Early detection of peripheral blood cell signature in children developing beta-cell autoimmunity at a young age

Henna Kallionpää[1*], Juhi Somani[2*], Soile Tuomela[1*], Ubaid Ullah[1*], Rafael de Albuquerque[1], Tapio Lönnberg[1], Elina Komsi[1], Heli Siljander[3,4], Jarno Honkanen[3,4] Taina Härkönen[3,4], Aleksandr Peet[5,6], Vallo Tillmann[5,6], Vikash Chandra[3,7], Mahesh Kumar Anagandula[8], Gun Frisk[8], Timo Otonkoski[3,7], Omid Rasool[1], Riikka Lund[1], Harri Lähdesmäki[2ǂ], Mikael Knip[3,4,9,10ǂ], Riitta Lahesmaa[1ǂ#]

[1]Turku Centre for Biotechnology, University of Turku and Åbo Akademi University, Turku, Finland

[2]Department of Computer Science, Aalto University School of Science, Espoo, Finland

[3]Children's Hospital, University of Helsinki and Helsinki University Hospital, Helsinki, Finland

[4]Research Programs Unit, Diabetes and Obesity, University of Helsinki, Helsinki, Finland

[5]Department of Pediatrics, University of Tartu, Tartu, Estonia

[6]Children's clinic of Tartu, Tartu University Hospital, Tartu Estonia

[7]Research Programs Unit, Molecular Neurology and Biomedicum Stem Cell Centre, Faculty of Medicine, University of Helsinki, Helsinki 00014, Finland

[8]Department of Immunology, Genetics and Pathology, Uppsala University, Sweden

[9]Folkhälsan Research Center, Helsinki, Finland

[10]Tampere Center for Child Health Research, Tampere University Hospital, Tampere, Finland

*These authors contributed equally to this work ǂShared senior authorship

#Corresponding author: Riitta Lahesmaa, riitta.lahesmaa@btk.fi, Phone: +358 29 450 2415

Abstract

The appearance of Type 1 diabetes (T1D)-associated autoantibodies is the first and only measurable parameter to predict progression toward T1D in genetically susceptible individuals. However, autoantibodies indicate an active autoimmune reaction, wherein the immune tolerance is already broken. Therefore, there is a clear and urgent need for new biomarkers that predict the onset of the autoimmune reaction preceding autoantibody positivity or reflect progressive beta-cell destruction. Here we report the mRNA-sequencing-based analysis of 306 samples including fractionated samples of CD4+ and CD8+ T cells as well as CD4-CD8- cell~~s~~ fractions and unfractionated PBMC samples longitudinally collected from seven children that developed beta-cell autoimmunity (Cases) at a young age and their matched controls. We identified transcripts, including interleukin-32 (*IL32*) that were upregulated before T1D-associated autoantibodies appeared. Single cell RNA-seq studies reveal that high IL32 in Case samples were contributed mainly by activated T cells and NK cells. Further, we showed that IL32 expression can be induced by a virus and cytokines in pancreatic islets and beta-cells, respectively. The results provide a basis for early detection of aberrations in the immune system function before T1D and suggest a potential role for *IL32* in the pathogenesis of T1D.

Introduction

Family and sibling studies in Type 1 diabetes (T1D) have implicated a firm genetic predisposition to a locus containing HLA class I and class II genes on chromosome 6 suggesting a role for CD4+ as well as CD8+ T cells in T1D pathogenesis (1–3). As much as 30-50% of the genetic risk is conferred by HLA class II molecules, which are crucial in antigen presentation to CD4+ T cells. Further, CD4+ cells reactive to beta-cell antigen peptides are found in peripheral blood and the pancreas, and typically secrete the cytokine IFNγ (4,5). CD4+ cells orchestrate adaptive immune responses, including that of antibody secreting B cells as well as cytotoxic CD8+ T cells. Indeed circulating autoantibodies against beta-cell antigens may appear years before the clinical onset. Further, a cytolytic CD4+ subtype might directly contribute to target cell killing (6).

Although HLA class II is associated with the development of autoantibodies, HLA class I seems to be more strongly linked to disease progression (7). Histological analysis of pancreatic sections of cadaveric donors with T1D revealed that HLA class I is highly expressed in islets (8,9). Moreover, CD8+ cells are the most abundant cell type during insulitis (10), and the islets contain CD8+ cells specific for T1D autoantigens (11). Thus, the autoimmune cascade in T1D might be initiated by self-reactive CD4+ cells that activate B cells to produce autoantibodies that target the beta-cells and unleash the cytotoxic activity of the autoreactive CD8+ cells. The environmental factors triggering and driving the autoimmunity in T1D are poorly defined, but the disease has been associated with viral infections (12), diet in early childhood (13), and reduced diversity of gut microbiota (14).

Currently, the appearance of T1D-associated autoantibodies is the first and only measurable parameter to predict progression toward T1D in genetically susceptible individuals. Although the disease progression rate varies considerably, children with genetic HLA risk expressing at least two T1D autoantibodies will very likely progress to clinical disease during the next 15 years (15). However, autoantibodies are poor prognostic markers for the timing of the clinical presentation of T1D. The appearance of autoantibodies indicates an active autoimmune reaction, wherein the immune tolerance is already broken. Therefore, there is a clear and urgent need for new biomarkers that predict the onset of the autoimmune reaction preceding autoantibody positivity or reflect progressive beta-cell destruction. Such markers would present a window for early intervention aimed at complete disease prevention. Earlier, we reported changes in whole-blood transcripts and serum proteins before the detection of diabetes-associated antibodies in children who later progressed to T1D (16,17). Therefore, we hypothesized that a comprehensive analysis of the transcriptome of longitudinal cellular samples including CD4+ and CD8+ T cells will lead to the identification of new early biomarkers.

Research Design and Methods

Study cohort

Samples were collected as part of the DIABIMMUNE study (FP7 grant no. 202063) from Finnish (n=10) and Estonian (n=4) participants (**Supplementary Table 1**). The HLA-DR-DQ genotypes were analysed as described earlier (18). 836 children with HLA-DR-DQ risk allele were monitored and sampled at 3, 6, 12, 18, 24 and 36 months of age. The study protocols were approved by the ethical committees of the participating hospitals, and the parents gave their written informed consent. Autoantibodies against insulin (IAA), glutamic acid decarboxylase (GADA), islet antigen-2 (IA-2A), and zinc transporter 8 (ZnT8A) were measured from serum with specific radiobinding assays (19). Islet cell antibodies (ICA) were analysed with immunofluorescence in autoantibody-positive subjects. The cut-off values were based on the 99th percentile in non-diabetic children, which were 2.80 relative units (RU) for IAA, 5.36 RU for GADA, 0.78 RU for IA-2A and 0.61 RU for ZnT8A. The detection limit in the ICA assay was 2.5 Juvenile Diabetes Foundation units (JDFU). A sample was considered seropositive when any of the autoantibodies exceeded the thresholds.

Sample collections

At each study visit, 8 ml of blood was drawn in sodium-heparin tubes (Vacutainer, 368480, BD). PBMCs were isolated by Ficoll-Paque centrifugation (17-1440-03 GE Healthcare), and were suspended in RPMI 1640 medium (42401-018, Gibco) supplemented with 10% DMSO (0231-500 ml, Thermo Scientific), 5% human AB serum (IPLA-SERAB-OTC, Innovative Research), 2 mM L-glutamine (G7513, Sigma-Aldrich), and 25 mM gentamicin (G-1397 Sigma-Aldrich). After overnight incubation at -80°C, samples were stored in liquid nitrogen (-180°C). For fractionation, PBMC samples were thawed quickly in a 37°C water bath, quantitated for cell numbers and viability. On an average 90% cells were viable. Magnetic antibody-coupled beads were used for sequential positive enrichment of CD4+ and CD8+ cells (11331D and 11333D Invitrogen). RNA was isolated from the samples with AllPrep kit (80224, Qiagen), and quantity and quality were determined using Qubit RNA assay (Q32852, Invitrogen) and Bioanalyzer 2100 (Agilent), respectively.

Bulk RNA-seq of PBMC and other fractions

At least 80 ng of total RNA was processed for RNA-seq with TruSeq Stranded mRNA Library Prep kit (RS-122-2101, Illumina). The sequencing was carried out with Illumina HiSeq2500 instrument using TruSeq v3; 2 x 100 bp chemistry. The average sequencing depth was around 51 million reads. Quality control was performed using FastQC (version 0.10.0). All the samples passed the quality criteria. The reads were aligned to the human reference transcriptome, GRCh37 assembly version 75 using TopHat (version 2.0.10) (20). Average mapping percentage was 93. The concordant pairs percentage was about 89. The aligned reads were counted with *htseq-count* (HTSeq 0.6.1; overlap mode of 'intersection-strict') (21). The read counts of genes were normalized using the trimmed means of the M-values (TMM implemented in the *edgeR* (22). Coding, noncoding information were taken from ensembl. Differential expression analyses were conducted separately for coding and non-coding genes, using the *edgeR* (22). The variance of the data was estimated using the trended dispersion method. Further filtering step retained only those genes as differentially expressed (DE) had that |median $\log_2$FC| > 0.5 and had more than 65% samples across all individuals regulated in the same direction (i.e., up- or down-regulated). These filtering steps were added to discard false positives that may arise due to the heterogeneity of the samples due to normal variation, which is non-related to T1D and outliers. A flow chart of the scheme of analysis has been shown in **Supplementary Fig. 1**.

Single-cell RNA-seq (scRNA-seq)

The concentrations of the PBMC samples varied from 0.55 to 1.80 x 10^6 cells/ml. From each sample, we aimed at the recovery of 5000 single cells, loading approximately 9000 cells on the Chromium Controller using Single Cell 3′ Solution v2 reagents and following manufacturer's instructions (CG00052 Rev B, 10x Genomics)._scRNA-seq sample processing was carried out in three batches on consecutive days using the same lot of reagents and chips for all samples. The cDNA was further amplified using a Veriti Thermal Cycler (Applied Biosystems/Thermo Fisher), followed by clean-up (SPRIselect kit, Beckman Coulter). Finally, enzymatic fragmentation, end repair, A-tailing, adaptor ligation and PCR were performed to produce indexed libraries, which were sequenced with Illumina HiSeq 3000 (one sample / lane) using paired end sequencing and 26 + 98 bp read-length configuration. The data were processed using the Cell Ranger pipeline version 2.0.0 yielding on average 2546 viable cells per sample, and 114,309 reads per cell.

The reads were aligned to the human reference genome (hg19) using STAR (23). The mean raw reads per cell varied 57-200 k. QC analysis and further exploration was done using Seurat (24). After filtering steps, 18,396 cells expressing 20,830 genes were retained. For details on the filtering steps please see “Supplementary Material”. The data were normalized using Seurat’s default. Highly variable genes (HVGs) were selected for principal component analysis (PCA). The top 20 PCs were used in the graph-based clustering. To identify marker genes for each cluster, cells of a single cluster were compared to the cells of all other clusters combined. A gene was considered a marker of a cluster if it was expressed in at least 25% of the cells of either of the two groups and the logFC between the cluster and all other clusters was at least 0.25.

For trajectory analysis**,** the pooled cells were ordered in pseudotime (i.e., placed along a trajectory corresponding to a type of biological transition, such as differentiation) using Monocle 2 (25). The analysis was performed on cells specifically from CD4+ and CD8+ T-cell clusters. For the details on the trajectory analyses, please see “Supplementary Material”.

RT-PCR analysis

For PBMC samples, 50 ng of total RNA was treated with DNaseI (Invitrogen), and cDNA was synthesized with Transcriptor First Strand cDNA Synthesis Kit (Roche). For isoform-specific (*IL32α*, *β*, and γ) assay, qPCR analysis was performed in triplicate runs using SYBR Select master mix (Applied Biosystems). ΔCt values were calculated relative to *EF1α*. For CD4+ T cells and pancreatic islets, RNA was isolated using the RNeasy Mini Kit (74106, Qiagen) and RNeasy Plus Mini Kit (74134, Qiagen), respectively. Purified RNA was treated with DNaseI and cDNA was synthesized with SuperScript II Reverse Transcriptase (18064014, Invitrogen). For the detection of global *IL32* qPCR reactions were run using a custom TaqMan Gene Expression Assay reagent (#AJ5IQA9, Thermo Scientific) in duplicate and in two separate runs. ΔCt values were calculated relative to *GAPDH*. The amplification was monitored with QuantStudio 12K Flex Real-Time PCR System, under the following PCR conditions: 10 minutes at 95 °C, followed by 40 cycles of 15" at 95 °C and 60" at 60 °C and analysed with QuantStudio Software on Thermo Cloud.

For EndoC-βH1 cells data, cDNA was synthesized using the Maxima first-strand cDNA synthesis kit as per manufacturer's recommendations (Thermo Fisher Scientific). All reactions were performed in duplicates on at least three biological replicates. *Cyclophilin-A* was used as an endogenous control. Primer sequences are presented in **Supplementary Table 2**.

ELISA

To measure secreted IL-32 levels we used IL-32 duoset ELISA kit (R&D Systems, (DY3040-05 and DY008) following manufacturer's instructions.

Intracellular staining and flow cytometry

The cells were fixed for 10 minutes in Fix buffer I (BD, 557870), followed by 45 minutes permeabilization using ice-cold permeabilization buffer III (BD, 558050). The cells were stained using APC-conjugated IL-32α antibody (R&D, IC30402A) and FITC-conjugated IFNγ antibody (Invitrogen, MHCIFG01) in PBS containing 0.5% FCS. The data were acquired in BD Fortessa and analysed using FlowJo (version 10.4.2).

EndoC-βH1 cell culture

The EndoC-βH1 human beta-cell line was obtained from Univercell Biosolution S.A.S., France. The cells were cultured as described (26). EndoC-βH1 cells were stimulated with either IL-32γ alone (100 ng/ml, R&D Systems) or in combination with a cocktail of IL-1β (5 ng/ml, R&D Systems) and IFN-γ (50 ng/ml, R&D Systems) for 24 h. RNA samples were collected at the end of each treatment and analysed by RT-qPCR.

Human CD4 T-cell isolation and culturing

CD4+ T cells were isolated from cord-blood collected from neonates born in Turku University Hospital and were cultured in IMDM containing 1%AB serum in absence (Th0) or presence (Th1) of 2.5ng/ml if IL-12 (R&D). Cells were activated with plate bound CD3 (0.5 µg/well of a 24 well-plate) and soluble CD28 (0.5 µg/ml), both from Immunotech, with or without 50 ng/ml rIL-32-γ (R&D). 12 ng/ml IL-2 was added at 48h. For IFNγ neutralization, anti-IFNγ antibody (10 µg/ml, R&D: MAB285) was used. For reactivation, cells were treated with 5ng/ml PMA (Calbiochem) and 0.5pg/ml Ionomycin (Sigma) for 5h.

Human pancreatic islets, their infection with Coxsackie B Virus

Human islets were isolated from pancreases obtained from brain dead organ donors and purified by handpicking to a purity of > 90%. Islet culturing and virus infection with Coxsackie B virus-1 (CBV-1-7-10796 (CBV-1-7) was performed as described (27). Islets were collected at the day 4 timepoint, and RNA was extracted using the RNeasy Plus Mini Kit or the AllPrep DNA/RNA Mini Kit (Qiagen). For RNA-seq, 100 ng of total RNA from three donors was used for library preparation according to Illumina TruSeq RNA Sample Preparation v2 Guide (part # 15026495). The high quality of the libraries was confirmed with Agilent Bioanalyzer 2100 and Qubit Fluorometric Quantitation (Life Technologies). The libraries were pooled in two pools and run in 2 lanes on the Illumina HiSeq 2500 instrument using 2 x 100 bp.

Results

Fractionation of PBMC sample into CD4+, CD8+ and CD4-CD8- cellular subsets reveals distinct and overlapping gene expression signatures

We performed RNA-seq of 306 longitudinal samples ~~of~~ including unfractionated PBMCs, as well as CD4 enriched (CD4+), CD8 enriched (CD8+), and CD4 and CD8 cell depleted (CD4-CD8-) cell fractions from seven Case-Control pairs (**Table 1**). The seven Case children who developed T1D-related autoantibodies (Aab+) were selected from the DIABIMMUNE Birth Cohort (18), where HLA-susceptible children are sampled at 3–36 months of age (**Fig. 1A**). All seven children developed T1D-associated autoantibodies by the age of 2 years (**Table 1**) and four of them developed clinical T1D between the ages of 2.4 and 3.7 years. For each Case, an autoantibody-negative Control child was matched for gender, date and place of birth, and HLA-conferred risk category.

The samples clustered according to the cell fraction (**Fig. 1B**) and the clustering was not affected by Case-Control status or sampling age, indicating that cell fraction–specific differences dominated over variation derived from other factors (**Supplementary Fig. 2A and 2B**). When CD4+, CD8+ and CD4-CD8- samples from Controls were compared to the unfractionated PBMC samples, 889, 399, and 1002 genes were specifically in CD4+ (e.g., *CD28*, *CTLA4*), CD8+ (e.g., *CD8A*, *CD8B*, *KLRK1*), and CD4-CD8- (e.g., *IL1A*, *IL1B*, *IL6)* fractions, respectively (**Fig. 1C** and **Supplementary Table 3**). CD4+ and CD8+ fractions shared 1815 DE genes, of which 1803 genes (99%) were concordant (either up or down in both fractions) (**Supplementary Fig. 2C, Supplementary Table 3**). In summary, fractionation of the PBMC population based on the T-cell phenotype allowed improved detection of DE genes and enabled identification of cell subset–specific gene expression signatures.

RNA-seq analysis identifies transcriptomic changes associated with beta-cell autoimmunity

Comparison of Case samples to their respective Controls identified 51, 69, 143 and 85 genes as DE (FDR<0.05) in CD4+, CD8+, CD4-CD8- and

PBMC fractions, respectively (**Supplementary Table 4**); total of 278 unique DE genes in one or more fractions (**Fig. 2A**). Six genes *AMICA1*, *BTN3A2*, *IL32*, *RPSAP15*, *RPSAP58* and *WASH7P* were upregulated in the Cases in all four fractions (**Fig. 2A**). Only 16% of the DE genes have previously been reported as DE in genetically susceptible prediabetic children using microarrays (16,28,29) or RT-PCR (30–32), confirming dysregulation of these genes in children progressing to T1D. Besides protein-coding genes, 54 non-coding genes, including three antisense, two sense intronic, seven enhancer and 18 promoter-associated lncRNAs, were DE. To our knowledge, none of these lncRNAs has been linked to the aetiology of T1D (16,28–32).

Hierarchical clustering identifies co-regulated gene expression clusters associated with T1D autoimmunity

Gene- and sample-wise hierarchical clustering for each cell fraction, including unfractionated PBMCs (also referred to as a fraction henceforth) identified a cluster, upregulated in the Case samples in all four fractions (**Fig. 2B and Supplementary Fig. 3A-D**). Interestingly, this cluster consistently contained *IL32* and *BTN3A2*, along with other fraction-specific genes (**Fig. 2C**). In the CD8+ fraction, expression of a distinct cluster, including *IFNG,* was lower in most of the Case samples than Control samples (**Supplementary Fig. 3B**). Surprisingly, in the PBMC fraction, we detected Case-specific upregulation of a cluster, including insulin (*INS*), glucagon (*CGC*) and regulin 1 alpha (*REG1A*) transcripts (**Supplementary Fig. 3D**), which are predominantly expressed in the pancreas.

To explicitly define coregulated genes in these clusters, we calculated Euclidean distances for *IL32* (in each fraction), *IFNG* (in CD8+ fraction), and *INS* (in PBMC fraction) and considered the genes with a median Euclidean distance < 2.5 across all Case-Control pairs to be co-clustering with the gene of interest (**Supplementary Table 5A**). In three of the four fractions, the *IL32* cluster included *BTN3A2*, *AMICA1*, *LARS* and *RSU1* (**Fig. 2C**). *IL32*, *AMICA1* and *BNT3A2* show concerted gene expression profiles in CD4+ samples (**Fig. 2D**). In at least two of four fractions, this cluster also comprised *TRBV4-1*, *TMEM14C*, *UROS*, *WASH7P*, *BTN3A3*, *CARD8*, *CCDC167* and *LINC01184*. The profile of these and other interesting genes are shown in **Supplementary Fig. 4A-AB.** Upon examining the overrepresented transcription factor binding sites (TFBS) on the promoters of *IL32* cluster genes, the V$IK_Q5_01 motif bound by Ikaros (IKZF1) was revealed to be among the enriched TFBS shared in both the CD4+ and PBMC fractions (**Supplementary Table 5B**). IKZF1 has been genetically associated with T1D (33). The T1D-associated risk allele rs10272724 (T) increases IKZF1 transcript level (34).

IFNG cluster of the CD8+ cells included *TBX21* (codes for TBET), *BHLHE40*, and *ZEB2*, transcription factors expressed in CD8+ T cells (35), as well as *NKG7*, *OASL*, and *KLRD1* (**Supplementary Table 5A**). ZEB2 has been reported to drive terminal effector CD8+ cell differentiation together with T-bet (36). In the PBMC fraction, *GCG* and *REG1A* were coregulated with *INS* (**Supplementary Table 4A, Supplementary Fig. 5**).

Transcriptional changes preceding the appearance of T1D-related autoantibodies are enriched in the CD8+ T-cell fraction

To identify changes that occur immediately before the first detection of T1D-related autoantibodies (i.e., seroconversion), we performed a separate differential expression analysis for the samples drawn at most 12 months before seroconversion. Altogether 121 coding and non-coding genes were DE in Cases, as compared to their matched Controls (**Supplementary Table 4** and **Supplementary Fig. 6**). Notably, more than half of these (58%) were detected only in the CD8+ fraction. Besides *IL32*, only two other genes were common to all fractions *RPSAP58*, and *RPSAP15*, both being the pseudogenes with unknown functions with very similar expression profiles **(Supplementary Fig. 4M-T)**.

Higher *IL32* expression in Cases was validated using qRT-PCR. Interestingly, all three major isoforms (*IL32α*, *IL-32β* and *IL32γ*) were upregulated in PBMC samples in all the Case children at each of the time points including 3 months (**Fig. 3A** and **Supplementary Fig. 7**). Among these isoforms, *IL-32γ* was expressed at the highest level, followed by *IL-32β* and *IL-32α*.

Single-cell RNA sequencing (scRNA-seq) identifies T and NK cells as the *IL32* high population

To specify the cell populations responsible for the *IL32* and *INS* signatures, we performed scRNA-seq on four selected Case and their nearest matched

Control PBMC samples where the expression of *IL32* or *INS* was high (or low) based on the bulk RNA-seq data (**Supplementary Table 6**). Unsupervised clustering of 18,396 single cells from all eight PBMC scRNA-seq runs identified 13 clusters (**Fig. 3B and Supplementary Fig. 8**). The two largest clusters expressing high *CCR7* were merged as one cluster of naive T cells reducing the number of clusters to 12. Clusters named as *RGCC+ T cells*, *CD62L+ T cells*, and *Activated Th cells* expressed lower levels of *CCR7*. *Activated CD8+ T cells* cluster expressed high levels of *CD8A* and *CD8B* as well as *NKG7* and two separate clusters of CD8+ T cells expressing either granulysin or granzyme A were observed (*Activated GNLY+ CD8+ T cells* and *Activated GZMA+ CD8+ T cells*, respectively). A subcluster of Activated GZMA+ CD8+ cells had higher expression of cell-cycle genes (e.g., *STMN1*, *TUBA1B*) and was named Activated proliferating GZMA+ CD8+ T cells. An NK cell cluster was positive for expression of *CD56*, *NKG7*, and *GNLY* and negative for *CD8A* and *CD3E*. A B-cell cluster was identified by the expression of *MS4A1*, *CD79A* and *CD79B*, whereas the Monocyte/DCs cluster was composed of cells expressing *CD14* or *FCGR3A, LYZ and TYROBP*. Interestingly, the expression of many HLA class II molecules was as high in B cells as in monocytes, suggesting high antigen-presentation potential.

The contribution of different Case or Control samples to the cells in a given cellular population (cluster) varied from cluster to cluster (**Supplementary Fig. 9** and **10A-B**). The naive T cells cluster was dominated by the cells from the Control samples ($p<0.05$) whereas the

Monocyte/DC cluster had more cells from Cases (p<0.005, **Supplementary Fig. 10B**). Case 9, with the highest *IL32* expression levels in the bulk RNA-seq data, dominated the *CD62L+ T-cell* cluster, *Activated NK cells*, and most clearly, *Activated and proliferating GZMA+ CD8+ T cell* clusters (**Supplementary Fig. 10B**). Conversely, Control children 5 and 9 seemed to dominate the cluster of *Developing T cells* expressing pre-T-cell receptor *PTCRA* suggesting the presence of immature T cells in those samples.

