scientific data

Check for updates

OPEN

DATA DESCRIPTOR

Bulk and single-cell transcriptome datasets of the mouse fetal and adult rete ovarii and surrounding tissues

Dilara N. Anbarci[1], Rebecca O'Rourke[2], Yu Xiang[1], Derek T. Peters[1], Blanche Capel[1] & Jennifer McKey[2] ✉

The rete ovarii (RO) is an epithelial structure that arises during development in close proximity to the ovary and persists throughout adulthood. However, the functional significance of the RO remains elusive, and it is absent from recent discussions of female reproductive anatomy. The RO comprises three regions: the intraovarian rete within the ovary, the extraovarian rete in the periovarian tissue, and the connecting rete linking the two. We hypothesize that the RO plays a pivotal role in ovarian homeostasis and responses to physiological changes. To begin to uncover the nature and function of RO cells, we conducted transcriptomic profiling of the RO. This study presents three datasets, and reports our analysis and quality control approaches for bulk, single-cell, and nucleus-level transcriptomics of the fetal and adult RO tissues using the *Pax8-rtTA;Tre-H2B-GFP* mouse line, where all RO regions express nuclear GFP. The integration and rigorous validation of these datasets will advance our understanding of the RO's roles in ovarian development, female maturation, and adult female fertility.

Background & Summary

The rete ovarii (RO) is an epithelial structure that develops in close association with the ovary during fetal life and remains in adulthood in mammals[1,2]. Despite the significant architecture of this ovarian appendage, and the fact that it is highly conserved in mammals[3–5], the function of the RO has not yet been determined, and it has disappeared from recent descriptions of female reproductive anatomy. The RO is the female homolog to the rete testis, thought to arise from the mesonephric tubules[1,6]. The RO is divided into three regions; the intraovarian rete (IOR), which resides inside the ovary, the extraovarian rete (EOR) located in the periovarian tissue, and the connecting rete (CR), which links the EOR and IOR[6]. Using the mouse as a model system we developed tissue-clearing and 3D-imaging methods using lightsheet microscopy[7] that allowed us to observe the RO in unprecedented detail[2]. The bipotential rete structure first appears in both sexes as a PAX8+ population of cells at the interface between the dorsal aspect of the gonad and the mesonephros[8,9]. This population of PAX8+ cells was recently shown to give rise to a subset of gonadal supporting cells in both sexes[9], and remains after gonadal sex determination as the IOR in females. The EOR begins to develop from the mesonephric tubules as a blind tubular epithelium that connects to the CR starting around E14.5[2]. In the adult, the IOR has regressed to a smaller population of cells within the ovary, while the EOR has significantly expanded into a large network of tubules[2,10]. The RO is in a unique location between the ovary and extraovarian milieu, where vascular and neuronal networks enter the ovary. We hypothesize that proximity to vascular and neuronal networks might allow the RO to sense homeostasis and convey information to the adult ovary. To better understand the nature and function of cells within the RO, we performed unbiased high-throughput transcriptome analysis of the RO and tissues that surround it. Here, we report three datasets that sequence the fetal and adult RO transcriptome in bulk, and at the single cell and single nucleus level[11]. We use these datasets to identify RO-specific gene expression signatures, and to further characterize gene expression differences between the three regions of the RO during development and in the adult. To allow for accurate capture and enrichment of RO cells in our

[1]Department of Cell Biology, Duke University Medical Center, Durham, NC, 27710, USA. [2]Section of Developmental Biology, Department of Pediatrics, University of Colorado Anschutz Medical Campus, Aurora, CO, 80045, USA. ✉e-mail: jennifer.mckey@cuanschutz.edu

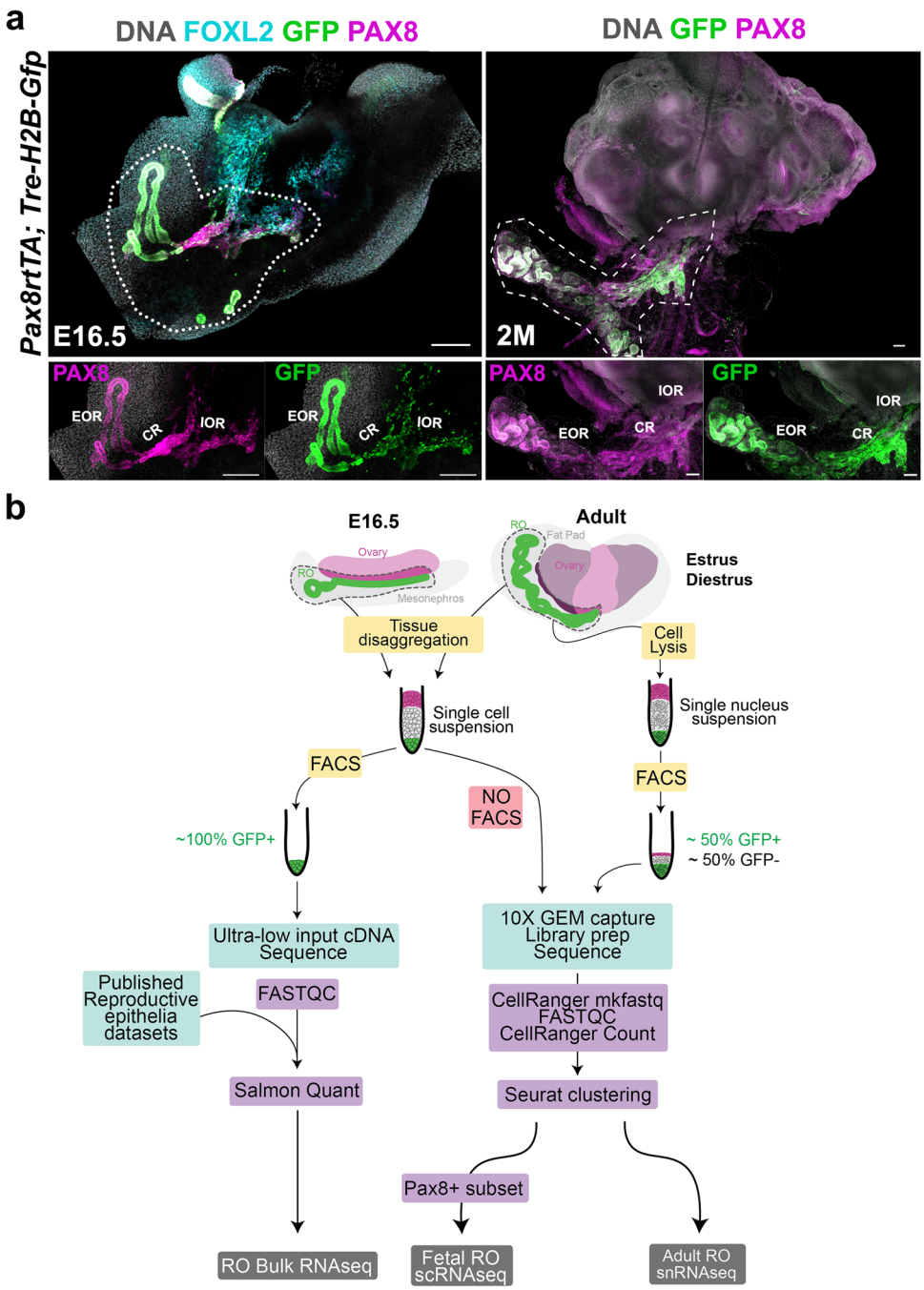

Fig. 1 Overview of experimental design. (**a**) Maximum intensity projection of confocal Z-stack images of the ovary/mesonephros/oviduct complex of the E16.5 *Pax8rtTA; Tre-H2b-GFP* mouse embryo (*left panel*), and the ovary/fat pad complex of the 2-month-old *Pax8rtTA; Tre-H2b-GFP* mouse. The rete ovarii is labeled with PAX8 antibodies in *magenta* and GFP antibodies in *green* in both panels, and the ovary is labeled with FOXL2 antibodies in *cyan* in the E16.5 (left panel). Dotted lines around the rete ovarii estimate the amount of surrounding tissue collected for further analysis. *Scale bars, 100um*. EOR, extraovarian rete; CR, connecting rete; IOR, intraovarian rete; MT, mesonephric tubules. (**b**) Illustration of the workflow and analysis pipelines for the generation of each dataset. The areas outlined by the dotted lines in the E16.5 and adult ovary represent the RO and surrounding mesenchyme. These areas were dissected as close to the ovary as possible, but a small fraction of ovarian tissue remained in the sample, represented by the brighter pink area. Tissues were dissociated and single cell suspensions were generated for downstream processing.

sequencing samples, we used the *Pax8-rtTA; Tre-H2B-GFP* mouse line, in which all regions of the RO express nuclear GFP at all stages. Figure 1 shows the expression domain of GFP in the fetal (Figs. 1a, *E16.5*) and adult (Fig. 1a, 2-*Month-old, 2 M*) mouse ovary, along with an illustration of the sample collection and analysis workflow for each dataset generated (Fig. 1b). These datasets complement recent publications that specifically focused

on the intraovarian region of the RO in mouse and human[9,12,13], and fill a critical gap in knowledge by providing data for all three regions of the RO in both adult and fetal stages. Integration and careful validation of all these datasets will pave the way towards understanding the roles of the RO, and its function in ovary development, female maturation, and adult female fertility.

