

Single-cell transcriptomic analysis of allergen-specific T cells in allergy and asthma

Grégory Seumois, Ciro Ramírez-Suástegui, Benjamin J. Schmiedel, Shu Liang, Bjoern Peters,
Alessandro Sette, Pandurangan Vijayanand

Supplementary Materials

Fig. S1. Experimental design and gating strategy to isolate HDM-reactive memory T_H and T_{REG} cells.

Fig. S2. Expression of differentially expressed genes in HDM⁺ cells.

Fig. S3. Single cell filtering of doublets and low HDM-reactive T cells.

Fig. S4. Single-cell clustering analysis using Seurat.

Fig. S5. Distribution of cell frequency for the 7 HDM⁺ T_H clusters for the 24 subjects.

Fig. S6. Co-expression of T_H1- or T_H17-specific signature genes.

Fig. S7. Donor and disease-groups cell distribution in each of HDM⁺ T_H clusters.

Fig. S8. Co-expression of T_HIFNR-specific signature genes in publicly available dataset.

Fig. S9. T_{REG} disease-related differences.

Fig. S10. Proportions of HDM-reactive T_{REG} subsets amongst disease groups.

Fig. S11. T_H2 single-cell analysis.

Provided as separate supplementary files:

Table S1. Study subject details (Excel spreadsheet).

Table S2. Differential gene expression analysis in bulk populations (Excel spreadsheet).

Table S3. Single cell analysis (Excel spreadsheet).

Table S4. List of genes used for Gene Set Enrichment Analysis (Excel spreadsheet).

Table S5. Raw data file with statistics (Excel spreadsheet).

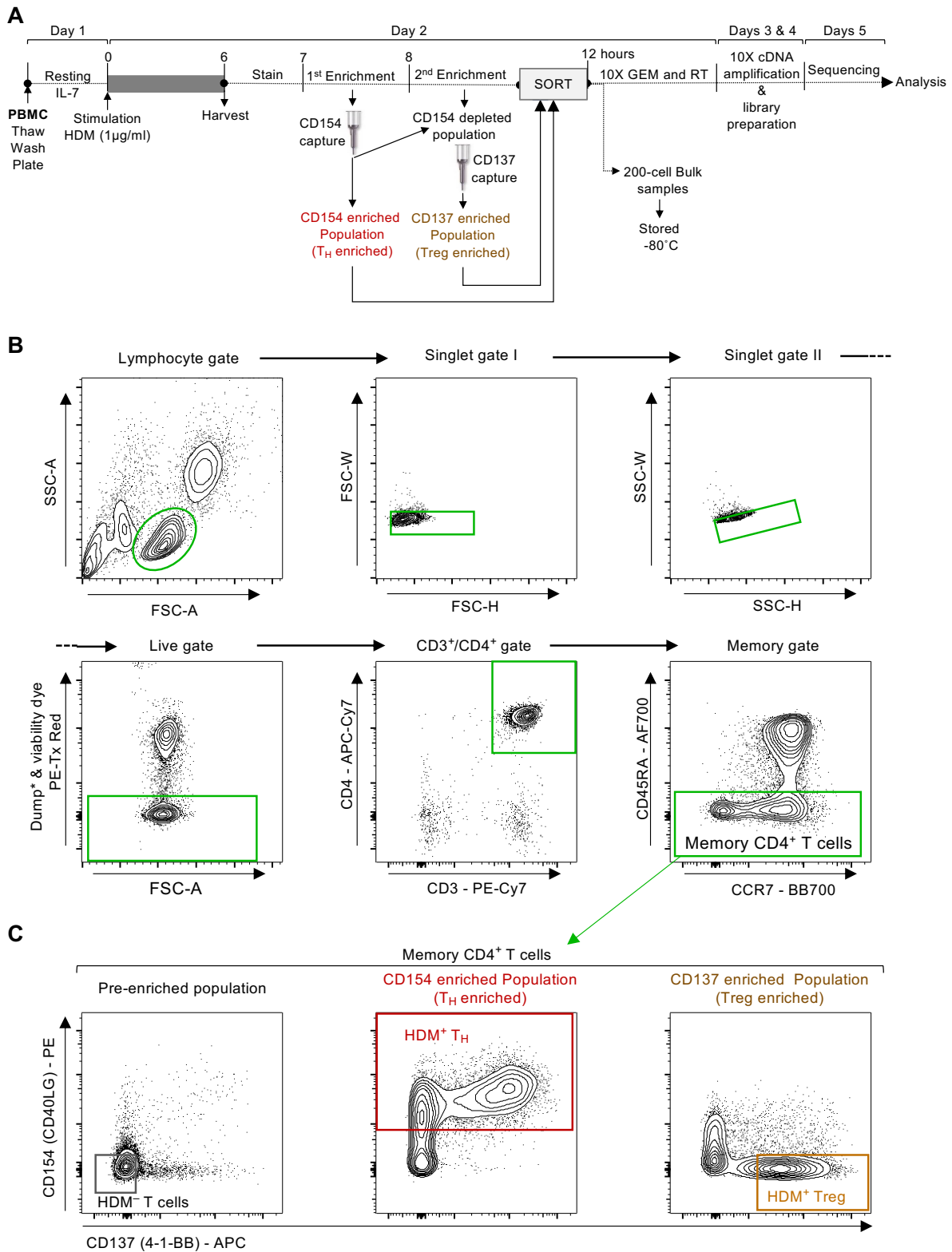


Fig. S1. Experimental design and gating strategy to isolate HDM-reactive memory T_H and T_{reg} cells. (A) Experimental design of activation with House Dust Mite peptide pool (1 $\mu\text{g/ml}$), enrichment of population, sorting, library preparation, and sequencing. (B) Gating strategy to sort lymphocytes, singlets (double gate Width vs Height forward (FSC) and side scatter (SSC)), alive, $CD3^+ CD4^+$, and memory $CD45RA^- CCR7^+$ cell population. (C) Gating strategy to sort human HDM $^+ T_H$ ($CD154^+$) and HDM $^+ T_{reg}$ ($CD154^- CD137^+$) cells from enriched populations as well as the control HDM-non-reactive ($CD154^- CD137^-$) population from pre-enriched samples.

