SLAS Technology 39 (2026) 100429

Contents lists available at ScienceDirect

SLAS Technology

journal homepage: www.elsevier.com/locate/slast

Full Length Article

Improving genotyped functional screening: A versatile closed-tube PCR for Illumina library generation reduces background sequences in multiplexed sequencing

Sami Oksanen [a,b,*], Tuomas Huovinen [a], Urpo Lamminmäki [a,b]

[a] Department of Life Technologies, University of Turku, Kiinamyllynkatu 10, Turku, 20520, Finland
[b] InFLAMES Research Flagship, University of Turku, Turku, 20014, Finland

ARTICLE INFO

Dataset link: https://www.ebi.ac.uk/ena/browser/view/PRJEB83324, https://github.com/sjoutu/closed-tube-barcoding/

Keywords:
NGS
Barcoding
Functional genotyped screening
Antibody engineering
PCR chimera
Background
Artifact

ABSTRACT

We present an improved version of DNA indexing platform allowing geno-phenotype analysis of all binder protein clones arrayed for high-throughput screening. The method was optimized for two primer pairs with differing annealing temperatures (>10 °C), resulting in plate and well ID barcoding in a single, closed-tube PCR. As compared to our earlier hierarchical indexing, the closed-tube approach enhanced the top-to-second sequence count ratio by threefold and decreased background sequences from 72% to 43% of total sequences. Sample cross-contamination (or index hopping) decreased from 14% to negligible levels, and chimera formation was reduced nearly sixfold. Additionally, by splitting the sequencing adapters between target and indexing primers, the closed-tube method produces sequencing-ready Illumina libraries with fewer artifact sequences.

This method is particularly beneficial for amplicons with high sequence homology, such as synthetic antibody libraries, where chimeras and other background sequences are commonly encountered in Illumina sequencing with highly multiplexed indexing schemes. The reduction in these artifacts ensures more accurate results, improving the reliability of downstream analyses (e.g. diversity or enrichment calculations) and allows a higher number of multiplexed samples. Furthermore, the platform is adaptable to novel binder scaffolds, such as nanobodies or DARPins, by designing two new target amplification primers.

1. Introduction

Next generation sequencing (NGS) is widely used in protein engineering e.g. to identify genotypes of the functionally analyzed binder clones [1,2], evaluate synthetic library quality [3] and track enrichment of specific binders during phage display selection campaigns [4]. In addition, an increasing number of protein language models, such as ImmuneBuilder, AlphaFold or AntiFormer [5–7], are being developed to predict the structure and other properties of proteins based on their sequences. To train high-precision models, it is essential to have high-quality sequence-linked functional data of various proteins.

Tagging DNA samples with unique barcode sequences enables simultaneous sequencing of numerous samples in a single run with next generation sequencing platforms [8]. Pooling a larger number of samples significantly reduces both the cost and labor associated with preparing and sequencing individual samples. For Illumina sequencing, different strategies have been proposed for barcoding of individual samples. These methods include incorporation of either single [9], dual [10] or quadruple barcode sequences [11] via ligation or PCR reactions to the amplicons.

Illumina sequencing is a powerful tool for sequencing large numbers of samples with high accuracy. However, index hopping is a well-known issue, where the index sequences are inadvertently swapped between samples [12]. Switching of indices introduces background sequences to the data, which can lead to incorrect assignment of reads to the sample. This contamination complicates accurate sample identification and can obscure true biological signals, particularly in studies with low-abundance targets or highly multiplexed libraries. Formation of chimeric DNA molecules [13,14] during PCR is a substantial source of background sequences. The rate of such chimeric products resulting from unintentional splicing of incomplete DNA fragments from different origins can exceed 70% in low diversity libraries [15]. Another cause of index hopping is the presence of unbound, free-floating index adapters on the flow cell, that can bind to the DNA fragments during the amplification [16]. The problem is more severe with the newer sequencing platforms utilizing patterned flow cells [17] and ExAmp chemistry [18] introduced in 2015, with background sequence rates exceeding up to 10% [12]. Other sources for inaccuracies in multiplexing

* Corresponding author at: Department of Life Technologies, University of Turku, Kiinamyllynkatu 10, Turku, 20520, Finland.
E-mail address: sami.j.oksanen@utu.fi (S. Oksanen).

https://doi.org/10.1016/j.slast.2026.100429
Received 28 July 2025; Received in revised form 24 March 2026; Accepted 11 May 2026
Available online 16 May 2026

include indexing adapter cross-contamination [19] and carry-over from previous sequencing runs [20]. The number of background sequences is significantly reduced through implementation of dual indexing [21]. One of the most effective strategies for preventing sample misassignment is the use of unique dual indexes (UDI) on both adapters, combined with unique molecular identifiers (UMIs) [16,22]. While the use of unique dual indexing greatly enhances the identification of correct sequences by providing distinct sample identifiers, it does not eliminate the generation of background sequences. Additionally, this approach is not economically feasible for large-scale multiplexing. For instance, multiplexing 9216 samples using UDIs would require a total of 18,432 primers, or 96 pre-mixed primer plates. Calculated with the specific primer lengths and current list prices of the primer supplier used in this study, such approach would demand an investment of over 268,800 € (2800 € per plate). Conversely, a combinatorial strategy achieves the same 9216-sample capacity using only two plates (96 × 96) for a total of 5600 €, representing a 98% reduction in upfront reagent costs. In addition to using UDIs, Illumina recommends removing free adapters from library preparations and pooling samples only immediately prior to sequencing to help mitigate index hopping [23].

Earlier, we reported a method to link the sequence information on full-length variable domains to thousands of antibody clones screened for their antigen binding in Fab format [2]. In this method, the barcoding was performed with two subsequent PCR reactions resulting in quadruple indexing, where inner indices encode for the clone's spatial location on a 96 well plate and outer indices for the plate. Although we successfully connected the binding phenotype with the variable domain genotype, we observed multiple distinct sequences arising from the same index positions. Consequently, a filtering method was devised to identify the correct sequences with high confidence. In order to scale up the system above ten thousand parallelly analyzed samples, the amount of background sequences needs to be significantly reduced.

In 2022, Wittmann et al. introduced evSeq [24], a universal barcoding system that utilizes two subsequent PCRs to tag amplicons with 96 × 96 unique barcodes. Although the authors justifiably presented the method as a cost-effective alternative to traditional indexing, emphasizing accuracy through high read depth, the workflow provides a critical, inherent benefit, that was not explicitly addressed in the study: the mitigation of PCR chimeras. By maintaining an arrayed format throughout the barcoding process and avoiding pre-sequencing pooling of the samples, evSeq bypasses the competitive template environments that facilitate cross-variant chimera formation. This protection stems directly from the physical partitioning of the workflow rather than reliance on post-sequencing bioinformatic filters. However, the workflow requires opening the reaction vessels after the initial target-specific amplification to supplement the barcoding primers, introducing a risk of cross-contamination. Furthermore, as the resulting amplicons contained only partial Illumina sequencing adapters, they require additional cycles of PCR to make them ready for sequencing. This reintroduces the risk of chimeric PCR product formation.

In this study, we report an improved, closed-tube genotyped functional screening platform for antibodies by adapting the method described by Wittmann et al. [24]. A detailed comparison of the indexing strategies is provided in Fig. 1D. Our approach involves amplifying and barcoding the antibody variable domain sequences with closed-tube PCR, without the need of intermediate reagent supplementation. The reaction contains both the target-specific and indexing primer pairs designed to operate in different annealing temperatures. Initially, the variable domains are amplified using a low number of cycles and a low concentration of target primers that include partial Illumina TruSeq adapter sequences. Barcoding is initiated by lowering the annealing temperature, which allows the indexing primers, containing the remaining adapter sequences, to hybridize effectively. Following this, the samples are pooled and size-selected for sequencing on the Illumina MiSeq platform. To validate the developed method, we functionally screened, barcoded, and sequenced four anti-dengue virus 2 non-structural protein 1 (DENV-2 NS1) Fab libraries from our original publication [2], comparing the results to those of our initial barcoding method, highlighting the enhanced sequence clarity achieved with the new approach. Additionally, to facilitate the detection of PCR chimeras, we analyzed 450 clones from a synthetic Fab library enriched with phage display against SpyCatcher-protein.

2. Results

2.1. Development and validation of the closed-tube indexing

To enhance sequence clarity and increase throughput in our genotyped functional screening platform, we developed a versatile, closed-tube dual barcoding PCR method, later referred as "closed-tube method". This approach enables the generation of sequencing-ready Illumina amplicon libraries from 9216 antibody clone variable domains in a closed-tube PCR, eliminating the need for multiple PCR reactions and intermediate purification steps. To achieve this, we designed two sets of target-specific primers for the V_H and V_L domains, respectively, with high melting temperatures (~69 °C). Each primer consisted of a variable domain-hybridizing region and an overhang sequence containing partial Illumina TruSeq adapter. We also designed 96 (forward) + 96 (reverse) universal barcoding primers, each incorporating a unique seven nucleotides long index sequence and the rest of the sequencing adapter, including a shared sequence for hybridization to the amplified target, as well as the P5 (forward) or P7 (reverse) grafts for MiSeq flow cell attachment. The indexing primers were designed with a significantly lower melting temperature (~56 °C), allowing both target-specific and indexing primers to function in the same PCR reaction by adjusting the annealing temperature for sequential amplification. Both the new closed-tube and original quadruple barcoding (later referred as "quad-barcoding" [2]) strategies are illustrated in Fig. 1 and the designed primers are listed in Supplementary Table 1, found in Appendix A.

