JNCI 16-0349R2

Article

Intrinsic subtype and therapeutic response among HER2-positive breast tumors from the NCCTG (Alliance) N9831 trial

Edith A. Perez, M.D.[1,2], Karla V. Ballman, Ph.D.[3], Afshin Mashadi-Hossein, M.S.[4], Kathleen S. Tenner, B.S.[3], Jennifer M. Kachergus, M.S.[2], Nadine Norton, Ph.D.[2], Brian M. Necela, Ph.D.[2], Jennifer M. Carr, B.S.[2], Sean Ferree, Ph.D.[4], Charles M. Perou, Ph.D.[5], Frederick Baehner, M.D.[6], Maggie Chon U. Cheang, Ph.D.[5,7], E. Aubrey Thompson, Ph.D.[2]

[1]Department of Medicine, Division of Hematology/Oncology, and Department of Cancer Biology, Mayo Clinic Comprehensive Cancer Center, Mayo Clinic, Jacksonville, FL

[2]Department of Cancer Biology, Mayo Clinic Comprehensive Cancer Center, Mayo Clinic, Jacksonville, FL

[3]Department of Health Science Research, Division of Biostatistics and Bioinformatics, Mayo Clinic Comprehensive Cancer Center, Mayo Clinic, Rochester, MN

[4]Nanostring, Inc., Seattle, WA

[5]Departments of Genetics and Pathology & Laboratory Medicine, Lineberger Comprehensive Cancer Center, University of North Carolina, Chapel Hill, NC.

[6]Genonic Health, Inc. Redwood CA

[7]Clinical Trials and Statistics Unit, The Institute of Cancer Research, London, UK

Corresponding author: E. Aubrey Thompson, Ph.D. Departments of Cancer Biology, Griffin Cancer Research Bld., Rm 214, Mayo Clinic, 4500 San Pablo Rd, Jacksonville, FL 32224; phone: (904) 953-6226; fax: (904) 953-6233; e-mail: thompson.aubrey@mayo.edu

ABSTRACT

Background: **Genomic data from HER2+ tumors were analyzed to assess the association between intrinsic subtype and clinical outcome in a large, well-annotated patient cohort.**

Methods: Samples from the NCCTG (Alliance) N9831 trial were analyzed using the Prosigna™ algorithm on the NanoString® platform to define intrinsic subtype, Risk of Recurrence scores, and Risk categories for 1392 HER2+ tumors. Subtypes were evaluated for recurrence-free survival (RFS) using Kaplan-Meier and Cox model analysis following adjuvant chemotherapy (n=484) or chemotherapy plus trastuzumab (n=908). All statistical tests were two-sided.

Results: Patients with HER2+ tumors from N9831 were primarily scored as HER2-enriched (72.1%) These individuals received statistically significant benefit from trastuzumab (HR=0.68, 95%CI=0.52-0.89, p=0.005), as did the patients (291/1392) with Luminal-type tumors (HR=0.52, 95%CI=0.32-0.85, p=0.01). Patients with Basal-like tumors (97/1392) did not have statistically significantly better RFS when treated with trastuzumab and chemotherapy compared to chemotherapy alone (HR=1.06, 95%CI=0.53-2.13, p=0.87).

Conclusions: The majority of clinically-defined HER2-positive tumors were classified as HER2-enriched or Luminal using the Prosigna algorithm. Intrinsic subtype alone cannot replace conventional histopathological evaluation of HER2 status, since many tumors that are classified as Luminal A or Luminal B will benefit from adjuvant trastuzumab if that subtype is accompanied by HER2 overexpression. However, among tumors that overexpress HER2, we speculate that assessment of intrinsic subtype may influence treatment, particularly with respect to evaluating alternative therapeutic approaches for that subset of HER2-positive tumors of the Basal-like subtype.