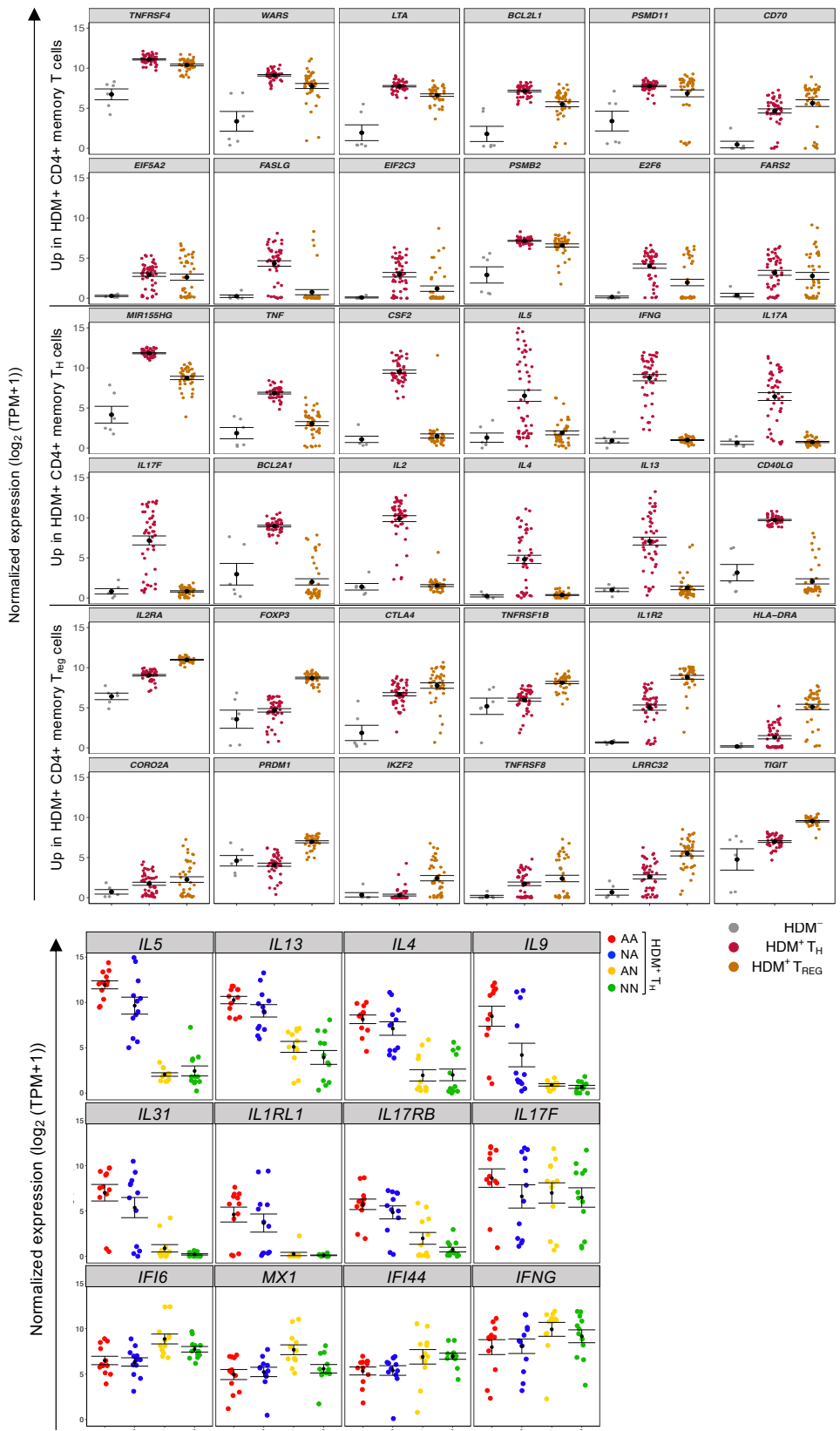


Fig. S2. Expression of differentially expressed genes in HDM⁺ cells. Top - Standard error dot plots of genes differentially up-regulated in HDM⁺ CD4⁺ memory effector (T_H) and regulatory (T_{reg}) T cells. Each dot represents a sample of 200 cells; color indicates the sorted subset. Bottom - Standard error dot plots for disease-specific differentially expressed genes in HDM⁺ CD4⁺ memory T_H cells. Each dot represents a sample of 200 cells; color indicates the sorted disease group. Black dots represent mean value expression, error bars represent standard error to the mean expression.

