ORIGINAL ARTICLE

immunology
Insights into immunity in health and disease

Distinguishing activated T regulatory cell and T conventional cells by single-cell technologies

Julia Reinhardt[1] | Virag Sharma[1,2] | Antigoni Stavridou[1] | Annett Lindner[1,2] | Susanne Reinhardt[3] | Andreas Petzold[3] | Mathias Lesche[3] | Fabian Rost[1,4] | Ezio Bonifacio[1,2] | Anne Eugster[1]

[1]Center for Regenerative Therapies Dresden, Faculty of Medicine, TU Dresden, Dresden, Germany

[2]German Center for Diabetes Research (DZD), Faculty of Medicine, Paul Langerhans Institute Dresden of Helmholtz Centre Munich at University Clinic Carl Gustav Carus of TU Dresden, Dresden, Germany

[3]Center for Molecular and Cellular Bioengineering (CMCB), DRESDEN-concept Genome Center, Technische Universität Dresden, Dresden, Germany

[4]Center for Information Services and High-Performance Computing (ZIH), TU Dresden, Dresden, Germany

Correspondence
Anne Eugster, Center for Regenerative Therapies Dresden, Faculty of Medicine, TU Dresden, Dresden, Germany.
Email: anne.eugster@tu-dresden.de

Senior author: Anne Eugster

Abstract

Resting conventional T cells (Tconv) can be distinguished from T regulatory cells (Treg) by the canonical markers FOXP3, CD25 and CD127. However, the expression of these proteins alters after T-cell activation leading to overlap between Tconv and Treg. The objective of this study was to distinguish resting and antigen-responsive T effector (Tconv) and Treg using single-cell technologies. CD4$^+$ Treg and Tconv cells were stimulated with antigen and responsive and non-responsive populations processed for targeted and non-targeted single-cell RNAseq. Machine learning was used to generate a limited set of genes that could distinguish responding and non-responding Treg and Tconv cells and which was used for single-cell multiplex qPCR and to design a flow cytometry panel. Targeted scRNAseq clearly distinguished the four-cell populations. A minimal set of 27 genes was identified by machine learning algorithms to provide discrimination of the four populations at >95% accuracy. In all, 15 of the genes were validated to be differentially expressed by single-cell multiplex qPCR. Discrimination of responding Treg from responding Tconv could be achieved by a flow cytometry strategy that included staining for CD25, CD127, FOXP3, IKZF2, ITGA4, and the novel marker TRIM which was strongly expressed in Tconv and weakly expressed in both responding and non-responding Treg. A minimal set of genes was identified that discriminates responding and non-responding CD4$^+$ Treg and Tconv cells and, which have identified TRIM as a marker to distinguish Treg by flow cytometry.

KEYWORDS
activation, CD4 cell, T cell, transcriptomics, Treg

INTRODUCTION

T-cell responses to antigen (Ag) are an essential component of the adaptive immune response. The type of T-cell response is classified by the cell type, transcription factors and cytokine production, and is influenced by the state of the antigen presenting cells, the nature of the Ag target and the environment where Ag presentation takes place. A key aspect of the CD4[+] T-cell response in this context is the balance between effector and regulatory T cells (Treg) [1-3]. Under resting conditions, Treg can be distinguished from effector T cells by their expression of the transcription factor FOXP3 [4], constitutive CD25 expression [5,6] and low CD127 expression in Treg [7]. However, upon activation, effector cells upregulate FOXP3 and CD25 and downregulate CD127, leading to a substantial overlap between these two cell types [8]. Single-cell RNA (scRNA) profiling has the potential to identify distinct profiles of both cell types during resting and activating conditions [9,10]. Technologies range from targeted qPCR or scRNA sequencing panels to untargeted scRNAseq, such as the 10X Genomics and SMARTseq technologies. The objective of this study was to distinguish resting and Ag-responding T effector (Tconv) and Treg using single-cell technologies.

MATERIALS AND METHODS

Subjects and PBMC isolation

Human samples from healthy adult blood donors were obtained as buffy coats (BC) from the Deutsches Rotes Kreuz for the isolation of peripheral blood mononuclear cells (PBMC) by density centrifugation. The use of the samples was approved by ethics committee with informed consent of the donors (EK240062016).

Isolation and stimulation of conventional and regulatory T cells

MACS was used to isolate CD4[+] T cells with the CD4[+] T cell Isolation Kit and CD4[−] T cells with CD4 Microbeads (Miltenyi Biotec). CD4[+] cells were stained (CD4-FITC (RPA-T4, BD); CD25-PE (M-A251, BD); CD127-eFluor 450 (eBioRDR5, eBioscience); 7AAD (BD)), washed and the CD4[+]CD25[dim/−] CD127[+] and CD4[+]CD25[+]CD127[lo] cells sorted and isolated by FACS (ARIAII, BD). CD4[+]CD25[dim/−]CD127[+] cells (Tconv) were stained with proliferation dye eFluor 450 (5 μM) and CD4[+]CD25[+]CD127[lo] cells (Treg) with eFluor 670 (10 μM) for 10 min at 37°C. After washing with cold RPMI + 10%

HS, the two populations were pooled in proportions initially measured in PBMCs and added to the non-CD4[+] cell population (Table S1A). Tetanus toxoid (Sanofi Pasteur) or influenza (Begripal 2014/2015; Novartis Vaccines and Diagnostics GmbH) Ags were added at 1 μl/ml and cells incubated for 5 days in a 96-well U-bottom plate. Cells were stained with CD8-APC-H7 (SK1, BD), CD4-FITC (RPA-T4, BD), CD25-PE (M-A251, BD), CD45RO-PE-Cy7 (UCHL1, BD) and 7AAD (BD) for the analysis of activation and proliferation of the T-cell subsets and sorting of proliferated responding (Dye[dim]CD25[hi]) and non-responding Tconv (eFluor 450) or Treg (eFluor 670) cells with very stringent gating, either as single cells (for analysis by SMARTseq or Biomark) or as bulk (for analysis by Rhapsody).

For validation experiments, the CD4[+] Tconv and Treg cells were sorted and isolated by FACS as described above and labelled with eFluor 670 (5 μM). After washing, 50 000 Tconv or Treg were cultured separately with 100 000 unlabelled CD4[+] T-cell-depleted PBMC from the same donors and 10 ng/ml Staphylococcal Enterotoxin B (SEB; Sigma-Aldrich) for 5 days in a 96-well U-bottom plate with 0·05 IU of IL-2 in the Treg culture for survival. For analysis by Multiplex qPCR by Biomark, cells were stained CD3-BUV395 (SK7, BD); CD4-BV786 (OKT4, Biolegend); CD25-BV650 (M-A251, BD); CD127-BUV737 (HIL-7R-M21, BD) and Tconv or Treg responding (Dye[dim]CD25[+]) and Tconv non-responding (Dye[bright]CD25[−]) and Treg non-responding (Dye[bright]CD25[+]) cells were single cell sorted. For FACS, cells were stained for additional surface markers with CD7-PE-Cy7 (CD7-6B7, Biolegend), CD49d-APC-Cy7 (9F10, Biolegend) and IL1R2-FITC (34141, Thermofisher Scientific) together with Fixable Viability Dye (eFluor 506, eBioscience) and for intracellular markers with FOXP3-BV421 (206D, Biolegend), IKZF2-PE/Dazzle (22F6, Biolegend) and TRIM-PE (TRIM-4, Biolegend) after fixation and permeabilization using the FoxP3/Transcription Factor Staining Buffer set (eBioscience).

Gene expression analysis

Rhapsody

The four-cell populations from one donor were bulk sorted into separate tubes pre-coated with PBS containing 4% BSA and a Sample Tag per population. Cells were stained with the Sample Tags as described by the manufacturer (BD™ Single-Cell Multiplex Kit, BD-Biosciences), reassembled into one tube and processed using the Human Immune Response Panel and the BD

Rhapsody Single-Cell Analysis System (BD-Biosciences) according to the manufacturer's instruction. The libraries were sequenced in 75-bp paired end mode on the Illumina NextSeq500 platform and the Illumina HiSeq 2500 platform to obtain approximately 40mio fragments for the mRNA libraries and approximately 0.5mio fragments
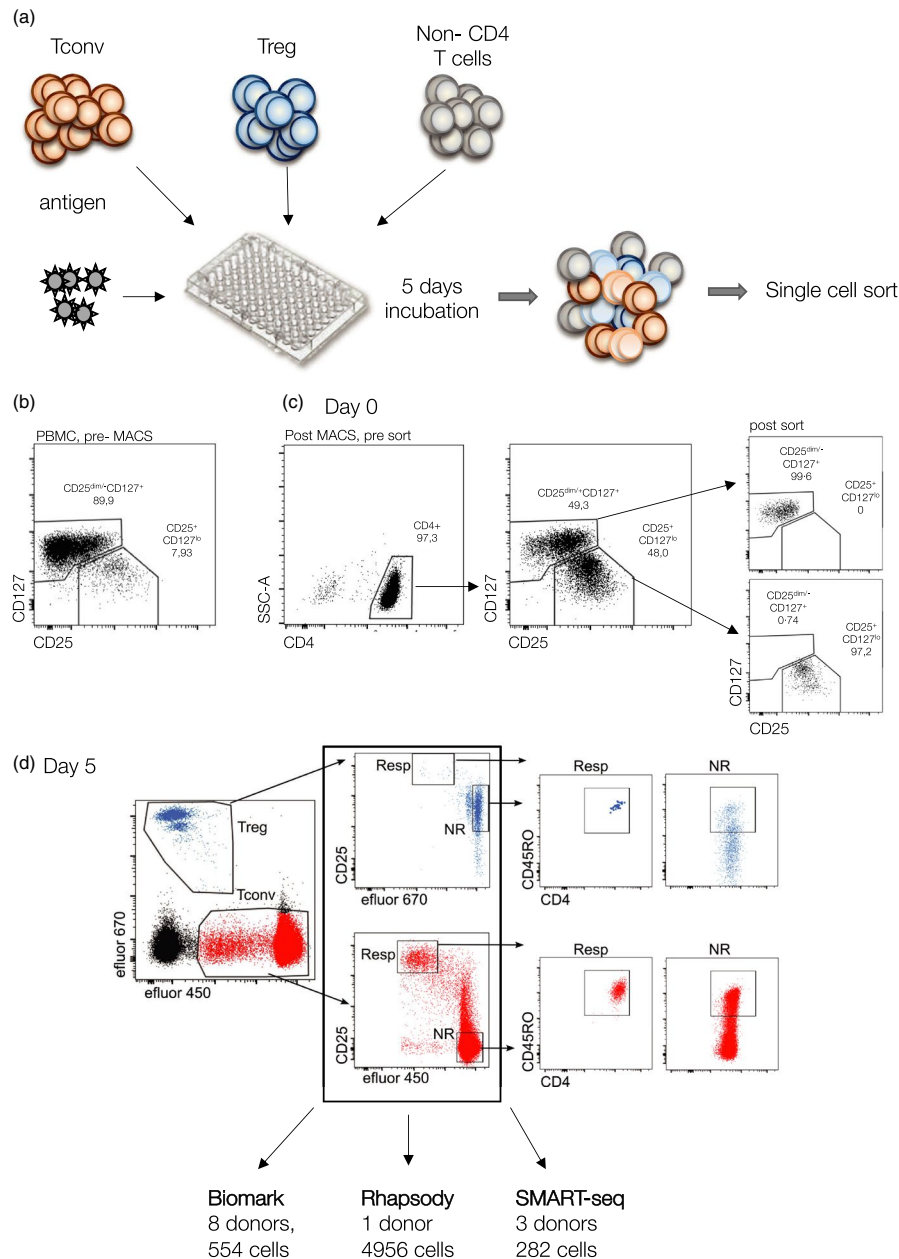