Methods

Experimental overview. We set out to sequence the transcriptome of the RO at E16.5 and in the 2-3 month-old (2 M) adult. To facilitate the capture of RO cells, we took advantage of the *Pax8rtTA; Tre-H2b-GFP* transgenic line, in which all regions of the RO express GFP during fetal development and in the adult (Fig. 1a). We collected samples of the RO and surrounding tissue from these mice and processed them for either Bulk RNAseq (E16.5 and adult), single-cell RNAseq (E16.5), and single-nucleus RNAseq (adult) (Fig. 1b). Standard pipelines were used in each case. For BulkRNAseq, we collected GFP+ cells using FACS analysis. Since the RO represents so few cells, we used the Ultra-Low input SMARTseq kit provided by Takara Clontech, to prepare cDNA libraries (Fig. 1b, l*eft*). For the E16.5 single-cell sample, we did not FAC-sort the cells, but rather collected the entire ovarian capsule for single-cell capture using 10X Genomics Chromium (Fig. 1b, m*iddle*). For the adult samples at Estrus and Diestrus, we learned from our E16.5 experience that it would be better to enrich the sample for RO cells and chose to FAC-sort approximately 50%GFP+ and 50%GFP-nuclei for snRNA sequencing. All cDNA libraries were sequenced using a NovaSeq 6000, resulting FASTQ read files were quality checked using FASTQC software and standard downstream analysis was performed. For Bulk RNAseq, transcript abundance and differential gene expression were obtained using Salmon and DESeq 2, while 10X CellRanger and Seurat software were used to analyze single cell and single nucleus datasets (Fig. 1b).

Mice. All experiments were performed on *Pax8rtTA; Tre-H2b-Gfp* (PTG) female mice. The *Pax8-rtTA* (*B6.Cg-Tg(Pax8-rtTA2S*M2)1Koes/J;* RRID: IMSR_JAX:007176) and *Tre-H2BGFP (Tg(tetO-HIST1H2BJ/GFP)47Efu/J*; RRID:IMSR_JAX:005104) lines were previously described[14,15] and maintained on a mixed CD-1 / C57BL/6 J background. To collect embryos at specific developmental stages, males were set up in timed matings with several females. Females were checked daily for the presence of a vaginal plug. Date of the plug was considered embryonic day 0.5. For adult timepoints, female carriers of the PTG alleles were weaned at 28 days and kept in single-sex housing until they reached 8–10 weeks old. Adult females for adult timepoints or pregnant dams for fetal timepoints were given a constant and exclusive doxycycline diet of 625 mg/kg (Teklad Envigo TD.01306) 3 days prior to tissue collection to induce GFP expression in *Pax8*+ cells. For collection of adult tissue at specific stages of the estrous cycle, vaginal swabs were collected and vaginal cytology was examined to determine estrous stage based on previously described standard metrics[16]. Female mice were estrous-tracked for a week prior to tissue collection to ensure they displayed typical cycling. Secondary validation of the estrous stage was performed by visual analysis of uterine swelling during tissue collection. Any mismatched observations resulted in tissue being omitted from further analysis. All mice were housed and handled in accordance with National Institutes of Health guidelines, in a barrier facility maintained at a temperature of 22 ± 0.1 °C, 30–70% humidity, within individually ventilated cages (Allentown; PNC75JU160SPCD3), and a controlled 12 h on / 12 h off light cycle. All experiments were conducted with the approval of the Duke University Medical Center Institutional Animal Care and Use Committee (IACUC protocol # A089-20-04 9 N).

Whole-mount Immunostaining for confocal imaging. E16.5 ovary/mesonephros/Müllerian duct complexes and adult ovary/fat pad complexes were dissected in PBS-/- and fixed for 30 minutes (fetal) or 1 hour (adult) at room temperature in 4%PFA/PBS. Following two 15 minute PBS washes, samples were gradually dehydrated in MeOH dilutions (25% MeOH/PBS; 50% MeOH/PBS; 75% MeOH/PBS; 100% MeOH) for 15 minutes (fetal) or 30 minutes (adult) each at room temperature. Samples were stored at −20 °C in 100% MeOH until required for staining. Before staining, samples were gradually rehydrated into PBS through 10 minutes (fetal) or 20 minutes (adult) washes in a reverse methanol gradient (75% MeOH/PBS; 50% MeOH/PBS; 25% MeOH/PBS), and transferred to PBS 0.1% Triton X-100 for 30 min. Samples were then transferred to blocking solution (PBS; 1% Triton X-100; 10% horse serum) for one hour, and incubated at 4 °C overnight (3 nights for adult) in primary antibodies diluted in blocking solution (chicken anti-GFP, 1:1000, RRID AB_300798; rabbit anti PAX8, 1:500, RRID AB_2236705; goat anti FOXL2, 1:250, RRID AB_2106188). The next day, samples were washed three times for 30 minutes in PBS 0.1% Triton X-100 and incubated at 4 °C overnight in secondary antibodies and Hoechst vital dye diluted 1:500 in blocking solution (AF647 Donkey anti rabbit, RRID AB_2492288; AF488 donkey anti chicken, RRID AB_2340375; Cy3 Donkey anti goat; RRID AB_2307351). On day 3, samples were washed twice for 15 minutes in PBS 0.1% Triton X-100 and transferred to PBS at 4 °C until ready to mount for confocal imaging (1 hour – 48 hours). Confocal images were captured in the longitudinal plane on a Zeiss LSM880 confocal microscope and the affiliated Zen software (Carl Zeiss, Inc., Germany) using a 10X objective.

Fetal tissue collection. Embryonic day (E)16.5 fetuses were harvested from pregnant *Tre-H2BGFP*$^{tg/tg}$ dams crossed to a *Pax8rtTA*$^{tg/+}$ male. Embryos positive for both alleles were identified by the presence of green fluorescence in the urogenital epithelia. Male fetuses were discarded, and ovary/mesonephros/oviduct complexes from females were dissected in RNase-free PBS-/-, with as little removal of surrounding tissue as possible to ensure retention and survival of RO cells. The GFP signal was then used to perform close dissection of the RO and surrounding ovarian capsule, and the oviduct was discarded. To ensure capture of the IOR and CR, only the dorso-lateral-most portion of the ovary was conserved. The samples were kept on ice until all fetuses were dissected.

Adult tissue collection. Ovaries and periovarian fat pads (in which the adult RO is embedded) were harvested from *Pax8rtTA*; *Tre-H2BGFP*$^{tg/tg}$ adult females. The RO was located within the fat pad by green fluorescence and closely dissected in RNase-free PBS-/-, with removal of as much surrounding tissue as possible to ensure enrichment of RO cells. To ensure capture of the IOR and CR, the dorso-lateral-most portion of the ovary was conserved, and the rest of the ovary was discarded. The samples were kept on ice until all samples were dissected. For single-nucleus RNA sequencing (snRNAseq) analysis of the adult RO, samples were collected from 5 mice in early diestrus and 5 mice in estrus (total of 10 ROs in each group), snap frozen in liquid nitrogen, and stored at −80 °C until further processing.

Bulk RNA sequencing. For each timepoint (E16.5 and 2month), three biological replicates were processed as follows. In each replicate, 4-5 ROs were pooled, thus 2-3 females are represented in each biological replicate. For the E16.5 timepoint, two different litters were used, replicate 1 is a single litter, and replicates 2 and 3 are from the second litter. Samples were washed in RNase-free PBS-/- and incubated at 37 °C in fresh 1X TrypLE (Thermo Fisher Scientific, catalog #12563029) for 20 minutes. TrypLE was then aspirated off, and samples were resuspended in chilled RNase-free PBS with 3% bovine serum albumin (BSA) and pipetted gently up and down for about 2 min to disaggregate cells. Cells were pipetted through a 0.32μm cell strainer into FAC-sorting filter tubes (Corning Falcon cat # 352235). Single-cell suspensions were then FAC-sorted at the Duke Flow Cytometry Shared Resource core on a B-C Astrios Sorter, based on presence or absence of GFP. Cells were sorted into PBS, pelleted, and resuspended in 10.5ul Clontech Lysis buffer (Takara Bio Inc. Cat # 635013). Lysed samples were stored at −80 °C until transferred to the Duke Center for Genomic and Computational Biology Sequencing and Genomic Technologies core facility for cDNA library preparation and next-generation sequencing. RNA quality control was measured on a TapeStation (Agilent Technologies) using High Sensitivity RNA ScreenTape (Agilent Technologies). All samples passed quality control. Sample libraries were prepared with SMART-Seq v4 Ultra Low Input RNA Kit (Takara Clontech Kit Cat# 63488). Libraries were sequenced on a NovaSeq 6000 (Illumina) as 50 bp paired-end reads (~46 M reads/sample). Quality control was performed using FastQC (v 0.11.8). Reads were mapped to the GRCm39 mouse genome using Salmon with default settings[17]. Mapped reads were annotated using the GRCm39 Ensembl Mus musculus gene annotation reference (release 110). Read abundance from Salmon (TPM) values were used for downstream quality control and gene expression analysis. For comparison to other reproductive tissues, FASTQs from published bulk RNAseq datasets were obtained from E16.5 ovary[18] (16.5dpc rep 1,2,3 from GSE117590), adult ovary[19] (control 1, 2, 3 from GSE101906), adult ovarian surface epithelium and adult oviduct[20] (Normal OSE and normal FTE reps 1,2,3 from GSE125016). Each FASTQ was reanalyzed using the same Salmon pipeline as for our RO datasets for accurate gene expression comparison.

