Title: Differential regulation of activation and differentiation processes in CD4+ T cell populations and their single-cell level heterogeneity

Authors:

Alla Bradley[1], Tetsuo Hashimoto[1] and Masahiro Ono[2§]

Affiliations:

[1] Faculty of Life and Environmental Sciences, University of Tsukuba, Tsukuba, Japan

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

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

Abstract

In T cells, T cell receptor (TCR) signalling initiates downstream transcriptional mechanisms for T cell activation and differentiation. Foxp3-expressing regulatory T cells (Treg) require TCR signals for their suppressive function and maintenance in the periphery. It is, however, unclear how TCR signalling controls the transcriptional programme of Treg. Since most of studies identified the transcriptional features of Treg in comparison to naïve T cells, the relationship between Treg and non-naïve T cells including memory-phenotype T cells (Tmem) and effector T cells (Teff) is not well understood. Here we dissect the transcriptomes of various T cell subsets from independent datasets using the multidimensional analysis method Canonical Correspondence Analysis (CCA). We show that resting Treg share gene modules for activation with Tmem and Teff. Importantly, Tmem activate the distinct transcriptional modules for T cell activation, which are uniquely repressed in Treg. The activation signature of Treg is dependent on TCR signals, and is more actively operating in activated Treg. Furthermore, by analysing single cell RNA-seq data from tumour-infiltrating T cells, we identified the common shared transcriptional modules for T cell activation, including *CTLA-4*, in activated Treg and activated Teff. Moreover, we identified distinct FOXP3-driven and T follicular helper-like transcriptional modules in activated FOXP3+ Treg and FOXP3- Teff, respectively. Collectively, we reveal the multidimensional identities and single cell-level heterogeneity of Treg, identifying the differential regulation of the activation and differentiation gene modules in Treg, Tmem and Teff during homeostasis and in the tumour microenvironment.

Introduction

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

In fact, TCR signalling plays a critical role in Treg. Studies using TCR transgenic showed that Treg require TCR activation for *in vitro* suppression (8). Foxp3 binds to the enhancer regions that have been opened by TCR signals, which explains a major part of the Treg-type chromatin structure (9). In fact, continuous TCR signals are required for Treg function, because the conditional deletion of the TCR-α chain in Treg abrogates the suppressive activity of Treg and eliminates their activated or effector-Treg type phenotype (10, 11). It is, however, unclear how TCR signals contribute to the Treg-type transcriptional programme, and whether TCR signals are operating in all Treg

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

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

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

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

Multidimensional analysis is an effective approach to address this problem, allowing to systematically investigate relationships between more than 2 populations (e.g. based on transcriptional similarities) (31). The prototype methods include Principal Component Analysis (PCA), Correspondence Analysis (CA) (32) and Multidimensional Scaling (33). In the application to genomic data, these methods measure distances (i.e. similarities) between samples and/or genes using different metrics, and thereby visualise the relationships between samples and/or genes in a reduced dimension, typically either in 2- or 3-dimensions, providing means to explore and investigate data

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

In this study, we investigate the multidimensional features of Treg in comparison to other CD4+ T cells including Teff, Tmem, and naïve T cells under normal or pathological conditions. Here we aim to identify the differential regulation of transcriptional modules for T cell activation and differentiation in these populations, and to reveal systems and molecular mechanisms behind the differential regulation. Furthermore, using a new single cell CCA approach, we investigate the single-cell level heterogeneity of CD4+ T cells including Treg and effector T cells, identifying the differentially regulated gene modules and the dynamic and gradational changes in transcriptomes of individual T cells.

Materials and Methods

Conventional CCA (Gene-oriented analysis)

CCA canonically analyses two independent microarray or RNA-seq datasets (34). Briefly, gene expression data of the standardised main dataset (**S**) is linearly regressed onto the explanatory variable(s) (**D**), which identifies the interpretable part of the main dataset ("Constrained data", **S***). When only one explanatory variable is used, the CA algorithm of CCA assigns numerical values to cell samples and genes so that the dispersion of samples is maximised (uncorrelated information components), providing a one-dimensional solution (34). In order to use the solution as a scoring system, CCA score (i.e. Axis 1 score) was multiplied by the single biplot value, which indicates positive or negative correlation to Axis 1, which ensures that cells and genes with high scores have high positive correlations to the explanatory variable. When two or more explanatory variables are used, the CA algorithm then performs SVD on **S***, creating new matrices (i.e. sample and gene score matrices). These scores are sorted into new uncorrelated axes α_k, along which the entire set of scores generated by SVD, are distributed. The first axis holds the greatest amount of information (largest variations, precisely, *inertia*). The map approach enables the comparison of more than two explanatory variables, while the regression process in CCA allows the analysis across two different experiments (34). Biplot values of the CCA result are shown by arrows on the CCA map. CCA provides a map that shows the correlations between samples of interest, explanatory variables, and genes. Highly correlated components are closely positioned on the map.

In the application of CCA to population data, transcriptomic datasets of peripheral CD4+ T cells (including Treg, naïve, memory, draining LN and non-dLN and tissue effector CD4+ T cells) were processed by CCA and the cross-level relationships between components at three different levels, namely, immunological process, gene and cell, were analysed. Note that the same genes must be used in both transcriptome dataset matrices (intersect). The main dataset is projected onto the explanatory variable dataset, thus the genes in common to both datasets comprise the interpretable part of the main data. Mathematical operation implemented in the CCA algorithm produces immunological process (explanatory variable), gene and cell sample scores. The results are visualized as the 3-dimensional CCA solution on the CCA map (i.e. CCA triplot) that shows the relationships between cell subsets, genes and immunological processes.

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

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

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

Explanatory variables for conventional CCA

Explanatory variables for CCA were prepared as follows. Differentially-expressed genes were selected by a moderated t-test result using the Bioconductor package, *limma*. The top-ranked differentially expressed genes (according to their *p*-values) were used for making the explanatory variables. The T cell activation explanatory variable were defined by the difference in gene expression between anti-CD3/CD28-stimulated (17 h) $CD4^+$ T cells and untreated $CD4^+$ T cells from GEO42276 (36). Precisely, genes were selected by FDR <0.01 and log 2 fold change (> 1 or < -1) in the comparison of the gene expression profile of the activated and resting T cells. For Figure 1A, the expression data of GSE15907 (37) was regressed onto the log2 fold change of activated CD4+ T cells (17 h after activation) and naïve CD4+ T cells from GSE42276 as 'T cell activation signature' explanatory variable, and Correspondence Analysis was performed for the regressed data and the correlation analysis was done between the new axis and the explanatory variable. For Figure 1C-D, the expression data of GSE15907 (37) was regressed onto the 'T cell activation signature' explanatory variable described

above, in combination with retrovirally gene-transduced T cells for Foxp3 and Runx1 effects (GSE6939 (38)): the log2 fold change of Foxp3-transduced and empty vector-transduced CD4+T cells as the 'Foxp3 effects' explanatory variable, and a log2 fold change of Runx1-transduced and empty vector-transduced CD4+T cells as 'Runx1 effects' explanatory variable. Correspondence Analysis was performed for the regressed data and the correlation analysis was done between the new axis and the explanatory variable.

Choice of explanatory variables by SC4A

SC4A aims to identify a set of genes that make the dispersion of cell populations maximum in the CCA solution. To achieve this, all the combinations of genes will be used as explanatory variables and tested for discriminating each two populations using CCA. During each combinatorial cycle, two genes are chosen from the total selected genes for all defined single-cell populations in the main dataset and tested for their correlations to one defined cell population vs all other T cells. In the analysis of Figure 8, the following two cell populations were analysed by the combinatorial CCA: (1) Activated T cells vs Resting T cells; (2) FOXP3+ cells vs FOXP3- cells; (3) BCL6+ cells (as Tfh-like T cells) vs BCL6- cells. The most correlated gene to each population (Activated T cells, Resting T cells, FOXP3+ cells, or BCL6+ cells) was identified, and these 4 genes were used as explanatory variables in the final output of SC4A in Figure 8.

Data pre-processing and other statistical methods

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

Heatmaps were generated the CRAN package *gplots*. Venn diagram was generated using the R code, overLapper.R, which was downloaded from the Girke lab at Institute for Integrative Genome Biology (http://faculty.ucr.edu/~tgirke/Documents/R_BioCond/My_R_Scripts/overLapper.R). Gene lists were compared for enriched pathways in the REACTOME pathway database using the Bioconductor packages *ReactomePA* and *clusterProfiler*. Violin plots were generated by the Bioconductor package *ggplot2*. The inside part of the violin plot shows the median and interquartile range (IQR) of the original gene expression data. The lineage curve was constructed by clustering SC4A/CCA sample scores using an expectation–maximization (EM) algorithm (39), and the nodes of these clusters were

identified by constructing a minimum spanning tree using the Bioconductor package *Slingshot* (40).

Results

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

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

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

The Treg transcriptome is characterized by the repression of a part of the activation genes in Tmem transcriptome

Next, we determine the modules of genes that are differentially regulated between Treg and Tmem, in order to reveal the multidimensional identity of Treg and Tmem transcriptomes. Specifically, we asked if the Axis 2 captured the differential transcriptional regulations between Tmem and Treg. Importantly, Axis 2 represents Foxp3-driven and Runx1-driven transcriptional effects, which are correlated with Treg and Tmem/Teff, respectively (**Figure 3A**). This suggests that Axis 2 provides a 'scoring system' for regulatory vs effector functions. Thus, the genes in Axis 1-low (precisely, genes above 25 percentile for positive correlations with Tact) were identified as *Tact genes*, which were subsequently classified into Axis 2-positive (i.e. positive correlations with Foxp3 and Treg) [designated as "*Tact-Foxp3 genes*"; top left quadrant of CCA gene space in Figure 1D] and Axis 2-negative genes (i.e. positive correlations with Runx1 and Tmem/Teff) [designated as "*Tact-Runx1 genes*"; bottom left quadrant of CCA gene space in Figure 1D] (**Figure 3A**). Tact-Runx1 genes contain genes linked to T cell activation (e.g. *Mki67*), effector functions (e.g. *Tbx21*), and Tfh (e.g. *Bcl6*, *Pdcd1*), while Tact-Foxp3 genes contain "Treg markers" such as *Il2ra* (CD25) and *Tnfrsf18* (GITR) (**Figure 2B**). These encouraged us to further analyse the gene space of the CCA solution, aiming to capture the unique regulation of activation genes in Foxp3+ Treg in comparison to Foxp3- non-naive T cells which include Tmem and Teff.

Intriguingly, heatmap analysis showed that both Treg and Tmem expressed Tact-Foxp3 genes at high levels, compared to naïve and effector T cells (**Figure 3B**). On the other hand, Tact-Runx1 genes were selectively downregulated in Treg, while their expressions were sustained in Tmem (**Figure 3C**). In other words, the repression of Tact-Runx1 genes was the major feature of Treg in comparison to Tmem (**Figure 2A**). Interestingly, comparable selective downregulation of Tact-Runx1 genes was observed in Teff as well (**Figure 3C**). This suggests that the set of activation genes operating in Teff is different from the ones in Tmem, and that Tmem and Treg share more activation genes than Treg-Teff and Tmem-Teff (**Figure 3B and 3C**). These collectively indicate that the Treg-ness is composed of the induction of the Treg-Tmem shared activation genes and the repression of Tmem-specific genes, defining the multidimensional identity of Treg.

While the overall activation levels of Treg and Tmem are similar to the ones of Teff at transcriptional level (Figure 1B), when explained by the prototype signature of activation in CD4+ T cells (i.e. the explanatory variable Tact), the compositions of the activation genes are different between Treg, Tmem and Teff (as captured by Figure 3B and 3C). Importantly, many of these activation genes are shared between Treg and Tmem, but not with Teff. The closer similarity between resting Treg and Tmem, compared to Teff, is not surprising, considering that both resting Treg and Tmem are considered to be at the resting steady-state, while Teff are more recently activated and executing effector functions, which presumably require unique sets of genes. These features were not captured by standard t-test analysis (**Supplementary Fig 1**), presumably due to the lack of multidimensional perspective.

