



Sketching T cell atlases in the single-cell era: challenges and recommendations

Itana Bojović^{1,2,a}, António GG Sousa^{1,2,a} , Sini Junttila^{1,2} & Laura L Elo^{1,2,3} 

¹ Turku Bioscience Centre, University of Turku and Åbo Akademi University, Turku, Finland

² InFLAMES Research Flagship Centre, University of Turku, Turku, Finland

³ Institute of Biomedicine, University of Turku, Turku, Finland

Keywords

atlas, integration, multi-omics, single-cell, T cells

Correspondence

Laura L Elo, Turku Bioscience Centre, University of Turku and Åbo Akademi University, Tykistökatu 6, 20520 Turku, Finland.

E-mail: laura.elo@utu.fi

^aEqual first-authors.

Received 14 March 2025;

Revised 23 May 2025;

Accepted 26 May 2025

doi: 10.1111/imcb.70040

Immunology & Cell Biology 2025; 1–15

Abstract

Recent advances in single-cell technologies have enabled the creation of comprehensive cell atlases, reference maps of various cell types within organisms. Here we specifically focus on T cell atlases, which offer a detailed catalog of the adaptive immune system at single-cell resolution. As such, they capture cellular diversity, functional states, and spatial dynamics across tissues, developmental stages, and disease conditions. Given the central role of T cells in orchestrating immune responses, their dysregulation underpins autoimmune disorders, cancer progression and failed immunotherapies. Therefore, a unified T cell atlas is critical for decoding such disease mechanisms, identifying therapeutic targets, and advancing personalized treatments. In this article, we explore the latest advances in T cell atlases, describing breakthroughs in multi-omics technologies, spatial profiling and computational frameworks that resolve transcriptional, epigenetic and proteomic heterogeneity. We also address persistent challenges and highlight strategies to address these gaps. Finally, we discuss emerging frontiers set to reshape our understanding of T cell dynamics in both health and diseases. Together, these insights underscore the transformative potential of T cell atlases in reconstructing precision immunology and accelerating therapeutic innovation.

INTRODUCTION

T cells or lymphocytes are key players of the adaptive immune system, providing cell-mediated immunity against harmful pathogens or cancer cells. They continuously monitor the human body to maintain immune homeostasis, but dysregulated T cells may fail to target and kill malignant cells and may even trigger an autoimmune response that harms the body's own tissues. This thin balance makes the phenotypic characterization of different T cell populations of utmost importance to understand human health and advance immunotherapies. In recent years, the expansion of single-cell technologies, which profile molecular features at the single-cell level in an unprecedented way, has revealed an overlooked diversity of T cell populations and functional states shaped by development, tissue-specific functions and immune activation. This suggests that more remains to

be discovered as technologies continue to improve and studies scale up.

The functional plasticity of T cell subsets, such as the one observed in T helper cell populations—Th1, Th2, Th17, Tfh, Treg—highlights the need to catalog their various types and functional states to better understand their roles in immunity. This can be achieved by constructing comprehensive T cell atlases, mapping cells through development¹ and across multiple tissues.² Such atlases enable the identification of phenotypically relevant cell states and their associated differential expression programs, providing a unique resource for studying disease mechanisms and developing therapeutic strategies³ (Figure 1a). This approach can enhance the identification of precise therapeutic targets and establish a foundation for evaluating drug effects. In addition, T cell atlases serve as a tool to monitor and infer immune compatibility in tissue engineering and cell therapy.⁴

Similarly, they can provide insights into transplant medicine by identifying specific T cell populations influencing the body's response to transplanted organs or tissues.⁵ They also can improve our understanding of the immunology of aging and provide insights into the enhancement of immune responses in older adults.⁶

The expansion in T cell atlasing has been driven by breakthroughs in single-cell RNA sequencing (scRNA-seq) and subsequently in other single-cell omics methods. These developments have been underpinned by innovations in technology and computational analysis.^{7,8} Simultaneously, research efforts are converging on integrating scRNA-seq with a range of omics methods.⁹ Accordingly, pairing scRNA-seq with other omics strategies allows the examination of genomic, epigenomic and protein-level information. This integrative approach gives a holistic systems-level view of individual cells, potentially revealing the interaction of different molecular layers that define cell types and states. This is particularly important given the challenges in accurately delineating cell states from transcriptomics data due to the high degree of transcriptional plasticity in cells. Additionally, single-cell transcriptomics can be coupled with spatial technologies and offers a view into cellular localization and intercellular interactions within tissue environments.¹⁰

As single-cell techniques continue to evolve, researchers can now better explain cellular integrity at both individual and spatial levels. As a result, it is possible to

determine cell type, state, transition, location, lineage, as well as cell–cell interactions. Additionally, further advances are paving the way for more accurate predictions of cellular perturbations and their impact on cell fate. Despite this potential, significant biological, technical and computational challenges persist (Figure 1b). Nevertheless, declining costs and rapid advances in computational methods are improving both accessibility and precision. Therefore, it remains to be seen how these advancements will enhance the detection and characterization of rare and small T cell populations, which are one of the key challenges in the creation of a global T cell atlas.

In this article, we review the current state and existing gaps in human T cell atlases in health and disease. In addition, we provide recommendations and describe the opportunities and challenges for building the next generation of T cell atlases.

CELL ATLASING — CHALLENGES AND RECOMMENDATIONS

Cell atlasing is the comprehensive cataloging of cellular heterogeneity within a defined space and time for a specific biological system of interest.¹¹ This process involves exhaustive sampling of the biological system to increase the likelihood of capturing all cell types and states, including rare populations. The success of this Herculean task depends on the effective integration of

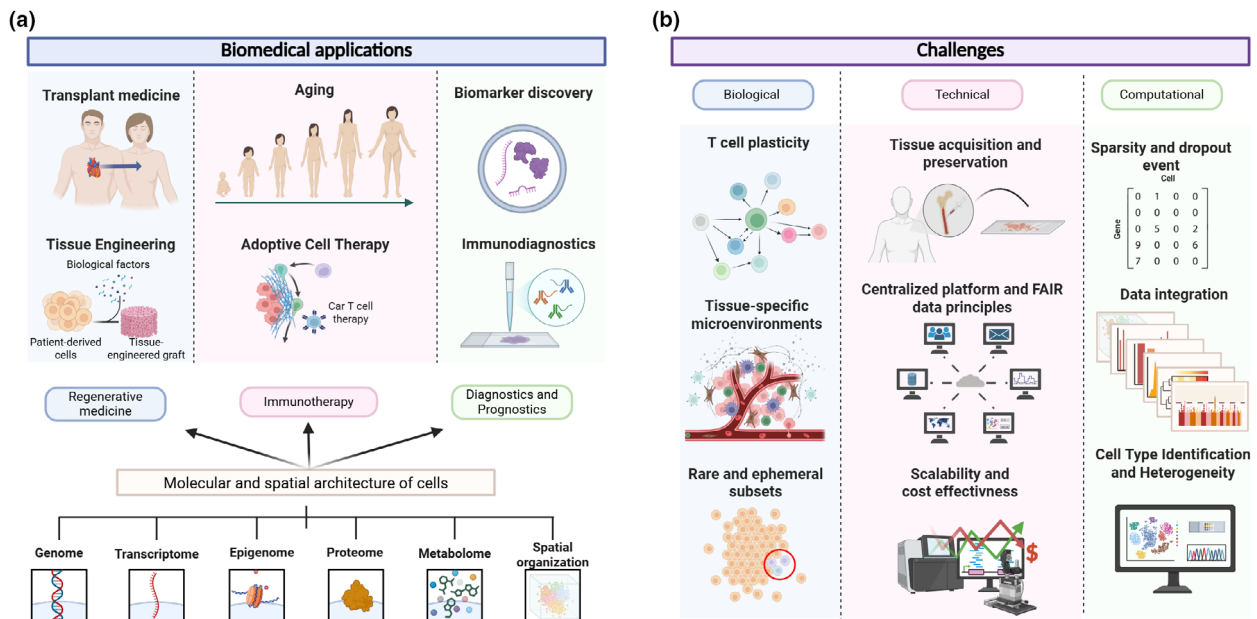


Figure 1. T cell atlas: applications and challenges. **(a)** Examples of biomedical applications enabled by profiling the molecular and spatial architecture of T cells in tissues. **(b)** Challenges encountered in building a T cell atlas. This figure was created with [Biorender.com](https://biorender.com).

multiple single-cell datasets from several donors, which are often obtained using different sampling, preservation, library preparation and sequencing protocols.¹¹ The main steps in building a cell atlas are (1) defining the scope, (2) data collection, (3) data integration, (4) quality assessment, (5) annotation and (6) data sharing. The three broad challenges faced in building a T cell atlas are the underlying biological complexity of T cells, the technological limitations of current methodologies, and sufficient computing resources (Figure 1b).

