Supplemental Materials

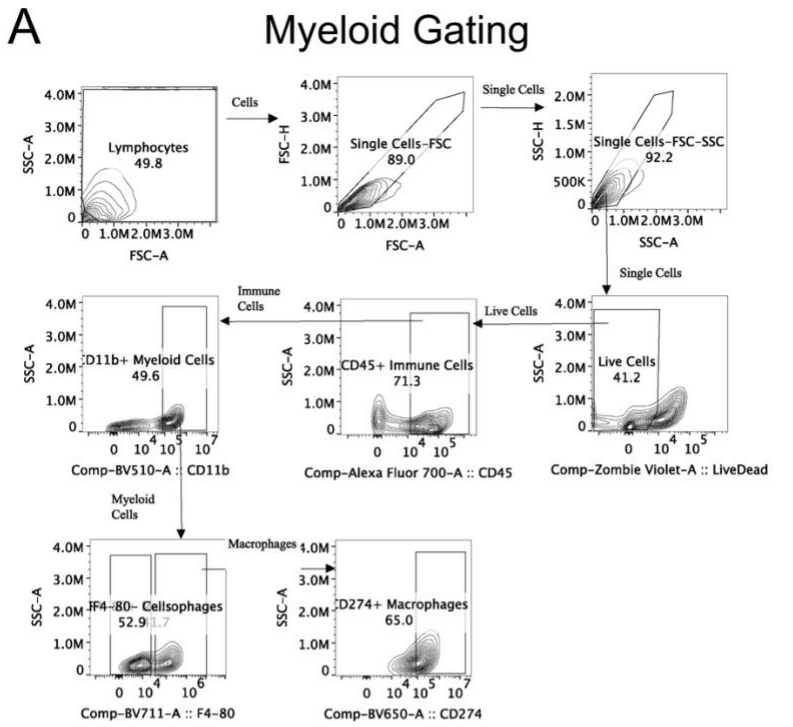
Synthetic Immunological Niche Reveals Early Immune Dysregulation and Stratifies Therapeutic Response in Type 1 Diabetes

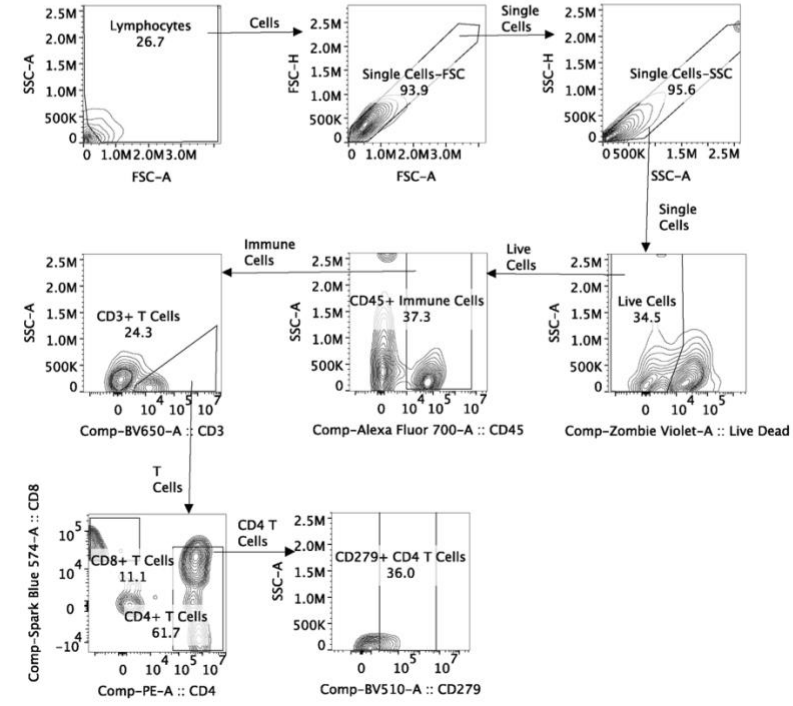
Jyotirmoy Roy[1], Yifei Jiang[1], Runbo Mao[2], Jessica L. King[1], Amod Talekar[1], Donglin Liu[2], Lillian Holman[1], Haoxuan Zeng[2], Xin Luo[2], Laila M. Rad[1], Peter Sajjakulnukit[3], Brianna Ha[4], Kai Liu[2], Elizabeth J. Bealer[1], Bryan Wonski[1], Kathryn Kang[1], Sean T. Carey[1], Luke F. Bugada[1], Brooke M. Smiley[1], Shahzad Sohail[1], Antonio Holmes[1,5], Dominik Awad[3,7], Aaron H. Morris[1], Costas A. Lyssiotis[3,6,7], Jie Liu[2], Lonnie D. Shea[1,4,8,*]

Affiliations:

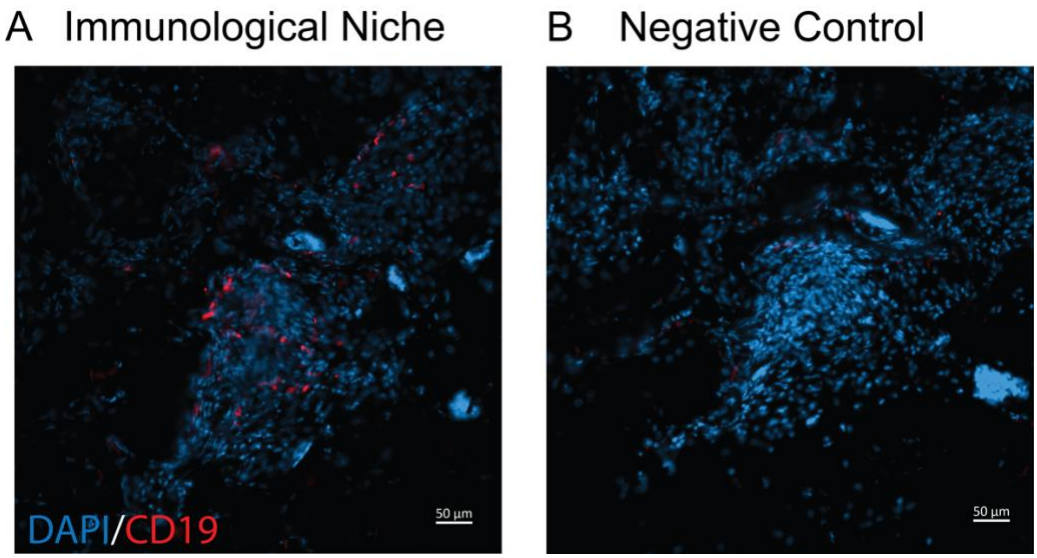
[1] Department of Biomedical Engineering, University of Michigan, Ann Arbor, USA
[2] Gilbert S. Omenn Department of Computational Medicine & Bioinformatics, University of Michigan, Ann Arbor, USA
[3] Department of Molecular & Integrative Physiology, University of Michigan, Ann Arbor, USA
[4] Department of Chemical Engineering, University of Michigan, Ann Arbor, USA
[5] Michigan Institute for Clinical & Health Research, University of Michigan, Ann Arbor, USA
[6] Department of Internal Medicine, Division of Gastroenterology and Hepatology, University of Michigan, Ann Arbor, USA
[7] Rogel Cancer Center, University of Michigan, Ann Arbor, USA
[8]Department of Surgery, University of Michigan, Ann Arbor, USA
* Corresponding author: ldshea@umich.edu

