Available online at www.sciencedirect.com

ScienceDirect

Current Opinion in Immunology

CD4+ teff cell heterogeneity: the perspective from single-cell transcriptomics

David Zemmour[1], Evgeny Kiner[1,2] and Christophe Benoist

Single-cell transcriptomics (scRNAseq) holds the promise to generate definitive atlases of cell types. We review scRNAseq studies of conventional CD4+ αβ T cells performed in a variety of challenged contexts (infection, tumor, allergy) that aimed to parse the complexity and representativity of previously defined CD4+ T cell types, lineages, and cosmologies. With a few years' experience, the field has realized the difficulties and pitfalls of scRNAseq. With the very high-dimensionality of scRNAseq data, subset definitions based on low-dimensionality marker combinations tend to fade or blur: cell types prove more complex than expected; transcripts of key defining transcripts (cytokines, chemokines) are distributed as broad and partially overlapping continua; boundaries with innate lymphocytes are blurred. Tissue location and activation, either cytokine-driven or TCR-driven, determine Teff heterogeneity in sometimes unexpected ways. Emerging techniques for lineage and trajectory tracing, and RNA-protein connections, will further help define the space of differentiated CD4+ T cell heterogeneity.

Address
Department of Immunology, Harvard Medical School, and Evergrande Center for Immunologic Diseases, Harvard Medical School and Brigham and Women's Hospital, Boston, MA 02115, USA

Corresponding author: Benoist, Christophe (cbdm@hms.harvard.edu)
[1] These authors contributed equivalently.
[2] Present address: Immunai, 180 Varick Street, NYC, NY 10014, USA.

Current Opinion in Immunology 2020, **63**:61–67

This review comes from a themed issue on **Lymphocyte development and activation**

Edited by **Claude-Agnès Reynaud** and **Stephen M Hedrick**

For a complete overview see the Issue and the Editorial

Available online 4th April 2020

https://doi.org/10.1016/j.coi.2020.02.004

Introduction

CD4+ T cells initiate and orchestrate immune responses to a wide array of pathogens by detecting non-self antigens, help humoral and cellular responses, orchestrate tissue repair, and have even more esoteric effects on the homeostasis of organ systems outside the classic confines of Immunology. This range of functions and response abilities has long raised the question of their heterogeneity: are diverse functions all mediated by one multifunctional cell, or do distinct subsets possess specific effector functions, like helping B cell responses or provoking delayed-type hypersensitivity. Single-cell transcriptomics holds the promise to reopen this long-standing question from a radically different perspective and to provide a genome-scale parsing of true identities in the CD4+ T cell subsets.

Almost half a century ago, it was shown that different T cell clones can have different functions, with a division of labor between cells responsible for cell-mediated immunity and for help to antibody responses [1,2]. Mosmann *et al.* [3] were the first to classify different CD4+ T cell clones by the cytokines they produce, a fundamental notion that is followed to this day, and spread to classifications of other immunocytes such as innate lymphoid cells (ILCs), gamma delta T cells or non-conventional T cells. They analyzed independent T cell clones in culture and noted that their cytokine secretion profiles could distinguish them in relation to their functional ability [3]: Th1 cells responsible for cellular immunity against intracellular pathogens, Th2 cells that orchestrated the antibody response by B cells [4]. It was soon realized that naïve T cells could also be induced to adopt these biased cytokine-producing phenotypes by culture with cocktails of inducing cytokines [5].

The CD4+ T effector (Teff) cosmology has since expanded beyond the Th1/Th2 dichotomy, which was quickly found to be too constraining, with identification of Th17, Tfh, Th9, or Th22 (and further 'sub-subsets' thereof, like Th17*, Tfh2, etc), and with the reallocation of several functions (i.e. Tfh are now seen as the main B cell helpers). These distinctions remain primarily based on cytokine but now bolstered by 'master' transcription factors which are thought to anchor these programs [6]. Moreover, sub-subsets such as pathogenic and non-pathogenic or homeostatic and inflammatory Th17s have been identified [7–9]. Conventional gene expression profiling of *in vitro* polarized T cell cultures was used to identify differentially expressed signatures [10–14], or differential epigenetic marks and chromatin states [15,16], but usually from cells derived in biasing tissue culture, rather than from *ex-vivo* cell populations. Importantly, it has proven impossible to identify cell surface markers that reliably and consistently identified Th-x subsets (some chemokine receptors have been proposed, but proved not to have unequivocal expression).

In hindsight, was there an obligate reason why cytokine expression had to be the organizing principle for $CD4^+$ T subset distinction? Cytokines are certainly essential to T cell function, but other functionally important molecules might have served. Indeed, an orthogonal frame of reference for Teff heterogeneity was developed, based primarily on organismal location and the expression of chemokine receptors and adhesion molecules that drive it [17,18], distinctions that have direct relevance to the rapid detection of pathogens or inflammation states in the tissue where they occur. As for $CD8^+$ T cells, $CD4^+$ T cells have been distinguished as T_N, T_{CM}, T_{EM}, and T_{RM} (but also less frequently used categories like T_{EMRA}, T_{PM}, and so on [19••], a classification orthogonal to the Th-x nomenclature and that also relies on very few (low dimensionality) flags.

