Branching developmental pathways through high dimensional single cell analysis in trajectory space

Denis Dermadi*,‡,1,2, Michael Bscheider‡,1,2, Kristina Bjegovic[2], Nicole H. Lazarus[1,2], Agata Szade[1,2], Husein Hadeiba[2], Eugene C. Butcher*[1,2]

[1] Laboratory of Immunology and Vascular Biology, Department of Pathology, School of Medicine, Stanford University, Stanford, CA, United States.

[2] The Center for Molecular Biology and Medicine, Veterans Affairs Palo Alto Health Care System and the Palo Alto Veterans Institute for Research (PAVIR), Palo Alto, CA, United States.

* shared correspondence

‡ shared authorship

High-dimensional single cell profiling coupled with computational modeling holds the potential to elucidate developmental sequences and define genetic programs directing cell lineages. However, existing algorithms have limited ability to elucidate branching developmental paths or to identify multiple branch points in an unsupervised manner. Here we introduce the concept of "trajectory space", in which cells are defined not by their phenotype but by their distance along nearest neighbor trajectories to every other cell in a population. We implement a tSpace algorithm, and show that multidimensional profiling of cells in trajectory space allows unsupervised reconstruction of developmental pathways, and in combination with existing biological knowledge can be used to infer the identity of progenitor populations and of the most differentiated subsets within samples. Applied to high dimensional flow and mass cytometry data, the method faithfully reconstructs known branching pathways of thymic T cell development, and reveals patterns of tonsillar B cell development and of B cell migration. Applied to single cell transcriptomic data, the method unfolds the complex developmental sequences and genetic programs leading from intestinal stem cells to specialized epithelial phenotypes. Profiling of

complex populations in high-dimensional trajectory space should prove useful for hypothesis generation in developing cell systems.

Precursor cells give rise to differentiated progeny through branching developmental pathways. Single cell technologies hold the promise of elucidating the developmental progression and defining underlying transcriptomic drivers and modulators. Mass cytometry (CyTOF) and single cell RNA-seq (scRNAseq) can capture a high-dimensional profile of a "cellular snapshot" within analyzed tissue that contains all developing, renewing and differentiated cell populations. High-dimensional profiles of cells can then be computationally aligned to reveal developmental relationships. Several algorithms have been proposed to model developmental trajectories, but most require *a priori* knowledge (e.g. selection of starting cell(s)), or are limited in their capacity to visualize multiple branches or to analyze large datasets[1-8].

Here we show that developmental pathways can be reconstructed from single cell profiles by analyzing cells in "trajectory space", in which each cell is represented by a profile or vector of its distance along nearest neighbor pathways to every other cell. The concept is illustrated in Fig. 1a, with a schematic example of 10 cells derived from cell A and analyzed with two phenotypic markers. Cells H and E are phenotypically similar but arise from different developmental sequences and thus are developmentally distant. A matrix of directionless cell-to-cell distances along the developmental pathways is constructed. Standard dimensionality reduction tools (e.g. principal component analysis (PCA)) are used to visualize and explore cell relationships in this novel trajectory space. As illustrated, the method reconstitutes the correct branching developmental sequences of cells in the simple example.

To implement the concept, we developed a tSpace algorithm. Its application to single cell datasets relies on the assumptions that (i) developmental processes are gradual, (ii) all developmental stages are represented in the data and (iii) markers used to profile cells are regulated and sufficiently informative to distinguish different developmental pathways. Starting with cell profiles (phenotypes), tSpace identifies the (*k*) nearest neighbors of every cell, constructs a nearest neighbor (*NN*) graph that provides connections to all cells in the dataset, and then calculates distances from each cell to every other cell in

the population along NN connections. tSpace determines distances within the graph using Wanderlust[1], an algorithm that takes advantage of waypoints and implements a weighing scheme to reduce "short-circuits" in selecting optimal paths (Methods). In experimental datasets, we find that Wanderlust refines developmental branches and reduces apparent "noise" in calculated trajectories (Fig. S1).

In analyses of large datasets, calculation of distances from every cell to every other cell can be computationally intensive and impractical. In this case, tSpace first defines cell clusters using K means or self-organizing map (SOM) algorithms, and then calculates distances along trajectories from one cell from each cluster to every other cell within the dataset. We find that in most biological samples evaluated, a matrix of trajectory distances from a relatively small number of cells (e.g. 100 – 1000) is sufficient to capture developmental pathways, although the required number is expected to depend on the complexity of the branching structure of cellular sequences (Fig. S1).

