Pairing of T-cell receptor chains via emulsion PCR

Turchaninova M.A.[1]*, Britanova O.V.[1]*, Bolotin D.A.[1]*, Shugay M.[1]*, Putintseva E.V.[1], Staroverov D.B.[1], Sharonov G.[1], Shcherbo D.[1,2], Zvyagin I.V[1,4], Mamedov I.Z.[1,4], Linnemann C.[3], Schumacher T.N.[3], Chudakov D.M. [1,4]**

[1]Shemiakin-Ovchinnikov Institute of Bioorganic Chemistry, Russian Academy of Science, Moscow, Russia

[2]Evrogen JSC, Moscow, Russia;

[3]Division of Immunology, The Netherlands Cancer Institute, Amsterdam, The Netherlands;

[4]CEITEC MU, Masaryk University, Brno, Czech Republic.

*These authors contributed equally to this work.

**Corresponding author: Chudakov D.M., Shemiakin-Ovchinnikov Institute of Bioorganic Chemistry, RAS, Miklukho-Maklaya 16/10, 117997, Moscow, Russia; E-mail: ChudakovDM@mail.ru

Fax: +7 (499) 330-70-56

Abstract

Our ability to analyze adaptive immunity and engineer its activity has long been constrained by our limited ability to identify native pairs of heavy-light antibody chains and alpha-beta T-cell receptor (TCR) chains – both of which comprise coupled "halves of a key", collectively capable of recognizing specific antigens. Here we report a cell-based emulsion RT-PCR approach that allows the selective fusion of the native pairs of amplified TCR alpha and beta chain genes for complex samples. A new type of PCR suppression technique was developed that makes it possible to amplify the fused library with minimal noise for subsequent analysis by high-throughput paired-end Illumina sequencing. With this technique, single analysis of a complex blood sample allows identification of multiple native TCR chain pairs. This approach may be extended to identify native antibody chain pairs and, more generally, pairs of mRNA molecules that are co-expressed in the same living cells.

Introduction

Antibodies and T-cell receptors (TCRs), the weapons of adaptive immunity capable of selective recognition of specific antigens, represent an invaluable resource for biological studies and medical applications [1, 2]. Analysis of the native TCR and antibody repertoires is also of fundamental importance for our understanding of adaptive immunity in health and disease [3]. Because of this, methods for efficient identification of their functional units - native pairs of heavy/light antibody or alpha/beta TCR chains - have been highly desired for decades.

A series of different approaches have been developed to this end, including hybrid cells [4], single cell PCR [5-7] and frequency-based pairing [8]. Native pairs of TCR or antibody chains can also be identified after culturing of lymphocyte clones [5], or sorting of narrow antigen-specific populations of T cells [9] or B cells [10]. While these approaches are feasible to identify native chain pairs, they can only be performed serially for a limited number of clones, thereby prohibiting comprehensive analysis of most biological samples. Phage and yeast display technologies [11-13], although efficient for isolation of antigen-specific antibodies, rely on random pairing and do not provide information on the native pairs of chains. Furthermore, those approaches that could yield massive output based on the amplification and assembly of genes in fixed cells [14, 15] have not demonstrated feasibility for the analysis of complex samples. Recently, a microwell plate assay was reported to pair human immunoglobulin heavy and light chains for high-throughput sequencing [16]. However, this assay is quite laborious and requires the fabrication of custom chips, while its efficiency is not fully clear.

Here we report a new approach for identifying TCR alpha-beta pairing, based on specific reverse-transcription of TCR alpha and beta chain mRNA, PCR amplification and subsequent coupling via overlap-extension [17] — all performed within emulsion droplets that each contain a single T cell. We overcame the primary obstacle of random, non-specific overlap-extension during post-emulsion amplification reactions with a new PCR-suppression technique that we invented to specifically and efficiently block the 3' ends of the free, non-overlapped alpha and beta chain PCR products. Notably, the whole technology is simple to implement and does not require any special equipment.

Results and Discussion

Reverse transcription and amplification from cells in emulsion

The key requirement for the success of our approach is the ability to generate a representative library of TCR alpha and beta genes that are fused within emulsion reactions, wherein individual T cells are each contained within a separate minute droplet of sufficient reaction volume (**Fig. 1A**).

We employed a strategy of starting from RNA, which benefits from a high abundance of molecules encoding TCR alpha and beta chains and universal primer annealing sites on the constant regions of the TCR chains [18, 19]. The emulsion PCR protocol based on ABIL EM 90 surfactant was adapted from Ref. [20], with revisions to allow for single-stage, cell-based reverse transcription and amplification reactions. Control experiments with DAPI-stained cells and a mixture of green and red fluorescent-labeled emulsions demonstrated high emulsion stability (<0.1% fusion of droplets after thermocycling), sufficient abundance of droplets for our reaction (millions), and an undetectable percentage of droplets containing more than one cell (**Materials and Methods**).

We designed a multiplex primer mix (**Supporting Information Table 1**) to prime cDNA synthesis, amplify and then fuse via overlap-extension [17] multiple TCR alpha and TRBV7-family beta genes, starting from T cells (**Fig. 1B**).

