

Bioinformatics approaches to profile the tumor microenvironment for immunotherapeutic discovery

Trevor Clancy¹, Ruth Dannenfelser^{2,3}, Olga Troyanskaya^{2,3}, Karl Johan Malmberg^{1,4,5}, Eivind Hovig^{2,7,8}, and Vessela Kristensen^{9,10,11}

¹ Department of Cancer Immunology, Institute for Cancer Research, The Norwegian Radium Hospital, Oslo University Hospital, Norway

² Department of Computer Science, Princeton University, Princeton, New Jersey.

³ Lewis-Sigler Institute for Integrative Genomics, Princeton University, Princeton, New Jersey

⁴ K.G Jebsen Center for Cancer Immunotherapy, Institute of Clinical Medicine, University of Oslo, Oslo, Norway.

⁵ Center for Infectious Medicine, Department of Medicine, Karolinska Institutet, Stockholm, Sweden.

² Department of Tumor Biology, Institute for Cancer Research, The Norwegian Radium Hospital, Oslo University Hospital, Norway.

⁷ Biomedical Research Group, Department of Informatics, Faculty of Mathematics and Natural Sciences, University of Oslo

⁸ Institute of Cancer Genetics and Informatics, The Norwegian Radium Hospital, Oslo University Hospital

⁹ Department of Clinical Molecular Biology (EpiGen), Division of Medicine, Akershus University Hospital, Lørenskog, and University of Oslo, Oslo, Norway

¹⁰ Department of Genetics, Institute for Cancer Research, The Norwegian Radium Hospital, Oslo University Hospital, Oslo, Norway

¹¹ K.G. Jebsen Centre for Breast Cancer Research, Institute for Clinical Medicine, Faculty of Medicine, University of Oslo, Norway

Corresponding author:

Vessela Kristensen, Department of Clinical Molecular Biology (EpiGen), Institute of Clinical Medicine, University of Oslo, P. O. Box 28 Ahus, 1478 Lørenskog, Norway. E-mail:

v.n.kristensen@medisin.uio.no

Summary

In the microenvironment of a malignancy, tumor cells do not exist in isolation, but rather in a diverse ecosystem consisting not only of heterogeneous tumor-cell clones, but also normal cell types such as fibroblasts, vasculature, and an extensive pool of immune cells at numerous possible stages of activation and differentiation. This results in a complex interplay of diverse cellular signaling systems, where the immune cell component is now established to influence cancer progression and therapeutic response. It is experimentally difficult and laborious to comprehensively and systematically profile these distinct cell types from heterogeneous tumor samples in order to capitalize on potential therapeutic and biomarker discoveries. One emerging solution to address this challenge is to computationally extract cell-type specific information directly from bulk tumors. Such *in silico* approaches are advantageous because they can capture both the cell-type specific profiles and the tissue systems level of cell-cell interactions. Accurately and comprehensively predicting these patterns in tumors is an important challenge to overcome, not least given the success of immunotherapeutic drug treatment of several human cancers. This is especially challenging for subsets of closely related immune cell phenotypes with relatively small gene expression differences, which have critical functional distinctions. Here we outline the existing and emerging novel bioinformatics strategies that can be used to profile the tumor immune landscape.

Immune cell diversity – an expanding universe of repertoires in the TME

The human immune system encompasses the set of processes and cells that protect the body by responding to invading cells. Components of the immune system have one critical task, they must be able to recognize and distinguish between self and non-self such as non-self antigens from harmful bacteria, viruses, fungi, parasites, and tumors. More specifically, immune cells are responsible for maintaining tissue homeostasis by identifying potentially harmful non--

self-molecules, neutralizing targets by adaptive cytotoxic and phagocytic responses against invading, malignant or stressed/damaged cells. Since pathogens can rather quickly evolve and adapt, the immune system must be particularly sophisticated and robust. A top level simplified description divides the immune system into two main layers: the innate immune and adaptive immune response. The innate system is composed of the first responders in immune defense and includes neutrophils, eosinophils, basophils, macrophages, dendritic cells, mast cells and natural killer cells. Many of these cell types have documented presence in the tumor microenvironment. The innate response is triggered by several possible events; usually distress signals caused by injured or dying cells. Once activated, for the most part, innate immune cells act in a non-specific manner attacking all pathogens. The cell types in the innate response have diverse roles. Some play facilitator roles, such as mast cells. They recruit more immune cells to the microenvironment of the inflamed tissue site, by secreting inflammatory cytokines and chemokines. Other innate immune cells are responsible for neutralizing threats by engulfing and digesting invaders, such as macrophages, while for example granulocytes neutralize through the granulated secretion of cytotoxic molecules in a cell-to-cell interaction with a pathogen.

The adaptive response has evolved to take effect over a more gradual time period compared to the almost instantaneous non-specific innate immune response. The hallmark of cells within the adaptive immune system is their expression of unique rearranged immune receptors. Hence, the adaptive immune response is more targeted, as it recognizes pathogen's specific antigens. Critically, the adaptive immune cells are often driven towards an adaptive cellular memory phenotype, which provides the opportunity to instigate an immediate response upon any future recognition of the pathogen specific antigen.

Lymphocytes, more particularly the small lymphocytes, B cells and T cells, are the main components in the adaptive immune response. Each of these cell types may exist in one of

three main stages of activation; naive cells that have yet to encounter an antigen, effector cells which are currently fighting an antigen, and memory cells which remain after a previous attack ready to recognize the same pathogen again. Besides sharing these three main stages of activation, the key effector cells of the adaptive immune system, B cells and T cells, have diverse functional phenotypes, both within and between these immune cell lineages.

B cells have been shown to play a role in immune surveillance during tumor progression. Their primary responsibility is to produce specific antibodies to tag and neutralize antigens [1]. With the assistance of CD4⁺ T-helper cells, B cells can differentiate into antibody-producing factories (plasma cells) or memory cells ready for secondary exposures. In terms of pathogen recognition, a key difference between B cells and T cells stems from the ability to recognize antigens. B cells can recognize antigens without assistance, whereas T cells typically require some preprocessing of the antigen before they can recognize them.

T cells can be further subdivided into a large and diverse lineage tree of specific types, such as cytotoxic-T-lymphocytes (CTL), T-helper 1 cells (Th1), T-helper 2 cells (Th2), T-helper 17 cells, T follicular helper cells (Tfh), and regulatory T cells (Treg). CTLs are responsible for killing infected and damaged or dysfunctional cells through the secretion of cytotoxins, namely perforin and granzyme. CTLs are major players in cancer immunology because they are primarily responsible in the first wave recognition and elimination of tumor cells.

The T-helper cells can be conceptually perceived as the moderators of the immune response. They secrete cytokine and chemokine signaling molecules that direct pathogen-specific responses. Th1 cells produce IFN γ , which is necessary for the regulation of immunogenic responses, antigen processing and presentation of tumor cells, for the function of macrophages and for the production of CTLs. Th2 cells secrete interleukins, such as IL4, IL5, IL6, IL9, IL10, and IL13, that help increase B cell production and B cell differentiation into plasma cells. More

generally, Th1 cells signal to increase the immune response to attack compromised cells, and Th2 cells signal to increase the immune response to attack pathogens outside of the cells. It is also important to note that the Th1 and Th2 cell signaling networks are involved in positive and negative feedback loops, whereby the components in the complex network of immune cell types in the tumor microenvironment (TME) regulate other [2].

The complexity of unraveling this infiltrate in the TME is magnified when considering the diversity of highly similar immune cell subsets. For example, it was recently demonstrated with the use of 28 molecular surface markers, that there is an estimated up to 30,000 phenotypic subsets of NK cells within each donor and more than 100,000 NK cell phenotypes among the entire panel of 17 donors [3]. The emergence of a mass cytometry technology, which combines the workflow of flow cytometry with the precision of mass spectrometry made this observation possible [4] and creates new bioinformatics challenges related to profiling the repertoire of immune cells at the single-cell level. These are outside the scope of this overview, but are covered in a recent comprehensive review on the topic [5]. These technology platforms hold great potential to capture the enormous diversity of the immune cell repertoire in the TME, however the advantages of using bioinformatics to predict accurate immune signatures from bulk native tumor tissue is that the functional intracellular and intercellular transcriptome profiles are preserved, whereas in the purified immune cells the molecular profiles will have inherently different patterns when isolated from native tissue. The remarkable diversity revealed by these and similar recent studies pose significant research challenges for the *in silico* identification of distinct immune cell subsets that are predictive of a clinical outcome or response to therapy in biomarker discovery.

The properties of immunosurveillance in the TME and the dynamic anti-tumor immune response

During the dynamic stages of cancer progression, when normal or primary tumor cells progress toward aggressive metastatic cell populations; mutated cell-surface markers, metabolism, cytokine and chemokine profiles may change from a state of self to a non-self inflamed state. This transition toward an altered composition of immune activation in the TME may radically alter its immune-cell composition. This process of the immune-cell component being shaped in the TME during cancer progression has been effectively summarized during the resurgence in our understanding of the anti-tumor immune response over 10 years ago as the “3E’s of cancer immune-editing”- elimination, equilibrium and escape [6]. When considering the scale of diversity and dynamics of immune cells in the TME during immune-editing, it becomes clear that there is dire need for developing bioinformatics tools to capture this information in a manner which is more fit-for-purpose.