To test the ability of tSpace to correctly determine developmental relations and reveal branch points, we analyzed data from different species and tissues generated with commonly used single cell platforms: fluorescence or mass cytometry and scRNAseq

Thymic T cell development in the thymus is well established (Fig. 1b) and allowed us to validate tSpace performance in a defined developmental system. We generated flow cytometric profiles of mouse thymocytes using a panel of 13 antibodies (Supplementary Table 1). Our panel detects early T-cell populations (so-called "double negative" populations DN1-DN4, which lack CD4 and CD8 and are distinguished by CD44 and CD25 expression), double positive (DP) $CD4^+CD8^+$ cells, and CD4 or CD8 single positive (SP) T-cells including poised thymic emigrant phenotype cells, regulatory T cells ($CD4^+$, $CD25^+$, $Foxp3^+$) and a small fraction of SP T-cells expressing CD44, an activation and memory marker[8,9]. We manually gated on these subsets (Fig. S2)[9]. Unsupervised tSpace analysis reveals the expected bifurcation of CD4 vs CD8 lineages from the dominant DP population in thymopoiesis and correctly positions T-cell populations from early (DN2) to mature thymic emigrant phenotype T cells in known developmental relationships (Fig. 1c). DN1 cells were not present in the dataset. In addition to the

expected major bifurcation of CD4 vs CD8 cells arising from the dominant DP pool, the analysis reveals branching of regulatory T cells (Foxp3$^+$) from the SP CD4 stage of CD4 branch. In contrast to methods based on clustering, tSpace highlights a developmental continuum of cells allowing exploration of intermediate populations. For example, tSpace visualizes DP cells in transition to the more mature SP CD4 and CD8 T cells. The transitional cells co-express CD4 and CD8 but some have upregulated TCRβ and CD3ε, a characteristic of positively selected cells[10]. Conventional clustering using t-SNE identifies the major subsets, but does not clarify developmental relationships (Fig. 1d).

To evaluate expression of markers during development of CD4 cells, we manually gated on cells along the path from DN2 cell to CD4 thymic poised emigrants (inset Fig. 1c), aligned them along a linear trajectory (Methods) and evaluated expression in a heatmap. The results capture regulation of the markers as cells progress towards maturity, recapitulating studies of the phenotypic stages of thymic T cell development. Examined protein expression trends confirm previously reported upregulation of CCR9 in DN3 cells and reveal notably stronger expression in DN4-DP transitioning cells. CCR9 chemokine receptor binds CCL25 and promotes T-cell cortical positioning[11].

Single cell analyses hold the potential to provide insights into patterns of cell development in settings not accessible to experimental manipulation, as in the human. We applied tSpace to the development of B cells in human tonsils. Naïve (IgD$^+$) B cell differentiation towards Immunoglobulin A or G (IgA, IgG) class-switched memory or plasma cells has been investigated. However, the sequence of class switch and fate determining decision points is still not entirely understood[12,13]. We used a panel of mass labeled antibodies that detects ~25 markers of B cell subsets and maturation (Supplementary Table 2) to stain human tonsils and blood. We gated on B cells and plasmablasts (Fig. S3) and applied tSpace (Fig. 2a).

tSpace analysis provided trajectories leading to the 4 most prominent terminal tonsil populations, IgG and IgA class switched memory cells and plasmablasts (PB). The first principal component in trajectory space (tPC1) delineates the transition from naïve to germinal center cells (GCC, Fig. 2b); tPC2 the differentiation of memory or plasma cells; tPC4 pathways to IgA vs IgG class switched cells (Fig. 2c). Early naïve cells

express CXCR5 and CCR6 which mediates lymphoid tissue entry. A broad strand of cells connects naïve IgD^+ B cells to the cluster of proliferating GC centroblasts and centrocytes (Fig. 2a, S4a-c). Along this path from naïve cells, IgD and IgM are downregulated as cells transition to $CD38^+CD77^+$ centroblasts. There are clear trajectories from GCC to class switched PB and memory cells. CD27 is upregulated in memory B cell branches (Fig. S4d-f). Subsets of tonsil memory cells express CXCR3 and CLA (Fig S4e-h). CLA is induced during immune responses associated with squamous epithelial surfaces including the oral mucosa, and squamous epithelial cells interdigitate into tonsillar lymphoid tissue[14-16]. Thus, the CLA^+ memory cells may be generated de novo from GCC in the tonsil. CD38, present on activated B cells and GCC, is further induced and CD19 and CD20 are lost in developing plasma cells (Fig. S4i-j).

The pathways from GC to differentiated IgA and IgG PB are well delineated along tight branches. In contrast, class switched IgG^+ and IgA^+ memory B cells are relatively dispersed in trajectory space (Fig. 2c, Fig. S4e-f): they constitute a “cloud” of cells some of which branch from the GC pool as mentioned, while others are closer in trajectory space to the path from naïve to GCs. This shows that tSpace does not “force” cells into specific developmental sequences or paths. Since cell alignment in trajectory space does not intrinsically provide directional information, the presence of IgG and IgA expressing B cells “near” the naïve to GC path would be consistent either with class switching of B cells during the naïve to GC transition, or with integration of previously committed memory cells into the developmental pathway to GC. Low expression of CD27 and retention of naïve markers CCR6, CXCR5 and $\alpha4\beta7$ on the class switched cells adjacent to the “naïve to GC” sequence is most consistent with the former interpretation (Fig. S5). While class switch recombination is normally attributed to the GC reaction, in experimental settings class switching can occur prior to GC formation, and it is observed in T-independent B cell responses as well[17]. tSpace analysis raises the possibility that, even in steady state human tonsil, some activated B cells make the class switch decision prior to becoming GC. This corresponds to reports that unswitched memory B cells with high receptor affinity can emerge from naïve B cells [19,20]. In contrast to their class-switched counterparts, IgM memory cells ($CD27^+$, $CD38^-$) appeared more closely connected to naïve (IgM^+, IgD^+, $CD27^-$, $CD38^-$) cells in most tPCs, with tPC2 specifically expanding this trajectory (Fig. 2b).

