e-*Blood*

IMMUNOBIOLOGY

The enhancer and promoter landscape of human regulatory and conventional T-cell subpopulations

Christian Schmidl,[1] Leo Hansmann,[1] Timo Lassmann,[2,3] Piotr J. Balwierz,[4] Hideya Kawaji,[2,3,5] Masayoshi Itoh,[2,3,5] Jun Kawai,[2,5] Sayaka Nagao-Sato,[2] Harukazu Suzuki,[2,3] Reinhard Andreesen,[1,6] Yoshihide Hayashizaki,[2,5] Alistair R. R. Forrest,[2,3] Piero Carninci,[2,3] Petra Hoffmann,[1,6] Matthias Edinger,[1,6] and Michael Rehli,[1,6] for the FANTOM consortium

[1]Department of Internal Medicine III, University Hospital Regensburg, Regensburg, Germany; [2]RIKEN Omics Science Center, Yokohama, Kanagawa, Japan*; [3]RIKEN Center for Life Science Technologies, Division of Genomic Technologies, Yokohama, Kanagawa, Japan; [4]Biozentrum, University of Basel, and Swiss Institute of Bioinformatics, Basel, Switzerland; [5]RIKEN Preventive Medicine and Diagnosis Innovation Program, Wako, Saitama, Japan; and [6]Regensburg Centre for Interventional Immunology, Regensburg, Germany

Key Points

- Transcription and enhancer profiling reveal cell type–specific regulome architectures and transcription factor networks in conventional and regulatory T cells.

CD4$^+$CD25$^+$FOXP3$^+$ human regulatory T cells (Tregs) are essential for self-tolerance and immune homeostasis. Here, we describe the promoterome of CD4$^+$CD25^{high}CD45RA$^+$ naïve and CD4$^+$CD25^{high}CD45RA$^-$ memory Tregs and their CD25$^-$ conventional T-cell (Tconv) counterparts both before and after in vitro expansion by cap analysis of gene expression (CAGE) adapted to single-molecule sequencing (HeliScopeCAGE). We performed comprehensive comparative digital gene expression analyses and revealed novel transcription start sites, of which several were validated as alternative promoters of known genes. For all in vitro expanded subsets, we additionally generated global maps of poised and active enhancer elements marked by histone H3 lysine 4 monomethylation and histone H3 lysine 27 acetylation, describe their cell type–specific motif signatures, and evaluate the role of candidate transcription factors STAT5, FOXP3, RUNX1, and ETS1 in both Treg- and Tconv-specific enhancer architectures. Network analyses of gene expression data revealed additional candidate transcription factors contributing to cell type specificity and a transcription factor network in Tregs that is dominated by FOXP3 interaction partners and targets. In summary, we provide a comprehensive and easily accessible resource of gene expression and gene regulation in human Treg and Tconv subpopulations. (*Blood*. 2014;123(17):e68-e78)

Introduction

Thymus-derived CD4$^+$CD25$^+$ regulatory T cells (Tregs) are crucial for the maintenance of peripheral self-tolerance and immune homeostasis by suppressing a wide variety of immune responses.[1] Expression of the transcription factor FOXP3 is indispensable for this suppressive activity because loss-of-function mutations in the gene cause lethal autoimmune diseases in mice and humans.[2,3] Their ability to avert unwanted immune reactions after adoptive transfer in a number of different settings, including autoimmune diseases and allogeneic hematopoietic stem cell or organ transplantation,[4-8] also renders Tregs highly interesting for clinical application. A pivotal prerequisite for their clinical use is the possibility to isolate and expand in vitro a Treg population with stable phenotypic and functional characteristics. In this respect, several groups showed that Tregs can be reliably identified by their low expression level of the interleukin-7R alpha chain (IL-7Rα; CD127).[9,10] However, we and others showed that this Treg population is heterogeneous with respect to several developmental markers, including CD45RA, and that only CD45RA$^+$ naïve Tregs retain high and homogenous FOXP3 expression and remain fully functional during in vitro expansion.[11-14] In contrast, CD45RA$^-$ Tregs partially downregulate FOXP3 during in

vitro culture, which correlates with a loss of suppressive activity and a tendency to differentiate into T cells with a proinflammatory phenotype.[13,15,16] These data raise questions about Treg stability, plasticity, and inherited subset properties and, in view of clinical applications, demand an in-depth molecular characterization of regulatory mechanisms. Previous genome-wide studies on human Tregs and Tconvs described differential FOXP3 binding in activated Tregs and Tconvs and its impact on gene expression.[17] However, most of the analyses of transcription factors (TFs) regulating Treg function were performed only in rodents or restricted to single loci in human cells, and comprehensive genome-wide studies have not yet been done. To better understand regulatory networks in human CD4$^+$ T-cell subpopulations, we applied cap analysis of gene expression (CAGE) adapted to single-molecule sequencing (HeliScopeCAGE) to CD4$^+$CD25^{high}CD45RA$^+$ naïve and CD4$^+$CD25^{high}CD45RA$^-$ memory Tregs and their CD25$^-$ conventional T cell (Tconv) counterparts both before and after in vitro expansion. In in vitro expanded cells, we extended the promoter analyses by mapping poised and active enhancers (marked by histone H3 lysine 4 monomethylation (H3K4me1) and histone H3 lysine 27 acetylation (H3K27ac), respectively,

There is an Inside *Blood* Commentary on this article in this issue.

throughout the genome. Enhancers show cell type–specific enrichment of TF binding motifs, and we highlight the global role of FOXP3, RUNX1, ETS1, and STAT5 in cell type–specific enhancer architecture and gene regulation by chromatin immunoprecipitation sequencing (ChIP-seq). The integrated analysis of promoter and enhancer activities identifies subset-specific gene regulation networks and provides a valuable resource on gene expression and regulation in Treg and Tconv subsets. This work is part of the Functional Annotation of Mammalian Genome 5 (FANTOM5) project. Data downloads, genomic tools, and copublished manuscripts are summarized at http://fantom.gsc.riken.jp/5/.

Methods

Cells

CD4[+]CD25[−]CD45RA[+], CD4[+]CD25[−]CD45RA[−], CD4[+]CD25[high]CD45RA[+], and CD4[+]CD25[high]CD45RA[−] T cells were isolated[13] and in vitro expanded as previously described.[18] Jurkat cells (human T-cell leukemia) were cultured as previously described.[19] Collection of blood cells from healthy donors was performed in compliance with the Declaration of Helsinki. All donors signed informed consent. The procedure was approved by the local ethical committee (reference numbers 92-1782 and 09/066b).

RNA preparation

RNA for HeliScopeCAGE and rapid amplification of cDNA ends-polymerase chain reaction (RACE-PCR) was isolated by using the miRNeasy RNA isolation kit (Qiagen, Hilden, Germany).

HeliScopeCAGE sequencing and data analysis

HeliScopeCAGE sequencing and sequence alignment were performed as part of the FANTOM5 project.[20] Normalization of individual tag libraries was done by using the common power-law distribution approach.[21] Expression data for annotated coding or noncoding genes (according to GencodeV10) were extracted by collecting normalized tag counts in regions −500 to +200 relative to all annotated transcription start sites (TSSs) associated with a single gene ID. Digital gene expression (DGE) analysis of normalized data was performed by using edgeR[22] as further outlined in the supplemental Methods (available on the *Blood* Web site).

3′ and 5′RACE-PCR

Complementary DNA (cDNA) from the RNA of T-cell subpopulations was generated with the SMARTer RACE cDNA Amplification Kit (Clontech, Saint-Germain-en-Laye, France) according to the manufacturers' instructions. RACE was performed by using a gene-specific primer with the samples specified in Figure 4 and supplemental Figure 3. When no distinct fragment sizes were observed, the PCR product was diluted and amplified with a nested gene-specific primer, and single PCR bands were directly sequenced. Primer sequences are listed in supplemental Table 1.

