nature portfolio

https://doi.org/10.1038/s41586-026-10363-4

Supplementary information

Spatial atlas of diabetic kidney disease reveals a B cell-rich subgroup

In the format provided by the authors and unedited

Spatial Atlas of Diabetic Kidney Disease Reveals a B-Cell-Rich Subgroup

Bernhard Dumoulin[1,2,3,4], Jonathan Levinsohn[1,2,3,4,5], Konstantin Klötzer[1,2,3,4,6], Chenyu Li[1,2,3,4], Liran Mao[7,8,9], Eunji Ha[1,2,3,4], Samer Mohandes[1,2,3,4], Thao Nguyen[16], Luca Paruzzo[14], Daigoro Hirohama[1,2,3,4], Victoria Fang[19], Vijay G Bhoj[9], Hamideh Parhiz[17,18], Magaiver Andrade-Silva[1,2,3,4], Amin Abedini[1,2,3,4], Andi Bergeson[1,2,3,4], Daniel Traum[2,4], Michael J. May[20], Klaus H Kaestner[2,4], Marco Ruella[13,14], Fiona Elizabeth McAllister[16], A Ari Hakimi[13], Mingyao Li[8,9], Matthew Palmer[9], E. John Wherry[10,11], Christopher A. Hunter[12], TRIDENT consortium*, Katalin Susztak[1,2,3,4]

Affiliations:

[1]Renal, Electrolyte, and Hypertension Division, Department of Medicine, University of Pennsylvania, Perelman School of Medicine, Philadelphia, PA 19104, USA

[2]Institute for Diabetes, Obesity, and Metabolism, University of Pennsylvania, Perelman School of Medicine, Philadelphia, PA 19104, USA

[3]Penn/CHOP Kidney Innovation Center, University of Pennsylvania, Perelman School of Medicine, Philadelphia, PA 19104, USA

[4]Department of Genetics, University of Pennsylvania, Perelman School of Medicine, Philadelphia, PA 19104, USA

[5]Department of Pediatrics, Division of Nephrology, Children's Hospital of Philadelphia, Philadelphia, PA 19104, USA

[6]Division of Nephrology, Department of Internal Medicine, Medical University of Graz, Graz, 8036, Austria

[7]Graduate Group in Genomics and Computational Biology, Perelman School of Medicine, University of Pennsylvania, Philadelphia, PA, USA

[8]Statistical Center for Single-Cell and Spatial Genomics, Department of Biostatistics, Epidemiology and Informatics, University of Pennsylvania Perelman School of Medicine, Philadelphia, PA, USA

[9]Department of Pathology and Laboratory Medicine, University of Pennsylvania, Perelman School of Medicine, Philadelphia, Pennsylvania, USA

[10]Institute for Immunology and Immune Health, Perelman School of Medicine, Philadelphia, PA 19104, USA

[11]Department of Systems Pharmacology and Translational Therapeutics, Perelman School of Medicine, Philadelphia, PA 19104, USA

[12]Department of Pathobiology, University of Pennsylvania School of Veterinary Medicine, Philadelphia, PA, 19104 USA

[13]Urology Service, Department of Surgery, Memorial Sloan Kettering Cancer Center, New York, New York, USA.

[14]Lymphoma Program, Abramson Cancer Center, University of Pennsylvania, Philadelphia

[15]Cell Therapy and Transplant Program, Abramson Cancer Center, University of Pennsylvania, Philadelphia

[16]Calico Life Sciences LLC, South San Francisco, USA.

[17]Department of Systems Pharmacology and Translational Therapeutics, Perelman School of Medicine, University of Pennsylvania, Philadelphia

[18]Penn Institute for RNA Innovation, University of Pennsylvania, Philadelphia, PA, USA

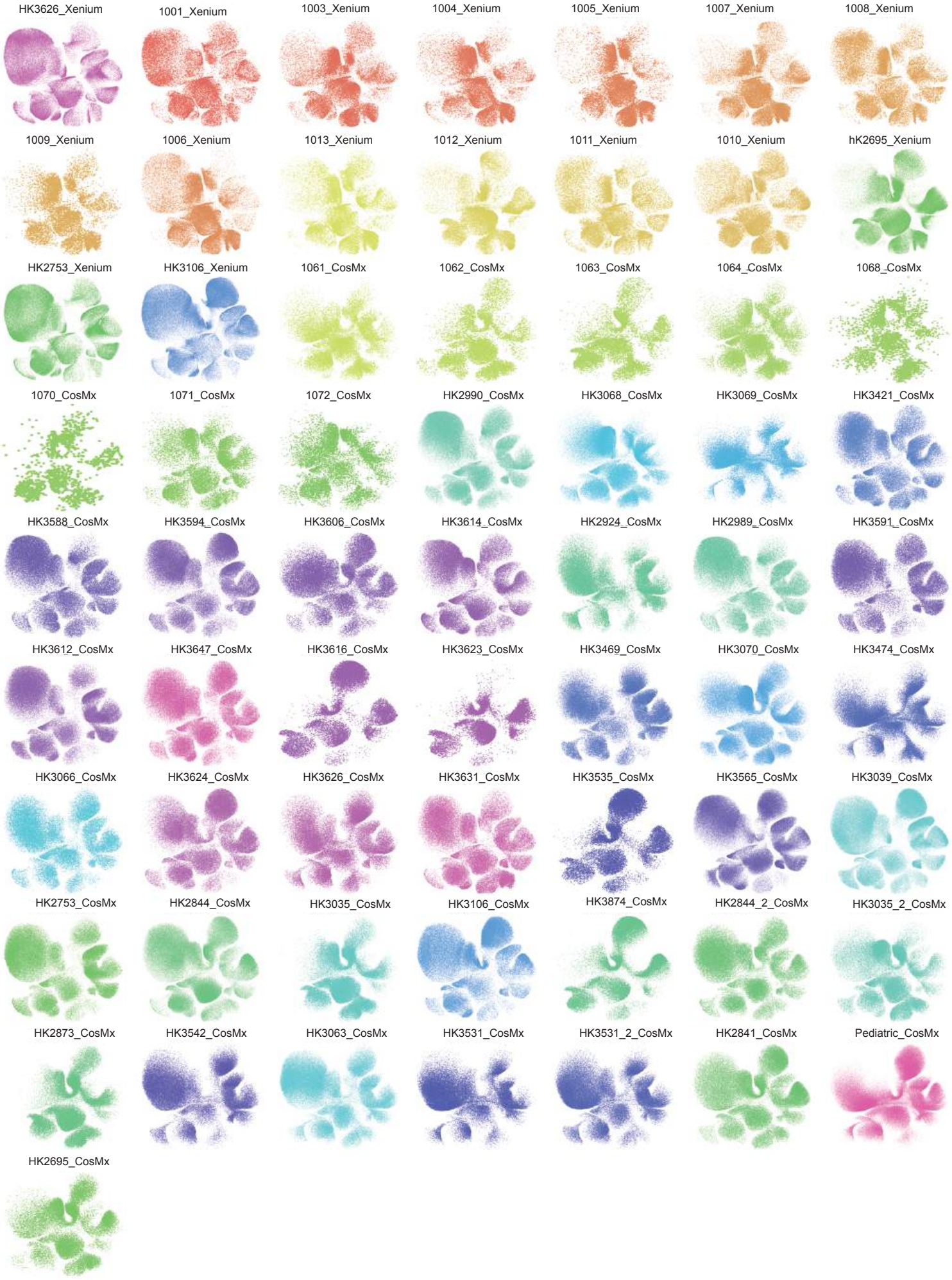
[19]Department of Dermatology, Perelman School of Medicine, University of Pennsylvania, Philadelphia, PA 19104, USA.

[20]Department of Biomedical Sciences, School of Veterinary Medicine, University of Pennsylvania, Philadelphia, PA 19104, USA.

*A list of authors and their affiliations appears at the end of the paper.

