Transcriptomic Profiling of Human Effector and Regulatory T Cell Subsets Identifies Predictive Population Signatures

Barbara Höllbacher, Thomas Duhen, Samantha Motley, Maria M. Klicznik, Iris K. Gratz and Daniel J. Campbell

*ImmunoHorizons* 2020, 4 (10) 585-596
doi: https://doi.org/10.4049/immunohorizons.2000037
http://www.immunohorizons.org/content/4/10/585

This information is current as of May 13, 2021.

**Supplementary Material**
http://www.immunohorizons.org/content/suppl/2020/10/09/4.10.585.DCSupplemental

**References**
This article cites 28 articles, 7 of which you can access for free at:
http://www.immunohorizons.org/content/4/10/585.full#ref-list-1

**Email Alerts**
Receive free email-alerts when new articles cite this article. Sign up at:
http://www.immunohorizons.org/alerts
Transcriptomic Profiling of Human Effector and Regulatory T Cell Subsets Identifies Predictive Population Signatures

Barbara Höllbacher,*1 Thomas Duhen,*1 Samantha Motley,* Maria M. Klicznik,† Iris K. Gratz,*+,‡ and Daniel J. Campbell*,§

*Immunology Program, Benaroya Research Institute, Seattle, WA 98101; †Department of Biosciences, University of Salzburg, 5020 Salzburg, Austria; ‡EB House Austria, Department of Dermatology, University Hospital of the Paracelsus Medical University, 5020 Salzburg, Austria; and §Department of Immunology, University of Washington School of Medicine, Seattle, WA 98195

ABSTRACT

After activation, CD4+ Th cells differentiate into functionally specialized populations that coordinate distinct immune responses and protect against different types of pathogens. In humans, these effector and memory Th cell subsets can be readily identified in peripheral blood based on their differential expression of chemokine receptors that govern their homeostatic and inflammatory trafficking. Foxp3+ regulatory T (Treg) cells can also be divided into subsets that phenotypically mirror each of these effector populations and share expression of key transcription factors and effector cytokines. In this study, we performed comprehensive transcriptional profiling of 11 phenotypically distinct Th and Treg cell subsets sorted from peripheral blood of healthy individuals. Despite their shared phenotypes, we found that mirror Th and Treg subsets were transcriptionally dissimilar and that Treg cell populations showed limited transcriptional diversity compared with Th cells. We identified core transcriptional signatures shared across all Th and Treg cell populations and unique signatures that define each of the Th or Treg populations. Finally, we applied these signatures to bulk Th and Treg RNA-sequencing data and found enrichment of specific Th and Treg cell populations in different human tissues. These results further define the molecular basis for the functional specialization and differentiation of Th and Treg cell populations and provide a new resource for examining Th and Treg specialization in RNA-sequencing data. ImmunoHorizons, 2020, 4: 585–596.

INTRODUCTION

The outcome of adaptive immune responses is dictated in large part by the activity of CD4+ Th cells. Th cells differentiate into functionally specialized populations that, through secretion of effector cytokines, coordinate the activities of immune and stromal cells to control different types of pathogens and dangerous toxins (1). Thus IFN-γ–producing Th1 cells are essential for control of intracellular pathogens; Th2 cells that produce IL-4, IL-5, and IL-13 mediate protection against helminth infection and poisonous venoms; and Th17 cells that produce IL-17 and IL-22 provide immunity to extracellular bacterial and fungal pathogens such as Staphylococcus aureus and Candida albicans. More recently described Th populations include Th22 cells (2, 3), which produce...
IL-22 but not IL-17 and appear to function specifically within the skin during tissue-repair responses, and Th1/17 cells that share key features with both Th1 and Th17 cells including coproduction of IFN-γ and IL-17, and are enriched in cells producing GM-CSF (4, 5). In human blood, these Th subsets can be readily identified based on their differential expression of chemotactrant receptors and adhesion molecules that control their specific migration to distinct inflammatory sites and likely sites of pathogen entry.

