THE UNIVERSITY OF TEXAS
MD Anderson
~~Cancer~~ Center
Making Cancer History®

Date: July 18, 2023

Cancer Immune Monitoring and Analysis Center
Department of Translational Molecular Pathology
The University of Texas at MD Anderson Cancer Center
Contact PI: Ignacio Ivan Wistuba (e-mail: iiwistuba@mdanderson.org)

Performance lab:
Translational Molecular Pathology and Immunoprofiling Lab (TMP-IL)
Ignacio Ivan Wistuba, MD (Director)
Beatriz Sanchez Espiridion, PhD (Project Director)
Cara Haymaker, PhD (Director)
Katarzyna Tomczak, PhD (Leading Scientist: performed the assay and data analysis)
Karen Millerchip, BS, SCYM (ASCP) (Laboratory Coordinator)
Kunal Rai, Ph.D. (Collaborator)
J Jack Lee, Ph.D., M.S., D.D.S. (Biostatistics)
Jiexin Zhang, MS (Biostatistics)
Ken Chen, PhD (Bioinformatics)

Single Cell RNA Sequencing with TCR Enrichment Validation Report, version 2.1

This report describes the analytical validation of single cell RNA sequencing with matched single cell TCR sequencing of cryopreserved, characterized melanoma tumor cells and corresponding Tumor-Infiltrating Lymphocytes (TIL) isolated from the two metastatic melanoma patients.

Table 1. Summary of analytical validation findings for scRNA-seq profiling

Accuracy	Accurate assay performance determined by fragment size/peak distribution of libraries (Figure 1, 2) and QC plots of sequenced data (Figure 5), obtained QC metrics (Table 3, 4) compared to a reference sample (10x Genomics).
Precision: Inter-assay	Technical reproducibility across analyzed samples. Gene expression patterns between technical replicates (n=2), and assessment of ability to detect the relevant information in 1:10 dilution of TIL in mixed sample (Figure 6 - 16, also see Table 3, 4).
Precision: Intra-assay	Plots and detection of cellular characteristic pattern (Figure 8). Differential analysis shows common patterns of gene expression (Figure 10 - 13; also see Table 3, 4).
Analytical sensitivity	Targeting 500 – 10,000 cells, depending on collection and preparation method and recovery rate. Here, we targeted 3,000 cells for library preparation (see ‘Table 3, 4: estimated number of cells).
Analytical sensitivity including interfering substances	Preservation and storage condition (fresh vs frozen) of samples could affect cell quality (10x Genomics technical notes). Here we utilized an optimized gold standard freezing method verified in the lab across a broad range of primary tumor and cell line samples.

Reportable range	Sample type dependent.
Reference interval (normal range)	Generated data represent the reference ranges of QC metrics of publicly available PBMC (10x Genomics), see Table 3.
Standardization, harmonization, reproducibility, and ruggedness	Standardized with a pooled library of unique indexed samples, sequenced on one flow cell type on NovaSeq6000 Sequencer to minimize technical variation. All analysis for a given sample set performed against the same reference data base.
Quality control and improvement procedures	Experiments carried out with 2 technical and 2 biological replicates (depending on the cells accessibility). Processing of single cell RNA sequencing data and matched TCR-seq was done with the cell ranger (10x Genomics) pipeline. The QC metrics are summarized in Table 3, 4 and Figure 6. Doublets were assessed and removed (Figure 7).
Any other performance	N/A

1. Purpose of the assay.

Profiling sequencing techniques at the single cell level [1], greatly exceeded the standard bulk sequencing [2], allowing to dissect the extremely heterogenous nature of tumors and the equally diverse immune repertoire in a population [3]. Moreover, it enables clonal evolution tracking, assessing tumor microenvironment remodeling status and development of cancer therapy resistance. The application of single-cell sequencing technologies, in combination with existing genomics methods, has greatly facilitated high-resolution interrogation of the genome [4], transcriptome, proteome [5, 6] and epigenome [7, 8]. 10x Genomics (https://www.10xgenomics.com) is one of the leading single cell platforms for integrated solution for RNA and DNA analysis currently extensively applied by the scientific community.

Here, our focus was to validate 10x Genomics Chromium Single Cell Immune Profiling Solution (https://www.10xgenomics.com/product-list/#vdj), an approach to measure the immune repertoire information and gene expression from the same cell. The single cell profiling of full-length (5'UTR to constant region) and paired T-cell receptor (TCR) transcripts is widely used for studying the transcriptome at high resolution of tens of thousands of single cells in parallel. Moreover, 5' single cell gene expression profiling is a method relying on next-generation sequencing (NGS) library construction and generated counts can be analyzed using 10x Genomics developed pipeline followed by downstream analysis and visualization in 10x Genomics Loupe Browser and/or in R. Here, we utilized primary tumor cell lines expanded from metastatic melanoma resections and the corresponding Tumor-Infiltrating Lymphocytes (TIL) to assess cell type unique patterns underlying gene expression differences at the single cell level.

2. Materials & Methods.

Materials:

The 5' single cell RNA-seq assay paired with TCR-enrichment step was performed using Chromium Next GEM Single Cell V(D)J Reagents Kits version 1.1 (10x Genomics) on 2 sets of cryopreserved primary melanoma cell lines and their corresponding TIL expanded from 2 metastatic melanoma patients (#2559 and #2765), obtained and expanded from each tumor as previously described [9].

Table 2. Samples

Batch Number	Sample #	Library ID, Downstream Analysis Sample Name
Set 1	1.	Mel 2559 (100%), M2559
	2.	TIL 2559 (100%), T2559
	3.	Mel 2559 (90%) + TIL 2559 (10%), MT2559
Set 2	4.	Mel 2765 (100%), M2765
	5.	TIL 2765 (100%), T2765
	6.	Mel 2765 (90%) + TIL 2765 (10%), MT2765

In order to test whether it's possible to detect a small fraction of TIL in heterogenous samples, the assay validation was performed by spiking autologous TIL into the matched melanoma tumor cell line (10% TIL + 90% tumor cell line), see Table 2.

Methods: detailed methodology, equipment and reagents details can be found within following documents:

- CG000053_CellPrepGuide_RevC.pdf,
- CG000207_ChromiumNextGEMSingleCellV_D_J_ReagentKits_v1.1_UG_RevE,
- https://support.10xgenomics.com/single-cell-gene-expression/software/pipelines/latest/output/summary and other related documentation deposited on https://support.10xgenomics.com/single-cell-vdj

The Single Cell V(D)J protocols provides comprehensive, scalable solutions for measuring immune repertoire information and gene expression from the same cell. Assay enables to profile full-length (5' UTR to constant region), along with paired T-cell receptor (TCR), or B-cell immunoglobulin (Ig) transcripts. The capacity of the assay allows for profiling from 100-10,000 individual cells per sample. A pool of ~750,000 barcodes are sampled separately to index each cell's transcriptome. It does so by partitioning thousands of cells into nanoliter-scale Gel Beads-in-emulsion (GEMs), where all generated cDNA share a common 10x Barcode. Libraries are generated and sequenced and 10x Barcodes are used to associate individual reads back to the individual partitions. This document outlines the protocol to generate 5' Gene Expression (5'GEX) library along with a V(D)J (T-cell enriched) library amplified from cDNA from the same cells, using The Chromium Next GEM Single Cell V(D)J Reagent Kits v1.1.

Briefly, each sample was thawed according to the optimized protocol and assessed for the cell count and viability (expected >= 80%) using Trypan Blue as wells as AO/PI staining in the Automated Cell Counter (Nexcelom Bioscience). Specific viability for melanoma tumor line and TIL was recorded as: 95%, 98% for Set 1 and 90%, 95% for Set 2, respectively. Next, for each sample the 1000 cells/uL stock was prepared. Part of each stock was directly loaded to the 10x chip for GEM generation and was further referred as 100% of cells of each sample. For loading volumes see table in the corresponding guide (CG000207) for targeting 3,000 cells. The mixture of TIL with cancer cells in 1:10 ratio was prepared by mixing 900uL of cancer cell sample with 100uL of corresponding TIL sample for each case, and appropriate volume was loaded to 10x chip for GEM generation.

(1) **GEMs generation & Barcoding** – nanoliter-scale Gel Beads-in-emulsion (GEMs) were generated by combining barcoded Single Cell V(D)J 5' Gel Beads v1.1, a Master Mix with cells, and Partitioning Oil on Chromium Next GEM Chip G, and run in 10x Genomics Chromium Controller. To achieve single cell resolution, cells were delivered at a limiting dilution, such that the majority (~90 – 99%) of generated GEMs contains no cell, while the remainder largely contain a single cell. Immediately following GEM generation, the Gel Bead was dissolved and any co-partitioned cell was lysed. Oligonucleotides containing (i) an Illumina R1 sequence (read 1 sequencing primer), (ii) a 16 nt 10x Barcode, (iii) a 10 nt unique molecular identifier (UMI), and (iv) 13 nt template switch oligo (TSO) were released and mixed with the cell lysate and a Master Mix containing reverse transcription (RT)

reagents and poly(dT) RT primers. Incubation of the GEMs produces 10x Barcoded, full-length cDNA from poly-adenylated mRNA was performed.