To validate the closed-tube barcoding method, we amplified the V_L and V_H domain sequences either separately or in same reaction from parental anti-NS1 Fab 49A3 plasmid DNA using either closed-tube (both primer pairs in reaction from the beginning) or two-step protocols (indexing primers added after the first PCR has finished, similarly to Wittmann et al. [24]). In addition to non-template controls (NT), we included a reaction without indexing primers. In electrophoresis analysis, we observed clear, correct sized PCR-products of 500 and 550 bps (V_H and V_L, respectively) with both closed-tube and two-step protocols and saw no difference in the amplification efficiency, as shown in Fig. 2. We observed no visible amplification in the NT or reactions without the indexing primers, explained by specific amplification as well as low number of cycles and concentration of primers in comparison to the indexing step (10 cycles/20 nM vs. 25 cycles/200 nM, respectively). In addition, amplification of both V_L and V_H in same reaction was possible similarly to our original barcoding. Equally efficient, specific amplification of V_H and V_L with all 96 + 96 indexing primers was observed with the closed-tube protocol (Supplementary Figure 1, Appendix A).

For further evaluation, we cloned the barcoded amplicons into the pUC19 vector and Sanger sequenced three clones from each sample, including V_L, V_H, and co-amplified V_L & V_H from both closed-tube and two-step methods. Out of 18 sequencing reactions, 16 yielded successful results, producing high-quality, full-length amplicons of dual-barcoded V_L or V_H regions. Eleven samples were entirely accurate, exhibiting error-free indexing adapters and variable domain sequences. Four sequences contained a single nucleotide substitution, and one had a deletion, all occurring at random positions within the adapters. These errors likely originated from PCR amplification using Taq polymerase or from primer synthesis, as the initial test primers were ordered as standard, desalted oligonucleotides. Based on these observations, we ordered the target-specific primers used in this study as high quality

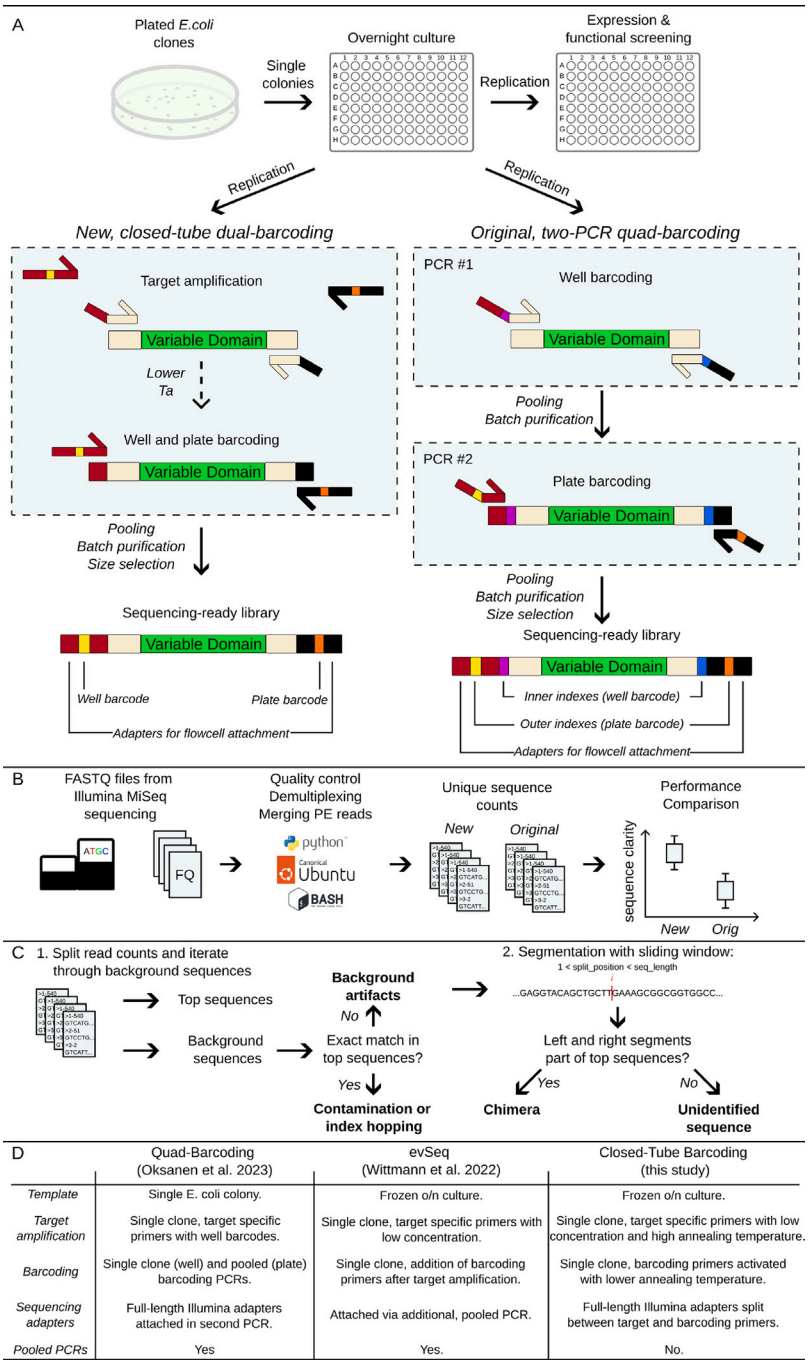

Fig. 1. Study design. (A) Schematic of the novel closed-tube (left) and original quad-barcoding (right) genotyped functional screening strategies for unique Fab clones. Single clones are inoculated from overnight cultures into both functional screening and barcoding PCR reactions. The closed-tube method employs target-specific and universal indexing primers with a high melting temperature difference, enabling target amplification and barcoding in a single reaction by adjusting the annealing temperature. Following PCR, all samples are pooled, purified, and size-selected for sequencing. The quad-barcoding involves two consecutive PCR steps: the first barcodes well coordinates in a target-specific PCR, followed by DNA purification and the addition of plate barcodes in a second PCR step. Both methods generate sequencing-ready Illumina TruSeq libraries. (B) NGS data analysis pipeline. Indexes of two different lengths (from the new and original barcoding methods) were combined in a single sequencing pool. Half of the indexes were demultiplexed manually using regular expression matching. After merging the paired-end reads, unique sequences were quantified per sample, and the sequence clarity – defined as the ratio of the most abundant sequence to secondary and total reads – was compared between the two methods. (C) Background sequence identification. First, the most abundant sequences from each sample are compiled into "top list" for subsequent matching. Remaining sequences (background) are then classified as contaminant, chimeric or background artifact sequences. If an exact match is found from the top list, it is marked as a contaminant. If not, the sequence is split with a sliding window approach; if both segments have match in top list, it is labeled chimeric, otherwise as unidentified sequence. (D) Side-to-side comparison of our original Quad-Barcoding (left), evSeq (Wittmann et al. 2022, center) and current Closed-Tube Barcoding (right) methods.

ultramer oligos. Among the six reactions where both V_H and V_L were amplified, half of the sequences corresponded to V_H, and the other half to V_L. We did not observe any difference in quality or other aspects between the closed-tube and two-step methods, strongly supporting use of former, as it simplifies the process significantly. The alignment of annotated Sanger sequences is shown in Fig. 3.

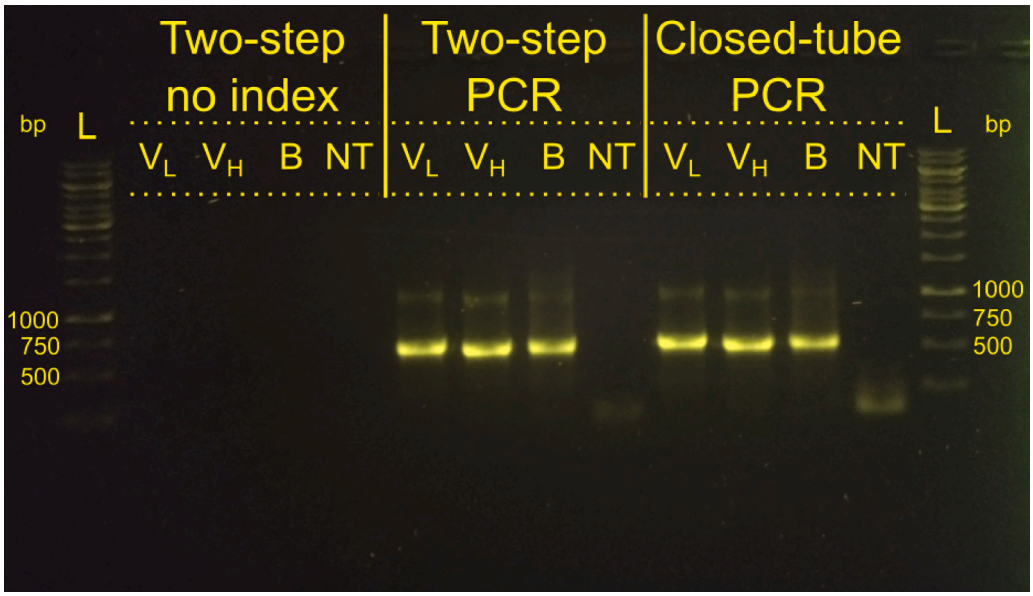

Fig. 2. Agarose gel electrophoresis analysis to test the closed-tube method. The PCR amplification was performed for Fab 49A3 variable light (V_L), variable heavy (V_H) or both domains in same reaction (B). The two-step barcoding protocol (Wittmann et al. [24]) was compared to the novel closed-tube barcoding. From left to right: amplification with only target specific primers; amplification by adding the indexing primers after target amplification PCR; amplification with both primer pairs present in the reaction from the very beginning. L = GeneRuler 1 kb ladder, NT = no-template control.