Insulin, glucagon, or *REG1A* expression were not detected even in the *INS*-high samples of Cases 5 and 9, leaving the origin of these transcripts in bulk RNA-seq as an open question. In contrast, *IL32* expression was clear, and as expected, it was explicitly over-expressed in the Case samples (**Supplementary Fig. 11**). *IL32* was expressed at a very low level in Monocyte/DC, B cells, and *Developing T cell* clusters, however, it was expressed at higher levels by both the T cells and the NK cells **(Fig. 3C)**.

To further define the relationship of *IL32* expression and T-cell activation status, we performed separate trajectory analyses for the CD4+ and CD8+ T cells. The less activated precursor populations (naive and RGCC+ T cells), which detect CD4 and CD8 transcripts in low abundances, were used as starting point for the trajectory analyses. The results revealed three major cellular branches (I-III) in the data both in CD4+ as well as CD8+ T cells (**Fig. 3D-I**). The branch I consisted mainly of naive T cells, among which cells from the Control samples were enriched (**Fig. 3E and**

H, Supplementary Fig. 12). In contrast, the highest levels of *IL32* were expressed by cells close to the end points of branches II and III, corresponding to more advanced stages of differentiation (**Fig. 3F** and **I, Supplementary Fig. 12**).

IL-32 and IFNγ are co-expressed by Th1 cells

To further study IL-32 expression, we measured intracellular IL-32 expression at protein level in CD4+ T cells isolated from human umbilical cord blood. Cells were either activated through CD3/CD28 in the absence of cytokine (Th0) or were differentiated towards a Th1 cell lineage for 72h. IL-32 was induced upon activation and, unlike IFNγ, was expressed both in Th0 as well as Th1 cells **(Fig. 4A)**. Interestingly in Th1 cells, most IFNγ-producing cells were also positive for IL-32 (**Fig. 4A; Supplementary Fig. 13A**) and the proportions of IL-32-positive cells and the per cell IL-32 levels were higher in IFNγ-producing Th1 cells than in Th0 cells (**Fig. 4B-C**). Furthermore, neutralization of IFNγ significantly reduced IL-32 secretion by Th1 cells (**Fig. 4D**) confirming that IFNγ positively regulates *IL32* expression. IL-32 expression was also induced by IL-32 itself in Th1 cells, both at the RNA level (**Fig. 4E**) as well as in the culture supernatant upon 48 h re-stimulation after seven days of polarization in Th1 condition (**Fig. 4F**).

Pancreatic beta-cells can express IL32 in response to cytokine stimulation and viral infection

To study how the elevated *IL32* expression may influence beta-cell function, we treated human EndoC-βH1 beta-cell line for 24 h with either recombinant IL-32γ alone or in combination with the pro-inflammatory cytokines IL-1β and IFNγ. In agreement with earlier published data on pancreatic ductal cancer cell lines (37), IL-1β and IFNγ significantly induced *IL32* expression in human EndoC-βH1 cells (**Fig. 4G)** However, addition of IL-32γ did not i) further enhance the IL-1β- and IFNγ-induced *IL32* expression, ii) the expression of inflammatory cytokines *TNFA, IL6* and *IL8* (**Fig. 4G**), iii) the expression of ER stress marker genes (ATF3, ATF4, ATF6, HSPA5, CHOP, sXBP1) (**Supplementary Fig. 13B**) in EndoC-βH1 cells. Furthermore, the IL-32γ treatment did not affect the expression of beta-cell–specific genes, such as *INS*, *MAFA* or *PDX1* (**Supplementary Fig. 13C**). These results suggest that, while IL-32 does not appear to directly affect the survival or the differentiation status of the beta-cells, beta-cells actively contribute to inflammation in the islets by secreting IL-32 upon stimulation by cytokines.

Coxsackie B viruses are beta-cell trophic viruses that have been linked to the development of T1D (38–43). To study the possible trigger of *IL32* expression in beta-cells, we infected purified human pancreatic islets of three cadaveric donors with Coxsackie B virus CBV1-7 strain. Infection by the virus led to the induction of *IL32* expression in the islets (**Fig. 4H**). We further validated this finding in the three islet samples used for RNA-seq as well as one additional islet sample using qRT-PCR assays and found a consistent increase in the IL32 expression upon CBV1-7 infection (**Fig. 4I**).

Taken together these results suggest that upon a viral infection **(Fig. 4H-I)** or a cytokine rush **(Fig. 4H)**, beta-cells may upregulate IL-32 secretion contributing to inflammation.

Discussion

We identified a panel of novel molecular players detected early in children who developed T1D-associated autoantibodies or even the clinical disease at a young age. Since the immunological changes related to T1D are known to be strongest among the T1D cases diagnosed at an early age (44), focusing on this age group should enhance the possibility to detect aberrations in the immune system predisposing to the disease. In this study, unbiased RNA-seq of CD4+ and CD8+ cells revealed many T1D-associated DE transcripts not previously reported. Analysis of the PBMC population offers an excellent overview of stable gene expression patterns but, at the same time, appears to mask some of the subtle fraction-specific changes. Such changes included upregulation of *CD52* detected only in the CD4+ cell fraction and downregulation of the *IFNG* and associated transcription factors ZEB2, TBX21 and ZNF683 detected specifically in the CD8+ cells. Further studies are needed to understand whether at-risk children have defects in formulating effector CD8+ response, or their effector CD8+ cells have homed to the sites of inflammation in the pancreas.

We selected *IL32* as our candidate for functional studies because it has not been linked to seroconversion before, it is easy to measure with available assays from clinical samples, and as a secreted molecule it can potentially affect the function of several cell types in paracrine and systemic fashion. Increased expression of IL32 in Cases across many cell types before seroconversion suggest that *IL32* is a critical member of the immunological signature characteristic for children developing beta-cell autoimmunity.

IL-32 is expressed by many immune and epithelial cells and has been described to be proinflammatory (45). However, to our knowledge, it has not been associated with human beta-cell autoimmunity. In contrast, *IL32* is downregulated in CD4+ T cells from recently diagnosed adult T1D patients (46) which along with our findings suggests a dynamic changes in immune cell signalling during the pathogenesis of the disease. On the other hand, IL-32 overexpression was observed in synovial biopsies of patients with rheumatoid arthritis (47), in inflamed mucosa of inflammatory bowel disease patients (48), and in the serum of myasthenia gravis patients (49) indicating a connection between IL32 and autoimmunity in general. In T cells, IL-32 is induced by T-cell activation, and it modulates human CD4+ T-cell effector function by promoting Th1 and Th17 responses (50). Both Th1 and Th17 cells have been linked to the T1D pathogenesis in both human and mouse(50). The *IL32* gene has been identified only in higher mammals, excluding rodents. Nonetheless, human IL-32γ transgenic mice exhibit impaired glucose tolerance, increased levels of IFNγ and other proinflammatory cytokines in the pancreas, as well as accelerated streptozotocin-induced experimental T1D

(51). No specific cell-surface receptor for IL-32 has been identified, but it may act through cell-surface integrins or proteinase-3 (52).

Our results showed that *IL32* was often co-regulated with genes previously linked to autoimmunity. For example, the BTN3 gene cluster reside in the extended MHC Class I locus. Further, BTN3 genes have been associated with T1D in a genetic screen, especially in the case of BTN3A2 (53). AMICA1, is a plasma membrane protein involved in lymphocyte migration through its interaction with Coxsackie-adenovirus receptor (CAR) expressed in epithelial cells and has been associated with multiple sclerosis (54). An analogous scenario could be envisaged for T1D: CAR is expressed by the pancreatic islet cells, including beta-cells (42), and its expression is elevated in autoantibody-positive individuals and T1D patients (55) suggesting that it might help recruit T cells to the islets. Interestingly, the findings point to human-specific phenomena not detectable in mouse models as IL-32 and the BTN3 protein family are not encoded by the mouse genome.

The strength of our study is that the children studied here comprise a homogeneous population with the early appearance of T1D-associated autoantibodies. Increasing evidence suggests that T1D can be subdivided to different phenotypes, e.g. characterized by age-dependent B-cell infiltration in the pancreas (56), defect in Coxsackievirus-induced antibody response in children with early insulin autoimmunity (57), or rapid versus slow progression to clinical disease (58). Thus, our results may not apply to “late progressors”, adolescents, and adults. Although the analysis of the global transcriptome of T-cell subsets of prediabetic children over the period of seroconversion is unique, a limitation of the current study is the analysis of only seven Aab+ children. The results of this study need to be validated and expanded on a larger cohort of prediabetic children but serve as a starting point for better understanding of immunological changes preceding the clinical onset of the disease. In the future, we are interested in addressing if our findings on cellular level are reflected also in IL-32 levels in plasma as well as to study if IL-32 alone or in combination of other identified molecules would have sufficient sensitivity and specificity as early indicators for T1D.

Acknowledgements

The authors are grateful for families for their participation in the DIABIMMUNE study. The DIABIMMUNE study group is acknowledged for their excellent collaborations, work with the families, and collection of the samples for the study. Marjo Hakkarainen, Sarita Heinonen, Päivi Junni and Elina Louramo are acknowledged for their skilful assistance in the laboratory. NGS sequencing was performed at the Finnish Functional Genomics Centre (FFGC), Turku, part of the Biocenter Finland network. We thank Professor Satu Mustjoki and her team at University of Helsinki for advice in designing scRNA-seq experiments and Riina Kaukonen at the FFGC for her careful execution. This work was financially supported by the Juvenile Diabetes Research Foundation (JDRF), The Academy of Finland

(AoF) Centre of Excellence in Molecular Systems Immunology and Physiology Research (SyMMyS) 2012-2017 (grant no. 250114), AoF Personalized Medicine Program (grant no. 292482), and the AoF grants 294337, 292335, 319280 and 314444, the Sigrid Jusélius Foundation, The Diabetes Research Foundation (Diabetestutkimussäätiö), Innovative 984 Medicines Initiative 2 Joint Undertaking under grant agreement No 115797 (INNODIA). This 985 Joint Undertaking receives support from the Union's Horizon 2020 research and innovation 986 programme and "EFPIA", 'JDRF" and "The Leona M. and Harry B. Helmsley Charitable 987 Trust" and The NOVO Nordisk Foundation. The DIABIMMUNE study was supported by the European Union Seventh Framework Programme (grant no. 202063). TL was supported by the AoF (Decision 311081).

Author Contributions

JS conducted bioinformatic analyses. HK, ST and UU led biological interpretation of the results. HK, ST, JS and UU drafted the manuscript. HK, JS and UU prepared the figures. HK was responsible for supervising EK. TH, HS, JH, AP and VT were responsible for sample collection, sample storage, and further clinical information of the children. RLu was responsible for study design, cell fractionation, sample analysis and data production. TL provided expertise in scRNA-seq study design, sample and data analysis, and interpretation of the results. RA, EK and OR were responsible for the isoform-specific *IL32* RT-PCR assay and the intracellular IL-32 staining in T cells and interpretation of the results. VC and TO carried out the experiments and interpreted the results of the studies in pancreatic beta-cells. MKA and GF were responsible for experiments on virus-infected pancreatic islets. HL was responsible for computational data analysis, interpretation of the results, editing the manuscript and supervising JS. MK was responsible for the DIABIMMUNE study design, sample collection, sample storage, clinical information of the children, directing of the clinical study, interpreting the results and editing the manuscript. RL was responsible for study design, sample and data analysis, interpretation of the results, writing the manuscript and supervision of the study. All authors contributed to the final version of the manuscript.

Guarantor Statement

RL and HL are the guarantors of this work, and had full access to all the data in the study and takes responsibility for the integrity of the data and the accuracy of the data analysis.

Prior Presentation Information

The results described in this study have not be presented in any conference/proceedings elsewhere.

Conflict of Interests

The authors declare that they have no conflict of interest.

Data and Resource Availability

All the raw data will be deposited to European genome-phenome archive (EGA) for access. The study does not involve any non-commercial reagents and tools.

REFERENCES

1. Todd JA, Bell JI, McDevitt HO. HLA-DQ beta gene contributes to susceptibility and resistance to insulin-dependent diabetes mellitus. Nature. 1987 Oct;329(6140):599–604.
2. Nejentsev S, Howson JM, Walker NM, Szeszko J, Field SF, Stevens HE, et al. Localization of type 1 diabetes susceptibility to the MHC class I genes HLA-B and HLA-A. Nature. 2007 Dec;450(7171):887–92.
3. Todd JA. Etiology of type 1 diabetes. Immunity. 2010 Apr;32(4):457–67.
4. Babon JA, DeNicola ME, Blodgett DM, Crevecoeur I, Buttrick TS, Maehr R, et al. Analysis of self-antigen specificity of islet-infiltrating T cells from human donors with type 1 diabetes. Nat Med. 2016 Dec;22(12):1482–7.
5. Delong T, Wiles TA, Baker RL, Bradley B, Barbour G, Reisdorph R, et al. Pathogenic CD4 T cells in type 1 diabetes recognize epitopes formed by peptide fusion. Science. 2016 Feb;351(6274):711–4.
6. Takeuchi A, Saito T. CD4 CTL, a Cytotoxic Subset of CD4(+) T Cells, Their Differentiation and Function. Front Immunol. 2017 Feb;8:194.
7. Lipponen K, Gombos Z, Kiviniemi M, Siljander H, Lempainen J, Hermann R, et al. Effect of HLA class I and class II alleles on progression from autoantibody positivity to overt type 1 diabetes in children with risk-associated class II genotypes. Diabetes. 2010 Dec;59(12):3253–6.
8. Foulis AK, Farquharson MA, Hardman R. Aberrant expression of class II major histocompatibility complex molecules by B cells and

hyperexpression of class I major histocompatibility complex molecules by insulin containing islets in type 1 (insulin-dependent) diabetes mellitus. Diabetologia. 1987 May;30(5):333–43.

9. Richardson SJ, Rodriguez-Calvo T, Gerling IC, Mathews CE, Kaddis JS, Russell MA, et al. Islet cell hyperexpression of HLA class I antigens: a defining feature in type 1 diabetes. Diabetologia. 2016 Nov;59(11):2448–58.

10. Willcox A, Richardson SJ, Bone AJ, Foulis AK, Morgan NG. Analysis of islet inflammation in human type 1 diabetes. Clin Exp Immunol [Internet]. 2009;155(2):173–81. Available from: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=2675247&tool=pmcentrez&rendertype=abstract

11. Coppieters KT, Dotta F, Amirian N, Campbell PD, Kay TW, Atkinson MA, et al. Demonstration of islet-autoreactive CD8 T cells in insulitic lesions from recent onset and long-term type 1 diabetes patients. J Exp Med. 2012 Jan;209(1):51–60.

12. Rodriguez-Calvo T, Sabouri S, Anquetil F, von Herrath MG. The viral paradigm in type 1 diabetes: Who are the main suspects? Autoimmun Rev. 2016 Oct;15(10):964–9.

13. Virtanen SM. Dietary factors in the development of type 1 diabetes. Pediatr Diabetes. 2016 Jul;17 Suppl 2:49–55.

14. Knip M, Siljander H. The role of the intestinal microbiota in type 1 diabetes mellitus. Nat Rev. 2016 Mar;12(3):154–67.

15. Ziegler AG, Rewers M, Simell O, Simell T, Lempainen J, Steck A, et al. Seroconversion to multiple islet autoantibodies and risk of progression to diabetes in children. JAMA. 2013 Jun;309(23):2473–9.

16. Kallionpää H, Elo LL, Laajala E, Mykkänen J, Ricaño-Ponce I, Vaarma M, et al. Innate immune activity is detected prior to seroconversion in children with HLA-conferred type 1 diabetes susceptibility. Diabetes. 2014;63(7):2402–14.

17. Moulder R, Bhosale SD, Erkkilä T, Laajala E, Salmi J, Nguyen E V, et al. Serum Proteomes Distinguish Children Developing Type 1 Diabetes in a Cohort With HLA-Conferred Susceptibility. Diabetes [Internet]. 2015;64(6):2265–78. Available from: http://diabetes.diabetesjournals.org/content/64/6/2265.abstract

18. Peet A, Kool P, Ilonen J, Knip M, Tillmann V, Group DS. Birth weight in newborn infants with different diabetes-associated HLA genotypes in three neighbouring countries: Finland, Estonia and Russian Karelia. Diabetes Metab Res Rev. 2012 Jul;28(5):455–61.

19. Cianciaruso C, Phelps EA, Pasquier M, Hamelin R, Demurtas D, Ahmed MA, et al. Primary Human and Rat beta-Cells Release the Intracellular Autoantigens GAD65, IA-2, and Proinsulin in Exosomes Together With Cytokine-Induced Enhancers of Immunity. Diabetes. 2017 Feb;66(2):460–73.

20. Kim D, Pertea G, Trapnell C, Pimentel H, Kelley R, Salzberg SL. TopHat2: accurate alignment of transcriptomes in the presence of insertions, deletions and gene fusions. Genome Biol. 2013 Apr;14(4):R36-2013-14-4-r36.

21. Anders S, Pyl PT, Huber W. HTSeq-A Python framework to work with high-throughput sequencing data. Bioinformatics. 2015;31(2):166–9.

22. Robinson MD, McCarthy DJ, Smyth GK. edgeR: A Bioconductor package for differential expression analysis of digital gene expression data. Bioinformatics. 2009;26(1):139–40.

23. Dobin A, Davis CA, Schlesinger F, Drenkow J, Zaleski C, Jha S, et al. STAR: ultrafast universal RNA-seq aligner. Bioinformatics. 2013 Jan;29(1):15–21.

24. 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(5):411–20.

25. Qiu X, Mao Q, Tang Y, Wang L, Chawla R, Pliner HA, et al. Reversed graph embedding resolves complex single-cell trajectories. Nat Methods. 2017 Oct;14(10):979–82.

26. Ravassard P, Hazhouz Y, Pechberty S, Bricout-Neveu E, Armanet M, Czernichow P, et al. A genetically engineered human pancreatic beta cell line exhibiting glucose-inducible insulin secretion. J Clin Invest. 2011 Sep;121(9):3589–97.

27. Anagandula M, Richardson SJ, Oberste MS, Sioofy-Khojine AB, Hyoty H, Morgan NG, et al. Infection of human islets of Langerhans with two strains of Coxsackie B virus serotype 1: assessment of virus replication, degree of cell death and induction of genes involved in the innate immunity pathway. J Med Virol. 2014 Aug;86(8):1402–11.

28. Reynier F, Pachot A, Paye M, Xu Q, Turrel-Davin F, Petit F, et al. Specific gene expression signature associated with development of

autoimmune type-I diabetes using whole-blood microarray analysis. Genes Immun. 2010 Apr;11(3):269–78.

29. Ferreira RC, Guo H, Coulson RM, Smyth DJ, Pekalski ML, Burren OS, et al. A type I interferon transcriptional signature precedes autoimmunity in children genetically at risk for type 1 diabetes. Diabetes. 2014 Jul;63(7):2538–50.

30. Jin Y, Sharma A, Bai S, Davis C, Liu H, Hopkins D, et al. Risk of type 1 diabetes progression in islet autoantibody-positive children can be further stratified using expression patterns of multiple genes implicated in peripheral blood lymphocyte activation and function. Diabetes. 2014 Jul;63(7):2506–15.

31. Reinert-Hartwall L, Honkanen J, Salo HM, Nieminen JK, Luopajärvi K, Härkönen T, et al. Th1/Th17 plasticity is a marker of advanced β cell autoimmunity and impaired glucose tolerance in humans. J Immunol [Internet]. 2015;194(1):68–75. Available from: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=4273995&tool=pmcentrez&rendertype=abstract

32. Heninger AK, Eugster A, Kuehn D, Buettner F, Kuhn M, Lindner A, et al. A divergent population of autoantigen-responsive CD4(+) T cells in infants prior to beta cell autoimmunity. Sci Transl Med. 2017 Feb;9(378):10.1126/scitranslmed.aaf8848.

33. Swafford AD, Howson JM, Davison LJ, Wallace C, Smyth DJ, Schuilenburg H, et al. An allele of IKZF1 (Ikaros) conferring susceptibility to childhood acute lymphoblastic leukemia protects against type 1 diabetes. Diabetes. 2011 Mar;60(3):1041–4.

34. Ram R, Mehta M, Nguyen QT, Larma I, Boehm BO, Pociot F, et al. Systematic Evaluation of Genes and Genetic Variants Associated with Type 1 Diabetes Susceptibility. J Immunol (Baltimore, Md 1950). 2016 Apr;196(7):3043–53.

35. Arsenio J, Kakaradov B, Metz PJ, Kim SH, Yeo GW, Chang JT. Early specification of CD8+ T lymphocyte fates during adaptive immunity revealed by single-cell gene-expression analyses. Nat Immunol. 2014 Apr;15(4):365–72.

36. Dominguez CX, Amezquita RA, Guan T, Marshall HD, Joshi NS, Kleinstein SH, et al. The transcription factors ZEB2 and T-bet cooperate to program cytotoxic T cell terminal differentiation in response to LCMV viral infection. J Exp Med. 2015 Nov;212(12):2041–56.

37. Nishida A, Andoh A, Inatomi O, Fujiyama Y. Interleukin-32 expression in the pancreas. J Biol Chem. 2009 Jun;284(26):17868–76.

38. Dotta F, Censini S, van Halteren AG, Marselli L, Masini M, Dionisi S, et al. Coxsackie B4 virus infection of beta cells and natural killer cell insulitis in recent-onset type 1 diabetic patients. Proc Natl Acad Sci U S A. 2007 Mar;104(12):5115–20.

39. Krogvold L, Edwin B, Buanes T, Frisk G, Skog O, Anagandula M, et al. Detection of a low-grade enteroviral infection in the islets of langerhans of living patients newly diagnosed with type 1 diabetes. Diabetes. 2015 May;64(5):1682–7.

40. Laitinen OH, Honkanen H, Pakkanen O, Oikarinen S, Hankaniemi MM, Huhtala H, et al. Coxsackievirus B1 is associated with induction of

beta-cell autoimmunity that portends type 1 diabetes. Diabetes. 2014 Feb;63(2):446–55.

41. Richardson SJ, Willcox A, Bone AJ, Foulis AK, Morgan NG. The prevalence of enteroviral capsid protein vp1 immunostaining in pancreatic islets in human type 1 diabetes. Diabetologia. 2009 Jun;52(6):1143–51.

42. Ylipaasto P, Klingel K, Lindberg AM, Otonkoski T, Kandolf R, Hovi T, et al. Enterovirus infection in human pancreatic islet cells, islet tropism in vivo and receptor involvement in cultured islet beta cells. Diabetologia. 2004 Feb;47(2):225–39.

43. Oikarinen S, Tauriainen S, Hober D, Lucas B, Vazeou A, Sioofy-Khojine A, et al. Virus antibody survey in different European populations indicates risk association between coxsackievirus B1 and type 1 diabetes. Diabetes. 2014 Feb;63(2):655–62.

44. Shields BM, Mcdonald TJ, Oram R, Hill A, Hudson M, Leete P, et al. C-Peptide Decline in Type 1 Diabetes Has Two Phases : An Initial Exponential Fall and a Subsequent Stable Phase. Diabetes Care. 2018;41(July):1486–92.

45. Kim SH, Han SY, Azam T, Yoon DY, Dinarello CA. Interleukin-32: a cytokine and inducer of TNFalpha. Immunity. 2005 Jan;22(1):131–42.

46. Orban T, Kis J, Szereday L, Engelmann P, Farkas K, Jalahej H, et al. Reduced CD4+ T-cell-specific gene expression in human type 1 diabetes mellitus. J Autoimmun. 2007 Jun;28(4):177–87.

47. Joosten LA, Netea MG, Kim SH, Yoon DY, Oppers-Walgreen B, Radstake TR, et al. IL-32, a proinflammatory cytokine in rheumatoid arthritis. Proc Natl Acad Sci U S A. 2006 Feb;103(9):3298–303.

48. Shioya M, Nishida A, Yagi Y, Ogawa A, Tsujikawa T, Kim-Mitsuyama S, et al. Epithelial overexpression of interleukin-32alpha in inflammatory bowel disease. Clin Exp Immunol. 2007 Sep;149(3):480–6.

49. Na SJ, So SH, Lee KO, Choi YC. Elevated serum level of interleukin-32alpha in the patients with myasthenia gravis. J Neurol. 2011 Oct;258(10):1865–70.

50. Walker LS, von Herrath M. CD4 T cell differentiation in type 1 diabetes. Clin Exp Immunol. 2016 Jan;183(1):16–29.

51. Jhun H, Choi J, Hong J, Lee S, Kwak A, Kim E, et al. IL-32gamma overexpression accelerates streptozotocin (STZ)-induced type 1 diabetes. Cytokine. 2014 Sep;69(1):1–5.

52. Xin T, Che M, Duan L, Xu Y, Gao P. Interleukin-32: its role in asthma andpotential as a therapeutic agent. Respir Res. 2018;19:124.

