reach a final concentration of 1-100 μg/ml. 250 μl of Matrigel solution was added to each well of a 24 well tissue culture plate and the plate was placed in 4°C overnight. The final Matrigel concentration was 0.1-10 μg/cm2. After coating overnight, the Matrigel solution was aspirated and the well was washed once with PBS before seeding cells. GSCs were seeded on Matrigel (0.1 μg/cm2) to induce domed-shaped colonies. We then transferred the colonies to Matrigel (10 μg/cm2) after which cells at the border could sense adjacent unoccupied space and spontaneously undergo EMT and migrate outwards. The proportion and distance that cells migrated were observed.

Small molecule inhibitors were immediately added after colony transfer and the cells were treated for 24 to 48 hours. RepSox (ab142139, Abcam), UNC0638 (ab142157, Abcam), Nintedanib (HY-50904, MedChemExpress LLC) and Erlotinib (SML2156, Sigma) were dissolved in DMSO. The final concentration of each inhibitor was: RepSox 5 μM, UNC0638 2 μM, Nintedanib 0.5 μM and Erlotinib 5 μM. Medium containing DMSO was used as a control.

In vitro small molecule inhibitor treatment and Fluorescence-activated cell sorting (FACS)

For small molecule inhibitor treatment subjected to flow cytometry, complete GSC media containing freshly diluted inhibitors were refreshed every 2 days. Medium containing DMSO was used as a control. Flow cytometry was performed after 4 days of drug treatment. For retinoic acid (RA) treatment, GSCs were cultured in complete GSC medium containing 0.5 μM RA (R2625, Sigma) for 20 hours before flow cytometry analysis. Medium containing DMSO was used as a control.

GSCs were subjected to flow cytometry analysis after trypsin digestion. For analysis, cells were incubated for 20 min at 4°C in staining buffer with Alexa Fluor 647 anti-mouse CDH1 antibody (147307, Biolegend) or PerCP-Vio700 anti-mouse CD90.2 antibody (130-102-204, Miltenyi Biotec). Flow cytometry was performed on a BD FACSAria Fusion Flow Cytometer (BD Biosciences). Data were analyzed by FlowJo software (FlowJo, LLC).

Immunofluorescence staining and confocal microscopy

Cells cultured on Matrigel-coated coverslips or testis sections on microscope slides were fixed in 4% (v/v) paraformaldehyde for 15 min at room temperature and rinsed three times in PBS for 5 min before staining. Cells were permeabilized by treating in PBS containing 0.2% Triton X-100 for 15 min. Cells were blocked for 2 hours in PBS containing 10% normal donkey serum (NDS) (Jackson Immunoresearch). Primary and secondary antibodies and respective dilutions were listed in Supplementary Table S2. Primary antibodies were applied overnight at 4°C. Appropriate secondary antibodies were applied for one hour at room temperature. Cell nuclei were counterstained with DAPI (1 μg/ml, Sigma). All confocal images were captured by Olympus FV1200 confocal microscope. FV10-ASW software (Olympus) was employed to create overlays of colors.

Statistical analysis

Statistical analysis. Assessment of statistical significance was performed using two-tailed unpaired t-tests, one-way ANOVA with Tukey multiple comparisons tests or Chi-squared tests. Statistical analysis was performed using GraphPad Prism v8. Associated P values are indicated as follows: *$P < 0.05$; **$P < 0.01$; ***$P < 0.001$; ****$P < 0.0001$; not significant (ns) $P > 0.05$.

Data availability

All data in this study have been deposited in the Gene Expression Omnibus (GEO). Histone from ref[56] are available from GSE131656. Hi-C data from ref[64] is available from GSE147536.

Author Contributions

JL, HCS, RMH and TLL conceived and designed the study. JL and HCS performed experiments. JL, HCS, RMH and TLL analyzed and interpreted data. JL, HCS, RMH and TLL wrote the paper. JL, HCS, ACSL, WTL, KWJN, TTHC, RMH and TLL reviewed and edited the manuscript.

Acknowledgments

We thank Tommy Lo for scATAC-Seq experimental assistance. We also thank the SBS Core Laboratories in The Chinese University of Hong Kong for technical support. This study was supported by General Research Fund (CUHK 14120418) to TLL.

Figure legends

Figure 1. Gene expression signatures of epithelial-like (EL) and mesenchymal-like (ML) germline stem cells (GSCs).

A. Fluorescence images of GSCs cultured on Matrigel show heterogeneity in the Oct4-GFP level.

B. Gene Ontology (GO) terms significantly enriched in each cluster identified from Mfuzz soft clustering ($p \leq 0.05$). Representative genes for each cluster are listed in the right column. Cluster 4 and 7 showed no significant enrichment in GO terms.

C. Bright-field images of GSC cultured on different concentrations of Matrigel coating after 14 days showing distinct domed and flat morphology. Scale bars: 100 μm.

D. Quantification of CDH1+ proportion in domed GSCs and flat GSCs by flow cytometry analysis ($p < 0.05$, $n = 3$). Significance was determined by an unpaired two-tailed t-test.

E. Representative bright-field and immunofluorescence images of domed and flat GSC colonies from at least 5 images per group. Scale bars: 20 μm.

F. Matrigel-precoated membrane filter insert was used to measure in vitro invasiveness. After 48 hours of incubation, the area invaded by cells was measured. Quantification of invasive cells ($p < 0.001$, $n \geq 3$). Significance was determined by an unpaired two-tailed t-test.

G. Heatmap of 381 differentially expressed genes (DEGs) identified by comparing EL-GSC and ML-GSC ($p < 0.05$, $FC > 2$).

H. Top 5 upstream regulators predicted by Ingenuity pathway analysis (IPA).

I. Gene ontology (GO) terms significantly enriched for DEGs upregulated in ML-GSCs compared with EL-GSCs (Red: Molecular Function, Blue: Biological Process, Green: Cellular Component) ($p \leq 0.05$).

J. GSEA showing the enrichment of EMT Hallmark gene set and Extracellular matrix organization gene set in ML-GSC compared to EL-GSC.

K. The expression level of genes involved in TGFβ-induced EMT pathway in EL-GSCs and ML-GSCs. * FDR < 0.01. ** FDR < 0.05. *** FDR < 0.001.

Figure 2. Single-cell chromatin state of germline stem cells (GSCs).

A. The workflow of scATAC-seq to measure single nuclei accessibility on BioRad SureCell

ATAC-Seq platform.

B. Quality assessment metrics for scATAC-Seq libraries. The fragment size distribution plot shows enrichment around 100 and 200 bp, indicating nucleosome-free and mono-nucleosome-bound fragments (Left panel). Plot of TSS enrichment score versus the total number of unique fragments. Only cells lying in the upper right quadrant (marked by dashed lines) are retained for downstream analysis (Right panel).

C. De novo identification of cell states by SnapATAC visualized using uniform manifold approximation and projection (UMAP).

D. Gene activity score of EMT marker genes and SSC self-renew and differentiation genes shown in UMAP.

E. Open chromatin signatures of EL-GSC and ML-GSC identified from bulk ATAC-seq were enriched in distinct locations shown in UMAP.

F. Variability of TF motifs calculated by chromVar.

G. Violin plot (Upper panel) and UMAP (Lower panel) showing the motif accessibility of five of the most variable motifs.

Figure 3. Pseudotime analysis revealed dynamic TF regulation on the EMT-like process in GSC.

A. De novo identification of two pseudotime trajectories capturing EMT-like process visualized using UMAP in 3D space.

B. scATAC-Seq profiles are ordered by pseudotime, corresponding to the epithelial-mesenchymal axis starting from SSC-1 subset to Diff-SPG subset through Prog-1/2/3 subset (Left) or SSC-2 subset (Right).

C. Gene score of Cdh1 and Vim in individual cells ordered by pseudotime in Trajectory 1 (Left) and 2 (Right).

D. Smoothened heatmap showing dynamic gene score (Left panel) and motif accessibility (Right panel) of indicated TFs along pseudotime for gene-motif pairs in Trajectory 1.

E. Smoothened heatmap showing dynamic gene score (Left panel) and motif accessibility (Right panel) of indicated TFs along pseudotime for gene-motif pairs in Trajectory 2.

F. Gene score (Left panel) and motif accessibility (Right panel) of selected TFs identified in D and E ordered by pseudotime. Colored lines correspond to the trajectories shown in 3D UMAP above.

Figure 4. Regulatory topics identified by chromatin accessibility associated with E-M heterogeneity.

A. cisTopic cell UMAP colored by the SnapATAC cluster assignment.

B. Heatmap showing Topic-cell enrichment revealed by cisTopic analysis.

C. cisTopic cell UMAP color-coded by the normalized topic score of selected EMT associated topics.

D. GREAT analysis of regions included in epithelial-like and mesenchymal-like topics.

E. Genomic distribution of epithelial-like and mesenchymal-like topic regions revealed by ChromHMM.

F. Aggregated scATAC-seq profiles showing three regions from epithelial-like topics and three regions from mesenchymal-like topics. ChromHMM tracks indicate chromatin-state based on histone modification profile from Id4-GFP-bright and Id4-GFP-dim spermatogonia (GSE131656). Emission profile (Right) from a 12-State model is based on the five histone modifications studied.

Figure 5. Cis-coaccessibility analysis revealed putative enhancers in GSC.

A. The schematic diagram illustrating the approach to link enhancer to target gene.

B. The Circos plot represents significantly enriched pathways and GO terms for biological process associated with putative enhancer-promoter links identified. Outside the circle, genes and significantly enriched pathways together with GO terms are indicated.

C. Inferred links for *Notch1*, *Spry4*, *Lin28a*, *Bmp7*. Red and blue lines indicate enhancer-promoter links and other links respectively. ChromHMM tracks indicate chromatin-states related to Figure 4.

D. Call of SEs by Rank ordering of super-enhancers (ROSE), which takes into account enhancer ranks by H3K27ac signals from Id4-GFP-bright spermatogonia ChIP-seq profile (GSE131656).

