



Review

Identification of non-cancer cells from cancer transcriptomic data[☆]Michele Bortolomeazzi¹, Mohamed Reda Keddar¹, Francesca D. Ciccarelli*, Lorena Benedetti*

Cancer Systems Biology Laboratory, The Francis Crick Institute, London NW1 1AT, UK
 School of Cancer and Pharmaceutical Sciences, King's College London, London SE11UL, UK



ARTICLE INFO

Keywords:

Tumour microenvironment
 RNA-seq
 Gene expression profiles
 Cell-specific signatures
 GSEA
 Deconvolution

ABSTRACT

Interactions between cancer cells and non-cancer cells composing the tumour microenvironment play a primary role in determining cancer progression and shaping the response to therapy. The qualitative and quantitative characterisation of the different cell populations in the tumour microenvironment is therefore crucial to understand its role in cancer. In recent years, many experimental and computational approaches have been developed to identify the cell populations composing heterogeneous tissue samples, such as cancer. In this review, we describe the state-of-the-art approaches for the quantification of non-cancer cells from bulk and single-cell cancer transcriptomic data, with a focus on immune cells. We illustrate the main features of these approaches and highlight their applications for the analysis of the tumour microenvironment in solid cancers. We also discuss techniques that are complementary and alternative to RNA sequencing, particularly focusing on approaches that can provide spatial information on the distribution of the cells within the tumour in addition to their qualitative and quantitative measurements.

This article is part of a Special Issue entitled: Transcriptional Profiles and Regulatory Gene Networks edited by Dr. Federico Manuel Giorgi and Dr. Shaun Mahony.

1. Introduction

Cancers arising from epithelial cells account for 80–90% of all solid cancers [1]. However, cancer cells do not grow in isolation. The malignant epithelium is in fact surrounded by stromal cells including fibroblasts, immune and endothelial cells which altogether form the tumour microenvironment (TME). Stromal cells in the TME sustain and regulate tumour growth, immune evasion and drug resistance mechanisms [2]. In the past decade, the interest of the cancer research community in the TME has progressively grown because of its role in new therapies that target the host immune system [2–4]. In particular, T cells are able to recognise and eliminate tumour cells. However, tumours develop resistance mechanisms preventing T cell activation. Immunotherapies currently used in the clinic have two main mechanisms of action. They either boost the immune response by activating T cells or they restore the immune response that has been inactivated during tumour growth. Anticancer vaccines and chimeric antigen receptor T cells represent successful attempts to activate the anticancer immune response [3]. Immune checkpoint inhibitors release the brakes imposed by tumour cells on T cells, restoring the host antitumour

immune response. These drugs are already successfully applied to treat a variety of tumours, including melanoma, lymphoma, lung, renal cell, head and neck squamous, bladder, liver and gastro-oesophageal cancers [3,5]. However, despite their encouraging success, still many patients do not respond to immunotherapy or develop resistance over time. Understanding TME complexity is therefore essential to predict which patients would benefit from immunotherapy, in full agreement with a personalised approach to cancer therapy.

Tumour infiltrating immune cells can be either beneficial or detrimental for cancer development depending on their localisation, abundance and function. For instance, the presence of CD8⁺ T cells and T helper cells is usually associated with good prognosis [2,6] while myeloid derived suppressor cells are predictive of bad outcome [7]. Therefore, the detailed characterisation of immune infiltrates is being progressively incorporated into the clinical practice [6,8]. A method widely used in the clinic to estimate the abundance of tumour infiltrating immune cells is the haematoxylin and eosin (H&E) staining. Although this staining is not specific for any particular cell type, it has proven to be clinically relevant for several cancer types [6]. For instance, high levels of lymphocyte infiltration estimated from H&E

[☆] This article is part of a Special Issue entitled: Transcriptional Profiles and Regulatory Gene Networks edited by Dr. Federico Manuel Giorgi and Dr. Shaun Mahony.

* Corresponding authors.

E-mail addresses: francesca.ciccarelli@crick.ac.uk (F.D. Ciccarelli), lorena.benedetti@crick.ac.uk (L. Benedetti).

¹ Equal contribution.

staining are predictive of better prognosis in non-small cell lung cancer [9]. More cell-specific methods for the clinical quantification of immune cells include the combination of up to five antibodies to detect the presence of different immune cell populations using immunohistochemistry (IHC) or immunofluorescence (IF) [6]. These chromogenic or fluorescent labelling-based approaches also provide some spatial information on how epithelial and stromal cells are distributed within the tumour. This analysis is however restricted to the small portion of the tumour that can be sliced from a formalin-fixed paraffin embedded (FFPE) cancer block. It therefore may not be representative of the whole tumour mass. Moreover, the number of cell populations that can be identified is limited due to the small number of markers that can be tested. In this respect, serial IF constitutes a major improvement allowing several rounds of sequential staining of the same sections using up to 12 antibodies [10]. Similarly, high-parameter flow cytometry can profile up to 27 markers in disaggregated cells from several centimetres of tumour mass [11]. These approaches are still being developed and are not yet part of the clinical practice.

Similarly, approaches based on the quantification of protein expression with mass-spectrometry can also reveal detailed profiles of the tumour immune infiltrates. Imaging mass cytometry (IMC) [12] and multiplex ion beam imaging (MIBI) [13] allow the simultaneous identification of up to 40 markers in about 1mm² of tissue area. IMC and MIBI provide spatial information on the distribution of cells within the tissue, which adds additional layers of relevant information. Other methods rely on mRNA quantification either using fluorescent probes, like the NanoString nCounter [14], or next-generation sequencing (NGS). NanoString nCounter can be applied to slices of FFPE or fresh frozen (FF) tissues leading to the quantification of up to 800 markers. NGS-based approaches like RNA sequencing (RNA-seq) can be applied to bulk cancer samples or to previously isolated single cells. Despite not providing any spatial information, RNA-seq enables a comprehensive and unbiased characterisation of tumour infiltrating immune cells [15,16]. Moreover, the latest advancements in the field of transcriptomics are beginning to provide spatial resolution ranging from a few cells to subcellular levels [17,18].

In this review, we describe the main methods currently used to quantify tumour-infiltrating cell populations, with a particular focus on those based on bulk and single cell RNA sequencing (scRNA-seq). We also comment on alternative and complementary approaches that are emerging for TME characterisation.

2. RNA sequencing of cancer samples

RNA-seq allows the quantification of gene expression and enables the profiling of a number of genes far greater than other approaches based on probes or antibodies. In the context of cancer biology, RNA-seq is a useful tool for tumour classification, patient stratification and for studying response to therapy, [19,20].

2.1. Bulk RNA sequencing

Bulk RNA-seq refers to the sequencing of RNA from the bulk cancer mass and it consists of four steps (Fig. 1A).

The first step is the extraction of RNA from either FF or FFPE cancer samples. FF samples yield higher quantity and better quality RNA and are thus preferentially used in large scale sequencing projects such as The Cancer Genome Atlas (TCGA). However, the vast majority of samples archived in hospital cancer biobanks are FFPE tissue blocks [21]. Paraffin embedding and long-term storage are known to cause the fragmentation of nucleic acids, while crosslinking is a direct consequence of formalin fixation. This usually leads to low quantity and bad quality RNA [21]. A de-modification step in which the RNA is heated in amine-rich or organocatalytic buffers can be performed to revert formaldehyde linkages and improve RNA quality [21]. Independently of the sample source, the quality of the extracted RNA is a

key factor for all downstream analysis and should be carefully evaluated. The main RNA quality metric is the 28 s:18 s rRNA ratio, generally expressed as a RNA integrity number (RIN), with a higher value indicating more intact RNA. While there is no consensus on the RIN value to be used as a quality threshold, generally RIN values below 5 can negatively impact the library preparation and sequencing steps [22].

The second step is the depletion of rRNAs that usually constitute > 80% of the total RNA. There are several approaches for rRNA depletion, depending on RNA quality [23]. In one of them, mRNA is enriched through poly-A enrichment using oligo-dT beads. This method generates high quality expression data that strongly correlates with measurements from independent techniques such as microarrays [24]. However, it requires high quality and intact input RNA, because the capture is done with a poly-T primer against the 3' end of the transcript. Thus, it is not always suitable for FFPE samples [23]. mRNA enrichment can also be achieved through exon capture probes after cDNA synthesis. In a comparative study with matched FF and FFPE tissue, the best correlation between FF and FFPE expression data was obtained with exon capture RNA [23]. However, the coverage is mostly limited to the captured sequences. This is because the RNA is partially fragmented so the exonic probes will pull down small fragments containing the target sequences. This approach allows to recover RNA fractions of > 98% of the exome [25]. Alternatively, rRNAs can be removed with techniques based on hybridisation, duplex digestion, or not-so-random RT-PCR priming [23].

In the third step the RNA is fragmented, generally by heat digestion with divalent cations. Finally, in the fourth step the fragmented RNA is converted into cDNA and ligated to adapters to generate the library for sequencing. The most commonly used NGS platforms for RNA-seq are HiSeq and MiSeq Illumina.

2.2. Single-cell RNA sequencing

The recent development of high-throughput scRNA-seq technologies allows to profile the transcriptome of thousands of individual cells per sample [26] (Fig. 1B). These approaches mostly differ in the techniques used for single-cell isolation. Cells can be isolated by fluorescence-activated cell sorting (FACS), as in the MARS-Seq method, which performs scRNA-seq on thousands of cells previously sorted into 384-well plates [27]. Alternatively, single cells can be separated using microfluidic chambers. This is achieved through micron-scale well arrays (as in Seq-Well [28]) or by separating cells in aqueous microdroplets forming an emulsion with an oil phase (as in Chromium [29], Drop-seq [30] and inDrop [31]). Microfluidic-based methods require lower reaction volumes and enable the screening of up to hundreds of thousand cells at lower costs [26]. Cell shape and stickiness (for example of fibroblasts or cancer cells) can affect the efficiency of these methods, biasing single cell capture towards certain cell types over others [32]. Due to the high number of cells, sequencing depth is limited to around 50,000 reads/cell, which is sufficient for clustering and identifying different populations [33].

FFPE samples represent a major challenge for scRNA-seq because the tissue cannot be disaggregated to obtain single cells. However, single cells can be isolated using a computer-guided laser capture microdissection (LCM) system. Although this approach has a throughput of hundreds of cells only, it offers the advantage that each sequenced cell can be mapped back to its original location in the tissue [34].