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

Tact-Runx1 genes included many cell cycle-related genes (e.g. *Ccna1*, *Cdca2*, and *Chek2*), suggesting that these cells are in cell cycle and proliferating (**Figure 3E**). The higher expression of *Mki67* and *Fos* suggests that these Tmem cells had been activated by TCR signals *in vivo* before the analysis. Tact-Runx1 genes also included the transcription factors *Tbx21*, *Maf*, *Hif1a*, and *Bcl6*, which have roles in Th1, Th2, Th17, and Tfh differentiation, respectively (42-44). In accordance with this, the Tfh markers *Cxcr5* and *Pdcd1* were specifically expressed by Tmem. These results are compatible with the model that Treg and Tmem constitute the self-reactive T cell population that have constitutive activation status (7), and that the major function of Foxp3 is to modify the constitutive activation processes by repressing a part of the activation gene modules (i.e. Tact-Runx1 genes) (**Figure 4**).

The activated status of Treg is TCR signal dependent

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

In order to further address whether the TCR signal-dependent activation signature of Treg is constitutively maintained or specifically induced by *in vivo* activation events (presumably as tonic TCR signals (7)), we analysed the RNA-seq dataset of *in vivo* activated Treg (Ref. (16), **Table 1**). The dataset was generated by depleting a part of Treg by Diphtheria toxin (DT) using bone marrow chimera of $Foxp3^{GFPCreERT2}$*:Rosa26YFP* and $Foxp3^{GFP\ DTR}$ (16). The DT treatment depletes DT receptor (DTR)-expressing Treg from $Foxp3^{GFP\ DTR}$, and thus induces a transient inflammation through the reduction of Treg. Van der Veeken *et al* thus analysed resting Treg from untreated mice (rTreg), activated Treg from mice with recent depletion (11 days before the analysis)

in an inflammatory condition (aTreg), and “memory” Treg (mTreg) from mice with a distant depletion (60 days before the analysis) (**Figure 5C**). As expected, the CCA analysis using the T cell activation variable showed that aTreg had higher activation scores than both rTreg and mTreg (**Figure 5D**). This indicates that the activation mechanisms are more actively operating in activated Treg in an inflammatory environment.

In order to further dissect the activation signature of Treg, we obtained the lists of differentially expressed genes (DEG) between WT Treg vs *TCRa* KO Treg (designated as *TCR-dependent genes*), and between aTreg and rTreg (designated as *aTreg-specific genes*, see Methods). Interestingly, 94/286 genes of Tact-Runx1 genes (Tmem-specific activation genes, repressed in resting Treg) are also used during the activation of Treg (**Figure 6A**), while only 8/119 of Tact-Foxp3 genes (used by Tmem and resting Treg) are induced during the activation of Treg (**Figure 6B**). This indicates that the activation of Treg does not enhance the genes that are used in resting Treg, but induces the expression of the Tmem-specific genes that are suppressed in resting Treg. On the other hand, 51/286 of Tact-Runx1 and 19/119 of Tact-Foxp3 genes are regulated by TCR signalling by TCRa KO (**Figure 6A and 6B**), suggesting that the activation status of resting Treg and Tmem may be sustained by TCR signals. Pathway analysis showed that Tact-Runx1 and aTreg-specific genes were enriched for cell-cycle related pathways. In contrast, Tact-Foxp3 genes were enriched for pathways related to signal transduction only (**Figure 6C**). Collectively, the results above suggest that resting Treg are maintained by TCR and cytokine signalling, and that the

activation of Treg induces the transcriptional activities of Tact-Runx1 genes, which promote proliferation and cell division.

Tumour-infiltrating FOXP3+ Treg are more enriched in activated T cells than resting cells by single cell CCA

The analyses above led us to hypothesise that the activation status of Treg is variable at the single cell level in physiological settings. In order to address this predicted heterogeneity in Treg, we investigated single cell data and further addressed how the activation mechanisms are operating in Treg at the single cell level. In order to address the differential regulations of activation mechanisms in individual Treg and related T cells, firstly, the features of individual cells needed to be characterized in a data-oriented manner, as no annotation data were available for individual single cells. CCA is a powerful method for identifying biological meanings, and we applied CCA to the single cell RNA-seq data of tumour-infiltrating T cells from human patients (Ref. (35) **Table 1**).

Firstly, we applied the standard CCA to the single cell RNA-seq data of $CD4^{+}CD3^{+}$ T cells (single-cell T cell samples of Treg and CD4+ non-Treg cells with unknown individual activation and differentiation statuses), using the explanatory variables of activated conventional CD4+ T cells (Tact) and resting T cells (Trest; GSE15390, **Table 1**), aiming to define activated T cells and resting T cells by the correlations to these two variables (**Figure 7A**) in Axis 1. Here we used these two variables, Tact and Trest, instead of the log2 fold changes between the two populations (i.e. T cell activation variable,

which produces a 1-dimensional CCA solution visualized as a single axis), because we aimed to identify an additional major differentiation process(es) in the Axis 2 (i.e. two explanatory variables produce a 2-dimensional result). In the single cell space of the CCA solution, the majority of FOXP3+ T cells had negative gene scores in the Axis 1, i.e. showing high positive correlations to the T cell activation variable (**Figure 7B**). Here, CCA Axis 1 x (-1) is designated as the *Activation Score*. Thus, using the Axis 1 score and FOXP3 expression, the following 4 subpopulations were defined: “Activated FOXP3+”, “Resting FOXP3+”, “Activated FOXP3-”, and “Resting FOXP3-” (**Figure 7B**).

Next, we aimed to determine whether activated Treg are more activated than resting Treg at the single cell level, or whether Treg are enriched with activated T cells. Establishing the T cell activation score by Axis 1 score (as the correlation to Tact in Figure 4B, see Methods), FOXP3^{+} Treg had significantly higher scores than FOXP3- non-Treg on average, as indicated by the higher median in the violin plots and greater density of samples with higher CCA gene scores for T cell activation variable (**Figure 7C**). Using the CCA definition of Activated and Resting Treg and non-Treg in Figure 4B, the activation score neatly captured the activated status of single cells, allocating high positive and negative scores to Activated and Resting cells, respectively (**Figure 7D**). Importantly, there was no significant difference between Activated FOXP3+ and Activated FOXP3- cells and between Resting FOXP3+ and Resting FOXP3- cells (**Figure 7D**), indicating that in tumour microenvironment, Treg cells are as activated as non-Treg CD4+ T cells, which may include Teff, Tmem, and Tfh. These results suggest that the Treg population have an activated signature because FOXP3+ cells are enriched

with T cells that have recognised antigens and received TCR signals (i.e. activated T cells), and that TCR signals can induce FOXP3 in these antigen-experienced T cells. Alternatively, but not exclusively, FOXP3+ T cells may have high-affinity TCRs to self-MHC and/or tumour antigens and be more prone to be activated. In fact, strikingly, 32.5% of activated T cells by the CCA result expressed FOXP3, while only 8.2% of resting T cells expressed FOXP3. In other words, FOXP3 expression occurred more frequently in activated T cells.

In the gene space of the CCA solution, genes with strong correlations to activated FOXP3^{+} T cells included *FOXP3* itself and common Treg markers such as *CTLA4* and *IL2RA* (CD25), which were found in the upper left quadrant (Axis 1-negative Axis 2-positive). Interestingly, the lower left quadrant (Axis 1-negative Axis 2-negative) contained more Tfh-like or effector-like molecules *PDCD1* (PD-1), *BCL6*, *IL21*, and *IFNG.* The chemokine receptors *CCR5* and *CCR2* had negative scores in Axis 1 (i.e. correlated with Tact), while *CCR7* had a high positive score in Axis 1 (i.e. correlated with Trest) (**Figure 7E**).

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

Next, we aimed to identify major differentiation and activation processes in the single cell transcriptomes above. To this end, we have employed a new CCA approach using Single cell analysis (Single Cell Combinatorial CCA, SC4A), which aims to construct a CCA model of single cell data and thereby to

identify major differentiation/activation processes and the underlying gene regulations (**Figure 8A,** see Methods and Supplementary Text). Firstly, we classified single cells into Activated and Resting cells, and FOXP3+ Treg and FOXP3- non-Treg, and thereby identified the following 4 processes as putative differentiation and activation processes in the dataset: T cell activation (Activated cells), and naïve-ness (Resting cells), FOXP3-driven process (Activated FOXP3+), and Tfh-like process (Activated FOXP3-) (Figure 4). Secondly, based on their high scores in the CCA solution (i.e. either high positive or high negative scores in either Axis 1 or 2 in Figure 4E) and abundant expressions in major populations (Figure 4F-4L), we selected 12 candidate genes (*CCR7*, *CCR5*, *CCR4*, *IL2RA*, *IL2RB*, *CTLA4*, *ICOS*, *TNFRSF4*, *TNFRSF9*, *FOXP3*, *BCL6*, *PDCD1*) as the candidate genes for the four processes. From these genes, we identified the most positively correlated gene to each of the 4 processes using the combinatorial CCA (i.e. test all the combinations of the variables by CCA and obtain the most correlated gene for each population; see Method). Thus, *PDCD1, FOXP3*, *CTLA4*, and *CCR7* were identified as the most correlated gene for Activated FOXP3-, Activated FOXP3+, Activated T cells, and Resting T cells, respectively (**Supplementary Figure 3**), which represent the four immunological processes (see above). Finally, using these 4 genes as explanatory variable, we applied CCA to the single cell transcriptomes, obtaining the solution of the SC4A approach.

The single cell space of the SC4A solution showed that Activated and Resting T cells had negative and positive scores, respectively (**Figure 8B**). This indicates that Axis 1 represents the T cell activation vs naïve-ness. Single cells were successfully clustered into Activated FOXP3^{+} Treg, Activated

FOXP3- non-Treg, and Resting T cells. Resting FOXP3+ Treg and Resting FOXP3- T cells were mostly overlapped (**Figure 8C**), indicating that the major features in the dataset dominated the difference between these two resting T cell groups. Importantly, the explanatory variable CTLA4, which represents the T cell activation process, was highly correlated with both Activated FOXP3^{+} Treg and Activated FOXP3- non-Treg at the middle, indicating its neutral position in terms of Tfh and Treg activation processes. As expected, the variable CCR7, which represents naïve-ness, was correlated with both Resting FOXP3+ Treg and Resting FOXP3- T cells. The explanatory variable PDCD1, which represents the Tfh-like process, was highly correlated with Activated FOXP3- non-Treg cells, while the variable FOXP3 was correlated with Activated FOXP3+ Treg. Thus, the single cell transcriptomes were modelled by the correlations between gene expression, single cells, and the expression of the 4 key genes, which represent the 4 immunological processes (**Figure 8B and 8C**). Principal Component Analysis and t-distributed stochastic neighbor embedding (t-SNE) did not provide insights into such cross-level relationships or clear separations of the populations (**Supplementary Figure 4**).

Next, in order to understand the relationship between the T cell activation signature and FOXP3-driven and Tfh-like processes (**Figure 8C**), we aimed to identify and characterise genes with high correlations to these processes, which were represented by CTLA4, FOXP3, and PDCD1, by analysing the gene space of the final output of SC4A (see Methods). As expected, the Tfh genes, *IL21* and *BCL6* (45), were highly correlated with *PDCD1*. *IL2RA* (CD25) is a Treg marker (46) and was highly correlated with *FOXP3*. *IL7R*

and *BACH2* are known to be associated with naïve T cells (47, 48), and were positively correlated with CCR7, which represents the naïve-ness (**Figure 8C**). Thus we identified *FOXP3-driven Treg genes* and *Tfh-like genes* according to their high correlation to the *FOXP3* and the *PDCD1* explanatory variables, respectively, while we designated as *Activation genes* the genes that have high correlations with the *CTLA4* variable, including *LAG3* and *CCR5*, and were positioned around 0 in Axis 2.

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

The analyses above strongly suggested that there are two major differentiation pathways for those tumour-infiltrating T cells, which are regulated by FOXP3-driven and Tfh-like processes. In order to identify these lineages, we applied an unsupervised clustering algorithm to the sample space of the SC4A/CCA result (Figure 8C), and identified 6 clusters, to which a pseudotime method (49) was applied, “lineage” curves were constructed (**Figure 8D;** see **Methods**). Importantly, the lineage curves had a bifurcation point at the Cluster 2, which leads to the two distinct differentiation pathways, Tfh-like and FOXP3-driven differentiation. Since cells may change and mature their phenotypes in different dynamics between these two lineages, we designated Tfh-like-associated and FOXP3-associated pseudotime as Tfh-pseudotime and FOXP3-pseudotime (**Figure 8D**).