Predictably, the immediate product of cell-based RT-PCR emulsion reactions represented a complex mixture of PCR products, in which overlap-extended products of interest (*i.e.*, fused TCR alpha and beta chain genes) were undetectable. Two sequential nested PCR amplification reactions [21] successfully yielded overlap-extended PCR product (**Supporting Information Fig. 1**). However, PCR amplification of a control mix of TCR alpha and beta chain genes that were independently amplified from PMBC in two separate emulsion reactions produced fused product with the same efficiency (**Fig. 1D, lanes 1** and **2**). Thus, random overlap-extension of abundant TCR alpha and beta amplicons released from emulsion was apparently predominant, resulting in loss of cell-specific information amidst random noise.

Noise-free amplification of paired TCR genes using a new type of PCR suppression

To prevent random overlap-extension and mega-priming by non-fused TCR alpha and beta PCR products after extraction from emulsion, it was necessary to find a way to efficiently block their 3'-ends. We made several unsuccessful attempts toward this end, using terminal deoxynucleotidyl transferase (TdT) to non-specifically elongate 3' ends, biotin labeling of internal

overlapping primers to filter non-overlapped products, increased primer annealing temperatures during nested amplification and dilution of emulsion-derived PCR products. None of these procedures proved sufficient, yielding no more than a three-cycle delay in the appearance of random background signal for treated overlap-extension PCR reactions relative to untreated reactions (data not shown).

Finally, we developed the new and efficient PCR-suppression approach described schematically in **Fig. 1C**. To "kick out" the 3' ends of undesirable DNA molecules, we added an excess of oligonucleotide that is complimentary to the 3' end, and also provides a template for further elongation with seven nonsense nucleotides. As a consequence, elongated products lose homology to their counterpart molecule and thus cannot enter the overlap-extension reaction efficiently.

The presence of 3.2 µM of two such blocking oligonucleotides complementary to the 3'-ends of the overlapping TCR alpha and beta PCR products efficiently suppressed amplification when we combined the products of two separate emulsion reactions, each performed with oligonucleotide sets for either TCR alpha or beta amplification (**Fig. 1D, lane 5**). At the same time, amplification starting from the product of a full emulsion reaction performed with all oligonucleotides for alpha and beta amplification and overlap-extension remained efficient (**Fig. 1D, lane 4**). Higher concentrations of blocking oligonucleotides suppressed amplification of the pre-fused product of interest (**Supporting Information Fig. 2**), and we used a concentration of 3.2 µM for subsequent experiments.

This new type of PCR suppression thus solves the most challenging issue of overlap-extension-based approaches for identifying co-expressed gene pairs: the high level of noise resulting from random overlap-extension after the emulsion stage. Due to this innovation, the library of paired TCR alpha and beta chains produced within each emulsion can be further selectively amplified and analyzed using next-generation sequencing.

As an additional control for emulsion stability, we performed experiments in which two separate emulsions containing oligonucleotide mix for amplification of either TCR alpha or beta genes were combined after reverse transcription (so that the amplification step was performed for the combined emulsions). In this setup, the inclusion of blocking oligonucleotides similarly suppressed the generation of fused product, indicating negligible fusion of cell-containing droplets during amplification (**Fig. 1D lane 6**).

Identification of native TCR chain pairs within a human blood sample

To demonstrate the power of the approach, we performed eight independent emulsion reactions each starting from 1 x 10^6 peripheral blood mononuclear cells (PBMCs) obtained from a human donor. To provide an internal control, we added approximately 1,000 cells of a cultured T-cell clone (C1A) with known TCR alpha and beta chains to three emulsion reactions. In addition, we performed two control reactions with cells without emulsion in order to provide statistics for the random pairing of TCR alpha and beta chains in solution. We further amplified the resulting PCR products with our new PCR-suppression method and analyzed the amplicons by paired-end Illumina 2 x 150 bp sequencing. In order to extract CDR3 alpha and beta information and correct for PCR errors, we subsequently analyzed the raw Illumina data with recently described TCR analysis approaches [19].

Hundreds of CDR3 pairs were selectively and prominently enriched in the emulsion reactions, indicating specific pairing, in contrast to the control reaction where the relative abundance of pairs was nearly equal to the frequency-expected random pairing (**Fig. 2**, **Supporting Information Fig. 3**). Of those, multiple TCR alpha-beta pairs were selectively enriched within several emulsion reactions (**Table 1**), providing independent cross-confirmation. Notably, there were no contradictory cross-pairings between these repeatedly identified pairs, which generally confirms the accuracy of the whole analysis pipeline. More than 700 alpha-beta pairs were unambiguously identified in single emulsion reactions (**Supporting Information Table 2, Materials and Methods**).

An independent set of experiments revealed two major TCR alpha-beta pairs for sorted CMV (NLV)-specific CD8[+] T cells obtained from the same donor (**Supporting Information Fig. 4, Supporting Information Table 3, Materials and Methods**). Both of these pairs, P1 and P4, were successfully identified in multiple emulsion reactions (**Table 1**). We also identified our internal control clone C1A in two of the three emulsion reactions, indicating the approximate limits of the method's sensitivity in its current implementation and demonstrating that 1,000 clonal T cells per million PBMCs (approximately 0.2% of all T cells) is roughly sufficient to repeatedly detect a clone in parallel emulsion reactions. This sensitivity is adequate for applications where one is looking for highly represented TCRs, especially in sorted T-cell subpopulations of interest.

