circVDJ-seq for T cell clonotype detection in single-cell and spatial multi-omics

Izabela Plumbom[†,1,2], Benedikt Obermayer[†,2], Raphael Raspe[3], Anna Pascual-Reguant [5,6], Ilan Theurillat[1], Tancredi Massimo Pentimalli[1,3], Yu-Hsin Hsieh[1,2], Marine Gil[1,2], Carola Dietrich[1,2], Michaela Seeger-Zografakis [1,2], Claudia Quedenau[1,2], Jeannine Wilde[1,2], Caroline Braeuning[1,2], Cornelius Fischer[1,2], Markus Schuelke[3], Leif S. Ludwig[1,2], Angelika Eggert[3], Nikolaus Rajewsky[1], Tatiana Borodina[1,2], Dieter Beule[2], Janine Altmueller[1,2], Helena Radbruch[3], Anja Hauser[3,4], Thomas Conrad[*,1,2]

[1] *Max Delbrück Center for Molecular Medicine in the Helmholtz Association, Berlin Germany*

[2] *Berlin Institute of Health (BIH) at Charité - Universitätsmedizin Berlin, Germany*

[3] *Charité - Universitätsmedizin Berlin, Germany*

[4] *German Rheumatism Research Center Berlin (DRFZ), Germany*

[5] *Centre Nacional d'Anàlisis Genòmica, Barcelona*

[6] *Universitat de Barcelona (UB), Barcelona, Spain*

[†] *Izabela Plumbom and Benedikt Obermayer contributed equally to this work*

* Correspondence: Thomas.Conrad@mdc-berlin.de

Abstract

Monitoring T cell clonality in human tissues provides important insights into adaptive immune response mechanisms in cancer, infectious diseases, and autoimmunity. However, retrieving VDJ sequence information from single-cell and spatial transcriptomics workflows with 3'-barcoding of cDNA remains resource-intensive or requires specialized sequencing equipment. Here, we introduce circVDJ-seq for simplified and cost-efficient TCR profiling from 3'-directed workflows such as single-nucleus RNA sequencing, RNA+ATAC multi-omics, and spatial transcriptomics. Application of circVDJ-seq to freshly resected neuroblastomas, and post-mortem lymph nodes affected by pneumonia or COVID-19 reveals distinct immune microenvironments and T cell clonality patterns, highlighting broad utility across diverse clinical contexts.

Keywords

TCR sequencing, single-cell immune profiling, adaptive immunity, spatial transcriptomics, tumor microenvironment

Background

Antigen-specific T cells are key mediators of the adaptive immune response and play a central role in autoimmune disorders, cancer, and infectious disease. T cell antigen specificity and clonal identity are determined by the T cell receptor (TCR), which is composed of either an α/β or a γ/δ heterodimer, with α/β being the most common form of TCR found on CD4+ helper and CD8+ cytotoxic T cells. The highly diverse repertoire of reactive T cells is generated by somatic recombination of the variable (V), diversity (D), and joining (J) TCR gene segments, which together define the sequence and antigen binding specificity of the Complementarity-Determining Region 3 (CDR3) that directly contacts the antigen presented by the Major Histocompatibility Complex (MHC) [1]. In humans, the TCRA locus includes approximately 50 functional Vα and 61 Jα gene segments, while the TCRB locus contains 52 Vβ, 2 Dβ, and 13 Jβ gene segments. During recombination, the CDR3 sequence is further diversified by random nucleotide insertions or deletions at the junctions between gene segments, leading to an estimated diversity of 10^{15} to 10^{20} possible TCR α/β combinations [2,3].

Sequencing based TCR profiling methods can determine individual T-cell repertoires at single-cell resolution and enable the monitoring of T cell clonotype dynamics in various disease settings, and advances in single-cell multi-omics analysis have enabled a more comprehensive understanding of T cell biology. Particularly, paired TCR sequencing with mRNA expression within the same cell has established a direct correlation between TCR repertoire and cellular phenotypes, illuminating T cell development and function [4–6]. Commonly used approaches for single-cell TCR profiling employ 5'-barcoded scRNA-seq strategies, as 3'-directed scRNA-seq approaches fail to recover the VDJ information located in the first 500 nucleotides of TCR transcripts [5]. At the same time, advanced multi-omics approaches such as paired RNA and ATAC sequencing, or spatial transcriptomics workflows such as Visium (10X Genomics) or Slide-Seq [7], are based on 3'-barcoding of transcripts and therefore lose VDJ information during fragmentation-based library preparation. Nevertheless, information on clonotypic transcriptional regulation and lymphocyte clonality within human tissues remains highly valuable as it provides important insights into adaptive immune response mechanisms and enables the identification and characterization of antigen-specific T cell clones, which can inform the development of targeted therapeutics [8].

Several recent studies have introduced modified workflows to recover VDJ sequence information from 3'-barcoded single-cell and spatial transcriptomics workflows [8–12]. However, these methods are resource-intensive as they rely on the amplification of TCRα and TCRβ sequences by multiplexed PCR with hundreds of primers that target all possible V segments, sometimes

combined with prior hybridization capture of TCRα and TCRβ cDNA molecules, and often require specialized equipment for long-read sequencing. Here, we introduce circVDJ-seq, a simplified workflow for paired full-length TCR sequencing from 3'-barcoded cDNA that massively reduces the number of required gene-specific primers and obviates the need for specialized sequencing equipment. We show that circVDJ-seq efficiently recovers T cell clonotypes from 3'-barcoded cDNA libraries and is compatible with the RNA+ATAC Multiome and Visium assays (10X Genomics). The identified TCR sequences and clonotype frequencies are highly concordant with results obtained from the commonly used 5'-directed Immuneprofiling v2 workflow (5'IPv2, 10X Genomics). We demonstrate that circVDJ-seq can be used to spatially resolve T cell clonotypes even in challenging tissue samples such as autopsy-derived lymph nodes from non-COVID-19 pneumonia and COVID-19 patients. We finally reveal contrasting tumor immune microenvironments in two freshly resected high- and low-risk neuroblastoma patient samples (HR-NB / LR-NB), with exclusion of a suppressive immune environment from the tumor compartment in HR-NB, and prominent myeloid infiltration in LR-NB, highlighting the broad applicability of circVDJ-seq in diverse clinical settings.

Results

circVDJ-seq simplifies TCR profiling from 3'-barcoded cDNA libraries

Here we present circVDJ-seq, a simple and cost-efficient workflow for TCR profiling from 3'-barcoded single-cell cDNA libraries that is independent of large oligonucleotide panels for PCR or hybridization capture, and does not require prior knowledge of all possible V gene sequences, or specialized sequencing equipment. cDNA is first tagged with homologous overhangs to facilitate circularization via Gibson assembly [18] and generate a cDNA library in which the VDJ region is juxtaposed to the unique molecular identifier (UMI) and cell barcode at the 3' end of cDNA molecules (**Fig. 1A**). A VDJ library is then amplified via nested PCR using primers against the TCRa and TCRb constant and 3'UTR regions (**Fig. 1A, B; Supplementary Table 1**). Subsequent library preparation steps, including fragmentation, end repair, A-tailing, adapter ligation, and Index PCR, follow the 5'IPv2/3 TCR library preparation workflow and can be performed with the commercial assay kit and adjusted index primers (**Fig. 1C; Supplementary Table 1**).

To validate the approach, we created triplicate circVDJ-seq libraries from archived cDNA that had previously been generated with the 3'GEX v3.1 assay (10X Genomics) from PBMCs of a healthy donor (**Fig. 1B, C**). The circVDJ read structure mirrors the outputs obtained from the 5'IPv2/3 assay. We were therefore able to readily process circVDJ-seq data with the VDJ CellRanger pipeline after replacing the cell barcode (CBC) whitelist with barcodes from the 3'GEX v3.1 assay.

The resulting output data was further processed with Dandelion for improved VDJ contig annotation [17]. On average, 93% of circVDJ sequence reads contained a valid CBC, and 83% of reads mapped to a VDJ gene (**Supplementary Table 2**). 73% and 84% of cells contained a TCRa or TCRb contig, respectively, yielding an average of 57% of cells with paired clonotype information. CBCs, CDR3 sequences, and CBC+CDR3 combinations strongly overlapped between circVDJ replicates (**Fig. 1D**). While circVDJ-seq data can readily be processed with other tools, combining CellRanger and Dandelion resulted in better reproducibility of technical replicates than processing the data with CellRanger alone, MiXCR [19], or TRUST4 [20] (**Supplementary Fig. 1**). Importantly, clonotypes determined by circVDJ replicates were readily recovered in relevant T cell clusters when projecting the 3' gene expression data onto a PBMC reference [21] (**Fig. 1E**), and clonotype frequencies were highly consistent across technical replicates (**Fig. 1F**).

circVDJ-seq reliably detects TCR clonotypes from single-cells and nuclei in multi-omics assays

Linking TCR sequence and gene expression states with information about accessible chromatin regions in the same cells could provide crucial insights into the epigenetic regulation of T cell development and function. We therefore processed PBMCs from the same donor with the ATAC+RNA Multiome workflow (MO) to test to what extent clonotype information could be recovered from the isolated cell nuclei required for that assay. MO circVDJ-seq sequence data were processed with CellRanger updated for the MO CBC whitelist, followed by Dandelion [17]. Reads in the MO-derived libraries contained about 85% valid CBCs, about 74% could be mapped to a VDJ gene, (**Supplementary Table 2**), and resulting CBC+CDR3 combinations showed very high overlap in triplicate circVDJ-seq libraries (**Supplementary Fig. 2**). To further benchmark the results obtained from 3'GEX v3.1 or the MO assay against an established 5'-directed workflow for single-cell TCR profiling, we next used the 5'IPv2 assay to generate a reference dataset using PBMCs from the same donor. Upon direct comparison, 5'IPv2 and circVDJ-seq from 3'GEX efficiently recovered T cell clonotype information, with circVDJ-seq nearly matching overall clonotype recovery of the 5'IPv2 kit (87% vs 97%) (**Fig. 2A**). At the same time, the number of T cells with assigned clonotypes was reduced to 36% in the MO assay where we also obtained fewer cells with paired clonotype information due to a lower recovery of TCRa contigs, most likely caused by reduced transcript abundance in single nuclei compared to whole cells.

