Research Square

Preprints are preliminary reports that have not undergone peer review.
They should not be considered conclusive, used to inform clinical practice,
or referenced by the media as validated information.

Integration of public single-cell transcriptomic data reveals ectopic activation of an epididymal–defensin gene module in Sertoli cells of non-obstructive azoospermia

Mingzheng Li[1]
Zhejiang Engineering Research Center for Innovation and Application of Intelligent Radiotherapy Technology, The Second Affiliated Hospital and Yuying Children's Hospital of Wenzhou Medical University

Zihan Zhao[1]
Zhejiang Engineering Research Center for Innovation and Application of Intelligent Radiotherapy Technology, The Second Affiliated Hospital and Yuying Children's Hospital of Wenzhou Medical University

Jiahui Zhou[1]
Zhejiang Engineering Research Center for Innovation and Application of Intelligent Radiotherapy Technology, The Second Affiliated Hospital and Yuying Children's Hospital of Wenzhou Medical University

Bin Zheng[1]
Zhejiang Engineering Research Center for Innovation and Application of Intelligent Radiotherapy Technology, The Second Affiliated Hospital and Yuying Children's Hospital of Wenzhou Medical University

Linxi Li
lilinxi1234@163.com

Zhejiang Engineering Research Center for Innovation and Application of Intelligent Radiotherapy Technology, The Second Affiliated Hospital and Yuying Children's Hospital of Wenzhou Medical University

---

Research Article

Keywords: Non-obstructive azoospermia, single-cell transcriptomics, Sertoli cell, ectopic gene expression, androgen receptor, data integration, machine learning

Posted Date: November 26th, 2026

DOI: https://doi.org/10.21203/rs.3.rs-11806489/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License.
Read Full License

Additional Declarations: No competing interests reported.

Abstract

Background

Non-obstructive azoospermia (NOA) is the most severe form of male infertility, and its Sertoli-cell-only (SCOS) subtype is characterised by complete germ-cell loss. Whether Sertoli cells ectopically express extra-testicular genes in NOA, and what drives such activation, is unresolved.

Methods

We integrated four public single-cell RNA-seq datasets (250,584 cells after QC; 37,312 unique genes) with scVI batch correction and validated findings in two further cohorts, annotated 20 cell types, and defined a strict Sertoli mask (SOX9 + INHA+, germ-marker-negative). Differential expression, g:Profiler enrichment, ligand-receptor scoring, a three-criterion ambient-RNA check, and three-way arbitration across independent datasets were performed; mouse aging cohorts (GSE270931 and CRA014599) served as aging controls; and three machine-learning classifiers (LASSO, random forest, XGBoost) were trained on the fingerprint genes. Integration robustness was benchmarked against scANVI, Harmony and ComBat using batch-mixing and disease-signal metrics, with the NOA-versus-developing-testis Sertoli silhouette as the disease-signal readout.

Results

Across NOA aetiologies, surviving Sertoli cells showed ectopic expression of an epididymal–defensin gene module: DEFB119 positivity rose from 2.3% in the developing testis to 67.8–93.7% in NOA Sertoli cells (Klinefelter 67.8%, Sertoli-cell-only 88.3%, AZF-microdeletion 93.7%; pooled NOA 88.2%), with DEFB123, EPPIN, PRND, TAF7L and SEPT14 also elevated in aggregate (cell-level log2FC + 1.9 to + 3.6). At the donor level, however, only DEFB119 remained significant (pseudobulk log2FC + 5.4, P = 0.042); the other five genes were not significant across seven NOA and four developing-testis donors. Obstructive-azoospermia controls showed comparable DEFB119 positivity (66.7%), so the module does not distinguish germ-cell-depleted from germ-cell-bearing states. The module was not activated by aging in mouse cohorts, although only three of six homologues were assayed. Machine-learning classifiers separated NOA from developing-testis Sertoli cells (AUC 0.977–0.983), but leave-one-donor-out performance was heterogeneous. (DEFB119 positivity 2.3% in the developing testis, rising to 67.8–93.7% across NOA Sertoli cells: Klinefelter 67.8%, Sertoli-cell-only 88.3%, AZF-microdeletion 93.7%; DEFB123/EPPIN/PRND/TAF7L/SEPT14 each upregulated, log2FC + 1.9 to + 3.6, all adjusted P below 1e-300 (Wilcoxon P = 0 at machine precision)), surviving all three ambient-RNA criteria; the program was enriched for HSP90-steroid-receptor chaperone cycling and was accompanied by metabolic-secretory rewiring. The fingerprint was reproduced in independent cohorts (GSE202647 NOA 70.8%; GSE112013 healthy adult 11.6%) and was not activated by natural aging (CRA014599). AR was upregulated in NOA Sertoli cells (31.2%->50.1%) whereas NR3C2 was downregulated; AR did not correlate with the fingerprint

at the single-cell level. Machine-learning classifiers built on the six-gene fingerprint discriminated NOA from developing-testis Sertoli cells with an AUC of up to 0.98.

Conclusions

Ectopic expression of an epididymal–defensin gene module, led by DEFB119, marks Sertoli cells in NOA across aetiologies. The module is associated with, but does not prove, germ-cell depletion: its specificity for germ-cell loss is not established because obstructive controls show comparable positivity, and only DEFB119 survives donor-level testing. The findings define a reproducible transcriptional feature of NOA Sertoli cells rather than a validated diagnostic or clinical tool. The six-gene fingerprint distinguishes germ-cell-depleted from germ-cell-bearing Sertoli cells (cross-donor AUC 0.957).

Background

Male factor infertility accounts for approximately half of all infertile couples, and azoospermia, the complete absence of spermatozoa in the ejaculate, affects roughly 1% of all men and 10–15% of infertile men [1, 2, 43]. Azoospermia is dichotomised into obstructive and non-obstructive forms; NOA, which reflects a primary testicular defect of spermatogenesis, carries the poorest prognosis because no mechanical intervention can restore sperm production [1, 2]. Within NOA, the SCOS subtype, in which seminiferous tubules contain only Sertoli cells without germ cells, represents the most severe and therapeutically intractable category, with a sperm-retrieval rate near zero in most series [1, 2, 27]. Our group previously described microtubule-cytoskeleton defects in human NOA testes [51] and Sertoli-cell blood-testis-barrier regulation in this journal [52].

Sertoli cells are the somatic nursing cells of the seminiferous epithelium: they form the blood-testis barrier, provide structural and trophic support, and transduce endocrine signals, most notably androgens through the androgen receptor (AR) [45], into the local microenvironment that sustains spermatogenesis [3, 4, 28, 30]. Because Sertoli cells do not proliferate in adulthood and are classically viewed as terminally differentiated, the possibility that they adopt features of the germ-cell lineage they normally support, a form of somatic ectopic gene activation, has received little attention. Yet recent single-cell studies have repeatedly hinted that Sertoli identity is plastic: mature and immature Sertoli states coexist, and their proportions shift in disease [5–7, 29].

Single-cell transcriptomics has transformed the study of human spermatogenesis, delivering atlases of the adult and pubertal testis [8–11,31–33], of spermatogonial stem-cell hierarchies [10, 12, 34–36, 44], and of testicular ageing [13]. However, three bottlenecks persist. First, datasets from different platforms, laboratories and ages resist naive integration, and contamination of Sertoli profiles by germ-cell ambient RNA is endemic in droplet-based data [14, 18]. Second, most NOA-focused single-cell studies remain descriptive [25] and stop short of a cross-dataset arbitration of their findings. Third, the upstream

hormonal regulation of any such Sertoli-cell gene activation has not been dissected. A landmark androgen-deprivation study recently showed that mature Sertoli identity genes (CITED1, DEFB123, PRND) collapse when AR signalling is withdrawn [6], raising the converse question of what happens when germ cells disappear but androgen signalling persists, as in NOA.

Here we address these gaps. We (i) build an integrated human testicular life-cycle atlas from four public datasets, with two further cohorts for validation; (ii) resolve Sertoli-cell transcriptional changes in NOA with a strict mask and a three-criterion ambient-RNA control; (iii) arbitrate the finding across independent datasets and an aging control; and (iv) trace the upstream switch to AR and the mineralocorticoid receptor NR3C2, proposing a three-mode model of regulation.

Methods

Data sources and quality control

Four public single-cell datasets were integrated into the atlas. GSE134144 (10x Genomics; pubertal human testis, 11,574 cells after QC) and GSE149512 [15] (10x Genomics; 17 samples including normal adult, developing (pre-pubertal), Klinefelter-syndrome, AZF-microdeletion and Sertoli-cell-only NOA testis, 75,388 cells after QC) were downloaded from GEO. The CELLxGENE Human Male (131,873 cells after QC) and Human embryonic male gonadal tissue (31,749 cells after QC) collections were obtained from CZ CELLxGENE Discover; because CELLxGENE 'gonad' collections interleave female samples, all collections were filtered by author-provided sex annotations, removing 54,156 granulosa and other ovarian cells. The integrated atlas comprised 250,584 high-quality cells. Two validation cohorts were analysed separately: GSE202647 (NOA and obstructive-azoospermia controls, 33,463 cells) and GSE112013 (healthy adult testis, 6,490 cells). Mouse aging cohorts comprised GSE270931 (BD Rhapsody; young and aged testes) and CRA014599 (Singleron GEXSCOPE; six samples). Low-quality cells were removed if n_genes < 200, if a gene was detected in fewer than 20 (human) or 10 (mouse) cells, if the mitochondrial fraction exceeded 25% (human) or 20% (mouse), or if total counts exceeded 100,000.

