ELSEVIER

Available online at www.sciencedirect.com

ScienceDirect

Current Opinion in
Immunology

High-throughput genomic profiling of tumor-infiltrating leukocytes

Aaron M Newman[1,2] and Ash A Alizadeh[1,2,3,4]

Tumors are complex ecosystems comprised of diverse cell types including malignant cells, mesenchymal cells, and tumor-infiltrating leukocytes (TILs). While TILs are well known to play important roles in many aspects of cancer biology, recent developments in immuno-oncology have spurred considerable interest in TILs, particularly in relation to their optimal engagement by emerging immunotherapies. Traditionally, the enumeration of TIL phenotypic diversity and composition in solid tumors has relied on resolving single cells by flow cytometry and immunohistochemical methods. However, advances in genome-wide technologies and computational methods are now allowing TILs to be profiled with increasingly high resolution and accuracy directly from RNA mixtures of bulk tumor samples. In this review, we highlight recent progress in the development of *in silico* tumor dissection methods, and illustrate examples of how these strategies can be applied to characterize TILs in human tumors to facilitate personalized cancer therapy.

Addresses
[1] Institute for Stem Cell Biology and Regenerative Medicine, Stanford University, Stanford, CA, USA
[2] Division of Oncology, Department of Medicine, Stanford Cancer Institute, Stanford University, Stanford, CA, USA
[3] Stanford Cancer Institute, Stanford University, Stanford, CA, USA
[4] Division of Hematology, Department of Medicine, Stanford Cancer Institute, Stanford University, Stanford, CA, USA

Corresponding authors: Newman, Aaron M (amnewman@stanford.edu) and Alizadeh, Ash A (arasha@stanford.edu)

Current Opinion in Immunology 2016, **41**:77–84

This review comes from a themed issue on **Cancer immunology: genomics & biomarkers**

Edited by **Ton N Schumacher** and **Nir Hacohen**

http://dx.doi.org/10.1016/j.coi.2016.06.006

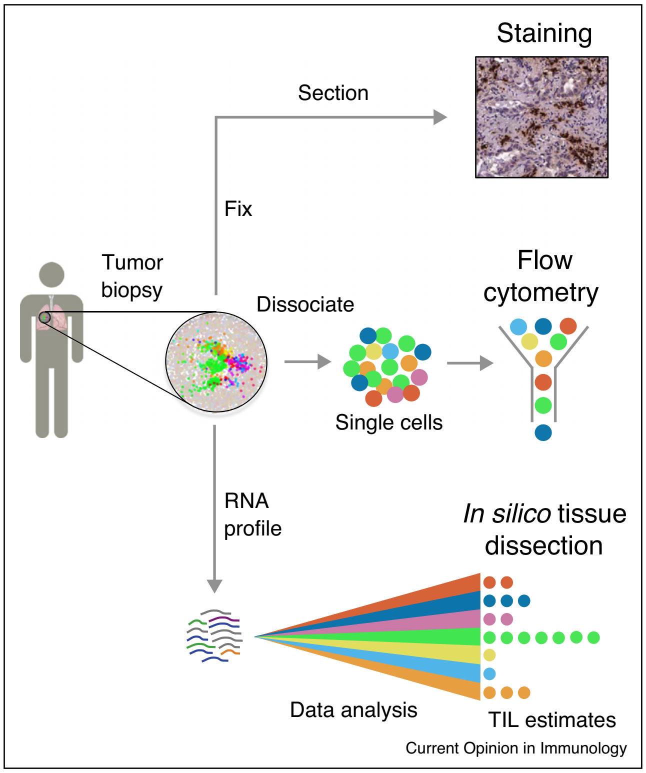
Introduction

TILs are critical determinants of cancer clinical outcomes and play important roles in tumor growth, cancer progression, and response to therapy [1••,2–5,6••,7,8••,9]. In recent years, novel immunotherapies have achieved unprecedented success in harnessing TILs to target human tumors [10–13]. For example, monoclonal antibodies that block PD-1/PD-L1 signaling can elicit durable anti-tumor T cell responses in some patients [12]. However, the majority of patients receiving these therapies either fail to achieve a long-term benefit or never respond. Several studies have found positive correlations between response to PD-1/PD-L1 blockade and immunological features of a patient's tumor prior to treatment, including higher levels of tumor-infiltrating CD8 T cells [5] and estimated tumor neoantigen load [14]. However, the predictive strength of these candidate biomarkers for therapeutic efficacy is currently suboptimal and their biological significance is only partially understood [13,15]. A better understanding of the key relationships between TILs, tumor subtypes, clinical parameters, and diverse therapies would facilitate the development of improved biomarkers and individualized treatments.