Mature effector and memory cells leave their sites of antigen activation and circulate via the blood to distant organs and tissues[14]. We reasoned that trajectories should thus link terminally differentiated cells, ready to exit their site of generation, with progeny cells in blood. Indeed, when we applied tSpace to combined blood and tonsil B cells datasets, blood PB aligned at the termini of tonsillar IgG and IgA PB branches (Fig. 2d). In contrast, blood memory cells and naïve IgD^+ cells overlapped extensively with their tonsillar counterparts in trajectory space, presumably reflecting their recirculation and interchange between lymphoid tissues and blood (Fig. 2d). The results show that the approach can unfold inter-organ transitions (i.e. migration patterns) of immune cells in settings where experimental analyses of leukocyte trafficking are challenging, as in humans. However, more extensive panels of trafficking associated receptors, or single cell gene expression analyses, will be necessary to increase the accuracy of inter-organ trajectories.

Single cell RNAseq is emerging as a powerful tool for the characterization of cell populations and provides rich cellular profiles for studying cell relationships. We applied tSpace to published scRNAseq data from mouse intestinal epithelial cells[18]. Intestinal epithelium forms the single-cell layer separating the lumen of small intestine from intestinal lamina propria. Almost all cells in the epithelium have a short life-span of about 4-7 days[19] and continuous renewal is driven by division of $Lgr5^+$ crypt base columnar (CBC) cells residing in the bottom of the intestinal crypts. The cells further divide in the transit-amplifying (TA) zone of the crypt and differentiate into absorptive (enterocyte) or secretory (Goblet cell, Paneth cell, enteroendocrine (EE) cell) lineages.

We ran tSpace on the set of 3521 cells from Yan, K. S. *et al.*[18] using 2420 variable genes (Methods). tSpace clearly delineates absorptive/enterocyte and secretory/EE development, both arising from $Lgr5^+$ CBC cells, and positions cell types in developmentally meaningful relations (Fig. 3a, S6, for cell labels see Methods). To assess the validity of tSpace alignments we isolated the early segment of the EE branch, preceding the differentiation of various terminal EE subsets, along with the enterocyte trajectory (Fig. 3a-c) and examined gene expression of selected hallmarks of intestinal differentiation[20] along these

developmental trajectories. The Wnt agonist *Lgr5* and its homolog *Lgr4* are in resting CBC and dividing slow-cycling CBC (sc-CBC) cells, but the analysis shows that *Lgr4* expression is retained in post-mitotic cells differentiating towards absorptive enterocytes from cycling TA (c-TA), suggesting that in addition to its known role in proliferation of TA cells[21] it may contribute to enterocyte fate or specification. Additionally, we confirm *Ascl2*[22], *OlfM4*[23] and *Prom1*[24] as robust markers of the presumptive crypt populations (CBC to TA cells) and show that the expression of *Prom1* extends into the TA pool, confirming previous findings[25] (Fig. 3d, S7a).

tSpace clearly positions *Dll1*-expressing cells in trajectory space between CBC cells and mature EE populations (Fig. 3d). These cells resemble previously described short-lived secretory progenitors (slEEP)[26], which upregulate EE lineage specification genes *Neurog3*, *Neurod1* and *Neurod2*[26,27]. In the original analysis of this scRNAseq data, these slEEP were labeled either as cycling stem cells (cSC) or Goblet cells (Fig. 3d, S6, S7A), reflecting the fact that the existing analytical tools applied (t-SNE, SPADE) failed to define these cells either as a discrete subset or as a precursor population. Their location in trajectory space however clearly suggests that slEEPs give rise to all other EE subsets. In this prediction, tSpace analysis is consistent with published fate mapping studies[26-28]. Thus, the patterns of gene expression and cell positioning in trajectory space mirror observations from decades of research on intestinal development (Fig. 3d, S7a).

tSpace segregates some sc-CBC and c-TA cells to the early EE or enterocyte branches, suggesting that they are already developing towards if not committed to EE or enterocyte fates. In order to identify transcription factors (TF) that might specify fate within these early progenitors, we initially aligned the EE and enterocyte trajectories (Fig. 3b-c) using dynamic time warping[29], and compared TFs between aligned segments (Fig. S7b). Four TF modules were identified (Fig. S7c, 3e). The first is expressed among early dividing cSC and TA, presumably crypt cells, cells that are shared between the stem to EE and stem to enterocyte trajectories (M1). TFs in this module are involved in proliferation and DNA maintenance (e.g. *Ccna2, Cdk2, Fancd2, Rbl1*). Three additional modules differentiate the two branches (M2-M4, Fig. S7c, 3E). The M2 TFs (e.g. *Foxa2, Foxa3, Neurog3, Sox4, Sox9*) are expressed by early cells but maintained

after stages 3-4 (late CBC cells) selectively in the EE branch; these TFs have been associated with endocrine and pancreatic development and may coordinate secretory pathways within intestinal enteroendocrine cells[30]. Among M2 TFs, tSpace reveals specific high expression of *Sox4* in slEEP cells, suggesting a role in EE specification[31] (Fig. 3e, S6c,e).

