The landscape of mouse epididymal cells defined by the single-cell RNA-Seq

Jianwu Shi[1,], Mengmeng Sang[1], Gangcai Xie[1*], Hao Chen[1 *]

[1] Institute of Reproductive Medicine, Medical School of Nantong University, Nantong, PR China

*To whom correspondence should be addressed:

Institute of Reproductive Medicine, Medical School, Nantong University Qixiu Road 19, Nantong, China, 226001

Email: chenhao@ntu.edu.cn to Hao Chen or gangcai@ntu.edu.cn to Gangcai Xie

ABSTRACT

Spermatozoa acquire their fertilizing ability and forward motility properties during epididymal transit. Although lots of attempts elucidating the functions of different cell types in epididymis, the composition of epididymal tubal and cell types are still largely unknown. Using single-cell RNA sequence, we analyzed the cell constitutions and their gene expression profiles of adult epididymis derived from caput, corpus and cauda epididymis with a total of 12,597 cells. This allowed us to elucidate the full range of gene expression changes during epididymis and derive region-specific gene expression signatures along the epididymis. A total of 7 cell populations were identified with all known constituent cells of mouse epididymis, as well as two novel cell types. Our analyses revealed a segment to segment variation of the same cell type in the three different part of epididymis and generated a reference dataset of epididymal cell gene expression. Focused analyses uncovered nine subtypes of principal cell. Two subtypes of principal cell, c0.3 and c.6 respectively, in our results supported with previous finding that they mainly located in the caput of mouse epididymis and play important roles during sperm maturation. We also showed unique gene expression signatures of each cell population and key pathways that may concert epididymal epithelial cell-sperm interactions. Overall, our single-cell RNA seq datasets of epididymis provide a comprehensive potential cell types and information-rich resource for the studies of epididymal composition, epididymal microenvironment regulation by the specific cell type, or contraceptive development, as well as a gene expression roadmap to be emulated in efforts to achieve sperm maturation regulation in the epididymis.

Key words: Epididymis, scRNA-seq, Principal cell sub-populations

INTRODUCTION

The epididymis is a critical male sex organ that plays key roles in sperm transport, maturation and storage (Cooper and Yeung, 2010; Cornwall, 2009). Spermatozoa from testes acquires their motility and fertilization ability when they transit through the epididymis. In the epididymis, the sperm plasma membrane is subjected to sequential biochemical and proteomic modifications as a result of subtle interactions with components of the extracellular environment in the epididymal lumen.

The epididymis epithelium supports a luminal environment that promotes sperm maturation, and each region of the duct (caput, corpus, and cauda) is believed to play a unique role during sperm transition (Breton and Brown, 2013; Breton et al., 2016). The luminal secretions from caput and corpus epididymis are beneficial for the acquisition of sperm motility and fertilizing ability (Turner, 1995), cauda epididymal secretions are important for the proper storage of spermatozoa for several days in conditions that preserve their fertility (Jones and Murdoch, 1996). Each of these regions have been demonstrated to possess a distinct pattern of gene expression related to physiological functions which are important in the different steps of sperm maturation (Dube et al., 2007; Guyonnet et al., 2009; Johnston et al., 2005; Thimon et al., 2007). This compartmentalized gene expression triggers segment-specific protein secretions into the luminal fluid and, in turn, generates microenvironments optimized for each step of sperm maturation.

The composition of the intra luminal milieu is controlled by the surrounding pseudostratified epithelium, which is composed of five cell types, for now, possessing distinct physiological functions including principal cells, basal cells, clear/narrow cells, apical cells and halo cells all along the organ (Bernard Robaire, 2006; Cornwall, 2009; Sullivan et al., 2019). Studies of region-specific epididymal proteins showed that certain cell type was able to express quite different classification of genes, which contributes to the different physiologic functions of the segments (Cornwall, 2009; Dacheux and Dacheux, 2014). With the development of single-cell RNA sequencing (scRNA-seq), numbers of organs were analyzed in mammalian (Dong et al., 2018; Han et al., 2018; Tabula Muris et al., 2018), including male and female reproductive organ such as testis (Green et al., 2018; Guo et al., 2018; Wang et al., 2018) and ovary (Fan et al., 2019; Zhang et al., 2018) but not epididymis. However, the exact cell composition and gene expression repertoire of each cell population in epididymis are less characterized.

In this study, we applied microfluidic based scRNA-seq to the analyses of 12,597 cells derived from caput, corpus and cauda of mouse epididymis. Cell clustering analysis revealed five known cell types and more importantly discovered two novel epididymis cell sub-populations. Remarkably, nine sub-populations of principle cells were found in our study, which revealed unexpectedly heterogeneous in epididymis principle cells. Finally, our study represents the first regional transcriptome profiling of mouse epididymis at single cell level, which is important to the future study of the

spatial microenvironments of the epididymis, and is also important for our understanding of epididymal diseases.

MATERIALS AND METHODS

Animals and epididymis sample collection

Sample collection was carried out under license in accordance with the Guidelines for Care and Use of Laboratory Animals of China and all protocols were approved by the Institutional Review Board of Nantong University. Five eight-week-old wildtype C57Bl/6J mice were used in this study. After sacrificing the mice, the epididymis were dissected and divided into three regions (caput, corpus, and cauda) and immediately stored in the GEXSCOPE™ Tissue Preservation Solution (Singleron Biotechnologies) at 2-8 ºC.

Tissue dissociation and single cell suspension preparation

Prior to tissue dissociation, the specimens were washed with Hanks Balanced Salt Solution (HBSS) for three times and minced into 1–2 mm pieces. The tissue pieces were digested in 2ml GEXSCOPE™ Tissue Dissociation Solution (Singleron Biotechnologies) at 37℃ for 15min in a 15ml centrifuge tube with continuous agitation. Following digestion, a 40-micron sterile strainer (Corning) was used to separate cells from cell debris and other impurities. The cells were centrifuged at 1000 rpm, 4ºC, for 5 minutes and cell pellets were resuspended in 1ml PBS (HyClone). To remove red blood cells, 2 mL GEXSCOPE™ Red Blood Cell Lysis Buffer (Singleron) was added to the cell suspension and incubated at 25°C for 10 minutes. The mixture was then centrifuged at 1000 rpm for 5 min and cell pellet

re-suspended in PBS. Cells were counted with TC20 automated cell counter (Bio-Rad).

Single cell RNA sequencing library preparation

The concentration of single-cell suspension was adjusted to 1×10^5 cells/mL in PBS. Single cell suspension was then loaded onto a microfluidic chip (part of Singleron GEXSCOPE™ Single Cell RNAseq Kit, Singleron Biotechnologies) and single cell RNA-seq libraries were constructed according to manufacturer's instructions (Singleron Biotechnologies). The resulting single cell RNAseq libraries were sequenced on Illumina HiSeq X10 instrument with 150bp paired end reads.

Single cell RNA-seq data analyses

In general, following steps were included in the analyses: raw sequencing data preprocessing, cell barcodes extraction, reads genomic alignment, unique molecular identifier (UMI) counting, cell sub-population discovery and sub-population gene markers finding. Firstly, reads with low sequencing qualities were filtered and the sequencing adapters were trimmed by Fastp (Chen et al., 2018) software (fastp 0.19.5) using default settings. Then, umi_tools (Smith et al., 2017) was used to identify and extract cell barcodes with the settings of cell number of 5000 and error correction threshold of 1. Next, STAR genomic mapper was applied to map the extracted reads to the mouse Gencode genome (GRCm38.primary_assembly.genome.fa, version M18). And then, featureCounts was used to assign exon-level reads based on Gencode gene annotation (gencode.vM18.primary_assembly.annotation.gtf, version M18). Furthermore, the UMIs for each gene were counted by umi_tools with the editing

distance threshold of 1. Finally, cell sub-population discovery and sub-population markers finding were analyzed by Seurat 3.0 (Stuart et al., 2019), and only the cells with the number of feature RNAs between 500 and 6000, and the percentage of mitochondrial less than 0.5 were included for further analyses. The normalization of the expression data was performed by sctransform (SCT) algorithms (https://github.com/ChristophH/sctransform), and 3000 genes were selected for the integration of sub-populations from epididymis caput, corpus and cauda regions.

Gene enrichment and transcription factor analyses

For each sub-population of epididymis or epididymal principle cells, the sub-population (or cell cluster) specific marker genes were selected for gene enrichment analyses, and the analyses were performed by clusterProfiler (Yu et al., 2012), which including gene ontology analysis. The file including 1675 mouse transcripton factors was downloaded from Riken Transcription Factor Database (http://genome.gsc.riken.jp/TFdb/). For each sub-population of epididymal principle cell, the average expression score was calculated based on SCT normalized expression values. The Maximal Regional Expression Difference (MRED) was defined as the difference between the maximal and the minimal expression scores among the three epididymis regions for all principle sub-populations. The 95% quantile value of the MREDs was used as the cutoff for the selection of the regional differentially expressed transcription factors.

