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 **Special section: Cancer immunology: Genomics & biomarkers**

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

For a complete overview see the Issue and the Editorial

Available online 30th June 2016

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

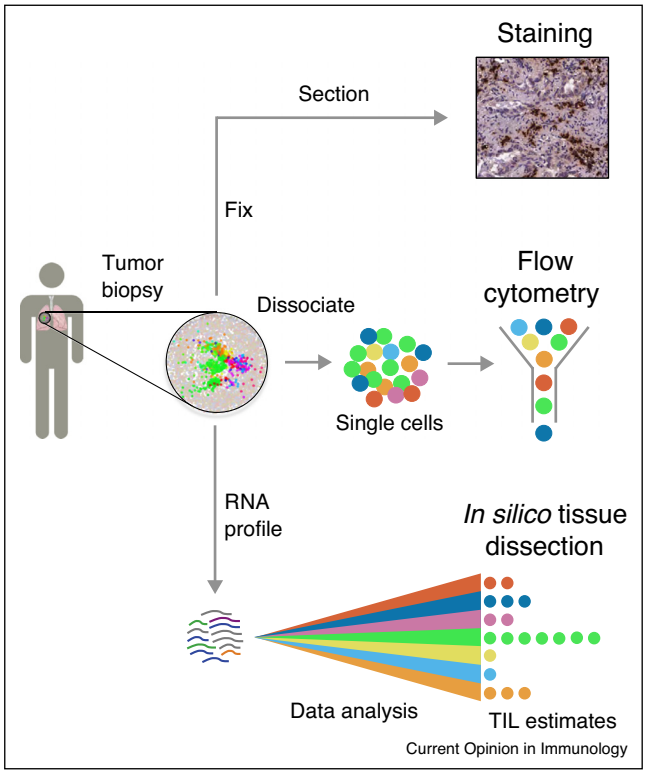
Introduction

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•].

Table 1

Feature comparison of selected *in silico* tissue dissection methods for TIL characterization in bulk tumor transcriptomes. Only deconvolution methods that infer non-negative cell type proportions are shown. The deconvolution methods were evaluated with respect to unknown content, random noise, and resolution of closely related cell types in prior work [41•], and the results are reproduced here. 'Single sample analysis' refers to methods that estimate cell type proportions in a manner that is independent of the combination of mixtures analyzed. Y, yes; N, no; ND, not determined; NA, not applicable; SPEC, Subset Prediction from Enrichment Correlation; ssGSEA, Single Sample Gene Set Enrichment Analysis; DSA, Digital Sorting Algorithm; MMAD, Microarray Microdissection with Analysis of Differences; LLSR, Linear Least Squares Regression (non-negative fractions removed as in [24••]); QP, Quadratic Programming.

Class	Method	Input: Marker genes	Input: Signature matrix	Output: Gene enrichment	Output: Cell type proportions	Robust to >50% unknown content?	Robust to random noise or multi-cell type perturbation?	Marker gene expression obtained from mixture samples?	Resolution of closely related cell types?	Significance analysis for gene enrichment or cell type proportions?	Single sample analysis?	Reference	Availability
Enrichment methods	Cluster analysis	NA	NA	NA	NA	ND	ND	NA	N	NA	N	e.g., [23]	NA
	TIL meta-genes	Y	N	Y	N	ND	ND	Y	N	NA	Y	e.g., [31]	NA
	SPEC	Y	N	Y	N	ND	ND	Y	N	Y	Y	[48]	R: SPEC clip.med.yale.edu/SPEC
	ssGSEA	Y	N	Y	N	ND	ND	Y	N	Y	Y	[47•]	GenePattern, R: GSVA
Deconvolution methods	DSA	Y	N	N	Y	N	N	Y	N	N	N	[17]	R: CellMix or DSA https://github.com/zhandong/DSA
	MMAD	Y	Y	N	Y	N	N	Y	N	N	Y	[37]	Matlab http://sourceforge.net/projects/mmad/
	CIBERSORT	N	Y	N	Y	Y	Y	N	Y	Y	Y	[41•]	Online, R, Java http://cibersort.stanford.edu
	LLSR	N	Y	N	Y	N	N	N	N	N	Y	[24••]	R: CellMix
	PERT	N	Y	N	Y	N	Y	N	N	N	N	[30]	Octave code
	QP	N	Y	N	Y	N	N	N	N	N	Y	[28•]	R: CellMix

