Clinical Chemistry 66:9
1228–1237 (2020)

Ultrasensitive DNA Immune Repertoire Sequencing Using Unique Molecular Identifiers

Gustav Johansson,[a,b,c] Melita Kaltak,[a] Cristiana Rîmniceanu,[d] Avadhesh K. Singh,[d] Jan Lycke,[e] Clas Malmeström,[e] Michael Hühn,[c] Outi Vaarala,[f,†1] Susanna Cardell,[d] and Anders Ståhlberg[a,b,g*]

BACKGROUND: Immune repertoire sequencing of the T-cell receptor can identify clonotypes that have expanded as a result of antigen recognition or hematological malignancies. However, current sequencing protocols display limitations with nonuniform amplification and polymerase-induced errors during sequencing. Here, we developed a sequencing method that overcame these issues and applied it to γδ T cells, a cell type that plays a unique role in immunity, autoimmunity, homeostasis of intestine, skin, adipose tissue, and cancer biology.

METHODS: The ultrasensitive immune repertoire sequencing method used PCR-introduced unique molecular identifiers. We constructed a 32-panel assay that captured the full diversity of the recombined T-cell receptor delta loci in γδ T cells. The protocol was validated on synthetic reference molecules and blood samples of healthy individuals.

RESULTS: The 32-panel assay displayed wide dynamic range, high reproducibility, and analytical sensitivity with single-nucleotide resolution. The method corrected for sequencing-depended quantification bias and polymerase-induced errors and could be applied to both enriched and nonenriched cells. Healthy donors displayed oligoclonal expansion of γδ T cells and similar frequencies of clonotypes were detected in both enrichment and nonenriched samples.

CONCLUSIONS: Ultrasensitive immune repertoire sequencing strategy enables quantification of individual and specific clonotypes in a background that can be applied to clinical as well as basic application areas. Our approach is simple, flexible, and can easily be implemented in any molecular laboratory.

Introduction

T and B lymphocytes undergo clonal expansion upon activation. The monitoring of clonality has immense importance in the diagnosis and follow-up of hematological diseases *(1–3)* and is a fundamental tool for the understanding of antigen specificity in immunity to infections, cancer, and in autoimmunity. Today, a comprehensive assessment of clonality is feasible using next-generation sequencing (NGS)-based approaches *(3, 4)*. Clinically, there is a need for improved methodological specificity and sensitivity to determine the immune repertoire, for example, to detect acute lymphoblastic leukemia and to monitor minimal residual disease *(1, 5)*. Moreover, in studies of immune disorders, such as autoimmunity and tumor immunity, improved immune repertoire analysis will facilitate the identification of involved immune cells.

T cells are divided into two major subtypes based on their antigen receptor, the αβ and γδ T cells. The gene loci encoding the γ and δ chains of the γδ T-cell receptor (TCR) display fewer gene segments for recombination, but the potential repertoire of γδ T cells is larger than that of αβ T cells, due to possible usage of two D-segments in the TCRδ chain. Despite intense research, only a few ligands and their recognition by the γδ TCR are known *(6, 7)*. Thus, there is a need for improved understanding of factors that drives γδ T-cell activation and clonal expansion.

[a]Sahlgrenska Center for Cancer Research, Department of Laboratory Medicine, Institute of Biomedicine, Sahlgrenska Academy at University of Gothenburg, Gothenburg, Sweden; [b]Wallenberg Centre for Molecular and Translational Medicine, University of Gothenburg, Gothenburg, Sweden; [c]Translational Science & Experimental Medicine, Research and Early Development, Respiratory, Inflammation and Autoimmune (RIA), Gothenburg, Sweden; [d]Department of Microbiology and Immunology, Institute of Biomedicine, Sahlgrenska Academy at University of Gothenburg, Medicinaregatan 7A, University of Gothenburg, Gothenburg, Sweden; [e]Department of Clinical Neuroscience, Institute of Neuroscience and Physiology, Sahlgrenska Academy at University of Gothenburg, Gothenburg, Sweden; [f]Respiratory Inflammation and Autoimmunity, IMED Biotech Unit, AstraZeneca, Gothenburg, Sweden; [g]Department of Clinical Genetics and Genomics, Sahlgrenska University Hospital, Gothenburg, Sweden.
* Address correspondence to this author at: Sahlgrenska Center for Cancer Research, Department of Laboratory Medicine, Institute of Biomedicine, Sahlgrenska Academy at University of Gothenburg, Gothenburg 413 90, Sweden. E-mail: anders.stahlberg@gu.se.
[†]Present address: Clinicum, University of Helsinki, Helsinki 00014, Finland
Received March 13, 2020; accepted June 25, 2020.
DOI: 10.1093/clinchem/hvaa159

The immune repertoire can be determined by sequencing the third complementary determining region (CDR3), by targeted amplification of either DNA *(8)* or mRNA *(4, 9)*. A challenge when analyzing cell clonality with these standard NGS-based approaches is that the exact number and size of clones cannot be determined due to uneven amplification, and further that clonotypes with similar sequences cannot be reliably distinguished from sequence errors introduced during library construction and sequencing *(2)*. Several methods have been developed to overcome these limitations, including the use of synthetic spiked molecules to correct for quantification biases *(10)* and bioinformatical approaches to reduce PCR-induced errors *(11)*.

Another strategy to correct for sequencing errors and to quantify molecules is ultrasensitive sequencing using unique molecular identifiers (UMIs) *(12)*. To date, UMIs have mostly been applied to mRNA when profiling the immune receptor repertoire of B and T cells to remove PCR duplicates and improving sequence accuracy *(12, 13)*. However, analysis of clonal size based on mRNA has the disadvantage that the number of transcripts per cell is not constant *(14–16)*. The reverse transcription efficiency is also variable between sequences *(17)* and 100 to 1000 times more prone to errors than high fidelity DNA polymerases *(18–20)*. Therefore, to enable quantitative clonality analysis, DNA is preferred over mRNA.

UMIs can be added to DNA by either ligation- or PCR-based approaches *(21, 22)*. Recently, Chovanec et al. reported a ligation-based approach for immunoglobulin repertoire sequencing *(23)*. Ligation-based UMI approaches require that target DNA be captured before the analysis. Another limitation is that molecules are lost due to limited ligation efficiency *(24)*. In comparison, PCR-based UMI approaches are experimentally simpler and potentially also more sensitive, since they do not suffer from ineffective capture and ligation steps. Ligation-based UMI approaches were first developed because it is challenging to introduce UMIs in PCR primers due to a massive formation of nonspecific PCR products caused by the random nucleotide sequence of UMIs. This problem is solved by methods that either shield the UMI in secondary structures *(25)* or by an extra purification step enriching DNA with incorporated UMIs *(21)*. However, no PCR-based UMI approach is currently available for immune repertoire profiling.