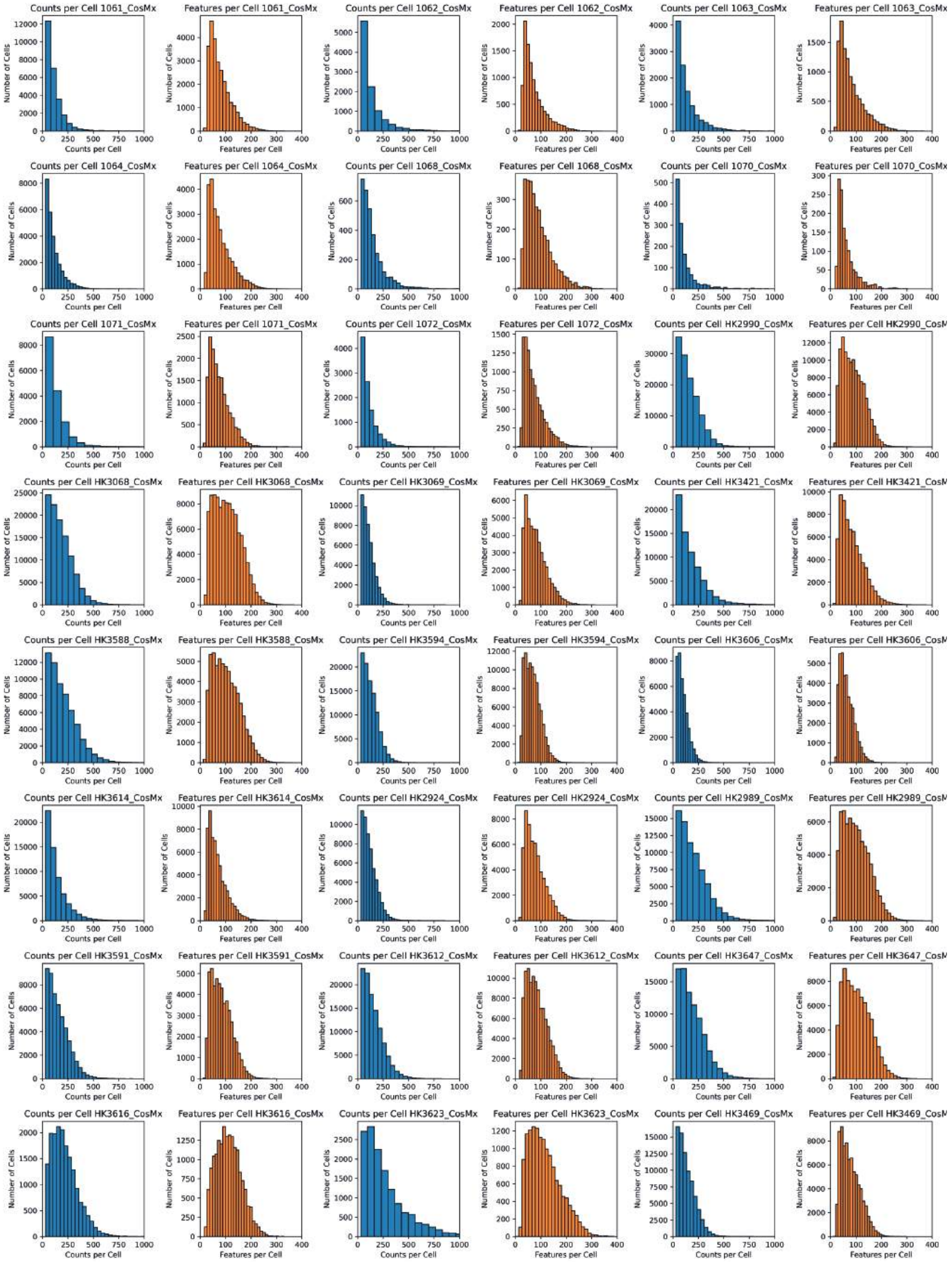
Contact Information:

Katalin Susztak, MD, PhD
Professor of Medicine
University of Pennsylvania, Perelman School of Medicine
3400 Civic Center Blvd,
Smilow Translational building 12-123,
Philadelphia, PA 19104
Phone: (215) 898-2009
ksusztak@pennmedicine.upenn.edu

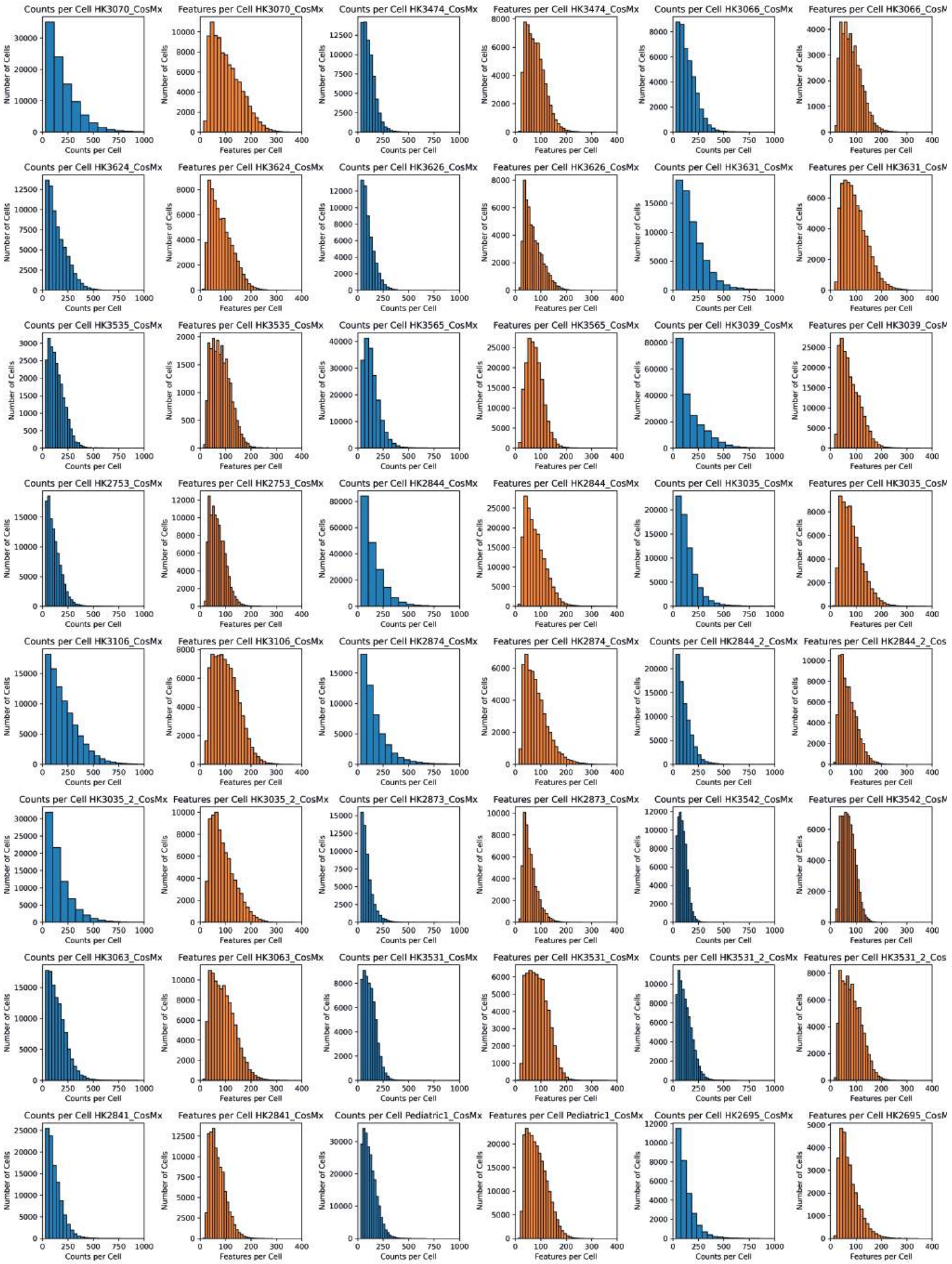


Supplementary Figure 1. Uniform Manifold Approximation and Projection (UMAP) plots of all individual samples included in the dataset. Each panel displays the two-dimensional UMAP embedding of cells from a single sample. Cell colors correspond to the sample identifier.

