Identification and Tracking of Alloreactive T Cell Clones in Rhesus Macaques Through the RM-scTCR-Seq Platform

Ulrike Gerdemann1†, Ryan A. Fleming1†, James Kaminski1,2,3,4†, Connor McGuckin1, Xianliang Rui1, Jennifer F. Lane1, Paula Keskula1, Lorenzo Cagnin1, Alex K. Shalek2,3,4‡, Victor Tkachev1‡ and Leslie S. Kean1*†

1 Division of Pediatric Hematology/Oncology, Boston Children’s Hospital; Department of Pediatric Oncology, Dana Farber Cancer Institute and Harvard Medical School, Boston, MA, United States, 2 Broad Institute of MIT and Harvard, Cambridge, MA, United States, 3 Department of Chemistry, Institute for Medical Engineering and Science (IMES), and Koch Institute for Integrative Cancer Research, MIT, Cambridge, MA, United States, 4 Ragon Institute of Massachusetts General Hospital (MGH), MIT and Harvard, Cambridge, MA, United States

T cell receptor (TCR) clonotype tracking is a powerful tool for interrogating T cell mediated immune processes. New methods to pair a single cell’s transcriptional program with its TCR identity allow monitoring of T cell clonotype-specific transcriptional dynamics. While these technologies have been available for human and mouse T cells studies, they have not been developed for Rhesus Macaques (RM), a critical translational organism for autoimmune diseases, vaccine development and transplantation. We describe a new pipeline, ‘RM-scTCR-Seq’, which, for the first time, enables RM specific single cell TCR amplification, reconstruction and pairing of RM TCR’s with their transcriptional profiles. We apply this method to a RM model of GVHD, and identify and track in vitro detected alloreactive clonotypes in GVHD target organs and explore their GVHD driven cytotoxic T cell signature. This novel, state-of-the-art platform fundamentally advances the utility of RM to study protective and pathogenic T cell responses.

Keywords: TCR sequencing, single cell sequencing, alloreactive T cells, Rhesus Macaque (Macaca mulatta), GVHD

INTRODUCTION

Expression of the highly polymorphic (up to 1018 specificities) (1) T cell receptor (TCR) allows T cells to specifically recognize foreign and self-antigens presented in the context of the Major Histocompatibility Complex (MHC) by antigen presenting cells. This diversity is achieved by complex genetic recombination of the T cell alpha and beta chain during thymic T cell development, and results in a unique molecular barcode for each T cell. These unique TCR sequences enable studies of clonal- and diversity- dynamics in models of transplantation, infection, tumor immunology and vaccine development (2–7). More recently TCR tracking has been combined with high throughput single cell RNA sequencing (scRNA-Seq), enabling transcriptional profiling of distinct T cell clonotypes within and across samples (8, 9).
To date, clonotype tracking technologies have been developed for both mouse and human TCRs, facilitating studies of T cell clonal dynamics at an unprecedented level of molecular sensitivity (9–11). However, both murine and human models have drawbacks: The inbred nature of mice and the evolutionary distance between the murine and human immune systems can restrict the clinical applicability of murine results. Moreover, the study of human single cell TCR data continues to focus predominantly on the peripheral blood, being restricted by inherent challenges in accessing human tissues during health and disease. This can result in an incomplete picture of important immunological processes when using human samples.

To address these limitations, nonhuman primate (NHP) models (particularly using rhesus macaques (RM)) have been essential (12), and have enabled critical advances for some of the most pressing human needs, including the development of vaccines against SARS-CoV-2 (13–17), the study of HIV pathogenesis (18–20), and in advancing our mechanistic understanding of solid organ transplant (21, 22) and hematopoietic stem cell transplant (HCT) (23–29). For HCT, NHP models have enabled the translation of several novel therapeutic strategies to the clinic for the prevention of graft-versus-host disease (GVHD), the major cause of transplant-related mortality after HCT (30, 31).

The identification and tracking of TCR clonotypes are of critical importance for each of the clinical settings described above, in order to understand the molecular mechanisms driving immune protection or disease pathogenesis. For acute GVHD (aGVHD), dissecting the link between T cell clonal architecture and disease is particularly relevant, given that donor-derived alloreactive T cells are the main culprits in the aGVHD-mediated destruction of the skin, intestine, and liver (32–34). Accurate identification and characterization of alloreactive T cell clonotypes is essential to identifying the molecular drivers of aGVHD, and doing so in aGVHD-target tissues as well as the peripheral blood remains a major challenge. Until the work described in this manuscript, the identification and tracking of individual TCR clonotypes in NHP models (in particular, RM) has been limited by a lack of robust methodology for single cell TCR sequencing adapted to this essential pre-clinical model. The key limitations facing the field were the lack of effective RM-specific primers that could successfully amplify the TCR region in a highly multiplexed fashion, as well as the incomplete annotation of the RM genome throughout the TCR alpha and beta chain region, which created a major barrier to accurate TCR chain reconstruction.

We now describe the design and validation of RM-specific primers compatible with the human 10x Genomics single cell sequencing platform, which accurately amplifies the RM TCR alpha and beta regions. Amplified fragments aligned to our custom-assembled and annotated RM TCR reference permit, for the first time, the in vitro and in vivo identification and tracking of RM derived T cell clonotypes. Pairing of in vivo identified alloreactive clonotypes with their transcriptional profile revealed a highly activated cytotoxic CD8+ T cell signature during NHP aGVHD.

