

The intra-tumoural T cell receptor repertoire – steps towards a useful clinical biomarker

Gayathri Nageswaran¹⁺, Suzanne Byrne¹⁺, Selvaraju Veeriah³, Benny Chain^{*1,2}

1. Division of Infection and Immunity, UCL, London
2. Department of Computer Science, UCL, London
3. The Cancer Institute, UCL, London

* Corresponding author. Address Division of Infection and Immunity, Cruciform Building, UCL, Gower St. WC1E 6BT email : b.chain@ucl.ac.uk

+ these two authors contributed equally to the work

Running head : tumour T cell receptor repertoires

Abstract

Adaptive immunity recognises and responds to tumours, although they are part of the immunological 'self'. T cells, both CD4+ and CD8+ play a key role in the process, and the specific set of receptors which recognise tumour antigens therefore has the potential to provide prognostic biomarkers for tracking tumour growth after cancer therapy, including immunotherapy. Most published data on the T cell repertoire continues to rely on commercial proprietary methods, which often do not allow access to the raw data, and are difficult to validate. We describe an open-source protocol for amplifying, sequencing and analysing T cell receptors which is economical, robust, sensitive and versatile. The key experimental step is the ligation of a single stranded oligonucleotide to the 3' end of the T cell receptor cDNA, which allows easy amplification of all possible rearrangements using only a single set of primers per locus, while simultaneously introducing a unique molecular identifier to label each starting cDNA molecule. After sequencing, this molecular identifier can be used to correct both sequence errors and the effects of differential PCR amplification efficiency, thus producing a more accurate measure of the true T cell receptor frequency within the sample. Samples are then tagged with unique pairs of indices, facilitating robotic scale-up and significantly reducing cross-sample contamination from index hopping. This method has been applied to the analysis of tumour infiltrating lymphocytes and matched peripheral blood samples from patients with a variety of solid tumours.

Key words : T cell receptors; tumour infiltrating T cells; unique molecular identifier;

1. Introduction

The successes of checkpoint immunotherapy for several cancers has prompted a revival of interest in the T cell immune response to tumour antigens, and specifically in using the T cell receptor (TCR) repertoire as a prognostic or diagnostic marker of immune-therapy success and failure [1]. The rapid advances in high throughput DNA sequencing (HTS) over the past decades, and specifically the introduction of reliable massively parallel technologies (reviewed in [2]) have opened the way for increasingly robust and extensive BCR and TCR repertoire studies. However, the enormous diversity of the repertoire arising from a complex imprecise somatic recombination mechanism provides major challenges to reliable, and especially quantitative sequencing of the TCR repertoire. In particular, the combination of multiple rounds of PCR, and the relatively high sequencing error of current NextGen sequencing introduce significant amplification bias and inflate the observed diversity. Several approaches to tackling these challenges have been published and a careful comparison of the advantages and disadvantages of available techniques has recently been published [3]. Logistic considerations such as cost, ease of use, robustness and versatility, as well as more scientific issues such as accuracy and coverage will determine which solution is optimal for different laboratories.

Here we describe a modified and improved version of an experimental pipeline previously developed and published by our laboratory [4, 5]. A computational pipeline with which to analyse the data is presented elsewhere [5–7]. The special feature of the protocol is that it uses single stranded cDNA ligation mediated by T4 RNA ligase to incorporate unique molecular identifiers (UMIs)[8–10]. UMIs can be used both for sequence error correction and to mitigate for inherent PCR heterogeneity [7, 11]. This allows quantitative estimates of TCR gene abundance.

An important additional consideration, especially for clinical applications, is the ability to scale the pipeline to deal with large number of samples, so as to allow robust statistical analysis. In order to achieve this, we have modified the protocol to incorporate robotic steps, especially for the numerous bead purification steps. The sequencing platform also imposes cost and scale limitations on the protocol and the clinical projects. Over the past five years, we have successfully transitioned from small scale TCR repertoire analysis on an Illumina Miseq, to intermediate throughput on the Nextseq and most recently high throughput on the Novoseq. However, inter-sample contamination becomes an increasing problem as more and more samples are sequenced together. Index hopping is a well-known source of such contamination [12]. We have therefore modified the protocol to incorporate unique dual index multiplexing, and shown that this significantly improves inter-sample mixing. Overall, we have now successfully used this method to characterise thousands of TCR repertoires (comprising billions of sequences) from tumour infiltrating lymphocyte (TIL) suspensions and unfractionated frozen tumour tissue, as well as whole blood or fractionated Peripheral Blood Mononuclear Cells (PBMC) from a broad range of solid and haematological cancers.

2. Materials

Materials

Prepare any primers or reagents that require the addition of water using RNase free water. Store all reagents according to manufacturer's instructions.

1. The full list of oligonucleotides is shown in Table 1 and Table 2. We order all oligonucleotides from Merck. We use reverse phase cartridge purified grade unless otherwise indicated. All

oligonucleotides are stored at -20°C, and defrosted on wet ice before use. Avoid repeat freeze thawing by using small aliquots of all reagents.

2. Nuclease free water (Sigma, #W4502)
3. Wet Ice
4. 96-Well PCR Plate, Semi-Skirted with Straight Edges (StarLab, #I1402-9800)
5. RQ1 RNase free DNase kit including 10xbuffer, DNase and DNase stop buffer (Promega, #M6101)
6. dNTP Mix (10 mM each) (Thermo Fisher Scientific, #R0192)
7. SuperScript™ IV Reverse Transcriptase kit (Thermo Fisher Scientific, #18090200), includes SuperScript™ IV RT (40,000 units total at 200U/μL), 5X RT Buffer, 0.1M DTT.
8. RNasin Ribonuclease Inhibitor (Promega, #N2115). Alternative RNase inhibitors such as RNaseOUT (Thermo fisher Scientific, #10777019) can also be used.
9. Molecular grade Ethanol (Fisher Chemical, #E/0600DF/17)
10. T4 RNA Ligase 1 (ssRNA Ligase) kit including T4 RNA Ligase reaction buffer, Adenosine-5'-Triphosphate (ATP) and PEG 8000 50% (New England Biolabs, #M0204S)
11. 20mg/ml BSA (New England Biolabs, #B9000S)
12. Hexamine Cobalt (III) Chloride (HCC) stock solution (100 mM, can be stored at RT; Sigma-Aldrich, #481521)
13. Agencourt AMPure XP beads (Beckman Coulter, #A63881)
14. Phusion® High-Fidelity DNA Polymerase kit including HF Buffer (New England BioLabs, #M0530L)
15. MicroAmp™ Fast Optical 96-Well Reaction Plate with barcode (Thermo Fisher Scientific, #4346906)
16. SYBR™ Green I Nucleic Acid Gel Stain, 10,000X concentrate in DMSO (ThermoFisher #S7585).
Make a stock solution by diluting this 1 : 100 in DMSO and store in the dark at -20°C.
17. ROX Reference Dye (Thermo Fisher Scientific, #12223012)

