nature genetics

Letter

https://doi.org/10.1038/s41588-022-01268-9

Multimodal single-cell and whole-genome sequencing of small, frozen clinical specimens

In the format provided by the authors and unedited

Supplementary information:

Multi-modal single-cell and whole-genome sequencing of small, frozen clinical specimens

Yiping Wang*, Joy Linyue Fan*, Johannes C. Melms* *et al.*

Spatial transcriptomics library preparation and sequencing

Spatial transcriptomics libraries were prepared using *Slide-seq V2* reagents[1]. The oligonucleotides and reagents used for library generation are listed in **Supplementary Table 5**. Frozen tissue sections were cut at 10 µm thickness using a Leica CM1950 cryostat and placed on a *Slide-seq V2* puck (mounted on glass) with mapped bead coordinates. The tissue was then thawed briefly to fully adhere to the puck, and the puck was placed in a 1.5 ml tube with hybridization buffer (6x SSC solution with RNAse out) and incubated at room temperature for 15 min to allow binding of RNA to capture oligos. For first-strand synthesis, the puck was then removed from the hybridization buffer, dipped once in 1x Maxima RT buffer to wash of excess hybridization buffer and placed in a fresh tube with reverse transcription mix (115 µl water, 40 µl Maxima 5x RT buffer, 20 µl dNTPs (10mM), 5µL RNase Inhibitor, 10 µl template switch oligo (50 µM), 10 µl Maxima H Minus reverse transcriptase). The reaction was then incubated for 30 min at room temperature followed by a 90 min incubation at 52 °C. Thereafter, the tissue was digested for 30 min at 37°C by adding 200 µl 2x tissue digestion buffer (200 mM Tris-Cl pH8, 400 mM NaCl, 4% SDS, 10 mM EDTA, and 32 U ml^{-1} proteinase K). After incubation the solution was pipetted 20x to dissociate the individual beads of the puck and 200 µl wash buffer (10 mM Tris pH 8.0, 1 mM EDTA and 0.01% Tween-20) were added. The beads were collected by centrifugation for 2 min at 3000 x g, supernatant was discarded and the beads were washed once more with 200 µL wash buffer followed by 200 µL 10 mM Tris-HCL. After the final wash and centrifugation, beads were resuspended in 200 µl ExoI mix (170 µl water, 20 µl ExoI buffer and 10 µl ExoI) and incubated at 37 °C for 30 min. After ExoI incubation the beads were centrifuged for 2 min at 3000 x g and resuspended in wash buffer. This step was repeated three times. After the final wash the beads were resuspended in 200 µl 0.1N NaOH and incubated for 5 min at room temperature. After the incubation, 200 µL wash buffer were added and the beads were pelleted for 2 min at

3000 x g. After centrifugation, the beads were resuspended in 200 µl TE buffer and pelleted again for 2 min at 3000 x g. After removing as much TE buffer as possible, the beads were resuspended in second strand synthesis buffer (133 µl water, 40 µl Maxima 5x RT, 20 µl dNTPs (10 mM), 2 µl dN-SMRT oligo (1 mM), and 5 µl Klenow enzyme) and incubated for 1 hour at 37°C. After second strand synthesis, 200 µl wash buffer were added and the beads were centrifuged for 2 min at 3000 x g. This was repeated three times. After the final wash in wash buffer, the beads were washed once more in water and pelleted again. The water was removed, and the beads were resuspended in PCR buffer (25 µl Terra PCR direct buffer, 1 µL TruSeq PCR handle primer (100 µM), 1 µl SMART PCR primer (100 µM), 1 µL Terra polymerase, and 22 µL water). The reaction was mixed 8x by pipetting and incubated with the following PCR program: 98°C for 2 minutes, followed by 4 cycles of 98°C for 20 seconds, 65°C for 45 seconds, 72°C for 3 min, 9 cycles of 98°C for 20 seconds, 67°C for 45 seconds, 72°C for 3 min, and a final extension at 72°C for 5 min followed by a 4°C hold stage. The PCR product was then purified twice using SPRIselect beads at 0.6x ratio according to manufacturer instructions and eluted in 20 µl water. The PCR product was then quantified using Agilent D5000 high sensitivity tapestation reagents. Sequencing libraries were prepared using Nextera XT ((Illumina, FC-131-1096) reagents. 600 pg DNA were carried forward in a 20 µL reaction for tagmentation with Nextera TD buffer (10 µl) and Amplicon tagment enzyme (5 µl) with the remaining volume being filled up with water. Tagmentation was performed in a preheated thermocycler at 55°C for 5 min. After incubation, 5 µl neutralization buffer were added and the reaction was mixed by pipetting and incubated for 5 min. Thereafter the indexing reaction was set up by adding 15 µl Nextera PCR mix, 8 µl water and 1 µl P5 TruSeq PCR oligo (10 µM) and 1 µl Nextera N70x P7 oligo (10µM), The reaction was mixed by pipetting and incubated with the following PCR program: 72°C for 3 min, 98°C for 30 sec, 12 cycles of 95°C for 10 seconds, 55°C for 30 seconds, 72°C for 30 sec, followed by a final extension at 72°C for 5 min and a 4°C hold stage. The indexed PCR product was then purified using SPRIselect beads at 0.6x ratio according to manufacturer instructions and eluted in 10 µl water, quantified using Agilent D5000 high sensitivity Tapestation reagents and sequenced on an Illumina NovaSeq S4 instrument.

