Modeling Limit of Detection for Dominant T Cell Receptor Clones Reveals Comparable Measures Across RNA and DNA Based Sequencing

Sagar N. Sharma BS[1,2], Samuel C. Butler BS[1], Jennifer M. Bone PhD[1], Pranav Murthy MS[1], Michael T. Lotze MD[1,2]

[1]Department of Surgery, University of Pittsburgh, Pittsburgh, PA, USA; [2]Nurix Therapeutics, San Francisco, CA

Dominant Tumor-Infiltrating Lymphocytes (TIL) Clones are Tumor Reactive

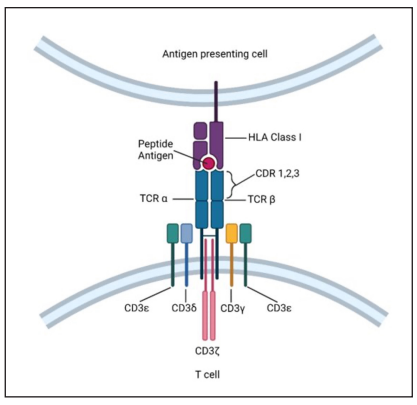

Figure 1. The T Cell Receptor (TCR) complex is a heterodimer composed of an α and β chain along with CD3-associated proteins. Complimentary-determining regions (CDRs) are related to binding with peptide antigens loaded on the human leukocyte antigen (HLA) complex. Sequencing of the CDR3 region of the β chain allows for identification of unique clones and thus quantification of the T cell repertoire.

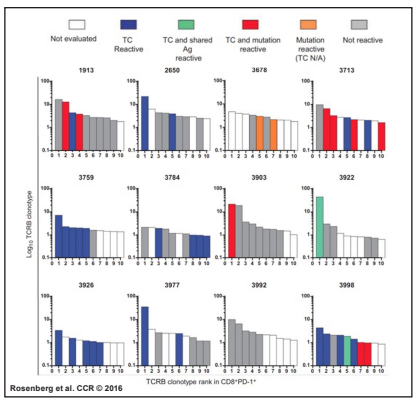

Figure 2. CD8+ PD-1+ T cells were isolated from tumor digest obtained from 12 patients with metastatic melanoma. DNA based β chain sequencing was combined with pairSEQ to identify the cells' αβ pairing. Transduction of top 10 most abundant TCRs revealed that 11 out of 12 patients possess at least 2 or more of these TCRs that show *in vitro* reactivity.

TCRβ Sequencing: RNA vs DNA

Method	RNA[1]	DNA[2]
Minimum input quantity for PBMC	300-500ng	3380-4700ng
CDR3 sequencing	Functional sequences	Includes introns, exons, and nonproductive sequences
CDR3 expression	Multiple transcripts per cell	Single copy per cell
Error control	Unique Molecular Identifiers (UMI)	Synthetic TCRG template
Cost per sample	$1375	$2200
TCR chains	β α δ γ	β, α/δ, γ
Time to service	2 weeks	2 weeks
Data sharing	By request	IMMUNE-ACCESS Public Database

[1]iRepertoire, Inc. [2]Adaptive Biotechnologies

Table 1. A comparison of some of the key similarities and differences in RNA vs DNA based sequencing methods. Calculation of pricing for Adaptive includes computational analysis and expedited data fee. Adaptive offers services for either β, α and δ, or γ chain sequencing, whereas iRepertoire offers any custom TCR chain combination

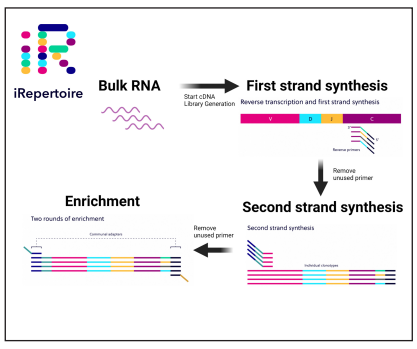

Figure 3. Sample bulk RNA is first reverse transcribed into cDNA for the synthesis of the first strand using a gene-specific primer mix. Physical and chemical means remove residual primers to reduce the number of primer-dimers. A second PCR round completes the cDNA synthesis followed by another wash step. Two rounds of enrichment are completed using the communal adapter sequences introduced earlier before pooling for sequencing. Unique molecule identifiers (UMI) are included in the initial PCR steps to reduce amplification bias.

