Elucidating T Cell Activation-Dependent Mechanisms for Bifurcation of Regulatory and Effector T Cell Differentiation by Multidimensional and Single-Cell Analysis

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

[1] *Faculty of Life and Environmental Sciences, University of Tsukuba, Tsukuba, Japan,* [2] *Department of Life Sciences, Faculty of Natural Sciences, Imperial College London, London, United Kingdom*

In T cells, T cell receptor (TCR) signaling 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 signaling controls the transcriptional program 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 at the cell population level, 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 using a new CCA-based method, single-cell combinatorial CCA, we analyzed unannotated single-cell RNA-seq data from tumor-infiltrating T cells, and revealed that FOXP3 expression occurs predominantly in activated T cells. Moreover, we identified FOXP3-driven and T follicular helper-like differentiation pathways in tumor microenvironments, and their bifurcation point, which is enriched with recently activated T cells. Collectively, our study reveals the activation mechanisms downstream of TCR signals for the bifurcation of Treg and Teff differentiation and their maturation processes.

Keywords: regulatory T cells, single-cell analysis, gene expression, T cell receptor signaling, canonical correspondence analysis

INTRODUCTION

T cell receptor (TCR) signaling 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 signaling induces further T cell activation, proliferation, and differentiation (2). In addition, IL-2 signaling has key roles in immunological tolerance (2). This is partly mediated through CD25-expressing regulatory T cells (Treg), which constitutively express FoxP3, the lineage-specific transcription factor of Treg (3), and suppress

the activities of other T cells (4). Intriguingly, TCR signaling also induces the transient expression of FoxP3 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 signaling plays a critical role in Treg. Studies using TCR transgenic mice showed that Treg require TCR activation for *in vitro* suppression (8). The binding of Foxp3 protein to chromatin occurs mainly in the enhancer regions that have been opened by TCR signals (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 (eTreg) phenotype (10, 11). It is, however, unclear how TCR signals contribute to the Treg-type transcriptional program, 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.

The majority of Treg have a unique memory phenotype including CD45RB^{low}, while some of them have relatively a naïve phenotype. Previously, our theoretical study showed the potential relationship between Treg and memory-like T cells (memory-phenotype T cells; Tmem) (7), and intriguingly, the surface phenotype of Tmem is CD44^{high}CD45RB^{low}CD25$^-$ (12), which is similar to CD25$^-$ Treg, apart from Foxp3 expression and suppressive activity (13, 14). Tmem may include both antigen-experienced memory T cells (15) and self-reactive T cells (16). In fact, CD44^{high}CD45RB^{low} Tmem do not develop in TCR transgenic mice with the *Rag* deficient background, indicating that they require agonistic TCR signals in the thymus (17). In addition, a study using a fate-mapping approach showed that a minority of Treg naturally lose Foxp3 expression and join the Tmem fraction (18). These suggest that, upon encountering 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). In addition, Treg share some features with effector T cells (Teff) as well: Teff express CD25 and CTLA-4 (19), the latter of which is also known as a Treg marker (20). Thus, Treg have a close relationship with Tmem and Teff, which indicates the possibility that many known features of Treg may be in fact shared with Tmem and Teff, since the experimental evidence for these features were obtained by using naïve T cells (Tnaïve) as the control for Treg.

In order to understand these interrelated CD4$^+$ T cell subsets, the following two approaches are required. First, it is critical to understand the common and distinct features of these subsets including Treg, naïve T cells, and other non-naïve T cells, which are composed of Teff and Tmem. The analysis of transcriptomes from these subsets using multidimensional analysis will objectively disentangle the relationship between these interrelated T cell populations. Second, in order to understand the heterogeneity within each T cell population and the regulations of lineage commitment and plasticity in individual cells and across different populations, the analysis of single-cell transcriptomes is expected to provide useful insights.

Heterogeneity within the Treg population has been previously addressed through further classifying Treg into subpopulations, according to the origin [thymic Treg, peripheral Treg, visceral adipose tissue Treg (21)], the transcription factor expression and ability to control inflammation [Th1-Treg (22) and Th2-Treg (23), and T follicular regulatory T cells (24)], and their activation status [activated Treg (aTreg)/eTreg, resting Treg (rTreg), and memory-type Treg (mTreg) (25)]. 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 (25, 26), and their differentiation is controlled by the transcription factors Blimp-1, IRF4, and Myb (27, 28). Human Treg can be classified into naïve Treg (CD25$^+$ CD45RA$^+$ Foxp3$^+$) and eTreg (CD25$^+$ CD45RA$^-$ Foxp3$^+$) (29). However, such classifications are based on manual gating, which cannot fully use the power of multidimensional data, and computational clustering may be more effective for understanding flow cytometric data from heterogeneous T cells (30). Furthermore, given the recent advancement of single-cell technologies, single-cell transcriptome analysis of Treg together with non-naïve T cells (Teff and/or Tmem) is expected to reveal the dynamic changes in the activation and differentiation statuses in individual T cells and thereby provide new insights into the heterogeneity of Treg.

Multidimensional analysis is an effective approach to disentangle the relationship of closely related multiple T cell populations, allowing to systematically investigate the relationships between more than two cell populations (31). Commonly used multidimensional methods include principal component analysis (PCA), correspondence analysis (CA) (32), and multidimensional scaling (33). These methods intend to measure distances (e.g., similarities) between samples and/or genes, and thereby visualize the relationships between samples and/or genes in a reduced dimension, typically either in 2D or 3D space, 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 of CA, canonical correspondence analysis (CCA) for microarray data (31) and RNA-seq data (34), which allows to quantitatively analyze the activities of certain immunological processes in T cell subsets by analyzing the dataset that have analyzed T cells subsets (main dataset) and another dataset that represents the immunological processes of interest (explanatory variables).

In this study, we investigate the features of Treg in comparison to other CD4$^+$ T cell populations including Teff, Tmem, and naïve T cells at cell population and single-cell levels. Here, we aim to identify the differential regulation of transcriptional modules for T cell activation and differentiation in these populations by analyzing multidimensional datasets and to understand the feature of these T cell subsets at the cell population level. Furthermore, we have extended the application of CCA to single-cell analysis of unannotated cells [single-cell combinatorial CCA (SC4A)] and revealed the dynamic regulation of T cell activation-induced differentiation processes in tumor-infiltrating T cells at the single-cell level.

MATERIALS AND METHODS

Conventional CCA (Gene-Oriented Analysis)

In the application of CCA to transcriptome data, the same genes must be used in both transcriptome dataset matrices. Genes are treated as the "sites" of measurements, and the expression of each gene is considered to occur in each sample, providing gene expression as variable. In other words, gene expression is defined as the amounts of transcripts that occur at each site in the genome (i.e., gene) (rows), producing the gene expression for each sample as variable (column). Thus, explanatory variable represents the inclination to differentiation/activation process at each site in the genome, which corresponds to *environmental gradients* in the analysis of ecological data by ter Braak (34). The CCA will identify the part of main data that can be interpreted by explanatory variables (*constrained space*) using linear regression, and subsequently, the regressed data will be analyzed by CA, producing new CCA sample and gene spaces, in which the distance represents similarities between cells (and between genes). Finally, the correlation coefficients between explanatory variables and the new gene space will be determined for the visualization of the immunological processes (**Figure 1A**).

Mathematically, RNA-seq data $X \in R^{p \times m}$ is the measurement of m cell samples for the expression of p genes. The j-th column $x_j = (x_{1j}, x_{2j}, \ldots, x_{pj})^T$ is the expression data of the j-th sample with the expression of p genes, where T indicates transposed vector. Meanwhile, the explanatory data $\mathbf{Z}$ will be obtained to provide k explanatory variables, i.e., $\mathbf{Z} \in R^{p \times k}$. $\mathbf{Z}$ is scaled and standardized (i.e., mean = 0 and variance = 1). For optimizing the measurement of similarities by distance, $\mathbf{X}$ is standardized in the χ^2 metrics, using the sums of transcripts in each sample (i.e., column sums, c), the sums of transcripts in each gene (i.e., row sums, r), and the grand total of transcripts in $\mathbf{X}$ (n), the abundance matrix $\mathbf{P}$ is defined: $\mathbf{P} = 1/n\ \mathbf{X}' - \mathbf{r}\ \mathbf{c}^T$, and the standardized matrix $\mathbf{S} = \mathbf{D_r}^{-1/2}\ \mathbf{P}\ \mathbf{D_c}^{-1/2}$, where $\mathbf{D_r}$ and $\mathbf{D_c}$ are the diagonal matrices of $\mathbf{r}$ and $\mathbf{c}$ [i.e., $\mathbf{Dr} = diag(\mathbf{r})$, $\mathbf{Dc} = diag(\mathbf{c})$]. Next, $\mathbf{S}$ is projected onto $\mathbf{Z}$ by linear regression while weighting by the sums of transcripts in each gene, using the projection matrix $\mathbf{Q} = \mathbf{D_r}^{1/2}\ \mathbf{Z}\left(\mathbf{Z}^T\ \mathbf{D_r}\ \mathbf{Z}\right)^{-1}\ \mathbf{Z}^T\ \mathbf{D_r}^{1/2}$, and thus, the constrained space $\mathbf{S}^\star = \mathbf{Q}\ \mathbf{S}$ (35). Next, CCA performs singular value decomposition (SVD) of $\mathbf{S}^\star = \mathbf{U}\ \mathbf{D_\alpha}\ \mathbf{V}^T$, where $\mathbf{U}^T\ \mathbf{U} = \mathbf{V}^T\ \mathbf{V} = \mathbf{I}$, and $\mathbf{D_\alpha}$ is the diagonal matrix of singular values in descending order ($\alpha_1 \geq \alpha_2 \geq \ldots$). Note that eigenvalue $\lambda_j = \alpha_j^2$ ($j = 1, \ldots, J$), and J is the minimum value of p-1, m-1, and k. Thus, CCA analyses the constrained space and provides new axes where the dispersion of samples and that of genes are maximized. Sample (cell) scores are defined as $\mathbf{D_r}^{-1/2}\ \mathbf{V}\ \mathbf{D_\alpha}$ or $\mathbf{D_r}^{1/2}\ \mathbf{V}$ (when explanatory variable is only one). Because $\mathbf{U}$ becomes a linear combination of explanatory variables (36), weighted average scores (WA scores) G_{wa} (37) are used as gene score and are obtained by projecting the abundance matrix $\mathbf{P}$ onto sample scores while weighting by row sums (i.e., transcript abundance): $G_{wa} = \mathbf{D_r}^{-1}\ \mathbf{P}\ \mathbf{V}\ \mathbf{D_\alpha}^{-1}$ (or $\mathbf{D_r}^{-1}\ \mathbf{P}\ \mathbf{V}$, when explanatory variable is only one).

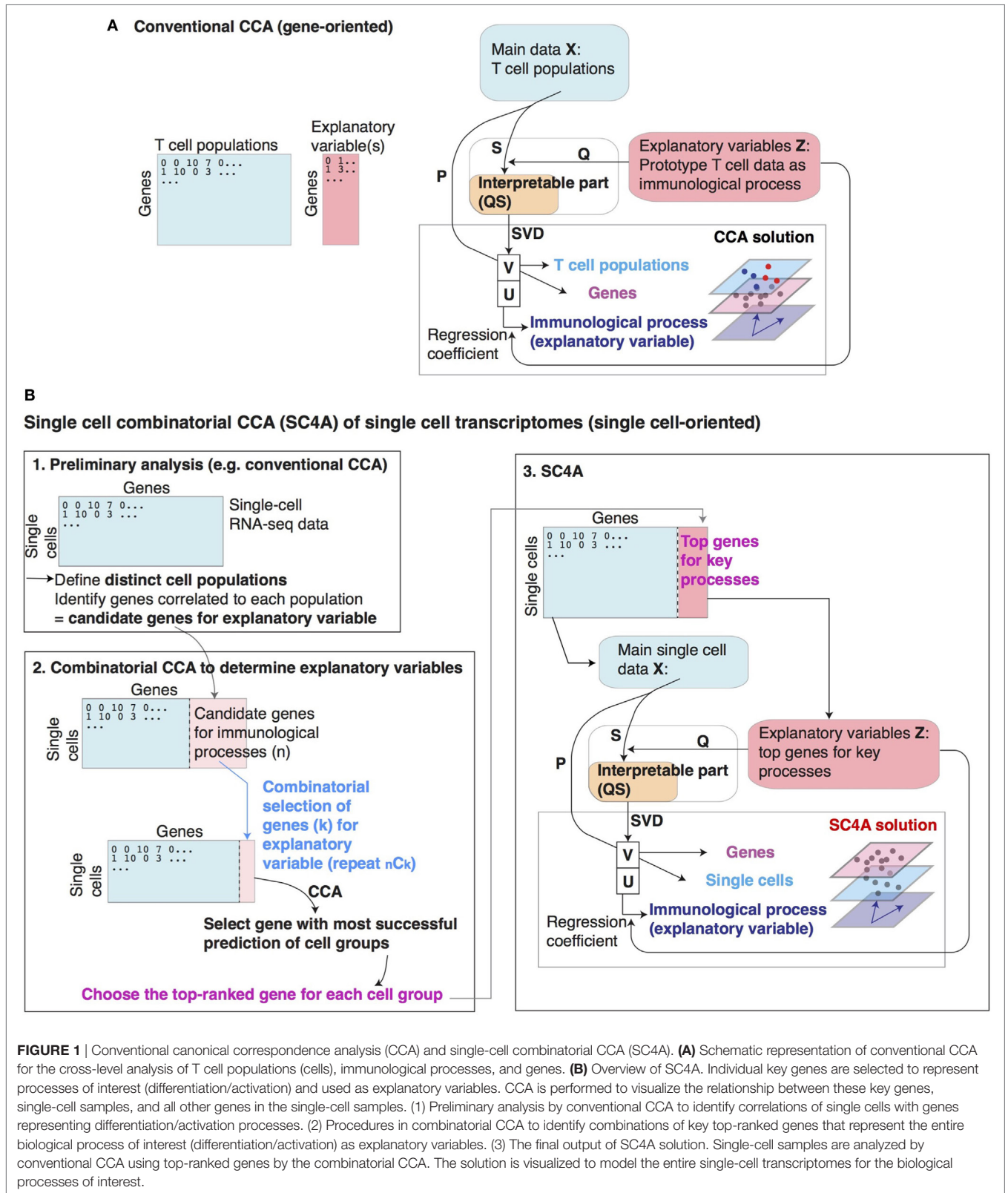In order to visualize the relationship between explanatory variables and T cells in the CCA solution, the explanatory variables $\mathbf{Z}$ will be linearly regressed to each axis of $\mathbf{U}$. Suppose $\mathbf{Z}' = [d_1, \ldots, d_k]$ is weighted explanatory variables by the transcript abundance (i.e., $\mathbf{D_r}$), and the n-th column of $\mathbf{Z}'$ is a vector with the length p, $d_n = (d_{1n}, d_{2n}, \ldots, d_{pn})^T$, the biplot values will be obtained by the correlation coefficient $\rho = cov(d_n, u_j)/\sigma(d_n)\sigma(u_j)$ ($n = 1, \ldots, k$; $j = b$), where $b = min(p$-1, m-1, $k)$. In the 1D CCA solution, the single biplot value can either be +1 or −1, indicating the direction (increasing/decreasing) of correlated genes in the explanatory variable against that in the main dataset. In order to use the 1D solution as a scoring system, the CCA score (i.e., Axis 1 score) is multiplied by the single biplot value, which indicates positive or negative correlation to Axis 1, ensuring that cells and genes with high scores have high-positive correlations to the explanatory variable.

