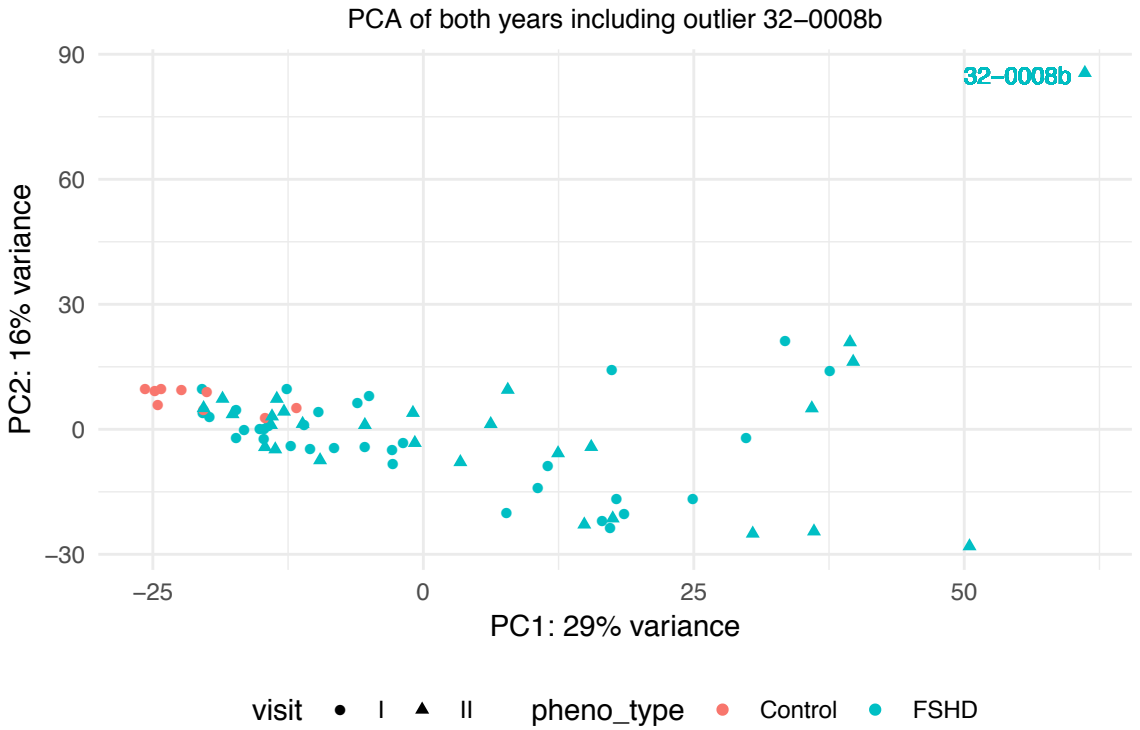
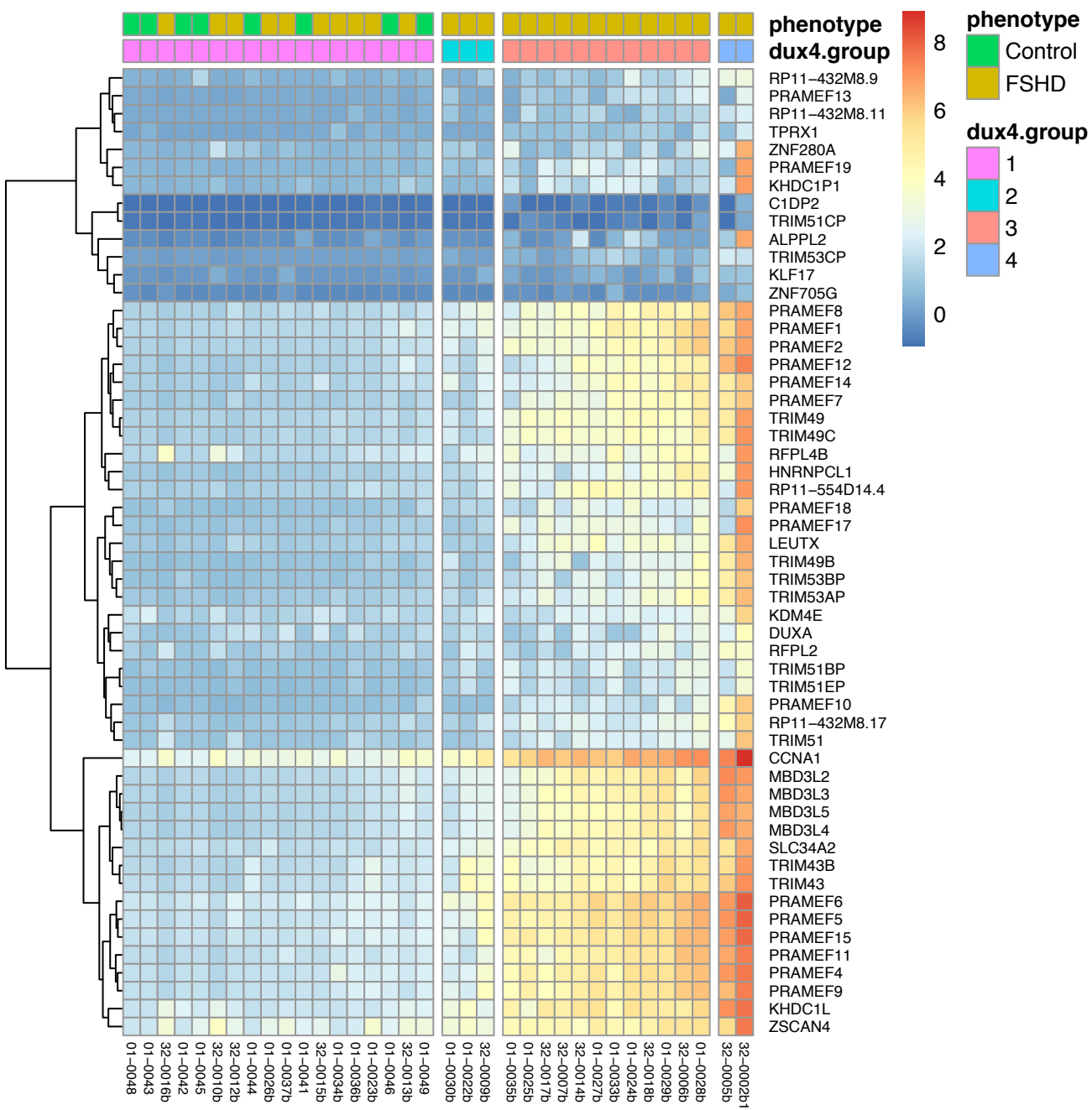
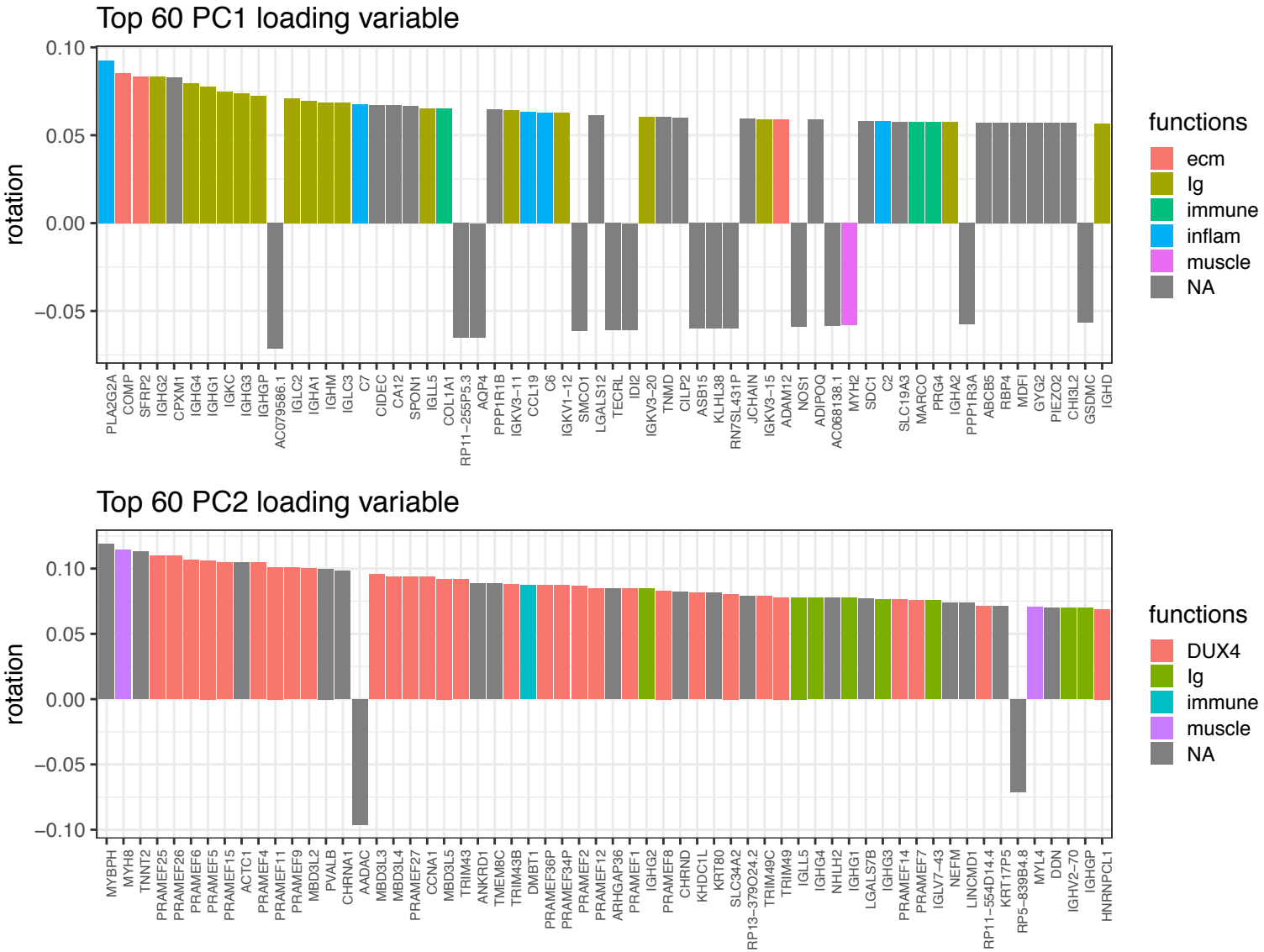


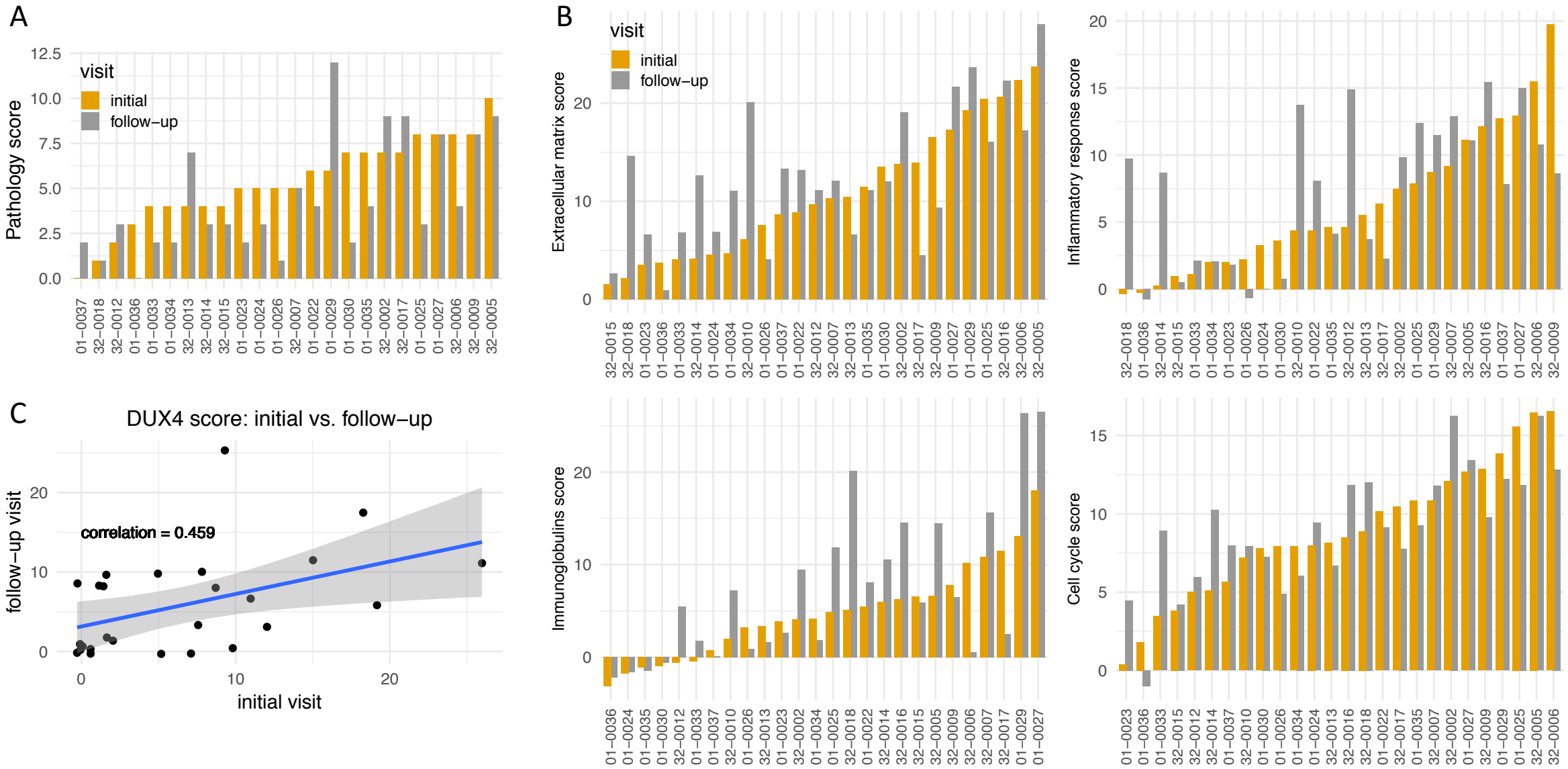
Suppl Figure S1. Principal component analysis plot showing 32-0008b is an outlier among all control, initial and follow-up visit FSHD biopsy RNA-seq samples.



