## Supplementary Information for

# Harnessing Expressed Single Nucleotide Variation and Single Cell RNA Sequencing to Define Immune Cell Chimerism in the Rejecting Kidney Transplant

Andrew F. Malone, Haojia Wu, Catrina Fronick, Robert Fulton, Joseph P. Gaut, Benjamin D. Humphreys

Benjamin D. Humphreys, MD, PhD

Division of Nephrology

Washington University School of Medicine

660 S. Euclid Ave., CB 8129

St Louis, MO 63110

humphreysbd@wustl.edu

**This PDF file includes:**

- Supplementary methods
- Figures S1 to S7
- Table S1

**Other supplementary materials for this manuscript include the following:**

Datasets S1 to S5

## Supplementary Methods

### Biopsy Samples

Research core biopsy samples were obtained at the time of indication kidney transplant biopsy at Washington University under an Institutional Review Board approved protocol. Biopsy tissue was placed in Wisconsin buffer and placed on ice prior to immediate preparation for study.

### Single Cell isolation and Library Preparation

The renal biopsy was minced into small pieces with a razor blade and incubated at 37°C in freshly prepared dissociation buffer containing 0.25% trypsin and 40 U/ml DNase I in phosphate buffered saline. Dissociated cells were harvested every 10 minutes by filtering the cell suspension through a 70-μm cell strainer (pluriSelect) into 10% FBS on ice. The residual biopsy tissue trapped on the cell strainer was dissociated once again with 1 ml dissociation buffer for 10 minutes and passed through the cell strainer into the same FBS buffer from the first collection. We repeated this dissociation procedure three times until most of the tissue had been dissociated into single cells (total dissociation time was 30-40 minutes). Finally, cells were collected by centrifugation at 400×g for 5 minutes, resuspended in suspension buffer containing 0.1% BSA, and strained through a 40-μm cell mesh (pluriSelect) to further remove cell clumps and large fragments. Cell viability was approximately 90% for the biopsy used in this study as assessed by Trypan Blue staining.

### Generation of cDNA libraries for scRNA-seq

Cell suspensions were loaded in to 10X Chromium lanes for a target cell capture number of 10,000 cells per lane. Two lanes were used per biopsy. The 10X Chromium libraries were prepared according to the manufacturers protocol. The 5-prime kits were used and T cell VDJ libraries were also prepared according to manufacturer's protocol.

### Next Generation Sequencing of scRNA-seq libraries

The concentration of each prepared 10x single cell library was accurately determined through qPCR utilizing the KAPA library Quantification Kit according to the manufacturer's protocol (Roche) to produce cluster counts appropriate for the Illumina NovaSeq6000 instrument. Paired end sequence data (2x150) on the S4 flow cell was generated targeting 50,000 total read pairs per cell for gene expression and 5,000 total read pairs per cell for VDJ sequence. The Cellranger 3.0 pipeline was used to generate BAM files and gene expression matrices for each biopsy.

### IDT Exome Sequencing (Whole-exome Sequencing) Methods

250ng of genomic DNA was fragmented on the Covaris LE220 instrument targeting 250bp inserts. Automated dual indexed libraries were constructed with the KAPA HTP library prep kit (Roche) on the SciClone NGS platform (Perkin Elmer). Ten libraries were pooled at an equimolar ratio by mass prior to the hybrid capture targeting a 5μg library pool. The library pool was hybridized with the xGen Exome Research Panel v1.0 reagent (IDT Technologies) that spans 39 Mb target region (19,396 genes) of the human genome. The libraries were hybridized for 16-18 hours at 65°C followed by stringent washing to remove spuriously hybridized library fragments. Enriched library fragments were eluted with streptavidin-coated magnetic beads and amplified with KAPA HiFi Polymerase prior to sequencing. PCR cycle optimization was performed to prevent over amplification of the libraries. The concentration of each captured library pool was accurately determined through qPCR utilizing the KAPA library Quantification Kit according to the manufacturer's protocol (Roche) to produce cluster counts appropriate for the Illumina NovaSeq6000 instrument. Approximately 5Gb of paired end sequence data (2x150) on the S4 flow cell was generated targeting 50x coverage per sample. Alignment was

performed using BWA-MEM. MarkDuplicates with Picard followed by BaseQualityRecalibrate (BQSR) with GATK was then performed in accordance with the functional equivalence paper published by the Centers for Common Disease Genomics (CCDG)[1]. Files were then converted to CRAM format.