53. Viken MK, Blomhoff A, Olsson M, Akselsen HE, Pociot F, Nerup J, et al. Reproducible association with type 1 diabetes in the extended class I region of the major histocompatibility complex. Genes Immun. 2009 Jun;10(4):323–33.

54. Alvarez JI, Kebir H, Cheslow L, Chabarati M, Larochelle C, Prat A. JAML mediates monocyte and CD8 T cell migration across the brain endothelium. Ann Clin Transl Neurol. 2015 Sep;2(11):1032–7.

55. Hodik M, Anagandula M, Fuxe J, Krogvold L, Dahl-Jorgensen K, Hyoty H, et al. Coxsackie-adenovirus receptor expression is enhanced in pancreas from patients with type 1 diabetes. BMJ open diabetes Res care. 2016 Nov;4(1):e000219.

56. Leete P, Willcox A, Krogvold L, Dahl-Jorgensen K, Foulis AK, Richardson SJ, et al. Differential Insulitic Profiles Determine the Extent of beta-Cell Destruction and the Age at Onset of Type 1 Diabetes. Diabetes. 2016 May;65(5):1362–9.

57. von Toerne C, Laimighofer M, Achenbach P, Beyerlein A, de Las Heras Gala T, Krumsiek J, et al. Peptide serum markers in islet autoantibody-positive children. Diabetologia. 2017 Feb;60(2):287–95.

58. Achenbach P, Hummel M, Thümer L, Boerschmann H, Höfelmann D, Ziegler AG. Characteristics of rapid vs slow progression to type 1 diabetes in multiple islet autoantibody-positive children. Diabetologia. 2013;56(7):1615–22.

FIGURE LEGENDS

Figure 1. **Fractionation of PBMC sample into CD4+, CD8+ and CD4-CD8- cellular subsets reveals distinct and overlapping gene expression signatures. A)** Outline of the sample collection and cell fractionation. **B)** tSNE (t-Distributed Stochastic Neighbor Embedding) visualization of the log2-transformed expression data (without any filtering steps) coloured according to cell fraction information. **C)** Number of DE genes, when CD4+, CD8+ and CD4-CD8- fractionated samples were compared to their original PBMC aliquots. The functionally important fraction-specific upregulated genes are highlighted in red. Analysis was restricted to healthy Controls only. For the gene lists, see **Supplementary Table 3**.

Figure 2. **RNA-seq analysis identifies transcriptomic changes associated with beta cell autoimmunity. A)** Number and overlap of DE genes between Cases and Controls identified in cell fractions analysed. Genes shared between all four fractions are highlighted. **B)** Heatmap of the genes DE in CD4+ T cells between the Cases and Controls. Values are presented as $\log_2$FC (truncated between [-2, 2]) between each Case-Control pair at each timepoint (3–36 months) and standardized to the mean of each gene. Genes co-regulated with *IL32* (< 2.5 Euclidean distance) are marked with red box and text. Additional information about the samples is marked on top of the heatmap. ‘Before/After SC’ informs whether the Case-sample was collected before (Before SC) or after seroconversion (After SC). ‘Pair Info’ provides the case-control pair information. The ‘SC / T1D’ annotation indicates whether the Case has progressed to clinical T1D diagnosis (T1D) or not (SC). **C)** Number and overlap of *IL32* co-clustered genes in indicated cell fractions. Genes regulated at least in two fractions are highlighted. **D)** Profiles of *IL32*, *AMICA1* and *BNT3A2* in CD4+ samples, presented in $\log_2$ RPKM scale. For individual profiles, see **Supplementary Fig. 4**. The Case-Control pairs are grouped according to the diagnosis of the Cases. T1D= Case has been diagnosed with clinical T1D, SC=Case has seroconverted to autoantibody positivity.

Figure 3. scRNA-seq of PBMCs identifies T and NK cells as *IL32* high populations

A) Expression of *IL32γ* isoform in longitudinal PBMC samples of Cases and their Controls (n=7+7), assayed by qRT-PCR. For alpha and beta isoforms, please see **Supplementary Fig. 7. B)** tSNE clusters from the pooled data from all scRNA-seq samples (4 Cases and 4 Controls, in total 18 396 cells). Clusters are named according to the expression of classical marker genes, such as *CD8A* (for details and marker gene list, please see **Supplementary Fig. 8**; for contribution of each sample per cluster, please refer to **Supplementary Fig. 9** and **10**. **C)** Expression of *IL32* in the 12 cell clusters (natural logarithm transformation with addition of 1). For Case-Control comparison, please see **Supplementary Fig. 11**. **D-F)** trajectories emerging when using the data from CD4+ cells and the precursor cells, as well as **G-I)** from CD8+ and the precursor cells. Here, precursor cells refer to cells from the naive and RGCC+ T cell clusters. For the trajectory analysis of all the cells from all clusters as well as the breakdown of each individual cluster, please see **Supplementary Fig. 12**. In **D)** and **G)**, cells are coloured based on the contributions from different tSNE clusters. In **E)** and **H),** cells are coloured by the Case (orange) or Control (grey) status. In **F)** and **I),** cells are coloured by the intensity of IL32 expression ($\log_{10}$ transformation with addition of 0.1).

Figure 4. Virus- and cytokine induce *IL32* expression by pancreatic beta cells

A) Representative FACS dot plots showing IFN-γ and IL-32 double staining in Th0 and Th1 polarized CD4+ cells. Staining controls and two other replicates are shown in **Supplementary Fig. 13A.** Percent IL-32 positive cells as well as Median Fluorescence Intensity (MFI) data (mean+/- SEM) from all the three replicates are shown in **B) and C),** respectively. Statistical significance was determined by paired two tailed t-test. **D)** IL-32 secretion in culture supernatant as measured by ELISA. Cells were cultured in Th0/1 condition for 72 h in the presence (+) or absence (-) of anti-IFNγ. The expression plotted is relative to Th0 (-). Statistical significance was determined by paired two tailed t-test. **E)** *IL32* expression in non-polarized Th0 cells and cells differentiated to Th1 for 72h in the presence (+) or absence (-) of IL-32γ as measured by the Taqman assay. The expression is calculated relative to *EEF1A*. Statistical significance was determined by unpaired two tailed t-test. **F)** IL-32 secretion in culture supernatant as measured by ELISA. Cells were cultured in Th0/1 condition for 7 days in the presence (+) or absence (-) of

IL-32γ, followed by washing and re-stimulation by PMA and ionomycin for 48 h. The expression plotted is relative to Th0 (-). Statistical significance was determined by paired two tailed t-test. **G)** Expression of the *TNFA* and *IL6* or *IL8* and *IL32* genes when the EndoC-βH1 cells were stimulated with IL32γ alone or in combination with other inflammatory cytokines for 24 h. The fold-change is calculated compared to non-treated (NT) cells. The results shown here are from four independent biological replicates (mean +/- SEM). Statistical significance was determined by paired two tailed t-test. **H)** *IL32* expression as measured in an RNA-seq experiment where pancreatic islets were infected with CBV1-7. Statistical significance was determined by EdgeR. **I)** *IL32* expression in virus infected pancreatic islets as measured by RT-qPCR Taqman assay. The expression is calculated as 2^-(dCt). The statistical significance is determined by paired two-tailed t-test. *= p-value <0.05, ** = p-value <0.01, and *** = FDR<0.001.

TABLES

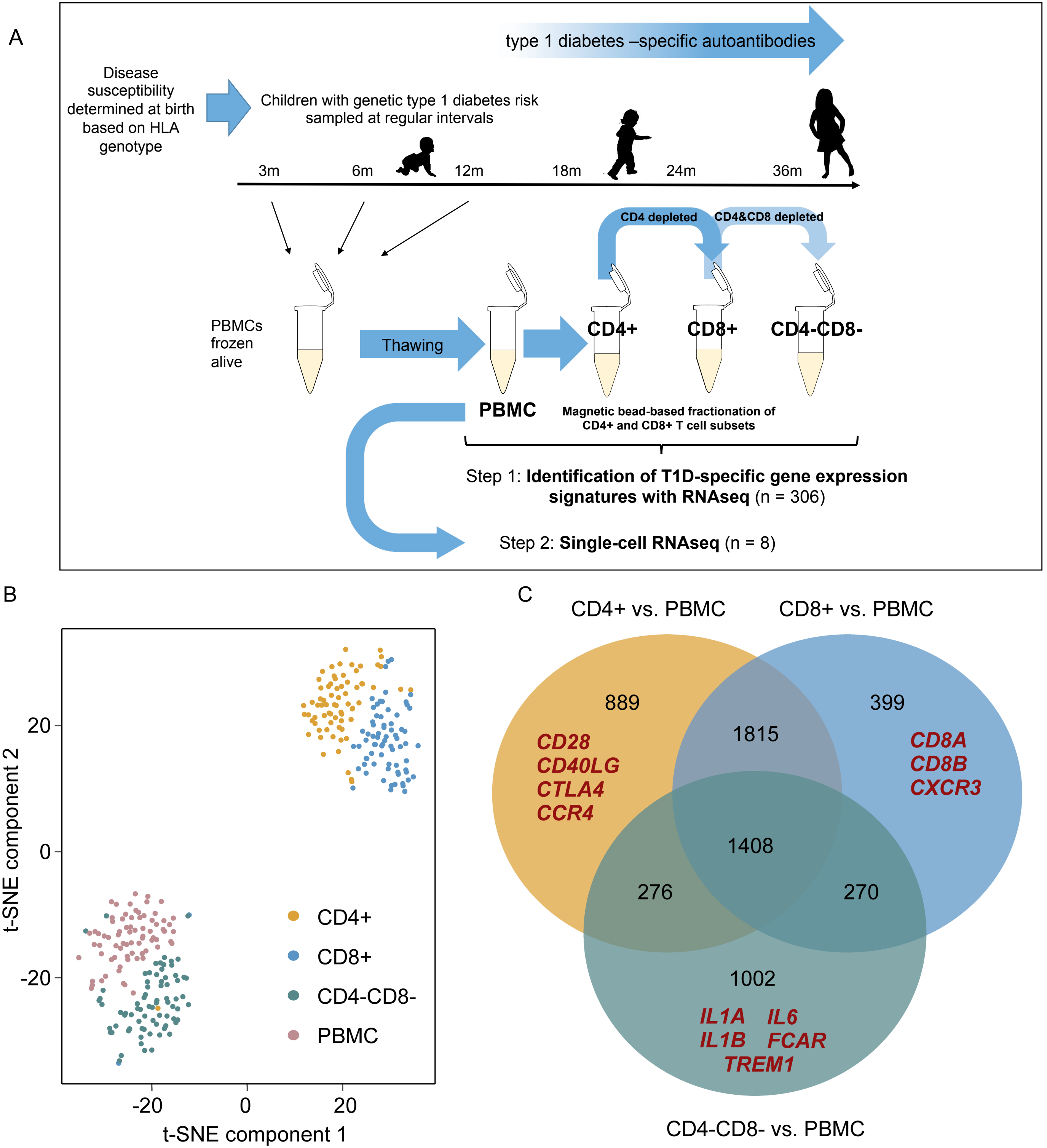
Table 1. Summary of the Case and Control children sampled at the age of 3–36 months.

Case #	Gender	Seroconversion* age	First autoantibodies	Age at T1D diagnosis	Matched control #
Case 1	Female	12 mo	IAA, GADA	3.2 y	Control 1
Case 2	Male	12 mo	IAA	-	Control 2
Case 3	Male	18 mo	IAA, ICA	3.7 y	Control 3
Case 5	Female	24 mo	IAA, IA-2A, ZnT8A, ICA	2.6 y	Control 5

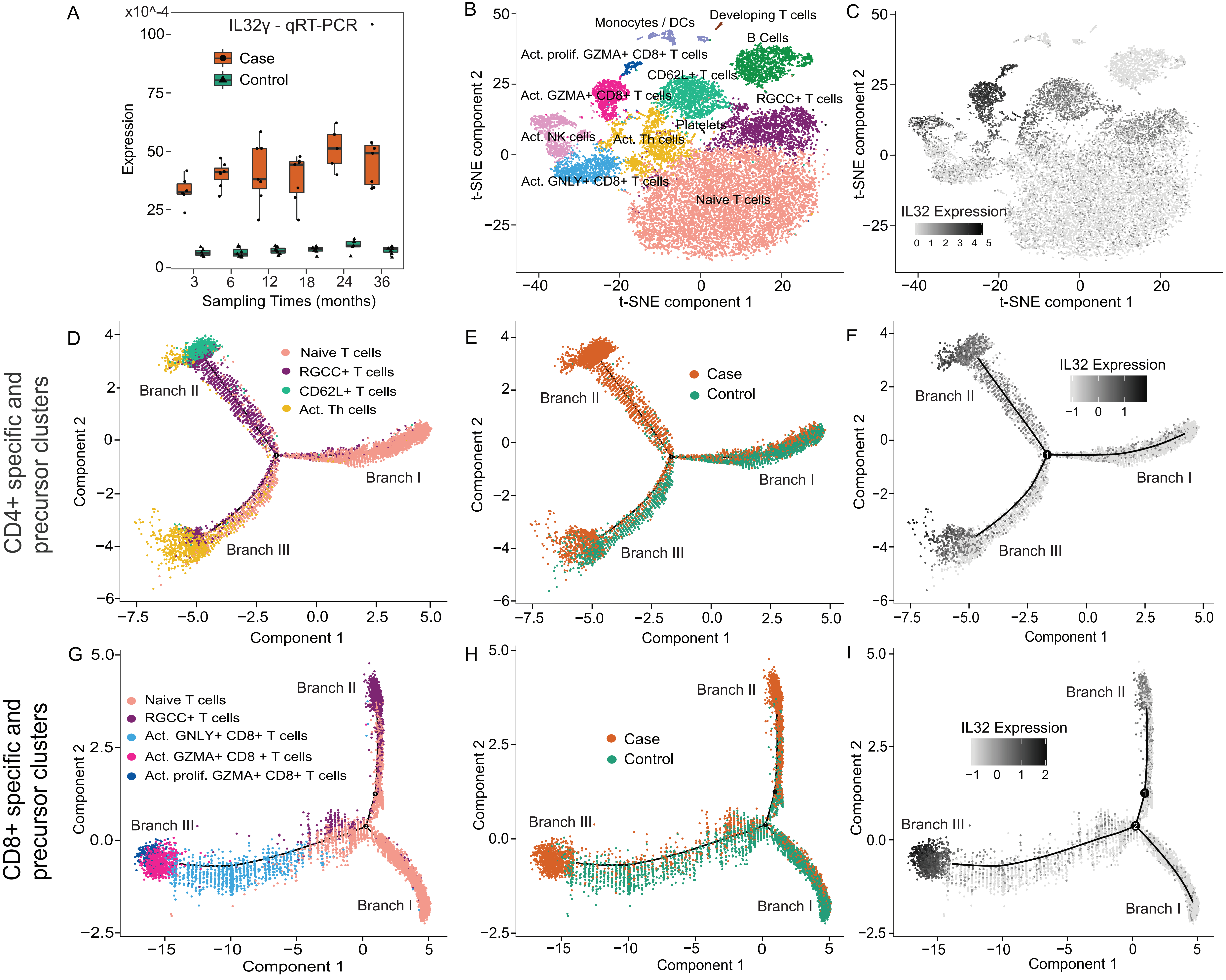
Case 9	Male	18 mo	IAA, GADA, ICA	-	Control 9
Case 10	Male	12 mo	IAA, GADA	-	Control 10.1 Control 10.2
Case 11	Female	18 mo	GADA	2.4 y	Control 11

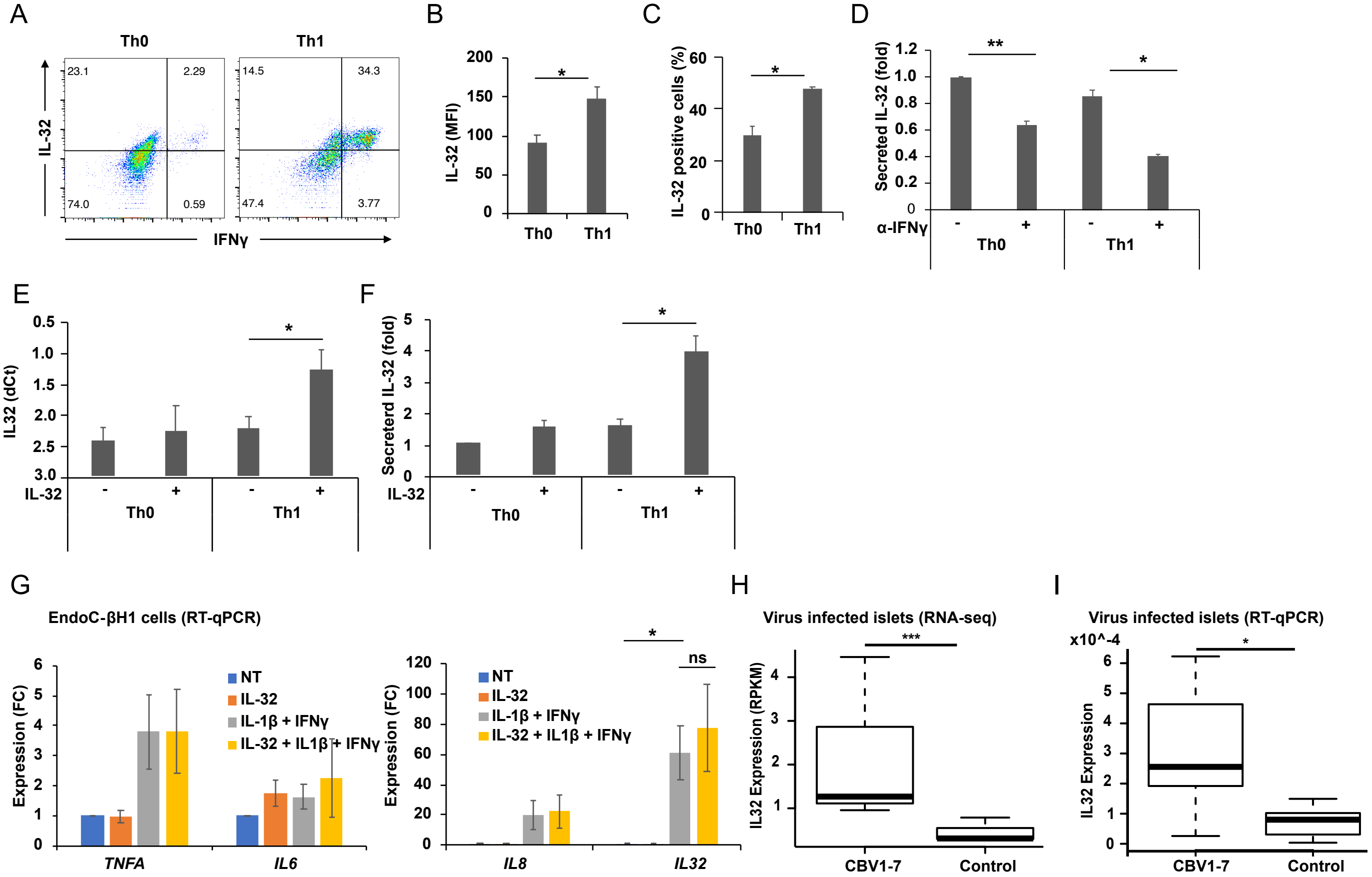
*First detection of T1D-associated autoantibodies.

For further details, see **Supplementary Table 1**.


A
type 1 diabetes –specific autoantibodies
Disease susceptibility determined at birth based on HLA genotype
Children with genetic type 1 diabetes risk sampled at regular intervals
3m
6m
12m
18m
24m
36m
CD4 depleted
CD4&CD8 depleted
PBMCs frozen alive
Thawing
CD4+
CD8+
CD4-CD8-
PBMC
Magnetic bead-based fractionation of CD4+ and CD8+ T cell subsets
Step 1: Identification of T1D-specific gene expression signatures with RNAseq (n = 306)
Step 2: Single-cell RNAseq (n = 8)
B
t-SNE component 2
20
0
-20
-20
0
20
t-SNE component 1
CD4+
CD8+
CD4-CD8-
PBMC
C
CD4+ vs. PBMC
CD8+ vs. PBMC
889
1815
399
CD28
CD40LG
CTLA4
CCR4
CD8A
CD8B
CXCR3
1408
276
270
1002
IL1A
IL6
IL1B
FCAR
TREM1
CD4-CD8- vs. PBMC

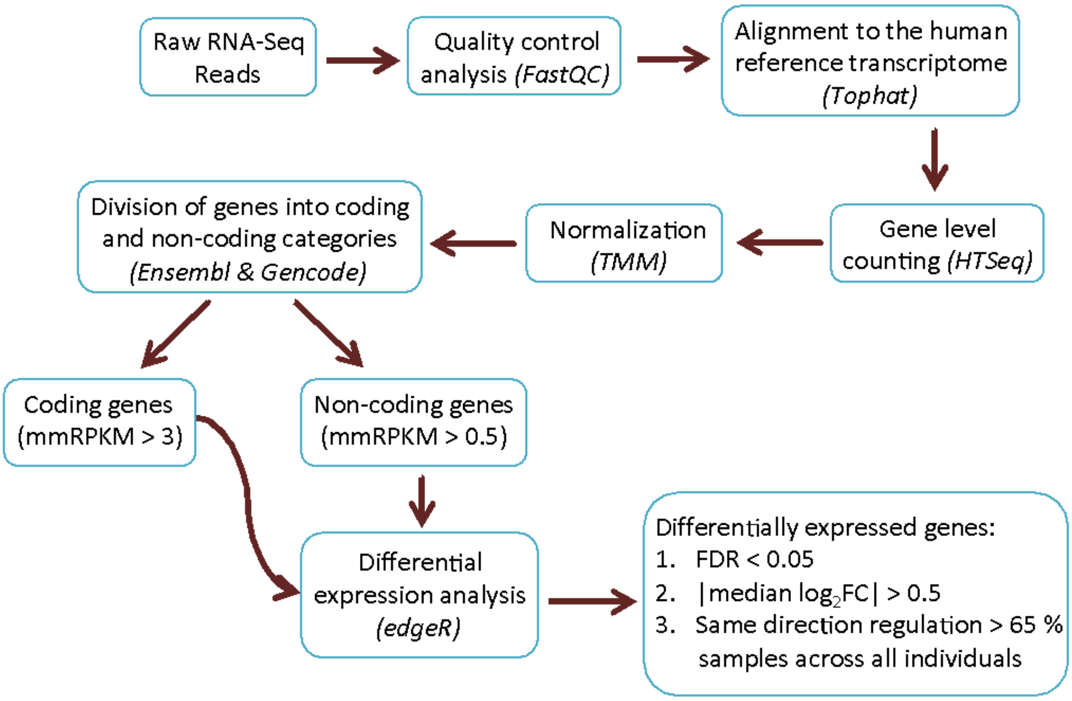
A
CD4+
CD8+
CD4-CD8-
PBMC
26
48
109
46
5
0
19
2
2
6
3
3
2
3
4
AMICA1
BTN3A2
IL32
RPSAP15
RPSAP58
WASH7P
C
CD4+
CD8+
CD4-CD8-
PBMC
7
1
45
39
0
0
4
0
1
2
2
1
1
0
2
IL32
BTN3A2
UROS
WASH7P
AMICA1
TRBV4-1
TMEM14C
LARS
RSU1
BTN3A3
CARD8
CCDC167
LINC01184
D
CD4+ cells: IL32 + AMICA1 + BTN3A2
T1D - Pair 1
T1D - Pair 3
T1D - Pair 5
T1D - Pair 11
SC - Pair 2
SC - Pair 9
SC - Pair 10
log2 (RPKM Values)
6
4
2
-12
0
12
24
Seroconversion-Centered Times (months)
Gene Name
IL32
AMICA1
BTN3A2
Group
Case
Control
B
Before/After SC
Pair Info
SC / T1D
FHL1
CTD−3092A11.1
RP11−632K20.2
XXyac−YRM2039.2
ENSG00000215513
LINC00665
SH2D3C
CTA−29F11.1
NAPRT1
DIP2A−IT1
SNX25
GDPD5
ABCA1
LZTS3
ABCG1
HES4
RP11−589C21.6
GADD45G
SPRY1
THBS1
TYMP
APOBEC3A
FAM26F
RPSAP58
RPSAP15
WASH7P
AP006222.2
TMEM8A
RP11−706O15.3
RP11−747H7.3
PRKXP1
AC092580.4
PDPR
CPT1A
GZMA
IL32
TRBV4−1
LINC01259
H1F0
CTA−445C9.15
SH3BP1
ENSG00000269996
LARS
RGS14
C17orf62
BTN3A2
TMEM14C
CD52
AMICA1
DGKQ
UROS
Pair3_18months
Pair 2_36months
Pair1_6months
Pair9_12months
Pair5_24months
Pair5_36months
Pair10_12months
Pair3_36months
Pair9_3months
Pair9_24months
Pair9_6months
Pair9_18months
Pair2_3months
Pair2_6months
Pair2_12months
Pair2_18months
Pair2_24months
Pair3_3months
Pair3_6months
Pair3_12months
Pair11_3months
Pair9_36months
Pair11_24months
Pair11_36months
Pair10_36months
Pair10_6months
Pair10_18months
Pair1_18months
Pair1_24months
Pair1_36months
Pair1_3months
Pair1_12months
Pair5_3months
Pair5_6months
Pair5_12months
Pair5_18months
2
1
0
−1
−2
Before/After SC
Before SC
After SC
Pair Info
Pair 1
Pair 2
Pair 3
Pair 5
Pair 9
Pair 10
Pair 11
SC / T1D
SC
T1D


