Single-cell characterization of the tumor microenvironment in advanced stage Lung Adenocarcinoma

Caroline McCoach[1], Julia Rotow[1], Ashley Maynard[2], Lincoln Harris[2], David Naeger[1], Yaron Gesthalter[1], K. Pallav Kolli[1], Jonathan Weissman[1], Collin Blakely[1], Spyros Darmanis[2], Trever Bivona[1]

[1]University of California, San Francisco, USA, [2]Chan- Zuckerberg Biohub, USA

Abstract

The tumor microenvironment in Lung Adenocarcinoma is diverse and poorly characterized. Here we use single-cell RNA-seq to profile the cell types and subpopulations present in late stage LAUD patients. Cell clustering was performed in high dimensional space, followed by cluster identification using cell-type specific marker genes. Putative tumor clusters were identified and their mutational profile assessed. Multiple ALK/EML4 fusion cells were identified, as well as EGFR^{ex19} mutants. Several single-patient clusters contain multiple driver gene mutations, for example KRAS/EGFR. We also dissected the immune component of the tumor microenvironment, focusing specifically on T-cells. We assembled T-cell receptor sequences for 1500 cells, and identified 115 clonally expanded populations, including one clonogroup shared across three patients. Future work aims to assess epitope specificity of these expanded T-cells, providing insights into the diversity of T-cell function in the tumor microenvironment.

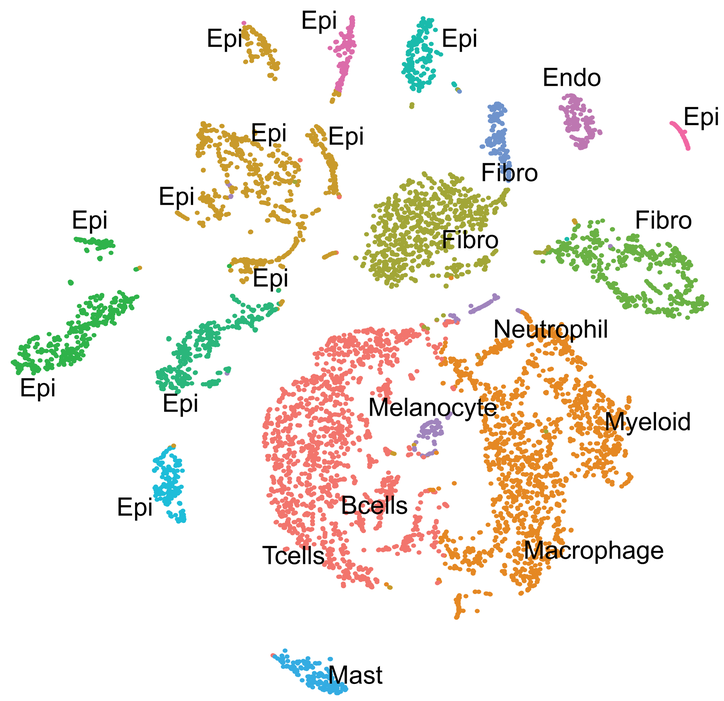

Figure 2. *tSNE for all cells included in analysis, with cluster labels*. Cluster identities were determined via differential gene expression test and manual annotation.

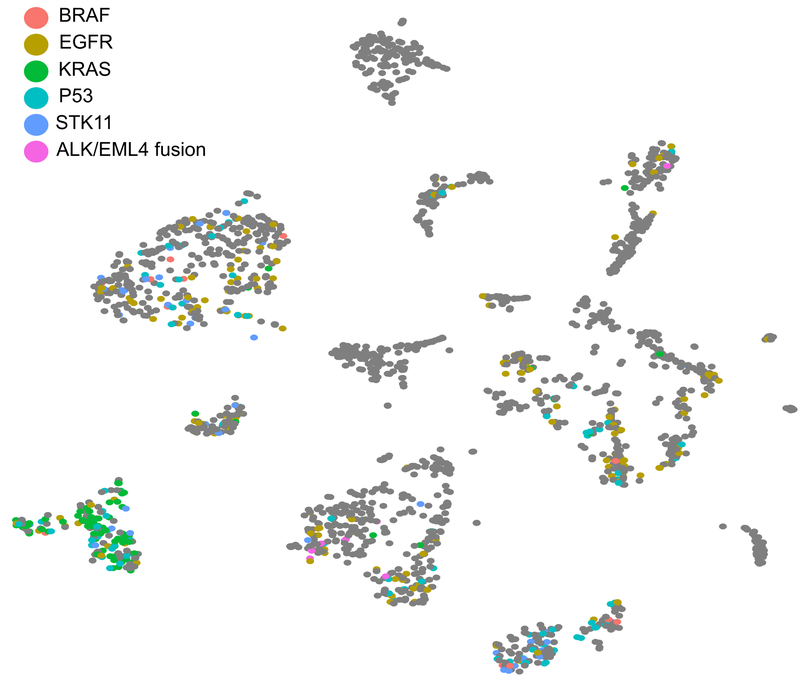

Figure 4. *tSNE for all epithelial cells, colored by LAUD driver gene mutation*. Mutations were called with GATK [6], then filtered via the COSMIC database [2], to isolate LAUD specific hits to genes of interest. ALK/EML4 fusion transcripts were identified with STAR-fusion [3].