18. 1.5% Agarose-dye-free- w/internal standards-Pippin prep- 250bp- 1.5kb -10/pk (Labtech, #CDF1510)
19. Biomek AP96 P20 Pre-Sterile Pipette tips (Beckman coulter, 717255)
20. Biomek AP96 P250 Sterile Tips with filter (Beckman coulter, 717253)
21. Micro-Pipettes (10µl, 20µl, 200µl and 1000µl)
22. 1.5ml DNase/RNase-free, low-retention microcentrifuge tubes (Thermo Scientific, #11569914)
23. PCR Plate seals (4titude, #4ti-0500); used for sealing the plates prior to loading onto the thermocycler (e.g. Ligation, PCR1 and PCR2).
24. Plate seals (StarLab, #E2796-0794); cheaper plate seals used to seal the plate for steps not involving thermocycler (e.g. purifications, or bead elutions).
25. Optical Plate seals (Applied Biosystem, #4311971); used for real-time PCR.
26. StarChill PCR Rack, Purple/Pink (StarLab, #E2396-2004)
27. Desktop Centrifuge/ Plate centrifuge
28. Electric Pipette
29. 96 well Thermocycler
30. 96well Magnetic Stand (e.g. Invitrogen, #12331D)
31. 96well qPCR Thermocycler
32. Qubit dsDNA HS Assay Kit (Thermo Fisher Scientific, #Q32854)
33. Qubit RNA BR Assay Kit (Thermo Fisher Scientific, #Q10211)
34. Qubit Assay Tubes (Thermo Fisher Scientific, #Q32856)
35. Qubit Fluorometer (ThermoFisher, using components 30-32). Alternative fluorometric DNA quantitation technologies could also be used, if they are of equivalent or greater sensitivity and accuracy.
36. High Sensitivity D1000 ScreenTape (Agilent Technologies, #5067-5584)
37. High Sensitivity D1000 Reagents (Agilent Technologies, #5067-5585)

38. RNA ScreenTape (Agilent Technologies, #: 5067-5576)
39. RNA ScreenTape Sample Buffer (Agilent Technologies, #5067-5577)
40. TapeStation Loading Tips (Agilent Technologies, # 5067-5599)
41. Optical strip tubes (Agilent Technologies, #401428)
42. Optical strip caps (Agilent Technologies, #401425)
43. Tape station 2200 or 4200 (Agilent Genomics, using reagents 34-37). Alternative nucleic acid quality control platforms could be used (e.g. Bioanalyzer, Agilent Technologies), providing they accurately determine the size (in base pairs) of input DNA libraries and/or RNA samples.
44. 1.5% Agarose- dye-free- w/ internal standards- Pippin Prep- 250 bp -1.5kb- 10/pk (Labtech, #CDF1510)

3. Methods

A diagrammatic overview of the pipeline is shown in Figure 1.

3.1 Isolation of TILs from human tumour samples.

These methods have been used for both tumor resection samples and core tissue biopsies. Fresh tissue samples are digested with collagenase (100ul/mL, Sigma) and DNase I (0.2 mg/mL, Roche) at 37°C for 30 minutes. Samples can be subsequently homogenised using a gentleMACS Dissociator (Miltenyi Biotech) and filtered through a 0.7µm cell mesh. Leukocytes are enriched by gradient centrifugation with Ficoll-paque PLUS (GE Healthcare) and re-suspended in 350 µl of RLT buffer (or 75 µl RLT buffer if $\leq 1 \times 10^5$ cells, Qiagen). Following resuspension, samples should be mixed by either vortexing or pipetting, and subsequently cryopreserved at -80°C. RNA can also be successfully extracted from cryopreserved TILs or PBMC following washing of cryopreserved cells post thawing and re-suspension in RLT buffer.

3.2 Extraction of RNA from snap frozen tumour tissue, frozen TILs or PBMC. Extraction of RNA from snap frozen tumour tissue involves firstly, the homogenization of tumour tissue using a hand held homogenizer (Qiagen) followed by RNA extraction using an AllPrep DNA/RNA Mini kit. Extracted RNA is eluted in 200 µl of water. RNA extraction from frozen TILs or PBMC is performed following density-gradient centrifugation using Ficoll-Paque PLUS (GE Healthcare). Following lysis in RLT buffer, RNA is extracted from TILs or PBMC using the Qiagen RNeasy kits. The RNA is eluted in 14 µl (Qiagen Micro kit used for fewer than 5×10^5 cells) and 30µl (Qiagen mini kit). The RNA is quantified fluorometrically (e.g. using the Qubit with RNA BR reagents, ThermoFisher Scientific) and integrity can be assessed using a Bioanalyzer or TapeStation (Agilent Genomics).

3.3 Residual DNA removal. Any residual genomic DNA is removed by DNase treatment. The RNA should be diluted as necessary in RNase free water to give a final volume of 8 μ l, aiming for a final total amount of 1 - 1000 ng. For total tumour tissue samples, 500-1000 ng RNA is normally used. We have found 1000 ng gives the largest number of TCRs, where available, but smaller amounts still give good number of sequences. If the RNA has been treated with DNase separately (for instance during RNA extraction), this step is omitted, and the protocol starts at step 3.4 with 11 μ l of total RNA. All RNA and reagents are kept on wet ice during the protocol.

1. Place a 96-well plate on a 96-well Chill rack.
2. Transfer the required amount of RNA (max 8 μ l), into a well of a 96-well plate. For concentrated samples, make up to 8 μ l with RNase free water. For tissue samples which have low RNA concentrations, up to 16 μ l of sample can be used. See Note e.
3. Add 1 μ l of RQ1 10x Buffer and 1 μ l of RQ1 DNase into each sample, mixing by gentle pipetting.
4. Apply PCR plate seal and centrifuge briefly to ensure all liquids are at the bottom of the well. Place the plate into a 96-well thermocycler at 37°C for 30 minutes.
5. Briefly centrifuge to collect condensation and add 1 μ l of RQ1 DNase Stop buffer to each sample, and reseal.
6. Briefly centrifuge the plate and then place the plate into the thermocycler at 65°C for 10 minutes to inactivate the DNase.

3.4 Reverse transcription.

1. Briefly centrifuge the plate and add 4 μ l of RNase free water, 1.5 μ l of oligonucleotides Alpha RC2 and Beta RC2 (Table 1), and 1.5 μ l of dNTPs per sample. For N samples (where N>1), a master mix should be prepared, containing 1.1N times the volumes above. The mix should be thoroughly mixed by pipetting or vortexing, and then 8.5 μ l of mix should be added to each sample. Seal.
2. Briefly centrifuge the plate and then place into the thermocycler at 65°C for 5 minutes.
3. Immediately transfer the plate to ice, and leave for at least 1 minute.
4. Briefly centrifuge the plate and add 6 μ l of 5X SSIV Buffer, 1.5 μ l of 0.1 M DTT, 1.5 μ l of RNasin and 1.5 μ l of SSIV RT to each sample. For N samples (where N>1), a master mix should be prepared, containing 1.1N times the volumes above. The mix should be thoroughly mixed by pipetting or vortexing, and then 10.5 μ l of mix should be added to each sample. The total sample volume at this stage should be 30 μ l.