Against this background, several groups have begun to apply single-cell (sc) RNAseq to probe T cell populations more deeply in the context of various challenges, to chart the relationships between subsets, to search for unrecognized heterogeneity within subsets, or perchance to discover entirely new planets in the $CD4^+$ system. We will focus here on effector CD4 T cells but occasionally refer to regulatory T cells and $CD8^+$ T cells for comparisons or parallels.

Caveats and pitfalls

Several good recent reviews have discussed the revolutionary potential of single-cell RNAseq in parsing the complexity of cell-types and cell-states [20–22]. Overcoming the averaging that occurs during gene expression profiling of sorted cells, scRNAseq has the potential to reveal hidden heterogeneity in cell populations, characterize distinct cell subsets, and map the trajectories of cell differentiation. So much exciting is on promise.

The field having a few years' experience, it is important to avoid rose-colored glasses and be conscious of the caveats and pitfalls of scRNAseq. Not so much the question of its sensitivity, as one still hears occasionally: low abundance transcripts are indeed difficult to analyze robustly in individual cells, owing to the high probability of technical dropouts as well as biological noise in the low expression range. But the scRNAseq process (in its different variants) is actually remarkably faithful: aggregation of cells into microclusters, and/or imputation of missing values [23••,24–26], yields profiles akin to conventional RNA-seq, showing that rare transcripts are correctly detected.

A key obstacle is that scRNAseq data are simply difficult to grasp and handle: defining a cell-type based on a multidimensional transcriptome is more difficult, conceptually and experimentally, than from a handful of flow cytometry markers. Computational pipelines have been developed to process, verify, and analyze scRNAseq data and are well codified by now (e.g. Seurat [27••]). But the follow-up, the interpretation of the results is far less codified and can be misleading. For instance, 2-dimensional projections (tSNE, UMAP) are commonly used to summarize and make human-readable multidimensional complexity. But those maps are by no means to be taken as gold-standard representations of the reality (play with https://distill.pub/2016/misread-tsne/ for sobering examples). Similarly, partition clustering tools produce groupings (e.g. the colored outputs of common processing pipelines) that may, or may not, have anything to do with biological reality. Clustering tools accurately distinguish distant cells, neurons from hepatocytes, B from T cells, but are far less trustworthy when trying to parse closely related cells — one should remember that a clustering algorithm will *always* return clusters (conditioned by its set parameters), even when there are not any biologically compelling or rational ones [28]. Inferring precursor-product relationships between cells using the early 'trajectory' tools can be hazardous (fine to map intermediates in a transition sequence that is known beforehand, haphazard if trying to infer an unknown relationship) [29••]. Annotation of cell types (regions in a 2D projection, or clusters) is surprisingly challenging; quick and superficial annotation based on recognizing a handful of favorite transcripts is too often used, but rigorous annotation based on defining signatures is much more challenging [30]. Finally, the common experience is that distinction between cells that one would imagine should be readily distinguished (e.g. NK and T cells) are often not straightforward: T cells and ILCs differ fundamentally by the rearrangement and expression of TCR genes, but much less when the rest of the functional transcriptome is taken into account. Whether a cell's identity should be defined by one or two key functional genes that deeply condition the cell's function, or by the fingerprint of its entire transcriptome, is a question that the field is still grappling with.

Did scRNAseq reveal novel $CD4^+$ T cell-types: 'Tis'

Discovering new cell-types is a long-established tradition in Immunology, and a novel $CD4^+$ T cell would be a valuable trophy. In other domains, scRNAseq has allowed the discovery of rare but important cell types, primarily epithelial: CFTR-expressing ionocytes in the lung surface, tuft-like cells in the thymic epithelium [31–33]. For $CD4^+$ T cells, there may be one such revelation: in several scRNAseq studies, a minor but very distinct population of $CD4^+$ T cells showed an inordinately high expression of Interferon-stimulated Genes (ISGs), with no other notable characteristic. These cells (hereafter 'Tis') (for T interferon signature) were observed in a range of immunologic challenges: in airways after allergic sensitization with house dust mites or *Alternaria* extract [34••,35••], in kidney infiltrates of lupus nephritis [36••], or in colon lamina propria of *Salmonella*-infected or *Citrobacter*-infected mice [37]. With the exception of the lupus

kidney, these are not contexts where Type-1 interferon would be expected. Tis were rare or absent in lymphoid tissues at baseline; however, indicating that they are induced or summoned during active immune defenses. The origin of their high ISG expression is mysterious. Their Interferon (IFN) signaling pathway may be unusually sensitive; Tibbitt *et al.* [34••] showed that the Type-I IFN receptor is not overexpressed in Tis, but other signaling molecules might be involved (e.g. MyD88). Alternatively, Tis cells may reside in micro-compartments that contain high levels of IFN, a reaction that can occur in any tissue. Interestingly, though, a similar population was also seen observed in thymic CD4⁺ single-positive, suggesting that Tis may differentiate as an independent branch in the thymus [38]. Finally, the interferon-activated transcriptional module may have been activated in Tis but never switched off. Tis function is totally unknown (and may prove difficult to explore, given their characteristics and rarity), but one might speculate that constitutively high ISG expression may have pathological consequences (e.g. participating in lupus pathogenesis?).

Discrete subsets or a continuum of states?

Several studies took the unbiased approach of analyzing the entire pool of CD4⁺ T cells in a given location, in the context of a tumor or some immune challenge [34••,36••,37,39,40••,41••]. A strong effect of location was commonly observed, an anticipated conclusion based on much conventional profiling.