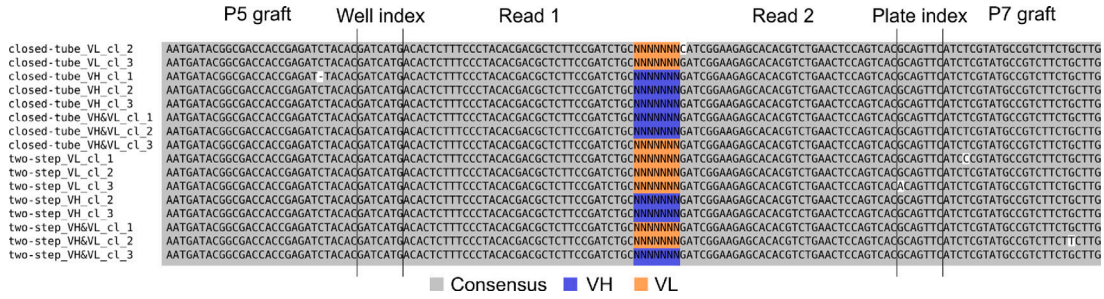

Fig. 3. Sanger sequencing of the indexing PCR amplicons. The alignment shows fully barcoded V_H and V_L domain amplicons produced with closed-tube and two-step methods. Consensus sequence is marked with gray background, V_H with blue and V_L with orange, both indicated with "NNNNNNN". The Illumina adapter domains are named on top of the sequences.

2.2. Analysis and comparison of the NGS data

To evaluate performance of the closed-tube barcoding method, we indexed and sequenced a total of 4320 samples. The same samples were run with the previously developed quad-barcoding method to enable comparison of the levels of background sequence noise within unique indices. Majority of the samples derived from the four anti-DENV-2 NS1 site-saturation mutagenesis libraries, consisting of 3600 new Fab clones, 120 positive control wells, and 120 empty control wells. Additionally, we included a more diverse set of 480 samples (including 15 positive control wells and 15 empty control wells) from a synthetic antibody library selected against SpyCatcher. Due to showing higher diversity in several positions within the amplicons, these clones were considered to be better suited for monitoring potential PCR chimera formation.

The sequencing generated 13.3 million paired-end reads, with an average Phred score of 33.6. Of these, 6.4 million reads (48%) were correctly demultiplexed by the Illumina software and contained barcodes used in quad-barcoding. The remaining "undetermined" reads were demultiplexed manually, as the Illumina software is unable to process index sequences of two different lengths within the same run. Upon demultiplexing, we found that 3.8 million (28%) of the total reads originated from closed-tube barcoding. Notably, 90% of the undetermined reads contained one or more outer indexes of quad-barcoding, while only 11.5% contained indexes closed-tube method. The imbalance of sequences might suggest an error in library quantification before pooling amplicons from the two methods.

We successfully merged 3.73 million (98.2%) paired-end reads from closed-tube and 6.05 million (94.5%) paired-end reads from quad-barcoding using the PEAR software. For each sample, we calculated the read counts for unique sequences and extracted the total read counts, as well as the counts of the two most abundant sequences. We used the ratio of the two most abundant sequences (later referred as "top-to-second ratio") as a metric for sequence clarity, similarly as we did earlier [2]. As expected, the total read counts per sample were higher for the quad-barcoding (median = 1093, interquartile range, IQR = 925) compared to the closed-tube method (median = 497, IQR = 977). However, the closed-tube method demonstrated significantly greater sequence clarity, as indicated by the top-to-second ratio (median = 70, IQR = 123) compared to the quad-barcoding (median = 27, IQR = 41), with statistical significance ($p < 0.001$). Additionally, the ratio of the top sequence to the total read counts was higher with the closed-tube method ($p = 0.0017$), further supporting the improved performance of the developed barcoding approach. The sequence clarity comparison is illustrated in Fig. 4.

To further investigate the origin of background sequences detected with both methods, we conceived a Python script to categorize these sequences, specifically from anti-SpyCatcher clones, into three types:

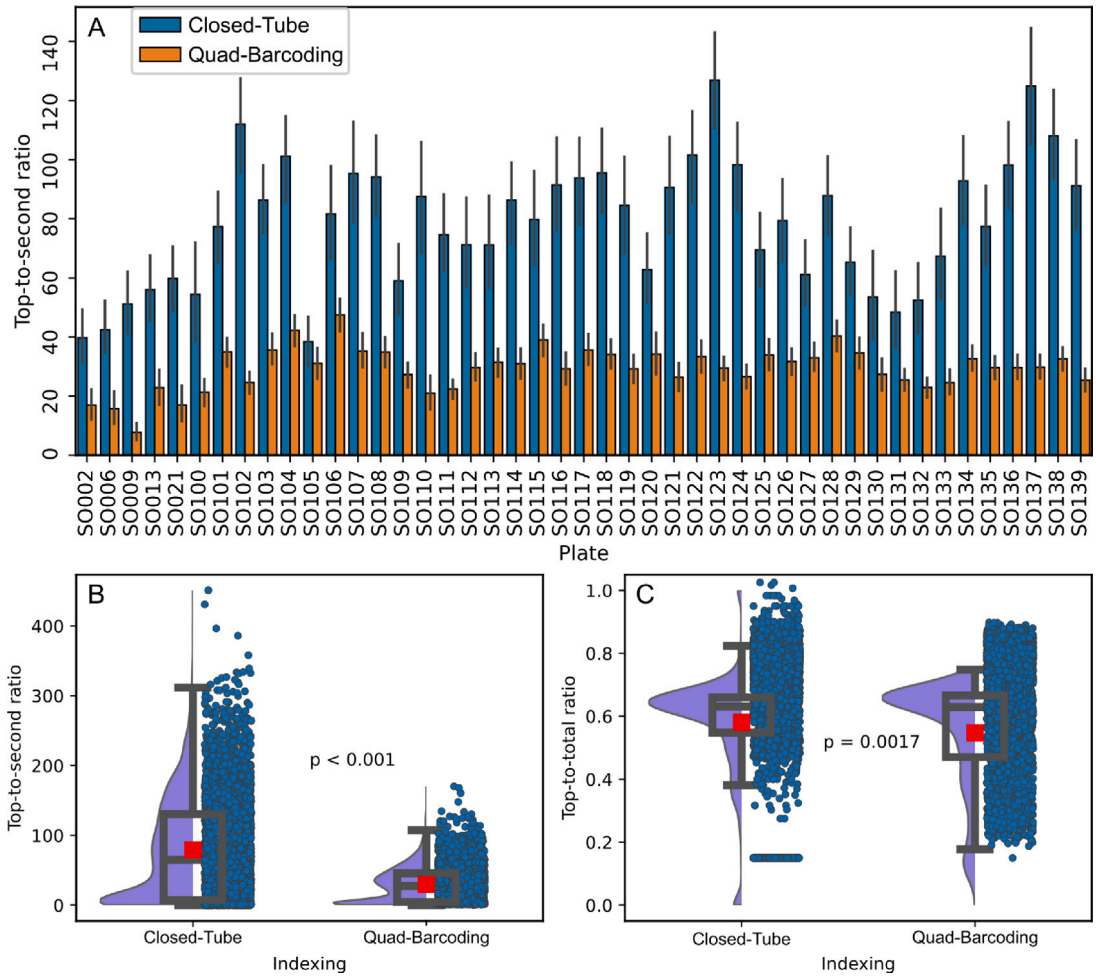

Fig. 4. Sequence clarity comparison between the closed-tube and quad-barcoding methods. The ratios of the most abundant sequence count ratio to second most abundant sequence (A and B) and to total sequence counts (C) demonstrate statistically significant improvement in sequence clarity. Stormcloud plots (B and C) display individual samples (blue points), their average (red square), and the distribution shown as a density plot (purple) and a box plot (black). The box plot includes the median (central line), interquartile range (IQR, box spans Q1 to Q3), and whiskers extending up to 1.5 times the IQR.

(1) contamination, if the sequence had an exact match with a list of the most abundant sequences, indicating potential contamination from neighboring wells or index hopping during Illumina sequencing; (2) chimera, if a split sequence – divided into left and right segments using a sliding window – had a match in the top sequence list; or (3) unidentified sequence, most likely originating from PCR or sequencing errors. The categorization process is illustrated in Fig. 1C.