Sample preparation for single cell RNA sequencing of the fetal RO. To maximize the number of fresh cells we had available on the day of the experiment, we combined female embryos from two litters at E16.5 and one litter at E17.5 for this experiment, as we did not expect major differences in the transcriptome of the RO between these stages. A total of 28 ROs were collected as described above, pooled and incubated for 12 minutes in 450ul Trypsin 0.05% with 50ul 2.5% Collagenase Type IV at 37 °C. 500ul of chilled PBS 0.3% BSA were then added for mechanical disaggregation by pipetting gently up and down for about 2 minutes. Some tissue chunks were still present, thus the sample was incubated for another 6 minutes in 100ul Trypsin 0.05% with 50ul 2.5% Collagenase Type IV at 37 °C. The sample was then pelleted by gentle centrifugation (5 minutes 500 × g 4 °C) and resuspended in 500ul Red Blood Cell lysis buffer (eBioscience, cat #00-4333-57) and incubated for 3 minutes at room temperature. Cells were pelleted by gentle centrifugation and resuspended in 120ul PBS 0.3% BSA before passing through a 0.32 um cell strainer into a clean 1.5 mL tube. 10ul of the cell suspension was collected for viability assessment using Trypan blue and a hemocytometer. Manual counts determined that cell viability was 87.3%, with 49,500 live cells (~400 cells/ul).

Sample preparation for single nucleus RNA sequencing of adult RO. Pooled snap frozen ROs at diestrus (N = 10) and estrus (N = 10) were further separated into two subgroups for gentle (N = 5) and harsh (N = 5) dissociation to ensure sensitive and highly adherent cell types were present in the final single nucleus suspension. Samples were then processed independently as described in the demonstrated protocol for nuclei isolation provided by 10X Genomics (https://assets.ctfassets.net/an68im79xiti/6x4KMzpIgPgkje01sR1Xgr/9cfb7d859985e5c479aec4e0e501f903/CG000124_Demonstrated_Protocol_Nuclei_isolation_RevE.pdf). Briefly, frozen samples were placed in a dounce with 1 mL of cell lysis buffer (10 mM Tris-HCl pH 7.4; 10 mM NaCl; 3 mM $MgCl_2$; 0.1%NP40) and left to incubate for 2 minutes on ice. Another 1.5 mL of Cell Lysis Buffer was added, tissue was gently homogenized using the dounce, and left to incubate for another 3 minutes on ice. Finally, another 3 mL lysis buffer was added and tissue was homogenized. For the light dissociation subgroups, tissue was gently homogenized, and tissue clumps remained in the solution, while for the harsh dissociation subgroups, the tissue was homogenized until no visible tissue clumps were left. After dissociation, nucleus suspensions were passed through a 70 um cell strainer into a 50 mL conical tube, and filtered again through 40 um filter tips (Millipore Sigma Flowmi® Cell Strainers; BAH136800040) into a 15 ml conical tube. Nucleus suspensions were then pelleted by gentle centrifugation (5 minutes, 500 × g, 4 °C), supernatants were removed, and nuclei pellets were resuspended in 1 mL chilled wash buffer (PBS 1%BSA, 0.2U/ul RNase out, Thermo Fisher cat #0777019). Samples were transferred to 1.5 mL tubes and pelleted by gentle centrifugation, supernatants were removed and pellets were resuspended in 1 mL chilled wash buffer and transferred to FACS filter tubes (Corning Falcon cat # 352235). NucRed live 647 ready probe (Thermo Fisher Scientific, cat # R37106) was added to each tube 30 minutes prior to FAC-sorting to label live nuclei. Single nucleus suspensions were FAC-sorted at the Duke Flow Cytometry Shared Resource core on a B-C Astrios Sorter. NucRed-negative cells were discarded and NucRed-positive cells were sorted based on presence or absence of GFP. Our aim was to collect 50% GFP+ (putative RO) and 50% GFP- cells.

Sample	Timepoint	Read length	Million Read-pairs	Salmon Mapping Rate	SRA Accession #
RO-E16_Rep1_R1	Fetal	50 bp	84.1	86%	SRR26313450
RO-E16_Rep1_R2		50 bp			
RO-E16_Rep2_R1		50 bp	53.6	87%	SRR26313449
RO-E16_Rep2_R2		50 bp			
RO-E16_Rep3_R1		50 bp	58.4	83%	Outlier - not in the repository
RO-E16_Rep3_R2		50 bp			
RO-2M_Rep1_R1	Adult	50 bp	61.1	68%	SRR26313453
RO-2M_Rep1_R2		50 bp			
RO-2M_Rep2_R1		50 bp	73.8	81%	SRR26313452
RO-2M_Rep2_R2		50 bp			
RO-2M_Rep3_R1		50 bp	65.2	86%	Outlier - not in the repository
RO-2M_Rep3_R2		50 bp			

Table 1. Rete ovarii Bulk RNAsequencing run Information.

Final nucleus counts were Estrus: 18054 GFP+ and 16619 GFP-; diestrus: 18337 GFP+; 18337 GFP-. GFP+ and GFP- nuclei for each stage were sorted into a single tube containing 700ul PBS, 1% BSA, pelleted by gentle centrifugation, and resuspended in 35ul PBS, 0.05% BSA.

10X Chromium single cell/nucleus capture, library preparation and sequencing. To capture, label, and generate cDNA libraries of individual cells and nuclei, the 10X genomics Chromium Single Cell 3' Library and Gel Bead Kit v3 following the 10X Genomics User Guide (https://assets.ctfassets.net/an68im79xiti/4tjk4KvXzTWgTs8f3tvUjq/2259891d68c53693e753e1b45e42de2d/CG000183_ChromiumSingleCell3__v3_UG_Rev_C.pdf) was used. Briefly, the single cell / nucleus suspensions, RT-PCR master mix, gel beads and partitioning oil were loaded into a Single Cell A Chip 10X genomics chip, placed into the Chromium controller, and the Chromium single cell A program was run to generate GEMs (Gel Bead-In-EMulsion) that contain RT-PCR enzymes, cell lysates and primers for sequencing, barcoding, and poly-DT sequences. GEMs were then transferred to PCR tubes and the RT-PCR reaction was run to generate barcoded single-cell identified cDNA. Barcoded cDNA was used to make sequencing libraries. Sequencing was performed on an Illumina NovaSeq 6000 S-Prim using paired end 150 cycles 2 × 150 reads by the Duke Center for Genomic and Computational Biology Sequencing and Genomic Technologies core facility.

10X single cell and single nucleus RNAseq data analysis. The 10X Genomics Cellranger (v3.1.0) mkfastq software was used for FASTQ generation, quality control was performed using FastQC (v 0.11.8), and CellRanger Count (cellranger-3.1.0 for fetal single cell and cellranger-7.0.0 for adult single nucleus) was used for alignment, filtering, barcode counting, and UMI counting of the single cell/nuclei FASTQs. The mm10-3.0.0 and mm10-2020-A transcriptomes were used as references for the fetal single cell and adult single nucleus analysis respectively. For the snRNAseq data, the argument "include-introns true" was added to account for nuclear RNA. Seurat (v4.3.0.1)[21,22] was used for cluster analysis using R (v4.1.2) in RStudio software (2023.06.1 + 524). The E16.5 scRNAseq dataset was filtered to remove cells with <200 genes, >7500 genes or percent mitochondrial genes >10%, scaled with cell cycle regression, and clusters were found using 42 dims resolution = 0.8. Sub-clusters for *Pax8* + (PxPos object) clusters (17, 14, 18, 8) were found by subsetting the data based on *ident* and standard Seurat analysis of the PxPos dataset with 10 dims and resolution = 0.5. The snRNAseq data from adult estrus and diestrus samples were merged and integrated using the FindIntegrationAnchors and IntegrateData functions of Seurat, and the integrated dataset was filtered to remove cells with <200 genes, >7500 genes or percent mitochondrial genes >10% and clusters were found using 30 dims and resolution = 0.5.

Data Records

The sequencing data from this study have been uploaded to the National Center for Biotechnology Information (NCBI) Gene Expression Omnibus with accession ID GSE244849. This includes the raw.fastq.gz files for E16.5 (3 samples with 2 fastq.gz files each: read 1 and read 2) and adult bulk RNAseq data (3 samples with 2 fastq.gz files each: read 1 and read 2); raw.fastq.gz files for the E16.5 scRNAseq sample (1 sample with 3 fastq.gz files each: read 1, read 2, index) and raw.fastq.gz files for the adult snRNAseq data (2 samples with 3 fastq.gz files each: read 1, read 2, index)[11]. We also provide the output from Salmon transcript quantification as *quant.sf* files with transcript abundance values for each replicate of our fetal and adult RO bulk RNAseq. These can be found in GEO accession ID GSE244849 as a supplementary file for each replicate[11]. In addition, we provide the *filtered_feature_bc_matrix* output folders from CellRanger Count for the E16.5 scRNAseq and adult snRNAseq datasets. On the GEO accession page, processed *filtered_feature_bc_matrix* files for the single cell and single nucleus experiments can be found within the folder for each sample. The Salmon transcript quantification (quant.sf) for the RO bulk RNAseq datasets can be found within the supplemental file named GSE244849_RAW.tar[11]. Raw fastq.gz files for each sequencing run can be found under the sample section using the SRA for each sample or by clicking the link labeled 'SRA Run Selector'. The SRA accession numbers in Tables 1 and 2 can also be used to search the datasets directly on the SRA run browser. The quant.sf files for the reanalyzed E16.5 ovary, adult

	E16.5 scRNA	2 M snRNA Diestrus	2 M snRNA Estrus
CellRanger count	7,952	2,964	2,792
Mean Reads per Cell	125,706	164,427	172,023
Valid Barcodes	96.90%	95.20%	94.80%
Valid UMIs	99.90%	99.90%	99.90%
Reads Mapped to Genome	92.80%	91.20%	92.10%
Reads Mapped Confidently to Genome	88.50%	88.30%	89.20%
Reads Mapped Confidently to Intergenic Regions	4.50%	4.50%	4.40%
Reads Mapped Confidently to Intronic Regions	23.60%	60.70%	59.20%
Reads Mapped Confidently to Exonic Regions	60.40%	23.10%	25.70%
Reads Mapped Confidently to Transcriptome	56.20%	39.50%	41.20%
Reads Mapped Antisense to Gene	2.50%	44.10%	43.40%
SRA Accession Number	SRR26313456	SRR26313454	SRR26313455

Table 2. Cell Ranger metrics on fetal RO scRNAsequencing and adult RO snRNAsequencing runs.

ovary, adult OSE, and adult oviduct datasets are available as a single download (ReprodEpith_SalmonQuant.zip) on FigShare[23] at https://doi.org/10.6084/m9.figshare.25193786.