5. Briefly centrifuge the plate and then place into the thermocycler at 55°C for 20 minutes followed by 80°C for 10 minutes.
6. Briefly centrifuge the plate. Add 30µl of AMPure beads to each well and mix by pipetting up and down at least 10 times (avoid creating bubbles) and leave at room temperature for 5 minutes. Note that the AMPure XP beads should be used at room temperature; remove from the fridge 30 minutes prior to use.
7. Briefly centrifuge the plate and then place onto a magnetic stand (see Materials) and leave for 2 minutes or until a pellet is formed.
8. Remove the clear liquid from the sample (ensuring the tips of the pipette does not touch the beads, and that no beads are aspirated).
9. Add 200 µl of freshly prepared 80% ethanol to each sample and leave for 30 seconds at RT.
10. Remove the ethanol, discard and add another 200 µl of 80% ethanol to each sample and leave for 30 seconds at RT.
11. Remove all ethanol from the sample, ensuring no liquid is left behind in the well.
12. Remove the plate from the magnet and dry at RT for 5 minutes. Samples can also be dried by placing plate on a heat block at 37°C; however, the samples need to be carefully monitored to prevent over drying which manifests by cracks appearing in the bead pellet.
13. Add 12µl of water into each sample and pipette mix at least 10 times. Leave at RT for 2 minutes.
14. Briefly centrifuge the plate and store on ice while preparing the ligation mix.

3.5 Ligation

1. Prepare a solution containing 1 mg/ml BSA and 10 mM HCC from stocks. To make 10 µl of solution, mix 8.5µl of RNase free water, 1µl of HCC stock (100 mM) and 0.5µl of BSA stock (20 mg/ml). 3 µl is required per sample.
2. Mix 3 µl of T4 RNA ligase buffer, 3µl of 1 mg/ml BSA/ 10 mM HCC solution, 1 µl of ATP, 1 µl SP2-I8 ligation oligonucleotide, 2 µl of T4 RNA Ligase and 10 µl of PEG 8000 50% per sample (in that order). For multiple samples make a master mix as above. The PEG solution is very viscous; allow it to first equilibrate to room temperature, and pipette very slowly with a wide-mouth tip. The mixture should be thoroughly mixed by pipetting to ensure uniformity. Remove samples from ice and add ligation mix.
3. Cover the plate with a plate seal, briefly centrifuge and then place into the thermocycler at 16°C for 16-23 hours (20 hours is preferred). Inactivate ligase by incubating at 65°C for 10 minutes and then store at 4-10°C until purification (for a maximum of 24 hrs). Do not freeze at this stage.

4. Briefly centrifuge the plate and add 70 μ l of water to each sample and mix. Seal and briefly centrifuge.
5. Add 50 μ l of AMPure XP beads to each sample, mix by pipetting up and down at least 10 times (avoid creating bubbles) and leave at room temperature for 5 minutes. Note that the AMPure XP beads should always be used at room temperature; remove from the fridge 30 minutes prior to use.
6. Briefly centrifuge the plate and then place onto a magnetic stand (see Materials) and leave for 2 minutes or until a pellet is formed.
7. Remove the clear liquid from the sample (ensuring the tips of the pipette does not touch the beads, and that no beads are aspirated).
8. Add 300 μ l of freshly prepared 80% ethanol to each sample and leave for 30 seconds at RT. A larger volume is added at this step in order to further dilute the PEG in the ligation.
9. Remove the ethanol, discard and add another 300 μ l of 80% ethanol to each sample and leave for 30 seconds at RT.
10. Remove all ethanol from the sample, ensuring no liquid is left behind in the well.
11. Remove the plate from the magnet and dry at RT for 5 minutes. Samples can also be dried by placing plate on a heat block at 37°C; however, the samples need to be carefully monitored to prevent over drying which manifests by cracks appearing in the bead pellet.
12. Add 33 μ l of water into each sample and pipette mix at least 10 times. Leave at RT for 2 minutes.
13. Briefly centrifuge the plate and place onto the plate magnet for 2 minutes.
14. Transfer 31 μ l of the sample into a clean 96-well plate.

3.6 First amplification - PCR 1. Initial amplification of ligated TCRs

1. Add 10 μ l of 5x HF Buffer, 2.5 μ l oligonucleotide Alpha RC1 (10 μ M, Table 1), 2.5 μ l Beta RC1 (10 μ M) (which is a premixed equimolar mixture of Beta RC1.1 and 1.2, Table 1), 2.5 μ l oligonucleotide SP2 (10 μ M), 1 μ l dNTPs (10 mM) and 0.5 μ l of Phusion Polymerase to each sample. For N samples (where N>1), a master mix should be prepared, containing 1.1N times the volumes above. The mix should be thoroughly mixed by pipetting or vortexing, and then 19 μ l of mix should be added to each sample. The total sample volume at this stage should be 50 μ l.
2. Seal and briefly centrifuge the plate and then place into the thermocycler, and amplify using the cycler program shown in Figure 2.
3. Briefly centrifuge the plate and add 40 μ l of prewarmed AMPure XP beads into each sample, mix by pipetting at least 10 times. Seal and leave at room temperature for 5 minutes.

4. Briefly centrifuge the plate and then place onto the Magnetic Stand and leave for 2 minutes or until a pellet is formed.
5. Purify following the procedure described in section 3.4, steps h-l. Use 200 μ l ethanol for each wash step.
6. Add 65 μ l of water into each sample and pipette mix at least 10 times. Seal and leave at room temperature for 2 minutes.
7. Briefly centrifuge the plate and place onto the plate magnet for 2 minutes.
8. Transfer 31 μ l of each sample into a fresh 96-well plate and label the plate "Alpha".
9. Transfer another 31 μ l of each sample into another fresh 96-well plate and label it "Beta".
The TCRalpha and TCRbeta amplifications can be performed on the same plate (e.g. Columns 1-6 contain Alpha samples and columns 7-12 contain Beta samples).