Supplemental Figures

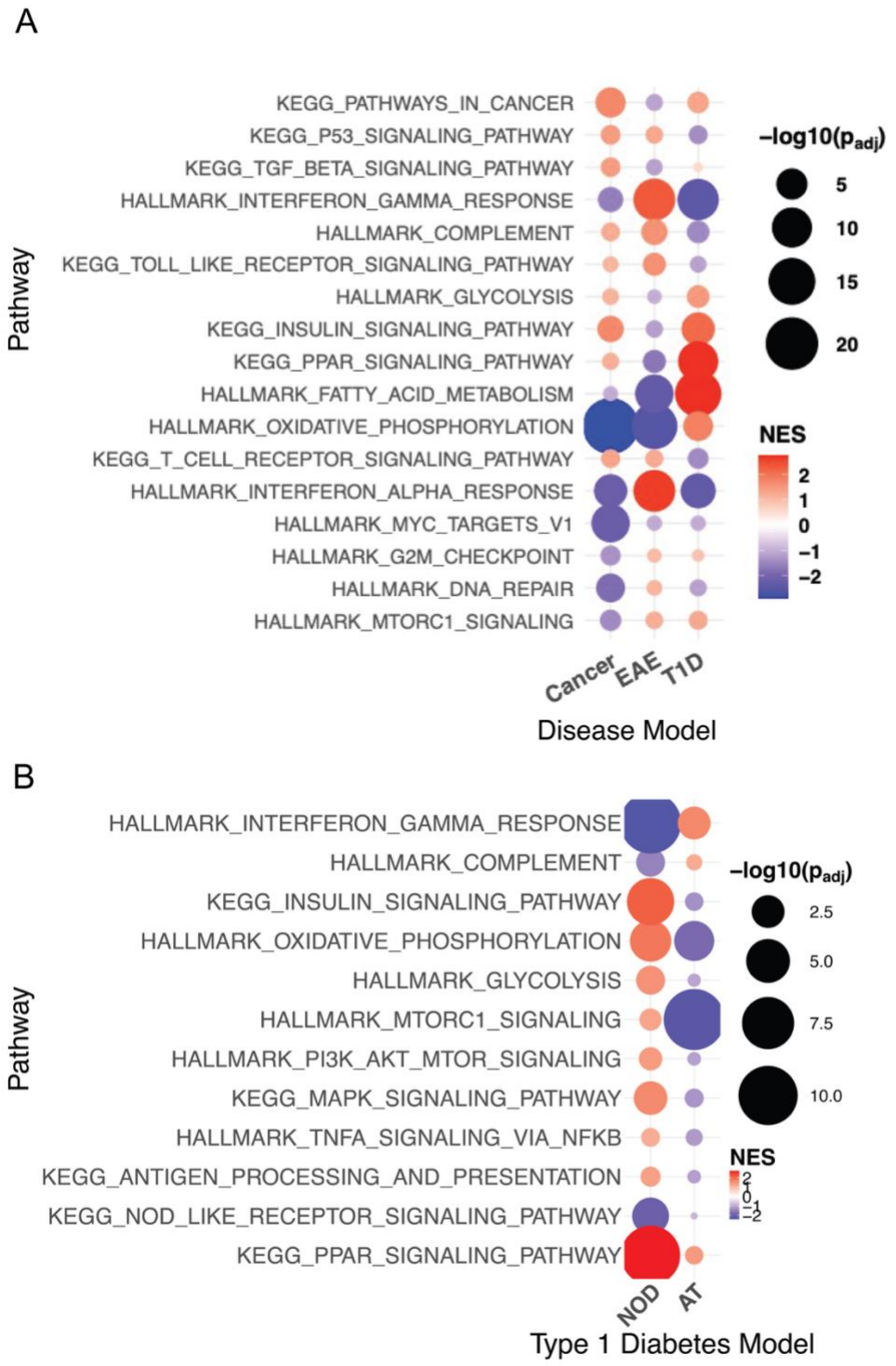




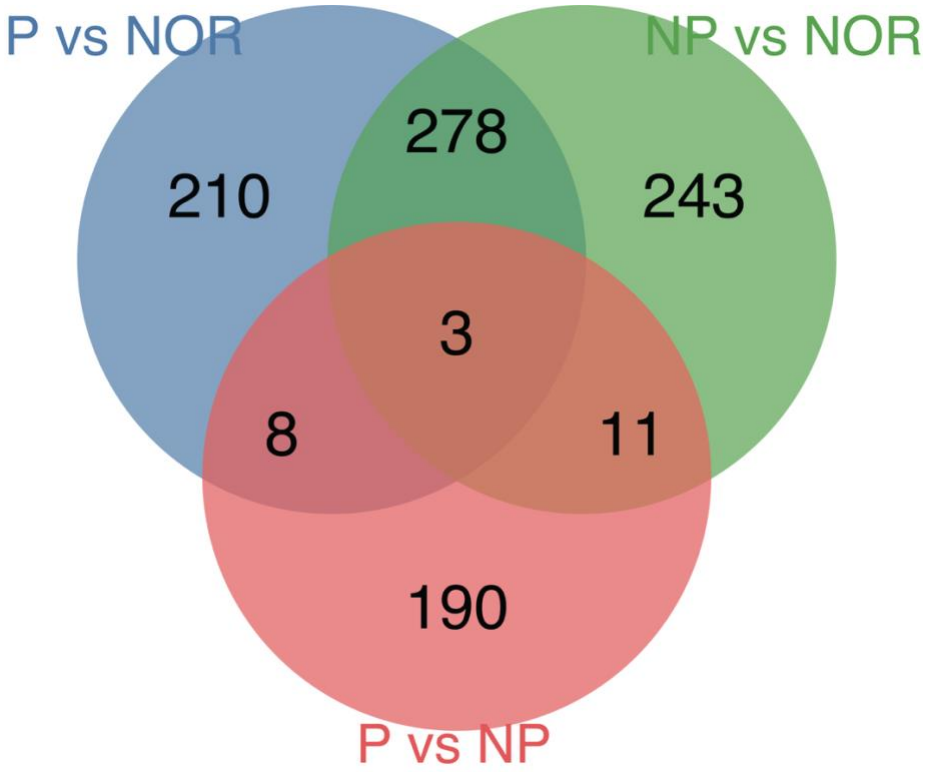
Supplemental Figure 1. Flow Cytometry Gating Strategy. Flow cytometry gating strategy for (**A**) Myeloid and (**B**) Lymphoid panel.