**Generation of Seurat Objects**

Cellranger output files were used to generate count matrices using Read10X() in R. Seurat v3 CreateSeuratObject() was then used and the final object was refined to include cells with genes expressed in at least 3 cells, 200-2500 genes per cell detected and less than 25% mitochondrial genes per cell (Fig. S1). The resolution and number of principle components used in the final Seurat object for each biopsy varied. Each biopsy was pre-clustered using an excessive PC number and high resolution to create a high number of clusters. Cluster defining genes based on the FindAllMarkers() function were then examined. Clusters containing 2 or more different cell type defining markers were removed as doublet cells.

| Cell Type | Cell Type Defining Marker Genes | | | | | |
|---|---|---|---|---|---|---|
| Proximal tubule | SLC34A1 | LRP2 | CUBN | | | |
| Loop of Henle | SLC12A1 | UMOD | SLC14A2 | | | |
| Principle cell | AQP2 | SCNN1G | | | | |
| Intercalated cell | SLC4A1 | SLC26A4 | | | | |
| Endothelial cell | PECAM1 | VWF | | | | |
| Stroma | Col1A1 | ACTA2 | PDGFRB | | | |
| Lymphocytes | CD58 | CD3E | CCL5 | IL7R | | |
| Macrophages | FCGR3A | CD14 | CD68 | C1QA | C1QB | C1QC |

Clusters with a very low average number of genes per cell were removed as debris clusters if no single cluster defining gene was identified. Once doublet clusters were removed the final object was recreated using a lower resolution and PC number.

| **Activated EC - select markers** | | |
|---|---|---|
| | **avg_logFC** | **p_val_adj** |
| ACKR1 | 1.912641502 | 0 |
| CCL14 | 1.514404961 | 2.74E-192 |
| HLA-A | 0.318347062 | 2.46E-88 |
| HLA-C | 0.290345748 | 1.63E-89 |
| HLA-DPA1 | 0.397149923 | 3.75E-64 |
| HLA-DQA1 | 0.541857251 | 2.54E-91 |
| VCAM1 | 0.527646972 | 4.09E-26 |
| TGFBR2 | 0.42654239 | 1.12E-78 |
| | | |
| **Activated PT - select markers** | | |
| | **avg_logFC** | **p_val_adj** |
| CCL2 | 1.458782518 | 0 |
| HLA-A | 0.730987498 | 0 |
| HLA-B | 1.010111178 | 0 |
| HLA-C | 0.686793105 | 0 |
| ITGB1 | 0.373605583 | 4.9338E-164 |
| ITGB8 | 0.392534968 | 2.7486E-134 |

| TNFRSF12A | 0.98407814 | 0 |
| --- | --- | --- |
| IRF1 | 1.002454404 | 0 |

Activated endothelial cells (EC) differentially express HLA class I and II genes, adhesion molecules and proinflammatory genes. Similarly, activated proximal tubular cells (PT) were defined by differential expression of class I HLA genes, integrins, pro-adhesion molecules and proinflammatory genes.[2-5]

**Generation of Integrated Seurat Object**
An integrated dataset was created using the standard Seurat v3 integration analysis pipeline. Each input object was log-normalized prior to variable feature selection based on a variance stabilizing transformation to find the 2000 most variable genes. 30 principle components were calculated for PC analysis and the final integrated object was constructed using 11 PCs.

**Donor Recipient Cell Origin – Demuxlet Pipeline[6]**
Variant call format files (.vcf) were created from the CRAM output file from the WES pipeline using GATK best practices. A single .vcf file was created for each donor-recipient pair. These files were filtered for high quality common variants (allele frequency > 0.05). An aggregated .BAM file was created from the .BAM files from each of the 2 single cell 10X lanes used for each biopsy. The aggregated .BAM file and the .vcf file from each biopsy was used as input for the demuxlet pipeline using demuxlet default parameters and alpha of 0, 0.5 and 1. Each cell in the final integrated data set was then annotated as donor or recipient using the best column from the demuxlet output data.

**Identification of differentially expressed genes**
Differentially expressed genes were identified by comparing the transcriptional profile of donor and recipient cells within the macrophage or T cell clusters in the integrated dataset using FindMarkers() with logfc.threshold=0.25 and min.pct=0.25. Also, differentially expressed genes were identified by comparing the transcriptional profile of T cells from rejecting biopsies and T cells from non-rejecting biopsies using the same method.

**Correlation plots and DotPlots**
The cor() and corrplot() functions in R were used for a Pearson correlation of gene expression between donor and recipient macrophages and rejecting and non-rejecting macrophages using a significance level of 0.05. The Seurat function DotPlot with default parameters was used to compare gene expression of top genes between donor and recipient cells and rejecting and non-rejecting cells.

**Pathway Analysis**
Significant differentially expressed genes (adjusted p value < 0.05) from the Seurat function FindMarkers() (minimum percentage of cells with gene = 0.25, log fold change threshold of 0.25) were used for pathway analysis. GO analysis was performed on the differentially expressed genes using the ToppGene Suite (https://toppgene.cchmc.org). Non-significant GO terms (Bonferroni p value above 0.05) were removed from analysis. For T cell pathway analysis, pathways for which only CD3 and CD2 gene expression was common to the reference pathway gene set were excluded from analysis. Significant enriched GO terms (defined by defined by Bonferroni P-value <0.05) were summarized by REVIGO[7] and visualized by treemap R package.