3.7 Second amplification - PCR 2. Addition of Illumina adapters and unique dual indices.

1. Prepare a master mix for PCR2. For N samples (where $N > 1$), a master mix should be prepared, containing 1.1N times the volumes above. The mix should be thoroughly mixed by pipetting or vortexing. Prepare different master mixes for TCRalpha and TCRbeta amplification, but the samples can be processed in parallel.
2. For each sample, the master mix contains 10 μ l of 5x HF Buffer, 1 μ l dNTPs and 0.5 μ l of Phusion polymerase, and either 2.5 μ l SP1-6HN-aRC1 (1 μ M, Table 1) for each alpha amplification, or 2.5 μ l SP1-6HN-bRC1.1/1.2 (1 μ M, Table 1) for each beta amplification. As shown in Table 1, SP1-6HN-bRC1.1/1.2 is a premixed equimolar mixture of SP1-6HN-bRC1.1 and SP1-6HN-bRC1.2. A total of 14 μ l master mix is added to each well.
3. Add 2.5 μ l of ONE SP1_I_P5 index primer (Table 1) and 2.5 μ l of the MATCHING SP2_I_P7 index primer (Table 1) to each sample. There are 96 unique combination of SP1 and SP2 (Table 2) available. Two biological samples to be sequenced together in the same run cannot have the same SP1 / SP2 index combination. However, alpha and beta genes from the same sample can be indexed with the same index pairs, and the sequences separated at analysis.
4. The total volume in each sample with the solution and indexes should now be 50 μ l. Centrifuge the plate briefly and place on the thermocycler. The thermocycler program is shown in Table 3.
5. Purify following the procedure described in section 3.4, steps h-l.
6. Add 40 μ l pre-warmed AMPure XP beads into each sample, mix by pipetting at least 10 times, seal and leave at room temperature for 5 minutes.

7. Briefly centrifuge and then place plate on the Magnetic Stand, and leave for 2 minutes or until a pellet is formed.
8. Purify following the procedure described in section 3.4, steps h-l.
9. Elute each sample in 30 μ l of water and pipette mix at least 10 times. Seal and leave at room temperature for 2 minutes.
10. Briefly centrifuge the plate and place onto the plate magnet for 2 minutes.

3.8 Third amplification - PCR 3 (real time PCR). Amplification of full length TCR transcript library.

1. Transfer 27.2 μ l of each sample into a MicroAmp™ Optical 96-Well Reaction Plate.
2. Make a 1 in 50 dilution of SYBR Green diluted stock (see Materials, 17) by mixing 1 μ l of stock SYBR green and 49 μ l of water. This should be made up fresh immediately before use, since SYBR green is poorly soluble in water.
3. Add 10 μ l of 5x HF buffer, 5 μ l of SYBR green dilution, 1.25 μ l of dNTP's, 1 μ l of ROX, 2.5 μ l of P5 primer, 2.5 μ l of P7 primer and 0.5 μ l of Phusion Polymerase to each sample and place the optical plate seal on the plate. For N samples (where N>1), a master mix should be prepared, containing 1.1N times the volumes above. The mix should be thoroughly mixed by pipetting or vortexing, and then 22.75 μ l of the mix should be added to each sample. The total sample volume at this stage should be 50 μ l.
4. Briefly centrifuge the plate and then place into a real time thermocycler, and amplify using the cycler program 3 shown in Table 3.. The number of cycles is dependent on how long the qPCR takes to reach the signal threshold. We routinely use a StepOne Thermocycler (ThermoFisher), but any machine can be used provided it is possible to follow the progress of the amplification in real time, and to stop the program when the signal reaches a predefined threshold. Some representative amplification plots are shown in Figure 2.
5. Once the samples have reached the predefined threshold value, transfer the plate to a standard thermocycler for a final 5 minute extension at 72° and then stop the thermocycler. Briefly centrifuge the plate.
6. Transfer each sample into a 96-Well PCR plate (the purification cannot be carried out in the qPCR plates as the ethanol wash volume exceeds the capacity of the well). Add 40 μ l of AMPure XP beads (at RT) to each sample, mix by pipetting up and down at least 10 times (avoid creating bubbles). Seal and leave at room temperature for 5 minutes.
7. Purify following the procedure described in section 3.4, steps h-l.
8. Elute each sample in 30 μ l of water. Seal and leave at room temperature for 2 minutes.
9. Briefly centrifuge the plate and place onto the plate magnet for 2 minutes.

10. Transfer 28 μ l of each purified library into a labelled 1.5ml polypropylene low binding tube and store at -20°C until required.

3.9 Library quality control and sequencing

1. The concentration of each DNA sample after amplification is quantified by spectroscopy (dsDNA high-sensitivity Qubit kit, ThermoFisher Scientific). Samples are also analysed by micro-electrophoresis using a high sensitivity D1000 TapeStation screen tape (Agilent). Typical profiles for libraries prepared from tumour and blood are shown in Figure 2.
2. The final libraries are prepared by mixing samples with different indices, so as to yield the same final concentration of each library in the pooled library. The actual concentration required may vary according to sequencing platform as detailed below.
3. After pooling, the final pooled library is further purified by electrophoresis, which we have found yields much better sequencing quality. We use a Pippin Prep electrophoresis machine (Sage science), collecting fragments of between 350 and 750 base pairs (fig 3).
4. For Illumina MiSeq sequencing, the final concentration of each sample in the pooled library is 4 nM. The final library is then diluted, denatured and sequenced as per manufacturer's instructions. Typically, 12 samples are run in parallel, yielding approximately 1-2 million reads per sample when using a v2 kit (2x251 bp paired end). However, greater depth can be achieved by running fewer samples per flow cell where needed. The final library concentration for loading is between 10-15pM. 5-7% PhiX DNA at 14-20pM is added to the library to increase diversity and improve cluster recognition.
5. For Illumina NextSeq sequencing, the final library is also prepared at 4nM. Typically at least 80 samples can be run on the flow cell, yielding 1-2 million reads per sample, when using the Mid-Output v2.5 300 cycle kit. Sample is loaded at 1.2 pM with 22% of PhiX DNA at 1.8 pM. Denaturation and sequencing follows standard Illumina protocols. For maximum analysis sensitivity, the standard 150 PE parameters are altered to give 201 bp for the forward read (R1) and 101 bp for the reverse read (R2).
6. For Illumina Novaseq sequencing, the final library concentration is 4 nM as above for NextSeq. If, after pooling and purification, the library concentration is below 4nM, the library can be reamplified for 4-6 cycles, using the same primers and cycler parameters as in PCR3 (see section 3.8) followed by a bead purification and quantification as above. Under these conditions we typically get 2-3 million reads per sample.

