An economical, quantitative and robust protocol for high throughput T cell receptor sequencing from tumour or blood.

Imran Uddin[1+], Kroopa Joshi[2+], Theres Oakes[1] , James M. Heather [1], Charles Swanton[3,4] on behalf of the TRACERx consortium[x], Benny Chain [1*]

1. Division of Infection and Immunity, UCL, London

2. Immune Regulation and Tumour Immunotherapy Group, UCL Cancer Institute, UCL, London

3. Cancer Research UK Lung Cancer Centre of Excellence, UCL Cancer Institute, UCL, London

4. Translational Cancer Therapeutics Laboratory, The Francis Crick Institute, London, UK

* 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

x TRAcking Cancer Evolution through therapy (Rx). Full list of members in Appendix 1.

Running Head : Sequencing the T cell receptor repertoire

Key words : T cell receptor; high throughput sequencing; lung cancer; adaptive immune response; immunotherapy

Abstract

The T cell receptor repertoire provides a window to the cellular adaptive immune response within a tumour, and has the potential to identify specific and personalised biomarkers for tracking host responses during cancer therapy, including immunotherapy. We describe a 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. This method has been applied to the analysis of unfractionated human tumour lysates, subpopulations of tumour infiltrating lymphocytes and peripheral blood samples from patients with a variety of solid tumours.

1. Introduction

The recent dramatic successes of checkpoint immunotherapy for several cancers has prompted a re-examination of the T cell immune response to tumour antigens. One approach to evaluating and tracking this response is by studying the T cell receptor (TCR) repertoire. The TCR is generated during differentiation in the thymus by a process of somatic DNA recombination of the α and β (or in a minor population the γ and δ) TCR genes. The imprecise rearrangement of either V and J segments or V, D, J segments of alpha and beta chains respectively, gives rise to a huge number of different complementarity determining region 3 (CDR3) sequences, which determine the antigen specificity of individual T cell clones.

Although the mechanisms for the generation of diversity have been studied in great detail, the very diversity of the sequences coding for the receptors prevented a global analysis of the repertoire of B or T cell antigen receptors using conventional DNA sequencing techniques. The rapid advances in high throughput DNA sequencing (HTS) over the past decades, and specifically the introduction of reliable massively parallel technologies (reviewed in ***[1]***) have opened the way for increasingly robust and extensive BCR and TCR repertoire studies. Repertoire analysis provides a powerful tool for the study of both fundamental and translational immunology, and are increasingly being used to study the anti-tumoural immune response to solid cancers ***[2–7]***. Since each T cell and its descendants express a unique rearranged receptor, TCR sequencing provides an opportunity to

precisely 'track' T cell clones through a period of time, for example, following immunotherapeutic intervention***[8–13]***. Moreover, TCR sequencing has proven valuable in the context of haematological malignancies for early and sensitive detection of minimal residual disease (MRD) in patients with acute or chronic lymphoblastic leukaemia ***[14, 15]***, information critical for the clinical management of these patients.

Despite its promise, repertoire analysis provides many experimental and computational challenges, for which a variety of solutions have been proposed ***[16–19]***. Logistic considerations such as cost, ease of use, robustness and versatility, as well as more scientific issues such as accuracy and coverage may contribute to which solution is optimal for different laboratories. We present here an experimental pipeline for TCR repertoire analysis, which we and others are using for the analysis of repertoires in solid tumours and blood. A computational pipeline with which to analyse the data is presented elsewhere ***[20, 21]***. The defining feature of the protocol is that it uses single-stranded cDNA ligation mediated by T4 RNA ligase, which we have demonstrated to have a high efficiency, in order to incorporate unique molecular identifiers (UMIs)***[22–24]***. UMIs can be used both for sequence error correction and to mitigate for inherent PCR heterogeneity ***[25]***. This allows quantitative estimates of TCR gene abundance. We have successfully used this method to characterise the TCR repertoire from tumour infiltrating lymphocyte (TIL) suspensions and unfractionated frozen tumour tissue from non-small cell lung carcinoma (NSCLC), metastatic melanoma and bladder cancer.

