PORTLAND PRESS

Review Article

Unlocking immune-mediated disease mechanisms with transcriptomics

Emma de Jong and Anthony Bosco

Telethon Kids Institute, Centre for Health Research, The University of Western Australia, Nedlands, Western Australia, Australia

Correspondence: Emma de Jong (emma.dejong@telethonkids.org.au)

OPEN ACCESS

The transcriptome represents the entire set of RNA transcripts expressed in a cell, reflecting both the underlying genetic and epigenetic landscape and environmental influences, providing a comprehensive view of functional cellular states at any given time. Recent technological advances now enable the study of the transcriptome at the resolution of individual cells, providing exciting opportunities to characterise cellular and molecular events that underpin immune-medicated diseases. Here, we draw on recent examples from the literature to highlight the application of advanced bioinformatics tools to extract mechanistic insight and disease biology from bulk and single-cell transcriptomic profiles. Key considerations for the use of available analysis techniques are presented throughout.

Introduction

Often, new discoveries are facilitated through technological advances. Some prominent examples include improvements in microscope technology that enabled the first observations of white blood cells and platelets [1], and the development of monoclonal antibodies for phenotypic characterisation of cells and the development of therapeutics [2]. Finally, the advent of next-generation sequencing revolutionised the study of the transcriptome, B-cell and T-cell receptor repertoires [3], and is arguably one of the most exciting technological advances in recent history, as reflected in the exponential rise in publications citing RNA-sequencing (RNA-Seq) over the past decade. While the transcriptome represents just one layer of biology that ultimately determines the form and function of a cell, its dynamic nature reflects both the effects of environmental influences as well as the underlying genetic and epigenetic landscape. Thus, the transcriptome is a middle-ground rich in information that can reveal important insights into disease biology. The advent of new technologies that perform RNA-Seq at single-cell resolution [4] enables the profiling of transcriptional states that underpin cellular and phenotypic states with unprecedented clarity. In this regard, transcriptomics lends itself particularly well to immunology; a field dedicated to understanding the behaviour of and functional interactions between diverse cell types. While transcriptomics provides exciting opportunities to unlock disease mechanisms, it also brings significant challenges, particularly around analysing and interpreting the vast volume of data generated employing an ever-growing array of analysis tools. Here, we present an overview of bioinformatics tools that can yield novel and mechanistic insights into the biology of immune-mediated diseases.

Gene expression signatures reveal immunophenotypes, biomarkers, and avenues for therapeutic intervention

A common application of transcriptomics is to compare gene expression profiles between two or more groups of samples, i.e. between resting versus stimulated immune cells, in samples collected from subjects during acute illness versus convalescence, or in subjects with or without clinical manifestations of the disease. Asthma is a pertinent example of how this approach can inform on disease mechanisms and motivate the development of novel treatment strategies. Asthma is a remarkably

heterogenous disease, arising through complex gene–environment interactions, yielding multiple phenotypes. These phenotypes often reflect differential activation of discrete inflammatory pathways, the prototypical example being the identification of Th2-high and Th2-low forms of atopic asthma through assessment of IL-13 response genes in airway brushings [5]. Importantly, Th2-low asthmatics were unresponsive to inhaled cortico-steroid therapy, linking molecular immunophenotypes with clinically relevant features. Since the seminal paper by Woodruff et al., transcriptomic profiling has been performed on various tissues including sputum [6,7], blood [8,9], bronchoalveolar lavage [10], bronchial biopsies [11], and epithelial brushings [12–14] to reveal molecular immunophenotypes beyond Th2-high and Th2-low subgroups, particularly in severe asthmatics; a subgroup of patients with the poorly controlled disease for which new and effective therapies are desperately needed. These studies typically employ unsupervised analysis approaches, where samples are clustered into sub-groups based on their gene expression profiles without considering any characteristics of the samples (Figure 1A). When performing unsupervised analyses, care must be taken to ensure that the analysis is driven by signatures that capture disease biology and is not confounded by unrelated sample characteristics (e.g. age, gender, ethnicity, medication use, variations in cellular composition). Jevnikar et al. [12] recently identified a novel subset of asthma patients characterised by high expression of IL-6 trans-signalling pathway genes within bronchial epithelium through unsupervised clustering. These patients were characterised by frequent exacerbations and exhibited reduced expression of epithelial barrier function genes alongside increased airways inflammation. This finding has important implications for clinical trials aiming to target the IL-6 pathway in exacerbation-prone asthmatics. Indeed, the discovery of distinct molecular phenotypes in asthma has motivated a shift in clinical trials toward endotype-based strategies (reviewed [15]). The PrecISE study (Precision Interventions for Severe and/or Exacerbation-prone Asthma) [16] is one such example, where six different treatments are now being tested across severe asthma phenotypes defined on the basis of variations in blood eosinophils, IL-6 levels in plasma, fractional exhaled nitric oxide, and genotypes in an adaptive Phase 2 clinical trial. These treatments target a variety of inflammatory mechanisms and include a tyrosine kinase inhibitor, anti-IL-6, a Janus kinase 1-selective inhibitor, medium-chain triglycerides, a modulator of the cystic fibrosis transmembrane conductance regulator (CFTR) protein, and Broncho-Vaxom; an endotoxin-low lyophilised extract from eight major respiratory bacterial pathogens. Embedding transcriptomics into adaptive clinical trial designs will enable the matching of disease endotypes with response signatures for each therapeutic agent, which will accelerate precision medicine approaches.