FIGURE 1 Mixed PBMC culture assay. A schematic representation of TT-stimulation assay composition: FACS isolated CD4$^+$CD25$^{dim/-}$ (Tconv, red cells) and CD4$^+$CD25$^+$CD127^{lo} (Treg, blue cells) were stained with different proliferation dyes and mixed back together with non-CD4 cells (grey cells) in a typical PBMC composition. TT-Ag (grey star shaped) was added to the culture. After incubation for 5 days, proliferating and non-proliferating Tconv and Treg (proliferation dye dim and high cells, respectively) were single cell sorted by FACS. B Gating of CD4$^+$CD25$^{dim/-}$CD127$^+$ and CD4$^+$CD25$^+$CD127^{lo} from an exemplary staining of PBMCs; shown is the CD25 and CD127 gate after selecting for CD4$^+$ cells. C FACS gating strategy for Tconv and Treg. Cell sorting on day 0 was performed using FACS to isolate CD4$^+$CD25$^{dim/-}$ CD127$^+$ Tconv and CD4$^+$CD25$^+$CD127^{lo} Treg (middle plot) from a MACS enriched population of CD4+ cells (left plot). Post-sort purity is shown in the right plots. D FACS gating strategy for Tconv and Treg on day 5 after stimulation. Tconv and Treg were identified by their different proliferation dyes, efluor 670 (Treg, blue) and efluor 450 (Tconv, red). Cells were divided into respondings (Resp) and non-respondings (NR) by their CD25 expression and dye intensity. Resp and NR were further characterized by the composition of memory cells (CD45RO$^+$, right plots). The number of donors and the number of cells included in the data analysis for each technology is shown at the bottom. Gating statistics are shown within the plots for B and C, and in Table 1 for D

for the sample tag libraries. The BD Rhapsody targeted analysis pipeline (v1.3; https://bitbucket.org/CRSwDev/cwl/src) was used to process the raw sequencing data. Cell labels and unique molecular indices were identified from the R1 reads. The respective R2 reads were mapped against the human BD Rhapsody Immune Response Panel sequences and against the human BD Rhapsody Sample Tag sequences #5-8. Data pre-processing, dimension reduction, clustering (default) and differential genes expression analysis were done using the Seurat package (Seurat 2.3.).The 21 genes in the panel were used for analysis. After normalization and scaling, principal component analysis (PCA) was used to perform dimensionality reduction of all the data and then projected with Uniform Manifold Approximation and Projection for Dimension reduction (UMAP). Differentially expressed genes were found using the FindMarker function with a logfc threshold of 0·3 and requiring the expression of the gene in ≥30% of the cells.

SMARTseq

Single cells from the four-cell populations from three donors were sorted into 96-well plates containing 2 μL of nuclease-free water with 0·2% Triton X-100 and 4 U murine RNase inhibitor (NEB), centrifuged and frozen at −80°C. The workflow was based on the previously described SMARTseq2 protocol [11] with some modifications. After thawing, 2 μl of the primer mix (5 mM dNTP (Invitrogen), 0·5 μM oligo-dT primer, 4 U murine RNase inhibitor) were added to each well. The reverse transcription reaction was performed using RNase inhibitor (9 U) and Superscript II (90 U) at 42°C for 90 min, followed by an inactivation step at 70°C for 15 min. The number of pre-amplification PCR cycles was increased to 22 and the amplified cDNA was purified using Sera-Mag SpeedBeads (GE Healthcare) and DNA eluted. 0·7 ng of pre-amplified cDNA was used for library preparation (Nextera XT library preparation kit, Illumina) in a 5 μL volume. Illumina indices were added

TABLE 1 Features of targeted, semi-targeted and whole transcriptome single-cell gene expression methods used

	Biomark	Rhapsody	SMARTseq
Company	Fluidigm	BD, Illumina	Takara, Illumina
Approach	targeted	semi-targeted	whole transcriptome
Methodological steps	cDNA synthesis (poly(A)), multiplex PCR, qPCR	cDNA synthesis (poly(A)), multiplex PCR, sequencing	SMARTer first strand cDNA synthesis, cDNA amplification, sequencing
Number of analysed cells	96	1000–20 000[a]	96/384
Number of input cells required	1	Minimum of 1000	1
Processing	Can be postponed, cells can be frozen after sort	Directly after FACS sort	Can be postponed, cells can be frozen after sort
Number of genes	Up to 96	259 or 399[b]	>25 000 (3000–6000)
Reads/cell	NA	2000–20 000	0·5 million
Developmental requirements	customized panel	Ready to use	Ready to use
Processing time	1 day	3–8 weeks	3–8 weeks
Costs/cell (€) *	16	1·59/0·11[c]	9·76/4·55[d]

Note: Costs per cell without personnel and overheads.

The characteristics of the three methods used are shown. Company: commercial companies providing reagents and/or technology; Approach: distinction between a targeted (only a restricted number of selected genes (here 48)), a semi-targeted (a commercial panel, including 399 Immune-related or 259 T cell- related genes) and a whole transcriptome (single-cell mRNA seq) approach; Number of cells analysed: cells analysed per experimental unit (Biomark: PCR plate, Rhapsody: cartridge, SMARTseq: PCR plate); Number of input cells required: the minimal number of cells that can be run per experimental unit; Processing: immediate processing required or freezing of cells possible; Number of genes: genes that can be detected with the method (Biomark: one chip allows the analysis of 96 samples and 96 genes; Rhapsody: the number of genes depends on the panel used); Reads per cell: sequencing depth usually applied; Developments required: customized panel needs to be developed for Biomark including primer design, primer efficiency testing and mutual primer inhibition testing; Processing time: time required from the sample to the raw data obtention; Costs/cell: processing costs per cell, not including personnel and overhead costs.

[a]Up to 12 samples with Barcodes can be pooled and analysed.

[b]259 genes in the T-Cell-Expression- and 399 genes in Immune response panel.

[c]Costs when 1000/20 000 cells are analysed.

[d]Costs when a 96-/384-well plate is used.

during the PCR (72°C for 3 min, 98°C for 30 s, 12 cycles of [98°C for 10 s, 63°C for 20 s and 72°C for 1 min], and 72°C 5 min) with 1× KAPA Hifi HotStart Ready Mix and 0·7 µM of dual indexing primers. After PCR, the libraries were quantified, pooled in equimolar amounts and purified twice with Sera-Mag SpeedBeads. The libraries were
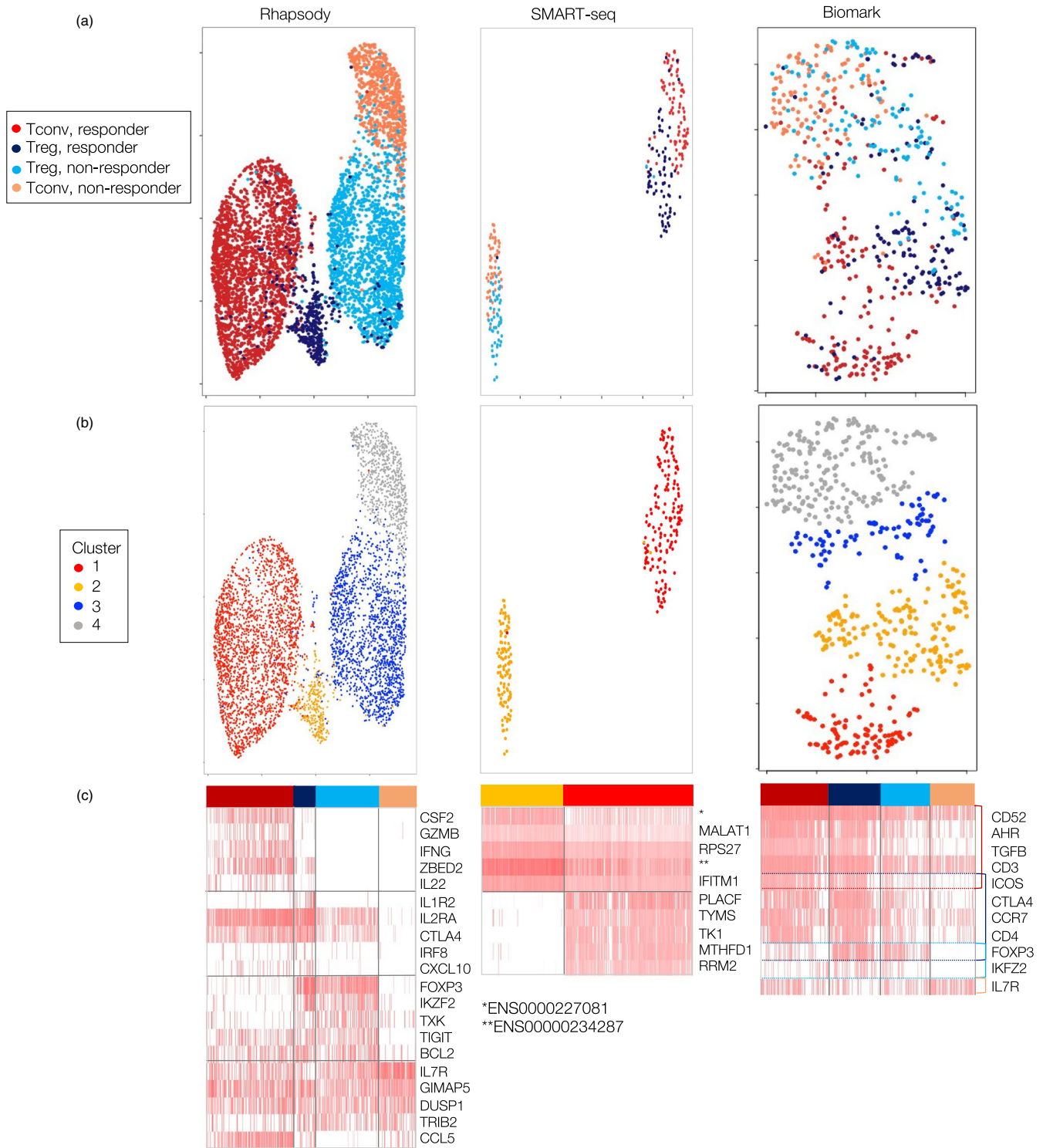

