Supplemental File 1, Table 1s:

Summary of feature selection and sample classification methods or supervised machine learning methods used for CR prediction analysis.

Supplemental File 2, Figure 2s:

Summary of CR Prediction Performance – Table shows maximum performance of various class prediction methods in predicting CR in the validation dataset using IFM II dataset (a) and Mulligan et al. dataset (b).

Supplemental File 3, Table 2s:

To look for CR prediction in those with early onset, we have now performed CR modeling based on time-points when CR was achieved, using data from HOVON trial. Specifically, HOVON study determined CR status at the end of 3rd and 8th cycle of induction therapy (CR3 and CR8 respectively) and at the end of protocol (CR20). Based on available clinical response data, patients with CR and no CR were regrouped in *early-onset* (CR3 and CR8) and late *CR* (CR20) groups. Also, patients remaining in CR were classified under *sustained CR* group. CR modeling was performed using seven different methods with 10-fold cross-validation and permutation analysis as described in the manuscript. Using this analysis we did not observe significant improvement in CR prediction in these newly regrouped subsets. Supplementary table 2 shows summary of CR prediction analysis. Near zero sensitivity with 100% specificity in CR3 modeling is attributed to skewed distribution of CR versus No CR samples. Group with disproportionately few cases (CR) fail to contribute in prediction modeling.

Supplemental File 4, Appendix:

We assessed performance of CR prediction separately in high and low GEP risk groups as defined by proliferation index (PI) and cytogenetic abnormalities.[29-30] HOVON study (Broyl A et al., 2010) used proliferation index to define low and high risk GEP groups. Cross-validated CR modeling analysis showed limited accuracy of GEP in CR prediction in both, high (61-68%) and low (57-67%) risk GEP groups. As anticipated, we observed further drop in performance when these cross-validated signatures were used in validation datasets – either a training-test split from the same HOVON trial or independent IFM dataset.

#end