Several groups have used methods that explicitly calculate a measure of leukocyte enrichment within a transcriptome sample [8••,42,46,47•,48]. Because these methods are rank-based, they are well suited for data from multiple transcriptome measurement platforms and robust to both technical and biological noise [47•,48]. For example, Single Sample Gene Set Enrichment Analysis (ssGSEA) is the main algorithm within ESTIMATE, a cross-platform method for inferring immune and stromal content within tumor GEPs [34]. Angelova and colleagues employed pre-ranked GSEA to analyze gene sets from 28 TIL subsets for survival associations in patients with colorectal cancer [46]. Rooney and colleagues employed ssGSEA for a systematic analysis of TIL-associated GEPs across diverse cancer types, but noted that ssGSEA results should not be interpreted as cell type proportions [8••]. Indeed, enrichment methods may have difficulty discriminating some TIL subsets owing to non-specific marker genes. Despite this caveat, enrichment approaches are simple to use, robust to platform-specific noise, and effective for inferring the presence of immune subsets with highly distinct expression signatures.

Gene expression deconvolution

Gene expression deconvolution is an emerging technique for cell composition analysis, including for characterizing TIL composition of bulk tumors. Analogous to '*in silico* flow cytometry,' deconvolution methods aim to computationally resolve a mixed GEP into its component cell types, thereby performing a virtual tissue dissection. As a result, these approaches can directly infer cell type proportions in complex tissues. Expression deconvolution methods generally require an input 'signature matrix' **G** consisting of marker genes and their expression values for several cell types of interest. A biological mixture **m** can then be modeled as a system of linear equations in which $\mathbf{m} = \mathbf{G} \times \mathbf{f}$, where **f** is a vector containing the fraction of each cell subset from **G** in **m**. While methods have been proposed to determine **f**, **G** or both [16••,17,21•,22,24••,25–27,28•,29,30,32,35–37,39•,41•,49], many approaches estimate **f** given **m** and **G**. Here, we focus on deconvolution techniques that can enumerate TIL proportions (**f**). For a discussion of other methods, including those that infer cell-type specific gene expression differences when cell fractions (**f**) are known [26], we refer the reader to [16••].

Deconvolution methods that focus on leukocytes were originally developed to analyze peripheral blood, where the majority of mixture content can be explained by a blend of normal immune reference profiles. For example, in an influential study, Abbas and colleagues established an iterative ordinary least-squares method to solve for **f** along with a strategy for signature matrix creation [24••]. They compiled a signature matrix containing 18 variably purified leukocyte subsets [50] and applied it to reveal new insights into Systemic Lupus Erythematosus (SLE) by whole blood deconvolution, including the specific expansion and activation of monocytes, NK cells, and T helper cells in SLE. Gong and colleagues extended this work using quadratic programming to optimally solve for **f** while enforcing non-negativity constraints (i.e., cell fractions ≥ 0). The authors demonstrated advantages of this approach for deconvolution of clinical blood samples, including for accurately measuring changes in leukocyte trafficking associated with Fingolomid (FTY720) therapy [28•]. These 'first generation' methods were instrumental in establishing deconvolution as a viable approach for estimating leukocyte proportions from RNA admixtures.

Cell deconvolution for solid tumors

The phenotypic and compositional diversity of leukocytes in solid tumors presents several challenges for deconvolution methods. First, solid tumors are heterogeneous mixtures of cell types, many of which will not be 'known' to the deconvolution engine (e.g., malignant cells or rare mesenchymal elements). Thus, ideal methods should be robust to unknown mixture content and technical/biological noise. Second, given the wide assortment of TILs with potential clinical utility, the method should scale to multiple cell types, including functionally defined leukocyte subsets with similar expression signatures. Third, because overall immune content can be highly variable in solid tissues, the method should produce a significance metric to help evaluate confidence in the deconvolution results.