2. **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. We order all oligonucleotides from Sigma-Aldrich (now Merck). We use HPLC 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. Collagenase (*Clostridium histolyticum*, Sigma)
3. RNAse free water
4. Wet Ice
5. 96-Well PCR Plate, Semi-Skirted with Straight Edges (StarLab)
6. RQ1 RNAse-free DNAse kit including 10X buffer, DNAse and DNAse stop buffer (Promega)
7. dNTP Mix (10 mM each) (Thermo Fisher Scientific)
8. Superscript III Reverse transcriptase kit including first strand buffer, DTT and reverse transcriptase (Thermo Fisher Scientific). Superscript IV has also been used successfully.
9. RNasin Ribonuclease Inhibitor (Promega). Alternative RNase inhibitors such as RNaseOUT (Thermo fisher Scientific) can also be used.
10. MinElute Spin kit including Buffer PB, Buffer PE and MinElute columns (Qiagen)
11. Molecular grade Ethanol
12. T4 RNA Ligase 1 (ssRNA Ligase) kit including T4 RNA Ligase reaction buffer, Adenosine-5'-Triphosphate (ATP) and Polyethylene glycol (PEG) 8000 50% (New England Biolabs)
13. 10 mg/ml bovine serum albumin (BSA;New England Biolabs)
14. Hexaammine Cobalt (III) Chloride (HCC) stock solution (100 mM, can be stored at room temperature; Sigma-Aldrich)
15. An aqueous solution containing 1 mg/ml BSA and 10 mM HCC. To make 10 µl of solution, mix 8 µl of RNase free water, 1µl of HCC stock (Materials 14) and 1µl of BSA stock (Materials 13).
16.
17. Agencourt AMpure XP beads (Beckman Coulter)
18. Phusion® High-Fidelity DNA Polymerase kit including HF Buffer (New England BioLabs)
19. MicroAmp™ Fast Optical 96-Well Reaction Plate with barcode (Thermo Fisher Scientific)

20. SYBR™ Green I Nucleic Acid Gel Stain, 10,000X concentrate in DMSO (ThermoFisher). Make a stock solution by diluting this 1:100 in DMSO and store in the dark at -20 °C.
21. ROX Reference Dye (Thermo Fisher Scientific)
22. Micro-Pipettes (10 µl, 20 µl, 200 µl and 1000 µl)
23. 1.5 ml DNase/RNase-free, low-retention microcentrifuge tubes (Thermo Scientific)
24. PCR Plate seals (4titude); used for sealing the plates prior to loading onto the thermocycler (e.g. Ligation, PCR1 and PCR2).
25. Plate seals (StarLab); cheaper plate seals used to seal the plate for steps not involving thermocycler (e.g. purifications, or bead elutions).
26. Optical Plate seals (Applied Biosystem); used for real-time PCR.
27. StarChill PCR Rack, Purple/Pink (StarLab)
28. Desktop Centrifuge/ Plate centrifuge
29. Electric Pipette
30. 96-well Thermocycler
31. 96-well Magnetic Stand (e.g. Invitrogen)
32. 96-well qPCR Thermocycler
33. Qubit dsDNA HS Assay Kit (Thermo Fisher Scientific)
34. Qubit RNA BR Assay Kit (Thermo Fisher Scientific)
35. Qubit Assay Tubes (Thermo Fisher Scientific)
36. Qubit Fluorometer (ThermoFisher, using components 30-32). Alternative fluorometric DNA quantitation technologies could also be used, if they are of equivalent or have greater sensitivity and accuracy.
37. High Sensitivity D1000 ScreenTape (Agilent Technologies)
38. High Sensitivity D1000 Reagents (Agilent Technologies)
39. RNA ScreenTape (Agilent Technologies)
40. RNA ScreenTape Sample Buffer (Agilent Technologies)

41. 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 length of input DNA libraries and/or RNA samples.