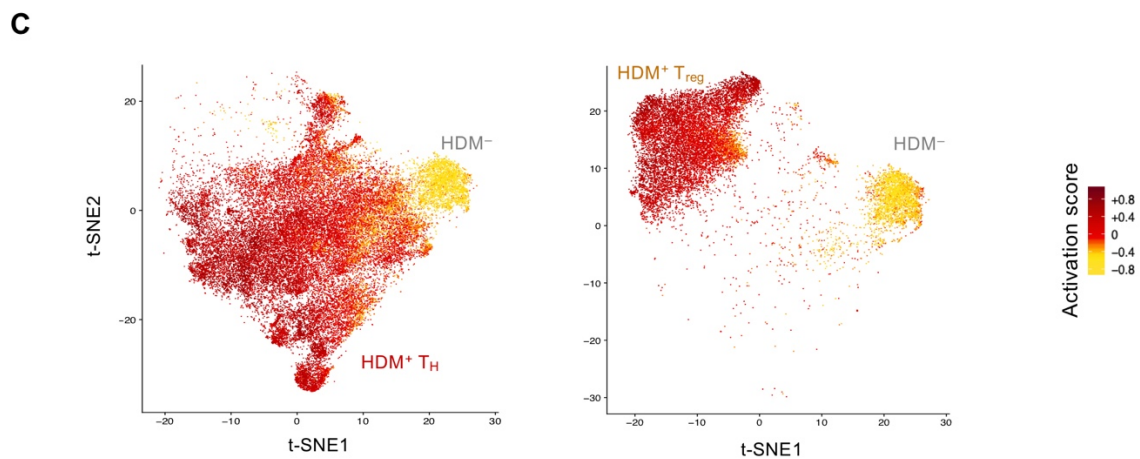
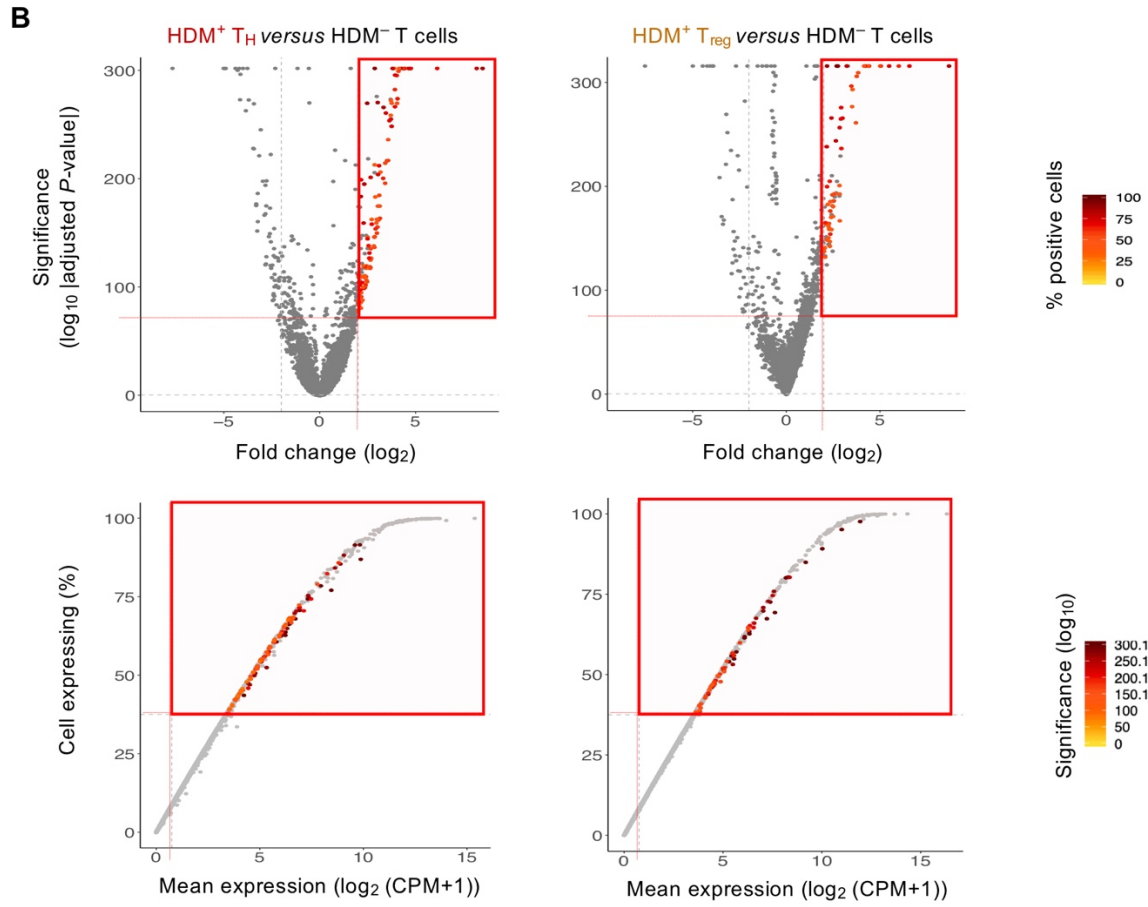
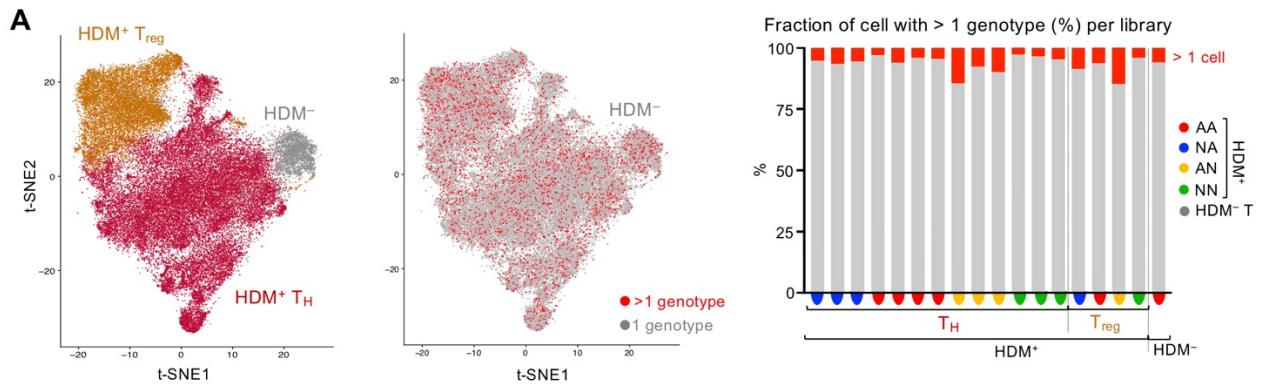


Fig. S3. Single cell filtering of doublets and low HDM-reactive T cells. (A) *Left* – t-SNE visualization of all single cell datasets. *Center* – t-SNE visualization of cells identified with more than one genotype (red, more than one cell) from the entire single-cell dataset. *Right* - Demuxlet classification of doublets (droplet with more than 1 genotype identified) as a fraction of all single cell datasets per sequencing library. (B) *Upper panels* - Volcano plots show differential gene expression analysis between HDM⁻ vs HDM⁺ T_H (left), and HDM⁻ vs HDM⁺ T_{reg} (right); red dotted lines represent threshold of selection: horizontally - Benjamini-Hochberg-adjusted *P*-value ≤ 0.05 and vertically, log₂ fold change ≥ 2. *Lower panels* – Dot plots represent the distribution of genes ordered based on their mean expression value in function of the percentage of cells expressing the given gene in HDM⁺ T_H (left), HDM⁺ T_{reg} (right). Red dotted lines represent threshold of selection: horizontally, log₂ mean expression ≥ 0.75 CPM and vertically, fraction of cells expressing the given genes > 37.5 %. Red boxes represent area of selection considered to identify “activation signature genes”. Non-grey colored dots are selected genes that responded to the 4 criteria of selection listed above. Selected genes are colored based on the % of cells expressing the given gene (upper panels) and log₂ mean of expression value (lower panels). (C) t-SNE visualization activation signature scores in HDM⁻ with either HDM⁺ T_H cells (left), or HDM⁺ T_{reg} cells (right); each dot represents a cell and is colored based on the score given by *AddModuleScore* from Seurat.