In fact, the expression of Activation genes was progressively increased in the common clusters for the two pseudotime (i.e. the Clusters 1 and 2) and

throughout the rest of the FOXP3-pseudotime and the early phase of Tfh-like differentiation in Tfh-pseudotime (i.e. Cluster 3), while it was suppressed towards the end of Tfh-pseudotime **(**Cluster 4; **Figure 8E**). Given that Tfh-pseudotime is correlated with PDCD1 expression (**Figure 8C**), this suggests that *PDCD1* expression and the Tfh-effector process is induced during earlier phases, and that the activation processes in *PDCD1*^{high} T cells with are suppressed, presumably through PD1-PDL1 interactions in the tumour environment (50). Interestingly, FOXP3-driven genes had similar dynamics to Activation genes in both FOXP3- and Tfh-pseudotime (**Figure 8F**). In contrast, Tfh-like genes were mostly suppressed throughout FOXP3-pseudotime, while they were progressively induced throughout Tfh-pseudotime (**Figure 8G**). These differential regulations of two gene modules resonate with those of Tact-Foxp3 (which are expressed by both Treg and Tmem) and Tact-Runx1 genes (which are expressed specifically by Tmem only) (Figure 3). In fact, FOXP3 expression is weakly induced in some cells in the bifurcating Cluster 2, and is progressively increased at and beyond the Cluster 5 (**Figure 8H**). In contrast, RUNX1 is highly expressed in the common Clusters 1 and 2, while it is specifically suppressed in the early phase of FOXP3-pseudotime (Cluster 5, **Figure 8I**). By analysing other key genes used as CCA explanatory variables, both CTLA4 and PDCD1 were induced at the bifurcating point, Cluster 2, and onwards in both of the lineages, while PDCD1 expression was specifically suppressed througohout FOXP3-pseudotime (**Figure 8J and 8K**). CCR7 is highly expressed in the relatively naïve cells, Cluster 1, and moderately downregulated at the bifurcation point, Cluster 2,

and suppressed beyond that in both FOXP3- and Tfh-pseudotime (**Figure 8L**).

These results collectively support the model that constant activation processes in the tumour microenvironment promote terminal differentiation of the Treg- and Tfh-like lineages in both previously committed and non-committed lineages of T cells. Interestingly, the Cluster 2 is the bifurcation point, in which T cells show moderate activation and are engaged in decision-making about their cell fate. This understanding was possible because SC4A effectively annotated genes and cells and thereby allowed to identify new cell populations.

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

Lastly, we aimed to identify new combinations of marker genes in order to demonstrate the strength of the current approaches. The SC4A identified the two lineages of T cells, and their potential differentiation dynamics, FOXP3- and Tfh-pseudotime: the FOXP3-driven pathway differentiates Treg via the clusters 1-2-5-6, while the Tfh-like pathway differentiates Teff via the clusters 1-2-3-4 (Figure 9A). Since Activation genes (Figure 8C) are shared by early phases of Tfh-like and FOXP3-driven differentiation (Figure 8E), we took the intersect of these genes and the Tact-Foxp3 genes, which were expressed by both resting Tmem and resting Treg in the immgen dataset (Figure 3). Thus, we obtained *DUSP4* and *NFAT5*, which were in fact induced in cells at the activated bifurcating Cluster 2 and onwards (**Figure 9B**). Similarly, in order to

identify Treg-specific genes, we took the intersect of FOXP3-driven genes (Figure 8C) and the Tact-Foxp3 gene, and thereby obtained *CCR8* and *IL2RA*. These genes were induced highly and progressively in Treg-lineage cells throughout FOXP3-pseudotime, while suppressed across Tfh-pseudotime (**Figure 9C**). Next, in order to identify activated non-Treg (Tfh-like)-specific genes, we took the intersect of Tfh-like genes (Figure 8C) and the Tact-Runx1 genes, which are expressed in resting Tmem but suppressed in resting Treg (Figure 2). These genes contained *BCL6* and *KCNK5*, which were progressively induced across Tfh-pseudotime, while suppressed in FOXP3-pseudotime (**Figure 9D**).

Lastly, in order to make the newly obtained knowledge easily accessible to experimental biologists, we showed the expression of *NFAT5*, *IL2RA*, *CCR8*, *BCL6*, and *KCKN5* in the tumour-infiltrating T cells (**Figure 9E**). The common activation gene *NFAT5*+ in fact captured 43% of Treg-lineage cells (i.e. cells in the Clusters 5 and 6) and 52% of Tfh-like-lineage cells (i.e. cells in the Clusters 3 and 4). The Treg-specific genes *IL2RA* and *CCR8* were expressed by the majority of T cells, whether NFAT-positive or negative. In contrast, the Tfh-like-specific genes *BCL6* and *KCKN5* were expressed by a majority of Tfh-like-lineage cells and were not expressed in Treg-lineage cells (**Figure 9E**). Collectively, these results indicate that the SC4A analysis successfully decomposed the gene regulations for T cell activation and Treg and effector T cell differentiation, identifying new cell populations, which include activated cells at the bifurcation point, early and late phases of Treg and Tfh-like differentiation, and their feature genes. In addition, although there must be considerable differences between resting T cells in the secondary lymphoid

organs and between human and mice, our study successfully identified the shared activation processes and the conserved genes that are differentially used between the Treg- and the Teff-lineage cells.

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 (51). The eTreg fraction includes the $GITR^{hi}PD\text{-}1^{hi}CD25^{hi}$ "Triple-high" eTreg that express high CD5 and Nur77 expressions, which indicate that they have received strong TCR signals (17). In humans, $CD25^{hi}CD45RA^{-}FOXP3^{hi}$ eTreg highly express Ki67 (52), indicating that these cells were recently activated. Given that TCRs of Treg have higher affinities to self-antigens (53), these eTreg may have the most self-reactive TCRs during homeostasis. Alternatively, the eTreg subset may have immediately 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 4B and Figure 6B). 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 the IL-2 and common gamma chain

cytokine receptors (i.e. *IL2RA*, *IL2RB*, *IL15RA*, *IL4R*, and *IL2RG*), DNA replication licensing factor (e.g. MCM2), and transcription factors such as *PRDM1* (BLIMP1) and *IRF4* (which control the differentiation and function of eTreg (19)). These gene modules are distinct from the Tfh-like genes and the activation genes (Figure 8), and may be controlled specifically by FOXP3 under strong TCR signals. /These genes expressions/ The expression of these genes is variable /are various/ within the FOXP3+ T cells, suggesting that the transcriptional activities of these genes are dynamically regulated over time in tumour-infiltrating Treg. Thus, single cell-level analysis is becoming a key technology to address the heterogeneity of Treg. To our knowledge, this study is one of the first single cell analyses of Treg transcriptomes, while we find that, during the review process of this manuscript, another study addressing Treg heterogeneity by single cell RNA-seq was deposited at a preprint server (54)).

The shared activation genes between activated FOXP3+ Treg and FOXP3- non-Treg contain apoptosis-related genes (e.g. *CASP3*, *BAD*), which may be differentially controlled between Treg and non-Treg at the protein level. For example, activated FOXP3- non-Treg express *DUSP6* (Figure 8E), which is a negative regulator of JNK-induced apoptosis through BIM activation, while FOXP3 suppresses *DUSP6* expression and promote the apoptosis mechanism (55). In addition, the activation genes include transcription factors such as *TBX21* (T-bet) and *BATF*. Although *TBX21* is sometimes thought to be a Th1-specific gene, it is upregulated immediately after T cell activation (56). *BATF* was identified as a critical factor for the differentiation and accumulation of tissue-infiltrating Treg (57). These activation genes may be

required when T cells are activated and differentiate into either Treg or Teff. Further studies are required to investigate the temporal sequences of these differentiation events *in vivo*.

Although the effects of TCR signals on Tmem were not directly examined, considering that Tmem are self-reactive and their differentiation is dependent on the recognition of cognate antigens in the thymus (7), these results collectively suggest that the activation signature of Tmem is also dependent on TCR signals, as is the signature of Treg (Figure 2F). Intriguingly, some Treg may lose their Foxp3 expression and become ex-Treg, which are enriched in CD44^{hi} effector T cells or Tmem (30). In contrast, a Tmem population (precisely, Foxp3^{-}CD44^{hi}CD73^{hi}FR4^{hi} T cells) efficiently express Foxp3 during lymphopenia (58). These findings support the feedback control model that Foxp3 expression can be induced in Tmem and sustained in Treg as a regulatory feedback mechanism for TCR signals (7). Given the variations in the activated status in individual Treg and Tmem, single cell analysis will be required to address this problem. For example, although Samstein *et al.* showed that DNA hypersensitivity sites in Treg are similar to those in activated T cells (9), it is possible that DNA hypersensitivity sites are variable between individual Treg, and that Tmem may have a similar chromatin structure to Treg.

Importantly, our analysis showed that Tmem-specific activation-induced genes (i.e. Tact-Runx1 genes) are uniquely repressed in Treg. The repression is likely to be mediated by the interaction between Foxp3 and other transcription factors that regulate the expression of the Tmem-specific

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

While CTLA-4 is commonly recognised as a Treg marker, it is upregulated in all activated T cells, thus CTLA-4 is also a marker of activated T cells (41). CTLA-4 is in fact expressed by only a subset of resting Treg (59), which may be more activated and proliferating *in vivo* (60). In fact, our study shows that CTLA-4 is expressed by non-Treg activated T cells including resting Tmem (Figure 2D) and FOXP3- Tfh-like effector T cells in the tumour microenvironment (Figure 4G and 6C). Our SC4A analysis also identified CTLA-4 as a molecule representing the activation process of CD4+ T cells. These findings support that CTLA-4 is primarily a marker for general T cell activation, rather than Treg-specific marker, and that Treg are highly activated T cells with Foxp3 and CTLA-4 expression. In order to address this problem, the *in vivo* dynamics of CTLA-4 expression need to be investigated. We anticipate that single cell analysis will reveal the dynamics of CTLA-4 expression and T cell activation levels in resting Treg and other activated T cells *in vivo*.

In contrast, PD-1 (*PDCD1*) was specifically expressed by activated FOXP3- non-Treg cells in the tumour microenvironment. The co-expression of *BCL6* and *IL21* in some of these PD-1+ cells indicates that Tfh differentiation occurs

in the tumour microenvironment, presumably through the repeated and chronic exposure to quasi-self antigens (i.e. tumour antigens). Interestingly, the Tfh signature has been identified in type-I diabetes including mice and humans (61). Intriguingly, the Tfh-like genes include cell-cycle related genes (e.g. *CDK6*), immediate early transcription factors (*NFATC1*, *EGR2/3*), and RNA-processing genes (e.g. *DICER1*). The significance of these gene modules should be addressed in the future studies. It should be emphasised, however, that PD-1 in the tumour microenvironment may constitute immunoregulatory mechanisms as well, which prevent effective tumour immunity (50). Further experimental investigations are required to address this problem.

SC4A is a useful method to identify distinct clusters of T cells and the correlated genes to each cluster, and thereby to reveal characteristic cell groups and their active gene modules, while retaining the single-cell level variations. We also showed that SC4A and CCA results can be further analysed by the pseudotime approach. Since SC4A/CCA provides functional annotations to cell groups and gene clusters, the understanding of the pseudotime axis is effective, as shown in the current study. However, it is emphasised that pseudotime is not a measurement of the time-dependent events, but rather is that of similarities between samples (62), and the conclusions in this study require future studies with a new experimental system to analyse time-dependent events in vivo. In order to make the current SC4A/CCA approach accessible, we visualised single cell data using a flow-cytometry style (Figure 9). Although currently reliable antibodies are not available for those intracellular candidate genes, and the expression of protein

and transcripts may not be synchronized, the current study showed that the power of single cell RNA-seq and the current SC4A/CCA approach. The current limitation of SC4A is that it is computationally expensive (i.e. requires several hours for each analysis using a standard desktop), and the improvement of the computational algorithm using a low-level language will be beneficial. Importantly, SC4A is most effective when used together with in-depth knowledge of immunology and gene regulation, which will facilitate the interpretation of CCA results and explanatory variable selection. Thus, it is hoped that these tools will be used widely by experimental immunologists with good understandings of the biological significance of results, as well as adequate competence in computational analysis, which will enable to ask questions involving multidimensional problems such as multiple T cell subsets.