RESULTS

ScRNA-seq analyses revealed a novel sperm interaction epididymal sub-population

Although quite a few studies elucidated the segment-specific function of epididymis (Belleannee et al., 2012; Sipila and Bjorkgren, 2016; Thimon et al., 2007), the comprehensive cell population of region-specific epididymal cells is still largely unknown. To define the cell types in mouse epididymis, we chose scRNA-seq to examine cells from fresh isolated caput, corpus and cauda region of the epididymis. A total of 12,597 cells from three regions of epididymis (Figure 1A and 1B) were conducted for scRNA-seq and subsequent analyses. Using Seurat analysis, we identified seven cell clusters with five known cell types in the mouse epididymis, including principal cell, clear/narrow cell, basal cell, sperm and T cells (Figure 1A and 1C, and more known marker gene expression in Figure S1).

Regarding the cell clusters (C1 and C5) without known epididymis marker genes expressed, we did GO enrichment or DAVID gene enrichment analyses. Based on GO enrichment analysis, C5 was identified as vascular cells, the marker genes of which are enriched with “vascular development”, “regulation of angiogenesis” and so on (Figure S2). Interestingly, we identified a novel epididymis cell types (C1), the marker genes of which were enriched with DAVID terms “secreted”, “signal peptide”, “protein digestion and absorption” et al (Figure 2B). Furthermore, Epididymis Secretory Sperm Binding Protein (Col6a1) gene was highly expressed in this novel epididymis sub-population (C1), which indicates its possible interaction with sperm

cells in epididymis (gene expression pattern illustrated in Figure 2A). And the highly specific expression of Olfactomedin Like 3 (Olfml3) in C1 novel cell type also revealed its possible intercellular functionalities (all top 10 C1 highly specific expressed genes can be found in Figure 2A). Interestingly, we found that the cell proportion of this novel epididymal sub-population (C1) was highest in epididymal cauda region, while lowest in caput region (Figure 2C).

For all the epididymis sub-populations (C0-C6), we identified population-level specifically expressed marker genes (Top 10 marker genes showed in Figure 1D, and all marker genes can be found in Table S1), which provides important marker genes for further study of epididymis sub-populations.

Single cell RNA-Seq analyses recapitulate principal cell sub-populations

Our clustering analysis of the total epididymal cells suggested that the principal cell population is itself comprised of several sub-populations (Figure1A). To further distinguish its sub-populations, we performed clustering on principal cells (Figure 3A). Nine different sub-populations of principal cell were revealed, which were defined as c0.1-c0.8 (Figure 3A). The pan-principal cell marker Aqp9 (Carvajal et al., 2018; Pastor-Soler et al., 2001) was found in all 9 clusters (Figure 3C), as expected. The enriched genes in each of these 9 principal cell were listed in the supplementary Table S2.

Principal cell is the major epididymal cell which composes the entire tubules and Aqp9 is the common recognized principal cell marker gene(Pastor-Soler et al., 2010). However, a comprehensive census of principal cell in the specific epididymal region

has been hampered by their marker gene rarity. To address these issues, we applied single cell genetic strategies and screened out top ten marker genes for each subsets of principal cell (Figure 4A). Representative marker genes for each principal cell subsets were visualized in UMAP space (Figure 4B). Consistently, the cell population c0.3 and c0.6 have been reported in the previous studies (Ma et al., 2013; Xie et al., 2013), which were identified through the marker genes Lcn5 and Spink13 respectively (Figure 4A&B). Furthermore, the cell subset c0.3 was the most abundant cells at the caput region of epididymis with the high expressions of Defb 15, 29 and 30 (figure 5A&B). GO analysis also revealed its functional role of defense response, suggesting its predominant role in defensin family regulation (Figure 5C).

Apart from reported cell subsets, the seven remaining clusters represent novel populations, corresponding to extracellular structure organization(c0.0), small GTPase mediated signal transduction (c0.4), regulation of protein serine/threonine kinase activity(c0.5), epithelial cell proliferation (c0.7) and nuclear division or organelle fission(c0.8) process analyzed by GO enrichment. The represented GO terms associated with the differentially expressed genes in each cell cluster (Figure 4C) manifested diverse roles of the principal cell in regulating epididymal microenvironments, albeit three cell subsets remained unknown GO processes.

Epididymal cell population proximal-distal distributions

In the total epididymal cell study, we found spatial distribution patterns of the seven epididymal cell populations. Four cell populations (C0: Principle Cell; C3: Clear/Narrow Cells; C4: Basal Cells; C5: Sperm cells) were distally reduced, while two

populations (C1: Novel epididymal cell identified in this study, C2: Halo/T Cells) were distally increased (Figure 2C). Intriguingly, expressions of lots genes showed segment-specific or distally changed pattern (Bjorkgren and Sipila, 2019; Sipila and Bjorkgren, 2016), suggesting their expression patterns might be altered along with the cell number changes. The spatial distribution changes of each epididymal cell population might help to reveal their functionalities.

In the principle sub-population study, unexpectedly, we discovered that expression of mitochondrial genes was significantly higher in the epididymal corpus and cauda regions comparing to the caput region (Figure 3B). This observation suggests that the energy requirements in the distal epididymal regions might be stronger. For each principle sub-population, we also did cell population regional distribution studies. In total, three epididymal principle sub-populations (c0.0, c0.1, c0.4) were observed as distally increased, and six sub-populations (c0.2, c0.3, c0.5, c0.6, c0.7, c0.8) were distally decreased (Figure 3D).

Based on the total cell population and principle sub-population spatial distribution studies, we suggested that the different microenvironments of the three epididymal regions might be due to the cell population differences. And the epididymal proximal-distal microenvironments differences might also suggest region-dependent functionalities in mouse epididymis for sperm maturation.

Regional regulation patterns identified in the principle sub-populations

Finally, we systematically studied the transcriptional regulations in mouse epididymal principle-cell sub-populations (Figure 6). Based on the global distribution

of the MREDs for the expressed genes (Figure S3), we found 459 transcription factors (TFs) were spatially differentially expressed, and the expression of 177 TFs of them was distally decreased or increased (Table S3). We found genes, such as Enpp2 and Bmyc, were highly regional differentially expressed (Top 10 reginal differentially expressed TFs shown in Figure 6A). Enpp2 had been annotated as both of phosphodiesterase and phospholipase, and might be important for the catalyzing of the production of lysophosphatidic acid in extracellular fluids (ENPP2 in GENECARDS). Based on TF-TF expression correlation and network study (Figure 6B, F), Nupr1 gene was found to be highly correlated with other TFs, which indicates its important role in the principle cell regulation network. Interestingly, we found three patterns of regional expression of TFs, some TFs were both regionally and sub-population specifically expressed, such as Nupr1 and Pdlim1 (Figure 6C), some TFs such as Lhx1, Stat5b and Meis1 were highly specifically expressed at certain epididymal region and principle sub-population (Figure 6D), and some TFs such as Runx2, Scand1 and Mafb (Figure 6E) were highly reginal expressed no matter which principle sub-population they were in. This information highlighted the complex transcriptional regulation patterns in epididymal principle sub-populations, and the regulation of which dependents on both of the epididymal location and sub-population types.

Discussion

The mammalian epididymis is composed of convoluted and interconnected regions, each of which contains a complex mosaic of spatially intermixed cells. Overwhelming studies have attempted to characterize the cell types in the epididymis

based on cell location, morphology, connection, function, and marker gene expression (Da Silva et al., 2011; Shum et al., 2011; Shum et al., 2014). However, there is still a lack of knowledge regarding the epididymal cell compositions and its gene regulations from caput to cauda of epididymis. So we, for the first time, profiled the single cell RNA sequencing of the whole mouse epididymis, which included 12,597 epididymal single cells from caput, corpus and cauda regions. In total, we identified six epididymal cell populations according to known marker genes or GO enrichment analysis. And interestingly, we discovered one novel epididymal cell population that cannot be identified as known cell types. This novel cell population is enriched with Epididymis Secretory Sperm Binding Protein (Col6a1) and Olfactomedin Like 3 (Olfml3), which indicates its possible interaction with sperms and might play an important role in sperm maturation. Unexpectedly, the epididymal principal cell, as the major cell type in epididymis, can be further categorized into nine sub-populations, and all of which express classical principal marker gene Aqp9 (Pastor-Soler et al., 2010). Notably, those principal sub-populations exhibit proximal-distal distribution variations in mouse epididymis, which suggests their functional divergence for sperm maturation.

In the past decades, emerging studies endeavored to elucidate the regulation of region-specific gene expression in the epididymis and its potential contribution for sperm maturation using varies strategy including genetic regulation (Johnston et al., 2005; Oh et al., 2006; Zhang et al., 2006), small RNA(Anderson et al., 2015), epigenetic control(Sipila and Bjorkgren, 2016) and androgen-responsive regulations (Belleannee et al., 2012; Pihlajamaa et al., 2014). Region specific gene expression of

290 epididymis in the mouse and human was also studied by microarray analysis(Johnston
291 et al., 2005; Zhang et al., 2006) . On the other hand, it was plausible that the
292 epididymosome was contributed to the sperm modifications during the sperm transient
293 in epididymis (Sharma et al., 2018; Trigg et al., 2019). Recently, James et al.
294 demonstrated the expression profiles of segment-specific human epididymal epithelial
295 cells (Browne et al., 2016). The human epididymal epithelial cells were isolated and
296 differentiated cultured from caput, corpus and cauda and then underwent RNA-seq.
297 However, the specific cell constitution and contribution which locate in the different
298 region are less characterized. Our results depicted the cell landscape from each region
299 of the adult mouse epididymis and in-depth analysis of the gene expression pattern of
300 each cell population, suggesting more complexity roles of cell-specific regulation of
301 epididymis besides segment-specific modulation.

