Ultra-Efficient Short Read Sequencing of T Cell Receptor Repertoires

Janelle M. Montagne[1], Xuwen Alice Zheng[1], Iago Pinal-Fernandez[2,3], Jose C. Milisenda[4], Lisa Christopher-Stine[5], Thomas E. Lloyd[3], Andrew L. Mammen[2,3], & H. Benjamin Larman[1,*]

[1]Division of Immunology, Pathology Department, Johns Hopkins University School of Medicine, Baltimore, MD, USA.

[2]Muscle Disease Unit, Laboratory of Muscle Stem Cells and Gene Regulations, National Institute of Arthritis and Musculoskeletal and Skin Diseases, National Institutes of Health, Bethesda, MD, USA.

[3]Department of Neurology, Johns Hopkins University School of Medicine, Baltimore, MD, USA.

[4]Internal Medicine Department, Hospital Clinic, Universitat de Barcelona, Barcelona, Spain and Centro de Investigación Médica en Red Enfermedades Raras.

[5]Division of Rheumatology, Johns Hopkins University School of Medicine, Baltimore, MD, USA.

Abstract:

T cell receptor (TCR) repertoire sequencing is increasingly employed to characterize adaptive immune responses. However, current TCR sequencing methodologies are complex and expensive, limiting the scale of feasible studies. Here we present Framework Region 3 AmplifiKation sequencing (FR3AK-seq), a simplified multiplex PCR-based approach for the ultra-efficient analysis of TCR complementarity determining region 3 (CDR3) repertoires. By using minimal primer sets targeting a conserved region adjacent to CDR3, undistorted amplicons are analyzed via short read, single-end sequencing. We find that FR3AK-seq is sensitive and quantitative, performing comparably to two industry standards. FR3AK-seq was utilized to quickly and inexpensively characterize the T cell infiltrates of muscle biopsies obtained from 145 patients with idiopathic inflammatory myopathies and controls. A cluster of related TCRs was identified in samples from patients with sporadic inclusion body myositis, suggesting the presence of a shared antigen-driven response. The ease and minimal cost of FR3AK-seq removes critical barriers to routine, large-scale TCR CDR3 repertoire analyses.

Introduction:

T cell receptor (TCR) repertoire analysis has emerged as a powerful tool for examining adaptive immune responses. TCR repertoires, generated by the process of V(D)J recombination, encompass the T cell clones within a given individual or sample. The unique TCR of each clone defines its antigen specificity, and TCRs can be associated with distinct cellular phenotypes and tissue occupancy. Notably, TCR repertoires encompass dynamic populations of cells that represent past and current immune exposures and reactivities.[1] Development of next generation sequencing (NGS)-based technologies over the last decade has enabled the unprecedented analysis of TCR repertoires.[2–5]

TCR recognition requires antigen processing and presentation of peptide epitopes on autologous major histocompatibility complex (MHC) molecules. Analysis of TCR repertoires both within and between individuals can be indicative of disease status, prior infections or immunizations, and individual-specific attributes of epitope selection.[6–12] Clonally expanded or tissue-infiltrating T cell clones provide evidence of antigen-directed immune responses; integrating TCR repertoire analyses with phenotypic measurements (e.g. single cell transcriptional profiling[13] or flow cytometric sorting of T cell subpopulations[14,15]) can further illuminate qualitative aspects of T cell responses (**Fig 1a**).

A variety of methods have been used to generate TCR libraries for NGS, including multiplex PCR, 5'-RACE, and target enrichment.[2–5,16,17] Generally, these methodologies prioritize sequencing of the hypervariable complementarity determining region 3 (CDR3) of the TCR beta (TCRB) chain. The TCRB CDR3 harbors the greatest amount of sequence diversity, confers most of the antigen specificity to a TCR, and can be used as a unique surrogate for T cell clonal identity.[18–20] Despite their utility, however, current repertoire sequencing approaches remain complex and expensive. These limitations have greatly hindered the utilization of TCR repertoire-based analyses, particularly for studies involving large numbers of samples. We therefore sought to develop a

40 simple, quantitative, multiplex PCR-based approach that prioritizes the ultra-efficient analysis of
41 TCRB CDR3 sequences.

42 The sequence diversity of the TCRB chain variable domain, which in humans is encoded by >120
43 unique TCRBV alleles, necessitates the use of complex primer pools for multiplex PCR
44 amplification. Primer competition and differential amplification efficiencies distort clonal amplicon
45 abundance, confounding quantification.[21] Sophisticated techniques have been developed to
46 computationally correct amplification bias, such as the use of spike-in library standards or the
47 incorporation of unique molecular identifiers (UMIs).[22–24] Adaptive Biotechnologies has become
48 a current industry leader by offering a spike-in standard corrected, multiplex PCR-based assay.
49 At a cost of $500-$1100 per sample, however, the scale of feasible studies using this assay has
50 been greatly constrained. Unique molecular identifier (UMI) based approaches, including the
51 Immunoverse assay (ArcherDX), provide quantitation and enhanced sequence accuracy by
52 resolving amplification and sequencing errors.[23–28] However, UMI approaches typically require
53 multi-day protocols, complicated analytical pipelines, and deeper sequencing to obtain sufficient
54 sampling of distorted libraries.[17,29] By designing maximally compact primer sets and a streamlined
55 workflow, we have essentially eliminated PCR amplification bias while maintaining high sequence
56 level accuracy. Lower sequencing depth requirements, coupled with CDR3 analysis via single-
57 end, short-read (100 nucleotide) sequencing, dramatically reduces the cost associated with this
58 approach to TCR repertoire analysis.

59 Here, we benchmark our simplified multiplex PCR-based assay, which we call Framework Region
60 3 AmplifiKation Sequencing (FR3AK-seq), against two different industry standards for quantitative
61 TCR repertoire sequencing: Adaptive Biotechnologies' multiplex PCR-based hsTCRB
62 immunoSEQ assay and ArcherDX's UMI-based Immunoverse HS TCR assay. We found that
63 FR3AK-seq data, which was analyzed using open-source software, was in excellent agreement
64 with both. Finally, we illustrate the utility of FR3AK-seq to characterize T cell responses within

biopsied muscle tissue from a cohort of 145 patients with idiopathic inflammatory myopathies and controls.

Results:

Minimal primer sets targeting the TCRB framework region 3 (FR3) efficiently amplify CDR3 sequences

We hypothesized that minimizing the number of potentially competing V-gene primers, while also minimizing amplicon length, would maximize the efficiency of TCRB CDR3 region amplification. To identify candidate primer binding sequences upstream and proximal to CDR3, we constructed a phylogenetic tree of all functional TCRB FR3 regions available from the Immunogenetics (IMGT/GENE-DB) reference directory.[30] This revealed a high degree of sequence conservation within the 3' terminus of FR3 (**Fig 1b**). This homology may reflect a critical role for these sequences in the optimal presentation of the CDR3 loop. The 3' 20 nucleotides of these IMGT FR3s were therefore extracted and used as target primer binding sequences. Remarkably, the 128 functional TCRBV alleles reported by IMGT are represented by only 47 distinct 3'-terminal 20-mers. Beyond their sequence homology, additional advantages of targeting these sequences include minimal CDR3 amplicon length (and in turn optimal amplification efficiency) and the ability to sequence using short (100 nucleotide) single end reads initiated proximal and upstream of CDR3. We refer to this approach to TCRB CDR3 analysis as FR3 AmplifiKation sequencing, or "FR3AK-seq".

We developed a greedy algorithm to automate the design of three candidate sets of primers, each targeting the 47 distinct TCRB FR3 3' 20-mers (**Fig S1**): one primer set lacks any universal bases or mismatches ("Full" set), one primer set lacks universal bases but allows a single mismatch to occur ("1MM" set), and one primer set contains parsimonious incorporation of up to 3 universal

89 (inosine) bases and allows one mismatch ("Compact" set). No mismatches were permitted within
90 the five nucleotides of the 3' end of the primers, as a precaution to minimize interference with
91 polymerase extension.[31–33] The Full set consists of 47 primers (equal to the number of unique
92 FR3 3' 20-mers), the 1MM set consists of 34 primers, and the Compact set consists of 20 primers
93 (**Table S1**). We calculated the expected efficiency of each primer to amplify each FR3 sequence;
94 these efficiencies are visualized in the form of a clustered heatmap (**Fig 1c**). Importantly, each of
95 the 47 unique 3' FR3 20 mers are predicted to be amplified efficiently by at least one primer in
96 each set.

97

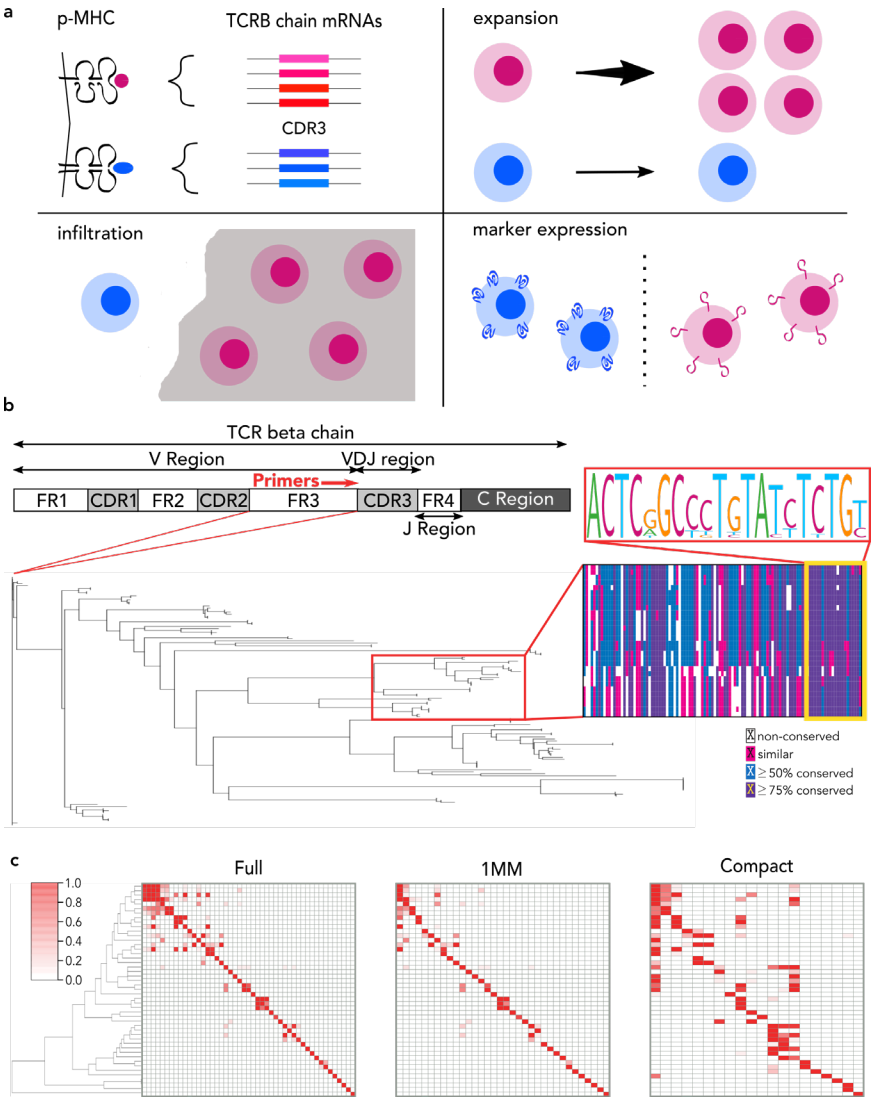


Figure 1: Utility of T cell receptor beta (TCRB) chain sequencing and FR3AK-seq multiplex PCR primer design. **a.** TCRB chain sequencing can be used to (i) identify related CDR3s that may share antigen specificity, (ii) detect expanded clones, (iii) identify tissue infiltrating clones, and (iv) link CDR3 sequences with T cell phenotypes. **b**. A schematic of the TCRB chain. A neighbor joining phylogenetic tree of FR3s extracted from the 128 functional human TCRBV alleles in the IMGT/GENE-DB reference directory and multiple sequence alignment from a dominant FR3 sequence cluster (red box on tree) identifies homology within the 3' ~20 nucleotides (yellow box). **c.** Clustered heatmaps for the Full, 1MM, and Compact primer sets showing primer amplification efficiencies for each primer across each of the 47 unique 3' 20 nucleotides of the human TCRBV FR3 region. Inosines were considered exact matches for the Compact set. Y axis: 47 unique TCRBV FR3s. X axis: primers. The order of the primers in each heatmap can be found in **Table S1**. The order of unique TCRBV FR3s is identical to the order of the Full primers for all heatmaps.

111 **TCRB FR3 primers amplify CDR3 sequences with minimal bias**

112 Each primer was tested separately and within its set for the ability to produce amplicon from
113 peripheral blood mononuclear cell (PBMC) cDNA. Agarose gel electrophoresis revealed a PCR
114 product at the expected size of ~220 base pairs (**Fig S2a**). Importantly, incorporation of the
115 inosine base did not prevent amplification by the KAPA Fast HotStart *Taq* DNA Polymerase. While
116 lower PCR primer annealing temperature might increase off-target priming, it may also reduce
117 unwanted bias in amplification from primers containing a single nucleotide mismatch.[34] We
118 therefore utilized gradient PCR to assess the effect of annealing temperature on the number of
119 unique CDR3 sequences detected using the Compact primer set. As expected, more unique
120 CDR3 clones were recovered at lower annealing temperatures (**Fig S2b**), and a similar trend was
121 observed for the 1MM primer set. The annealing temperature that produced the maximum number
122 of unique clones was identified for each set as follows: Full: 52°C, 1MM: 47°C, Compact: 42°C.
123 Use of a hemi-nested RT-PCR strategy (**Fig S3**) resulted in no apparent amplification of non-
124 TCRB sequences, even at the lowest PCR annealing temperature.

125 We next quantified PCR amplification bias inherent to the three primer sets. cDNA first underwent
126 20 cycles of PCR. An aliquot of this product was diluted 2^{10}-fold and then subjected to 10 more
127 cycles of PCR; sequencing was then performed on both amplicons and the resulting data sets
128 compared to each other. In this experiment, significant per-cycle amplification bias would manifest
129 as discordance. However, the concordance remains high for all three primer sets, suggesting that
130 FR3AK-seq primers negligibly bias the CDR3 repertoire during cycles of PCR amplification (ρ =
131 0.997 for the Full primer set and ρ = 0.998 for the 1MM primer set, **Fig 2a**; ρ = 0.997 for the
132 Compact primer set, **Fig S2c**).

133 We then performed rarefaction analysis to determine the relationship between RNA input amount,
134 sequencing depth, and the number of unique clones detected using FR3AK-seq. We found that
135 diversity discovery is saturated at ~1×10^6 reads for 100 ng of purified peripheral blood T cell RNA

136 (**Fig 2b**), providing the rationale for RNA input amounts and sequencing depths used in
137 subsequent analyses. We also assessed the sensitivity of FR3AK-seq to detect rare clones by
138 spiking varying amounts of monoclonal Jurkat RNA (from 1 up to 10,000 cellular equivalents) into
139 a background of 400 ng PBMC RNA (~20% T cells). We found that FR3AK-seq detected the
140 Jurkat CDR3 sequence at all input amounts at the expected abundances, while the abundance
141 of the top three PBMC T cell clones remained constant across the dilution series (**Fig 2c**). The
142 Jurkat sequence abundance was used to calculate cellular equivalents for the top three PBMC
143 clones in each sample. Notably, the coefficient of variation for these cellular equivalents was
144 below 1% in the 1,000 Jurkat cell equivalent spike-in samples (CV = 0.897%). This provided the
145 basis for the subsequent use of 1,000 cellular equivalents of spike-in cells.