Earlier studies of T cell receptor repertoires have shown that the same TCRb chain can be paired with different TCRa chains in the same individual [22,23], and that two TCRa chains can be co-expressed in the same cell [24,25]. In our own data, we found only rare instances of such

promiscuous pairing (below 5%; **Supplementary Fig. 3**). For fair comparison of clonotype recovery between different assays, we therefore decided to impute missing TCRa or TCRb chains from the most frequent pairing observed in our data. Although distinct populations of cells from the same donor were processed in each experiment, 4% of all clonotypes identified by 5'IPv2 were also found by circVDJ-seq from 3'-directed scRNA-seq after merging technical replicates, and 1.5% were found by circVDJ-seq from the MO assay (**Fig. 2B**). However, these recurring clonotypes represented 34% and 26% of cells, as 19 and 15 of the top 20 most abundant clones detected in the 5'IPv2 workflow were also recovered by circVDJ-seq from 3'GEX or MO cDNA, respectively. Consequently, the relative abundance of individual clones was highly correlated between circVDJ and 5'IPv2 (R=0.93, and R=0.78, respectively; **Fig. 2C, D; Supplementary Fig. 4**), and the associated gene expression data reproducibly identified expanded clones as CD8 effector memory T cells. Interestingly, clonotype detection by circVDJ-seq showed similar efficiency in two unrelated PBMC samples that underwent a mild PFA fixation prior to MO processing as used in DOGMA-seq [26], highlighting the potential for parallel high-throughput TCR profiling alongside four separate genomic modalities at the single-cell level (**Supplementary Fig. 5**).

Slight variations in individual clonotype abundances observed between MO circVDJ-seq and 5'IPv2 measurements from the same donor could be caused to varying degrees by fluctuations during multiple sub-sampling of cells, differences in the source material (whole cells vs nuclei), or by inaccuracy of the circVDJ-seq workflow itself. To exclude the latter possibility, we next validated the results from MO circVDJ-seq using an alternative approach. To this end, we directly subjected MO cDNA to long read sequencing using MAS-ISO-seq without prior enrichment or amplification of VDJ sequences [27]. The overall number of detected clonotypes was reduced in MAS-ISO-seq compared to circVDJ-seq, which was expected since circVDJ-seq is a targeted approach and PacBio sequencing has limited sequencing depth (**Fig. 2E**). Importantly, however, clonotype abundances were highly correlated (R=0.84, $p<2.2e^{-16}$) between both methods (**Fig. 2F, Supplementary Fig. 4, Supplementary Fig. 6**). Together, these data indicate that circVDJ-seq can faithfully recover TCR sequence information from 3'-barcoded cDNA libraries generated from whole cells, or from workflows that use single nuclei as starting material, such as the MO assay.

Spatial circVDJ-seq detects T cell clonal expansion in autopsy-derived tissue

We next explored the utility of circVDJ-seq for spatial T cell clonotype mapping in autopsy-derived tissue samples. We recently identified a CCL21 to CCR7 signaling axis linked to the accumulation of exhausted T cells in ectopic lymphoid structures during prolonged lung immunopathology in

COVID-19 by applying Visium gene expression profiling to lung and lymph node samples from deceased patients [13]. We generated triplicate circVDJ-seq libraries from Visium cDNA of a chronic COVID-19 post mortem lung sample from the same cohort that showed prominent immune cell infiltration and fibrotic areas with exacerbated collagen expression. We processed circVDJ-seq sequence reads with CellRanger VDJ after exchanging the CBC whitelist with Visium spot barcodes. Interestingly, although overall mapping statistics appeared similar to data from single-cells or nuclei (**Supplementary Table 2**), CellRanger VDJ did not report any spots with a successfully reconstructed TCR chain. This was potentially caused by the lower number of spot barcodes compared to single-cell barcodes, which might interfere with effective UMI thresholding. However, further processing of the raw CellRanger output with Dandelion [17] efficiently reconstructed TCRa and TCRb chains together with their spatial addresses **(Fig. 3A, B)**.

The identified T cell clonotypes were highly concordant between technical replicates (**Fig. 3C**), and the proportion of spots with detected TCR clones per cluster correlated with TCR UMI counts in the corresponding Visium GEX data (**Fig. 3D**). This analysis revealed the highest T cell abundance in plasma cell and alveolar spots, and in close proximity to stroma cluster 7 and cluster 8 spots (**Fig. 3D, E**), as calculated by relative co-occurrence probability between spots with TCR and spots of other clusters within a certain radius [28]. circVDJ-seq further revealed substantial expansion of a single clonotype with TCRb CDR3 amino acid sequence CASSHENQPQHF (**Fig. 3F**). We previously observed T cell activation markers in COVID-19 lung draining lymph nodes (dLNs) with no signs of active viral infection, suggesting that T cell activation persists in dLNs after the infection has been resolved [13]. We therefore expanded our circVDJ-seq analysis to archived Visium cDNA from an acute, chronic, and prolonged COVID-19, and a non-COVID-19 pneumonia dLN from the same cohort [13], to determine the respective extent of clonal T cell expansion (**Fig. 4A, B**). Cluster-wise clonotype detection again strongly correlated with TCR gene expression UMIs in all four samples (**Fig. 4C**), and identified TCR clones were most abundant in the T cell zone surrounding B cell follicles (**Fig. 4B, C**). At the same time, we observed very low levels of clonal expansion in COVID-19 LNs (**Fig. 4D**). In contrast, the non-COVID-19 control sample showed very pronounced clonal T cell expansion (**Fig. 4D**), potentially caused by non-COVID-19 pneumonia and a possible squamous cell carcinoma lymph node metastasis of this donor, as suggested by concomitant keratin expression. Together, these data indicated that circVDJ-seq can efficiently retrieve contrasting clonotype dynamics even from challenging samples such as autopsy derived tissue.

Spatial circVDJ-seq reveals differential T-Cell infiltration in high- and low-risk neuroblastoma

To further explore its clinical utility, we next applied circVDJ-seq to freshly resected neuroblastoma samples. Neuroblastomas are neural crest cell-derived pediatric tumors and represent the third most common type of cancer in children. NB cases can be categorized into high-, middle, and low risk disease based on patient age, histology, and genetic alterations like DNA ploidy and N-myc amplification [29]. Stage IVs neuroblastomas constitute a special group of low-risk tumors that are disseminated but often spontaneously regress. While patients with low-risk disease have a very good prognosis, the 5-year survival rate of high-risk patients remains at ~50%. To test if circVDJ-seq can reveal differences in the immune tumor microenvironment between distinct NB risk groups, we first generated Visium libraries from a stage IV high-risk (HR) and a stage IVs low-risk (LR) neuroblastoma, both without MYCN amplification.

Joint analysis of spatial gene expression revealed four tumor cell clusters and five stroma cell clusters with differential enrichment in the LR- and HR-NB sample (**Fig. 5A-C**). LR-NB tumor clusters 0, 1 and 3 showed high expression of the known low-risk markers PRPH and NRCAM [30,31] and several genes associated with favorable prognosis were additionally highly expressed in LR-NB cluster 3, including CCNL2, WSB1, PCBP4, and HAND2-AS1 [32–37]. In contrast, the predominant HR-NB tumor cluster 2 showed downregulation of neuronal differentiation markers, low levels of PLXNA4 and high expression of sympathetic neuropeptide NPY and NEAT1, which have all been linked to poor prognosis [38–41].

Besides differences in tumor cell states, both samples also displayed substantial differences in TME composition. LR-NB showed higher abundance of cluster 4 with high expression of CD74, APOE and APOC1, reflecting immune cell infiltration [42,43]. At the same time, immune cluster 4 spots in HR-NB showed higher expression of CD163, TGFB, CSF1R and VEGFB, indicating the presence of immunosuppressive tumor-associated M2 macrophages (**Fig. 5D**). Stromal cluster 5 was found in both LR- and HR-NB and contained markers of tumor associated fibroblasts (CAFs) SPARC and IGFBP7 [44] [45–47], but with higher expression of the cancer-associated fibroblast-secreted extracellular matrix (ECM) glycoprotein IGFBP7 in HR-NB (**Fig. 5D**). Stromal clusters 6 and 8 were strongly enriched in the HR-NB sample. Myeloid makers and higher levels of mitochondrial genes in cluster 6 might point towards metabolic stress of intratumoral myeloid cells [48], while cluster 8 showed signs of stress response and inflammation in a reactive stroma and high levels of MT2A indicative of CAFs linked to disease progression and poor prognosis [49,50]. Taken together, this analysis revealed substantial differences in tumor cell state and

microenvironment, with a less aggressive cancer state and enhanced immune infiltration in LR-NB, and signs of immunosuppression, inflammatory stress and CAFs in the HR-NB patient sample.

We next generated duplicate circVDJ-seq libraries from each sample and processed the sequenced reads with CellRanger VDJ and Dandelion. This analysis revealed vastly different levels of T cell clonotypes in the HR- and LR-NB samples with 116 and 16 identified T cell clones, respectively (Fig. 5A; F). The number of spots that contain T cell clones in HR-NB strongly correlated with TCRab UMI counts in individual gene expression clusters (**Fig. 5E**), showing the highest enrichment in blood vessels and immune cluster 4. circVDJ-seq revealed substantial differences in cell type composition of immune cluster 4 between the LR-NB and HR-NB patient sample, with strongly elevated T cell counts and clonal amplification in HR-NB compared to low T cell counts in LR-NB (**Fig. 5E, F**). In the HR-NB sample, cluster 4 spots mostly co-occurred with blood vessels, other immune cluster 4 spots, and stromal spots, but were mostly excluded from the vicinity of the predominant HR-NB tumor cluster 2 (**Fig. 5G, top left panel**). The same exclusion from the HR-NB tumor area was also observed when the analysis was centered on all spots that include a T cell clone (**5G, top right panel**). Conversely, in the LR-NB sample immune cluster 4 spots were not excluded from the LR-enriched tumor clusters 0 and 1, but appeared more distant from LR-NB tumor cluster 3. However, centering the analysis on the few spots that contain a T cell clone showed no spatial exclusion from the tumor area in general. Taken together, these analyses indicated substantial differences in the immune microenvironments of the HR- and LR-NB sample, with a more pronounced myeloid immune infiltration in LR-NB, and a higher abundance of amplifying T cells in HR-NB, which remained largely excluded from the tumor area and were accompanied by immunosuppressive M2 macrophages and signs of a stressed reactive stroma.