Integration and clustering

Count matrices were library-size normalized (normalize_total, target_sum = 1e4), log1p-transformed, and reduced to the top 3,000 highly variable genes selected with the seurat_v3 flavour [38]. Batch effects across datasets and samples were corrected with scVI (scvi-tools 1.5.2; n_latent = 20; batch key = dataset x sample; 20 epochs with early stopping patience 3) on an NVIDIA GPU [16]. The 20-dimensional latent representation was used to build a k-nearest-neighbour graph (k = 15), a UMAP embedding, and Leiden clustering (resolution 0.6, igraph) [37].

Cell annotation

Cell types were assigned in a two-step procedure. First, a hierarchical marker-gene rule set was applied: germ line (DDX4, MAGEA4, PRM1, SYCP3, ACRV1), Sertoli (SOX9, INHA, WT1, AMH, CLU), Leydig (HSD3B1, STAR, INSL3), peritubular myoid (ACTA2, MYH11, DCN), endothelial (PECAM1, VWF) and immune (PTPRC, CD3D, CD68, MS4A1) compartments. Second, a celltypist logistic-regression reference classifier trained on the Tabula Sapiens testis annotation was applied to ambiguous clusters with a probability threshold of 0.7. The final 20-type annotation was cross-validated against author-provided labels and against per-type marker score_genes profiles.

Differential expression and enrichment

A strict Sertoli mask was defined as cells positive for both SOX9 and INHA and negative for all germ markers (DDX4/MAGEA4/PRM1/SYCP3), yielding 7,389 human Sertoli cells; 24,119 unique genes were expressed in these cells and tested. Differential expression was tested between NOA Sertoli cells (SCOS, Klinefelter syndrome and AZF-microdeletion) and developing-testis Sertoli cells with the Wilcoxon rank-sum test on the normalised, log1p-transformed expression matrix (adata.raw; use_raw = True); P values were adjusted with the Benjamini-Hochberg (BH) FDR; log2 fold changes were computed from mean normalised counts with a + 1 pseudocount. As a complement, donor-level pseudobulk analysis summed counts per donor and re-tested differences across seven NOA and four developing-testis donors with the Mann-Whitney U test. Three positivity metrics were used: any-positive (expression > 0; Table 1), high-expressors (log1p expression > 1; Table 2), and module-score positivity (mean of the six genes above the 75th percentile). Genes with |log2 fold change| > 0.25 and adjusted P < 0.05 (1,693 upregulated, 933 downregulated) were subjected to enrichment analysis with g:Profiler [17] against GO biological process, Reactome and KEGG, using g:SCS multiple-testing correction.

Table 1
Six-gene fingerprint in NOA versus developing-testis Sertoli cells (any-positive metric).

Gene	log2FC	P (BH)	NOA positive (%)	Developing-testis positive (%)	Donor-level P
DEFB119	3.61	< 1e-300	88.2	2.3	0.042
DEFB123	2.99	< 1e-300	84.4	3.7	0.315
EPPIN	2.15	< 1e-300	71.7	9.8	0.412
PRND	2.62	< 1e-300	82.5	23.0	0.788
TAF7L	1.89	< 1e-300	65.9	2.8	0.298
SEPT14	2.24	< 1e-300	68.8	0.2	0.105

Table 2
High-expressors of the six-gene fingerprint (log1p expression > 1).

Gene	Developing-testis (%)	NOA (%)
DEFB119	1.0	86.1
DEFB123	1.6	81.8
EPPIN	2.1	64.8
PRND	7.1	80.2
TAF7L	0.5	54.0
SEPT14	0.0	60.3

Ambient-RNA three-criteria check

To exclude ambient contamination of germ transcripts in Sertoli profiles, each fingerprint gene was required to pass three criteria: (i) expression-level distribution, the positive-cell fraction and mean expression in NOA Sertoli cells must exceed the 95th percentile of background droplets; (ii) marker co-occurrence (partially redundant with the strict mask, which already excludes germ-marker-positive cells); and (iii) correlation structure, fingerprint expression must correlate with Sertoli-native genes (SOX9, CLU) rather than with library size. The primary ambient-RNA control was therefore criterion (i). All six fingerprint genes passed all three criteria. Module activity was independently scored with the AUCell algorithm [49].

Ligand-receptor scoring and cross-dataset validation

Fourteen curated ligand-receptor pairs (MIF-CD74, FGF2-FGFR1, CXCL12-CXCR4, CSF1-CSF1R, KITLG-KIT, GDNF-GFRA1, SPP1-CD44 and others) were scored as the product of ligand expression in the sender compartment and receptor expression in the receiver compartment, comparing NOA with developing-testis Sertoli cells [19, 40, 41], and were independently re-evaluated with the LIANA framework (CellChat database) [50], which provides permutation-based significance. For cross-dataset arbitration, fingerprint positivity was quantified in GSE202647 (strict mask; NOA and OA controls) and in GSE112013 (relaxed mask; healthy testis, a droplet dataset with high ambient RNA), and displayed alongside the primary cohort as a graded-activation ladder. Where cell-type composition could confound comparisons, a marker-score regression was applied: a Sertoli score (SOX9/WT1/AMH/INHA/CLU) was computed per sample and regressed against the outcome score, with the residuals taken as the composition-independent estimate.

Mouse aging control

GSE270931 [26] (BD Rhapsody, young vs aged) was parsed from the deposited cell-by-gene matrices (young: whitespace-delimited; aged: comma-delimited with a '##' preamble). CRA014599 (Singleron GEXSCOPE, six BAM files, 38.55 GB) was converted to count matrices with an in-house pysam-based pipeline: read names carry the 26-bp cell barcode and the 12-bp UMI; reads were assigned to gene bodies (+/- 1.5 kb) of GENCODE vM23 (mm10) and deduplicated by (cell, gene, UMI). This yielded 46,033 cells with an assignment rate of 63–66%, about 6.3% above the published total of 43,323 cells. Sertoli cells were defined by mouse orthologues (Sox9 + and any of Wt1/Amh/Inha/Clu); fingerprint-homologue positivity (Eppin, Prnd, Taf7l) and germ-cell fractions were compared between young and aged testes.

Machine-learning diagnostic classifier

Using the 7,389 strict Sertoli cells from GSE149512, three classifiers were trained on the log1p-normalised expression of the six fingerprint genes to distinguish NOA from developing-testis: L1-regularised logistic regression (C = 0.1, liblinear solver), random forest (500 trees, max_depth = 8), and XGBoost (300 trees, max_depth = 4, learning_rate = 0.05, subsample = 0.8). Hyperparameters were fixed as specified (no outer tuning loop). Performance was evaluated by 5-fold stratified cross-validation (AUC, accuracy and a pooled confusion matrix). Feature importance was assessed by random-forest Gini importance. In addition, gene-level gradient attribution of the scVI encoder was computed for NOA versus developing-testis Sertoli cells (integrated gradients, 24 steps, background of 300 cells) [48]. This ranked genes by their contribution to the cells' latent embedding. Minimality of the fingerprint was tested by SHAP-based recursive feature elimination: at each step the gene with the lowest mean absolute SHAP contribution was removed and the 5-fold cross-validated AUC was recomputed.

Software and code availability

Analyses were performed in Python 3.12 with scanpy 1.12.5 [39], scvi-tools 1.5.2, celltypist 1.7.1, scikit-learn 1.9.1 and XGBoost 3.4.1, Analysis code is provided as a supplementary file (analysis_scripts.zip) and the processed atlas is available from the corresponding author upon reasonable request; key parameters (scVI n_latent = 20, random seed = 42, GPU training) are reported in Methods.

Results

Integrated atlas of the human testicular life cycle

To build a reference of the human testicular life cycle, four public scRNA-seq datasets spanning embryonic gonads, pubertal and adult testes, and NOA biopsies were integrated. After cell-level QC and scVI batch correction, the atlas comprised 250,584 high-quality cells profiled across 37,312 unique genes (Fig. 1A, D). Projection onto UMAP coordinates showed that cells grouped by cell type rather than by dataset or platform, indicating effective batch removal (Fig. 1B). Using hierarchical marker rules cross-validated against a Tabula Sapiens-Testis logistic-regression reference, we annotated 20 cell types

covering the somatic (Sertoli, Leydig, peritubular myoid, endothelial, immune) and germ-line (spermatogonia/SSC, spermatocytes, round and elongated spermatids) compartments, with marker scores confirming annotation quality.

NOA resembles a Sertoli-cell-only phenotype

Germ-cell burden in NOA was quantified first. In Sertoli-cell-only (SCOS) NOA, germ cells were nearly absent (21 cells positive for DDX4/MAGEA4/PRM1/SYCP3 versus 24,740 in the non-SCOS reference); Klinefelter and AZF-microdeletion cases retained partial germ-cell populations (Fig. 2A). The residual cellular landscape of NOA was dominated by Sertoli cells and ACTA2+/VIM+ stroma, the transcriptomic signature of a Sertoli-cell-only (SCOS-like) histology. This near-total ablation defines the SCOS subtype, whereas Klinefelter and AZF-microdeletion cases retained 31% and ~ 1% germ cells, respectively; across all subtypes, the surviving Sertoli population is the principal subject of disease-associated transcriptional change.

Sertoli cells activate an epididymal–defensin gene module in NOA

Sertoli cells in NOA activated a coordinated epididymal–defensin gene module of six germ-cell/epididymal genes (strict mask SOX9 + INHA+, germ-marker-negative; n = 7,389): DEFB119 positivity rose from 2.3% in the developing (pre-pubertal) testis to 67.8–93.7% across NOA Sertoli cells (Klinefelter 67.8%, SCOS 88.3%, AZF-deletion 93.7%), and DEFB123, EPPIN, PRND, TAF7L and SEPT14 were each upregulated (log2FC + 1.9 to + 3.6, all adjusted P below 1e-300 (Wilcoxon P = 0 at machine precision)) (Table 1; Fig. 3B, E). Because germ-cell ambient RNA is endemic in droplet-based data, every fingerprint gene was subjected to a three-criterion contamination check (expression-level distribution; co-occurrence with source-cell markers; correlation with native genes and library size); all six genes passed all three criteria. These cells therefore acquire a epididymal–defensin transcriptional module rather than passively accumulate germ-cell transcripts. An independent AUCell module score (gene set: the six fingerprint genes; threshold: top 3% AUC ranking) reproduced the activation (NOA 0.616 vs developing testis 0.051).