ChIP-seq and data analysis

ChIP and library construction were done essentially as described[23] by using antibodies against H3K4me1 (Abcam), H3K27ac (Abcam), STAT5 (Santa Cruz), ETS1 (Santa Cruz), RUNX1 (Abcam), and FOXP3 (Novus). Sequence tags were mapped to the human reference genome (GRCh37/hg19) by using Bowtie software.[24] Downstream analysis of uniquely mapped tags, including quality control, peak calling, and motif analysis, was performed as described by using Hypergeometric Optimization of Motif EnRichment (HOMER) software tools.[25] ChIP-seq data are deposited at the National Center for Biotechnology Information (NCBI) Gene Expression Omnibus (GEO) database (GSE43119). Clonal tags in TF ChIP-seq experiments were removed as described before.[26] A summary of all experiments can be found in supplemental Table 2, and a more detailed description of computational analyses is provided in the supplemental Methods. A University of California Santa Cruz
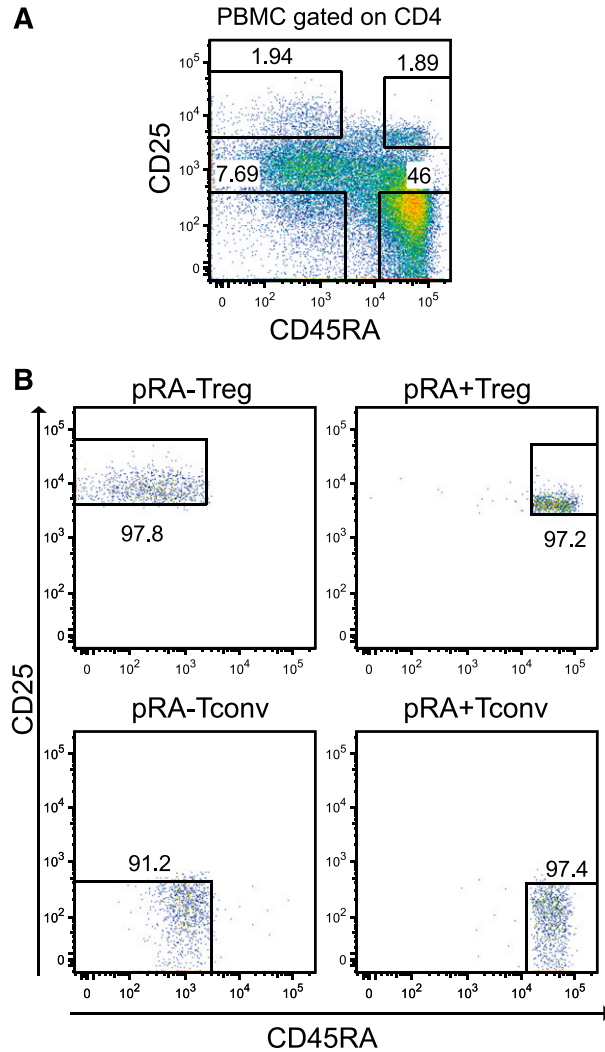
Figure 1. T-cell isolation and expansion. (A) Sorting strategy for CD4[+]CD25[high]CD45RA[+] (pRA+Treg), CD4[+]CD25[high]CD45RA[−] (pRA–Treg), CD4[+]CD25[−]CD45RA[+] (pRA+Tconv) and CD4[+]CD25[−]CD45RA[−] (pRA–Tconv) from human peripheral blood mononuclear cells (PBMCs) as described in "Methods". (B) Cells were reanalyzed after sorting on the FACSAria.

(UCSC) Genome Browser track hub of the entire data set can be found at http://www.ag-rehli.de/NGSdata.htm.

De novo motif analyses

Enriched sequence motifs were de novo extracted from regions surrounding differentially expressed CAGE clusters determined by the edgeR bioconductor software package ($P \leq .01$ for pairwise comparisons; −300 to +50 bp from cluster center) using HOMER. For ChIP-seq data sets, enhancers (distal H3K4me1 and H3K27ac regions, defined as being located at least 1000 bp away from GencodeV10 annotated TSSs) were extracted by using a fixed region size of 1 kb for replicate samples (and with a twofold tag enrichment in one sample in comparison with the other in case of cell type–specific enhancers). Motifs were extracted from 1-kb enhancer regions by using HOMER. A more detailed description of motif analyses is provided in the supplemental Methods.

Network analysis

We inferred regulatory inputs of genes differentially expressed in the T-cell subsets by applying a feature selection approach similar to the one used in the software program Genie3[27] as outlined in the supplemental Methods. The networks were constructed from the top 50 TFs defined by the number of target
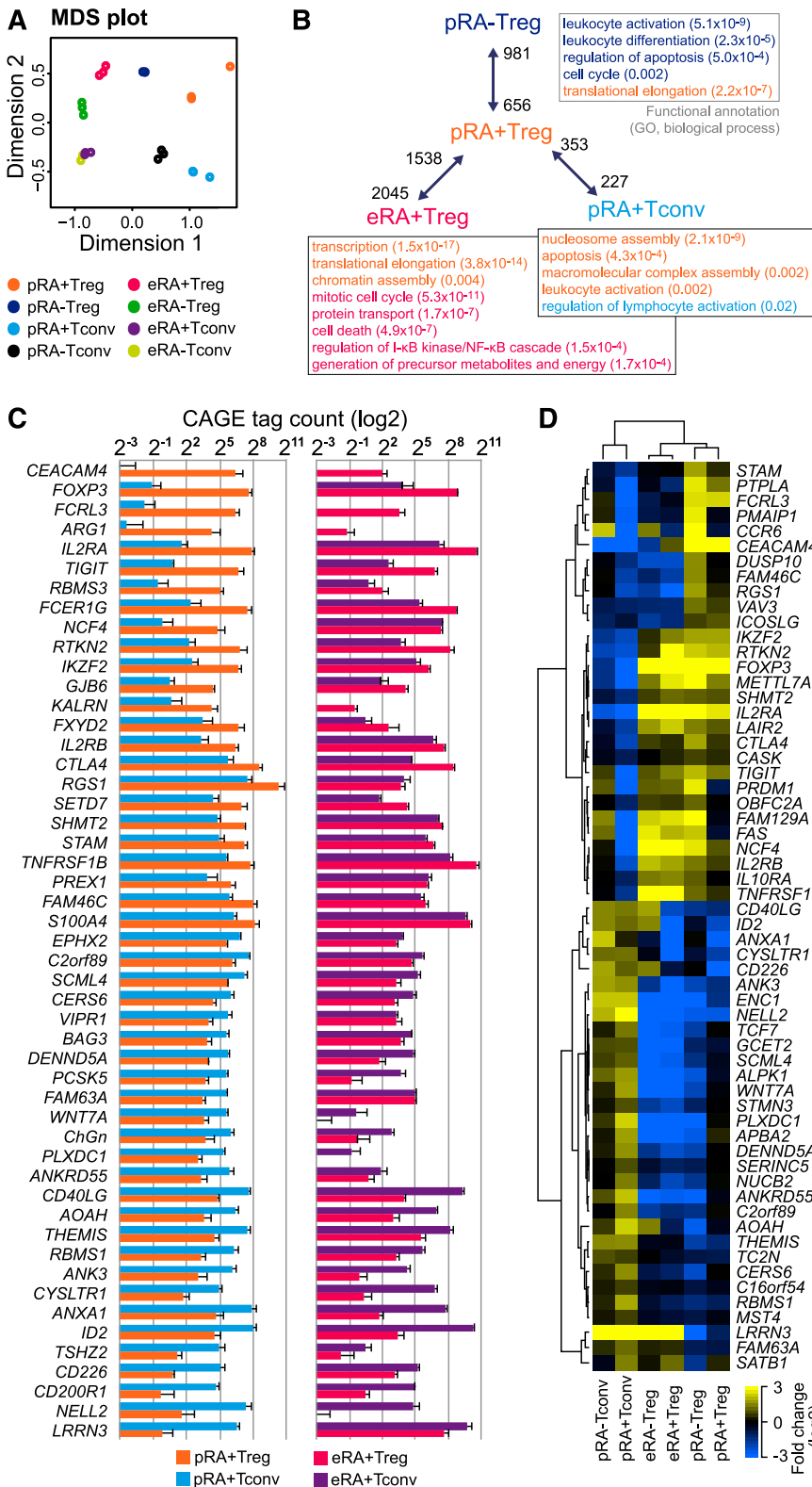
Figure 2. HeliScopeCAGE-based DGE analysis. (A) A multidimensional scaling (MDS) plot for replicate HeliScopeCAGE-based DGE data shows distance of samples based on tag distribution in expressed genes. (B) Differential gene expression between pRA+Treg and the indicated T-cell types. Numbers indicate the number of significantly upregulated genes in either cell type (5% false discovery rate cutoff) for each pairwise comparison. Boxes contain generic gene ontology (GO) terms (color coding refers to the corresponding cell type, numbers in parentheses correspond to enrichment q values) enriched in each differential gene set. A complete list of all expressed genes and corrected q values for all relevant pairwise comparisons is provided in supplemental Table 3. (C) DGE data for the top 50 significantly differentially expressed genes in a pairwise comparison of pRA+Treg vs pRA+Tconv. Shown are data for both primary and expanded RA+ cells. (D) A Treg core signature displaying unsupervised hierarchical clustering of genes differing highly significantly in expression between pTreg and pTconv in both CD45RA⁺ naïve and CD45RA⁻ memory subpopulations. eTreg were included in the clustering to visualize expression changes of core Treg genes after in vitro expansion. Values were log₂ transformed and normalized to the geometric mean of tag counts in pTreg and pTconv subpopulations for every gene.