In addition to unveiling distinct immunophenotypes ripe for therapeutic intervention, transcriptomics is a powerful tool for biomarker discovery. In contrast with the unsupervised analyses used to detect novel immuno-phenotypes, biomarker discovery focuses on supervised analysis techniques to perform group-wise comparisons (Figure 1A). A clinical setting in which transcriptomic biomarkers have important diagnostic, prognostic and therapeutic utility is sepsis, a notoriously complex and life-threatening disease. In this regard, Sweeney et al. [17] leveraged publicly available gene expression data derived from >660 whole blood/buffy coat samples to generate an 11-gene signature that could distinguish sepsis from non-infectious systemic inflammation; mean area under the receiver operating characteristic curve (AUC) of 0.87 (range 0.7–0.98). Accurate identification of infectious from sterile systemic inflammation has critical implications for guiding appropriate therapy, avoiding the unnecessary use of antibiotics. In addition, a recent study utilised transcriptomic data across 21 sepsis cohorts to develop four predictive models of 30-day mortality with AUCs between 0.765 and 0.89, each using <20 genes [18]. Each model was validated using multiple external cohorts across both community- and hospital-acquired infections, a critical step for evaluating the performance of any biomarker. Furthermore, when combined with clinical severity scores, each of the prognostic models significantly outperformed those based on clinical features alone, highlighting the prognostic value of gene expression-based features to improve risk stratification for sepsis patients. Collectively, these studies demonstrate the value in assessing levels of gene expression through both supervised and unsupervised approaches. However, evaluation of the transcriptome on a gene-by-gene basis is limited because genes do not exist or function in isolation, they function as components of interconnected systems [19].

Mapping gene co-expression networks to unveil unique immunobiological insights

Network analysis provides a holistic framework to elucidate the topology and functional organisation of gene expression programs. Notably, changes in network structure can occur in the absence of changes in mean expression levels. Therefore, network analysis methods capture covert disease-associated signatures that remain
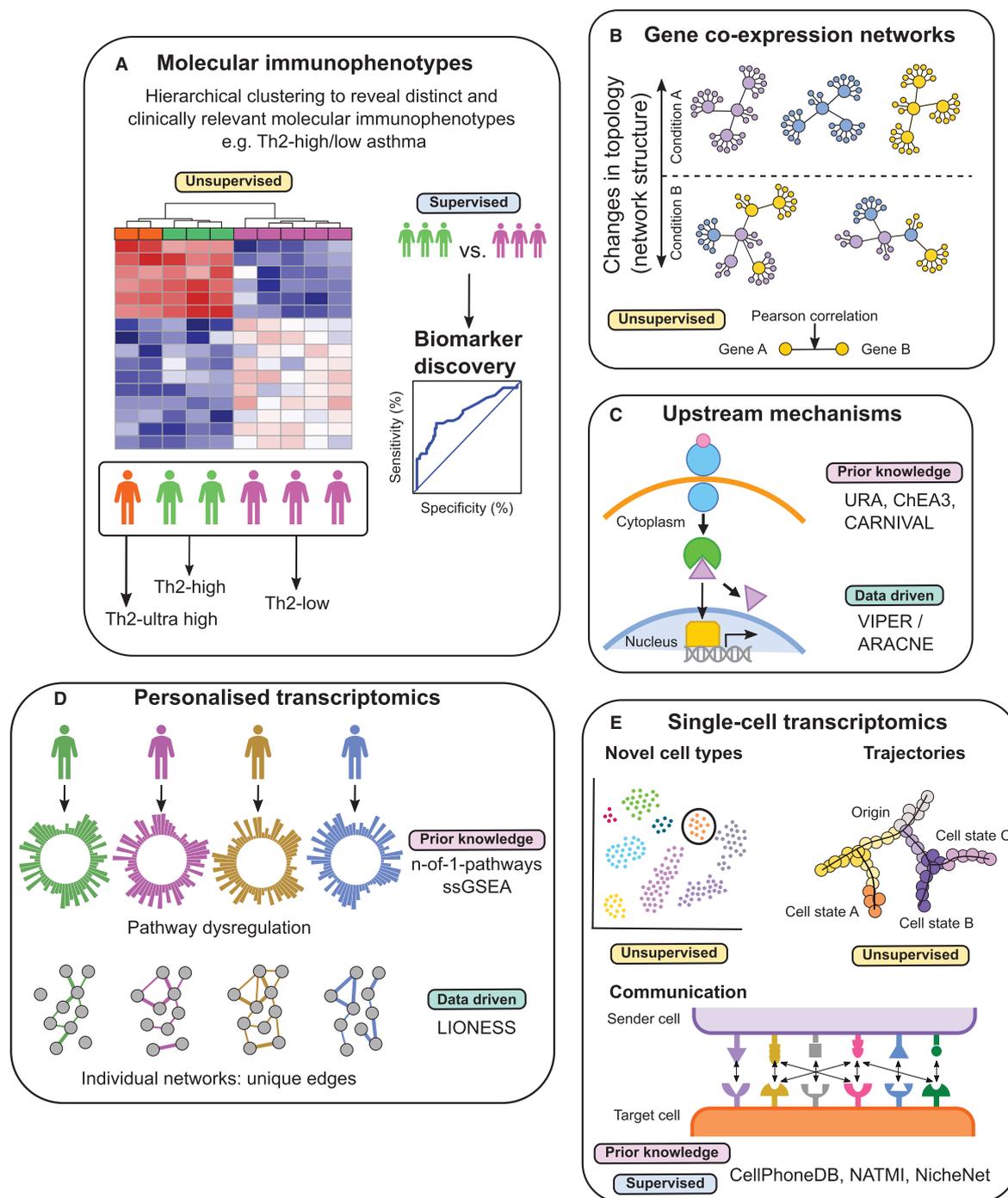
Figure 1. An overview of applications for transcriptomics to better understand disease biology and extract mechanistic insight.

Part 1 of 2

(**A**) Unsupervised hierarchical clustering can reveal novel molecular phenotypes which may inform endotype-based therapeutics. Gene expression signatures can also be used to derive biomarkers through supervised analyses. (**B**) Unsupervised gene co-expression networks can capture disease-associated changes in the network topology that may remain undetected through evaluation of gene expression levels alone. The use of both prior knowledge-based and data-driven bioinformatic tools can provide mechanistic insights into observed transcriptional changes (**C**), and offer a personalised view of the transcriptome in terms of both specific biological pathways, and network topology (**D**). (**E**) Single-cell RNA-sequencing provides opportunities for discovery through the identification of novel cell subsets, analysis of transitional states, and the study of cell–cell communication through ligand–receptor signalling. URA, Upstream Regulator Analysis; ChEA3, ChIP-X

Figure 1. An overview of applications for transcriptomics to better understand disease biology and extract mechanistic insight.