a

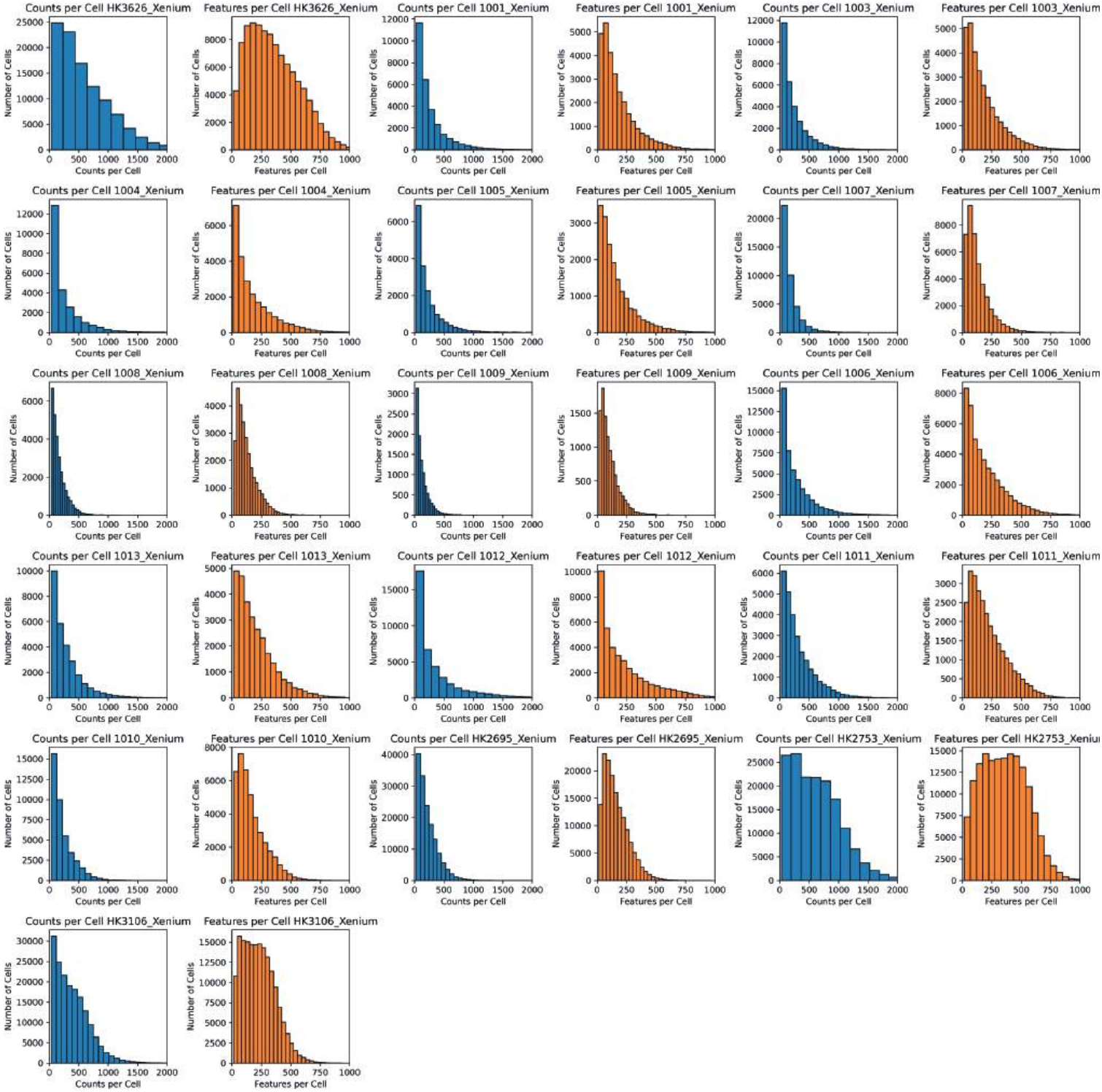


Supplementary Figure 2. Counts and Features CosMx. Histograms showing the distribution of total transcript counts (blue) and detected genes (Features) per cell across individual CosMx samples. For each sample, two histograms are shown side by side: the left panel (blue) displays the number of cells per binned count value, and the right panel shows the number of cells per binned gene count (orange).

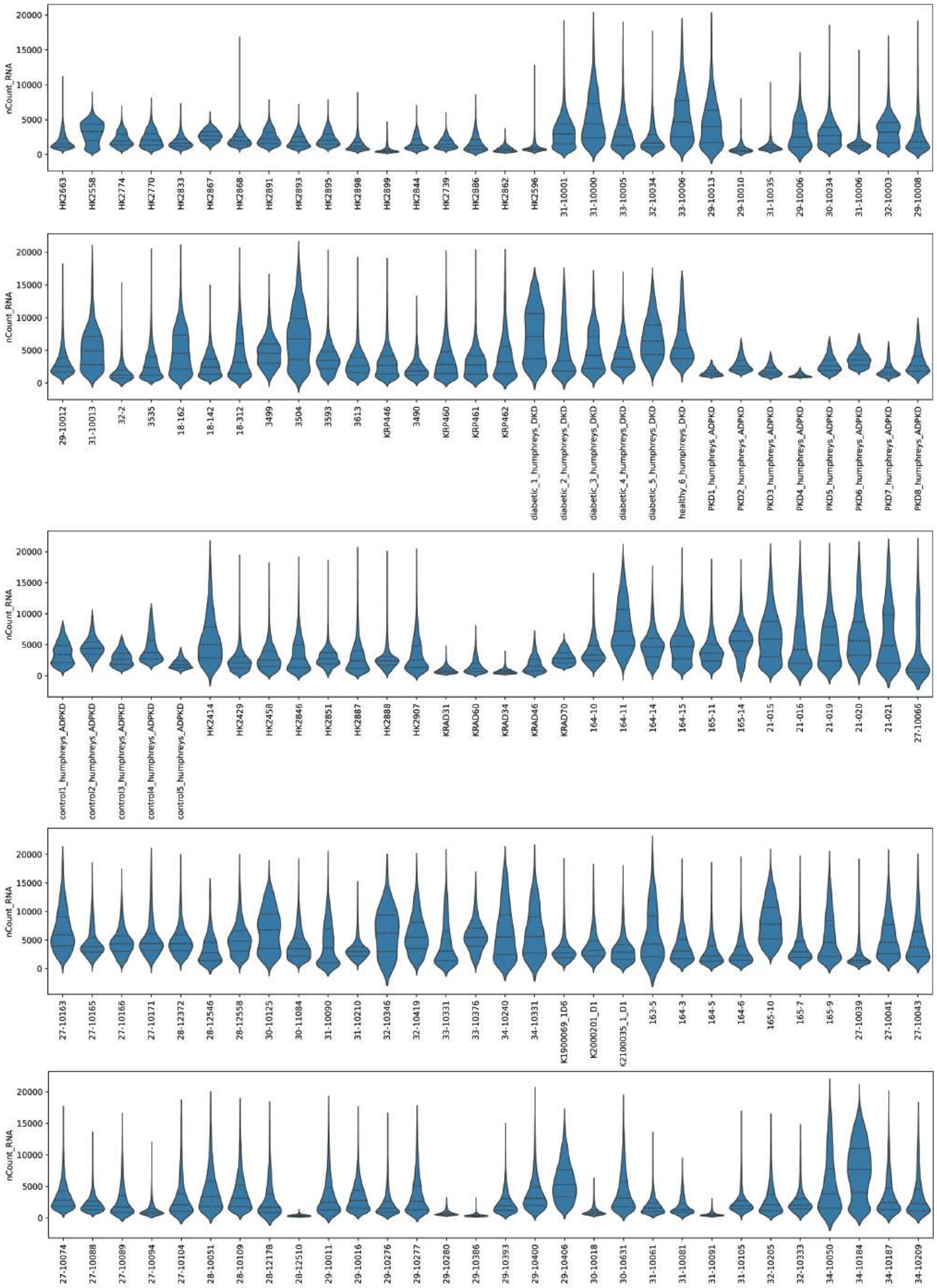


Supplementary Figure 3. Counts and Features CosMx II. Histograms showing the distribution of total transcript counts (blue) and detected genes (Features) per cell across individual CosMx samples. For each sample, two histograms are shown side by side: the left panel (blue) displays the number of cells per binned count value, and the right panel shows the number of cells per binned gene count (orange).