FIGURE 2 Genes differentially expressed between TT responding, non-responding, Tconv and Treg cells. (a) UMAP visualization of cells after integrating the three technologies, Rhapsody, SMARTseq and multiplex qPCR by Biomark. Cells are coloured according to their cell type (responding Tconv: dark red, non-responding Tconv: coral, responding Treg: dark blue and non-responding Treg: royal blue). (b) UMAP visualization as in (a). Cell coloured according to clusters found. (c) Differential expression of the top five most variable genes between the four cell types (Rhapsody and Biomark) and responding and non-responding cells (SMARTseq) are shown in one heatmap for the three technologies. Cell Types are colour-coded as in A for Rhaspody and Biomark and according to clusters for SMARTseq

sequenced on the NextSeq 500 Illumina platform to obtain 75-bp single-end reads aiming at an average sequencing depth of 0·5 million reads per cell. Alignment of the reads to the reference human genome (hg38) was done with GSNAP (v2018-05-30) and Ensembl gene annotation (version 87) was used to detect splice sites. The aligned reads were quantified with featureCounts (v1·6.2) from the Subread package and the same Ensembl annotation was used to generate a counts-matrix. Cells expressing only a few genes were filtered out from the counts-matrix using the clean.counts function in SCDE [12] (min.lib. size = 1000, min.reads = 1, min.detected = 1). Further processing of the counts-matrix was performed using the following R packages: SingleCellExperiment [13,14] and scater [15]. Briefly, the counts-matrix was loaded and a single-cell experiment object was constructed. The ERCC spike-in counts were added as an alternative experiment attributed to the object. Dimensionality reduction was performed and the data were visualized using UMAP [16] implemented in the R package umap, version 0.2.3.1. Differentially expressed genes were identified using the findMarkers function from the scran package with the default settings while blocking for plate technical confounder.

Multiplex qPCR by Biomark

Single cells were sorted into 96-well PCR plates containing 5 µl EB buffer (Qiagen), immediately snap-frozen on dry ice and stored at −80°C. Multiplex qPCR was performed as described [17] with some modifications. cDNA was synthesized using Quanta qScript™ cDNA Supermix

directly on cells. Total cDNA was pre-amplified for 20 cycles (1 × 95°C for 8 min, 95°C for 45 s, 49°C with 0·3°C increment/cycle for 1 min, and 72°C for 1·5 min) and 1 × 72°C for 7 min with TATAA GrandMaster Mix (TATAA Biocenter) in a volume of 35 µL in the presence of the primer pairs for 40 genes (25 nM final concentration for each primer as described [17] but without *CCR10*, *CCR3*, *GATA3*, *IL17F*, *EOMES* and *NFTAC* and for validation using the primers listed in Table S2). Pre-amplified cDNA (10 µl) was treated with 1·2 U of exonuclease I and expression quantified by RT-PCR on a Biomark™ HD System (Fluidigm Corporation) using the 96·96 Dynamic Array IFC and the GE 96x96 Fast PCR + Melt protocol with SsoFast EvaGreen Supermix and Low ROX (BIO RAD) and 5 µM of primers for each assay. Raw data were analysed using the Fluidigm Real-Time PCR analysis software. Pre-processing and data analysis were conducted using KNIME 3.7.0, R version 3.5.1 and RStudio version 1.2.1335 (RStudio). For pre-processing, a linear model was used to correct for confounding effects potentially introduced through processing batches. In brief, batch effects (dummy coding for each plate/batch) were modelled jointly with dose effects by regressing out the effect of plates on each individual gene while controlling for dose to obtain a corrected gene expression dataset [18]. The data from eight subjects were pooled. Dimension reduction was performed using UMAP as described above. Clustering was performed with hclust and ward.D2. To find genes significantly differing between two conditions in qPCR data, the Hurdle model was applied for regression taking count data with over dispersion into account [19]. To find cluster marker genes, the FindMarker function was used, implemented in scanpy.

Clusters	Responder T conv (%)	Responder T reg (%)	Non-responder Treg (%)	Non-responder T conv (%)
Rhapsody				
1	98·7	12·9	0·2	0·0
2	0·4	72·8	0·3	0·0
3	0·6	14·0	96·1	2·9
4	0·2	0·3	3·4	97·1
SMARTseq				
1	100	98·8	0·0	3·7
2	0·0	1·2	100	96·3
Biomark				
1	48·3	10·1	0·0	0·0
2	29·0	63·0	19·2	3·3
3	11·4	12·3	30·0	7·5
4	11·4	14·5	50·8	89·2

TABLE 2 Frequencies from each cell type found per UMAP cluster after Rhapsody, SMARTseq and Biomark analysis

Machine learning

Support Vector Machine (SVM) and Logistic Regression (LR) were used with their implementation in Python's scikit library. For feature selection, we used the Recursive Feature Elimination with Cross Validation (RFECV) algorithm, also implemented in Python's scikit library. Since a larger number of data points was obtained from the Rhapsody analyses, the Rhapsody data were used for feature selection and building models. The following parameters were obtained after modelling: (i) Accuracy—how many cells were correctly predicted; (ii) Precision, defined as TruePositive/(TruePositive + FalsePositive); and (iii) Recall, defined as TruePositive/(TruePositive + FalseNegative).
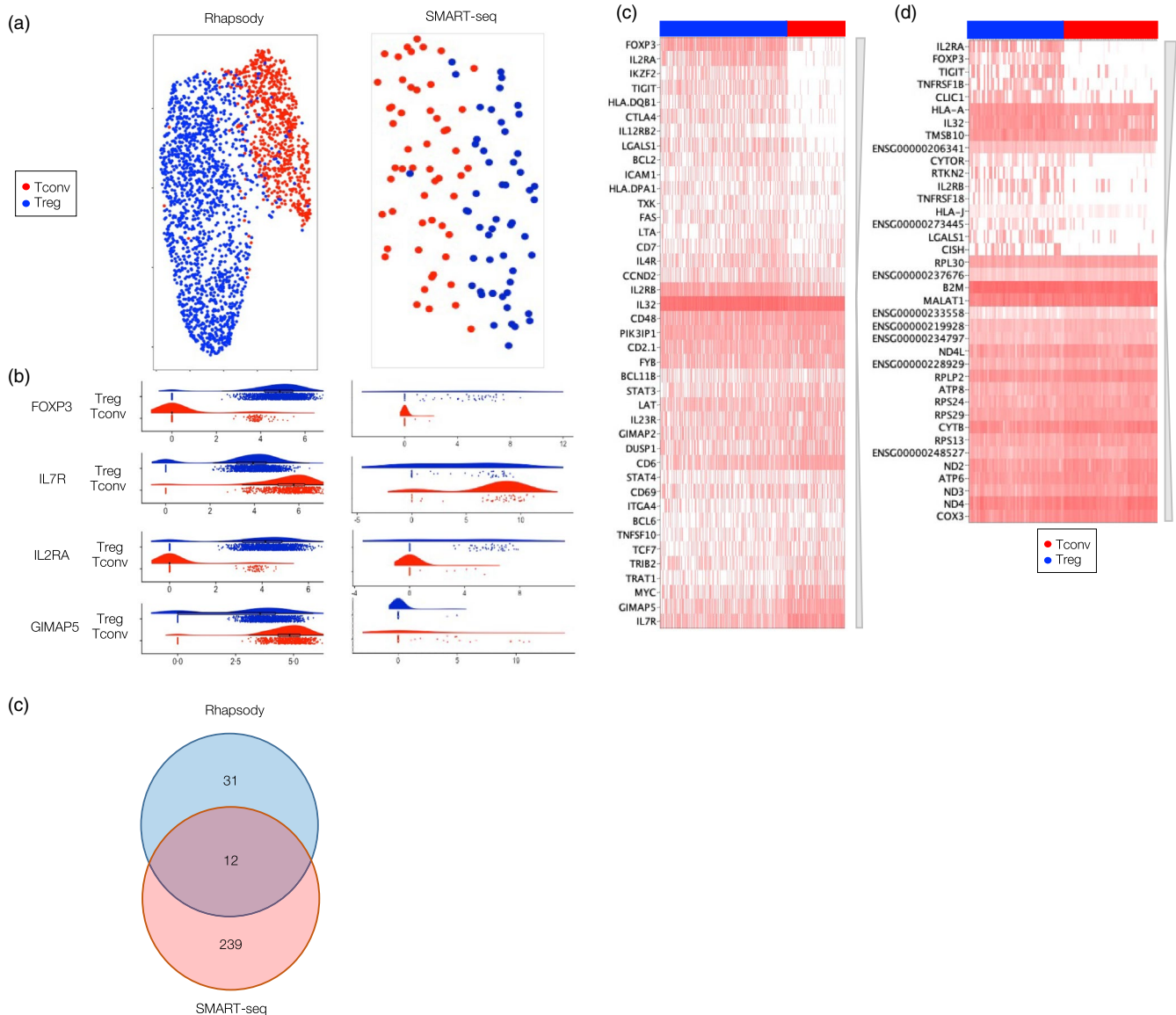