Data and code availability

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

Acknowledgement

We thank Dr David Bending for valuable comments on the manuscript. M.O. is a David Phillips Fellow (BB/J013951/2) from the Biotechnology and Biological Sciences Research Council (BBSRC), and is also supported by a pump-priming grant from Cancer Research Centre of Excellence, Imperial

College London/the Institute of Cancer Research. A.B is supported by the Japanese Ministry of Education, 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.

Kretschmer, L. Klein, T. Korn, S. Sakaguchi, and M. Schmidt-Supprian. 2014. Continuous T Cell Receptor Signals Maintain a Functional Regulatory T Cell Pool. *Immunity* 41: 722-736.

12. Feuerer, M., J. A. Hill, K. Kretschmer, H. von Boehmer, D. Mathis, and C. Benoist. 2010. Genomic definition of multiple ex vivo regulatory T cell subphenotypes. *Proceedings of the National Academy of Sciences of the United States of America* 107: 5919-5924.

13. Koch, M. A., G. Tucker-Heard, N. R. Perdue, J. R. Killebrew, K. B. Urdahl, and D. J. Campbell. 2009. The transcription factor T-bet controls regulatory T cell homeostasis and function during type 1 inflammation. *Nature immunology* 10: 595-602.

14. Zheng, Y., A. Chaudhry, A. Kas, P. deRoos, J. M. Kim, T. T. Chu, L. Corcoran, P. Treuting, U. Klein, and A. Y. Rudensky. 2009. Regulatory T-cell suppressor program co-opts transcription factor IRF4 to control T(H)2 responses. *Nature* 458: 351-356.

15. Linterman, M. A., W. Pierson, S. K. Lee, A. Kallies, S. Kawamoto, T. F. Rayner, M. Srivastava, D. P. Divekar, L. Beaton, J. J. Hogan, S. Fagarasan, A. Liston, K. G. Smith, and C. G. Vinuesa. 2011. Foxp3+ follicular regulatory T cells control the germinal center response. *Nature medicine* 17: 975-982.

16. van der Veeken, J., A. J. Gonzalez, H. Cho, A. Arvey, S. Hemmers, C. S. Leslie, and A. Y. Rudensky. 2016. Memory of Inflammation in Regulatory T Cells. *Cell* 166: 977-990.

17. Wyss, L., B. D. Stadinski, C. G. King, S. Schallenberg, N. I. McCarthy, J. Y. Lee, K. Kretschmer, L. M. Terracciano, G. Anderson, C. D. Surh, E. S. Huseby, and E. Palmer. 2016. Affinity for self antigen selects Treg cells with distinct functional properties. *Nature immunology* 17: 1093-1101.

18. Dias, S., A. D'Amico, E. Cretney, Y. Liao, J. Tellier, C. Bruggeman, F. F. Almeida, J. Leahy, G. T. Belz, G. K. Smyth, W. Shi, and S. L. Nutt. 2017. Effector Regulatory T Cell Differentiation and Immune Homeostasis Depend on the Transcription Factor Myb. *Immunity* 46: 78-91.

19. Cretney, E., A. Xin, W. Shi, M. Minnich, F. Masson, M. Miasari, G. T. Belz, G. K. Smyth, M. Busslinger, S. L. Nutt, and A. Kallies. 2011. The transcription factors Blimp-1 and IRF4 jointly control the differentiation and function of effector regulatory T cells. *Nature immunology* 12: 304-311.

20. Sakaguchi, S., M. Miyara, C. M. Costantino, and D. A. Hafler. 2010. FOXP3+ regulatory T cells in the human immune system. *Nature reviews. Immunology* 10: 490-500.

21. Fujii, H., J. Josse, M. Tanioka, Y. Miyachi, F. Husson, and M. Ono. 2016. Regulatory T Cells in Melanoma Revisited by a Computational Clustering of FOXP3+ T Cell Subpopulations. *Journal of immunology*.

22. Andersen, P., and B. Smedegaard. 2000. CD4+ T-Cell Subsets That Mediate Immunological Memory to Mycobacterium tuberculosis Infection in Mice. *Infection and Immunity* 68: 621-629.

23. Coleman, M. M., C. M. Finlay, B. Moran, J. Keane, P. J. Dunne, and K. H. G. Mills. 2012. The immunoregulatory role of CD4+FoxP3+CD25− regulatory T cells in lungs of mice infected with Bordetella pertussis. *FEMS Immunology & Medical Microbiology* 64: 413-424.

24. Ono, M., J. Shimizu, Y. Miyachi, and S. Sakaguchi. 2006. Control of autoimmune myocarditis and multiorgan inflammation by glucocorticoid-induced TNF receptor family-related protein(high), Foxp3-expressing CD25+ and CD25- regulatory T cells. *Journal of immunology* 176: 4748-4756.

25. Linsley, P. S., J. L. Greene, P. Tan, J. Bradshaw, J. A. Ledbetter, C. Anasetti, and N. K. Damle. 1992. Coexpression and functional cooperation of CTLA-4 and CD28 on activated T lymphocytes. *The Journal of experimental medicine* 176: 1595-1604.

26. Attridge, K., and L. S. Walker. 2014. Homeostasis and function of regulatory T cells (Tregs) in vivo: lessons from TCR-transgenic Tregs. *Immunological reviews* 259: 23-39.

27. Bunce, C., and E. B. Bell. 1997. CD45RC Isoforms Define Two Types of CD4 Memory T Cells, One of which Depends on Persisting Antigen. *The Journal of experimental medicine* 185: 767-776.

28. Min, B., R. McHugh, G. D. Sempowski, C. Mackall, G. Foucras, and W. E. Paul. 2003. Neonates Support Lymphopenia-Induced Proliferation. *Immunity* 18: 131-140.

29. Curotto de Lafaille, M. A., A. C. Lino, N. Kutchukhidze, and J. J. Lafaille. 2004. CD25- T cells generate CD25+Foxp3+ regulatory T cells by peripheral expansion. *Journal of immunology* 173: 7259-7268.

30. Miyao, T., S. Floess, R. Setoguchi, H. Luche, H. J. Fehling, H. Waldmann, J. Huehn, and S. Hori. 2012. Plasticity of Foxp3(+) T cells reflects promiscuous Foxp3 expression in conventional T cells but not reprogramming of regulatory T cells. *Immunity* 36: 262-275.

31. Ono, M., R. J. Tanaka, M. Kano, and T. Sugiman. 2013. Visualising the Cross-Level Relationships between Pathological and Physiological Processes and Gene Expression: Analyses of Haematological Diseases. *PLoS One* 8: e53544.

32. Greenacre, M. 2008. *Correspondence Analysis in Practice.* Chapman & Hall/CRC,, London.

33. Tzeng, J., H. H.-S. Lu, and W.-H. Li. 2008. Multidimensional scaling for large genomic data sets. *BMC Bioinformatics* 9: 179.

34. Ono, M., R. J. Tanaka, and M. Kano. 2014. Visualisation of the T cell differentiation programme by Canonical Correspondence Analysis of transcriptomes. *BMC genomics* 15: 1028.

35. Tirosh, I., B. Izar, S. M. Prakadan, M. H. Wadsworth, 2nd, D. Treacy, J. J. Trombetta, A. Rotem, C. Rodman, C. Lian, G. Murphy, M. Fallahi-Sichani, K. Dutton-Regester, J. R. Lin, O. Cohen, P. Shah, D. Lu, A. S. Genshaft, T. K. Hughes, C. G. Ziegler, S. W. Kazer, A. Gaillard, K. E. Kolb, A. C. Villani, C. M. Johannessen, A. Y. Andreev, E. M. Van Allen, M. Bertagnolli, P. K. Sorger, R. J. Sullivan, K. T. Flaherty, D. T. Frederick, J. Jane-Valbuena, C. H. Yoon, O. Rozenblatt-Rosen, A. K. Shalek, A. Regev, and L. A. Garraway. 2016. Dissecting the multicellular ecosystem of metastatic melanoma by single-cell RNA-seq. *Science* 352: 189-196.

36. Wakamatsu, E., D. Mathis, and C. Benoist. 2013. Convergent and divergent effects of costimulatory molecules in conventional and regulatory CD4+ T cells. *Proceedings of the National Academy of Sciences of the United States of America* 110: 1023-1028.

37. Painter, M. W., S. Davis, R. R. Hardy, D. Mathis, C. Benoist, and C. Immunological Genome Project. 2011. Transcriptomes of the B and T lineages compared by multiplatform microarray profiling. *Journal of immunology* 186: 3047-3057.

38. Ono, M., H. Yaguchi, N. Ohkura, I. Kitabayashi, Y. Nagamura, T. Nomura, Y. Miyachi, T. Tsukada, and S. Sakaguchi. 2007. Foxp3 controls regulatory T-cell function by interacting with AML1/Runx1. *Nature* 446: 685-689.

39. Chen, W., and R. Maitra. 2017. EMCluster: EM Algorithm for Model-Based Clustering of Finite Mixture Gaussian Distribution. R Package.

40. Street, K., D. Risso, R. B. Fletcher, D. Das, J. Ngai, N. Yosef, E. Purdom, and S. Dudoit. 2017. Slingshot: Cell lineage and pseudotime inference for single-cell transcriptomics. *bioRxiv*.

41. Walker, L. S., and D. M. Sansom. 2015. Confusing signals: recent progress in CTLA-4 biology. *Trends Immunol* 36: 63-70.

42. Szabo, S. J., B. M. Sullivan, C. Stemmann, A. R. Satoskar, B. P. Sleckman, and L. H. Glimcher. 2002. Distinct Effects of T-bet in T_H1 Lineage Commitment and IFN-γ Production in CD4 and CD8 T Cells. *Science* 295: 338-342.

43. Dang, E. V., J. Barbi, H. Y. Yang, D. Jinasena, H. Yu, Y. Zheng, Z. Bordman, J. Fu, Y. Kim, H. R. Yen, W. Luo, K. Zeller, L. Shimoda, S. L. Topalian, G. L. Semenza, C. V. Dang, D. M. Pardoll, and F. Pan. 2011.

Control of T(H)17/T(reg) balance by hypoxia-inducible factor 1. *Cell* 146: 772-784.

44. Ho, I. C., M. R. Hodge, J. W. Rooney, and L. H. Glimcher. 1996. The proto-oncogene c-maf is responsible for tissue-specific expression of interleukin-4. *Cell* 85: 973-983.

45. Vinuesa, C. G., M. A. Linterman, C. C. Goodnow, and K. L. Randall. 2010. T cells and follicular dendritic cells in germinal center B-cell formation and selection. *Immunological reviews* 237: 72-89.

46. Sakaguchi, S., T. Yamaguchi, T. Nomura, and M. Ono. 2008. Regulatory T cells and immune tolerance. *Cell* 133: 775-787.

47. Seddon, B., and R. Zamoyska. 2002. TCR and IL-7 Receptor Signals Can Operate Independently or Synergize to Promote Lymphopenia-Induced Expansion of Naive T Cells. *The Journal of Immunology* 169: 3752-3759.

48. Roychoudhuri, R., K. Hirahara, K. Mousavi, D. Clever, C. A. Klebanoff, M. Bonelli, G. Sciume, H. Zare, G. Vahedi, B. Dema, Z. Yu, H. Liu, H. Takahashi, M. Rao, P. Muranski, J. G. Crompton, G. Punkosdy, D. Bedognetti, E. Wang, V. Hoffmann, J. Rivera, F. M. Marincola, A. Nakamura, V. Sartorelli, Y. Kanno, L. Gattinoni, A. Muto, K. Igarashi, J. J. O'Shea, and N. P. Restifo. 2013. BACH2 represses effector programs to stabilize T(reg)-mediated immune homeostasis. *Nature* 498: 506-510.

49. Lonnberg, T., V. Svensson, K. R. James, D. Fernandez-Ruiz, I. Sebina, R. Montandon, M. S. Soon, L. G. Fogg, A. S. Nair, U. Liligeto, M. J. Stubbington, L. H. Ly, F. O. Bagger, M. Zwiessele, N. D. Lawrence, F. Souza-Fonseca-Guimaraes, P. T. Bunn, C. R. Engwerda, W. R. Heath, O. Billker, O. Stegle, A. Haque, and S. A. Teichmann. 2017. Single-cell RNA-seq and computational analysis using temporal mixture modelling resolves Th1/Tfh fate bifurcation in malaria. *Sci Immunol* 2.

50. Okazaki, T., S. Chikuma, Y. Iwai, S. Fagarasan, and T. Honjo. 2013. A rheostat for immune responses: the unique properties of PD-1 and their advantages for clinical application. *Nature immunology* 14: 1212-1218.

