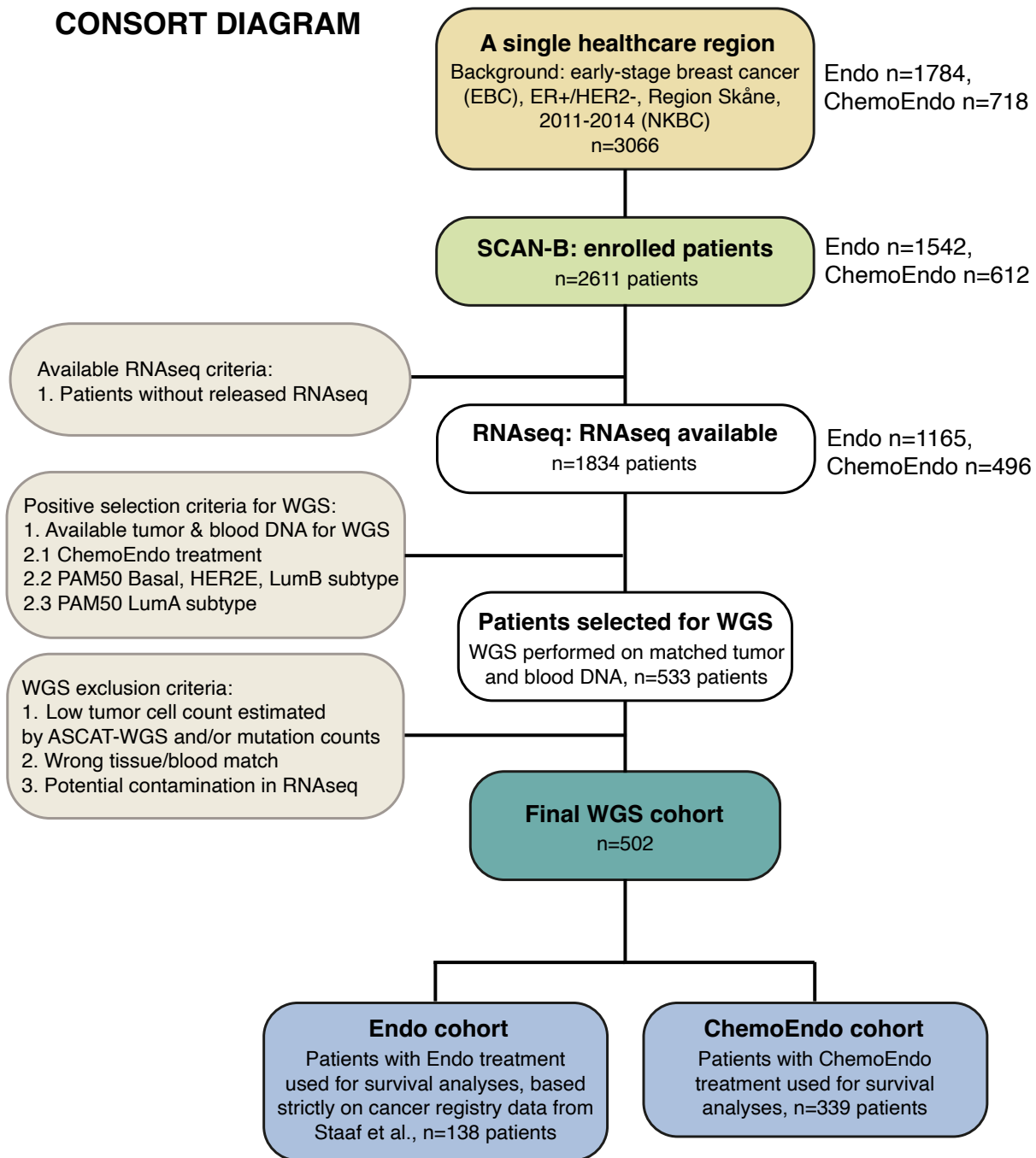


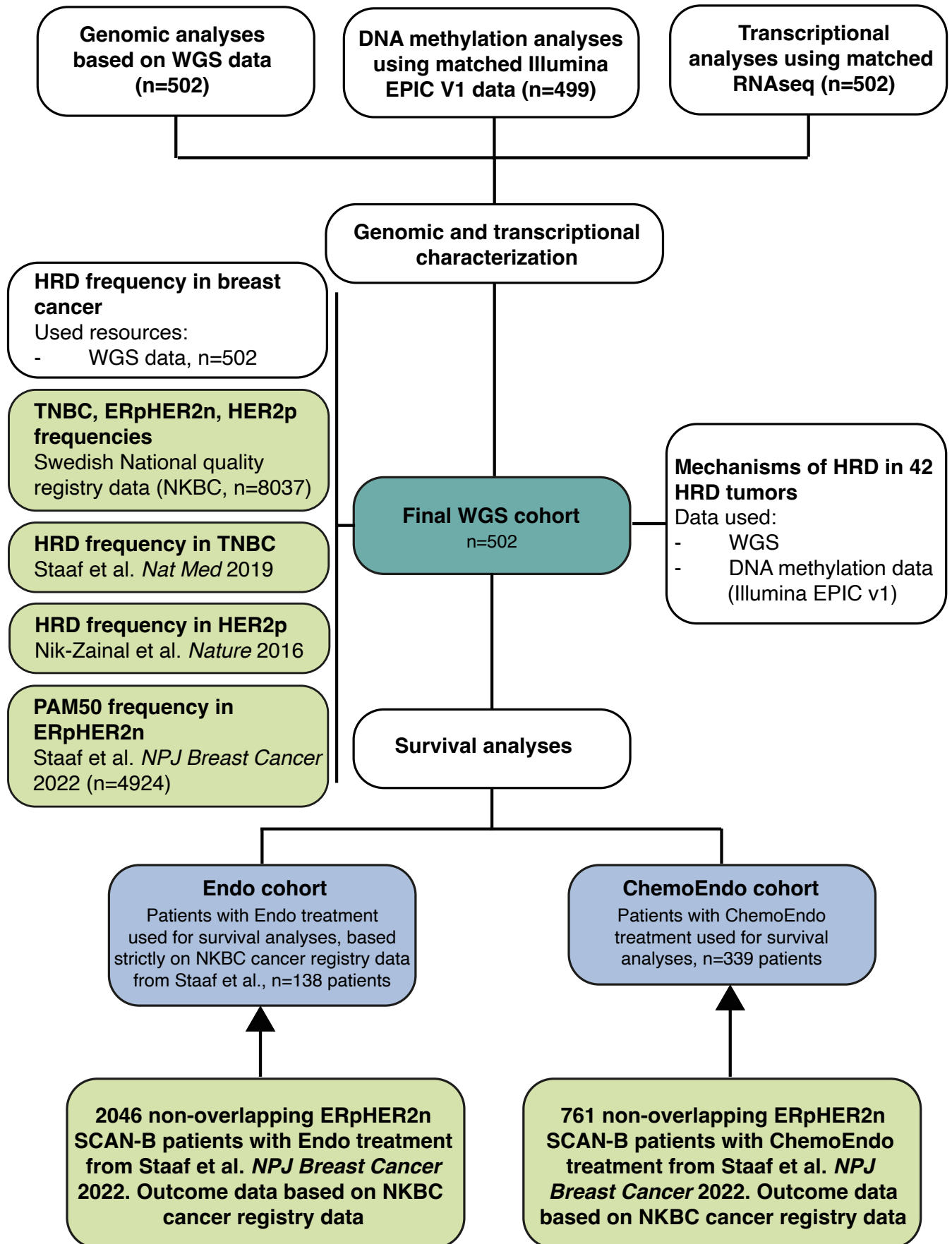
A)

CONSORT DIAGRAM



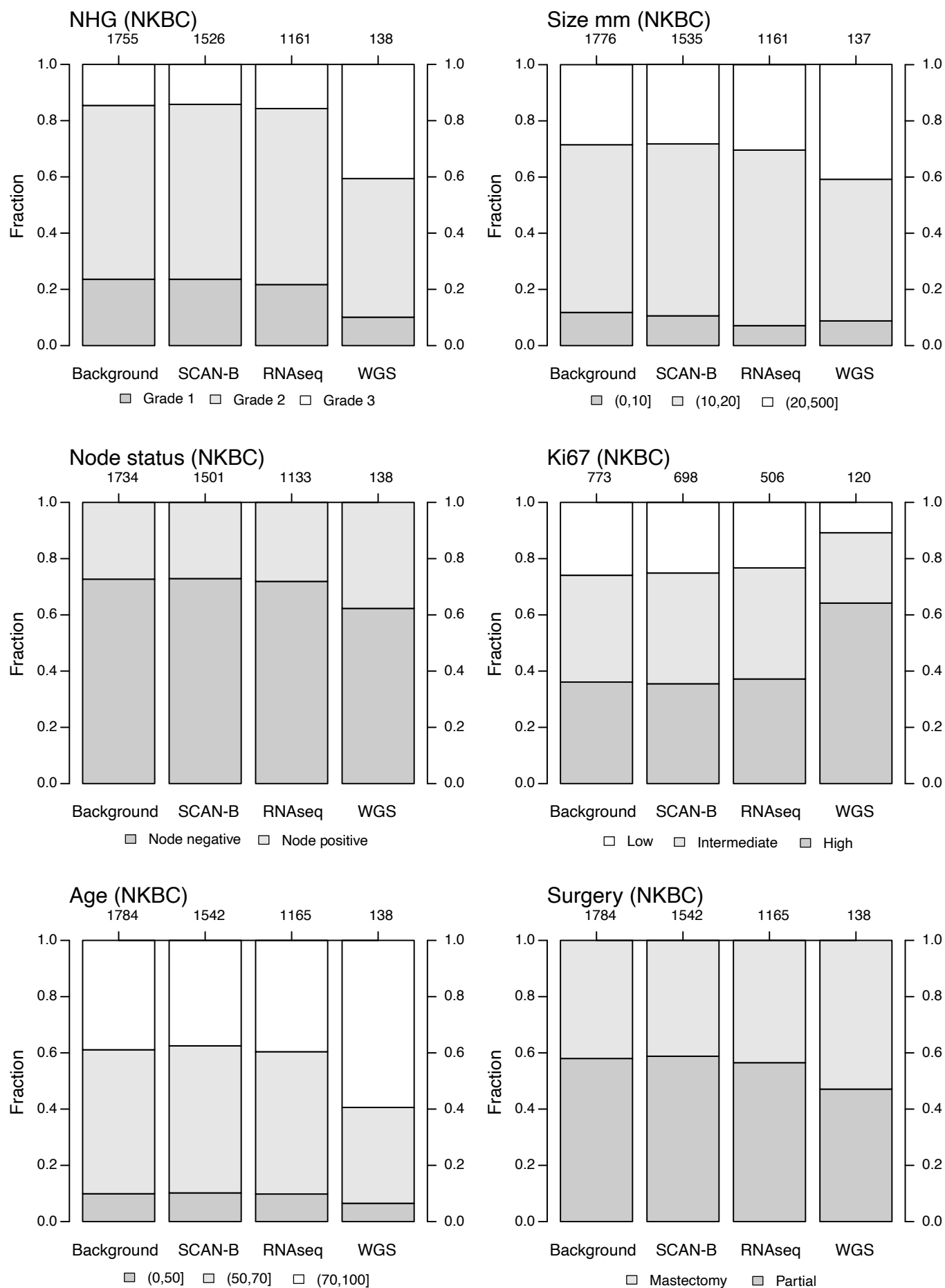
Supplementary Figure S1. CONSORT and treatment group representativity. (A) CONSORT diagram. NKBC: Swedish National Quality Registry for Breast cancer. **(B)** Study overview, with key analyses and external cohorts used. **(C)** Representativity of the WGS profiled Endo treatment group compared to background patient populations diagnosed during 2010-2014 for different clinical variables obtained from the NKBC registry in the catchment region (RS: Skane healthcare region). EBC: Early breast cancer (surgically treated). Top axis in barplots corresponds to sample numbers. **(D)** Representativity of the WGS profiled ChemoEndo treatment group compared to background patient populations for different clinical variables as in panel C. Top axis in barplots corresponds to sample numbers.