FIGURE 3 Gene expression differences between Tconv and Treg non-responding cells. (a) UMAP visualization of both cell types. Tconv are coloured in red, Treg in blue. (b) Raindot-plots showing the expression of exemplary genes significantly differing between the two cell types and shared between both (*FOXP3*, *IL7R*, *IL2RA* and *GIMAP5*), SMARTseq and Rhapsody. Genes significantly increased in Tregs are shown first. Cell Types are colour-coded as in (a). *y*-axis shows values after processing raw data, differing for each technology. (c) The Venn diagram represents the number of DE genes found by the methods and the number of genes shared between them. (d, e) Shown are heatmaps with differentially expressed genes. All genes significantly differing between Treg and Tconv in non-responding TT-stimulated cells using the Rhapsody (d), or the 20 top upregulated and downregulated genes found using SMARTseq (e). From top to bottom are genes with the highest fold change (FC) in Tregs and from bottom to top in Tconv. Tconv are shown in red, Treg in blue

RESULTS

In vitro isolation of non-responding and responding conventional and regulatory T cells

An in vitro model was established to distinguish and isolate Tconv and Treg human CD4$^+$ T cells after activation (Figure 1). After isolation and separate labelling of Treg and Tconv cells with eFluor670 or eFluor450, cells were reunited along with the non-CD4$^+$ T cells so as to mimic a classical PBMC stimulation assay, and the cell mixture was stimulated with Ag for 5 days (Figure 1b and Table S1A). The dye^{dim}CD25^{high} for responding Treg (eFluor670) and Tconv cells (eFlour450), the dye^{bright}CD25^{high} as non-responding Treg cells, and the dye^{bright}CD25^{low} as non-responding Tconv cells were sorted and processed (Figure 1d and Table S1B,C). Responding cells against the model Ag tetanus toxoid CD45RO$^+$ and were observed for both Treg and Tconv cells (Figure 1d).

mRNA profiles distinguish non-responding and responding conventional and regulatory T cells

Three methods were used to profile mRNA (Table 1). RNAseq was performed using the Rhapsody technology (Rhapsody) and next-generation sequencing using SMARTseq2 (SMARTseq). Multiplex RT-PCR using the Biomark Fluidics System (Biomark) with a previously established panel was also examined. Each method was able to discriminate cell types to a certain degree (Figure 2a).

The greatest discrimination between the four cell types was achieved by Rhapsody, with clear separation of the Treg and Tconv cells confirming the purity of the populations (Figure 2a) and four clusters, each containing the majority of a distinct cell type (Figure 2b and Table 2).

The genes that best discriminated Treg from Tconv were *FOXP3*, *IKFZ2* and *TXK*, whereas the top genes discriminating responding and non-responding cells were *IL2RA* and *CTLA4* (Table S3). Responding Tconv were the only cells that could be separated from the others by a set of almost uniquely expressed genes including *CSF2*, *GZMB*, *IFNG*, *ZBED2* and *IL22* (Figure 2c and Table S3).

A clear distinction between responding and non-responding cells was obtained with SMARTseq, resulting in two characteristic clusters (Figure 2). Several genes were almost exclusively expressed in the responding cells (Figure 2c). These two characteristic clusters were maintained when analysing the three donors separately, and the majority of the cell types were assignable to the corresponding cluster. The largest fraction of incorrectly assigned cells was the non-responding Treg, with 27·8%, 11·1% and 5·6% of the cells from donors 1, 2 and 3, respectively, found in the cluster containing responding cells (Table S4).

The Biomark panel genes were less discriminatory with overlap between the four cell types observed in the four UMAP clusters (Figure 2c and Table S3). Therefore, the Rhapsody and SMARTseq data were used for subsequent comparisons.

Genes distinguishing non-responding regulatory and conventional T cells

The Rhapsody and SMARTseq profiles were examined to find differentially expressed genes between non-responding Treg and Tconv. Although *FOXP3* alone provided reasonable discrimination with the Rhapsody technology, genes known to differ between the two cell types at the protein level (e.g. *FOXP3*, *IL7R*, *IL2RA*) showed some degree of overlap in their gene expression level between the non-responding Treg and Tconv as measured by the Rhapsody and the SMARTseq technologies (Figure 3 and Table S5). Therefore, we explored

FIGURE 4 Machine learning for discriminatory genes between non-responding Treg and Tconv (a) Plot showing the performance of the Recursive Feature Elimination with Cross-Validation (RFECV) approach using the Random Forest Algorithm to identify the most discriminatory genes to distinguish non-responding Treg from Tconv cells. The *y*-axis shows the accuracy (determined by identifying the cells that are correctly classified) and the *x*-axis shows the number of input features (genes). The light blue shaded area represents the variability of cross-validation, one standard deviation above and below the mean accuracy score shown in the curve. (b) A bar plot showing the relative contribution of the 13 input features that were selected and used for training classifiers using both the algorithms—Support Vector Machine (SVM) and Logistic Regression (LR). (c) Receiver Operating Curve (ROC) for the classification of non-responding Treg from Tconv using Support Vector Machine (SVM) from the features determined using the RFE-RF algorithm. (d) ROC for the classification of non-responding Treg from Tconv using Logistic Regression (LR) from the features determined using the RFE-RF algorithm. Red curves correspond to performance of the model that was built using the Tetanus-Ag-stimulated cells (trained with 80% of the input data, tested with the remaining 20%) and blue dashed curves correspond cells that responded to stimulation with the influenza Ag. (e and f) Shown are heatmaps with the 13 selected signature genes and the gene expression data obtained by SMARTseq after stimulation with the tetanus (f) or the influenza Ag (f). Tconv are shown in red, Treg in blue

(a)

(b)

(c)

(d)

(e)

(f)
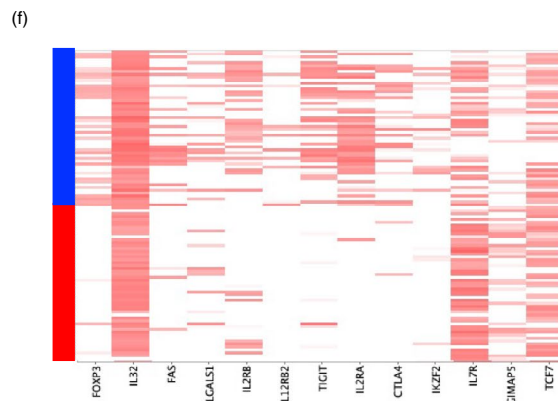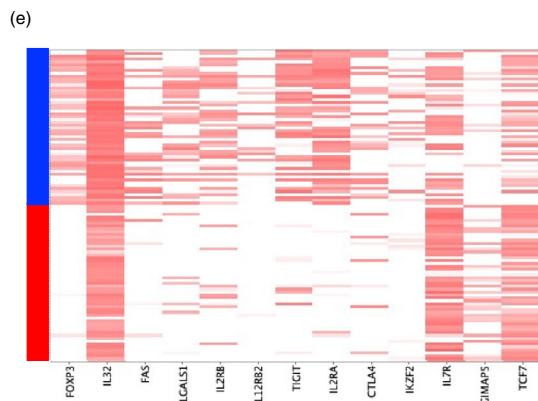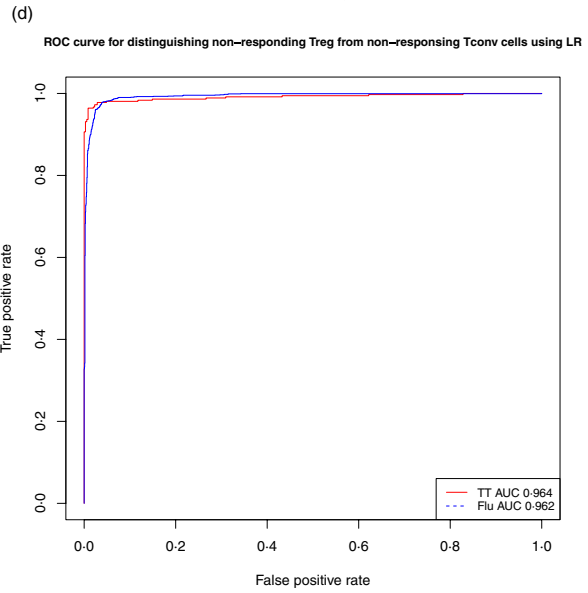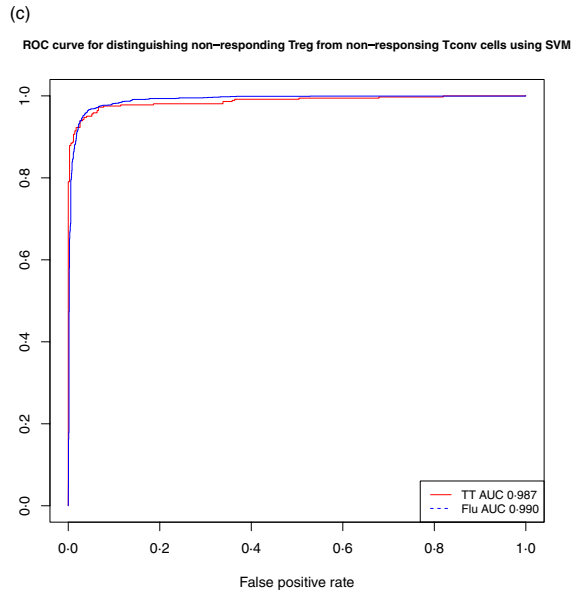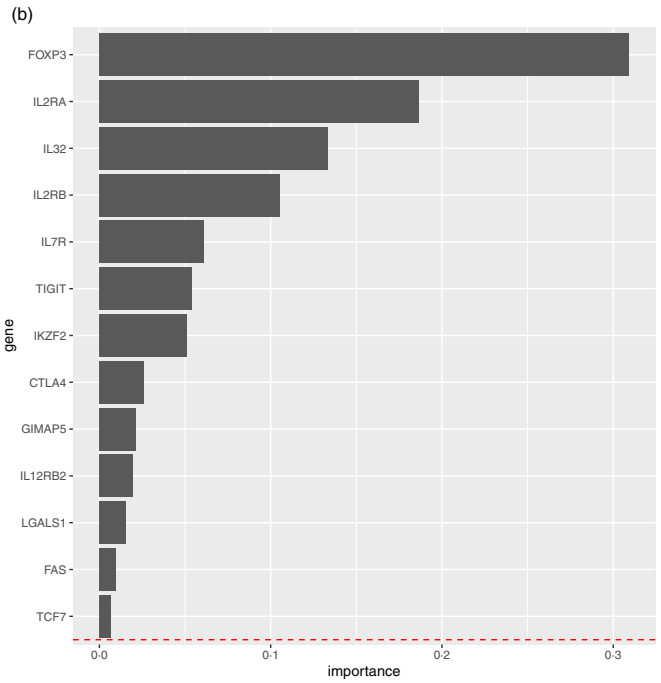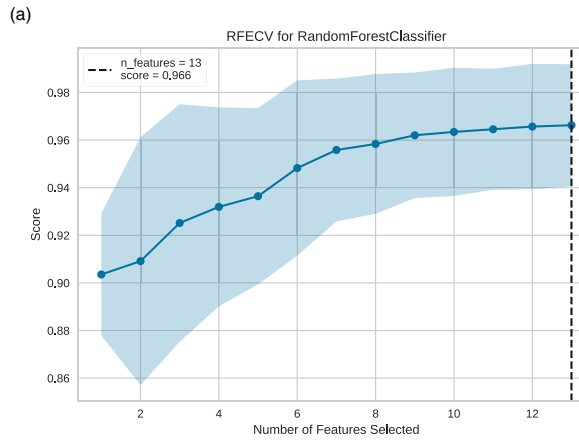
two machine learning algorithms. We used the Rhapsody data from non-responding cells stimulated with TT and identified 13 genes with significant differential expression between the cell types found by Rhapsody and either additionally by SMARTseq or Biomark (Table S5 and Figure S1) to develop a Recursive Feature Elimination with

