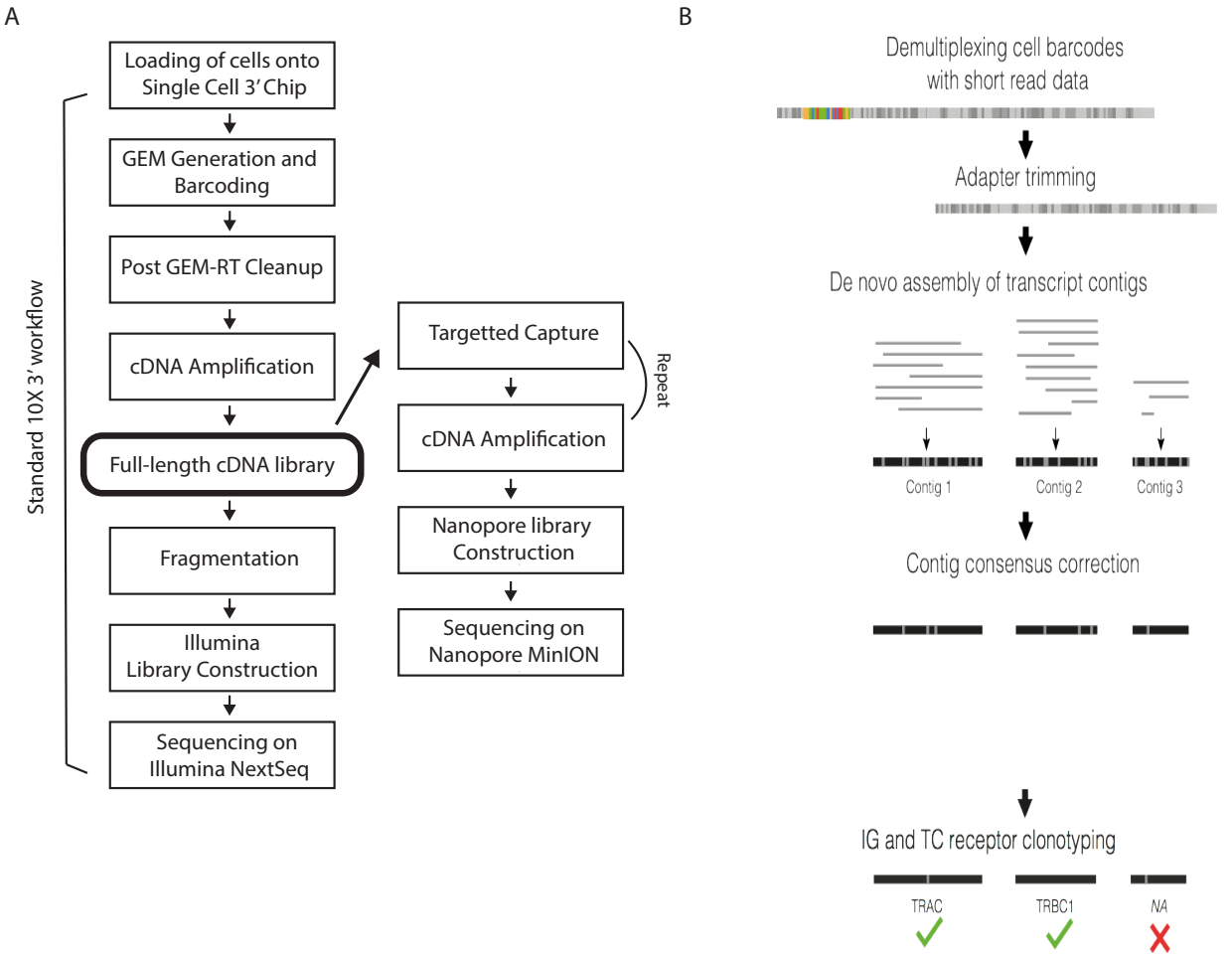
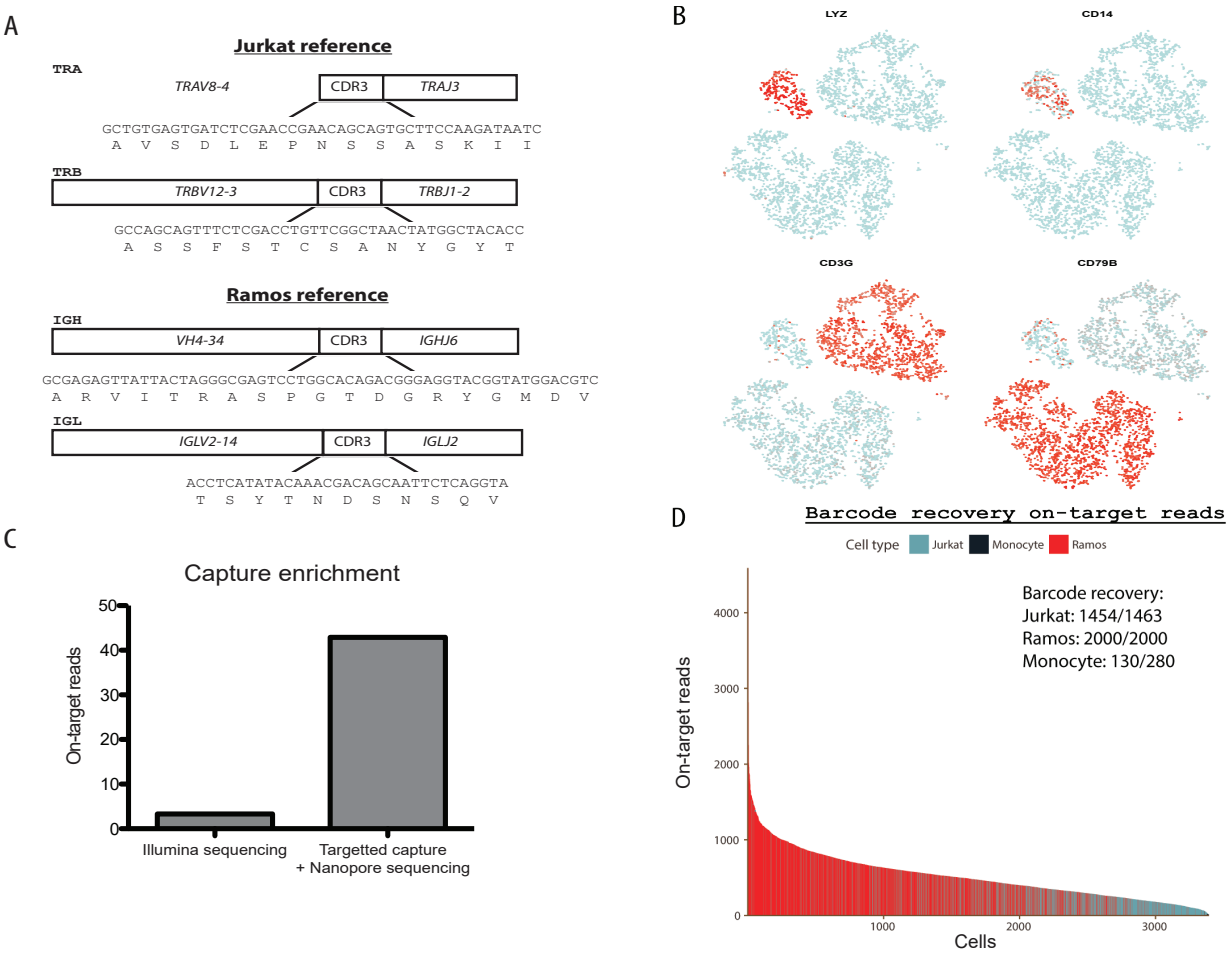
RAGEseq: Supplementary material