INTRODUCTION

With the advent of high dimensional technologies for quantifying expression of large numbers of genes, it has become clear that breast tumors exhibit a considerable range of molecular/genomic heterogeneity, both between and among the clinically defined cohorts of hormone receptor-positive, HER2-positive, and triple negative tumors. The most thoroughly characterized classifier of molecular heterogeneity is the PAM50 signature [1, 2] which uses the expression of 50 genes to stratify breast tumors into four major classes: Basal-like, HER2-enriched, Luminal A, and Luminal B. There is a strong presupposition that these biological processes will prove to be informative of clinical/pathological features of breast tumors [2-6].

Our studies have focused upon defining molecular heterogeneity among HER2-positive (HER2+) breast tumors. We used the NanoString® Prosigna™ algorithm and PAM50 genes to define intrinsic subtypes and assess their distribution and association with outcome within tumors from the North Central Cancer Treatment Group [NCCTG (Alliance)] N9831 trial of patients with early stage HER2+ tumors [7, 8], randomized to receive adjuvant chemotherapy or chemotherapy plus trastuzumab. Our primary analytical focus was to test the hypothesis that benefit from adjuvant chemotherapy combined with trastuzumab, compared to benefit from adjuvant chemotherapy alone, varies as a function of intrinsic subtype.

MATERIALS AND METHODS

Patients

The NCCTG (Alliance) N9831 trial enrolled 3505 patients who were randomized to 3 study arms: Arm A - chemotherapy (doxorubicin and cyclophosphamide followed by weekly paclitaxel); Arm B - chemotherapy followed by 52 weeks of weekly trastuzumab; and Arm C -

chemotherapy with 12 weeks of trastuzumab concurrent with paclitaxel followed by 40 more weeks of weekly trastuzumab alone, as previously described in detail [7]. Patients in both trastuzumab-treated arms (B and C) received 52wk of trastuzumab and showed statistically significant improvement in relapse-free survival [7, 8], compared to patients who received chemotherapy alone (Arm A). We therefore combined patients from Arms B and C into a single trastuzumab-treated group (Arm B/C) for subsequent analysis. All patients included in this analysis were assessed for central HER2 IHC and/or FISH as part of the N9831 parent study, as previously described [7]. Informed written consent to these studies was obtained from all patients. Studies were carried out under Mayo Clinic Institutional Review Board protocols 954-00 and 13-000290.

Statistical Analysis

The main objective of the analysis was to determine whether the effect of trastuzumab qualitatively differed for any of the intrinsic subtypes. If differing treatment effects were observed among the subtypes, a test for interaction was to be performed to determined whether this observation was statistically significant. The primary endpoint for outcome analysis was recurrence-free survival (RFS), defined as the time from randomization to breast cancer recurrence (local, regional, or distant recurrence of breast cancer or breast cancer related death). The time to event for patients who died without recurrence was considered censored at the time of death. Cox proportional hazard models were used to generate point estimates of hazard ratios (HRs) and corresponding 95% confidence intervals (95%CI) to assess the benefit of trastuzumab for RFS comparisons. The assumption of proportional hazards was verified with a Cox proportional hazard model containing the treatment indicator and the interaction of the indicator with the log of time. The test of proportionality involves a likelihood ratio test of the model containing the interaction term against the model that contains the treatment variable only. A non-significant p-value of this test suggests there is not sufficient evidence to reject the

assumption of proportional hazards. The p-value of this test was 0.88, suggesting that there is not sufficient evidence to reject the assumption of proportional hazards.Wald Chi-square p-values were calculated for the Cox hazard ratios. Comparisons were made both within and between treatment arms. The interaction term of Basal-like vs Non-Basal-like by treatment arm was determined using a Cox model that also included the main effects; this interaction was tested because the effect of trastuzumab appeared qualitatively different in this subtype. Kaplan-Meier plots were used to depict the proportion of patients free from breast cancer recurrence as a function of time. Chi-square tests for nominal categories or Cochran-Mantel-Haenszel tests for ordered categories were performed to ascertain differences in eligible N9831 patients included in the analysis with those that were excluded. All statistical tests were two-sided and a P value of less than 0.05 was considered statistically significant.

