RESEARCH

Open Access

A time-resolved meta-analysis of consensus gene expression profiles during human T-cell activation

Michael Rade[1*] , Sebastian Böhlen[2], Vanessa Neuhaus[2], Dennis Löffler[1], Conny Blumert[1], Maximilian Merz[3], Ulrike Köhl[1,4], Susann Dehmel[2], Katherina Sewald[2†] and Kristin Reiche[1,4,5†]

†Katherina Sewald and Kristin Reiche shared the last authorship.

*Correspondence: michael.rade@izi.fraunhofer.de

[1] Department of Diagnostics, Fraunhofer Institute for Cell Therapy and Immunology, Perlickstraße 1, 04103 Leipzig, Germany
[2] Department of Preclinical Pharmacology and In-Vitro Toxicology, Fraunhofer Institute for Toxicology and Experimental Medicine, ITEM, Nikolai-Fuchs-Strasse 1, 30625 Hannover, Germany
[3] Department of Hematology, Cellular Therapy, Hemostaseology and Infectiology, University Hospital of Leipzig, Leipzig, Germany
[4] Institute for Clinical Immunology, Leipzig University, Johannisallee 30, 04103 Leipzig, Germany
[5] Center for Scalable Data Analytics and Artificial Intelligence (ScaDS.AI), Dresden, Leipzig, Germany

Abstract

Background: The coordinated transcriptional regulation of activated T-cells is based on a complex dynamic behavior of signaling networks. Given an external stimulus, T-cell gene expression is characterized by impulse and sustained patterns over the course. Here, we analyze the temporal pattern of activation across different T-cell populations to develop consensus gene signatures for T-cell activation.

Results: Here, we identify and verify general biomarker signatures robustly evaluating T-cell activation in a time-resolved manner. We identify time-resolved gene expression profiles comprising 521 genes of up to 10 disjunct time points during activation and different polarization conditions. The gene signatures include central transcriptional regulators of T-cell activation, representing successive waves as well as sustained patterns of induction. They cover sustained repressed, intermediate, and late response expression rates across multiple T-cell populations, thus defining consensus biomarker signatures for T-cell activation. In addition, intermediate and late response activation signatures in CAR T-cell infusion products are correlated to immune effector cell-associated neurotoxicity syndrome.

Conclusion: This study is the first to describe temporally resolved gene expression patterns across T-cell populations. These biomarker signatures are a valuable source, e.g., monitoring transcriptional changes during T-cell activation with a reasonable number of genes, annotating T-cell states in single-cell transcriptome studies, or assessing dysregulated functions of human T-cell immunity.

Keywords: T-cell activation, Non-negative matrix factorization, Transcriptome, Gene expression, Temporal gene profiles, Biomarkers, Time series

Background

T-cells are a subgroup of lymphocytes and mediate the adaptive cellular immune response. T-cell priming and initiation of activation to a foreign antigen requires engagement of T-cell receptors (TCR) with the cognate peptide-MHC complex

Rade *et al. Genome Biology*   (2023) 24:287

Page 2 of 26

presented by antigen-presenting cells (APCs), inducing co-stimulatory signals such as CD28 and OX40, and cytokines released by APCs that drive T-cell differentiation [1, 2]. These processes initiate coordinated transcriptional changes of genes. T-cells exit the quiescent state by down-regulation of genes associated with maintenance of resting status [3–5] followed by metabolic reprogramming to provide energy for growth. This includes, for example, an increased mitochondrial biogenesis, glycolysis, or oxidative phosphorylation, which is coupled with the tricarboxylic acid (TCA) cycle. Furthermore, T-cells respond to autocrine or paracrine Interleukin 2 (IL2) signaling and initiate effector functions, such as cytokine production [3, 6]. Another important aspect is post-transcriptional regulation defining T-cell rewiring and maintaining T-cell quiescence. Post-transcriptional regulation determines protein production by regulating RNA splicing, mRNA stability, RNA localization, and translation machinery [7, 8]. All these processes must be coordinated in a temporal fashion.

In the past years, time series of transcriptome-wide studies (microarray, bulk RNA-sequencing (RNA-Seq), and single-cell RNA-Seq) assessed activation and differentiation of T-cells into distinct T-cell populations from healthy donors [9–19]. Although these studies provided enormous insights into molecular mechanisms during T-cell activation and differentiation, no transcriptome study with a time-resolved description of the transient expression changes from TCR stimulation via signal pathways to proliferation across T-cell populations is reported. Given that the distribution of T cell populations differs largely between healthy individuals [20] and from immune condition to immune condition (e.g., cytokine storm or allergy), we aim at time-resolved consensus gene expression profiles characterizing T cell activation to proliferation across lineage choices.

Applications of time-resolved T-cell biomarker signatures that are independent from the underlying inflammatory disease condition are wide-spread. For example, there is a need for human-based in vitro models that allow objective assessment of T-cell dependent inter-individual immune responses during non-clinical development of new candidates for immunotherapies. Adoptive cancer immunotherapies [21] represent a great opportunity to treat cancer but also come with the challenge of predicting immunogenicity prior to first-in-human studies. Prediction of immunogenicity of these novel therapies is based on the generation of an adaptive immune response and can be linked to an early key event, the T-cell activation [22], for which T-cell activation across lineages is considered as an important endpoint [23]. Thus, gene signatures robustly monitoring T-cell activation across different T-cell populations, if developed as biomarker gene signatures, i.e., as "measurable and evaluable indicators for the underlying" [24] T-cell activation processes, enables robust monitoring of inter-individual immune responses in non-clinical models.

Further, advances in manufacturing processes and increasing patient numbers suitable for cellular immunotherapies like autologous CAR T-cell therapy reinforce the importance of robust biomarker signatures predicting the toxicity and/or efficacy of CAR T-cells [25]. Response rates of CAR T-cell therapies strongly depend on the cellular fitness of CAR T-cells in the infusion product. Biomarker signatures as consensus gene expression profiles characterizing temporal cellular changes throughout

T-cell activation could function as T-cell intrinsic descriptors correlating with the efficacy and safety of CAR T-cells.

Lastly, characterizing a consensus gene expression profile for human T-cell activation is of value to reveal characteristics of an overall immune response no matter the stressor or immune condition, thus, enabling systems immunology to identify by which the T-cell response of an individual varies from the consensus under specific conditions [20].

Hence, we describe biomarker signatures as consensus gene expression profiles characterizing temporal cellular changes throughout T-cell activation. In this study, we provide a novel and unbiased time-resolved set of genes whose expression patterns correlate strongly with T-cell stimulation processes. We performed a meta-analysis of T-cell activation using time-series transcriptome datasets to develop kinetic consensus gene signatures across multiple T-cell populations. Our results characterize gene expression changes in a coordinated and temporal fashion and provide general biomarker signatures for human T-cell activation.

Results

Towards time-resolved consensus gene expression signatures for T-cell activation

To identify time-resolved consensus gene expression profiles across T-cell populations, we followed a three-step approach. Firstly, we used a *Discovery Set* of publicly available transcriptome data. We identified genes with a significant combined effect size among T-cell populations and studies in at least one time point of T-cell activation. T-cells were activated by stimulation with anti-CD3/anti-CD28 coated beads or in the presence of stimulation beads and differentiating cytokines (see Additional file 1: Fig. S1). Secondly, we combined genes into classes of coherent metagenes by unsupervised decomposition. Thirdly, we confirmed the time-resolved consensus signatures in 2 independent *Verification Sets* (Fig. 1).

The *Discovery Set* included transcriptome-wide expression data of in vitro kinetic models of CD4$^+$ T-cells, previously profiled by publicly available time series microarrays [10, 11] or RNA-Seq [12–14] studies. In these studies, CD4$^+$ T-cells were obtained from healthy donors consisting of either adult PBMCs or neonatal cord blood. The CD4$^+$ T-cells were collected between 0 (before activation) and 72 h after activation with anti-CD3/anti-CD28 coated beads only (Th0) or in the presence of differentiating cytokines, leading to polarization towards Th1, Th2, Th17, and iTreg T-cell fates. The datasets of Th0 and iTreg consist of multiple studies and sample sources (PBMCs and cord blood). After removing potential batch effects across studies, we observed no significant correlation between the study or sample source and principal components (see Additional file 1: Fig, S4). Except for Th1, 9 overlapping time points of activation were available in the T-cell populations (Fig. 1). Overall, we used 224 samples with up to 12 replicates per time point of activation and CD4$^+$ T-cell population in the *Discovery Set*.

For verification, we re-analyzed an independent RNA-Seq dataset [15] (referred to as *Memory T-cell Verification Set*) that included unactivated as well as with anti-CD3/CD28 beads activated memory CD4$^+$ T-cells from 24 healthy donors between 2 and 72 h (overall 184 samples).

In addition, we isolated Pan T-cells from 4 healthy donors (for MACS isolation strategy of Pan T-cells, see Additional file 1). RNA-Seq was performed at 5 different time

Rade *et al. Genome Biology*   (2023) 24:287

Page 4 of 26
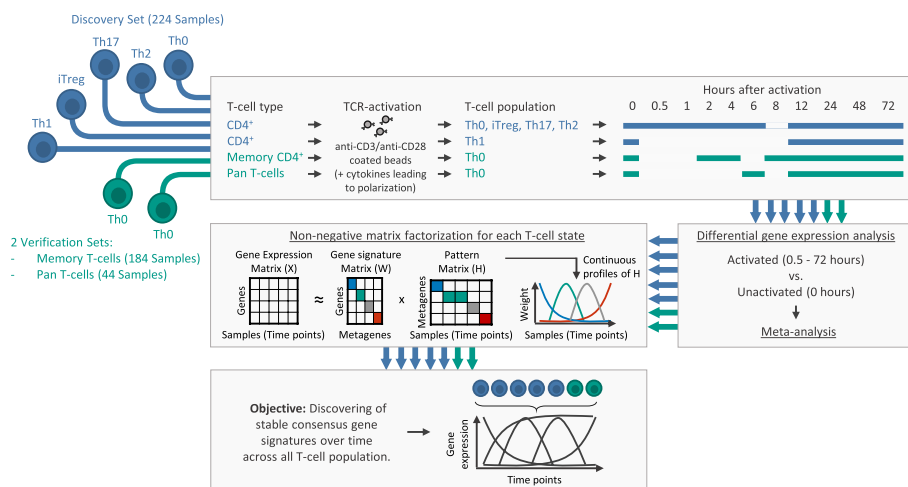
Fig. 1 Workflow for discovering and verifying of temporal consensus gene expression signatures of T-cells. We analyzed transcriptome data from T-cells after in vitro activation with anti-CD3/anti-CD28 coated beads (Th0) or in the presence of stimulation beads and differentiating cytokines (polarization towards Th1, Th2, Th17, and iTreg T-cell fates). For each T-cell population, we performed DGEA to find DE (FDR < 0.05) genes of activated T-cell populations at different analysis time points compared to unactivated T-cell populations (time series of gene expression arrays and RNA-seq data). To find DE genes with a significant combined effect size (FDR < 0.05) across CD4$^+$ T-cell populations from the *Discovery Set* (highlighted in blue), we conducted a meta-analysis. Only DE genes with a significant combined effect size in at least one contrast (0.5 to 72 h vs. 0 h) across the available populations (4 populations for time course 0.5 to 6 h of activation, 5 populations for time course 12 to 72 h of activation) were used for NMF. We conducted NMF to infer expression changes over time and to discover stable continuous metagenes (i.e., sets of genes with similar expression patterns across the analysis time points) across all T-cell populations. For verification of the temporal consensus gene signature, we analyzed 2 independent RNA-Seq datasets (highlighted in green). The *Discovery-* and *Memory T-cell Verification Set* are based on publicly available datasets

points of anti-CD3/anti-CD28 activation (6, 12, 24, 48, and 72 h) and from unactivated (0 h) Pan T-cells. We additionally cultured and sequenced Pan T-cells as negative controls in a serum-free medium without anti-CD3/anti-CD28 activation at the same time points as mentioned above. We refer to this dataset as the *Pan T-cell Verification Set*. Overall, 44 samples were analyzed (Figs. 1 and 5A).