In recent years, several groups have implemented methods that reduce the impact of technical/biological noise on deconvolution results [17,29,30,37,41•]. For example, Digital Sorting Algorithm (DSA) [17] and Microarray Microdissection with Analysis of Differences (MMAD) [37] can each infer the expression levels of user-specified marker genes directly from RNA mixture samples. Such methods are therefore robust to technical/biological heterogeneity in marker gene expression, provided these genes maintain their cell type-specificity in independent mixture samples. However, lineage-specific marker genes are not always trivial to define, especially in complex tissues and when considering closely related TIL subsets.

One technique for accommodating closely related cell types within a deconvolution framework is to employ regularization in solving the linear system [51]. This allows correlated predictor variables (i.e., similar cell types) to be assigned coefficients (i.e., cell fractions) that more closely reflect their actual proportions, thereby minimizing the risk of 'drop out' problems owing to multicollinearity [41•,52]. For example, to profile a large diversity of cell types, Altboum and colleagues introduced a deconvolution approach based on elastic net regularization called Digital Cell Quantifier (DCQ) [36], and observed improvement through regularization. Indeed, regularization was found to improve the performance of several distinct

deconvolution methods [39•], suggesting the general utility of this approach.

To overcome the main challenges outlined above, we recently described CIBERSORT, an approach for enumerating cell proportions based on nu-support vector regression (ν-SVR) [41•]. ν-SVR combines feature selection with a linear loss function and *L2*-norm regularization, rendering it robust to unknown mixture content, noise, and highly correlated cell subsets [41•]. To facilitate evaluation of the results, CIBERSORT also provides a confidence metric in the deconvolution output. We defined a signature matrix containing 22 mature hematopoietic subsets and benchmarked >50% of subsets against gold standard methods on real biological mixture samples, including solid tumors [6,41•]. When applied to GEPs from 25 tumor types in a pan-cancer analysis of thousands of bulk tumors, CIBERSORT revealed complex associations between 22 leukocyte subsets and clinical outcomes. Predictions linking tumor-infiltrating neutrophils and plasma cells to overall survival were validated by microscopy and IHC in lung adenocarcinoma [6]. Several groups have successfully applied CIBERSORT for cell deconvolution in various complex tissues, including microarray studies of peripheral blood [53], lung biopsies [54], triple negative breast cancers [55], and glioblastomas [56], and RNA-seq analyses of endometrial cancers [57].

Each method described in this review has inherent strengths and weaknesses. Therefore, to assist readers in selecting the most appropriate technique for their application of interest, Table 1 summarizes the key attributes of several approaches.

Calibration of in *silico* TIL profiling methods

Many TIL-associated genes are not cell type-specific. For example, CD4 is highly expressed not only by CD4 T cells, but also by monocyte and macrophage subsets [41•,50]. To establish baseline performance, we recommend that marker genes (and signature matrices) be validated by analyzing independently generated GEPs of (1) the same leukocyte phenotypes, (2) different leukocyte phenotypes, and (3) samples devoid of immune content. At a minimum, deconvolution methods should be able to robustly resolve all cell types within a given signature matrix; if external datasets are not available, this can be accomplished by leave-one-out cross validation.

In silico benchmarking, while important, is not an effective substitute for validation against gold standard experimental methods. While comprehensive evaluation is likely impractical, we highly recommend that a subset of leukocyte subsets be compared against their ground truth proportions in real biological specimens. For comparisons against flow cytometry, expression profiling should ideally be performed on RNA derived from the cell suspension, rather than from the bulk tissue, to avoid alterations in cellular representation.