Defining the scope

Ideally, the scope should not be too broad, to avoid strong or nested batch effects that could mask subtle biological signals, nor too narrow, to ensure comprehensive characterization of cellular phenotypes within the biological system under study. Defining the scope involves selecting the cell type(s) (phenotypic heterogeneity), tissue(s) (space) and developmental phase(s) (time) from the biological system of interest (organism level) under steady (healthy) and/or perturbed (disease) conditions. For example, the cell atlas built by Park *et al.*¹ cataloged thymic cells across the human lifespan.

Data collection

The collection of datasets is guided by the defined scope and may rely solely on newly generated data,² publicly available datasets, or both.¹ Generating new data allows for the standardization of protocols, minimizing technical variation. Although incorporating public datasets avoids cumbersome repetition and potentially enriches the atlas by increasing coverage (e.g. donors and cell heterogeneity) and including otherwise inaccessible samples, it also presents several challenges. Raw data are often unavailable—especially in human studies—limiting use to processed data. Datasets annotated with older reference genome versions can lead to “misalignment” when compared with newer versions or may lack annotations for relevant features, such as long non-coding RNAs.¹² Quality control steps may have removed underrepresented populations, such as double-positive T cells, flagged as doublets.

Single-cell sequencing technologies used across datasets may vary significantly, which can greatly impact downstream analyses. Key differences between plate-based (low-throughput, high-coverage) and droplet-based (high-throughput, low-coverage) methods influence cell and transcript capture as well as dropout rates.^{13,14} Although immune cells from blood do not require tissue dissociation, immune cells from other tissues do, in order

to obtain a viable cell suspension. The cell suspension protocol should be standardized to minimize artifacts arising from differences in cell viability and cell type composition.¹⁵ Additionally, some samples from problematic tissues may only be available through single-nucleus RNA sequencing (snRNA-seq), which lacks cytoplasmic transcripts such as mitochondrial RNAs.¹³ While snRNA-seq is unaffected by cell size and avoids stress-related gene expression artifacts caused by dissociation,¹³ it presents challenges when integrating with scRNA-seq.

Datasets available in single-cell repositories (e.g. GEO, Single Cell Portal) are not always provided in the most flexible format—raw counts. Instead, they are often transformed (e.g. Transcripts Per Million), which can hinder the downstream application of count-dependent computational methods, such as Scrublet¹⁶ or scDblFinder¹⁷ for doublet detection or scVI for integration.¹⁸ Metadata associated with public datasets may be missing, aggregated, or incomplete, often requiring harmonization across datasets.¹¹ All these factors should be carefully considered for each potential dataset, as they will significantly influence downstream steps, particularly integration.

Examples of data modalities

While single-cell transcriptomics remains the cornerstone for cell atlas construction, integrating complementary modalities can substantially enhance their biological resolution and functional utility (Figure 2a). A comprehensive cell architecture must describe several interconnected features to accurately capture the full complexity of T cells. These include cellular function, structural characteristics, developmental and differentiation status, interactions with other cells, metabolic profiles and regulatory mechanisms (Figure 2b). Capturing these features often relies on a combination of advanced technologies and rigorous data analysis approaches, where key bioinformatics steps critically shape the outcomes (Table 1).

Cell surface markers provide complementary information for identification of cell types, as conventional short-read scRNA-seq alone is sometimes insufficient to resolve the full heterogeneity of T cell subsets.²⁷ Cellular Indexing of Transcriptomes and Epitopes by sequencing (CITE-seq) allows the simultaneous measurement of gene expression and surface markers within the same cell and has been applied to overcome this limitation.^{28,29} T cells possess unique T cell receptors (TCRs) that enable recognition of various antigens and act as intrinsic signatures for detecting clonal expansion, diversity and migration.

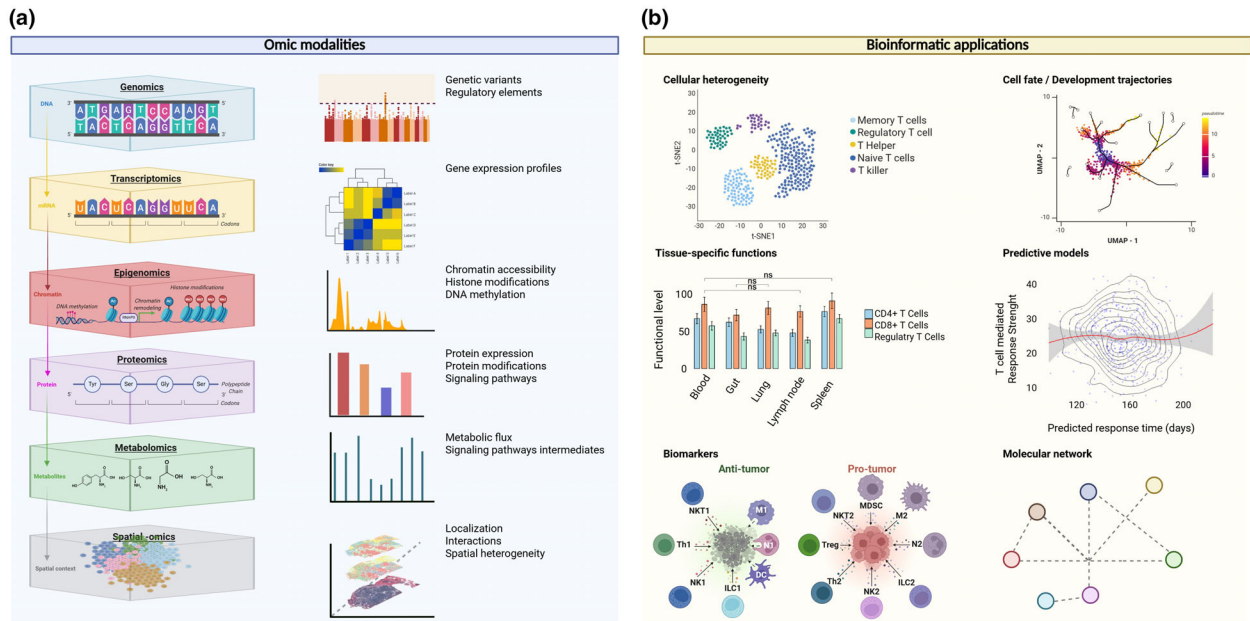


Figure 2. Biological insights provided by different (a) omics and (b) bioinformatic analyses. This figure was created with [Biorender.com](https://biorender.com).

Table 1. Key bioinformatic steps, considerations, challenges and methods for assembling a cell atlas.