METHODS

NHP

This study was conducted in strict accordance with (USDA) United States Department of Agriculture regulations and the recommendations in the Guide for the Care and Use of Laboratory Animals of the National Institutes of Health. It was approved by the Massachusetts General Hospital and Biomere Animal Care and Use Committees. T cells were obtained from healthy colony RM or from animals who underwent allogenic HCT in the setting of previously published studies (28).

Mixed Lymphocyte Reaction (MLR)

RM PBMCs were isolated from whole blood by Ficoll gradient centrifugation, and then used for MLR assays either immediately, or after liquid nitrogen cryopreservation in 10% dimethyl sulfoxide (DMSO)/90% fetal bovine serum (FBS). At the time of the MLR, stimulator PBMCs were irradiated with 3500 cGy of (137Cs) radiation. Responder PBMCs were stained with cell trace violet (CTV, Invitrogen) as per manufacturer’s instructions. 2×10^5 T cell-enriched responder PBMCs along with an equal number of stimulator PBMCs were added to each well in a 96-well plate (Corning) in X-vivo-15 medium (BioWhitaker) supplemented with 10% FBS (Irvine Sciific) and incubated at 37°C for a total of 5 days. Cell culture media change was performed on day 3 of the culture. At the end of 5 days, cells were stained with an extracellular antibody for CD3 (clone SP34-2), CD20 (clone 2H7) and CD14 (clone M5E2, all antibodies from BD Bioscience) for 20min at 4°C, and high-, medium- and non-proliferating CD3+ T cells (identified based on the dilution of CTV) were sorted and processed for single cell sequencing using the Chromium Next GEM Single Cell 5’ Reagent Kit v1 with optimized RM primers as described below.

NHP HCT, Necropsy and Tissue Processing

RM HCT, necropsy, and tissue processing were performed as previously described (28, 31). Single cell suspensions resulting from tissue processing were sorted on a live CD3+ and CD14/CD20- population and prepared for single cell sequencing.

Sample Processing, Primer Design and PCR Amplification

All products used in the preparation of these samples, excluding the custom primer set designed and described herein, are available from 10x Genomics. While samples in this publication were prepared using the Next GEM 5’ v2 Gel Beads with the i7 Multiplex Plate (Single Index), we also validated primers and PCR conditions compatible with the newer Next GEM Single Cell 5’ v2 Sample Index Plate TT (Dual Index) (see below).

Single Cell V(D)J v1: GEMs were generated using v2 Gel Beads and the v1 Target Enrichment kit substituting the custom primers for the off-the-shelf human/mouse T Cell Mix 1 and 2 premixed primers. Sample Indexing was performed using the i7 Multiplex Kit plate (10x Genomics, PN-120262). TCR alpha and beta primers were designed to perform in a nested PCR approach.
to optimize target enrichment of the respective constant regions (Table 1). TRAC inner and outer primers were 297bp apart, and the beta chain inner and outer primers were 357bp apart. Primers were optimized to match the melting temperature of the 10x v1 forward primers for annealing to a known sequence within the gel beads in emulsion (GEMs) according to the human 10x Genomics Chromium Single Cell V(D)J Reagent Kits User Guide for target enrichment (https://assets.ctfassets.net/an68im79xiti/31W4aZ0J8C2ipxzTrcQ1Wn/72c5c2bc3dd7784f4f5218ca345ec04/CG000086_ChromiumSingleCellV_D_J_ ReagentKits_UG_RevM.pdf). Primer sequences are listed in Table 1.

PCR was performed with an initial denaturation temperature of 98°C for 45 seconds, followed by denaturation at 98°C for 20 seconds. Annealing was optimal at 67°C for 30 seconds with extension of 72°C for 1 min. A total of 20 PCR cycles were performed with a final extension step at 72°C for 1 min.

Next GEM Single Cell 5’ v2: GEMs were also generated using v2 reagent kits for GEM generation and library preparation while substituting the custom RM primers for the off-the-shelf human/mouse T Cell Mix 1 and 2 premixed primers. Sample Indexing was performed using the Dual Index Kit TT (10x Genomics PN:1000215). Primers (Table 2) were optimized to match the melting temperature of the 10x v2 forward primers for annealing to a known sequence within the gel beads in emulsion (GEMs) according to the human 10x Genomics Chromium Single Cell V(D)J Reagent Kits User Guide for target enrichment (https://assets.ctfassets.net/an68im79xiti/57JaTECQBNBSpypDz80ucdi/92da9f62521e1a1f5c781b a234bd0a0f5/CG0000331_ChromiumNextGEMSingleCell5-v2_ UserGuide_RevC.pdf). TRAC inner and outer primers were 357bp apart, and the beta chain inner and outer primers were 357bp apart. Primer sequences are listed in Table 2.

PCR was performed with an initial denaturation temperature of 98°C for 45 seconds, followed by denaturation at 98°C for 20 seconds. Annealing was optimal at 62°C for 30 seconds with extension of 72°C for 1 min. A total of 20 PCR cycles were performed with a final extension step at 72°C for 1 min. Sequencing results utilizing v2 primers are listed in Table S1.