Discussion

TCR profiling provides essential insights into adaptive immunity, informing diagnosis, prognosis, and therapeutic approaches in cancer, infection, and autoimmune diseases. Here, we introduced circVDJ-seq, a streamlined and cost-efficient method for comprehensive TCR clonotype analysis from widely used 3'-barcoded transcriptomics assays. Unlike existing methods requiring complex primer sets or costly sequencing technologies, circVDJ-seq simplifies clonotype retrieval through cDNA circularization and amplification of conserved TCR constant regions. This reduces reagent and sequencing costs, but more importantly also eliminates inherent biases of the multiplexed PCR reaction that may lead to uneven amplification of different TCR clones [51]. In addition,

circVDJ-seq can be applied to species with inaccurate or incomplete genome annotations of the TCR variable genes.

We show that circVDJ-seq accurately recovers TCR sequence information encoded in the cDNA libraries that serve as input material, evidenced by highly correlating results from technical circVDJ-seq replicates, by parallel long-read sequencing of the original cDNA template, and by 5'-directed TCR profiling of PBMCs from the same donor. While circVDJ-seq recovers clonotypes efficiently, sensitivity depends heavily on the quality of the initial cDNA libraries, emphasizing the importance of optimal sample preparation protocols. Besides timely acquisition and processing of biological samples, the underlying cDNA amplification method needs to amplify a full-length cDNA molecule, for instance via a template switch mechanism, with workflows that use random hexamers for second strand priming leading to shortened cDNA molecules that lack complete VDJ information.

While a 5'-directed strategy will provide the highest sensitivity for paired clonotype identification due to fewer library preparation steps, circVDJ-seq can efficiently reconstruct the clonotype composition of archived cDNA generated with 3'-directed scRNA-seq protocols, or from inherently 3'-directed assays such as the ATAC+RNA MO workflow. Our results from MO circVDJ-seq also show that TCR clonotype information can be retrieved from single nuclei. This enables the identification of the most abundant TCR clones, albeit with lower proportion of a chains and a reduced overall number of recovered clonotypes owing to the lower number of captured mRNA molecules in nuclei compared to whole cells. Nevertheless, single nucleus RNAseq has been the method of choice for a large range of biological samples that are collected via snap freezing, e.g. during surgery, with circVDJ-seq enabling the retrospective analysis of clonal T cell amplification in such samples.

Similarly, circVDJ-seq can be combined with any spatial transcriptomics method that is based on polyA-capture of mRNAs and template switching, such as Visium, Slide-seq [52,53] or Stereo-seq [54]. Using the Visium workflow for initial cDNA library generation we demonstrated that circVDJ-seq can provide important insights into T cell biology. Our spatial circVDJ-seq analysis identified distinct immune landscapes in neuroblastoma subtypes, with notable exclusion of T cells from tumor compartments in high-risk disease, suggesting the presence of immunosuppressive microenvironments previously documented in aggressive pediatric cancers. This observation is also consistent with recent spatial transcriptomics studies on high-risk neuroblastoma that have shown spatial compartmentalization with malignant tumor cells often spatially segregated from immune cells [48].

We furthermore observed enhanced clonal amplification in lung draining lymph nodes from cancer-associated pneumonia compared to prolonged COVID-19, illustrating the utility of circVDJ-seq as a versatile tool for TCR profiling in various tissue contexts including post-mortem autopsy material, which was especially encouraging. Here, the integrity of the starting material is of particular importance, since degradation of mRNA molecules will make TCR information inaccessible. Timely tissue processing and the testing of RNA integrity are therefore advisable in this context. The examples shown here are based on standard Visium with 50µm spot diameter and corresponding to multiple cells per spot. However, newer workflows such as STOMICS or Visium HD 3' have reached true single-cell resolution combined with enhanced mRNA capture efficiency, and circVDJ-seq can be readily applied to those workflows.

Conclusions

circVDJ-seq is a simple and robust assay for TCR profiling from 3'barcoded cDNA libraries. It is broadly applicable to single-cell multi omics workflows as well as spatial transcriptomics and does not require large pools of PCR primers, capture oligos, or specialized long-read sequencing equipment, enabling easier access to single-cell and spatial TCR repertoire profiling. We envision that circVDJ-seq will facilitate widespread and cost-effective TCR profiling across diverse single-cell and spatial transcriptomics platforms and clinical contexts.

Methods

Patient recruitment and sample collection

The PBMCs, lung and lung draining lymph node (dLN) samples included in this study were collected in the Department of Neuropathology of Charité - Universitätsmedizin Berlin as part of the COVID-19 autopsy Biobank. Donor identities were encoded at the hospital before sharing for sample processing or data analysis. All COVID-19 donors are part of the previously published cohort in [13]. All COVID-19 donors tested positive for COVID-19 in oropharyngeal swabs at the time of hospital admission. The Neuroblastoma samples were collected during surgical resections at the University Hospital Cologne. All relevant characteristics and clinical information of the donors are presented in Supplementary Table 2.

Generation of 3'GEX single-cell RNA-seq libraries from human PBMCs

Single cell capturing and library construction were performed with the Chromium Next GEM Single-cell 3' Reagent Kit v3.1 (10X Genomics) according to the manufacturer's instructions. Briefly, a droplet emulsion targeting 10,000 cells was generated in a microfluidic Next GEM Chip

G, with each droplet containing a single cell and 10X chemistry for cell lysis, 3'-barcoding, and reverse transcription of released mRNA. The purified cDNA was then amplified using PCR. 1/4th of the cDNA library was subjected to standard library preparation and sequencing on a NovaSeq 6000 instrument (Illumina, USA).

Generation of 5'GEX single-cell RNA-seq libraries from human PBMCs

Single cell capturing and library construction were performed with the Chromium Next GEM Single-cell 5' Kit v2 (10X Genomics) according to the manufacturer's instructions. In short, a droplet emulsion targeting 10,000 cells was generated in a microfluidic Next GEM Chip K, with each droplet containing a single cell and 10X chemistry for cell lysis, 5'-barcoding and reverse transcription of released mRNA. The purified cDNA was then amplified using PCR. 1/4th of the cDNA library was subjected to GEX and TCR library preparation, respectively, followed by sequencing on a NovaSeq 6000 instrument (Illumina, USA).

Generation of single-cell Multiome ATAC+RNA libraries from human PBMCs

Single nuclei were isolated from PBMCs as described in the demonstrated protocol CG000365 Rev C (Nuclei Isolation for Single-cell Multiome ATAC + Gene Expression Sequencing, 10X Genomics). Isolated nuclei were then processed with the Chromium Single-cell Multiome ATAC + Gene Expression v1 assay (10X Genomics) according to the manufacturer's instructions. In short, isolated nuclei were transposed in bulk solution, followed by partitioning into a droplet emulsion using a microfluidic chip, with each droplet containing a single cell and 10X chemistry for nuclear lysis, 3'-barcoding, and reverse transcription of released mRNA and indexing of transposed DNA. Both ATAC and gene expression (GEX) libraries were then generated from the same pool of pre-amplified transposed DNA/cDNA and subsequently sequenced on a NovaSeq 6000 instrument (Illumina, USA).

Fixed whole-cell 10× Genomics Multiome experiments were performed as previously described for DOGMA-seq [14]. Briefly, cells were fixed in 0.1 % formaldehyde for 5 min at room temperature, quenched with 0.125M glycine, and permeabilized for 3 min on ice in lysis buffer (10 mM Tris-HCl pH 7.4, 10 mM NaCl, 3 mM $MgCl_2$, 0.1 % NP-40, 1 % BSA, 1 mM DTT, 1 U μL^{-1} RNase inhibitor). Cells were then washed three times in 1 mL of wash buffer (same composition as lysis buffer but without NP-40). After centrifugation for 5 min at 500 × g, the supernatant was removed, and the cells were resuspended in 1× Diluted Nuclei Buffer (10× Genomics). Permeabilized cells were processed using the Chromium Next GEM Single-cell Multiome ATAC + Gene Expression workflow (10× Genomics, CG000338-Rev E) with the following modification. After SPRI cleanup

of the pre-amplification PCR product, beads were eluted in 100 µL EB buffer instead of 140 µL. From this eluate, 25 µL were used for ATAC-seq library construction, and 35 µL were used for cDNA amplification.

Spatial transcriptomics

Visualization of gene expression in lung tissue was performed using 10× Visium spatial gene expression kit (10× Genomics) following the manufacturer's protocol. Briefly, control and COVID-19 lung samples from donors categorized based on disease durations were cut into 10 µm sections using an MH560 cryotome (ThermoFisher, Waltham, Massachusetts, USA), and mounted on 10X Visium slides, which were pre-cooled to −20 °C. The sections were fixed in pre-chilled methanol for 30 min, stained with CD45-AF647, CD31-AF594 and DAPI for 30 min and imaged using an LSM 880 confocal microscope (Zeiss). The sections were then permeabilized for 10 min and the captured mRNA was reverse transcribed and spatially 3'-barcoded followed by cDNA library amplification. Short-read sequencing libraries were constructed from 1/4th of the cDNA using the 10x Genomics Visium Spatial Gene Expression 3' Library Construction V1 Kit. Libraries were sequenced on a NovaSeq 6000 instrument (Illumina, USA). Frames around the capture area on the Visium slide were aligned manually and spots covering the tissue were manually selected based on the immunofluorescence staining, using Loupe Browser 5.1.0 software (10× Genomics). Sequencing data was mapped to GRCh38-2020-A reference transcriptome using the Space Ranger software (version 1.3.0, 10× genomics) to derive a feature spot-barcode expression matrix.