Category	Consideration	Challenges	Methods	References
Quality control	Initial QC steps to ensure high-quality data	Data loss from threshold-based QC	Adaptive QC models, data-driven thresholding	19
Contamination	Ambient molecules (RNA, DNA, proteins) from dying cells, debris, or reagents	Background noise	Prior knowledge of negative cell markers, statistical frameworks for background noise reduction	
Sparsity & dropout handling	Differentiate biological from technical zeros in gene and protein expression data	High sparsity in expression matrices	Imputation, non-imputation (i.e., binarization), pseudo-bulk analysis	20,21
Normalization	Mitigate bias due to technical variability (sequencing depth, capture efficiency, library size, among others)	High batch effect variability across platforms	Scaling, normalization, generalized linear models, mixed methods, deep learning-based methods	22
Feature selection	Selecting most relevant features reduces dimension and redundancy, thus enhance performance in predictive models in downstream analyses	Managing high feature numbers, computational scaling	Filters (statistical criteria), wrappers (predictive models), embedded methods (model training), hybrid and deep learning-based approaches	23
Dimensionality reduction	Revealing meaningful patterns, improving clustering accuracy and facilitate data visualization	Loss of information, interpretability, scalability	Linear, nonlinear, matrix factorization, graph-based, probabilistic, deep learning-based methods	24
Batch correction/ data integration	Cross platform and cross-dataset integration	Harmonizing data, batch effects, technical noise	Global, linear embedding, graph-based and deep learning-based models	

For detailed descriptions of available methods and best practices in single-cell analysis, consult the following online manuals: "Orchestrating single-cell analysis with Bioconductor"²⁵ (<https://bioconductor.org/books/release/OSCA>) and "Single-cell best practices"²⁶ (<https://www.sc-best-practices.org>).

Integrating TCR-seq with scRNA-seq reveals how TCR variability shapes the functional and physiological landscape of T cell populations. However, the vast majority of TCRs are unique to each individual across time and are rarely shared³⁰ making it challenging to extract common immune response signatures from TCR-seq data, thus hampering its utility in T cell atlases.

Given the substantial costs of library preparation and sequencing, single-cell genomics remains an infrequent choice for single-cell atlases. As it surpasses transcriptomics in sensitivity for detecting genetic variants, the best results are achieved when they are used in combination,³¹ enhancing the sensitivity for identifying point mutations and structural variations, thus enabling the identification of rare subpopulations and detailed lineage mapping. While this has been particularly impactful in cancer research—revealing evolutionary trajectories of primary tumors and the tumor microenvironment^{32,33}—it also has direct relevance to T cell atlases. Specifically, incorporating single-cell genomics can uncover pathogenic T cell clones, resolve clonotype diversity, and map antigen-driven T cell evolution in both health and disease, thus enriching T cell atlases with clinically relevant immune heterogeneity.³⁴

Single-cell proteomics directly measures protein abundance, captures post-translational modifications, reveals protein–protein interactions, and construes functional states and spatial contexts that cannot be properly derived from transcriptomic data. Over the last years, single-cell proteomics has advanced rapidly; nevertheless, significant challenges remain in achieving comprehensive detection. A central limitation stems from the inherently low abundance of proteins in the cells. Unlike nucleic acids, they cannot be amplified, rendering their detection and quantification particularly demanding. Therefore, highly sensitive instrumentation and proper sample preparation are needed, along with standardized workflows.

Epigenomic profiling reveals key regulatory regions, including non-coding elements such as enhancers. These modifications are integral to T cell differentiation and the regulation of cell-type specific gene expression.³⁵ For instance, single-cell epigenomic profiling has revealed alterations in cis- and trans-regulatory elements that orchestrate gene expression in specific T cell subsets.³⁶ Transcriptional and epigenetic modifications collectively define attractor states within the cellular landscape, corresponding to distinct cell types. Thus, it is no surprise that epigenomics is one of the most used complementary approaches to single-cell transcriptomics in atlasing.

Metabolites have a key role in immune response acting as critical activation, differentiation and epigenetic remodeling regulators.³⁷ Single-cell metabolomics is arguably the most complex omics technology, despite being one of the fastest and most cost-effective.³⁸ This complexity arises from the dynamic nature of the metabolome, which is highly susceptible to perturbations at every stage, beginning with sample collection that often disrupts the native state.³⁸ Furthermore, external factors such as nutrition, environment and circadian rhythms can significantly alter the innate metabolomic profiles.³⁹ In addition, identification and quantification of metabolites have long been significant challenges in metabolomics, largely due to chemical diversity and structural complexity but also the lack of technology for amplification, which restricts the detection of low-abundance metabolites.⁴⁰ However, ongoing efforts to expand and refine metabolite databases are addressing these limitations.^{41–43}

Spatial contextualization that encompasses direct measurement of molecular parameters within intact tissue holds great potential by preserving spatial context and cellular interactions. It can be integrated with different omics technologies, enabling a multi-modal understanding of molecular processes. Among the omics approaches, spatial transcriptomics and spatial proteomics have seen the most widespread adoption. Further exploration of spatial metabolomics could provide valuable insights, particularly given the profound influence of the native microenvironment on cellular biochemistry, especially within the tumor microenvironment.^{38,44} These methods still need to overcome challenges of cost, trade-offs between sensitivity, throughput and resolution (e.g. lack of true single-cell resolution), as well as imaging complexity and complex data analysis workflows.¹⁰ To address some of these limitations, bioinformatic approaches have been developed to integrate spatial information with scRNA-seq data, enabling the prediction of spatial distributions of undetected transcripts and improving cell type deconvolution.⁴⁵ Despite these advances, the application of spatial single-cell technologies to T cell atlasing remains limited, particularly because commonly used sample types such as blood and non-mucosal tissues lack spatial organization and contain relatively few T cells. Therefore, this article focuses on non-spatial single-cell technologies, with a particular emphasis on transcriptomics. If the atlas being constructed includes multi-modal data, such as antibody-derived tags (e.g. from CITE-seq) or TCR repertoires, these modalities should be consistently considered across all stages of the atlas construction process: (1) defining the scope, (2)

data collection, (3) data integration, (4) quality assessment, (5) annotation and (6) data sharing.

Data integration

Integration is the key step in cell atlasing. Integration is a broad concept used to refer to the identification of shared cell populations and states across heterogeneously generated datasets. For example, “integration” has been used to describe the identification of shared cell populations within the same modality across different biological conditions (e.g. healthy versus diseased) and/or batches (e.g. sequencing runs), or across modalities of single-cell data (e.g. RNA-seq and ATAC-seq).⁴⁶ Integrating cells across the same modality involves specifying a batch covariate, that is, a variable representing non-biological experimental confounding effects, such as sample processing or library preparation protocols. A biological variable may instead be provided as the batch covariate, such as species or disease status, when performing cross-species or cross-tissue integration tasks.¹¹ For example, the cross-tissue immune cell atlas integrated ~330 K cells by donor ($n = 12$; Dominguez Conde *et al.*²), with each donor contributing multiple tissue samples processed through three distinct protocols, and libraries prepared with or without hashtags using different chemistries. In contrast, the pan-cancer tumor-infiltrating T cell atlas integrated ~400 K cells from 316 donors by dataset ($n = 35$; Zheng *et al.*⁴⁷), each dataset representing a unique combination of independent study ($n = 22$), cancer type ($n = 21$), and library preparation method ($n = 4$). In addition to the batch covariate, some integration methods, such as scVI,¹⁸ allow the “correction” for other covariates (e.g. age or sex). If cell type annotations are available, these can be used to guide integration by using semi-supervised integration methods such as scANVI⁴⁸ or STACAS.⁴⁹ Ultimately, a key consideration when choosing an integration method is the reusability of its output model, allowing for reference mapping of additional studies once the model is built. For example, scArches relies on a conditional autoencoder-based model,⁵⁰ while Coralys uses logistic regression models.⁵¹

Quality assessment

Assessing integration is fundamental to ensure that integration neither distorted the data by masking subtle biological variation nor left technical noise uncorrected. Approaches to assess the quality of integration include low-dimensional reduction techniques, expression of canonical cell type markers, expression of cluster markers, pre- versus post-integration clustering comparison,

automatic cell type annotation, and quantification of bio-conservation and batch-correction metrics. Low-dimensional representations are useful for visually inspecting the overlap between batches. Drawing strong conclusions should, however, be avoided, as they provide an oversimplified view of the data, being unable to preserve high-dimensional distances.⁵² Canonical cell type markers can help validate and annotate well-defined clusters across batches, whereas cluster markers identified through differential expression analysis across clusters aid in validating fine-grained cell populations that may represent distinct cell states.⁴⁶ Comparison of clusters from individual batches with those from the batch-corrected data ensures that rare populations—particularly those shared across only a few batches—are preserved during integration. Automatic cell type annotation methods, such as Azimuth,²⁸ CellTypist,² or ArchMap,⁵³ provide fast and comparable annotations that can help validate integration and guide the annotation step. Finally, metrics quantifying biological conservation and batch correction, such as the cell-specific mixing score from the CellMixS Bioconductor package⁵⁴ or those implemented in the scib python package,⁵⁵ provide objective measures of the effectiveness of integration.