A
IL32γ - qRT-PCR
x10^-4
Case
Control
Expression
100
75
50
25
0
3
6
12
18
24
36
Sampling Times (months)
B
Monocytes / DCs
Developing T cells
B Cells
Act. prolif. GZMA+ CD8+ T cells
CD62L+ T cells
Act. GZMA+ CD8+ T cells
RGCC+ T cells
Platelets
Act. NK cells
Act. Th cells
Act. GNLY+ CD8+ T cells
Naive T cells
t-SNE component 2
t-SNE component 1
50
25
0
−25
−40
−20
0
20
C
IL32 Expression
0 1 2 3 4 5
t-SNE component 2
t-SNE component 1
CD4+ specific and precursor clusters
D
Naive T cells
RGCC+ T cells
CD62L+ T cells
Act. Th cells
Branch I
Branch II
Branch III
Component 2
Component 1
4
2
0
−2
−4
−6
−7.5
−5.0
−2.5
0.0
2.5
5.0
E
Case
Control
Branch I
Branch II
Branch III
Component 2
Component 1
F
IL32 Expression
−1 0 1
Branch I
Branch II
Branch III
Component 2
Component 1
CD8+ specific and precursor clusters
G
Naive T cells
RGCC+ T cells
Act. GNLY+ CD8+ T cells
Act. GZMA+ CD8 + T cells
Act. prolif. GZMA+ CD8+ T cells
Branch I
Branch II
Branch III
Component 2
Component 1
5.0
2.5
0.0
−2.5
−15
−10
−5
0
5
H
Case
Control
Branch I
Branch II
Branch III
Component 2
Component 1
I
IL32 Expression
−1 0 1 2
Branch I
Branch II
Branch III
Component 2
Component 1


A
Th0
Th1
IL-32
IFNγ
23.1
2.29
74.0
0.59
14.5
34.3
47.4
3.77
B
IL-32 (MFI)
200
150
100
50
0
*
Th0
Th1
C
IL-32 positive cells (%)
60
40
20
0
*
Th0
Th1
D
Secreted IL-32 (fold)
1.2
1.0
0.8
0.6
0.4
0.2
0
**
*
α-IFNγ
-
+
-
+
Th0
Th1
E
IL32 (dCt)
0.5
1.0
1.5
2.0
2.5
3.0
*
IL-32
-
+
-
+
Th0
Th1
F
Secreterd IL-32 (fold)
5
4
3
2
1
0
*
IL-32
-
+
-
+
Th0
Th1
G
EndoC-βH1 cells (RT-qPCR)
Expression (FC)
6
5
4
3
2
1
0
NT
IL-32
IL-1β + IFNγ
IL-32 + IL1β + IFNγ
TNFA
IL6
Expression (FC)
120
100
80
60
40
20
0
*
ns
NT
IL-32
IL-1β + IFNγ
IL-32 + IL1β + IFNγ
IL8
IL32
H
Virus infected islets (RNA-seq)
IL32 Expression (RPKM)
4
3
2
1
***
CBV1-7
Control
I
Virus infected islets (RT-qPCR)
x10^-4
IL32 Expression
6
5
4
3
2
1
0
*
CBV1-7
Control

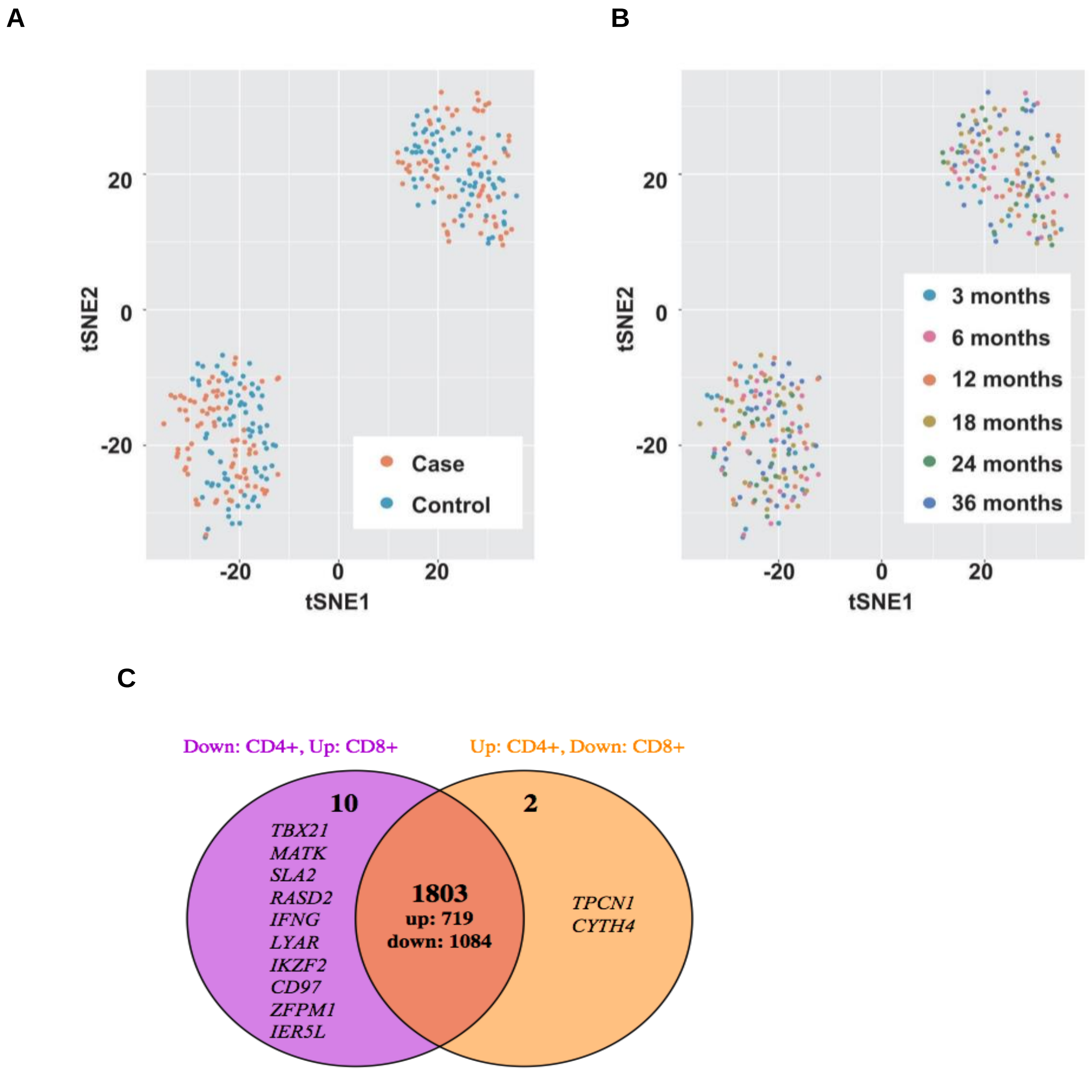
Supplementary Figures

Early detection of peripheral blood cell signature in children developing beta cell autoimmunity at a young age


Raw RNA-Seq Reads
Quality control analysis (FastQC)
Alignment to the human reference transcriptome (Tophat)
Gene level counting (HTSeq)
Normalization (TMM)
Division of genes into coding and non-coding categories (Ensembl & Gencode)
Coding genes (mmRPKM > 3)
Non-coding genes (mmRPKM > 0.5)
Differential expression analysis (edgeR)
Differentially expressed genes:
1. FDR < 0.05
2. |median log2FC| > 0.5
3. Same direction regulation > 65 % samples across all individuals

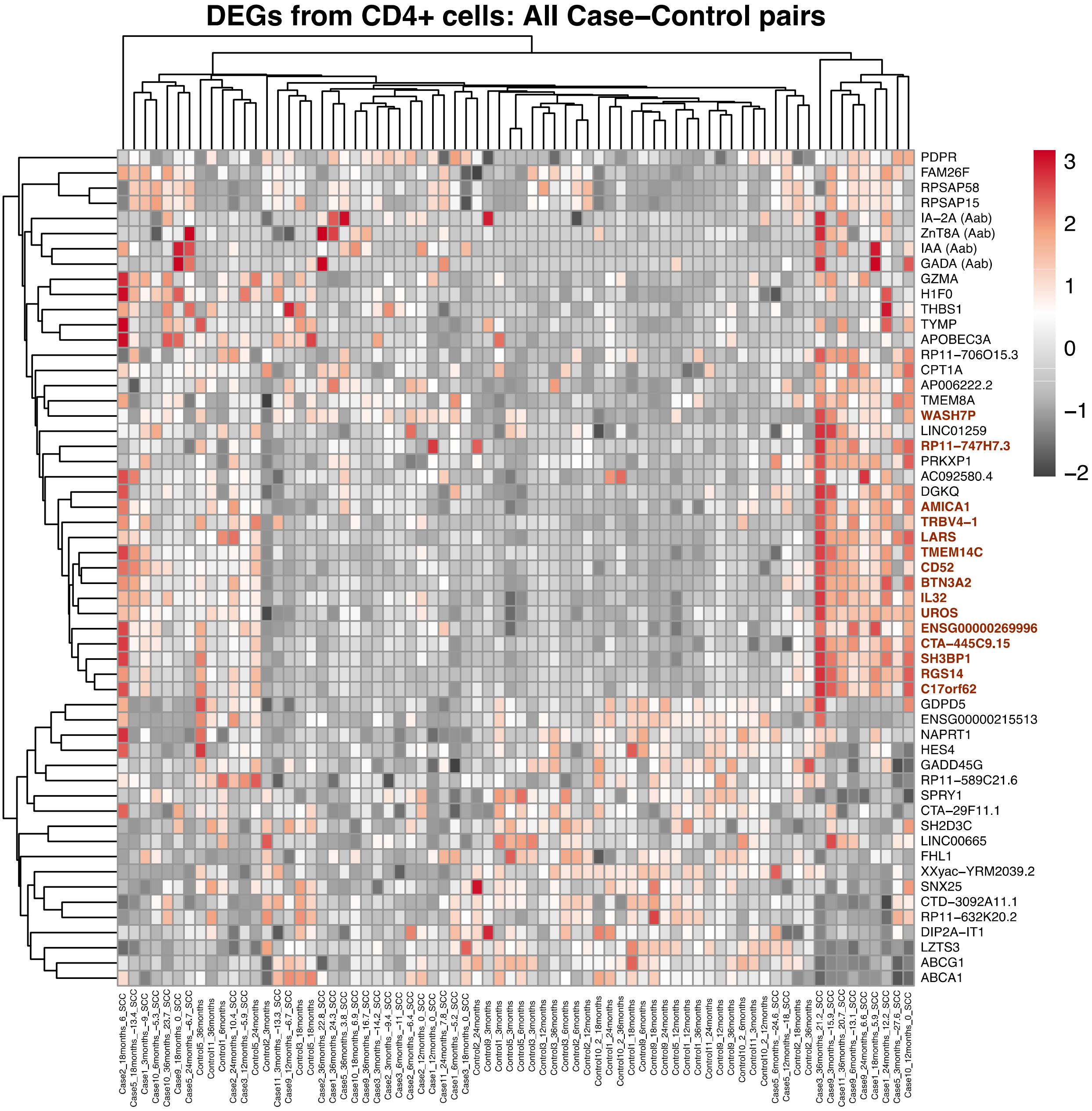
Supplementary Figure 1. Related to Figure 2.

Flow chart depicting the steps taken in the differential expression analyses of the RNA-seq data in this study.



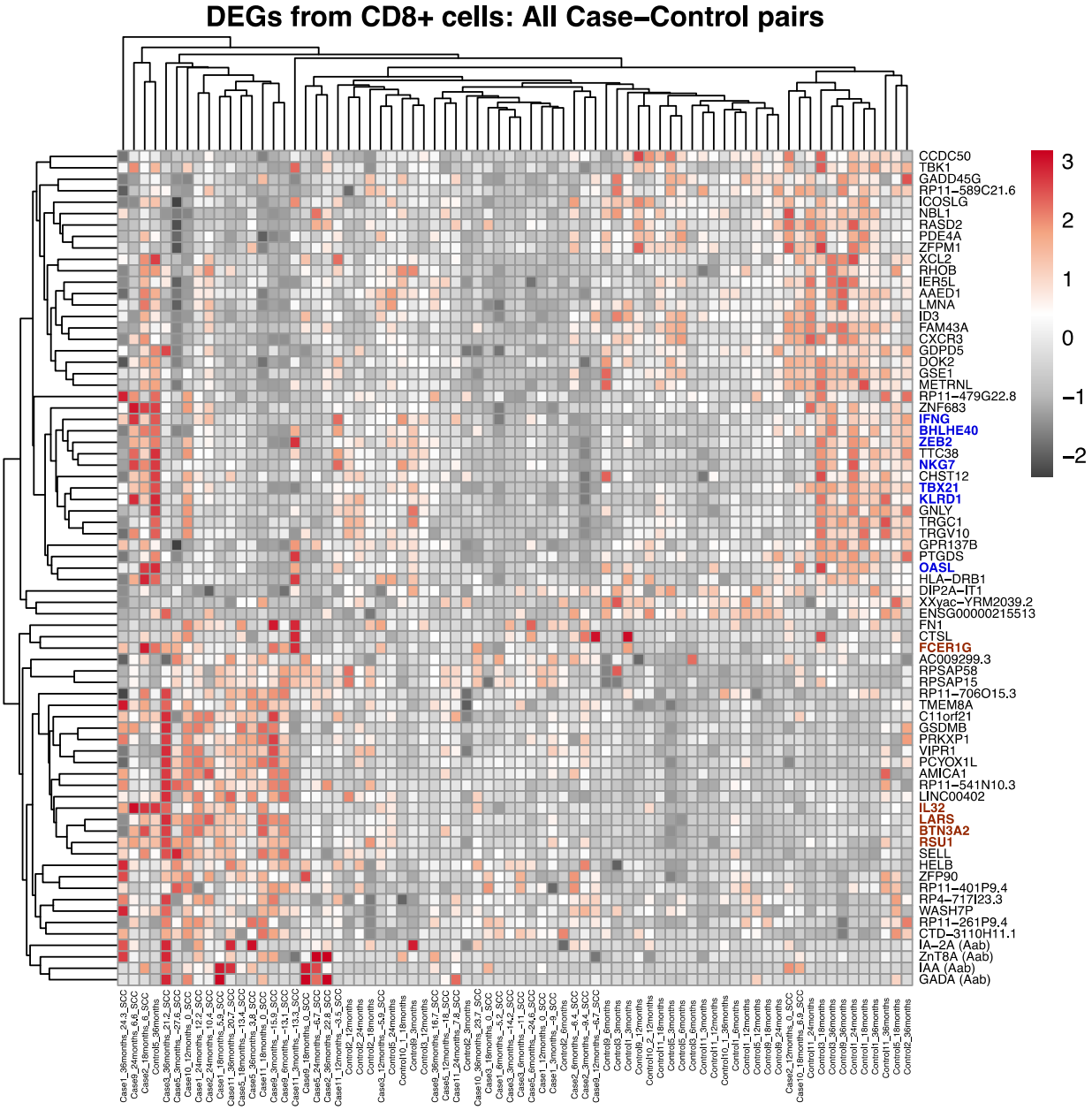
Supplementary Figure 2. Related to Figure 1.

A-B) tSNE (t-Distributed Stochastic Neighbor Embedding) visualization of the log2-transformed expression data from all cell fractions and all genes. **Figure 1** was colored according to cell fractions, and here the colouring of the samples is done according to **A)** Case and Control status and **B)** age at sample collection. For further sample information, see **Table 1** and **Supplementary Table 1**. **C)** Venn diagram expanding on the 1815 genes found DE in both CD4+ vs PBMC and CD8+ vs PBMC analyses (**Figure 1C**). Here, the intersection represents the genes regulated in the same direction. For full lists of genes, see **Supplementary Table 2.**



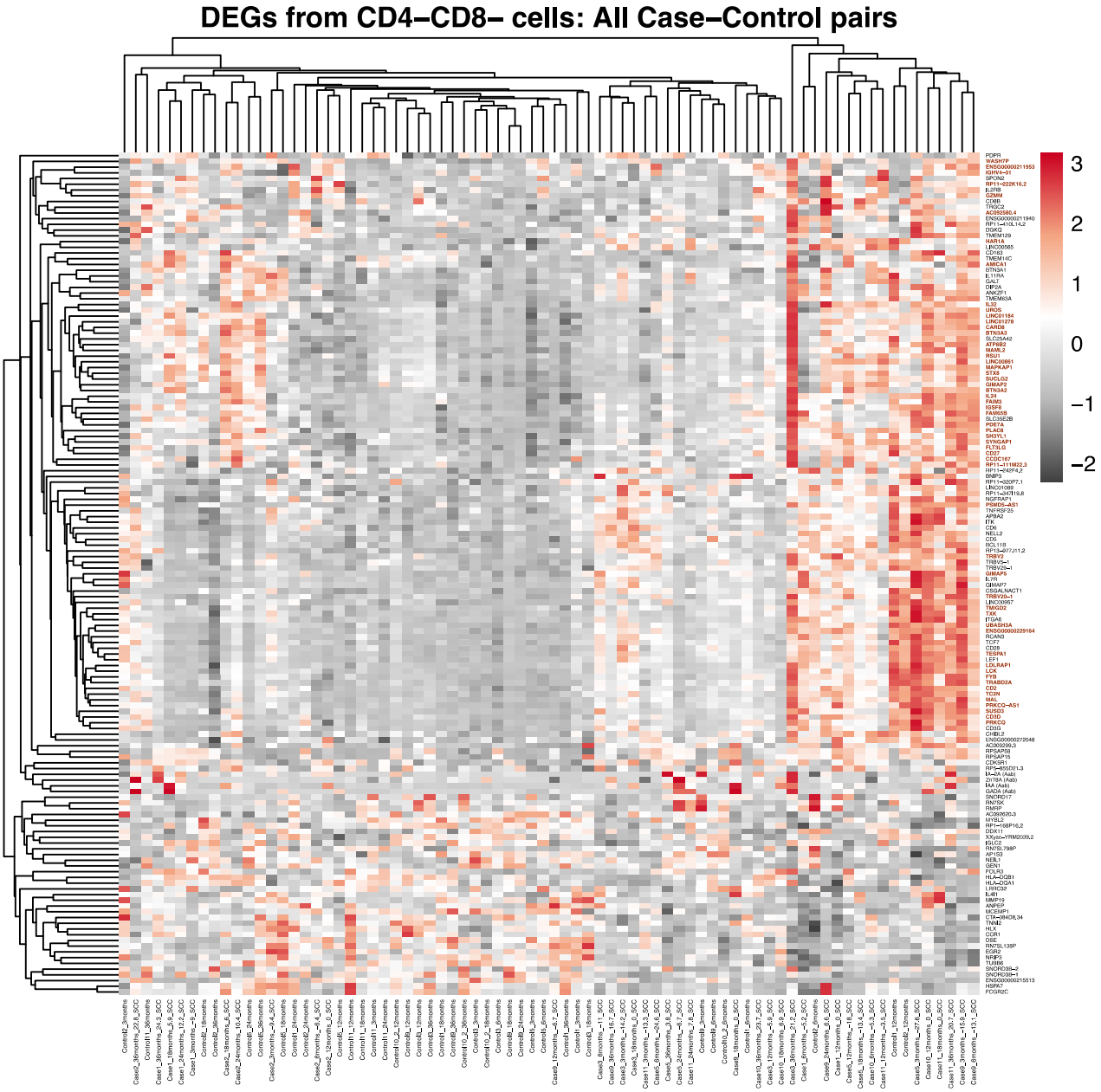
Supplementary Figure 3A. Related to Figure 2.

Hierarchical clustering of the levels of standardized autoantibodies (IAA, IA-2A, ZnT8A, and GADA) and the 51 differentially expressed (DE) genes between the Cases and Controls detected in the CD4+ fraction. Each gene's expression was standardized across samples from each case-control pair individually. Genes with an Euclidean distance (ED)< 2.5 to IL-32 (co-clustering results from k-means clustering) are marked with red text (Supplementary table 4). The samples labels along the x-axis include the sample number, case/control indicator, age of sampling in months, and months to (negative no. of months) or from (positive no. of months) seroconversion time. Here, SCC stands for seroconversion-centered, which is why the months to/from seroconversion are negative or positive.



Supplementary Figure 3B. Related to Figure 2.

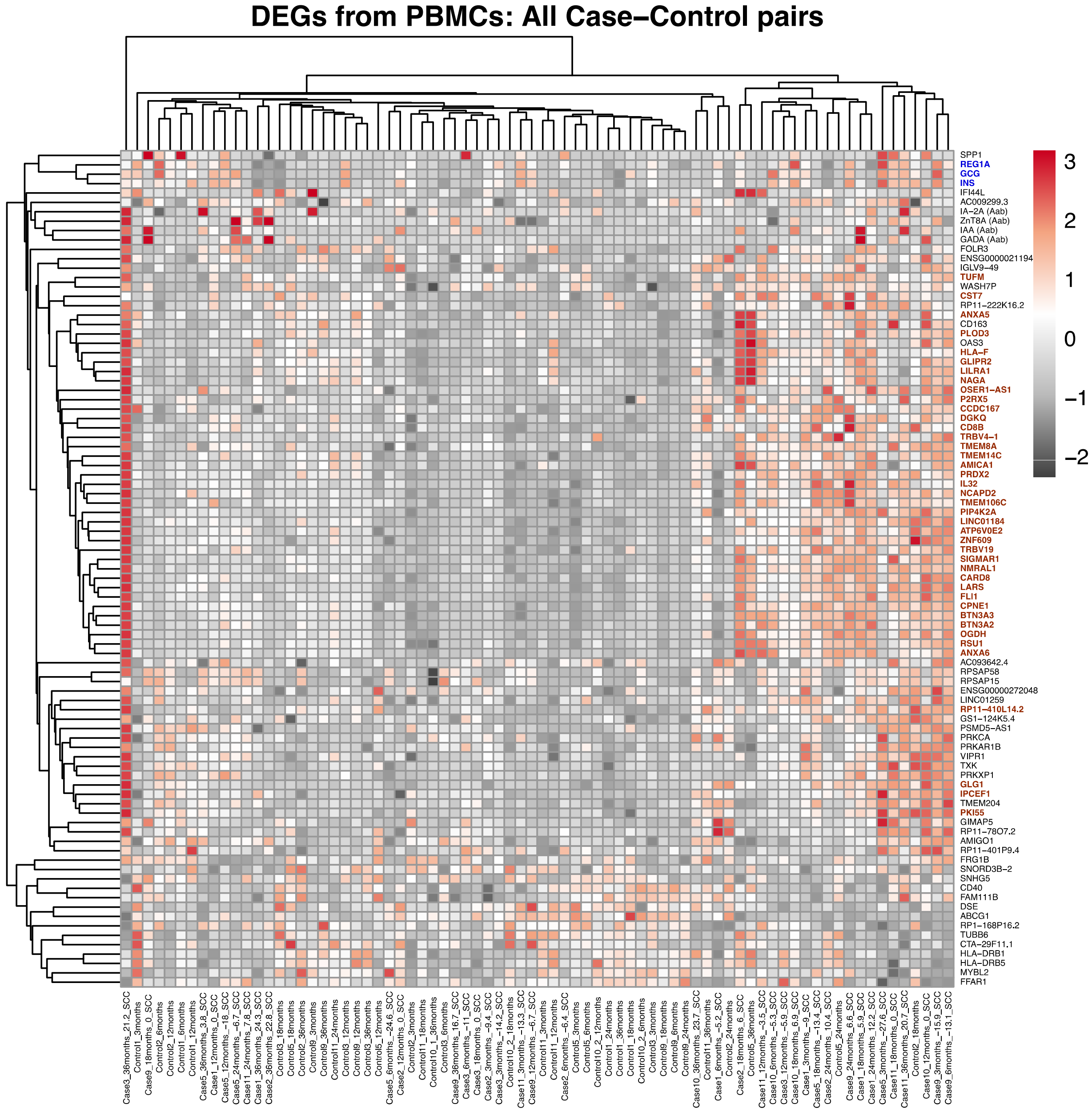
Hierarchical clustering of the levels of standardized autoantibodies (IAA, IA-2A, ZnT8A, and GADA) and the 69 DE genes between the Cases and Controls detected in the CD8+ fraction. Each gene's expression was standardized across samples from each case-control pair individually. Genes with an Euclidean distance (ED) < 2.5 to IL-32 (co-clustering results from k-means clustering) are marked with red text and those with ED < 2.5 to IFNG are marked with blue text (Supplementary table 4). The samples labels along the x-axis include the sample number, case/control indicator, age of sampling in months, and months to (negative no. of months) or from (positive no. of months) seroconversion time. Here, SCC stands for seroconversion-centered, which is why the months to/from seroconversion are negative or positive.