In contrast to anti-DENV-2 NS1 dataset, we obtained more than twice as many anti-SpyCatcher V_H reads using closed-tube method versus quad-barcoding (450,286 vs 210,156, respectively). In addition to the improved top-to-second ratio, we also observed a significant reduction in the proportion of background sequences relative to the total number of sequences in each sample, with medians of 43% (IQR = 14.4) for closed-tube and 72.2% (IQR = 36.7) for quad-barcoding. Specifically, the proportion of chimeric sequences decreased from 13.6% (IQR = 13.6) to 2.4% (IQR = 2.2), and neighboring contaminants were reduced from 14.3% (IQR = 16.5) to 0.0% (IQR = 0.3). We also found that most background sequences resulting from closed-tube method were labeled as unidentified sequences, likely originating from PCR (single nucleotide errors or complex chimeras) or sequencing errors. These errors were slightly more prominent with the closed-tube than with quad-barcoding method with medians of 40.0% (IQR = 11.2) and 35.7% (IQR = 13.2), respectively. All analyzed differences were statistically significant, as determined by the Mann–Whitney U test. A summary of this comparison is provided in Table 1.

2.3. Genotyped functional screening of anti-DENV-2 NS1 sublibraries

To assess the success of anti-DENV-2 NS1 sublibrary generation through site saturation mutagenesis, we identified the genotypes of all quality-filtered clones using ANARCI. We observed the genotype counts to be evenly distributed based on codon usage and frequency. As anticipated, we found an increase in the parental genotype count across each sublibrary. All possible amino acids are represented in all four mutated positions. The distributions for all four sublibraries are illustrated in Fig. 5A.

We normalized the antigen-binding signal of the genotype-identified Fab clones to their expression level, which we measured using a similar time-resolved fluorescence immunoassay and quantified based on standard curves of purified Fab 49A3 (Supplementary Figure 2, Appendix A). Consistent with our previously reported results from the same sublibraries, we identified genotypes with lower affinities, such as those carrying glycine at heavy chain position 98 (H98, Kabat numbering scheme). Based on the assumption that clones exhibiting higher antigen binding than the parental clone (as indicated by a trendline on concentration vs. antigen binding illustrated in Fig. 5B) might possess superior properties, we selected several for further analysis. We tested the clones by performing antigen titration assays similar to before (data not shown). However, none of the tested clones demonstrated improved affinity compared to the parental clone 49A3.

Table 1
Comparison of background sequences identified from samples barcoded with closed-tube and quad-barcoding methods. Values are represented as medians and interquartile ranges (IQR) of percentages of the total read count for each sample. *P*-value was calculated using two-sided Mann–Whitney U test.

	Closed-tube N = 452 Median	IQR	Quad-barcoding N = 479 Median	IQR	Mann–Whitney U test
Total background sequences	43.0%	14.4	72.2%	36.7	P < 0.001
↪Contaminants	0.0%	0.3	14.3%	16.5	P < 0.001
↪Artifacts	42.9%	12.3	52.8%	23.6	P < 0.001
↪Chimeras	2.4%	2.2	13.6%	13.6	P < 0.001
↪Unidentified	40.0%	11.2	35.7%	13.2	P < 0.001

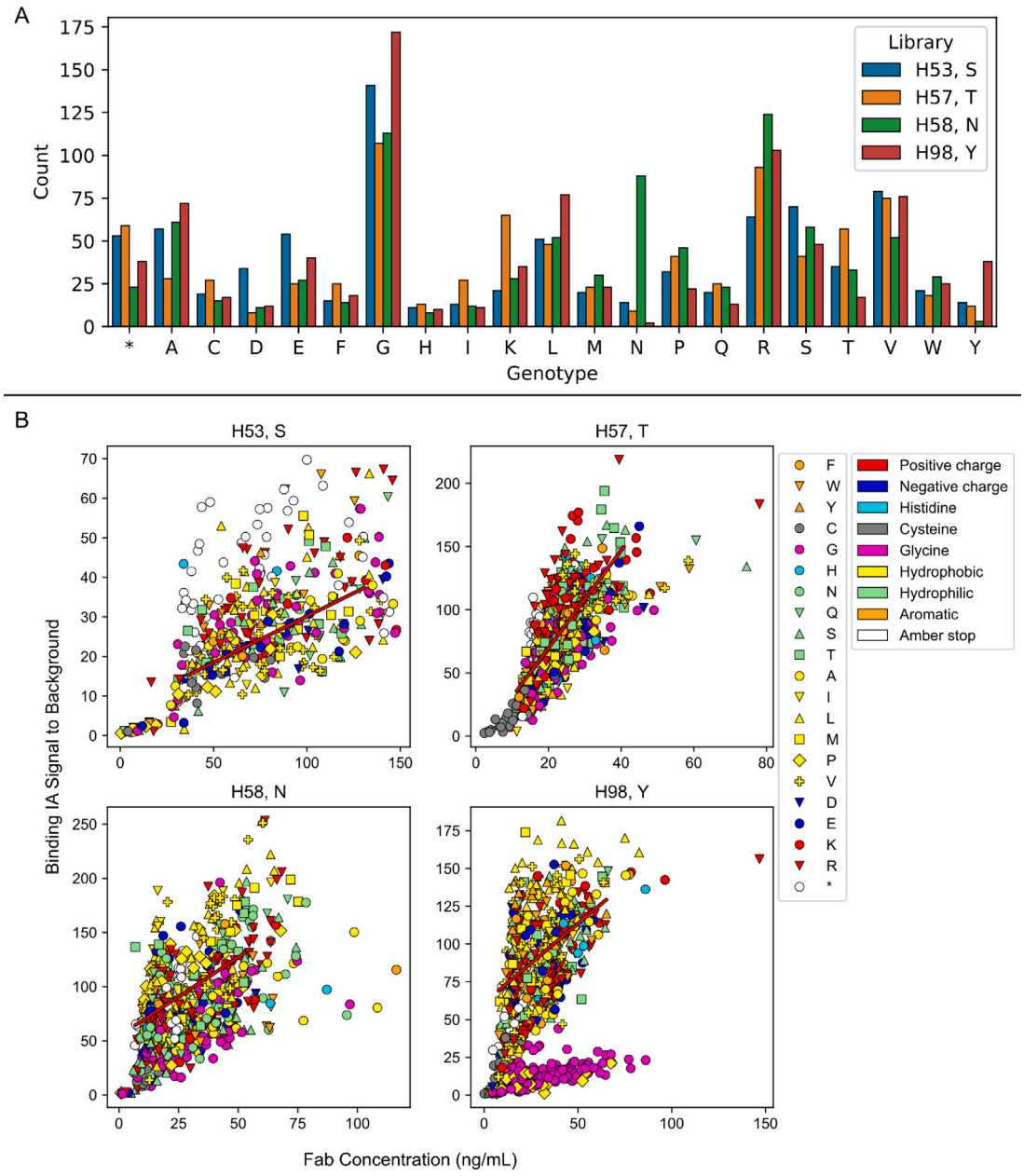

Fig. 5. Genotyped functional screening of four anti-DENV-2 NS1 sublibraries. (A) Genotype count distribution for sublibraries H53, H57, H58, and H98, represented in blue, orange, green, and red, respectively. The parental genotype for each sublibrary is indicated in the legend. (B) Functional screening results, illustrating the relationship between antigen binding and Fab concentration. Each subplot displays the corresponding sublibrary and parental genotypes at the top. Amino acids and their properties are color- and shape-coded as shown in the legend. A trend line for parental genotypes is represented by a red line.

3. Discussion

In this study, we present a versatile, streamlined version of our original genotyped functional screening platform, capable of linking genotype and functional data from thousands of screened proteins. Using a closed-tube PCR with 96 + 96 unique index primers, we drastically improved the sequence clarity, simplifying the identification of the true sequences from background noise. By adapting the universal approach proposed by Wittmann et al. [24], we made the barcoding adaptable to different templates; now, only two new primers are required, should the template sequence change. We removed the need for stepwise addition of the primers, reducing the risk of cross-contamination between PCR wells. This was achieved by designing primers with sufficiently high, i.e. 10 °C, T_m difference to enable consecutive target amplification and indexing in a closed-tube reaction. As the method yielded clear products of the expected size (Fig. 2), no further optimization was performed; however, a temperature gradient, or optimization of the touch-down PCR conditions may be beneficial for other templates to balance specificity and yield. Additionally, we distributed the full-length sequencing adapter sequences between the target and indexing primers, enabling direct generation of sequencing-ready libraries compatible with the Illumina MiSeq platform. The system streamlines sample indexing and mitigates the chimera formation by ensuring that only a single DNA template is present in each reaction.