Given the complexity of the integration task, benchmarking methods is common to identify the one that best balances batch correction and biological conservation. Domínguez Conde *et al.*² selected BBKNN⁵⁶ over scVI¹⁸ for constructing their human cross-tissue immune cell atlas, as it demonstrated superior performance. In contrast, Sikkema *et al.*⁵⁷ identified scANVI⁴⁸ as the top performer among 12 tested methods for building their human lung atlas. Filippov *et al.*⁵⁸ validated their aging blood immune atlas, integrated with scVI,¹⁸ against Harmony,⁵⁹ finding that both methods performed similarly. Community-driven benchmarking platforms such as Omnibenchmark⁶⁰ and Open Problems⁶¹ provide useful resources and guidelines for benchmarking integration. If the assessment shows subpar results, a new round of integration should be performed using a different set of parameters or an alternative method.

Annotation

Annotation consists of identifying the cell type identity of individual cells by means of clustering, automatic annotation, or a combination of these two. Clustering-based annotation is recommended for poorly characterized biological systems, such as tertiary lymphoid structures, whereas automatic annotation is recommended for well-known tissues, like blood, for which good reference data exist.

There are three types of approaches for automatic annotation: marker-based, reference-based, and a combination of the two. Marker-based methods, such as Garnett,⁶² scType⁶³ or MMoChI,²⁹ rely on predefined cell type markers—either positive or both positive and negative—to annotate cells. They generally provide accurate high-level annotations, where well-curated markers are available, but often lack low-level annotations.⁶⁴ On the contrary, reference-based methods such as SingleR,⁶⁵ Azimuth,²⁸ CellTypist,² and ArchMap⁵³ are more flexible, allowing the identification of both cell types and subsets at lower levels, but are more computationally demanding. To overcome this, CellTypist (<https://www.celltypist.org>), Azimuth (<https://azimuth.hubmapconsortium.org>) and ArchMap (<https://www.archmap.bio>) provide web applications allowing users without coding experience to annotate their datasets against available references. In both marker- and reference-based approaches, the user may rely on a pre-existing list of markers or a reference model, or provide their own list or reference. Since the goal of building an atlas is to comprehensively catalog the cellular landscape within a given scope, identification of fine-grained cell populations is important, as they may represent rare or overlooked subsets. Therefore, an approach combining clustering with automatic annotation is recommended.^{29,58} Regardless of the approach chosen for cell annotation, classifications must follow best practices in the field, following cell type ontology recommendations⁶⁶ to enable comparison, reuse, and hierarchical classification. Annotation is an iterative process that often requires multiple rounds of classification and assessment to fine-tune cell type labels. The assessment aims to validate annotations based on marker expression, whether canonical or identified through differential expression analysis. Ultimately, one must be able to isolate all cell types and functional states *in vitro* from the respective tissue samples using the identified markers.

Data sharing

The process of building a single-cell atlas culminates in its publication. Publishing the atlas requires sharing the data and accompanying materials: code deposition (e.g. GitHub or GitLab), sharing integrated model (e.g. ICP, variational autoencoder), processed data and metadata deposition (e.g. GEO, HCA Data Portal, custom website) and methodology description (article, repositories). In summary, sharing data should follow the FAIR principles. The published atlas should be easily findable—unique identifier—and accessible by submitting raw and processed data to repositories such as GEO or Single Cell

Portal. Processed data from public datasets should be deposited in public platforms like Figshare or Zenodo to obtain a unique digital object identifier. Sharing files solely through institutional or custom websites/servers is not recommended, as there are no guarantees regarding their long-term availability and security. To ensure interoperability and reusability, the atlas should ideally be shared in a single-cell object format, such as SingleCellExperiment²⁵ or SeuratObject²⁸ for R, and anndata⁶⁷ for python. This allows for easy distribution of assays, metadata, embeddings and parameters used to build the atlas in a single instance. Additionally, it is imperative to submit the scripts to common repositories such as GitHub or GitLab. Objects and scripts should be distributed under a license that promotes reuse and ensures accessibility, such as CC-BY (Creative Commons Attribution) or an open-source software license (e.g. MIT, GPL). Finally, sharing the atlas through a queryable online platform such as CZ CELLxGENE increases visibility and reusability. A good example following these practices is the human cross-tissue immune cell atlas.²

CURRENT HUMAN T CELL ATLASES

Existing single-cell atlases comprehensively profiling T cell populations have focused on either healthy^{1,68–71} or diseased^{47,72–80,81–86} samples (Table 2). The disease conditions studied can be broadly categorized into cancer,^{47,72–80} infections,^{81,82} and (auto)immune disorders.^{83–86} Park *et al.* sampled thymic cellular diversity at the gene expression level throughout human life, cataloging more than 50 cell populations, including 12 representing T cell subsets, which comprise conventional (e.g. CD4⁺ and CD8⁺ T cells) and unconventional (e.g. CD8 α and γ δ T) populations.¹ This allowed the authors to discover a previously uncharacterized population of GNG4⁺ CD8 α T cells in the thymic medulla and identify a TCR α V-J bias in CD8⁺ T cells using single-cell TCR analysis. Extending this work to other lymphoid and non-lymphoid tissues, Wong *et al.*⁶⁸ and Domínguez Conde *et al.*² assembled human cross-tissue immune cell atlases, analyzing surface markers with mass cytometry and gene expression with scRNA-seq. Both studies characterized T cell populations and identified tissue-specific molecular signatures. The latter study included TCR repertoire data for the T cell compartment comprising 185 806 T cells, categorized into 15 cell types representing different states, allowing the description of clonotype distribution across human tissues. A key difference between these studies is the identification of naive and central memory T cells, which, among other markers, relies on the expression of alternative isoforms of the *PTPRC* gene, resulting in the

Table 2. Description of published healthy and diseased T cell atlases that include fine-grained cell populations.

Atlas	References	Tissue	Resolution	Study finding	Analytical methods	Number of cells	Number of donors	Cancer/disease	Data availability
Enhancers across helper T cell diversity	71	Blood	136 (not annotated)	Genetic risk of immune-mediated diseases	5' snRNA-seq 5' scRNA-seq CITE-seq snATAC-seq + 3' snRNA-seq	1 M	56	Healthy	NBDC: hum0350 JGA: JGAS000689 GEA and DDBI: PRIDB13816 Synapse: syn49637038
General longitudinal blood immune cell atlas	70	PMBC	41	Age related alterations in PBMCs	scRNA-seq scTCR-seq scBCR-seq Protein feature barcoding	2 M	166	Healthy	
T cell differentiation atlas	69	PBMC	14	Regulatory elements that control CD8 ⁺ T cell subset gene expression	RNA-seq ATAC-seq CRISPRI	Not applicable	24	Healthy	GEO: GSE179613
General thymus cell atlas	1	Thymus	21	T cell development across lifetime in thymus and bias in human TCR repertoire formation	scRNA-seq scTCR-seq	256 K	24	Healthy	ArrayExpress: E-MTAB-8581 https://developmentcellatlas.ncl.ac.uk Not available
T cell atlas in healthy tissues	68	8 tissues	75	T cell trafficking and functional markers for tissue specificity	CytoF	Not indicated	31	Healthy	
General T cell lymphoma atlas	72	Skin	5	TH2-like programs in malignant T cells, spatial interactions with B cells, and tumor-supportive microenvironment in cutaneous T cell lymphoma	scRNA-seq TCR-seq DNA-seq Bulk RNA-seq Spatial transcriptomics	420 K	45	Cutaneous T cell lymphoma	https://collections.cellatlas.io/ctcl
T cell acute lymphoblastic leukemia	73	Bone marrow PBMC	9	Cellular and genetic factors contributing to resistance and relapse in T cell acute lymphoblastic leukemia	scRNA-seq CITE-seq scATAC-seq bulk RNA-seq	604 K	48	T cell acute lymphoblastic leukemia	dbGaP: phs003432
Pre-infusion CAR T atlas	74	PBMC	17	Pre-infusion CAR T cells with high type 2 functionality correlation with long-term leukemia remission	scRNA-seq CITE-seq Multiplexed secretomic assay	696 K	88	Acute lymphoblastic leukemia	GEO: GSE262072

(Continued)

Table 2. Continued.