DEGs from CD4–CD8– cells: All Case–Control pairs

Supplementary Figure 3C. Related to Figure 2.

Hierarchical clustering of the levels of standardized autoantibodies (IAA, IA-2A, ZnT8A, and GADA) and the 143 DE genes between the Cases and Controls detected in the CD4-CD8- fraction. The expression of each gene was standardized across samples from each case-control pair individually. Genes with an Euclidean distance (ED) < 2.5 to IL-32 (co-clustering results from k-means clustering) are marked with red text (Supplementary table 4). The samples labels along the x-axis include the sample number, case/control indicator, age of sampling in months, and months to (negative no. of months) or from (positive no. of months) seroconversion time. Here, SCC stands for seroconversion-centered, which is why the months to/from seroconversion are negative or positive.

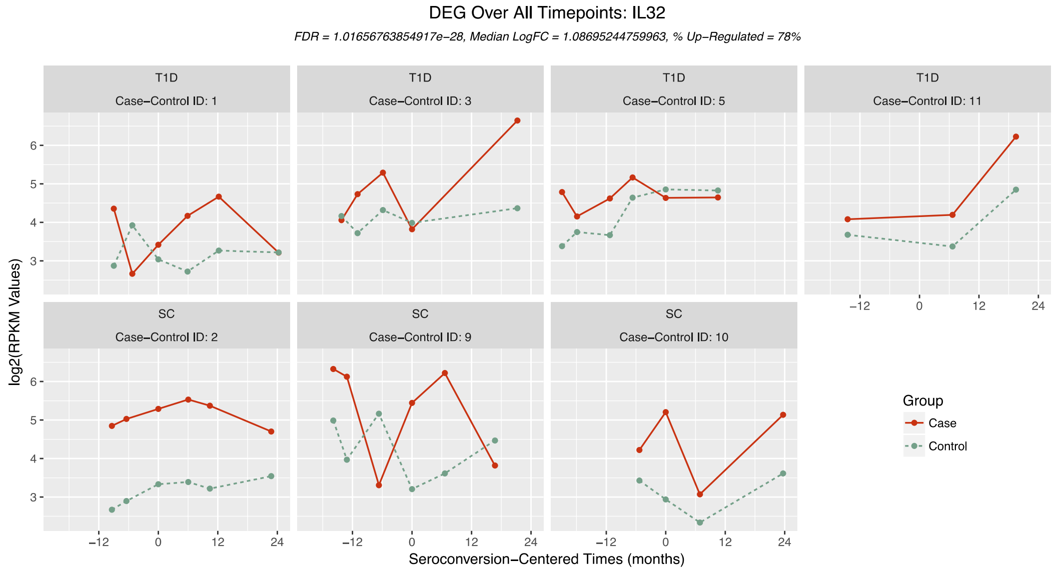


Supplementary Figure 3D. Related to Figure 2.

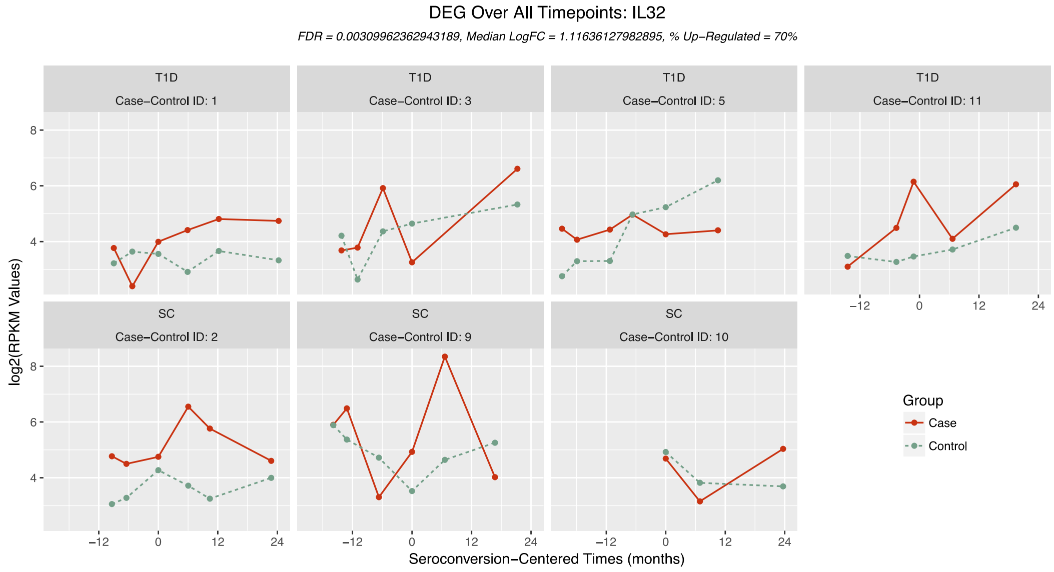
Hierarchical clustering of the levels of standardized autoantibodies (IAA, IA-2A, ZnT8A, and GADA) and the 85 DE genes between the Cases and Controls detected in the PBMC population. Each gene's expression was standardized across samples from each case-control pair individually. Genes with an Euclidean distance (ED) < 2.5 to IL-32 (co-clustering results from k-means clustering) are marked with red text and those with ED < 2.5 to INS are marked with blue text (Supplementary table 4). The samples labels along the x-axis include the sample number, case/control indicator, age of sampling in months, and months to (negative no. of months) or from (positive no. of months) seroconversion time. Here, SCC stands for seroconversion-centered, which is why the months to/from seroconversion are negative or positive.

Supplementary Figure 4A-W. Related to Figure 2.

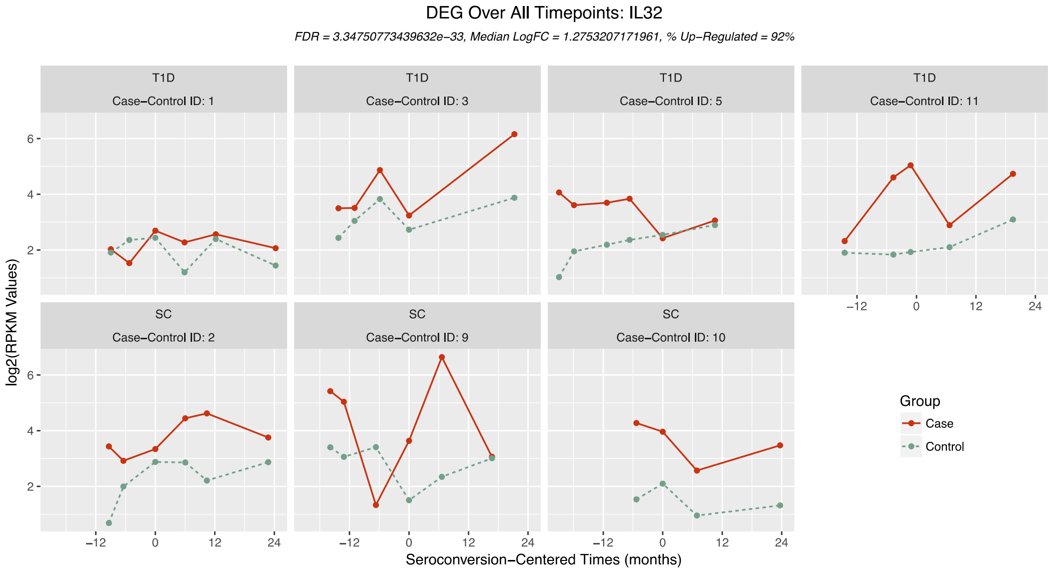
Expression profile plots of genes highlighted in the manuscript:



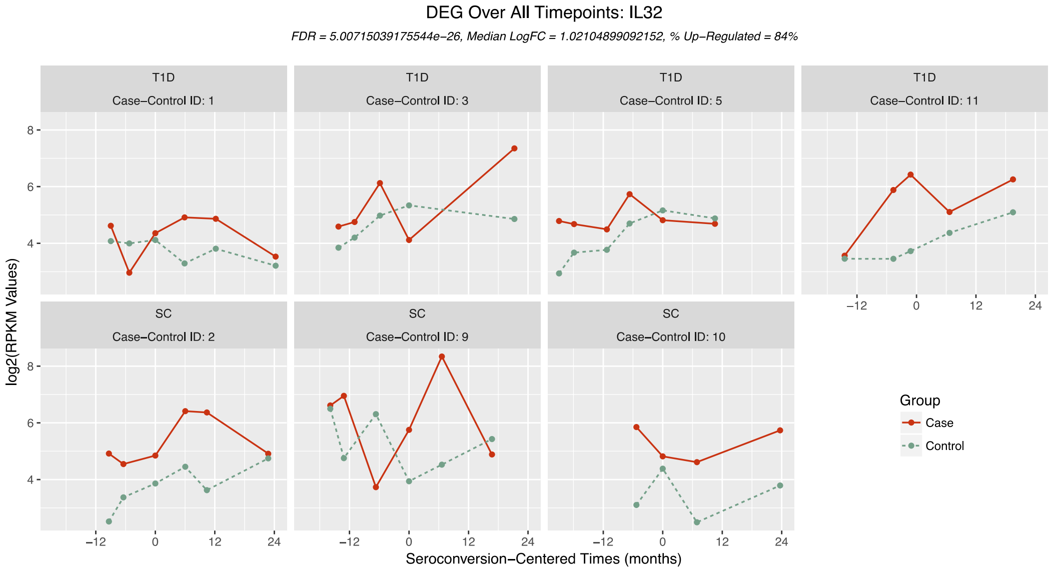
A) Expression levels of IL-32 gene in CD4+ cells.



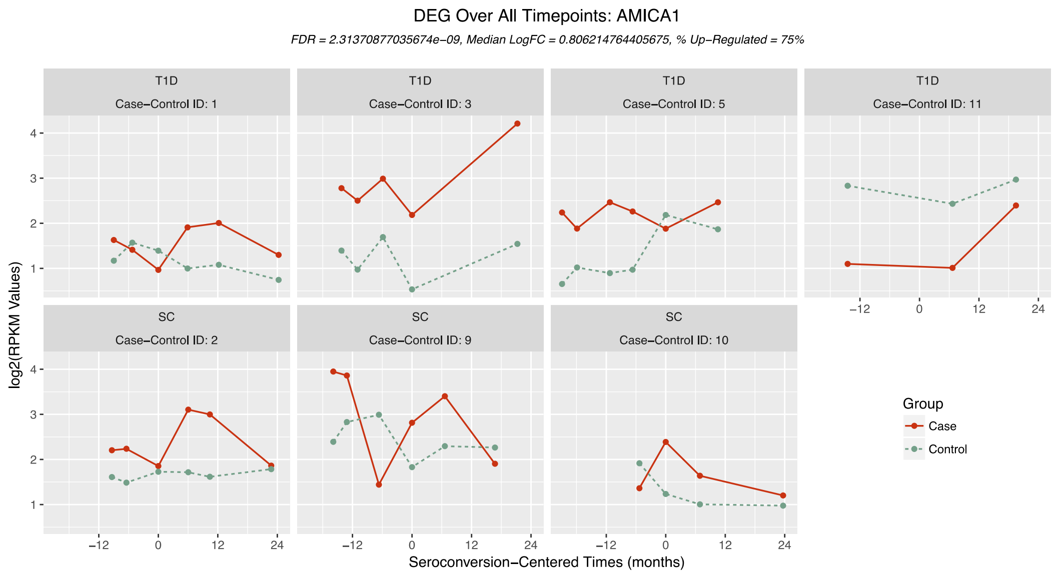
B) Expression levels of IL-32 gene in CD8+ cells.



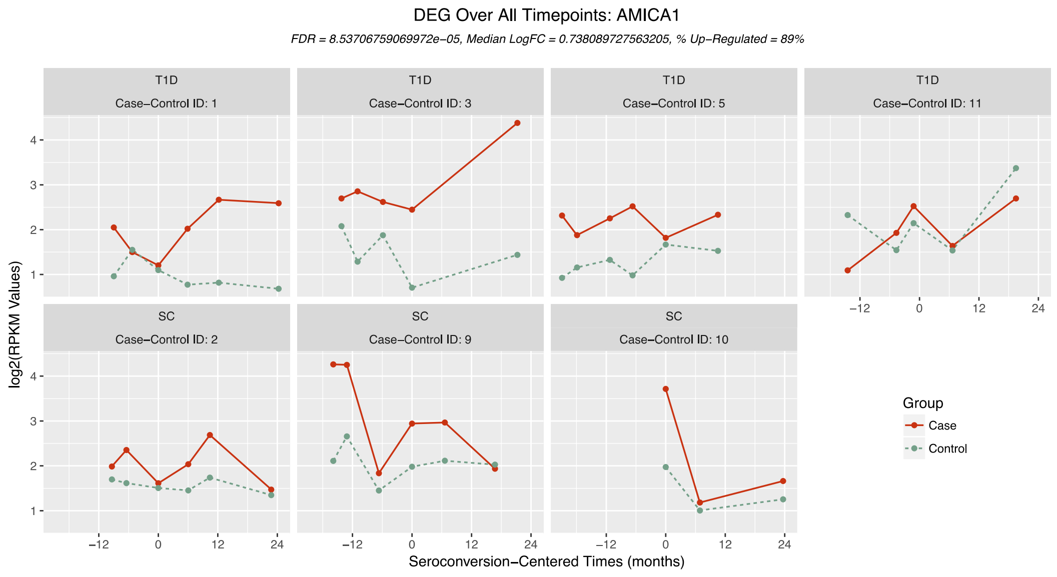
C) Expression levels of IL-32 gene in CD4-CD8- cells.



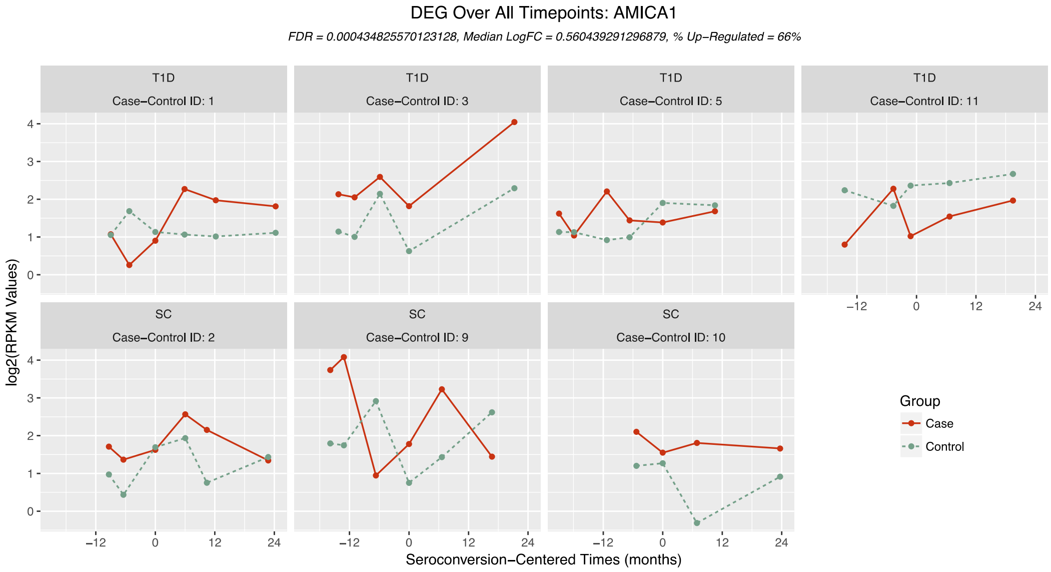
D) Expression levels of IL-32 gene in PBMCs.



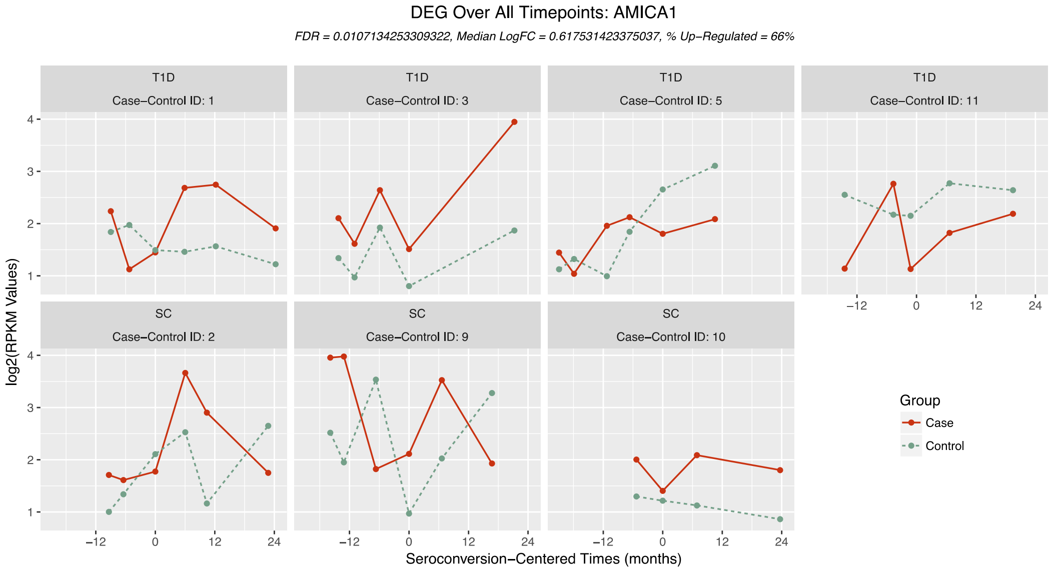
E) Expression levels of AMICA1 gene in CD4+ cells.



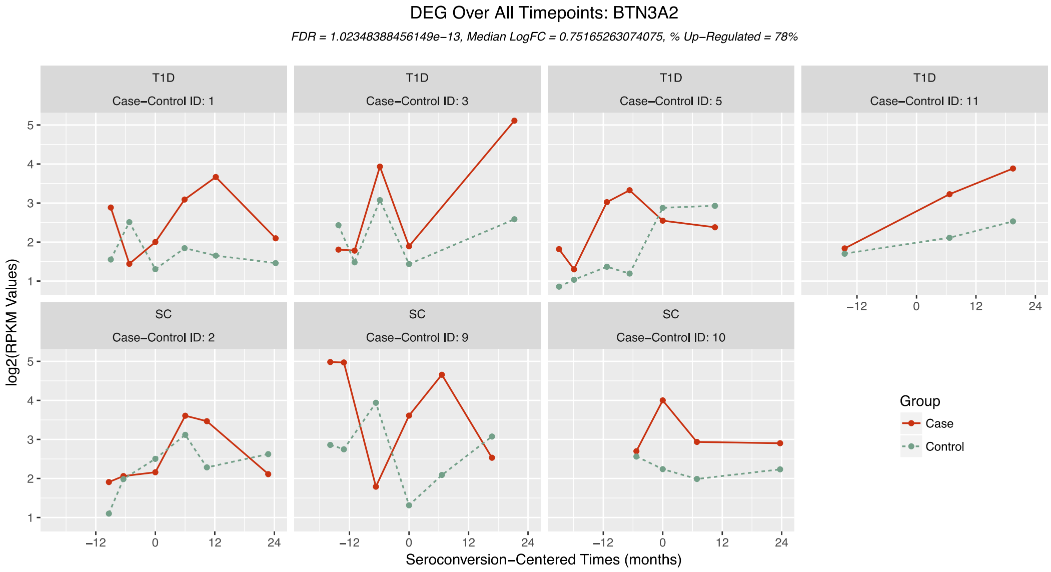
F) Expression levels of AMICA1 gene in CD8+ cells.



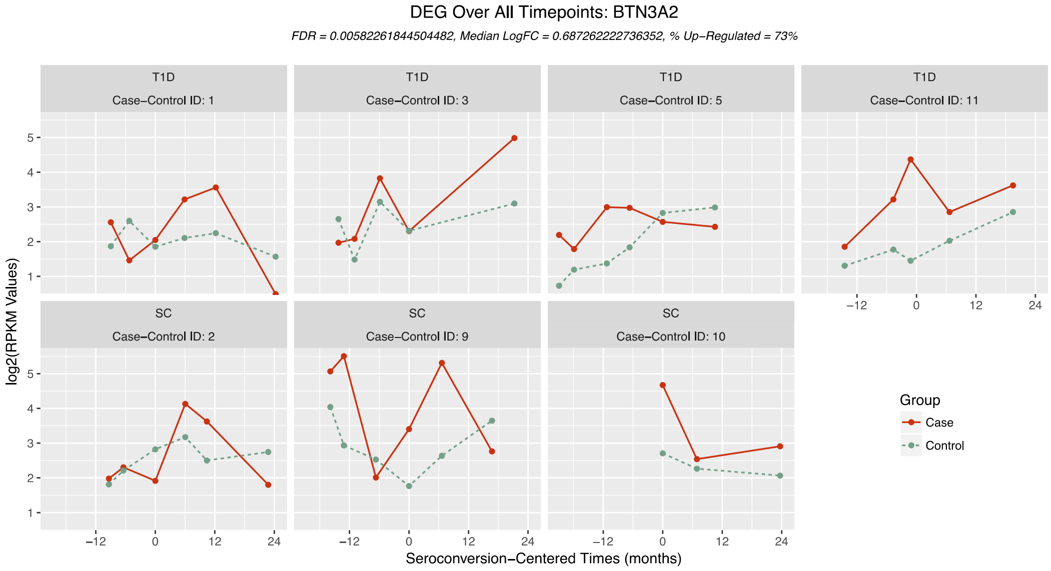
G) Expression levels of AMICA1 gene in CD4-CD8- cells.



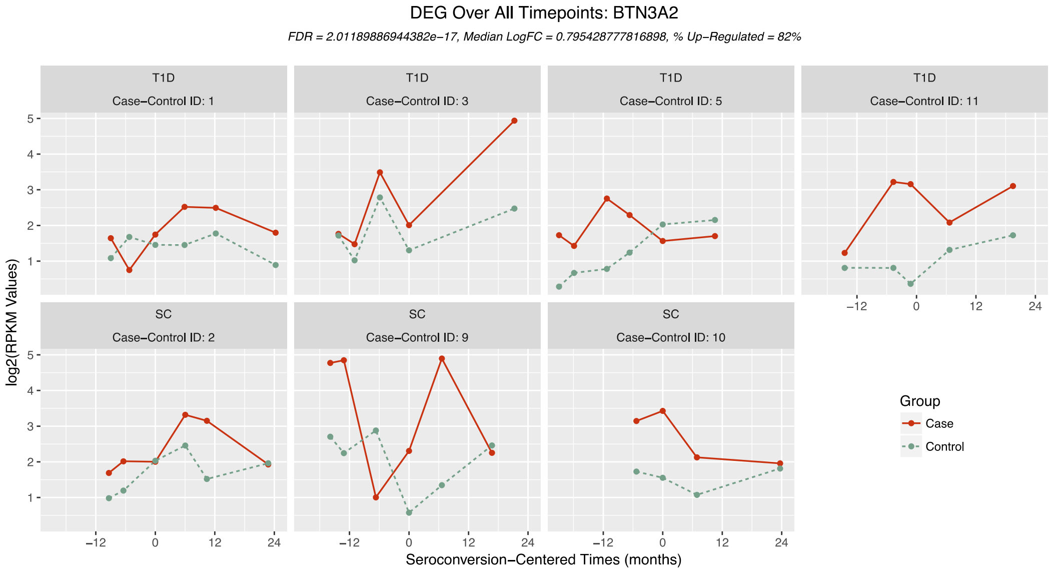
H) Expression levels of AMICA1 gene in PBMCs.