Part 2 of 2

Enrichment Analysis 3; CARNIVAL, CAusal Reasoning pipeline for Network identification using Integer VALue programming; VIPER, Virtual Inference of Protein-activity by Enriched Regulon analysis; ARACNE, Algorithm for the Reconstruction of Accurate Cellular Networks; ssGSEA, single-sample Gene Set Enrichment Analysis; LIONESS, Linear Interpolation to Obtain Network Estimates for Single Samples.

undetected by other analysis methods [20,21]. Network theory lends itself to transcriptomics, where gene products (mRNA transcripts) represent network nodes and relationships between genes represent edges (or connections) between nodes (Figure 1B). These relationships can be drawn from prior knowledge or derived from gene co-expression patterns across samples. This latter approach is used by one of the most widely used tools in this space: weighted gene correlation network analysis (WGCNA) [22]. Briefly, WGCNA is an unsupervised algorithm that constructs a network-based on Pearson correlations between gene pairs across the data. The correlation matrix is raised to a power to achieve scale-free topology, i.e. the correlation structure of the network follows a power-law degree distribution where a subset of genes known as 'hubs' dominate the connectivity patterns. Finally, functionally enriched modules of highly correlated genes are identified through hierarchical clustering. By focusing on the correlation structure of the data, WGCNA can unveil mechanisms that are not detected through differential expression analysis. Jones et al. [23] demonstrate this principle in a recent longitudinal study of CD4+ T-cell responses to house dust mite (HDM) during HDM-specific immunotherapy in allergic rhinitis patients. The authors reported that no differentially expressed genes were identified with respect to treatment effect during immunotherapy. Yet through network analysis, they demonstrated that initially topologically isolated type 1 interferon, Th2 and IL-2/Treg networks were rewired over time to form a single highly interconnected network by the end of immunotherapy, concomitant with symptom stabilisation relative to baseline. Drawing further examples from asthma, Peters et al. recently applied WGCNA to sputum transcriptomic data to better understand the contributions of distinct immune cell types to type-2 inflammation (T2) high and T2-low asthma phenotypes [24]. The authors identified a network of 92 genes enriched for gene expression signatures derived from eosinophils, mast cells/basophils and monocyte-derived dendritic cells, which correlated with both asthma status and lung function. Based on this 92-gene network, hierarchical clustering of study participants revealed separation between T2-high and T2-low asthmatics, but also identified a new third 'ultra-high' T2 phenotype. Furthermore, a network module enriched for cytotoxic CD8+ T-cells was significantly under-expressed in the T2-low asthmatics. While WGCNA provided the foundation for the analysis performed by Peters et al., the immunological insights gained relied on delineating the contributions of each cell type to the gene expression networks. The authors generated their own cell type-specific gene signatures through analysis of publicly available data from purified cell populations; however, it is worth noting that tools such as CIBERSORTx [25], xCell [26], and EPIC [27] are available as a form of 'digital cytometry' to infer the contributions of leukocyte populations to gene expression profiles, albeit predominantly designed for analysis of malignant tissues. An important consideration for WGCNA is the sample size required to generate robust results; at least 30 samples are required, however, larger sample sizes will lead to more informative results. Furthermore, while WGCNA provides a global view of gene co-expression networks, additional analyses are required to identify the causal pathways that drive gene network patterns.

Identifying molecular drivers of transcriptome regulation

Cellular states are determined by complex and co-ordinated signalling cascades that are tightly regulated under homeostatic conditions, and that adapt during perturbations. While differential gene expression and co-expression network analyses can identify the molecular changes that underpin phenotypic states, they cannot predict what is driving these changes. Several tools are available in this regard, some leveraging prior knowledge to generate causal networks, while others use data-driven approaches to infer regulatory mechanisms (Figure 1C). Upstream Regulator Analysis is one of the most comprehensive prior knowledge-based tools for predicting upstream molecular drivers of downstream expression patterns [28]. Here, a variety of molecule types (e.g. ligands, receptors, kinases, transcription factors (TFs), microRNAs, drug compounds) are considered candidate upstream transcriptional regulators based on experimental evidence, and through comparing their observed influence on gene expression with a user-defined list of input genes and their associated log-fold

changes, an overlap *p*-value is calculated for each regulator alongside an activation *Z*-score, which considers the direction of change in expression (up/down-regulation) to predict whether a regulator is activated or inhibited. Zemek et al. [29] employed this approach on a gene signature associated with positive responses to immune checkpoint blockade therapy in murine mesothelioma prior to the onset of treatment, which was predicted to be driven by activation of STAT1 and inhibition of IL-10. By administering the STAT1 activator IFNγ and blocking IL-10 to mimic this molecular phenotype, tumours were sensitised to immune checkpoint blockade leading to improved cure rates. Other prior knowledge-based tools include ChEA3 (ChIP-X Enrichment Analysis 3) [30] which ranks TFs based on pooled information across six TF-target gene set libraries to identify those responsible for observed changes in gene expression, and CARNIVAL (CAusal Reasoning pipeline for Network identification using Integer VALue programming) which takes an alternative approach by integrating both TF and pathway activity scores to infer the entire signalling cascade [31]. While prior knowledge-based tools can provide valuable mechanistic insight, they rely on data generated from diverse experimental systems, and a limitation of this approach is that the data lack context-specificity and are biased towards the identification of well-characterised pathways. Data-driven approaches whilst less accurate, are better suited to identify novel, context-dependant mechanisms, and in this domain, VIPER (Virtual Inference of Protein-activity by Enriched Regulon analysis) [32] is one of the best performing tools. VIPER analysis entails the construction of networks between TF and target genes using ARACNE (Algorithm for the Reconstruction of Accurate Cellular Networks) [33], and then for each TF, a regulon is defined comprising the complete set of predicted downstream target genes for that TF. The regulons are then scored for differential enrichment between experimental conditions or on a single-sample basis using gene set analysis methods, to infer the biological activity of each TF. A limitation of VIPER is that hundreds (or thousands) of samples are required for accurate network construction using ARACNE, however, in many applications, it is possible to use large data sets available in the public domain derived from the same tissue type to construct a tissue-specific network suitable for VIPER.