Ciucci *et al.* [41••] performed scRNAseq on splenic activated (CD44^{hi}) T cells, 7 days after LCMV infection, polyclonally or after tetramer selection of cells reacting to the immunodominant GP66 peptide. Although polyclonal splenocytes appear to split into clusters according to conventional Th1/Tfh/Tmem phenotypes, the GP66-specific population proved more continuous, with few or no well-defined and reproducible clusters (perhaps an illustration of issues with default clustering and coloring), and with overlapping representations of transcripts classically used to identify T subsets (*Tbx21*, *Bcl6*, *Ccr7*). In a study of activated CD4⁺ T cells from bronchoalveolar lavage after house dust mite re-challenge [34••], clustering only poorly resolved cell subsets (other than Tregs and Tis). Cells producing IL13 and IL5 were predominantly found in one corner of the tSNE plot, but the expression of *Gata3* (highly represented) or *Tbx21* (less common) was spread among the clusters. Similarly, in a study of colon lamina propria T cells after challenge with a panel of bacteria or parasites known to induce Th1, 17 or 2 biased responses, Kiner *et al.* [37] found that Teff could not be parsed into discrete Th cell clusters, even with pliable artificial intelligence tools, cytokine and other signature transcripts being distributed along biased but continuous gradients. Variance in expression was tied to that infecting pathogen, more than to the cytokines produced or to any classic Th signature. Thus, in these several conditions where pathogenic challenges might have been expected to elicit distinct Th-x subsets, more complex populations of responding CD4⁺ T cells were observed [34••,37,41••]; cell distributions were biased in accordance with the challenge, but with far 'messier' representations of key transcripts, and no distinct cell-types.

Several studies undertook single-cell analyses in tumors in order to study tumor heterogeneity as well as the heterogeneity of the immune response to tumors. Zheng *et al.* [39] profiled single T cells from hepatocellular carcinoma patients and found several clusters among CD4⁺ T cells, including naive CD4s, Tregs, Teffs, exhausted T cells, as well as cytotoxic CD4⁺ T cells. As expected from conventional profiling, they observed a strong tissue effect that differentiated the transcriptomes of peripheral blood T cells and tumor-infiltrating T cells. Azizi *et al.* [40••] profiled total CD45⁺ cells from breast tumor patients and described 21 CD4⁺ clusters, which were split into 9 naive, 7 central memory, 15 effector memory, and 5 Treg clusters (overclustered?). Tumor-specific signatures were not found in discrete cell state clusters, but rather along 'gradual trends of variation', or a single continuous trajectory.

Other scRNAseq studies drilled down into subsets of T cells, identified by knockin reporters driven by cytokine-encoding genes and/or marker combinations, aiming to uncover further heterogeneity within CD4⁺ T cell subtypes. In a pioneering study, Gaublomme *et al.* [42••] used scRNAseq on IL17-producing cells derived *in vitro* and from the central nervous system of EAE-bearing mice. This allowed for the unbiased characterization of cell-states within Th17 cells, which had been previously described to vary phenotypically, with regard to their exposure to IL23 and on their pathogenicity [8]. Indeed, variation within Th17s was observed, *in vivo* and *in vitro*, but as a phenotypic continuum rather than discrete subsubsets. The variation could be decomposed into several components: overall activation; a module containing classic Th1 transcripts including *Ifng*, and one linked to the IL23 signature or pathogenicity. These results were paralleled by IL17-producing CD8⁺ T cells in the skin, after challenge with *Staphylococcus epidermidis* [43] where a significant proportion of the cells could also co-express type 2 cytokines. Two other studies similarly revealed heterogeneity within IL10-producing CD4⁺ T cells in anti-CD3 treated or LCMV infected mice [44,45]. Xin *et al.* showed a dominant population of IL10-producing cells with several characteristics related to Th1 cells (*Tbx21*, *Id2*) but also an outlier cluster with classic Tfh transcripts (*Il21*, *Cxcr5*, *Bcl6*). Building on the scRNAseq data, elegantly constructed mixed bone-marrow chimeras were then used to show the functional relevance of these IL10-producing Tfh in the humoral response to LCMV.

These studies, focused on cells making an individual cytokine, illustrated that cytokine production does not result from a single and discrete cell population.

Gowthaman *et al.* [35••] took apart the Tfh subset (strenuously purified as $CD44^{hi}PD1^{hi}CXCR5^{hi}$) elicited in an intranasal model of allergic sensitization. What might have been expected to be a well-defined group of cells actually contained a number of distinct cell-types: some Tis cells (confirming that the Tis phenotype cuts across marker-identified subsets), groups of cells superficially annotated as IL4-producing Tfh2, and a small but distinct group of cells with dual production of IL4 and IL13 (follow-up experiments showed that IL13+ cells were indeed essential for allergic sensitization and anaphylactic IgE production; other allergy models may use other phenotypes for Type 2 cytokine production, however). In a similar vein, Patil *et al.* [46] sorted classic fractions of effector/memory cells (T_{CM}, T_{EM}, T_{EMRA}) from human donors originating from different continents, also based on low dimensionality flow cytometry criteria (CD45RA and CCR7). The T_{EMRA} pool proved to contain several different clusters, although it was not clear how much of the clustering was driven by inter-donor differences versus true cell-states. Some of these clusters were hypothesized to represent effector or precursor pools, based on the differential representation of some homeostatic or precursor transcripts, although these relationships remain to be substantiated experimentally. Thus, both these studies illustrate that previously defined cell-types that rest on limited marker combinations actually hide far more extensive complexity when analyzed in the multidimensional transcript space.

