



HAL
open science

Benchmarking of T cell receptor repertoire profiling methods reveals large systematic biases

Pierre Barennes, Valentin Quiniou, Mikhail Shugay, Evgeniy Egorov, Alexey Davydov, Dmitriy Chudakov, Imran Uddin, Mazlina Ismail, Theres Oakes, Benny Chain, et al.

► To cite this version:

Pierre Barennes, Valentin Quiniou, Mikhail Shugay, Evgeniy Egorov, Alexey Davydov, et al.. Benchmarking of T cell receptor repertoire profiling methods reveals large systematic biases. *Nature Biotechnology*, 2021, 39 (2), pp.236-245. 10.1038/s41587-020-0656-3 . hal-03838554

HAL Id: hal-03838554

<https://hal.sorbonne-universite.fr/hal-03838554v1>

Submitted on 25 Feb 2023

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Distributed under a Creative Commons Attribution - NoDerivatives 4.0 International License

1 Systematic study of T-cell receptor repertoire profiling reveals large 2 methodological biases: lessons from a multicenter study

3

4 Pierre Barennes^{1,2}, Valentin Quiniou^{1,2}, Mikhail Shugay^{3,4,5,6}, Evgeniy S. Egorov⁴, Alexey N.

5 Davydov⁶, Dmitriy M. Chudakov^{3,4,5,6}, Imran Uddin⁷, Mazlina Ismail⁷, Theres Oakes⁷, Benny

6 Chain⁷, Anne Eugster⁸, Karl Kashofer⁹, Peter P. Rainer¹⁰, Samuel Darko¹¹, Amy Ransier¹¹, Daniel

7 C. Douek¹¹, David Klatzmann^{1,2}, Encarnita Mariotti-Ferrandiz^{1*}

8

9 1. Sorbonne Université, INSERM, Immunology-Immunopathology-Immunotherapy (i3), Paris,
10 France

11 2. AP-HP, Hôpital Pitié-Salpêtrière, Biotherapy (CIC-BTi) and Inflammation-Immunopathology-
12 Biotherapy Department (i2B), Paris, France

13 3. Center of Life Sciences, Skoltech, Moscow, Russia

14 4. Genomics of Adaptive Immunity Department, Shemyakin and Ovchinnikov Institute of
15 Bioorganic Chemistry, Moscow, Russia

16 5. Center for Precision Genome Editing and Genetic Technologies for Biomedicine, Pirogov Russian
17 National Research Medical University, Moscow, Russia

18 6. Adaptive Immunity Group, Central European Institute of Technology, Brno, Czechia

19 7. Division of Infection and Immunity, University College London, United Kingdom

20 8. DFG-Centre for Regenerative Therapies Dresden, Faculty of Medicine Carl Gustav Carus,
21 Technische Universität Dresden, Fetscherstrasse 105, 01307 Dresden, Germany

22 9. Diagnostic and Research Institute of Pathology, Medical University of Graz, Graz, Austria

23 10. Division of Cardiology, Medical University of Graz, Graz, Austria

24 11. Vaccine Research Center, National Institute of Allergy and Infectious Diseases, National
25 Institutes of Health, Bethesda, MD, United States

26

27 *Corresponding author:

28 Encarnita Mariotti-Ferrandiz, PhD

29 encarnita.mariotti@sorbonne-universite.fr

30

31

32

33 **Acknowledgments:** We are grateful to M. Barbie for providing the human samples. This work
34 benefited from equipment and services from the iGenSeq core facility, at ICM.

35 **Funding:** This work was supported the ERC-Advanced TRiPoD (322856), LabEx Transimmunom
36 (ANR-11-IDEX-0004-02) and RHU iMAP (ANR-16-RHUS-0001) grants to DK. MS and DMC were
37 supported by a grant from the Ministry of Science and Higher Education of the Russian
38 Federation (075-15-2019-1789). This work was funded in part by the intramural program of the
39 National Institute of Allergy and Infectious Diseases (DCD). BC was supported by the National
40 Institute for Health Research UCL Hospitals Biomedical Research. AE was supported by DFG
41 CRTD (FZ 111). AND was supported by the Ministry of Education, Youth and Sports of the Czech
42 Republic under the project CEITEC 2020 (LQ1601). EMF, KK and PPR were supported by the
43 European Research Area Network – Cardiovascular Diseases (ERA-CVD, JCT2018, AIR-MI
44 Consortium) program.

45 **Author contributions:** PB, VQ, ESE, AND, IU, MI, TO, AE, SD, AR, KK, PR performed the
46 experiments and raw data pre-processing. PB, VQ, MS and MI analyzed the data. EMF, DMC,
47 BC, DCD and DK designed the experiments. PB, DK and EMF wrote the manuscript with input
48 from all authors. DK and EMF conceived the study, which was supervised by EMF. DK, BC, AE,
49 DCD, DMC, KK and MS obtained funding for the study.

50 **Competing Interests statement:** DMC and MS are cofounders of MiLaboratory LLC.

51 **SUMMARY**

52 Accurate profiling of T-cell receptor (TCR) repertoires is key to monitoring adaptive immunity.
53 We systematically compared TCR sequences obtained with 9 methods applied to aliquots of
54 the same T-cell sample. We observed marked differences in accuracy and intra- and inter-
55 method reproducibility for alpha (TRA) and beta (TRB) TCR chains. Most methods showed lower
56 ability to capture TRA than TRB diversity. Low RNA input generated non-representative
57 repertoires. Results from 5'RACE-PCR methods were consistent among themselves, while
58 differing from the RNA-based multiplex-PCR results. gDNA-based multiplex-PCR methods also
59 differed from each other. Using an in silico meta-repertoire generated from 108 replicates, we
60 found that one gDNA-based method and two non-UMI RNA-based methods were more
61 sensitive than UMI methods in detecting rare clonotypes, despite the better clonotype
62 quantification accuracy of the latter. This study delineates the advantages and limitations of
63 different TCR sequencing methods, which should help the study, diagnosis and treatment of
64 human diseases.

65 INTRODUCTION

66 T-cell receptors (TCR), which drive T-cell activation by antigenic peptide recognition, are
67 heterodimers formed by an α and a β chain¹ produced by somatic V(D)J rearrangements during
68 thymopoiesis² of 47V and 61J functional TRA genes and 48V, 2D, 12J functional TRB genes³. The
69 stochastic V(D)J recombination generates a combinatorial diversity that is further increased by
70 random nucleotide excision and addition at the V(D)J junctions. The independent
71 recombination and subsequent pairing of TRA and TRB chains add an additional level of
72 combinatorial diversity. Recently, computational chain pairing experiments suggested that the
73 potential diversity of the paired repertoire is $\sim 2 \times 10^{19}$ TCRs⁴, while the number of different TRB
74 clonotypes in an individual has been estimated to range from 10^6 to 10^8 ⁵⁻⁷. The TCR repertoire
75 is dynamic, as lymphocytes are continuously generated, die and expand in response to
76 stimulation, and reflects both an individual's immune potential and history.

77 Analysis of the TCR repertoire by deep sequencing (TCRseq) is increasingly used to measure
78 lymphocyte dynamics in health, in pathological contexts such as autoimmune disease,
79 infections and cancer⁸⁻¹⁴, and following interventions such as vaccination^{11,15-18} and
80 immunotherapy¹⁹⁻²², with the goal of identifying TCR biomarkers of disease or of clinical
81 response to treatment and to stratify patients for precision medicine²³. These diverse
82 applications have different requirements in terms of sensitivity, specificity and depth.
83 Accurately capturing the TCR repertoire therefore presents great challenges. A large number
84 of TCRseq methods have been developed. They are all complex multistep protocols, and each
85 step may have a profound impact on the TCRseq data and hence on their interpretation²⁴.

86 Methods can be broadly classified as DNA- or RNA-based, and the latter can be categorized as
87 using multiplex PCR (mPCR) with panels of V and J primers^{5,25,26} or using rapid amplification of
88 cDNA-ends by PCR (RACE-PCR)^{14,27-29} optionally incorporating unique molecular identifiers

89 (UMI) to limit PCR amplification bias and sequencing errors^{14,29–31}. Each method has potential
90 advantages and limitations^{27,32–35}. Specifically, DNA-based methods are believed to be more
91 quantitative and can be used in situations where RNA quality may not be guaranteed. In
92 contrast, RNA-based methods are believed to be more sensitive because of the presence of
93 multiple mRNA copies per cell, and also are more amenable to UMI incorporation³⁶. However,
94 the relative robustness and accuracy of the different approaches have not been systematically
95 compared. Here, we compared 9 different TCRseq library preparation protocols by analyzing
96 the TCR repertoire of aliquots of the same T-cell sample.

97

98 RESULTS

99 Experimental design to evaluate the robustness of human T-cell receptor repertoire analysis

100 We set out to compare 9 different academic or commercial protocols for library preparation
101 and sequencing (**Supplementary material and methods; Supplementary Table 1**) based either
102 on RACE-PCR (RACE-1 to RACE-6) or on multiplex-PCR (mPCR-1 to 3). We sequenced nucleic
103 acids from CD4⁺CD25⁻CD127⁺ effector T-cells (**Supplementary Fig.1a**) sorted from two healthy
104 donors (experiments A&B). In experiment A, we evaluated the accuracy and sensitivity of the
105 different methods by spiking donor A T-cell RNA (RACE-1 to RACE-6 and mPCR-3) or DNA
106 (mPCR-1 and mPCR-2) aliquots with different amounts of RNA or DNA from Jurkat cells
107 (**Supplementary Fig.1b**). In experiment B, we analyzed the impact of decreasing amounts of the
108 input material quantity by processing donor B RNA aliquots of 100 ng and 10 ng (**Supplementary**
109 **Fig.1c**). In both experiments, the CD4⁺CD25⁻CD127⁺ T-cells were sorted, and the RNA and DNA
110 were extracted and aliquoted in a single laboratory. Triplicates of aliquots were distributed to
111 service providers and academic laboratories. Raw and/or pre-filtered sequences data were all
112 processed using MiXCR³⁷.