genes in each T-cell subset (listed in supplemental Table 3) by using the Search Tool for the Retrieval of Interacting Genes (STRING) 9.0 database.[28]

Reporter plasmid construction and purification

The native as well as new CTLA4 and FOXP3 TSSs were amplified from genomic DNA by using PCR primers listed in supplemental Table 1. The PCR

fragments were cloned in the pGL4.10 vector (Promega). Plasmids were isolated and purified by using the EndoFree Plasmid Kit (Qiagen).

Transient DNA transfection

Jurkat cells were transfected and treated with phorbolmyristate acetate and ionomycin as described elsewhere.[19] Firefly luciferase activity values were

normalized against Renilla activity and correspond to at least 3 independent experiments measured in duplicates.

Results

To analyze gene expression and the exact promoter locations of T-cell subpopulations, we subjected 3 biologic replicates of highly purified primary (prefix "p") CD4$^+$CD25^{high}CD45RA$^+$ naïve Tregs (pRA+Treg), CD4$^+$CD25^{high}CD45RA$^-$ memory Tregs (pRA–Treg), CD4$^+$CD25$^-$CD45RA$^+$ naïve Tconvs (pRA+Tconv), and CD4$^+$CD25$^-$CD45RA$^-$ memory Tconvs (pRA–Tconv) to HeliScopeCAGE. Additionally, we in vitro expanded all subpopulations as previously described[13] and also subjected them to HeliScopeCAGE sequencing (prefix "e"). The sorting strategy with representative fluorescence-activated cell sorter plots is charted in Figure 1.

Overall, 11 022 protein-coding and 1168 noncoding genes were expressed with at least 1 tag per million (see "Methods") in 1 cell type. We performed DGE analysis of protein-coding genes by using edgeR[22] for pairwise comparisons between subpopulations. A multidimensional scaling plot clustered replicates together and clearly separated Tregs from Tconvs in one dimension and in vitro expanded from primary cells in the other dimension (Figure 2A). Results of pairwise comparisons for pRA+Treg (along with functional annotation of differentially regulated genes) are summarized in Figure 2B. Figure 2C shows a representative DGE analysis with the top 50 significantly differentially expressed protein-coding genes between pRA+Treg and pRA+Tconv. As expected, well-characterized signature genes appeared in the top ranked list, for example, *FOXP3*, *IL2RA*, *CTLA4*, *IKZF2*, and *CD40LG*. Plots of top-regulated genes of additional pairwise comparisons are displayed in supplemental Figure 1, and the complete DGE data set, including the statistical evaluation of differentially expressed protein coding genes, is available in supplemental Table 3. We next defined a core pTreg gene signature by identifying a set of genes that was statistically significantly differentially expressed in both pRA+Treg and pRA+Tconv and in pRA–Treg vs pRA–Tconv comparisons. In addition, we compared the expression levels of these 61 core genes in pTreg to those in eTreg (Figure 2D). This Treg core signature comprises intensively studied Treg markers as well as genes only recently discovered to be essential to foster or restrict Treg function such as *THEMIS* and *SATB1*.[29,30] In addition, the list also includes several genes less well described in the Treg context such as *LAIR2*, *METTL7A*, and *RTKN2* as being upregulated and *TCF7* (*TCF-1*), *ANK3*, *NELL2*, and *ANXA1* as being downregulated in pTreg and eTreg. Interestingly, an alternative transcript isoform corresponding to CAGE promoter 4 of *RTKN2* (p4@RTKN2), a gene that was previously reported to be expressed in lymphocytes,[31] showed exclusive expression in Tregs when compared with all other samples of the FANTOM5 expression atlas.[20]

Since Tregs are intensively studied for future clinical applications, we were particularly interested in the differential expression of Treg-specific effector molecules. The genes *CTLA4*, *IL2RA* (CD25), *TGFB1*, *TIGIT*, and *TNFRSF10B* were expressed in all Treg populations (Figure 3A). In contrast, several genes were not expressed in pRA+Treg but were upregulated in pRA–Treg, namely *TNFRSF18* (GITR), *LAG3*, *GZMA*, *IL10*, *FGL2*, and *ENTPD1* (CD39). Interestingly, eRA+Treg resembled pRA+Treg in many respects except for their higher expression of *GITR*, *LGAS1*, and *IL2RA*. Only a few genes were expressed exclusively in eRA–Treg, namely
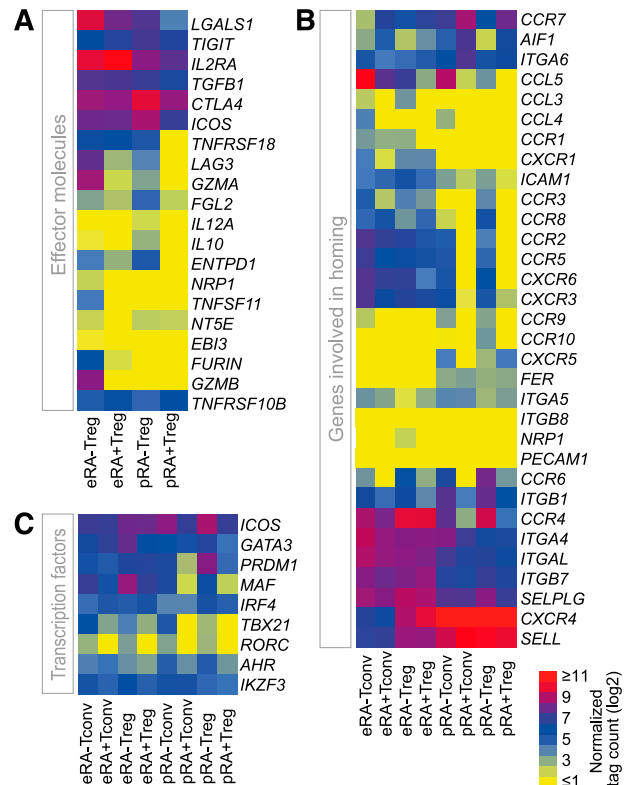
Figure 3. Effector molecule, homing receptor, and TF expression in T-cell subpopulations. Normalized gene expression data (log$_2$ transformed) of (A) Treg effector molecules, (B) genes involved in homing, and (C) TFs. Data are presented as a heatmap with yellow, blue, and red representing low, intermediate, and high expression, respectively.

TNFSF11, *NRP1*, *EBI3* (IL-35 subunit), and *GZMB*. However, since this population shows heterogeneity,[15,16] unequivocal identification of the cells expressing these effector genes is not possible. Another crucial factor in adoptive T-cell therapy is the potential ability of the cells to home to specific locations in the host.[32] Transcripts encoding CCR2, CCR5, and CCR8 that mediate migration to inflamed tissues as well as the skin-, mucosa-, liver- and intestine-homing receptors CCR10, CXCR6, and CCR9 were not expressed in pRA+Treg, but were present in pRA–Treg (Figure 3B). Notably, eRA+Treg resembled pRA–Treg cells more than pRA+Treg with regard to their CCR expression with the exception of *CCR6*, *CCR9*, and *CCR10*. In contrast to pTreg populations, eTregs do not express *CXCR5*, encoding a receptor described for homing to B-cell follicles and germinal centers. Recent publications suggest that Tregs express specific TFs of other T-cell lineages that drive specialized gene expression programs in order to suppress the corresponding T helper (Th) cell–associated inflammation.[33-35] We therefore investigated whether such TFs are already expressed in pTregs and whether their expression changes upon in vitro expansion. In contrast to pRA+Treg, pRA–Treg expressed TFs of other T-cell lineages, albeit at low levels, as shown in Figure 3C exemplarily for *MAF*, *TBX21*, and *RORC*. The Th2 TF genes *AHR*, *PRDM1*, and *GATA3* were expressed in both Treg populations but to a higher degree in pRA–Tregs. Of note, upon in vitro expansion, eRA+Tregs upregulate the Th1 and Th2 TF genes *TBX21* and *MAF*, respectively. However, the highest expression of Th2-associated TFs was observed in eRA–Tregs.