42. DNAse I (Roche)

43. GentleMACS Dissociator (Miltenyi)

44. Ficoll-paquePlus (GE Healthcare)

45. RNeasy Mini or micro kits (Qiagen)

46. AllPrep DNA/RNA Mini kit (Qiagen)

3. Methods

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

3.1 Isolation of lymphocytes from human tumour samples.

The protocols for lymphocyte extraction and RNA extraction (section 3.2) have been developed using resected tissue from non-small cell lung carcinoma (NSCLC) collected as part of the TRACERx study ***[26, 27]***. The same methods have also been used with core tissue biopsies. For lymphocyte extraction, fresh tissue samples are digested with collagenase (100 µl/mL) and DNAse I (0.2 mg/mL)) at 37 °C for 30 min. Samples can be subsequently homogenised using a gentleMACS Dissociator and filtered through a 0.7-µm 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)(buffer in RNeasy kits, Materials 46) . 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 PBMCs 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 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 PBMCs is performed following density-gradient centrifugation using Ficoll-Paque PLUS. 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 is used for fewer than 5 x 10^5 cells) or 30µl (Qiagen mini kit). The RNA is quantified fluorometrically (e.g. using the Qubit with RNA BR reagents) and integrity can be assessed using a Bioanalyzer or Tapestation.

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 µl, aiming for a final total amount of 1-1000 ng. For total tumour tissue samples, 500-1000 ng RNA are normally used (we have found 1000 ng to be optimal, where available). 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 µl 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 µl) into a well of a 96-well plate. For concentrated samples, make up to 8 µl with RNAse free water.
3. Add 1 µl of RQ1 10x Buffer and 1 µl 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 min.
5. Briefly centrifuge to collect condensation and add 1 µl 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 min to inactivate the DNase.

3.4 Reverse transcription.

1. Briefly centrifuge the plate and add 4 µl of RNase free water, 1.5 µl of oligonucleotides Alpha RC2 and Beta RC2 (Table 1), and 1.5 µl 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 µl of mix should be added to each sample. Seal the plate.
2. Briefly centrifuge the plate and then place it into the thermocycler at 65 °C for 5 min.
3. Immediately transfer the plate to ice, and leave for at least 1 min.
4. Briefly centrifuge the plate and add 6 µl of 5x FS Buffer, 1.5 µl of 0.1 M DTT, 1.5 µl of RNasin and 1.5 µl of SS III 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 µl of mix should be added to each sample. The total sample volume at this stage should be 30 µl.
5. Briefly centrifuge the plate and then place into the thermocycler at 55 °C for 30 minfollowed by 70 °C for 15 min.
6. Briefly centrifuge the plate and add 150 µl of Buffer PB (MinElute kit, Qiagen) to each sample, mixing well. Apply the plate seal and briefly centrifuge the plate.
7. Transfer the sample (180 µl) into a MinElute column (with collection tube). Proceed with DNA purification as detailed in MinElute kit instructions.
8. Elute the purified DNA in 10.5 µl of water (in order to allow the small volume of water enough time to dissolve the DNA leave on filter of the MinElute column for 1 min before centrifugation). Centrifuge the column at 12,000 g force for 1 min. Purified cDNA should now be in 10 µl at the bottom of a polypropylene tube. Discard the MinElute column.

3.5 Ligation

1. Transfer 10 µl of cDNA from step 3.3 into a new 96-well plate.
2. Mix 3 µl of T4 RNA ligase buffer, 3 µl of 1 mg/ml BSA/ 10 mM HCC solution (Materials 14), 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 RT, and pipette very slowly with a wide-mouth tip. The mixture should be thoroughly mixed by pipetting to ensure uniformity.
3. Cover the plate with a plate seal, briefly centrifuge and then place into the thermocycler at 16 °C for 23 h. Inactivate ligase by incubating at 65 °C for 10 min and then store at 4-10 °C until purification (for a maximum of 24 h).
4. Briefly centrifuge the plate and add 70 µl of water to each sample and mix. Seal and briefly centrifuge.
5. Add 50 µl of AMPure XP beads to each sample, mix by pipetting up and down at least 10 times (avoid creating bubbles) and leave at RT for 5 min. Note that the AMPure XP beads should be used at RT; remove from the fridge 30 min prior to use.
6. Briefly centrifuge the plate and then place onto a magnetic stand (see Materials) and leave it for 2 minor until a pellet is formed.
7. Remove the clear liquid from the sample (ensuring the tip of the pipette does not touch the beads, and that no beads are aspirated).
8. Add 300 µl of freshly prepared 80% ethanol to each sample and leave for 30 s 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 200 µl of 80% ethanol to each sample and leave for 30 s 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 min. Samples can also be dried by placing the 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 32.5 µl of water into each sample and pipette mix at least 10 times. Leave at RT for 2 min.
13. Briefly centrifuge the plate and place onto the plate magnet for 2 min.
14. Transfer 31 µl of the sample into a clean 96-well plate.