### Reference Transcriptome

The reference transcriptome (Data S4) was created using Cellranger version 6.0.0 with the following steps:

1. We first filtered Ensembl’s gtf file for v104 of the macaca mulatta gene annotation using cellranger’s mkgtf command with the following attributes:

   - attribute=gene_biotype:protein_coding – attribute=gene_biotype:incRNA – attribute=gene_biotype:miRNA – attribute=gene_biotype:antisense – attribute=gene_biotype:IG_LV_gene – attribute=gene_biotype:IG_V_gene – attribute=gene_biotype:IG_J_gene – attribute=gene_biotype:IG_C_gene – attribute=gene_biotype:IG_Pseudogene – attribute=gene_biotype:TR_V_gene – attribute=gene_biotype:TR_J_gene – attribute=gene_biotype:TR_C_gene

2. We next edited cellranger’s mkref program to use more than one thread using the command:

   ```
   sed -i `88s/num_threads/#num_threads/`/cellranger-6.0.0/bin/rna/mkref
   ```

3. We then created the reference transcriptome using cellranger’s mkref program: cellranger-6.0.0/bin/mkref –genomes=MMulatta_Ensmbel_v104 –fasta=MMulatta_Ensmbel_v104_combined.fasta –genes=MMulatta_Ensmbel_v104_combined_filtered.gtf –nthreads=32 –memgb=128

### VDJ Reference

1. We used version 6.0.2 of Cellranger to create the initial VDJ reference based on IMGT (35, 36).

2. We used the following command to generate the initial reference: cellranger-6.0.2/lib/bin/fetch-imgt –genome vdj:IMGT_rhesus –species “Macaca mulatta”.

3. This version of the VDJ reference did not include constant regions, so we added the following sequences - obtained from IMGT and Ensembl - to the reference:

   ```
   >340TRAC*01 IMGT|TRAC|C-REGION|TR|TRA|None|01
   AATGATACGGCGACCACCGAGA-TCT
   ```

### Table 1 | Primer sequences optimized for GEM-scTCR-Seq v1.

| Primer                | Sequence                              |
|-----------------------|---------------------------------------|
| Forward Primer 1 (10x Genomics) | AATGATACGGCGACCACCGAGA-TCTCACTCTTTTCCCTACACGAGCGCTC |
| Forward Primer 2 (10x Genomics) | AATGATACGGCGACCACCGAGA-TCTCACTCTTTTCCCTACACGAGCGCTC |
| Alpha Outer Primer     | GGAGCGACACTTCTACACGAGCGCGAGAGAGAGAG |
| Alpha Inner Primer     | TGTCTGTGTATGATACGGCGACCAAGAGAGAGAGAG |
| Beta Outer Primer      | CGGTATTCCAGCAGACACACACCATCTCATGTC |
| Beta Inner Primer      | CGGTATTCCAGCAGACACACACCATCTCATGTC |

### Table 2 | Primer sequences optimized for GEM-scTCR-Seq v2.

| Primer                | Sequence                              |
|-----------------------|---------------------------------------|
| Forward Primer 1 (10x Genomics) | GATCTACACTCTTTTCCCTACACGAGCGCTC |
| Forward Primer 2 (10x Genomics) | GATCTACACTCTTTTCCCTACACGAGCGCTC |
| Alpha Outer Primer     | GGAGCGACACTTCTACACGAGCGCGAGAGAGAGAG |
| Alpha Inner Primer     | TGTCTGTGTATGATACGGCGACCAAGAGAGAGAGAG |
| Beta Outer Primer      | CGGTATTCCAGCAGACACACACCATCTCATGTC |
| Beta Inner Primer      | CGGTATTCCAGCAGACACACACCATCTCATGTC |
as discussed in the Results section, the clonotypes used for the alpha chain and beta chain Morisita heatmaps (Figure 2, panels E, F) were identified by grouping together T cells on the basis of the CDR3 region in the alpha chain or beta chain. This is to provide a more stringent test of the specificity of clonotype calling, as Cellranger’s default settings do not allow for clonotypes to be shared across different donors. The Morisita index was calculated using the “vegan” package in R.

**Differential Expression (DE) Tests**

DE tests were carried out as Wilcoxon Rank Sum tests in Seurat, using the default settings. Results with an adjusted p value <0.05 were considered significant.

**Scoring of Cells With Gene Signatures Using VISION**

We used VISION to apply gene signatures to the normalized expression data for the analyzed T cells. We used the C7 collection of immunological signatures from MSigDB, and ten signatures from the C2 curated collection that contained the key phrases “allo” or “graft” to identify potential alloreactivity signatures. To test for differences in signature enrichment between cells belonging to MLR+ and MLR- clonotypes within an organ, we used the Wilcoxon test.

**RESULTS**

**Efficient Amplification of RM TCR Alpha and Beta Chains**

Immunological studies in RM have been performed for decades, and have significantly advanced our understanding of adaptive T cell immunity in health and disease. However, in-depth characterization of RM T cell clonality and associated T cell function on a single cell level have been hindered by a paucity of available methods, particularly the ability to efficiently amplify and reconstruct the RM alpha and beta TCR chains. Here we present a robust platform for RM TCR amplification leveraging the well-established human 5’ 10x Genomics single cell sequencing platform. To accomplish this, we optimized primer pairs adapted to utilize a nested PCR approach analogous to that of the 10x human and mouse workflows, thereby maximizing the specificity of amplification, with primer sets targeting the constant regions of the RM alpha and beta chains (‘TRAC’ and ‘TRBC’, respectively). The primers targeting the beta chain were specifically designed to amplify both TRBC1 and TRBC2. Primer length, CG content and annealing temperatures were adjusted to be used with the forward primers supplied in the human 5’ 10x Genomics kit. Cycling conditions were modified from the original 10x protocol to allow optimal amplification of the targeted region in RM (Tables 1, 2). We have named this new resource ‘RM-scTCR-Seq’.

RM-scTCR-Seq primers annealed to the alpha and beta RM constant region and amplified transcripts covering the complete alpha and beta loci including the variable V and J regions for the alpha as well as V, D and J regions for the beta locus.
Our assembled contigs extend past the V region, which likely represents unannotated UTR sequences. Such UTR sequences are annotated in the human and mouse VDJ references (40) and can be seen in the assembled contigs for 10x Genomics’ human and mouse sample datasets.