Additionally, one of the clones captured in a single emulsion reaction consisted of the alpha and beta variants CILDNNNDMRF/CASSLAPGATNEKLFF-2, a combination that was previously identified among sorted NLV-specific cells, confirming the correctness of this pairing (P30, **Supporting Information Tables 2** and **3**). Strikingly, exactly the same amino-acid CDR3

sequences for the TCR alpha and beta chains were earlier reported for the NLV-specific T-cell clone for another patient [9].

Accuracy of pairing for the four known clones was generally high in all emulsion reactions, with the average ratio of native alpha-beta pairs to the most abundant mispairing of 334:1 (**Fig. 2** and **Supporting Information Table 4**).

Concluding remarks

In summary, the independent confirmation of four TCR alpha-beta chain pairs identified in a complex PBMC mix (**Table 1**) allows us to conclude that the described emulsion-based approach for identifying immune receptor gene pairs works reliably. Having demonstrated the soundness of this concept, we recognize that substantial technical optimization will be necessary in order to efficiently capture chain-pairs for rare clones and for the full diversity of TCR V gene segments. Relatively simple improvements like the use of advanced live cell emulsion generators, such as those manufactured by RainDance Technologies, will dramatically expand the power of this method compared to manually-generated emulsions, yielding increased efficiency of all reactions by consistently generating standardized droplets of sufficient volume. We believe that future iterations of our platform will greatly facilitate analysis of adaptive immune responses and the production of TCR- and antibody-based tools for biology and medicine.

Materials and methods

Emulsion RT-PCR. PBMCs were isolated from the peripheral blood sample of a 45-year-old man (informed consent obtained) by Ficoll-Paque (Paneco) density gradient centrifugation. For each of the eight emulsion and two control reactions, 1×10^6 PBMCs were pre-incubated overnight with 10 U/ml IL-2 (Roche), which increases TCR mRNA expression level (our observations, data not shown). Emulsions were generated at 25 °C as described previously [20] from 900 µl oil-surfactant (2% ABIL EM 90 and 0.05% Triton X-100 in mineral oil) and 100 µl aqueous phase. The aqueous phase included PBMCs in 10 µl 150 mM NaCl, PCR mix (7.5 U Encyclo polymerase (Evrogen), 5 µl 10x Encyclo buffer, 5 U MMLV reverse transcriptase (Evrogen), 3.5 mM $MgCl_2$, 1.4 mM DTT; In four emulsion reactions, it was substituted with single stage RT-PCR mix, Evrogen), 0.5 g/L BSA, 30 U RNasin (Promega), 2.4 mM dNTP mix, and oligonucleotide primers. Primer set included 0.2 µM each: BC_synth_rev, AC_synth_rev, BV7_for, and AV_for_uni, as well as a mix of the following 13 TRAV primers (0.02 µM each): AV14_for, AV26-1/4_for, AV26-2/4_for, AV39/24_for, AV38-2_for, AV13-2_for, AV25-2_for, AV41/22_for, AV8_for, AV29/23_for, AV12-3_for, AV27_for, and AV5_for. All oligonucleotides used in the study and their functions are listed in **Supporting Information Table 1**.

Preliminary experiments demonstrated that MMLV reverse transcriptase survives heating at 65°C for 2 min, while T cells are efficiently destroyed by such incubation (data not shown). Thus, we employed this step in our emulsion reactions as a temperature shock to burst living cells and release RNA into the droplet volume (**Fig. 1A**).

For the sequential reactions shown in **Fig. 1B**, we employed the following temperature regimen: 30 min incubation at 50°C for specific cDNA priming and synthesis, followed by 27 cycles of PCR amplification and overlap extension at 94°C (10 s), 53°C (10 s) and 72°C (20 s). The two reverse primers specific to the constant regions of the TCR alpha and beta chains (BC_synth_rev and AC_synth_rev) served a triple function: specific priming of cDNA synthesis, PCR amplification of variable fragments and PCR amplification of the overlapped-extended (fused) chains within emulsion. The direct primers (named "for" in the **Supporting Information Table 1**) were specific to the variable TCR alpha and beta gene segments and served a double function: PCR amplification of variable fragments and introduction of complementary sequences for the overlap-extension reaction. These primers were designed to anneal specifically at corresponding TCR V gene segments with a complementary region of sufficient length for the overlap-extension reaction.

1 mM EDTA was added immediately after the emulsion reaction to block polymerase activity. Emulsion product was extracted using diethyl ether/ethyl acetate/diethyl ether as described previously [20] and purified using a PCR cleanup kit (Qiagen).