Finally, DGE was also performed for noncoding genes. Among the most significantly upregulated transcripts in RA+Tregs compared
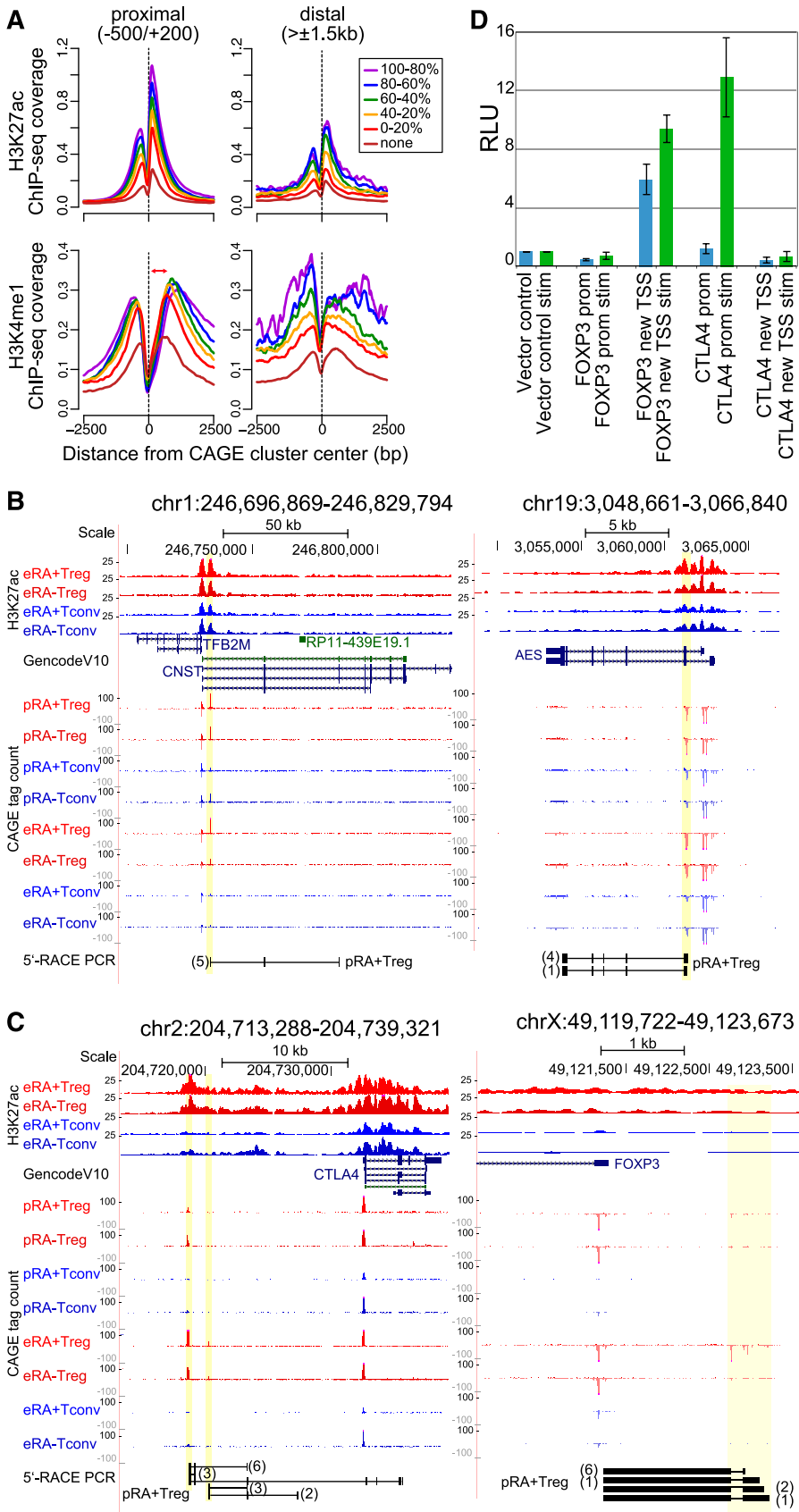
Figure 4. Novel CAGE clusters. (A) Histograms for genomic distance distributions of H3K27ac (top panels) and H3K4me1 tag counts (bottom panels) are shown for eRA+Treg centered across expression-binned CAGE clusters across a 5-kb genomic interval contingent on their location relative to annotated TSSs. CAGE clusters represent a merge of all T-cell subpopulations, and clusters with no associated CAGE tags in Tregs (none) were separated from the expressed ones, which were ranked according to CAGE cluster activity and binned into 5 equal-sized groups (quintiles). Expression bins are color coded as indicated. The average H3K27ac deposition peaks at the nucleosome immediately downstream of the CAGE cluster and shows a dip just upstream of CAGE clusters, indicating a nucleosome-free region. The average H3K4me1 deposition is broader. Marking of the nucleosome immediately downstream of the CAGE cluster decreases with increased TSS activity (as indicated by the red arrow), reflecting the higher degree of H3K4 methylation (trimethylation; me3) associated with highly active promoters. Of note, CAGE clusters distal from annotated promoters do show the highest H3K4me1 deposition at the nucleosome immediately downstream of the TSS, suggesting that many of them may represent enhancers. (B-C) 5′-RACE confirms the presence of spliced transcripts from novel CAGE TSSs. UCSC browser graphics are shown for the indicated genomic positions, including H3K27ac signal of expanded populations, GencodeV10 gene annotation, and CAGE signals for all 8 T-cell subpopulations. In the bottom of each panel, results from 5′-RACE-PCR of the indicated cell types are aligned (only products representing novel TSSs are shown). Numbers of sequenced clones are indicated in brackets. (D) Relative luciferase activity (RLU) of promoter construct for the annotated (prom) and novel FOXP3 and CTLA4 TSSs (new TSS) in unstimulated and phorbolmyristate acetate/ionomycin-stimulated (stim) Jurkat T cells.

with RA+Tconv populations is *CTC-231011.1*, the host transcript for mir146a, a micro RNA that is involved in Treg-mediated control of Th1 responses in the murine system[36] (supplemental Figure 2). In addition, we identified several uncharacterized noncoding genes that are differentially expressed between subsets and might be involved in subset-specific function (summarized in supplemental Table 4).

CAGE clusters not annotated to known promoters could represent promoters of new genes, alternative promoters of known genes, or TSSs of enhancer RNAs.[37-39] As shown in Figure 4A, the average distribution of H3K27ac and H3K4me1 (two histone modifications that demarcate open chromatin around promoters and enhancers) relative to CAGE-cluster centers close to known T-cell–expressed promoters correlated with CAGE expression levels. Interestingly, nonannotated clusters (distal to GencodeV10 promoters) displayed a similar correlation, which strongly indicates functionality for a large fraction of genomic elements associated with such CAGE peaks. The different distribution of H3K4me1 signals downstream of CAGE TSSs suggests that distal nonannotated clusters may contain a large fraction of transcribed enhancers (Figure 4A). To highlight the potency of HeliScopeCAGE to detect novel promoters in T cells, we performed 5′-RACE-PCR for several selected samples in which annotated promoter and CAGE TSSs differed (Figure 4B and supplemental Figure 3A). In addition to the Treg-specific TSS for RTKN2 mentioned above, we found a second Treg-exclusive nonannotated upstream TSS producing a spliced RNA (supplemental Figure 3B). Interestingly, novel CAGE TSSs were also found at the well-studied Treg signature genes *CTLA4* and *FOXP3* (Figure 4C). The Treg-specific TSS upstream of *CTLA4* yielded inter alia transcripts that extended into the annotated *CTLA4* gene and could potentially encode for a novel CTLA4 isoform. Additional evidence from 3′- and 5′-RACE-PCR of the *CTLA4* upstream TSSs are displayed in supplemental Figure 3C. At the *FOXP3* locus, a conserved cluster of additional Treg-specific TSSs was found approximately 1 kb upstream of the annotated *FOXP3* promoter. 5′-RACE from the native *FOXP3* promoter/5′-untranslated region confirmed spliced transcripts extending to this novel upstream TSS cluster. Reporter assays using upstream sequences of the alternative *FOXP3* TSS but not the novel *CTLA4* TSS showed high luciferase activity when transfected in Jurkat cells, suggesting general activity of the newly discovered *FOXP3* TSS with the capacity to increase transcription after stimulation (Figure 4D). In summary, these results demonstrate that nonannotated CAGE clusters in T cells can indeed represent functional promoters. Intriguingly, we also identified TSSs with Treg-exclusive expression as well as new TSSs in proximity to well-studied key Treg genes.

