ARTICLE

Comparative single-cell analysis of biopsies clarifies pathogenic mechanisms in Klinefelter syndrome

Eisa Mahyari,[1,7] Jingtao Guo,[2,7,*] Ana C. Lima,[1] Daniel P. Lewinsohn,[1] Alexandra M. Stendahl,[1] Katinka A. Vigh-Conrad,[1] Xichen Nie,[2,3] Liina Nagirnaja,[1] Nicole B. Rockweiler,[4] Douglas T. Carrell,[2,5,6] James M. Hotaling,[2,6] Kenneth I. Aston,[2] and Donald F. Conrad[1,4,*]

Summary

Klinefelter syndrome (KS), also known as 47, XXY, is characterized by a distinct set of physiological abnormalities, commonly including infertility. The molecular basis for Klinefelter-related infertility is still unclear, largely because of the cellular complexity of the testis and the intricate endocrine and paracrine signaling that regulates spermatogenesis. Here, we demonstrate an analysis framework for dissecting human testis pathology that uses comparative analysis of single-cell RNA-sequencing data from the biopsies of 12 human donors. By comparing donors from a range of ages and forms of infertility, we generate gene expression signatures that characterize normal testicular function and distinguish clinically distinct forms of male infertility. Unexpectedly, we identified a subpopulation of Sertoli cells within multiple individuals with KS that lack transcription from the *XIST* locus, and the consequence of this is increased X-linked gene expression compared to all other KS cell populations. By systematic assessment of known cell signaling pathways, we identify 72 pathways potentially active in testis, dozens of which appear upregulated in KS. Altogether our data support a model of pathogenic changes in interstitial cells cascading from loss of X inactivation in pubertal Sertoli cells and nominate dosage-sensitive factors secreted by Sertoli cells that may contribute to the process. Our findings demonstrate the value of comparative patient analysis in mapping genetic mechanisms of disease and identify an epigenetic phenomenon in KS Sertoli cells that may prove important for understanding causes of infertility and sex chromosome evolution.

Introduction

Klinefelter syndrome (KS) (47, XXY) has an incidence rate of 1 in 650 male births.[1] There are several clinical manifestations of KS pathology, which include spermatogenic failure, endocrine dysregulation, and neuropsychological, cognitive, behavioral, and executive function impairment, as well as increased morbidity and mortality.[1] Although KS was first described in 1942, the precise molecular mechanisms that cause infertility in KS are still unclear.[2–4] KS is also the most commonly recognized genetic cause of male infertility. Male infertility is often characterized by the failure of producing viable sperm, and although there are several clinically recognized causes, the majority of cases remain idiopathic.[5] Some of the features of KS pathology are shared by other forms of male infertility and gonadal disruption, such as idiopathic non-obstructive azoospermia (NOA) (e.g., MIM: 258150), cryptorchidism (MIM: 219050), and hypogonadotropic hypogonadism (e.g., MIM: 308700). These features include elevated levels of follicle-stimulating hormone (FSH) and luteinizing hormone (LH), degeneration of germ cells, Leydig cell hyperplasia, and thickening of the walls of the seminiferous tubules. The fact that these changes are shared among different disease conditions with very distinct etiology indicates that these are most likely secondary consequences and not indicative of the primary causes. One feature fairly specific to KS is fibrosis and hyalinization of the seminiferous tubules. While these major histological changes are rarely observed in older fertile adults, in KS-affected individuals, the onset is at puberty and can quickly involve the entire testis.[3,6] During this process, Sertoli cells (SCs) and germ cells are completely lost, replaced by extensive deposition of collagen and other extracellular matrix (ECM) components. In the late 1960s and early 1970s, Frøland and Skakkebaek identified two distinct classes of tubules in histology from KS donors, which they classified as type A and type B.[7–9] These two types differ in the appearance and organization of SCs: type A tubules are characterized by "differentiated" SCs with normal organization and type B tubules contain what are described as "immature" SCs, oftentimes strongly disorganized. In advanced stages of KS histopathology, the tubular walls completely degenerate. The extent of tubule degeneration observed in adult KS-affected individuals is variable, and in some cases, relatively normal tubules supporting

[1]Division of Genetics, Oregon National Primate Research Center, Oregon Health and Science University, Portland, OR 97006, USA; [2]Andrology and IVF Laboratory, Division of Urology, Department of Surgery, University of Utah School of Medicine, Salt Lake City, UT 84108, USA; [3]Department of Oncological Sciences, University of Utah School of Medicine, Salt Lake City, UT 84112, USA; [4]Department of Genetics, Washington University School of Medicine, St. Louis, MO 63110, USA; [5]Department of Human Genetics, University of Utah School of Medicine, Salt Lake City, UT 84112, USA; [6]Department of Obstetrics and Gynecology, University of Utah School of Medicine, Salt Lake City, UT 63110, USA
[7]These authors contributed equally
*Correspondence: jingtao.guo@hci.utah.edu (J.G.), conradon@ohsu.edu (D.F.C.)
https://doi.org/10.1016/j.ajhg.2021.09.001.

spermatogenesis can be observed mixed among degenerating ones. The cell type(s) responsible for initiating this process are unclear, as are the molecular mediators.

The majority of genetic studies of testis[10–12] and infertility[13–15] have focused on bulk transcriptomics, which mask the complexity of the various cell types that exist in the testis. Recent studies of KS have attempted to address this with cellularity-matched control samples,[14,16] but standard analysis methods are still sensitive to quantitative, albeit smaller, differences in cell-type proportions among samples with this approach, and bulk RNA sequencing (RNA-seq) obscures contributions from rare cell populations. These studies identified hundreds of genes with expression changes in KS testis; however, there was limited overlap among identified gene sets, making it difficult to interpret the molecular pathology of KS. Although it is thought the extra X chromosome of KS cells must be the root cause, analyses to date have indicated that all KS cells experience X chromosome inactivation (XCI). In theory, XCI should erase the dosage imbalance of X-linked genes. Thus, many mechanistic theories have focused on either the role of X-linked genes that escape epigenetic silencing or the more general problems that may occur in cells that have whole-chromosome aneuploidy (e.g., as a result of cellular responses to aneuploidy). Emerging single-cell RNA-seq (scRNA-seq) methods have enabled the quantitative characterization of cell populations within the testis and can clearly contribute new insights into the pathologic mechanisms of KS. Pioneering scRNA-seq studies of KS have emerged in the past year, hinting at a special role for the Sertoli cell in KS pathology,[17–19] but insights have been limited by the small number of donors and suitable analysis tools. An scRNA-seq framework is needed to delineate the role of the X chromosome in KS and separate expression changes that are general responses to spermatogenic impairment versus causes of KS-specific pathology.

Here, we describe an approach based on joint analysis of multiple scRNA-seq datasets and demonstrate the potential of this approach to improve identification and interpretation of individual-specific defects in KS and idiopathic infertility. To dissect the pathological biology in KS, we combined and analyzed 26,300 single-cell transcriptomes derived from the testis of 12 carefully selected donors, chosen to allow us to isolate expression changes that are attributes of the specific defects of KS versus changes that are general hallmarks of defects in spermatogenesis and endocrine dysfunction. Our primary analysis framework is called sparse decomposition of arrays (SDA), a matrix factorization method that we find very useful for summarizing large, multi-donor scRNA-seq data into manageable sets of co-regulated genes, known as SDA components. We have recently shown the value of SDA in interpreting testis pathology in mice,[20] and herein we expand on its application by showing utility in humans. First, we describe the landscape of gene expression in our human testis atlas through the lens of SDA components and then identify KS-specific components that provide insight into KS-specific pathology. Our results highlight specific cells, pathways, and even individual genes that are likely to contribute to KS pathology. Finally, we describe and validate the surprising discovery that X chromosomes in SCs of KS donors lose XIST expression during postpubertal development, apparently leading to "X-reactivation" in these cells. We describe how our findings could fit together to explain a cascade of cause and effect that leads to some of the distinct features of KS that have eluded explanation until now. Along with this publication, we provide a web-based interface to our dataset—the Human Infertility Single-cell Transcription Atlas (HISTA)—with many features for exploring these data and results (web resources).

Material and methods

scRNA-seq experiments

Human testis biopsies were obtained from five affected individuals, with written consent, through the University of Utah Andrology laboratory (IRB approved protocol #00075836). Samples were obtained by testicular biopsy of 1–2 mm of tissue. The biopsy samples include one idiopathic azoospermic individual (INF1), two Klinefelter syndrome-affected individuals (KS1 and KS2), and one individual with secondary infertility treated as a control (CNT-U4). Three adult control samples (CNT-U1, CNT-U2, CNT-U3) were obtained from three cadaveric donors provided by Intermountain Donor Service, from which we took samples of ~2 cm × 2 cm in size. Each sample was prepared and sequenced as a separate library (10× Genomics single-cell 3′ gene expression library prep v.2). About 300 M total reads were sequenced from each library with either an Illumina Nova-seq 150 × 150 pair end (KS2) or Illumina Hi-seq 2500 at 125 × 125 pair end (all other samples). Raw sequencing data were processed with the CellRanger software package with default settings prior to further analysis.

Published data

We obtained published data for three additional adult control individuals (CNT-H1, CNT-H2, CNT-H3) from GEO (GEO: GSE109037), which were prepared on the 10× Genomics Chromium platform, processed with CellRanger, and aligned to the Human reference genome NCBI build 38 (GRCh38). Two juvenile samples (JUV1 and JUV2; ~1,500 cells total) were downloaded from GEO (GEO: GSE120506), and they were also sequenced by 10× Genomics and processed with CellRanger (v.1.2.1) with the STAR aligner to the GRCh38 reference. These juvenile samples, described previously,[12] were obtained from 13-month-old deceased donors. Published digital gene expression matrices (DGEs) for additional KS donors, generated by 10× library sequencing, were downloaded from GEO (GEO: GSE130151 and GSE149512); for comparison equivalency in validating sections of our results, five control samples and three infertile azoospermic samples were acquired in addition to the three KS donors of the GSE149512 dataset.

Data integration and pre-processing

Unique molecular identifier (UMI) count matrices were merged with Seurat as a starting point for all other analyses and storage

of all metadata and processed data.[21] For "conventional" analyses with the Seurat package, our quality control (QC) process was to remove cells with more than 150,000 total UMI count, less than 200 total UMI count, or more than half of their reads mapping to mitochondrial genes; starting from 26,751 cells, we retained 26,300. For SDA analyses of our QC process, we used DropSim (data and code availability), removing cells with fewer than 1,000 genes, reducing the total cells that enter SDA (and related downstream analyses) to 26,093.

Sparse decomposition of arrays (SDA) analysis

We performed SDA runs over a range of component numbers (20, 50, 100, 150, 200) and found, for this data, that 150 components optimally captures both normal and pathological signals of the testis scRNA-seq amalgam dataset at hand. We ran SDA with 26,300 cells until convergence (i.e., plateau observed in free energy, starting at 0 and increasing per iteration), and maximum iterations was set to 10,000. By comparing cell-scoring patterns across available replicates, in addition to functional assessment of the gene loadings, components were manually labeled as "batch" or "biological." As a QC filter, we removed components that only identified a single cell (singleton) or components with an outlier (>95th percentile) as the "max-score." Note that SDA components have arbitrary signs and must be interpreted through the combination of gene and cell signs. Gene loadings and cell scores with concordant signs result in a positive expression contribution from a component, whereas discordant signs result in negative contribution. Finally, we constructed a new gene by cell matrix, DGE_{SDA}, by taking the dot-product of the gene loading and cell score matrices after removing the unwanted components. To perform differential expression (DE) analysis on this new DGE_{SDA}, we used Seurat as intended for standard pipeline analysis. SDA imputation can result in small but negative values, which were adjusted to 0 prior to analysis.

SDA component "fingerprint" heatmap

We used a reduced subset of the SDA cell score matrix, removing the noise-associated components, to enumerate the number of cells that score positive or negative per component. These cell counts were tabled against the experimental condition and then normalized for the number of cells contributed by each condition. Next, we ran a chi-square test on the cell-contribution-normalized count table to obtain Pearson's residuals. The residuals represent both magnitude and direction of enrichment or depletion, and we transformed them by the inverse-sine transform (arcsinh) to balance the range of the residuals prior to plotting them as a heatmap; the components on the x axis can be ordered by hierarchical clustering or binned by the cell type that the components map to. One utility of this heatmap is to provide a barcode-like identifier for each condition, showing similarity between the affected individuals and control individuals. Furthermore, it can be used to navigate the SDA components to identify co-regulated gene signatures of both pathological and normal origins. Because of the "soft-clustering" nature of SDA, these enrichments or deficits not only highlight particular cell types that can be achieved by clinical phenotyping but also enumerate the cells at the scope of gene regulation.

Differential expression (DE) analysis

To broadly identify canonical cell types of the testis, we performed DE analysis for cell type identification on the normalized expression data. This was done with the Seurat library of functions;[21] specifically, we utilized the FindAllMarkers function with parameters suited for picking marker genes (min.pct > 0.6, logFC.thresh > 1, min.diff.pct > 0.3), combining runs from two separate tests (Wilcoxon and MAST). For all other DE analyses, we utilized the DGE_{SDA} matrix with the same Seurat set of tools and methods. For each cell type DE analysis, we first filtered for those cells and reprocessed the new Seurat object to obtain new cell-type-specific analyses, including top variable gene identification and dimensionality reduction (principal-component analysis [PCA], t-distributed stochastic neighbor embedding [tSNE], UMAP). This enabled us to identify DE genes and discriminate the subclusters of each cell type. In analyses where Seurat's integration was used, such as during validation, formal DE was performed on the RNA assay as recommended by the original authors.

Generation and analysis of integrated cell set

In order to validate and replicate signals derived from the original SDA analysis, we created a larger, "integrated" dataset consisting of additional samples from two studies published during the course of the project.[18,19] First, we annotated the cell types in these new datasets by performing label transfer with the Seurat package, using our original amalgam dataset as the reference. Next, we performed a manual QC to optimally capture the cell types of interest, such as Leydig cells (LCs) and SCs. Each sample underwent QC, normalization, and variable feature identification independently. Cells with fewer than 700 or more than 5,000 genes read were removed. Additionally, cells with more than 20,000 reads were removed. We then performed one integration for all of the samples by using the Seurat functions FindIntegrationAnchors and IntegrateData with the canonical correlation analysis (CCA) reduction. The integrated data then underwent the usual Seurat pipeline and other analyses described in the methods. We performed pseudotime analysis on this newly integrated dataset by using Monocle.[22] We assigned each post-QC single-cell library a score reflecting the expression of genes correlated with cell-cycle stage by using the Seurat function CellCycleScoring. CellCycleScoring requires a list of genes with known expression values across the cell cycle, which were obtained from published results.[23] For the LC analyses, we removed CNT-H2 because of low LC count, as well as the two JUV samples.

Analysis of expressed genetic variation on the X chromosome

In order to characterize X chromosome copy number in the scRNA-seq data, we screened X-linked transcripts for evidence of heterozygous genotypes. This analysis required read-level data from the original BAM files for various samples; in addition to the files generated by this study, Hermann et al. was available publicly (GEO: GSE109037), and the data from Laurentino et al. was provided directly by the authors. We used Samtools to sort and index the X-chromosome-subsetted BAM files.[24] We used Subset-bam to create new BAM files consisting of all reads coming from either LCs or SCs. Potential PCR duplicates were removed via samtool's rmdup function. Next, we used bcftools mpileup to identify the variant loci.[25] To clean up the variant calls, we retained variants in exonic regions of the X chromosome (hg38-gencode34 gene models) and excluded variants in pseudogenes, segmental duplications, pseudo-autosomal region 1 (PAR1) (GRCh38:1:278147), pseudo-autosomal region 2 (PAR2) (155701383:156030895), and X-transposed regions (88981596:93141104). The four possible

genotype calls (alt/ref, alt/alt, ref/ref, and no call) were then tabulated in the downstream analyses contrasting genotype frequencies among donors, conditions, and cell types.

Cell-cell communication analysis

Inference of cell-cell signaling activity was performed with CellChat, an R package for inference of cell signaling via scRNA-seq data.[26] CellChat uses a database of known receptor-ligand interactions to first select cell signaling genes with differential expression among cell groups in the input data. The resulting gene list is then used for identification of pairs of cell types that show expression patterns suggestive of intercellular communication, and the significance of the expression pattern is assessed by permutation test. The probability of intercellular communication is described in this study as the "amount of signaling activity" between two cell types. Analysis was performed according to default settings, except in the case of computeCommunProb, where the population.size option was set to "FALSE."

Histology: Hematoxylin and eosin (H&E), fluorescence *in situ* hybridization (FISH), and immunofluorescence (IF)

All histology images were acquired in an Olympus VS120 bright field and fluorescent slide scanner equipped with DAPI, EGFP, Cy3, Cy5, and Cy7 filters with a 40× (numerical aperature 0.95; 0.17 μm/pixel) objective or in a Leica SP5 AOBS multi-spectral confocal microscope equipped with five excitation lasers (405, 458, 476, 488, 514, 561, 594, and 633 nm) with a 40× (HCX PL APO lambda blue/40.0 × 1.25 Oil/UV) objective.

H&E

For H&E staining, deparaffinized and rehydrated 5 mm sections were incubated in hematoxylin for 3 min and rinsed with running tap water for 5 min. Afterward, the sections were dipped in acid alcohol (0.5% v/v hydrochloric acid in 70% ethanol), washed with distilled H_2O, and incubated in eosin (Poly Scientific, cat# 176) for 30 s. The sections were dehydrated before mounting with Histomount (National Diagnostics, cat# 12954910).

FISH

Paraffin blocks of donor tissue from normal human endometrium and testis, as well as from testis biopsies of individuals with XXY karyotype, were obtained from the Oregon Health and Science University Biolibrary. Japanese macaque testis was collected at the Oregon National Primate Research Center from a scheduled necropsy. Tissue was fixed in 4% paraformaldehyde overnight, washed in 1× PBS, dehydrated in a series of ethanol, and embedded in paraffin. Unless otherwise stated, we used two adjacent 5 μm paraffin sections per slide (with and without probe incubation). All assays were performed using the RNAscope Multiplex Fluorescent Reagent Kit v.2 and ancillary products, with 1 h of slide baking time and 15 min of both protease plus treatment and target antigen retrieval. Details about reagents and concentrations used can be found in Table S1.

For the detection of *XIST* (MIM: 314670), three adjacent tissue sections were used per slide: (1) for FISH, (2) for IF, and (3) for negative control probe and no primary control. We used endometrium tissue, which has XX karyotype and *XIST*+ cells, as a positive control to validate the use of the mRNA probe. FISH was performed following the RNAscope Multiplex Fluorescent Reagent Kit v.2 Assay protocol (document 323100-USM) and probe was visualized with Opal 650. Instead of counterstaining with DAPI, slides were rinsed in deionized water (diH_2O) for immunofluorescence. In brief, slides were incubated in blocking solution (10% normal donkey serum, 1% BSA, 1× PBS) for 15 min, in primary antibody for 1 h, in secondary antibody for 20 min, and in 1 μg/mL Hoechst 33342 for 5 min. These incubations were performed at 37°C in a humid chamber in the HybEZ oven (HybEZ II Hybridization System; Advanced Cell Diagnostics, cat# 321711). To stain the boundaries of the seminiferous tubules, we used an antibody against Actin Alpha 2—ACTA2—and detected it with a secondary conjugated with Alexa Fluor 594. Antibodies were diluted in PBST (0.1% BSA, 0.05% Tween 20, 1× PBS).

Leydig cell mRNA markers *INSL3* (MIM: 146738), *DLK1* (MIM: 176290), and *SHROOM2* (MIM: 300103) were detected in multiplex following the RNAscope Multiplex Fluorescent Reagent Kit v.2 Assay protocol (document 323100-USM) with Opal 520, Opal 570, and Opal 650, respectively, for probe visualization. *PPIB* (C3 of 3-plex positive control probe) and *DapB* probes were used in a control slide as positive and negative control, respectively, and visualized with Opal 620.

MKI67 (MIM: 176741) and *INSL3* mRNAs were detected in multiplex following the RNAscope Multiplex Fluorescent v.2 Assay combined with immunofluorescence protocol (document MK 51-150) with Opal 570 and Opal 650, respectively, for probe visualization. *PPIB* (C3 of 3-plex positive control probe) and *DapB* probes were used in a control slide as positive and negative control, respectively, and visualized with Opal 620. Additionally, the no probe adjacent section on the test slide was used as a negative control.

IF

To confirm Sertoli cell identity of the cells inside the tubules with different *XIST* expression, we stained sections of the same sample for SOX9 (SC marker in testis) by immunofluorescence. Tissue was deparaffinized in xylene and rehydrated through a series of decreasing concentrations of ethanol. Tissue was permeabilized in 0.1% Triton X-100 (in 1× PBS), heat-induced antigen retrieval was performed in Universal Antigen Retrieval Reagent, and the sample was incubated in blocking solution (10% normal donkey serum, 1% BSA, 1× PBS). Primary antibodies were incubated overnight at 4°C, secondary antibodies were incubated for 1 h at room temperature, and tissue was counterstained with 1 μg/mL Hoechst 33342.

Image analysis of *XIST*

Images were processed via Fiji (ImageJ 1.53c).[27] Regions of interest (ROIs) for tubule boundaries were manually defined with the ACTA2 image, combined, and aligned to the adjacent section with the *XIST* probe. The background was removed from the images (rolling ball: 50 px) of *XIST* and nuclei channels, which were then processed in individual ImageJ scripts. In brief, we pre-processed the nuclei image with contrast-limited adaptive histogram equalization (CLAHE) and median filtering to improve quality and remove noise, following the recommendations of Win et al.[28] Nuclei within each tubule ROI were then segmented with Otsu's threshold[29] followed by watershed and particle detection. On the *XIST* channel, we used a difference of Gaussians filter to improve local contrast and detection of small punctate signal and performed Otsu's threshold followed by watershed and particle detection. Tables were obtained, containing an entry for each particle detected, which summarized the area, XY centroid, and tubule identification number.