Although essential for protection against infection, dysregulated Th cell responses are pathogenic in immune-mediated diseases. These include organ-specific autoimmune diseases such type-1 diabetes (Th1), psoriasis (Th17, Th22), and multiple sclerosis (Th1/17), as well as allergic hypersensitivities and asthma (Th2). The activities of Th cells are restrained by T regulatory (Treg) cells, a subset of CD4+ T cells that constitutively expresses the IL-2R α-chain CD25 and the transcription factor Foxp3. Treg cells dampen the activation and function of Th cells via multiple mechanisms, including production of anti-inflammatory cytokines such as IL-10, IL-35, and TGF-β, CTLA4-mediated blockade of T cell costimulation, sequestration of effector cytokines, metabolic disruption of Th cells, competition for peptide/MHC complexes, and direct cell lysis (6). Importantly, we and others have shown that Treg cells are not homogenous but instead, like Th cells, can be divided into diverse subsets (7). These subsets phenotypically “mirror” each of the major Th populations, and the majority of Treg cells in each subset share expression of lineage-defining transcription factors such as T-bet and RORγt (8). These Treg cells can even produce effector cytokines such as IFN-γ and IL-17 (8, 9). These shared features suggest that Th and Treg cells may be more similar than previously appreciated. We also showed that production of the immunosuppressive cytokine IL-10 was limited to Th1- and Th17-like Treg cells, suggesting that different Treg cell subsets may employ specialized immunoregulatory mechanisms to modulate different types of inflammatory responses.

To address these issues, we performed comprehensive transcriptional profiling of 11 distinct CD4+ T cell subsets sorted from the peripheral blood of three healthy donors. We found that although they shared many phenotypic features, mirror Th and Treg cell populations were transcriptionally divergent and that the Treg cell populations showed limited transcriptional diversity compared with Th cells. We further identified core transcriptional signatures shared across all Th or Treg cells, as well as unique signatures that define each of the Th or Treg cell populations. Finally, we applied these signatures to RNA-sequencing (RNA-seq) data from bulk Th and Treg cells to show enrichment of specific Th and Treg cell populations in healthy and diseased human tissues.

**MATERIALS AND METHODS**

**Flow cytometric sorting**

Blood samples were obtained from healthy donors participating in the Benaroya Research Institute Immune-Mediated Disease Registry. Informed consent was obtained from all subjects according to institutional review board–approved protocols at Benaroya Research Institute and following the Declaration of Helsinki. CD4+CD25high Treg cells were enriched from PBMCs after staining with PE-cyanine 5 (PE-Cy5)–labeled anti-CD25 Ab (BioLegend), followed by positive selection using anti-PE and anti-Cy5 microbeads (Miltenyi Biotec). On the negative fraction, CD4+CD25− Th cells were purified by positive selection with CD4–specific microbeads (Miltenyi Biotec). Memory T cell subsets were sorted to more than 97% purity as CD4+CD45RA−RO−CD127− CD25− using allopheococyanin-780–conjugated anti-CD45RA Ab (eBioscience), Alexa 700–conjugated anti-CD45RO Ab (BioLegend), v450–conjugated anti-CD127 Ab (BD Bioscience), PE-Cy5–conjugated anti-CD25 Ab (BioLegend), and Qdot655–conjugated anti-CD4 Ab (eBioscience). Abs used for sorting of memory Th and Treg cell subsets were as follows: PE-Cy7–conjugated anti-CCR6 Ab (BioLegend), PE-conjugated anti-CCR10 Ab (R&D Systems), PerCP/Cy5.5–conjugated anti-CCR4 Ab (BioLegend), and Alexa Fluor 488–conjugated anti-CXCR3 Ab (BD Bioscience). Cells were sorted with a FACSARia II (BD Biosciences).

**Library construction and RNA-seq**

RNA-seq libraries were constructed from up to 100 ng of total RNA using the TruSeq RNA Sample Prep Kit v2 (Illumina). Libraries were clustered on a flow cell using the TruSeq Paired-end Cluster Kit, v3 using a cBot-clustering instrument (Illumina), followed by paired-end sequencing on a HiScanSQ (Illumina) for 50 cycles in either direction. After the run was completed, the reads were demultiplexed and FASTQs were generated for each sample using CASAVA.