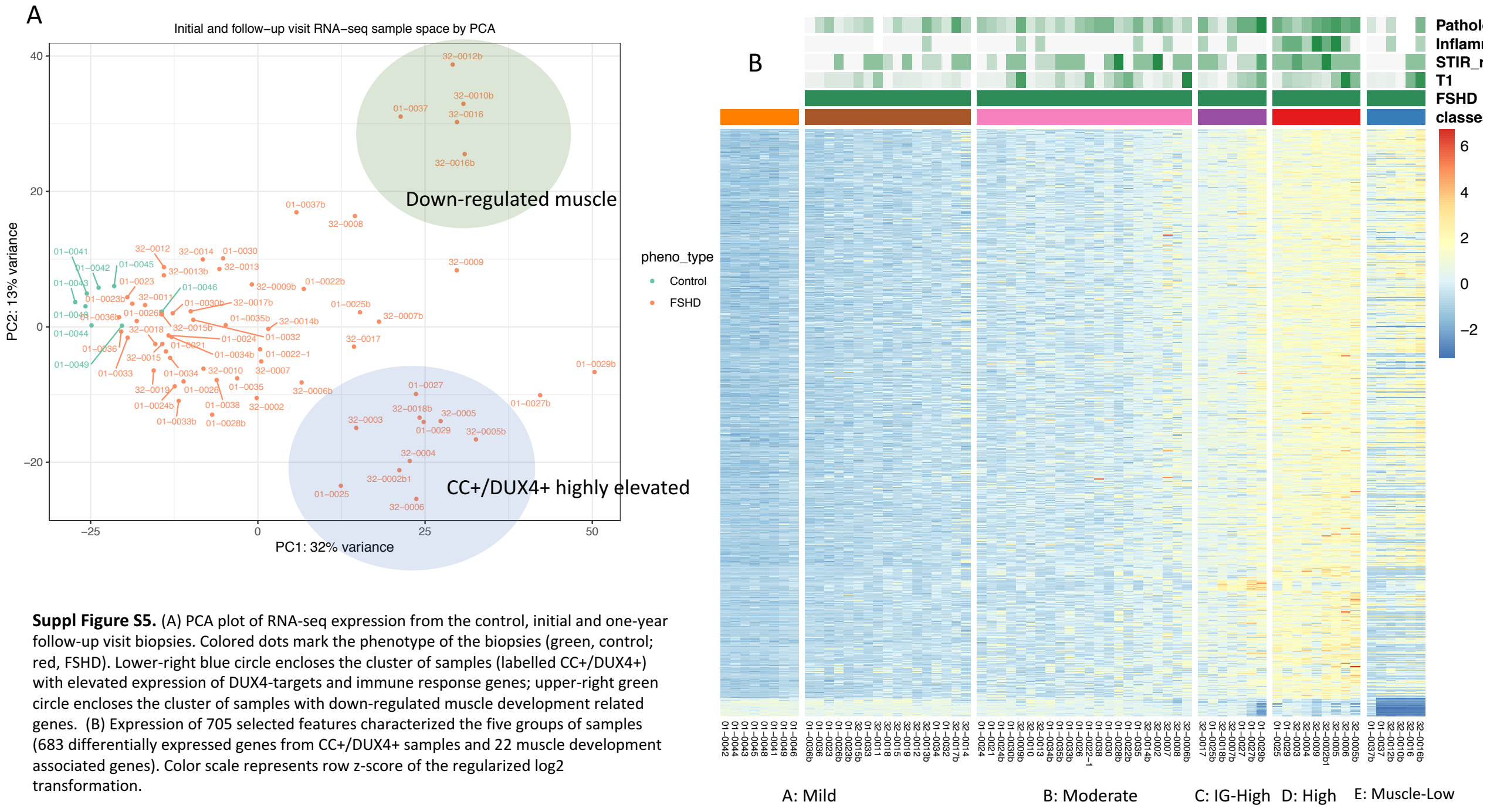
Suppl Figure S2. RNA-seq expression of 54 DUX4 robust biomarkers from the control and one-year follow-up FSHD biopsies. Color scale represents row z-score of the regularized log2 transformation. Vertical gaps divide DUX4 group 1, 2, 3, and 4 that categorized based on the expression of four DUX4 candidate biomarkers.