TABLE 3 Performance of Machine learning algorithms to distinguish cell types after analysis with Rhapsdoy and SMARTseq

Comparison	Ag	Algorithm	Accuracy	Precision	Recall	AUC	Cells training set	Cells test set	Input Genes	Signature Genes	Method
Non-responder T conv vs T reg	TT	LR	0·972	0·980	0·964	0·992	2845	712	13	FOXP3, IL2RA, IL32, IL2RB, IL7R, TIGIT, IKZF2, CTLA4, GIMAP5, IL12RB2, LGALS1, FAS, TCF7	Rhapsody
	TT	SVM	0·954	0·974	0·934	0·987	2845	712			
	Flu	LR	0·967	0·971	0·962	0·993	NA	3647			
	Flu	SVM	0·952	0·977	0·928	0·990	NA	3647			
	TT	SVM	0·818	0·714	1	0·95	86	22	NA		SMARTseq
	TT	LR	0·864	0·769	1	0·908	86	22			
	Flu	SVM	0·869	0·81	0·962	0·943	108	108			
	Flu	LR	0·879	0·823	0·962	0·947	108	108			
T reg responder vs non-responder	TT	LR	0·986	0·992	0·992	0·999	1766	422	46	GAPDH, ITGAE, ICOS, IL2RA, TYMS, LGALS1, IRF4, ANXA5, AURKB, BAX, CXCR3, LAP3	Rhapsody
	TT	SVM	0·989	0·997	0·989	0·998	1766	422			
	Flu	LR	0·972	0·995	0·974	0·995	NA	2018			
	Flu	SVM	0·987	0·997	0·989	0·995	NA	2018			
	TT	SVM	0·893	1	0·812	1	110	28	NA		SMARTseq
	TT	LR	0·929	0·938	0·938	0·979	110	28			
	Flu	SVM	0·928	0·974	0·905	0·97	138	138			
	Flu	LR	0·913	0·929	0·929	0·964	138	138			
Responder T conv vs T reg	TT	LR	0·974	0·945	0·873	0·971	2119	530	46	FOXP3, IKZF2, ITGA4, TRAT1, LGALS1, IL1R2, CD7	Rhapsody
	TT	SVM	0·966	0·969	0·797	0·990	2119	530			
	Flu	LR	0·976	0·910	0·793	0·961	NA	2121			
	Flu	SVM	0·975	0·904	0·793	0·966	NA	2121			
	TT	SVM	0·943	0·938	0·938	0·987	136	35	NA		SMARTseq
	TT	LR	0·914	0·882	0·938	0·98	136	35			
	Flu	SVM	0·924	0·902	0·954	0·966	171	171			
	Flu	LR	0·912	0·891	0·943	0·959	171	171			

Note: Ag: the antigen used to stimulate the cells (Tetanus Toxoid (TT) or Influenza (Flu)); Algorithm: the algorithm used (Support Vector Machine (SVM) and Logistic Regression (LR)); AUC: the area under the curve obtained by plotting false-positive against true-positive classifications; Cells training set: the number of cells used in the training set; Cells test set: the number of cells used in the test set; Genes: the genes used for modelling.

Cross Validation (RFECV) algorithm (Figure 4a,b). These 13 most discriminatory genes (*FOXP3*, *IL2RA*, *IL32*, *IL2RB*, *IL7R*, *TIGIT*, *IKZF2*, *CTLA4*, *GIMAP5*, *IL12RB2*, *LGALS1*, *FAS* and *TCF7*) were used in SVM and LR machine learning algorithms on the data split into a training set (2845 of the 3557 available cells—80% of the data) and a test set (712 cells—20% of the data). The SVM algorithm correctly attributed 95·4% of the 712 cells from the test set to the correct cell type (Accuracy: 0·954; Precision: 0·974; Recall: 0·934; AUC: 0·987, Figure 4c and Table 3) and the LR algorithm 97·2% of the cells (Accuracy: 0·972; Precision:

0·98; Recall: 0·964; AUC: 0·992, Figure 4d and Table 3). The same 13 genes were used on the SMARTseq data and the algorithms were able to correctly attribute a high percentage of cells to their type (SVM: Accuracy, 0·925; Precision, 0·871; Recall, 1·0; AUC, 0·995; LR: Accuracy, 0·963; Precision, 0·946; Recall, 0·981; AUC, 0·992; Table 3). These 13 genes were expressed in a characteristic and cell-type-specific manner (Figure 4e).

Application of the algorithms to 3647 cells non-responding to a stimulation with influenza Ag and analysed by Rhapsody attributed 95·2% of the cells to the
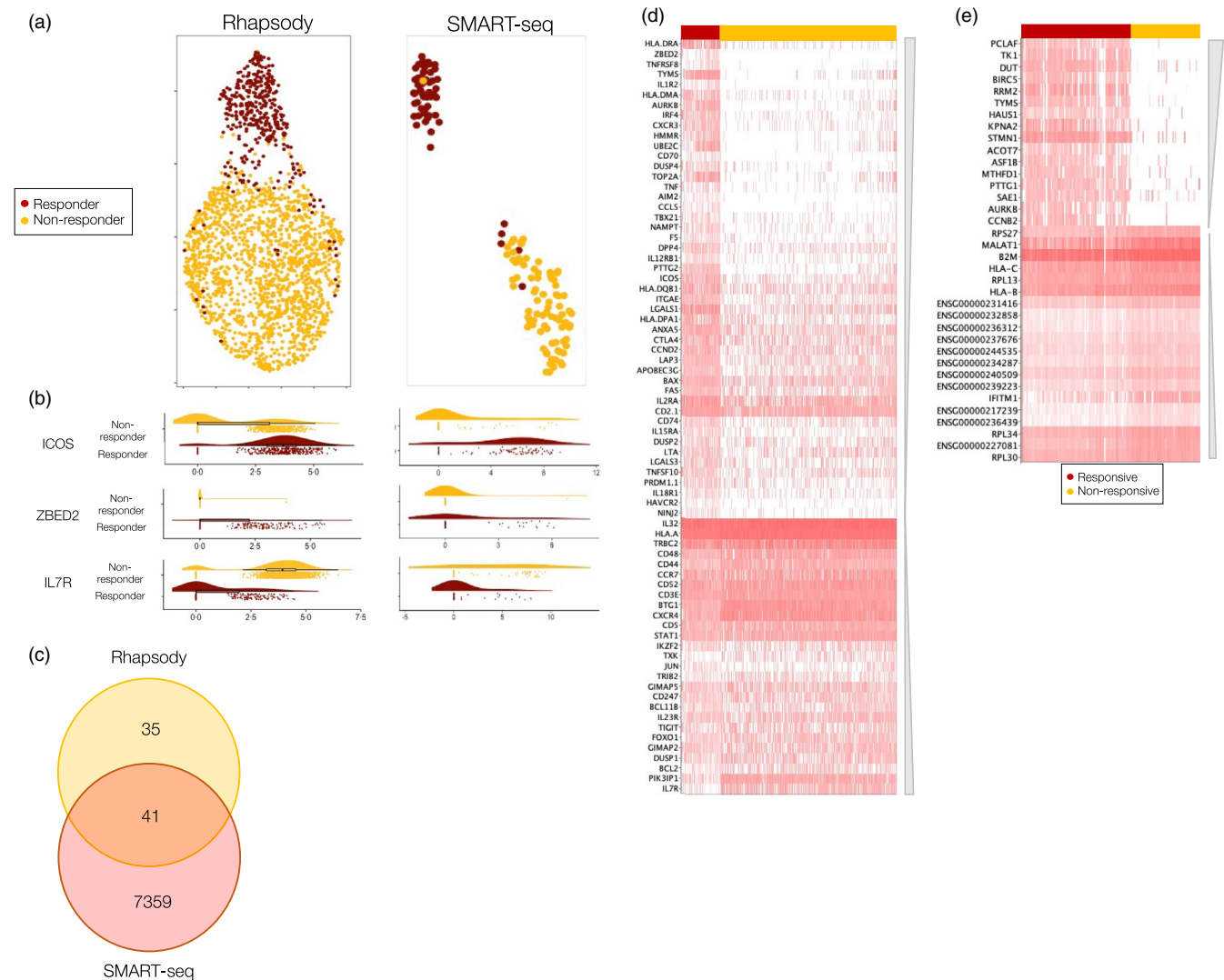