Recent studies demonstrated the possibility of identifying regulator TFs by epigenetic "fingerprinting" of cell type–specific enhancers.[23] Hence, we initially analyzed active enhancer candidates characterized by promoter-distal enrichment of H3K27ac in eRA+Treg and expanded CD4⁺CD25⁻ Tconv (eTconv), in which the freshly sorted cells naturally mainly contain naïve T cells (>90%). This identified 6822 and 7112 promoter-distal and putative enhancer regions, respectively. De novo motif analysis of candidate enhancers yielded a broad panel of highly enriched DNA sequences frequently similar to known TF consensus motifs (supplemental Figure 4). In addition to analyzing complete enhancer sets, we also determined motif fingerprints in eRA+Treg- and eTconv-specific candidate enhancers (twofold difference in H3K27ac signal). eTconv candidate enhancers (4531 regions) were clearly dominated by an ETS, RUNX, and IRF motif signature, whereas eRA+Treg candidate enhancers (3962 regions)

were less enriched in RUNX motifs but showed specific enrichment for KLF, FOX, and STAT5 motifs (Figure 5A). Many TFs corresponding to the extracted de novo motifs were shown to play crucial roles in Treg development and function, but the global association of these factors with enhancer landscapes in human Tregs has not been studied before. To confirm the in silico-derived motif signatures, we generated TF-binding data by using ChIP-seq for the possible regulators ETS1, RUNX1, STAT5, and FOXP3 in eRA+Treg and eTconv. Figure 5B shows a representative 60-kb region of the *IL2RA* gene in which a large fraction of TF binding events is subset specific. The global distribution of binding sites is illustrated in supplemental Figure 5, and the proportion of subset-specific binding sites is provided in Figure 5C. We then evaluated the ChIP-seq signal strengths of TFs in the cell type–specific enhancers. First, we merged both eRA+Treg- and eTconv-derived peaks of the corresponding TF and overlapped the merged set with the respective cell type–specific enhancers. We then counted the TF ChIP-seq reads in the overlapping regions. In line with the observed over-representation of motifs in specific enhancers, we indeed observed an enrichment of STAT5 ChIP-seq signal in eRA+Treg-specific enhancers (Figure 5D). In contrast, ETS1 and RUNX1 ChIP-seq tags were both highly enriched in eTconv enhancers, which parallels their motif distribution. Interestingly, FOXP3 ChIP-seq signals were equally enriched in FOXP3 binding sites overlapping with eRA+Treg- and eTconv-specific enhancers. Taken together, we identified candidate enhancers in eRA+Treg and eTconv and could identify the key regulators ETS1, STAT5, RUNX1, and FOXP3 participating in cell type–specific enhancer landscapes based on their de novo motif signatures.

Next, we extended the histone profiling to characterize enhancer elements in all subpopulations. We generated additional maps of H3K4me1 and H3K27ac for eRA+Treg, eRA–Treg, eRA+Tconv, and eRA–Tconv from two independent donors that were also used for CAGE profiling. We then isolated distal cell type–specific regions for H3K4me1 and H3K27ac in pairwise comparisons and investigated their correlation with the expression of adjacent genes. As shown in the bubble plot representations in Figure 6A, subset-specific enhancers were significantly associated with higher tag counts in neighboring CAGE clusters of the same cell type ($P < .001$, Wilcoxon signed rank test), suggesting that these regions indeed represent good candidates for subset-specific enhancer. We detected similar correlations even in highly similar populations (eg, eRA+Treg vs eRA–Treg) (supplemental Figure 6), resembling observations for closely related monocyte subpopulations[40] and confirming the positive correlation of enhancer regions with gene expression. These data sets provide the most comprehensive candidate enhancer maps of human CD4⁺ T-cell subpopulations so far. To facilitate easy access to both HeliScopeCAGE and all ChIP-seq data sets, we generated a UCSC browser track hub that can be found at www.ag-rehli.de/NGSdata.htm.

Having determined subset-specific *cis*-regulatory regions from CAGE clusters as well as from H3K4me1/H3K27ac enhancers, we systematically determined their motif composition to deduce possible regulators. Different *cis*-regulatory elements (CAGE-TSSs, H3K4me1-poised or H3K27ac-active enhancers) showed similar but nonidentical motif compositions (supplemental Figure 7). We combined nonredundant sets of significantly enriched motifs for *cis*-regulatory regions of eRA+Treg versus eRA+Tconv subpopulations and matched these with differentially expressed candidate TFs (Figure 6B). We observed cell type–specific overrepresentation of JUN, STAT, PAX, NFE, KLF, and forkhead motifs in eRA+Treg and E2F, CREB, TCF, GTF, and helix-loop-helix (HLH) motifs in eRA+Tconv. Notably,