Towards personalised immunology through single-sample transcriptomics

The analyses considered above focus on group-wise comparisons, and in general, these methods are not well suited to elucidate disease mechanisms that are operating in small numbers of patients. Where inter-sample heterogeneity is viewed as a limitation in group-wise analyses, it becomes the focus of single-subject analyses (Figure 1D). As we continue to transition toward personalised medicine (the understanding of disease and the tailoring of treatments at an individual level), the insights gained through single-subject analyses will be invaluable. In the context of transcriptomics, several tools have emerged for this purpose including *n*-of-1-pathways [34], ssGSEA (single-sample Gene Set Enrichment Analysis) [35], and LIONESS (Linear Interpolation to Obtain Network Estimates for Single Samples) [36,37]. The *n*-of-1-pathways algorithm [34] requires a pair of samples from a single subject and generates a personal profile of dysregulated pathways. Jones et al. demonstrate the utility of *n*-of-1-pathways to study paediatric viral bronchiolitis, where transcriptomic profiles of PBMC collected during an acute episode as well as after convalescence were analysed on a single-subject basis, followed by unsupervised clustering to reveal two clear subgroups of study participants [38]. These groups exhibited striking variations in the expression of innate immune pathways (interferon, inflammation, toll-like receptors) effectively yielding a 'low-response' group which importantly, was not separable based on available clinical metrics including steroid use, pathogen type or age, highlighting the power of *n*-of-1-pathways to reveal novel insights. In a similar vein, ssGSEA [35] calculates an enrichment score for each sample gene set pair, reflecting the level of co-ordinated up- or down-regulation of that gene set (provided by the user). Kobayashi et al. demonstrate how ssGSEA can be used to derive 'immunogram' scores for solid tumours based on 10 gene sets spanning pathways integral to immune function and cancer biology, to deliver a personalised view of cancer-immunity status and identify potential avenues for therapeutic intervention [39]. Finally, in contrast with *n*-of-1-pathways and ssGSEA which interrogate predefined biological pathways, the recently developed tool LIONESS [36,37] provides an opportunity to infer single-sample networks from experimental data. To achieve this, LIONESS first constructs an aggregate network across the entire dataset, and then systematically re-builds the network *n* times, leaving out a different sample upon each re-build. By subtracting each $n − 1$ network from the aggregate network, LIONESS estimates the relative influence of an individual sample, yielding sample-specific network edge weights. Using transcriptional profiles from 153 lymphoblastic cell lines, the authors illustrate the power of single-sample networks to find biologically meaningful data structures not

otherwise identifiable based on aggregate networks or evaluation of gene expression levels. Specifically, the data separated into two distinct groups based on clustering of individual network edge weights, which were strikingly different from the clusters that formed based on gene expression levels. Comparisons of the most highly targeted genes (those with the highest sum of edge weights) in each network-based cluster revealed enrichment for ∼50 biological pathways, whereas no pathways were enriched in the differentially expressed genes between expression-based clusters. A general limitation of personalised analyses is the potential for incidental findings, and naturally, there is a higher degree of uncertainty when the statistical universe is a single subject.

Single-cell transcriptomics: a revolutionary tool for immunology

Single-cell RNA-sequencing (scRNA-seq) is a major technological advance that overcomes many of the limitations of 'bulk' profiling, providing not only phenotypic, but functionally relevant information on individual cells, enabling researchers to study the underlying heterogeneity that exists within and between cell populations with unprecedented clarity. The Human Cell Atlas [40,41] is an international collaboration aiming to create maps of all human cells, largely through single-cell transcriptomics. The availability of single-cell atlases across biological and clinical states will enable the identification of novel cell subsets, shed light on ontologies and differentiation trajectories, and elucidate cell-to-cell interactions that govern tissue and organ function (Figure 1E). In a landmark paper, Villani et al. [42] present an unbiased re-classification system for human dendritic cell and monocyte subsets based on the transcriptomic profiles from ∼2400 individual HLA-DR+ cells from healthy adult donors. This work has major implications for the field, establishing a new taxonomy to capture the previously unknown heterogeneity of dendritic cells comprising six distinct subsets, and monocytes comprising four distinct subsets. In a disease-relevant setting, scRNA-seq of the human bronchial epithelium has recently led to the discovery of the pulmonary ionocyte; a rare cell type with high CFTR gene expression [43]. The authors demonstrate that these cells were responsible for the majority of CFTR activity in airway epithelium, are dependent on Notch signalling and express high levels of the transcription factor FOXI1; a master regulator of the proton pump V-ATPase suggesting a role for the ionocyte in the regulation of luminal pH, and a potential target for therapeutic intervention in cystic fibrosis [44]. Furthermore, the authors identified two distinct differentiation trajectories of conducting airway cells in an animal model of airway injury and repair using a method called SPRING [45], which generates k-nearest neighbour graphs to visualise the continuum of cell states based on their transcriptional profiles. A similar tool in this space is Monocle [46], an unsupervised algorithm that uses pseudotemporal ordering of single cells to capture transitional cellular states such as differentiation. Kwok et al. [47] recently employed Monocle to delineate the fate of granulocyte-monocyte progenitors, and combined this with flow cytometry to identify novel early committed (proNeu1) and intermediate (proNeu2) neutrophil progenitors. In a sepsis model, the authors demonstrate the preferential and specific expansion of proNeu1 cells during early systemic inflammation at the expense of monocyte differentiation, yielding new insights into the pathogenesis of sepsis. Single-cell genomics has also unveiled new insights into the complexity and heterogeneity of autoimmune diseases such as systemic lupus erythematosus (SLE), lupus nephritis and rheumatoid arthritis. Nehar-Belaid et al. [48] profiled ∼276 000 PBMC from children with SLE and matched healthy controls, and through sub-clustering of leukocytes based on their molecular profiles identified the expansion of monocyte, dendritic cell (conventional and plasmacytoid), CD4+ and CD8+ T-cell, NK cell, B-cell and plasma cell sub-clusters in SLE patients that specifically expressed high levels of interferon-stimulated genes. While an elevated blood type I interferon gene signature is a common feature of SLE [49], through scRNA-seq Nehar-Belaid et al. were able to pinpoint the precise cellular compartments altered in this regard, reporting novel leukocyte variations associated with SLE that were most extreme with increased disease severity. Similarly, through scRNA-seq of renal biopsies from patients with lupus nephritis, Der et al. [50] observed that high expression of a type I IFN signature alongside fibrotic pathway genes in tubular cells was significantly associated with non-response to treatment. The authors note the potential clinical utility of the fibrotic gene signature as an early prognostic marker; logistic regression using four of these genes could predict response to treatment 6 months post-biopsy (accuracy 92%, AUC 0.96) even though tuberinterstitial damage (linked to poor prognosis) was undetected via histological scoring for some of the treatment-refractory patients at the time of sampling. In the context of rheumatoid arthritis, Zhang and colleagues integrated scRNA-seq data with bulk RNA-seq and mass/flow cytometry to comprehensively map the functional cellular landscape of synovial tissue and delineate populations driving joint inflammation [51]. They identified the specific expansion of pro-inflammatory monocytes producing IL1B,