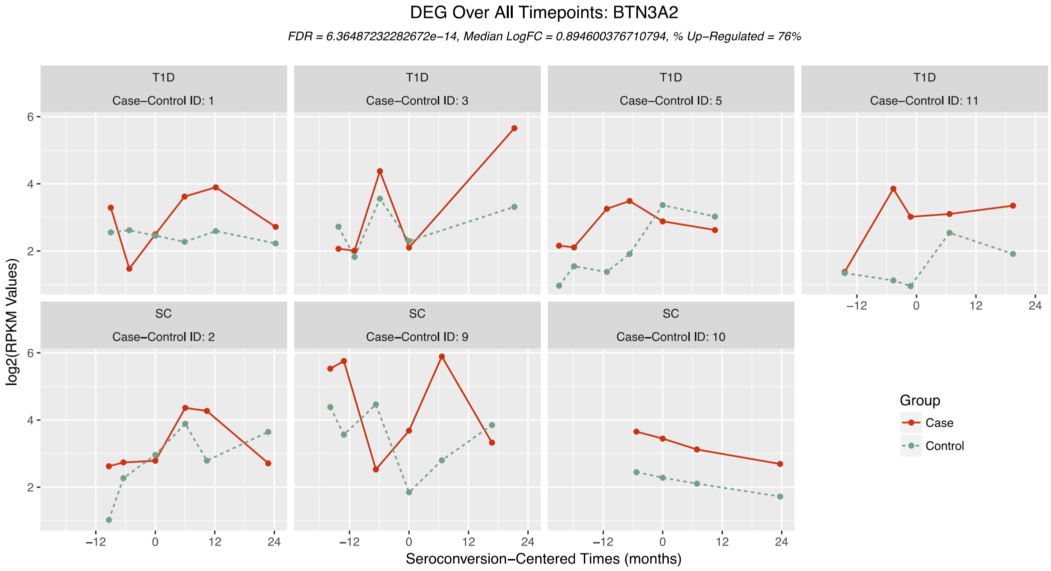
I) Expression levels of BTN3A2 gene in CD4+ cells.



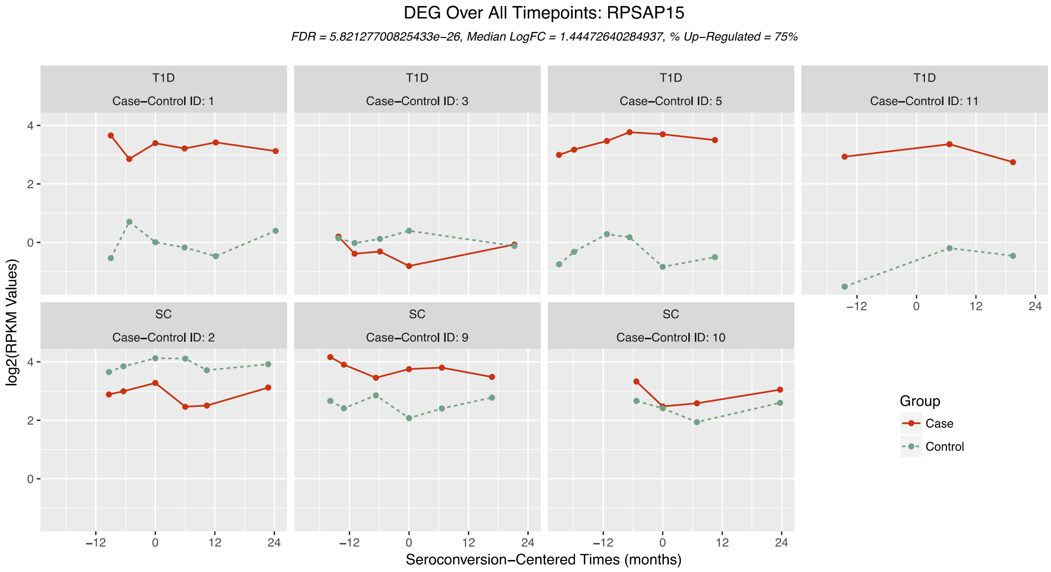
J) Expression levels of BTN3A2 gene in CD8+ cells.



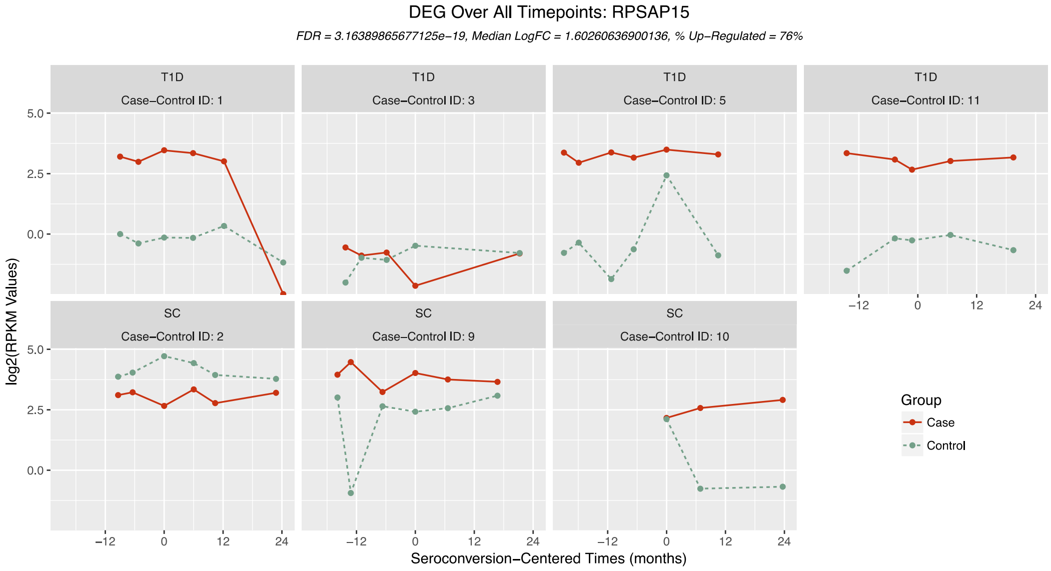
K) Expression levels of BTN3A2 gene in CD4-CD8- cells.



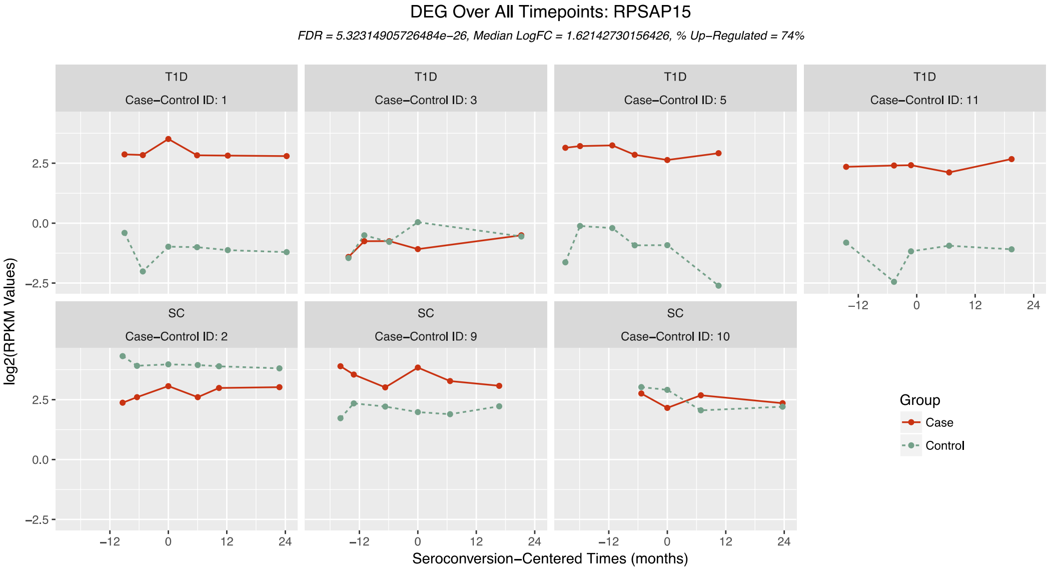
L) Expression levels of BTN3A2 gene in PBMCs.



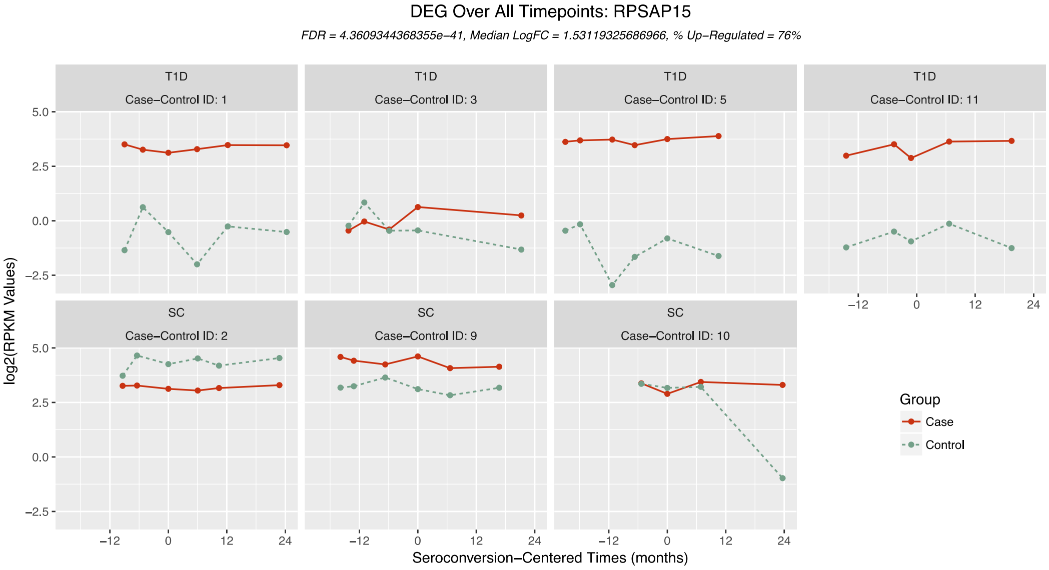
M) Expression levels of RPSAP15 gene in CD4+ cells.



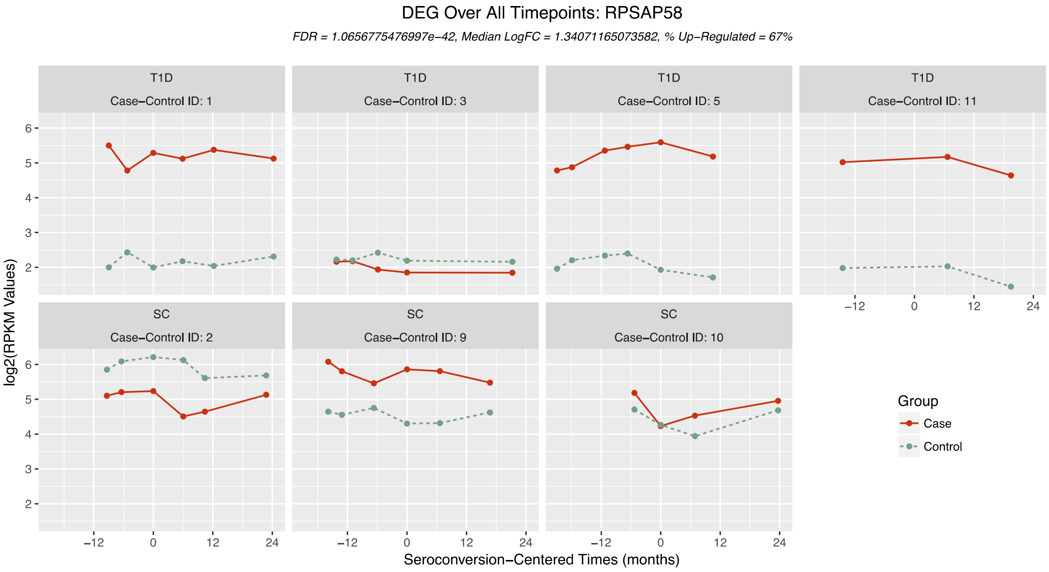
N) Expression levels of RPSAP15 gene in CD8+ cells.



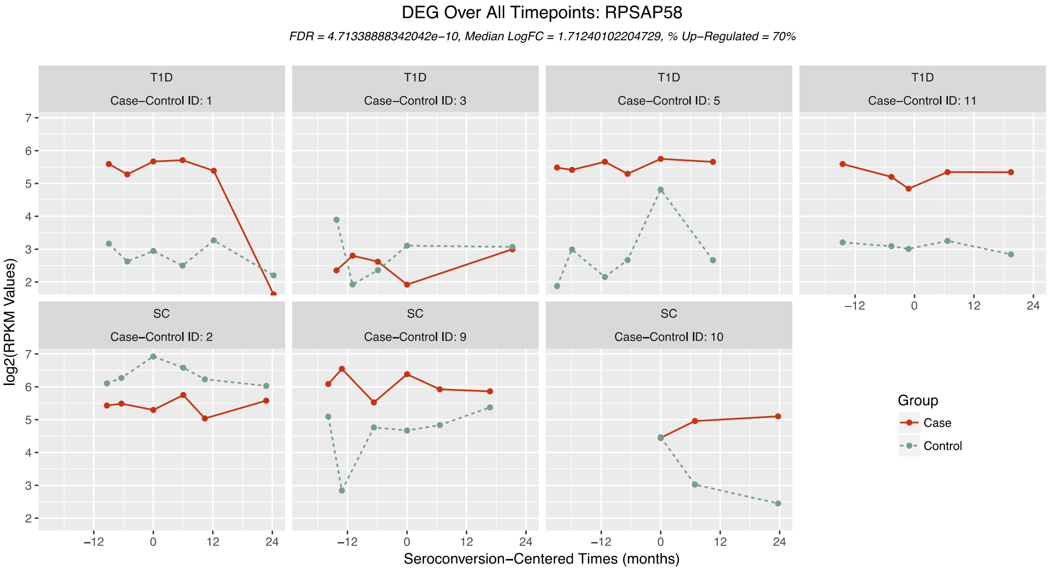
O) Expression levels of RPSAP15 gene in CD4-CD8- cells.



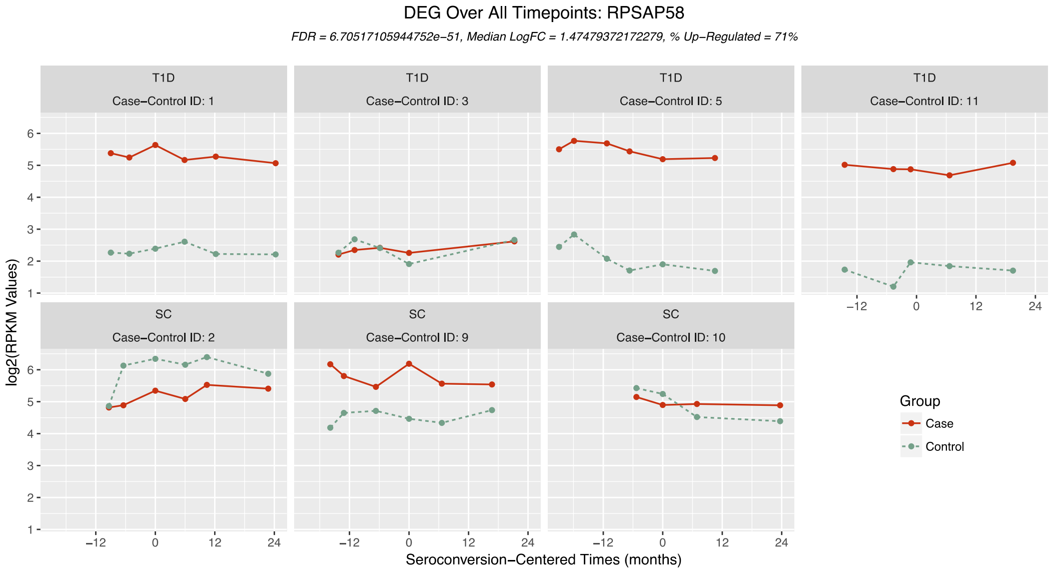
P) Expression levels of RPSAP15 gene in PBMCs.



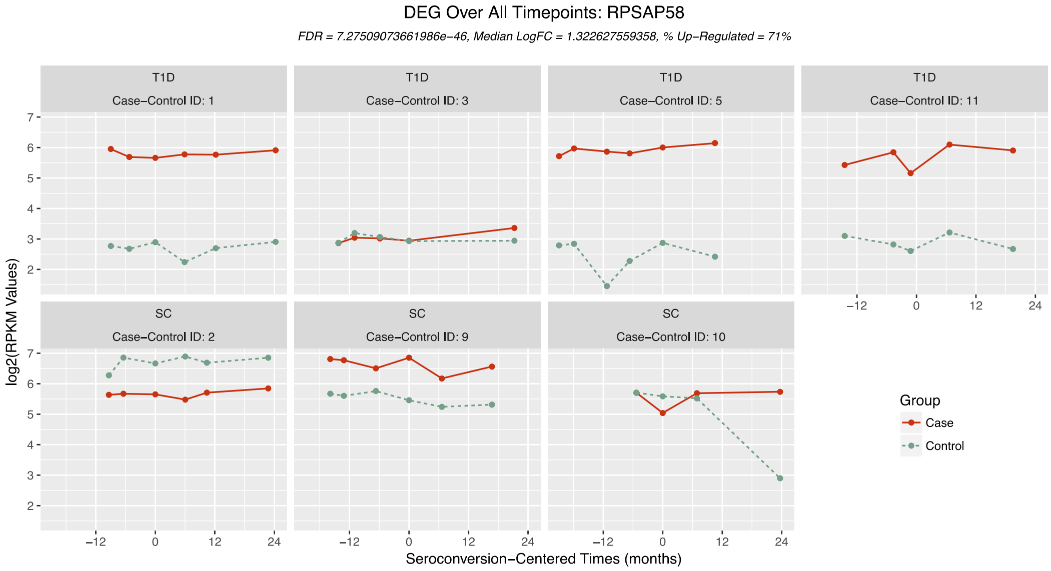
Q) Expression levels of RPSAP58 gene in CD4+ cells.



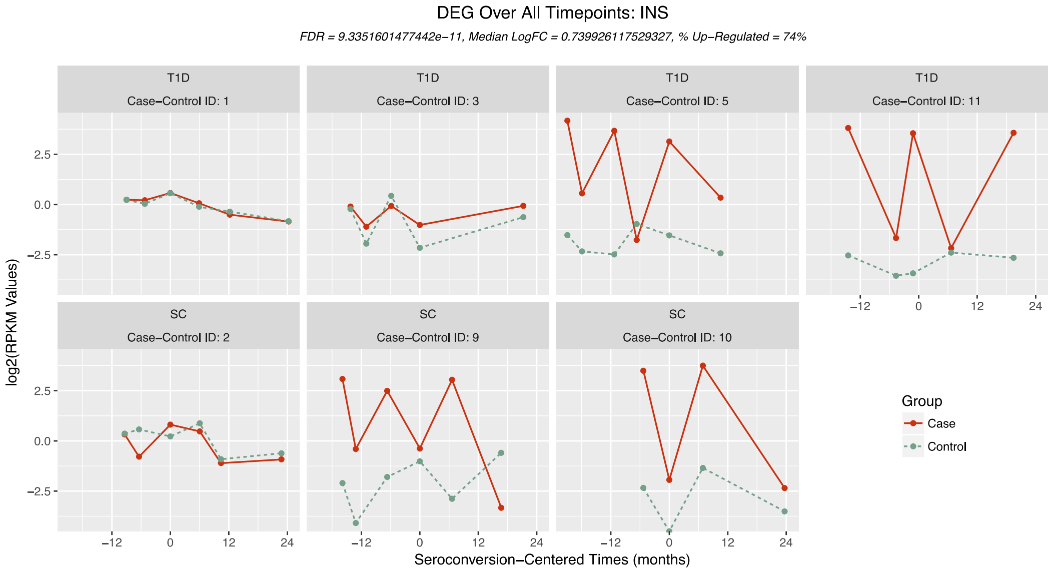
R) Expression levels of RPSAP58 gene in CD8+ cells.



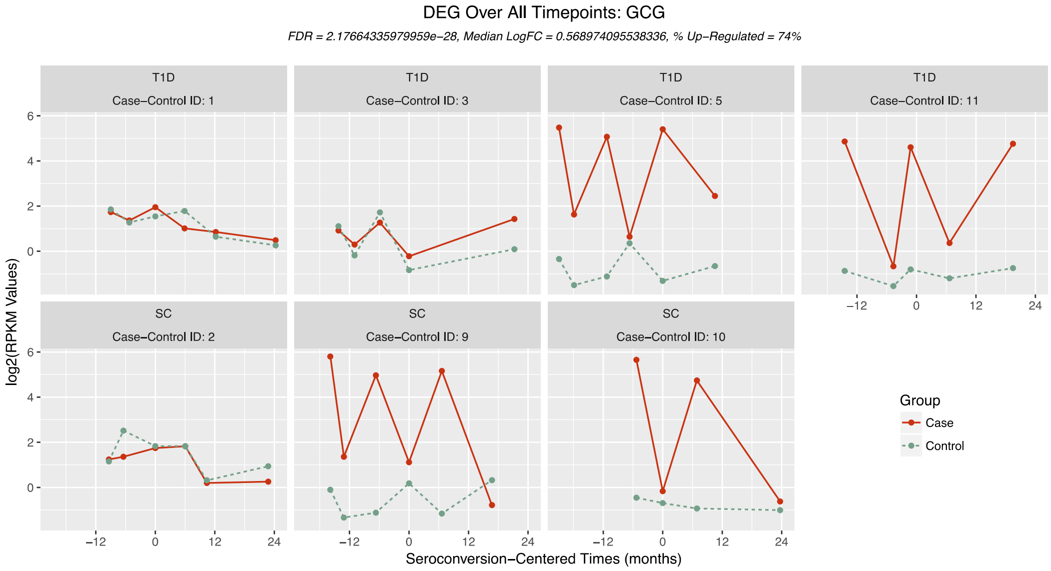
S) Expression levels of RPSAP58 gene in CD4-CD8- cells.



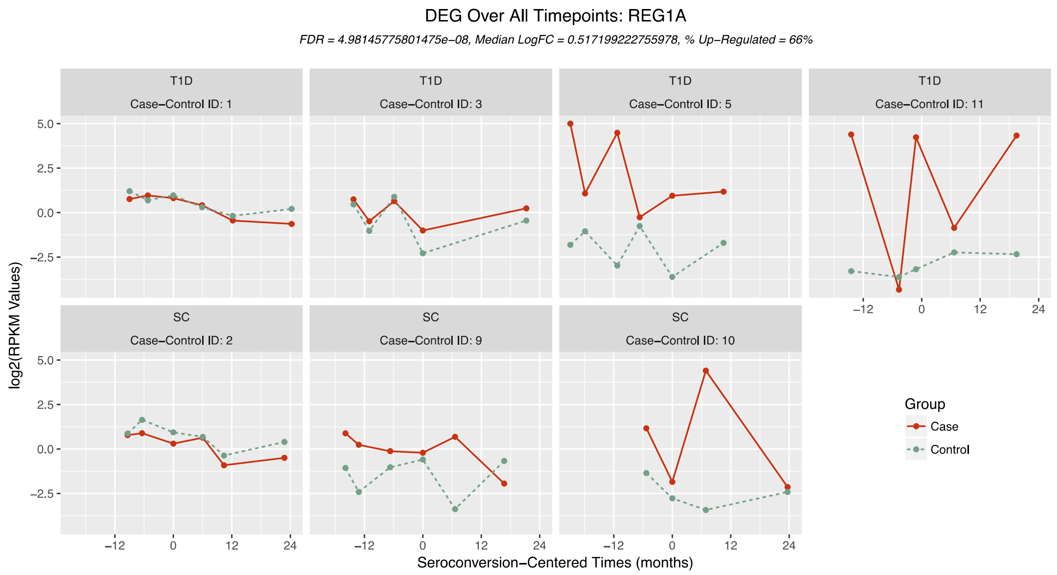
T) Expression levels of RPSAP58 gene in PBMCs.



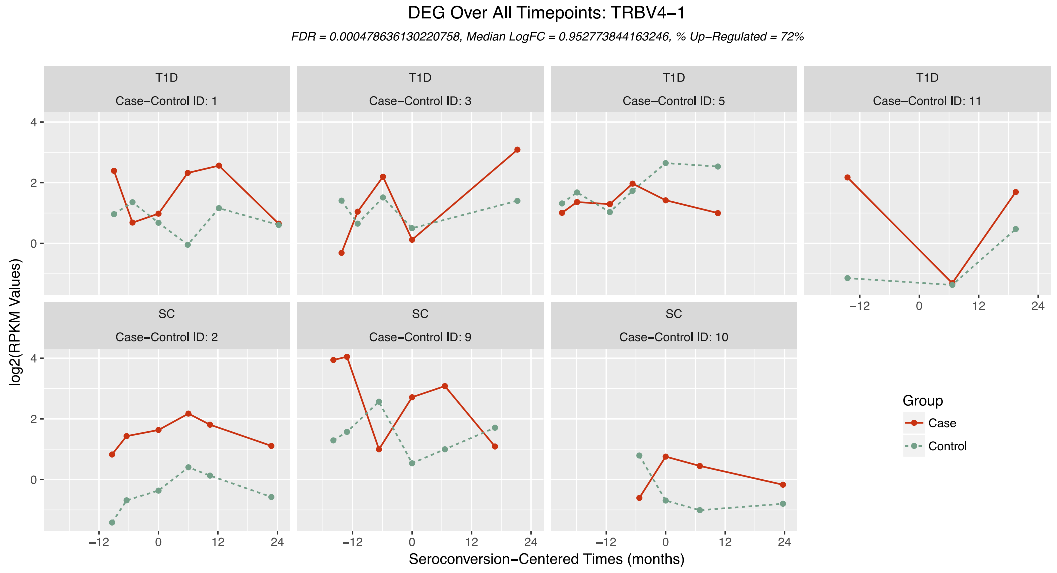
U) Expression levels of INS gene in PBMCs.



