COMBINING CANCER PATIENT SPATIAL TRANSCRIPTOMICS AND BULK RNA-SEQ DATA TO DRIVE INSIGHTS INTO NSCLC

FPN: 2301P
Translational research

Julia Bischof, Monika Schoeppler, Bita Motamedi-Baniassad, Nicole Kerstedt, Kerstin A. David, Jonathan Woodsmith

indivumed services
A CROWN BIOSCIENCE Company

Indivumed Therapeutics

INTRODUCTION

Combining cancer patient spatial transcriptomics (ST) and bulk RNA-Seq allows for a more nuanced understanding of the gene expression profiles. While large-scale sequencing projects have revealed some NSCLC cancer drivers, they cannot fully address the complexity of the heterogeneous tissue composition present in patients. ST enables the measurement of gene expression in specific regions within a tissue sample, complementing and refining insights gained from bulk RNA-Seq.

METHODS

Collection of samples: Tumor tissues were globally collected using a standardized protocol, minimizing the ischemia time until freezing in liquid nitrogen. To ensure the quality of the samples, all tissues were Hematoxilin and Eosin stained and subjected to a pathological QC. Samples need to be invasive, have a tumor content of >= 30 % and Necrosis <= 30 %. Normal tissues were processed in parallel and need to be free of tumor and representative regarding the tumor tissue to be included. Approximately 10 mg tissue were taken for nucleic acid extraction and protein lysate preparation each. To account for tumor heterogeneity, pathological QCs were performed on two sections, before and after taking the analysis material. The tissues stay frozen during the entire process.

RNA-Seq and Spatial Transcriptomics analyses were performed on one NSCLC patient, adenocarcinoma (ADC), stage IV. RNA-Seq differential expression was additionally done on 52 ADCs.

Nucleic Acid Extraction and Quality Assessment: Frozen tissue slices were mixed with beta-mercaptoethanol containing sample buffer and homogenized using the BeadBug system. RNA was extracted using the Qiagen AllPrep Universal Kit according to the manufacturer's instructions. RNA concentration was quantified using Qubit RNA BR assay respectively. RNA quality was assessed using Agilent High-Sensitivity RNA ScreenTape kit respectively. RNAs need to have a RIN >= 4 or a DV200 >= 60 to be selected for library preparation.

RNA-Seq: For whole transcriptome sequencing, RNA samples were depleted of the ribosomal RNA using the Ribo Zero Kit (Illumina) and library preparation was performed using the TruSeq Stranded Total RNA Kit (Qiagen). The library preparation kit was used according to manufacturer's instructions. Sequencing was performed on a NovaSeq6000 system (Illumina).
Whole transcriptome sequencing datasets have >= 100 million total reads with less than 20% of ribosomal origin and >=20 million reads mapping to mRNAs according to Ensembl reference. Ribosomal depletion was performed to remove nuclear rRNA and mt-rRNA. NGS data was aligned against Grch38 genome assembly.

10x Genomics Visium Spatial Gene Expression Manual Workflow for FFPE: One FFPE section was mounted on a Visium Spatial Gene Expression Slide containing spatially barcoded capture primers. The FFPE section was first H&E stained and scanned to visualize the histological tissue structure, followed by human whole transcriptome probe hybridization, targeting 17,943 human protein coding genes by 18,630 probe pairs. According to hybridization probes were ligated, followed by tissue permeabilization and probe release. Ligated probes were than captured via spatially barcoded primers and extended followed by Visium Spatial Gene Expression FFPE Library Preparation. The final library was sequenced at a concentration of 300 pM on a NovaSeq™6000 using a S1 v1.5 flowcell (Illumina) to reach a minimum sequencing depth of 25,000 read pairs per tissue covered spot by our sub-provider (IMGM Laboratories). Sequencing data containing the specific spatially sequencing barcodes were analyzed using the Space Ranger analysis pipeline (version 1.3.0) and the Loupe Browser (version 6.4.1) to assign the gene expression data to the corresponding histological positions in the tissue section.

Bioinformatical analyses: RNA-Seq differential and co- expression was based on normalized readcount data (TPM: transcripts per million).

10x Genomics data was integrated in R using Seurat package [1]. To remove artifacts, tumor and normal data were normalized separately, using SCTransform method. Both datasets were then combined for UMAP clustering [2] and differential expression analysis.

[1] Hao, Y. et al.: "Integrated analysis of multimodal single-cell data", Cell, 2021, 184(13); https://doi.org/10.1016/j.cell.2021.04.048, R package version 4.3.0.1

[2] Konopka, T.: "umap: Uniform Manifold Approximation and Projection", 2023, R package version 0.2.10.0