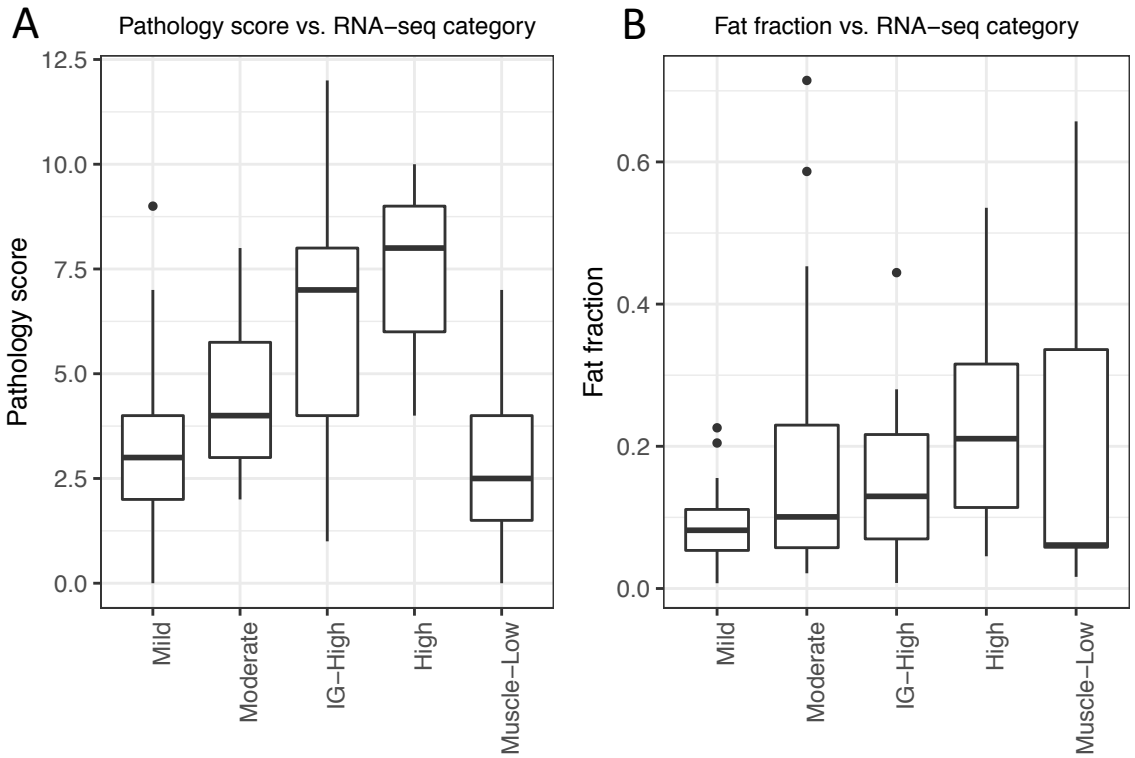
Suppl Figure S3. Top 60 loading variables of first and second principal components (PC1 and PC2) of the follow-up RNA-seq samples. Y-axis represents the PCA loading values of the variables indicating the how much the variable contributes to the sample variance. Colors indicate the functional category of the variables. This figure conveys that the sample variance captured by PC1 were mostly contributed by genes associated inflammatory/immune response, extracellular matrix and immunoglobulins. The top PC2 loading variables that accounted for 13% of the sample variance were associated with immunoglobulins, DUX4 and muscle development.