B) Study layout



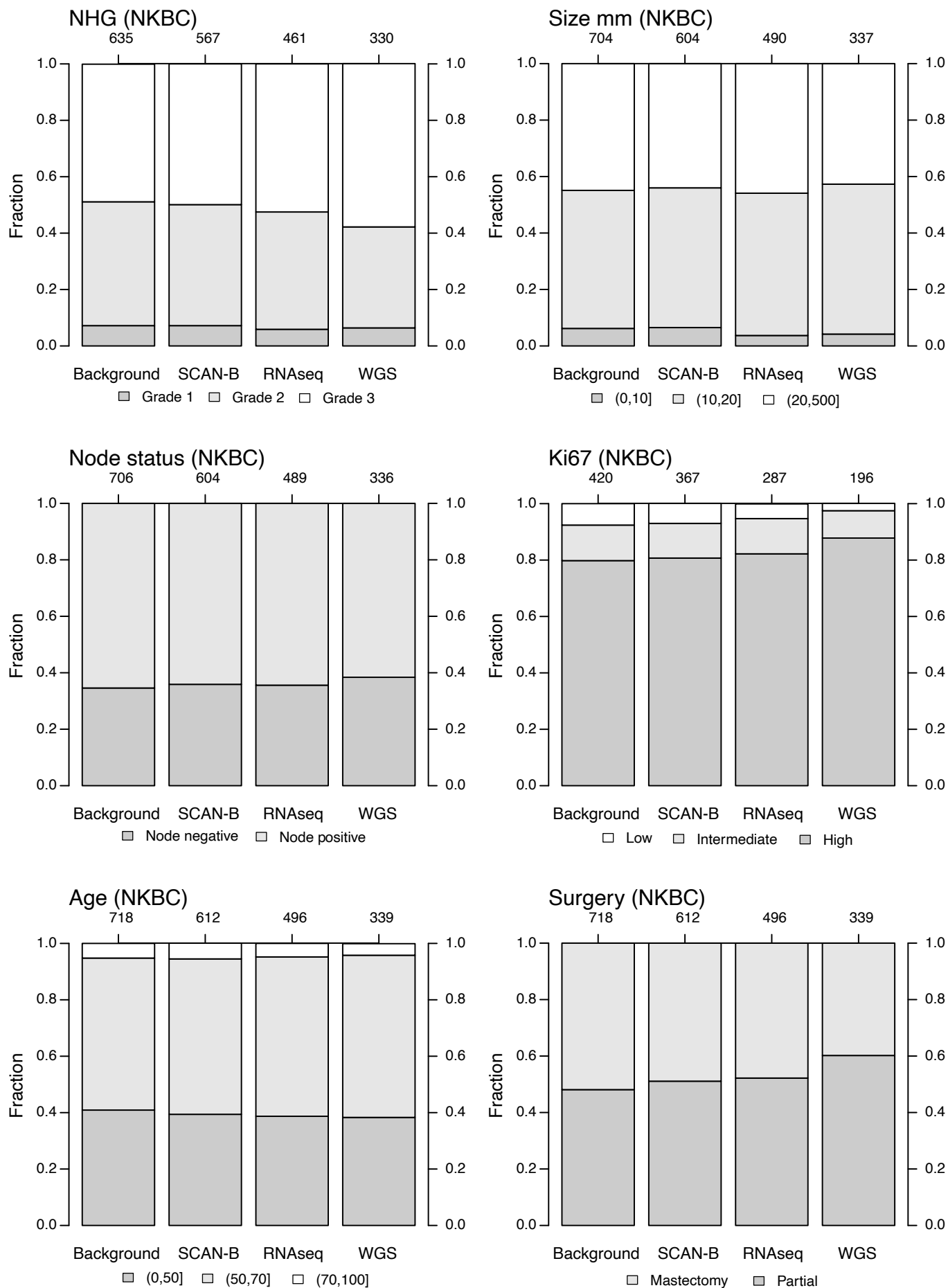
c)

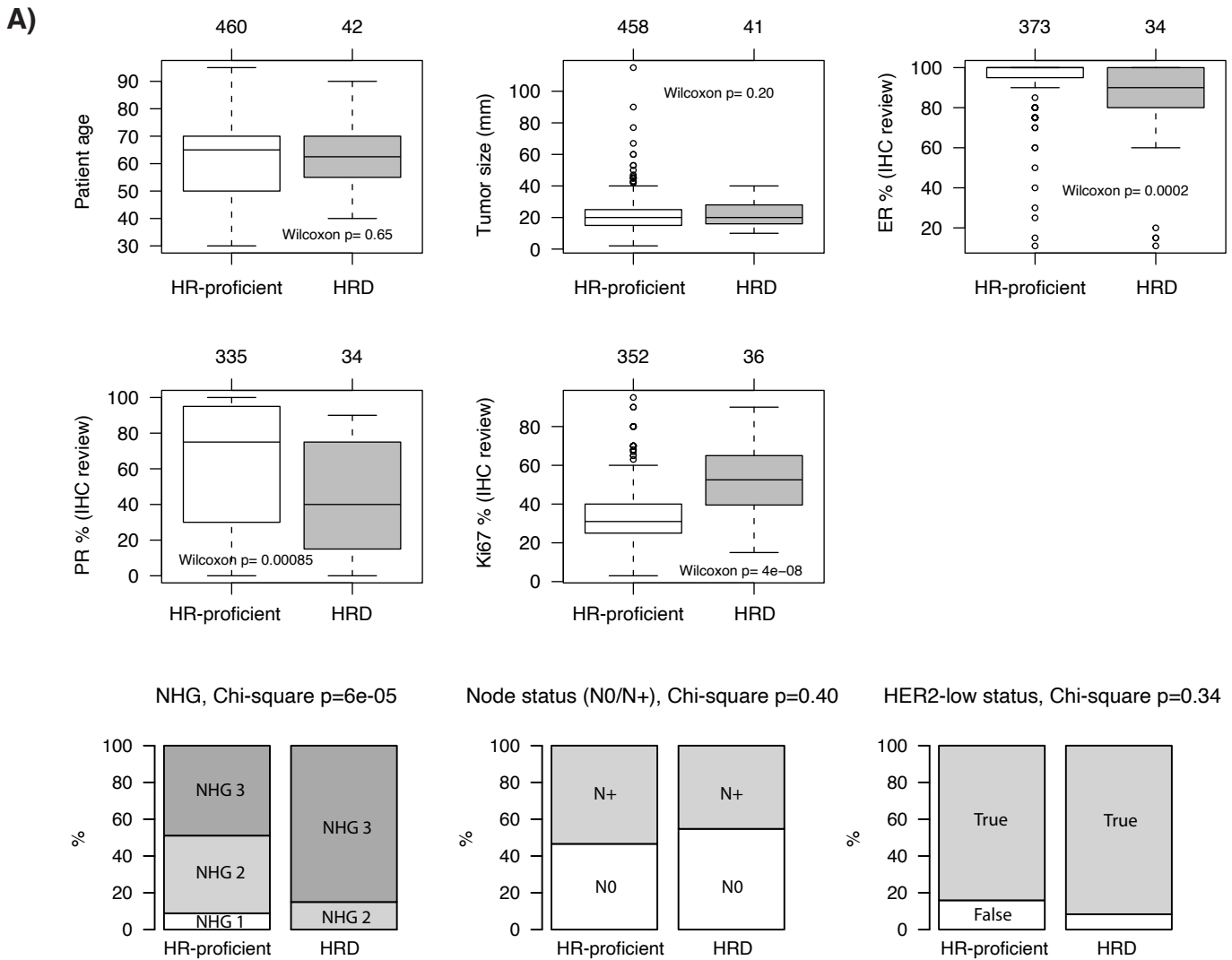
NKBC EBC, RS, ERpHER2n, 2010–2014, Endo



D)

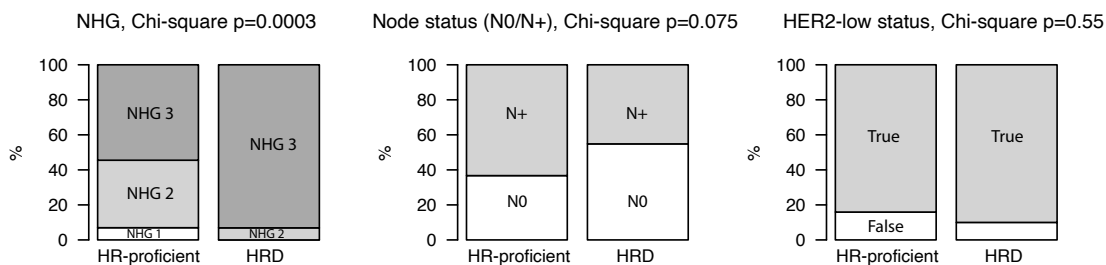
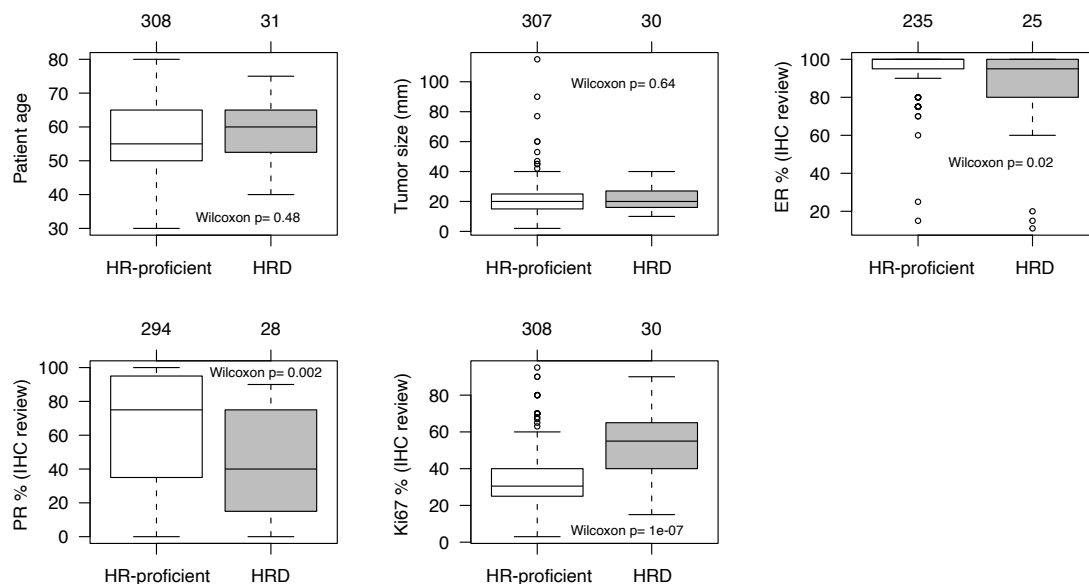
NKBC EBC, RS, ERpHER2n, 2010–2014, ChemoEndo



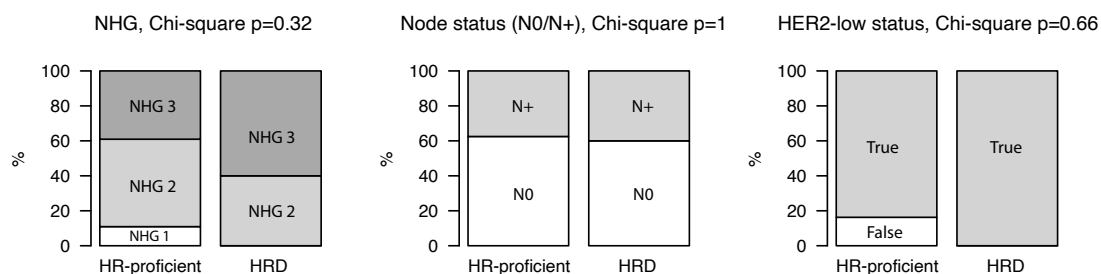
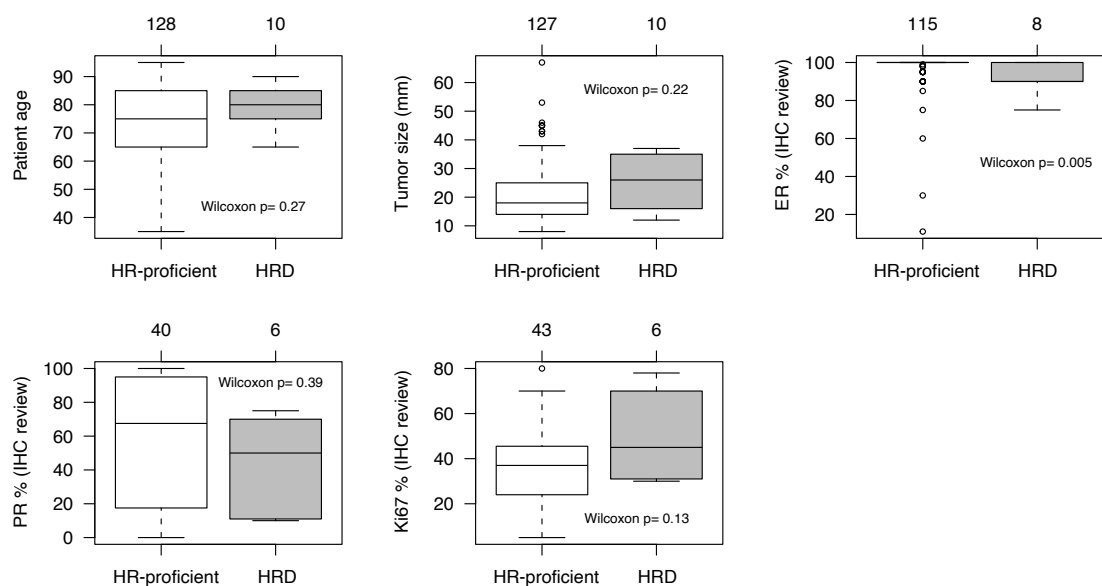


Supplementary Figure S2. Clinicopathological characteristics of HRD and HR-proficient ERpHER2n tumors in 502 WGS analyzed patients divided by treatment status. (A) Characteristics for all 502 patients. Not all patients have review data available. **(B)** Characteristics for 339 ChemoEndo treated patients. **(C)** Characteristics for 138 Endo treated patients. Not all patients have review data available. NHG: Nottingham grade index. N0: lymph node negative. N+: lymph node positive. Boxplot elements correspond to: i) center line = median, ii) box limits = upper and lower quartiles, iii) whiskers = 1.5x interquartile range. In boxplots, top axes indicate group sizes.