302 Based on the current knowledge, the epididymal epithelium is composed of five
303 distinct cell types: principal, narrow/clear, apical, Halo (T cell) and basal cells
304 (Bernard Robaire, 2006; Cornwall, 2009; Sullivan et al., 2019). Supporting previous
305 studies, we found all known cell types based on the reported cell-specific marker genes
306 (Figure 1). In detail, Cell population C0 was defined as principal cell; cell population
307 C2 was referred to as Halo cell; cell population C4 was corresponding to basal cell, as
308 well as cell population C6 was identified as sperm. Cell population C3 was identified as
309 Clear/Narrow cell due to lack specific cell markers. Intriguingly, the cell number of
310 these known cell populations was divergent along with the different region. For
311 example, cell population of C0, C3, C4 and C5 was gradually decreased, whereas cell

312 population of C1 and C2 was elevated from caput to cauda. Convincingly, our results of
313 cell population C3 dramatically reduced in corpus and cauda were in line with previous
314 study that cell/Narrow cell were exclusively in the caput region (Bernard Robaire, 2006;
315 Breton et al., 2016).

316 Besides specific expression of Col6a1 and Olfml3, the novel epididymal cell
317 population C1 is also significantly enriched with extracellular structure organization，
318 extracellular matrix organization and epithelial cell proliferation. The top 10
319 expressed genes in C1 are Clec3b, Olfml3, Col6a1, Gsn, Pcolce, Igfbp6,
320 Serping1, C7 and Col1a1. Consisting with GO analysis, all these genes were
321 predominantly localized in the extracellular compartment according to the
322 information from GeneCards (Stelzer G, 2016). In particularly, C-Type Lectin
323 Domain Family 3 Member B (clec3b) was reported to control cell proliferation
324 in clear cell renal cell carcinoma(Liu et al., 2018) as well as low clec3b in
325 exosomes derived from HCC promoted cell migration, invasion and
326 epithelial–mesenchymal transition(Dai et al., 2019), indicating its contribution
327 to the maintenance of epididymal epithelium. Recent study found that Gelsolin
328 (Gsn) was the major secreted protein in the distal region of the bovine
329 epididymis(Belleannee et al., 2011; Dacheux and Dacheux, 2014). In line with
330 previous finding, our results demonstrated that the number of C1 cells was
331 elevated from caput to cauda epididymis, suggesting its potential involvement
332 for the sperm storage. According to the GO enrichment analysis, the population of
333 C5 cells was correspond to the regulation of vasculature development and

334 angiogenesis, which might be responsible for the communication between epididymal
335 cells and vasculature cells the epididymis.

336 Specifically, principal cell is the most abundant cell type in the epididymal tubule,
337 which accounts for ~80% of the epithelium(Cornwall, 2009). However, little is known
338 about the differences of region-specific principal cell as well as the transcriptome
339 profiling in the three regions of epididymis. Our data demonstrated for the first time
340 the multiple cell subsets of principal cells with distinct marker genes in each
341 sub-population. Clustering analysis focusing on the principal cells derived from caput,
342 corpus and cauda epididymis revealed nine major cell types, two of which had been
343 reported to be located at the caput region, and such localization is also confirmed in
344 our study. The marker gene Lcn5 was highly and specifically expressed in the
345 sub-population of principal cells c0.3, which mainly located at the caput region of
346 epididymis (Xie et al., 2013). And another marker gene spink13 in c0.6 cell
347 population was expressed in the principal cell of initial segment of caput and was
348 involved in the regulation of sperm acrosome reaction (Ma et al., 2013).

349 The traditional perspective of epididymis being a well differentiated organ is not
350 thought to have stem cell, nor its epididymal epithelial cells to divide in adult
351 (Bernard Robaire, 2006). The epididymal epithelial cells consists of four cell types
352 including principal cells, basal cell, clear cells and narrow cells. Recently, several
353 studies have been investigated the possible proliferation of epididymal epithelial cells
354 both in vitro and in mammalian animals. The isolated bovine epididymal cells were
355 stimulated to be proliferation when co-cultured with spermatozoa(Reyes-Moreno et

356 al., 2008). Bernal-Manas et al. reported the cell proliferation of epididymal epithelium
357 (mainly clear and principal cells) of *sus domesticus* epididymis with PCNA
358 immunochemistry staining (Bernal-Manas et al., 2014). Another study demonstrated
359 that the basal cells gained the capacity of proliferation after efferent duct ligation
360 (EDL) in mice (Kim et al., 2015). The most recent studies found that the epididymal
361 epithelia cells was able to rapidly grow after efferent duct ligation or orchidectomy.
362 Basal cells was considered as a competitive candidate for the epididymal stem cell,
363 albeit more convincing results need to be provided (Pinel et al., 2019). These results
364 suggested the proliferative possibility of epididymal epithelial cells during epididymal
365 cells dynamics. Of note, our data showed that cell subset c0.7 of principal cell was
366 predicted as progenitor/stem cells with significantly enrichment in the epithelial cell
367 proliferation and cell growth according to gene ontology analysis, suggesting that it
368 might serve as stem cell or progenitor cells to keep the intact of epididymis. The
369 proliferative cell population c0.7 from our data is comparable to previous results,
370 although the cell source was fundamentally different. Future studies will aim to
371 investigate whether the population c0.7 play a critical role in epididymal epithelium
372 homeostasis and its potential stem cell capability.

373 Previous study reported the apical mitochondria-rich cells in both mouse and
374 human epididymis These cells were believed to be distinct from principal cells and
375 involved in the holocrine secretion and acidification of epididymal fluid
376 (Martinez-Garcia et al., 1995). In contrast with previous studies, we surprisingly
377 found the mitochondria percentage of principal cell was significantly increased in

corpus and cauda region compared to caput region of epididymis. Furthermore, one of sub-population of principal cell (c0.4) is highly enriched with the expression of mitochondria genes such as mt-Nd1, mt-Co1, mt-Nd5, mt-Nd2, mt-Cytb, mt-Nd6, mt-Nd4, mt-Rnr2, Nktr and mt-Rnr1, most of which are believed to belong to the core subunits of mitochondrial membrane respiratory chain NADH dehydrogenase (Complex I) and Cytochrome C Oxidase complex, that is essential for the maintenance of mitochondrial functions and oxidative phosphorylation. Our data suggested the apical mitochondria-rich principle cell might play more important role in the corpus/cauda than caput region of epididymis. The detailed characterizations for the c0.4 cell sub-population still need to be further elucidated.

Epididymal cells secret plenty of the proteins, small RNA and extracellular vesicles into the epididymal lumen to regulate sperm maturation (Bjorkgren and Sipila, 2019). DNA binding transcription factors (TFs) play center roles in the regulation of gene expression. Therefore, we further explored the differentially expressed TFs and their expression profiles. Among TFs that were abundant in the majority of principal sub-population cells, some of them are related to epididymis function in our data, such as bmyc and enpp2. The bmyc, member of the Myc family to inhibit myc, was found to be predominately expressed in the caput region (Cornwall et al., 2001) and regulated by androgen and testicular factors. Bmyc gene knockout (KO) experiment showed that the KO mice had both of smaller testes and epididymis(Turunen et al., 2012). Enpp2 was detected in the epithelial cells in rat epididymis (Belleannee et al., 2010). Besides, recent study demonstrated the expression patterns of primary cultured human

epididymal epithelial cells derived from caput, corpus and cauda regions (Browne et al., 2016). They reported a total of 90 deferentially expressed TFs associated with epididymis epithelial cells. Compatible with previous study, expression profile of several TFs was identical such as creb3l1, tbx3, atf3 etc. However, more TFs (eg. Hlf, esr1, pou2f2, fosb and cdkn2c) displayed the inconsistent with previous results. Reasons for the inconsistency might be on account of the different of species between human and mouse. On the other hand, cultured cells may not fully epitomize the cells in the original organ because of its loss of cell communications and other hormonal regulations.

Taken together, our data provide an indispensable resource, at single cell resolution, for a comprehensive cell atlas of the epididymis. Combine with further studies on the detailed mechanisms of epididymal cells communications and microenvironments hemostasis for sperm maturation, it will shed light on the understanding of normal epididymis functionalities and the cellular mechanisms of epididymis diseases that cause male infertility.

Authors' contributions

HC conceived and designed the experiments. JWS, MMS and GCX performed the experiments and analyzed the data. JWS, GCX and HC wrote the paper.

Funding

This work was supported in part by grants from the National Key Research and Development Program of China (No. 2018YFC1003602 to HC); National Natural Science Foundation of China (81671432 and 81871202 to HC, 31900484 to GCX);

Natural Science Foundation of Jiangsu Province (BK20190924 to GCX) and the startup R&D funding of Nantong University (03083011 to HC).

Conflict of Interest

The authors declare that they have no conflict of interest.