Data tables generated by the ImageJ scripts were analyzed in the R programming language. To compare *XIST* expression in two

different classes of tubules (thin and thick), we counted nuclei and RNA spots per tubule and plotted its ratio. We performed analysis of variance (ANOVA) to evaluate the statistical significance of differences in *XIST* expression resulting from tubule type.

Results

For this study, we generated testis scRNA-seq data from two individuals with KS (KS1 and KS2) and two individuals with other forms of infertility (INF1 and INF2) and combined these with published data from adult control individuals (CNTs) and two juveniles (JUVs) with normal spermatogenesis (Figure 1A; Table S2; material and methods). Subject INF1 was diagnosed with idiopathic NOA, whereas subject INF2 was diagnosed with retrograde oligozoospermia with secondary infertility because of ejaculatory dysfunction (INF2). All individuals were subject to a full andrological workup, and we were also able to perform testis histology on CNTU1–U3 and KS1 (Table S2; Figure S1; material and methods). LH and FSH were highly elevated for KS1, KS2, and INF1, whereas hormone levels were in the normal range for INF2. INF1 and INF2 were previously exome sequenced as part of GEMINI, a large cohort study to identify novel genetic causes of male infertility, and no convincing diagnostic variants were identified in this donor (L.N., K.I.A., and D.F.C., unpublished data). For the purposes of this study, we considered INF2 as a control individual for normal testis function and combined this sample with other control samples. Importantly, five of the individuals (CNTH1–H3, JUV1–2) were sequenced as part of two prior projects, some by a different laboratory.[11,30] Strong batch effects were apparent when we simply combined these published data with our scRNA-seq (Figure 1B). We constructed an analysis pipeline to identify and remove these batch effects prior to downstream analysis (Figure 1B, Figure S2, Figure S3; material and methods). We also used a more traditional approach of hard clustering and DE analysis to provide additional annotation of the data (Figure 1C).

SDA enables identification and interpretation of individual-specific pathology

In brief, SDA[31] identifies co-regulated genes in the context of specific cells, not limited by cell type boundaries, represented as the gene-loading and the cell-score matrices, respectively (see additional details in Figure 2A; material and methods). SDA is a model-based method used for decomposition of a matrix of cell-by-gene expression data into components. Each component is defined by two vectors: one quantifies which genes are active in that component (the "gene loadings") and the other quantifies the relative activity of the component in each cell (the "cell scores"). We found some accessory information to be essential for component interpretation, summarized in cartoon form in Figure 2B. Using SDA, we factorized the combined 12-donor scRNA-seq expression matrix into 150 components (supplemental note, Table S3). We provide the full SDA dataset and the accessory data for component inspection in our interactive web tool HISTA, available online (web resources).

Of the retained 75 SDA components, 37 of the cell scores were enriched for germ cells, 33 were enriched for somatic cells, and five were enriched for mixed cell-type components (Figure 2C). Judging from the distribution of cell scores across donors, the majority of the components summarize variation in normal gene expression programs of healthy cells (Figure 2D). It is of interest to identify components with individual- or condition-specific cell scores, as these components may provide information about causes and consequences of individual-specific pathology. Because SDA components are defined at the level of single cells, the use of SDA is robust to differences in cellularity between case and control testes, a problem that has historically plagued analyses of bulk testis tissue. In our data, 20% of SDA components were primarily mapping to KS cells, indicating numerous sets of genes with coordinated expression specific to KS cells. A much smaller proportion of components were found specific to INF1 with NOA (3%), and 4% of the components seem to be broadly shared among all infertile donors (Figure 2D).

To provide a higher resolution overview of SDA cell scores across donors and conditions, we devised a simple statistic to summarize components with unusual distributions across affected individuals called the cell score residual (material and methods). The cell score residual indicates donors that have an excess or deficit of cells loading on each component, compared to random expectation, and can be used to identify patterns related to pathology (Figure 2E). The pattern of residuals across INF2 is similar to adult control individuals (CNTs), consistent with the fact that INF2 has obstructive azoospermia and relatively unperturbed testicular function, providing further justification to consider this donor as a control for normal spermatogenesis. INF1, KS1, and KS2 are very similar, consistent with severe spermatogenic impairment in these individuals. The JUV donors provide further context by highlighting which components in KS and INF1 donors may correspond to development defects in somatic cells.

Three subtypes of Leydig cells coexist in adults

Leydig cells had eight spatially separate subpopulations according to tSNE analysis (Figures 3A and 3B). Three of these were specific to KS donors, one specific to JUVs, and one specific to INF1. One common model of human Leydig cell development defines four distinct types of cells in the ontogeny of Leydig cells: stem cells, progenitor LCs (PLCs), immature LCs (ILCs), and mature LCs (MLCs).[32] PLCs, ILCs, and MLCs can be readily identified and distinguished by morphology and marker gene expression. We hypothesized that some of the distinct LC clusters corresponded to these different developmental stages. By

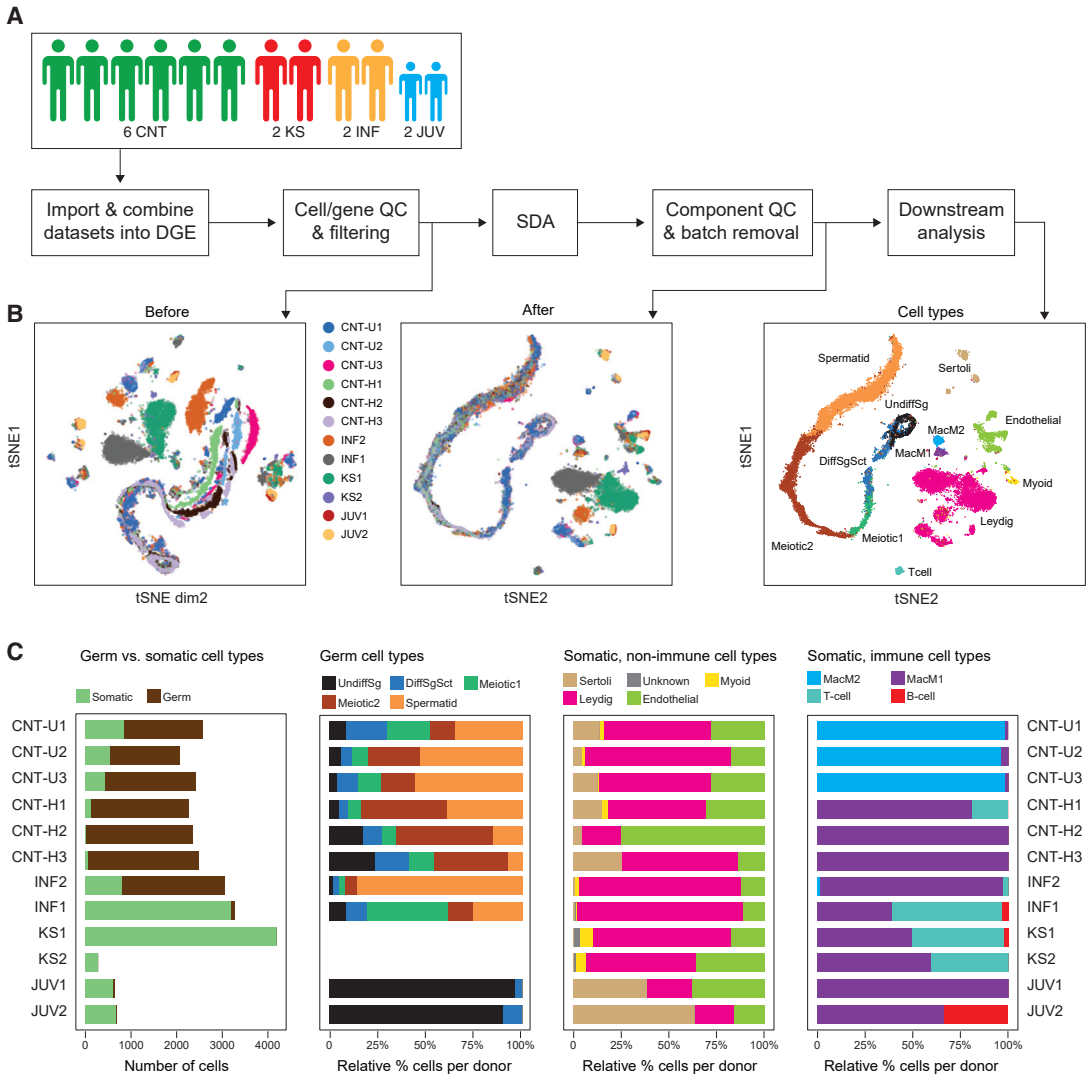


Figure 1. Construction of a human testis single-cell expression atlas

(A) Pre-processing pipeline: starting with the raw counts, an amalgam digital gene expression (DGE) matrix is produced. Cell-wise and gene-wise quality control (QC) is performed for removal of aberrant cells and undetected genes (material and methods). DropSim normalization is performed for correction of transcript-level differences across the cells as well as scaling of the genes. The normalized DGE matrix is then passed to SDA. The components are then assessed for divergent scoring across replicates (batch) as well as QC filters. The batch-removed DGE matrix is then used for downstream analysis.

(B) tSNE plots of the SDA cell score matrices before and after batch removal. Colors identify the donors ("before" and "after") or cell types ("cell types").

(C) The frequency of cell types identified across the donors. Abbreviations: CNT prefixes represent control donors; UndiffSg, undifferentiated spermatogonia; DiffSgSct, differentiating spermatogonia and pre-leptotene spermatocytes; meiotic 1, first spermatocyte cluster; meiotic 2, second spermatocyte cluster; spermatid, round and elongating spermatids; MacM1, M1 macrophages; MacM2, M2 macrophages.

inspecting the expression of known marker genes, we could clearly assign populations 3, 2, and 7 as PLCs, ILCs, and MLCs, respectively (Figures 3C and 3D). PLCs, which express the steroidogenic genes *CYP11A1* (MIM: 118485) and *HSD3B2* (MIM: 613890), but lack *HSD17B6* (MIM: 606623) expression, formed one distinct cluster with contributions from control, INF1, and KS donors. MLCs were clearly recognizable as a single subpopulation with robust expression of *LHCGR* (MIM: 152790), numerous elements of the testosterone production pathway, and the marker *INSL3*. Essentially all control cells were assigned to clusters 3, 2, and 7. Cluster 4 is exclusively juveniles, and all JUV cells were assigned to this population. INF1 and KS cells are mixed in with the PLC and MLC clusters; however, they are largely missing from the ILC cluster (2). Because of this, we reason that the remaining four clusters (0, 1, 5, and 6) are likely to be ILCs affected by pathology, and indeed, these clusters express multiple markers of ILCs such as *DLK1*, *IGF1* (MIM: 147440), and *HSD11B1* (MIM: 600713).

To further confirm the interpretation of these clustering results, we selected one marker each from clusters 3, 2, and

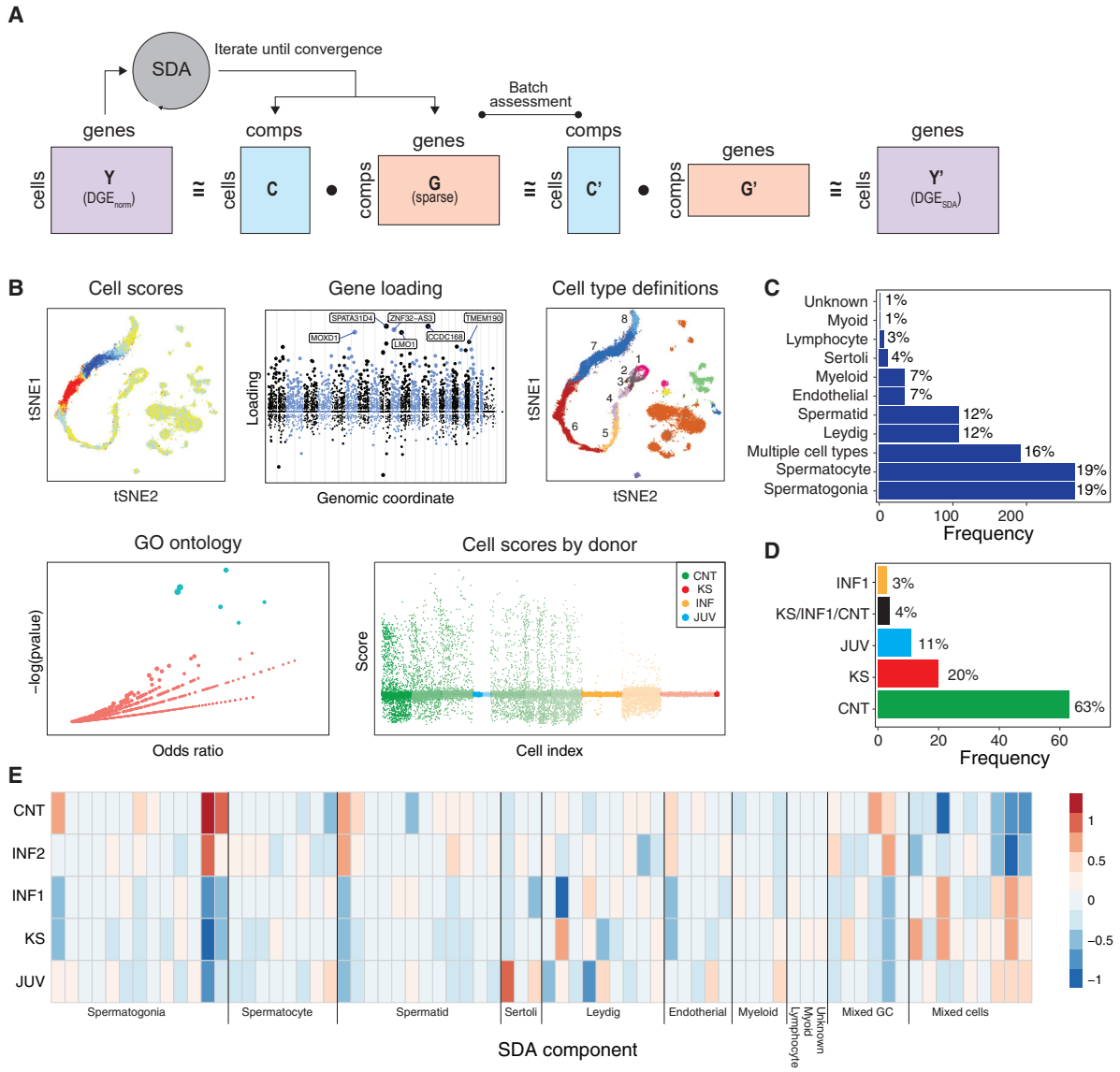


Figure 2. Overview of SDA analysis

(A) SDA is used to decompose the input DGE matrix into cells-by-components (C) and genes-by-components (G) matrices (material and methods). The number of components is defined by the user, and for this analysis, it is set to 150. After removal of batch-related components, we retained 75 components, reflected by new matrices, C′ and G′. A batch-corrected DGE matrix was obtained as the dot-product of C′ and G′, producing a new matrix, DGE_{SDA}, which was used for all downstream analysis.

(B) Schematic overview of the sources of information used to interpret each SDA component.

(C) Distribution of cellular expression patterns for all retained SDA components.

(D) Distribution of the primary source of cells loading on each SDA component.

(E) Heatmap of cell score residuals for each component, stratified by donor and sorted by cell type. Strong positive residuals reflect an excess of cells with loading on a particular SDA component, while negative reflects the converse. Small variations around 0 are expected due to chance.

7 for experimental validation by RNA FISH with healthy control tissue. For cluster 3 (PLCs), we selected *SHROOM2*, a poorly characterized gene that we predicted to have high specificity for PLCs; for cluster 2 (ILCs), we selected *DLK1*, and for cluster 3 (MLCs), we selected *INSL3*. We obtained clear and crisp staining for all three markers from human testis. *SHROOM2* appeared as a bright, punctate nuclear signal within peritubular cells, interstitial cells, and rarely, within spermatogonia (Figure 3E). *SHROOM2*+ nuclei from peritubular cells were flattened, elongated, and regularly spaced around the tubules (Figure 3E). *SHROOM2*+ peritubular cells were also apparent in similar numbers and configuration within the testis of *Macaca fuscata* (Figure S4). Multiplex staining for *SHROOM2*, *DLK1*, and *INSL3* produced positive signal in at least half of interstitial cells, the majority of which were positive for only a single probe, suggesting that PLCs, ILCs, and MLCs co-exist in adult testis. Cells positive for only one of each marker were observed in close proximity within the same interstitial space (Figures 3E and 3F). Occasionally double-positive

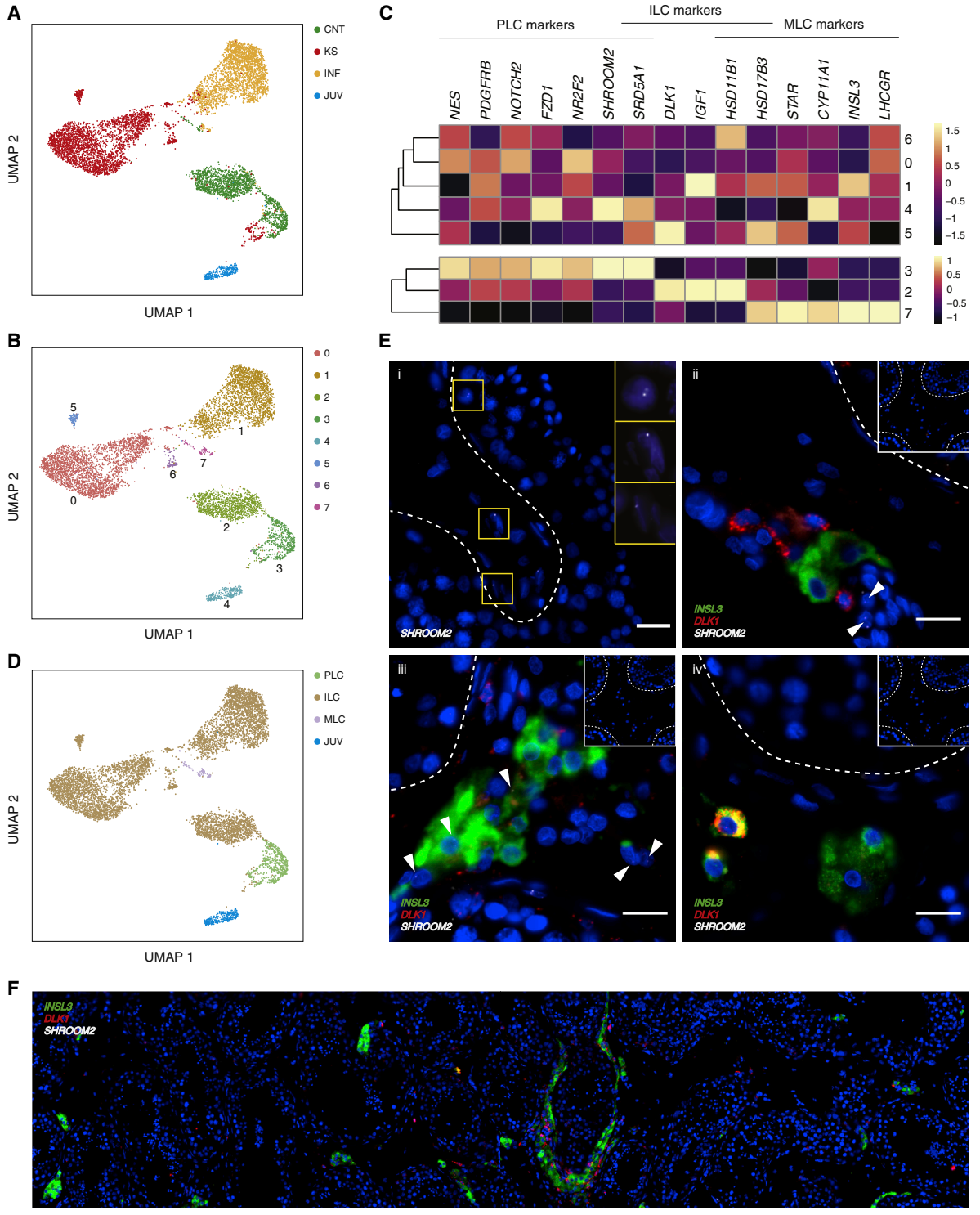


Figure 3. Three subtypes of Leydig cells coexist in adults

(A) UMAP representation of LC clusters colored by donor type.

(B) UMAP representation of LC subclusters identified by semi-supervised clustering. Each subcluster is labeled with a distinct number.

(C) Expression heatmap of markers used to interpret LC subclustering; row numbers correspond to subcluster labels used in (B). PLCs, progenitor LCs; ILCs, immature LCs; MLCs, mature LCs.

(D) The same data as shown in (A) and (B) but labeled with inferred LC subtypes.

(E) RNA fluorescent *in situ* hybridization (FISH) of LC markers in normal adult human testis. (Ei) *SHROOM2* (gray signal), a candidate marker for PLCs identified from the scRNA-seq data, can be found in cells within the tubules, in peritubular cells, and in interstitial cells. An example of each type is delimited with a yellow box and magnified (insets). (Eii–Eiv) Expression of *INSL3*, *DLK1*, and *SHROOM2* appeared to be mutually exclusive within interstitial space. In (Eiv), however, some double-positive cells were observed, such as these

(legend continued on next page)

cells were observed, perhaps representing transitional states (Figures 3E and 3F).