The M3 and M4 TFs are expressed preferentially in the enterocyte branch. The M3 TFs activate lipid and cholesterol metabolism (e.g. *Cebpb, Klf5, Nr5a2,* Fig. S7c, Fig. 3e), known to be important for mature enterocytes[32,33]. *Nfe2l2* and *Maf*, part of the M4 module, suggest associated activation of *Nfe2l2/Nrf2*-antioxidant response element (ARE) pathway[34]. Enterocytes utilize short fatty acids as a source of energy, and fatty acid metabolism generates reactive oxygen species (ROS). ROS are also abundant in the intestinal lumen[35]. Upregulation of the *Nfe2l2*-ARE pathway may help protect differentiating enterocytes from oxidative damage[35]. Taken together (Fig. 3f), the trajectory analysis shows that rapidly proliferating sc-CBC and c-TA cells (crypt cells) are already heterogeneous and express gene programs leading to secretory vs. absorptive phenotypes.

In conclusion, we have presented the concept of trajectory space and its implementation in the tSpace algorithm for elucidation of branching developmental pathways and mechanisms from single cell profiles. tSpace performs well across different biological systems and platforms and reveals known and novel biology. A number of methods for aligning cells in developmental sequences have been described, and recent reviews highlight their unique features and limitations[8]. tSpace compares favorably to published algorithms: It allows non-supervised discovery and exploration of branching developmental sequences. It does not require prior knowledge, selection of minimal parameter sets or clustering. It is capable of positioning rare cells in proper developmental alignment, and is applicable to large datasets. The output is deterministic and intuitive. We believe that tSpace will prove useful to the rapidly growing field of singe cell analysis.

Acknowledgments. Mass cytometry analysis for this project was done on Cyrano instrument in the Stanford Shared FACS Facility, obtained by S10OD016318-01 NIH grant. We thank Grace X.Y. Zheng for

provided scRNAseq data, and Dr. Saara Ollila and Dr. Joni Nikkanen for discussions. This work was supported by NIH grants R37-AI047822 and R01-AI093981 to ECB and R01-AI109452 to HH, and by pilot awards under ITI Seed 122C158 and CCSB grant U54-CA209971. MB was supported by fellowships from the German Research Foundation (DFG, BS56/1-1) and the Crohn's and Colitis Foundation of America. AS was supported by the Mobility Plus fellowship from the Ministry of Science and Higher Education, Poland (1319/MOB/IV/2015/0).

Author contributions. DD wrote the algorithm, supervised computational analyses, and interpreted intestinal data; KB wrote parts of the algorithm; DD & HH designed and interpreted the thymus study; MB & NHL performed and MB analyzed the tonsil B cell study; AS prepared human tissues; DD, MB, ECB wrote the manuscript; HH provided advice; ECB conceived the trajectory space concept and supervised the project.

The authors have declared no conflict of interest.

References

1. Bendall, S. C. *et al.* Single-cell trajectory detection uncovers progression and regulatory coordination in human B cell development. *Cell* **157,** 714–725 (2014).
2. Setty, M. *et al.* Wishbone identifies bifurcating developmental trajectories from single-cell data. *Nature Biotechnology* 1–14 (2016). doi:10.1038/nbt.3569
3. Paul, F. & Amit, I. Plasticity in the transcriptional and epigenetic circuits regulating dendritic cell lineage specification and function. *Current Opinion in Immunology* **30,** 1–8 (2014).
4. Zunder, E. R., Lujan, E., Goltsev, Y., Wernig, M. & Nolan, G. P. A continuous molecular roadmap to iPSC reprogramming through progression analysis of single-cell mass cytometry. *Cell Stem Cell* **16,** 323–337 (2015).
5. Haghverdi, L., Büttner, M., Wolf, F. A., Buettner, F. & Theis, F. J. Diffusion pseudotime robustly reconstructs lineage branching. *Nat. Methods* **13,** 845–848 (2016).

6. Lönnberg, T. *et al.* Single-cell RNA-seq and computational analysis using temporal mixture modelling resolves Th1/Tfh fate bifurcation in malaria. *Sci Immunol* **2,** eaal2192 (2017).

7. Simonds, E. F. *et al.* Extracting a cellular hierarchy from high-dimensional cytometry data with SPADE. *Nature Biotechnology* **29,** 886–891 (2011).

8. Cannoodt, R., Saelens, W. & Saeys, Y. Computational methods for trajectory inference from single-cell transcriptomics. *Eur. J. Immunol.* **46,** 2496–2506 (2016).

9. Shah, D. K. & Zúñiga-Pflücker, J. C. An overview of the intrathymic intricacies of T cell development. *J. Immunol.* **192,** 4017–4023 (2014).

10. Brodeur, J.-F., Li, S., Damlaj, O. & Dave, V. P. Expression of fully assembled TCR-CD3 complex on double positive thymocytes: synergistic role for the PRS and ER retention motifs in the intra-cytoplasmic tail of CD3epsilon. *Int. Immunol.* **21,** 1317–1327 (2009).

11. Wurbel, M.-A., Malissen, B. & Campbell, J. J. Complex regulation of CCR9 at multiple discrete stages of T cell development. *Eur. J. Immunol.* **36,** 73–81 (2006).

12. De Silva, N. S. & Klein, U. Dynamics of B cells in germinal centres. *Nature Publishing Group* **15,** 137–148 (2015).

