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

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 DD & 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.

High-dimensional single cell profiling coupled with computational modeling holds the potential to elucidate developmental sequences and define genetic programs directing cell lineages. Here we introduce an approach to the discovery and exploration of developmental pathways based on 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 complex developmental sequences. tSpace is robust, scalable, and implements a global approach to trajectory analysis that attempts to preserve both local and distant relationships in developmental pathways. 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.

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 distances along nearest neighbor pathways to every other cell. The concept is illustrated in Fig. 1a, with a schematic example of several 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 dense matrix of cell-to-cell distances along the developmental pathways is constructed. Standard dimensionality reduction tools [e.g. principal component analysis (PCA), UMAP] 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; calculates distances from each cell to every other cell in the population along NN connections; and exports a dense matrix of N x T (number of cells x number of calculated “trajectories”, vectors of cell-to-cell distances within the manifold) dimensions. tSpace determines the distances within the KNN graph using Wanderlust[1], an algorithm that takes advantage of subgraphs and waypoints, and which implements a weighting scheme to reduce “short-circuits” in selecting optimal paths (Fig. S1). The Wanderlust algorithm has been described in detail[1]. It significantly improves the definition of branching pathways even in simple flow cytometry datasets (Fig. S1). We outline the effects of varying user-defined Wanderlust parameters on tSpace in the Supplementary Methods (Fig. S2). tSpace detection of developmental branches is robust over a range of input parameters, allowing implementation of default settings that work well in different applications. The tSpace output also provides principal component and UMAP embedding of cells in trajectory space, suitable for visualization and biological exploration of developmental pathways. It also exports the matrix of trajectories (cell-to-cell distances) useful for analysis of e.g. gene/protein expression changes along isolated linear developmental branches.

For samples with large number of cells (N), tSpace has the option of calculating fewer trajectories, but it is important that these trajectories start from cells well distributed throughout phenotypical space. K-means clustering identifies groups of cells that are well distributed within phenotypical space, and we calculate trajectories from one cell from each cluster. The clusters are not used for further analysis. As illustration of this feature, tSpace recovers thymic T cell maturation branches from as few as 25 – 100 trajectories (Fig. S1).