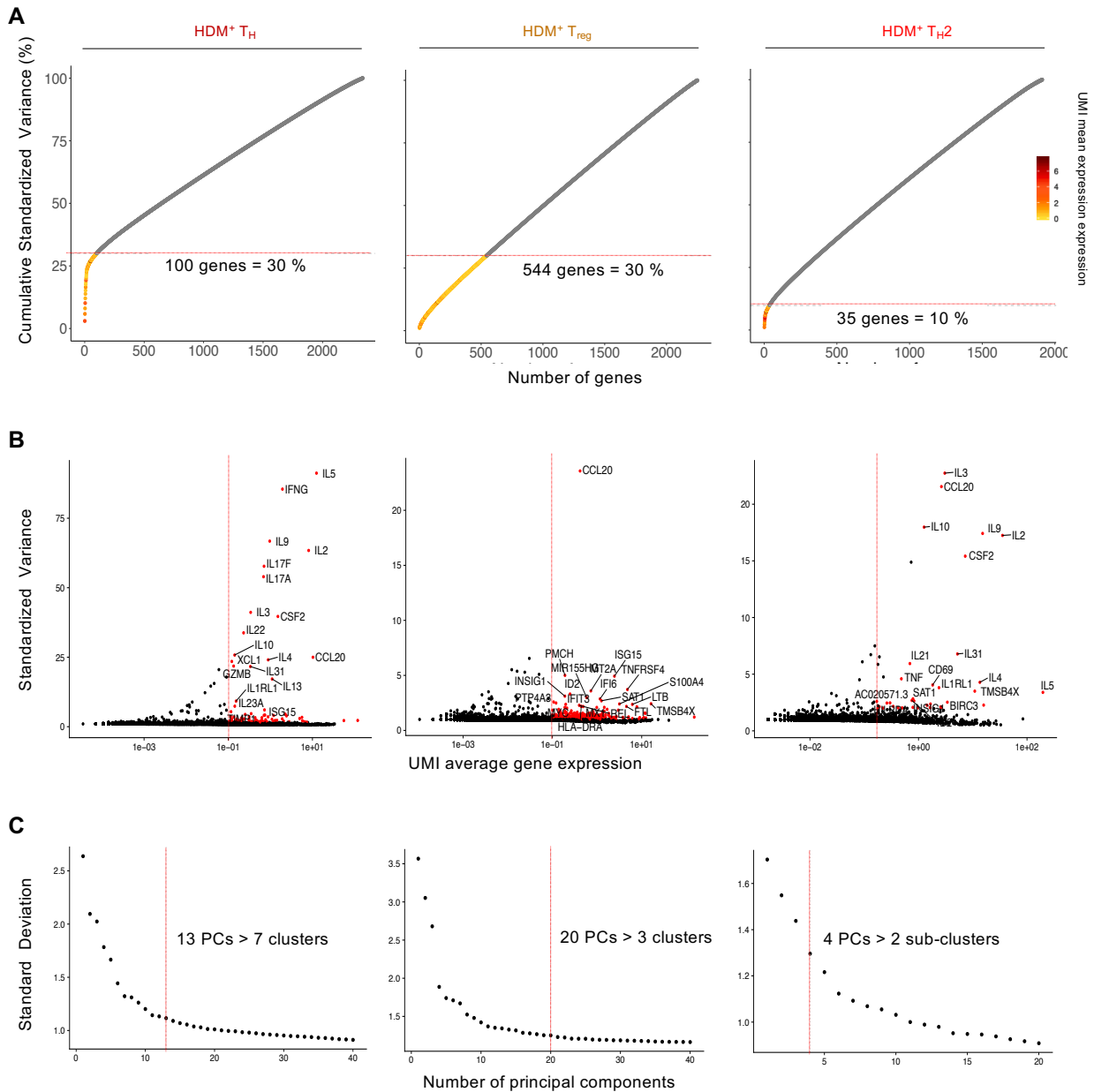


Fig. S4. Single-cell clustering analysis using Seurat. (A) Dot plots represent cumulative plot of standardized variance for the top highly variable genes within the HDM⁺ T_H cells (N = 2338, left), HDM⁺ T_{reg} cells (N = 2250, center) and HDM⁺ T_{H2} cells (N = 2085, right). Genes are ranked from highest to lowest standardized variance values. Genes below the red dotted lines are selected highest variable genes explaining 30 % (for HDM⁺ T_H cells, HDM⁺ T_{reg} cells) or 10 % (HDM⁺ T_{H2} cells) of the cumulative standardized variance for the 2085 highest variable genes. Genes are colored based on the UMI mean expression given by Seurat. (B) Scatter plots represent the distribution of genes (dots) ordered based on their UMI mean expression values in function of their individual standardized variance values. Red dots are the most variable genes with a UMI mean expression value > 0.1. The 20 most variable genes are labelled. (C) Scatter plots represent the standardized variation for first principal components (40 for HDM⁺ T_H cells and HDM⁺ T_{reg} cells; 20 for HDM⁺ T_{H2} cells). Red arrows indicate the number of PCs selected for clustering analysis in respect of Seurat methodology. Number of clusters are indicated.