**T Cells VDJ Immune Cell Profiling**

T cell VDJ sequence was analyzed using VDJ Loupe Browser. Clones were defined as cells with complete VDJ sequence from both alpha and beta chains. Using cell barcodes T cell clones were identified in the final integrated dataset.

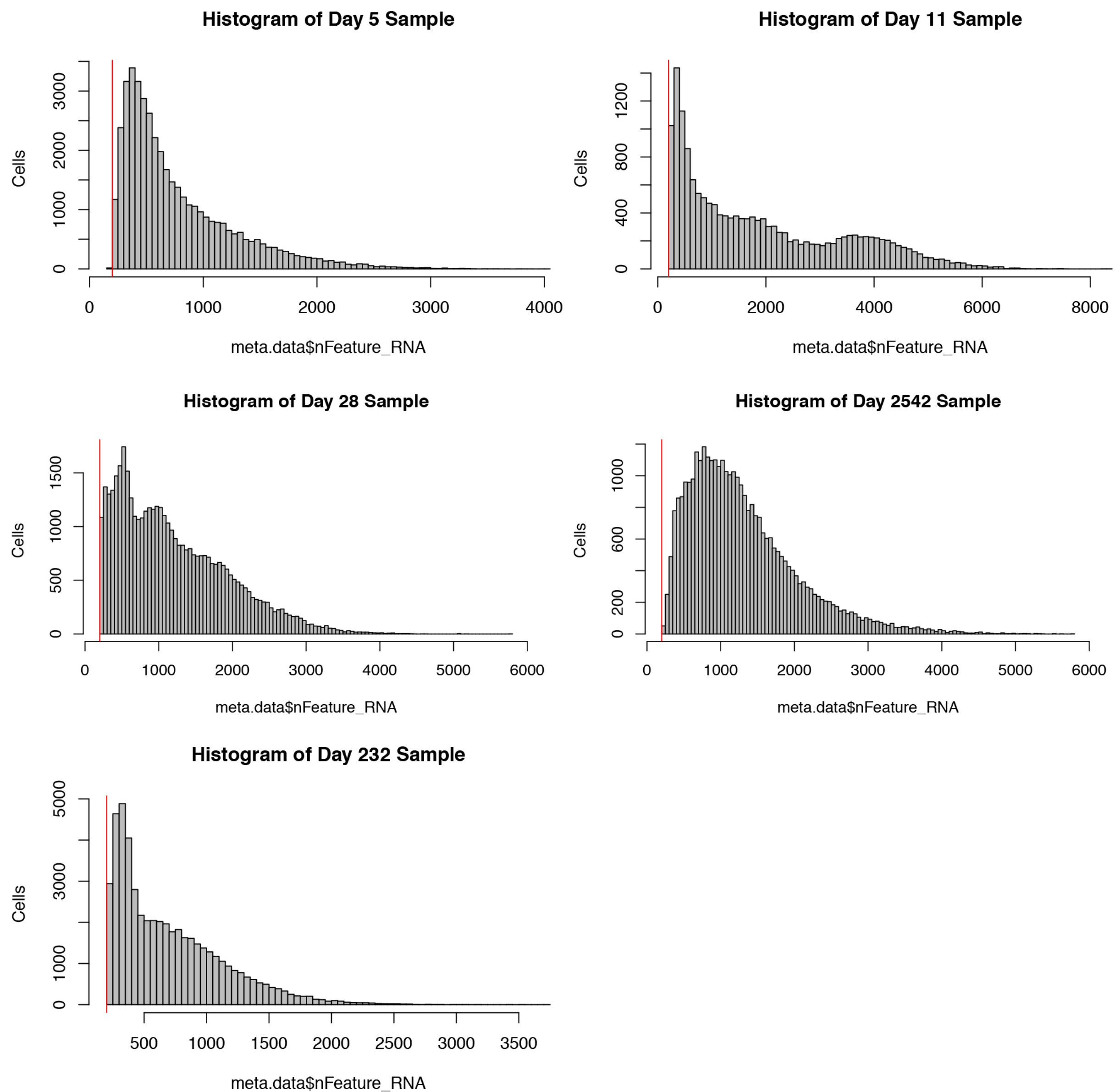


Figure S1. Histograms of all droplets with at least one gene per droplet for each sample in the dataset. Data is from the meta.data@nFeature_RNA slot of the initial object created by CreateSeuratObject() from Cellranger output 'filtered_feature_bc_matrix'. Vertical red line demotes lower limit cut off of 200 genes per cell used for quality control filtering of the samples.