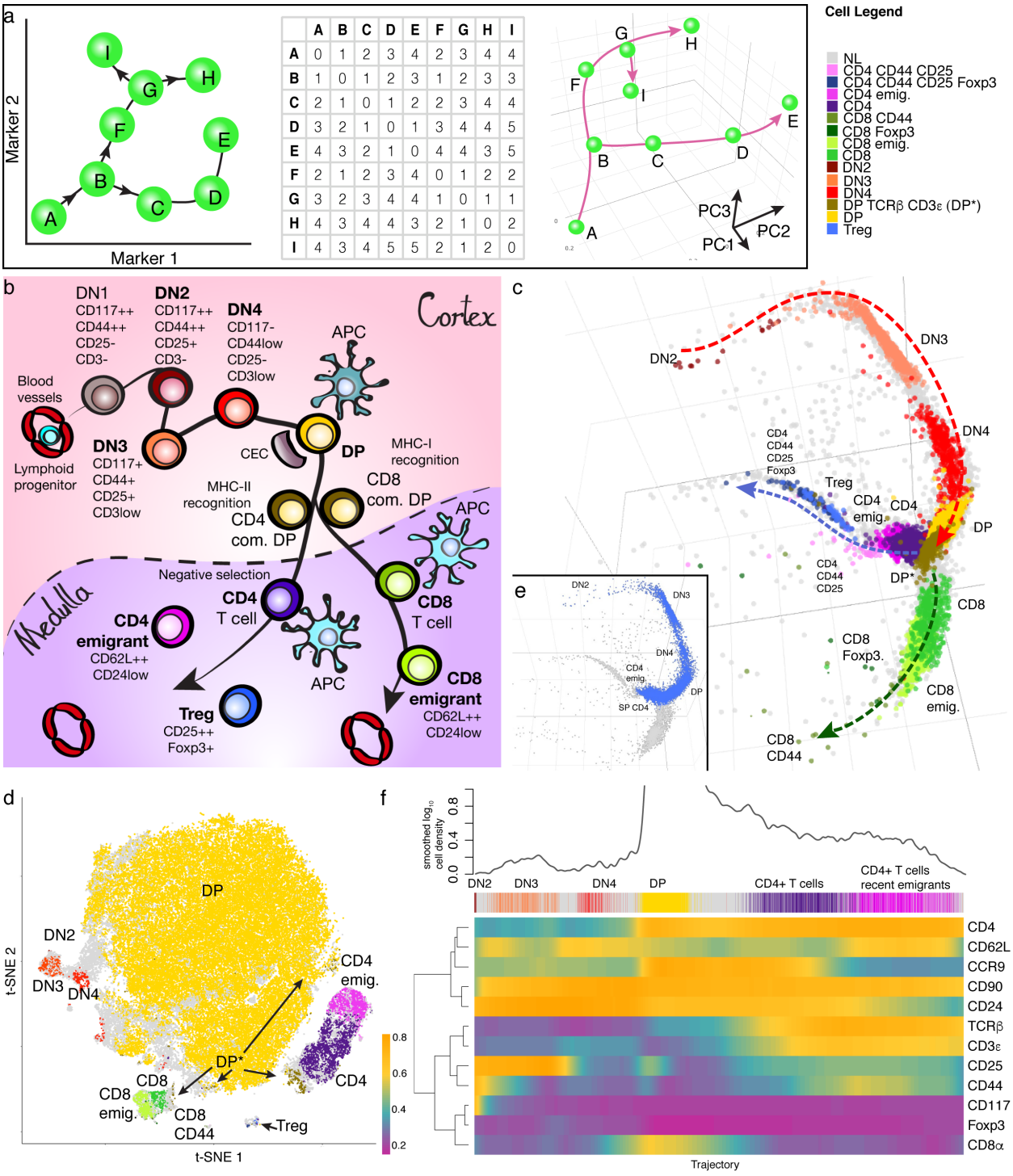

Figure 1. tSpace orders thymic T cells in correct developmental trajectories and recovers expression patterns of markers of T cell differentiation. a A simple example illustrates the concept of trajectory space. The "cells" are marked with the letters (A-I) and their developmental sequences with arrows. A matrix of cell-to-cell distances along developmental paths is created. Visualization of cell positions in this 'trajectory space' 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.

To test the ability of tSpace to correctly determine developmental relations and reveal branch points, we analyzed data 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 allows validation of tSpace in a defined 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. We manually gated on these subsets and labeled them (Fig. S3)[2]. Unsupervised tSpace analysis reveals the expected bifurcation of CD4 vs CD8 lineages from the dominant DP population in thymopoiesis and correctly positions T cells 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[3]. Conventional clustering, based on measured markers, using t-SNE identifies the major subsets, but does not clarify developmental relationships (Fig. 1d).

The tSpace output allows evaluation of expression of markers along developmental paths. To illustrate this for CD4 cell differentiation, we manually gated on cells along the path from DN2 cell to CD4 thymic poised emigrants (inset Fig. 1c), identified and averaged trajectories in the exported tSpace matrix that started from early DN2 cells, and displayed marker expression by cells vs their trajectory distance from DN2 cells in a heatmap (Methods). The results capture regulation of the markers as cells progress towards maturity, recapitulating known phenotypic progression of thymic T cell development and highlighting details of transitional states. For example, protein expression trends confirm upregulation of CCR9 in DN3 cells but reveal notably stronger expression in DN4-DP transitioning cells. CCR9 chemokine receptor binds CCL25 and promotes T cell cortical positioning[4].

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[5,6]. We used a panel of mass labeled antibodies that detects ~25 markers of B cell subsets and maturation (Supplementary Table 2) to stain human tonsil lymphocytes. We applied tSpace (Fig. 2a) and subsequently gated on B cell subsets for visualization (Fig. S4).