51. Rosenblum, M. D., S. S. Way, and A. K. Abbas. 2016. Regulatory T cell memory. *Nature reviews. Immunology* 16: 90-101.

52. Miyara, M., Y. Yoshioka, A. Kitoh, T. Shima, K. Wing, A. Niwa, C. Parizot, C. Taflin, T. Heike, D. Valeyre, A. Mathian, T. Nakahata, T. Yamaguchi, T. Nomura, M. Ono, Z. Amoura, G. Gorochov, and S. Sakaguchi. 2009. Functional delineation and differentiation dynamics of human CD4+ T cells expressing the FoxP3 transcription factor. *Immunity* 30: 899-911.

53. Weissler, K. A., and A. J. Caton. 2014. The role of T-cell receptor recognition of peptide:MHC complexes in the formation and activity of Foxp3(+) regulatory T cells. *Immunological reviews* 259: 11-22.

54. Miragaia, R. J., T. Gomes, A. Chomka, L. Jardine, A. Riedel, A. N. Hegazy, I. Lindeman, G. Emerton, T. Krausgruber, J. Shields, M. Haniffa, F. Powrie, and S. A. Teichmann. 2017. Single cell transcriptomics of regulatory T cells reveals trajectories of tissue adaptation. *bioRxiv*.

55. Tai, X., B. Erman, A. Alag, J. Mu, M. Kimura, G. Katz, T. Guinter, T. McCaughtry, R. Etzensperger, L. Feigenbaum, D. S. Singer, and A. Singer. 2013. Foxp3 transcription factor is proapoptotic and lethal to developing regulatory T cells unless counterbalanced by cytokine survival signals. *Immunity* 38: 1116-1128.

56. Sahni, H., S. Ross, A. Barbarulo, A. Solanki, C. I. Lau, A. Furmanski, J. I. Saldana, M. Ono, M. Hubank, M. Barenco, and T. Crompton. 2015. A genome wide transcriptional model of the complex response to pre-TCR signalling during thymocyte differentiation. *Oncotarget* 6: 28646-28660.

57. Hayatsu, N., T. Miyao, M. Tachibana, R. Murakami, A. Kimura, T. Kato, E. Kawakami, T. A. Endo, R. Setoguchi, H. Watarai, T. Nishikawa, T. Yasuda, H. Yoshida, and S. Hori. 2017. Analyses of a Mutant Foxp3 Allele Reveal BATF as a Critical Transcription Factor in the Differentiation and Accumulation of Tissue Regulatory T Cells. *Immunity* 47: 268-283 e269.

58. Kalekar, L. A., S. E. Schmiel, S. L. Nandiwada, W. Y. Lam, L. O. Barsness, N. Zhang, G. L. Stritesky, D. Malhotra, K. E. Pauken, J. L. Linehan, M. G. O'Sullivan, B. T. Fife, K. A. Hogquist, M. K. Jenkins, and D. L. Mueller. 2016. CD4(+) T cell anergy prevents autoimmunity and generates regulatory T cell precursors. *Nature immunology* 17: 304-314.

59. Takahashi, T., T. Tagami, S. Yamazaki, T. Uede, J. Shimizu, N. Sakaguchi, T. W. Mak, and S. Sakaguchi. 2000. Immunologic self-tolerance maintained by CD25(+)CD4(+) regulatory T cells constitutively expressing cytotoxic T lymphocyte-associated antigen 4. *The Journal of experimental medicine* 192: 303-310.

60. Tang, A. L., J. R. Teijaro, M. N. Njau, S. S. Chandran, A. Azimzadeh, S. G. Nadler, D. M. Rothstein, and D. L. Farber. 2008. CTLA4 expression is an indicator and regulator of steady-state CD4+ FoxP3+ T cell homeostasis. *Journal of immunology* 181: 1806-1813.

61. Kenefeck, R., C. J. Wang, T. Kapadi, L. Wardzinski, K. Attridge, L. E. Clough, F. Heuts, A. Kogimtzis, S. Patel, M. Rosenthal, M. Ono, D. M. Sansom, P. Narendran, and L. S. Walker. 2015. Follicular helper T cell

signature in type 1 diabetes. *The Journal of clinical investigation* 125: 292-303.

62. Haghverdi, L., M. Buttner, F. A. Wolf, F. Buettner, and F. J. Theis. 2016. Diffusion pseudotime robustly reconstructs lineage branching. *Nat Methods* 13: 845-848.

Figure Legends

Figure 1. Identification of the activation signature in Treg and Tmem by CCA of T cell populations

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

Figure 2. Identification of the Foxp3-independent activation signature in Treg by CCA of T cell populations

The microarray dataset of peripheral CD4+ T cells (GSE15907) was analysed using the T cell activation variable and the variables for retroviral *Foxp3* transduction and *Runx1* transduction as explanatory variables. (**A**) The CCA solution was visualised by a biplot where CD4+ T cell samples are shown by closed circles (see legend) and the explanatory variables are shown by blue arrows. Percentage indicates that of the variance accounted for by the inertia

of the axes (i.e. the amount of information (eigenvalue) retained in each axis). (**B**) Gene biplot of the 2D CCA solution in (C) showing the relationships between genes (grey circles) and the explanatory variables (blue arrows). Selected key genes are annotated.

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

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

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

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

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

(**A**) The experimental design for the TCR dataset. CD44hi activated Treg and CD44lo naïve-like Treg were obtained from TCRa KO or WT mice and analysed for 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 *TCRa* 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*^{GFPCreERT2}*: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.

(**A**) Venn diagram analysis of TCR genes (DEGs between TCRa KO and WT Treg), activated Treg genes (DEGs between aTreg and rTreg), and Tact-Foxp3 and Tact-Runx1 genes (see Figure 3). (**B**) Pie chart showing the number of the genes in the intersect between the TCR genes and either Tact-Foxp3 or Tact-Runx1 genes, and that of the rest (designated as TCR-specific genes). (**C**) Pie chart showing the numbers of aTreg gene-specific genes and the intersects with Tact-Foxp3 and Tact-Runx1 genes (**D**) Pathway analysis of the gene clusters in (C). The colours of texts correspond to the ones in the Venn Diagram in (C). The gene clusters that are not shown did not have any enrichment.

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

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

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

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

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

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

(Figure 3). (E) The expression of selected genes in the tumour-infiltrating T cells were shown by a 2-dimensional plot in a flow cytometric style. Data from Treg-lineage cells (the cluster 5 and 6, upper panels) and Tfh-like lineage cells (the cluster 3 and 4, lower panels). The gene in x-axis (NFAT5) is from the activation gene group (B), while y-axis shows genes from either the FOXP3-Treg group (C) or the Tfh-like/Tmem group (D). Thresholds and quadrant gates were determined in an empirical manner using density plot.

Accession number	Short description	Reference	Description of animal models	Timing of cell harvest	Cell purification strategy and sorting markers	Tissue origin
GSE15907	T cell subsets	Immunological Genome Project; Painter *et al*., 2011	Primary cells from multiple immune lineages are isolated *ex-vivo*, primarily from young adult B6 male mice, and double-sorted to >99% purity.	6 weeks	FACS; Treg: Foxp3, CD4, CD25, GFP CD4 Tmem (scLN): TCRb CD44^{high} CD122^{lo} CD4 CD4 Tmem (spleen): CD4+ CD8- CD25- CD44^{hi} CD122^{lo} CD4 Tmem ( ) CD44^{hi}62L^{o} Activated Tmem44^{hi}62L^{lo} (spleen): CD4 TCRb CD44^{hi} CD62L^{lo} Naïve CD4 (scLN): CD4+ CD8- CD25- 62L^{hi} 44lo Naïve CD4 (mLN): CD4+ CD8- CD25- 62L^{hi} 44^{lo} Naïve CD4 (PP): TCR+ CD4+ CD44^{lo} CD62L^{hi} Naïve CD4 (spleen): CD4+ CD8- CD25- 62L^{hi} 44^{lo} CD4+ BDC (scLN): CD4+ CD8- BDC+ CD4+ BDC (pancreas): CD4+ CD8- BDC+ CD4+ BDC (pancreatic LN): CD4+ CD8- BDC+ Foxp3-Tnaive (spleen): CD4+ CD8- GFP- CD44^{lo}	Spleen, subcutaneous LNs, mesenteric LN, Peyer's patches

GSE83315	aTreg data	Van der Veeken *et al*., 2016	Diphtheria toxin (DT) was administered to Foxp3$^{GFP-DTR}$ mice to deplete Treg cells and induce expansion/activation of pro-inflammatory effector CD4+ T cells. After DT clearance, new Treg cells are generated. Mixed bone marrow chimeras generated with 90% Foxp3-GFP-DTR /10% Foxp3-GFP-CRE-ERT2XR26Y bone marrow were used to deplete 90% of Treg (thereby inducing inflammation and generation of *de novo* activated Treg) and irreversibly label 10% remaining inflammation-experienced Treg (mTreg).	Day 0 ('resting Treg), day 11 (activated Treg), day 60 (memory Treg)	FACS; CD4, Foxp3, CD44, CD62L, GFP	Spleen and peripheral lymph nodes
GSE61077	TCR KO data	Levine *et al*., 2014	8-10 week mice from TracFL x Foxp3$^{YFP-Cre}$ (tamoxifen-inducible deletion of TCR in Treg. Tamoxifen was administered on days 0 and 1.	Day 9 after tamoxifen administration	FACS TCRβ, CD4, Foxp3, CD44, CD62L	Lymph nodes
GSE42276	T cell activation	Wakamatsu *et*	Conventional CD4+ T cells	8 weeks	FACS;	Spleen, lymph node

		al., 2013	from C57BL/6J male mice were stimulated by anti-CD3 and anti-CD28 for 20h and 48h and data were pooled. 0h unstimulated samples were used as control.		DAPI−CD45R−CD8a−CD11b/c−CD4+ GFP+	
GSE6939	RV-transduced T cells	Ono *et al*., 2007	Primary CD4+ T cells were obtained from BALB/c mice and purified into CD4+ naive T cells (GITRlowCD25-CD4+). CD4+ naive T cells were activated by… and retrovirally gene transduced with Runx1 (AML1), wild type Foxp3, and empty vector as control. Mice: BALB/c, ~8 weeks old. Cells: CD25-GITRlowCD4+ cells were isolated from lymph nodes and spleen, and stimulated by mitomycin-treated Thy1(-) splenocytes and anti-CD3 antibody. On the following day, cells were transfected by retrovirus carrying either Foxp3 or Runx1. Two days after transduction, GFP+	60 hours after transfection	FACS ; GFP	Lymph nodes and spleen

			cells were sorted..			
GSE72056	Single cell analysis of tumour-infiltrating T cells	Tirosh *et al*., 2016	Single cell RNA-seq analysis of human melanoma tumour samples. Freshly resected samples were disaggregated to generate single cell suspensions of mixed cells of unknown identities. Individual viable immune (CD45+) and nonimmune (CD45–) cells (including malignant and stromal cells) were recovered from the single cell suspension by flow cytometry. Single cells were profiled by single-cell RNA-seq.	Single cells were obtained within 45min of tumour resection	FACS; CD45	
GSE15390	Human activated T cells and Treg	Beyer *et al*., 2011	Conventional CD4+ T cells were obtained from whole blood of healthy human donors. T cells were activated for 5 days with CD3+CD28-coated beads; unstimulated naïve T cells as control.	5 days	MACS/FACS; CD4, CD25, CD127	

Figure 1

A

Gene-oriented CCA of T cell populations and immunological processes

Two transcriptome data

T cell populations

Genes

```
0 0 10 7 0...
1 10 0 3 ...
...
```

Explanatory variable(s)

Genes

```
0 1..
1 3..
...
```

Main data:
T cell populations

Intersect

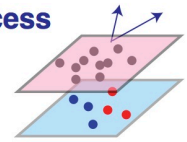
Explanatory variable:
Prototype T cell data as immunological process

CCA solution = 3 levels:

Immunological process

Gene

Cell



→ Visualise relationships between genes, cell populations, and immunological processes