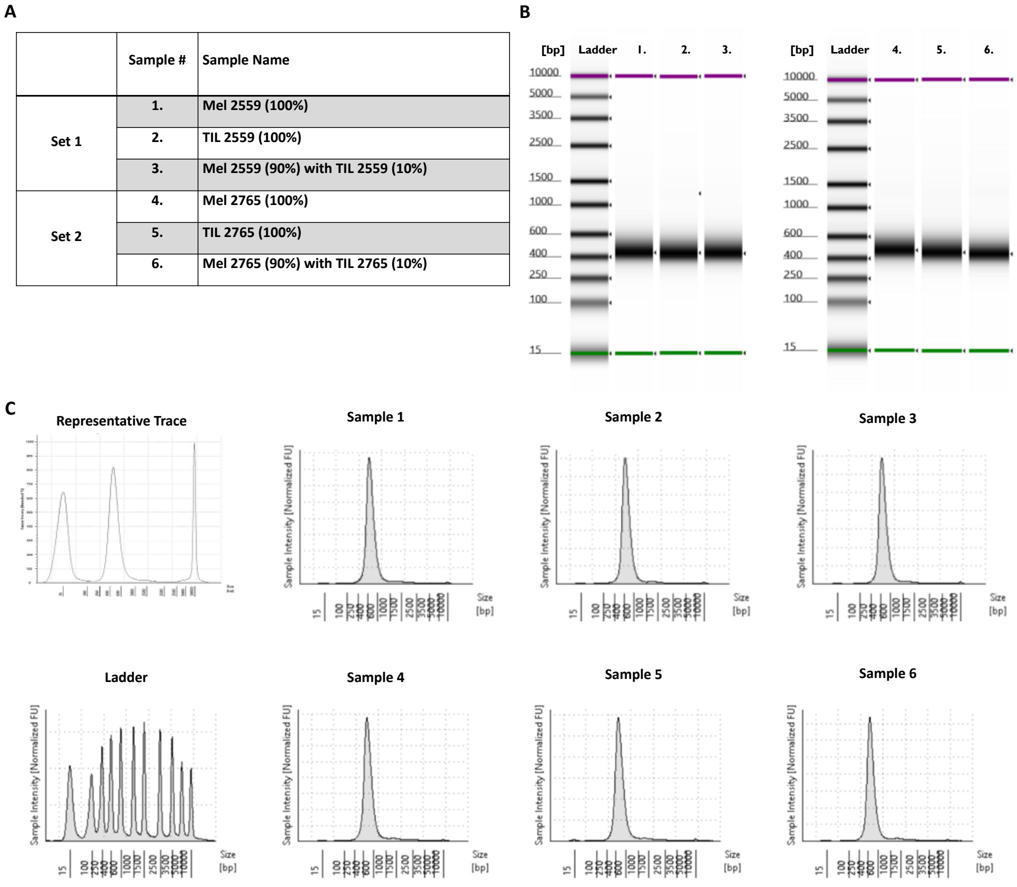
(2) **Post GEM Incubation Cleanup** – After incubation, the GEMs were broken, and pooled fractions were recovered. Silane magnetic beads were used to purify the 10x Barcoded first-strand cDNA from the post GEM-RT reaction mixture, which includes leftover biochemical reagents and primers.

(3) **cDNA Amplification & QC** – 10x Barcoded, full-length cDNA was amplified via PCR with primers against common 5' and 3' ends added during GEM-RT. Amplification generates sufficient material to construct multiple libraries from the same cells; here T cell enriched libraries and 5' Gene Expression libraries were generated.
(4) **Target Enrichment from cDNA** – Full-length V(D)J segments (10x Barcoded) were enriched from amplified cDNA via PCR amplification with primers specific to the TCR regions. P5 was added during enrichment. The TCR-enrichment was performed on Sample 2 (sample with 100% TIL) and Sample 3 (sample with 10% TIL) of each set (Set 1 and Set 2). Additionally, two input conditions: standard 2uL of cDNA (condition A) and 5uL of cDNA (condition B) were tested, aiming to better understand any potential performance benefits of addition of more cDNA and as an additional replicate. The cancer line samples (0% TIL) were not used for V(D)J assessment.

(5) **TCR Enriched Library Construction** – 50ng of TCR enriched cDNA was used at that step. Enzymatic fragmentation and size selection were used to generate variable length fragments that collectively span the V(D)J segments of the enriched TCR transcripts prior to library construction. P7, a sample index, and an Illumina R2 sequence (read 2 primer sequence) were added via End Repair, A-tailing, Adaptor Ligation, and Sample Index PCR. The final libraries contain the P5 and P7 priming sites used in Illumina sequencing.

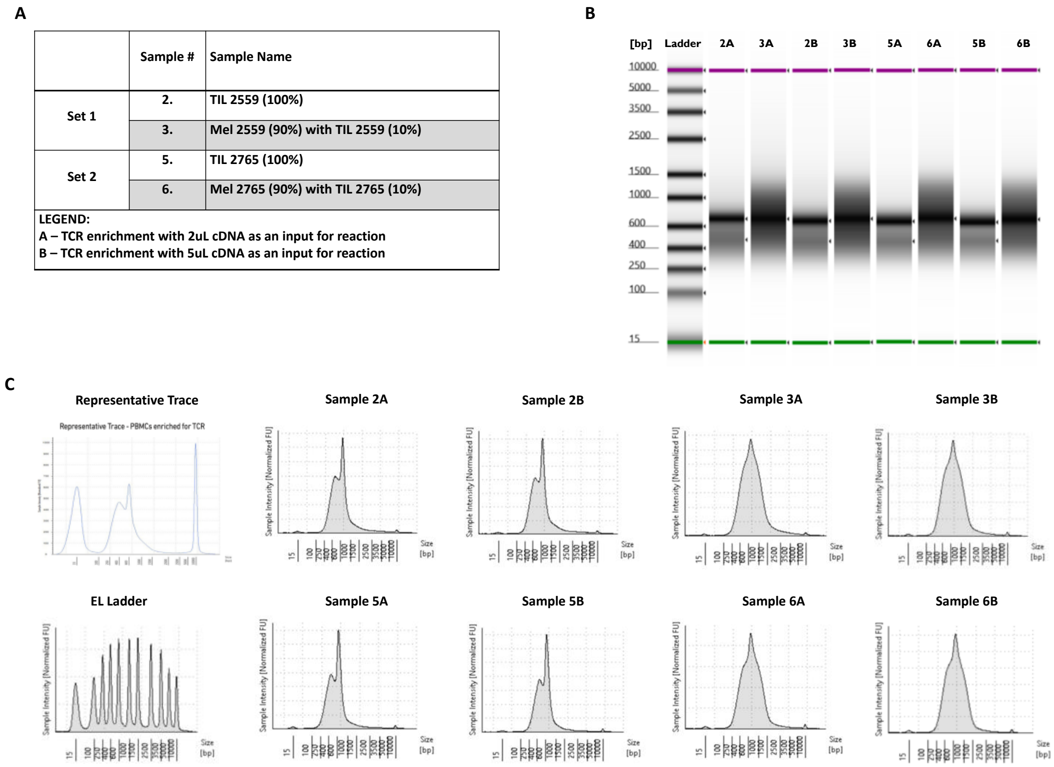
(6) **5'Gene Expression (GEX) Library Construction** – 50ng of cDNA (from step 3) was used. Enzymatic fragmentation and size selection were used to optimize the cDNA amplicon size prior to 5' GEX library construction. P5, P7, a sample index, and Illumina R2 sequence (read 2 primer sequence) were added via End Repair, A-tailing, Adaptor Ligation, and Sample Index PCR. The final libraries contain the P5 and P7 priming sites used in Illumina sequencers.

The library QC (4200 Tape Station High Sensitivity (HS) D5000 Assay; Agilent) of prepared 5'GEX and TCR-enriched libraries is indicated in the Figure 1 and Figure 2, respectively, along with the reference guide criteria (10x Genomics).



	Sample #	Sample Name
Set 1	1.	Mel 2559 (100%)
	2.	TIL 2559 (100%)
	3.	Mel 2559 (90%) with TIL 2559 (10%)
Set 2	4.	Mel 2765 (100%)
	5.	TIL 2765 (100%)
	6.	Mel 2765 (90%) with TIL 2765 (10%)

Figure 1. QC of 5'GEX libraries of scRNA-seq assay. (A) Sample annotation. (B) Fragment length distribution pattern and (C) peaks distribution using Agilent 4200 Tape Station HS D5000 assay.



	Sample #	Sample Name
Set 1	2.	TIL 2559 (100%)
	3.	Mel 2559 (90%) with TIL 2559 (10%)
Set 2	5.	TIL 2765 (100%)
	6.	Mel 2765 (90%) with TIL 2765 (10%)
LEGEND: A – TCR enrichment with 2uL cDNA as an input for reaction B – TCR enrichment with 5uL cDNA as an input for reaction		

Figure 2. QC of TCR enriched libraries of scRNA-seq assay. (A) Sample annotation. (B) Fragment length distribution pattern and (C) peaks distribution using Agilent 4200 Tape Station HS D5000 assay.

(7) **Sequencing** – Illumina-ready sequencing libraries were pooled and sequenced at the recommended depth & run parameters. The generated libraries (Figure 3) contain unique indexes which allowed pooling all libraries together maintaining an equal molarity ratio and sequencing on the same sequencing lane of the Novaseq 6000 sequencer (ATGC MDACC core facility; NIH 1S10OD024977-01) to minimize technical variation. The 10:1 ratio, as for 5'GEX to TCR-enriched libraries was applied covering the optimum sequencing depth of 50,000 read pairs per cell for 5'GEX and 5,000 of TCR-enriched libraries. Libraries were sequenced using the following read length format: 26bp Read1 and 91bp Read2. Chromium Single Cell V(D)J enriched libraries and 5' Gene Expression libraries comprise standard Illumina paired-end constructs which begin with P5 and end with P7. 16 bp 10x Barcodes are encoded at the start of Read 1, while sample index sequences are incorporated as the i7 index read. Read 1 and Read 2 are standard Illumina sequencing primer sites used in paired-end sequencing. Read 1 is used to sequence 16 bp 10x Barcodes and 10 bp UMI. Sequencing these libraries produced a standard Illumina BCL data output folder.

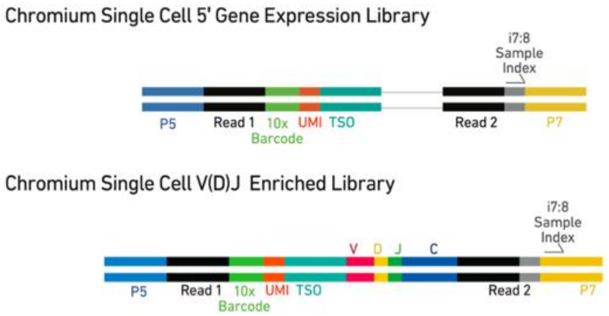


Figure 3. Chromium Single Cell 5' Gene Expression (5'GEX) and V(D)J (TCR-enriched) library.

3. Data Analysis and Assay Performance Assessment.

The Cell Ranger pipeline for immune profiling developed by 10x Genomics (see: https://support.10xgenomics.com/single-cell-vdj/software/pipelines/3.1/what-is-cell-ranger was used to processed all raw sequencing files (BCL data).