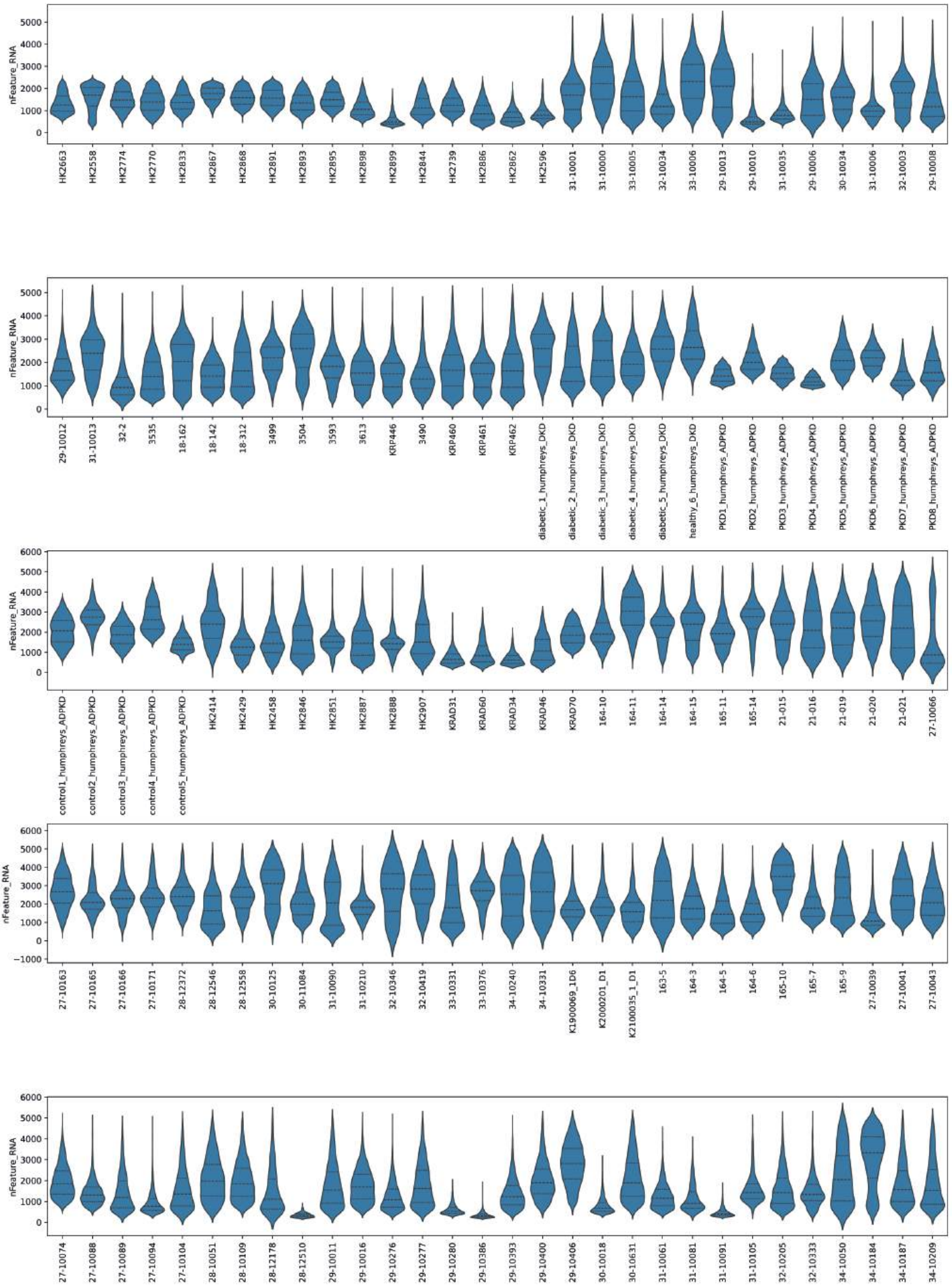
Counts and Features Xenium 5k



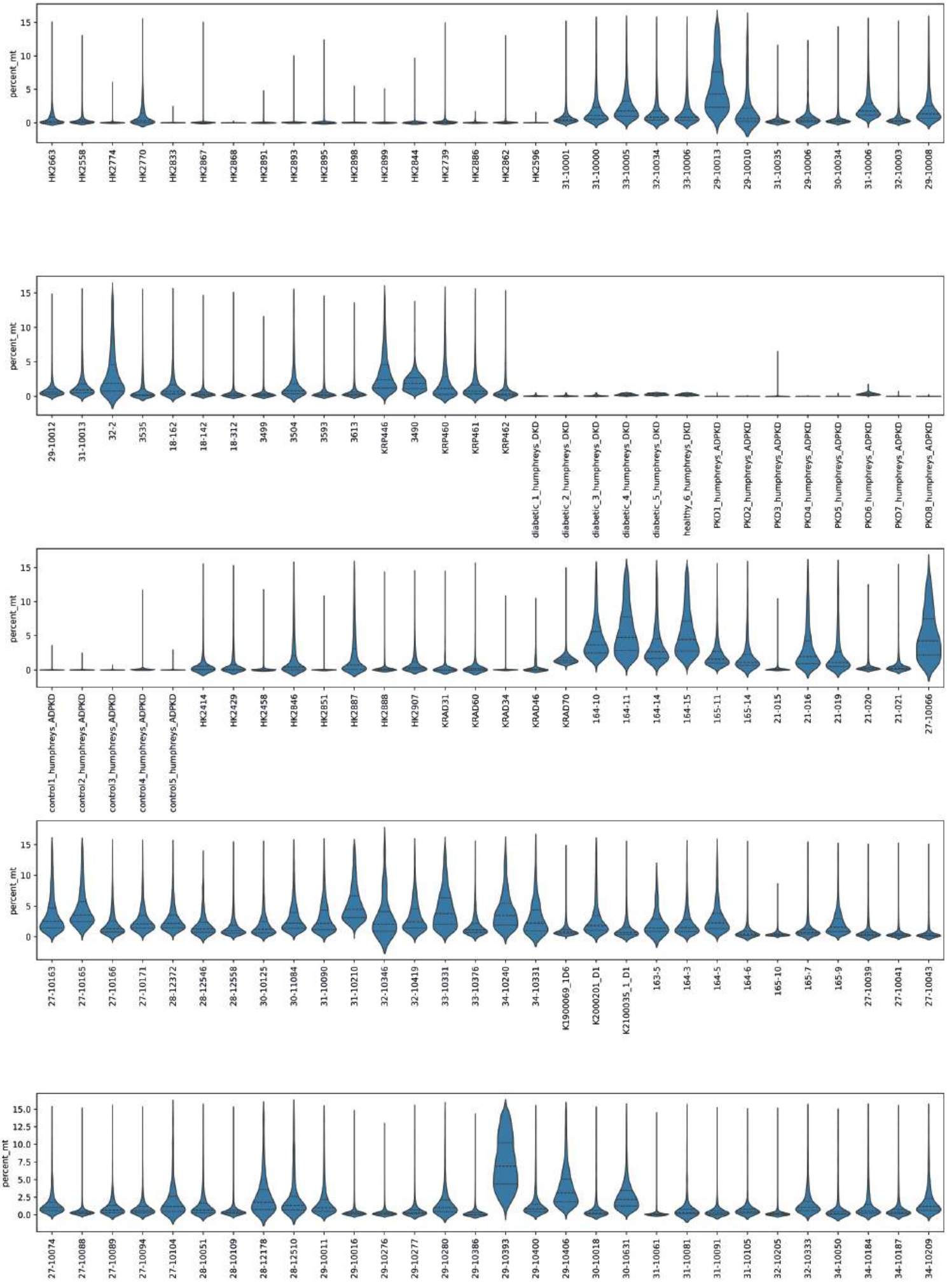
Supplementary Figure 4. Counts and Features Xenium. Histograms showing the distribution of total transcript counts (blue) and detected genes (Features) per cell across individual Xenium samples. For each sample, two histograms are shown side by side: the left panel (blue) displays the number of cells per binned count value, and the right panel shows the number of cells per binned gene count (orange).