Library construction for low pass whole genome sequencing
20-50 ng genomic DNA served as input for low pass whole genome sequencing (lp-WGS) library preparation with the Lotus DNA Library Prep Kit (Integrated DNA Technologies - IDT, Coralville, IA) using xGen™ Stubby Adapter and UDI Primer Pairs (IDT, #10005921) according to manufacturer institutions with the following specifications: In the enzymatic preparation program, samples were held at 32°C for 9 minutes. Adapters were not diluted in the ligation step and the optional PCR (5 cycles) and cleanup step were performed so the stubby adapters could be amplified with primers to incorporate the index sequences and P5 and P7 sequences. After library cleanup with AMPure beads (Beckman Coulter, Brea, CA), the libraries were quantified with Qubit 1X dsDNA HS Assay Kit (ThermoFisher; #Q33231) and Tapestation D5000 HS tapes (Agilent) and stored at -20°C until sequencing.

Processing FASTQ files into gene expression matrices
Demultiplexed FASTQ files from raw single-cell/-nucleus RNA sequencing reads were aligned to the human GRCh38 genome, and gene counts were quantified using Cell Ranger 'count'. Reads mapping to introns and exons were counted in order to increase the number of genes detected per nucleus and number of nuclei passing quality control, leading later on to improved cell-type identification. To include introns during read mapping, we followed the instructions provided by 10X Genomics (https://support.10xgenomics.com/single-cell-gene-expression/software/pipelines/latest/advanced/references), and created a 'pre-mRNA' human GRCh38 reference genome, using the Cell Ranger command 'mkref' with a customized gene transfer format (GTF) file, in which all transcripts were labeled as exons.

Removal of background noise in gene expression matrices
We used the 'remove-background' function of *CellBender* (v0.2.0)[2] on *CellRanger*-generated 'raw_feature_bc_matrix.h5' files, in order to remove technical ambient-RNA counts and empty droplets from the gene expression matrices. The parameter 'expected-cells' was obtained from the *CellRanger* metric 'Estimated Number of Cells', while the parameter 'total-droplets-included' was set to the midpoint of a plateau in the barcode-rank plot, based on visual inspection.

Quality control and filtering

After processing by *CellBender*, expression matrices were processed individually in R (v4.0.2) using *Seurat* (v4.0.1)[3]. For sequentially collected specimens from the patient on anti-PD1 therapy, we applied filters to keep only cells with 300-10000 genes, 400-30000 UMIs, and <10% of mitochondrial reads. For uveal melanoma liver metastasis samples, we kept cells with at least 200 genes and <10% of mitochondrial reads and excluded cells based on high gene (ranging 2,300-10,000) and/or UMI (ranging 4,000-75,000) counts. Finally, for all other datasets, we kept cells with at least 300 genes and <20% of mitochondrial reads. Additionally, Scrublet[4] was used for doublet removal, with an expected doublet rate of 9.6%. Filtered gene-barcode matrices were then normalized with the 'NormalizeData' function using the 'LogNormalize' method. The top 2,000 variable genes were identified using the 'vst' method in the 'FindVariableFeatures' function. Gene expression matrices were scaled and centered using 'ScaleData'. We performed principal component analysis (PCA) as well as uniform manifold approximation and projection (UMAP) dimension reduction using the top 30 principal components. We also used the AddModuleScore function of Seurat to calculate the expression of a stress-associated gene signature (**Supplementary Table 4**).