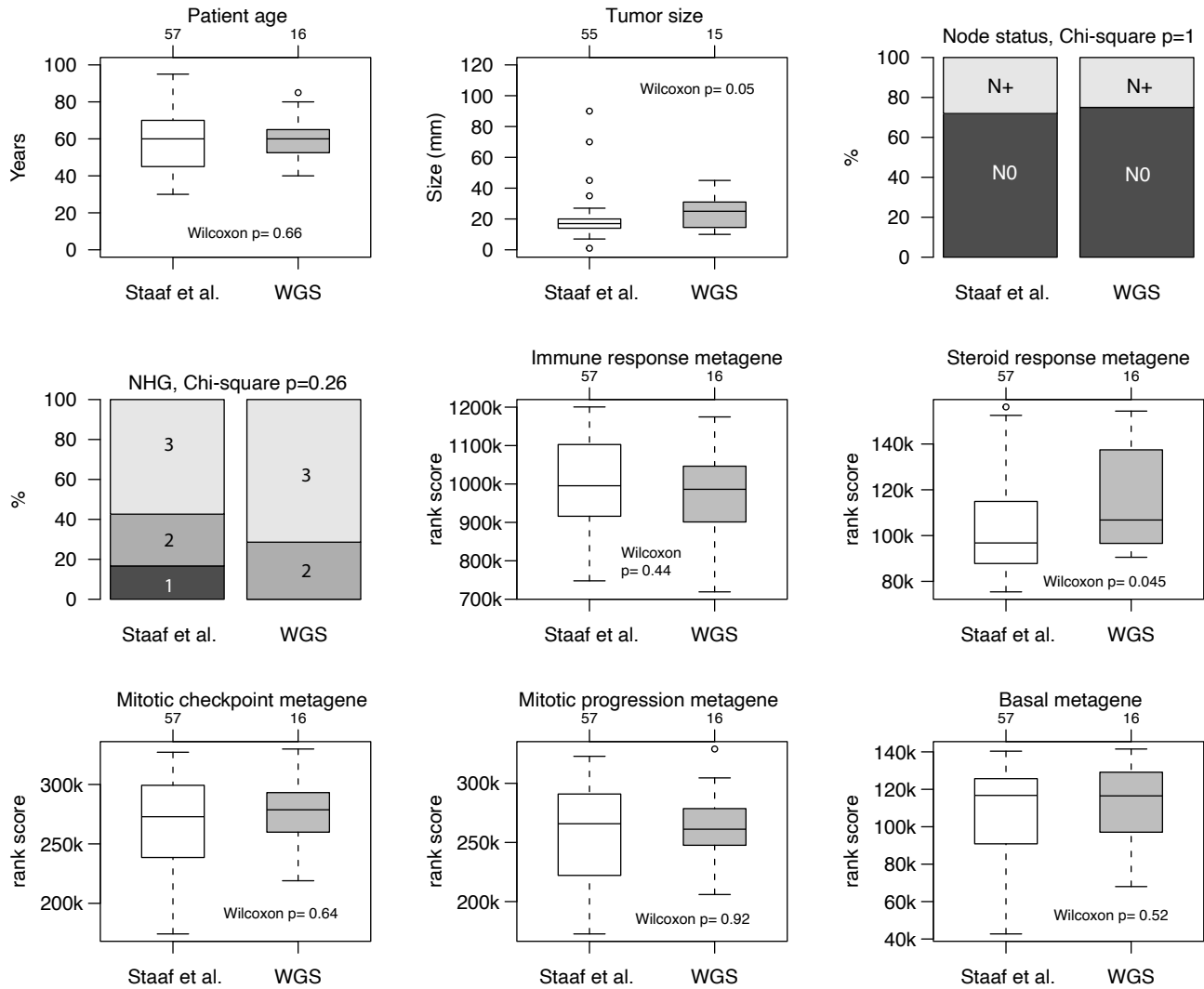
B)



C)

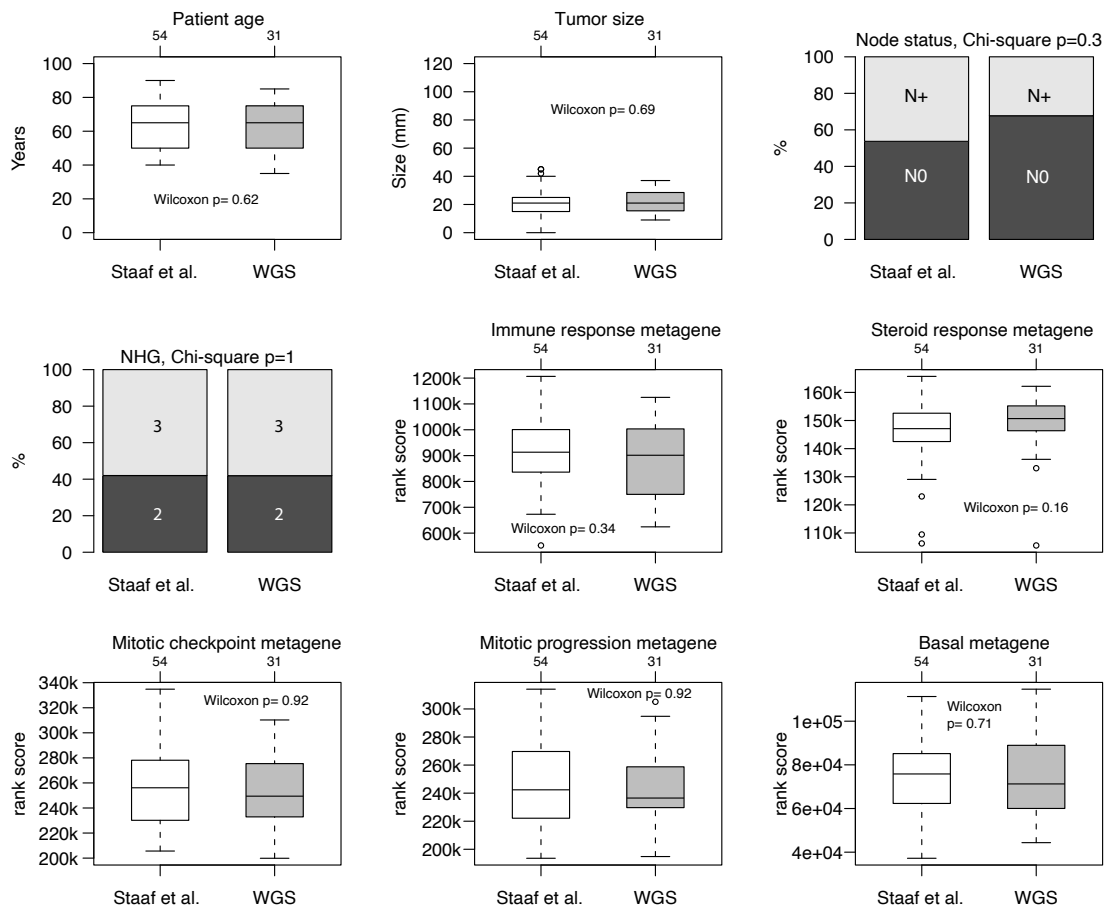


A) PAM50 Basal

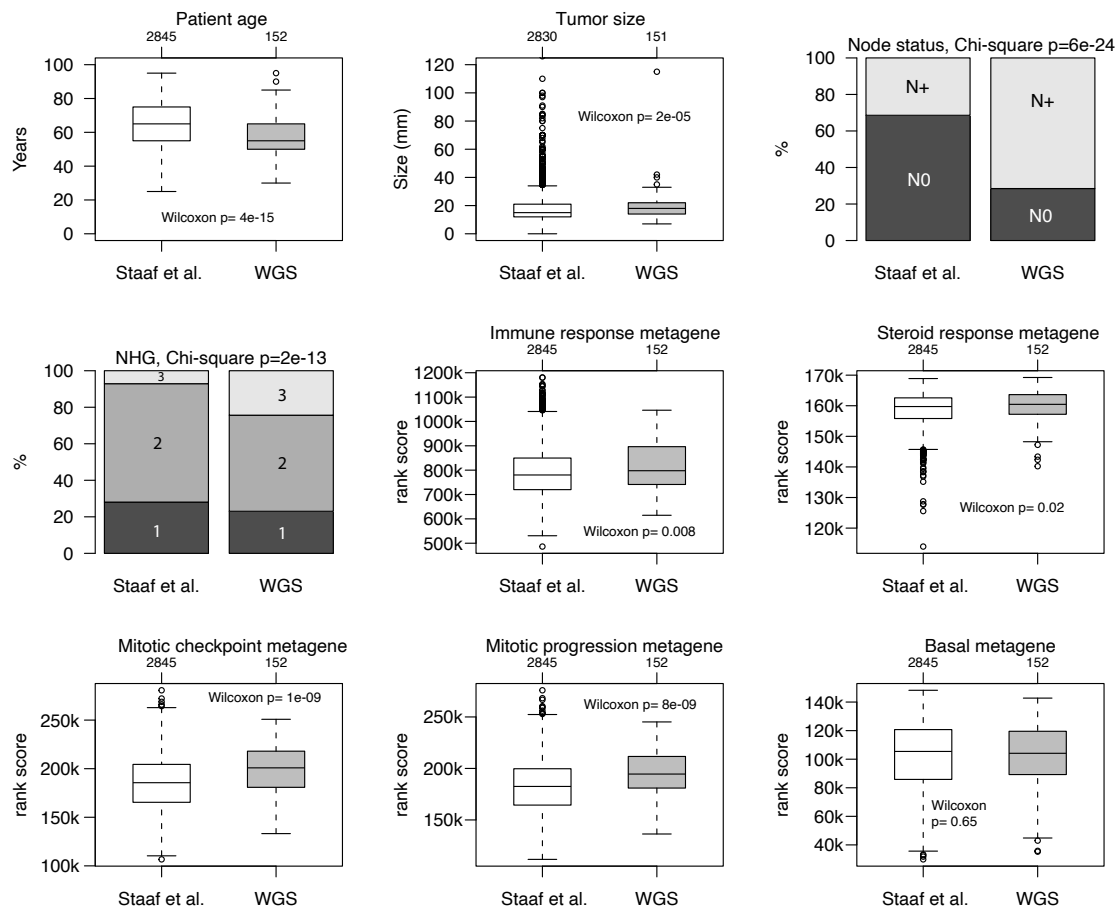


Supplementary Figure S3. Representativity of PAM50 subtypes in the ERpHER2n WGS cohort versus PAM50 subtypes in 4427 unrelated ERpHER2n tumors from the study by Staaf et al. (NPJ Breast Cancer 2022) regarding clinicopathological characteristics and expression of biological mRNA metagenes from Fredlund et al. (BCR 2012). The Chi-square test was used to test statistical significance for categorical data, whereas Wilcoxon's test was used for continuous data and rank-scores. **(A)** PAM50 Basal. **(B)** PAM50 HER2E. **(C)** PAM50 LumA. **(D)** PAM50 LumB. **(E)** PAM50 Normal. While certain characteristics in panel D (LumB) are significant it should be noted that the larger group sizes here can cause smaller differences to become statistically significant (e.g. age and nodal status). NHG: Nottingham grade index (grade 1, 2, 3). N0: lymph node negative. N+: lymph node positive.