Generation of cirVDJseq libraries from 3'-barcoded cDNA

Homologous ends for Gibson assembly (TCRGOT_1 and TCRGOT_12; **Supplementary Table S1**) were added to 10-15ng of 3'-barcoded cDNA through PCR using the KAPA HiFi HotStart ReadyMix PCR Kit with the following temperature cycle: 95°C for 3 min, five cycles of 98°C for 20 s, 65°C for 30 s, 72°C for 2 min, and final extension at 72°C for 5 min. Amplified cDNA was purified using 0.8X AMPure XP beads and eluted in 20 µL EB. Circularization was carried out at 50°C for 1 h with NEBuilder® HiFi DNA Assembly Master Mix in a 200 µL reaction (1x CutSmart Buffer) followed by 0.9X AMPure XP bead clean up and elution in 20 µL EB. Remaining linear DNA was removed by incubation with 6U Lambda exonuclease at 37°C for 30 min, followed by inactivation at 75°C for 10 min. Circularized cDNA was purified using 0.8X AMPure XP beads and eluted in 20µl EB. Next, nested PCR was performed with primers targeting the 3' UTR and constant regions of TCR α and β chains (**Supplementary Table S1**). PCR conditions for the first amplification (12-15 cycles) were: 95°C for 3 min, 98°C for 20 s, 62°C for 30 s, and 72°C for 1 min, followed by a

final extension at 72°C for 1 min. Amplified products were purified using 0.8X AMPure XP beads. A second amplification (10 cycles) used the same conditions, followed by purification and concentration with 0.9X AMPure XP beads. Successful VDJ amplification was assessed using the Agilent 4200 TapeStation High Sensitivity DNA assay and quantified by Qubit dsDNA HS assay. 2-3ng of amplified VDJ cDNA was used to generate final circVDJ-seq libraries using the library construction kit from 10X Genomics (PN1000190). Briefly, VDJ cDNA was subjected to enzymatic fragmentation, end-repair, and A-tailing at 32°C for 2 minutes followed by heat inactivation at 65°C for 30 minutes. Subsequently, adapter ligation was carried out by incubating the sample with ligation buffer, adapter oligos, and DNA ligase at 20°C for 15 minutes. The ligated products were purified using a 0.8X SPRIselect bead cleanup. Indexed amplification was performed using indexing primers shown in **Supplementary Table 1**, with cycling conditions of 98°C for 45 seconds; 8 cycles of 98°C for 20 seconds, 54°C for 30 seconds, 72°C for 20 seconds; followed by a final extension at 72°C for 1 minute. Post-PCR products were cleaned with 0.9X SPRIselect beads and eluted in elution buffer. Final libraries were assessed using the Agilent 4200 TapeStation High Sensitivity DNA assay and quantified by Qubit dsDNA HS assay. In addition, libraries were quantified using qPCR and sequenced with a NextSeq 550 System Mid-Output Kit with read configuration 90–28–10-10 using custom sequencing primers indicated in **Supplementary Table 1**.

Primer Design for Multiplex PCR

Primers targeting the 3’ UTR and constant regions of TCR α and β chains were designed to facilitate the detection of rearranged TCR genes post-cDNA circularization. Sequences for human TRAC, TRBC1, and TRBC2 genes were obtained from the IMGT/GENE-DB database. Primer3 was used for primer selection.

Single-cell and spatial transcriptomics data analysis

Due to the involvement of multiple collaborating groups, single-cell, Multiome or Visium gene expression and/or chromatin accessibility data were processed with different versions of CellRanger or SpaceRanger against the GRCh38 reference genome, as summarized in the following table:

material	sample(s) / method	program (version)	command used
PBMC	5’IP v2	CellRanger v6.1.2	CellRanger multi
	3’	CellRanger v6.1.1	CellRanger multi

	MO	CellRanger-arc v2.0.0	CellRanger-arc count
	MO + PFA	CellRanger-arc v2.0.2	CellRanger-arc count
neuroblastoma	HR + LR	SpaceRanger v1.3.0	SpaceRanger count
lung tissue	chronic COVID-19	SpaceRanger v1.3.1	SpaceRanger count
lymph node	control/acute/chronic/prolonged	SpaceRanger v1.3.1	SpaceRanger count

PBMC gene expression data was analyzed in R (v4.3.2) using Seurat (v4.4.0). We used all cells with less than 10% mitochondrial RNA and with at least 250 and at most 8000 genes. We used sctransform (v0.3.5) and the MapQuery workflow to transfer cell type annotations from a PBMC reference [15]. We removed cells with a layer 1 cell type prediction score below 0.75, predicted monocytes, and doublets predicted by DoubletFinder [16].

Visium data for neuroblastoma, lung tissue, and lymph node samples were loaded into R, normalized with sctransform, and subjected to PCA, neighbor graph construction, clustering and UMAP reduction. Neuroblastoma datasets were merged before PCA.

circVDJ-seq data analysis

Technical replicates of PBMC data generated using different assays were processed individually but also jointly after pooling reads from replicates for better comparison of assays or processing tools.

CellRanger processing

We modified CellRanger v6.0.0 to allow for cell barcodes from other assays than 5'IP by exchanging the file CellRanger-6.0.0/lib/python/CellRanger/barcodes/737K-august-2016.txt for corresponding files in the same directory, namely 3M-february-2018.txt (single-nuc or 3' assay), 737K-arc-v1.txt (multi-ome assay) or visium-v1.txt (Visium V1). We further switched fastq files for read mates 1 and 2 and then ran CellRanger vdj using the refdata-CellRanger-vdj-GRCh38-alts-ensembl-5.0.0 reference provided by 10X.

MiXCR processing

We used MiXCR v4.1.2 with the following command line mixcr -Xmx64g analyze 10x-vdj-tcr --species hsa /path/to/read_2.fastq.gz /path/to/read_1.fastq.gz output --tag-pattern "^(CELL:N{16})(UMI:N{12})\^(R2:*)" --threads 8 -M refineTagsAndSort.whitelists.CELL="file:3M-february-2018.txt" for the 3' assay and the corresponding barcode whitelist files for the other assays, and 10nt UMIs for 5'IPv2.

TRUST processing

We used TRUST4 v1.1.5 with the following command line:

TRUST4/run-trust4 -f /path/to/hg38_bcrtcr.fa --ref /path/to/human_IMGT+C.fa -u /path/to/read_2.fastq.gz --barcode /path/to/read_1.fastq.gz --barcodeRange 0 15 + --barcodeWhitelist 3M-february-2018.txt --UMI /path/to/read_1.fastq.gz --umiRange 16 25 + --od output -t 8 –repseq

For the MAS-ISO-seq data, we used the tagged refined sorted and de-duplicated bam file from the extracted mRNA sequence as read mate 2, and cell barcode + UMI as read mate 1, and used TRUST4 as follows:

TRUST4/run-trust4 -f /path/to/hg38_bcrtcr.fa --ref /path/to/human_IMGT+C.fa -1 /path/to/read1.fastq.gz -2 /path/to/read2.fastq.gz --barcode /path/to/read1.fastq.gz --UMI /path/to/read1.fastq.gz --readFormat bc:0:15,um:16:27 -o output -t 8

Cell barcodes returned by TRUST for MAS-ISO-seq data were reverse-complemented.

Dandelion processing

We used Dandelion v0.3.7 [17] and followed the re-annotation vignette, i.e. we used ddl.pp.format_fastas(...) and then ddl.pp.reannotate_genes(.., loci="tr", reassign_dj=True, org='human', extended=True, flavour='strict', min_j_match=7, min_d_match=9, v_evalue=0.0001, d_evalue=0.001, j_evalue=0.0001, dust='no', db='imgt')

CircVDJ-seq data post-processing and filtering

We implemented an additional filtering strategy to identify and remove spurious contigs with identical cell barcode and CDR3 sequences that originated from different samples sequenced on the same flowcell and likely resulted from index hopping, as demultiplexing of the first set of libraries had to performed in single-index mode due to failure of the initial custom index 2 sequencing primer. For this, we assigned a clone as primary (defined via cell barcode and junction

aa sequence) if it either contained more than 1% of the total reads (for replicates derived from the same cDNA sequenced on the same flowcell), or if it originated from the sample with the highest read count for that clone on that flowcell. We meanwhile updated the index 2 primer design and validated sequencing of circVDJ-seq libraries in dual index mode, so that filtering for index hopping can be omitted in future analyses.

To further account for spurious signals likely created during PCR amplification, we extracted UMI sequences for all reads assigned to a clone by CellRanger, grouped UMIs within Hamming distance 1, and counted the number of perfectly matching reads (using the CIGAR string) as well as the fraction of reads aligning to the 3' end of the contig. We then assigned revised UMI and read counts to a clone by counting only UMIs containing at least one perfectly matching read but with less than half of reads aligning to the 3' end.

Data analysis and statistics

Filtered dandelion output was combined with gene expression data by matching cell barcodes, keeping only productive and primary contigs, and filtering out secondary α or β chains for single-cell (but not Visium) data.

For imputation of missing α or β chains, we collected the most frequently observed combination across all our PBMC data generated using different assays and replaced a missing chain with the partner from that combination.

For the co-occurrence analysis, we re-created the co-occurrence function from squidpy in R. Briefly, we created a distance matrix from the tissue coordinates of all spots and then assessed the co-occurrence probability as the ratio of the probability to observe a spot of a certain type within a radius around spots of a given index type by the probability of observing spots of a certain type overall. This calculation was repeated multiple times for random subsets of 50% of the respective index spots to obtain error bars from standard deviation.

Abbreviations

ATAC: Assay for Transposase-Accessible Chromatin

BSA: Bovine Serum Albumin

CAF: Cancer-Associated Fibroblast

CBC: Cell Barcode

cDNA: Complementary DNA

CDR3: Complementarity-Determining Region 3

circVDJ-seq: Circularized VDJ sequencing

COVID-19: Coronavirus Disease 2019

DTT: Dithiothreitol

dLN: draining Lymph Node

EB: Elution Buffer

ECM: Extracellular Matrix

GEX: Gene Expression

HD: High Definition (in context of spatial transcriptomics)

HR-NB: High-Risk Neuroblastoma

IMGT: International Immunogenetics Information System

5'IP v2/3: Immune Profiling Version 2/3

LR-NB: Low-Risk Neuroblastoma

MAS-ISO-seq: Multiplexed Amplification of Specific Isoforms Sequencing

MHC: Major Histocompatibility Complex

MO: Multiome (combined RNA and ATAC sequencing)

PBMC: Peripheral Blood Mononuclear Cell

PCA: Principal Component Analysis

PCR: Polymerase Chain Reaction

PFA: Paraformaldehyde

qPCR: Quantitative PCR

RNA: Ribonucleic Acid

RT-PCR: Reverse-Transcription PCR

scRNA-seq: Single-cell RNA Sequencing

SPRI: Solid-Phase Reversible Immobilization

STOMICS: Spatial Transcriptomics Omics

TCR: T Cell Receptor

TRAC: T Cell Receptor Alpha Constant Region

TRBC: T Cell Receptor Beta Constant Region

UMI: Unique Molecular Identifier

UTR: Untranslated Region

VDJ: Variable (V), Diversity (D), Joining (J)

UMAP: Uniform Manifold Approximation and Projection

Declarations

Ethics approval and consent to participate

This study was conducted in accordance with the Declaration of Helsinki and with the approval of the Ethics Committee of the Charité (EA2/066/20, EA1/317/20, EA2/144/15, and EA1/005/21), the Charité - BIH COVID-19 research board, and the Ethics committee of the Medical Faculty of the University of Cologne (#9764 and #04-049). Autopsies were performed on the legal basis of §1 926 SRegG BE of the Autopsy Act of Berlin and §25(4) of the German Infection Protection Act. Autopsy consent was obtained from the families of the patients.