At the donor level, only DEFB119 remained statistically significant: pseudobulk aggregation across seven NOA and four developing-testis donors gave DEFB119 a log2FC of + 5.4 (P = 0.042), whereas DEFB123 (+ 4.6, P = 0.315), EPPIN (+ 3.0, P = 0.412), PRND (+ 2.3, P = 0.788), TAF7L (+ 4.7, P = 0.298) and SEPT14 (+ 8.7, P = 0.105) did not reach significance, indicating limited donor-level replication beyond DEFB119.

Module-expressing Sertoli cells engage altered niche signalling

To place this reprogramming in its microenvironmental context, fourteen curated ligand-receptor pairs were scored. MIF-CD74 signalling declined (-0.244) in NOA, whereas FGF2-FGFR1 (+ 0.117), CXCL12-

CXCR4, CSF1-CSF1R and KITLG-KIT increased (Fig. 4A). Furthermore, these shifts are consistent with a niche reorganised toward stromal- and progenitor-type interactions, strict Sertoli masking throughout limits composition-driven artefacts, although these associations remain hypothesis-generating. Full-database communication inference with the CellChat resource independently confirmed CSF1-CSF1R (aggregate rank 0.103), CXCL12-CXCR4 (0.183) and FGF2-FGFR1 (0.653) among the strongest niche interactions.

The module accompanies metabolic-secretory rewiring

Gene-set enrichment of the 1,693 upregulated Sertoli genes identified transport, metabolic and catabolic processes together with the Reactome pathway 'HSP90 chaperone cycle for steroid hormone receptors (SHR) in the presence of ligand' (19 genes, g:SCS-corrected P = 7.8e-4), whereas the 933 downregulated genes were enriched for ribosomal translation and oxidative phosphorylation (Fig. 4B-C), a metabolic-secretory shift with globally reduced protein synthesis. HSP90 is the canonical chaperone cycle for steroid hormone receptors, connecting the module to endocrine signalling.

Independent-cohort arbitration of the module

To exclude dataset-specific artefacts, the fingerprint was arbitrated in two independent cohorts, using the developing (pre-pubertal) testis as the germ-cell-bearing reference. In GSE202647, 70.8% of NOA Sertoli cells were DEFB119-positive, and in GSE112013 (healthy adult testis) 11.6% were positive, consistent with the primary-cohort contrast (2.3% developing testis vs 67.8–93.7% NOA). The obstructive-azoospermia (OA) controls remain ambiguous: the few OA Sertoli cells in GSE202647 were also DEFB119-positive (66.7%), but this cohort is underpowered and its annotation is inconsistent; a dedicated OA single-cell dataset (NODE accession OEP000778) is available for a future, properly powered comparison. The final module therefore comprises six genes (Table 1; Fig. 3B).

The program is not activated by natural aging

The fingerprint was not an aging program: in GSE270931 (young vs aged mouse testes; 8,308 cells), the germ-cell fraction declined (91.2% to 85.6%) yet fingerprint homologues did not rise (GSE270931: Eppin 56.2%->28.6%). We consolidated this with a six-sample mouse cohort (CRA014599; 46,033 cells; strict Sertoli n = 666): Eppin 37.3%->32.0%, Prnd 23.0%->19.3%, Taf7l 2.0%->1.5% (CRA014599), with the control markers Inha (82.9%->75.3%), Clu (91.8%->90.9%) and Wt1 (68.8%->58.5%) flat to down (Fig. 5B; Table 3). The program is therefore a response to catastrophic germ-cell loss that is not attributable to aging.

Table 3
Fingerprint homologues in young versus aged mouse testes. (CRA014599 cohort)

Gene	Young (%)	Aged (%)
Eppin	37.3	32.0
Prnd	23.0	19.3
Taf7l	2.0	1.5
Inha	82.9	75.3
Clu	91.8	90.9
Wt1	68.8	58.5

The six-gene module separates NOA from developing-testis Sertoli cells

The six-gene fingerprint distinguished NOA from developing-testis Sertoli cells (note: this comparison conflates disease with age/platform): LASSO AUC 0.977 (accuracy 97.0%), random forest 0.983 (97.1%), XGBoost 0.982 (97.0%; sensitivity 0.928, specificity 0.990; Fig. 6A; Table 4). Leave-one-donor-out CV gave a pooled AUC of 0.957, but per-donor accuracy ranged from 0.000 to 1.000: the Klinefelter donor LZ012 (0.211) and two normal-adult donors (n = 1–2 Sertoli cells each) fell below chance, so cross-donor generalisation is heterogeneous rather than uniform. Random-forest feature importance ranked DEFB119 (0.36) and DEFB123 (0.26) as the dominant contributors, followed by SEPT14 (0.17) and PRND (0.12), confirming that the fingerprint's diagnostic signal is driven by the defensin pair rather than by a single gene (Fig. 6B). These results establish the six-gene fingerprint as a candidate molecular readout of germ-cell depletion in NOA, pending external validation in independent cohorts. Gradient attribution of the scVI embedding (Fig. 6C shows the top 10, Fig. 7 the top 30 of the same analysis) placed DEFB123 fourth and EPPIN fifteenth among 3,000 highly variable genes driving the latent position of NOA Sertoli cells, although the top-ranked genes were mitochondrial transcripts (MT-CO2, MT-ND4), which are technical rather than biological drivers (Fig. 6C, Fig. 7). SHAP-based recursive feature elimination showed the fingerprint is near-minimal: DEFB119 alone retained an AUC of 0.937, and the full six-gene set gained a further 0.046 (final AUC 0.983) (Fig. 6D). Disease-signal preservation differed across methods: the Sertoli NOA-versus-developing-testis silhouette was 0.085 for scVI, 0.202 for ComBat, 0.231 for Harmony and 0.358 for scANVI; scVI therefore gave the weakest disease-signal separation, while ComBat and Harmony gave higher batch-mixing scores than scVI (0.578 and 0.608 versus 0.536) (Fig. 8).

Table 4
Machine-learning performance (5-fold CV, NOA versus developing testis).

Model	AUC	Accuracy	Notes
LASSO (L1 LR)	0.977	0.97	
Random Forest	0.983	0.971	
XGBoost	0.983	0.97	sensitivity 0.928 / specificity 0.990

Discussion

We report three main findings. First, germ-cell depletion across NOA aetiologies (SCOS, Klinefelter syndrome and AZF-microdeletion) leaves Sertoli cells numerically dominant. Second, surviving Sertoli cells ectopically activate an epididymal–defensin gene module, the six-gene fingerprint, that passes all three ambient-RNA controls and is reproduced across NOA aetiologies and independent datasets, and is low in the developing testis and unaltered by aging. Third, this module accompanies AR upregulation and NR3C2 downregulation at the cell level, suggesting a hormonal switch.

These findings extend a rapidly maturing single-cell literature. The adult human testicular cell atlas of Guo et al. [8] and subsequent atlases of puberty [9], neonatal and adult development [10], mammalian spermatogenesis [11], and cross-species germ-cell programs [12, 32, 33] established the reference cell-type landscape within which our analysis operates; our contribution is the disease dimension, a strict-mask, ambient-controlled comparison that none of the reference atlases performed. The spermatogonial stem-cell field has likewise converged on a single-cell description of the stem-cell hierarchy [10, 34–36], including its metabolic maturation [36] and its mouse roadmap [32, 44]. Mechanistically, our AR-dependent model is consistent with the classical Sertoli-cell literature: androgen action through the somatic compartment is required for adult Sertoli function [28–30], and hormonal control of germ-cell development operates through the Sertoli cell [45]. The graded fingerprint further aligns with the emerging view that Sertoli maturation states are plastic [5, 6]. Clinically, the capacity of the fingerprint to stratify NOA is directly relevant to surgical sperm retrieval, where patient selection remains the dominant determinant of outcome [27, 42, 43].

The fingerprint genes deserve individual scrutiny. DEFB119 and DEFB123 are primate beta-defensins canonically expressed in the epididymis [20]; their activation in Sertoli cells has been interpreted either as a marker of Sertoli maturation or, here, as ectopic adoption of a germ-cell/epididymal program. EPPIN and PRND (doppel) are epididymal/spermatogenic genes [21, 22]; TAF7L is an X-linked, testis-enriched transcription factor [24]; SEPT14 is a sperm-morphology gene whose mutation causes teratozoospermia [23]. Their coordinate activation in Sertoli cells therefore constitutes ectopic expression of an epididymal–defensin gene module.

SEPT14 deserves specific discussion, because its direction contradicts the literature: it is downregulated in spermatogenic failure, yet we observed 0.2%->68.8% any-positive (0.0%->60.3% high-expressors) in NOA Sertoli cells. We interpret this as ectopic activation of an epididymal–defensin gene module in Sertoli cells. Immunohistochemistry or RNA-FISH will test this directly.

The regulatory model integrates our data with the recent demonstration that mature Sertoli identity genes (CITED1, DEFB123, PRND) collapse upon androgen deprivation and NR3C2 knockdown [6]. Our quick check (Table 5) shows AR up at the cell level (31.2%->50.1%; cell-level log2FC + 0.71, adjusted P 2.60e-19; donor-level log2FC -0.28, P = 0.53), with NR3C2 downregulated (cell-level log2FC -0.19, adjusted P 2.55e-41; donor-level − 3.3, P = 0.072); however, AR expression did not consistently correlate with fingerprint expression at the single-cell level (Spearman rho − 0.055 to + 0.141), so the AR-dependence of the module remains correlational and hypothetical. We propose a three-mode model: Mode A (physiological; androgen-sufficient with germ cells present; mature program at intermediate level); Mode B (ADT; AR signal withdrawn; program collapses to an immature state); Mode C (NOA; catastrophic germ-cell loss; AR up, NR3C2 down; program switches to an AR-driven mature/hypermature Sertoli state). This reconciles our results with both the immature-Sertoli depletion reported in NOA [5] and the ADT collapse [6].