Reference:

Anderson, A.L., Stanger, S.J., Mihalas, B.P., Tyagi, S., Holt, J.E., McLaughlin, E.A., and Nixon, B. (2015). Assessment of microRNA expression in mouse epididymal epithelial cells and spermatozoa by next generation sequencing. Genom Data *6*, 208-211.

Belleannee, C., Da Silva, N., Shum, W.W., Brown, D., and Breton, S. (2010). Role of purinergic signaling pathways in V-ATPase recruitment to apical membrane of acidifying epididymal clear cells. Am J Physiol Cell Physiol *298*, C817-830.

Belleannee, C., Labas, V., Teixeira-Gomes, A.P., Gatti, J.L., Dacheux, J.L., and Dacheux, F. (2011). Identification of luminal and secreted proteins in bull epididymis. J Proteomics *74*, 59-78.

Belleannee, C., Thimon, V., and Sullivan, R. (2012). Region-specific gene expression in the epididymis. Cell Tissue Res *349*, 717-731.

Bernal-Manas, C.M., Beltran-Frutos, E., Ferrer, C., Seco-Rovira, V., Pinart, E., Briz, M.D., Bonet, S., Canteras, M., and Pastor, L.M. (2014). Cell proliferation in the seminiferous and epididymal epithelia of Sus domesticus. Theriogenology *81*, 702-711.

Bernard Robaire, B.T.H., Marie-Claire Orgebin-Crist (2006). The Epididymis. In Knobil and Neill's Physiology of Reproduction, J.D. Neill, ed. (Elservier), pp. 1071-1148.

Bjorkgren, I., and Sipila, P. (2019). The impact of epididymal proteins on sperm function. Reproduction.

Breton, S., and Brown, D. (2013). Regulation of luminal acidification by the V-ATPase. Physiology (Bethesda) *28*, 318-329.

Breton, S., Ruan, Y.C., Park, Y.J., and Kim, B. (2016). Regulation of epithelial function, differentiation, and remodeling in the epididymis. Asian J Androl *18*, 3-9.

Browne, J.A., Yang, R., Leir, S.H., Eggener, S.E., and Harris, A. (2016). Expression profiles of human epididymis epithelial cells reveal the functional diversity of caput, corpus and cauda regions. Mol Hum Reprod *22*, 69-82.

Carvajal, G., Brukman, N.G., Weigel Munoz, M., Battistone, M.A., Guazzone, V.A., Ikawa, M., Haruhiko, M., Lustig, L., Breton, S., and Cuasnicu, P.S. (2018). Impaired male fertility and abnormal epididymal epithelium differentiation in mice lacking CRISP1 and CRISP4. Sci Rep *8*, 17531.

Chen, S., Zhou, Y., Chen, Y., and Gu, J. (2018). fastp: an ultra-fast all-in-one FASTQ preprocessor. Bioinformatics *34*, i884-i890.

462 Cooper, T.G., and Yeung, C.-H. (2010). Physiology of Sperm Maturation and
463 Fertilization. In Andrology, pp. 61-85.
464 Cornwall, G.A. (2009). New insights into epididymal biology and function. Hum
465 Reprod Update *15*, 213-227.
466 Cornwall, G.A., Collis, R., Xiao, Q., Hsia, N., and Hann, S.R. (2001). B-Myc, a
467 proximal caput epididymal protein, is dependent on androgens and testicular factors for
468 expression. Biol Reprod *64*, 1600-1607.
469 Da Silva, N., Cortez-Retamozo, V., Reinecker, H.C., Wildgruber, M., Hill, E., Brown,
470 D., Swirski, F.K., Pittet, M.J., and Breton, S. (2011). A dense network of dendritic cells
471 populates the murine epididymis. Reproduction *141*, 653-663.
472 Dacheux, J.L., and Dacheux, F. (2014). New insights into epididymal function in
473 relation to sperm maturation. Reproduction *147*, R27-42.
474 Dai, W., Wang, Y., Yang, T., Wang, J., Wu, W., and Gu, J. (2019). Downregulation of
475 exosomal CLEC3B in hepatocellular carcinoma promotes metastasis and angiogenesis
476 via AMPK and VEGF signals. Cell Commun Signal *17*, 113.
477 Dong, J., Hu, Y., Fan, X., Wu, X., Mao, Y., Hu, B., Guo, H., Wen, L., and Tang, F.
478 (2018). Single-cell RNA-seq analysis unveils a prevalent epithelial/mesenchymal
479 hybrid state during mouse organogenesis. Genome Biol *19*, 31.
480 Dube, E., Chan, P.T., Hermo, L., and Cyr, D.G. (2007). Gene expression profiling and
481 its relevance to the blood-epididymal barrier in the human epididymis. Biol Reprod *76*,
482 1034-1044.
483 Fan, X., Bialecka, M., Moustakas, I., Lam, E., Torrens-Juaneda, V., Borggreven, N.V.,
484 Trouw, L., Louwe, L.A., Pilgram, G.S.K., Mei, H.*, et al.* (2019). Single-cell
485 reconstruction of follicular remodeling in the human adult ovary. Nat Commun *10*,
486 3164.
487 Green, C.D., Ma, Q., Manske, G.L., Shami, A.N., Zheng, X., Marini, S., Moritz, L.,
488 Sultan, C., Gurczynski, S.J., Moore, B.B.*, et al.* (2018). A Comprehensive Roadmap of
489 Murine Spermatogenesis Defined by Single-Cell RNA-Seq. Dev Cell *46*, 651-667
490 e610.
491 Guo, J., Grow, E.J., Mlcochova, H., Maher, G.J., Lindskog, C., Nie, X., Guo, Y., Takei,
492 Y., Yun, J., Cai, L.*, et al.* (2018). The adult human testis transcriptional cell atlas. Cell
493 Res *28*, 1141-1157.
494 Guyonnet, B., Marot, G., Dacheux, J.L., Mercat, M.J., Schwob, S., Jaffrezic, F., and
495 Gatti, J.L. (2009). The adult boar testicular and epididymal transcriptomes. BMC
496 Genomics *10*, 369.
497 Han, X., Wang, R., Zhou, Y., Fei, L., Sun, H., Lai, S., Saadatpour, A., Zhou, Z., Chen,
498 H., Ye, F.*, et al.* (2018). Mapping the Mouse Cell Atlas by Microwell-Seq. Cell *172*,
499 1091-1107 e1017.
500 Johnston, D.S., Jelinsky, S.A., Bang, H.J., DiCandeloro, P., Wilson, E., Kopf, G.S., and
501 Turner, T.T. (2005). The mouse epididymal transcriptome: transcriptional profiling of
502 segmental gene expression in the epididymis. Biol Reprod *73*, 404-413.
503 Jones, R.C., and Murdoch, R.N. (1996). Regulation of the motility and metabolism of
504 spermatozoa for storage in the epididymis of eutherian and marsupial mammals.
505 Reprod Fertil Dev *8*, 553-568.

506 Kim, B., Roy, J., Shum, W.W., Da Silva, N., and Breton, S. (2015). Role of testicular
507 luminal factors on Basal cell elongation and proliferation in the mouse epididymis. Biol
508 Reprod *92*, 9.
509 Liu, J., Liu, Z., Liu, Q., Li, L., Fan, X., Wen, T., and An, G. (2018). CLEC3B is
510 downregulated and inhibits proliferation in clear cell renal cell carcinoma. Oncol Rep
511 *40*, 2023-2035.
512 Ma, L., Yu, H., Ni, Z., Hu, S., Ma, W., Chu, C., Liu, Q., and Zhang, Y. (2013). Spink13,
513 an epididymis-specific gene of the Kazal-type serine protease inhibitor (SPINK) family,
514 is essential for the acrosomal integrity and male fertility. J Biol Chem *288*,
515 10154-10165.
516 Martinez-Garcia, F., Regadera, J., Cobo, P., Palacios, J., Paniagua, R., and Nistal, M.
517 (1995). The apical mitochondria-rich cells of the mammalian epididymis. Andrologia
518 *27*, 195-206.
519 Oh, J., Lee, J., Woo, J.M., Choi, E., Park, I., Han, C., Baek, N., Lee, H., Kim, D.H., and
520 Cho, C. (2006). Systematic identification and integrative analysis of novel genes
521 expressed specifically or predominantly in mouse epididymis. BMC Genomics *7*, 314.
522 Pastor-Soler, N., Bagnis, C., Sabolic, I., Tyszkowski, R., McKee, M., Van Hoek, A.,
523 Breton, S., and Brown, D. (2001). Aquaporin 9 expression along the male reproductive
524 tract. Biol Reprod *65*, 384-393.
525 Pastor-Soler, N.M., Fisher, J.S., Sharpe, R., Hill, E., Van Hoek, A., Brown, D., and
526 Breton, S. (2010). Aquaporin 9 expression in the developing rat epididymis is
527 modulated by steroid hormones. Reproduction *139*, 613-621.
528 Pihlajamaa, P., Sahu, B., Lyly, L., Aittomaki, V., Hautaniemi, S., and Janne, O.A.
529 (2014). Tissue-specific pioneer factors associate with androgen receptor cistromes and
530 transcription programs. EMBO J *33*, 312-326.
531 Pinel, L., Mandon, M., and Cyr, D.G. (2019). Tissue regeneration and the epididymal
532 stem cell. Andrology *7*, 618-630.
533 Reyes-Moreno, C., Laflamme, J., Frenette, G., Sirard, M.A., and Sullivan, R. (2008).
534 Spermatozoa modulate epididymal cell proliferation and protein secretion in vitro. Mol
535 Reprod Dev *75*, 512-520.
536 Sharma, U., Sun, F., Conine, C.C., Reichholf, B., Kukreja, S., Herzog, V.A., Ameres,
537 S.L., and Rando, O.J. (2018). Small RNAs Are Trafficked from the Epididymis to
538 Developing Mammalian Sperm. Dev Cell *46*, 481-494 e486.
539 Shum, W.W., Ruan, Y.C., Da Silva, N., and Breton, S. (2011). Establishment of
540 cell-cell cross talk in the epididymis: control of luminal acidification. J Androl *32*,
541 576-586.
542 Shum, W.W., Smith, T.B., Cortez-Retamozo, V., Grigoryeva, L.S., Roy, J.W., Hill, E.,
543 Pittet, M.J., Breton, S., and Da Silva, N. (2014). Epithelial basal cells are distinct from
544 dendritic cells and macrophages in the mouse epididymis. Biol Reprod *90*, 90.
545 Sipila, P., and Bjorkgren, I. (2016). Segment-specific regulation of epididymal gene
546 expression. Reproduction *152*, R91-99.
547 Smith, T., Heger, A., and Sudbery, I. (2017). UMI-tools: modeling sequencing errors in
548 Unique Molecular Identifiers to improve quantification accuracy. Genome Res *27*,
549 491-499.