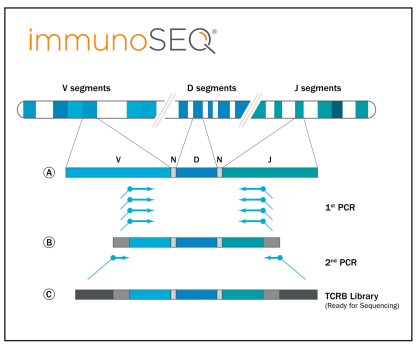

Figure 4. A synthetic TCRG repertoire was created to represent all possible VJ-gene combinations to titrate primers before the first round of PCR to reduce amplification bias. Sample bulk DNA is mixed with multiplex V- and J- primers before going through two rounds of PCR and the addition of barcodes and adaptor sequences for sequencing. Post amplification, the synthetic template can be further used to computationally correct any residual bias.

Experimental Overview: Modeling Limit of Detection

Sample	MNCs	PBMCs	Dilution Factor
1	5 million	0	-
2	0	5 million	-
3	5	5 million	10^6
4	50	5 million	10^5
5	500	5 million	10^4
6	5000	5 million	10^3

Table 2. Mononuclear cells (MNCs) from a single donor were serially diluted in a diverse pool of peripheral blood mononuclear cells (PBMCs). Sample 1 and 2 were used as reference repertoires for MNCs and PBMCs, respectively.

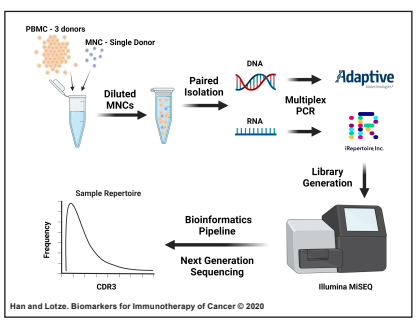

Figure 5. RNA and DNA was isolated and sent for cDNA library generation and Next-Generation Sequencing (NGS) by iRepertoire and Adaptive, respectively. The proprietary bioinformatic protocols were followed to generate CDR3 maps that contained the amino acid sequence and its respective frequency within the sample.

Top 10 Most Abundant Clones Capture Head of CDR3 Distribution

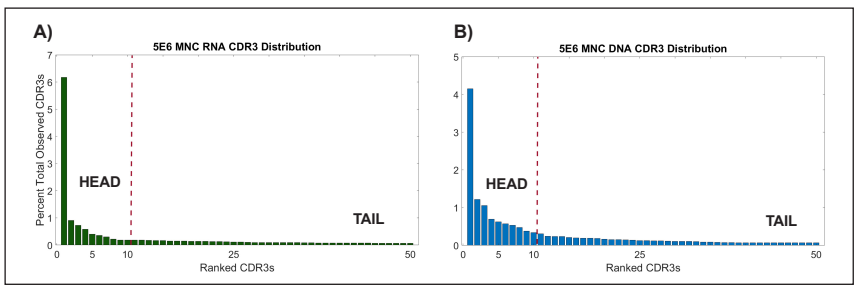

Figure 6. Shown are the top 50 most abundant CDR3 sequences for the undiluted RNA (**A**) and DNA (**B**) MNC samples. The x-axis denotes the rank or relative abundance of the CDR3, and the y-axis shows the percent of the sample's repertoire that CDR3 sequence occupies. For example, the most abundant MNC clone in the RNA CDR3 distribution makes up 6.177% of the total reads from that sample. The dashed red line denotes the head (left) or top 10 most abundant CDR3s and the tail (right) illustrates the decreasingly abundant CDR3s. Clones to the left of the cutoff are deemed dominant for the purposes of all following analyses.