Taken together, these results imply that Teff states are non-discrete, vary between tissues of residence, and are strongly influenced by cytokine-induced or TCR-induced activating signals. The variation in gene expression between T cells distributes as broad continua, rather than as discrete entities. Conversely, populations that were thought to be well-defined turned out to be more complex than anticipated. We have proposed elsewhere [37] a 'Cloud Model' for Teff cell heterogeneity, wherein cytokines or other key functional or regulatory molecules are expressed along partially overlapping gradients, rather than in discrete Th-x entities. This concept is different from the 'plasticity' that has been often evoked over the last decade, in that plasticity implies defined states or lineages between which T cells could switch. The Cloud Model raises conceptual difficulties. It is easier to think of interactions between well-defined and consistent cell types than between proteoform cell populations, in which the expression of any one effector or regulatory molecule is loosely distributed. It raises practical difficulties for experimentation: for instance, it was comfortable to sort CCR6+ cells to test their activity and properties, under the assumption that they represented a coherent functional cell entity. But how does one sort a segment of a cloud?

Do variations in TCR clonotypes drive $CD4^+$ T heterogeneity?

T cells see the world through their TCR. Other factors (inductive influences in specific tissue locations, microbe-derived ligands, cytokines, etc) can sway T cell phenotypes, but a key question is how much of the phenotypic choice rests in the characteristics of signals caused by TCR engagement (affinity, avidity, duration). ScRNA-seq, and in particular its recent advances [40••,47••,48], allows the determination of TCR sequences (both chains) while evaluating the cells' transcriptomes. Several scRNAseq studies have addressed these questions, in localized settings (tumors, infection, immunization) where clonotype expansion allows the comparison of several cells that express the same TCR (noting that this focus on expanded clones may introduce an experimental bias).

The overall conclusion is that the TCR does influence phenotypes, but to varying degrees according to location. Very demonstrative were the results of Azizi *et al.* in breast cancer infiltrates, where cells expressing the same αβ TCR clonotype tended to occupy confined regions of the tSNE projection, although to variable extent for different TCRs [40••,49••]. A multiple regression procedure estimated that 30–50% of the T cell diversity could be explained by TCR clonotype. Similarly, among ~3000 $CD4^+$ T cells infiltrating basal cell carcinoma tumors, Yost *et al.* found no clonotypic overlap between cell regions annotated as Treg, Tfh and Th17, contrary to some CD8 subsets that transitioned from memory or effector to activated phenotypes [50••]. In contrast, Pratama *et al.* [51] reported some sharing of TCR clonotypes between Treg and Tconv pools in the colonic lamina propria of ex-germ-free mice colonized with a single microbe. Several of the highly represented clones evidenced by Tu *et al.* [48] after immunization with HPV-E7 were also spread throughout the phenotypic space. Overall, these studies show that the TCR is not the only driver of T cell heterogeneity but definitely an important factor.

Mechanistically, TCR signals may control the transcriptional programs directly or indirectly. Early TCR signaling is almost always the most variable gene module in single-cell surveys of T cell populations [42••,44,45,49••,52,53]. Zemmour *et al.* [49••] tested the connection between signaling intensity and phenotypes in Tregs, using Nr4a1-gfp reporter mice in which the reporter intensity was directly related to the strength of TCR signals [54]. Cells belonging to different windows of GFP intensity were analyzed by scRNAseq. GFP levels proved highly related to the position of the cells in the multidimensional phenotypic space. These results suggest that TCR signals are intrinsically

related to phenotypes. In contrast, the same range of phenotypes was found in Treg cells of the adipose tissue, whether they were polyclonal or expressed a single transgenic TCR [53], showing that extrinsic cues can dominate in some contexts.

What lies ahead?

Single-cell transcriptomics is a recent and very fast-moving technology, and novel approaches to study T cells at a single-cell level are continuously emerging. 'CITE-Seq' uses DNA-tagged antibodies that recognize molecules on the cell surface, and the tags are then quantitated in the same microfluidic process as mRNAs [47••,55,56]. This allows the simultaneous quantitation of surface proteins and mRNAs and can be scaled to tens or even hundreds of cell surface proteins. CITE-seq promises to be invaluable for T cell biology, connecting complex structures in the transcriptome with the surface markers that can be used to select cells for functional experiments. Along the same lines, tagging reagents that enable input multiplexing ('hash-tags') can be used for creative experimental designs, combining controls with experimental variants, time-shifted or dose-shifted points, to enable direct comparisons and alleviate dreaded batch effects [37,47••,55]. Single-cell analysis of chromatin accessibility is also coming of age [57], with new versions that associate chromatin and transcriptome data on the same cells. T cell differentiation and heterogeneity can be probed at the more 'primordial' chromatin level, free of confounders from variation in RNA stability or transcriptional bursting. Finally, spatially resolved transcriptomics (e.g. Refs. [58,59]) will shed light on the locations within organs where T cells of particular phenotypes live (building here on an already extensive knowledge from multi-color imaging), and on transcriptional changes happening in T cells *in situ*. One should acknowledge that spatial transcriptomics may be of more value to study immobile cells (neurons) than for T cells that are highly mobile, constantly probing other cells in search of excitement. Finally, new experimental supports for cell and lineage tracing (e.g. Refs. [60••,61,62,63]), and more performant computational tools [29••] should build on early explorations [64] and allow a robust analysis of differentiation trajectories, what paths and fluctuations are followed by a T cell during activation, holding a key for understanding the molecular underpinnings of T cell differentiation and immune or autoimmune responses.