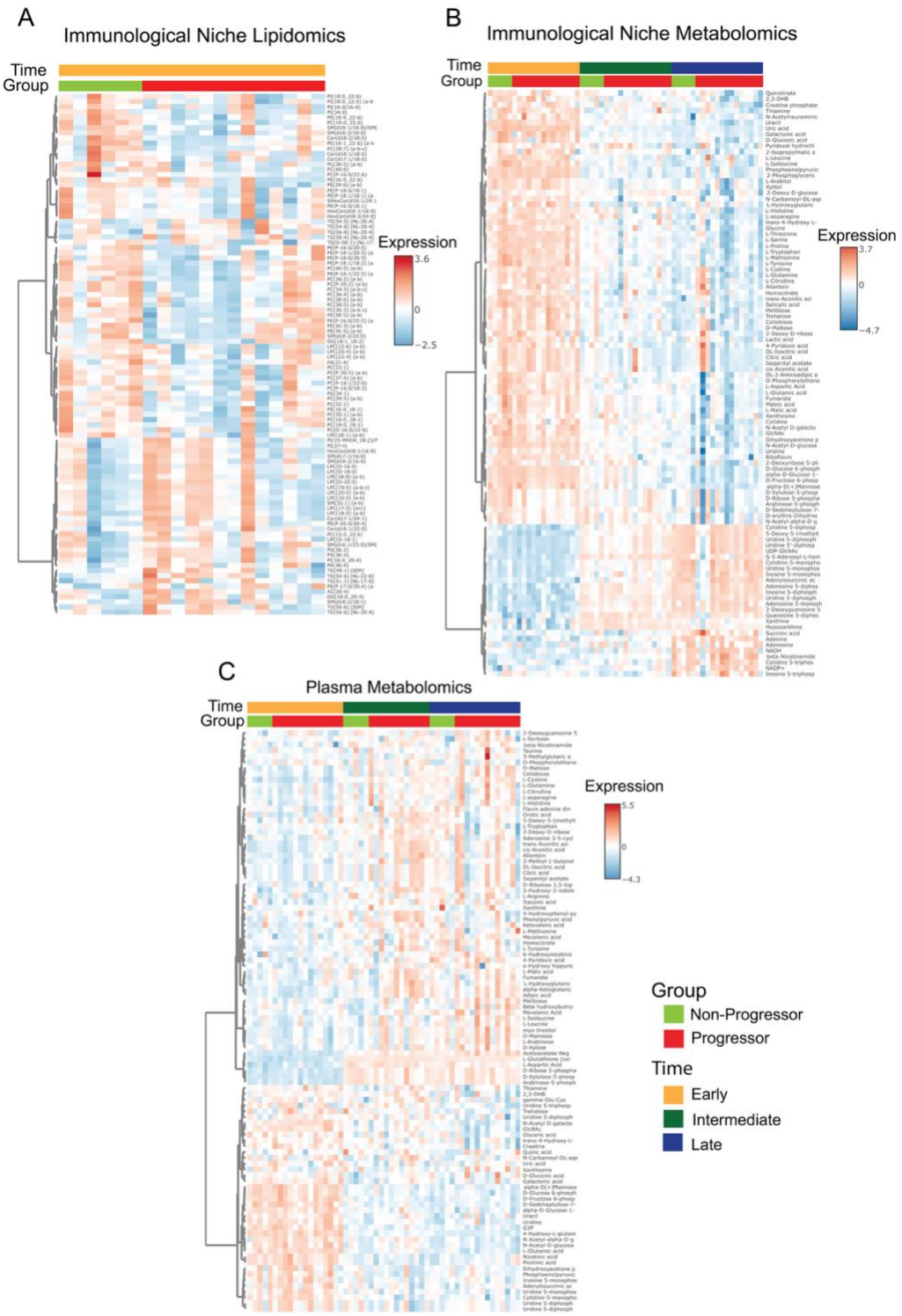
Supplemental Figure 2. Immunohistochemistry of Immunological Niche Reveals Presence of B Cells. (**A**) IHC Staining for CD19 shows presence of B cells in the immunological Niche. (**B**) No CD19 signal was observed in negative control (Scale bar, 50µm).



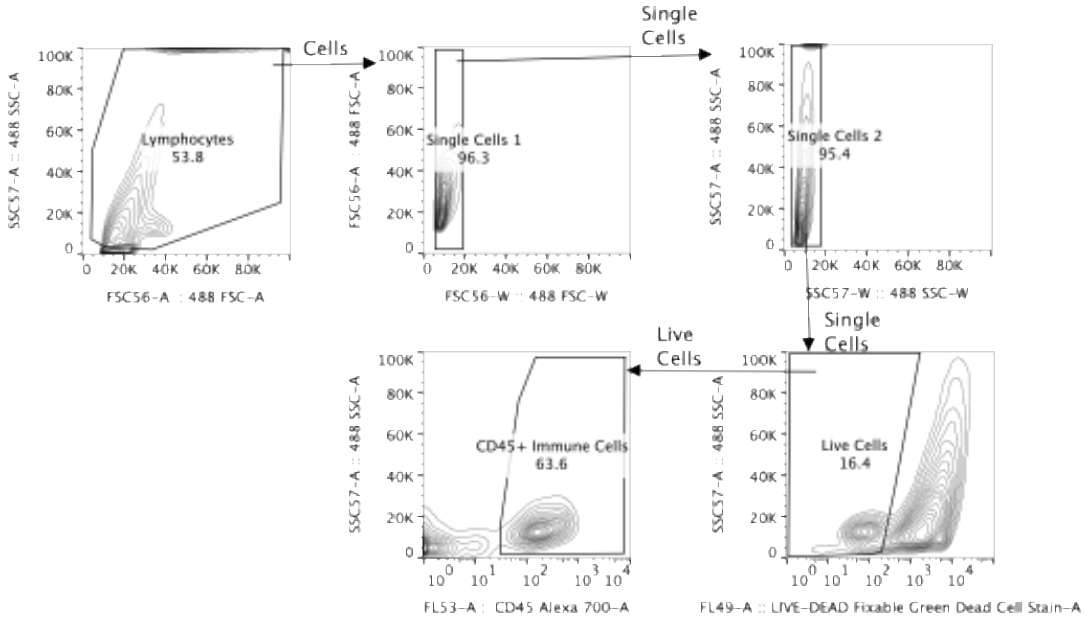
Supplemental Figure 3. Geneset Enrichment Analysis of IN transcriptomics- Dotplot showing Normalized Enrichment Score and significance for selected pathways for- a) Early Stage NOD T1D vs 4T1 Breast Cancer and Experimental Autoimmune Encephalomyelitis. Inflammation disease comparison b) Early Stage NOD vs Adoptive Transfer (AT) T1D models.



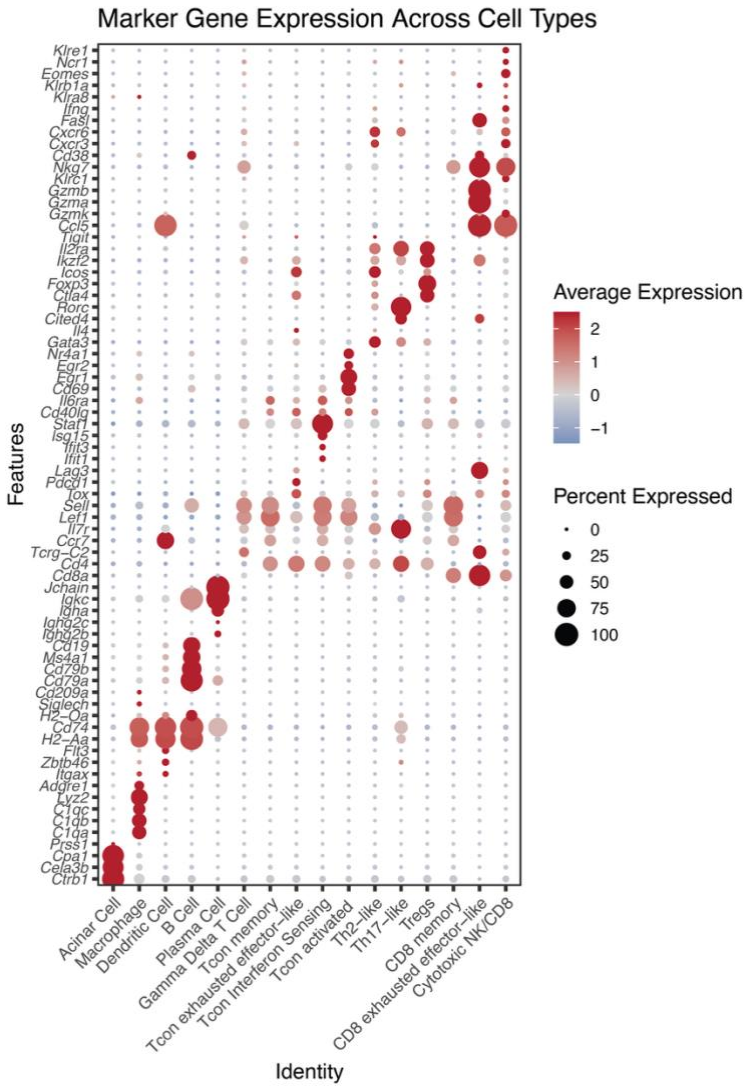
Supplemental Figure 4. Comparison of IN based Differentially Expressed Genes (DEGs) in NOR vs NOD. Venn Diagrams comparing the differentially expressed genes between NOR (n = 4), NOD progressors (P, n = 24) and non-progressors (NP, n = 16) at Early Stage (6 weeks of age).