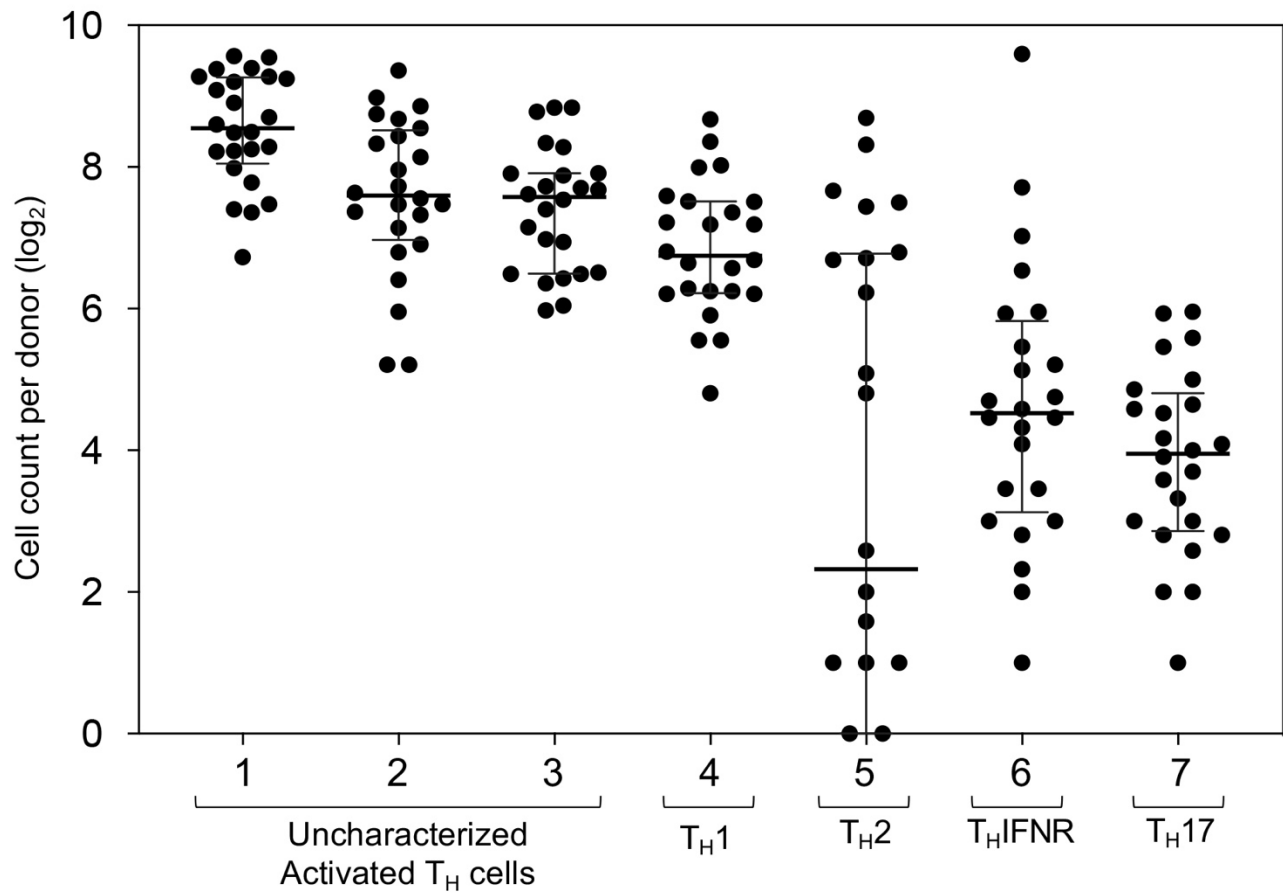


Fig. S5. Distribution of cell frequency for the 7 HDM⁺ T_H clusters for the 24 subjects. Dot plot represents the distribution of log₂ cell count per HDM⁺ T_H clusters for each subject. Central bar represents the median and error bars represent inter-quartile distribution of data.

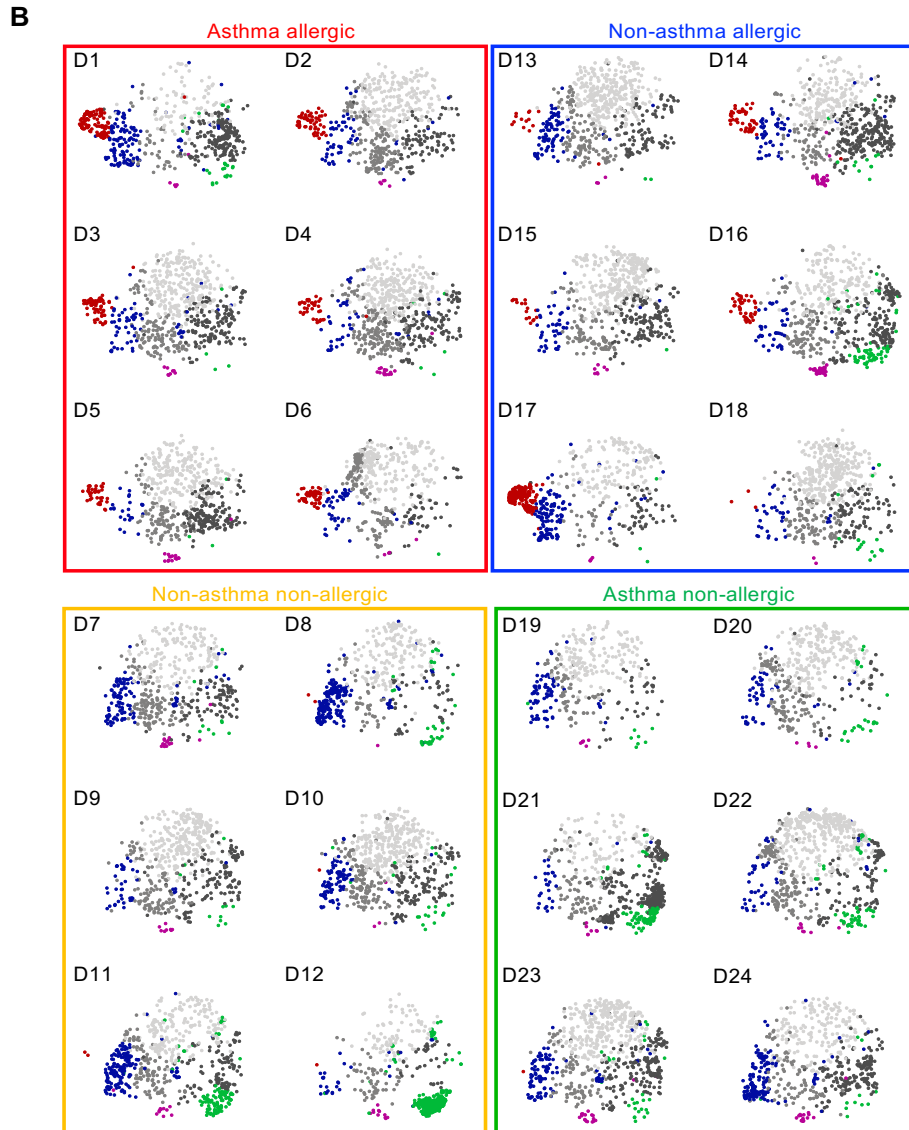
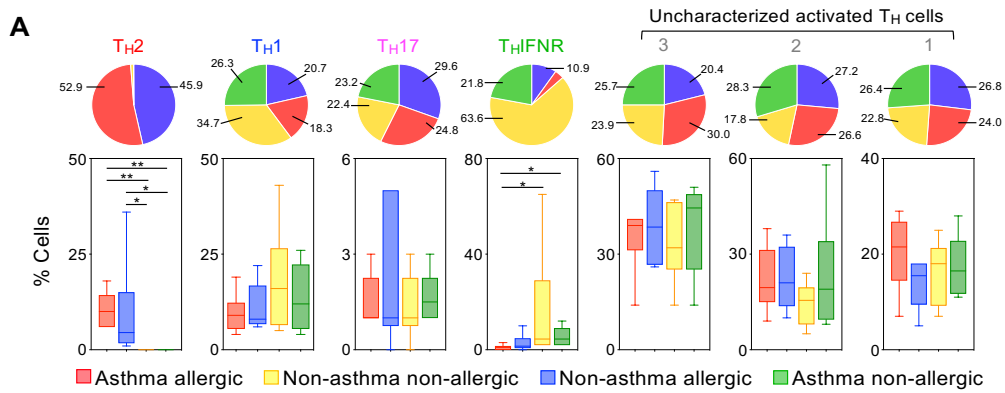