E. Venn diagram showing significant overlapping of super-enhancers identified from ROSE and high density cis co-accessibility networks (CCANs).

F. Aggregated scATAC-seq profiles showing SE regions near *Nanos2* with inferred TAD domains[64] and co-accessible CREs.

G. Aggregated scATAC-seq profiles showing SE regions near *Tgfb1* with inferred TAD domains[64] and co-accessible CREs.

Figure 6. Role of G9a in the GSC EMT-like process.

A. Representative bright-field and immunofluorescence images of Matrigel-based GSC culture stained for G9a and H3K9me2 (from at least 5 images per group). The arrowheads indicate flat cells with stronger G9a and H3K9me2 protein levels. Scale bars: 20 μm.

B. Flow cytometry analysis of epithelial marker CDH1 in UNC0638-treated GSCs.

C. A schematic diagram of in vitro assay for studying spontaneous EMT progression. Domed GSC colonies (green) are isolated from low concentration Matrigel culture and then transferred to high concentration Matrigel culture. The colonies are subsequently treated with small molecule inhibitors and cell migration is monitored (blue).

D. Bright-field images of domed GSC colonies treated with small molecule inhibitors (RepSox 5 μM, UNC0638 2 μM) or vehicle (DMSO) for 24 hours. Dash lines indicate colony boundary at 0 hour.

E. Differentially expressed genes were identified among the control group and UNC0638 treated groups. *Z*-score-normalized ratios of the gene expression of 435 DEGs were subjected to soft clustering with Mfuzz software. The 4 clusters represent different expression profiles.

F. Gene Ontology (GO) terms significantly enriched in each cluster identified from Mfuzz soft clustering ($p \leq 0.05$). Representative genes for each cluster are listed in the right column.

G. GSEA showing the enrichment of EMT Hallmark gene set and TGF-β signaling Hallmark gene set in non-UNC0638-treated GSCs.

H. Venn diagram showing significant overlapping of DEGs that upregulated in ML-GSC (ML-GSC vs EL-GSC) and downregulated after G9a inhibition.

I. Schematic diagram in vivo G9a inhibition. UNC0638 was administered in neonatal mice on PND3-5 and samples were collected on PND6.

J. Representative confocal images of testis sections from UNC0638-treated and control (DMSO) mice at PND6. Oct4-GFP expression was observed in undifferentiated germ cells. Cell nuclei were stained with DAPI. Arrows indicate germ cells not attached to the basement membrane. Scale bars 20 μm.

K. Quantification of percentage of Oct4-GFP+ spermatogonia attached to basement membrane on PND6 ($p < 0.001$, UNC0638 = 6 testes, DMSO = 3 testes). Significance was determined by an unpaired two-tailed t-test.

Figure 7. Chromatin accessibility changes associated with G9a inhibition.

A. Normalized pseudo-bulk accessibility tracks showing differential chromatin accessibility of selected EMT associated genes between control and UNC0638 treatment group.

B. Scatter plot showing the transcription factor (TF) activity dynamics between control and UNC0638 treatment group. Y-axis represents the differences in TF activity and the x-axis indicates the number of cleavage events (log10 transformed). TFs with significant differential activity identified by HINT are colored as red (adjusted p-value <0.05 and the number of cleavage events > 150).

C. Transcription factor footprints (average ATAC-seq signal around predicted binding sites) for selected TFs. Logo of underlying sequences is shown below.

D. UMAP projection of scATAC-seq profiles generated by SnapATAC from control and UNC0638-treated GSCs.

E. Motif activity of HES1 in control and UNC0638-treated GSCs shown in UMAP. Logo of the HES1 motif is shown on the right.

F. Western blotting of HES1 in control (DMSO) and UNC0638-treated (2 μM) cells at 26, 50 and 64 hours after treatment.

G. Heatmap showing Topic-cell enrichment combining both the control and UNC0638 treatment groups revealed by cisTopic analysis.

H. Violin plot of the normalized topic score of topics that lose accessibility (Topic 2, 7 and 14) or gain accessibility (Topic 3 and 15) during UNC0618 treatment.

I. GO analysis of regions including in topics with differential accessibility between control and UNC0638 treatment group.

J. Schematic diagram illustrating that the migration of gonocyte towards the basement membrane to establish SSC pool requires G9a to regulate *Hes1* (Top). This process is inhibited upon the addition of G9a inhibitor UNC0638 (Bottom).

References

1. Culty, M. Gonocytes, the forgotten cells of the germ cell lineage. *Birth Defects Res. C Embryo Today* **87,** 1–26 (2009).
2. Manku, G. & Culty, M. Mammalian gonocyte and spermatogonia differentiation: recent advances and remaining challenges. *Reproduction* **149,** R139-57 (2015).
3. Saitou, M. & Yamaji, M. Primordial germ cells in mice. *Cold Spring Harb. Perspect. Biol.* **4,** (2012).
4. Yoshida, S., Sukeno, M. & Nabeshima, Y.-I. A vasculature-associated niche for undifferentiated spermatogonia in the mouse testis. *Science* **317,** 1722–1726 (2007).
5. Kitadate, Y. *et al.* Competition for mitogens regulates spermatogenic stem cell homeostasis in an open niche. *Cell Stem Cell* **24,** 79-92.e6 (2019).
6. Pradella, D., Naro, C., Sette, C. & Ghigna, C. EMT and stemness: flexible processes tuned by alternative splicing in development and cancer progression. *Mol. Cancer* **16,** 8 (2017).
7. Jolly, M. K. *et al.* Stability of the hybrid epithelial/mesenchymal phenotype. *Oncotarget* **7,** 27067–27084 (2016).
8. Chen, T., You, Y., Jiang, H. & Wang, Z. Z. Epithelial-mesenchymal transition (EMT): A biological process in the development, stem cell differentiation, and tumorigenesis. *J. Cell. Physiol.* **232,** 3261–3272 (2017).
9. Goossens, S. *et al.* The EMT regulator Zeb2/Sip1 is essential for murine embryonic hematopoietic stem/progenitor cell differentiation and mobilization. *Blood* **117,** 5620–5630 (2011).
10. Chng, Z., Teo, A., Pedersen, R. A. & Vallier, L. SIP1 mediates cell-fate decisions between neuroectoderm and mesendoderm in human pluripotent stem cells. *Cell Stem*

Cell **6,** 59–70 (2010).

11. van Roy, F. & Berx, G. The cell-cell adhesion molecule E-cadherin. *Cell. Mol. Life Sci.* **65,** 3756–3788 (2008).
12. Serrano-Gomez, S. J., Maziveyi, M. & Alahari, S. K. Regulation of epithelial-mesenchymal transition through epigenetic and post-translational modifications. *Mol. Cancer* **15,** 18 (2016).
13. Liu, D. *et al.* Identification and characterization of mesenchymal-epithelial progenitor-like cells in normal and injured rat liver. *Am. J. Pathol.* **185,** 110–128 (2015).
14. Lin, Y., Ma, Q., Li, L. & Wang, H. The CXCL12-CXCR4 axis promotes migration, invasiveness, and EMT in human papillary thyroid carcinoma B-CPAP cells via NF-κB signaling. *Biochem. Cell Biol.* **96,** 619–626 (2018).
15. Zheng, C.-H., Chen, X.-M., Zhang, F.-B., Zhao, C. & Tu, S.-S. Inhibition of CXCR4 regulates epithelial mesenchymal transition of NSCLC via the Hippo-YAP signaling pathway. *Cell Biol. Int.* **42,** 1386–1394 (2018).
16. Wei, C., Lin, H. & Cui, S. The forkhead transcription factor FOXC2 is required for maintaining murine spermatogonial stem cells. *Stem Cells Dev.* **27,** 624–636 (2018).
17. He, Y., Xie, H., Yu, P., Jiang, S. & Wei, L. FOXC2 promotes epithelial-mesenchymal transition and cisplatin resistance of non-small cell lung cancer cells. *Cancer Chemother. Pharmacol.* **82,** 1049–1059 (2018).
18. Hollier, B. G. *et al.* FOXC2 expression links epithelial-mesenchymal transition and stem cell properties in breast cancer. *Cancer Res.* **73,** 1981–1992 (2013).
19. Liao, J. *et al.* Revealing cellular and molecular transitions in neonatal germ cell differentiation using single cell RNA sequencing. *Development* **146,** (2019).
20. Santibanez, J. F. Transforming growth factor-Beta and urokinase-type plasminogen activator: dangerous partners in tumorigenesis-implications in skin cancer. *ISRN*

Dermatol **2013,** 597927–597927 (2013).

21. Hammoud, S. S. *et al.* Transcription and imprinting dynamics in developing postnatal male germline stem cells. *Genes Dev.* **29,** 2312–2324 (2015).
22. La, H. M. *et al.* Identification of dynamic undifferentiated cell states within the male germline. *Nat. Commun.* **9,** 2819 (2018).
23. Hermann, B. P. *et al.* The Mammalian Spermatogenesis Single-Cell Transcriptome, from Spermatogonial Stem Cells to Spermatids. *Cell Rep.* **25,** 1650-1667.e8 (2018).
24. Choi, N. Y. *et al.* A novel feeder-free culture system for expansion of mouse spermatogonial stem cells. *Mol. Cells* **37,** 473–479 (2014).
25. Paye, A. *et al.* EGFR activation and signaling in cancer cells are enhanced by the membrane-bound metalloprotease MT4-MMP. *Cancer Res.* **74,** 6758–6770 (2014).
26. Liu, Y. J., Xu, Y. & Yu, Q. Full-length ADAMTS-1 and the ADAMTS-1 fragments display pro- and antimetastatic activity, respectively. *Oncogene* **25,** 2452–2467 (2006).
27. Ishii, K., Kanatsu-Shinohara, M., Toyokuni, S. & Shinohara, T. FGF2 mediates mouse spermatogonial stem cell self-renewal via upregulation of Etv5 and Bcl6b through MAP2K1 activation. *Development* **139,** 1734–1743 (2012).
28. Liu, B. *et al.* TLR2 promotes human intrahepatic cholangiocarcinoma cell migration and invasion by modulating NF-κB pathway-mediated inflammatory responses. *FEBS J.* **283,** 3839–3850 (2016).
29. Leight, J. L., Wozniak, M. A., Chen, S., Lynch, M. L. & Chen, C. S. Matrix rigidity regulates a switch between TGF-β1-induced apoptosis and epithelial-mesenchymal transition. *Mol. Biol. Cell* **23,** 781–791 (2012).
30. Chen, X.-F. *et al.* Transforming growth factor-β1 induces epithelial-to-mesenchymal transition in human lung cancer cells via PI3K/Akt and MEK/Erk1/2 signaling pathways. *Mol. Biol. Rep.* **39,** 3549–3556 (2012).