Dominant CDR3s are Reproducibly Detected

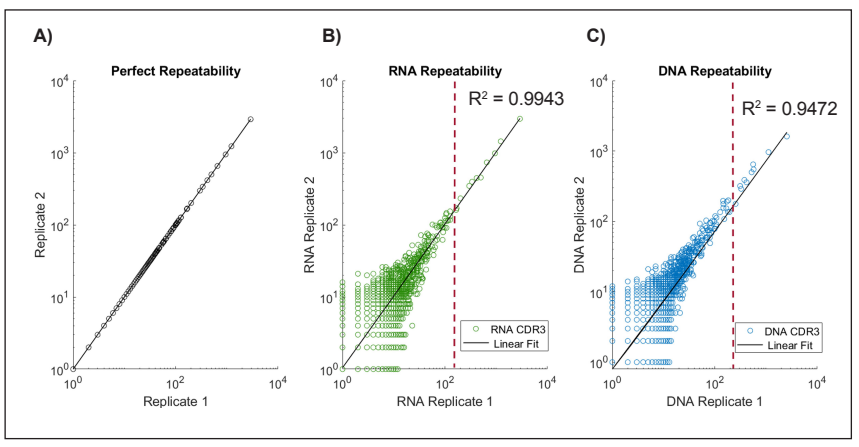

Figure 7. Open circles on plots (**A-C**) represent a unique CDR3 sequence where the x- or y-value would be its abundance on a log scale in replicate 1 and 2, respectively. The right most CDR3s are the most abundant. Perfect reproducibility (**A**) is a straight line through the origin where the same CDR3s have the exact same abundance in each replicate. Plots for RNA (**B**) and DNA (**C**) show significant deviation from linearity with CDR3s that have a low abundance. The dashed red line divides each plot into the top 10 clones (right) and the rest (left). In both RNA and DNA, the top 10 clones were highly linear (R^2 = 0.9943 and 0.9472, respectively). The rest of the clones for RNA and DNA were poorly correlated between the replicates (R^2 = 0.5001 and 0.5991, respectively).

RNA and DNA Detect Dominant MNC CDR3s and Maintain Sample Hierarchy

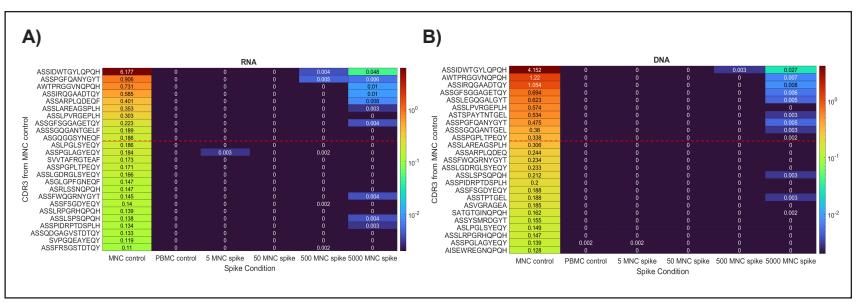

Figure 8. Sharing of the 25 most abundant clones from the undiluted reference MNC sample is shown. Each value shown is the percent of that sample's repertoire that is occupied by the specific CDR3 amino acid sequence that borders the left side of the MNC control column. The dashed red line delineates the top 10 (top) from the rest of the clones (bottom). All values ≤ 0.001% were considered equal to zero for visual clarity. There is no significant sharing between the MNC and PBMC reference columns for both RNA (**A**) and DNA (**B**). Samples with MNCs diluted at factors of 10^6 and 10^5 did not harbor detectable dominant clones. The most dominant clonotype from the MNC control (ASSIDWTGYLQPQH) was the most dominant of shared clones found in the 5000(1:10^3) and 500(1:10^4) MNC spike in samples for both RNA and DNA. The hierarchy of the original MNC sample is maintained in both the 10^3 and 10^4 diluted MNC sample for RNA and DNA. Clones outside the top 10 had little to no detectable signal.

Summary

Conclusions

- The top 10 CDR3s are reproducible and capture a substantial portion of the head of the long tail CDR3 distribution.
- The LOD for diluted populations is 1:10^3 for the most abundant clones and 1:10^4 for the top clone.
- Detection of dominant clones is consistent across DNA and RNA based methods.

Limitations

- Dilution factors 10^2 and 10^1 were not tested and may have provided a clearer illustration of sample hierarchy maintenance.
- Cutoff for dominant clone classification should be validated on another test cohort.

Future Directions

- Due to codon degeneracy, use of nucleotide sequences may be a more accurate method of detecting shared clones.
- Screening shared TCRs through a public database may help ensure shared clones are truly private.

Presented at the 24th Annual TRCCC Meeting, June 6-8, 2022, Seven Springs PA, USA