In the elimination phase the perturbed state in the TME and disruption of the normal immune homeostasis may activate the innate immune system leading to the recruitment of some of the general innate immune-cell types to the TME. These cells of the innate immune system may recognize ligands on the tumor either by the emergent inflammation or the cellular transformation process in the progressing tumor. Regardless of the specific mechanisms of the various possible cell-cell interactions between innate immune cells in the TME, the most frequent outcome is a significant production of IFN- γ that promotes further recruitment of innate immune-cell subsets and enhance tumor cell killing. The increased supply of antigenic material from dead tumor cells stimulates recruitment of immune-cells of the adaptive immune response. For example, tumor-specific CD4⁺ and CD8⁺ T adaptive cells are recruited to the TME and participate in the killing of antigen-positive tumor cells. The cytokine IL-2 production by the CD4⁺ T helper maintains the activation of the cytotoxic CD8⁺ T cells

function and viability of the tumor-specific CD8⁺ T cells. The increased IFN- γ also enhances the antigen presentation of tumor cells themselves, allowing for efficient recognition by the CD8 T cells and their eradication of the tumor [7]. If tumor elimination is carried out to completion, the sculpting of tumor development by the immune system comes to an end and the tumor is completely eradicated. However in progressing tumors the immunoediting process enters the equilibrium phase [6] where the activity of CD8⁺ T cells and IFN- γ production continues to kill tumor cells while simultaneously tumor cells with a high mutated phenotype are evolutionarily selected to survive. Over potentially a long time period, phenotypic alterations caused by external stimuli, genomic and epigenomic changes that are beneficial for overall tumor survival will be selected for and the tumor may have a diminished potential to be immunogenic in this equilibrium phase. Eventually in the escape phase of cancer immunoediting, the tumor bypasses an operative immune system in the TME and can grow unchecked and cancer cells survival, invasion, and dissemination of tumor cells is enhanced. The transition to immune escape may occur due to various possible signaling, genomic and epigenetic perturbations in the tumor [6, 8].

The type, density and spatial location of immune cells in a tumor are dynamic features, which are both patient-specific and highly unique for each tumor lesion [2]. Yet, a few general patterns have emerged from advanced immunophenotyping studies. B cells are more likely to be found in the invasive margin of the tumor and the areas adjacent to the tumor tissue. T cells, specifically the CTLs are typically found in deep in the core of the tumor, with fewer in the invasive margin and surrounding tissue [9]. For a progressing tumor to reach immune escape it usually has recruited an immune suppressive environment in the TME. Understanding the mechanisms of how tumors may promote the recruitment of immune suppressive cells is an active area of research, and is important for development of anti-cancer therapies. The actions of distinct populations of certain T regulatory cells, myeloid derived suppressor cells (MDSC),

and macrophages are highly implicated in this process. These suppressor cells are present in the TME as heterogeneous population of immature cells that suppress innate and adaptive immunity. The MDSCs utilize a variety of mechanisms to suppress T cell activation and promotes a transformation of the TME to promote tumor growth [10]. The presence and activation of MDSCs in the TME well characterized; however, the tumor-promoting mechanism and functional aspects of MDSCs on cancer cells is poorly defined [11]. The bioinformatics tools to profile for immunosuppressive cells in the TME at the resolution of capturing detailed subpopulation that have precisely defined mechanisms of action are in their infancy. The continuous development and improvements on the bioinformatics tools reviewed here may add to the growing body of evidence indicating that upon immune escape, immune cells with a suppressive phenotype contribute to broad and complex immunosuppressive network in the TME.

All tumors elicit an immune response, but some more so than others. One recent idea is that the number of unique cancer antigens expressed by tumors explain this phenomenon [12]. If a tumor cell expresses entirely tolerated antigens, it is more difficult for the immune system to recognize it as a threat. In an additional layer of complexity, tumors actively block the immune system by expressing proteins that interfere with antigen presentation, immune cell recruitment, activation, and differentiation. These “checkpoint” proteins, which are conventionally used to regulate the immune response, are maliciously used by the tumor as defense. Some recent well studied examples are programmed cell death protein 1 (PD1) and cytotoxic T lymphocyte antigen 4 (CTLA4), which both negatively regulate T cell responses, particularly inhibiting CTLs. Treatments that block the expression of PD1 and CTLA4 have been particularly promising in eliminating tumors in melanoma, prostate, lung, and other cancers [12-14].

The TME composition as a biomarker for a successful anti-tumor response

Recent discoveries that examine how immune cell activation is regulated in the TME, have allowed for the characterization of targetable “checkpoint” proteins which have translated to therapeutic effect [15]. Long term positive clinical outcomes and survival greater than 10 years have been observed [16] in a small subset of patients with metastatic melanoma. These developments lead us on the path to deliver curative therapies to cancer patients, but in order to maximize the potential of discovering new checkpoint control drugs and understand how their molecular mode of action operates, especially in non-responders, we need a way to thoroughly examine the TME prior to, during and after therapy [17]. The bioinformatics tools outlined in this review may be beneficial for this type of profiling in future areas of TME clinical research.

The TME is sculpted in a spatiotemporal fashion by the process of the three stages of immune editing [18], ranging from (a) distinct properties of immune cells in the TME to carry out tumor killing, leading to (b) an equilibrium between promotion of tumor cell proliferation and tumor killing by immune cells, and finally to a critical phase of (c) tumor escape, whereby an immunosuppressive environment in the TME has arisen. This immunosuppressive environment potentially stimulates growth and deactivation of immune cells, and regulates the tumor cell intrinsic ligand interaction by checkpoint receptors on immune cells. As previously described [19], these acts of immunosurveillance by key immune cell phenotypes infiltrated into the TME can allow tumors to exploit normal immune checkpoints that normally prevent aberrant T cell responses. The aforementioned checkpoint PD-1 receptor is expressed on activated T cells, where it functions as an antagonist to T cell receptor signaling after binding the PD-L1 or PD-L2 ligand. Activation of the PD-1 immune checkpoint limits auto-immunity during inflammatory responses [20]. Evidence that PD-1 activation prevents T cell-mediated killing of tumor cells began with the observation that while normal human tissue lacks PD-L1 expression, PD-L1 is abundantly expressed in many tumors [21]. These and many similar observations

motivated a therapeutic effort to activate an adaptive immune response against tumors, by blocking the checkpoint interaction using monoclonal antibodies targeted against PD-1 or the PD-L1 ligand [22, 23]. Targeting the immune checkpoint blockade in the TME has produced durable responses in patients with advanced melanoma, renal cancer, and non-small-cell lung cancer [23], supporting the importance of this pathway in linking cancer progression to tumor immune editing.

The presence of tumor infiltrating lymphocytes (TILs) in the TME of melanomas is strongly associated with expression of PD-L1 [24]. A study in 46 patients with metastatic melanoma found that response to the PD-1 inhibitor pembrolizumab was associated with the pre-treatment presence of PD-L1 expressing immune cell infiltrates in the TME, not exclusively at the tumor margin [25]. A recent phase I trial using the anti-PD-L1 antibody MPDL3280A in 277 patients found that elevated PD-L1 expression in infiltrating immune cells, but not in tumor cells, was associated with elevated objective response rates [26]. The most consistent result so far has been that patients lacking tumor immune cell infiltrates do not benefit from checkpoint inhibitor therapy [19]. However, much remains to be determined about which patients will benefit from these approaches [19].

The need for bioinformatics tools to predict immune cells from the TME

It is now well established that clinical outcome and therapeutic response is governed by the dynamic presence of immune cells in the TME [27]. Our improved understanding of the impact of immune cell infiltration into the TME leads us ever more toward the design of effective combination immunotherapy strategies [28, 29]. The basis for this has been established through numerous studies using conventional laboratory assays [30, 31], and more recently, with whole transcriptomics approaches [31-33]. Notably, Galon *et al.* were among the first to report that the type, density and spatial location of infiltrating CD8⁺ T cells in human epithelial tumors

were features associated with positive clinical outcome in colorectal cancer [5]. These and other similar recent observations [34] have inspired the momentum toward creating a formalized diagnostic test called the ImmunoScore [28, 35] that is now being investigated in a series of clinical studies [6]. For example in the case of colorectal cancer, the Immunoscore staging proved to be an effective predictor of disease recurrence, beyond that of the convention microsatellite instability [34].

Despite these recent advancements in the field, significant challenges need to be overcome for accurate and comprehensive profiling of the immune landscape of the TME for biomarker and therapeutic discoveries. It is laborious, time consuming, expensive, and difficult to automatically and thoroughly scrutinize the presence of multiple immune cells in tumors with fidelity by laboratory techniques such as immunohistochemistry. Computational methods are therefore desired to not only predict the network landscape of general immune cell types (*e.g.*, CD4 vs. CD8), but also identify and precisely define immune cell states at higher resolution (various stages of differentiation or activation), as opposed to general classifications of immune cells such as "pro-tumor" and "antitumor". It is becoming increasingly clear that in order to achieve such optimal automated procedures, systems biology [36-38] and bioinformatics [39] approaches can assist by making predictions of the immune cell component from –omics data to guide this discovery process.