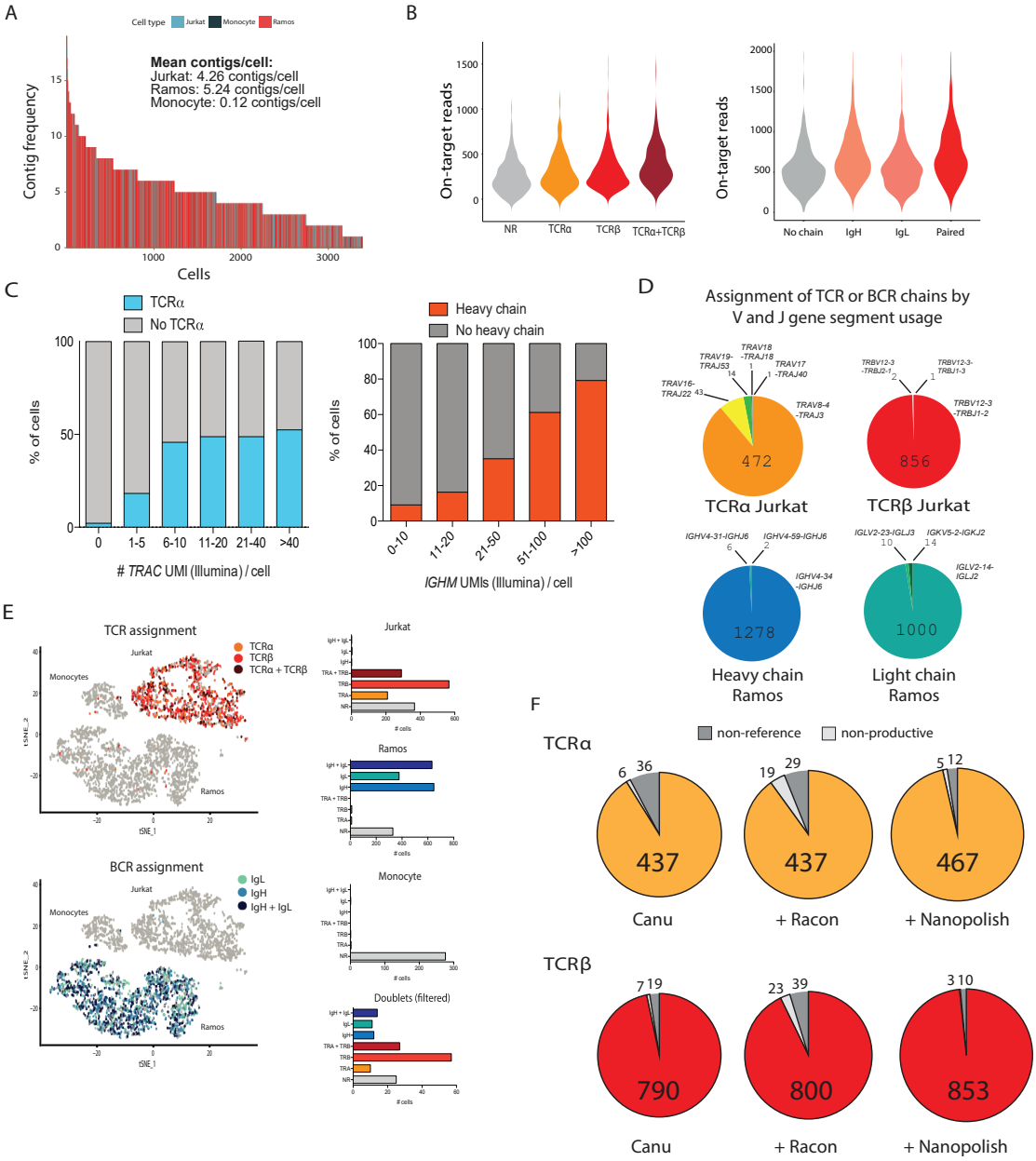
Description: Supplementary figures and supplementary tables.