13. Dufaud, C. R., McHeyzer-Williams, L. J. & McHeyzer-Williams, M. G. Deconstructing the germinal center, one cell at a time. *Current Opinion in Immunology* **45,** 112–118 (2017).

14. Seong, Y. *et al.* Trafficking receptor signatures define blood plasmablasts responding to tissue-specific immune challenge. *JCI Insight* **2,** e90233 (2017).

15. Rott, L. S., Briskin, M. J., Andrew, D. P., Berg, E. L. & Butcher, E. C. A fundamental subdivision of circulating lymphocytes defined by adhesion to mucosal addressin cell adhesion molecule-1. Comparison with vascular cell adhesion molecule-1 and correlation with beta 7 integrins and memory differentiation. *J. Immunol.* **156,** 3727–3736 (1996).

16. Kantele, A. *et al.* Cutaneous lymphocyte antigen expression on human effector B cells depends on the site and on the nature of antigen encounter. *Eur. J. Immunol.* **33,** 3275–3283 (2003).

17. Stavnezer, J. & Schrader, C. E. IgH chain class switch recombination: mechanism and regulation. *J. Immunol.* **193,** 5370–5378 (2014).

18. Yan, K. S. *et al.* Intestinal Enteroendocrine Lineage Cells Possess Homeostatic and Injury-Inducible Stem Cell Activity. *Cell Stem Cell* **21,** 78–90.e6 (2017).

19. Barker, N., van Oudenaarden, A. & Clevers, H. Identifying the stem cell of the intestinal crypt: strategies and pitfalls. *Cell Stem Cell* **11,** 452–460 (2012).

20. Clevers, H. The Intestinal Crypt, A Prototype Stem Cell Compartment. *Cell* **154,** 274–284 (2013).

21. Mustata, R. C. *et al.* Lgr4 is required for Paneth cell differentiation and maintenance of intestinal stem cells ex vivo. *EMBO Rep.* **12,** 558–564 (2011).

22. van der Flier, L. G. *et al.* Transcription factor achaete scute-like 2 controls intestinal stem cell fate. *Cell* **136,** 903–912 (2009).

23. van der Flier, L. G., Haegebarth, A., Stange, D. E., van de Wetering, M. & Clevers, H. OLFM4 is a robust marker for stem cells in human intestine and marks a subset of colorectal cancer cells. *Gastroenterology* **137,** 15–17 (2009).

24. Zhu, L. *et al.* Prominin 1 marks intestinal stem cells that are susceptible to neoplastic transformation. *Nature* **457,** 603–607 (2009).

25. Itzkovitz, S. *et al.* Single-molecule transcript counting of stem-cell markers in the mouse intestine. *Nat. Cell Biol.* **14,** 106–114 (2011).

26. van Es, J. H. *et al.* Dll1+ secretory progenitor cells revert to stem cells upon crypt damage. *Nat. Cell Biol.* **14,** 1099–1104 (2012).

27. Schonhoff, S. E., Giel-Moloney, M. & Leiter, A. B. Neurogenin 3-expressing progenitor cells in the gastrointestinal tract differentiate into both endocrine and non-endocrine cell types. *Dev. Biol.* **270,** 443–454 (2004).

28. Jenny, M. *et al.* Neurogenin3 is differentially required for endocrine cell fate specification in the intestinal and gastric epithelium. *EMBO J.* **21,** 6338–6347 (2002).

29. Giorgino, T. Computing and Visualizing Dynamic Time Warping Alignments in R: The dtw Package. 1–24 (2009).

30. The Tabula Muris Consortium, Quake, S. R., Wyss-Coray, T. & Darmanis, S. Single-cell transcriptomic characterization of 20 organs and tissues from individual mice creates a Tabula Muris. *bioRxiv* 237446 (2017). doi:10.1101/237446

31. Gracz, A. *et al.* Sox4 drives intestinal secretory differentiation toward tuft and enteroendocrine fates. (2017). doi:10.1101/183400

32. Yen, C.-L. E., Nelson, D. W. & Yen, M.-I. Intestinal triacylglycerol synthesis in fat absorption and systemic energy metabolism. *The Journal of Lipid Research* **56,** 489–501 (2015).

33. Degirolamo, C., Sabbà, C. & Moschetta, A. Intestinal nuclear receptors in HDL cholesterol metabolism. *The Journal of Lipid Research* **56,** 1262–1270 (2015).

34. Itoh, K. *et al.* An Nrf2/small Maf heterodimer mediates the induction of phase II detoxifying enzyme genes through antioxidant response elements. *Biochem. Biophys. Res. Commun.* **236,** 313–322 (1997).

35. Ferrebee, C. B. *et al.* Organic Solute Transporter α-β Protects Ileal Enterocytes From Bile Acid-Induced Injury. *Cellular and Molecular Gastroenterology and Hepatology* **5,** 499–522 (2018).

36. Tetteh, P. W. *et al.* Replacement of Lost Lgr5-Positive Stem Cells through Plasticity of Their Enterocyte-Lineage Daughters. *Cell Stem Cell* **18,** 203–213 (2016).