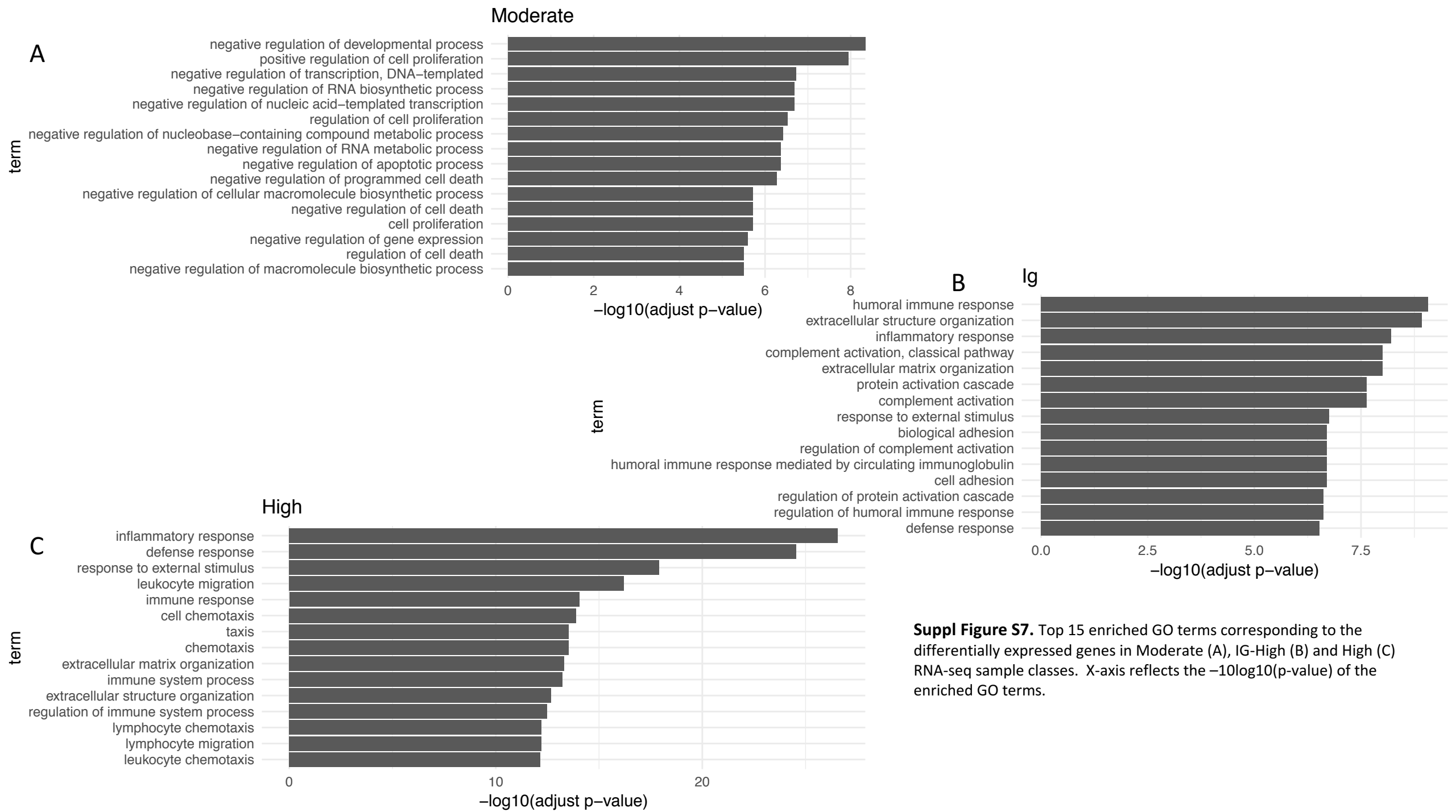
Suppl Figure S4. (A) Pathology score and (B) non-DUX4 marker scores over two time points for each individual. Yellow marks the initial visits and grey the one-year follow-up. The extracellular matrix score was calculated based on PLA2GA, COL19A1, COMP, and COL1A1; immune/inflammatory response score based on CCL18, CCL13, C6 and C7; immunoglobulins score based on IGHA1, IGHG4 and IGHGP; cell cycle score based on CCNA1, CDKN1A and CDKN2A. (C) Scatter plot of DUX4 score in the initial vs. follow-up visits. Blue line presents the linear regression between the two visits; the grey shadow is the 95% confidence interval of the linear regression.