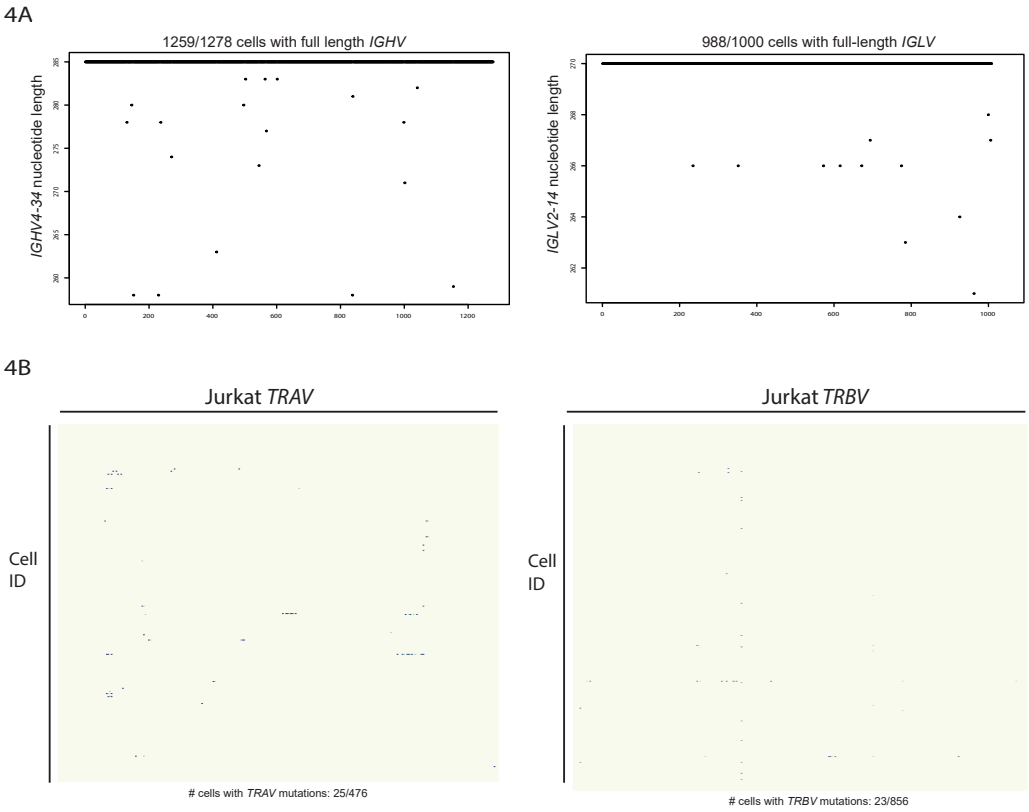
Supplementary Figure 1. RAGE-Seq experimental protocol and computational pipeline. A) Detailed experimental workflow of RAGE-Seq. Single-cell suspensions are loaded onto 3' library chips for the Chromium Single Cell 3' Library according to the manufacturer's recommendations (10X Genomics). Single cells are partitioned into Gel Beads in Emulsion (GEMs) with cell lysis and barcoded reverse transcription of RNA. Following isolation of GEMs from the Chromium instrument, full-length cDNA is PCR amplified. At this point the cDNA library is split into two fractions. The first fraction undergoes the remainder of the Chromium workflow which involves shearing, adapter ligation and Illumina sequencing. The second fraction undergoes two rounds of targeted capture of TCR and BCR genes followed by ONT library preparation using the 1D adapter ligation sequencing kit and sequencing on the MinION platform. (B) computational pipeline of RAGE-Seq.