Atlas	References	Tissue	Resolution	Study finding	Analytical methods	Number of cells	Number of donors	Cancer/disease	Data availability
Atlas of CD19 chimeric antigen receptor T cells	75	PBMC	26	Elevated glycolysis gene expression in CAR T cells correlates with complete remission at three months	scRNA-seq	417 K	59	Relapsed/refractory large B-cell lymphoma	CellXGene: https://cellxgene.cziscience.com/collections/d9a9760c-e43d-404d-86aa-447c9a1e9120 https://singlecell.manderson.org/TCM/
Pan-cancer atlas of T cells	76	13 tissues	32	Stress response state across different tumor types	scRNA-seq Spatial profiling	308 K	324	16 cancer types	GEO: GSE31519, GSE176078
General breast cancer immune cell atlas	77	Breast	11	Unique T cell-B cell interaction that fosters a more immunosuppressive profile	scRNA-seq	118 K (T cell: 93 K)	21	Breast tumours	GSA-BDC: HRA001867
General urothelial carcinoma cell atlas	78	Urothelial (urinary tract)	6	Basal upper tract urothelial carcinoma features exhausted T cells interacting with immunosuppressive macrophages, driving immune evasion and affecting therapy response	scRNA-seq	67 K (T cells: 21 K)	12	Upper tract urothelial carcinoma	GSA-BDC: HRA000704 OMIX: OMIX001073
General gastric cancer cell atlas	79	Stomach Peripheral blood	20	Tc17-driven T cell exhaustion and stromal-immune crosstalk drive immunosuppression in gastric cancer	scRNA-seq scTCR-seq scBCR-seq bulk DNA-seq bulk RNA-seq	166.5 K	10	Gastric tumor	EGA: EGAC00001001616 GEO: GSE149614
Hepatocellular carcinoma multicellular ecosystem atlas	80	Liver Lymph node Portal vein thrombus	10	Early tertiary lymphoid structures enrichment with antitumor central memory T cells	scRNA-seq	25 K	10	Hepatocellular carcinoma	GSA: PRJCA001702 GEO: GSE156728 http://cancer-pku.cn:3838/PanC_T
Pan-cancer atlas of T cells	47	15 tissues	41	Landscape of tumor-infiltrating T cells (heterogeneity, dynamics and key pathways of CD8+ T cells exhaustion)	scRNA-seq scTCR-seq	398 K	316	21 cancer types	

(Continued)

Table 2. Continued.

Atlas	References	Tissue	Resolution	Study finding	Analytical methods	Number of cells	Number of donors	Cancer/disease	Data availability
Sepsis T cell atlas	81	PBMC	32	Dynamic changes of T cell subsets during sepsis onset and progression	scRNA-seq	911.6 K	16	Sepsis	GEO: GSE1673639 , GSE17545310
HIV immune cell atlas	82	PBMC	8	Analysis of gene signatures associated with immune cell exhaustion during human immunodeficiency virus infection	scRNA-seq	66.5 K	9	Human immunodeficiency virus	GEO: GSE157829
"TCR-first" T cell atlas	83	Multi-tissue	49	The utility of a TCR-first approach	scRNA-seq scTCR-seq	494 K	90	Multi-disease	Zenodo repository: https://doi.org/10.5281/zenodo.10809382
General skin Immune-Stromal Atlas	84	Skin	9	Immune and stromal network's link to disrupted epithelial differentiation in atopic dermatitis	scRNA-seq scVDJ-seq scATAC-seq	707 K	19	Scleroderma Atopic dermatitis	GEO: GSE158432 , GSE153760 , GSE147424 https://developmental-cellatlas.io/diseased-skin
Intrahepatic T cell atlas	85	Liver	13	Tissue-resident, naive-like CD4 ⁺ T cells are poised for TH17 differentiation	scRNA-seq scTCR-seq CITE-seq	22 K	11	Primary sclerosing cholangitis	ArrayExpress: E-MTAB-10143
Colonic CD8 ⁺ T cells in ulcerative colitis	86	Colon	14	The dynamic interplay between cell states and their interactions with different epithelial cell subtypes in ulcerative colitis	scATAC-seq scRNA-seq scV(D)J-seq	8.6 K	6	Inflammatory bowel diseases	GEO: GSE148837 , GSE148505
Human cell atlas portal Single cell atlas Human immunology project consortium				https://www.humancellatlas.org/ https://www.singlecellatlas.org/ https://hipc-dashboard.org/					

Background colors denote study classification: green for healthy tissue, light orange for cancer, blue for infections, and light purple for (auto)immune diseases. Some diseased cell atlases also include healthy cells. "Resolution level" refers to the number of T cell clusters or states reported in the original publications and is not directly comparable across studies. Additionally, the table concludes with integrative resources that aggregate multiple independent cell atlas studies. Databases referenced in these studies are as follows: dbGap, The database of Genotypes and Phenotypes; DDBJ, DNA Data Bank of Japan; GEA, Genomic Expression Omnibus; GEO, Gene Expression Omnibus; GSA, Genome Sequence Archive; GSA-BDC, Genome Sequence Archive in BIG Data Center (Beijing Institute of Genomics); JGA, Japanese Genotype-phenotype Archive; NBDC, National Bioscience Database Center Human Database.

expression of either CD45RA or CD45RO protein. Since this distinction cannot be reliably quantified using scRNA-seq, naive and central memory T cell populations appeared grouped in Domínguez-Conde *et al.*²

Additionally, several other healthy atlases have been published, comprising circulating T cells with different scopes.^{58,69–71} Giles *et al.* built a bulk epigenetic and transcriptional cell atlas of 14 sorted circulating T cell populations, providing higher resolution for CD8⁺ T cell populations (10 populations versus 4 for CD4⁺ T cells).⁶⁹ Among these, a population of PD1⁺CD39⁺CD8⁺ T cells with an exhausted phenotype was identified as the epigenetically closest to tumor-infiltrating T cells (TILs). This helped identify cytotoxic and WNT signaling signatures as key differentiators of responders and non-responders, respectively, in anti-PD1-treated melanoma patients. Oguchi *et al.*⁷¹ constructed a multimodal CD4⁺ T cell atlas integrating gene expression and enhancer activity with chromatin accessibility (snATAC-seq) for approximately one million cells. This study aimed to investigate immune-mediated diseases and pinpointed the cell-type specificity of disease-associated single nucleotide polymorphisms (SNPs) within the identified bidirectionally transcribed candidate enhancers. Finally, Terekhova *et al.*⁷⁰ and Filippov *et al.*⁵⁸ assembled human atlases of aging peripheral blood mononuclear cells (PBMCs). Terekhova *et al.*⁷⁰ profiled approximately two million PBMCs from 166 healthy donors aged 25–85 years. Only certain T cell subsets (TRAV1-2⁻CD8⁺ T cells, MAIT and $\gamma\delta$ T cells) showed statistically significant changes over the human lifespan. These results were further validated by Filippov *et al.*,⁵⁸ who found a decrease in CD8⁺ naive T cells and MAIT cells with age.