Conflict of interest statement

Nothing declared.

Acknowledgements

We thank Drs N. Hacohen. S. Eisenbarth, G. LeGros, V. Pandurangan and P. Czarnewsk for discussions. Work in the authors' laboratory on this topic was supported by grants from the National Institutes of HealthAI150686, AI116834 and AI125603 to CB, and AI072073 to the ImmGen consortium. DZ and EK were supported by a PhD fellowship from the Boehringer Ingelheim Fonds. All authors declare no competing interest.

References and recommended reading

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

- •• of outstanding interest

1. Liew FY, Parish CR: **Lack of a correlation between cell-mediated immunity to the carrier and the carrier-hapten helper effect**. *J Exp Med* 1974, **139**:779-784.
2. Araneo BA, Marrack PC, Kappler JW: **Functional heterogeneity among the T-derived lymphocytes of the mouse. II. Sensitivity of subpopulations to anti-thymocyte serum**. *J Immunol* 1975, **114**:747-751.
3. Mosmann TR, Cherwinski H, Bond MW, Giedlin MA, Coffman RL: **Two types of murine helper T cell clone. I. Definition according to profiles of lymphokine activities and secreted proteins**. *J Immunol* 1986, **136**:2348-2357.
4. Bottomly K: **A functional dichotomy in CD4+ T lymphocytes**. *Immunol Today* 1988, **9**:268-274.
5. Seder RA, Paul WE: **Acquisition of lymphokine-producing phenotype by CD4+ T cells**. *Annu Rev Immunol* 1994, **12**:635-673.
6. Zhu J, Paul WE: **CD4 T cells: fates, functions, and faults**. *Blood* 2008, **112**:1557-1569.
7. Ghoreschi K, Laurence A, Yang X-P, Tato CM, McGeachy MJ, Konkel JE, Ramos HL, Wei L, Davidson TS, Bouladoux N *et al.*: **Generation of pathogenic T(H)17 cells in the absence of TGF-β signalling**. *Nature* 2010, **467**:967-971.
8. Lee Y, Awasthi A, Yosef N, Quintana FJ, Xiao S, Peters A, Wu C, Kleinewietfeld M, Kunder S, Hafler DA *et al.*: **Induction and molecular signature of pathogenic TH17 cells**. *Nat Immunol* 2012, **13**:991-999.
9. Omenetti S, Bussi C, Metidji A, Iseppon A, Lee S, Tolaini M, Li Y, Kelly G, Chakravarty P, Shoaie S *et al.*: **The intestine harbors functionally distinct homeostatic tissue-resident and inflammatory Th17 cells**. *Immunity* 2019, **51**:77-89.e6.
10. Lu B, Zagouras P, Fischer JE, Lu J, Li B, Flavell RA: **Kinetic analysis of genomewide gene expression reveals molecule circuitries that control T cell activation and Th1/2 differentiation**. *Proc Natl Acad Sci U S A* 2004, **101**:3023-3028.
11. Rogge L, Bianchi E, Biffi M, Bono E, Chang SY, Alexander H, Santini C, Ferrari G, Sinigaglia L, Seiler M *et al.*: **Transcript imaging of the development of human T helper cells using oligonucleotide arrays**. *Nat Genet* 2000, **25**:96-101.
12. Chtanova T, Kemp RA, Sutherland AP, Ronchese F, Mackay CR: **Gene microarrays reveal extensive differential gene expression in both CD4(+) and CD8(+) type 1 and type 2 T cells**. *J Immunol* 2001, **167**:3057-3063.
13. Yosef N, Shalek AK, Gaublomme JT, Jin H, Lee Y, Awasthi A, Wu C, Karwacz K, Xiao S, Jorgolli M *et al.*: **Dynamic regulatory network controlling TH17 cell differentiation**. *Nature* 2013, **496**:461-468.
14. Ciofani M, Madar A, Galan C, Sellars M, Mace K, Pauli F, Agarwal A, Huang W, Parkhurst CN, Muratet M *et al.*: **A validated regulatory network for Th17 cell specification**. *Cell* 2012, **151**:289-303.
15. Wei G, Wei L, Zhu J, Zang C, Hu-Li J, Yao Z, Cui K, Kanno Y, Roh T-Y, Watford WT *et al.*: **Global mapping of H3K4me3 and H3K27me3 reveals specificity and plasticity in lineage fate determination of differentiating CD4+ T cells**. *Immunity* 2009, **30**:155-167.
16. Miraldi ER, Pokrovskii M, Watters A, Castro DM, De Veaux N, Hall JA, Lee J-Y, Ciofani M, Madar A, Carriero N *et al.*: **Leveraging chromatin accessibility for transcriptional regulatory network inference in T Helper 17 Cells**. *Genome Res* 2019, **29**:449-463.
17. Masopust D, Vezys V, Marzo AL, Lefrançois L: **Preferential localization of effector memory cells in nonlymphoid tissue**. *Science* 2001, **291**:2413-2417.