3.6 First amplification - PCR 1

1. Add 10 µl of 5x HF Buffer, 2.5 µl of oligonucleotide Alpha RC1 (10 µM, Table 1), 2.5 µl of oligonucleotide Beta RC1 (10 µM) (which is a premixed equimolar mixture of Beta RC1.1 and 1.2, Table 1), 2.5 µl of oligonucleotide SP2 (10 µM), 1 µl of dNTPs (10 mM) and 0.5 µl 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 µl of mix should be added to each sample. The total sample volume at this stage should be 50 µl.
2. Seal and briefly centrifuge the plate and then place it into the thermocycler, and amplify using the cycler program shown in Figure 2.
3. Briefly centrifuge the plate and add 40 µl of prewarmed AMPure XP beads into each sample, mix by pipetting at least 10 times. Seal and leave at RT for 5 min.
4. Briefly centrifuge the plate and then place onto the Magnetic Stand and leave for 2 minor until a pellet is formed.
5. Purify following the procedure described in section section 3.5, steps 7-11. Use 200 µl ethanol for each wash step.
6. Add 65 µl of water into each sample and pipette mix at least 10 times. Seal and leave at room temperature for 2 min.
7. Briefly centrifuge the plate and place onto the plate magnet for 2 min.
8. Transfer 31 µl of each sample into a fresh 96-well plate and label the plate “Alpha”.
9. Transfer another 31 µl of each sample into another fresh 96-well plate and label it “Beta” The Alpha and Beta samples can be transferred onto the same plate, as long as the two sets

of samples are kept physically separate (e.g. Columns 1-6 contain Alpha samples and columns 7-12 contain Beta samples).

3.7 Second amplification - PCR 2

1. The alpha and beta samples can be processed in parallel. Add 2.5 µl of SP1-6N-Ix-aRC1 (1 µM, Table 1) to each sample in the alpha plate, or 2.5 µl of SP1-6N-Ix-bRC1 to each sample in the beta plate. Note that, as shown in Table 1, SP1-6N-Ix-bRC1 is a premixed equimolar mixture of SP1-6N-Ix-bRC1.1 and SP1-6N-Ix-bRC1.2.

   This oligonucleotide introduces a random hexamer at the start of the sequence to augment nucleotide diversity and improve cluster calling during Illumina sequencing. It also introduces a six or eight base pair index read (Ix, which can be chosen from the standard Illumina index sequence sets) which allows multiplexing of many different samples on the same sequencing run.
2. Add 2.5 µl of P7-Iy-SP2 (10 µM, Table 1) to each sample. This oligonucleotide introduces an additional six or eight base pair index to allow further multiplexing. Choose different Ix and Iv index pairs for different samples that are to be sequenced on the same sequencing run. Alpha and beta from the same sample can be indexed with the same index pairs, and separated out at analysis.
3. Add 10 µl of 5x HF Buffer, 2.5 µl of SP1-P5 (10 µM, Table 1), 1 µl of dNTPs and 0.5 µl 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 14 µl of mix should be added to each sample. The total sample volume at this stage should be 50 µl.
4. Seal and briefly centrifuge alpha and beta plates, then place into a thermocycler, and amplify using the cycler program shown in Figure 3.

5. Briefly centrifuge each plate and then add 40 µl of pre-warmed AMPure XP beads into each sample, mix by pipetting at least 10 times, seal and leave at RT for 5 min.
6. Briefly centrifuge and then place plate on the Magnetic Stand, and leave for 2 minutes or until a pellet is formed.
7. Purify following the procedure described in section 3.5, steps 7-11.
8. Elute each sample in 30 µl of water and pipette mix at least 10 times. Seal and leave at RT for 2 min.
9. Briefly centrifuge the plate and place onto the plate magnet for 2 min.

3.8 Third amplification - PCR 3 (real time PCR).

1. Transfer 27.2 µl 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, 2.17) by mixing 1 µl of stock SYBR green and 49 µl of water. This should be made up fresh immediately before use, since SYBR green is poorly soluble in water.
10. Add 10 µl of 5x HF buffer, 5 µl of SYBR green dilution, 1.25 µl of dNTPs, 1 µl of ROX, 2.5 µl of P5 primer, 2.5 µl of P7 primer and 0.5 µl 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 µl of the mix should be added to each sample. The total sample volume at this stage should be 50 µl. Apply the optical plate seal.
11. Briefly centrifuge the plate and then place into a real time thermocycler, and amplify using the cycler program shown in Figure 4. 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 5.