As T cells play a crucial role in cancer, many cell atlases focusing on T cell heterogeneity have been assembled for cancers such as lymphoma,⁷² T cell acute lymphoblastic leukemia (T-ALL⁷³), triple-negative breast cancer,⁷⁷ urothelial carcinoma,⁷⁸ and gastric cancer,⁷⁹ as well as cancer therapies like CAR T-cells^{74,75} (see Table 2). To decipher pan-cancer gene expression patterns in TILs, Zheng *et al.*⁴⁷ and Chu *et al.*⁷⁶ constructed pan-cancer T cell atlases comprising 21 and 16 cancer types, respectively, with over 300 donors and more than 300 000 cells each. Through trajectory analyses, the atlases enabled Zheng *et al.*⁴⁷ to identify two developmental pathways leading to T cell exhaustion, one progressing through effector memory T cells and the other through tissue-resident memory T cells. Similarly, Chu *et al.*⁷⁶ identified an overlooked stressed T cell state characterized by the expression of heat shock genes, which was further validated in histological samples through spatial transcriptomics analyses and found to be

associated with a poorer response to immune checkpoint therapy.

Disease focused atlases with a scope beyond cancer include those for primary sclerosing cholangitis,⁸⁵ ulcerative colitis,⁸⁶ and human immunodeficiency virus (HIV).⁸² The transcriptomic atlas from Wang *et al.*⁸² analyzed PBMCs from both healthy and HIV-infected donors, profiling immune populations, including eight distinct T-cell subsets. This enabled the identification of an HIV-specific KLRG1⁺ CD8⁺ T cell population with an exhausted phenotype, which could be restored by blocking the inhibitory receptor KLRG1. In a colonic CD8⁺ T-cell atlas integrating scRNA-seq, TCR repertoire analysis and CyTOF data from healthy individuals and ulcerative colitis patients, Corridoni *et al.*⁸⁶ identified an IL-26 expressing CD8⁺ T cell population with an innate-like phenotype in ulcerative colitis. This population exhibited anti-inflammatory properties, attenuating acute colitis in a humanized IL-26 transgenic mouse model. Similarly, Poch *et al.*⁸⁵ identified a tissue-resident naive-like CD4⁺ T cell population, with a propensity to acquire a Th17 polarization state, which was expanded in primary sclerosing cholangitis patients, based on their intrahepatic T-cell atlas comprising 22 198 transcriptomes.

Studies employing single-cell methods are often limited by the narrow demographic scope of sampled donors. Beyond the limited availability of healthy samples, there is also a need for geographical, gender and ethnic diversity, since ~90% of human genomic data collected to date comes from individuals of European origin.⁸⁷ For the time being, most published studies only report information regarding the number and age of donors. Overall, existing T cell atlases have proven invaluable by enabling the identification of phenotypically relevant cell states in both health and disease, elucidating developmental trajectories, and exploring disease-related mechanisms with potential therapeutic applications.

FUTURE PERSPECTIVES

The future of T cell atlases heavily relies on collaborative work. Advancing our understanding of T cells across diverse tissues and conditions requires pooling expertise, data and technologies. This effort is supported by the continuous development and enhancement of computational tools, which enable the analysis and integration of the vast and complex datasets generated. Deep learning methods have recently proven to be highly effective in single-cell analysis.⁸⁸ A critical next frontier lies in developing predictive models to simulate cellular behaviors, disease trajectories, and response to

environmental changes, drug treatments, or genetic perturbations.

The development of T cell atlases will benefit not only from advances in computational single-cell analysis methods but also from continuous improvements in technological platforms. For instance, 10x Genomics has recently launched a new chromium product GEM-X, designed for lower cost and higher cell profiling per run. These advancements also include compatibility with previously inaccessible sample types, such as fresh, frozen, PFA-fixed tissue, whole blood and FFPE samples. Improvements in sample recovery now allow researchers to work more effectively with limited samples like small biopsies and flow-sorted cells.

Transcriptome profiling is currently the mainstay of T cell atlases, but technological advances and cost reduction may soon enable routine integration of multiple other omics layers. The shift from descriptive profiling to the functional characterization of cellular states is already underway, and this trend is expected to accelerate with the further development of next-generation approaches. The commercialization of third-generation sequencing (Oxford Nanopore Technologies and PacBio single-molecule real-time sequencing) holds great promise for improved characterization of genomic regions, including structural variations, repeated regions and long-range interactions. Furthermore, omics technologies aim to minimize assumptions, improving the accuracy of inferred biological mechanisms. Simultaneous profiling of diverse biomolecules will uncover direct molecular interactions. These advancements promise to enable systems-level insights into intracellular and extracellular regulatory networks.

T cell states are highly dynamic, and capturing those transitions without cell destruction would allow T cell atlases to collect more detailed, time-based data. Although advances such as live cell imaging paired with single-cell RNA-seq offer glimpses into these dynamics, significant issues persist.^{89,90} There are also some advances in direct visualization of chromatin structure in the cell nucleus^{90,91} but we are about to see the future of live cell imaging that will enable the tracking of chromatin changes relative to transcriptional and morphological transitions.

In the future, we also expect an expansion of T cell atlases across different species for the estimation of evolutionary relationships, including conservation and divergence of T cell function. Comparative analyses of immune repertoires, activation pathways and microenvironmental crosstalk across phylogenies will deepen our understanding of adaptive immunity's evolutionary logic.

ACKNOWLEDGMENTS

AGGS and LLE were supported by the European Union's Horizon 2020 research and innovation programme under the Marie Skłodowska-Curie grant agreement no.: 955321. AGGS was also supported by the University of Turku, Åbo Akademi University, Turku Graduate School (UTUGS). LLE reports grants from the Research Council of Finland (329278, 341342, 364700), the Sigrid Juselius Foundation and the Cancer Foundation Finland during the conduct of the study. Our research is also supported by Biocenter Finland and ELIXIR Finland. Open access publishing facilitated by Turun yliopisto, as part of the Wiley - FinELib agreement.

AUTHOR CONTRIBUTIONS

Itana Bojović: Conceptualization; visualization; writing – original draft; writing – review and editing. **António GG Sousa:** Conceptualization; writing – original draft; writing – review and editing. **Sini Junntila:** Conceptualization; supervision; writing – review and editing. **Laura L Elo:** Conceptualization; funding acquisition; supervision; writing – review and editing.

CONFLICT OF INTEREST

The authors declare no conflict of interest.

REFERENCES

1. Park JE, Botting RA, Domínguez Conde C, *et al.* A cell atlas of human thymic development defines T cell repertoire formation. *Science* 2020; **367**: eaay3224.
2. Domínguez Conde C, Xu C, Jarvis LB, *et al.* Cross-tissue immune cell analysis reveals tissue-specific features in humans. *Science* 2022; **376**: eabl5197.
3. Rood JE, Maartens A, Hupalowska A, Teichmann SA, Regev A. Impact of the human cell atlas on medicine. *Nat Med* 2022; **28**: 2486–2496.
4. Panina Y, Karagiannis P, Kurtz A, Stacey GN, Fujibuchi W. Human cell atlas and cell-type authentication for regenerative medicine. *Exp Mol Med* 2020; **52**: 1443–1451.
5. Osadolor S, Solez K, Moghe I. Sun-476 the human cell atlas and artificial intelligence: kidney medicine in the digital age and beyond. *Kidney Int Rep* 2020; **5**: S391.
6. Mogilenko DA, Shchukina I, Artyomov MN. Immune ageing at single-cell resolution. *Nat Rev Immunol* 2022; **22**: 484–498.
7. Wang S, Sun ST, Zhang XY, *et al.* The evolution of single-cell RNA sequencing technology and application: progress and perspectives. *Int J Mol Sci* 2023; **24**: 2943.
8. Wang R, Peng G, Tam PPL, Jing N. Integration of computational analysis and spatial transcriptomics in single-cell studies. *Genomics Proteomics Bioinformatics* 2023; **21**: 13–23.

