# **Global gene expression of NSCLC TIL**

G. Dilman1, I. Bogdanova1, E. Esaulova2

1 Bioinformatics Institute, Kantemirovskaya st. 2A, 197342, St. Petersburg, Russia

2 BioNTech US

**Introduction**

Programmed cell death protein 1 (PD-1) and its ligand, programmed cell death ligand 1 (PD-L1) play a central role in inhibiting immune responses to tumor cells by reducing T cell activation, proliferation, and cytotoxic activity [1]. Inhibition of PD-1 or PD-L1 has been associated with the restoration of an effective immune response against cancer cells [2]. In particular, such therapy promotes the normal immune response of CD8 T cells, however factors in the tumor microenvironment can inhibit these T-cell reactions. Research in the field of single-cell transcriptomics has revealed global T-cell dysfunction in tumor infiltrating lymphocytes (TIL). However, most TILs do not recognize tumor antigens [3], and little is known about mutation-associated neoantigens (MANA) specific TIL transcription programs. Justina X. Caushi et al. studied the transcription profile of TIL. Their work is described in the article ‘Transcriptional programs of neoantigen-specific TIL in anti-PD-1-treated lung cancers' [4]. One of the objectives of this work was to evaluate the global gene expression of NSCLC TIL in order to establish whether gene expression profiling of total TIL was sensitive enough in distinguishing the pathologic response to PD-1 blockade. The aim of our research work was the reproduction of the study of the transcription profile of tumor-infiltrating lymphocytes (TIL) described in the article.

**Methods**

Cell Ranger v3.1.0 [5] was used by the authors to demultiplex the FASTQ reads, align them to the GRCh38 human transcriptome, and extract their cell and unique molecular identifier (UMI) barcodes. The output of this pipeline is a digital gene expression (DGE) matrix for each sample, which records the number of UMIs for each gene that are associated with each cell barcode.

Preprocessed single-cell RNA-seq data was downloaded from GEO (GSE176021). The count matrix was read into an AnnData object, which holds many slots for annotations and different representations of the data. It also comes with its own HDF5-based file format: .h5ad. Scanpy v1.9.2 [6] was used for analyzing single-cell gene expression.

Six patients with and without major pathologic response were randomly selected. Due to the lack of computing power, each sample was subsampled up to 1500 cells before quality control (QC) and then all samples were combined into a single table for further analysis.

The quality of cells was then assessed based on (1) the number of genes detected per cell and (2) the proportion of mitochondrial gene/ribosomal gene counts. Low-quality cells were filtered if the number of detected genes was below 250 or above 3× the median absolute deviation away from the median gene number of all cells. Cells were filtered out if the proportion of mitochondrial gene counts was higher than 10% or the proportion of ribosomal genes was less than 10%. Mitochondrial genes (annotated with the prefix ‘MT-’), high abundance lincRNA genes, genes linked with poorly supported transcriptional models (annotated with the prefix ‘RP-’) and TCR (TR) genes (TRA/TRB/TRD/TRG, to avoid clonotype bias) were removed from further analysis. In addition, genes that were expressed in less than five cells were excluded.

We continued to use Scanpy to normalize the raw count data, identify highly variable features, scale features (normalization and algorithmization), and integrate samples. PCA (principal component analysis) was performed based on the most variable features identified. As there were more than ten thousand genes in analysis, we needed strong dimension reduction for later clusterization. PCA reveals the main axes of variation and denoises the data. Gene features associated with type I Interferon (IFN) response, immunoglobulin genes and specific mitochondrial related genes were excluded from clustering to avoid cell subsets driven by the above genes. Dimension reduction was done using the UMAP.

Leiden graph-clustering method was used for clusterization (community detection based on optimizing modularity) by Traag et al. (2018).

Clusters were labeled based on the expression of the canonical immune cell markers taken from the article.

**Results**

Our work was more educational in nature and allowed us to master processing single-cell RNA-seq data. We learned how to work with data from GEO presented as 3 count matrix files, carried out QC, single-cell data integration and clustering using Scanpy, assessed TIL expression profile.

**References:**

[1] Bardhan K., Anagnostou T., Boussiotis V.A. The PD1:PD-L1/2 Pathway from Discovery to Clinical Implementation. *Front. Immunol.* 2016;7:550. doi: 10.3389/fimmu.2016.00550.

[2] Han Y., Liu D., Li L. PD-1/PD-L1 Pathway: Current Researches in Cancer. *Am. J. Cancer Res.* 2020;10:727–742.

[3] Simoni Y, et al. Bystander CD8+ T cells are abundant and phenotypically distinct in human tumour infiltrates. *Nature.* 2018;557:575–579.

[4] Caushi JX, Zhang J, Ji Z, Vaghasia A, Zhang B, Hsiue EH, Mog BJ, Hou W, Justesen S, Blosser R, Tam A, Anagnostou V, Cottrell TR, Guo H, Chan HY, Singh D, Thapa S, Dykema AG, Burman P, Choudhury B, Aparicio L, Cheung LS, Lanis M, Belcaid Z, El Asmar M, Illei PB, Wang R, Meyers J, Schuebel K, Gupta A, Skaist A, Wheelan S, Naidoo J, Marrone KA, Brock M, Ha J, Bush EL, Park BJ, Bott M, Jones DR, Reuss JE, Velculescu VE, Chaft JE, Kinzler KW, Zhou S, Vogelstein B, Taube JM, Hellmann MD, Brahmer JR, Merghoub T, Forde PM, Yegnasubramanian S, Ji H, Pardoll DM, Smith KN. Transcriptional programs of neoantigen-specific TIL in anti-PD-1-treated lung cancers. Nature. 2021 Aug;596(7870):126-132. doi: 10.1038/s41586-021-03752-4. Epub 2021 Jul 21. Erratum in: Nature. 2021 Oct;598(7881):E1. PMID: 34290408; PMCID: PMC8338555.

[5] Zheng, Grace X.Y., Terry, Jessica M., [...] Bielas, Jason H. (2017). Massively parallel digital transcriptional profiling of single cells. *Nature Communications*. 8: 1-12, [doi:10.1038/ncomms14049](https://doi.org/doi:10.1038/ncomms14049)

[6] SCANPY: large-scale single-cell gene expression data analysis

F. Alexander Wolf, Philipp Angerer, Fabian J. Theis. *Genome Biology* 2018 Feb 06. doi: [10.1186/s13059-017-1382-0](https://doi.org/10.1186/s13059-017-1382-0).