Here we demonstrate a clear enhancement in sequence clarity, with the top-to-second sequence count ratio increasing from 29 in the quad-barcoding method to 82 in the improved, closed-tube method. In addition, we observed a substantial decrease in the number of total background sequences, dropping from 72% to 43%. More specifically, sample cross-contaminations or index hopping – where a true sequence from a neighboring well appears as a background sequence in another well – decreased from 14% to nearly negligible level. Chimera formation, defined as a recombination of two real sequences into hybrid artifacts, was reduced almost six-fold. Given that our sequencing library comprises Fab variable domain sequences with identical framework genes, it has inherently low diversity. The observed rates of chimeric and other background sequences align with those reported in other low-diversity libraries, such as the up to 70% rate of background sequence seen with 16S rRNA sequencing by Haas et al. [15]. The reduction in cross-contamination is likely due to both target amplification and indexing occurring in a closed-tube reaction, where only one DNA template is present. This setup, also utilized by Wittmann et al. [24], minimizes chimera formation, as chimeras typically form from the recombination of highly similar DNA molecules [14], which are effectively absent here. However, the formation of more complex chimeras cannot be excluded. Additionally, the new closed-tube method operates with fewer PCR cycles (10 + 25 cycles for the improved method compared to 35 + 8 cycles in the original), which by itself reduces the rate of chimera formation. This effect aligns with our previous findings, where fewer cycles effectively mitigated chimeric artifacts [2]. In our studies, the PCR wells were sealed with adhesive sealing tape instead of cap strips, which can leak if not applied with precision, leading to potential sample contamination. This could partly explain the low amounts of chimeric sequences observed with the closed-tube method.

Other strategies to further reduce background sequence pollution include using more thorough purification methods, such as polyacrylamide gel electrophoresis (PAGE), to eliminate non-bound sequencing adapters, which have been shown to contribute to index hopping [16]. The unbound adapters can also be blocked with Illumina Free Adapter Blocking Reagent [25] to prevent them from amplifying library sequences and leading to miss-assignment of the indices. Additionally, unique dual indexing (UDI) – where unique indices are added at both ends of the amplicon – can help differentiate true sequences from backround sequences [22]. However, while UDI can improve clarity, it does not address the root cause of these artifacts and is likely cost-prohibitive for highly multiplexed sequencing.

While index hopping and chimera formation are well-known issues in low-diversity library sequencing, they are easily overlooked. Particularly in applications such as antibody sequencing, where calculations of library diversity and enrichment are critical, it is essential to pay careful attention to this issue. Studies that utilize next-generation sequencing for antibody repertoire analysis often rely on sequence counts to assess diversity or enrichment [3,4]. However, these methods can be impacted by artifacts such as chimeras, which may lead to misleadingly inflated diversity estimates. For instance, without careful attention to these issues, a significant portion of sequences could represent artifacts rather than true unique variants, compromising the reliability of downstream analyses. Therefore, while NGS-based approaches have advanced antibody research, awareness and mitigation of these artifacts are essential to ensure accurate results. PCR free methods for indexing and library generation [26] should be considered.

In conclusion, we have successfully improved our original genotyped functional screening platform to more robust and yield better resolution between true and background sequences, making identification of true clone sequences with higher confidence possible. This enhanced sequence clarity suggests that the new closed-tube barcoding method provides more reliable and accurate sequencing data, making it a superior option for future applications. The universal set of 96 + 96 indexing primer set can be applied for barcoding other amplicons than Fabs as well, e.g. nanobodies [27] or DARPin molecules [28] by changing only the pair of high-T_m target-specific primers in the setup. The sequence clarity is now sufficient for increasing the sample size even further, however additional optimization for e.g. number of PCR cycles or sample purification methods could help to further reduce the number of background sequences.

4. Methods

4.1. Bacterial strains, vectors and cloning

We used *E. coli* XL-1 Blue (recA1 endA1 gryA96 thi-1 hsdR17 supE44 relA1 lac [F' proAB lacI q ZΔM15 Tn10 (Tet r)]) (Stratagene, USA) for soluble Fab expression with pAK400 vector [29]. For Sanger sequencing we cloned the indexed variable domain sequences to pUC19 vector (New England Biolabs, USA). For cloning, we used restriction enzymes SfiI, LguI, HindIII and BamHI and T4 DNA ligase according to the manufacturer's recommendations, all purchased from Thermo Fisher Scientific, USA. We conducted transformation to *E. coli* with electroporation using Gene Pulser Xcell (Bio-Rad, USA) at 1250 V, 25 μ F, 200 Ohm, with 1 mm Gene Pulse Cuvettes (Bio-Rad, USA). After electroporation, the cells recovered in 1 mL of SOC media (20 g/L Tryptone, 5 g/L Yeast extract, 0.5 g/L NaCl, 2.5 mM KCl, 10 μ M MgCl2, 3.6 g/L D-Glucose, pH 7.0) for 1 h at 37 °C with shaking at 300 rpm, and were inoculated to agar plates with 1% glucose and suitable antibiotics (10 μ g/mL tetracycline and 25 μ g/mL chloramphenicol for pAK400 or 100 μ g/mL of ampicillin for pUC19 vector). We purchased all used antibiotics from Sigma Aldrich (USA).

4.2. Antibody library construction

To assess performance of the new barcoding method, we reconstructed four anti-Dengue virus-2 nonstructural protein 1 (later referred as "anti-DENV-2 NS1") affinity maturation sub-libraries from our previous study using the same protocol [2]. Briefly, we used DNA from the anti-DENV-2 NS1 Fab 49A3 (discovered at the Department of Life Technologies, University of Turku) as a template to generate libraries, each with a single codon randomized in CDRH2 (Kabat positions H53, H57, and H58) or CDRH3 (Kabat position H98) using NNK primers ordered from Sigma Aldrich, USA. Our previous screening indicated that these positions might be the most conducive to beneficial mutations, making them prime targets for deeper investigation. We cloned the generated libraries directly into the pAK400 expression vector for screening. To

evaluate the platform with more diverse, phage display-selected Fab clones, we included five 96-well plates containing glycerol-stored *E. coli* expressing phage display selected anti-SpyCatcher Fabs, which we also studied previously. Detailed information on library generation, phage display selections and Fab conversions is found in our original publication [2].

For DNA purification (minipreps, gel extractions and PCR purification) we used GeneJet kits (Thermo Scientific, USA) according to the manufacturer's recommendations, unless otherwise stated. We bought Sanger sequencing as service from Macrogen Inc. (South Korea). To facilitate indexing of Fab variable domains with both well and plate barcode sequences in single PCR reaction, we adapted method and index sequences developed by Wittmann et al. [24] with some modifications. Firstly, to enable target amplification and indexing in same reaction and without need for subsequent primer addition, we developed target specific primers with 12 – 14 °C higher annealing temperature. Secondly, we changed the indexing adapter sequence to match our original Illumina TruSeq based scheme and split it between target and indexing primers to yield full length, sequencing ready amplicons without need for additional PCR or ligation steps. We designed the target specific primers to contain Fab variable domain hybridization sequence (heavy, T_a = 62 °C or light, T_a = 64 °C) and the first 34 bases of the adapter sequence as overhang. We then designed universal 96 + 96 well (forward) and plate (reverse) barcoding primers (Ta = 50 °C) with remaining sequencing adapter sequence, including unique eight base long index sequence. We ordered all primers from Integrated DNA Technologies (USA); target specific primers as 4 nmole Ultramer™DNA Oligos (initial testing was done with standard oligos) and index primers as standard oligos, prediluted to 100 μ M in IDT buffer pH 8.0 in 96 well plate format. All primer sequences are listed in Supplementary Table 1, found in Appendix A.

To validate the primers and assess viability of the developed closed-tube method, we compared it to the original two-step method, described by Wittmann et al. [24], by indexing variable heavy, light or both in same reaction from 1 ng of anti-DENV-2 NS1 49A3 Fab template DNA [2] in 10 μL PCR reactions. In closed-tube method, both primer pairs are present in the reaction from the beginning, whereas in two-step the universal indexing primers are added to the reaction after target amplification. In addition to a non-template control for both methods, we included a control without supplementation of the indexing primers. Both PCR reactions consisted of 0.025 U/μL FirePol DNA polymerase (Solis Biodyne, Estonia), 1 ×Buffer BD, 1.5 mM MgCl2, 200 μ M dNTP, 20 nM target specific and 200 nM universal indexing primers. We used touchdown PCR with 10 cycles of 95 °C (30 s), 64→54 °C (60 s, −1 °C/cycle) and 72 °C (30 s) with a 95 °C (5 min) initial denaturing at the beginning to amplify the variable domain sequences. The indexing PCR was initiated directly (closed-tube) or after supplementing the reaction with index primers (two-step) by lowering the annealing temperature to 50 °C and cycling for additional 25 cycles of 95 °C (30 s), 50 °C (60 s) and 72 °C (35 s) with a 72 °C (5 min) elongation step at the end. We evaluated the amplification with 1% agarose gel electrophoresis by running 5 μL from each reaction at 100 V for 40 min. Finally, we tested all the new 96 + 96 barcoding primers by amplifying V_H and V_L sequences of Fab 49A3 separately using the one-step protocol and running the same electrophoresis analysis as described above.