9. Subramanian I, Verma S, Kumar S, Jere A, Anamika K. Multi-omics data integration, interpretation, and its application. *Bioinform Biol Insights* 2020; **14**: 117793221989905.
10. Bressan D, Battistoni G, Hannon GJ. The dawn of spatial omics. *Science* 2023; **381**: eabq4964.
11. Hrovatin K, Sikkema L, Shitov VA, *et al.* Considerations for building and using integrated single-cell atlases. *Nat Methods* 2025; **22**: 41–57.
12. Andrabi SBA, Kalim UU, Palani S, *et al.* Long noncoding RNA LIRIL2R modulates FOXP3 levels and suppressive function of human CD4⁺ regulatory T cells by regulating IL2RA. *Proc Natl Acad Sci USA* 2024; **121**: e2315363121.
13. Ding J, Adiconis X, Simmons SK, *et al.* Systematic comparison of single-cell and single-nucleus RNA-sequencing methods. *Nat Biotechnol* 2020; **38**: 737–746.
14. Yamawaki TM, Lu DR, Ellwanger DC, *et al.* Systematic comparison of high-throughput single-cell RNA-seq methods for immune cell profiling. *BMC Genomics* 2021; **22**: 66.
15. Regan C, Preall J. Practical considerations for single-cell genomics. *Curr Protoc* 2022; **2**: e498.
16. Wolock SL, Lopez R, Klein AM. Scrublet: computational identification of cell doublets in single-cell transcriptomic data. *Cell Syst* 2019; **8**: 281–291.e9.
17. Germain PL, Lun A, Garcia Meixide C, Macnair W, Robinson MD. Doublet identification in single-cell sequencing data using scDblFinder. *F1000Res* 2021; **10**: 979.
18. Lopez R, Regier J, Cole MB, Jordan MI, Yosef N. Deep generative modeling for single-cell transcriptomics. *Nat Methods* 2018; **15**: 1053–1058.
19. Subramanian A, Alperovich M, Yang Y, Li B. Biology-inspired data-driven quality control for scientific discovery in single-cell transcriptomics. *Genome Biol* 2022; **23**: 267.
20. Xu J, Cui L, Zhuang J, *et al.* Evaluating the performance of dropout imputation and clustering methods for single-cell RNA sequencing data. *Comput Biol Med* 2022; **146**: 105697.
21. Qiu P. Embracing the dropouts in single-cell RNA-seq analysis. *Nat Commun* 2020; **11**: 1169.
22. Cuevas-Diaz Duran R, Wei H, Wu J. Data normalization for addressing the challenges in the analysis of single-cell transcriptomic datasets. *BMC Genomics* 2024; **25**: 444.
23. Yang P, Huang H, Liu C. Feature selection revisited in the single-cell era. *Genome Biol* 2021; **22**: 321.
24. Sun S, Zhu J, Ma Y, Zhou X. Accuracy, robustness and scalability of dimensionality reduction methods for single-cell RNA-seq analysis. *Genome Biol* 2019; **20**: 269.
25. Amezquita RA, Lun ATL, Becht E, *et al.* Orchestrating single-cell analysis with Bioconductor. *Nat Methods* 2020; **17**: 137–145.
26. Heumos L, Schaar AC, Lance C, *et al.* Best practices for single-cell analysis across modalities. *Nat Rev Genet* 2023; **24**: 550–572.
27. Nathan A, Baglaenko Y, Fonseka CY, Beynor JJ, Raychaudhuri S. Multimodal single-cell approaches shed light on T cell heterogeneity. *Curr Opin Immunol* 2019; **61**: 17–25.
28. Hao Y, Hao S, Andersen-Nissen E, *et al.* Integrated analysis of multimodal single-cell data. *Cell* 2021; **184**: 3573–3587.e29.
29. Caron DP, Specht WL, Chen D, *et al.* Multimodal hierarchical classification of CITE-seq data delineates immune cell states across lineages and tissues. *Cell Rep Methods* 2024; **5**: 100938.
30. Chu ND, Bi HS, Emerson RO, *et al.* Longitudinal immunosequencing in healthy people reveals persistent T cell receptors rich in highly public receptors. *BMC Immunol* 2019; **20**: 19.
31. Rusch M, Nakitandwe J, Shurtleff S, *et al.* Clinical cancer genomic profiling by three-platform sequencing of whole genome, whole exome and transcriptome. *Nat Commun* 2018; **9**: 3962.
32. Turajlic S, Xu H, Litchfield K, *et al.* Deterministic evolutionary trajectories influence primary tumor growth: TRACERx renal. *Cell* 2018; **173**: 595–610.e11.
33. Cai Y, Lu Z, Chen C, *et al.* An atlas of genetic effects on cellular composition of the tumor microenvironment. *Nat Immunol* 2024; **25**: 1959–1975.
34. Lim J, Chin V, Fairfax K, *et al.* Transitioning single-cell genomics into the clinic. *Nat Rev Genet* 2023; **24**: 573–584.
35. Henning AN, Roychoudhuri R, Restifo NP. Epigenetic control of CD8⁺ T cell differentiation. *Nat Rev Immunol* 2018; **18**: 340–356.
36. Satpathy AT, Granja JM, Yost KE, *et al.* Massively parallel single-cell chromatin landscapes of human immune cell development and intratumoral T cell exhaustion. *Nat Biotechnol* 2019; **37**: 925–936.
37. Wilfahrt D, Delgoffe GM. Metabolic waypoints during T cell differentiation. *Nat Immunol* 2024; **25**: 206–217.
38. Seydel C. Single-cell metabolomics hits its stride. *Nat Methods* 2021; **18**: 1452–1456.
39. Grosjean E, Simonneaux V, Challet E. Reciprocal interactions between circadian clocks, food intake, and energy metabolism. *Biology* 2023; **12**: 539.
40. Ali A, Davidson S, Fraenkel E, *et al.* Single cell metabolism: current and future trends. *Metabolomics* 2022; **18**: 77.
41. Palmer A, Phapale P, Chernyavsky I, *et al.* FDR-controlled metabolite annotation for high-resolution imaging mass spectrometry. *Nat Methods* 2017; **14**: 57–60.
42. METASPACE. The metabolite annotation engine and community-populated knowledgebase of spatial metabolomes. 2024 [cited March 9, 2025]. <https://metaspace2020.eu>.
43. Wadie B, Stuart L, Rath CM, Drotleff B, Mamedov S, Alexandrov T. METASPACE-ML: context-specific metabolite annotation for imaging mass spectrometry using machine learning. *Nat Commun* 2024; **15**: 9110.
44. Anderson NM, Simon MC. The tumor microenvironment. *Curr Biol* 2020; **30**: R921–R925.
45. Kleino I, Frolovaité P, Suomi T, Elo LL. Computational solutions for spatial transcriptomics. *Comput Struct Biotechnol J* 2022; **20**: 4870–4884.