For a detailed overview of all datasets used for this study, see Additional file 1: Table S1 and Fig. S1. Principal component analysis of all analyzed samples for all T-cell populations revealed no critical outlier (Additional file 1: Fig. S5 and Fig. S3).

Time-resolved transcriptome-wide changes depict common trends across CD4$^+$ T-cell populations

For each CD4$^+$ T-cell population in the *Discovery Set* and at each time point of activation, we assessed differential gene expression between polarized or activated T-cells without differentiating cytokines and CD4$^+$ T-cells before activation. The maximum number of significantly differentially expressed (DE) genes among all 5 CD4$^+$ T-cell populations was 3641 after 24 h of activation (Fig. 2A and see Additional file 1: Fig. S7 for a more detailed overview). We observed coherent signs of fold changes for more than 96.5% of all genes identified as DE in at least 2 populations (Fig. 2B). We observed 124 genes that were significant in all populations at a given activation time point and had the same logFC direction in all but one population (Additional file 1: Fig. S8).
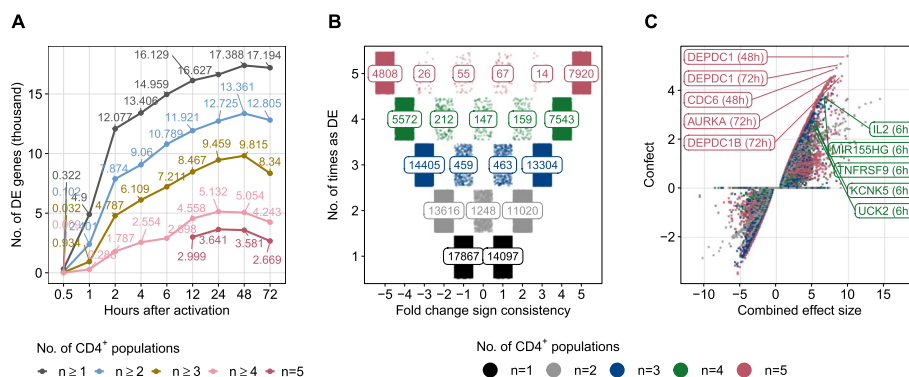
Fig. 2 Common trends across CD4+ T-cell populations from the *Discovery Set*. **A** Inverse cumulative distributions of DE genes (FDR < 0.05) in the *Discovery Set* for each time point during activation compared to unactivated CD4+ T-cells. The colored curves represent the number of DE genes in at least *n* T-cell populations. Analysis time points 12 to 72 h were available for all 5 CD4+ T-cell populations. **B** Depicting the number of T-cell populations in which genes were identified as DE (*y*-axis) and the number of genes with consistent fold changes across T-cell populations (*x*-axis). For example, a gene that is significantly differentially expressed in 3 T-cell populations after 12 h of activation, of which it is down-regulated in 2 populations, will get a fold change sign consistency of $1 + (-2) = -1$ (*x*-axis). An example of how to read the numbers: 7920 genes (top right) were identified as DE in all 5 T-cell populations. These DE genes were also upregulated in all 5 T-cell populations at the same time point of activation. Colors represent the number of T-cell populations in which genes were significantly differentially expressed. **C** Shown are DE genes with a combined effect size identified in the meta-analysis using a random effect model in at least 2 T-cell populations. The *x*-axis represents the combined effect size, the *y*-axis the "confect" value, a confident inner bound of the calculated combined effect size (see the "Methods" sction). Genes that do not show a significant combined effect size (FDR > 0.05) have a "confect" value of 0

Next, we conducted a meta-analysis to identify DE genes with a significant combined effect across CD4+ T-cell populations in the *Discovery Set*. For all DE genes at least 2 populations at the same time point of activation, we estimated a combined effect size by applying a random-effects model (see the "Methods" section). With an FDR < 0.05, we identified 12,581 DE genes with a significant combined effect size across all 5 CD4+ T-cell populations after 12 to 72 h of activation, the majority (90.4%) of which were in at least 2 time points. Overall, 94.2% of all genes that were DE in at least 4 CD4+ T-cell populations had also a significant combined effect size (Additional file 1: Fig. S9D).

The 5 highest ranked genes from the meta-analysis across 4 (0.5 to 6 h) and 5 (12 to 72 h) T-cell populations are depicted in Fig. 2C. IL2, TNF Receptor Superfamily Member 9 (TNFRSF9), and the miR-155 host gene (MIR155HG) were among the 5 highest ranked genes after 6 h of activation across 4 T-cell populations. The most prominent genes across 5 T-cell populations included cell cycle-associated genes such as AURKA, CDC6, DEPDC1B, and DEPDC1 [26–29]. The 20 highest ranked genes with a significant combined effect size and corresponding forest plots for each activation time point are described in Additional file 1: Fig. S11. The 30 highest enriched Reactome pathways and GO terms (FDR < 0.05) for DE genes with a significant combined effect size are shown in Additional file 1: Fig. S12 and S13. Antigen stimulation of T-cell receptor (TCR) signaling to nuclear factor kappa B (NF-κB) is required for T-cell proliferation and polarization and is represented among the enriched Reactome pathways. All DE genes enriched in TCR signaling are shown in Additional file 1: Fig. S14. GO terms of T-cell activation were significantly enriched mainly after one hour of activation.

Rade *et al. Genome Biology*   (2023) 24:287

Page 6 of 26

Unsupervised decomposition revealed coherent metagenes across CD4[+] T-cell populations in the *Discovery Set*

To identify coherent temporal expression profiles for each CD4[+] T-cell population in the *Discovery Set*, we employed non-negative matrix factorization (NMF), an unsupervised decomposition approach. NMF allows a part-based representation by reducing the dimensionality of the expression data and describing the samples as a composition of metagenes (i.e., sets of genes with similar expression patterns across time points). Based on consensus clustering as qualitative measurement combined with the cophenetic correlation coefficients and silhouette scores as quantitative measures, NMF resulted in the greatest stability with 3–5 metagenes for each CD4[+] T-cell population (Additional file 1: Fig. S16).

The resulting pattern matrix represents the relative weights of metagene *k* in sample *m*. A graphical representation of the metagene patterns is depicted in Fig. 3A. We observed different temporal metagene profiles for considered time points. All CD4[+] T-cell populations were characterized by a sustained repressed and late response to activation and at least one intermediate expression pattern. In addition, metagenes showed consistent temporal peaks across all populations, which justifies an alignment of metagenes between CD4[+] populations.

The gene signature matrix provided gene sets for each metagene, which can be linked to biological processes by enrichment analysis or functional annotation. The top 15 significantly enriched Reactome pathways and Gene Ontology (GO) terms are shown in Additional file 1: Fig. S18 and S19. FOXO-mediated transcription was identified in metagenes M1 in 4 out of 5 CD4[+] T-cell populations. T-cell activation GO terms were significantly enriched in the intermediate metagene M2 with an expression pattern peak at 2 h, while metabolic processes were enriched in intermediate metagene M3 with an expression pattern peak at 4 to 12 h. Gene sets from the late response metagenes M5 were strongly enriched for cell cycle-associated GO terms and pathways from the Reactome database. The 10 highest ranked genes for each metagene and T-cell population are shown in Fig. 3B (Additional file 1: Fig. S17 for the 25 highest ranked genes).

Consensus gene expression profiles for CD4[+] T-cells

To obtain a more robust time series and to explore the consensus transcriptional regulators of T-cell activation, we combined metagenes with similar expression patterns among CD4[+] T-cell populations in the *Discovery Set*. This means that the metagenes had to be temporally consistent across all CD4[+] T-cell populations. This resulted in a set of genes with similar expression profiles over the course across CD4[+] T-cell populations. Furthermore, to refine the variety of gene expression profiles, we defined identity and shared metagene. For illustrative purposes, we developed a metagene landscape reflecting the relative relationship of genes and samples to metagenes (Fig. 4A). As an example, we used IL2, IL2RA, CD4, and NF-κB (NFKBID), which are part of our consensus gene expression profiles of the *Discovery Set* and depicted the mean metagene weights from the CD4[+] T-cell populations as dots in the metagene landscape. CD4 was highly expressed in all T-cell populations. However, we observed that CD4 embedded between metagene M1 and M5, which represents the sustained repressed and late response

Rade *et al. Genome Biology*   (2023) 24:287

Page 7 of 26
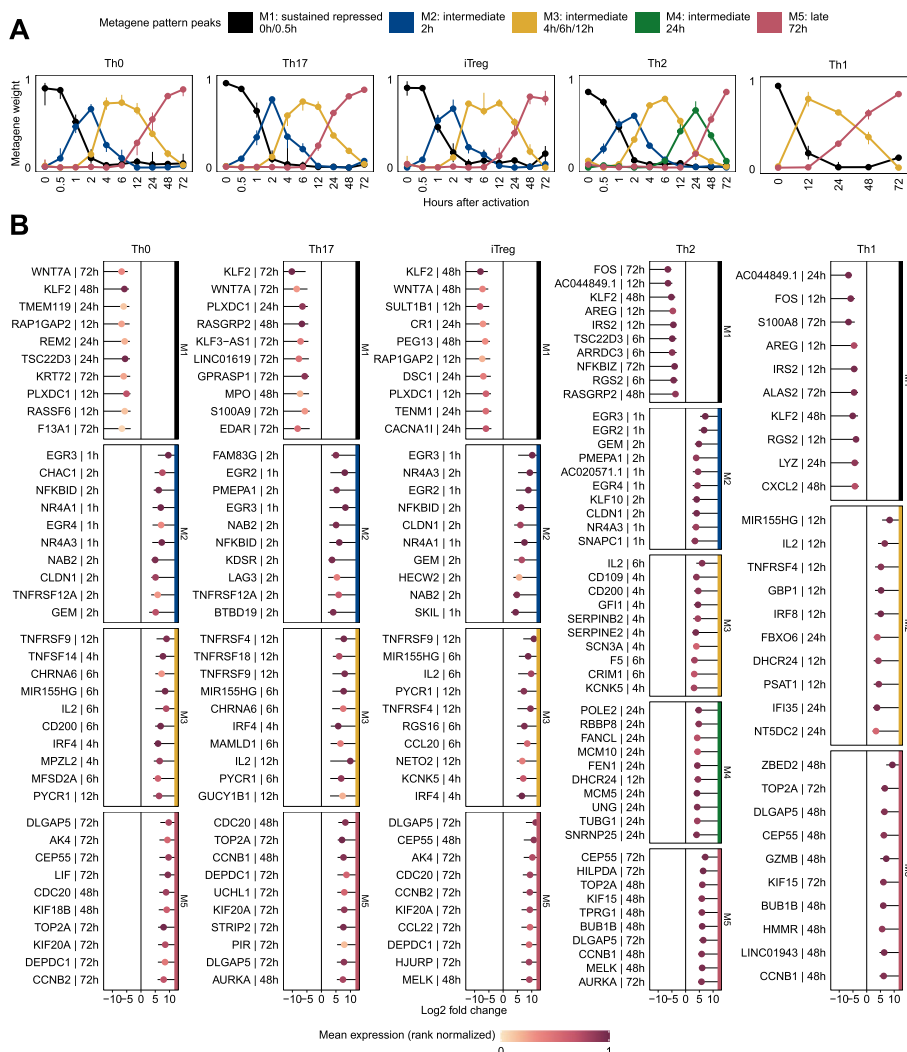
Fig. 3 Temporal profiles obtained from NMF. **A** The pattern matrix for each CD4⁺ T-cell population from the *Discovery Set* is shown as continuous profiles, with samples assigned to time points of activation. We scaled each column in the pattern matrix to sum up to one. Dots depict median weights for all samples from identical analysis time points. Vertical lines represent interquartile ranges. We annotated and colored the metagenes based on their maximum median values across all analysis time points. The time point with the maximum median value is depicted in the legend. **B** Top 10 genes associated with metagenes for each T-cell population. For each CD4⁺ T-cell population and gene used for NMF, we used the highest absolute "confect" value estimated in the DGEA across all contrasts (e.g., 12 h vs. 0 h). Genes are ranked by "confect" values. Dots represent log2 fold changes for contrasts with the highest absolute "confect" value. The time point to the right of the gene represents the contrast with the highest absolute "confect" value. For example, "12 h" represents the following contrast: a sample group activated for 12 h compared to unactivated samples of the same group. The color of the dots corresponds to rank normalized average expression values of the activation group in contrast with the highest absolute "confect." The inner end of the horizontal line shows the "confect" value (inner confidence bound). NFKBID denotes NF-κB

metagene and was expressed at a higher level compared to the other samples (Fig. 4B). On the other hand, IL2 was embedded near metagene M3, which represents an intermediate response metagene and is expressed at higher levels in samples at 6 to 24 h after activation (Fig. 4B) compared with the other samples. This indicates that there are set of genes associated with more than one metagene. Therefore, we extended the definition of

Rade *et al. Genome Biology*   (2023) 24:287

Page 8 of 26

metagenes and genes associated with them. Genes with high weights in one metagene from the gene signature matrix were referred to as genes linked to identity metagenes. Genes with higher weights in more than one metagene were referred to as genes linked to shared metagenes (see the "Methods" section).