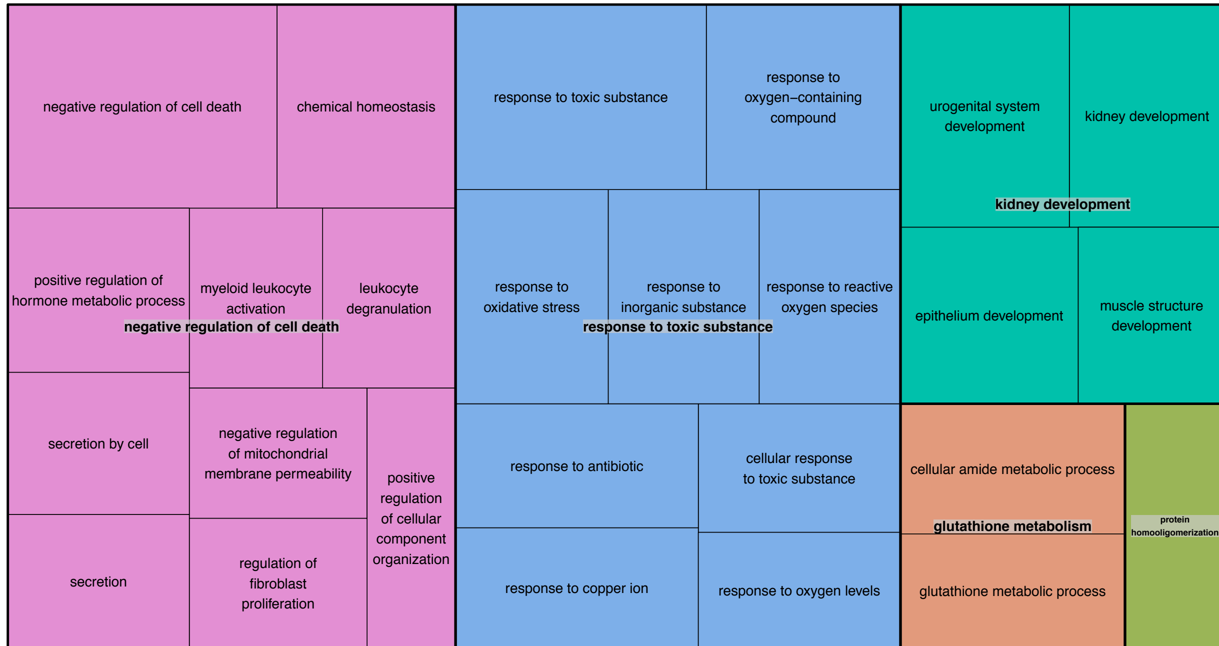


Figure S2. Gene ontology terms for differentially expressed genes that define stressed tubular cell cluster. Top terms were negative regulation of cell death and responses to external toxins including oxidative stress.