Leydig cell expression patterns are consistent with differentiation in adults

To further boost the sample size and replicate the signals observed in our original data, we attempted to integrate LC data and SC data from two other published studies,[18,19] which adds an additional four KS-affected individuals, three NOA-affected individuals, and five control individuals (material and methods). The optimal solution we identified for this integration was not able to preserve disease-specific clustering, but the three major subpopulations of PLCs, ILCs, and MLCs were readily apparent (Figures 4A–4C). Because of logistical limitations, this integrated dataset was not used for SDA analysis but strictly for replicating and clarifying the biological signals identified in the initial SDA dataset. We distinguish between the two datasets by using the terms "SDA dataset" (n = 12 donors) and "integrated dataset" (n = 24 donors).

Using the integrated dataset, we found reproducible differences in the composition of LC subtypes when comparing KS to control individuals (Figure 4D; Table S4). The proportion of ILCs is higher in KS compared to control individuals (81% versus 69%), while the proportion of both PLCs and MLCs is smaller (18% versus 30% and 1.1% versus 1.4%, respectively). When considering just the two datasets with both affected individuals and control individuals (Zhao and Utah/Mahyari), the proportions estimated for both studies are very similar (Table S4). Like in KS, the proportion of MLCs appears to be lower in individuals with NOA compared to control individuals (0.8% versus 1.4%). However, unlike KS, there was no clear pattern in the proportion of PLCs and ILCs among NOA datasets, perhaps reflecting a heterogeneity in disease mechanisms in NOA that is not present in KS.

Using transcription-based scoring of cell-cycle states, we categorized all cells in our atlas into one of three cell-cycle states—G1/G0, G2M, or S phase (material and methods). Aggregating across all donors, inferred cell-cycle state was not randomly distributed across the three LC subpopulations (Chi-square = 156, $p < 1 \times 10^{-32}$). While 100% of adult control MLCs were annotated as G1/G0, the proportion dropped to 90% of PLCs and 85% of ILCs (Figure 4E). Strikingly, these relative proportions were replicated across the KS and INF/NOA subsets, and there were greater overall numbers of cells out of G1/G0 in KS and INF/NOA (Figure 4E). Overall, these numbers indicate that LC development from PLCs to MLCs may be occurring in adult life and that the proportion of LCs undergoing division and differentiation is higher in the testis of men with KS and idiopathic NOA.

To further confirm these findings, we performed RNA FISH on healthy adult testis with a probe for *MKI67*, a well-known marker of cell proliferation.[33] *MKI67* was present in the nuclei of both spermatogonia and interstitial LCs (Figure 4F, Figure S5). Finally, we performed pseudotime analysis, a method for ordering a set of developmentally related cells according to developmental stage (materials and methods). This analysis ordered the LCs according to their expected developmental relationship—PLCs, followed by ILCs, and then MLCs (Figure 4G, Figure S6). Intriguingly, ILCs spanned the broadest amount of developmental space, while MLCs spanned the smallest, consistent with a transcriptionally homogeneous post-mitotic cell population.

Leydig cell SDA components capture variation in cellular development and function

SDA identified ten components with high cell scores from KS donors: three LC components (36, 85, and 147), two endothelial cell (EC) components (27 and 100), two immune cell (IC) components (139 and 8), and three components affecting multiple cell types (22, 57, and 121). The KS cells in our data are depleted of SCs and lack germ cells; as a result, there is less power to detect the gene expression signatures of these cell types.

Fourteen SDA components demonstrated variable expression across LC subpopulations (Figure 5). On the basis of their expression patterns, we subdivided these SDA components into five groups: PLCs, MLCs, all donors, KS-specific, and shared by the azoospermic donors (INF1, KS1, and KS2). The first four groups reflect expression variation in normal testicular function, while the last two highlight pathology. These SDA components do not simply reflect differences in expression across space or time of genes involved in the same biological processes but represent distinct sets of genes that are each enriched for specific functional annotations (Figure S7).

SDA components 57 and 132 highlight genes that are expressed solely in the PLC subset and genes that are developmentally regulated during differentiation from PLC to ILC, respectively. Strikingly, among the top 100 genes loading on 57, many are genes more typically associated with fibroblasts and muscle function, such as *TAGLN* (MIM: 600818), *MYH11* (MIM: 160745), *ACTA2* (MIM: 102620), *LMOD1* (MIM: 602715), and *ACTN1* (MIM: 102575). This is consistent with a recent finding that undifferentiated mesenchymal cells pass through a "myofibroblast" intermediate during *in vitro* differentiation to Leydig cells.[34] SDA component 132 indicates that as PLCs transition to ILCs, they reduce expression of cytoskeleton-related proteins and increase expression of many genes involved in the production of ECM, the complement pathway, TGF-β signaling, and regulation of inflammatory

DLK1+, *INSL3*+ cells, which may represent a transitional intermediate between ILCs and MLCs. (Inset) A field from the experimental negative control (DapB probe). White arrows point to *SHROOM2*+ cells. White dashed lines mark the approximate location of tubule limits. Nuclei are stained in blue by Hoechst33342. Scale bars represent 20 μm.
(F) A broader field of the slide used to generate images in (E).

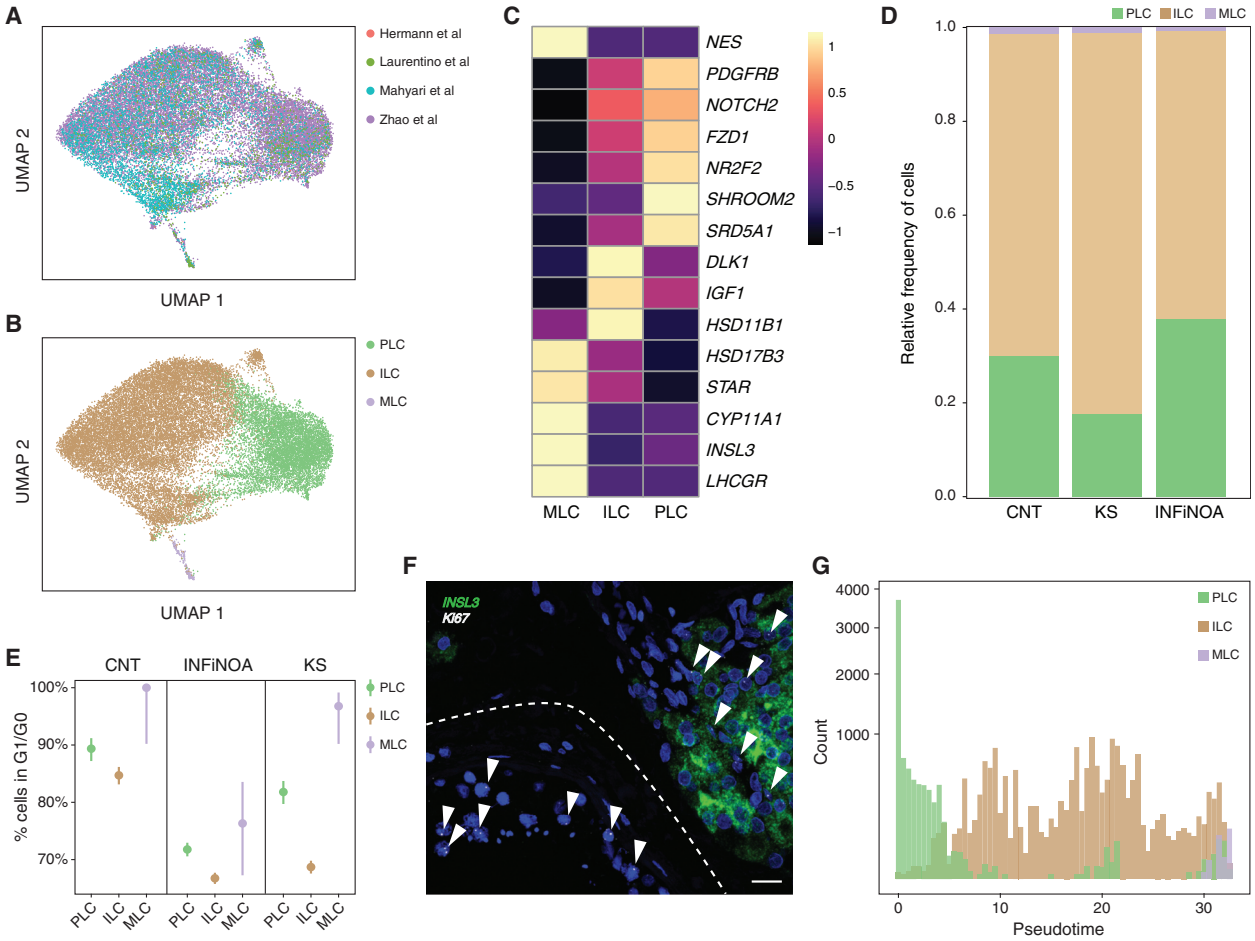


Figure 4. Additional evidence that three Leydig cell subtypes coexist in adults and are developmentally related

(A) UMAP projection of cells from an integrated LC dataset combining data from three prior studies[11,18,19] with data generated in this study (material and methods). All analyses described in this figure are based on this integrated dataset.

No obvious batch structure corresponding to publications is apparent in the 2D UMAP projection after integration, although LC subpopulations corresponding to disease states are lost.

(B) Three primary subpopulations were produced by semi-supervised clustering, corresponding to PLCs, ILCs, and MLCs.

(C) Heatmap of LC marker expression levels from subpopulations defined in (B).

(D) Frequency distribution of LC subtypes found in each donor population. CNT, control (n = 10); KS, Klinefelter syndrome-affected individuals (n = 5); INFiNOA, idiopathic non-obstructive azoospermia-affected individuals (n = 4).

(E) Cell-cycle analysis of LCs. The proportion of cells in G1/G0 was tabulated separately for PLCs, ILCs, and MLCs and compared across conditions.

(F) RNA FISH of *MKI67*, a marker of cell proliferation, using normal adult human testis. *MKI67*-positive nuclei (white arrows) were observed inside tubules and intermingled among INSL3+ interstitial cells, consistent with cell division of LCs in adults.

(G) Pseudotime analysis of LC scRNA-seq data orders cells in a manner consistent with the proposed cell type definitions.

response. SDA component 118 clearly marks the MLC subpopulation; it has *LHCGR*, *INSL3*, *STAR* (MIM: 600617), and *INHA* (MIM: 147380) among the highest loading genes and functional enrichments for categories involving steroid metabolism and transport (Figure S7).

SDA components 36, 85, and 147 are general indicators of spermatogenic failure

Because the signals of SDA components 36, 85, and 147 have high cell scores for all donors with spermatogenic failure but lower scores on the other donor cells, we interpret these as reflecting expression changes secondary to the lack of developing germ cells. The basis for these expression changes should include elevated LH signaling[35] but is likely to include dysregulation of other signaling mechanisms. SDA component 147 is highly enriched for genes involving response to stimulus, including a large number of immediate early genes important for mediating extracellular signaling: *JUN* (MIM: 165160), *JUNB* (MIM: 165161), *FOS* ([MIM: 164810), *FOSB* (MIM: 164772), and *IER2*. The transcription factors *MAFB* (MIM: 608968) and *NR2F1* (MIM: 132890) are highly upregulated in the infertile individuals. Genes with negative loading on SDA component 147 have reduced expression in KS and INF1 ILCs compared to all other LCs. These are functionally enriched for genes involved in ECM (*BGN* [MIM: 301870], *TNS1* [MIM: 600076], *FBLN5* [MIM: 604580], and *PRELP* [MIM: 601914]) and, interestingly, ossification (*DLK1*, *OMD*

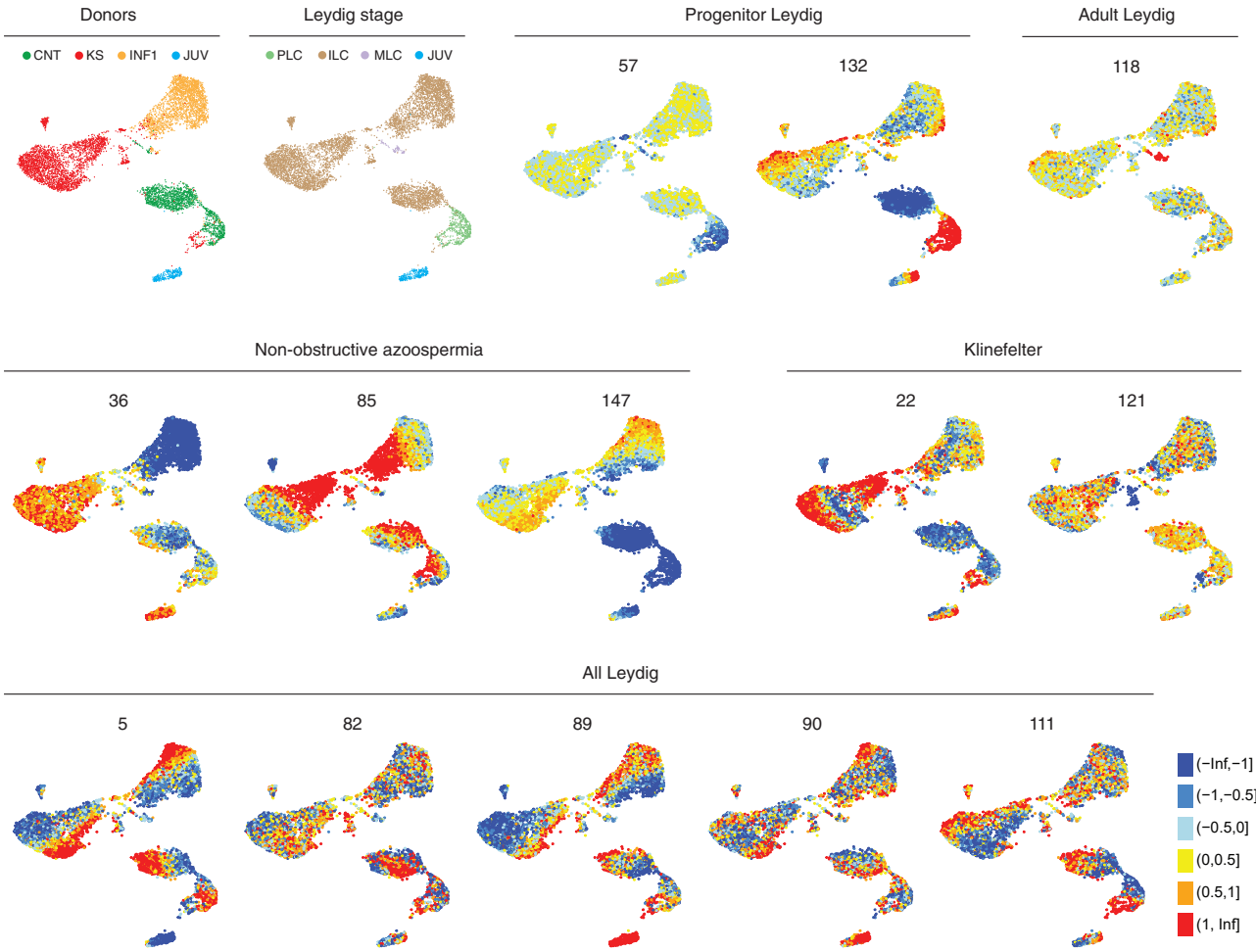


Figure 5. SDA cell scores projected on a UMAP map for all SDA components specific to LCs
SDA components are organized by expression pattern into five groups described in the text.

[MIM: 618926], *OSR1* [MIM: 608891], *PKDCC* [MIM: 614150], and *SFRP1* [MIM: 604156]).

Genes that load positively on SDA component 85 are upregulated in KS and INF1 ILCs compared to all other LCs. These are highly enriched for genes involved in the regulation of RNA processing and splicing (*CCNL1* [MIM: 613384], *DDX17* [MIM: 60849], *PNISR* [MIM: 616653], and *FUS* [MIM: 137070]). As mentioned above, SDA component 85 is significantly correlated with cell cycle state and thus may reflect proliferation of LCs that is sometimes observed in cases of spermatogenic failure.[36]

SDA components 36 and 22 may identify Leydig cell response to loss of Sertoli cells

KS is unusual among genetic causes of spermatogenic failure: beyond the loss of germ cells, it has a very striking phenotype of tubular degeneration, including loss of SCs and tubule boundaries, that begins at the onset of puberty.[3] By including an individual with idiopathic azoospermia in our analysis (INF1), we were able to separate out signals in LCs that reflect the loss of SCs from those that reflect loss of germ cells. Two SDA components, 36 and 22, show strong differential expression between INF1 and KS ILCs.

SDA component 36 represents genes that are inversely regulated in INF1 and KS ILCs: genes that are positively loaded in KS are negatively loaded in INF1. Interestingly, all JUV LCs also load positively on this component, which suggests that mature SC-derived signaling may be the source of modulation of SDA component 36 genes because mature SCs are absent or depleted from both JUV tubules and KS tubules (Figure 5). Consistent with this hypothesis, we noted five secreted signaling factors, *RSPO3* (MIM: 610574), *IGFBP5* (MIM: 146734), *ADM* (MIM: 103275), *WFIKKN2* (MIM: 610895), and *INHBA* (MIM: 147290), in the top 15 positively loaded genes of this component, and we characterize these in more detail below.

Functional annotation of SDA component 22 identified responses to unfolded/topologically incorrect proteins and protein refolding, as well as interactions with the immune system (Figure S7). The positively loaded genes were significantly enriched for immediate early-response genes. Combined, these results suggested a pathology that results in inflammation and immune engagement. Among the top

genes of SDA component 22 was *SERPINE1* (MIM: 173360), which encodes PAI-1, the primary inhibitor of tissue plasminogen activator, which was highly upregulated in ILCs from KS but not INF1. Upregulation of *SERPINE1* is expected to block plasmin activity and hence block the degradation of fibrous ECM components such as fibronectin and laminins.

Expression changes identified by SDA replicate with additional donors

To confirm the expression changes identified by SDA, we performed conventional DE analysis on the "integrated" dataset, which increased the number of individuals with KS from two to six, the number of individuals with idiopathic NOA from one to four, and the number of control individuals from seven to ten. We confirmed that this expression signature of genes on component 36 replicated with the larger set of cases and controls in our "integrated dataset." Seventy of the top 100 positive gene loadings on component 36 were differentially expressed between KS and NOA (q-value < 0.05); likewise, 68/100 of the top negative gene loadings were significantly differentially expressed (Figure S8, Table S5).

Signaling factors specific to KS Sertoli cells are dysregulated

SCs formed three distinct subpopulations in our global clustering analysis: one adult population and two clusters comprised primarily of JUV cells (Figure 6A). We equate these to the two immature SC subpopulations that we previously reported, IMM1 and IMM2.[30] We identified three SDA components specific to SCs. SDA component 130 is expressed in all SCs and consists of classic markers of SC lineage such as *SOX9* (MIM: 608160), *AMH* (MIM: 600957), *INHA*, and *CLU* (MIM: 185430). SDA component 73 distinguishes the two immature subpopulations from the adult one. SDA component 122 distinguishes IMM2 from IMM1 and adult SCs. Notably, unlike LCs, we didn't identify a distinct cluster of KS SCs. To expand the representation of this important cell type in our analysis, we again integrated our data with other published results, this time by using SCs (material and methods). Subclustering and pseudotime analysis of this integrated SC dataset revealed additional subpopulations of cells, largely grouped by pubertal state and disease state (Figures 6B and 6C).

When considering only genes expressed in at least 10% of SCs, DE analysis identified 2,865 genes with at least 1.5-fold change in expression between KS and CNT (q-value < 0.05, Wilcoxon test). However, 716 (25%) of these changes can be also identified when comparing NOA and CNT, indicating that a large fraction of changes are simply secondary responses to spermatogenic impairment (Figure 6D). Notably, some of the most strongly downregulated genes in both KS and NOA are genes highly expressed in meiotic and postmeiotic germ cells (e.g., *STAG3* [MIM: 608489], *SYCP2* [MIM: 604105], *PRM1* [MIM: 182880], and *PRM2* [MIM: 182890]), indicating that such transcripts are either naturally found within SCs or were present as a result of technical artifacts. Six genes showed very strong upregulation in KS (>16-fold increase) but little or no change in NOA: *CTRB1* (MIM: 118890), *ENG* (MIM: 131195), *MRAP* (MIM: 609196), *NKAIN4* (MIM: 612873), *PLAC9* (MIM: 612857), and *TMSB4X* (MIM: 300159). *TMSB4X* is particularly notable because thymosin has been shown to stimulate secretion of GNRH from the hypothalamus.[37] In our original SDA analysis, *GNRH1* (MIM: 152760) expression was greatly reduced in KS SCs ($p < 2 \times 10^{-12}$) compared to CNT, but levels in INF1 and CNT were indistinguishable ($p = 0.13$, Figure S10). However, *GNRH1* was not included in some of the datasets used for the "integrated" analysis and thus we are not able to replicate this signal.

Numerous gene ontology (GO) categories were significantly enriched in genes with DE along the KS branch of SC development (Figure S9). Among the top categories were RNA-binding and mitochondrial metabolism, unfolded protein response, and response to stimulus. Notably, the most significant enrichments were for genes encoding secreted proteins and proteins localized to extracellular organelles such as exosomes.