146

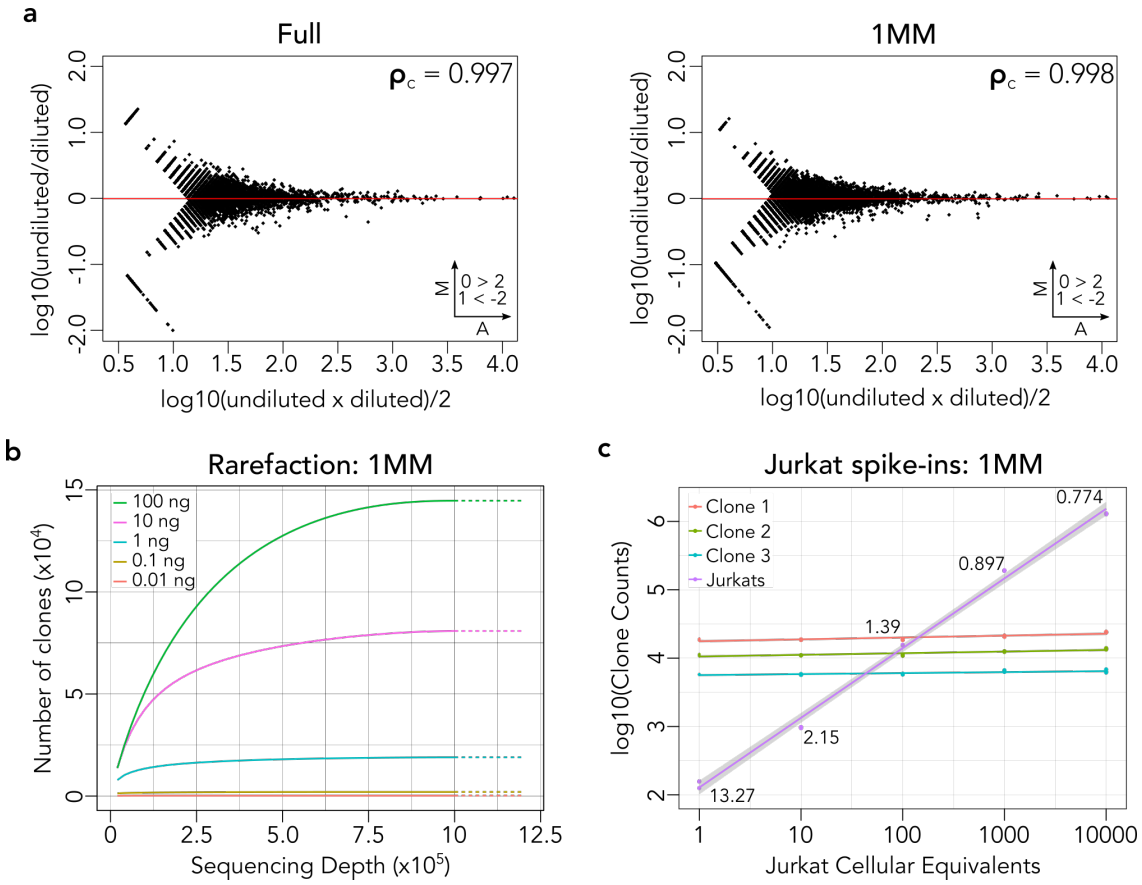


Figure 2: FR3AK-seq multiplex PCR performance. **a.** MA plots of 2^{10} dilution experiments for the Full and 1MM primer sets demonstrated high linear concordance (Full: 0.997, 1MM: 0.998, Compact in **Fig S2c**), indicating negligible PCR bias. Clones with ≥10 counts in either dataset were compared. Insets indicate number of clones above or below Y axis limits (set to 2 and –2 for visualization). **b.** Rarefaction analysis using T cell RNA purified from PBMCs showed the relationship between RNA input, number of clones detected, and sequencing depth using FR3AK-seq. **c.** The spiked-in Jurkat CDR3 sequence was detected at the expected abundances using FR3AK-seq (performed in triplicate). Estimated abundances of the top three T cell clones were consistent among triplicates and across spiked-in samples. 95% confidence intervals for regression lines and coefficients of variation for the cellular equivalents calculated for the T cell clones are shown.

162 **FR3AK-seq performs comparably to a multiplex PCR-based industry standard**

163 T cell RNA from two donors, A and B, as well as a third sample comprised of 90% Donor A RNA
164 and 10% Donor B RNA (sample "C"), were used to generate cDNA for benchmarking studies
165 against the current multiplex PCR-based industry standard, the immunoSEQ hsTCRB "Deep
166 Resolution" sequencing service offered by Adaptive Biotechnologies. Separate cDNA aliquots
167 were also subjected to FR3AK-seq analysis using each of the three primer sets. The open source
168 software MiXCR v2.1.11 was used to define and quantify CDR3 sequences from both assays for
169 direct comparison.[35] The technical reproducibility of the immunoSEQ assay was quantified by
170 comparing the abundances of Donor A's unique CDR3 sequences against their corresponding
171 abundances in sample C (concordance correlation coefficient, ρ = 0.989, left panel **Fig 3a**). The
172 same analysis was applied to the FR3AK-seq data sets, which were generated using each of the
173 three FR3 primer sets. Equally high measures of internal concordance were observed (ρ = 0.990
174 for the 1MM primer set, right panel **Fig 3a;** ρ = 0.990 for the Full primer set and ρ = 0.987 for the
175 Compact primer set, **Fig S4a**). Importantly, the concordance between immunoSEQ and FR3AK-
176 seq measurements of CDR3 abundance was also very high (ρ = 0.975, **Fig 3b**, **Fig S4b**),
177 providing further evidence that FR3AK-seq amplifies CDR3s with negligible bias. The majority of
178 even moderately abundant immunoSEQ-defined CDR3 sequences were also detected using any
179 of the three FR3 primer sets (**Fig 3c**). Additionally, the 9:1 mixture composed of cells from Donor
180 A and Donor B (sample C) was used to determine quantification accuracy of the FR3 primer sets
181 relative to the immunoSEQ assay (**Fig 3d**, **Fig S4c**). These data indicate that all three FR3 primer
182 sets provide CDR3 quantification comparable to the industry standard immunoSEQ assay, with
183 the 1MM primer set performing most accurately.

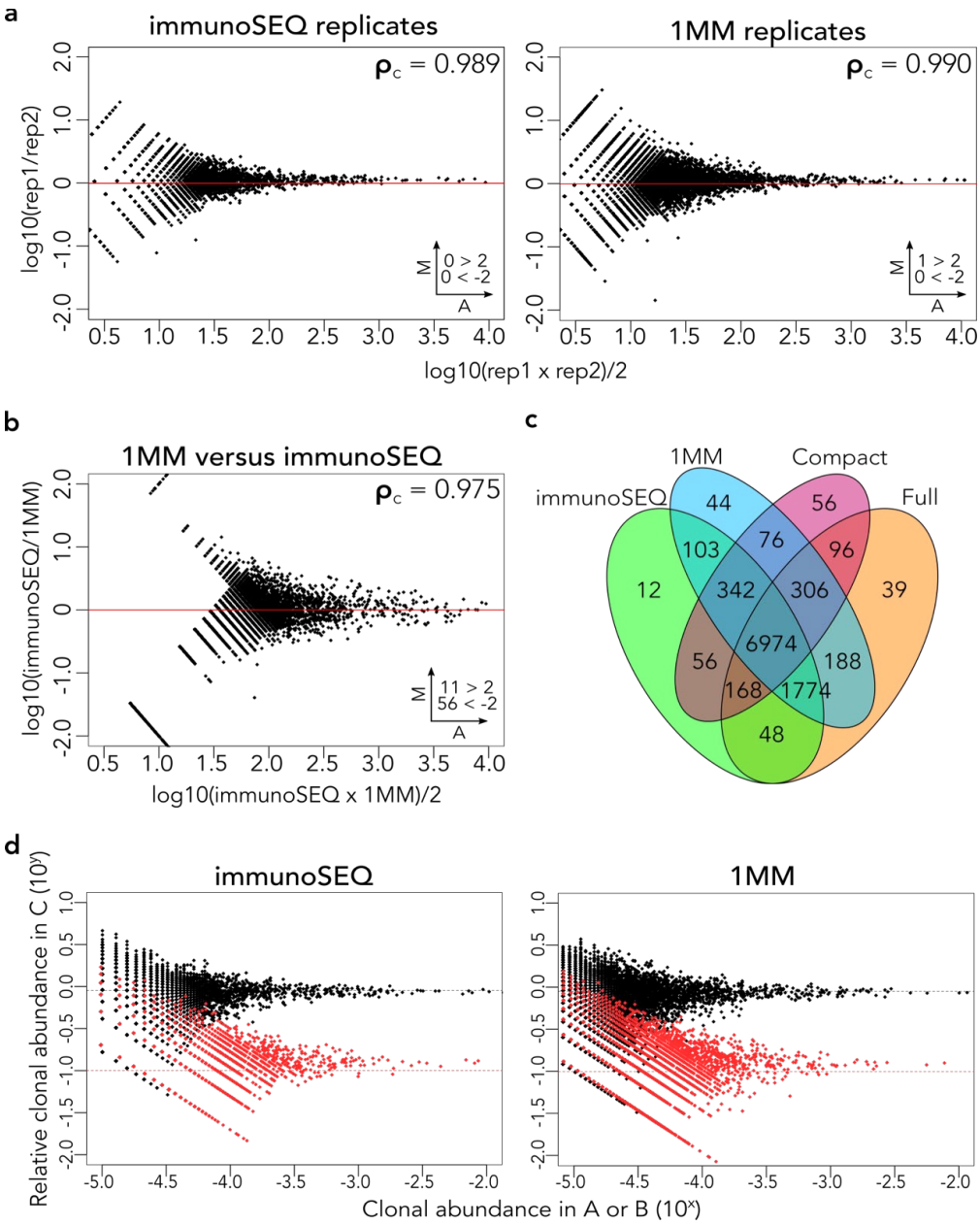


Figure 3: FR3AK-seq performs comparably to the multiplex PCR-based immunoSEQ assay from Adaptive Biotechnologies. **a.** MA plots comparing technical replicates of the immunoSEQ and 1MM FR3AK-seq assays. **b.** MA plot comparing the 1MM FR3AK-seq assay against the immunoSEQ assay. Clones with ≥10 counts in either dataset were included. **c.** A Venn diagram shows overlapping and non-overlapping CDR3 sequences detected using immunoSEQ and all three FR3AK-seq primer sets. CDR3 sequences were included in this analysis if they had a clone count ≥10 in any of the four data sets. **d.** Relative frequencies of Donor A (black points) and Donor B (red points) in sample C determined using the immunoSEQ assay or the 1MM FR3AK-seq assay. Insets in **a** and **b** indicate number of clones above or below Y axis limits (set to 2 and -2 for visualization).

194 **FR3AK-seq performs comparably to a unique molecular identifier (UMI)-based industry**
195 **standard**

196 To assess whether FR3AK-seq achieves results comparable to UMI-based methodologies, we
197 compared the FR3AK-seq 1MM and immunoSEQ datasets to one obtained using ArcherDX's
198 UMI-based Immunoverse HS TCR assay. PBMC RNA from the same samples A, B, and C
199 described above were subjected to Immunoverse analysis for comparison with the FR3AK-seq
200 and immunoSEQ datasets. The technical reproducibility of the Immunoverse assay was similar
201 to that of both immunoSEQ and FR3AK-seq (ρ = 0.967, **Fig 4a**). Additionally, concordance
202 between FR3AK-seq and Immunoverse was high (ρ = 0.900, upper panel **Fig 4b**). Relatively high
203 concordance was also observed when comparing immunoSEQ to Immunoverse (ρ = 0.886, lower
204 panel **Fig 4b**). Crucially, the majority of clones detected by the Immunoverse assay were also
205 detected by both immunoSEQ and FR3AK-seq (**Fig 4c**). These data demonstrate that FR3AK-
206 seq has a high sensitivity for detecting clones of the correct sequence (4352/4403, 98.8% of total
207 clones detected by Immunoverse). Strikingly, a large number (2803) of clones were detected by
208 both immunoSEQ and FR3AK-seq but not Immunoverse. Rarefaction analysis of the FR3AK-seq
209 and Immunoverse datasets explained this difference in clonal detection sensitivity, in that FR3AK-
210 seq sampled more clones at all sequencing depths (**Fig 4d**). Although FR3AK-seq achieves better
211 coverage of the repertoire than Immunoverse across sequencing depths, we wanted to quantify
212 how many of the FR3AK-seq detected clones may be the result of false discovery. To address
213 this question, we performed FR3AK-seq 1MM analysis of 1,000 cellular equivalents of Jurkat cell
214 RNA in triplicate. Since the Jurkat cell line is monoclonal, the frequencies of subdominant clones,
215 presumably generated by PCR or sequencing errors, can be used to assess the rate of false
216 positive discovery. 99.98% of MiXCR-defined CDR3s mapped to the correct reported Jurkat
217 TCRB sequence in each sample. We found an average of 50 non-Jurkat clones across triplicates,
218 of which only one was moderately abundant (≥10 clone counts) in each. This secondary clone

219 was the same between replicates and represented on average 0.0067% of total MiXCR-defined
220 CDR3s. Using default MiXCR settings, FR3AK-seq analysis is therefore associated with a very
221 low rate of false discovery due to PCR or sequencing errors. Modification of MiXCR thresholds
222 can be used to adjust the stringency with which these subdominant clones are grouped. We also
223 compared the CDR3 lengths of clones detected using each sequencing platform, and found that
224 FR3AK-seq's short (100 nucleotide) reads did not bias the distribution of CDR3 lengths recovered
225 (**Fig 4e**).

226

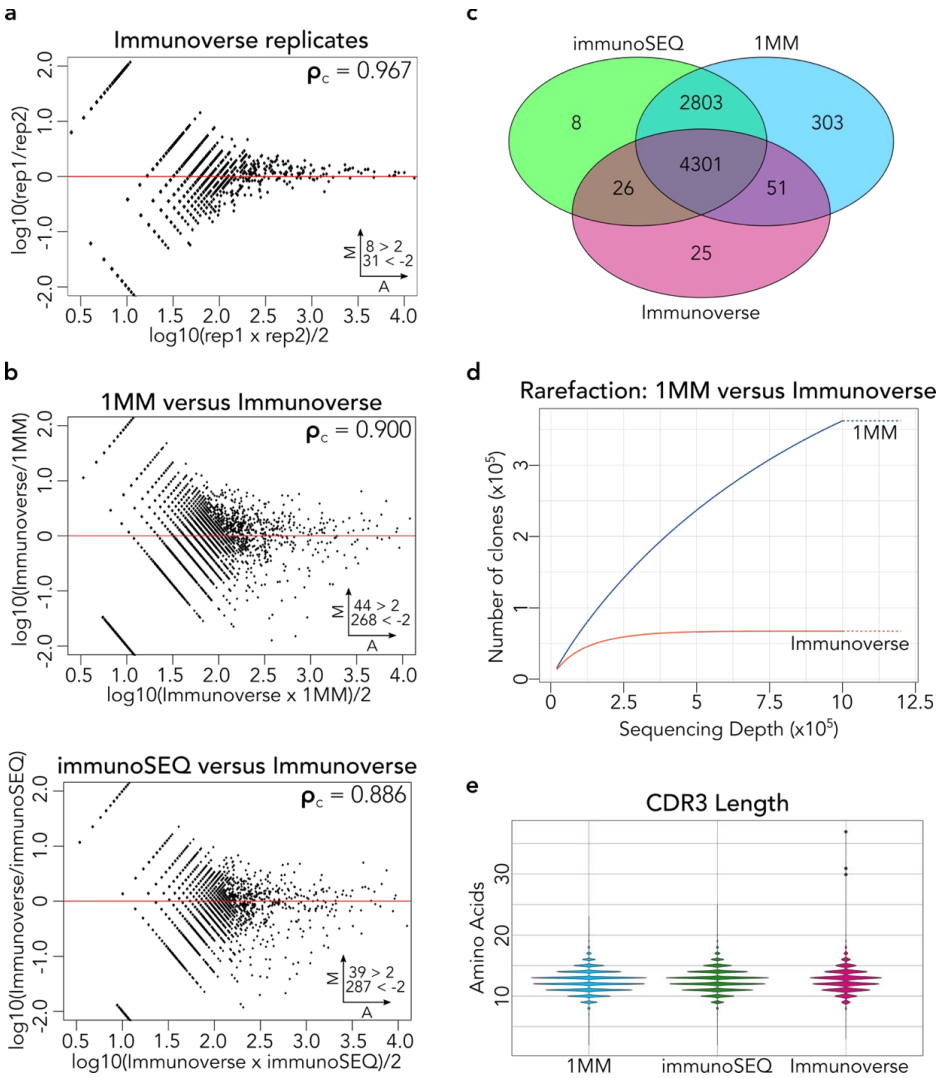


Figure 4: FR3AK-seq performs comparably to the UMI-based ArcherDX Immunoverse assay. **a.** MA plot comparing Immunoverse technical replicates. **b.** MA plots comparing the 1MM FR3AK-seq assay against the Immunoverse assay (top panel) and Adaptive Biotechnologies' immunoSEQ assay against the Immunoverse assay (bottom panel). Clones with ≥10 counts in either dataset were included (using pre-deduplicated counts for the Immunoverse dataset). **c.** A Venn diagram shows overlapping and non-overlapping CDR3 sequences detected using Immunoverse, immunoSEQ, and the 1MM FR3AK-seq primer set. CDR3 sequences were included in this analysis if they had a clone count of ≥10 in any of the three data sets. **d.** Rarefaction analysis comparing the 1MM FR3AK-seq assay to the Immunoverse assay using 100 ng of T cell RNA (~250,000 cells) for 1MM FR3AK-seq analysis and 200 ng of T cell RNA (~500,000 cells) for Immunoverse analysis. 1MM: ~1.2×10^6 reads, Immunoverse: ~1.4×10^6 reads. **e.** Violin plots comparing CDR3 lengths captured by the short-read 1MM FR3AK-seq versus longer-read immunoSEQ and Immunoverse assays. Insets in **a** and **b** indicate number of clones above or below Y axis limits (set to 2 and -2 for visualization).