12. Once the samples have reached the predefined threshold value, transfer the plate to a standard thermocycler for a final 5 min extension at 72 °C and then stop the thermocycler. Briefly centrifuge the plate.
13. 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 µl 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 RT for 5 min.
14. Purify following the procedure described in section 3.5, steps 7-11.
15. Elute each sample in 30 µl of water. Seal and leave at room temperature for 2 min.
16. Briefly centrifuge the plate and place onto the plate magnet for 2 min.
17. Transfer 28 µl of each purified library into a labelled 1.5 ml 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 6.
2. The final libraries are prepared by mixing samples with different indices. Improved sequencing is obtained if the final pooled library is further purified by electrophoresis. We use a Pippin Prep (Sage science), collecting fragments of between 350 and 750 base pairs (bp).
3. For Illumina MiSeq sequencing, each sample is present at 4 nM in the final library, and the library is then diluted, denatured and a subsample is sequenced. 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 10-15 pM. 5-7% PhiX DNA at 14-20 pM is added to the library to increase diversity and improve cluster recognition. Denaturation and sequencing follows the standard Illumina protocols.

4. For Illumina NextSeq sequencing, the final library is also prepared at 4 nM. Typically, at least 60 samples can be run on the flow cell, yielding 1-2 million reads per sample, when using the Mid-Output v2 300 cycle kit. Sample is loaded at 1 pM with 15-20% of PhiX DNA at 1.8 pM. Denaturation and sequencing follows standard Illumina protocols.

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 mRNA. The method is potentially sensitive to changes in mRNA 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 ***[28–31]***. Unpublished experiments from our laboratory suggest that each TCR alpha or beta chain mRNA is present in at least 50 copies in each 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. PCR2 introduces several features required for subsequent sequencing on the Illumina platform, including the first sequencing primer SP1, an Illumina index for sequence multiplexing 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. A second multiplexing index sequence is also introduced immediately 3' of the constant region. The option to introduce two independent indices is important when considering high throughput experiments with many samples, which becomes even more important when using larger throughput machines such as the NextSeq, HiSeq or NovaSeq series. Rotating through different indexes across multiple runs also protects against cross contamination, in which sequencing-compatible molecules can carry over from one run to a later one.

5. 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 £14 (approximately 16-18 Euros or US dollars) per sample, excluding sequencing. The most expensive component is the reverse transcriptase SuperScript III. 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.

Acknowledgements

This research was funded by grants from CancerResearchUK, the UK MRC, the Rosetrees Trust, the Stoneygate Foundation, a grant to BC from Unilever PLC, and supported by the National Institute for Health Research UCL Hospitals Biomedical Research. JH was supported by an MRC studentship. K.J. is funded by the Sam Keen Foundation and UCL/Royal Marsden Hospital NIHR Biomedical Research Centres. We are grateful to the TRACERx consortium, funded by Cancer Research UK (http://www.cruklungcentre.org/Research/TRACERx) for giving us access to the tumour samples which the amplified TCR libraries illustrated in figs 5 and 6 were prepared.

References

1. Heather JM and Chain B (2016) The sequence of sequencers: The history of sequencing DNA. Genomics 107:1–8

2. Pasetto A, Gros A, Robbins PF, et al (2016) Tumor- and Neoantigen-Reactive T-cell Receptors Can Be Identified Based on Their Frequency in Fresh Tumor. Cancer Immunol Res 4:734–43

3. Gros A, Robbins PF, Yao X, et al (2014) PD-1 identifies the patient-specific CD8+ tumor-reactive repertoire infiltrating human tumors. J Clin Invest 124:2246–2259

4. Gerlinger M, Quezada SA, Peggs KS, et al (2013) Ultra-deep T-cell receptor sequencing reveals the complexity and intratumour heterogeneity of T-cell clones in renal cell carcinomas. 231:424–432