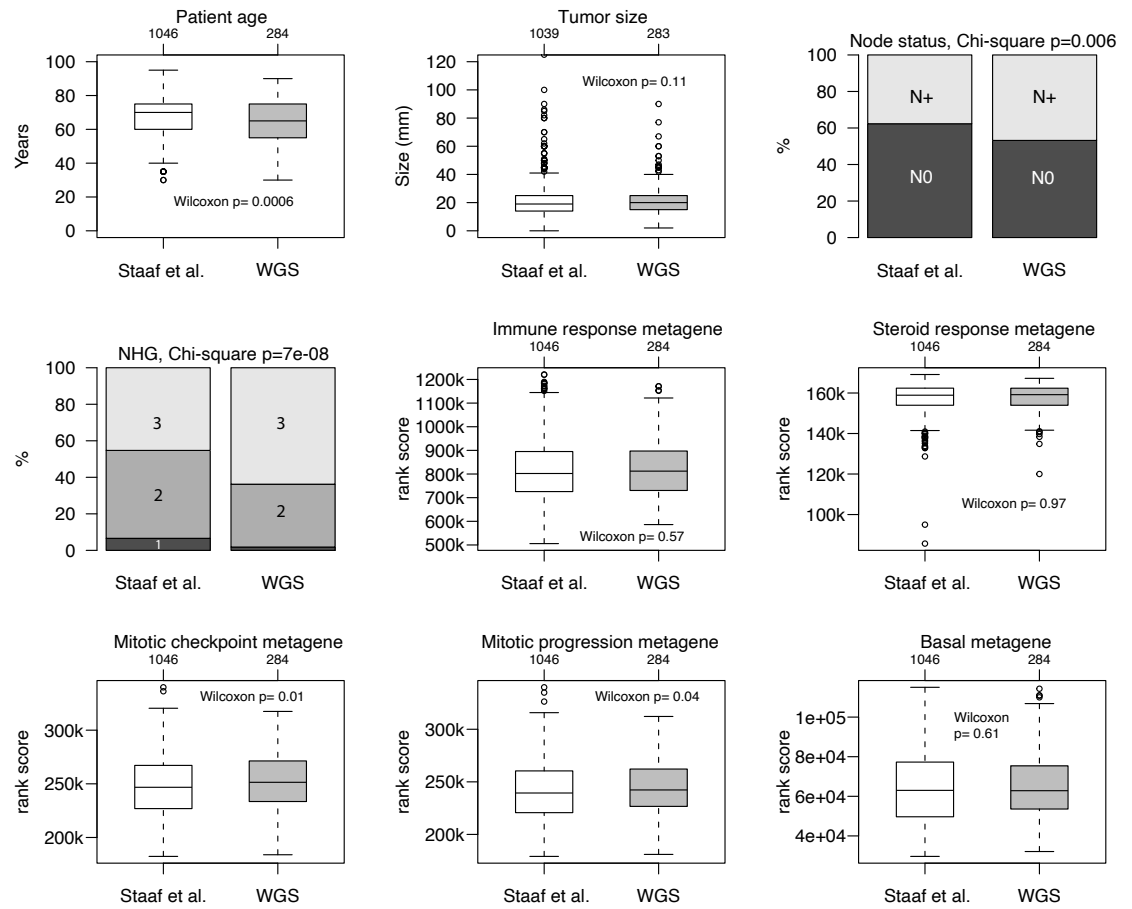
B) PAM50 HER2E



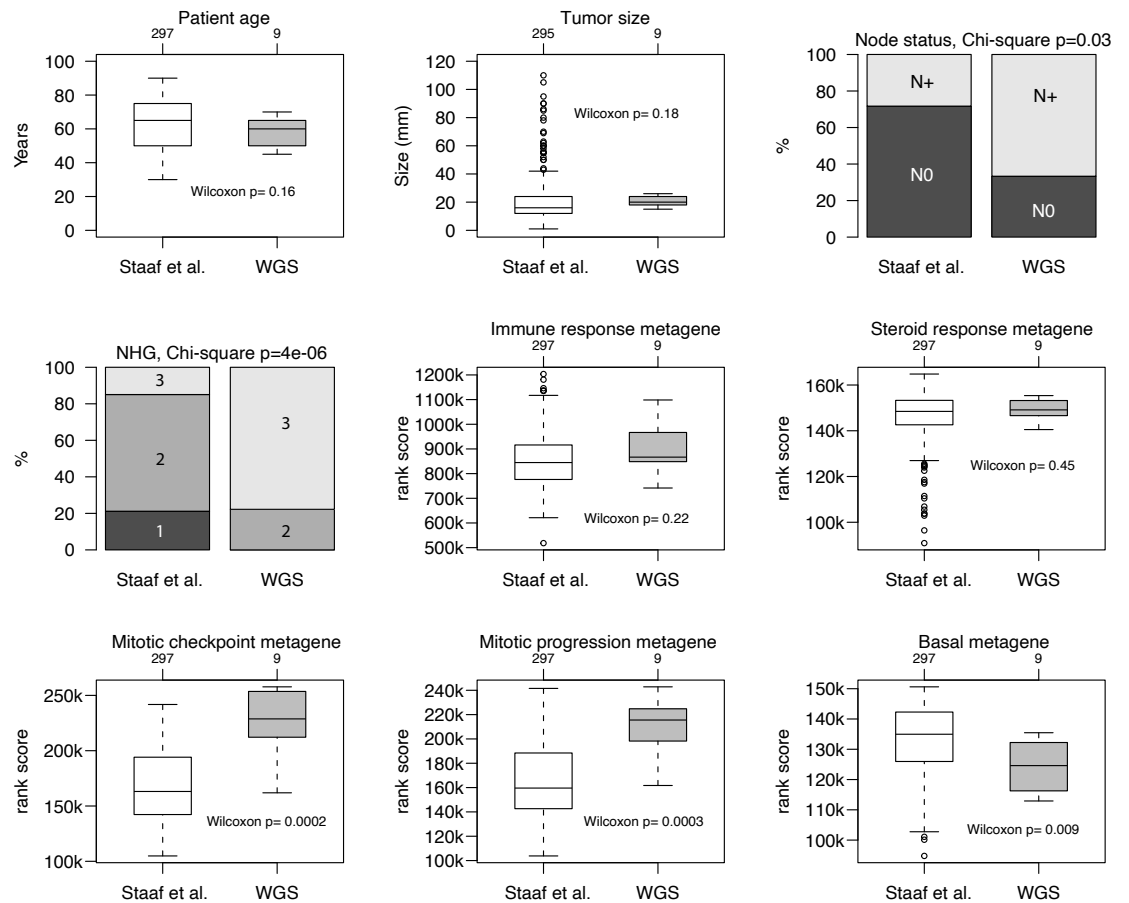
C) PAM50 LumA



D) PAM50 LumB



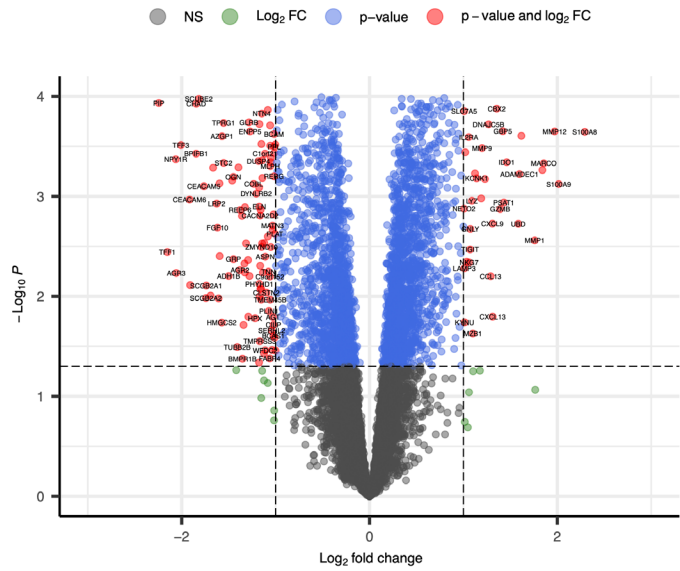
E) PAM50 Normal



A)

Ttest FDR DEGs HRD: All

DEG, nbr genes plotted= 11758 N significant genes= 3400

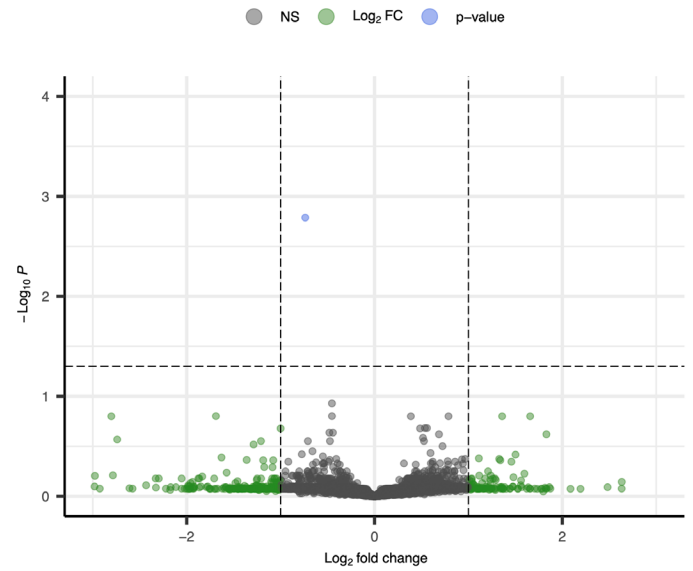


FC cutoff, 1; p-value cutoff, 0.05

B)

Ttest FDR DEGs HRD: Her2

DEG, nbr genes plotted= 12101 N significant genes= 1

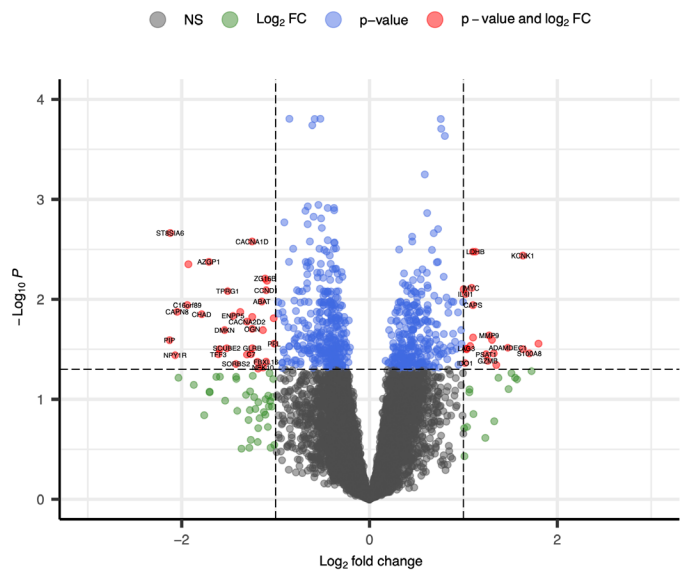


FC cutoff, 1; p-value cutoff, 0.05

C)

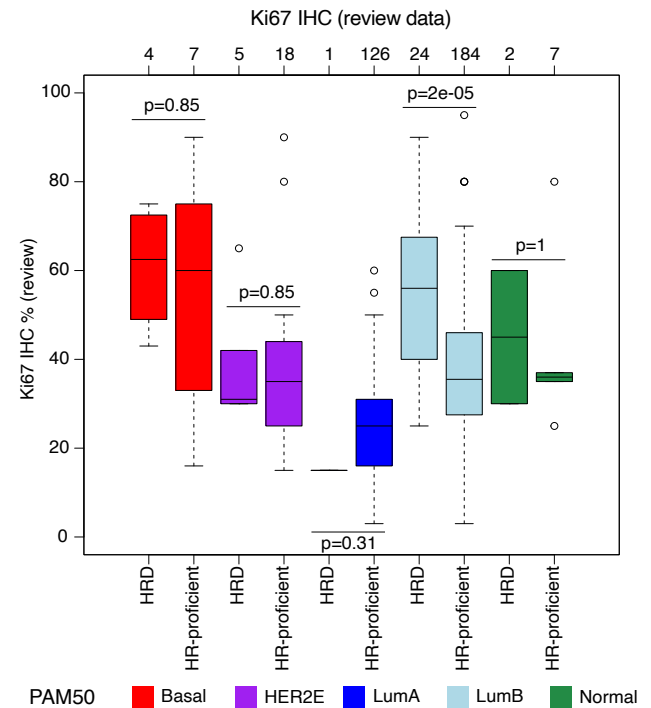
Ttest FDR DEGs HRD: LumB

DEG, nbr genes plotted= 11730 N significant genes= 664



FC cutoff, 1; p-value cutoff, 0.05

D)