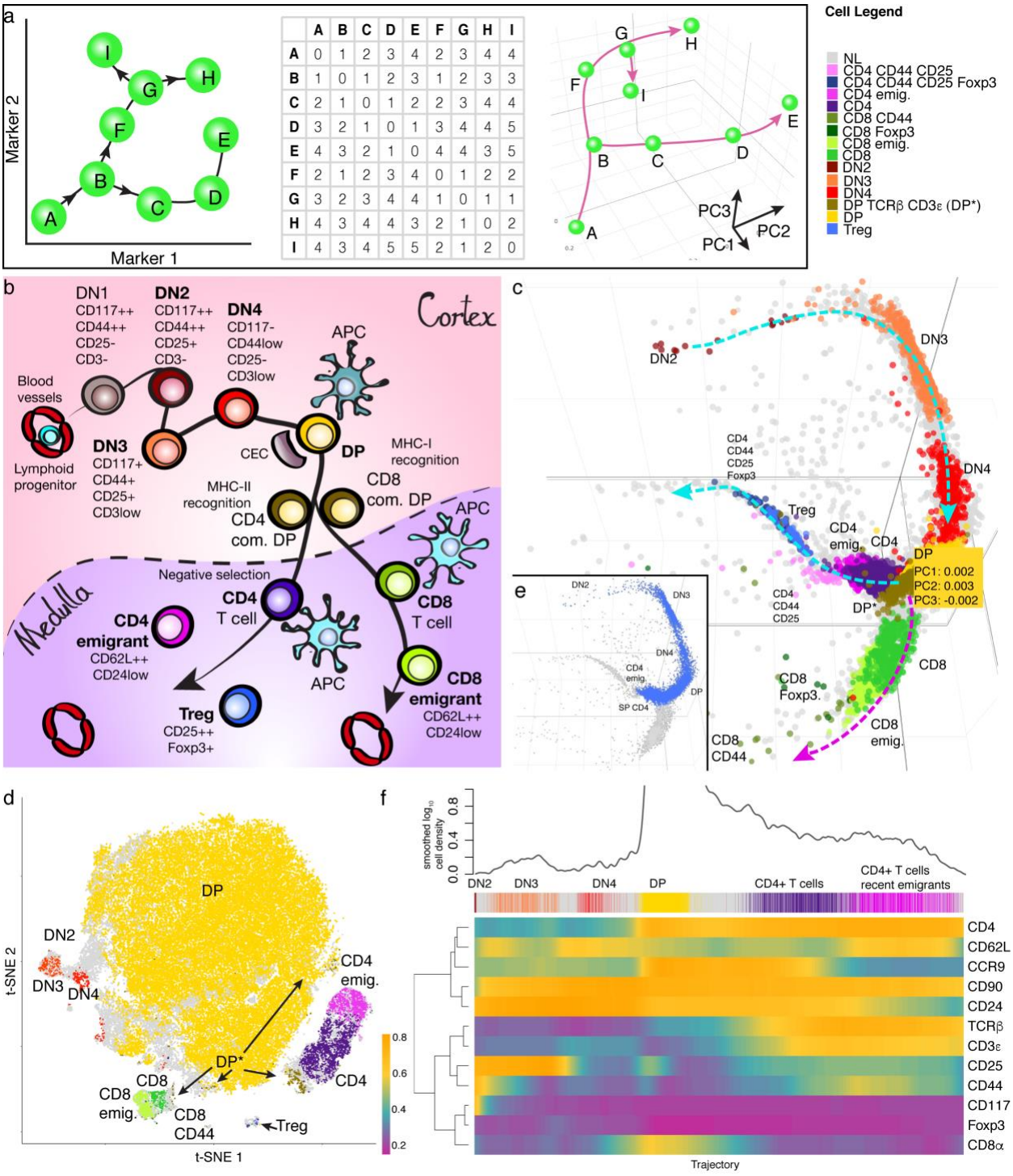

Figure 1. tSpace orders thymic T-cells in correct developmental trajectories and recovers expression patterns of markers of T-cell differentiation. a A simple example illustrates the concept of trajectory space. The "cells" are marked with the letters (A-I) and their developmental sequences with arrows. A matrix of cell to cell distances along developmental paths is created. Visualization of cell positions in this "trajectory space" cells recapitulates branching

developmental pathways. **b** An overview of thymic T-cell development with immunophenotypic markers, starting from T-cell progenitors through poised thymic emigrants and specialized T-cells (e.g. Treg $CD4^+$, $CD25^+$, $Foxp3^+$). Populations labeled in bold were detected in FACS data. **c** Unsupervised tSpace analysis of T-cell development in mouse thymus restores developmental relations between conventional T-cell populations. **d** t-SNE of thymic T-cells defines clusters but not developmental relationships. It fails to identify a unique DP* transitional population between DP and SP CD4 or CD8 T-cells, indicating loss of sensitivity to distinguish smaller populations when close to abundant ones (for example comparison of the more dominant DP from the DP* populations in C: tSpace and D: t-SNE). **e** Isolated trajectory from DN2 precursors to CD4 thymic emigrants. **f** Smoothed expressions of measured markers along isolated trajectory (E) reveals patterns of protein regulation during T-cell differentiation. The position of manually defined subsets along the isolated trajectory is shown as a reference above the heatmap. The abundance of DP cells, seen as a broad peak (tip not shown) of cell density in the trajectory, correlates with increased CCR9. DN - double negative T-cells; APC - antigen presenting cell; CD4 emig. - $CD4^+$ T-cell poised emigrants; CD8 emig. - $CD8^+$ T-cell poised emigrants.