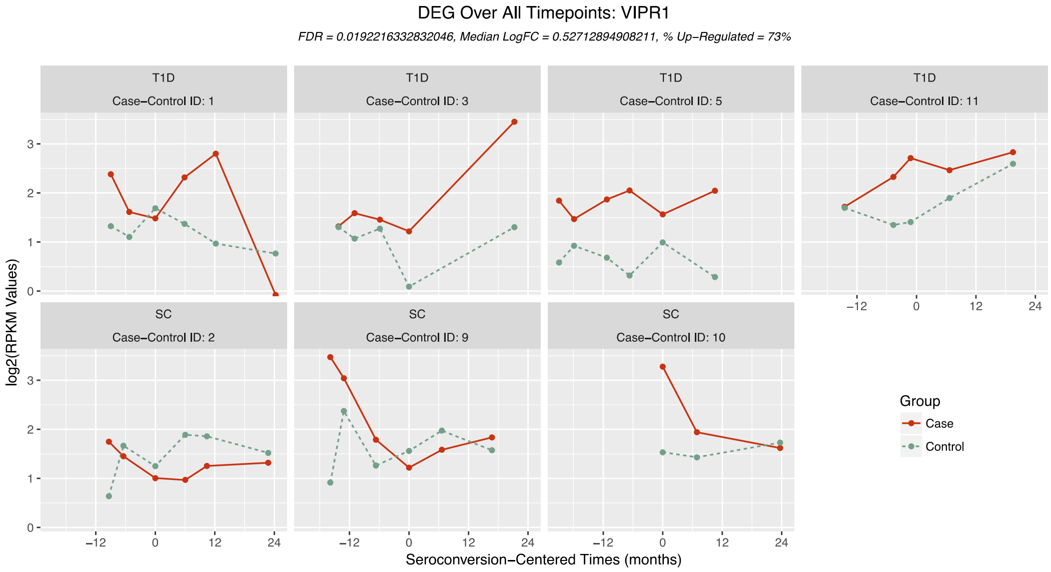
V) Expression levels of GCG gene in PBMCs.



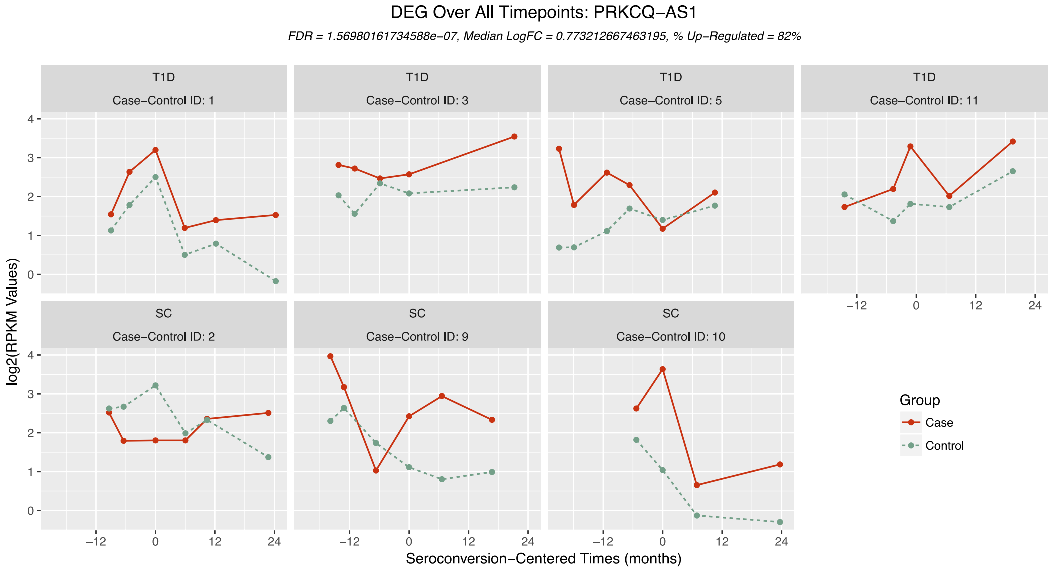
W) Expression levels of REG1A gene in PBMCs.



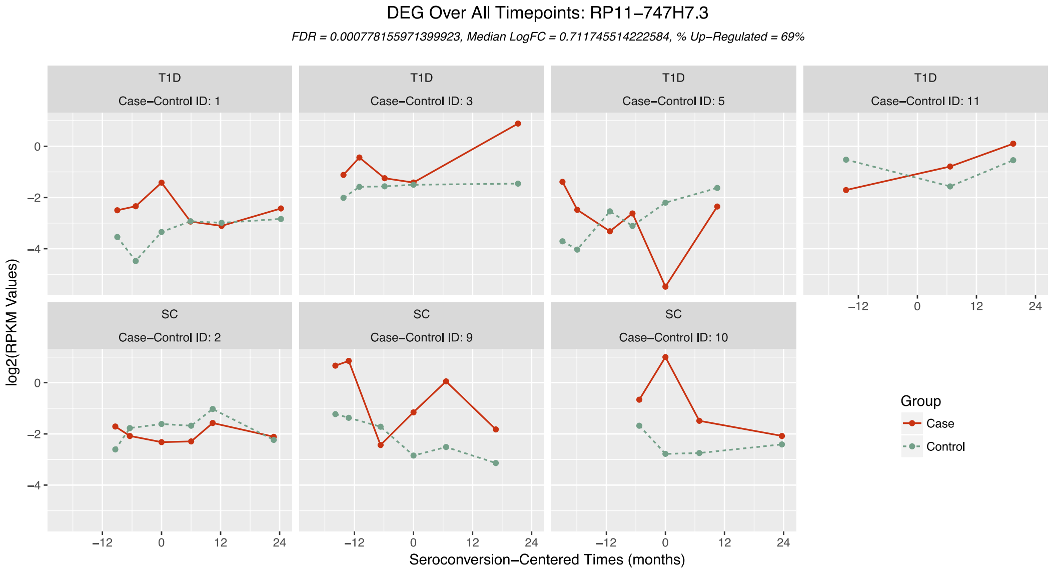
X) Expression levels of TRBV4-1 gene in CD4+ cells.



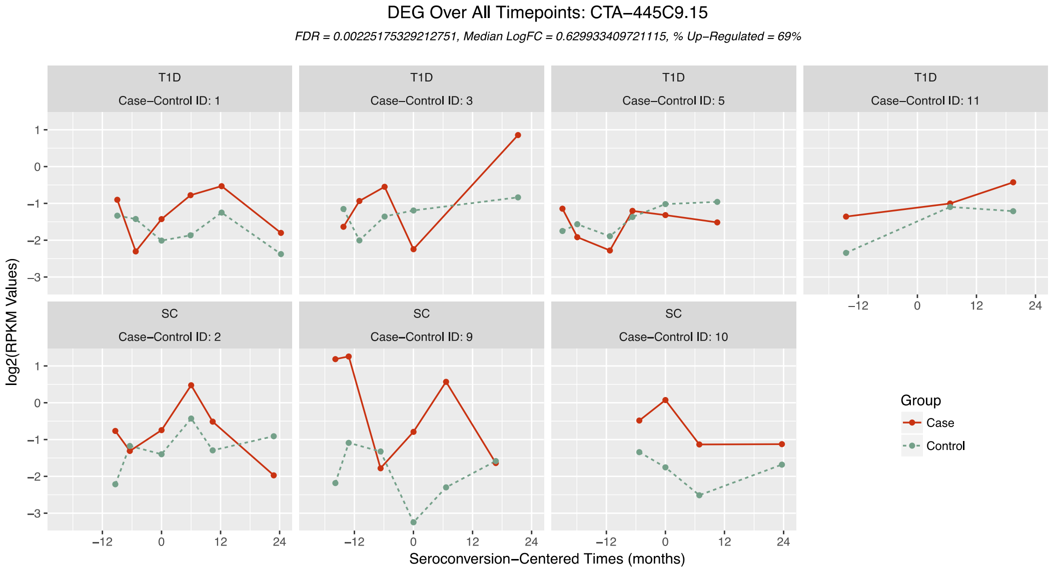
Y) Expression levels of VIPR1 gene in CD8+ cells.



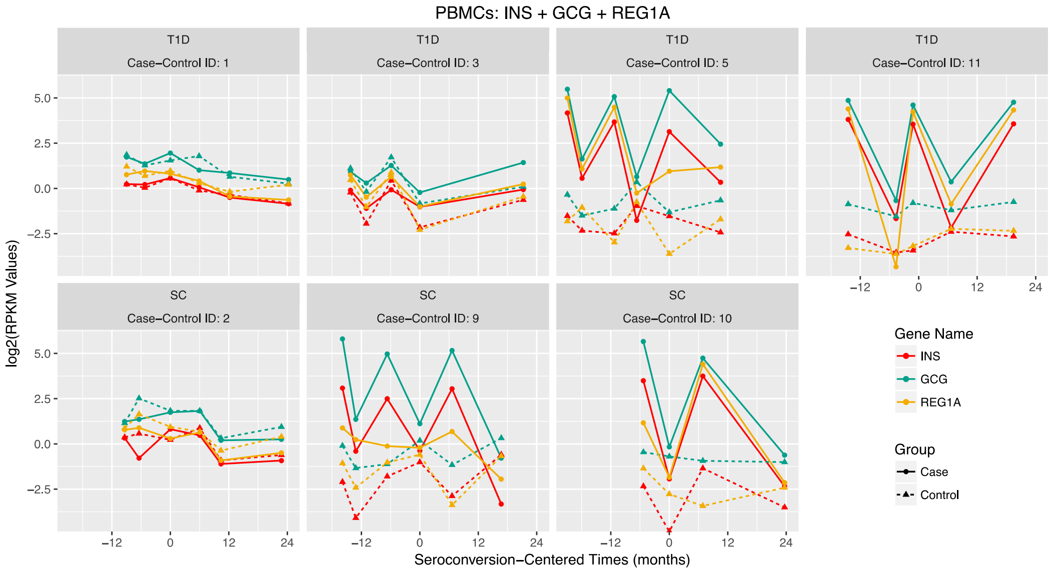
Z) Expression levels of PRKCQ-AS1 gene in CD4-CD8- cells.



AA) Expression levels of RP11-747H7.3 gene in CD4+ cells.

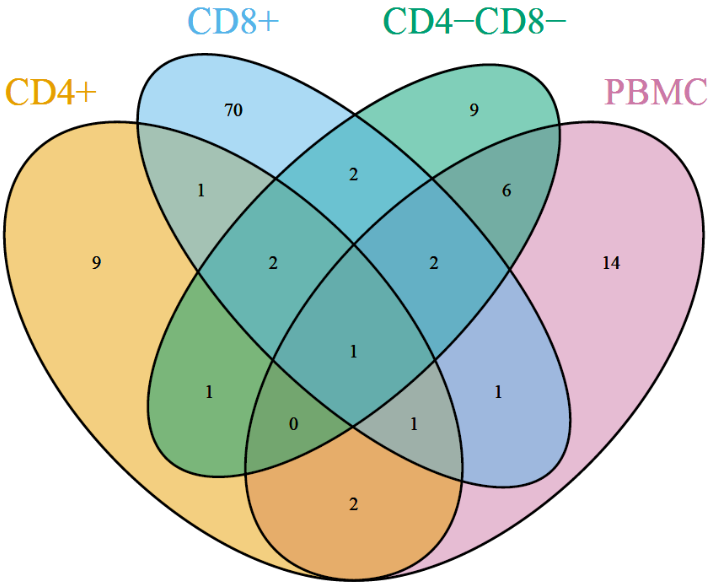


AB) Expression levels of CTA-445C9.15 gene in CD4+ cells.



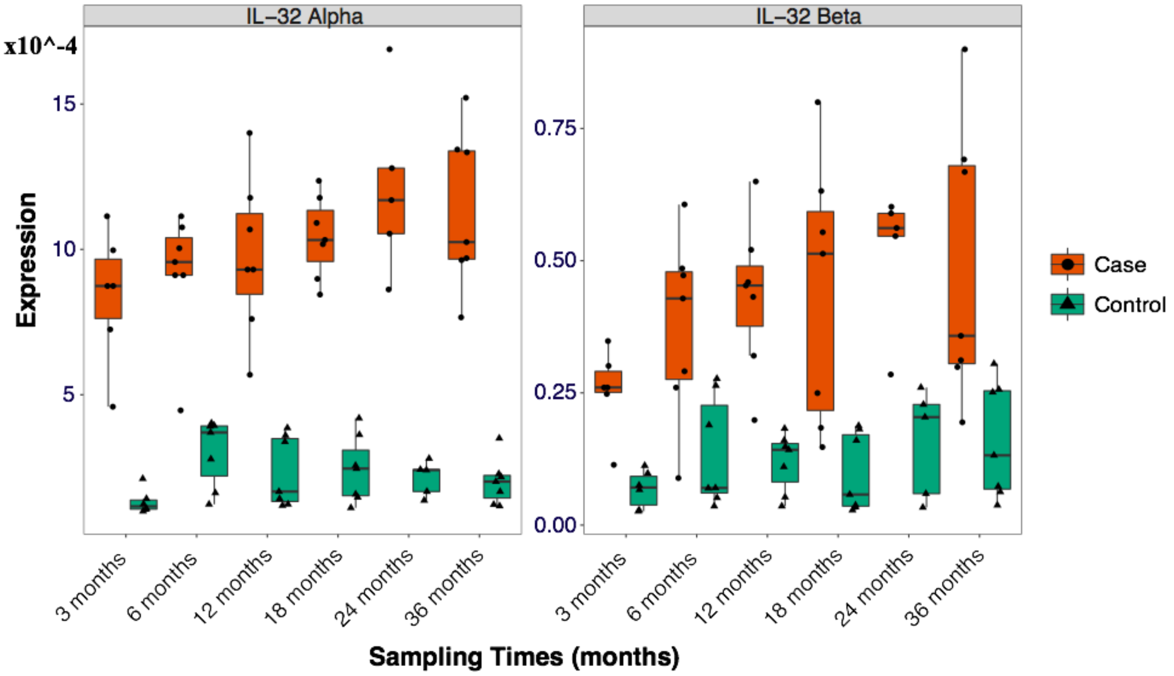
Supplementary Figure 5. Related to Figure 2.

PBMC-specific co-regulation of pancreatic transcripts Insulin (INS), Glucagon (CGC) and Regulin 1 alpha (REG1A). Profiles of *INS*, *CGC* and *REG1A* show concerted gene expression profiles. For individual profiles, see Supplementary Figure 4.



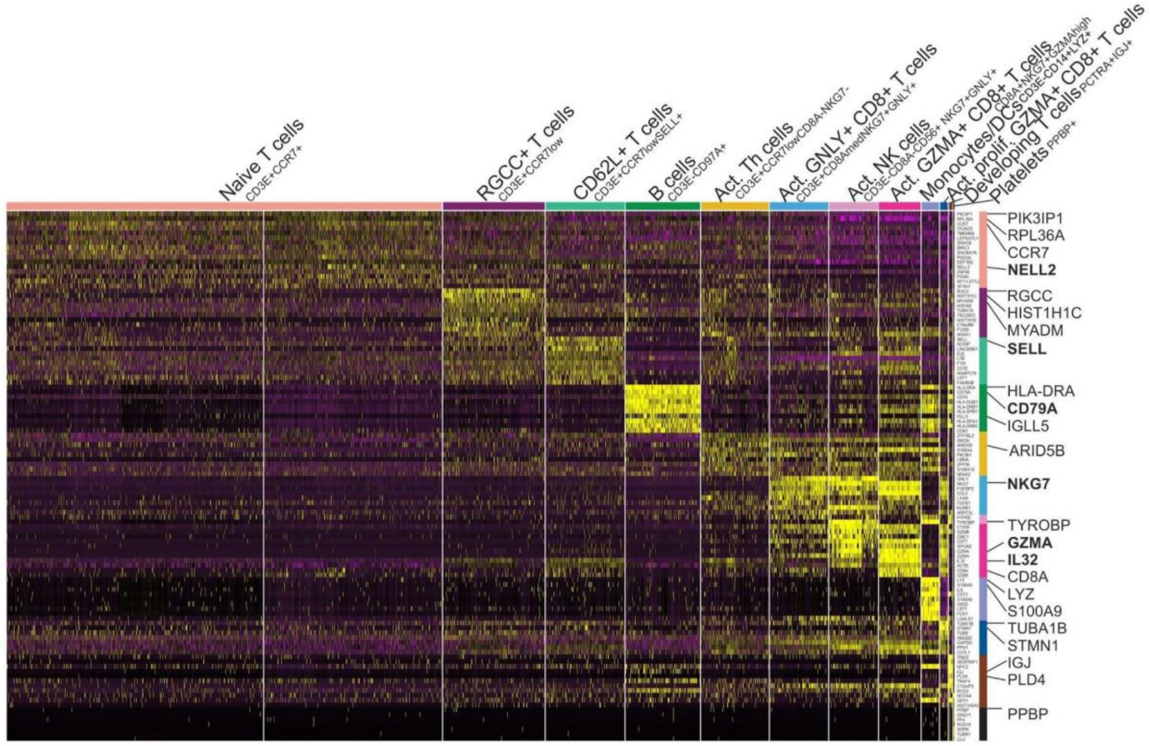
Supplementary Figure 6. Related to Figure 2.

Number of DE genes between the Cases and Controls in the time-window of 12 months before seroconversion (RPKM > 3 for coding genes and RPKM < 0.5 for non-coding genes, Up- or downregulated in ≥ 65% of the Cases). For complete listing, see Supplementary Table 3 columns “12 mo before SC”.



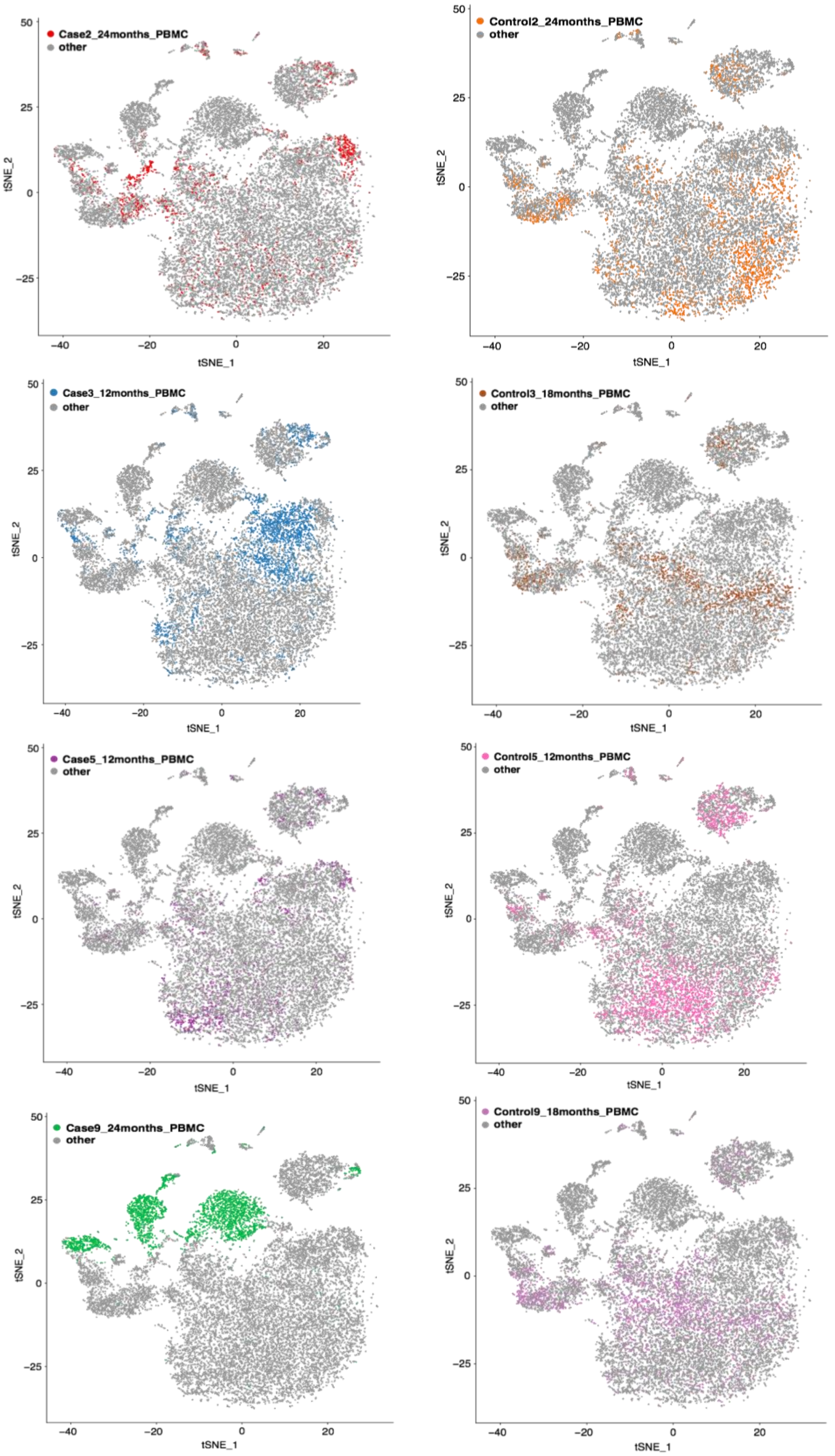
Supplementary Figure 7. Related to Figure 3A.

mRNA expression of IL32 isoforms analysed in PBMC samples of Cases (n=7) and their matched Controls (n=7) by qRT-PCR. For expression level plot of IL32γ isoform, please refer to **Figure 3A.**



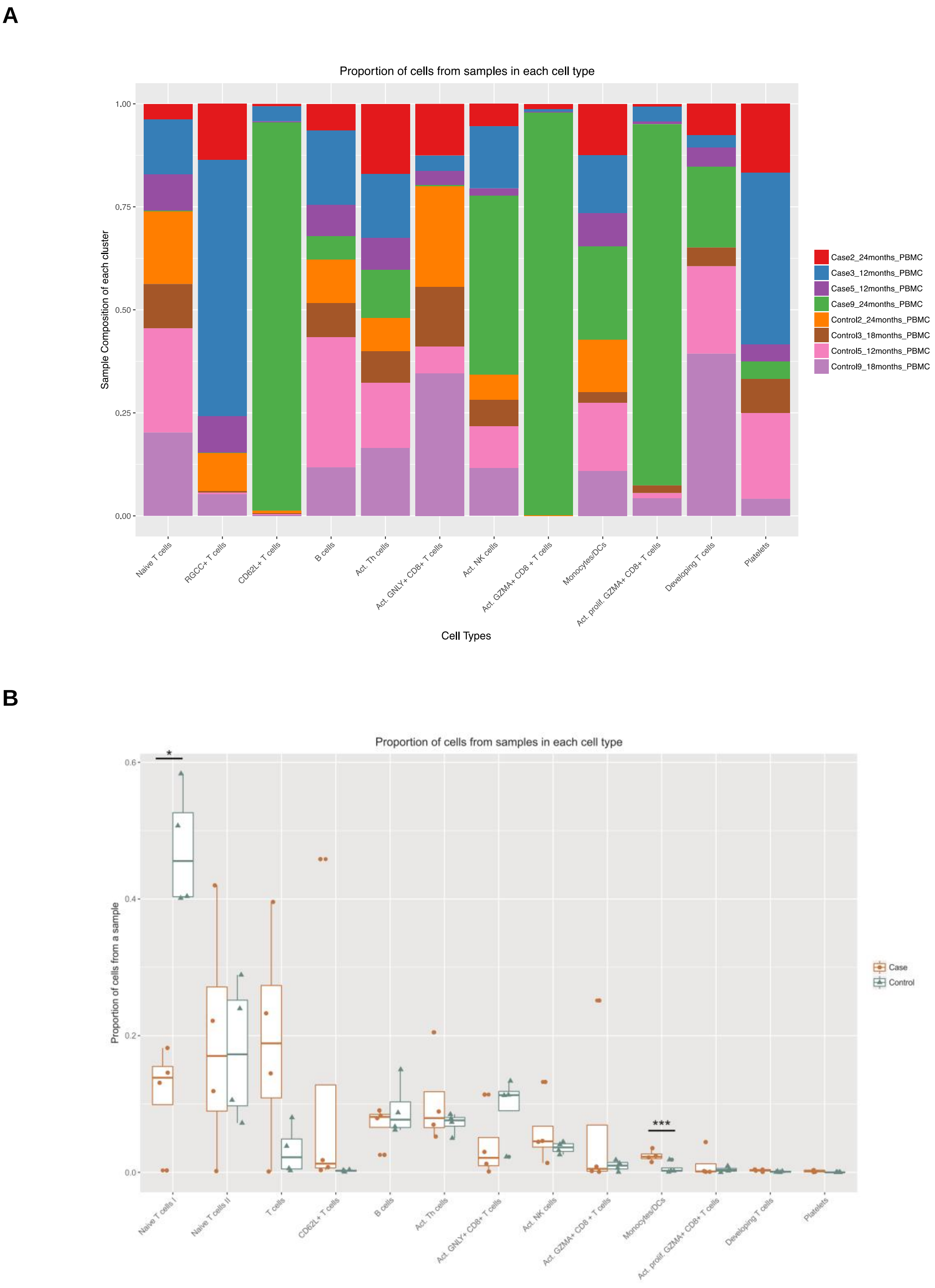
Supplementary Figure 8. Related to Figure 3.