Figure 1. *Analysis pipeline for high-throughput, agnostic variant calling.*

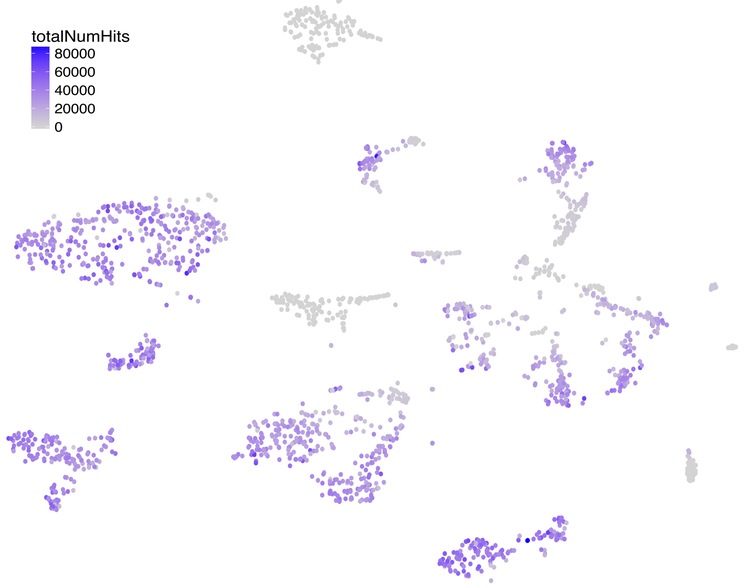

Figure 3. *tSNE for all epithelial cells, colored by mutational burden.* Mutations were called with GATK [6]. Note that SNPs and small indels are reflected in these scores, and not large-scale chromosomal aberrations.

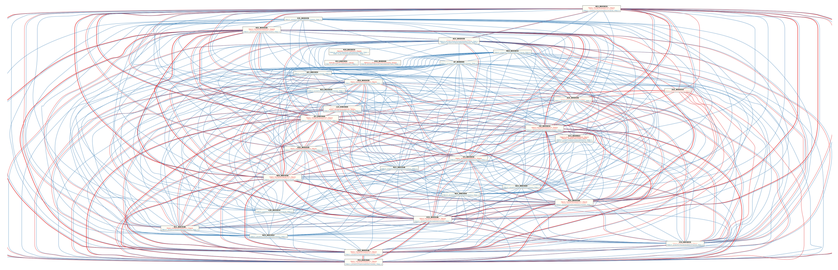

Figure 5. *Expanded T-cell clonotype network, containing cells from across three patients.* TCR sequences were assembled with Tracer [4].

Introduction

The composition of the tumor microenvironment is increasingly seen as essential to the disease progression. Here we use single-cell RNA-seq to characterize the tumor microenvironment in Lung adenocarcinoma. A significant challenge we faced was computationally classifying neoplastic cells from among a mixed population of tumorogenic and healthy cells. There is no single established method for this task – thus we developed a three part approach. First, cluster occupancy was determined, with the assumption that malignant clusters would be derived entirely from a single patient (data not shown). Next, we established CNV profiles for each of our clusters, with the assumption that malignant clusters would contain more large-scale chromosomal insertions/deletions (data not shown). Finally, we assessed mutational burden. We used a novel computational pipeline consisting GATK for variant calling, followed by relevancy filtering with COSMIC. Our goal was to assess the extent to which clustering is driven by overall mutational burden, and by LAUD driver gene mutations such as EGFR and KRAS. We also assembled TCR sequences and searched for expanded clonotypes, in an attempt to better understand immune presence and response within the tumor microenvironment.

Methods

Patient biopsies were collected at various timepoints during treatment, dissociated, sorted and subjected to single-cell sequencing (Smart-seq2). Cell/counts tables were generated, and Seurat was used for clustering and dimensionality reduction [1]. An epithelial-only subset was generated for further analysis. Cluster occupancy was calculated with custom R scripts. CNV profiles were calculated as in [5]. SNP/indel calling was done with GATK [6]. Variant hits were filtered according to experimentally validated SNPs/indels found in the COSMIC database [2]. Fusion transcripts were identified with STAR-fusion [3]. TRACER was used for T-cell receptor sequence assembly [4].

Results

Epithelial clusters appear to separate, to some extent, based on mutational burden. We believe mutational burden will serve as a valid indicator for cancer cell identity, and plan to validate this method on a negative control, *ie*. healthy lung tissue. We identified several single-patient epithelial clusters with multiple hits to LAUD driver genes. The conventional logic is that a given tumor should contain only one driver gene mutation, thus this finding is quite unexpected. We were able to assemble TCRs for a large portion of our patients, and see evidence of clonal expansion within the tumor microenvironment. In addition we see several clonotypes present across multiple patients. Future work will assess these clonotypes and ask whether they are associated with a known antigen response (*eg*. HPV) or represent a novel Tcell class.

References

1. Butler, *et. al.* (2018) *Integrating single-cell transcriptomic data across different conditions, technologies, and species*. Nature Biotechnology, *36*: 411-420.
2. Forbes, *et. al.* (2017) *COSMIC: somatic cancer genetics at high-resolution*. Nucleic Acids Research, *45*: D777-D783.
3. Haas, *et. al.* (2016). *STAR-Fusion: Fast and accurate fusion transcript detection from RNA-seq*. bioRxiv, *120295*; doi: https://doi.org/10.1101/120295
4. Stubbington, *et. al.* (2016). *T cell fate and clonality inference from single-cell transcriptomes*. Nature Methods, *13*: 329-332.
5. Tirosh, *et. al.* (2016). *Dissecting the multicellular ecosystem of metastatic melanoma by single-cell RNA-seq*. Science, *352*: 189-196.
6. Van der Auwera, *et. al.* (2013). *From FastQ Data to High-Confidence Variant Calls: The Genome Analysis Toolkit Best Practices Pipeline*. Current Protocols in Bioinformatics, *43*: 11.10.1-11.10.33