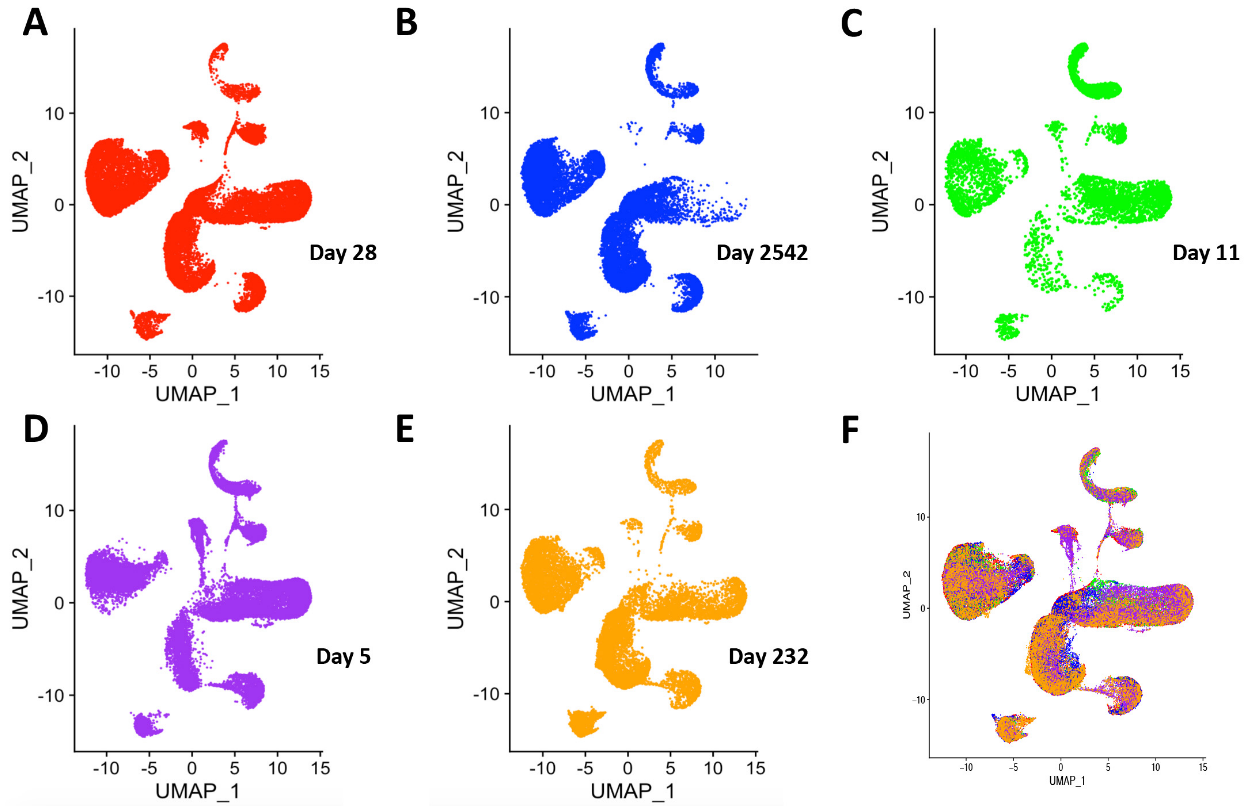


Figure S3. A+B+C+D+E) UMAP visualization of each individual biopsy in the dataset demonstrating equal coverage of all cell clusters. F) UMAP visualization of the whole dataset grouped by original biopsy identity.

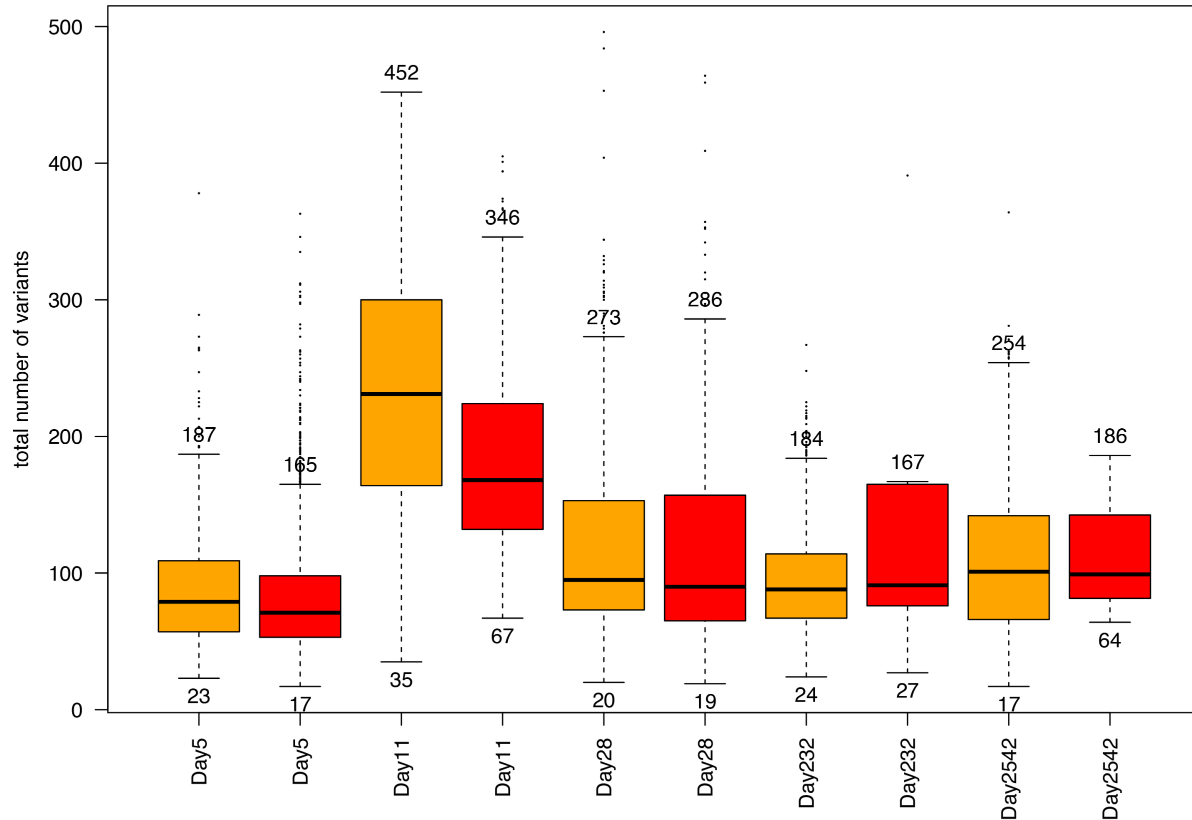


Figure S4. Total number of SNVs overlapping at least one read in immune cells grouped by five biopsy samples. The median number of SNVs overlapping at least one read used to call origin in a cell was between 78 and 233. Numbers represent the maximum and minimum number of read overlapping SNPs for each sample. Yellow boxes represent recipient cells and red boxed represent donor cells.