Assignment of intrinsic subtype

We constructed a NanoString® custom CodeSet that includes the PAM50 genes plus appropriate housekeeping genes (B2M, GAPDH, POLR2A, UBC, YWHAZ) for normalization purposes. The Prosigna™ subtyping algorithm is trained for use with Prosigna™ CodeSet. To use this algorithm for subtyping N9831 data, adjustment factors that account for differences between the custom CodeSet used in the N9831 study and the Prosigna™ CodeSet were estimated and applied to the data. To this end, 30 samples, which were processed with the custom CodeSet, were re-processed using the Prosigna™ CodeSet. For each probe, this provided 30 pairs of counts (corresponding to the two CodeSets), which were then used to estimate the adjustment factor between the two CodeSets. Specifically, for each of the PAM50 probes, first, the ratio of the normalized counts from the custom CodeSet relative to their corresponding normalized counts from the Prosigna™ CodeSet were calculated, resulting in 30 ratios for each of the probes. Then, for each probe, the median of the 30 ratios were taken as the adjustment factor. Subsequently, the resultant 50 adjustment factors (one for each PAM50

gene) were applied to the PAM50 counts in the original dataset and subtyping was performed on the adjusted counts per standard Prosigna™ algorithm. Quantitative real time PCR data on mRNA abundance, proliferative scores, and mitotic indices were abstracted from a published study on the OncotypeDX analysis of N9831 samples[9].

RESULTS

Assignment of risk scores and intrinsic subtype using Prosigna™

Samples from 1426 eligible patients had sufficient RNA for testing and were analyzed on the NanoString® platform to measure the abundance of mRNAs that define the intrinsic subtype patterns. Thirty four samples were excluded for quality control issues, mostly low total gene counts; 1392 remaining samples (Arm A: 484, Arm B: 494, Arm C: 414) were analyzed for intrinsic subtype in a blinded fashion. Patient demographics for samples that were included as well as those from whom RNA was not available are given in **Supplementary Table 1**.

Distributions of intrinsic subtypes, Risk of Recurrence scores (ROR), and Risk categories (High, Intermediate, Low) as a function of treatment arm are given in **Table 1**, and clinical/pathological features of the tumors by treatment are in **Supplementary Table 2**. There were small, but statistically significant differences in nodal status between the two arms (p=0.04). However, the observed increase in node 0 and node 1-3 patients in Arm B/C did not translate into differences in Risk category (**Table 1**).

The patients enrolled in N9831 were centrally confirmed as HER2+. The majority of the tumors (1004/1392, 72.1%) were classified by Prosigna™ as HER2-enriched. The remaining tumors were more or less equally distributed among the Basal-like (97/1392, 7.0%), Luminal A (132/1392, 9.5%), or Luminal B (159/1392, 11.4%) subtypes.

Association between risk of recurrence score, risk category, and RFS

The Prosigna™ algorithm reports a risk of recurrence (ROR) score that combines intrinsic subtype, tumor size, and proliferation score into a single metric. The majority of the N9831 samples (1045/1392) had ROR scores >70, where low risk is considered 1-40, intermediate risk 40-60, and high risk >60. We did not observe a statistically significant association between RFS and ROR, expressed as a discontinuous, quantile score, among patients who received chemotherapy alone (Arm A, **Figure 1A**, p=0.40) or patients who received chemotherapy plus trastuzumab (Arms B/C, **Figure 1B**, p=0.37).

Tumor size is incorporated into ROR to stratify tumors into High, Intermediate, and Low Risk categories. Almost all of the N9831 samples were classified as High Risk (1335/1392, **Table 1**). The Kaplan-Meier analysis suggested a quantitative tendency towards improved RFS in the few Low/Intermediate Risk patients who received chemotherapy alone; however, this tendency did not achieve statistical significance (Arm A, **Figure 1C**, p=0.13). Low/Intermediate Risk patients who received trastuzumab appeared to receive statistically significant benefit (Arm B/C, **Figure 1D** p=0.01). Among the tumors with Low/Intermediate Risk, those classified as estrogen receptor and/or progesterone receptor-positive seldom recurred. Among 30 such patients, only 1 recurrence was reported, compared to the overall recurrence rate of about 1 in 4 (317/1392) for all patients in this study (Chi Square p=0.01).