Post-emulsion amplification with PCR suppression. Emulsion-derived DNA was further amplified in two sequential PCR amplification reactions (15 cycles/25-fold dilution/17-20 cycles) with nested primers specific to the constant regions of the TCR alpha and beta genes (**Supporting Information Fig. 1, Supporting Information Table 1**). Blocking oligonucleotides were added at concentration 3.2 µM to suppress the non-overlapped segments of TCR alpha and beta genes, as shown in **Fig. 1C**. To minimize potential side-effects in PCR, the 3'-ends of the blocking oligonucleotides were phosphate-protected. Taq polymerase (Evrogen) was used at these steps to exclude proofreading activity that could alter efficiency of PCR suppression. The temperature regimen for the 1st post-emulsion amplification was 94°C for 30 s followed by 15 cycles of 94°C (10 s) and 72°C (30 s); the regimen for the 2nd post-emulsion amplification was 94°C for 30 s followed by 17-20 cycles of 94°C (10 s), 62°C (10 s) and 72°C (20 s). We analyzed the resulting library of paired TCR alpha/beta fragments (average length ~400 bp) via paired-end 150 nt Illumina sequencing. Raw data will be deposited on NCBI Sequence Read Archive (SRA).

Emulsion controls

1. After 40-60 min of incubation in Encyclo PCR buffer (Evrogen), most PMBC cells were positively stained with 5% trypan blue solution (PanEco, Russia), indicating that they were either dead or at least perforated. However, the whole RT-PCR reaction worked well starting from both centrifuged cells and from the supernatant after such incubation, indicating that although the RNA leaks into solution, the RNasin keeps it intact. To avoid contamination of leaked TCR alpha- and beta-encoding RNA between droplets, we therefore performed manual emulsification immediately after placing the cells into pre-mixed PCR buffer. Manual emulsification itself takes only 2–4 minutes [20], while only a minor percentage of PBMCs were positively stained with trypan blue after being incubated in PCR buffer for 7 minutes.

2. To achieve both separation of T cells and completion of further RT-PCR reactions in individual volumes, appropriate droplet size and emulsion stability had to be maintained. To analyze these parameters, we generated two emulsions with the water phase stained by either green (FITC) or red (TAMRA) fluorescent oligonucleotides, mixed them and subjected them to the RT-PCR temperature regimen. We analyzed these emulsion aliquots using fluorescent microscopy to

calculate average droplet size, percentage of droplets of a volume sufficient to carry a T cell, and percentage of droplets that fused during thermocycling and thus fluoresced in both channels. Basing on analysis of >4,000 individual droplets, we estimated that a 50 µl reaction volume consists of approximately 2.5×10^8 droplets with an average volume of 210 fl, of which approximately 9×10^6 droplets have volume >500 fl—roughly sufficient to include a T cell and a required volume of reaction mix. These experiments also demonstrated minimal (< 0.1%) fusion of droplets during thermocycling.

3. In another control experiment, we examined 5×10^5 DAPI-stained PMBCs put in emulsion within 50 µl reaction buffer (colored by carboxyfluorescein), and observed cell-containing droplets but failed to identify droplets containing multiple cells among the > 50,000 droplets we manually screened.

4. Finally, we performed control experiments to test emulsion stability, using a mixture of two emulsions each containing oligonucleotide mix for amplification of either TCR alpha or beta chains. We performed cell disruption, reverse transcription and inactivation of reverse transcriptase for the two emulsions separately, and then combined them. Thus, amplification and overlap-extension reactions were performed in the same tube; if fusion of cell-containing droplets was occurring, this should have led to generation of full, fused TCR alpha-beta product. As shown in **Figure 1D lane 6**, this reaction performed poorly when the TCR alpha and beta RT reactions were performed separately, indicating negligible fusion between cell-containing droplets during amplification. In this control experiment, we intentionally joined emulsions only after reverse transcription; therefore, fusion of a small primer-containing droplet to a T cell-containing droplet could not provide primers for cDNA synthesis of a second TCR chain at the reverse transcription stage. After heat-inactivation of reverse transcriptase, fusion of small droplets with primers therefore did not lead to generation of a paired TCR alpha/beta product. Thus this control was specifically for the background noise produced by the fusion of T cell-containing droplets.

Basing on the above analyses, we postulated that the ABIL EM90-based emulsion provides sufficient stability under the temperature regimen used, with a sufficient number of >500-fl droplets potentially suitable for the living cell RT-PCR and low toxicity for cells during the emulsification timetable we employed.

NGS data analysis

To extract CDR3 alpha and beta regions and correct PCR errors that introduce artificial sequence diversity and lead to biased read counts, raw Illumina data were analyzed as previously described [19]. For each emulsion reaction, we obtained between 300,000 and 1.2 million high quality, paired-end reads containing both TCR alpha and beta CDR3 sequence. To minimize noise from random events such as leakage through PCR suppression or overlap-extension of incomplete PCR products during library amplification, we discarded low-count pairs ($\leq$0.01% of total reads in each sample). We identified alpha-beta TCR CDR3 pairs that were significantly and selectively enriched within each emulsion reaction based on their read counts as follows.

First, we computed an expected value for the random pairing of each possible pair as:

#alpha_reads * #beta_reads / #total_reads

where #alpha_reads and #beta_reads are total counts of reads with a given CDR3 alpha and CDR3 beta in a sample, respectively, and #total_reads is the total number of reads in a sample. To evaluate the strength of pairing (*i.e.*, to check if the difference between observed and expected read count is significant) we computed *p*-values based on hypergeometric distribution, approximated by the worst of the two *p*-values obtained from the following binomial distributions:

P = P$_{hypergeom}$(#pair_reads/f|#alpha_reads/f,#beta_reads/f,#total_reads/f)

where #pair_reads is a number of reads for each pair and f is a normalization factor—a reads-per-droplet parameter empirically chosen as f=150, at which no significant events were observed in control reactions. Most of the significantly enriched events in a given emulsion had a very low *p*-value, and the final list of pairs all had $p < 10^{-10}$ in corresponding emulsions.