B

Main data:
Peripheral CD4+ T cell populations

Intersect

Explanatory variable:
T cell activation

Solution: CCA 1D space

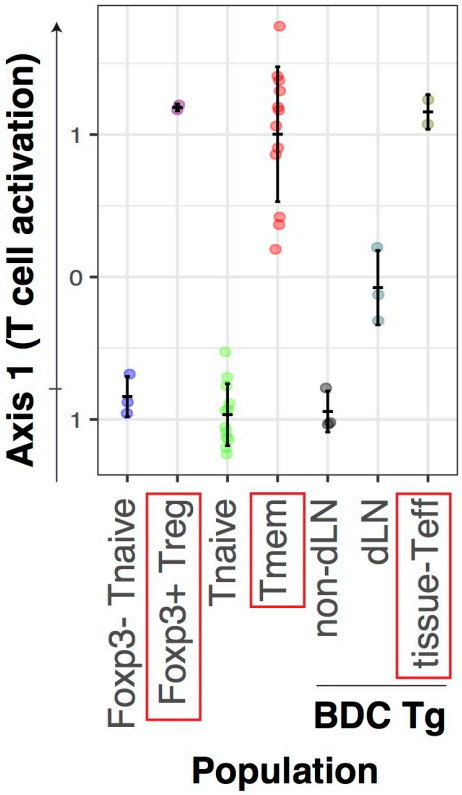


Figure 2

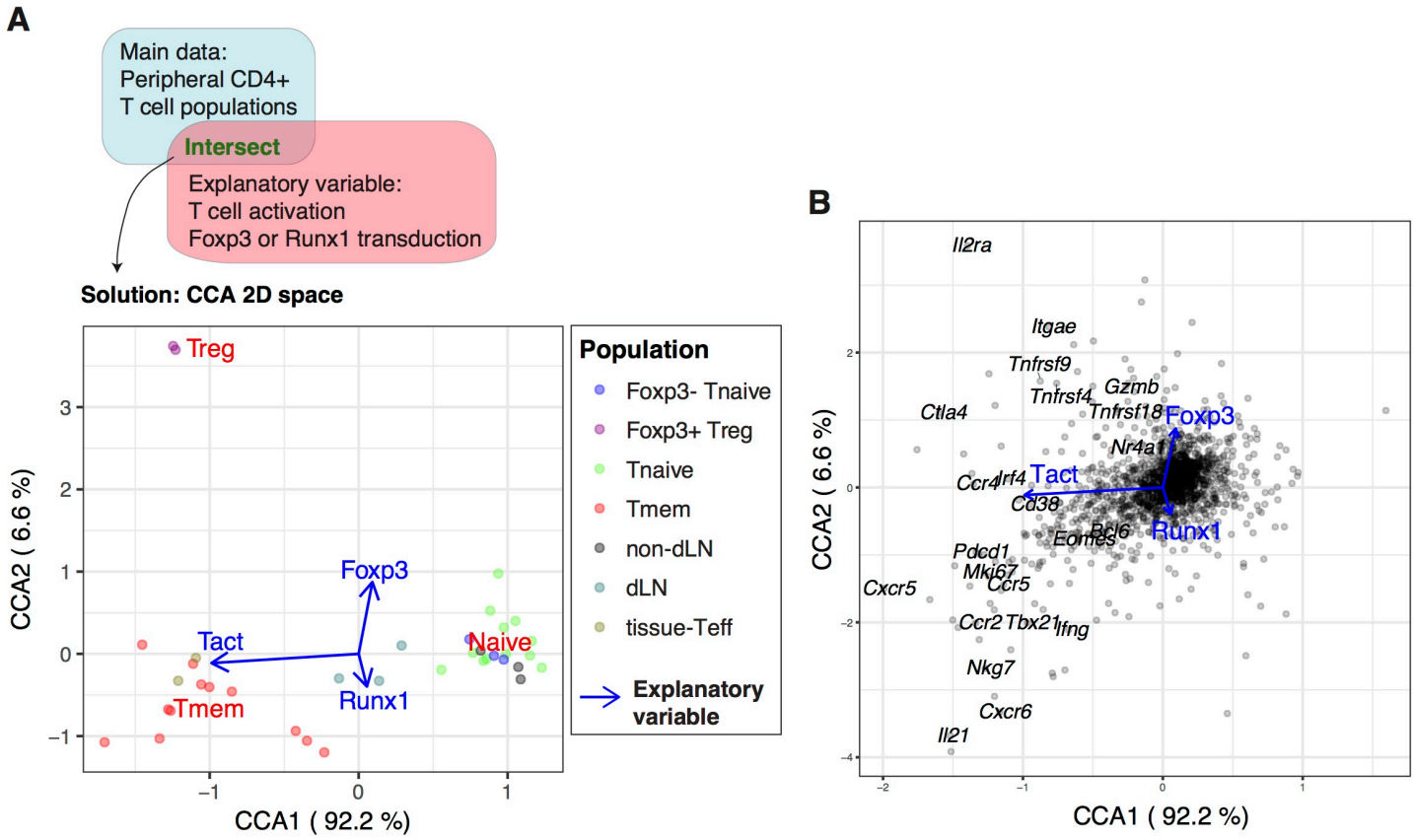

A
Main data:
Peripheral CD4+
T cell populations
Intersect
Explanatory variable:
T cell activation
Foxp3 or Runx1 transduction
Solution: CCA 2D space
Treg
Foxp3
Tact
Naive
Tmem
Runx1
CCA2 (6.6 %)
CCA1 (92.2 %)
Population
Foxp3- Tnaive
Foxp3+ Treg
Tnaive
Tmem
non-dLN
dLN
tissue-Teff
Explanatory variable
B
Il2ra
Itgae
Tnfrsf9
Gzmb
Tnfrsf4
Tnfrsf18
Ctla4
Foxp3
Nr4a1
Ccr4
Irf4
Tact
Cd38
Bcl6
Runx1
Eomes
Pdcd1
Mki67
Ccr5
Cxcr5
Ccr2
Tbx21
Ifng
Nkg7
Cxcr6
Il21
CCA2 (6.6 %)
CCA1 (92.2 %)

Figure 3

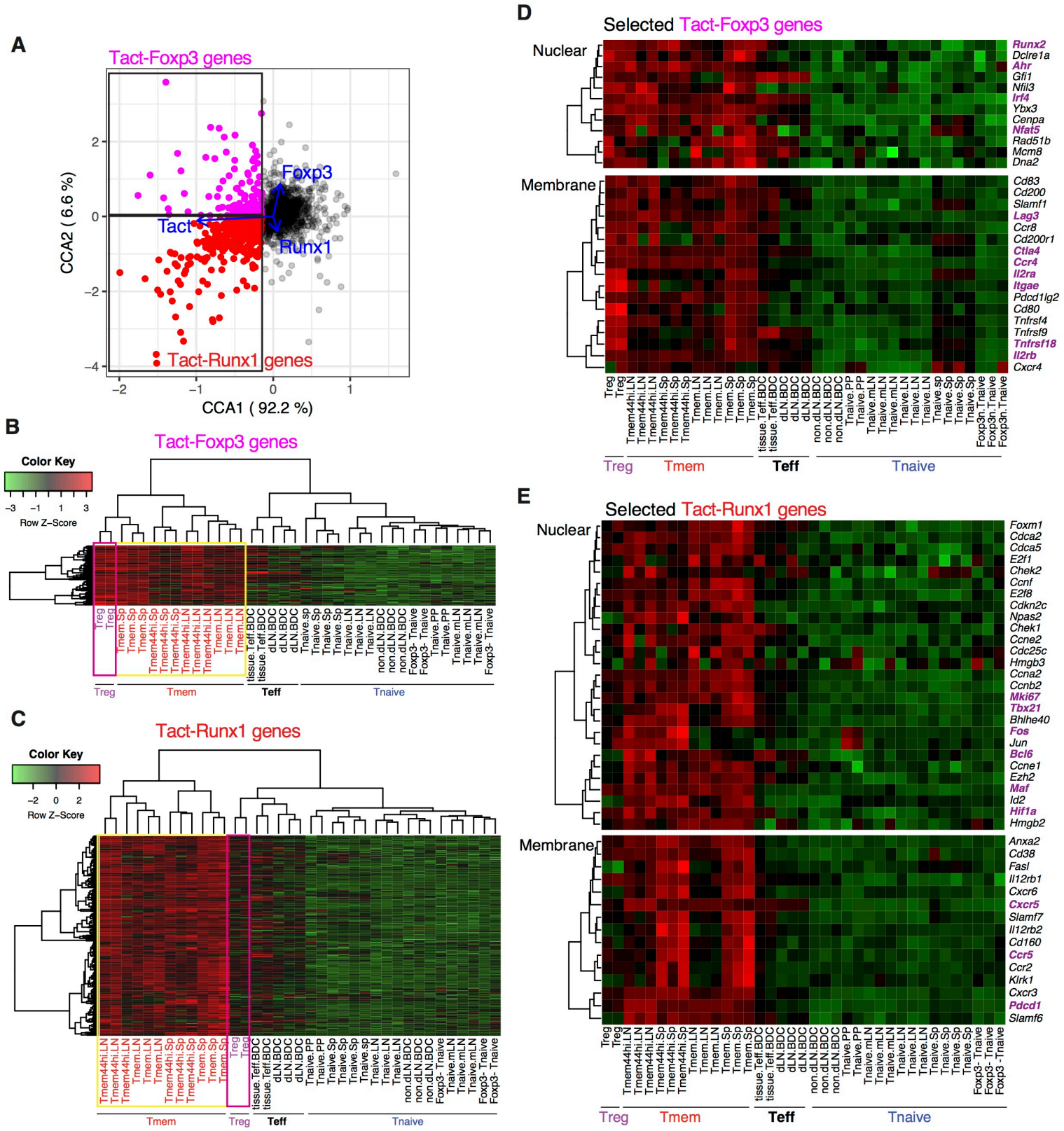

A
Tact-Foxp3 genes
Foxp3
Tact
Runx1
Tact-Runx1 genes
CCA2 (6.6 %)
CCA1 (92.2 %)
B
Tact-Foxp3 genes
Color Key
Row Z-Score
Treg
Tmem
Teff
Tnaive
C
Tact-Runx1 genes
Color Key
Row Z-Score
Tmem
Treg
Teff
Tnaive
D
Selected Tact-Foxp3 genes
Nuclear
Runx2
Dclre1a
Ahr
Gfi1
Nfil3
Irf4
Ybx3
Cenpa
Nfat5
Rad51b
Mcm8
Dna2
Membrane
Cd83
Cd200
Slamf1
Lag3
Ccr8
Cd200r1
Ctla4
Ccr4
Il2ra
Itgae
Pdcd1lg2
Cd80
Tnfrsf4
Tnfrsf9
Tnfrsf18
Il2rb
Cxcr4
Treg
Tmem
Teff
Tnaive
E
Selected Tact-Runx1 genes
Nuclear
Foxm1
Cdca2
Cdca5
E2f1
Chek2
Ccnf
E2f8
Cdkn2c
Npas2
Chek1
Ccne2
Cdc25c
Hmgb3
Ccna2
Ccnb2
Mki67
Tbx21
Bhlhe40
Fos
Jun
Bcl6
Ccne1
Ezh2
Maf
Id2
Hif1a
Hmgb2
Membrane
Anxa2
Cd38
Fasl
Il12rb1
Cxcr6
Cxcr5
Slamf7
Il12rb2
Cd160
Ccr5
Ccr2
Klrk1
Cxcr3
Pdcd1
Slamf6
Treg
Tmem
Teff
Tnaive

