nature immunology

ARTICLES

https://doi.org/10.1038/s41590-021-00910-8

Check for updates

Single-cell analysis of FOXP3 deficiencies in humans and mice unmasks intrinsic and extrinsic CD4$^+$ T cell perturbations

David Zemmour[1,2], Louis-Marie Charbonnier[3], Juliette Leon[1,4], Emmanuelle Six[4], Sevgi Keles[5], Marianne Delville[4,6], Mehdi Benamar[3], Safa Baris[7,8], Julien Zuber[4,6], Karin Chen[9,10], Benedicte Neven[4,6], Maria I. Garcia-Lloret[11], Frank M. Ruemmele[4,6], Carlo Brugnara[3], Nadine Cerf-Bensussan[4,6], Frederic Rieux-Laucat[4], Marina Cavazzana[12,13,14], Isabelle André[4,14], Talal A. Chatila[3], Diane Mathis[1]✉ and Christophe Benoist[1]✉

***FOXP3* deficiency in mice and in patients with immune dysregulation polyendocrinopathy enteropathy X-linked (IPEX) syndrome results in fatal autoimmunity by altering regulatory T (T_{reg}) cells. CD4$^+$ T cells in patients with IPEX syndrome and *Foxp3*-deficient mice were analyzed by single-cell cytometry and RNA-sequencing, revealing heterogeneous T_{reg}-like cells, some very similar to normal T_{reg} cells, others more distant. Conventional T cells showed no widespread activation or helper T cell bias, but a monomorphic disease signature affected all CD4$^+$ T cells. This signature proved to be cell extrinsic since it was extinguished in mixed bone marrow chimeric mice and heterozygous mothers of patients with IPEX syndrome. Normal T_{reg} cells exerted dominant suppression, quenching the disease signature and revealing in mutant T_{reg}-like cells a small cluster of genes regulated cell-intrinsically by FOXP3, including key homeostatic regulators. We propose a two-step pathogenesis model: cell-intrinsic downregulation of core FOXP3-dependent genes destabilizes T_{reg} cells, de-repressing systemic mediators that imprint the disease signature on all T cells, furthering T_{reg} cell dysfunction. Accordingly, interleukin-2 treatment improved the T_{reg}-like compartment and survival.**

T_{reg} cells that express the transcription factor FOXP3 are dominant negative regulators of many facets of the immune system, controlling immune responses and enforcing peripheral tolerance to self, symbiotic commensals and fetal antigens[1,2]. They also have extra-immunological roles in maintaining tissue homeostasis outside of the purely immunological context[3]. T_{reg} cells share a common transcriptional signature, genes differentially expressed relative to their conventional CD4$^+$ counterparts (T_{conv}), in mice and humans[4–6]. Much of the T_{reg} signature is controlled by FOXP3, the lineage's defining transcription factor (TF), although FOXP3 is neither fully necessary nor sufficient to establish the T_{reg} transcriptome, requiring input from transcriptional cofactors[7,8]. Consistent with their pleiotropic functions, T_{reg} cells show a range of phenotypic variation, differing by activation state, effector capabilities and tissue localization[3,9]. T_{reg} heterogeneity has been further refined by single-cell transcriptomics (scRNA-seq)[6,10–12].

IPEX syndrome is perhaps the prototype of monogenic autoimmune syndromes[13], resulting from FOXP3 loss of function (LOF) and thus T_{reg} cell deficiency[1,14]. X-linked, it is transmitted by asymptomatic carrier females to their male progeny who usually present early after birth with a constellation of autoimmune manifestations dominated by severe enteropathy, eczematous dermatitis and type 1 diabetes, with some less common manifestations (nephropathy, hypoparathyroidism and antibody-mediated cytopenias)[15–18]. There are several causes to the variable severity observed in patients with IPEX syndrome. First, complete LOF mutations are generally more deleterious than missense mutations; and mutations in the DNA-binding forkhead or dimerization (leucine zipper) domains are generally more severe than N-terminal mutations, in keeping with results from our alanine-scan dissection of FOXP3 (refs. [8,17]). But manifestations and severity can vary between patients with the same mutation[17,19], suggesting that genetic modifiers, environmental influences or immunological history modify the disease course in each patient. Mice of the spontaneous *scurfy* mutant line, or carrying engineered *Foxp3* LOFs, show a similar rapidly lethal phenotype, dominated by enteropathy, dermatitis and lymphoproliferation[1,20]; partial or slower disease appears in mice with mild *Foxp3* missense mutations[8,21,22].

Also contributing to this variability, the relationship between FOXP3 and T_{reg} cells is now recognized not to be obligate. Several

[1]Department of Immunology, Harvard Medical School, Boston, MA, USA. [2]Department of Pathology, Brigham and Women's Hospital, Harvard Medical School, Boston, MA, USA. [3]Division of Immunology, Boston Children's Hospital, and Department of Pediatrics, Harvard Medical School, Boston, MA, USA. [4]INSERM UMR 1163, University of Paris, Imagine Institute, Paris, France. [5]Division of Pediatric Allergy and Immunology, Meram Medical Faculty, Necmettin Erbakan University, Konya, Turkey. [6]Necker Children's Hospital, Assistance Publique-Hôpitaux de Paris (AP-HP), Paris, France. [7]Division of Pediatric Allergy and Immunology, Marmara University School of Medicine, Istanbul, Turkey. [8]Istanbul Jeffrey Modell Diagnostic and Research Center for Primary Immunodeficiencies, Istanbul, Turkey. [9]Division of Allergy and Immunology, Department of Pediatrics, University of Utah School of Medicine, Salt Lake City, UT, USA. [10]Division of Immunology, Department of Pediatrics, University of Washington and Seattle Children's Research Institute, Seattle, WA, USA. [11]Division of Allergy, Immunology, and Rheumatology, Department of Pediatrics, David Geffen School of Medicine, University of California, Los Angeles, Los Angeles, CA, USA. [12]Biotherapy Department and Clinical Investigation Center, Assistance Publique Hopitaux Paris, Inserm, Paris, France. [13]Université de Paris, Paris, France. [14]Institut Imagine, Paris, France. ✉e-mail: cbdm@hms.harvard.edu; cbdm@hms.harvard.edu

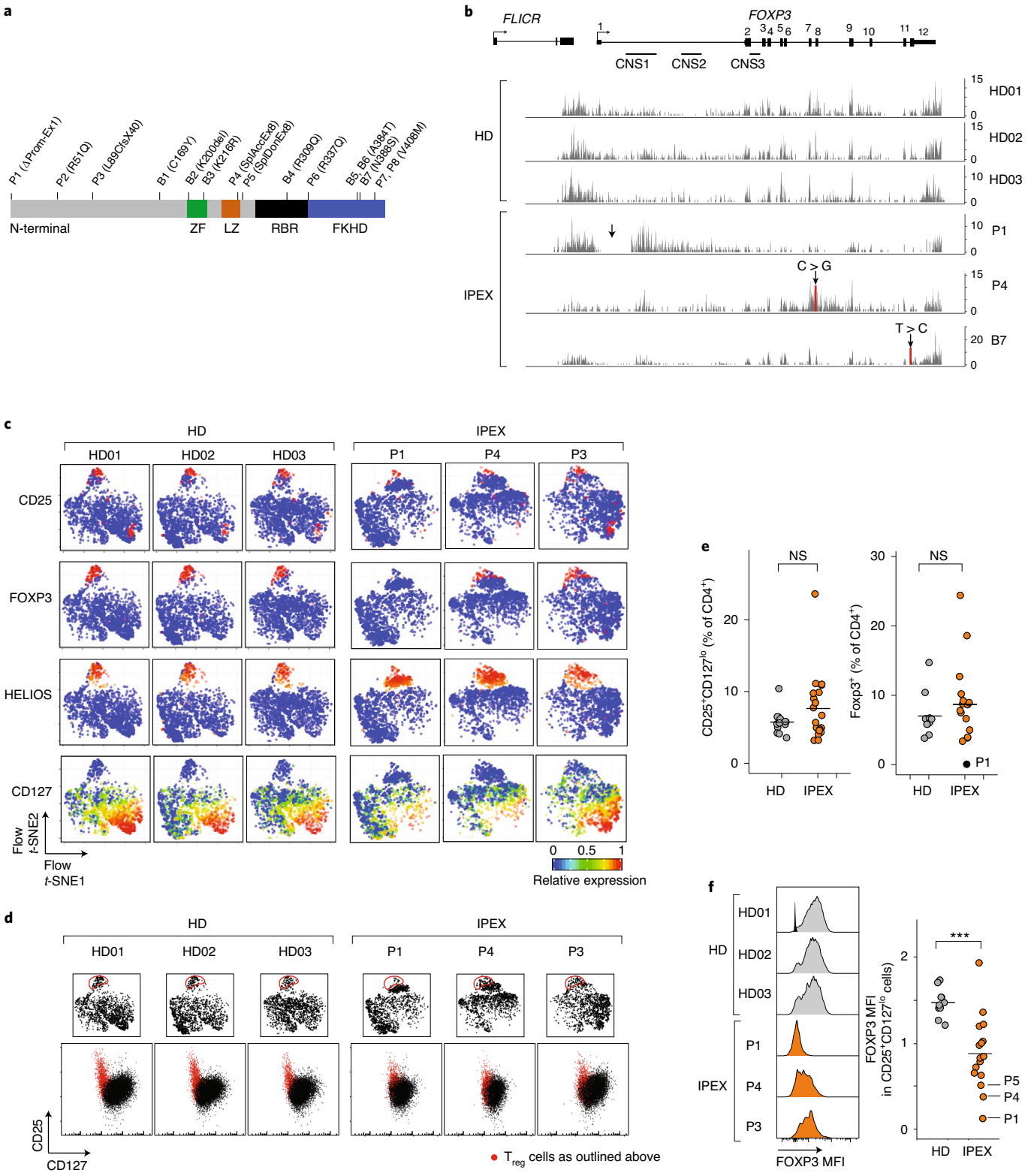

Fig. 1 | Identification of T_{reg}-like cells in IPEX syndrome by flow cytometry. a, Position of FOXP3 mutations in IPEX cohorts ($n=15$ IPEX). ZF, zinc finger; LZ, leucine zipper; RBR, RUNX1-binding region; FKHD, forkhead. **b**, Mapping of RNA-seq reads from CD4$^+$CD25$^+$CD127^{lo} cells to *FOXP3* locus in representative samples (Supplementary Fig. 1). Arrow indicates mutation. **c**, Flow-*t*-SNE plot of CD4$^+$ T cells from anti-CD3, CD4, CD25, CD127, HELIOS, CD45RA and FOXP3 staining in representative samples (full set in Supplementary Fig. 3). Color represents scaled expression. **d**, Flow cytometric anti-CD25/CD127 plots of CD4$^+$ T cells (bottom); red, cells gated in the flow-*t*-SNE plot (top). **e**, Proportion of CD25$^+$CD127^{lo} cells and total FOXP3$^+$ cells in HD and IPEX (each dot is a sample) (Supplementary Fig. 2a), NS, nonsignificant, two-sided Student's *t*-test ($n=15$ IPEX, $n=13$ HD). **f**, FOXP3 expression in CD25$^+$CD127^{lo} HD and IPEX cells. Representative cytometry profiles (left) (Supplementary Fig. 2b for all other samples; one unstained control overlaid with HD01) and quantitation (right) (each dot is a sample; $^{***}P<0.001$, two-sided Student's *t*-test, $n=15$ IPEX, $n=13$ HD).

reports have documented the existence of T_{reg}-like cells in the absence of viable *Foxp3* in *scurfy* mice[23–25] and in some patients with IPEX syndrome[26–28]. In addition, FOXP3 is not exclusive to T_{reg} cells: it is also expressed, albeit at lower levels than in T_{reg} cells, early after T_{conv} cell activation[29–32] and scRNA-seq showed some FOXP3-positive cells otherwise similar to T_{conv} cells[6].

The advent of single-cell transcriptomics opened the potential to revisit the alterations of CD4+ T cells in patients with IPEX syndrome, which were previously difficult to address as the markers potentially used to sort T_{reg} or T_{conv} cells may be themselves perturbed. We have thus performed a deep profiling analysis in patients with IPEX syndrome that associates flow cytometry, single-cell (for resolution) and conventional (for depth) RNA-seq. Complementary analyses in *scurfy* mice assessed the generality of the conclusions, eschewing the inevitable confounders of patient material and permitting experimental manipulations to trace the source of the perturbations. The results paint an unexpected landscape of IPEX T cells, with a complex mix of T_{reg}-like cells, some healthy like and others more perturbed, a surprisingly narrow intrinsic signature of FOXP3 but a dominant 'IPEX signature' that affects both T_{reg} and T_{conv} cells and suggests a feed-forward loop in T cell perturbations culminating in clinical disease.

Results

Study cohorts. This study included a primary and a replication cohort for confirmation and refinement (including some re-sampling of initial patients). Altogether, we analyzed peripheral blood mononuclear cells (PBMCs) from 15 patients with IPEX syndrome and 15 healthy donors (HDs) collected at two reference clinical centers (Necker Hospital, Paris and Children's Hospital, Boston; Supplementary Table 1 and Fig. 1a). HDs had no significant medical history and were recruited during well-child visits or orthopedic follow-ups. IPEX presentation was typical, with the first symptoms appearing neonatally for most (1–8 weeks). As in other cohorts[15–18], enteropathy (12 out of 14), dermatitis (10 out of 14), food allergy (8 out of 13) and diabetes (6 out of 14) were most common, with less frequent kidney, neurological or pulmonary involvement. Four patients were untreated at the time of sampling, but most were managed by immunosuppression (mainly mTOR inhibitors); four of them later received bone marrow transplants. Blood samples were collected at different ages (9 months to 26 years) during routine visits, with no concurrent acute events. Five patients were analyzed at two or three time points (6–12 months apart) to assess the stability of the transcriptional characteristics.

FOXP3 mutations were confirmed by sequencing. Most were missense mutations in different domains, two of them represented twice (V408M and the common A384T). One mutation (P1) was a large deletion spanning the promoter and intron 1; two others affected the exon 8 splice acceptor and donor sites (P4 and P5, respectively). The mutations' impact was seen in the RNA-seq reads at the *FOXP3* locus of sorted CD4+CD25^{hi}CD127^{lo} cells (see below). For HDs, the traces were very reproducible, with reads piling up at the exons, the conserved noncoding sequence (CNS) 1 and the lncRNA *FLICR* (Fig. 1b and Supplementary Fig. 1). Profiles from patients were variable; generally normal for the missense mutations; for P1, aberrantly initiated transcripts piled up around CNS1 with essentially no exonic reads; for P4, the splice mutation in exon 8 disturbed later exons and introns. Notably though, the *FOXP3* locus was active in all IPEX cells, at far greater levels than in T_{conv} cells (Supplementary Fig. 1). *FLICR* transcripts varied markedly in patient cells, unrelated to the proportion of *FOXP3* exonic reads, in line with its independent transcriptional regulation[33].