A subset of the N9831 patients (445 from Arm A and 397 from Arm C) has previously been analyzed using the 21 gene OncotypeDX™ panel[9]. As with the Prosigna™ analysis, almost all of the patients were classified as high risk (763/842). A number of molecular features associated with subtype were derived from the 21 gene signature, including ERBB2 mRNA, proliferative score, and Ki67 mRNA. Percent cells staining 3+ for HER2 was extracted from the

clinical record, and mitotic index was assessed by histological examination, as shown in **Table 2**. The proliferative markers associated with each subtype were ranked Basal-like>HER2-enriched≈Luminal B>Luminal A. HER2 expression generally followed the pattern HER2-enriched>Luminal A>Luminal B≈>Basal-like. Among the four subtypes, the Basal-like tumors exhibited the lowest HER2 expression and the highest proliferation, features that might be expected to influence outcome following HER2-targeted therapy.

Association between intrinsic subtype and RFS

A statistically significant association between RFS and intrinsic subtype was observed when all patients (irrespective of therapy) were evaluated (**Figure 2A**). Patients with Basal-like tumors exhibited statistically significantly worse RFS compared to patients with HER2-enriched, Luminal A, or Luminal B subtypes. When intrinsic subtype was evaluated within each treatment arm, there was no statistically significant association between subtype and RFS in the patients who received chemotherapy alone (**Figure 2B**, log rank p=0.79; HR_{Basal}=reference; HR_{HER2}=0.75, 95%CI=0.42-1.33, p=0.32; HR_{LumA}=0.76, 95% CI=0.35-1.64, p=0.32; HR_{LumB}=0.83, 95%CI=0.41-1.69, p=0.61). However, a statistically significant association was observed between outcome and intrinsic subtype among patients who received trastuzumab (**Figure 2C**, log rank p=0.007; HR_{Basal}=reference; HR_{HER2}=0.49, 95%CI=0.30-0.78, p=0.003; HR_{LumA}=0.47, 95%CI=0.25-0.90, p=0.02; HR_{LumB}=0.35, 95%CI=0.18-0.68, p=0.002).

RFS as a function of treatment and subtype

The data in **Figure 2** suggest that there is a relationship between intrinsic subtype and outcome and that this relationship resides primarily in those patients who received adjuvant trastuzumab. We therefore evaluated the relationship between RFS and treatment within each subtype. Trastuzumab appeared to have little or no statistically significant association with RFS in

HER2+ patients of the Basal-like subtype (**Figure 3A**, HR=1.06, 95%CI=0.53-2.13, p=0.87). Patients with HER2-enriched genomic profiles received statistically significant benefit from trastuzumab, compared to chemotherapy alone (**Figure 3B**, HR=0.68. 95%CI=0.52-0.89, p=0.005). Patients with Luminal A profiles exhibited a quantitative trend towards improved RFS, although this trend did not achieve statistical significance (blue curve vs green curve in **Figure 3C**, HR=0.62, 95%CI=0.30-1.26, p=0.20). HER2+ tumors with Luminal B profiles benefitted statistically significantly (red curve vs brown curve in **Figure 3C**, HR=0.45, 95%CI=0.22-0.88, p=0.02). Due to the small number of events in patients with Luminal-type tumors, we combined the Luminal A and B tumors into a single cohort of Luminal-type tumors for subsequent analyses. Trastuzumab provided statistically significant benefit in patients whose tumors exhibited Luminal profiles (**Figure 3D**, HR=0.52, 95%CI=0.32-0.85, p=0.01).