In the case of FACS or microfluidic-based methods, cells are bar-coded during the cDNA synthesis step using beads bound to primers containing a cell-specific barcode, a poly-T capture sequence, and a Unique Molecular Identifier (UMI). While cell-specific barcodes are identical within each cell-containing droplet or well, UMI sequences are different and allow the counting of individual mRNA molecules. This reduces the effects of duplicates that can be generated during cDNA amplification [35]. Since a poly-T capture sequence is used, only the 3'

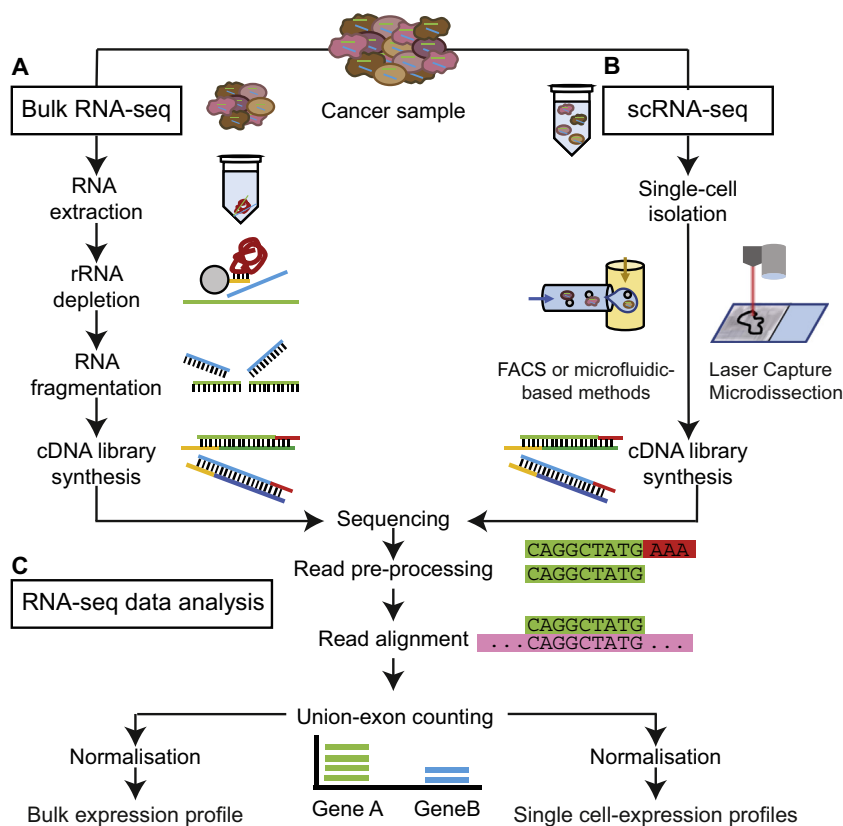


Fig. 1. Workflows of bulk and scRNA-seq experiments. (A) Bulk RNA-seq of solid tumours is based on four steps: RNA extraction from the cancer tissue, rRNA depletion, RNA fragmentation, and cDNA library synthesis for sequencing. (B) scRNA-seq from solid tumour samples requires single cell isolation either through FACS or microfluidics-based methods or laser capture microdissection. cDNA libraries from individual cells are then synthesised and sequenced. (C) Analytical approaches for the quantification of gene expression for bulk RNA-seq and scRNA-seq data. After pre-processing, the reads are aligned to the reference transcriptome or genome. Reads mapping to the exons are counted and normalised to generate gene expression profiles. FACS = fluorescence-activated cell sorting.

RNAs ends are sequenced [26]. This achieves a throughput of thousands to hundreds of thousands of cells, despite the limitations imposed by sequencing cost and capacity. The main drawbacks of poly-T capture beads are that low abundance transcripts may be lost mainly due to limited capture efficiency [36] and it is not suitable for detecting mutations or splicing variants [33].

In the case of LCM, isolated cells are generally sequenced at lower throughput using full-length scRNA-seq protocols, like Smart-seq2 [37]. These methods use Switching Mechanism at 5'End of RNA Template (SMART) chemistry. In this respect the recently developed Smart-3SEQ protocol is particularly suited for FFPE samples [38].

2.3. RNA-seq data analysis

Conceptually similar analytical approaches can be applied to the quantification of gene expression from either bulk RNA-seq or scRNA-seq data [36] (Fig. 1C). In fact, although scRNA-seq-specific methods have started to be developed [36], bulk RNA-seq analysis tools are still successfully applied to scRNA-seq data [39].

The input data for the quantification of gene expression are the raw sequencing reads, which undergo pre-processing to remove adaptor sequences, trim poor-quality bases, and discard low-quality reads, usually derived from poor quality RNA. Libraries with a high number of low-quality reads have lower complexity. This affects the detection of lowly expressed genes and can negatively bias the quantification of gene expression [22].

Pre-processed reads are then aligned to either the reference transcriptome or genome. Aligned reads may undergo post-mapping quality control to evaluate sequence overrepresentation and fragment-size biases. Finally, reads mapping to the exons are counted using union-exon counting methods. In the case of bulk RNA-seq, read counts are normalised to account for gene length and library size and obtain the sample gene expression profile. Different types of gene expression measures can be used, including Reads Per Kilobase of transcript per

Million mapped reads (RPKM), Fragments Per Kilobase of transcript per Million mapped reads (FPKM), or Transcripts Per Million (TPM). In the case of scRNA-seq a direct molecule counting based on UMIs can be performed providing an absolute measure of gene expression [35]. If UMIs are not used, scRNA-seq specific normalisation tools can be applied [39]. Moreover, several quality control metrics are usually used to exclude cells with too few or degraded RNA or cell doublets accidentally captured in the same reaction chamber. For instance, bad quality reads or a large percentage of unmapped reads in scRNA-seq can be an index of RNA degradation. Also, high mitochondrial-to-nuclear gene mapping ratio or low mRNA abundance are linked to apoptotic or damaged cells which have lost most of their cytoplasmic mRNAs [39].

The resulting gene expression data from either bulk or scRNA-seq can be used as input for determining the cell-type composition in the sample. Furthermore, scRNA-seq data can be used to define refined cell-type specific expression profiles. These applications of bulk and scRNA-seq are described in detail in the next sections.

3. Quantification of non-cancer cells from bulk transcriptomic data

The bulk transcriptomic profile of a cancer sample is an admixture of transcripts from cancer and non-cancer cells. It therefore offers a qualitative and quantitative representation of the diverse cell types that are present in the sample.

In recent years, many computational approaches have been designed to estimate the abundance of the various cell populations of the TME from bulk transcriptomic data [15] (Table 1). Such approaches leverage on reference signatures consisting of either marker genes and/or expression profile matrices that are specific for a given cell population. Therefore, to quantify the non-cancer component of the TME from cancer expression data, it is paramount to derive robust marker genes or profile matrices. These are generated using a signature derivation pipeline that consists of three main steps (Fig. 2). In the first step, gene

Table 1

Examples of approaches for the quantification of tumour-infiltrating cells from bulk transcriptomic data. For each approach, we report the underlying mathematical method, the type of expression data used to derive the signatures, the total number of marker genes included in the signatures and the final number of non-cancer cell populations considered. Only methods that implement their own reference signatures and that have been applied to the analysis of cancer samples are reported. ssGSEA = single sample gene set enrichment analysis, GSVA = gene set variation analysis.

Approach	Computational method	Source of expression data	Marker genes (n)	Cell populations (n)
Angelova et al. [40]	ssGSEA	Microarray	812	31
Charoentong et al. [41]	ssGSEA	Microarray	782	28
ConsensusTME [42]	ssGSEA	Microarray, bulk RNA-seq	Cancer type specific	18
xCell [43]	ssGSEA and spillover compensation	Microarray, bulk RNA-seq	10,808	64
Tamborero et al. [44]	Scoring (GSVA)	Microarray, bulk RNA-seq	401	16
MCP-counter [45]	Log-transformed geometric mean of expression	Microarray	522	10
Danaher et al. [46]	Log-transformed geometric mean of expression	Microarray, bulk RNA-seq	60	14
ImSig [47]	Arithmetic mean of expression	Microarray, bulk RNA-seq	318	7
CIBERSORT [48]	Deconvolution, nu support vector regression	Microarray	547	22
TIMER [49]	Deconvolution, constrained least square fitting	Microarray	Cancer type specific	6
EPIC [50]	Deconvolution, constrained least square fitting	scRNA-seq	118	10
quanTIseq [51]	Deconvolution, constrained least square fitting	Bulk RNA-seq	153	10

expression data of the cell populations of interest are collected from gene expression databases (e.g. GEO, IRIS, ArrayExpress) and/or from the literature (Fig. 2A). In the second step, these expression data are curated and normalised to allow their comparative analysis (Fig. 2B). Finally, cell type-specific markers (Fig. 2C) or reference expression profile matrices (Fig. 2D) are derived from the normalised transcriptional profiles of cell populations.

3.1. Cell type-specific signatures based on marker genes

A marker gene signature consists of a set of genes that should be expressed specifically by the cell population represented by that signature. The first approaches that were developed to build cell type-specific marker gene signatures used microarray data of purified cell populations (Table 1). One of the first large-scale efforts to build such signatures used microarray-derived expression profiles of immune cells sorted from different tissues, including peripheral blood and bone marrow [40]. Differentially expressed marker genes across cell populations were then identified using ANOVA and further refined by applying a fold-change threshold based on their median expression.

Furthermore, as the marker genes of a cell population are expected to be co-expressed, only those with an average correlation coefficient between all other markers of the same population of at least 0.6 were kept. Following this approach, a final set of 812 immune-related marker genes was obtained. The signatures derived from these markers were then used to estimate the abundance of 31 colorectal cancer-infiltrating immune cell populations [40]. The same pipeline was later applied to build signatures for 28 immune cell populations used to characterise the TME of TCGA tumours [41].

Another approach based on signatures derived from microarray data of purified stromal populations is MCP-counter [45]. In this case, however, the area under the curve (AUC) and the signal-to-noise ratio were used in addition to the expression fold-change threshold to select the marker genes. In addition, the signatures were derived taking the hierarchical classification of immune cells into account. This allowed the generation of robust signatures for both parental populations (e.g. all T cells) and subpopulations (e.g. CD8⁺ T cells). In total, 522 marker genes were derived to define ten stromal cell populations. MCP-counter was applied to estimate the abundance of these populations in a large dataset of non-hematopoietic human tumours [45]. A recent

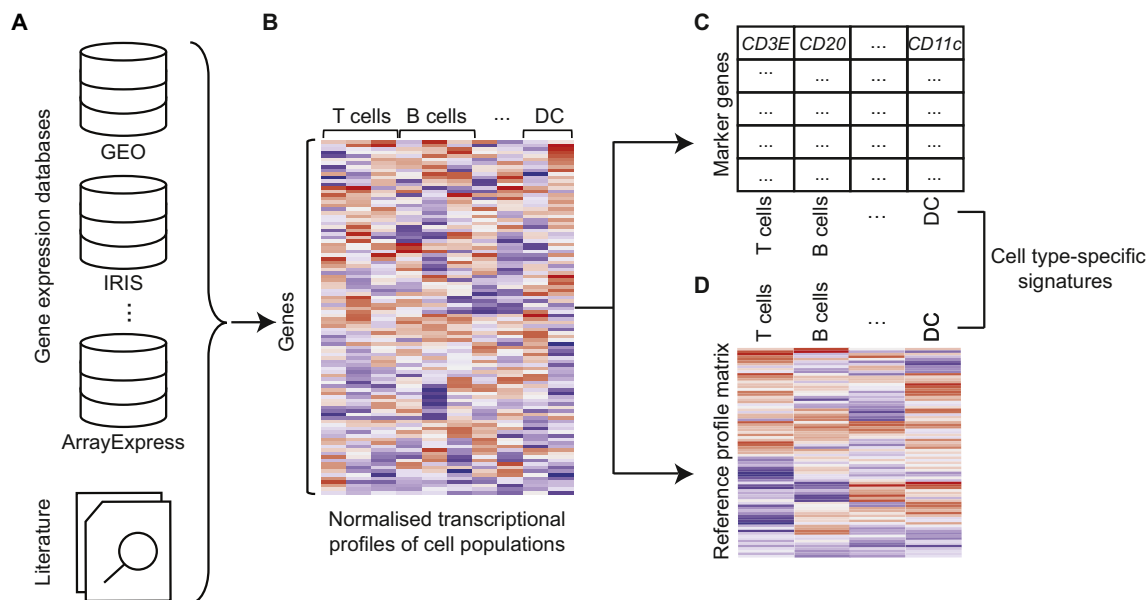


Fig. 2. Computational framework to derive reference signatures. (A) Gene expression data of purified cell populations and marker genes are collected from gene expression databases and/or the literature. (B) They are then normalised to derive cell type-specific transcriptional profiles. (C) Profiles are used to derive cell type-specific reference marker genes through differential expression and correlation analyses. (D) Alternatively, the transcriptional profiles can be aggregated to generate reference expression profile matrices. GEO = Gene Expression Omnibus database, IRIS = Immune Response In Silico database, DC = dendritic cells.

benchmark study found MCP-counter particularly reliable for the comparison of immune infiltrates across samples due to the robustness of its signatures [52]. It performed particularly well in the quantification of B cells, CD8⁺ T cells, macrophages, natural killer (NK) cells and cancer-associated fibroblasts (CAFs).