Table 5
Androgen-pathway genes in NOA Sertoli cells.

Gene	Developing-testis positive (%)	NOA positive (%)	Correlation (single-cell)
AR	31.2	50.1	cell-level log2FC + 0.71; donor-level − 0.28 (P = 0.53)
NR3C2	22.8	4.8	cell-level log2FC -0.19; donor-level − 3.3 (P = 0.072)
NR3C1	28.1	19.5	cell-level log2FC + 0.01; donor-level − 1.8 (P = 0.16)

Clinically, the six-gene fingerprint is a quantitative ruler of the spermatogenic state: low in health, not activated by aging, strongly activated in NOA. It provides a research-grade transcriptional readout of the Sertoli response to germ-cell loss; because obstructive controls show comparable positivity, it does not yet determine sperm-retrieval feasibility.

Two mechanistic bridges follow from the metabolic findings. First, the enrichment of HSP90-steroid-receptor chaperone cycling connects the module to the Sertoli cytoskeleton. Actin- and microtubule-based platforms are established HSP90 clients, which we have reviewed previously in the context of cell polarity and spermatogenesis [46]. Second, the metabolic-secretory rewiring, together with the known role of Sertoli cells as the principal testicular lactate source, raises the possibility that histone lactylation couples this activation to the epigenetic state of the surviving epithelium [47]; both predictions are directly testable in the NOA tissue context.

Limitations. Diffusion-pseudotime trajectory inference did not resolve the germ lineage monotonically (elongated spermatids formed a separate branch) and was therefore omitted. Normal adult Sertoli cells are scarce in the deposited data (n = 3 after strict masking), so the germ-cell-complete reference is the developing (pre-pubertal) testis rather than normal adult testis, introducing an age confound. SSC/Sertoli boundary annotation remains imperfect; GSE202647 has small and inconsistent OA controls, whose DEFB119 positivity approaches that of NOA and is therefore treated as underpowered rather than as evidence against specificity; communication scoring is composition-confounded; AR evidence is correlational and weak (single-cell Spearman rho − 0.055 to + 0.141; activity and nuclear localisation unmeasured); direct AR binding at fingerprint promoters lacks independent confirmation; and cross-species validation covers 3/6 fingerprint genes (Eppin, Prnd, Taf7l); DEFB119 and DEFB123 are primate-specific beta-defensins without established one-to-one mouse orthologues, so cross-species validation covers only the three assayable homologues (Eppin, Prnd, Taf7l).

Conclusions

Ectopic expression of an epididymal–defensin gene module, led by DEFB119, marks Sertoli cells in NOA across aetiologies. The module is associated with, but does not prove, germ-cell depletion: its specificity for germ-cell loss is not established because obstructive controls show comparable positivity, and only DEFB119 survives donor-level testing. The findings define a reproducible transcriptional feature of NOA Sertoli cells rather than a validated diagnostic or clinical tool. The six-gene fingerprint distinguishes germ-cell-depleted from germ-cell-bearing Sertoli cells (cross-donor AUC 0.957).

Abbreviations

AR
androgen receptor
AUC
area under the receiver-operating-characteristic curve
AZF
azoospermia factor
BH
Benjamini-Hochberg
DE
differential expression
FDR
false discovery rate
GO
Gene Ontology
HSP90
heat shock protein 90
KEGG

Kyoto Encyclopedia of Genes and Genomes
KS
Klinefelter syndrome
LASSO
least absolute shrinkage and selection operator
LODO
leave-one-donor-out
ML
machine learning
NOA
non-obstructive azoospermia
QC
quality control
RF
random forest
RFE
recursive feature elimination
SCOS
Sertoli-cell-only syndrome
scVI
single-cell variational inference
SHAP
SHapley Additive exPlanations
UMAP
uniform manifold approximation and projection
XGBoost
extreme gradient boosting.

Declarations

Ethics approval and consent to participate

This study is a secondary analysis of publicly available datasets and does not involve new human samples or patients.

Consent for publication

Not applicable.

Competing interests

The authors declare no competing interests.

Funding

Author Contribution

ML retrieved and quality-controlled the public single-cell datasets, performed the data analyses, and prepared the figures and tables; LL conceived and designed the study, supervised the analyses, and wrote the manuscript; ZZ, JZ and BZ contributed to data retrieval and verification. All authors read and approved the final manuscript.

Acknowledgements

Not applicable.

Data Availability

All data analysed in this study are publicly available: scRNA-seq datasets GSE134144 [53], GSE149512 [54], GSE202647 [55], GSE112013 [56] and GSE270931 [26] from the NCBI Gene Expression Omnibus; CRA014599 [57] from the CNCB GSA; and the CZ CELLxGENE Discover collections [58]. The full Sertoli-cell differential-expression table is provided as Table S1. Analysis code is provided as a supplementary file (analysis_scripts.zip).

References

1. Wosnitzer M, Goldstein M, Hardy MP. Rev Azoospermia Spermatogenesis. 2014;4:e28218. PMID 25105055.
2. Meistrich ML. Effects of chemotherapy and radiotherapy on spermatogenesis in humans. Fertil Steril. 2013;100(5):1180–6. PMID 24012199.
3. Wang RS, Yeh S, Tzeng CR, Chang C. Androgen receptor roles in spermatogenesis and fertility: lessons from testicular cell-specific androgen receptor knockout mice. Endocr Rev. 2009;30(2):119–32. PMID 19176467.
4. Sharpe RM, McKinnell C, Kivlin C, Fisher JS. Proliferation and functional maturation of Sertoli cells, and their relevance to disorders of testis function in adulthood. Reprod (Cambridge England). 2003;125(6):769–84. PMID 12773099.
5. Wang S, Wang H, Jin B, Yan H, Zheng Q, Zhao D. scRNA-seq and scATAC-seq reveal that Sertoli cell mediates spermatogenesis disorders through stage-specific communications in non-obstructive azoospermia. eLife. 2025;13. PMID 40371706.

6. Guo Y, Xu K, Zhou Q, Du L, Dang C, Liu G et al. Multi-omic analysis of androgen-deprived human testes reveals Sertoli cell regulation and niche reprogramming. Nat Commun. 2026;17(1). PMID 42660954.
7. Zhang Z. Composition-Aware Analysis of Public Testis Transcriptomes Identifies Reproducible Tissue-State Signatures in Nonobstructive Azoospermia. Omics: a journal of integrative biology. 2026. PMID 42844842.
8. Guo J, Grow EJ, Mlcochova H, Maher GJ, Lindskog C, Nie X, et al. The adult human testis transcriptional cell atlas. Cell Res. 2018;28(12):1141–57. PMID 30315278.
9. Guo J, Nie X, Giebler M, Mlcochova H, Wang Y, Grow EJ, et al. The Dynamic Transcriptional Cell Atlas of Testis Development during Human Puberty. Cell Stem Cell. 2020;26(2):262–e2764. PMID 31928944.
10. Sohni A, Tan K, Song HW, Burow D, de Rooij DG, Laurent L, et al. The Neonatal and Adult Human Testis Defined at the Single-Cell Level. Cell Rep. 2019;26(6):1501–e15174. PMID 30726734.
11. Hermann BP, Cheng K, Singh A, Roa-De La Cruz L, Mutoji KN, Chen IC, et al. The Mammalian Spermatogenesis Single-Cell Transcriptome, from Spermatogonial Stem Cells to Spermatids. Cell Rep. 2018;25(6):1650–e16678. PMID 30404016.
12. Rabbani M, Zheng X, Manske GL, Vargo A, Shami AN, Li JZ, et al. Decoding the Spermatogenesis Program: New Insights from Transcriptomic Analyses. Annu Rev Genet. 2022;56:339–68. PMID 36070560.
13. Nie X, Munyoki SK, Sukhwani M, Schmid N, Missel A, Emery BR, et al. Single-cell analysis of human testis aging and correlation with elevated body mass index. Dev Cell. 2022;57(9):1160–e11765. PMID 35504286.
14. Young MD, Behjati S. SoupX removes ambient RNA contamination from droplet-based single-cell RNA sequencing data. GigaScience. 2020;9(12). PMID 33367645.
15. Zhao L, Yao C, Xing X, Jing T, Li P, Zhu Z, et al. Single-cell analysis of developing and azoospermia human testicles reveals central role of Sertoli cells. Nat Commun. 2020;11(1):5683. PMID 33173058.
16. Lopez R, Regier J, Cole MB, Jordan MI, Yosef N. Deep generative modeling for single-cell transcriptomics. Nat Methods. 2018;15(12):1053–8. PMID 30504886.
17. Raudvere U, Kolberg L, Kuzmin I, Arak T, Adler P, Peterson H, et al. g:Profiler: a web server for functional enrichment analysis and conversions of gene lists (2019 update). Nucleic Acids Res. 2019;47(W1):W191–8. PMID 31066453.
18. Yang S, Corbett SE, Koga Y, Wang Z, Johnson WE, Yajima M, et al. Decontamination of ambient RNA in single-cell RNA-seq with DecontX. Genome Biol. 2020;21(1):57. PMID 32138770.
19. Efremova M, Vento-Tormo M, Teichmann SA, Vento-Tormo R. CellPhoneDB: inferring cell-cell communication from combined expression of multi-subunit ligand-receptor complexes. Nat Protoc. 2020;15(4):1484–506. PMID 32103204.