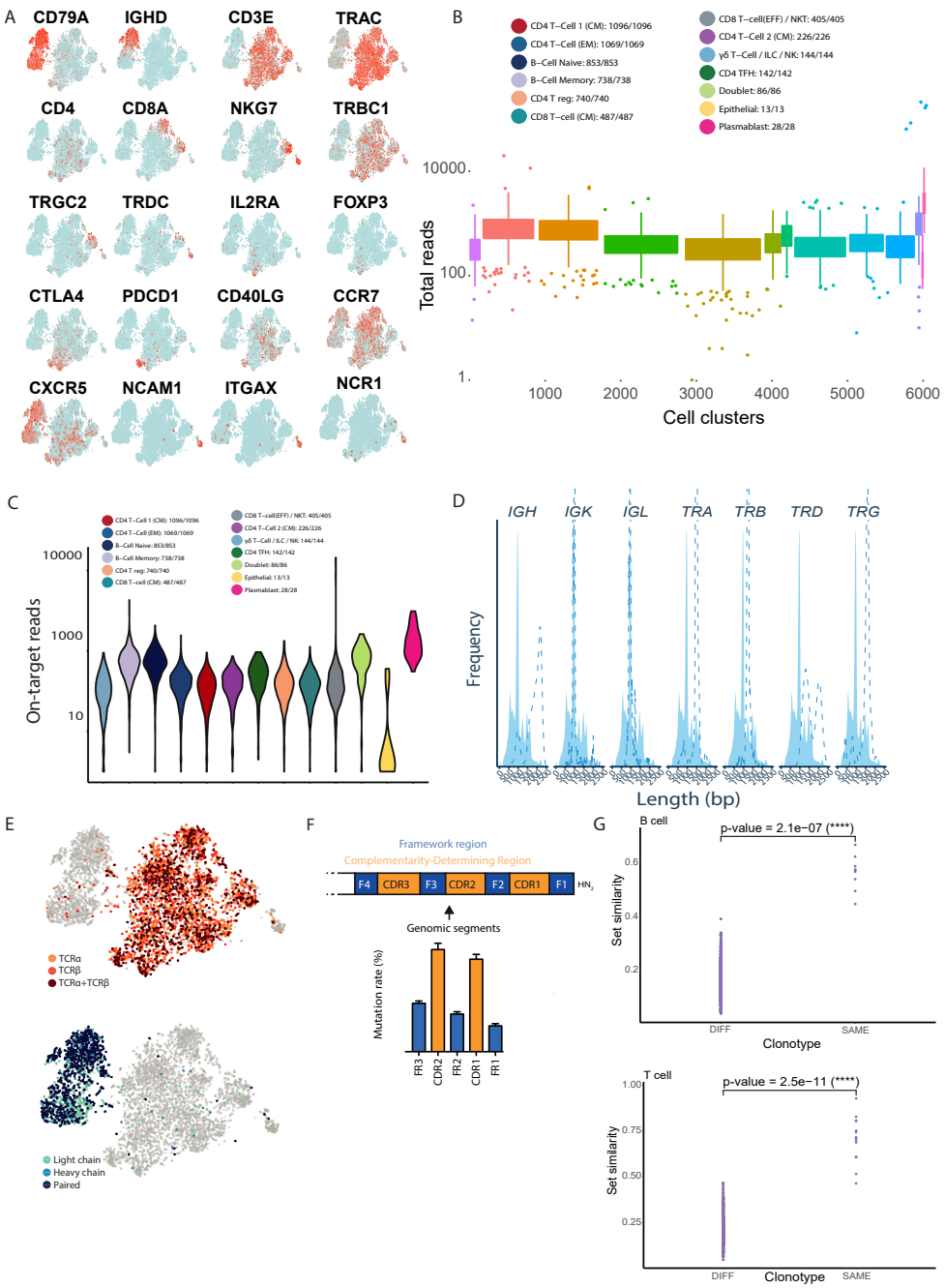
Supplementary Figure 2. RAGE-Seq cross-sequencing platform quality control measurements A) Reference CDR3, V and J gene segments that encode the TCRα and TCRβ chains of Jurkat or the immunoglobulin heavy and light chains of Ramos. For Ramos the most abundant heavy chain CDR3 (1123/1278 cells) and light chain CDR3 sequences (104/937) were chosen as the reference. B) tSNE of key canonical gene expression markers used to identify cell type of cluster. *CD3G* was used to identify Jurkat cells, *CD79B* was used to identify Ramos cells and *LYZ* and *CD14* were used to identify monocyte cells. (C) The relative enrichment of targeted capture of antigen receptor genes. On-target reads were determined by the percentage of total Nanopore or Illumina sequencing reads that align to TRA, TRB, IGH, IGL and IGK constant region genes. (D) Nanopore cell barcode recovery for Nanopore reads that are on-target. Mean on-target reads per cell type: Jurkat, 309; Ramos, 646; Monocyte, 1.49.



