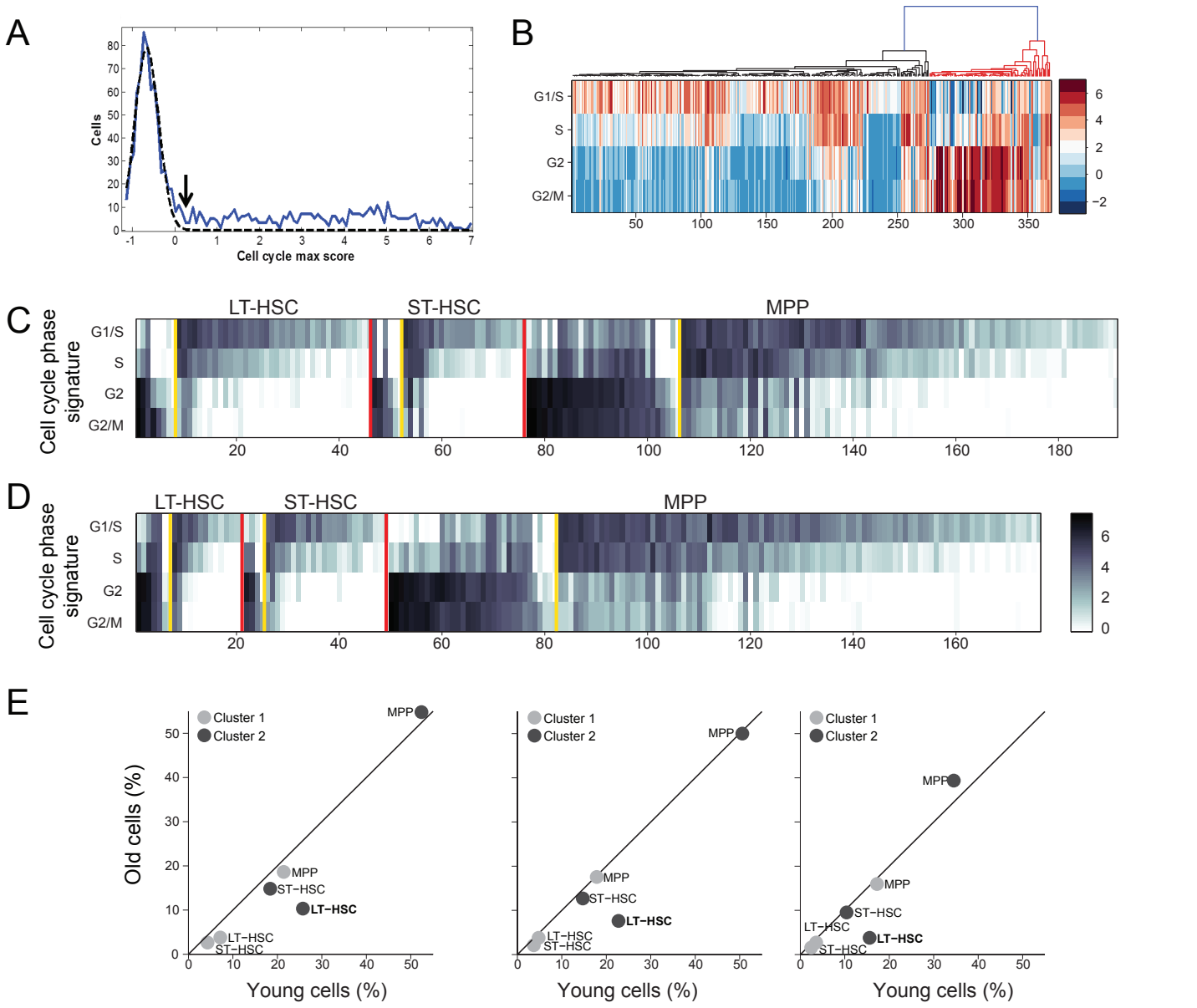

Figure S3. Clustering cells by cell cycle expression programs

(A) Signature based distinction of cycling and non-cycling cells. Shown is the distribution of the maximum cell scores (of four cell cycle signatures) (blue) and a least-squares Gaussian regression (black) performed with zero weights for all points above zero in order to fit the main peak while disregarding the long right tail. The fitted Gaussian distribution was used to define a p-value for each cell and the black arrow represents the corresponding threshold with FDR=0.05. (B) Two robust clusters of cycling cells. Shown is a dendrogram (top) generated by average-linkage hierarchical clustering of the 367 cycling cells from (A) using correlation over the four cell cycle scores (average expression (log2(TPM+1) in each cell; heatmap bottom) as a distance metric (Methods). A strong separation to two clusters is apparent that preferentially express G1/S+S genes (black portion of the dendrogram) and G2/M+M genes (red portion of the dendrogram), respectively. (C, D) Cell cycle state in different cell types and ages. Shown are the scores (greyscale bar) for each of the four cell cycle signatures (rows) in each cell type (columns; separated by red vertical lines) from either young (C) or old (D) cycling cells. Within each cell type, cells are separated by their membership of the two major clusters in (B) (separated by yellow lines). (E) The depletion of G1/S cells in old LT-HSCs is observed in the cell cycle two-cluster analysis and is robust to the threshold of identifying cycling cells. The analysis in (A) was repeated with different thresholds for classifying cells as cycling, corresponding to an FDR of 0.4, 0.05, or 10-7, followed by hierarchical clustering as in (B) and definition of two clusters (a G2/M Cluster 1 and a G1/S Cluster 2). Each graph shows percentage of cells from each cell type in each cluster in young vs. old. Three graphs represent data for three FDR thresholds (from left to right 0.4, 0.05, 10-7). Despite the wide range of thresholds, the percentage of cells in cluster 1 was almost not affected. The percentage of cells in cluster 2 was affected, but the impact of aging was qualitatively robust. The frequency of cluster 2 cells in young LT-HSCs varied from 16% to 26%, and in old LT-HSCs from 4% to 10%, and these differences between young and old were significant (P<0.01) in all cases, while none of the other aging-dependent differences were significant.