Furthermore, we used the "Housekeeping and Reference Transcript Atlas database" [30] to remove 183 housekeeping genes from the consensus signatures constitutively expressed in different tissues and cell types (Fig. 5F).

The resulting consensus gene expression profiles, separated into identity and shared metagenes-associated genes, are shown in Fig. 4C. In addition to the 4 identity metagenes, we were able to identify another 6 shared metagenes. The top significantly enriched GO term for biological processes (FDR < 0.05) for genes linked to identity and shared metagenes indicate a distinct functional context (Fig. 4E). Metagene M2 is enriched with GO terms of T-cell activation regulation, intracellular receptor signaling pathway, and leukocyte cell–cell adhesion. Genes from the shared metagene M1/M2 are enriched with GO terms of cellular responses to stimuli and MAPK cascade. Metagene M3 and M5 are linked to metabolic processes and cell cycle terms, respectively (see Additional file 1: Fig. S21 for the top enriched Reactome pathways and Additional file 2 for expression profiles of all genes from the consensus signatures and all associated pathways/terms).

Verification and mapping to a wider variety of T-cell transcriptional landscape

To verify the consensus gene expression profiles and additionally represent a wider variety of the T-cell landscape, we used *2 Verification Sets*. Firstly, we used RNA-Seq activation kinetics of memory CD4$^+$ T-cells from Gutierrez-Arcelus et al. [15] (referred to as *Memory T-cell Verification Set*). Secondly, we performed time series RNA-seq of activated and unactivated Pan T-cells from 4 healthy donors (Fig. 5A). We used this dataset as Pan T-cell Verification Set to exclude from the consensus signature those genes whose expression profiles deviated from conditions as close as possible to human blood (i.e.,

(See figure on next page.)

Fig. 4 Consensus gene expression profiles for CD4$^+$ T-cells from the *Discovery Set*. **A** Metagene landscape. We embedded all samples relative to temporally coherent metagenes using the pattern matrix. All genes that we used for NMF were embedded relative to the temporally coherent metagenes using the gene signature matrix and depicted as a density map. For IL2, IL2RA, CD4, and NF-κB (NFKBID) we calculated the average metagene weights across the CD4$^+$ T-cell populations and depicted them as dots in the metagene landscape. **B** Each sample is colored by the rank normalized gene expression of the corresponding gene. **C** We grouped the consensus expression profiles over the course by genes associated with identity and shared metagenes. Each boxplot represents one CD4$^+$ T-cell population from the *Discovery Set*. The *y*-axis depicts the standardized median expression of genes from samples with identical analysis time points. The number in parentheses represents the number of genes for the corresponding metagene. **D** Top 15 genes associated with identity metagenes. For each gene belonging to the consistent metagenes, we used the highest absolute "confect" value estimated in the meta-analysis. Genes were ranked according to their absolute highest "confect" value. Diamonds represent the combined effect size from the meta-analysis for the activation time point with the highest absolute "confect" value. The time point to the right of the gene represents the time point of the highest absolute "confect" value. The inner end of the horizontal line shows the "confect" value (inner confidence bound). **E** The top 10 (sorted by rich factor) significantly enriched GO terms of biological processes (FDR < 0.05) of metagene-associated genes identified by enrichment analysis. The dot size indicates the rich factor, which is the number of metagene-associated genes in the GO term divided by the number of background genes of the term. Colors indicate adjusted *p*-values of significantly enriched GO terms
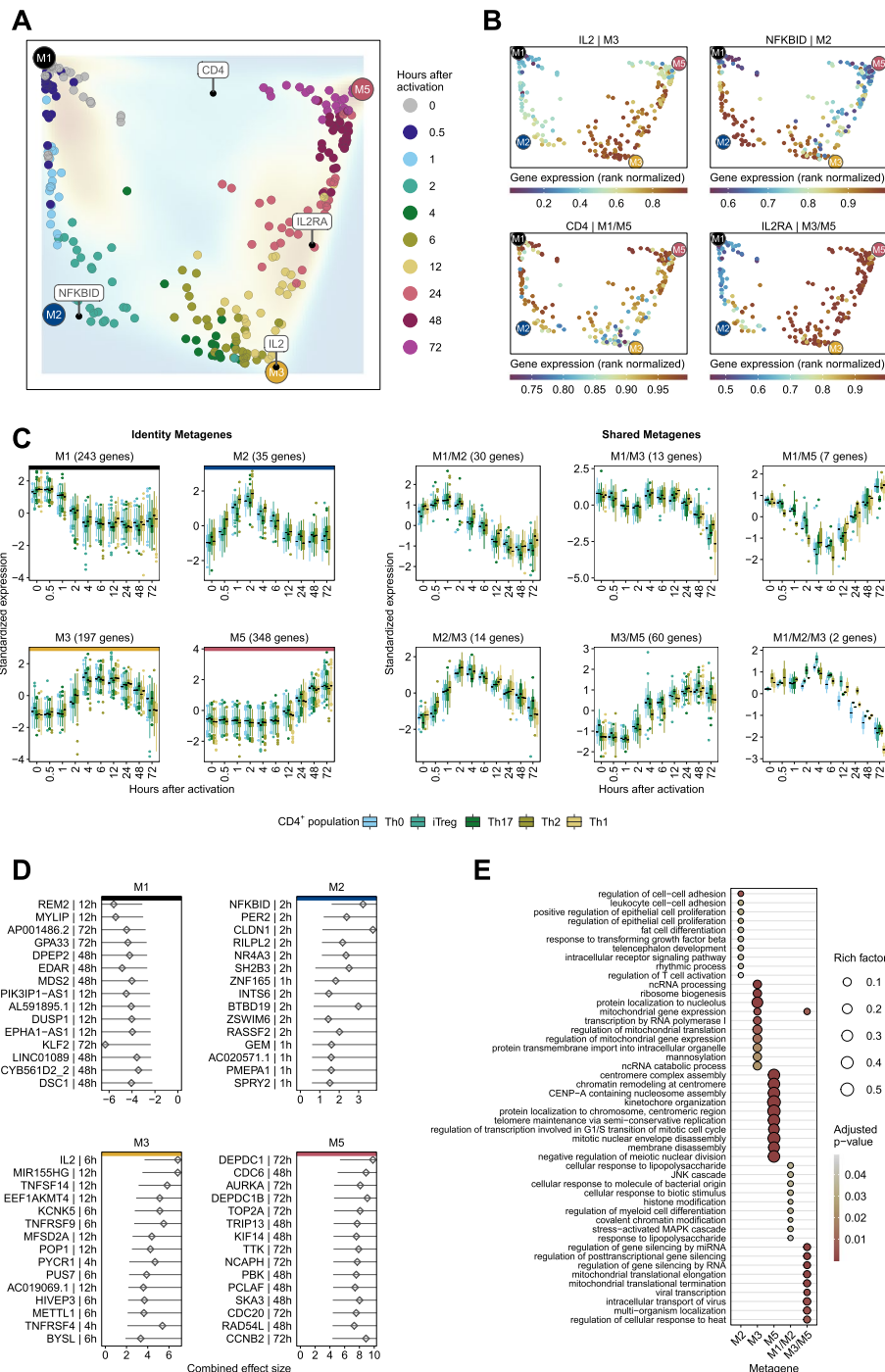
Fig. 4 (See legend on previous page.)

conditions where cells are not sorted prior to analysis). The fractions of the Pan T-cell populations are shown in Fig. 5B. To find stable metagenes for each activation kinetic of the 2 *Verification Sets,* we used the same procedure as for the *Discovery Set* (Additional file 1: Fig. S16). The temporal profiles of the *Verification Sets* calculated from the pattern matrix and metagene-associated genes are shown in Additional file 1: Fig. S22.

Rade *et al. Genome Biology*   (2023) 24:287

Page 10 of 26
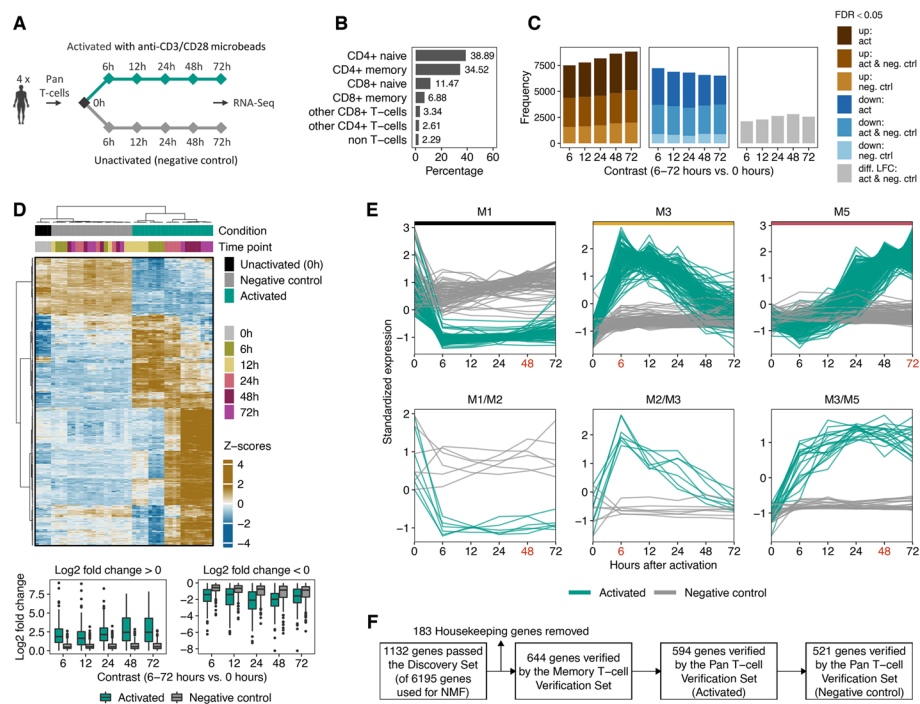
Fig. 5 Verification of consensus temporal gene signature. **A** Experimental design for the *Pan T-cell Verification Set*. We performed RNA Sequencing of Pan T-cells from 4 healthy donors at 5 different time points after anti-CD3/anti-CD28 activation (6 to 72 h) and of unactivated (0 h) T-cells. In addition, we sequenced Pan T-cells for the same time points without activation as negative controls. **B** Depicted are the fractions of Pan T-cell populations before activation. 200,000 cells were analyzed using seven human blood donors (see Additional file 1 for details). **C** For each contrast (6 to 72 h of activation and without activation vs. 0 h), we performed a DGEA. The brown and blue bar plots depict the number of DE genes (FDR < 0.05) for each contrast that is up- (brown) or down-regulated (blue) under activated (act), unactivated (neg. ctrl) conditions, or both (act and neg. ctrl). Gray bar plots show the number of DE genes under activated and unactivated conditions without consistent log2 fold change. **D** Hierarchical clustering of DE genes from the *Pan T-cell Verification Set* in activated and unactivated conditions. Only genes from the consensus signatures that passed the activation kinetics of both *Verification Sets* are shown. Euclidean distance with Ward clustering was applied to visualize the similarity between samples. Each column represents a sample, each row represents a gene. The *y*-axis depicts the standardized median CPM expression values of genes from samples with identical analysis time points. Bottom panel: For each contrast and condition, boxplots of log2 fold changes are shown. **E** For the *Pan T-cell Verification Set* the temporal expression pattern of genes from the consensus signatures that passed the 2 *Verification Sets* (activation and negative control kinetics) are shown. CPM values were *z*-score standardized. Identity metagenes are highlighted with colored horizontal bars (M1, M3, and M5). The red-colored time points indicate the maximum centroid threshold (see the "Methods" section). Only metagenes with more than 10 genes are shown. **F** Flowchart depicting the number of genes passing each filter step

Next, we compared activated and unactivated (negative control) gene expression kinetics of T-cells from the *Pan T-cell Verification Set*. We performed DGEA to find differentially expressed genes for activated and unactivated T cells at 6 to 72 h compared to unactivated T-cells at 0 h. DGEA revealed that across all contrasts (6 to 72 h vs. 0 h), an average of about 2700 genes were significantly up- or down-regulated (FDR < 0.05) in both anti-CD3/CD28 bead-activated and unactivated conditions (Fig. 5C). Significantly enriched Reactome pathways for DE genes under activated and unactivated conditions are shown in Additional file 1: Fig. S25. For the unactivated conditions, we observed that pathways associated with metabolic processes, transcription, and proliferation were significantly enriched for downregulated DE genes. Even though mostly disjunct enriched

pathways were observed for upregulated DE genes under activated conditions, they are also associated with metabolic processes, transcription, and proliferation. Heatmap visualization of DE genes in activated and unactivated conditions showed that temporal impulse patterns were not observed under unactivated conditions (Additional file 1: Fig. S26). Although we observed two sustained temporal patterns under unactivated conditions, samples with identical analysis time points were not clustered together.