sublining fibroblasts (THY1[+], HLA-DR[hi]) producing IL6, peripheral and follicular helper T-cells (PD-1[+]) and autoimmune-associated B-cells (ITGAX[+]TBX21[+]) that also expressed interferon-stimulated genes within arthritic synovia. Interestingly, while the up-regulation of some inflammatory mediators was restricted to discrete subpopulations (e.g. IL1B and IL6), others including TNF were globally increased within synovial cell types, providing a deeper understanding of the relative contributions of different cell populations to joint inflammation.

Finally, the advent of scRNA-seq provides new opportunities to understand cellular interactions through cell-to-cell communication networks that are mediated by ligand–receptor signalling, which has become a field of study in its own right (reviewed [52]). Understanding interactions between cells of the adaptive and innate immune system, and interactions between immune and non-immune cells within diseased organs, malignant cells and even transplanted tissues plays a critical role in disease progression, resolution, and chronicity. Here, we highlight three methods for studying cell–cell communication: CellPhoneDB [53,54], NATMI (Network Analysis Toolkit for Multicellular Interactions) [55], and NicheNet [56]. All three are built on known ligand–receptor interactions but are each unique in their approach. CellPhoneDB compares the expression of ligands in one cell type with expression of receptors in another, to generate a ranked list of active ligand–receptor pairs. Uniquely, CellPhoneDB considers the protein architecture of ligand–receptor pairs, recognising that they often exert functions as heteromeric complexes. The application of CellPhoneDB to bronchial biopsies from healthy and asthmatic individuals demonstrates unique airway cell–cell communication networks in asthma, dominated by interactions between Th2 and epithelial or mesenchymal cell populations that are absent in healthy lung tissues, including some novel epithelial-Th2 ligand–receptor pairs (*CXCL2-DPP4*, *CXCL17-GPR35*) [57]. These findings extend previous observations from bulk RNA-Seq of highly interlinked epithelial-Th2 gene networks specific to airway mucosa of atopic asthmatics [58]. NATMI also compares the expression of ligands and receptors between cell clusters, but its advantage lies in the use of connectomeDB2020 (currently the largest literature curated database for ligand–receptor pairs containing 2293 interactions) and its ability to generate and visualise global cell–cell communication networks through systematic assessment of ligand and receptor expression in every cell cluster. NATMI also allows for direct comparison of two distinct cell–cell communication networks to identify condition-specific interactions. Whilst NATMI provides a global view of intercellular communication networks, NicheNet leverages a prior network model to link cell-to-cell communication though specific signalling paths (Ligand–Receptor-TF-Target gene) to predict responses in target cells. Across all three methods, accurate cell type annotations are critical for delineating cell–cell interactions and generating valid novel hypotheses. Collectively, these studies demonstrate the power of scRNA-seq to elucidate the cellular and molecular events that underpin inflammation. However, some important limitations of the technology should be considered. Firstly, sample processing protocols required for tissue dissociation can introduce artefacts into the data (e.g. activation of stress response genes) [59]. Secondly, the data are sparse and zero-inflated, resulting in many false negatives which complicate downstream statistical and computational analyses. Even so, as the cost of scRNA-seq declines and analysis techniques evolve, this technology will cement itself as an integral tool for surveying the functional immune repertoire, and the intersection of cellular and molecular mechanisms that govern immunopathologies.

Conclusion

Remarkable advances in genomic profiling technologies and bioinformatics have revolutionised the study of gene expression, which in the context of immunology has led to a deeper understanding of disease biology, disease heterogeneity, and the identification of therapeutic targets. Furthermore, the adaptation of RNA-Seq to individual cells provides many opportunities for discovery, as the unbiased analysis of cell populations can reveal novel functional subsets and transitional states. A diverse range of tools are available to interrogate gene expression data, with advantages and limitations that should be carefully considered. In this regard, there are several challenges when applying transcriptomics to the study of immune-mediated diseases. First, projects should be based on well-designed questions, with careful consideration of the specimen type (e.g. blood, biopsy, swabs), immunological context (e.g. baseline versus challenge; inactive/active disease), and the collection of comprehensive sample metadata (e.g. demographic and clinical variables, medication use, technical variables). Second, samples should be randomised and processed in large batches to minimise the potential for batch effects to confound analysis. Third, a data analysis plan should be developed prior to collecting the samples, to ensure that the sample size is large enough to address the aims of the study, considering disease heterogeneity and patient subgroups. Fourth, mechanistic insights derived from computational analyses represent predictions that require experimental validation. Finally, while transcriptomics alone is a powerful tool to

uncover biological insight, technologies for studying additional layers of biology in the same sample are maturing such as single-sample multi-OMICS, which together have the potential to transform our understanding of immune-mediated diseases.

Perspectives

- The transcriptome is a dynamic layer of biological regulation, the study of which has significantly advanced our understanding of immunological processes and immune-mediated diseases. Furthermore, transcriptomics can generate clinically relevant features in the form of disease biomarkers and endotypes that may be stratified for targeted therapeutics.

- Analyses that move beyond differential expression to elucidate gene network patterns and driver genes can yield a unique level of insight, alongside personalised approaches that support single-subject analyses.

- scRNA-seq will replace bulk profiling and will challenge current paradigms pertaining to the basic mechanisms and principles that govern the functionality of the immune system in health and disease. Furthermore, as technologies for studying additional layers of biology in the same sample mature (e.g. epigenome, proteome, metabolome), this will enable a much deeper understanding of immune regulation.

Competing Interests

The authors declare that there are no competing interests associated with the manuscript.

Author Contributions

Conceptualisation and drafting of the manuscript: E.d.J. and A.B.; Critical revision of manuscript: E.d.J. and A.B.