Next we verified the specificity of pairing based on a pairing factor (PF), described as:

PF = #pair_reads/(#alpha_reads + #beta_reads - #pair_reads)

For the identification of highly expanded T cell clones, those pairs that failed to fulfill the criteria of PF > 40% in at least two emulsion reactions were filtered out. In those cases in which several T-cell clones carry the same TCR beta but different alpha chains [22], we iteratively computed this score starting from the most abundant pair, then removing it with all its reads, continuing with the next most abundant pair and so on.

To extract reliable information concerning TCR chain pairing for rare clones, we selected pairs that were unambiguously identified in only one emulsion reaction, and for which alpha and beta partners were also exclusively observed in this particular emulsion reaction (no cross-pairing with other partners in any emulsion). We filtered out those pairs that were observed in more than one emulsion preparation or failed to fulfill the criteria of PF > 80% for the joined data of all emulsion reactions. This eliminated noise that could originate from cross-contamination of minor clones by

more abundant clones at any stage of paired library generation, and yielded 732 identified TCR chain pairs, from 4 to 402 pairs per emulsion, 92 pairs per emulsion at average (**Supporting Information Table 2**).

Independent identification of the native TCR alpha/beta pairs

To identify native pairs of TCR alpha and beta chains within a human PBMC sample by independent method, we performed the following set of experiments.

PBMCs were isolated from the peripheral blood samples by Ficoll-Paque (Paneco) density gradient centrifugation, incubated with CD8-specific antibody (clone YTC 182.20, Abcam), CD4-specific antibody (clone S3.5, Invitrogen), and HLA-A*02-CMV (NLVPMVATV) pelimer (Sanquin) for 20 min at room temperature, then washed twice with PBS and sorted on a MoFlo cell sorter (DakoCytomation). RNA was extracted using TRIZOL reagent (Life Technologies), and TCR alpha and beta chain libraries were generated using template-switch technology as previously described [18].

In the first sorting experiment we cloned and sequenced 50 TCR alpha and 50 TCR beta chain genes. Three variants of TCR alpha and two variants of TCR beta complementarity-determining regions were identified (CDR3, **Supporting Information Table 3**). Among these variants, the predominant CDR3 alpha CIRDNNNDMRF and CDR3 beta CASSLAPGATNEKLFF-1 were suggested as a probable functional pair, a notion confirmed by the identification of a nearly identical NLV-specific T cell clone in an independent study [9]. Note that these data were obtained before the design of emulsion reactions. The TRBV-7 family of the CASSLAPGATNEKLFF-1 clone was thus intentionally included so that we initially had an intrinsic control.

To confirm the correctness of these TCR chain pairings, we repeatedly sorted NLV-specific CD8+ T cells from the same patient. We sorted more than 100,000 cells with a 97% purity (**Supporting Information Fig. 4**), and then performed deep TCR alpha and beta profiling on an Illumina MiSeq (1/10 of a run) using a quantitative RNA-based technique for the generation of CDR3 libraries [18]. More than 20,000 CDR3-containing sequences were obtained and analyzed for TCR alpha and beta libraries. This deep profiling revealed the same major TCR alpha and beta CDR3 variants as previous cloning and sequencing analysis, but also provided sufficient statistics for rough frequency-based pairing [8]. CDR3 variant CASSLAPGATNEKLFF-1 constituted approximately 90% of all NLV-specific TCR beta sequences, nearly equal to the cumulative input of TCR alpha variants CIRDNNNDMRF and CAGYSGGGADGLTF. Thus these data strongly suggested P1 and P4 TCR alpha-beta pairs (**Supporting Information Table 3**).

Acknowledgments

Competing financial interests

Pairing technology is a property of Evrogen JSC, Moscow, Russia. SD is employed by Evrogen JSC. CDM has a share in Evrogen JSC.

Figure 1. Cell-based emulsion RT-PCR technique for identifying TCR alpha-beta chain pairing. (**A**) Flow chart illustrating the process. (**B**) Released TCR alpha and beta mRNAs are reverse-transcribed at 50°C with specific primers, amplified and overlap-extended within each droplet. (**C**) Reaction products are extracted from the emulsion and fused molecules of interest are selectively amplified, while non-fused molecules are suppressed by blockers. (**D**) Post-emulsion amplification of fused TCR molecules using: (**lanes 1, 4**) product of emulsion reaction that initially contained all primers from the start (αβ), (**lanes 2, 5**) combined products of two separate emulsions, each initially containing only primers for either TCR alpha or beta amplification (as a control for random overlap extension after emulsion), and (**lanes 3, 6**) combined product of two separate emulsions that initially contained primers for either alpha or beta TCR amplification but were joined after reverse transcription step (as a control for random overlap-extension due to fusion of cell-containing emulsion droplets during PCR). PCR suppression (lanes 4-6) eliminated random overlap extension efficiently compared to reactions performed without such suppression (lanes 1-3). (D) Data shown are representative of three experiments performed.