tSpace analysis defines developmental sequences leading to the 4 terminal tonsil B cell populations, IgG and IgA class switched memory cells and plasmablasts (PB). The first trajectory space principal component (tPC1) delineates the transition from naïve to germinal center cells (GCC, Fig. 2b); tPC2 the differentiation of memory or plasma cells; tPC3 and tPC4 pathways to IgA vs IgG class switched cells (Fig. 2c). Early naïve cells express CXCR5 which mediates lymphoid tissue entry and follicle homing (Fig S5). A broad strand of cells connects naïve IgD^+ B cells to the cluster of proliferating GC centroblasts and centrocytes (Fig. 2a, S5a-c). Along this path from naïve cells, IgD and IgM are downregulated as cells transition to $CD38^+CD77^+$ centroblasts. There are also clear trajectories from GCC to class switched PB and memory cells. CD27 is upregulated in memory B cell branches (Fig. S5d-f), along with trafficking receptor CLA (Fig. S5e-h) that is induced in response to immune challenge at squamous epithelial surfaces including the oral mucosa. CXCR3, implicated in lymphocyte homing to inflamed tissues[7,8] is coordinately 'upregulated' in the transition to memory cells as well. Appearance of these receptors along trajectories from GCC to memory cells suggest that CLA and CXCR3 may be induced in response to environmental signals in the 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. S5i-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. S5e-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 constrain or "force" cells into specific developmental sequences or paths, but instead positions each cell in a likely developmental context with all others even when cell transitions are biologically or phenotypically diffuse. 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. S6). While class switch recombination is normally attributed to the GC reaction, in some mouse models class switching can occur prior to GC formation, and it is observed in T-independent B cell responses as well[9]. tSpace analysis suggests that, in steady state human tonsil, activated B cells can undergo IgA or IgG class switching and conversion to memory cells without transiting through the GC reaction. In contrast to their class-switched counterparts, IgM memory cells ($CD27^+$, $CD38^-$) are more closely connected to naïve (IgM^+, IgD^+, $CD27^-$, $CD38^-$) cells in most tPCs, with tPC2 specifically expanding this trajectory (Fig. 2b). Thus, tSpace recapitulates known pathways of tonsil B cell development and differentiation, presents evidence that human B cells can follow alternative paths that have only been described in animal studies, and reveals the developmental stage(s) and transitions at which tissue and inflammation-specific trafficking receptors are induced.

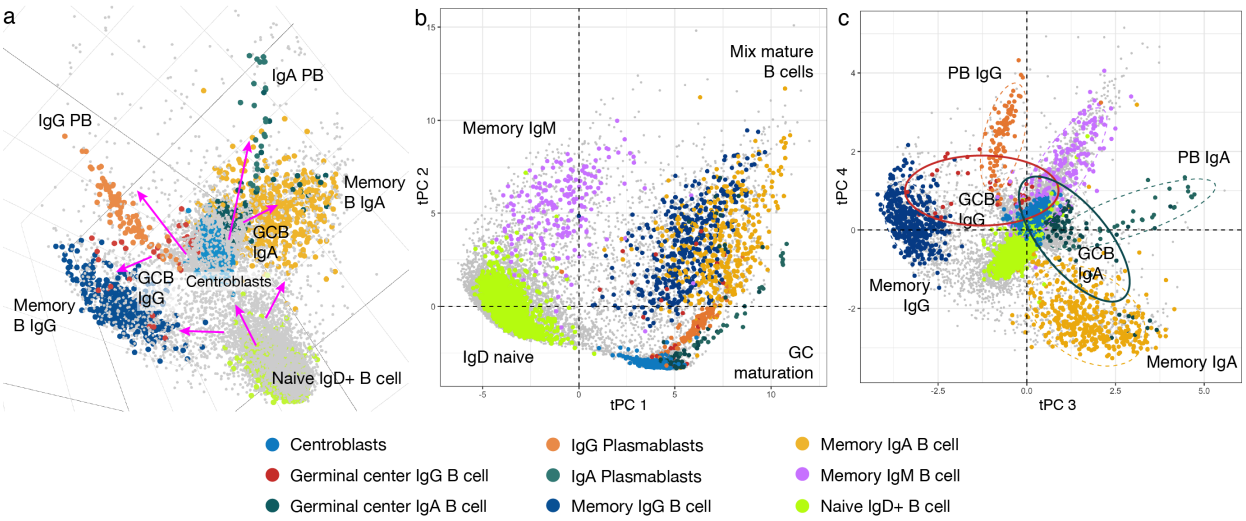

Figure 2. tSpace analysis of B cell differentiation in tonsils and inter-organ trajectories with blood. a tSpace unravels maturation paths of B cells starting from naïve B cells in tonsil throughout GC into memory B cells and plasmablasts (PB). Magenta arrows mark suggested directionalities based on 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.