**RNA-seq analysis**

Base-calling was performed automatically by Illumina real time analysis software and demultiplexing was performed with the program Casava. One 3′-end base was removed from all reads, followed by quality-based trimming from both ends until minimum base quality for each read was ≥30. Tophat aligned reads to GRCh38, using Ensembl annotation release number 77, and the read counts per Ensembl gene ID were computed with featureCounts. Sequencing, alignment, and quantitation metrics were obtained for FASTQ, BAM/SAM, and count files using FastQC, Picard, TopHat, Samtools, and htseq-count. All samples passed quality check with mapped reads with duplicates >80%, median coefficient of variation coverage <0.8 and total fastq reads of >5 Mio. Protein-coding transcripts with a minimum of 1 counts per million (CPM) in at least 5% of the total number of libraries were retained and the ensemble gene IDs mapped to HGNC gene symbols. Count data were normalized using edgeR’s trimmed means of M values (10). For linear modeling, we computed a coefficient for each subset with the naive population as reference and accounted for donor variation as random factor using the R package limma (11). To determine differences between the Th and Treg cells that are common for all subsets, we set up pairwise contrasts between mirror Th and Treg subsets and examined the overlap. For visualizations such as principal component analysis...
(PCA) and expression plots, this is approximated with the limma function `removeBatchEffect`. Cluster means were computed with the k-means function for three total clusters, transformed into PCA space, and added to the PCA plot. For gene expression dot plots the base of two was raised to the power of rembatch-corrected values (which were originally taken from output of the `voomWithQualityWeights` function) to unlog them. The p values for the Venn diagrams of differentially expressed genes between Tregs and their respective Th counterpart were adjusted using global Benjamini-Hochberg adjustment of FDR = 0.05 and summarized with the `limma::decideTests` function including a cutoff of absolute log2 fold-change >1. Heatmaps are based on log-transformed expression values, were z-scaled by rows, and were plotted using the R package ComplexHeatmap (12). Euclidean distances between all samples were computed with default settings from the `stats::dist()` function and plotted as a heatmap with manually defined sample order. To compare the heterogeneity of Th and Treg populations, we recomputed the distances of Treg to Treg samples, Th to Th samples, and Treg to Th samples separately. The results are summarized and plotted as density estimates using ggplot. Unique subset signatures were determined by making contrasts of a given subset with all other Th or Treg populations (excluding Th1/17 cells), and genes that were differentially expressed at adjusted p value < 0.05 between a given subset and all other subsets were termed to be part of a population signature.

To evaluate the relative expression of signature gene sets in other samples, we rank ordered all genes for a given samples and computed the mean rank for the significantly upregulated genes in each signature. Gene expression data of IFN-γ- and IL-10-producing Treg populations from the public dataset GSE116283 (https://www.ncbi.nlm.nih.gov/geo/) were accessed through Gene Expression Omnibus (GEO) and filtered for genes with CPM in at least 25% of the samples.

Gene expression matrix and series matrix of breast cancer and PBMC samples were retrieved through the GEO accession number GSE89225 (https://www.ncbi.nlm.nih.gov/geo/). The samples run on the IonTorrent platform were downloaded and assigned to their corresponding HGNC gene keys. Gene expression data were filtered for protein-coding genes that were also expressed in our dataset and those with CPM >1 in at least 25% of the samples.

Data accessibility
RNA-seq data generated for this study are available from GEO as SuperSeries GSE149090.

RESULTS
RNA-seq analysis of human CD4+ T cell populations
To comprehensively profile human CD4+ Th and Treg cells, we performed RNA-seq on 11 distinct CD4+ T cell populations sorted directly from peripheral blood of three healthy individuals on the basis of surface markers and chemokine receptors as previously described (8) (Fig. 1). Th cells were identified as CD127+CD25 and were sorted into naive (CD45RA+), as well as Th1 (CD45RA−CXCR3−CCR6−CCR4−CCR10−), Th1/17 (CD45RA−CXCR3−CCR6−CCR4−CCR10−), Th17 (CD45RA−CXCR3−CCR6−CCR4−CCR10−), and Th2 (CD45RA−CXCR3−CCR6−CCR4−CCR10−) cell fractions. Treg1, Treg17, Treg1/17, Treg2, and Treg22 populations were sorted within the CD127+CD25+ T cell using the same markers (Fig. 1). Expression of genes encoding the surface receptors used for cell sorting faithfully segregated into the various Th and Treg transcriptional profiles, and analysis of genes encoding key transcription factors associated with Treg cells (FOXP3), or with Th1/Treg1 (TBX21), Th17/Treg17 (RORC), and Th2/Treg2 (GATA3) subsets were expressed as expected within the appropriate cell populations (Fig. 2A). Thus, the sorting strategy...
FIGURE 2. Identification of core Th and Treg cell signatures.