5. Chen Z, Zhang C, Pan Y, et al (2016) T cell receptor β-chain repertoire analysis reveals intratumour heterogeneity of tumour-infiltrating lymphocytes in oesophageal squamous cell carcinoma. J Pathol 239:450–8

6. Reuben A, Gittelman R, Gao J, et al (2017) TCR Repertoire Intratumor Heterogeneity in Localized Lung Adenocarcinomas: An Association with Predicted Neoantigen Heterogeneity and Postsurgical Recurrence. Cancer Discov 7:1088–1097

7. Emerson RO, Sherwood AM, Rieder MJ, et al (2013) High-throughput sequencing of T-cell receptors reveals a homogeneous repertoire of tumour-infiltrating lymphocytes in ovarian cancer. J Pathol 231:433–440

8. Tumeh PC, Harview CL, Yearley JH, et al (2014) PD-1 blockade induces responses by inhibiting adaptive immune resistance. Nature 515:568–71

9. Snyder A, Nathanson T, Funt SA, et al (2017) Contribution of systemic and somatic factors to clinical response and resistance to PD-L1 blockade in urothelial cancer: An exploratory multi-omic analysis. PLoS Med 14:e1002309

10. Robert L, Tsoi J, Wang X, et al (2014) CTLA4 blockade broadens the peripheral T-cell receptor repertoire. Clin Cancer Res 20:2424–32

11. Cha E, Klinger M, Hou Y, et al (2014) Improved Survival with T Cell Clonotype Stability After Anti-CTLA-4 Treatment in Cancer Patients. Sci Transl Med 6:238ra70-238ra70

12. Kvistborg P, Philips D, Kelderman S, et al (2014) Anti-CTLA-4 therapy broadens the melanoma-reactive CD8+ T cell response. Sci Transl Med 6:254ra128-254ra128

13. Riaz N, Havel JJ, Makarov V, et al (2017) Tumor and Microenvironment Evolution during Immunotherapy with Nivolumab. Cell 171:934–949.e15

14. Wu D, Sherwood A, Fromm JR, et al (2012) High-throughput sequencing detects minimal residual disease in acute T lymphoblastic leukemia. 4:134ra63

15. Logan AC, Gao H, Wang C, et al (2011) High-throughput VDJ sequencing for quantification of minimal residual disease in chronic lymphocytic leukemia and immune reconstitution assessment. Proc Natl Acad Sci 108:21194–21199

16. Hou X-L, Wang L, Ding Y-L, et al (2016) Current status and recent advances of next generation sequencing techniques in immunological repertoire. Genes Immun 17:153–64

17. Robins H (2013) Immunosequencing: applications of immune repertoire deep sequencing. Curr Opin Immunol 25:646–52

18. Six A, Mariotti-Ferrandiz ME, Chaara W, et al (2013) The past, present, and future of immune repertoire biology - the rise of next-generation repertoire analysis. Front Immunol 4:1–16

19. Heather JM, Ismail M, Oakes T, et al (2017) High-throughput sequencing of the T-cell receptor repertoire: pitfalls and opportunities. Brief Bioinform bbw138

20. 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. 29:542–550

21. 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. Front Immunol 8:1267

22. Weinstein JA, Jiang N, White RA, et al (2009) High-throughput sequencing of the zebrafish antibody repertoire. Science 324:807–10

23. Shugay M, Britanova O V, Merzlyak EM, et al (2014) Towards error-free profiling of immune repertoires. Nat Methods 11:653–5

24. 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. Front Immunol 4:1–10

25. Best K, Oakes T, Heather JM, et al (2015) Computational analysis of stochastic heterogeneity

in PCR amplification efficiency revealed by single molecule barcoding. Sci Rep 5:14629

26. Jamal-Hanjani M, Hackshaw A, Ngai Y, et al (2014) Tracking Genomic Cancer Evolution for Precision Medicine: The Lung TRACERx Study. PLoS Biol 12:e1001906

27. Jamal-Hanjani M, Wilson GA, McGranahan N, et al (2017) Tracking the Evolution of Non–Small-Cell Lung Cancer. N Engl J Med 376:2109–2121

28. Gallard A, Foucras G, Coureau C, et al (2002) Tracking T cell clonotypes in complex T lymphocyte populations by real-time quantitative PCR using fluorogenic complementarity-determining region-3-specific probes. J Immunol Methods 270:269–80