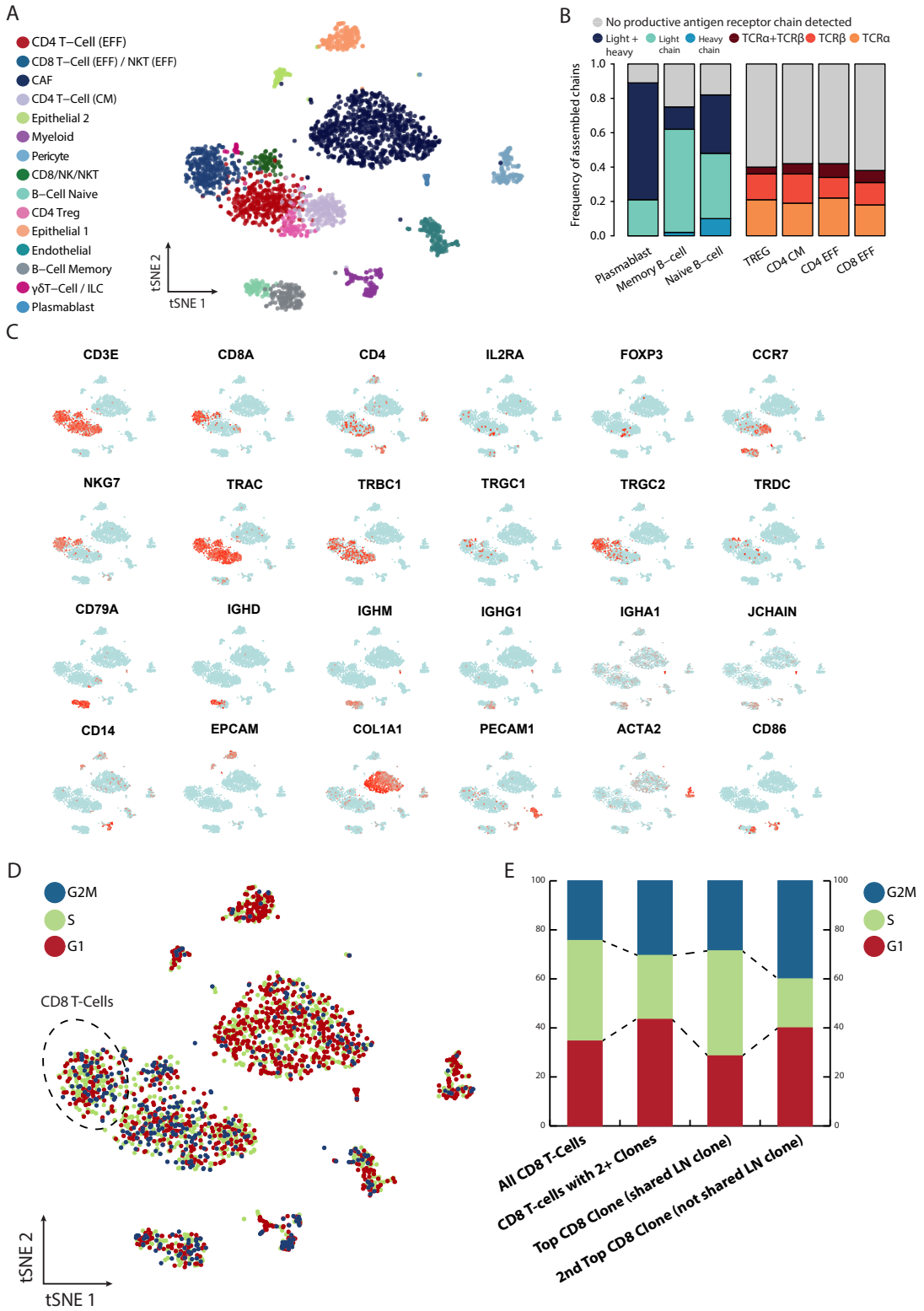
Supplementary Figure 3. Quality control measurements of antigen receptor assembly. (A) Mean number of contigs assembled per cell. Each bar corresponds to an individual cell. (B) The number of on-target Nanopore reads for Jurkat (left panel) or Ramos (right panel) grouped by the recovery of TCR chains or BCR chains, respectively. NR, no receptor. Only those TCR and BCR chains that match their reference V and J gene and contain in-frame CDR3 sequences and lack stop codons, termed a productive clonotype, were assigned. (C) The recovery of Jurkat cells assigned a TCRα chain (left panel) or Ramos cells assigned a Immunoglobulin heavy chain (right panel) as a function of the number of Illumina UMIs *TRAC* (Jurkat) or *IGHM* (Ramos) genes per cell. (D) Assignment of TCR chains to Jurkat cells or BCR chains to Ramos cells based on their V and J gene segment usage. Shown in each pie graph is the number of cells expressing the designated V and J genes. Only productive clonotypes are assigned. (E) tSNE plot of Jurkat, Ramos and monocyte cells (Fig. 2A) assigned TCR (top panel) or BCR (bottom panel) chains. Right panels show the total number of cells assigned different chains for each cell type. Doublets (n=136 cells) are not shown on the tSNE plots and were filtered out based on high gene count (see Methods). (F) Accuracy of CDR3 sequences of Jurkat cells at each stage of contig assembly and polishing. Shown are the number of cells assigned a CDR3 sequence that match the reference *TRA* or *TRB* CDR3. 'Non-reference' refers to a CDR3 sequence that does not match the reference Jurkat CDR3. 'Non-productive' refers to TCR chains with a CDR3 sequence that is out-of-frame or contains stop-codons and are usually filtered from the dataset. Only TCR chains that match the Jurkat reference V and J gene segments are assigned.