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[10]. 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[11] 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 separate 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.*[10] using 2420 variable genes (Methods). tSpace clearly delineates absorptive/enterocyte and secretory/EE developmental paths, both arising via transit amplifying (TA) cells from Lgr5$^+$ CBC cells (Fig. 3a, Fig. S7c-d). Goblet and Paneth cells define short branches from the proliferating TA pool. tSpace positions *Dll1*-expressing cells in trajectory space between CBC cells and mature EE populations (Fig. 3a, shaded grey rectangle). These cells express genes that define short-lived secretory progenitors (sIEEP)[12], which upregulate EE lineage specification genes *Neurog3*, *Neurod1* and *Neurod2* (Fig. 3d)[13,14]. Consistent with their location in tSpace projection, sIEEP are well-documented precursors of EE cells[12–14]. The EE branch proceeds through EE3 cells, recently identified as EE intermediates, giving rise to specialized mature enteroendocrine subsets[10] (for cell labels see Methods and Fig. S8). Interestingly, sparse intermediates link a single tuft cell population to both CBC/TA and to EE3 cells. While the number of intermediates defining these two pathways to tuft cells would suggest caution in interpretation, it is noteworthy that a dual origin of tuft cells (directly from Lrg5+ CBC cells but also from EE3 enteroendocrine cells) has been proposed from multiple lines of evidence[10,15]. tSpace performed well in comparison with SPADE[16], a minimum spanning tree (MST) algorithm applied to visualize trajectory relationships in the original analysis of this scRNAseq dataset. SPADE[10] and tSpace both delineate the major CBC to enterocyte and EE branches, the relationship of goblet and Paneth cells to CBC/TA, and the terminal branching of EE subsets. However, SPADE provides a 2D representation of a MST structure, an approach that is inherently challenged by non-tree-

like developmental paths, such as the paths from EE3 and CBC that converge on tuft cells (Figs. 3a, S7c-d). While tSpace identifies a single tuft cell pool with dual connections, SPADE analysis forced tufts cells into two disconnected populations, one arising from CBC cells and the other from intermediates leading to EE3 cells. SPADE also failed to detect or properly position slEEP on the path to EE cells.[10] In contrast to tSpace which positions each cell in trajectory space, SPADE and related MST algorithms rely on prior definition of cell clusters and limited gene sets, features which run the risk of missing or mislabeling important cell intermediates[10]. slEEP were defined either as cycling CBC or goblet cells in the published analysis, and were subsumed in biologically inappropriate clusters (Figs. S8a, S9a).

The ability of tSpace to position all cells (not just clusters or cluster centers) in developmental relationships allows additional interesting insights. tSpace trajectories reveal a common 'trunk' leading to secretory and absorptive branches, but many sc-CBC and c-TA cells (defined Fig. S8) actually segregate to the early EE or enterocyte branches, suggesting that they are already developing towards if not committed to EE or enterocyte fates. To illustrate the application of tSpace to explore developmental progression of gene expression in this context, we isolated trajectories within the tSpace distance matrix starting from the Lrg5+ CBC cell population, gated on cells of the enterocyte branch and cells within the early segment of the EE branch preceding EE3 (Fig. 3a-c), and plotted gene expression of cells vs their distance from CBC cells (Fig. 3d). We focused initially on genes for known hallmarks of intestinal differentiation (Fig. 3d, S9a)[17]. The analysis confirms *Ascl2*[18], *OlfM4*[19] and *Prom1*[20] as robust markers of the presumptive crypt populations (CBC to TA cells) and reveals that the expression of *Prom1* extends into the TA pool, confirming previous findings[21] (Fig. 3d, S9a). 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[22] it may contribute to enterocyte fate or specification.

