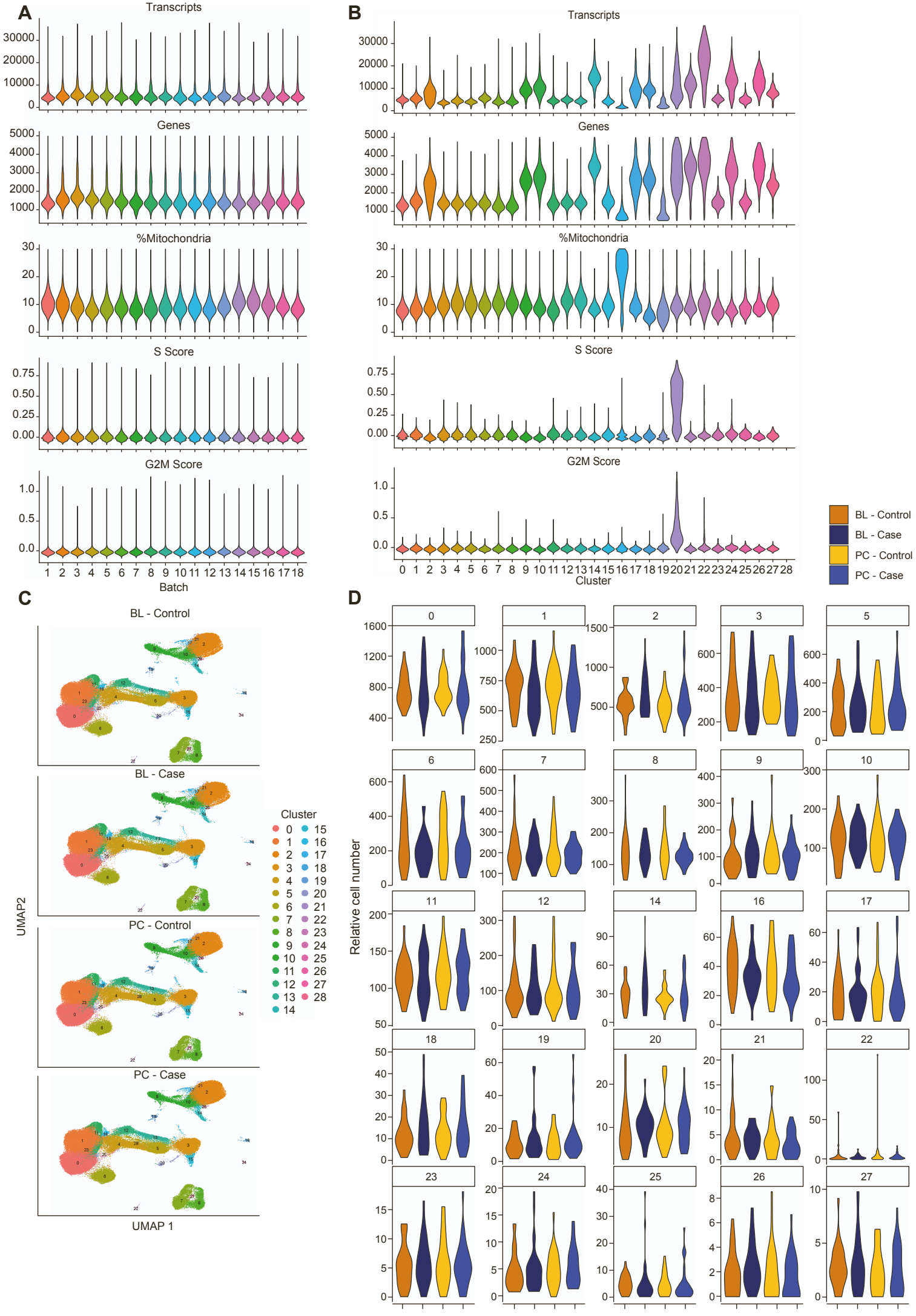
Cell Reports Medicine, Volume *5*

Supplemental information

Single-cell transcriptomics of the immune system in ME/CFS at baseline and following symptom provocation

Luyen Tien Vu, Faraz Ahmed, Hongya Zhu, David Shing Huk Iu, Elizabeth A. Fogarty, Yeonui Kwak, Weizhong Chen, Carl J. Franconi, Paul R. Munn, Ann E. Tate, Susan M. Levine, Jared Stevens, Xiangling Mao, Dikoma C. Shungu, Geoffrey E. Moore, Betsy A. Keller, Maureen R. Hanson, Jennifer K. Grenier, and Andrew Grimson