Here, we developed an ultrasensitive immune repertoire sequencing approach, sequencing the rearranged *T-cell receptor delta* (TRD) locus in γδ T cells. To minimize the formation of nonspecific PCR products, we used the concept of SiMSen-Seq *(26)* that protects the UMI inside a hairpin loop that opens and closes its secondary structure in a temperature-dependent manner.

Material and Methods

PRIMER DESIGN AND ASSAY VALIDATION

To capture the full diversity of the T-cell receptor δ repertoire, we designed target-specific primers for all 8 variable (*TRDV*) and 4 joining (*TRDJ*) genes associated with the TRD locus in the International immunogenetics information system gene database *(27)* (Fig. 1A). Primers were designed using National Center for Biotechnology Information's tool Primer-Blast *(28)* as described *(25)*. Forward primers targeted the downstream part of the *TRDV* genes, and the reverse primer targeted the downstream part of the *TRDJ* genes (Fig. 1B). Primers were designed to bind the nonrearranged parts of each segment amplifying the CDR3 region (Table 1). Each assay was validated using quantitative PCR and fragment analysis to ensure high efficiency (90% to 110%) and analytical specificity, respectively. Full details on assay validation of target specific primers by using synthetic reference material can be found in Supplemental Material and Method.

LIBRARY CONSTRUCTION AND SEQUENCING

DNA were barcoded in a 10 μL reaction, containing 0.2 U Phusion Hot start II DNA polymerase, 1x Phusion High-Fidelity Buffer (both #F549S, Thermo Fisher Scientific), 0.2 mM dNTP (#D7295, Sigma-Aldrich), 0.5 M L-carnitine inner salt (#C0158, Sigma-Aldrich), 40 nM of each barcode primer (polyacrylamide gel electrophoresis purified, Integrated DNA Technologies) (Table 1) and target DNA. The following temperature program was used on a T100 Thermal cycler (Bio-Rad Laboratories): 98 °C for 30 sec, 3 cycles of amplification (98 °C for 10 sec, 62 °C for 6 min, 72 °C for 30 sec, all ramping rates were 4 °C/sec), and 65 °C for 15 min and 95 °C for 15 min. Twenty microliters of 45 ng/μL *Streptomyces griseus* protease (#P5147, Sigma-Aldrich) dissolved in 10 mM Tris, 1 mM EDTA-buffer (pH 8.0, #AM9849, Thermo Fisher) was added to each well at the start of the 15 min incubation step at 65 °C to degrade the polymerase, reducing the formation of nonspecific PCR products. Illumina adapters were added in a second PCR step. Ten microliters of barcoded PCR product was amplified in a 40 μL reaction containing, 1x Q5 Hot Start High-Fidelity Master Mix (#M0494, New England BioLabs) and 400 nM of each Illumina Adaptor index primer (desalted, Sigma-Aldrich, Supplemental Table 1) using the following thermal program on a T100 Thermal cycler; 98 °C for 3 min, 30 to 40 cycles of amplification (98 °C for 10 sec, 80 °C for 1 sec, 72 °C for 30 sec, 76 °C for 30 sec, all ramping rate of 0.2 °C/sec). Final reaction concentrations are shown for each PCR. A quality control to quantify the amount of barcoded DNA using

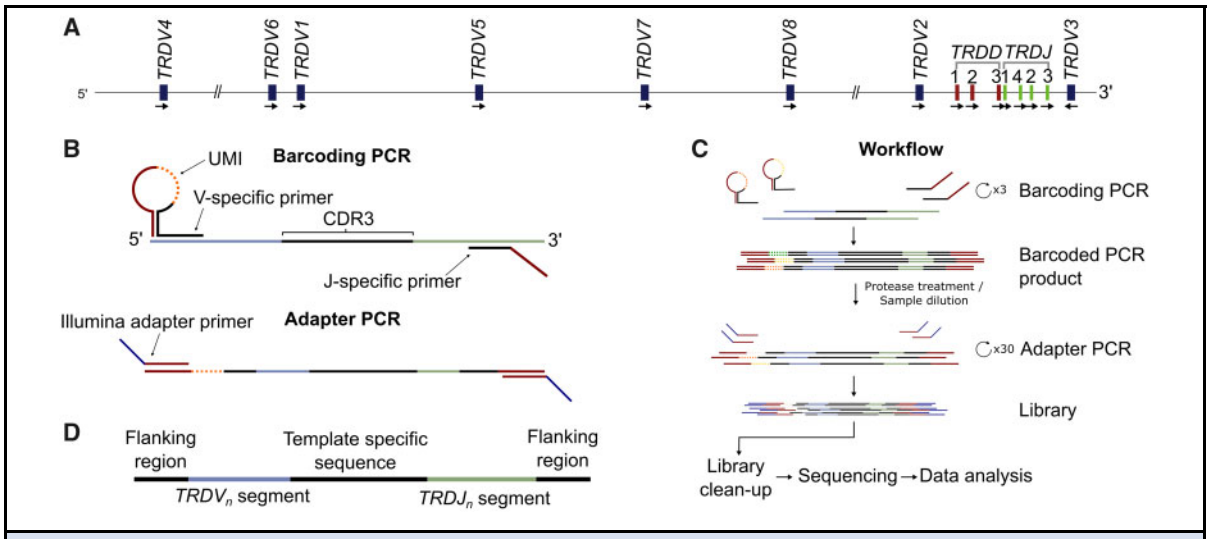

Fig. 1. Illustration of δ T-cell receptor repertoire sequencing. a) Schematic overview of the unrearranged TRD locus, arrows show transcription direction. b) SiMSen-Seq consists of two rounds of PCR. In the first barcoding PCR, target primers bind to the V and J genes, generating specific PCR products with UMIs and flanked adapter sequences. In the second adapter PCR, barcoded DNA is amplified with Illumina adapter primers. c) The experimental workflow for VDJ-sequencing. d) Design of synthetic reference molecules used for assay validation.

Table 1. Primer sequence information.

Primer name	Target specific sequence 5′->3′	Strand	Start[a]	Stop[a]
Target V-region forward primer				
TRDV1	GCGAAATCCGTCGCCTTAAC	Forward	22096543	22096562
TRDV2	ACTTGCACCATCAGAGAGAGATG	Forward	22422991	22423013
TRDV3	TCCAGTAAGGACTGAAGACAGTG	Reverse	22469085	22469063
TRDV4/TRAV14	CCAGAAGGCAAGAAAATCCGC	Forward	21924565	21924585
TRDV5/TRAV29	CTTAAACAAAAGTGCCAAGCACC	Forward	22163785	22163807
TRDV6/TRAV23	GCAGTTCTCATCGCATATCATGG	Forward	22086894	22086916
TRDV7/TRAV36	AGACCGGAGACTCGGCCAT	Forward	22227216	22227234
TRDV8/TRAV38-2	GGGGATGCCGCGATGTAT	Forward	22281712	22281729
Target J-region reverse primers				
TRDJ1	CACAGTCACACGGGTTCCTT	Reverse	22450132	22450113
TRDJ2	CGATGAGTTGTGTTCCCTTTCCAA	Reverse	22456733	22456710
TRDJ3	AGTTTGATGCCAGTTCCGAAA	Reverse	22459142	22459122
TRDJ4	GTTGTACCTCCAGATAGGTTCCT	Reverse	22455293	22455271
Barcode forward primer				
5′-GGACACTCTTTCCCTACACGACGCTCTTCCGATCTNNNNNNNNNNNNNATGGGAAAGAGTGTCC-V-region forward primer-3′				
Barcode reverse primer				
5′-GTGACTGGAGTTCAGACGTGTGCTCTTCCGATCT-J-region reverse primer-3′				

[a]Nucleotide position from GRCh38/hg38, chromosome 14 (*40*).

qPCR was preformed and described in the Supplemental Material and Methods.