Patterns of expression of transcription factors (TF) along the trajectories suggest combinations of them that might specify fate within these early progenitors, and/or control downstream cell specialization. At least 4 distinct TF modules were identified (Methods) based on their patterns of regulation along the EE or enterocyte branches (M1-M4, Fig. 3e, S9c). Genes for proliferation and DNA maintenance (M1, e.g. *Ccna2, Cdk2, Fancd2, Rbl1*) are expressed by dividing sc-CBC and TA "early" along the trajectory, as expected. A second module of TF genes is also expressed by early cells but is maintained selectively in the EE branch: these include TF's associated with endocrine and pancreatic development (e.g. *Foxa2, Foxa3, Neurog3, Sox4, Sox9*) that may coordinate secretory pathways within intestinal enteroendocrine cells[23]. Interestingly, among these, tSpace reveals an unexpectedly high and selective expression of *Sox4* in slEEP cells, suggesting it as a novel candidate contributor to EE specification[24] (Fig. 3e, S8c, e). Modules 3 and 4 TFs are expressed preferentially in the enterocyte branch. They include TF involved in lipid and cholesterol metabolism required for mature enterocytes (e.g. *Cebpb, Klf5, Nr5a2,* Fig. 3e, S9c)[25,26], but also *Nfe2l2* and *Maf* associated with the activation of *Nfe2l2/Nrf2*-antioxidant response element (ARE) pathway[27]. Enterocytes utilize short fatty acids as a source of energy, and fatty acid metabolism generates reactive oxygen species (ROS), and ROS are also abundant in the intestinal lumen[28]. Upregulation of the *Nfe2l2*-ARE pathway may help protect differentiating enterocytes from oxidative damage[28]. The analysis also identifies *Isx* and *Ski* (both within M3) as putative markers of TA cells selectively within the early enterocyte developmental branch.

Overall, the patterns of gene expression and cell positioning in trajectory space reflect observations from decades of research on intestinal development (Fig. 3d, S9a), but also suggest refinements to current

understanding (e.g. many dividing c-TAs in the crypt already express gene programs leading to secretory vs. absorptive phenotypes). tSpace analyses also highlight novel pathways and mechanisms of cell differentiation, including genes for transcription factors and receptors that may regulate cell fate (Fig. 3f).