Spurred by the increased availability of –omics technologies, we have currently entered an era of digital cancer medicine; whereby diverse and global molecular profiling may guide therapeutic discovery to predict response to immunotherapy [40-42]. Here, we will outline how bioinformatics deconvolution strategies used to analyze transcriptome data have the potential to provide a cell-subset centered view of the immune system's role in the TME.

Diversity of immune cells in the TME and their impact on immunotherapeutic discovery

One of the major hurdles to overcome in understanding the role of the TME and how immunotherapeutic modulation can be better exploited in personalized medicine, is to uncover the precise mechanisms of how therapy may affect distinct populations of immune cells within the TME [17]. This is especially important when mapping out the detailed landscape of diverse immune cell types in the TME prior to, during and after therapy [43, 44]. It has been shown [45] for many tumor types that the presence of a lymphocytic infiltrate, across a range of cancers types, may be a positive predictor of clinical outcome [44]. The immune cell landscape in the TME includes a broad range of cell types, many of which have revealed transcriptome predictive signatures detected from bulk tumor tissue (see Figure 1). These transcriptome gene signatures correspond to cell types such as natural killer (NK) CD56dim cells, CD4⁺ T helper cells, CD4⁺ T regulatory cells, memory T cells (CD8⁺CD45RO⁺ cells), CD8⁺ T cells, myeloid derived suppressor cells (MDSCs), macrophages, and B cells [31, 46]. The effect that diverse infiltrated immune cells have on prognosis and response to therapy has often been described as being paradoxical, and very much dependent on the tumor stage and tumor type [47]. This is especially true of the members of the adaptive immune response and their roles during tumor progression [48]. Both CD8⁺ cytotoxic T cells (CTL) and NK cells harbor anti-tumorigenic properties and are among the main regulators of tumor immune surveillance [31,41], and T helper 1 cells also contribute to tumor cell clearance, by favoring the differentiation of M1-polarized macrophages, which may have tumoricidal properties [42]. In contrast, while B cells contribute to anti-tumor immunity in the acute phase of a malignancy; chronic inflammatory B cell signaling can be pro-tumorigenic [49, 50]. Indeed, Th2-driven B cell differentiation exerts pro-tumorigenic activities through the recruitment of MDSCs, including monocytes and mast cells, and the polarization of macrophages to an M2 phenotype [15,43]. The combination of high levels of tumor-associated macrophages (TAMs), robust Th2

responses and low CTL/NK cell infiltration have been used to create an immune cell signature in breast cancer to predict patient survival [45] (reviewed in [44]). Further, it has been recently shown that chemotherapy stimulates macrophage differentiation and the subsequent cross talk between macrophages and cytotoxic T cells in the tumor may contribute to response to chemotherapy [45,46]. TAMs are heterogeneous cell populations in their own right, with specific localization in the tumors and gene expression signatures [37].

The accurate prediction of high-resolution biomarkers from the TME will no doubt have valuable impacts in areas of personalized medicine related to prognosis, patient stratification, patient monitoring, and drug safety and efficacy studies. Dedicated bioinformatics pipelines will be critical in the regard [5], especially those which have undergone a degree of validation and those which integrate multiple high-throughput platforms in the same workflow, taking advantage of the technologies which analyze bulk complex tissue and single cell technologies in the same setting [51].

Additionally, the spatial temporal context of infiltrating immune cells can have a clinical impact [47]. It has been shown that not only the presence, but also the type, density, and spatial location of the lymphocytic infiltrate may be of prognostic significance [31]. However, such features may be difficult to capture from transcriptome analysis of a single biopsy. Recently the spatiotemporal features between distinct immune cell types were demonstrated to be predictive of patient survival in colorectal cancer [31]. This was achieved by integrating bulk tumor transcriptomics data and immune gene expression signatures with tissue microarrays to capture some dynamic features of the immune landscape over time. It is clear however, that in order to achieve spatiotemporal profiling at levels of high resolution (various stages of activation, differentiation and function in a patient's repertoire), we will need to perform high-throughput transcriptome profiling at the single cell level [51], while simultaneously preserving the spatial location of distinct immune cells in the bulk tumor tissue [52]. This will come with an added

layer of challenges giving rise to the next generation of bioinformatics tools to profile TME beyond those described in this review.

Transcriptomics enabled determination of cell states from the TME

In the methods reviewed here some form of gene expression data in the form of gene expression arrays or RNAseq is required for *in silico* immune cell profiling. Reference profiles can be obtained from studies that measure gene expression in specific isolated and purified specific immune cell populations. Very often the expression pattern of genes across immune cells can also reflect the phylogeny of the immune cell lineage [53], making it informative for elucidating a general immune cell phenotype but more challenging to distinguish between similar immune cell subsets. Transcriptome profiling of immune cell lineages provides opportunities for identifying the wide range of gene signatures that define diverse subsets and functional states during the immune response, above and beyond cluster of differentiation (CD) antigens which are the usual molecular markers used experimentally for immune cell subsets. The number of immune cell types assayed and the number of non-immune cell types used to compare these patterns complicates direct use of these differential expression based signatures. Careful selection of both the methods and the type of datasets compared are needed to allow for the capture of expression differences which are truly distinct to immune cell subsets, as opposed to those expressed widely across the many cell types partaking in general cellular functions. Corrections for external effects, such as differences between and within experimental platforms and protocols used by different labs and experimental setups must also be considered in these workflows. Several groups have tried to reduce these confounding effects by using methods to intelligently combine expression data from multiple studies or by carrying out large scale studies of immune cell lineages to reveal patterns of genes expression that recapitulate the lineages and immune cell differentiation processes [54, 55]. Such approaches have been used to characterize the immune cell expression signatures in peripheral

blood, identifying key transcription factors responsible for lineage commitment of immune cell subsets [56].

One of the largest initiatives for profiling the immune cell lineage is the work being done by The Immunological Genome Consortium (ImmGen), to profile approximately 270 different immune cell subsets in mice [57]. ImmGen is a joint effort between immunologists and computational biologists to transcriptomically profile the murine immune system using carefully controlled methods of sample collection and data analysis. The result is an immune transcriptome data collection platform where careful standardized operating procedures have been applied to result in a high quality dataset that appears relatively complete. The resources from this project offer data that pertain to the activation, development, heterogeneity, and function of the immune cell lineages and their distinct subsets. However, the direct translation of findings from mouse to human, and comprehensive assessment of the accuracy in validated studies of these resources remains a challenge and a barrier to application in cancer immunotherapy trials. One analytical strategy that is emerging approaches this problem by looking at regulatory network modules of genes instead of one-dimensional large gene lists, by identifying co-regulated gene modules [55]. Such a modular (a group of interconnected signaling molecules) approach to characterizing immune cell repertoires in patients represents one new paradigm for the systems level characterization of immune cell subsets. Other gene selection strategies have emerged, whereby text mining the medical literature combined with protein interaction network analysis is used to capture signature genes of distinct immune cell phenotypes and applied to make clinically relevant predictions from the TME [58].

The selection of gene signatures for immune phenotypes has often relied on differential gene expression [59, 60], tissue-specific expression of genes [31, 33, 60], applying cutoffs to expression levels in cellular states [61], or inferring network modules of co-expressed genes [32, 62]. Typically, immune cell specific genes have been identified on the basis of higher gene

expression across all immune cells compared to a selection of non-immune tissues [54]. In general, these approaches can be considered as being primarily centered on the principle that higher expression in an immune cell type is likely to define a cell's distinct properties[2]. However, this feature is only one of many features that may be used to identify immune genes in complex tissue [54]. Methods that computationally dissect the immune cell component of the TME through reliance on gene expression module prediction alone have proved useful. However, more sophisticated tools are needed to accurately differentiate between highly similar immune cell subsets in complex tissue [63, 64].

Overview of computational tools to profile immune cells from the TME

Efforts to quantify rare cell types from bulk heterogeneous cell admixtures of transcriptomes began as early as 2001 [65]. The majority of methods which succeed the framework first proposed by Venet *et al.* are based on the linearity assumption that the expression of each gene in a mixture of cell types is a weighted average of the expression values that would exist for pure populations of those cell types. Some of these early efforts were first developed in model organisms [66]. As with most of the current deconvolution strategies, these early methods were easily applicable to any transcriptomics dataset including studies in cancer [67]. The main drawback to these linear based deconvolution methods is that they can only give relative estimates of composition and thus require a priori knowledge of cell type composition. While the accuracy and robustness of these methods have improved since the effort by Venet et al. these challenges still remain.

Another problem that deconvolution methods, and all other TME profiling methods, must overcome is the lack of well-defined strategies and gold standards to benchmark accuracy. Successful methods must produce results that can be applied to and replicated across multiple platforms and technologies. Special care must be taken to examine the performance across

methods that give high resolution pictures of the immune cell landscape, as it is not possible using current tools to accurately carry out a high resolution analysis, going deep into the immune cell lineage at different stages of activation, effector function, and differentiation for immune cells in a patients repertoire.