113 We obtained from $5 \cdot 10^5$ to $2 \cdot 10^6$ reads per aliquot depending on the method (**Supplementary**
 114 **Fig.2a-b**). Numbers of unique V, J and VJ sequences as well as UMI distribution for RACE-1 and
 115 RACE-2 (**Supplementary Fig.2a-c**) were comparable between all the methods. Numbers of TCR
 116 sequences and clonotypes were correlated in a method-dependent manner, but not globally,
 117 suggesting that the sequencing depth required for a given number of clonotypes is method-
 118 dependent (**Supplementary Fig.2d**).

119

120 **Replicability and reproducibility differ among methods**

121 For each method, we first analyzed the proportion of reads that were identified as TCRs (**Fig.1a**
 122 **and Supplementary Fig.2**). For 7/9 methods, we observed 20 to 60% of non-aligned reads,
 123 which were mainly explained by no V and/or J sequence identification. TCR sequences had a
 124 high-quality score (phred > 30, **Fig.1b**) and contained less than 1% PCR errors (**Fig.1c**), except
 125 for RACE-2, RACE-6, mPCR-2 and mPCR-3. Note that these parameters could not be assessed

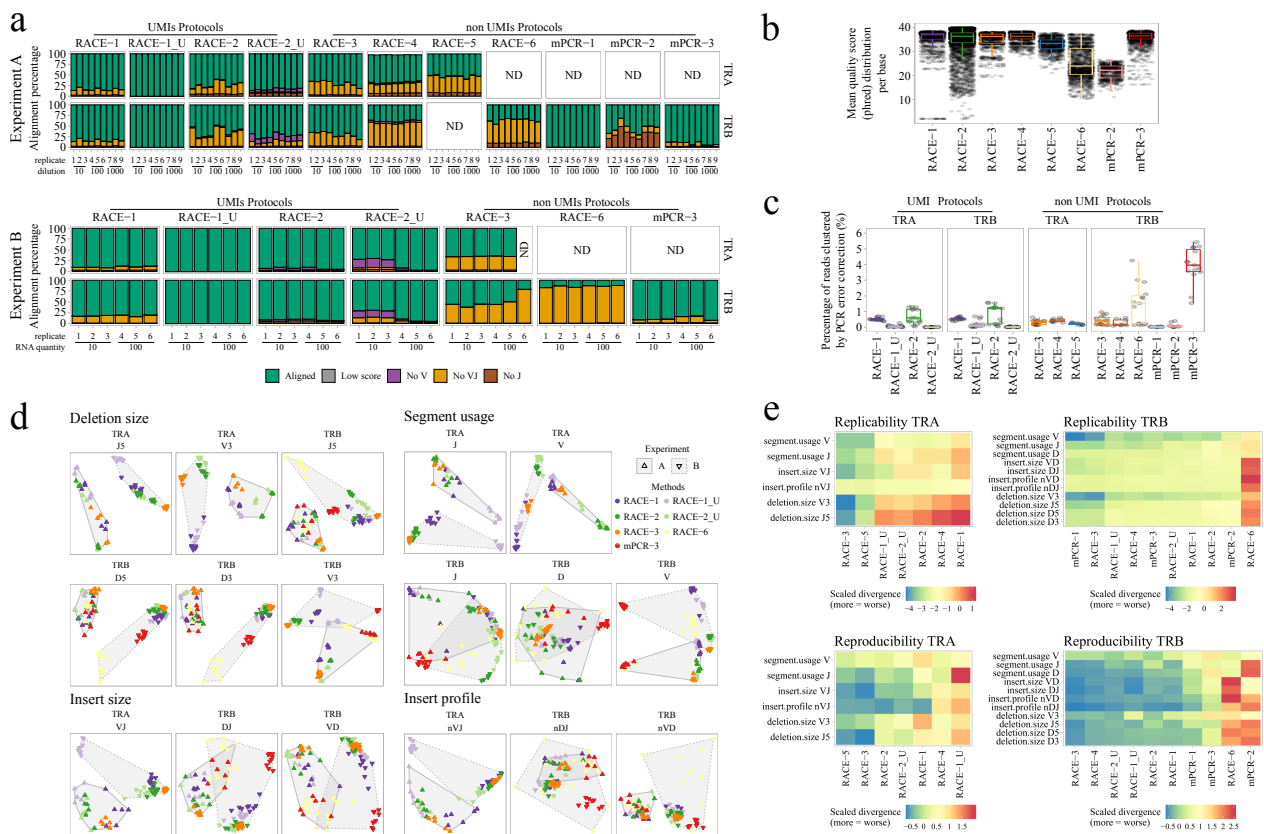


Fig. 1: Performance statistics and VDJ rearrangement model of each method for experiments A and B.

126 for one of the commercialized mPCR-1 for which undisclosed proprietary pre-processing of the
127 data is performed.

128 Using a VDJ rearrangement model (Methods), we computed 17 rearrangement parameters for
129 TRA and TRB sequences from experiments A&B (**Supplementary Fig.3**) and calculated Jensen-
130 Shannon Divergence (JSD) distances between samples per parameter. Multi-Dimensional
131 Scaling (MDS, **Fig.1d**) showed that, within each experiment, samples obtained with the same
132 method clustered together, suggesting that each method imposed its methodological imprint
133 on the repertoire profile.

134 We further compared the different library methods' replicability (i.e. the similarity among data
135 obtained with the same method) and reproducibility (i.e. the similarity among data obtained
136 with different methods) using JSD as a measure of the distance between datasets³⁸. **Figure 1e**
137 showed that for TRB, both the replicability and reproducibility of RACE-6 and mPCR-2 are lower
138 than for all the other methods tested. However, when considering TRA, replicability is higher
139 for RACE-3 and RACE-5 and reproducibility is higher for RACE-3, RACE-5 and RACE-2 (with and
140 without UMI). Since RACE-6 showed extremely low replicability for TRB samples and was not
141 reproduced by any other methods, we excluded it from further analysis. Altogether, our results
142 showed that many fundamental parameters of the TCR repertoire, as well as inter-sample
143 replicability and reproducibility, vary between the different methods tested.

144

145 **The observed TRBV gene usage varies between RACE- and multiplex-PCR RNA-based methods.**

146 We compared the TRBV usage obtained from the sequencing data with the percentage of TRBV
 147 protein expression quantified by flow cytometry (FC) (**Fig.2a and Supplementary Figs.4a-b**).
 148 mPCR-1 data were highly correlated with FC data (**Fig.2b**, $R^2 > 0.9$, $P < 5 \cdot 10^{-12}$), which likely
 149 reflects the undisclosed proprietary filtering by the provider. All other methods also showed a
 150 significant R^2 Pearson correlation score ranging from 0.4 to 0.8, $P < 0.05$) with TRBV protein
 151 expression (**Fig.2a-b**), except for mPCR-3 ($R^2 < 0.2$, $P > 0.05$). The Pearson correlation of TRBV
 152 gene usage within replicates prepared with the same method (**Fig.2c**) was high ($R^2 > 0.9$).
 153 However, clustering showed that mPCR-3 formed a distinct cluster with a low correlation score
 154 ($R^2 < 0.5$) with other methods. The RACE methods data were highly correlated between each
 155 other ($R^2 > 0.8$), except RACE-1 and RACE-1_U, which had a lower correlation ($0.6 < R^2 < 0.7$).
 156 mPCR-1 and mPCR-2 formed an independent “DNA cluster” with an $R^2 > 0.6$ when compared to
 157 RACE replicates and a low correlation with mPCR-3 ($R^2 < 0.4$). This low correlation with mPCR-3
 158 could in part be explained by a skewed TRBV9, TRBV29-1 and TRBV20-1 usage (**Supplementary**
 159 **Fig.4c**). Spearman correlation scores were higher between FC data and mPCR-3 as well as RACE-
 160 1, and globally between the methods (**Supplementary Fig.4d-e**). In summary, RACE-PCR

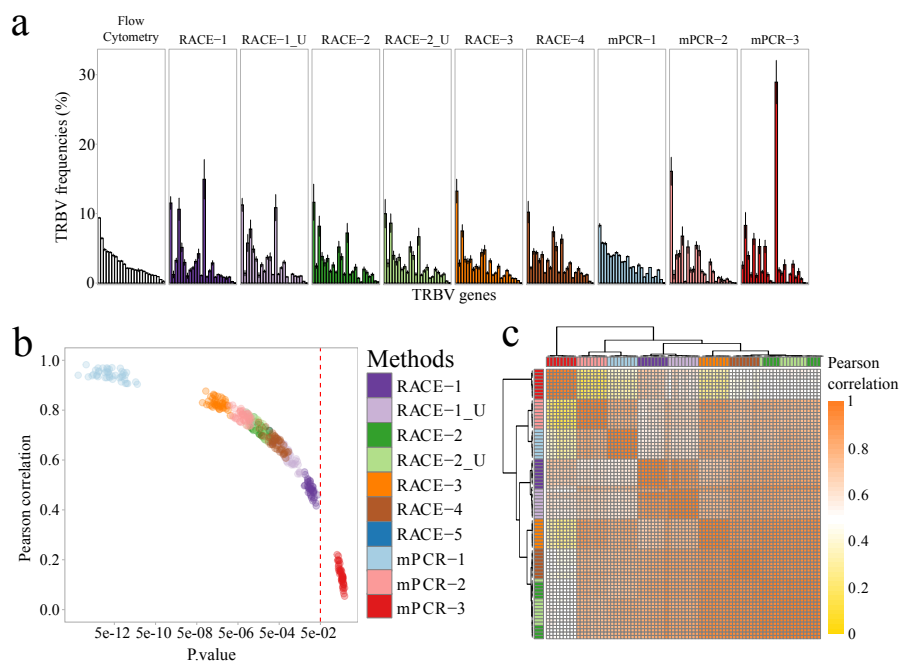


Fig. 2: TRBV usage comparison between flow cytometry and TCRseq.