Identification of T_{reg}-like cells in patients with IPEX syndrome. T_{reg} cells are typically identified by flow cytometry as CD25^{hi}CD127^{lo}, a combination that overlaps well with intracellular FOXP3 expression[34]. Such cells have been reported in patients with IPEX syndrome[26–28], but we first adopted a more global approach to identify T_{reg}-like cells as expression of these two markers could be perturbed by FOXP3 deficiencies. We analyzed CD4+ PBMCs by multiparameter flow cytometry for a larger panel of T_{reg} markers and intracellular staining for FOXP3 and HELIOS (Supplementary Fig. 2). The data from individual staining experiments, including HD and IPEX samples, were then integrated in two-dimensional (2D) space by *t*-distributed stochastic neighbor embedding (*t*-SNE; Fig. 1c and Supplementary Fig. 3). T_{reg} cells from HD donors clustered tightly, with characteristic FOXP3, CD25 and HELIOS expression. For IPEX donors, the same region of the plots contained cells that also expressed CD25 and HELIOS, but variable FOXP3 (predictably absent for P1, more normal for P4). Indeed, 'gating' T_{reg}-like cells on the *t*-SNE showed that they fell in the expected position of the CD25/CD127 plot (Fig. 1d and Supplementary Fig. 2a), although with lower CD25 in IPEX samples. The proportions of T_{reg}-like cells and generally of FOXP3+ cells, were well conserved in patients with IPEX syndrome (Fig. 1e and Supplementary Fig. 2b), although levels of FOXP3 protein were variably reduced (Fig. 1f and Supplementary Fig. 2b). Other than the low or null levels resulting from promoter or splicing mutations, there was no discernible relationship between FOXP3 intensity and domains affected by each mutation. These cytometry results confirmed that cells with surface T_{reg} characteristics, but with low or absent FOXP3, could be identified in every patient with IPEX syndrome.

Widespread perturbations of IPEX CD4+ T cells transcriptomes. Given the presence of T_{reg}-like CD25^{hi}CD127^{lo} cells in PBMCs from patients with IPEX syndrome, we performed population low-input RNA-seq after sorting these and T_{conv} (CD25^{-}CD127^{hi}) cells, as a preliminary to single-cell profiling and to help anchor its interpretation (first and replication cohorts were profiled independently). Several observations emerged. First, the classic T_{reg} signature was well conserved in T_{reg} cells from IPEX donors (Fig. 2a), including prototype transcripts (*IL2RA*, *LRRC32*, *CTLA4*). In ranked fold change (FC) plots that display the T_{reg} signature for individual donors, the distribution of signature genes was tight for HDs but more variable in

Fig. 2 | Transcriptional changes in IPEX T_{reg} and T_{conv} cells by population RNA-seq. Population RNA-seq was performed on sorted T_{conv} and T_{reg}-like cells from HD ($n=12$) and IPEX ($n=10$) donors. **a**, The T_{reg}/T_{conv} FC in HDs (*x* axis) and IPEX (*y* axis); T_{reg} signature[5] genes are highlighted. **b**, Ranked FC plots of T_{reg} signature transcripts for individual donors, ranked according to mean FC in all HD (blue dots). FC values for each donor (black dots) computed from the donor's T_{reg} versus the mean of HD T_{conv} cells. **c**, Index and CV of T_{reg} up signature transcripts (each dot is a sample). ***$P<0.001$, two-sided Student's *t*-test . **d**, FC versus *P* value (volcano) plot comparing normalized expression in all IPEX to all HD samples. Genes with differential expression (two-sided Student's *t*-test $P<0.01$, FC > 2) are highlighted (452 up genes and 1,032 down genes) and numbers are shown. **e**, Heat map of the expression ratio for IPEX signature genes defined in **d** and in T_{reg} cells from each donor, computed against mean expression in HD T_{reg} cells (each cohort was computed against its own HD T_{reg} set). **f**, Comparison of mean IPEX effect (all IPEX versus all HD samples) in T_{reg} (*x* axis) versus T_{conv} (*y* axis) cells. **g**, Same FC/FC plot as in **f**, but highlighted with representative signatures of tumor-infiltrating T_{reg} cells, activated T_{reg} versus resting T_{reg} cells and IFN-I induced genes. **h**, Same volcano plot as in **d**, but with highlights from a representative CD4+ T activation signature (NS, hypergeometric test). **i**, Heat map, for the IPEX signature genes defined in **d**, of their overlap with the pathways and signatures that significantly overlap with IPEX signature (hypergeometric test, $P<0.001$). Present genes are shown by tick marks, color coded by their IPEX/HD FC. Up, upregulated; down (DN), downregulated.

IPEX T_{reg} cells, ranging from quasi-normal (P7) to markedly affected (P6, P8; Fig. 2b and Supplementary Fig. 4a), perhaps surprisingly, P1 with the complete LOF mutation did not show the strongest reduction in T_{reg} signature genes. The T_{reg} signature intensity score seemed better preserved in T_{reg} cells from IPEX donors than its coefficient of variation (CV) (Fig. 2b,c), suggesting an instability of the T_{reg}

signature in absence of a fully functional FOXP3. These indices did not correlate significantly with clinical outcomes, although patients under immunosuppression trended to a lower index (Supplementary Fig. 4b); this correlation is confounded, however, as patients with more severe *FOXP3* mutations and thus a lower T_{reg} up index, are clinically more affected and more likely to receive strong treatment. The IPEX mutations did not affect all signature genes equally; some T_{reg} signature genes were actually over-expressed in IPEX T_{reg} cells (*DUSP4*, *LRRC32*, *CTLA4*), whereas most were downregulated as expected (Supplementary Fig. 4c).

It is worth mentioning that no patient's T_{reg} or T_{conv} cells showed unusual expression of *IL4* or *IL5*, transcripts, which have been found to be paradoxically upregulated in response to forced expression of mutant FOXP3 (refs. [8,22]). More generally, there was no specific induction of cytokine genes in T_{conv} cells from patients with IPEX syndrome, which might have denoted a loss of T_{reg} cell control (Supplementary Fig. 4d).

We then assessed more generally the impact of IPEX mutations on transcriptomes of T_{reg}-like cells. Widespread differences were observed (Fig. 2d), reproducibly in the two cohorts (Supplementary Fig. 5a and Supplementary Table 2). This 'IPEX signature' was consistent in every patient, involving all the same genes, albeit at variable overall intensity (Fig. 2e; one HD did show partial bias, perhaps because of unrecognized pathology). It did not correlate with the T_{reg} signature intensity score (Supplementary Fig. 5b) and was strikingly similar in T_{reg} and T_{conv} cells (Fig. 2f), indicating a global impact on CD4$^+$ T cells that transcended the sole effect of FOXP3 in T_{reg} cells (only 17 genes of the T_{reg} signature belonged to this IPEX signature; Supplementary Table 2). There was no marked relationship between the main clinical parameters, including current corticosteroid or rapamycin treatment and the IPEX signature score (Supplementary Fig. 5c), which was also present in untreated patients with IPEX syndrome. The IPEX signature was not accompanied by activation of endogenous retroviruses, which might plausibly be reactivated (Supplementary Fig. 5d). Enrichment analysis showed significant overlap between the IPEX signature and several gene expression signatures of CD4$^+$ T cells (Fig. 2g–i and Supplementary Table 3), but notably not with the signatures of T cell activation (Fig. 2h), again indicating that the absence of T_{reg} suppression did not result in wholesale T cell activation. If anything, a T_{reg} activation signature[35] was downregulated in IPEX cells (Fig. 2g). Transcripts differentially expressed in tumor-infiltrating T_{reg} cells (two independent studies) were counter-regulated in IPEX CD4$^+$ T cells, as were interferon (IFN)-stimulated genes (Fig. 2g,i). This broad change was consistent with the upregulation of major response regulators such as *RELB*, *KRAS* or *HIF1*. Overall, our results indicated that the T_{reg} signature was in large part maintained in T_{reg} cells from patients with IPEX syndrome, albeit with a notably high degree of variability. More unexpected was the peculiar transcriptomic footprint shared by IPEX T_{reg} and T_{conv} cells, which might result from integration of extracellular cues, but not from a generic T cell activation state.

Different types of T_{reg} cells in patients with IPEX syndrome. From these preliminary indications, we proceeded to the crux of the work, applying scRNA-seq to identify with unbiased profiling the actual types of FOXP3-deficient T_{reg}-like cells, which might be blurred by the averaging inherent in population profiling or have escaped recognition because of shifts in CD25 or CD127 markers. We sorted total CD4$^+$ T cells to yield granular information on both T_{reg} and T_{conv} pools and performed scRNA-seq on the two cohorts as above (a total of 52,776 cells passing quality control (Supplementary Table 4) from 11 IPEX and 11 HD donors). All results were observed in both the initial and replication cohorts, but are combined below for simplicity. Experimental confounders were minimized by multiplexing IPEX and HD samples in the same scRNA-seq libraries, using DNA-coded tags[36]. Residual intercohort and interexperiment effects were corrected using the canonical correlation analysis (CCA) and *k*-nearest neighbor-based integration methodology[37] (Fig. 3a).

In the uniform manifold approximation and projection (UMAP) of the integrated data, CD4$^+$ T cells partitioned sharply according to disease; cells from HD donors clustered together, whereas those from patients with IPEX syndrome were much more dispersed (Fig. 3a), each tending to form an island distant to various degrees from the HD group (Fig. 3b). These distances were not batch artifacts (similar HD/IPEX partitions were seen in each batch; Supplementary Fig. 6a). Replicate samples from three patients drawn >1 year apart mapped to the same regions, as did samples from patient P7 collected before and after immunosuppressive treatment (Supplementary Fig. 6b), indicating that the different locations were patient-specific and not consequences of environmental or treatment variables. Conventional and scRNA-seq reflected the same perturbation, as UMAP1 scores axis that partitioned IPEX and HD samples corresponded with the IPEX signature above (Supplementary Fig. 6c).

We then sought to deconvolute, at single-cell resolution, T_{reg} cells among these CD4$^+$ T cells. T_{reg} cells were identified in HD and IPEX samples in an unsupervised approach, using community detection in the reciprocal principal-component analysis (PCA) integration network (implemented in Seurat v.3 (ref. [37])), also supported by activity at the *FOXP3* locus. Three distinct populations of CD4$^+$ T cells could be identified (Supplementary Fig. 6d): two T_{conv} populations (resting and activated) and one T_{reg} population (confirmed by expression of T_{reg} signature genes *FOXP3*, *IKZF2*, *CTLA4*, *IL2RA* and *TIGIT*; Supplementary Fig. 7a). When T_{reg} cells thus identified were displayed onto the UMAP space, HD T_{reg} cells formed a tight cluster for all HD samples (Fig. 3c), which colocalized with *FOXP3* expression (yellow dots on Fig. 3c). In IPEX samples, T_{reg}-like cells were similarly identified, in proportions equivalent to controls, even for fully FOXP3-deficient P1 (Supplementary Fig. 7b), but they split into two groups (Fig. 3c and Supplementary Fig. 8a); some (hereafter referred to as type A IPEX T_{reg} cells) mapped in the same cluster as HD T_{reg} cells, whereas others were clustered away in their respective patient-specific 'island' (type B). Both were truly T_{reg} cells, expressing *FOXP3* mRNA (Fig. 3c), core T_{reg} transcripts (*IL2RA*, *CTLA4*, *IKZF2* and low *IL7R*) and the T_{reg} signature (Fig. 3d). Type A and type B T_{reg} cells did differ, with higher levels of T_{reg} signature genes in type A T_{reg} cells and higher representation in type B cells of the IPEX signature (Fig. 3e and Supplementary Fig. 8b). Finally, proportions varied between patients with IPEX syndrome (Fig. 3f); type A T_{reg} cells dominated in some, almost to HD levels, but were far less abundant in others. This proportion was stable in independent

Fig. 3 | scRNA-seq reveals the heterogeneous effect of FOXP3 deficiency in IPEX T_{reg} cells. scRNA-seq was performed on peripheral blood CD4$^+$ T cells from IPEX ($n=11$) and HDs ($n=11$). **a**, 2D UMAP plot of all CD4$^+$ cells from IPEX and HD samples (52,776 cells altogether, merged with CCA and knn-graph; Methods). **b**, Same UMAP as **a**, color-coded by individual donor (see also Supplementary Fig. 6d). **c**, Same UMAP as **a**, with four representative HD and IPEX donors (see Supplementary Fig. 8 for other donors). Blue, green and red cells are resting T_{conv}, activated T_{conv} and T_{reg} cells, respectively. *FOXP3*-expressing cells (RNA) are yellow. Type A and B T_{reg} cells in IPEX donors are indicated by an arrow. **d**, Normalized counts expression of *IL2RA*, *CTLA4*, *IKZF2*, *IL7R* and the T_{reg} signature in T_{reg} cells from HD and IPEX (type A and B) and T_{conv} cells; each dot is a sample. $^*P<0.05$, Student's *t*-test. **e**, Average expression of the IPEX signature in type A and B IPEX T_{reg} cells and HD T_{reg} cells (normalized counts). $^{***}P<10^{-3}$, $^{****}P<10^{-4}$, two-sided Student's *t*-test. **f**, Proportion of type A T_{reg} cells in total T_{reg} cells for each sample. **g**, Proportion of type A T_{reg} cells plotted against the average UMAP1 dimension for each sample.

samplings of the same patient and related to the intensity of the IPEX signature in T_{reg} and T_{conv} cells; patients with the highest index had the lowest fraction of type A T_{reg} cells (Fig. 3g).