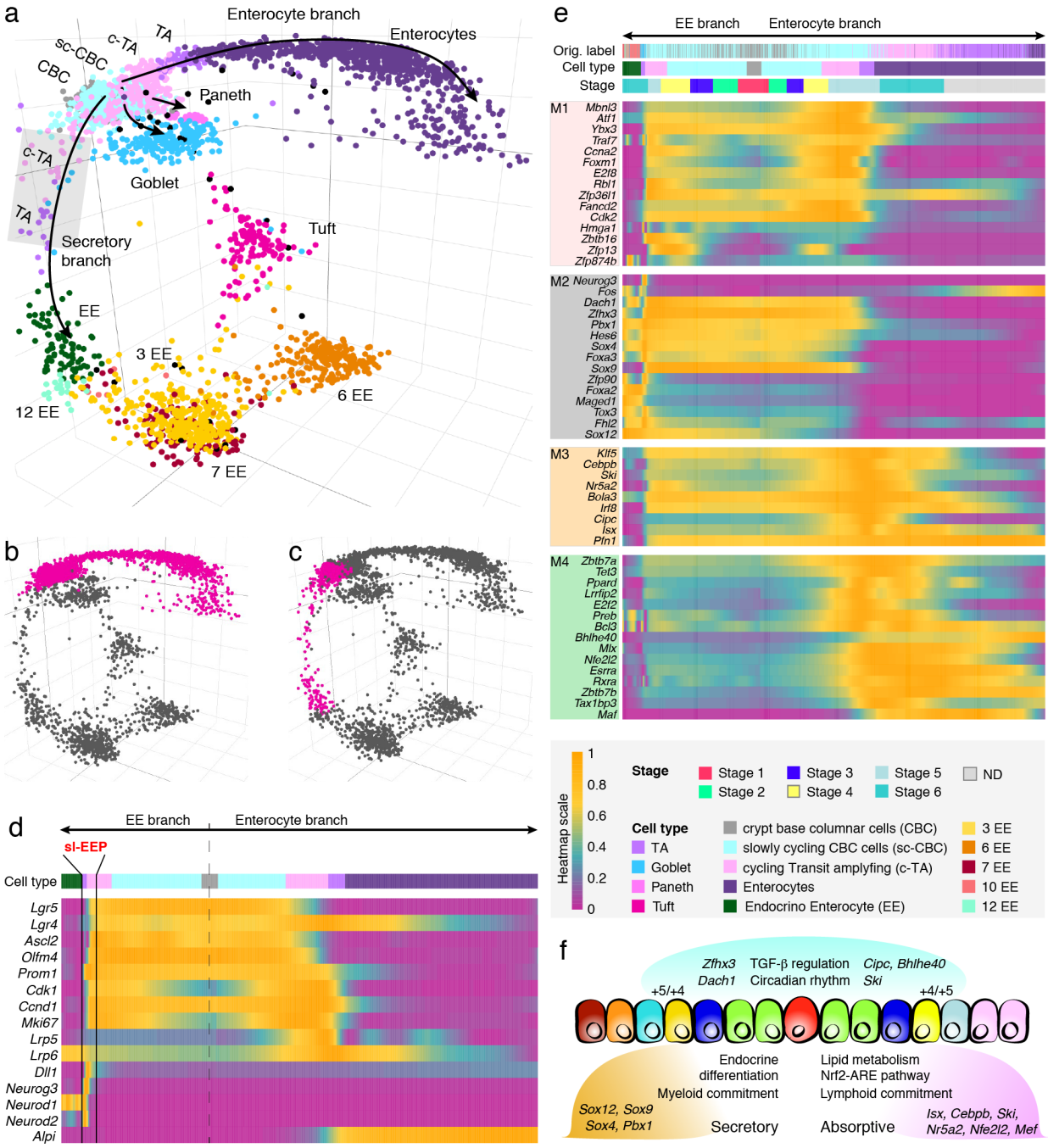

Figure 3. tSpace analysis of mouse small intestinal epithelial cells based on scRNAseq data. a tSpace separates trajectories to enterocytes, enteroendocrine (EE), Paneth and goblet cells. CBC and TA subsets were defined by our analysis, as described in Fig. S8., other subsets are labeled as in Yan *et al.*[10]. Shaded rectangle highlights the position of short-lived EE progenitors (slEEP) cells. **b** Isolated enterocyte trajectory. **c** Isolated EE trajectory. **d** Expression patterns of selected genes[17] (known markers or regulators of intestinal crypt development; expanded gene list in Fig. S9a) along the isolated trajectories. **e** Four detected transcription factor modules in early trajectories, identified by comparing gene expression between cells at similar stages in the two trajectories (Methods): M1

comprises TFs involved in cell cycle and genome integrity expressed in precursor populations (early in the shared trajectory). M2 and M3-M4 differentiate the two lineages and comprise TF's that may determine cell fate or specialization (see text). Cell stage (Methods) and cell identities defined here (Cell type) or in Yan *et al*. (Original labels) are indicated above the heatmap. ND –fully differentiated enterocytes, not used in trajectory alignment. **f** Summary of differences between two branches suggested by gene regulation along the trajectories. 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. Expression of *Dll1* and *Sox4* in EE development and *Apli* in enterocyte differentiation, mark specific progenitor cells located within the +4/+5 position in the intestinal crypt according to the literature[12,24,29]; clear peaks are seen in their expression along the trajectories (Fig. 3d-e) near the TA to differentiated cell transitions, likely representing these specific progenitor populations.

The tSpace approach is conceptually similar to that of isomap[30]. Both methods provide a global approach to dimensionality reduction, designed to preserve manifold geometry at all scales. Both algorithms determine geodesic distances along a KNN graph. Isomap embeds the resulting distance matrix in low dimensions using multidimensional scaling (MDS). It has been successfully applied to diverse high dimensional datasets[31–34], but it has not been adopted for high dimensional single cell analyses, perhaps because of well-described limitations. The algorithm is computationally expensive. This has been addressed in part by 'landmark isomap', which approximates the large global computation in isomap by calculating distances from a set of randomly selected 'landmark' cells. To ensure uniform sampling of the manifold, we modify this approach in tSpace by selecting individual cells from each of T K-means clusters, where T is the number of trajectories to be calculated. We show that linear trajectories (distance vectors) calculated from 100 – 250 well-distributed starting cells are sufficient to recapitulate cellular relationships in each of the datasets here. Isomap suffers also from sensitivity to "short circuit" errors if K is too large or if noise in the data positions cells aberrantly between valid branches or populations in the manifold. Short circuits pose a problem with the Dijkstra algorithm, used in isomap to calculate shortest paths between cells. We take advantage of the Wanderlust, which refines distances and avoids "short circuits" by using subgraph averaging and weighting in shortest path calculations based on waypoints[1]. We show that tSpace with Wanderlust improves the definition of developmental paths (Fig. S1). Finally, isomap uses MDS, a memory intensive algorithm, for dimensionality reduction and visualization of manifold relationships. Based on our desire as biologists for display modalities that allow robust exploration of the manifold from different perspectives, we prefer principal component projections of the tSpace matrix. PCA is computationally inexpensive. The first 3 tPCs often embody the most important developmental branches, but higher tPCs can also reveal critical biological processes. We illustrate for example the parallel pathways of IgA vs IgG memory and plasma cell development in tonsil B cells, which dominate the 4th tPC. The R version of tSpace also implements dimensionality reduction with UMAP, and we show for our scRNAseq example that UMAP embedding of trajectory space reveals developmental branches better than UMAP embedding of the original gene expression matrix.