Figure 5. Cell type–specific enhancers. (A) Motif composition of cell type–specific active enhancers in eRA+Treg compared with eTconv. Shown are extracted de novo motifs, their hypergeometric *P* value, and the best matching known motif families (with the similarity score to the best matching known motif in parentheses). (B) An overview of the produced ChIP-seq data are illustrated by a UCSC genome browser graph (http://genome.ucsc.edu/) of a 60-kb region of the *IL2RA* promoter. (C) Venn diagrams for genome-wide TF binding events of ETS1, FOXP3, RUNX1, and STAT5 in eRA+Treg and eTconv. (D) ChIP-seq signal strength of STAT5, FOXP3, ETS1, and RUNX1 in eRA+Treg- and eTconv-specific active H3K27ac-marked enhancer candidates.
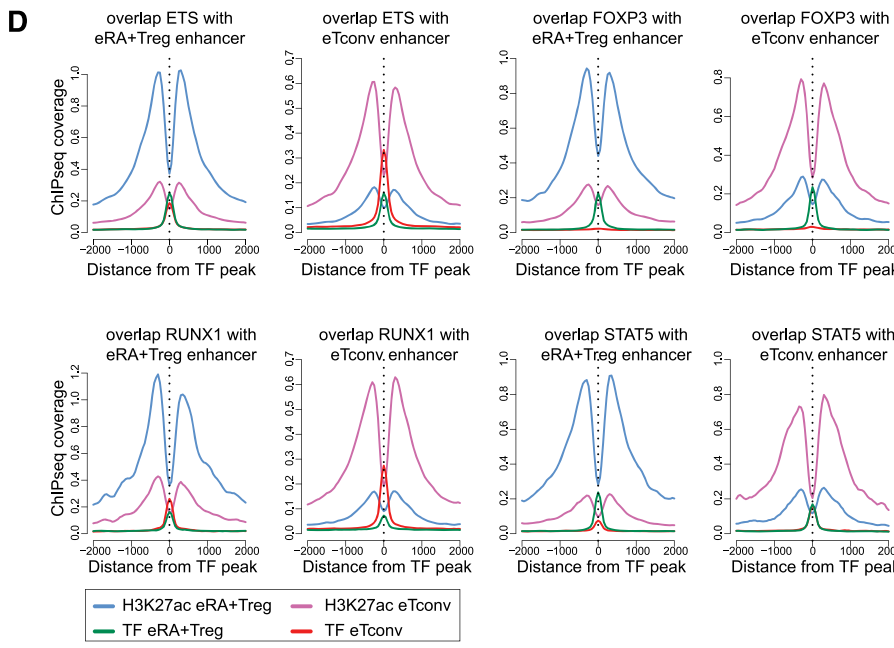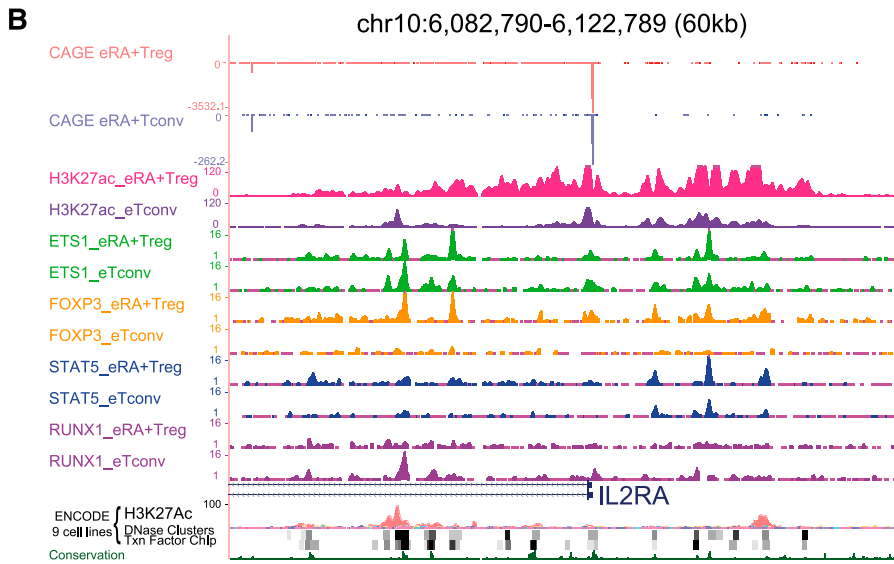
Figure 6. Correlation of cell type–specific enhancers to gene expression and potential regulators of T-cell subsets. (A) Bubble plot representations of CAGE TSS activity around eRA+Treg vs eRA+Tconv enhancer candidate regions showing at least twofold different H3K27ac or H3K4me1 signals. The bubble plots encode 3 quantitative parameters per CAGE cluster: distance from the putative enhancer, log10 fold change in CAGE cluster tag count between eRA+Treg and eRA+Tconv (y-axis), and the absolute CAGE cluster tag count of the T-cell subset with the highest expression level (bubble diameter). There is a clear bias for the putative enhancer elements to associate with CAGE clusters upregulated in the corresponding cell type ($P < .001$, Wilcoxon signed rank test). (B) A nonredundant combined set of de novo motifs from CAGE clusters and candidate enhancers in eRA+Treg vs eRA+Tconv. Shown are motifs with high similarity to known TF motifs. Absolute tag counts (log2 transformed) of differentially expressed TFs matching a de novo motif are presented as colored boxes with yellow, blue, and red representing low, intermediate, and high expression, respectively. (C) STRING-based networks (confidence views) of the top TFs predicted to have the strongest regulatory input in human Treg subsets compared with their Tconv counterparts (additional networks are shown in supplemental Figure 8). Only connected edges are included. Complete TF lists are provided in supplemental Table 5. Direct interaction partners or targets of FOXP3 are marked as indicated.
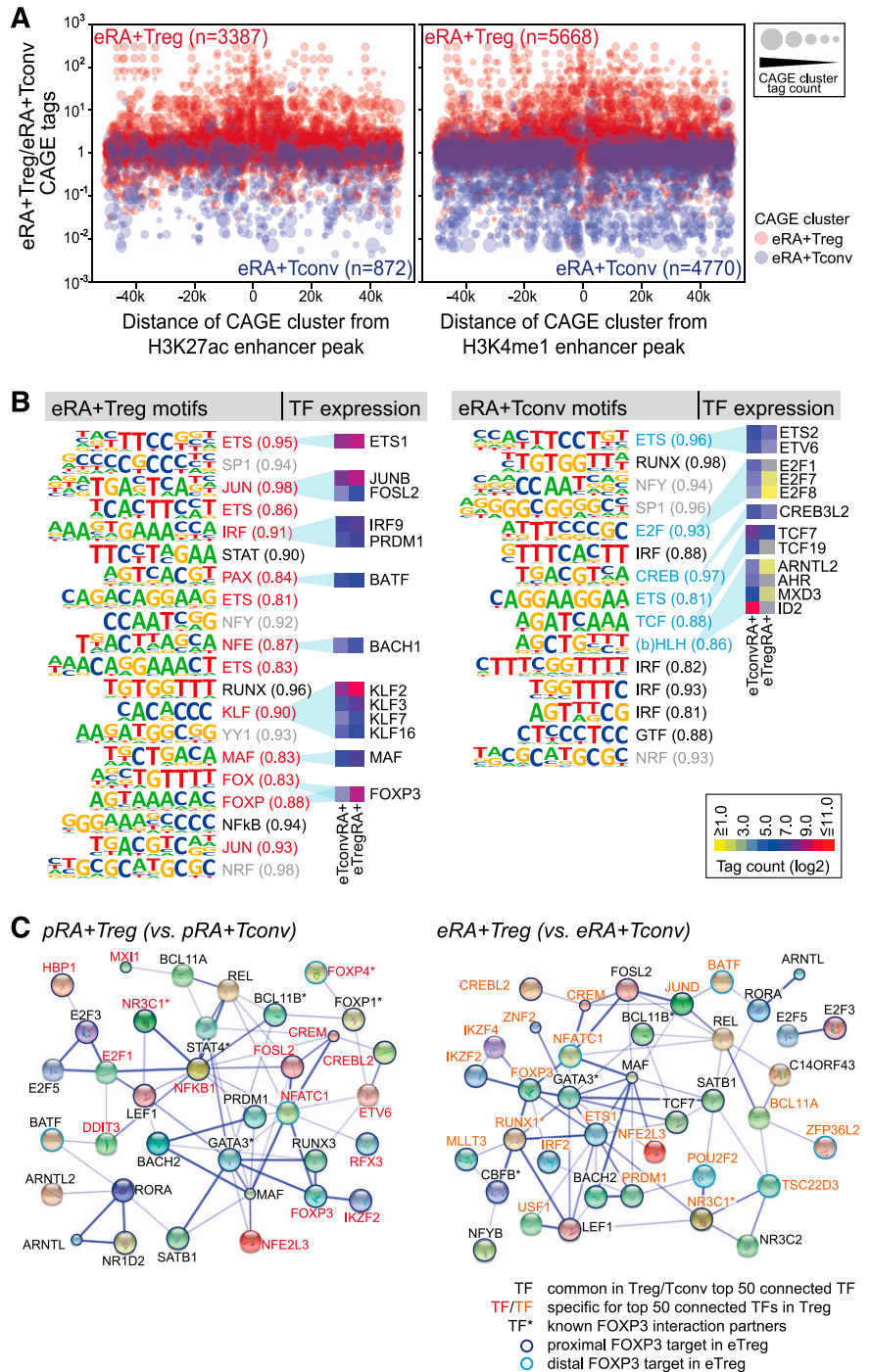
many specific signature motifs had a corresponding regulated TF candidate—JUNB/FOSL2 (JUN motif), BATF (PAX/ATF motif), BACH1 (NFE motif), KLF2/3/7/19 (KLF motif), MAF (MAF motif), FOXP3 (FOX and FOXP motif)—in the eRA+Treg set and E2F1/7/8 (E2F motif), CREB2L (CREB motif), TCF7/19 (TCF motif) and ARNTL/AHR/MXD3/ID2 (potentially binding bHLH motif) for the eRA+Tconv set of motifs. Notably, not all motifs are expected to correlate with differential expression of a TF since the activities of some TFs are regulated by messenger RNA stability, protein degradation, or posttranslational modifications such as acetylation or phosphorylation (eg, STAT and nuclear factor κB proteins). As an

alternative approach to identifying regulators that are most important in describing the expression pattern of subset-specific target genes, we used a regularized random forest-based regression analysis.[27] Here, TFs are selected on the basis of differential gene expression data that are predicted to be the most important in describing the differential expression pattern of target genes. Networks (based on the STRING database of known and predicted protein interactions)[28] of the top TFs relevant for T-cell subsets (relative to the indicated counterparts) are shown in Figure 6C and supplemental Figure 8 and highlight the presence of Treg- and Tconv-specific regulatory inputs. In line with previous work, the pTreg and eTreg networks included FOXP3 as well

as several direct FOXP3 interaction partners[41] or targets and also novel candidate regulators. In line with the stable expression of core Treg genes in pTreg and eTreg, the top regulators predicted to control differentially expressed genes in primary vs expanded cells do not include FOXP3 (supplemental Figure 8B).