Libraries were purified using the Agencourt AMPure XP system (#A63881, Beckman Coulter), according to the manufacturer's instructions with a bead to sample ratio of 1:1. Library quality and quantification were assessed on a Fragment Analyzer using the HS NGS Fragment kit (#DNF-474, Agilent) and analyzed using the PROsize software version 3.0 (Agilent), according to the manufacturer's instructions.

Final quantification of library pool was performed by qPCR using NEBNext Library Quant Kit (#E7630, New England Biolabs), according to manufacturer's instructions. Sequencing was performed on a Miniseq using midi output reagent kit with 20% added Phix control v3 (#FC-420-1004 and #FC-110-3001, Illumina) and 1 pM library. We performed 150 cycles of paired-end sequencing.

DATA ANALYSIS

Sequencing data were processed using the molecular identifier guided error correction (MIGEC) pipeline, version 1.2.9, using only paired-end reads and overlap-max-offset set to 100 *(12)*. In summary, UMIs were extracted from raw sequencing reads. Data were then grouped according to their UMIs and assembled to consensus reads with applied error correction. In the nomenclature used here, all reads amplified with the same UMI sequence form a UMI family, and all UMI families with identical TRD sequence add up to a clonotype. Post-processing was performed using modified scripts from VDJtools, version 1.2.1 *(30)* and TcR R programming package, version v2.2.4.1 *(31)*. More information on data analysis can be found in Supplemental Material and Methods.

DATA AVAILABILITY

All sequencing data can be found in the Sequence Read Archive database with the accession number PRJNA596380. Processed data, information about figure generation and details about specific analysis are available at Mendeley data *(32)*.

Results

DEVELOPMENT OF A SEQUENCING APPROACH TARGETING THE δ-CHAIN IN γδ T CELLS USING UNIQUE MOLECULAR IDENTIFIERS

To develop an ultrasensitive sequencing approach of the rearranged TRD locus in γδ T cells (Fig. 1A), we applied targeted sequencing using UMIs. Target primers were designed for 8 *TRDV* genes and 4 *TRDJ* genes (Table 1). The 12 nucleotides long UMI was incorporated between the target primer and the adapter sequence. The method consisted of 2 rounds of PCR. In the first PCR step, all target DNA was barcoded in 3 cycles of amplification. In the second PCR, the barcoded DNA was amplified with Illumina adapter primers (Figs 1, B and C and Supplemental Table 1).

To validate the efficiency of each target primer combination, we used 32 synthetic gBlock DNA molecules containing the target gene segment of each primer combination (Fig. 1D and Supplemental Table 2). All target primer pairs showed 90% to 110% PCR efficiency using quantitative PCR (qPCR) (Supplemental Table 3). The analytical specificity of all combinations was also tested on 20 ng of genomic DNA from breast cancer cell line T47D that had the TRD locus in germline configuration. None of the primer pair combinations produced any specific PCR products using Fragment Analyzer analysis (data not shown).

Next, we tested the same target primers, but with hairpin protected UMI added (Fig. 1B). We analyzed the formation of barcoded PCR-products of each primer pair combination by qPCR, followed by melting curve analysis using a standard curve, ranging from 10 000 to 16 molecules with 5-fold dilution steps. Quantitative PCR analysis showed that the dynamic range of all assays spanned the entire range, and the melting curve analysis indicated that specific PCR products were formed (Supplemental Fig. 1, A and B). The presence of correct library size was also validated by Fragment Analyzer analysis (Supplemental Fig. 1C). At lower DNA concentrations, the amount of nonspecific PCR products increased, but specific PCR products were still generated.

VALIDATION OF A SIMPLE, ROBUST AND FAST SEQUENCING PROTOCOL TARGETING THE TRD GENES

To determine amplification efficiency, sensitivity, and reproducibility of the final 32-plex assay targeting TRD gene rearrangements, we performed standard curves of synthetic gBlock molecules, ranging from 2×10^7 to 20 molecules per target sequence. To test the overall PCR efficiency and dynamic range of the 32-plex assay, we first performed barcoding PCR and then quantified the barcoded PCR product with qPCR using the Illumina adapter primers. The overall PCR efficiency for the 32-plex assay was 101% (Fig. 2A). Next, we sequenced the libraries generated from 2000, 200, and 20 synthetic reference DNA standard molecules to evaluate the performance of each primer pair combination (Fig. 2B). The sequencing efficiency ranged between 98% and 117% for all assays (Supplemental Table 4). Because different *TRDV* and *TRDJ* genes share sequence homology, we tested for off-target amplification. Synthetic gBlock DNA standards were matched by its unique template specific sequence to the primer that amplified