Five main categories of bioinformatics tools to extract cell type specific information from tissue derived -omics have been outlined recently in a comprehensive overview [63]. Broadly, these classifications can be simplified into two groups depending on the input they require: tools that start with gene expression matrices of cell types or immune marker gene lists, and those that start only with proportions in the expression data they hope to analyze. Here, we've focused primarily on the former and further abstract the potential application of these tools into two main categories defined according to their utility to profile the TME. The first category consists of tools used to solely predict tumor purity through the profiling of the tumor, immune, and stromal component that stop short of a full cell type resolution. The second category consists of tools designed to predict fractions of more distinct immune cell types from complex tissue. Both have usefulness in aiding to define the immune cell component of the TME to guiding our understanding of the immune-modulatory role of TME in therapy response.

Bioinformatics profiling tumor purity: broad capture of immune and stromal component in the TME

Heterogeneity in the TME presents significant consequences for transcriptomic analyses in cancer biology, leading to increased risk of incorrect or misleading inferences of the immune/stromal composition. Some of the applications to measure and extract signatures of tumor purity have been developed with the focused intent of predicting the non-tumor expression profiles and proportions from a tumor transcriptome. The tools most often used to calculate tumor purity and determine the immune or stromal fraction from a bulk tumor

transcriptome are summarized in Table 1 and Figure 2. Some of these approaches have already been used to find potential therapeutic and prognostic biomarkers [59, 68, 69]. The ESTIMATE method [59] in particular, has been integrated into the The Cancer Genome Atlas's (TCGA) standard pipelines and has been applied to predict the general fractions of immune and stromal components of the tumor, as well as the tumor purity in a sample. The ESTIMATE method integrates publicly available datasets such as the TCGA in its application to the tumor purity prediction, to derive stromal and immune gene signatures, which are then applied to calculate enrichment in tumor samples. The cancer related applications of the tools outlined in Table 1 have proved quite effective in quantifying the stromal and general leukocyte fractions in bulk tumor transcriptomes and relating this information to clinical outcome [70]. ESTIMATE, for example, was recently effectively used to characterize immune subtype signatures among four distinct molecular subtypes of colorectal cancer (microsatellite instability immune, hypermutated, microsatellite unstable, and strong immune activation) [71].

DeMix predicts the proportion of tumor and stromal cell samples using a linear mixture model [72], accounting for the transcripts contributed separately by the stromal component and the epithelial component of a tumor sample. Unlike other traditional deconvolution methods, DeMix does not require advanced stromal and epithelial profiles, but requires at least one gene with expression specific to either cell type. DeMix was applied recently in multi-platform analysis of estimating tumor purity in 333 primary prostate carcinomas, in an analysis that characterized the heterogeneity among primary prostate cancers leading to the identification of potentially therapeutic molecular targets. [73]. PurBayes, a Bayesian mixture modeling approach that estimates tumor purity and sub-clonality, was one of the first of these approaches that was modeled specifically for RNAseq data [74].

Overall, the application of such immune and stromal prediction tools applied to tumor transcriptomes may allow us to interrogate the state-of-the-art molecular classification

paradigms, and move more systematically toward routine profiling of the TME [75]. The use of tumor purity prediction tools outlined in Table 1 has helped to demonstrate that the stromal and immune composition of tumors is complex and highly dynamic. However, they arguably offer a limited perspective on the profiles of distinct and clearly defined immune cell phenotypes potentially present in the TME. In addition, it is important to note that a direct comparison of the performance of these tools is yet subject to an interrogation using the same benchmarked dataset, under control conditions and where the identical raw data processing and bioinformatics pipelines are applied, before the deconvolution step. A summary of the most common bioinformatics approaches capable of quantifying tumor, immune, and stromal proportions are summarized in Table 1.

Tools designed to predict distinct immune cell types from the TME

The prediction of precise signatures of clearly defined immune cell populations in tumors is one of even greater complexity compared to that of normal solid tissues, due to the dynamic presence of clonal and hierarchical subclonal populations. Furthermore, as the TME background composition of normal and tumor cells will vary significantly, the intrinsic heterogeneity of the immune cells themselves adds to the considerable consequences of incorrect or misleading downstream inferences and biological assumptions made in the analyses [76]. Despite the many limitations described above, it is now becoming increasingly possible to computationally predict the proportion of different cell populations from complex tissue admixtures. There are several bioinformatics tools that have emerged in recent years [63, 77] attempting to tackle these problems [64]. As with all TME profiling methods these require some degree of prior knowledge: defined cell phenotypes, the known proportion of the defined cell types in the tumor, or cell type specific reference genes with or without their expression profile (see Table 2 and Figure 2). Several approaches, summarized comprehensively in recent review articles [63, 77] and summarized with methodological categories in Table 2, estimate

relative fractions of individual cell types within a sample using gene expression profiles that are characteristic for each cell type. Although most of the bioinformatics tools for deconvolution require known cell type proportions or marker gene expression profiles, current efforts are seeking to minimize the necessity of prior knowledge [72], by focusing on the identification of distinct immune cell marker genes using data mining approaches [58, 78]. Popular bioinformatics approaches which capitalized on such recent improvements were Dsection [79] and csSAM [80]. Dsection adopts a probabilistic approach, and csSAM uses linear regression. Many approaches in Table 1 may not be optimal when the reference immune cell expression profiles are not representative of the primary populations of cells in a very heterogeneous sample. This is a problem particularly prominent for deconvolution strategies applied in the TME [81]. PERT was a method developed to address this by perturbing all input reference profiles, while the model predicts proportions of the reference populations in a heterogeneous sample [81]. Another recent advancement is the development of semi-supervised machine learning applications to the deconvolution challenges. Nanodissection [82], for example, starts with a small set of marker genes of varying quality and large pre-assembled expression compendia, and utilizes support vector machines within an iterative framework to identify a set of cell type specific transcripts [82]. Cell type specific transcripts identified using this method were successfully used to assay the presence of B cells, cytotoxic T-lymphocytes, T-helper 1, and T-helper 2 cells in breast tissue yielding insights into TP53 mutation driven infiltration differences [83], and lymphocyte invasion in IC10/basal-like breast tumors is associated with wild type TP53]. CIBERSORT, a deconvolution method, also uses a supervised machine-learning framework, in this instance to solve a linear equation for the cell types composing the mixture [64]. CIBERSORT relies on a precisely defined signature gene matrix of immune cell types based on differential gene expression analysis as input, before deconvolving the mixture using a robust implementation of a support vector regression

algorithm. When applied to quantify immune cell subsets in complex transcriptome mixtures, including solid tumors, it claimed to outperform six [79, 81, 84-87] of the total of 11 other methods outlined in Table 2, with respect to noise and other cell lines during *in silico* simulations [64]. CIBERSORT has recently been used in a clinical setting to identify prognostic genes and leukocyte subsets within and across 25 cancer histologies from the TCGA database [83].

Choosing the optimal deconvolution strategy from those outlined in Tables 1 and 2 requires basic understanding of the assumptions of the underlying models, the input required, and the type of output desired. The field at present has primarily concentrated on deconvolution challenges from complex tissue using transcriptomics data. Future interesting developments in the field may offer improved success by offering the integration of other –omics (proteomics, miRNAs ,*etc*) molecular profiles to improve the prediction power of the deconvolution tools .If only a general estimate of tumor purity is desired, one can choose from the any of available methods. When more specific cell type estimates are needed, one must weigh the pros and cons of the input and models, as well as the desired context of the output. The current state-of-the-art TME deconvolution tools to do not cater for detailed predictions on the specific pharmacologically relevant secreted agents (lipids, cytokines, growth factors, *etc*), nor do they predict their capacity to secrete functionally important extracellular vesicles loaded with important pharmacological relevant agents (such as mRNA and miRNA, *etc*). Indeed, deconvolution the pharmacologically relevant signaling systems from the TME remains and critical challenge for future bioinformatics studies in this field [19]. However, this review may be used as a starting point for understanding the basics needed to apply deconvolution to the TME and define these challenges more precisely.

Conclusions

The current state-of-the-art computational predictions of distinct immune cells from the TME can be performed within a certain degree of precision. The results are for the most part constrained by the inputs, *a priori* estimates of the proportional representation of immune cell types in each sample or known marker genes with cell type specificity. Accuracy is limited by the quality of this prior knowledge. The general deconvolution task brings with it algorithmic, computational and experimental validation challenges which current research is actively overcoming to achieve the accurate *in silico* prediction of clearly defined cell types from the TME. Many of these tools have been developed and tested to date on peripheral blood, the primary source of samples in human immunology studies. The tools appear to generalize well to other tissue and data types, with good performance. The recent emergence of single-cell – omics experimental technologies hold great potential to offer deconvolution at the resolution of the high-resolution immune cells repertoires of patients, with applicability to deconvolution to the TME, wherever the single cells can be obtained. Although the current approaches outlined here provide predictions at a tissue-level resolution, a combination of single cell and current tissue level technologies will generate a powerful synergy through the accurate prediction of immune cells at high resolution in the TME without disrupting the important tissue level context of the TME admixture [77].