Figure 4

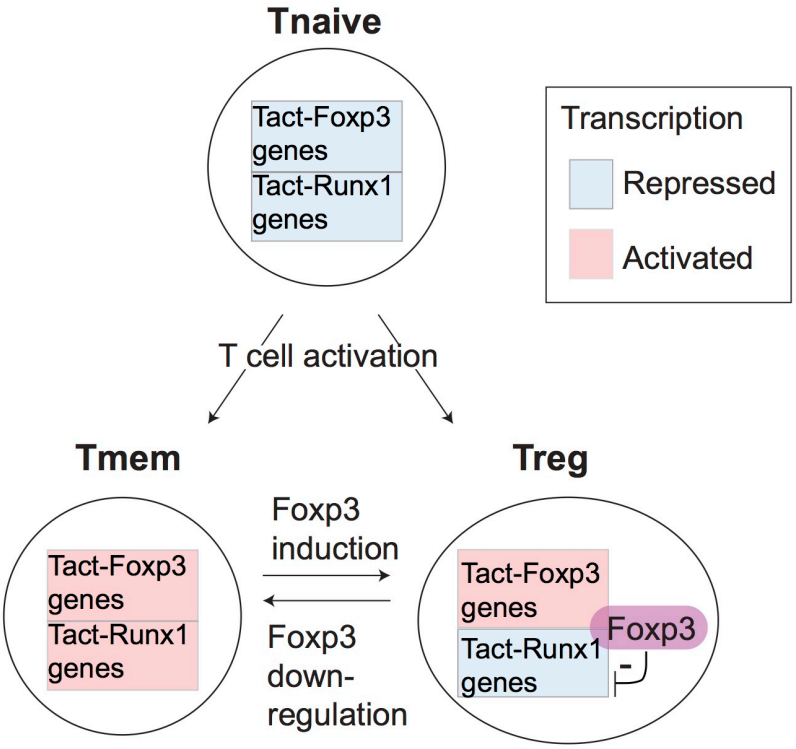

Tnaive
Tact-Foxp3 genes
Tact-Runx1 genes
Transcription
Repressed
Activated
T cell activation
Tmem
Treg
Foxp3 induction
Foxp3 down-regulation
Tact-Foxp3 genes
Tact-Runx1 genes
Tact-Foxp3 genes
Tact-Runx1 genes
Foxp3

Figure 5

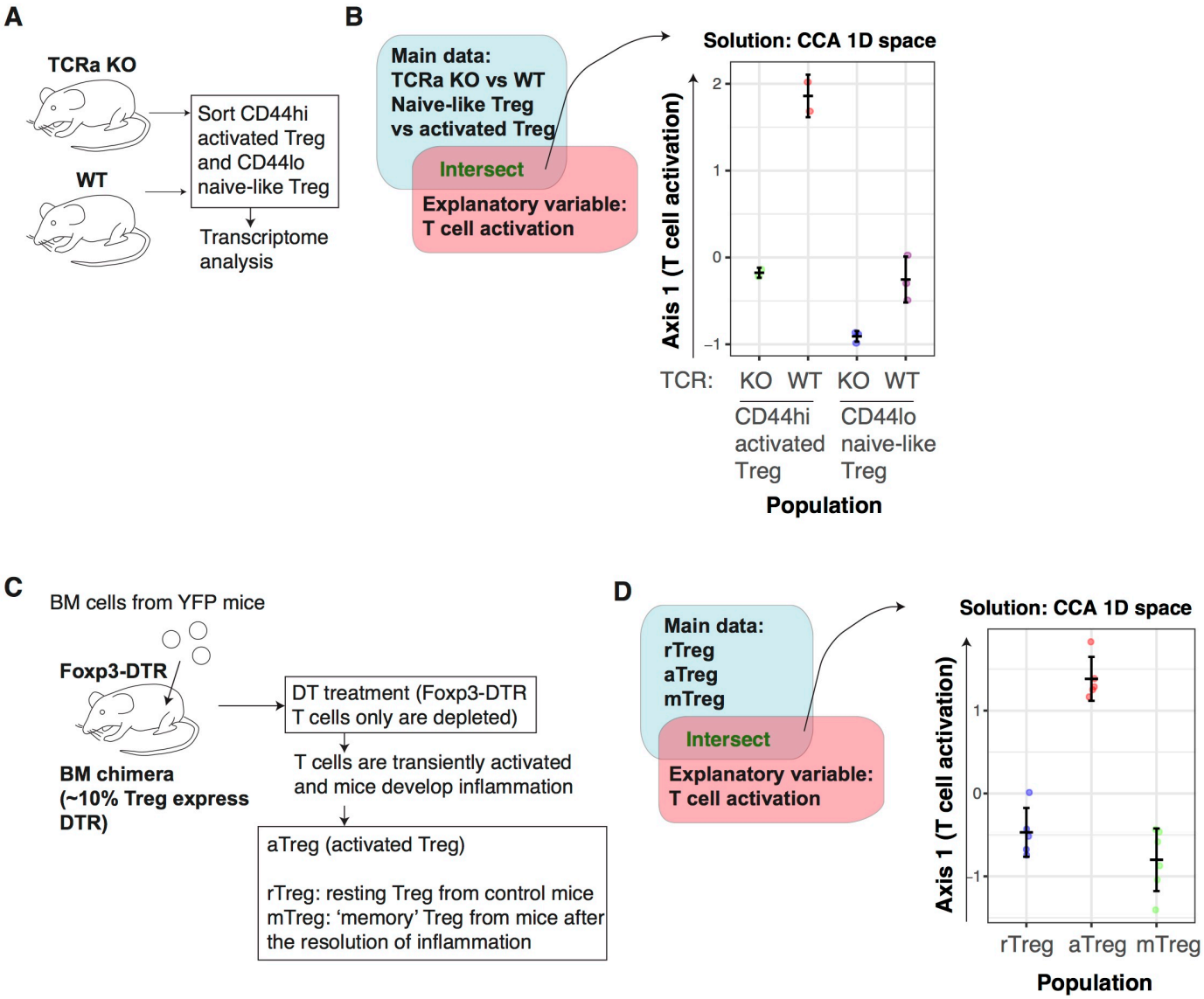

A
TCRa KO
WT
Sort CD44hi activated Treg and CD44lo naive-like Treg
Transcriptome analysis
B
Main data: TCRa KO vs WT Naive-like Treg vs activated Treg
Intersect
Explanatory variable: T cell activation
Solution: CCA 1D space
Axis 1 (T cell activation)
2
1
0
−1
TCR: KO WT KO WT
CD44hi activated Treg
CD44lo naive-like Treg
Population
C
BM cells from YFP mice
Foxp3-DTR
BM chimera (~10% Treg express DTR)
DT treatment (Foxp3-DTR T cells only are depleted)
T cells are transiently activated and mice develop inflammation
aTreg (activated Treg)
rTreg: resting Treg from control mice
mTreg: 'memory' Treg from mice after the resolution of inflammation
D
Main data: rTreg aTreg mTreg
Intersect
Explanatory variable: T cell activation
Solution: CCA 1D space
Axis 1 (T cell activation)
1
0
−1
rTreg aTreg mTreg
Population

Figure 6

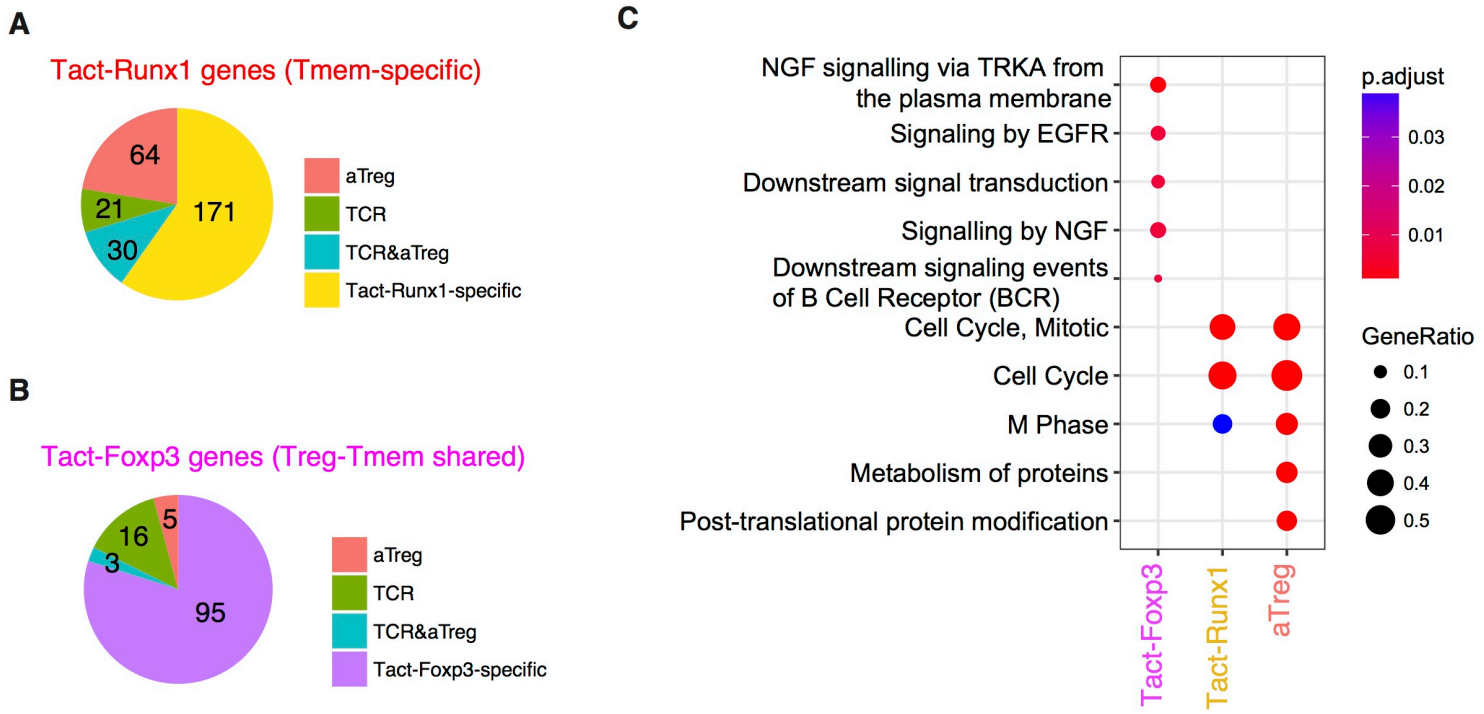
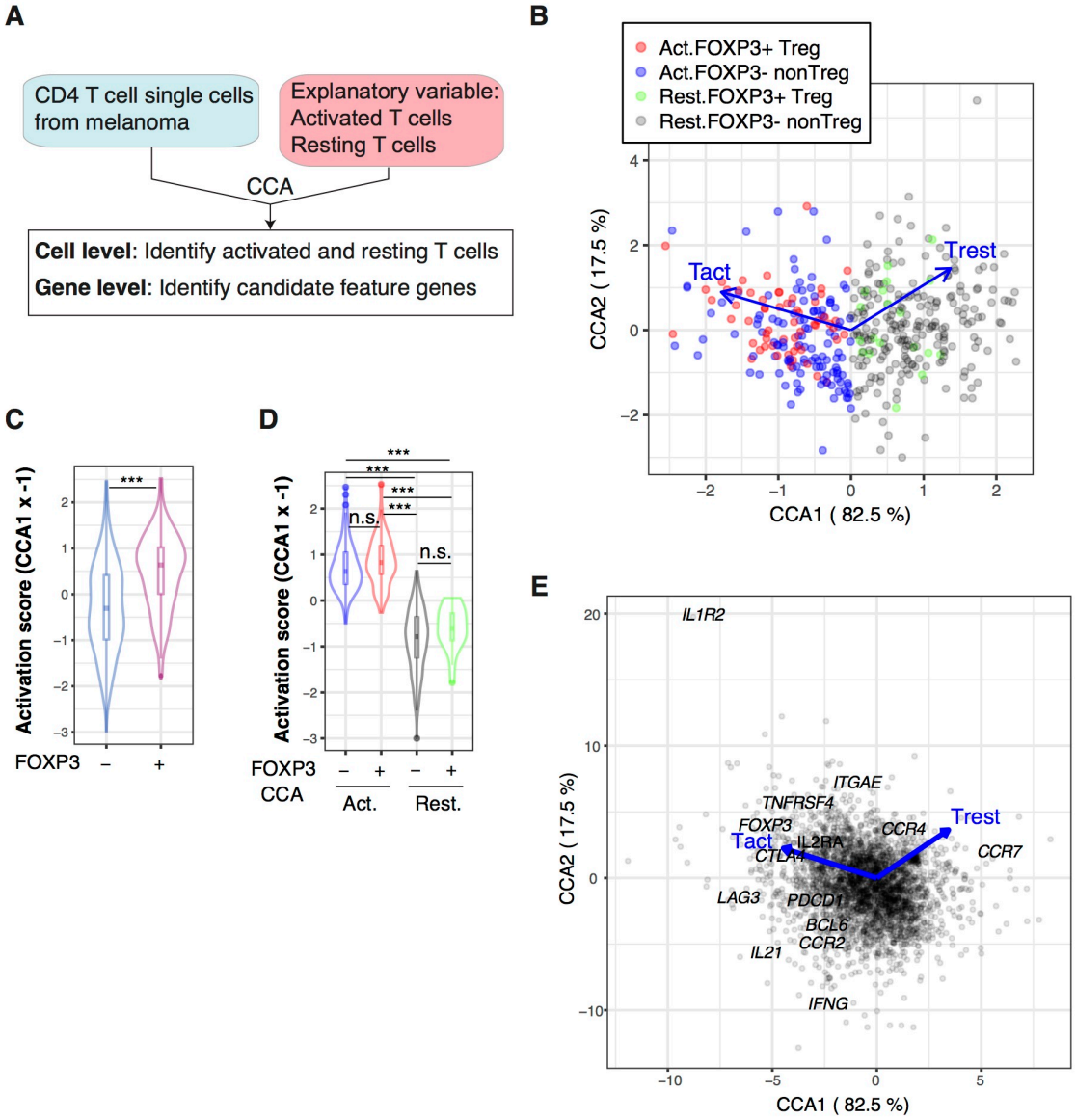