31. Li, C.-L. *et al.* Fibronectin induces epithelial-mesenchymal transition in human breast cancer MCF-7 cells via activation of calpain. *Oncol. Lett.* **13,** 3889–3895 (2017).

32. Li, X.-L. *et al.* Integrin β4 promotes cell invasion and epithelial-mesenchymal transition through the modulation of Slug expression in hepatocellular carcinoma. *Sci. Rep.* **7,** 40464 (2017).

33. Xu, J., Lamouille, S. & Derynck, R. TGF-beta-induced epithelial to mesenchymal transition. *Cell Res.* **19,** 156–172 (2009).

34. Massagué, J. TGFbeta in Cancer. *Cell* **134,** 215–230 (2008).

35. Lamouille, S., Xu, J. & Derynck, R. Molecular mechanisms of epithelial-mesenchymal transition. *Nat. Rev. Mol. Cell Biol.* **15,** 178–196 (2014).

36. Fang, R. *et al.* Fast and Accurate Clustering of Single Cell Epigenomes Reveals *Cis* -Regulatory Elements in Rare Cell Types. *BioRxiv* (2019). doi:10.1101/615179

37. Chen, H. *et al.* Assessment of computational methods for the analysis of single-cell ATAC-seq data. *Genome Biol.* **20,** 241 (2019).

38. Schep, A. N., Wu, B., Buenrostro, J. D. & Greenleaf, W. J. chromVAR: inferring transcription-factor-associated accessibility from single-cell epigenomic data. *Nat. Methods* **14,** 975–978 (2017).

39. Li, H. J. *et al.* NKX6.1 functions as a metastatic suppressor through epigenetic regulation of the epithelial-mesenchymal transition. *Oncogene* **35,** 2266–2278 (2016).

40. Zhang, Y., Feng, X. H. & Derynck, R. Smad3 and Smad4 cooperate with c-Jun/c-Fos to mediate TGF-beta-induced transcription. *Nature* **394,** 909–913 (1998).

41. Matson, C. K. *et al.* The mammalian doublesex homolog DMRT1 is a transcriptional gatekeeper that controls the mitosis versus meiosis decision in male germ cells. *Dev. Cell* **19,** 612–624 (2010).

42. Williams, K. A. *et al.* Prostate cancer susceptibility gene HIST1H1A is a modulator of

androgen receptor signaling and epithelial to mesenchymal transition. *Oncotarget* **9,** 28532–28546 (2018).

43. Zhou, Z. J. *et al.* MACROD2 deficiency promotes hepatocellular carcinoma growth and metastasis by activating GSK-3β/β-catenin signaling. *NPJ Genom Med* **5,** 15–15 (2020).

44. Hogarth, C. A., Mitchell, D., Small, C. & Griswold, M. EGR4 displays both a cell- and intracellular-specific localization pattern in the developing murine testis. *Dev. Dyn.* **239,** 3106–3114 (2010).

45. Tourtellotte, W. G., Nagarajan, R., Auyeung, A., Mueller, C. & Milbrandt, J. Infertility associated with incomplete spermatogenic arrest and oligozoospermia in Egr4-deficient mice. *Development* **126,** 5061–5071 (1999).

46. Yu, J. *et al.* Identification of seven genes essential for male fertility through a genome-wide association study of non-obstructive azoospermia and RNA interference-mediated large-scale functional screening in Drosophila. *Hum. Mol. Genet.* **24,** 1493–1503 (2015).

47. Chang, K.-C., Shieh, B. & Petrash, J. M. Influence of aldose reductase on epithelial-to-mesenchymal transition signaling in lens epithelial cells. *Chem. Biol. Interact.* **276,** 149–154 (2017).

48. Hou, Y. *et al.* DNA demethylation and USF regulate the meiosis-specific expression of the mouse Miwi. *PLoS Genet.* **8,** e1002716 (2012).

49. Dolfini, D., Andrioletti, V. & Mantovani, R. Overexpression and alternative splicing of NF-YA in breast cancer. *Sci. Rep.* **9,** 12955 (2019).

50. Bezzecchi, E., Ronzio, M., Dolfini, D. & Mantovani, R. NF-YA Overexpression in Lung Cancer: LUSC. *Genes (Basel)* **10,** (2019).

51. Guo, J. *et al.* Chromatin and Single-Cell RNA-Seq Profiling Reveal Dynamic Signaling

and Metabolic Transitions during Human Spermatogonial Stem Cell Development. *Cell Stem Cell* **21,** 533-546.e6 (2017).

52. Okada, R. *et al.* Expression profile of NOTCH3 in mouse spermatogonia. *Cells Tissues Organs (Print)* **204,** 283–292 (2017).

53. Li, X., Cao, Y., Li, M. & Jin, F. Upregulation of HES1 promotes cell proliferation and invasion in breast cancer as a prognosis marker and therapy target via the AKT pathway and EMT process. *J. Cancer* **9,** 757–766 (2018).

54. Hollenhorst, P. C. RAS/ERK pathway transcriptional regulation through ETS/AP-1 binding sites. *Small GTPases* **3,** 154–158 (2012).

55. Gautam, M., Mathur, A., Khan, M. A., Majumdar, S. S. & Rai, U. Transcriptome analysis of spermatogenically regressed, recrudescent and active phase testis of seasonally breeding wall lizards Hemidactylus flaviviridis. *PLoS ONE* **8,** e58276 (2013).

56. Cheng, K. *et al.* Unique epigenetic programming distinguishes regenerative spermatogonial stem cells in the developing mouse testis. *iScience* 101596 (2020). doi:10.1016/j.isci.2020.101596

57. Bravo González-Blas, C. *et al.* cisTopic: cis-regulatory topic modeling on single-cell ATAC-seq data. *Nat. Methods* **16,** 397–400 (2019).

58. Hasegawa, K. & Saga, Y. FGF8-FGFR1 signaling acts as a niche factor for maintaining undifferentiated spermatogonia in the mouse. *Biol. Reprod.* **91,** 145 (2014).

59. Faisal, I. *et al.* Transcription factor USF1 is required for maintenance of germline stem cells in male mice. *Endocrinology* **160,** 1119–1136 (2019).

60. Shlyueva, D., Stampfel, G. & Stark, A. Transcriptional enhancers: from properties to genome-wide predictions. *Nat. Rev. Genet.* **15,** 272–286 (2014).

61. Pliner, H. A. *et al.* Cicero Predicts cis-Regulatory DNA Interactions from Single-Cell

Chromatin Accessibility Data. *Mol. Cell* **71,** 858-871.e8 (2018).

62. Souza, N. P. *et al.* Isolation and molecular characterization of spermatogonia from male Sprague-Dawley rats exposed in utero and postnatally to dibutyl phthalate or acrylamide. *Toxicol. Mech. Methods* **29,** 488–498 (2019).

63. Chakraborty, P. *et al.* LIN28A marks the spermatogonial progenitor population and regulates its cyclic expansion. *Stem Cells* **32,** 860–873 (2014).

64. Luo, Z. *et al.* Reorganized 3D genome structures support transcriptional regulation in mouse spermatogenesis. *iScience* **23,** 101034 (2020).

65. Xu, J. *et al.* AIP1-mediated actin disassembly is required for postnatal germ cell migration and spermatogonial stem cell niche establishment. *Cell Death Dis.* **6,** e1818 (2015).

66. Peixoto, P. *et al.* EMT is associated with an epigenetic signature of ECM remodeling genes. *Cell Death Dis.* **10,** 205 (2019).

67. Stowers, R. S. *et al.* Matrix stiffness induces a tumorigenic phenotype in mammary epithelium through changes in chromatin accessibility. *Nat. Biomed. Eng.* **3,** 1009–1019 (2019).

68. Liu, S. *et al.* G9a is essential for EMT-mediated metastasis and maintenance of cancer stem cell-like characters in head and neck squamous cell carcinoma. *Oncotarget* **6,** 6887–6901 (2015).

69. Vedadi, M. *et al.* A chemical probe selectively inhibits G9a and GLP methyltransferase activity in cells. *Nat. Chem. Biol.* **7,** 566–574 (2011).

70. Jechlinger, M. *et al.* Autocrine PDGFR signaling promotes mammary cancer metastasis. *J. Clin. Invest.* **116,** 1561–1570 (2006).

71. Duan, L. *et al.* VEGFC/VEGFR3 axis mediates TGFβ1-induced epithelial-to-mesenchymal transition in non-small cell lung cancer cells. *PLoS ONE* **13,** e0200452

(2018).

72. Tulchinsky, E., Demidov, O., Kriajevska, M., Barlev, N. A. & Imyanitov, E. EMT: A mechanism for escape from EGFR-targeted therapy in lung cancer. *Biochim. Biophys. Acta Rev. Cancer* **1871,** 29–39 (2019).

73. Tomlinson, D. C., Baxter, E. W., Loadman, P. M., Hull, M. A. & Knowles, M. A. FGFR1-induced epithelial to mesenchymal transition through MAPK/PLCγ/COX-2-mediated mechanisms. *PLoS ONE* **7,** e38972 (2012).