Consent for publication

Not applicable.

Availability of data and materials

The data generated / analyzed here will be made available via a public repository.

Competing interests

The authors declare no competing interests.

Funding

This study was in part supported by grants from BIH and the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) within CRC1588, project number 493872418.

Authors' contributions

I.P. and T.C. conceived the circVDJseq workflow. I.P. performed circVDJseq experiments. B.O. developed circVDJseq analysis pipelines. Y.H., C.B., M.G., M.SZ. and C.D. generated single-cell RNAseq and Multiome data and provided cDNA for circVDJseq. I.T., T.M.P. and N.R. generated Neuroblastoma Visium data and provided cDNA for circVDJseq. A.PR. and A.H. generated lung and lymph node Visium data and provided cDNA for circVDJseq. C.Q., J.W. and T.B. performed long- and short-read sequencing. B.O., R.R. and C.F. performed data analysis. I.P., B.O., R.R., H.R., A.H, and T.C. interpreted the data. I.P., B.O., and R.R. prepared figure panels. T.C. wrote the manuscript. A.H., A.E., M.S., and H.R. contributed clinical specimens. L.S.L. and J.A. commented on the manuscript. All authors reviewed the final manuscript.

Acknowledgements

We thank the patients who participated in this study. We thank Rogier Versteeg and Vedran Franke for critical reading of the manuscript. We thank the Deutsche Forschungsgemeinschaft (DFG; German Research Foundation) and the BIH for financial support.

References

1. Bassing CH, Swat W, Alt FW. The Mechanism and Regulation of Chromosomal V(D)J Recombination. Cell. 2002;109:S45–55.

2. Lythe G, Callard RE, Hoare RL, Molina-París C. How many TCR clonotypes does a body maintain? J Theor Biol. 2016;389:214–24.

3. Zarnitsyna VI, Evavold BD, Schoettle LN, Blattman JN, Antia R. Estimating the diversity, completeness, and cross-reactivity of the T cell repertoire. Front Immunol. 2013;4:68766.

4. Stubbington MJT, Lönnberg T, Proserpio V, Clare S, Speak AO, Dougan G, et al. T cell fate and clonality inference from single-cell transcriptomes. Nat Methods. 2016;13:329–32.

5. Guo X, Zhang Y, Zheng L, Zheng C, Song J, Zhang Q, et al. Global characterization of T cells in non-small-cell lung cancer by single-cell sequencing. Nat Med. 2018;24:978–85.

6. Zheng C, Zheng L, Yoo JK, Guo H, Zhang Y, Guo X, et al. Landscape of Infiltrating T Cells in Liver Cancer Revealed by Single-Cell Sequencing. Cell. 2017;169:1342-1356.e16.

7. Rodriques SG, Stickels RR, Goeva A, Martin CA, Murray E, Vanderburg CR, et al. Slide-seq: A scalable technology for measuring genome-wide expression at high spatial resolution. Science. 2019;363:1463–7.

8. Engblom C, Thrane K, Lin Q, Andersson A, Toosi H, Chen X, et al. Spatial transcriptomics of B cell and T cell receptors reveals lymphocyte clonal dynamics. Science. 2023;382:eadf8486.

9. Li S, Sun J, Allesøe R, Datta K, Bao Y, Oliveira G, et al. RNase H–dependent PCR-enabled T-cell receptor sequencing for highly specific and efficient targeted sequencing of T-cell receptor mRNA for single-cell and repertoire analysis. Nature Protocols 2019 14:8. 2019;14:2571–94.

10. Tu AA, Gierahn TM, Monian B, Morgan DM, Mehta NK, Ruiter B, et al. TCR sequencing paired with massively parallel 3′ RNA-seq reveals clonotypic T cell signatures. Nature Immunology 2019 20:12. 2019;20:1692–9.

11. Benotmane JK, Kueckelhaus J, Will P, Zhang J, Ravi VM, Joseph K, et al. High-sensitive spatially resolved T cell receptor sequencing with SPTCR-seq. Nature Communications 2023 14:1. 2023;14:1–17.

12. Singh M, Al-Eryani G, Carswell S, Ferguson JM, Blackburn J, Barton K, et al. High-throughput targeted long-read single-cell sequencing reveals the clonal and transcriptional landscape of lymphocytes. Nature Communications 2019 10:1. 2019;10:1–13.

13. Mothes R, Pascual-Reguant A, Koehler R, Liebeskind J, Liebheit A, Bauherr S, et al. Distinct tissue niches direct lung immunopathology via CCL18 and CCL21 in severe COVID-19. Nature Communications 2023 14:1. 2023;14:1–16.

14. Mimitou EP, Lareau CA, Chen KY, Zorzetto-Fernandes AL, Hao Y, Takeshima Y, et al. Scalable, multimodal profiling of chromatin accessibility, gene expression and protein levels in single-cells. Nature Biotechnology 2021 39:10. 2021;39:1246–58.

15. Hao Y, Hao S, Andersen-Nissen E, Mauck WM, Zheng S, Butler A, et al. Integrated analysis of multimodal single-cell data. Cell. 2021;184:3573-3587.e29.

16. McGinnis CS, Murrow LM, Gartner ZJ. DoubletFinder: Doublet Detection in Single-Cell RNA Sequencing Data Using Artificial Nearest Neighbors. Cell Syst. 2019;8:329-337.e4.

17. Suo C, Polanski K, Dann E, Lindeboom RGH, Vilarrasa-Blasi R, Vento-Tormo R, et al. Dandelion uses the single-cell adaptive immune receptor repertoire to explore lymphocyte developmental origins. Nature Biotechnology 2023 42:1. 2023;42:40–51.

18. Nam AS, Kim KT, Chaligne R, Izzo F, Ang C, Taylor J, et al. Somatic mutations and cell identity linked by Genotyping of Transcriptomes. Nature. 2019;571:355–60.

19. Bolotin DA, Poslavsky S, Mitrophanov I, Shugay M, Mamedov IZ, Putintseva E V., et al. MiXCR: software for comprehensive adaptive immunity profiling. Nature Methods 2015 12:5. 2015;12:380–1.

20. Song L, Cohen D, Ouyang Z, Cao Y, Hu X, Liu XS. TRUST4: immune repertoire reconstruction from bulk and single-cell RNA-seq data. Nature Methods 2021 18:6. 2021;18:627–30.

21. Hao Y, Hao S, Andersen-Nissen E, Mauck WM, Zheng S, Butler A, et al. Integrated analysis of multimodal single-cell data. Cell. 2021;184:3573-3587.e29.

22. Lee ES, Thomas PG, Mold JE, Yates AJ. Identifying T Cell Receptors from High-Throughput Sequencing: Dealing with Promiscuity in TCRα and TCRβ Pairing. PLoS Comput Biol. 2017;13:e1005313.

23. Stubbington MJT, Lönnberg T, Proserpio V, Clare S, Speak AO, Dougan G, et al. T cell fate and clonality inference from single-cell transcriptomes. Nature Methods 2016 13:4. 2016;13:329–32.

24. Padovan E, Casorati G, Dellabona P, Meyer S, Brockhaus M, Lanzavecchia A, et al. Expression of Two T Cell Receptor α Chains: Dual Receptor T Cells. Science. 1993;262:422–4.

25. Han A, Glanville J, Hansmann L, Davis MM. Linking T-cell receptor sequence to functional phenotype at the single-cell level. Nature Biotechnology 2014 32:7. 2014;32:684–92.

26. Miller TE, Lareau CA, Verga JA, DePasquale EAK, Liu V, Ssozi D, et al. Mitochondrial variant enrichment from high-throughput single-cell RNA sequencing resolves clonal populations. Nature Biotechnology 2022 40:7. 2022;40:1030–4.

27. Al'Khafaji AM, Smith JT, Garimella K V., Babadi M, Popic V, Sade-Feldman M, et al. High-throughput RNA isoform sequencing using programmed cDNA concatenation. Nature Biotechnology 2023 42:4. 2023;42:582–6.

28. Palla G, Spitzer H, Klein M, Fischer D, Schaar AC, Kuemmerle LB, et al. Squidpy: a scalable framework for spatial omics analysis. Nature Methods 2022 19:2. 2022;19:171–8.

29. Brodeur GM. Neuroblastoma: biological insights into a clinical enigma. Nature Reviews Cancer 2003 3:3. 2003;3:203–16.

30. Jansky S, Sharma AK, Körber V, Quintero A, Toprak UH, Wecht EM, et al. Single-cell transcriptomic analyses provide insights into the developmental origins of neuroblastoma. Nature Genetics 2021 53:5. 2021;53:683–93.

31. Wachowiak R, Mayer S, Suttkus A, Martynov I, Lacher M, Melling N, et al. CHL1 and NrCAM are Primarily Expressed in Low Grade Pediatric Neuroblastoma. Open Medicine. 2019;14:920.

32. Li HL, Huang DZ, Deng T, Zhou LK, Wang X, Bai M, et al. Overexpression of cyclin L2 inhibits growth and enhances chemosensitivity in human gastric cancer cells. Asian Pac J Cancer Prev. 2012;13:1425–30.

33. Chen QR, Bilke S, Wei JS, Greer BT, Steinberg SM, Westermann F, et al. Increased WSB1 copy number correlates with its over-expression which associates with increased survival in neuroblastoma. Genes Chromosomes Cancer. 2006;45:856–62.

34. Castaño Z, Vergara-Irigaray N, Pajares MJ, Montuenga LM, Pio R. Expression of αCP-4 inhibits cell cycle progression and suppresses tumorigenicity of lung cancer cells. Int J Cancer. 2008;122:1512–20.

35. Yan W, Scoumanne A, Jung YS, Xu E, Zhang J, Zhang Y, et al. Mice deficient in poly(C)-binding protein 4 are susceptible to spontaneous tumors through increased expression of ZFP871 that targets p53 for degradation. Genes Dev. 2016;30:522–34.

36. Huang Z, Wang Z, Xia H, Ge Z, Yu L, Li J, et al. Long noncoding RNA HAND2-AS1: A crucial regulator of malignancy. Clinica Chimica Acta. 2023;539:162–9.