Outstanding issues

Despite the utility of computational methods for TIL characterization, there are several issues that will require further investigation. First, while progress toward 'deep deconvolution' [16••] of multiple cell types has been made [36,41•], the maximum phenotypic repertoire and analytical sensitivity of each *in silico* dissection method remains an open question. The results will depend on several factors, including on features of the selected algorithm (Table 1), the robustness of TIL marker genes, the level of unknown content and/or noise in the mixture, and the expression profiling platform (e.g., microarray or RNA-seq). For example, all methods are expected to benefit from the lower noise levels and increased dynamic range of RNA-seq [58,59]. Second, because *in silico* methods cannot count individual cells in bulk tissue GEPs, differences in cellular RNA content may lead to estimation biases. Efforts to address this issue are under way (e.g., [37]). Third, while we have observed accurate assessment of TIL fractions using normal leukocyte reference profiles [6,41•], it remains possible that significant drifts in marker gene expression between TILs and normal leukocyte subsets will lead to less accurate results. Methods that determine TIL expression profiles directly from tumor samples can overcome this problem (e.g., [17,43•]). As a corollary, the architectural pattern and spatial distribution of individual TILs may be important for their function in a manner not captured by bulk tumor transcriptomes [1,4,5,60]. If so, extending reference profiles to include sorted TIL subsets from bulk tumors and from specific spatial compartments should alleviate this issue. Finally, a key advantage of *in silico* tumor dissection is its potential to analyze formalin fixed, paraffin embedded (FFPE) specimens [41•]. Further studies are needed to define optimal FFPE processing techniques and better characterize the impact of RNA degradation on technical performance.

Outlook

Infiltration of tumors by distinct immune subsets is a hallmark of cancer. Therefore, a better understanding of TIL biology across human tumor types, both at important clinical milestones and in relation to therapy, will facilitate the discovery of predictive and prognostic biomarkers, and new immunotherapeutic targets. *In silico* tissue dissection is an emerging class of techniques for large-scale characterization of tumor cellular heterogeneity [6••,8••,34,42,43•,61]. Because these approaches do not typically rely on surface markers, antibodies, tissue disaggregation (for preparation of single cell suspensions), or viably preserved cells, they have potential to complement traditional methods while enabling TIL enumeration from fresh, frozen, and fixed tumor specimens (Figure 1). When integrated with complementary data, including somatic and epigenetic alterations, B-cell and

T-cell receptor profiles, and imaging, *in silico* tissue dissection will lead to more comprehensive portraits of human tumors. In the near future, we expect such analyses to transform cancer therapy and management.

Acknowledgements

We are grateful to A. Gentles for critically reading the manuscript. Supported by grants from the Doris Duke Charitable Foundation (AAA), the Damon Runyon Cancer Research Foundation (AAA), the B&J Cardan Oncology Research Fund (AAA), the Ludwig Institute for Cancer Research (AAA), NIH grant 1K99CA187192-01A1 (AMN), NIH grant PHS NRSA 5T32 CA09302-35 (AMN), US Department of Defense grant W81XWH-12-1-0498 (AMN) and a grant from the Siebel Stem Cell Institute and the Thomas and Stacey Siebel Foundation (AMN).

References and recommended reading

Papers of particular interest, published within the period of review, have been highlighted as:

- • of special interest
- •• of outstanding interest

1. •• Galon J, Costes A, Sanchez-Cabo F, Kirilovsky A, Mlecnik B, Lagorce-Pagès C, Tosolini M, Camus M, Berger A, Wind P *et al.*: **Type, density, and location of immune cells within human colorectal tumors predict clinical outcome**. *Science* 2006, **313**:1960-1964.
Levels of tumor-infiltrating lymphocytes, along with their identities and spatial distribution patterns, were found to influence colorectal cancer clinical outcomes independently of standard clinical indices.

2. Hanahan D, Coussens LM: **Accessories to the crime: functions of cells recruited to the tumor microenvironment**. *Cancer Cell* 2012, **21**:309-322.

3. Junttila MR, de Sauvage FJ: **Influence of tumour microenvironment heterogeneity on therapeutic response**. *Nature* 2013, **501**:346-354.