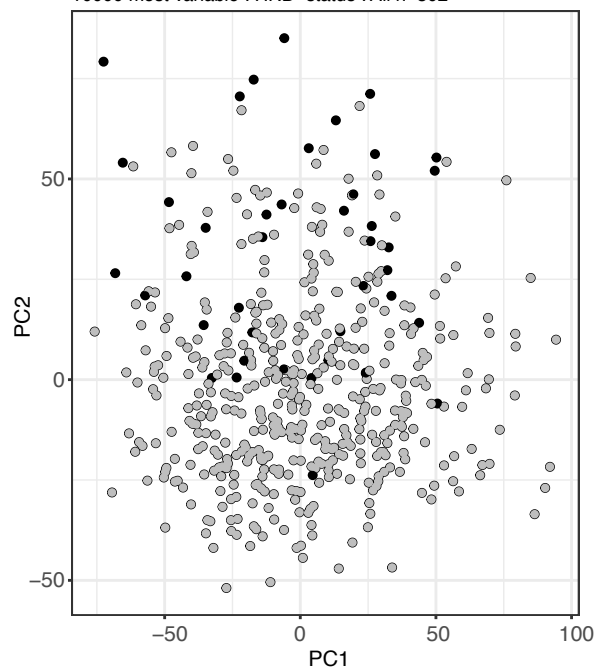


PAM50 Basal HER2E LumA LumB Normal

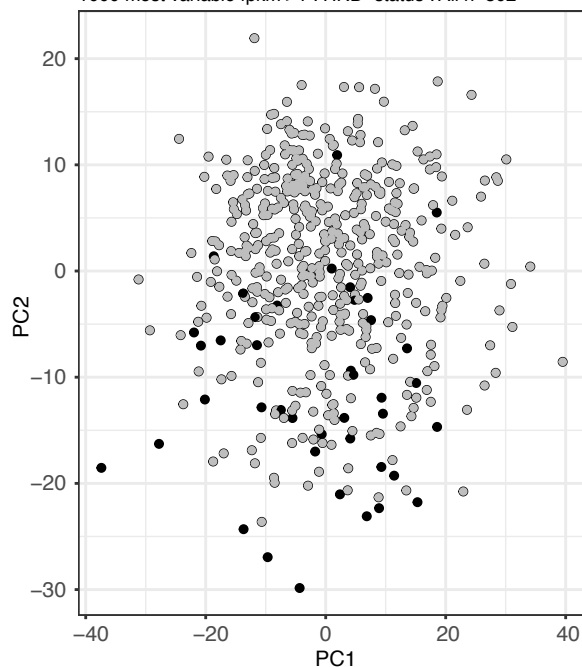
Supplementary Figure S4. Gene expression analysis of ERpHER2n tumors with respect to HRD status. (A) Volcano plot showing the results of the differential gene expression analysis for HRD-status in all 502 tumors. (B) Volcano plot showing the results of the differential gene expression analysis for HRD-status in HER2E tumors. (C) Volcano plot showing the results of the differential gene expression analysis for HRD-status in LumB tumors. In A-C, the fold change (FC) cut-off was 1, and the p-value FDR cut-off 0.05. (D) Ki67 IHC expression based on clinical review data for tumors stratified by PAM50 and HRD status, showing higher Ki67 levels in HRD LumB tumors. Two-sided p-values calculated using Wilcoxon's test. (E) Principal component analysis (PCA) of FPKM data in all ERpHER2n tumors using the top 10000 most varying genes. HRD tumors are marked by black points. Principal component 1 and 2 (PC1 and PC2, respectively) are shown. (F) PCA plot of all tumors using the 1000 most variant genes with a minimum FPKM level >1. (G) PCA plot of LumB tumors using the 10000 most variant genes. (H) PCA plot of LumB tumors using the 1000 most variant genes with a minimum FPKM level >1.

E)

10000 most variable : HRD-status : All n=502

**F)**

1000 most variable fpkm >1 : HRD-status : All n=502

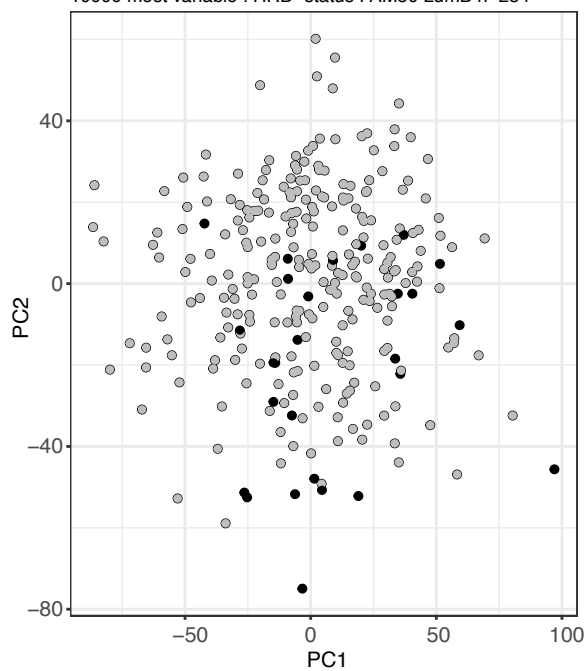


HRD status

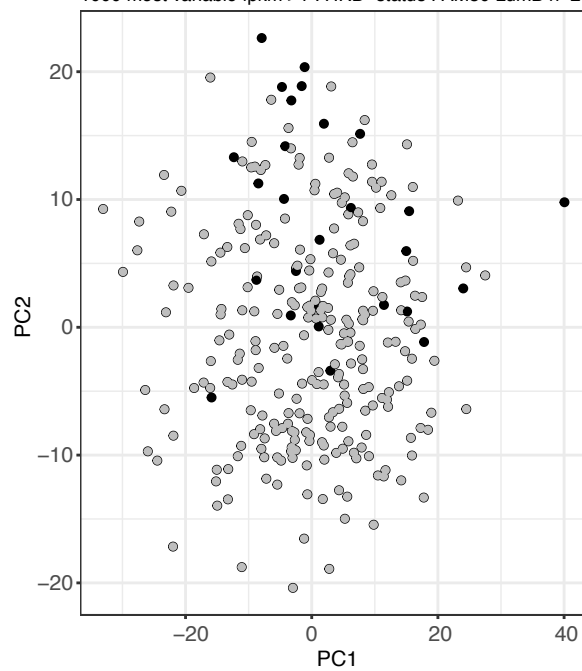
● HRP ● HRD

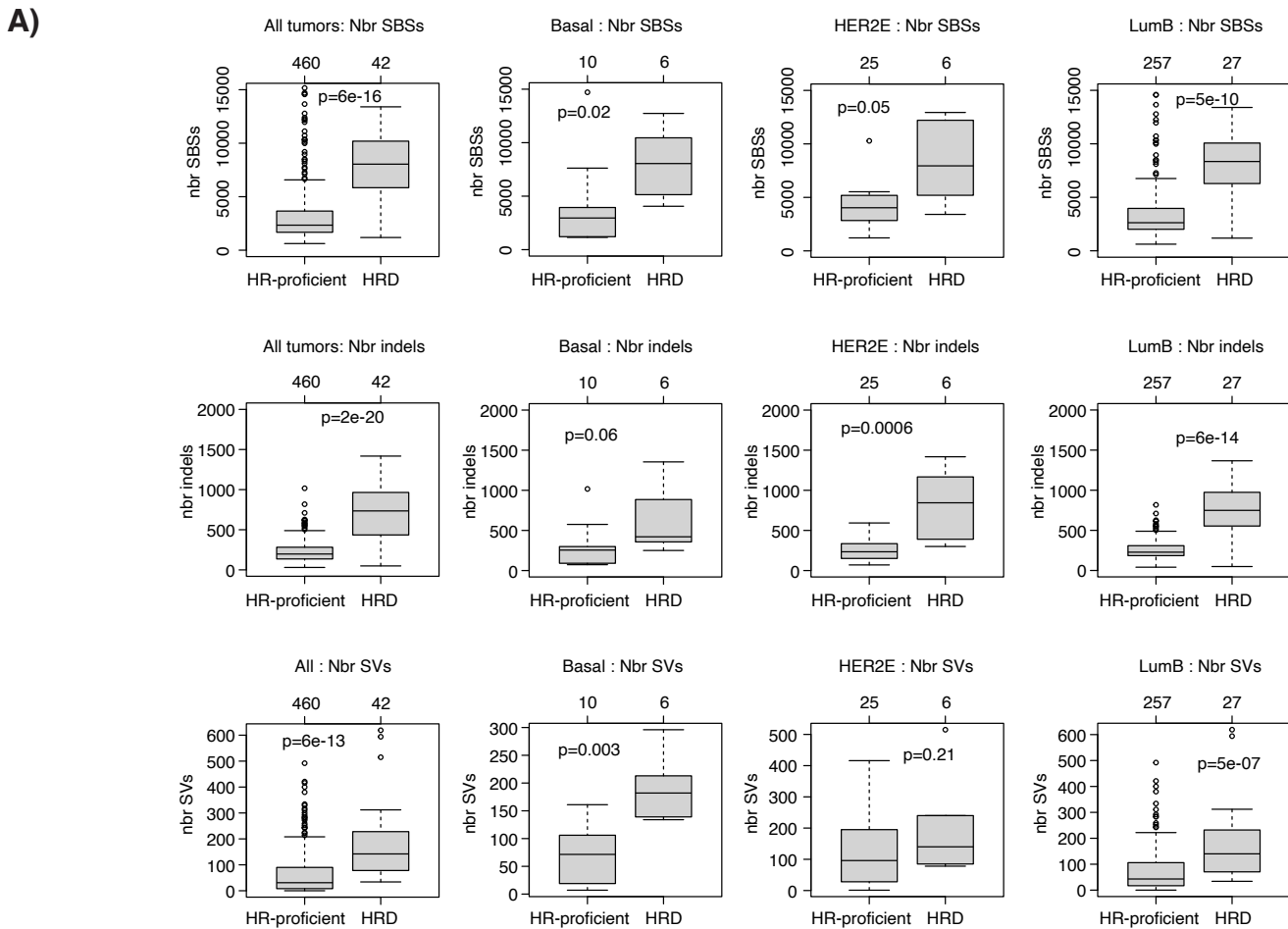
G)

10000 most variable : HRD-status PAM50 LumB n=284

**H)**

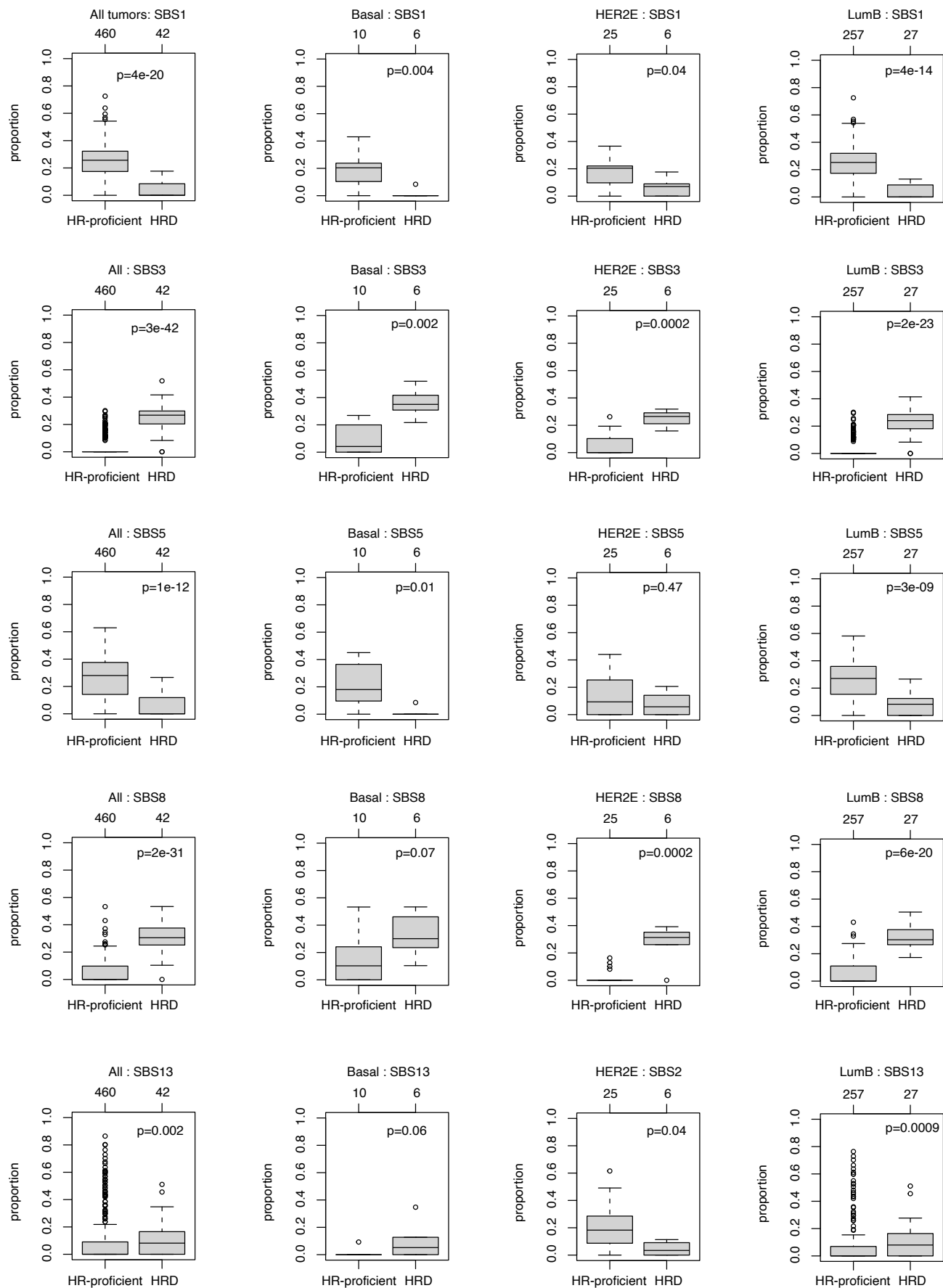
1000 most variable fpkm >1 : HRD-status PAM50 LumB n=284





