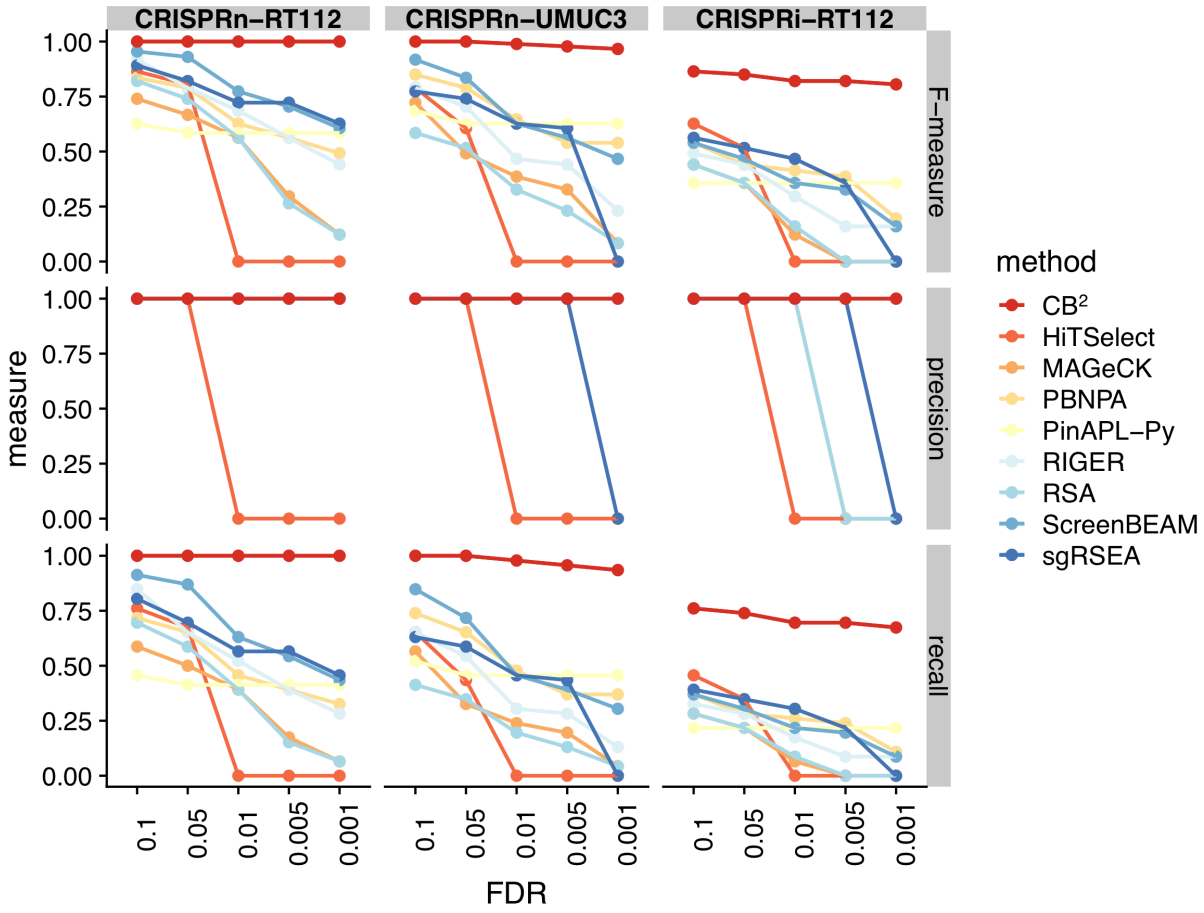

Supplemental Figure 7. CB^2 outperformed all other methods in F-1, precision and recall measures on Evers et al. (2016)'s screen data. F1-score (top), precision (middle) and recall (bottom) for each method on three benchmark datasets are presented as a function of FDR cut-off values.