A Cox model was used to evaluate RFS among Basal and Non-Basal tumors. The HRs for RFS of Basal-like tumors in Arm A did not appear to be statistically significantly different from those of Non-Basal tumors after chemotherapy (**Table 3A**, Arm A Basal-like vs Arm A Non-Basal, HR=0.76, 95%CI=0.43-1.34, p=0.34). In addition, RFS was not statistically significantly different between Basal-like tumors after chemotherapy versus after chemotherapy plus trastuzumab (**Table 3B**, Arm A Basal-like vs Arm B/C Basal-like, HR=1.06, 95%CI=0.53-2.13, p=0.87). As expected, RFS among the Non-Basal cohort was statistically significantly improved by trastuzumab (**Table 3B**, Arm A Non-Basal vs Arm B/C Non-Basal, HR=0.65, 95%CI=0.51-0.82, p<0.001). Moreover, RFS after trastuzumab was statistically significantly better among Non-Basal-like tumors (**Table 3A**, Arm B/C Basal-like vs Arm B/C Non-Basal-like, HR=0.47, 95%CI=0.29-0.74, p=0.001). However, the test for interaction of treatment arm (A vs B/C) by tumor subtype (Basal vs Non-Basal) did not achieve statistical significance (p=0.20), which may or may not be due to the small number of Basal-like tumors in our sample cohort.

Hormone receptor status and outcome among Non-Basal tumors

Since Basal-like tumors are generally estrogen receptor (ER) and/or progesterone receptor (PR)-negative, we evaluated the association between subtype, outcome, and ER/PR status. Patients with ER/PR-positive or ER/PR-negative non-Basal-like tumors had similar RFS after chemotherapy (**Table 3C**, Arm A ER/PR-neg vs Arm A ER/PR-pos, HR=0.90, 95%CI=0.63-1.28, p=0.55). Likewise, ER/PR status was not statistically significantly associated with outcome after trastuzumab treatment of Non-Basal tumors (**Table 3C**, Arm B/C ER/PR-neg vs Arm B/C ER/PR-pos, HR=0.84, 95% CI=0.62-1.14, p=0.27). Moreover, both patients with ER/PR-positive and negative tumors with Non-Basal features appeared to benefit from trastuzumab (Arm A ER/PR-neg vs Arm B/C ER/PR-neg, HR=0.67, 95% CI=0.47-0.85, p=0.02; Arm A ER/PR-pos vs Arm B/C ER/PR-pos, HR=0.62, 95% CI=0.45-0.85, p=0.003).

DISCUSSION

Intrinsic subtype, defined by abundance of transcripts that correspond to the PAM50 cohort of genes, has gained wide acceptance as a tool for stratifying breast tumors based on the extent to which these tumors exhibit expression profiles that correspond to Basal-like, HER2-enriched, Luminal A, or Luminal B samples. All of the samples included in this analysis were centrally evaluated for HER2 protein or gene copy number at Mayo Clinic[7]. Thus, our data reflect molecular heterogeneity within clinically well-define HER2+ tumors, such that some HER2+ tumors exhibit Basal-like or Luminal gene expression profiles. These results are consistent with others published elsewhere[10-12]. Evaluation of several molecular features associated with subtype suggests that there are statistically significant differences in HER2 expression and proliferation. Most notably, the Basal-like tumors exhibited statistically significantly lower HER2 mRNA and percent of cells staining 3+ for HER2, coupled with high proliferative markers. Both of these features might plausibly be associated with outcome.

To our knowledge, this is the first report of the use of the Prosigna™ algorithm to define intrinsic subtype distribution and outcome following adjuvant trastuzumab therapy of HER2+ tumors. Our data are broadly consistent with analyses of intrinsic subtype among HER2+ tumors from The Cancer Genome Atlas [10], the NOAH neoadjuvant [11], and the NSABP B31 adjuvant trastuzumab trials [12], in which the research PAM50 algorithm was used to assess molecular heterogeneity among tumors that are clinically defined as HER2+. The key question is to what extent this molecular heterogeneity is associated with clinical outcome following trastuzumab. As reported in the NOAH and NSABP B31/PAM50 analysis we observed that HER2+ tumors with HER2-enriched profiles are generally responsive to trastuzumab. Thus, the HER2-enriched profile is predictive of better outcome following adjuvant trastuzumab. The data with the HER2+ Luminal subtypes are somewhat more difficult to evaluate. These patients tend to do very well and few events were recorded. We observed a qualitative trend towards increased RFS in trastuzumab-treated patients with both Luminal A and Luminal B tumors, although statistical significance was not achieved with the Luminal A subgroup. Nevertheless, the data suggest that HER2+ patients whose tumors express Luminal signatures likely receive additional benefit from trastuzumab, above and beyond that received from chemotherapy alone.