4. Adams S, Gray RJ, Demaria S, Goldstein L, Perez EA, Shulman LN, Martino S, Wang M, Jones VE, Saphner TJ *et al.*: **Prognostic value of tumor-infiltrating lymphocytes in triple-negative breast cancers from two phase III randomized adjuvant breast cancer trials: ECOG 2197 and ECOG 1199**. *J Clin Oncol* 2014, **32**:2959-2966.

5. Tumeh PC, Harview CL, Yearley JH, Shintaku IP, Taylor EJ, Robert L, Chmielowski B, Spasic M, Henry G, Ciobanu V *et al.*: **PD-1 blockade induces responses by inhibiting adaptive immune resistance**. *Nature* 2014, **515**:568-571.

6. •• Gentles AJ, Newman AM, Liu CL, Bratman SV, Feng W, Kim D, Nair VS, Xu Y, Khuong A, Hoang CD *et al.*: **The prognostic landscape of genes and infiltrating immune cells across human cancers**. *Nat Med* 2015, **21**:938-945.
Large-scale application of gene expression deconvolution to analyze TIL composition and survival associations in 25 cancer types spanning thousands of bulk human tumors.

7. Green MR, Kihira S, Liu CL, Nair RV, Salari R, Gentles AJ, Irish J, Stehr H, Vicente-Duenas C, Romero-Camarero I *et al.*: **Mutations in early follicular lymphoma progenitors are associated with suppressed antigen presentation**. *Proc Natl Acad Sci U S A* 2015, **112**:E1116-E1125.

8. •• Rooney MS, Shukla SA, Wu CJ, Getz G, Hacohen N: **Molecular and genetic properties of tumors associated with local immune cytolytic activity**. *Cell* 2015, **160**:48-61.
Genes enriched in cytotoxic T cells and NK cells were used to dissect the cytolytic activity of infiltrating immune cells in bulk tumor samples from 18 cancers. Cytolytic activity was related to important tumor genomic features, revealing new insights into resistance mechanisms against anti-tumor immunity.

9. Mlecnik B, Bindea G, Kirilovsky A, Angell HK, Obenauf AC, Tosolini M, Church SE, Maby P, Vasaturo A, Angelova M *et al.*: **The tumor microenvironment and Immunoscore are critical determinants of dissemination to distant metastasis**. *Sci Transl Med* 2016, **8** 327ra326.

10. Chao MP, Alizadeh AA, Tang C, Myklebust JH, Varghese B, Gill S, Jan M, Cha AC, Chan CK, Tan BT *et al.*: **Anti-CD47 antibody synergizes with rituximab to promote phagocytosis and eradicate non-Hodgkin lymphoma**. *Cell* 2010, **142**:699-713.

11. Willingham SB, Volkmer JP, Gentles AJ, Sahoo D, Dalerba P, Mitra SS, Wang J, Contreras-Trujillo H, Martin R, Cohen JD *et al.*: **The CD47-signal regulatory protein alpha (SIRPa) interaction is a therapeutic target for human solid tumors**. *Proc Natl Acad Sci U S A* 2012, **109**:6662-6667.

12. Ribas A: **Releasing the brakes on cancer immunotherapy**. *N Engl J Med* 2015, **373**:1490-1492.

13. Schalper KA, Kaftan E, Herbst RS: **Predictive biomarkers for PD-1 axis therapies: the hidden treasure or a call for research**. *Clin Cancer Res* 2016, **22**:2102-2104.

14. McGranahan N, Furness AJ, Rosenthal R, Ramskov S, Lyngaa R, Saini SK, Jamal-Hanjani M, Wilson GA, Birkbak NJ, Hiley CT *et al.*: **Clonal neoantigens elicit T cell immunoreactivity and sensitivity to immune checkpoint blockade**. *Science* 2016, **351**:1463-1469.

15. Hugo W, Shi H, Sun L, Piva M, Song C, Kong X, Moriceau G, Hong A, Dahlman KB, Johnson DB *et al.*: **Non-genomic and immune evolution of melanoma acquiring MAPKi resistance**. *Cell* 2015, **162**:1271-1285.