We found 43 secreted signaling factors in the full list of genes differentially expressed between KS and CNT; 39/43 of these were downregulated in KS compared to control individuals. The gene with the largest increase, *MIF* (MIM: 153620), showed a 5-fold increase over control individuals; this same gene was also highlighted as differentially expressed in the Zhao et al. study, but we also find evidence for this change with just the SDA dataset. Among the downregulated signaling factors were *INHBB* (MIM: 147390), *DHH* (MIM: 605423, and *KITLG* (MIM: 184745); numerous cytokines (*CXCL1* [MIM: 155730], *CXCL2* [MIM: 139110], *CXCL3* [MIM: 139111], *CXCL16* [MIM: 605398], *IL1A* [MIM: 147760], *IL7* [MIM: 146660], *IL13* [MIM: 147683], *TNFSF4* [MIM: 603594], and *TNFSF9* [MIM: 606182]); and WNTs (*WNT3* [MIM: 165330] and *WNT5A* [MIM: 164975]).

Twenty-three receptors were identified with significant expression changes in KS (q-value < 0.05, Wilcoxon test) but not NOA; 70% were decreases in expression. Among the receptors with KS-specific downregulation were WNT receptors (*FZD3* [MIM: 606143], *FZD4* [MIM: 604579], and *FZD8* [MIM: 606146]), FGF receptors (*FGFR1* [MIM: 136350], *FGFR2* [MIM: 176943], and *FGFR3* [MIM: 134934]), and IGF receptors (*IGF1R* [MIM: 147370] and *IGF2R* [MIM: 147280]).

XIST expression is lost in Sertoli cells of postpubertal KS donors

The long non-coding transcript *XIST* is involved in XCI and is normally inactive in XY cells. As expected, *XIST* was robustly and consistently expressed in all KS cell populations, except, much to our surprise, SCs, where *XIST* expression was similar to control levels in all three KS donors (Figure 6E). Projecting *XIST* expression levels onto

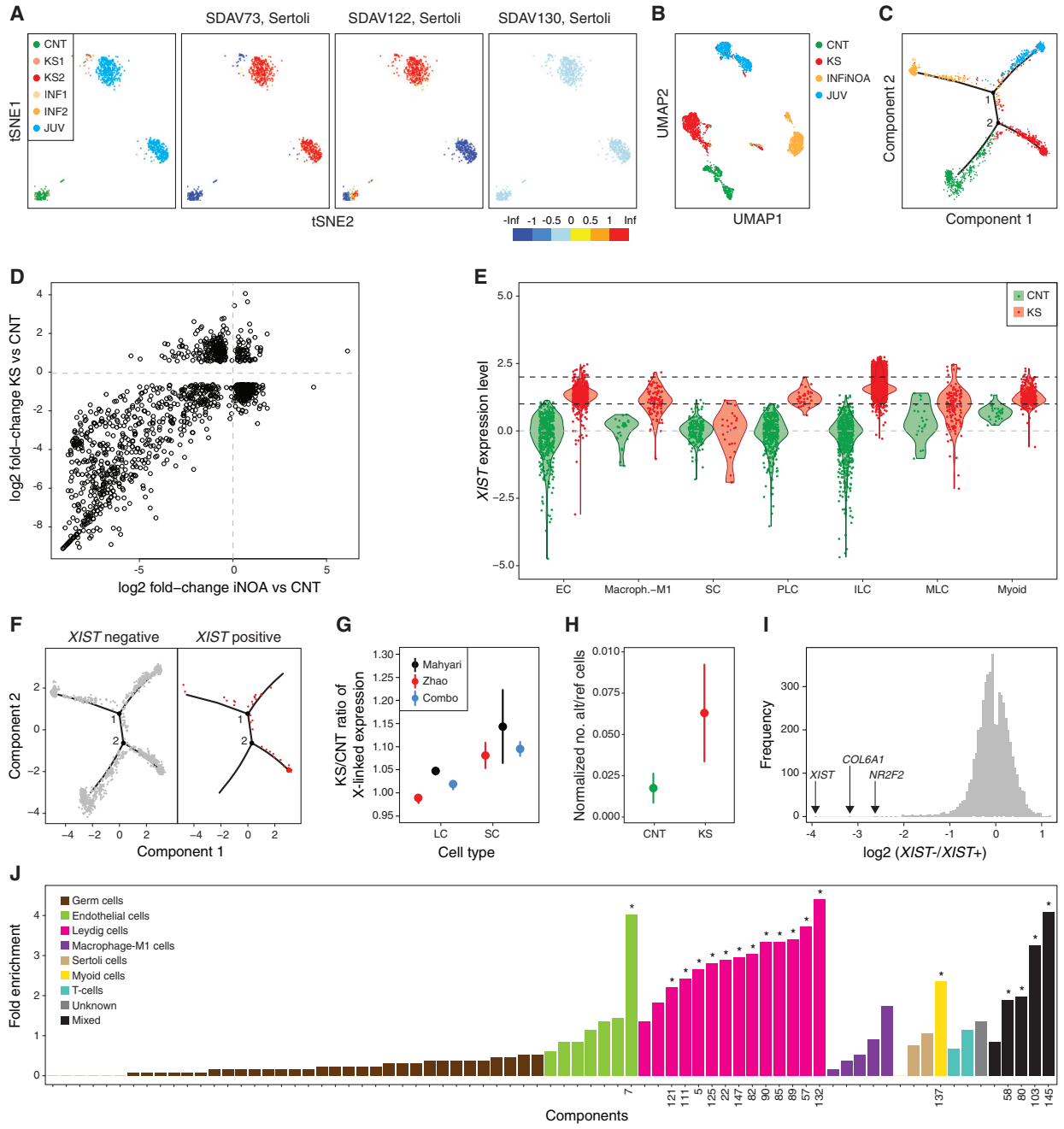


Figure 6. Analysis of SCs

(A) tSNE plots of the "SDA" dataset, indicating the donor labels for Sertoli cells (SCs), followed by the cell scores for the three SDA components with primary loading on SCs.

(B) UMAP summary of SC data from the "integrated" dataset.

(C) Diffusion analysis of data from (B).

(D) Comparison of gene-level expression differences between KS and CNT (y axis) and NOA and CNT (x axis) shows a large fraction of changes shared by KS and NOA.

(E) Expression of *XIST* in CNT and KS cell populations.

(F) Position of *XIST*+ and *XIST*− SCs in pseudotime trajectories from (C).

(G) Relative expression of all X-linked genes in KS compared to CNT cells. LCs, Leydig cells; SCs, Sertoli cells; Mahyari, ratios calculated with the "SDA" dataset of two KS donors and six control donors; Zhao, ratios calculated with published[18] data from three KS and five CNT donors; combo, ratios calculated with the combined Zhao and Mahyari data. Error bars represent 95% confidence intervals.

(H) We performed discovery and genotyping of germline genetic variation by using the RNA-seq reads for all cells annotated as SCs. We observed a 3-fold increase in the proportion of heterozygous genotype calls from the X chromosome in KS SCs compared to CNT SCs, consistent with loss of X inactivation in KS SCs. Error bars represent 95% confidence intervals.

(I) Histogram of log2-fold changes comparing expression in *XIST*− SCs to expression in *XIST*+ SCs.

(J) Differential expression analysis of *XIST*+ and *XIST*− cells identified 227 significant genes. When we assessed the distribution of these genes across all SDA component gene loadings, we found them to be strongly enriched on LC components. The asterisk equals an SDA component with enrichment for DE genes at $p < 0.05$ by Fisher's exact test. Components are ordered and colored by primary cell type of expression.

the pseudotime graph of KS SCs, we did not observe a simple correlation with development (Figure 6F). To clarify whether loss of *XIST* expression was matched with the expected increase in expression of X-linked genes, we compared the proportion of total reads mapping to X-linked genes between CNT and KS libraries (Figure 6G). In LCs, a cell type where *XIST* is robustly expressed in KS samples, we estimated there was a 1.9% increase in transcript abundance of X-linked genes, most likely corresponding to increased expression of genes known to escape X inactivation.[38] In SCs, X-linked expression was significantly higher (10.5% increase) but still well below the theoretical expected value of a 100% increase. These effect estimates showed the same trends when stratifying the data by study source (LC ratio 0.99–1.05; SC ratio 1.06–1.14; Figure 6G).

One possible explanation for SC-specific loss of *XIST* is that SCs are much more likely to revert to a euploid XY karyotype in XXY individuals. We tested this explanation by genotyping single-nucleotide variants (SNVs) in X-linked transcripts from both KS and control SCs. If the *XIST*− SCs of KS donors are euploid, then the rate of heterozygous SNVs inferred from KS cells should be identical to that from XY cells; however, we observed a 3-fold excess of heterozygous SNVs in KS compared to control individuals (Figure 6H). This suggests that the loss of *XIST* expression in SCs does not result from the reversion to a euploid XY karyotype.

While the large majority of SCs in our integrated data were *XIST*−, we identified rare *XIST*+ SCs robustly expressing *SOX9* and numerous other SC markers. To further understand possible causes and effects of *XIST* deactivation, we performed DE analysis of *XIST*+ and *XIST*− cells from KS donors and identified 276 differentially expressed genes at a q-value of 0.05 (Table S6). The majority (n = 262) were autosomal genes. Among X-linked genes, we saw significant increases in dosage of *FATE1* (MIM: 300450), *LINC00684, HMGN5* (MIM: 300385), *CITED1* (MIM: 300149), *HPRT1* (MIM: 308000), *BEX1* (MIM: 300690), *NDUFA1* (MIM: 300078), *SSR4* (MIM: 300090), and *TCEAL5*. All of these genes are normally expressed in SCs, and to our knowledge, none of these genes are known to escape XCI in normal cells. Curiously, the two genes most strongly co-expressed with *XIST* in SCs were *COL6A* (MIM: 120220) and *NR2F2* (MIM: 107773) (Figure 6I). The latter is a key interstitial cell marker that we find broadly expressed in all LCs and peritubular myoid cells (PMCs). Similarly, a large fraction of the other differentially expressed genes identified are normally expressed in LC lineages but not SC lineages (Figure 6J). These results indicate that loss of *XIST* expression in KS SCs may be related to problems in cell-fate specification within these cells.

XIST expression is associated with tubule thickness

To validate and further clarify our inferences from scRNA-seq, we obtained testis tissue from a KS donor not involved in the original scRNA-seq data generation and characterized the frequency and spatial context of *XIST*+ cells by using immunofluorescent staining with antibody and RNA probes (material and methods). The overall appearance of the section reflected clear signs of KS pathology. We identified numerous "type A" seminiferous tubules with normal shape and morphology interspersed with highly abnormal "type B" tubules with extremely thick tubule walls, which are associated with Leydig cell hyperplasia and fibrosis. While there were no clear signs of germ cells in the tubules examined, the cells that remained within the tubules were positive for SC marker SOX9+ (Figure 7A). Consistent with the scRNA-seq data, we saw a much larger number of *XIST*−, SOX9+ cells than *XIST*+, SOX9+ cells. Interestingly, the *XIST*+ cells were much more likely to be found in type A tubules compared to the thick, abnormal type B tubules ($p = 0.0024$, Figures 7B and 7C). These findings collectively highlight the altered transcriptional profile and identity of SCs in abnormal seminiferous tubules in KS.

Signaling pathways are altered in KS

During analysis of SDA components enriched in KS cell populations, we noted that several ligands and receptors for cell signaling pathways were among the top differentially expressed genes. Among LCs, *RSPO3*, *IGFBP5*, *ADM*, *WFIKKN2*, and *INHBA* are secreted signaling factors in the top 15 genes most enriched in KS ILCs compared to INF1 ILCs (i.e., on SDA component 36) and thus may provide insights into the specific pathology of KS (Figures 8A–8C). RSPO3 is a potentiator of WNT signaling (Figure 8B). Of all the somatic cells in our testis atlas, we found that SCs were the only cell type to robustly express *LGR3* (MIM: 603372) and *ZNRF3* (MIM: 612062), encoding the receptors for RSPO3 (Figure 8A). Control SCs also expressed a number of other WNT pathway components, including *FZD3*, *FRZB*, and *CTNNB1* (MIM: 116806) (Figure 8A). Further, expression of these WNT signaling components was essentially abolished in KS SCs; thus, we posit that upregulation of *RSPO3* in KS LCs may be a response to loss of some normal output of WNT signaling produced by SCs.

Under normal conditions, IGFBP5 binds to IGF1 and IGF2 to modulate their signaling activity. However, recent experimental evidence has demonstrated that overexpression of IGFBP5 increases the expression of pro-fibrotic genes, including itself, leading to a feed-forward loop that can trigger or maintain fibrosis in pulmonary tissues.[39] ADM binds to CALCR on ECs but has also been shown to stimulate INHBB secretion from SCs.[40] *WFIKKN2* acts as a regulator for a wide variety of growth factors in the TGF-β superfamily.[41] The precise effect of WFIKKN2 interaction depends on the target; it can antagonize GDF8 and GDF11 activity or fine-tune the spatial localization of BMP2, BMP4, and TGF-β1. INHBA has diverse functions in the testis, and homodimerizes to form Activin-A, a powerful stimulator of FSH secretion. The adverse effects of Activin-A overexpression include fibrosis.[42]

These five genes were identified by manual inspection of gene lists. In order to systematically quantify changes in

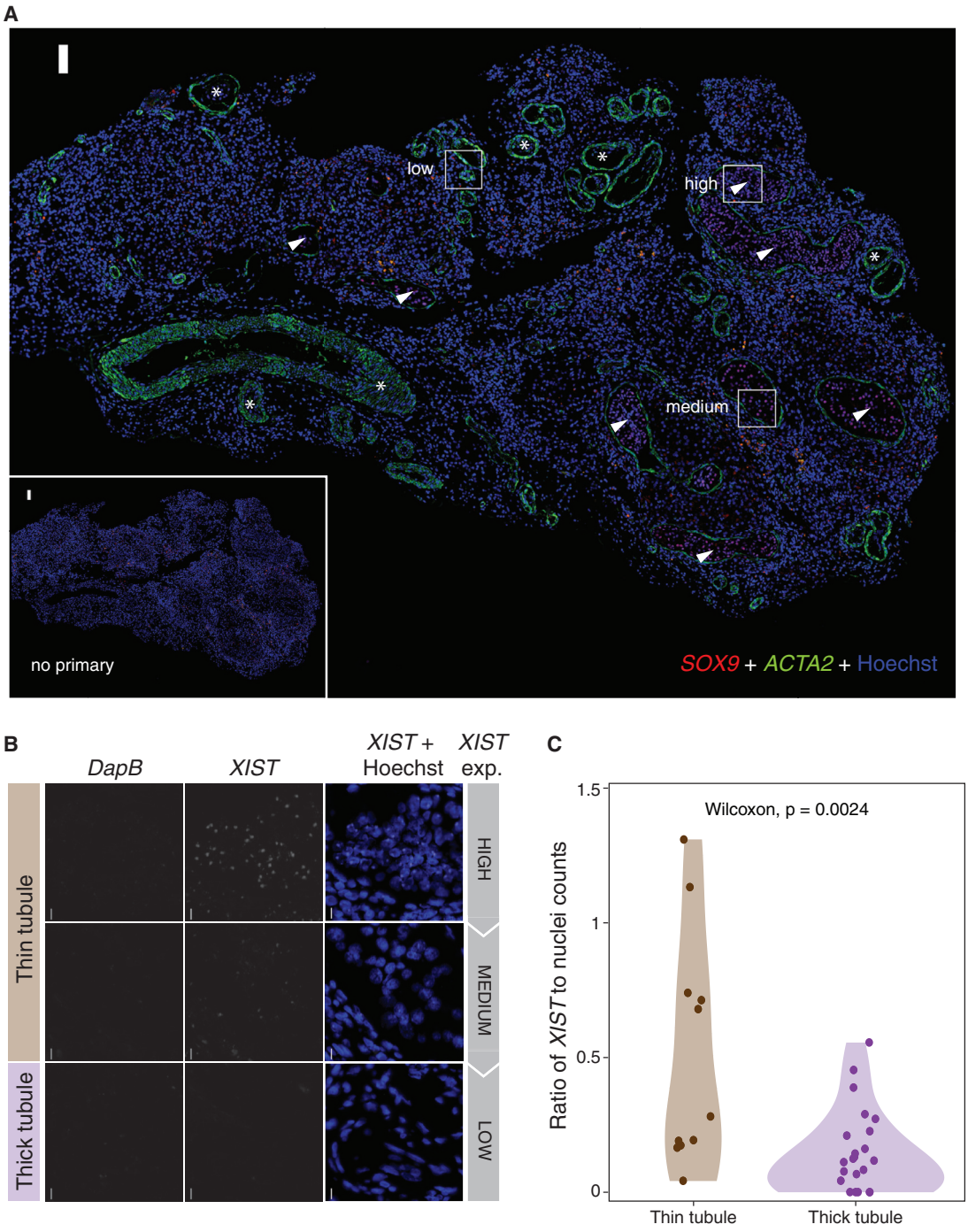


Figure 7. *XIST*− SCs are enriched in tubules with thick walls

(A) Immunofluorescence staining of a KS testis biopsy using antibodies against SOX9, a marker of Sertoli cells, and ACTA2, a smooth muscle actin highly expressed by peritubular myoid cells that surround the seminiferous tubules. The staining revealed two subpopulations of seminiferous tubules: one with tubule walls of normal thickness ("thin") and one with thicker walls, perhaps the result of accumulation of ECM components and/or myoid cell hyperplasia ("thick"). Cells within "thin" tubules were SOX9+, confirming their identity as Sertoli cells. Representative tubules with low, medium, and high XIST expression are highlighted for further analysis.

(B) Adjacent section of the same tissue shown in (A), stained for *XIST* using an RNA FISH probe. Shown are representative close-ups of thin tubules, showing Sertoli cells with medium and high *XIST* expression, as well as a close-up of a thick tubule, showing Sertoli cells with low *XIST* expression. The color of the boxes in (A) indicates the spots adjacent to the zoom boxes in (B), which are marked with corresponding colors. DapB, negative control probe.

(C) We used digital image analysis to quantify the number of *XIST*+ spots in thin tubules (n = 11 tubules) and thick tubules (n = 19 tubules). *XIST*+ spot numbers were normalized to the number of nuclei in each tubule. The number of *XIST*+ spots was higher in thin tubules compared to thick (p = 0.0024, Wilcoxon test).

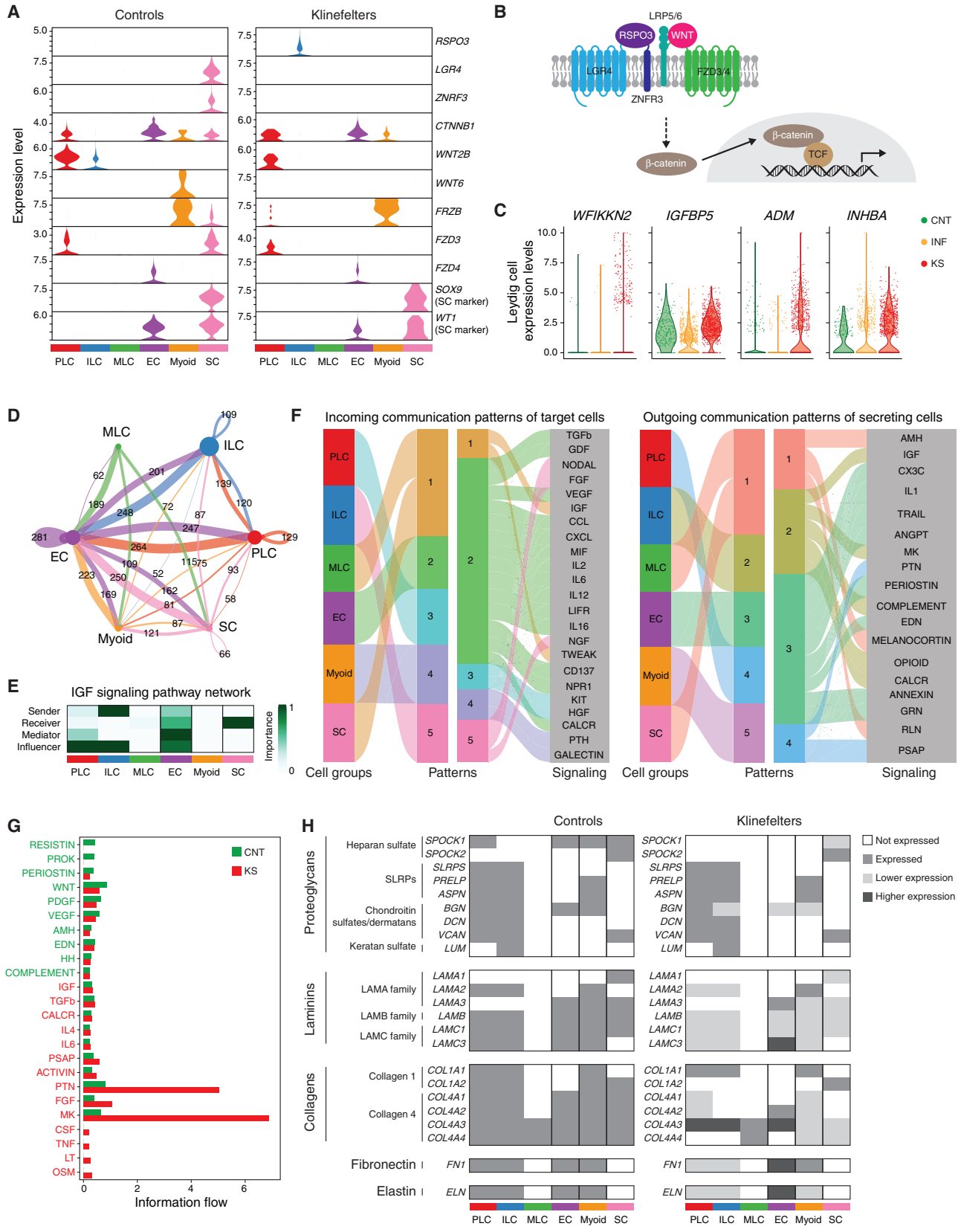


Figure 8. Dissection of intercellular signaling and ECM production in KS testis

(A) We identified five secreted signaling factors in the top 15 genes of SDA component 36, reflecting genes with upregulation in KS LCs compared to LCs from control individuals and from an individual with idiopathic NOA. Among these was *RSPO3*, a potentiator of WNT signaling. Among control cells, the RSPO3 receptors *LGR4* and *LZNRF3* were only expressed in SCs; expression of these genes and other downstream targets of the WNT pathway that are normally expressed in SCs are lost in KS SCs.