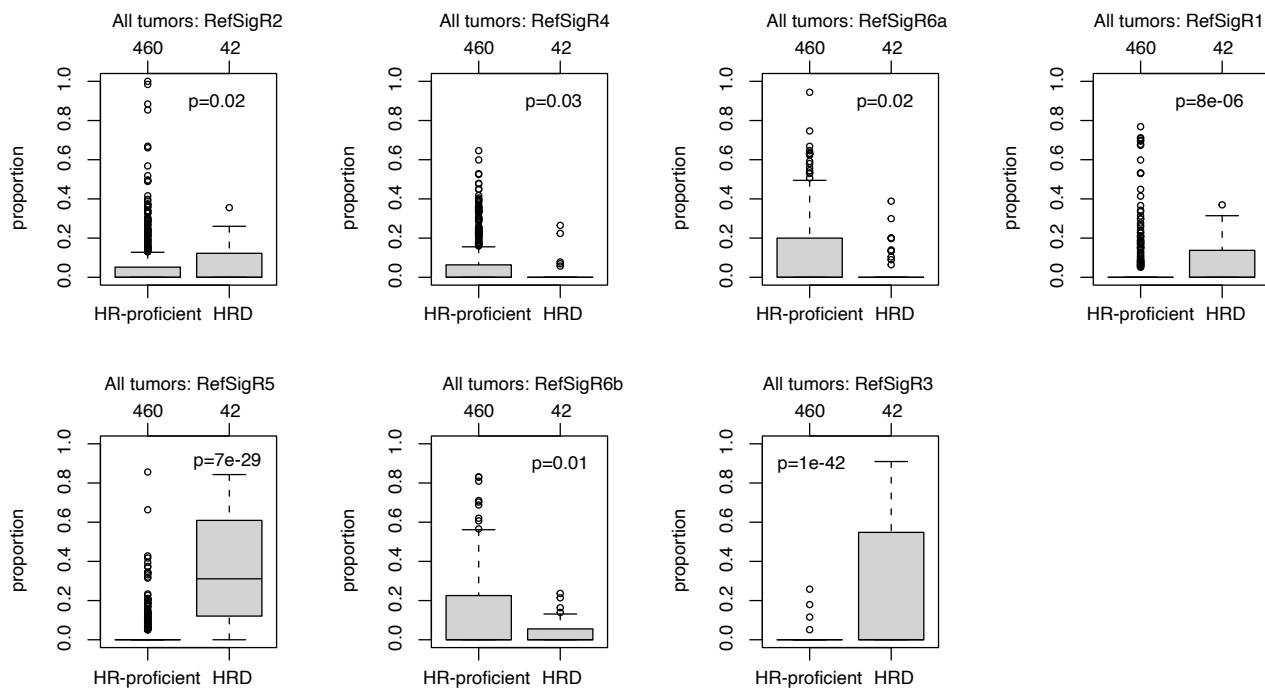
Supplementary Figure S5. Genetic features of HRD ERpHER2n tumors. (A) Number of single base substitutions (SBSs), indels, and structural rearrangements (SVs) in HRD versus HR-proficient tumors stratified by PAM50 subtype (LumA and Normal excluded due to few HRD cases). Two-sided p-values are calculated using Wilcoxon's test. **(B)** Exposure to SBS signatures for all tumors, Basal, HER2E, and LumB. Only SBS1, SBS3, SBS5, SBS8, and SBS2/13 are shown as these were the signatures with significant differences between groups. Two-sided p-values are calculated using Wilcoxon's test. **(C)** Exposure to SV signatures for all tumors, Basal, HER2E, and LumB subtypes. Only signatures significant or borderline nonsignificant are shown. Two-sided p-values are calculated using Wilcoxon's test. **(D)** Exposure to indel signatures for all tumors, Basal, HER2E, and LumB subtypes. Only signatures with significant differences are shown. Two-sided p-values are calculated using Wilcoxon's test. **(E)** WGS estimated tumor ploidy, fraction of the genome altered by copy number alterations (CN_FGA), fraction of the genome altered by LOH (LOH_FGA), and number of ASCAT segments on chromosome 1-22 per tumor called as gain or loss for all tumors, Basal, HER2E, and LumB subtypes. **(F)** Barplot of FDR adjusted Wilcoxon's test p-values (-log10 transformed) for 25 copy number signatures (CN1-25) between HR-proficient and HRD tumors stratified by PAM50 subtype. E.g. for CN12, significant p-values (FDR>0.05) was found for the comparison of HRD vs HR-proficient tumors in the total cohort and in LumB tumors specifically. **(G)** Left: frequency of gene driver events based on somatic SBSs and indels in PAM50 Basal tumors stratified by HRD status. Right: frequency of gene driver events based on somatic structural rearrangements (SVs) in PAM50 Basal tumors stratified by HRD status. Only genes detected in at least two tumors in at least one group are shown. **(H)** As in G but for PAM50 HER2E tumors. **(I)** Principal component analysis (PCA) of tumor purity adjusted DNA methylation data (beta values) for three different CpG contexts in all tumors and PAM50 subtypes (LumA excluded as only one tumor was HRD). For each context, the 5000 most variant CpGs in each subgroup of samples were selected for PCA. Black dots represent tumors with an HRDetect HRD classification. First two principal components (PC1 and PC2) shown. **(J)** PCA of tumor purity adjusted DNA methylation data (beta values) for three different CpG contexts in LumB HRD tumors. For each context, the 5000 most variant CpGs in the subgroup were selected for PCA. Samples are colored based on their proposed HRD inactivation mechanism. First two principal components (PC1 and PC2) shown.

B)

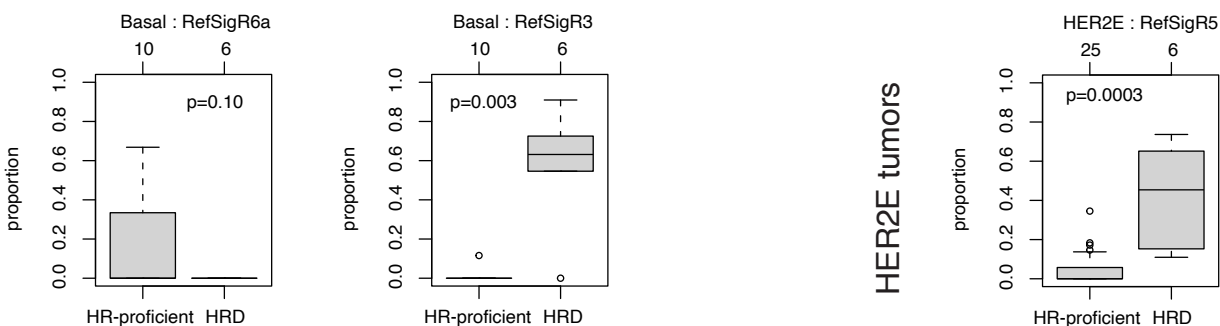


c)

All tumors

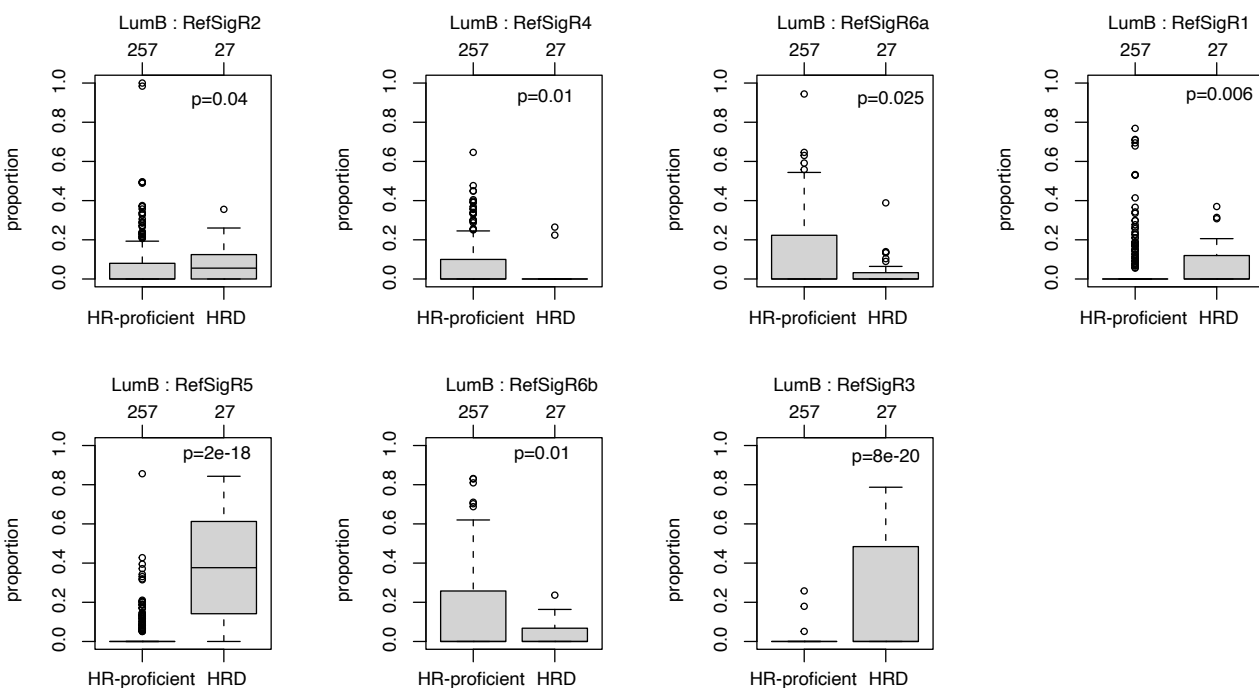


Basal tumors



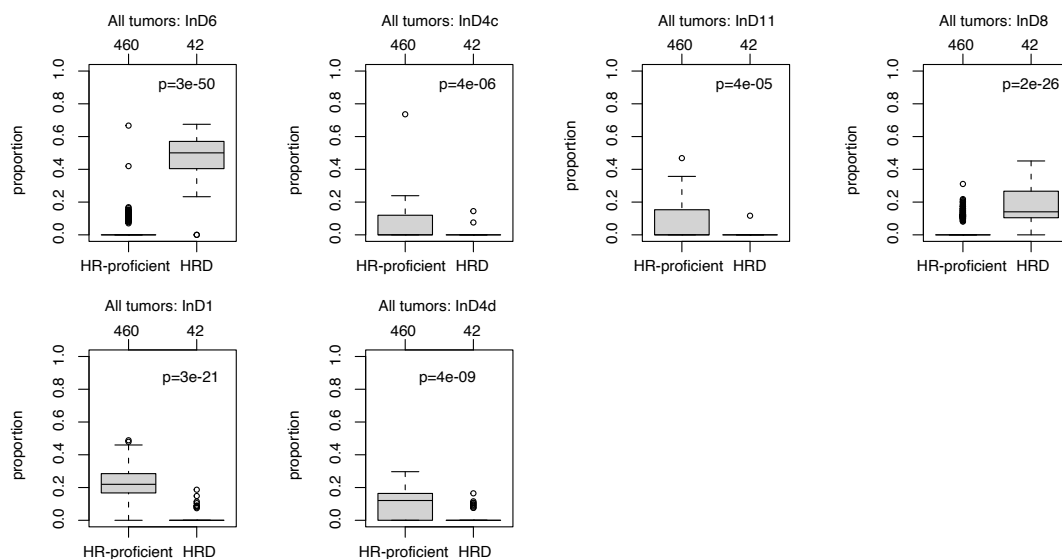
HER2E tumors

LumB tumors

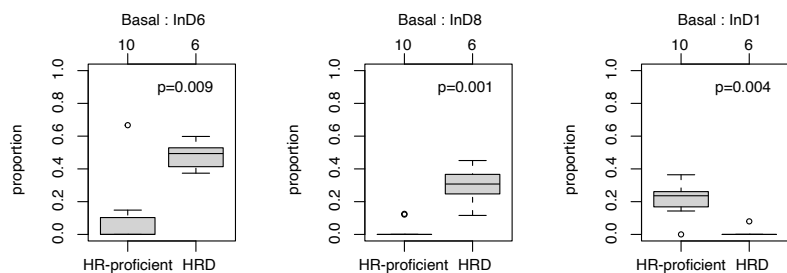


D)