Figure 2. Observed frequencies for the 400 theoretically possible pairings of the 20 most abundant CDR3 alpha and beta chains, Data are depicted as (A, B) 3D graphs or (C, D) circular plots, generated using Circos [23]. Frequencies of each observed pair are shown for both (A, C) an emulsion and (B, D) a control reaction, with frequency-expected random pairing subtracted. (C) Numbers (P1-P6) indicate experimentally identified pairs (**Table 1**). (C, D) For the circular plots, segment sizes indicate rank of chain abundance in control, segment color shows CDR3 alpha (orange) and beta (blue) sequences, ribbon widths are scaled to the number of reads for a pair in a specific emulsion, and ribbon colors match those of the P1-P6 pairs shown in the corresponding 3D graph.

Table 1. Cross-confirmed pairs of TCR alpha and beta chains.

#	TRAV	CDR3 alpha	TRAJ	TRBV	CDR3 beta*	TRBJ	#em[a]
P1[b]	**26-2**	**CIRDNNNDMRF**	**43**	**7-6**	**CASSLAPGATNEKLFF-1**	**1-4**	**8**
P2	4	CLVNTGGGNKLTF	10	7-9	CASSLVEAGDPYEQYF	2-7	6
P3	38-1	CAFMYPQGGSEKLVF	57	7-9	CASGSGLAGGELFF	2-2	6
P4	**27**	**CAGYSGGGADGLTF**	**45**	**7-6**	**CASSLAPGATNEKLFF-1**	**1-4**	**4**
P5	13-1	CAAIGSEKLVF	57	7-3	CASSFGNRENIQYF	2-4	3
P6	38-2	CAYIPMDSNYQLIW	33	7-8	CASSTIGGPVNTGELFF	2-2	4
P7	38-2	CAYRLRGRGSNYKLTF	53	7-3	CASSPTRNTEAFF	1-1	2
P8	8-2/8-4	CAAGFQKLVF	8	7-9	CASSLVTSGNSYEQYF	2-7	3
P9	4	CLVATGAGSYQLTF	28	7-2	CASSLSAYEQYF	2-7	2
P10	26-1	CIVRVAFSGGYNKLIF	4	7-8	CASSSPIFLPGNEQFF	2-1	2
P11	38-1	CAFTNAGGTSYGKLTF	52	7-9	CASSLAGLNYEQYF	2-7	2
P12	13-1	CAASITGGGNKLTF	10	7-2	CASSLATNEKLFF	1-4	2
P13	22	CAVAVNFGNEKLTF	48	7-2	CASSFIVYEQYF	2-7	2
P14	41	CAVVKGAQKLVF	54	7-2	CASSFTTNTGELFF	2-2	2
P15	8-4	CAVTKNRDDKIIF	30	7-9	CASSLDRQGNREQYF	2-7	2
P16	26-1	CIGRFSDGQKLLF	16	7-2	CASSRRLPGETQYF	2-5	2
P17	25	CAGPRTGNQFYF	49	7-3	CASSSGQEVGELFF	2-2	2
P18	8-4	CAVSDRGDDKIIF	30	7-9	CASSLASGLAEEQFF	2-1	2
P19	8-4	CAVSDRGSARQLTF	22	7-7	CASSLSKGAYEQYF	2-7	2
P20	38-1	CAIRKGNNNDMRF	43	7-3	CASSLNIA~EFETQYF	2-5	2
P21	38-2	CAYRSVLRDDKIIF	30	7-2	CASSLPTGNNSPLHF	1-6	2
P22	26-1	CIVSSYGQNFVF	26	7-9	CASSLARDQETQYF	2-5	2
P23	38-2	CALIGNQFYF	49	7-8	CASSLRAGSEDTQYF	2-3	3
P24	41	CAVNLYNNNDMRF	43	9	CASSEWTGAETQYF	2-5	3
P25	4	CLSQWF~ARQLTF	22	7-2	CASSLADSPLHF	1-6	3
P26	8-2	CVVSGPKGSNTGKLIF	37	12-4	CASSVANWADTQYF	2-3	2
P27[c]	**5**	**CAEMGGSNYKLTF**	**53**	**7-9**	**CASSGTYEQYF**	**2-7**	**2**
P28	4	CLVGLGGQAGTALIF	15	7-6	CASSLQQGSYEQYF	2-7	2
P29	26-1	CIVRANDYKLSF	20	9	CASSVNLGTPYEQYF	2-7	2
P30[d]	**26-2**	**CILDNNNDMRF**	**43**	**7-6**	**CASSLAPGATNEKLFF-2**	**1-4**	**1**