(A) Expression analysis of the indicated chemokine receptors and transcription factors that define the different T cell populations. Values represent batch-effect–corrected counts per million. (B) PCA of all samples run on all expressed genes. Cluster centers representing naive T cells (1), Th cells (2), and Treg cells (3) returned by k-means clustering. (C) Venn diagrams showing the overlap in genes that were significantly (adjusted p value, \(p < 0.05\) and absolute log2 fold-change \(>1\)) upregulated (left) or downregulated (right) when comparing Treg cells with their mirror Th counterpart as indicated. Numbers indicate genes that were up-/downregulated in all comparisons (core of the Venn diagrams), or genes that were differentially expressed in only one comparison. (D) Heatmap showing expression of 31 genes identified by Pesenacker et al. as activation-independent Treg markers in the indicated samples. Bolded genes were significantly differentially expressed in all comparisons of mirror Th and Treg populations. (E) Volcano plots showing pairwise comparisons of mirror Th and Treg cell populations. Genes identified by Pesenacker et al. (13) as higher in Treg cells are highlighted in red, those higher in Th cells in blue. The \(\chi^2\) test statistic and \(p\) values indicate whether the sets of up- and downregulated identified by Pesenacker et al. (13) are differentially distributed in the indicated Treg cell versus Th cell comparison.
we employed properly isolated functionally specialized Th and Treg cell populations and our data provide a comprehensive transcriptional analysis of in vivo differentiated Th and Treg cells.

**Th and Treg cells are transcriptionally distinct**

Using PCA, we compared the overall transcriptional profile of all Th and Treg subsets (Fig. 2B). K-means clustering was used to identify three distinct clusters, which corresponded to samples from naïve Th cells (cluster 1), memory Th cells (cluster 2), and Treg cells (cluster 3). Thus, despite their phenotypical similarity and shared expression of key lineage-defining transcription factors, mirror Th and Treg cell populations (e.g., Th1 and Treg cells) are transcriptionally distinct and more closely resemble other Th and Treg cells. To define core Th and Treg transcriptional signatures, we performed pairwise comparisons between each of the individual Th and Treg cell mirror pairs, and identified 294 genes that were significantly higher (log2 fold-change >1, adjusted p value <0.05) in all Treg cell populations and 492 that were more highly expressed in all Th populations (Fig. 2C). Analysis of these gene sets revealed many genes previously identified as differentially expressed in Treg versus Th cells (13, 14) (Supplemental Table I). These included FOXP3, CTLA4, ENTPD1 (CD39), IKZF2 (Helios), and TNFRSF9 that are preferentially expressed in Treg cells and BHLHE40, CD40LG, ID2, and IL2 that are more highly expressed in Th cells.

Because Treg cells are largely specific for either autoantigens or for components of the microbiome present on barrier surfaces, they are subject to chronic stimulation and many of the genes previously associated with Treg cells are activation-induced genes that are not actually Treg cell specific (13). Consistent with this, among genes more highly expressed in Treg cells, several were genes we previously found to be downregulated in subjects treated with the costimulation blocking drug abatacept (CTLA4-Ig) (15), including ARHGAPI1A, CENPE, DUSP4, ER1I, GXYLT1, HLLS, NUSAPI, PMAIP1, SGMS1, and TOP2A (Supplemental Table I). To account for these and other activation-induced genes, Pesenacker et al. (13) compared the transcriptomes of resting and activated Th and Treg cells, and identified a 31-gene “activation-independent” Treg cell signature. Of these, 21 were differentially expressed in the expected direction in all five Th/Treg cell comparisons (bolded in Fig. 2D), and of the remaining genes, three (ICA1, ZC2HClA, and ABCB1) were significant in four out of five comparisons (Fig. 2E).