The majority of the HER2+ tumors were classified as high risk, by both Prosigna™ and OncotypeDX™. It is unclear if the small subset of patients with Low/Intermediate Risk, estrogen and/or progesterone receptor-positive early stage HER2+ tumors require adjuvant trastuzumab. The number of such patients enrolled in N9831 was small.Additional studies will be required to assess the relationship between Risk category and RFS among patients with hormone-receptor positive tumors.

The data indicate that Non-Basal HER2+ tumors (i.e., those tumors with HER2-enriched, Luminal A or Luminal B profiles) are responsive to trastuzumab, irrespective of hormone receptor status. A subset, about 7%, of HER2+ tumors in the N9831 study exhibited gene

expression profiles that resemble those of Basal-like tumors. Similar results were observed in analysis of 1711 samples from The Cancer Genome Atlas Network, 2012 #16062} and METABRIC databases (14.1% Basal-like)[13] and 1579 samples from the NSABP B31 trial (6.5% Basal-like) [12]. The data presented here suggest that HER2+ tumors with Basal-like features may benefit less from adjuvant trastuzumab. This suggestion is consistent with the molecular features of these tumors, which express low levels of HER2 and appear to be highly proliferative. However, the conclusion that patients with such tumors will not benefit from trastuzumab is not warranted by the data. Although the hazard ratio for RFS following trastuzumab is not statistically significantly different from that observed with chemotherapy alone (HR=1.06, p=0.87), the 95% confidence intervals are very large (0.53 to 2.13), and the test for interaction between treatment and subtype (Basal vs Non-Basal) was not statistically significant. Moreover, Pogue-Geile et al. reported that Basal-like tumors in the NSABP B31 trial received statistically significant benefit from adjuvant trastuzumab [12]. There are some differences in the algorithms used to call intrinsic subtype in our study and that reported by Pogue-Geile. They reported disease free survival, whereas we report breast cancer relapse free survival. Perhaps more importantly, they used a modified version of the PAM50 algorithm, rather than the Prosigna™ algorithm used in our studies. These two algorithms use different normalization strategies; and they called 47% of the HER2+ tumors in B31 as HER2-enriched, whereas we called 73% of the N9831 HER2+ samples as HER2-enriched. Furthermore, both studies were limited in power by the relative small number of Basal-like tumors that were identified. Given these considerations, we are inclined to the most conservative conclusion: that additional studies will be required to sort out the role of intrinsic subtype in predicting response to HER2-targeted therapy in the adjuvant setting.

There are a couple of limitations of this study. The first is that it is a retrospective analysis of samples from less than half of the patient enrolled on N9831. This due to the fact

that the submission of patient samples was optional for this study and that some samples were dropped due to poor quality. It is recognized that this could bias the results. A comparison of patients with samples used in this analysis to patients not included in the analysis only showed some clinical-pathological differences (e.g. patients with larger tumors are included in this analysis), but there was no difference observed in the treatment effect between the patients included in this study and patients not included in this study. Another limitation is that some of the identified intrinsic subtype groups had a relatively small number of patients and events. This limits the power of the analyses in these subgroups, which makes statistically non-significant results hard to interpret: they could be reflective of no real difference between the groups or could be due to lack of power to detect a difference of interest.

Finally, our data indicate that evaluation of HER2 status by IHC and/or FISH is an essential component of clinical management of breast cancer patients. Intrinsic subtype alone cannot replace conventional histopathological evaluation of HER2 status, since many tumors that are classified as Luminal A or Luminal B will benefit from adjuvant trastuzumab if that subtype is accompanied by HER2 overexpression. However, among tumors that overexpress HER2, we speculate that assessment of intrinsic subtype may guide further treatment development, particularly with respect to evaluating alternative therapeutic approaches for that subset of HER2+ tumors of the Basal-like subtype. Given the relative rarity of such tumors, additional studies will be required to evaluate this idea.