In addition to electrophoresis, we verified the success of the indexing with Sanger sequencing. Briefly, after closed-tube and two-step PCRs similar to above, we ran the samples on 1% agarose gel at 90 V for 1 h, followed by gel extraction of the correct sized DNA. We then amplified the DNA with Illumina adapter P5 and P7 specific primers to add BamHI and HindIII restriction sites (TH260_BamHI-P5: 5 'cggggatccAATGATACGGCGACCACCGAG 3' and TH261_HindIII-P7: 5' tgcaagcttCAAGCAGAAGACGGCATACGAGAT 3') using similar reaction mix for Fire Pol DNA Polymerase as before and thermal cycling of 25 cycles of 95 °C (30 s), 56 °C (60 s) and 72 °C (40 s) with a 95 °C (5 min) initial denaturing at the beginning and 72 °C (5 min) final elongation step at the end. Amplicons were purified with Zymo DNA Clean & Concentrator-5 kit (Zymo Research, USA) and cloned into pUC19 vector by digesting them with BamHI and HindIII, followed with gel extraction and finally ligation using T4 DNA ligase. We transformed the constructs directly using the ligation reactions into XL1-Blue *E .coli* by electroporation and after recovery plated the cells along with empty vector control as described earlier. Next morning, we inoculated three colonies from V_H, V_L and V_H+V_L indexed with both methods into 5 mL of SB media (30 g/L Tryptone, 20 g/L Yeast extract, 10 g/L MOPS, pH 7.0), allowed them to grown o/n, used them for minipreps and sent the DNA to sequencing with forward primer LMB: 5' ATGTGCTGCAAGGCGATTAAG 3'.

4.3. Expression and indexing of anti-DENV-2 NS1 and anti-SpyCatcher libraries

To compare the novel closed-tube barcoding method with our previously established quad-barcoding, we conducted genotyped screening of same Fab clone populations side by side with both methods. For this, we screened ten 96 well plates from each of the four anti-DENV-2 NS1 sublibraries (total of 3840 wells, including 120 control wells with Fab 49A3 and 120 empty control wells). To further validate the new closed-tube method with more sequence diverse clones, we also indexed and sequenced (but did not re-screen for antigen binding) clones from five anti-SpyCatcher plates (stored in glycerol at -70 °C) from our previous study (S002, S006, S009, S013 and S021). We selected the anti-SpyCatcher plates based on the high number of antigen binding clones identified in the original study.

For soluble Fab expression, we inoculated single colonies of XL-1 E. coli from agar plates to primary culture well on 96-well plate containing 200 μL SB media with 1% D-glucose and antibiotics, and incubated the cells overnight at 37 °C, 900 rpm. After incubation we inoculated 4 μL of cells from overnight culture into fresh 180 μL of SB supplemented with 0.05% D-glucose and antibiotics and incubated them again at 37 °C with standard shaking for 4 h. We stored the original culture plates at -70 °C after adding glycerol to final concentration of 16%. We induced Fab expression by addition of IPTG to final concentration of 200 μ M, and lowering temperature to 26 °C for 16 h. In the morning, we pelleted the cells by centrifugation at 4 °C, 3200 g for 30 min and then lyzed the pelleted cells by resuspending them in lysis buffer (1 mg/mL lysozyme from chicken egg white L6876 (Sigma Aldrich, United Kingdom), 0.025 U/μL Pierce Universal Nuclease (Thermo Fisher Scientific, USA) in PBS (20 mM sodium phosphate, 300 mM sodium chloride, pH 7.4), incubating them 30 min in room temperature in slow agitation, and finally by freezing the lysate at -70 °C. Lysate was clarified by centrifugation before the immunoassays.

We carried out the Fab variable heavy domain indexing either using overnight cultured cells (quad-barcoding for anti-DENV-2 NS1 clones) or using glycerol-stored cells (both versions of indexing for anti-SpyCatcher and the closed-tube for anti-DENV-2 NS1 clones). For both barcoding methods, we pipetted 1 μL of cells from either overnight culture or glycerol stored plates to 10 μL PCR reactions on 96 well plates. The quad-barcoding PCR was done with similar protocol as in the original publication [2]. The well barcoding PCR consisted of 0.025 U/μL FirePol DNA polymerase, 1x Buffer BD, 1.5 mM MgCl2, 200 μ M dNTP and 200 nM custom indexing primers. For thermocycling we used 35 cycles of 95 °C (30 s), 56 °C (60 s) and 72 °C (60 s) with a 95 °C (15 min) initial denaturing at the beginning and 72 °C (4 min) final elongation step at the end. We pooled 5 μL from each well and batch PCR purified the DNA. We did the subsequent plate barcoding PCR in 20 μL volume with 2 ng of template DNA, 0.02 U/μL Q5 High-Fidelity DNA Polymerase (New England Biolabs, USA), 1x Q5 Reaction Buffer, 200 μ M dNTP, and 0.5 μ M of both primer from TruSeq Custom Amplicon kit (Illumina inc., USA). Cycling was 8

cycles of 98 °C (10 s), 67 °C (15 s) and 72 °C (30 s) with a 98 °C (3 min) initial denaturing at the beginning and 72 °C (4 min) final elongation step at the end. We then pooled 15 μL from each sample and purified the DNA in batch, similarly to above. The remaining PCR product we used as quality control and used for electrophoresis analysis by running on 1% agarose gel at 70 V for 90 min. We carried out the novel closed-tube barcoding PCR with similarly as stated in previous section describing the method validation. After the reaction on the 96 well plate, we pooled 5 μL from each well together and batch purified the DNA as with original method. We further pooled the fully indexed, sequencing ready amplicons generated with the new method in equal weight ratios of DNA.

For library size selection, we gel extracted the fully indexed amplicons from 2% agarose gel, after running electrophoresis at 60 V for 2 h and combined the libraries in equal molar ratios. Sequencing was bought as service from Finnish Functional Genomic Center (Turku, Finland), where library quality was first assessed using 2100 Bioanalyzer with High Sensitivity DNA Assay kit (Agilent, USA) and quantified using Qubit (Thermo Fisher Scientific, USA). 9 pM library supplemented with 10% PhiX Sequencing Control V3 (Illumina, USA) was sequenced on MiSeq using MiSeq Reagent Kit v3 (600-cycle) (Illumina).

4.4. Functional screening of the anti-NS1 fab clones

To obtain normalized functional data, we analyzed both antigen binding and Fab expression from the lysates in parallel time-resolved fluorescence (TRF) based immunoassays. For both assays we used Goat anti-Human (Fab specific, later referred as "GAH") IgG (Sigma Aldrich, USA) coated plates. GAH plates were prepared by passive coating on yellow C12 MaxiSorp plates (Thermo Fisher Scientific, USA), described in detail in supplementary methods in Appendix A. All dilutions for immunassays were done using Assay Buffer Red and washes using Uniogen Wash buffer, both purchased from Uniogen (Finland). Delfia Plate Wash (Wallac, Turku Finland) instrument was used for plate washing. All incubations were done at room temperature with slow shaking. Immunoassay plates were read with Hidex Sense Multimode plate reader (Hidex, Finland). As an antigen for antigen binding immunoassay we used biotinylated DENV-2 NS1 (later referred as "bio-NS1") that we purchased from The Native Antigen Company (UK) and biotinylated with EZ-Link NHS-PEG4-Biotin (Thermo Fisher Scientific, USA) with 12×molar excess of biotin, according to the manufacturers protocol.

In antigen binding assay, we first added 100 μL of thawed and cleared lysates with 1:5 dilution into prewashed GAH plate wells and incubated plates for one hour. After washing the wells two times, we supplemented the wells with 100 μL of bio-NS1 diluted to 36 ng/mL (EC_{20} of the parental 49A3 Fab) and allowed the plates to incubate for another hour, followed by washing the wells two times. To detect the bound antigen, we added 100 μL of filtered (∅ 0.22 μ m), N1-Europium-labeled streptavidin (labeled at Department of Life Technologies, University of Turku) diluted to 200 ng/mL per well, incubated the reactions for 15 min and washed the wells again twice. After 10 min incubation with supplemented 200 μL of Delfia Enhancement solution per well the TRF signals were measured using Hidex Sense plate reader. For each plate we conducted the expression assay in parallel with similar protocol to above, except after washing off the unbound Fab from GAH plates, 100 μL filtered, N1-Europium-labeled anti-hFab IgG 2A11 (purchased from Hytest, Finland and labeled at Department of Life Technologies, University of Turku) diluted to 200 ng/mL per well. We then incubated the wells for one hour, washed them twice and measured the TRF after addition of DES, similarly to antigen binding assay. We performed the same expression assay each day of screening in parallel as triplicates for purified Fab 49A3 with concentrations of 0, 0.41, 1.23, 11.11, 33.33, 100 and 300 ng/mL to convert the measured TRF into actual Fab concentrations for normalization of the antigen binding data.