Supplemental figure 1

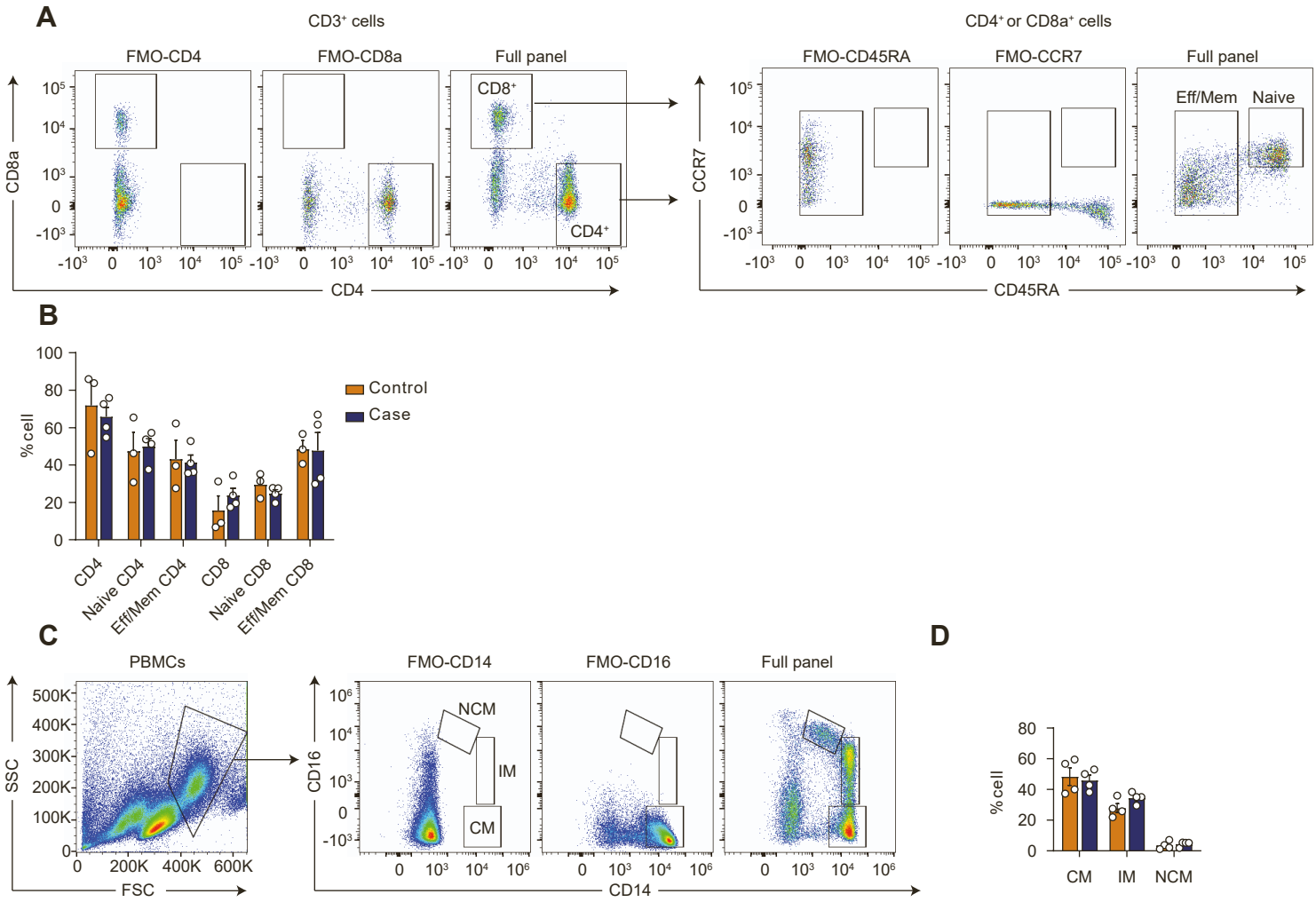


Supplemental Figure 1. Single cell transcriptomics of the ME/CFS immune system

Related to Figure 1

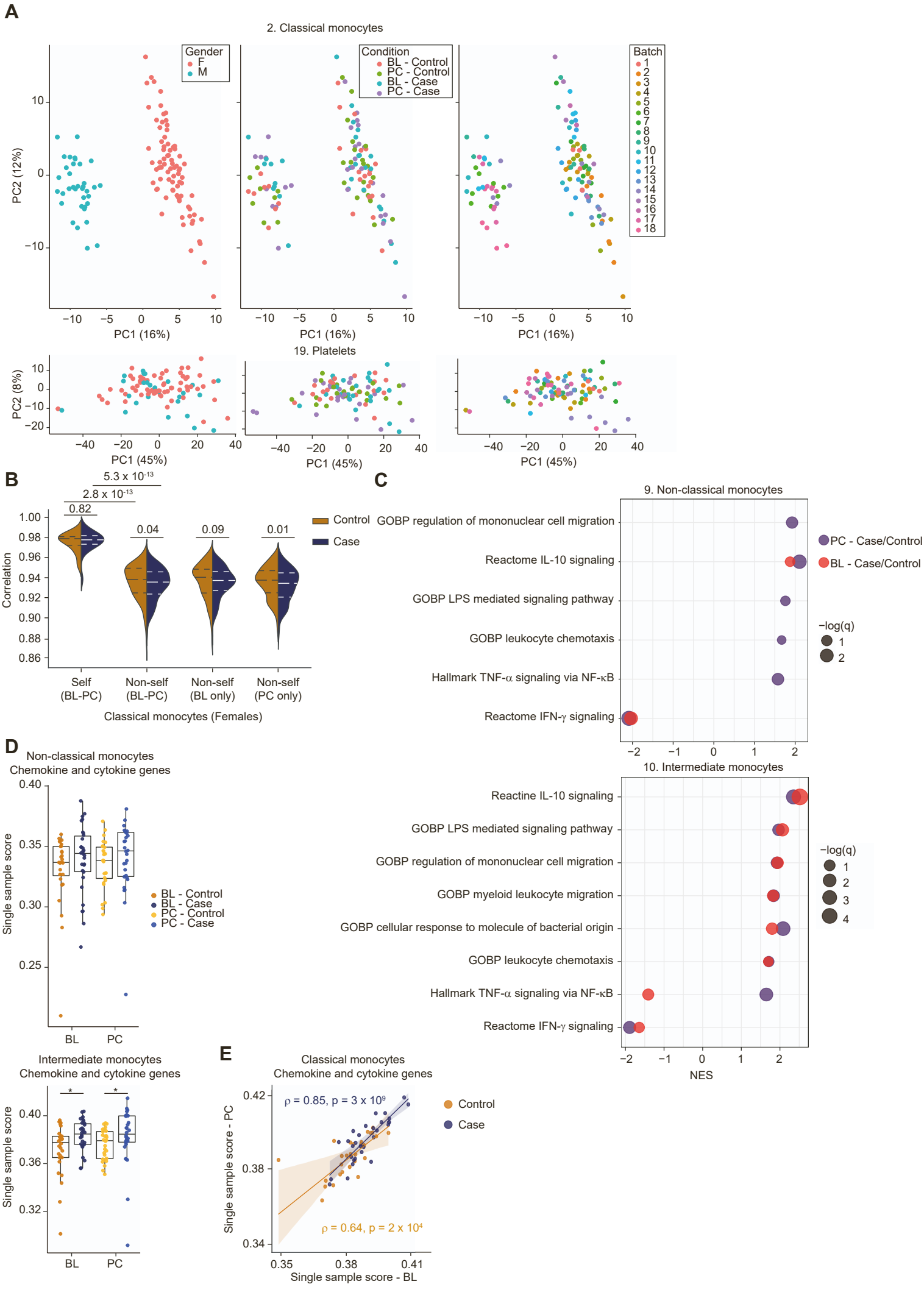
(A) Violin plot showing quality control metrics (from top to bottom: transcripts per cell, genes per cell, percent mitochondrial reads per cell, S Score and G2M score) for each batch of samples (x-axis) processed on the 10x Genomics Chromium instrument. (B) Violin plot showing quality control metrics for each cluster (x-axis). (C) UMAP split by condition (Case-Baseline, Case-post-CPET, Control-Baseline, Control-post-CPET), showing representation of all clusters in each condition. (D) Relative cell numbers between cohorts across all clusters (numbered as per Figure 1D) except clusters 4, 13, and 15 (see Figure 1F). Panels represent data from 28 healthy controls and 30 ME/CFS cases.

Supplemental figure 2



Supplemental Figure 2. Flow cytometry analysis of T cell and monocyte subsets
Related to Figure 1
(A) Representative flow cytometry analysis of CD3+ T cells from PBMCs. Naïve and effector/memory (Eff/Mem) cells were separated based on CD45RA and CCR7 expression (right panel). For correct gating, we also stained cells with antibody cocktails that omit a single antibody per channel (fluorescent minus one – FMO). Left and middle panel are FMO controls omitting the indicated antibody. (B) Percentage of different T cell subsets at baseline between controls and ME/CFS cases, measured for 4 individuals per group; bar graphs represent mean ± SEM. (C) Representative flow cytometry analysis of monocytes from PBMCs. Classical (CM), intermediate (IM) and non-classical monocytes (NCM) were identified based on CD14 and CD16 expression; otherwise as described in panel E. (D) Percentage of different monocyte subsets in controls and ME/CFS cases at baseline, measured for 4 individuals per group; bar graphs represent mean ± SEM.