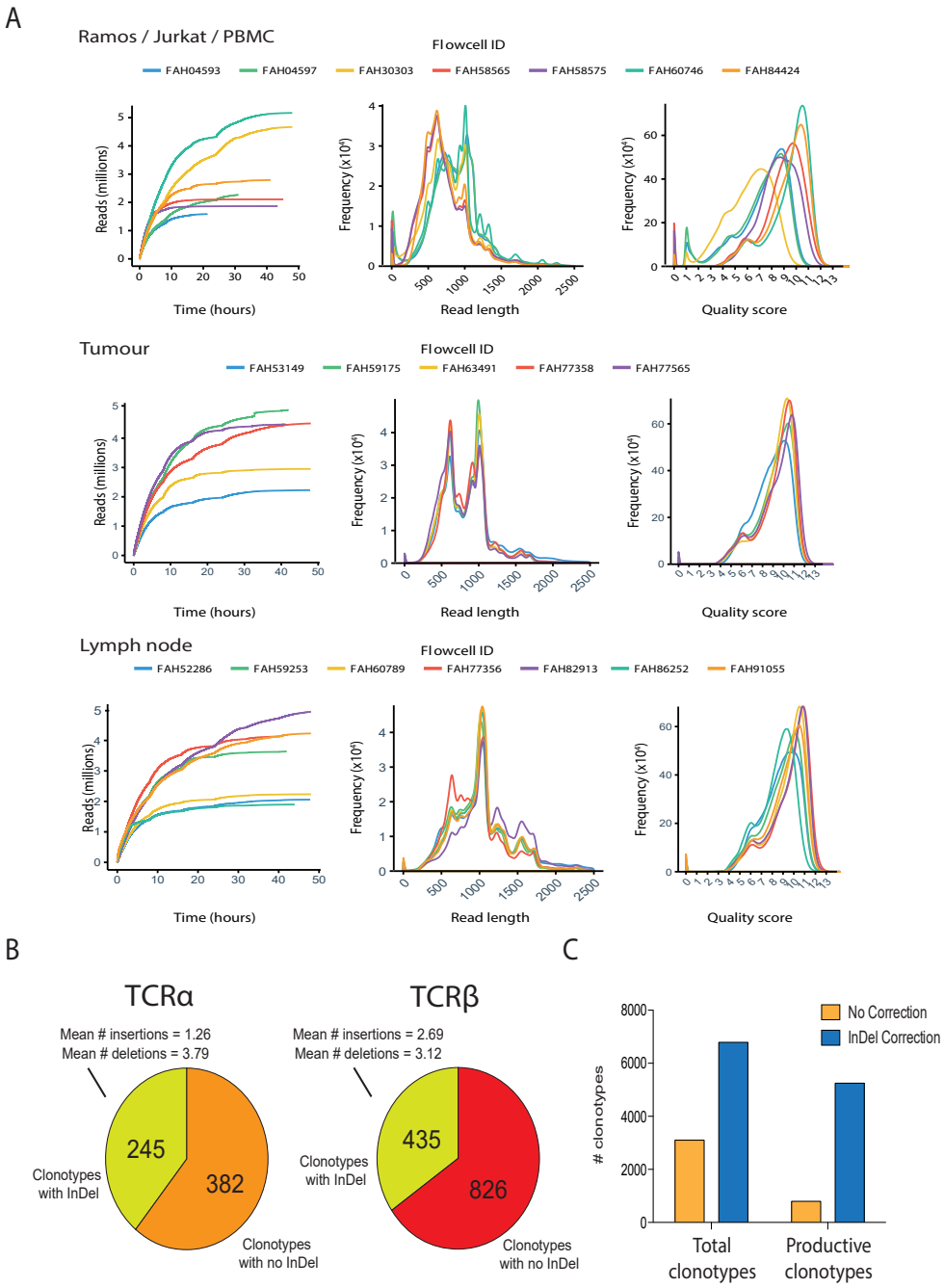
Supplementary Figure 4. Quality control measurements of calling somatic hypermutation in individual cells. A) Nucleotide length of Ramos heavy chain (left) and light chain (right) V regions. The maximum length of the entire IGVH4-34 (left) or IGLV2-13 (right) gene is indicated. (B) Heatmap of the V regions of individual Jurkat cells encoding the TCRα (right) and TCRβ (left) chain. Each row represents an individual cell and each column a nucleotide position in the respective V gene. Light blue represents synonymous nucleotide substitutions while dark blue represents non-synonymous nucleotide substitutions, when compared to germline TRAV8-4 and TRBV12-3 sequences.