Main steps of cell ranger pipeline (v3.1.0) relevant to Immune Profiling data analysis are as indicated (Figure 4):

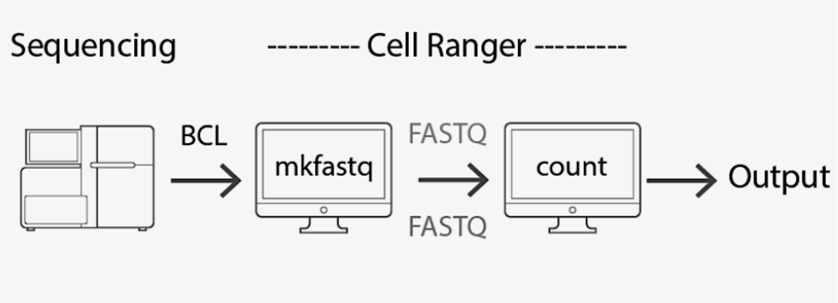


Figure 4. Data analysis processing steps using cell ranger pipeline.

(1) **cellranger mkfastq** demultiplexes raw base call (BCL) files generated by Illumina sequencers into FASTQ files. It is a wrapper around Illumina's bcl2fastq, with additional useful features that are specific to 10x libraries and a simplified sample sheet format.

(2) **cellranger vdj** takes FASTQ files from cellranger mkfastq for V(D)J libraries and performs sequence assembly and paired clonotype calling. It uses the Chromium cellular barcodes and UMIs to assemble V(D)J transcripts per cell. Clonotypes and CDR3 sequences are output as a .vloupe file which can be loaded into Loupe V(D)J Browser.

(3) **cellranger count** takes FASTQ files for 5' Gene Expression libraries and performs alignment, filtering, barcode counting, and UMI counting. It uses the Chromium cellular barcodes to generate feature-barcode matrices, determine clusters, and perform gene expression analysis. Among other, the cellranger count pipeline outputs a .cloupe file which can be loaded into Loupe Browser for an interactive visualization, clustering, and differential expression analysis. The .cloupe files output by cellranger count and the .vloupe files output by cellranger vdj can be overlaid for an integrated analysis with Loupe Browser and Loupe V(D)J Browser.
The cellranger count pipeline also generates the filtered counts matrices which can be used for comprehensive downstream data analysis in R.

For each sample a web summary file was generated with the QC metrics of pre-aggregated data which are **the initial point of reference for determining sample performance and sensitivity** of the scRNA-seq and scTCR-seq assays. The pre-aggregated data QC parameters for each sample are summarized in Table 3-4, along with a reference PBMC sample (fresh PBMC - 10K retrieved from https://support.10xgenomics.com/single-cell-vdj/datasets/3.1.0/vdj_nextgem_hs_pbmc3, PBMCs of a Healthy Donor (Next GEM v1.1), 10x Genomics) with indicated reference ranges of each parameter.

Table 3. QC metrics of pre-aggregated data 5'GEX libraries web summary files.

	Reference PBMC3-10K (fresh)*	**Sample 1**	**Sample 2**	**Sample 3**	**Sample 4**	**Sample 5**	**Sample 6**
Pipeline Version	3.1.0	3.1.0	3.1.0	3.1.0	3.1.0	3.1.0	3.1.0
CELLS PARAMETERS							
Estimated number of cells	11,715 (at least 65% recovery of targeted range 100 – 10,000)	3,376	3,343	4,456	4,260	2,728	3,410
Mean reads per cell	49,285	71,928	71,232	52,459	57,437	89,567	73,597
Median genes per cell	1,404	4,590	2,303	3,966	1,302	2,282	1,916
Total Genes Detected	21,413	20,450	17,270	21,098	20,879	17,397	21,207
Median UMI Counts per Cell	4,077	23,060	6,319	16,883	2,971	6,278	4,576
SEQUENCING PARAMETERS							
Number of reads	577,375,360 (depends on the targeted number of cells)	242,829,736	238,127,886	233,755,850	244,682,036	244,338,298	250,966,861
Valid barcodes	90.7%	93.2%	92.0%	94.4%	94.0%	92.8%	94.2%
Valid UMIs	100.0% (>75%)	99.9%	99.8%	99.9%	100.0%	99.9%	99.9%
Sequencing Saturation	84.7%	40.5%	80.2%	37.1%	41.9%	84.5%	45.9%
Q30 Bases in Barcode	94.1%	97.6%	97.6%	97.5%	97.6%	97.6%	97.6%

Q30 Bases in RNA Read	90.0%	93.3%	93.4%	93.2%	92.9%	93.0%	93.1%
Q30 Bases in Sample Index	91.3%	96.1%	94.4%	95.3%	95.3%	94.0%	92.6%
Q30 Bases in UMI	93.8%	97.5%	97.5%	97.4%	97.5%	97.5%	97.5%
MAPPING							
Transcriptome	GRCh38	hg19	hg19	hg19	hg19	hg19	hg19
Reads Mapped to Genome	91.5%	92.0%	92.9%	91.7%	90.5%	91.1%	91.2%
Reads Mapped Confidently to Genome	82.8%	82.9%	80.9%	84.0%	82.4%	84.8%	83.3%
Reads Mapped Confidently to Intergenic Regions	4.5%	5.1%	5.2%	5.5%	8.0%	5.1%	8.3%
Reads Mapped Confidently to Intronic Regions	7.0%	6.1%	8.3%	6.2%	10.3%	7.9%	10.2%
Reads Mapped Confidently to Exonic Regions	71.4%	71.8%	67.4%	72.3%	64.2%	71.7%	64.8%
Reads Mapped Confidently to Transcriptome	64.0%	65.5%	61.0%	65.4%	58.6%	65.5%	59.2%
Reads Mapped Antisense to Gene	3.9%	2.7%	3.1%	3.4%	2.4%	2.9%	2.3%

Legend: * PBMCs of a Healthy Donor (Next GEM v1.1): https://support.10xgenomics.com/single-cell-vdj/datasets/3.1.0/vdj_nextgem_hs_pbmc3

The TCR QC metrics analysis of pre-aggregated data, as observed in 5'GEX library, showed very good sensitivity based on the detected number of cells when compared 100% TIL samples to 10% spiked of TIL in cancer line samples (see Table 4: Estimated number of cells). There was no difference observed based on the different cDNA input volumes (condition A and B) for the TCR enrichment step and obtained results.

Table 4. QC metrics of pre-aggregated data V(D)J (TCR-enriched libraries) web summary files.

	Reference PBMC3-10K (fresh)*	**Sample 2 A/B**	**Sample 3 A/B**	**Sample 5 A/B**	**Sample 6 A/B**
Pipeline Version	3.1.0	3.1.0	3.1.0	3.1.0	3.1.0
CELLS PARAMETERS					
Estimated number of cells	6,648 (at least 65% recovery of targeted range 100 – 10,000)	A: 2,710 B: 2,698	A: 350 B: 346	A: 2,165 B: 2,159	A: 209 B: 209
Mean reads per cell	13,000	A: 38,660 B: 34,109	A: 211,512 B: 214,251	A: 36,560 B: 40,193	A: 370,030 B: 326,236
Mean Used Read Pairs per Cell	8,665	A: 25,271 B: 21,742	A: 22,008 B: 25,019	A: 23,681 B: 25,144	A: 49,136 B: 48,324
Fraction Reads in Cells	84.5%	A: 74.6% B: 74.2%	A: 15.3% B: 16.9%	A: 74.2% B: 73.5%	A: 21.1% B: 23.2%

Number of Cells with Productive V-J Spanning Pair	5,615	A: 911 B: 910	A: 115 B: 116	A: 836 B: 888	A: 78 B: 78
SEQUENCING					
Number of reads	86,425,410	A: 104,768,406 B: 92,026,284	A: 74,029,131 B: 74,130,739	A: 79,151,423 B: 86,776,238	A: 77,336,347 B: 68,183,341
Valid barcodes	92.9%	A: 96.0% B: 95.2%	A: 81.5% B: 81.5%	A: 96.4% B: 95.8%	A: 76.8% B: 76.7%
Q30 Bases in Barcode	97.3%	A: 97.5% B: 97.4%	A: 96.6% B: 96.6%	A: 97.5% B: 97.4%	A: 96.5% B: 96.5%
Q30 Bases in RNA Read 1	95.7%	A: 94.9% B: 94.8%	A: 92.3% B: 92.4%	A: 95.2% B: 95.1%	A: 92.4% B: 92.6%
Q30 Bases in Sample Index	95.2%	A: 94.3% B: 91.4%	A: 94.1% B: 95.2%	A: 91.3% B: 94.1%	A: 95.4% B: 95.6%
Q30 Bases in UMI	97.3%	A: 97.4% B: 97.4%	A: 96.9% B: 96.9%	A: 97.4% B: 97.4%	A: 96.8% B: 96.9%
MAPPING					
Transcriptome	GRCh38	hg19	hg19	hg19	hg19
ENRICHMENT					
Reads Mapped to Any V(D)J Gene	78.5%	A: 89.3% B: 88.5%	A: 16.0% B: 18.2%	A: 90.8% B: 89.7%	A: 24.1% B: 26.6%
Reads Mapped to TRA	23.9%	A: 30.4% B: 32.8%	A: 4.7% B: 5.4%	A: 37.4% B: 39.0%	A: 9.6% B: 10.5%
Reads Mapped to TRB	54.4%	A: 58.9% B: 55.7%	A: 11.3% B: 12.8%	A: 53.5% B: 50.7%	A: 14.5% B: 16.1%
V(D)J EXPRESSION					
Median TRA UMIs per Cell	4.00	A: 10.00 B: 11.00	A: 8.50 B: 8.50	A: 12.00 B: 12.00	A: 11.00 B: 11.00
Median TRB UMIs per Cell	10.00	A: 27.00 B: 28.00	A: 22.00 B: 23.00	A: 20.00 B: 21.00	A: 20.00 B: 21.00
V(D)J ANNOTATION					
Cells with Productive V-J Spanning Pair	84.5%	A: 33.6% B: 33.7%	A: 32.9% B: 33.5%	A: 38.6% B: 41.1%	A: 37.3% B: 37.3%
Cells with Productive V-J Spanning (TRA, TRB) Pair	84.5%	A: 33.6% B: 33.7%	A: 32.9% B: 33.5%	A: 38.6% B: 41.1%	A: 37.3% B: 37.3%
Paired Clonotype Diversity	3,424.38	A: 8.00 B: 8.04	A: 7.61 B: 7.74	A: 12.81 B: 14.54	A: 10.83 B: 10.83
Cells with TRA Contig	91.4%	A: 99.0% B: 99.0%	A: 98.9% B: 98.8%	A: 98.4% B: 98.2%	A: 97.1% B: 97.1%
Cells with TRB Contig	98.8%	A: 98.2% B: 98.4%	A: 98.0% B: 98.0%	A: 96.1% B: 95.6%	A: 96.2% B: 96.7%
Cells with CDR3-annotated TRA Contig	89.0%	A: 98.9% B: 98.8%	A: 98.3% B: 98.6%	A: 98.0% B: 97.7%	A: 97.1% B: 97.1%
Cells with CDR3-annotated TRB Contig	98.2%	A: 97.2% B: 97.5%	A: 96.9% B: 97.1%	A: 94.5% B: 94.1%	A: 95.7% B: 96.2%
Cells with V-J Spanning TRA Contig	90.8%	A: 98.8% B: 98.8%	A: 98.9% B: 98.8%	A: 97.7% B: 97.8%	A: 96.7% B: 97.1%