Supplemental figure 3

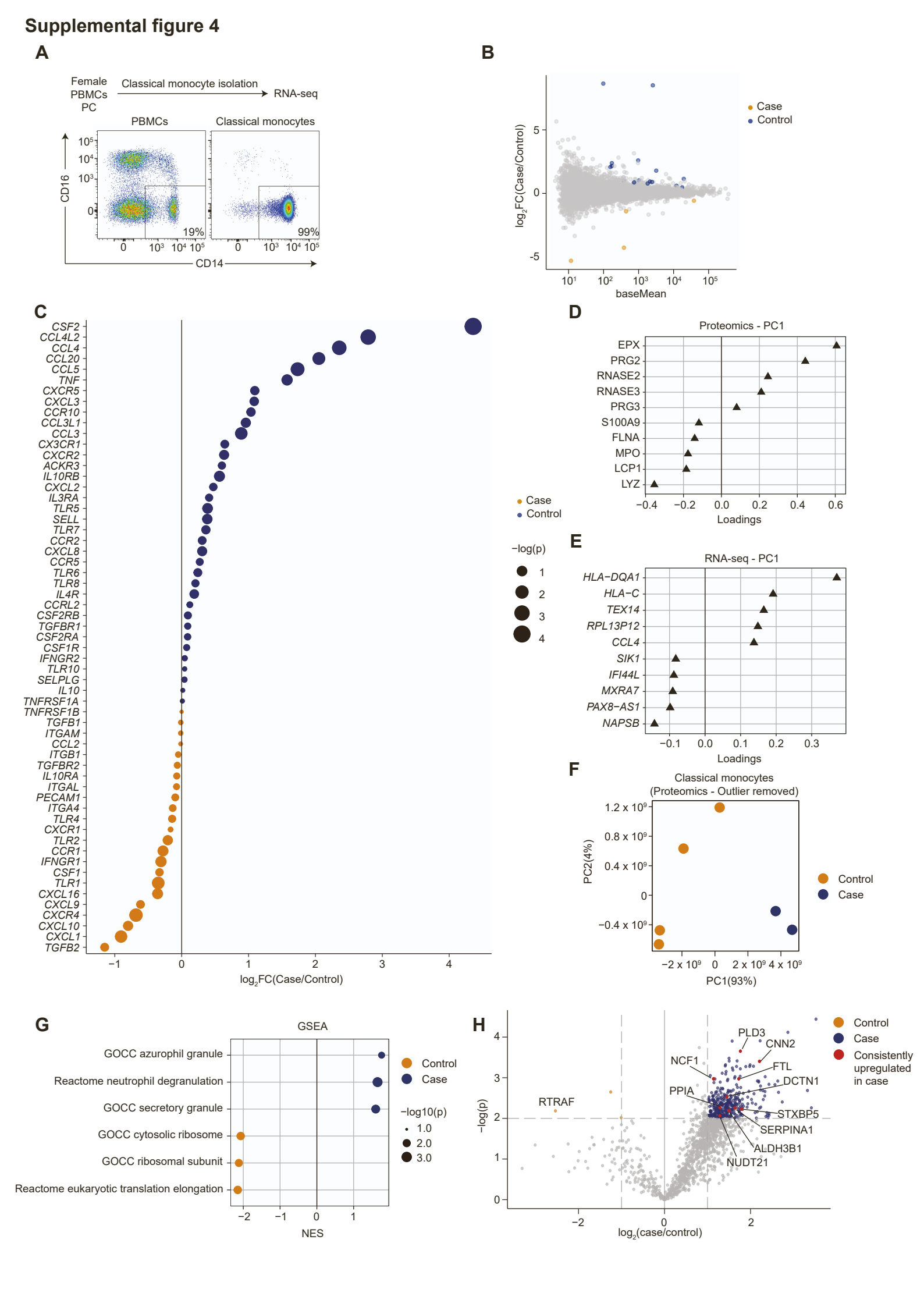


Supplemental Figure 3. Dysregulation of immune cells in ME/CFS

Related to Figure 2

(A) PCA plots for pseudobulk analysis of gene expression for cluster 2 (classical monocytes, top) and cluster 19 (platelets, bottom). Left panels are colored by sex, center panels are colored by condition (Case-Baseline, Case-post-CPET, Control-Baseline, Control-post-CPET), and right panels are colored by batch (Chromium processing). In most clusters and as shown for cluster 2, sex explains the first principal component of variation in the gene expression profiles. Cluster 19 is unique in not showing a strong sex bias. (B) Distribution of pairwise correlation (Spearman) values for cluster 2 gene expression profiles. Correlations were compared for each individual at the two time points (Self BL-PC) and between individuals at different time points (Non-self BL only and Non-self PC only) and between time points (Non-self BL-PC). Mann-Whitney U test was conducted to compare the distributions between controls and cases within each subset and between controls (BL-PC) and cases (BL-PC) with p-value shown on top of each graph. (C) GSEA results for comparisons of case versus control cohorts at baseline or post-CPET for clusters 9 (top) and 10 (bottom) for the same gene sets shown in Figure 2D, when the result is statistically significant (q-value < 0.05). Dots are sized to denote significance and colored to indicate the timepoint for the comparison of case versus control (Baseline or post-CPET); x-axis indicates normalized enrichment score (NES). (D) Single-sample scores for pseudobulk profiles for clusters 9 (top) and 10 (bottom) generated using the same list of genes as in Figure 2E (leading-edge genes from Figure 2D). * p-value < 0.05. (E) Correlation (Spearman) for single-sample scores for each individual, comparing baseline (x-axis) to post-CPET (y-axis). All panels represent data from 28 healthy controls and 30 ME/CFS cases.