241 **FR3AK-seq enables inexpensive cohort-scale repertoire studies**

242 Muscle biopsies from 145 patients with idiopathic inflammatory myopathies (IIMs, 124) as well as
243 healthy controls (9) and non-IIM controls (12) were analyzed using the 1MM primer set at a total
244 cost of ~$20 per sample (**Table S2**). IIM patients included those with dermatomyositis (DM, 40),
245 immune-mediated necrotizing myopathy (IMNM, 49), inclusion body myositis (IBM, 14), and anti-
246 synthetase syndrome (ASyS, 21) (**Table S3**). For absolute T cell number estimation, 1,000 cell
247 equivalents of RNA derived from a clonal tumor infiltrating T cell line[36] were spiked into each
248 sample after RNA purification. Aggregated CD4 and CD8 sequencing reads from RNA-seq
249 analysis of the same samples (performed separately and reported elsewhere) tightly correlated
250 with FR3AK-seq based estimates of cellular equivalents (**Fig 5a**). In subsequent analyses, CDR3
251 sequences present at levels at or below one cell equivalent per biopsy were considered
252 'bystander' T cell clones, unlikely to be involved in the disease process. By this metric, the number
253 of non-bystander T cells were elevated in each disease subgroup versus the controls (**Fig 5b**). In
254 agreement with previous immunohistochemical analyses, IBM biopsies contained particularly
255 high levels of T cell infiltrates.[37–39] We determined this to be the case both in terms of absolute T
256 cell number and clonal diversity (**Fig 5b, Fig S5a**). No distinct patterns of J gene usage were
257 found among patient subgroups (**Fig S6a**), and CDR3 hydrophobicity and lengths were similar
258 between patient subgroups (**Fig S6b-c**).

259 The open source software GLIPH (grouping of lymphocyte interactions by paratope hotspots)[10]
260 was next used to search for clusters of related CDR3 sequences within the aggregated set of
261 non-bystander CDR3s (**Table S4**). 143 clusters composed of sequences contributed by at least
262 3 individuals were identified. We first examined clusters that had a disproportionate number of
263 sequences contributed by single individuals. **Fig 5c** provides an example of a cluster dominated
264 by related CDR3 sequences contributed by a single ASyS patient. Compared to CDR3 sequences
265 from the same patient which were not in any cluster, the clonal abundances of these CDR3s were

266 significantly increased, suggestive of a polyclonal, shared antigen-driven expansion. These
267 results suggest that GLIPH is able to cluster related CDR3 sequences that have been detected
268 and quantified using FR3AK-seq.

269 We next determined whether any of the GLIPH clusters were composed of sequences contributed
270 disproportionately by individuals from one of the IIM subgroups. Indeed, we found a GLIPH cluster
271 that was composed exclusively of sequences contributed by IBM patients. Specifically, 5 of the
272 12 IBM patients with non-bystander T cells contributed at least one sequence to this cluster ($p =$
273 1.4×10^{-6}, **Fig 5d**). Remarkably, all CDR3 sequences in this IBM cluster are linked to a single J-
274 chain (TRBJ2-2*01), providing additional support for shared antigen recognition**.** Taken together,
275 this study demonstrates the utility of FR3AK-seq in probing T cell responses at cohort scale.

276

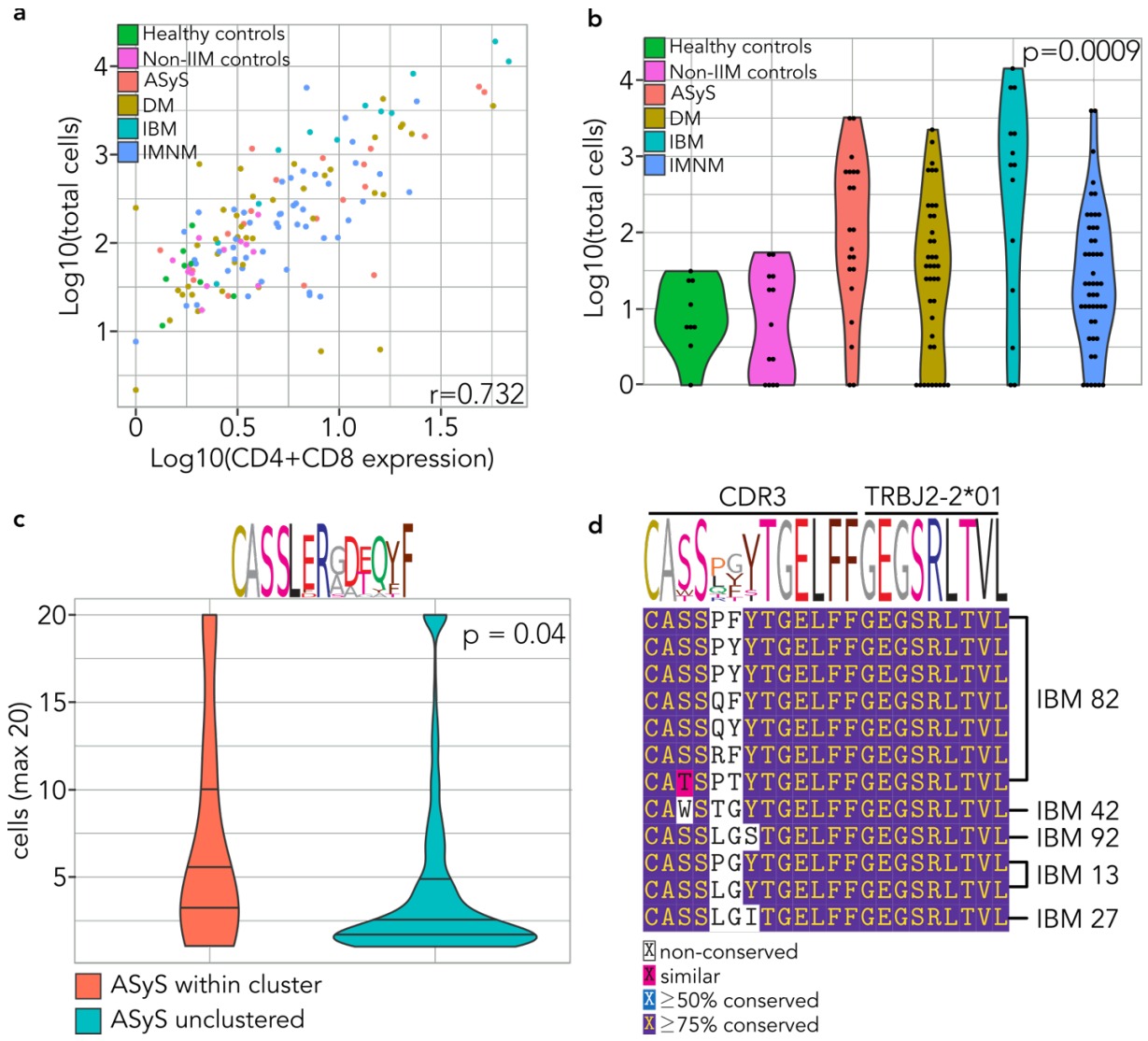


Figure 5: FR3AK-seq detects related T cell responses in subsets of patients with idiopathic inflammatory myopathies. **a.** Scatterplot of estimated T cell counts per patient versus CD4 and CD8 expression levels obtained from bulk mRNA-seq. Each patient data point is colored according to their IIM association. Pearson correlation coefficient (r) is shown. **b.** Distributions of cell counts per patient for each IIM and controls. Kruskal-Wallis p-value is provided. **c.** A CDR3 cluster enriched for sequences from a single patient with ASyS was identified using GLIPH software ($p = 1.4x10^{-7}$). The amino acid sequence logo of the CDR3 region is shown above the violin plot. The clones within this cluster are more expanded than corresponding unclustered clones from the same patient. Mann-Whitney p = 0.04. Cell counts were set to a maximum of 20 for visualization. **d.** An IBM-exclusive cluster was also identified using GLIPH, encompassing 5 of 12 IBM patients who had non-bystander T cell infiltrates (41.7%, $p=1.4x10^{-6}$). Multiple sequence alignment (using MUSCLE) is shown for the corresponding MiXCR-defined CDR3s and the J chain allele (TRBJ2-2*01).

291 **Discussion**

292 Although the utility of TCR repertoire sequencing to characterize T cell responses has been well
293 established, the lack of a streamlined, quantitative, inexpensive, and non-proprietary assay has
294 limited the scope and scale of feasible studies. By rationally designing minimal sets of primers
295 that target the 3' terminus of FR3, we have developed a multiplex PCR-based approach for ultra-
296 efficient library preparation and sequencing of TCRB CDR3 repertoires. By minimizing
297 amplification bias (via reduction of primer number and amplicon length), the resulting sequencing
298 libraries quantitatively capture clonal abundance distributions with an accuracy comparable to
299 both multiplex PCR-based and UMI-based industry standards. While the library preparation and
300 sequencing strategy presented here already bring the per-sample cost to ~$20 (**Table S2**), we
301 expect that alternative approaches to library preparation, reduced sequencing depth, and
302 declining sequencing costs will likely enable another ~10-fold reduction in per-sample cost. In
303 addition to reducing the overall cost of obtaining CDR3 sequences, we have developed the
304 FR3AK-seq workflow to require minimal expertise, effort, and time. Quantitative sequencing
305 libraries can easily be prepared within a single day. Comparisons of protocols and cost between
306 FR3AK-seq, immunoSEQ, and Immunoverse are summarized in **Table S5**.

307 Our approach to TCRB repertoire analysis should be generalizable to additional immune receptor
308 types, as well as to those of non-human species (see **Table S6** for *Mus musculus* TCRB primers).
309 Although these mouse primers have not been as extensively validated as the human set, they
310 have been confirmed to efficiently amplify CDR3 sequences from mouse tissues. We have
311 additionally used the same FR3 primer design and analysis principles to target human TCR alpha,
312 gamma, and delta repertoires, although these primers have not yet been validated (**Table S7**).
313 We have also designed a set of TCRB J region reverse primers (**Table S8**) that perform
314 comparably to our constant region reverse primer using cDNA templates (ρ = 0.949, **Fig S7**).
315 These will be potentially useful for analysis of FFPE samples which suffer from extensive RNA

316 degradation. Interestingly, during the preparation of this manuscript, the Euroclonality Group
317 (creators of the BIOMED2 primer set[40]) published a new set of TCR primers for next-generation
318 sequencing, including some that are designed in the 5' portion of FR3,[41] providing additional
319 validation for an FR3-targeted approach to TCR repertoire sequencing.

320 V gene usage is often of interest, for example to detect usage skewing and for functional studies
321 involving TCR cloning. FR3AK-seq prioritizes small amplicons to reduce both PCR bias and cost.
322 Therefore, by design, amplicons do not contain easily recoverable V gene sequence information.
323 However, it is theoretically possible to make inferences on each CDR3's associated V genes,
324 based on the pattern of FR3 primer usage. We expect these inferences to be more accurate for
325 more abundant clones. When analysis of V gene usage is desired, complementary techniques
326 (e.g. 5'RACE) can be parsimoniously utilized to capture this information. One may wish to use
327 FR3AK-seq to track unique CDR3 sequences of interest over time, across tissues, and/or after
328 FACS analysis – these FR3AK-seq detected CDR3 sequences can then be readily associated
329 with a V gene by merging data sets.

330 To demonstrate the use of FR3AK-seq for efficient cohort scale analysis of TCRB CDR3
331 repertoires, we characterized the muscle-infiltrating T cells present within muscle biopsies
332 obtained from 145 inflammatory muscle disease patients and controls. TCRB CDR3 sequence
333 clustering indicated both donor and disease specific antigen-driven T cell responses. Importantly,
334 the majority of T cell clones detected in IBM muscle biopsies were not hyperexpanded, in contrast
335 to the hypothesis that infiltrating IBM T cells originate from a clonal T cell large granular
336 lymphocytic leukemia (TLGLL) (**Fig S5b**).[42,43] Future studies will utilize FR3AK-seq in combination
337 with *in vitro* stimulation to characterize the antigenic determinants of these clonal clusters.

338

339 **Methods:**

340 **FR3 primer design**

341 FR3 sequences from all functional human TCRB V alleles were downloaded from the IMGT/LIGM-
342 DB reference directory (non-functional and pseudogenes were excluded).[30] Multiple sequence
343 alignment (MSA) was performed using MUSCLE and UPGMA (Unweighted Pair Group Method
344 with Arithmetic Mean) and neighbor-joining phylogenetic trees were generated via the R package
345 msa.[44] The 3' 20 nucleotides of the TCRB chain FR3s were subsequently used to design three
346 sets of primers as described in the text. We automated design of the 1MM and the Compact
347 primer sets with an algorithm outlined in **Fig S1**, available on Github
348 (https://github.com/jmmontagne/FR3AK-seq). Human TCRB primer sequences can be found in
349 **Table S1**.

350 **PBMC preparation, T cell purification, RNA extraction, and cDNA synthesis**

351 PBMCs from Donors A and B were isolated by Ficoll-Paque (GE Healthcare, Pittsburgh, PA)
352 gradient centrifugation and cryopreserved. T cells were purified from thawed PBMCs using the
353 EasySep Human T cell enrichment kit (STEMCELL Technologies, Cambridge, MA). RNA was
354 extracted from purified human T cells using the RNeasy Plus Minikit (Qiagen, Germantown, MD).
355 0.5 μg Donor B RNA was mixed with 4.5 μg Donor A RNA to create sample C for quantitation
356 experiments. 4 μg each of Donor A, Donor B, and sample C RNA was reverse transcribed using
357 a TCRB chain constant region reverse primer with Superscript III First-Strand Synthesis System
358 (Invitrogen, Waltham, MA). cDNA was column purified with the Oligo Clean and Concentrator Kit
359 (Zymo Research, Irvine, CA). 100 ng of purified cDNA from each sample was sent to Adaptive
360 Biotechnologies (Seattle, WA) for the immunoSEQ hsTCRb Deep sequencing service or used for
361 FR3AK-seq PCR. For the ArcherDX (Boulder, CO) Immunoverse analysis, RNA was extracted
362 from matched Donor A and B PBMCs and mixed at the same 9:1 ratio to create sample C. 800

363 ng of RNA from each sample was input into the Immunoverse assay. The concordance between
364 FR3AK-seq and Immunoverse remained high despite the use of PBMC RNA rather than T cell
365 RNA for the Immunoverse assay. All primer sequences are provided in **Table S1**.