Cells with V-J Spanning TRB Contig	98.3%	A: 37.1% B: 37.8%	A: 36.9% B: 37.3%	A: 50.2% B: 52.4%	A: 47.8% B: 47.8%
Cells with Productive TRA Contig	86.8%	A: 98.5% B: 98.1%	A: 98.3% B: 98.6%	A: 97.0% B: 97.0%	A: 95.7% B: 95.7%
Cells with Productive TRB Contig	97.7%	A: 35.2% B: 35.7%	A: 34.6% B: 35.0%	A: 41.7% B: 44.1%	A: 41.6% B: 41.6%

Legend: * PBMCs of a Healthy Donor (Next GEM v1.1): https://support.10xgenomics.com/single-cell-vdj/datasets/3.1.0/vdj_nextgem_hs_pbmc3
A - indicates a replicate of the sample with 2uL cDNA as an input for TCR enrichment reaction
B - indicates a replicate of the sample with 5uL cDNA as an input for TCR enrichment reaction

In addition, the **barcode rank plot** (knee plot) is another way of assessing the assay performance. The plot shows the distribution of barcode counts and which barcodes were inferred to be associated with cells. The y-axis is the number of UMI counts mapped to each barcode and the x-axis is the number of barcodes below that value. A steep drop-off is indicative of good separation between the cell-associated barcodes and the barcodes associated with empty partitions. Since barcodes can be associated with cells based on their UMI count or by their RNA profiles, some regions of the graph can contain both cell-associated and background-associated barcodes. The color of the graph represents the local density of barcodes that are cell-associated. The barcode rank plots for each sample are summarized in Figure 5.

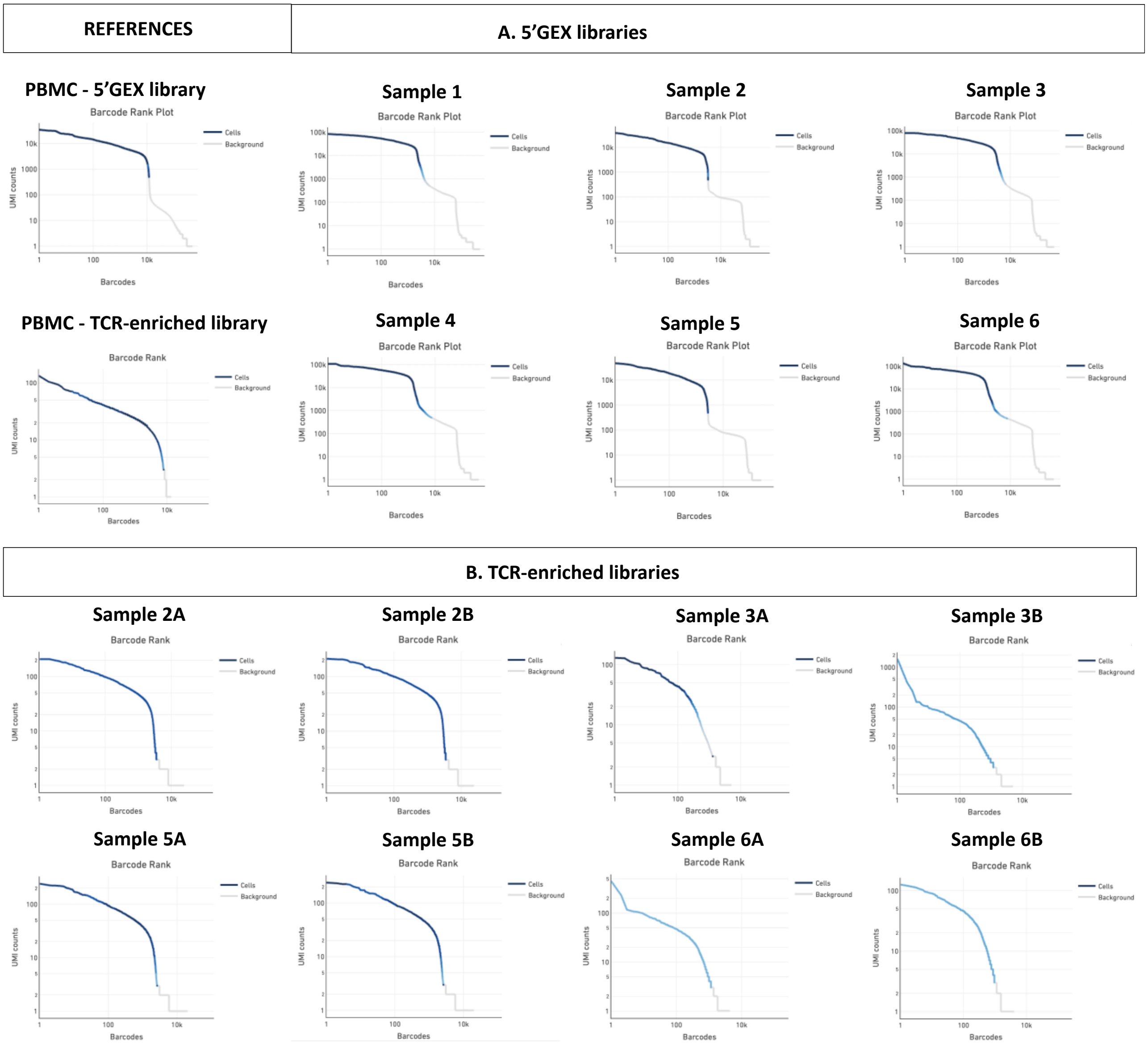


Figure 5. QC Plot of pre-aggregated data – Barcode Rank plot for 5’GEX libraries (A; Samples 1-6) and for TCR enriched libraries (B; Samples 2,3,5,6 with corresponding conditions A and B described in the text). PBMC data (10x Genomics) was used as a reference.

The Loupe VDJ Browser (v.3.0.0) were used for initial TCRseq downstream analysis and visualization of top overlapping clonotypes.

Downstream scRNA-seq analysis in R

The downstream analysis, including the quality control steps, normalization, downstream analysis, and visualization, were performed in R (version 4.2.1) [10] using Seurat (v3.1.0) [11] R package. The cells with low reads, features (genes) below 200 and 20% mitochondrial fraction were removed (Figure 6). The data were normalized using default “LogNormalize” method and corrected for the presence of estimated 2% doublets rate (Figure 7) using DoubletFinder [12]. Top variable 2,000 features were selected to identify integrating features. Principal component analysis was performed on scaled, log-transformed, library-size-normalized UMI matrices using variable gene sets. Dimensionality reduction and visualization were also performed with the UMAP (Uniform Manifold Approximation and Projection) . Graph-based clustering was performed to identify clusters using the first 20 principal components. Top transcripts per cluster were identified using a Wilcoxon rank sum test. Significantly differentially expressed transcripts were selected using adjusted $P < 0.05$ (Benjamini–Hochberg method) and were considered while annotating the clusters (Figure 8).

Downstream scTCR-seq analysis in R

scRepertoire [13], an R package designated to process data derived from the 10x Genomics Chromium Immune Profiling for both T-cell receptor (TCR) and immunoglobulin (Ig) enrichment workflows, was applied to perform comprehensive scTCR-seq analysis and visualization including combining scRNA-seq and immune profiling data.