161 methods and gDNA-based mPCR methods showed comparable TRBV usage results, in contrast
162 with the mPCR-3 RNA based method.

163

164 **Robustness of TRA and TRB detection is method-dependent**

165 We compared the similarity and composition of the 1% most predominant clonotypes
166 (1%_MPC) detected by each method. The Morisita-Horn similarity index (MH) was calculated
167 for each replicate across all the methods for both TRA (**Fig.3a-left**) and TRB sequences (**Fig.3a-**
168 **right**). TRA repertoires from RACE-3 and RACE-5 clustered together, inter- and intra-replicates
169 having a high degree of similarity ($MH \approx 0.8$). RACE-1, RACE-2 and RACE-4 have a lower inter-
170 and intra-method similarity ($0.2 < MH < 0.5$), but a higher similarity with RACE-3 and RACE-5.
171 Comparable clustering was obtained with the Jaccard similarity index (JSI), a measure
172 independent of clonotype frequency (**Supplementary Fig.5a**). For the TRB repertoires, MH
173 scores were low when comparing RACE and mPCR protocols ($MH \approx 0.36$), but high within the
174 RACE cluster ($0.6 > MH > 0.9$). There was poor similarity between the results of the three mPCR
175 methods, regardless of the template. Differences between RACE and mPCR methods
176 disappeared when calculating the JSI, suggesting a bias in clonotype frequency, as expected
177 when comparing RNA- with DNA-based methods, but less when comparing RNA-based
178 methods. Similar results were obtained by iteratively increasing the percentage of clonotypes
179 (**Supplementary Fig.5b**). Rényi diversity profiles (**Supplementary Fig.5c**) showed comparable
180 results for TRB with all the methods, but the diversity of TRA varied depending on the method.
181 However, the potential diversity estimated using Chao extrapolation was variable between
182 methods (**Supplementary Fig.5d**).

183 To test a possible bias in capturing the TRA diversity for some methods, we pooled and
184 compared the three spiking replicates per method from experiment A, as suggested by Greiff

185 et al.²⁴. The MH similarity significantly increased for all the RACE-based methods for TRA
 186 (Fig.3b-top) (except RACE-3) and for TRB (Fig.3b-bottom), with the TRA MH similarity remaining
 187 lower than that of TRB. Similar observations were made for mPCR replicates. This suggests that
 188 for a given depth of sequencing, the TRB diversity is better captured than that of TRA.

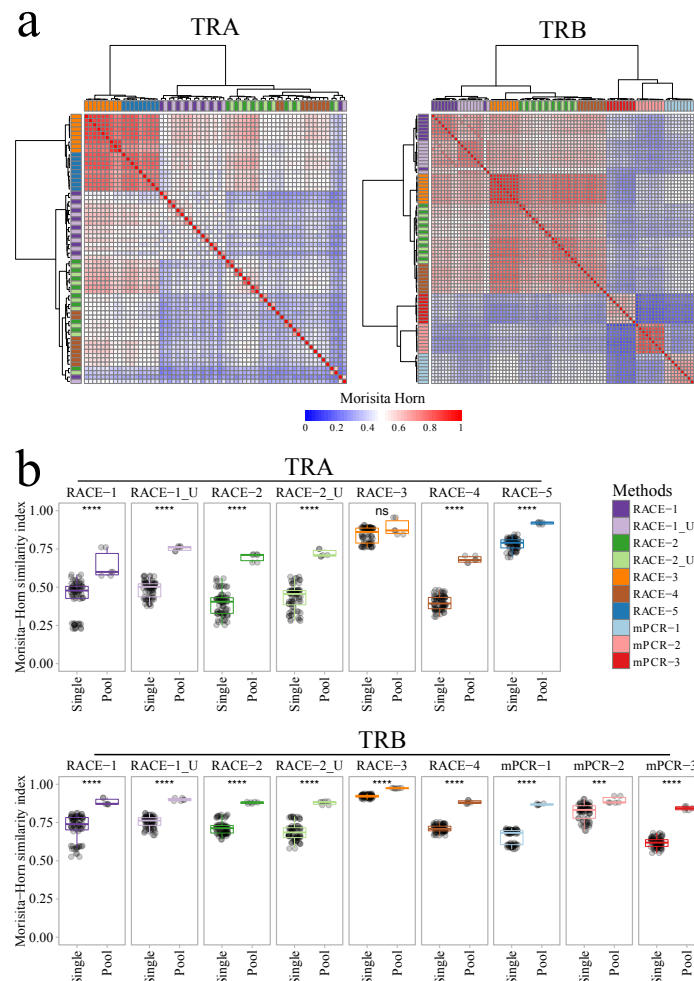


Fig. 3: The reproducibility of detection of major TCR clonotypes by different methods.

189

190 **Detection sensitivity of rare TCRs depends on the method**

191 To determine the accuracy of the different library amplifications for different clonotype
 192 frequencies, we compared the observed frequencies of the TCR from the Jurkat spike-in to their
 193 theoretical frequencies of 1/10, 1/100 and 1/1000. (Supplementary Fig.1b). TRA observed
 194 frequencies were on average 3 times lower than expected (Fig.4a-top; Supplementary Table 2

195 and Supplementary Fig.6a). In contrast, TRB frequencies were on average 3 times higher than
 196 the theoretical percentage, except for mPCR-1 (Fig.4a-bottom; Supplementary Table 2 and
 197 Supplementary Fig.6a). For most of the methods, except RACE-1_U, RACE-4 and mPCR-3, the
 198 ratio between the different dilutions was maintained, as shown by the mean slope values close
 199 to 1 (Fig.4b).

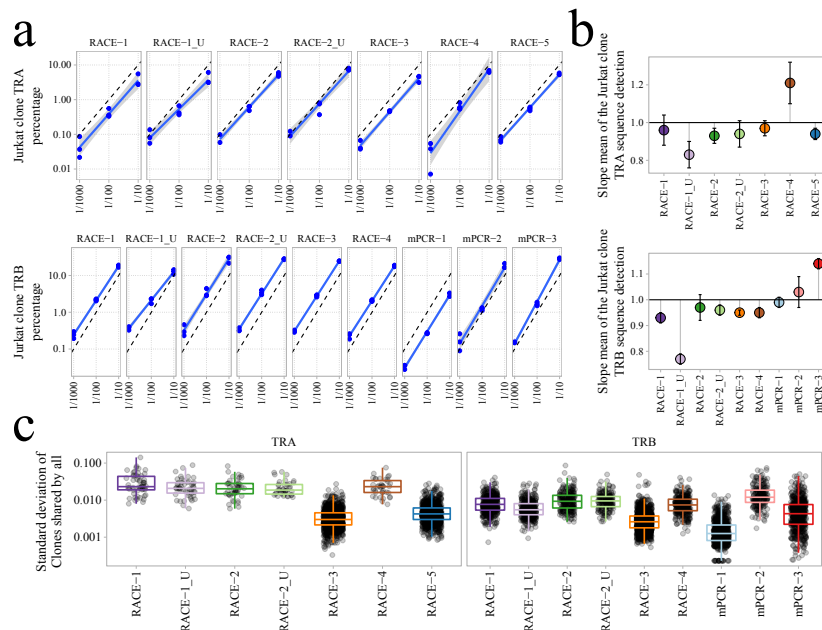


Fig. 4: Sensitivity of TCR sequence detection by different methods.

200 We then compared the inter-sample variation in clonal frequency for those TCR sequences
 201 shared between all replicates of an individual method (excluding the Jurkat clone). **Figure 4c**
 202 represents the standard deviation of the frequency of each shared clonotype (dots) per method
 203 (see details in Supplementary Fig.6b-d). For TRA, RACE-3 and RACE-5 had the highest number
 204 of clonotypes shared between the 9 replicates and the lowest standard deviation. For TRB, all
 205 the methods captured a high number of shared clonotypes, and mPCR-1 and RACE-3 had the
 206 lowest standard deviation. Finally, pooling all the clonotypes from all the replicates, we
 207 identified 9 TRA and 31 TRB clonotypes shared by all the replicates of all methods,
 208 corresponding to the most predominant clonotypes (Supplementary Fig.7). RACE-3, RACE-5
 209 (both RNA-based) and mPCR-1 (DNA_based) showed the lowest inter-sample variability in TCR
 210 frequency.

211

212 **The quantity of starting material impacts TCR diversity capture**

213 One major limitation when analyzing TCR repertoire is the number of T-cells that can be
214 analyzed. Focusing on 4 RNA-based methods, we analyzed the influence of input RNA quantity
215 on TRA and TRB repertoires (**Supplementary Fig.1c**). We compared two sets of samples, one
216 containing 10 ng or 100 ng (corresponding to 10^4 and 10^5 cells, respectively). For all the
217 methods, the richness was higher with large (100 ng) than small (10 ng) samples
218 (**Supplementary Fig.8a**). Rényi diversity profiles (**Supplementary Fig.8b**) showed that when
219 $\alpha < 2$ (i.e. when the diversity metric is influenced by rare clones), the diversity of small
220 samples is less than that of larger ones. In contrast, at $\alpha = 2$ (Simpson index) or above,
221 diversity profiles of both samples overlap. Thus, a low RNA input influences the number of rare
222 TCR sequences detected, but not the distribution of the more abundant TCRs.