More recent approaches started to use cancer RNA-seq expression data to derive marker gene signatures. For example, xCell [43] employs signatures derived from RNA-seq, microarray and Cap Analysis of Gene Expression datasets of tumours and normal tissues from different sources. Unlike other methods, xCell uses more than one signature for each considered population. Signatures were first derived from each data source individually based on marker gene overexpression analysis with different thresholds. Then, for each data source, the top three signatures were kept based on the t-statistic of their enrichment scores (ES) between the cell population they define and all the others. A total of 489 signatures were obtained to define 64 cell populations, making xCell the broadest and most comprehensive quantification approach to date. xCell was applied to characterise infiltrates in TCGA and TARGET data [43]. In the comparative study cited above [52], xCell resulted particularly suitable to estimate the abundance of CD4⁺ T cells, T regulatory cells and endothelial cells.

In addition to deriving *ex novo* signatures, cancer RNA-seq data has also been used to refine pre-existing signatures to make them more specific for the quantification of infiltrates in tumour samples. Danaher et al. [46] were the first to derive signatures from an initial compendium of 14 previously published immune cell signatures. Using bulk RNA-seq data from 24 TCGA cancer types, the authors measured the co-expression patterns of markers associated with a given signature using a pairwise similarity metric. Then, they built a pairwise similarity matrix for each cancer type and applied hierarchical clustering using the average similarity values across the 24 cancer types. They only considered as final markers for a specific cell type the genes with the highest co-expression patterns across tumours. By using bulk RNA-seq data from the TME, the differences between intratumoral and purified immune cell expression patterns are accounted for [46].

A very similar RNA-seq dataset from TCGA was used to select the most representative signatures from an initial list of marker gene sets obtained from three literature sources [44]. The specificity of the initial signatures was assessed through a correlation analysis using the signature ESs instead of marker gene expression as in other approaches. For each literature source, a pairwise correlation matrix was computed for all the ES of the signatures across the TCGA samples. Sources were discarded when the overall correlation picture of their signatures poorly agreed with biological knowledge. For instance, sources with signatures from cell populations known to be highly co-infiltrated, but that resulted to be negatively correlated, were discarded. Compared to Danaher et al., this approach is less susceptible to the quality of gene expression data, since the correlations are done on the ES values. This strategy yielded a curated set of 16 immune signatures defined by 401 marker genes that were then used to characterise the immune infiltrates in the same TCGA cohort [44].

ConsensusTME [42] is a more inclusive approach as compared to the others because it integrated pre-existing signatures instead of refining them separately. For each cell population, a new set of markers was obtained combining previously defined sets. Additionally, genes whose expression showed a correlation coefficient higher than -0.2 with tumour purity scores derived from 32 TCGA cancer types were filtered out. This step was justified because the correlation of gene expression with tumour purity is indicative of the fact that cancer cells may express these marker genes thus invalidating their specificity for a particular stromal population [42].

In addition to using expression profiles from purified cell populations or refining previous signatures, gene sets can also be derived *ex novo* from bulk transcriptomic data. For instance, ImSig [47] relies on a collection of immune signatures derived from microarray datasets of healthy and disease human samples. For each dataset, a gene

correlation network was computed and subsequent clustering was performed to identify modules of co-expressed genes. These modules were then manually annotated to identify those corresponding to immune cell types and extract 318 associated marker genes defining seven immune cell populations. ImSig was applied to characterise the immune infiltrates in TCGA samples [47].

3.2. Cell type-specific signatures based on profile matrices

Instead of sets of marker genes, cell type-specific signatures can also consist of reference expression profile matrices of marker genes in a particular cell population. CIBERSORT [48] was the first tool to use a curated signature matrix of reference expression profiles to estimate the proportion of 22 immune cell populations. Marker genes were first selected from microarray expression data of isolated immune cells using differential expression analysis and fold-change ranking. The expression value of each marker gene and immune cell population in the reference matrix was defined as the median expression of that gene across all transcriptome profiles for that population [48]. TIMER [49] uses a different expression profile matrix for each one of 23 TCGA cancer types to estimate the abundance of six immune cell populations. In this case, marker genes were collected from the Immune Response In Silico database [53] and filtered out if positively correlated with TCGA tumour purity. Expression profiles of isolated immune cells were then obtained from the Human Primary Cell Atlas [54]. For each immune cell type, the reference profile was taken as the median expression of the filtered marker genes across corresponding transcriptome profiles. Unlike the profile matrices of the above methods, EPIC [50] was the first to use a profile matrix derived from scRNA-seq data of primary and non-lymphoid metastatic melanoma samples. Marker genes were identified by differential expression analysis and the resulting profile of a cell type was taken as the average expression of corresponding markers. Out of the considered stromal populations, EPIC was recommended for the deconvolution of B cells, CD4⁺ and CD8⁺ T cells, NK cells, CAFs and endothelial cells [52]. quanTIseq [51] was the first method to derive its signature matrix entirely from bulk RNA-seq data of purified cell populations. Marker genes were selected based on their differential expression between cell types and filtered out if highly expressed in tumour cells. The reference profile of each cell population was computed as the median expression over corresponding RNA-seq purified profiles. The approach was found to be particularly suitable for the deconvolution of regulatory and CD8⁺ T cells [52]. Notably, quanTIseq implements a whole RNA-seq data processing pipeline, from read pre-processing to TME cell type quantification. This avoids technical differences between the bulk tumour sample and the reference profile matrix.

3.3. Computational methods for the quantification of tumour-infiltrates

The cell type-specific signatures derived from either marker genes or profile matrices can then be used to quantify non-cancer cells of the TME. Computational approaches developed so far can be broadly divided into gene set scoring approaches and deconvolution approaches.

Gene set scoring approaches leverage on marker gene signatures to provide relative abundance scores indicative of how enriched a cell population of interest is in the bulk tumour sample. Most of these approaches implement Gene Set Enrichment Analysis (GSEA) methods to quantify cell populations defined by their corresponding marker gene set in each individual sample. In these GSEA-based methods, genes from bulk transcriptomic data are first ranked in decreasing order of their expression. Cell populations are then considered to be enriched or depleted if their marker genes are among the top or bottom expressed genes, respectively. An example of GSEA-based methods is single-sample GSEA (ssGSEA) [55] that computes an ES in each sample by ranking the genes according to their absolute expression value. ESs are calculated for every pair of sample and marker gene set. This is

achieved by integrating the difference between the empirical cumulative distribution of the rank-normalised gene expression inside and outside the gene set [55]. ssGSEA was directly used for the characterisation of the TME in several cancer types [40–42]. xCell uses ssGSEA for the calculation of the raw enrichment score of a cell population, which is then adjusted through a spillover technique to correct for cell type collinearity [43]. xCell is therefore less prone to background predictions, i.e. the artificial abundance estimation of cell types that are actually absent. For this reason, it was recommended for use when the main interest is to identify the presence of a particular cell population in the sample [52]. Unlike ssGSEA, Gene Set Variation Analysis (GSVA) [56] still applies GSEA but accounts for expression variability across large and heterogeneous datasets. It uses a non-parametric estimation of the cumulative density function of the expression profile of each gene. GSVA has been used to quantify tumour-immune infiltrates and characterise the immunophenotypes of TCGA samples [44].

Other gene set scoring methods that are not based on GSEA use the log-transformed geometric [45,46] or arithmetic [47] mean of the normalised marker gene expression values in the tumour sample (Table 1). Although these methods are more dependent on the quality of gene expression data than GSEA-based methods, they provide abundance scores that are directly proportional to marker gene expression [46]. This facilitates their interpretation. For instance, if marker genes associated to a particular cell population are twice as expressed in sample A than in sample B, one can infer that this cell population is twice as abundant in A than in B (assuming the absence of aberrant expression of any of those markers by some tumour cells). This fold change would not be reflected by GSEA-based approaches as they provide scores computed from gene ranks.

Deconvolution approaches estimate the fraction of each cell population in the sample from transcriptomic data using both marker gene sets and expression profile matrices. These methods consider the expression profile of a heterogeneous tissue sample as the sum of the expression profiles of the composing cell populations weighted by their relative fractions [57]. Deconvolution can be partial to find only the fraction of each cell population, or complete to derive also the associated expression profiles [57]. Partial deconvolution requires a reference expression profile matrix containing an aggregate of the expression profile of each marker gene. It is usually based on least square regression to minimise the differences between the bulk expression values and the product of the reference expression profiles with the estimated fractions [57]. Tools implementing least square regression include PERT [58], DeconRNASeq [59], TIMER [60], EPIC [50], and quanTIseq [51]. Machine learning based on nu-support vector regression (nu-SVR) has also been applied in the context of partial deconvolution, such as CIBERSORT [48] and Mysort [61]. Although nu-SVR was a first step towards handling outlier gene expression values, the recently proposed FARDEEP [62] was the first approach to directly address this issue. FARDEEP uses an adaptive least trimmed square model to detect and remove outliers prior to cell fraction estimation and thereby increase estimation robustness. All these partial deconvolution methods rely on a linear model of gene expression that considers the total bulk mRNA as the sum of the mRNAs of the composing cell populations. However, solving deconvolution equations on the linear scale is not always efficient [63]. This is because RNA-seq data generally have a skewed asymmetric distribution with a longer right tail of highly expressed genes. To account for this skewedness in gene expression data, dtangle [63] implements a multivariate logistic model that solves the linearly-modelled deconvolution problem on the logarithmic scale.

Complete deconvolution approaches, also known as unsupervised methods, estimate both cell fractions and their expression profiles [57]. Most of these methods are based on non-negative matrix factorisation that factorises the bulk expression profiles as the product of non-negative cell fractions and cell-specific profiles. Examples of tools implementing non-negative matrix factorisation include deconf [64] and a semi-supervised algorithm that incorporates prior knowledge of cell

type-specific markers [65]. Other approaches that also use cell type-specific markers are based on quadratic programming [66] or on maximum likelihood estimation [67]. Recently, DeMixT [68] has been developed to de-convolute bulk RNA-seq cancer data into tumour and stromal components. DeMixT considers the input data as a linear additive model of tumour and stroma. Then, their relative proportions and corresponding expression profiles are estimated using the iterated conditional modes algorithm and a gene-set-based component merging approach [68].

3.4. Limitations of TME quantification from bulk transcriptomic data

Both gene set scoring- and deconvolution-based approaches present several limitations when characterising the TME from bulk tumour data.

First, as mentioned above, the scores derived from gene set scoring approaches cannot be interpreted as cell type proportions within the sample. One of the reasons for this is that the sizes of the marker gene sets can be highly variable, biasing the scoring towards larger sets. Thus, gene set scoring approaches do not allow intra-sample comparisons of different cell populations. This is partially solved in deconvolution-based approaches as they provide cellular fractions that can be related to cell population abundances both within and across samples.

Second, most cell type-specific signatures are derived from expression data of cell populations that were isolated from non-cancer tissues, generally peripheral blood. This is likely to affect the abundance estimation in bulk tumour samples for at least two reasons. First, the immune cell composition varies across cancers [15]. Second, some marker genes can be expressed also by tumour cells [43]. Some approaches reduce these biases by incorporating tumour-specific expression profiles when constructing cell type-specific signatures.