We now have entered an era in TME biology where we can begin to computationally predict for the presence of distinct immune cell types in the complex heterogeneous tissues of tumors, guiding us toward diagnostic and therapeutic discovery. Methods which attempt the prediction of the network landscape of immune cells in the tumor microenvironment [58, 88-90], may improve our understanding of therapy-responsive and therapy-resistant phenotypes [36, 91, 92]. Moving forward, it will be of great benefit to systematically profile for not only patterns of one

immune cell population, but to capture immune cell networks in tumors with higher resolution;
to facilitate therapeutic discovery of immune-modulatory drugs

References

- [1] Murphy K, Travers P, Walport M, Janeway C. *Janeway's immunobiology*. Garland Science: New York 2008.
- [2] Clancy T, Hovig E. Profiling networks of distinct immune-cells in tumors. *BMC Bioinformatics*, 2016; 17: 263.
- [3] Horowitz A, Strauss-Albee DM, Leipold M, Kubo J, Nemat-Gorgani N, Dogan OC, Dekker CL, Mackey S, Maecker H, Swan GE, Davis MM, Norman PJ, Guethlein LA, Desai M, Parham P, Blish CA. Genetic and environmental determinants of human NK cell diversity revealed by mass cytometry. *Sci Transl Med*, 2013; 5: 208ra145.
- [4] Spitzer MH, Nolan GP. Mass Cytometry: Single Cells, Many Features. *Cell*, 2016; 165: 780-91.
- [5] Saey Y, Gassen SV, Lambrecht BN. Computational flow cytometry: helping to make sense of high-dimensional immunology data. *Nat Rev Immunol*, 2016; 16: 449-62.
- [6] Dunn GP, Old LJ, Schreiber RD. The three Es of cancer immunoediting. *Annu Rev Immunol*, 2004; 22: 329-60.
- [7] Shankaran V, Ikeda H, Bruce AT, White JM, Swanson PE, Old LJ, Schreiber RD. IFN γ and lymphocytes prevent primary tumour development and shape tumour immunogenicity. *Nature*, 2001; 410: 1107-11.
- [8] Dijkstra KK, Voabil P, Schumacher TN, Voest EE. Genomics- and Transcriptomics-Based Patient Selection for Cancer Treatment With Immune Checkpoint Inhibitors: A Review. *JAMA Oncol*, 2016.
- [9] Giraldo NA, Becht E, Remark R, Damotte D, Sautes-Fridman C, Fridman WH. The immune contexture of primary and metastatic human tumours. *Curr Opin Immunol*, 2014; 27: 8-15.
- [10] Parker KH, Beury DW, Ostrand-Rosenberg S. Myeloid-Derived Suppressor Cells: Critical Cells Driving Immune Suppression in the Tumor Microenvironment. *Adv Cancer Res*, 2015; 128: 95-139.
- [11] Malek E, de Lima M, Letterio JJ, Kim BG, Finke JH, Driscoll JJ, Giralto SA. Myeloid-derived suppressor cells: The green light for myeloma immune escape. *Blood Rev*, 2016.
- [12] Blankenstein T, Coulie PG, Gilboa E, Jaffee EM. The determinants of tumour immunogenicity. *Nat Rev Cancer*, 2012; 12: 307-13.
- [13] Hamid O, Robert C, Daud A, Hodi FS, Hwu WJ, Kefford R, Wolchok JD, Hersey P, Joseph RW, Weber JS, Dronca R, Gangadhar TC, Patnaik A, Zarour H, Joshua AM, Gergich K, Elassaiss-Schaap J, Algazi A, Mateus C, Boasberg P, Tumei PC, Chmielowski B, Ebbinghaus SW, Li XN, Kang SP, Ribas A. Safety and tumor responses with lambrolizumab (anti-PD-1) in melanoma. *N Engl J Med*, 2013; 369: 134-44.
- [14] Small EJ, Tchekmedyian NS, Rini BI, Fong L, Lowy I, Allison JP. A pilot trial of CTLA-4 blockade with human anti-CTLA-4 in patients with hormone-refractory prostate cancer. *Clin Cancer Res*, 2007; 13: 1810-5.
- [15] Sharma P, Allison JP. Immune checkpoint targeting in cancer therapy: toward combination strategies with curative potential. *Cell*, 2015; 161: 205-14.
- [16] Luke JJ, Ott PA. PD-1 pathway inhibitors: the next generation of immunotherapy for advanced melanoma. *Oncotarget*, 2015; 6: 3479-92.
- [17] Klemm F, Joyce JA. Microenvironmental regulation of therapeutic response in cancer. *Trends Cell Biol*, 2015; 25: 198-213.
- [18] Dunn GP, Bruce AT, Ikeda H, Old LJ, Schreiber RD. Cancer immunoediting: from immunosurveillance to tumor escape. *Nat Immunol*, 2002; 3: 991-8.
- [19] Quigley DA, Kristensen V. Predicting prognosis and therapeutic response from interactions between lymphocytes and tumor cells. *Mol Oncol*, 2015; 9: 2054-62.
- [20] Barber DL, Wherry EJ, Masopust D, Zhu B, Allison JP, Sharpe AH, Freeman GJ, Ahmed R. Restoring function in exhausted CD8 T cells during chronic viral infection. *Nature*, 2006; 439: 682-7.

- [21] Dong H, Strome SE, Salomao DR, Tamura H, Hirano F, Flies DB, Roche PC, Lu J, Zhu G, Tamada K, Lennon VA, Celis E, Chen L. Tumor-associated B7-H1 promotes T-cell apoptosis: a potential mechanism of immune evasion. *Nat Med*, 2002; 8: 793-800.
- [22] Topalian SL, Hodi FS, Brahmer JR, Gettinger SN, Smith DC, McDermott DF, Powderly JD, Carvajal RD, Sosman JA, Atkins MB, Leming PD, Spigel DR, Antonia SJ, Horn L, Drake CG, Pardoll DM, Chen L, Sharfman WH, Anders RA, Taube JM, McMiller TL, Xu H, Korman AJ, Jure-Kunkel M, Agrawal S, McDonald D, Kollia GD, Gupta A, Wigginton JM, Sznol M. Safety, activity, and immune correlates of anti-PD-1 antibody in cancer. *N Engl J Med*, 2012; 366: 2443-54.
- [23] Topalian SL, Drake CG, Pardoll DM. Targeting the PD-1/B7-H1(PD-L1) pathway to activate anti-tumor immunity. *Curr Opin Immunol*, 2012; 24: 207-12.
- [24] Taube JM, Anders RA, Young GD, Xu H, Sharma R, McMiller TL, Chen S, Klein AP, Pardoll DM, Topalian SL, Chen L. Colocalization of inflammatory response with B7-h1 expression in human melanocytic lesions supports an adaptive resistance mechanism of immune escape. *Sci Transl Med*, 2012; 4: 127ra37.
- [25] Tumeh PC, Harview CL, Yearley JH, Shintaku IP, Taylor EJ, Robert L, Chmielowski B, Spasic M, Henry G, Ciobanu V, West AN, Carmona M, Kivork C, Seja E, Cherry G, Gutierrez AJ, Grogan TR, Mateus C, Tomasic G, Glaspy JA, Emerson RO, Robins H, Pierce RH, Elashoff DA, Robert C, Ribas A. PD-1 blockade induces responses by inhibiting adaptive immune resistance. *Nature*, 2014; 515: 568-71.
- [26] Herbst RS, Soria JC, Kowanetz M, Fine GD, Hamid O, Gordon MS, Sosman JA, McDermott DF, Powderly JD, Gettinger SN, Kohrt HE, Horn L, Lawrence DP, Rost S, Leabman M, Xiao Y, Mokatrinn A, Koeppen H, Hegde PS, Mellman I, Chen DS, Hodi FS. Predictive correlates of response to the anti-PD-L1 antibody MPDL3280A in cancer patients. *Nature*, 2014; 515: 563-7.
- [27] Jochems C, Schlom J. Tumor-infiltrating immune cells and prognosis: the potential link between conventional cancer therapy and immunity. *Exp Biol Med (Maywood)*, 2011; 236: 567-79.
- [28] Smyth MJ, Ngiow SF, Ribas A, Teng MW. Combination cancer immunotherapies tailored to the tumour microenvironment. *Nat Rev Clin Oncol*, 2016; 13: 143-58.
- [29] Sathyanarayanan V, Neelapu SS. Cancer immunotherapy: Strategies for personalization and combinatorial approaches. *Mol Oncol*, 2015; 9: 2043-53.
- [30] Galon J, Costes A, Sanchez-Cabo F, Kirilovsky A, Mlecnik B, Lagorce-Pages C, Tosolini M, Camus M, Berger A, Wind P, Zinzindohoue F, Bruneval P, Cugnenc PH, Trajanoski Z, Fridman WH, Pages F. Type, density, and location of immune cells within human colorectal tumors predict clinical outcome. *Science*, 2006; 313: 1960-4.
- [31] Bindea G, Mlecnik B, Tosolini M, Kirilovsky A, Waldner M, Obenauf AC, Angell H, Fredriksen T, Lafontaine L, Berger A, Bruneval P, Fridman WH, Becker C, Pages F, Speicher MR, Trajanoski Z, Galon J. Spatiotemporal dynamics of intratumoral immune cells reveal the immune landscape in human cancer. *Immunity*, 2013; 39: 782-95.
- [32] Linsley PS, Speake C, Whalen E, Chaussabel D. Copy number loss of the interferon gene cluster in melanomas is linked to reduced T cell infiltrate and poor patient prognosis. *PLoS One*, 2014; 9: e109760.
- [33] Angelova M, Charoentong P, Hackl H, Fischer ML, Snajder R, Krogsdam AM, Waldner MJ, Bindea G, Mlecnik B, Galon J, Trajanoski Z. Characterization of the immunophenotypes and antigenomes of colorectal cancers reveals distinct tumor escape mechanisms and novel targets for immunotherapy. *Genome Biol*, 2015; 16: 64.
- [34] Mlecnik B, Bindea G, Angell HK, Maby P, Angelova M, Tougeron D, Church SE, Lafontaine L, Fischer M, Fredriksen T, Sasso M, Bilocq AM, Kirilovsky A, Obenauf AC, Hamieh M, Berger A, Bruneval P, Tuech JJ, Sabourin JC, Le Pessot F, Mauillon J, Rafii A, Laurent-Puig P, Speicher MR, Trajanoski Z, Michel P, Sesboue R, Frebourg T, Pages F, Valge-Archer V, Latouche JB, Galon J. Integrative Analyses of Colorectal Cancer Show Immunoscore Is a Stronger Predictor of Patient Survival Than Microsatellite Instability. *Immunity*, 2016; 44: 698-711.

