Prediction Of Intratumor Transcriptional Heterogeneity From Bulk Tumors

Agnieszka Kraft[1,2], Auguste Lefevre[1], Michaela B Kirschner[2], Mayura Meerang[2], Isabelle Opitz[2], Mitchell P Levesque[3,4], TUPRO Consortium, Valentina Boeva[1,5,6]

[1]ETH Zürich, Department of Computer Science, Institute for Machine Learning, Zurich, Switzerland
[2]Department of Thoracic Surgery, University Hospital Zurich, Zurich, Switzerland
[3]Department of Dermatology, University Hospital Zurich, Zurich, Switzerland
[4]Faculty of Medicine, University of Zurich, Zurich, Switzerland
[5]Swiss Institute of Bioinformatics (SIB), Zurich, Switzerland
[6]Institut Cochin, Inserm U1016, CNRS UMR 8104, Université Paris Descartes UMR-S1016, Paris, France

Abstract
In recent years tumor cell composition was associated with patients' survival and response to treatment. Although there were many studies on the composition of tumor microenvironment (TME), for many tumors the knowledge about differences between transcriptional profiles of cancer cells (here referred to as Intratumor Heterogeneity, ITH) is still limited. Previous studies on melanoma[1,2], gastric adenocarcinoma[3] and neuroblastoma[4] identified subpopulations of mesenchymal-like malignant cells expressing markers of epithelial-to-mesenchymal transformation (EMT), which were often associated with worse prognosis, metastases and resistance to treatment[3,4]. On the other hand, a vast majority of cancers have not been studied in this aspect, mostly due to the limited availability of scRNA-seq data. The majority of available methods for bulk data deconvolution assume prior knowledge of mixed components (DeMix[5], UNDO[6], ESTIMATE[7]) or require pure expression of components provided as a reference (DeconRNAseq[8], CIBERSORT[9]). More recent method[10] proposes an approach based on cancer subtyping into distinct molecular clusters, but does not address the possibility that tumors may consist of a mixture of cancer-cell subtypes. To overcome these limitations, we propose an unsupervised method to predict and analyze cancer-cell subpopulations from bulk data. By using matched bulk RNA-seq and scRNA-seq data from skin melanoma patients we established an approach which successfully separates bulk signal into respective components, separates cancer-cell subpopulations from TME-associated ones and measures ITH as a gradient between proportions of identified cancer-cell subpopulations within each tumor. Our method characterizes identified cancer-cell components by association with patients' survival, enriched pathways and driver mutations. Finally, we apply our method to 33 TCGA datasets and provide an overview of cancer-cells transcriptional profiles on a pan-cancer scale.

Keywords cancer, deconvolution, bulk RNA-seq, intratumor heterogeneity

Relevant topics
(1) Graphical representation of biological knowledge
(2) Multi-scale networks (genome, epigenome, transcriptome, proteome, metabolome, lipidome..)
(3) Single-cell data and network inference

References

1. Rambow, F. *et al.* Toward Minimal Residual Disease-Directed Therapy in Melanoma. *Cell* **174**, 843–855.e19 (2018).
2. Tirosh, I. *et al.* Dissecting the multicellular ecosystem of metastatic melanoma by single-cell RNA-seq. *Science* **352**, 189–196 (2016).
3. Zhang, M. *et al.* Dissecting transcriptional heterogeneity in primary gastric adenocarcinoma by single cell RNA sequencing. *Gut* **70**, 464–475 (2021).
4. Boeva, V. *et al.* Heterogeneity of neuroblastoma cell identity defined by transcriptional circuitries. *Nat. Genet.* **49**, 1408–1413 (2017).
5. Ahn, J. *et al.* DeMix: deconvolution for mixed cancer transcriptomes using raw measured data. *Bioinformatics* vol. 29 1865–1871 (2013).
6. Wang, N. *et al.* UNDO: a Bioconductor R package for unsupervised deconvolution of mixed gene expressions in tumor samples. *Bioinformatics* **31**, 137–139 (2015).
7. Yoshihara, K. *et al.* Inferring tumour purity and stromal and immune cell admixture from expression data. *Nat. Commun.* **4**, 2612 (2013).
8. Gong, T. & Szustakowski, J. D. DeconRNASeq: a statistical framework for deconvolution of heterogeneous tissue samples based on mRNA-Seq data. *Bioinformatics* **29**, 1083–1085 (2013).
9. Newman, A. M. *et al.* Robust enumeration of cell subsets from tissue expression profiles. *Nature Methods* vol. 12 453–457 (2015).
10. Wang, L. *et al.* A reference profile-free deconvolution method to infer cancer cell-intrinsic subtypes and tumor-type-specific stromal profiles. *Genome Med.* **12**, 24 (2020).