4. Notes

1. The protocol uses RNA rather than DNA as starting material, since this allows straightforward introduction of UMIs. The use of RNA also increases the likely coverage of the repertoire since each cell contains several molecules of TCR messenger RNA. The method is potentially sensitive to changes in RNA message associated with T cell activation and differentiation. However, several studies have demonstrated that such changes are small and transitory, and are therefore unlikely to have a major impact on the repertoire [13–17]. Quantitative PCR experiments from our laboratory (Oakes et al., 2017b) suggest that each TCR alpha or beta chain mRNA is present in 100-300 copies per cell.
2. The ligation of an oligonucleotide to the 3' end of the cDNA permits introduction of UMIs before PCR amplification, and also avoids the need for multiplex PCR and the biases this can generate. This greatly increases the flexibility of the protocol because T cells from different species or gamma/delta TCRs – or indeed any target transcript – can be easily processed simply by changing the constant region primers. Although T4 RNA ligase is often described as having low efficiency for single stranded DNA, we were unable to discover any publications with direct measurements of this efficiency. We therefore measured ligation efficiency using full-length cDNA extracted from Jurkat cells, and found it to be in the order of 10%. Thus at least one in ten TCR cDNA molecules should be ligated after overnight ligation using this method.
3. We carried out the amplification protocol in three steps. All PCR steps were carried out using Phusion high fidelity proof-reading polymerase. We observed that we could significantly increase efficiency by introducing a washing step after four cycles of PCR1. This was not due to inherent low efficiency of this PCR reaction, because amplification of an appropriate standard (a previously cloned TCR) showed PCR efficiencies >1.9. We speculate that residual oligonucleotide from the ligation step may remain after the first bead washing step and interfere with the amplification. Having this first round of amplification prior to separate alpha/beta processing also mitigates against losing lower frequency TCRs of one chain due to being sampled exclusively into the opposite chain's fraction. The third PCR was carried out using a qPCR SYBR Green protocol, so that amplification could be observed in real time. In this way, the amplification could be stopped when the amount of product reaches a pre-determined threshold, corresponding to a sufficient yield of DNA for sequencing. This avoids excessive amplification, which could bias against some rare TCR sequences in the starting pool, and minimises the burden of erroneous sequence production.
4. In some cases, the concentration of RNA may be too low to allow sufficient RNA (typically 500 – 1000 ng for non-immunological tissues) to be added in 8 µl (step 3.3b). We have

successfully double the reaction volume, starting with 16 μ l RNA, and then doubling all reagent volumes appropriately until and including step 3.4e (addition of Ampure Beads to purify ligation mixture). The subsequent wash steps, elution and all further steps then return to the standard protocol.

5. Lower molecular weight bands in tissue samples. As shown in fig 2, right panels, amplification of some samples, especially tissue samples where T cells are relatively scarce, often results in libraries with a number of lower molecular weight bands. Some of these bands may represent partial TCR mRNA amplification, while others may be non-TCR contamination. Our experience is that useful TCR sequence data is nevertheless obtained, provided at least some full sized TCR band (around 650 bp) can be seen in the library profile. Additional rounds of amplification using P5/P7 primers can sometimes be used to give higher yields of full length TCR.
6. PCR2 introduces several features required for subsequent sequencing on the Illumina platform, including the sequencing primers SP1 and SP2, two Illumina indices for multiplexing on the sequencer and the two Illumina anchor sequences P5 and P7. A random hexamer is introduced immediately downstream of SP1 to increase diversity at the beginning of the first sequence run, which greatly improves cluster identification.
7. Dual-indexing and cross-contamination. In the earlier version of this protocol, the libraries including two indices, but the index pairs were not unique (samples within one sequencing run could share a P5 or P7 sequence). However, we noted that this resulted in higher sharing between unrelated samples in the same sequencing run, compared to sharing between unrelated samples in different sequencing runs. "Index hopping" is a well-established cause of inter-sample overlap, especially in the higher capacity sequencing machines such as the Illumina Novogene. We therefore modified the extension PCR to introduce unique pairs of dual indices, as recommended by Illumina. We have observed much lower overlap between unrelated samples using this protocol modification. We also typically include negative samples (no RNA) in each library, and these samples record very few, or in some cases no reads, suggesting intra-well contamination is low or absent. We have so far used eight base pair dual indices using the Illumina index panel shown in Table 2, which allows up to 96 samples to be sequenced in parallel. Extension to 10 base pair indices should be straight forward if required. We routinely use the same index pair for alpha and beta samples, and distinguish the sequences during Decombinator analysis.
8. The cost of the protocol is an important consideration when large numbers of samples are to be processed. The current total cost of materials is in the order of £30-40 (approximately 35-

40Euros or 40-45 US dollars) per sample, excluding sequencing. The most expensive component is the reverse transcriptase SuperScript IV. We have not experimented with other reverse transcriptases but it is possible the total cost could be further reduced by using a different source of this enzyme.

- 9.** Robotics. By using an automated workstation which has interchangeable tools to perform different functions and complete multiple tasks, a larger number of samples can be prepared for sequencing with a much quicker turnaround. The use of automation, especially for the magnetic bead cleaning operations, also greatly reduces the amount of manual pipetting required, and we have observed a significantly lower risk of repetitive strain injury. The Biomek 4000 automated liquid handling has been validated for use the TCR pipeline described above. The method that is used on the machine is the same method that would be carried out manually however it is inputted into the Biomek software and carried out through a controller. The software is user friendly and is a linear step-based method building application. By using this software, minimal interface with the machine is needed during the run. Consumables for the Biomek 4000 are listed in the material section above. The components for the PCRs must be prepared by the handler for each master mix, however the purification after PCR and aliquoting the SP1 and SP2 indexes is all carried out automatically. The same 96-well plates can be used on the Biomek 4000 as for manual library preparation; however the pipette tips used are specific to the Biomek 4000. As the Biomek 4000 can perform the majority of the pipetting and handling tasks the only other equipment needed when using the Biomek 4000 is a plate centrifuge, a thermocycler for the PCR runs such as SimpliAmp and a qPCR machine for the final PCR.
- 10.** Our experience of using the Biomek4000 robotic platform has highlighted a number of important quality control issues. It is important to check visually that 80% ethanol is removed from samples after ethanol washes, taking extra care at wells around the outside of plate are clear of ethanol, and that the AMPure beads are fully mixed. The PEG8000 solution is viscous. It should be thoroughly mixed before dispensing, should be added last to the ligation mix, and the pipetting steps (both aspiration and dispensing) should be slower than for the wash steps. The half-skirted 96 well plate which is used for most steps of the pipeline is compatible with the Biomek4000. However for PCR3, a 96 well plate that is compatible with the qPCR machine must be used such as MicroAmp 96 well plate. An additional transfer into these plates is therefore required before the qPCR step. At the end of the final step, it is worthwhile introducing an additional visual check that all wells have an equal volume of solution.

Acknowledgements

This research was funded by grants from CancerResearchUK, the UK MRC, the Rosetrees Trust, and supported by the National Institute for Health Research UCL Hospitals Biomedical Research.