In our PCA analysis (Fig. 2B), all Treg cell populations clustered tightly together (cluster 3), whereas there was substantially more spreading within the cluster of memory Th cell populations (cluster 2), indicating that, overall, Treg cells are transcriptionally less diverse than the Th cells. To further examine this, we calculated the Euclidian distance of all samples relative to each other to quantitatively assess their overall similarity (Fig. 3). As evidenced in the PCA, Treg and Th subsets were highly divergent, with a mean Euclidian distance in these comparisons of 129.7. The Euclidian distance comparing all the Th cell samples to each other had a bimodal distribution, with a peak of highly similar samples corresponding to comparisons between the same Th populations isolated from the three donors (e.g., comparing all Th1 samples, mean Euclidian distance 59.06), and a second significantly more-distant peak corresponding to comparisons between the different types of Th cell populations (mean Euclidian distance 82.44, p < 2 × 10^-16, Welch two sample t test). By contrast, we found that the Euclidian distance between different Treg populations was relatively uniform, and that the mean Euclidian distance between all of the Treg populations was significantly smaller than observed in the comparisons of the Th populations (mean Euclidian distance 69.71, p < 2 × 10^-16, Welch two sample t test). Thus, despite their similar degree of phenotypic heterogeneity, the different Treg cell populations are transcriptionally less diverse than their Th counterparts.

**Functional specialization of Treg cells**

The differentiation of Th cells into specialized subsets is accompanied by acquisition of distinct effector functions that coordinate different types of immune responses. Similarly, a large number of effector mechanisms are used by Treg cells to dampen inflammation and prevent autoimmunity. These include production of anti-inflammatory cytokines, expression of coinhibitory receptors, cytokine sequestration, modulation of pro- and anti-inflammatory metabolites, and direct target cell lysis. This raises the intriguing possibility that, like Th cells, different Treg cell
FIGURE 4. Functional and molecular specialization of Treg cell populations.

(A) Expression analysis of genes controlling different functions of Treg cells as indicated. (B) Expression analysis of transcription factors implicated in Treg specialization. Values represent batch-effect–corrected CPM. Significance levels were determined by ANOVA followed by Tukey honestly significant difference for testing pairwise differences. Significant differences with the naive samples are not highlighted.

https://doi.org/10.4049/immunohorizons.2000037
subsets may employ specialized immunoregulatory mechanisms to modulate different types of inflammatory responses. Indeed, we previously demonstrated that production of the anti-inflammatory cytokine IL-10 was restricted to Treg1 and Treg17 cells (8). To further examine the functional specialization of Treg cells, we assessed expression of a set of key genes selected from the literature that control Treg cell activity in different settings (Fig. 4A). Consistent with our previous findings, IL10 expression was largely limited to the Treg1, Treg17, and Treg1/17 populations, and these also selectively expressed the effector cytokines IFNG and IL17A, which has previously been reported for human Treg cells. Treg1 cells also most highly expressed the coinhibitory receptors LAG3 and HAVCR2 (TIM-3), and the cytolytic effectors GZMA and GZMK. By contrast, expression the TGF-β-activating molecule LRRC32 (GARP) was highest in Treg2 cells, the decoy IL-1R IL1R2 was most highly expressed by Treg17 and Treg1/17 cells, and HPGD, which is used by Treg cells to degrade the proinflammatory PG PGE2 (16), was most highly expressed in Treg22 cells. Other genes implicated in Treg function, including TGFB1, ITGAV, ITGB8, IL12A (a component of the anti-inflammatory cytokine IL-35), CTLA4, PDCD1 (PD1), TIGIT, PREDI, and ENTPD1 (CD39) did not show strong preferential expression in any Treg population, and NTSE (CD73) was barely detected in these ex vivo–isolated Treg cells.