Thus, scRNA-seq analysis revealed a subset of T_{reg} cells that closely resemble normal T_{reg} cells and another with more extensive perturbations. The origin of these two distinct T_{reg} populations in

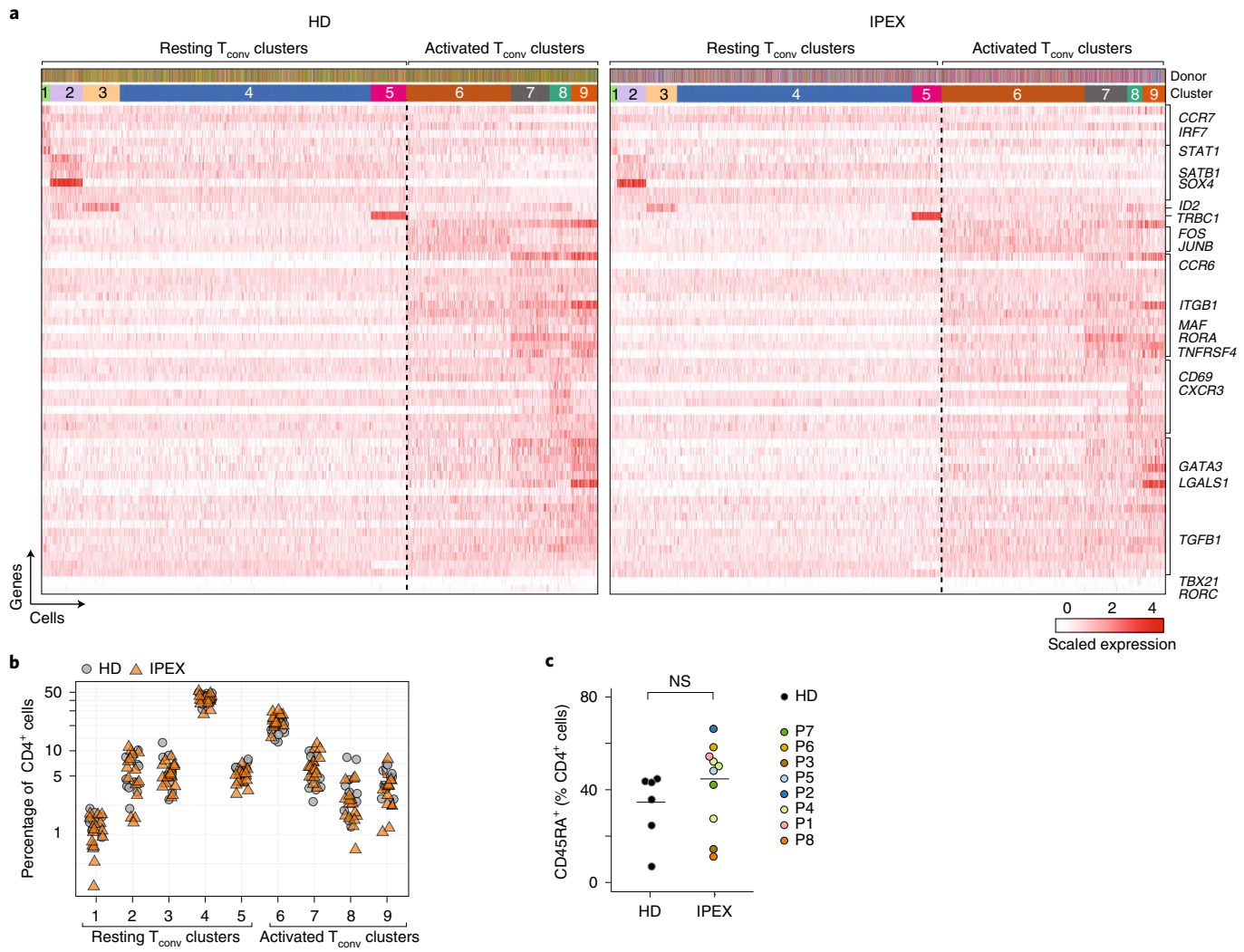

Fig. 4 | Normal T_{conv} phenotypes in patients with IPEX syndrome. a, From the scRNA-seq T_{conv} datasets (HDs and IPEX combined), the IPEX signature was regressed out (Methods) and biclustering was performed to define clusters in resting and activated T_{conv} cells, plotted as a heat map for expression of the most characteristic genes for each T_{conv} cluster (each vertical line represents one cell). Donor origin and cluster annotations for every cell (top). **b**, Proportion of resting and activated T_{conv} clusters in total CD4$^+$ cells in HD and IPEX in scRNA-seq data. **c**, Proportion of CD45RA$^+$ (resting) CD4$^+$ T cells determined by flow cytometry in IPEX and HDs. NS, two-sided Student's *t*-test ($n = 11$ HDs, 11 IPEX).

patients with IPEX syndrome was unclear. To exclude maternal microchimerism (wild-type (WT) T_{reg} cells from the mother could have a competitive advantage in the IPEX offspring), we checked female-specific *XIST* transcripts; no *XIST*-positive cells were found among any patient's T_{reg} cells. A striking feature of type A T_{reg} cells was their marked downregulation of the IPEX signature.

IPEX does not affect T_{conv} phenotypes. T_{reg} cells can affect T_{conv} polarization in many ways[9] and it is generally assumed that T_{reg} cell deficiency in patients with IPEX syndrome leads to excessive T_{conv} cell activation and differentiation and an over-representation of their effector states[38,39]. Unsupervised clustering of T_{conv} cells performed in the shared reciprocal PCA space, after regressing out the IPEX effect, distinguished nine clusters (five resting and four activated/memory types, judging from the expression of characteristic markers and transcription factors such as *CCR7*, *SATB1*, *CD69*, *TBX21* and *GATA3*; Fig. 4a). Their distribution and the expression of defining transcripts, was strikingly similar for HD and IPEX samples (Fig. 4a,b). This observation was confirmed by flow cytometry, which showed similar ranges of CD45RA$^+$ cells among CD4$^+$ T cells in IPEX and HD PBMCs (Fig. 4c). Thus, beyond the shared signature, IPEX disease seemed to have limited impact on other phenotypic aspects of circulating T_{conv} cells.

Mixed populations of *Foxp3*-active cells in *Foxp3*-deficient mice. *FOXP3* deficiency in patients with IPEX syndrome thus led to a global change affecting all CD4$^+$ T cells and to a diversity of T_{reg}-like cells. To assess mutational impacts in a setting devoid of genetic or therapeutic variables, we performed parallel analyses on *Foxp3*-deficient mice (the previously described *Foxp3*$^{\Delta EGFPcre}$ inactivating reporter line[25], hereafter denoted Δ*Foxp3*), in which FOXP3 protein is absent, but *Foxp3* locus activity can be detected.

We used scRNA-seq to analyze the diversity of CD4$^+$ T splenocytes in Δ*Foxp3* mice and WT controls (four mice per group, 18,000 cells altogether, again multiplexing samples in the same libraries; Supplementary Table 4). As in human patients, cells from WT and Δ*Foxp3* mice clustered separately on the UMAP, here again reflecting a generic ΔFoxp3 signature (Fig. 5a) present in every mutant mouse and affecting T_{reg} as well as T_{conv} cells (Supplementary Fig. 9a). This cross-cutting '*ΔFoxp3* signature' was confirmed by

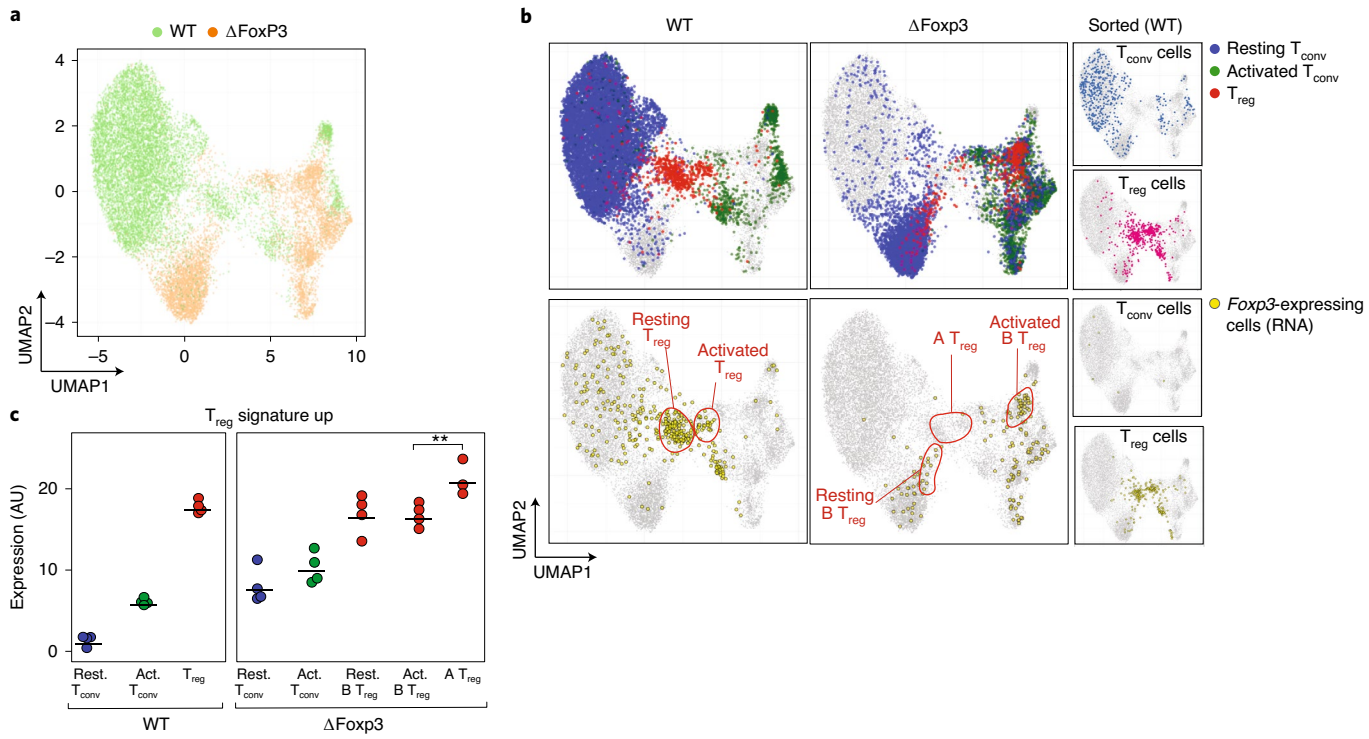

Fig. 5 | scRNA-seq reveals heterogeneous effects of *Foxp3* ablation in Δ*Foxp3* mice. scRNA-seq was performed on CD4$^+$ T cells from WT ($n=4$) and Δ*Foxp3* ($n=4$) mice (18,097 cells altogether). **a**, 2D UMAP plot of CD4$^+$ single-cell transcriptomes for WT (green) and ΔFoxP3 (orange). **b**, Same UMAP as **a**. Resting T_{conv} (blue), activated T_{conv} (green) and T_{reg} cells (red) (assigned by graph-based clustering on the merged dataset after regressing out the Δ*Foxp3* signature; Methods) are highlighted (top); cells with an active *Foxp3* locus (*GFP* or *Cre* transcripts detected) are colored yellow (bottom). Small inserts (right) show sorted control WT T_{conv} and WT T_{reg} cells, sorted and included as spike-in controls. **c**, T_{reg} signature expression in resting T_{conv}, activated T_{conv} and T_{reg} cells in WT and Δ*Foxp3* mice (normalized counts). $^{**}P<10^{-2}$, two-sided Student's *t*-test. GFP, green fluorescent protein.

population RNA-seq of Δ*Foxp3* T_{reg} cells. The same perturbation was found in Δ*Foxp3* T_{reg} cells from spleen and lung, showing that it extended to parenchymal locations, specifically in a site relevant to scurfy disease (Supplementary Fig. 9b).

Unsupervised clustering in the shared CCA space identified, for WT CD4$^+$ cells, resting and activated T_{conv} cells and a tight group of T_{reg} cells (blue, green and red, respectively; Fig. 5b). These assignments were confirmed by the location of sorted T_{conv} or T_{reg} cells (Fig. 5b), by differential expression of canonical genes (for example *Ccr7*, *Sell*, *Cd44*, *Tbx21*; Supplementary Fig. 9c and Supplementary Table 5) and at the *Foxp3* locus (as *Cre* and *Foxp3* transcripts; Fig. 5b). Note how T_{conv} clusters were similarly structured in Δ*Foxp3* and WT mice, as they had been in human patients (Supplementary Fig. 9c).

As for patients with IPEX syndrome, T_{reg}-like cells of Δ*Foxp3* mice were multiform (Fig. 5b); a minor fraction of type A T_{reg} cells closely resembled normal T_{reg} cells, in the same small proportions as in the most complete IPEX deficiencies; a larger proportion of type B T_{reg} cells that mapped into resting and active areas of the UMAP (Fig. 5b), also expressing *Foxp3* (Fig. 5b). Both expressed the T_{reg} signature, highest for type A T_{reg} cells (Fig. 5c). Together, these T_{reg}-like populations accounted for 7% of total CD4$^+$ T cells, as in WT T_{reg} cells (Supplementary Fig. 9d).

Thus, T_{reg}-like cells in *Foxp3*-deficient mice showed the same heterogeneity as in human patients with IPEX syndrome; a mix of healthy and altered T_{reg} cells, but with a dominant transcriptional signature that cut across both T_{reg} and T_{conv} CD4$^+$ T cells.

Cell-intrinsic and extrinsic impact of *Foxp3* deficiency. In humans and mice, FOXP3 deficiency resulted in a heterogenous mix of T_{reg}-like cells and a strong disease-specific signature unexpectedly shared by T_{reg} and T_{conv} cells. The latter suggested cell-extrinsic influences on the transcriptomes of CD4$^+$ T cells, which we investigated in two ways.

First, we analyzed FOXP3-deficient T_{reg} cells present in the same environment as WT cells, in blood CD4$^+$ T cells from two mothers of patients with IPEX syndrome. X inactivation in females occurs with equal probability, such that half of the T cells in carrier mothers inactivate the mutant X chromosome and half the WT X chromosome. In the presence of T_{reg} cells expressing WT FOXP3, cell-extrinsic signals should be repressed, unmasking the true cell-intrinsic effect of *FOXP3* mutations. We performed scRNA-seq on blood CD4$^+$ T cells from mothers of patients P5 and B2, analyzing X chromosome reads that overlapped known single-nucleotide polymorphism variants to assign the active X chromosome of each cell to either the patient's mutant haplotype or the nontransmitted WT (Fig. 6a and Methods). T_{reg} cells in which the mutant chromosome was active were present in roughly equal proportions to their WT counterparts, as was the case for T_{conv} cells. In the UMAP, cells with the mutant and WT ChrX active mapped to the same areas for both T_{reg} and T_{conv} cells (Fig. 6b) and all T_{reg} cells were type A, irrespective of the active ChrX. The IPEX signature induced in mutant T_{conv} cells was fully extinguished (Fig. 6c), with no significant gene expression differences between T_{conv} with WT or mutant ChrX, highlighting the absence of a cell-intrinsic effect of FOXP3 in unchallenged T_{conv} pools. The T_{reg} up signature genes that were dysregulated in P5 or B2 T_{reg} cells generally normalized in mother's T_{reg} cells expressing the same mutant FoxP3 (Fig. 6d). These results indicate that the signals generating the IPEX signature are cell extrinsic.