74. Tan, K., Song, H.-W. & Wilkinson, M. F. Single-cell RNAseq analysis of testicular germ and somatic cell development during the perinatal period. *Development* **147,** (2020).

75. Li, Z. *et al*. scOpen: chromatin-accessibility estimation of single-cell ATAC data. *BioRxiv* (2019). doi:10.1101/865931

76. Wang, Z., Li, Y., Kong, D. & Sarkar, F. H. The role of Notch signaling pathway in epithelial-mesenchymal transition (EMT) during development and tumor aggressiveness. *Curr. Drug Targets* **11,** 745–751 (2010).

77. Jiang, Y.-G. *et al*. Role of Wnt/beta-catenin signaling pathway in epithelial-mesenchymal transition of human prostate cancer induced by hypoxia-inducible factor-1alpha. *Int. J. Urol.* **14,** 1034–1039 (2007).

78. Bhowmick, N. A. *et al*. Transforming growth factor-beta1 mediates epithelial to mesenchymal transdifferentiation through a RhoA-dependent mechanism. *Mol. Biol. Cell* **12,** 27–36 (2001).

79. He, N. *et al*. Extracellular Matrix can Recover the Downregulation of Adhesion Molecules after Cell Detachment and Enhance Endothelial Cell Engraftment. *Sci. Rep.* **5,** 10902 (2015).

80. Siu, M. K. Y. & Cheng, C. Y. Extracellular matrix and its role in spermatogenesis. *Adv.*

Exp. Med. Biol. **636,** 74–91 (2008).

81. Mori, S. *et al.* Enhanced Expression of Integrin αvβ3 Induced by TGF-β Is Required for the Enhancing Effect of Fibroblast Growth Factor 1 (FGF1) in TGF-β-Induced Epithelial-Mesenchymal Transition (EMT) in Mammary Epithelial Cells. *PLoS ONE* **10,** e0137486 (2015).

82. Maezawa, S., Yukawa, M., Alavattam, K. G., Barski, A. & Namekawa, S. H. Dynamic reorganization of open chromatin underlies diverse transcriptomes during spermatogenesis. *Nucleic Acids Res.* **46,** 593–608 (2018).

83. Sohni, A. *et al*. The Neonatal and Adult Human Testis Defined at the Single-Cell Level. *Cell Rep*. **26,** 1501-1517.e4 (2019).

84. Wang, H. & Unternaehrer, J. J. Epithelial-mesenchymal Transition and Cancer Stem Cells: At the Crossroads of Differentiation and Dedifferentiation. *Dev. Dyn.* **248,** 10–20 (2019).

85. An, J., Zheng, Y. & Dann, C. T. Mesenchymal to epithelial transition mediated by CDH1 promotes spontaneous reprogramming of male germline stem cells to pluripotency. *Stem Cell Reports* **8,** 446–459 (2017).

86. Song, H.-W. *et al.* The homeobox transcription factor RHOX10 drives mouse spermatogonial stem cell establishment. *Cell Rep.* **17,** 149–164 (2016).

87. Tachibana, M., Nozaki, M., Takeda, N. & Shinkai, Y. Functional dynamics of H3K9 methylation during meiotic prophase progression. *EMBO J.* **26,** 3346–3359 (2007).

88. Tachibana, M. *et al.* G9a histone methyltransferase plays a dominant role in euchromatic histone H3 lysine 9 methylation and is essential for early embryogenesis. *Genes Dev.* **16,** 1779–1791 (2002).

89. Casciello, F., Windloch, K., Gannon, F. & Lee, J. S. Functional role of g9a histone methyltransferase in cancer. *Front. Immunol.* **6,** 487 (2015).

90. Hsiao, S.-M. *et al*. The H3K9 Methyltransferase G9a Represses E-cadherin and is Associated with Myometrial Invasion in Endometrial Cancer. *Ann. Surg. Oncol.* **22 Suppl 3,** S1556-65 (2015).

91. Polyak, K. & Weinberg, R. A. Transitions between epithelial and mesenchymal states: acquisition of malignant and stem cell traits. *Nat. Rev. Cancer* **9,** 265–273 (2009).

92. Wang, S.-C. *et al.* Hes1 triggers epithelial-mesenchymal transition (EMT)-like cellular marker alterations and promotes invasion and metastasis of nasopharyngeal carcinoma by activating the PTEN/AKT pathway. *Oncotarget* **6,** 36713–36730 (2015).

93. Dang, N.-N. *et al.* Abnormal overexpression of G9a in melanoma cells promotes cancer progression via upregulation of the Notch1 signaling pathway. *Aging (Albany NY)* **12,** 2393–2407 (2020).

94. Gao, F. *et al.* Hes1 promotes cell proliferation and migration by activating Bmi-1 and PTEN/Akt/GSK3β pathway in human colon cancer. *Oncotarget* **6,** 38667–38680 (2015).

95. Ohbo, K. *et al.* Identification and characterization of stem cells in prepubertal spermatogenesis in mice. *Dev. Biol.* **258,** 209–225 (2003).

96. Garcia, T. & Hofmann, M.-C. Isolation of undifferentiated and early differentiating type A spermatogonia from Pou5f1-GFP reporter mice. *Methods Mol. Biol.* **825,** 31–44 (2012).

97. Fischer, A. H., Jacobson, K. A., Rose, J. & Zeller, R. Cryosectioning Tissues. *Cold Spring Harb. Protoc.* **2008,** pdb.prot4991-pdb.prot4991 (2008).

98. Picelli, S. *et al.* Full-length RNA-seq from single cells using Smart-seq2. *Nat. Protoc.* **9,** 171–181 (2014).

99. Li, B. & Dewey, C. N. RSEM: accurate transcript quantification from RNA-Seq data with or without a reference genome. *BMC Bioinformatics* **12,** 323 (2011).

100. Robinson, M. D., McCarthy, D. J. & Smyth, G. K. edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics* **26,** 139–140 (2010).

101. Kumar, L. & E Futschik, M. Mfuzz: a software package for soft clustering of microarray data. *Bioinformation* **2,** 5–7 (2007).

102. Chen, J., Bardes, E. E., Aronow, B. J. & Jegga, A. G. ToppGene Suite for gene list enrichment analysis and candidate gene prioritization. *Nucleic Acids Res.* **37,** W305-11 (2009).

103. Subramanian, A. *et al.* Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. *Proc Natl Acad Sci U S A* **102,** 15545–15550 (2005).

104. Corces, M. R. *et al.* An improved ATAC-seq protocol reduces background and enables interrogation of frozen tissues. *Nat. Methods* **14,** 959–962 (2017).

105. Ewels, P. A. *et al.* The nf-core framework for community-curated bioinformatics pipelines. *Nat. Biotechnol.* **38,** 276–278 (2020).

106. Zhang, Y. *et al.* Model-based analysis of ChIP-Seq (MACS). *Genome Biol.* **9,** R137 (2008).

107. Manor, M., Blickstein, I., Ben-Arie, A., Weissman, A. & Hagay, Z. Case series of labor induction in twin gestations with an Intrauterine Balloon catheter. *Gynecol. Obstet. Invest.* **47,** 244–246 (1999).

108. Granja, J. M. *et al.* ArchR: An integrative and scalable software package for single-cell chromatin accessibility analysis. *BioRxiv* (2020). doi:10.1101/2020.04.28.066498

109. Yu, G., Wang, L.-G., Han, Y. & He, Q.-Y. clusterProfiler: an R package for comparing biological themes among gene clusters. *OMICS* **16,** 284–287 (2012).

110. Cavalcante, R. G. & Sartor, M. A. annotatr: genomic regions in context. *Bioinformatics*

33, 2381–2383 (2017).

111. McLean, C. Y. *et al.* GREAT improves functional interpretation of cis-regulatory regions. *Nat. Biotechnol.* **28,** 495–501 (2010).