All genes from the consensus signatures that we identified based on the *Discovery Set* were verified for temporal consistency by the *Verification Sets* (see the "Methods" section). This results in 594 genes. However, of those 457 also showed significant expression changes in at least one contrast in negative controls. Unsupervised clustering of those genes resulted in distinct expression clusters according to conditions, but partly similar temporal patterns among conditions (see Fig. 5D). The majority showed only minor log2 fold changes in negative controls compared to unactivated conditions at 0 h. About 37% of all DE genes had a log2 fold change > 1 in at least one contrast under unactivated conditions (Fig. 5D, bottom panel).

Therefore, we compared activated T-cells with negative controls at the time point with maximum absolute distance to the centroid of the metagenes (see the "Methods" section). We only retained genes from the consensus signatures with significant differential expression between activated and negative controls (FDR < 0.05) and an absolute "confect" value > 1 at the selected time point. The temporal expression patterns of genes passing this filter step are shown in Fig. 5E (see Additional file 1: Fig. S27 for all metagenes and Fig. S28 for genes expression profiles not passing this filter step).

Overall, passing all filtering steps, this resulted in 521 genes being included in the final consensus gene signatures (Fig. 5F). For the highest ranked genes, we provide an overview linking genes to the metagenes and the top enriched Reactome pathways/GO terms in Additional Fig. S29 and S30 (see Additional file 2, which shows an overview of all genes and their temporal expression as an interactive HTML document). In addition, we also provide a comprehensive characterization of genes passing the *Memory T-cell Verification Set* in Additional file 1: Fig. S23 and Additional file 2, respectively.

Consensus gene expression profiles are enriched in anti-CD19 CAR T-cell products from patients with low-grade ICANS

We re-analyzed a scRNA-Seq study of autologous axicabtagene ciloleucel (axi-cel) anti-CD19 CAR T-cell infusion products (including non-transduced T-cells) from 24 patients with LBCL [31]. As adverse events related to CAR T-cell therapy, 12 patients developed high-grade immune effector cell-associated neurotoxicity syndrome (ICANS, grade 3–4), whereas the other 12 patients developed low-grade ICANS (grade 0–2). The aim of this re-analysis was to investigate whether there is a significant difference in the expression of the temporal consensus gene signatures between low-grade and high-grade ICANS. Deng et al. observed only few significant differences in the gene expression of CAR^+ T-cells compared to non-transduced T-cells. Therefore, they did not separate CAR^+ and CAR^- T-cells in their analyses, which is also our strategy in the following.

Firstly, we analyzed whether metagenes were associated with cell clusters. Among all $CD8^+CD4^-$ and $CD8^-CD4^+$ T-cells, 3 metagenes (M5, M3, and M3/M5) with at least 10 genes were present in the scRNA-Seq data (Fig. 6A). We grouped the T-cells

Rade *et al. Genome Biology*   (2023) 24:287

Page 12 of 26
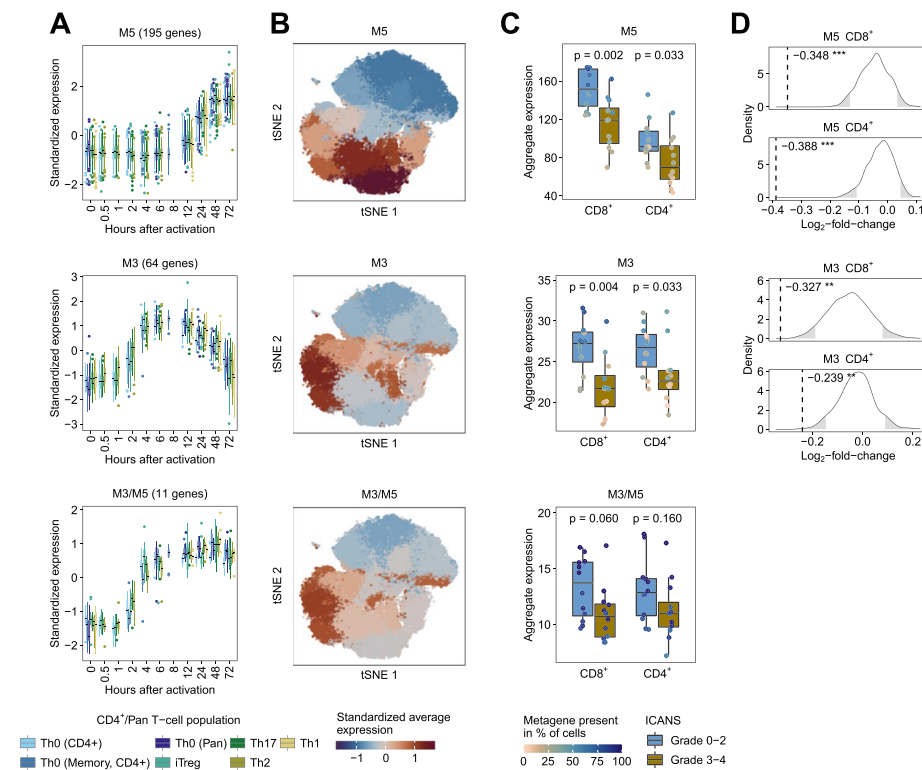
Fig. 6 Re-analysis of data scRNA-Seq data of autologous anti-CD19 CAR T-cell infusion products from 24 patients with LBCL. **A**–**C** Metagenes with at least 10 genes among the most highly variable genes in the scRNA-Seq dataset were analyzed. **A** The boxplots in the left panel show grouped consensus expression profiles over the course by genes associated with identity and shared metagenes. Each boxplot represents one T-cell population from the *Discovery and Verification Sets*. The y-axis depicts the standardized median expression of genes from samples with identical analysis time points. The number in brackets above the boxplots indicates the number of highly variable genes of the metagene present in the scRNA-Seq data. **B** We embedded $CD8^+CD4^-$ and $CD8^-CD4^+$ T-cells from 24 patients into a two-dimensional space by the t-distributed stochastic neighbor embedding (tSNE) method. The colors indicate the standardized average expression of the metagenes for each cluster (the same value is assigned to all cells in a cluster). **C** For each metagene and patient, we calculated aggregated expression (summed average expression) for $CD8^+CD4^-$ and $CD8^-CD4.^+$ cells. Differences in aggregated expression between patients with low- and high-grade ICANS were evaluated using the Wilcoxon rank-sum test. The colors of the dots in the boxplots indicate the percentage of cells for each patient in which the corresponding metagene is present. We considered a metagene as present in a cell if at least 25% of all associated genes had at least one UMI count. **D** For each metagene and T-cell population that was significant ($p < 0.05$) in the Wilcoxon rank-sum test, we generated a null distribution in order to confirm the results (see the "Methods" section). Dashed vertical lines indicate median log2 fold change of aggregated expression between low- and high-grade ICANS patients of the metagene. Rejection regions (empirical *p*-value < 0.05) are highlighted in grey (* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$)

using commonly deployed analytical methods and visualized resultant clusters into a two-dimensional space. Each cluster is colored by the standardized average expression of the metagenes (Fig. 6B). We performed the same procedure with known T-cell state markers (see the "Methods" section) and markers associated with T-cell molecular mechanisms [32] (Additional file 1: Fig. S31A).

We observed that metagene M5 is present in cell clusters associated with the cell cycle signatures G1/S and G2/M (Additional file 1: Fig. S31B). Metagenes M3 and M3/M5 were present in cell clusters linked to the TCA cycle, glycolysis, Treg, and

hypoxia/HIF signatures. All metagenes that were linked to cell clusters were associated with the CD8 lineage (Additional file 1: Fig. S31B).

For each metagene and patient, we calculated the aggregated expression of $CD8^+CD4^-$ and $CD8^-CD4^+$ cells. We evaluated differences in aggregated expression between patients with low- and high-grade ICANS using the Wilcoxon rank-sum test (Fig. 6C). We observed significant differences for metagene M5 ($CD8^+$ $p = 0.002$ and $CD4^+$ $p = 0.033$) and M3 ($CD8^+$ $p = 0.004$ and $CD4^+$ $p = 0.033$). To confirm this finding, we generated a null distribution by randomly drawing gene sets of the same size as the metagenes (see the "Methods" section) and found that metagenes performed significantly ($p < 0.05$) better than randomly generated gene sets (Fig. 6D). We also tested whether patient characteristics correlate with aggregate expression of the significant metagenes. Patient age correlated significantly (p-value $= 0.045$) with the aggregated expression of the gene set of metagene M3 in $CD8^+$ T-cells (Additional file 1: Fig. S33).

We conducted the same analysis as above for the adverse event, cytokine release syndrome (CRS), and also for patient outcome. However, when comparing aggregate expression between patients with low-grade and high-grade CRS or between complete response and partial response/progressive disease, we did not observe significant differences (Additional file 1: Fig. S34).

For T-cell state markers as well as markers for T-cell molecular mechanisms associated with the same cell clusters as the metagenes, only the cell cycle signatures G1/S and G2/M showed significant differences between low- and high-grade ICANS in $CD8^+$ cells (Additional file 1: Fig. S32A, Wilcoxon rank-sum test). However, only the G1/S signature performed significantly better than random sets (Additional file 1: Fig. S32B).

For markers not associated with the same cell clusters as the metagenes, the pro-inflammatory $CD8^+$ signature, the cytolytic effector pathway, and the glucose deprivation signature showed also significant differences in aggregated expression between patients with low- and high-grade ICANS (Additional file 1: Fig. S32A, Wilcoxon rank-sum test). All three were confirmed by comparison to randomly generated gene sets (Additional file 1: Fig. S32B).

Discussion

We analyzed kinetic changes in gene expression profiles of human T-cells isolated from PBMCs or cord blood after activation with anti-CD3/anti-CD28 only or after subsequent polarization with cytokines. To provide temporal details of the coordinated multiple functions before and after activation of T-cells, we used in vitro kinetic models profiled by RNA sequencing and microarray platforms. We used a *Discovery-* and two *Verification Sets* consisting of multiple T-cell populations to find DE genes after activation. In the *Discovery Set*, we identified a reasonably large number of genes with consistent signs of log2 fold change across different T-cell populations (Fig. 2B). This points to non-negligible overlaps in transcriptional responses upon activation of T-cells, no matter of polarization conditions. We obtained genes with a significant combined effect size across different $CD4^+$ T-cell populations by a meta-analysis. Based on these genes, we conducted NMF to aggregate the temporal expression profile of each T-cell population into metagenes, i.e., sets of temporally co-expressed genes.