Second, we constructed bone marrow chimeras (BMCs) by transfer of equal proportions of congenically marked stem cells

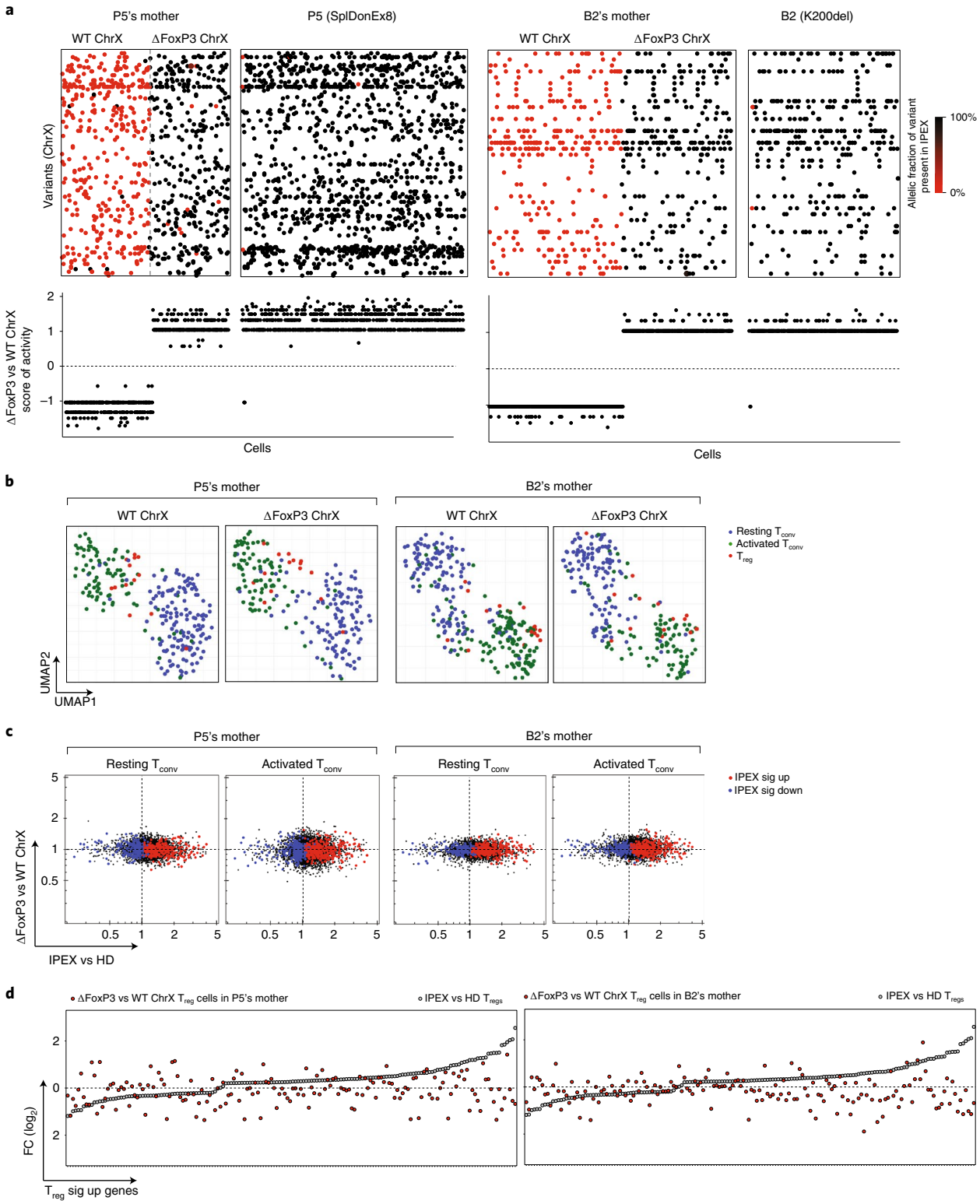

Fig. 6 | Cell-intrinsic and extrinsic effects of *FOXP3* deficiency in T_{reg} and T_{conv} cells of IPEX female carriers (mothers). scRNA-seq was performed on sorted and hashtagged blood CD4$^+$ T cells from two IPEX mothers: P5's and B2's mothers (2,880 cells altogether). **a**, Because of X chromosome (ChrX) inactivation in females, a mix of FOXP3-deficient (Δ*FOXP3*) and -proficient (WT) T cells are present in IPEX mothers. Biclustering heat map, showing for each single cell (columns) the allelic fraction of X chromosome variants carrying the FOXP3 mutation (present in P5 or B2) (top). FOXP3-deficient versus -proficient X chromosome score of activity for each single cell (bottom). **b**, 2D UMAP plot of all CD4$^+$ cells from P5's and B2's mothers split by active X chromosome (FOXP3-proficient (left); FOXP3-deficient (right)). Resting T_{conv}, activated T_{conv} and T_{reg} cells are highlighted in blue, green and red, respectively. **c**, scRNA-seq data for T_{conv} cells were collapsed and expression ratios between FOXP3-deficient and -proficient T_{conv} cells (resting or activated) were calculated for IPEX (*x* axis) and in P5's and B2's mothers (*y* axis). Upregulated and downregulated IPEX signature genes are indicated in red and blue, respectively. **d**, FC plots of FOXP3-deficient versus -proficient T_{reg} cells in P5's and B2's mothers, T_{reg} up signature genes, ranked according to FC in IPEX versus HDs (gray dots).

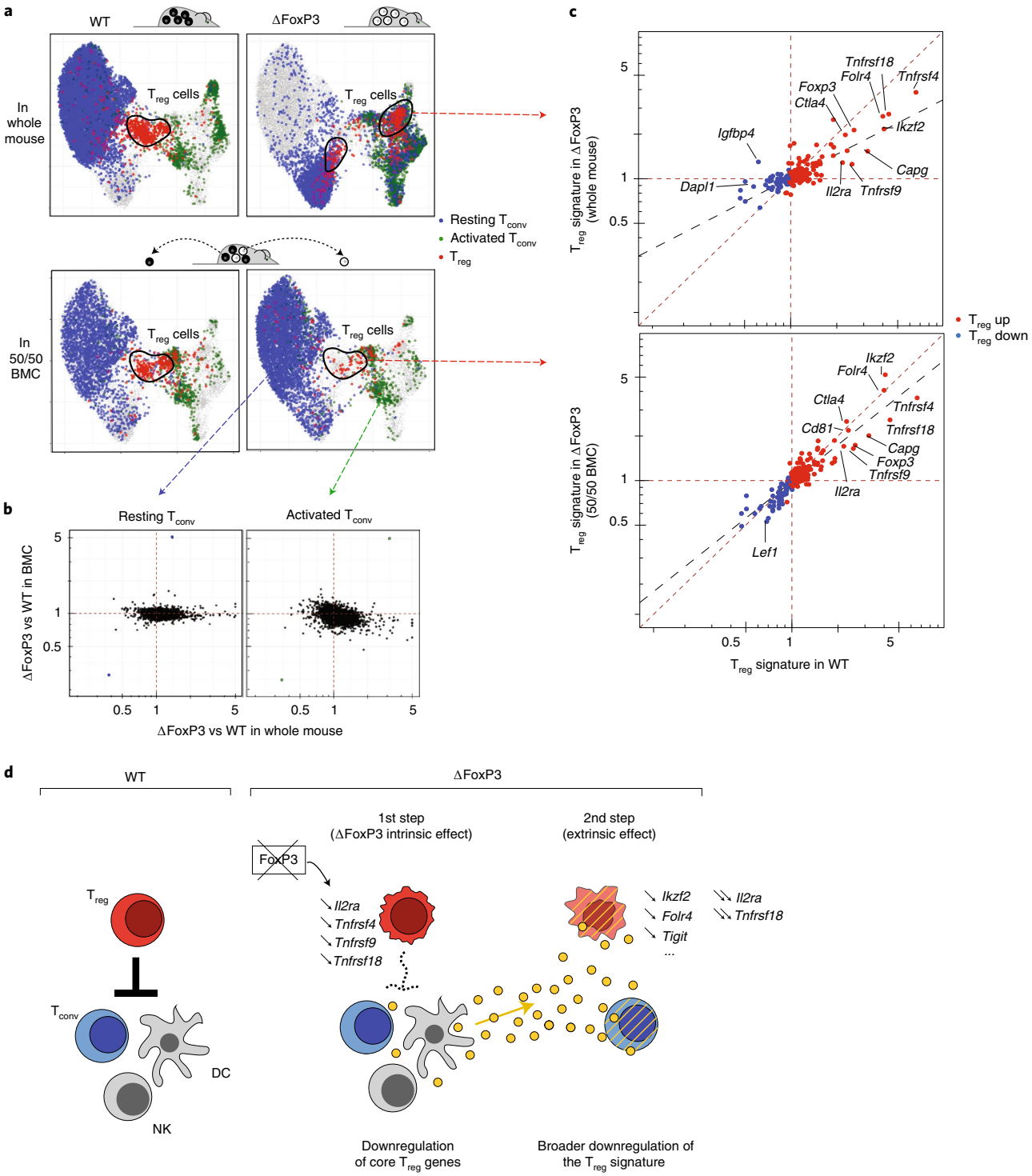

Fig. 7 | Cell-intrinsic and extrinsic effects of *Foxp3* deficiency in T_{reg} and T_{conv} cells in mice. scRNA-seq was performed on sorted and hashtagged CD4$^+$ T cells from 50/50 mixed BMC mice (WT and Δ*Foxp3* hematopoietic stem cells, congenically marked) (8,556 cells altogether, $n=3$ mice). **a**, Same UMAP as shown in Fig. 5b, integrating WT and Δ*Foxp3* CD4$^+$ cells from whole mice (top) or from chimeras (bottom). Resting T_{conv} (blue), activated T_{conv} (green) and T_{reg} cells (red) (identified by graph-based clustering on the merged dataset after regressing out the Δ*Foxp3* signature; Methods). **b**, scRNA-seq data for T_{conv} cells were collapsed and expression ratios between WT and Δ*Foxp3* T_{conv} were calculated for whole mice (*x* axis) or for 50/50 BMCs (*y* axis); for the latter, WT and Δ*Foxp3* cells originated from the same host. **c**, T_{reg}/T_{conv} ratio in Δ*Foxp3* cells (*y* axis) in whole mice (top) and in 50/50 BMCs (bottom), both plotted against the same T_{reg}/T_{conv} expression ratio in WT cells (*x* axis). T_{reg} signature genes are highlighted. Dashed lines (linear regression) represent the slope of the fit. **d**, Two-step model of the impact of FOXP3 deficiency in mice and humans. First, the absence of FOXP3 downregulates a few core T_{reg} signature genes in T_{reg} cells (cell-intrinsic effect), leading to perturbed control of other immune cells (which might include known T_{reg} targets such as T_{conv} cells, dendritic cells (DCs), natural killer (NK) cells, or others), unleashing a secondary intercellular response (for example, cytokines and cell surface modulators), which imprints the dominant IPEX signature on T_{reg} and T_{conv} cells and further affects T_{reg} functionality, amplifying disease pathology.

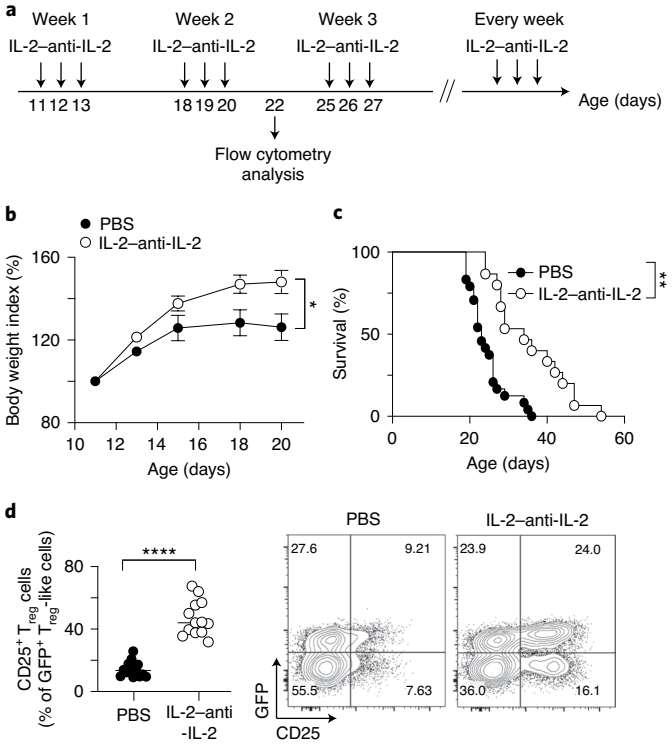

Fig. 8 | IL-2 therapy mitigates *Foxp3*-deficiency disease. a, IL-2 treatment protocol: daily injection of IL-2–anti-IL-2 complex for 3 d consecutively every week, starting 11 d after birth. **b**, Body weight index growth curves of vehicle (PBS) and IL-2 (IL-2–anti-IL-2)-treated Δ*Foxp3* mice (mean ± s.e.m.). $^{*}P < 0.05$, one-way analysis of variance, $n = 10$ and 19 (PBS and IL-2 groups, respectively). **c**, Survival curves of vehicle (PBS) and IL-2 (IL-2–anti-IL-2)-treated Δ*Foxp3* mice. $^{**}P < 10^{-4}$, log-rank test, $n = 24$ and 15 (PBS and IL-2 groups, respectively). **d**, Proportion of $CD25^{+}GFP^{+}$ T_{reg} cells (of GFP^{+} T_{reg}-like cells) in vehicle (PBS) and IL-2 (IL-2–anti-IL-2)-treated Δ*Foxp3* mice (day 22). Representative flow cytometric CD25/GFP plots of $CD4^{+}$ T cells (right). $^{****}P < 10^{-4}$, two-sided Student's *t*-test, $n = 14$ and 13 (PBS and IL-2 groups, respectively).

from WT and Δ*Foxp3* mice, a protocol that prevents the scurfy-like disease that appears after reconstitution with *Foxp3*-deficient stem cells[40]. As in hemizygous human mothers, cell-extrinsic transcriptional hallmarks should be reverted in genetically *Foxp3*-deficient cells by the presence of WT cells in the same mouse. Ten weeks after reconstitution, 5,556 $CD4^{+}$ T cells from either Δ*Foxp3* or WT compartments of BMC mice were profiled by multiplexed scRNA-seq (three mice per group, with data aligned to the same UMAP space as in Fig. 5).