- [35] Angell H, Galon J. From the immune contexture to the Immunoscore: the role of prognostic and predictive immune markers in cancer. *Curr Opin Immunol*, 2013; 25: 261-7.
- [36] Chaussabel D, Baldwin N. Democratizing systems immunology with modular transcriptional repertoire analyses. *Nat Rev Immunol*, 2014; 14: 271-80.
- [37] Benoist C, Germain RN, Mathis D. A plaidoyer for 'systems immunology'. *Immunol Rev*, 2006; 210: 229-34.
- [38] Janes KA, Wang CC. Bringing systems biology to cancer, immunology and infectious disease. *Genome Biol*, 2014; 15: 407.
- [39] Kidd BA, Peters LA, Schadt EE, Dudley JT. Unifying immunology with informatics and multiscale biology. *Nat Immunol*, 2014; 15: 118-27.
- [40] Maecker HT, McCoy JP, Nussenblatt R. Standardizing immunophenotyping for the Human Immunology Project (vol 12, pg 191, 2012). *Nature Reviews Immunology*, 2012; 12.
- [41] Galon J, Mlecnik B, Bindea G, Angell HK, Berger A, Lagorce C, Lugli A, Zlobec I, Hartmann A, Bifulco C, Nagtegaal ID, Palmqvist R, Masucci GV, Botti G, Tatangelo F, Delrio P, Maio M, Laghi L, Grizzi F, Asslaber M, D'Arrigo C, Vidal-Vanaclocha F, Zavadova E, Chouchane L, Ohashi PS, Hafezi-Bakhtiari S, Wouters BG, Roehrl M, Nguyen L, Kawakami Y, Hazama S, Okuno K, Ogino S, Gibbs P, Waring P, Sato N, Torigoe T, Itoh K, Patel PS, Shukla SN, Wang Y, Kopetz S, Sinicrope FA, Scripcariu V, Ascierto PA, Marincola FM, Fox BA, Pages F. Towards the introduction of the 'Immunoscore' in the classification of malignant tumours. *J Pathol*, 2014; 232: 199-209.
- [42] Chaussabel D, Pascual V, Banchereau J. Assessing the human immune system through blood transcriptomics. *BMC Biol*, 2010; 8: 84.
- [43] Gajewski TF. The Next Hurdle in Cancer Immunotherapy: Overcoming the Non-T-Cell-Inflamed Tumor Microenvironment. *Semin Oncol*, 2015; 42: 663-71.
- [44] Galon J, Angell HK, Bedognetti D, Marincola FM. The continuum of cancer immunosurveillance: prognostic, predictive, and mechanistic signatures. *Immunity*, 2013; 39: 11-26.
- [45] Kristensen VN, Vaske CJ, Ursini-Siegel J, Van Loo P, Nordgard SH, Sachidanandam R, Sorlie T, Warnberg F, Haakensen VD, Helland A, Naume B, Perou CM, Haussler D, Troyanskaya OG, Borresen-Dale AL. Integrated molecular profiles of invasive breast tumors and ductal carcinoma in situ (DCIS) reveal differential vascular and interleukin signaling. *Proc Natl Acad Sci U S A*, 2012; 109: 2802-7.
- [46] Bindea G, Mlecnik B, Angell HK, Galon J. The immune landscape of human tumors: Implications for cancer immunotherapy. *Oncoimmunology*, 2014; 3: e27456.
- [47] Fridman WH, Pages F, Sautes-Fridman C, Galon J. The immune contexture in human tumours: impact on clinical outcome. *Nat Rev Cancer*, 2012; 12: 298-306.
- [48] de Visser KE, Eichten A, Coussens LM. Paradoxical roles of the immune system during cancer development. *Nat Rev Cancer*, 2006; 6: 24-37.
- [49] Grivennikov SI, Greten FR, Karin M. Immunity, inflammation, and cancer. *Cell*, 2010; 140: 883-99.
- [50] Shalapour S, Karin M. Immunity, inflammation, and cancer: an eternal fight between good and evil. *J Clin Invest*, 2015; 125: 3347-55.
- [51] Chattopadhyay PK, Gierahn TM, Roederer M, Love JC. Single-cell technologies for monitoring immune systems. *Nat Immunol*, 2014; 15: 128-35.
- [52] Crosetto N, Bienko M, van Oudenaarden A. Spatially resolved transcriptomics and beyond. *Nat Rev Genet*, 2015; 16: 57-66.
- [53] Clancy T, Hovig E. Differential protein network analysis of the immune cell lineage. *Biomed Res Int*, 2014; 2014: 363408.
- [54] Abbas AR, Baldwin D, Ma Y, Ouyang W, Gurney A, Martin F, Fong S, van Lookeren Campagne M, Godowski P, Williams PM, Chan AC, Clark HF. Immune response in silico (IRIS): immune-specific genes identified from a compendium of microarray expression data. *Genes Immun*, 2005; 6: 319-31.
- [55] Chaussabel D, Quinn C, Shen J, Patel P, Glaser C, Baldwin N, Stichweh D, Blankenship D, Li L, Munagala I, Bennett L, Allantaz F, Mejias A, Ardura M, Kaizer E, Monnet L, Allman W, Randall H,

- Johnson D, Lanier A, Punaro M, Wittkowski KM, White P, Fay J, Klintmalm G, Ramilo O, Palucka AK, Banchereau J, Pascual V. A modular analysis framework for blood genomics studies: application to systemic lupus erythematosus. *Immunity*, 2008; 29: 150-64.
- [56] Watkins NA, Gusnanto A, de Bono B, De S, Miranda-Saavedra D, Hardie DL, Angenent WG, Attwood AP, Ellis PD, Erber W, Foad NS, Garner SF, Isacke CM, Jolley J, Koch K, Macaulay IC, Morley SL, Rendon A, Rice KM, Taylor N, Thijssen-Timmer DC, Tijssen MR, van der Schoot CE, Wernisch L, Winzer T, Dudbridge F, Buckley CD, Langford CF, Teichmann S, Gottgens B, Ouwehand WH, Bloodomics C. A HaemAtlas: characterizing gene expression in differentiated human blood cells. *Blood*, 2009; 113: e1-9.
- [57] Kim CC, Lanier LL. Beyond the transcriptome: completion of act one of the Immunological Genome Project. *Curr Opin Immunol*, 2013; 25: 593-7.
- [58] Clancy T, Hovig E. Profiling networks of distinct immune-cells in tumors. *BMC Bioinformatics*, 2016; in press.
- [59] Yoshihara K, Shahmoradgoli M, Martinez E, Vegesna R, Kim H, Torres-Garcia W, Trevino V, Shen H, Laird PW, Levine DA, Carter SL, Getz G, Stemke-Hale K, Mills GB, Verhaak RG. Inferring tumour purity and stromal and immune cell admixture from expression data. *Nat Commun*, 2013; 4: 2612.
- [60] Bolen CR, Uduman M, Kleinstein SH. Cell subset prediction for blood genomic studies. *BMC Bioinformatics*, 2011; 12: 258.
- [61] Chambers SM, Boles NC, Lin KY, Tierney MP, Bowman TV, Bradfute SB, Chen AJ, Merchant AA, Sirin O, Weksberg DC, Merchant MG, Fisk CJ, Shaw CA, Goodell MA. Hematopoietic fingerprints: an expression database of stem cells and their progeny. *Cell Stem Cell*, 2007; 1: 578-91.
- [62] Linsley PS, Chaussabel D, Speake C. The Relationship of Immune Cell Signatures to Patient Survival Varies within and between Tumor Types. *PLoS One*, 2015; 10: e0138726.
- [63] Shen-Orr SS, Gaujoux R. Computational deconvolution: extracting cell type-specific information from heterogeneous samples. *Curr Opin Immunol*, 2013; 25: 571-8.
- [64] Newman AM, Liu CL, Green MR, Gentles AJ, Feng W, Xu Y, Hoang CD, Diehn M, Alizadeh AA. Robust enumeration of cell subsets from tissue expression profiles. *Nat Methods*, 2015; 12: 453-7.
- [65] Venet D, Pecasse F, Maenhaut C, Bersini H. Separation of samples into their constituents using gene expression data. *Bioinformatics*, 2001; 17 Suppl 1: S279-87.
- [66] Lu P, Nakorchevskiy A, Marcotte EM. Expression deconvolution: a reinterpretation of DNA microarray data reveals dynamic changes in cell populations. *Proc Natl Acad Sci U S A*, 2003; 100: 10370-5.
- [67] Buess M, Nuyten DS, Hastie T, Nielsen T, Pesich R, Brown PO. Characterization of heterotypic interaction effects in vitro to deconvolute global gene expression profiles in cancer. *Genome Biol*, 2007; 8: R191.
- [68] Shen Q, Hu J, Jiang N, Hu X, Luo Z, Zhang H. contamDE: differential expression analysis of RNA-seq data for contaminated tumor samples. *Bioinformatics*, 2016; 32: 705-12.
- [69] Anghel CV, Quon G, Haider S, Nguyen F, Deshwar AG, Morris QD, Boutros PC. ISOpureR: an R implementation of a computational purification algorithm of mixed tumour profiles. *BMC Bioinformatics*, 2015; 16: 156.
- [70] Isella C, Terrasi A, Bellomo SE, Petti C, Galatola G, Muratore A, Mellano A, Senetta R, Cassenti A, Sonetto C, Inghirami G, Trusolino L, Fekete Z, De Ridder M, Cassoni P, Storme G, Bertotti A, Medico E. Stromal contribution to the colorectal cancer transcriptome. *Nat Genet*, 2015; 47: 312-9.
- [71] Guinney J, Dienstmann R, Wang X, de Reynies A, Schlicker A, Soneson C, Marisa L, Roepman P, Nyamundanda G, Angelino P, Bot BM, Morris JS, Simon IM, Gerster S, Fessler E, De Sousa EMF, Missiaglia E, Ramay H, Barras D, Homicsko K, Maru D, Manyam GC, Broom B, Boige V, Perez-Villamil B, Laderas T, Salazar R, Gray JW, Hanahan D, Tabernero J, Bernards R, Friend SH,