20. Patil AA, Cai Y, Sang Y, Blecha F, Zhang G. Cross-species analysis of the mammalian beta-defensin gene family: presence of syntenic gene clusters and preferential expression in the male reproductive tract. Physiol Genomics. 2005. PMID 16033865.
21. Allais-Bonnet A, Pailhoux E. Role of the prion protein family in the gonads. Front cell Dev biology. 2014;2:56. PMID 25364761.
22. Long Y, Gu A, Yang H, Ji G, Han X, Song L, et al. Distribution of Eppin in mouse and human testis. Mol Med Rep. 2011;4(1):71–5. PMID 21461566.
23. Wang HY, Chen KR, Yeh BC, Li WD, Wu SR, Ching YH, et al. SEPT14 complexes maintain sperm morphogenesis and function. FASEB journal: official publication Federation Am Soc Experimental Biology. 2025;39(4):e70414. PMID 39982757.
24. Li JY, Liu YF, Xu HY, Zhang JY, Lv PP, Liu ME, et al. Basonuclin 1 deficiency causes testicular premature aging: BNC1 cooperates with TAF7L to regulate spermatogenesis. J Mol Cell Biol. 2020;12(1):71–83. PMID 31065688.
25. Wu X, Lu M, Yun D, Gao S, Sun F. Long-read single-cell sequencing reveals the transcriptional landscape of spermatogenesis in obstructive azoospermia and Sertoli cell-only patients. QJM: monthly J Association Physicians. 2024;117(6):422–35. PMID 38192002.
26. Gene Expression Omnibus accession GSE270931. Gene expression profile at single cell level of testicular cells from young and aged mice. https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE270931
27. Tournaye H, Krausz C, Oates RD. Novel concepts in the aetiology of male reproductive impairment. Lancet Diabets Endocrionol. 2017;5(7):544–53. PMID 27395771.
28. Griswold MD. 50 years of spermatogenesis: Sertoli cells and their interactions with germ cells. Biol Reprod. 2018;99(1):87–100. PMID 29462262.
29. Smith LB, Walker WH. The regulation of spermatogenesis by androgens. Semin Cell Dev Biol. 2014;30:2–13. PMID 24598768.
30. De Gendt K, Verhoeven G. Tissue- and cell-specific functions of the androgen receptor revealed through conditional knockout models in mice. Mol Cell Endocrinol. 2012;352(1–2):13–25. PMID 21871526.
31. Hammoud SS, Low DH, Yi C, Carrell DT, Guccione E, Cairns BR. Chromatin and transcription transitions of mammalian adult germline stem cells and spermatogenesis. Cell Stem Cell. 2014;15(2):239–53. PMID 24835570.
32. Green CD, Ma Q, Manske GL, Shami AN, Zheng X, Marini S, et al. A Comprehensive Roadmap of Murine Spermatogenesis Defined by Single-Cell RNA-Seq. Dev Cell. 2018;46(5):651–e66710. PMID 30146481.
33. Shami AN, Zheng X, Munyoki SK, Ma Q, Manske GL, Green CD, et al. Single-Cell RNA Sequencing of Human, Macaque, and Mouse Testes Uncovers Conserved and Divergent Features of Mammalian Spermatogenesis. Dev Cell. 2020;54(4):529–e54712. PMID 32504559.

34. de Rooij DG. The nature and dynamics of spermatogonial stem cells. Development. 2017;144(17):3022–30. PMID 28851723.
35. Tan K, Wilkinson MF. Human Spermatogonial Stem Cells Scrutinized under the Single-Cell Magnifying Glass. Cell Stem Cell. 2019;24(2):201–3. PMID 30735645.
36. Voigt AL, Dardari R, Su L, Lara NLM, Sinha S, Jaffer A, et al. Metabolic transitions define spermatogonial stem cell maturation. Hum Reprod (Oxford England). 2022;37(9):2095–112. PMID 35856882.
37. Hao Y, Hao S, Andersen-Nissen E, Mauck WM 3rd, Zheng S, Butler A, et al. Integrated analysis of multimodal single-cell data. Cell. 2021;184(13):3573–e358729. PMID 34062119.
38. Stuart T, Butler A, Hoffman P, Hafemeister C, Papalexi E, Mauck WM 3, et al. Comprehensive Integration of Single-Cell Data. Cell. 2019;177(7):1888–e190221. PMID 31178118.
39. Wolf FA, Angerer P, Theis FJ. SCANPY: large-scale single-cell gene expression data analysis. Genome Biol. 2018;19(1):15. PMID 29409532.
40. Jin S, Guerrero-Juarez CF, Zhang L, Chang I, Ramos R, Kuan CH, et al. Inference and analysis of cell-cell communication using CellChat. Nat Commun. 2021;12(1):1088. PMID 33597522.
41. Sang-Aram C, Browaeys R, Seurinck R, Saeys Y. Unraveling cell-cell communication with NicheNet by inferring active ligands from transcriptomics data. Nat Protoc. 2025;20(6):1439–67. PMID 40038548.
42. Brant A, Schlegel PN. Modern surgical treatment of azoospermia. Curr Opin Urol. 2023;33(1):39–44. PMID 36301052.
43. Kimmins S, Anderson RA, Barratt CLR, Behre HM, Catford SR, De Jonge CJ, et al. Frequency, morbidity and equity - the case for increased research on male fertility. Nat Rev Urol. 2024;21(2):102–24. PMID 37828407.
44. Busada JT, Velte EK, Serra N, Cook K, Niedenberger BA, Willis WD, et al. Rhox13 is required for a quantitatively normal first wave of spermatogenesis in mice. Reprod (Cambridge England). 2016;152(5):379–88. PMID 27486269.
45. O'Shaughnessy PJ. Hormonal control of germ cell development and spermatogenesis. Semin Cell Dev Biol. 2014;29:55–65. PMID 24598767.
46. Li L, Gao Y, Chen H, Jesus T, Tang E, Li N, et al. Cell polarity, cell adhesion, and spermatogenesis: role of cytoskeletons. F1000Research. 2017;6:1565. 10.12688/f1000research.11421.1.
47. Zhang D, Tang Z, Huang H, Zhou G, Cui C, Weng Y, et al. Metabolic regulation of gene expression by histone lactylation. Nature. 2019;574(7779):575–80. PMID 31645732.
48. Gold MP, Reyes M, Diamant N, Kuo T, Hajiramezanali E, Newburger JW. Scoring gene importance by interpreting single-cell foundation models. Nature biotechnology. 2026(in press). PMID 42204361.
49. Aibar S, González-Blas CB, Moerman T, Huynh-Thu VA, Imrichova H, Hulselmans G, et al. SCENIC: single-cell regulatory network inference and clustering. Nat Methods. 2017;14:1083–6. PMID 28991892.

50. Dimitrov D, Turei D, Garrido-Rodriguez M, Burmedi PL, Nagai JS, Boys C, et al. Comparison of methods and resources for cell-cell communication inference from single-cell RNA-Seq data. Nat Commun. 2022;13:3224. PMID 35680885.
51. Li L, Mao B, Ben S, Li H, et al. Defects of microtubule cytoskeletal organization in NOA human testes. Reproductive biology Endocrinol. 2022;20(1):154. PMID 36329464.
52. Wu X, Li L, et al. Map-1a regulates Sertoli cell BTB dynamics through the cytoskeletal organization of microtubule and F-actin. Reproductive biology Endocrinol. 2024;22(1):78. 10.1186/s12958-024-01204-y.