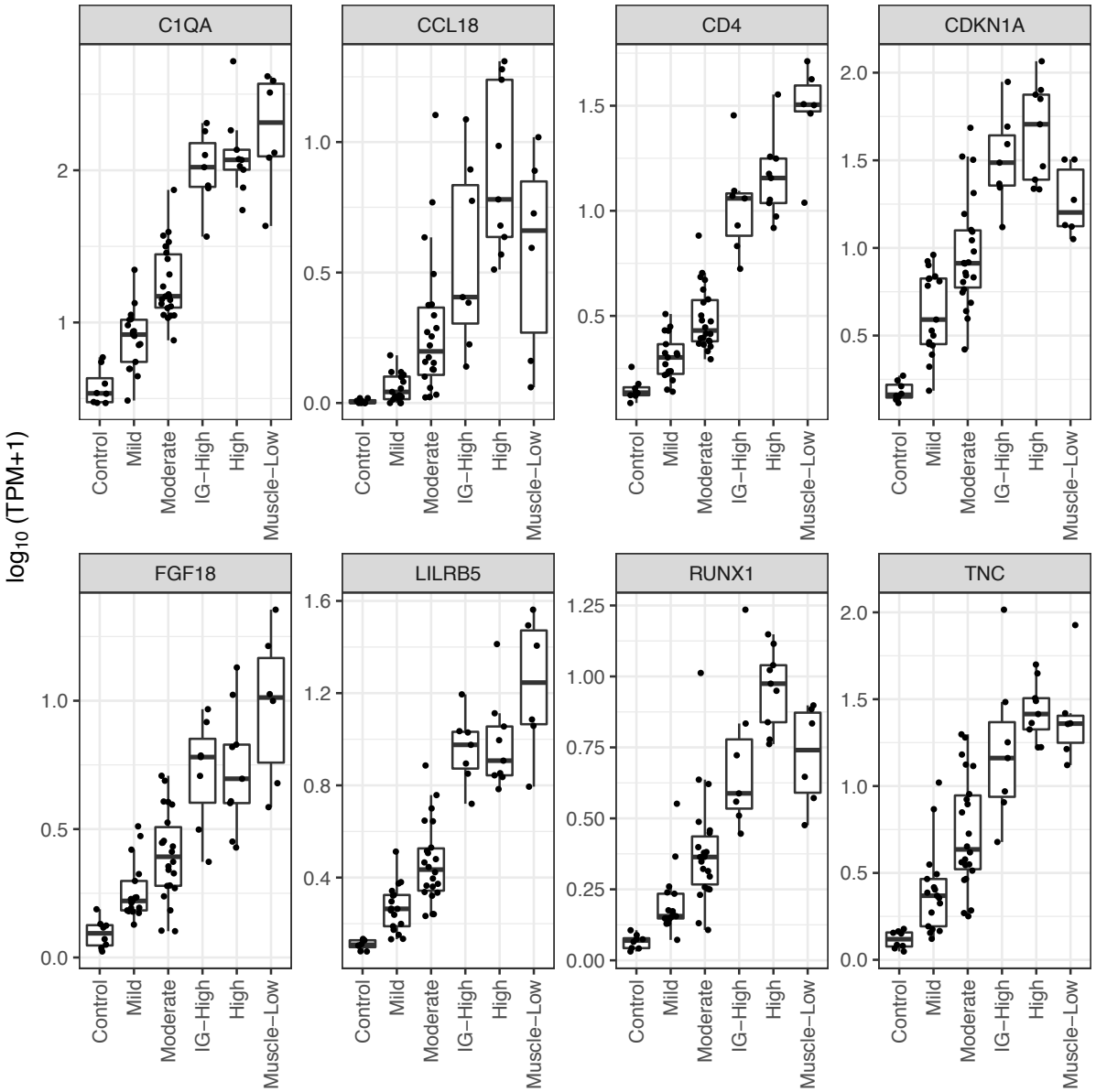
Suppl Figure S5. (A) PCA plot of RNA-seq expression from the control, initial and one-year follow-up visit biopsies. Colored dots mark the phenotype of the biopsies (green, control; red, FSHD). Lower-right blue circle encloses the cluster of samples (labelled CC+/DUX4+) with elevated expression of DUX4-targets and immune response genes; upper-right green circle encloses the cluster of samples with down-regulated muscle development related genes. (B) Expression of 705 selected features characterized the five groups of samples (683 differentially expressed genes from CC+/DUX4+ samples and 22 muscle development associated genes). Color scale represents row z-score of the regularized log2 transformation.



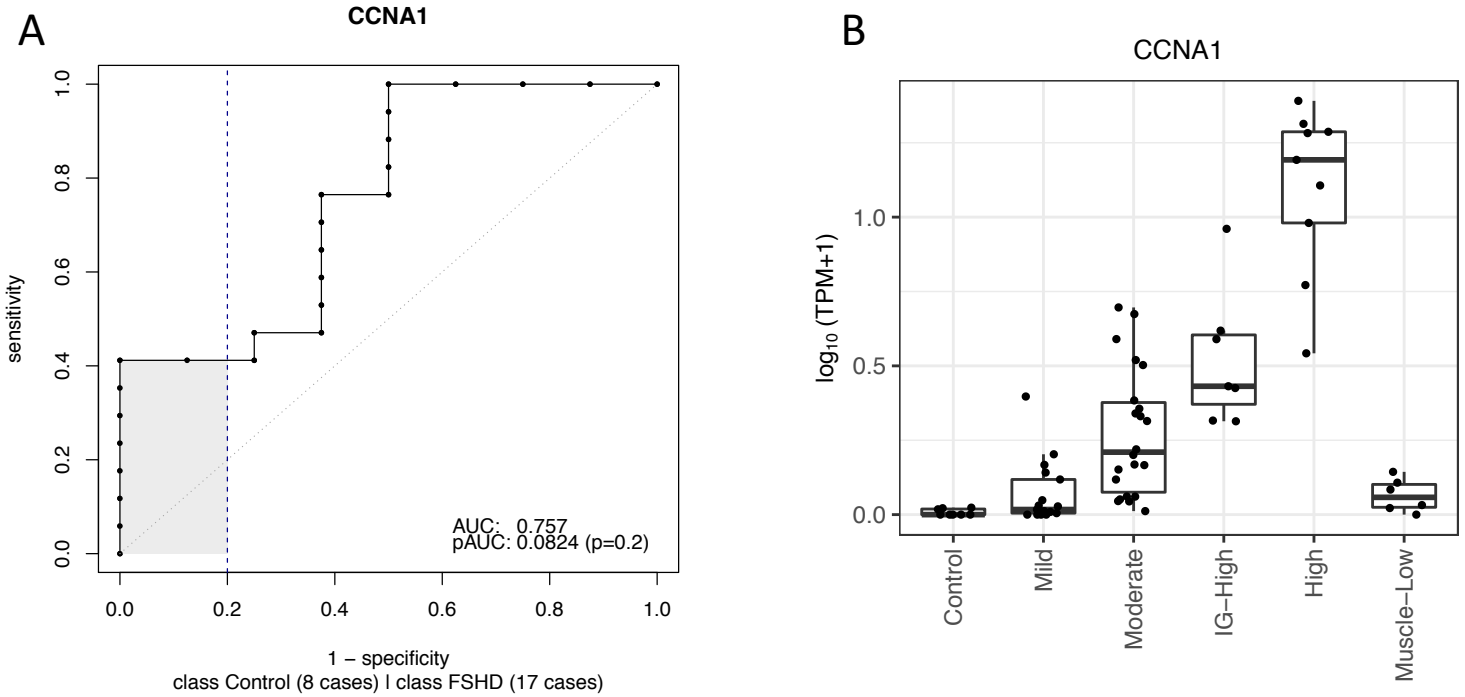
Suppl Figure S6. Boxplot of Pathology score (A) and T1 fat fraction (B) for each FSHD RNA-seq sample classes represented with minimum, first quartile, median, third quartiles, and maximum.