- Laurent-Puig P, Medema JP, Sadanandam A, Wessels L, Delorenzi M, Kopetz S, Vermeulen L, Tejpar S. The consensus molecular subtypes of colorectal cancer. *Nat Med*, 2015; 21: 1350-6.
- [72] Ahn J, Yuan Y, Parmigiani G, Suraokar MB, Diao L, Wistuba, II, Wang W. DeMix: deconvolution for mixed cancer transcriptomes using raw measured data. *Bioinformatics*, 2013; 29: 1865-71.
- [73] Cancer Genome Atlas Research N. The Molecular Taxonomy of Primary Prostate Cancer. *Cell*, 2015; 163: 1011-25.
- [74] Larson NB, Fridley BL. PurBayes: estimating tumor cellularity and subclonality in next-generation sequencing data. *Bioinformatics*, 2013; 29: 1888-9.
- [75] Dunne PD, McArt DG, Bradley CA, O'Reilly PG, Barrett HL, Cummins R, O'Grady T, Arthur K, Loughrey M, Allen WL, McDade S, Waugh DJ, Hamilton PW, Longley DB, Kay EW, Johnston PG, Lawler M, Salto-Tellez M, Van Schaeybroeck S. Challenging the cancer molecular stratification dogma: Intratumoral heterogeneity undermines consensus molecular subtypes and potential diagnostic value in colorectal cancer. *Clin Cancer Res*, 2016.
- [76] Zhao Y, Simon R. Gene expression deconvolution in clinical samples. *Genome Med*, 2010; 2: 93.
- [77] Yadav VK, De S. An assessment of computational methods for estimating purity and clonality using genomic data derived from heterogeneous tumor tissue samples. *Brief Bioinform*, 2015; 16: 232-41.
- [78] Clancy T, Pedicini M, Castiglione F, Santoni D, Nygaard V, Lavelle TJ, Benson M, Hovig E. Immunological network signatures of cancer progression and survival. *BMC Med Genomics*, 2011; 4: 28.
- [79] Erkkila T, Lehmusvaara S, Ruusuvaara P, Visakorpi T, Shmulevich I, Lahdesmaki H. Probabilistic analysis of gene expression measurements from heterogeneous tissues. *Bioinformatics*, 2010; 26: 2571-7.
- [80] Shen-Orr SS, Tibshirani R, Khatri P, Bodian DL, Staedtler F, Perry NM, Hastie T, Sarwal MM, Davis MM, Butte AJ. Cell type-specific gene expression differences in complex tissues. *Nat Methods*, 2010; 7: 287-9.
- [81] Qiao W, Quon G, Cszasz E, Yu M, Morris Q, Zandstra PW. PERT: a method for expression deconvolution of human blood samples from varied microenvironmental and developmental conditions. *PLoS Comput Biol*, 2012; 8: e1002838.
- [82] Ju W, Greene CS, Eichinger F, Nair V, Hodgin JB, Bitzer M, Lee YS, Zhu Q, Kehata M, Li M, Jiang S, Rastaldi MP, Cohen CD, Troyanskaya OG, Kretzler M. Defining cell-type specificity at the transcriptional level in human disease. *Genome Res*, 2013; 23: 1862-73.
- [83] Quigley D, Silwal-Pandit L, Dannenfelser R, Langerod A, Volland HK, Vaske C, Siegel JU, Troyanskaya O, Chin SF, Caldas C, Balmain A, Borresen-Dale AL, Kristensen V. Lymphocyte Invasion in IC10/Basal-Like Breast Tumors Is Associated with Wild-Type TP53. *Mol Cancer Res*, 2015; 13: 493-501.
- [84] Abbas AR, Wolslegel K, Seshasayee D, Modrusan Z, Clark HF. Deconvolution of blood microarray data identifies cellular activation patterns in systemic lupus erythematosus. *PLoS One*, 2009; 4: e6098.
- [85] Gong T, Szustakowski JD. DeconRNASeq: a statistical framework for deconvolution of heterogeneous tissue samples based on mRNA-Seq data. *Bioinformatics*, 2013; 29: 1083-5.
- [86] Liebner DA, Huang K, Parvin JD. MMAD: microarray microdissection with analysis of differences is a computational tool for deconvoluting cell type-specific contributions from tissue samples. *Bioinformatics*, 2014; 30: 682-9.
- [87] Li Y, Xie X. A mixture model for expression deconvolution from RNA-seq in heterogeneous tissues. *BMC Bioinformatics*, 2013; 14 Suppl 5: S11.
- [88] Thomas A, Giaccone G. Why has active immunotherapy not worked in lung cancer? *Ann Oncol*, 2015.
- [89] Maio M, Grob JJ, Aamdal S, Bondarenko I, Robert C, Thomas L, Garbe C, Chiarion-Sileni V, Testori A, Chen TT, Tschaike M, Wolchok JD. Five-year survival rates for treatment-naive patients with advanced melanoma who received ipilimumab plus dacarbazine in a phase III trial. *J Clin Oncol*, 2015; 33: 1191-6.

- [90] Schadendorf D, Hodi FS, Robert C, Weber JS, Margolin K, Hamid O, Patt D, Chen TT, Berman DM, Wolchok JD. Pooled Analysis of Long-Term Survival Data From Phase II and Phase III Trials of Ipilimumab in Unresectable or Metastatic Melanoma. *J Clin Oncol*, 2015; 33: 1889-94.
- [91] Gustafson MP, Lin Y, LaPlant B, Liwski CJ, Maas ML, League SC, Bauer PR, Abraham RS, Tollefson MK, Kwon ED, Gastineau DA, Dietz AB. Immune monitoring using the predictive power of immune profiles. *J Immunother Cancer*, 2013; 1: 7.
- [92] Davis MM. A Prescription for Human Immunology. *Immunity*, 2008; 29: 835-838.
- [93] Wang N, Gong T, Clarke R, Chen L, Shih le M, Zhang Z, Levine DA, Xuan J, Wang Y. UNDO: a Bioconductor R package for unsupervised deconvolution of mixed gene expressions in tumor samples. *Bioinformatics*, 2015; 31: 137-9.
- [94] Quon G, Morris Q. ISOLATE: a computational strategy for identifying the primary origin of cancers using high-throughput sequencing. *Bioinformatics*, 2009; 25: 2882-9.
- [95] Kuhn A, Thu D, Waldvogel HJ, Faull RL, Luthi-Carter R. Population-specific expression analysis (PSEA) reveals molecular changes in diseased brain. *Nat Methods*, 2011; 8: 945-7.
- [96] Zuckerman NS, Noam Y, Goldsmith AJ, Lee PP. A self-directed method for cell-type identification and separation of gene expression microarrays. *PLoS Comput Biol*, 2013; 9: e1003189.