Funding: These studies were supported in part by grants from the National Cancer Institute (CA129949), the Breast Cancer Research Foundation, the 26.2 with Donna Foundation, and NCI Breast SPORE CA58223 (to C.M.P.). Additional support for infrastructure was derived from the Mayo Clinic Comprehensive Cancer Center support grant CA15083 and the Mayo Foundation.

Notes

The funders had no role in the design of the study; the collection, analysis, and interpretation of the data; the writing of the manuscript; and the decision to submit the manuscript for publication.

Drs. Edith A. Perez, E. Aubrey Thompson, and Karla V. Ballman had full access to all the data in the study: and these individuals take full responsibility, jointly and severally, for the integrity of the data and the accuracy of the data analysis.

Potential conflicts of interest include Nanostring employment for Drs. Mashadi-Hossein and Ferree. Dr. Perou is a board of director's member, consultant, and stock holder in Bioclassifier LLC as well as a patent application on PAM50 assay with royalties paid to Nanostring Technologies. Dr. Cheang has a patent "Gene expression profiles to predict breast cancer outcomes (PAM50 classifier)" issued. Dr. Baehner is an employee of Genomic Health, Inc. All other authors declare no conflicts of interest.

Dr. Perez was involved in the conception and design of the manuscript as well as provision of the study material or patients. Collection and assembly of data were provided by Drs. Ballman and Perez. Dr. Thompson had overall responsibility for data accumulation, analysis, and interpretation. All authors contributed to the analysis and interpretation of data as well as

drafting and final approval of the manuscript, and all authors agree to be accountable for all aspects of the work.

REFERENCES

1. Perou CM, Sorlie T, Eisen MB, et al. Molecular portraits of human breast tumours. *Nature.* 2000;406(6797):747-752.
2. Parker JS, Mullins M, Cheang MC, et al. Supervised risk predictor of breast cancer based on intrinsic subtypes. *J Clin Oncol.* 2009;27(8):1160-1167.
3. Perou CM, Parker JS, Prat A, Ellis MJ, Bernard PS. Clinical implementation of the intrinsic subtypes of breast cancer. *Lancet Oncol.* 2010;11(8):718-719; author reply 720-711.
4. Chia SK, Bramwell VH, Tu D, et al. A 50-gene intrinsic subtype classifier for prognosis and prediction of benefit from adjuvant tamoxifen. *Clin Cancer Res.* 2012;18(16):4465-4472.
5. Dowsett M, Sestak I, Lopez-Knowles E, et al. Comparison of PAM50 risk of recurrence score with oncotype DX and IHC4 for predicting risk of distant recurrence after endocrine therapy. *J Clin Oncol.* 2013;31(22):2783-2790.
6. Esserman LJ, Berry DA, Cheang MC, et al. Chemotherapy response and recurrence-free survival in neoadjuvant breast cancer depends on biomarker profiles: results from the I-SPY 1 TRIAL (CALGB 150007/150012; ACRIN 6657). *Breast Cancer Res Treat.* 2012;132(3):1049-1062.
7. Perez EA, Romond EH, Suman VJ, et al. Four-year follow-up of trastuzumab plus adjuvant chemotherapy for operable human epidermal growth factor receptor 2-positive breast cancer: joint analysis of data from NCCTG N9831 and NSABP B-31. *J Clin Oncol.* 2011;29(25):3366-3373.
8. Perez EA, Romond EH, Suman VJ, et al. Trastuzumab plus adjuvant chemotherapy for human epidermal growth factor receptor 2-positive breast cancer: planned joint analysis

of overall survival from NSABP B-31 and NCCTG N9831. *Journal of clinical oncology : official journal of the American Society of Clinical Oncology.* 2014;32(33):3744-3752.