Technical Validation

Bulk RNAseq of the fetal and adult rete ovarii. The first dataset we produced and validated was bulk RNA sequencing of FAC-sorted GFP+ cells for three biological replicate samples of pooled ROs (4-5 per pool) from *Pax8-rtTA; TreH2bGFP* females at E16.5 and 2 months (Table 1). Paired-end 50 bp sequencing runs performed within the standard quality range, and we achieved 53–84 M read pairs for each sample (Table 1). We used FASTQC[24] and Rqc[25] for FASTQ quality control, and found that the Mean per Read sequence quality passed the standard quality threshold of 35 for all samples (Fig. 2a). We then used Salmon software (16) to map reads to the mouse genome (build GRCm39) and quantify transcript expression. Salmon mapping rate ranged from 68% to 87% across all samples (Table 1). The output from Salmon was opened into RStudio software for count normalization (Fig. 2b). These results showed that the normalized counts for all samples fell in the same range, except for the adult replicate 3, which had on average lower counts. Principal component analysis (PCA, Fig. 2c), and sample distance analysis (Fig. 2d) showed that Fetal RO replicate 3 was a major outlier in the experiment, and that Adult RO replicate 3 was somewhat of an outlier. We have removed these replicates from further analysis of the data.

Comparison between the bulk transcriptome of the RO and other reproductive epithelia. To confirm that our bulk RNA seq of the RO represented epithelial cell gene expression signatures related to the reproductive system, we compared our dataset with published bulk transcriptomes of other reproductive epithelia and the ovary. We accessed and re-analyzed published bulk RNAseq datasets of E16.5 ovary[18], adult ovary[19], adult ovarian surface epithelium and adult oviduct[20], as these were all expected to be the most closely related to the RO, and differential gene expression was likely to point us towards RO-specific genes. Each dataset had three replicates, which were analyzed alongside our RO datasets using Salmon quantification (Fig. 3). Log10-normalized counts were in the same range for all samples (except, as noted before, Adult RO replicate 3) (Fig. 3a). PCA and sample distance analysis showed that all replicates from each cell type clustered together (Fig. 3a,c). Interestingly, the fetal and adult RO (Fig. 3b, *teal* and *cherry*, respectively) clustered close to each other, but further away from the other reproductive epithelia (Fig. 3b). This could be due to a batch bias in the data generation, since they were sequenced separately from the other samples. On the other hand, the adult ovary and adult ovarian surface epithelium (OSE) clustered closely despite being generated in two different studies (Fig. 3b, *yellow* and *orange*, respectively). This was to be expected since the ovary dataset likely included OSE. In addition, these tissues are very closely related, and the OSE is a source of ovarian cells[26,27]. The fetal ovary and adult oviduct (Fig. 3b, *green* and *purple*, respectively) were the furthest from the other tissues (Fig. 3b,c), with the fetal ovary clustering almost equidistant to the fetal RO and to the adult ovary. Intriguingly, the adult RO samples appeared closer to the adult OSE than to the adult ovary. This re-analysis of published datasets presents the transcriptome of reproductive epithelia, all analyzed with the same parameters, and will be a helpful comparative resource for the field of reproductive epithelium development and homeostasis.

Single-cell RNAseq of the fetal RO. To provide more detailed insight into the transcriptome of the RO and surrounding cells, we next performed and validated single-cell RNA sequencing of the RO. We collected ROs from E16.5 *Pax8rtTA; TreH2BGFP* females, dissecting as close to the RO as possible, but intentionally leaving some surrounding tissue from the ovarian capsule and mesonephric tubules. This would allow one to identify markers of other poorly described components of the ovarian capsule, such as the mesovarium and mesonephric mesenchyme (Fig. 1). In addition, including this tissue meant one could use the dataset[11] to interrogate the data for potential interactions between cells of the RO and surrounding tissue. We produced single cell suspensions and performed cell capture and barcoding using the 10X Genomics Chromium. After library preparation and sequencing, we performed QC on the FASTQ read files, studied the CellRanger QC output, and found that the reads passed QC thresholds (Table 2). We imported the outputs from CellRanger counts into R as a Seurat object and began the standard Seurat single-cell analysis pipeline with the whole sample. Standard QC metrics for single

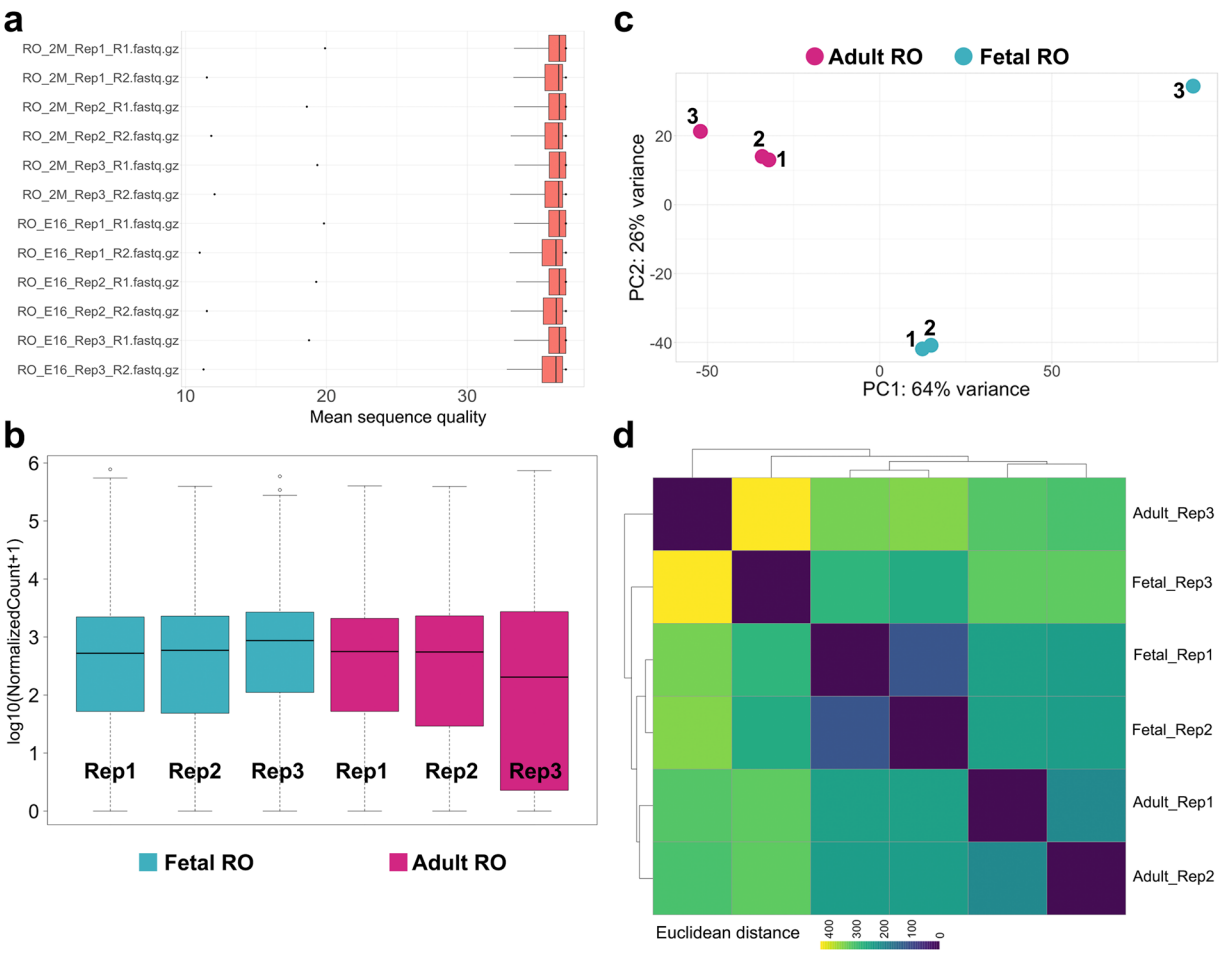

Fig. 2 Technical validation of RO Bulk RNAseq datasets. (**a**) Boxplot illustrating mean per-read quality of the sequences present in each fastq.gz file. (**b**) Boxplot illustrating normalized counts for each sample on the logarithmic scale. Adult RO replicates are depicted in *cherry*, fetal RO replicates in *teal*. Note all samples follow a similar distribution except for Adult_RO rep3, which appears to be an outlier. (**c**) Scatter plot illustrating Principal Component Analysis (PCA) of adult and fetal RO datasets. Each dot represents a single replicate for each timepoint (*cherry*, adult RO; *teal*, fetal RO). In each case, replicates 1 and 2 cluster more closely than replicate 3. In particular, fetal RO rep3 appears to be a true outlier. (**d**) Heatmap showing sample distance analysis for all adult RO and fetal RO replicates. Replicates 1 and 2 for each stage share more similarities than replicates 3, as illustrated by the deeper teal hues in the bottom right third of the heatmap, and the yellow hues at the top left. The color scale represents the Euclidean distance between samples. Rep1, replicate 1; Rep2, replicate 2; Rep3, replicate 3.