4.5. Data analysis

We used FASTQC v0.11.9 [30] in conjunction with MultiQC v1.13 [31] to assess the quality of FASTQ files after each analysis step. Base calling was done using bcl2fastq v2 software to include the index sequences in FASTQ headers. This was done because we were using two different lengths of index sequences (7 and 8 nucleotides in closed-tube and quad-barcoding, respectively) and thus the Illumina software could not directly demultiplex all reads. The V_H sequences from closed-tube method were demultiplexed to separate files for each clone manually using Python v3.10.14 with a custom script utilizing regular expressions. We then merged and pre-filtered the read pairs using PEAR v0.9.11 [32] with parameters -v 25 -m 600 -n 300 -q 20 -y 8G -j 12. Plate barcodes (TruSeq indexes from quad-barcoding) were automatically demultiplexed by Illumina software. To demultiplex the well barcodes and trim the sequences to contain only the desired V_H sequence, we utilized another custom Python script, which looped through the merged FASTQ files while matching the sequences to correct samples using regular expression patterns. We trimmed the sequences from closed-tube method with the same script. As a result, files were converted into FASTA format. We then calculated the unique sequence counts with yet another Python script and stored the total, top, and second most abundant read counts along with the sequences of latter two to a data frame. We used the ratios of these sequence counts for filtering out samples with low sequence count or unclear primary sequence, similarly to what we did in our previous study [2], as shown in Formula (1) and Formula (2). Finally, we translated the DNA sequences to amino acid sequences and Kabat numbered the antibody residues with ANARCI [33] via Python api Abnumber v0.3.7. Similarly to before, we discarded the sequences not recognized as immunoglobulin sequences by ANARCI, in addition to sequences with different lengths from parental. The Python and UNIX codes and scripts describing the data analysis can be found from GitHub (https://github.com/sjoutu/closed-tube-barcoding/).

$$Filter\ 1: \frac{\text{Top sequence count}_{\text{SAMPLE}}}{\text{Total sequence count}_{\text{SAMPLE}}} > \frac{\text{Top sequence count}_{\text{EMPTY}}}{\text{Total sequence count}_{\text{EMPTY}}} \quad (1)$$

$$Filter\ 2: \frac{\text{Top sequence count}_{\text{SAMPLE}}}{\text{Second sequence count}_{\text{SAMPLE}}} > 2 \quad (2)$$

To identify the origin of background sequences and compare the two methods, we conceived a Python script to categorize these sequences. First, we compiled a list of top sequences from anti-SpyCatcher samples that passed the quality filters. This list served as a reference for the categorization process. The script then iterated through the sequences of each sample, classifying them first as contaminant or background artifact sequences and the latter further as verified chimeras or unidentified sequences. A sequence was labeled as a contaminant if it had an exact match in the top sequence list. If no exact match was found, the resulting background sequence was split into two at all possible positions using a sliding window approach. If both segments of the split sequence matched entries in the top sequence list, it was classified as a verified chimera. Sequences that did not fit either condition were labeled as unidentified sequence. The unidentified sequences include single nucleotide errors and possibly more complex chimeras, formed from multiple DNA molecules. After classifying the background sequences, we calculated their proportions relative to the total sequence count for each sample. The medians of these proportions were then compared between the closed-tube and quad-barcoding methods using the Mann-Whitney U test.

We analyzed the functional immunoassay data using Python. For each sample passing the filters, signal to background values of both antigen binding and Fab concentration TRF was calculated by dividing the measured time-resolved fluorescence counts with the counts from empty well control of the same plate. We fitted four parameter logistic

curves on the concentration immunoassay standards and used the resulting function to resolve the Fab concentrations of each screened Fab clones. The genotypes of each filter passing clones were identified using ANARCI. To analyze the success of the site-saturation mutagenesis library generation, we counted the unique genotypes of each library. Genotype counts and the genotyped functional screening results of the anti-NS1 libraries were plotted using seaborn.

CRediT authorship contribution statement

Sami Oksanen: Writing – review & editing, Writing – original draft, Visualization, Software, Methodology, Investigation, Formal analysis, Data curation, Conceptualization. **Tuomas Huovinen:** Writing – review & editing, Supervision, Conceptualization. **Urpo Lamminmäki:** Writing – review & editing, Supervision, Resources, Conceptualization.

Declaration of competing interest

The authors declare the following financial interests/personal relationships which may be considered as potential competing interests: Sami Oksanen reports financial support was provided by Emil Aaltonen Fundation. If there are other authors, they declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgments

This study was funded by the University of Turku Graduate School (UTUGS), Emil Aaltosen Säätiö sr (stipend number: 240140) and the Research Council of Finland's Flagship InFLAMES (decision numbers: 337530, 357910 and 358823). The Illumina MiSeq sequencing was performed in Finnish Functional Genomics Centre supported by University of Turku, Åbo Akademi University and Biocenter Finland. We wish to acknowledge CSC – IT Center for Science, Finland, for computational resources used during this study.

Appendix A. Supplementary data

Supplementary material related to this article can be found online at https://doi.org/10.1016/j.slast.2026.100429.

Data availability

The raw NGS data generated and analyzed in this study have been deposited in the European Nucleotide Archive (ENA) at EMBL-EBI under accession number PRJEB83324 (https://www.ebi.ac.uk/ena/browser/view/PRJEB83324). The UNIX and Python scripts can be found from authors github page at https://github.com/sjoutu/closed-tube-barcoding/.

References

[1] Fahad AS, Timm MR, Madan B, Burgomaster KE, Dowd KA, Normandin E, Gutiérrez-González MF, Pennington JM, Souza MOD, Henry AR, Laboune F, Wang L, Ambrozak DR, Gordon IJ, Douek DC, Ledgerwood JE, Graham BS, Castilho LR, Pierson TC, Mascola JR, DeKosky BJ. Functional profiling of antibody immune repertoires in convalescent zika virus disease patients. Front Immunol 2021;12. URL: https://www.frontiersin.org/journals/immunology/articles/10.3389/fimmu.2021.615102.

[2] Oksanen S, Saarinen R, Korkiakoski A, Lamminmäki U, Huovinen T. Genotyped functional screening of soluble fab clones enables in-depth analysis of mutation effects. Sci Rep | 2023;13:13107. http://dx.doi.org/10.1038/s41598-023-40241-2, URL: https://doi.org/10.1038/s41598-023-40241-2.

[3] Glanville J, Zhai W, Berka J, Telman D, Huerta G, Mehta GR, Ni I, Mei L, Sundar PD, Day GM, Cox D, Rajpal A, Pons J. Precise determination of the diversity of a combinatorial antibody library gives insight into the human immunoglobulin repertoire. Proc Natl Acad Sci USA 2009;106:20216–21. http://dx.doi.org/10.1073/PNAS.0909775106, URL: https://pubmed.ncbi.nlm.nih.gov/19875695/.

[4] Ravn U, Gueneau F, Baerlocher L, Osteras M, Desmurs M, Malinge P, Magistrelli G, Farinelli L, Kosco-Vilbois MH, Fischer N. By-passing in vitro screening—next generation sequencing technologies applied to antibody display and in silico candidate selection. Nucleic Acids Res 2010;38:e193. http://dx.doi.org/10.1093/NAR/GKQ789, https://www.ncbi.nlm.nih.gov/pmc/articles/PMC2995085/.

[5] Abanades B, Wong WK, Boyles F, Georges G, Bujotzek A, Deane CM. ImmuneBuilder: Deep-learning models for predicting the structures of immune proteins. Commun Biology 2023 6:1 2023;6:1–8. http://dx.doi.org/10.1038/s42003-023-04927-7, URL: https://www.nature.com/articles/s42003-023-04927-7.

[6] Jumper J, Evans R, Pritzel A, Green T, Figurnov M, Ronneberger O, Tunyasuvunakool K, Bates R, Žídek A, Potapenko A, Bridgland A, Meyer C, Kohl SA, Ballard AJ, Cowie A, Romera-Paredes B, Nikolov S, Jain R, Adler J, Back T, Petersen S, Reiman D, Clancy E, Zielinski M, Steinegger M, Pacholska M, Berghammer T, Bodenstein S, Silver D, Vinyals O, Senior AW, Kavukcuoglu K, Kohli P, Hassabis D. Highly accurate protein structure prediction with AlphaFold. Nat 2021 596:7873 2021;596:583–9. http://dx.doi.org/10.1038/s41586-021-03819-2, URL: https://www.nature.com/articles/s41586-021-03819-2.

[7] Wang Q, Feng Y, Wang Y, Li B, Wen J, Zhou X, Song Q. AntiFormer: graph enhanced large language model for binding affinity prediction. Brief Bioinform 2024;25. http://dx.doi.org/10.1093/BIB/BBAE403.

[8] Smith AM, Heisler LE, St.Onge RP, Farias-Hesson E, Wallace IM, Bodeau J, Harris AN, Perry KM, Giaever G, Pourmand N, Nislow C. Highly-multiplexed barcode sequencing: an efficient method for parallel analysis of pooled samples. Nucleic Acids Res 2010;38:e142. http://dx.doi.org/10.1093/NAR/GKQ368.

[9] Craig DW, Pearson JV, Szelinger S, Sekar A, Redman M, Corneveaux JJ, Pawlowski TL, Laub T, Nunn G, Stephan DA, Homer N, Huentelman MJ. Identification of genetic variants using bar-coded multiplexed sequencing. Nat Methods 2008 5:10 2008;5:887–93. http://dx.doi.org/10.1038/nmeth.1251, URL: https://www.nature.com/articles/nmeth.1251.

[10] Glenn TC, Nilsen RA, Kieran TJ, Sanders JG, Bayona-Vásquez NJ, Finger JW, Pierson TW, Bentley KE, Hoffberg SL, Louha S, Leon FJG-D, Portilla MADR, Reed KD, Anderson JL, Meece JK, Aggrey SE, Rekaya R, Alabady M, Belanger M, Winker K, Faircloth BC. Adapterama I: Universal stubs and primers for 384 unique dual-indexed or 147,456 combinatorially-indexed illumina libraries (itru & inext). PeerJ 2019;2019. http://dx.doi.org/10.7717/PEERJ.7755/SUPP-31,.