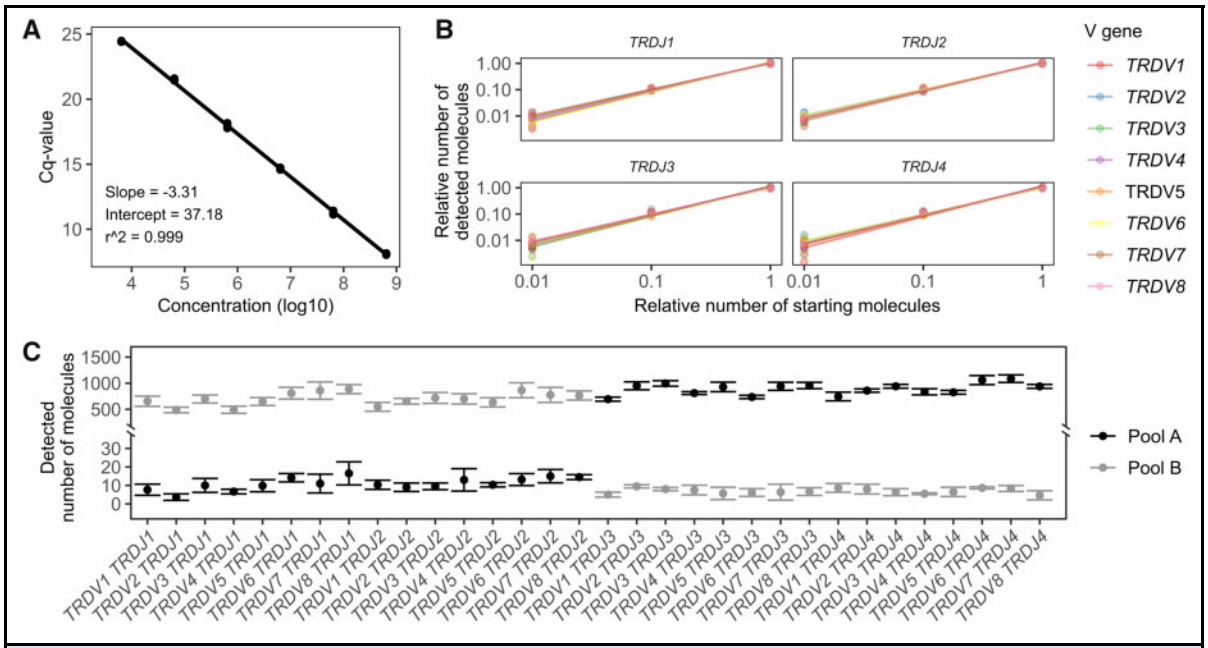

Fig. 2. Performance of 32-plex assay targeting TRD sequences. a) Dynamic range of 32-plex assay. Quantitative PCR on barcoded synthetic gBlock molecules, ranging from 2 · 10^7 to 20 molecules per standard with 10-fold dilution steps. Cycle of quantification value (Cq-value) is shown, n = 3 b) Individual assay performance. The number of molecules quantified by sequencing using. 2000, 200, and 20 synthetic gBlock molecules per assay. For visualization purpose, all standard curves are normalized by the mean value of the 2000 synthetic gBlock samples. The standard curve of each assay is shown, n = 3. c) Assay sensitivity and reproducibility. Sequencing of 2 pools (A and B) of synthetic gBlock molecules with approximately 10 molecules for half of the assays and 1000 molecules for the other half. Synthetic gBlock molecules that were in minority in the first pool were in majority in the second pool. Mean ± SD is shown, n = 3.

respective molecule (Supplemental Fig. 2). All *TRDV* primers showed high specificity where only 7 out of 175 103 (0.004%) barcoded families were amplified by the wrong *TRDV* target primer. For the *TRDJ* primers, the corresponding number was 705 out of 175 961 (0.4%) barcoded families, which was expected due to higher sequence similarity between J genes compared to V genes.

To further determine the analytical sensitivity to detect rare TRD clones, we generated pools of synthetic gBlock standards where we sequenced 10 standard molecules per assay for 16 assays (160 molecules in total) in a background of 1000 standard molecules per assay for the remaining 16 individual assays (16 000 molecules in total). Hence, the analytical sensitivity of each primer pair combination was tested at a ratio of 1:1616 (~0.06%). We quantified the pools of synthetic gBlock molecules by qPCR (Supplemental Fig. 3) and then by sequencing (Fig. 2C). Sequencing data showed that 1000, as well as 10 molecules, were reproducibly detected for all assays. Some variations between the assays may be explained by dilution artifacts as observed by qPCR (Supplemental Fig. 3).

IMMUNE REPERTOIRE SEQUENCING OF TRD USING ENRICHED γδ T-CELLS

To validate our 32-plex assay on human samples, we analyzed the genomic DNA extracted from γδ T-cells enriched from buffy coats using immunomagnetic cell separation. Sequencing libraries were constructed from 500, 100, and 25 ng γδ T-cell DNA and then sequenced. The total number of productive TRD molecules for each combination of rearranged TRD locus showed a linear correlation with the amount of analyzed DNA (Fig. 3A). The 500, 100, and 25 ng libraries generated mean 97.1%, 93.2%, and 76.1% on target reads, respectively. Non-aligned reads were almost exclusively generated from primer-dimers (mean 99.8%).

Supplemental Fig. 4A displays the 10 most commonly produced clonotypes. One advantage of using UMIs is that it could correct for amplification biases between molecules. Supplemental Fig. 4B shows the non-uniform amplification of different UMI families, which is agreement with other approaches using UMIs (*21*, *26*). Several specific UMIs were only observed once, while some specific UMIs were detected more