In this mixed setting, many of the characteristics of Δ*Foxp3* $CD4^{+}$ T cells essentially disappeared (Fig. 7a), whereas WT cells were unchanged. First, on the UMAP plot, resting and activated T_{conv} cells from Δ*Foxp3* compartments shifted and overlapped with WT cells from the same mice. Accordingly, the Δ*Foxp3* signature was flattened out for mutant cells in the BMC setting (Fig. 7b and Supplementary Fig. 10a). These results demonstrated the cell-extrinsic origin of the Δ*Foxp3* signature in *Foxp3*-deficient T_{conv} cells. Δ*Foxp3* cells were partially outcompeted by WT T_{reg} cells (1:3) in the BMC mice (Supplementary Fig. 10b), but their phenotypes were also corrected; they were all type A T_{reg} cells that overlapped with WT T_{reg} cells on the UMAP (Fig. 7a) and their expression of the T_{reg} signature was almost completely normalized (Fig. 7c and Supplementary Table 5). Only 13 genes remained as under-expressed in the Δ*Foxp3* T_{reg} compartment, including classic FOXP3 target loci (*Il2ra*, *Tnfrsf4*, *Tnfrsf9*, *Tnfrsf18* and *Capg*); no genes were upregulated, confirming that FOXP3 is mainly a transcriptional activator. The other T_{reg} signature genes were otherwise normally expressed, contrasting with their general downregulation in Δ*Foxp3* mice.

Complementing FoxP3 deficiency with IL-2. Thus, the dominant suppressive effect of WT T_{reg} cells curtailed the Δ*Foxp3* signature in T_{reg} and T_{conv} cells, revealing a narrower intrinsic effect of *Foxp3* on a minor proportion of the T_{reg} signature, highlighting both intrinsic and extrinsic consequences of the *Foxp3* deficiency in the disease context. We thus propose a 'two-step' model of the IPEX/scurfy disease (Fig. 7d). First, the intrinsic effect of *Foxp3* in T_{reg} cells dysregulates a few core T_{reg} genes, which unleashes a systemic response. Secondarily, this milieu imprints broad changes on both T_{reg} and T_{conv} cells (IPEX/Δ*Foxp3* signatures) and further destabilizes T_{reg} signature expression and dampens T_{reg} effector function.

To provide a mechanistic proof of concept, we reasoned that enhancing an essential core T_{reg} gene in Δ*Foxp3* mice might restore T_{reg} functionality. *Il2ra* is one of the genes intrinsically regulated by FOXP3 above (see elsewhere[8,40,41]) and is preferentially expressed in type A T_{reg} cells (Fig. 7c). Interleukin (IL)-2, its ligand, is the key trophic cytokine for T_{reg} cells and induces Il2ra expression in a positive feed-forward loop[42]. IL-2 administration might thus complement the deficit due to the missing FoxP3 and, in part, revert the T_{reg} deficiency. Therefore, we evaluated the potential of serial IL-2 injections (stabilized in a complex with an anti-IL-2 antibody[43]) into Δ*Foxp3* pups (three consecutive daily injections per week starting at 11 d of age; Fig. 8a). Compared to the phosphate-buffered saline (PBS) control, IL-2–anti-IL-2 complexes partially prevented weight loss (Fig. 8b) and prolonged survival (Fig. 8c) of Δ*Foxp3* mice, at least for a period of time. The treatment also expanded the $CD25^{+}$ fraction within T_{reg}-like cells, thus presumably growing the type A fraction (Fig. 8d). These results show that reverting one of the core deficiencies resulting from the Δ*Foxp3* mutation does improve T_{reg} function, at least for some time and additionally suggest that IL-2 therapy might be used to mitigate Δ*Foxp3* disease.

Discussion

These results provide unprecedented cellular and genomic resolution of a primary immunodeficiency and reveal unexpectedly multifarious molecular and cellular consequences of FOXP3 deficiency. There were surprisingly limited cell-intrinsic perturbations in T_{reg} cells, associated with a dominant and monomorphic signature of cell-extrinsic origin, which cuts across the transcriptome of all $CD4^{+}$ T cells and amplifies T_{reg} perturbations.

We thus propose a two-step model of the disease's molecular pathogenesis. FOXP3 is actually important for only very few T_{reg} genes; when the milieu was normalized by the presence of WT T_{reg} cells the cell-intrinsic impact of the Δ*Foxp3* deficiency was exerted on only a handful of genes (*Il2ra*, *Tnfrsf4*, *Tnfrsf9*, *Tnfrsf18*, *Capg*, *Ikzf2* and *Ctla4*). This surprisingly short list corresponds to a 'core set' of genes that are expressed by all T_{reg} cells[6] and are directly transactivated by and bind FOXP3 (refs. [8,44]). These genes encode the major homeostatic regulator of T_{reg} cells (*IL2RA*) and several members of the tumor necrosis factor (TNF) receptor superfamily, which are also connected to T_{reg} homeostasis and function[45]. We propose that their downregulation is the first step (accordingly, *IL2RA* deficiency also causes an IPEX-like disease). This initiates a systemic reaction which constitutes the second step of the FOXP3 deficiency syndrome; a broadly shared signature that marks both T_{reg} and T_{conv} cells, further perturbs T_{reg} signature transcripts and amplifies in a vicious circle the defect in T_{reg}-suppressive activity.

What, then, is this IPEX signature, the hundreds of genes that were impacted equivalently in T_{reg} and T_{conv} cells? It was shared among all patients with IPEX syndrome, with different intensities

rather than qualitative differences, stressing its common etiology. It proved stable over several years in each multiply tested patient, before and after (and unrelated to) treatment. The genes involved were not simply T cell-activation genes, as might have been expected from a loss of suppression (activation-induced transcripts were actually repressed). We hypothesize that it is due to an inductive element(s), delivered via cytokine or cell–cell contact, normally repressed by T_{reg} cells but unleashed by FOXP3 deficiency. This signal might originate from other T cells, or from other immunocytes, much as dendritic and natural killer cells are the first responders to acute T_{reg} ablation in diphthera toxin receptor models[46,47]. The presence of WT T_{reg} cells in mixed BMC mice or in heterozygous females would restore the negative feedback, thus reinstating better T_{reg} function, evoking the 'infectious tolerance' concept[48], with the presence of normal suppressors improving the tone and function of defective ones. Pathway and ontology analysis revealed no clear match, except for effects on a few IFN- and TNF-induced genes. There was, on the other hand, an intriguing anti-correlation for genes over-expressed in tumor-infiltrating T_{reg} cells[49], where $CD4^+$ T cells from patients with IPEX syndrome turned off much of the tumor T_{reg} signature.

This principal effect cutting across both T_{reg} and T_{conv} cells evokes the debated hypothesis that FOXP3 deficiency has an intrinsic effect in T_{conv} cells, associated with the transient induction of FOXP3 in activated T_{conv} cells[29–32]. The mixed chimeras and the mother–son pairs demonstrated that FOXP3 deficiency affected T_{conv} cells extrinsically. However, it remains an open question whether this disease signature contributes to pathology via further dampening of T_{reg} function or through T_{conv} defects.

That FOXP3 is not absolutely required for T_{reg} differentiation and homeostasis is now well established, with the existence of T_{reg}-like cells described repeatedly in mice and humans lacking FOXP3 (refs. [23–28]). These 'T_{reg} wannabes' were reported to maintain some T_{reg} features (self-reactivity, partial T_{reg} signature, activity at the *FOXP3* locus) while acquiring some T_{conv} characteristics (no in vitro anergy, cytokine expression). The unexpected insight emerging from our single-cell analysis of deficient humans and mice was the wide array of cells with T_{reg}-like characteristics and/or an active *FOXP3* locus. Rather than a single population of T_{reg} wannabes, several distinct populations were present: first, a large component of type B T_{reg} cells, with many T_{reg} transcriptional characteristics, but perturbed by the IPEX signature; and second (and most mysterious), the subset of type A T_{reg} cells that seemed almost unaffected relative to WT T_{reg} cells (full T_{reg} signature, active *FOXP3* and no IPEX signature), as if these cells had somehow become nonresponsive to the systemic influence, generating escape variants, for example by dampening receptors or signaling or because they reside in protected niches. But then, why was this the case only for T_{reg} cells?

These results show that it can be misleading to infer the transcriptional footprint of a TF and its mechanistic causality in disease, solely from the transcriptome of deficient cells, as it can be perturbed as here by dominant cell-extrinsic effects. When the extrinsic effects were blocked in the chimeras, the true core signature of *Foxp3* proved much smaller than the first analysis of the deficiency had suggested. IPEX syndrome is a rare disease and our cohorts were not powered to robustly detect clinical correlates of the patients' cellular and transcriptional characteristics. On the other hand, there was no obvious correlation between clinical severity indicators and the integrated gene expression metrics. This finding is congruent with the notion that molecular severity of the FOXP3 mutations in patients with IPEX syndrome only loosely correlates with clinical severity[19,50]; here, the null mutation in patient P1 did not determine the most severe disease.

There are some implications of our findings for therapeutic strategies in IPEX syndrome. Current management, when bone marrow transplantation is not an option, involves immunologic dampening via immunosuppressants. Our results might suggest harnessing those type A T_{reg} cells that are present in the patients by sustaining their homeostasis or complementing the small set of primary FOXP3 targets identified here. Indeed, the effect of the IL-2–anti-IL-2 treatment validates this proposition (it would be interesting to see whether combination of IL-2 with other activators of the core gene set, for example TNF family members, would further improve efficacy). Alternatively, damage might be avoided in patients by blocking the signaling mechanism that imparts the dominant IPEX signature to all T cells.

In conclusion, the new landscape of T_{reg} and T_{conv} cells revealed in patients with IPEX syndrome by single-cell analysis and their correction in mixed cell contexts, have opened a new perspective on the disease and on the role of FOXP3 and T_{reg} cells in immune homeostasis.

Online content

Any methods, additional references, Nature Research reporting summaries, source data, extended data, supplementary information, acknowledgements, peer review information; details of author contributions and competing interests; and statements of data and code availability are available at https://doi.org/10.1038/s41590-021-00910-8.

Received: 22 July 2020; Accepted: 26 February 2021;
Published online: 8 April 2021

References

1. Josefowicz, S. Z., Lu, L. F. & Rudensky, A. Y. Regulatory T cells: mechanisms of differentiation and function. *Annu. Rev. Immunol.* **30**, 531–564 (2012).
2. Wing, J. B., Tanaka, A. & Sakaguchi, S. Human $FOXP3^+$ regulatory T cell heterogeneity and function in autoimmunity and cancer. *Immunity* **50**, 302–316 (2019).
3. Panduro, M., Benoist, C. & Mathis, D. Tissue Tregs. *Annu. Rev. Immunol.* **34**, 609–633 (2016).
4. Hill, J. A. et al. Foxp3 transcription-factor-dependent and -independent regulation of the regulatory T cell transcriptional signature. *Immunity* **27**, 786–800 (2007).
5. Ferraro, A. et al. Interindividual variation in human T regulatory cells. *Proc. Natl Acad. Sci. USA* **111**, E1111–E1120 (2014).
6. Zemmour, D. et al. Single-cell gene expression reveals a landscape of regulatory T cell phenotypes shaped by the TCR. *Nat. Immunol.* **19**, 291–301 (2018).
7. Ono, M. Control of regulatory T-cell differentiation and function by T-cell receptor signalling and Foxp3 transcription factor complexes. *Immunology* **160**, 24–37 (2020).
8. Kwon, H. K., Chen, H. M., Mathis, D. & Benoist, C. Different molecular complexes that mediate transcriptional induction and repression by FoxP3. *Nat. Immunol.* **18**, 1238–1248 (2017).
9. Campbell, D. J. & Koch, M. A. Phenotypical and functional specialization of $FOXP3^+$ regulatory T cells. *Nat. Rev. Immunol.* **11**, 119–130 (2011).
10. Li, C. et al. TCR transgenic mice reveal stepwise, multi-site acquisition of the distinctive fat-Treg phenotype. *Cell* **174**, 285–299 (2018).
11. Dispirito, J. R. et al. Molecular diversification of regulatory T cells in nonlymphoid tissues. *Sci. Immunol.* **3**, eaat5861 (2018).
12. Miragaia, R. J. et al. Single-cell transcriptomics of regulatory T cells reveals trajectories of tissue adaptation. *Immunity* **50**, 493–504 (2019).
13. Powell, B. R., Buist, N. R. & Stenzel, P. An X-linked syndrome of diarrhea, polyendocrinopathy, and fatal infection in infancy. *J. Pediatr.* **100**, 731–737 (1982).
14. Ramsdell, F. & Ziegler, S. F. FOXP3 and scurfy: how it all began. *Nat. Rev. Immunol.* **14**, 343–349 (2014).
15. Barzaghi, F., Passerini, L. & Bacchetta, R. Immune dysregulation, polyendocrinopathy, enteropathy, X-linked syndrome: a paradigm of immunodeficiency with autoimmunity. *Front. Immunol.* **3**, 211 (2012).
16. d'Hennezel, E., Bin, D. K., Torgerson, T. & Piccirillo, C. A. The immunogenetics of immune dysregulation, polyendocrinopathy, enteropathy, X-linked (IPEX) syndrome. *J. Med. Genet.* **49**, 291–302 (2012).
17. Duclaux-Loras, R. et al. Clinical heterogeneity of immune dysregulation, polyendocrinopathy, enteropathy, X-linked syndrome: a French multicenter retrospective study. *Clin. Transl. Gastroenterol.* **9**, 201 (2018).
18. Gambineri, E. et al. Clinical, immunological, and molecular heterogeneity of 173 patients with the phenotype of immune dysregulation,