RESULTS

Integration and clustering of 10xGenomics tumor and normal data into R

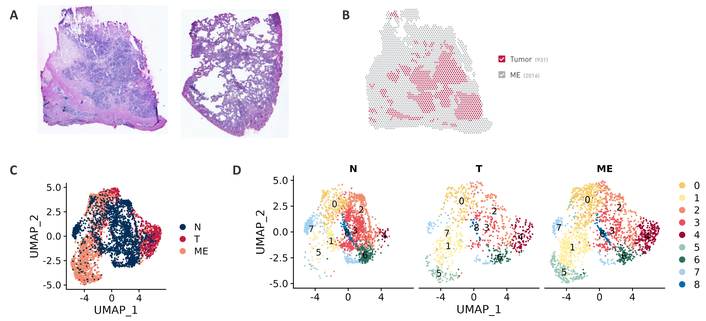

Figure 1: A) H&E scans of tumor (T, left) and normal (N, right) tissue. **B)** Identification of tumor and microenvironment (ME), including necrosis using Loupe Browser, based on 10xGenomics ST data. Number of spots under tissue were 2947 (T) and 2169 (N); mean reads per shot were 138,100 (T) and 129,120 (N) and median genes per spot were 1640 (T) and 1420 (N). **C)** UMAP clustering of ST data in R. N, T and ME datapoints can be differentiated in the clustering, as well as **D)** good separation into nine clusters with different representation in T, ME and N tissue.

RESULTS

Good correlation between both datasets, but bulk RNA-Seq includes a wider spectrum of genes and provides an overview about co-expressed and pathway relevant differentially expressed genes

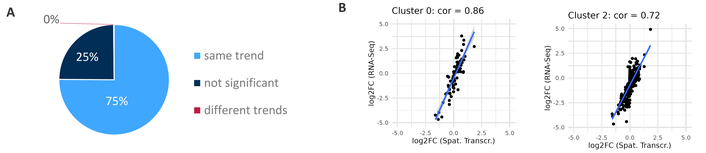

Figure 2: A) Bulk RNA-Seq and ST showed strong overall agreement between expression profiles, facilitating refinement of disease specific alterations observed in large NSCLC cohorts through ST data. **B)** Further log2 fold changes between tumor and normal correlated positively between patient specific ST and bulk RNA-Seq data (2/9 clusters shown as examples).

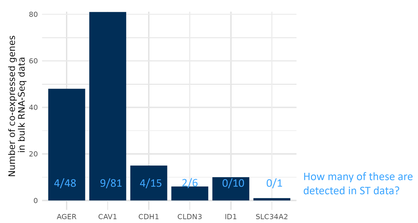

Figure 3: Co-expression analysis of bulk RNA-Seq data, filtered for positive correlation greater than 0.7 and genes known from NSCLC relevant publications (OpenTargets literature score > 0.6, specific for NSCLC related literature). AGER and CAV1 show biggest co-expression clusters, but only few of those genes are also detected in ST data.

RESULTS

Spatially resolved data can provide insights into the identification and assignment of cancer-relevant genes that primarily show only minor regulation in bulk RNA-Seq data

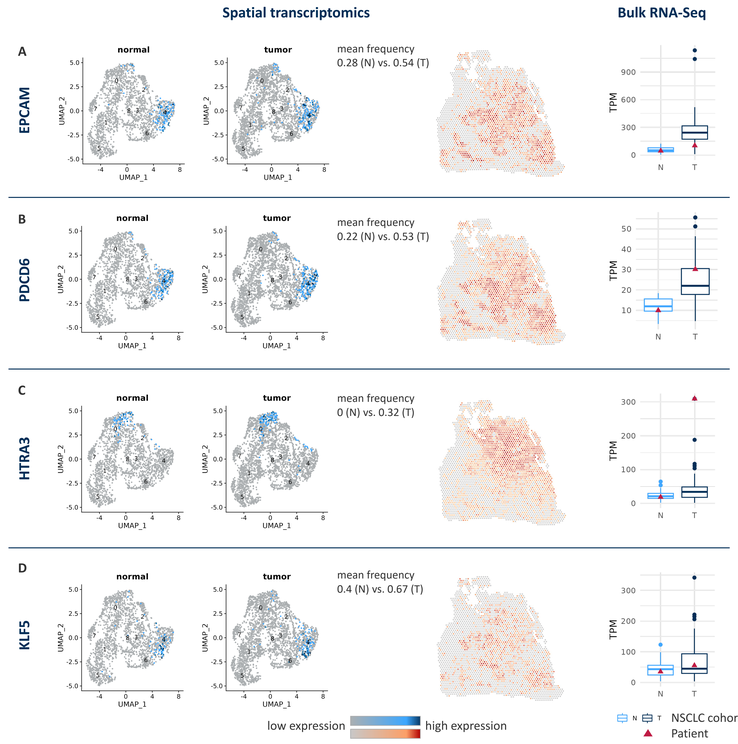

Figure 4: Patient-specific gene expression in specific tumor regions was identified and could guide future precision medicine approaches: for example tumor-related genes **A)** EPCAM (positive regulation of cell proliferation) or **B)** PDCD6 (angiogenesis) were higher expressed in the tumor center, whereas **C)** HTRA3 which suppresses tumor cell invasiveness shows higher expression in the microenvironment. **D)** Further the transcription factor KLF5 is higher expressed in tumor regions but shows no strong regulation in bulk data.

CONCLUSION

A more refined in situ understanding of gene expression profiles in the NSCLC microenvironment, combined with large cohort data will help guide therapeutic target selection. The combination not only refines target expression cell types (e.g. primary tumor or immune system) but also provide a strong indicator for target gene expression variability within patients' tumors for improved precision medicine approaches.

Contact: Julia Bischof (Bischof.Julia@indivumed.com).