We assessed the quality of our primers using summary statistics from 10x’s ‘Cellranger’ software (a pipeline for alignment, quantification, and evaluation of 10x Genomics-generated single-cell libraries), and found that our primers consistently amplify high-quality reads from the TCR alpha and beta chains. For this assessment, we prepared peripheral blood mononuclear cells (PBMCs) from three representative RMs (animal IDs: R.276, R.279, R.319) and used these PBMCs to perform RM-scTCR-Seq. We observed that 49-52% of the aligned reads mapped to the alpha chain and 22-25% of the reads aligned to the beta chain (Figure 1B, ii and Table S2) in the three RMs, with >=73% of total reads mapping to any V(D)J gene (including to gamma and delta chain sequences).

**Figure 1** | RM primers amplify reads from alpha and beta chains, and enable TCR reconstruction. (A) Pile up of reads on the alpha chain (top) and beta chain (bottom) from a reconstructed RM TCR. (B) Barplots of “Cellranger vdj” summary statistics for three RM PBMC T cell samples. (i) Percent of reads in each sample that map to the alpha chain of the TCR, (ii) Percent of reads in each sample that map to the beta chain of the TCR, (iii) Percent of cells (defined as droplets identified by Cellranger as containing a valid cell) with a reconstructed CDR3-annotated alpha chain contig, (iv) Percent of cells (defined as droplets identified by Cellranger as containing a valid cell) with a reconstructed CDR3-annotated beta chain contig, (v) Percent of cells with a productive (meaning no premature stop codons) alpha chain and a productive beta chain, both of which span the start of a V sequence to the end of a J sequence.
In addition to demonstrating that the sequencing reads mapped to known TCR loci, we were also able to reconstruct the alpha and beta chains from these reads (Figure 1Biii, iv and Table S2). The Complementarity-determining region (CDR) 3 is the most variable of the CDR regions, and spans the V and J regions; thus, identification of this sequence is useful as a benchmark for reconstruction quality (41). As such, an ideal reconstruction of the alpha chain or beta chain would completely span the V-to-J region, and enable inference of the CDR3 sequence (42, 43). Using the preceding criteria as our standard for a high-quality alpha or beta TCR chain reconstruction, we observed that 50 to 67% of the 10X droplets identified as cells by Cellranger produced a full length, CDR3-annotated alpha chain (Figure 1Biii and Table S2). Inspecting the beta chain, we found that a high-quality beta chain could be constructed for 83% to 94% for these same cells (Figure 1Biv and Table S2). Full reconstruction of the TCR requires both an alpha chain and a beta chain that span the V and J regions, and ideally are productive (no premature stop codons). We identified both alpha and beta chains in 32 to 47% of cells examined (Figure 1Bv and Table S2).

These data demonstrate, for the first time, efficient amplification of the RM TCR alpha and beta region by utilizing optimized RM primer pairs compatible with the human 5’ 10x Genomics single cell sequencing platform. Amplified reads enabled high quality beta chain reconstruction in almost all, and reconstruction of full TCRs (alpha and beta chains) in a significant number of sequenced cells.

**RM-scTCR Seq Enables Tracking of T Cell Clonotypes in Allogenic Mixed Lymphocyte Reactions**

Identification of T cell clonotypes using in vitro assays or in vivo disease models enables the evaluation of the clonal repertoire, clonal dynamics, and the tracking of specific T cells over time. To validate our single cell TCR sequencing and analysis pipeline, we first utilized an in vitro allo-proliferation assay: the mixed lymphocyte reaction (MLR) (44–47). To perform these studies, PBMCs from 4 RM were irradiated with 3500 cGy and used to stimulate “responder” PBMCs from MHC-disparate RM. The responder PBMCs were labeled with cell trace violet (CTV), a cell dye which dilutes with proliferation. After five days of allo-stimulation, responder T cell were sorted into three populations: CTV High (non-proliferating T cells), CTV Mid (intermediate proliferation, 3-5 cell divisions) and CTV Low (highly proliferating T cells, demonstrating > 7 cell divisions by flow cytometry), (Figures 2A, B). We applied the RM-scTCR-Seq pipeline to analyze TCR diversity, measured by the nucleotide sequence of samples between different donors, using either alpha chain MI (MI < 0.08) or beta chain MI (MI <0.001), demonstrate the reproducibility of a single repertoire between different samples (48, 50) between all pairs of samples from the MLR experiment. We analyzed these data using the CDR3 regions of the alpha chain (Figure 2E) and the beta chain (Figure 2F).

As has been previously demonstrated, the vast majority of TCR clonotypes in our assay were private to the individual animal, with low levels of similarity when comparing TCR clonotypes between different animals (MI < 0.08, indicating minimal overlap in the TCR repertoire, Figures 2E, F). As expected, the highest level of similarity was observed between responder PBMCs in the ‘Blood’ sample (T cells isolated from peripheral blood pre MLR) and T cells in the non-proliferating population (‘CTV high’). Importantly we were also able to identify clonotypes that existed in both the Blood sample and the high-proliferating population (Figures 2E, F), substantiating the ability of RM-scTCR-Seq to track clonotypes as they proliferated in vitro. The low Morisita indices for comparison of samples between different donors, using either alpha chain MI (MI < 0.08) or beta chain MI (MI <0.001), demonstrate the specificity of the RM-scTCR-Seq approach (Figures 2E, F) and validate that clonotype tracking can be performed by utilizing the TCR alpha or beta chains.