Figure 1

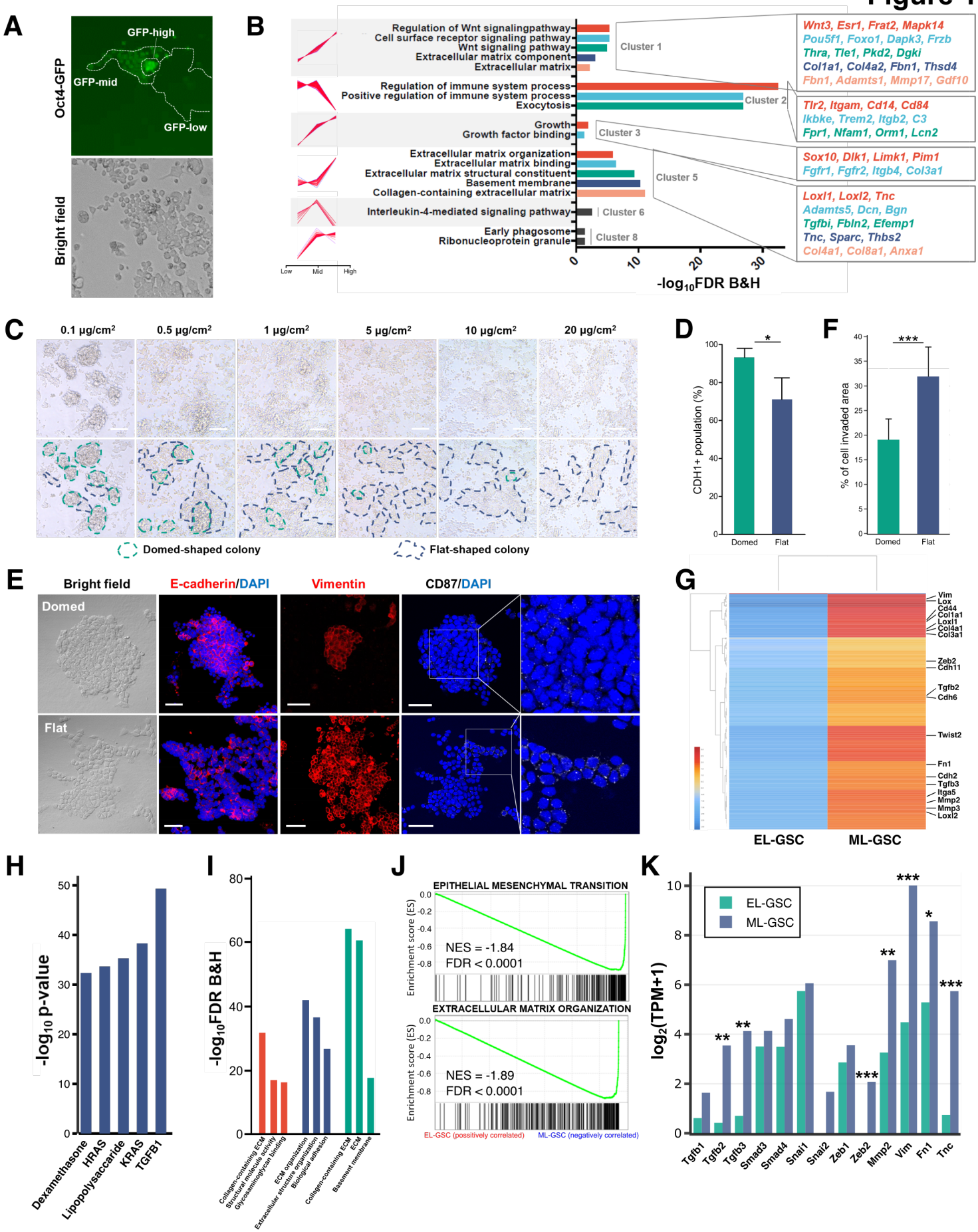

A
Oct4-GFP
GFP-high
GFP-mid
GFP-low
Bright field
B
Regulation of Wnt signalingpathway
Cell surface receptor signaling pathway
Wnt signaling pathway
Extracellular matrix component
Extracellular matrix
Cluster 1
Regulation of immune system process
Positive regulation of immune system process
Exocytosis
Cluster 2
Growth
Growth factor binding
Cluster 3
Extracellular matrix organization
Extracellular matrix binding
Extracellular matrix structural constituent
Basement membrane
Collagen-containing extracellular matrix
Cluster 5
Interleukin-4-mediated signaling pathway
Cluster 6
Early phagosome
Ribonucleoprotein granule
Cluster 8
Low
Mid
High
0
10
20
30
-log10FDR B&H
Wnt3, Esr1, Frat2, Mapk14
Pou5f1, Foxo1, Dapk3, Frzb
Thra, Tle1, Pkd2, Dgki
Col1a1, Col4a2, Fbn1, Thsd4
Fbn1, Adamts1, Mmp17, Gdf10
Tlr2, Itgam, Cd14, Cd84
Ikbke, Trem2, Itgb2, C3
Fpr1, Nfam1, Orm1, Lcn2
Sox10, Dlk1, Limk1, Pim1
Fgfr1, Fgfr2, Itgb4, Col3a1
Loxl1, Loxl2, Tnc
Adamts5, Dcn, Bgn
Tgfbi, Fbln2, Efemp1
Tnc, Sparc, Thbs2
Col4a1, Col8a1, Anxa1
C
0.1 μg/cm2
0.5 μg/cm2
1 μg/cm2
5 μg/cm2
10 μg/cm2
20 μg/cm2
Domed-shaped colony
Flat-shaped colony
D
CDH1+ population (%)
Domed
Flat
F
% of cell invaded area
Domed
Flat
E
Bright field
E-cadherin/DAPI
Vimentin
CD87/DAPI
Domed
Flat
G
Vim
Lox
Cd44
Col1a1
Loxl1
Col4a1
Col3a1
Zeb2
Cdh11
Tgfb2
Cdh6
Twist2
Fn1
Cdh2
Tgfb3
Itga5
Mmp2
Mmp3
Loxl2
EL-GSC
ML-GSC
H
-log10 p-value
Dexamethasone
HRAS
Lipopolysaccaride
KRAS
TGFB1
I
-log10FDR B&H
Collagen-containing ECM
Structural molecule activity
Glycosaminoglycan binding
ECM organization
Extracellular structure organization
Biological adhesion
Collagen-containing ECM
ECM
Basement membrane
J
EPITHELIAL MESENCHYMAL TRANSITION
Enrichment score (ES)
NES = -1.84
FDR < 0.0001
EXTRACELLULAR MATRIX ORGANIZATION
NES = -1.89
FDR < 0.0001
EL-GSC (possitively correlated)
ML-GSC (negatively correlated)
K
EL-GSC
ML-GSC
log2(TPM+1)
Tgfb1
Tgfb2
Tgfb3
Smad3
Smad4
Snai1
Snai2
Zeb1
Zeb2
Mmp2
Vim
Fn1
Tnc

Figure 2

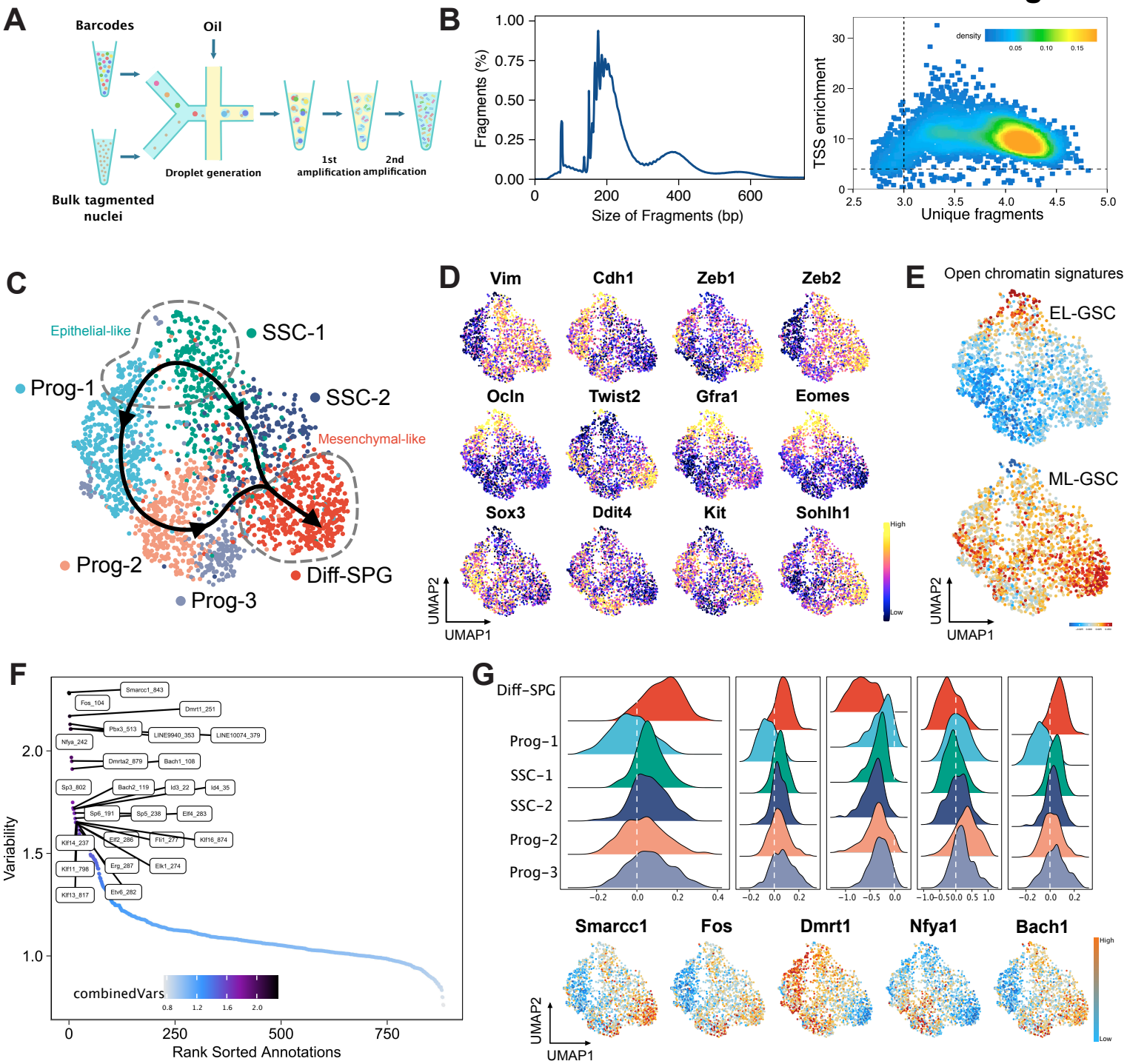

A
Barcodes
Oil
Droplet generation
1st amplification
2nd amplification
Bulk tagmented nuclei
B
Fragments (%)
Size of Fragments (bp)
TSS enrichment
Unique fragments
density
C
Epithelial-like
SSC-1
Prog-1
SSC-2
Mesenchymal-like
Prog-2
Diff-SPG
Prog-3
D
Vim
Cdh1
Zeb1
Zeb2
Ocln
Twist2
Gfra1
Eomes
Sox3
Ddit4
Kit
Sohlh1
High
Low
UMAP2
UMAP1
E
Open chromatin signatures
EL-GSC
ML-GSC
UMAP2
UMAP1
F
Variability
Rank Sorted Annotations
combinedVars
Smarcc1_843
Fos_104
Dmrt1_251
Ptx3_513
LINE9940_353
LINE10074_379
Nfya_242
Dmrta2_879
Bach1_108
Sp3_802
Bach2_119
Id3_22
Id4_35
Sp6_191
Sp5_238
Elf4_283
Klf14_237
Elf2_286
Fli1_277
Klf16_874
Klf11_798
Erg_287
Elk1_274
Klf13_817
Etv6_282
G
Diff-SPG
Prog-1
SSC-1
SSC-2
Prog-2
Prog-3
Smarcc1
Fos
Dmrt1
Nfya1
Bach1
High
Low
UMAP2
UMAP1