Supplemental figure 4

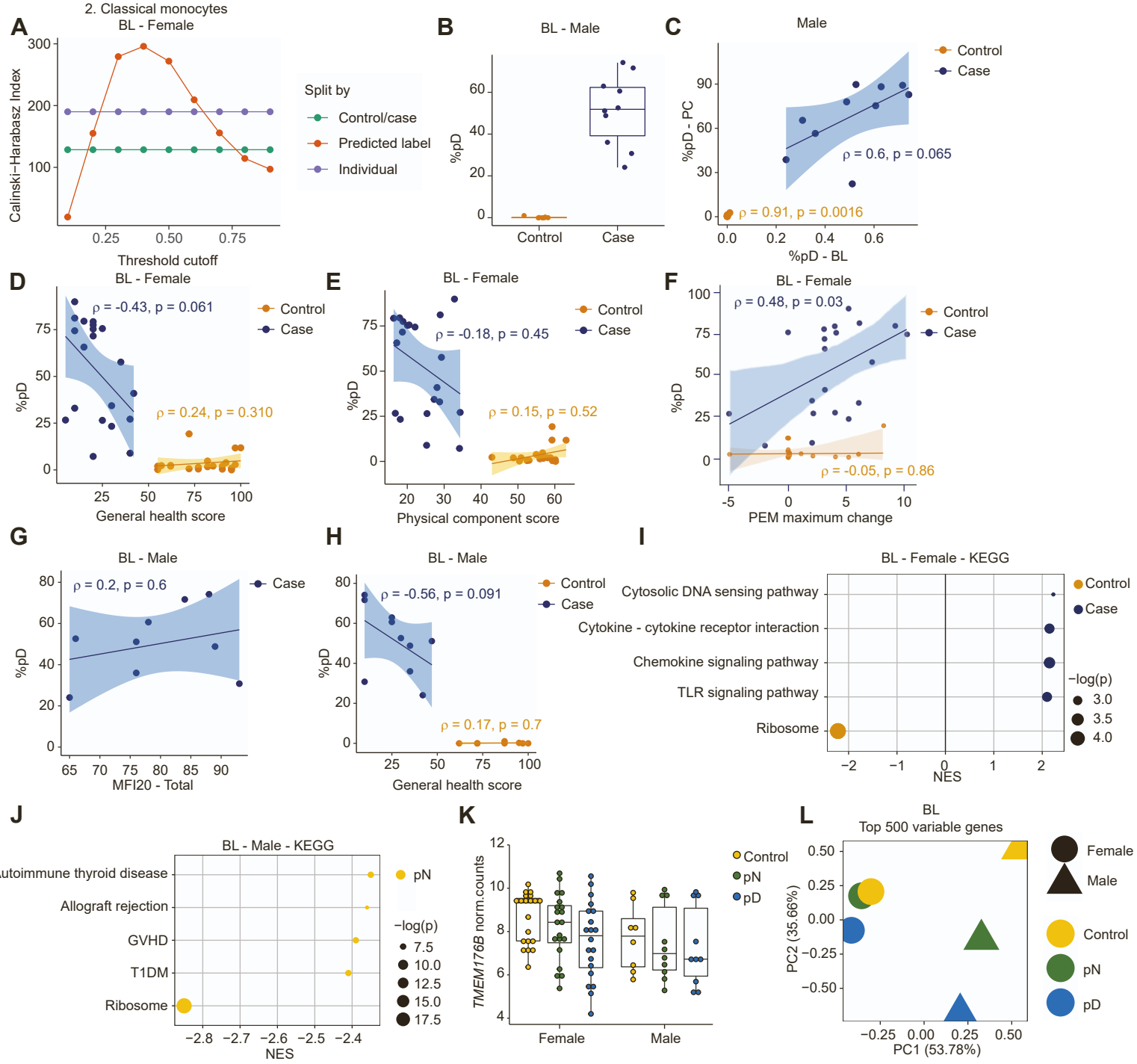


Supplemental Figure 4. Complete transcriptomics of classical monocytes

Related to Figure 3

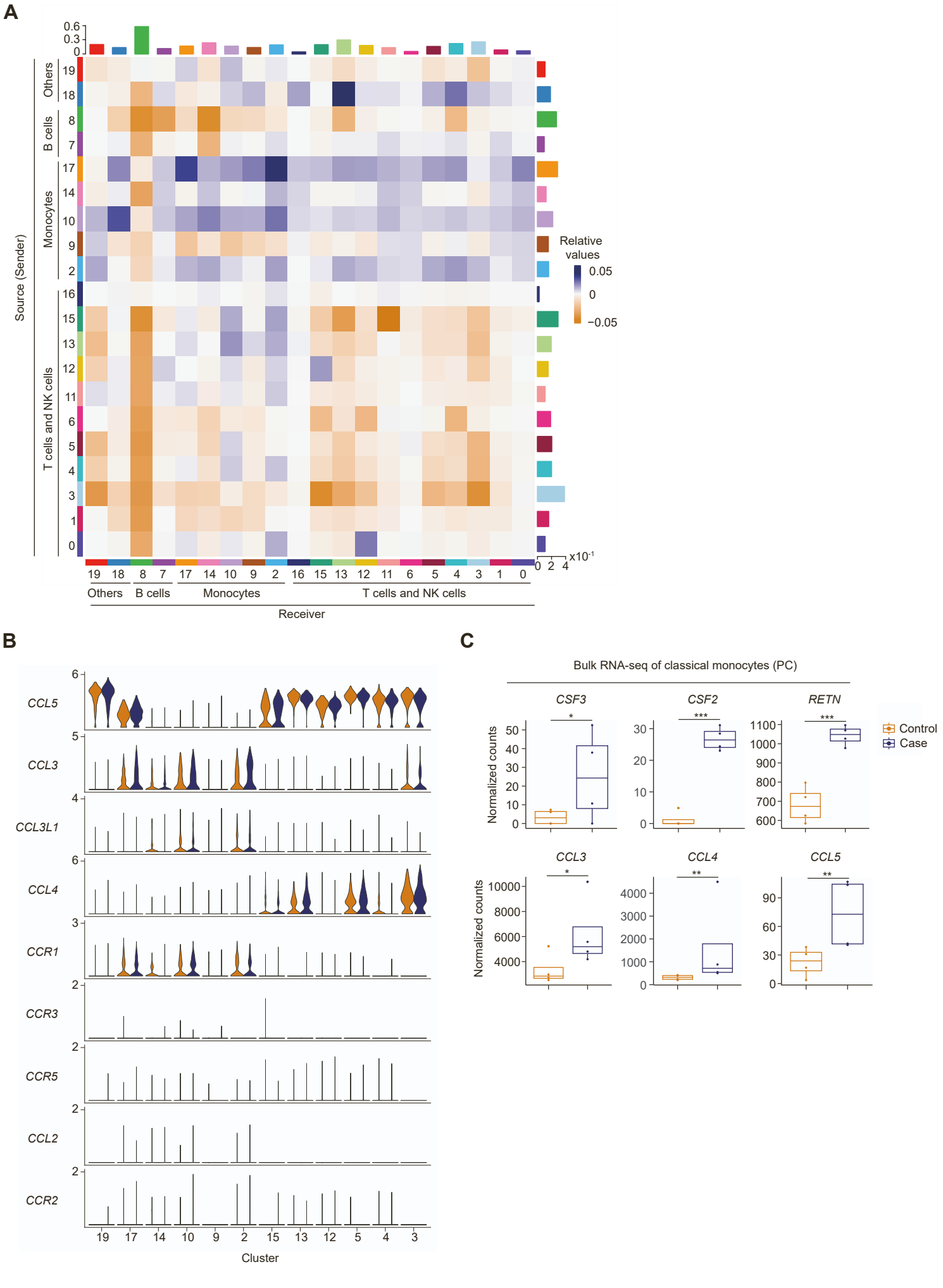
(A) Purification of classical monocytes from PBMCs. PBMC samples from female cases and controls collected post-CPET was utilized for classical monocyte isolation and bulk RNA-seq (top); all individuals were distinct from those profiled with scRNA-seq. Flow cytometry analysis confirmed enrichment of classical monocytes (CD14+CD16-, bottom). (B) MA plot showing average expression of genes (x-axis, DEseq2 baseMean) and the log_2-fold change between case and control groups (y-axis). Dots are color-coded to indicate statistically significant differential expression and the group with higher relative expression (case, control; blue, yellow, respectively) at adjusted-$p < 0.05$. (C) Differential expression of genes associated with monocyte migration and differentiation for classical monocytes collected as in Figure 3A between case and control groups. Dots are sized to denote significance (p-values) and colored to reflect the group with higher relative expression (case, control; blue, yellow, respectively); x-axis indicates log_2-fold change (case/control). (D)Top 5 genes contributing to positive and negative side of PC1 of proteomic PCA (Figure 3D). (E) Top 5 genes contributing to positive and negative side of PC1 of RNA-seq PCA (Figure 3A). (F) PCA of two patient and four control proteome profiles from classical monocytes, excluding outlier sample in Figure 3D. (G) Significantly enriched gene sets between patient and control cohorts by GSEA, excluding outlier sample in Figure 3D. Dots are sized to denote significance (adjusted p-values); x-axis indicates NES. (H) Differentially expressed proteins between patients and control cohorts, excluding outlier sample in Figure 3D. Red dots represent consistently upregulated proteins in cases between analyses (see Figure 3F). y-axis represents adjusted p-values. 2 cohorts of 4 healthy controls and 4 ME/CFS cases (all females) at baseline and post-CPET were chosen for proteome (panels D and F-H) and transcriptome (panels A-C and E) profiling, respectively.