Open Access

Open access for this article was enabled by the participation of the University of Western Australia in an all-inclusive *Read & Publish* pilot with Portland Press and the Biochemical Society under a transformative agreement with CAUL.

Abbreviations

LIONESS, Linear Interpolation to Obtain Network Estimates for Single Samples; scRNA-seq, single-cell RNA-sequencing; SLE, systemic lupus erythematosus; TFs, transcription factors; VIPER, Virtual Inference of Protein-activity by Enriched Regulon analysis; WGCNA, weighted gene correlation network analysis.

References

1 Hajdu, S.I. (2003) A Note from History: The Discovery of Blood Cells. *Ann. Clin. Lab. Sci.* **33**, 237–238 PMID:12817630
2 Shepard, H.M., Phillips, G.L., Thanos C, D. and Feldmann, M. (2017) Developments in therapy with monoclonal antibodies and related proteins. *Clin. Med.* **17**, 220–232 https://doi.org/10.7861/clinmedicine.17-3-220
3 Liu, X. and Wu, J. (2018) History, applications, and challenges of immune repertoire research. *Cell Biol. Toxicol.* **34**, 441–457 https://doi.org/10.1007/s10565-018-9426-0
4 Picelli, S. (2017) Single-cell RNA-sequencing: the future of genome biology is now. *RNA Biol.* **14**, 637–650 https://doi.org/10.1080/15476286.2016.1201618
5 Woodruff, P.G., Modrek, B., Choy, D.F., Jia, G., Abbas, A.R., Ellwanger, A. et al. (2009) T-helper type 2–driven inflammation defines major subphenotypes of asthma. *Am. J. Respir. Crit. Care Med.* **180**, 388–395 https://doi.org/10.1164/rccm.200903-0392OC
6 Kuo, C.-H.S., Pavlidis, S., Loza, M., Baribaud, F., Rowe, A., Pandis, I. et al. (2017) T-helper cell type 2 (Th2) and non-Th2 molecular phenotypes of asthma using sputum transcriptomics in U-BIOPRED. *Eur. Respir. J.* **49**, 1602135 https://doi.org/10.1183/13993003.02135-2016
7 Rossios, C., Pavlidis, S., Hoda, U., Kuo, C.-H., Wiegman, C., Russell, K. et al. (2018) Sputum transcriptomics reveal upregulation of IL-1 receptor family members in patients with severe asthma. *J. Allergy Clin. Immunol.* **141**, 560–570 https://doi.org/10.1016/j.jaci.2017.02.045
8 Yeh, Y.L., Su, M.W., Chiang, B.L., Yang, Y.H., Tsai, C.H. and Lee, Y.L. (2018) Genetic profiles of transcriptomic clusters of childhood asthma determine specific severe subtype. *Clin. Exp. Allergy* **48**, 1164–1172 https://doi.org/10.1111/cea.13175

9   Bigler, J., Boedigheimer, M., Schofield, J.P.R., Skipp, P.J., Corfield, J., Rowe, A. et al. (2017) A severe asthma disease signature from gene expression profiling of peripheral blood from U-BIOPRED cohorts. *Am. J. Respir. Crit. Care Med.* **195**, 1311–1320 https://doi.org/10.1164/rccm.201604-0866OC

10   Weathington, N., O'Brien, M.E., Radder, J., Whisenant, T.C., Bleecker, E.R., Busse, W.W. et al. (2019) BAL cell gene expression in severe asthma reveals mechanisms of severe disease and influences of medications. *Am. J. Respir. Crit. Care Med.* **200**, 837–856 https://doi.org/10.1164/rccm.201811-2221OC

11   Kuo, C.S., Pavlidis, S., Loza, M., Baribaud, F., Rowe, A., Pandis, I. et al. (2017) A transcriptome-driven analysis of epithelial brushings and bronchial biopsies to define asthma phenotypes in U-BIOPRED. *Am. J. Respir. Crit. Care Med.* **195**, 443–455 https://doi.org/10.1164/rccm.201512-2452OC

12   Jevnikar, Z., Östling, J., Ax, E., Calvén, J., Thörn, K., Israelsson, E. et al. (2019) Epithelial IL-6 trans-signaling defines a new asthma phenotype with increased airway inflammation. *J. Allergy Clin. Immunol.* **143**, 577–590 https://doi.org/10.1016/j.jaci.2018.05.026

13   Östling, J., van Geest, M., Schofield, J.P.R., Jevnikar, Z., Wilson, S., Ward, J. et al. (2019) IL-17-high asthma with features of a psoriasis immunophenotype. *J. Allergy Clin. Immunol.* **144**, 1198–1213 https://doi.org/10.1016/j.jaci.2019.03.027

14   Khoo, S.K., Read, J., Franks, K., Zhang, G., Bizzintino, J., Coleman, L. et al. (2019) Upper airway cell transcriptomics identify a major new immunological phenotype with strong clinical correlates in young children with acute wheezing. *J. Immunol.* **202**, 1845–1858 https://doi.org/10.4049/jimmunol.1800178

15   Svenningsen, S. and Nair, P. (2017) Asthma endotypes and an overview of targeted therapy for asthma. *Front. Med.* **4**, 158 https://doi.org/10.3389/fmed.2017.00158

16   Ivanova, A., Israel, E., LaVange, L.M., Peters, M.C., Denlinger, L.C., Moore, W.C. et al. (2020) The precision interventions for severe and/or exacerbation-prone asthma (PrecISE) adaptive platform trial: statistical considerations. *J. Biopharm. Stat.* **30**, 1026–1037 https://doi.org/10.1080/10543406.2020.1821705

17   Sweeney, T.E., Shidham, A., Wong, H.R. and Khatri, P. (2015) A comprehensive time-course-based multicohort analysis of sepsis and sterile inflammation reveals a robust diagnostic gene set. *Sci. Transl. Med.* **7**, 287ra271 https://doi.org/10.1126/scitranslmed.aaa5993

18   Sweeney, T.E., Perumal, T.M., Henao, R., Nichols, M., Howrylak, J.A., Choi, A.M. et al. (2018) A community approach to mortality prediction in sepsis via gene expression analysis. *Nat. Commun.* **9**, 694 https://doi.org/10.1038/s41467-018-03078-2