37. Voth H, Oberthuer A, Simon T, Kahlert Y, Berthold F, Fischer M. Identification of DEIN, a Novel Gene with High Expression Levels in Stage IVS Neuroblastoma. Molecular Cancer Research. 2007;5:1276–84.

38. Abualsaud N, Caprio L, Galli S, Krawczyk E, Alamri L, Zhu S, et al. Neuropeptide Y/Y5 Receptor Pathway Stimulates Neuroblastoma Cell Motility Through RhoA Activation. Front Cell Dev Biol. 2021;8:627090.

39. Galli S, Naranjo A, Van Ryn C, Tilan JU, Trinh E, Yang C, et al. Neuropeptide Y as a Biomarker and Therapeutic Target for Neuroblastoma. Am J Pathol. 2016;186:3040.

40. Hu Z, Xu W, Wang H, Li M, Wang J, Sun C, et al. CARM1-induced lncRNA NEAT1 synchronously activates MYCN and GalNAcT-I to accelerate the progression of neuroblastoma. Gene. 2025;938:149164.

41. Yang B, Ye X, Wang J, Xia S. Long noncoding RNA nuclear-enriched abundant transcript 1 regulates proliferation and apoptosis of neuroblastoma cells treated by cisplatin by targeting miR-326 through Janus kinase/signal transducer and activator of transcription 3 pathway. Neuroreport. 2020;31:1189–98.

42. Li RQ, Yan L, Zhang L, Zhao Y, Lian J. CD74 as a prognostic and M1 macrophage infiltration marker in a comprehensive pan-cancer analysis. Scientific Reports 2024 14:1. 2024;14:1–18.

43. Ren L, Yi J, Yang Y, Li W, Zheng X, Liu J, et al. Systematic pan-cancer analysis identifies APOC1 as an immunological biomarker which regulates macrophage polarization and promotes tumor metastasis. Pharmacol Res. 2022;183:106376.

44. Alcaraz LB, Mallavialle A, Mollevi C, Boissière-Michot F, Mansouri H, Simony-Lafontaine J, et al. SPARC in cancer-associated fibroblasts is an independent poor prognostic factor in non-metastatic triple-negative breast cancer and exhibits pro-tumor activity. Int J Cancer. 2023;152:1243–58.

45. Li D, Xia L, Huang P, Wang Z, Guo Q, Huang C, et al. Cancer-associated fibroblast-secreted IGFBP7 promotes gastric cancer by enhancing tumor associated macrophage infiltration via FGF2/FGFR1/PI3K/AKT axis. Cell Death Discovery 2023 9:1. 2023;9:1–12.

46. Rupp C, Scherzer M, Rudisch A, Unger C, Haslinger C, Schweifer N, et al. IGFBP7, a novel tumor stroma marker, with growth-promoting effects in colon cancer through a paracrine tumor–stroma interaction. Oncogene 2015 34:7. 2014;34:815–25.

47. Brown KM, Xue A, Gill AJ, Hugh TJ. IGFBP7 and POSTN Are Prognostic Biomarkers in the Cancer Associated Stroma of Colorectal Liver Metastases. HPB. 2021;23:S124–5.

48. Patel AG, Ashenberg O, Collins NB, Segerstolpe Å, Jiang S, Slyper M, et al. A spatial cell atlas of neuroblastoma reveals developmental, epigenetic and spatial axis of tumor heterogeneity. bioRxiv. 2024;2024.01.07.574538.

49. Shimizu M, Koma YI, Sakamoto H, Tsukamoto S, Kitamura Y, Urakami S, et al. Metallothionein 2a expression in cancer-associated fibroblasts and cancer cells promotes esophageal squamous cell carcinoma progression. Cancers (Basel). 2021;13:4552.

50. Habel N, Hamidouche Z, Girault I, Patiño-García A, Lecanda F, Marie PJ, et al. Zinc chelation: a metallothionein 2A's mechanism of action involved in osteosarcoma cell death and chemotherapy resistance. Cell Death & Disease 2013 4:10. 2013;4:e874–e874.

51. Carlson CS, Emerson RO, Sherwood AM, Desmarais C, Chung MW, Parsons JM, et al. Using synthetic templates to design an unbiased multiplex PCR assay. Nat Commun. 2013;4:1–9.

52. Rodriques SG, Stickels RR, Goeva A, Martin CA, Murray E, Vanderburg CR, et al. Slide-seq: A scalable technology for measuring genome-wide expression at high spatial resolution. Science. 2019;363:1463–7.

53. Stickels RR, Murray E, Kumar P, Li J, Marshall JL, Di Bella DJ, et al. Highly sensitive spatial transcriptomics at near-cellular resolution with Slide-seqV2. Nat Biotechnol. 2021;39:313–9.

54. Chen A, Liao S, Cheng M, Ma K, Wu L, Lai Y, et al. Spatiotemporal transcriptomic atlas of mouse organogenesis using DNA nanoball-patterned arrays. Cell. 2022;185:1777-1792.e21.

Figures

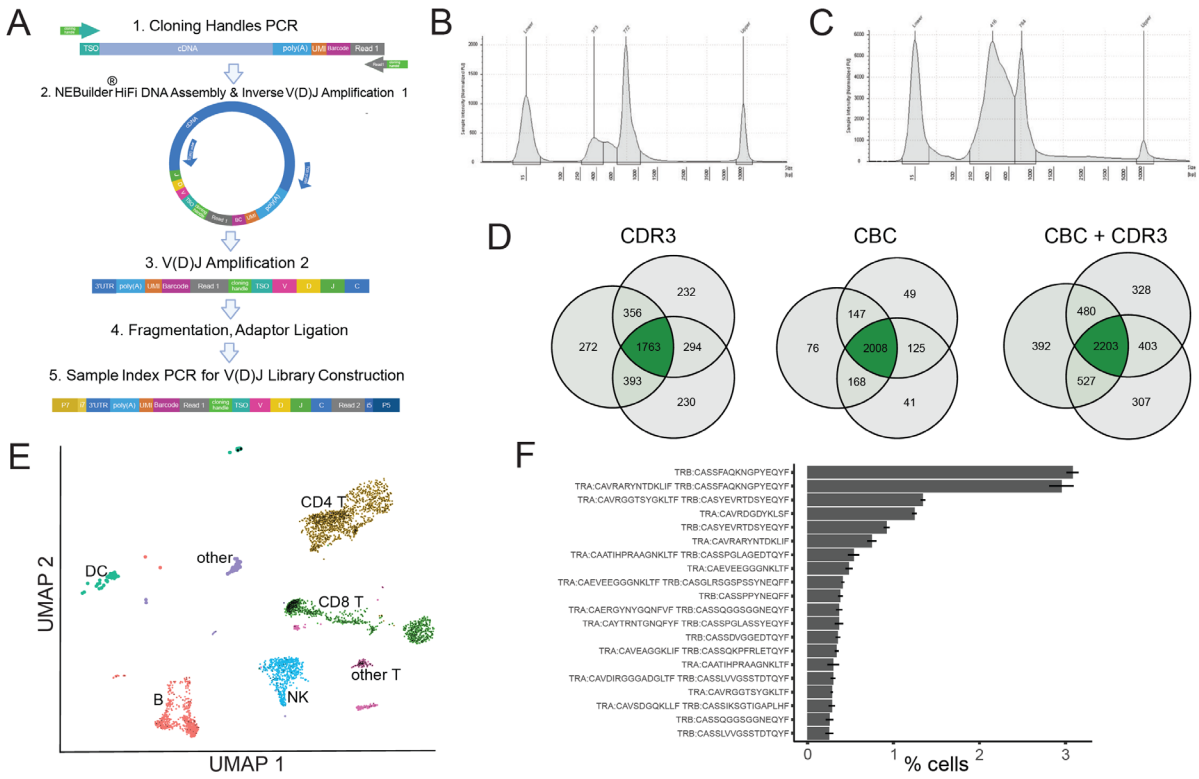


Fig. 1 circVDJ-seq enables TCR profiling with 3'-directed single-cell RNAseq

A) circVDJ-seq workflow: both ends of the cDNA library are appended by short homologous sequences to enable circularization via Gibson assembly and reposition 3'-cell barcodes next to the 5'end of cDNA molecules. This is followed by TCR-specific cDNA amplification, fragmentation,

and short-read library amplification. B) Amplified TCR cDNA library from human PBMCs. Initial cDNA was generated with the 3'GEX v3.1 assay (10X Genomics). C) TCR short-read sequencing library generated from (B). D) overlap of identified CDR3 sequences, cell barcodes (CBCs), and CDR3+CBC combinations in three technical circVDJ-seq replicates derived from the same cDNA library. E) UMAP plot showing gene expression data from human PBMCs with identified TCR clones highlighted as black dots. F) TCR clonotype frequencies, error bars from the s.e.m. across three circVDJ-seq replicates.

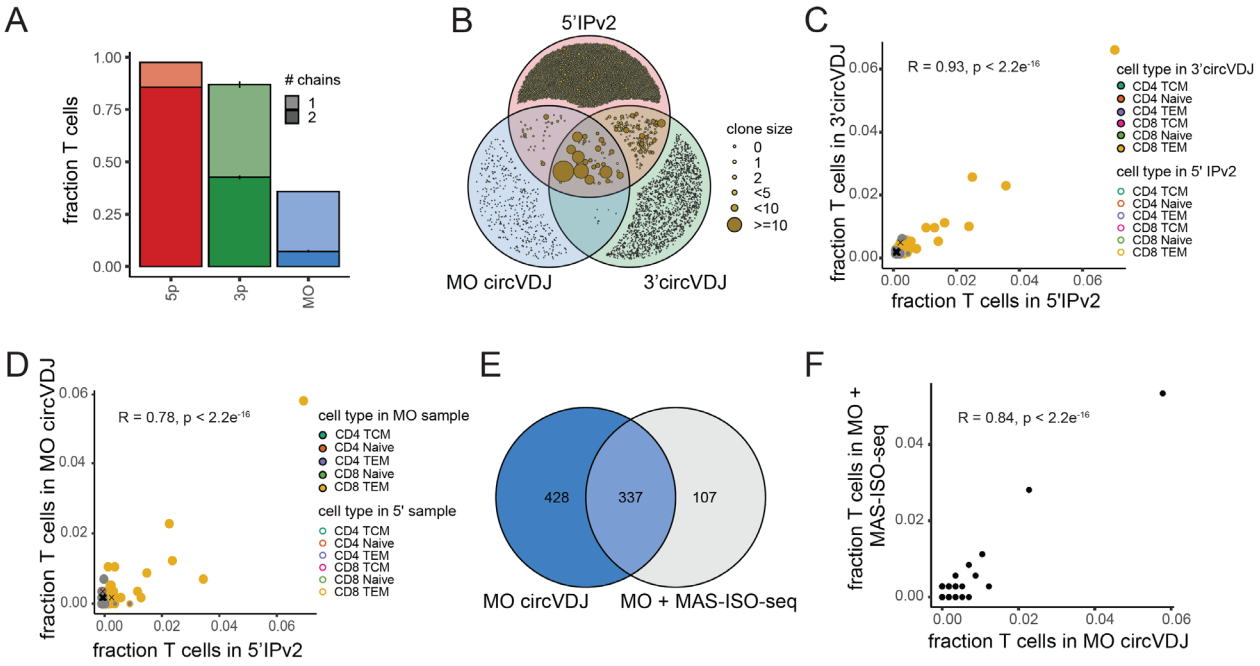


Fig. 2 Robust TCR clonotype identification across single-cell multi-omics workflows