Supplemental figure 5



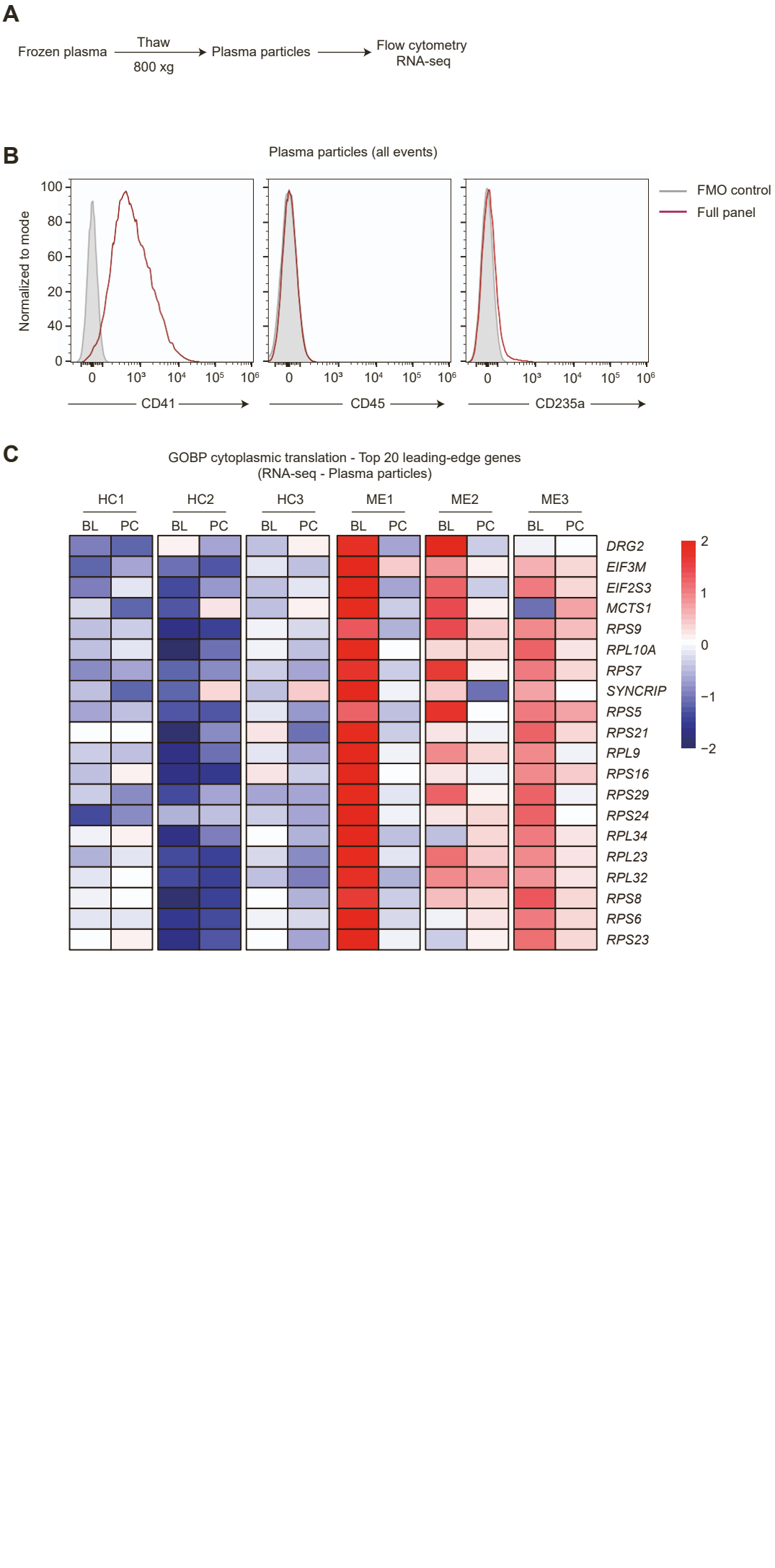
Supplemental Figure 5. Heterogeneity in classical monocyte cells from ME/CFS patients
Related to Figure 4
(A) Calinski-Harabasz Index comparing performance of classification for using different cutoffs for predictions, using female classical monocytes at baseline. The threshold with the maximum CH Index value for the algorithm was selected (0.4). (B) Performance of grouping predicted normal and diseased cells for male samples: percentage of cells predicted as diseased (pD) per sample; males only, examined at baseline. (C) Correlation (Spearman) between percentage of pD cells per sample, comparing paired baseline and post-CPET values per individual; males only. (D) Correlation (Spearman) between general health score and percentage of pD cells per sample; females only, examined at baseline. (E) Correlation (Spearman) between SF-36 physical component score and percentage of pD cells per sample; females only, examined at baseline. (F) Correlation (Spearman) between maximum change in PEM symptom severity and percentage of pD cells per sample; females only, examined at baseline. (G) Correlation (Spearman) between MFI-20 total score and percentage of pD cells per sample; males only, examined at baseline. (H) Correlation (Spearman) between general health score and percentage of pD cells per sample; males only, examined at baseline. (I) Top 5 gene sets differentially enriched in GSEA comparing average $\log_2$-fold change (FindMarkers) in case and control cells from females at baseline. Dots are color-coded to indicate enrichment in case (blue) or control (yellow) cells and sized to indicate corrected P values. (J) Top 5 gene sets differentially enriched in GSEA comparing predicted normal (pN) and pD cells from males at baseline. Dots are color-coded to indicate enrichment in pN (yellow) cells and sized to indicate corrected P values. (K) Expression of *TMEM176B* (y-axis) across indicated groups (X-axis) per sample, aggregating expression of cells from controls (yellow), pN cells from cases (green), and pD cells from cases (blue), partitioned by sex, all at baseline. (L) PCA (principal components 1 and 2) of pseudobulk profiles from aggregated subsets of cells from controls (yellow), pN cells from cases (green), and pD cells from cases (blue), partitioned by sex, all at baseline. Panels A, D-F, and I represent data from the female cohort with 20 healthy controls and 20 ME/CFS cases. Panel B, C, G, H, and J represent data from the male cohort with 8 healthy controls and 10 ME/CFS cases. Panels K and L represent data from the full cohort with partition based on sex. All panels represent data at baseline except C which compares baseline to post-CPET.