366 **Polymerase Chain Reaction (PCR) and sequencing library preparation**

367 100 ng TCRB chain cDNA was used as template for PCR with KAPA2G Fast Multiplex Mix (Kapa
368 Biosystems, Wilmington, MA). Forward TCRB FR3 primers from the Full, 1MM, or Compact
369 primer sets were used with a single nested TCR beta chain constant region reverse primer at 0.2
370 μM per primer. Samples underwent 30 cycles of PCR for visualization by agarose gel
371 electrophoresis, or 20 cycles for NGS library preparation ("PCR1"). We compared the
372 performance of Herculase II Fusion DNA polymerase (Agilent, Santa Clara, CA) and KAPA2G
373 Fast Multiplex Mix for PCR1 and found that both enzymes perform comparably ($\rho = 0.966$, **Fig**
374 **S8**). Herculase II is therefore an acceptable enzyme alternative for PCR1 if desired.

375 20 cycles of PCR2 were performed on PCR1 product (2 μL of PCR1 added to 18 μL of PCR2
376 master mix, which contained 0.25 μM each i5 and i7 sample barcoding primers) to incorporate
377 sample barcodes and Illumina sequencing adaptors using Herculase II Fusion DNA Polymerase.
378 Equal volumes of barcoded PCR2 products (5 μL each) were pooled and PCR column purified
379 using QIAquick PCR Purification Kit (Qiagen). Libraries were quantified using KAPA Library
380 Quantification Kit for Illumina Platforms (Kapa Biosystems). Barcoding primer sequences can be
381 found in **Supplementary File 1**. See **Fig S3** for a schematic of our sequencing strategy.

382 **Sequencing and CDR3 analysis**

383 Sequencing was performed on an Illumina NextSeq 500 (immunoSEQ benchmarking
384 comparisons), MiSeq (Immunoverse samples), or HiSeq 2500 (IBM muscle biopsy analysis).
385 CDR3s were identified and quantified using MiXCR v2.1.11 software[35] with default parameters
386 except as noted in **Table S9.** Data obtained from Adaptive Biotechnologies' immunoSEQ assay

387 were re-analyzed using MiXCR v2.1.11 and identical parameters for comparison to our own data.
388 For reanalysis, full nucleotide sequences for each CDR3 from the Adaptive Biotechnologies
389 ImmunoSEQ dataset were expanded to repeat as many times as indicated by the corresponding
390 "count" column. This file, with each clonal nucleotide sequence represented as many times as its
391 "count" column, was converted to FASTA format for compatibility with MiXCR. The sequences
392 obtained from ArcherDX's Immunoverse assay were expanded to repeat as many times as their
393 pre-deduplicated clone counts and then matched back to their deduplicated clone counts after
394 reanalysis with MiXCR v2.1.11. Clone count cutoffs for each analysis are as described in the
395 figure legends and refer to the pre-deduplicated counts for the Immunoverse assay (Note:
396 deduplicated clone counts were used for all comparisons, while pre-deduplicated counts were
397 only used to filter the dataset to ≥ 10 counts). Information regarding the total number of clones,
398 the number of singletons, and the number of clones occurring at ≥10 counts for FR3AK-seq,
399 immunoSEQ, and Immunoverse can be found in **Table S10**.

400 **Repertoire analysis using muscle biopsies from patients with idiopathic inflammatory**
401 **myopathies (IIMs)**

402 RNA isolated from muscle biopsies taken from 145 IIM patients or controls was reverse
403 transcribed, PCR amplified, sequenced, and analyzed as described above. Healthy control
404 muscle biopsies were obtained from healthy volunteers at the Skeletal Muscle Biobank of the
405 University of Kentucky, while non-IIM controls were obtained for clinical purposes at the Johns
406 Hopkins Neuromuscular Pathology Laboratory and were normal after pathologic evaluation. We
407 added 1,000 cell equivalents of RNA from a clonal tumor infiltrating T cell line as a spike-in to
408 each sample.[36] Sequencing of 6 spike-in only samples provided the CDR3 sequences to remove
409 from downstream analyses of samples containing muscle biopsy RNA. The spike-in counts were
410 used for cell number quantification within patient samples, which distinguished 'bystander' versus

411 ‘non-bystander’ T cell clones. One non-IIM control muscle biopsy was sequenced in duplicate,
412 making the total number of samples 146.

413 CDR3 sequences obtained from these samples were analyzed for disease-specific clusters using
414 the GLIPH 1.0 group discovery algorithm (see **Table S9** for parameters)**.**[10] To determine the
415 statistical significance of each cluster, we performed chi-square analysis on all clusters with at
416 least three contributing individuals, followed by Benjamini-Hochberg multiple comparison
417 correction. This analysis was performed for both the number of sequences contributed by each
418 disease subgroup to each cluster, as well as the number of patients in each disease group
419 represented within each cluster. P-values ≤0.05 after multiple test correction were considered
420 statistically significant.

421

422 **Author Contributions:**

423 Conceptualization and experimental design: J.M.M., H.B.L.; performing experiments: J.M.M.; data
424 analysis and software: J.M.M., X.A.Z.; human specimen procurement and analysis: I.P-F., J.C.M.,
425 L.C-S., T.E.L., A.L.M., H.B.L.; data interpretations: J.M.M., X.A.Z., T.E.L., A.L.M., H.B.L.; paper
426 writing: J.M.M., H.B.L. All authors reviewed the paper.

427

428 **Competing Interests:**

429 H. Benjamin Larman is a consultant for Tscan Therapeutics, which seeks to develop T cell
430 receptor based cellular therapies.

431

432

433 **Acknowledgements**:

434 The authors would like to thank Dr. Erik Wright for his advice on primer design, Dr. Suzanne L.
435 Topalian for the TIL1558 cell line, and Dr. Maximilian F. Konig for his review of the manuscript.
436 We would also like to thank the Johns Hopkins Medical Institute Deep Sequencing and Microarray
437 Core Facility and the Johns Hopkins Institute of Genetic Medicine Genetics Resources Core
438 Facility. This work was supported by NIH grant U24AI118633 and a Prostate Cancer Foundation
439 Young Investigator Award.

440

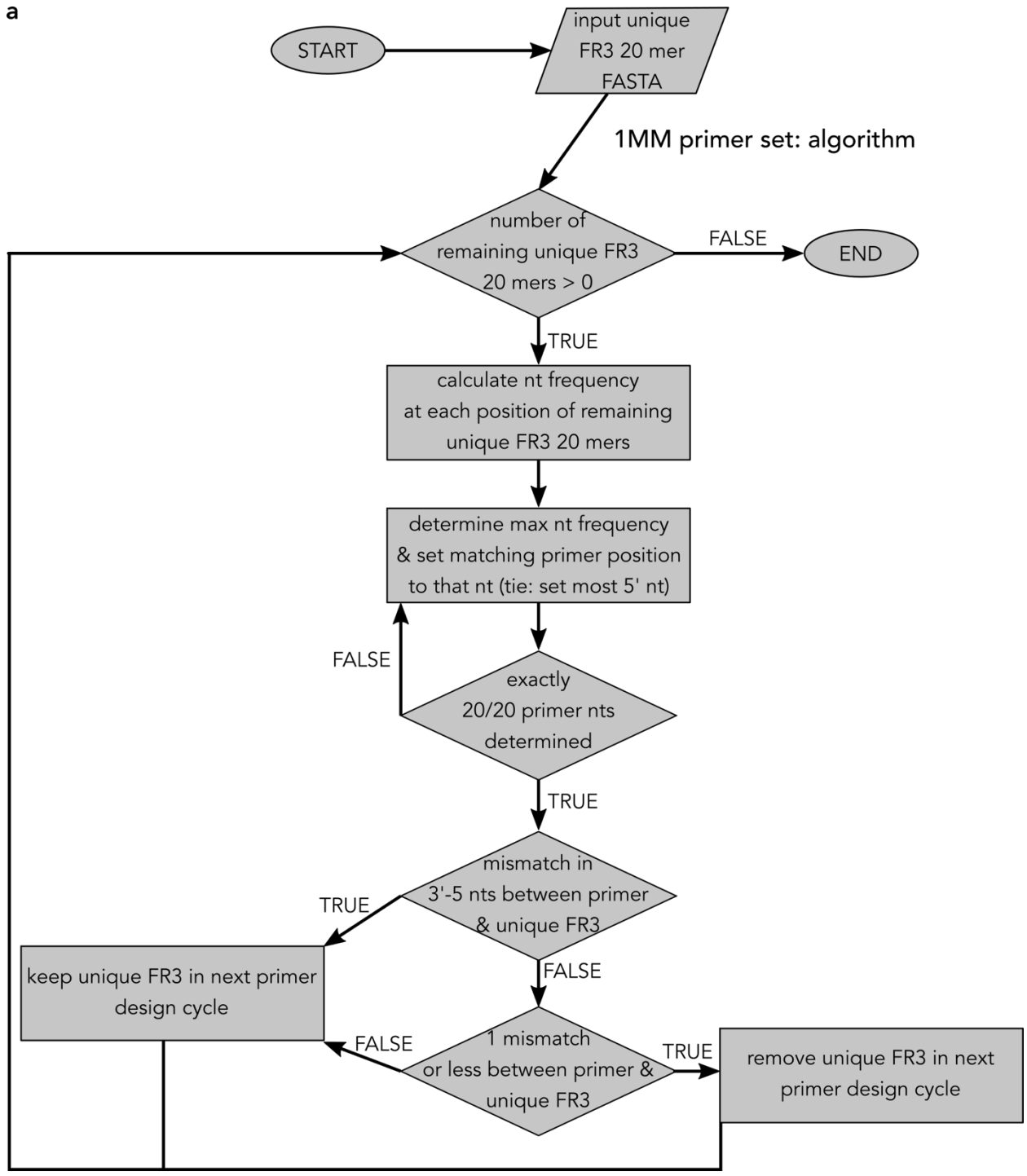
441 **References**:

442 1. Attaf, M., Huseby, E. & Sewell, A. K. αβ T cell receptors as predictors of health and
443 disease. *Cell. Mol. Immunol.* **12**, 391–399 (2015).
444 2. Freeman, J. D., Warren, R. L., Webb, J. R., Nelson, B. H. & Holt, R. A. Profiling the T-cell
445 receptor beta-chain repertoire by massively parallel sequencing. *Genome Res.* **19**, 1817–
446 1824 (2009).
447 3. Robins, H. S. *et al.* Comprehensive assessment of T-cell receptor beta-chain diversity in
448 alphabeta T cells. *Blood* **114**, 4099–4107 (2009).
449 4. Weinstein, J. A., Jiang, N., White, R. A., Fisher, D. S. & Quake, S. R. High-throughput
450 sequencing of the zebrafish antibody repertoire. *Science* **324**, 807–810 (2009).
451 5. Boyd, S. D. *et al.* Measurement and clinical monitoring of human lymphocyte clonality by
452 massively parallel VDJ pyrosequencing. *Sci. Transl. Med.* **1**, 12ra23 (2009).
453 6. Muraro, P. A. *et al.* T cell repertoire following autologous stem cell transplantation for
454 multiple sclerosis. *J. Clin. Invest.* **124**, 1168–1172 (2014).
455 7. Gros, A. *et al.* PD-1 identifies the patient-specific $CD8^+$ tumor-reactive repertoire infiltrating
456 human tumors. *J. Clin. Invest.* **124**, 2246–2259 (2014).
457 8. Beausang, J. F. *et al.* T cell receptor sequencing of early-stage breast cancer tumors
458 identifies altered clonal structure of the T cell repertoire. *Proc. Natl. Acad. Sci. U. S. A.* **114**,
459 E10409–E10417 (2017).
460 9. Emerson, R. O. *et al.* Immunosequencing identifies signatures of cytomegalovirus exposure
461 history and HLA-mediated effects on the T cell repertoire. *Nat. Genet.* **49**, 659–665 (2017).
462 10. Glanville, J. *et al.* Identifying specificity groups in the T cell receptor repertoire. *Nature* **547**,
463 94–98 (2017).
464 11. Dash, P. *et al.* Quantifiable predictive features define epitope-specific T cell receptor
465 repertoires. *Nature* **547**, 89–93 (2017).
466 12. DeWitt, W. S. *et al.* Human T cell receptor occurrence patterns encode immune history,
467 genetic background, and receptor specificity. *eLife* **7**, (2018).
468 13. Han, A., Glanville, J., Hansmann, L. & Davis, M. M. Linking T-cell receptor sequence to
469 functional phenotype at the single-cell level. *Nat. Biotechnol.* **32**, 684–692 (2014).
470 14. Morris, H. *et al.* Tracking donor-reactive T cells: Evidence for clonal deletion in tolerant
471 kidney transplant patients. *Sci. Transl. Med.* **7**, 272ra10 (2015).
472 15. Theil, A. *et al.* T cell receptor repertoires after adoptive transfer of expanded allogeneic
473 regulatory T cells. *Clin. Exp. Immunol.* **187**, 316–324 (2017).
474 16. Mamedov, I. Z. *et al.* Quantitative tracking of T cell clones after haematopoietic stem cell
475 transplantation. *EMBO Mol. Med.* **3**, 201–207 (2011).
476 17. Rosati, E. *et al.* Overview of methodologies for T-cell receptor repertoire analysis. *BMC*
477 *Biotechnol.* **17**, 61 (2017).
478 18. Rudolph, M. G., Stanfield, R. L. & Wilson, I. A. How TCRs bind MHCs, peptides, and
479 coreceptors. *Annu. Rev. Immunol.* **24**, 419–466 (2006).
480 19. Turner, S. J., Doherty, P. C., McCluskey, J. & Rossjohn, J. Structural determinants of T-cell
481 receptor bias in immunity. *Nat. Rev. Immunol.* **6**, 883–894 (2006).
482 20. Miles, J. J., Douek, D. C. & Price, D. A. Bias in the αβ T-cell repertoire: implications for
483 disease pathogenesis and vaccination. *Immunol. Cell Biol.* **89**, 375–387 (2011).
484 21. Kebschull, J. M. & Zador, A. M. Sources of PCR-induced distortions in high-throughput
485 sequencing data sets. *Nucleic Acids Res.* **43**, e143 (2015).
486 22. Carlson, C. S. *et al.* Using synthetic templates to design an unbiased multiplex PCR assay.
487 *Nat. Commun.* **4**, 2680 (2013).
488 23. Shugay, M. *et al.* Towards error-free profiling of immune repertoires. *Nat. Methods* **11**,
489 653–655 (2014).