[a] #em indicates the number of emulsions in which a pair was identified.
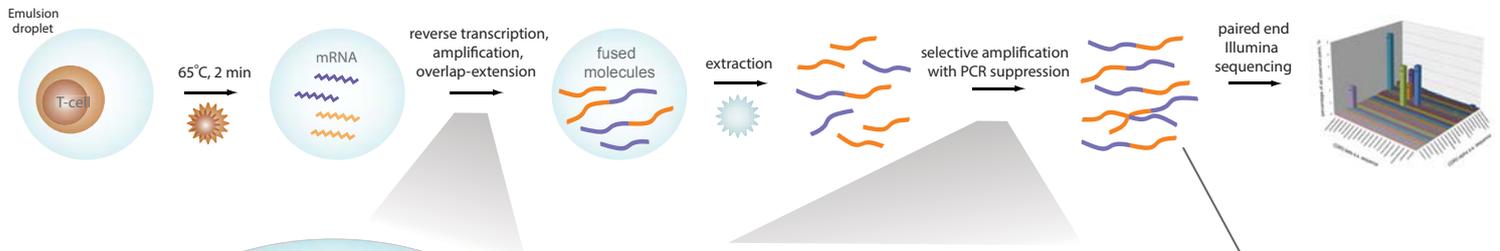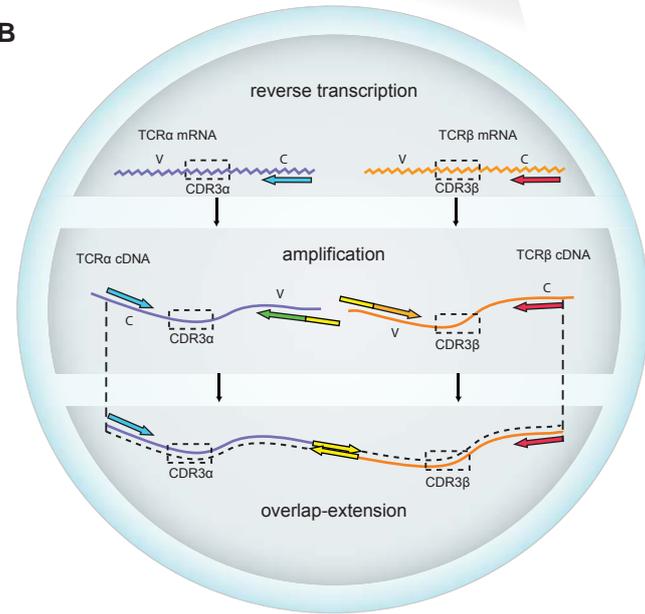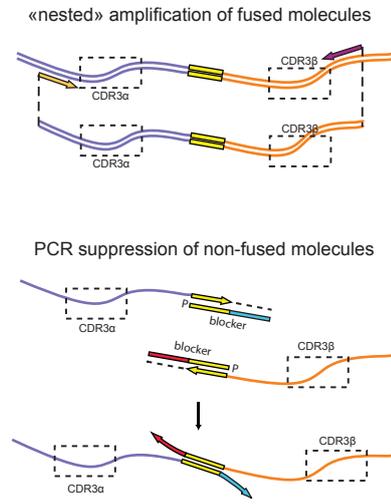[b] Pairs additionally confirmed by independent methods are shown in bold (P1, P4, P27, P30).
[c] Pair 27 is a control clone (C1A) spiked into emulsion reactions 5, 6, and 8.
[d] Pair 30 was identified in only one emulsion, but was additionally confirmed by cloning and analysis of sorted NLV-specific T cells, and was earlier reported in another patient [9].

References:

1   **Leavy, O.,** Therapeutic antibodies: past, present and future. *Nat Rev Immunol.* 2010. **10**: 297.

2   **Schumacher, T. N.,** T-cell-receptor gene therapy. *Nat Rev Immunol.* 2002. **2**: 512-519.

3   **Miles, J. J., Douek, D. C. and Price, D. A.,** Bias in the alphabeta T-cell repertoire: implications for disease pathogenesis and vaccination. *Immunol Cell Biol.* 2011. **89**: 375-387.

4   **Schwaber, J. and Cohen, E. P.,** Human x mouse somatic cell hybrid clone secreting immunoglobulins of both parental types. *Nature.* 1973. **244**: 444-447.

5   **Lagerkvist, A. C., Furebring, C. and Borrebaeck, C. A.,** Single, antigen-specific B cells used to generate Fab fragments using CD40-mediated amplification or direct PCR cloning. *Biotechniques.* 1995. **18**: 862-869.

6   **Babcook, J. S., Leslie, K. B., Olsen, O. A., Salmon, R. A. and Schrader, J. W.,** A novel strategy for generating monoclonal antibodies from single, isolated lymphocytes producing antibodies of defined specificities. *Proc Natl Acad Sci U S A.* 1996. **93**: 7843-7848.

7   **Meijer, P. J., Andersen, P. S., Haahr Hansen, M., Steinaa, L., Jensen, A., Lantto, J., Oleksiewicz, M. B.** *et al.***,** Isolation of human antibody repertoires with preservation of the natural heavy and light chain pairing. *J Mol Biol.* 2006. **358**: 764-772.

8   **Reddy, S. T., Ge, X., Miklos, A. E., Hughes, R. A., Kang, S. H., Hoi, K. H., Chrysostomou, C.** *et al.***,** Monoclonal antibodies isolated without screening by analyzing the variable-gene repertoire of plasma cells. *Nat Biotechnol.* 2010. **28**: 965-969.