19   Jones, A.C. and Bosco, A. (2017) Using network analysis to understand severe asthma phenotypes. *Am. J. Respir. Crit. Care Med.* **195**, 1409–1411 https://doi.org/10.1164/rccm.201612-2572ED

20   de la Fuente, A. (2010) From 'differential expression'to 'differential networking'–identification of dysfunctional regulatory networks in diseases. *Trends Genet.* **26**, 326–333 https://doi.org/10.1016/j.tig.2010.05.001

21   Hudson, N.J., Dalrymple, B.P. and Reverter, A. (2012) Beyond differential expression: the quest for causal mutations and effector molecules. *BMC Genom.* **13**, 356 https://doi.org/10.1186/1471-2164-13-356

22   Langfelder, P. and Horvath, S. (2008) WGCNA: an R package for weighted correlation network analysis. *BMC Bioinform.* **9**, 559–559 https://doi.org/10.1186/1471-2105-9-559

23   Jones, A.C., Anderson, D., Troy, N.M., Mallon, D., Hartmann, R., Serralha, M. et al. (2020) Rewiring of gene networks underlying mite allergen-induced CD4 + Th-cell responses during immunotherapy. *Allergy* **75**, 2330–2341 https://doi.org/10.1111/all.14265

24   Peters, M.C., Ringel, L., Dyjack, N., Herrin, R., Woodruff, P.G., Rios, C. et al. (2019) A transcriptomic method to determine airway immune dysfunction in T2-high and T2-low asthma. *Am. J. Respir. Crit. Care Med.* **199**, 465–477 https://doi.org/10.1164/rccm.201807-1291OC

25   Newman, A.M., Steen, C.B., Liu, C.L., Gentles, A.J., Chaudhuri, A.A., Scherer, F. et al. (2019) Determining cell type abundance and expression from bulk tissues with digital cytometry. *Nat. Biotechnol.* **37**, 773–782 https://doi.org/10.1038/s41587-019-0114-2

26   Aran, D., Hu, Z. and Butte, A.J. (2017) Xcell: digitally portraying the tissue cellular heterogeneity landscape. *Genome Biol.* **18**, 220 https://doi.org/10.1186/s13059-017-1349-1

27   Racle, J. and Gfeller, D. (2020) EPIC: a tool to estimate the proportions of different cell types from bulk gene expression data. bioinformatics for cancer immunotherapy. *Methods Mol. Biol.* **2120**, 233–248 https://doi.org/10.1007/978-1-0716-0327-7_17

28   Kramer, A., Green, J., Pollard, Jr, J. and Tugendreich, S. (2014) Causal analysis approaches in ingenuity pathway analysis. *Bioinformatics* **30**, 523–530 https://doi.org/10.1093/bioinformatics/btt703

29   Zemek, R.M., De Jong, E., Chin, W.L., Schuster, I.S., Fear, V.S., Casey, T.H. et al. (2019) Sensitization to immune checkpoint blockade through activation of a STAT1/NK axis in the tumor microenvironment. *Sci. Transl. Med.* **11**, eaav7816 https://doi.org/10.1126/scitranslmed.aav7816

30   Keenan, A.B., Torre, D., Lachmann, A., Leong, A.K., Wojciechowicz, M.L., Utti, V. et al. (2019) ChEA3: transcription factor enrichment analysis by orthogonal omics integration. *Nucleic Acids Res.* **47**, W212–W224 https://doi.org/10.1093/nar/gkz446

31   Liu, A., Trairatphisan, P., Gjerga, E., Didangelos, A., Barratt, J. and Saez-Rodriguez, J. (2019) From expression footprints to causal pathways: contextualizing large signaling networks with CARNIVAL. *npj Syst. Biol. Appl.* **5**, 40 https://doi.org/10.1038/s41540-019-0118-z

32   Alvarez, M.J., Shen, Y., Giorgi, F.M., Lachmann, A., Ding, B.B., Ye, B.H. et al. (2016) Functional characterization of somatic mutations in cancer using network-based inference of protein activity. *Nat. Genet.* **48**, 838–847 https://doi.org/10.1038/ng.3593

33   Margolin, A.A., Nemenman, I., Basso, K., Wiggins, C., Stolovitzky, G., Favera R, D. et al. (2006) ARACNE: an algorithm for the reconstruction of gene regulatory networks in a mammalian cellular context. *BMC Bioinform.* **7**, S7 https://doi.org/10.1186/1471-2105-7-S1-S7

34   Gardeux, V., Achour, I., Li, J., Maienschein-Cline, M., Li, H., Pesce, L. et al. (2014) 'N-of-1-pathways' unveils personal deregulated mechanisms from a single pair of RNA-Seq samples: towards precision medicine. *J. Am. Med. Inform. Assoc.* **21**, 1015–1025 https://doi.org/10.1136/amiajnl-2013-002519

35   Barbie, D.A., Tamayo, P., Boehm, J.S., Kim, S.Y., Moody, S.E., Dunn, I.F. et al. (2009) Systematic RNA interference reveals that oncogenic KRAS-driven cancers require TBK1. *Nature* **462**, 108–112 https://doi.org/10.1038/nature08460

36   Kuijjer, M.L., Hsieh, P.-H., Quackenbush, J. and Glass, K. (2019) Lionessr: single sample network inference in R. *BMC Cancer* **19**, 1003 https://doi.org/10.1186/s12885-019-6235-7

37   Kuijjer, M.L., Tung, M.G., Yuan, G., Quackenbush, J. and Glass, K. (2019) Estimating sample-specific regulatory networks. *Iscience* **14**, 226–240 https://doi.org/10.1016/j.isci.2019.03.021

38   Jones, A.C., Anderson, D., Galbraith, S., Fantino, E., Gutierrez Cardenas, D., Read, J.F. et al. (2019) Personalized transcriptomics reveals heterogeneous immunophenotypes in children with viral bronchiolitis. *Am. J. Respir. Crit. Care Med.* **199**, 1537–1549 https://doi.org/10.1164/rccm. 201804-0715OC

39   Kobayashi, Y., Kushihara, Y., Saito, N., Yamaguchi, S. and Kakimi, K. (2020) A novel scoring method based on RNA-Seq immunograms describing individual cancer-immunity interactions. *Cancer Sci.* **111**, 4031 https://doi.org/10.1111/cas.14621