18. Sallusto F, Lenig D, Förster R, Lipp M, Lanzavecchia A: **Two subsets of memory T lymphocytes with distinct homing potentials and effector functions**. *Nature* 1999, **401**:708-712.

19. •• Jameson SC, Masopust D: **Understanding subset diversity in T cell memory**. *Immunity* 2018, **48**:214-226
An interesting discussion of phenotypic variation with memory T cells, and the limits of sub-subsetology.

20. Stubbington MJT, Rozenblatt-Rosen O, Regev A, Teichmann SA: **Single-cell transcriptomics to explore the immune system in health and disease**. *Science* 2017, **358**:58-63.

21. Papalexi E, Satija R: **Single-cell RNA sequencing to explore immune cell heterogeneity**. *Nat Rev Immunol* 2018, **18**:35-45.

22. Yost KE, Chang HY, Satpathy AT: **Tracking the immune response with single-cell genomics**. *Vaccine* 2019 http://dx.doi.org/10.1016/j.vaccine.2019.11.035.

23. •• Kharchenko PV, Silberstein L, Scadden DT: **Bayesian approach to single-cell differential expression analysis**. *Nat Methods* 2014, **11**:740-742
An early and thorough bayesian statistical analysis of scRNAseq data modeling missing data for differential expression analysis.

24. Hou W, Ji Z, Ji H, Hicks SC: **A systematic evaluation of single-cell RNA-sequencing imputation methods**. *bioRxiv* 2020 http://dx.doi.org/10.1101/2020.01.29.925974.

25. Huang M, Wang J, Torre E, Dueck H, Shaffer S, Bonasio R, Murray JI, Raj A, Li M, Zhang NR: **SAVER: gene expression recovery for single-cell RNA sequencing**. *Nat Methods* 2018, **15**:539-542.

26. van Dijk D, Sharma R, Nainys J, Yim K, Kathail P, Carr AJ, Burdziak C, Moon KR, Chaffer CL, Pattabiraman D *et al.*: **Recovering gene interactions from single-cell data using data diffusion**. *Cell* 2018, **174**:716-729.

27. •• Stuart T, Butler A, Hoffman P, Hafemeister C, Papalexi E, Mauck WM 3rd, Hao Y, Stoeckius M, Smibert P, Satija R: **Comprehensive integration of single-cell data**. *Cell* 2019, **177**:1888-1902.e21
A powerful and widely used pipeline to perform common bioinformatic analyses of scRNAseq data, for example, differential gene expression, clustering, and dimensionality reduction (tSNE/UMAP).

28. Kiselev VY, Andrews TS, Hemberg M: **Challenges in unsupervised clustering of single-cell RNA-seq data**. *Nat Rev Genet* 2019, **20**:273-282.

29. •• Saelens W, Cannoodt R, Todorov H, Saeys Y: **A comparison of single-cell trajectory inference methods**. *Nat Biotechnol* 2019, **37**:547-554
The authors provide guidelines for choosing trajectory methods based on a systematic comparison of more than 70 algorithms.

30. Aran D, Looney AP, Liu L, Wu E, Fong V, Hsu A, Chak S, Naikawadi RP, Wolters PJ, Abate AR *et al.*: **Reference-based analysis of lung single-cell sequencing reveals a transitional profibrotic macrophage**. *Nat Immunol* 2019, **20**:163-172.

31. Bornstein C, Nevo S, Giladi A, Kadouri N, Pouzolles M, Gerbe F, David E, Machado A, Chuprin A, Tóth B *et al.*: **Single-cell mapping of the thymic stroma identifies IL-25-producing tuft epithelial cells**. *Nature* 2018, **559**:622-626.

32. Montoro DT, Haber AL, Biton M, Vinarsky V, Lin B, Birket SE, Yuan F, Chen S, Leung HM, Villoria J *et al.*: **A revised airway epithelial hierarchy includes CFTR-expressing ionocytes**. *Nature* 2018, **560**:319-324.

33. Miller CN, Proekt I, von Moltke J, Wells KL, Rajpurkar AR, Wang H, Rattay K, Khan IS, Metzger TC, Pollack JL *et al.*: **Thymic tuft cells promote an IL-4-enriched medulla and shape thymocyte development**. *Nature* 2018, **559**:627-631.

34. •• Tibbitt CA, Stark JM, Martens L, Ma J, Mold JE, Deswarte K, Oliynyk G, Feng X, Lambrecht BN, De Bleser P *et al.*: **Single-cell RNA sequencing of the T helper cell response to house dust mites defines a distinct gene expression signature in airway Th2 cells**. *Immunity* 2019, **51**:169-184
Parsing of Type2 responses after house dust mite sensitization, with the deepest analysis of the origin of the novel Tis cell-type.

35. •• Gowthaman U, Chen JS, Zhang B, Flynn WF, Lu Y, Song W, Joseph J, Gertie JA, Xu L, Collet MA *et al.*: **Identification of a T follicular helper cell subset that drives anaphylactic IgE**. *Science* 2019, **365**
scRNAseq analysis of mouse Tfh ($CD44^{hi}PD1^{hi}CXCR5^{hi}$) after allergic sensitization with *Alternaria* extract, revealing a particular subpopulation, perhaps specific to this model, of high IL4/IL13 producers. A nice illustration of functional follow-up to the scRNAseq discovery.