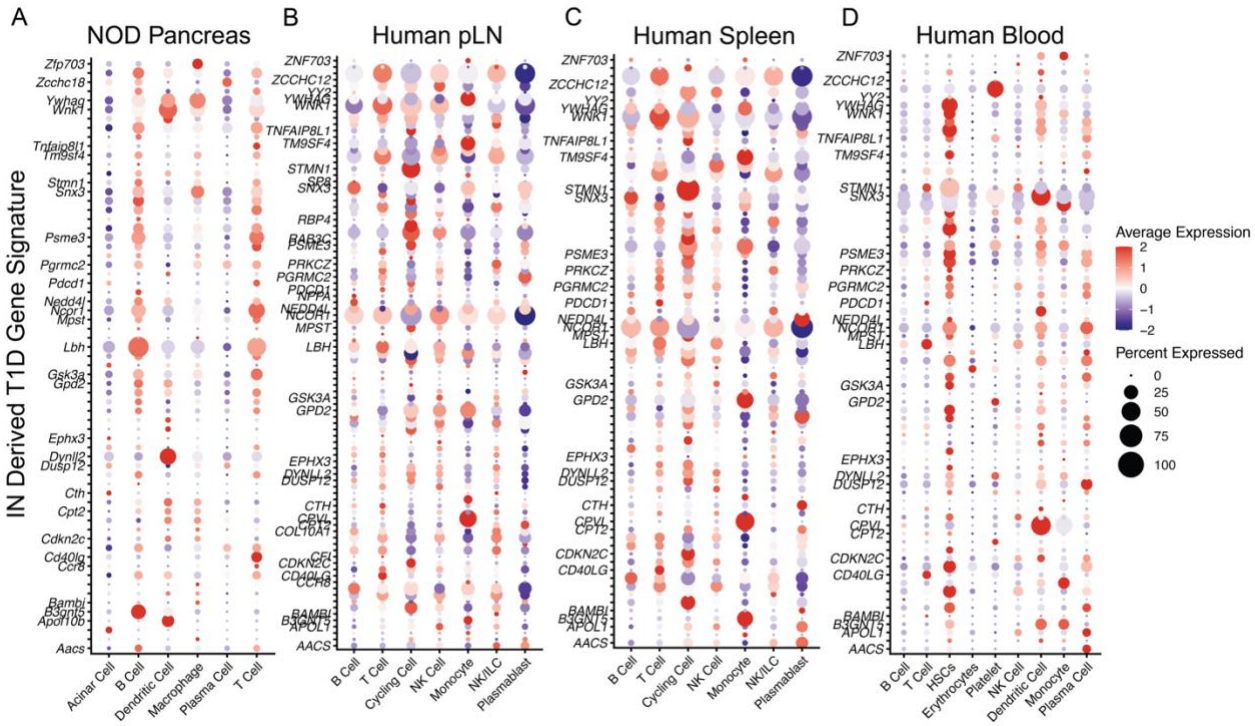
Supplemental Figure 5. Lipidomic and Metabolomic profiling of the Immunological Niche and Plasma. (**A**) Lipid profiles at the early stage between NOD progressors (n = 13) vs non-progressors (n = 6). (**B-C**) Polar metabolite profiles across early, intermediate, and late stages in progressors and non-progressors: (**B**) IN (n=14/5 at each stage) and (**C**) plasma (n = 14/5, 12/5, and 13/5, respectively).



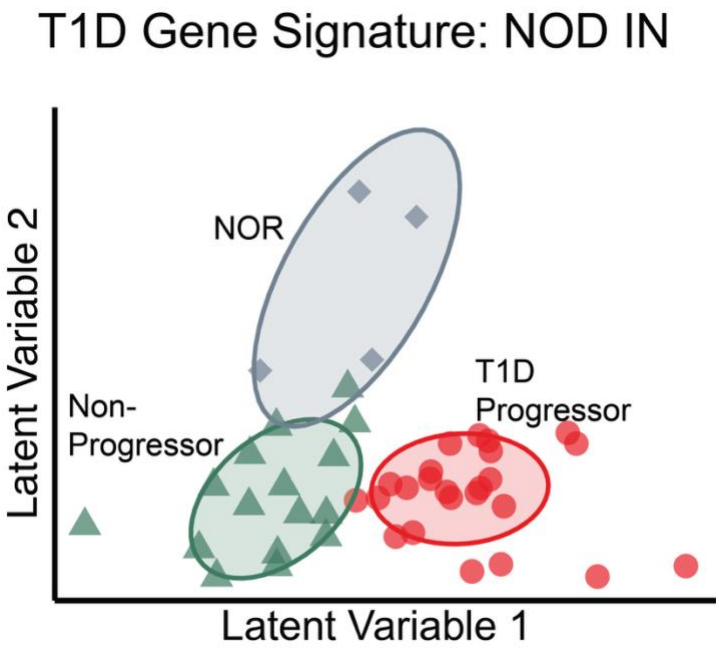
Supplemental Figure 6. Gating Strategy for NOD Pancreas Immune Cell Sorting using FACS. Immune cells were sorted based of Live CD45+ gating.



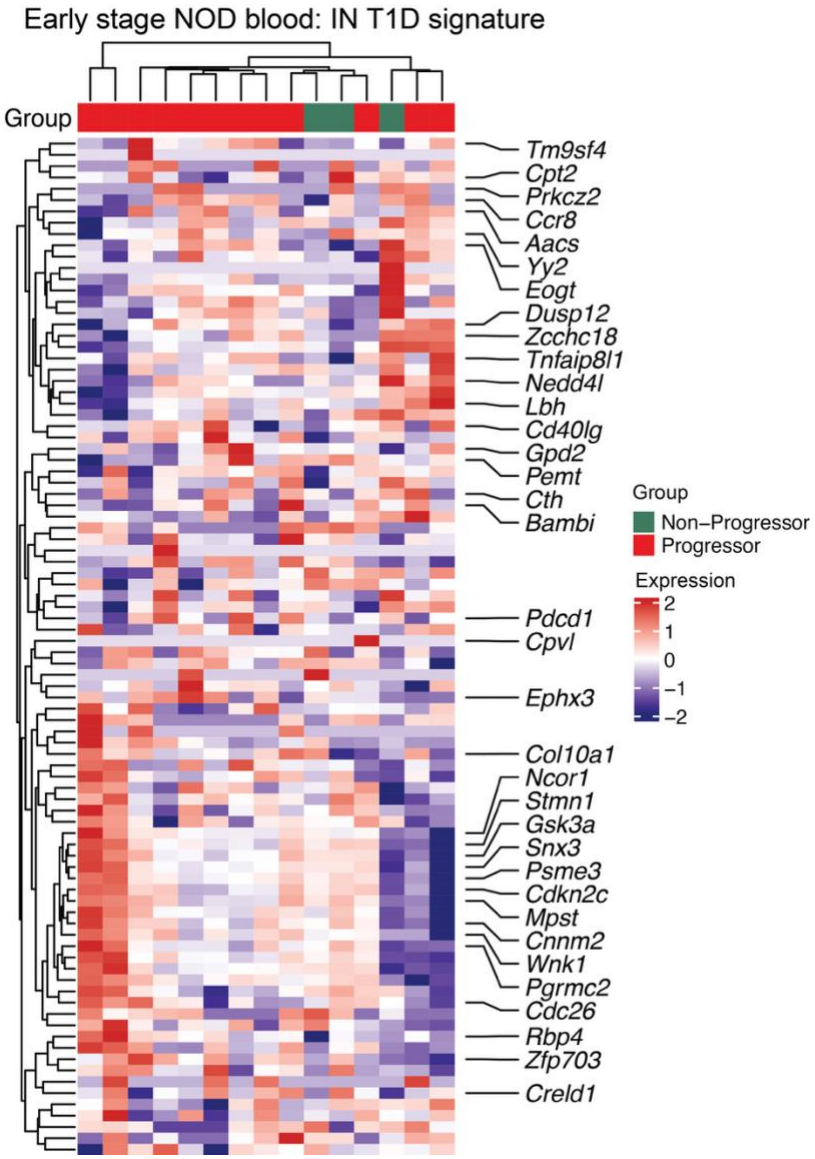
Supplemental Figure 7. Dot plot of canonical marker genes used for annotation of pancreas cell types. Scaled expression levels and frequency of canonical genes are shown for each identified NOD pancreatic cells.