Third, most partial deconvolution approaches rely on static cell type-specific signature matrices that assume constant expression profiles of the cell populations across samples. This assumption neglects sample-specific variations in time and space [57]. Moreover, given the variability and diversity of the TME, it is likely that not all cell populations are accounted for by the existing quantification approaches. In addition, not all cell populations considered by these approaches are necessarily present in the cancer samples (referred to as background predictions [52]). As a result, partial deconvolution approaches may produce under- or over-estimated cell fractions [57]. To address this, some methods avoid restricting the cell fraction estimation to the populations under consideration [50,51,69]. Instead, they estimate the fraction of uncharacterised cells within the tumour bulk to provide more accurate estimations.

Fourth, mRNA abundances across cell types are often neglected by partial deconvolution methods when estimating TME cell fractions. Only EPIC [50] and quanTIseq [51] correct for this by normalising each estimated cell type abundance by a corresponding scaling factor representing the mRNA content of that cell type. Therefore, these methods allow a more reliable comparison of cell population abundances as they can be interpreted as actual cell fractions. Both methods were recommended for immuno-oncology applications as their fractions are comparable both across and within samples [52]. An alternative approach, ABIS [70], used a reference profile matrix normalised for cell type-specific mRNA abundance instead of correcting estimated abundances by a scaling factor. However, ABIS was derived from and applied to blood-derived expression data, and has not been applied to cancer transcriptomic data yet.

Finally, often only a small set of cell populations is used to benchmark quantification approaches. This is because experimental techniques to derive ground-truth quantifications (such as flow cytometry) allow the simultaneous profiling of a limited number of cell types [43,48]. Recently, five deconvolution-based and two scoring-based approaches were systematically compared by assessing their performance on estimating nine stromal populations [52]. Four metrics

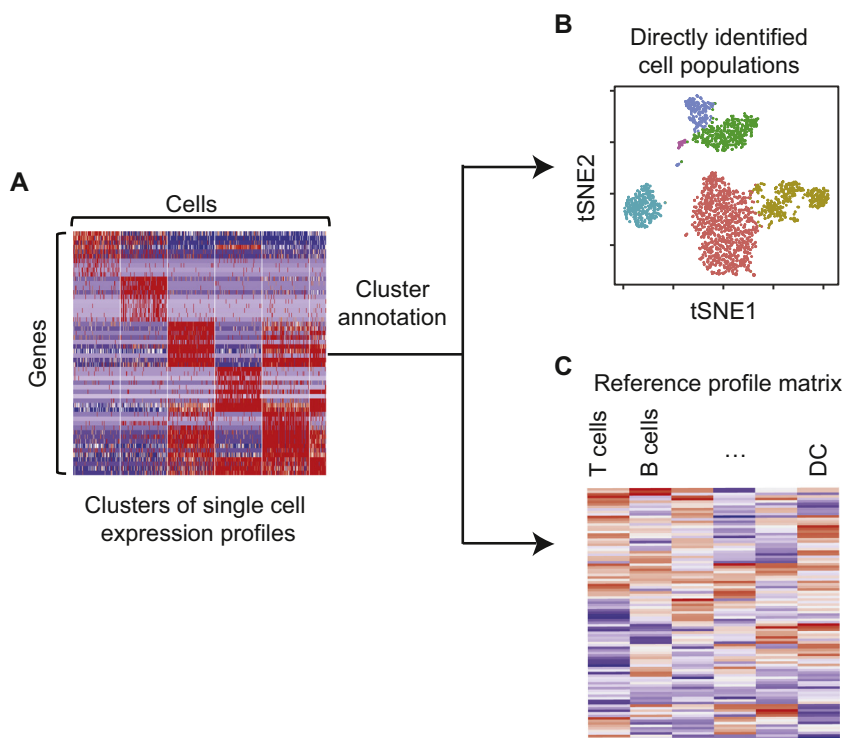


Fig. 3. Single-cell RNA-seq for the identification of TME cell populations. (A) Clustered scRNA-seq profiles of cancer samples are annotated according to the expression of known marker genes. (B) Cell populations can be directly identified from the annotated clusters, and visualised after dimensionality reduction. (C) Alternatively, the annotated clusters can also be used to derive high-resolution reference profile matrices. DC = dendritic cells, tSNE = t-distributed Stochastic Neighbour Embedding.

assessed each quantification approach: predictive performance, minimal detection fraction, background predictions, and spillover effect on both real and simulated bulk RNA-seq datasets. Spillover effect measures the over-estimation of a cell population due to the inaccurate estimation of others. Interestingly, the performance of the tested approaches varied across cell types, with poor performances on CD4⁺ T cells and dendritic cells, overall. Deconvolution-based approaches were found to be more likely to estimate minimal immune cell fractions even when these were absent (i.e. background predictions).

4. Quantification of non-cancer cells from single-cell transcriptomic data

The last five years have seen an increasing number of studies applying scRNA-seq to characterise the TME across different cancer types [19]. scRNA-seq data can be used to identify the sequenced cells directly or to generate reference expression profile matrices to deconvolute bulk RNA-seq data (Fig. 3).

TME cell populations can be directly quantified from scRNA-seq data by clustering and annotating the resulting clusters according to the expression of cell type-specific marker genes (Fig. 3A). This allows to assign the clusters to specific stromal cell populations (Fig. 3B). This approach has been used to profile the tumour infiltrates in melanoma [71], hepatocellular carcinoma [72], breast [73–75], colorectal [76] and lung cancer [77].

Cell type-specific reference expression profile matrices can also be derived from scRNA-seq data (Fig. 3C). Deconvolution-based quantification methods can then use these matrices to estimate the abundance of different cell populations from bulk tumour transcriptomic data. For example, EPIC [50] used a reference matrix derived from expression profiles of a melanoma scRNA-seq dataset [71]. This approach was later extended to two other scRNA-seq datasets (normal blood [29] and ovarian cancer [78]) to build five reference expression profile matrices [78]. Each matrix was obtained using a different strategy to average gene expression across and within single cell datasets and cell types. Then, CIBERSORT was applied with each reference matrix [48] and reference marker genes from independent sources [48,71,78]. The best

deconvolution results for ten stromal and two cancer cell types, on both simulated and real bulk RNA-seq data were obtained by averaging expression within both cell types and datasets [78]. This highlights the strong dependence of deconvolution methods on the quality of reference profile matrices. The newest version of CIBERSORT, CIBERSORTx [79], accounts for this dependency by allowing the use of reference signatures obtained from single cells or from bulk expression data. CIBERSORTx also uses nu support vector regression to estimate cell type proportions. However, before deconvolution, it performs normalisation and batch correction of platform-specific variations between the reference signatures and the bulk RNA-seq data [79].

5. Other approaches for the quantification of tumour infiltrating cells

Alternative approaches that are not based on tumour gene expression profiles can also be used to characterise the TME. They are usually based on the multidimensional analysis of proteins or RNAs from thousands of cells detected either from solid tissue sections or disaggregated tissues (Table 2). These methods can be categorised in four main groups according to the detection technology, namely chromogenic or fluorescent labelling, mass-spectrometry, and DNA probes coupled with bulk- or single-cell sequencing.

5.1. Chromogenic or fluorescent labelling methods

The latest development in IHC and IF allow multiplexed assays of tens of markers through repeated rounds of staining. They also permit the analysis of large regions of interest at high spatial resolution. For instance, the IHC-based approach SIMPLE [88] was used to quantify the association between mono-myelocytic and exhausted T-cell density and the response to GVAX vaccination in pancreatic ductal adenocarcinomas [89]. Similarly, the MultiOmyx IF platform [90] was applied to investigate resistance to rituximab-CHOP in diffuse large B-Cell lymphoma, pointing to high PD-1 expressing CD8⁺ T cells and PD-L1 expressing macrophages as mediators of resistance.

Flow cytometry employs a fluidic system and fluorescent labelling

Table 2

Non-transcriptomic approaches for the quantification of tumour-infiltrating cells. For each method, we report its detection technology and compatibility with FFPE samples, the type and the maximum number of measurable markers, and its throughput per run. For techniques providing spatial information, we also report their spatial resolution. FFPE = formalin-fixed paraffin-embedded, IHC = immunohistochemistry, IF = immunofluorescence, cyTOF = cytometry by time-of-flight, IMC = imaging mass cytometry, MIBI = multiplexed ion beam imaging, DSP = digital spatial profiling.

Method	Technology	FFPE	Markers	Throughput per run	Spatial resolution
Multiplex IHC [80]	Chromogenic-antibodies	Y	< 12 proteins	~500 mm ² /run	< 1 μm
Multiplex IF [81]	Fluorescent-antibodies	Y	< 50 proteins	~500 mm ²	< 1 μm
Flow Cytometry [82]	Fluorescent-antibodies	Y	< 28 proteins	~10 ⁷ cells	N
cyTOF [83]	Mass spectrometry	Y	< 40 proteins	~10 ⁷ cells	N
IMC [12]	Mass spectrometry	Y	< 40 proteins	~ 1 mm ²	1 μm
MIBI [13]	Mass spectrometry	Y	< 50 proteins	~ 1 mm ²	0.2 μm
Spatial transcriptomics [17]	DNA probes Bulk DNA-seq	N	> 1500 genes	1007 spots/slide	200 μm
NanoString DSP [84]	DNA probes	Y	< 40 proteins	600 μm ²	10 μm
REAP-seq [85]	Bulk DNA-seq	N	< 90 genes	~4000 cells	N
	scDNA-seq		82 surface markers		
Abseq [86]	DNA probes	N	> 600 genes	> 10,000 cells	N
	scDNA-seq		30 surface markers		
CITE-seq [87]	DNA probes	N	~1500 genes	~10,000 cells	N
	scDNA-seq		> 20 surface markers		

to isolate and characterise cells according to the expression of 5–15 markers. Since flow cytometry is not destructive, the cells can be used for further analyses. Due to its intrinsic robustness, flow cytometry is often used to validate computational methods that quantify TME populations [50]. Recently developed high dimensional flow-cytometry techniques allow the quantification of up to 28 markers in single cells [82]. Routine flow cytometry analyses cannot be applied to this technique as it requires specialised computational tools for the unbiased identification of cell populations from this larger number of markers [82].

5.2. Mass spectrometry

Mass cytometry, also known as cytometry by time-of-flight (cyTOF) [83], similarly to flow cytometry, uses a fluidic system to isolate cells. However, cyTOF marker detection is based on time-of-flight (TOF) mass spectrometry instead of fluorescence. Cells are first labelled with heavy metal-tagged antibodies, which are then distinguished according to the atomic mass of the associated metal ions. Ion counts are acquired across the mass spectra, and combined to form events as in flow-cytometry experiments. These events are then thresholded according to signal intensity across all channels to discard events caused by debris. After this filtering the event data is exported in the standard FCS format used also for flow cytometry [91]. cyTOF was recently used to characterise the TME of breast cancer leading to the identification of TME features that can be used for patient stratification [92]. Because cyTOF does not rely on fluorophores, the detection specificity is not reduced by spectral overlap and autofluorescence. This increases the number of markers that can be quantified in a single experiment. However, the limiting factor is the number of pure heavy metal isotopes available (Table 2). Despite higher specificity, cyTOF has still lower throughput than flow-cytometry (< 1000 cells/s compared to about 10,000 cells/s). Both cyTOF and high-throughput flow cytometry data differ from scRNA-seq data in two main aspects: the number of analysed cells is much higher ~10⁷, and the possibility to quantify only up to about 40 markers (Table 2). After gating to remove doublets and select only intact single cells, a multidimensional analysis of the single cell data can be performed. First unsupervised clustering is employed to group cells into different subpopulations. Then, differential cell population abundance and/or differential marker expression across different conditions can be analysed. Finally, the different cell populations and the expression of markers of interest can be visualised using dimensionality reduction approaches [91].