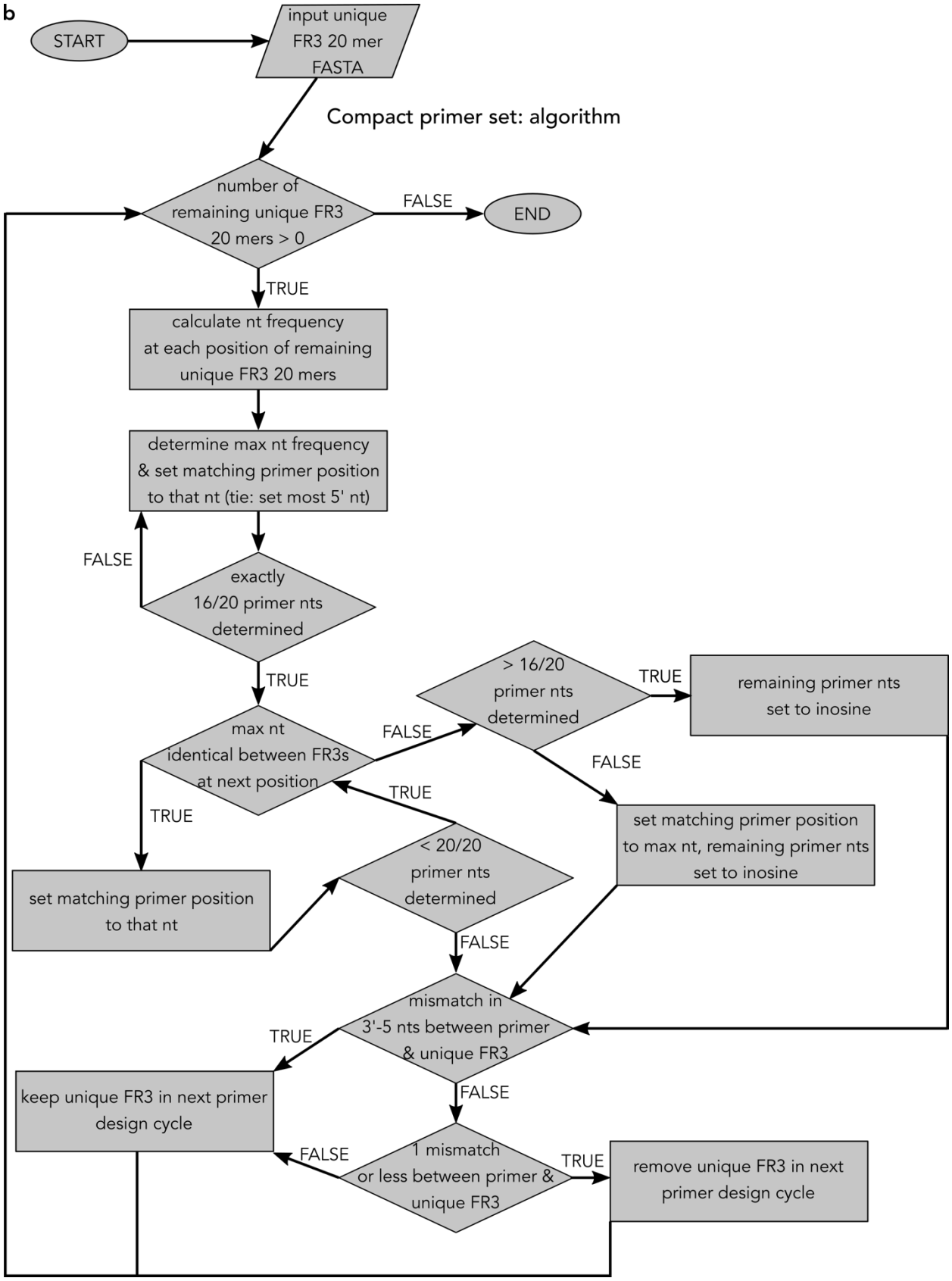
24. Peng, Q., Vijaya Satya, R., Lewis, M., Randad, P. & Wang, Y. Reducing amplification artifacts in high multiplex amplicon sequencing by using molecular barcodes. *BMC Genomics* **16**, 589 (2015).
25. Vander Heiden, J. A. *et al.* pRESTO: a toolkit for processing high-throughput sequencing raw reads of lymphocyte receptor repertoires. *Bioinformatics* **30**, 1930–1932 (2014).
26. Shugay, M. *et al.* VDJtools: Unifying Post-analysis of T Cell Receptor Repertoires. *PLoS Comput. Biol.* **11**, (2015).
27. Oakes, T. *et al.* Quantitative Characterization of the T Cell Receptor Repertoire of Naïve and Memory Subsets Using an Integrated Experimental and Computational Pipeline Which Is Robust, Economical, and Versatile. *Front. Immunol.* **8**, (2017).
28. Ma, K.-Y. *et al.* Immune Repertoire Sequencing Using Molecular Identifiers Enables Accurate Clonality Discovery and Clone Size Quantification. *Front. Immunol.* **9**, 33 (2018).
29. Egorov, E. S. *et al.* Quantitative profiling of immune repertoires for minor lymphocyte counts using unique molecular identifiers. *J. Immunol. Baltim. Md 1950* **194**, 6155–6163 (2015).
30. Giudicelli, V., Chaume, D. & Lefranc, M.-P. IMGT/GENE-DB: a comprehensive database for human and mouse immunoglobulin and T cell receptor genes. *Nucleic Acids Res.* **33**, D256-261 (2005).
31. Simsek, M. & Adnan, H. Effect of single mismatches at 3'-end of primers on polymerase chain reaction. *J. Sci. Res. Med. Sci.* **2**, 11–14 (2000).
32. Wu, J.-H., Hong, P.-Y. & Liu, W.-T. Quantitative effects of position and type of single mismatch on single base primer extension. *J. Microbiol. Methods* **77**, 267–275 (2009).
33. Wright, E. S. *et al.* Exploiting extension bias in polymerase chain reaction to improve primer specificity in ensembles of nearly identical DNA templates. *Environ. Microbiol.* **16**, 1354–1365 (2014).
34. Ishii, K. & Fukui, M. Optimization of annealing temperature to reduce bias caused by a primer mismatch in multitemplate PCR. *Appl. Environ. Microbiol.* **67**, 3753–3755 (2001).
35. Bolotin, D. A. *et al.* MiXCR: software for comprehensive adaptive immunity profiling. *Nat. Methods* **12**, 380–381 (2015).
36. Deng, L. *et al.* Structural basis for the recognition of mutant self by a tumor-specific, MHC class II-restricted T cell receptor. *Nat. Immunol.* **8**, 398–408 (2007).
37. Arahata, K. & Engel, A. G. Monoclonal antibody analysis of mononuclear cells in myopathies. I: Quantitation of subsets according to diagnosis and sites of accumulation and demonstration and counts of muscle fibers invaded by T cells. *Ann. Neurol.* **16**, 193–208 (1984).
38. Fyhr, I. M., Moslemi, A. R., Lindberg, C. & Oldfors, A. T cell receptor beta-chain repertoire in inclusion body myositis. *J. Neuroimmunol.* **91**, 129–134 (1998).
39. Cavazzana, I., Fredi, M., Selmi, C., Tincani, A. & Franceschini, F. The Clinical and Histological Spectrum of Idiopathic Inflammatory Myopathies. *Clin. Rev. Allergy Immunol.* **52**, 88–98 (2017).
40. van Dongen, J. J. M. *et al.* Design and standardization of PCR primers and protocols for detection of clonal immunoglobulin and T-cell receptor gene recombinations in suspect lymphoproliferations: Report of the BIOMED-2 Concerted Action BMH4-CT98-3936. *Leukemia* **17**, 2257–2317 (2003).
41. Brüggemann, M. *et al.* Standardized next-generation sequencing of immunoglobulin and T-cell receptor gene recombinations for MRD marker identification in acute lymphoblastic leukaemia; a EuroClonality-NGS validation study. *Leukemia* **33**, 2241–2253 (2019).
42. Greenberg, S. A., Pinkus, J. L., Amato, A. A., Kristensen, T. & Dorfman, D. M. Association of inclusion body myositis with T cell large granular lymphocytic leukaemia. *Brain* **139**, 1348–1360 (2016).

43. Greenberg, S. A. *et al.* Highly differentiated cytotoxic T cells in inclusion body myositis. *Brain J. Neurol.* (2019) doi:10.1093/brain/awz207.
44. Bodenhofer, U., Bonatesta, E., Horejš-Kainrath, C. & Hochreiter, S. msa: an R package for multiple sequence alignment. *Bioinforma. Oxf. Engl.* **31**, 3997–3999 (2015).

546 **SUPPLEMENTARY FIGURES:**



547

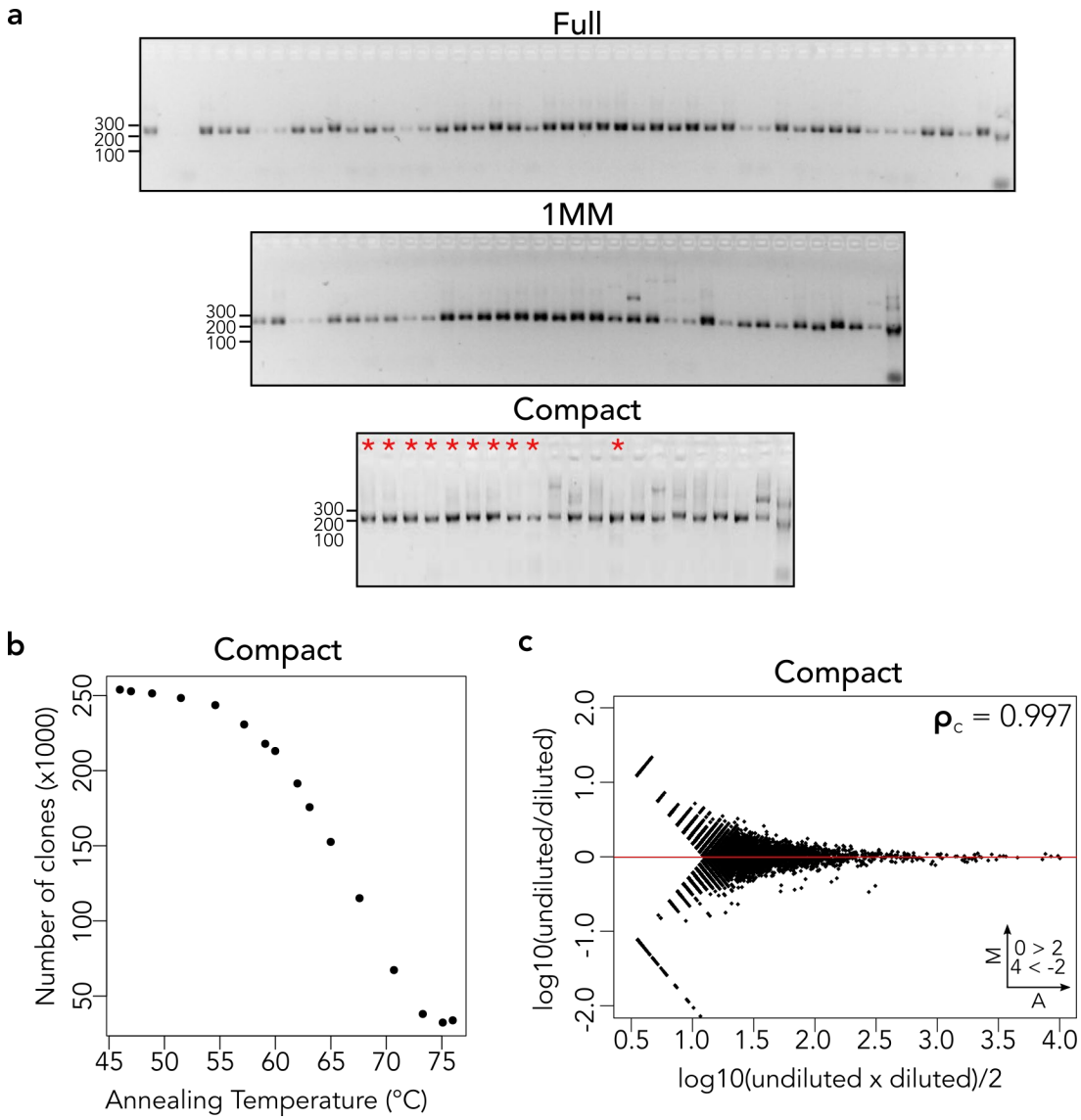

b
START
input unique FR3 20 mer FASTA
Compact primer set: algorithm
number of remaining unique FR3 20 mers > 0
FALSE
END
TRUE
calculate nt frequency at each position of remaining unique FR3 20 mers
determine max nt frequency & set matching primer position to that nt (tie: set most 5' nt)
FALSE
exactly 16/20 primer nts determined
TRUE
> 16/20 primer nts determined
TRUE
remaining primer nts set to inosine
FALSE
max nt identical between FR3s at next position
FALSE
TRUE
TRUE
set matching primer position to max nt, remaining primer nts set to inosine
< 20/20 primer nts determined
set matching primer position to that nt
FALSE
mismatch in 3'-5 nts between primer & unique FR3
TRUE
keep unique FR3 in next primer design cycle
FALSE
1 mismatch or less between primer & unique FR3
FALSE
TRUE
remove unique FR3 in next primer design cycle

548

549 **Figure S1:** FR3AK-seq primer design algorithm schematic. **a.** 1MM primer set design algorithm. Input is a
550 FASTA formatted file containing each unique FR3 20-mer nucleotide (nt) sequence. Nt frequencies at each
551 of the 20 FR3 positions were calculated. The position with the maximum nt frequency in the unique FR3
552 20-mers was used and the corresponding position in the primer was set to that nt. In the case of ties, the
553 most 5' position was set first. Once all 20 nts of the primer were determined, unique FR3 20-mers were
554 removed from the next round of primer design only if they: i) did not have a mismatch in the most 3'-5 nts
555 (as described in the text) and ii) had 1 mismatch or an exact match to the designed primer. The primer
556 design loop continued until no unique FR3 20-mer sequences remained. **b.** Compact primer set design
557 algorithm. Primers were designed as in **a** until 16/20 primer nts were determined. At this point, if the position
558 with the maximum nt frequency in the unique FR3 20-mers was identical to the number of remaining FR3
559 20-mers, then these nts were set to these positions. If they were not and exactly 16/20 nts in the primer
560 were already determined, then the nt with the maximum frequency was set to the corresponding primer
561 position and the remaining 3 nt positions in the primer were set to inosines. If they were not and >16/20 nts
562 in the primer were already determined, then the remaining nt positions were set to inosines. In this way, we
563 allowed for up to 3 inosines and 1 mismatch. Once all 20 nts of the primer were determined, the remaining
564 design was identical to **a**, where we considered each inosine a match to any nt in the FR3 regardless of its
565 position in the primer.

566

567



568

569 **Figure S2:** FR3AK-seq multiplex PCR performance. **a.** Individual primers from all three sets produce
570 amplicons of the expected size from human PBMC cDNA (~220 bp). Asterisks indicate inosine-containing
571 primers. Order of primers can be found in **Table S1**. **b.** The number of clones detected versus annealing
572 temperature using the Compact primer set is shown. A similar trend was observed for the 1MM primer set.
573 To account for differences in number of reads between samples, FASTQ sequencing files were trimmed to
574 identical line numbers before analysis. **c.** MA plot of 2^{10} dilution experiment for the Compact primer set
575 demonstrated high linear concordance (ρ=0.997), indicating negligible PCR bias. Clones with ≥10 counts
576 in either dataset were compared. Insets indicate number of clones above or below Y axis limits (set to 2
577 and –2 for visualization).