40   Regev, A., Teichmann, S., Rozenblatt-Rosen, O., Stubbington, M., Ardlie, K., Amit, I. et al. The human cell atlas white paper. arXiv preprint arXiv:181005192 2018

41   Rozenblatt-Rosen, O., Stubbington, M.J., Regev, A. and Teichmann, S.A. (2017) The human cell atlas: from vision to reality. *Nat. News* **550**, 451 https://doi.org/10.1038/550451a

42   Villani, A.-C., Satija, R., Reynolds, G., Sarkizova, S., Shekhar, K., Fletcher, J. et al. (2017) Single-cell RNA-seq reveals new types of human blood dendritic cells, monocytes, and progenitors. *Science* **356**, eaah4573 https://doi.org/10.1126/science.aah4573

43   Plasschaert, L.W., Žilionis, R., Choo-Wing, R., Savova, V., Knehr, J., Roma, G. et al. (2018) A single-cell atlas of the airway epithelium reveals the CFTR-rich pulmonary ionocyte. *Nature* **560**, 377–381 https://doi.org/10.1038/s41586-018-0394-6

44   Martin, S.L., Saint-Criq, V., Hwang, T.-C. and Csanády, L. (2018) Ion channels as targets to treat cystic fibrosis lung disease. *J. Cyst. Fibros* **17**, S22–S27 https://doi.org/10.1016/j.jcf.2017.10.006

45   Weinreb, C., Wolock, S. and Klein, A.M. (2018) SPRING: a kinetic interface for visualizing high dimensional single-cell expression data. *Bioinformatics* **34**, 1246–1248 https://doi.org/10.1093/bioinformatics/btx792

46   Trapnell, C., Cacchiarelli, D., Grimsby, J., Pokharel, P., Li, S., Morse, M. et al. (2014) The dynamics and regulators of cell fate decisions are revealed by pseudotemporal ordering of single cells. *Nat. Biotechnol.* **32**, 381 https://doi.org/10.1038/nbt.2859

47   Kwok, I., Becht, E., Xia, Y., Ng, M., Teh, Y.C., Tan, L. et al. (2020) Combinatorial single-cell analyses of granulocyte-monocyte progenitor heterogeneity reveals an early uni-potent neutrophil progenitor. *Immunity* **53**, 303–318.e305 https://doi.org/10.1016/j.immuni.2020.06.005

48   Nehar-Belaid, D., Hong, S., Marches, R., Chen, G., Bolisetty, M., Baisch, J. et al. (2020) Mapping systemic lupus erythematosus heterogeneity at the single-cell level. *Nat. Immunol.* **21**, 1094–1106 https://doi.org/10.1038/s41590-020-0743-0

49   Eloranta, M.-L. and Rönnblom, L. (2016) Cause and consequences of the activated type I interferon system in SLE. *J. Mol. Med.* **94**, 1103–1110 https://doi.org/10.1007/s00109-016-1421-4

50   Der, E., Suryawanshi, H., Morozov, P., Kustagi, M., Goilav, B., Ranabothu, S. et al. (2019) Tubular cell and keratinocyte single-cell transcriptomics applied to lupus nephritis reveal type I IFN and fibrosis relevant pathways. *Nat. Immunol.* **20**, 915–927 https://doi.org/10.1038/s41590-019-0386-1

51   Zhang, F., Wei, K., Slowikowski, K., Fonseka, C.Y., Rao, D.A., Kelly, S. et al. (2019) Defining inflammatory cell states in rheumatoid arthritis joint synovial tissues by integrating single-cell transcriptomics and mass cytometry. *Nat. Immunol.* **20**, 928–942 https://doi.org/10.1038/ s41590-019-0378-1

52   Armingol, E., Officer, A., Harismendy, O. and Lewis, N.E. (2020) Deciphering cell–cell interactions and communication from gene expression. *Nat. Rev. Genet.* **22**, 71–88 https://doi.org/10.1038/s41576-020-00292-x

53   Vento-Tormo, R., Efremova, M., Botting, R.A., Turco, M.Y., Vento-Tormo, M., Meyer, K.B. et al. (2018) Single-cell reconstruction of the early maternal–fetal interface in humans. *Nature* **563**, 347–353 https://doi.org/10.1038/s41586-018-0698-6

54   Efremova, M., Vento-Tormo, M., Teichmann, S.A. and Vento-Tormo, R. (2020) CellPhoneDB: inferring cell-cell communication from combined expression of multi-subunit ligand-receptor complexes. *Nat. Protoc.* **15**, 1484–1506 https://doi.org/10.1038/s41596-020-0292-x

55   Hou, R., Denisenko, E., Ong, H.T., Ramilowski, J.A. and Forrest, A.R. (2020) Predicting cell-to-cell communication networks using NATMI. *Nat. Commun.* **11**, 5011 https://doi.org/10.1038/s41467-020-18873-z

56   Browaeys, R., Saelens, W. and Saeys, Y. (2020) Nichenet: modeling intercellular communication by linking ligands to target genes. *Nat. Methods* **17**, 159–162 https://doi.org/10.1038/s41592-019-0667-5

57   Braga FA, V., Kar, G., Berg, M., Carpaij, O.A., Polanski, K., Simon, L.M. et al. (2019) A cellular census of human lungs identifies novel cell states in health and in asthma. *Nat. Med.* **25**, 1153–1163 https://doi.org/10.1038/s41591-019-0468-5

58   Jones, A.C., Troy, N.M., White, E., Hollams, E.M., Gout, A.M., Ling, K.-M. et al. (2018) Persistent activation of interlinked type 2 airway epithelial gene networks in sputum-derived cells from aeroallergen-sensitized symptomatic asthmatics. *Sci. Rep.* **8**, 1511 https://doi.org/10.1038/ s41598-018-19837-6

59   Denisenko, E., Guo, B.B., Jones, M., Hou, R., de Kock, L., Lassmann, T. et al. (2020) Systematic assessment of tissue dissociation and storage biases in single-cell and single-nucleus RNA-seq workflows. *Genome Biol.* **21**, 130 https://doi.org/10.1186/s13059-020-02048-6