223 Finally, we evaluated the inter-sample similarity as a function of RNA input quantity by
224 calculating the MH index with either the TRVJ combination usage (VJ_usage), all clonotype
225 frequencies (Overall), or with the frequencies of the 1% most predominant clonotype
226 (1%_MPC) (**Supplementary Fig.8c-middle**). For TRA, the similarity between 10 ng replicates was
227 lower at the level of VJ usage and of all clonotypes compared with that between 100 ng
228 replicates (**Supplementary Fig.8c-top&bottom**). For TRB, the results were comparable
229 regardless of the quantity (MH>0.5). When focusing on the 1% MPC, the similarity was
230 comparable regardless of the quantity for both TRA and TRB. These results indicated that RNA
231 quantity impacts rare clonotype detection.

232

233 Reliability and sensitivity of each method highlighted using an in silico meta-repertoire

234 One unavoidable issue when aiming at capturing the diversity of a repertoire is sampling, i.e.
235 only a fraction of the cells are analyzed and then a fraction of their nucleic acids²⁴. To better
236 assess the ability of each method robustly to capture rare and frequent clonotypes, we took
237 advantage of the fact that altogether we generated 45 TRA and 63 TRB replicates of the same
238 cell sample. We aggregated these results to generate an in silico meta-repertoire. To ensure
239 the accuracy of the TCR sequences composing this meta-repertoire, we removed singletons
240 and kept clonotypes found by at least 3 methods.

241 We first analyzed how many of the clonotypes present in this meta-repertoire were detected
242 by each method. For TRA (**Fig.5a-left**), RACE-3 and RACE-5 datasets included up to 50% of the
243 meta-repertoire clonotypes (MRC) compared to 10 to 20% for the other RACE method datasets.
244 Similar results were found for TRB (**Fig.5a-right**). We then computed for each method the
245 fraction of MRC found in 0, 1, 2, 3 etc. up to 9 replicates. The dot-heatmaps (**Fig.5b**) showed
246 that for TRA, RACE-3 and RACE-5 clearly outperformed the other methods, capturing up to 40%
247 of the MRC in all 9 replicates (**Fig.5b-left**; Replicate number=9) and missing (i.e. never captured
248 in any of the 9 replicates) less than 1% of the MRC (**Fig.5b-left**; Replicate number=0). The other
249 RACE protocols detected only 1% of MRC in all 9 replicates and missed 15 to 20% of the MRC

250 (Fig.5b-left). In contrast, there was much less difference between the methods for TRB (Fig.5b-
 251 right).

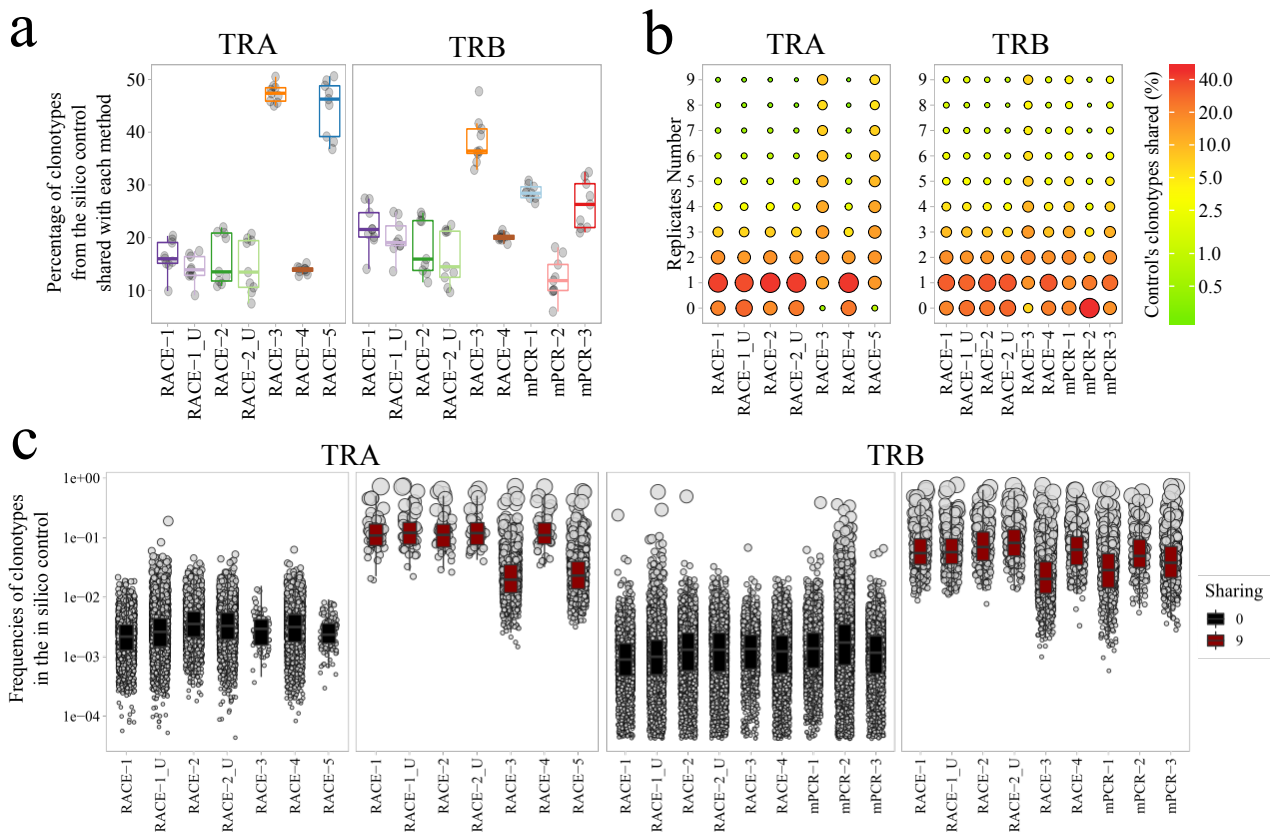


Fig. 5: Sharing with robust and representative meta-repertoire.

252 Finally, we analyzed the frequency of MRC TCRs that were detected or not by each method
 253 (Fig.5c and Supplementary Fig.9). For TRA (Fig.5c-left), the frequency of MRC found in 9
 254 replicates (red boxplots) ranged from 1% to 0.001% for RACE-3 and RACE-5 and from 1% to
 255 0.05% for the other methods. In contrast, clonotypes not detected in any replicates (black
 256 boxplots) were present at 10- to 100-fold lower abundance. A similar overall pattern was seen
 257 for TRB, although the frequencies were shifted to a lower range. This analysis suggested that
 258 RACE-3 and RACE-5 had increased sensitivity, and hence were able to detect a larger proportion
 259 of clonotypes at lower abundances. These differences were more evident for TRA than for TRB
 260 (Fig.5c-right). The other methods compared behaved very similarly to each other. Importantly,
 261 those results were independent of sample size (Supplementary Fig.10).

262

263 DISCUSSION

264 Interpreting the TCR repertoire is an increasingly important tool in understanding the
265 underlying causes of immune-mediated diseases and in assisting the development of new
266 immunotherapeutic strategies. However, despite hundreds of TCRseq studies in the last decade
267 using a variety of different methodologies, there has been no systematic study comparing
268 them.

269 In this work, we compared methods developed by academics, at a time when there was little
270 or no reliable commercial service provision, with some currently available commercial
271 methods. Both RNA- and gDNA-based methods were included. To avoid mis-implementation of
272 protocols, each method (including appropriate pre-processing of sequence data) was
273 performed by the laboratory or commercial provider (except for kit providers) that developed
274 them.

275 Unexpectedly, some consistent differences were observed in TRBV usage when compared to
276 FC measurement of TRBV-encoded proteins, especially for RNA-based profiling. This might
277 reflect bias in amplification of RNA transcripts according to their expression levels, more
278 efficient transcription of some V genes, or differences in nonsense-mediated decay³⁹. Further
279 studies, using single-cell RNAseq may shed light on this phenomenon.

280 Working with human samples often imposes limits on the number of available T-cells. Notably,
281 lymphopenia is a common feature in people undergoing treatment (transplantation,
282 immunosuppressive therapy) or with autoimmune disease⁴⁰ and infections. Additionally, T-cell
283 subsets of interest, as well as available counts of tumor-infiltrating T-cells, may be limited.
284 Therefore, it is important to identify which methods provide reliable TCRseq profiles for small
285 numbers of T-cells. In this context, we observed that, regardless of the method, starting from

286 a highly polyclonal population, the initial amount of material is critical to obtaining
287 representative results, notably in terms of diversity and rare clone detection.

288 Although our study focused on polyclonal CD4 T-cells from healthy repertoires, we analyzed a
289 wide range of global and sequence-specific repertoire parameters, including V(D)J gene usage,
290 junctional diversity, repertoire diversity and sequence sharing. These parameters are all
291 relevant to any other alpha/beta T-cell populations, as indeed are all parameters routinely used
292 to analyze repertoires of samples from pathological and clinical human samples⁴¹.

293 Because our study incorporated multiple replicates tested with each method, we were able to
294 explore method replicability, i.e. the ability of each method to reproduce the same repertoire
295 from different sub-samples from the same individual. Our results showed that, except mPCR-
296 3, all the methods provided consistent results among replicates. We also evaluated the
297 reproducibility, i.e. the extent to which different methods record the same results when applied
298 to the same sample. We observed a low degree of TRB clonotype overlap between repertoires
299 amplified from gDNA and RNA (cDNA), perhaps reflecting differences in gDNA and RNA copy
300 numbers. The four RACE methods produced relatively similar repertoires as revealed by the
301 Morisita-Horn index. The mPCR on gDNA showed low reproducibility between methods,
302 suggesting that the choice of multiplexing primers might bias the amplification of some
303 clonotypes, as suggested previously³⁴. However, most RACE methods (not tested for mPCR)
304 had a lower efficiency in capturing TRA rather than TRB diversity, which may reflect the 2- to 3-
305 fold lower number of TRA transcripts than TRB transcripts³¹.