References

1. Joshi K, Milighetti M, Chain BM (2021) Application of T cell receptor (TCR) repertoire analysis for the advancement of cancer immunotherapy. *Current opinion in immunology* 74:1–8. <https://doi.org/10.1016/j.coi.2021.07.006>
2. Heather JM, Chain B (2016) The sequence of sequencers: The history of sequencing DNA. *Genomics* 107:1–8. <https://doi.org/10.1016/j.ygeno.2015.11.003>
3. Barennes P, Quiniou V, Shugay M, et al (2021) Benchmarking of T cell receptor repertoire profiling methods reveals large systematic biases. *Nature biotechnology* 39:236–245. <https://doi.org/10.1038/s41587-020-0656-3>
4. Uddin I, Woolston A, Peacock T, et al (2019) Quantitative analysis of the T cell receptor repertoire. *Methods in enzymology* 629:465–492. <https://doi.org/10.1016/bs.mie.2019.05.054>
5. Oakes T, Heather JM, Best K, et al (2017) Quantitative Characterization of the T Cell Receptor Repertoire of Naïve and Memory Subsets Using an Integrated Experimental and Computational Pipeline Which Is Robust, Economical, and Versatile. *Frontiers in Immunology* 8:1267. <https://doi.org/10.3389/fimmu.2017.01267>
6. Thomas N, Heather J, Ndifon W, et al (2013) Decombinator: a tool for fast, efficient gene assignment in T-cell receptor sequences using a finite state machine. *Bioinformatics* 29:542–550. <https://doi.org/10.1093/bioinformatics/btt004>
7. Peacock T, Heather JM, Ronel T, Chain B (2021) Decombinator V4: an improved AIRR compliant-software package for T-cell receptor sequence annotation? *Bioinformatics (Oxford, England)* 37:876–878. <https://doi.org/10.1093/bioinformatics/btaa758>
8. Weinstein JA, Jiang N, White RA, et al (2009) High-throughput sequencing of the zebrafish antibody repertoire. *Science (New York, NY)* 324:807–10. <https://doi.org/10.1126/science.1170020>
9. Shugay M, Britanova O V, Merzlyak EM, et al (2014) Towards error-free profiling of immune repertoires. *Nature methods* 11:653–5. <https://doi.org/10.1038/nmeth.2960>
10. Mamedov IZ, Britanova O V., Zvyagin I V., et al (2013) Preparing Unbiased T-Cell Receptor and Antibody cDNA Libraries for the Deep Next Generation Sequencing Profiling. *Frontiers in Immunology* 4:1–10. <https://doi.org/10.3389/fimmu.2013.00456>
11. Best K, Oakes T, Heather JM, et al (2015) Computational analysis of stochastic heterogeneity in PCR amplification efficiency revealed by single molecule barcoding. *Scientific reports* 5:14629. <https://doi.org/10.1038/srep14629>

12. Costello M, Fleharty M, Abreu J, et al (2018) Characterization and remediation of sample index swaps by non-redundant dual indexing on massively parallel sequencing platforms. BMC genomics 19:332. <https://doi.org/10.1186/s12864-018-4703-0>
13. Gallard A, Foucras G, Coureau C, Guéry J-C (2002) Tracking T cell clonotypes in complex T lymphocyte populations by real-time quantitative PCR using fluorogenic complementarity-determining region-3-specific probes. Journal of immunological methods 270:269–80
14. Lindsten T, June CH, Thompson CB (1988) Transcription of T cell antigen receptor genes is induced by protein kinase C activation. Journal of immunology (Baltimore, Md : 1950) 141:1769–74
15. Paillard F, Sterkers G, Vaquero C (1990) Transcriptional and post-transcriptional regulation of TcR, CD4 and CD8 gene expression during activation of normal human T lymphocytes. The EMBO journal 9:1867–72
16. Paillard F, Sterkers G, Vaquero C (1992) Correlation between up-regulation of lymphokine mRNA and down-regulation of TcR, CD4, CD8 and Ick mRNA as shown by the effect of CsA on activated T lymphocytes. Biochemical and biophysical research communications 186:603–11
17. Oakes T, Heather JM, Best K, et al (2017) Quantitative Characterization of the T Cell Receptor Repertoire of Naïve and Memory Subsets Using an Integrated Experimental and Computational Pipeline Which Is Robust, Economical, and Versatile. Frontiers in Immunology 8:. <https://doi.org/10.3389/FIMMU.2017.01267>

Table 1 Oligonucleotide sequences for TCR library preparation

Name	Sequence	Protocol step	Modification
	HUMAN		
Alpha_RC2	GAGTCTCTCAGCTGGTACACG	3.4	
Beta_RC2	ACACAGCGACCTCGGGTGGGAA	3.4	
6N_I8.1_6N_I8.1_SP2 (M13 ligation oligo)	³¹ P-NNN NNN ATCACGAC NNN NNN CCAGGGTTTTCCAGTCACGAC-C3	3.5	5' phosphate; 3' C3 spacer
Alpha RC1	ACGGCAGGGTCAGGGTTCTGGATAT	3.6	
Beta RC1.1 ²	GGTGGGAACACCTTGTTTCAGGTCCTC	3.6	
Beta RC1.2 ²	ATCTACACTGGTGGGAACACGTTTTTCAGGTCCTC	3.6	
SP1-6N-aRC1 ^{3,4}	ACACTCTTTCCCTACACGACGCTCTTCCGATCT HNHNNH ACGGCAGGGTCAGGGTTCTGGATAT	3.7	
SP1-6N-bRC1.1 ^{2,3,4}	ACACTCTTTCCCTACACGACGCTCTTCCGATCTHNHNNH GGTGGGAACACCTTGTTTCAGGTCCTC	3.7	
SP1-6N-bRC1.2 ^{2,3,4}	ACACTCTTTCCCTACACGACGCTCTTCCGATCTHNHNNH GGTGGGAACACGTTTTTCAGGTCCTC	3.7	
SP2_I_P7 ^{4,5}	CAAGCAGAAGACGGCATAACGAGAT [i7RC] GTGACTGGAGTTCAGACGTGTGCTCTTCCGATC Eg: CAAGCAGAAGACGGCATAACGAGAT <u>CCGCGGTT</u> GTGACTGGAGTTCAGACGTGTGCTCTTCCGATC	3.7	

SP1_I_P5 ⁴⁵	AATGATACGGCGACCACCGAGATCTACAC[<u>i5</u>]ACACTCTTCCCTACACGAC GCTC Eg: AATGATACGGCGACCACCGAGATCTACAC <u>AGCGCTAG</u> ACACTCTTCCCTACACGACGCTC	3.7	
P5 ⁵	AATGATACGGCGACCACCGAGATCTACACT	3.8	
P7 ⁵	CAAGCAGAAGACGGCATAACGAGAT	3.8	

1. The ligation oligonucleotide is phosphorylated at the 5' terminus, and blocked using a C3 spacer at the 3' terminus.
2. Two beta RC1 primers are required, as there are two isoforms of the beta constant region gene in the human genome. The two primers should be synthesised separately, and then made up as a stock solution containing 10 µM of each primer.
3. N represents a position at which any base should be inserted randomly. H represents a position at which A or C or T should be inserted.
4. The I refers to the index number. Underlined bases represent a representative index sequence, and can be substituted by any of the Illumina indices (Table 2). The indices allow multiple libraries to be combined for sequencing, and then deconvoluted for analysis.