**In Vitro Predicted Alloreactive T Cell Clonotypes Exhibit an In Vivo aGVHD Expression Program in the Spleen and Liver**

We next performed in vivo clonotype tracking in HCT recipients who developed aGVHD, in order to trace specific TCRs previously identified as allo-proliferating in an in vitro MLR. To accomplish this, we first performed MLR assays using donor and recipient PBMCs from transplants in which recipients subsequently developed severe aGVHD (28). For these assays, responder PBMCs were from donor animals R.276, R.279, R.319, and stimulator PBMCs were from recipient animals R.227, R.276 and R.312, respectively (28). Terminal analysis of spleen and liver infiltrating donor T cells was performed 8 days after HCT, during active aGVHD (28). Sorted T cells were analyzed with CTV low samples, after applying the Hutcheson t-test was significant with a p-value of < 0.01 for all four samples tested (49). Using the RM-scTCR-Seq pipeline, we were also able to track individual clonotypes throughout the three MLR peaks (Figure 2D) using either their alpha or their beta chain clonotype. While we are able to identify individual clones and track clonal changes over time, it should be noted that the 10X platform is limited in its ability to measure large-scale T cell population-based clonal expansion and deletion, given that 10X is designed to analyze ~10,000 cells. Bulk RM TCR Seq, which has the capacity to measure millions of input cells, will be more suitable to answer these population-based questions. A RMspecific platform for bulk TCRseq is currently under development by our research group.

To further assess our ability to accurately track T cell clonotypes in vitro, and to quantify TCR similarities/differences between T cells isolated from different animals, we calculated the Morisita Index (‘MI’, measuring the similarity of two repertoires, or, the reproducibility of a single repertoire between different samples) (48, 50) between all pairs of samples from the MLR experiment. We analyzed these data using the CDR3 regions of the alpha chain (Figure 2E) and the beta chain (Figure 2F).
scRNA-Seq using the 10X platform, and clonotype-tracking was performed using the RM-scTCR-Seq pipeline. The number of cells sequenced and the number of cells with unique alpha and beta TCR chains are found in Table S2.

We first assessed our ability to track clonotypes across samples by applying the MI to all of the scTCR-Seq libraries (Figure 3A). As expected, in all three animals, measurable TCR clonal overlap could be detected in T cells sorted from recipient liver and spleen, consistent with infiltration of activated T cell clonotypes into both of these organs. Overlap of alloreactive (highly proliferating) clonotypes from the MLR with clonotypes derived from both the spleen and the liver was also evident (Figure 3A). Next, we

![Figure 2](image-url)

**FIGURE 2** | Application of RM primers to an MLR experiment demonstrate the ability to track T cell clonotypes over across samples with high accuracy. (A) Experimental schema for MLR experiments. (B) CTV staining and sorting strategy for one representative RM. (C) Shannon Index for each of the sorted samples from four individual MLRs. (D) Representative graph of T cell clonotypes detected in the peripheral blood (Pre-MLR) and from all three sorted samples (CTV-High, CTV-Mid, CTV-Low) for R.330. (E) Heatmap of Morisita Index for all MLR and peripheral blood samples, using the CDR3 region of the alpha chain to group cells into clonotypes. (F) Heatmap of Morisita Index for all MLR and peripheral blood samples, using the CDR3 region of the beta chain to group cells into clonotypes.
FIGURE 3 | Allo-proliferative T cell clonotypes identified in an MLR exhibit aGVHD-specific gene expression programs in the spleen and the liver. (A) Heatmap of the Morisita Index for all analyzed samples. (B) Percentage of cells assigned to a specific T cell clonotype using RM-scTCR-Seq that also had high-quality gene expression data using the 10X gene expression platform. (C) UMAP of cells from each donor NHP, colored by sample type. Blood = T cells sorted from the blood. MLR = T cells subjected to a MLR. Spleen = T cells purified from the spleen during severe aGVHD. Liver = T cells purified from the liver during severe aGVHD. (D) Volcano plots of differential expression between MLR+ and MLR- detected clonotypes within one liver sample. (E) Volcano plots of differential expression between MLR+ and MLR- detected clonotypes within two spleen samples. (F) Violin plot of GZMB expression between MLR+ (red) and MLR- (blue) cells within two spleen samples (from R.279 and R.319) and one liver sample (from R.279). (G) Violin plot of T cells scored with the Ichiba Graft Versus Host Disease signature (35), colored by the detection of a clonotype in MLR+ (red) or MLR- (blue) samples. P <0.05 for the comparisons in the liver, and both spleen samples. (H) Violin plot of T cells scored with the KEGG Graft Versus Host Disease signature (51) colored by the detection of a clonotype in MLR+ (red) or MLR- (blue) samples. P <0.001 for the comparisons in the liver, and both spleen samples.
confirmed that the majority of cells with trackable TCRs by RM-scTCR-Seq cells also had a corresponding barcode in the scRNA-Seq library after quality control and filtering. We found that at least 75% of the cells with a valid scTCR-Seq library (assessed by Cellranger) in each sample also had a valid corresponding scRNA-Seq barcode (Figure 3B) confirming the efficiency of the assay. After alignment, filtering based on quality metrics, and dimensionality reduction, we plotted a UMAP of the analyzed cells (n = 23,618 total cells analyzed from three RM) separated according to each transplant recipient (Figure 3C and Table S2 - GEX Cell Count After Filtering).