Fig. S7. Donor and disease-groups cell distribution in each of HDM⁺ T_H clusters. (A) Top - Pie charts show the proportion of cell per disease groups for each cluster. Proportions are calculated after sub-sampling disease groups to equal cell numbers (N = 3,720). **Bottom** - Box and whisker plots show percentage of cells in each HDM⁺ T_H cluster in each disease group (N = 5 or 6). Center line, median value; box, quartiles; whisker lines, 10th and 90th percentiles. *, P < 0.05; **, P < 0.01; p values calculated based pair-wise non-parametric, Kruskal-Wallis test. **(B)** t-SNE visualization of Seurat clustering analysis as shown in Figure 2B for each 6 subjects of the 4 disease groups (colored frame). Cells are colored according to cluster as in Figure 2B.

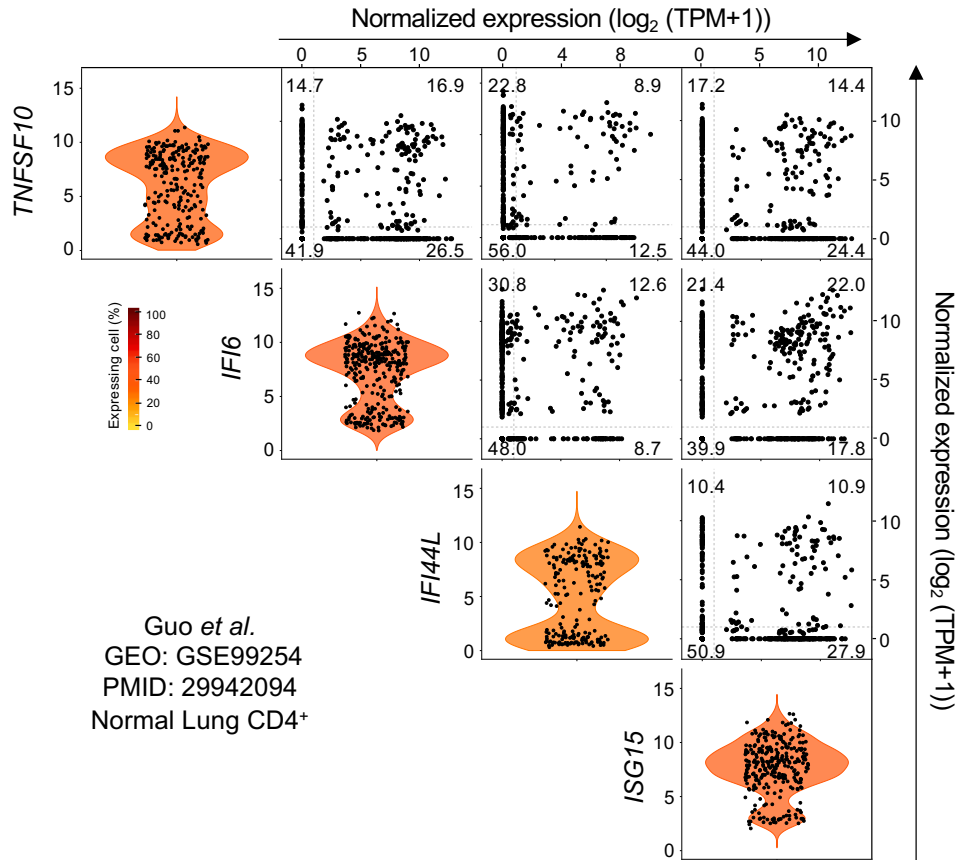


Fig. S8. Co-expression of T_H1IFNR-specific signature genes in publicly available dataset. Scatter plots show co-expression of publicly available CD4⁺ lung single-cell RNA-seqs for T_H1IFNR representative signature genes (*IFI6*, *IFI44L*, *ISG15*) as well as for *TNFSF10*. On the left of every row, violin plots show distribution of expression for individual genes (color represents the % of cells expressing the gene). Each scatter plot is split in 4 quadrants (dotted lines at 1 TPM) and percentage of cells present in each quadrant is shown.