NMF does not inherently preserve relative temporal ordering in pattern matrix, because metagenes are treated as nominal variables. Hence, we ordered samples in the pattern matrices according to time points in order to classify the expression pattern of each T-cell population into sustained repressed, intermediate, and late response metagenes.

To obtain a more robust time series, we combined metagenes with similar expression patterns from each CD4$^+$ T-cell population in the *Discovery Set* and confirmed the resulting consensus gene signatures by the *Verification Sets*. We are aware that due to the lack of analysis time points (0.5 to 6 h for Th1 and 0.5 to 4 h for Th0 in the *Pan T-cell Verification Set*), a limitation of our study is that genes of the Th1 and Th0 populations might be falsely assigned to metagenes.

Krüppel-like factor 2 (KLF2) is among the highest ranked genes associated with the sustained repressed metagene (M1) across all T-cell populations (Additional file 1: Fig. S29). KLF2 plays an important role in the regulation of the T-cell homeostasis by promoting naïve-T-cell quiescence [33]. Its expression decreases rapidly after TCR activation, and it has been observed that KLF2 protein degradation and subsequent loss of KLF2 mRNA occurs during T-cell stimulation [33, 34]. KLF2 and NF-κB are reciprocal antagonists [35, 36]. NF-κB is the highest ranked gene linked to the intermediate metagene M2 (Additional file 1: Fig. S29). The expression patterns of metagene M1 and M2 here illustrate the dual effect of KLF2 and NF-κB.

EGR2 (members of the early growth response family) ist among the prominent genes of the intermediate response metagene M2. This transcription factor is the target of NFAT transcription factors [37, 38], acts as a transcriptional repressor, and impairs IL2 transcription in anergy [39]. IL2 is the highest ranked gene for the intermediate metagene M3 (pattern peak at 6-12 h) in the *Discovery Set* and *Pan T-cell Verification Set*. However, in the *Memory T-cell Verification Set*, IL2 is associated with the intermediate metagene M2/M3 (pattern peak at 2-4 h). This classification might be of biological origin because memory T-cells were activated in the *Memory T-cell Verification Set*, which may lead to a faster response to the activation signal. Otherwise, we were not able to compare the expression peak of IL2 in the *Memory T-cell Verification Set* with the *Discovery Set* or *Pan T-cell Verification Set*, because the *Memory T-cell Verification Set* lacks the activation time point at 6 h. In addition to IL2, this observation was also made for TNFRSF9, an activation-induced co-stimulatory molecule of the tumor necrosis factor receptor superfamily that is upregulated on activated T cells and antigen-presenting cells, including dendritic cells, macrophages, and B cells [40–44] (for expression profile, see Additional file 2). Although it would be of interest to address the molecular mechanisms responsible for memory T-cells responding faster to stimulation than naïve T-cells, only two data sets analyzed explicitly mentioned that the isolated CD4$^+$ T-cells were naïve T-cells [13, 14]. Since the other expression data sets are Pan CD4$^+$ T-cells, a direct comparison of possible metagene time shifts between naive and memory T cells is not feasible.

The highest ranked gene for the shared metagene M3/M5 is interleukin-2 receptor (IL2RA), which plays a crucial role in immune homeostasis [45] (Additional file 1: Fig. S29). AURKA, CDC6, TOP2A, and DEPDC1 are the most prominent genes for the late response metagene M5, which are mainly involved in cell cycle processes [26–29].

Although filtering metagenes specific to a T-cell population provides added value, we did not include this analysis in our study. Since the Th1, Th2, and Th17 populations consist of only three biological replicates, we cannot guarantee sufficient robustness. Further kinetic studies of these populations with overlapping time points are needed to develop T-cell population-specific time series signatures.

For the *Pan T-cell Verification Set*, in addition to the activation kinetics, we also sequenced unactivated Pan T-cells after 6 to 72 h, which we used as negative controls. With this strategy, we were able to exclude genes with similar expression in negative control and activated Pan T-cells from the consensus gene signatures. It is worth noting that we observed an average of 10,760 DE genes across the contrasts for the negative controls (Fig. 5C). This indicates an unspecific regulation due to cultivation effects and emphasizes the importance of using matched negative controls in time series experiments.

Our study provides temporal consensus signatures of T-cell regulatory dynamics from healthy donors that could also be useful for defining T-cell states in disease, as proposed in Szabo et al. [46]. We demonstrated this by re-analysis of a scRNA-Seq study by Deng et al. [31] in which autologous axi-cel anti-CD19 CAR T-cell infusion products from 24 patients with LBCL were assessed by single-cell RNA sequencing. Because CAR T cells are also cultured and expanded with anti-CD3/anti-CD28 beads during manufacturing [47], we analyzed the activation states of the infusion product after manufacturing before the CAR T cells were infused into patients. Metagenes M3 and M5, which are associated with metabolic processes and proliferation, respectively, were significantly enriched in infusion products of patients with low-grade ICANS compared to patients with high-grade ICANS. These observations are novel with respect to ICANS and counterintuitive at first sight. Anti-CD19 CAR T-cell proliferation, activation, and maximal expansion are positively correlated with ICANS grade in vivo [48]. However, in the analyzed dataset from Deng et al., we assessed infusion products instead of in vivo activated CAR T-cells following infusion. High-grade ICANS occurred in the median 4 days after infusion, underlining that activation, proliferation and maximal expansion of CAR T-cells most likely occurred after infusion and subsequent activation of CAR T-cells through antigen exposure on target malignant B-cells. Therefore, our findings propose that infusion of already activated and proliferative CAR T-cells is not linked to occurrence of high-grade ICANS. It is crucial to validate these findings by further analyses including longitudinally isolated CAR T cells from patients' peripheral blood. Notably, the metagenes were trained using samples derived from healthy donors, and not from heavily pretreated LBCL patients [31], which is important to consider. We did not find a significant correlation between metagenes and CRS. However, this could be due to the fact that the high-grade CRS group (grade 3–4) included 4 patients, while the low-grade group (grade 1–2) consisted of 18 patients. Deng et al. showed that CAR T-cell products which had a significant enrichment of cells with a CD8 T-cell exhaustion lead to a partial response or progressive disease, whereas cell products enriched with the memory phenotype were correlated to a complete remission of LBCL. This phenomenon was also confirmed in chronic lymphoblastic leukemia patients treated with anti-CD19 CAR-T cells [49]. However, we did not find a significant correlation between metagenes and patient outcomes. Overall, our activation signature could be helpful for evaluating CAR T-cell products prior or after the manufacturing process.

Of note, expression peaks of the signatures are interpreted relative to the unactivated condition since we did not filter genes according to low expression prior to activation.

We are aware that the consensus signatures developed are based on an artificial immune response where T-cells were activated with anti-CD3/anti-CD28 coated beads. This experimental design only partially reflects the human immune response in vivo. Therefore, the expression pattern of the genes from our signatures might behave differently under in vivo conditions. For this reason, we are investigating the temporal resolution of the recall immune response to antigens in an ongoing study.

Further, with the increasing number of available temporal transcriptome-wide studies, we expect to transfer our concept of defining and validating temporal biomarker signatures to differentiation processes of distinct T-cell populations. This was not the aim of the current study due to a small number of replicates (see Additional file 1: Fig. S1). Low sample sizes limit the separation of data sets into disjunct training and verification data sets, a substantial requirement for the development of biomarker gene signatures.

Conclusion

In this study, we identified and verified general biomarker signatures robustly evaluating T-cell activation in a time-resolved manner. It describes general temporal changes in gene expression patterns upon T-cell activation no matter of the polarization condition. It serves as a resource for studies of human T-cell immunity in disease or immunotherapies to classify T-cell states and may be used to define optimal T-cell biomarkers in dependence of experimental analysis time points. For easy access and interpretation of the consensus gene signatures, we provide an interactive HTML document (Additional file 2). In addition, the developed method for analyzing time-series transcriptome data can be applied to other cell types.

Methods

Data sources used for the development of the consensus gene signature

We downloaded the publicly available RNA-Seq and microarray data sets from the Sequence Read Archive (SRA) [50] and NBCI's Gene Expression Omnibus (GEO) [51], respectively. The meta-analysis consisted of one *Discovery* and 2 *Verification Sets* (Fig. 1A). For the *Discovery Set*, we used the following sources: Raw sequencing data in FASTQ format from the RNA-Seq projects (GSE52260, GSE90569, GSE94396, GSE96538) [12–14] were obtained using the prefetch and fastq-dump commands in the SRA Toolkit v2.9.2 [52]. Raw CEL files from Affymetrix Human Genome U133 Plus 2.0 Arrays for GSE17974 [10] and GSE32959 [11] were obtained using the R package GEO-query v2.58.0 [53]. Since some of the analyzed datasets examined gene expression of CD4$^+$ T-cells at multiple time points of activation, we decided that one time point must be available for at least 3 CD4$^+$ T-cell populations. With the exception of CD4$^+$ cells cultured under Th1 cell polarization condition [11] overall 9 different time points were analyzed for each population. Activated (Th0) CD4$^+$ T-cells comprised 4, iTreg 3, and the other populations one dataset each. A detailed description of the experimental conditions and RNA-Seq or microarray technical specifications for each dataset is shown in Additional file 1: Table S1 and Fig. S1.

We downloaded the RNA-Seq gene counts from the study Gutierrez-Arcelus et al. (GSE140244) [15] and used them as a *Memory T-cell Verification Set*.

For the *Pan T-cell Verification Set* (GSE197067), we activated Pan T-cells from 4 healthy individuals with anti-CD3/CD28 beads (for MACS isolation strategy of Pan T-cells, see Additional file 1). Briefly, 2×10^5 Pan T-cells were seeded in a 96 U-well plate with quadruplicates per condition. 2×10^5 Dynabeads human T-Activator CD3/CD28 (Gibco, 11132D) per sample were washed with PIB and were then resuspended in T cell medium (CG-DC-medium supplemented with 5% human AB serum and 1% penicillin/streptomycin). 2×10^5 CD3/CD28 beads were added to activate T-cells and wells were filled up to 200 μL total volume. Unstimulated T-cells were used as controls. T-cells were harvested at different time points (0 h, 6 h, 12 h, 24 h, 48 h, 72 h). Therefore, unactivated and activated T-cells were collected and centrifuged (10 min, 300 xg, room temperature). Seven hundred microliters QIAzol were added and samples were stored at $-80\,°C$ till RNA isolation. We performed RNA-seq before activation (0 h) and 6, 12, 24, 48, and 72 h after activation. In addition, as a control, we sequenced Pan T-cells for the same time series without activation conditions. For experimental details, see Additional file 1.

To ensure a uniform notation, we refer throughout this article to T-cells during activation condition only (Th0) or polarization condition leading to different T-cell fates (Th1, Th2, Th17, and iTreg) as T-cell populations. It should be noted that we did not include CD8 T-cells in our analysis because we could not find suitable time series data for this T-cell lineage in our literature search.

Pre-processing and normalization

To facilitate the multi-step analysis of the RNA sequencing datasets, we applied the workflow-manager uap v0.0.1 [54]. A detailed description of all processing steps from FASTQ files to gene quantification can be found in Additional file 1. The according configuration files for uap are available in Additional file 3. For pre-processing of microarray datasets, we used the R package affy v1.68.0. Gene expression of microarray and RNA-Seq datasets was quantified using the human reference gene annotation GENCODE v29. Expression data of the same T-cell population and platform were normalized in one junk (see Additional file 1 for more details).

To visualize the gene expression in Fig. 4C, we performed cross-platform normalization using feature-specific quantile normalization (FSQN) [55]. For this purpose, we used the batch-corrected RNA-seq data (see Additional file 1) as a reference to normalize the expression arrays.

Differential gene expression analysis

We performed a differential gene expression analysis (DGEA) for each contrast (activated T-cells at a given time point compared to unactivated T-cells) using the empirical Bayes moderated t-test implemented in the package limma v3.46.0 [56, 57]. For each contrast, p-values were adjusted using the method of Benjamini-Hochberg (or false discovery rate (FDR)) [58]. A gene was considered as significantly differentially expressed (DE) if the FDR-adjusted p-value was < 0.05. DE genes were ranked using the Topconfects v1.6.0 R Bioconductor package [59] (see Additional file 1 for a more detailed method description of the DGEA). The same procedure was carried out for the negative

controls from the *Pan T-cell Verification Set* with the following contrasts: Firstly, unactivated Pan T-cell at 6 to 72 h compared to unactivated Pan T-cells at 0 h. Secondly, activated Pan T-cells at a given time point after activation compared to unactivated Pan T-cells at the same time point. A volcano plot representation of the DGEA is shown in Additional file 1: Fig. S6.