29. Lindsten T, June CH, and Thompson CB (1988) Transcription of T cell antigen receptor genes is induced by protein kinase C activation. J Immunol 141:1769–74

30. Paillard F, Sterkers G, and Vaquero C (1990) Transcriptional and post-transcriptional regulation of TcR, CD4 and CD8 gene expression during activation of normal human T lymphocytes. EMBO J 9:1867–72

31. Paillard F, Sterkers G, and Vaquero C (1992) Correlation between up-regulation of lymphokine mRNA and down-regulation of TcR, CD4, CD8 and lck mRNA as shown by the effect of CsA on activated T lymphocytes. Biochem Biophys Res Commun 186:603–11

Legends

Figure 1. Diagrammatic overview of library preparation. Numbers in brackets refer to numbered sections in text.

Figure 2: PCR 1 thermocycler program set-up.

Figure 3: PCR 2 thermocycler program set-up.

Figure 4: PCR 3 real-time thermocycler program set-up.

Fig 5. Real time PCR library amplification (PCR3, section 3.8). Yellow and green lines represent alpha and beta TCR genes amplified from 1000 ng of lung tumour RNA.

Fig 6. Electrophoretic (Tapestation) analysis of amplified TCR libraries prepared from 1000 ng RNA extracted from lung tumour samples. 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/µl 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 400 bp, as demonstrated in panel B, shaded arrow). 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.

Appendix 1. The TRACERx consortium

Jason Lester	Anita Amadi	Madava Djearaman
Fiona Morgan	Marianne Nicolson	Yvonne Summers
Malgorzata Kornaszewska	Keith Kerr	Raffaele Califano
Richard Attanoos	Shirley Palmer	Paul Taylor
Haydn Adams	Hardy Remmen	Rajesh Shah
Helen Davies	Joy Miller	Piotr Krysiak
Dean Fennell	Keith Buchan	Kendadai Rammohan
Jacqui Shaw	Mahendran Chetty	Eustace Fontaine
John Le Quesne	Lesley Gomersall	Richard Booton
Apostolos Nakas	Sara Lock	Matthew Evison
Sridhar Rathinam	Azmina Verjee	Phil Crosbie
William Monteiro	Babu Naidu	Stuart Moss
Hilary Marshall	Gerald Langman	Faiza Idries
Louise Nelson	Simon Trotter	leena Joseph
Jonathan Bennett	Mary Bellamy	Paul Bishop
Joan Riley	Hollie Bancroft	Anshuman Chaturvedi
Lindsay Primrose	Amy Kerr	Anne Marie Quinn
Luke Martinson	Salma Kadiri	Helen Doran
Girija Anand	Joanne Webb	Angela leek
Sajid Khan	Gary Middleton	Phil Harrison