cell data include the number of genes per cell (nFeature_RNA), which allows the distinction between lysed, empty cells (nFeature_RNA < 200) and cell doublets (nFeature_RNA > 7500), and the percentage of mitochondrial genes, which acts as a proxy for quantifying cell death (optimal percent.mt < 20) (Fig. 4a,b). Figure 4a shows that a majority of cells in our sample fall within the optimal quality range (Fig. 4a, d*otted rectangle*). Figure 4b shows violin plots for each QC metric and illustrates that most of our cells are within the optimal range. Nonetheless, we opted to filter out low quality cells (cells with <200 genes, >7500 genes or percent mitochondrial genes >10%) for further analysis, as recommended in the Seurat user guide. We then performed Seurat Cluster analysis and identified 21 clusters. In a separate analysis using the Mm10 + Gfp reference genome for the analysis, we found that only 7.63% of the captured cells expressed the *GFP* transcript (Table 3). Analysis of *Pax8* expression revealed that we had captured very few *Pax8*+ cells, and only 3.4% of the cells expressed *Pax8* transcripts (Fig. 4c,Table 3). This could be in part due to the sparse nature of single-cell sequencing technology, which does not accurately reflect all transcripts expressed in all cells. In addition, trancripts coding for transcription factors such as PAX8 tend to be expressed at very low levels. Nonetheless, the *Pax8*+ cells mapped to distinct clusters. We used standard markers, prior knowledge, and the EnrichR online gene ontology database to identify putative cell types in each of the 21 clusters (Fig. 4d,e). We found that among these clusters, the *Pax8*+ cells mapped to IOR/Granulosa; CR; EOR and mesonephric tubules (Fig. 4e). To analyze these cells further, we created a new dataset by sub-clustering these *Pax8*+ clusters into a new Seurat object which we called PxPos. This subset was expected to contain only granulosa cells, cells of the RO, and cells of the mesonephric tubules. Standard Seurat Analysis on this subset revealed five independent *Pax8*+ clusters (Fig. 4f). We used prior knowledge and the EnrichR gene ontology database to

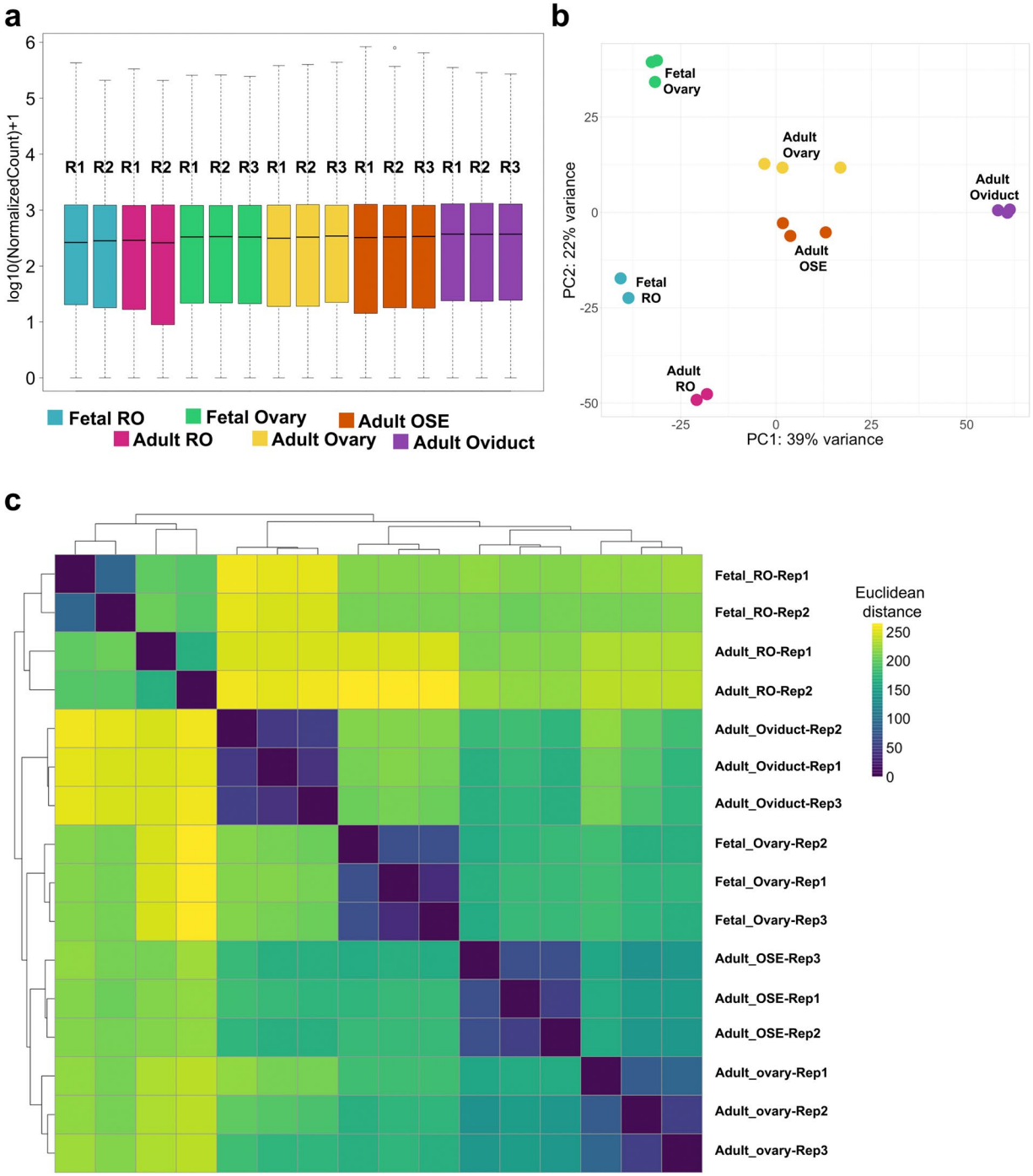

Fig. 3 Technical validation of reproductive epithelia Bulk RNAseq datasets. (**a**) Boxplot illustrating normalized counts for each sample on the logarithmic scale. Adult RO replicates are depicted in *cherry*, fetal RO replicates in *teal*, fetal ovary replicates in *green*, adult ovary replicates in *yellow*, adult ovarian surface epithelium (OSE) replicates in *orange*, and adult oviduct (ovi) samples in *purple*. Note removal of Fetal_RO_Rep3 and Adult_RO rep3, which appeared to be outliers. (**b**) Scatter plot illustrating Principal Component Analysis (PCA) of all reproductive epithelia datasets. Each dot represents a single replicate for each tissue and timepoint (*cherry*, adult RO; *teal*, fetal RO; *green*, fetal ovary; *yellow*, adult ovary; *orange*, adult OSE; *purple*, adult oviduct). In each case, all replicates for a tissue cluster together more closely than with any other tissue. (**c**) Heatmap showing sample distance analysis for all replicates of each tissue and timepoint. The color scale represents the Euclidean distance between samples. Rep1, replicate 1; Rep2, replicate 2; Rep3, replicate 3.

identify putative cell types for each cluster. *Foxl2* and *Nr5a1* are granulosa markers that are expressed in cells of the IOR (*Foxl2*+ and *Nr5a1*+)[2,9]. Cells of the CR express *Nr5a1*+ and GFRa1[2,28]. Thus, we labeled the *Foxl2*+/*Nr5a1*+/*Gfra1*- cluster IOR, and the *Foxl2*-/*Nr5a1*+/*Gfra1*+ cluster CR[28]. We had previously found that at E16.5,