Stelzer G, R.R., Plaschkes I, Zimmerman S, Twik M, Fishilevich S, Iny Stein T, Nudel R, Lieder I, Mazor Y, Kaplan S, Dahary D, Warshawsky D, Guan - Golan Y, Kohn A, Rappaport N, Safran M, and Lancet D. (2016). The GeneCards Suite: From Gene Data Mining to Disease Genome Sequence Analysis (Current Protocols in Bioinformatics), pp. 1.30.31-31.30.33.

Stuart, T., Butler, A., Hoffman, P., Hafemeister, C., Papalexi, E., Mauck, W.M., 3rd, Hao, Y., Stoeckius, M., Smibert, P., and Satija, R. (2019). Comprehensive Integration of Single-Cell Data. Cell *177*, 1888-1902 e1821.

Sullivan, R., Legare, C., Lamontagne-Proulx, J., Breton, S., and Soulet, D. (2019). Revisiting structure/functions of the human epididymis. Andrology *7*, 748-757.

Tabula Muris, C., Overall, c., Logistical, c., Organ, c., processing, Library, p., sequencing, Computational data, a., Cell type, a., Writing, g., *et al.* (2018). Single-cell transcriptomics of 20 mouse organs creates a Tabula Muris. Nature *562*, 367-372.

Thimon, V., Koukoui, O., Calvo, E., and Sullivan, R. (2007). Region-specific gene expression profiling along the human epididymis. Mol Hum Reprod *13*, 691-704.

Trigg, N.A., Eamens, A.L., and Nixon, B. (2019). The contribution of epididymosomes to the sperm small RNA profile. Reproduction.

Turner, T.T. (1995). On the epididymis and its role in the development of the fertile ejaculate. J Androl *16*, 292-298.

Turunen, H.T., Sipila, P., Strauss, L., Bjorkgren, I., Huhtaniemi, I., and Poutanen, M. (2012). Loss of Bmyc results in increased apoptosis associated with upregulation of Myc expression in juvenile murine testis. Reproduction *144*, 495-503.

Wang, M., Liu, X., Chang, G., Chen, Y., An, G., Yan, L., Gao, S., Xu, Y., Cui, Y., Dong, J., *et al.* (2018). Single-Cell RNA Sequencing Analysis Reveals Sequential Cell Fate Transition during Human Spermatogenesis. Cell Stem Cell *23*, 599-614 e594.

Xie, S., Xu, J., Ma, W., Liu, Q., Han, J., Yao, G., Huang, X., and Zhang, Y. (2013). Lcn5 promoter directs the region-specific expression of cre recombinase in caput epididymidis of transgenic mice. Biol Reprod *88*, 71.

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

Zhang, J.S., Liu, Q., Li, Y.M., Hall, S.H., French, F.S., and Zhang, Y.L. (2006). Genome-wide profiling of segmental-regulated transcriptomes in human epididymis using oligo microarray. Mol Cell Endocrinol *250*, 169-177.

Zhang, Y., Yan, Z., Qin, Q., Nisenblat, V., Chang, H.M., Yu, Y., Wang, T., Lu, C., Yang, M., Yang, S., *et al.* (2018). Transcriptome Landscape of Human Folliculogenesis Reveals Oocyte and Granulosa Cell Interactions. Mol Cell *72*, 1021-1034 e1024.

Figure Legend

Figure 1. Single cell RNA sequencing of mouse epididymis regions. (A) Seven epididymis cell populations identified. Left: UMAP of the cell clustering based on all cells. Right: UMAP of the cell clustering based on cells from each epididymis region.(B) Number of single cells identified in each region.(C) Cell populations with known marker genes. (D) Top 10 marker genes for each cell population.

Figure 2: Study of the novel epididymis sub-population C1. (A) Specific expression of C1 sub-population marker genes. (B) Dotplot representation of DAVID enriched terms (DAVID terms selected based on FDR < 10^(-5)). (C) Increasing of C1 sub-population in the epididymal distal regions. The spatial proportion of each sub-population is also illustrated here.

Figure 3. Sub-populations of principle cells identified. (A) UMAP representation of the sub-populations. Left: All single cells identified in principle cells (C0). Right: Split by epididymis regions. (B) Higher expression of mitochondrial genes in corpus and cauda. Only principle cells were included in this comparison. (C) Principle cell marker Aqp9 expression in each sub-population. (D) Patterns Of epididymis sub-population spatial distributions.

Figure 4. Features of principle cell sub-populations. (A) Heatmap showing top 10 marker genes for each sub-population. (B) UMAP showing representative marker genes for the principle sub-populations. (C) GO enrichment analysis for each sub-population.

Figure 5. Study of principle cell c0.3 sub-population. (A) c0.3 epididymis sub-population is enriched in the caput region of epididymis. Y-axis: proportion of cells from c0.3 (total: cells in principle cell population) in each region. (B) Specific expression of Defensin genes in c0.3 principle sub-population. (C) GO enrichment analysis for the marker genes in c0.3 principle sub-population. Genes related to "Defense response to bacterium" are significantly enriched.

Figure 6. Regional regulation patterns in epididymal principle sub-populations. (A) Top 10 regional differentially expressed transcription factors (TFs) for each principle sub-population. (B) Heatmap representation of the correlation matrix for the regional differentially expressed TFs. Spearman's *rho* was used for correlation calculation. (C) Examples of TFs with high regional and sub-population expression variations. (D) Examples of TFs with high specificities of regional and sub-population expression. (E) Examples of TFs specifically expressed in one epididymal region. (F) TF-TF correlation network showing the network of Nupr1.

Figure S1. Expression of known marker genes. Marker genes from six known epididymis cell types were illustrated in UMAP plot.

638

639 **Figure S2. GO enrichment analysis for epididymis C5 cell population.** GO terms
640 related to vascular were enriched.

641

642 **Figure S3. Density plot of MREDs of all expressed genes.** 95% quantile value is
643 highlighted by green dash line.