The SP1 sequence is used by the Illumina sequencing protocol to initiate read 1. The SP2 sequence is used to initiate reads 2 (index only) and 3 (reverse). P5 and P7 sequences are adaptors necessary for binding of the DNA fragment to the flow cell.

Table 2 : Illumina Dual Indices

Index name	i7 Bases in Adapter	i7 Bases for Sample Sheet	i5 Bases for Sample Sheet NovaSeq, MiSeq, HiSeq 2000/2500	i5 Bases for Sample Sheet iSeq, MiniSeq, NextSeq, HiSeq 3000/4000
UDI0001	AACCGCGG	<u>CCGCGGTT</u>	AGCGCTAG	CTAGCGCT
UDI0002	GGTTATAA	<u>TTATAACC</u>	GATATCGA	TCGATATC
UDI0003	CCAAGTCC	<u>GGACTTGG</u>	CGCAGACG	CGTCTGCG
UDI0004	TTGGACTT	<u>AAGTCCAA</u>	TATGAGTA	TACTCATA
UDI0005	CAGTGGAT	<u>ATCCACTG</u>	AGGTGCGT	ACGCACCT
UDI0006	TGACAAGC	<u>GCTTGTC A</u>	GAACATAC	GTATGTTC
UDI0007	CTAGCTTG	<u>CAAGCTAG</u>	ACATAGCG	CGCTATGT
UDI0008	TCGATCCA	<u>TGGATCGA</u>	GTGCGATA	TATCGCAC
UDI0009	CCTGAACT	<u>AGTTCAGG</u>	CCAACAGA	TCTGTTGG
UDI0010	TTCAGGTC	<u>GACCTGAA</u>	TTGGTGAG	CTCACCAA
UDI0011	AGTAGAGA	<u>TCTCTACT</u>	CGCGGTTC	GAACCGCG
UDI0012	GACGAGAG	<u>CTCTCGTC</u>	TATAACCT	AGGTTATA
UDI0013	AGACTTGG	<u>CCAAGTCT</u>	AAGGATGA	TCATCCTT
UDI0014	GAGTCCAA	<u>TTGGACTC</u>	GGAAGCAG	CTGCTTCC
UDI0015	CTTAAGCC	<u>GGCTTAAG</u>	TCGTGACC	GGTCACGA
UDI0016	TCCGGATT	<u>AATCCGGA</u>	CTACAGTT	AACTGTAG
UDI0017	CTGTATTA	<u>TAATACAG</u>	ATATTCAC	GTGAATAT
UDI0018	TCACGCCG	<u>CGGCGTGA</u>	GCGCCTGT	ACAGGCCG
UDI0019	ACTTACAT	<u>ATGTAAGT</u>	ACTCTATG	CATAGAGT
UDI0020	GTCCGTGC	<u>GCACGGAC</u>	GTCTCGCA	TGCGAGAC
UDI0021	AAGGTACC	<u>GGTACCTT</u>	AAGACGTC	GACGTCTT
UDI0022	GGAACGTT	<u>AACGTTCC</u>	GGAGTACT	AGTACTCC
UDI0023	AATTCTGC	<u>GCAGAATT</u>	ACCGGCCA	TGGCCGGT
UDI0024	GGCCTCAT	<u>ATGAGGCC</u>	GTTAATTG	CAATTAAC
UDI0025	ATCTTAGT	<u>ACTAAGAT</u>	AACCGCGG	CCGCGGTT
UDI0026	GCTCCGAC	<u>GTCGGAGC</u>	GGTTATAA	TTATAACC
UDI0027	ATACCAAG	<u>CTTGGTAT</u>	CCAAGTCC	GGACTTGG
UDI0028	GCGTTGGA	<u>TCCAACGC</u>	TTGGACTT	AAGTCCAA
UDI0029	CTTCACGG	<u>CCGTGAAG</u>	CAGTGGAT	ATCCACTG
UDI0030	TCCTGTAA	<u>TTACAGGA</u>	TGACAAGC	GCTTGTC A
UDI0031	AGAATGCC	<u>GGCATTCT</u>	CTAGCTTG	CAAGCTAG
UDI0032	GAGGCATT	<u>AATGCCTC</u>	TCGATCCA	TGGATCGA
UDI0033	CCTCGGTA	<u>TACCGAGG</u>	CCTGAACT	AGTTCAGG
UDI0034	TTCTAACG	<u>CGTTAGAA</u>	TTCAGGTC	GACCTGAA
UDI0035	ATGAGGCT	<u>AGCCTCAT</u>	AGTAGAGA	TCTCTACT
UDI0036	GCAGAATC	<u>GATTCTGC</u>	GACGAGAG	CTCTCGTC
UDI0037	CACTACGA	<u>TCGTAGTG</u>	AGACTTGG	CCAAGTCT
UDI0038	TGTCGTAG	<u>CTACGACA</u>	GAGTCCAA	TTGGACTC
UDI0039	ACCACTTA	<u>TAAGTGGT</u>	CTTAAGCC	GGCTTAAG
UDI0040	GTTGTCCG	<u>CGGACAAC</u>	TCCGGATT	AATCCGGA
UDI0041	ATCCATAT	<u>ATATGGAT</u>	CTGTATTA	TAATACAG