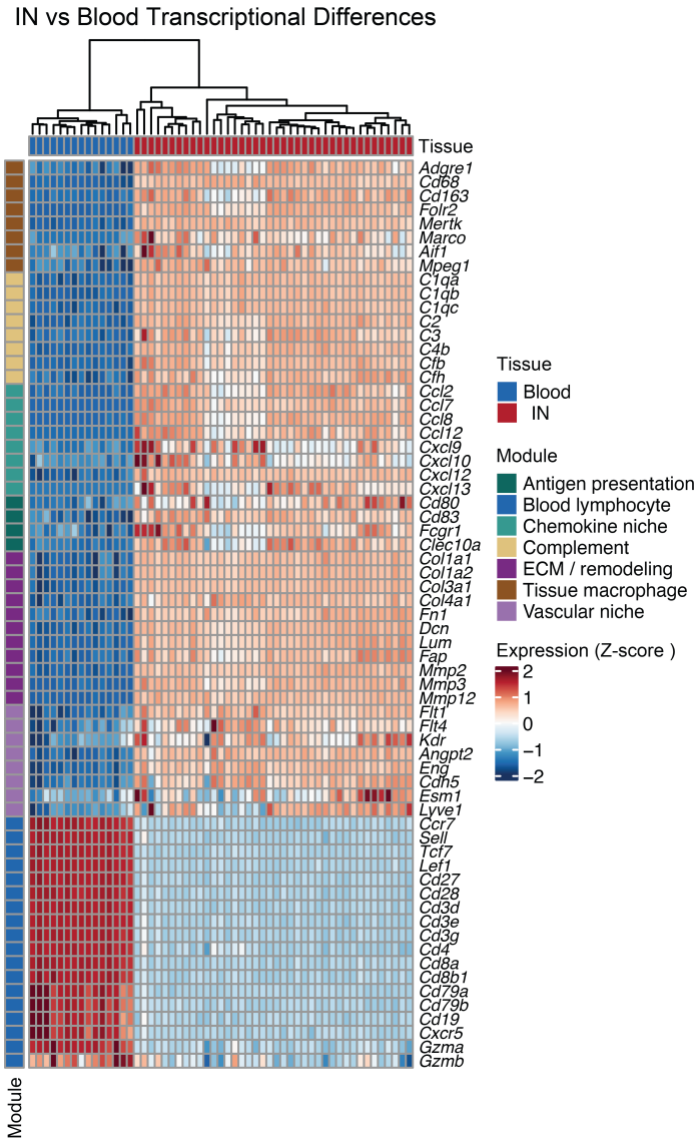
Supplemental Figure 8. Mapping of IN derived T1D gene signature. Scaled expression levels and frequency of T1D gene signature (or their human orthologs) for a) NOD Pancreas (n=11 mice), and b) Human pLN (n=19 donors), c) spleen (n=17 donors), d) PBMCs (n=77 donors) from T1D and non-diabetic individuals



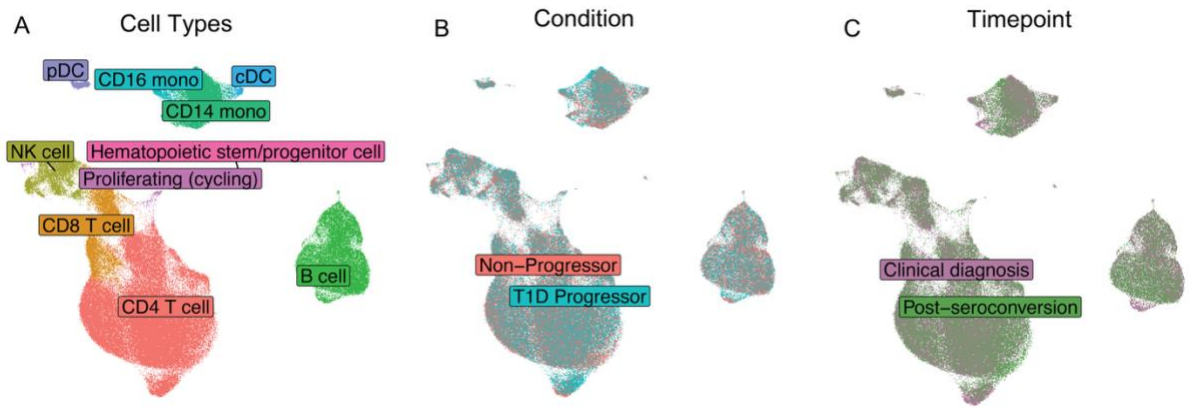
Supplemental Figure 9. PLS-DA of IN T1D gene signature with Non-obese Resistant (NOR) group. Two-dimensional PLS-DA score plot based on the IN-derived T1D gene signature in early-stage NOD progressors (n = 24), NOD non-progressors (n = 16), and diabetes-resistant NOR mice (n = 4). Each point represents an individual mouse. Ellipses indicate ~68% multivariate normal contour for each group.



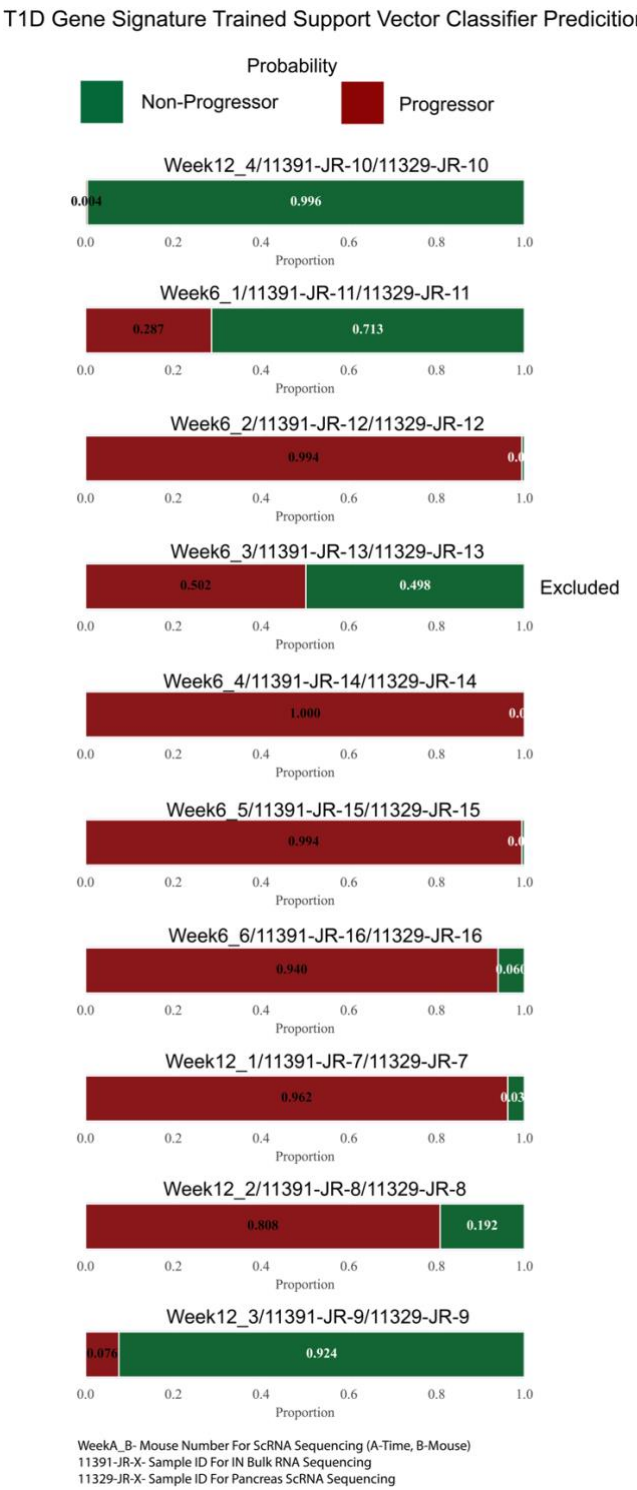
Supplemental Figure 10. Expression of IN T1D gene signature in NOD Blood at early stage. Heatmap of 90 out of the 100 IN-based T1D signature genes at early stage in blood from NOD progressors (n = 12) vs non-progressors (n = 3). Heatmap colors represent row-wise Z-scored expression from variance-stabilized data.