9   **Trautmann, L., Rimbert, M., Echasserieau, K., Saulquin, X., Neveu, B., Dechanet, J., Cerundolo, V.** *et al.***,** Selection of T cell clones expressing high-affinity public TCRs within Human cytomegalovirus-specific CD8 T cell responses. *J Immunol.* 2005. **175**: 6123-6132.

10   **Franz, B., May, K. F., Jr., Dranoff, G. and Wucherpfennig, K.,** Ex vivo characterization and isolation of rare memory B cells with antigen tetramers. *Blood.* 2011. **118**: 348-357.

11   **Hoogenboom, H. R., Griffiths, A. D., Johnson, K. S., Chiswell, D. J., Hudson, P. and Winter, G.,** Multi-subunit proteins on the surface of filamentous phage: methodologies for displaying antibody (Fab) heavy and light chains. *Nucleic Acids Res.* 1991. **19**: 4133-4137.

12   **Marks, J. D., Hoogenboom, H. R., Bonnert, T. P., McCafferty, J., Griffiths, A. D. and Winter, G.,** By-passing immunization. Human antibodies from V-gene libraries displayed on phage. *J Mol Biol.* 1991. **222**: 581-597.

13   **Bowley, D. R., Jones, T. M., Burton, D. R. and Lerner, R. A.,** Libraries against libraries for combinatorial selection of replicating antigen-antibody pairs. *Proc Natl Acad Sci U S A.* 2009. **106**: 1380-1385.

14   **Embleton, M. J., Gorochov, G., Jones, P. T. and Winter, G.,** In-cell PCR from mRNA: amplifying and linking the rearranged immunoglobulin heavy and light chain V-genes within single cells. *Nucleic Acids Res.* 1992. **20**: 3831-3837.

15   **Chapal, N., Bouanani, M., Embleton, M. J., Navarro-Teulon, I., Biard-Piechaczyk, M., Pau, B. and Peraldi-Roux, S.,** In-cell assembly of scFv from human thyroid-infiltrating B cells. *Biotechniques.* 1997. **23**: 518-524.

16   **DeKosky, B. J., Ippolito, G. C., Deschner, R. P., Lavinder, J. J., Wine, Y., Rawlings, B. M., Varadarajan, N.** *et al.***,** High-throughput sequencing of the paired human immunoglobulin heavy and light chain repertoire. *Nat Biotechnol.* 2013. **31**: 166-169.

17   **Ho, S. N., Hunt, H. D., Horton, R. M., Pullen, J. K. and Pease, L. R.,** Site-directed mutagenesis by overlap extension using the polymerase chain reaction. *Gene.* 1989. **77**: 51-59.

18   **Mamedov, I. Z., Britanova, O. V., Bolotin, D. A., Chkalina, A. V., Staroverov, D. B., Zvyagin, I. V., Kotlobay, A. A.** *et al.***,** Quantitative tracking of T cell clones after haematopoietic stem cell transplantation. *EMBO Mol Med.* 2011. **3**: 201-207.

19   **Bolotin, D. A., Mamedov, I. Z., Britanova, O. V., Zvyagin, I. V., Shagin, D., Ustyugova, S. V., Turchaninova, M. A.** *et al.***,** Next generation sequencing for TCR repertoire profiling: platform-specific features and correction algorithms. *Eur J Immunol.* 2012. **42**: 3073-3083.

20   **Williams, R., Peisajovich, S. G., Miller, O. J., Magdassi, S., Tawfik, D. S. and Griffiths, A. D.,** Amplification of complex gene libraries by emulsion PCR. *Nat Methods.* 2006. **3**: 545-550.

21   **Porter-Jordan, K., Rosenberg, E. I., Keiser, J. F., Gross, J. D., Ross, A. M., Nasim, S. and Garrett, C. T.,** Nested polymerase chain reaction assay for the detection of cytomegalovirus overcomes false positives caused by contamination with fragmented DNA. *J Med Virol.* 1990. **30**: 85-91.

22   **Arstila, T. P., Casrouge, A., Baron, V., Even, J., Kanellopoulos, J. and Kourilsky, P.,** Diversity of human alpha beta T cell receptors. *Science.* 2000. **288**: 1135.

23   **Krzywinski, M., Schein, J., Birol, I., Connors, J., Gascoyne, R., Horsman, D., Jones, S. J.** *et al.***,** Circos: an information aesthetic for comparative genomics. *Genome Res.* 2009. **19**: 1639-1645.

A

Emulsion droplet — T-cell — 65°C, 2 min — mRNA — reverse transcription, amplification, overlap-extension — fused molecules — extraction — selective amplification with PCR suppression — paired end Illumina sequencing

B

reverse transcription

TCRα mRNA
V
CDR3α
C

TCRβ mRNA
V
CDR3β
C

amplification

TCRα cDNA
C
CDR3α
V

TCRβ cDNA
V
CDR3β
C

overlap-extension

CDR3α
CDR3β

C

«nested» amplification of fused molecules

CDR3α
CDR3β
CDR3α
CDR3β

PCR suppression of non-fused molecules

CDR3α
blocker
P
blocker
P
CDR3β

CDR3α
CDR3β

D

PCR suppression

1 2 3 4 5 6

fused product
blockers

αβ
α + β after RT-PCR
α + β after RT
αβ
α + β after RT-PCR
α + β after RT