16. •• Shen-Orr SS, Gaujoux R: **Computational deconvolution: extracting cell type-specific information from heterogeneous samples**. *Curr Opin Immunol* 2013, **25**:571-578.
Review of *in silico* deconvolution approaches developed prior to 2014, including their scope, potential applications, and limitations.

17. Zhong Y, Wan YW, Pang K, Chow LM, Liu Z: **Digital sorting of complex tissues for cell type-specific gene expression profiles**. *BMC Bioinformatics* 2013, **14**:89.

18. Bendall SC, Simonds EF, Qiu P, Amir el AD, Krutzik PO, Finck R, Bruggner RV, Melamed R, Trejo A, Ornatsky OI *et al.*: **Single-cell mass cytometry of differential immune and drug responses across a human hematopoietic continuum**. *Science* 2011, **332**:687-696.

19. Robins HS, Ericson NG, Guenthoer J, O'Briant KC, Tewari M, Drescher CW, Bielas JH: **Digital genomic quantification of tumor-infiltrating lymphocytes**. *Sci Transl Med* 2013, **5**:214ra169.

20. Angelo M, Bendall SC, Finck R, Hale MB, Hitzman C, Borowsky AD, Levenson RM, Lowe JB, Liu SD, Zhao S *et al.*: **Multiplexed ion beam imaging of human breast tumors**. *Nat Med* 2014, **20**:436-442.

21. • Venet D, Pecasse F, Maenhaut C, Bersini H: **Separation of samples into their constituents using gene expression data**. *Bioinformatics* 2001, **17(Suppl. 1)**:S279-S287.
This early paper formalized the problem of gene expression deconvolution as a system of linear equations. It also presented methods for solving the linear system and provided anecdotes illustrating the utility of the approach.

22. 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**:10370-10375.

23. Dave SS, Wright G, Tan B, Rosenwald A, Gascoyne RD, Chan WC, Fisher RI, Braziel RM, Rimsza LM, Grogan TM *et al.*: **Prediction of survival in follicular lymphoma based on molecular features of tumor-infiltrating immune cells**. *New Engl J Med* 2004, **351**:2159-2169.

24. •• 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.
First paper to show how expression deconvolution can be applied to enumerate leukocyte composition in whole blood samples and reveal new insights into a human disease. Methods for expression deconvolution and signature matrix generation are presented.

25. Repsilber D, Kern S, Telaar A, Walzl G, Black GF, Selbig J, Parida SK, Kaufmann SH, Jacobsen M: **Biomarker discovery in**

heterogeneous tissue samples — taking the in-silico deconfounding approach. *BMC Bioinformatics* 2010, **11**:27.

26. 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**:287-289.

27. Gaujoux R, Seoighe C: **Semi-supervised nonnegative matrix factorization for gene expression deconvolution: a case study**. *Infect Genet Evol* 2012, **12**:913-921.

28. • Gong T, Hartmann N, Kohane IS, Brinkmann V, Staedtler F, Letzkus M, Bongiovanni S, Szustakowski JD: **Optimal deconvolution of transcriptional profiling data using quadratic programming with application to complex clinical blood samples**. *PLoS One* 2011, **6**:e27156.
The linearity assumptions made by microarray expression deconvolution approaches are likely to hold in RNA-seq space.

29. 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**: 945-947.

30. Qiao W, Quon G, Csaszar 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.

31. Bindea G, Mlecnik B, Tosolini M, Kirilovsky A, Waldner M, Obenauf AC, Angell H, Fredriksen T, Lafontaine L, Berger A *et al.*: **Spatiotemporal dynamics of intratumoral immune cells reveal the immune landscape in human cancer**. *Immunity* 2013, **39**:782-795.

32. Gong T, Szustakowski JD: **DeconRNASeq: a statistical framework for deconvolution of heterogeneous tissue samples based on mRNA-Seq data**. *Bioinformatics* 2013, **29**:1083-1085.

33. Nagalla S, Chou JW, Willingham MC, Ruiz J, Vaughn JP, Dubey P, Lash TL, Hamilton-Dutoit SJ, Bergh J, Sotiriou C *et al.*: **Interactions between immunity, proliferation and molecular subtype in breast cancer prognosis**. *Genome Biol* 2013, **14**: 1-18.