Other approaches leverage mass spectrometry with heavy metal ion-tagged antibodies for the imaging of FF or FFPE tissues. The best-known

examples are IMC [12] and MIBI [13], which differ mainly in the way the heavy metal ions are separated from the tissue slide. IMC uses a UV laser to ablate pre-selected areas of the tissue slide and the resulting gas is then ionised with inductively coupled plasma before TOF mass spectrometry is applied [12]. MIBI instead relies on a primary ion beam to liberate the heavy metals chelated to the antibodies as secondary ions. These ions are then analysed with sector field [13] or TOF mass analysers. The ion counts obtained from rasterising the slide with the laser or the ion beam are finally used to reconstruct a multidimensional image composed of one layer per ion/marker. The tissue areas scanned in both IMC and MIBI are much smaller than those acquired with multiplex IHC and IF. However, they provide greater sensitivity, with at least five orders of magnitude of linear dynamic range, and can use a higher number of markers. MIBI can reach a resolution higher than IMC (of < 500 nm as compared to about 1 μm). In contrast, IMC has faster scan sampling times which makes it suitable for the ablation of larger areas and has been further adapted to quantify mRNAs from FFPE tissues [93]. After removing background noise, IMC and MIBI images can be used to identify single cells with image segmentation techniques. Then, for each of these cells, the expression values of each marker can be extracted to obtain a matrix similar to those derived from cyTOF or high dimensional flow cytometry. This matrix, generally containing a much lower number of cells, generally in the order of 10³ per image, can be analysed with unsupervised clustering. Finally, the spatial information contained in the images can be leveraged to identify significant cell-cell interactions through neighbourhood analysis, or by investigating the localisation of specific cell population in the tissue.

Both IMC and MIBI have been applied to TME characterisation. For example, MIBI revealed a positive correlation between the expression of immunoregulatory proteins and the tumour-immune composition and organisation in triple negative breast cancer [94]. IMC enabled the analysis of the relationship between CD8⁺ T cell infiltration, the extracellular domain of HER2, and response to trastuzumab in breast cancer [95].

5.3. DNA probes coupled with bulk sequencing

Two recently developed techniques, spatial transcriptomics [17] and NanoString digital spatial profiling (DSP) [84], can quantify gene expression in specific areas of tissue samples. Both methods rely on DNA or DNA-RNA probes coupled with fluorescent labelling to retain spatial information of gene expression.

In spatial transcriptomics, this is achieved through an array of 100 μm-large spots of spatially barcoded oligo-dT probes. After placing the tissue on the array, mRNAs can be reverse-transcribed directly in

situ and then sequenced. The spatial barcode sequences from the array probes are retained in the RNA-seq reads and this allows to trace them back to the original spot in the tissue. Spatial transcriptomics requires intact RNA (therefore it cannot currently be applied to FFPE samples) and cannot reach single cell resolution. However, it has a higher throughput than, for example, multiplexed sequential FISH techniques (about 1000 spots per sample able to detect > 1500 genes per spot). Spatial transcriptomics also provides greater flexibility than other in situ sequencing approaches, as it does not require customised instruments. Spatial transcriptomic data consists of an expression matrix where each row corresponds to a gene and each column corresponds to a spot coordinate. An integration of scRNA-seq and spatial transcriptomics was recently applied to study the spatial composition of the TME in pancreatic ductal adenocarcinoma [96].

NanoString DSP [84] relies on the NanoString nCounter platform [14] to quantify antibody-bound proteins or hybridised transcripts using specific photocleavable DNA probes. The probes are then hybridised with complementary fluorescent-labelled RNA probes. To obtain spatial information three consecutive slides are used, one for IHC or in situ RNA hybridisation to visualise the tissue and select the area of interest and the other two for protein and RNA quantification. After the tissue area is selected, the photocleavable probes are released with UV light and collected by microcapillary aspiration for quantification with the NanoString nCounter platform. Area selection in NanoString DSP is flexible ranging from simple to complex shapes associated with tissue compartments or single cells. This approach can be used in both FF and FFPE samples, but the number of markers is limited to about 40 proteins and 90 transcripts [97]. NanoString DSP read counts are normalised using spike-in probes to account for capture and amplification efficiency. Moreover, since ROIs differ in size both within and across samples, area normalisation is also applied. Additionally, ROI background is corrected through the addition of negative RNA probes and isotype antibodies; while transcripts and antibodies against cellular proteins address differences in cellularity across ROIs. The output data consists of a matrix with the normalised intensities of the protein and mRNA markers in each ROI [97]. NanoString DSP has been recently applied to quantify 32 proteins and 82 transcripts in tumour and stromal regions of non-small cell lung cancer [97].

The characterisation of tissue regions from marker expression is achieved by processing the expression matrices with dimensionality reduction followed by hierarchical clustering. The clustered features can then be placed back on the tissue images to relate them with tissue architecture [98].

5.4. DNA probes coupled with single-cell sequencing

In the past three years single-cell approaches that integrate transcriptomics and cell-surface protein quantification have emerged [20]. These approaches quantify protein expression in single cells through DNA-tagged antibodies. In parallel they allow RNA expression profiling in the same cell using microdroplet- or microwell- based scRNA-seq [99]. The most widely used technologies implementing this approach are REAP-seq [85], Abseq [86] and CITE-seq [87]. REAP-seq is based on the Chromium sequencing platform [29] and, while it has a relatively low throughput (about 4000 cells per run), it allows the quantification of up to 82 different proteins. Abseq relies on the BD Rhapsody sequencing platform [100] and can quantify of up to 600 genes and 30 proteins in > 10,000 cells [86]. Finally, CITE-seq [87] uses either Drop-seq [30] or other microdroplet-based technologies to measure about 1500 genes and 20 proteins in > 10,000 cells [101]. These high throughput scRNA-seq methods allow to perform multimodal RNA-protein analyses on large single-cell datasets. For example, CITE-seq has been used to characterise rare immune cell phenotypes by splitting scRNA-seq derived clusters into subsets with high and low expression of specific surface markers [87].

6. Conclusion

The success of cancer immunotherapy has led to an increased interest in the fine characterisation of TME composition. This is indeed the first step to understand how the TME influences response to therapy [2]. In addition to a better knowledge of the interactions between cancer and non-cancer cells, TME characterisation can also be exploited as biomarker for patient stratification and prognosis. For example, the quantification of tumour infiltrating CD3⁺ and CD8⁺ T cells using digital pathology from IHC slides has a validated prognostic value for predicting colorectal cancer recurrence [8]. This measure, called Immunoscore, represents the first step towards the adoption of standardised immune-based assays for colorectal cancer classification. Other similar efforts are extending this approach to a broader set of cancer types [6].

Despite their undoubted utility, the incorporation of technically sophisticated methods that allow a thorough analysis of the TME in the clinical setting is still challenging. Indeed, these techniques are usually expensive, highly sensitive to the quality of the input material and require specialised expertise for their analysis. This is particularly the case for the more recent approaches such as high throughput scRNA-seq and mass spectrometry-based imaging. Moreover, the turnaround time is often not compatible with the decision-making process of the clinical practice. Further efforts are needed to harmonise the depth and specificity of the TME analysis achieved in the research setting to the requirements of time and cost-effective clinical assays.

Future developments in the characterisation of the TME should incorporate spatial information and integrate different types of omic data. Emerging approaches have already started to link the expression of marker genes to their localisation within the tissue enabling a deeper understanding of the tumour-TME interactions. However, these approaches are currently limited to tiny regions of the tumour that may not be representative of the whole tumour mass. Increasing the tissue area that can be analysed will also increase the robustness of the results. Similarly, new technologies enabling the simultaneous multi-omic analyses of the TME are being developed, particularly in the single cell setting. They combine scRNA-seq, scDNA-seq, single cell T and B cell receptor sequencing, single cell epigenomics and small and non-coding scRNA-seq [101]. In combination with functional studies, these techniques will enable a further in-depth description of all the cell populations constituting the TME and their interactions [20].

We are at the beginning of an exciting era where technological innovations can effectively contribute to improve not only our understanding of cancer biology, but also the way we treat cancer patients.

Author contributions

All authors contributed to write the manuscript.

Funding

This study has received funding from Cancer Research UK [C43634/A25487], the Cancer Research UK King's Health Partners Centre at King's College London [C604/A25135 and C604/A25189], the CRUK City of London Centre Award [C7893/A26233], and the European Union's Horizon 2020 research and innovation programme under the Marie Skłodowska-Curie grant agreement No CONTRA-766030.

Transparency document

The [transparency document](#) associated with this article can be found in the online version.

Declaration of competing interest

The authors declare that no potential conflicts of interest were

disclosed.