Treg cell specialization depends on their selective expression of a set of transcription factors that alter their migration and function. In addition to T-bet (TBX21), GATA3 (GATA3) and RORyt (RORC) that promote Treg1, Treg2, and Treg17 differentiation, other transcription factors implicated in control of Treg cell function include IRF4, BACH2, RORA, PPARy, BCL6, IKZF2 (Helios), IKZF4 (Eos), ID2, ID3, and FOXO1. Indeed, IRF4, RORA, IKZF2, IKZF4, and ID3 were significantly differentially expressed between

![Image of transcriptional signatures in Th and Treg cells.](https://doi.org/10.4049/immunohorizons.2000037)
various Treg cell subsets, and this may contribute to their molecular specialization. Expression of PPARG was highest in Treg17 cells, but this did not reach statistical significance. By contrast, BCL6 and FOXO1 were expressed equally in all Treg populations whereas ID2 and BACH2 showed limited expression in all Treg populations relative to naive T cells, and therefore these factors are unlikely to contribute to their phenotypic and functional heterogeneity.

**Identification of subset-specific transcriptional signatures in Th and Treg cells**

Although the functional specialization of Th cells was first described over 30 y ago, their complex phenotypes, specialized functions and pathways of differentiation are still being defined. Much of our understanding of human Th cell differentiation has been derived from cells differentiated in vitro, and these may differ substantially from cells differentiated in vivo. Therefore, to gain further insight into the functional specialization of Th cells, we identified the unique transcriptional signature for each Th subset based on genes significantly up- or downregulated in comparison with all other Th subsets (Fig. 5A). Because Th1/17 cells are a hybrid population with characteristics of both Th1 and Th17 cells (4), we excluded them from these comparisons to more readily define the Th1 and Th17 subset signatures. Analysis of these signatures revealed selective expression of genes that were previously implicated in Th cell specialization, localization, and function including IFNG, IL12RB2, TBX21, and EOMES that are selectively expressed in Th1 cells; IL12RB1 and ABCB1 in Th17 cells; GATA3, PTGDR2, and IL4R in Th2 cells; and ITGAE, CD9, and CD101 in Th22 cells. We also identified genes in each of the subset signatures that may differentially influence cell metabolism (decreased HK2 expression in Th1 cells), cell signaling (increased SOCS2 expression in Th17 cells), and cell function (increased TNFSF11 [RANKL] expression in Th2 cells). Similarly, by comparing the transcriptional profiles of the Treg subsets (excluding Treg1/17 cells), we identified population-specific transcriptional signatures in Treg cells (Fig. 5B). A full list of upregulated and downregulated genes in all Th and Treg population signatures is found in Supplemental Table II.

![FIGURE 6. Shared transcriptional signatures of mirror Th and Treg cell populations.](https://doi.org/10.4049/immunohorizons.2000037)
To assess expression of the gene signatures between different populations, we ranked all genes in order of expression for each sample and determined the mean gene list rank for each set of signature genes in all populations examined. For this, we focused on genes that were upregulated in each of the signatures and therefore positively identified a given population. This analysis demonstrated that the signatures of both the Th1 and Th17 populations were also shared in the Th1/17 cells (Fig. 6A), consistent with the Th17/17 population having hybrid characteristics of both Th1 and Th17 cells. Analysis of the gene signatures of mirror Th and Treg populations revealed little overlap in the identity of the differentially expressed genes (Fig. 6B). However, examining the mean gene list rank showed that in all cases, Th cell subset signatures were significantly enriched in the corresponding Treg cell subset, and that Treg subset signatures were significantly enriched in the corresponding Th populations (Fig. 6C). Thus, despite their overall transcriptional dissimilarity (as highlighted in Fig. 2), the Th and Treg cell populations do share common phenotype-associated transcriptional programs.