A) Fraction of T cells with assigned clonotypes in 5'IPv2, 3'circVDJ-seq and MO circVDJ-seq. Error bars (for 3' and MO assays) from s.e.m. across 3 technical replicates. B) Venn diagram shows overlap between the clonotypes recovered by all three methods. Sizes of overlapping clones are based on the frequencies observed with 5'IPv2. C) Scatter plot shows the frequencies of individual T cell clones detected with 3'circVDJ-seq or 5'IPv2 in PBMCs from the same donor. Fill and border color indicate the cell types assigned to each clone by the respective matched gene expression library, and crosses mark clones that are assigned to differing cell types in both assays. D) Scatter plot shows clonotype frequencies detected with MO circVDJ-seq and 5'IPv2 in PBMCs from the same donor as in C). E) Number of T cell clones detected by MO circVDJ-seq and MAS-ISO long-read sequencing of the same MO cDNA. F) Scatter plot shows T cell clonotype frequencies detected by MO circVDJ-seq and MAS-ISO-seq of the same MO cDNA.

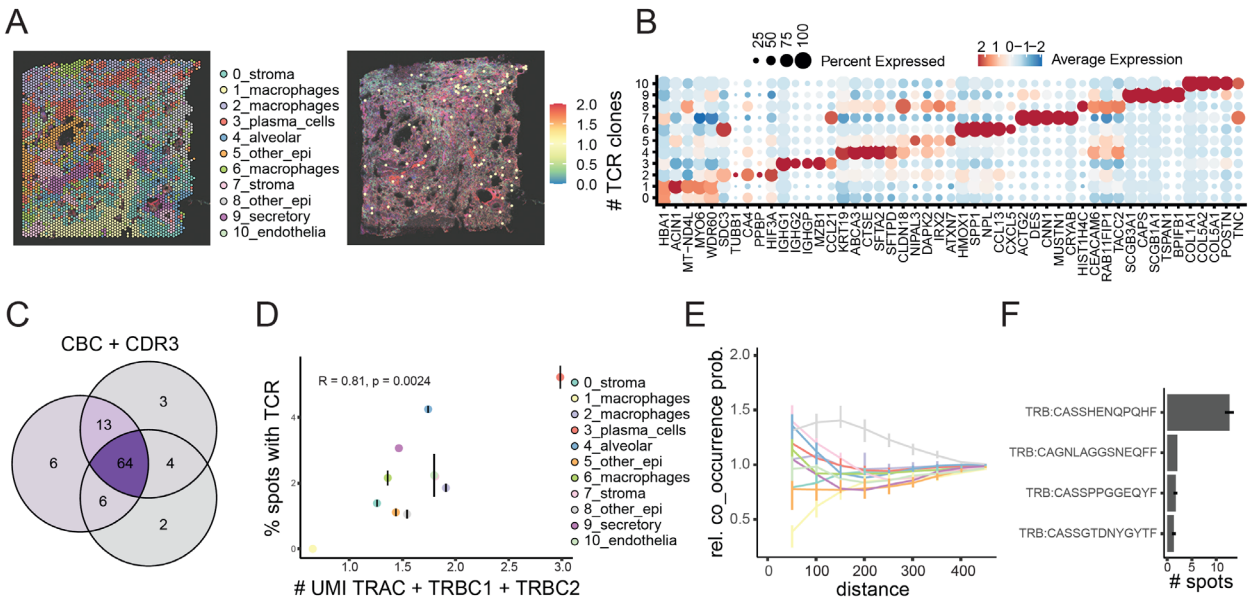


Fig. 3 spatial circVDJ-seq reveals clonal T cell expansion in autopsy derived COVID-19 lung tissue

A) Gene expression clusters and TCR clones identified by Visium and circVDJ-seq in chronic COVID-19 lung (chronic case 5). B) marker genes for clusters shown in A. C) Venn diagram shows overlap between the CBC and CDR3 combinations detected in technical circVDJ-seq replicates. D) Correlation between Visium TCR gene expression UMI counts and the fraction of spots with circVDJ-seq TCR clone found in the same clusters. Error bars indicate s.e.m. across technical triplicates. E) Relative co-occurrence probabilities of spots with TCR clonotypes with spots from other clusters within a certain distance. Error bars indicate std. deviation across random subsamples of index spots. F) Reproducible detection of identified TCR clonotypes, error bars correspond to three circVDJ-seq replicates. Error bars indicate s.e.m. across technical triplicates.

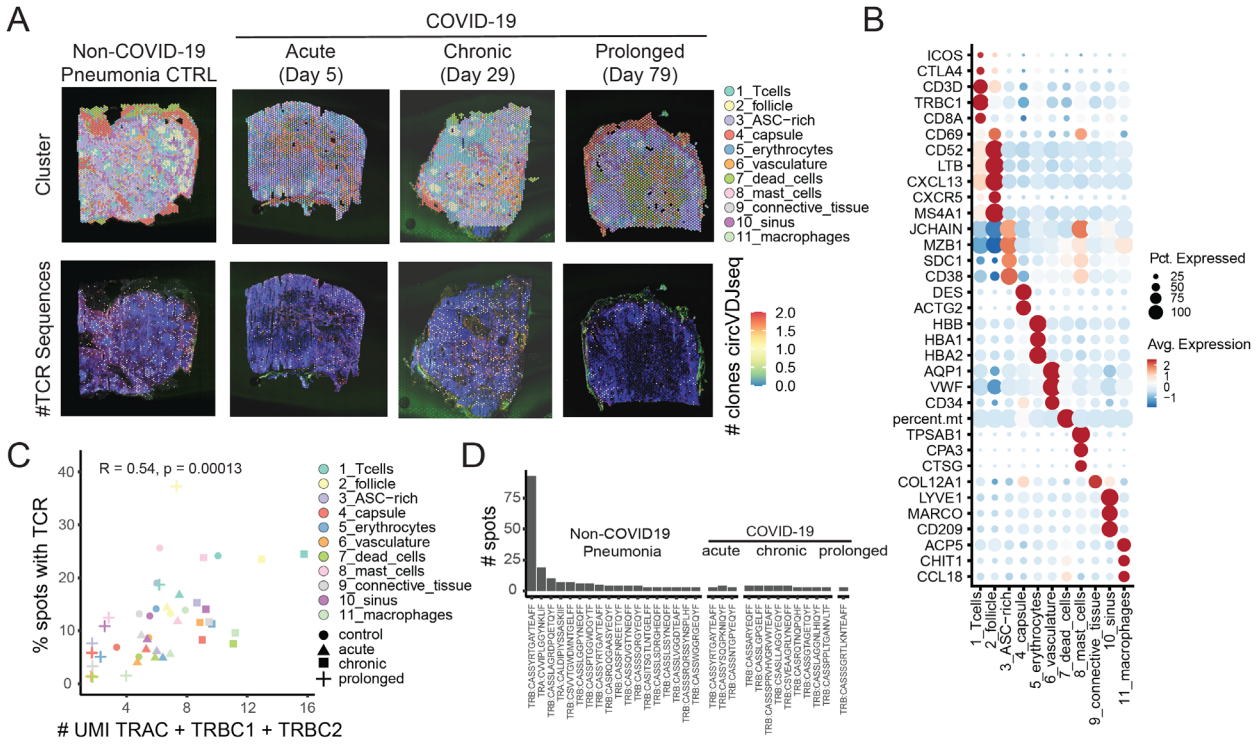


Fig. 4 spatial circVDJ-seq shows distinct clonal T cell expansion in lung draining lymph nodes

A) Visium gene expression clustering and identified TCR clones in autopsy-derived lymph nodes from non-COVID-19 pneumonia and squamous cell carcinoma (control case 2). B) gene expression marker genes for clusters shown in A. C) Correlation between TCR gene expression UMI counts and percentage of spots with identified TCR clones for individual gene expression clusters. D) TCR clonotype frequencies for clones identified in >2 spots.