Katrina Moore

Rachael Waddington

Juliette Novasio

Fiona Blackhall

Jane Rogan

Elaine Smith

Caroline Dive

Ged Brady

Dominic Rothwell

Sakshi Gulati

Francesca Chemie

Jonathan Tugwood

Jackie Pierce

David Lawrence

Davide Patrini

Martin Hayward

Nikolaos Panagiotopoulos

Shyam Kolvekar

Mary Falzon

Elaine Borg

Gemma Hector

Amy Smith

Marie Aranda

Charles Swanton

Mariam Jamal-Hanjani

Gareth Wilson

Nicolai juul Birkbak

Thomas Watkins

Nicholas McGranahan

Christopher Abbosh

Rachel Rosenthal

Stuart Horswell

Richard Mitter

Mickael Escudero

Aengus Stewart

Andrew Rowan

Crispin Hiley

Jacki Goldman

Asia Ahmed

Magali Taylor

Junaid Choudhary

Penny Shaw

Raju Veeriah

Justyna Czyzewska-Khan

Diana Johnson

Joanne Laycock

Robert Hynds

Kroopa Joshi

Andrew Furness

Assma Ben Aissa

Yien Ning Sophia Wong

Andy Georgiou

Mariana Werner

Sunderland

James Reading

Sergio Quezada

Karl Peggs

Teresa Marafioti

Marco Novelli

Dahmane Oukrif

Sam Janes

Martin Forster

Tanya Ahmad

Siow Ming Lee

Peter van Loo

Javier Herrero

John Hartley

Richard Kevin Stone

Tamara Denner

Marta Costa

Sharmin Begum

Ben Phillimore

Tim Chambers

Emma Nye

Sofia Graca

Sophie Ward

Greg Elgar

Maise Al-Bakir

Dawn Carnell

Ruheena Mendes

Jeremy George

Neal Navani

Dionysis Papadatos-Pastos

Marco Scarci

Pat Gorman

Helen Lowe

Leah Ensell

David Moore

Mairead MacKenzie

Maggie Wilcox

Harriet Bell

Allan Hackshaw

Yenting Ngai

Sean Smith

Nicole Gower

Christian Ottensmeier

Serena Chee

Benjamin Johnson

Aiman Alzetani

Emily Shaw

Eric Lim

Paulo De Sousa

Monica Tavares Barbosa

Andrew Nicholson

Alex Bowman

Simon Jordan

Alexandra Rice

Hilgardt Raubenheimer

Chiara Proli

Maria Elena Cufari

John Carlo Ronquillo

Angela Kwayie

Harshil Bhayani

Morag Hamilton

Yusura Bakar

Natalie Mensah

Lyn Ambrose

Anand Devaraj

Silviu Buderi

Jonathan Finch

Leire Azcarate	Sarah Danson	Sarah Feeney
Hema Chavan	Jonathan Bury	Julius Asante-Siaw
Sophie Green	John Edwards	Peter Russell
Hillaria Mashinga	Jennifer Hill	Teresa Light
Kelvin Lau	Sue Matthews	Nikolaos Panagiotopoulos
Michael Sheaff	Yota Kitsanta	Tracey Horey
Peter Schmid	Kim Suvarna	Kevin Blyth
John Conibear	Michael Shackcloth	Craig Dick
Veni Ezhil	John Gosney	Alan Kirk
Vineet Prakash	Pieter Postmus	

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	**[1]P**-NNN NNN ATCACGAC NNN NNN ATCACGAC AGA TCG GAA GAG CAC ACG TCT GAA CTC CAG-**C3**	3.5	5'phosphate; 3'C3 spacer
Alpha RC1	ATCTACACTACGGCAGGGTCAGGGTTCTGGATAT	3.6	
Beta RC1.1[2]	TACACTGGTGGGAACACCTTGTTCAGGTCCTC	3.6	
Beta RC1.2[2]	ATCTACACTGGTGGGAACACGTTTTTCAGGTCCTC	3.6	
SP1-6N-Ix-1-aRC1[3,4]	ACACTCTTTCCCTACACGACGCTCTTCCGATCTNNNNNNATCACGACGGCAGGGTCAGGGTTCTGGATAT	3.7	
SP1-6N-Ix-1-bRC1.1[2,4]	ACACTCTTTCCCTACACGACGCTCTTCCGATCTNNNNNNATCACGGGTGGGAACACCTTGTTCAGGTCCTC	3.7	
SP1-6N-Ix-1-bRC1.2[2,4]	ACACTCTTTCCCTACACGACGCTCTTCCGATCTNNNNNNATCACGGGTGGGAACACGTTTTTCAGGTCCTC	3.7	
P7-Iy-SP2[4]	CAAGCAGAAGACGGCATACGAGATATCACGGTGACTGGAGTTCAGACGTGTGCTCTTCCGATC	3.7	
SP1-P5[6]	AATGATACGGCGACCACCGAGATCTACACTCTTTCCCTACACGACGCTCTTCC	3.5	
SP2[6]	GTGACTGGAGTTCAGACGTGTGCTCTTCCGATCT	3.6	
P5[6]	AATGATACGGCGACCACCGAGATCTACACT	3.8	
P7[6]	CAAGCAGAAGACGGCATACGAGAT	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 can be inserted randomly
4. The “x” refers to the index number. Underlined bases represent a representative index sequence, and can be substituted by any of the Illumina indices. The indices allow multiple libraries to be combined for sequencing, and then deconvoluted for analysis.
5. The “y” refers to the index number. We have used “y” to indicate the index number, to highlight that the index sequences at the P5 and P7 end of the amplicon can be different, thus allowing many potential combinations of primers to be used. The underlined bases represent a representative index sequence, and can be substituted by any of the Illumina indices.
6. 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.

Alan Kirk