polyendocrinopathy, enteropathy, X-linked (IPEX) syndrome. *Front. Immunol.* **9**, 2411 (2018).
19. Barzaghi, F. et al. Long-term follow-up of IPEX syndrome patients after different therapeutic strategies: an international multicenter retrospective study. *J. Allergy Clin. Immunol.* **141**, 1036–1049 (2018).
20. Godfrey, V. L., Wilkinson, J. E. & Russell, L. B. X-linked lymphoreticular disease in the scurfy (*sf*) mutant mouse. *Am. J. Pathol.* **138**, 1379–1387 (1991).
21. Wan, Y. Y. & Flavell, R. A. Regulatory T-cell functions are subverted and converted owing to attenuated Foxp3 expression. *Nature* **445**, 766–770 (2007).
22. Van Gool, F. et al. A mutation in the transcription factor *Foxp3* drives T helper 2 effector function in regulatory T cells. *Immunity* **50**, 362–377 (2019).
23. Lin, W. et al. Regulatory T cell development in the absence of functional Foxp3. *Nat. Immunol.* **8**, 359–368 (2007).
24. Gavin, M. A. et al. Foxp3-dependent programme of regulatory T-cell differentiation. *Nature* **445**, 771–775 (2007).
25. Charbonnier, L. M. et al. Functional reprogramming of regulatory T cells in the absence of Foxp3. *Nat. Immunol.* **20**, 1208–1219 (2019).
26. Bacchetta, R. et al. Defective regulatory and effector T cell functions in patients with FOXP3 mutations. *J. Clin. Invest.* **116**, 1713–1722 (2006).
27. Otsubo, K. et al. Identification of FOXP3-negative regulatory T-like ($CD4^+CD25^+CD127^{low}$) cells in patients with immune dysregulation, polyendocrinopathy, enteropathy, X-linked syndrome. *Clin. Immunol.* **141**, 111–120 (2011).
28. Boldt, A. et al. Differences in FOXP3 and CD127 expression in Treg-like cells in patients with IPEX syndrome. *Clin. Immunol.* **153**, 109–111 (2014).
29. Walker, M. R. et al. Induction of FoxP3 and acquisition of T regulatory activity by stimulated human $CD4^+$. *J. Clin. Invest.* **112**, 1437–1443 (2003).
30. Gavin, M. A. et al. Single-cell analysis of normal and FOXP3-mutant human T cells: FOXP3 expression without regulatory T cell development. *Proc. Natl Acad. Sci. USA* **103**, 6659–6664 (2006).
31. Allan, S. E. et al. Activation-induced FOXP3 in human T effector cells does not suppress proliferation or cytokine production. *Int. Immunol.* https://doi.org/10.1093/intimm/dxm014 (2007).
32. McMurchy, A. N. et al. A novel function for FOXP3 in humans: intrinsic regulation of conventional T cells. *Blood* **121**, 1265–1275 (2013).
33. Zemmour, D. et al. *Flicr*, a long noncoding RNA, modulates Foxp3 expression and autoimmunity. *Proc. Natl Acad. Sci. USA* **114**, E3472–E3480 (2017).
34. Seddiki, N. et al. Expression of interleukin (IL)-2 and IL-7 receptors discriminates between human regulatory and activated T cells. *J. Exp. Med.* **203**, 1693–1700 (2006).
35. Pesenacker, A. M. et al. A regulatory T-cell gene signature is a specific and sensitive biomarker to identify children with new-onset type 1 diabetes. *Diabetes* **65**, 1031–1039 (2016).
36. Stoeckius, M. et al. Cell hashing with barcoded antibodies enables multiplexing and doublet detection for single cell genomics. *Genome Biol.* **19**, 224 (2018).
37. Stuart, T. et al. Comprehensive integration of single-cell data. *Cell* **177**, 1888–1902 (2019).
38. Bakke, A. C., Purtzer, M. Z. & Wildin, R. S. Prospective immunological profiling in a case of immune dysregulation, polyendocrinopathy, enteropathy, X-linked syndrome (IPEX). *Clin. Exp. Immunol.* **137**, 373–378 (2004).
39. Ziegler, S. F. FOXP3: of mice and men. *Annu. Rev. Immunol.* **24**, 209–226 (2006).
40. Fontenot, J. D., Gavin, M. A. & Rudensky, A. Y. Foxp3 programs the development and function of $CD4^+CD25^+$ regulatory T cells. *Nat. Immunol.* **4**, 330–336 (2003).
41. Hori, S., Nomura, T. & Sakaguchi, S. Control of regulatory T cell development by the transcription factor Foxp3. *Science* **299**, 1057–1061 (2003).
42. Malek, T. R. & Ashwell, J. D. Interleukin 2 upregulates expression of its receptor on a T cell clone. *J. Exp. Med.* **161**, 1575–1580 (1985).
43. Boyman, O. et al. Selective stimulation of T cell subsets with antibody-cytokine immune complexes. *Science* **311**, 1924–1927 (2006).
44. Samstein, R. M. et al. Foxp3 exploits a pre-existent enhancer landscape for regulatory T cell lineage specification. *Cell* **151**, 153–166 (2012).
45. Remedios, K. A. et al. The TNFRSF members CD27 and OX40 coordinately limit T_H17 differentiation in regulatory T cells. *Sci. Immunol.* **3**, eaau2042 (2018).
46. Kim, J. M., Rasmussen, J. P. & Rudensky, A. Y. Regulatory T cells prevent catastrophic autoimmunity throughout the lifespan of mice. *Nat. Immunol.* **8**, 191–197 (2007).
47. Sitrin, J. et al. Regulatory T cells control NK cells in an insulitic lesion by depriving them of IL-2. *J. Exp. Med.* **210**, 1153–1165 (2013).
48. Cobbold, S. & Waldmann, H. Infectious tolerance. *Curr. Opin. Immunol.* **10**, 518–524 (1998).
49. Plitas, G. et al. Regulatory T cells exhibit distinct features in human breast cancer. *Immunity* **45**, 1122–1134 (2016).
50. Gambineri, E. et al. Clinical and molecular profile of a new series of patients with immune dysregulation, polyendocrinopathy, enteropathy, X-linked syndrome: inconsistent correlation between forkhead box protein 3 expression and disease severity. *J. Allergy Clin. Immunol.* **122**, 1105–1112 (2008).

Methods

Mice. *Foxp3*$^{IRES\text{-}GFP}$/B6 (ref. [51]), *Foxp3*$^{\Delta EGFPiCre}$*xR26*^{YFP} (Δ*Foxp3*) and *Foxp3*$^{\Delta IRES\text{-}GFP}$/B6 (Δ*Foxp3*) mice on the C57Bl/6J background were maintained in our colony. Except when specified, 3-week-old male mice were used in this study (Supplementary Table 4). Mice were housed under specific-pathogen-free conditions and all experimentation was performed following the animal protocol guidelines of Harvard Medical School and Boston Children's Hospital (protocol 02954).

Human cohorts. Male IPEX and healthy donor whole-blood samples were obtained under protocols reviewed and approved by the local Institutional Review Boards at each center (Boston Children's Hospital 04-09-113R, Necker/Imagine C 15-13_CODECOH_AR, HMS IRB15-0504). Anonymized clinical data included age at onset and at blood sampling, clinical symptoms (enteropathy, diabetes, eczema, other autoimmune diseases and allergy), ancillary testing, treatments and long-term outcomes (Supplementary Table 1). Two independent cohorts of samples were profiled (Supplementary Table 1). Each cohort contains samples from both centers. Cohort 1 samples were processed in two different scRNA-seq runs and cohort 2 samples (replication cohort) were processed in three scRNA-seq runs. Samples from each cohort were sorted and sequenced as one batch for population RNA-seq (Supplementary Table 4).

Peripheral blood mononuclear cell isolation. Whole blood was collected in K_2 EDTA tubes and processed within a few hours. An equal volume of room-temperature PBS/2 mM EDTA was mixed into 15 ml of blood and carefully layered over 14 ml Ficoll Hypaque solution (GE Healthcare). After centrifugation for 30 min at 900*g* (with no break), at room temperature, the mononuclear cell layer was washed three times with excess HBSS (Gibco) for 10 min at 400*g* and resuspended in 2 ml of HBSS. The pellet was resuspended in 90% FBS-10% dimethylsulfoxide, 20 million cells ml^{-1}, 1 ml per vial, cooled progressively in isopropyl alcohol (Mr Frosty Freezing Container, Thermo Fisher Scientific) for 24 h and stored long term in liquid nitrogen.

Vials were thawed in 10 ml 10% FBS RPMI and cells were washed (500*g* for 5 min) and resuspended in FACS buffer (phenol red-free RPMI, 2% FBS, 0.1% azide and 10 mM HEPES, pH 7.9). After cell counting, samples were allocated for flow cytometry, population RNA-seq and scRNA-seq.

Flow cytometric profiling of IPEX and HDs. Cells were stained for flow cytometry in 100 µl of FACS buffer (phenol red-free DMEM, 2% FBS, 0.1% azide and 10 mM HEPES, pH 7.9) for 10 min with 10 µl FcBlock (supernatant of clone 2.4G2, ATCC HB-197, hybridoma cultures) and the following antibodies: CD3 BV605 or A700 (2 µl, OKT3, BioLegend), CD4 PerCP-Cy5.5 (2 µl, OKT4, BioLegend), CD25 PE-Cy7 (3 µl, BC96, BioLegend), CD127 A488 (3 µl, A019D5, BioLegend) and CD45RA PB (3 µl, HI100, BioLegend). After permeabilization/fixation for 2 h on ice, FOXP3 (APC, clone PCH101; BioLegend; 2 µl) and anti-HELIOS PE (PE, clone 22F6, BioLegend, 2 µl) staining was performed overnight at 4 °C in the dark in accordance with the manufacturer's instructions (eBioscience FOXP3/Transcription Factor Staining Buffer set). Data were recorded on an LSRII flow cytometer (BD Biosciences) and analyzed using FlowJo v.10. FOXP3 mean fluorescence intensity was normalized to the mean of HDs. To generate *t*-SNE plots from flow cytometry data, compensated, scaled data in the lymphocyte/singlet/CD3$^+$/CD4$^+$ gate were exported in .csv format from FlowJo. In R, the matrix containing CD3, CD4, FOXP3, HELIOS, CD25, CD127 and CD45RA expression was centered and scaled (by marker) before performing PCA using the prcomp function and *t*-SNE projections were calculated with package Rtsne (dims = 2, perplexity = 50, check_duplicates = F, pca = F, max_iter = 500).

Low-input RNA-seq of human and mouse samples. *RNA-seq libraries.* RNA-seq was performed with the standard ImmGen low-input protocol (www.immgen.org). Human T_{reg} and T_{conv} cells were sorted as DAPI$^-$CD4$^+$CD25^{hi}CD127^{lo} and CD4$^+$CD25$^-$CD127^{hi}, respectively on a Moflo Astrios Cell Sorter (Beckman Coulter). Mouse T_{reg} and T_{conv} cells were sorted as DAPI$^-$TCRβ$^+$CD4$^+$G/YFP$^+$CD25$^+$ and DAPI$^-$TCR$^+$CD4$^+$G/YFP$^-$, respectively. A total of 1,000 cells were double-sorted directly into 5 µl of lysis buffer (TCL Buffer (QIAGEN) supplemented with 1% 2-mercaptoethanol). Smart-seq2 libraries were prepared as previously described[52] with slight modifications. Briefly, total RNA was captured and purified on RNAClean XP beads (Beckman Coulter). Polyadenylated messenger RNA was then selected using an anchored oligonucleotide(dT) primer (50 –AAGCAGTGGTATCAACGCAGAGTACT30VN-30) and converted to complementary DNA via reverse transcription. First-strand cDNA was subjected to limited PCR amplification followed by Tn5 transposon-based fragmentation using the Nextera XT DNA Library Preparation kit (Illumina). Samples were then PCR amplified for 12 cycles using barcoded primers such that each sample carried a specific combination of eight base Illumina P5 and P7 barcodes for subsequent pooling and sequencing. Paired-end sequencing was performed on an Illumina NextSeq 500 using 2 × 38-bp reads with no further trimming.

Reads were aligned to the human genome (GENCODE GRCh38 primary assembly and gene annotations v.27) or to the mouse genome (mm10) with STAR 2.5.4a (https://github.com/alexdobin/STAR/releases). Ribosomal RNA gene annotations were removed from a .gtf (general transfer format) file. Gene-level quantification was calculated by featureCounts (http://subread.sourceforge.net/). Raw read counts tables were normalized by the median of ratios method with the DESeq2 package from Bioconductor (https://bioconductor.org/packages/release/bioc/html/DESeq2.html) and then converted to .gct and .cls format. Samples with fewer than 1 million uniquely mapped reads were excluded from normalization to mitigate the effect of poor quality samples on normalized counts. Genes with a minimum read count of five in all replicates of a population (31,448 human genes) were retained. A pseudo count of one was added and $\log_2$-transformed before quantile normalization. Quantile-normalized counts were converted back to a linear scale.

Signatures and indexes. IPEX signature genes were identified by computing the ratio of expression in T_{reg} or T_{conv} cells of all patients with IPEX syndrome versus all HDs (*P* value as a simple uncorrected Student's *t*-test). IPEX-up indices for each individual were calculated by selecting the 100 transcripts most over-expressed in IPEX T_{reg} cells overall (FC > 3), calculating their differential expression in T_{reg} cells of each patient relative to the mean of all HD T_{reg} cells and averaging the $\log_2$ of these FCs (and similarly for IPEX-down indices from transcripts with IPEX / HD < 0.37). The T_{reg} up indices were similarly calculated (expression in T_{reg} cells of each individual over mean expression in all HDs, average the $\log_2$ of these FCs).

Gene set enrichment analysis. Gene signatures were curated from published datasets (references in the signature name)[6]. Human and mouse T_{reg} signatures have been reported[4,5]. Data were downloaded from Gene Expression Omnibus (https://www.ncbi.nlm.nih.gov/geo/). Only datasets containing replicates were used. To reduce noise, genes with a CV between biological replicates <0.6 in either comparison groups were selected. Up- and downregulated transcripts were defined as having an FC in gene expression >1.5 or <0.6 and a Student's *t*-test *P* value <0.05. Other signatures were obtained from MSigDB C7 Immunologic signatures collection[53]. Gene set enrichment analysis with IPEX signature was performed using the hypergeometric distribution and type I error was controlled using false discovery rate. Signatures with *P* value <0.001 (all with false discovery rate < 6%) are reported.

scRNA-seq analysis of human PBMC samples. scRNA-seq was performed in several batches (different experiment dates in Supplementary Table 1 and Supplementary Fig. 4): two for cohort 1 and three for cohort 2. Cohort 1 samples were profiled with the 10X Genomics Single Cell 3′ Reagent kit (V2 chemistry), cohort 2 samples with the 10X Genomics Single Cell 3′ Reagent kit (V3 chemistry) and sample barcoding with DNA-tagged antibodies ('hashtagging')[36]. See also Supplementary Table 4.

Cell sorting and pooling using cell hashtagging. Cells were stained in 100 µl of FACS buffer (phenol red-free RPMI, 2% FBS, 0.1% azide and 10 mM HEPES, pH 7.9) for 10 min with 10 µl FcBlock (homemade) and the following antibodies: CD3 A700 (2 µl, OKT3, BioLegend), CD4 PerCP-Cy5.5 (2 µl, OKT4, BioLegend), CD25 PE-Cy7 (3 µl, BC96, BioLegend), CD127 A488 (3 µl, A019D5, BioLegend), 5 µl (2.5 µg) of a unique hashtag antibody (TotalSeq-A0251 to A0258 anti-human hashtag 1 to 8 antibody). A total of 8,000 DAPI$^-$CD3$^+$CD4$^+$ T cells were single sorted using the Moflo Astrios Cell Sorter (70-µm nozzle, Beckman Coulter) in 30 µl of PBS–BSA 0.1%. Samples with different hashtag antibodies were sorted in the same tube and the total volume was adjusted to 30 µl.

scRNA-seq libraries. Cells were encapsulated in one channel per sample (cohort 1 runs) or in one channel per pool (cohort 2 samples) of a 10X Genomics Chromium instrument and libraries were constructed with the Single Cell 3′ Reagent kit (V2 for cohort 1 and V3 chemistry for cohort 2) (https://support.10xgenomics.com/single-cell-gene-expression/library-prep/). Libraries were sequenced on the NextSeq 500 platform (28/8/0/91, Read1/i7/i5/Read2). Gene counts were obtained by aligning reads to the hg38 transcriptome using CellRanger software (v.3.0.2) (10X Genomics) using default parameters.