36. •• Arazi A, Rao DA, Berthier CC, Davidson A, Liu Y, Hoover PJ, Chicoine A, Eisenhaure TM, Jonsson AH, Li S *et al.*: **The immune cell landscape in kidneys of patients with lupus nephritis**. *Nat Immunol* 2019, **20**:902-914
A large cohort analysis of immunocytes in an autoimmune context from the AMP consortium. Tis cells are sharply distinguishable, but other $CD4^+$ T cells do not cluster according to classic subsets (no sign of Th1 or Th2, putative Tfh tend to segregate with Tregs).

37. Evgeny Kiner, Elijah Willie, Brinda Vijaykumar, Hugo Schmutz, Jodie Chandler, Graham LeGros, Sara Mostafavi, Diane Mathis, Christophe Benoist, and the Immunological Genome Project Consortium: **Gut CD4+ T cell phenotypes: a continuum molded by microbes, not by Th archetypes**. *Submitted* 2020.

38. Hemmers S, Schizas M, Azizi E, Dikiy S, Zhong Y, Feng Y, Altan-Bonnet G, Rudensky AY: **IL-2 production by self-reactive CD4 thymocytes scales regulatory T cell generation in the thymus**. *J Exp Med* 2019, **216**:2466-2478.

39. Zheng C, Zheng L, Yoo J-K, Guo H, Zhang Y, Guo X, Kang B, Hu R, Huang JY, Zhang Q *et al.*: **Landscape of infiltrating T cells in liver cancer revealed by single-cell sequencing**. *Cell* 2017, **169**:1342-1356.e16.

40. •• Azizi E, Carr AJ, Plitas G, Cornish AE, Konopacki C, Prabhakaran S, Nainys J, Wu K, Kiseliovas V, Setty M *et al.*: **Single-cell map of diverse immune phenotypes in the breast tumor microenvironment**. *Cell* 2018. 0
A comprehensive study of immunocytes that infiltrate breast tumors, with creative and extensive bioinformatic analysis, highlighting the relationship between TCR and phenotypic differentiation.

41. •• Ciucci T, Vacchio MS, Gao Y, Tomassoni Ardori F, Candia J, Mehta M, Zhao Y, Tran B, Pepper M, Tessarollo L *et al.*: **The emergence and functional fitness of memory CD4+ T cells require the transcription factor Thpok**. *Immunity* 2019, **50**:91-105.e4
A focused transcriptome and chromatin analysis of the $CD4^+$ T cell response, and the role of Zbtb7b, early after LCMV infection (polyclocal or after tetramer selection for cells that bind the immunodominant GP66 peptide). Illustrates continua in differentiated phenotypes, and the temptation to revert to simple classifications from low-dimension data.

42. •• Gaublomme JT, Yosef N, Lee Y, Gertner RS, Yang LV, Wu C, Pandolfi PP, Mak T, Satija R, Shalek AK *et al.*: **Single-cell genomics unveils critical regulators of Th17 cell pathogenicity**. *Cell* 2015, **163**:1400-1412
The first scRNAseq dissection of CD4+ T cells, parsing Th17 cell induction and responses during experimental autoimmune encephalomyelitis, revealing continuous distribution of phenotypic heterogeneity within a cell-type.

43. Harrison OJ, Linehan JL, Shih H-Y, Bouladoux N, Han S-J, Smelkinson M, Sen SK, Byrd AL, Enamorado M, Yao C *et al.*: **Commensal-specific T cell plasticity promotes rapid tissue adaptation to injury**. *Science* 2019, **363**.

44. Brockmann L, Soukou S, Steglich B, Czarnewski P, Zhao L, Wende S, Bedke T, Ergen C, Manthey C, Agalioti T *et al.*: **Molecular and functional heterogeneity of IL-10-producing CD4+ T cells**. *Nat Commun* 2018, **9**:5457.

45. Xin G, Zander R, Schauder DM, Chen Y, Weinstein JS, Drobyski WR, Tarakanova V, Craft J, Cui W: **Single-cell RNA sequencing unveils an IL-10-producing helper subset that sustains humoral immunity during persistent infection**. *Nat Commun* 2018, **9**:5037.

46. Patil VS, Madrigal A, Schmiedel BJ, Clarke J, O'Rourke P, de Silva AD, Harris E, Peters B, Seumois G, Weiskopf D *et al.*: **Precursors of human CD4+ cytotoxic T lymphocytes identified by single-cell transcriptome analysis**. *Sci Immunol* 2018, **3**.

47. •• Mimitou EP, Cheng A, Montalbano A, Hao S, Stoeckius M, Legut M, Roush T, Herrera A, Papalexi E, Ouyang Z *et al.*: **Multiplexed detection of proteins, transcriptomes, clonotypes**

and CRISPR perturbations in single cells. *Nat Methods* 2019, **16**:409-412
Describes ECCITE-seq, the successor to CITEseq, a dazzling demonstration of the multimodal information that can be garnered from single-cell analysis (transcriptome, surface proteome, CRISPR mutations, and hash-tagging for mixed population profiling).

48. Tu AA, Gierahn TM, Monian B, Morgan DM, Mehta NK, Ruiter B, Shreffler WG, Shalek AK, Love JC: **TCR sequencing paired with massively parallel 3' RNA-seq reveals clonotypic T cell signatures**. *Nat Immunol* 2019, **20**:1692-1699.