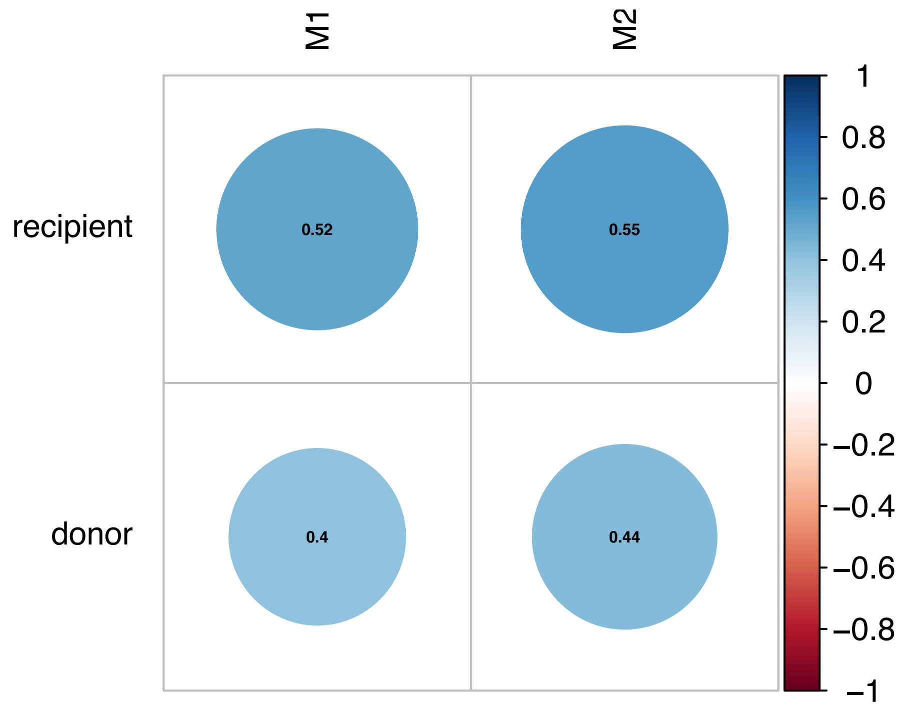


Figure S5. Gene expression from donor and recipient macrophages were correlated with human M1 (n=4) and M2 (n=4) macrophage bulk RNA-seq gene expression external datasets. K. Y. Gerrick *et al.*, Transcriptional profiling identifies novel regulators of macrophage polarization. *PloS one* **13**, e0208602 (2018).