We then characterized cells isolated from the spleen and liver by whether they could be matched with an alloproliferating MLR clonotype (termed ‘MLR+’) or not (termed ‘MLR-’). Each MLR+ cluster incorporated at least 37 cells (Table 3). Using these definitions of MLR+ and MLR-, we conducted a gene expression test of MLR+ versus MLR- T cells within two spleen samples and one liver sample with large numbers of MLR clonotypes, to identify differentially expressed (DE) genes. The volcano plots in Figures 3D, E demonstrate these DE genes, and highlight the top 5 annotated genes in each comparison (top genes with only an Ensembl ID are also identified). This analysis identified several known markers of alloreactivity in each sample. GZMB, a well-recognized aGVHD-associated gene (28, 52) was identified as a DE gene in both spleen samples (Table S3). We further interrogated the expression of GZMB at the level of each individual cell, to understand its variability within the MLR+ and MLR- populations. As expected, a violin plot of GZMB (Figure 3F) demonstrated that the spleen MLR+ samples had higher, less variable expression of GZMB compared to the spleen MLR- samples, which demonstrated a wider distribution of expression, with a larger proportion of low-expressing cells. To further evaluate gene expression in the MLR+ and MLR- cells, we used VISION (53) to score each cell using a set of previously curated immunological and alloreactivity signatures. As shown in Figures 3G, H, MLR+ cells from the liver and spleen enriched for the Ichiha: Graft versus Host Disease gene signature compared to MLR- cells (Figure 3G, p < 0.05) (35) as well as for the KEGG Graft versus host disease signature (Figure 3H, p < 0.01) (51).

These results demonstrate the feasibility of combining RM-scTCR-Seq and scRNA-Seq to identify and interrogate alloreactive T cell clonotypes in aGVHD target organs in RM, providing proof-of-concept of the in vivo utility of the RM-scTCR-Seq pipeline.

### DISCUSSION

TCR repertoire analysis has become a fundamental tool to understand the immune responses to infections and vaccines, as well as the immunopathogenesis of rejection after solid organ transplantation, and GVHD after HCT (2, 46, 47, 54–58). In murine and human studies, the introduction of VDJ target enrichment combined with single cell sequencing platforms has enabled the interrogation of transcriptional profiling of single T cells at an unprecedented level of accuracy (42, 43). However, paired single T cell receptor and transcriptional studies in RM have been impeded by the lack of a validated platform to amplify and reconstruct alpha and beta TCR sequences (59–61). This study demonstrates successful identification and tracking of RM single T cell clonotypes by utilizing customized primers compatible with the human 5’ 10x Genomics single cell sequencing platform, and by significantly improving alignment of RM sequencing results by creating a custom reference for the RM TCR alpha and beta chains.

Prior to the work described herein, two major limitations significantly hindered the development of T cell clonotype tracking in RM. They were (1): The lack of optimized, single cell sequencing-compatible primers covering both the TCR alpha and beta region for RM; (2) The poor genomic annotation of the alpha and beta loci in RM, which resulted in inadequate reconstruction of amplified VDJ PCR reads. To address both of these technical barriers, we have designed a robust set of RM-specific TCR primer pairs which anneal to the constant region of the alpha and beta TCR loci, and which are compatible with the commonly used single cell sequencing platform from 10x Genomics. Optimizing PCR cycling and temperature conditions resulted in efficient amplification of the alpha and beta regions of the RM TCR and generated productive TCR reconstructions. Additionally, we created an updated reference for the TCR alpha and beta chain region, enhancing our ability to annotate the assembled RM TCR’s.

We demonstrated amplification of highly specific TCR regions with >73% of total reads mapping to any V(D)J gene. We note that this mapping efficiency is somewhat less than the ~91% of reads mapping to the human VDJ reference, in a human dataset provided by 10x (www.10xgenomics.com/resources/dataset). The difference in the alignment rate may be explained by differing levels of annotation for these organisms: 10x provides human references which have been refined over time, but this resource does not exist for RM, which necessitated our creation of a new reference (see Methods) by combining elements of the Ensembl and IMGT references, and then carefully filtering out sequences identified by 10x’s enclone software that contained errors (such as premature stop codons, see Methods). As annotation of the RM TCR improves, we anticipate some gains in read mapping. As mentioned earlier, we have noticed that UTR regions are annotated in the human and mouse VDJ references, and we expect that addition of these sequences to the RM reference will improve alignment rates. We additionally tested whether our reads adequately covered all genomic V(D) regions and confirmed that in the 30 primary VDJ samples analyzed, reads mapped to 91% (179 of 196) of all V(D)J regions from our reference (Table S5). Additionally, Figure S1 provides

### TABLE 3 | Sample sizes for MLR+ and MLR- cells in GEX data.

| NHP   | Organ | MLR Status | Cell Count |
|-------|-------|------------|------------|
| R.279 | Liver | MLR-       | 4,025      |
| R.279 | Liver | MLR+       | 58         |
| R.279 | Spleen| MLR-       | 2,773      |
| R.279 | Spleen| MLR+       | 37         |
| R.319 | Spleen| MLR-       | 4,632      |
| R.319 | Spleen| MLR+       | 39         |
detailed information regarding alpha-chain, beta-chain, and CDR3 region length, which were all with the expected length range, further supporting the robustness of this method.