The map approach enables the comparison of two or more explanatory variables, while the regression process in CCA allows the analysis across two different experiments (34). Explanatory variables (biplot values) 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. Since CCA is based on linear operations only, the interpretation of CCA solution is relatively straightforward using CCA biplot, which allows to directly compare the cell and gene spaces (34).

Importantly, the CCA in this article is different from a more commonly used multivariate method, canonical correlation analysis, which aims to identify shared correlation structures by maximizing the correlations between the two datasets using cross-covariance matrices, and thereby map samples from two datasets in the same sample space (38).

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 (*Tact*) was defined by the difference in gene expression between anti-CD3/CD28-stimulated (17 h) CD4$^+$ T cells and untreated naïve CD4$^+$ T cells from GSE42276 (39). Precisely, genes were selected by FDR < 0.01 and log2 fold change (> 1 or < −1) in the comparison of the gene expression profile of the activated and resting T cells. For the 1D CCA of T cell populations, the expression data of GSE15907 (40) were regressed onto gene values in *Tact* representing the change in gene expression following T cell activation, and CA was performed for the regressed data and the correlation analysis was done between the new axis and the explanatory variable. For the 2D CCA of T cell populations the expression data of GSE15907 (40) were regressed onto *Tact*, in combination with *Foxp3* and *Runx1* explanatory variables representing the effects of *Foxp3* and *Runx1* expression on CD4$^+$ T cells [GSE6939; (41)]. *Foxp3* explanatory variable is the log2 fold change of *Foxp3*-transduced naïve CD4$^+$ T cells and empty

FIGURE 1 | Conventional canonical correspondence analysis (CCA) and single-cell combinatorial CCA (SC4A). **(A)** Schematic representation of conventional CCA for the cross-level analysis of T cell populations (cells), immunological processes, and genes. **(B)** Overview of SC4A. Individual key genes are selected to represent processes of interest (differentiation/activation) and used as explanatory variables. CCA is performed to visualize the relationship between these key genes, single-cell samples, and all other genes in the single-cell samples. (1) Preliminary analysis by conventional CCA to identify correlations of single cells with genes representing differentiation/activation processes. (2) Procedures in combinatorial CCA to identify combinations of key top-ranked genes that represent the entire biological process of interest (differentiation/activation) as explanatory variables. (3) The final output of SC4A solution. Single-cell samples are analyzed by conventional CCA using top-ranked genes by the combinatorial CCA. The solution is visualized to model the entire single-cell transcriptomes for the biological processes of interest.

vector-transduced CD4$^+$ T cells. *Runx1* explanatory variable is the log2 fold change of *Runx1*-transduced naïve CD4$^+$ T cells and empty vector-transduced CD4$^+$ T cells. Subsequently, CA was applied to the regressed data and the correlation analysis was performed between the new axes and the explanatory variables.

Single-Cell Data Pre-Processing for CCA and SC4A

RNA-seq expression data of GSE72056 were obtained from single-cell suspension of tumor cells with unknown activation and differentiation statuses (42). 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 (30) for subsequent analysis.

Single-Cell Combinatorial CCA

Single-cell combinatorial CCA is a composite approach to understand non-annotated single-cell data by identifying distinct populations of cells and the differentiation processes that are correlated with these populations. Since a T cell population is usually identified by a single lineage specification factor, in the application of SC4A, such a factor will represent the cell population and their differentiation process (**Figure 1B**). The advantage of this approach is that SC4A uses the gene expression data of a part of the dataset analyzed, and thus the regression analysis of CCA becomes more efficient because of the absence of between-experimental variation, which is usually significant in cross-dataset analysis (34).

Single-cell combinatorial CCA is performed by the following three steps: (1) identification of putative cell populations and candidate genes for explanatory variables by conventional CCA, (2) combinatorial CCA to identify the top-ranked genes to be used as explanatory variables, and (3) the final CCA solution using the selected genes as explanatory variables.

Importantly, SC4A uses the same single cells as samples (rows), rather than genes, and analyzes the expression of genes as variables (columns), in order to cross-analyze main data and the explanatory variables (i.e., selected genes). In other words, gene expression in each single cell is cross-analyzed between two sets of genes in SC4A, while the expression of each gene is cross-analyzed between two sets of cell samples in conventional CCA.

Preliminary Analysis

The aim of the preliminary analysis is to identify the putative cell populations and candidate genes for explanatory variables, and conventional CCA is a useful method to do this, because candidate genes can be identified by their correlation to each putative cell population. Considering that the final output is most effectively understood by visualization using 2D (showing correlation between explanatory variable and either samples or genes) or 3D (showing correlation between explanatory variable, samples, and genes) plot, up to 4 cell populations will be identified, and up to 5–10 genes for each population will be identified by their correlation to the population (**Figure 1B**).

Combinatorial CCA

Here, 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.

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

Single-Cell Combinatorial CCA

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

Mathematically, SC4A is defined as follows. Single-cell RNA-seq data $\mathbf{X} \in \mathbf{R}^{p \times m}$ is the measurement of m genes from p single cells. The j-th column $x_j = (x_{1j}, x_{2j}, \ldots, x_{pj})^{\mathrm{T}}$ is the expression data of the j-th gene from p single cells, where T indicates transposed vector. In SC4A, by choosing a set of k genes for explanatory variable, $\mathbf{X}' \in \mathbf{R}^{p \times (m-k)}$ will be analyzed by CCA using $\mathbf{Z} \in \mathbf{R}^{p \times k}$ as explanatory variables. As in the algorithm for CCA, using the sums of transcripts for each gene (i.e., column sums, $\mathbf{c}$), and the sums of transcripts in each single cell (i.e., row sums, $\mathbf{r}$), and the grand total of transcripts of $\mathbf{X}$ (n), the abundance matrix $\mathbf{P} = 1/n$ $\mathbf{X}' - \mathbf{r}\,\mathbf{c}^{\mathrm{T}}$, and the standardized matrix $\mathbf{S} = \mathbf{D}_r^{-1/2}\,\mathbf{P}\,\mathbf{D}_c^{-1/2}$, where $\mathbf{D}_r$ and $\mathbf{D}_c$ are the diagonal matrices of $\mathbf{r}$ and $\mathbf{c}$, respectively. Meanwhile, explanatory data are scaled and standardized (i.e., mean = 0 and variance = 1) to obtain $\mathbf{Z}$. $\mathbf{S}$ is linearly regressed onto $\mathbf{Z}$ by the projection matrix $\mathbf{Q} = \mathbf{D}_r^{1/2}\,\mathbf{Z}\left(\mathbf{Z}^{\mathrm{T}}\,\mathbf{D}_r\,\mathbf{Z}\right)^{-1}\mathbf{Z}^{\mathrm{T}}\,\mathbf{D}_r^{1/2}$, and the constrained space $\mathbf{S}^{\star} = \mathbf{Q}\,\mathbf{S}$. Next, SVD is applied to $\mathbf{S}^{\star} = \mathbf{U}\,\mathbf{D}_\alpha\,\mathbf{V}^{\mathrm{T}}$, where $\mathbf{U}^{\mathrm{T}}\,\mathbf{U} = \mathbf{V}^{\mathrm{T}}\,\mathbf{V} = \mathbf{I}$, and $\mathbf{D}_\alpha$ is the diagonal matrix of singular values in descending order ($\alpha 1 \geq \alpha 2 \geq \ldots$). Gene scores are defined as $\mathbf{D}_r^{-1/2}\,\mathbf{V}\,\mathbf{D}_\alpha$ or $\mathbf{D}_r^{1/2}\,\mathbf{V}$. Weighted average scores (WA scores) $\mathbf{G}_{wa}$ are used as single-cell scores (see Conventional CCA (Gene-Oriented Analysis)) and are obtained by projecting $\mathbf{P}$ onto sample scores while weighting by row sums (i.e., transcript amounts in each single cell): $\mathbf{G}_{wa} = \mathbf{D}_r^{-1}\,\mathbf{P}\,\mathbf{V}\,\mathbf{D}_\alpha^{-1}$ or $\mathbf{D}_r^{-1}\,\mathbf{P}\,\mathbf{V}$. Biplot values will be obtained by calculating the correlation coefficient of weighted explanatory variables $\mathbf{Z} = [\mathbf{d}_1, \ldots, \mathbf{d}_k]$ ($\mathbf{Z}$ is weighted by $\mathbf{D}_r$) and $\mathbf{U} = (\mathbf{u}_1, \ldots, \mathbf{u}_k)$, that is, $\rho = cov(\mathbf{d}_n, \mathbf{u}_j)/\sigma(\mathbf{d}_n)\sigma(\mathbf{u}_j)(n = 1, \ldots, k; j = b)$, where $b = min(p\text{-}1, m\text{-}1, k)$.

Choice of Explanatory Variables by SC4A

Single-cell combinatorial CCA 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 tumor-infiltrating lymphocytes, the following two cell populations were analyzed by the combinatorial CCA: (1) activated T cells vs resting T cells; (2) FOXP3$^+$ cells vs FOXP3$^-$ cells; (3) BCL6$^+$ cells [as T follicular helper (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 four genes were used as explanatory variables in the final output of SC4A.

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 moderated *t*-statistics. DEG for activated CD44^{hi} and resting CD44^{lo} 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, normalized, and log-transformed using standard techniques (34).

Heatmaps were generated using 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 show kernel density plots (outside) and the median and interquartile range (inside) of the original gene expression data and were generated by the Bioconductor package *ggplot2*. The lineage curve was constructed by clustering SC4A/CCA sample scores using an expectation–maximization algorithm (43), and the nodes of these clusters were identified by constructing a minimum spanning tree using the Bioconductor package *Slingshot* (44).

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 rTreg 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 analyzed resting and activated conventional T cells ("T cell subset data" and "T cell activation data" in **Table 1**). Thus, we aimed to visualize 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 1D, and the cell sample scores of CCA (represented by Axis 1) provide "T cell activation score" (see Materials and 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 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 mice (i.e., with inflammation in the islets) had high scores (**Figure 2**). These CCA results indicate that Treg are as "activated" as Tmem and tissue-infiltrating activated Teff at the transcriptomic level.

Next, we addressed whether the highly "activated" status of Treg is dependent on Foxp3. Since Foxp3 suppresses Runx1-mediated transcriptional activities (41), 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 Materials and Methods). The CCA solution was 3D, while the first two axes explained the majority of variance (98.8%, **Figure 3A**). As expected, Tmem, tissue-infiltrating Teff, and Treg had high 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 3A**). By analyzing 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 Tfh cells, including *Cxcr5*, *Pdcd1* (PD-1) *Il21, Ifng, Tbx21* (T-bet), and *Mki67* (Ki-67) (**Figure 3B**). 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 3B**). 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 for Tmem

Next, we investigated the modules of genes that are differentially regulated between Treg and Tmem, in order to understand the multidimensional identity of Treg and Tmem transcriptomes (i.e., how these populations can be defined in comparison to all relevant populations). Specifically, we asked if 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 4A**). This suggests that Axis 2 provides
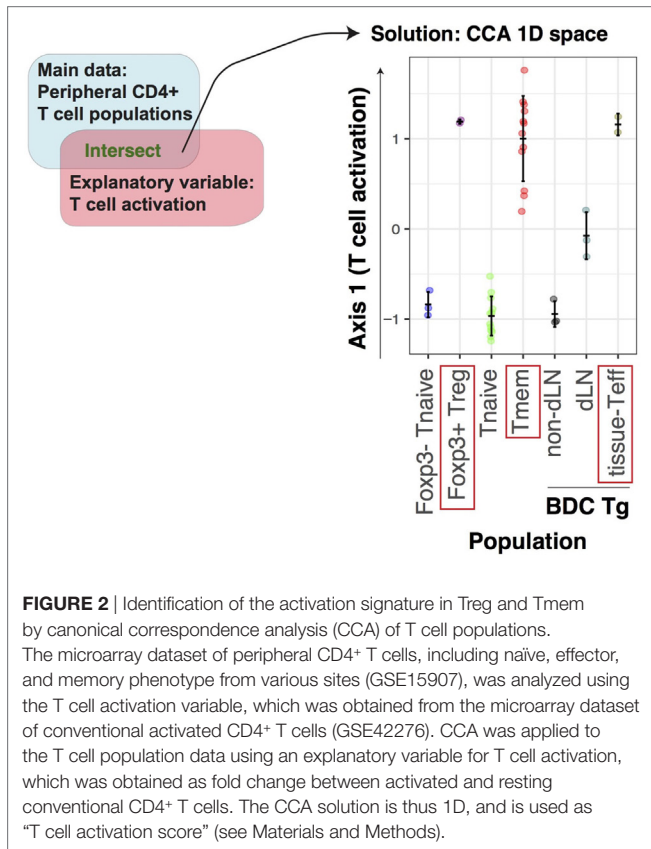
TABLE 1 | Datasets used in this study.