A
Tact-Runx1 genes (Tmem-specific)
64
21
30
171
aTreg
TCR
TCR&aTreg
Tact-Runx1-specific
B
Tact-Foxp3 genes (Treg-Tmem shared)
5
16
3
95
aTreg
TCR
TCR&aTreg
Tact-Foxp3-specific
C
NGF signalling via TRKA from the plasma membrane
Signaling by EGFR
Downstream signal transduction
Signalling by NGF
Downstream signaling events of B Cell Receptor (BCR)
Cell Cycle, Mitotic
Cell Cycle
M Phase
Metabolism of proteins
Post-translational protein modification
Tact-Foxp3
Tact-Runx1
aTreg
p.adjust
0.03
0.02
0.01
GeneRatio
0.1
0.2
0.3
0.4
0.5

Figure 7


A
CD4 T cell single cells from melanoma
Explanatory variable: Activated T cells Resting T cells
CCA
Cell level: Identify activated and resting T cells
Gene level: Identify candidate feature genes
B
Act.FOXP3+ Treg
Act.FOXP3- nonTreg
Rest.FOXP3+ Treg
Rest.FOXP3- nonTreg
Tact
Trest
CCA2 (17.5 %)
CCA1 (82.5 %)
C
Activation score (CCA1 x -1)
FOXP3
D
Activation score (CCA1 x -1)
FOXP3
CCA
Act.
Rest.
n.s.
E
IL1R2
ITGAE
TNFRSF4
FOXP3
IL2RA
CTLA4
CCR4
CCR7
LAG3
PDCD1
BCL6
CCR2
IL21
IFNG
Tact
Trest
CCA2 (17.5 %)
CCA1 (82.5 %)

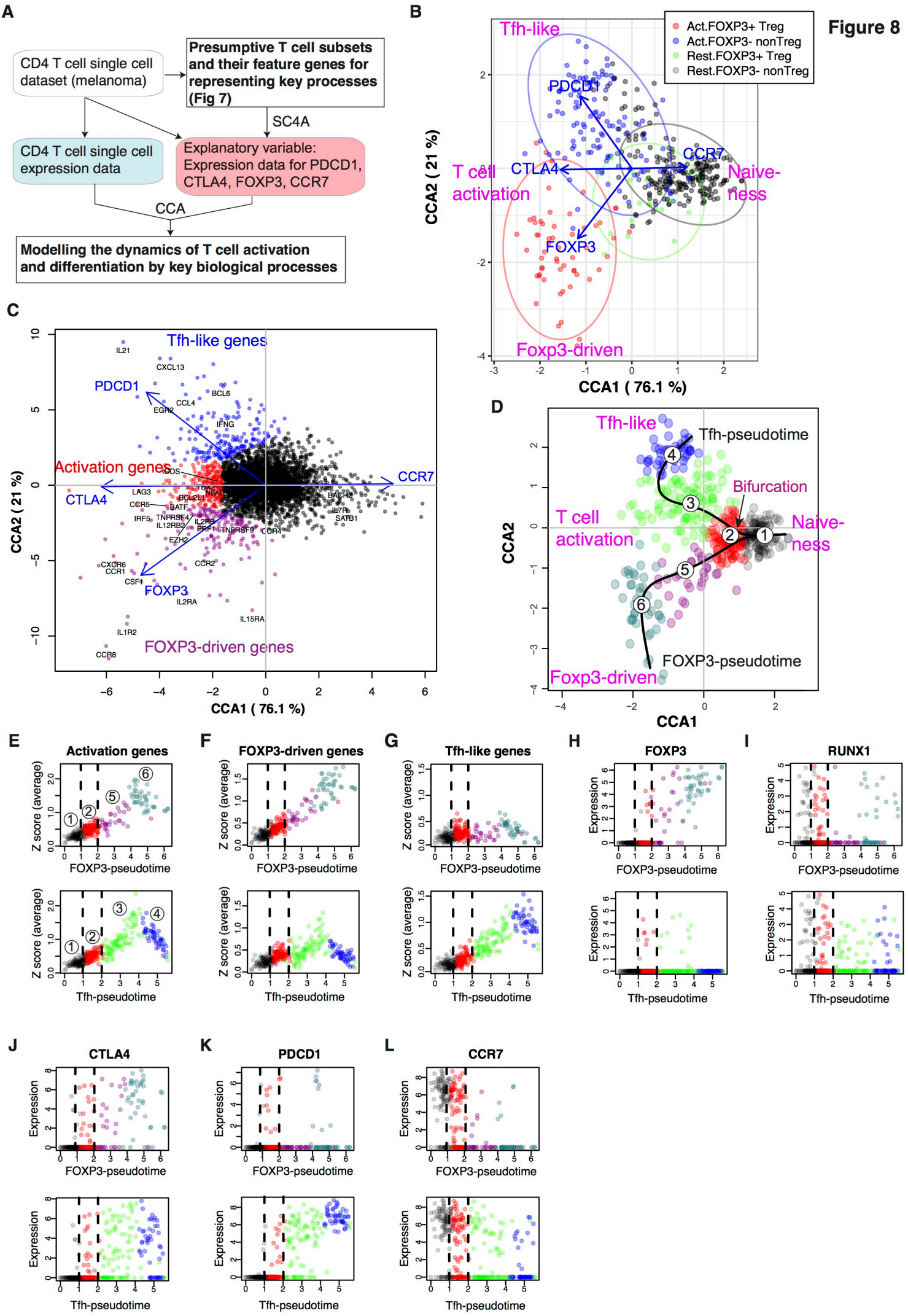
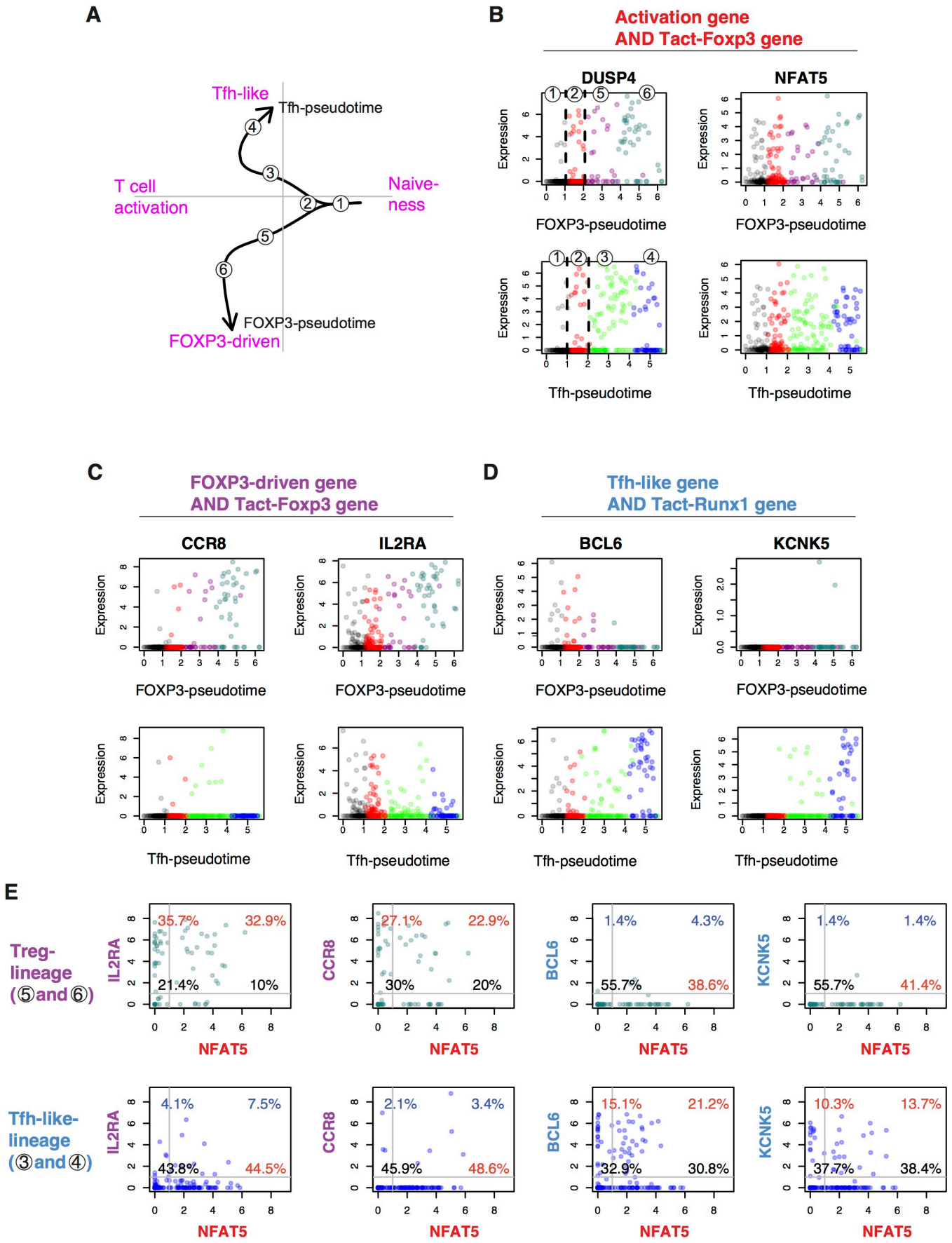


Figure 8
A
CD4 T cell single cell dataset (melanoma)
Presumptive T cell subsets and their feature genes for representing key processes (Fig 7)
SC4A
CD4 T cell single cell expression data
Explanatory variable: Expression data for PDCD1, CTLA4, FOXP3, CCR7
CCA
Modelling the dynamics of T cell activation and differentiation by key biological processes
B
Act.FOXP3+ Treg
Act.FOXP3- nonTreg
Rest.FOXP3+ Treg
Rest.FOXP3- nonTreg
Tfh-like
PDCD1
CCR7
T cell activation
CTLA4
Naive-ness
FOXP3
Foxp3-driven
CCA2 (21 %)
CCA1 (76.1 %)
C
Tfh-like genes
PDCD1
Activation genes
CCR7
CTLA4
FOXP3
FOXP3-driven genes
CCA2 (21 %)
CCA1 (76.1 %)
D
Tfh-like
Tfh-pseudotime
Bifurcation
T cell activation
Naive-ness
FOXP3-pseudotime
Foxp3-driven
CCA2
CCA1
E
Activation genes
F
FOXP3-driven genes
G
Tfh-like genes
H
FOXP3
I
RUNX1
Z score (average)
Expression
FOXP3-pseudotime
Tfh-pseudotime
J
CTLA4
K
PDCD1
L
CCR7

Figure 9


A
Tfh-like
Tfh-pseudotime
T cell activation
Naive-ness
FOXP3-pseudotime
FOXP3-driven
B
Activation gene AND Tact-Foxp3 gene
DUSP4
NFAT5
Expression
FOXP3-pseudotime
Tfh-pseudotime
C
FOXP3-driven gene AND Tact-Foxp3 gene
CCR8
IL2RA
D
Tfh-like gene AND Tact-Runx1 gene
BCL6
KCNK5
E
Treg-lineage (⑤and⑥)
Tfh-like-lineage (③and④)
IL2RA
CCR8
BCL6
KCNK5
NFAT5
35.7%
32.9%
21.4%
10%
27.1%
22.9%
30%
20%
1.4%
4.3%
55.7%
38.6%
1.4%
1.4%
55.7%
41.4%
4.1%
7.5%
43.8%
44.5%
2.1%
3.4%
45.9%
48.6%
15.1%
21.2%
32.9%
30.8%
10.3%
13.7%
37.7%
38.4%