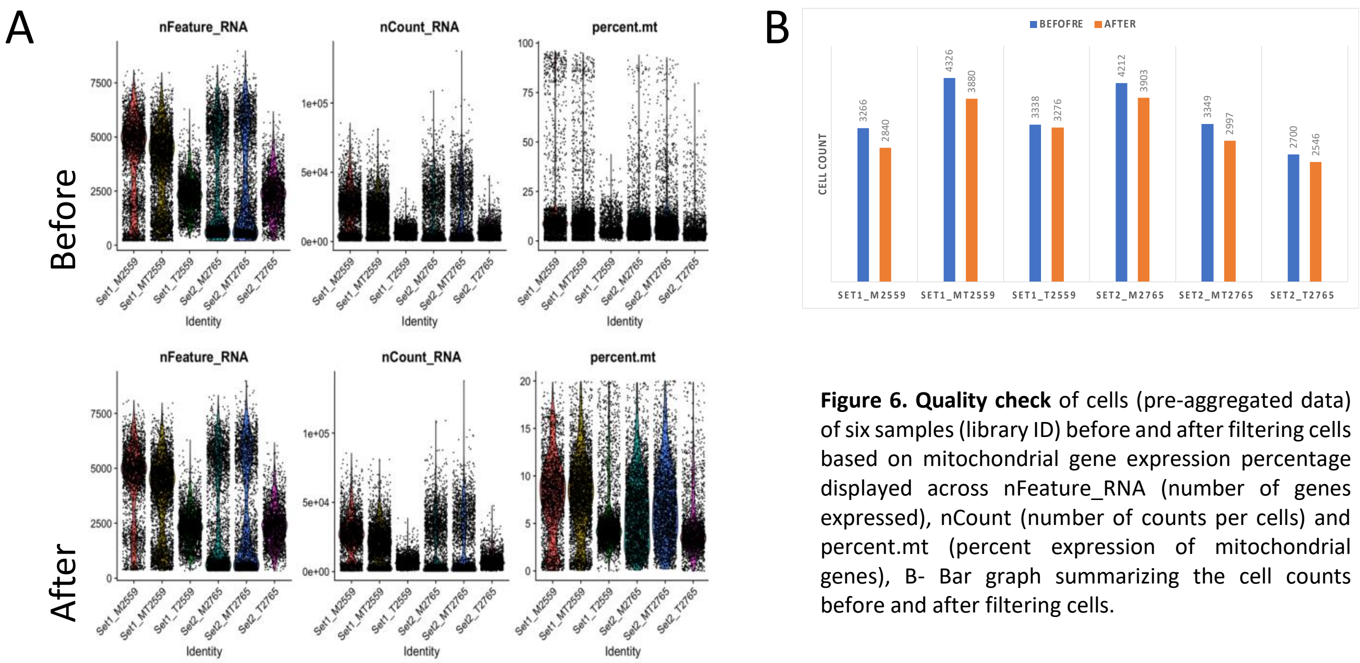
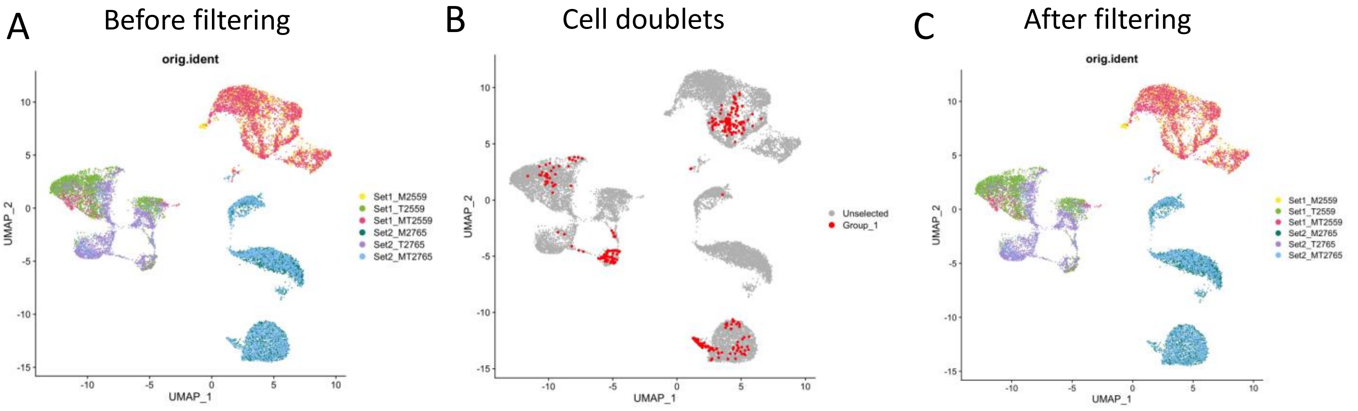


Figure 6. Quality check of cells (pre-aggregated data) of six samples (library ID) before and after filtering cells based on mitochondrial gene expression percentage displayed across nFeature_RNA (number of genes expressed), nCount (number of counts per cells) and percent.mt (percent expression of mitochondrial genes), B- Bar graph summarizing the cell counts before and after filtering cells.

Before: 20775 features across 21191 cells
After: 20775 features across 19442 cells



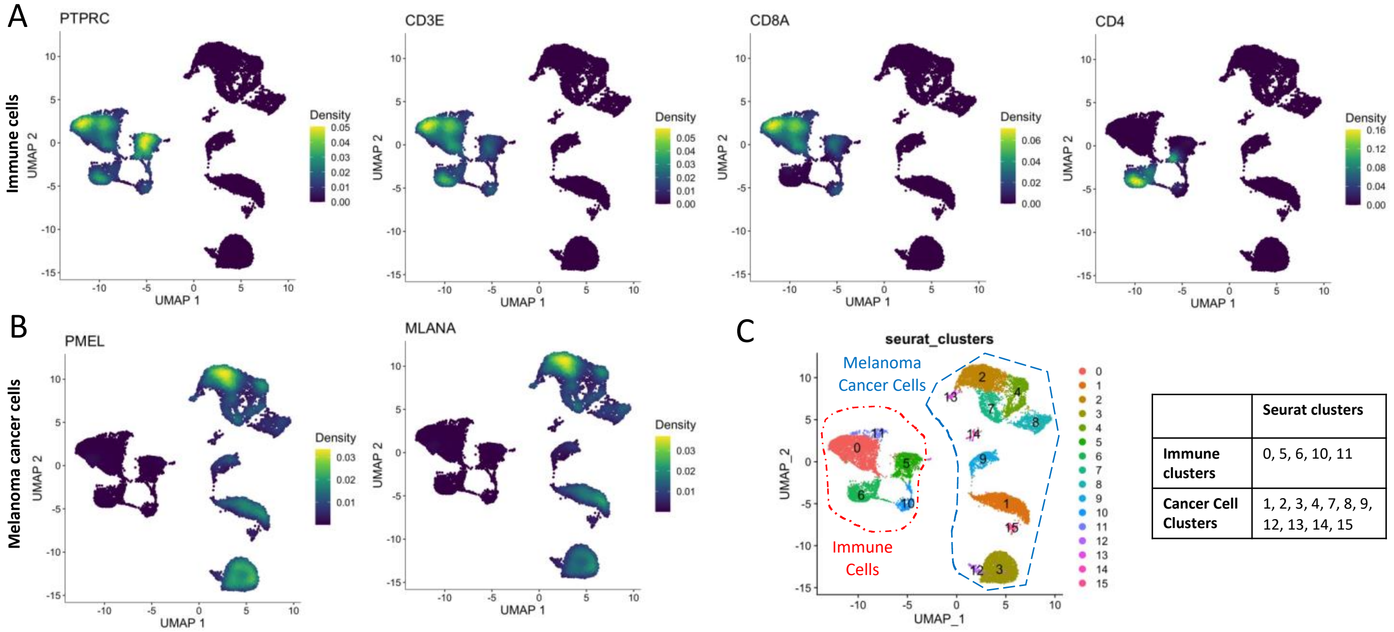
D

Set #	Set 1			Set 2		
Sample Name	**M2559**	**MT2559**	**T2559**	**M2765**	**MT2765**	**T2765**
Before	2840	3880	3276	3903	2997	2546
After	2796	3814	3223	3840	2950	2503
Cells removed	**44 (1.5%)**	**66 (1.7%)**	**53 (1.6%)**	**63 (1.6%)**	**47 (1.6%)**	**43 (1.7%)**

Figure 7. Cell Doublets removal. UMAP clustering of cells (aggregated data) of six samples (library ID) (A) before and (C) after removing predicted cell doublets, marked in red (B). Summarized counts are indicated in table (D).

The precision of intra- and inter- assay of pooled samples (post-aggregation) across different batches (set 1 and set 2), was assessed within further analysis. We applied Uniform Manifold Approximation and Projection (UMAP) method [14]. The final, post-filtering UMAP clustering based on the samples name (library ID) is displayed on Figure 7C which indicates separate clustering for TIL (immune cells) from both sets when comparing to melanoma cancer cells, and additional separation within two sets of melanoma cancer cells.

Additionally, UMAP density displaying gene expression of key immune markers *(PTPRC, CD3E, CD8A, CD4* genes; Figure 8A) and key melanoma cancer cells markers (*PMEL, MLANA*; Figure 8B) highlights the main division of analyzed population of aggregated cells.



	Seurat clusters
Immune clusters	0, 5, 6, 10, 11
Cancer Cell Clusters	1, 2, 3, 4, 7, 8, 9, 12, 13, 14, 15

Figure 8. Identification of Immune and Cancer Cell populations. A - UMAP density plot of aggregated six samples based across immune markers (*PTPRC, CD3E, CD8A, CD4* genes), B – UMAP density plot of aggregated six samples based on sample name (library ID) across melanoma related markers (PMEL, MLANA genes), C- UMAP plot of aggregated six samples based on identified Seurat clusters marked 0-15. Seurat clusters with immune population are marked with red dashed line, and for melanoma cancer cells marked with blue dashed line, summary of related clusters are presented in adherent table.

We performed graph-based clustering (Figure 8C), an algorithm which seeks to find highly connected "modules" in the graph and identified 16 (0-15) seurat clusters which helped dividing the cells to immune and melanoma cancer cell populations. Based on the distribution of the clusters especially within spiked in TIL samples in cancer cells, we identified a similar clusters' pattern across immune population (Figure 9).