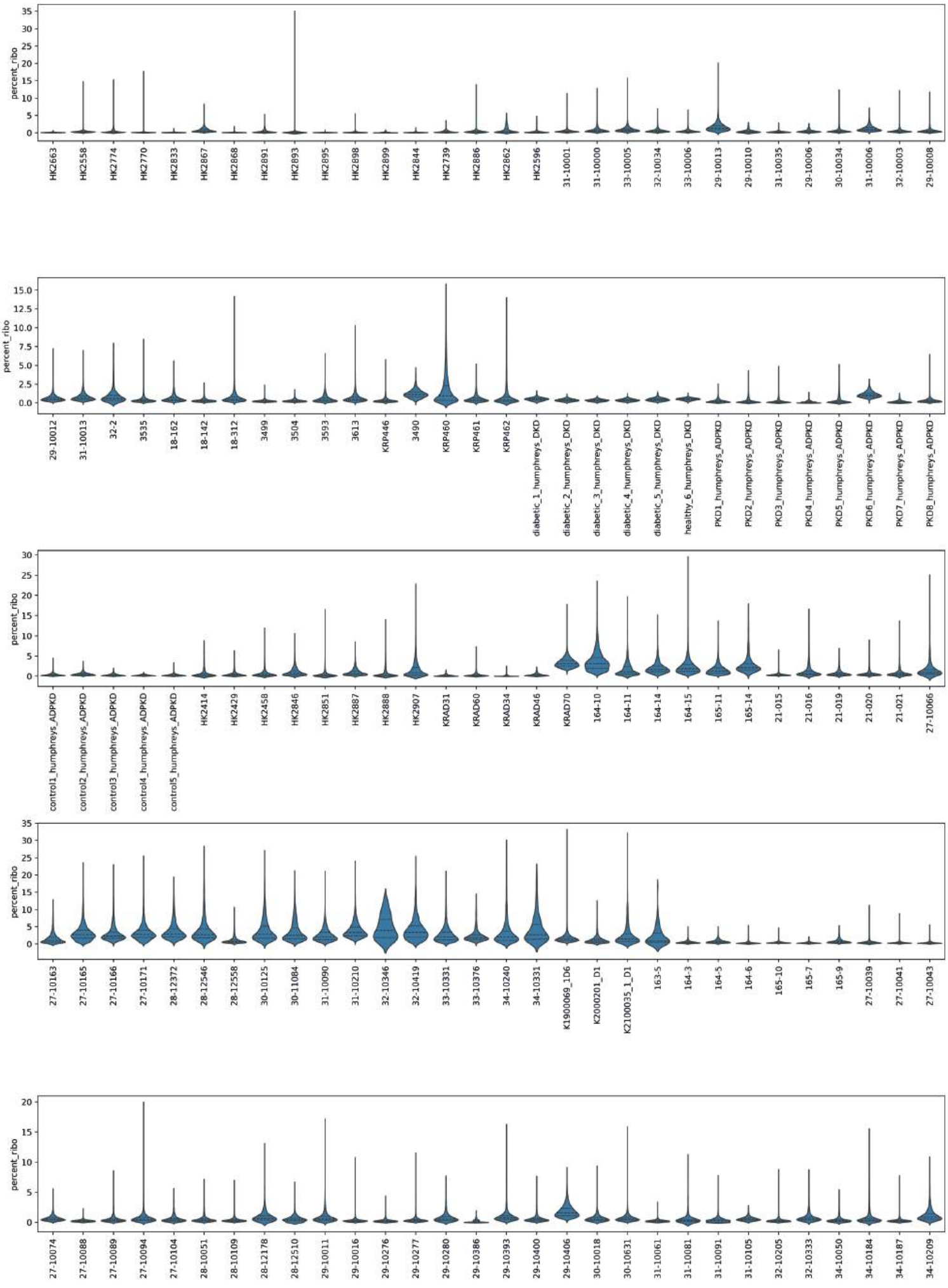
Supplementary Figure 5. Total transcripts counts per sample in snRNAseq dataset: Violin plots showing the distribution of total transcript counts (nCount_RNA) per cell, grouped by samples for all samples in the single nuclear sequencing dataset. Each violin represents the count distribution within a single sample, with inner quartile lines indicating the interquartile range and median.



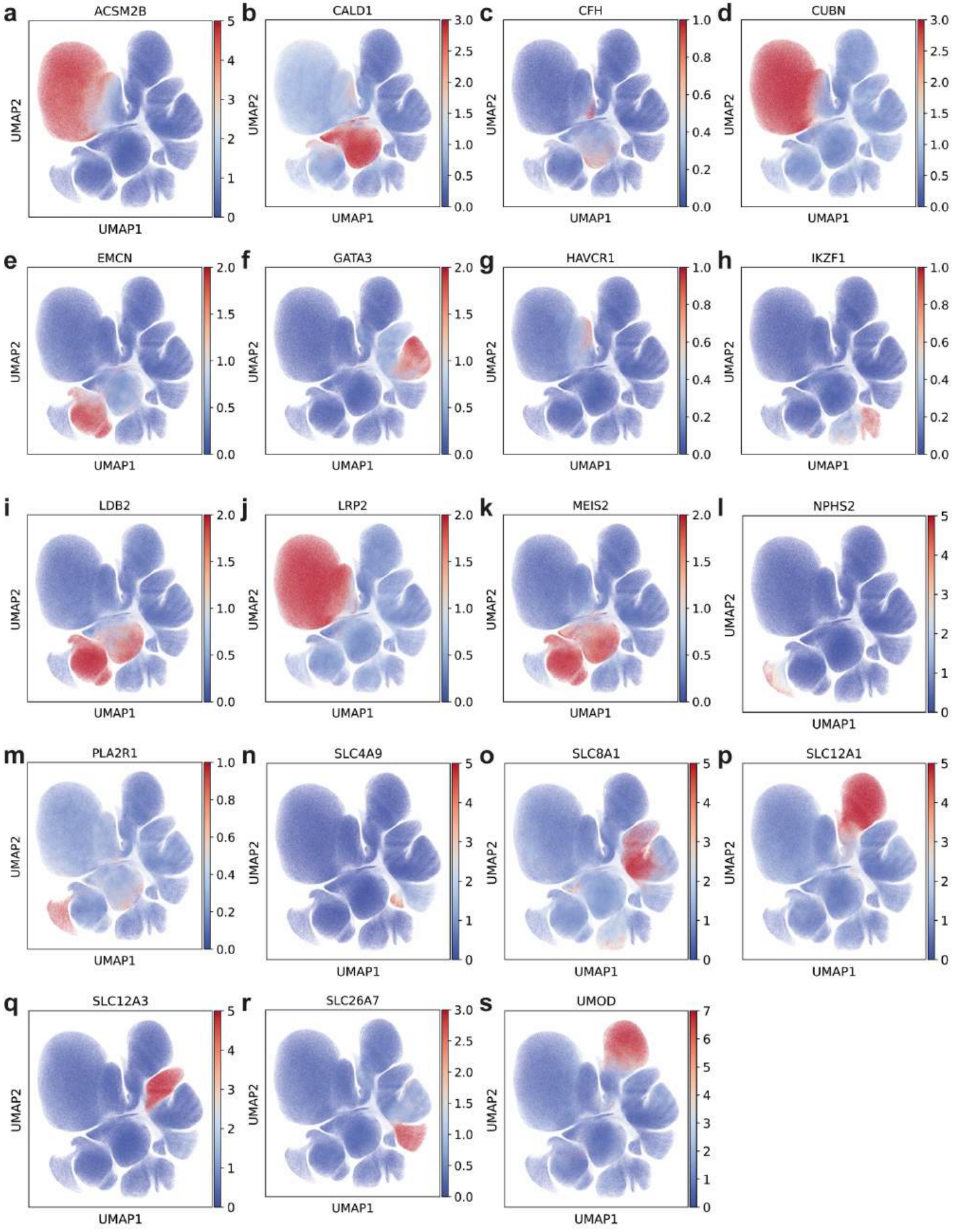
Supplementary Figure 6. Transcript feature counts per sample in snRNAseq dataset: Violin plots showing the distribution of total features per cell, grouped by samples for all samples in the single nuclear sequencing dataset. Each violin represents the count distribution within a single sample, with inner quartile lines indicating the interquartile range and median.