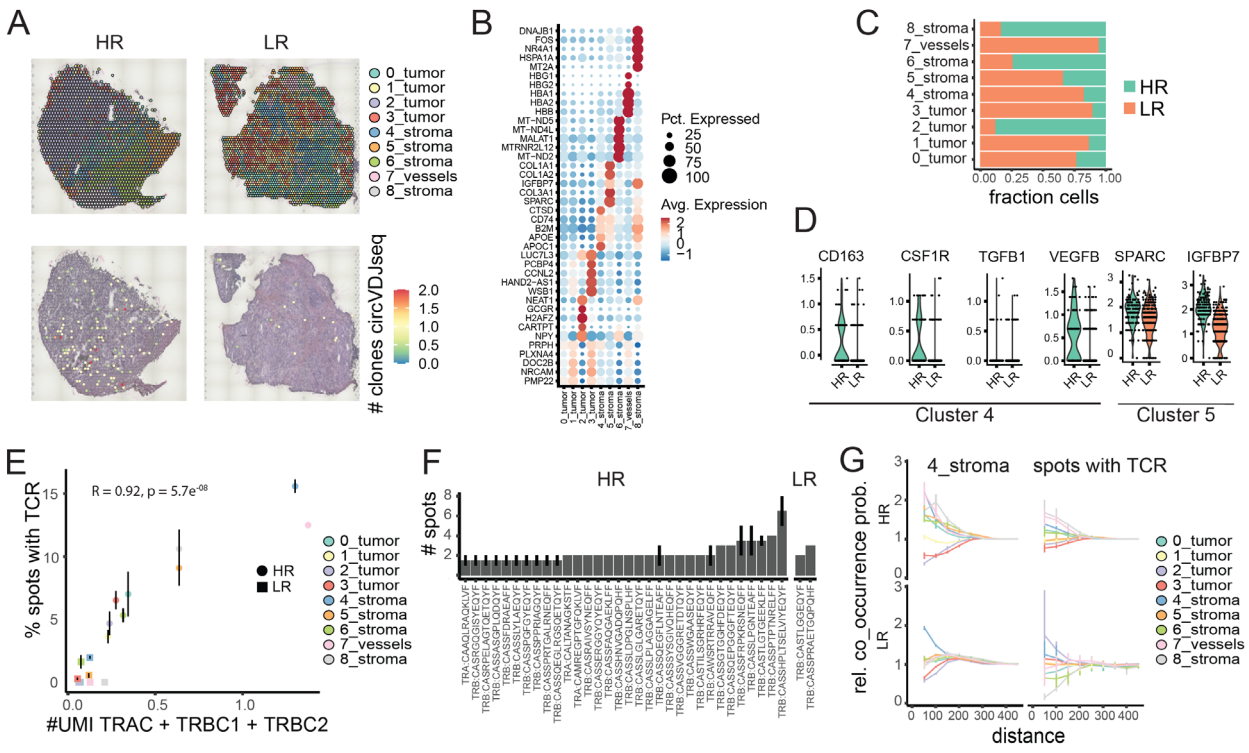
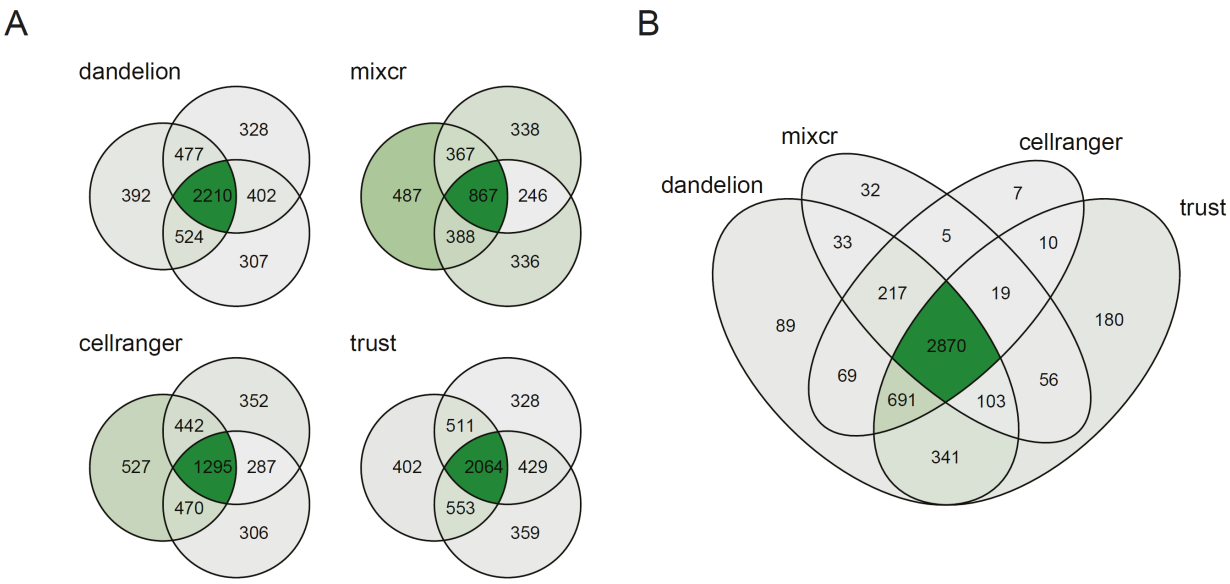


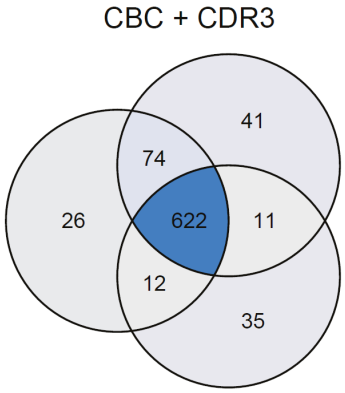
Fig. 5 spatial circVDJ-seq reveals distinct T cell dynamics in HR and LR Neuroblastoma

A) Gene expression clusters (top) and TCR clones (bottom) identified by Visium and circVDJ-seq of high risk (HR) and low risk (LR) neuroblastoma. B) Marker genes for clusters shown in A. C) Cluster composition of HR-NB and LR-NB sample for the clusters in A. D) Expression of CD163, CSF1R, TGFB1 and VEGFB in cluster 4, and SPARC and IGFBP7 in cluster 5 in HR and LR neuroblastoma. E) Correlation between Visium TCR gene expression UMI counts and the fraction of spots with a circVDJ-seq TCR clone in individual gene expression clusters. Error bars indicate s.e.m. across technical duplicates. F) Frequencies of spatially expanded clonotypes in the HR and LR neuroblastoma sample. Error bars indicate s.e.m. across technical duplicates. G) Relative co-occurrence probabilities of spots from cluster 4 (left panel) and of spots with TCR clonotypes (right panel) with spots from other clusters within a certain distance in HR and LR neuroblastoma. Error bars indicate std. deviation across random subsamples of index spots.



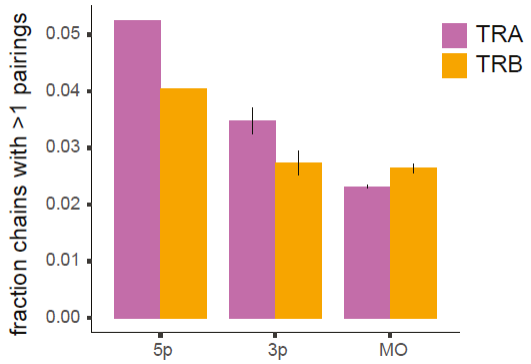
Supplementary Fig. 1

A) Venn diagrams of the overlap of CBC+CDR3 combinations between technical replicates for different VDJ sequencing data processing pipelines. B) Venn diagram of overlap between CBC+CDR3 combinations obtained on pooled replicates for different processing methods.



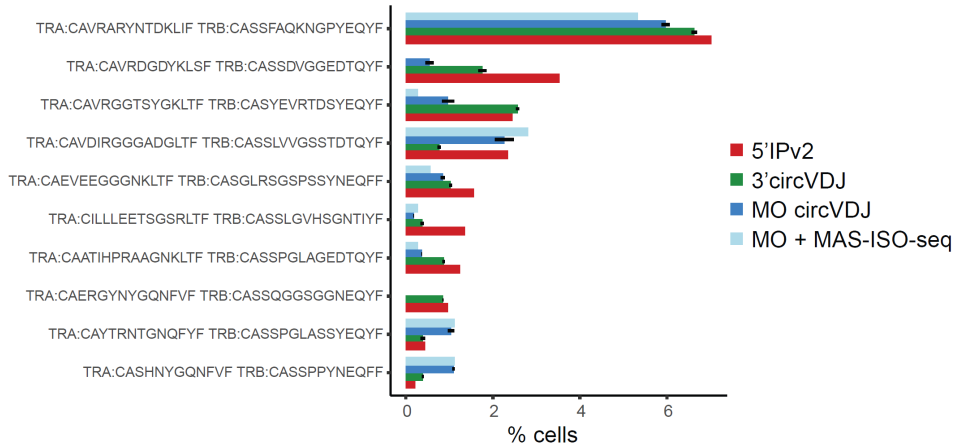
Supplementary Fig. 2

Venn Diagram of the overlap of CBC+CDR3 combinations between circVDJ-seq replicates of ATAC+RNA MO data.



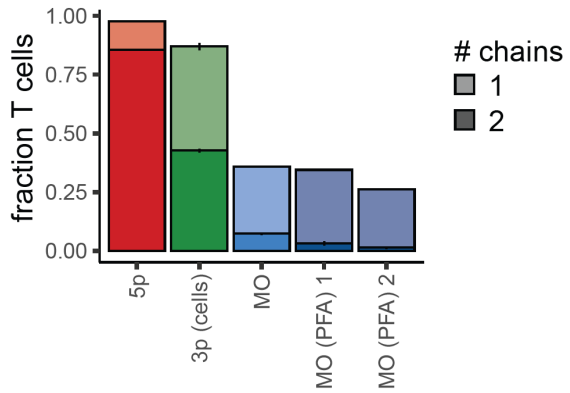
Supplementary Fig. 3

Frequency of TCRα or TCRβ chains that are detected in more than one pairing in 5'IP v2 or circVDJ-seq data from 3' or MO cDNA. Error bars from s.e.m. across technical replicates.



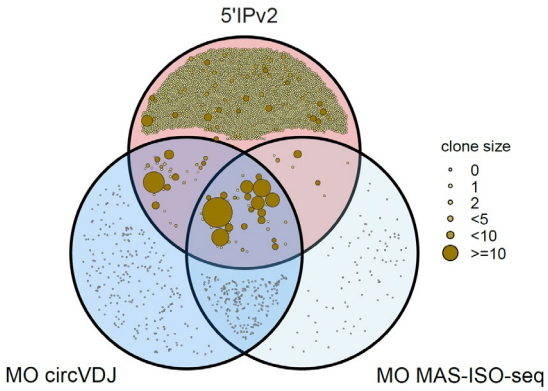
Supplementary Fig. 4

Clone sizes for the top 10 TCRα+β clones (missing chains imputed) detected in 5'IP v2, circVDJ-seq from 3' or MO cDNA, or MAS-ISO-seq data from MO cDNA, respectively. Error bars from s.e.m. across technical replicates.



Supplementary Fig. 5

Fraction of T cells with assigned clonotypes in 5'IPv2, 3'v3.1 circVDJ-seq and MO circVDJ-seq as in Fig. 2A) together with two additional MO datasets using mild PFA fixation. Error bars (for 3' and MO assays) from s.e.m. across 3 technical replicates.



Supplementary Fig. 6

Venn diagram as in Fig. 2B shows overlap between the clonotypes recovered by 5' 5'IPv2 as well as circVDJ-seq and MAS-ISO-seq of the same MO cDNA. Sizes of overlapping clones are based on the frequencies observed with 5'IPv2.