Accession number	Short description	Reference	Description of animal models	Timing of cell harvest	Cell purification strategy and sorting markers	Tissue origin	Figures
GSE15907	T cell subsets	Immunological Genome Project (40)	Primary cells from multiple immune lineages are isolated *ex vivo*, primarily from young adult B6 male mice (WT, Foxp3GFP, or BDC Tg mice), and double-sorted to >99% purity	6 weeks	Flow cytometric sorting Treg (spleen): Foxp3GFP$^+$ CD25$^+$ CD4$^+$ Tmem [subcutaneous (sc)LN, spleen]: TCRβ$^+$ CD44^{high} CD122^{lo} CD25$^-$ CD4$^+$ CD44^{hi}CD62L^{lo} Tmem (scLN, spleen): Foxp3GFP$^-$ TCRβ$^+$ CD44^{hi} CD62L^{lo} CD4$^+$ Naïve CD4 (scLN): CD25$^-$ CD62L^{hi} CD44^{lo} CD4$^+$ Naïve CD4 (mesenteric (m) LN): CD25$^-$ CD62L^{hi} CD44^{lo} CD4$^+$ Naïve CD4 (Peyer's patches): TCRβ$^+$ CD44^{lo} CD62L^{hi} CD4$^+$ Naïve CD4 (spleen): CD25$^-$ CD62L^{hi} CD44^{lo} CD4$^+$ Foxp3$^-$ Tnaive (spleen): Foxp3GFP$^-$ CD44^{lo} CD4$^+$ Non-draining lymph node (dLN), BDC (scLN): BDC$^+$ CD4$^+$ dLN BDC (pancreatic LN): CD4$^+$ BDC$^+$ Tissue-Teff, BDC (pancreas): BDC$^+$ CD4$^+$ *Exclusion markers include PI, CD8, CD11b, CD11c, CD19, CD49b, Gr-1, Ter119	Spleen, subcutaneous LNs, mesenteric LN, Peyer's patches, pancreatic LN, pancreas	**Figure 2** **Figure 3** **Figure 4**
GSE83315	Activated Treg (aTreg) data	(25)	Mixed bone marrow (BM) chimeras were generated with 90% *Foxp3$^{GFP-DTR}$*/10% *Foxp3-$^{GFP-CRE-ERT2}$ Rosa26YFP* BM. Diphtheria toxin (DT) was administered at day 0 to these chimeric mice in order to deplete *Foxp3$^{GFP-DTR}$* Treg cells and induce expansion/activation of effector CD4$^+$ T cells and Treg, thereby inducing inflammation. Subsequently, tamoxifen was administered at days 3 and 4 to irreversibly label *Foxp3*-expressing *Foxp3-$^{GFP-CRE-ERT2}$ Rosa26YFP* T cells with YFP. Resting Treg (rTreg), aTreg, and 'memory' Treg (mTreg) were isolated at day 0, 11, and 60, respectively, based on the dynamics of inflammation (CD4$^+$ T cell number is normalized by day 60)	Day 0 (rTreg), day 11 (aTreg), day 60 (mTreg) after DT treatment	Flow cytometric sorting rTreg: CD4, Foxp3$^-$ GFP aTreg and mTreg: CD4, Foxp3$^-$ GFP, YFP	Spleen and peripheral LN	**Figures 6C,D**
GSE61077	T cell receptor (TCR) KO data	(10)	8- to 10-week-old mice from *Trac$^{flox/WT}$* × *Foxp3$^{ERT2-Cre}$* tamoxifen-inducible deletion of TCRα in Treg. Tamoxifen was administered on days 0, 1, and 3	Day 14 after the first tamoxifen administration	Flow cytometric sorting TCRβ$^+$ CD4$^+$ Foxp3$^+$ CD44$^{high/low}$ CD62L$^{low/high}$	LN	

(Continued)

TABLE 1 | Continued

Accession number	Short description	Reference	Description of animal models	Timing of cell harvest	Cell purification strategy and sorting markers	Tissue origin	Figures
GSE42276	T cell activation	(39)	Conventional CD4+ T cells from C57BL/6J male mice were stimulated by anti-CD3 and anti-CD28 for 20 and 48 h and data were pooled 0 h unstimulated samples were used as control	8 weeks	Flow cytometric sorting DAPI− CD45R− CD8a− CD11b/c− CD4+ GFP+	Spleen, LN	**Figure 2** **Figure 3** **Figure 4** **Figure 6**
GSE6939	RV-transduced T cells	(41)	Cells: T cells from LN and spleen of 8 week-old BALB/c mice and purified into CD4+ naive T cells (GITR^low CD25−CD4+), which were subsequently activated by anti-CD3 and antigen-presenting cells [mitomycin-treated Thy1(−) splenocytes] in the presence of interleukin (IL)-2. On the following day, T cells were retrovirally gene transduced with Runx1 (AML1), wild type Foxp3, and empty vector as control	60 h after transfection	Flow cytometric sorting CD4, GFP Exclusion marker: PI	Spleen, LN	**Figure 3** **Figure 4**
GSE72056	Single-cell analysis of tumor-infiltrating T cells	(42)	Single-cell RNA-seq analysis of human melanoma tumor 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 45 min of tumor resection	Flow cytometric sorting CD45	Human melanoma tissues	**Figure 8** **Figure 9** **Figure 10**
GSE15390	Human activated and resting T cells	(45)	Resting T cells (CD25− CD4+ T cells; GSM386262, GSM386264, and GSM386266) were obtained from whole blood of healthy human donors Activated T cells (GSM777695) were prepared by stimulating CD25− CD4+ T cells for 24 h with CD3 and IL-2	Freshly sorted from buffy coat; or cultured for 24 h	Magnetic and flow cytometric sorting CD4, CD25, CD127	Human PBMC	**Figure 8**

FIGURE 2 | Identification of the activation signature in Treg and Tmem by canonical correspondence analysis (CCA) of T cell populations. The microarray dataset of peripheral CD4+ T cells, including naïve, effector, and memory phenotype from various sites (GSE15907), was analyzed using the T cell activation variable, which was obtained from the microarray dataset of conventional activated CD4+ T cells (GSE42276). 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 1D, and is used as "T cell activation score" (see Materials and Methods).

a "scoring system" for regulatory vs effector functions. Thus, the genes in Axis 1-low (precisely, genes above the 25th percentile for positive correlations with Tact) were identified as *Tact genes*. These genes 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 3B**) 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 3B**) (**Figure 4A**). Tact-Runx1 genes contain genes linked to T cell activation (e.g., *Mki67*), effector functions (e.g., *Tbx21*), and Tfh differentiation (e.g., *Bcl6*, *Pdcd1*), while Tact-Foxp3 genes contain "Treg markers" such as *Il2ra* (CD25) and *Tnfrsf18* (GITR) (**Figure 3B**).

Intriguingly, heatmap analysis showed that both Treg and Tmem expressed Tact-Foxp3 genes at high levels, compared to naïve and effector T cells (**Figure 4B**). On the other hand, Tact-Runx1 genes were selectively downregulated in Treg, while their expressions were sustained in Tmem (**Figure 4C**). In other words, the repression of Tact-Runx1 genes was the major feature of Treg in comparison to Tmem, and Tact-Foxp3 genes are the activation genes, the expression of which is induced by T cell activation in both Treg and Tmem, and is sustained or enhanced even in the presence of Foxp3. Interestingly, comparable selective downregulation of Tact-Runx1 genes was observed in Teff as well (**Figure 4C**). 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

(**Figures 4B,C**). These results collectively indicate that the Tregness is composed of the induction of the Treg–Tmem shared activation genes (i.e., Tact-Foxp3 genes) and the Foxp3-mediated repression of Tmem-specific genes (i.e., Tact-Runx1 genes), defining the multidimensional identity of Treg.
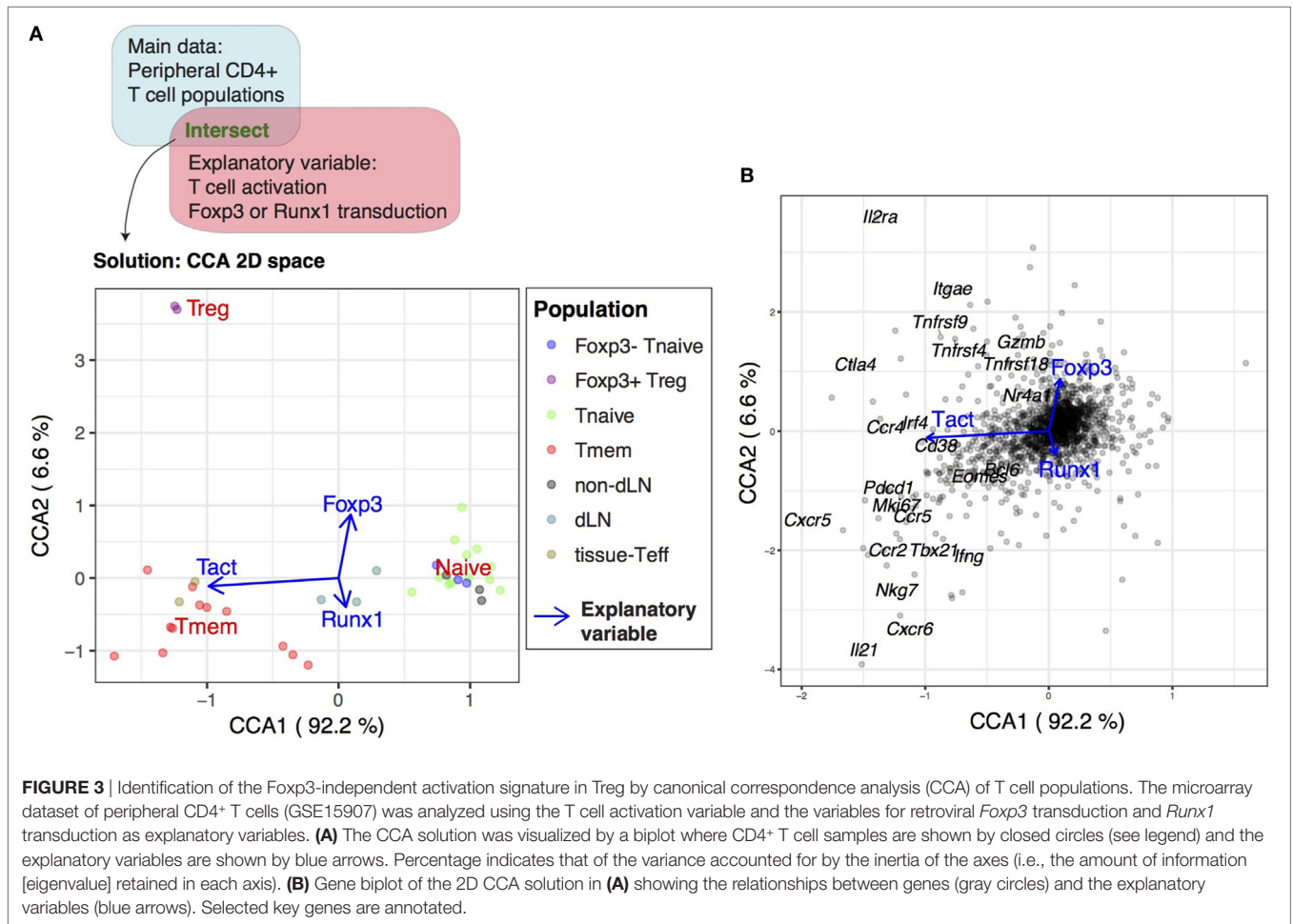
While the overall activation levels of Treg and Tmem are similar to the ones of the tissue-infiltrating Teff at transcriptional level (**Figure 2**), when explained by the prototype signature of activation in CD4+ T cells (i.e., the explanatory variable Tact), the compositions of the activation genes are different between Treg, Tmem, and Teff (**Figures 4B,C**). Importantly, many of these activation genes are shared between Treg and Tmem, but not with Teff. The closer similarity between rTreg and Tmem, compared to Teff, is not surprising, considering that both rTreg and Tmem are at the resting state, while Teff are more recently activated and executing effector functions. In addition, the distinct features of Teff may also include their capacity of tissue infiltration and the effects of the microenvironment. These features were not captured by standard *t*-test analysis (Figure S1 in Supplementary Material).

Tact-Foxp3 genes included the transcription factors *Nfat5*, *Runx2*, and *Ahr*, which were expressed by most of Tmem cells as well (**Figure 4D**). 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 4D**). 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 (46), 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., *Ccna2*, *Cdca2*, and *Chek2*), suggesting that these cells are in cell cycle and proliferating (**Figure 4E**). The higher expression of *Mki67* and *Fos* in Tmem suggests that these 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 (47–49). In accordance with this, the Tfh markers *Cxcr5* and *Pdcd1* were specifically expressed by Tmem, suggesting that Tmem are heterogeneous populations and composed of Th and Tfh cells. These results are compatible with the model that Treg and Tmem constitute the self-reactive T cell population that have a 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 5**).