References

- [1] S. Frank, *Dynamics of Cancer: Incidence, Inheritance, and Evolution*, Princeton University Press, Place Published, 2007.
- [2] M. Binnewies, E.W. Roberts, K. Kersten, V. Chan, D.F. Fearon, M. Merad, L.M. Coussens, D.I. Gabrilovich, S. Ostrand-Rosenberg, C.C. Hedrick, R.H. Vonderheide, M.J. Pittet, R.K. Jain, W. Zou, T.K. Howcroft, E.C. Woodhouse, R.A. Weinberg, M.F. Krummel, Understanding the tumor immune microenvironment (TIME) for effective therapy, *Nat. Med.* 24 (2018) 541–550.
- [3] M.F. Sanmamed, L. Chen, A paradigm shift in cancer immunotherapy: from enhancement to normalization, *Cell* 175 (2018) 313–326.
- [4] A. Ribas, J.D. Wolchok, Cancer immunotherapy using checkpoint blockade, *Science* 359 (2018) 1350–1355.
- [5] J.J. Havel, D. Chowell, T.A. Chan, The evolving landscape of biomarkers for checkpoint inhibitor immunotherapy, *Nat. Rev. Cancer* 19 (2019) 133–150.
- [6] S. Hendry, R. Salgado, T. Gevaert, P.A. Russell, T. John, B. Thapa, M. Christie, K. van de Vijver, M.V. Estrada, P.I. Gonzalez-Ericsson, M. Sanders, B. Solomon, C. Solinas, G. Van den Eynde, Y. Allory, M. Preusser, J. Hainfellner, G. Pruneri, A. Vingiani, S. Demaria, F. Symmans, P. Nuciforo, L. Comerma, E.A. Thompson, S. Lakhani, S.R. Kim, S. Schnitt, C. Colpaert, C. Sotiriou, S.J. Scherer, M. Ignatiadis, S. Badve, R.H. Pierce, G. Viale, N. Sirtaine, F. Penault-Llorca, T. Sugie, S. Fineberg, S. Paik, A. Srinivasan, A. Richardson, Y. Wang, E. Chmielik, J. Brock, D.B. Johnson, J. Balko, S. Wienert, V. Bossuyt, S. Michiels, N. Ternes, N. Burchardi, S.J. Luen, P. Savas, F. Klauschen, P.H. Watson, B.H. Nelson, C. Criscitiello, S. O'Toole, D. Larsimont, R. de Wind, G. Curigliano, F. Andre, M. Lacroix-Triki, M. van de Vijver, F. Rojo, G. Floris, S. Bedri, J. Sparano, D. Rimm, T. Nielsen, Z. Kos, S. Hewitt, B. Singh, G. Farshid, S. Loibl, K.H. Allison, N. Tung, S. Adams, K. Willard-Gallo, H.M. Horlings, L. Gandhi, A. Moreira, F. Hirsch, M.V. Dieci, M. Urbanowicz, I. Brcic, K. Korski, F. Gaire, H. Koepfen, A. Lo, J. Giltman, M.C. Rebelatto, K.E. Steele, J. Zha, K. Emancipator, J.W. Juco, C. Denkert, J. Reis-Filho, S. Loi, S.B. Fox, Assessing tumor-infiltrating lymphocytes in solid tumors: a practical review for pathologists and proposal for a standardized method from the international immuno-oncology biomarkers working group: part 2: TILs in melanoma, gastrointestinal tract carcinomas, non-small cell lung carcinoma and mesothelioma, endometrial and ovarian carcinomas, squamous cell carcinoma of the head and neck, genitourinary carcinomas, and primary brain tumors, *Adv. Anat. Pathol.* 24 (2017) 311–335.
- [7] B. Weide, A. Martens, H. Zelba, C. Stutz, E. Derhovanessian, A.M. Di Giacomo, M. Maio, A. Sucker, B. Schilling, D. Schadendorf, P. Büttner, C. Garbe, G. Pawelec, Myeloid-derived suppressor cells predict survival of patients with advanced melanoma: comparison with regulatory T cells and NY-ESO-1- or melan-A-specific T cells, *Clin. Cancer Res.* 20 (2014) 1601.
- [8] F. Pagès, B. Mlecnik, F. Marliot, G. Bindea, F.-S. Ou, C. Bifulco, A. Lugli, I. Zlobec, T.T. Rau, M.D. Berger, I.D. Nagtegaal, E. Vink-Börger, A. Hartmann, C. Geppert, J. Kolwelter, S. Merkel, R. Grützmann, M. Van den Eynde, A. Jouret-Mourin, A. Kartheiser, D. Léonard, C. Remue, J.Y. Wang, P. Bavi, M.H.A. Roehrl, P.S. Ohashi, L.T. Nguyen, S. Han, H.L. MacGregor, S. Hafezi-Bakhtiari, B.G. Wouters, G.V. Masucci, E.K. Andersson, E. Zavadova, M. Vocka, J. Spacek, L. Petruzella, B. Konopasek, P. Dünd, H. Skalova, K. Nemejcova, G. Botti, F. Tatangelo, P. Delrio, G. Ciliberto, M. Maio, L. Laghi, F. Grizzi, T. Fredriksen, B. Buttard, M. Angelova, A. Vasaturo, P. Maby, S.E. Church, H.K. Angell, L. Lafontaine, D. Bruni, C. El Sissy, N. Haicheur, A. Kirilovsky, A. Berger, C. Lagorce, J.P. Meyers, C. Paustian, Z. Feng, C. Ballesteros-Merino, J. Dijkstra, C. van de Water, S. van Lent-van Vliet, N. Knijn, A.-M. Muşină, D.-V. Scripcariu, B. Popivanova, M. Xu, T. Fujita, S. Hazama, N. Suzuki, H. Nagano, K. Okuno, T. Torigoe, N. Sato, T. Furuhashi, I. Takemasa, K. Itoh, P.S. Patel, H.H. Vora, B. Shah, J.B. Patel, K.N. Rajvik, S.J. Pandya, S.N. Shukla, Y. Wang, G. Zhang, Y. Kawakami, F.M. Marincola, P.A. Ascierto, D.J. Sargent, B.A. Fox, J. Galon, International validation of the consensus Immunoscore for the classification of colon cancer: a prognostic and accuracy study, *Lancet* 391 (2018) 2128–2139.
- [9] M. Rakaee, T.K. Kilvaer, S.M. Dalen, E. Richardsen, E.-E. Paulsen, S.M. Hald, S. Al-Saad, S. Andersen, T. Donnem, R.M. Bremnes, L.-T. Busund, Evaluation of tumor-infiltrating lymphocytes using routine H&E slides predicts patient survival in resected non-small cell lung cancer, *Hum. Pathol.* 79 (2018) 188–198.
- [10] R.E. Parra, A. Francisco-Cruz, I.I. Wistuba, State-of-the-art of profiling immune contexture in the era of multiplexed staining and digital analysis to study paraffin tumor tissues, *Cancers*, 11, (2019).
- [11] C.M. Mousset, W. Hobo, R. Woestenenk, F. Preijers, H. Dolstra, A.B. van der Waart, Comprehensive phenotyping of T cells using flow cytometry, *Cytometry A*, 2019.
- [12] C. Giesen, H.A.O. Wang, D. Schapiro, N. Zivanovic, A. Jacobs, B. Hattendorf, P.J. Schöffler, D. Grolimund, J.M. Buhmann, S. Brandt, Z. Varga, P.J. Wild, D. Günther, B. Bodenmiller, Highly multiplexed imaging of tumor tissues with subcellular resolution by mass cytometry, *Nat. Methods* 11 (2014) 417.
- [13] M. Angelo, S.C. Bendall, R. Finck, M.B. Hale, C. Hitzman, A.D. Borowsky, R.M. Levenson, J.B. Lowe, S.D. Liu, S. Zhao, Y. Natkunam, G.P. Nolan, Multiplexed ion beam imaging of human breast tumors, *Nat. Med.* 20 (2014) 436.
- [14] G.K. Geiss, R.E. Bumgarner, B. Birditt, T. Dahl, N. Dowidar, D.L. Dunaway, H.P. Fell, S. Ferree, R.D. George, T. Grogan, J.J. James, M. Maysuria, J.D. Mitton, P. Oliveri, J.L. Osborn, T. Peng, A.L. Ratcliffe, P.J. Webster, E.H. Davidson, L. Hood, K. Dimitrov, Direct multiplexed measurement of gene expression with color-coded probe pairs, *Nat. Biotechnol.* 26 (2008) 317.
- [15] F. Finotello, Z. Trajanoski, Quantifying tumor-infiltrating immune cells from transcriptomics data, *Cancer Immunol. Immunother.* 67 (2018) 1031–1040.
- [16] F. Petitprez, C.-m. Sun, L. Lacroix, C. Sautès-Fridman, A. de Reyniès, W.H. Fridman, Quantitative Analyses of the Tumor Microenvironment Composition and Orientation in the Era of Precision Medicine, *Frontiers in Oncology* 8 (2018) 390.
- [17] F. Salmén, P.L. Ståhl, A. Mollbrink, J.F. Navarro, S. Vickovic, J. Frišén, J. Lundberg, Barcoded solid-phase RNA capture for spatial transcriptomics profiling in mammalian tissue sections, *Nat. Protoc.* 13 (2018) 2501–2534.
- [18] J.H. Lee, E.R. Daugharthy, J. Scheiman, R. Kalhor, T.C. Ferrante, R. Terry, B.M. Turczyk, J.L. Yang, H.S. Lee, J. Aach, K. Zhang, G.M. Church, Fluorescent in situ sequencing (FISSEQ) of RNA for gene expression profiling in intact cells and tissues, *Nat. Protoc.* 10 (2015) 442–458.
- [19] F. Valdes-Mora, K. Handler, A.M.K. Law, R. Salomon, S.R. Oakes, C.J. Ormandy, D. Gallego-Ortega, Single-cell transcriptomics in cancer immunobiology: the future of precision oncology, *Front. Immunol.* 9 (2018) 2582.
- [20] F. Finotello, F. Eduati, Multi-omics profiling of the tumor microenvironment: paving the way to precision immuno-oncology, *Front. Oncol.* 8 (2018) 430.
- [21] L.C. Wehmas, C.E. Wood, R. Gagne, A. Williams, C. Yauk, M.M. Gosink, D. Dalmás, R. Hao, R. O'Leone, S. Hester, Demodifying RNA for transcriptomic analyses of archival formalin-fixed paraffin-embedded samples, *Toxicol. Sci.* 162 (2017) 535–547.
- [22] Q. Sheng, K. Vickers, S. Zhao, J. Wang, D.C. Samuels, O. Koues, Y. Shyr, Y. Guo, Multi-perspective quality control of Illumina RNA sequencing data analysis, *Briefings in Functional Genomics* 16 (2016) 194–204.
- [23] J. Li, C. Fu, T.P. Speed, W. Wang, W.F. Symmans, Accurate RNA sequencing from formalin-fixed cancer tissue to represent high-quality transcriptome from frozen tissue, *JCO precision oncology* 2 (2018) 1–9.
- [24] Y. Guo, Q. Sheng, J. Li, F. Ye, D.C. Samuels, Y. Shyr, Large scale comparison of gene expression levels by microarrays and RNAseq using TCGA data, *PLoS One* 8 (2013) e71462.
- [25] I. Illumina, Illumina, Exon Capture Protocol, 2017.
- [26] X. Zhang, T. Li, F. Liu, Y. Chen, J. Yao, Z. Li, Y. Huang, J. Wang, Comparative analysis of droplet-based ultra-high-throughput single-cell RNA-Seq systems, *Molecular Cell*, 73 e135 (2019) 130–142.
- [27] D.A. Jaitin, E. Kenigsberg, H. Keren-Shaul, N. Elefant, F. Paul, I. Zaretsky, A. Mildner, N. Cohen, S. Jung, A. Tanay, I. Amit, Massively parallel single-cell RNA-Seq for marker-free decomposition of tissues into cell types, *Science* 343 (2014) 776.
- [28] T.M. Gierahn, M.H. Wadsworth II, T.K. Hughes, B.D. Bryson, A. Butler, R. Satija, S. Fortune, J.C. Love, A.K. Shalek, Seq-Well: portable, low-cost RNA sequencing of single cells at high throughput, *Nat. Methods* 14 (2017) 395.
- [29] G.X.Y. Zheng, J.M. Terry, P. Belgrader, P. Ryvkin, Z.W. Bent, R. Wilson, S.B. Ziraldo, T.D. Wheeler, G.P. McDermott, J. Zhu, M.T. Gregory, J. Shuga, L. Montesclaros, J.G. Underwood, D.A. Masquelier, S.Y. Nishimura, M. Schnall-Levin, P.W. Wyatt, C.M. Hindson, R. Bharadwaj, A. Wong, K.D. Ness, L.W. Beppu, H.J. Deeg, C. McFarland, K.R. Loebe, W.J. Valente, N.G. Ericson, E.A. Stevens, J.P. Radich, T.S. Mikkelsen, B.J. Hindson, J.H. Bielas, Massively parallel digital transcriptional profiling of single cells, *Nat. Commun.* 8 (2017) 14049.
- [30] Z. Evan, A. Macosko, R. Basu, J. Satija, K. Nemes, M. Shekhar, I.T. Goldman, Allison R. Bialas, N. Kamitaki, Emily M. Martersteck, John J. Trombetta, David A. Weitz, Joshua R. Sanes, Alex K. Shalek, A. Regev, Steven A. McCarroll, Highly parallel genome-wide expression profiling of individual cells using nanoliter droplets, *Cell* 161 (2015) 1202–1214.
- [31] A.M. Klein, L. Mazutis, I. Akartuna, N. Tallapragada, A. Veres, V. Li, L. Peshkin, D.A. Weitz, M.W. Kirschner, Droplet barcoding for single-cell transcriptomics applied to embryonic stem cells, *Cell* 161 (2015) 1187–1201.
- [32] A.A. Kolodziejczyk, J.K. Kim, V. Svensson, J.C. Marioni, S.A. Teichmann, The technology and biology of single-cell RNA sequencing, *Mol. Cell* 58 (2015) 610–620.
- [33] Q.H. Nguyen, N. Pervolarakis, K. Nee, K. Kessenbrock, Experimental considerations for single-cell RNA sequencing approaches, *Frontiers in Cell and Developmental Biology* 6 (2018) 108.
- [34] S. Nichterwitz, G. Chen, J. Aguila Benitez, M. Yilmaz, H. Storvall, M. Cao, R. Sandberg, Q. Deng, E. Hedlund, Laser capture microscopy coupled with Smart-seq2 for precise spatial transcriptomic profiling, *Nat. Commun.* 7 (2016) 12139.
- [35] S. Islam, A. Zeisel, S. Joost, G. La Manno, P. Zajac, M. Kasper, P. Lönnerberg, S. Linnarsson, Quantitative single-cell RNA-seq with unique molecular identifiers, *Nat. Methods* 11 (2013) 163.
- [36] A.A. AlJanahi, M. Danielsen, C.E. Dunbar, An introduction to the analysis of single-cell RNA-sequencing data, *Molecular Therapy - Methods & Clinical Development* 10 (2018) 189–196.
- [37] S. Picelli, Å. Björklund, O.R. Faridani, S. Sagasser, G. Winberg, R. Sandberg, Smart-seq2 for sensitive full-length transcriptome profiling in single cells, *Nat. Methods* 10 (2013) 1096.
- [38] J.W. Foley, C. Zhu, P. Jolivet, S.X. Zhu, P. Lu, M.J. Meaney, R.B. West, Gene-expression profiling of single cells from archival tissue with laser-capture microdissection and Smart-3SEQ, *bioRxiv*, (2019) 207340.
- [39] R. Rostom, V. Svensson, S.A. Teichmann, G. Kar, Computational approaches for interpreting scRNA-seq data, *FEBS Lett.* 591 (2017) 2213–2225.
- [40] M. Angelova, P. Charoentong, H. Hackl, M.L. Fischer, R. Snajder, A.M. Krogsdam, M.J. Waldner, G. Bindea, B. Mlecnik, J. Galon, Z. Trajanoski, Characterization of the immunophenotypes and antigenomes of colorectal cancers reveals distinct tumor escape mechanisms and novel targets for immunotherapy, *Genome Biol.* 16 (2015) 64.
- [41] P. Charoentong, F. Finotello, M. Angelova, C. Mayer, M. Efremova, D. Rieder, H. Hackl, Z. Trajanoski, Pan-cancer immunogenomic analyses reveal genotype-immunophenotype relationships and predictors of response to checkpoint blockade, *Cell Rep.* 18 (2017) 248–262.