Supplementary Figure 5. Additional measurements of RAGE-seq on a human lymph node. (A) tSNE of key canonical gene expression markers used to identify cell type of cluster. (B) The recovery of cell barcodes following Nanopore sequencing for each cell population identified by sc-RNA-seq. (C) The number of on-target Nanopore reads for each cell population identified in the lymph node. (D) The nucleotide length for assembled antigen receptor transcripts for each receptor chain identified across all cells in the lymph node. The overall Nanopore sequencing read length distribution is shown as shaded and the assembled contigs as dashed lines. (E) Assignment of TCR and BCR chains for each cell in the lymph node. (F) Mutation rate of the framework and complementarity regions of the heavy chain that have been assigned to memory B cells. (G) Jaccard set similarity score of top 250 raw UMI gene counts across all B cells (top) and T cells (bottom) with shared (same) V(D)J sequences and those with dissimilar V(D)J sequences within each respective cell type cluster. Significance was calculated via the corrected Wilcoxon test.



Supplementary Figure 6. Additional measurements of RAGE-seq on a tumour. (A) tSNE plot associated with 3' gene expression profiling of a tumour sample captured on the 10X Chromium platform. (B) Assignment of TCR chains to each T population and BCR chains to each B cell population identified in (A). (C) tSNE plots of key canonical gene expression markers used to identify cell types in (A). D) Cell cycle phase of all cells in Tumour with tSNE structure overlay. E) Proportion of cell cycle phase of all CD8 T-cells cells and expanded clones within tumour. Top CD8 clone: "TRBV7-9 TRBJ2-3: ASSLAGRVPGDTQY" and second top CD8 clone: "TRBV7-9 TRBJ2-2: ASSLELTGELF".