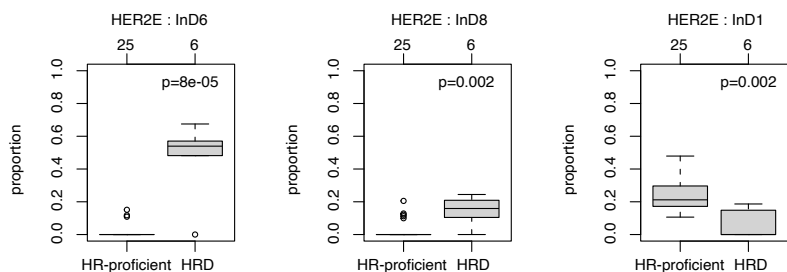
All tumors



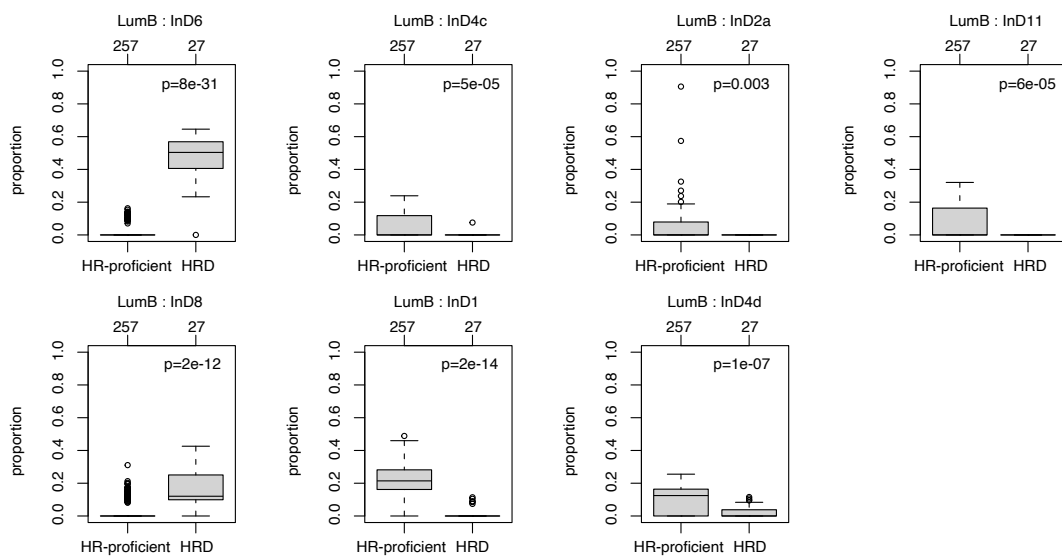
Basal tumors



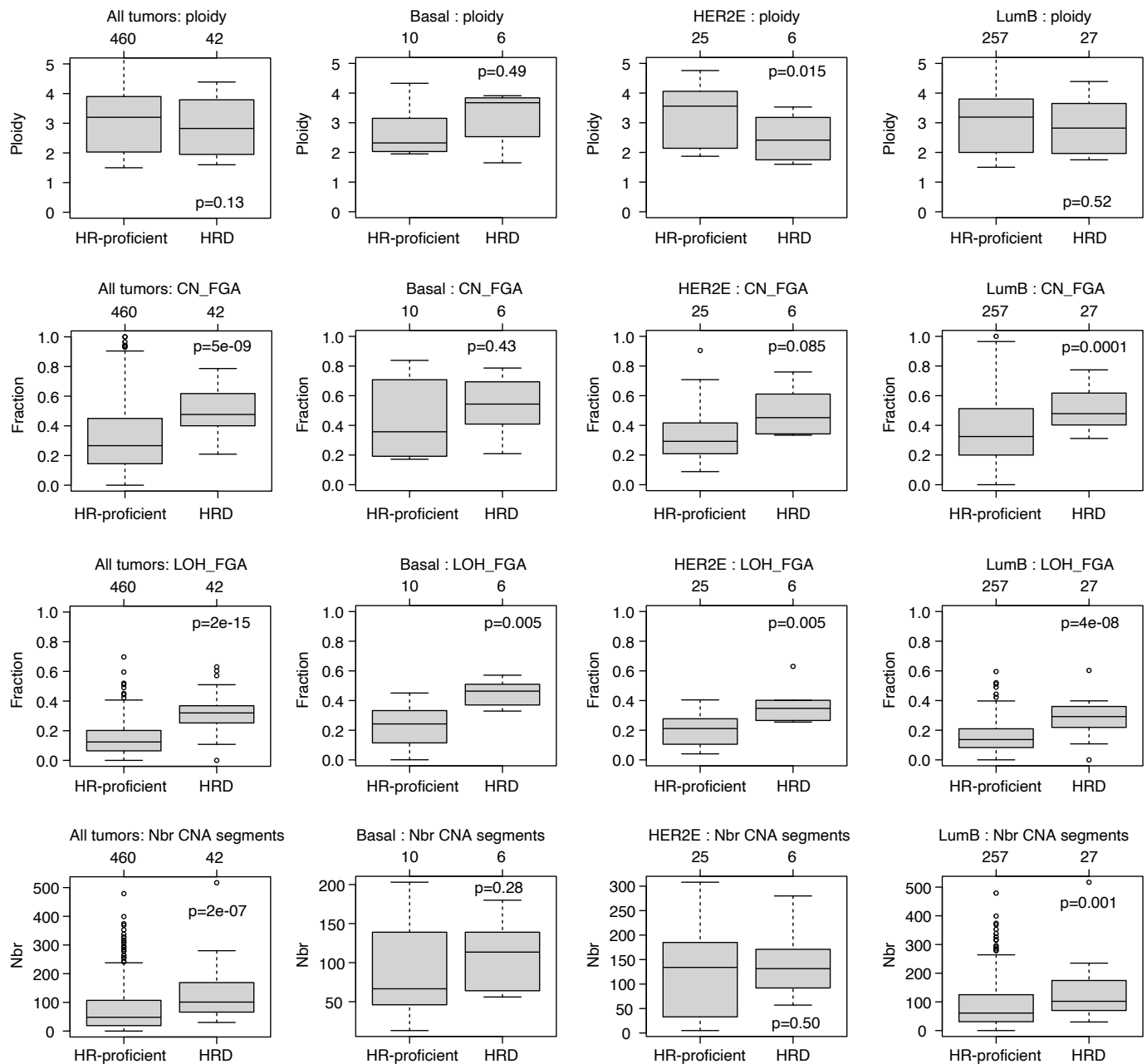
HER2E tumors



LumB tumors

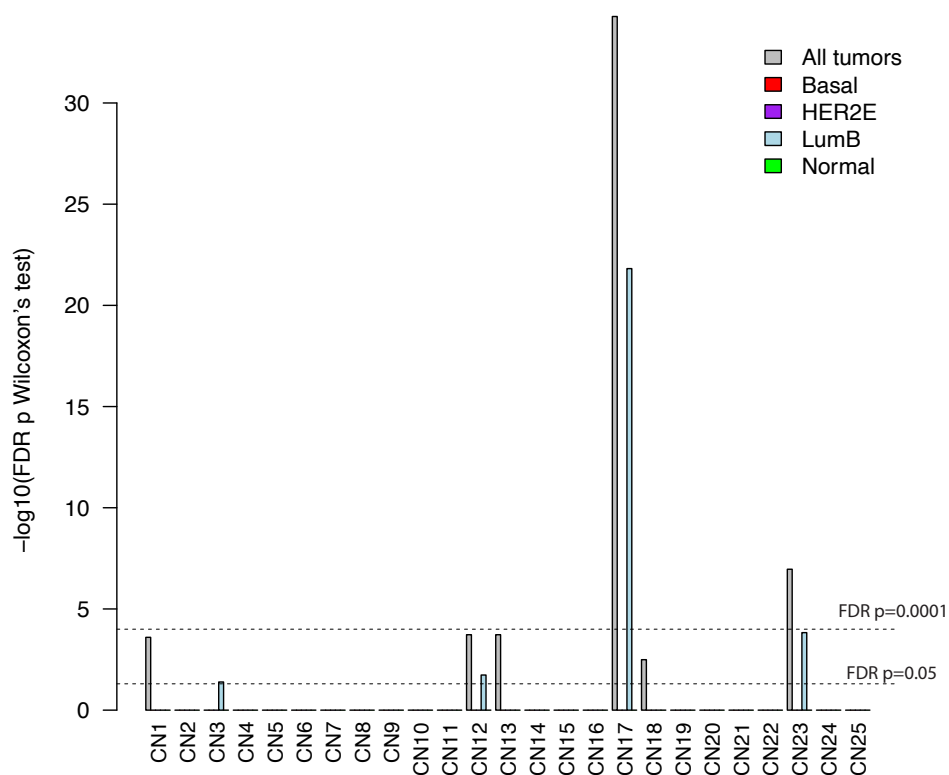


E)



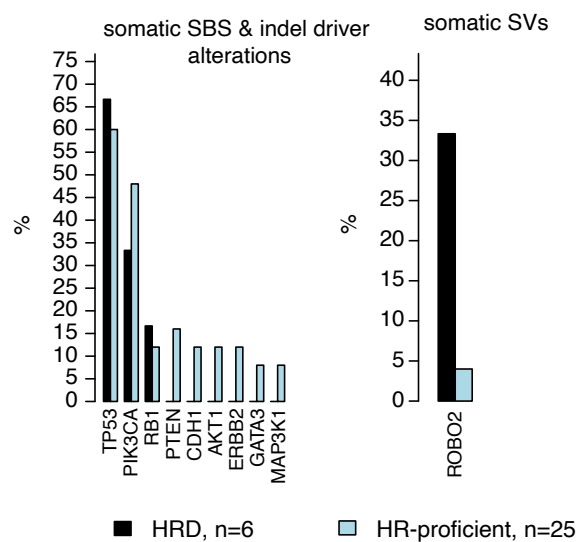
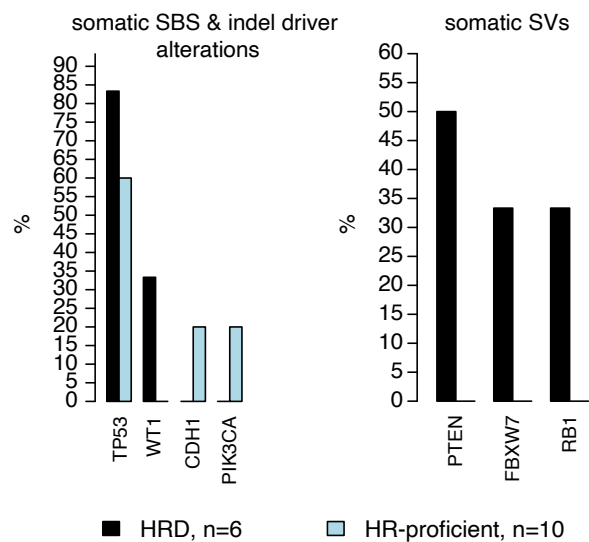
F)