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 CD44hiCD62Llo aTreg (CD44hi aTreg) and CD44loCD62Lhi naïve-like Treg (CD44lo naïve Treg) from inducible *Tcra* KO or WT mice (TCR KO data, **Table 1**; **Figure 6A**) using the T cell

FIGURE 3 | Identification of the Foxp3-independent activation signature in Treg by canonical correspondence analysis (CCA) of T cell populations. The microarray dataset of peripheral CD4+ T cells (GSE15907) was analyzed using the T cell activation variable and the variables for retroviral *Foxp3* transduction and *Runx1* transduction as explanatory variables. **(A)** The CCA solution was visualized 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 **(A)** showing the relationships between genes (gray circles) and the explanatory variables (blue arrows). Selected key genes are annotated.
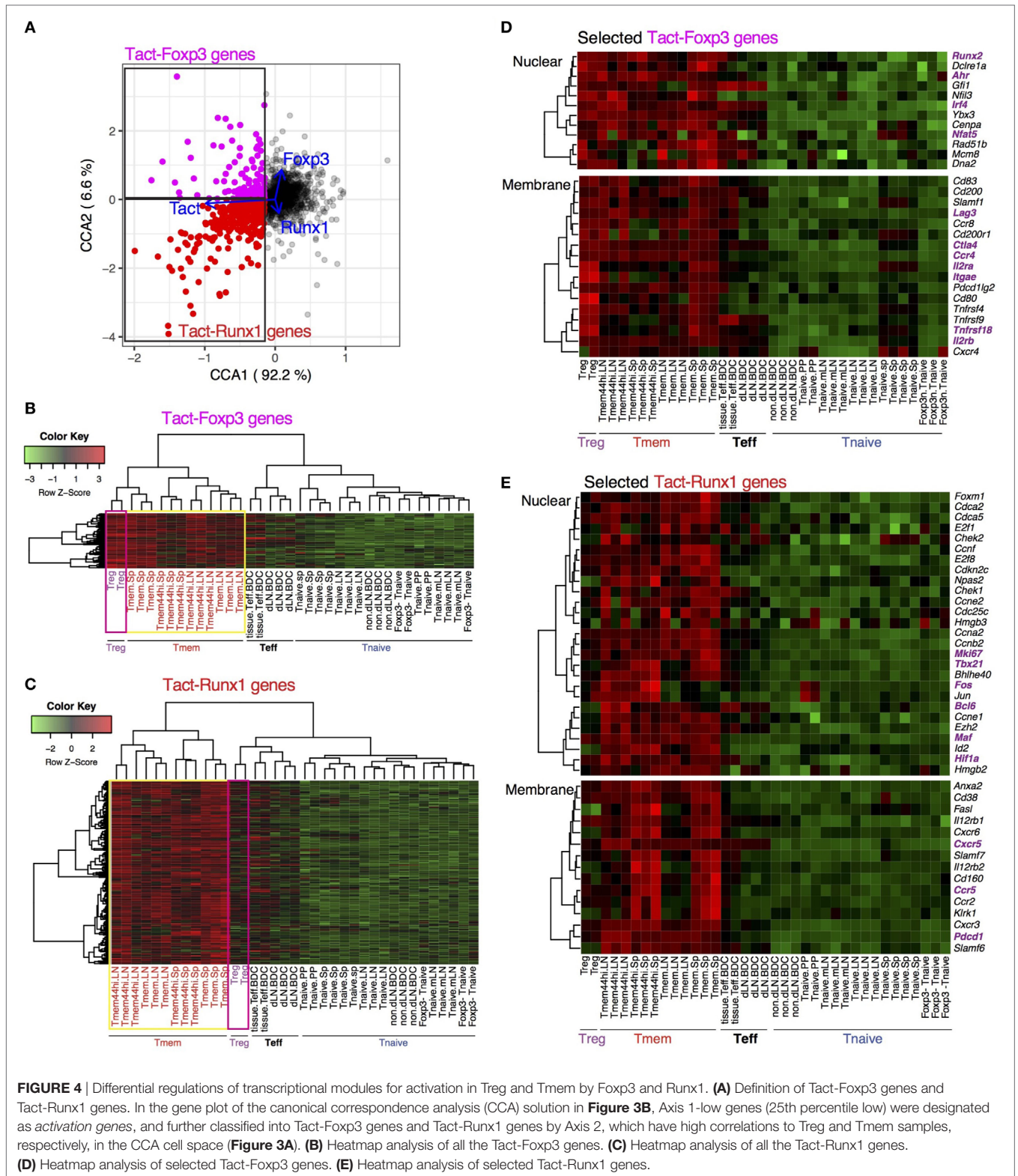
activation variable as explanatory variable. The CCA result showed that CD44^hi aTreg from WT mice only showed high activation scores, compared with all the other groups. Interestingly, *Tcra* KO CD44^lo naïve Treg showed the lowest scores, and these scores were lower than WT CD44^lo naïve Treg (**Figure 6B**). These results indicate that TCR signaling is required for the constitutive activation status of Treg, especially CD44^hi aTreg, and suggest that these aTreg are more enriched with the cells that received TCR signals recently, compared to CD44^lo naïve 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 analyzed the RNA-seq dataset of *in vivo* aTreg (25) (**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^GFPDTR* (25). The DT treatment depletes DT receptor (DTR)-expressing Treg from *Foxp3^GFP DTR*, and thus induces a transient inflammation through the reduction of Treg. *Foxp3^GFPCreERT2* allows to label Foxp3-expressing cells by YFP at the moment of tamoxifen administration. van der Veeken *et al.* thus analyzed rTreg from untreated mice (rTreg), aTreg from mice with recent depletion (11 days before the analysis) in an inflammatory condition (aTreg), and 'memory' Treg from mice
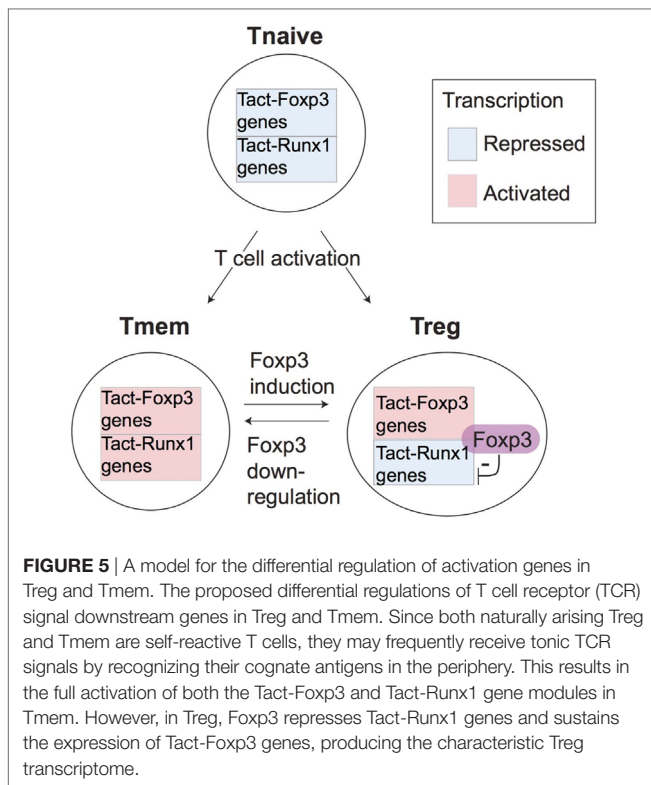
with a distant depletion (60 days before the analysis) (**Figure 6C**). As expected, the CCA analysis using the T cell activation variable showed that aTreg had higher activation scores than both rTreg and mTreg (**Figure 6D**). This indicates that the activation mechanisms are more actively operating in aTreg 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 Materials and Methods). Interestingly, 94/286 genes of Tact-Runx1 genes (Tmem-specific activation genes, repressed in rTreg) are also used during the activation of Treg (**Figure 7A**), while only 8/119 of Tact-Foxp3 genes (used by Tmem and rTreg) are induced during the activation of Treg (**Figure 7B**). This indicates that the activation of Treg does not enhance the genes that are used in rTreg, but induces the expression of the Tmem-specific genes that are suppressed in rTreg. On the other hand, 51/286 of Tact-Runx1 and 19/119 of Tact-Foxp3 genes are regulated by TCR signaling (**Figures 7A,B**), suggesting that the activation status of rTreg 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

FIGURE 4 | 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 canonical correspondence analysis (CCA) solution in **Figure 3B**, Axis 1-low genes (25th percentile low) were designated as *activation genes*, and 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 (**Figure 3A**). **(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.

contrast, Tact-Foxp3 genes were enriched for pathways related to signal transduction only (**Figure 7C**). Collectively, the results above suggest that rTreg are maintained by TCR and cytokine signaling, and that the activation of Treg induces the transcriptional activities of a part of Tact-Runx1 genes, which promote proliferation and cell division.

FIGURE 5 | A model for the differential regulation of activation genes in Treg and Tmem. The proposed differential regulations of T cell receptor (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 recognizing 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.

FOXP3 Expression More Frequently Occurs in Activated T Cells Than Resting Cells by Single-Cell CCA

The analyses above showed that Treg are on average more activated than naïve T cells and that the activation status of Treg can be variable. However, it is still unclear whether individual Treg are more activated than any naïve T cells at the single-cell level. The alternative hypothesis is that Treg are enriched with the T cells that have recognized their cognate antigens and have been activated. In order to determine this and thereby understand the dynamics of T cell regulation *in vivo*, we investigated the single-cell RNA-seq data of tumor-infiltrating T cells from human patients (42) (**Table 1**), and further enquired how the activation mechanisms are operating in Treg at the single-cell level.

First, we *in silico*-sorted FOXP3$^+$ and FOXP3$^-$ CD4$^+$ CD3$^+$ T cells from unannotated single-cell data from tumors, the tissues of which were dispersed and CD45$^+$ cells were sorted by flow cytometry without the use of any other lymphocyte markers (GSE72056, **Table 1**). Thus, the identities of individual single cells were needed to be identified in a data-oriented manner, and Treg and non-Treg cells in these tumor tissues had unknown activation and differentiation statuses. Thus, we applied CCA to the *in silico*-sorted single-cell T cell data using the explanatory variables of human activated conventional CD4$^+$ T cells (*Tact*) and resting T cells (*Trest*; GSE15390, **Table 1**), aiming to define individual single cells according to their level of activation by their correlations to these two variables (**Figure 8A**). Here, we used these two variables, *Tact* and *Trest*, in order to generate a 2D CCA solution, instead of a single explanatory variable that

represents T cell activation by the log2 fold change in gene expression between activated and resting CD4$^+$ T cells (c.f. **Figures 2** and **6**), which produces a 1D CCA solution visualized as a single axis, because we aimed to identify any additional major differentiation process(es) in the Axis 2. The explanatory variables *Tact* and *Trest* are both captured by Axis 1 because they represent two poles of one continuum—the spectrum of activation—ranging from "resting" to "activated" cell state. Thus, the CCA aimed to sort the single cells according to their individual levels of activation along the spectrum of activation, capturing the heterogeneity in activation levels in single cells. Remarkably, in the cell sample space of the CCA solution, the majority of FOXP3$^+$ T cells were positively correlated with the T cell activation variable *Tact* (**Figure 8B**), and thus had negative scores in the Axis 1 (**Figure 8B**). Here, CCA Axis 1 × (−1) score is designated as the T cell activation score. Thus, using the activation score and FOXP3 expression, the following four subpopulations were defined: "activated FOXP3$^+$," "resting FOXP3$^+$," "activated FOXP3$^-$," and "resting FOXP3$^-$" (**Figure 8B**).

Next, we aimed to determine whether individual activated FOXP3$^+$ Treg are more activated than activated FOXP3$^-$ non-Treg at the single-cell level. According to the T cell activation score established by the CCA solution in **Figure 8B**, FOXP3$^+$ Treg had significantly higher T cell activation scores than FOXP3$^-$ non-Treg on average, as indicated by the higher median in the violin plots and greater density of samples with higher T cell activation scores (**Figure 8C**), confirming the results by bulk cell analysis (**Figure 2**). Using the CCA definition of activated and rTreg and non-Treg established in **Figure 8B**, the T cell activation score neatly captured the activated status of single cells, allocating high positive and negative scores to activated and resting cells, respectively (**Figure 8D**). Importantly, there was no significant difference between activated FOXP3$^+$ and activated FOXP3$^-$ cells and between resting FOXP3$^+$ and resting FOXP3$^-$ cells (**Figure 8D**), indicating that in the tumor microenvironment, Treg cells are as activated as non-Treg CD4$^+$ T cells, which may be enriched with Teff. Strikingly, 32.5% of activated T cells expressed FOXP3, while only 8.2% of resting T cells expressed FOXP3 in **Figure 8B**. In other words, FOXP3 expression occurred more frequently in activated T cells. Given that the activation signature of Treg is dependent on TCR signals (**Figure 6**), these results suggest that FOXP3 expression occurs predominantly in the activated T cells that have recognized the tumor antigens and received TCR signals, as a negative feedback mechanism to suppress the effector response against the tumor antigens (7). Alternatively, but not exclusively, FOXP3$^+$ T cells may have high-affinity TCRs to self-MHC and/or tumor antigens and be more prone to activation (10).
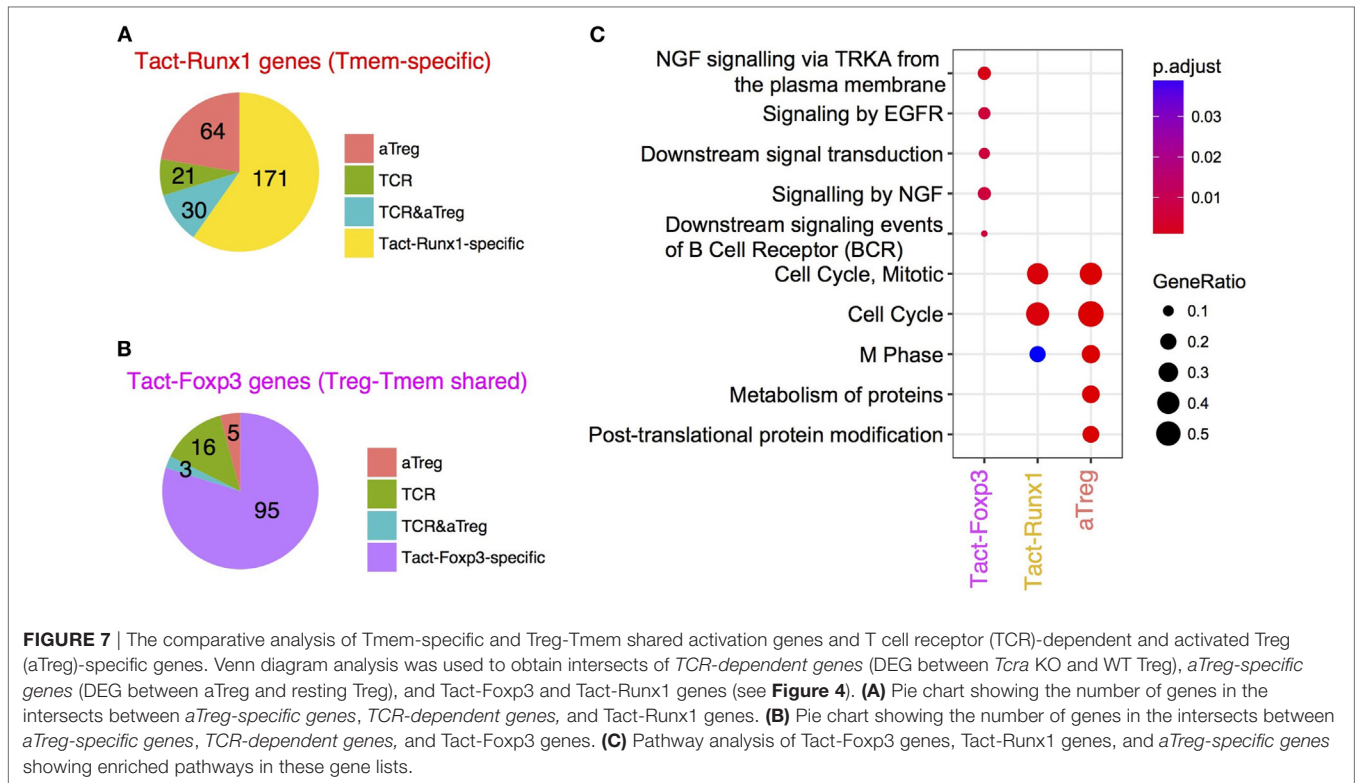
In the gene space of the CCA solution, genes with strong correlations to activated FOXP3$^+$ T cells included *FOXP3* itself and common Treg markers such as *CTLA4* and *IL2RA* (CD25), which were found in the upper left quadrant (Axis 1-negative Axis 2-positive). Interestingly, the lower left quadrant (Axis 1-negative Axis 2-negative) contained more Tfh-like or effector-like molecules *PDCD1* (PD-1), *BCL6*, *IL21*, and *IFNG*. The chemokine receptor *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 8E**).