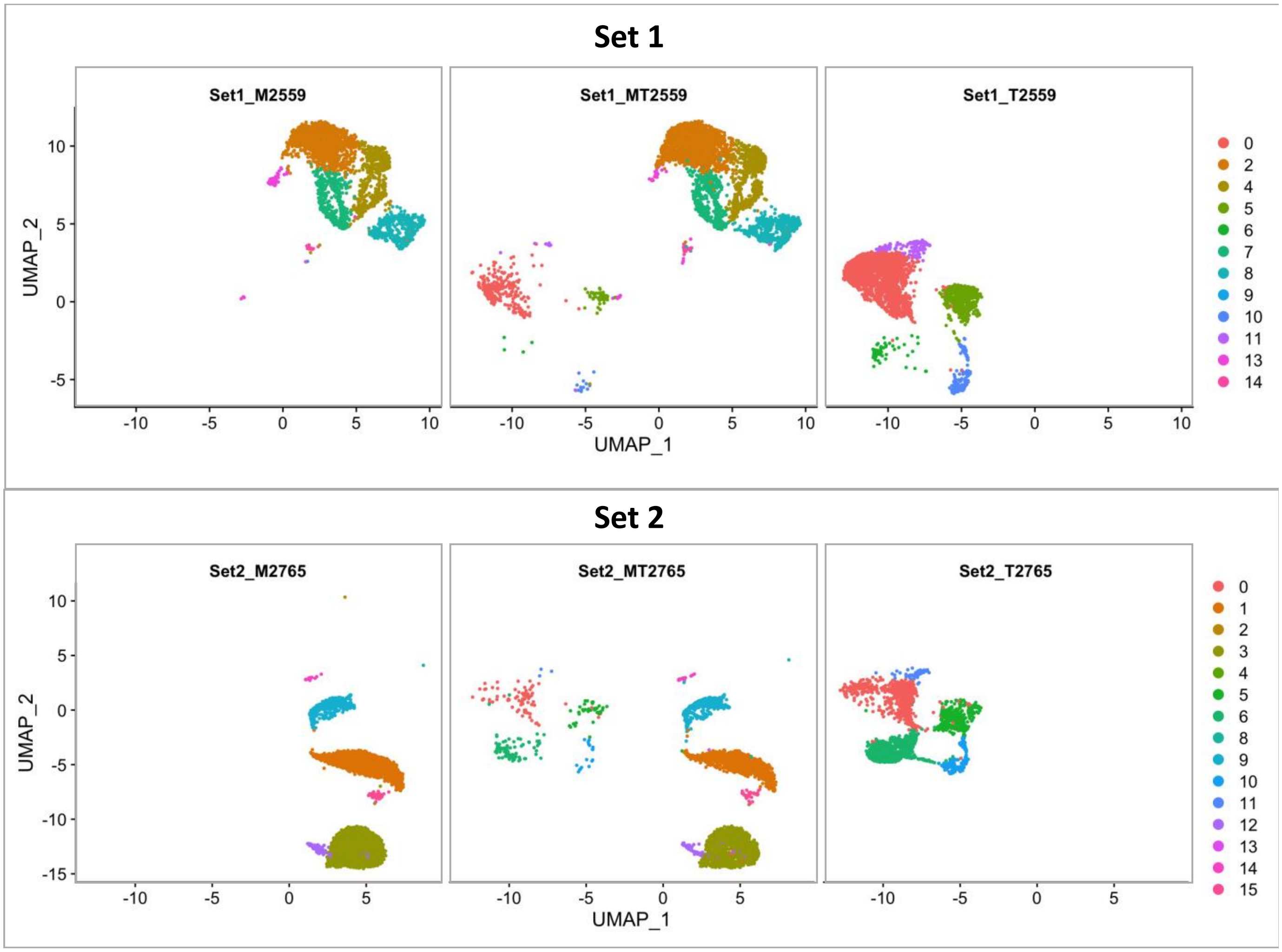


Figure 9. Similarity of Immune and Cancer Cell populations across obtained clusters. UMAP plot of aggregated six samples Set 1 (A) and Set 2 (B) with a split view, cells are colored based on the identified clusters.

The elevated gene expression of selected genes associated with analyzed samples additionally confirmed differential pattern among analyzed marked populations (Figure 10) and the reproducibility of the results.

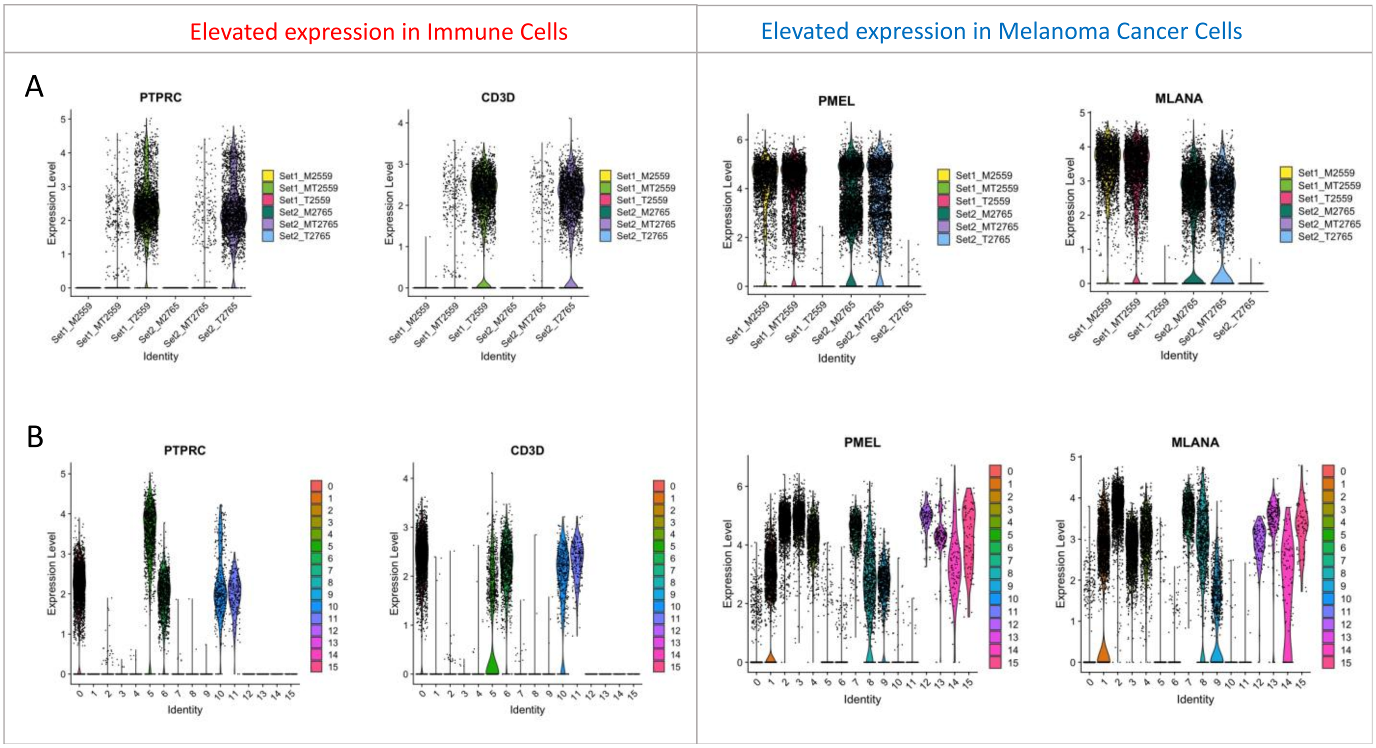
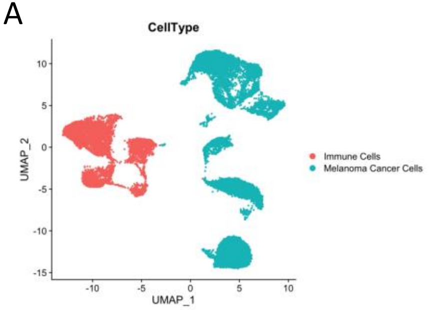


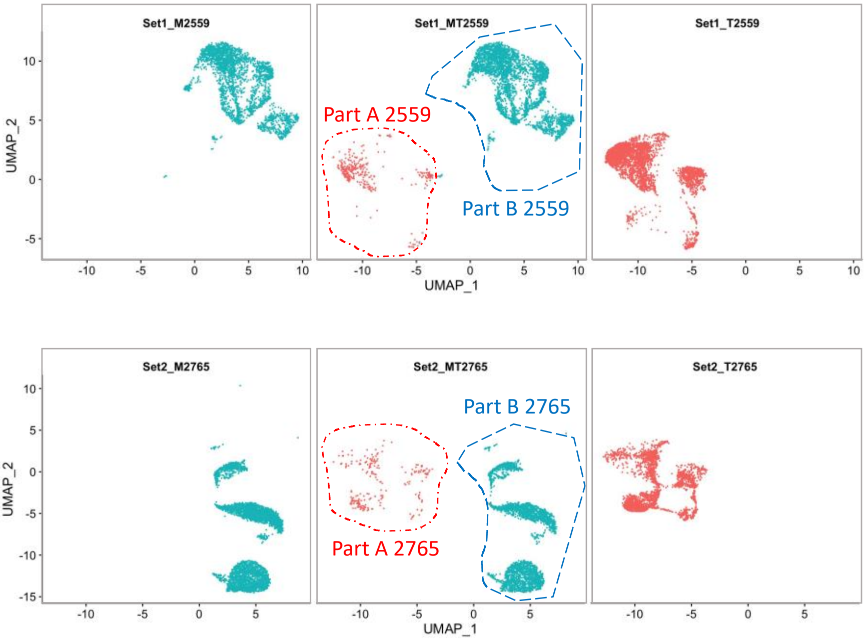
Figure 10. Differential gene pattern across Immune and melanoma cancer cell clusters. Violin plots displaying gene expression of selected characteristic immune genes (*PTPRC, CD3D*) and melanoma cancer cell genes (*PMEL, MLANA*); data displayed by library ID (A) and graph-based generated clusters (B).

Next, to see profiles of each replicate within each set, we split cells of the mixed sample into two separate parts (part A – corresponding to immune cells and part B – corresponding to melanoma cancer cells, see Figure 11). The separate, split view clearly shows the immune population of TIL within the mixed samples in each set of melanoma cancer cells with spiked TIL. The immune cells percentage is indicated in the adherent table of figure 11C).