Figure 3

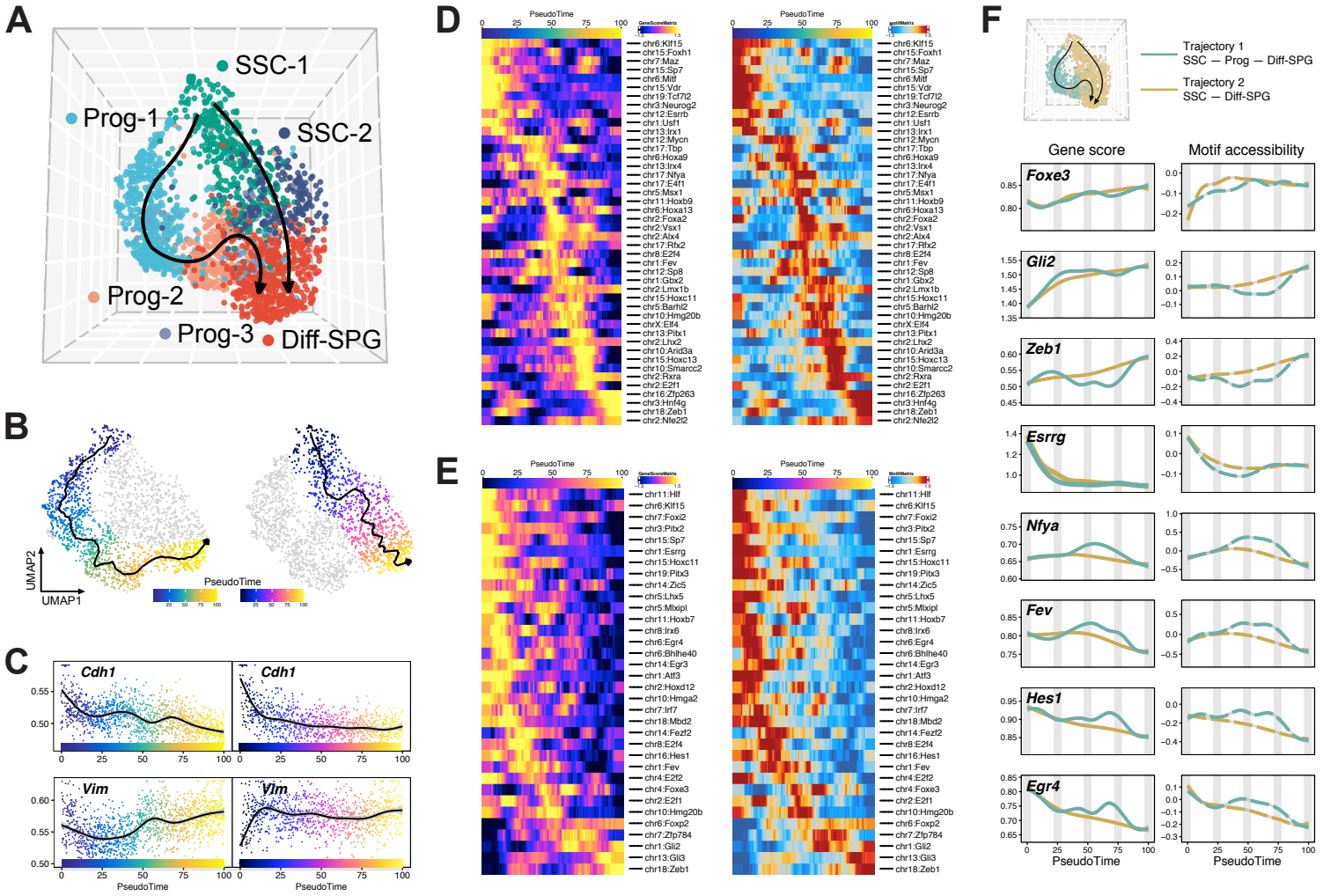

A
SSC-1
Prog-1
SSC-2
Prog-2
Prog-3
Diff-SPG
B
UMAP2
UMAP1
PseudoTime
C
Cdh1
Vim
PseudoTime
D
E
F
Trajectory 1
SSC — Prog — Diff-SPG
Trajectory 2
SSC — Diff-SPG
Gene score
Motif accessibility
Foxe3
Gli2
Zeb1
Esrrg
Nfya
Fev
Hes1
Egr4
PseudoTime