- [42] A. Jiménez-Sánchez, O. Cast, M. Miller, Comprehensive benchmarking and integration of tumour microenvironment cell estimation methods, *bioRxiv*, (2018) 437533.
- [43] D. Aran, Z. Hu, A.J. Butte, xCell: digitally portraying the tissue cellular heterogeneity landscape, *Genome Biol.* 18 (2017) 220.
- [44] D. Tamborero, C. Rubio-Perez, F. Muiños, R. Sabarinathan, J.M. Piulats, A. Muntassel, R. Dienstmann, N. Lopez-Bigas, A. Gonzalez-Perez, A pan-cancer landscape of interactions between solid tumors and infiltrating immune cell populations, *Clin. Cancer Res.* 24 (2018) 3717–3728.
- [45] E. Becht, N.A. Giraldo, L. Lacroix, B. Buttard, N. Elarouci, F. Petitprez, J. Selves, P. Laurent-Puig, C. Sautès-Fridman, W.H. Fridman, A. de Reyniès, Estimating the population abundance of tissue-infiltrating immune and stromal cell populations using gene expression, *Genome Biol.* 17 (2016) 218.
- [46] P. Danaher, S. Warren, L. Dennis, L. D'Amico, A. White, M.L. Disis, M.A. Geller, K. Odunsi, J. Beechem, S.P. Fling, Gene expression markers of tumor infiltrating leukocytes, *Journal for ImmunoTherapy of Cancer* 5 (2017) 1–15.
- [47] A.J. Nirmal, T. Regan, B.B. Shih, D.A. Hume, A.H. Sims, T.C. Freeman, Immune cell gene signatures for profiling the microenvironment of solid tumors, *Cancer Immunology Research*, 44 (2018) canimm.0342.2018.
- [48] A.M. Newman, C.L. Liu, M.R. Green, A.J. Gentles, W. Feng, Y. Xu, C.D. Hoang, M. Diehn, A.A. Alizadeh, Robust enumeration of cell subsets from tissue expression profiles, *Nat. Methods* 12 (2015) 453–457.
- [49] B. Li, E. Severson, J.C. Pignon, H. Zhao, T. Li, J. Novak, P. Jiang, H. Shen, J.C. Aster, S. Rodig, S. Signoretti, J.S. Liu, X.S. Liu, Comprehensive analyses of tumor immunity: implications for cancer immunotherapy, *Genome Biol.* 17 (2016) 1–16.
- [50] J. Racle, K. de Jonge, P. Baumgaertner, D.E. Speiser, D. Gfeller, Simultaneous enumeration of cancer and immune cell types from bulk tumor gene expression data, *eLife* 6 (2017).
- [51] F. Finotello, C. Mayer, C. Plattner, G. Laschober, D. Rieder, H. Hackl, A. Krogsdam, M.V. Posch, D. Wilflingseder, S. Soppor, D. Johnson, Y. Xu, Y. Wang, M.E. Sanders, M.V. Estrada, P. Ericsson-gonzalez, J. Balko, N.D. Miranda, Z. Trajanoski, *quantIseq: Quantifying Immune Contexture of Human Tumors*, 2017.
- [52] G. Sturm, F. Finotello, F. Petitprez, J.D. Zhang, J. Baumbach, W.H. Fridman, M. List, T. Anechik, Comprehensive evaluation of transcriptome-based cell-type quantification methods for immuno-oncology, *Bioinformatics* 35 (2019) i436–i445.
- [53] A.R. Abbas, D. Baldwin, Y. Ma, W. Ouyang, A. Gurney, F. Martin, S. Fong, M. van Lookeren Campagne, P. Godowski, P.M. Williams, A.C. Chan, H.F. Clark, Immune response in silico (IRIS): immune-specific genes identified from a compendium of microarray expression data, *Genes & Immunity* 6 (2005) 319–331.
- [54] N.A. Mabbott, J. Baillie, H. Brown, T.C. Freeman, D.A. Hume, An expression atlas of human primary cells: inference of gene function from coexpression networks, *BMC Genomics* 14 (2013) 632.
- [55] D.A. Barbie, P. Tamayo, J.S. Boehm, S.Y. Kim, S.E. Moody, I.F. Dunn, A.C. Schinzel, P. Sandy, E. Meylan, C. Scholl, S. Fröhling, E.M. Chan, M.L. Sos, K. Michel, C. Mermel, S.J. Silver, B.A. Weir, J.H. Reiling, Q. Sheng, P.B. Gupta, R.C. Wadlow, H. Le, S. Hoersch, B.S. Wittner, S. Ramaswamy, D.M. Livingston, D.M. Sabatini, M. Meyerson, R.K. Thomas, E.S. Lander, J.P. Mesirov, D.E. Root, D.G. Gilliland, T. Jacks, W.C. Hahn, Systematic RNA interference reveals that oncogenic KRAS-driven cancers require TBK1, *Nature* 462 (2009) 108–112.
- [56] S. Hänzelmann, R. Castelo, J. Guinney, GSEA: gene set variation analysis for microarray and RNA-Seq data, *BMC Bioinformatics* 14 (2013) 7.
- [57] F. Avila Cobos, J. Vandesompele, P. Mestdagh, K. De Preter, Computational deconvolution of transcriptomics data from mixed cell populations, Oxford University Press, *Bioinformatics*, 2018, pp. 1969–1979.
- [58] W. Qiao, G. Quon, E. Cszasz, M. Yu, Q. Morris, P.W. Zandstra, PERT: a method for expression deconvolution of human blood samples from varied microenvironmental and developmental conditions, *PLoS Comput. Biol.* 8 (2012) e1002838.
- [59] T. Gong, J.D. Szustakowski, DeconRNASeq: a statistical framework for deconvolution of heterogeneous tissue samples based on mRNA-Seq data, *Bioinformatics* 29 (2013) 1083–1085.
- [60] T. Li, J. Fan, B. Wang, N. Traugh, Q. Chen, J.S. Liu, B. Li, X.S. Liu, TIMER: a web server for comprehensive analysis of tumor-infiltrating immune cells, *Cancer Res.* 77 (2017) e108–e110.
- [61] H.H.-S. Lu, S.-Y. Su, W.-Y. Kuo, W.-C. Chung, S.-H. Chen, J.-M. Ho, C.-Y. Lin, A gene profiling deconvolution approach to estimating immune cell composition from complex tissues, *BMC Bioinformatics* 19 (2018) 154.
- [62] Y. Hao, M. Yan, B.R. Heath, Y.L. Lei, Y. Xie, Fast and robust deconvolution of tumor infiltrating lymphocyte from expression profiles using least trimmed squares, *PLoS Comput. Biol.* 15 (2019) e1006976.
- [63] G.J. Hunt, S. Freytag, M. Bahlo, J.A. Gagnon-Bartsch, dtangle: accurate and robust cell type deconvolution, *Bioinformatics* 35 (12) (2018) 2093–2099.
- [64] D. Reipsilber, S. Kern, A. Telaar, G. Walzl, G.F. Black, J. Selbig, S.K. Parida, S.H. Kaufmann, M. Jacobsen, Biomarker discovery in heterogeneous tissue samples – taking the in-silico deconvolution approach, *BMC Bioinformatics* 11 (2010) 27–42.
- [65] R. Gaujoux, C. Seoighe, Semi-supervised nonnegative matrix factorization for gene expression deconvolution: a case study, *Infect. Genet. Evol.* 12 (2012) 913–921.
- [66] Y. Zhong, Y.-W. Wan, K. Pang, L.M. Chow, Z. Liu, Digital sorting of complex tissues for cell type-specific gene expression profiles, *BMC Bioinformatics* 14 (2013) 89.
- [67] D.A. Liebner, K. Huang, J.D. Parvin, MMAD: microarray microdissection with analysis of differences is a computational tool for deconvoluting cell type-specific contributions from tissue samples, *Bioinformatics* 30 (2014) 682–689.
- [68] Z. Wang, S. Cao, J.S. Morris, J. Ahn, R. Liu, S. Tyekucheva, F. Gao, B. Li, W. Lu, X. Tang, I.I. Wistuba, M. Bowden, L. Mucci, M. Loda, G. Parmigiani, C.C. Holmes, W. Wang, Transcriptome Deconvolution of Heterogeneous Tumor Samples with Immune Infiltration, *iScience* 9 (2018) 451–460.
- [69] C. Maurer, S.R. Holmstrom, J. He, P. Laise, T. Su, A. Ahmed, H. Hibshoosh, J.A. Chabot, P.E. Oberstein, A.R. Sepulveda, J.M. Genkinger, J. Zhang, A.C. Iuga, M. Bansal, A. Califano, K.P. Olive, Experimental microdissection enables functional harmonisation of pancreatic cancer subtypes, *Gut* 68 (2019) 1034.
- [70] G. Monaco, B. Lee, W. Xu, S. Mustafah, Y.Y. Hwang, C. Carré, N. Burdin, L. Visan, M. Ceccarelli, M. Poidinger, A. Zippelius, J. Pedro de Magalhães, A. Larbi, RNA-Seq signatures normalized by mRNA abundance allow absolute deconvolution of human immune cell types, *Cell Reports*, 26 e1627 (2019) 1627–1640.
- [71] I. Tirosh, B. Izar, S.M. Prakadan, M.H. Wadsworth, D. Treacy, J.J. Trombetta, A. Rotem, C. Rodman, C. Lian, G. Murphy, M. Fallahi-Sichani, K. Dutton-Regester, J.-R. Lin, O. Cohen, P. Shah, D. Lu, A.S. Genshaft, T.K. Hughes, C.G.K. Ziegler, S.W. Kazer, A. Gaillard, K.E. Kolb, A.-C. Villani, C.M. Johansson, A.Y. Andreev, E.M. Van Allen, M. Bertagnolli, P.K. Sorger, R.J. Sullivan, K.T. Flaherty, D.T. Frederick, J. Jané-Valbuena, C.H. Yoon, O. Rozenblatt-Rosen, A.K. Shalek, A. Regev, L.A. Garraway, Dissecting the multicellular ecosystem of metastatic melanoma by single-cell RNA-seq, *Science* 352 (2016) 189–196.
- [72] C. Zheng, L. Zheng, J.K. Yoo, H. Guo, Y. Zhang, X. Guo, B. Kang, R. Hu, J.Y. Huang, Q. Zhang, Z. Liu, M. Dong, X. Hu, W. Ouyang, J. Peng, Z. Zhang, Landscape of infiltrating T cells in liver cancer revealed by single-cell sequencing, *Cell* 169 (2017) 1342–1356 (e1316).
- [73] P. Savas, B. Virassamy, C. Ye, A. Salim, C.P. Mintoff, F. Caramia, R. Salgado, D.J. Byrne, Z.L. Teo, S. Dushyanthen, A. Byrne, L. Wein, S.J. Luen, C. Poliness, S.S. Nightingale, A.S. Skandarajah, D.E. Gyorki, C.M. Thornton, P.A. Beavis, S.B. Fox, C. Kathleen Cuningham, Foundation Consortium for Research into Familial Breast, P.K. Darcy, T.P. Speed, L.K. Mackay, P.J. Neeson, S. Loi, Single-cell profiling of breast cancer T cells reveals a tissue-resident memory subset associated with improved prognosis, *Nat. Med.* 24 (2018) 986–993.
- [74] W. Chung, H.H. Eum, H.O. Lee, K.M. Lee, H.B. Lee, K.T. Kim, H.S. Ryu, S. Kim, J.E. Lee, Y.H. Park, Z. Kan, W. Han, W.Y. Park, Single-cell RNA-seq enables comprehensive tumour and immune cell profiling in primary breast cancer, *Nat. Commun.* 8 (2017) 15081.
- [75] E. Azizi, A.J. Carr, G. Plitas, A.E. Cornish, C. Konopacki, S. Prabhakaran, J. Nainys, K. Wu, V. Kiseliovas, M. Setty, K. Choi, R.M. Fromme, P. Dao, P.T. McKenney, R.C. Wasti, K. Kadaveru, L. Mazutis, A.Y. Rudensky, D. Pe'er, Single-cell map of diverse immune phenotypes in the breast tumor microenvironment, *Cell*, 174 e1236 (2018) 1293–1308.
- [76] H. Li, E.T. Courtois, D. Sengupta, Y. Tan, K.H. Chen, J.J.L. Goh, S.L. Kong, C. Chua, L.K. Hon, W.S. Tan, M. Wong, P.J. Choi, L.J.K. Wee, A.M. Hillmer, I.B. Tan, P. Robson, S. Prabhakar, Reference component analysis of single-cell transcriptomes elucidates cellular heterogeneity in human colorectal tumors, *Nat. Genet.* 49 (2017) 708.
- [77] X. Guo, Y. Zhang, L. Zheng, C. Zheng, J. Song, Q. Zhang, B. Kang, Z. Liu, L. Jin, R. Xing, R. Gao, L. Zhang, M. Dong, X. Hu, X. Ren, D. Kirchhoff, H.G. Roeder, T. Yan, Z. Zhang, Global characterization of T cells in non-small-cell lung cancer by single-cell sequencing, *Nat. Med.* 24 (2018) 978–985.
- [78] M. Schelker, S. Feau, J. Du, N. Ranu, E. Klipp, G. MacBeath, B. Schoeberl, A. Raue, Estimation of immune cell content in tumour tissue using single-cell RNA-seq data, *Nat. Commun.* 8 (2017) 2032.
- [79] A.M. Newman, C.B. Steen, C.L. Liu, A.J. Gentles, A.A. Chaudhuri, F. Scherer, M.S. Khodadoust, M.S. Esfahani, B.A. Luca, D. Steiner, M. Diehn, A.A. Alizadeh, Determining cell type abundance and expression from bulk tissues with digital cytometry, *Nat. Biotechnol.* 37 (7) (2019) 773–782.
- [80] E.C. Stack, C. Wang, K.A. Roman, C.C. Hoyt, Multiplexed immunohistochemistry, imaging, and quantitation: a review, with an assessment of Tyramide signal amplification, multispectral imaging and multiplex analysis, *Methods* 70 (2014) 46–58.
- [81] J.-R. Lin, M. Fallahi-Sichani, J.-Y. Chen, P.K. Sorger, Cyclic immunofluorescence (CycIF), a highly multiplexed method for single-cell imaging, *Current Protocols in Chemical Biology* 8 (2016) 251–264.
- [82] K.M. McKinnon, Flow cytometry: an overview, *Curr. Protoc. Immunol.* 120 (2018) 5.1.1–5.1.11.
- [83] D.R. Bandura, V.I. Baranov, O.I. Ornatsky, A. Antonov, R. Kinach, X. Lou, S. Pavlov, S. Vorobiev, J.E. Dick, S.D. Tanner, Mass cytometry: technique for real time single cell multitarget immunoassay based on inductively coupled plasma time-of-flight mass spectrometry, *Anal. Chem.* 81 (2009) 6813–6822.
- [84] A. White, N. Elliott, Y. Liang, S.E. Warren, J.M. Beechem, Spatially resolved, multiplexed digital characterization of protein distribution and abundance in formalin-fixed, paraffin-embedded (FFPE) diffuse large B cell lymphoma tissue sections based on the Nanostring® digital spatial profiling (DSP) technology, *Blood* 130 (2017) 1196.
- [85] V.M. Peterson, K.X. Zhang, N. Kumar, J. Wong, L. Li, D.C. Wilson, R. Moore, T.K. McClanahan, S. Sadekova, J.A. Klappenbach, Multiplexed quantification of proteins and transcripts in single cells, *Nat. Biotechnol.* 35 (2017) 936.
- [86] P. Shahi, S.C. Kim, J.R. Haliburton, Z.J. Gartner, A.R. Abate, ASeq: ultrahigh-throughput single cell protein profiling with droplet microfluidic barcoding, *Sci. Rep.* 7 (2017) 44447.
- [87] M. Stoeckius, C. Hafemeister, W. Stephenson, B. Houck-Loomis, P.K. Chattopadhyay, H. Swerdlow, R. Satija, P. Smibert, Simultaneous epitope and transcriptome measurement in single cells, *Nat. Methods* 14 (2017) 865.
- [88] G. Glass, J.A. Papin, J.W. Mandell, Simple: a sequential immunoperoxidase labeling and erasing method, *Journal of Histochemistry & Cytochemistry* 57 (2009) 899–905.
- [89] T. Tsujikawa, S. Kumar, R.N. Borkar, V. Azimi, G. Thibault, Y.H. Chang, A. Balter, R. Kawashima, G. Choe, D. Sauer, E. El Rassi, D.R. Clayburgh, M.F. Kulesz-Martin,

