Figure 1. Size exclusion chromatography separates serum-derived EVs into fraction with distinct proteomes. (A) Particle concentration and average particle size of 12 size exclusion chromatography (SEC) column fractions (F) from 5 patients (P.CF1-4, H.CF1). (B) Transmission electron micrograph of serum-derived EVs from F6 to F12. Scale bar represents 200 nm. (C) Number of proteins identified in each sample from EVs in F6 to F12. (D) Number of proteins identified in multiple fractions. (E) Two-dimensional principal component analysis (PCA) of the EV proteomes of each sample coloured by fraction. (F) The log2 label-free quantification (LFQ) intensity of plasma membrane-associated (left panel) and cytosolic EV proteins (middle panel), and non-EV co-isolated proteins (right panel). Coloured dots indicate replicates (n=7). Box plots show mean log2 intensity and error bars show standard error. (G) Venn diagram of all identified proteins that were present in over 50% of replicates within combined F6 to F9 and F10 to F12. (H) Two-dimensional PCA of the proteomes of the F6 to F9 and F10 to F12 fractions. Colour indicates fraction combination and dots indicate replicates (n=6). (I) Functional enrichment analysis of biological process and cellular component gene ontologies proteins only identified in F6 to F9. Colour indicates -log10 adjusted p-value. n = 6. (J) Hierarchical clustering heatmap of all small RNAs identified within EVs including miRNAs. Rows indicate small RNAs and columns indicate samples. Color indicates the row Z-score (normalised expression relative to mean of the row) with yellow indicating higher expression and purple indicating lower expression. Distance between samples was the Euclidean distance and hierarchical clustering was calculated with Ward’s method. (K) Venn diagram of the number of sEV miRNAs identified in our study and in Zhao et al. (2020). n=12.​

Figure 2. Differential Expression Analysis of CFRD, IGT, NGT pre-modulator and post-modulator samples. (A) Volcano plots of differentially abundant proteins (top row) and differentially abundant transcripts (bottom row) between premodulator samples of CFRD, IGT and NGT ​  
(B) Heatmaps showing the z-scores of most significantly affected IPA pathways associated with the differentially abundant proteins between post-modulator and pre-modulator samples. The corresponding volcano plots are shown in Supplementary Figure 2B1. Positive z-score indicates pathway activation and negative z-score indicates pathway inhibition. The rows indicate the IPA pathways and columns are the 4 comparisons : (i) all post-modulator samples Vs all pre-modulator sample (ii) post-modulator Vs pre-modulator CFRD samples (iii) post-modulator Vs pre-modulator IGT samples (iv) post-modulator Vs pre-modulator NGT samples​  
Heatmap on the left shows pathways commonly affected in the 4 comparisons and the heatmap on the right shows other top pathways from each of the comparison. The top pathways are defined in terms of highest absolute value for z-score ​

(C) Shift from CFRD to NGT –  The boxplot of the top 10 upregulated proteins and UMAP dimensionality reduction plot of transcriptomics show that post-modulator CFRD samples are closer to pre-modulator NGT samples in comparison to pre-modulator CFRD samples. ​  
BH-adjusted p-values identified using Wilcoxon test for the significance of the expression difference between the pre-modulator CFRD, pre-modulator NGT and between post-modulator CFRD, pre-modulator NGT are shown in the boxplot. For 8 out of the top 10 upregulated proteins there is a significant difference between pre-modulator CFRD and pre-modulator NGT, but no significant difference between post-modulator CFRD and pre-modulator NGT.​  
In the UMAP plot, the pre-modulator CFRD samples are scattered along the bottom right, whereas the pre-modulator NGT and post-modulator CFRD samples are scattered primarily across the top left ​

Figure 3A : Best biomarker mean AUC heatmap : The cells show the mean AUC values for the identified best biomarkers across transcriptomics and proteomics. Each cell also shows the classification model that resulted in the best performance. The mean AUC values from all classification models used and other considered biomarker sets are provided in Supplementary Figure 3A