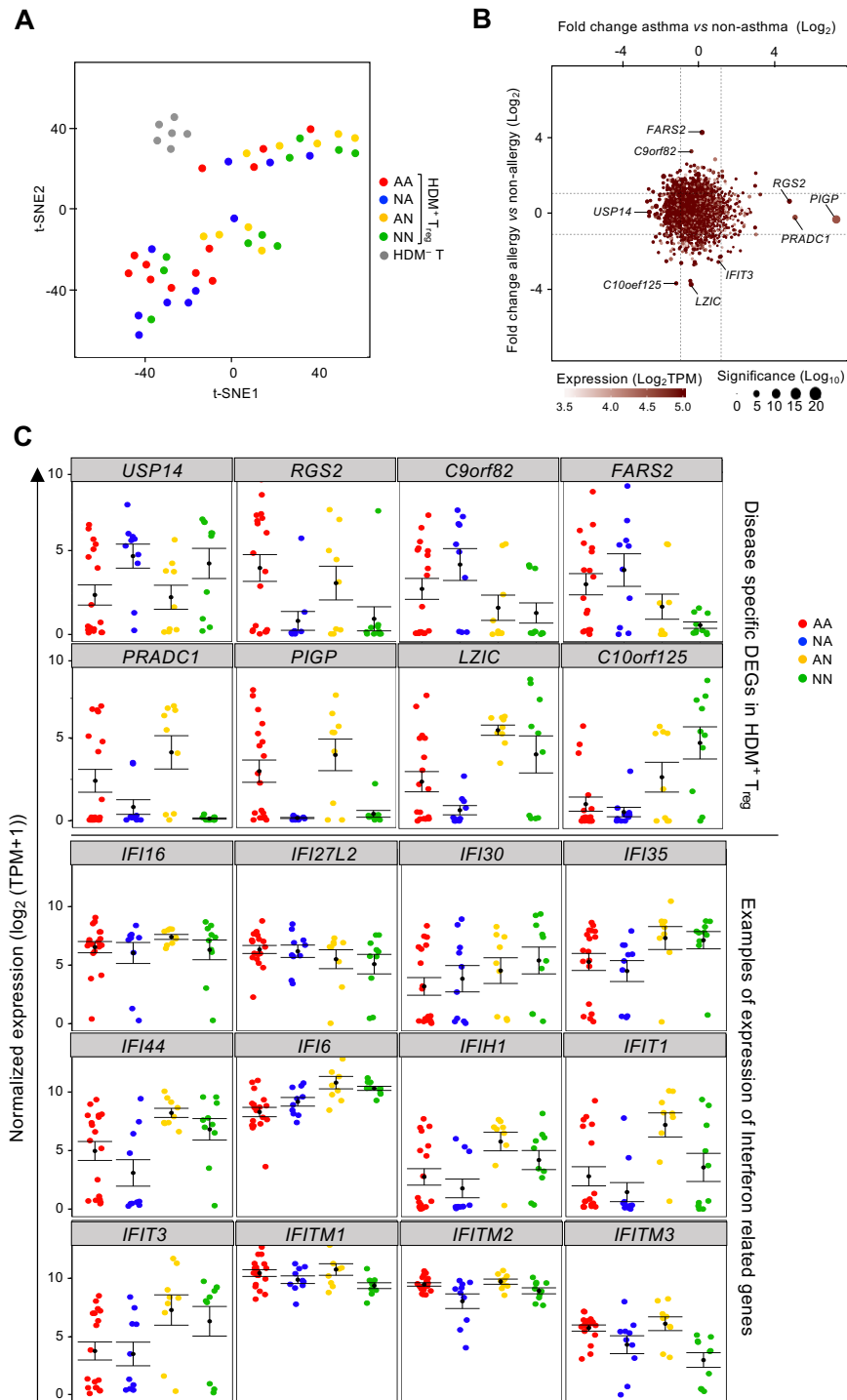


Fig. S9. T_{reg} disease-related differences. (A) t-SNE plot of bulk RNA-seq datasets for HDM⁻ T samples (N = 5) and HDM⁺ T_{reg} samples colored by disease group, where each dot represents one RNA-seq data from one subject (12 AA, 10 NA, 9 AN, 10 NN). (B) Scatter plot shows log₂-fold change in expression of significantly differentially expressed genes between asthma (N = 21) versus non-asthma (N = 20) (x axis) and allergic (N = 22) versus non-allergic (N = 19) (y axis) T_{reg} samples. (C) Standard error dot plots for disease-specific differentially expressed and interferon-related genes in HDM⁺ T_{reg} cells. Each dot represents a sample of 200 cells; color indicates the sorted disease group. Black dots represent mean value expression, error bars represent standard error to the mean expression.