644

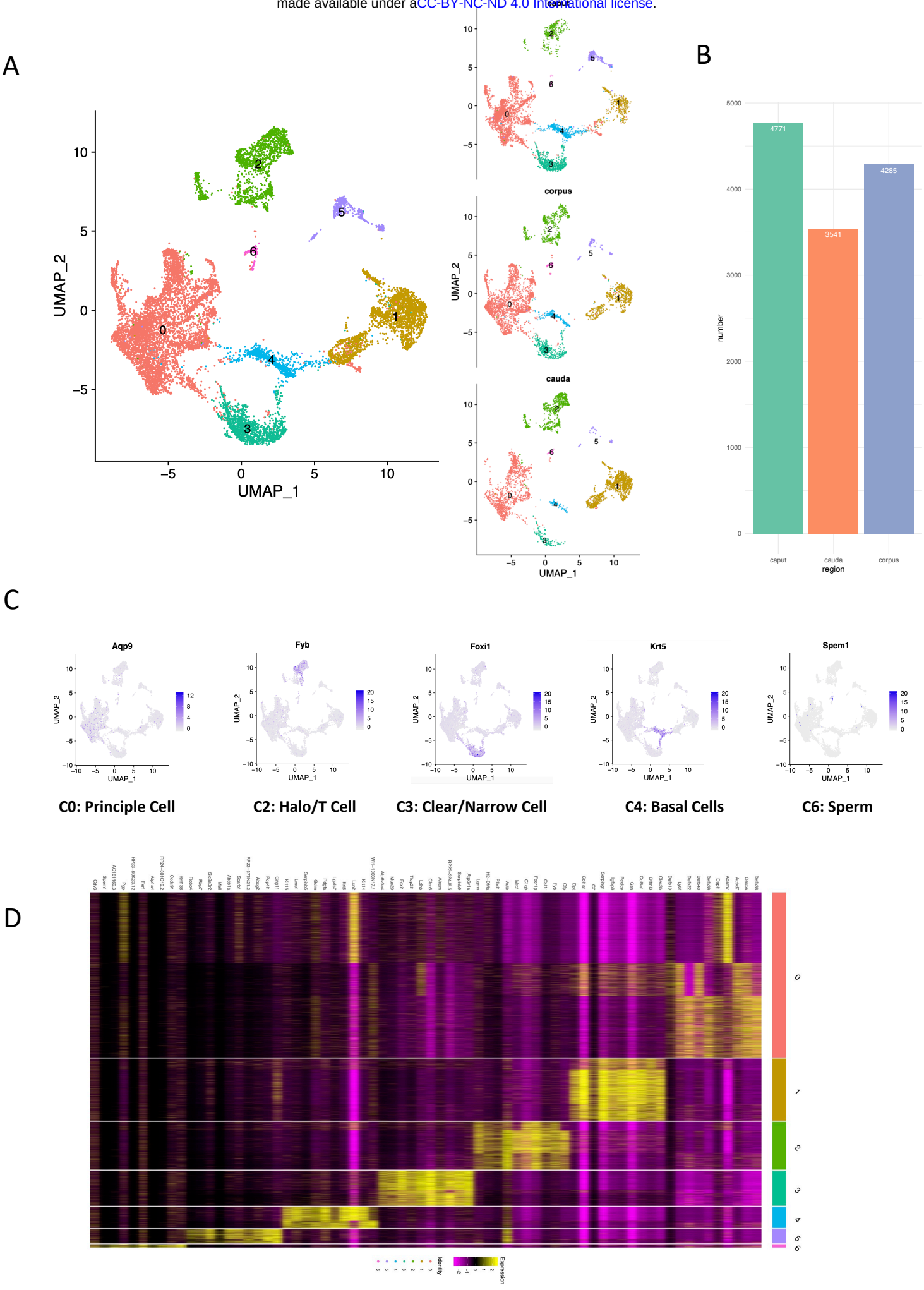


Figure 1

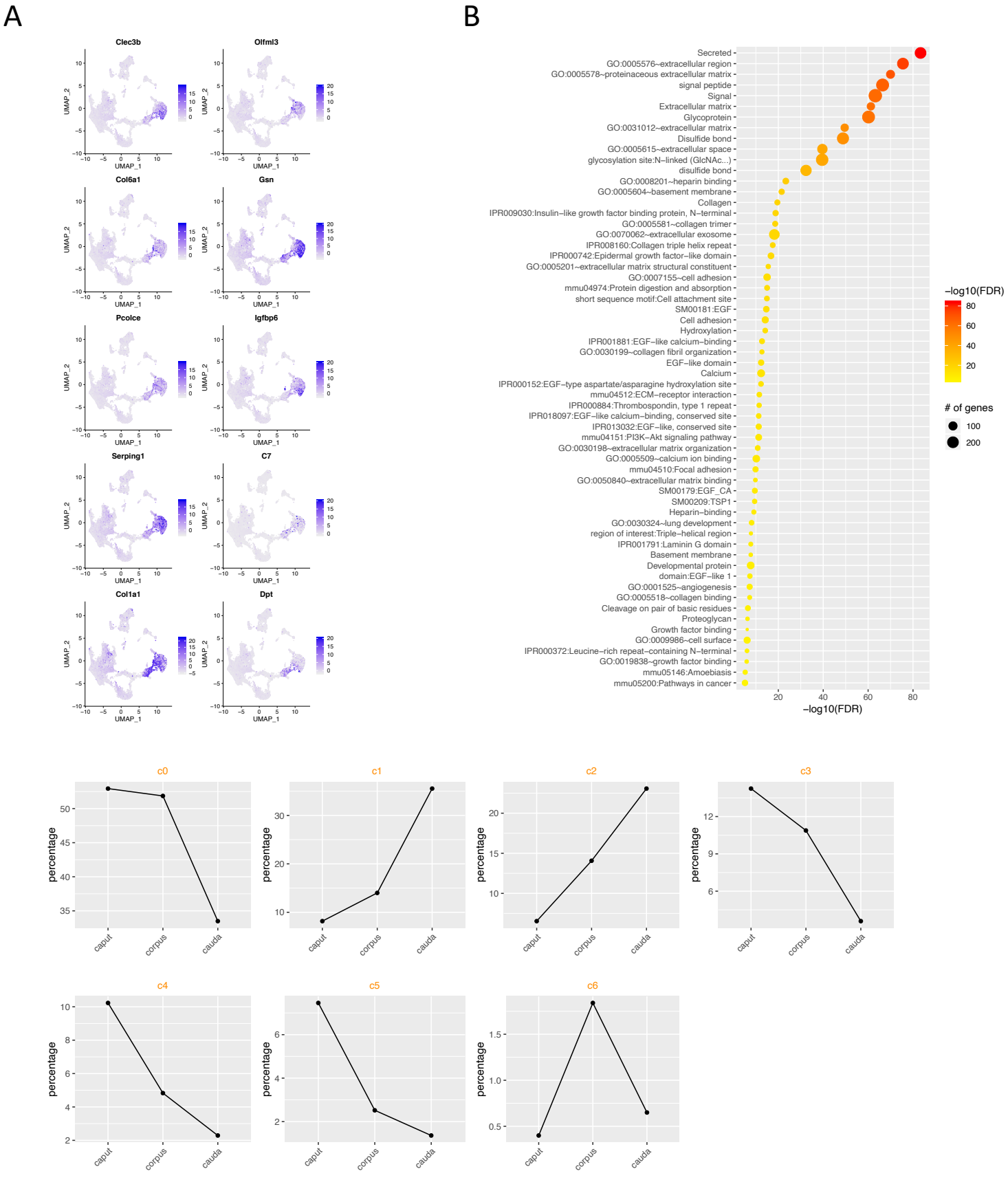


Figure 2

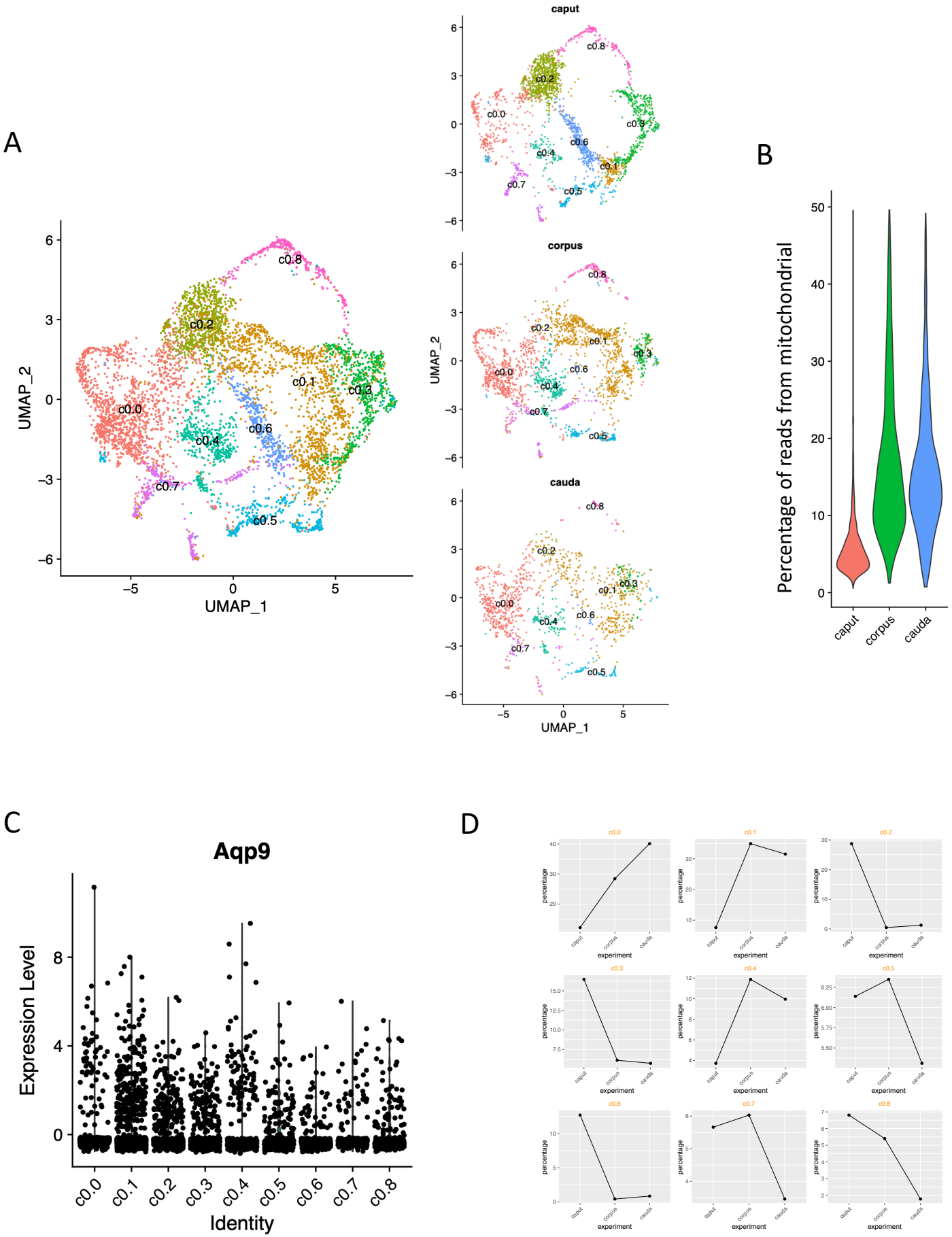