49. •• Zemmour D, Zilionis R, Kiner E, Klein AM, Mathis D, Benoist C: **Single-cell gene expression reveals a landscape of regulatory T cell phenotypes shaped by the TCR**. *Nat Immunol* 2019, **19**:291-301
The first scRNAseq analysis of cells sorted according to signal strength, showed the importance of TCR signaling strength in restricting T cell phenotypes.

50. •• Yost KE, Satpathy AT, Wells DK, Qi Y, Wang C, Kageyama R, McNamara KL, Granja JM, Sarin KY, Brown RA *et al.*: **Clonal replacement of tumor-specific T cells following PD-1 blockade**. *Nat Med* 2019, **25**:1251-1259
A comprehensive analysis of TCR clonality and single-cell transcriptomics revealing the origin, stability and expansion of intra-tumoral clonotypes at baseline and after anti-PD1 treatment of basal and squamous cell carcinomas.

51. Pratama A, Schnell A, Mathis D, Benoist C: **Developmental and cellular age direct conversion of CD4+ T cells into RORγ+ or Helios+ colon Treg cells**. *J Exp Med* 2020, **217**.

52. Fan X, Moltedo B, Mendoza A, Davydov AN, Faire MB, Mazutis L, Sharma R, Pe'er D, Chudakov DM, Rudensky AY: **CD49b defines functionally mature Treg cells that survey skin and vascular tissues**. *J Exp Med* 2018, **215**:2796-2814.

53. Li C, DiSpirito JR, Zemmour D, Spallanzani RG, Kuswanto W, Benoist C, Mathis D: **TCR transgenic mice reveal stepwise, multi-site acquisition of the distinctive fat-Treg phenotype**. *Cell* 2018, **174**:285-299.

54. Moran AE, Holzapfel KL, Xing Y, Cunningham NR, Maltzman JS, Punt J, Hogquist KA: **T cell receptor signal strength in Treg and iNKT cell development demonstrated by a novel fluorescent reporter mouse**. *J Exp Med* 2011, **208**:1279-1289.

55. Stoeckius M, Hafemeister C, Stephenson W, Houck-Loomis B, Chattopadhyay PK, Swerdlow H, Satija R, Smibert P: **Simultaneous epitope and transcriptome measurement in single cells**. *Nat Methods* 2017, **14**:865-868.

56. Peterson VM, Zhang KX, Kumar N, Wong J, Li L, Wilson DC, Moore R, McClanahan TK, Sadekova S, Klappenbach JA: **Multiplexed quantification of proteins and transcripts in single cells**. *Nat Biotechnol* 2017, **35**:936-939.

57. Satpathy AT, Granja JM, Yost KE, Qi Y, Meschi F, McDermott GP, Olsen BN, Mumbach MR, Pierce SE, Corces MR *et al.*: **Massively parallel single-cell chromatin landscapes of human immune cell development and intratumoral T cell exhaustion**. *Nat Biotechnol* 2019, **37**:925-936.

58. Ståhl PL, Salmén F, Vickovic S, Lundmark A, Navarro JF, Magnusson J, Giacomello S, Asp M, Westholm JO, Huss M *et al.*: **Visualization and analysis of gene expression in tissue sections by spatial transcriptomics**. *Science* 2016, **353**:78-82.

59. Maniatis S, Äijö T, Vickovic S, Braine C, Kang K, Mollbrink A, Fagegaltier D, Andrusivová Ž, Saarenpää S, Saiz-Castro G *et al.*: **Spatiotemporal dynamics of molecular pathology in amyotrophic lateral sclerosis**. *Science* 2019, **364**:89-93.

60. •• Weinreb C, Rodriguez-Fraticelli A, Camargo FD, Klein AM: **Lineage tracing on transcriptional landscapes links state to fate during differentiation**. *Science* 2020, **367**:eaww3381
A critical look on the potential pitfalls of trajectory inference from single snapshot scRNAseq data, and a theoretical approach derived from physical dynamics illustrating the various putative trajectories that cells can follow.

61. Ludwig LS, Lareau CA, Ulirsch JC, Christian E, Muus C, Li LH, Pelka K, Ge W, Oren Y, Brack A *et al.*: **Lineage tracing in humans enabled by mitochondrial mutations and single-cell genomics**. *Cell* 2019, **176**:1325-1339.e22.

62. Rodriguez-Fraticelli AE, Wolock SL, Weinreb CS, Panero R, Patel SH, Jankovic M, Sun J, Calogero RA, Klein AM, Camargo FD: **Clonal analysis of lineage fate in native haematopoiesis**. *Nature* 2018, **553**:212-216.

63. Weinreb C, Wolock S, Tusi BK, Socolovsky M, Klein AM: **Fundamental limits on dynamic inference from single-cell snapshots**. *Proc Natl Acad Sci U S A* 2018, **115**:E2467-E2476.

64. Lönnberg T, Svensson V, James KR, Fernandez-Ruiz D, Sebina I, Montandon R, Soon MSF, Fogg LG, Nair AS, Liligeto U *et al.*: **Single-cell RNA-seq and computational analysis using temporal mixture modelling resolves Th1/Tfh fate bifurcation in malaria**. *Sci Immunol* 2017, **2**.