Supplemental Figure 11. Transcriptomic differences between the IN and blood in early-stage NOD mice. Heatmap showing genes associated with immune and tissue-remodeling programs in IN (n = 40) and blood samples (n = 15) from NOD mice at early stage. Colors represent row-wise Z-scored expression calculated from variance-stabilized data.



Supplemental Figure 12. Longitudinal single-cell transcriptomic analysis of human blood from the TRIGR clinical trial. UMAP projections colored by (**A**) annotated cell type; (**B**) disease status; and (**C**) sampling time point. Blood samples are from T1D progressors and age-matched non-progressors at post-seroconversion (progressor/non-progressor, n = 45/44) and clinical diagnosis (n = 46/46).



Supplemental Figure 13. Prediction of NOD as Progressor vs Non-Progressor Using IN Derived T1D 100 Gene Signature. Probability of being Progressor as predicted by the Support Vector Classifier trained on IN Derived T1D 100 Gene Signature. Samples classified with >60% probability were taken for downstream analysis of the pancreas single cell RNA sequencing data.

Supplemental Methods

Animal Study Design

For the EAE model, the INs were used in murine experimental autoimmune encephalomyelitis as previously described (1). In brief, EAE was induced by adoptive transfer of T cells reactive to myelin oligodendrocyte glycoprotein (MOG) peptide 35 to 55 and control animals received an adoptive transfer of T cells reactive to ovalbumin (OVA) peptide 323 to 339. Mice received a subcutaneously implanted IN 14 days before the adoptive transfer of T cells on Day 0. INs were explanted 6 days after the adoptive transfer before clinical symptoms were present. For the 4T1 model, female BALB/c mice (7-8 weeks) underwent subcutaneous IN implantation followed by orthotopic mammary fat pad inoculation 4T1 cells, with tumor progression monitored for 21 days, as previously described (2). The adoptive transfer T1D model was established by activating splenocytes from 6-week-old NOD BDC2.5 donor mice with cognate peptide in vitro, followed by intraperitoneal transfer of 5×10^5 activated cells into age-matched NOD.scid recipients, as previously described (3). The INs were explanted at the time of transfer and 10 days post-transfer.

Bulk RNA Seq Analysis

Frozen immunological niche were lysed in 800 µL of TRIzol reagent (Fisher Scientific, Cat. No. 15596026) and homogenized at 15,000 rpm. RNA was isolated from the homogenized solution with a direct-zol RNA Miniprep Plus (Fisher Scientific, Cat. No. NC1047980) kit following the manufacturer's instructions. The concentration of the isolated RNA was measured using a NanoDrop and the samples were submitted to the Advanced Genomics Core at the University of Michigan for bulk RNA sequencing. Library preparation was performed using ribosomal RNA

depletion for total RNA library construction. Sequencing was carried out on the Illumina NovaSeqXPlus. BCL Convert Conversion Software v4.0 (Illumina) was used to generate de-multiplexed Fastq files.

Raw gene-level count matrices from three independent bulk RNA-seq cohorts were jointly processed for differential gene expression (DEG) analysis. Data for two of these cohorts were obtained from previously published work in King et al.(4). Samples were stratified by disease stage as follows: early (6 weeks of age; 16 non-progressors and 24 Progressors), intermediate (12-14 weeks; 10 non-progressors and 10 progressors), and late (17-20 weeks; 4 non-progressors and 10 progressors). All downstream analyses were performed independently within each disease stage. For each stage, genes that were unexpressed across all samples or sparsely detected (zero counts in >85% of samples) were removed. To restrict analyses to well-annotated and biologically interpretable transcripts, genes annotated as RIKEN or pseudogenes were excluded prior to differential expression testing. Differential expression analysis was performed using DESeq2 (v1.50.2), incorporating cohort batch as a covariate in the design formula. Log_2 fold changes were estimated for progressors relative to non-progressors using the Wald test. Genes with an absolute log_2 fold change ≥ 1 and Wald p-value ≤ 0.05 were considered differentially expressed. Gene set enrichment analysis (GSEA) was performed separately for each disease stage using clusterProfiler, with gene sets obtained from MSigDB, including Hallmark and KEGG pathway collections (5). All the genes were ranked by the DESeq2 Wald statistic (Progressor vs. Non-Progressor) and pathways with adjusted p ≤ 0.1 and absolute normalized enrichment score (|NES|) ≥ 1 were considered significantly enriched. Complete GSEA results for all stages are provided in Supplemental Table 1**.** For visualization of pathways, a curated subset of immune, signaling, and metabolic pathways that were significantly enriched in at least one disease stage were selected.

To assess the disease specificity of differentially expressed genes (DEGs), we analyzed IN-associated transcriptomic datasets from additional inflammatory disease models and compared them to the early stage T1D niche signature. For the 4T1 breast cancer model, bulk RNA-seq data were obtained from Orbach et al. and differential expression at IN was assessed between healthy and tumor-bearing mice at the pre-symptomatic Day 14 timepoint (2). For the Experimental Autoimmune Encephalomyelitis (EAE) model, IN samples collected at the pre-symptomatic Day 6 stage were similarly compared between diseased and healthy conditions. DEGs and enriched pathways for these models were identified using the same analytical pipeline described above. DEGs and pathway enrichment results from these datasets were compared specifically to the early-stage T1D DEGs. In addition, for comparison with a highly T cell-driven T1D context, adoptive transfer model data were obtained from King et al.(3), in which NOD.scid mice received either diabetogenic BDC2.5 T cells or ovalbumin (OVA) control T cells. Differential expression analysis was performed on niche samples collected at day 10 post-transfer, and the resulting gene and pathway signatures were compared against early-stage T1D results. DEGs and enriched pathways from the 4T1 cancer, EAE and Adoptive Transfer T1D model are available in Supplemental Table 2.

For identifying an early transcriptional signature predictive of T1D progression, filtered and limma based batch corrected genes that were significantly differentially expressed between progressor and non-progressor at the early stage (Welch p-value ≤ 0.05) were selected as input features to a Partial Least Squares–Discriminant Analysis (PLS-DA) model (6). Top 100 genes with the highest Variable Importance in Projection (VIP) scores were selected as the core discriminatory signature, which robustly separated progressors from non-progressors at the early timepoint.