306 Finally, sensitivity is important for the study of circulating blood T-cells, especially when the
307 goal is to track a few expanded clones associated with infection or autoimmunity, or in
308 response to treatment. However, assessing sensitivity based on sample overlap is a complex
309 performance metric, since it is impacted by experimental variability, but also by sampling. In

310 order to tackle this problem directly, we generated an in silico meta-repertoire which provided
311 a more robust platform with which to directly compare the sensitivity performance of the
312 different methods. Interestingly, using this standard, we found that two non-UMI methods
313 (RACE-3 and RACE-5) had greater sensitivity than UMI-based methods (RACE-1 and RACE-2)
314 and were able to detect clonotypes at a 10-fold lower frequency. In part, this results from the
315 reads-per-UMI cutoff, which may lead to a decrease in observed TCR diversity if sequencing
316 coverage is not sufficient. For example, introducing a hard cutoff which discards all UMIs with
317 less than 5 reads leads to a decrease in observed TCR diversity. UMI-based methods may be
318 more accurate for assessing clonotype frequency, in line with their use to quantify and correct
319 for PCR errors and bias^{42,43}. Furthermore, a threshold of 2-4 reads per UMI efficiently protects
320 against artefacts and cross-sample contamination⁴⁴, which becomes critical with tighter cluster
321 density on modern Illumina machines. UMI-based methods may require several replicates or
322 higher sequencing coverage to consistently and unambiguously identify rare TCR sequence
323 clonotypes. Noteworthy, both RACE-1 and RACE-2 methods performed better after UMI
324 correction (see Table 1).

325 Such in silico standards may be of value in further comparative TCRseq method evaluation,
326 although ideally synthetic repertoires recapitulating at least the extent of the TRAVJ and TRBVJ
327 combinations and distributions may provide an even more robust alternative. Two such
328 approaches have been proposed for specific clone detection in Minimal Residual Diseases^{45,46}
329 as well as for the BCR, but not TCR, repertoire⁴⁷, still at a very low diversity level. The
330 construction of such gold standard repertoires is currently very costly and remains a major
331 challenge that the Adaptive Immune Receptor Repertoire Community (AIRR-C)⁴⁸, engaged in
332 AIRR-seq standardization⁴⁹⁻⁵¹, may tackle in the future. Finally, in this study some data were
333 pre-processed using proprietary (mPCR-1, mPCR-3) or published^{30,52} (RACE-1_U and RACE-2_U)

334 tools and then aligned and error-corrected using MiXCR (v2.1.10)³⁷. To further optimize TCR
335 data accuracy, it would also be interesting to benchmark available software analysis tools,
336 especially regarding UMI analysis and sequence alignment. Our datasets generated using
337 different methods should be a valuable complement to using datasets generated purely in
338 vitro^{53,54}.

339 In conclusion, the take-home messages from this work are the following. Firstly, there are
340 satisfactory TCRseq methods based on either DNA or RNA input, and in both cases the amount
341 of material impacts both diversity and the detection of rare clones. Secondly, various methods
342 are optimal for detecting maximal diversity, while others most accurately quantify the
343 abundance of specific clonotypes. For the latter, UMI-based methods are potentially more
344 accurate, although they could miss relevant but rare clones. In contrast, non-UMI RACE
345 methods are more sensitive in capturing rare clones, especially for TRA. Thirdly, the availability
346 of raw data is crucial in allowing reliable and reproducible in-depth analyses of TCR repertoires;
347 the mPCR-1 service provider does not provide access to raw sequence data, while mPCR-1 and
348 mPCR-3 do not disclose the proprietary pre-processing filters. In contrast, the RACE-2 provider
349 provides raw data and all preprocessing algorithms. We summarized our results as well as
350 practical aspects in Table 1. Regarding the results, we calculated for each method a rank value
351 for Replicability, reliability and sensitivity based on various measures (**Table 1** and
352 **Supplementary file**). We also summarized cost per sample, presence of controls or standards,
353 format of the method and raw data availability. The Table 1 highlight the advantages and
354 disadvantages of the different methods which could serve as guidance for end-users. Improved
355 and more sophisticated data analyses are essential to extract the full power of TCR repertoire
356 data. We anticipate that now that TCR sequencing has come of age, the next key developments

357 in the field will come from novel methods of data analysis, as has been the case in the related
 358 field of global transcriptomics.

| TR chain | Method | Replicability | Reliability | Sensitivity | Cost per sample | Controls & standards | Format type | fastq data availability |
|----------|----------|---------------|-------------|-------------|-----------------|----------------------|----------------|-------------------------|
| TRA | RACE-1 | 7 | 4 | 4 | ~230 | - | lab protocol | YES |
| | RACE-1_U | 4 | 5 | 4 | ~230 | UMI | lab protocol | YES |
| | RACE-2 | 5 | 4 | 5 | 230-280 | - | service or kit | YES |
| | RACE-2_U | 4 | 5 | 5 | 230-280 | UMI | service or kit | YES |
| | RACE-3 | 3 | 2 | 3 | ~150 | - | kit | YES |
| | RACE-4 | 5 | 6 | 4 | ~150 | - | lab protocol | YES |
| TRB | RACE-5 | 2 | 3 | 3 | ~300 | - | lab protocol | YES |
| | mPCR-1 | 3 | 3 | 3 | ~350-550* | synthetic TCRs | service or kit | NO |
| | mPCR-2 | 6 | 7 | 7 | ~230 | - | lab protocol | YES |
| | mPCR-3 | 5 | 5 | 3 | ~350-550* | - | service or kit | YES |
| | RACE-1 | 6 | 5 | 4 | ~230 | - | lab protocol | YES |
| | RACE-1_U | 4 | 6 | 5 | ~230 | UMI | lab protocol | YES |
| | RACE-2 | 6 | 6 | 6 | 230-280 | - | service or kit | YES |
| | RACE-2_U | 6 | 6 | 7 | 230-280 | UMI | service or kit | YES |
| | RACE-3 | 2 | 2 | 3 | ~150 | - | kit | YES |
| RACE-4 | 3 | 5 | 4 | ~150 | - | lab protocol | YES | |

359 **Table 1: Comparative performance of the nine TCRseq molecular methods.**

360 MATERIAL AND METHODS

361 *Blood effector T cell isolation*

362 Peripheral blood mononuclear cells (PBMC) from two healthy blood donors (Etablissement
363 Français du sang; French Blood Center) were obtained with written informed consent for
364 biomedical research. The experiments carried out were in conformity with the Helsinki
365 Declaration on Biomedical Research. Donors A (experiment A) and B (experiment B) were both
366 men, 36 and 54 years old, respectively. CD3⁺CD4⁺CD127⁺CD25⁻ cells (CD4⁺ T effector cells) were
367 sorted at the Sorbonne Université laboratory as follows: CD4⁺ cells were isolated by
368 Lymphoprep (Stemcell®) density gradient and positive selection using the Dynabeads™ CD4
369 Positive Isolation Kit (Invitrogen®). Enriched CD4⁺ T-cells were then labeled with anti-CD3⁺,
370 CD4⁺, CD127⁺ and CD25⁺ antibodies and effector T-cells were sorted on a FACS ARIA II with a
371 purity > 95% (**Supplementary Fig.1a**).

372

373 *Jurkat cell culture*

374 The Jurkat cell line with a known TCR (TRAV8-4-CAVSDLEPNSSASKIIF-TRAJ3; TRBV12-3-
375 CASSFSTCSANYGYTF-TRBJ1-2) (clone E6-1), from ATCC, was grown in 5% CO₂, in RPMI 1640
376 medium, supplemented with 10% (v/v) fetal bovine serum (FBS), 2 mM L-glutamine, 50 U/mL
377 penicillin, and 50 µg/mL streptomycin at the Sorbonne Université laboratory.

378

379 *RNA and DNA extraction*

380 In experiment A, DNA and RNA were both extracted using TRIzol Reagent (Invitrogen®) from 5
381 million Jurkat cells and 20 million CD4⁺ T effector cells and, in experiment B, only RNA was
382 extracted using the RNAqueous-Kit (Invitrogen®) from 7.2 million CD4⁺ T effector cells following
383 the manufacturer's recommendations. DNA concentration and RNA concentration were

384 measured on a NanoDrop1000 (Thermo Scientific™) and RNA integrity was determined on a
385 Bioanalyzer (Agilent®) with measurements higher than 8. RNA and DNA extraction and
386 validation were performed at the Sorbonne Université laboratory.

387

388 *Aliquot preparation for method comparison*

389 In experiment A, 100 ng of RNA or DNA from the CD4⁺ effector T-cells sorted from donor A was
390 split into 3 aliquots that were spiked with different amounts of RNA or DNA from the Jurkat cell
391 line, at ratios of 1/10, 1/100 and 1/1000. Each spiked aliquot was further split into 3 and all
392 replicates were processed by all methods tested (7 for RNA and 2 for DNA; **Supplementary**
393 **Fig.1b**). With experiment B, we analyzed the impact of the input material quantity. RNA from
394 sorted CD4⁺ effector T-cells of donor B was extracted, split into 15 aliquots of 100 ng each and
395 15 aliquots of 10 ng each and processed in triplicate using 5 of the RNA-based methods
396 (**Supplementary Fig.1c**). Aliquots were prepared at the Sorbonne Université laboratory and sent
397 to the partners.