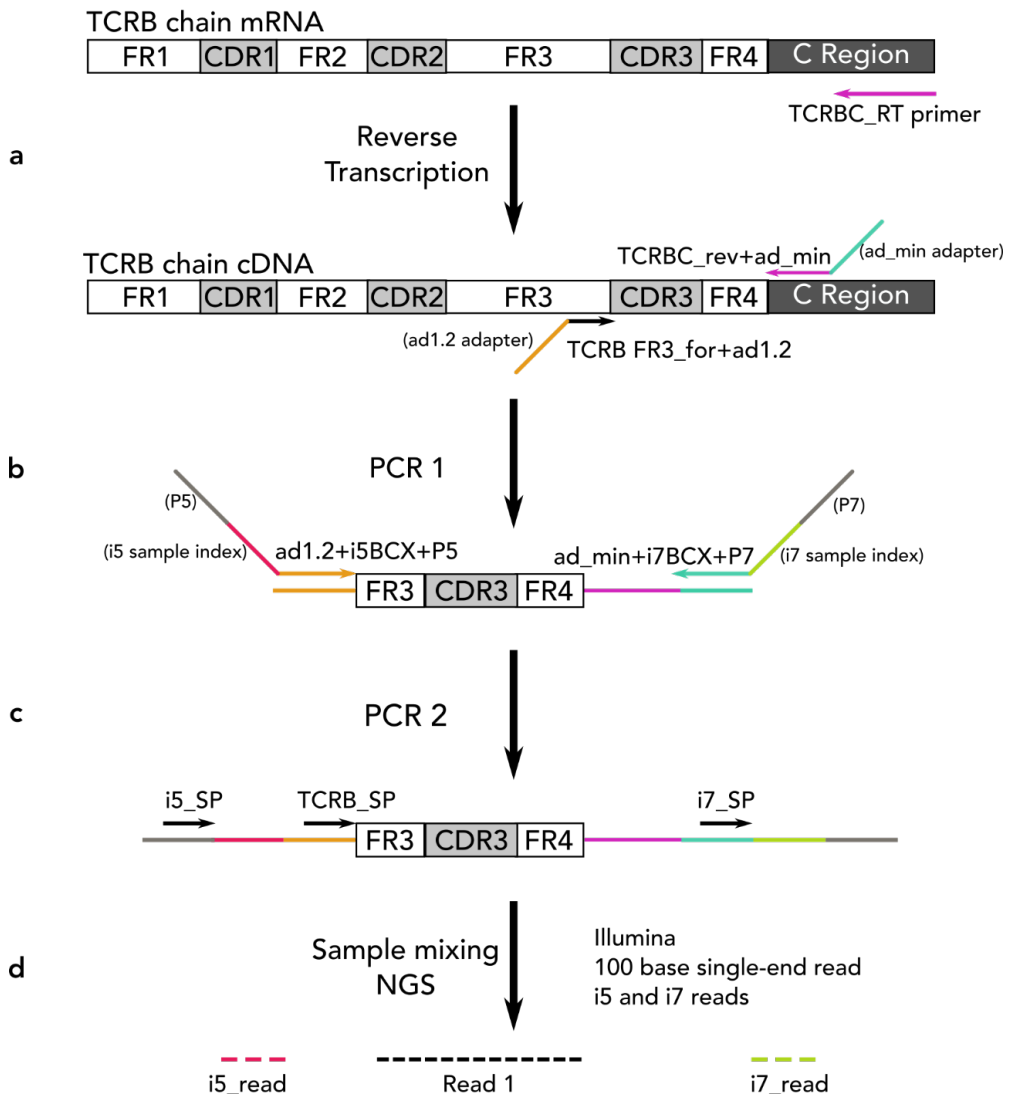


Figure S3: FR3AK-seq dual index sequencing strategy. **a.** Reverse transcription using a TCRB constant region primer generates cDNA for use as template in **b.** PCR1 using our forward TCRB FR3 primers and a hemi-nested TCRB constant region reverse primer. FR3 primers have 5' ad1.2 adapter sequences and the constant region primer has a 5' ad_min adapter sequence. **c.** PCR2 using PCR1 product as a template and indexed ad1.2 and ad_min primers, including P5 and P7 adapters for Illumina sequencing. Samples are mixed upon completion of PCR2 prior to **d.** Next Generation Sequencing (NGS): Illumina 100-base single-end reads along with index 5 (i5) and index 7 (i7) reads. Dual indexes are used for demultiplexing reads from each sample and can be found in **Supplementary File S1**. No index read mismatches were permitted during sample demultiplexing. Primer and adapter sequences can be found in **Table S1**.

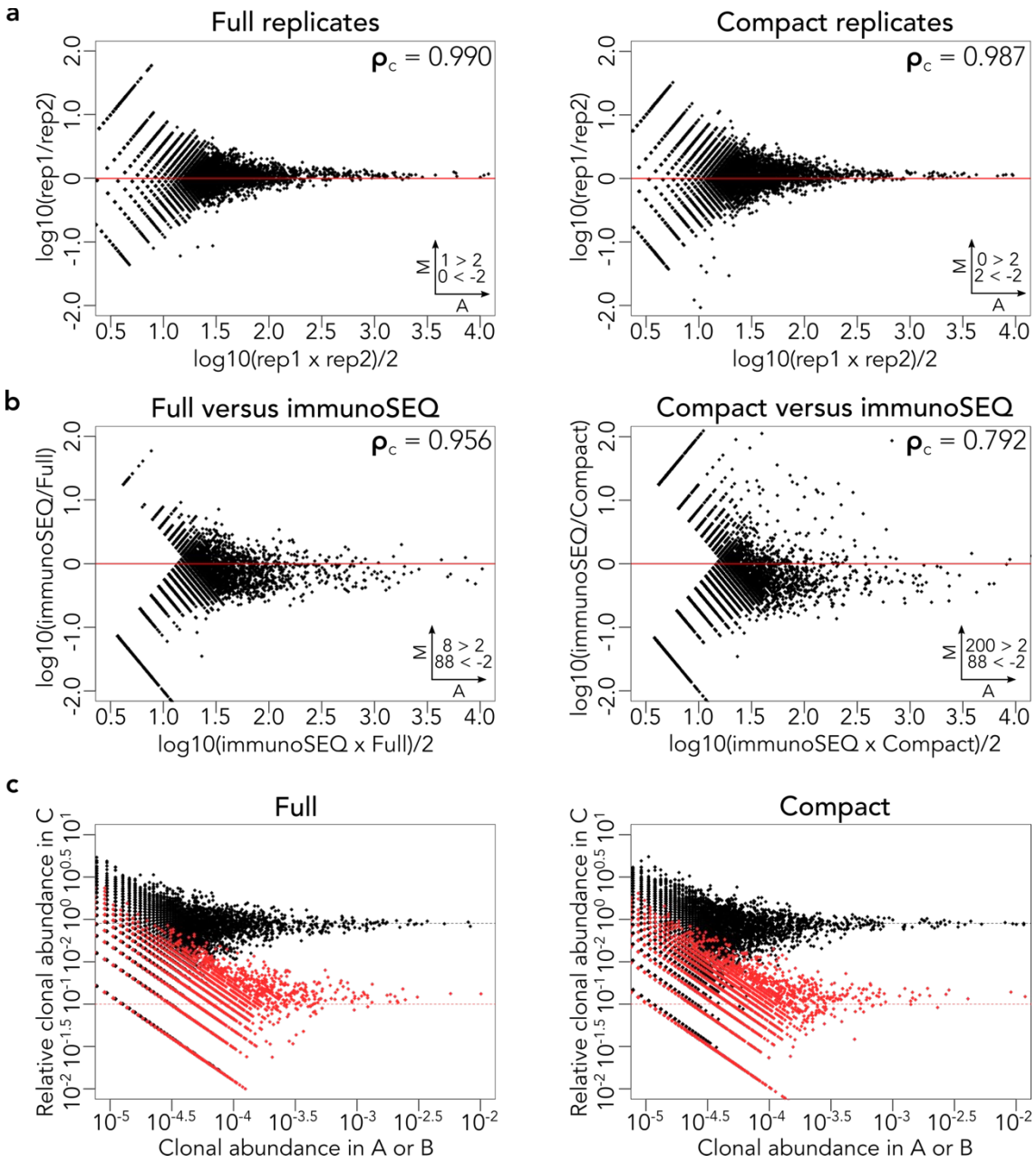


Figure S4: FR3AK-seq Full and Compact primer sets perform comparably to immunoSEQ. **a.** MA plots showing technical replicates of the Full (left panel) and Compact (right panel) FR3AK-seq assays. **b.** MA plot comparing the Full (left panel) and Compact (right panel) FR3AK-seq assay against the immunoSEQ assay. Clones with ≥10 counts in either dataset were included. **c.** Relative frequencies of Donor A (black points) and Donor B (red points) in sample C determined using the Full (left panel) and Compact (right panel) FR3AK-seq assay. Insets in **a** and **b** indicate number of clones above or below Y axis limits (set to 2 and –2 for visualization)

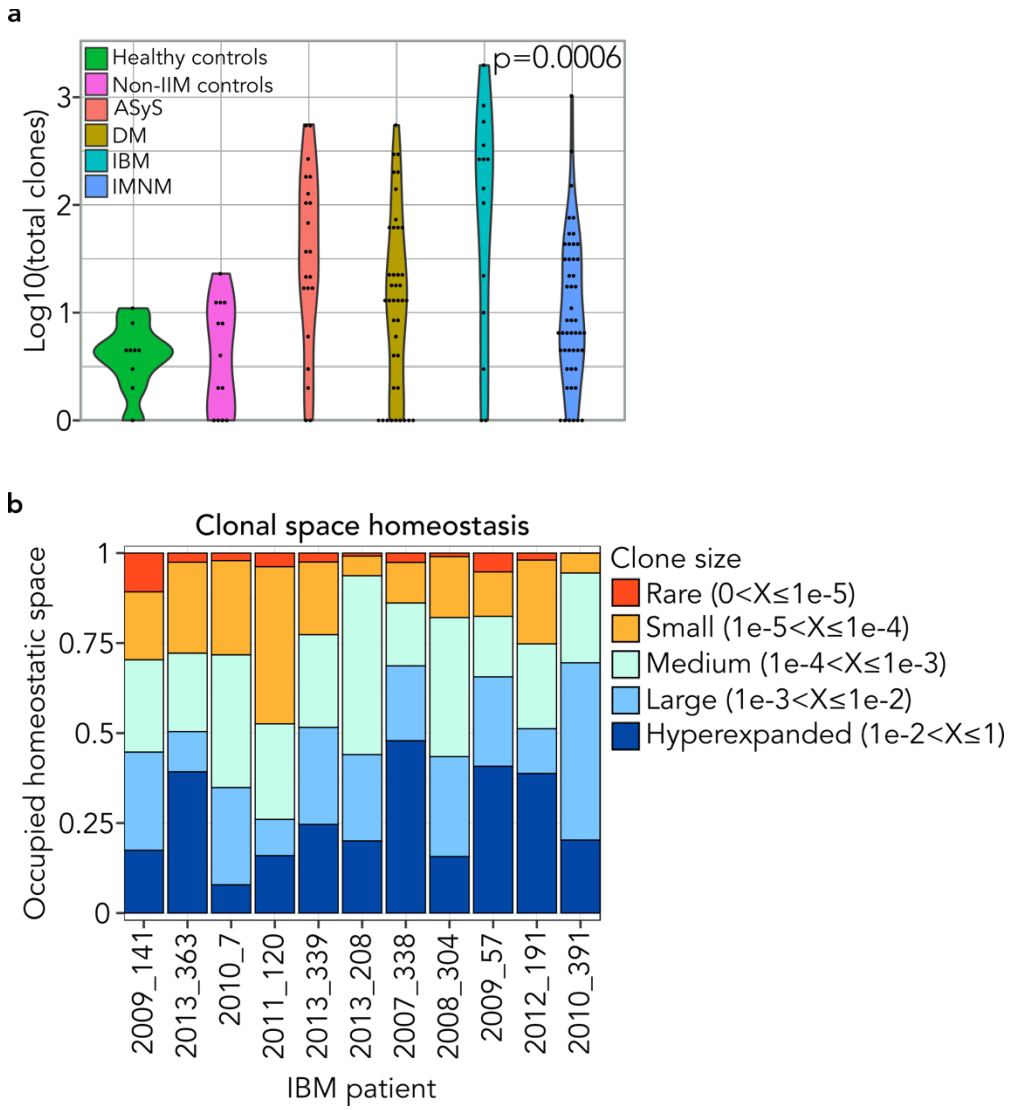


Figure S5: a. Distributions of unique clones per patient for each IIM subgroup and controls. Kruskal-Wallis p-value is shown. **b.** Clonal space homeostasis for each IBM donor with at least one clone present at >2 T cell equivalents in their muscle biopsy. Criteria for clone size are shown.

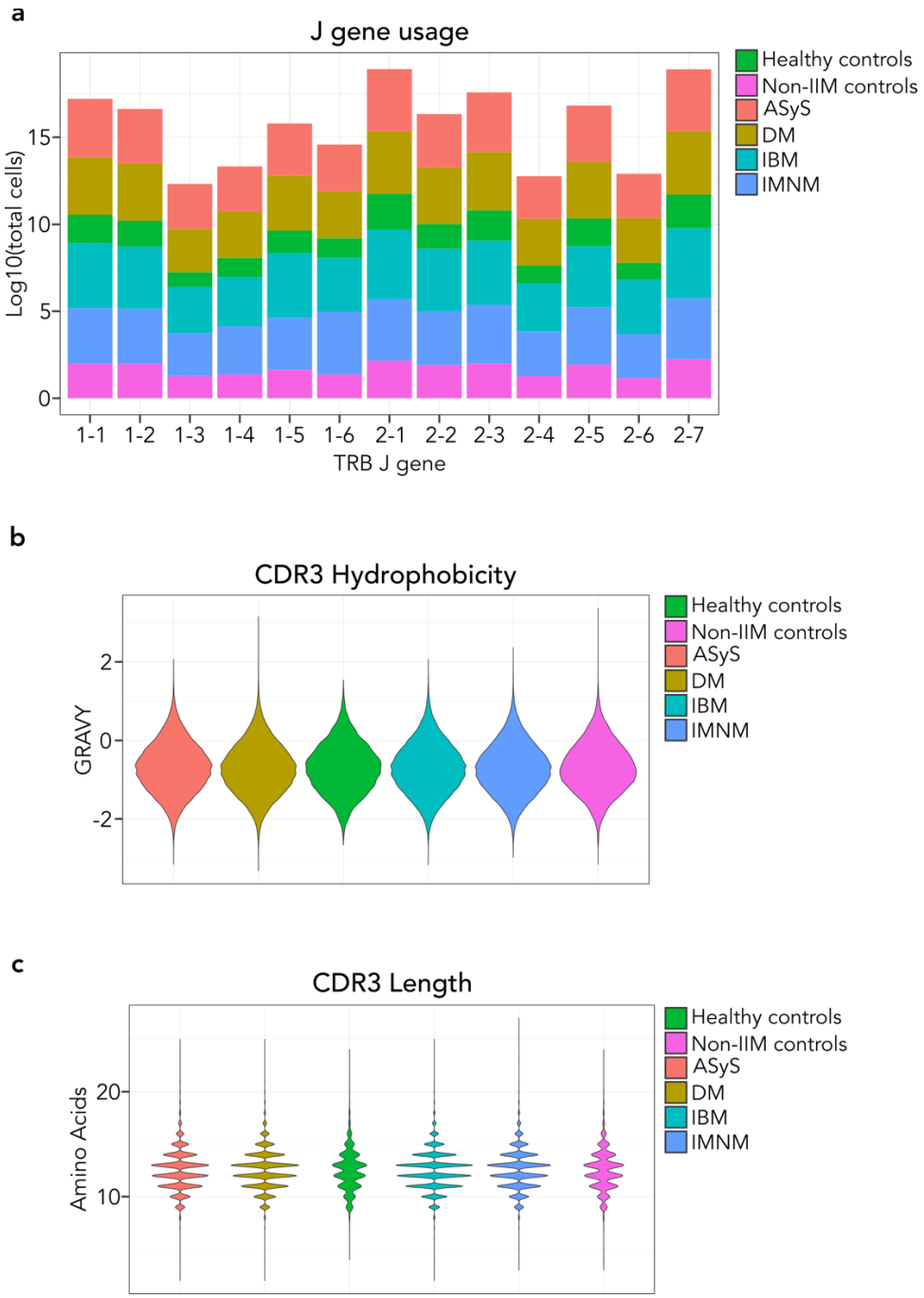


Figure S6: Summary statistics of FR3AK-seq TCRB repertoires from each IIM subgroup and controls. **a.** Stacked bar plot of J gene usage for each IIM subgroup and controls. **b.** Grand Average of Hydropathy (GRAVY) for CDR3s from each IIM subgroup and controls. **c.** CDR3 length comparisons between each IIM subgroup and controls.

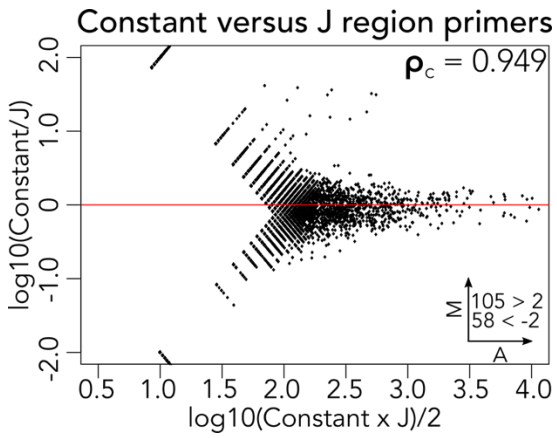


Figure S7: MA plot demonstrating the performance of the FR3AK-seq constant region reverse PCR1 primer as compared to a set of J region reverse primers. cDNA was used as a starting material along with the 1MM forward primer set with PCR conditions as described in the Methods section. CDR3 sequences were included in this analysis if they had a clone count of ≥10 in either data set. J region primers can be found in **Table S8**.