Sufficient alignment was also demonstrated by the ability of the RM-scTCR-Seq pipeline to detect allo-proliferating clonotypes using an in vivo RM model of aGVHD. While bulk TCR-Seq is well-suited for providing a comprehensive profile of the diversity of T cell clonotypes in a given sample due to its ability to profile large numbers of cells (59, 61), scTCR-Seq is unique in its ability to be paired with other single-cell technologies, allowing users to tag each T cell with its clonal identity, and measure each cell’s gene expression profile as performed here (9). These clonotypes can also be interrogated for their protein expression or epigenetic profiles, or other features of interest (62, 63). Most importantly, these individual clonotypes can be tracked in vivo, to better map the clonal architecture of protective versus pathogenic T cell responses, in target tissues that are far more accessible in NHP than in patient samples. In this study utilizing RM-scTCR-Seq, we were able to detect allo-proliferative clonotypes in in vitro MLR experiments and track their in vivo alloreactive and cytotoxic potential by transcriptional analysis of tissue-infiltrating T cells. We expect that future RM-scTCR-Seq studies from GVHD target tissues including the GI tract and skin will provide further insights into mechanisms controlling T cell alloreactivity, and enhance our ability to identify GVHD-specific targetable pathways.

Given the central importance of NHP models to our understanding of protective and pathogenic immunity, and to the development of novel immune-based therapeutics, the application of RM-scTCR-Seq is expected to have wide application, and further advance the ability of NHP to model human disease.

**DATA AVAILABILITY STATEMENT**

The original contributions presented in the study are publicly available. This data can be found here: https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE190268.

**REFERENCES**

1. Arstila TP, Casrouge A, Baron V, Even J, Kanellopoulos J, Kourilsky P. A Direct Estimate of the Human Alphabeta T Cell Receptor Diversity. *Science* (1999) 286(5441):958–61. doi: 10.1126/science.286.5441.958

2. Suessmuth Y, Mukherjee R, Watkins B, Koura DT, Finstermeier K, Desmarais C, et al. CMV Reactivation Drives Posttransplant T-Cell Reconstitution and Results in Defects in the Underlying TCRbeta Repertoire. *Blood* (2015) 125(25):3835–50. doi: 10.1182/blood-2015-03-631853

3. Li B, Li T, Pignon JC, Wang B, Wang J, Shukla SA, et al. Landscape of Tumor-Infiltrating T Cell Repertoire of Human Cancers. *Nat Genet* (2016) 48(7):725–32. doi: 10.1038/ng.3581

4. Dziubianau M, Hecht J, Kuchenbecker L, Sattler A, Rodelsperger C, et al. TCR Repertoire Analysis by Next Generation Sequencing Allows Complex Differential Diagnosis of T Cell-Related Pathology. *Am J Transplant* (2013) 13(11):2842–54. doi: 10.1111/ajt.12431

5. Pogorelyy MV, Minervina AA, Touzel MP, Sycheva AL, Komech EA, Kovalenko EI, et al. Precise Tracking of Vaccine-Responding T Cell Clones Reveals Convergent and Personalized Response in Identical Twins. *Proc Natl Acad Sci USA* (2018) 115(50):12704–9. doi: 10.1073/pnas.1809642115

6. Freeman JD, Warren RL, Webb JR, Nelson BH, Holt RA. Profiling the T-Cell Receptor Beta-Chain Repertoire by Massively Parallel Sequencing. *Genome Res* (2009) 19(10):1817–24. doi: 10.1101/gr.092924.109

7. Robins HS, Campegher PV, Srivastava SK, Wacher A, Turtle CJ, Kahsai O, et al. Comprehensive Assessment of T-Cell Receptor Beta-Chain Diversity in Alphabeta T Cells. *Blood* (2009) 114(19):4099–107. doi: 10.1182/blood-2009-04-217604

8. Carter JA, Preall JB, Grigaityte K, Goldfuss SJ, Jeffery E, Briggs AW, et al. Single T Cell Sequencing Demonstrates the Functional Role of Alphabeta TCR Pairing in Cell Lineage and Antigen Specificity. *Front Immunol* (2019) 10:1516. doi: 10.3389/fimmu.2019.01516

9. Stubbington MJT, Lomberg T, Proserpio V, Clare S, Speak AO, Dougan G, et al. T Cell Fate and Clonality Inference From Single-Cell Transcriptomes. *Nat Methods* (2016) 13(4):329–32. doi: 10.1038/nmeth.3800

10. Buggert M, Vella LA, Nguyen S, Wu VH, Chen Z, Sekine T, et al. The Identity of Human Tissue-Emigrant CD8(+) T Cells. *Cell* (2020) 183(7):1946–61.e15. doi: 10.1016/j.cell.2020.11.019

11. Merkenschlager J, Finkin S, Ramos V, Kraft J, Cipolla M, Nowosad CR, et al. Dynamic Regulation of TFH Selection During the Germinal Centre Reaction. *Nature* (2021) 591(7850):458–63. doi: 10.1038/s41586-021-03187-x

**ETHICS STATEMENT**

The animal study was reviewed and approved by Massachusetts General Hospital Animal Care and Use Committee and Biomere Animal Care and Use Committee.

**AUTHOR CONTRIBUTIONS**

Conceptualization: UG, RAF, VT, and NK. Methodology: UG, RAF, JK, VT, and CM. Investigation: RAF, UG, JK, XR, LC, JFL, and CM. Visualization: JK and UG. Supervision: NK, AS, PK, and LC. Writing—original draft: UG, JK, and NK. Writing—review & editing: UG, JK, RAF, XR, JFL, PK, LC, VT, AKS, and NK. All authors contributed to the article and approved the submitted version.