CN signature proportions vs HRDetect



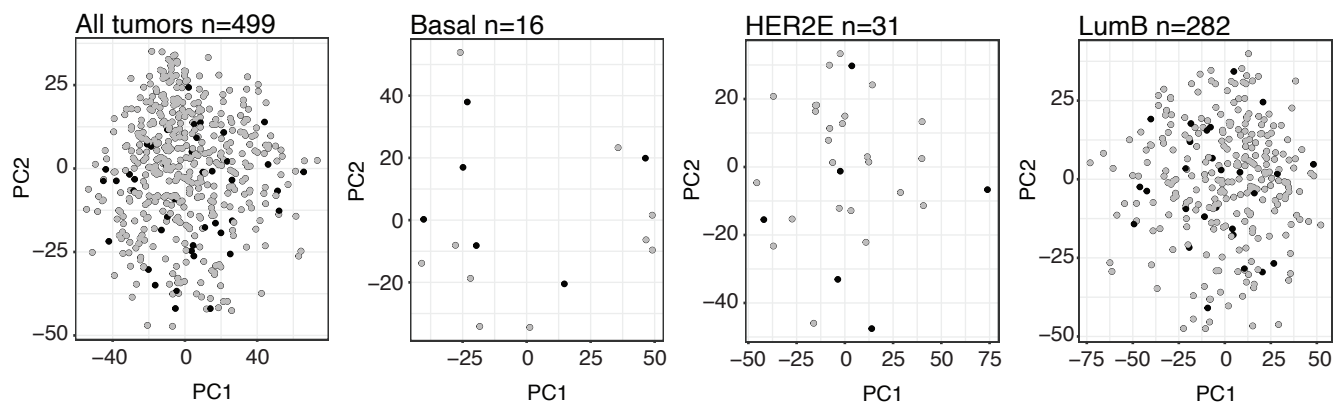
G) PAM50 Basal

H) PAM50 HER2E

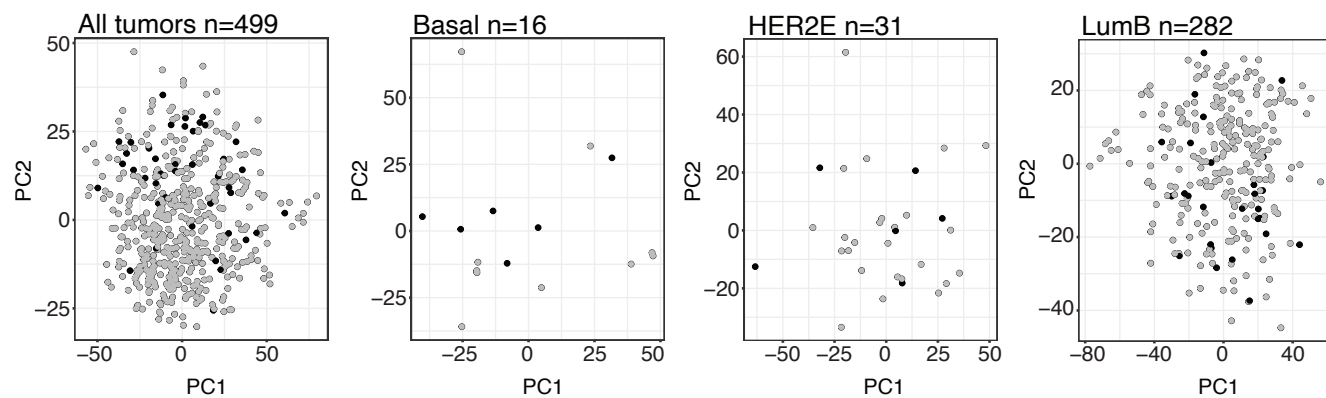


I)

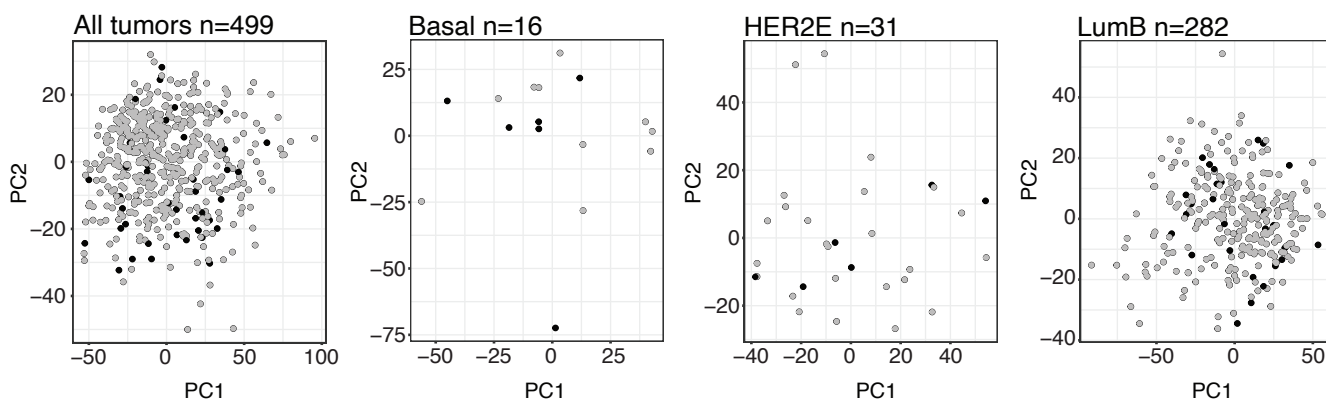
distal-ATAC



proximal-ATAC

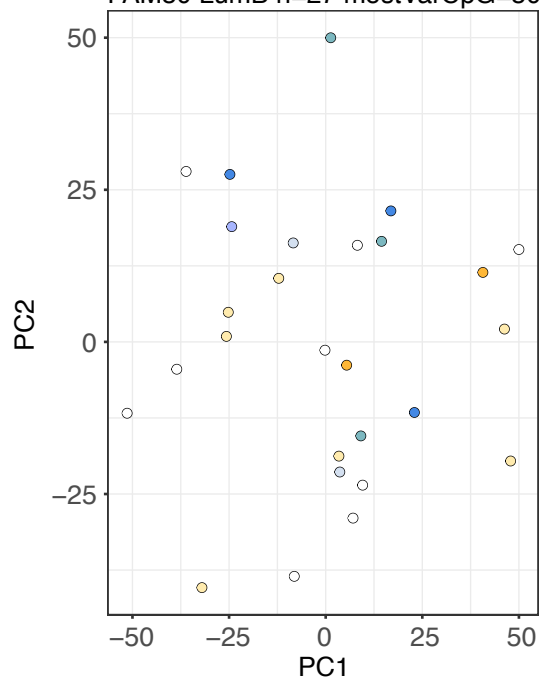


promoter-ATAC

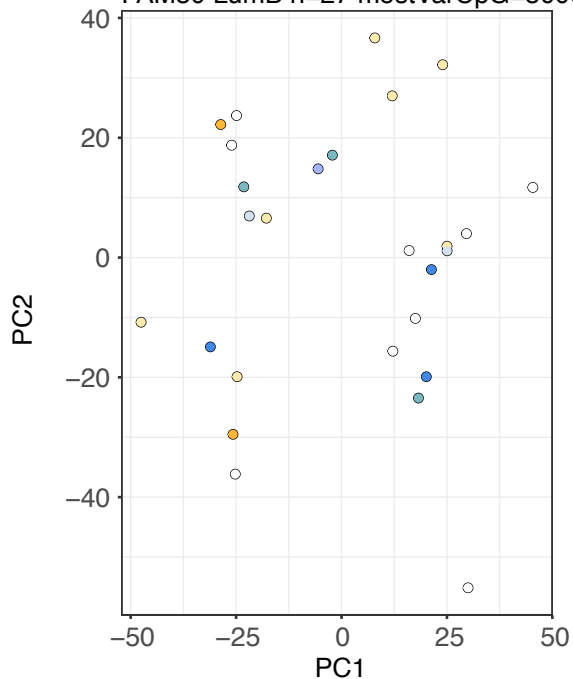


J)

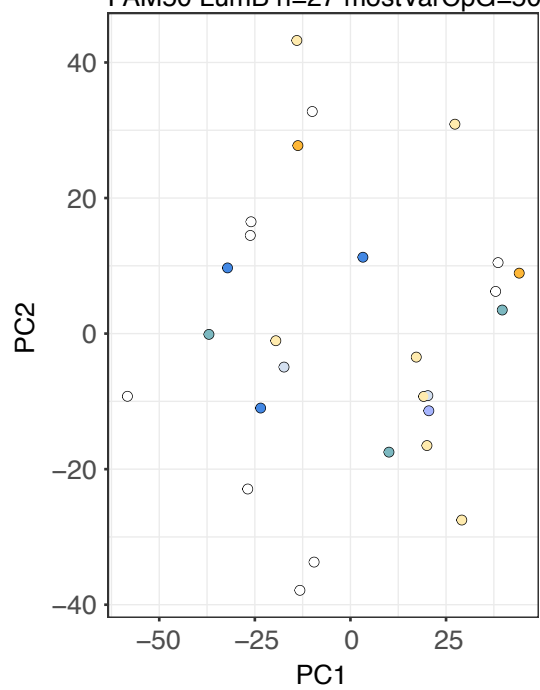
distal-ATAC: HRD inactivation
PAM50 LumB n=27 mostVarCpG=5000



proximal-ATAC: HRD inactivation
PAM50 LumB n=27 mostVarCpG=5000



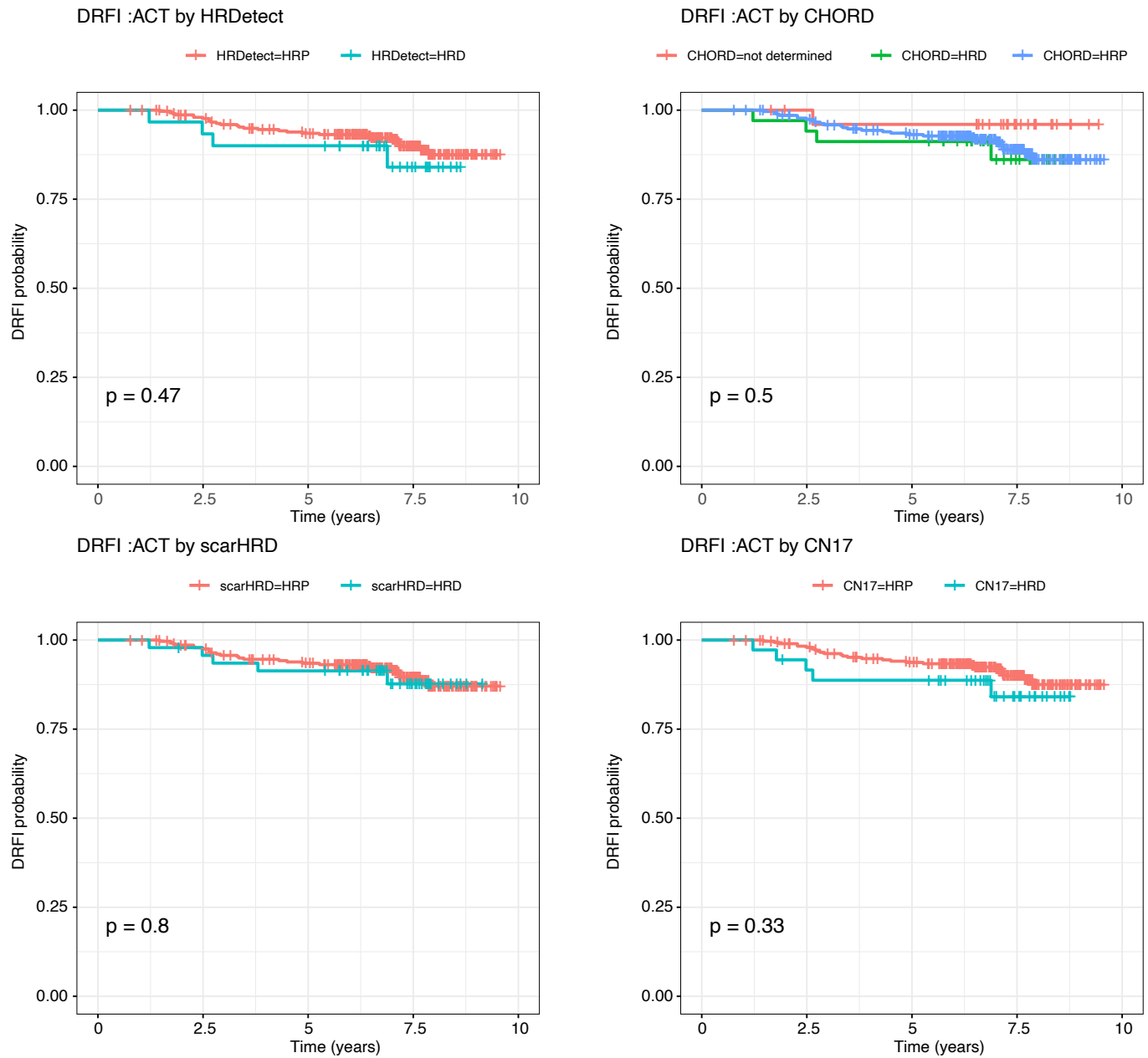
promoter-ATAC: HRD inactivation
PAM50 LumB n=27 mostVarCpG=5000



Inactivation mechanism

- BRCA1promoter
- BRCA1somatic
- BRCA2germline
- BRCA2somatic
- PALB2somatic
- RAD51Cpromoter
- Unknown

A)



Supplementary Figure S6. Survival analysis of alternative HRD methods and multivariate Cox regression in ChemoEndo ERpHER2n BC. (A) Kaplan-Meier plots of alternative HRD classification methods using distant relapse-free interval (DRFI) as clinical endpoint in ChemoEndo patients. P-values computed using the log-rank test. **(B)** Forest plot of hazard ratios and 95% confidence intervals for a multivariate Cox regression model including a merged molecular class (PAM50 LumA yes/no and HRD status), Nottingham grade index (NHG), binary age, and binary lymph node status in ChemoEndo treated ERpHER2n patients using distant relapse-free interval (DRFI) as clinical endpoint. In this analysis, WGS-analyzed ChemoEndo treated patients was merged with 761 non-overlapping ChemoEndo treated patients with unknown HRD status stratified by their PAM50 LumA status (LumA or not-LumA). For the 761 patients the DRFI data based on cancer registry data from, while for the WGS analyzed samples clinical review DRFI data was used as endpoint.

B)

Multivariate ERpHER2n: ChemoEndo: DRFI-review+DRFI-registry