Figure 4

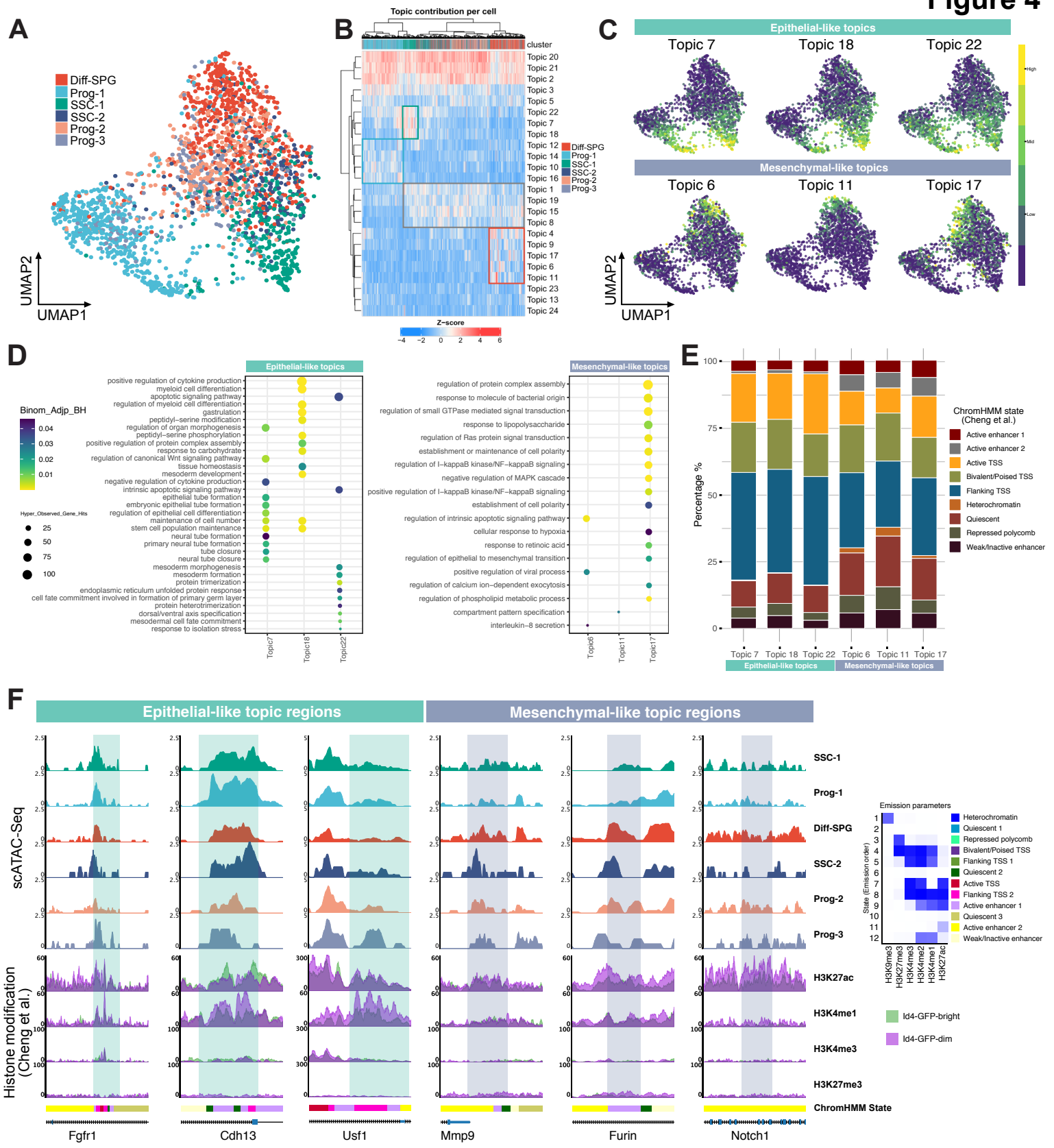

A
Diff-SPG
Prog-1
SSC-1
SSC-2
Prog-2
Prog-3
UMAP2
UMAP1
B
Topic contribution per cell
cluster
Topic 20
Topic 21
Topic 2
Topic 3
Topic 5
Topic 22
Topic 7
Topic 18
Topic 12
Topic 14
Topic 10
Topic 16
Topic 1
Topic 19
Topic 15
Topic 8
Topic 4
Topic 9
Topic 17
Topic 6
Topic 11
Topic 23
Topic 13
Topic 24
Z-score
C
Epithelial-like topics
Topic 7
Topic 18
Topic 22
Mesenchymal-like topics
Topic 6
Topic 11
Topic 17
D
Epithelial-like topics
Mesenchymal-like topics
Binom_Adjp_BH
Hyper_Observed_Gene_Hits
positive regulation of cytokine production
myeloid cell differentiation
apoptotic signaling pathway
regulation of myeloid cell differentiation
gastrulation
peptidyl−serine modification
regulation of organ morphogenesis
peptidyl−serine phosphorylation
positive regulation of protein complex assembly
response to carbohydrate
regulation of canonical Wnt signaling pathway
tissue homeostasis
mesoderm development
negative regulation of cytokine production
intrinsic apoptotic signaling pathway
epithelial tube formation
embryonic epithelial tube formation
regulation of epithelial cell differentiation
maintenance of cell number
stem cell population maintenance
neural tube formation
primary neural tube formation
tube closure
neural tube closure
mesoderm morphogenesis
mesoderm formation
protein trimerization
endoplasmic reticulum unfolded protein response
cell fate commitment involved in formation of primary germ layer
protein heterotrimerization
dorsal/ventral axis specification
mesodermal cell fate commitment
response to isolation stress
regulation of protein complex assembly
response to molecule of bacterial origin
regulation of small GTPase mediated signal transduction
response to lipopolysaccharide
regulation of Ras protein signal transduction
establishment or maintenance of cell polarity
regulation of I−kappaB kinase/NF−kappaB signaling
negative regulation of MAPK cascade
positive regulation of I−kappaB kinase/NF−kappaB signaling
establishment of cell polarity
regulation of intrinsic apoptotic signaling pathway
cellular response to hypoxia
response to retinoic acid
regulation of epithelial to mesenchymal transition
positive regulation of viral process
regulation of calcium ion−dependent exocytosis
regulation of phospholipid metabolic process
compartment pattern specification
interleukin−8 secretion
E
Percentage %
ChromHMM state (Cheng et al.)
Active enhancer 1
Active enhancer 2
Active TSS
Bivalent/Poised TSS
Flanking TSS
Heterochromatin
Quiescent
Repressed polycomb
Weak/Inactive enhancer
Topic 7
Topic 18
Topic 22
Topic 6
Topic 11
Topic 17
Epithelial-like topics
Mesenchymal-like topics
F
Epithelial-like topic regions
Mesenchymal-like topic regions
scATAC-Seq
Histone modification (Cheng et al.)
SSC-1
Prog-1
Diff-SPG
SSC-2
Prog-2
Prog-3
H3K27ac
H3K4me1
H3K4me3
H3K27me3
ChromHMM State
Emission parameters
State (Emission order)
H3K9me3
H3K27me3
H3K4me3
H3K4me2
H3K4me1
H3K27ac
Heterochromatin
Quiescent 1
Repressed polycomb
Bivalent/Poised TSS
Flanking TSS 1
Quiescent 2
Active TSS
Flanking TSS 2
Active enhancer 1
Quiescent 3
Active enhancer 2
Weak/Inactive enhancer
Id4-GFP-bright
Id4-GFP-dim
Fgfr1
Cdh13
Usf1
Mmp9
Furin
Notch1

Figure 5

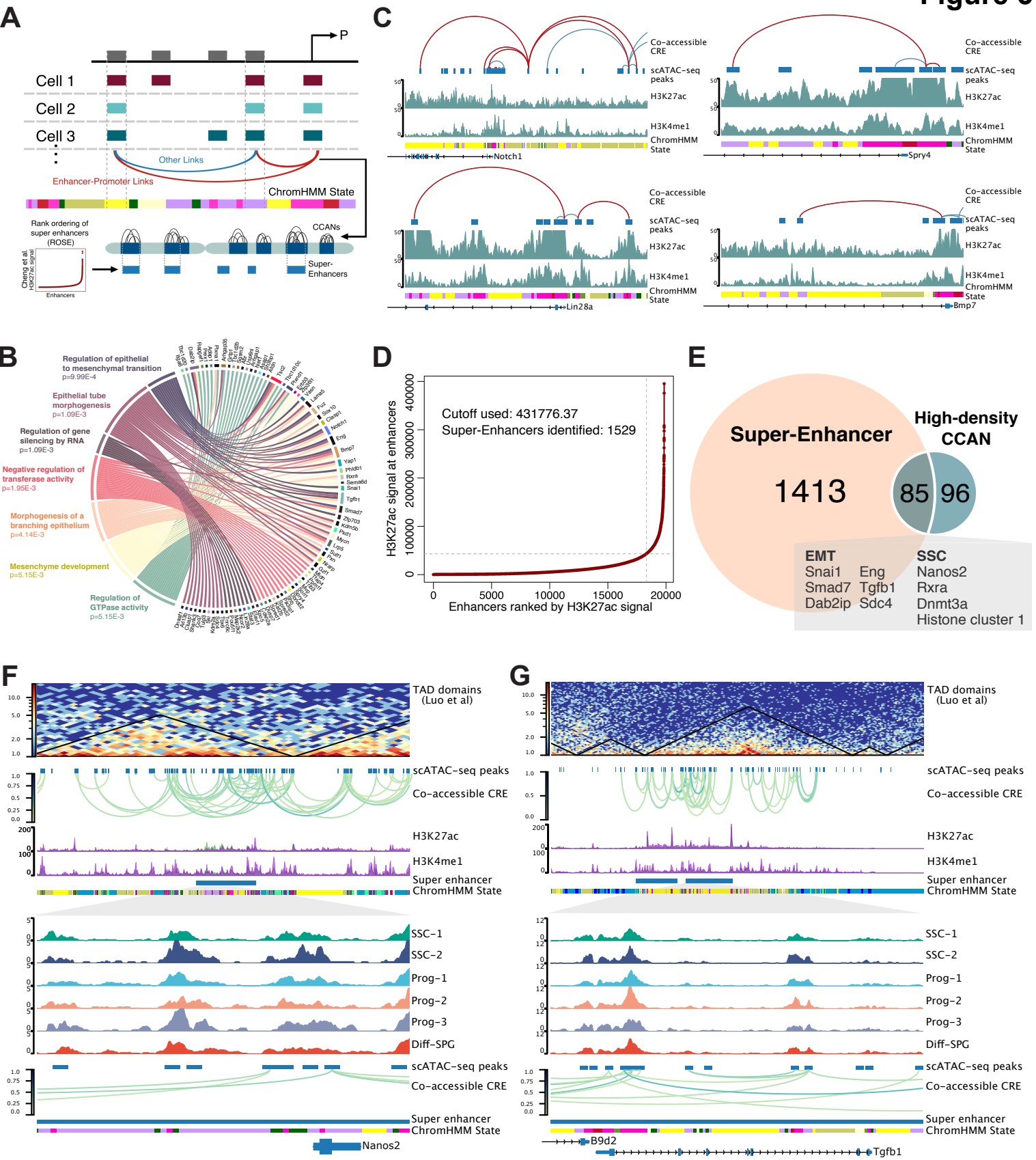

A
P
Cell 1
Cell 2
Cell 3
Other Links
Enhancer-Promoter Links
ChromHMM State
Rank ordering of super enhancers (ROSE)
Cheng et al. H3K27ac signal
Enhancers
CCANs
Super-Enhancers
B
Regulation of epithelial to mesenchymal transition p=9.99E-4
Epithelial tube morphogenesis p=1.09E-3
Regulation of gene silencing by RNA p=1.09E-3
Negative regulation of transferase activity p=1.95E-3
Morphogenesis of a branching epithelium p=4.14E-3
Mesenchyme development p=5.15E-3
Regulation of GTPase activity p=5.15E-3
C
Co-accessible CRE
scATAC-seq peaks
H3K27ac
H3K4me1
ChromHMM State
Notch1
Spry4
Lin28a
Bmp7
D
Cutoff used: 431776.37
Super-Enhancers identified: 1529
H3K27ac signal at enhancers
Enhancers ranked by H3K27ac signal
E
Super-Enhancer
1413
High-density CCAN
85
96
EMT
Snai1
Eng
Smad7
Tgfb1
Dab2ip
Sdc4
SSC
Nanos2
Rxra
Dnmt3a
Histone cluster 1
F
TAD domains (Luo et al)
scATAC-seq peaks
Co-accessible CRE
H3K27ac
H3K4me1
Super enhancer
ChromHMM State
SSC-1
SSC-2
Prog-1
Prog-2
Prog-3
Diff-SPG
Nanos2
G
B9d2
Tgfb1