(legend continued on next page)

signaling networks in KS cells, we used CellChat[26] to infer autocrine and paracrine signaling network activity among ECs, PMCs, LCs, and SCs (material and methods). CellChat provides tools to characterize cell-to-cell signaling from scRNA-seq data on the basis of a curated database of receptor-ligand interactions. Beginning from a database of 140 signaling pathway families, we identified 62 families that are likely to be active among the somatic cells that we profiled in normal adult testes (Figure 8D). Using network mediation analysis, we identified the dominant "sender" and "receiver" of each signaling pathway. For example, in the insulin growth factor (IGF) signaling network, ILCs are the primary "senders" and SCs are the primary "receivers" (Figure 8E). Interestingly, when considering all pathways, ECs and PLCs were the dominant senders and receivers among the six cell types considered here. MLCs are almost exclusively senders, acting as a terminal endpoint of LC development unlikely to be directly affected by changes in other cell types.

Using non-negative matrix factorization, we summarized these 62 signaling pathways into five patterns that describe the major cell types that are involved in producing the ligands ("outgoing patterns") or receptors ("incoming patterns") for each pathway (material and methods). Unique outgoing patterns were mapped to ECs, ILCs, and SCs, while PLCs, MLCs, and PMCs were mapped to a single pattern (Figure 8F). Unique incoming patterns were mapped to ECs and PMCs, while PLCs and ILCs shared one pattern, as did MLCs and SCs. The specific pathways associated with each pattern were consistent with existing knowledge of testicular cell signaling, and, using the scRNA-seq expression matrix, we were able to confirm the expected gene expression patterns for the uncharacterized pathways inferred by CellChat.

Next, we performed the same inference of signaling pathway activity by using the KS scRNA-seq data, this time identifying 72 pathways active, 16 of which were not identified in the control data. Comparing the activity of pathways identified in both KS and control individuals, we found 47 pathways with greater activity in KS and 32 pathways with lower activity (Figure 8G, Table S7). The top five upregulated pathways were MK (6.2× increase), FGF (4.4×), PTN (4×), Activin (2.7×), and PSAP (2.7×).

Endothelial and Leydig cells may collaborate to drive fibrosis in KS

Because of the range of LC physiological states detected across donors, these data are a rich resource for building models of how ECM is built and modified in the interstitial space. We assessed the expression patterns of laminins, collagens, proteoglycans, fibrin, elastin, matrix metalloproteinases, and other matricellular proteins to construct a picture of ECM dynamics in aging and disease (Table S8, Figure 8H).

Among adult control individuals, LCs are the primary generators of proteoglycans, expressing 14/21 genes tested, followed by PMCs (n = 9), SCs (n = 7), and ECs (n = 5). Intriguingly, of the LCs, PLCs and ILCs seem to be the predominant producers of proteoglycans; JUVs expressed only six and MLCs only expressed the activin receptor *TGFBR3* (MIM: 600742) (Table S8). *SPOCK1* (MIM: 602264), whose product is the protease inhibitor testican, is uniquely expressed only in PLCs, while *LUM*, encoding lumican, is specifically expressed in ILCs. Expression of all three laminin families was detected among the full set of somatic cells, and *LAMA1* (MIM: 150320) was highly and specifically expressed in SCs (Figure 8H). We also interrogated the 16 major types of collagen (I–XVI) and found expression of all except II and XI. While there is broad expression of many collagens across the four primary somatic cell types, there is structure to the expression. Specifically, LCs have the highest average expression for 13 collagen genes, ECs and PMCs both have the highest expression for six genes, and only *COL9A1* appears specific to SCs. The genes *FN1* (MIM: 135600) and *ELN* (MIM: 130160), encoding other major components of the ECM, were robustly expressed in LCs, PMCs, and ECs but not SCs.

We performed detailed comparisons of expression for these ECM-related genes between KS and control individuals. Surprisingly, the general trend was for expression of these genes to be either lower or unchanged in KS with a

(B) RSPO3 interacts with other components of the WNT signaling pathway, some shown in (A), to regulate beta-catenin signaling.
(C) Among the other top signaling factors of SDA component 36 are *WFIKKN2*, *IGFBP5*, *ADM*, and *INHBA*, which all have expression that is specific to KS LCs or highly enriched in KS LCs compared to LCs in controls and idiopathic NOA (INF).
(D) Number of significant ligand-receptor pairs identified between pairs of testicular cell types. The color of the edge indicates the cell type origin of the ligand. Significant autocrine relationships are indicated by a loop that begins and ends at the same node.
(E) Example heatmap showing the importance of each cell type based on network centrality measures of IGF signaling in control cells. ILCs appear to be the primary secretor of IGF proteins and SCs appear to be the primary receiver on the basis of expression levels of IGF receptors.
(F) Non-negative matrix factorization (NNMF) was used to organize signaling networks into five patterns of incoming communication based on co-expression of receptors. Likewise, a separate NNMF analysis was used for organization of outgoing communication into five patterns on the basis of co-expression of secreted proteins.
(G) Significant signaling pathways were ranked on the basis of their differences in information flow within the inferred networks for control individuals (CNT) and KS-affected individuals. Pathway labels indicate the condition with greater network activity; green labels are networks with higher activity in CNT and red labels are networks with higher activity in KS.
(H) A semiquantitative heatmap comparing the expression level of ECM components in KS cell populations against the baseline expression level in controls. Expressed, expressed at the same level in control and KS donors; higher/lower expression, higher/lower expression in KS donors compared to controls.

small number of notable exceptions (Figure 8H). *FN1*, *ELN*, *LAMC3* (MIM: 604349), and eight collagen genes (including *COL4A3* [MIM: 120070]) were upregulated in ECs, while *COL4A3* and *COL7A1* (MIM: 120120) were upregulated in LCs. The overexpression of *COL4A3* is of particular interest, as epitopes of COL4A3 are the antigens recognized by autoantibodies in the autoimmune disorder Goodpasture syndrome.[43] It is well known that in KS, and other forms of spermatogenic impairment, pathology is often accompanied by an increase in immune cell numbers and activity in the interstitium.[44]

Additional SDA components identify other KS-specific signatures

For completeness, we briefly mention highlights of SDA components that load primarily on ECs, PMCs, T cells, and macrophages (Figure S11). EC SDA components captured differences in signaling and activation (SDA components 31 and 100), between juvenile and adult cells (SDA component 56), and between arterial and venous cells (SDA component 27). We identified only one SDA component clearly marking PMCs, 137, which seems to indicate KS-specific changes shared by KS1 and KS2 but not observed in any other donor. The genes *PLN* (MIM: 172405) and *RERGL* are good markers for this component; they have high expression in KS donor cells but low expression in other donors. PLN regulates the Ca^{2+} pump in muscle cells, causing changes in contractility and relaxation upon extracellular signals like epinephrine. There are seven SDA components loading on immune cell populations; some of these clearly separate lymphocyte and myeloid lineages, and some distinguish cells detected in KS and INF1 from all other donors. The positive gene loadings on SDA component 14 clearly identify MHC-II antigen presentation genes of M1 macrophages, while the negative gene loadings correspond to the M2 alternative-activation state. Interestingly, the positive scored cells in this component are the KS and INF1 samples, whereas the negative scored cells are CNT. It is likely this differential explains the extra stress on the cells of the testis system that forces a polarization of the macrophages to deal with the inflammation and wound healing common among the pathologies. Finally, we identified one SDA component, 50, that, on the basis of GO enrichment and gene loadings, may mark a cell population that has not been described in prior human testis cell atlases (Figure S11). There are 119 cells in this population, which are primarily detected in KS donors. The existence of this cluster as a distinct cell population will need experimental verification to clarify whether this population exists in normal tissues, reflects a pathological state, or is a technical artifact.

Discussion

KS was first recognized as a chromosomal disorder in 1957.[45] As the oldest known genetic form of male infertility, it has been a paradigm for the male infertility community, leading to decades of research that has very carefully described the timing and nature of physiological changes in XXY males but has yet to provide a conclusive causal molecular mechanism of infertility. Here, we have combined scRNA-seq with new statistical methods to identify cell populations, pathways, and individual genes that are likely to be important pieces in the pathology of KS. We demonstrated that our approach was useful for identifying numerous sets of genes with expression patterns specific to men with idiopathic NOA, men with KS, or both. We wish to highlight that, while it is certain that the KS- and NOA-affected individuals differ in their X chromosome counts, there is likely to be heterogeneity in the pathology of the NOA-affected individuals that will influence our ability to reliably define expression changes specific to KS. Future analysis of this type will benefit from using NOA case definitions based on etiology and not phenotype, which in turn should lead to more reproducible conclusions. It is likely that a comparative approach such as the one pursued here will be useful for dissecting pathogenic mechanisms in complex tissues affected by other heterogeneous genetic disorders.

The key finding of our work may be the surprising observation that *XIST* expression is lost in the SCs of postpubertal KS-affected individuals. We validated that this loss of expression appears to be functional, and X-linked genes show a concomitant increase in expression. Interestingly, we estimated that loss of *XIST* expression only leads to a 11%–14% increase in expression output from the X chromosome in SCs, much less than the theoretical increase of 100%. This is consistent with prior observations in female mice lacking *XIST*, where the increase in X-linked expression was in the range of 14%–36%, findings that have been attributed to an as-yet-undescribed backup system for dosage compensation.[46] Our observations are consistent with old histological data that SCs of type A tubules invariably lacked Barr bodies, indicating no XCI, while type B tubules could contain a mixture of Barr-body-positive and -negative SCs.[9] Curiously, the relatively straightforward interpretation of the Barr body data as informative about XCI in KS does not appear to have been pursued in the past.

We found that *XIST*+ SCs ectopically express numerous autosomal genes that are primarily found in LCs. Taken together, these results add to a growing number of observations that define the unusual epigenetic plasticity of SCs. Conditional deletion of the transcription factor *DMRT1* (MIM: 602424) from adult SCs causes them to transform to granulosa cells, the female analog of SCs.[47] Conditional deletion of the transcription factor *WT1* (MIM: 607102) from embryonic SCs leads to a phenotype similar to fetal Leydig cells, and complete loss of seminiferous tubules, producing testis histology that is reminiscent of KS.[48,49] The fact that X chromosome copy number influences the epigenetic state of the SC but not other cells of the testis presents an interesting puzzle and suggests that X

chromosome inactivation and sex determination may have a shared evolutionary history with the differentiation program of somatic cells of the gonads.

These results bring clarification to the cell types that are involved in KS pathology and what the temporal sequence of events might be. SCs are likely to be the key cell type involved in initiating the testis pathology of KS. Loss of *XIST* expression in SCs may be triggered by changes in the endocrine environment at puberty, perhaps in response to increased testosterone production from LCs or another diffusible signal. Loss of SC function and differentiation status could explain defects in germ cell development simply by loss of one or more of the numerous SC processes that are essential for spermatogenesis. De-differentiation or trans-differentiation caused by changes in X-linked gene dosage could be disastrous in the context of testis homeostasis. Testicular cell types re-use the same signaling pathways multiple times at different stages of development.[50] Primary defects in the SC, including loss of expected signals or gain of unexpected signals, would in turn stimulate secondary changes in expression from interstitial cells, such as those reflected in the SDA components 22 and 36. Expression programs that are upregulated in the interstitium could lead to local increased production of ECM components by interstitial cells attempting to foster SC maturation and/or assembly of the tubules. This is consistent with the finding of an increase of *XIST*− SCs in thick tubules compared to thin, as increased KS tubule thickness has been shown to reflect the accumulation of the ECM components fibronectin, collagen I, and collagen IV.[51] This interpretation of KS pathology is largely ignoring the role of germ cells, largely because we have so few germ cells to study from the KS donors in our study and, thus, little idea of gene expression changes in such cells. It is certainly possible that there are primary causes of KS that occur in XXY germ cells that are currently invisible to us; such events would presumably either directly or indirectly lead to germ loss, either by cell-autonomous effects or through SC dysfunction.

The model of KS pathology that we favor invokes the concept of reciprocal, direct communication between SCs and LCs, ECs, and perhaps other interstitial cell types, mediated by diffusible factors in the testis, or at least indirect communication that is confined to the testis. Relatively little research has been done to identify diffusible signals from SCs to LCs, but there is evidence that such signals may exist.[52] Experiments that have applied treatments to a single testis, or even individual seminiferous tubules, found effects on LCs that were too locally restricted to be mediated by systemic changes in LH abundance.[53–55] We saw numerous genes related to WNT signaling pathways downregulated in KS SCs compared to control SCs (e.g., Figures 8A–8C). Curiously, in a recent scRNA-seq of KS testis, the authors concluded that the WNT pathway was upregulated in individuals affected by idiopathic NOA, and they found that treating SCs from individuals with NOA with a WNT inhibitor helped promote SC development and function.[18] Our analysis of genes with strong expression changes in KS SCs, but not INF1 SCs, identified *GNRH1* as a possible candidate that may be an important signaling factor between SCs and LCs responsible for maintaining testis homeostasis because of both changes in *GNRH1* expression directly and very strong upregulation of the X-linked gene *TMSB4X*, whose product has been shown to stimulate secretion of GNRH from the hypothalamus.[37] High-affinity GNRH receptors exist on LCs,[56] and stimulation of LCs with GNRH leads to downregulation of these receptors[57] and inhibition of testosterone production.[52] More work is needed to pick up these old research threads and better clarify the nature and function of the factors that mediate intercellular communication in the testis.

More broadly, cell-cell signaling is an essential aspect of testis homeostasis, and imbalances in signaling most likely contribute to disease in other forms of spermatogenic impairment as well. We were able to highlight several strongly dysregulated pathways with clear involvement of specific cell types, but our results suggest that the overall picture is quite complex, as 62 distinct signaling pathways were detected in our data from normal testis, and much more work is needed to understand these signaling pathways. One urgent need is to provide more background data on stimulus-response experiments where cultured primary cells are exposed to known ligands and then assayed for expression response. Co-culture and organoid systems will also be useful. Analysis methods that are capable of assessing the impact of quantitative changes in signaling activity also need to be developed. Emerging methods for spatial analysis of transcription could be of great value for further dissection of KS pathology by transforming the observation of signaling between cell types into an observation of signaling between cell states and contexts.

In conclusion, we provide a new transcriptional map of normal and abnormal human testis tissue explorable through our Shiny app tool, HISTA (web resources). We provide a framework to uncover molecular pathology to understand cellular function and dysregulation in the human testis and use this to dissect pathology in the KS testis. Our approach will be useful for both future exploration of single-cell genomic data and the development of computational tools to improve current approaches for the diagnosis and classification of individuals with infertility.

Data and code availability

scRNA-seq data generated for this project have been deposited at GEO under accession number GEO: GSE169062. The code for running the HISTA browser is available from github and zenodo: https://github.com/eisascience/HISTA (https://doi.org/10.5281/zenodo.4433041). The list of human cell-cycle genes for scoring single-cell libraries is available at https://raw.githubusercontent.com/hbc/tinyatlas/master/cell_cycle/Homo_sapiens.csv. The Dropsim R package for normalization is available at https://github.com/marchinilab/dropsim.

Supplemental information

Supplemental information can be found online at https://doi.org/10.1016/j.ajhg.2021.09.001.

Acknowledgments

This work was supported by funding from the National Institutes of Health (R01HD078641 and P50HD096723). Research reported in this publication was supported by the Office of the Director of the National Institutes of Health under award number P51OD011092 to the Oregon National Primate Research Center. The ONPRC Integrated Pathology Core provided support services for the research. The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health. We thank Arpiar Saunders, David Zarkower, and members of the Conrad Lab and Zarkower Lab for helpful discussions. We thank Sandra Laurentino and Nina Neuhaus for sharing data.

Declaration of interests

The authors declare no competing interests.

Received: March 17, 2021
Accepted: August 31, 2021
Published: October 7, 2021

Web resources

HISTA browser, https://conradlab.shinyapps.io/HISTA
OMIM, https://www.omim.org/
R Programming Environment, https://www.R-project.org/
Utility Plugins - GDSC ImageJ Toolsets, http://www.sussex.ac.uk/gdsc/intranet/microscopy/UserSupport/AnalysisProtocol/imagej/utility

References

1. Kanakis, G.A., and Nieschlag, E. (2018). Klinefelter syndrome: more than hypogonadism. Metabolism *86*, 135–144.
2. Klinefelter, H.F., Reifenstein, E.C., and Albright, F. (1942). Syndrome Characterized by Gynecomastia, Aspermatogenesis without A-Leydigism, and Increased Excretion of Follicle-Stimulating Hormone1. J. Clin. Endocrinol. Metab. *2*, 615–627.
3. Aksglaede, L., Wikström, A.M., Rajpert-De Meyts, E., Dunkel, L., Skakkebaek, N.E., and Juul, A. (2006). Natural history of seminiferous tubule degeneration in Klinefelter syndrome. Hum. Reprod. Update *12*, 39–48.
4. Aksglaede, L., and Juul, A. (2013). Testicular function and fertility in men with Klinefelter syndrome: a review. Eur. J. Endocrinol. *168*, R67–R76.
5. Forti, G., and Krausz, C. (1998). Clinical review 100: Evaluation and treatment of the infertile couple. J. Clin. Endocrinol. Metab. *83*, 4177–4188.
6. Heller, C.G., and Nelson, W.O. (1945). Hyalinization of the seminiferous tubules associated with normal or failing Leydig-cell function. Discussion of relationship to eunuchoidism, gynecomastia, elevated gonadotrophins, depressed 17-ketosteroids and estrogens. J. Clin. Endocrinol. *5*, 1–12.
7. Skakkebaek, N.E. (1969). Two types of tubules containing only Sertoli cells in adults with Klinefelter's syndrome. Nature *223*, 643–645.
8. Froland, A. (1969). Klinefelter's syndrome. Clinical, endocrinological and cytogenetical studies. Dan. Med. Bull. *16* (*Suppl 6*), 1–108.
9. Frøland, A., and Skakkebaek, N.E. (1971). Dimorphism in sex chromatin pattern of Sertoli cells in adults with Klinefelter's syndrome: correlation with 2 types of "Sertoli-cell-only" tubules. J. Clin. Endocrinol. Metab. *33*, 683–687.
10. Griswold, M.D. (2016). Spermatogenesis: The Commitment to Meiosis. Physiol. Rev. *96*, 1–17.
11. Hermann, B.P., Cheng, K., Singh, A., Roa-De La Cruz, L., Mutoji, K.N., Chen, I.C., Gildersleeve, H., Lehle, J.D., Mayo, M., Westernströer, B., et al. (2018). The Mammalian Spermatogenesis Single-Cell Transcriptome, from Spermatogonial Stem Cells to Spermatids. Cell Rep. *25*, 1650–1667.e8.
12. Guo, J., Grow, E.J., Mlcochova, H., Maher, G.J., Lindskog, C., Nie, X., Guo, Y., Takei, Y., Yun, J., Cai, L., et al. (2018). The adult human testis transcriptional cell atlas. Cell Res. *28*, 1141–1157.
13. D'Aurora, M., Ferlin, A., Di Nicola, M., Garolla, A., De Toni, L., Franchi, S., Palka, G., Foresta, C., Stuppia, L., and Gatta, V. (2015). Deregulation of sertoli and leydig cells function in patients with Klinefelter syndrome as evidenced by testis transcriptome analysis. BMC Genomics *16*, 156.
14. Winge, S.B., Dalgaard, M.D., Jensen, J.M., Graem, N., Schierup, M.H., Juul, A., Rajpert-De Meyts, E., and Almstrup, K. (2018). Transcriptome profiling of fetal Klinefelter testis tissue reveals a possible involvement of long non-coding RNAs in gonocyte maturation. Hum. Mol. Genet. *27*, 430–439.
15. D'Aurora, M., Ferlin, A., Garolla, A., Franchi, S., D'Onofrio, L., Trubiani, O., Palka, G., Foresta, C., Stuppia, L., and Gatta, V. (2017). Testis Transcriptome Modulation in Klinefelter Patients with Hypospermatogenesis. Sci. Rep. *7*, 45729.
16. Winge, S.B., Dalgaard, M.D., Belling, K.G., Jensen, J.M., Nielsen, J.E., Aksglaede, L., Schierup, M.H., Brunak, S., Skakkebæk, N.E., Juul, A., et al. (2018). Transcriptome analysis of the adult human Klinefelter testis and cellularity-matched controls reveals disturbed differentiation of Sertoli- and Leydig cells. Cell Death Dis. *9*, 586.
17. Winge, S.B., Soraggi, S., Schierup, M.H., Rajpert-De Meyts, E., and Almstrup, K. (2020). Integration and reanalysis of transcriptomics and methylomics data derived from blood and testis tissue of men with 47,XXY Klinefelter syndrome indicates the primary involvement of Sertoli cells in the testicular pathogenesis. Am. J. Med. Genet. C. Semin. Med. Genet. *184*, 239–255.
18. Zhao, L., Yao, C., Xing, X., Jing, T., Li, P., Zhu, Z., Yang, C., Zhai, J., Tian, R., Chen, H., et al. (2020). Single-cell analysis of developing and azoospermia human testicles reveals central role of Sertoli cells. Nat. Commun. *11*, 5683.
19. Laurentino, S., Heckmann, L., Di Persio, S., Li, X., Meyer Zu Hörste, G., Wistuba, J., Cremers, J.-F., Gromoll, J., Kliesch, S., Schlatt, S., and Neuhaus, N. (2019). High-resolution analysis of germ cells from men with sex chromosomal aneuploidies reveals normal transcriptome but impaired imprinting. Clin. Epigenetics *11*, 127.
20. Hore, V., Viñuela, A., Buil, A., Knight, J., McCarthy, M.I., Small, K., and Marchini, J. (2016). Tensor decomposition for