**FUNDING**

UG is supported by the Helen Gurley Brown Foundation and the National Cancer Center. JK is supported by the NIH (5T32HL007574-40) and the National Cancer Center. VT is supported by an ASTCT New Investigator Award and the CIBMTR/Be The Match Foundation Amy Strelzer Manasevit Research Program Award. AS is supported by the Searle Scholars Program, the Beckman Young Investigator Program, a Sloan Fellowship in Chemistry, and the NIH (5U24AI111867, 2R01HL095791). NK is supported by NIH grants U19 AI105173, R01 HL095791, P01 HL158504, and by a Leukemia and Lymphoma Society TRP grant.

**SUPPLEMENTARY MATERIAL**

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fimmu.2021.804932/full#supplementary-material
51. Kanehisa M, Goto S. KEGG: Kyoto Encyclopedia of Genes and Genomes. Nucleic Acids Res (2000) 28(1):27–30. doi: 10.1093/nar/28.1.27
52. Graubert TA, Russell JH, Ley TJ. The Role of Granzyme B in Murine Models of Acute Graft-Versus-Host Disease and Graft Rejection. Blood (1996) 87(4):1232–7. doi: 10.1182/blood.V87.4.1232.bloodjournal8741232
53. DeTomaso D, Jones MG, Subramaniam M, Ashua T, Ye CJ, Yosef N. Functional Interpretation of Single Cell Similarity Maps. Nat Commun (2019) 10(1):4376. doi: 10.1038/s41467-019-12235-0
54. Joshi K, Milighetti M, Chain BM. Application of T Cell Receptor (TCR) Repertoire Analysis for the Advancement of Cancer Immunotherapy. Curr Opin Immunol (2021) 74:1–8. doi: 10.1016/j.coi.2021.07.006
55. Smith C, Corvino D, Beagley L, Rehan S, Neller MA, Crooks P, et al. T Cell Repertoire Remodeling Following Post-Transplant T Cell Therapy Coincides With Clinical Response. J Clin Invest (2019) 129(11):5020–32. doi: 10.1172/JCI128323
56. Nikolich-Zugich J, Slifa MK, Messaoudi I. The Many Important Facets of T-Cell Repertoire Diversity. Nat Rev Immunol (2004) 4(2):123–32. doi: 10.1038/nri1292
57. Babel N, Stervbo U, Reinke P, Volk HD. The Identity Card of T Cells: Clinical Utility of T-Cell Receptor Repertoire Analysis in Transplantation. Transplantation (2019) 103(8):1544–55. doi: 10.1097/TP.0000000000002776
58. Meier JA, Haque M, Fawaz M, Abdeen H, Coffey D, Towlerton A, et al. T Cell Repertoire Evolution After Allogeneic Bone Marrow Transplantation: An Organizational Perspective. Biol Blood Marrow Transplant (2019) 25(5):868–82. doi: 10.1016/j.bbmt.2019.01.021
59. Brochu HN, Tseng E, Smith E, Thomas MJ, Jones AM, Diveley KR, et al. Systematic Profiling of Full-Length Ig and TCR Repertoire Diversity in Rhesus Macaque Through Long Read Transcriptome Sequencing. J Immunol (2020) 204(12):3434–44. doi: 10.4049/jimmunol.1901256
60. Rosenfeld R, Zvi A, Winter E, Hope R, Isaiah O, Mazor O, et al. A Primer Set for Comprehensive Amplification of V-Genes From Rhesus Macaque Origin Based on Repertoire Sequencing. J Immunol Methods (2019) 465:67–71. doi: 10.1016/j.jim.2018.11.011
61. Li Z, Liu G, Tong Y, Zhang M, Xu Y, Qin L, et al. Comprehensive Analysis of the T-Cell Receptor Beta Chain Gene in Rhesus Monkey by High Throughput Sequencing. Sci Rep (2015) 5:10092. doi: 10.1038/srep10092
62. Ma S, Zhang B, LaFave LM, Earl AS, Chiang Z, Hu Y, et al. Chromatin Potential Identified by Shared Single-Cell Profiling of RNA and Chromatin. Cell (2020) 183(4):1103–16 e20. doi: 10.1016/j.cell.2020.09.056
63. Angermueller C, Clark SJ, Lee HJ, Macaulay IC, Teng MJ, Hu TX, et al. Parallel Single-Cell Sequencing Links Transcriptional and Epigenetic Heterogeneity. Nat Methods (2016) 13(3):229–32. doi: 10.1038/nmeth.3728

Conflict of Interest: AS reports compensation for consulting and/or SAB membership from Merck, Honeycomb Biotechnologies, Cellarity, Repertoire Immune Medicines, Hoviosne, Third Rock Ventures, Ochre Bio, FL82, and Dahlia Biosciences unrelated to this work. AS has received research support from Merck, Novartis, Leo Pharma, Janssen, the Bill and Melinda Gates Foundation, the Moore Foundation, the Pew-Stewart Trust, Foundation MIT, the Chan Zuckerberg Initiative, Novo Nordisk and the FDA unrelated to this work. LK is on the scientific advisory board for HiFiBio and Mammoth Biosciences. She reports research funding from Kymab Limited, Magenta Therapeutics, BlueBird Bio, and Regeneron Pharmaceuticals. She reports consulting fees from Equilum, FortySeven Inc, Novartis Inc, EMD Serono, Gilleed Sciences, Vertex Pharmaceuticals, and Takeda Pharmaceuticals. LK reports grants and personal fees from Bristol Myers Squibb that are managed under an agreement with Harvard Medical School.

The remaining authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher’s Note: All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Copyright © 2022 Gerdemann, Fleming, Kaminski, McGuckin, Rui, Lane, Keskula, Cagnin, Shulek, Tkachev and Kean. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.