Figures

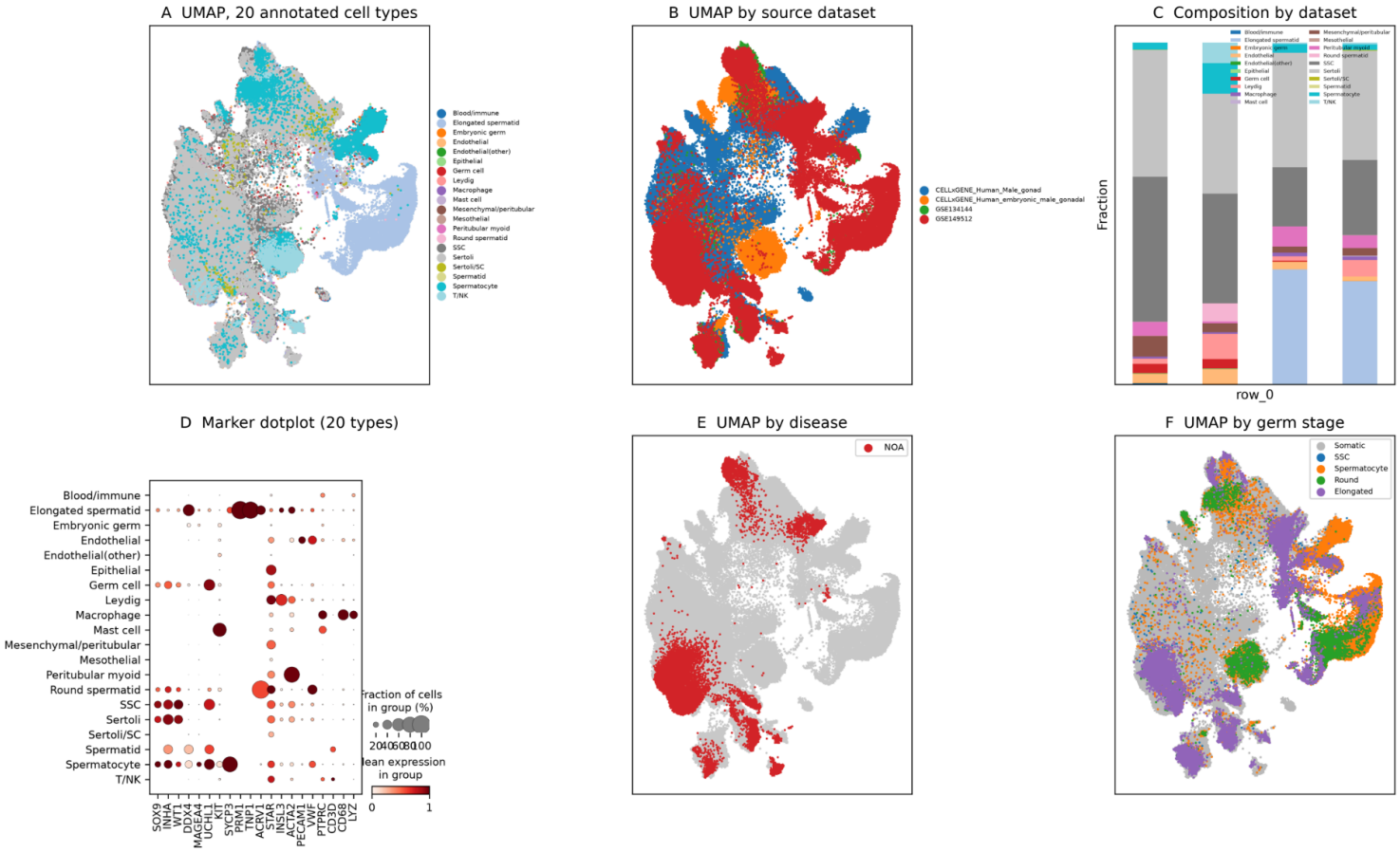


Figure 1

Integrated human testicular atlas. (A) UMAP of 250,584 cells coloured by 20 annotated cell types. (B) Same embedding coloured by source dataset, showing scVI batch integration. (C) Cell-type composition. (D) Marker dotplot validating the 20 annotations. (E) UMAP coloured by disease (NOA in red). (F) UMAP coloured by germ-stage class.

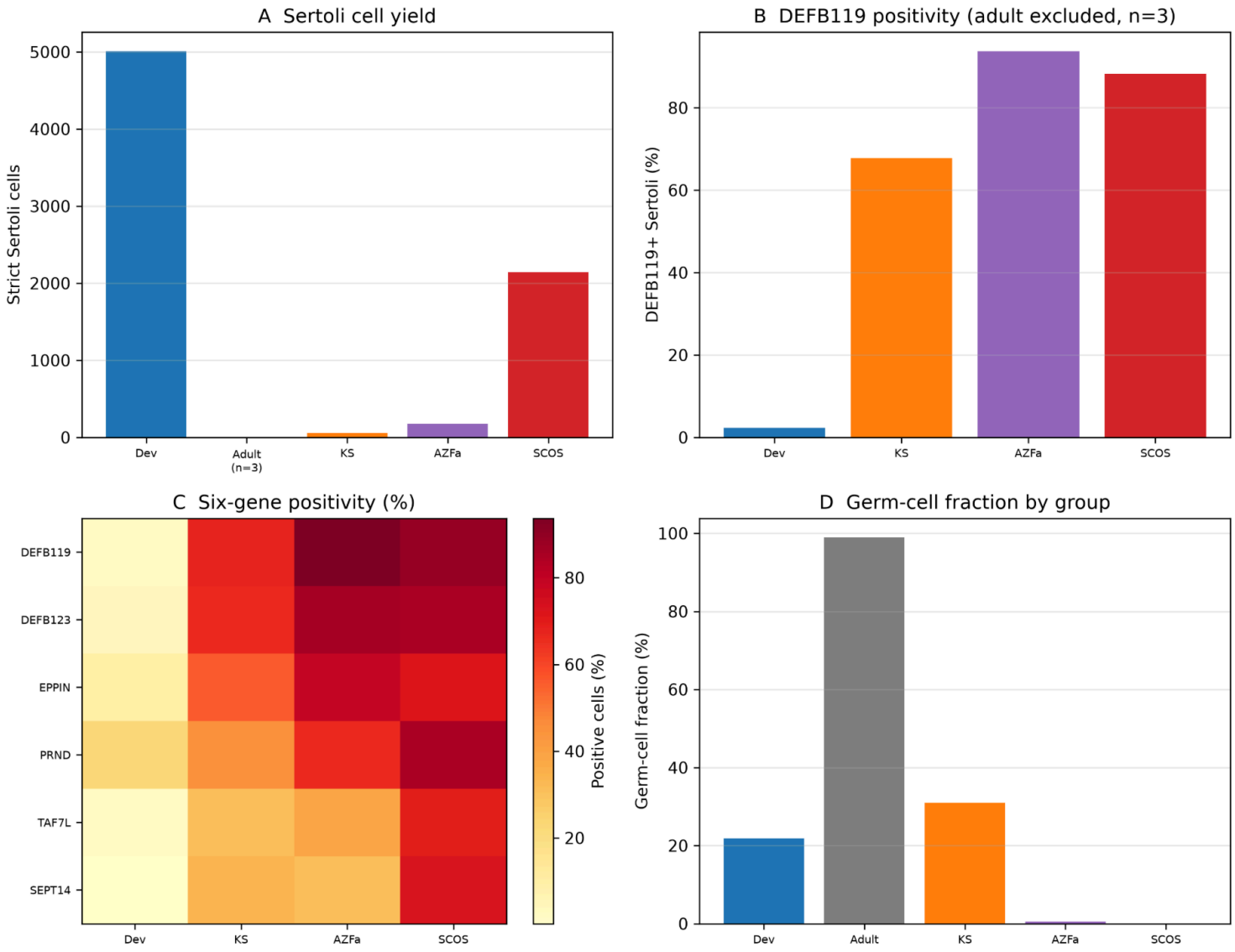


Figure 2

NOA aetiologies and the ectopic epididymal–defensin gene module. (A) Strict-Sertoli cell yield per group (normal adult: 3 cells). (B) DEFB119 positivity in strict Sertoli cells (normal adult excluded; only 3 cells available). (C) Six-gene positivity heatmap by group. (D) Germ-cell fraction by group.

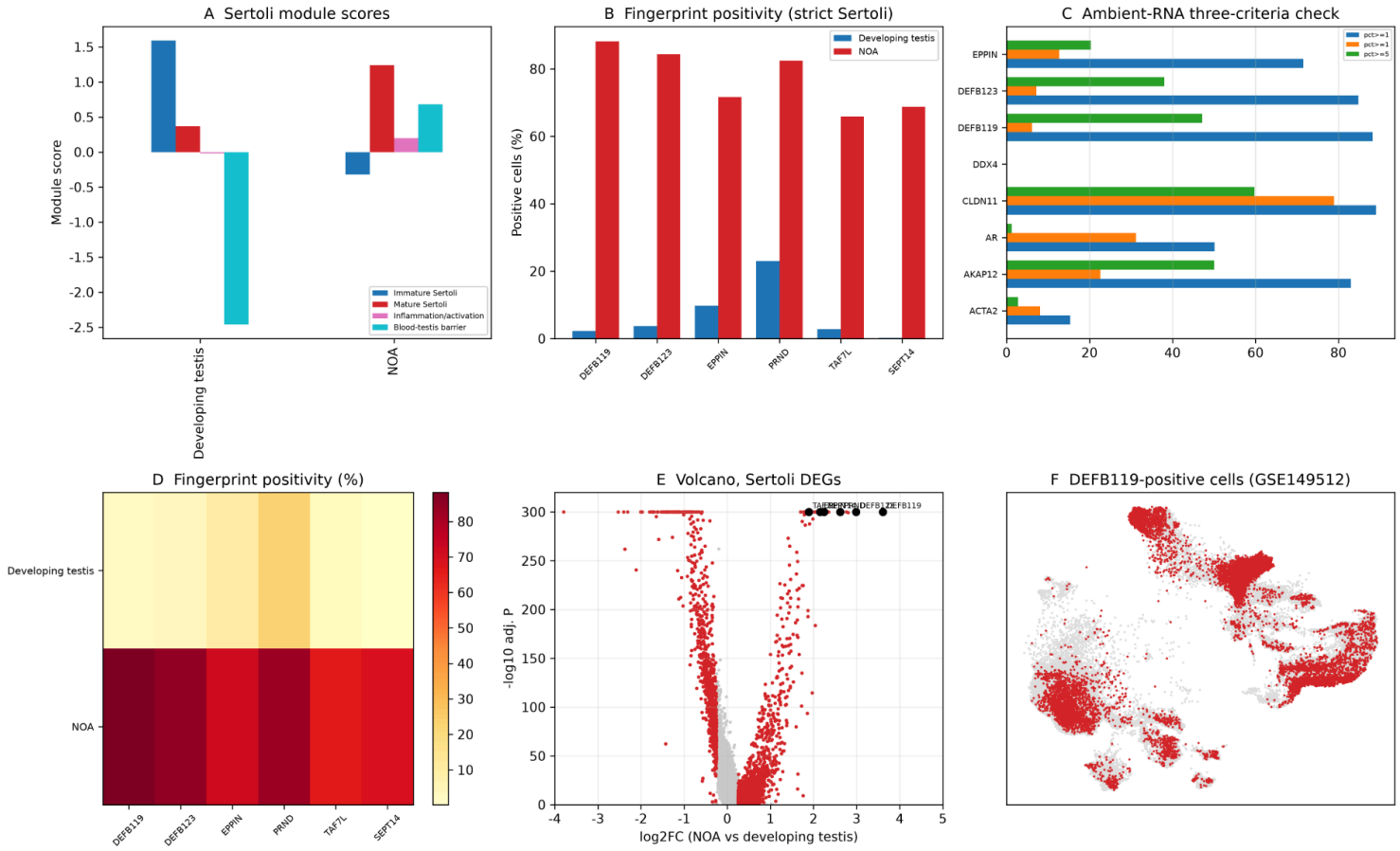


Figure 3

Sertoli ectopic epididymal–defensin gene activation. (A) Sertoli module scores in NOA vs developing-testis. (B) Six-gene fingerprint positivity (strict Sertoli, GSE149512). (C) Ambient-RNA three-criteria check. (D) Per-group positivity summary of the six fingerprint genes (two rows). (E) Volcano plot of Sertoli DEGs with fingerprint genes in red. (F) UMAP highlighting DEFB119-positive cells.