Supplemental figure 6



Supplemental Figure 6. Intercellular signaling in the circulating ME/CFS immune system
Related to Figure 5
(A) Heatmap of differential interaction strengths between cell types at baseline, from female samples following down-sampling. Top bar plot indicates aggregate interaction strength of incoming signals to indicated clusters (X-axis); right bar plot indicates aggregate interaction strength of outgoing signals from indicated clusters (Y-axis). Positive values (blue) indicate increased signaling strength in cells from ME/CFS patients compared to controls; negative values (orange) indicate decreased signaling strength. (B) Violin plots of log-normalized expression levels for genes annotated under the CCL pathway in CellChatDB, per cluster, showing female cells at baseline in the control (orange) and ME/CFS (blue) cohorts. (C) Expression levels of indicated genes in RNA-seq of female classical monocytes at post-CPET for cases and controls (see Figure 3). * p-value < 0.05, ** p-value < 0.01, *** p-value <0.001. Panels A and B represent data from the female cohort with 20 healthy controls and 20 ME/CFS cases at baseline.

Supplemental figure 7



Supplementary figure 7. Plasma particle isolation

Related to Figure 6

(A) Schema describing plasma particle isolation using low speed centrifugation for flow cytometry and RNA-seq, (B) Representative histograms of platelet (CD41), leukocytes (CD45), and red blood cells (CD235a) on the surface of plasma particles, compared between full panel and fluorescence minus one (FMO; omitting the antibody identified on the x-axis) control panels. (C) Heatmap of the top 20 leading-edge genes from RNA-seq of plasma particles between 3 healthy controls (HC) and 3 ME/CFS cases (ME) from the enriched GSEA termed “GOBP cytoplasmic translation”. Gene expression values are row-normalized. Panel C represents data from a cohort of 3 healthy controls and 3 ME/CFS cases (all females) at both time points.