34. Yoshihara K, Shahmoradgoli M, Martinez E, Vegesna R, Kim H, Torres-Garcia W, Trevino V, Shen H, Laird PW, Levine DA *et al.*: **Inferring tumour purity and stromal and immune cell admixture from expression data**. *Nat Commun* 2013, **4**:2612.

35. 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.

36. Altboum Z, Steuerman Y, David E, Barnett-Itzhaki Z, Valadarsky L, Keren-Shaul H, Meningher T, Mendelson E, Mandelboim M, Gat-Viks I *et al.*: **Digital cell quantification identifies global immune cell dynamics during influenza infection**. *Mol Syst Biol* 2014, **10**:720.

37. 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**:682-689.

38. Chikina M, Zaslavsky E, Sealfon SC: **CellCODE: a robust latent variable approach to differential expression analysis for heterogeneous cell populations**. *Bioinformatics* 2015, **31**: 1584-1591.

39. • Mohammadi S, Zuckerman N, Goldsmith A, Grama A: **A critical survey of deconvolution methods for separating cell-types in complex tissues**. In ArXiv 2015, http://arxiv.org/abs/1510.04583v1.
Recent technical overview and comparative assessment of expression deconvolution methods.

40. Moffitt RA, Marayati R, Flate EL, Volmar KE, Loeza SG, Hoadley KA, Rashid NU, Williams LA, Eaton SC, Chung AH *et al.*: **Virtual microdissection identifies distinct tumor- and stroma-specific subtypes of pancreatic ductal adenocarcinoma**. *Nat Genet* 2015, **47**:1168-1178.

41. • 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**:453-457.
A gene expression deconvolution approach (CIBERSORT) that is robust to unknown mixture content, noise, and closely related cell types. CIBERSORT also provides a measure of statistical confidence in the results.

42. Senbabaoglu Y, Winer AG, Gejman RS, Liu M, Luna A, Ostrovnaya I, Weinhold N, Lee W, Kaffenberger SD, Chen YB *et al.*: **The landscape of T cell infiltration in human cancer and its association with antigen presenting gene expression**. *bioRxiv* 2015. http://dx.doi.org/10.1101/025908.

43. • Tirosh I, Izar B, Prakadan SM, Wadsworth MH 2nd, Treacy D, Trombetta JJ, Rotem A, Rodman C, Lian C, Murphy G *et al.*: **Dissecting the multicellular ecosystem of metastatic melanoma by single-cell RNA-seq**. *Science* 2016, **352**:189-196.
Single cell transcriptome profiling was use to define TIL and stromal marker genes directly from melanoma tumors. This approach should overcome potential issues with *in silico* TIL dissection related to unknown expression differences between TILs and normal leukocyte counterparts.

44. • Ju W, Greene CS, Eichinger F, Nair V, Hodgin JB, Bitzer M, Lee YS, Zhu Q, Kehata M, Li M *et al.*: **Defining cell-type specificity at the transcriptional level in human disease**. *Genome Res* 2013, **23**:1862-1873.
Interesting *in silico* approach for defining novel cell type-enriched marker genes without the need for purification techniques, such as flow cytometry.

45. Hong WJ, Warnke R, Chu G: **Immune signatures in follicular lymphoma**. *N Engl J Med* 2005, **352**:1496-1497 (author reply).

46. Angelova M, Charoentong P, Hackl H, Fischer ML, Snajder R, Krogsdam AM, Waldner MJ, Bindea G, Mlecnik B, Galon J *et al.*: **Characterization of the immunophenotypes and antigenomes of colorectal cancers reveals distinct tumor escape mechanisms and novel targets for immunotherapy**. *Genome Biol* 2015, **16**:64.