Figure 6

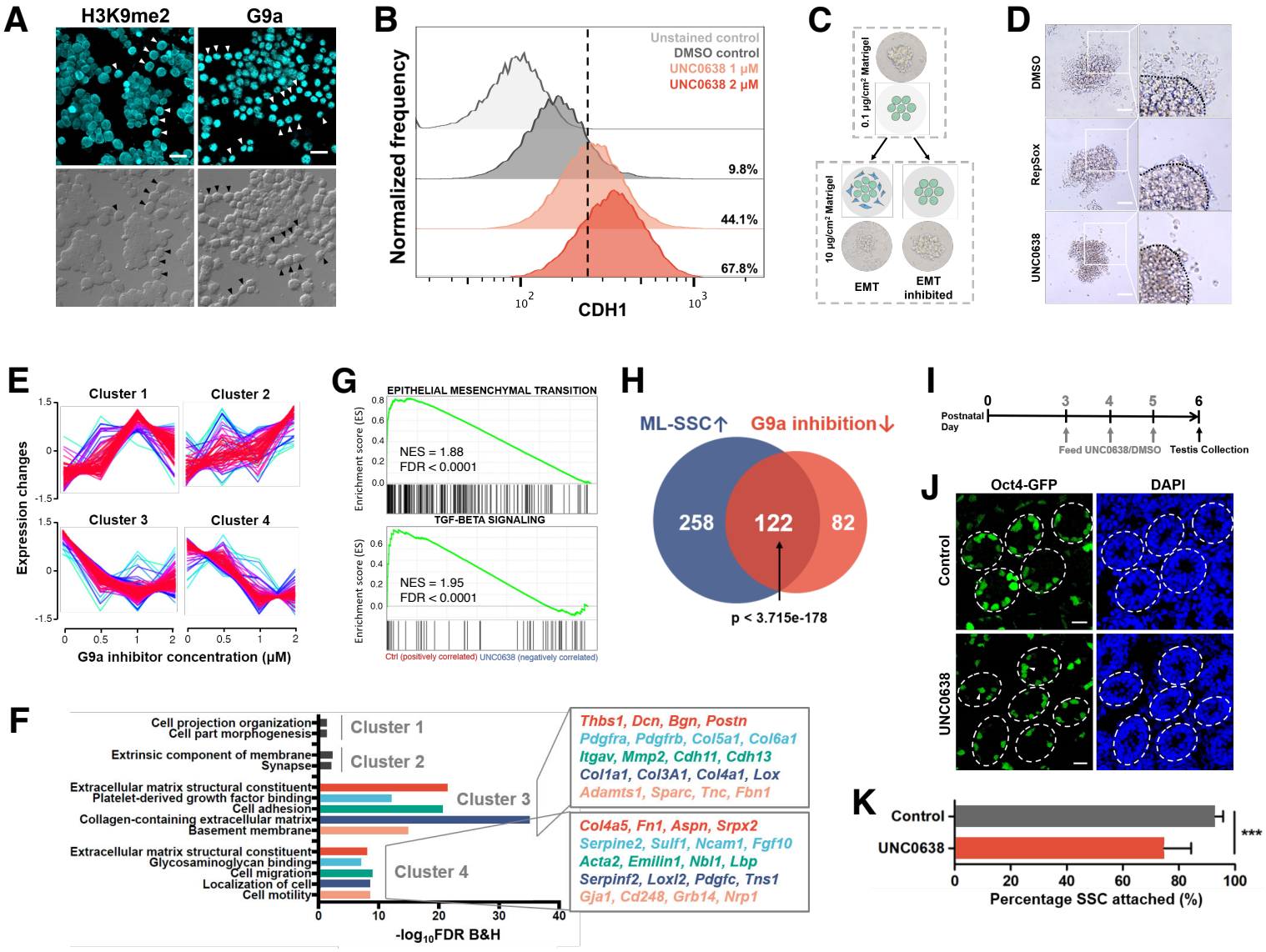
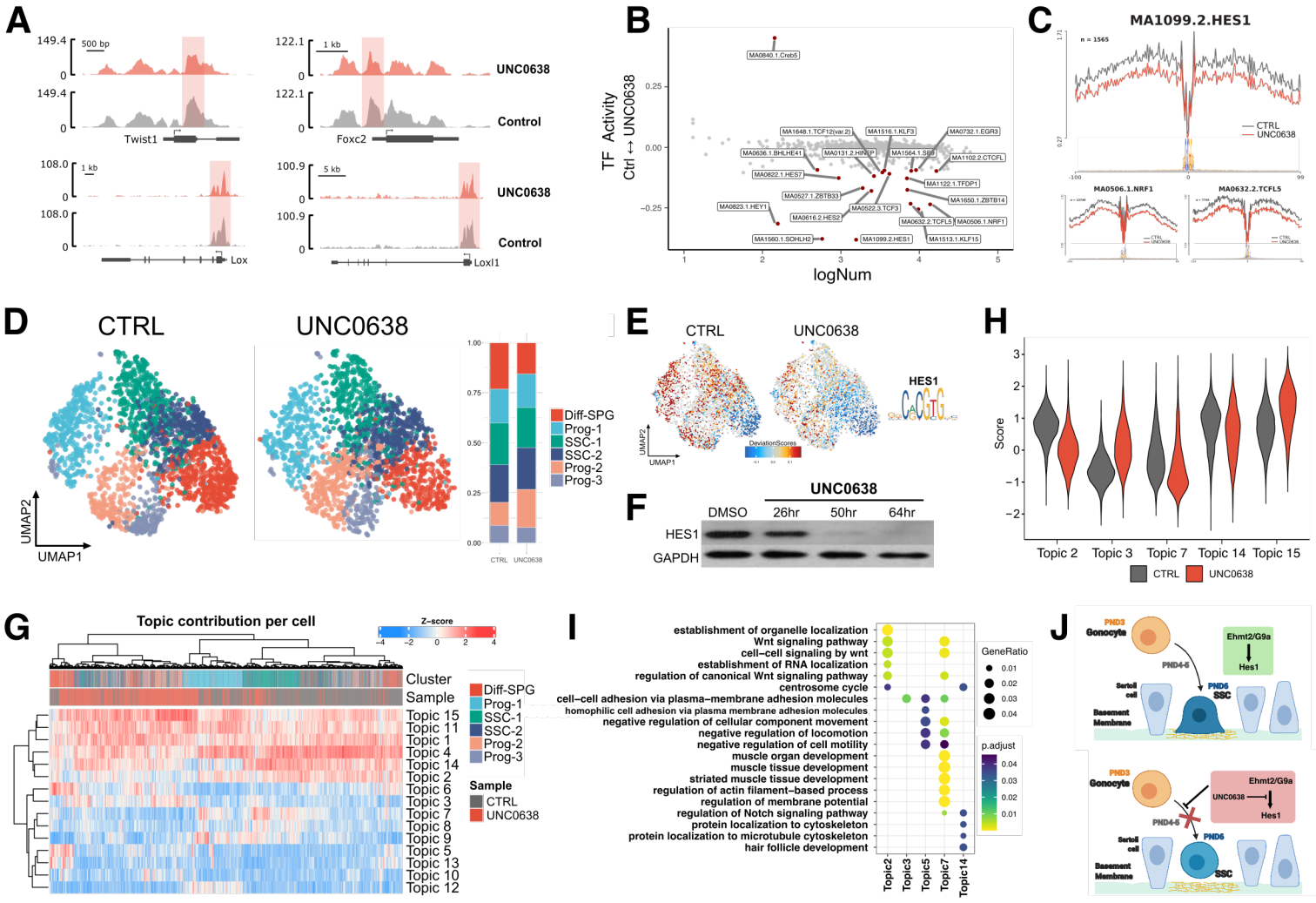

A
H3K9me2
G9a
B
Normalized frequency
Unstained control
DMSO control
UNC0638 1 μM
UNC0638 2 μM
9.8%
44.1%
67.8%
CDH1
C
0.1 μg/cm² Matrigel
10 μg/cm² Matrigel
EMT
EMT inhibited
D
DMSO
RepSox
UNC0638
E
Cluster 1
Cluster 2
Cluster 3
Cluster 4
Expression changes
G9a inhibitor concentration (μM)
G
EPITHELIAL MESENCHYMAL TRANSITION
NES = 1.88
FDR < 0.0001
TGF-BETA SIGNALING
NES = 1.95
FDR < 0.0001
Enrichment score (ES)
Ctrl (positively correlated) UNC0638 (negatively correlated)
H
ML-SSC↑
G9a inhibition↓
258
122
82
p < 3.715e-178
I
Postnatal Day
0
3
4
5
6
Feed UNC0638/DMSO
Testis Collection
J
Oct4-GFP
DAPI
Control
UNC0638
F
Cell projection organization
Cell part morphogenesis
Extrinsic component of membrane
Synapse
Extracellular matrix structural constituent
Platelet-derived growth factor binding
Cell adhesion
Collagen-containing extracellular matrix
Basement membrane
Extracellular matrix structural constituent
Glycosaminoglycan binding
Cell migration
Localization of cell
Cell motility
Cluster 1
Cluster 2
Cluster 3
Cluster 4
-log10FDR B&H
Thbs1, Dcn, Bgn, Postn
Pdgfra, Pdgfrb, Col5a1, Col6a1
Itgav, Mmp2, Cdh11, Cdh13
Col1a1, Col3A1, Col4a1, Lox
Adamts1, Sparc, Tnc, Fbn1
Col4a5, Fn1, Aspn, Srpx2
Serpine2, Sulf1, Ncam1, Fgf10
Acta2, Emilin1, Nbl1, Lbp
Serpinf2, Loxl2, Pdgfc, Tns1
Gja1, Cd248, Grb14, Nrp1
K
Control
UNC0638
***
Percentage SSC attached (%)

Figure 7


A
149.4
500 bp
0
149.4
0
Twist1
122.1
1 kb
0
122.1
0
Foxc2
UNC0638
Control
108.0
1 kb
0
108.0
0
Lox
100.9
5 kb
0
100.9
0
Loxl1
UNC0638
Control
B
TF Activity
Ctrl ↔ UNC0638
logNum
MA0840.1.Creb5
MA1648.1.TCF12(var.2)
MA1516.1.KLF3
MA0732.1.EGR3
MA0636.1.BHLHE41
MA0131.2.HINFP
MA1564.1.SP8
MA1102.2.CTCFL
MA0822.1.HES7
MA1122.1.TFDP1
MA0527.1.ZBTB33
MA1650.1.ZBTB14
MA0823.1.HEY1
MA0522.3.TCF3
MA0616.2.HES2
MA0632.2.TCFL5
MA0506.1.NRF1
MA1560.1.SOHLH2
MA1099.2.HES1
MA1513.1.KLF15
C
MA1099.2.HES1
n = 1565
CTRL
UNC0638
MA0506.1.NRF1
MA0632.2.TCFL5
D
CTRL
UNC0638
UMAP2
UMAP1
Diff-SPG
Prog-1
SSC-1
SSC-2
Prog-2
Prog-3
CTRL
UNC0638
E
CTRL
UNC0638
HES1
DeviationScores
UMAP2
UMAP1
F
UNC0638
DMSO
26hr
50hr
64hr
HES1
GAPDH
H
Score
Topic 2
Topic 3
Topic 7
Topic 14
Topic 15
CTRL
UNC0638
G
Topic contribution per cell
Z-score
Cluster
Sample
Topic 15
Topic 11
Topic 1
Topic 4
Topic 14
Topic 2
Topic 6
Topic 3
Topic 7
Topic 8
Topic 9
Topic 5
Topic 13
Topic 10
Topic 12
Diff-SPG
Prog-1
SSC-1
SSC-2
Prog-2
Prog-3
Sample
CTRL
UNC0638
I
establishment of organelle localization
Wnt signaling pathway
cell−cell signaling by wnt
establishment of RNA localization
regulation of canonical Wnt signaling pathway
centrosome cycle
cell−cell adhesion via plasma−membrane adhesion molecules
homophilic cell adhesion via plasma membrane adhesion molecules
negative regulation of cellular component movement
negative regulation of locomotion
negative regulation of cell motility
muscle organ development
muscle tissue development
striated muscle tissue development
regulation of actin filament−based process
regulation of membrane potential
regulation of Notch signaling pathway
protein localization to cytoskeleton
protein localization to microtubule cytoskeleton
hair follicle development
Topic2
Topic3
Topic5
Topic7
Topic14
GeneRatio
p.adjust
J
PND3
Gonocyte
Ehmt2/G9a
Hes1
PND4-6
Sertoli cell
PND6
SSC
Basement Membrane
UNC0638