FIGURE 6 | The activation signature of Treg is dependent on T cell receptor (TCR) signaling. **(A)** The experimental design for the TCR dataset. CD44hi activated Treg (aTreg) and CD44lo naïve Treg were obtained from *Tcra* KO or WT mice and analyzed by transcriptome analysis. **(B)** Canonical correspondence analysis (CCA) was applied to the transcriptome data of CD44loCD62hi naïve and CD44hiCD62lo aTreg 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 aTreg dataset. Bone marrow (BM) cells were obtained from *Foxp3GFPCreERT2:Rosa26YFP* mice (YFP mice), and transferred into *Foxp3GFP DTR* mice (Foxp3-DTR mice), in order to make BM chimera, in which ~10% of Treg expressed DTR. Subsequently, diphtheria toxin was administered to these BM chimeras, which depleted Foxp3-DTR cells but not donor cells. This treatment induced a transient activation of T cells and inflammation *in vivo*. 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 rTreg, *in vivo* aTreg, and mTreg from the *aTreg* data (GSE83315), with T cell activation variable as explanatory variable.

Identification of Tfh-Like Differentiation and Foxp3-Driven Processes and the Common Activation Process in Tumor-Infiltrating T Cells

Next, we aimed to identify major differentiation and activation processes in the single-cell transcriptomes above. To this end, we have developed a new CCA approach for single-cell analysis (SC4A), which aims to visualize major differentiation/activation processes and the underlying gene regulations (**Figure 9A**, see Materials and Methods; **Figure 1B**). First, we classified single cells into the four populations (activated and resting cells, and FOXP3⁺ Treg and FOXP3⁻ non-Treg; **Figure 8B**), and thereby identified the following four 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 8**). Second, 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 8E**) and abundant expressions in FOXP3⁺ and FOXP3⁻ cells (data not shown), we selected 12 genes (*CCR7, CCR5, CCR4, IL2RA, IL2RB, CTLA4, ICOS, TNFRSF4,*

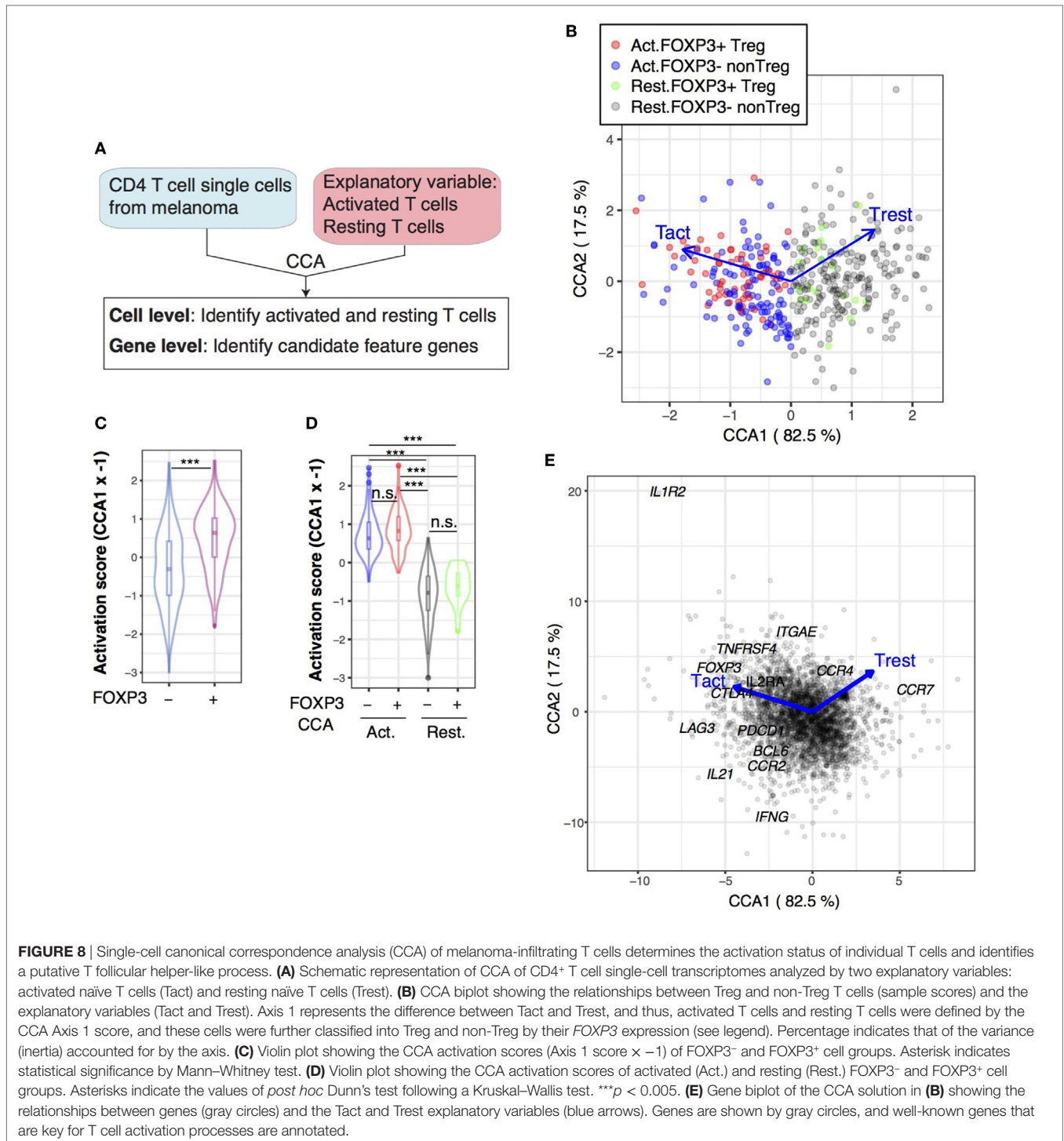**FIGURE 7** | The comparative analysis of Tmem-specific and Treg-Tmem shared activation genes and T cell receptor (TCR)-dependent and activated Treg (aTreg)-specific genes. Venn diagram analysis was used to obtain intersects of *TCR-dependent genes* (DEG between *Tcra* KO and WT Treg), *aTreg-specific genes* (DEG between aTreg and resting Treg), and Tact-Foxp3 and Tact-Runx1 genes (see **Figure 4**). **(A)** Pie chart showing the number of genes in the intersects between *aTreg-specific genes*, *TCR-dependent genes,* and Tact-Runx1 genes. **(B)** Pie chart showing the number of genes in the intersects between *aTreg-specific genes*, *TCR-dependent genes,* and Tact-Foxp3 genes. **(C)** Pathway analysis of Tact-Foxp3 genes, Tact-Runx1 genes, and *aTreg-specific genes* showing enriched pathways in these gene lists.

TNFRSF9, *FOXP3*, *BCL6*, and *PDCD1*) as the candidate genes for the four processes. From these genes, we identified the most positively correlated gene to each of the four processes using the combinatorial CCA, which tests all the combinations of variables by CCA and obtains the most correlated gene for each population (see Materials and Methods). Thus, *PDCD1, FOXP3, CTLA4*, and *CCR7* were identified as the most correlated genes for activated FOXP3⁻, activated FOXP3⁺, activated T cells, and resting T cells, respectively (Figure S2 in Supplementary Material), which represent the four immunological processes (see above). Finally, using these four genes as explanatory variables, 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 9B**). This indicates that Axis 1 represents 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 9B**), 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 modeled by the correlations between gene expression, single cells, and the expression of the four key genes, which represent the four immunological processes (**Figures 9B,C**). PCA and t-distributed stochastic neighbor embedding (t-SNE) did not provide insights into such cross-level relationships or clear separations of the populations (Figure S3 in Supplementary Material).

Next, in order to understand the relationship between the T cell activation signature and FOXP3-driven and Tfh-like processes (**Figures 9B**), we aimed to identify and characterize genes with high correlations to these processes, which were represented by CTLA4, FOXP3, and PDCD1 explanatory variables, by analyzing the gene space of the final output of SC4A (**Figure 9C**; see Materials and Methods). As expected, the Tfh genes, *IL21* and *BCL6* (50), were highly correlated with PDCD1 explanatory variable. *IL2RA* (CD25) is a Treg marker (51) and was highly correlated with FOXP3 explanatory variable. *IL7R* and *BACH2* are known to be associated with naïve T cells (52, 53) and were positively correlated with CCR7 explanatory variable, which represents the naïve-ness (**Figure 9C**). Thus, we defined *FOXP3-driven Treg genes* (magenta circles) and *Tfh-like genes* (blue circles) according to their high correlation to the FOXP3 and the PDCD1 explanatory variables, respectively, while we designated as *activation genes* (red circles) the genes that have high correlations with the CTLA4 variable, including *LAG3* and *CCR5*, which were positioned around 0 in Axis 2 (**Figure 9C**).

FIGURE 8 | Single-cell canonical correspondence analysis (CCA) of melanoma-infiltrating T cells determines the activation status of individual T cells and identifies a putative T follicular helper-like process. **(A)** Schematic representation of CCA of CD4+ T cell single-cell transcriptomes analyzed by two explanatory variables: activated naïve T cells (Tact) and resting naïve T cells (Trest). **(B)** CCA biplot showing the relationships 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 × −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 (gray circles) and the Tact and Trest explanatory variables (blue arrows). Genes are shown by gray circles, and well-known genes that are key for T cell activation processes are annotated.

Identification of the Bifurcation Point of Activated T Cells That Leads to Tfh-Like and Treg Differentiation in Tumor-Infiltrating T Cells

The analyses above strongly suggested that there are two major differentiation pathways for those tumor-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 9B**), and identified six clusters, to which a pseudotime method (54) was applied, constructing "lineage curves" (**Figure 9D**; see Materials and Methods). Importantly, the lineage curves had a bifurcation point at Cluster II, 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 9D**).

In fact, the expression of *activation genes* was progressively increased in the shared clusters (i.e., Clusters I and II) for the two