multiple-tissue gene expression experiments. Nat. Genet. *48*, 1094–1100.
21. 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.e21.
22. Qiu, X., Mao, Q., Tang, Y., Wang, L., Chawla, R., Pliner, H.A., and Trapnell, C. (2017). Reversed graph embedding resolves complex single-cell trajectories. Nat. Methods *14*, 979–982.
23. Tirosh, I., Izar, B., Prakadan, S.M., Wadsworth, M.H., 2nd, Treacy, D., Trombetta, J.J., Rotem, A., Rodman, C., Lian, C., Murphy, G., et al. (2016). Dissecting the multicellular ecosystem of metastatic melanoma by single-cell RNA-seq. Science *352*, 189–196.
24. Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., Marth, G., Abecasis, G., Durbin, R.; and 1000 Genome Project Data Processing Subgroup (2009). The Sequence Alignment/Map format and SAMtools. Bioinformatics *25*, 2078–2079.
25. Li, H. (2011). A statistical framework for SNP calling, mutation discovery, association mapping and population genetical parameter estimation from sequencing data. Bioinformatics *27*, 2987–2993.
26. Jin, S., Guerrero-Juarez, C.F., Zhang, L., Chang, I., Ramos, R., Kuan, C.H., Myung, P., Plikus, M.V., and Nie, Q. (2021). Inference and analysis of cell-cell communication using CellChat. Nat. Commun. *12*, 1088.
27. Schindelin, J., Arganda-Carreras, I., Frise, E., Kaynig, V., Longair, M., Pietzsch, T., Preibisch, S., Rueden, C., Saalfeld, S., Schmid, B., et al. (2012). Fiji: an open-source platform for biological-image analysis. Nat. Methods *9*, 676–682.
28. Win, K.Y., Choomchuay, S., Hamamoto, K., and Raveesunthornkiat, M. (2018). Comparative Study on Automated Cell Nuclei Segmentation Methods for Cytology Pleural Effusion Images. J. Healthc. Eng. *2018*, 9240389.
29. Otsu, N. (1979). A Threshold Selection Method from Gray-Level Histograms. IEEE Trans. Syst. Man Cybern. *9*, 62–66.
30. Guo, J., Nie, X., Giebler, M., Mlcochova, H., Wang, Y., Grow, E.J., Kim, R., Tharmalingam, M., Matilionyte, G., Lindskog, C., et al.; DonorConnect (2020). The Dynamic Transcriptional Cell Atlas of Testis Development during Human Puberty. Cell Stem Cell *26*, 262–276.e4.
31. Jung, M., Wells, D., Rusch, J., Ahmad, S., Marchini, J., Myers, S.R., and Conrad, D.F. (2019). Unified single-cell analysis of testis gene regulation and pathology in five mouse strains. eLife *8*, 8.
32. Zirkin, B.R., and Papadopoulos, V. (2018). Leydig cells: formation, function, and regulation. Biol. Reprod. *99*, 101–111.
33. Miller, I., Min, M., Yang, C., Tian, C., Gookin, S., Carter, D., and Spencer, S.L. (2018). Ki67 is a Graded Rather than a Binary Marker of Proliferation versus Quiescence. Cell Rep. *24*, 1105–1112.e5.
34. Shen, Y.-C., Shami, A.N., Moritz, L., Larose, H., Manske, G.L., Ma, Q., Zheng, X., Sukhwani, M., Czerwinski, M., Sultan, C., et al. (2021). TCF21^{+} mesenchymal cells contribute to testis somatic cell development, homeostasis, and regeneration in mice. Nat. Commun. *12*, 3876.
35. Nieschlag, E. (2013). Klinefelter syndrome: the commonest form of hypogonadism, but often overlooked or untreated. Dtsch. Arztebl. Int. *110*, 347–353.
36. Aksglaede, L., Skakkebaek, N.E., Almstrup, K., and Juul, A. (2011). Clinical and biological parameters in 166 boys, adolescents and adults with nonmosaic Klinefelter syndrome: a Copenhagen experience. Acta Paediatr. *100*, 793–806.
37. Rebar, R.W., Miyake, A., Low, T.L., and Goldstein, A.L. (1981). Thymosin stimulates secretion of luteinizing hormone-releasing factor. Science *214*, 669–671.
38. Tukiainen, T., Villani, A.-C., Yen, A., Rivas, M.A., Marshall, J.L., Satija, R., Aguirre, M., Gauthier, L., Fleharty, M., Kirby, A., et al. (2017). Landscape of X chromosome inactivation across human tissues. Nature *550*, 244–248.
39. Nguyen, X.-X., Muhammad, L., Nietert, P.J., and Feghali-Bostwick, C. (2018). IGFBP-5 Promotes Fibrosis via Increasing Its Own Expression and That of Other Pro-fibrotic Mediators. Front. Endocrinol. (Lausanne) *9*, 601.
40. Chan, Y.-F., Tang, F., and O, W.-S. (2008). Adrenomedullin in the rat testis. II: Its production, actions on inhibin secretion, regulation by follicle-stimulating hormone, and its interaction with endothelin 1 in the Sertoli cell. Biol. Reprod. *78*, 780–785.
41. Monestier, O., and Blanquet, V. (2016). WFIKKN1 and WFIKKN2: "Companion" proteins regulating TGFB activity. Cytokine Growth Factor Rev. *32*, 75–84.
42. de Kretser, D.M., O'Hehir, R.E., Hardy, C.L., and Hedger, M.P. (2012). The roles of activin A and its binding protein, follistatin, in inflammation and tissue repair. Mol. Cell. Endocrinol. *359*, 101–106.
43. Hellmark, T., and Segelmark, M. (2014). Diagnosis and classification of Goodpasture's disease (anti-GBM). J. Autoimmun. *48-49*, 108–112.
44. Willems, M., Vloeberghs, V., Gies, I., De Schepper, J., Tournaye, H., Goossens, E., and Van Saen, D. (2020). Testicular immune cells and vasculature in Klinefelter syndrome from childhood up to adulthood. Hum. Reprod. *35*, 1753–1764.
45. Jacobs, P.A., and Strong, J.A. (1959). A case of human intersexuality having a possible XXY sex-determining mechanism. Nature *183*, 302–303.
46. Yang, L., Kirby, J.E., Sunwoo, H., and Lee, J.T. (2016). Female mice lacking Xist RNA show partial dosage compensation and survive to term. Genes Dev. *30*, 1747–1760.
47. Matson, C.K., Murphy, M.W., Sarver, A.L., Griswold, M.D., Bardwell, V.J., and Zarkower, D. (2011). DMRT1 prevents female reprogramming in the postnatal mammalian testis. Nature *476*, 101–104.
48. Zhang, L., Chen, M., Wen, Q., Li, Y., Wang, Y., Wang, Y., Qin, Y., Cui, X., Yang, L., Huff, V., and Gao, F. (2015). Reprogramming of Sertoli cells to fetal-like Leydig cells by Wt1 ablation. Proc. Natl. Acad. Sci. USA *112*, 4003–4008.
49. Gao, F., Maiti, S., Alam, N., Zhang, Z., Deng, J.M., Behringer, R.R., Lécureuil, C., Guillou, F., and Huff, V. (2006). The Wilms tumor gene, Wt1, is required for Sox9 expression and maintenance of tubular architecture in the developing testis. Proc. Natl. Acad. Sci. USA *103*, 11987–11992.
50. Heinrich, A., and DeFalco, T. (2020). Essential roles of interstitial cells in testicular development and function. Andrology *8*, 903–914.
51. Van Saen, D., Vloeberghs, V., Gies, I., De Schepper, J., Tournaye, H., and Goossens, E. (2020). Characterization of the stem cell niche components within the seminiferous tubules in testicular biopsies of Klinefelter patients. Fertil. Steril. *113*, 1183–1195.e3.

52. De Kretser, D.M. (1982). Sertoli cell—Leydig cell interaction in the regulation of testicular function. Int. J. Androl. *5*, 11–17.
53. Risbridger, G.P., Kerr, J.B., and de Kretser, D.M. (1981). Evaluation of Leydig cell function and gonadotropin binding in unilateral and bilateral cryptorchidism; evidence for local control of Leydig cell function by the seminiferous tubule. Biol. Reprod. *24*, 534–540.
54. Risbridger, G.P., Kerr, J.B., Peake, R.A., and de Kretser, D.M. (1981). An assessment of Leydig cell function after bilateral or unilateral efferent duct ligation: further evidence for local control of Leydig cell function. Endocrinology *109*, 1234–1241.
55. Aoki, A., and Fawcett, D.W. (1978). Is there a local feedback from the seminiferous tubules affecting activity of the Leydig cells? Biol. Reprod. *19*, 144–158.
56. Sharpe, R.M., and Fraser, H.M. (1980). Leydig cell receptors for luteinizing hormone releasing hormone and its agonists and their modulation by administration or deprivation of the releasing hormone. Biochem. Biophys. Res. Commun. *95*, 256–262.
57. Arimura, A., Serafini, P., Talbot, S., and Schally, A.V. (1979). Reduction of testicular luteinizing hormone/human chorionic gonadotropin receptors by [D-Trp6]-luteinizing hormone releasing hormone in hypophysectomized rats. Biochem. Biophys. Res. Commun. *90*, 687–693.

The American Journal of Human Genetics, Volume *108*

Supplemental information

Comparative single-cell analysis of biopsies clarifies pathogenic mechanisms in Klinefelter syndrome

Eisa Mahyari, Jingtao Guo, Ana C. Lima, Daniel P. Lewinsohn, Alexandra M. Stendahl, Katinka A. Vigh-Conrad, Xichen Nie, Liina Nagirnaja, Nicole B. Rockweiler, Douglas T. Carrell, James M. Hotaling, Kenneth I. Aston, and Donald F. Conrad

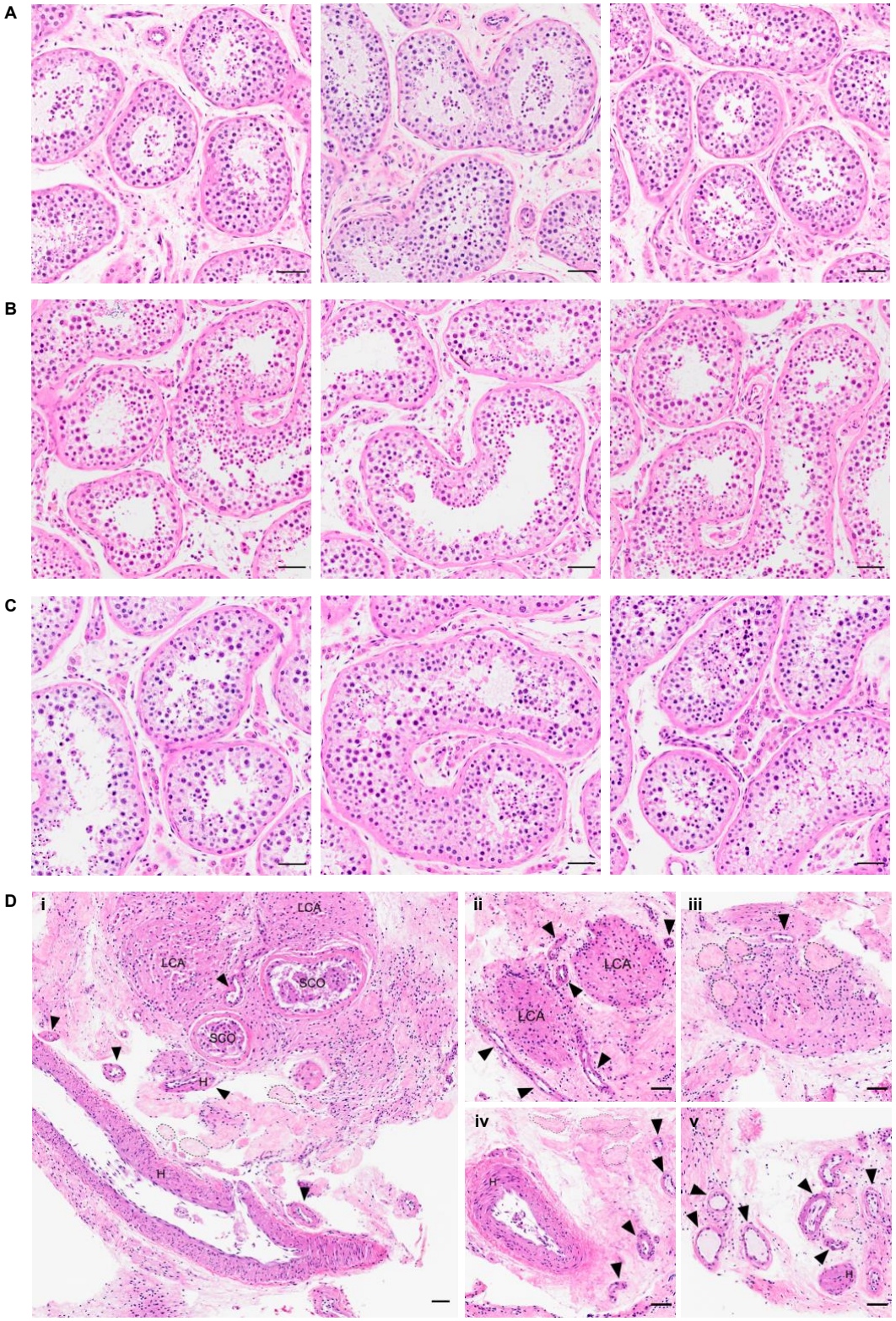


Figure S1. H&E staining of tissue samples. Control samples show normal testicular histology (A:Utah-D1; B:Utah-D2; C:Utah-D3). Three fields were obtained for each

sample showing the presence of seminiferous tubules with active spermatogenesis. The KS1 patient histology (D) presents the typical morphological abnormalities observed in Klinefelter Syndrome (fields D: i-v). Specifically, we can detect tubules with only Sertoli cells (SCO), empty tubules (arrowheads), some exhibiting the hyalinization of the tubule walls (H) that is characteristic of this syndrome. There is also an accumulation of degenerated/phantom tubules (dashed lines) and of Leydig cell aggregates throughout the tissue (LCA). Scale bars represent 50μm.

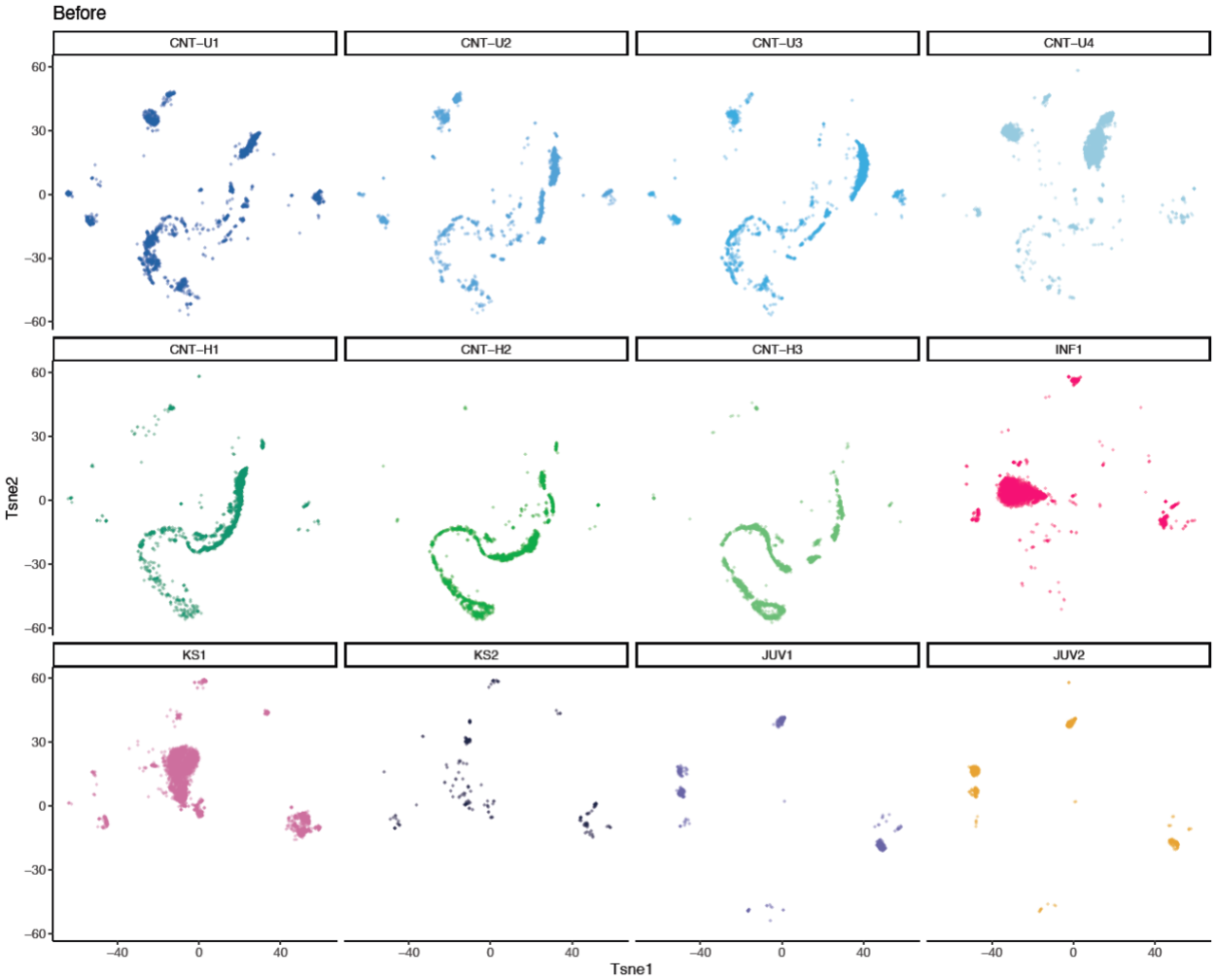


Figure S2. tSNE projection of original scRNA-seq data prior to QC and batch correction. This panel shows the same data as the first panel of Figure 1B, but split by donor, to better appreciate the contribution of each donor to the overall data.

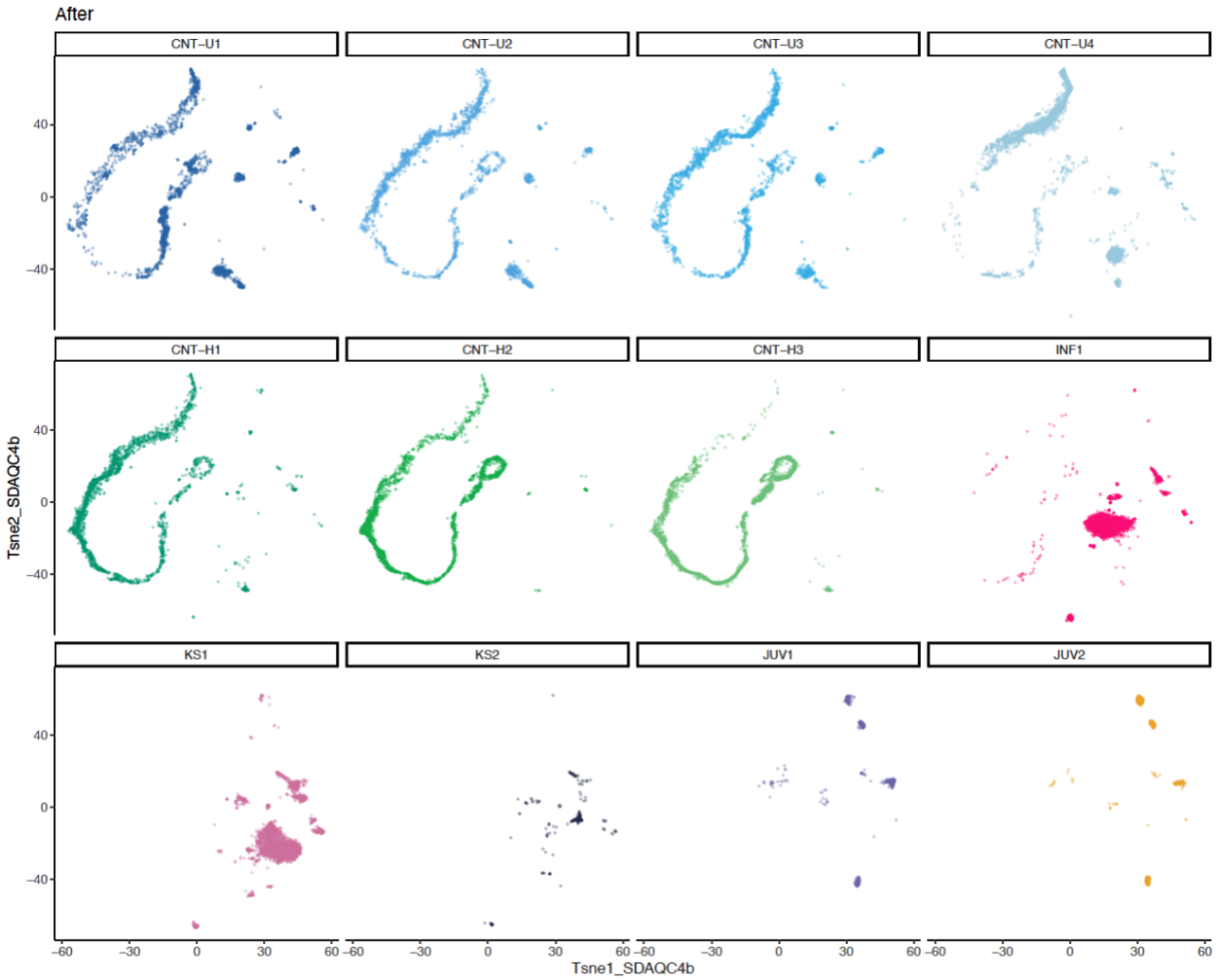


Figure S3. tSNE projection of original scRNA-seq data after QC and batch correction. This panel shows the same data as the second panel of Figure 1B, but split by donor, to better appreciate the contribution of each donor to the overall data.