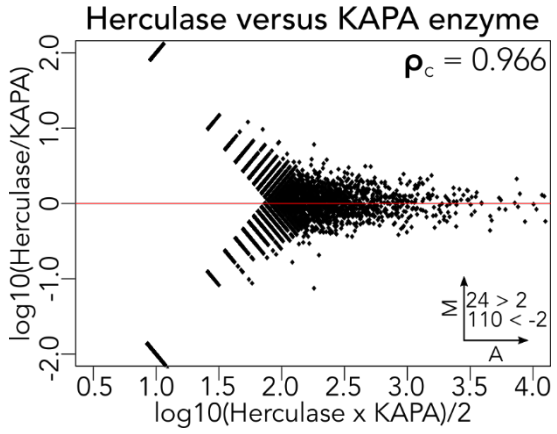


Figure S8: MA plot demonstrates that Herculase II Fusion DNA polymerase performs comparably to KAPA 2G Fast Multiplex Mix in PCR1. CDR3 sequences were included in this analysis if they had a clone count of ≥10 in either data set.

618 **SUPPLEMENTARY TABLES:**

Table S1: FR3AK-seq human TCRB primers			
TCRB constant region primers			
Sequence	**Step**		
accagtgtggccttttgggtgtg	Reverse transcription		
cttttgggtgtgggagatctctg	Reverse primer: PCRI		
Adapter sequence for TCRB constant region PCRI primer (add to 5' end)			
Sequence	**Adapter name**		
ctggagttcagacgtgtgctcttccgatct	ad_min		
Adapter sequence for FR3 primers (add to 5' end)			
Sequence	**Adapter name**		
ggagctgtcgttcactc	AD1.2		
Full: TCRB FR3 primers			
Primer Sequence	**IMGT Accession | Associated V allele**	**Unique FR3 & Heatmap order**	**Gel order**
actccgccgtgtatctctgt	X61441|TRBV7-8*02 M11953|TRBV7-8*01 M27384|TRBV7-8*03	1	25
actctgccgtgtatctctgt	X74843|TRBV7-3*04	2	20
actcggccgtgtatctctgt	X58797|TRBV11-3*02 X61442|TRBV7-2*01 L36190|TRBV7-2*02 X58796|TRBV11-2*03 M27387|TRBV7-2*04 U03115|TRBV11-3*01 L36092|TRBV11-2*01 U07975|TRBV7-2*03 M33235|TRBV11-2*02 MGT000021|TRBV11-2*04 AB305924|TRBV11-3*04	3	47
actcagccgtgtatctctgt	M13550|TRBV7-3*05 X61440|TRBV7-3*01	4	4
actcagctgtgtatctctgt	L36092|TRBV7-4*01	5	36
actcagccgtgtatctccgt	M97943|TRBV7-3*02	6	45
actcagccgcgtatctccgt	AF009660|TRBV7-3*03	7	15
actcggccatgtatcgctgt	X58806|TRBV7-6*02 L36092|TRBV7-6*01	8	23
actcagccatgtatcgctgt	L36092|TRBV7-7*01 X57607|TRBV7-7*02	9	37
actcggccatgtatctctgt	M33233|TRBV11-1*01 L14854|TRBV7-9*07 M27385|TRBV7-9*05 M14261|TRBV7-9*04 AF009663|TRBV7-9*03 M62377|TRBV11-3*03 L36092|TRBV7-9*01	10	26

	M15564|TRBV7-9*02 X74844|TRBV7-9*06		
actcagccgtgtatctctac	U03115|TRBV17*01 IMGT000021|TRBV17*02	11	34
actcagccctgtatctctgc	U07977|TRBV4-1*01 M13855|TRBV4-1*02	12	5
acttggctgtgtatctctgt	X61444|TRBV7-1*01	13	33
actcggccctgtatctctgc	X58812|TRBV4-3*02 L06888|TRBV4-3*03 X57616|TRBV4-3*04 U07978|TRBV4-3*01	14	27
actcggccctgtatctctgt	X58811|TRBV4-2*02 U07975|TRBV4-2*01 X57611|TRBV5-5*02 X57615|TRBV5-4*02 X58803|TRBV5-8*02 L36092|TRBV5-8*01 X58804|TRBV5-4*04 L36092|TRBV5-5*01 L36092|TRBV5-4*01 AF009660|TRBV5-3*02 X58801|TRBV5-5*03 S50547|TRBV5-4*03 X61439|TRBV5-3*01	15	1
actcggccctctatctctgt	L36092|TRBV5-6*01 L36092|TRBV5-7*01	16	18
actcggccctttatctttgc	L36092|TRBV5-1*01 M14271|TRBV5-1*02	17	40
acacggcactgtatctctgc	L36092|TRBV23-1*01	18	39
actcagctgtgtacttctgt	K02546|TRBV12-4*01 X07192|TRBV12-3*01 M14264|TRBV12-4*02	19	31
actcagctgtgtatttttgt	X07223|TRBV12-5*01	20	14
actcagctttgtatttctgt	L36092|TRBV9*01 M27380|TRBV9*03 AF009660|TRBV9*02	21	19
actctgctgtgtatttctgt	L06889|TRBV3-1*02 U07977|TRBV3-1*01	22	28
actcagccatgtacttctgt	M64351|TRBV2*03 L36092|TRBV2*01 M62379|TRBV2*02	23	44
actcagccctgtacttctgt	U03115|TRBV13*01 M62378|TRBV13*02	24	9
actctggcttctatctctgt	L36092|TRBV30*01 M13554|TRBV30*04 L06893|TRBV30*05 Z13967|TRBV30*02	25	22
acgcagccatgtacctgtgt	X58800|TRBV15*02 IMGT000021|TRBV15*04	26	43
acacagccatgtacctgtgt	U03115|TRBV15*01	27	3

acgcagccatgtaccagtgt	M62376|TRBV15*03	28	42
attcggcagcttatttctgt	L36092|TRBV18*01	29	16
attctggagtttatttctgt	X57722|TRBV14*02 X06154|TRBV14*01	30	46
attcagcagtgtatttttgt	L26054|TRBV16*03 L26231|TRBV16*01	31	7
agacatctgtgtacttctgt	AF009662|TRBV6-6*02 X58815|TRBV6-6*03 L33102|TRBV10-3*04 L36092|TRBV6-6*01 U17047|TRBV10-3*02 U03115|TRBV10-3*01 AF009660|TRBV6-4*02 X61446|TRBV6-1*01 L06892|TRBV6-6*05 X74848|TRBV6-6*04 L33101|TRBV10-3*03 X61653|TRBV6-4*01 L36092|TRBV6-5*01	32	29
agacatctgtgtacttgtgt	L36092|TRBV6-8*01	33	21
aaacatctgtgtacttctgt	U07978|TRBV6-3*01 X61445|TRBV6-2*01	34	24
agacatctgtatacttctgt	X61447|TRBV6-9*01	35	30
agacttctgtttacttctgt	L36092|TRBV6-7*01	36	10
agacctctctgtacttctgt	L36092|TRBV27*01	37	6
agacatctatgtacctctgt	U08314|TRBV28*01	38	38
agacagctctttacttctgt	M11951|TRBV24-1*01 IMGT000021|TRBV24-1*02	39	8
agacatctgtatatttctgc	L36092|TRBV10-1*01 AF009660|TRBV10-1*02	40	35
agacatctgtgtatttctgc	L36092|TRBV10-2*01 IMGT000021|TRBV10-2*03	41	32
atacctctcagtacctctgt	L36092|TRBV25-1*01	42	41
cgacagctttctatctctgt	U48259|TRBV19*02 M97725|TRBV19*03 L36092|TRBV19*01	43	2
gcttctacatctgcagtgct	X72719|TRBV20-1*02	44	17
acagcagcttctacatctgc	D13088|TRBV20-1*06 X57604|TRBV20-1*05 M14263|TRBV20-1*04 M11955|TRBV20-1*01 X74852|TRBV20-1*07 M11954|TRBV20-1*03	45	12
acagcagcatatatctctgc	L36092|TRBV29-1*01 X04926|TRBV29-1*03 M13847|TRBV29-1*02	46	13
cccgctcccagacatctgtg	U17048|TRBV10-2*02	47	11
1MM: TCRB FR3 primers			

Primer Sequence	Heatmap Order	Gel Order
actcggccgtgtatctctgt	1	25
actcagctgtgtatctctgt	2	19
actcagccgcgtatctccgt	3	9
actcagccatgtatcgctgt	4	20
actcagccgtgtatctctac	5	26
actcagccctgtatctctgc	6	1
acttggctgtgtatctctgt	7	18
actcggccctctatctctgt	8	11
actcggccctttatctttgc	9	22
acacggcactgtatctctgc	10	30
actcagctgtgtacttctgt	11	16
actcagctgtgtatttttgt	12	28
actcagctttgtatttctgt	13	12
actctgctgtgtatttctgt	14	14
actcagccctgtacttctgt	15	5
actctggcttctatctctgt	16	27
acgcagccatgtacctgtgt	17	24
acgcagccatgtaccagtgt	18	23
attcggcagcttatttctgt	19	10
attctggagtttatttctgt	20	29
attcagcagtgtatttttgt	21	3
agacatctgtgtacttctgt	22	15
agacatctgtgtacttgtgt	23	13
agacttctgtttacttctgt	24	6
agacctctctgtacttctgt	25	2
agacatctatgtacctctgt	26	21
agacagctctttacttctgt	27	4
agacatctgtgtatttctgc	28	17
atacctctcagtacctctgt	29	34
cgacagctttctatctctgt	30	33
gcttctacatctgcagtgct	31	31
acagcagcttctacatctgc	32	7
acagcagcatatatctctgc	33	8
cccgctcccagacatctgtg	34	32
Compact: TCRB FR3 primers		
Primer Sequence	**Heatmap Order**	**Gel Order**
actcigciitgtatctctgt	1	1
actcagccgigtatctccgt	2	7
actcggccatgtatcgctgt	3	11
actcagccgtgtatctctac	4	10
actcigccctitatctitgc	5	5
acacggcactgtatctctgc	6	16
actcagctgtgtatttttgt	7	14
actctggcttctatctctgt	8	12
acicagccatgtacciitgt	9	4
attcggcagiitatttitgt	10	8

attctggagtttatttctgt	11	15
agacatctitgtacititgt	12	3
agaciictitttacttctgt	13	13
actcagciitgtacttctgt	14	2
agacatctgtitatttctgc	15	6
atacctctcagtacctctgt	16	20
cgacagctttctatctctgt	17	19
gcttctacatctgcagtgct	18	17
acagcagcttitaiitctgc	19	9
cccgctcccagacatctgtg	20	18
PCR2 primers		
Primer Sequence: X refers to barcode nucleotides	**Adapters**	
caagcagaagacggcatacgagatXXXXXXXXgtgactggagttcagacgtgtgctcttccgatct	P7_ad_min	
aatgatacggcgaccaccgagatctacacXXXXXXXXacactctttccgatctggagctgtcgttcactc	P5_AD1.2	

Table S1: Human TCRB FR3AK-seq primers and adapter sequences.

Table S2: FR3AK-seq cost per sample			
Step	**Item**	**Price**	**Cost per sample**
Reverse Transcription	Invitrogen Superscript III first-strand synthesis system	$467.00 (50 reactions)	$9.34
cDNA clean up	Zymo Research Oligo Clean & Concentrator	$423.00 (384 preps)	$1.10
PCR1	KAPA2G Fast Multiplex Mix	$295.88 (6.25 mL/ 624-20μL reactions)	$0.47
PCR2	Herculase II Fusion DNA Polymerase	$475.00 (400 uL, 2000-20 uL reactions)	$0.24
Reverse Transcription, 1MM PCR1, PCR2	Primers	$204.75 (36 primers, 6250 reactions PCR1/RT) $11.38 (2 primers, 5000 reactions PCR2)	$0.03 $0.002
Plastics	Pipet tips, tubes, etc.	$1.00/sample	$1.00
Sequencing (HiSeq 2500)	100 nt single-end Illumina reads, 213 samples at ~$1x10^6$ reads per sample	$1600.00 (100bp, per lane)	$7.51
TOTAL			**$19.69**

Table S2: Cost per sample for FR3AK-seq. HiSeq cost calculations were based on the Johns Hopkins Genetics Resources Core Facility High Throughput Sequencing Center (internal pricing) at an estimated

622 $1x10^6$ reads per sample. Note that these calculations do not include overhead, personnel costs, or time

623 spent by the researcher.

Table S3: Idiopathic Inflammatory Myopathy and Control Muscle Biopsy Cohort						
	Controls (n=21)	DM (n=40)	ASyS (n=21)	IMNM (n=49)	IBM (n=14)	Total (n=145)
Female	71% (15)	53% (9)	89% (8)	67% (18)	54% (7)	66% (57)
White	82% (9)	71% (12)	44% (4)	56% (15)	83% (10)	66% (50)
Black	18% (2)	18% (3)	44% (4)	37% (10)	17% (2)	28% (21)
Other race	0% (0)	12% (2)	11% (1)	7% (2)	0% (0)	7% (5)
Age	50.6 (13.9)	59.7 (13.4)	46.2 (8.4)	51.8 (13.4)	66.1 (11.6)**	54.6 (14.0)

624 * p<0.05,**p<0.01,***p<0.001

625 **Table S3:** Description of the IIM muscle biopsy cohort. Chi-squared or Fisher´s exact tests were used to

626 compare the dichotomous variables of each one of the clinical groups with the rest. Student's T test was

627 used to compare the age of each one of the clinical groups with the rest.

Table S4: GLIPH summary	
Input IIM FR3AK-seq data	
Number of samples with non-bystander T cell clones	122 (23 excluded)
Number of non-bystander T cell clones included in analysis	11351
Number of bystander T cell clones	278121
GLIPH Output	
Total number of GLIPH clusters	527
Total number of GLIPH clusters with at least 3 individuals contributing a T cell clone	143
Number of non-bystander T cell clones clustered by GLIPH	2125

628 **Table S4:** Summary of IIM and control muscle biopsy FR3AK-seq data used in GLIPH analysis. Bystander

629 clones refer to those appearing at a frequency of less than or equal to 1 cellular equivalent.

630

631

632

633

634

Table S5: Comparisons between platforms							
Platform	**ImmunoSEQ hsTCRB service (Adaptive Biotechnologies)**		**ImmunoSEQ hsTCRB kit (Adaptive Biotechnologies)**		**Immunoverse (ArcherDX)**		**FR3AK-seq**
Method	Multiplex PCR with bias correction		Multiplex PCR with bias correction		UMI		Multiplex PCR
Time	• <1 day to prepare sample and ship to company • Processed data within ~2-3 weeks		• <1 day to prepare sample • ~1 day to obtain processed data		• ~2 days for library preparation • ~1 day to obtain processed data		• ~1 day for library preparation • ~1 day to obtain processed data
Hands-on user steps	• Reverse transcription (if using cDNA)		• 3 step library preparation (if using cDNA) • Sequencing • Upload fastq files to immunoSEQ analyzer for bias correction		• 9 step library preparation • Sequencing • Upload fastq files to ArcherDX portal for deduplication analysis		• 3 step library preparation • Sequencing • MiXCR analysis
Cost per sample	Reverse transcription	$9.87	Kit	$629.00	Kit	$240.00	$19.69 (see **Table S2**)
	Deep Sequencing Service	$1100	MiSeq	$44.00	MiSeq	$44.00	
	TOTAL: $1109.87		**TOTAL**: $642.02		**TOTAL:** $284.00		**TOTAL:** $19.69

Table S5: Comparisons between Adaptive Biotechnologies' immunoSEQ assay (both service and kit), ArcherDX's Immunoverse assay, and FR3AK-seq. All prices are approximate. MiSeq cost calculations were based on the Johns Hopkins Transcriptomics and Microarray Core (JHU internal pricing) at an estimated $1x10^6$ reads per sample.