[11] Glenn TC, Pierson TW, Bayona-Vásquez NJ, Kieran TJ, Hoffberg SL, Thomas JC, Lefever DE, Finger JW, Gao B, Bian X, Louha S, Kolli RT, Bentley KE, Rushmore J, Wong K, Shaw TI, Rothrock MJ, McKee AM, Guo TL, Mauricio R, Molina M, Cummings BS, Lash LH, Lu K, Gilbert GS, Hubbell SP, Faircloth BC. Adapterama II: Universal amplicon sequencing on illumina platforms (TaggiMatrix). PeerJ 2019;2019:e7786. http://dx.doi.org/10.7717/PEERJ.7786/SUPP-8, URL: https://peerj.com/articles/7786.

[12] Sinha R, Stanley G, Gulati GS, Ezran C, Travaglini KJ, Wei E, Chan CKF, Nabhan AN, Su T, Morganti RM, Conley SD, Chaib H, Red-Horse K, Longaker MT, Snyder MP, Krasnow MA, Weissman IL. Index switching causes "spreading-of-signal" among multiplexed samples in illumina HiSeq 4000 DNA sequencing. BioRxiv 2017;125724. http://dx.doi.org/10.1101/125724, https://www.biorxiv.org/content/10.1101/125724v1 https://www.biorxiv.org/content/10.1101/125724v1.abstract.

[13] Meyerhans A, Vartanian JP, Wain-Hobson S. DNA recombination during PCR. Nucleic Acids Res 1990;18:1687–91. http://dx.doi.org/10.1093/NAR/18.7.1687, URL: https://academic.oup.com/nar/article/18/7/1687/1096541.

[14] Brakenhoff RH, Schoenmakers JG, Lubsen NH. Chimeric cDNA clones: a novel PCR artifact.. Nucleic Acids Res 1991;19:1949. http://dx.doi.org/10.1093/NAR/19.8.1949, URL: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC328135/.

[15] Haas BJ, Gevers D, Earl AM, Feldgarden M, Ward DV, Giannoukos G, Ciulla D, Tabbaa D, Highlander SK, Sodergren E, Methé B, DeSantis TZ, Petrosino JF, Knight R, Birren BW. Chimeric 16s rRNA sequence formation and detection in sanger and 454-pyrosequenced PCR amplicons. Genome Res 2011;21:494–504. http://dx.doi.org/10.1101/GR.112730.110, https://genome.cshlp.org/content/21/3/494.full https://genome.cshlp.org/content/21/3/494 https://genome.cshlp.org/content/21/3/494.abstract.

[16] Costello M, Fleharty M, Abreu J, Farjoun Y, Ferriera S, Holmes L, Granger B, Green L, Howd T, Mason T, Vicente G, Dasilva M, Brodeur W, DeSmet T, Dodge S, Lennon NJ, Gabriel S. Characterization and remediation of sample index swaps by non-redundant dual indexing on massively parallel sequencing platforms. BMC Genomics 2018;19:1–10. http://dx.doi.org/10.1186/S12864-018-4703-0/TABLES/2, URL: https://bmcgenomics.biomedcentral.com/articles/10.1186/s12864-018-4703-0.

[17] Lin S, Wu Y-S, Gunderson K, Moon JA. US20120316086a1 - patterned flow-cells useful for nucleic acid analysis - google patents. 2012, URL: https://patents.google.com/patent/US20120316086A1/en.

[18] Shen M-JR, Boutell JM, Stephens KM, Ronaghi M, Gunderson KL, VENKATESAN BM, Bowen MS, Vijayan K. WO2013188582a1 - kinetic exclusion amplification of nucleic acid libraries - google patents. 2013, URL: https://patents.google.com/patent/WO2013188582A1/enIt.pdf.

[19] van der Valk T, Vezzi F, Ormestad M, Dalén L, Guschanski K. Index hopping on the illumina hiseqx platform and its consequences for ancient DNA studies. Mol Ecol Resour 2020;20:1171–81. http://dx.doi.org/10.1111/1755-0998.13009, https://onlinelibrary.wiley.com/doi/full/10.1111/1755-0998.13009 https://onlinelibrary.wiley.com/doi/abs/10.1111/1755-0998.13009 https://onlinelibrary.wiley.com/doi/10.1111/1755-0998.13009.

[20] Nelson MC, Morrison HG, Benjamino J, Grim SL, Graf J. Analysis, optimization and verification of illumina-generated 16s rrna gene amplicon surveys. PLoS One 2014;9:e94249. http://dx.doi.org/10.1371/JOURNAL.PONE.0094249, URL: https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0094249.

[21] Kircher M, Sawyer S, Meyer M. Double indexing overcomes inaccuracies in multiplex sequencing on the illumina platform. Nucleic Acids Res 2012;40:e3. http://dx.doi.org/10.1093/NAR/GKR771.

[22] MacConaill LE, Burns RT, Nag A, Coleman HA, Slevin MK, Giorda K, Light M, Lai K, Jarosz M, McNeill MS, Ducar MD, Meyerson M, Thorner AR. Unique, dual-indexed sequencing adapters with UMIs effectively eliminate index cross-talk and significantly improve sensitivity of massively parallel sequencing. BMC Genomics 2018;19:1–10. http://dx.doi.org/10.1186/S12864-017-4428-5/FIGURES/6, URL: https://bmcgenomics.biomedcentral.com/articles/10.1186/s12864-017-4428-5.

[23] Illumina Inc. Index hopping | intro & how to minimize it. 2024, URL: https://assets.illumina.com/content/illumina-marketing/en/techniques/sequencing/ngs-library-prep/multiplexing/index-hopping.html.

[24] Wittmann BJ, Johnston KE, Almhjell PJ, Arnold FH. Evseq: Cost-effective amplicon sequencing of every variant in a protein library. ACS Synth Biology 2022;11:1313–24. http://dx.doi.org/10.1021/ACSSYNBIO.1C00592/ASSET/IMAGES/LARGE/SB1C00592_0005.JPEG, URL: https://pubs.acs.org/doi/full/10.1021/acssynbio.1c00592.

[25] Illumina Inc. Illumina free adapter blocking reagent. 2024, URL: https://support.illumina.com/sequencing/sequencing_kits/illumina-free-adapter-blocking-reagent.html.

[26] Fantini M, Pandolfini L, Lisi S, Chirichella M, Arisi I, Terrigno M, Goracci M, Cremisi F, Cattaneo A. Assessment of antibody library diversity through next generation sequencing and technical error compensation. PLoS One 2017;12:e0177574. http://dx.doi.org/10.1371/JOURNAL.PONE.0177574, URL: https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0177574.

[27] Salvador JP, Vilaplana L, Marco MP. Nanobody: outstanding features for diagnostic and therapeutic applications. Anal Bioanal Chem 2019;411:1703–13. http://dx.doi.org/10.1007/S00216-019-01633-4/FIGURES/3, URL: https://link.springer.com/article/10.1007/s00216-019-01633-4.

[28] Binz HK, Stumpp MT, Forrer P, Amstutz P, Plückthun A. Designing repeat proteins: Well-expressed, soluble and stable proteins from combinatorial libraries of consensus ankyrin repeat proteins. J Mol Biol 2003;332:489–503. http://dx.doi.org/10.1016/S0022-2836(03)00896-9.

[29] Krebber A, Bornhauser S, Burmester J, Honegger A, Willuda J, Bosshard HR, Plückthun A. Reliable cloning of functional antibody variable domains from hybridomas and spleen cell repertoires employing a reengineered phage display system. J Immunol Methods 1997;201:35–55. http://dx.doi.org/10.1016/S0022-1759(96)00208-6, https://pubmed.ncbi.nlm.nih.gov/9032408/ https://pubmed.ncbi.nlm.nih.gov/9032408/?itool=EntrezSystem2.PEntrez.Pubmed.Pubmed_ResultsPanel.Pubmed_RVDocSum&ordinalpos=8.

[30] Andrews S. FastQC: A quality control tool for high throughput sequence data. 2010, [Online]. Available online at: http://www.bioinformatics.babraham.ac.uk/projects/fastqc/, https://Www.Bioinformatics.Babraham.Ac.Uk/Projects/Fastqc/.

[31] Ewels P, ans Magnusson M, Lundin S, Käller M. Multiqc: summarize analysis results for multiple tools and samples in a single report. Bioinformatics 2016;32:3047–8. http://dx.doi.org/10.1093/BIOINFORMATICS/BTW354, URL: https://app.dimensions.ai/details/publication/pub.1010049881.

[32] Zhang J, Kobert K, Flouri T, Stamatakis A. PEAR: a fast and accurate illumina paired-end read merger. Bioinformatics 2014;30:614. http://dx.doi.org/10.1093/BIOINFORMATICS/BTT593, https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3933873/.

[33] Dunbar J, Deane CM. ANARCI: antigen receptor numbering and receptor classification. Bioinformatics 2016;32:298–300. http://dx.doi.org/10.1093/BIOINFORMATICS/BTV552, URL: https://academic.oup.com/bioinformatics/article/32/2/298/1743894.