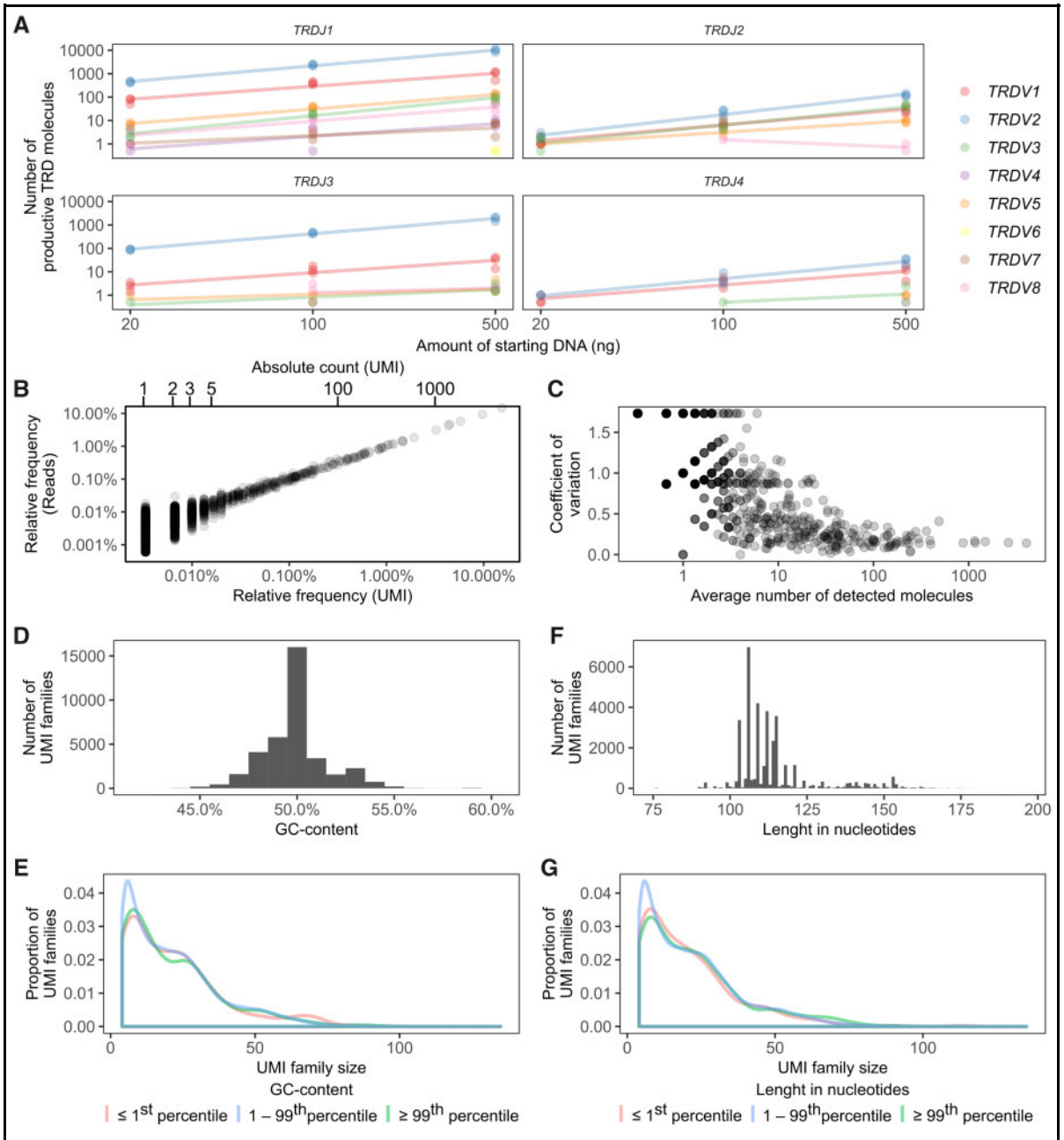

Fig. 3. Immune repertoire sequencing of γδ T cells. a) Number of productive TRD molecules for each assay, versus starting amount of γδ T-cell DNA. The linear correlation of each assay is shown, n = 3. b) Correction of molecule quantification using UMIs. Relative frequencies of clonotypes using raw sequencing reads (y-axis) versus using UMI (x-axis) are shown. Absolut molecule count based on UMI is shown on top x-axis. Data from a representative sample is show. c) Sequencing reproducibility and sampling ambiguity. Coefficient of variation versus average number of barcode families detected is shown. Data are from 500 ng γδ T-cells DNA, n = 3. d) Distribution of molecules with different GC-content. e) UMI family size distribution in relation to different GC-content. 1st and 99th percentile is at 46.0% and 54.4% GC content, respectively. f) Distribution of molecules with different amplicon length. g) UMI family size distribution in relation to amplicon length, 1st and 99th percentile is at 92 and 156 nucleotides, respectively. Raw sequencing reads were used in c), f) andi).

than 100 times. This quantitative bias was corrected since all reads with the same UMI originated from the same molecule and were collapsed into one molecule. Figure 3B illustrates the errors when not using UMI. For example, individual clonotypes with a single UMI sometimes displayed more than 10-fold variability in frequency when analyzed with raw sequencing reads. Figure 3C shows the reproducibility when detecting distinct number of molecules using UMIs. The coefficient of variation increased when detecting rare clonotypes, which agreed with sampling ambiguity when detecting few molecules. Furthermore, we detected no correlations between reads per UMI and amplicon length but a small decrease of reads per UMI for high GC-content (Fig. 3, D–G), supporting that our quantitative sequencing approach was largely independent of sequence context.

IMMUNE REPERTOIRE SEQUENCING OF TRD IN HEALTHY DONORS

To characterize individual TRD immune repertoires, we analyzed γδ T-cells, enriched by fluorescence-activated cell sorting, from peripheral blood mononuclear cells (PBMC) of 10 healthy individuals. The number of productive CDR3 sequences detected by sequencing correlated with the estimated amount of productive CDR3 γδ T-cell molecules loaded (Fig. 4A). The TRD repertoire diversity was different between the 10 individuals (Fig. 4B). Donors 1 and 10 displayed no single clonotype with a frequency above 10%, while donors 3, 4, 6, 7, 8, and 9 were highly oligoclonal, where the top 5 clonotypes represented more than 50% of all productive TRD molecules detected. The distributions of clonotype sizes are shown in Supplemental Fig. 5. The observed variation in TRD repertoire diversity of healthy donors agreed with reported data *(33, 34)*. The most common TRD rearrangement used among the different clonotypes was between *TRDV2* and *TRDJ1* (Supplemental Fig. 6). To validate our genomic approach at the cellular level, we determined the frequencies of *TRDV1* and *TRDV2* using FACS (Fig. 4C). The frequencies detected by FACS correlated significantly ($P < 0.01$) with immune repertoire sequencing.

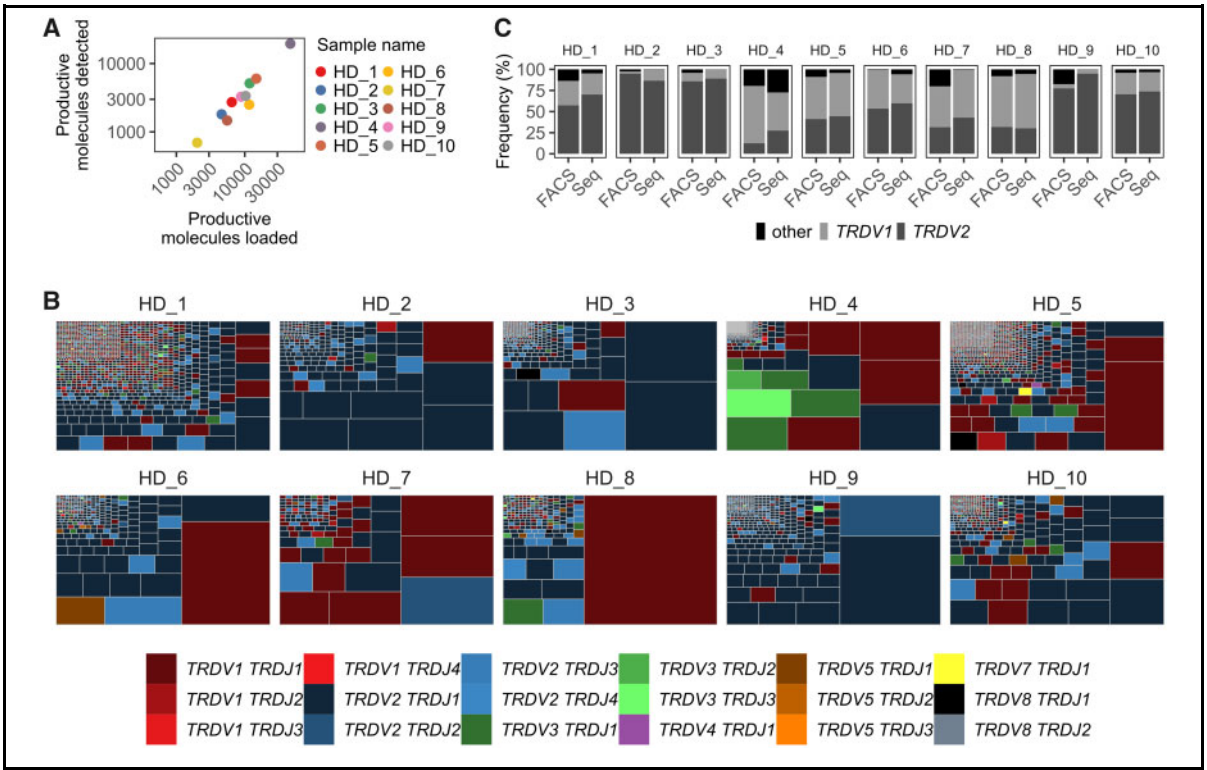

Fig. 4. The immune repertoire of TRD in healthy individuals. a) The number of productive TRD molecules detected by sequencing compared with the amount of γδ T-cell DNA used, assuming 139 productive molecules per nanogram DNA. b) Treemapping of all clonotypes across 10 healthy individuals. Each square represents a unique clonotype. The area of the square indicates the clonotype frequency and the color shows which V and J genes was used. c) Comparison of gene usage analyzed by FACS and immune repertoire sequencing (Seq). The usage of *TRDV1* and *TRDV2* as frequencies are shown. The Spearman's rank correlation coefficient is 0.88 ($P < 0.01$) for *TRDV1* and 0.94 ($P < 0.01$) for *TRDV2* when comparing FACS with sequencing.