Discussion

Using state-of-the-art sequencing technologies, this study provides insights into the regulatory landscapes of human T-cell subpopulations in an unprecedented depth. In addition to primary naïve and memory CD4[+] Tregs and Tconvs, we also characterized ex vivo expanded T cells, which represent clinically important cell products.

Previous studies suggested that human CD4[+]CD25[high] Tregs comprise a heterogeneous cell population, with CD45RA[−] memory Tregs expressing TFs and cytokines characteristic of nonregulatory proinflammatory T-cell lineages, a feature not observed in CD45RA[+] naïve Tregs.[12,13,15,16,42] In addition, several murine Treg populations displayed a specific gene expression profile ex vivo, depending on their tissue origin and homing receptor repertoire.[43,44] Prompted by these observations, we studied differences in the expression of effector molecules, homing receptors, and TFs between naïve and memory Treg subpopulations and in particular between primary and expanded cell populations. Notably, eRA+Treg, which retain a stable Treg phenotype even after extensive in vitro expansion,[13] do not dramatically change their Treg-specific effector molecule repertoire. Furthermore, TF networks controlling genes up- or downregulated by ex vivo culture do not involve the Treg master TF FOXP3. Expanded RA+Treg, however, alter the expression of several homing receptors, such as CCR2, CCR5, CCR8, CCR10, CXCR6, CCR9, or CXCR5. These changes are likely to influence not only the migratory properties of transferred eTreg but also their further developmental fate. The detailed CAGE expression data also allowed us to refine the core signature of human Tregs, which includes well-known key Treg genes and also novel candidate genes such as *LAIR2*, *METTL7A*, and *RTKN2* that share the unique Treg signature expression pattern but have a yet unknown role in Treg biology.

In line with recent comprehensive CAGE and RNA sequencing studies that identified tens of thousands of novel TSSs that represent promoters of functional protein coding and noncoding transcripts,[20,37] our data provide a rich source for discovering novel and/or alternative promoters, as exemplified by *CTLA4* and *FOXP3* in this study. Although their biologic significance needs to be evaluated, both alternative TSSs were expressed specifically in Tregs, which is in line with the previously observed Treg-specific hypomethylation of these regions.[19]

Even more regulatory input is thought to originate from enhancers, which display greater diversity than promoters. These regulatory elements are distributed across the genome in a cell type–specific manner and are designated by the deposition of H3K4me1 and H3K27ac.[45-47] By comparative analysis of H3K27ac and H3K4me1 patterns, we were able to identify numerous candidate enhancers in eTreg and eTconv. Computational analysis of enhancers allowed us to extract sequences that matched consensus-binding motifs of TFs known to be essential for Treg and Tconv function. In comparisons between eRA+Treg and eTconv, these included forkhead, ETS, STAT, IRF, JUN/AP1, KLF, and RUNX motifs. STAT5 was shown to bind and directly regulate expression of the *Foxp3* gene in mice, a finding that was also suggested for human

Tregs by indirect evidence.[48,49] In addition, ETS1 and RUNX proteins were also described to regulate Treg gene expression in mice and humans.[50-53] Validating the computational predictions, we observed increased binding of ETS1, RUNX1, and STAT5 to either eRA+Treg- or eTconv-specific enhancers in correspondence to their motif distribution. At least in ex vivo expanded cells, RUNX1 and ETS1 were more frequently associated with eTconv enhancers than eRA+Treg enhancers, which is in contrast to their established significance in Treg development and function. Interestingly, in eTregs, FOXP3 occupies eRA+Treg-specific enhancers and also potential enhancers in eTconv, raising the possibility that FOXP3 negatively regulates the expression of Tconv-associated genes through distal regulatory elements. This observation is in line with a recent report showing that FOXP3 in Tregs is recruited to Treg- and Tconv-shared preexisting DNase-accessible sites to exert its lineage-specifying functions.[26] In addition to candidate regulators derived from motif analyses, we also predicted subset-specific TF networks solely on the basis of DGE data. Notably, many TFs that are predicted to provide regulatory input in Tregs are either part of the FOXP3 protein complexes identified in a recent study,[41] or likely targets of FOXP3 with binding sites either at the proximal promoter or at distal elements, confirming the central role of this factor in specifying the Treg-specific gene expression program.

In summary, we generated the most comprehensive dataset so far on TSS location, gene expression, TF binding, and enhancer profiling of human CD4[+] T-cell subpopulations. These integrated genome-wide data provide a unique resource for studies on human T-cell differentiation and, in particular, for advancing the use of clinical T-cell products. As a whole, these data might prove helpful in elucidating drug-specific mechanisms and effects, such as those that modulate signaling pathways or the epigenetic status of T cells. For instance, with global maps of histone acetylation as presented here, the impact of different histone deacetylase inhibitors on the acetylation status of immunologically relevant genes in Tregs and Tconvs can be studied. This will help improve in vitro expansion strategies and, hence, the therapeutic potential of human Treg products.

Authorship

Contribution: C.S. performed experiments and computational analyses and wrote the manuscript; L.H. performed experiments; P.H., R.A., and M.E. contributed to planning, supervision, and manuscript writing; P.J.B. performed computational analyses; M.I. and S.N.-S. were responsible for CAGE data production; T.L. was responsible for tag mapping and performed network analyses; H.K. and J.K. managed the data handling; P.C., H.S., Y.H., and A.R.R.F. were responsible for FANTOM5 management and concept; and M.R. initiated, planned, and supervised the study and contributed to manuscript writing.

Conflict-of-interest disclosure: The authors declare no competing financial interests.

The current affiliation for C.S. is Research Center for Molecular Medicine of the Austrian Academy of Sciences, Vienna, Austria.

The current affiliation for L.H. is Department of Microbiology and Immunology, Stanford School of Medicine, Stanford, CA.

Correspondence: Michael Rehli, Department of Internal Medicine III, University Hospital Regensburg, D-93042 Regensburg, Germany; e-mail: michael.rehli@ukr.de.

References

1. Sakaguchi S. Naturally arising CD4+ regulatory t cells for immunologic self-tolerance and negative control of immune responses. *Annu Rev Immunol.* 2004;22:531-562.

2. Brunkow ME, Jeffery EW, Hjerrild KA, et al. Disruption of a new forkhead/winged-helix protein, scurfin, results in the fatal lymphoproliferative disorder of the scurfy mouse. *Nat Genet.* 2001; 27(1):68-73.

3. Bennett CL, Christie J, Ramsdell F, et al. The immune dysregulation, polyendocrinopathy, enteropathy, X-linked syndrome (IPEX) is caused by mutations of FOXP3. *Nat Genet.* 2001;27(1): 20-21.

4. Edinger M, Hoffmann P, Ermann J, et al. CD4+ CD25+ regulatory T cells preserve graft-versus-tumor activity while inhibiting graft-versus-host disease after bone marrow transplantation. *Nat Med.* 2003;9(9):1144-1150.

5. Edinger M, Hoffmann P. Regulatory T cells in stem cell transplantation: strategies and first clinical experiences. *Curr Opin Immunol.* 2011; 23(5):679-684.

6. Hoffmann P, Ermann J, Edinger M, Fathman CG, Strober S. Donor-type CD4(+)CD25(+) regulatory T cells suppress lethal acute graft-versus-host disease after allogeneic bone marrow transplantation. *J Exp Med.* 2002;196(3):389-399.

7. Nadig SN, Wieckiewicz J, Wu DC, et al. In vivo prevention of transplant arteriosclerosis by ex vivo-expanded human regulatory T cells. *Nat Med.* 2010;16(7):809-813.

8. Brusko TM, Putnam AL, Bluestone JA. Human regulatory T cells: role in autoimmune disease and therapeutic opportunities. *Immunol Rev.* 2008; 223:371-390.

9. Seddiki N, Santner-Nanan B, Martinson J, et al. Expression of interleukin (IL)-2 and IL-7 receptors discriminates between human regulatory and activated T cells. *J Exp Med.* 2006;203(7): 1693-1700.

10. Liu W, Putnam AL, Xu-Yu Z, et al. CD127 expression inversely correlates with FoxP3 and suppressive function of human CD4+ T reg cells. *J Exp Med.* 2006;203(7):1701-1711.

11. Valmori D, Merlo A, Souleimanian NE, Hesdorffer CS, Ayyoub M. A peripheral circulating compartment of natural naive CD4 Tregs. *J Clin Invest.* 2005;115(7):1953-1962.