Hashtag libraries. Hashtag libraries were made separately as described by Stoeckius et al.[36] (https://citeseq.files.wordpress.com/2019/02/cell_hashing_protocol_190213.pdf). In brief, at the cDNA amplification step in the Single Cell 3′ Reagent kit protocol, the yield of HTO (Hashtag Oligo) products was increased using an 'additive' primer to cDNA PCR. Hashtag-derived cDNAs (<180 bp) and mRNA-derived cDNAs (>300 bp) were then separated using 0.6× SPRI bead selection. The supernatant contained the hashtag-derived cDNA that was purified with two rounds of 2× SPRI beads. The sequencing oligonucleotides were added by PCR which also amplified the Hashtag library. Libraries were sequenced on the NextSeq 500 platform (28/8/0/91, Read1/i7/i5/Read2).

Hashtag count matrices were obtained from CITE-Seq-Count package (https://zenodo.org/record/2590196) using default parameters. Each droplet from the gene count matrix was matched to a hashtag using the HTODemux function from the Seurat v.3.1.2 package. Doublets (droplets with two hashtags) were excluded and cells were assigned to the max hashtag signal. The hashtag count data were also analyzed by *t*-SNE for a visual check (clear separated clusters for each hashtag). All single cells from the gene count matrix were matched unambiguously to a single hashtag (and therefore their original donor).

Quality control. Cells with fewer than 500 (V2) or 1,000 (V3) counts were excluded. Dead cells with more than 10% of counts mapping to mitochondrial genes were excluded. Doublets were excluded using scrublet (doublet score > 0.3)[54]. Finally, cells were annotated using singleR[55] with the BluePrintEncode reference data. Cells that were not annotated as $CD4^+$ cells were excluded.

Batch correction. Batch correction was performed using the Integration method in Seurat V3 as described by Stuart et al.[37] and used only to visualize the whole dataset with UMAP. Cohort 1 and cohort 2 samples were first normalized independently using the SCTransform function in Seurat V3 with parameters to regress out the following variables: experiment date, percent of mitochondrial gene mapping and the number of counts of each single cell (SCTransform(vars.to.regress = c('experiment_date','percent_mito', 'nCount_RNA'), variable.features.*n* = 500). Integration of both cohorts together was then performed using the top 500 variable genes in both cohorts. Each cohort was projected into a common CCA space. Anchors (robust pairwise cell correspondence between datasets) were found by knn and snn graphs (FindIntegrationAnchors function) and they were used to transform the data in an integrated space. We used this integrated space, reduced to the top 40 principal components (PCs) by PCA to visualize with UMAP[56]. The number of significant PCs was determined by comparison to PCA over a randomized matrix, as described by Klein et al.[57]. Of note, another batch correction method, scAlign[58] produced similar results.

Clustering. We found shared clusters across HD and IPEX samples ('regressing out the IPEX effect') using the Integration method in Seurat V3 as described by Stuart et al.[37]. This time, each sample was first normalized independently using the SCTransform function in Seurat V3 with parameters to regress out the following variables: percent of mitochondrial gene mapping and the number of counts of each single cell (SCTransform(vars.to.regress = c('percent_mito', 'nCount_RNA'), variable.features.*n* = 500). Integration of samples together was then performed using the top 500 variable genes. Each cohort was projected into a common reciprocal PCA space. Anchors (robust pairwise cell correspondence between samples) were found by knn and snn graphs (FindIntegrationAnchors) and they were used to transform the data in a shared space. We used this shared space and reduced it to the top 32 most significant PCs by PCA for clustering[56]. In this space, a shared nearest neighbor graph was constructed from a *k*-nearest-neighbor graph ($k = 20$) by pruning cell–cell edges with less than 1/15 neighbor overlap. Community detection using the Louvain algorithm at a resolution of 0.5 found 11 clusters. Automated annotation using singleR with the BluePrintEncode reference data and manual annotation using canonical markers (see Figs. 3 and 4 and Supplementary Fig. 7) clearly distinguished resting T_{conv}, activated T_{conv} and T_{reg} cells. Type A IPEX T_{reg} cells were defined as T_{reg} cells in IPEX samples that overlapped with the HD T_{reg} cells (top left quadrant in UMAP). Type B IPEX T_{reg} cells were all the others.

Differential gene expression taking into account technical and biological variables. We used limma-trend[59], as benchmarked elsewhere[60]. Briefly, default trimmed mean of *m*-value normalization from edgeR was applied to the SCTransform-corrected count matrix. A linear model was fitted to the data using a contrast matrix including confounding variables (lmFit function). To find cluster-defining markers shared across HD and IPEX samples (regressing out the IPEX effect) or markers between type A and B IPEX T_{reg} cells, the contrast matrix contained the sample origin as a confounding variable. Empirical Bayes method was then used to estimate the overall trend of gene expression variance and adjust the genewise residual variances toward this global trend (less variance for genes trending high and more for those trending low) (eBayes(trend = TRUE, robust = TRUE)). The TopTable function was then used to extract the differential statistics corresponding to the contrast of interest. Benjamini–Hochberg correction was applied to control type I error. Adjusted *P* values <0.05 were deemed significant.

Analysis of active and inactive X chromosomes carrying FOXP3 mutation in IPEX mother cells. scRNA-seq was performed as described in the previous paragraph of peripheral blood $CD4^+$ cells from two mothers of patients with IPEX syndrome: P5's and B2's mothers (see also Supplementary Table 4). A total of 919 cells from P5's mother and 1,961 cells from B2's mother were analyzed after quality control. Because X inactivation in females occurs randomly during embryogenesis, a mix of FOXP3-deficient (Δ*FOXP3*) and -proficient (WT) T cells are present in IPEX mothers (female carriers). We sought to identify them using X chromosome variants in patients with IPEX syndrome (P5 and B2) and their mother. For each sample, we started from the CellRanger bam file (possorted_genome_bam.bam), converted it back to fastq (bamtofastq-1.2.0 --nthreads 32 possorted_genome_bam.bam fastq) and aligned the reads using STAR v.020201 ($STAR_PATH/STAR --genomeDir GRCh38_fasta --readFilesIn $READ2 --runThreadN 32 --outSAMtype BAM SortedByCoordinate) to the GRCh38 human genome to be able to run bcftools. Variants on the X chromosome were called using bcftools v.1.10.2 (bcftools mpileup -Q 30 -A -x -Ou --threads 32 -r X -f $FASTA_REF Aligned.sortedByCoord.out.bam | bcftools call -mv -Ov -o calls.vcf). We then used scAlleleCount (https://github.com/barkasn/scAlleleCount) to obtain for each single cell the reference and alternate allele count for the X chromosome variants (cell × variant matrices). We then constructed the IPEX (ΔFoxp3) X chromosome haplotype (present in sons and carrying the FOXP3 mutation) and the complementary haplotype (present in the mother and carrying the 'WT' FOXP3). To do so, we only focused on high confidence variants that were present both in the mother and in the son with a quality >10. We then filtered out the uninformative variants (variants with 100% allelic fraction in the mothers, present in both X chromosome). Using the son as a reference, we constructed the IPEX (ΔFoxp3) X chromosome haplotype and the complementary WT haplotype. For each single cell the distances to the ΔFoxp3 and the WT haplotypes were computed and we used the ratio as a score to identify whether the active X chromosome harbored WT or mutant FOXP3 (Fig. 6). There was no evidence of X chromosome recombination in mothers. In IPEX, 3 out of 815 were assigned to an active WT chromosome in P5 and 2 out of 1,612 in B2 (misclassification or microchimerism). Overall, 469 P5's mother's cells and 513 B2's mother's cells were confidently assigned to a ΔFoxp3 or WT active X chromosome (Fig. 6).

Bone marrow chimeras. C57BL/6J mice were irradiated with 10 Gy and reconstituted with equal proportions of congenically labeled T cell-depleted bone marrow cells from WT (*Foxp3*$^{ires\text{-}gfp}$x*CD45.1*) and Δ*Foxp3* (*Foxp3*$^{\Delta EGFPiCre}$*xR26*$^{-YFP}$*xCD45.1/CD45.2*) mice. Bone marrow cells were collected from femurs, tibias and hip bones from two WT and two Δ*Foxp3* male mice. After red blood cell lysis with ACK for 1 min at 4 °C, T cells were depleted; bone marrow single-cell suspensions were incubated with 20 µl of biotinylated anti-CD3e antibodies (OKT3, BioLegend) for 10 min in 2 ml of MACS buffer (PBS, FBS 0.5%, EDTA 2 mM), washed and then incubated with 200 µl of streptavidin beads (Dynabeads Biotin Binder, 11047; Thermo Fisher Scientific) for 20 min in 5 ml of MACS buffer. Isolation of the $CD3^-$ population was performed after three magnet incubations for 2 min. A total of 4 million cells (2 million WT and 2 million Δ*Foxp3*) were injected intravenously in each mouse. Mice were treated for 2 weeks with trimethoprim–sulfamethoxazol and analyzed 10 weeks later.

scRNA-seq profiling of $CD4^+$ T splenocytes cells in mice and bone marrow chimeras. scRNA-seq profiling of mouse samples was performed in two experiments (see also Supplementary Table 4). In the first experiment, two WT mice and two Δ*Foxp3* mice were profiled with the 10X Genomics Single Cell 3′ Reagent kit (V2 chemistry) at one channel per mouse. In the second experiment, $CD4^+$ T cells from three BMCs, two WT mice and two Δ*Foxp3* mice were pooled with sorted WT T_{reg} and T_{conv} cells as controls and profiled with the 10X Genomics Single Cell 3′ Reagent kit (V3 chemistry). Both experiments were analyzed together, after batch correction.

scRNA-seq libraries. Spleens were collected. After red blood cell lysis with ACK for 1 min at 4 °C, 30% of splenocytes (~30 million cells) were stained for sorting by flow cytometry in 200 µl of FACS buffer (PBS, 2% BSA) for 20 min at 4 °C in the dark, with the following antibodies (1:100 dilution): 100 µl of FcBlock (homemade), TCRβ PE-Cy7 (H57–597; BioLegend), CD4 PerCP-Cy5.5 (GK1.5; BioLegend), DAPI, CD45.1 PE-Cy7 (clone A20, BioLegend) and CD45.2 A700 (clone 104, BioLegend). Whole Δ*Foxp3* and WT mouse $CD4^+$ T cells were sorted as $DAPI^-TCR\beta^+CD4^+$. For BMC mice, Δ*Foxp3* $CD4^+$ T cells were sorted as $DAPI^-TCR\beta^+CD4^+CD45.1^+CD45.2^+$. WT $CD4^+$ T cells were sorted as $DAPI^-TCR\beta^+CD4^+CD45.1^+CD45.2^-$. A total of 40,000–50,000 cells were single sorted in 50 µl of PBS–BSA 0.05% and the total volume was adjusted to a concentration of 1,000 cells μl^{-1} after cell counting with a hemocytometer. In the first experiment, cells from each tube were encapsulated in one channel of a 10X Chromium instrument and libraries were constructed with a Single Cell 3′ Reagent kit (V2 chemistry) (https://support.10xgenomics.com/single-cell-gene-expression/library-prep/). In the second experiment, to pool samples for cell hashtagging, cells were sorted in 50 µl of PBS–BSA 0.05% in two different tubes (5,000 to 10,000 sorted cells for each sample) for sequencing in two 10X lanes and libraries were constructed with a Single Cell 3′ Reagent kit (V3 chemistry) (see also Supplementary Table 4).

Libraries were sequenced on the NextSeq 500 platform (28/8/0/91, Read1/i7/i5/Read2). Gene counts were obtained by aligning reads to the mm10 transcriptome using CellRanger software (v.3.0.2) (10X Genomics) using default parameters. The mm10 transcriptome was complemented with the transgene sequences of *Ires–Gfp*, *Yfp* and *Cre* to map reads to the *Foxp3* locus in *Foxp3*$^{ires\text{-}gfp}$/B6 and *Foxp3*$^{\Delta GFPiCre}$*xR26*$^{-YFP}$/B6 mice. In *Foxp3*$^{ires\text{-}gfp}$/B6 mice, *Foxp3* locus expression was calculated as the sum of reads mapping to *Foxp3* and *Ires–Gfp*. In *Foxp3*$^{\Delta GFPiCre}$*xR26*$^{-YFP}$/B6 mice, *Foxp3* locus expression was calculated as the sum of reads of mapping to *Foxp3* and *Cre* (not GFP because the read 2 length was not long enough to reach the *GFP* sequence).

Hashtag libraries. See similar section in the human analysis (scRNA-seq analysis of human PBMC samples)

Quality control. See similar section in the human analysis (scRNA-seq analysis of human PBMC samples). SingleR was used with the ImmGen reference data.

Batch correction. Experiment 1 and 2 were profiled with 10X V2 and V3, respectively. Batch correction was performed using the Integration method in

Seurat V3 as described by Stuart et al.[37] and only used to visualize the whole mouse dataset with UMAP. Experiment 1 (10X V2) and experiment 2 (10X V3) samples were first normalized independently using the SCTransform function in Seurat V3 with parameters to regress out the following variables: percent of mitochondrial gene mapping and the number of counts of each single cell (SCTransform(vars.to.regress = c('percent_mito', 'nCount_RNA'), variable.features.n = 500). Integration of both experiments together was then performed using the top 500 variable genes in both cohorts. Each cohort was projected into a common CCA space. Anchors were found by knn and snn graphs (FindIntegrationAnchors) and they were used to transform the data in an integrated space. We used this integrated space, reduced to the top 50 PCs by PCA to visualize with UMAP. The number of significant PCs was determined by comparison to PCA over a randomized matrix, as described by Klein et al.[57].