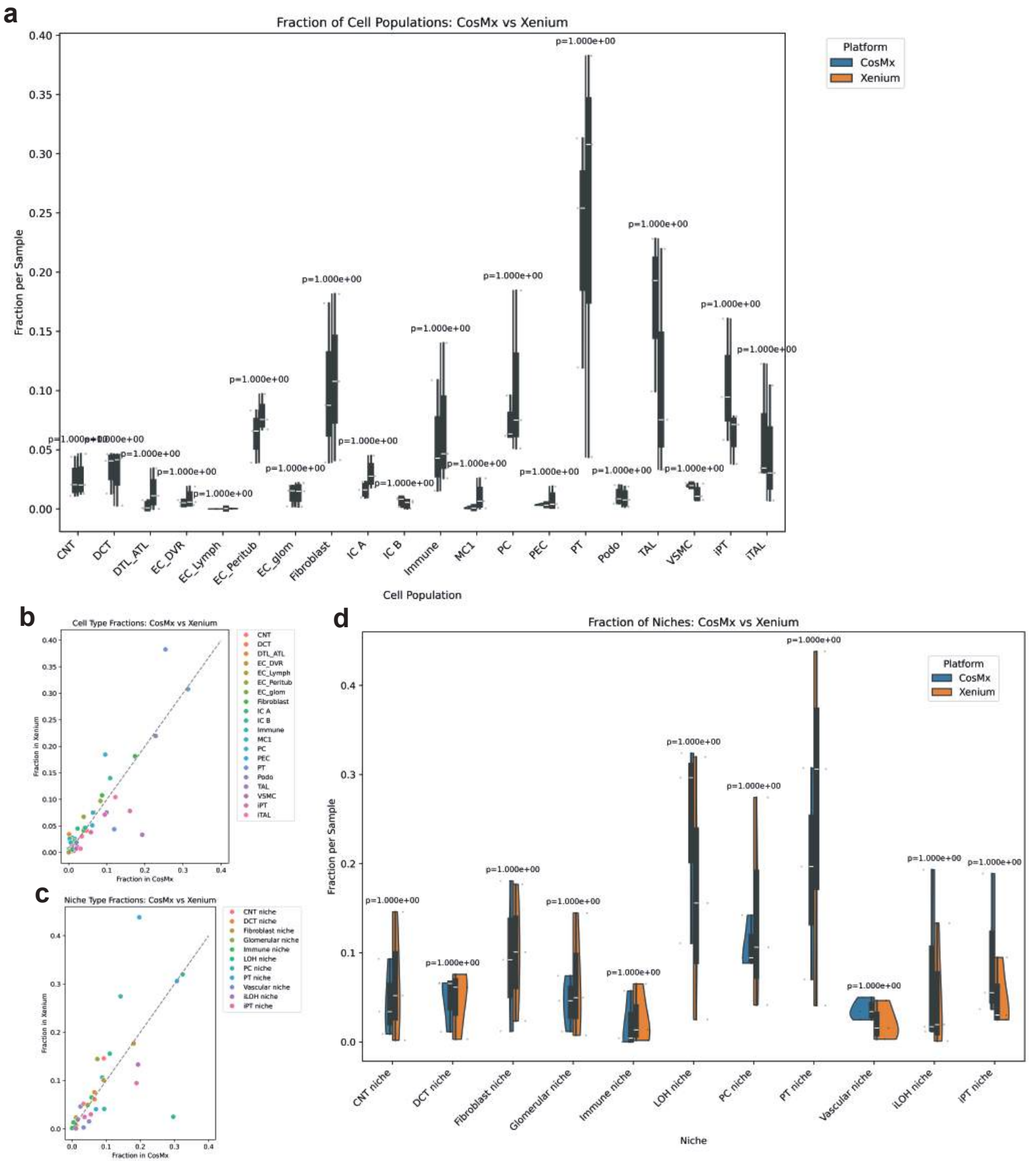
Supplementary Figure 7. Mitochondrial counts per sample in snRNAseq dataset: Violin plots showing the distribution of percentage of mitochondrial genes per cell, grouped by samples for all samples in the single nuclear sequencing dataset. Each violin represents the count distribution within a single sample, with inner quartile lines indicating the interquartile range and median.



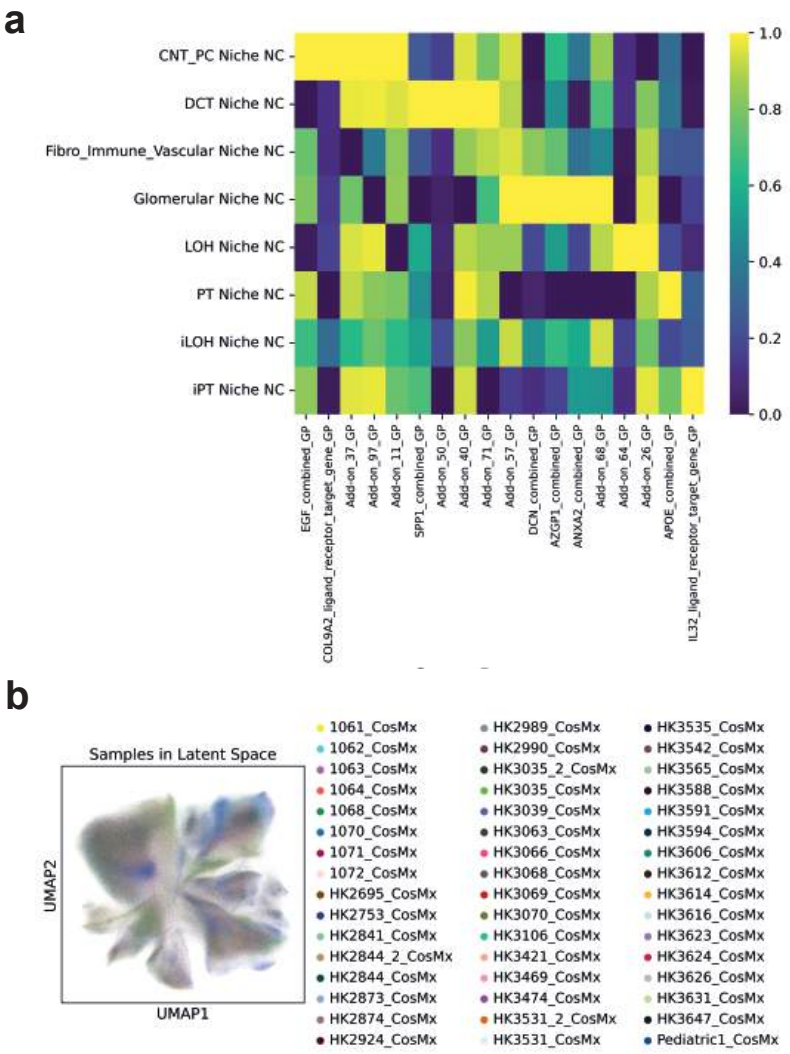
Supplementary Figure 8. Ribosomal counts per sample in snRNAseq dataset: Violin plots showing the distribution of percentage of ribosomal genes per cell, grouped by samples for all samples in the single nuclear sequencing dataset. Each violin represents the count distribution within a single sample, with inner quartile lines indicating the interquartile range and median.



