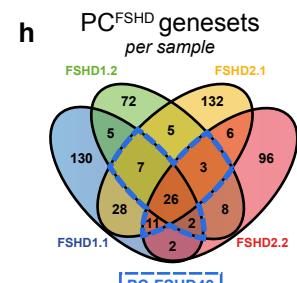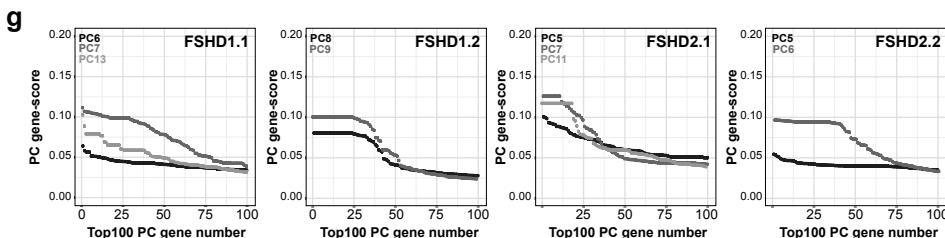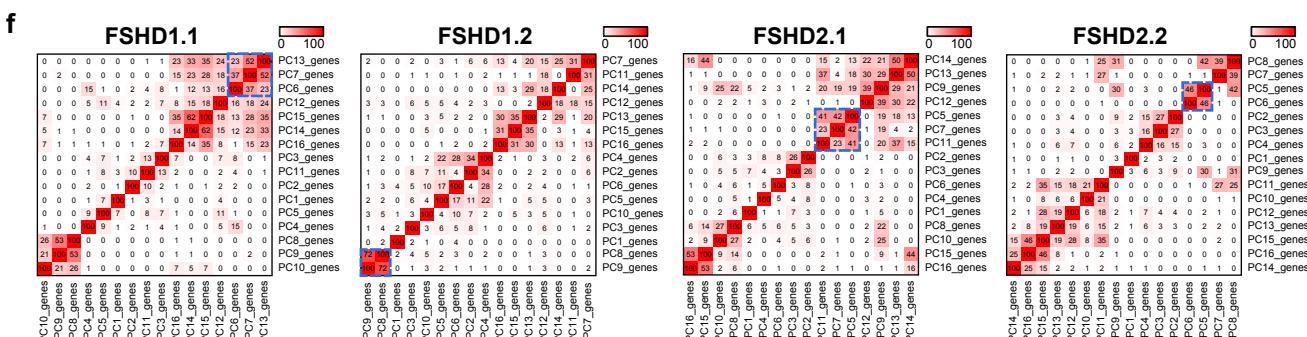**Figure S6. PCA-based identification of an FSHD-specific layer of complexity.**

a) Scree plots for the first 100 PCs per sample, with the selected first sixteen highlighted in red. **b)** An overlap matrix heatmap of the PC1 top-100 gene lists of each sample. Numbers indicate the number of genes overlapping between the two gene lists. **c)** Predicted protein association network of all genes in the PC1 gene lists of all six samples combined (STRING-db, version 10.5)(5). Thickness of the connections indicate association confidence score (only associations with confidence score >0.4 depicted). Non-connected proteins not depicted. Color highlights are a selection of the pathways identified in the STRING gene set enrichment tests (Fisher's exact test followed by multiple testing correction).(5) **d)** Gene set enrichment analysis (EnrichR) results for significantly affected 'GO Biological Processes' for all genes in the PC1 gene lists of all six samples combined. The graph only shows the significantly affected processes (adjusted pvalue <0.05). **e)** An overview of the number of DUX4 biomarkers being present in the top-100 gene lists of the first 16 PCs per sample. The selected PCs for generating the PCFSHD (PC-FSHD49 gene set) are highlighted with a red box. **f)** An overlap matrix heatmap of the first sixteen PC top-100 gene lists per sample. Numbers indicate the number of genes overlapping between the two PC gene lists. The heatmap is clustered based on overlap-scores. The selected DUX4 biomarker containing PCs are highlighted by the blue dash-lined boxes. **g)** Gene contribution plots for the selected DUX4 biomarker-containing PCs per sample. **h)** Venn diagram plotting the overlap between the combined DUX4 biomarker-containing PC gene lists per sample. Genes identified in at least three out of four FSHD samples were selected as the PC-FSHD49 gene set (marked by the blue dash-lined shape) (see also Table S3).