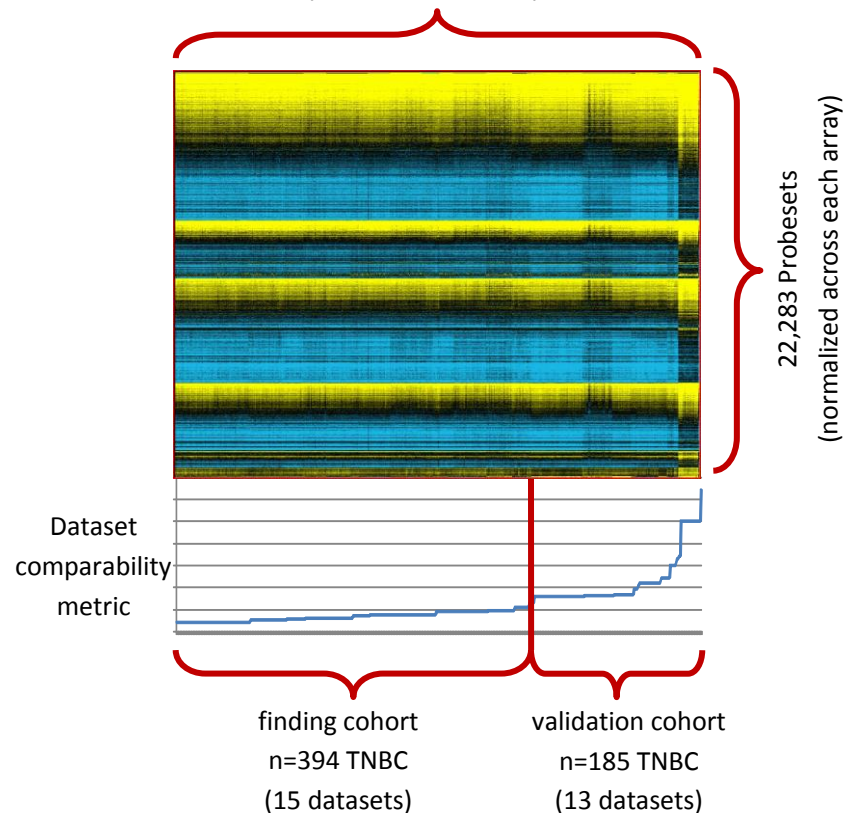


28 Affymetrix datasets encompassing n=579 Triple Negative Breast Cancers  
(Affymetrix U133A array data)



**Supplementary Figure S1: Selection of the TNBC finding cohort from multiple datasets based on dataset comparability**

Triple negative breast cancers (TNBC, n=579) from 28 datasets were sorted by dataset according to a dataset comparability metric (horizontally). Shown are the full array data of normalized Affymetrix U133A microarrays. The 15 most comparable datasets encompassing n=394 TNBC samples were subsequently used as a finding cohort and the remaining 13 datasets (n=185 TNBC samples) withheld as validation cohort.