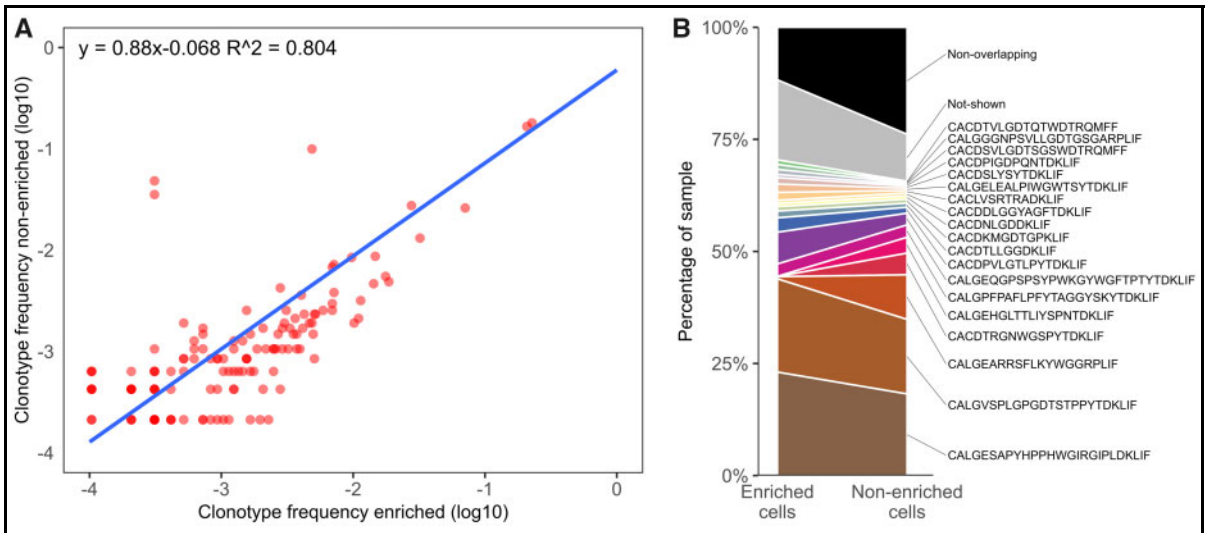

Fig. 5. Comparison of immune repertoire sequencing of enriched and non-enriched γδ T cells from PBMCs. Peripheral blood mononuclear cells (PBMCs) from one individual were split in half: one sample was enriched for γδ T cells, while the other was analyzed directly without cell enrichment. a) Each point is a unique clonotype, and its frequencies as non-enriched and enriched are shown. Line shows linear regression. b) Relative frequencies of the 20 most commonly detected clonotypes in the γδ T cell enriched and non-enriched samples, respectively. Amino acid sequence of the CRD3 region for each clonotype is shown. Non-overlapping (black) is the cumulative frequency of all clonotypes only detected in one of the samples. Non-shown (gray) is the cumulative frequency of all overlapping clones that are not among the top 20 clonotypes.

IMMUNE REPERTOIRE SEQUENCING USING NON-ENRICHED CELLS

One major advantage of targeted PCR is that cell enrichment is potentially not needed. To compare sequencing from non-enriched PBMC and enriched γδ T cells, PBMC from a buffy coat were isolated and split into 2 equal aliquots: one sample was enriched for γδ T cells using negative selection with magnetic beads, while the other sample was analyzed directly without further cell enrichment. We sequenced 100 ng DNA of the γδ T-cell enriched sample and 1 μg of the non-enriched sample. In total, 9925 and 4836 productive TRD molecules were identified in the γδ T-cell enriched and the non-enriched samples, respectively (Fig. 5, A and B). The clonotype frequencies between non-enriched and enriched cells correlated linearly ($R^2 = 0.80$). However, 3 outlier clonotypes were identified. These clonotypes were abundantly expressed in non-enriched cells, while their relative frequencies in enriched cells were low. A drawback of analyzing non-enriched γδ T cells is that partially rearranged loci and unproductively rearranged alleles of TRD in αβ T cells may be amplified. Indeed, the number of out-of-frame sequences was higher in non-enriched cells (44.5%) compared to enriched γδ T cells (15.6%).

Discussion

The TCR repertoires of γδ T cells are the least explored among B and T lymphocytes. Notably, among all immune cells investigated, infiltrating γδ T cells show the strongest association with favorable outcome in a meta-analysis of 25 different malignancies *(35, 36)*. Here, we developed an ultrasensitive method for immune repertoire sequencing that increases the quantification accuracy by multiple order of magnitudes for low-frequency clones. By utilizing UMIs, we increased sequence accuracy enabling individual clonotypes to be reliable detected and quantified in a background of similar clonotypes *(12)*. We also show that sequence libraries could be reliably generated from 15 ng enriched γδ T cells DNA and 640 synthetic molecules, equivalent to 4.6 ng enriched γδ T cells DNA. Improved quantification accuracy was achieved by counting the number of UMIs, which related to the original number of DNA molecules and, therefore, the number of cells. Consequently, PCR-introduced amplification biases were avoided. Furthermore, current bioinformatics, such as MIGEC, can also handle PCR-introduced errors in the UMI, avoiding an overestimation of the clonotype size *(37)*. The method minimizes sequencing errors by the

construction of consensus sequences. Compared to ultrasensitive detection of allele variants, such as mutation analysis, immune repertoire sequencing is experimentally more challenging since the sequences of individual target DNA molecules are different from each other, including variable GC-content and length. Allele variant analysis regularly only includes the detection of 2 different sequences, often with a single nucleotide variant. Hence, assay optimization and amplification performance are fundamental for reliable immune repertoire sequencing. The 32-plex assay showed a small GC-content bias, but this error was corrected using UMIs as long as the original molecule is amplified.