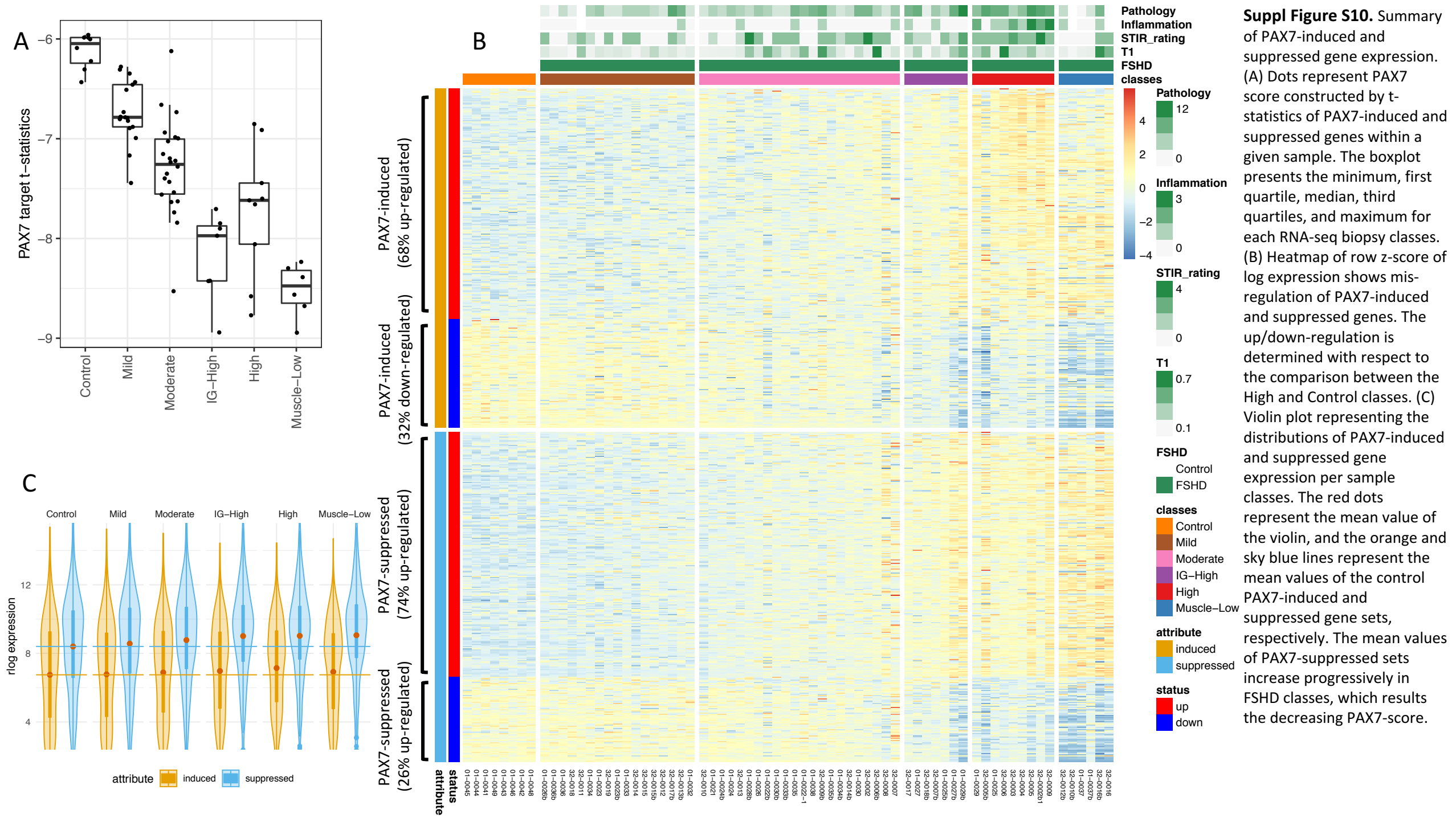
Suppl Figure S7. Top 15 enriched GO terms corresponding to the differentially expressed genes in Moderate (A), IG-High (B) and High (C) RNA-seq sample classes. X-axis reflects the –10log10(p-value) of the enriched GO terms.