FIGURE 9 | Continued
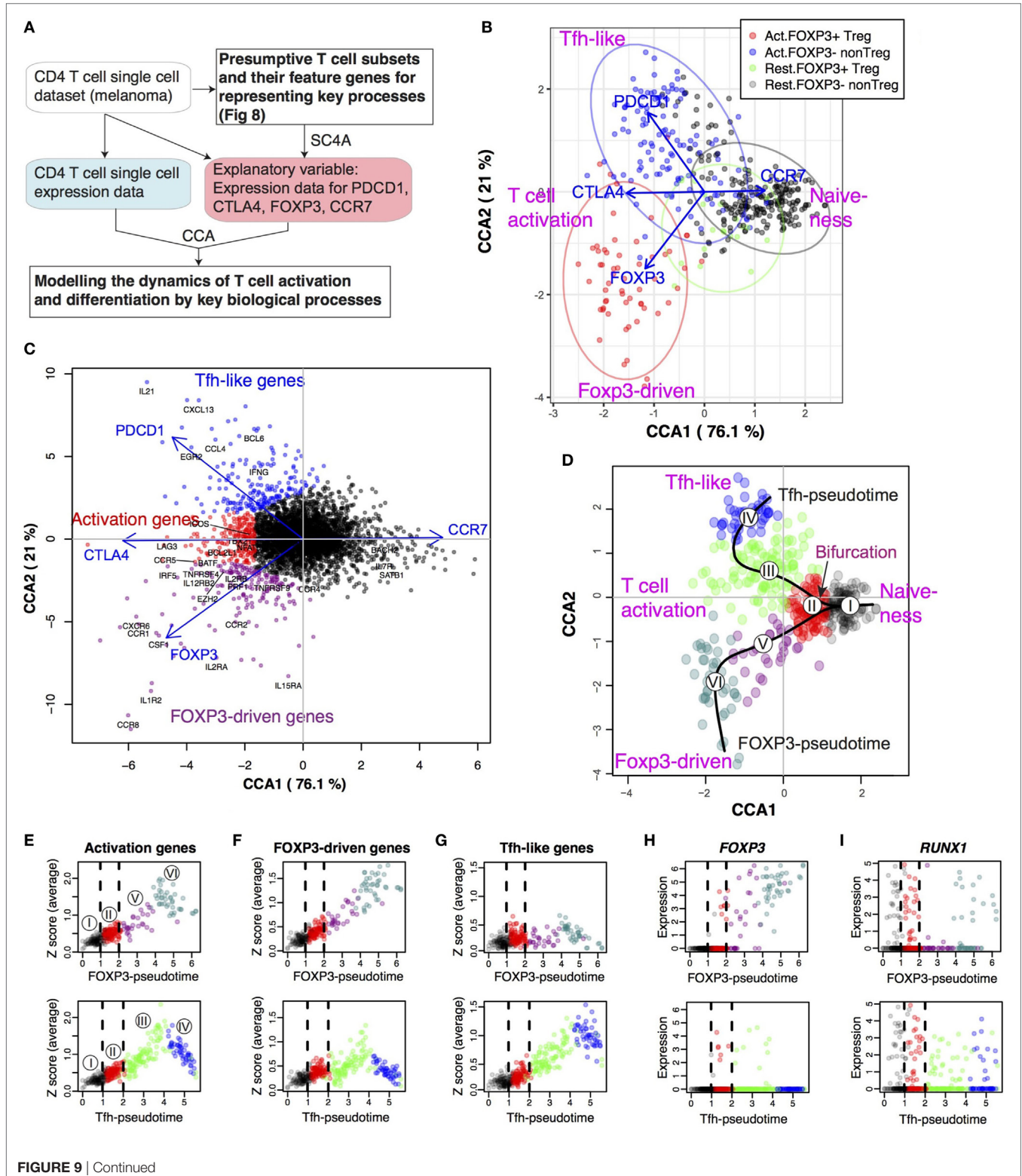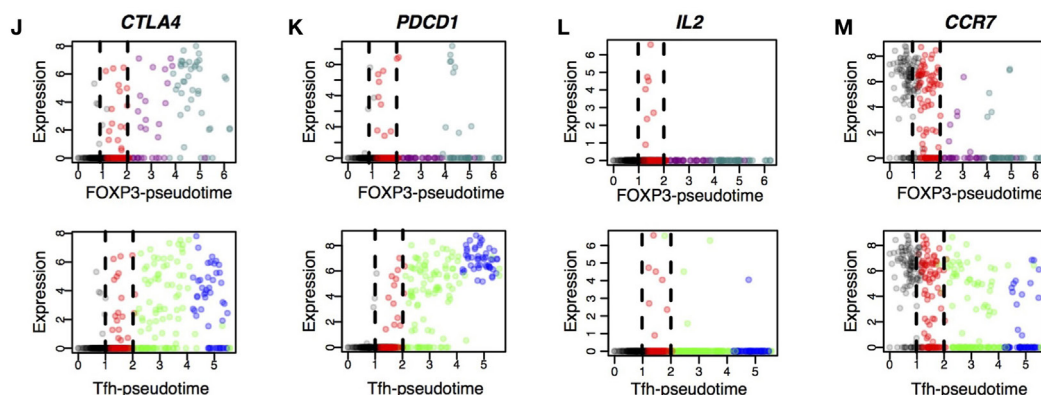
FIGURE 9 | Single-cell combinatorial CCA (SC4A) identifies the bifurcation point of activated T cells that leads to T follicular helper (Tfh)-like and Treg differentiation in tumor-infiltrating T cells. SC4A was applied to the single-cell data of tumor-infiltrating T cells, and four genes (CTLA-4, CCR7, 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 the analysis. The single-cell data from the melanoma were analyzed by SC4A to identify the most effective combinations of explanatory variables for dispersing the four T cell populations identified in **Figure 8**. These genes were used as explanatory variables to analyze the rest of the single-cell data as main dataset. Thus, the single-cell level dynamics of T cell differentiation and activation are modeled 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 six clusters by an unsupervised clustering algorithm. These clusters were further analyzed for pseudotime inference. **(E–G)** The average gene expression was plotted against each pseudotime (upper: FOXP3-pseudotime; lower: Tfh-pseudotime). The bifurcation point (Cluster II) is emphasized by broken lines. The numbers in circle indicate the cluster number. Gene expression was standardized, and the sum of the standardized expression was obtained for **(E)** Activation genes, **(F)** FOXP3-driven genes, and **(G)** Tfh-like genes **(C)**. **(H–M)** The expression of key genes was plotted against each pseudotime.

pseudotimes, and throughout the rest of the FOXP3-pseudotime and the early phase of Tfh-like differentiation (i.e., Cluster III) in Tfh-pseudotime, while it was suppressed toward the end of Tfh-like differentiation (Cluster IV; **Figure 9E**) in Tfh-pseudotime. Given that Tfh-pseudotime is correlated with *PDCD1* expression (**Figure 9C**), this suggests that *PDCD1* expression and the Tfh-effector process are induced during the earlier phases of effector T cell activity, and that the activation processes in *PDCD1*^high T cells are suppressed, presumably through PD1–PDL1 interactions in the tumor environment (55). Interestingly, *FOXP3-driven genes* had similar dynamics to *activation genes* in both FOXP3-pseudotime and Tfh-pseudotime (**Figure 9F**). By contrast, *Tfh-like genes* were mostly suppressed throughout FOXP3-pseudotime, while they were progressively induced throughout Tfh-pseudotime (**Figure 9G**). These differential regulations of two gene modules resonate with those of Tact-Foxp3 genes (which are expressed by both Treg and Tmem) and Tact-Runx1 genes (which are expressed specifically in Tmem, and repressed in Treg) (**Figure 4**). In fact, *FOXP3* expression is weakly induced in some cells in the bifurcating Cluster II and the early phase of Tfh-like differentiation (Cluster III) in Tfh-pseudotime and is progressively increased at and beyond Cluster V in FOXP3-pseudotime (**Figure 9H**).

RUNX1 is highly expressed in the common Clusters I and II and is downregulated in the transition from Clusters II to III in Tfh-pseudotime, and from Clusters II to V in FOXP3-pseudotime (**Figure 9I**), which is compatible with the known dynamics of RUNX1 expression: Runx1 is downregulated when naïve

CD4+ T cells differentiate into activated/effector cells following TCR signaling (56). By analyzing other key genes used as CCA explanatory variables, *CTLA4* was induced at the bifurcating point, Cluster II, and onward in both of the lineages at equivalent expression levels (**Figure 9J**), reflecting the activated status of both effector Tfh-type cells and Treg. Importantly, CTLA4 is a marker of Treg as well as activated effector T cells, where it acts as a negative regulator of T cell proliferation (57).

PDCD1 expression was also induced at the bifurcating point, and throughout Tfh-pseudotime, but specifically suppressed in the early phase of FOXP3-pseudotime (**Figure 9K**), which is compatible with the known dynamics that PDCD1 is transiently upregulated in activated CD4+ T cells as a negative regulatory mechanism to restrain proinflammatory immune responses and maintain peripheral tolerance (58). Further supporting this dynamic perspective, *IL2* expression occurs mainly in Cluster II, indicating that these cells are enriched with the T cells that recently recognized antigens (59) (**Figure 9L**). Consistently, the expression of the naïve T cell marker CCR7 was the highest in the cells with a relatively naïve phenotype in shared Cluster I and was moderately downregulated in the early and late phase of Tfh-pseudotime, and suppressed in most Treg in FOXP3-pseudotime (**Figure 9M**).

These results collectively support the model that constant activation processes in the tumor microenvironment promote terminal differentiation of the Treg- and Tfh-like lineages in both previously committed and non-committed lineages of T cells. Interestingly, Cluster II is the bifurcation point, in which

T cells show moderate activation and together with simultaneous expression of FOXP3 and Tfh-like genes, as well as RUNX1 and PDCD1 expression. These cells are most probably engaged in decision-making about their cell fate and the cell type-specific usage of these genes—whether their transcriptional mechanisms would be used to generate a proinflamatory or regulatory response. 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 demonstrate the utility of the current approach by discovering exemplary marker genes that distinguish cells in FOXP3⁻ and Tfh-pseudotime (i.e., the FOXP3-driven pathway I–II–V–VI, and the Tfh-like pathway I–II–III–IV) (**Figure 10A**), and to identify the T cell subpopulations by a flow cytometric visualization of single-cell data. Since *activation genes* (**Figure 9C**) are shared by early phases of Tfh-like and FOXP3-driven differentiation (**Figure 9E**), we took the intersect of these genes and the Tact-Foxp3 genes, which were expressed by both resting Tmem and rTreg in mice (**Figure 4**). *DUSP4* and *NFAT5* were such genes and were in fact induced in cells at the activated bifurcating Cluster II and onward in both lineages (**Figure 10B**). Similarly, in order to identify a marker to distinguish Treg- and Tfh-like cells, firstly, we identified *CCR8* and *IL2RA* in the intersect of FOXP3-driven genes (**Figure 9C**) and the Tact-Foxp3 genes, which were induced highly and progressively in Treg-lineage cells throughout FOXP3-pseudotime, while mostly suppressed across Tfh-pseudotime (**Figure 10C**). By contrast, *BCL6* and *KCNK5* [found in the intersect of Tfh-like genes (**Figure 9C**) and the Tact-Runx1 genes, which are expressed in resting Tmem but suppressed in rTreg (**Figure 4**)] were progressively induced across Tfh-pseudotime, while suppressed in FOXP3-pseudotime (**Figure 10D**).

Lastly, in order to make the newly obtained knowledge easily accessible to experimental immunologists, we showed the expression of *NFAT5*, *IL2RA*, *CCR8*, *BCL6*, and *KCNK5* in the tumor-infiltrating T cells in a flow cytometric format (**Figure 10E**). The common activation gene *NFAT5* in fact captured the majority of Treg-lineage cells (i.e., cells in the Clusters V and VI) and Tfh-like-lineage cells (i.e., cells in the Clusters III and IV). The expression of the Treg-specific genes *IL2RA* and *CCR8* occurred in the majority of FOXP3⁺ Treg-lineage cells, whether *NFAT5*-positive or negative, but not in most of Tfh-like-lineage cells. By contrast, the Tfh-like-specific genes *BCL6* and *KCNK5* were expressed by the majority of Tfh-like-lineage cells but were not expressed in Treg-lineage cells (**Figure 10E**).

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 humans 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. We thus identified a shared systems-level mechanism for the differential regulation of activation and differentiation processes in CD4⁺ T cell populations.

DISCUSSION

Resting Treg showed an activated status, comparable to that of Teff and Tmem at the population level (**Figure 2**), which is consistent with the previous reports that Treg receive TCR signals more frequently than other T cells (60) and their epigenomic features are similar to the ones of Tact (9). The activation signature of Treg was more remarkable in CD44^hi^CD62L^lo^ aTreg than CD44^lo^CD62L^hi^ naïve Treg. CD44^hi^CD62L^lo^ Treg are also identified as eTreg, which may have enhanced immunosuppressive activities (61). The eTreg fraction includes the GITR^hi^PD-1^hi^CD25^hi^ "Triple-high" eTreg that have high CD5 and Nur77 expressions, which indicates that they have received strong TCR signals (26). In humans, CD25^hi^CD45RA⁻FOXP3^hi^ eTreg highly express Ki67 (62), indicating that these cells were recently activated. Given that TCRs of Treg have higher affinities to self-antigens (63), these eTreg may have the most self-reactive TCRs during homeostasis. Alternatively, the eTreg subset may have recently received strong TCR signals and upregulated activation markers, and such cells may enter a resting state 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 tumor-infiltrating Treg include FOXP3⁺ T cells with various levels of activation (**Figures 8 and 9**). It is plausible that, in the physiological polyclonal settings, the variations in the activated status of individual Treg may be due to the TCR affinity to its cognate antigen, the availability of cognate antigen, and the strength and duration of TCR signals. Our SC4A analysis identified the *FOXP3*-driven genes, which are specific to activated FOXP3⁺ cells and include IL-2 and common gamma chain cytokine receptors (i.e., *IL2RA*, *IL2RB*, *IL15RA*, *IL4R*, and *IL2RG*), DNA replication licensing factors (e.g., MCM2), and transcription factors such as *PRDM1* (BLIMP1) and *IRF4* [which control the differentiation and function of eTreg (28)]. These gene modules are distinct from the Tfh-like genes and the activation genes (**Figure 9**) and may be controlled specifically by *FOXP3* under strong TCR signals. The expression of these genes is variable within the FOXP3⁺ T cells, suggesting that the transcriptional activities of these genes are dynamically regulated over time in tumor-infiltrating Treg. Thus, single-cell level analysis is becoming a key technology to address the heterogeneity of Treg. This study is one of the first single-cell analyses of Treg transcriptomes [during the review process of this manuscript, two studies that report the single cell transcriptomes of Treg were published (64) or deposited at a preprint-server (65)].

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
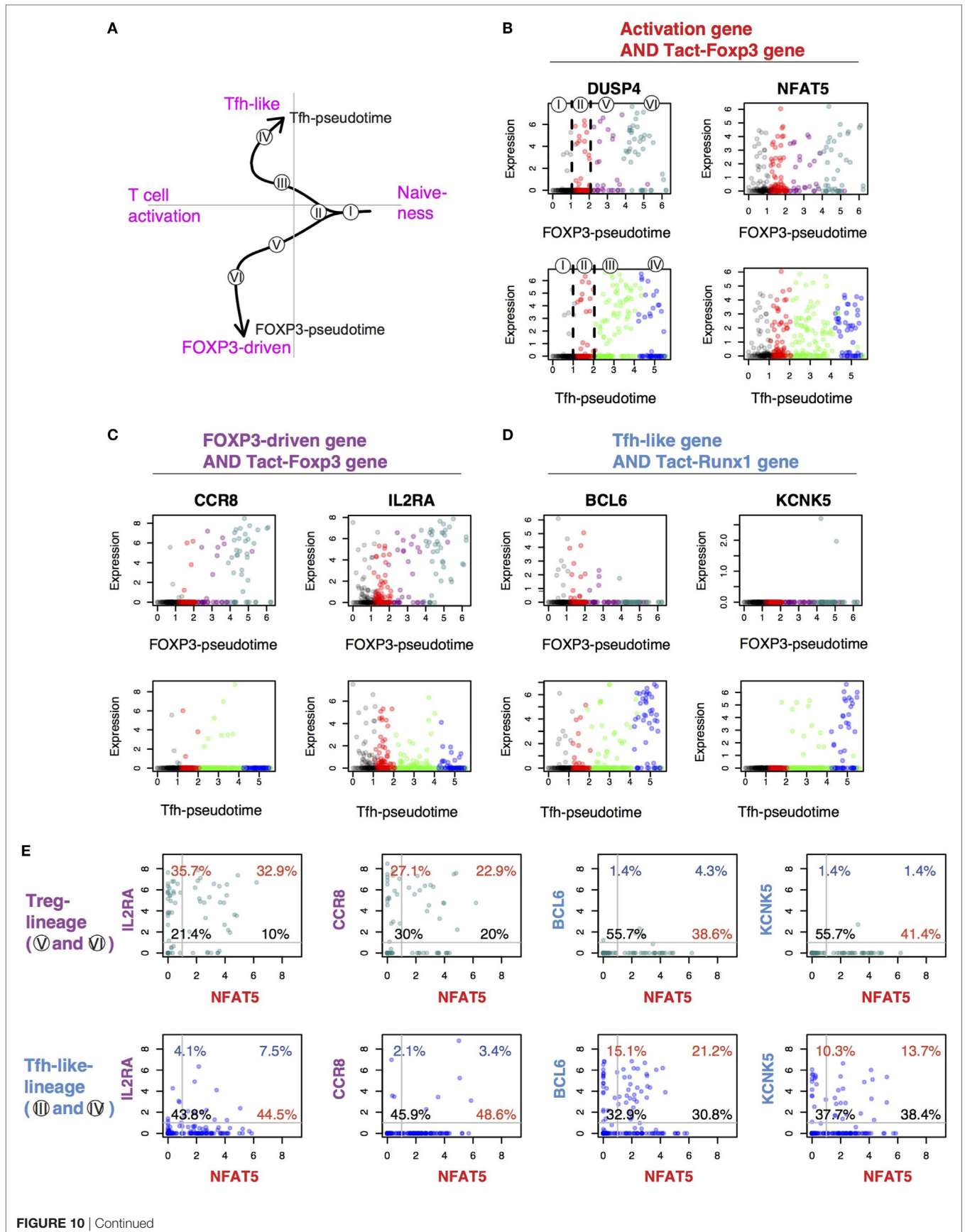
FIGURE 10 | Continued