A



B

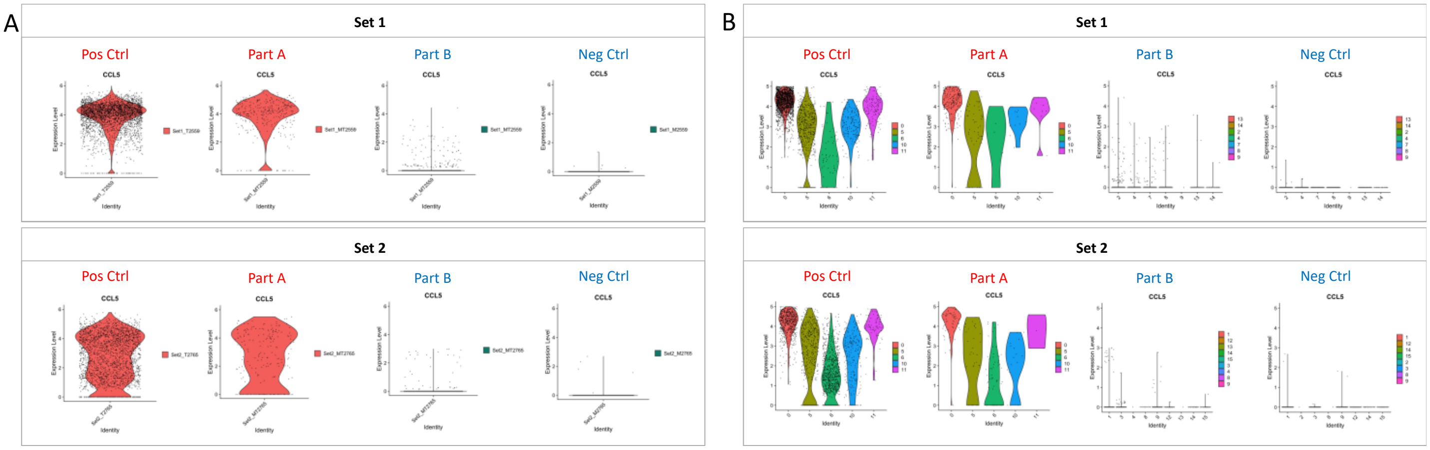


C

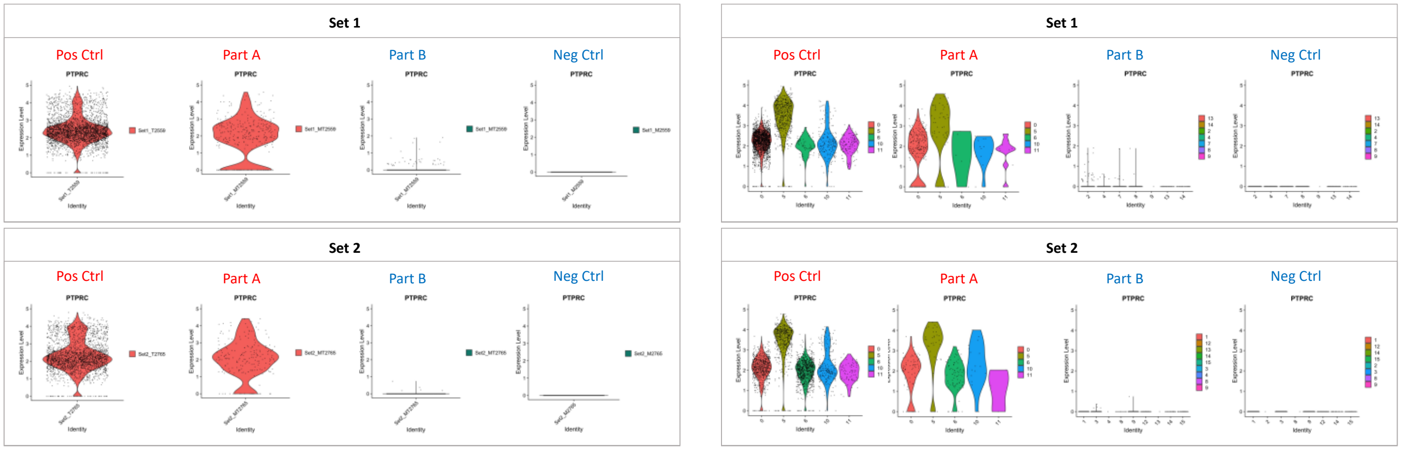
Set #	Sample Name	Cell #	Immune Cell (PART A)	Melanoma Cancer Cells (PART B)
Set1	MT2559	3814	337 (8.8%)	3477 (91.2%)
Set2	MT2765	2950	214 (7.25%)	2736 (92.75%)

Figure 11. Identification of Immune cell contribution across spiked in samples. A - UMAP plot of aggregated six samples marked based on the Immune and Melanoma cancer cells annotation; B - UMAP plot of aggregated six samples with a split view per sample, cells are colored based on annotation as in A with marked Part A (Immune cells) and Part B (Melanoma cancer cells); C - Summarized counts contribution for each set.

The examples of top differential genes of selected Seurat clusters across samples before and after splitting are shown on Figure 12.



Cluster 0: CCL5 (Avg_log2FC = 4.08, P_val_adj = 0.000000e+00)



Cluster 5: PTPRC (Avg_log2FC = 3.5, P_val_adj = 0.000000e+00)

Figure 12. Reproducibility of gene expression patterns. Violin plots of selected top significant genes of exemplary clusters (*CCL5* – cluster 0, *PTPRC* – cluster 5); A- data displayed by library ID of samples (left panel) with split version of spiked in samples as Part A (Immune cells) and Part B (Melanoma cancer cells), with positive and Negative controls (Pos Ctrl, Neg Ctrl) corresponding to individual samples. ; B- data displayed by Seurat clusters (right panel) with split version of spiked in samples as Part A (Immune cells) and Part B (Melanoma cancer cells), with positive and Negative controls (Pos Ctrl, Neg Ctrl) corresponding to individual samples.

Additionally, the TCR analysis confirmed the presence of a good separation of the Immune population when comparing 100% TIL samples (positive control) to spiked in cancer samples with 10% of TIL (see Table 4, Figure 13, 14).

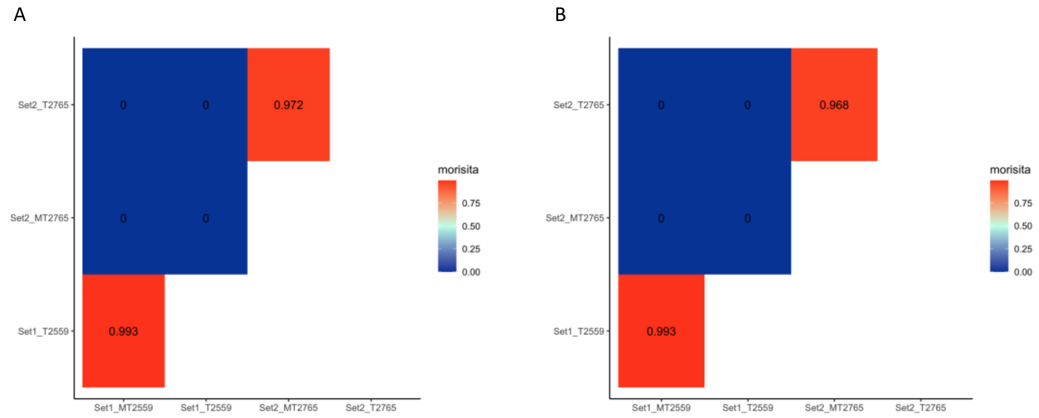


Figure 13. Reproducibility of scTCR-seq patterns across analyzed samples (A- 2ul input of cDNA, B- 5ul input of cDNA) – Morisita overlap index**.**

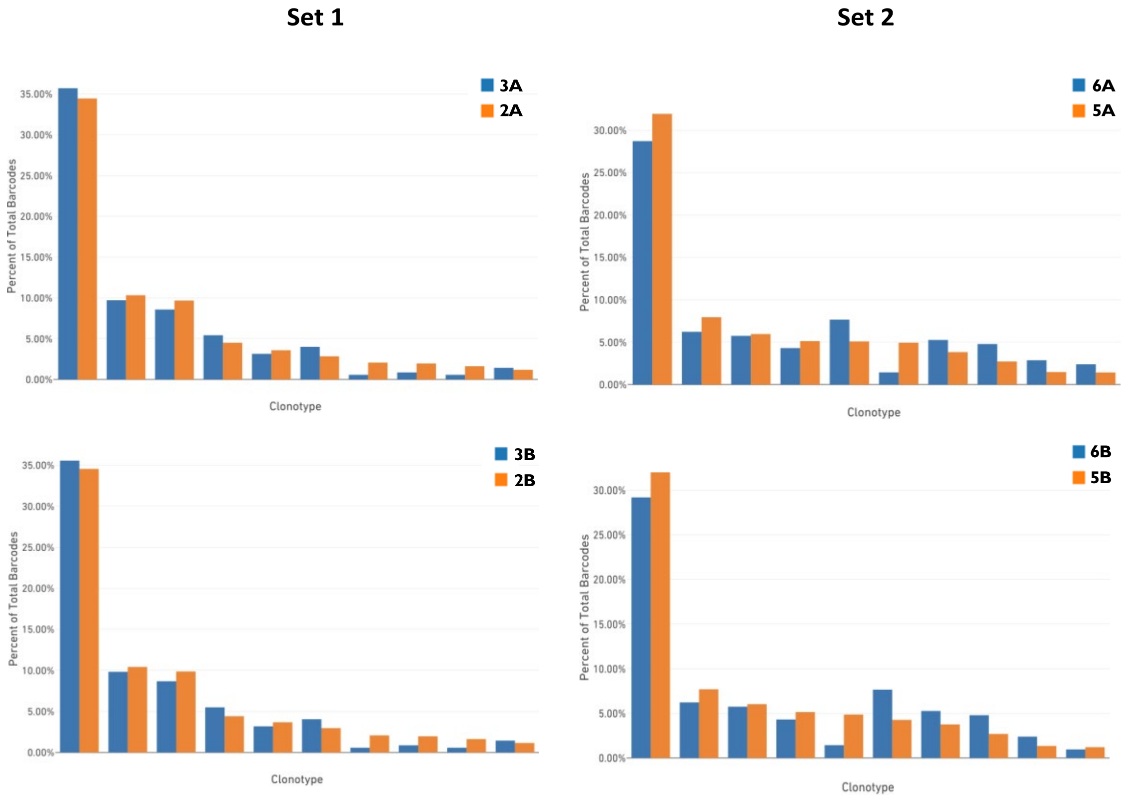


Figure 14. TOP 10 clonotypes across analyzed samples of Set 1 and Set 2 (#2559 and #2765, respectively) - V(D)J loupe browser analysis. Samples 2 and 5 correspond to 100% TIL, samples 3 and 6 corresponds to spiked in cancer samples with 10% of TIL. Condition A - indicates a replicate of the sample with 2uL cDNA as an input for TCR enrichment reaction; condition B - indicates a replicate of the sample with 5uL cDNA as an input for TCR enrichment reaction. Percent of total barcodes is indicated (y-axis).

The precision of intra- and inter- TCR assay and biological reproducibility was additionally confirmed by overlay of the scTCR-seq with the scRNA-seq gene expression data overlapping number and type of shared expanded clonotypes (Table 5, Figure 15, 16).

Table. 5. Summary of cell distribution across different categories of T cell expansion of cells with TCRA+TCRB chains as plotted in Figure 16.