Table S6: FR3AK-seq *Mus musculus* TCRB primers	
***Mus musculus* TCRB constant region primers**	
Sequence	**Step**
tggctcaaacaaggagaccttgg	Reverse transcription
ggtggagtcacatttctcagatcc	Reverse primer: PCRI
Full: *Mus musculus* TCRB FR3 primers	
Primer Sequence	**IMGT Accession | Associated V allele**
acagcagtatctatttctgt	X16695|TRBV30*01
agatgaccgtttttctctgt	X01643|TRBV19*02
actctgctgtgtatctctgt	AJ249819|TRBV4*02

	X56725|TRBV4*01
agacatctttgtacttctgt	AE000663|TRBV13-1*02 M15618|TRBV13-1*01
aaggcagaaccttgtactgc	X01642|TRBV1*02
agatggccgtttttctctgt	AE000664|TRBV19*01 AJ249821|TRBV19*03
actctggcttctacctctgt	X03277|TRBV31*01 M26418|TRBV31*02
actcagctgtctatttttgt	AJ249820|TRBV5*05 M20878|TRBV5*04 X00438|TRBV5*02 X02779|TRBV5*03 AE000663|TRBV5*01
acagaggcttatatctctgt	AE000664|TRBV20*01 M11859|TRBV20*02
attcagctgtgtacttctgt	X16691|TRBV21*01
actcagcggtgtatctttgt	X14388|TRBV16*04 M15459|TRBV16*03 M13670|TRBV16*02 L29434|TRBV16*01
actcagcactgtacttgtgc	AE000664|TRBV23*01
agggcagaaccttgtactgc	AE000663|TRBV1*01
agacagctgtatatttctgt	M15616|TRBV13-3*01
actcagctgtgtacttctgt	X03865|TRBV3*02 AE000663|TRBV3*01
actcagcactgtacctctgt	L48997|TRBV24*04 M16203|TRBV24*03 AJ249822|TRBV26*02 K02548|TRBV26*01 M61184|TRBV24*01
actcagctgtgtatctgtgt	AE000664|TRBV15*01
agacatcagtgtacttctgt	AJ250103|TRBV13-2*05 M26417|TRBV13-2*03 AJ249823|TRBV13-2*04 M15617|TRBV13-2*01 M27350|TRBV13-2*02
attctgccatgtacctctgt	AE000664|TRBV17*01
agacatctgtgtacttctgt	X00696|TRBV29*02 AE000664|TRBV29*01
actcggccacatacttctgt	AE000663|TRBV2*01
actctgccgtgtacttctgt	M15613|TRBV12-2*01 X02782|TRBV12-2*02
actctgctatgtacttctgt	M30881|TRBV12-1*02 M15614|TRBV12-1*01
acacagccacctatctctgt	AE000664|TRBV14*01

Table S6: *Mus musculus* TCRB FR3AK-seq primers. Adapter sequences can be found in **Table S1**.

Table S7: FR3AK-seq human TCRA, TCRG, and TCRD primers	
human TCRA: constant region primers	
Sequence	**Step**
cggtgaataggcagacagacttg	Reverse transcription
cagacttgtcactggatttagag	Reverse primer: PCRI
Full: human TCRA FR3 primers	
Primer Sequence	**IMGT Accession | Associated V allele**
atgctgctgtgtactactgc	AE000660|TRAV26-2*01
actcagccacctacttctgt	M17660|TRAV23/DV6*02 AE000660|TRAV23/DV6*01 M97704|TRAV23/DV6*03
actcagctatctacttctgt	AE000659|TRAV5*01
atgctgctgagtacttctgt	AE000659|TRAV8-3*01 AE000660|TRAV8-7*01 M35617|TRAV8-3*02
atgcgtctgagtacttctgt	L06885|TRAV8-3*03
atacaggccactacctctgt	X04957|TRAV27*02
actccgctgtgtacttctgt	AE000659|TRAV9-1*01
actcagccacctacctctgt	X58736|TRAV21*02 AE000660|TRAV21*01
atacaggcctctacctctgt	AE000660|TRAV27*01 D13075|TRAV27*03
actctgccgtgtactactgc	AE000660|TRAV18*01
actcagctgtctacttttgt	M17658|TRAV13-2*02 AE000659|TRAV13-2*01
actcagcaatgtatttctgt	AE000659|TRAV14/DV4*02 M21624|TRAV14/DV4*03
atgtaggaacctacttctgt	AE000660|TRAV25*01
attcagccacctatttctgt	AE000659|TRAV7*01
atgcaggcatctacctctgt	AE000660|TRAV34*01
acacagctgagtacttctgt	AE000659|TRAV8-1*01
attcagccacatacctctgt	M17661|TRAV24*02
atgtaggcatctacttctgt	AE000660|TRAV35*01 X58738|TRAV35*02
acgcggctgagtacttctgt	M12959|TRAV8-4*04 X63455|TRAV8-4*05 M17650|TRAV8-2*02 M12423|TRAV8-4*02 D13077|TRAV8-4*03 AE000659|TRAV8-2*01 AE000659|TRAV8-4*01 K02777|TRAV8-4*06
actcaggcgtttatttctgt	AE000660|TRAV22*01
actcggccgtctacctctgt	Z46643|TRAV36/DV7*04 X58767|TRAV36/DV7*03 X61070|TRAV36/DV7*02
acactgcgatgtatttctgt	AE000661|TRAV38-1*01 L06880|TRAV38-1*04 M95394|TRAV38-1*03

	M64355|TRAV38-1*02
actcagcaatgtacttctgt	M21626|TRAV14/DV4*01
actcagcagtatacttctgt	AE000660|TRAV19*01
actcagccgtgtactactgt	X73521|TRAV40*01
acacggctgagtacttctgt	X02850|TRAV8-6*01 AE000659|TRAV8-6*02
actcagccacatacctctgt	AE000660|TRAV24*01
actcagctacctacctctgt	AE000659|TRAV6*01 X58747|TRAV6*02 Z49060|TRAV6*03 Y10410|TRAV6*05 Y10409|TRAV6*04
actctgcagtgtacttctgt	S81645|TRAV29/DV5*02 AE000660|TRAV29/DV5*01
acactgctgtgtactattgc	AE000660|TRAV26-1*01 M27371|TRAV26-1*02 L06886|TRAV26-1*03
actcggccatctacctctgt	AE000660|TRAV36/DV7*01
actctgcctcttacttctgc	X04939|TRAV1-1*02 AE000658|TRAV1-1*01
tctctgccacctacttctgt	AE000661|TRAV39*01
actctgcctcttacctctgt	AE000658|TRAV1-2*01
acactgctgtgtactactgc	AE000658|TRAV4*01
actccgctttgtacttctgt	AE000658|TRAV3*01
actcagccacttatctctgt	X70305|TRAV20*04 S60789|TRAV20*03 AE000660|TRAV20*01 X68696|TRAV20*02
atgctgctgtttactactgt	AE000658|TRAV2*01 M17659|TRAV2*02
acccggctgagtacttctgt	M17665|TRAV8-4*07
attcagccacctacctctgt	AE000659|TRAV12-2*01 AE000659|TRAV12-1*01 M17657|TRAV12-1*02 M17656|TRAV12-3*02 M81774|TRAV12-2*02 X04946|TRAV12-2*03 AE000659|TRAV12-3*01
actctgccgtctacatctgt	AE000661|TRAV41*01
actcaggaacctacttctgc	L06883|TRAV30*03 AE000660|TRAV30*01 X58768|TRAV30*02
atgccgcgatgtatttctgt	AE000661|TRAV38-2/DV8*01
attcagcctcctacatctgt	AE000659|TRAV10*01
actcggctgtctacttctgt	AE000659|TRAV13-1*01 X04954|TRAV13-1*02
acactgcttcttacttctgt	AE000660|TRAV17*01
actcagcggtgtacttctgt	L06882|TRAV9-2*04 X58745|TRAV9-2*02 L06881|TRAV9-2*03

	AE000659|TRAV9-2*01
actcagccatgtattactgt	AE000659|TRAV16*01
1MM: human TCRA FR3 primers	
actccgctgtgtacttctgt	
actcagcggtgtacttctgt	
actcggccgtctacctctgt	
actcagccacctacctctgt	
acgcggctgagtacttctgt	
actcagcagtatacttctgt	
actcagctatctacttctgt	
actctgccgtctacatctgt	
actcagcaatgtacttctgt	
actctgccgtgtactactgc	
acactgcttcttacttctgt	
attcagccacctatttctgt	
actctgcctcttacctctgt	
actcggctgtctacttctgt	
atacaggcctctacctctgt	
actcagccgtgtactactgt	
acactgcgatgtatttctgt	
atgctgctgtgtactactgc	
attcagcctcctacatctgt	
actctgcagtgtacttctgt	
atgccgcgatgtatttctgt	
atgctgctgtttactactgt	
actcagccacttatctctgt	
acactgctgtgtactactgc	
atgcaggcatctacctctgt	
actcagccatgtattactgt	
actctgcctcttacttctgc	
atgtaggcatctacttctgt	
acactgctgtgtactattgc	
actcaggcgtttatttctgt	
actcagctgtctacttttgt	
actcaggaacctacttctgc	
atgcgtctgagtacttctgt	
atgctgctgagtacttctgt	
atgtaggaacctacttctgt	
acacagctgagtacttctgt	
tctctgccacctacttctgt	
attcagccacatacctctgt	
Compact: human TCRA FR3 primers	
actcigcigtitacttctgt	
actcagccaiitacitctgt	
acicigctgagtacttctgt	
actcagccitgtaitictgt	
actctgccgiitactictgc	
aticaggciictacctctgt	
aiictgctgtgtactaitgc	

attcagccicctaiitctgt	
acactgcgatgtatttctgt	
atgtaggiaictacttctgt	
actcaggcgtttatttctgt	
actctgcctcttacctctgt	
actcagctgtctacttttgt	
atgccgcgatgtatttctgt	
actcaggaacctacttctgc	
atgcgtctgagtacttctgt	
acactgcttcttacttctgt	
tctctgccacctacttctgt	
atgctgctgtttactactgt	
human TCRG: constant region primers	
Sequence	**Step**
tatgttccagccttctggag	Reverse transcription
gccttctggagctttgtttcagc gccttctggagttttgtttcagc	Reverse primers: PCRI
Full: human TCRG FR3 primers	
Primer Sequence	**IMGT Accession | Associated V allele**
atgaggtggtgtaccactgt	Y11227|TRGV11*01 AF159056|TRGV11*02
attctgggttctattactgt	M12949|TRGV1*01
actttggggtctattactgt	IMGT000011|TRGV2*03
actctggggtctattactgt	M27337|TRGV2*02 M13429|TRGV2*01 M13434|TRGV8*01
acatagctacctactactgt	X15274|TRGV9*02 X07205|TRGV9*01
acatggccgtttactactgt	X74798|TRGV10*02 X07206|TRGV10*01
attctggggtctattactgt	X13355|TRGV5*01 X04038|TRGV3*02 M13430|TRGV3*01
actctggagtctattactgt	X13354|TRGV4*02 X15272|TRGV4*01
1MM: human TCRG FR3 primers	
actctggggtctattactgt	
atgaggtggtgtaccactgt	
acatggccgtttactactgt	
acatagctacctactactgt	
attctgggttctattactgt	
Compact: human TCRG FR3 primers	
aititggigtctattactgt	
acatagctacctactactgt	
atgaggtggtgtaccactgt	
acatggccgtttactactgt	
human TCRD: constant region primers	
Sequence	**Step**

gaattccttcaccagacaagcgac	Reverse transcription
ccagacaagcgacatttgttcc	Reverse primer: PCRI
Full: human TCRD FR3 primers	
Primer Sequence	**IMGT Accession | Associated V allele**
attcagcaaagtacttttgt	M22198|TRDV1*01
atgccgcgatgtatttctgt	AE000661|TRAV38-2/DV8*01
actctgcagtgtacttctgt	S81645|TRAV29/DV5*02 AE000660|TRAV29/DV5*01
actcggccatctacctctgt	AE000660|TRAV36/DV7*01
actcagccacctacttctgt	M97704|TRAV23/DV6*03 Y10411|TRAV23/DV6*04 AE000660|TRAV23/DV6*01 M17660|TRAV23/DV6*02
actcagcaatgtatttctgt	AE000659|TRAV14/DV4*02 M21624|TRAV14/DV4*03
actcggccgtctacctctgt	Z46643|TRAV36/DV7*04 X58767|TRAV36/DV7*03 X61070|TRAV36/DV7*02
acagtgccacttactactgt	M23326|TRDV3*01 X15261|TRDV3*02
atgaagggtcttactactgt	X15207|TRDV2*01 Y13426|TRDV2*02 AE000661|TRDV2*03
actcagcaatgtacttctgt	M21626|TRAV14/DV4*01
1MM: human TCRD FR3 primers	
actcggccgtctacctctgt	
actcagcaatgtacttctgt	
actctgcagtgtacttctgt	
atgccgcgatgtatttctgt	
atgaagggtcttactactgt	
actcagccacctacttctgt	
acagtgccacttactactgt	
attcagcaaagtacttttgt	
Compact: human TCRD FR3 primers	
actcggccitctacctctgt	
actcagcaaiitaittctgt	
atgccgcgatgtatttctgt	
actctgcagtgtacttctgt	
atgaagggtcttactactgt	
acagtgccacttactactgt	
attcagcaaagtacttttgt	

Table S7: Human TCRA, TCRG, and TCRD FR3AK-seq primers. Adapter sequences can be found in **Table S1**

Table S8: FR3AK-seq human TCRB J region primers	
Primer Sequence	**IMGT Accession | Associated J allele**
ccagcactgagagccgggtc	X02987|TRBJ2-4*01
ccagaaccaggagtcctccg	X02987|TRBJ2-2P*01
cgagcaccaggagccgcgtg	X02987|TRBJ2-5*01
ccagtacggtcagcctagag	X02987|TRBJ2-2*01
cgagcactgtcagccgggtg	X02987|TRBJ2-3*01
ctgtgaccgtgagcctggtg	X02987|TRBJ2-7*02 M14159|TRBJ2-7*01
ctacaacggttaacctggtc	K02545|TRBJ1-2*01
ctacaacagtgagccaactt	M14158|TRBJ1-3*01
ctgtcacagtgagcctggtc	M14158|TRBJ1-6*01 L36092|TRBJ1-6*02
ctacaactgtgagtctggtg	K02545|TRBJ1-1*01
ctaggatggagagtcgagtc	M14158|TRBJ1-5*01
ccaagacagagagctgggtt	M14158|TRBJ1-4*01
ccagcacggtcagcctgctg	X02987|TRBJ2-6*01
ctagcacggtgagccgtgtc	X02987|TRBJ2-1*01

Table S8: Human TRBJ FR3AK-seq primers. Adapter sequences can be found in **Table S1**.

Table S9: Software parameters	
MiXCR v2.1.11	
Parameter	**Rationale**
assemble -OaddReadsCountOnClustering=true	makes error correction algorithms to combine clone abundancies (all "children" clone counts added to head clone)
assemble --index	For IIM samples only; links reads in fastq file to CDR3s
exportClones --chains TRB	exports TCRB chain sequences only
GLIPH 1.0	
Algorithm step **--inputs required**	**Parameters** **Input file(s) column names**
1) ./gliph/bin/gliph-group-discovery.pl --tcr TCR_TABLE	default TCR_TABLE = only CDR3b sequences
2) ./gliph/bin/gliph-group-scoring.pl --convergence_file TCR_TABLE-convergence-groups.txt --clone_annotations FULL_TCR_TABLE --motif_pval_file TCR_TABLE.minp.ove10.txt	default TCR_TABLE-convergence-groups.txt = from step 1 FULL_TCR_TABLE = CDR3b, Patient, Counts TCR_TABLE.minp.ove10.txt Counts = from step 1

Table S9: MiXCR and GLIPH parameters used in analysis. Default parameters were used except as noted. Only CDR3b sequences were used with the gliph-group-discovery algorithm. The gliph-group-scoring algorithm was used to match patient and counts with original CDR3b sequences.

Table S10: Summary of Donor A clones across platforms			
Platform	**ImmunoSEQ hsTCRB service (Adaptive Biotechnologies)**	**Immunoverse (ArcherDX)**	**FR3AK-seq**
Total clones detected	394797	67274	362288
Number of singletons (clone count = 1)	345880	20570 (pre-deduplication) 31668 (post-deduplication)	198635
Number of clones present at ≥ 10 counts	2702	10689 (pre-deduplication) 1080 (post-deduplication)	7189

Table S10: Summary of Donor A CDR3 sequences detected across platforms.