398

399 *Flow Cytometry*

400 V β identification was performed on enriched CD4⁺ effector T-cells from experiment A (see
401 *Blood effector T cell isolation* for enrichment procedure) stained with the IOtest Beta Mark TR
402 Repertoire Kit (Beckman Coulter®) according to the manufacturer's protocol as well as with
403 CD4-APC, CD127-BV421, CD25-PECy7. Data acquisition was performed on a Cytoflex®
404 (Beckman Coulter®) using CytExpert® software. FlowJo® was used for data analysis. V β
405 frequencies were calculated on CD4⁺CD25⁻CD127⁺ gated cells (**Supplementary Fig.4a-b**).
406 Staining was performed at the Sorbonne Université laboratory.

407

408 ***TCR library preparation and sequencing***

409 The nine protocols for TCR library preparation compared in this study were selected according
410 to at least one the following criteria: published use by groups other than the one who
411 developed it (mPCR-1, mPCR-3, RACE-1, RACE-2, RACE-4 and RACE-5), (ii) their association with
412 well-known analysis tools (RACE-1, RACE-2, mPCR-2) and (iii) commercially available (RACE-2,
413 RACE-3, mPCR-1, mPCR-3). Sequencing protocols were harmonized taking into account
414 published recommendations or recommendations provided by the manufacturer of
415 commercial kits or by the owner or users of the protocol. All protocols are detailed in
416 **Supplementary material and methods.**

417 ***TCR deep sequencing data processing***

418 FASTQ raw data files were obtained from each method, except for Multiplex-1 & 2, for which
419 we obtained, respectively, FASTA file and FASTQ files following proprietary pre-processing. For
420 RACE-1 and RACE-2, UMI pre-processing was performed following protocols published
421 elsewhere^{30,31,52}. FASTQ and FASTA files were then processed for TRB and TRA sequence
422 annotation using the MiXCR software³⁷ (v2.1.10) with RNA-Seq default parameters (*-p rna-seq*
423 *-s hsa*) as available online. MiXCR extracts TRA and TRB repertoire providing correction of PCR
424 and sequencing errors.

425

426 ***Data analysis***

427 Statistical comparisons and multivariate analyses were performed using R software version
428 3.5.0 (www.r-project.org). We used the ggplot2 package to generate figures⁵⁵, except
429 heatmaps. More complex analyses are detailed in the next section.

430

431 ***Comparing VDJ rearrangement statistics***

432 An empirical VDJ rearrangement model for each method was built as follows. We analyzed
433 clonotype tables to obtain comprehensive statistics of VDJ rearrangements including the
434 frequencies of V/D/J segment usage, number of added N Bases (namely “insert profile”, i.e. the
435 probability distribution of having A/T/G/C inserted in the N-region of CDR3 given that we
436 observe a certain base inserted before it) and V/J segment trimming bases, with the IGoR
437 package⁵⁶. This model is built in a 'greedy' way in the sense that it uses best alignments provided
438 by MiXCR rather than running expectation maximization procedures as described in Murugan
439 et al.⁵⁷. We utilized the Jensen-Shannon divergence (JSD) between distributions of VDJ usage
440 to define the following two statistics that we use for comparative analysis of different TCRseq
441 methods: 1) *replicability* measured as the distance between different samples produced by the
442 same protocol and 2) *reproducibility* measured as the distance between samples produced by
443 two different protocols. MDS used for sample mapping was performed on rank-transformed
444 distances to avoid the distorting effect of outliers. All the analyses involve VDJ usage inferred
445 from weighted data (TCR clonotype is weighted by its frequency in the sample) to account for
446 TCRseq method amplification biases.

447

448 ***Similarity analysis***

449 Pearson and Spearman correlations, the Morisita-Horn index⁵⁸ (MH) and the Jaccard similarity
450 index⁵⁹ (JSI) were used to assess the similarity between samples. The MH index takes into
451 account the relative abundance of species in the sample, while the JSI is a measure of the
452 intersection between two populations relative to the size of their union, and is independent of
453 relative abundances. Both indices vary between 0 (no overlap) and 1 (perfect overlap). JSI and
454 MH were calculated using the DIVO package⁶⁰ on R. In order to discriminate indices represented

455 by a heatmap with the pheatmap package⁶¹, we used a different set of colors. The Pearson and
456 Spearman correlations are presented as yellow/white/orange (Fig.2c and Supplementary
457 Fig.4e), MH is presented as blue/white/red (Fig.3a) and JSI is presented as purple/yellow/green
458 (Supplementary Fig.5a).

459

460 *Diversity profiling*

461 The diversity was analyzed using two indices. Rényi entropy⁶² is a generalization of Shannon
462 entropy, which increases when both species richness and evenness are high. Rényi entropy is
463 a function of a parameter α spanning from (i) the species richness ($\alpha=0$), which corresponds to
464 the number of clonotypes regardless of their abundance, to (ii) the clonal dominance ($\alpha \rightarrow +\infty$),
465 corresponding to the frequency of the most predominant clonotype. For $\alpha=1$, the Shannon
466 diversity index is computed. The exponential of the Rényi entropy corresponds to the actual
467 number of clonotypes in the datasets⁶³ and is used to build a diversity profile⁶⁴. It was
468 computed using the entropy package⁶⁵ on R. ChaoE⁶⁶ index was calculated with the iNEXT
469 package⁶⁷ as a measure of extrapolation of the possible number of clonotypes based on the
470 observed clonotypes. Rarefaction curves were interpolated from 0 to the current sample size
471 and then extrapolated to the size of the largest of samples, allowing comparison of diversity
472 estimates. Interpolation and extrapolation were based on ChaoE multinomial models⁶⁸.

473

474 *Meta-repertoire construction*

475 We generated an in silico meta-repertoire from the sequences obtained from the 108
476 replicates (45 for TRA and 63 for TRB). This meta-repertoire, for each chain, was designed to
477 minimize biases by (i) pooling all clonotypes from the 9 datasets and removed singletons to
478 avoid introducing noise due to PCR errors, (ii) Selecting non-reprocessed datasets, meaning

479 before UMI, (iii) keeping only clonotypes found by at least 3 different methods to avoid bias
480 toward one particular method. The threshold was defined to reach a dataset size as close as
481 possible to the original datasets to avoid additional sampling, (iv) normalizing the size of each
482 dataset to the lowest dataset to ensure the same weighting for each method. Completion of
483 the representative meta-repertoire was achieved by pooling all the datasets. This generated a
484 pooled dataset of 14 458 TRA and 18 735 TRB clonotypes.

485

486 **Data Availability**

487 All the fastq data obtained in this study, including the Jurkat Clone E6-1 (ATCC®TIB-152™) cell
488 line TCR alpha and beta sequences, were deposited in the NCBI Sequence Read Archive
489 repository following MiAIRR standard recommendations⁴⁷ under the BioProject ID
490 PRJNA548335. Source data for TCRVb flow cytometry data are provided as **Supplementary**
491 **Fig.4a-b**. All other data are available from the corresponding author upon request.

492

493 **Code Availability**

494 All software packages and programs are publicly available and open source. Scripts used to
495 analyze the data with MiXCR are available from <https://mixcr.milaboratory.com> ; Decombinator
496 from <https://github.com/innate2adaptive/Decombinator>; MiGEC from
497 <https://github.com/mikessh/migec>; detailed VDJ rearrangement statistics scripts are available
498 from <https://github.com/antigenomics/repseq-protocol-comparison>. There is no restriction on
499 the use of the code or data.

500 REFERENCES

- 501 1. Chien, Y. H., Gascoigne, N. R., Kavaler, J., Lee, N. E. & Davis, M. M. Somatic recombination
502 in a murine T-cell receptor gene. *Nature* **309**, 322–326 (1984).
- 503 2. Davis, M. M. & Bjorkman, P. J. T-cell antigen receptor genes and T-cell recognition. *Nature*
504 **334**, 395–402 (1988).
- 505 3. Lefranc, M.-P. Nomenclature of the Human T Cell Receptor Genes. *Curr. Protoc. Immunol.*
506 **40**, A.10.1-A.10.23 (2000).
- 507 4. Dupic, T., Marcou, Q., Walczak, A. M. & Mora, T. Genesis of the $\alpha\beta$ T-cell receptor. *PLOS*
508 *Comput. Biol.* **15**, e1006874 (2019).
- 509 5. Robins, H. S. *et al.* Comprehensive assessment of T-cell receptor β -chain diversity in $\alpha\beta$ T
510 cells. *Blood* **114**, 4099–4107 (2009).
- 511 6. Warren, R. L. *et al.* Exhaustive T-cell repertoire sequencing of human peripheral blood
512 samples reveals signatures of antigen selection and a directly measured repertoire size of
513 at least 1 million clonotypes. *Genome Res.* **21**, 790–797 (2011).
- 514 7. Qi, Q. *et al.* Diversity and clonal selection in the human T-cell repertoire. *Proc. Natl. Acad.*
515 *Sci.* **111**, 13139–13144 (2014).
- 516 8. Cui, J.-H. *et al.* TCR Repertoire as a Novel Indicator for Immune Monitoring and Prognosis
517 Assessment of Patients With Cervical Cancer. *Front. Immunol.* **9**, (2018).
- 518 9. Davis, M. M. The $\alpha\beta$ T Cell Repertoire Comes into Focus. *Immunity* **27**, 179–180 (2007).
- 519 10. Lindau, P. & Robins, H. S. Advances and applications of immune receptor sequencing in
520 systems immunology. *Curr. Opin. Syst. Biol.* **1**, 62–68 (2017).
- 521 11. Miles, J. J., Douek, D. C. & Price, D. A. Bias in the $[\alpha][\beta]$ T-cell repertoire:
522 implications for disease pathogenesis and vaccination. *Immunol. Cell Biol.* **89**, 375 (2011).