Supplementary Figure 7. Nanopore sequencing statistics. A) Number of reads, read length and quality score for cell line, Tumour and Lymph node samples. Colour denotes each flowcell the run was performed on (R9.4 or R9.5 chemistry) B) The total number of TCR chains assigned to Jurkat cells (n=1463) that carry InDels in their TRA or TRB V gene before InDel correction (See Methods). TCR chains include those with non-productive CDR3 sequences. C) The effect of InDel correction on antigen receptor chain recovery. Shown are all productive BCR and TCR chains recovered across all Jurkat, Ramos and monocyte cells. Total clonotypes refers to both productive and non-productive clonotypes.

Supplementary Table 1. Number of total Nanopore reads and de-multiplexed Nanopore reads per sample.

Sample	Total reads	Total on-target reads	Percent de-multiplexed
Cell line	20,346,396	3,805,076	18.7
Lymph node	21,721,568	3,380,621	19.9
Tumour	16,601,436	3,069,468	15.8

Supplementary Table 2. Comparison of RAGE-Seq against SmartSeq2

	# Jurkat cells	TRA recovery	TRB recovery	Paired	Cost per cell (AUD)
RAGE-Seq	1463	472	856	277	4.03
SmartSeq + VDJ-Puzzle	28	22	25	21	91.78

Supplementary Table 3. Cost-breakdown of RAGE-Seq for cell line experiment. RAGEseq had 3,743 cells in the final dataset.

Items	Supplier	Cost per cell (AUD)
RAGEseq		
Capture probes (NimbleGen)	Roche	0.31
SeqCap EZ accessory kit	Roche	0.013
Hybridisation and wash kit	Roche	0.0053
KAPA Hotstart HiFi ReadyMix	Kapa Biosystems	0.0031
Dynabeads M-270 Streptavidin	ThermoFisher	0.018
AMPure XP magnetic beads	Agencourt	0.0055
10X Chromium capture and library preparation	10X Genomics	0.74
MinION library prep and sequencing (x7)	Oxford Nanopore	1.64
NextSeq 500/550 Mid Output v2 kit (150 cycles)	Illumina	0.36
NextSeq 500/550 High Output v2 kit (150 cycles)	Illumina	0.94
SMARTseq		
Recombinant RNase Inhibitor	Clontech	2.04
SuperScript II Reverse Transcriptase	Invitrogen	3.48
KAPA Hotstart HiFi ReadyMix	Kapa Biosystems	1.22
Nextera XT Index Kit	Illumina	3.15
Nextera XT DNA Library Prep Kit	Illumina	39.0
AMPure XP magnetic beads	Agencourt	1.75
LabChip GX Touch 24	PerkinElmer	1.50
PicoGreen	ThermoFisher	0.82
dNTP Mix	ThermoFisher	0.14
Oligonucleotides	Exiqon	3.40
NextSeq 500/550 High Output v2 kit (300 cycles) ∼2.5 x 10^6 reads/cell	Illumina	35.28

Supplementary Table 4. Assignment of TCRα and TCRβ chains to cells assigned TCRγ and/or TCRδ chains.

Assigned chain (number of cells)	TCRα	TCRβ	TCR$\alpha\beta$
TCRδ only (14)	1	1	0
TCRγ only (84)	16	12	5
TCRγ + TCRδ (11)	0	0	0

Supplementary Table 5. List of samples, chemistries, flowcell identification numbers, and manufacturer software versions.

Sample	Flowcell ID	Flowcell chemistry	Kit	Albacore version
Tumour	FAH59175	FLO-MIN106	SQK-LSK108	2.2.7
	FAH63491			
	FAH77585			
	FAH77358			
	FAH53149			2.1.3
	FAH89946			Guppy
Lymph Node	FAH59253	FLO-MIN106	SQK-LSK108	2.2.7
	FAH60789			
	FAH77356			
	FAH82913			
	FAH52286			2.1.3
	FAH86252			Guppy
	FAH82560			
Ramos & Jurkat	FAH58575/FAH58565	FLO-MIN106	SQK-LSK108	2.2.7
	FAH60746			
	FAH84424			
	FAH04597	FLO-MIN107	LSK308 (basecalled SQK-LSK108)	2.1.3
	FAH04593		SQK-LSK108	
	FAH30303			