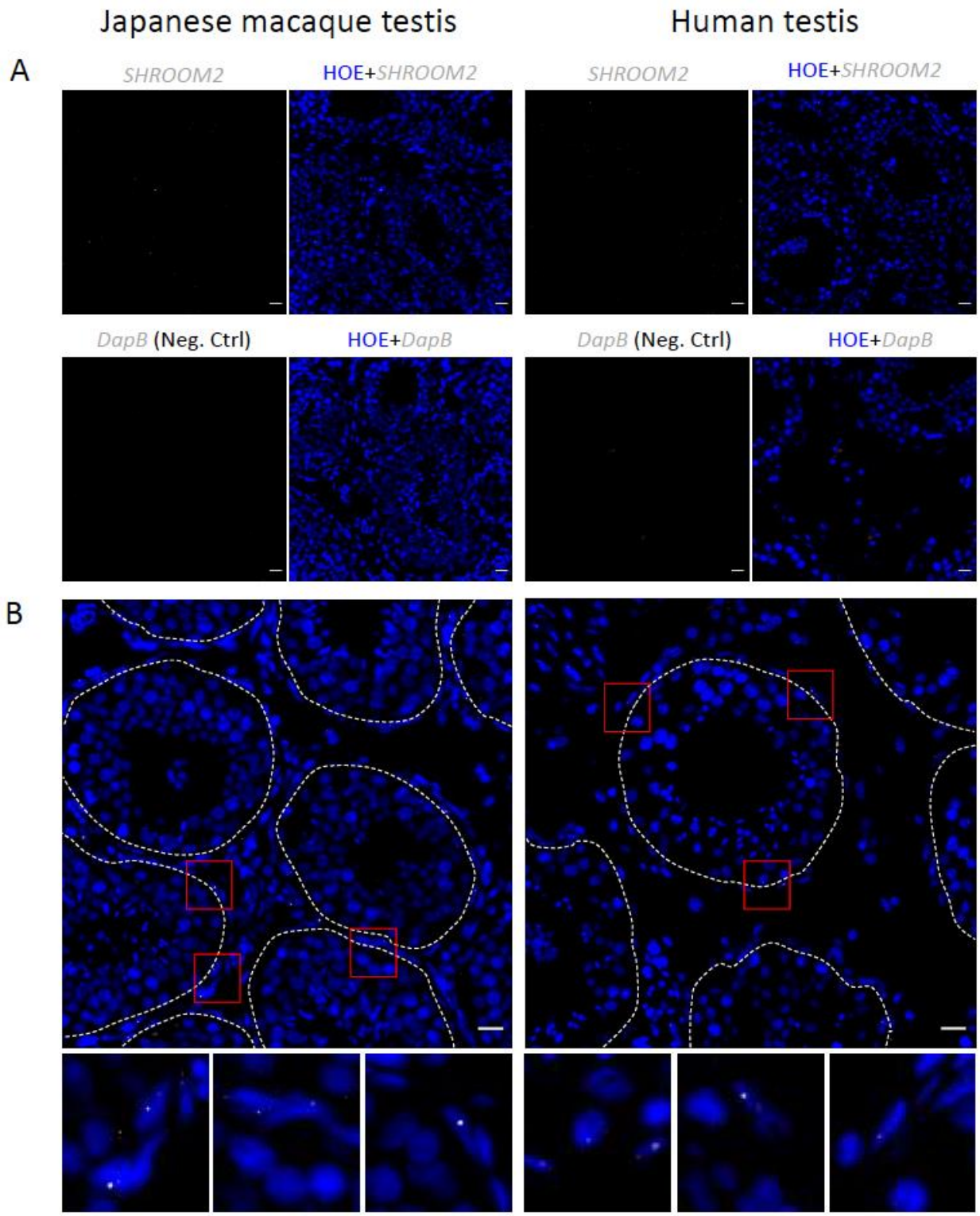

Japanese macaque testis
Human testis
A
SHROOM2
HOE+SHROOM2
SHROOM2
HOE+SHROOM2
DapB (Neg. Ctrl)
HOE+DapB
DapB (Neg. Ctrl)
HOE+DapB
B

Figure S4. *SHROOM2* expression is conserved between monkey and human testis. *SHROOM2* mRNA (in grey) was detected in Japanese macaque (left panels) and human (right panels) testis by FISH. Panel A shows the signal obtained in testis sections with the *SHROOM2* probe (top) compared to the lack of signal when using the negative control probe DapB (bottom) for both species. The expression of *SHROOM2* can be easily observed in panel B. DNA is stained in blue by Hoechst33342 and approximate tubule boundaries are marked in white dashed lines. In both species, some *SHROOM2* expression is seen in the tubules but most is in the interstitial space of the testis. Interestingly, a subtype of peritubular cells shows *SHROOM2* expression (red boxes zoomed below image), consistent with the expected location of progenitor Leydig cells (PLCs). Scale bars represent 20 μm.

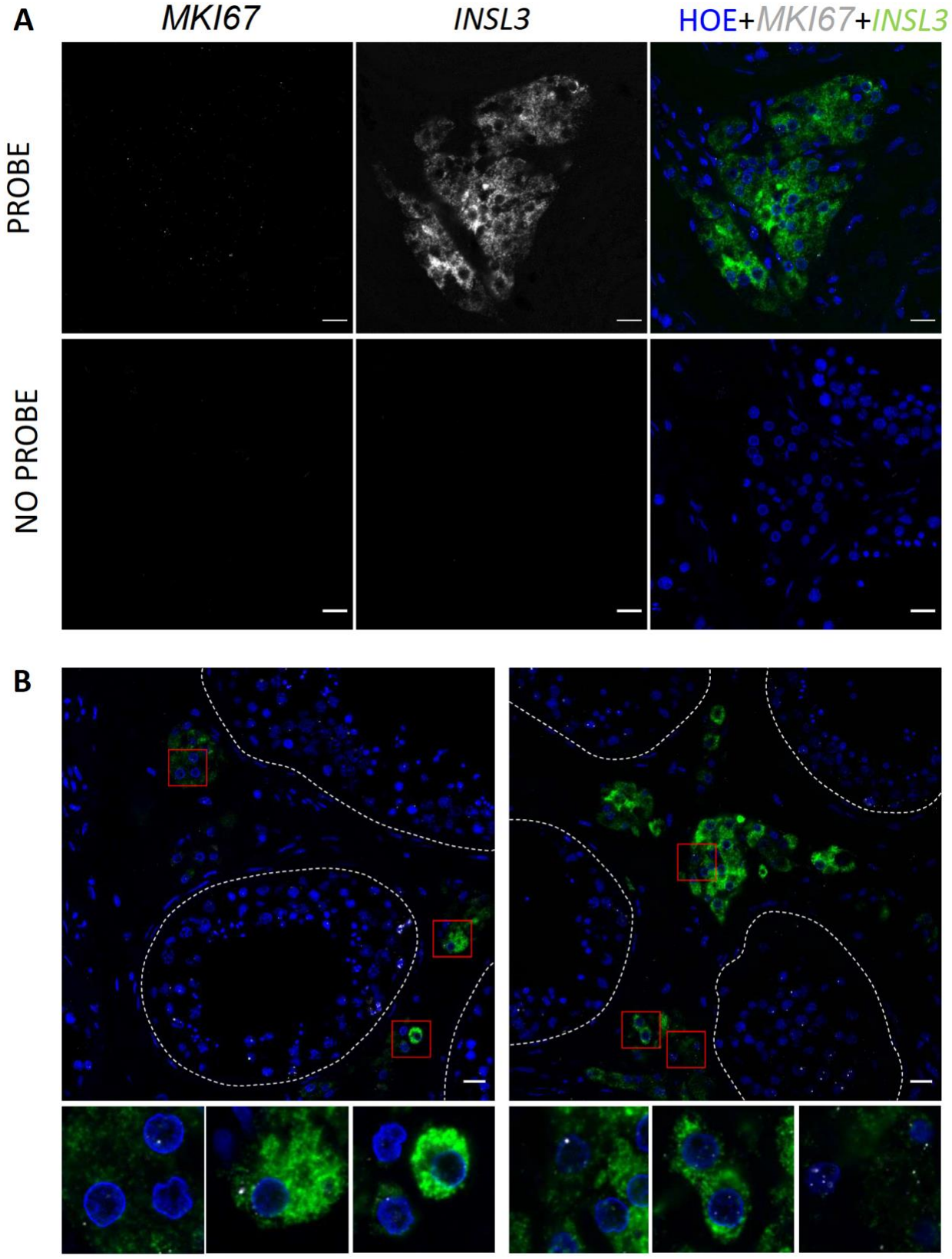

A
MKI67
INSL3
HOE+MKI67+INSL3
PROBE
NO PROBE
B

Figure S5. Leydig cell (LC) proliferation in the adult testis. Detection of *MKI67* (proliferation marker) and *INSL3* (LC marker) mRNAs in normal adult testis tissue by fluorescence *in situ* hybridization. Scale bars represent 20 µm. (A) Single channels and merged fields of the interstitial space staining positive for *MKI67* and *INSL3* mRNAs (PROBE). A random control field was obtained from an adjacent tissue section not incubated with probe (NO PROBE). (B) Different fields of the same section showing the expression *MKI67* (grey) and *INSL3* (green). Nuclei stained in blue. Approximate tubule boundaries marked by white dashed lines. As expected, *MKI67* stains stronger inside the tubules. Fields delimited by red boxes are zoomed at the bottom showing multiple examples of co-localization of *MKI67* and *INSL3*. These observations suggest that some LCs of the adult testis might be proliferating.

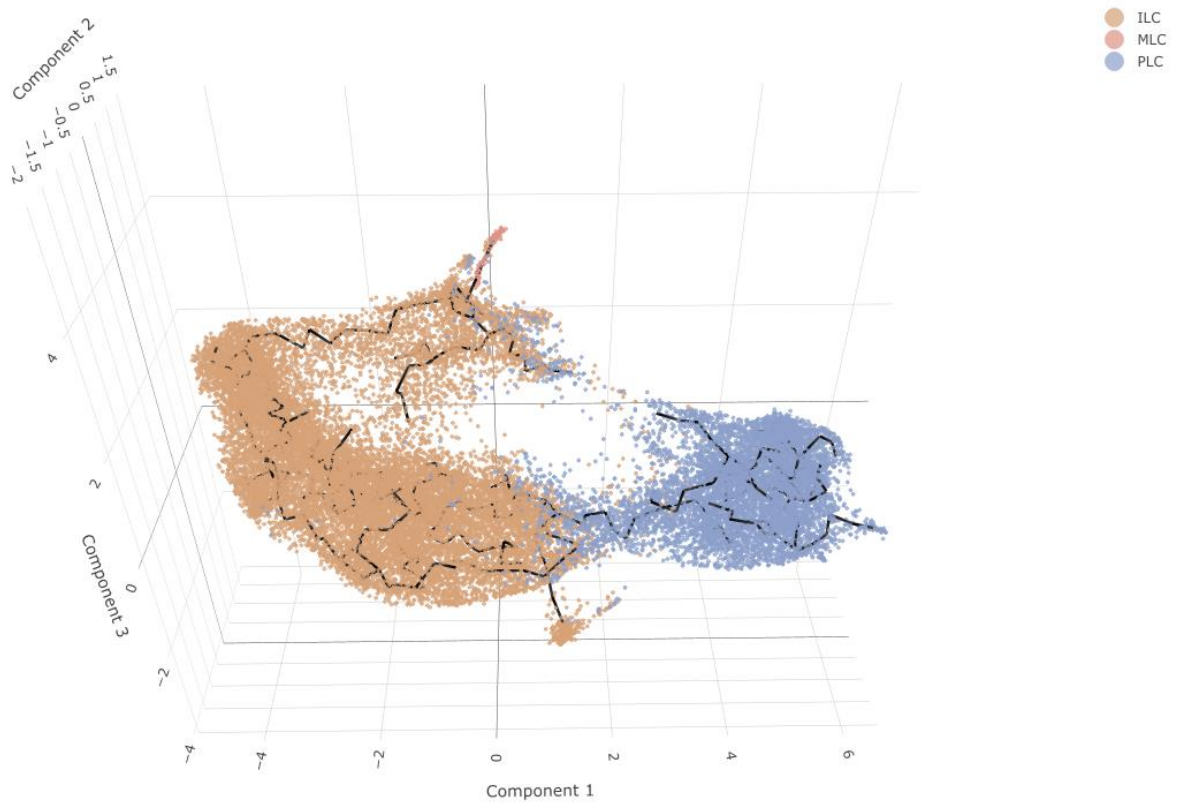


Figure S6. 3D reconstruction of pseudotime analysis for the integrated LC dataset. Note that the PLCs appear to encounter a bottleneck transitioning to ILCs. The MLC group is a small cluster with coordinates 0 (component 1) and between 2-4 (component 3), near the top center of the plot.

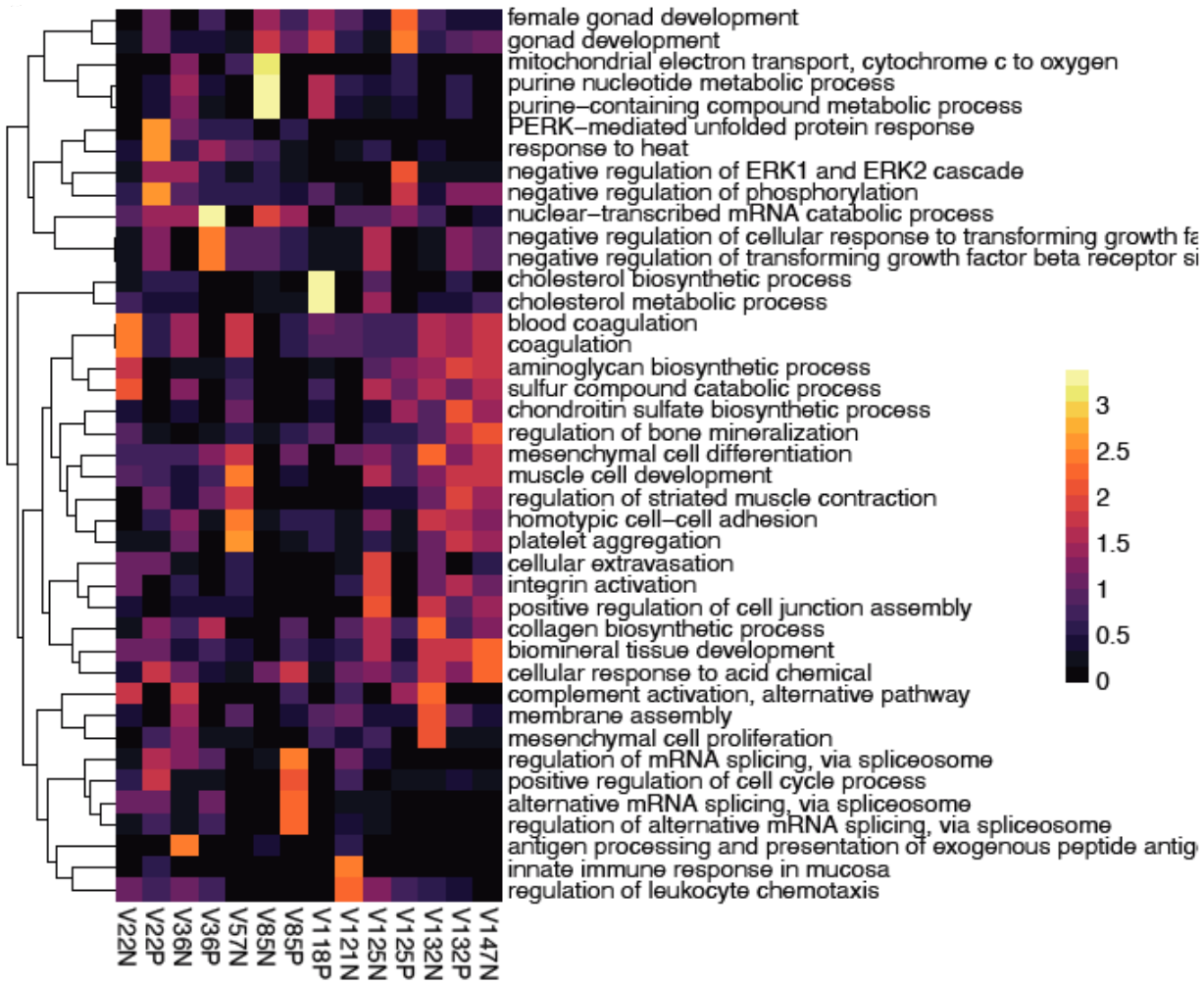
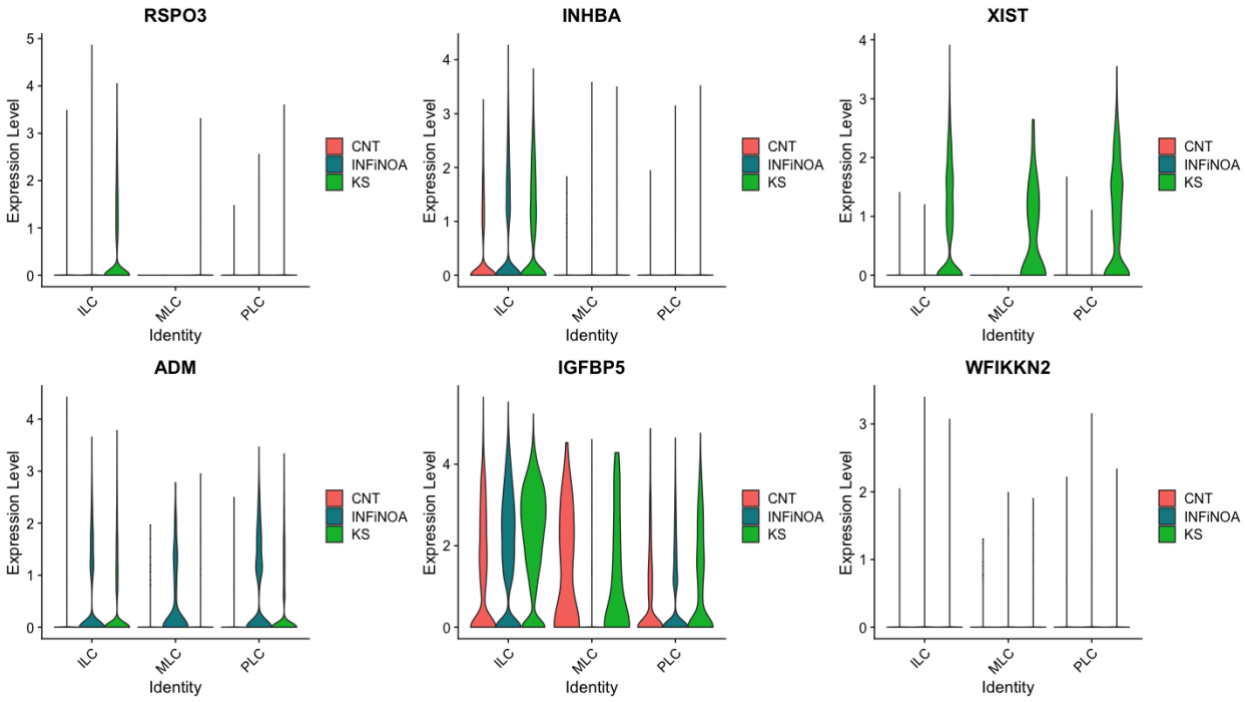


Figure S7. GO enrichment analysis of LC SDA components. Gene ontology enrichment of select SDA components indicated distinct functional enrichments among components. Scale is the log odds ratio of enrichment of SDA component genes compared to all genes in DGE. “N” and “P” suffixes on component names indicate gene sets with negative or positive component loadings, respectively.