Meta-analyses

For each DE gene in the *Discovery Set*, we calculated standardized effect sizes as Hedges' g values [60–62]. Briefly, this involves log2 fold changes of the DE genes for activated compared to unactivated $CD4^+$ T-cell populations at each analysis time point divided by the standard deviation as estimated with the empirical Bayes method in limma, followed by an adjustment for small sample sizes (see Additional file 1 for more details).

We estimated a combined effect size for each DE gene across the $CD4^+$ T-cell populations and time points in the *Discovery Set* using a random effects model. We used a random effects model because we did not assume that there is one true effect size, which is shared by all the included datasets, but rather a range of true effect sizes with additional sources or variation, such as different platforms (RNA-Seq and Microarray). The model was fitted with the restricted maximum-likelihood estimator in the R package metafor v2.4.0 [63]. We used the Topconfects v1.6.0 R Bioconductor package [59] to calculate for each estimated combined effect size and standard error a "confect" value. This "confect" value or confident effect size represents a confident inner bound of the calculated combined effect size by metafor while maintaining a given FDR of < 0.05. In this way, we ranked each gene according to its "confect." This is a more conservative way than ranking by raw effect size because it avoids ranking genes by the highest effect size, which can have a large within-group variability (see Additional file 1: Fig. S9A). We declared that at a given time point of activation, a DE gene had a significant combined effect size in at least 2 $CD4^+$ T-cell populations if the FDR-adjusted *p*-value was < 0.05. To assess the amount of heterogeneity post hoc the I^2 statistic was calculated using the metafor package (see Additional file 1: Fig. S9B).

Unsupervised decomposition

To infer biological patterns over time, non-negative matrix factorization (NMF) was performed. We performed NMF for each T-cell population (including unactivated T-cells) from the *Discovery Set* and *Verification Sets* separately. For the *Discovery Set*, the following set of genes were defined as input to NMF: All DE genes with a significant combined effect size in at least one contrast (e.g., 12 h vs. 0 h) across 4 (0.5 to 6 h) and 5 (12 to 72 h) T-cell populations. For the activation kinetics of the *Verification Sets*, we factorized the set of genes defined for the *Discovery Set.* However, we only used genes from this set that were also significantly differentially expressed in the Verification Sets. To preserve the linearity of NMF modeling, quantile normalized CPM values from RNA-Seq and normalized intensities from gene expression arrays were used in non-log space [37, 64]. We used the R-package NMF v0.23.0 [38] with the algorithm from Brunet et al. [65] based on the Kullback–Leibler divergence as objective with multiplicative updates rules to solve the following approximation:

$$X_{\mathbb{R}^{n \times m}} \approx W_{\mathbb{R}^{n \times k}} \times H_{\mathbb{R}^{k \times m}} \text{ such that } W, H \geq 0$$

The aim is to factorize a non-negative matrix X into two lower-rank matrices with strictly non-negative elements where *X* is the expression matrix whose rows contain the expression levels of the *n* genes in the *m* samples. Matrix *W (*referred to as gene signature matrix*)* has size *n * k*, where *n > k*. Each of the *k* columns defined a latent factor and was used to describe the original data in a sub-dimensional space. In the context of gene expression data, these latent factors were referred to as metagenes that reflected genes with similar expression patterns. Each entry w_{ij} contains the contribution for gene *i* to metagene *j*. Matrix *H* represented the pattern matrix and has size *k * m*, where the entry h_{ij} was the weight of metagene *i* in sample *j*.

The initialization of *H* and *W* was generated by random seeding, where the entries of each factor are drawn from a uniform distribution within the same range as the entries in the matrix *X*.

To determine the optimal factorization rank *k* (number of metagenes) for each T-cell population, we performed NMF from actual and randomized data by repeating the rank value in the interval 2–10. For each rank, 200 iterations were performed. We used consensus clustering as a qualitative measurement, cophenetic correlation coefficients, and average silhouette scores as a quantitative measure to assess the stability of the clusters [65, 66] (see Additional file 1 for a more detailed method description). As proposed in Brunet et al., we selected the factorization rank where the magnitude of the cophenetic correlation coefficient begins to fall. In addition, we used the average silhouette width to choose the optimal rank (see Additional file 1: Fig. S16). After determining the optimal rank, the pattern and gene signature matrix was obtained from the factorization that achieved the lowest approximation error for the selected rank in 200 runs.

Temporal categorization and visualization of metagenes

We used the entries in matrix *H*, representing the weights of metagene *i* in sample *j* for temporal annotation and visualization of metagenes (Fig. 3, top panel). Each column in the matrix *H* was scaled to sum up to one. For each metagene *i*, we calculated median weights for all samples from identical analysis time points. We annotated metagene *i* as sustained repressed, intermediate, or late response metagene based on its maximum median weight across all analysis time points.

Defining metagene-associated genes

We used the gene signature matrix to assign DE genes to annotated metagenes. Each row in matrix *W* was min–max normalized such that the values, which contain the contribution of gene *i* to metagene *j*, are in the interval [0,1]. The resulted bimodal distribution for each T-cell population is shown in Additional file 1: Fig. S15. Gene *i* was associated with metagene *j* when the normalized weight w_{ij} was > 0.5. In this way, we were able to identify genes that are specific to one metagene ($w_{ij} == 1$) and are referred to as identity metagenes or genes with higher weights (> 0.5) in more than one metagene (referred to as shared metagenes). We did not filter out genes post hoc because we factorized only informative genes, that is, genes with a significant combined effect size.

Consensus gene expression profiles from the *Discovery Set* and verification

For the consensus gene signatures from the *Discovery Set*, the annotated metagenes had to be temporally consistent across all CD4$^+$ T-cell populations. This means, for example, that all genes associated with the late-response metagene in one T-cell population must also be associated with the late-response metagene in all other T-cell populations. We used 2 *Verification Sets* to verify the consensus signatures (Fig. 4C) from the *Discovery Set* using the same procedure. Since only the analysis time points 12 to 72 h for the Th1 population (*Discovery Set*) and 6 to 72 h for the Th0 population (*Pan T-cell Verification Set*) are available, we could not make any conclusion about the time course of expression with respect to intermediate metagene 2 (expression peak after 2 h). Therefore, for the two populations mentioned, we excluded metagene 2 from this analysis.

For the *Pan T-cell Verification Set*, we used the time series negative controls (unactivated Pan T-cells after 6 to 72 h) to compare their gene expression profiles with the kinetics of the activated Pan T-cells. For each metagene from the *Discovery Set*, its centroid was calculated, defined as the median expression value of all genes for samples from identical analysis time points associated with the metagene. The calculation was performed with FSQN normalized data. For each centroid, we calculated distances between time points of activation (0.5 to 72 h) compared to unactivated (0 h) T-cells. Provided that the analysis time point is present in the *Pan T-cell Verification Set*, we selected the time point with the maximum absolute distance for filtering genes. The aim was to exclude genes with similar expression in negative control and activated Pan T-cells. We analyzed only genes that are part of the consensus gene signatures and with significant expression changes in negative controls (i.e., with FDR < 0.05 in at least one contrast when comparing unactivated Pan T-cells at 6 to 72 h vs. 0 h). At the selected time point, i.e., the time point with maximum absolute distance to the centroid of the metagene, we retained genes with absolute "confect" value > 1 (FDR < 0.05). "Confect" values and their significance between activated and negative controls at this given time point were estimated in the DGEA.

Metagene landscape

To visualize the relation of sample and gene to metagenes in a two-dimensional space, we followed the methodology described previously [67, 68]: We first concatenated the pattern matrices based on temporally coherent metagenes across all CD4$^+$ T-cell populations from the *Discovery Set* (the combined pattern matrix in shown in Additional file 1: Fig. S20). We calculated a pairwise distance matrix between the metagenes (rows of combined pattern matrix) using Euclidean distance. Then, we projected the distance matrix into two dimensions using the Sammon mapping method from the MASS R library [69]. The *x* and *y* coordinates for each metagene obtained by dimension reduction were standardized. Based on the combined pattern matrix *H*, we computed the x and y coordinates for each sample in the two-dimensional space according to:

$$S_{ix} = \frac{\sum_i H_{ij} M_{ix}}{\sum_i H_{ij}} S_{iy} = \frac{\sum_i H_{ij} M_{iy}}{\sum_i H_{ij}}$$

where M_{ix}, M_{iy} represents the x and y coordinates for metagene i. For gene embedding, we used the gene signature matrix of each CD4$^+$ T-cell population. Embedding of genes was performed with the same methodology as for sample embedding. The resulting metagene landscape is shown in Fig. 4A.

Pathway analysis

We performed enrichment analysis for DE genes with a significant combined effect size and for metagene-associated genes using the R Bioconductor package clusterProfiler v3.18.1 [70] on the Gene ontology (GO) database [71] for the ontology biological process and the Reactome database [72]. To compare the enrichment of pathways/GO terms between different contrasts and metagenes, we used the compareCluster() function of the clusterProfiler package. For GO analysis, we used the simplify() function of the clusterProfiler package with default parameters to remove redundant terms. The significance of enrichment was assessed by a hypergeometric test and adjusted p-values for multiple testing were calculated based on the Benjamini–Hochberg method. All pathways/GO terms with FDR < 0.05 are considered significantly over-represented. To colorize enriched pathways with the uppermost hierarchical level of Reactome database, we used the hierarchical pathway relationship file (ReactomePathwaysRelation.txt, available on www.reactome.org).

Re-analysis of scRNA-Seq data

We downloaded the raw read counts of the scRNA-Seq study by Deng et al. [31] under the GEO accession number GSE151511. As proposed in Deng et al. only cells with less than 7000 genes and less than 15% of reads mapped to mitochondrial genes were retained. Raw counts were normalized and transformed to log-space using the NormalizeData() function implemented in the R package Seurat v4.0.3 [73] with default settings. Only cells with expression of CD3D or CD3E or CD3G > 1 normalized count value were used for subsequent analysis. In addition, we filtered for CD8$^-$CD4$^+$ (15,936 cells) and CD8$^+$CD4$^-$ (59,002 cells) CAR T-cells based on the gene expression data using the following strategy: Given the normalized expression, one cell was considered as CD8 positive or negative if the normalized count value of CD8A or CD8B was > 1 or ≤ 1, respectively. One cell was considered as CD4 positive or negative if the normalized count value of CD4 was > 1 or ≤ 1, respectively.

We detected highly variable genes using the "vst" method of the FindVariableFeatures() function in Seurat at default settings, resulting in 3000 highly variable genes. Principal component analysis (PCA) was conducted using standardized and normalized highly variable genes with the function RunPCA implemented in Seurat. We used 39 principal components, which explained 95% of the variance, to integrate the dataset from different samples using RunHarmony() from the Harmony v1.0 R package [74]. Using the Harmony-corrected cell embeddings, we computed a shared nearest neighbors graph, as implemented in FindNeighbors() function in Seurat with default settings. Clustering was performed using the FindClusters() function in Seurat with the default setting. The Harmony-corrected cell embeddings were projected into a two-dimensional space using the t-distributed stochastic neighbor embedding (tSNE) method. For this, we used 40 Harmony-corrected components, which explained 95% of the variance.

Known T-cell state marker genes [75, 76] including 5 exhausted genes (CTLA4, HAVCR2, LAG3, PDCD1, TIGIT), 12 cytotoxic genes (PRF1, IFNG, GNLY, NKG7, GZMB, GZMA, GZMH, KLRK1, KLRB1, KLRD1, CTSW, CST7) and 5 naïve/memory genes (CCR7, TCF7, LEF1, SELL, CD44) were used to calculate the standardized average expression in each cluster. We also used gene signatures from Azizi et al. [32] to associate cell clusters with a functional context.