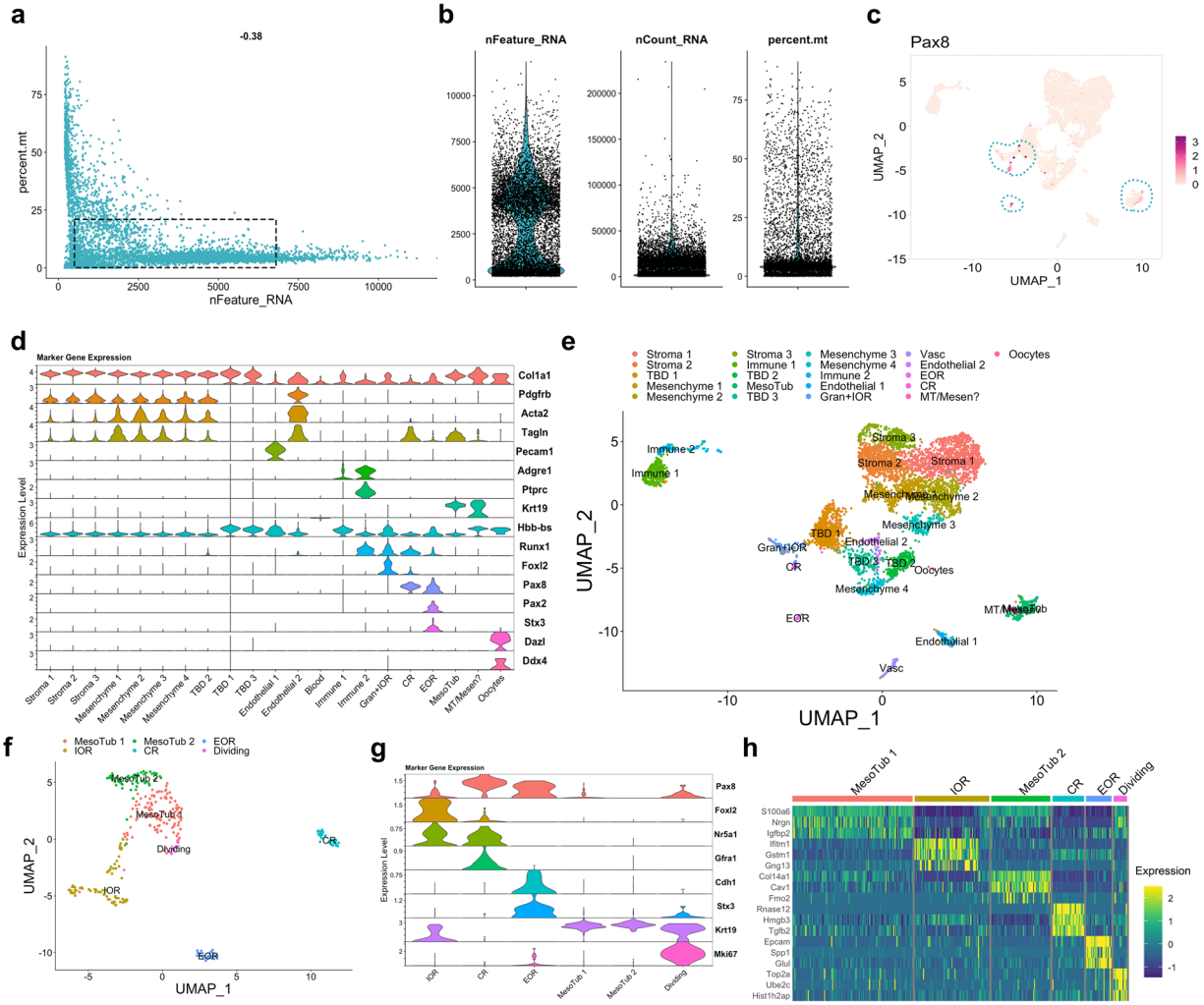

Fig. 4 Technical validation of the E16.5 scRNAseq dataset. (**a**) Scatter plot illustrating the distribution of the number of genes (nfeature_RNA) relative to the percentage of mitochondrial genes (percent. mito). Live cells of optimal quality have nFeature_RNA between 200 (empty cells) and 7500 (dublets), and have low expression of mitochondrial genes (used as a proxy for dying cells). The dotted rectangle represents the optimal quality threshold region. The −0.38 value at the top is the Pearson correlation between nFeature_RNA and percent. mt. (**b**) Violin plots illustrate the number of genes (nfeature_RNA), unique molecular identifier (UMI) (nCount_RNA), and the percentage of mitochondrial genes (percent.mt). (**c**) Seurat FeaturePlot showing expression of *Pax8* across the whole single-cell dataset. Darker dots represent higher *Pax8* expression levels. Dotted outlines highlight the clusters that express Pax8 (clusters # 8, 14, 17, 18). (**d**) Seurat Stacked Violin plot showing expression of marker genes in each cluster of the fetal single cell dataset. Putative cell types for each cluster were identified using known gene markers and Gene Ontology analysis. (**e**) Seurat UMAP plot showing cluster analysis of the whole single-cell sample at E16.5, using nDims = 42 and resolution = 0.8. 21 clusters were identified, and renamed based on putative cell type identification. Where cell type could not be inferred by prior knowledge or gene ontology databases, cluster was labeled "TBD – to be determined". (**f**) Seurat UMAP plot showing cluster analysis of the subset of *Pax8*+ clusters (# 8, 14, 17, 18), using nDims = 10 and resolution = 0.6. 6 clusters were found, and putative cell types for each cluster were identified using known gene markers and Gene Ontology analysis. (**g**) Seurat Stacked Violin plot showing expression of marker genes in each cluster of the *PxPos* subset dataset. (**h**) Heatmap illustrating expression differences for top 3 enriched genes from each cluster in the PxPos dataset. Dark blue hues depict low expression while green-yellow hues depict high expression.

	E16.5 scRNA	2 M snRNA Diestrus	2 M snRNA Estrus
Percent *Gfp*+	7.63%	8.03%	11.28%
Percent *Pax8*+	3.40%	16.37%	15.44%

Table 3. Distribution of *Gfp* and *Pax8* positive cells in the datasets.

E-Cadherin and STX3 are specifically expressed in the EOR[2,28], thus we called the *Cdh1*+/*Stx3*+ cluster EOR. Two clusters were enriched in *Krt19*, which is not found in the EOR at this stage but is present in the mesonephric tubules. *Stx3* is generally a marker of tubular epithelia, thus we determined that the *Stx3*+/*Krt19*+ cluster likely represents mesonephric tubules remaining in the ovarian capsule. Finally, the remaining cluster was enriched for *Mki67*, a marker of cell proliferation, as well as many cell cycles genes. We thus concluded that this cluster contained mostly dividing cell (Fig. 4g). Using the *FindAllMarkers* function of Seurat, we identified the top 3 enriched genes for each cluster in the PxPos dataset (Fig. 4h). Significant differential enrichment of specific genes in each cell cluster validated the quality of our clustering analysis in Seurat. Encouragingly, these genes were also found to be enriched in the bulk RNAseq dataset of the fetal RO, further validating the quality and consistency of our transcriptome datasets. These quality control steps demonstrate that this dataset is useful for interrogations of gene expression in the different regions of the RO. Spatial analysis of RNA expression using *in situ* hybridization will be required to validate the mapping of these clusters and regional markers. The code used to perform these analyses and generate the panels in Fig. 4 is available on our GitHub page (https://github.com/McKeyLab/RODatasets).

Single-nucleus RNAseq of the adult RO. We next produced and validated the single cell transcriptome of the adult RO. To optimize the representation of RO cells in the dataset, we chose to FAC-sort GFP+ cells from adult *Pax8rtTA; TreH2bGFP* females and mix GFP+ cells with GFP- cells in the final sample, to reach an optimal number of cells for 10X capture, and to allow for future analysis of the interaction between cells of the adult RO and surrounding tissue. Our first disaggregation trials to generate single cell suspensions of the adult RO revealed that the tubular EOR is strongly adherent and very difficult to dissociate without lysing the cells. We thus turned to single-nucleus RNA sequencing, which is recommended for difficult tissues. Finally, we chose to sequence the RO from mice at estrus and diestrus, to provide information on gene expression in the RO during different phases of the estrous cycle. Similar to the fetal scRNAseq, single nuclei were captured using the 10X Chromium, and cDNA libraries were prepared and sequenced. We began the analysis by performing QC on the FASTQ files and CellRanger outputs (Table 2), and both samples passed the QC thresholds. We then brought both datasets into Rstudio and merged them for analysis with Seurat. The QC metrics shown in the nFeature_RNA relative to percent.mt scatter plot (Fig. 5a) and violin plots (Fig. 5b) revealed that the majority of the cells in both samples were found within the optimal thresholds.

In this single-nucleus dataset, only RNAs found in the nucleus are represented. This explains the higher rate of intronic sequences and the higher rate of antisense reads identified in the single-nucleus dataset compared to the single-cell dataset of (Table 2). RNAs of mitochondrial genes are typically not present in the nucleus, thus the percent.mt was expected to be lower. In addition, because single-nucleus only captures nascent RNA in the nucleus, the number of RNAs (nCount_RNA) was also expected to be lower, as seen in the QC plots (Fig. 5a,b)[29]. The number of genes and reads sequenced per cell was lower in the snRNA-seq compared to the scRNA-seq approach. We used the Mm10 + *GFP* reference genome to determine whether our FACS approach had yielded a better enrichment in *GFP*+cells in this dataset. Surprisingly, we found that only 8.03% and 11.28% of the nuclei expressed the *GFP* transcript in the diestrus and estrus dataset, respectively (Table 3). FACS technology is based on active fluorescence and not transcription, thus it is possible that more nuclei were positive for the GFP protein but not actively expressing the transcript. It is also possible that the transcript was not being produced at high enough levels to capture at high rates with single nucleus sequencing and that single cell sequencing would have yielded GFP levels closer to 50% of the cells. In comparison, we did find a more significant enrichment in *Pax8*+ cells in these datasets compared to the fetal dataset, with 16.37 and 15.44% of the nuclei expressing *Pax8* transcripts at diestrus and estrus respectively. After QC, the estrus and diestrus datasets were integrated in Seurat, and the UMAP of the integrated dataset showed that there was significant overlap between the two stages, with only a few clustered cells found in one stage but not the other (Fig. 5c).

To validate that we had indeed captured the RO at estrus and diestrus, we investigated the expression of genes previously reported to change with the estrous cycle[30]. *Star* and *Rgcc* were shown to be upregulated during estrus in granulosa and theca cells, and this is indeed what we found in our dataset (Fig. 6a). During diestrus, *Lhcgr* and *Inhba* are upregulated[30], and this was also true in our dataset (Fig. 6b). Thus, we concluded that our dataset did capture expression in the RO and surrounding cells at estrus and diestrus. Seurat clustering yielded 19 clusters. We attempted to identify cell types for each cluster using our prior knowledge and the EnrichR gene ontology database[31–33]. The following markers were used to annotate the clusters: Epithelial markers (RO): *Krt8*; *Cdh1*, Epithelial progenitor marker (Hilum/OSE, RO): *Lgr5*; Ciliated epithelial marker: *Foxj1;* RO markers: *Pax8, Pax2, Gfra1, Rmst*; Granulosa cell markers: *Foxl2, Esr2;* Growing granulosa markers: *Amh, Nr5a2, Mki67;* Theca markers: *Nr5a1, Cyp17a1;* Immune cell markers: *Adgre; Ptprc;* Endothelial cell markers: *Pecam1, Emcn;* Adipose marker: *Adipoq;* Fibroblast / Myofibroblast markers: *Pdgfrb, Acta2, Col1a1;* Smooth muscle cell (SMC) markers: *Tagln, Cnn1* (Fig. 5d,e). Using the Seurat FindMarkers function, we identified the top 2 genes for each putative RO cluster (Fig. 5f). Significant differential enrichment of specific genes in each cell cluster validated the quality of our clustering analysis in Seurat. We are currently performing spatial validation for all the top genes, and hope this data will lead to exciting discoveries about the role and regulation of the RO.