47. • Barbie DA, Tamayo P, Boehm JS, Kim SY, Moody SE, Dunn IF, Schinzel AC, Sandy P, Meylan E, Scholl C *et al.*: **Systematic RNA interference reveals that oncogenic KRAS-driven cancers require TBK1**. *Nature* 2009, **462**:108-112.
This paper extended GSEA to single sample analysis (ssGSEA), which has since been widely applied for examining the statistical enrichment of gene sets in individual transcriptomes.

48. Bolen CR, Uduman M, Kleinstein SH: **Cell subset prediction for blood genomic studies**. *BMC Bioinformatics* 2011, **12**:258.

49. Gaujoux R, Seoighe C: **CellMix: a comprehensive toolbox for gene expression deconvolution**. *Bioinformatics* 2013, **29**:2211-2212.

50. Abbas AR, Baldwin D, Ma Y, Ouyang W, Gurney A, Martin F, Fong S, van Lookeren Campagne M, Godowski P, Williams PM *et al.*: **Immune response in silico (IRIS): immune-specific genes identified from a compendium of microarray expression data**. *Genes Immun* 2005, **6**:319-331.

51. Hastie T, Tibshirani R, Friedman J: *The Elements of Statistical Learning*. Springer, New York Inc.; 2001.

52. Farrar DE, Glauber RR: **Multicollinearity in regression analysis: the problem revisited**. *Rev Econ Stat* 1967, **49**:92-107.

53. Karpinski P, Frydecka D, Sasiadek MM, Misiak B: **Reduced number of peripheral natural killer cells in schizophrenia but not in bipolar disorder**. *Brain Behav Immun* 2016, **54**:194-200.

54. Araujo JM, Prado A, Cardenas NK, Zaharia M, Dyer R, Doimi F, Bravo L, Pinillos L, Morante Z, Aguilar A *et al.*: **Repeated observation of immune gene sets enrichment in women with non-small cell lung cancer**. *Oncotarget* 2016, **7**:20282-20292.

55. Vinayak S, Gray RJ, Adams S, Jensen KC, Manola J, Afghahi A, Goldstein LJ, Ford JM, Badve SS, Telli ML: **Association of increased tumor-infiltrating lymphocytes (TILs) with immunomodulatory (IM) triple-negative breast cancer (TNBC) subtype and response to neoadjuvant platinum-based therapy in PrECOG0105**. *ASCO Meeting Abst* 2014, **32**:1000.

56. Wang Q, Hu X, Muller F, Kim H, Squatrito M, Millelsen T, Scarpace L, Barthel F, Lin Y-H, Satani N *et al.*: **Tumor evolution of glioma intrinsic gene expression subtype associates with immunological changes in the microenvironment**. *bioRxiv* 2016. http://dx.doi.org/10.1101/052076.

57. Mehnert JM, Panda A, Zhong H, Hirshfield K, Damare S, Lane K, Sokol L, Stein MN, Rodriguez-Rodriquez L, Kaufman HL *et al.*: **Immune activation and response to pembrolizumab in POLE-mutant endometrial cancer**. *J Clin Invest* 2016 http://dx.doi.org/10.1172/JCI84940.

58. Wang C, Gong B, Bushel PR, Thierry-Mieg J, Thierry-Mieg D, Xu J, Fang H, Hong H, Shen J, Su Z *et al.*: **The concordance between RNA-seq and microarray data depends on chemical treatment and transcript abundance**. *Nat Biotech* 2014, **32**:926-932.

59. Zhao S, Fung-Leung W-P, Bittner A, Ngo K, Liu X: **Comparison of RNA-Seq and microarray in transcriptome profiling of activated T cells**. *PLoS ONE* 2014, **9**:e78644.

60. Yuan Y, Failmezger H, Rueda OM, Ali HR, Graf S, Chin SF, Schwarz RF, Curtis C, Dunning MJ, Bardwell H *et al.*: **Quantitative image analysis of cellular heterogeneity in breast tumors complements genomic profiling**. *Sci Transl Med* 2012, **4** 157ra143.

61. Aran D, Sirota M, Butte AJ: **Systematic pan-cancer analysis of tumour purity**. *Nat Commun* 2015, **6**:8971.