For each metagene, T-cell state marker, and gene signatures (referred to as gene sets), we conducted a permutation test. We generated a null distribution for each gene set by performing 1000 permutations. For each permutation, a random gene set was formed. Random gene sets are not part of genes associated with the original gene sets and have the same number of genes as the corresponding original gene set (based on 3000 present highly variable genes). For each random gene set and patient, we calculated the aggregate expression (summed average expression) of $CD8^+CD4^-$ and $CD8^-CD4^+$ cells. We then computed the log2 fold change in the median of aggregate expression between low- and high-grade ICANS patient groups. Based on the log2 fold change between low- and high-grade ICANS from the original gene set, we calculated the statistical significance using a left-tailed test for positive log2 fold changes or a right-tailed test for negative log2 fold changes.

Supplementary Information

The online version contains supplementary material available at https://doi.org/10.1186/s13059-023-03120-7.

Additional file 1. Supplementary methods and supplementary figures [87–105].

Additional file 2. Overview of all genes of the consensus signatures and their temporal expression as an interactive HTML document. In addition, the signaling pathways of the enrichment analyses for the consensus signatures are shown.

Additional file 3. Configuration files for the workflow-manager uap. These contain a detailed description of all processing steps from FASTQ files to gene quantification.

Additional file 4. Peer review history.

Availability of data and materials
The RNA-Seq and microarray datasets analyzed in this study are available in the NCBI GEO database (https://www.ncbi.nlm.nih.gov/geo/) with accession numbers GSE52260 [12, 77], GSE90569 [14, 78], GSE94396 [13, 79], GSE96538 [13, 80], GSE17974 [10, 81], GSE32959 [11, 82], and GSE140244 [15, 83]. Raw and normalized gene expression counts for the *Pan T-cell Verification Set* are available from GSE197067 [84]. In Additional file 2, we summarize all genes, including interactive

Rade *et al. Genome Biology*   (2023) 24:287

Page 23 of 26

visualization of expression profiles of individual genes of the consensus signatures. Processing and analysis code related to this study is deposited in our GitHub repository: https://github.com/fraunhofer-izi/Rade_et_al_tcell_2023 [85] and https://zenodo.org/records/10145756 [86].

Declarations

Ethics approval and consent to participate
Experimental methods using human cells (PBMCs) were approved by the ethics committee of the Hannover Medical School and are in compliance with the Helsinki Declaration of the world medical association. All of the donors gave written informed consent to the Blutspendedienst that their blood donation could also be used for scientific investigations.

Consent for publication
Not applicable.

Competing interests
The authors declare that they have no competing interests.

References

1. Smith-Garvin JE, Koretzky GA, Jordan MS. T cell activation. Annu Rev Immunol. 2009;27:591–619.
2. Malissen B, Grégoire C, Malissen M, Roncagalli R. Integrative biology of T cell activation. Nat Immunol. 2014;15:790–7.
3. Chapman NM, Boothby MR, Chi H. Metabolic coordination of T cell quiescence and activation. Nat Rev Immunol. 2020;20:55–70.
4. Liu JO. The yins of T cell activation. Sci STKE. 2005;2005:re1.
5. Sprent J, Surh CD. Normal T cell homeostasis: the conversion of naive cells into memory-phenotype cells. Nat Immunol. 2011;12:478–84.
6. Chapman NM, Chi H. Hallmarks of T-cell Exit from Quiescence. Cancer Immunol Res. 2018;6:502–8.
7. Piccirillo CA, Bjur E, Topisirovic I, Sonenberg N, Larsson O. Translational control of immune responses: from transcripts to translatomes. Nat Immunol. 2014;15:503–11.
8. Ricciardi S, Manfrini N, Alfieri R, Calamita P, Crosti MC, Gallo S, et al. The Translational Machinery of Human CD4+ T Cells Is Poised for Activation and Controls the Switch from Quiescence to Metabolic Remodeling. Cell Metab. 2018;28:895-906.e5.
9. Hess K, Yang Y, Golech S, Sharov A, Becker KG, Weng N-P. Kinetic assessment of general gene expression changes during human naive CD4+ T cell activation. Int Immunol. 2004;16:1711–21.
10. Elo LL, Järvenpää H, Tuomela S, Raghav S, Ahlfors H, Laurila K, et al. Genome-wide Profiling of Interleukin-4 and STAT6 Transcription Factor Regulation of Human Th2 Cell Programming. Immunity. 2010;32:852–62.
11. Aijö T, Edelman SM, Lönnberg T, Larjo A, Kallionpää H, Tuomela S, et al. An integrative computational systems biology approach identifies differentially regulated dynamic transcriptome signatures which drive the initiation of human T helper cell differentiation. BMC Genomics. 2012;13:572.
12. Tuomela S, Rautio S, Ahlfors H, Öling V, Salo V, Ullah U, et al. Comparative analysis of human and mouse transcriptomes of Th17 cell priming. Oncotarget. 2016;7:13416–28.
13. Schmidt A, Marabita F, Kiani NA, Gross CC, Johansson HJ, Éliás S, et al. Time-resolved transcriptome and proteome landscape of human regulatory T cell (Treg) differentiation reveals novel regulators of FOXP3. BMC Biol. 2018;16:47.
14. Ubaid Ullah N, Andrabi SBA, Tripathi SK, Dirasantha O, Kanduri K, Rautio S, et al. Transcriptional Repressor HIC1 Contributes to Suppressive Function of Human Induced Regulatory T Cells. Cell Rep. 2018;22:2094–106.
15. Gutierrez-Arcelus M, Baglaenko Y, Arora J, Hannes S, Luo Y, Amariuta T, et al. Allele-specific expression changes dynamically during T cell activation in HLA and other autoimmune loci. Nat Genet. 2020;52:247–53.
16. Soon MSF, Lee HJ, Engel JA, Straube J, Thomas BS, Pernold CPS, et al. Transcriptome dynamics of CD4+ T cells during malaria maps gradual transit from effector to memory. Nat Immunol. 2020;21:1597–610.
17. Richard AC, Lun ATL, Lau WWY, Göttgens B, Marioni JC, Griffiths GM. T cell cytolytic capacity is independent of initial stimulation strength. Nat Immunol. 2018;19:849–58.
18. Lönnberg T, Svensson V, James KR, Fernandez-Ruiz D, Sebina I, Montandon R, et al. Single-cell RNA-seq and computational analysis using temporal mixture modelling resolves Th1/Tfh fate bifurcation in malaria. Sci Immunol. 2017;2:eaal2192.
19. Soskic B, Cano-Gamez E, Smyth DJ, Ambridge K, Ke Z, Matte JC, et al. Immune disease risk variants regulate gene expression dynamics during CD4+ T cell activation. Nat Genet. 2022;54:817–26. https://doi.org/10.1038/s41588-022-01066-3.
20. Brodin P, Davis MM. Human immune system variation. Nat Rev Immunol. 2017;17:21–9. https://doi.org/10.1038/nri.2016.125.
21. Upadhaya S, Yu JX, Shah M, Correa D, Partridge T, Campbell J. The clinical pipeline for cancer cell therapies. Nat Rev Drug Discov. 2021;20:503–4. https://doi.org/10.1038/d41573-021-00100-z.
22. Rosenberg AS, Sauna ZE. Immunogenicity assessment during the development of protein therapeutics. J Pharm Pharmacol. 2018;70:584–94. https://doi.org/10.1111/jphp.12810.

23. Duke BR, Mitra-Kaushik S. Current In Vitro Assays for Prediction of T Cell Mediated Immunogenicity of Biotherapeutics and Manufacturing Impurities. J Pharm Innov. 2020;15:202–18. https://doi.org/10.1007/s12247-019-09412-5.

24. Biomarkers Definitions Working Group. Biomarkers and surrogate endpoints: preferred definitions and conceptual framework. Clin Pharmacol Ther. 2001;69:89–95.

25. Seliger B, Reiche K, Massa C, Fricke S, Schmiedeknecht G, Horn F, et al. Necessity for next-generation quality assessment of CAR T cell manufacturing and advanced therapy guidance. Cell Gene Ther Insights. 2020;1:163–8.

26. Willems E, Dedobbeleer M, Digregorio M, Lombard A, Lumapat PN, Rogister B. The functional diversity of Aurora kinases: a comprehensive review. Cell Div. 2018;13:7.

27. Borlado LR, Méndez J. CDC6: from DNA replication to cell cycle checkpoints and oncogenesis. Carcinogenesis. 2008;29:237–43.

28. Marchesi S, Montani F, Deflorian G, D'Antuono R, Cuomo A, Bologna S, et al. DEPDC1B coordinates de-adhesion events and cell-cycle progression at mitosis. Dev Cell. 2014;31:420–33.

29. Mi Y, Zhang C, Bu Y, Zhang Y, He L, Li H, et al. DEPDC1 is a novel cell cycle related gene that regulates mitotic progression. BMB Rep. 2015;48:413–8.

30. Hounkpe BW, Chenou F, Lima F de, Paula EV de. HRT Atlas v1.0 database: redefining human and mouse housekeeping genes and candidate reference transcripts by mining massive RNA-seq datasets. Nucleic Acids Res. 2021;49:D947-D955.

31. Deng Q, Han G, Puebla-Osorio N, Ma MCJ, Strati P, Chasen B, et al. Characteristics of anti-CD19 CAR T cell infusion products associated with efficacy and toxicity in patients with large B cell lymphomas. Nat Med. 2020;26:1878–87.

32. Azizi E, Carr AJ, Plitas G, Cornish AE, Konopacki C, Prabhakaran S, et al. Single-Cell Map of Diverse Immune Phenotypes in the Breast Tumor Microenvironment. Cell. 2018;174:1293-1308.e36.

33. Kuo CT, Veselits ML, Leiden JM. LKLF: A transcriptional regulator of single-positive T cell quiescence and survival. Science. 1997;277:1986–90.

34. Buckley AF, Kuo CT, Leiden JM. Transcription factor LKLF is sufficient to program T cell quiescence via a c-Myc-dependent pathway. Nat Immunol. 2001;2:698–704.

35. Turpaev KT. Transcription Factor KLF2 and Its Role in the Regulation of Inflammatory Processes. Biochemistry. 2020;85:54–67.

36. Jha P, Das H. KLF2 in Regulation of NF-\kappaB-Mediated Immune Cell Function and Inflammation. Int J Mol Sci. 2017;18:2383.

37. Avila Cobos F, Alquicira-Hernandez J, Powell JE, Mestdagh P, de Preter K. Benchmarking of cell type deconvolution pipelines for transcriptomics data. Nat Commun. 2020;11:5650.

38. Gaujoux R, Seoighe C. A flexible R package for nonnegative matrix factorization. BMC Bioinformatics. 2010;11:367.

39. Safford M, Collins S, Lutz MA, Allen A, Huang C-T, Kowalski J, et al. Egr-2 and Egr-3 are negative regulators of T cell activation. Nat Immunol. 2005;6:472–80.

40. Sharma A, Campbell M, Yee C, Goswami S, Sharma P. 77 - Immunotherapy of Cancer. In: Rich RR, Fleisher TA, Shearer WT, Schroeder HW, Frew AJ, Weyand CM, editors. Clinical Immunology. 5th ed. London: Elsevier; 2019. p. 1033- 1048.e1.

41. Alderson MR, Smith CA, Tough TW, Davis-Smith T, Armitage RJ, Falk B, et al. Molecular and biological characterization of human 4–1BB and its ligand. Eur J Immunol. 1994;24:2219–27.

42. Chester C, Sanmamed MF, Wang J, Melero I. Immunotherapy targeting 4–1BB: mechanistic rationale, clinical results, and future strategies. Blood. 2018;131:49–57.

43. Futagawa T, Akiba H, Kodama T, Takeda K, Hosoda Y, Yagita H, Okumura K. Expression and function of 4–1BB and 4–1BB ligand on murine dendritic cells. Int Immunol. 2002;14:275–86.

44. Pollok KE, Kim YJ, Hurtado J, Zhou Z, Kim KK, Kwon BS. 4–1BB T-cell antigen binds to mature B cells and macrophages, and costimulates anti-mu-primed splenic B cells. Eur J Immunol. 1994;24:367–74.