- 523 12. Schrama, D., Ritter, C. & Becker, J. C. T cell receptor repertoire usage in cancer as a
524 surrogate marker for immune responses. *Semin. Immunopathol.* **39**, 255–268 (2017).
- 525 13. Aoki, H. *et al.* TCR Repertoire Analysis Reveals Mobilization of Novel CD8+ T Cell Clones
526 Into the Cancer-Immunity Cycle Following Anti-CD4 Antibody Administration. *Front.*
527 *Immunol.* **9**, (2019).
- 528 14. Heather, J. M. *et al.* Dynamic Perturbations of the T-Cell Receptor Repertoire in Chronic
529 HIV Infection and following Antiretroviral Therapy. *Front. Immunol.* **6**, (2016).
- 530 15. Howson, L. J. *et al.* MAIT cell clonal expansion and TCR repertoire shaping in human
531 volunteers challenged with Salmonella Paratyphi A. *Nat. Commun.* **9**, 253 (2018).
- 532 16. Pogorelyy, M. V. *et al.* Precise tracking of vaccine-responding T cell clones reveals
533 convergent and personalized response in identical twins. *Proc. Natl. Acad. Sci. U. S. A.*
534 **115**, 12704–12709 (2018).
- 535 17. Qi, Q. *et al.* Diversification of the antigen-specific T cell receptor repertoire after varicella
536 zoster vaccination. *Sci. Transl. Med.* **8**, 332ra46–332ra46 (2016).
- 537 18. Sycheva, A. L. *et al.* Quantitative profiling reveals minor changes of T cell receptor
538 repertoire in response to subunit inactivated influenza vaccine. *Vaccine* **36**, 1599–1605
539 (2018).
- 540 19. Hogan, S. A. *et al.* Peripheral Blood TCR Repertoire Profiling May Facilitate Patient
541 Stratification for Immunotherapy against Melanoma. *Cancer Immunol. Res.* **7**, 77–85
542 (2019).
- 543 20. Hopkins, A. C. *et al.* T cell receptor repertoire features associated with survival in
544 immunotherapy-treated pancreatic ductal adenocarcinoma.
545 <https://insight.jci.org/articles/view/122092/pdf> (2018) doi:10.1172/jci.insight.122092.

- 546 21. Jin, Y. *et al.* TCR repertoire profiling of tumors, adjacent normal tissues, and peripheral
547 blood predicts survival in nasopharyngeal carcinoma. *Cancer Immunol. Immunother.* **67**,
548 1719–1730 (2018).
- 549 22. Wieland, A. *et al.* T cell receptor sequencing of activated CD8 T cells in the blood identifies
550 tumor-infiltrating clones that expand after PD-1 therapy and radiation in a melanoma
551 patient. *Cancer Immunol. Immunother.* **67**, 1767–1776 (2018).
- 552 23. Six, A. *et al.* The Past, Present, and Future of Immune Repertoire Biology – The Rise of
553 Next-Generation Repertoire Analysis. *Front. Immunol.* **4**, (2013).
- 554 24. Greiff, V., Miho, E., Menzel, U. & Reddy, S. T. Bioinformatic and Statistical Analysis of
555 Adaptive Immune Repertoires. *Trends Immunol.* **36**, 738–749 (2015).
- 556 25. Wang, C. *et al.* High throughput sequencing reveals a complex pattern of dynamic
557 interrelationships among human T cell subsets. *Proc. Natl. Acad. Sci.* **107**, 1518–1523
558 (2010).
- 559 26. Zhang, W. *et al.* IMonitor: A Robust Pipeline for TCR and BCR Repertoire Analysis. *Genetics*
560 **201**, 459–472 (2015).
- 561 27. Douek, D. C. *et al.* A Novel Approach to the Analysis of Specificity, Clonality, and
562 Frequency of HIV-Specific T Cell Responses Reveals a Potential Mechanism for Control of
563 Viral Escape. *J. Immunol.* **168**, 3099–3104 (2002).
- 564 28. Eugster, A. *et al.* Measuring T cell receptor and T cell gene expression diversity in antigen-
565 responsive human CD4+ T cells. *J. Immunol. Methods* **400–401**, 13–22 (2013).
- 566 29. Mamedov, I. Z. *et al.* Preparing Unbiased T-Cell Receptor and Antibody cDNA Libraries for
567 the Deep Next Generation Sequencing Profiling. *Front. Immunol.* **4**, (2013).
- 568 30. Shugay, M. *et al.* Towards error-free profiling of immune repertoires. *Nat. Methods* **11**,
569 653–655 (2014).

- 570 31. Oakes, T. *et al.* Quantitative Characterization of the T Cell Receptor Repertoire of Naïve
571 and Memory Subsets Using an Integrated Experimental and Computational Pipeline
572 Which Is Robust, Economical, and Versatile. *Front. Immunol.* **8**, (2017).
- 573 32. Liu, X. *et al.* Systematic Comparative Evaluation of Methods for Investigating the TCR β
574 Repertoire. *PLOS ONE* **11**, e0152464 (2016).
- 575 33. Rosati, E. *et al.* Overview of methodologies for T-cell receptor repertoire analysis. *BMC*
576 *Biotechnol.* **17**, (2017).
- 577 34. Dunn-Walters, D., Townsend, C., Sinclair, E. & Stewart, A. Immunoglobulin gene analysis
578 as a tool for investigating human immune responses. *Immunol. Rev.* **284**, 132–147 (2018).
- 579 35. Doenecke, A., Winnacker, E.-L. & Hallek, M. Rapid amplification of cDNA ends (RACE)
580 improves the PCR-based isolation of immunoglobulin variable region genes from murine
581 and human lymphoma cells and cell lines. *Leukemia* **11**, 1787–1792 (1997).
- 582 36. Nielsen, S. C. A. & Boyd, S. D. Human adaptive immune receptor repertoire analysis—
583 Past, present, and future. *Immunol. Rev.* **284**, 9–23.
- 584 37. Bolotin, D. A. *et al.* MiXCR: software for comprehensive adaptive immunity profiling. *Nat.*
585 *Methods* **12**, 380–381 (2015).
- 586 38. Yokota, R., Kaminaga, Y. & Kobayashi, T. J. Quantification of Inter-Sample Differences in T-
587 Cell Receptor Repertoires Using Sequence-Based Information. *Front. Immunol.* **8**, (2017).
- 588 39. Gudikote, J. P. & Wilkinson, M. F. T-cell receptor sequences that elicit strong down-
589 regulation of premature termination codon-bearing transcripts. *EMBO J.* **21**, 125–134
590 (2002).
- 591 40. Schulze-Koops, H. Lymphopenia and autoimmune diseases. *Arthritis Res. Ther.* **6**, 178–180
592 (2004).

- 593 41. Miho, E. *et al.* Computational Strategies for Dissecting the High-Dimensional Complexity
594 of Adaptive Immune Repertoires. *Front. Immunol.* **9**, (2018).
- 595 42. Kivioja, T. *et al.* Counting absolute numbers of molecules using unique molecular
596 identifiers. *Nat. Methods* **9**, 72–74 (2012).
- 597 43. Smith, T., Heger, A. & Sudbery, I. UMI-tools: modeling sequencing errors in Unique
598 Molecular Identifiers to improve quantification accuracy. *Genome Res.* **27**, 491–499
599 (2017).
- 600 44. Britanova, O. V. *et al.* Dynamics of Individual T Cell Repertoires: From Cord Blood to
601 Centenarians. *J. Immunol.* **196**, 5005–5013 (2016).
- 602 45. Brüggemann, M. *et al.* Standardized next-generation sequencing of immunoglobulin and
603 T-cell receptor gene recombinations for MRD marker identification in acute lymphoblastic
604 leukaemia; a EuroClonality-NGS validation study. *Leukemia* (2019) doi:10.1038/s41375-
605 019-0496-7.
- 606 46. Knecht, H. *et al.* Quality control and quantification in IG/TR next-generation sequencing
607 marker identification: protocols and bioinformatic functionalities by EuroClonality-NGS.
608 *Leukemia* (2019) doi:10.1038/s41375-019-0499-4.
- 609 47. Friedensohn, S. *et al.* Synthetic Standards Combined With Error and Bias Correction
610 Improve the Accuracy and Quantitative Resolution of Antibody Repertoire Sequencing in
611 Human Naïve and Memory B Cells. *Front. Immunol.* **9**, (2018).
- 612 48. Breden, F. *et al.* Reproducibility and Reuse of Adaptive Immune Receptor Repertoire Data.
613 *Front. Immunol.* **8**, 1418 (2017).
- 614 49. Rubelt, F. *et al.* Adaptive Immune Receptor Repertoire Community recommendations for
615 sharing immune-repertoire sequencing data. *Nature Immunology*
616 <https://www.nature.com/articles/ni.3873> (2017) doi:10.1038/ni.3873.

- 617 50. Olson, B. J. *et al.* sumrep: A Summary Statistic Framework for Immune Receptor
618 Repertoire Comparison and Model Validation. *Front. Immunol.* **10**, 2533 (2019).
- 619 51. Vander Heiden, J. A. *et al.* AIRR Community Standardized Representations for Annotated
620 Immune Repertoires. *Front. Immunol.* **9**, 2206 (2018).
- 621 52. Thomas, N., Heather, J., Ndifon, W., Shawe-Taylor, J. & Chain, B. Decombinator: a tool for
622 fast, efficient gene assignment in T-cell receptor sequences using a finite state machine.
623 *Bioinformatics* **29**, 542–550 (2013).
- 624 53. Zhang, Y. *et al.* Tools for fundamental analysis functions of TCR repertoires: a systematic
625 comparison. *Brief. Bioinform.* doi:10.1093/bib/bbz092.
- 626 54. Weber, C. R. *et al.* immuneSIM: tunable multi-feature simulation of B- and T-cell receptor
627 repertoires for immunoinformatics benchmarking. *Bioinformatics*
628 doi:10.1093/bioinformatics/btaa158.
- 629 55. Wickham, H. *ggplot2 - Elegant Graphics for Data Analysis.* (2016).
- 630 56. Marcou, Q., Mora, T. & Walczak, A. M. High-throughput immune repertoire analysis with
631 IGoR. *Nat. Commun.* **9**, (2018).
- 632 57. Murugan, A., Mora, T., Walczak, A. M. & Callan, C. G. Statistical inference of the
633 generation probability of T-cell receptors from sequence repertoires. *Proc. Natl. Acad. Sci.*
634 **109**, 16161–16166 (2012).
- 635 58. Horn, H. S. Measurement of ‘Overlap’ in Comparative Ecological Studies. *Am. Nat.* **100**,
636 419–424 (1966).
- 637 59. Jaccard, P. The distribution of the flora in the Alpine zone. *New Phytol.* **11**, 37–50 (1912).
- 638 60. Sadee, C., Pietrzak, M., Seweryn, M. & Rempala, G. *Tools for analysis of diversity and*
639 *similarity in biological system (Diversity and Overlap Analysis Package).* (2017).
- 640 61. Kolde, R. *Package ‘pheatmap’.* (2019).