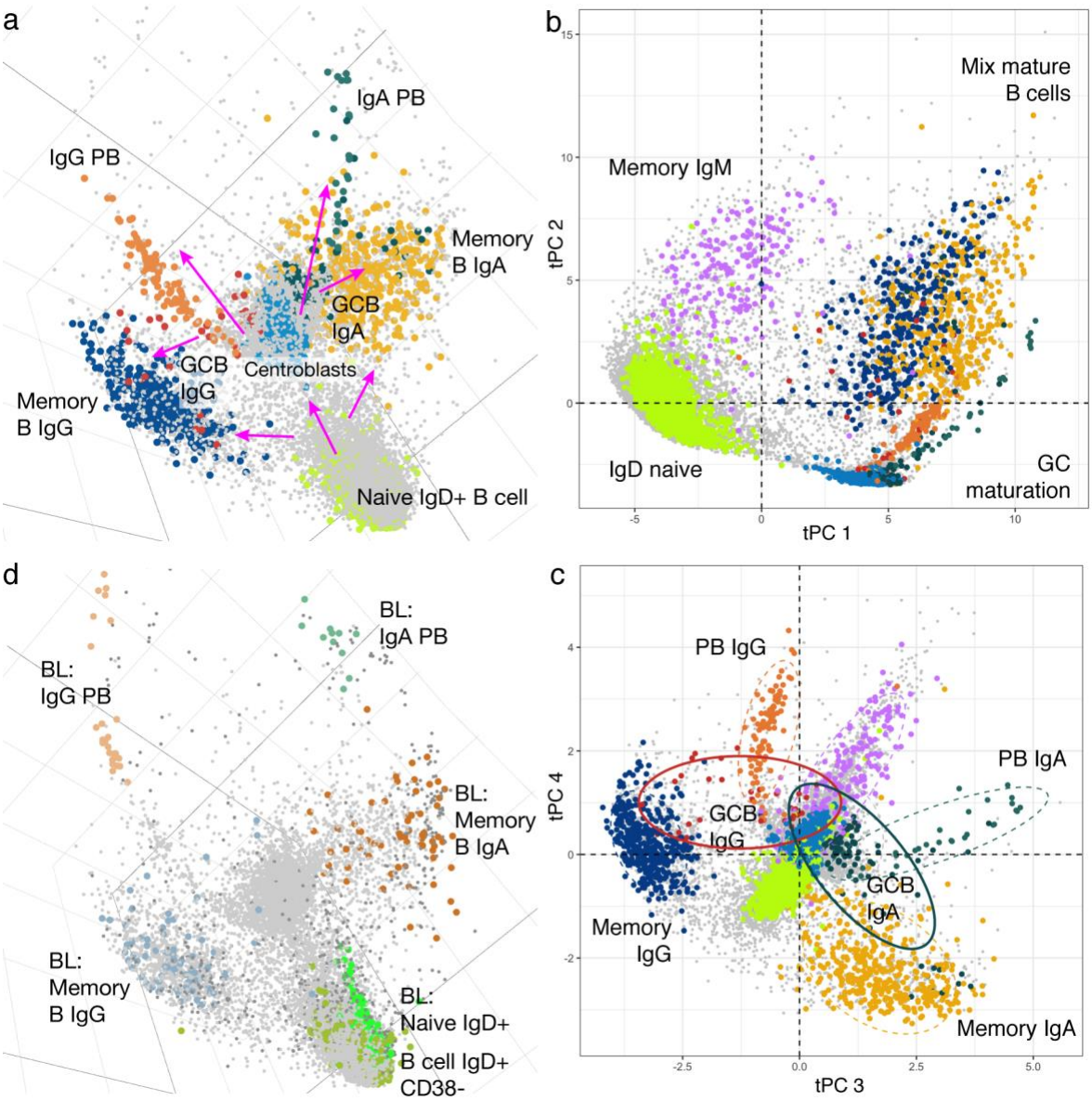

Figure 2. tSpace analysis of B cell differentiation in tonsils and inter-organ trajectories with blood. a tSpace unravels maturation paths of B cells starting from naïve B cells in tonsil throughout GC into memory B cells and plasmablasts (PB). Blue arrows mark suggested directionalities based on known biology. **b-c** Different principal components reveal branches and potential developmental relationships in tonsillar B cell maturation. Ellipses show 80% confidence intervals for indicated clusters. **d** Blood PB align as an extension of tonsillar PB trajectories, while recirculating blood memory B cells overlap with the major tonsil memory cell clouds. In D, tonsil B cells are in light gray.