	Hyperexpanded (100 < X <= 500)	Large (20 < X <= 100)	Medium (5 < X <= 20)	Small (1 < X <= 5)	Single (0 < X <= 1)
	3402				
	1501	736	538	400	227
Set 1	826	417	303	207	112
Set1_T2559	749	378	263	187	102
Set1_MT2559	77	39	40	20	10
Set1_M2559	0	0	0	0	0
Set 2	675	319	235	193	115
Set2_T2765	624	281	211	178	106
Set2_MT2765	51	38	24	15	9
Set2_M2765	0	0	0	0	0

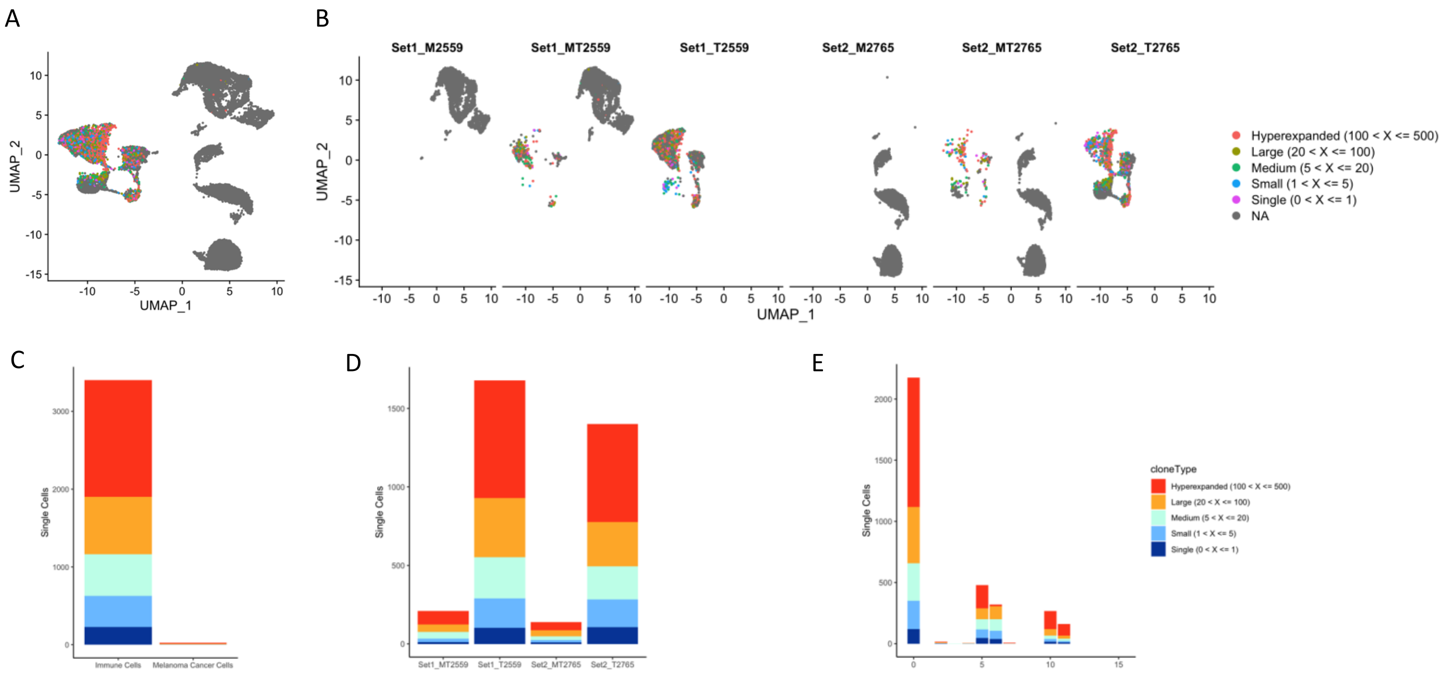


Figure 15. Clone type distribution across analyzed cells displayed as A - aggregated UMAP plot of 6 samples and B – aggregated UMAP plot in split view for individual samples, C - bar graph plotted across immune cells and melanoma cancer cells, D – Individual samples, E – Seurat clusters. Cells representing different expansion level: Hyperexpanded (100 < X <= 500), Large (20 < X <= 100), Medium (5 < X <= 20), Small (1 < X <= 5), Single (0 < X <= 1) are color coded.

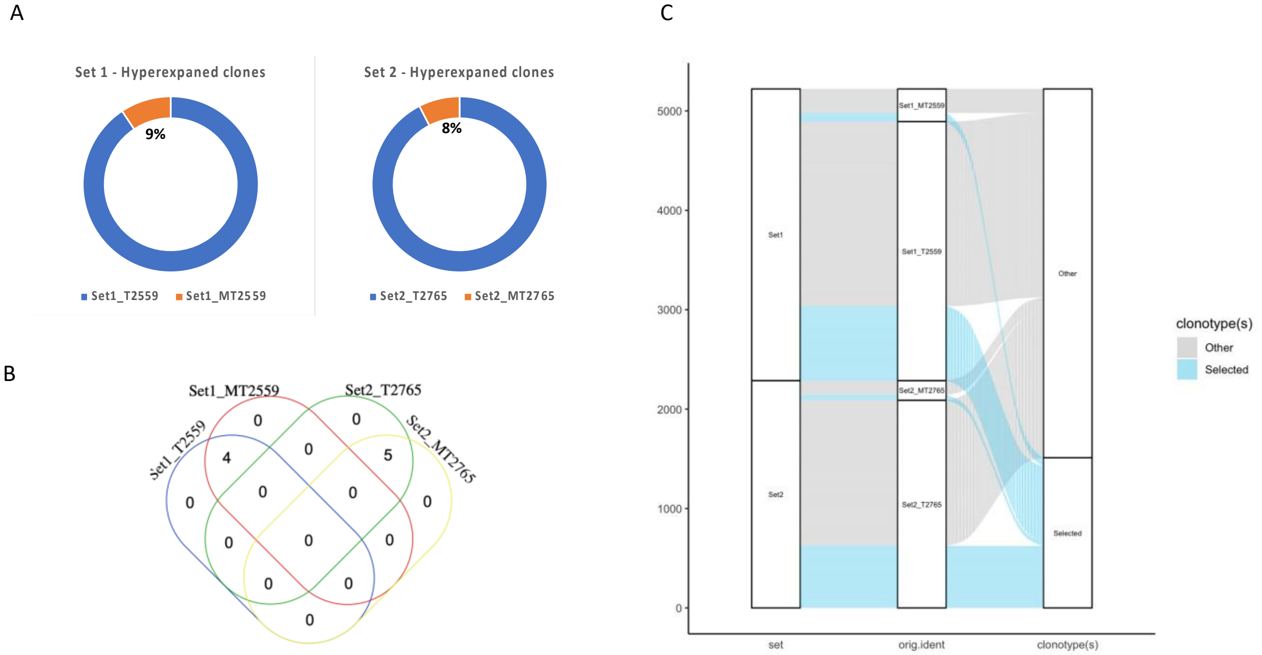


Figure 16. Distribution of Hyperexpanded Clones. A- Pie chart representing hyperexpanded clones for set 1 and set 2 across analyzed samples (see table 5); B – Venn Diagram displaying overlapped unique clone types identified across analyzed samples within the hyperexpanded clone population; C- alluvialClonotypes overlay plot of 9 unique clonotypes across set 1 and set 2.

Appendix:

- CG000053_CellPrepGuide_RevC,
- CG000207_ChromiumNextGEMSingleCellV_D_J_ReagentKits_v1.1_UG_RevE
- Cimac_specimen_collection_umbrella (Master protocol)

References:

1. Wang, Y. and N.E. Navin, *Advances and Applications of Single-Cell Sequencing Technologies.* Molecular Cell, 2015. **58**(4): p. 598-609.
2. Kulkarni, A., et al., *Beyond bulk: a review of single cell transcriptomics methodologies and applications.* Current Opinion in Biotechnology, 2019. **58**: p. 129-136.
3. Suva, M.L. and I. Tirosh, *Single-Cell RNA Sequencing in Cancer: Lessons Learned and Emerging Challenges.* Molecular Cell, 2019. **75**(1): p. 7-12.
4. Luquette, L.J., et al., *Identification of somatic mutations in single cell DNA-seq using a spatial model of allelic imbalance.* Nature Communications, 2019. **10**.
5. Stoeckius, M., et al., *Simultaneous epitope and transcriptome measurement in single cells.* Nature Methods, 2017. **14**(9): p. 865-+.
6. Stuart, T. and R. Satija, *Integrative single-cell analysis.* Nature Reviews Genetics, 2019. **20**(5): p. 257-272.
7. Granja, J.M., et al., *Single-cell multiomic analysis identifies regulatory programs in mixed-phenotype acute leukemia.* Nature Biotechnology, 2019. **37**(12): p. 1458-+.
8. Satpathy, A.T., et al., *Massively parallel single-cell chromatin landscapes of human immune cell development and intratumoral T cell exhaustion.* Nature Biotechnology, 2019. **37**(8): p. 925-+.
9. Forget, M.A., et al., *Prospective Analysis of Adoptive TIL Therapy in Patients with Metastatic Melanoma: Response, Impact of Anti-CTLA4, and Biomarkers to Predict Clinical Outcome.* Clinical Cancer Research, 2018. **24**(18): p. 4416-4428.
10. R Core Team, *R: A language and environment for statistical computing. R Foundation for Statistical Computing,* Vienna, Austria. 2021. URL https://www.R-project.org/.
11. Hao Y, Hao S, et al. I*ntegrated analysis of multimodal single-cell data*. Cell. 2021 Vol 184 | 13, P3573-3587.E29
12. McGinnis CS ety al D*oubletFinder: Doublet Detection in Single-Cell RNA Sequencing Data Using Artificial Nearest Neighbors.* Cell Stst. 2019 Apr 24;8(4):329-337.
13. Borcherding N. *scRepertoire: A toolkit for single-cell immune receptor profiling.* Version 2. F1000Res. 2020; 9:47.
14. Becht, E., et al., *Dimensionality reduction for visualizing single-cell data using UMAP.* Nature Biotechnology, 2019. **37**(1): p. 38-+.