True clonal variation is difficult to separate from sequencing errors *(11, 38)*. The use of UMIs addressed this issue, enabling accurate immune repertoire sequencing. The PCR-introduced UMIs corrected for all DNA polymerase-induced errors, except errors that occurred in the first barcoding PCR step *(21)* and non-polymerase induced errors, such as chemically modified bases *(39)*. With a PCR-based approach, there is also no need for target cell enrichment. We showed a linear correlation between the frequencies of clonotypes detected in enriched and non-enriched samples. Hence, immune repertoire DNA sequencing approaches without cell enrichments may be an interesting option, since it is both simpler and allows more sample types to be analyzed. Interestingly, we detected 3 clonotypes as outliers with more than a 10-fold higher frequency in the non-enriched sample compared to the enriched sample. The reason for this is unknown and needs to be further investigated, but one possibility may be that antibodies used in the depletion cocktail for negative selection also reacted to and depleted a subset of γδ T cells. However, to determine systematic differences between various sample types and enrichment techniques need further experiments.

The approach to add UMI by targeted PCR is both simpler and more efficient than ligation-based methods. Sequence libraries can be generated within 4 h, with limited hands-on time. Another advantage of PCR-based approaches is the possibility to choose a subset of target primers for specific applications. For example, in minimal residual disease in lymphoid malignancies where the immunoreceptor clonotype is known, and the detection of other immunoreceptor recombinations is of limited value, it is possible to monitor relevant clones longitudinally in a cost-effective manner. However, in exploratory studies, PCR-based approaches will not detect novel sequences that are not targeted by the primer sequences used. Here, a more unspecific ligation-based approach will be more suitable.

Single-cell analysis is another emerging method that enables detailed profiling of different clonotypes that may also provide additional information about transcriptomics, receptor chain pairing, and antigen affinity *(40)*. However, single-cell sequencing is experimentally and analytically more labor-intensive and requires a single-cell suspension that is not always feasible to generate. Therefore, our approach and single-cell analysis are complementary to each other. An immune repertoire sequencing approach can also be combined with approaches for nonrecombined DNA sequences targeting somatic mutations *(21, 25)*. Combined immune repertoire and somatic mutation analysis is useful in both T and B cell malignancies, but also in solid tumor entities. Here, cell-free tumor DNA purified from the plasma as well as cellular immune cell DNA from the buffy coat could be analyzed in parallel.

Supplemental Material

Supplemental material is available at *Clinical Chemistry* online.

Nonstandard Abbreviations NGS, next-generation sequencing; TCR, T-cell receptor CDR3, third complementary determining region UMI, unique molecular identifier PBMC, ,peripheral blood mononuclear cell qPCR, quantitative PCR; FACS, fluorescence-activated cell sorting Cq-Value, ,cycle of quantification value.

Human Genes: *TRD*, T cell receptor delta locus; *TRA*, T cell receptor alpha locus; *TRDV*, T cell receptor variable region; *TRDJ*, T cell receptor joining region

Author Contributions: *All authors confirmed they have contributed to the intellectual content of this paper and have met the following 4 requirements: (a) significant contributions to the conception and design, acquisition of data, or analysis and interpretation of data; (b) drafting or revising the article for intellectual content; (c) final approval of the published article; and (d) agreement to be accountable for all aspects of the article thus ensuring that questions related to the accuracy or integrity of any part of the article are appropriately investigated and resolved.*

A.K. Singh, statistical analysis, provision of study material or patients; J. Lycke, financial support, provision of study material or patients; S. Cardell, financial support.

Authors' Disclosures or Potential Conflicts of Interest: *Upon manuscript submission, all authors completed the author disclosure form. Disclosures and/or potential conflicts of interest:*

Employment or Leadership: G. Johansson, AstraZeneca; M. Hühn, AstraZeneca; O. Vaarala, AstraZeneca; A. Ståhlberg, Iscaff Pharma, SiMSen Diagnostics.

Consultant or Advisory Role: J. Lycke, Almirall, Teva, Biogen, Novartis, Genzyme/SanofiAventis; A. Ståhlberg, Iscaff Pharma, TATAA Biocenter, SiMSen Diagnostics.

Stock Ownership: M. Hühn, AstraZeneca; G. Johansson, AstraZeneca; O. Vaarala, AstraZeneca; A. Ståhlberg, Iscaff Pharma, TATAA Biocenter, SiMSen Diagnostics.

Honoraria: J. Lycke, travel support and/or lecture honoraria from Biogen, Novartis, Teva, Genzyme/SanofiAventis.

Research Funding: Johan Jansson Foundation for Cancer Research, Knut and Alice Wallenberg Foundation, Wallenberg Centre for Molecular and Translational Medicine, University of Gothenburg,

Gothenburg, Sweden, Swedish Cancer Society (19-0306), Swedish Research Council (2017-01392), Swedish Childhood Cancer Foundation (2017-0043), the Swedish state under the agreement between the Swedish government and the county councils, the ALF-agreement (716321), VINNOVA.

Patents: A. Ståhlberg, United States Patent No. 10,557,134, Swedish patent application No. 2050673-9.

Role of Sponsor: The funding organizations played no role in the design of study, choice of enrolled patients, review and interpretation of data, preparation of manuscript, or final approval of manuscript.