A


RSPO3
INHBA
XIST
ADM
IGFBP5
WFIKKN2
Expression Level
Identity
ILC
MLC
PLC
CNT
INFiNOA
KS

B

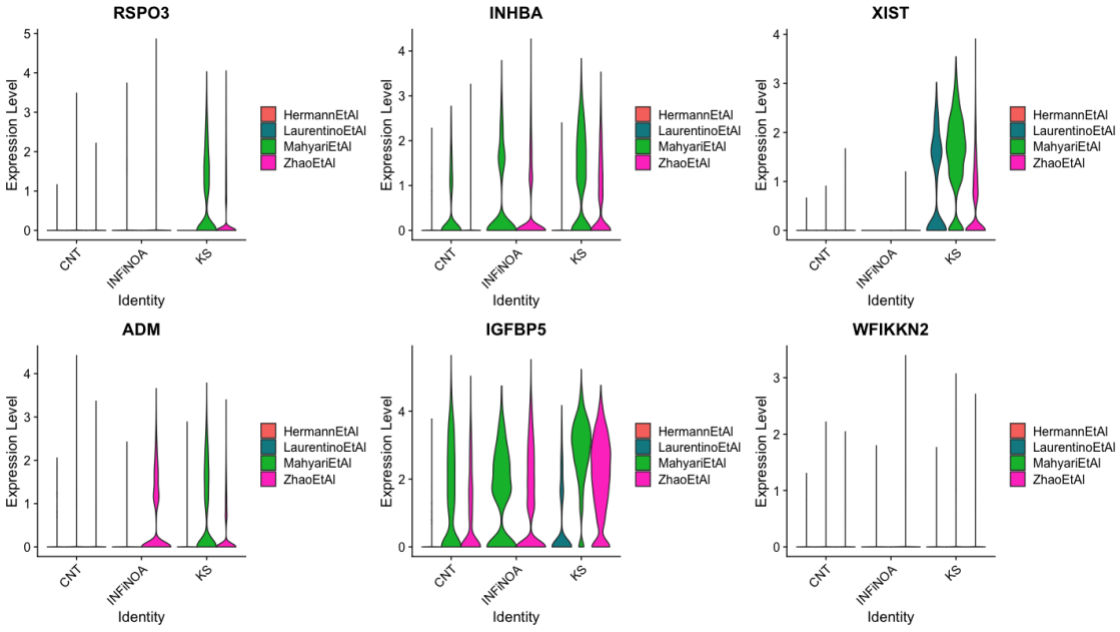

RSPO3
INHBA
XIST
ADM
IGFBP5
WFIKKN2
Expression Level
Identity
CNT
INFiNOA
KS
HermannEtAl
LaurentinoEtAl
MahyariEtAl
ZhaoEtAl

Figure S8. Replication of SDA36 expression changes using the integrated LC dataset. Genes with strong loading on SDA36 were interpreted as expression changes between KS and NOA LCs. As described in the main text, 70% (68%) of the top 100 positive (negative) genes loading on component 36 replicate as differentially expressed in the integrated dataset. Here we plot selected genes from SDA36, including *XIST* and 5 secreted signaling factors, (A) by LC subset and (B) by study of origin. These genes were all inferred to have higher expression in KS compared to NOA in the original SDA analysis, and all of them but *ADM* replicate in the integrated data (but note that *ADM* is still inferred to be upregulated in KS compared to controls, it just no longer appears KS-specific).

Figure S9. GO category enrichment trees for genes identified as differentially expressed between KS and CNT SCs. (A) Cellular compartment categories (B) Biological process categories (C) Molecular function categories.

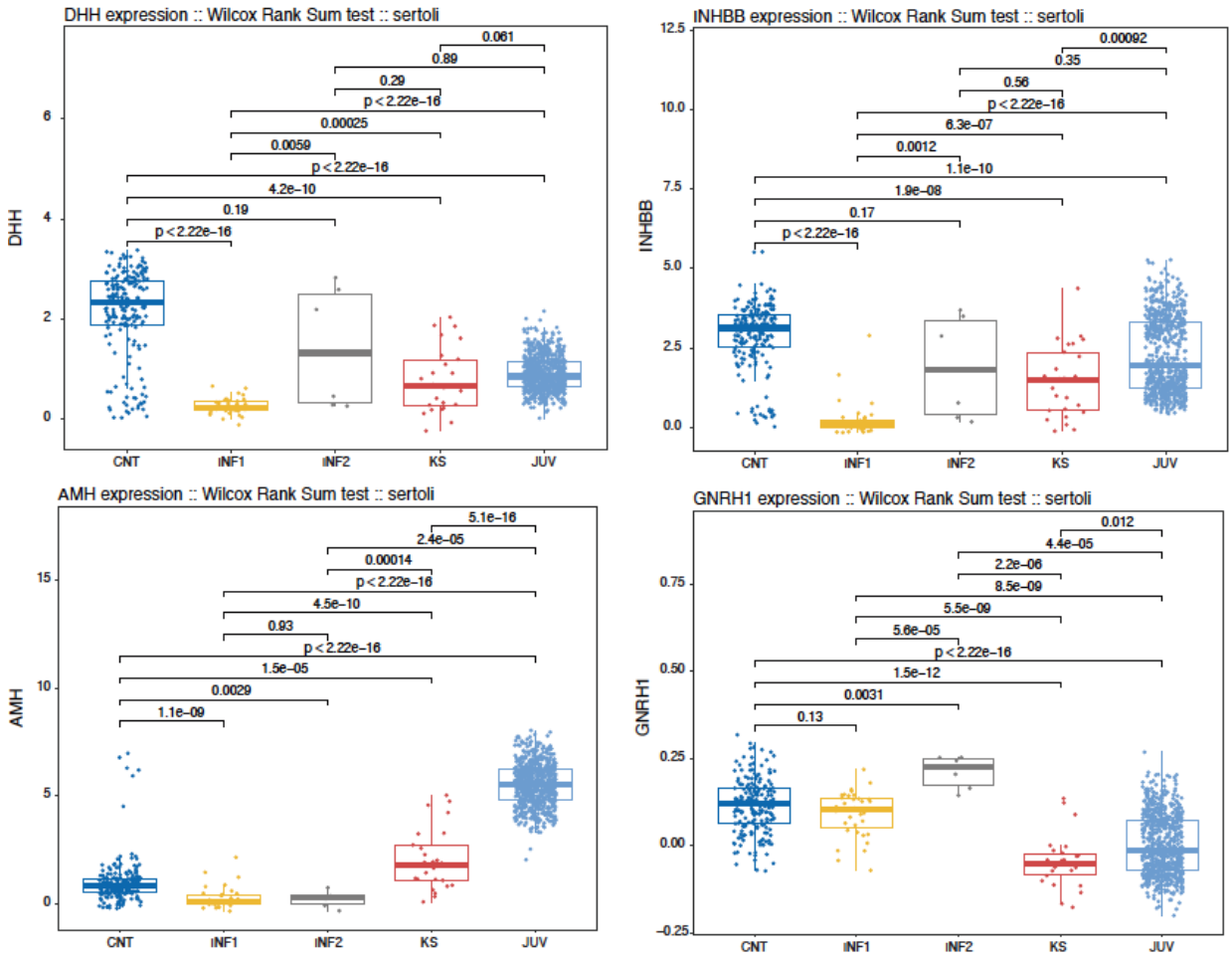


Figure S10. Expression levels of signaling factors in Sertoli Cells identified from our original “SDA” analysis. Box and whisker plots indicate single cell expression levels for four different genes across five different donor types. Expression levels were compared using Wilcoxon rank sum test, and p-values for each pairwise comparison are indicated on the panels. The observed changes in *DHH* and *INHBB* were replicated in the larger “integrated” dataset, while *AMH* did not appear significantly different from CNT. *GNRH1* was not reported in the Zhao et al. and thus we could not attempt replication.

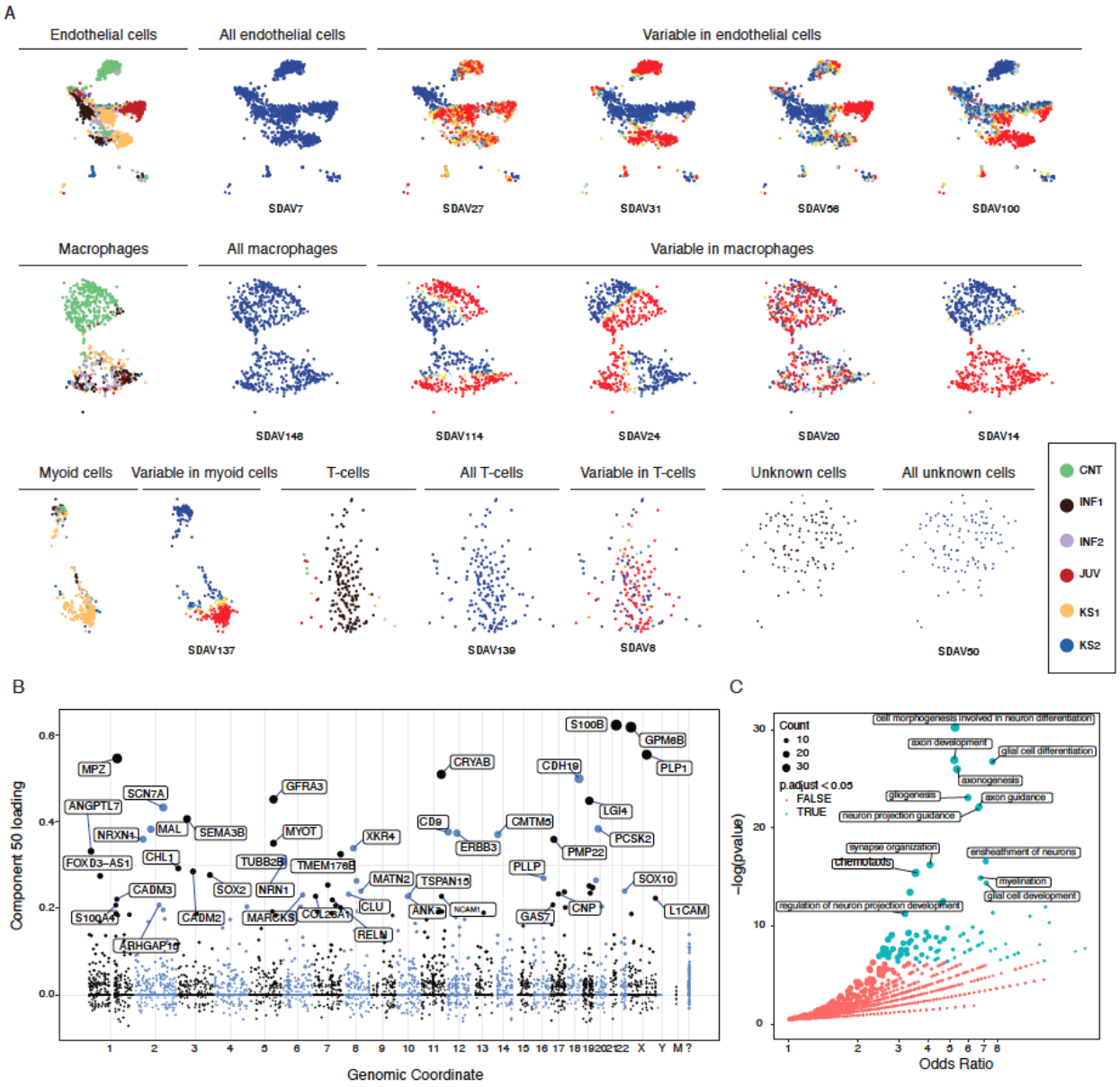


Figure S11. Summary of additional somatic SDA components. Overview of SDA components with cell scores mapping primarily to somatic cell types other than Leydig or Sertoli cells. Presentation as in Figure 5. For each cell type, the first tSNE plot is colored by donor (legend at the end of the subpanel). Each subsequent tSNE plot for a cell type corresponds to a single SDA component, and is colored by the cell score for that component (red= cell scores > 0, blue = cells scores < 0). Component numbers are labeled at the bottom of each tSNE. B) Gene loadings for SDA component 50, corresponding to an unknown population. C) Gene Ontology (GO) enrichments for the top 150 gene loadings of SDA component 50. Blue points correspond to GO categories with significant enrichments after multiple test correction.

Supplementary Note: Abridged SDA component annotations

A full description per component can be found in the complementary Excel file.

Components enriched and/or specific to interest given conditions:

Juv:

LC:

- SDA5: LC: Common pattern of scoring (+/-) between KS, INF1, INF2, and majority of CNT. But JUV only negatively scored. Suggests negative scoring is immature LC, vs positive as mature.
- SDA89: LC. JUV highest pos scored. INF1, INF2 and KS show variance higher than CNT. suggest mature vs immature LC.
- SDA90: LC. JUV highest pos scored. INF1, INF2 and KS show variance higher than CNT. suggest mature vs immature LC.

EC:

- SDA31: EC: CNT are mainly positively scored, INF1, INF2, and KS are scored both directions, but JUV is uniquely negatively scored.
- SDA56: EC. JUV uniquely score positive, whereas INF1, INF2, and KS are negative, CNT are near zero.

SC:

- SDA73: SC. JUV SC are uniquely scored differently than SC from other conditions.
- SDA122: SC. JUV have cells that score +/- . SC and LC score here, but SC show most pronounced scoring specific to the clusters.
- SDA130: SC. SC clusters are all scored positive, most cells from JUV followed by controls and relatively infrequent positively scored cells from the other cases.

KS:

LC:

- SDA36: LC mainly. The majority of INF1 cells are scored negatively whereas the majority of KS and some of the JUV cells are scored positively, above the background seen from certain CNT samples. The top positive loading in this component belongs to XIST amongst other genes that functionally map to translation and secretion whereas the top negative loadings identify mainly regulation of immune and inflammatory responses as well as antigen processing and presentation.
- SDA82: LC mainly. KS2 and INF2 most highest scored, but variance common between cases and controls.
- SDA85: LC. KS, INF1 and some INF2 similarly scored (above bkg of CNT). JUV and KS score Neg more than CNT. Other somatic cell types also scored. suggest common pathology between KS1 and KS2 as well as INF1 and INF2.
- SDA125: LC. INF2 scores negative whereas the controls have cells scoring positive; infrequent negatively scored cells from JUV and KS also.

EC:

- SDA27: EC: scores two distinct EC, the positively scored are found in all cases and controls, but the negatively scored are mostly from KS, INF1, INF2, and JUV.
- SDA100: EC. KS1 positive KS2 like Juv is bidirectional whereas the controls negative

Myoid:

- SDA57: Myoid - LC. a common scoring between the myoid in general and LC of specifically KS.
- SDA137: Myoid: There are two rare myoid clusters, which score oppositely with this component. The positive cells are from KS whereas the negative from KS, INF1, INF2, and some of the controls.

Mix:

- SDA22: Mix Soma. High-turnover and Apoptosis. The cells that score positive in this component are the KS, JUV, and INF1 whereas the controls lack this positive cell scoring.
- SDA121: Mix Soma. KS2 specific.
- SDA145: Mix Soma. KS have cells that are most positively scored, but positive as well as negative cells from all cases and controls present.
- SDA147: Mix. LC are the majority of negatively scored cells, but the positively scored cells span all soma. These positively scored cells are only in KS, INF1, and JUV. The negative scored cells are from all the cases and controls but with different magnitudes. INF1 negative scored cells have the smallest magnitude similar to JUV.

Immune:

- SDA139: Adaptive IC. T cell markers in positive loaded genes. the negative loaded genes span many cell types, but GO analysis map to functions related to antigen processing and presentation, myeloid-leukocyte immunity and regulation, granulocyte/neutrophil activation and response, and leukocyte differentiation, proliferation and cell-cell adhesion..

INF1 & 2:

GC:

- SDA4: GC. identifies a banding of positive and negative scored germ cells, late meiosis post pachytene. INF2 has pos, not neg scored cells, CNT have both directions, but INF1 has neither.

Immune:

- SDA14: Myeloid IC. Distinguishes two myeloid M1 and M2 activation states. Interestingly, the positive scored cells in this component are the KS1/2 and INF1/2 samples whereas the negative scored cells are CNT. It is likely this differential explains the extra stress the testis system of cells is under, that forces a polarization of the macrophages to deal with the inflammation and wound healing common amongst the pathologies.

LC:

- SDA36: LC mainly. Idiopathic INF1; but KS opposite scoring.
- SDA118: LC. Rare cluster that is closest to LC, and has similar genes as LC. Mostly from INF1, but one of the controls also has a few cells scoring positive. INF2 and KS contribute even less cells.
- SDA132: LC. One specific cluster of LC cells scores positive as well as other cells in the other other LC clusters. INF1 and KS lack the negatively scored cells found in controls and INF2.

CNTs & Normal Biology:

GC:

- SDA1: GC. identifies the late stage spermatids where sperm chromatin condensation genes as well as other key genes known for spermatid development are in the top loaded genes.
- SDA12: GC. distinguishes early undifferentiated spermatogonia from JUV vs CNT.
- SDA18: GC. scores mainly spermatocytes in the last part of meiosis as they become round spermatids. The CNTs were highly scored, INF1 lacks this pattern and INF2 had very few positively scored cells.
- SDA19: GC. scores spermatocytes in the last part of meiosis in a banding pattern, which is represented by the CNT and INF2 but INF1 is not scored at all.
- SDA21: GC. Identifies spermatocytes in meiosis I. INF1 lacks this but INF2 has the signature.
- SDA29: GC. Entry to meiosis I by differentiating sg.
- SDA35: GC. Undifferentiated sg. subsets.
- SDA40: GC. Meiosis I, banding showing differential programming.
- SDA43: GC. Differentiating sg. banding showing differential genetic co-regulation.
- SDA46: GC. Differentiating sg. banding showing differential genetic co-regulation.
- SDA47: GC. Pre-pachytene spermatocytes banding.
- SDA49: GC: Round and elongating spermatids.
- SDA51: GC. Score spermatocytes highest found in early meiosis I, but sg are also scored positive.
- SDA52: GC: Scores a clear split between cells in later stages of meiosis I vs spermatids in the next stage.
- SDA59: GC. Banding in meiosis I pre-pachytene, negative top loadings has many lincRNA and piRNA associated transcripts.
- SDA67: GC. Enriched for DNA binding genes, highlighting differentiating spermatogonia.
- SDA68: GC. highlights the germ cells from the last stages of meiosis I and includes round and early elongating spermatids. The top loaded genes include those related to acrosomal vesicles .
- SDA71: GC. Banding of spermatocytes in meiosis post pachytene.
- SDA74: GC. Banding of spermatocytes in meiosis post pachytene
- SDA92: GC. Round and elongating spermatids are scored positive, negative elsewhere.
- SDA94: GC. Banding in meiosis I pre-pachytene and beyond.
- SDA98: GC. Banding in meiosis I post-pachytene.
- SDA102: GC. Elongated spermatids are scored positive, but the terminal end of the progression bands negative. The positive and negative scored GC are from CNTs and INF2. INF1 shows rare positively scoring cells.
- SDA108: GC. Undifferentiated spermatogonia are negatively scored from all expected cases and controls.
- SDA109: GC. Middle set of elongating spermatids scored negatively flanked both directions by positive banding. INF2 has these negatively scored cells but not scored as strongly as controls.
- SDA110: GC. Meiosis I post pachytene stage negatively scored.

- SDA112: GC. Banding in early undifferentiated spermatogonia.
- SDA126: GC. Undifferentiated spermatogonia banding.
- SDA127: GC. Enriched for genes regulating early cell fate decisions given stimuli, highlighting a specific segment of the undifferentiated spermatogonial cells. The top negatively loaded genes include those related to hormone secretion and regulation as well as several homeobox and transcription factor genes.
- SDA138: GC. Banding of spermatids from expected cases and controls.
- SDA141: GC. Meiosis post pachytene spermatocytes negatively scored.
- SDA142: GC. This component identifies cells from undifferentiated sg. To pre-pachytene meiosis I, picking signatures related to chromatin and nucleosome assembly as well as DNA packing; the homeobox genes enriched in the positive loadings, coregulated here with several histone packing genes.
- SDA143: GC. Soma is most negatively scored followed by spermatogonia. A small section of meiosis I, likely just pachytene spermatocytes are scored positive. The negative loaded genes are enriched with ribosomal proteins.
- SDA146: GC. Banding in elongating spermatids. +/- scored cells from controls and INF2.
- SDA149: GC. Banding of undifferentiated and early differentiating spermatogonia.

Immune:

- SDA7: Myeloid IC. identifies Immune-myeloid as well as endothelial cells and the gene loadings suggests interactions between these cell times as well as other regulatory processes like wound healing.
- SDA8: Adaptive IC. identifies effector T cells vs regulator or memory-like state.
- SDA20: Myeloid IC. identifies differences between macrophage responses where the CNTs and INF2 are scored similarly but INF1 and KS1 and KS2 are scored opposite.
- SDA24: Myeloid IC: finds two different myeloid states.
- SDA114: IC. Myeloid and T cells +/- scored. GO annotation suggests that positive loaded genes are related to regulation of innate immunity whereas negative loaded genes suggest T cell activation and and response.

Mix:

- SDA13: Mix. scores several somatic and germ cells. The most highly scored cells are the early germ cells, followed by somatic cells from the rest of the donors. Only the CNT and INF2 retain non-scoring cells, suggesting a common pattern between them.
- SDA26: Mix. The positive scored cells are unique to CNT and undiff. sg. whereas the negatively score cells are mix of somatic and differentiating sg. cells.
- SDA54: Mix. GC up to mid-meiosis I are most negatively scored.
- SDA58: Mix. Mitochondrial gene expression.
- SDA70: Mix. The most negatively scored are the spermatogonia germ cells and LC, but other cell types are also scored negatively.
- SDA80: Mix. Every cell type is involved, but KS1 are scored most negative, possibly idiopathic.
- SDA103: Mix. Early germ cells are scored negatively whereas soma is positive.
- SDA148: Mix Soma. Immune cells (myeloid and t cells) are negatively scored whereas endothelial cells are positively scored both from all the cases and controls; highlighting endo-immune connection.

Unknown:

- SDA50: Unknon. This component scores a small cluster of cells which functionally map to neuronal functions.

Leydig:

- SDA111: LC. All cases and controls show +/- scored cells. INF1 and KS highest variance.