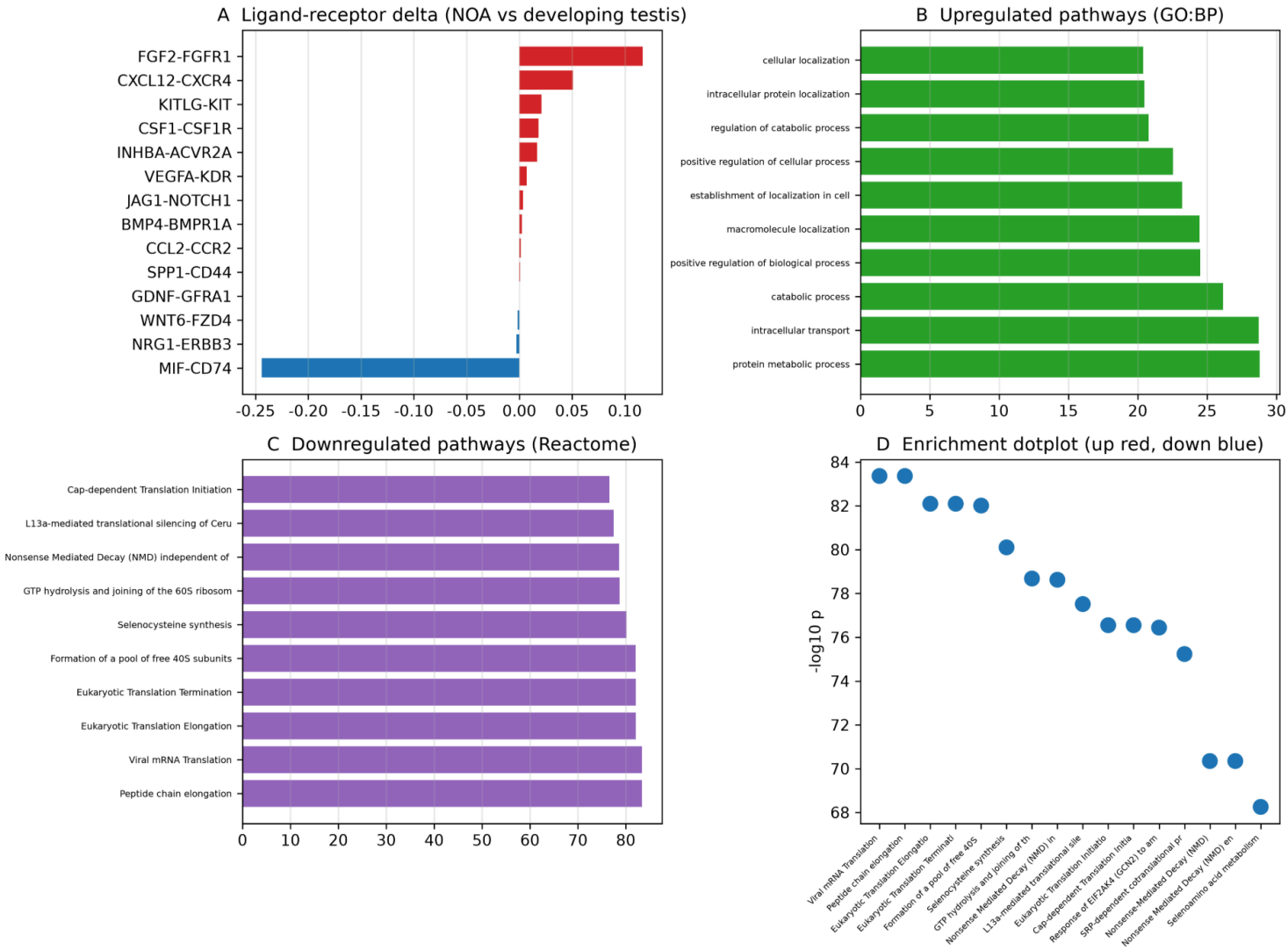


Figure 4

Niche signalling and enrichment. (A) Ligand-receptor changes. (B) Upregulated pathways. (C) Downregulated pathways. (D) Enrichment dotplot combining GO and Reactome terms.

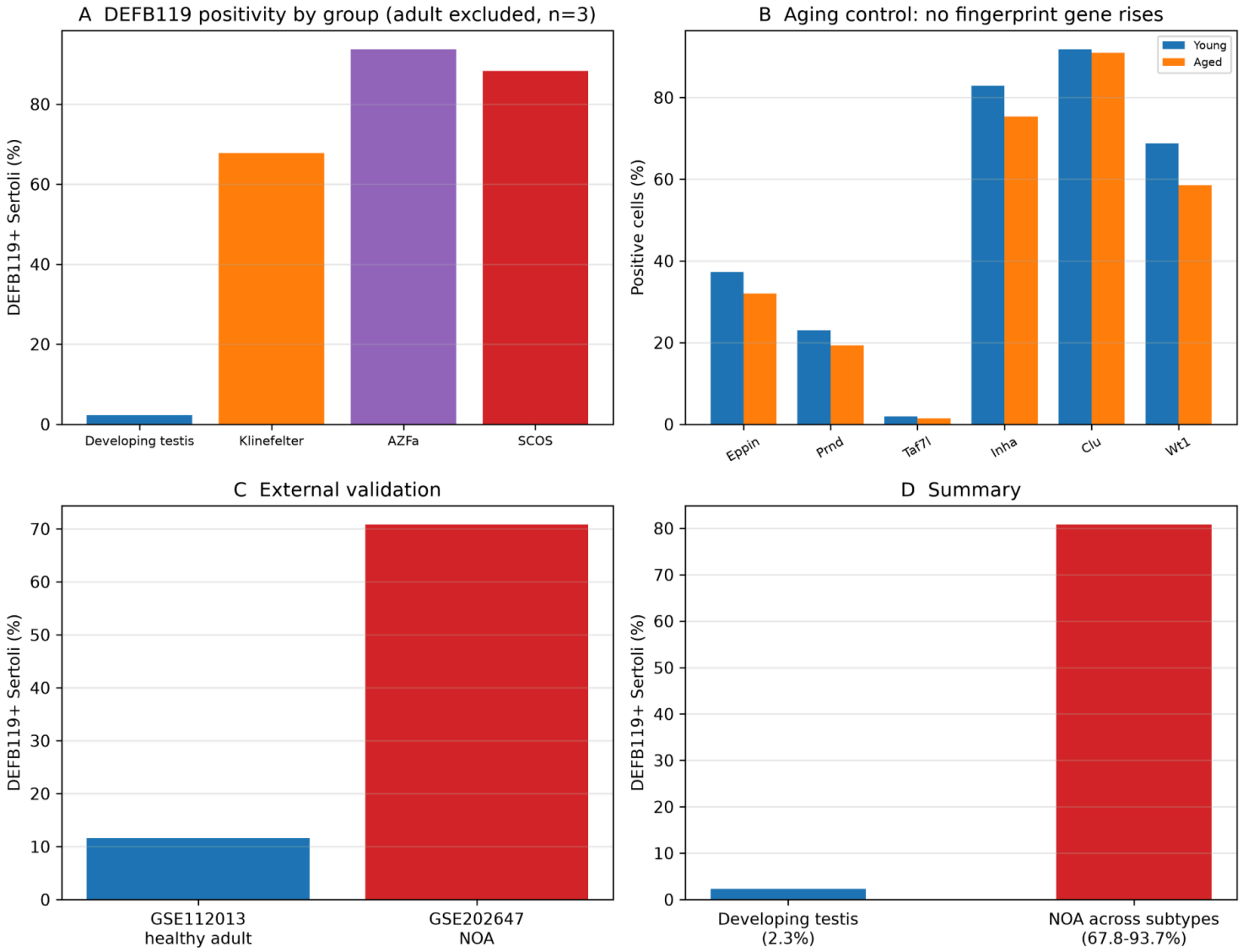


Figure 5

Three-way arbitration and aging control. (A) DEFB119 positivity across the four primary-cohort groups; the two external cohorts are shown in (C). (B) Fingerprint homologues in young vs aged mice (six samples). (C) Cross-cohort positivity ladder. (D) Graded activation summary (healthy vs NOA).