This gene signature was subsequently used to train a Support Vector Classifier (SVC) model to predict T1D progression at early stage. All machine learning analyses were conducted in Python (v3.11.1) using scikit-learn (v1.5.1) and pandas (v2.2.2). The dataset comprising 40 samples was stratified into a training and validation set into a 7:3 ratio using a stratified sampling method to preserve class balance. Model hyperparameters were optimized using a 5-fold cross validation on the training data and a linear SVC kernel, with regularization parameter C = 1 was selected. Model performance was subsequently evaluated across 1000 independent stratified train validation sets to assess robustness. This model was used to infer progression status of animals used for the pancreas single-cell sequencing using their early-stage niche-derived bulk RNA-sequencing profile. Out of 10 mice (6 harvested for pancreas was extracted at early timepoint and 4 for intermediate timepoint), 9 were confidently (>60% probability) predicted as progressor or non-progressor and used for downstream analysis of single cell RNA analysis. The progression assignments are summarized in Supplemental Figure 13.

Differential gene expression and gene set enrichment analyses for the early anti-TNF-α-treated cohort were performed using the same pipeline described above. To derive an Anti-TNF-α Response Score (ATRS), pathway-level activity scores were computed on a per-sample basis using Gene Set Variation Analysis (GSVA). GSVA scores for pathways associated with inflammatory signaling and immune-metabolic regulation that were significantly enriched (adjusted $p \leq 0.1$ and $|NES| \geq 1$) in both pre- and post-treatment comparisons were used as input features for an elastic-net logistic regression model ($\alpha = 0.7$), trained to distinguish anti-TNF responders from non-responders at the early treatment time point. Model regularization and feature selection were performed using leave-one-out cross-validation. The final model selected three pathways-KEGG_PEROXISOME, HALLMARK_TGF_BETA_SIGNALING, and HALLMARK_TNFA_SIGNALING_VIA_NFKB-which together defined the ATRS. The resulting model assigned positive weights to TNF-α/NF-κB and TGF-β signaling and a negative weight to

peroxisome metabolism. Sample-level ATRS values were calculated as a linear combination of GSVA pathway scores weighted by the corresponding model coefficients. Full input pathway selection list, model coefficients, and ATRS values are provided in Supplemental Table 6.

Human Tissue Single Cell Pre-Processing

We analyzed HPAP spleen and pancreatic lymph node (PLN) snMultiome (RNA+ATAC) datasets (7); no standalone scRNA-seq datasets were available for these tissues. Downstream analyses in this study (dimensional reduction, clustering, cell-type annotation, and differential expression) were performed using the RNA modality only. However, because the raw data were generated with the snMultiome protocol, nucleus identification and quality control were performed using joint RNA and ATAC metrics to mitigate modality-specific artifacts and ensure consistent filtering across heterogeneous donors and processing centers. Accordingly, we describe ATAC preprocessing and ATAC-derived QC metrics only to the extent that they informed nucleus-level QC and filtering.

RNA reads were aligned and quantified at the donor level, followed by barcode calling and decontamination. Putative cell-containing barcodes were identified using emptyDrops from DropletUtils (8). To mitigate ambient RNA contamination, CellBender “remove-background” outputs were loaded from the CellBender H5, retaining inferred latent variables. In the joint QC stage, we used CellBender-derived quantities such as cell probability and per-barcode post-decontamination UMI counts to compute the fraction of ambient signal removed and to support RNA-side filtering (9).

ATAC reads were adapter-trimmed (CTA), assessed with FastQC/MultiQC, and aligned to the reference genome using BWA-MEM with coordinate sorting by SAMtools (10–12). Read groups belonging to the same library were merged, duplicates were marked with Picard, and a high-

confidence ATAC alignment set was obtained by retaining properly paired autosomal reads with MAPQ ≥ 30 while removing unmapped, secondary, duplicate, and supplementary alignments. Broad ATAC peaks were called with MACS2, and these peaks together with the processed BAMs were used as inputs to ataqv to compute ATAC QC summaries, including TSS enrichment, mitochondrial read fraction, and high-quality autosomal accessibility (HQAA) as well as the data-driven suggested HQAA cutoff reported by ataqv, enabling sample-adaptive ATAC quality filtering across heterogeneous donors/centers (13, 14).

RNA and ATAC metrics were then integrated per nucleus, producing a single metrics table per donor that retained both modalities (RNA UMI depth, RNA %mitochondrial, CellBender-derived metrics; ATAC HQAA, ATAC TSS enrichment, ATAC %mitochondrial). Because HPAP snMultiome data are not uniformly generated across donors and centers, we avoided fixed global cutoffs and instead estimated several key thresholds per donor using data-driven procedures. In particular, we computed modality-aware mitochondrial thresholds by first diagnosing whether the underlying distribution was unimodal or multimodal via KDE peak counting, then applying either 2D histogram smoothing and Multi-Otsu segmentation to separate foreground high-quality nuclei from background in the joint space of accessibility/UMIs vs. %mitochondrial, or fallback 1D Multi-Otsu thresholding when the 2D segmentation was unstable. This same philosophy was used to set additional RNA-side cutoffs using Multi-Otsu on the observed distributions. We also performed a refined knee-plot analysis of barcode-rank vs. RNA UMI curves, including interpolation on a log scale and Savitzky–Golay smoothing to stabilize slope-based cliff detection in noisy donors.

After applying the RNA- and ATAC-side preprocessing and the joint, per-donor QC filtering, doublets were annotated on the QC-passing snMultiome nuclei using two complementary strategies to reduce modality-specific artifacts: AMULET and DoubletFinder (15). Concordance between the two calls was used to summarize doublet burden and to support downstream

cluster-level artifact screening. The filtered nuclei were then processed in Seurat using normalization, variable feature selection, PCA with PC selection guided by elbow plots, and initial UMAP clustering to establish baseline structure prior to correction. To mitigate non-biological structure driven by donor and processing covariates, we applied Harmony on the PCA space using covariates including HPAP donor ID, sex, BMI, age, ethnicity, and relocation center, producing corrected embeddings. Integration performance was assessed by comparing pre- and post-Harmony diagnostics, with a primary criterion that no cluster was dominated by a single metadata category, while preserving expected biological marker patterns (16). Finally, low-quality or artifactual clusters were flagged for removal based on high doublet proportion together with incoherent marker expression, and by abnormal cluster-level QC distributions defined as follows: for %chrMT, clusters were flagged when the fold change < 1/2 and the associated adjusted p-value < 0.05; for nUMI and nFeature, clusters were flagged when the fold change > 2 with an adjusted p-value < 0.05. For cell annotation, we leveraged an externally curated HPAP CITE-seq reference dataset with manual cell-type annotations as the reference panel. Cluster-level reference mapping was performed using SingleR to get the final annotation (17). Cells were called restricted to donors annotated as non-diabetic (ND) or stage 3 type 1 diabetes (T1D) based on annotation from PanKbase (18). Harmony-derived clusters were mapped to major immune cell types.

For the TRIGR PBMCs, donor-multiplexed PBMC scRNA-seq data were obtained from 98 TRIGR participants, including 49 children who progressed to T1D and 49 matched non-progressors, under EGA accession EGAD50000001257. The available archive comprised 98, 90, and 92 participant-timepoint samples at the three respective sampling windows. Raw FASTQ files from 24 10x Chromium 3′ v3 libraries were aligned to GRCh38/GENCODE v39 using STARsolo. Cell calling combined CellBender, EmptyDrops, and library-adaptive thresholds for UMI depth, mitochondrial fraction, and ambient-RNA removal. Donor identities

were assigned using demuxlet, and expression-based doublets and doublet-enriched clusters were removed using DoubletFinder and marker-based review. Sampling time points were reconstructed by exact matching of genotype identifier and normalized pool identifier to the archived sample metadata. The final annotated dataset contained 218,935 cells; 199,264 uniquely assigned singlets were retained for donor-level, longitudinal, and pseudobulk analyses, whereas 19,671 cells without a unique donor–timepoint assignment were excluded from these analyses.