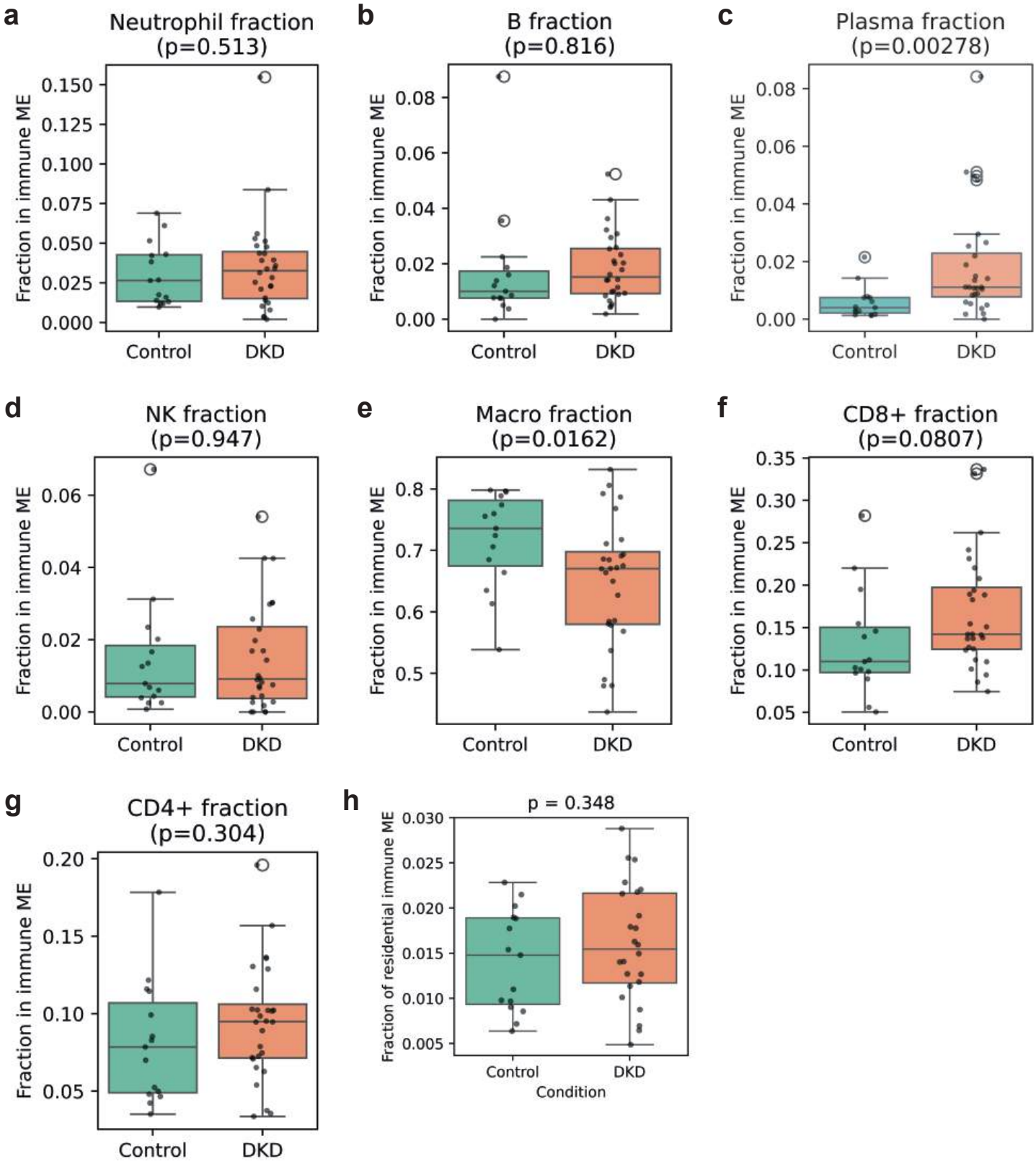
Supplementary Figure 9. Imputed gene expression of the spatial transcriptomics data. UMAP visualization showing the expression of ACSM2B (a), *CALD1* (b), *CFH* (c), *CUBN* (d), *EMCN* (e), *GATA3* (f), *HAVCR1* (g), *IKZF1* (h), *LDB2* (i), *LRP2* (j), *MEIS2* (k), *NPHS2* (l), *PLA2R1* (m), *SLC4A9* (n), *SLC8A1* (o), *SLC12A1* (p), *SLC12A3* (q), *SLC26A7* (r), *UMOD* (s) across all cells in the dataset. Color intensity represents imputed gene expression levels.



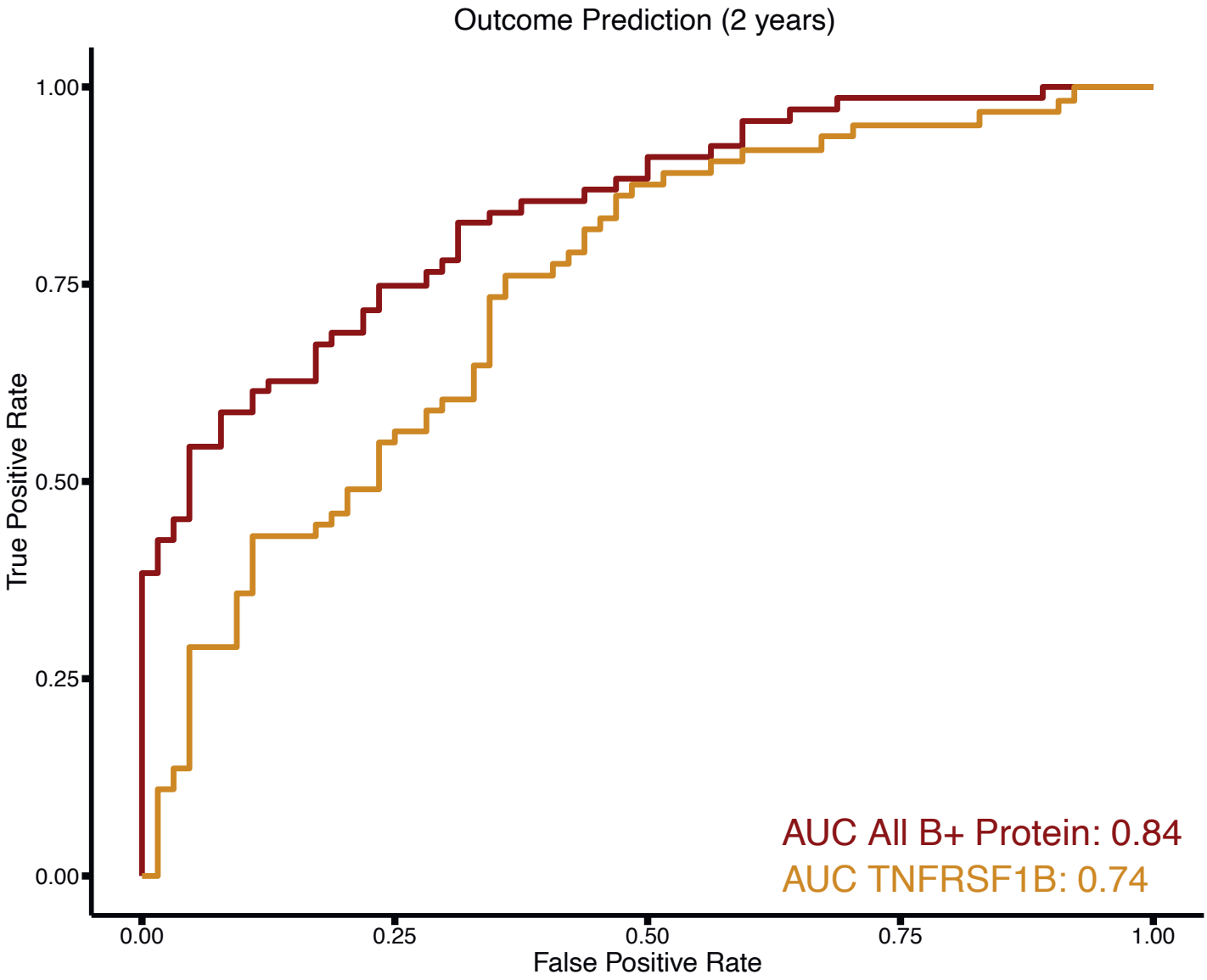
Supplementary Figure 10. Cell and Nichefractions CosMx vs. Xenium. To compare cell population representation across spatial transcriptomics platforms, we selected kidney samples profiled by both CosMx and Xenium (HK2695, HK3626, HK3106). Cells from these samples were subset from the full dataset, and the cell type annotation was retained. Each cell was labeled by its corresponding platform and patient identifier. For each sample and technology, we computed the fraction of cells assigned to each cell type by dividing cell type counts by the total number of cells per sample and platform. The resulting fractions were reshaped into a paired format to enable direct comparison between CosMx and Xenium measurements per cell type and donor. Statistical comparisons were performed using the two-sided Wilcoxon signed-rank test for each cell type. Fractions were visualized using split violin plots with overlaid strip plots to show paired values across platforms. Adjusted p-values were annotated for each cell population comparison. c, d: To compare niche representation across spatial transcriptomics platforms, we selected kidney samples profiled by both CosMx and Xenium (HK2695, HK3626, HK3106). Cells from these samples were subset from the full dataset, and the cell type annotation was retained. Each cell was labeled by its corresponding platform and patient identifier. For each sample and technology, we computed the fraction of cells assigned to each cell type by dividing cell type counts by the total number of cells per sample and platform. The resulting fractions were reshaped into a paired format to enable direct comparison between CosMx and Xenium measurements per cell type and donor. Statistical comparisons were performed using the two-sided Wilcoxon signed-rank test for each cell type. Fractions were visualized using split violin plots with overlaid strip plots to show paired values across platforms. Adjusted p-values were annotated for each cell population comparison.