45. Malek TR, Castro I. Interleukin-2 receptor signaling: at the interface between tolerance and immunity. Immunity. 2010;33:153–65. https://doi.org/10.1016/j.immuni.2010.08.004.

46. Szabo PA, Levitin HM, Miron M, Snyder ME, Senda T, Yuan J, et al. Single-cell transcriptomics of human T cells reveals tissue and activation signatures in health and disease. Nat Commun. 2019;10:4706.

47. Xu J, Melenhorst JJ, Fraietta JA. Toward precision manufacturing of immunogene T-cell therapies. Cytotherapy. 2018;20:623–38. https://doi.org/10.1016/j.jcyt.2017.12.007.

48. Morris EC, Neelapu SS, Giavridis T, Sadelain M. Cytokine release syndrome and associated neurotoxicity in cancer immunotherapy. Nat Rev Immunol. 2022;22:85–96. https://doi.org/10.1038/s41577-021-00547-6.

49. Fraietta JA, Lacey SF, Orlando EJ, Pruteanu-Malinici I, Gohil M, Lundh S, et al. Determinants of response and resistance to CD19 chimeric antigen receptor (CAR) T cell therapy of chronic lymphocytic leukemia. Nat Med. 2018;24:563–71. https://doi.org/10.1038/s41591-018-0010-1.

50. Leinonen R, Sugawara H, Shumway M. The Sequence Read Archive. Nucleic Acids Res. 2011;39:D19–21. Accessed Mar 2019.

51. Edgar R, Domrachev M, Lash AE. Gene Expression Omnibus: NCBI gene expression and hybridization array data repository. Nucleic Acids Res. 2002;30:207–10.

52. NCBI SRA Toolkit. http://ncbi.github.io/sra-tools/.

53. Davis S, Meltzer PS. GEOquery: a bridge between the Gene Expression Omnibus (GEO) and BioConductor. Bioinformatics. 2007;23:1846–7.

54. Kämpf C, Specht M, Scholz A, Puppel S-H, Doose G, Reiche K, et al. uap: reproducible and robust HTS data analysis. BMC Bioinformatics. 2019;20.

55. Franks JM, Cai G, Whitfield ML. Feature specific quantile normalization enables cross-platform classification of molecular subtypes using gene expression data. Bioinformatics. 2018;34:1868–74.

56. Ritchie ME, Phipson B, Di Wu, Hu Y, Law CW, Shi W, Smyth GK. limma powers differential expression analyses for RNA-sequencing and microarray studies. Nucleic Acids Res. 2015;43:e47.

57. Smyth GK. Linear models and empirical bayes methods for assessing differential expression in microarray experiments. Stat. Appl. Genet. Mol. Biol. 2004;3:Article3.

58. Benjamini Y, Hochberg Y. Controlling the false discovery rate: a practical and powerful approach to multiple testing. J. R. Stat. Soc. Series B Stat. Methodol. 1995;57:289–300.

59. Harrison PF, Pattison AD, Powell DR, Beilharz TH. Topconfects: a package for confident effect sizes in differential expression analysis provides a more biologically useful ranked gene list. Genome Biol. 2019;20:67.

60. Hedges LV. Fitting Categorical Models to Effect Sizes from a Series of Experiments. J Educ Behav Stat. 1982;7:119–37.

61. Ellis PD. The Essential Guide to Effect Sizes: Statistical Power, Meta-Analysis, and the Interpretation of Research Results. Cambridge: Cambridge University Press; 2010.

62. Cohn LD, Becker BJ. How meta-analysis increases statistical power. Psychol Methods. 2003;8:243–53.

63. Viechtbauer W. Conducting meta-analyses in R with the metafor package. J Stat Softw. 2010;36:1–48.

64. Zhong Y, Liu Z. Gene expression deconvolution in linear space. Nat. Methods. 2011;9:8–9; author reply 9.

65. Brunet J-P, Tamayo P, Golub TR, Mesirov JP. Metagenes and molecular pattern discovery using matrix factorization. Proc Natl Acad Sci U S A. 2004;101:4164–9.

66. Monti S, Tamayo P, Mesirov J, Golub T. Consensus Clustering: A Resampling-Based Method for Class Discovery and Visualization of Gene Expression Microarray Data. Mach Learn. 2003;52:91–118.

67. Wu Y, Tamayo P, Zhang K. Visualizing and Interpreting Single-Cell Gene Expression Datasets with Similarity Weighted Nonnegative Embedding. Cell Syst. 2018;7:656-666.e4.

68. Kim JW, Abudayyeh OO, Yeerna H, Yeang C-H, Stewart M, Jenkins RW, et al. Decomposing Oncogenic Transcriptional Signatures to Generate Maps of Divergent Cellular States. Cell Syst. 2017;5:105-118.e9.

69. Venables WN, Ripley BD. Modern Applied Statistics with S. New York: Springer; 2002.

70. Yu G, Wang L-G, Han Y, He Q-Y. clusterProfiler: an R package for comparing biological themes among gene clusters. OMICS. 2012;16:284–7.

71. Ashburner M, Ball CA, Blake JA, Botstein D, Butler H, Cherry JM, et al. Gene ontology: tool for the unification of biology. The Gene Ontology Consortium. Nat Genet. 2000;25:25–9.

72. Croft D, O'Kelly G, Wu G, Haw R, Gillespie M, Matthews L, et al. Reactome: a database of reactions, pathways and biological processes. Nucleic Acids Res. 2011;39:D691–7.

73. Butler A, Hoffman P, Smibert P, Papalexi E, Satija R. Integrating single-cell transcriptomic data across different conditions, technologies, and species. Nat Biotechnol. 2018;36:411–20.

74. Korsunsky I, Millard N, Fan J, Slowikowski K, Zhang F, Wei K, et al. Fast, sensitive and accurate integration of single-cell data with Harmony. Nat Methods. 2019;16:1289–96.

75. Wherry EJ. T cell exhaustion. Nat Immunol. 2011;12:492–9.

76. Guo X, Zhang Y, Zheng L, Zheng C, Song J, Zhang Q, et al. Global characterization of T cells in non-small-cell lung cancer by single-cell sequencing. Nat Med. 2018;24:978–85.

77. Tuomela S, Rautio S, Ahlfors H, Öling V, Salo V, Ullah U, et al. Comparative analysis of human and mouse transcriptomes of Th17 cell priming. Gene Expression Omnibus. 2016. https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE52260.

78. Ubaid Ullah N, Andrabi SBA, Tripathi SK, Dirasantha O, Kanduri K, Rautio S, et al. Transcriptional Repressor HIC1 Contributes to Suppressive Function of Human Induced Regulatory T Cells. Gene Expression Omnibus. 2018. https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE90569.

79. Schmidt A, Marabita F, Kiani NA, Gross CC, Johansson HJ, Éliás S, et al. Time-resolved transcriptome and proteome landscape of human regulatory T cell (Treg) differentiation reveals novel regulators of FOXP3. Gene Expression Omnibus. 2018. https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE94396.

80. Schmidt A, Marabita F, Kiani NA, Gross CC, Johansson HJ, Éliás S, et al. [Duplikat] Time-resolved transcriptome and proteome landscape of human regulatory T cell (Treg) differentiation reveals novel regulators of FOXP3. Gene Expression Omnibus. 2018. https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE96538.

81. Elo LL, Järvenpää H, Tuomela S, Raghav S, Ahlfors H, Laurila K, et al. Genome-wide Profiling of Interleukin-4 and STAT6 Transcription Factor Regulation of Human Th2 Cell Programming. Gene Expression Omnibus. 2010. https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE17974.

82. Aijö T, Edelman SM, Lönnberg T, Larjo A, Kallionpää H, Tuomela S, et al. An integrative computational systems biology approach identifies differentially regulated dynamic transcriptome signatures which drive the initiation of human T helper cell differentiation. Gene Expression Omnibus. 2012. https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE32959.

83. Gutierrez-Arcelus M, Baglaenko Y, Arora J, Hannes S, Luo Y, Amariuta T, et al. Allele-specific expression changes dynamically during T cell activation in HLA and other autoimmune loci. Gene Expression Omnibus. 2019. https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE140244.

84. Rade M, Böhlen S, Neuhaus V, Löffler D, Blumert C, Köhl U, et al. A time-resolved meta-analysis of consensus gene expression profiles during human T-cell activation. Gene Expression Omnibus. 2023. https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE197067.

85. Rade M, Böhlen S, Neuhaus V, Löffler D, Blumert C, Köhl U, et al. A time-resolved meta-analysis of consensus gene expression profiles during human T-cell activation. Github. 2023. https://github.com/fraunhofer-izi/Rade_et_al_tcell_2023.

86. Rade M, Böhlen S, Neuhaus V, Löffler D, Blumert C, Köhl U, et al. A time-resolved meta-analysis of consensus gene expression profiles during human T-cell activation. 2023. Zenodo. https://doi.org/10.5281/zenodo.10145756.

87. Gautier L, Cope L, Bolstad BM, Irizarry RA. affy–analysis of Affymetrix GeneChip data at the probe level. Bioinformatics. 2004;20:307–15. Accessed 24 Jul 2020.

88. Dai M, Wang P, Boyd AD, Kostov G, Athey B, Jones EG, et al. Evolving gene/transcript definitions significantly alter the interpretation of GeneChip data. Nucleic Acids Res. 2005;33:e175. Accessed 26 Oct 2018.

89. Brettschneider J, Collin F, Bolstad BM, Speed TP. Quality Assessment for Short Oligonucleotide Microarray Data. Technometrics. 2008;50:241–64. Accessed 29 Oct 2019.

90. Chockalingam S, Aluru M, Aluru S. Microarray Data Processing Techniques for Genome-Scale Network Inference from Large Public Repositories. Microarrays (Basel). 2016;5.

91. Illumina, Inc. bcl2fastq2 Conversion Software v2.19. 2017. Accessed 07 Mar 2017.
92. Schubert M, Lindgreen S, Orlando L. AdapterRemoval v2: rapid adapter trimming, identification, and read merging. BMC Res Notes. 2016;9:88. Accessed 12 Oct 2019.
93. Kim D, Paggi JM, Park C, Bennett C, Salzberg SL. Graph-based genome alignment and genotyping with HISAT2 and HISAT-genotype. Nat Biotechnol. 2019;37:907–15. Accessed 08 Jun 2017.
94. Anders S, Pyl PT, Huber W. HTSeq–a Python framework to work with high-throughput sequencing data. Bioinformatics. 2015;31:166–9. Accessed 01 Jan 2019.
95. Andrews S. Babraham Bioinformatics - FastQC A Quality Control tool for High Throughput Sequence Data. https://www.bioinformatics.babraham.ac.uk/projects/fastqc/. Accessed 04 Oct 2018.
96. Daley T, Smith AD. Predicting the molecular complexity of sequencing libraries. Nat Methods. 2013;10:325–7.
97. Wang L, Wang S, Li W. RSeQC: quality control of RNA-seq experiments. Bioinformatics. 2012;28:2184–5.
98. Wingett SW, Andrews S. FastQ Screen: A tool for multi-genome mapping and quality control. F1000Res. 2018;7:1338.
99. Langmead B, Salzberg SL. Fast gapped-read alignment with Bowtie 2. Nat Methods. 2012;9:357–9.
100. Ewels P, Magnusson M, Lundin S, Käller M. MultiQC: summarize analysis results for multiple tools and samples in a single report. Bioinformatics. 2016;32:3047–8.
101. McCarthy DJ, Chen Y, Smyth GK. Differential expression analysis of multifactor RNA-Seq experiments with respect to biological variation. Nucleic Acids Res. 2012;40:4288–97.
102. Robinson MD, Oshlack A. A scaling normalization method for differential expression analysis of RNA-seq data. Genome Biol. 2010;11:R25.
103. Law CW, Chen Y, Shi W, Smyth GK. voom: precision weights unlock linear model analysis tools for RNA-seq read counts. Genome Biol. 2014;15:R29.
104. McCarthy DJ, Smyth GK. Testing significance relative to a fold-change threshold is a TREAT. Bioinformatics. 2009;25:765–71.
105. Wilhelm BT, Landry J-R. RNA-Seq-quantitative measurement of expression through massively parallel RNA-sequencing. Methods. 2009;48:249–57.

Publisher's Note