Metabolomics and Lipidomics Analysis

For targeted metabolomics, the immunological niche was bead-homogenized, and metabolites were extracted using 80% methanol (v/v). Supernatants from each sample were collected based on weight normalization. The collected supernatants were then dried using a SpeedVac Vacuum Concentrator, reconstituted in 50% methanol (v/v) in water, Targeted metabolomics was performed on an Agilent 1290 Infinity II Binary Bio LC coupled with an Agilent 6495d QqQ mass spectrometer. The column used was an Agilent InfinityLab Poroshell 120 HILIC-Z, 2.1 x 150 mm, 2.7uM (p/n 683775-924). Method parameters are as follows: Solvent A is water + 20mM ammonium acetate (pH 9.3) + 5uM medronic acid. Solvent B is acetonitrile. Wash solvent is 1:1:1 water, acetonitrile, and methanol. The solvent gradient is 10% A – 90% B until 1 minute, 22% A – 78% B until 8 minutes, 40% A – 60% B until 12 minutes, 90% A – 10% B until 15 minutes, hold until 18 minutes, 10% A – 90% B until 19 minutes, hold until 23 minutes. Flow rate is 0.4mL/minutes. Column temperature is set to 15 C. Source parameters are optimized for the Agilent dMRM (dynamic multiple reaction monitoring) library. The dMRM library was acquired from Agilent for their HILIC-Z platform and further optimized to screen for 435 targets in both positive and negative ion modes.

Lipids were extracted from the immunological niche using a chilled biphasic solvent system (MTBE:methanol, 10:3, v/v). Samples were homogenized, phase-separated with water, and the upper organic phase was collected. Targeted lipidomics was performed on an Agilent 1290 Infinity II Binary Bio LC coupled with an Agilent 6495d QqQ mass spectrometer. The column used was an Agilent ZORBAX Eclipse Plus C18, 100 × 2.1 mm, 1.8 µm (p/n 959758-902). Method parameters are as follows: Solvent A consists of 10 mM ammonium formate and 5 µM Agilent deactivator additive (p/n 5191-3940) in 5:3:2 water: acetonitrile:2-propanol. Solvent B consists of 10 mM ammonium formate in 1:9:90 water: acetonitrile:2-propanol. Wash solvent is a 1:1 mixture of solvents A and B. The solvent gradient is 85% A – 15% B until 2.5 minutes, 50% A – 50% B until 2.6 minutes, 43% A – 57% B until 9 minutes, 30% A – 70% B until 9.1 minutes, 7% A – 93% B until 11 minutes, 4% A – 96% B until 11.1 minutes 0% A – 100% B until 12.2, 85% A –15% B until 16 minutes. Flow rate is 0.4mL/minutes. Column temperature is set to 45 C. Samples are kept at 20 C. The dMRM library was acquired from Agilent for targeted lipidomics platform and further optimized to screen for 665 targets in both positive and negative ion modes.

Data from both methods was preprocessed using Agilent MassHunter Workstation Quantitative Analysis for QQQ Version 12.1. 12.1. Raw data was normalized by median-centering, and log normalization followed by pareto scaling. Further fold-change and Student's T- Tests analysis was performed relative to control groups to achieve a relative abundance in Metaboanalyst (v. 5.0).

Supplemental Tables Description (Attached as Excel Files)

Supplemental Table 1. NOD Progressor vs Non-Progressor IN Transcriptomic GSEA Analysis

Stage-wise Gene Set Enrichment Analysis (GSEA) results for NOD progressors vs non-Progressors.

Supplemental Table 2. IN T1D Specificity Analysis

Differential expression and pathway analyses used to assess T1D specificity of the IN-derived early stage DEGs. This includes DEG lists from 4T1 Breast cancer, Experimental Autoimmune Encephalomyelitis (EAE), and adoptive transfer (AT) T1D models, along with corresponding GSEA-enriched pathways for each model. Diseased vs healthy was compared in each model.

Supplemental Table 3. PLS-DA VIP Scores for IN T1D Signature Genes

Top 100 genes ranked by their Variable Importance in Projection (VIP) scores.

Supplemental Table 4. Anti-TNF Cohort Blood Glucose Data

Longitudinal blood glucose measurements for anti-TNF-treated cohorts and untreated controls.

Supplemental Table 5. Anti-TNF Sensitive vs Resistant IN GSEA Analysis

GSEA analysis of gene expression of IN transcriptomics comparing anti-TNF sensitive and resistant groups at pre-treatment and post-treatment.

Supplemental Table 6. IN-Derived Anti-TNF Response Score (ATRS)

Input gene sets related to TNF for elastic net, obtained parameters used for calculating the ATRS for selected gene sets, and individual ATRS scores for each mice.

References

1. Rad LM, et al. Engineered immunological niche directs therapeutic development in models of progressive multiple sclerosis. *Proceedings of the National Academy of Sciences*. 2025;122(7):e2409852122.

2. Orbach SM, et al. An engineered niche delineates metastatic potential of breast cancer. *Bioengineering & Translational Medicine*. 2024;9(1):e10606.

3. King JL, et al. Polymer scaffolds delineate healthy from diseased states at sites distal from the pancreas in two models of type 1 diabetes. *Biotechnology and Bioengineering*. 2024;121(11):3600–3613.

4. King JL, et al. Longitudinal monitoring of type 1 diabetes progression to disease onset. *Science Advances*. 2026;12(5):eadw8946.

5. Wu T, et al. clusterProfiler 4.0: A universal enrichment tool for interpreting omics data. *Innovation (Camb)*. 2021;2(3):100141.

6. Ritchie ME, et al. limma powers differential expression analyses for RNA-sequencing and microarray studies. *Nucleic Acids Res*. 2015;43(7):e47.

7. Kaestner KH, et al. NIH Initiative to Improve Understanding of the Pancreas, Islet, and Autoimmunity in Type 1 Diabetes: The Human Pancreas Analysis Program (HPAP). *Diabetes*. 2019;68(7):1394–1402.

8. Lun ATL, et al. EmptyDrops: distinguishing cells from empty droplets in droplet-based single-cell RNA sequencing data. *Genome Biol*. 2019;20(1):63.

9. Fleming SJ, et al. Unsupervised removal of systematic background noise from droplet-based single-cell experiments using CellBender. *Nat Methods*. 2023;20(9):1323–1335.

10. Ewels P, et al. MultiQC: summarize analysis results for multiple tools and samples in a single report. *Bioinformatics*. 2016;32(19):3047–3048.

11. Li H. Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM [preprint]. 2013. https://doi.org/10.48550/arXiv.1303.3997.

12. Li H, et al. The Sequence Alignment/Map format and SAMtools. *Bioinformatics*. 2009;25(16):2078–2079.

13. Zhang Y, et al. Model-based Analysis of ChIP-Seq (MACS). *Genome Biol*. 2008;9(9):R137.

14. Orchard P, et al. Quantification, Dynamic Visualization, and Validation of Bias in ATAC-Seq Data with ataqv. *cels*. 2020;10(3):298-306.e4.

15. Thibodeau A, et al. AMULET: a novel read count-based method for effective multiplet detection from single nucleus ATAC-seq data. *Genome Biol*. 2021;22(1):252.

16. Stuart T, et al. Comprehensive Integration of Single-Cell Data. *Cell*. 2019;177(7):1888-1902.e21.

17. Aran D, et al. Reference-based analysis of lung single-cell sequencing reveals a transitional profibrotic macrophage. *Nat Immunol*. 2019;20(2):163–172.

18. 2121-LB: PanKbase Integrated Single-Cell Map—A Comprehensive Atlas of Human Pancreatic Islets Unlocking Insights into Type 1 and Type 2 Diabetes | Diabetes | American Diabetes Association [Internet]. https://diabetesjournals.org/diabetes/article/74/Supplement_1/2121-LB/159793/2121-LB-PanKbase-Integrated-Single-Cell-Map-A. Accessed February 18, 2026.