**Enrichment of Th and Treg transcriptional signatures in bulk RNA-seq data**

The identification of specific transcriptional signatures for each of the Th and Treg cell subsets from peripheral blood raises the possibility that the gene list rank approach could be used to functionally characterize bulk Th and Treg samples. To test this idea, we first applied the Treg subset signatures to ranked gene lists of RNA-seq samples from peripheral blood Treg cells that were sorted into four populations based on their expression of the cytokines IFN-γ and IL-10 (17). As expected, based on their preferential production of these cytokines, we found that genes specifically upregulated in Treg cells were significantly enriched in the IFN-γ^+IL-10^+ and IFN-γ^+IL-10^- cells, that Treg1 signature genes were also enriched in the IFN-γ^-IL-10^- cells, and that Treg2 and Treg22 signature genes were elevated in the IFN-γ^-IL-10^- cells (Fig. 7A). To further extend this approach, we applied our signatures to data from bulk Th and Treg cells isolated from either peripheral blood or from human breast carcinoma (18). In this case, Th cells from tumors were enriched in Th1 signature genes,
whereas those from PBMC were comparatively enriched in Th17 and Th2 signature genes. This is consistent with the initial characterization of these data reporting high expression of Th1 signature genes EOMES, GZMK, CXCR3, and IFNG in tumor-infiltrating Th cells (18). Similarly, Treg cells from the tumors were enriched in Treg1 signature genes but showed decreased expression of Treg17 and Treg22 signature genes. Lastly, we applied the Th cell signatures to RNA-seq data comparing gene expression by CCR7+ central memory T (TCM) cells and CCR7+ effector memory T (TEM) Th cells that express the skin-homing receptor cutaneous lymphocyte Ag (CLA) from peripheral blood with CLA+ TCM and TEM from skin. In this analysis, we found a significant enrichment of Th22 signature genes in both TCM and TEM populations from skin. Th22 cells in the blood express a set of surface adhesion molecules and chemokine receptors indicative of skin-tropism, including CLA, CCR4, and CCR10, and IL-22–producing T cells are enriched in the skin where this cytokine can act directly on keratinocytes to induce tissue-repair and antimicrobial responses (2, 3). Together, these examples demonstrate the utility of the gene signatures we have identified for functional analysis of bulk Th and Treg cells isolated from both healthy and diseased tissues.

**DISCUSSION**

The functional specialization of CD4+ T cells is the cornerstone of effective adaptive immune responses that prevent the growth and spread of various pathogens while limiting collateral tissue damage and autoimmunity (19). Through analysis of 11 human CD4+ T cell populations sorted directly ex vivo, we have comprehensively defined the transcriptional basis for human CD4+ T cell specialization which resulted in several important insights. We assessed the transcriptional relationship between the various CD4+ T cell populations and defined core Th and Treg gene signatures as well as unique signatures of each population that can be applied to analysis of existing RNA-seq data sets for functional profiling of human Th and Treg cells in different tissue sites.

The majority of Treg cells are believed to develop in the thymus upon encounter with self-antigen, and therefore, this can be thought of as the first branch point in the functional differentiation of CD4+ T cells. At barrier sites, Treg cells can also develop from conventional naive T cells activated in tolerogenic conditions to help enforce tolerance to harmless environmental Ags, including components of the skin and intestinal microbiomes (20). The fact that all of the Treg cell populations differentially expressed genes generally involved in T cell activation and proliferation is indeed consistent with each containing large fractions of highly reactive cells. Among these, it is interesting that each Treg population showed increased expression of the transcription factor TOX, which has recently been found to promote an exhausted phenotype in chronically stimulated CD8+ T cells and thus to prevent activation-induced cell death (21–23).

Interestingly, despite displaying similar phenotypic diversity, transcriptionally, the Treg populations were far more uniform then their Th counterparts. This likely reflects the dominant role of Foxp3 on the Treg transcriptome, and the ability of this single factor to convert potentially pathogenic effector T cells into potent Treg cells (24). Induction of Foxp3 in Treg cells initiates dramatic changes in gene expression that underlie their suppressive functions. Foxp3 activates a broad transcriptional program that controls Treg cell function and homeostasis, and includes genes such as CTLA4 and IL2RA (CD25). Foxp3 also inhibits expression of key effector cell molecules such as IL2, IFNG, and CD40LG that were more highly expressed in Th cells (14). Indeed, our analysis clearly shows that despite these shared properties, functionally specialized Th subsets are transcriptionally distinct from their Treg cell mirror counterparts. We therefore used each of these comparisons to identify “core” Th and Treg gene signatures that confirm and extend previous analysis of the specific Treg cell transcriptome. All Treg populations showed increased expression of core immunosuppressive genes (CTLA4, TIGIT, ITGB8, LRRC32, ENTPD1) and specific chemokine/cytokine receptors (CCR3, CXCR6, CSF2RB, IL1R1). Unlike Th cells that undergo extensive functional diversification based on differential expression of key effector cytokines, most of the key functional immunoregulatory molecules expressed by Treg cells were expressed to some degree by all of the Treg cell populations. Thus, the phenotypic diversity observed among Treg cell populations may function primarily to direct Treg cells with the appropriate specificities to sites of Th1-, Th2-, Th17-, or Th22-mediated inflammatory responses (7).