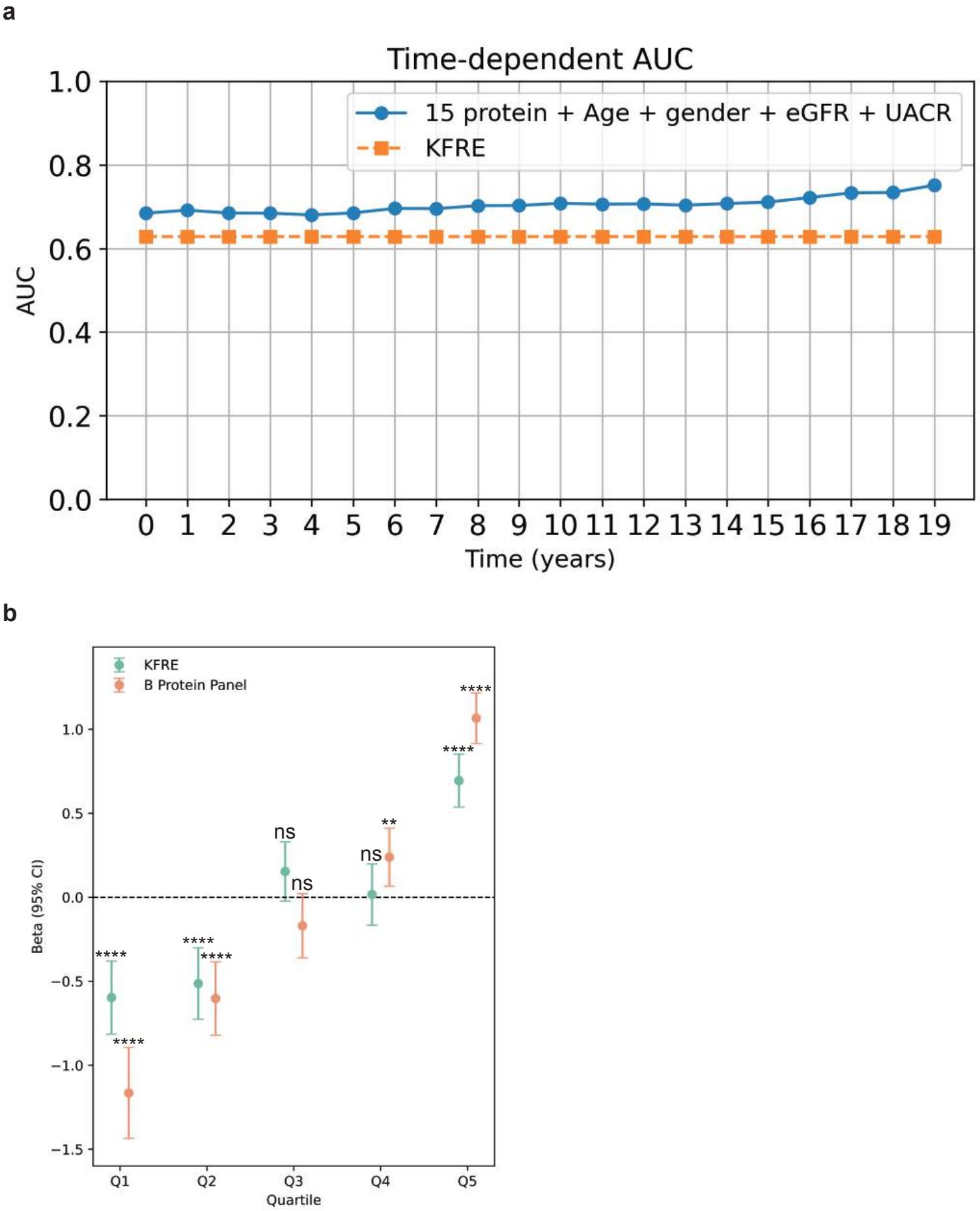
Supplementary Figure 11. Orthogonal niche definition and visualization using NicheCompass analysis. a: UMAP of integrated NicheCompass embeddings colored by celltypes. b: UMAP of integrated NicheCompass embeddings colored by samples. b: Gene programs (GPs) were inferred using NicheCompass and filtered based on activity across spatial neighborhoods. Differential GP testing was performed using a Bayesian framework, comparing each niche against all others (log Bayes factor > 2). The heatmap shows the normalized activity of significantly enriched GPs across annotated NicheCompass niches. Colors represent scaled average GP activity per niche.



Supplementary Figure 12. Immune cell fractions in residential immune microenvironment. Boxplots of immune cell subtype fractions in residential microenvironments from control (n = 15) and DKD samples (n = 29), normalized per sample and compared by two-sided Welch's t-test for a: Neutrophils, b: B cells, c: Plasma cells, d: NK cells, e: Macrophages, f: CD8+ T cells, g: CD4+ T cells. h: Boxplots of fraction of immune cells classified as residential immune cells of all immune cells, comparing Control samples (n = 15) vs. DKD samples (n = 29) using two-sided Welch's t-test.



Supplementary Figure 13. Plasma biomarkers in TRIDENT study. An elastic-net logistic regression classifier was trained, with class imbalance addressed by the Synthetic Minority Oversampling Technique (SMOTE). Model hyperparameters were tuned via 5-fold cross-validation to maximize AUROC. Proteins identified by the Elastic Net with non-zero coefficients in the refit model were incorporated as covariates in a Cox proportional hazards model of kidney survival (death, dialysis, or ≥40% eGFR decline). Model performance was compared to TNFRSF1B, a known biomarker of DKD.



Supplementary Figure 14. UKBB external validation: External validation of the circulating biomarker model in 3,309 UK Biobank participants with diabetes mellitus. Time-dependent AUC curves are shown for the Kidney Failure Risk Equation (KFRE; orange) and the protein-plus-clinical parameter model (blue), with the latter achieving higher discrimination (AUC = 0.7047 vs. 0.6286). b: Quartile-based β-coefficients (±95% CI) from Cox proportional hazards models for kidney disease progression, defined as death, ≥40% decline in eGFR, end-stage renal disease (ESRD), initiation of dialysis, or kidney transplantation. Results are shown for the B DKD protein set (orange) and the Kidney Failure Risk Equation (KFRE; green). Higher quartiles correspond to increasing biomarker or KFRE score levels.