Usage Notes

Potential limitations of the datasets. We used online cell type identification databases and prior knowledge to define labels for the clusters in the scRNAseq and snRNAseq datasets. However, it is important to note that the RO clusters have not yet been spatially validated, and thus labels may be inaccurate. In addition, the *Pax8rtTA; TreH2BGFP* is more highly expressed in the cells of the EOR and CR than in the IOR, so it is likely that we have a smaller representation of the IOR in our datasets. Nevertheless, the datasets described here provide a wealth of information on the transcriptome of all regions of the rete ovarii and surrounding tissue in fetal and

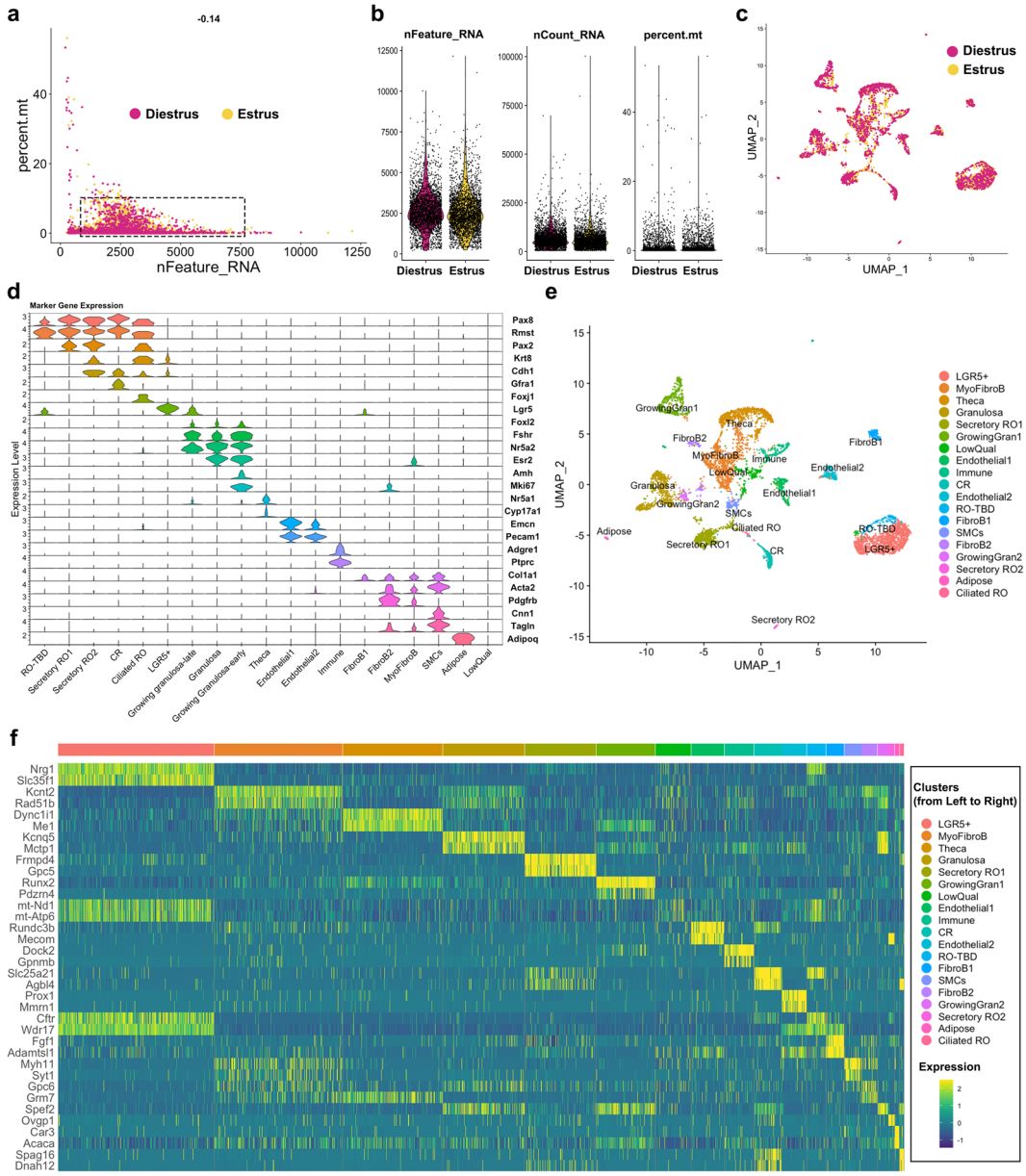

Fig. 5 Technical validation of the adult snRNAseq dataset. (**a,b**) Quality control of adult single-nucleus data at estrus (*yellow*) and diestrus (*cherry*). (**a**) Scatter plot illustrating the distribution of the number of genes (nfeature_RNA) relative to the percentage of mitochondrial genes (percent. mito). Live cells of optimal quality have nFeature_RNA between 200 (empty cells) and 7500 (dublets), and have low expression of mitochondrial genes (used as a proxy for dying cells). nFeature_RNA and percent.mt are lower here due to the sample being single nuclei. The dotted rectangle represents the optimal quality threshold region. The −0.14 value at the top is the Pearson correlation between nFeature_RNA and percent.mt. (**b**) Violin plots illustrate the number of genes (nfeature_RNA), unique molecular identifier (UMI) (nCount_RNA), and the percentage of mitochondrial genes (percent.mt). (**c**) Seurat UMAP plot showing analysis of the integrated single-nucleus dataset from Estrus (*yellow*) and Diestrus (*cherry*) adult samples, illustrating high overlap between Estrus and Diestrus samples. (**d**) Seurat Stacked Violin plot showing expression of marker genes in each cluster of the *integrated* dataset. Putative cell types for each cluster were identified using known gene markers and Gene Ontology analysis. (**e**) Seurat UMAP plot showing cluster analysis of the single-nucleus integrated dataset, using nDims = 30 and resolution = 0.5. 19 clusters were identified, numbered 0 to 19. (**f**) Heatmap illustrating expression for top 2 enriched genes from each cluster in the Integrated snRNAseq dataset. Dark blue hues depict low expression while green-yellow hues depict high expression. Cluster names are in the legend on the right side of the heatmap, in the order that they appear on the heatmap (top bar, from left to right).

adult mice. This work integrates well with recent efforts to better understand the origin and progenitor function of the intraovarian region of the RO in fetal mice and humans[9,12,13]. To the best of our knowledge, these are the first transcriptomic insights into the adult RO and the extraovarian regions of the RO. We are actively working

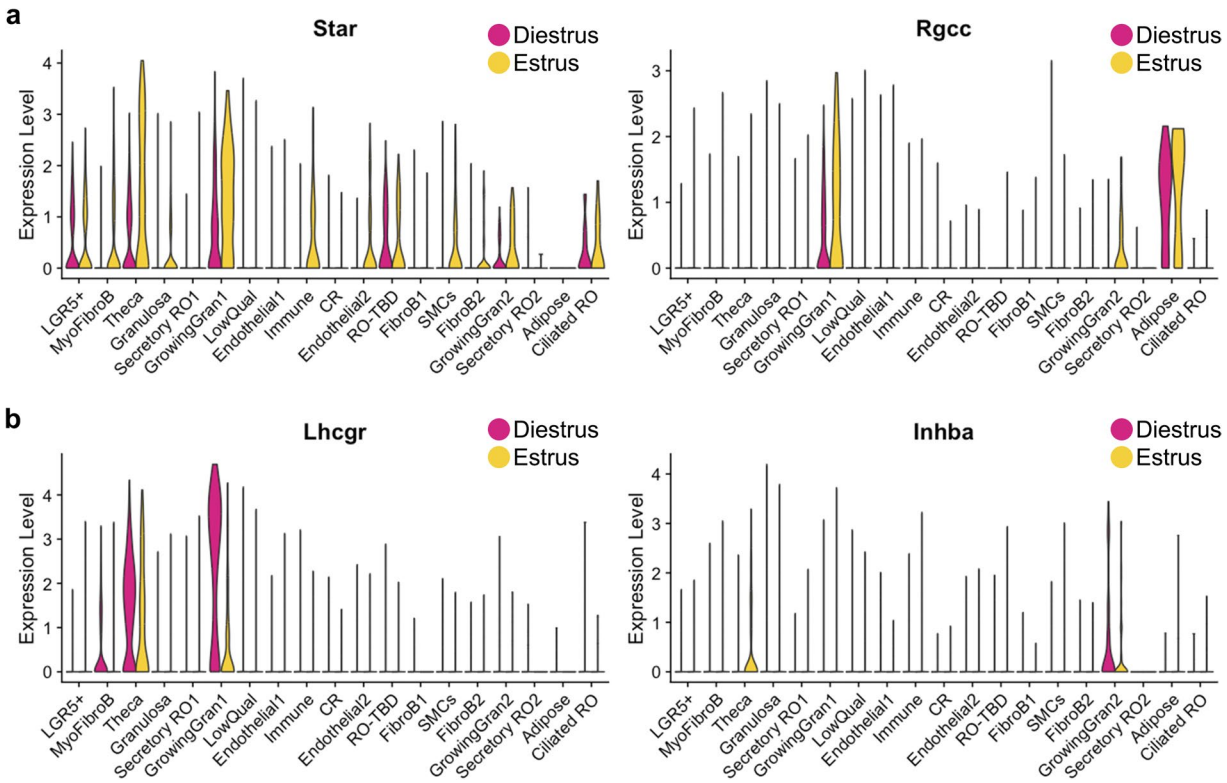

Fig. 6 Validation of estrous staging for snRNAseq dataset. (**a,b**) Seurat violin plots illustrating expression of known cycling genes in the integrated snRNA dataset at estrus (*yellow*) and diestrus (*cherry*). *Star* and *Rgcc* are enriched in granulosa and theca cells at estrus (**a**), while *Lhcgr* and *Inhba* are enriched in granulosa and theca cells at diestrus (**b**).

on spatial validation of the data presented here, with the goal of identifying candidate genes that may lead us to functions of the RO. Importantly, we hope these datasets become resources in the field of ovarian biology that may provide novel candidates for the investigation of idiopathic female subfertility.