In addition to the core Th and Treg signatures we defined, we also identified smaller subset-specific signatures by comparing all of the Th or Treg populations to each other. In this case, we found that these subset-specific signatures were largely shared between the phenotypically similar Th and Treg populations and, accordingly, these signatures encompass genes known to be direct targets of the lineage-defining transcription factors shared by these populations. The phenotypic and functional diversity of Th cells has been extensively studied. Although initially thought of as terminally differentiated subsets, it has become increasingly clear that there is significant plasticity among different Th populations (25). In particular, there is substantial plasticity in Th17 cells, which can adopt a hybrid Th1/17 phenotype in response to inflammatory cytokines such as IL-12 or IL-1β (5, 26). This trans-differentiation is most relevant in the context of autoimmune and inflammatory diseases such as multiple sclerosis and Crohn disease in which hybrid populations and/or “ex-Th17” cells have been implicated in pathogenesis (27, 28). Indeed, our transcriptional analysis of Th1/17 cells showed that they are hybrid population that shares the transcriptional signature of both Th1 and Th17 cells. This includes not only shared expression of RORC and TFBX21, but also the highest levels of the cytokine receptors IL23R and IL12RB2 and some unique genes such as the multidrug resistance transporter ABCB1 and the inhibitory receptor KLRB1 (CD161), similarly, Treg1/7 cells transcriptionally resembled both the Treg1 and Treg7 populations, indicating that there may be significant phenotypic plasticity among these populations as well.

URL: http://www.immunohorizons.org/
The identification and isolation of distinct Th and Treg populations based on differential chemokine receptor expression works well in CD4+ T cells isolated from healthy human blood. However, similar analyses from nonlymphoid tissues can be challenging due to changes in phenotype that occur upon cellular entry into certain tissue sites, cleavage of specific markers during enzymatic digestion required for isolation of T cells from some nonlymphoid tissues, and the small number of T cells obtained from clinical samples and tissue biopsies. Our identification of unique subset-specific transcriptional signatures provides the opportunity to reanalyze bulk RNA-seq data of tissue Th and Treg cells to determine if any of these signatures are enriched. We validate the usefulness of this approach in three independent data sets derived from cytokine producing Treg cells from peripheral blood, Th and Treg cells infiltrating breast carcinomas, and Th cells found in the skin. Thus, these signatures can be used to guide analyses of both bulk and single-cell RNA-seq data to assess Th and Treg cell specialization. This will be particularly useful in analyses of small clinical samples that are limiting for conventional flow cytometric or functional analyses.

Our comprehensive analysis of human CD4+ T cells provides a new framework for understanding the transcriptional basis of Treg and Th specialization, and an important resource for analysis of transcriptomic data from T cells in healthy and diseased tissues. Cross-referencing these transcriptional profiles with analyses of epigenetic modification and transcription factor binding in different Treg and Th populations will help to further define the molecular mechanisms that underlie the diversity of CD4+ T cells, and suggest new ways to manipulate specific pathways to tune Th or Treg responses in cancer, autoimmunity, and chronic infection.

DISCLOSURES

The authors have no financial conflicts of interest.

REFERENCES

5. Duhen, T., and D. J. Campbell. 2014. IL-1β promotes the differentiation of polyfunctional human CCR6+CXCR3+ Th17 cells that are specific for pathogenic and commensal microbes. J. Immunol. 193: 120–129.

https://doi.org/10.4049/immunohorizons.2000037