Until recently, flow cytometry and immunohistochemistry (IHC) have been the two most common approaches for profiling TILs in complex tissues (Figure 1). While both methods have significant utility, they also have notable limitations for high-resolution TIL characterization. For example, flow cytometry, like other single cell analysis methods (e.g., single cell RNA-seq), requires mechanical or enzymatic dissociation of solid tissues, which can distort TIL representation [6••,16••,17]. In contrast, IHC is directly applicable to solid tissues, but is generally limited to one marker (or cell type) per tissue section, restricting its scope to a small number of cell types. Finally, the reliance of both techniques on markers with available antibodies can complicate detection of some TILs, particularly those that require multiple such markers. While several recently reported techniques can overcome some of these issues through higher order multiplexing [18–20], methods that combine genomics with bioinformatics have significant potential to enable high resolution TIL assessment.

For over a decade, computational techniques have been applied to decipher cellular content directly from genomic profiles of mixture samples [8••,16••,17,21•,22,23,24••,25–27,28•,29–37,38,39•,40,41•,42,43•]. Here, we review recent developments and outstanding questions related to *in silico* dissection of TILs from bulk tumors (Figure 1). Given the emphasis in the field, we focus on tumor gene expression profiles (GEPs), although many of the methods and concepts described here could be extended to other high-dimensional genomic data types (e.g., methylation data). With further refinements, we expect that *in silico* tissue dissection will become a routine analytical technique for

Figure 1



Current and emerging techniques for evaluating TIL composition in solid tumors.

characterizing cellular heterogeneity in a variety of research and clinical settings.

In silico approaches for TIL profiling

Analytical methods for profiling TILs in bulk tumor transcriptomes can be broadly classified based on their reliance on (1) enrichment measures for genes associated with individual cell types or (2) algorithmic deconvolution of admixed transcriptomes to resolve composition. Regardless of their main analytical underpinnings, each of the methods reviewed in this article inherently requires prior knowledge of 'marker genes' enriched in each TIL subset of interest.

Typically, marker genes of specific leukocyte subsets are defined either from prior biological knowledge (e.g., established markers used for FACS or IHC), or by definition of differentially expressed genes after profiling functionally defined leukocyte subsets (whether directly purified from human tissues or following established approaches for *in vitro* differentiation/stimulation). To facilitate unbiased identification of robust cell type-specific markers, we generally favor the latter approach to systematically define differentially expressed GEPs. In some instances, reference GEPs may not be readily available, possibly due to their rarity and difficulties with efficient cell sorting. However, if a small number of lineage specific genes are already known, then it might be possible to derive additional marker genes using *in silico* nanodissection, a novel machine learning technique for predicting cell type-specific genes from GEP mixture data [44•]. We refer the reader to [16••,39•] for additional details of gene expression deconvolution methods, including marker gene selection methods and technical considerations. Table 1 summarizes various GEP enrichment and deconvolution methods, highlighting and comparing their key features.

Marker gene enrichment

Distinct gene expression programs underlie phenotypic variation among cell types in complex tissues. Therefore, one common approach for studying TILs is to measure the enrichment of immune-related genes in bulk tumor GEPs. In one early proof-of-principle study, Dave and colleagues analyzed 191 GEPs from untreated bulk follicular lymphoma tumors, and identified two clusters of prognostic genes that were significantly enriched in genes expressed by T cells, macrophages, and/or dendritic cells, but not B cells, suggesting tumor infiltration by non-malignant leukocytes [23]. More recently, Nagalla and colleagues analyzed nearly 2000 breast tumor GEPs and identified expression modules enriched in known immune genes. These genes were cross-referenced with normal leukocyte reference profiles to infer potential TIL identities [33].

Defining TIL-enriched gene clusters from bulk tumors can be relatively straightforward and can provide useful hints for follow-up studies as demonstrated by several groups. However, this approach cannot readily measure TIL composition, nor can it address noise in gene expression levels or effectively discriminate between TIL subsets with highly similar transcriptomes (e.g., naïve vs memory B cells), especially when these cells are rare [45]. One important step toward addressing these issues involves defining leukocyte-specific genes from reference GEPs of purified cell types and then evaluating these genes in bulk tumor samples. For example, Bindea and colleagues inferred TIL survival associations in colorectal cancer patients using genes that discriminate 24 normal leukocyte subsets [31]. In a broad analysis of tumor genomic and transcriptomic data across thousands of tumors profiled by The Cancer Genome Atlas (TCGA), Rooney and colleagues discovered new relationships between cytolytic activity and tumor genomic features using genes enriched in cytotoxic T cells and NK cells [8••]. More recently, Tirosh and colleagues used single cell RNA-seq to define genes enriched in specific TIL and stromal subsets from melanoma biopsies. Further analysis of these genes in bulk melanoma tumors revealed evidence for novel cell-cell interactions [43•].