FIGURE 10 | Identification of the conserved genes for the differential regulation of T follicular helper (Tfh)-like and Treg differentiation in activated T cells.
(A) The identified lineage curves and the bifurcation point in the tumor-infiltrating T cells. The number in circle indicates the cluster number in **Figure 9D**.
(B–D) The expression of selected feature genes was plotted against each pseudotime. Genes are from the intersect of **(B)** activation genes (**Figure 9C**) and Tact-Foxp3 genes (**Figure 4**), **(C)** FOXP3-driven genes (**Figure 9C**) and Tact-Foxp3 genes, and **(D)** Tfh-like genes (**Figure 9C**) and Tact-Runx1 genes (**Figure 4**).
(E) The expression of selected genes in the tumor-infiltrating T cells was shown by a 2D plot in a flow cytometric style. Data from Treg-lineage cells (Clusters V and VI, upper panels) and Tfh-like lineage cells (Clusters III and IV, 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.

FOXP3⁻ non-Treg express *DUSP6* (**Figure 10B**), which is a negative regulator of JNK-induced apoptosis through BIM activation, while *FOXP3* suppresses *DUSP6* expression and promotes the apoptosis mechanism (66). 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 (67). *BATF* was identified as a critical factor for the differentiation and accumulation of tissue-infiltrating Treg (68). 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 activation signature in Treg (**Figure 6B**). Intriguingly, some Treg may lose their Foxp3 expression and become ex-Treg, which are enriched in CD44^hi effector T cells or Tmem (18). By contrast, a Tmem population (precisely, Foxp3⁻CD44^hiCD73^hiFR4^hi T cells) efficiently express Foxp3 during lymphopenia (69). 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 express the same activated genes as Treg, while the additional Tmem-specific activation-induced genes (i.e., Tact-Runx1 genes) are uniquely repressed in Treg (**Figure 4**), which supports that Treg and Tmem are closely related populations (7). For example, fate-mapping experiments show that T cells that transiently express *Foxp3* contribute to the memory-like Foxp3⁻CD44^hi T cell population (18). Also, a memory-like T cell more efficiently generates Foxp3⁺ Treg than naïve T cells. Our findings (**Figure 4B**), together with these previous reports, are in accordance with the feedback control perspective proposed by Ono and Tanaka, whereby Treg and memory-like T cells are in a dynamic equilibrium and can change their phenotype to each other in order to fill the antigen niche and maintain CD4⁺ T cell homeostasis (7).

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 4C**). 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 (41), 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 recognized as a Treg marker, it is upregulated in all activated T cells, thus CTLA-4 is also a marker of activated T cells (46). CTLA-4 is in fact expressed by only a subset of rTreg (70), which may be more activated and proliferating *in vivo* (71). In fact, our study shows that CTLA-4 is expressed by non-Treg activated T cells including resting Tmem (**Figure 4D**) and FOXP3⁻ Tfh-like effector T cells in the tumor microenvironment (**Figure 9J**). 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. Importantly, although both FOXP3⁺ and FOXP3⁻ cells had the same relative level of activation (**Figure 8D**), the absolute number of FOXP3⁺ cells expressing *CTLA4* was lower than that of Tfh-type cells (**Figure 9J**), which suggests that therapeutic anti-CTLA4 antibodies (e.g., Ipilimumab) primarily target activated Tfh-like effector cells and thereby directly enhance their activities in tumor microenvironments. Future studies are required to experimentally investigate the *in vivo* dynamics of CTLA-4 expression in mice and humans.

By contrast, the expression of *PDCD1* was consistently high in all Tfh-like cells, while it was sparse among FOXP3⁺ cells (**Figure 9K**). The coexpression of *BCL6* and *IL21* in some of these PD-1⁺ cells indicates that Tfh differentiation occurs in the tumor microenvironment, presumably through the repeated and chronic exposure to quasi-self antigens (i.e., tumor antigens). Interestingly, the Tfh signature has been identified in type-I diabetes in both mice and humans (72). 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 future studies. However, the high *PDCD1* expression in Tfh-like cells may make them vulnerable to the negative immunoregulatory effect of PD-1 in tumor microenvironments (55). In fact, the most mature *PDCD1*^high Tfh-like cells (Cluster VI, **Figure 9K**) moderately decrease the expression levels of activation genes (**Figure 9E**), suggesting that these cells may have started to be regulated by PD1 ligands. Further experimental investigations are required to reveal how dynamically PD1 regulates T cells during the immune response.

Interestingly, *RUNX1* is completely repressed in the early phase of FOXP3-pseudotime (Cluster V) but re-expressed in the late phase of FOXP3-pseudotime (Cluster VI) (**Figure 9I**). Similar to RUNX1, some cells appear to be expressing *PDCD1* in the later phase of FOXP3-pseudotime in Treg-lineage cells. The reappearance of effector phenotype genes *RUNX1* and *PDCD1* in FOXP3[high] cells may indicate that these Treg are highly activated eTreg, which are actively participating in neutralizing the activities of effector T cells and Tfh. In addition, these FOXP3[high] cells may include ambivalent cells with both regulatory and effector features, or alternatively, be at the point of conversion to Tmem (**Figure 9K**). In bulk rTreg, *Pdcd1* was expressed at low levels in some Treg (**Figure 4E**) as well as PD-L2-encoding gene *Pdcd1lg2* (**Figure 4D**). Future studies are required to reveal the role of these cells.

This study has shown that SC4A and CCA results can be further analyzed by the lineage analysis using *pseudotime*, revealing the differentiation dynamics of tumor-infiltrating T cells as a function of arbitrary time. The CCA-pseudotime approach in this study has shown single-cell level heterogeneity of CD4[+] T cells including Treg and effector T cells, identifying the differentially regulated gene modules and the bifurcation point for T cell fates. Biologically, however, it should be emphasized that, because pseudotime is not a direct measurement of the time-dependent events, but rather is that of similarities between samples (73), future studies are required to analyze time-dependent events *in vivo*, ideally with a new experimental system to directly address the temporal dynamics.

Nevertheless, this study demonstrated the power of the SC4A/CCA approach to extract biological meaning from unannotated single-cell RNA-seq data, identifying the unique features of each T cell population and visualizing the relationships between T cell populations and genes without using annotation database. This has been made possible by the unique ability of CCA to add interpretable axes on the low dimensional plot of RNA-seq data. While a conventional dimension reduction method such as PCA requires an arbitrary interpretation of axes, CCA can thus improve the analysis of multidimensional RNA-seq data. In addition, SC4A has been shown to be effective in identifying 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. While the conventional CCA approach requires an explanatory dataset that represents the biological process of interest to be generated or identified, SC4A uses a single-cell dataset to select explanatory variables in the dataset according to their correlations to cell clusters. Although this step has been partially automated, the current SC4A method requires to pre-select ~20 candidate genes for explanatory variables due to the computational limitation (i.e., this process requires several hours for each analysis using a standard desktop). Potentially,

this process can be further improved and automated by combining SC4A with other algorithms (e.g., machine learning) and/or the use of annotation databases to identify the biologically meaningful genes that can model the entire single-cell data. In addition, further mathematical investigation of the method and the implementation of the computational algorithm using a low-level language may be beneficial. Although there is room for future improvements, our current study has shown that multidimensional data (i.e., data with many cell populations) and single-cell data can be better analyzed with the use of SC4A and CCA, allowing both hypothesis-driven investigation and exploration, and that the in-depth knowledge of immunology and gene regulation is useful for data analysis, in the same manner as it is required for experimental investigations. Thus, it is hoped that these tools will be widely used by experimental immunologists with a sound understanding of the biological significance of the 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.

AUTHOR CONTRIBUTIONS

MO conceived the conception of the work. AB and MO designed the work; wrote R codes and performed experiments, data analyses, and visualization; wrote the first draft of the manuscript. AB, TH, and MO wrote and completed the manuscript. TH and MO supervised the project.

ACKNOWLEDGMENTS

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at https://www.frontiersin.org/articles/10.3389/fimmu.2018.01444/full#supplementary-material.

REFERENCES

1. Macian F, Lopez-Rodriguez C, Rao A. Partners in transcription: NFAT and AP-1. *Oncogene* (2001) 20(19):2476–89. doi:10.1038/sj.onc.1204386

2. Malek TR. The biology of interleukin-2. *Annu Rev Immunol* (2008) 26:453–79. doi:10.1146/annurev.immunol.26.021607.090357

3. Hori S. Lineage stability and phenotypic plasticity of Foxp3(+) regulatory T cells. *Immunol Rev* (2014) 259(1):159–72. doi:10.1111/imr.12175

4. Setoguchi R, Hori S, Takahashi T, Sakaguchi S. Homeostatic maintenance of natural Foxp3(+) CD25(+) CD4(+) regulatory T cells by interleukin (IL)-2 and induction of autoimmune disease by IL-2 neutralization. *J Exp Med* (2005) 201(5):723–35. doi:10.1084/jem.20041982

5. Tran DQ, Ramsey H, Shevach EM. 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* (2007) 110(8):2983–90. doi:10.1182/blood-2007-06-094656

6. Chen W, Jin W, Hardegen N, Lei KJ, Li L, Marinos N, et al. Conversion of peripheral CD4+CD25− naive T cells to CD4+CD25+ regulatory T cells by TGF-β induction of transcription factor Foxp3. *J Exp Med* (2003) 198(12):1875–86. doi:10.1084/jem.20030152

7. Ono M, Tanaka RJ. Controversies concerning thymus-derived regulatory T cells: fundamental issues and a new perspective. *Immunol Cell Biol* (2016) 94(1):3–10. doi:10.1038/icb.2015.65

8. Thornton AM, Shevach EM. Suppressor effector function of CD4+CD25+ immunoregulatory T cells is antigen nonspecific. *J Immunol* (2000) 164(1):183–90. doi:10.4049/jimmunol.164.1.183

9. Samstein RM, Arvey A, Josefowicz SZ, Peng X, Reynolds A, Sandstrom R, et al. Foxp3 exploits a pre-existent enhancer landscape for regulatory T cell lineage specification. *Cell* (2012) 151(1):153–66. doi:10.1016/j.cell.2012.06.053

10. Levine AG, Arvey A, Jin W, Rudensky AY. Continuous requirement for the TCR in regulatory T cell function. *Nat Immunol* (2014) 15(11):1070–8. doi:10.1038/ni.3004

11. Vahl JC, Drees C, Heger K, Heink S, Fischer JC, Nedjic J, et al. Continuous T cell receptor signals maintain a functional regulatory T cell pool. *Immunity* (2014) 41(5):722–36. doi:10.1016/j.immuni.2014.10.012

12. Andersen P, Smedegaard B. CD4+ T-cell subsets that mediate immunological memory to *Mycobacterium tuberculosis* infection in mice. *Infect Immun* (2000) 68(2):621–9. doi:10.1128/IAI.68.2.621-629.2000

13. Coleman MM, Finlay CM, Moran B, Keane J, Dunne PJ, Mills KHG. The immunoregulatory role of CD4+FoxP3+CD25− regulatory T cells in lungs of mice infected with *Bordetella pertussis*. *FEMS Immunol Med Microbiol* (2012) 64(3):413–24. doi:10.1111/j.1574-695X.2011.00927.x

14. Ono M, Shimizu J, Miyachi Y, Sakaguchi S. Control of autoimmune myocarditis and multiorgan inflammation by glucocorticoid-induced TNF receptor family-related protein(high), Foxp3-expressing CD25+ and CD25- regulatory T cells. *J Immunol* (2006) 176(8):4748–56. doi:10.4049/jimmunol.176.8.4748

15. Bunce C, Bell EB. CD45RC isoforms define two types of CD4 memory T cells, one of which depends on persisting antigen. *J Exp Med* (1997) 185(4):767–76. doi:10.1084/jem.185.4.767

16. Min B, McHugh R, Sempowski GD, Mackall C, Foucras G, Paul WE. Neonates support lymphopenia-induced proliferation. *Immunity* (2003) 18(1):131–40. doi:10.1016/S1074-7613(02)00508-3

17. Curotto de Lafaille MA, Lino AC, Kutchukhidze N, Lafaille JJ. CD25- T cells generate CD25+Foxp3+ regulatory T cells by peripheral expansion. *J Immunol* (2004) 173(12):7259–68. doi:10.4049/jimmunol.173.12.7259

18. Miyao T, Floess S, Setoguchi R, Luche H, Fehling HJ, Waldmann H, et al. Plasticity of Foxp3(+) T cells reflects promiscuous Foxp3 expression in conventional T cells but not reprogramming of regulatory T cells. *Immunity* (2012) 36(2):262–75. doi:10.1016/j.immuni.2011.12.012