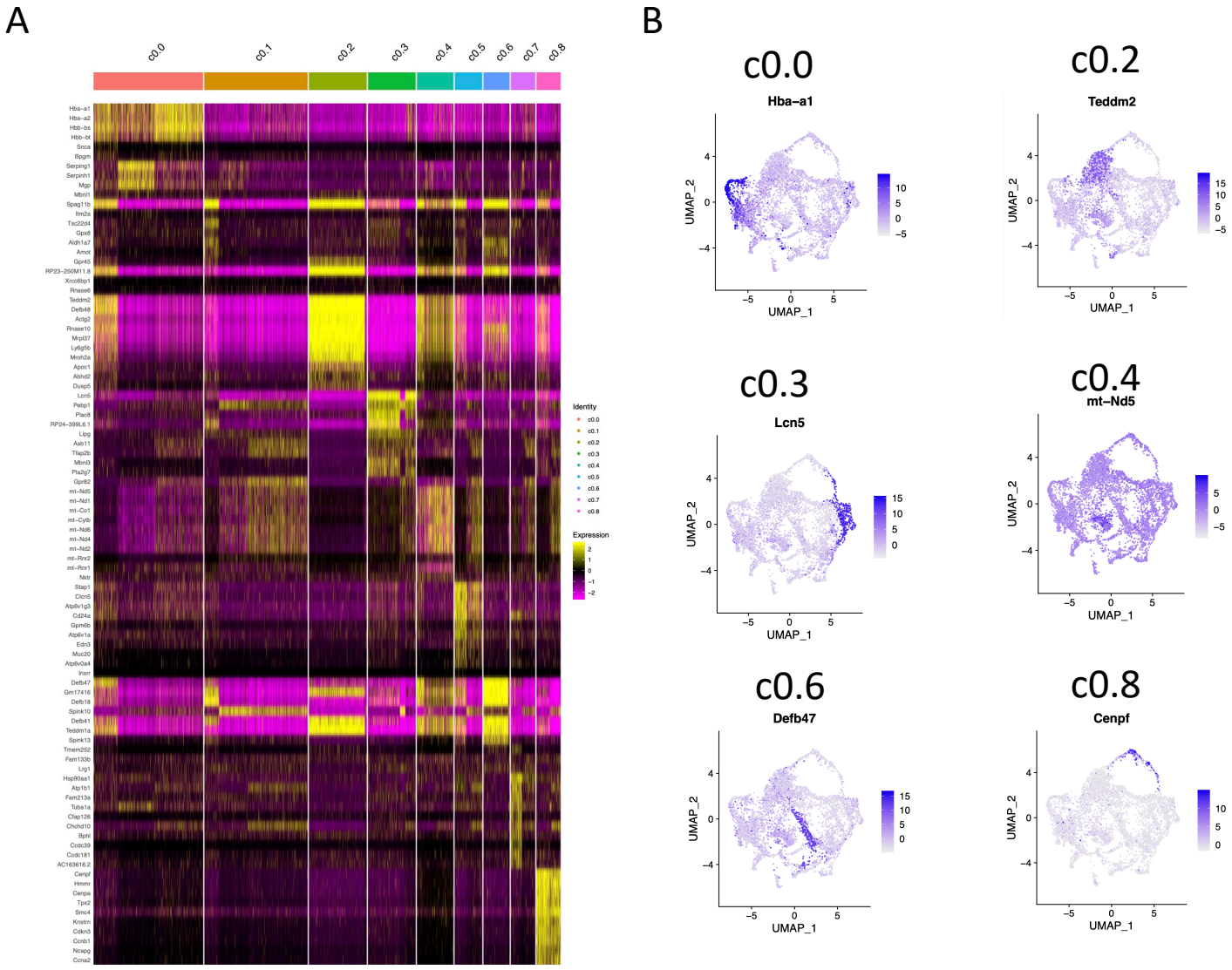


Figure 3



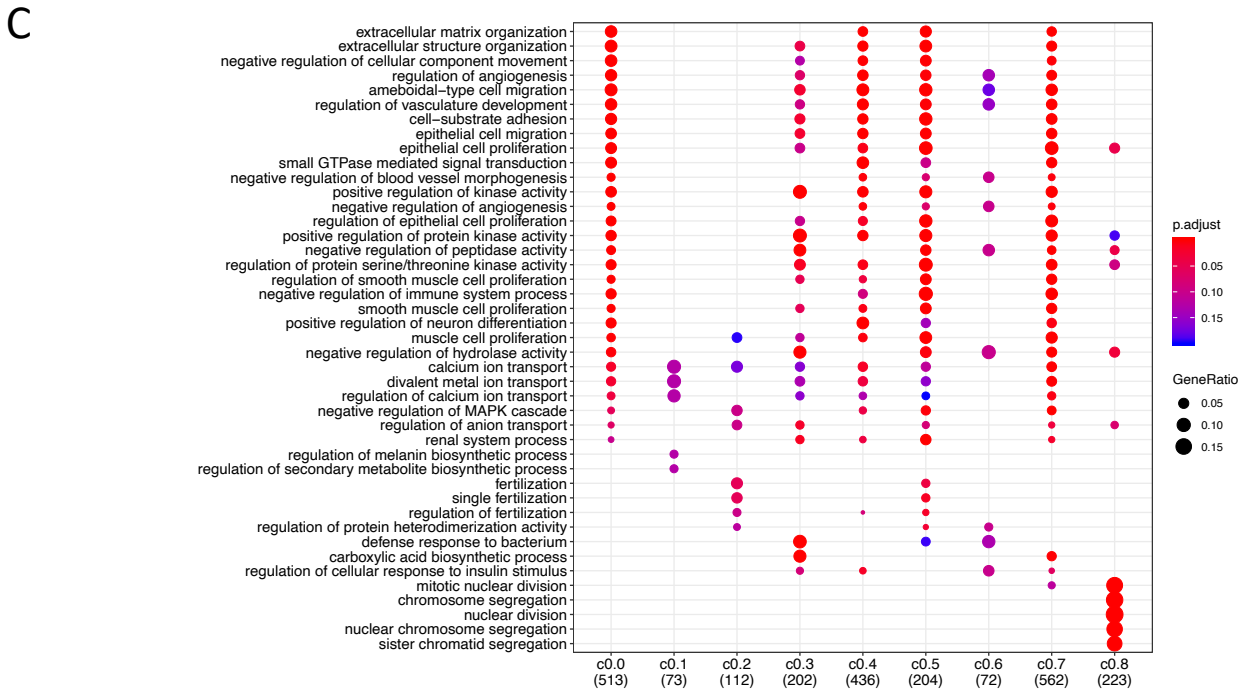
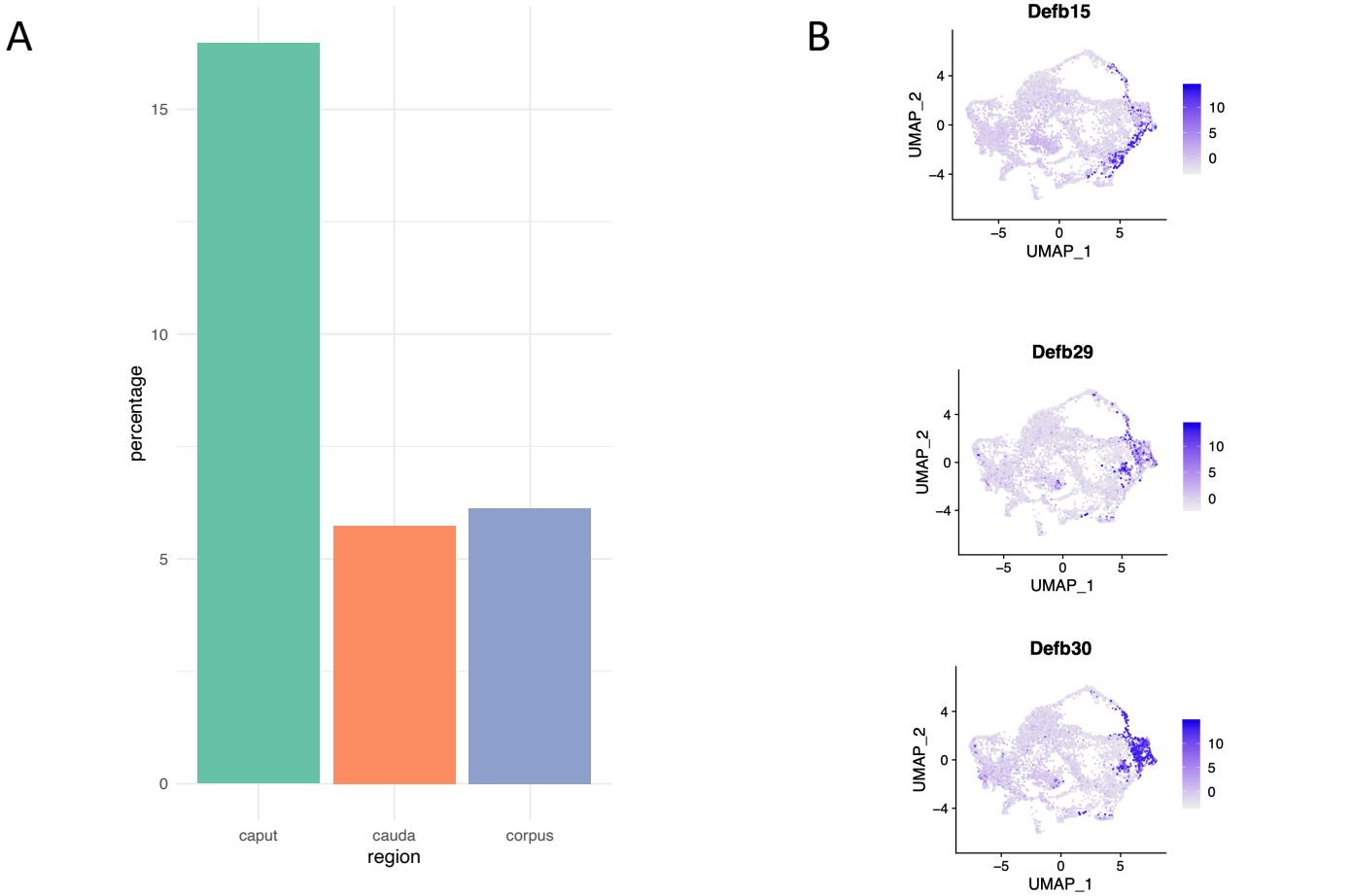