Analysis of Splicing Fractions

To compare the fractions of spliced and unspliced RNA in the NSCLC samples, we applied *velocyto*[5] (v0.17.17) The pipeline was run directly on outputs from *CellRanger*, and visualized using *scVelo*[6] (v0.2.3)

Identification of Differentially Expressed Genes

We applied *scanpy*'s function for ranking genes to identify differentially expressed genes. For each cluster (defined by K-means clustering), we compare the genes contained within the cluster to genes not contained in the cluster using a t-test with Benjamini-Hochberg correction. Genes are then ranked by scores, and the top 300 genes for each cluster are returned (excluding mitochondrial and ribosomal genes).

Manual cell-type annotation

We integrated and clustered samples within each dataset using *Seurat* and identified differential gene expression (DGE) between clusters using the FindAllMarkers function. We then manually annotated cell type clusters using the following list of marker genes (**Supplementary Table 4**), except for B-cells/Plasma cells, which were annotated based on expression of IG genes. This initial labeling resulted in the identification of major cell types such as melanoma cells, lung cancer cells, hepatocytes, T cells/NK cells, and others (**Supplementary Table 4**). Next, to annotate subtypes of T cells/NK cells only, we created a Seurat object containing only these cells, and reran scaling, PCA, UMAP dimension reduction, clustering and DGE analysis. The resulting clusters were again annotated manually as CD4+ T-cells, CD8+ cells, T-regs and others (**Supplementary Table 4**).

Copy number alteration (CNA) inference

Chromosomal CNA profiles of individual cells were inferred from transcriptional data using *inferCNV*[7] (v1.6.0). For each sample, we used cells that were identified as hematopoietic cells by *SingleR*[8] as a diploid reference to estimate CNAs in test cells. We applied a cutoff of 0.1 for the minimum average read counts per gene among reference cells/nuclei, set the clustering to 'subcluster', denoised the output using the default 'sd_amplifier' of 1.5, and ran Hidden Markov Models (HMM) to predict the CNA level.

TCR data processing and integration

TCR FASTQ files were aligned using *CellRanger* '*vdj*' (v6.1.1; 10X Genomics). We then used the combineTCR function the package scRepertoire to process filtered contig annotation files from the Cell Ranger output. Clonotypes which were annotated as having the same clonotype gene by combineTCR, and occurred in two or more samples within a dataset, were considered as shared within those samples.

Whole exome sequencing analysis of sequential treatment pre-treatment sample

We analyzed whole exome sequencing data using the *control-FREEC* pipeline, using parameters *window*=0, *ploidy*=2,3, *breakPointType*=4, *breakPointThreshold*=1.2, *noisyData*=TRUE, *readCountThreshold*=50, and *minimalCoveragePerPosition*=5[9].

Supplementary Information References

1. Stickels, R. R. *et al.* Highly sensitive spatial transcriptomics at near-cellular resolution with Slide-seqV2. *Nat Biotechnol* **39**, 313–319 (2021).
2. Fleming, S. J., Marioni, J. C. & Babadi, M. CellBender remove-background: a deep generative model for unsupervised removal of background noise from scRNA-seq datasets. *bioRxiv* 791699 (2019) doi:10.1101/791699.
3. Stuart, T. *et al.* Comprehensive Integration of Single-Cell Data. *Cell* **177**, 1888-1902.e21 (2019).
4. Wolock, S. L., Lopez, R. & Klein, A. M. Scrublet: Computational Identification of Cell Doublets in Single-Cell Transcriptomic Data. *Cell Syst* **8**, 281-291.e9 (2019).
5. La Manno, G. *et al.* RNA velocity of single cells. *Nature* **560**, 494–498 (2018).
6. Bergen, V., Lange, M., Peidli, S., Wolf, F. A. & Theis, F. J. Generalizing RNA velocity to transient cell states through dynamical modeling. *Nat Biotechnol* **38**, 1408–1414 (2020).
7. Tirosh, I. *et al.* Dissecting the multicellular ecosystem of metastatic melanoma by single-cell RNA-seq. *Science* **352**, 189–196 (2016).
8. Aran, D. *et al.* Reference-based analysis of lung single-cell sequencing reveals a transitional profibrotic macrophage. *Nat Immunol* **20**, 163–172 (2019).
9. Boeva, V. *et al.* Control-FREEC: a tool for assessing copy number and allelic content using next-generation sequencing data. *Bioinformatics* **28**, 423–425 (2012).