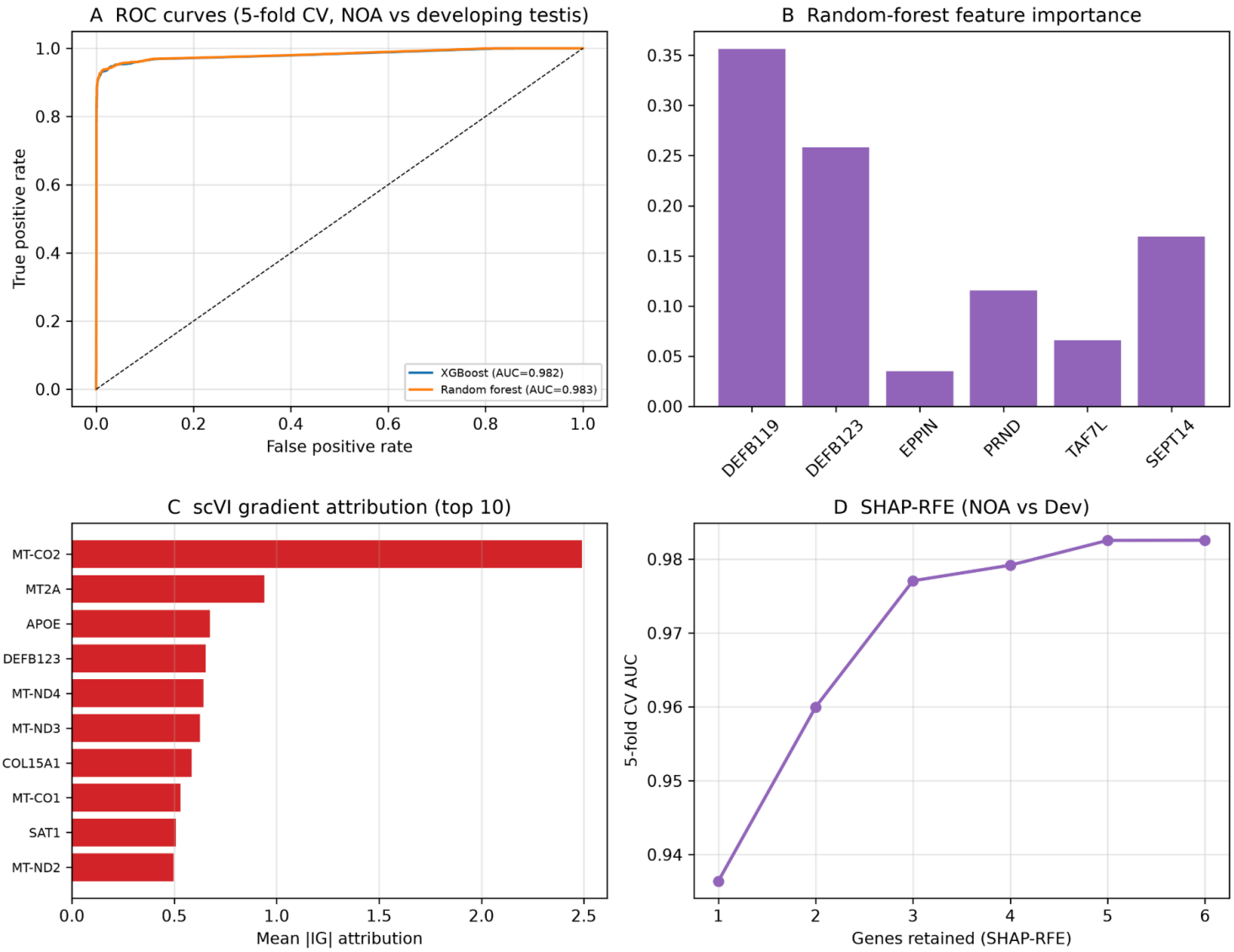


Figure 6

Machine-learning diagnostic classifier. (A) ROC curves for XGBoost and random forest (5-fold CV; LASSO shown in Table 4). (B) Random-forest feature importance. (C) scVI gradient attribution of the NOA Sertoli embedding (top 10). (D) SHAP-based recursive feature elimination: cross-validated AUC versus genes retained.

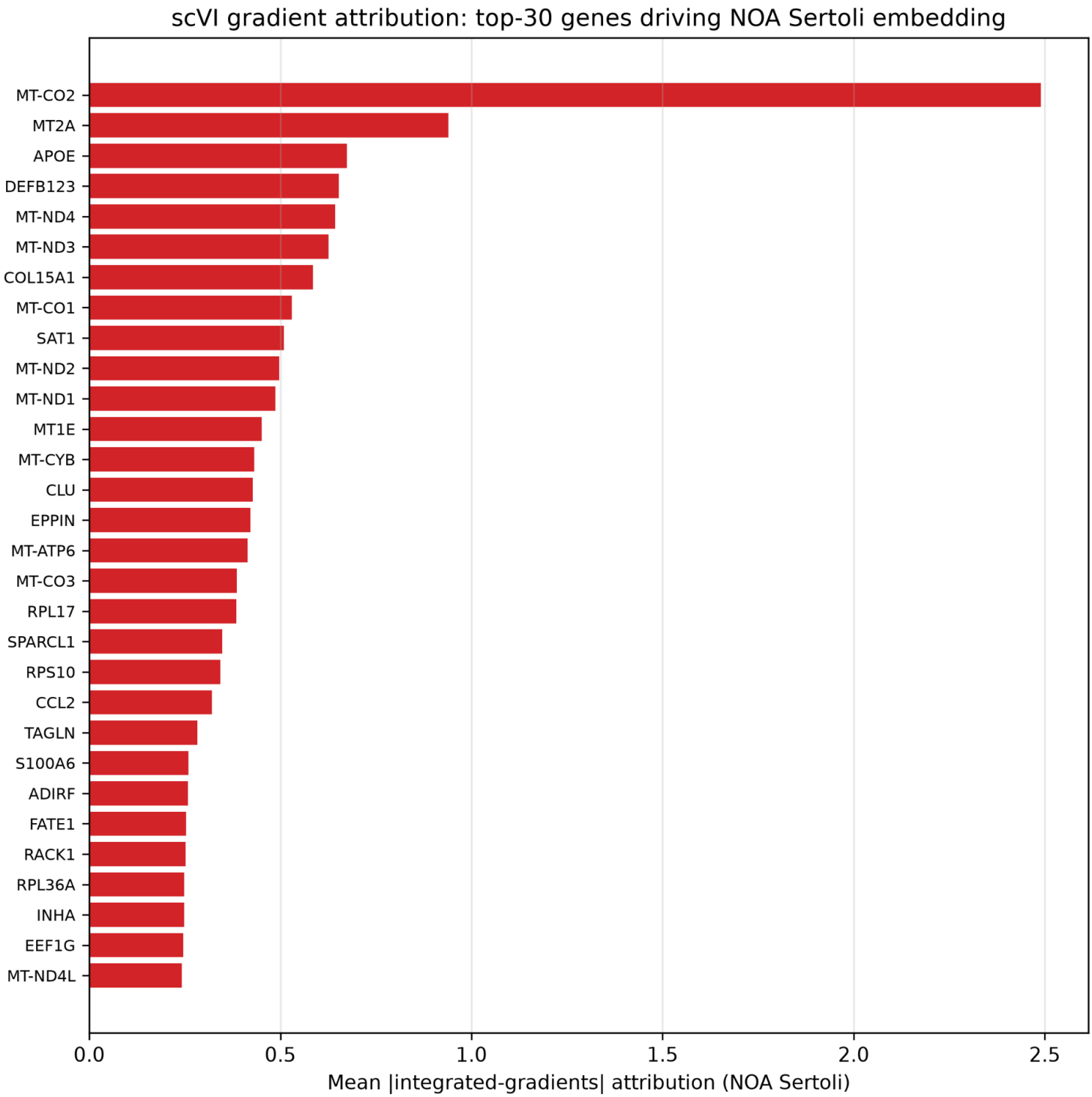


Figure 7

scVI gradient attribution. Mean absolute integrated-gradients attribution of the scVI encoder for NOA Sertoli cells; top-30 genes are shown.

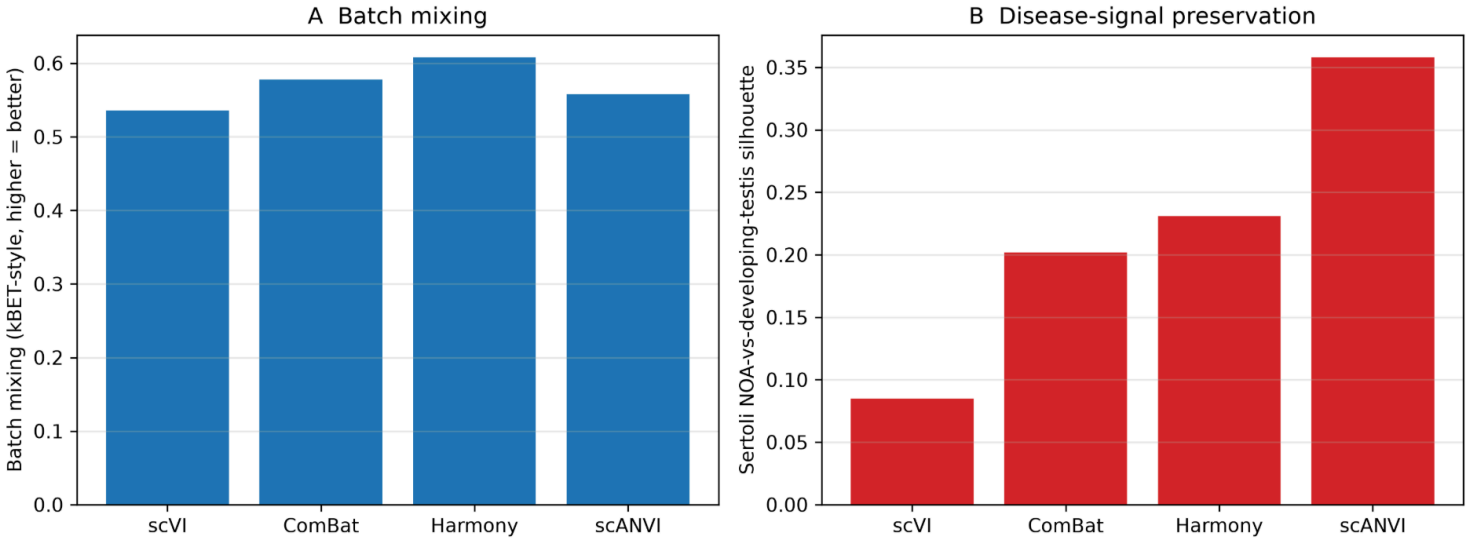


Figure 8

Integration benchmark. (A) Batch mixing (kBET-style, lower is better). (B) Sertoli NOA-vs-developing-testis silhouette (disease-signal preservation).

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- analysisscripts.zip