FIGURE 5 Gene expression differences between responding and non-responding Treg cells. (a) UMAP visualization of both cell types. Responding are coloured in dark red, non-responding in yellow. (b) Raindot-plots showing the expression of exemplary genes significantly differing between the two cell types and shared between SMARTseq and Rhapsody (*ICOS*, *ZBED2* and *IL7R*). Genes significantly increased in responding cells are shown first. Cell Types are colour-coded as in (a). *y*-axis shows values after processing raw data, differing for each technology. (c) The Venn diagram represents the number of DE genes found by each method and the number of genes shared between them. (d and e) Shown are heatmaps with all genes found to be significantly differing between responding and non-responding in TT stimulated Treg cells using the Rhapsody (d) or the 20 top upregulated and downregulated genes found using SMARTseq (e). From top to bottom are genes with the highest fold change (FC) in responding and from bottom to top in non-responding cells. Respondings are shown in dark red, non-respondings in yellow
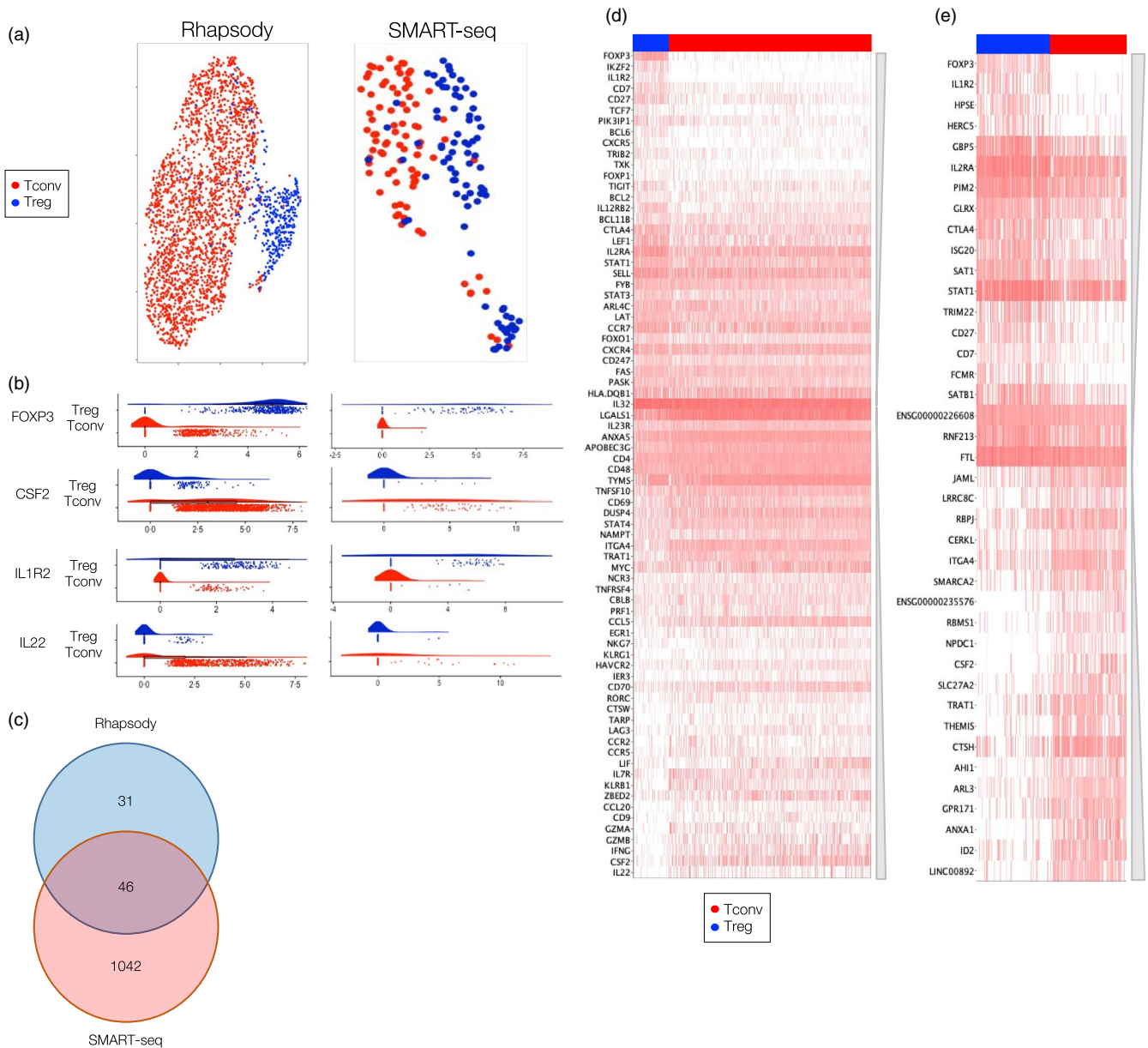
FIGURE 6 Gene expression differences between Tconv and Treg responding cells. (a) UMAP visualization of both cell types. Tconv are coloured in red, Treg in blue. (b) Raindot-plots show the expression of genes significantly differing between the two cell types and shared between both SMARTseq and Rhapsody (*FOXP3*, *CSF2*, *IL1R2* and *IL22*). Cell Types are colour-coded as in (a). *y*-axis shows values after processing raw data, differing for each technology. (c) The Venn diagram represents the number of DE genes found by each method and the number of genes shared between them. (d and e) Shown are all genes found to be significantly differing between Treg and Tconv in responding TT-stimulated cells using the Rhapsody (d) or the 20 top upregulated and downregulated genes found using SMARTseq (e). From top to bottom are genes with the highest fold change (FC) in Tregs and from bottom to top in Tconv. Tconv are shown in red, Treg in blue

correct cell type with SVM (Accuracy: 0·952; Precision: 0·977; Recall: 0·928; AUC: 0·990) and 96·7% with LR (Accuracy: 0·967; Precision: 0·971; Recall: 0·962; AUC: 0·993). The same procedure was repeated with SMARTseq data, which also demonstrated that the classifier was sensitive and specific in discriminating the two cell types, thereby confirming that both the machine learning algorithms can discriminate the two non-responding cell types exposed to different Ag (Figure 4f and Table 3).

Ag-responding and non-responding Treg cells have distinct scRNAseq profiles

The transcription profiles of responding Treg cells differed markedly from those of the non-responding cells (Figure 5 and Table S6). Upregulation of the Treg marker *ICOS* was observed using both technologies (and also using Biomark, see Figure S2). Other upregulated genes in either or both technologies included the transcription

factors *TBX21* and *RORC*, activation markers such as HLA class II genes, and genes associated with proliferation such as *HMMR*. *ZBED2* was rarely detected in non-responding Treg, but expressed in around 50% of the responding Treg. A number of genes including the Tconv marker *IL7RA* were downregulated in the responding Treg. Of note, each technology revealed a list of technology-specific DE genes (Figure 5c–e and Table S6).

Using a recursive feature elimination algorithm with 46 genes found by the Rhapsody technology as well as by either the SMARTseq or the Biomark technology, 12 marker genes (*GAPDH*, *ITGAE*, *ICOS*, *IL2RA*, *TYMS*, *LGALS1*, *IRF4*, *ANXA5*, *AURKB*, *BAX*, *CXCR3* and *LAP3*) were sufficient to classify the Treg into responding or non-responding. Using an 80–20 split of the data into training (1766 cells) and test (422 cells) sets, both SVM and LR machine learning algorithms could discriminate responding- from non-responding Tregs (SVM: Accuracy, 0·988; Precision, 0·997; Recall, 0·989; AUC, 0·998; LR: Accuracy, 0·986; Precision, 0·992; Recall, 0·992, AUC, 0·999; Table 3). Application of the algorithm to Treg cells stimulated with an influenza Ag attributed >97% of the 2018 analysed cells to their correct cell type (SVM: Accuracy, 0·987; Precision, 0·997; Recall, 0·989; AUC, 0·995; LR: Accuracy, 0·972; Precision, 0·995; Recall, 0·974; AUC, 0·995; Table 3). Application of the method to cells stimulated with either of the two Ags but sequenced with the SMARTseq technology also led to discrimination of the cell types with high sensitivity and specificity (Table 3).

Distinguishing responding regulatory and responding conventional T cells

Genes that distinguished responding Treg from responding Tconv were of particular interest. Both Rhapsody and SMARTseq could discriminate the majority of the cells from these two cell types (Figure 6a). The level of *FOXP3* expression was the most discriminatory single gene found with both technologies, with markedly higher expression in responding Tregs than responding Tconv (Figure 6b–e and Table S7). Cytokine genes were also discriminatory and present mainly in responding Tconv cells, with *IFNG*, *CSF2*, *IL22* and *IL32* differentially expressed in both technologies and *IL13* and *IL21* differentially expressed in SMARTseq. Similar genes were also identified by Biomark (Figure S3 and Table S6). A total of 46 DE genes were observed in both technologies, an additional 31 with Rhapsody only and 1042 with SMARTseq only (Figure 6c, Table S6). The recursive feature elimination algorithm using 46 genes with significant expression identified a minimum set of seven genes (*FOXP3*, *IKZF2*, *ITGA4*, *TRAT1*, *LGALS1*, *IL1R2* and *CD7*) that provided discrimination between the responding Treg and Tconv cells. Using an 80–20 split of the data into training (2119 cells) and test (530 cells) sets, both SVM and LR algorithms were able to achieve high specificity and sensitivity in identifying the two cell types (SVM: Accuracy, 0·966; Precision, 0·969; Recall, 0·797; AUC, 0·990; LR: Accuracy, 0·974; Precision, 0·945; Recall, 0·873, AUC, 0·971; Table 3). The algorithms were also able to discriminate influenza Treg and Tconv (SVM: Accuracy, 0·975; Precision, 0·904; Recall, 0·793; AUC, 0·961; LR: Accuracy, 0·976; Precision, 0·910; Recall, 0·793; AUC, 0·966; Table 3) and were also effective on SMARTseq generated data (Table 3).

From the seven genes allowing discrimination of responding Treg from Tconv, all except *IL1R2* were also found to discriminate non-responding Treg from responding Tconv by at least one method (Table S8). This comparison also revealed an unexpected increased expression of *IL7R* in non-responding Treg as compared with Tconv.