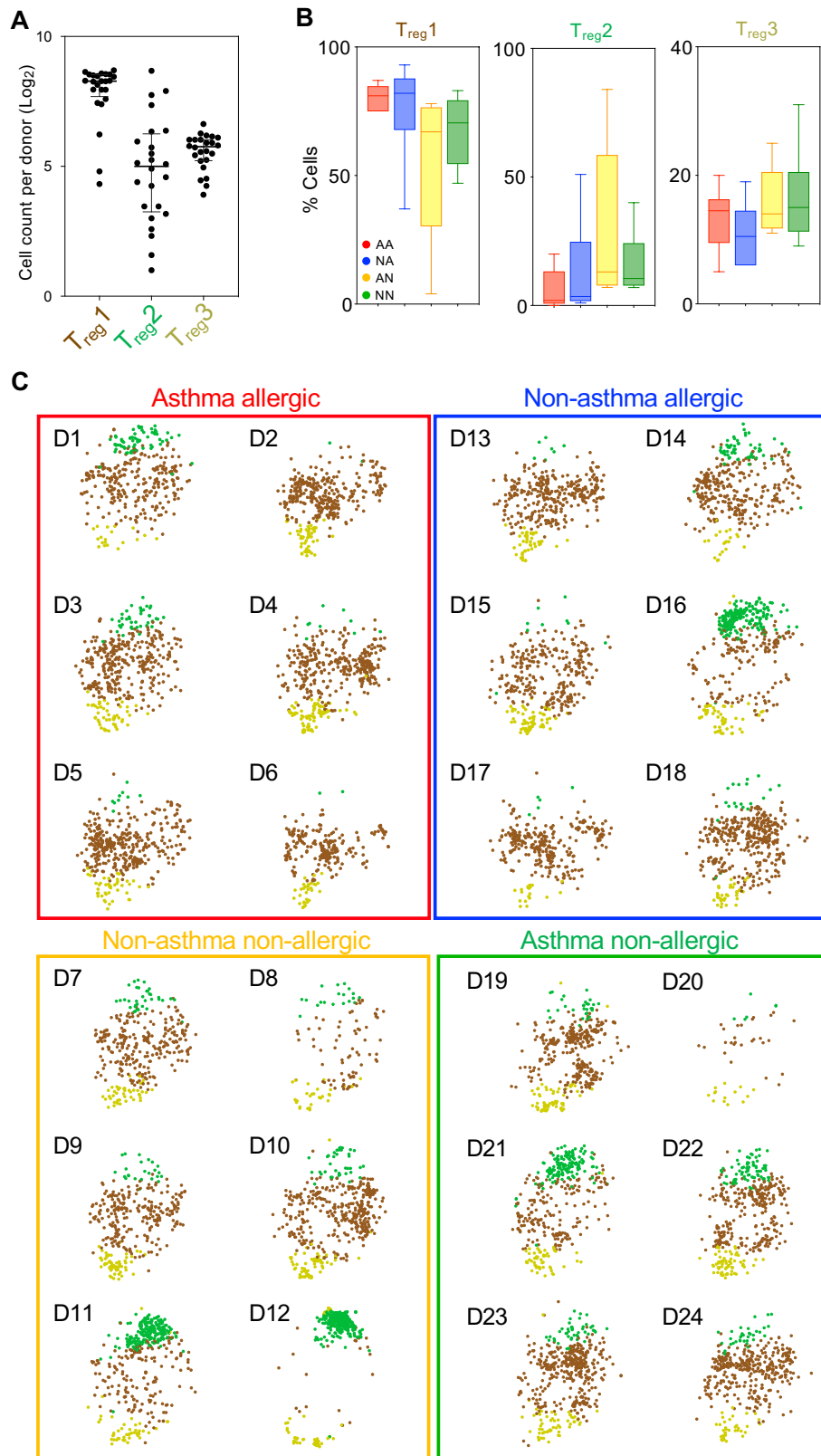


Fig. S10. Proportions of HDM-reactive T_{REG} subsets amongst disease groups. (A) Dot plot depicting the log₂ cell count per HDM⁺ T_{REG} clusters for each subject. Central bars represent the median and error bars represent inter-quartile distribution of data. (B) Pie charts show the proportion of cells per disease groups for each cluster. Proportions are calculated after sub-sampling disease groups to equal cell numbers (N = 2,180). (C) t-SNE visualization of Seurat clustering analysis as shown in Figure 4A for each 6 subjects of the 4 disease groups (colored frame). Cells are colored according to cluster as in Figure 4A.

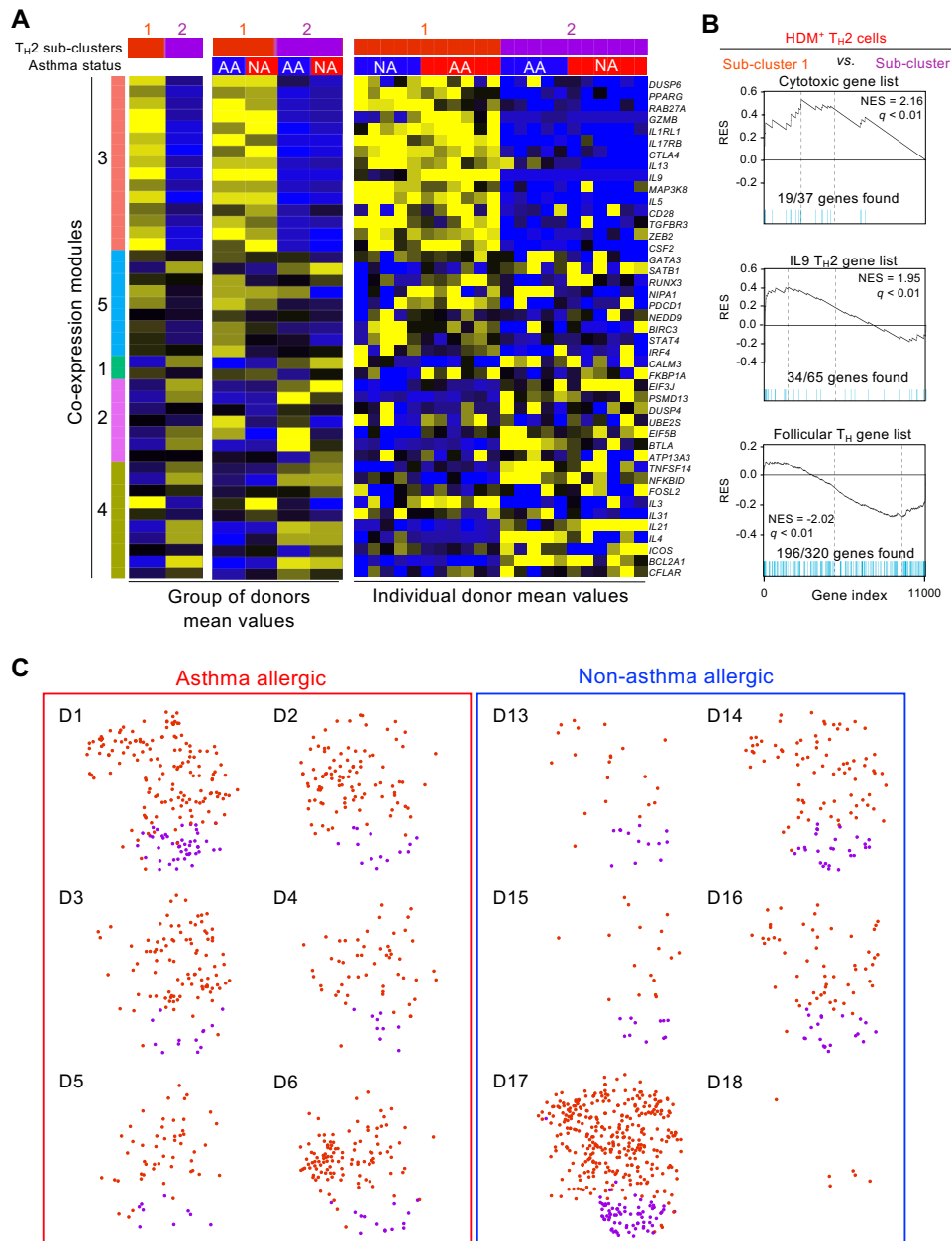


Fig. S11. T_H2 single-cell analysis. (A) Heatmap of row-wise z-score-normalized mean expression of selected module genes. From the left, columns represent the average expression per T_H2 sub-clusters, per disease group within each cluster, and per donor. (B) Gene Set Enrichment analysis of selected lists when comparing T_H2 sub-clusters. (C) t-SNE visualization of Seurat clustering analysis as shown in Figure 5C for each 6 subjects of the 2 disease groups (colored frame). Cells are colored according to cluster as in Figure 5C.