19. Linsley PS, Greene JL, Tan P, Bradshaw J, Ledbetter JA, Anasetti C, et al. Coexpression and functional cooperation of CTLA-4 and CD28 on activated T lymphocytes. *J Exp Med* (1992) 176(6):1595–604. doi:10.1084/jem.176.6.1595

20. Attridge K, Walker LS. Homeostasis and function of regulatory T cells (Tregs) in vivo: lessons from TCR-transgenic Tregs. *Immunol Rev* (2014) 259(1):23–39. doi:10.1111/imr.12165

21. Feuerer M, Hill JA, Kretschmer K, von Boehmer H, Mathis D, Benoist C. Genomic definition of multiple ex vivo regulatory T cell subphenotypes. *Proc Natl Acad Sci U S A* (2010) 107(13):5919–24. doi:10.1073/pnas.1002006107

22. Koch MA, Tucker-Heard G, Perdue NR, Killebrew JR, Urdahl KB, Campbell DJ. The transcription factor T-bet controls regulatory T cell homeostasis and function during type 1 inflammation. *Nat Immunol* (2009) 10(6):595–602. doi:10.1038/ni.1731

23. Zheng Y, Chaudhry A, Kas A, deRoos P, Kim JM, Chu TT, et al. Regulatory T-cell suppressor program co-opts transcription factor IRF4 to control T(H)2 responses. *Nature* (2009) 458(7236):351–6. doi:10.1038/nature07674

24. Linterman MA, Pierson W, Lee SK, Kallies A, Kawamoto S, Rayner TF, et al. Foxp3+ follicular regulatory T cells control the germinal center response. *Nat Med* (2011) 17(8):975–82. doi:10.1038/nm.2425

25. van der Veeken J, Gonzalez AJ, Cho H, Arvey A, Hemmers S, Leslie CS, et al. Memory of inflammation in regulatory T cells. *Cell* (2016) 166(4):977–90. doi:10.1016/j.cell.2016.07.006

26. Wyss L, Stadinski BD, King CG, Schallenberg S, McCarthy NI, Lee JY, et al. Affinity for self antigen selects Treg cells with distinct functional properties. *Nat Immunol* (2016) 17(9):1093–101. doi:10.1038/ni.3522

27. Dias S, D'Amico A, Cretney E, Liao Y, Tellier J, Bruggeman C, et al. Effector regulatory T cell differentiation and immune homeostasis depend on the transcription factor Myb. *Immunity* (2017) 46(1):78–91. doi:10.1016/j.immuni.2016.12.017

28. Cretney E, Xin A, Shi W, Minnich M, Masson F, Miasari M, et al. The transcription factors Blimp-1 and IRF4 jointly control the differentiation and function of effector regulatory T cells. *Nat Immunol* (2011) 12(4):304–11. doi:10.1038/ni.2006

29. Sakaguchi S, Miyara M, Costantino CM, Hafler DA. FOXP3+ regulatory T cells in the human immune system. *Nat Rev Immunol* (2010) 10(7):490–500. doi:10.1038/nri2785

30. Fujii H, Josse J, Tanioka M, Miyachi Y, Husson F, Ono M. Regulatory T cells in melanoma revisited by a computational clustering of FOXP3+ T cell subpopulations. *J Immunol* (2016) 196(6):2885–92. doi:10.4049/jimmunol.1402695

31. Ono M, Tanaka RJ, Kano M, Sugiman T. Visualising the cross-level relationships between pathological and physiological processes and gene expression: analyses of haematological diseases. *PLoS One* (2013) 8(1):e53544. doi:10.1371/journal.pone.0053544

32. Greenacre M. *Correspondence Analysis in Practice*. 2nd ed. London: Chapman & Hall/CRC (2008).

33. Tzeng J, Lu HH-S, Li W-H. Multidimensional scaling for large genomic data sets. *BMC Bioinformatics* (2008) 9(1):179. doi:10.1186/1471-2105-9-179

34. Ono M, Tanaka RJ, Kano M. Visualisation of the T cell differentiation programme by canonical correspondence analysis of transcriptomes. *BMC Genomics* (2014) 15:1028. doi:10.1186/1471-2164-15-1028

35. Greenacre M. *Canonical Correspondence Analysis in Social Science Research. Classification as a Tool for Research. Studies in Classification, Data Analysis, and Knowledge Organization*. Berlin, Heidelberg: Springer (2010). p. 279–91.

36. Oksanen J, Blanchet F, Friendly M, Kindt R, Legendre P, McGlinn D, et al. Vegan: Community Ecology Package. Technical report. *R Package Version 20–2*. (2011). Available from: https://cran.r-project.org/web/packages/vegan/index.html (Accessed: June 29, 2018).

37. Graffelman J, Tuft R. Site scores and conditional biplots in canonical correspondence analysis. *Environmetrics* (2003) 15(1):67–80. doi:10.1002/env.629

38. Butler A, Hoffman P, Smibert P, Papalexi E, Satija R. Integrating single-cell transcriptomic data across different conditions, technologies, and species. *Nat Biotechnol* (2018) 36(5):411–20. doi:10.1038/nbt.4096

39. Wakamatsu E, Mathis D, Benoist C. Convergent and divergent effects of costimulatory molecules in conventional and regulatory CD4+ T cells. *Proc Natl Acad Sci U S A* (2013) 110(3):1023–8. doi:10.1073/pnas.1220688110

40. Painter MW, Davis S, Hardy RR, Mathis D, Benoist C. Immunological genome project C. Transcriptomes of the B and T lineages compared by multiplatform microarray profiling. *J Immunol* (2011) 186(5):3047–57. doi:10.4049/jimmunol.1002695

41. Ono M, Yaguchi H, Ohkura N, Kitabayashi I, Nagamura Y, Nomura T, et al. Foxp3 controls regulatory T-cell function by interacting with AML1/Runx1. *Nature* (2007) 446(7136):685–9. doi:10.1038/nature05673

42. Tirosh I, Izar B, Prakadan SM, Wadsworth MH II, Treacy D, Trombetta JJ, et al. Dissecting the multicellular ecosystem of metastatic melanoma by single-cell RNA-seq. *Science* (2016) 352(6282):189–96. doi:10.1126/science.aad0501

43. *EMCluster: EM Algorithm for Model-Based Clustering of Finite Mixture Gaussian Distribution* [computer program]. R Package (2017).

44. Street K, Risso D, Fletcher RB, et al. Slingshot: cell lineage and pseudotime inference for single-cell transcriptomics. *bioRxiv*. (2017).

45. Beyer M, Thabet Y, Müller RU, Sadlon T, Classen S, Lahl K, et al. Repression of the genome organizer SATB1 in regulatory T cells is required for suppressive function and inhibition of effector differentiation. *Nat Immunol* (2011) 12(9):898–907. doi:10.1038/ni.2084

46. Walker LS, Sansom DM. Confusing signals: recent progress in CTLA-4 biology. *Trends Immunol* (2015) 36(2):63–70. doi:10.1016/j.it.2014.12.001

47. Szabo SJ, Sullivan BM, Stemmann C, Satoskar AR, Sleckman BP, Glimcher LH. Distinct effects of T-bet in TH1 lineage commitment and IFN-γ production in CD4 and CD8 T cells. *Science* (2002) 295(5553):338–42. doi:10.1126/science.1065543

48. Dang EV, Barbi J, Yang HY, Jinasena D, Yu H, Zheng Y, et al. Control of T(H)17/T(reg) balance by hypoxia-inducible factor 1. *Cell* (2011) 146(5):772–84. doi:10.1016/j.cell.2011.07.033

49. Ho IC, Hodge MR, Rooney JW, Glimcher LH. The proto-oncogene c-maf is responsible for tissue-specific expression of interleukin-4. *Cell* (1996) 85(7):973–83. doi:10.1016/S0092-8674(00)81299-4

50. Vinuesa CG, Linterman MA, Goodnow CC, Randall KL. T cells and follicular dendritic cells in germinal center B-cell formation and selection. *Immunol Rev* (2010) 237(1):72–89. doi:10.1111/j.1600-065X.2010.00937.x

51. Sakaguchi S, Yamaguchi T, Nomura T, Ono M. Regulatory T cells and immune tolerance. *Cell* (2008) 133(5):775–87. doi:10.1016/j.cell.2008.05.009

52. Seddon B, Zamoyska R. TCR and IL-7 receptor signals can operate independently or synergize to promote lymphopenia-induced expansion of naive T cells. *J Immunol* (2002) 169(7):3752–9. doi:10.4049/jimmunol.169.7.3752

53. Roychoudhuri R, Hirahara K, Mousavi K, Clever D, Klebanoff CA, Bonelli M, et al. BACH2 represses effector programs to stabilize T(reg)-mediated immune homeostasis. *Nature* (2013) 498(7455):506–10. doi:10.1038/nature12199

54. Lönnberg T, Svensson V, James KR, Fernandez-Ruiz D, Sebina I, Montandon R, et al. Single-cell RNA-seq and computational analysis using temporal mixture modelling resolves Th1/Tfh fate bifurcation in malaria. *Sci Immunol* (2017) 2(9):eaal2192. doi:10.1126/sciimmunol.aal2192

55. Okazaki T, Chikuma S, Iwai Y, Fagarasan S, Honjo T. A rheostat for immune responses: the unique properties of PD-1 and their advantages for clinical application. *Nat Immunol* (2013) 14(12):1212–8. doi:10.1038/ni.2762

56. Wong WF, Kurokawa M, Satake M, Kohu K. Down-regulation of Runx1 expression by TCR signal involves an autoregulatory mechanism and contributes to IL-2 production. *J Biol Chem* (2011) 286(13):11110–8. doi:10.1074/jbc.M110.166694

57. Buchbinder EI, Desai A. CTLA-4 and PD-1 pathways: similarities, differences, and implications of their inhibition. *Am J Clin Oncol* (2016) 39(1):98–106. doi:10.1097/COC.0000000000000239

58. McPherson RC, Konkel JE, Prendergast CT, Thomson JP, Ottaviano R, Leech MD, et al. Epigenetic modification of the PD-1 (Pdcd1) promoter in effector CD4(+) T cells tolerized by peptide immunotherapy. *Elife* (2014) 3:e03416. doi:10.7554/eLife.03416

59. Morgan DA, Ruscetti FW, Gallo R. Selective in vitro growth of T lymphocytes from normal human bone marrows. *Science* (1976) 193(4257):1007–8. doi:10.1126/science.181845

60. Moran AE, Holzapfel KL, Xing Y, Cunningham NR, Maltzman JS, Punt J, et al. T cell receptor signal strength in Treg and iNKT cell development demonstrated by a novel fluorescent reporter mouse. *J Exp Med* (2011) 208(6):1279–89. doi:10.1084/jem.20110308

61. Rosenblum MD, Way SS, Abbas AK. Regulatory T cell memory. *Nat Rev Immunol* (2016) 16(2):90–101. doi:10.1038/nri.2015.1

62. Miyara M, Yoshioka Y, Kitoh A, Shima T, Wing K, Niwa A, et al. Functional delineation and differentiation dynamics of human CD4+ T cells expressing the FoxP3 transcription factor. *Immunity* (2009) 30(6):899–911. doi:10.1016/j.immuni.2009.03.019

63. Weissler KA, Caton AJ. The role of T-cell receptor recognition of peptide: MHC complexes in the formation and activity of Foxp3(+) regulatory T cells. *Immunol Rev* (2014) 259(1):11–22. doi:10.1111/imr.12177

64. Zemmour D, Zilionis R, Kiner E, Klein AM, Mathis D, Benoist C. Single-cell gene expression reveals a landscape of regulatory T cell phenotypes shaped by the TCR. *Nat Immunol* (2018) 19(3):291–301. doi:10.1038/s41590-018-0051-0

65. Miragaia RJ, Gomes T, Chomka A, et al. Single cell transcriptomics of regulatory T cells reveals trajectories of tissue adaptation. *bioRxiv*. (2017).

66. Tai X, Erman B, Alag A, Mu J, Kimura M, Katz G, et al. Foxp3 transcription factor is proapoptotic and lethal to developing regulatory T cells unless counterbalanced by cytokine survival signals. *Immunity* (2013) 38(6):1116–28. doi:10.1016/j.immuni.2013.02.022

67. Sahni H, Ross S, Barbarulo A, Solanki A, Lau CI, Furmanski A, et al. A genome wide transcriptional model of the complex response to pre-TCR signalling during thymocyte differentiation. *Oncotarget* (2015) 6(30):28646–60. doi:10.18632/oncotarget.5796

68. Hayatsu N, Miyao T, Tachibana M, Murakami R, Kimura A, Kato T, et al. Analyses of a mutant Foxp3 allele reveal BATF as a critical transcription factor in the differentiation and accumulation of tissue regulatory T cells. *Immunity* (2017) 47(2):268–83.e9. doi:10.1016/j.immuni.2017.07.008

69. Kalekar LA, Schmiel SE, Nandiwada SL, Lam WY, Barsness LO, Zhang N, et al. CD4(+) T cell anergy prevents autoimmunity and generates regulatory T cell precursors. *Nat Immunol* (2016) 17(3):304–14. doi:10.1038/ni.3331

70. Takahashi T, Tagami T, Yamazaki S, Uede T, Shimizu J, Sakaguchi N, et al. Immunologic self-tolerance maintained by CD25(+)CD4(+) regulatory T cells constitutively expressing cytotoxic T lymphocyte-associated antigen 4. *J Exp Med* (2000) 192(2):303–10. doi:10.1084/jem.192.2.303

71. Tang AL, Teijaro JR, Njau MN, Chandran SS, Azimzadeh A, Nadler SG, et al. CTLA4 expression is an indicator and regulator of steady-state CD4+FoxP3+ T cell homeostasis. *J Immunol* (2008) 181(3):1806–13. doi:10.4049/jimmunol.181.3.1806

72. Kenefeck R, Wang CJ, Kapadi T, Wardzinski L, Attridge K, Clough LE, et al. Follicular helper T cell signature in type 1 diabetes. *J Clin Invest* (2015) 125(1):292–303. doi:10.1172/JCI76238

73. Haghverdi L, Buttner M, Wolf FA, Buettner F, Theis FJ. Diffusion pseudotime robustly reconstructs lineage branching. *Nat Methods* (2016) 13(10):845–8. doi:10.1038/nmeth.3971