Table 1: Deconvolution methods applicable to profile tumor purity

Bioinformatics Tool	Algorithm class	Input data type	Output	Availability	Reference
UNDO	Linear latent variable model	Expression profiles (optional matched normal)	Tumor-stroma mixed gene expressions	R package	[93]
contamDE	Linear equations	Expression profile (matched and unmatched normal sample)	Proportion estimation and differential expression of tumor purity	R package	[68]
ISOpureR	Linear equations and statistical model	Expression profile (matched and unmatched normal sample)	Proportion estimation and differential expression of tumor purity	R & MATLAB implementations	[69]
ISOLATE	Precursor to ISOpureR (generative statistical models)	Expression profile (matched and unmatched normal sample)	Proportion estimation and differential expression of tumor purity	Precursor to ISOpureR implementation	[94]
ESTIMATE	Gene set enrichment analysis	Expression profiles of tumor, stroma and immune signature genes (optional matched normal)	Tumor purity scores	R implementation	[59]
DeMix	Statistical (maximum likelihood)	Expression profiles of tumor, stroma and immune signature genes (optional matched normal)	Proportion estimation and differential expression of tumor purity	R implementation	[72]
PurBayes	Bayesian statistics	Expression profiles of tumor, stroma and matched normal signature genes	Tumor purity scores	R implementation	[74]

Table 2: Deconvolution methods also applicable to the identification of distinct immune cell subsets

Bioinformatics Tool	Algorithm class	Input data type	Output	Availability	Reference
DeconRNASeq	Linear Equations	Reference expression profiles (RNASeq)	Cell proportions	R- package	[85]
PSEA	Linear regression	Marker genes	Cell proportions	R- package	[95]
csSAM	Linear Regression	Known cell proportions	Cell-specific expression profiles	R- package	[80]
NMF	Unsupervised NMF + Information theory	Reference expression profiles, Prior information on cell proportions	Cell proportions, Cell-specific expression profiles	Matlab	[96]
DSA	Linear equations	Marker genes (RNASeq)	Cell proportions, Cell-specific expression profiles	R-package	[87]
MMAD	Linear model	Reference expression profiles Prior information on cell proportions	Cell proportions, Cell-specific expression profiles	MATLAB	[86]
PERT	Probabilistic Non-negative maximum likelihood model, on linear models	Reference expression profiles	Cell proportions	Octave	[81]

LLSR	Linear least squares regression	Reference expression profiles	Cell proportions	R	[84]
CIBERSORT	Support regression machines	Reference expression profiles	Cell proportions	Java, R and online tool	[64]
Nanodissection	Support vector machines	Reference expression profiles (training data)	Cell/Tissue specific gene expression	Online tool, C++ implementation	[82]
Dsection	Probabilistic method Bayesian framework	Reference expression profiles, Estimated info on cell proportions	Cell proportions, Cell-specific expression profiles	Online tool	[79]

Figure 1: Functions of the microenvironment in tumor development. A beneficial microenvironment is dominated by IFN γ , TNF α , and IL12 signaling, infiltrating CD4 $^+$ T cells, and tumor eliminating CD8 $^+$ T cells, NK, and M1 macrophages. In a tumor-promoting environment TGF β , IL10, IL13, and PDL1 signaling increase recruitment of regulatory T cells, which block the infiltration of cytotoxic NK and CD8 $^+$ T cells, increasing tumor cell proliferation.

Figure 2: General workflow of bioinformatics deconvolutions strategies from complex heterogeneous tumor tissue: the many classes of tools categorized in Table 1 may be further simplified into two main application categories. (1) Bioinformatics strategies to capture general signatures of tumor purity relative to general immune component and stromal purity. Generally the input into these tools requires expression profiles of the tumor, immune or stromal components being analyzed. (2) There is wide array of tools which address deconvolution of complex tissue with the goal of predicting the fraction of distinct cellular phenotypes, or clearly defined immune cell types such as T Cells, macrophages, B Cells, and Natural Killer cells, *etc.*

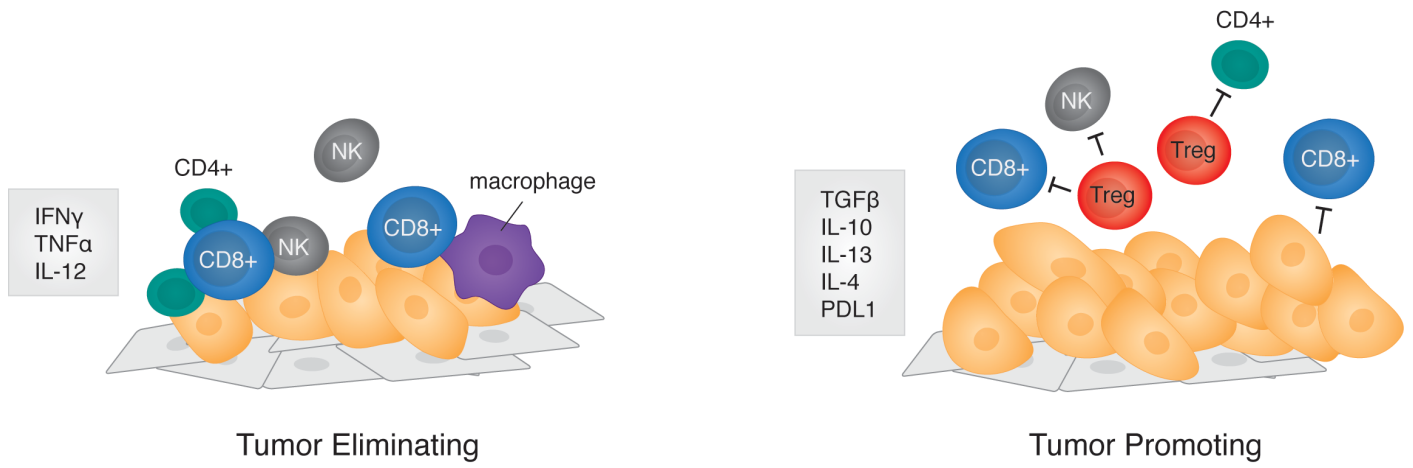


Figure 1

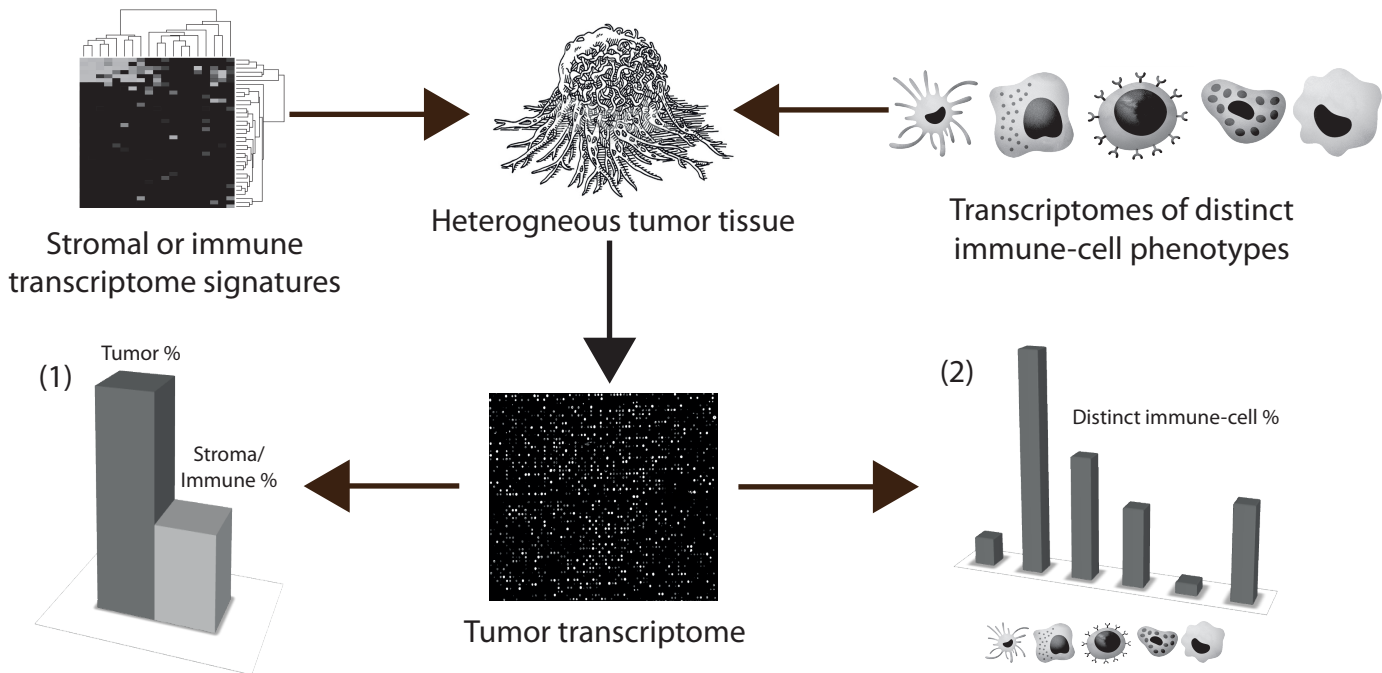


Table 1: Deconvoluting tumor and general stromal or immune cells

Table 2: Deconvoluting proportions of distinct immune-cells

Figure 2