Figure 4



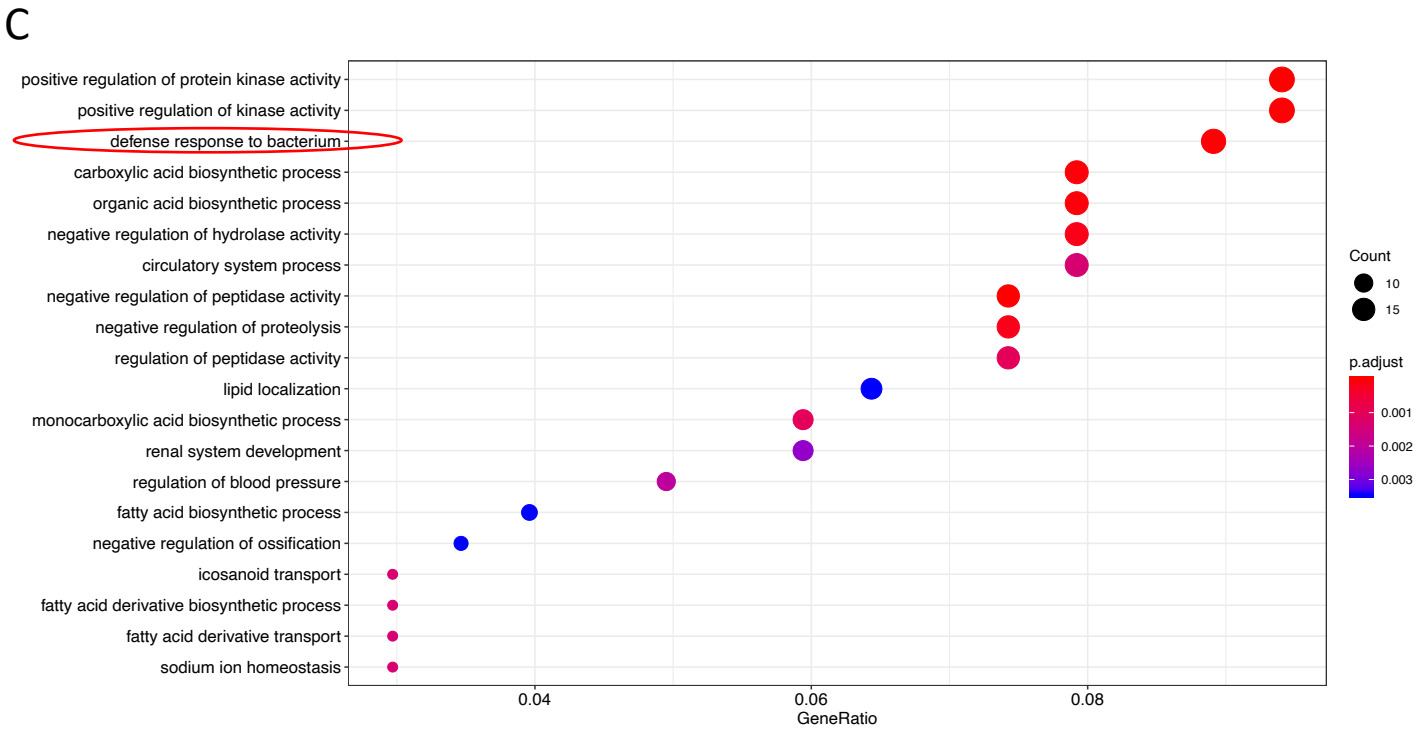


Figure 5

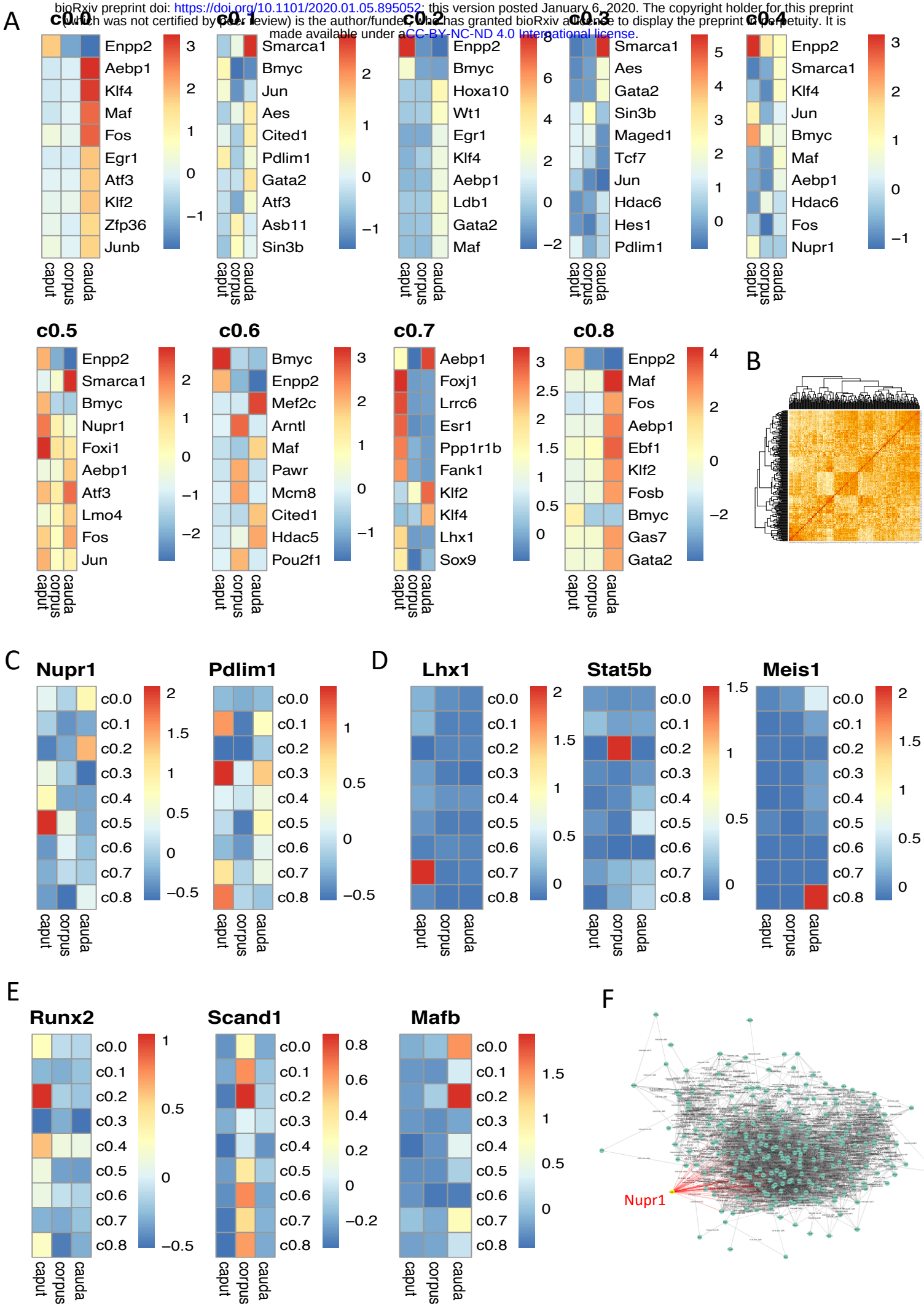


Figure 6