- E.R. Lutz, L. Zheng, E.M. Jaffee, P. Leyschok, A.A. Margolin, M. Mori, J.W. Gray, P.W. Flint, L.M. Coussens, Quantitative multiplex immunohistochemistry reveals myeloid-inflamed tumor-immune complexity associated with poor prognosis, *Cell Rep.* 19 (2017) 203–217.
- [90] M.J. Gerdes, C.J. Sevinsky, A. Sood, S. Adak, M.O. Bello, A. Bordwell, A. Can, A. Corwin, S. Dinn, R.J. Filkins, D. Hollman, V. Kamath, S. Kaanumalle, K. Kenny, M. Larsen, M. Lazare, Q. Li, C. Lowes, C.C. McCulloch, E. McDonough, M.C. Montalto, Z. Pang, J. Rittscher, A. Santamaria-Pang, B.D. Sarachan, M.L. Seel, A. Seppo, K. Shaikh, Y. Sui, J. Zhang, F. Ginty, Highly multiplexed single-cell analysis of formalin-fixed, paraffin-embedded cancer tissue, *Proc. Natl. Acad. Sci.* 110 (2013) 11982.
- [91] L.R. Olsen, M.D. Leipold, C.B. Pedersen, H.T. Maecker, The anatomy of single cell mass cytometry data, *Cytometry Part A* 95 (2019) 156–172.
- [92] J. Wagner, M.A. Rapsomaniki, S. Chevrier, T. Anzeneder, C. Langwieder, A. Dykgers, M. Rees, A. Ramaswamy, S. Muenst, S.D. Soysal, A. Jacobs, J. Windhager, K. Silina, M. van den Broek, K.J. Dedes, M. Rodríguez Martínez, W.P. Weber, B. Bodenmiller, A single-cell atlas of the tumor and immune ecosystem of human breast cancer, *Cell*, 177 e1318 (2019) 1330–1345.
- [93] D. Schulz, V.R.T. Zanutelli, J.R. Fischer, D. Schapiro, S. Engler, X.-K. Lun, H.W. Jackson, B. Bodenmiller, Simultaneous multiplexed imaging of mRNA and proteins with subcellular resolution in breast cancer tissue samples by mass cytometry, *Cell Systems*, 6 e25 (2018) 25–36.
- [94] L. Keren, M. Bosse, D. Marquez, R. Angoshtari, S. Jain, S. Varma, S.-R. Yang, A. Kurian, D. Van Valen, R. West, S.C. Bendall, M. Angelo, A structured tumor-immune microenvironment in triple negative breast cancer revealed by multiplexed ion beam imaging, *Cell*, 174 e1319 (2018) 1373–1387.
- [95] D.E. Carvajal-Hausdorf, J. Patsenker, K.P. Stanton, F. Villarroel-Espindola, A. Esch, R.R. Montgomery, A. Psyrrri, K.T. Kalogeras, V. Kotoula, G. Foutzilas, K.A. Schalper, Y. Kluger, D.L. Rimm, Multiplexed (18-Plex) measurement of signaling targets and cytotoxic T cells in trastuzumab-treated patients using imaging mass cytometry, *Clin. Cancer Res.* 25 (10) (2019) 3054–3062.
- [96] R. Moncada, F. Wagner, M. Chiodin, J.C. Devlin, M. Baron, C.H. Hajdu, D.M. Simeone, I. Yanai, Integrating single-cell RNA-Seq with spatial transcriptomics in pancreatic ductal adenocarcinoma using multimodal intersection analysis, *bioRxiv*, (2019) 254375.
- [97] J. Decalf, M.L. Albert, J. Ziai, New tools for pathology: a user's review of a highly multiplexed method for in situ analysis of protein and RNA expression in tissue, *J. Pathol.* 247 (2019) 650–661.
- [98] P.L. Ståhl, F. Salmén, S. Vickovic, A. Lundmark, J.F. Navarro, J. Magnusson, S. Giacomello, M. Asp, J.O. Westholm, M. Huss, A. Mollbrink, S. Linnarsson, S. Codeluppi, Å. Borg, F. Pontén, P.I. Costea, P. Sahlén, J. Mulder, O. Bergmann, J. Lundeberg, J. Frisén, Visualization and analysis of gene expression in tissue sections by spatial transcriptomics, *Science* 353 (2016) 78.
- [99] Y. Simoni, M.H.Y. Chng, S. Li, M. Fehlings, E.W. Newell, Mass cytometry: a powerful tool for dissecting the immune landscape, *Curr. Opin. Immunol.* 51 (2018) 187–196.
- [100] E.Y. Shum, E.M. Walczak, C. Chang, H.C. Fan, Quantitation of mRNA transcripts and proteins using the BD Rhapsody™ single-cell analysis system, Springer, Place Published, Single Molecule and Single Cell Sequencing, 2019, pp. 63–79.
- [101] Y. Hu, Q. An, K. Sheu, B. Trejo, S. Fan, Y. Guo, Single cell multi-omics technology: methodology and application, *Frontiers in Cell and Developmental Biology* 6 (2018) 28.