Querying our datasets. All datasets can be queried for expression of genes of interest on our shiny app at https://cuanschutz-devbio.shinyapps.io/McKey_rete_ovarii_shiny/. Alternatively, in our GitHub repository, we provide an R markdown file (ROGeneTest.rmd) to query all the datasets for expression of a specific gene. The output from this file includes a bar plot of TPMs from the bulk RO and reproductive epithelia analyses, and feature plots and violin plots for the fetal scRNAseq and the adult snRNAseq, for the full datasets and the *Pax8*+ subclusters.

Code availability

All quality control analyses were performed using FastQC (http://www.bioinformatics.babraham.ac.uk/projects/fastqc/). Bulk RNAseq analysis was performed using Salmon (https://combine-lab.github.io/salmon/) and DESeq 2 (https://github.com/thelovelab/DESeq2). Single-cell and single-nucleus analyses were performed using Cell Ranger (downloaded from 10x genomics) and Seurat (https://satijalab.org/seurat/). All code generated for the analyses presented here is available on the McKey Lab GitHub at https://github.com/McKeyLab/RODatasets.

Received: 5 December 2023; Accepted: 4 April 2024;
Published online: 13 April 2024

References

1. Byskov, A. G. The Anatomy and Ultrastructure of the Rete System in the Fetal Mouse Ovary. *Biology of Reproduction* **19**, 720–735 (1978).
2. McKey, J. *et al.* Integration of mouse ovary morphogenesis with developmental dynamics of the oviduct, ovarian ligaments, and rete ovarii. *eLife* **11**, e81088 (2022).
3. Archbald, L. F., Schultz, R. H., Fahning, M. L., Kurtz, H. J. & Zemjanis, R. Rete ovarii in heifers: a preliminary study. *J Reprod Fertil* **26**, 413–414 (1971).
4. Wenzel, J. G. & Odend'hal, S. The mammalian rete ovarii: a literature review. *Cornell Vet* **75**, 411–425 (1985).
5. Wenzel, J. G., Odend'hal, S. & Player, E. C. Histological and histochemical characterization of the bovine rete ovarii through the estrous cycle and gestation. *Anat Histol Embryol* **16**, 124–135 (1987).
6. Lee, S.-H. *et al.* Ovarian cysts in MRL/MpJ mice are derived from the extraovarian rete: a developmental study. *J. Anat.* **219**, 743–755 (2011).

7. McKey, J., Cameron, L. A., Lewis, D., Batchvarov, I. S. & Capel, B. Combined iDISCO and CUBIC tissue clearing and lightsheet microscopy for in toto analysis of the adult mouse ovary†. *Biol. Reprod.* **102**, 1080–1089 (2020).
8. Bunce, C., McKey, J. & Capel, B. Concerted morphogenesis of genital ridges and nephric ducts in the mouse captured through whole-embryo imaging. *Development* **148**, (2021).
9. Mayère, C. *et al*. *Origin, Specification and Differentiation of a Rare Supporting-like Lineage in the Developing Mouse Gonad*. https://doi.org/10.1101/2021.09.15.460431 (2021).
10. Rosenfeld, C. S. *et al*. The Differential Fate of Mesonephric Tubular-Derived Efferent Ductules in Estrogen Receptor-α Knockout Versus Wild-Type Female Mice*. *Endocrinology* **141**, 3792–3798 (2000).
11. Anbarci, D. N., O'Rourke, R., Capel, B. & McKey, J. *GEO* https://identifiers.org/geo/GSE244849 (2023).
12. Taelman, J. *et al*. Characterization of the human fetal gonad and reproductive tract by single-cell transcriptomics. *Dev Cell* **59**, 529-544 (2024)
13. Lardenois, A. *et al*. Single-cell transcriptome landscape of developing fetal gonads defines somatic cell lineage specification in humans. 2023.08.07.552336 Preprint at https://doi.org/10.1101/2023.08.07.552336 (2023).
14. Tumbar, T. *et al*. Defining the epithelial stem cell niche in skin. *Science* **303**, 359–363 (2004).
15. Traykova-Brauch, M. *et al*. An efficient and versatile system for acute and chronic modulation of renal tubular function in transgenic mice. *Nat Med* **14**, 979–984 (2008).
16. McLean, A. C., Valenzuela, N., Fai, S. & Bennett, S. A. L. Performing Vaginal Lavage, Crystal Violet Staining, and Vaginal Cytological Evaluation for Mouse Estrous Cycle Staging Identification. *JoVE (Journal of Visualized Experiments)* e4389, https://doi.org/10.3791/4389 (2012).
17. Patro, R., Duggal, G., Love, M. I., Irizarry, R. A. & Kingsford, C. Salmon provides fast and bias-aware quantification of transcript expression. *Nat Methods* **14**, 417–419 (2017).
18. Wang, J. *et al*. Comprehensive Transcriptomic Analysis of Mouse Gonadal Development Involving Sexual Differentiation, Meiosis and Gametogenesis. *Biol Proced Online* **21**, 20 (2019).
19. Hohos, N. M., Cho, K. J., Swindle, D. C. & Skaznik-Wikiel, M. E. High-fat diet exposure, regardless of induction of obesity, is associated with altered expression of genes critical to normal ovulatory function. *Mol Cell Endocrinol* **470**, 199–207 (2018).
20. Zhang, S. *et al*. Both fallopian tube and ovarian surface epithelium are cells-of-origin for high-grade serous ovarian carcinoma. *Nature Communications* **10**, 1–16 (2019).
21. Stuart, T. *et al*. Comprehensive Integration of Single-Cell Data. *Cell* **177**, 1888–1902.e21 (2019).
22. Butler, A., Hoffman, P., Smibert, P., Papalexi, E. & Satija, R. Integrating single-cell transcriptomic data across different conditions, technologies, and species. *Nat Biotechnol* **36**, 411–420 (2018).
23. McKey, J. ReprodEpith_SalmonQuant.zip. *Figshare* https://doi.org/10.6084/m9.figshare.25193786.v1 (2024).
24. Babraham Bioinformatics - FastQC A Quality Control tool for High Throughput Sequence Data. https://www.bioinformatics.babraham.ac.uk/projects/fastqc/ (2023).
25. Souza, W., de, Carvalho, B., de, S. & Lopes-Cendes, I. Rqc: A Bioconductor Package for Quality Control of High-Throughput Sequencing Data. *Journal of Statistical Software* **87**, 1–14 (2018).
26. Mork, L. *et al*. Temporal differences in granulosa cell specification in the ovary reflect distinct follicle fates in mice. *Biol. Reprod.* **86**, 37 (2012).
27. Ng, A. *et al*. Lgr5 marks stem/progenitor cells in ovary and tubal epithelia. *Nature Cell Biology* **16**, 745–757 (2014).
28. Anbarci, D. N., McKey, J., Levic, D. S., Bagnat, M. & Capel, B. Rediscovering the Rete Ovarii: a secreting auxiliary structure to the ovary. 2023.11.08.566085 Preprint at https://doi.org/10.1101/2023.11.08.566085 (2023).
29. Kang, R. B. *et al*. Single-nucleus RNA sequencing of human pancreatic islets identifies novel gene sets and distinguishes β-cell subpopulations with dynamic transcriptome profiles. *Genome Medicine* **15**, 30 (2023).
30. Morris, M. E. *et al*. A single-cell atlas of the cycling murine ovary. *Elife* **11**, e77239 (2022).
31. Chen, E. Y. *et al*. Enrichr: interactive and collaborative HTML5 gene list enrichment analysis tool. *BMC Bioinformatics* **14**, 128 (2013).
32. Kuleshov, M. V. *et al*. Enrichr: a comprehensive gene set enrichment analysis web server 2016 update. *Nucleic Acids Res* **44**, W90–97 (2016).
33. Xie, Z. *et al*. Gene Set Knowledge Discovery with Enrichr. *Current Protocols* **1**, e90 (2021).

Acknowledgements

The authors would like to thank Dr. Nicolas Devos from the Duke Center for Genomic and Computational Biology Sequencing and Genomic Technologies core facility for sequencing resources and assistance; members of the Duke Flow Cytometry facility for helpful guidance and FAC-sorting; and the Duke Office of Information Technology and the Duke Compute Cluster teams for the assistance and computational power to perform the initial sequencing analyses reported here. We are grateful to all members of the Capel laboratory, the McKey laboratory and the Roberson laboratory for helpful discussions and suggestions on the work presented here. This project was supported by grants from the National Institutes of Health #1R01HD090050 and #5R01HD039963 to BC; and #K99HD103778 and #R00HD103778 to JM. DNA was funded by the Duke School of Medicine.

Author contributions

D.N.A., Y.X., D.T.P. and J.M. carried out the experiments; D.N.A., R.O., and J.M. carried out the data processing and data analysis, R.O. formatted and deposited the data in appropriate repositories and designed the shiny app to query the data; D.N.A., B.C. and J.M. designed the experiments; D.N.A., B.C. and J.M. drafted, revised and edited the manuscript; B.C. and J.M. acquired funding for the project.

Competing interests

The authors declare no competing interests.

Additional information

Correspondence and requests for materials should be addressed to J.M.