UDI0042	GCTTGCGC	<u>GCGCAAGC</u>	TCACGCCG	CGGCGTGA
UDI0043	AGTATCTT	<u>AAGATACT</u>	ACTTACAT	ATGTAAGT
UDI0044	GACGCTCC	<u>GGAGCGTC</u>	GTCCGTGC	GCACGGAC
UDI0045	CATGCCAT	<u>ATGGCATG</u>	AAGGTACC	GGTACCTT
UDI0046	TGCATTGC	<u>GCAATGCA</u>	GGAACGTT	AACGTTCC
UDI0047	ATTGGAAC	<u>GTTCCAAT</u>	AATTCTGC	GCAGAATT
UDI0048	GCCAAGGT	<u>ACCTTGGC</u>	GGCCTCAT	ATGAGGCC
UDI0049	CGAGATAT	<u>ATATCTCG</u>	ATCTTAGT	ACTAAGAT
UDI0050	TAGAGCGC	<u>GCGCTCTA</u>	GCTCCGAC	GTCGGAGC
UDI0051	AACCTGTT	<u>AACAGGTT</u>	ATACCAAG	CTTGGTAT
UDI0052	GGTTCACC	<u>GGTGAACC</u>	GCGTTGGA	TCCAACGC
UDI0053	CATTGTTG	<u>CAACAATG</u>	CTTCACGG	CCGTGAAG
UDI0054	TGCCACCA	<u>TGGTGGCA</u>	TCCTGTAA	TTACAGGA
UDI0055	CTCTGCCT	<u>AGGCAGAG</u>	AGAATGCC	GGCATTCT
UDI0056	TCTCATTC	<u>GAATGAGA</u>	GAGGCATT	AATGCCTC
UDI0057	ACGCCGCA	<u>TGCGGCGT</u>	CCTCGGTA	TACCGAGG
UDI0058	GTATTATG	<u>CATAATAC</u>	TTCTAACG	CGTTAGAA
UDI0059	GATAGATC	<u>GATCTATC</u>	ATGAGGCT	AGCCTCAT
UDI0060	AGCGAGCT	<u>AGCTCGCT</u>	GCAGAATC	GATTCTGC
UDI0061	CAGTTCCG	<u>CGGAACTG</u>	CACTACGA	TCGTAGTG
UDI0062	TGACCTTA	<u>TAAGGTCA</u>	TGTCGTAG	CTACGACA
UDI0063	CTAGGCAA	<u>TTGCCTAG</u>	ACCACTTA	TAAGTGGT
UDI0064	TCGAATGG	<u>CCATTCGA</u>	GTTGTCCG	CGGACAAC
UDI0065	CTTAGTGT	<u>ACACTAAG</u>	ATCCATAT	ATATGGAT
UDI0066	TCCGACAC	<u>GTGTCGGA</u>	GCTTGCGC	GCGCAAGC
UDI0067	AACAGGAA	<u>TTCTGTIT</u>	AGTATCTT	AAGATACT
UDI0068	GGTGAAGG	<u>CCTTCACC</u>	GACGCTCC	GGAGCGTC
UDI0069	CCTGTGGC	<u>GCCACAGG</u>	CATGCCAT	ATGGCATG
UDI0070	TTCACAAT	<u>ATTGTGAA</u>	TGCATTGC	GCAATGCA
UDI0071	ACACGAGT	<u>ACTCGTGT</u>	ATTGGAAC	GTTCCAAT
UDI0072	GTGTAGAC	<u>GTCTACAC</u>	GCCAAGGT	ACCTTGGC
UDI0073	GTTAATTG	<u>CAATTAAC</u>	CGAGATAT	ATATCTCG
UDI0074	ACCGGCCA	<u>TGGCCGGT</u>	TAGAGCGC	GCGCTCTA
UDI0075	GGAGTACT	<u>AGTACTCC</u>	AACCTGTT	AACAGGTT
UDI0076	AAGACGTC	<u>GACGTCTT</u>	GGTTCACC	GGTGAACC
UDI0077	GTCTCGCA	<u>TGCGAGAC</u>	CATTGTTG	CAACAATG
UDI0078	ACTCTATG	<u>CATAGAGT</u>	TGCCACCA	TGGTGGCA
UDI0079	GCGCCTGT	<u>ACAGGCGC</u>	CTCTGCCT	AGGCAGAG
UDI0080	ATATTCAC	<u>GTGAATAT</u>	TCTCATTC	GAATGAGA
UDI0081	CTACAGTT	<u>AACTGTAG</u>	ACGCCGCA	TGCGGCGT
UDI0082	TCGTGACC	<u>GGTCACGA</u>	GTATTATG	CATAATAC
UDI0083	GGAAGCAG	<u>CTGCTTCC</u>	GATAGATC	GATCTATC
UDI0084	AAGGATGA	<u>TCATCCTT</u>	AGCGAGCT	AGCTCGCT
UDI0085	TATAACCT	<u>AGGTTATA</u>	CAGTTCCG	CGGAACTG
UDI0086	CGCGGTTT	<u>GAACCGCG</u>	TGACCTTA	TAAGGTCA
UDI0087	TTGGTGAG	<u>CTACCAA</u>	CTAGGCAA	TTGCCTAG
UDI0088	CCAACAGA	<u>TCTGTTGG</u>	TCGAATGG	CCATTCGA
UDI0089	GTGCGATA	<u>TATCGCAC</u>	CTTAGTGT	ACACTAAG
UDI0090	ACATAGCG	<u>CGCTATGT</u>	TCCGACAC	GTGTCGGA

UDI0091	GAACATAC	<u>GTATGTTC</u>	AACAGGAA	TTCCTGTT
UDI0092	AGGTGCGT	<u>ACGCACCT</u>	GGTGAAGG	CCTTCACC
UDI0093	TATGAGTA	<u>TACTCATA</u>	CCTGTGGC	GCCACAGG
UDI0094	CGCAGACG	<u>CGTCTGCG</u>	TTCACAAT	ATTGTGAA
UDI0095	GATATCGA	<u>TCGATATC</u>	ACACGAGT	ACTCGTGT
UDI0096	AGCGCTAG	<u>CTAGCGCT</u>	GTGTAGAC	GTCTACAC

Table 3 Cycler parameters for each PCR step

PCR	Step	Denaturation temp ¹	Denaturation time	Annealing temp ¹	Annealing time ¹	Extension temp ¹	Extension time	No of cycles
1	1	98	3 mins					1
	2	98	15 secs	69	30 secs	72	40 secs	4
	3					72	5 min	1
	4					10	Hold	
2	1	98	3 mins					1
	2	98	15 secs	69	30 secs	72	40 secs	6
	3					72	5 min	1
	4					10	Hold	
3	1	98	3 mins					1
	2	98	20 secs	69	30 secs	72	40 secs	Until hits threshold
	3					72	5 min	1
	4					10	Hold	

1. All temperatures in degrees centigrade.

2. Lid held at 112°C for all steps.

Legends

Figure 1. Diagrammatic overview of library preparation, labelled according to protocol steps.

Figure 2. PCR 3 (real time) amplification step (3.8). a) Thermocycler profiles. Yellow and green lines represent alpha and beta TCR genes amplified from 500 ng whole blood RNA (left panel) and 1000 ng of tissue RNA (right). b). Image of gel electrophoresis (Tapestation) analysis of amplified TCR libraries prepared from 1000 ng RNA extracted from lung tumour samples. c) Individual density trace for an individual lane of the gel in b.

Successful library preparation yields a major peak of approximately 650 bp (solid arrow) as predicted from the combined size of the recombined TCR variable gene, together with the short sequence of constant region and various adaptors/indices. The peak molecular size as determined on the Tape station is used to convert the concentration from ng/ μ l to nM. In some samples (especially when there are few T cells in the tissue) the library includes additional bands at lower molecular weights (at approximately 400bp, as demonstrated in right panels. If the concentration of the 650 bp band is insufficient for sequencing (see section 3.7), the library can be repurified and reamplified using the PCR3 protocol.

Figure 3. Workflow of the final purification on the PippinPrep (step 3.9c).