46. Argelaguet R, Cuomo ASE, Stegle O, Marioni JC. Computational principles and challenges in single-cell data integration. *Nat Biotechnol* 2021; **39**: 1202–1215.
47. Zheng L, Qin S, Si W, et al. Pan-cancer single-cell landscape of tumor-infiltrating T cells. *Science* 2021; **374**: abe6474.
48. Xu C, Lopez R, Mehlman E, Regier J, Jordan MI, Yosef N. Probabilistic harmonization and annotation of single-cell transcriptomics data with deep generative models. *Mol Syst Biol* 2021; **17**: e9620.
49. Andreatta M, Hérault L, Gueguen P, Gfeller D, Berenstein AJ, Carmona SJ. Semi-supervised integration of single-cell transcriptomics data. *Nat Commun* 2024; **15**: 872.
50. Lotfollahi M, Naghipourfar M, Luecken MD, et al. Mapping single-cell data to reference atlases by transfer learning. *Nat Biotechnol* 2022; **40**: 121–130.
51. Sousa AGG, Smolander J, Junttila S, Elo LL. Coralys enables sensitive identification of imbalanced cell types and states in single-cell data via multi-level integration. 2025 [cited March 9, 2025]. <https://doi.org/10.1101/2025.02.07.637023>.
52. Lause J, Berens P, Kobak D. The art of seeing the elephant in the room: 2D embeddings of single-cell data do make sense. *PLoS Comput Biol* 2024; **20**: e1012403.
53. Lotfollahi M, Bright C, Skorobogat R, et al. ArchMap: a web-based platform for reference-based analysis of single-cell datasets. 2024 [cited March 9, 2025]. <https://doi.org/10.1101/2024.09.19.613883>.
54. Lütge A, Zypřych-Walczak J, Brykczynska Kunzmann U, et al. CellMixS: quantifying and visualizing batch effects in single-cell RNA-seq data. *Life Sci Alliance* 2021; **4**: e202001004.
55. Luecken MD, Büttner M, Chaichoompu K, et al. Benchmarking atlas-level data integration in single-cell genomics. *Nat Methods* 2022; **19**: 41–50.
56. Polański K, Young MD, Miao Z, Meyer KB, Teichmann SA, Park JE. BBKNN: fast batch alignment of single cell transcriptomes. *Bioinformatics* 2020; **36**: 964–965.
57. Sikkema L, Ramírez-Suástegui C, Strobl DC, et al. An integrated cell atlas of the lung in health and disease. *Nat Med* 2023; **29**: 1563–1577.
58. Filippov I, Schauser L, Peterson P. An integrated single-cell atlas of blood immune cells in aging. *NPJ Aging* 2024; **10**: 59.
59. Korsunsky I, Millard N, Fan J, et al. Fast, sensitive and accurate integration of single-cell data with harmony. *Nat Methods* 2019; **16**: 1289–1296.
60. Mallona I, Luetge A, Carrillo B, et al. Omnibenchmark (alpha) for continuous and open benchmarking in bioinformatics. 2024 [cited March 12, 2025]. <https://doi.org/10.48550/arXiv.2409.17038>
61. Luecken MD, Gigante S, Burkhardt DB, et al. Defining and benchmarking open problems in single-cell analysis. 2024 [cited March 12, 2025]. <https://doi.org/10.21203/rs.3.rs-4181617/v1>.
62. Pliner HA, Shendure J, Trapnell C. Supervised classification enables rapid annotation of cell atlases. *Nat Methods* 2019; **16**: 983–986.
63. Ianevski A, Giri AK, Aittokallio T. Fully-automated and ultra-fast cell-type identification using specific marker combinations from single-cell transcriptomic data. *Nat Commun* 2022; **13**: 1246.
64. Abdelaal T, Michielsen L, Cats D, et al. A comparison of automatic cell identification methods for single-cell RNA sequencing data. *Genome Biol* 2019; **20**: 194.
65. Aran D, Looney AP, Liu L, et al. Reference-based analysis of lung single-cell sequencing reveals a transitional profibrotic macrophage. *Nat Immunol* 2019; **20**: 163–172.
66. Osumi-Sutherland D, Xu C, Keays M, et al. Cell type ontologies of the human cell atlas. *Nat Cell Biol* 2021; **23**: 1129–1135.
67. Virshup I, Rybakov S, Theis FJ, Angerer P, Wolf FA. Anndata: access and store annotated datamatrices. *J Open Source Softw* 2024; **9**: 4371.
68. Wong MT, Ong DEH, Lim FSH, et al. A high-dimensional atlas of human T cell diversity reveals tissue-specific trafficking and cytokine signatures. *Immunity* 2016; **45**: 442–456.
69. Giles JR, Manne S, Freilich E, et al. Human epigenetic and transcriptional T cell differentiation atlas for identifying functional T cell-specific enhancers. *Immunity* 2022; **55**: 557–574.e7.
70. Terekhova M, Swain A, Bohacova P, et al. Single-cell atlas of healthy human blood unveils age-related loss of NKG2C⁺GZMB⁻CD8⁺ memory T cells and accumulation of type 2 memory T cells. *Immunity* 2023; **56**: 2836–2854.e9.
71. Oguchi A, Suzuki A, Komatsu S, et al. An atlas of transcribed enhancers across helper T cell diversity for decoding human diseases. *Science* 2024; **385**: eadd8394.
72. Li R, Strobl J, Poyner EFM, et al. Cutaneous T cell lymphoma atlas reveals malignant TH2 cells supported by a B cell-rich tumor microenvironment. *Nat Immunol* 2024; **25**: 2320–2330.
73. Xu J, Chen C, Sussman JH, et al. A multiomic atlas identifies a treatment-resistant, bone marrow progenitor-like cell population in T cell acute lymphoblastic leukemia. *Nat Cancer* 2024; **6**: 102–122.
74. Bai Z, Feng B, McClory SE, et al. Single-cell CAR T atlas reveals type 2 function in 8-year leukaemia remission. *Nature* 2024; **634**: 702–711.
75. Li X, Henderson J, Gordon MJ, et al. A single-cell atlas of CD19 chimeric antigen receptor T cells. *Cancer Cell* 2023; **41**: 1835–1837.
76. Chu Y, Dai E, Li Y, et al. Pan-cancer T cell atlas links a cellular stress response state to immunotherapy resistance. *Nat Med* 2023; **29**: 1550–1562.
77. Ding S, Qiao N, Zhu Q, et al. Single-cell atlas reveals a distinct immune profile fostered by T cell-B cell crosstalk in triple negative breast cancer. *Cancer Commun* 2023; **43**: 661–684.
78. Liang Y, Tan Y, Guan B, et al. Single-cell atlases link macrophages and CD8⁺ T-cell subpopulations to disease progression and immunotherapy response in urothelial carcinoma. *Theranostics* 2022; **12**: 7745–7759.

79. Sun K, Xu R, Ma F, *et al.* scRNA-seq of gastric tumor shows complex intercellular interaction with an alternative T cell exhaustion trajectory. *Nat Commun* 2022; **13**: 4943.
80. Lu Y, Yang A, Quan C, *et al.* A single-cell atlas of the multicellular ecosystem of primary and metastatic hepatocellular carcinoma. *Nat Commun* 2022; **13**: 4594.
81. Jiang C, Chen J, Sun T, Xu J, Zhu H, Chen J. Unveiling dynamic changes and regulatory mechanisms of T cell subsets in sepsis pathogenesis. *Immunotargets Ther* 2024; **13**: 29–44.
82. Wang S, Zhang Q, Hui H, Agrawal K, Karris MAY, Rana TM. An atlas of immune cell exhaustion in HIV-infected individuals revealed by single-cell transcriptomics. *Emerg Microbes Infect* 2020; **9**: 2333–2347.
83. Mullan KA, Ha M, Valkiers S, *et al.* T cell receptor–centric perspective to multimodal single-cell data analysis. *Sci Adv* 2024; **10**: eadr3196.
84. Fiskin E, Eraslan G, Alora-Palli MB, *et al.* Multi-modal skin atlas identifies a multicellular immune-stromal community associated with altered cornification and specific T cell expansion in atopic dermatitis. 2023 [cited March 12, 2025]. <https://doi.org/10.1101/2023.10.29.563503>.
85. Poch T, Krause J, Casar C, *et al.* Single-cell atlas of hepatic T cells reveals expansion of liver-resident naive-like CD4⁺ T cells in primary sclerosing cholangitis. *J Hepatol* 2021; **75**: 414–423.
86. Corridoni D, Antanaviciute A, Gupta T, *et al.* Single-cell atlas of colonic CD8⁺ T cells in ulcerative colitis. *Nat Med* 2020; **26**: 1480–1490.
87. Rood JE, Wynne S, Robson L, *et al.* The human cell atlas from a cell census to a unified foundation model. *Nature* 2025; **637**: 1065–1071.
88. Ma Q, Xu D. Deep learning shapes single-cell data analysis. *Nat Rev Mol Cell Biol* 2022; **23**: 303–304.
89. Tsabar M, Lahav G. The single-cell yin and Yang of live imaging and transcriptomics. *Cell Syst* 2017; **4**: 375–377.
90. Huang S, Dai R, Zhang Z, *et al.* CRISPR/Cas-based techniques for live-cell imaging and bioanalysis. *Int J Mol Sci* 2023; **24**: 13447.
91. May D, Yun S, Gonzalez DG, *et al.* Live imaging reveals chromatin compaction transitions and dynamic transcriptional bursting during stem cell differentiation in vivo. *eLife* 2023; **12**: e83444.

© 2025 The Author(s). Immunology & Cell Biology published by John Wiley & Sons Australia, Ltd on behalf of the Australian and New Zealand Society for Immunology, Inc.

This is an open access article under the terms of the [Creative Commons Attribution](#) License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited.