

1 **Title: Elucidating the activation mechanisms for bifurcation**
2 **of regulatory and effector T cell fates by multidimensional single cell**
3 **analysis**

4

5

6

7

8 **Authors:**

9 Alla Bradley¹, Tetsuo Hashimoto¹ and Masahiro Ono^{2§}

10

11 **Affiliations:**

12 ¹ Faculty of Life and Environmental Sciences, University of Tsukuba,
13 Tsukuba, Japan

14 ² Department of Life Sciences, Faculty of Natural Sciences, Imperial College
15 London, Sir Alexander Fleming Building, Imperial College Road, London,
16 SW7 2AZ, United Kingdom

17

18 § Address correspondence and reprint requests to Dr Masahiro Ono,
19 Department of Life Sciences, Sir Alexander Fleming Building, Imperial College
20 Road, South Kensington, London, SW7 2AZ, United Kingdom. E-mail
21 address: m.ono@imperial.ac.uk

22

23

Abstract

2

3 In T cells, T cell receptor (TCR) signalling initiates downstream transcriptional
4 mechanisms for T cell activation and differentiation. Foxp3-expressing
5 regulatory T cells (Treg) require TCR signals for their suppressive function
6 and maintenance in the periphery. It is, however, unclear how TCR signalling
7 controls the transcriptional programme of Treg. Since most of studies
8 identified the transcriptional features of Treg in comparison to naïve T cells,
9 the relationship between Treg and non-naïve T cells including memory-
10 phenotype T cells (Tmem) and effector T cells (Teff) is not well understood.
11 Here we dissect the transcriptomes of various T cell subsets from
12 independent datasets using the multidimensional analysis method Canonical
13 Correspondence Analysis (CCA). We show that resting Treg share gene
14 modules for activation with Tmem and Teff. Importantly, Tmem activate the
15 distinct transcriptional modules for T cell activation, which are uniquely
16 repressed in Treg. The activation signature of Treg is dependent on TCR
17 signals, and is more actively operating in activated Treg. Furthermore, by
18 analysing single cell RNA-seq data from tumour-infiltrating T cells, we
19 revealed that FOXP3 expression occurs predominantly in activated T cells.
20 Moreover, we identified FOXP3-driven and T follicular helper (Tfh)-like
21 differentiation pathways in tumour microenvironments, and their bifurcation
22 point, which is enriched with recently activated T cells. Collectively, our study
23 reveals the activation mechanisms downstream of TCR signals for the
24 bifurcation of Treg and Teff differentiation and their maturation processes.

1 Introduction

2 T cell receptor (TCR) signalling activates NFAT, AP-1, and NF- κ B (1), which
 3 induces the transcription of Interleukin (IL)-2 and IL-2 receptor (R) α -chain
 4 (*Il2ra*, CD25). IL-2 signalling induces further T cell activation, proliferation and
 5 differentiation (2). In addition, IL-2 signalling has key roles in immunological
 6 tolerance (2). This is partly mediated through CD25-expressing regulatory T
 7 cells (Treg), which suppress the activities of other T cells (3). Intriguingly, TCR
 8 signalling also induces the transient expression of FoxP3, the lineage-specific
 9 transcription factor of Treg (4), in any T cells in humans (5), and in mice in the
 10 presence of IL-2 and TGF- β (6). These suggest that FoxP3 can be actively
 11 induced as a negative feedback mechanism for the T cell activation process,
 12 especially in inflammatory conditions in tissues (7). Thus, the T cell activation
 13 processes may dynamically control Treg phenotype and function during
 14 immune response and homeostasis.

15 In fact, TCR signalling plays a critical role in Treg. Studies using TCR
 16 transgenic mice showed that Treg require TCR activation for *in vitro*
 17 suppression (8). The binding of Foxp3 protein to chromatin occurs mainly in
 18 the enhancer regions that have been opened by TCR signals (9). In fact,
 19 continuous TCR signals are required for Treg function, because the
 20 conditional deletion of the TCR- α chain in Treg abrogates the suppressive
 21 activity of Treg and eliminates their activated or effector-Treg (eTreg)
 22 phenotype (10, 11). It is, however, unclear how TCR signals contribute to the
 23 Treg-type transcriptional programme, and whether TCR signals are operating

1 in all Treg cells or whether these are required only when Treg suppress the
2 activity of other T cells.

3 Heterogeneity of Treg has been previously addressed through classifying
4 Treg into subpopulations, according to the origin (thymic Treg, peripheral
5 Treg, visceral adipose tissue Treg (12)), the transcription factor expression
6 and ability to control inflammation (Th1-Treg (13) and Th2-Treg (14), and T
7 follicular regulatory T cells (15)), and their activation status (activated
8 Treg/eTreg, resting Treg, and memory-type Treg (16)). Among these Treg
9 subpopulations, of interest is eTreg, which are activated and functionally
10 mature Treg. Murine eTreg can be identified by memory/activation markers
11 such as CD44, CD62L, and GITR (16, 17), and their differentiation is
12 controlled by the transcription factors Blimp-1, IRF4 and Myb (18, 19). Human
13 Treg can be classified into naïve Treg (CD25+CD45RA+Foxp3+) and eTreg
14 (CD25+CD45RA-Foxp3+) (20). However, our recent computational study
15 showed that classical gating approach is not effective for understanding
16 multidimensional data, and that marker expression data may be rather
17 effectively analysed by the computational clustering approaches that aim to
18 understand the dynamics of marker expressions in Treg (21). Furthermore,
19 the recent advancement of single cell technologies has opened the door to
20 address the heterogeneity of Treg by their gene profiles at the single cell
21 level.

22 When addressing the single cell level heterogeneity, it is critical to analyse
23 activated effector T cells (Teff) and memory-like T cells (memory-phenotype T
24 cells; Tmem) together with Treg. The surface phenotype of Tmem is

1 CD44^{high}CD45RB^{low}CD25⁻ (22), which is similar to CD25⁻ Treg, apart from
 2 Foxp3 expression and suppressive activity (23, 24). In addition, Teff express
 3 CD25 and CTLA-4 (25), the latter of which is also known as a Treg marker
 4 (26). Tmem may include both antigen-experienced memory T cells (27) and
 5 self-reactive T cells (28). In fact, CD44^{high}CD45RB^{low} Tmem do not develop in
 6 TCR transgenic mice with the *Rag* deficient background, indicating that they
 7 require agonistic TCR signals in the thymus (29). In addition, a study using a
 8 fate-mapping approach showed that a minority of Treg naturally lose Foxp3
 9 expression and join the Tmem fraction (30). These suggest that, upon
 10 encountering cognate self-antigens, self-reactive T cells, which include Tmem
 11 and Treg, express and sustain Foxp3 expression as a negative feedback
 12 mechanism for strong TCR signals (7). Thus, Treg have a close relationship
 13 with Tmem and Teff. However, since most studies used naïve T cells (Tnaïve)
 14 as the control for Treg, many of known Treg-associated features may be in
 15 fact shared with Tmem and Teff.

16 Multidimensional analysis is an effective approach to address this problem,
 17 allowing to systematically investigate the relationships between more than two
 18 cell populations (e.g. based on transcriptional similarities) (31). The prototype
 19 methods include Principal Component Analysis (PCA), Correspondence
 20 Analysis (CA) (32) and Multidimensional Scaling (33). In the application to
 21 genomic data, these methods measure distances (i.e. similarities) between
 22 samples and/or genes using different metrics, and thereby visualise the
 23 relationships between samples and/or genes in a reduced dimension, typically
 24 either in two-dimensional (2D) or three-dimensional (3D) space, providing
 25 means to explore and investigate data (31). However, these multidimensional

1 methods are often not sufficiently powerful for hypothesis-driven research,
 2 and our previous studies developed a transcriptome analysis method using a
 3 variant of CA, Canonical Correspondence Analysis (CCA) for microarray data
 4 (31) and RNA-seq data (34). In this approach, two transcriptome datasets are
 5 canonically analysed: the correlations between cell samples in one dataset
 6 (main dataset) and the immunological processes (explanatory variables) in
 7 another dataset (explanatory dataset) are analysed based on their
 8 correlations to individual genes. Briefly, CCA uses a linear regression to
 9 identify the interpretable part (constrained space) of main data by explanatory
 10 variables, and visualises similarities between genes, cells, and explanatory
 11 variables using a singular value decomposition (SVD) solution within the
 12 interpretable space (34). Thus, CCA enables to investigate and identify the
 13 unique features of each T cell population, visualising the relationships
 14 between T cell populations.

15 In this study, we investigate the multidimensional features of Treg in
 16 comparison to other CD4⁺ T cells including Teff, Tmem, and naïve T cells
 17 under normal or pathological conditions. Here we aim to identify the
 18 differential regulation of transcriptional modules for T cell activation and
 19 differentiation in these populations, and to reveal systems and molecular
 20 mechanisms behind the differential regulation. Furthermore, using our new
 21 single cell combinatorial CCA (SC4A) approach, we investigate the single-cell
 22 level heterogeneity of CD4⁺ T cells including Treg and effector T cells,
 23 identifying the differentially regulated gene modules and the bifurcation point
 24 for T cell fates.

25

1 **Materials and Methods**

2

3 *Conventional CCA (Gene-oriented analysis)*

4 CCA canonically analyses two independent microarray or RNA-seq datasets
5 (34). Briefly, gene expression data of the standardised main dataset (**S**) is
6 linearly regressed onto the explanatory variable(s) (**D**), which identifies the
7 interpretable part of the main dataset ("Constrained data", **S***). When only one
8 explanatory variable is used, the CA algorithm of CCA assigns numerical
9 values to cell samples and genes so that the dispersion of samples is
10 maximised (uncorrelated information components), providing a one-
11 dimensional solution (34). The correlation analysis of explanatory data and
12 gene scores in the CA solution generates a biplot value, which, in the one-
13 dimensional solution, represents the explanatory variable score. In the one-
14 dimensional CCA solution, the single biplot value can either be +1 or -1,
15 indicating the direction (increasing/decreasing) of correlated genes in the
16 explanatory variable against that in the main dataset. In order to use the one-
17 dimensional solution as a scoring system, the CCA score (i.e. Axis 1 score) is
18 multiplied by the single biplot value, which indicates positive or negative
19 correlation to Axis 1, ensuring that cells and genes with high scores have high
20 positive correlations to the explanatory variable. When two or more
21 explanatory variables are used, the CA algorithm then performs SVD on **S***,
22 creating new matrices (i.e. sample and gene score matrices). These scores
23 are sorted into new uncorrelated axes α_k , along which the entire set of scores
24 generated by SVD is distributed. The first axis has the largest variation
25 (*inertia*) and thereby explains the greatest amount of information extracted by

1 the analysis. The map approach enables the comparison of two or more
2 explanatory variables, while the regression process in CCA allows the
3 analysis across two different experiments (34). Biplot values of the CCA result
4 are shown by arrows on the CCA map. CCA provides a map that shows the
5 correlations between samples of interest, explanatory variables, and genes.
6 Highly correlated components are closely positioned on the map.

7

8 Note that the same genes must be used in both transcriptome dataset
9 matrices. The main dataset is projected onto the explanatory variable dataset,
10 thus the genes in common to both datasets comprise the interpretable part
11 (intersect) of the main data. Mathematical operation implemented in the CCA
12 algorithm produces immunological process (explanatory variable), gene and
13 cell sample scores. The results are visualized as the 3-dimensional CCA
14 solution on the CCA map (i.e. CCA triplot) that shows the relationships
15 between cell subsets, genes and immunological processes. For example, in
16 the application of CCA to population-level (bulk) data, transcriptome datasets
17 of peripheral CD4+ T cells (including Treg, naïve, memory, draining LN and
18 non-dLN and tissue effector CD4+ T cells) were processed by CCA using
19 indicated explanatory variables (e.g. T cell activation) and the cross-level
20 relationships between components at three different levels (immunological
21 process, gene and cell) were analysed.

22

23 *Single cell data pre-processing and single cell CCA (single-cell oriented*
24 *analysis)*

1 RNA-seq expression data of GSE72056 was obtained from single-cell
2 suspension of tumour cells with unknown activation and differentiation
3 statuses (35). Genes with low variances and low maximal values were
4 excluded. In order to identify CD4⁺ T cells, single cell data were filtered by the
5 expression of *CD4* and *CD3E* to obtain only the *CD4*⁺*CD3E*⁺ single cells, and
6 also by k-means clustering of PCA gene plot to exclude outlier cells (21) for
7 subsequent analysis.

8 In the application of CCA to single cell data, importantly, the same single cells
9 are used in both main data and the explanatory variables (i.e. selected
10 genes). The main dataset is projected onto the explanatory variables,
11 visualising the relationships between single cells, genes and explanatory
12 variables, which represent major activation/differentiation processes in the
13 dataset.

14

15 *Explanatory variables for conventional CCA*

16 Explanatory variables for CCA were prepared as follows. Differentially-
17 expressed genes were selected by a moderated t-test result using the
18 Bioconductor package, *limma*. The top-ranked differentially expressed genes
19 (according to their *p*-values) were used for making the explanatory variables.
20 The T cell activation explanatory variable (*Tact*) was defined by the difference
21 in gene expression between anti-CD3/CD28-stimulated (17 h) CD4⁺ T cells
22 and untreated naïve CD4⁺ T cells from GSE42276 (36). Precisely, genes were
23 selected by FDR < 0.01 and log2 fold change (> 1 or < -1) in the comparison
24 of the gene expression profile of the activated and resting T cells. For the one-

1 dimensional CCA of T cell populations (Figure 1B), the expression data of
 2 GSE15907 (37) was regressed onto gene values in *Tact* representing the
 3 change in gene expression following T cell activation, and CA was performed
 4 for the regressed data and the correlation analysis was done between the
 5 new axis and the explanatory variable. For the two-dimensional CCA of T cell
 6 populations (Figure 2A and 2B), the expression data of GSE15907 (37) was
 7 regressed onto *Tact*, in combination with *Foxp3* and *Runx1* explanatory
 8 variables representing the effects of *Foxp3* and *Runx1* expression on CD4+ T
 9 cells (GSE6939 (38)). *Foxp3* explanatory variable is the log2 fold change of
 10 *Foxp3*-transduced naïve CD4+ T cells and empty vector-transduced CD4+T
 11 cells. *Runx1* explanatory variable is the log2 fold change of *Runx1*-transduced
 12 naïve CD4+ T cells and empty vector-transduced CD4+T cells. Subsequently,
 13 CA was applied to the regressed data and the correlation analysis was
 14 performed between the new axes and the explanatory variables.

15

16 SC4A

17 SC4A is a composite approach to understand non-annotated single cell data
 18 by identifying distinct populations of cells and the differentiation processes
 19 that are correlated with these populations. Since T cell population is usually
 20 identified by a single lineage specification factor, in the application of SC4A,
 21 such a factor will represent the cell population and their differentiation process
 22 (**Supplementary Figure 1A**). The advantage of this approach is that SC4A
 23 uses the gene expression data of a part of the dataset analysed, and thus the
 24 regression analysis of CCA becomes more efficient because of the absence

1 of between-experimental variation, which is usually significant in cross-dataset
2 analysis (34).

3 SC4A is performed by the following three steps: (1) identification of putative
4 cell populations and candidate genes for explanatory variables by standard
5 CCA; (2) combinatorial CCA to identify the top-ranked genes to be used as
6 explanatory variables; and (3) the final CCA solution using the selected genes
7 as explanatory variables.

8 *1. Preliminary analysis*

9 The aim of the preliminary analysis is to identify the putative cell populations
10 and candidate genes for explanatory variables, and standard CCA is a useful
11 method to do this, because candidate genes can be identified by their
12 correlation to each putative cell population. Considering that the final output is
13 most effectively understood by visualisation using 2-dimensional (showing
14 correlation between explanatory variable, samples and genes) or 3-
15 dimensional (showing correlation between explanatory variable and
16 samples/genes) plot, up to 4 cell populations will be identified, and up to 5- 10
17 genes for each population will be identified by their correlation to the
18 population (**Supplementary Figure 1B**).

19 *2. Combinatorial CCA*

20 Here SC4A aims to identify a set of genes that make the dispersion of cell
21 populations maximum in the CCA solution. To achieve this, all the
22 combinations of genes will be used as explanatory variables and tested for
23 discriminating each two populations using CCA. During each combinatorial

1 cycle, two genes are chosen from the total selected genes for all defined
2 single-cell populations in the main dataset and tested for their correlations to
3 one defined cell population vs all other T cells.

4 Correlation can be visualised as the degree angle measured between the
5 explanatory variable (gene) and the centroid of the defined cell population.
6 Out of the two genes, the gene with the smallest angle to the defined cell
7 population is the most correlated. All selected genes are tested in this
8 pairwise manner against all defined cell populations vs all other T cells to
9 identify the gene that is most highly and uniquely correlated to each defined
10 cell population. At each combinatorial CCA, the most correlated gene to each
11 cell population is identified using vector multiplication (**Supplementary Figure**
12 **1C**). The top-ranked genes are determined by F1 score (the harmonic mean
13 of precision and sensitivity) and the correlation to the population of interest.
14 When the top-ranked gene is different between F1 score and the correlation,
15 the most immunologically-meaningful gene can be chosen.

16 3. SC4A

17 Finally, CCA is performed using the genes that are selected by the
18 combinatorial CCA to be used as explanatory variables. Thus, the single cell
19 dataset will be explained by the expression of the set of chosen genes, each
20 of which uniquely correlates with a cell population and represents the
21 differentiation process of the cell population (**Supplementary Figure 1D**).

22 Algebraically, SC4A is defined as follows. Single cell RNA-seq data $\mathbf{X} \in \mathbf{R}^p \times$
23 m is the measurement of m genes from p single cells. The j -th column $x_j = (x_{1j}$

1 $x_{2j} \dots x_{pj})^T$ is the expression data of the j -th gene from p single cells, where T
2 indicates transposed vector. In SC4A, by choosing a set of k genes for
3 explanatory variable, $\mathbf{X}' \in \mathbf{R}^{p \times (m-k)}$ will be analysed by CCA using $\mathbf{Z} \in \mathbf{R}^{p \times k}$
4 as explanatory variables. As in the algorithm for CCA, \mathbf{X}' is standardised by
5 column sums (\mathbf{c}) and row sums (\mathbf{r}), i.e. $\mathbf{S} = \mathbf{D}_r^{-1/2} (1/n \mathbf{X} - \mathbf{r}\mathbf{c}^T) \mathbf{D}_c^{-1/2}$, where n
6 is the grand total of gene expression data, \mathbf{D}_r and \mathbf{D}_c are the diagonal
7 matrices of \mathbf{r} and \mathbf{c} , respectively. Meanwhile, \mathbf{Z} is scaled and standardised
8 (i.e. mean = 0 and variance = 1). \mathbf{S} is linearly regressed onto \mathbf{Z} by the
9 projection matrix $\mathbf{Q} = \mathbf{D}_r^{1/2} \mathbf{Z} (\mathbf{Z}^T \mathbf{D}_r \mathbf{Z})^{-1} \mathbf{Z}^T \mathbf{D}_r^{1/2}$, and the constrained space \mathbf{S}^*
10 = $\mathbf{Q} \mathbf{S}$. Next, CCA performs singular value decomposition (SVD) of $\mathbf{S}^* = \mathbf{U} \mathbf{D}_\alpha$
11 \mathbf{V}^T , where $\mathbf{U}^T \mathbf{U} = \mathbf{V}^T \mathbf{V} = \mathbf{I}$, and \mathbf{D}_α is the diagonal matrix of singular values in
12 descending order ($\alpha_1 \geq \alpha_2 \geq \dots$). Thus, SVD analyses the constrained space
13 and provides new axes where the dispersion of samples and that of genes are
14 maximised in the first axes. Gene scores are defined as $\mathbf{D}_r^{-1/2} \mathbf{U} \mathbf{D}_\alpha$ or $\mathbf{D}_r^{-1/2} \mathbf{U}$,
15 and sample scores for single cells are defined by weighted average scores
16 (WA scores) $\mathbf{S} \mathbf{V} \mathbf{D}_\alpha$, or $\mathbf{S} \mathbf{V}$.

17 4. Choice of explanatory variables by SC4A

18 SC4A aims to identify a set of genes that make the dispersion of cell
19 populations maximum in the CCA solution. To achieve this, all the
20 combinations of genes will be used as explanatory variables and tested for
21 discriminating each two populations using CCA. During each combinatorial
22 cycle, two genes are chosen from the total selected genes for all defined
23 single-cell populations in the main dataset and tested for their correlations to
24 one defined cell population vs all other T cells. In the analysis of Figure 8, the

1 following two cell populations were analysed by the combinatorial CCA: (1)
 2 Activated T cells vs Resting T cells; (2) FOXP3+ cells vs FOXP3- cells; (3)
 3 BCL6+ cells (as Tfh-like T cells) vs BCL6- cells. The most correlated gene to
 4 each population (Activated T cells, Resting T cells, FOXP3+ cells, or BCL6+
 5 cells) was identified, and these 4 genes were used as explanatory variables in
 6 the final output of SC4A in Figure 8.

7

8 *Data pre-processing and other statistical methods*

9 All microarray datasets were downloaded from GEO site, and normalized,
 10 where appropriate using the Bioconductor package *Affy*. Data were arranged
 11 into an expression matrix where each row corresponds with gene expression
 12 for each gene and each column corresponds with cell phenotype (sample).
 13 Data were log2-transformed and values above log2(10) were used for
 14 analysis. Differentially expressed genes (DEG) the TCR KO dataset and the
 15 aTreg dataset were identified by a moderated t-statistics. DEG for activated
 16 CD44hi and resting CD44lo Treg were combined. The CRAN package *vegan*
 17 was used for the computation of CCA. Gene scores used the *wa* scores of the
 18 CCA output by *vegan*. The Bioconductor package *limma* was used to perform
 19 a moderated t-test. RNA-seq data were preprocessed, normalised, and log-
 20 transformed using standard techniques (34).

21 Heatmaps were generated the CRAN package *gplots*. Venn diagram was
 22 generated using the R code, *overLapper.R*, which was downloaded from the
 23 Girke lab at Institute for Integrative Genome Biology
 24 (http://faculty.ucr.edu/~tgirke/Documents/R_BioCond/My_R_Scripts/overLapp)

1 [er.R](#)). Gene lists were compared for enriched pathways in the REACTOME
2 pathway database using the Bioconductor packages *ReactomePA* and
3 *clusterProfiler*. Violin plots shows kernel density plots (outside) and the
4 median and interquartile range (inside) of the original gene expression data,
5 and were generated by the Bioconductor package *ggplot2*. The lineage curve
6 was constructed by clustering SC4A/CCA sample scores using an
7 expectation–maximization (EM) algorithm (39), and the nodes of these
8 clusters were identified by constructing a minimum spanning tree using the
9 Bioconductor package *Slingshot* (40).

10

11

1

2 **Results**

3 *Identification of the Foxp3-independent activation signature in Treg and* 4 *memory-phenotype T cells*

5 Firstly, we investigated how T cell activation-related genes are differentially
6 regulated in resting Treg and other CD4⁺ T cell populations including Tmem
7 and Teff. To address this multidimensional problem, we applied CCA to the
8 microarray dataset of various CD4⁺ T cells using the explanatory variable for
9 the T cell activation process, which was obtained from the microarray dataset
10 that analysed resting and activated conventional T cells (“T cell subset data”
11 and “T cell activation data” in **Table 1**). Thus, we aimed to visualise the cross-
12 level relationships between genes, the T cell populations, and the T cell
13 activation process (**Figure 1A**). Using the single explanatory variable, the T
14 cell activation process, the solution of CCA is one-dimensional and the cell
15 sample scores of CCA (represented by Axis 1) provides “T cell activation
16 score” (see Methods), indicating the level of activation in each cell population
17 relative to the prototype signature of T cell activation, as defined by the
18 explanatory variable *Tact*. All the naïve T cell populations had low Axis 1
19 values (i.e. Foxp3⁻ T naïve cells (Tnaive); Tnaive, and non-draining lymph
20 node (dLN) T cells from BDC TCR transgenic (Tg) mice, which develop type I
21 diabetes). In contrast, Foxp3⁺ Treg, Tmem, and tissue-infiltrating Teff in the
22 pancreas from BDC Tg (i.e. with inflammation in the islets) had high scores
23 (**Figure 1B**). These results indicate that Treg are as “activated” as Tmem and
24 tissue-infiltrating activated Teff at the transcriptomic level by CCA.

1 Next, we addressed whether the highly “activated” status of Treg is dependent
2 on Foxp3. Since Foxp3 suppresses Runx1-mediated transcriptional activities
3 (38), we investigated the same T cell population dataset using the following
4 three explanatory variables: T cell activation (Tact), retroviral *Foxp3*
5 transduction (Foxp3) and *Runx1* transduction (Runx1) (see Methods). The
6 CCA solution was 3-dimensional, while the first two axes explained the
7 majority of variance (98.8%, **Figure 2A**). As expected, Tmem, tissue-
8 infiltrating Teff and Treg had low negative values and showed high
9 correlations to T cell activation (Tact) in Axis 1, whereas only Treg had high
10 correlations with the Foxp3 variable in Axis 2, while Tmem and Teff were
11 correlated with the Runx1 variable in Axis 2 (**Figure 2A**). By analysing the
12 gene space of the CCA solution, genes in the lower left quadrant (i.e. negative
13 in both Axes 1 and 2) were enriched with the genes that are involved in T cell
14 activation, effector functions, and T follicular helper cells (Tfh), including
15 *Cxcr5*, *Pdcd1*(PD-1) *Il21*, *Ifng*, *Tbx21* (T-bet), *Mki67* (Ki-67) (**Figure 2B**). On
16 the other hand, genes in the upper left quadrant (i.e. negative in Axis 1 and
17 positive in Axis 2) were enriched with Treg-associated genes including *Ctla4*,
18 *Il2ra* (CD25), *Itgae* (CD103), *Tnfrsf9* (4-1BB) and *Tnfrsf4* (OX40) (**Figure 2B**).
19 These results indicate that a set of activation genes are operating in all the
20 three non-naïve T cell populations (i.e. Treg, Teff and Tmem), while some of
21 them are more specific to Treg.

22

23

1 *The Treg transcriptome is characterized by the repression of a part of the*
2 *activation genes for Tmem*

3 Next, we determine the modules of genes that are differentially regulated
4 between Treg and Tmem, in order to understand the multidimensional identity
5 of Treg and Tmem transcriptomes (i.e. how these populations can be defined
6 in comparison to all relevant populations). Specifically, we asked if the Axis 2
7 captured the differential transcriptional regulations between Tmem and Treg.
8 Importantly, Axis 2 represents Foxp3-driven and Runx1-driven transcriptional
9 effects, which are correlated with Treg and Tmem/Teff, respectively (**Figure**
10 **3A**). This suggests that Axis 2 provides a ‘scoring system’ for regulatory vs
11 effector functions. Thus, the genes in Axis 1-low (precisely, genes above 25
12 percentile for positive correlations with Tact) were identified as *Tact genes*.
13 These genes were subsequently classified into Axis 2-positive (i.e. positive
14 correlations with Foxp3 and Treg) [designated as “*Tact-Foxp3 genes*”; top left
15 quadrant of CCA gene space in Figure 1D] and Axis 2-negative genes (i.e.
16 positive correlations with Runx1 and Tmem/Teff) [designated as “*Tact-Runx1*
17 *genes*”; bottom left quadrant of CCA gene space in Figure 1D] (**Figure 3A**).
18 Tact-Runx1 genes contain genes linked to T cell activation (e.g. *Mki67*),
19 effector functions (e.g. *Tbx21*), and Tfh (e.g. *Bcl6*, *Pdcd1*), while Tact-Foxp3
20 genes contain “Treg markers” such as *Il2ra* (CD25) and *Tnfrsf18* (GITR)
21 (**Figure 2B**).

22 Intriguingly, heatmap analysis showed that both Treg and Tmem expressed
23 Tact-Foxp3 genes at high levels, compared to naïve and effector T cells
24 (**Figure 3B**). On the other hand, Tact-Runx1 genes were selectively

1 downregulated in Treg, while their expressions were sustained in Tmem
2 (**Figure 3C**). In other words, the repression of Tact-Runx1 genes was the
3 major feature of Treg in comparison to Tmem, and Tact-Foxp3 genes are the
4 activation genes, the expression of which is induced by T cell activation in
5 both Treg and Tmem, and is sustained or enhanced even in the presence of
6 Foxp3. Interestingly, comparable selective downregulation of Tact-Runx1
7 genes was observed in Teff as well (**Figure 3C**). This suggests that the set of
8 activation genes operating in Teff is different from the ones in Tmem, and that
9 Tmem and Treg share more activation genes than Treg-Teff and Tmem-Teff
10 (**Figure 3B and 3C**). These results collectively indicate that the Treg-ness is
11 composed of the induction of the Treg-Tmem shared activation genes (i.e.
12 Tact-Foxp3 genes) and the Foxp3-mediated repression of Tmem-specific
13 genes (i.e. Tact-Runx1 genes), defining the multidimensional identity of Treg.

14 While the overall activation levels of Treg and Tmem are similar to the ones of
15 the tissue-infiltrating Teff at transcriptional level (Figure 1B), when explained
16 by the prototype signature of activation in CD4⁺ T cells (i.e. the explanatory
17 variable Tact), the compositions of the activation genes are different between
18 Treg, Tmem and Teff (as captured by Figure 3B and 3C). Importantly, many
19 of these activation genes are shared between Treg and Tmem, but not with
20 Teff. The closer similarity between resting Treg and Tmem, compared to Teff,
21 is not surprising, considering that both resting Treg and Tmem are at the
22 resting status, while Teff are more recently activated and executing effector
23 functions. In addition, the distinct features of Teff may also include their
24 capacity of tissue infiltration and the effects of the microenvironment. These

1 features were not captured by standard t-test analysis (**Supplementary Fig**
2 **1**).

3 Tact-Foxp3 genes included the transcription factors *Nfat5*, *Runx2*, and *Ahr*,
4 which were expressed by most of Tmem cells as well (**Figure 3D**). The Treg-
5 associated markers, *Il2ra* (CD25), *Itgae* (CD103), and *Tnfrsf18* (GITR) were
6 expressed not only by Treg but also by Tmem at moderate to high levels.
7 Notably, the expression of *Ctla4*, *Ccr4*, and *Lag3* was high in Treg and Tmem
8 cells, but it was repressed in Teff (**Figure 3D**). This suggests that Treg and
9 Tmem are in later stages of T cell activation, when the expression of CTLA-4
10 is induced as a negative feedback mechanism (41), while it is not induced in
11 tissue-infiltrating Teff, presumably because they are more recently activated
12 and actively proliferating.

13 Tact-Runx1 genes included many cell cycle-related genes (e.g. *Ccna1*,
14 *Cdca2*, and *Chek2*), suggesting that these cells are in cell cycle and
15 proliferating (**Figure 3E**). The higher expression of *Mki67* and *Fos* suggests
16 that these Tmem cells had been activated by TCR signals *in vivo* before the
17 analysis. Tact-Runx1 genes also included the transcription factors *Tbx21*,
18 *Maf*, *Hif1a*, and *Bcl6*, which have roles in Th1, Th2, Th17, and Tfh
19 differentiation, respectively (42-44). In accordance with this, the Tfh markers
20 *Cxcr5* and *Pdcd1* were specifically expressed by Tmem, suggesting that
21 Tmem are heterogenous populations and composed of these Th and Tfh
22 cells. These results are compatible with the model that Treg and Tmem
23 constitute the self-reactive T cell population that have constitutive activation
24 status (7), and that the major function of Foxp3 is to modify the constitutive

1 activation processes by repressing a part of the activation gene modules (i.e.
2 Tact-Runx1 genes) (**Figure 4**).

3

4 *The activated status of Treg is TCR signal dependent*

5 We next asked whether the constitutively “activated” status of Treg is
6 dependent on TCR signals. We applied CCA to the microarray data of
7 CD44^{hi}CD62L^{lo} activated Treg (CD44^{hi} activated Treg) and CD44^{lo}CD62L^{hi}
8 naïve-like Treg (CD44^{lo} naïve Treg) from inducible *Tcra* KO or WT (TCR KO
9 data, **Table 1, Figure 5A**) using the T cell activation variable as explanatory
10 variable. The CCA result showed that CD44^{hi} activated Treg from WT mice
11 only showed high activation scores, compared with all the other groups.
12 Interestingly, *TCRa* KO CD44^{lo} naïve-like Treg showed the lowest scores, and
13 were lower than WT CD44^{lo} naïve-like Treg (**Figure 5B**). These results
14 indicate that TCR signaling is required for the constitutive activation status of
15 Treg, especially CD44^{hi} activated Treg, and suggest that these activated Treg
16 are more enriched with the cells that received TCR signals recently,
17 compared to CD44^{lo} naïve-like Treg.

18 In order to further address whether the TCR signal-dependent activation
19 signature of Treg is constitutively maintained or specifically induced by *in vivo*
20 activation events (presumably as tonic TCR signals (7)), we analysed the
21 RNA-seq dataset of *in vivo* activated Treg (Ref. (16), **Table 1**). The dataset
22 was generated by depleting a part of Treg by Diphtheria toxin (DT) using bone
23 marrow chimera of *Foxp3*^{GFP^{CreERT2}}:*Rosa26YFP* and *Foxp3*^{GFP^{DTR}} (16). The
24 DT treatment depletes DT receptor (DTR)-expressing Treg from *Foxp3*^{GFP^{DTR}},

1 and thus induces a transient inflammation through the reduction of Treg.
 2 *Foxp3*^{GFPCreERT2} allows to label Foxp3-expressing cells by YFP at the moment
 3 of tamoxifen administration. Van der Veecken *et al* thus analysed resting Treg
 4 from untreated mice (rTreg), activated Treg from mice with recent depletion
 5 (11 days before the analysis) in an inflammatory condition (aTreg), and
 6 “memory” Treg (mTreg) from mice with a distant depletion (60 days before the
 7 analysis) (**Figure 5C**). As expected, the CCA analysis using the T cell
 8 activation variable showed that aTreg had higher activation scores than both
 9 rTreg and mTreg (**Figure 5D**). This indicates that the activation mechanisms
 10 are more actively operating in activated Treg in an inflammatory environment.

11 In order to further dissect the activation signature of Treg, we obtained the
 12 lists of differentially expressed genes (DEG) between WT Treg vs *Tcra* KO
 13 Treg (designated as *TCR-dependent genes*), and between aTreg and rTreg
 14 (designated as *aTreg-specific genes*, see Methods). Interestingly, 94/286
 15 genes of Tact-Runx1 genes (Tmem-specific activation genes, repressed in
 16 resting Treg) are also used during the activation of Treg (**Figure 6A**), while
 17 only 8/119 of Tact-Foxp3 genes (used by Tmem and resting Treg) are
 18 induced during the activation of Treg (**Figure 6B**). This indicates that the
 19 activation of Treg does not enhance the genes that are used in resting Treg,
 20 but induces the expression of the Tmem-specific genes that are suppressed
 21 in resting Treg. On the other hand, 51/286 of Tact-Runx1 and 19/119 of Tact-
 22 Foxp3 genes are regulated by TCR signalling (**Figure 6A and 6B**),
 23 suggesting that the activation status of resting Treg and Tmem may be
 24 sustained by TCR signals. Pathway analysis showed that Tact-Runx1 and
 25 aTreg-specific genes were enriched for cell-cycle related pathways. In

1 contrast, Tact-Foxp3 genes were enriched for pathways related to signal
2 transduction only (**Figure 6C**). Collectively, the results above suggest that
3 resting Treg are maintained by TCR and cytokine signalling, and that the
4 activation of Treg induces the transcriptional activities of Tact-Runx1 genes,
5 which promote proliferation and cell division.

6

7 *FOXP3 expression more frequently occurs in activated T cells than resting*
8 *cells by single cell CCA*

9 The analyses above showed that Treg are on average more activated than
10 naïve T cells and that the activation status of Treg can be variable. However,
11 it is still unclear whether individual Treg are activated than any naïve T cells at
12 the single cell level. The alternative hypothesis is that Treg are enriched with
13 the T cells that have recognised their cognate antigens and been activated. In
14 order to determine this and thereby understand the dynamics of T cell
15 regulation in vivo, we investigated the single cell RNA-seq data of tumour-
16 infiltrating T cells from human patients (Ref. (35) **Table 1**), and further
17 enquired how the activation mechanisms are operating in Treg at the single
18 cell level.

19 Firstly, we *in silico*-sorted FOXP3+ and FOXP3- CD4⁺CD3⁺ T cells from
20 unannotated single cell data from tumours, which tissues were dispersed and
21 CD45+ cells were sorted by flow cytometry without the use of any lymphocyte
22 markers (GSE72056, **Table 1**). Thus, the identities of individual single cells
23 were needed to be identified in a data-oriented manner, and Treg and non-
24 Treg cells in these tumour tissues had unknown individual activation and

1 differentiation statuses. Thus, we applied CCA to the *in silico*-sorted single
2 cell T cell data using the explanatory variables of activated conventional CD4+
3 T cells (*Tact*) and resting T cells (*Trest*; GSE15390, **Table 1**), aiming to define
4 individual single cells according to their level of activation by their correlations
5 to these two variables (**Figure 7A**). Here we used these two variables, *Tact*
6 and *Trest*, in order to generate a two-dimensional CCA solution, instead of a
7 single explanatory variable that represents T cell activation by the log2 fold
8 change in gene expression between activated and resting CD4+ T cells (*c.f.*
9 Figures 1-6), which produces a one-dimensional CCA solution visualised as a
10 single axis), because we aimed to identify any additional major differentiation
11 process(es) in the Axis 2. The explanatory variables *Tact* and *Trest* are both
12 captured by Axis 1 because they represent two poles of one continuum – the
13 spectrum of activation – ranging from ‘resting’ to ‘activated’ cell state. Thus
14 the CCA aimed to sort the single cells according to their individual levels of
15 activation along the spectrum of activation, capturing the heterogeneity in
16 activation levels in single cells. Remarkably, in the cell sample space of the
17 CCA solution, the majority of FOXP3+ T cells were positively correlated with
18 the T cell activation variable *Tact* (**Figure 7B**), and thus had negative scores
19 in the Axis 1 (**Figure 7B**). Here, CCA Axis 1 \times (-1) score is designated as the
20 T cell activation score. Thus, using the activation score and FOXP3
21 expression, the following four subpopulations were defined: “Activated
22 FOXP3+”, “Resting FOXP3+”, “Activated FOXP3-”, and “Resting FOXP3-”
23 (**Figure 7B**).

24 Next, we aimed to determine whether individual activated FOXP3+ Treg are
25 more activated than activated FOXP3- non-Treg at the single cell level

1 According to the T cell activation score established by the CCA solution in
2 Figure 7B, FOXP3⁺ Treg had significantly higher T cell activation scores than
3 FOXP3⁻ non-Treg on average, as indicated by the higher median in the violin
4 plots and greater density of samples with higher T cell activation scores
5 (**Figure 7C**), confirming the results by bulk cell analysis (Figure 1). Using the
6 CCA definition of activated and resting Treg and non-Treg established in
7 Figure 7B, the T cell activation score neatly captured the activated status of
8 single cells, allocating high positive and negative scores to activated and
9 resting cells, respectively (**Figure 7D**). Importantly, there was no significant
10 difference between Activated FOXP3⁺ and Activated FOXP3⁻ cells and
11 between Resting FOXP3⁺ and Resting FOXP3⁻ cells (**Figure 7D**), indicating
12 that in tumour microenvironment, Treg cells are as activated as non-Treg
13 CD4⁺ T cells, which may be enriched with Teff. Strikingly, 32.5% of activated
14 T cells expressed FOXP3, while only 8.2% of resting T cells expressed
15 FOXP3 in Figure 7B. In other words, FOXP3 expression occurred more
16 frequently in activated T cells. Given that the activation signature of Treg is
17 dependent on TCR signals (Figure 5), these results suggest that FOXP3
18 expression occurs predominantly in the activated T cells that have recognised
19 the tumour antigens and received TCR signals, as a negative feedback
20 mechanism to suppress the effector response against tumour antigen (7).
21 Alternatively, but not exclusively, FOXP3⁺ T cells may have high-affinity TCRs
22 to self-MHC and/or tumour antigens and be more prone to activation (10).

23 In the gene space of the CCA solution, genes with strong correlations to
24 activated FOXP3⁺ T cells included *FOXP3* itself and common Treg markers
25 such as *CTLA4* and *IL2RA* (CD25), which were found in the upper left

1 quadrant (Axis 1-negative Axis 2-positive). Interestingly, the lower left
 2 quadrant (Axis 1-negative Axis 2-negative) contained more Tfh-like or
 3 effector-like molecules *PDCD1* (PD-1), *BCL6*, *IL21*, and *IFNG*. The
 4 chemokine receptors *CCR5* and *CCR2* had negative scores in Axis 1 (i.e.
 5 correlated with *Tact*), while *CCR7* had a high positive score in Axis 1 (i.e.
 6 correlated with *Trest*) (**Figure 7E**).

7

8 *Identification of Tfh-like differentiation and Foxp3-driven processes and the*
 9 *common activation process in tumour-infiltrating T cells*

10 Next, we aimed to identify major differentiation and activation processes in the
 11 single cell transcriptomes above. To this end, we have developed a new CCA
 12 approach for single cell analysis (Single Cell Combinatorial CCA, SC4A),
 13 which aims to visualise major differentiation/activation processes and the
 14 underlying gene regulations (**Figure 8A**, see Materials and Methods). Firstly,
 15 we classified single cells into the four populations (Activated and Resting
 16 cells, and FOXP3+ Treg and FOXP3- non-Treg; Figure 7B), and thereby
 17 identified the following four processes as putative differentiation and activation
 18 processes in the dataset: T cell activation (Activated cells), and naïve-ness
 19 (Resting cells), FOXP3-driven process (Activated FOXP3+), and Tfh-like
 20 process (Activated FOXP3-) (Figure 7). Secondly, based on their high scores
 21 in the CCA solution (i.e. either high positive or high negative scores in either
 22 Axis 1 or 2 in Figure 7E) and abundant expressions in FOXP3+ and FOXP3-
 23 cells (data not shown), we selected 12 candidate genes (*CCR7*, *CCR5*,
 24 *CCR4*, *IL2RA*, *IL2RB*, *CTLA4*, *ICOS*, *TNFRSF4*, *TNFRSF9*, *FOXP3*, *BCL6*,

1 *PDCD1*) as the candidate genes for the four processes. From these genes,
 2 we identified the most positively correlated gene to each of the four processes
 3 using the combinatorial CCA, which tests all the combinations of variables by
 4 CCA and obtains the most correlated gene for each population; see Materials
 5 and Methods). Thus, *PDCD1*, *FOXP3*, *CTLA4*, and *CCR7* were identified as
 6 the most correlated genes for Activated FOXP3-, Activated FOXP3+,
 7 Activated T cells, and Resting T cells, respectively (**Supplementary Figure**
 8 **3**), which represent the four immunological processes (see above). Finally,
 9 using these four genes as explanatory variables, we applied CCA to the single
 10 cell transcriptomes, obtaining the solution of the SC4A approach.

11 The single cell space of the SC4A solution showed that Activated and Resting
 12 T cells had negative and positive scores, respectively (**Figure 8B**). This
 13 indicates that Axis 1 represents T cell activation vs naïve-ness. Single cells
 14 were successfully clustered into Activated FOXP3⁺ Treg, Activated FOXP3-
 15 non-Treg, and Resting T cells. Resting FOXP3+ Treg and Resting FOXP3- T
 16 cells were mostly overlapped (**Figure 8C**), indicating that the major features in
 17 the dataset dominated the difference between these two resting T cell groups.
 18 Importantly, the explanatory variable CTLA4, which represents the T cell
 19 activation process, was highly correlated with both Activated FOXP3⁺ Treg
 20 and Activated FOXP3- non-Treg at the middle, indicating its neutral position in
 21 terms of Tfh and Treg activation processes. As expected, the variable CCR7,
 22 which represents naïve-ness, was correlated with both Resting FOXP3+ Treg
 23 and Resting FOXP3- T cells. The explanatory variable PDCD1, which
 24 represents the Tfh-like process, was highly correlated with Activated FOXP3-
 25 non-Treg cells, while the variable FOXP3 was correlated with Activated

1 FOXP3+ Treg. Thus, the single cell transcriptomes were modelled by the
2 correlations between gene expression, single cells, and the expression of the
3 four key genes, which represent the four immunological processes (**Figure 8B**
4 **and 8C**). PCA and t-distributed stochastic neighbor embedding (t-SNE) did
5 not provide insights into such cross-level relationships or clear separations of
6 the populations (**Supplementary Figure 4**).

7 Next, in order to understand the relationship between the T cell activation
8 signature and FOXP3-driven and Tfh-like processes (Figure 8B and 8C), we
9 aimed to identify and characterise genes with high correlations to these
10 processes, which were represented by CTLA4, FOXP3, and PDCD1
11 explanatory variables, by analysing the gene space of the final output of
12 SC4A (Figure 8C; see Methods). As expected, the Tfh genes, *IL21* and *BCL6*
13 (45), were highly correlated with PDCD1 explanatory variable. *IL2RA* (CD25)
14 is a Treg marker (46) and was highly correlated with FOXP3 explanatory
15 variable. *IL7R* and *BACH2* are known to be associated with naïve T cells (47,
16 48), and were positively correlated with CCR7 explanatory variable, which
17 represents the naïve-ness (**Figure 8C**). Thus we defined *FOXP3-driven Treg*
18 *genes* (magenta circles) and *Tfh-like genes* (blue circles) according to their
19 high correlation to the FOXP3 and the PDCD1 explanatory variables,
20 respectively, while we designated as *Activation genes* (red circles) the genes
21 that have high correlations with the CTLA4 variable, including *LAG3* and
22 *CCR5*, which were positioned around 0 in Axis 2 (Figure 8C).

23

1 *Identification of the bifurcation point of activated T cells that leads to Tfh-like* 2 *and Treg differentiation in tumour-infiltrating T cells*

3 The analyses above strongly suggested that there are two major
4 differentiation pathways for those tumour-infiltrating T cells, which are
5 regulated by FOXP3-driven and Tfh-like processes. In order to identify these
6 lineages, we applied an unsupervised clustering algorithm to the sample
7 space of the SC4A/CCA result (Figure 8B), and identified 6 clusters, to which
8 a pseudotime method (49) was applied, constructing “lineage curves” (**Figure**
9 **8D**; see **Methods**). Importantly, the lineage curves had a bifurcation point at
10 Cluster II, which leads to the two distinct differentiation pathways, Tfh-like and
11 FOXP3-driven differentiation. Since cells may change and mature their
12 phenotypes in different dynamics between these two lineages, we designated
13 Tfh-like-associated and FOXP3-associated pseudotime as Tfh-pseudotime
14 and FOXP3-pseudotime (**Figure 8D**).

15 In fact, the expression of *Activation genes* was progressively increased in the
16 shared clusters (i.e. Cluster I and II) for the two pseudotimes, and throughout
17 the rest of the FOXP3-pseudotime and the early phase of Tfh-like
18 differentiation (i.e. Cluster III) in Tfh-pseudotime, while it was suppressed
19 towards the end of Tfh-like differentiation (Cluster IV; **Figure 8E**) in Tfh-
20 pseudotime. Given that Tfh-pseudotime is correlated with *PDCD1* expression
21 (**Figure 8C**), this suggests that *PDCD1* expression and the Tfh-effector
22 process are induced during the earlier phases of effector T cell activity, and
23 that the activation processes in *PDCD1*^{high} T cells are suppressed,
24 presumably through PD1-PDL1 interactions in the tumour environment (50).

1 Interestingly, *FOXP3-driven genes* had similar dynamics to *Activation genes*
 2 in both FOXP3-pseudotime and Tfh-pseudotime (**Figure 8F**). In contrast, *Tfh-*
 3 *like genes* were mostly suppressed throughout FOXP3-pseudotime, while
 4 they were progressively induced throughout Tfh-pseudotime (**Figure 8G**).
 5 These differential regulations of two gene modules resonate with those of
 6 Tact-Foxp3 genes (which are expressed by both Treg and Tmem) and Tact-
 7 Runx1 genes (which are expressed specifically in Tmem, and repressed in
 8 Treg) (Figure 3). In fact, *FOXP3* expression is weakly induced in some cells in
 9 the bifurcating Cluster II and the early phase of Tfh-like differentiation (Cluster
 10 III) in Tfh-pseudotime, and is progressively increased at and beyond Cluster V
 11 in FOXP3-pseudotime (**Figure 8H**).

12 *RUNX1* is highly expressed in the common Clusters I and II, and is
 13 downregulated in the transition from Cluster II to Cluster III in Tfh-pseudotime,
 14 and from Cluster II to Cluster V in FOXP3-pseudotime (**Figure 8I**), which is
 15 compatible with the known dynamics of *RUNX1* expression: Runx1 is
 16 downregulated when naïve CD4⁺ T cells differentiate into activated/effector
 17 cells following TCR signaling (51). By analysing other key genes used as CCA
 18 explanatory variables, *CTLA4* was induced at the bifurcating point, Cluster II,
 19 and onwards in both of the lineages at equivalent expression levels (**Figure**
 20 **8J**), reflecting the activated status of both effector Tfh-type cells and Treg.
 21 Importantly, *CTLA4* is a marker of Treg as well as activated effector T cells,
 22 where it acts as a negative regulator of T cell proliferation (52).

23 *PDCD1* expression was also induced at the bifurcating point, and throughout
 24 Tfh-pseudotime, but specifically suppressed in the early phase of FOXP3-

1 pseudotime (**Figure 8K**), which is compatible with the known dynamics that
 2 PDCD1 is transiently upregulated in activated CD4⁺ T cells as a negative
 3 regulatory mechanism to restrain proinflammatory immune responses and
 4 maintain peripheral tolerance (53). Further supporting this dynamic
 5 perspective, *IL2* expression occurs mainly in Cluster II, indicating that these
 6 cells are enriched with the T cells that recently recognised antigens (54)
 7 (**Figure 8L**). Consistently, the expression of the naïve T cell marker CCR7
 8 was the highest in the cells with a relatively naïve phenotype in shared Cluster
 9 I, and was moderately downregulated in the early and late phase of Tfh-
 10 pseudotime, and suppressed in most Treg in FOXP3-pseudotime (**Figure**
 11 **8M**).

12 These results collectively support the model that constant activation
 13 processes in the tumour microenvironment promote terminal differentiation of
 14 the Treg- and Tfh-like lineages in both previously committed and non-
 15 committed lineages of T cells. Interestingly, Cluster II is the bifurcation point,
 16 in which T cells show moderate activation and together with simultaneous
 17 expression of FOXP3 and Tfh-like genes, as well as RUNX1 and PDCD1
 18 expression. These cells are most probably engaged in decision-making about
 19 their cell fate and the cell type-specific usage of these genes – whether their
 20 transcriptional mechanisms would be used to generate a proinflammatory or
 21 regulatory response. This understanding was possible because SC4A
 22 effectively annotated genes and cells and thereby allowed to identify new cell
 23 populations.

24

1 *Identification of markers for the differential regulation of Tfh-like and Treg* 2 *differentiation in activated T cells*

3 Lastly, we aimed to demonstrate the utility of the current approach by
4 discovering exemplary marker genes that distinguish cells in FOXP3- and Tfh-
5 pseudotime (i.e. the FOXP3-driven pathway I-II-V-VI, and the Tfh-like pathway
6 I-II-III-IV) (**Figure 9A**), and identifying the T cell subpopulations by a flow
7 cytometric visualisation of single cell data. Since *Activation genes* (Figure 8C)
8 are shared by early phases of Tfh-like and FOXP3-driven differentiation
9 (Figure 8E), we took the intersect of these genes and the Tact-Foxp3 genes,
10 which were expressed by both resting Tmem and resting Treg in mice (Figure
11 3). *DUSP4* and *NFAT5* were such genes and in fact induced in cells at the
12 activated bifurcating Cluster II and onwards in both lineages (**Figure 9B**).
13 Similarly, in order to identify a marker to distinguish Treg and Tfh-like cells,
14 firstly, we identified *CCR8* and *IL2RA* in the intersect of FOXP3-driven genes
15 (Figure 8C) and the Tact-Foxp3 genes, which were induced highly and
16 progressively in Treg-lineage cells throughout FOXP3-pseudotime, while
17 mostly suppressed across Tfh-pseudotime (**Figure 9C**). In contrast, *BCL6* and
18 *KCNK5* (found in the intersect of Tfh-like genes (Figure 8C) and the Tact-
19 Runx1 genes, which are expressed in resting Tmem but suppressed in resting
20 Treg (Figure 3)) were progressively induced across Tfh-pseudotime, while
21 suppressed in FOXP3-pseudotime (**Figure 9D**).

22 Lastly, in order to make the newly obtained knowledge easily accessible to
23 experimental immunologists, we showed the expression of *NFAT5*, *IL2RA*,
24 *CCR8*, *BCL6*, and *KCNK5* in the tumour-infiltrating T cells in a flow cytometric

1 format (**Figure 9E**). The common activation gene *NFAT5* in fact captured the
 2 majority of Treg-lineage cells (i.e. cells in the Clusters V and VI) and Tfh-like-
 3 lineage cells (i.e. cells in the Clusters III and IV). The Treg-specific genes the
 4 expression of *IL2RA* and *CCR8* occurred in the majority of FOXP3+ Treg-
 5 lineage cells, whether NFAT-positive or negative, but not in most of Tfh-like-
 6 lineage cells. In contrast, the Tfh-like-specific genes *BCL6* and *KCNK5* were
 7 expressed by a majority of Tfh-like-lineage cells and were not expressed in
 8 Treg-lineage cells (**Figure 9E**).

9 Collectively, these results indicate that the SC4A analysis successfully
 10 decomposed the gene regulations for T cell activation and Treg and effector T
 11 cell differentiation, identifying new cell populations, which include activated
 12 cells at the bifurcation point, early and late phases of Treg and Tfh-like
 13 differentiation, and their feature genes. In addition, although there must be
 14 considerable differences between resting T cells in the secondary lymphoid
 15 organs and between humans and mice, our study successfully identified the
 16 shared activation processes and the conserved genes that are differentially
 17 used between the Treg- and the Teff-lineage cells, identifying a shared
 18 systems-level mechanism for the differentiation regulation of activation and
 19 differentiation processes in CD4+ T cell populations.

20

Discussion

Resting Treg showed an activated status, comparable to that of Teff and Tmem at the population level. In addition, the activation signature of Treg was more remarkable in CD44^{hi}CD62L^{lo} activated Treg than CD44^{lo}CD62L^{hi} naïve-like Treg. CD44^{hi}CD62L^{lo} Treg are also identified as eTreg, which may have enhanced immunosuppressive activities (55). The eTreg fraction includes the GITR^{hi}PD-1^{hi}CD25^{hi} “Triple-high” eTreg that have high CD5 and Nur77 expressions, which indicates that they have received strong TCR signals (17). In humans, CD25^{hi}CD45RA⁻FOXP3^{hi} eTreg highly express Ki67 (56), indicating that these cells were recently activated. Given that TCRs of Treg have higher affinities to self-antigens (57), these eTreg may have the most self-reactive TCRs during homeostasis. Alternatively, the eTreg subset may have recently received strong TCR signals and upregulated activation markers, and such cells may acquire a resting status at later time points. Future investigations by TCR repertoire analysis will answer this question.

Our study revealed the heterogeneity of FOXP3⁺ Treg at the single cell level, and showed that tumour-infiltrating Treg include FOXP3⁺ T cells with various levels of activation (Figure 7B and Figure 8C). It is plausible that, in the physiological polyclonal settings, the variations in the activated status of individual Treg may be due to the TCR affinity to its cognate antigen, the availability of cognate antigen, and the strength and duration of TCR signals. Our SC4A analysis identified the *FOXP3*-driven genes, which are specific to activated FOXP3⁺ cells and include IL-2 and common gamma chain cytokine receptors (i.e. *IL2RA*, *IL2RB*, *IL15RA*, *IL4R*, and *IL2RG*), DNA replication

1 licensing factors (e.g. MCM2), and transcription factors such as *PRDM1*
2 (*BLIMP1*) and *IRF4* (which control the differentiation and function of eTreg
3 (19)). These gene modules are distinct from the Tfh-like genes and the
4 activation genes (Figure 8), and may be controlled specifically by *FOXP3*
5 under strong TCR signals. The expression of these genes is variable within
6 the *FOXP3*⁺ T cells, suggesting that the transcriptional activities of these
7 genes are dynamically regulated over time in tumour-infiltrating Treg. Thus,
8 single cell-level analysis is becoming a key technology to address the
9 heterogeneity of Treg. To our knowledge, this study is one of the first single
10 cell analyses of Treg transcriptomes, while we find that, during the review
11 process of this manuscript, another study addressing Treg heterogeneity by
12 single cell RNA-seq was deposited at a preprint server (58)).

13 The shared activation genes between activated *FOXP3*⁺ Treg and *FOXP3*⁻
14 non-Treg contain apoptosis-related genes (e.g. *CASP3*, *BAD*), which may be
15 differentially controlled between Treg and non-Treg at the protein level. For
16 example, activated *FOXP3*⁻ non-Treg express *DUSP6* (Figure 9B), which is a
17 negative regulator of JNK-induced apoptosis through BIM activation, while
18 *FOXP3* suppresses *DUSP6* expression and promotes the apoptosis
19 mechanism (59). In addition, the activation genes include transcription factors
20 such as *TBX21* (T-bet) and *BATF*. Although *TBX21* is sometimes thought to
21 be a Th1-specific gene, it is upregulated immediately after T cell activation
22 (60). *BATF* was identified as a critical factor for the differentiation and
23 accumulation of tissue-infiltrating Treg (61). These activation genes may be
24 required when T cells are activated and differentiate into either Treg or Teff.

1 Further studies are required to investigate the temporal sequences of these
2 differentiation events *in vivo*.

3 Although the effects of TCR signals on Tmem were not directly examined,
4 considering that Tmem are self-reactive and their differentiation is dependent
5 on the recognition of cognate antigens in the thymus (7), these results
6 collectively suggest that the activation signature of Tmem is also dependent
7 on TCR signals, as is the activation signature in Treg (Figure 5B). Intriguingly,
8 some Treg may lose their Foxp3 expression and become ex-Treg, which are
9 enriched in CD44^{hi} effector T cells or Tmem (30). In contrast, a Tmem
10 population (precisely, Foxp3⁻CD44^{hi}CD73^{hi}FR4^{hi} T cells) efficiently express
11 Foxp3 during lymphopenia (62). These findings support the feedback control
12 model that Foxp3 expression can be induced in Tmem and sustained in Treg
13 as a regulatory feedback mechanism for TCR signals (7). Given the variations
14 in the activated status in individual Treg and Tmem, single cell analysis will be
15 required to address this problem. For example, although Samstein *et al.*
16 showed that DNA hypersensitivity sites in Treg are similar to those in
17 activated T cells (9), it is possible that DNA hypersensitivity sites are variable
18 between individual Treg, and that Tmem may have a similar chromatin
19 structure to Treg.

20 Importantly, our analysis showed that Tmem-specific activation-induced
21 genes (i.e. Tact-Runx1 genes) are uniquely repressed in Treg. The repression
22 is likely to be mediated by the interaction between Foxp3 and other
23 transcription factors that regulate the expression of the Tmem-specific
24 activation genes (Figure 3C). Interestingly, *Runx1* was associated with these

1 Tmem-specific genes. In fact, Foxp3 interacts with Runx1 and thereby
2 represses IL-2 transcription and controls the regulatory function of Treg (38),
3 and a significant part of the Foxp3-binding to active enhancers occurs through
4 the Foxp3-Runx1 interaction (9). These suggest that Runx1 may have a
5 unique role in the differentiation and maintenance of Tmem.

6 While CTLA-4 is commonly recognised as a Treg marker, it is upregulated in
7 all activated T cells, thus CTLA-4 is also a marker of activated T cells (41).
8 CTLA-4 is in fact expressed by only a subset of resting Treg (63), which may
9 be more activated and proliferating *in vivo* (64). In fact, our study shows that
10 CTLA-4 is expressed by non-Treg activated T cells including resting Tmem
11 (Figure 3D) and FOXP3- Tfh-like effector T cells in the tumour
12 microenvironment (Figure 7E and 8C). These findings support that CTLA-4 is
13 primarily a marker for general T cell activation, rather than Treg-specific
14 marker, and that Treg are highly activated T cells with *FOXP3* and *CTLA-4*
15 expression. Importantly, although both FOXP3+ and FOXP3- cells had the
16 same relative level of activation (Figure 7D), the absolute number of FOXP3+
17 cells expressing *CTLA4* was lower than that of Tfh-type cells (Figure 8J),
18 which suggests that therapeutic anti-CTLA4 antibodies (i.e. Ipilimumab)
19 primarily target activated Tfh-like effector cells and thereby directly enhance
20 their activities in tumour microenvironments. Future studies are required to
21 experimentally investigate the *in vivo* dynamics of CTLA-4 expression in mice
22 and humans.

23 In contrast, the expression of *PDCD1* was consistently high in all Tfh-like
24 cells, while it was sparse among FOXP3+ cells (Figure 8K). The co-

1 expression of *BCL6* and *IL21* in some of these PD-1+ cells indicates that Tfh
2 differentiation occurs in the tumour microenvironment, presumably through
3 the repeated and chronic exposure to quasi-self antigens (i.e. tumour
4 antigens). Interestingly, the Tfh signature has been identified in type-I
5 diabetes in both mice and humans (65). Intriguingly, the Tfh-like genes include
6 cell-cycle related genes (e.g. *CDK6*), immediate early transcription factors
7 (*NFATC1*, *EGR2/3*), and RNA-processing genes (e.g. *DICER1*). The
8 significance of these gene modules should be addressed in future studies.
9 However, the high *PDCD1* expression in Tfh-like cells may make them
10 vulnerable to the negative immunoregulatory effect of PD-1 in tumour
11 microenvironments (50). In fact, the most mature *PDCD1*^{high} Tfh-like cells
12 (cluster VI, Figure 8K) moderately decrease the expression levels of activation
13 genes (Figure 8E), suggesting that these cells may have started to be
14 regulated by PD1 ligands. Further experimental investigations are required to
15 reveal how dynamically PD1 regulates T cells during immune response.

16 Interestingly, *RUNX1* is completely repressed in the early phase of FOXP3-
17 pseudotime (Cluster V) but re-expressed in the late phase of FOXP3-
18 pseudotime the expression of *RUNX1* is significantly elevated (Cluster VI)
19 (**Figure 8I**). Similar to *RUNX1*, some cells appear to be expressing *PDCD1* in
20 the later phase of FOXP3-pseudotime in Treg-lineage cells. The
21 reappearance of effector phenotype genes *RUNX1* and *PDCD1* in FOXP3^{high}
22 cells may indicate that these Treg are highly activated effector Treg, which
23 may be actively participating in neutralizing the activities of effector T cells
24 and Tfh. Alternatively, these FOXP3^{high} cells may include ambivalent cells with
25 the characters of both regulatory and effector, or alternatively, be at the point

1 of conversion to Tmem (**Figure 8K**). In bulk resting Treg, *Pdcd1* was
2 expressed at low levels in some Treg (Figure 3E) as well as PD-L2-encoding
3 gene *Pdcd1lg2* (Figure 3D). Future studies are required to reveal the role of
4 these cells.

5 SC4A is a useful method to identify distinct clusters of T cells and the
6 correlated genes to each cluster, and thereby to reveal characteristic cell
7 groups and their active gene modules, while retaining the single-cell level
8 variations. We also showed that SC4A and CCA results can be further
9 analysed by the *pseudotime* approach (CCA-pseudotime). Since SC4A/CCA
10 provides functional annotations to cell groups and gene clusters, the
11 understanding of the pseudotime axis is effective, as shown in the current
12 study. However, given that pseudotime is not a direct measurement of the
13 time-dependent events, but rather is that of similarities between samples (66),
14 future studies are required to analyse time-dependent events *in vivo*, ideally
15 with a new experimental system to directly address the temporal dynamics. In
16 order to make the current SC4A/CCA approach accessible to experimental
17 immunologists, we visualised our single cell data findings using a flow-
18 cytometry style (Figure 9). Although reliable antibodies are currently not
19 available for those intracellular candidate genes, and the expression of protein
20 and transcripts may not be synchronized, the present study demonstrated the
21 power of the SC4A/CCA approach to extract biological meaning from
22 unannotated single cell RNA-seq data. The current limitation of SC4A is that it
23 is computationally expensive (i.e. requires several hours for each analysis
24 using a standard desktop), and the improvement of the computational
25 algorithm using a low-level language will be beneficial. Importantly, SC4A is

1 most effective when used together with in-depth knowledge of immunology
 2 and gene regulation, facilitates the interpretation of CCA results and
 3 explanatory variable selection. Thus, it is hoped that these tools will be widely
 4 used by experimental immunologists with a sound understanding of the
 5 biological significance of results, as well as adequate competence in
 6 computational analysis, which will enable to ask questions involving
 7 multidimensional problems such as multiple T cell subsets.

8

9 **Data and code availability**

10 All R codes are available upon request. Processed data will be provided upon
 11 reasonable requests to the corresponding author.

12

13 **Acknowledgement**

14 We thank Dr Cristina Leslie for providing us the details of their dataset
 15 GSE83315, Prof Lucy Walker for a useful discussion, and Dr David Bending
 16 for valuable comments on the manuscript. M.O. is a David Phillips Fellow
 17 (BB/J013951/2) from the Biotechnology and Biological Sciences Research
 18 Council (BBSRC), and is also supported by a pump-priming grant from
 19 Cancer Research Centre of Excellence, Imperial College London/the Institute
 20 of Cancer Research. A.B is supported by the Japanese Ministry of Education,
 21 Culture, Sports, Science and Technology (MEXT) PhD scholarship.

References

1. Macian, F., C. Lopez-Rodriguez, and A. Rao. 2001. Partners in transcription: NFAT and AP-1. *Oncogene* 20: 2476-2489.
2. Malek, T. R. 2008. The biology of interleukin-2. *Annu Rev Immunol* 26: 453-479.
3. Setoguchi, R., S. Hori, T. Takahashi, and S. Sakaguchi. 2005. Homeostatic maintenance of natural Foxp3(+) CD25(+) CD4(+) regulatory T cells by interleukin (IL)-2 and induction of autoimmune disease by IL-2 neutralization. *The Journal of experimental medicine* 201: 723-735.
4. Hori, S. 2014. Lineage stability and phenotypic plasticity of Foxp3(+) regulatory T cells. *Immunological reviews* 259: 159-172.
5. Tran, D. Q., H. Ramsey, and E. M. Shevach. 2007. Induction of FOXP3 expression in naive human CD4+FOXP3 T cells by T-cell receptor stimulation is transforming growth factor-beta dependent but does not confer a regulatory phenotype. *Blood* 110: 2983-2990.
6. Chen, W., W. Jin, N. Hardegen, K.-j. Lei, L. Li, N. Marinos, G. McGrady, and S. M. Wahl. 2003. Conversion of Peripheral CD4⁺CD25⁻ Naive T Cells to CD4⁺CD25⁺ Regulatory T Cells by TGF- β Induction of Transcription Factor *Foxp3*. *The Journal of experimental medicine* 198: 1875-1886.
7. Ono, M., and R. J. Tanaka. 2016. Controversies concerning thymus-derived regulatory T cells: fundamental issues and a new perspective. *Immunology and cell biology* 94: 3-10.
8. Thornton, A. M., and E. M. Shevach. 2000. Suppressor Effector Function of CD4⁺CD25⁺ Immunoregulatory T Cells Is Antigen Nonspecific. *The Journal of Immunology* 164: 183-190.
9. Samstein, R. M., A. Arvey, S. Z. Josefowicz, X. Peng, A. Reynolds, R. Sandstrom, S. Neph, P. Sabo, J. M. Kim, W. Liao, M. O. Li, C. Leslie, J. A. Stamatoyannopoulos, and A. Y. Rudensky. 2012. Foxp3 exploits a pre-existent enhancer landscape for regulatory T cell lineage specification. *Cell* 151: 153-166.
10. Levine, A. G., A. Arvey, W. Jin, and A. Y. Rudensky. 2014. Continuous requirement for the TCR in regulatory T cell function. *Nature immunology* 15: 1070-1078.
11. Vahl, J. C., C. Drees, K. Heger, S. Heink, Julius C. Fischer, J. Nedjic, N. Ohkura, H. Morikawa, H. Poeck, S. Schallenberg, D. Rieß, Marco Y. Hein, T. Buch, B. Polic, A. Schönle, R. Zeiser, A. Schmitt-Gräff, K.

- 1 Kretschmer, L. Klein, T. Korn, S. Sakaguchi, and M. Schmidt-Supprian.
2 2014. Continuous T Cell Receptor Signals Maintain a Functional
3 Regulatory T Cell Pool. *Immunity* 41: 722-736.
- 4 12. Feuerer, M., J. A. Hill, K. Kretschmer, H. von Boehmer, D. Mathis, and
5 C. Benoist. 2010. Genomic definition of multiple ex vivo regulatory T
6 cell subphenotypes. *Proceedings of the National Academy of Sciences*
7 *of the United States of America* 107: 5919-5924.
- 8 13. Koch, M. A., G. Tucker-Heard, N. R. Perdue, J. R. Killebrew, K. B.
9 Urdahl, and D. J. Campbell. 2009. The transcription factor T-bet
10 controls regulatory T cell homeostasis and function during type 1
11 inflammation. *Nature immunology* 10: 595-602.
- 12 14. Zheng, Y., A. Chaudhry, A. Kas, P. deRoos, J. M. Kim, T. T. Chu, L.
13 Corcoran, P. Treuting, U. Klein, and A. Y. Rudensky. 2009. Regulatory
14 T-cell suppressor program co-opts transcription factor IRF4 to control
15 T(H)2 responses. *Nature* 458: 351-356.
- 16 15. Linterman, M. A., W. Pierson, S. K. Lee, A. Kallies, S. Kawamoto, T. F.
17 Rayner, M. Srivastava, D. P. Divekar, L. Beaton, J. J. Hogan, S.
18 Fagarasan, A. Liston, K. G. Smith, and C. G. Vinuesa. 2011. Foxp3+
19 follicular regulatory T cells control the germinal center response.
20 *Nature medicine* 17: 975-982.
- 21 16. van der Veen, J., A. J. Gonzalez, H. Cho, A. Arvey, S. Hemmers, C.
22 S. Leslie, and A. Y. Rudensky. 2016. Memory of Inflammation in
23 Regulatory T Cells. *Cell* 166: 977-990.
- 24 17. Wyss, L., B. D. Stadinski, C. G. King, S. Schallenberg, N. I. McCarthy,
25 J. Y. Lee, K. Kretschmer, L. M. Terracciano, G. Anderson, C. D. Surh,
26 E. S. Huseby, and E. Palmer. 2016. Affinity for self antigen selects
27 Treg cells with distinct functional properties. *Nature immunology* 17:
28 1093-1101.
- 29 18. Dias, S., A. D'Amico, E. Cretney, Y. Liao, J. Tellier, C. Bruggeman, F.
30 F. Almeida, J. Leahy, G. T. Belz, G. K. Smyth, W. Shi, and S. L. Nutt.
31 2017. Effector Regulatory T Cell Differentiation and Immune
32 Homeostasis Depend on the Transcription Factor Myb. *Immunity* 46:
33 78-91.
- 34 19. Cretney, E., A. Xin, W. Shi, M. Minnich, F. Masson, M. Miasari, G. T.
35 Belz, G. K. Smyth, M. Busslinger, S. L. Nutt, and A. Kallies. 2011. The
36 transcription factors Blimp-1 and IRF4 jointly control the differentiation
37 and function of effector regulatory T cells. *Nature immunology* 12: 304-
38 311.
- 39 20. Sakaguchi, S., M. Miyara, C. M. Costantino, and D. A. Hafler. 2010.
40 FOXP3+ regulatory T cells in the human immune system. *Nature*
41 *reviews. Immunology* 10: 490-500.

- 1 21. Fujii, H., J. Josse, M. Tanioka, Y. Miyachi, F. Husson, and M. Ono.
2 2016. Regulatory T Cells in Melanoma Revisited by a Computational
3 Clustering of FOXP3+ T Cell Subpopulations. *Journal of immunology*.
- 4 22. Andersen, P., and B. Smedegaard. 2000. CD4+ T-Cell Subsets That
5 Mediate Immunological Memory to Mycobacterium tuberculosis
6 Infection in Mice. *Infection and Immunity* 68: 621-629.
- 7 23. Coleman, M. M., C. M. Finlay, B. Moran, J. Keane, P. J. Dunne, and K.
8 H. G. Mills. 2012. The immunoregulatory role of CD4+FoxP3+CD25-
9 regulatory T cells in lungs of mice infected with Bordetella pertussis.
10 *FEMS Immunology & Medical Microbiology* 64: 413-424.
- 11 24. Ono, M., J. Shimizu, Y. Miyachi, and S. Sakaguchi. 2006. Control of
12 autoimmune myocarditis and multiorgan inflammation by
13 glucocorticoid-induced TNF receptor family-related protein(high),
14 Foxp3-expressing CD25+ and CD25- regulatory T cells. *Journal of*
15 *immunology* 176: 4748-4756.
- 16 25. Linsley, P. S., J. L. Greene, P. Tan, J. Bradshaw, J. A. Ledbetter, C.
17 Anasetti, and N. K. Damle. 1992. Coexpression and functional
18 cooperation of CTLA-4 and CD28 on activated T lymphocytes. *The*
19 *Journal of experimental medicine* 176: 1595-1604.
- 20 26. Attridge, K., and L. S. Walker. 2014. Homeostasis and function of
21 regulatory T cells (Tregs) in vivo: lessons from TCR-transgenic Tregs.
22 *Immunological reviews* 259: 23-39.
- 23 27. Bunce, C., and E. B. Bell. 1997. CD45RC Isoforms Define Two Types
24 of CD4 Memory T Cells, One of which Depends on Persisting Antigen.
25 *The Journal of experimental medicine* 185: 767-776.
- 26 28. Min, B., R. McHugh, G. D. Sempowski, C. Mackall, G. Foucras, and W.
27 E. Paul. 2003. Neonates Support Lymphopenia-Induced Proliferation.
28 *Immunity* 18: 131-140.
- 29 29. Curotto de Lafaille, M. A., A. C. Lino, N. Kutchukhidze, and J. J.
30 Lafaille. 2004. CD25- T cells generate CD25+Foxp3+ regulatory T cells
31 by peripheral expansion. *Journal of immunology* 173: 7259-7268.
- 32 30. Miyao, T., S. Floess, R. Setoguchi, H. Luche, H. J. Fehling, H.
33 Waldmann, J. Huehn, and S. Hori. 2012. Plasticity of Foxp3(+) T cells
34 reflects promiscuous Foxp3 expression in conventional T cells but not
35 reprogramming of regulatory T cells. *Immunity* 36: 262-275.
- 36 31. Ono, M., R. J. Tanaka, M. Kano, and T. Sugiman. 2013. Visualising the
37 Cross-Level Relationships between Pathological and Physiological
38 Processes and Gene Expression: Analyses of Haematological
39 Diseases. *PLoS One* 8: e53544.
- 40 32. Greenacre, M. 2008. *Correspondence Analysis in Practice*. Chapman
41 & Hall/CRC,, London.

- 1 33. Tzeng, J., H. H.-S. Lu, and W.-H. Li. 2008. Multidimensional scaling for
2 large genomic data sets. *BMC Bioinformatics* 9: 179.
- 3 34. Ono, M., R. J. Tanaka, and M. Kano. 2014. Visualisation of the T cell
4 differentiation programme by Canonical Correspondence Analysis of
5 transcriptomes. *BMC genomics* 15: 1028.
- 6 35. Tirosh, I., B. Izar, S. M. Prakadan, M. H. Wadsworth, 2nd, D. Treacy, J.
7 J. Trombetta, A. Rotem, C. Rodman, C. Lian, G. Murphy, M. Fallahi-
8 Sichani, K. Dutton-Regester, J. R. Lin, O. Cohen, P. Shah, D. Lu, A. S.
9 Genshaft, T. K. Hughes, C. G. Ziegler, S. W. Kazer, A. Gaillard, K. E.
10 Kolb, A. C. Villani, C. M. Johannessen, A. Y. Andreev, E. M. Van Allen,
11 M. Bertagnolli, P. K. Sorger, R. J. Sullivan, K. T. Flaherty, D. T.
12 Frederick, J. Jane-Valbuena, C. H. Yoon, O. Rozenblatt-Rosen, A. K.
13 Shalek, A. Regev, and L. A. Garraway. 2016. Dissecting the
14 multicellular ecosystem of metastatic melanoma by single-cell RNA-
15 seq. *Science* 352: 189-196.
- 16 36. Wakamatsu, E., D. Mathis, and C. Benoist. 2013. Convergent and
17 divergent effects of costimulatory molecules in conventional and
18 regulatory CD4+ T cells. *Proceedings of the National Academy of
19 Sciences of the United States of America* 110: 1023-1028.
- 20 37. Painter, M. W., S. Davis, R. R. Hardy, D. Mathis, C. Benoist, and C.
21 Immunological Genome Project. 2011. Transcriptomes of the B and T
22 lineages compared by multiplatform microarray profiling. *Journal of
23 immunology* 186: 3047-3057.
- 24 38. Ono, M., H. Yaguchi, N. Ohkura, I. Kitabayashi, Y. Nagamura, T.
25 Nomura, Y. Miyachi, T. Tsukada, and S. Sakaguchi. 2007. Foxp3
26 controls regulatory T-cell function by interacting with AML1/Runx1.
27 *Nature* 446: 685-689.
- 28 39. Chen, W., and R. Maitra. 2017. EMCluster: EM Algorithm for Model-
29 Based Clustering of Finite Mixture Gaussian Distribution. R Package.
- 30 40. Street, K., D. Risso, R. B. Fletcher, D. Das, J. Ngai, N. Yosef, E.
31 Purdom, and S. Dudoit. 2017. Slingshot: Cell lineage and pseudotime
32 inference for single-cell transcriptomics. *bioRxiv*.
- 33 41. Walker, L. S., and D. M. Sansom. 2015. Confusing signals: recent
34 progress in CTLA-4 biology. *Trends Immunol* 36: 63-70.
- 35 42. Szabo, S. J., B. M. Sullivan, C. Stemmann, A. R. Satoskar, B. P.
36 Sleckman, and L. H. Glimcher. 2002. Distinct Effects of T-bet in
37 T_H1 Lineage Commitment and IFN- γ Production in CD4
38 and CD8 T Cells. *Science* 295: 338-342.
- 39 43. Dang, E. V., J. Barbi, H. Y. Yang, D. Jinasena, H. Yu, Y. Zheng, Z.
40 Bordman, J. Fu, Y. Kim, H. R. Yen, W. Luo, K. Zeller, L. Shimoda, S. L.
41 Topalian, G. L. Semenza, C. V. Dang, D. M. Pardoll, and F. Pan. 2011.

- 1 Control of T(H)17/T(reg) balance by hypoxia-inducible factor 1. *Cell*
2 146: 772-784.
- 3 44. Ho, I. C., M. R. Hodge, J. W. Rooney, and L. H. Glimcher. 1996. The
4 proto-oncogene c-maf is responsible for tissue-specific expression of
5 interleukin-4. *Cell* 85: 973-983.
- 6 45. Vinuesa, C. G., M. A. Linterman, C. C. Goodnow, and K. L. Randall.
7 2010. T cells and follicular dendritic cells in germinal center B-cell
8 formation and selection. *Immunological reviews* 237: 72-89.
- 9 46. Sakaguchi, S., T. Yamaguchi, T. Nomura, and M. Ono. 2008.
10 Regulatory T cells and immune tolerance. *Cell* 133: 775-787.
- 11 47. Seddon, B., and R. Zamoyska. 2002. TCR and IL-7 Receptor Signals
12 Can Operate Independently or Synergize to Promote Lymphopenia-
13 Induced Expansion of Naive T Cells. *The Journal of Immunology* 169:
14 3752-3759.
- 15 48. Roychoudhuri, R., K. Hirahara, K. Mousavi, D. Clever, C. A. Klebanoff,
16 M. Bonelli, G. Sciume, H. Zare, G. Vahedi, B. Dema, Z. Yu, H. Liu, H.
17 Takahashi, M. Rao, P. Muranski, J. G. Crompton, G. Punkosdy, D.
18 Bedognetti, E. Wang, V. Hoffmann, J. Rivera, F. M. Marincola, A.
19 Nakamura, V. Sartorelli, Y. Kanno, L. Gattinoni, A. Muto, K. Igarashi, J.
20 J. O'Shea, and N. P. Restifo. 2013. BACH2 represses effector
21 programs to stabilize T(reg)-mediated immune homeostasis. *Nature*
22 498: 506-510.
- 23 49. Lonnberg, T., V. Svensson, K. R. James, D. Fernandez-Ruiz, I. Sebina,
24 R. Montandon, M. S. Soon, L. G. Fogg, A. S. Nair, U. Liligeto, M. J.
25 Stubbington, L. H. Ly, F. O. Bagger, M. Zwiessele, N. D. Lawrence, F.
26 Souza-Fonseca-Guimaraes, P. T. Bunn, C. R. Engwerda, W. R. Heath,
27 O. Billker, O. Stegle, A. Haque, and S. A. Teichmann. 2017. Single-cell
28 RNA-seq and computational analysis using temporal mixture modelling
29 resolves Th1/Tfh fate bifurcation in malaria. *Sci Immunol* 2.
- 30 50. Okazaki, T., S. Chikuma, Y. Iwai, S. Fagarasan, and T. Honjo. 2013. A
31 rheostat for immune responses: the unique properties of PD-1 and their
32 advantages for clinical application. *Nature immunology* 14: 1212-1218.
- 33 51. Wong, W. F., M. Kurokawa, M. Satake, and K. Kohu. 2011. Down-
34 regulation of Runx1 expression by TCR signal involves an
35 autoregulatory mechanism and contributes to IL-2 production. *J Biol*
36 *Chem* 286: 11110-11118.
- 37 52. Buchbinder, E. I., and A. Desai. 2016. CTLA-4 and PD-1 Pathways:
38 Similarities, Differences, and Implications of Their Inhibition. *Am J Clin*
39 *Oncol* 39: 98-106.
- 40 53. McPherson, R. C., J. E. Konkel, C. T. Prendergast, J. P. Thomson, R.
41 Ottaviano, M. D. Leech, O. Kay, S. E. Zandee, C. H. Sweeney, D. C.

- 1 Wraith, R. R. Meehan, A. J. Drake, and S. M. Anderton. 2014.
2 Epigenetic modification of the PD-1 (Pdcd1) promoter in effector
3 CD4(+) T cells tolerized by peptide immunotherapy. *Elife* 3.
- 4 54. Morgan, D. A., F. W. Ruscetti, and R. Gallo. 1976. Selective in vitro
5 growth of T lymphocytes from normal human bone marrows. *Science*
6 193: 1007-1008.
- 7 55. Rosenblum, M. D., S. S. Way, and A. K. Abbas. 2016. Regulatory T
8 cell memory. *Nature reviews. Immunology* 16: 90-101.
- 9 56. Miyara, M., Y. Yoshioka, A. Kitoh, T. Shima, K. Wing, A. Niwa, C.
10 Parizot, C. Taflin, T. Heike, D. Valeyre, A. Mathian, T. Nakahata, T.
11 Yamaguchi, T. Nomura, M. Ono, Z. Amoura, G. Gorochoy, and S.
12 Sakaguchi. 2009. Functional delineation and differentiation dynamics of
13 human CD4+ T cells expressing the FoxP3 transcription factor.
14 *Immunity* 30: 899-911.
- 15 57. Weissler, K. A., and A. J. Caton. 2014. The role of T-cell receptor
16 recognition of peptide:MHC complexes in the formation and activity of
17 Foxp3(+) regulatory T cells. *Immunological reviews* 259: 11-22.
- 18 58. Miragaia, R. J., T. Gomes, A. Chomka, L. Jardine, A. Riedel, A. N.
19 Hegazy, I. Lindeman, G. Emerton, T. Krausgruber, J. Shields, M.
20 Haniffa, F. Powrie, and S. A. Teichmann. 2017. Single cell
21 transcriptomics of regulatory T cells reveals trajectories of tissue
22 adaptation. *bioRxiv*.
- 23 59. Tai, X., B. Erman, A. Alag, J. Mu, M. Kimura, G. Katz, T. Guinter, T.
24 McCaughy, R. Etzensperger, L. Feigenbaum, D. S. Singer, and A.
25 Singer. 2013. Foxp3 transcription factor is proapoptotic and lethal to
26 developing regulatory T cells unless counterbalanced by cytokine
27 survival signals. *Immunity* 38: 1116-1128.
- 28 60. Sahni, H., S. Ross, A. Barbarulo, A. Solanki, C. I. Lau, A. Furmanski, J.
29 I. Saldana, M. Ono, M. Hubank, M. Barenco, and T. Crompton. 2015. A
30 genome wide transcriptional model of the complex response to pre-
31 TCR signalling during thymocyte differentiation. *Oncotarget* 6: 28646-
32 28660.
- 33 61. Hayatsu, N., T. Miyao, M. Tachibana, R. Murakami, A. Kimura, T. Kato,
34 E. Kawakami, T. A. Endo, R. Setoguchi, H. Watarai, T. Nishikawa, T.
35 Yasuda, H. Yoshida, and S. Hori. 2017. Analyses of a Mutant Foxp3
36 Allele Reveal BATF as a Critical Transcription Factor in the
37 Differentiation and Accumulation of Tissue Regulatory T Cells.
38 *Immunity* 47: 268-283 e269.
- 39 62. Kalekar, L. A., S. E. Schmiel, S. L. Nandiwada, W. Y. Lam, L. O.
40 Barsness, N. Zhang, G. L. Stritesky, D. Malhotra, K. E. Pauken, J. L.
41 Linehan, M. G. O'Sullivan, B. T. Fife, K. A. Hogquist, M. K. Jenkins,
42 and D. L. Mueller. 2016. CD4(+) T cell anergy prevents autoimmunity

- 1 and generates regulatory T cell precursors. *Nature immunology* 17:
2 304-314.
- 3 63. Takahashi, T., T. Tagami, S. Yamazaki, T. Uede, J. Shimizu, N.
4 Sakaguchi, T. W. Mak, and S. Sakaguchi. 2000. Immunologic self-
5 tolerance maintained by CD25(+)CD4(+) regulatory T cells
6 constitutively expressing cytotoxic T lymphocyte-associated antigen 4.
7 *The Journal of experimental medicine* 192: 303-310.
- 8 64. Tang, A. L., J. R. Teijaro, M. N. Njau, S. S. Chandran, A. Azimzadeh,
9 S. G. Nadler, D. M. Rothstein, and D. L. Farber. 2008. CTLA4
10 expression is an indicator and regulator of steady-state CD4+ FoxP3+
11 T cell homeostasis. *Journal of immunology* 181: 1806-1813.
- 12 65. Kenefeck, R., C. J. Wang, T. Kapadi, L. Wardzinski, K. Attridge, L. E.
13 Clough, F. Heuts, A. Kogimtzis, S. Patel, M. Rosenthal, M. Ono, D. M.
14 Sansom, P. Narendran, and L. S. Walker. 2015. Follicular helper T cell
15 signature in type 1 diabetes. *The Journal of clinical investigation* 125:
16 292-303.
- 17 66. Haghverdi, L., M. Buttner, F. A. Wolf, F. Buettner, and F. J. Theis.
18 2016. Diffusion pseudotime robustly reconstructs lineage branching.
19 *Nat Methods* 13: 845-848.

21

Table 1. Datasets used in this study

Accession number	Short description	Reference	Description of animal models	Timing of cell harvest	Cell purification strategy and sorting markers	Tissue origin	Link to Figures	Link to materials and methods
GSE15907	T cell subsets	Immunological Genome Project; Painter <i>et al.</i> , 2011	Primary cells from multiple immune lineages are isolated <i>ex-vivo</i> , primarily from young adult B6 male mice (WT, Foxp3GFP or BDC Tg mice), and double-sorted to >99% purity.	6 weeks	Flow cytometric sorting; Treg (spleen): Foxp3GFP+ CD25+ CD4+ Tmem (subcutaneous (sc)LN, spleen): TCRβ+CD44 ^{high} CD122 ^{lo} CD25- CD4+ CD44 ^{hi} CD62L ^{lo} Tmem (scLN, spleen): Foxp3GFP-TCRβ+ CD44 ^{hi} CD62L ^{lo} CD4+ Naïve CD4 (scLN): CD25- 62L ^{hi} 44 ^{lo} CD4+ Naïve CD4 (mesenteric (m) LN): CD25- CD62L ^{hi} CD44 ^{lo} CD4+ Naïve CD4 (Peyer's patches): TCRβ+ CD44 ^{lo} CD62L ^{hi} CD4+ Naïve CD4 (spleen): CD25- CD62L ^{hi} CD44 ^{lo} CD4+ Foxp3- Tnaïve (spleen): Foxp3GFP- CD44 ^{lo} CD4+ Non-dLN, BDC (scLN): BDC+ CD4+ dLN BDC (pancreatic LN): CD4+ BDC+ Tissue-Teff, BDC (pancreas): BDC+ CD4+	Spleen, subcutaneous LNs, mesenteric LN, Peyer's patches, pancreatic LN, pancreas	Figure 1 Figure 2 Figure 3	NCBI GEO

					*Exclusion markers include PI, CD8, CD11b, CD11c, CD19, CD49b, Gr-1, Ter119			
GSE83315	aTreg data	Van der Veeken <i>et al.</i> , 2016	<p>Mixed bone marrow chimeras were generated with 90% <i>Foxp3^{GFP-DTR}</i> /10% <i>Foxp3-GFP-CRE-ERT2 Rosa26YFP</i> bone marrow. Diphtheria toxin (DT) was administered at day 0 to these chimeric mice in order to deplete <i>Foxp3^{GFP-DTR}</i> Treg cells and induce expansion/activation of effector CD4+ T cells and Treg, thereby inducing inflammation. Subsequently, tamoxifen was administered at days 3 and 4 to irreversibly label <i>Foxp3</i>-expressing <i>Foxp3-GFP-CRE-ERT2 Rosa26YFP</i> T cells with YFP. Resting Treg (rTr), activated Treg (aTr), and ‘memory’ Treg (mTr) were isolated at day 0, 11, and 60, respectively, based on the dynamics of inflammation (CD4+ T cell number is normalised by day 60).</p>	Day 0 (resting Treg), day 11 (activated Treg), day 60 (memory Treg) after DT treatment	Flow cytometric sorting; rTr: CD4, Foxp3-GFP aTr and mTr: CD4, Foxp3-GFP, YFP	Spleen and peripheral LN	Figure 5C and 5D	PMC NCBI GEO

GSE61077	TCR KO data	Levine <i>et al.</i> , 2014	8-10 week mice from <i>Trac^{fllox/WT}</i> x <i>Foxp3^{ERT2-Cre}</i> (tamoxifen-inducible deletion of TCR α in Treg. Tamoxifen was administered on days 0, 1 and 3.	Day 14 after the first tamoxifen administration	Flow cytometric sorting; TCR β + CD4+ Foxp3+ CD44 ^{high/low} CD62L ^{low/high}	LN		NCBI GEO
GSE42276	T cell activation	Wakamatsu <i>et al.</i> , 2013	Conventional CD4+ T cells from C57BL/6J male mice were stimulated by anti-CD3 and anti-CD28 for 20 h and 48 h and data were pooled. 0h unstimulated samples were used as control.	8 weeks	Flow cytometric sorting; DAPI-CD45R-CD8a-CD11b/c-CD4+ GFP+	Spleen, LN	Figure 1 Figure 2 Figure 3 Figure 5	NCBI GEO
GSE6939	RV-transduced T cells	Ono <i>et al.</i> , 2007	Cells: T cells from LN and spleen of 8 week-old BALB/c mice and purified into CD4+ naive T cells (GITR ^{low} CD25-CD4 ⁺), which were subsequently activated by anti-CD3 and antigen-presenting cells (mitomycin-treated Thy1(-) splenocytes) in the presence of IL-2. On the following day, T cells were retrovirally gene transduced with Runx1 (AML1), wild type Foxp3, and empty vector as control.	60 hours after transfection	Flow cytometric sorting; CD4, GFP Exclusion marker: PI	Spleen, LN	Figure 2 Figure 3	NCBI GEO

GSE72056	Single cell analysis of tumour-infiltrating T cells	Tirosh <i>et al.</i> , 2016	<p>Single cell RNA-seq analysis of human melanoma tumour samples.</p> <p>Freshly resected samples were disaggregated to generate single cell suspensions of mixed cells of unknown identities.</p> <p>Individual viable immune (CD45+) and nonimmune (CD45-) cells (including malignant and stromal cells) were recovered from the single cell suspension by flow cytometry.</p> <p>Single cells were profiled by single-cell RNA-seq.</p>	Single cells were obtained within 45 min of tumour resection	Flow cytometric sorting; CD45	Human melanoma tissues	Figure 7 Figure 8 Figure 9	NCBI GEO
GSE15390	Human activated T cells and Treg	Beyer <i>et al.</i> , 2011	<p>Resting T cells (CD25-CD4+ T cells; GSM386262, GSM386264, and GSM386266) were obtained from whole blood of healthy human donors.</p> <p>Activated T cells (GSM777695) were prepared by stimulating CD25-CD4+ T cells for 24 hours with CD3 and IL-2.</p>	Freshly sorted from buffy coat; or cultured for 24 hours	Magnetic and flow cytometric sorting; CD4, CD25, CD127	Human PBMC	Figure 7	PMC

1 **Figure Legends**

2

3 **Figure 1. Identification of the activation signature in Treg and Tmem by**

4 **CCA of T cell populations**

5 The microarray dataset of peripheral CD4+ T cells, including naïve, effector
6 and memory phenotype from various sites (GSE15907), was analysed using
7 the T cell activation variable, which was obtained by the microarray dataset of
8 conventional activated CD4+ T cells (GSE42276). **(A)** Schematic
9 representation of CCA for the cross-level analysis of T cell populations (cells),
10 immunological processes, and genes. **(B)** CCA was applied to the T cell
11 population data using an explanatory variable for T cell activation, which was
12 obtained as fold change between activated and resting conventional CD4+ T
13 cells. The CCA solution is thus one-dimensional, and is used as “T cell
14 activation score” (see Methods).

15

16 **Figure 2. Identification of the Foxp3-independent activation signature in**

17 **Treg by CCA of T cell populations**

18 The microarray dataset of peripheral CD4+ T cells (GSE15907) was analysed
19 using the T cell activation variable and the variables for retroviral *Foxp3*
20 transduction and *Runx1* transduction as explanatory variables. **(A)** The CCA
21 solution was visualised by a biplot where CD4+ T cell samples are shown by
22 closed circles (see legend) and the explanatory variables are shown by blue
23 arrows. Percentage indicates that of the variance accounted for by the inertia
24 of the axes (i.e. the amount of information (eigenvalue) retained in each axis).

1 (B) Gene biplot of the 2D CCA solution in (C) showing the relationships
2 between genes (grey circles) and the explanatory variables (blue arrows).
3 Selected key genes are annotated.

4

5 **Figure 3. Differential regulations of transcriptional modules for**
6 **activation in Treg and Tmem by Foxp3 and Runx1**

7 (A) Definition of Tact-Foxp3 genes and Tact-Runx1 genes. In the gene plot of
8 the CCA solution in Figure 2B, Axis 1-low genes (25 percentile low) were
9 designated as *activation genes*, which were further classified into Tact-Foxp3
10 genes and Tact-Runx1 genes by Axis 2, which have high correlations to Treg
11 and Tmem samples, respectively, in the CCA cell space (Figure 2A). (B)
12 Heatmap analysis of all the Tact-Foxp3 genes. (C) Heatmap analysis of all the
13 Tact-Runx1 genes. (D) Heatmap analysis of selected Tact-Foxp3 genes. (E)
14 Heatmap analysis of selected Tact-Runx1 genes.

15

16 **Figure 4. A model for the differential regulation of activation genes in**
17 **Treg and Tmem**

18 The proposed differential regulations of TCR signal downstream genes in
19 Treg and Tmem. Since both naturally-arising Treg and Tmem are self-reactive
20 T cells, they may frequently receive tonic TCR signals by recognising their
21 cognate antigens in the periphery. This results in the full activation of both the
22 Tact-Foxp3 and Tact-Runx1 gene modules in Tmem. However, in Treg,
23 Foxp3 represses Tact-Runx1 genes and sustains the expression of Tact-
24 Foxp3 genes, producing the characteristic Treg transcriptome.

25

Figure 5. The activation signature of Treg is dependent on TCR signalling

(A) The experimental design for the TCR dataset. CD44^{hi} activated Treg and CD44^{lo} naïve-like Treg were obtained from TCRα KO or WT mice and analysed by transcriptome analysis. (B) CCA was applied to the transcriptome data of CD44^{lo}CD62^{hi} naïve-like and CD44^{hi}CD62^{lo} activated Treg cell populations from inducible *TCRα* KO or WT (from the *TCR* KO data, GSE61077), using the T cell activation variable as the explanatory variable. This produces a 1D CCA solution, and the sample score was plotted (representing “T cell activation score”). (C) The experimental design for the activated Treg dataset. Bone marrow (BM) cells were obtained from *Foxp3*^{GFP}*CreERT2*:*Rosa26YFP* mice (YFP mice), and transferred into *Foxp3*^{GFP}*DTR* mice (Foxp3-DTR mice), in order to make BM chimera, in which ~10% of Treg expressed DTR. Subsequently, DT was administered to these BM chimera, which depleted Foxp3-DTR cells but not donor cells. This treatment induced a transient activation of T cells and inflammation *in vivo*. Activated Treg (aTreg) were obtained from these mice with inflammation, while resting Treg (rTreg) were from control mice, and memory Treg (mTreg) were from the mice after the resolution of inflammation. (D) 1D CCA sample score plot of transcriptomic data of resting Treg (rTreg), *in vivo* activated Treg (aTreg) and memory Treg (mTreg) from the *aTreg* data (GSE83315), with T cell activation signature as explanatory variable.

Figure 6. The comparative analysis of Tmem-specific and Treg-Tmem shared activation genes and TCR-dependent and activated Treg-specific genes

Venn diagram analysis was used to obtain intersects of *TCR-dependent genes* (DEG between TCRa KO and WT Treg), *aTreg-specific genes* (DEG between aTreg and rTreg), and Tact-Foxp3 and Tact-Runx1 genes (see Figure 3). **(A)** Pie chart showing the number of genes in the intersects between *aTreg-specific genes*, *TCR-dependent genes*, and Tact-Runx1 genes. **(B)** Pie chart showing the numbers of *aTreg-specific genes*, *TCR-dependent genes*, and Tact-Foxp3 genes. **(C)** Pathway analysis of Tact-Foxp3 genes, Tact-Runx1 genes and *aTreg-specific genes* showing enriched pathways in these gene lists.

Figure 7. Single cell CCA of melanoma-infiltrating T cells determines the activation status of individual T cells and identifies a putative Tfh-like process

(A) Schematic representation of CCA of CD4+ T cell single cell transcriptomes analysed by two explanatory variables: activated naïve T cells (Tact) and resting naïve T cells (Trest). **(B)** CCA biplot showing the relationship between Treg and non-Treg T cells (sample scores) and the explanatory variables (Tact and Trest). Axis 1 represents the difference between Tact and Trest, and thus, Activated T cells and Resting T cells were defined by the CCA Axis 1 score, and these cells were further classified into Treg and non-Treg by their *FOXP3* expression (see legend). Percentage

1 indicates that of the variance (inertia) accounted for by the axis. **(C)** Violin plot
 2 showing the CCA activation scores (Axis 1 score \times -1) of FOXP3- and
 3 FOXP3+ cell groups. Asterisk indicates statistical significance by Mann-
 4 Whitney test **(D)** Violin plot showing the CCA activation scores of Activated
 5 (Act.) and Resting (Rest.) FOXP3- and FOXP3+ cell groups. Asterisks
 6 indicate the values of post-hoc Dunn's test following a Kruskal Wallis test. ***
 7 $p < 0.005$. **(E)** Gene biplot of the CCA solution in (B) showing the relationships
 8 between genes (grey circles) and the Tact and Trest explanatory variables
 9 (blue arrows). Genes are shown by grey circles, and well-known genes that
 10 are key for T cell activation processes are annotated.

11

12

13 **Figure 8. SC4A identifies the bifurcation point of activated T cells that**
 14 **leads to Tfh-like and Treg differentiation in tumour-infiltrating T cells**

15 SC4A was applied to the single cell data of tumour-infiltrating T cells, and 4
 16 genes (CTLA-4, CCR7, FOXP3, and PDCD1) were chosen as explanatory
 17 variables to represent the T cell activation, resting, FOXP3-driven process,
 18 and Tfh-like process. **(A)** The design of analysis. The single cell data from the
 19 melanoma samples were analysed by SC4A to identify the most effective
 20 combinations of explanatory variables for dispersing the 4 presumptive T cell
 21 populations identified in Figure 7. These genes were used as explanatory
 22 variables to analyse the rest of the single cell data as main dataset. Thus, the
 23 single-cell level dynamics of T cell differentiation and activation are modelled
 24 by the key biological processes that are represented by the T cell populations
 25 and explanatory variables. **(B)** Single cell sample space of the final SC4A
 26 output showing correlations between single cell samples and the explanatory

1 variables **(C)** Gene space of the final SC4A output showing correlations
2 between genes and the explanatory variables. The genes that showed high
3 correlations to the PDCD1, CTLA4, and FOXP3 variables were identified as
4 Tfh-like genes, Activation genes, and FOXP3-driven genes, respectively. **(D)**
5 The identification of two differentiation processes as lineages and a
6 bifurcation point. The cells in the sample space of the SC4A output (B) were
7 classified into 6 clusters by an unsupervised clustering algorithm. These
8 clusters were further analysed for pseudotime inference. **(E-G)** The average
9 gene expression was plotted against each pseudotime (upper: FOXP3-
10 pseudotime; lower: Tfh-pseudotime). The bifurcation point (Cluster II) is
11 emphasised by broken lines. The numbers in circle indicate the cluster
12 number. Gene expression was standardised, and the sum of the standardised
13 expression was obtained for **(E)** Activation genes, **(F)** FOXP3-driven genes,
14 and **(G)** Tfh-like genes (see C). **(H-M)** The expression of key genes was
15 plotted against each pseudotime.

16

17 **Figure 9. Identification of the conserved genes for the differential**
18 **regulation of Tfh-like and Treg differentiation in activated T cells**

19 **(A)** The identified lineage curves and the bifurcation point in the tumour-
20 infiltrating T cells. The number in circle indicates the cluster number in Figure
21 8D. **(B-D)** The expression of selected feature genes was plotted against each
22 pseudotime. Genes are from the intersect of **(B)** Activation genes (Figure 8C)
23 and Tact-Foxp3 genes (Figure 3), **(C)** FOXP3-driven genes (Figure 8C) and
24 Tact-Foxp3 genes, and **(D)** Tfh-like genes (Figure 8C) and Tact-Runx1 genes
25 (Figure 3). **(E)** The expression of selected genes in the tumour-infiltrating T

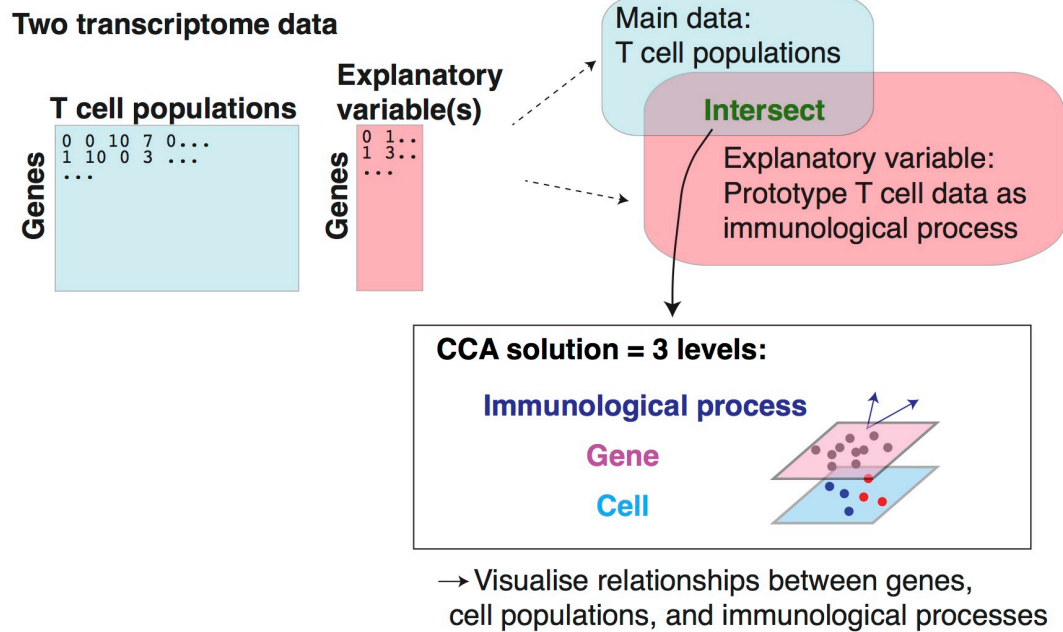
1 cells was shown by a 2-dimensional plot in a flow cytometric style. Data from
 2 Treg-lineage cells (Cluster V and VI, upper panels) and Tfh-like lineage cells
 3 (Cluster III and IV, lower panels). The gene in x-axis (NFAT5) is from the
 4 activation gene group (B), while y-axis shows genes from either the FOXP3-
 5 Treg group (C) or the Tfh-like/Tmem group (D). Thresholds and quadrant
 6 gates were determined in an empirical manner using density plot.

7

Figure 1

A

Gene-oriented CCA of T cell populations and immunological processes



B

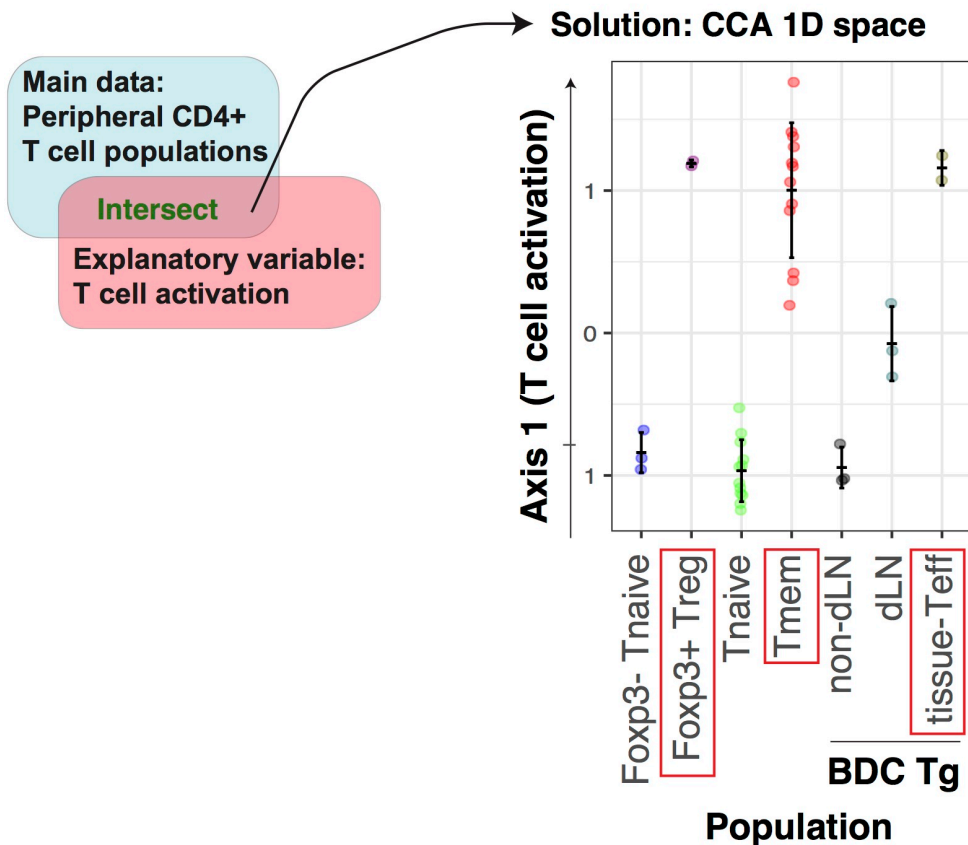
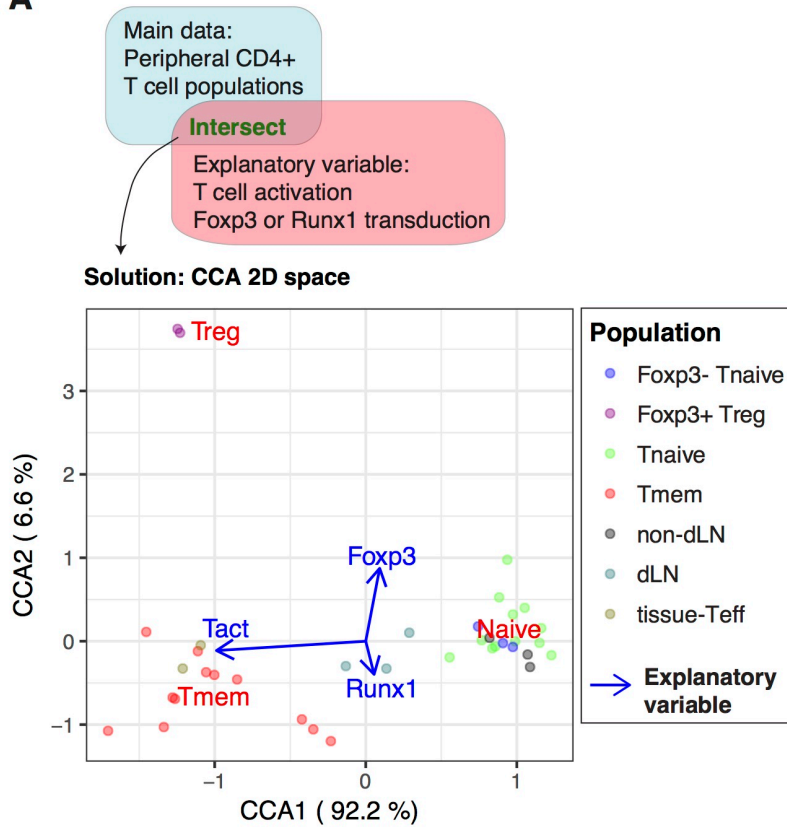


Figure 2

A



B

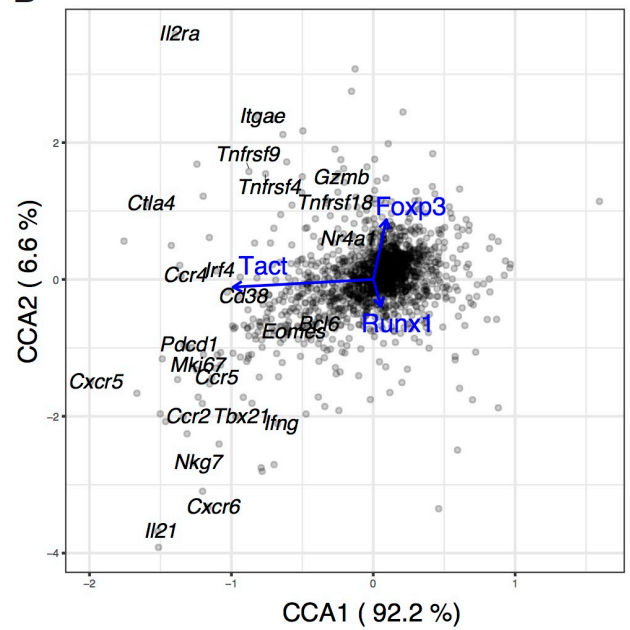


Figure 3

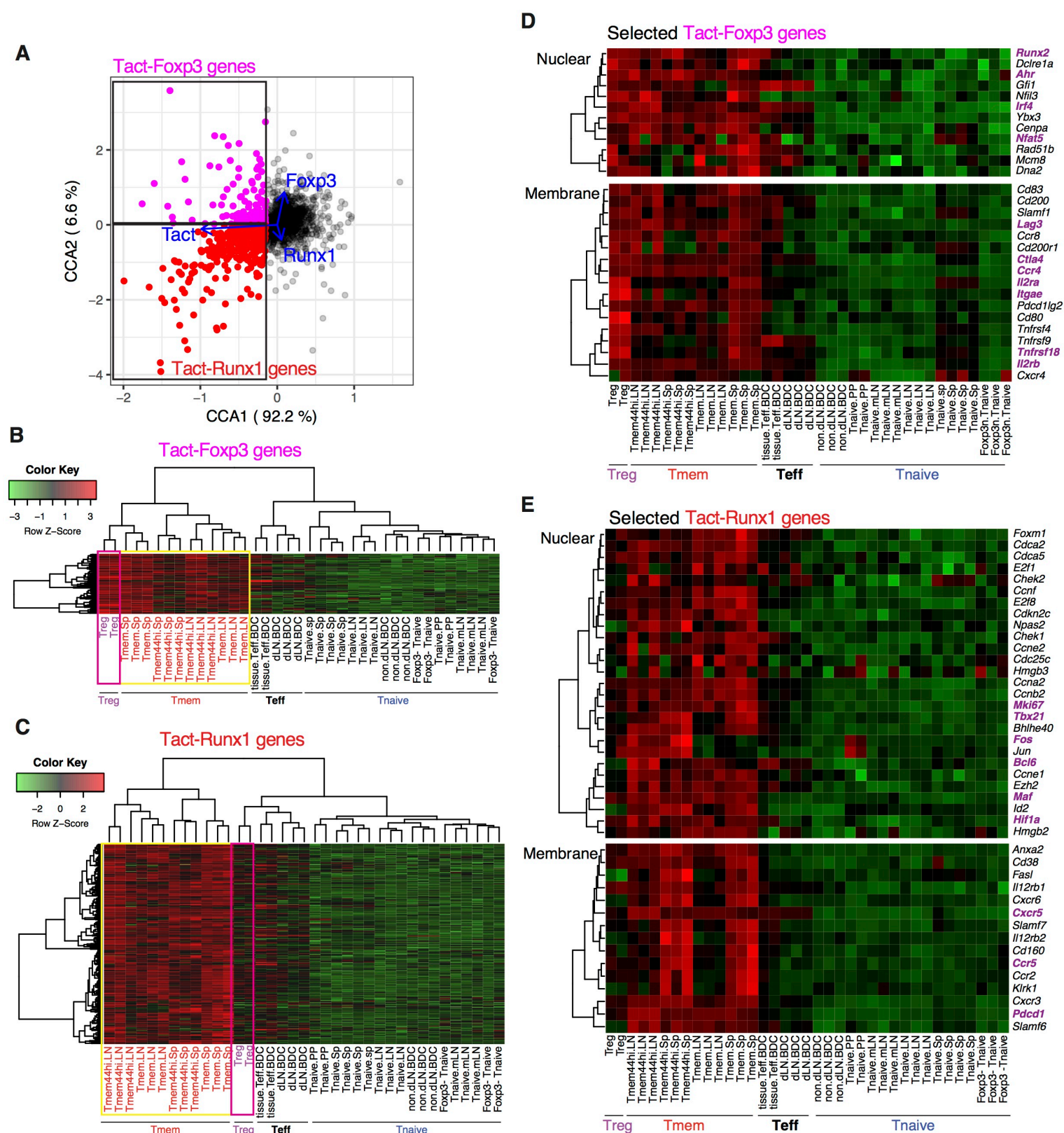


Figure 4

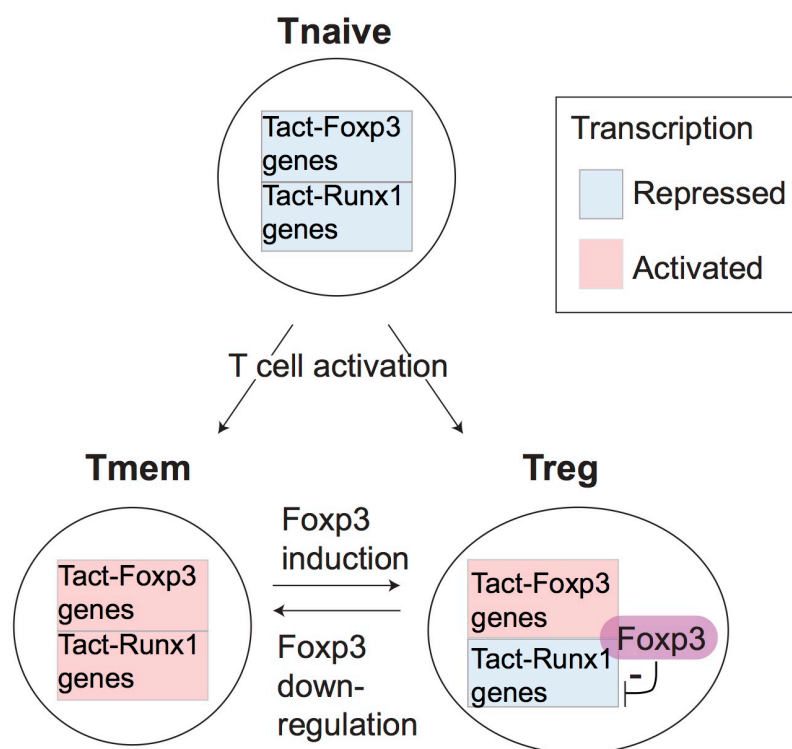


Figure 5

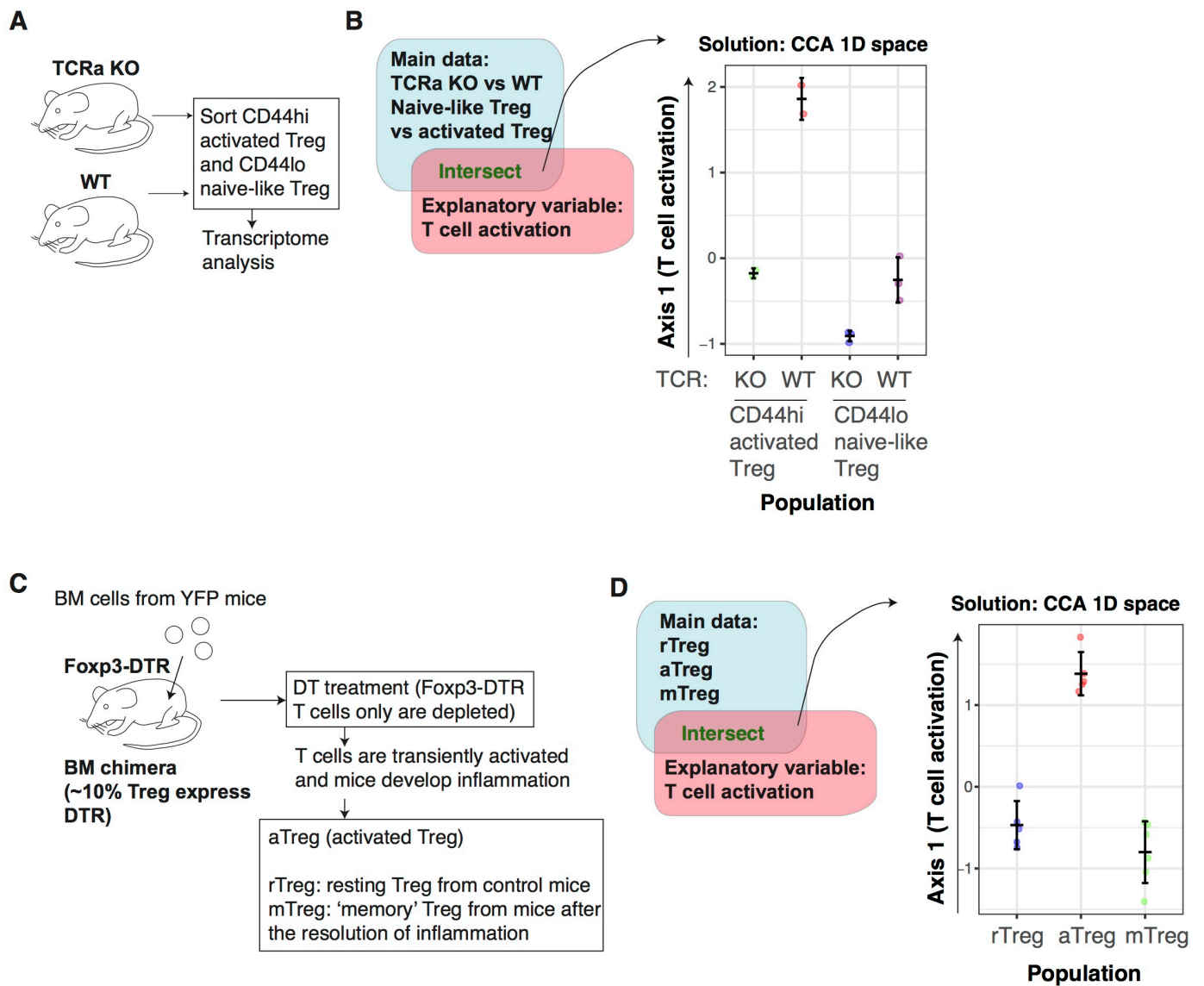


Figure 6

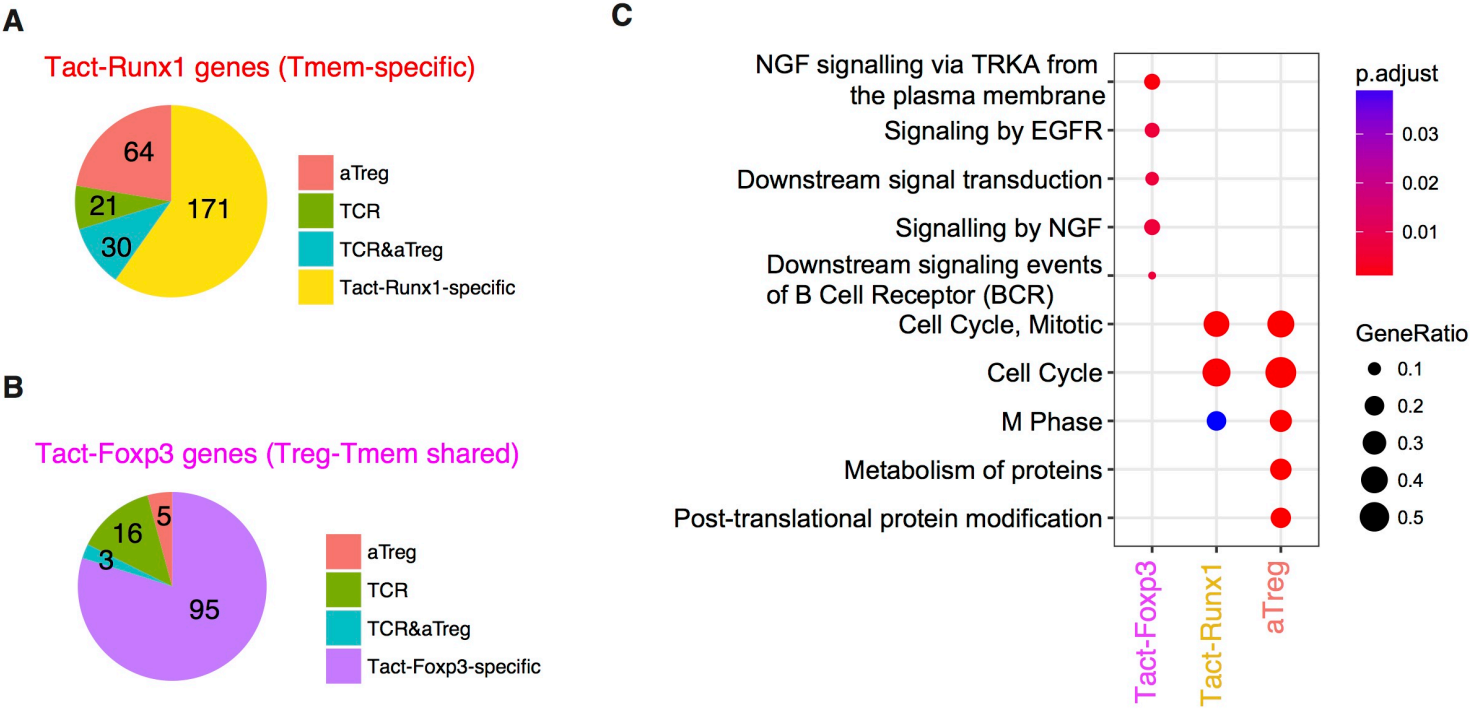


Figure 7

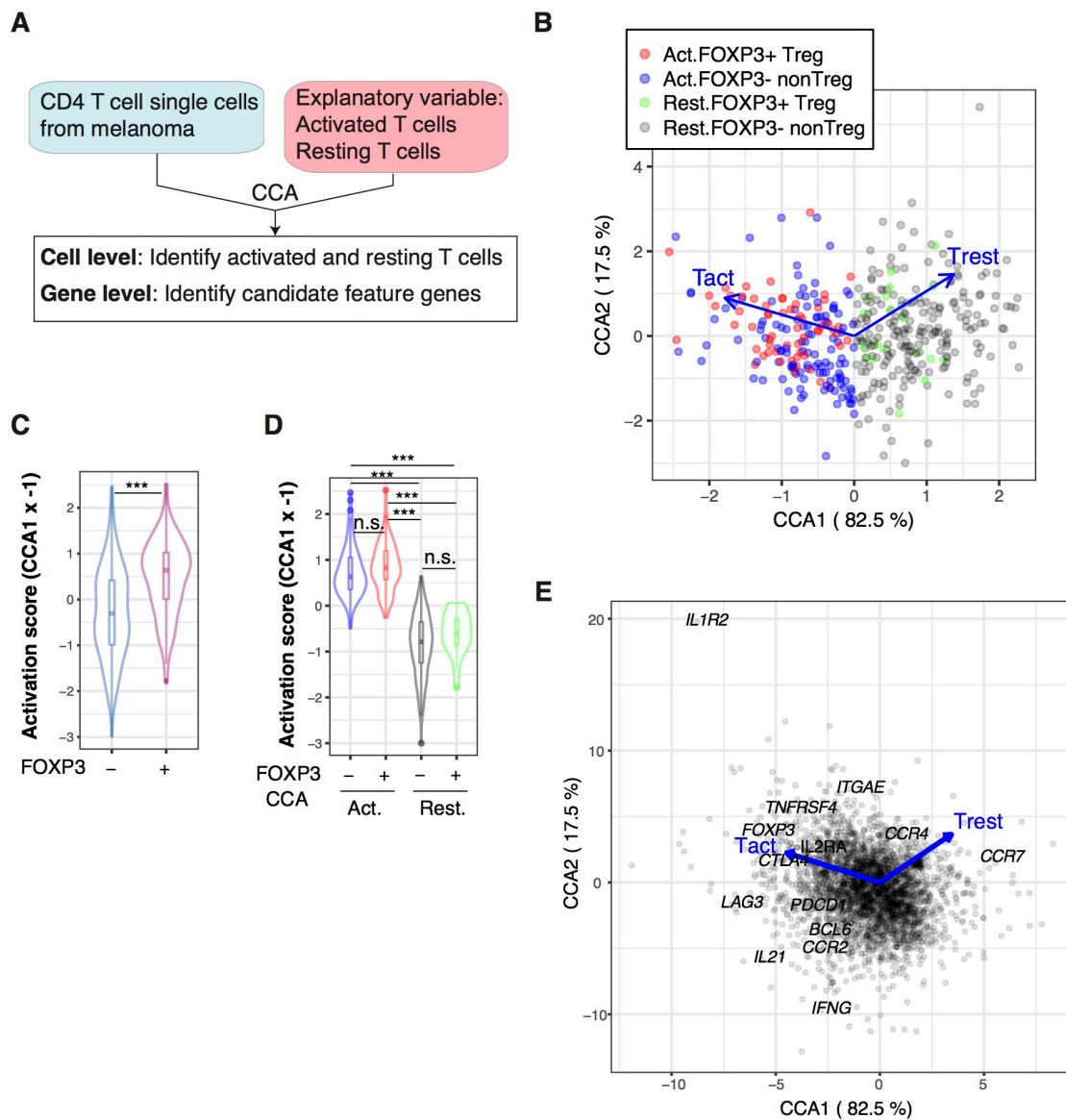


Figure 8

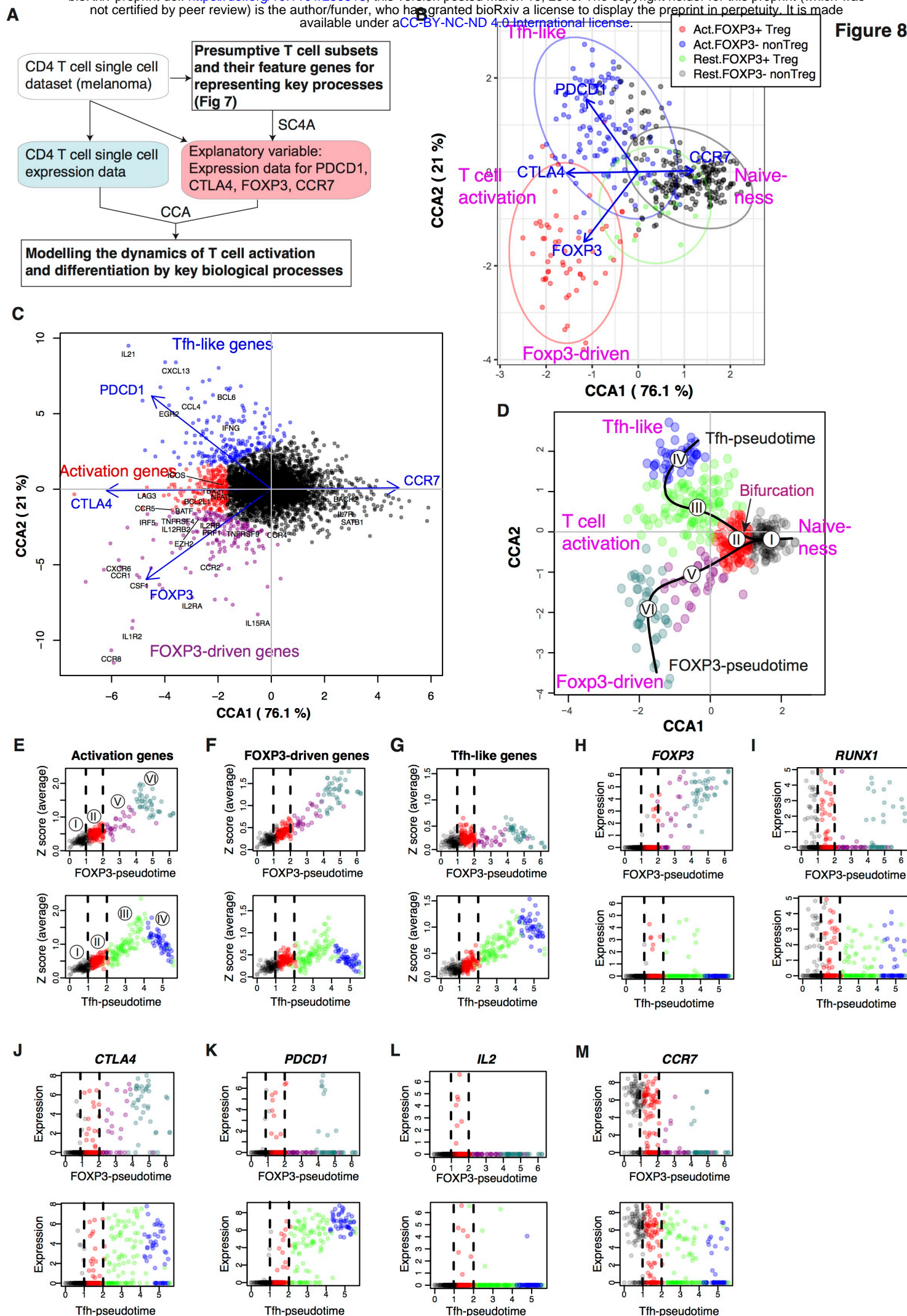


Figure 9