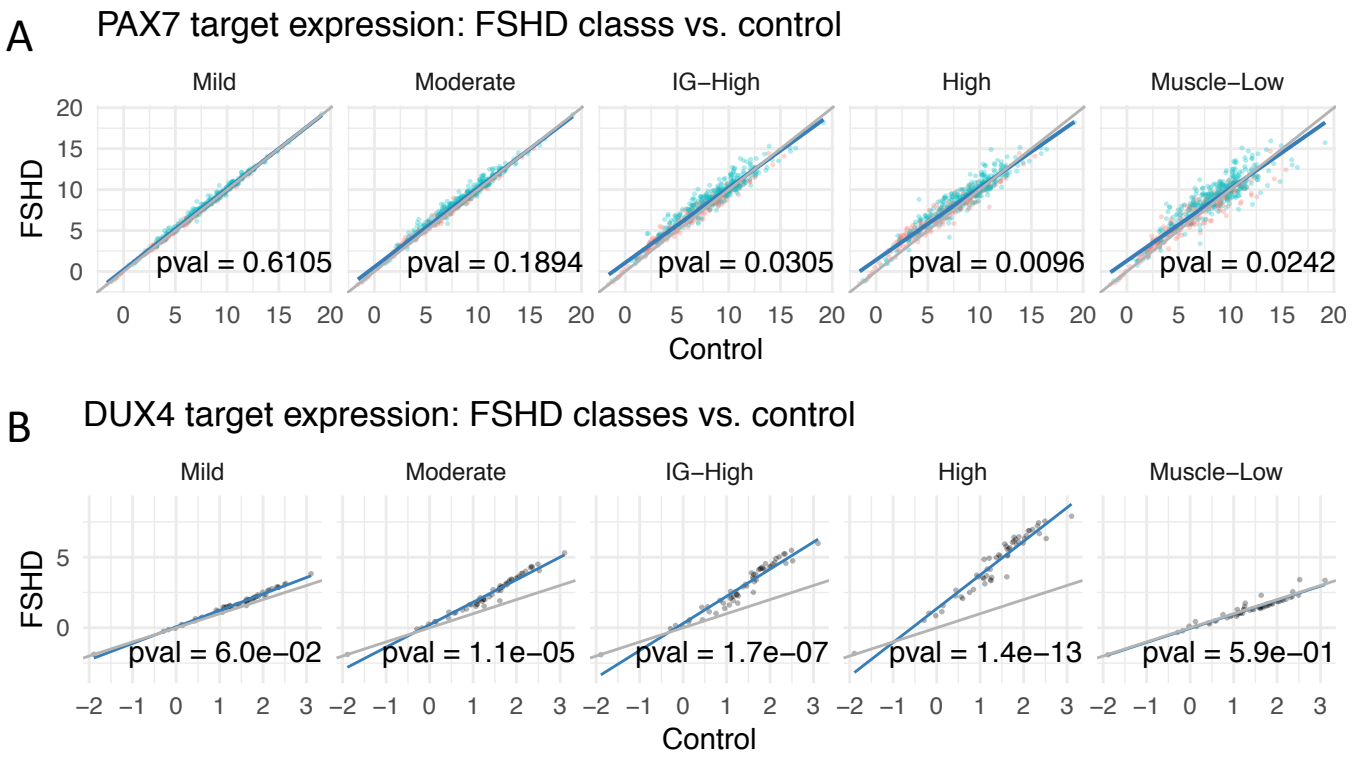
Suppl Figure S8. TPM expression of eight selected potential biomarkers per FSHD classes form the Mild, Moderate, IG-High to High classes and Muscle-Low. Each dot represents the expression from a biopsy sample and the boxplot displays the minimum, first quartile, median, third quartiles, and maximum. TPM, transcripts per million.



Suppl Figure S9. CCNA1's ROC curve and expression for each class of samples. (A) ROC curve of CCNA1 in discriminating 17 Mild class samples and eight controls with 0.757 AUC and 0.824 partial AUC (p=0.2, false positive rate). (B) TPM expression of CCNA1 for each class of samples presented by boxplots with minimum, first quartile, median, third quartiles, and maximum. TPM, transcripts per million.



Suppl Figure S10. Summary of PAX7-induced and suppressed gene expression. (A) Dots represent PAX7 score constructed by t-statistics of PAX7-induced and suppressed genes within a given sample. The boxplot presents the minimum, first quartile, median, third quartiles, and maximum for each RNA-seq biopsy classes. (B) Heatmap of row z-score of log expression shows mis-regulation of PAX7-induced and suppressed genes. The up/down-regulation is determined with respect to the comparison between the High and Control classes. (C) Violin plot representing the distributions of PAX7-induced and suppressed gene expression per sample classes. The red dots represent the mean value of the violin, and the orange and sky blue lines represent the mean values of the control PAX7-induced and suppressed gene sets, respectively. The mean values of PAX7-suppressed sets increase progressively in FSHD classes, which results the decreasing PAX7-score.



Suppl Figure S11. Scatter plot of PAX7-target (B) and DUX-target (C) average rlog gene expression in different FSHD classes vs. controls. The blue lines represent linear regression of FSHD classes and control samples; the grey lines represent as if gene expression in FSHD and control samples were in 1:1 relationship (x=y). The p-values were yielded by the two-sample Wilcoxon tests performed on different FSHD classes vs. control.