- 641 62. Renyi, A. On measures of information and entropy. *Proc. 4th Berkeley Symp. Math. Stat.*
642 *Probab.* 547–561 (1961).
- 643 63. Hill, M. O. Diversity and evenness: a unifying notation and its consequences. *Ecology* **54**,
644 427–432 (1973).
- 645 64. Chaara, W. *et al.* RepSeq Data Representativeness and Robustness Assessment by
646 Shannon Entropy. *Front. Immunol.* **9**, 1038 (2018).
- 647 65. Hausser, J. & Strimmer, K. *Estimation of Entropy, Mutual Information and Related*
648 *Quantities.* (2014).
- 649 66. Colwell, R. K. *et al.* Models and estimators linking individual-based and sample-based
650 rarefaction, extrapolation and comparison of assemblages. *J. Plant Ecol.* **5**, 3–21 (2012).
- 651 67. Hsieh, T. C., Ma, K. H. & Chao, A. *Package iNEXT: Interpolation and Extrapolation for*
652 *Species Diversity.* (2019).
- 653 68. Chao, A. *et al.* Rarefaction and extrapolation with Hill numbers: a framework for sampling
654 and estimation in species diversity studies. *Ecol. Monogr.* **84**, 45–67 (2014).

655

656

657 **FIGURE LEGENDS**

658 **Fig. 1: Performance statistics and VDJ rearrangement model of each method for experiments**
659 **A and B.**

660 **a**, The proportion of sequence reads aligned for TRA or TRB genes for each TCRseq replicate
661 per experiment (Experiment A, top, Experiment B, bottom). The bars represent the percentage
662 of TRA and TRB alignment, and the reason for alignment failure is color coded. **b**, Distribution
663 of the reads quality control (QC) for each method over all datasets, computed with fastQC
664 software (www.bioinformatics.babraham.ac.uk/projects/fastqc). **c**, Percentage of reads

665 collapsed after PCR error correction for all samples in the study. For each method, the MiXCR
666 clustering strategy was applied to correct for PCR errors and collapse reads. Each box-plot
667 represents the percentage of clustered reads. **d**, Multi-dimensional scaling (MDS) of V(D)J
668 recombination parameters. MDS was performed based on the Jensen-Shannon Divergence
669 (JSD) calculated between replicates on weighted VDJ segment usage (Segment usage), non-
670 template nucleotide insert size distributions (Insert size), V/D/J segment trimming distributions
671 (Deletion size), and nucleotide frequencies in N-inserts (Insert profile). JSD values were
672 transformed to rank for better visualization. Solid and dotted polygons outline samples from
673 experiments A and B, respectively. Colors represents the different methods as in B (only
674 methods used in both experiments are presented). **e**, Replicability and reproducibility of the
675 TRA and TRB repertoires for each method. The average JSD calculated in D (rows) for TRA (left)
676 and TRB (right) measured between replicates produced by the same method (Replicability, top)
677 or replicates of a given method and all other protocols (Reproducibility, bottom) was used as
678 distance metric to compare different protocols (columns). Columns are sorted according to the
679 mean scaled distance (averaged over all rows) from the lowest (best
680 replicability/reproducibility) to the highest (worst replicability/reproducibility). Distance values
681 are shown using a color scale. Jurkat TCR sequences were removed from datasets for this
682 analysis.

683

684 **Fig. 2: TRBV usage comparison between flow cytometry and TCRseq.**

685 **a**, Flow cytometry and TCRseq TRBV frequencies. Bar plots represent the TRBV frequencies
686 calculated from flow cytometry stained CD4⁺ T effector cells for the 24 TRBV for which
687 antibodies are available and from the TCRseq data, considering only clonotypes annotated for
688 the same 24 TRBV (original TRBV frequencies were used accordingly). Histograms of the 24

689 TRBV frequencies are organized by decreasing order using frequencies obtained by flow
690 cytometry as a reference reference (TRBV20-1, TRBV19, TRBV12-3/4, TRBV28, TRBV2, TRBV3-
691 1, TRBV30, TRBV6-5/9, TRBV9, TRBV5-1, TRBV4-1/2, TRBV27, TRBV29-1, TRBV6-6, TRBV11-2,
692 TRBV10-3, TRBV25-1, TRBV6-2, TRBV18, TRBV5-5, TRBV14, TRBV5-6, TRBV13, TRBV4-3). **b**,
693 TRBV usage correlation between flow cytometry and TCRseq. Pearson's correlation of the TRBV
694 frequencies between the 5 flow cytometry datasets and the 9 TCRseq replicates was calculated
695 for each method. The plot is represented by the correlation score (y-axis) and the *P*-value (x-
696 axis) of the correlation, allowing the classification of the methods. **c**, Heatmap of the Pearson
697 correlations between each replicate for the distribution of TRBV gene usage (n=62). The
698 Euclidean distance was used for hierarchical clustering as a color-coded matrix ranging from 0
699 (yellow, maximum dissimilarity) to 1 (orange, maximum similarity). Jurkat TCR sequences were
700 removed from datasets for this analysis.

701

702 **Fig. 3: The reproducibility of detection of major TCR clonotypes by different methods.**

703 **a**, Heatmaps of the Morisita-Horn similarity index (MH). MH scores were calculated between
704 each replicate across all methods for the top 1% of most predominant clonotypes (MPC) for
705 TRA (left) and TRB (right). The Euclidean distance was used for hierarchical clustering as a color-
706 coded matrix ranging from 0 (blue, maximum dissimilarity) to 1 (red, maximum similarity). **b**,
707 Comparison between individual replicates (Single) and pooled replicates (Pool) by the MH
708 similarity index. Datasets from replicates of the same dilution were pooled for each method to
709 get 1 pooled sample per dilution. Singletons (count=1) were removed; MH similarity scores
710 were calculated for the top 1% of most predominant clonotypes for TRA (left) and TRB (right).
711 Jurkat TCR sequences were removed from datasets for this analysis.

712

713 **Fig. 4: Sensitivity of TCR sequence detection by different methods.**

714 **a**, Jurkat clone percentage. Jurkat TRA (top) and TRB (bottom) clonotype percentages were
715 calculated for each experiment per dilution (1/10, 1/100 and 1/1000 spike-in) and are
716 represented by the blue dots. The blue line represents linear regression and the black dashed
717 line represents the theoretically expected percentage. **b**, Slope of the Jurkat tracking linear
718 regression. Slope was computed between dilution with standard deviation by method for TRA
719 (top) and TRB (bottom). **c**, Standard deviation of the clonotypes shared among the 9 replicates
720 (except Jurkat clone) per method, for TRA (left) and TRB (right).

721

722 **Fig. 5: Sharing with robust and representative meta-repertoire.**

723 **a**, Replicate sharing fraction in meta-repertoire repertoire (focus on meta-repertoire
724 clonotypes) for TRA (left) and TRB (right). The values represented correspond to the percentage
725 of clonotypes from each replicate per method found in the meta-repertoire, median and the
726 1st and 3rd quartiles are shown. **b**, Replicability of replicate methods with meta-repertoire for
727 TRA (left) and TRB (right). By chain, heatmaps on the left represent the fraction, which
728 corresponds to the percentage of meta-repertoire clonotypes found in 1 to 9 replicates per
729 method (0: unseen in any of the replicates). **c**, Distribution of meta-repertoire clonotypes in the
730 replicates by methods for TRA (left) and TRB (right). Each dot represents a meta-repertoire
731 clonotype and the boxplot represents the average frequencies. Black boxplots with
732 corresponding gray dots represent the unseen clonotypes (0) and red boxplots with
733 corresponding gray dots represent the clonotypes found by the 9 replicates (9). Each method
734 is represented independently. Jurkat TCR sequences were removed from datasets for this
735 analysis.

736

737 **Table 1: Comparative performance of the nine TCRseq molecular methods.** For each method,
738 an average rank score for TRA (top) and TRB (bottom) sequencing has been calculated for
739 Replicability, Reliability, and Sensitivity (three first column) and practical information have been
740 summarized (4 last columns). Ranks have been calculated as the average of the ranks for results
741 from Fig. 1e, 2c, 3b, 4c for “Replicability”; Fig. 1e, 2b, 4b, 5a, 5b for “Reliability”; Fig. 4c, 5b &
742 Supplementary Fig. 2a, 5c for “Sensitivity”. Rank values are comprised between 2 (best) and 7
743 (worst) and represented as bars with their values. Details are provided as Supplementary
744 information. Cost per sample” is expressed in USD as per current prices for a depth of 1 million
745 TCR sequences per sample on a 25 million reads sequencing format. The costs cover reagents
746 for library preparation to sequencing. *mPCR1 and mPCR3 price ranges correspond to the cost
747 for either purchasing kits (lowest price) or service up to sequencing and basic data analyses
748 from the provider.