A number of other methods for trajectory inference have been described and compared[35,36]. We highlight some of the practical features and limitations of tSpace and of published methods in Supplementary Table 3. tSpace has advantages over algorithms that use memory intensive MST methods to define branch points, as for large datasets these depend on downsampling of cells or calculation of relationships between clusters (rather than individual cells) to reduce computational complexity. Examples include slingshot[37], p-Creode[38] and SPADE[39]. As highlighted above in discussion of published analysis of intestinal epithelial cells, downsampling in MST-based methods holds inherent risks of obscuring important cell subsets, and in most algorithms fails to position each cell in developmental relationships.

tSpace avoids the loss of individual cell resolution associated with cell downsampling, while retaining the ability to reveal the developmental relationships of all cells to each other. Algorithms that focus on computationally defining branchpoints and tree structures can also limit appreciation of alternative pathways of differentiation represented by cells that bridge between dominant pathways. Moreover, in contrast to algorithms that rely only on local cell relationships, the global approach of tSpace estimates distant as well as nearby cell relationships. Indeed, the algorithm exports a dense matrix of meaningful cell-to-cell distances that represent measures of the extent of phenotypic change along developmental pathways. As illustrated in our examples, cells along specific developmental pathways and branches can be easily gated (isolated) in plots of tPCs using commonly available software such as Flowjo, JMP or in R (Methods). Trajectories starting from branch termini or other desired points within pathways are readily identified within the tSpace matrix, and plotted vs gene/protein expression to characterize changes in cell phenotypes along isolated developmental sequences (as in Figs. 1-3).

In conclusion, we have presented the concept of trajectory space and its implementation in the tSpace algorithm for elucidation of branching or convergent developmental pathways and mechanisms from single cell profiles. tSpace performs well across different biological systems and platforms and reveals known and novel biology. tSpace embodies a combination of useful features including 1) applicability to any type of data (proteomic, transcriptomic, etc.); 2) simplicity of use; 3) scalability and independence from the need for downsampling; 4) robustness to input parameters; 5) positioning of each individual cell in developmental relationships (allowing visualization of alternative or minor pathways of differentiation); 6) retention of distant as well as local cell relationships; and 7) independence from requirements of clustering or prior information. The tSpace outputs are reproducible, intuitive and amenable to exploration of biology (gene or protein expression, trajectory isolation, etc.). We believe that tSpace will prove useful to the rapidly growing field of singe cell analysis.

Funding. This work was supported by NIH grants R37-AI047822, R01-AI093981 and R01-CA228019 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).

The authors have declared no conflict of interest.

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

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. Shah, D. & Zuniga-Pflucker, J. An Overview of the Intrathymic Intricacies of T Cell Development. *J Immunol* **192,** 4017–4023 (2014).

3. 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 CD3ε. *Int Immunol* **21,** 1317–1327 (2009).

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

5. Silva, N. S. & Klein, U. Dynamics of B cells in germinal centres. *Nat Rev Immunol* **15,** 137–148 (2015).

6. Dufaud, C. R., McHeyzer-Williams, L. J. & McHeyzer-Williams, M. G. Deconstructing the germinal center, one cell at a time. *Curr Opin Immunol* **45,** 112–118 (2017).