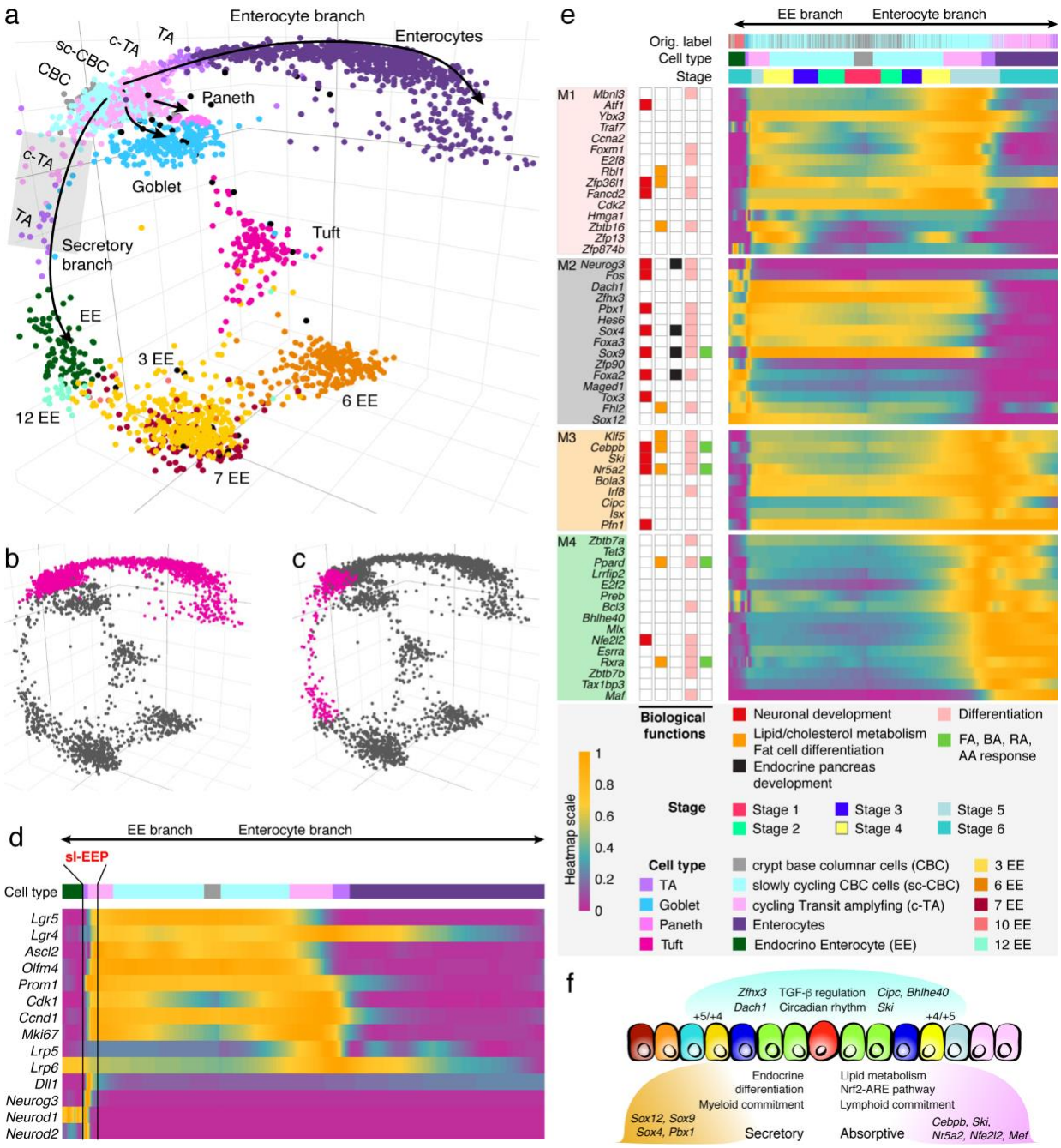

Figure 3. tSpace analysis of mouse small intestinal epithelial cells based on scRNAseq data. a tSpace separates trajectories to enterocytes, enteroendocrine (EE), Paneth and Goblet cells. For cell population labels see Methods and Fig. S6. Shaded area indicates short lived EE progenitors (slEEP) cells. **b** Isolated enterocyte trajectory. **c** Isolated EE trajectory. **d** Expression patterns of selected genes[20] (known markers or regulators of intestinal crypt development; expanded gene list in Fig. S7a) along the isolated trajectories confirm known biology. **e** Four detected transcription factor modules in early trajectories: M1 comprises of TFs involved in cell cycle and genome integrity,

while M2 and M3-M4 differentiate the two lineages. M2 characterizes the EE branch and consists of TFs involved in secretory (e.g. endocrine pancreas) development, while TFs in M3-M4 are modulated by fatty acids (FA), retinoic acid (RA), bile acids (BA), amino acids (AA) and upregulate lipid/cholesterol metabolism together with *Nrf2*/antioxidant response element (ARE) pathways. M3-M4 likely drive changes in TA cells that lead to enterocyte lineage commitment. **f** Summary of the detected changes between two lineages. Different genes in the TGF-β and circadian rhythm pathways are expressed in the two lineages (genes in blue above the cartoon). TFs enriched in the EE branch are involved in endocrine secretory cell development; while TFs associated with enterocyte commitment include regulators of lipid/cholesterol metabolism. Consistent with the literature, peak expressions of *Dll1 and Sox4* in EE branch (Fig. 3d-e), and *Apli* in enterocyte branch (Fig. S6c) mark specific progenitor cells: these may correlate to the +4/+5 position in the intestinal crypt[26,31,36]: We propose that lineage programs start to drive differentiation at this stage. The EE module M2 and enterocyte module M4 have TF's associated with myeloid vs lymphoid commitment in hematopoiesis, respectively.