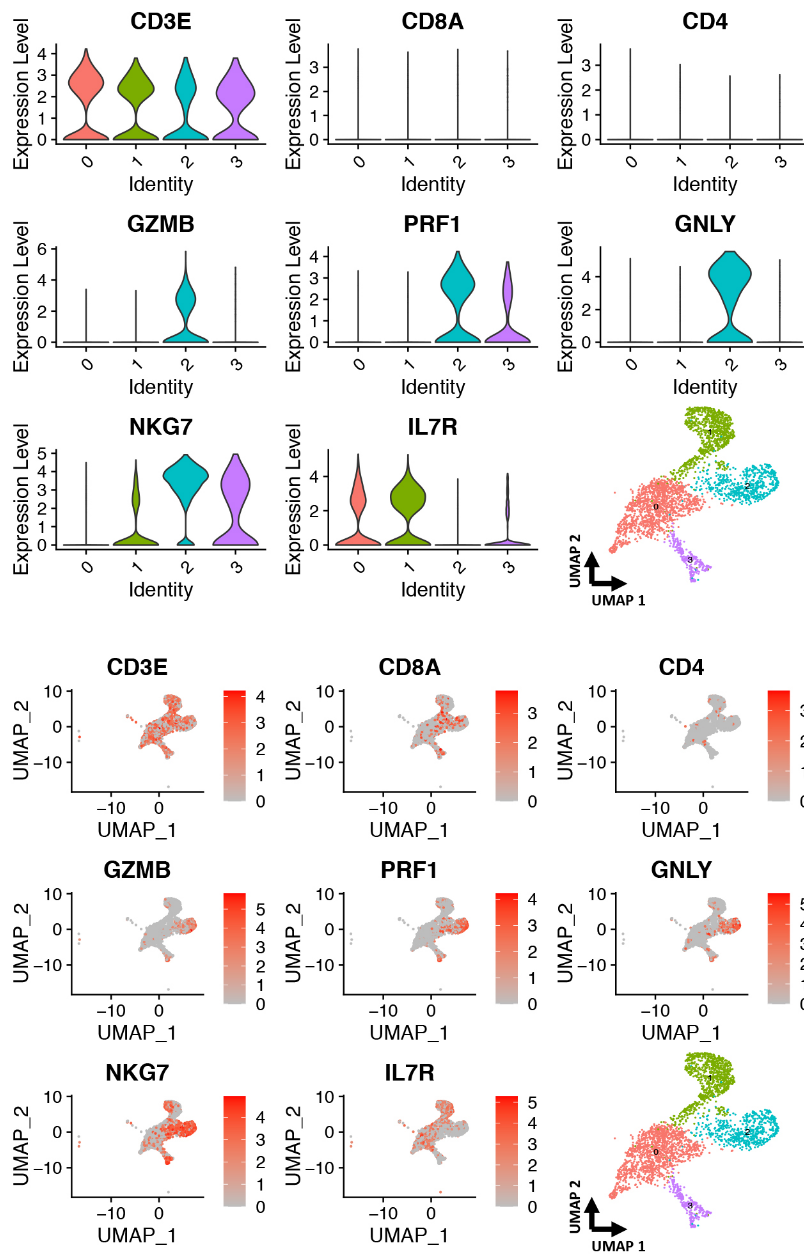


Figure S6. Violin plots and feature plots of common T cell genes across the T cell subclusters 0, 1, 2, 3. One was an activated T cell subcluster expressing *GZMB*, *PRF1* and *GNLY*. Two subclusters expressed *IL7R* and were likely CD4+ T cells however neither *CD8A* nor *CD4* were significantly expressed in any T cell subcluster. Of the two IL7R+ T cells clusters, one was predominantly donor origin and the other mixed origin. T cell subcluster 3 expressed *NKG7* most likely representing NKT cells.

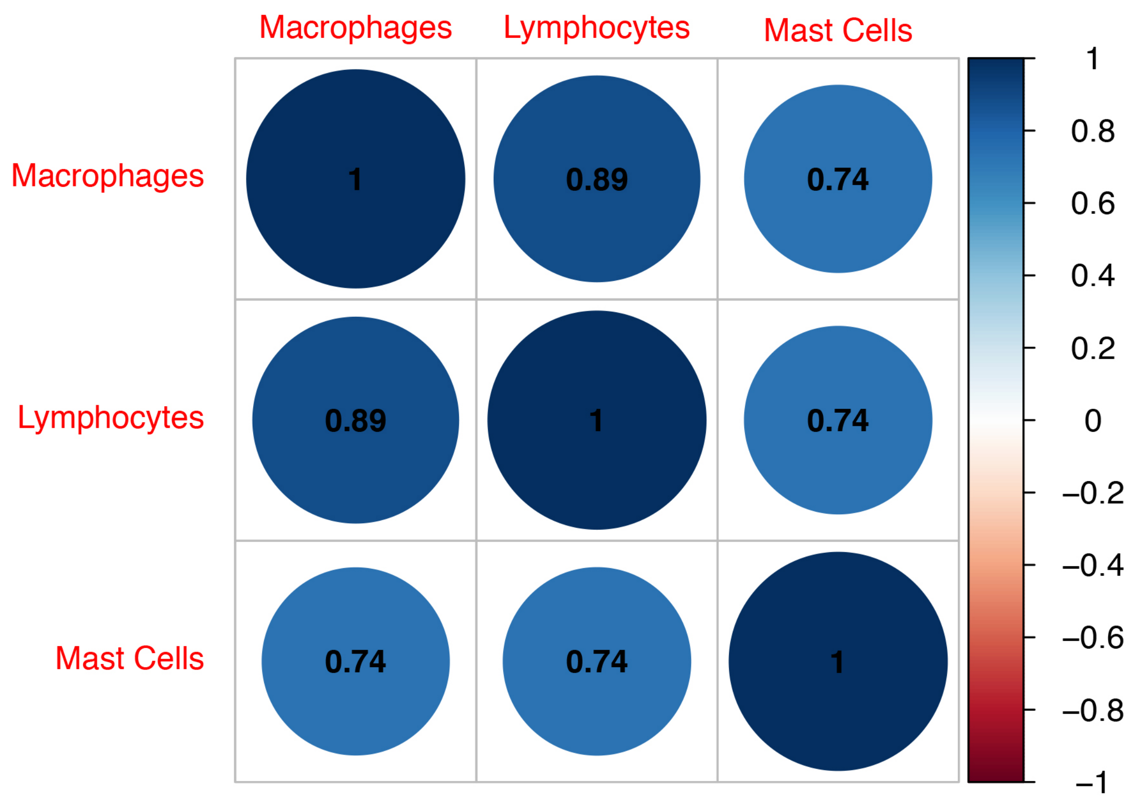


Figure S7. Spearman correlation between mast cell gene expression and macrophage gene expression and with lymphocyte gene expression.