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

8. Rott, L., Briskin, M., Andrew, D., Berg, E. & Butcher, E. 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 Baltim Md 1950* **156,** 3727–36 (1996).

9. Stavnezer, J. & Schrader, C. E. IgH Chain Class Switch Recombination: Mechanism and Regulation. *J Immunol* **193,** 5370–5378 (2014).

10. 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).

11. 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).

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

13. 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).

14. 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).

15. Gerbe, F., Legraverend, C. & Jay, P. The intestinal epithelium tuft cells: specification and function. *Cell Mol Life Sci* **69,** 2907–2917 (2012).

16. Anchang, B. *et al.* Visualization and cellular hierarchy inference of single-cell data using SPADE. *Nat Protoc* **11,** 1264 (2016).

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

18. van der Flier, L. G. *et al.* Transcription Factor Achaete Scute-Like 2 Controls Intestinal Stem Cell Fate. *Cell* **136,** 903–912 (2009).

19. 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).

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

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

22. 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).

23. Consortium, T. *et al.* Single-cell transcriptomics of 20 mouse organs creates a Tabula Muris. *Nature* **562,** 367–372 (2018).

24. Gracz, A. D. *et al.* SOX4 Promotes ATOH1-independent Intestinal Secretory Differentiation Toward Tuft and Enteroendocrine Fates. *Gastroenterology* (2018). doi:10.1053/j.gastro.2018.07.023

25. Yen, C.-L., Nelson, D. W. & Yen, M.-I. Intestinal Triacylglycerol Synthesis in Fat Absorption and Systemic Energy Metabolism. *J Lipid Res* **56,** jlr.R052902 (2014).

26. Degirolamo, C., Sabbà, C. & Moschetta, A. Intestinal nuclear receptors in HDL cholesterol metabolism. *J Lipid Res* **56,** 1262–1270 (2015).

27. Itoh, K. *et al.* An Nrf2/Small Maf Heterodimer Mediates the Induction of Phase II Detoxifying Enzyme Genes through Antioxidant Response Elements. *Biochem Bioph Res Co* **236,** 313–322 (1997).

28. Ferrebee, C. B. *et al.* Organic Solute Transporter α-β Protects Ileal Enterocytes From Bile Acid–Induced Injury. *Cell Mol Gastroenterology Hepatology* **5,** (2018).

29. 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).

30. Tenenbaum, J. B., de Silva, V. & Langford, J. C. A Global Geometric Framework for Nonlinear Dimensionality Reduction. *Science* **290,** 2319–2323 (2000).

31. Silva, V. D. & Tenenbaum, J. B. Global Versus Local Methods in Nonlinear Dimensionality Reduction. 721–728 (2003).

32. Mahecha, M. D., Martínez, A., Lischeid, G. & Beck, E. Nonlinear dimensionality reduction: Alternative ordination approaches for extracting and visualizing biodiversity patterns in tropical montane forest vegetation data. *Ecol Inform* **2,** 138–149 (2007).

33. Hannachi, A. & Turner, A. G. Isomap nonlinear dimensionality reduction and bimodality of Asian monsoon convection. *Geophys Res Lett* **40,** 1653–1658 (2013).

34. Stamati, H., Clementi, C. & Kavraki, L. E. Application of nonlinear dimensionality reduction to characterize the conformational landscape of small peptides. *Proteins Struct Funct Bioinform* **78,** 223–235 (2010).

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

36. Saelens, W., Cannoodt, R., Todorov, H. & Saeys, Y. A comparison of single-cell trajectory inference methods: towards more accurate and robust tools. *Biorxiv* 276907 (2018). doi:10.1101/276907

37. Street, K. *et al.* Slingshot: cell lineage and pseudotime inference for single-cell transcriptomics. *Bmc Genomics* **19,** 477 (2018).

38. Herring, C. A. *et al.* Unsupervised Trajectory Analysis of Single-Cell RNA-Seq and Imaging Data Reveals Alternative Tuft Cell Origins in the Gut. *Cell Syst* (2017). doi:10.1016/j.cels.2017.10.012

39. Qiu, P. *et al.* Extracting a cellular hierarchy from high-dimensional cytometry data with SPADE. *Nat Biotechnol* **29,** 886 (2011).