References

1. Gazzola A, Mannu C, Rossi M, Laginestra MA, Sapienza MR, Fuligni F, et al. The evolution of clonality testing in the diagnosis and monitoring of hematological malignancies. Ther Adv Hematol 2014;5:35–47.
2. Friedensohn S, Khan TA, Reddy ST. Advanced methodologies in high-throughput sequencing of immune repertoires. Trends Biotechnol 2017;35:203–14.
3. Wu D, Sherwood A, Fromm JR, Winter SS, Dunsmore KP, Loh ML, et al. High-throughput sequencing detects minimal residual disease in acute T lymphoblastic leukemia. Sci Transl Med 2012;4:134ra63.
4. Robins HS, Campregher PV, Srivastava SK, Wacher A, Turtle CJ, Kahsai O, et al. Comprehensive assessment of T-cell receptor beta-chain diversity in alphabeta T cells. Blood 2009;114:4099–107.
5. Sherwood AM, Desmarais C, Livingston RJ, Andriesen J, Haussler M, Carlson CS, et al. Deep sequencing of the human TCRγ and TCRβ repertoires suggests that TCRβ rearranges after $\alpha\beta$ and $\gamma\delta$ T cell commitment. Sci Transl Med 2011;3:1–8.
6. Willcox BE, Willcox CR. $\gamma\delta$ TCR ligands: the quest to solve a 500-million-year-old mystery. Nat Immunol 2019;20:121–8.
7. Vermijlen D, Gatti D, Kouzeli A, Rus T, Eberl M. $\gamma\delta$ T cell responses: How many ligands will it take till we know? Semin Cell Dev Biol 2018;84:75–86.
8. Wren D, Walker BA, Brüggemann M, Catherwood MA, Pott C, Stamatopoulos K, et al. Comprehensive translocation and clonality detection in lymphoproliferative disorders by next-generation sequencing. Haematologica 2017;102:e57–60.
9. Bashford-Rogers RJ, Palser AL, Idris SF, Carter L, Epstein M, Callard RE, et al. Capturing needles in haystacks: a comparison of B-cell receptor sequencing methods. BMC Immunol 2014;15:29.
10. Carlson CS, Emerson RO, Sherwood AM, Desmarais C, Chung M-W, Parsons JM, et al. Using synthetic templates to design an unbiased multiplex PCR assay. Nat Commun 2013;4:1–9.
11. Bolotin DA, Poslavsky S, Mitrophanov I, Shugay M, Mamedov IZ, Putintseva EV, et al. MiXCR: Software for comprehensive adaptive immunity profiling. Nat Methods 2015;12:380–1.
12. Shugay M, Britanova OV, Merzlyak EM, Turchaninova MA, Mamedov IZ, Tuganbaev TR, et al. Towards error-free profiling of immune repertoires. Nat Methods 2014;11:653–5.
13. He L, Sok D, Azadnia P, Hsueh J, Landais E, Simek M, et al. Toward a more accurate view of human B-cell repertoire by next-generation sequencing, unbiased repertoire capture and single-molecule barcoding. Sci Rep 2015;4:6778.
14. Bengtsson M, Ståhlberg A, Rorsman P, Kubista M. Gene expression profiling in single cells from the pancreatic Islets of Langerhans reveals lognormal distribution of mRNA levels. Genome Res 2005;15:1388–92.
15. Raj A, Peskin CS, Tranchina D, Vargas DY, Tyagi S. Stochastic mRNA synthesis in mammalian cells. PLoS Biol 2006;4:e309.
16. Chubb JR, Trcek T, Shenoy SM, Singer RH. Transcriptional pulsing of a developmental gene. Curr Biol 2006;16:1018–25.
17. Ståhlberg A, Håkansson J, Xian X, Semb H, Kubista M, Stahlberg A, et al. Properties of the reverse transcription reaction in mRNA quantification. Clin Chem 2004;50:509–15.
18. Arezi B, Hogrefe HH. *Escherichia coli* DNA polymerase III ε subunit increases Moloney murine leukemia virus reverse transcriptase fidelity and accuracy of RT–PCR procedures. Anal Biochem 2007;360:84–91.
19. Lundberg KS, Shoemaker DD, Adams MW, Short JM, Sorge JA, Mathur EJ. High-fidelity amplification using a thermostable DNA polymerase isolated from *Pyrococcus furiosus*. Gene 1991;108:1–6.
20. Barrioluengo V, Álvarez M, Barbieri D, Menéndez-Arias L. Thermostable HIV-1 group O reverse transcriptase variants with the same fidelity as murine leukaemia virus reverse transcriptase. Biochem J 2011;436:599–607.
21. Kinde I, Wu J, Papadopoulos N, Kinzler KW, Vogelstein B. Detection and quantification of rare mutations with massively parallel sequencing. Proc Natl Acad Sci U S A 2011;108:9530–5.
22. Schmitt MW, Kennedy SR, Salk JJ, Fox EJ, Hiatt JB, Loeb LA. Detection of ultra-rare mutations by next-generation sequencing. Proc Natl Acad Sci U S A 2012;109:14508–13.
23. Chovanec P, Bolland DJ, Matheson LS, Wood AL, Krueger F, Andrews S, et al. Unbiased quantification of immunoglobulin diversity at the DNA level with VDJ-seq. Nat Protoc 2018;13:1232–52.
24. Ranjan RK, Rajagopal K. Efficient ligation and cloning of DNA fragments with 2-bp overhangs. Anal Biochem 2010;402:91–2.
25. Ståhlberg A, Krzyzanowski PM, Egyud M, Filges S, Stein L, Godfrey TE. Simple multiplexed PCR-based barcoding of DNA for ultrasensitive mutation detection by next-generation sequencing. Nat Protoc 2017;12:664–82.
26. Ståhlberg A, Krzyzanowski PM, Jackson JB, Egyud M, Stein L, Godfrey TE. Simple, multiplexed, PCR-based barcoding of DNA enables sensitive mutation detection in liquid biopsies using sequencing. Nucleic Acids Res 2016;44:e105.
27. 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 2005;33:256–61.
28. Ye J, Coulouris G, Zaretskaya I, Cutcutache I, Rozen S, Madden TL. Primer-BLAST: A tool to design target-specific primers for polymerase chain reaction. BMC Bioinformatics 2012;13:134.
29. Kent WJ, Sugnet CW, Furey TS, Roskin KM, Pringle TH, Zahler AM, et al The human genome browser at UCSC. Genome Res 2002;12:996–1006.
30. Shugay M, Bagaev DV, Turchaninova MA, Bolotin DA, Britanova OV, Putintseva EV, et al. VDJtools: Unifying post-analysis of T cell receptor repertoires. PLOS Comput Biol 2015;11:e1004503.
31. Nazarov VI, Pogorelyy MV, Komech EA, Zvyagin IV, Bolotin DA, Shugay M, et al. tcR: an R package for T cell receptor repertoire advanced data analysis. BMC Bioinformatics 2015;16:175.
32. Johansson G. Processed data for ultrasensitive DNA immune repertoire sequencing using unique molecular identifiers. Mendeley Data doi:10.17632/56brpjxg7d.1.
33. Davey MS, Willcox CR, Joyce SP, Ladell K, Kasatskaya SA, McLaren JE, et al. Clonal selection in the human Vδ1 T cell repertoire indicates $\gamma\delta$ TCR-dependent adaptive immune surveillance. Nat Commun 2017;8:1–15.
34. Ravens S, Schultze-Florey C, Raha S, Sandrock I, Drenker M, Oberdörfer L, et al. Human $\gamma\delta$ T cells are quickly reconstituted after stem-cell transplantation and show adaptive clonal expansion in response to viral infection. Nat Immunol 2017;18:393–401.
35. Gentles AJ, Newman AM, Liu CL, Bratman SV, Feng W, Kim D, et al. The prognostic landscape of genes and infiltrating immune cells across human cancers. Nat Med 2015;21:938–45.
36. Tosolini M, Pont F, Poupot M, Vergez F, Nicolau-Travers M-L, Vermijlen D, et al. Assessment of tumor-infiltrating TCRVγ9Vδ2 $\gamma\delta$ lymphocyte abundance by deconvolution of human cancers microarrays. Oncoimmunology 2017;6:e1284723.
37. Smith T, Heger A, Sudbery I. UMI-tools: modeling sequencing errors in Unique Molecular Identifiers to improve quantification accuracy. Genome Res 2017;27:491–9.
38. Victora GD, Nussenzweig MC. Germinal centers. Annu Rev Immunol 2012;30:429–57.
39. Diehl F, Schmidt K, Durkee KH, Moore KJ, Goodman SN, Shuber AP, et al. Analysis of mutations in DNA isolated from plasma and stool of colorectal cancer patients. Gastroenterology 2008;135:489–98.
40. Stubbington MJT, Lönnberg T, Proserpio V, Clare S, Speak AO, Dougan G, et al T cell fate and clonality inference from single-cell transcriptomes. Nat Methods 2016;13:329–32.