Clustering. We found shared clusters across samples using the Integration method in Seurat V3 as described by Stuart et al.[37]. This time, each sample was first normalized independently using the SCTransform function in Seurat V3 with parameters to regress out the following variables: percent of mitochondrial gene mapping and the number of counts of each single cell (SCTransform(vars.to.regress = c('percent_mito', 'nCount_RNA'), variable.features.n = 500). Integration of samples together was then performed using the top 500 variable genes. Each cohort was projected into a common CCA space. Anchors were found by knn and snn graphs (FindIntegrationAnchors) and they were used to transform the data in a shared space. We used this shared space and reduced it to the top 28 most significant PCs by PCA for clustering. In this space, a shared nearest neighbor graph was constructed from a k-nearest-neighbor graph (k = 20) by pruning cell–cell edges with less than 1/15 neighbor overlap. Community detection using the Louvain algorithm at a resolution of 1.5 found 11 clusters. Automated annotation using singleR with the ImmGen reference data and manual annotation using canonical markers (Supplementary Fig. 9) clearly distinguished resting T_{conv}, activated T_{conv}, resting T_{reg} and activated T_{reg} cells. Type A IPEX T_{reg} cells were defined as T_{reg} cells in IPEX samples that overlapped with the HD T_{reg} cells. Type B IPEX T_{reg} cells were all the others.

Differential gene expression. See similar section in the human analysis (scRNA-seq analysis of human PBMC samples).

Flow cytometry analysis of mouse cells. A single-cell suspension was obtained from murine splenocytes after mechanical dissociation through a 40-µm strainer using a syringe plunger. Red blood cell lysis was performed using 1 ml of ACK (Lonza) for 1 min on ice. After careful perfusion with 5 ml of sterile PBS, the right lung was collected, minced and enzymatically dissociated in RPMI containing 0.5 mg ml^{-1} DNase I (Sigma), 1.5 mg ml^{-1} collagenase IV (Sigma), 10% FBS and 2 mM EDTA with constant stirring at 37 °C for 35 min. Single-cell suspensions were then filtered through a 70-µm strainer and washed twice with 10% FBS + 2 mM EDTA RPMI buffer. Lung CD4$^+$ T cells were enriched before staining and sorting, using negative magnetic selection (Stem Cell).

After Fc blocking, antibody staining was performed in ice-cold buffer (DMEM without phenol red, 2% FBS, 1 mM EDTA) for 15 min at a dilution of 1:100 with antibodies against CD45 (30-F11; BioLegend), CD4 (GK1.5; BioLegend), CD25 (PC61; BioLegend), TCRβ (H57-597; BioLegend), CD19 (SA011F11; BioLegend) and DAPI. Stained samples were then analyzed by flow cytometry on a BD Fortessa or Symphony analyzers and processed using FlowJo v.10 (Tree Star).

IL-2 treatment. The IL-2–anti-IL-2 monoclonal antibody complex was formed by incubating 1 µg of recombinant mouse IL-2 (BioLegend) and 9 µg of purified anti-mouse IL-2 monoclonal antibody (clone JES6-1A12) (Bio-X-cell) for 20 min at 37 °C in sterile PBS. This complex was administered for 3 d consecutively per week starting at day 11 after birth on the basis of 40 ng of IL-2 per gram weight of treated mice.

Statistics. Analysis was conducted using R-3.6.2. Heat maps were made with Morpheus (https://software.broadinstitute.org/morpheus) or the pheatmap package in R (https://cran.r-project.org/web/packages/pheatmap/index.html). All other plots were made with ggplot2 (ref. [61]). Statistical tests are described in their respective method section.

Reporting Summary. Further information on research design is available in the Nature Research Reporting Summary linked to this article.

Data availability

The data reported in this paper have been deposited in the Gene Expression Omnibus database under SuperSeries accession no. GSE168492: GSE166866 (human population RNA-seq), GSE166860 (mouse population RNA-seq), GSE167976 (human scRNA-seq) and GSE167575 (mouse scRNA-seq).

References

51. Bettelli, E. et al. Reciprocal developmental pathways for the generation of pathogenic effector T_H17 and regulatory T cells. *Nature* **441**, 235–238 (2006).
52. Picelli, S. et al. Full-length RNA-seq from single cells using Smart-seq2. *Nat. Protoc.* **9**, 171–181 (2014).
53. Godec, J. et al. Compendium of immune signatures identifies conserved and species-specific biology in response to inflammation. *Immunity* **44**, 194–206 (2016).
54. Wolock, S. L., Lopez, R. & Klein, A. M. Scrublet: computational identification of cell doublets in single-cell transcriptomic data. *Cell Syst.* **8**, 281–291 (2019).
55. Aran, D. et al. Reference-based analysis of lung single-cell sequencing reveals a transitional profibrotic macrophage. *Nat. Immunol.* **20**, 163–172 (2019).
56. Becht, E. et al. Dimensionality reduction for visualizing single-cell data using UMAP. *Nat. Biotechnol.* https://doi.org/10.1038/nbt.4314 (2018).
57. Klein, A. M. et al. Droplet barcoding for single-cell transcriptomics applied to embryonic stem cells. *Cell* **161**, 1187–1201 (2015).
58. Johansen, N. & Quon, G. scAlign: a tool for alignment, integration, and rare cell identification from scRNA-seq data. *Genome Biol.* **20**, 166 (2019).
59. Law, C. W., Chen, Y., Shi, W. & Smyth, G. K. Voom: precision weights unlock linear model analysis tools for RNA-seq read counts. *Genome Biol.* **15**, R29 (2014).
60. Soneson, C. & Robinson, M. D. Bias, robustness and scalability in single-cell differential expression analysis. *Nat. Methods* **15**, 255–261 (2018).
61. Wickham, H. *ggplot2: Elegant Graphics for Data Analysis* (Springer-Verlag, 2009).

Acknowledgements

We thank M. Levings and A. Rudensky for insightful discussions and K. Hattori, C. Araneo, K. Seddu and the Klarman Cell Observatory team for help with mice, cell sorting and single-cell profiling. This work was funded by grants from the National Institutes of Health to C. Benoist and D.M. (AI116834 and AI150686), T.A.C. (AI085090) and L.M.C. (AI153174); the Institut National de la Santé et Recherche Médicale; the European Union Seventh Framework (269037 and 261387) and Horizon 2020 (693762); and the Agence Nationale pour la Recherche (Investissement d'Avenir ANR-10-IAHU-01) to I.A., E.S., M.D., J.L., M.C., B.N., F.R.L., F.R. and N.C.B. J.L. was supported by an INSERM Poste d'Accueil and an Arthur Sachs scholarship.

Author contributions

D.Z., L.M.C., J.L. and M.B. performed the experiments; E.S., S.K., M.D., S.B., J.Z., K.C., B.N., M.I.G.L., F.R., N.C.B., F.R.L., M.C., I.A., T.A.C., L.M.C. and C. Bruganara provided samples and discussed interpretations; D.Z., L.M.C., J.L., T.A.C., I.A., C. Benoist and D.M. designed the study and analyzed and interpreted the data; D.Z., J.L., C. Benoist and D.M. wrote the manuscript.

Competing interests

The authors declare no competing interests.

Additional information

Supplementary information The online version contains supplementary material available at https://doi.org/10.1038/s41590-021-00910-8.

Correspondence and requests for materials should be addressed to D.M. or C.B.

Peer review information *Nature Immunology* thanks Fred Ramsdell and the other, anonymous, reviewer(s) for their contribution to the peer review of this work. Zoltan Fehervari was the primary editor on this article and managed its editorial process and peer review in collaboration with the rest of the editorial team.

Reprints and permissions information is available at www.nature.com/reprints.

nature research

Corresponding author(s): Christophe Benoist

Last updated by author(s): Feb 20, 2021

Reporting Summary

Nature Research wishes to improve the reproducibility of the work that we publish. This form provides structure for consistency and transparency in reporting. For further information on Nature Research policies, see our Editorial Policies and the Editorial Policy Checklist.

Statistics

For all statistical analyses, confirm that the following items are present in the figure legend, table legend, main text, or Methods section.

n/a	Confirmed
☐	☒ The exact sample size (*n*) for each experimental group/condition, given as a discrete number and unit of measurement
☐	☒ A statement on whether measurements were taken from distinct samples or whether the same sample was measured repeatedly
☐	☒ The statistical test(s) used AND whether they are one- or two-sided *Only common tests should be described solely by name; describe more complex techniques in the Methods section.*
☐	☒ A description of all covariates tested
☐	☒ A description of any assumptions or corrections, such as tests of normality and adjustment for multiple comparisons
☐	☒ A full description of the statistical parameters including central tendency (e.g. means) or other basic estimates (e.g. regression coefficient) AND variation (e.g. standard deviation) or associated estimates of uncertainty (e.g. confidence intervals)
☐	☒ For null hypothesis testing, the test statistic (e.g. *F*, *t*, *r*) with confidence intervals, effect sizes, degrees of freedom and *P* value noted *Give P values as exact values whenever suitable.*
☒	☐ For Bayesian analysis, information on the choice of priors and Markov chain Monte Carlo settings
☐	☒ For hierarchical and complex designs, identification of the appropriate level for tests and full reporting of outcomes
☐	☒ Estimates of effect sizes (e.g. Cohen's *d*, Pearson's *r*), indicating how they were calculated

Our web collection on statistics for biologists contains articles on many of the points above.

Software and code

Policy information about availability of computer code

Data collection	no software used
Data analysis	R-3.6.2; CellRanger software (v3.0.2) ; CITE-Seq-Count package (https://zenodo.org/record/2590196; Seurat v3.1.2 ; limma-trend; STAR 2.5.4a ; bamtofastq-1.2.0 ; bcftools v1.10.2 ; scAlleleCount ; pheatmap; ggplot2

For manuscripts utilizing custom algorithms or software that are central to the research but not yet described in published literature, software must be made available to editors and reviewers. We strongly encourage code deposition in a community repository (e.g. GitHub). See the Nature Research guidelines for submitting code & software for further information.

Data

Policy information about availability of data

All manuscripts must include a data availability statement. This statement should provide the following information, where applicable:

- Accession codes, unique identifiers, or web links for publicly available datasets
- A list of figures that have associated raw data
- A description of any restrictions on data availability

The data reported in this paper will be deposited in the Gene Expression Omnibus (GEO) database under accession no. GSE166866 (human RNAseq), GSE166860 (mouse RNAseq), GSEXXX (human scRNAseq), GSEXXX (mouse scRNAseq).

Field-specific reporting

Please select the one below that is the best fit for your research. If you are not sure, read the appropriate sections before making your selection.

☒ Life sciences ☐ Behavioural & social sciences ☐ Ecological, evolutionary & environmental sciences

For a reference copy of the document with all sections, see nature.com/documents/nr-reporting-summary-flat.pdf

Life sciences study design

All studies must disclose on these points even when the disclosure is negative.

Sample size	15 healthy donors and 15 IPEX patients (all available samples) (table S1); 4 WT and 4 deltaFoxp3 mice for scRNAseq analysis (for replications)
Data exclusions	Cells that did not pass standard single cell QC (see methods): not T cells, doublets, low-coverage (<500 UMIS)
Replication	biological replicates, technical replicates, all concordant.
Randomization	not applicable, comparison of disease vs healthy
Blinding	not applicable, comparison of disease vs healthy

Reporting for specific materials, systems and methods

We require information from authors about some types of materials, experimental systems and methods used in many studies. Here, indicate whether each material, system or method listed is relevant to your study. If you are not sure if a list item applies to your research, read the appropriate section before selecting a response.

Materials & experimental systems

n/a	Involved in the study
☐	☒ Antibodies
☒	☐ Eukaryotic cell lines
☒	☐ Palaeontology and archaeology
☐	☒ Animals and other organisms
☐	☒ Human research participants
☐	☒ Clinical data
☒	☐ Dual use research of concern

Methods

n/a	Involved in the study
☒	☐ ChIP-seq
☐	☒ Flow cytometry
☒	☐ MRI-based neuroimaging

Antibodies

Antibodies used	All from accessible commercial sources HUMAN: CD3 BV605 or A700 (OKT3, Biolegend), CD4 PerCP-Cy5.5 (OKT4, Biolegend), CD25 PE-Cy7 (BC96, Biolegend), CD127 A488 (A019D5, Biolegend), CD45RA PB (HI100, Biolegend) FOXP3 (APC, clone PCH101; Biolegend) and anti-HELIOS PE (PE, clone 22F6, Biolegend) (TotalSeqTM- A0251 to A0258 anti-human Hashtag 1 to 8 Antibody). MOUSE: TCRb PE-Cy7 (H57-597;BioLegend), CD4 PerCPCy5.5 (GK1.5; BioLegend), CD4 BV-605 (clone RM4-5, Biolegend,), CD25 PE-Cy7 (clone PC61, Biolegend), CD45.1 PE-Cy7 (clone A20, Biolegend), CD45.2 A700 (clone 104, Biolegend), CD19 (SA011F11; BioLegend).
Validation	validated by the manufacturer and in comparison to negative/isotype controls.

Animals and other organisms

Policy information about studies involving animals; ARRIVE guidelines recommended for reporting animal research

Laboratory animals	Foxp3ires-gfp/B6 and Foxp3-delta-EGFPiCrexR26-YFP (delta-Foxp3) mice ; C57BL/6J
Wild animals	no
Field-collected samples	no
Ethics oversight	animal protocol guidelines of Harvard Medical School and Boston Children's Hospital (HMS IACUC protocol 02954).

Note that full information on the approval of the study protocol must also be provided in the manuscript.

Human research participants

Policy information about studies involving human research participants

Population characteristics	Anonymized clinical data include age at onset and at blood sampling, clinical symptoms (enteropathy, diabetes, eczema, other autoimmune diseases and allergy), ancillary testing, treatments and long-term outcomes (Table S1). Two independent cohorts of samples were profiled (Table S1).
Recruitment	all available samples at 2 clinical centers: Boston Children's Hospital, Necker/Imagine
Ethics oversight	Informed consent obtained following the Institutional Review Boards : Boston Children's Hospital 04-09-113R, Necker/ Imagine C 15-13_CODECOH_AR, HMS IRB15- 0504).

Note that full information on the approval of the study protocol must also be provided in the manuscript.

Clinical data

Policy information about clinical studies

All manuscripts should comply with the ICMJE guidelines for publication of clinical research and a completed CONSORT checklist must be included with all submissions.

Clinical trial registration	NA
Study protocol	NA
Data collection	NA
Outcomes	NA

Flow Cytometry

Plots

Confirm that:

☒ The axis labels state the marker and fluorochrome used (e.g. CD4-FITC).

☒ The axis scales are clearly visible. Include numbers along axes only for bottom left plot of group (a 'group' is an analysis of identical markers).

☒ All plots are contour plots with outliers or pseudocolor plots.

☒ A numerical value for number of cells or percentage (with statistics) is provided.

Methodology

Sample preparation	frozen human peripheral blood mononuclear cells; mouse splenocytes
Instrument	LSRII flow cytometer (BD Biosciences)
Software	FlowJo v10
Cell population abundance	Lymphocyte/Singlet/CD3+/CD4+ gate 10% of PBMC
Gating strategy	Lymphocyte/Singlet/CD3+/CD4+ gate

☒ Tick this box to confirm that a figure exemplifying the gating strategy is provided in the Supplementary Information.