12. Miyara M, Yoshioka Y, Kitoh A, et al. Functional delineation and differentiation dynamics of human CD4+ T cells expressing the FoxP3 transcription factor. *Immunity.* 2009;30(6):899-911.

13. Hoffmann P, Eder R, Boeld TJ, et al. Only the CD45RA+ subpopulation of CD4+CD25high T cells gives rise to homogeneous regulatory T-cell lines upon in vitro expansion. *Blood.* 2006; 108(13):4260-4267.

14. Hoffmann P, Boeld TJ, Eder R, et al. Loss of FOXP3 expression in natural human CD4+ CD25+ regulatory T cells upon repetitive in vitro stimulation. *Eur J Immunol.* 2009;39(4): 1088-1097.

15. Hansmann L, Schmidl C, Kett J, et al. Dominant Th2 differentiation of human regulatory T cells upon loss of FOXP3 expression. *J Immunol.* 2012;188(3):1275-1282.

16. Schmidl C, Hansmann L, Andreesen R, Edinger M, Hoffmann P, Rehli M. Epigenetic reprogramming of the RORC locus during in vitro expansion is a distinctive feature of human memory but not naïve Treg. *Eur J Immunol.* 2011; 41(5):1491-1498.

17. Birzele F, Fauti T, Stahl H, et al. Next-generation insights into regulatory T cells: expression profiling and FoxP3 occupancy in Human. *Nucleic Acids Res.* 2011;39(18):7946-7960.

18. Hoffmann P, Eder R, Kunz-Schughart LA, Andreesen R, Edinger M. Large-scale in vitro expansion of polyclonal human CD4(+)CD25high regulatory T cells. *Blood.* 2004;104(3):895-903.

19. Schmidl C, Klug M, Boeld TJ, et al. Lineage-specific DNA methylation in T cells correlates with histone methylation and enhancer activity. *Genome Res.* 2009;19(7):1165-1174.

20. Forrest ARR, Kawaji H, Rehli M, et al. A promoter level mammalian expression atlas. *Nature.* 2014; doi:10.1038/nature13182.

21. Balwierz PJ, Carninci P, Daub CO, et al. Methods for analyzing deep sequencing expression data: constructing the human and mouse promoterome with deepCAGE data. *Genome Biol.* 2009;10(7): R79.

22. Robinson MD, McCarthy DJ, Smyth GK. edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics.* 2010;26(1):139-140.

23. Pham TH, Benner C, Lichtinger M, et al. Dynamic epigenetic enhancer signatures reveal key transcription factors associated with monocytic differentiation states. *Blood.* 2012;119(24): e161-e171.

24. Langmead B. *Aligning short sequencing reads with Bowtie. Curr Protoc Bioinform.* 2010;32: 11.7.1-11.7.14.

25. Heinz S, Benner C, Spann N, et al. Simple combinations of lineage-determining transcription factors prime cis-regulatory elements required for macrophage and B cell identities. *Mol Cell.* 2010; 38(4):576-589.

26. Samstein RM, Arvey A, Josefowicz SZ, et al. Foxp3 exploits a pre-existent enhancer landscape for regulatory T cell lineage specification. *Cell.* 2012;151(1):153-166.

27. Huynh-Thu VA, Irrthum A, Wehenkel L, Geurts P. Inferring regulatory networks from expression data using tree-based methods. *PLoS ONE.* 2010;5(9):e12776.

28. Szklarczyk D, Franceschini A, Kuhn M, et al. The STRING database in 2011: functional interaction networks of proteins, globally integrated and scored. *Nucleic Acids Res.* 2011;39(Database issue):D561-D568.

29. Chabod M, Pedros C, Lamouroux L, et al. A spontaneous mutation of the rat Themis gene leads to impaired function of regulatory T cells linked to inflammatory bowel disease. *PLoS Genet.* 2012;8(1):e1002461.

30. Beyer M, Thabet Y, Müller RU, 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.

31. Collier FM, Gregorio-King CC, Gough TJ, Talbot CD, Walder K, Kirkland MA. Identification and characterization of a lymphocytic Rho-GTPase effector: rhotekin-2. *Biochem Biophys Res Commun.* 2004;324(4):1360-1369.

32. Campbell DJ, Koch MA. Phenotypical and functional specialization of FOXP3+ regulatory T cells. *Nat Rev Immunol.* 2011; 11(2):119-130.

33. Chaudhry A, Rudra D, Treuting P, et al. CD4+ regulatory T cells control TH17 responses in a Stat3-dependent manner. *Science.* 2009; 326(5955):986-991.

34. 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.

35. Zheng Y, Chaudhry A, Kas A, et al. Regulatory T-cell suppressor program co-opts transcription factor IRF4 to control T(H)2 responses. *Nature.* 2009;458(7236):351-356.

36. Lu LF, Boldin MP, Chaudhry A, et al. Function of miR-146a in controlling Treg cell-mediated regulation of Th1 responses. *Cell.* 2010;142(6): 914-929.

37. Djebali S, Davis CA, Merkel A, et al. Landscape of transcription in human cells. *Nature.* 2012; 489(7414):101-108.

38. Kanamori-Katayama M, Itoh M, Kawaji H, et al. Unamplified cap analysis of gene expression on a single-molecule sequencer. *Genome Res.* 2011;21(7):1150-1159.

39. Andersson R, Gebhard C, Miguel-Escalada I, et al. An atlas of active enhancers across human cell types and tissues. *Nature.* 2014;doi:10.1038/ nature12787.

40. Schmidl C, Renner K, Peter K, et al. Transcription and enhancer profiling in human monocyte subsets. *Blood.* 2014;doi:10.1182/blood-2013-02-484188.

41. Rudra D, deRoos P, Chaudhry A, et al. Transcription factor Foxp3 and its protein partners form a complex regulatory network. *Nat Immunol.* 2012;13(10):1010-1019.

42. Ayyoub M, Deknuydt F, Raimbaud I, et al. Human memory FOXP3+ Tregs secrete IL-17 ex vivo and constitutively express the T(H)17 lineage-

specific transcription factor RORgamma t. *Proc Natl Acad Sci USA.* 2009;106(21): 8635-8640.

43. 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 USA.* 2010;107(13): 5919-5924.

44. Duhen T, Duhen R, Lanzavecchia A, Sallusto F, Campbell DJ. Functionally distinct subsets of human FOXP3+ Treg cells that phenotypically mirror effector Th cells. *Blood.* 2012;119(19):4430-4440.

45. Heintzman ND, Hon GC, Hawkins RD, et al. Histone modifications at human enhancers reflect global cell-type-specific gene expression. *Nature.* 2009;459(7243):108-112.

46. Rada-Iglesias A, Bajpai R, Swigut T, Brugmann SA, Flynn RA, Wysocka J. A unique chromatin signature uncovers early developmental enhancers in humans. *Nature.* 2011;470(7333): 279-283.

47. Creyghton MP, Cheng AW, Welstead GG, et al. Histone H3K27ac separates active from poised enhancers and predicts developmental state. *Proc Natl Acad Sci USA.* 2010;107(50): 21931-21936.

48. Yao Z, Kanno Y, Kerenyi M, et al. Nonredundant roles for Stat5a/b in directly regulating Foxp3. *Blood.* 2007;109(10):4368-4375.

49. Zorn E, Nelson EA, Mohseni M, et al. IL-2 regulates FOXP3 expression in human CD4+ CD25+ regulatory T cells through a STAT-dependent mechanism and induces the expansion of these cells in vivo. *Blood.* 2006; 108(5):1571-1579.

50. Polansky JK, Schreiber L, Thelemann C, et al. Methylation matters: binding of Ets-1 to the demethylated Foxp3 gene contributes to the stabilization of Foxp3 expression in regulatory T cells. *J Mol Med (Berl).* 2010;88(10): 1029-1040.

51. Mouly E, Chemin K, Nguyen HV, et al. The Ets-1 transcription factor controls the development and function of natural regulatory T cells. *J Exp Med.* 2010;207(10):2113-2125.

52. Bruno L, Mazzarella L, Hoogenkamp M, et al. Runx proteins regulate Foxp3 expression. *J Exp Med.* 2009;206(11):2329-2337.

53. Kitoh A, Ono M, Naoe Y, et al. Indispensable role of the Runx1-Cbfbeta transcription complex for in vivo-suppressive function of FoxP3+ regulatory T cells. *Immunity.* 2009;31(4): 609-620.