Validation of signature genes using the biomark technology and FACS

Genes discriminating responding and non-responding Treg and Tconv cells by machine learning were tested using the Biomark technology and by FACS. Treg and Tconv cells were isolated by FACS sorting of CD4$^+$CD25$^+$CD127^{low} and CD4$^+$CD25$^{dim/−}$CD127$^+$ cells, respectively, and each cell type was reunited separately with non-CD4$^+$ T cells. The cell mixture was stimulated with SEB and, after 5 days, the responding and the non-responding Treg or Tconv populations were isolated by FACS sorting single cells (Figure S4a,b). In all, 20 cells from each population were processed for gene expression profiling by Biomark qPCR (Figure 7). Of the 27 signature genes, 24 were successfully transferred to qPCR on Biomark. Six (*FOXP3*, *IL2RA*, *IL7R*, *IKZF2*, *GIMAP5* and *TCF7*) of 13 signature genes tested to distinguish non-responding Treg from non-responding Tconv differed by Biomark qPCR; 4 (*FOXP3*, *IKZF2*, *TRAT1* and *IL1R2*) of the 6 signature genes tested to distinguish responding Treg and responding Tconv differed by Biomark qPCR; and 8 (*GAPDH*, *ICOS*, *IL2RA*, *LGALS1*, *IRF4*, *AURKB*, *BAX* and *CXCR3*) of the 12 genes signature genes tested to distinguish responding from non-responding Treg differed by Biomark qPCR (Figure 7a). Projection with UMAP of this minimal set of significant genes allowed separation of the four cell types (Figure 7b).

We used markers selected by machine learning that were available for flow cytometry of responding Treg or Tconv after the same SEB-stimulation strategy as described for the Biomark analysis above. The staining
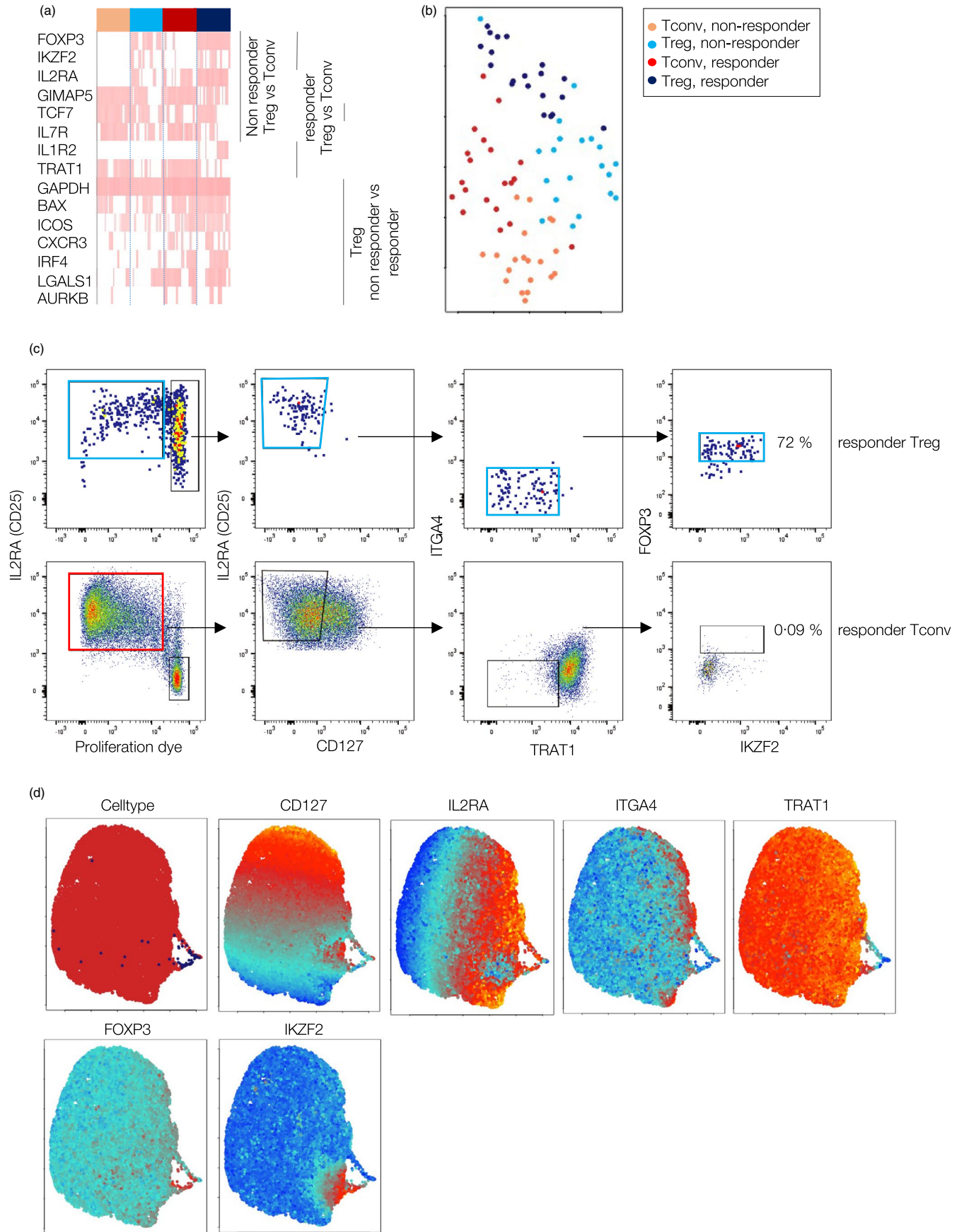
FIGURE 7 Verification of signature genes to distinguish cell types by Multiplex qPCR (Biomark) and FACS. (a) Heatmap showing the expression of signature genes in the four cell types (Tconv, non-responding (coral), Treg, non-responding (royal blue), Tconv, responding (dark red) and Treg, responding (dark blue)) measured by Biomark. The signature genes are ordered from top to bottom according to the cell type they denote: non-responding Treg versus Tconv, responding Treg versus Tconv and Treg, non-responding versus responding (as shown at the right of the heatmap). (b) UMAP visualization of the cell types after Biomark analysis using genes that were significantly different between cell types in the Biomark analysis (*FOXP3, IL2RA, IL7R, IKZF2, GIMAP5, TCF7, TRAT1, IL1R2, GAPDH, ICOS, LGALS1, IRF4, AURKB, BAX* and *CXCR3*); cells are coloured as in (a). (c) Exemplary FACS gating strategy allowing the distinction of responding Treg from responding Tconv after a 5-day stimulation with SEB using the markers IL2RA, CD127 (IL7R), TRAT1, ITGA4, FOXP3 and IKZF2. The top panels show the gates set to select Treg (blue gates) and the bottom panels shows the same gates applied to Tconv. The most left top and bottom panels show gating of responding cells (blue on top for Tregs and red on bottom for Tconv). Frequencies in the right panels refer to the frequency of $CD4^+CD25^{++}CD127^{low}TRIM^{low}FOXP3^{+++}IKZF2^{+++}$ cells out of all $CD4^+CD25^{++}CD127l^{ow}$ cells for the responding Treg (72%) and responding Tconv (0·09%). (d) UMAP visualization of responding Treg and Tconv cells from 1 donor analysed by FACS using compensated fluorescent intensities. In the top left panel, the cell types are shown (Tconv in red, Treg in blue). All the other panels show the expression of the markers used for the analysis. The colours represent a relative scale of fluorescent intensities, from dark blue (low) over light blue and red to yellow (high)

of the Treg markers FOXP3 ($p < 0·0001$) and IKZF2 ($p < 0·0001$) and the Tconv markers ITGA4 ($p < 0·0001$) and TRIM (TRAT1) ($p < 0·0001$) differed between the responding Treg and responding Tconv cells, although there was overlap between cell populations for each marker (Figure S4c). CD7 and IL1R2 were not consistently different in the two cell populations. A gating strategy with CD127 (IL7RA) and CD25 (IL2RA) and the four significant markers FOXP3, IKZF2, ITGA4 and TRIM (TRAT1) was used to identify the majority of the responding Treg and then applied to responding Tconv population (Figure 7c). The Treg-defined gate was able to identify 72% of the responding Treg as compared to 0·09% responding Tconv. These markers were also discriminatory when used in a UMAP projection (Figure 7d).

DISCUSSION

Ag-responding and non-responding Treg and Tconv cells were distinguished by the single-cell gene expression techniques. Algorithms based on the expression of less than 20 genes could accurately identify responding Tregs from non-responding Treg and responding Tconv cells.

We developed a mixed PBMC Ag-stimulation culture assay that allowed us to track differentially dye-labelled Treg and Tconv. Responding and non-responding Treg and Tconv could be identified and sorted via their label and dye dilution allowing single-cell transcriptomic analysis of the four populations. We used this approach to mimic *in vitro* antigen stimulation assays that use proliferation as their readout. We chose methods for semi-targeted (RNAseq on a relatively large panel of genes) and a non-targeted RNAseq approach for the gene expression analyses so that we could assess the merits of these methods in distinguishing a limited number of related cell types and provide validation of findings in multiple methods.

The semi-targeted Rhapsody method yielded four clusters, each highly enriched for one of the cell types. The non-targeted SMARTseq method yielded two very distinct clusters separating responding and non-responding cells. Examining the genes that distinguished the responding and non-responding clusters in the SMARTseq data suggests that the responding and non-responding discrimination was determined by a number of genes that were not present in the targeted methods. However, the targeted Rhapsody single-cell method appeared to have advantages in distinguishing the four related cell types. The advantages and disadvantages of the targeted Rhapsody and non-targeted 10x Genomics methods for high-resolution analysis of primary $CD4^+$ T cells have been discussed [20].

An important feature of this study was the ability to distinguish the responding and non-responding Treg and Tconv with algorithms that used data from a small set of genes. We focussed on the ability to discriminate the responding Tregs since these are relevant to tolerance inducing therapies. In total, 27 genes could distinguish the four $CD4^+$ T-cell populations. Responding Tregs were distinguished from non-responding Tregs using an algorithm based on 12 genes (*GAPDH, ITGAE, ICOS, IL2RA, TYMS, LGALS1, IRF4, ANXA5, AURKB, BAX, CXCR3* and *LAP3*) in both the Rhapsody and the SMARTseq methods. Responding Tregs were distinguished from responding Tconv using an algorithm based on seven genes (*FOXP3, IKZF2, ITGA4, TRAT1, LGALS1, IL1R2* and *CD7*). Therefore, by measuring expression of 18 genes, it was possible to provide an estimate of the frequency of Ag-responding Tregs within a mixed $CD4^+$ T-cell culture. In all, 24 of the genes were tested in the targeted Biomark qPCR and 15 were confirmed to distinguish the populations. Moreover, several of the genes also differed at the protein level as demonstrated by flow cytometry. It should, therefore, be possible to design efficient and cost-effective methods with

minimal manipulation and without the requirement for FACS isolation of cells from whole blood to obtain a measure of Ag-responding Tregs. A limitation of the study is that we only used proliferation and 5-day culture assays as our measure of response and it is unclear whether discriminatory algorithms could be developed for shorter stimulation assays, other assay types or *in vivo* responding cells.