## Table S1. Clinical Characteristics of the 5 Biopsies

| Rejection Status | Patient Age/Gender (yrs.)/(M/F) | Race | Time post Transplant (days) | Organ Type | Sex Matching | Induction | Biopsy Indication | Biopsy Diagnosis | Banff Scoring | Donor Specific Antibodies |
|---|---|---|---|---|---|---|---|---|---|---|
| No Rejection | 53F | AsA | 5 | DBD | M/F | TMG Ritux | AKI | Acute tubular injury, No rejection | t0,v0,i0, g0,ptc0, ct0,ci0, cg0,c4d0 | No |
| Rejection | 38M | C | 11 | DCD ABOi | M/M | TMG Ritux | AKI and proteinuria 2.1g | ABOi rejection, Thrombotic microangiopathy. No cellular rejection | t0,v0,i0, g0,ptc2, ci0,ct0, cg0,c4d3 | Yes anti-A 1:32 |
| No Rejection | 55M | AA | 28 | DCD | F/M | TMG | AKI and proteinuria 3.1g | Acute tubular injury. No rejection | t0,v0,i0, g0,ptc0, ci0,ct1, cg0,c4d0 | No |
| Rejection | 75M | C | 232 | LURKT | F/M | TMG | AKI and previous ABMR | Active ABMR, C4d is diffusely positive | t0,v0,i0, g2,ptc3, ci0,ct1, cg0,c4d3 | Yes B8(1190mfi); DQA1*05:01(17846mfi); DQA1*03(5153mfi) |
| Rejection | 35M | C | 2542 | LURKT | F/M | TMG | AKI and proteinuria 6.7g | Chronic Active ABMR, C4d is focally positive. IgA nephropathy consistent with recurrence | t0,v0,i0, g2, ptc3, ct1,ci1, cg3,c4d1 | Yes DR1(5834mfi); DR53(25460mfi); DQ2(7623) |

AsA, Asian American. C, Caucasian. AA, African American. DBD, brain death donor. DCD, donation after cardiac death. ABOi, ABO incompatible. LURKT, living unrelated kidney transplant. TMG, thymoglobulin. Ritux, rituximab.

**Additional data table (DatatableS1.cxv)**
Differential gene list defining each cell cluster in the whole dataset.

**Additional data table (DatatableS2.xls)**
Differentially expressed genes for recipient versus donor T cells and Differentially expressed genes for T cells from rejecting samples versus T cells from non-rejecting T cells.

**Additional data table (DatatableS3.csv)**
Differentially expressed genes for recipient versus donor macrophages.

**Additional data table (DatatableS4.csv)**
Pathway analysis for differential expressed (recipient and donor) macrophages genes using ToppFun tool.

**Additional data table (DatatableS5.csv)**
Pathway analysis for differential expressed (recipient vs donor and rejecting vs non-rejecting) T cells genes using ToppFun tool.

## Supplemental References


1. Regier, AA, Farjoun, Y, Larson, DE, Krasheninina, O, Kang, HM, Howrigan, DP, Chen, BJ, Kher, M, Banks, E, Ames, DC, English, AC, Li, H, Xing, J, Zhang, Y, Matise, T, Abecasis, GR, Salerno, W, Zody, MC, Neale, BM, Hall, IM: Functional equivalence of genome sequencing analysis pipelines enables harmonized variant calling across human genetics projects. *Nat Commun,* 9**:** 4038, 2018.
2. Egelkamp, J, Chichelnitskiy, E, Kühne, JF, Wandrer, F, Daemen, K, Keil, J, Bräsen, JH, Schmitz, J, Bellmàs-Sanz, R, Iordanidis, S, Katsirntaki, K, Hake, K, Akhdar, A, Neudörfl, C, Haller, H, Blume, C, Falk, CS: Back signaling of HLA class I molecules and T/NK cell receptor ligands in epithelial cells reflects the rejection-specific microenvironment in renal allograft biopsies. *American journal of transplantation : official journal of the American Society of Transplantation and the American Society of Transplant Surgeons,* 19**:** 2692-2704, 2019.
3. Zhang, X, Rozengurt, E, Reed, EF: HLA class I molecules partner with integrin beta4 to stimulate endothelial cell proliferation and migration. *Science signaling,* 3**:** ra85, 2010.
4. Cottone, S, Palermo, A, Vaccaro, F, Mulè, G, Guarneri, M, Arsena, R, Vadalà, A, Cerasola, G: Inflammation and endothelial activation are linked to renal function in long-term kidney transplantation. *Transplant international : official journal of the European Society for Organ Transplantation,* 20**:** 82-87, 2007.
5. Sanz, AB, Izquierdo, MC, Sanchez-Niño, MD, Ucero, AC, Egido, J, Ruiz-Ortega, M, Ramos, AM, Putterman, C, Ortiz, A: TWEAK and the progression of renal disease: clinical translation. *Nephrology, dialysis, transplantation : official publication of the European Dialysis and Transplant Association - European Renal Association,* 29 Suppl 1**:** i54-i62, 2014.
6. Kang, HM, Subramaniam, M, Targ, S, Nguyen, M, Maliskova, L, McCarthy, E, Wan, E, Wong, S, Byrnes, L, Lanata, CM, Gate, RE, Mostafavi, S, Marson, A, Zaitlen, N, Criswell, LA, Ye, CJ: Multiplexed droplet single-cell RNA-sequencing using natural genetic variation. *Nat Biotechnol,* 36**:** 89-94, 2018.

7. Supek, F, Bosnjak, M, Skunca, N, Smuc, T: REVIGO summarizes and visualizes long lists of gene ontology terms. *PLoS One,* 6**:** e21800, 2011.