Heatmap of the top 10 most highly expressed genes in the 13 clusters identified after Seurat clustering analysis of the pooled single-cell RNA-Seq data (4 Cases + 4 Controls) represented as a tSNE plot in **Figure 3B**. As the two biggest clusters were similar in their gene expression profiles, they were merged to form the Naive T cell cluster, leaving in total 12 cell clusters. Genes used in the annotation of the cell clusters are marked on the right column, where bolded genes are those that were also found to be DE between Cases and Controls in the bulk RNA-seq data analysis (**Supplementary Table 3**).



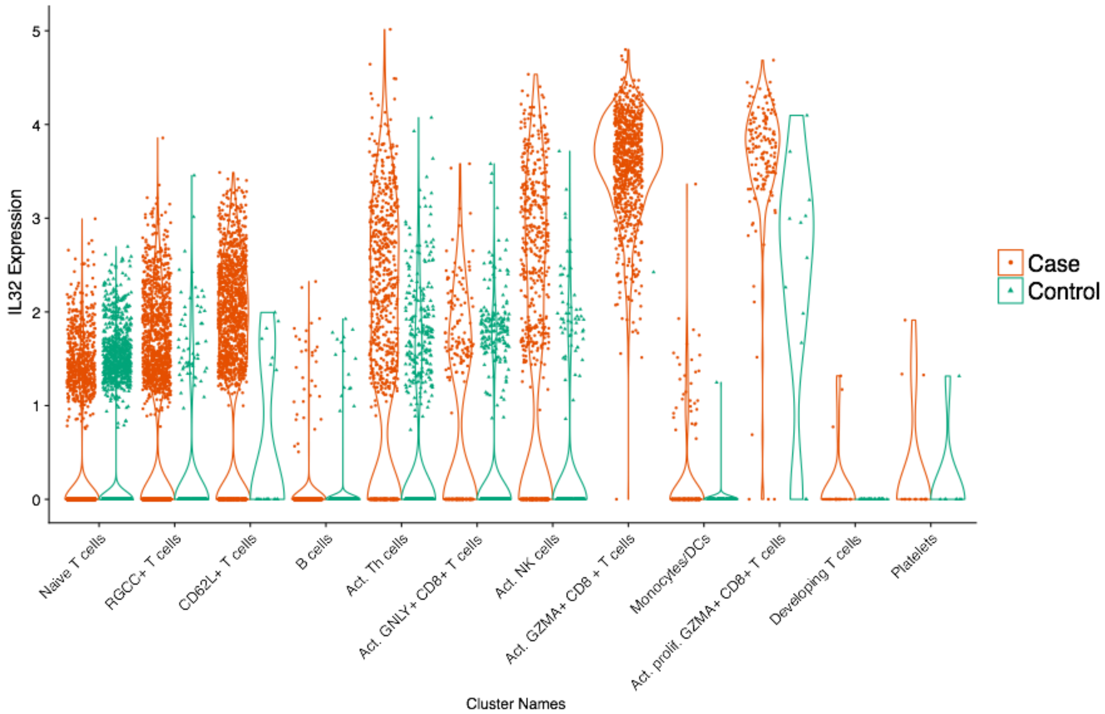
Supplementary Figure 9. Related to Figure 3.

Contribution of individual samples in the t-SNE visualization of pooled single-cell RNA-Seq data, presented in **Figure 3B** and **C**.



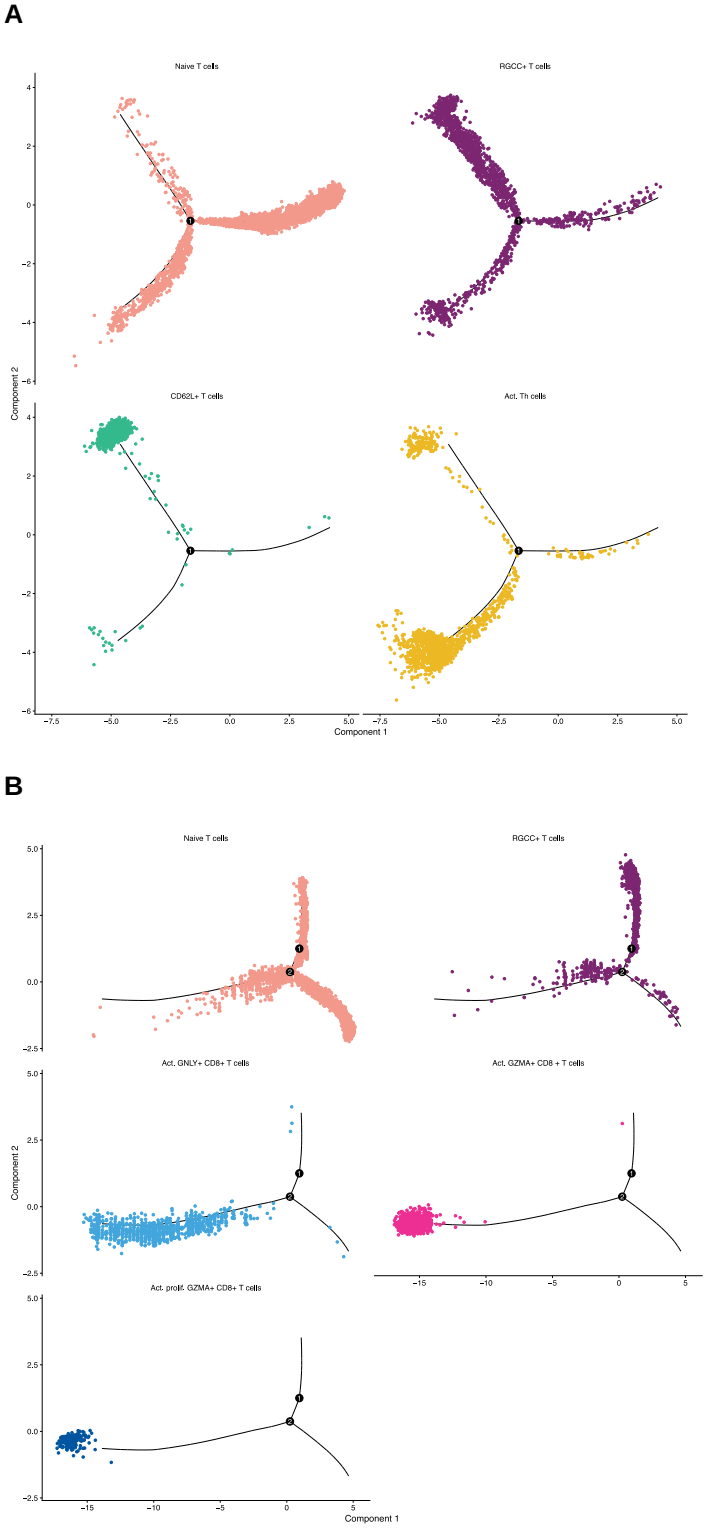
Supplementary Figure 10. Related to Figure 3.

A) Proportion of cells coming from individual samples per cluster (cluster-wise proportioning) **B)** Box-plot highlighting the proportions of cells per cluster in Case (orange) and Control (green) samples. * $p < 0.05$, *** $p < 0.005$ according to paired t-test of the sample-wise proportions of cells per cluster.



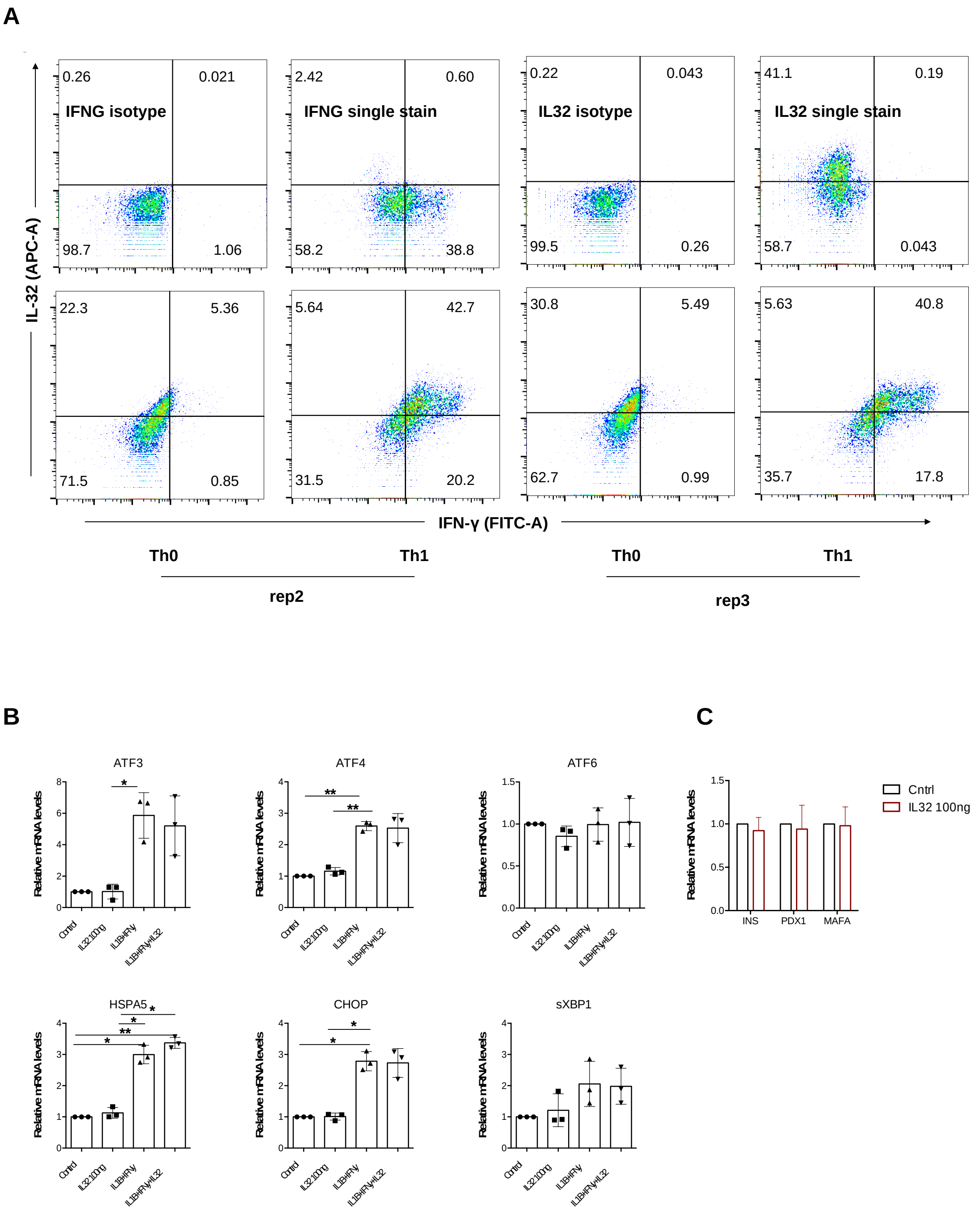
Supplementary Figure 11. Related to Figure 3.

Violin plot showing the expression of *IL32* in the 12 cell clusters identified from the single-cell RNA-Seq data, displayed separately for Cases (orange) and Controls (green).



Supplementary Figure 12. Related to Figure 3.

A) Trajectory in pseudotime of CD4+ specific and **B)** CD8+ specific cells along with the precursor cells plotted individually.



Supplementary Figure 13. Related to Figure 4.

A) Tow additional replicates of the Th0/Th1 intracellular staining data shown in **Figure 4C**. **B)** EndoC-βH1 cells were treated for 24 h with recombinant IL-32*γ* in presence and absence of IL-1β and IFNγ, and the expression of ER stress markers *ATF3*, *ATF4*, *ATF6*, *CHOP*, *HSPA5* and *sXBP1* was measured by RT-qPCR assay. **C)** Expression of endocrine marker genes *INS*, *PDX1* and *MAFA* was measured after treatment of EndoC-βH1 cells with 100 ng of IL-32γ for 24h. In **B-C** fold change is calculated as compared to non-treated (control) cells. Statistical significance was determined by Tukey's multiple comparisons test. * =p-value <0.05 while ** =p-value <0.01.

Supplementary Material

Bulk RNA-seq data analysis

RNA-seq data processing and analysis

Of the 306 RNA-seq samples (**Supplementary Table 1**), 298 were used for the differential expression analysis because some Case samples had more than one corresponding control samples. The average sequencing depth of the samples in this study was around 51 million paired-end reads. Quality control checks were performed on the raw RNA-seq data using FastQC (version 0.10.0, http://www.bioinformatics.babraham.ac.uk/projects/fastqc/). The reads were aligned to the human reference transcriptome, Human GRCh37 assembly version 75 (http://feb2014.archive.ensembl.org/index.html), using TopHat (version 2.0.10), where the default parameters of the software were retained. On average, approximately 93% of the reads from each sample in each fraction were mapped (average overall read mapping of samples in each cell type was CD4+: 93%, CD8+: 92.6%, CD4-CD8-: 93.4%, PBMC: 93.5%). This resulted in about 89% of concordant pair alignments (CD4+: 88.9%, CD8+: 88.93%, CD4-CD8-: 89.89%, PBMC: 89.89%). The aligned reads, with a mapping quality > 10, were counted at the gene level availing the *htseq-count* function from the HTSeq package and using the overlap resolution mode of 'intersection-strict' (htseq-count version 0.6.1). The read counts of genes were normalized using the trimmed means of the M-values (TMM) method, implemented in the software package *edgeR*, which adjusts for varying sequencing depths as well as normalizes for the RNA composition. Using the biotype information, the genes were divided into coding and non-coding categories. The biotype data for each gene were retrieved from the Ensemble database , and the descriptions of biotypes were taken from Gencode (http://www.gencodegenes.org/gencode_biotypes.html).

Filtering genes using RPKM values

First, the RPKM values were calculated for each gene in each sample of the analysis, where the length of the gene was taken to be the sum of the lengths of all its known exons. Second, a max-of-means RPKM value (mmRPKM) was computed for each gene to assess the overall expression of the gene in all the samples of the analysis. As the differential expression analyses in this study usually involved two groups (e.g., cases and controls, CD4+ and PBMCs), the max-of-means of RPKM value refers to: maximum(mean("RPKM values in Group 1"), mean("RPKM values in Group 2")). Subsequently, coding genes with mmRPKM > 3 and non-coding genes with mmRPKM > 0.5 were retained. These filtering criteria usually retained about 7000–7500 coding genes and 600–700 non-coding genes.

Differential expression analysis

Differential expression analyses were conducted separately for coding and non-coding genes, using the *edgeR* package. The variance of the data was estimated using the trended dispersion method.

Post-differential analysis filtering steps (only for paired sample analyses)

The *edgeR* output of differentially expressed (DE) genes with FDR < 0.05 from the paired sample analyses were further subjected to median $\log_2$FC filtering, where DE genes with a |median $\log_2$FC| > 0.5 were retained for downstream filtering step. The final filtering step retained only those genes as DE that had more than 65% samples across all individuals regulated in the same direction (i.e., up- or down-regulated). These filtering steps were added to discard false positives that may arise due to the heterogeneity of the samples due to normal variation, which is non-related to T1D and outliers. A visual depiction of the RNA-seq data processing and analysis pipeline has been shown in **Supplementary Figure 2**.

Analysis 1: Cell fraction vs PBMC

In this analysis, the expressions of genes from each cell fraction (i.e., CD4+, CD8+ and CD4-CD8-) were compared to those of the (paired) original PBMC population of Control children. Samples collected at all ages were included but were required to have expression data from both the fraction under analysis as well as PBMCs.

Analysis 2: Cases versus Controls – over all timepoints

The aim of this analysis was to identify genes that are differentially expressed in children who have seroconverted to autoantibody positivity (Cases) in comparison to those who have not (Controls). Each Case child was matched to a Control child, according to date of birth, HLA-risk class, gender and country. Case samples were compared to the samples from their matched Controls that were collected at the same age. In these analyses, other than for pairing purposes, the sampling ages were not utilized.

Differential expression analysis of Cases and Controls were compared with another method. The RNA-seq data of the filtered coding and non-coding genes were modelled using generalized linear mixed effects models (GLMMs) using *glmer()* function from *lme4* package (23). A random effect is added in this model for each child's samples. GLMM with the negative binomial likelihood was fit to the data using the *MASS* package, where the dispersion values per gene were obtained from the *edgeR*. For the filtered coding and non-coding genes from all fractions, the Spearman rank correlation coefficients of the results from the two methods ranged between 0.91 and 0.96 with an average of 0.936, indicating similar ranking of the genes after FDR correction in both the methods. Further details of RNA-seq data analysis can be found in the "Supplementary Material".

Analysis 3: Cases versus Controls – 12 months before the seroconversion window

This analysis is similar to **Analysis 2** in terms of the Case versus Control analysis set-up. However, to understand gene expression changes that take place in Cases right before seroconversion, this analysis compared only those Case samples that were taken at most 12 months before seroconversion with their matched Control samples.

For comparison, the RNA-Seq data of the filtered coding and non-coding genes were also modelled using generalized linear mixed effects models (GLMMs) using the same design as explained in **Analysis 2**. A random effect is added in this model for each child's samples. The *glmer()* function from *lme4* package was used here for modelling. GLMM with the negative binomial likelihood was fit to the data using the *MASS* package, where the dispersion values per gene were obtained from the *edgeR* **Analysis 2**. For the filtered coding and non-coding genes from all fractions, the Spearman rank correlation coefficients of the results from the two methods ranged between 0.91 and 0.96 with an average of 0.936, indicating similar ranking of the genes after FDR correction in both the methods.

Differential gene clustering

To find the genes and autoantibodies (together referred to as 'features' in this section) co-regulated/co-clustering with *IL32* in each cell-fraction, *k*-means clustering, followed with Euclidean distance based co-clustering selection criteria, was performed on the expression levels of coding and non-coding differentially expressed genes (**Analysis 2**) as well as on the autoantibodies. Due to the heterogeneity of the data and the disease, the clustering was done individually on each case and its matched control. Before clustering, the RPKM expression values of each gene and expression level of each autoantibody were $\log_2$ transformed to ensure approximately normal distribution of the values, and gene-wise standardized to make the features comparable.

For each possible number of clusters (i.e., from 2 to total number of features - 1), the features were clustered using the *k*-means clustering algorithm (*kmeans* function implemented in R *stats* package). Subsequently, using the resulting classification of features into clusters along with the Euclidean distance measures between the features, a silhouette score was calculated. The optimum number of clusters was chosen to be the one with the largest silhouette score. The features were then clustered into the "optimum number of clusters" using *k*-means clustering with 20 random sets of initialization values and sufficient iterations for convergence, where the configuration with minimum loss score was reported as the best clustering. Once clustered, the cluster containing *IL32* was considered the *IL32*-cluster with its co-regulated features.

To summarize over the *IL32*-clusters from the seven case-control pairs, a feature co-clustering with *IL32* in at least one case-control pair was considered to co-cluster with *IL32* if the median of its Euclidean distances to *IL32* across all pairs was below 2.5. This selection criteria, based on median Euclidean distance to *IL32,* ensured that only those features were considered to co-cluster with *IL32* that co-clustered with it in at least 5–6 case-control pairs (**Supplementary Table 4**).

As *IFNG*-cluster in CD8+ cells and *INS*-cluster in PBMCs were of specific interest also, the Euclidean distance-based summarization over the seven case-control pairs was repeated for these genes as well (**Supplementary Table 4**).

Transcription factor binding site analysis

Overrepresented transcription factor binding motifs on the promoters of *IL32* and its co-regulated genes were analysed with updated (2018) TRANSFAC database, using the Fmatch tool with default parameters (best supported promoter, -10,000 to +1000 bp around transcription start site) and a randomly selected gene set as a background. Afterwards the *p*-values were corrected for multiple testing using the Benjamini-Hochberg method. Results with FDR < 0.05 are presented in **Supplementary Table 4**.

Single-cell RNA-seq data processing and analysis

The Chromium single-cell 3' RNA-Seq data from four Case and four Control samples (**Supplementary Table 5**) was individually preprocessed using the Cell Ranger Single-Cell Software Suite. The reads were aligned to the human reference genome (hg19) using STAR and the data from non-cellular barcodes were filtered out. Across samples, the mean raw reads per cell varied between ~57 k to ~200 k (**Supplementary Table 5**). To identify rare cell types, the cells from different samples were pooled together using Cell Ranger's multi-library aggregation algorithm where the samples were normalized using subsampling normalization. The downsampling (subsampling normalization) of sample reads after pooling retained on average ~31 k confidently mapped reads per cell (from ~59 k raw reads per cell on average). These mapped to the median of 801 genes per cell. After the pooling, expression of 32,738 genes from 20,370 cells was obtained.

For QC analysis and further exploration of the single-cell RNA-Seq data the Seurat R package was used. Firstly, all the genes expressed in less than one cell and all the cells expressing less than 200 genes or more than 4000 genes were filtered out. Furthermore, any cells containing more than 5% of mitochondrial genes or a UMI count higher than 5000 but a gene count less than 500, were also filtered out. The latter filtering steps involved filtering of cells with high UMI count but low gene count on the basis of the gene count and UMI count relationship plots following the recommendations of Seurat tool. After these quality control filtering steps, 18,396 cells expressing 20,830 genes were retained for downstream analyses.

The filtered data were normalized using Seurat's default global-scaling normalization method, 'LogNormalize', and variation from uninteresting sources (i.e., the number of molecules detected and percentage of mitochondrial genes expressed per cell) was regressed out. To capture the heterogeneity of the single-cell data and cluster the cells, a set of highly variable genes (HVGs) was selected, whose average expression was above 0.0125, and dispersion above 0.5 resulting in ~1200 HVGs in pooled cell library. Principal component analysis (PCA) was then performed on the HVGs, and the resulting top 20 PCs were used in the graph-based clustering employed by Seurat, keeping other parameters as default.

To determine the cell type represented in each cluster, markers defining the clusters were determined via differential expression algorithm implemented in Seurat, where cells of a single cluster were compared to the cells of all other clusters combined. A gene was considered a marker of a cluster if it was expressed in at least 25% of the cells of either of the two clusters and the log fold change between the cluster and all other clusters was at least 0.25. On average, one to five genes were used as markers for each cluster (**Supplementary Figure 8**). On the basis of these cluster-specific markers, no biological difference was found in two of the 13 clusters, which both represented cells from naive T cells. Therefore, they were merged into a single cluster and were labeled as naive T cells, resulting in a total of 12 different clusters.

Single-cell RNA-seq trajectory analysis

The QC filtered pooled cells from the Seurat analysis were ordered in pseudotime (i.e., placed along a trajectory corresponding to a type of biological transition, such as differentiation) using Monocle 2. The trajectory analysis was performed on cells specifically from CD4+ (CD62L+ T cells and Act. Th cells) and CD8+ (Act. GNLY+ CD8+ T cells, Act. GZMA+ CD8+ T cells and Act. prolif. GZMA+ CD8+ T cells) T-cell clusters, using the cell typing information from the Seurat analysis. In both CD8+ and CD4+ specific cell ordering, cells identified as naive T cells or T cells were also included. The trajectory analysis in Monocle 2 has three major steps.

In the first step, all genes expressed in at least 1% of the cells were used in a principal component analysis, whose resulting top PCs (six in the case of CD8+ and 11 in the case of CD4+ specific single-cell trajectory analyses) were used to initialize the t-SNE ordination of the cells. Then, the *dpFeature* function was used to cluster the cells defined in the 2-D t-SNE space. Finally, the differential gene expression test of all genes expressed in more than 10 cells was performed between the clusters defined in the previous step as a way to extract the genes that distinguish them from each other. The top 1000 significant genes were then selected for subsequent steps of the analysis. The second step reduced the dimensionality of the data using the feature genes from the previous step and availing technique called reverse graph embedding (RGE) implemented in DDRTree algorithm. In the final step, cells were ordered along the trajectory by performing manifold learning on the tree from the second step.