The algorithms that separated Treg and Tconv included canonical Treg and Tconv markers such as *FOXP3*, *IKFZ*, *CTLA4*, *IL2RA* and *IL7R* plus a number of other genes so far not described as Treg or Tconv markers. The typical activation-induced genes *IL2RA* and *ICOS* were found in the signature allowing to distinguish the activation state of Treg, again together with a number of novel marker genes. Our forward approach, defining the transcriptome signatures from defined FACS sorted populations, is unique so far. Transcriptomic signatures for Tregs have been described for bulk cells, using defined populations of Treg and Tconv, memory or naïve, with or without stimulation and using Affimetrix and Nanostring. They contained 31 genes that included *FOXP3* and *IL7R* plus a number of other genes [21]. Others have applied single-cell approaches to whole human CD4$^+$ T cells [22] or to FACS-sorted Tconv and Treg from mouse and humans [23]. These studies also show the importance of canonical markers to distinguish Treg from Tconv. Zemmour et al further demonstrated a sizeable presence of furtive Treg, which shared features with Tconv. Although the fresh isolation by Zemmour et al and our in vitro culture and cell sorting approach differed substantially, it is possible that the Treg observed within the Tconv clusters represent furtive Treg. We expect that the markers described in our study will facilitate the definition of Treg and Tconv in response to an Ag stimulus or in resting conditions in similar studies in the future.

Novel genes found in Tregs were identified in our study. *ZBED2* was observed in around half of the responding Tregs but not in the non-responding Tregs. It had even higher expression in responding Tconv as compared to Treg. *ZBED2* is a sequence-specific transcriptional repressor of IFN-stimulated genes, which occurs through antagonism of IFN regulatory factor 1 (IRF1)-mediated transcriptional activation [24]. To our knowledge, *ZBED2* has not previously been reported in Tregs. Also, of interest are the *LGALS1* gene, that encodes the lectin Galectin-1, and the *CD7* gene, which encodes the Galectin-1 receptor. *LGALS1* was upregulated in responding Treg as compared to non-responding Treg and downregulated as compared to responding Tconv, as described, [25] whereas *CD7* was upregulated in responding and non-responding Treg compared to Tconv. Galectin-1 is reported to attenuate

NF-kB activation through a feedback loop mechanism and is expressed on T cells [26]. Another so far unidentified Treg-specific gene found preferentially expressed in Treg is *TXK*. The resting lymphocyte kinase Txk is a member of non-receptor tyrosine kinases that facilitated downstream signalling after TCR or other receptor activation and it so far only described to be expressed in T or NK cells [27]. *FYB* is another gene whose product regulates signalling downstream of the TCR and that has not yet been described to be preferentially expressed in Treg. The molecular adapter Fyb/Slap regulates integrin clustering and adhesion, coupling TCR stimulation and avidity modulation [28]. Here we find higher *FYB* expression in responding Treg than in responding Tconv. The *TRAT1* gene that encodes TRIM was particularly effective as both a transcriptional and FACS marker. It has not been previously described as a marker to distinguish Treg and Tconv. TRIM was only weakly expressed in both non-responding and responding Treg, but strongly expressed in non-responding and responding Tconv. Its addition to the FACS panel along with CD25, CD127, FOXP3 and IKFZ provided a very effective selection of responding Treg from responding Tconv. Although a membrane protein, its surface portion is very short and, therefore, the use of TRIM as a marker for sorting Treg may not be feasible.

The gene sets we described here provide an important basis to classify cell types from future whole CD4$^+$ T-cell single-cell transcriptome data and therefore are a useful resource to characterise T-cell responses in health and disease and after immune intervention.

CONFLICT OF INTEREST

There are no conflict of interest to be disclosed.

ETHICAL APPROVAL

The use of human samples was approved by ethics committee and informed consent of the donors was obtained (EK240062016).

ORCID

Anne Eugster https://orcid.org/0000-0001-8009-5959

REFERENCES

1. Trzupek D, Dunstan M, Cutler AJ, Lee M, Godfrey L, Jarvis L, et al. Discovery of CD80 and CD86 as recent activation markers on regulatory T cells by protein-RNA single-cell analysis.

Genome Med. 2020;12:55. https://doi.org/10.1186/s13073-020-00756-z

2. Piccirillo CA, d'Hennezel E, Sgouroudis E, Yurchenko E. CD4+Foxp3+ regulatory T cells in the control of autoimmunity: in vivo veritas. Curr Opin Immunol. 2008;20:655–62.

3. Sakaguchi S, Yamaguchi T, Nomura T, Ono M. Regulatory T cells and immune tolerance. Cell. 2008;133:775–87.

4. Hori S, Nomura T, Sakaguchi S. Control of regulatory T cell development by the transcription factor Foxp3. Science. 2003;299(5609):1057–1061.

5. Dieckmann D, Plottner H, Berchtold S, Berger T, Schuler G. Ex vivo isolation and characterization of CD4+CD25+ T cells with regulatory properties from human blood. J Exp Med. 2001;193:1303–10.

6. Baecher-Allan C, Brown JA, Freeman GJ, Hafler DA. CD4 + CD25 high regulatory cells in human peripheral blood. J Immunol. 2001;167(3):1245–1253.

7. Seddiki N, Santner-Nanan B, Martinson J, Zaunders J, Sasson S, Landay A, et al. Expression of interleukin (IL)-2 and IL-7 receptors discriminates between human regulatory and activated T cells. J Exp Med. 2006;203:1693–700.

8. Klein S, Kretz CC, Krammer PH, Kuhn A. CD127 low/and FoxP3 expression levels characterize different regulatory T-cell populations in human peripheral blood. J Invest Dermatol. 2010;130:492–9.

9. Chattopadhyay PK, Gierahn TM, Roederer M, Love JC. Single-cell technologies for monitoring immune systems. Nat Immunol. 2014;15:128–35.

10. Nathan A, Baglaenko Y, Fonseka CY, Beynor JI, Raychaudhuri S. Multimodal single-cell approaches shed light on T cell heterogeneity. Curr Opin Immunol. 2019;61:17–25.

11. Picelli S, Faridani OR, Björklund ÅK, Winberg G, Sagasser S, Sandberg R. Full-length RNA-seq from single cells using Smart-seq2. Nat Protoc. 2014;9:171–81.

12. Kharchenko PV, Silberstein L, Scadden DT. Bayesian approach to single-cell differential expression analysis. Nat Methods. 2014;11:740–2.

13. Lun ATL, McCarthy DJ, Marioni JC. A step-by-step workflow for low-level analysis of single-cell RNA-seq data with bioconductor. F1000Res. 2016;5:2122.

14. Amezquita RA, Lun ATL, Becht E, Carey VJ, Carpp LN, Geistlinger L, et al. Orchestrating single-cell analysis with bioconductor. Nat Methods. 2020;17:137–45.

15. McCarthy DJ, Campbell KR, Lun ATL, Scater WQF. Preprocessing, quality control, normalization and visualization of single-cell RNA-seq data in R. Bioinformatics. 2017;33:1179–86.

16. Becht E, McInnes L, Healy J, Dutertre CA, Kwok IWH, Ng LG, et al. Dimensionality reduction for visualizing single-cell data using UMAP. Nat Biotechnol. 2019;37:38–44.

17. Knoop J, Eugster A, Gavrisan A, Lickert R, Sedlmeier EM, Dietz S, et al. Maternal type 1 diabetes reduces autoantigen-responsive CD41 T cells in offspring. Diabetes. 2020;69:661–9.

18. Buettner F, Natarajan KN, Casale FP, Proserpio V, Scialdone A, Theis FJ, et al. Computational analysis of cell-to-cell heterogeneity in single-cell RNA-sequencing data reveals hidden subpopulations of cells. Nat Biotechnol. 2015;33:155–60.

19. McDavid A, Dennis L, Danaher P, Finak G, Krouse M, Wang A, et al. Modeling bi-modality improves characterization of cell cycle on gene expression in single cells. PLoS Comput Biol. 2014;10(7):e1003696.

20. Chen G, Ning B, Shi T. Single-cell RNA-seq technologies and related computational data analysis. Front Genet. 2019;10:317.

21. Pesenacker AM, Wang AY, Singh A, Gillies J, Kim Y, Piccirillo CA, et al. A Regulatory T-cell gene signature is a specific and sensitive biomarker to identify children with new-onset type 1 diabetes. Diabetes. 2016;65:1031–9.

22. Bradley A, Hashimoto T, Ono M. Elucidating T cell activation-dependent mechanisms for bifurcation of regulatory and effector T cell differentiation by multidimensional and single-cell analysis. Front Immunol. 2018;9:1444.

23. Zemmour D, Zilionis R, Kiner E, Klein AM, Mathis D, Benoist C. Single-cell gene expression reveals a landscape of regulatory T cell phenotypes shaped by the TCR article. Nat Immunol. 2018;19:291–301.

24. Somerville TDD, Xu Y, Wu XS, Maia-Silva D, Hur SK, de Almeida LMN, et al. ZBED2 is an antagonist of interferon regulatory factor 1 and modifies cell identity in pancreatic cancer. Proc Natl Acad Sci USA. 2020;117:11471–82.

25. Garín MI, Chu NC, Golshayan D, Cernuda-Morollón E, Wait R, Lechler RI. Galectin-1: A key effector of regulation mediated by CD4 +CD25+ T cells. Blood. 2007;109:2058–65.

26. Toscano MA, Campagna L, Molinero LL, Cerliani JP, Croci DO, Ilarregui JM, et al. Nuclear factor (NF)-κB controls expression of the immunoregulatory glycan-binding protein galectin-1. Mol Immunol. 2011;48(15–16):1940–9.

27. Schaeffer EM, Debnath J, Yap G, McVicar D, Liao XC, Littman DR, et al. Requirement for Tec kinases Rlk and ltk in T cell receptor signaling and immunity. Science. 1999;284:638–41.

28. Griffiths EK, Krawczyk C, Kong YY, Raab M, Hyduk SJ, Bouchard D, et al. Positive regulation of T cell activation and integrin adhesion by the adapter Fyb/Slap. Science. 2001;293:2260–3.

SUPPORTING INFORMATION

Additional supporting information may be found in the online version of the article at the publisher's website.