9. Perez EA, Baehner FL, Butler SM, et al. The relationship between quantitative human epidermal growth factor receptor 2 gene expression by the 21-gene reverse transcriptase polymerase chain reaction assay and adjuvant trastuzumab benefit in Alliance N9831. *Breast Cancer Res.* 2015;17(1):133.

10. Network TCGA. Comprehensive molecular portraits of human breast tumours. *Nature.* 2012;490(7418):61-70.

11. Prat A, Bianchini G, Thomas M, et al. Research-based PAM50 subtype predictor identifies higher responses and improved survival outcomes in HER2-positive breast cancer in the NOAH study. *Clinical cancer research : an official journal of the American Association for Cancer Research.* 2014;20(2):511-521.

12. Pogue-Geile KL, Song N, Jeong JH, et al. Intrinsic Subtypes, PIK3CA Mutation, and the Degree of Benefit From Adjuvant Trastuzumab in the NSABP B-31 Trial. *Journal of clinical oncology : official journal of the American Society of Clinical Oncology.* 2015.

13. Prat A, Carey LA, Adamo B, et al. Molecular features and survival outcomes of the intrinsic subtypes within HER2-positive breast cancer. *Journal of the National Cancer Institute.* 2014;106(8).

TABLES

Table 1. Comparison of Prosigna intrinsic subtype, Risk of Recurrence (ROR), and Risk Category in Arm A (chemotherapy alone) and Arm B/C (chemotherapy plus trastuzumab) of N9831.

Subtype	**Arm A**	**Arm B/C**	**p-value***
Basal	38 (7.9%)	59 (6.5%)	0.76
HER2	342 (70.7%)	662 (72.9%)	
Luminal A	47 (9.7%)	85 (9.4%)	
Luminal B	57 (11.7%)	102 (11.2%)	
ROR			
1. <70	140 (28.9%)	207 (22.8%)	0.03
2. 70-81	116 (23.9%)	220 (24.2%)	
3. 82-91	108 (22.3%)	240 (26.4%)	
4. 92-100	120 (24.7%)	241 (26.5%)	
Risk category			
High	462 (95.4%)	873 (96.1%)	0.54
Intermediate/Low	22 (4.5%)	35 (3.8%)	

* p-value from Chi-square test for nominal categories or Cochran-Mantel-Haenszel test for ordered categories.

Table 2: Basal-like HER2+ tumors exhibit lower ERBB2 expression and higher proliferation*

Subtype	ERBB2 mRNA (95% CI)	p-value†	% IHC 3+ cells (95%CI)	p-value†	Proliferative score (95% CI)	p-value†	Ki67 mRNA (95% CI)	p-value†	Mitotic score (95% CI)	p-value†
Basal-like (n=56)	10.8 (11.3,10.4)	ref	49.2 (61.9,36.5)	ref	6.91 (7.03,6.78)	ref	7.93 (8.07,7.79)	ref	2.55 (2.74,2.36)	ref
HER2-enriched (n=607)	13.0 (13.1,12.9)	<0.001	91.2 (92.9,89.4)	<0.001	6.52 (6.56,6.48)	<0.001	7.50 (7.55,7.45)	<0.001	2.01 (2.08,1.94)	<0.001
Luminal A (n=75)	11.7 (12.0,11.4)	0.002	64.6 (74.6,54.6)	0.03	5.78 (5.90,5.65)	<0.001	6.79 (695,6.62)	<0.001	1.37 (1.52,1.22)	<0.001
Luminal B (n=104)	11.1 (11.4,10.6)	0.11	45.8 (55.2,36.4)	0.67	6.56 (6.68,6.44)	<0.001	7.49 (7.61,7.37)	<0.001	1.56 (1.71,1.40)	<0.001
Luminal All (n=179)	11.4 (11.6,11.2)	0.01	54.2 (61.2,47.3)	0.42	6.23 (6.34,6.13)	<0.001	7.20 (7.30,7.09)	<0.001	1.48 (1.60,1.37)	<0.001

*Data are summarized from 842 samples from Arm A (445) and Arm C (397 that were analyzed using the OncotypeDX^{TM} 21 gene signature [9]. Proliferative score is the average of 5 proliferative marker genes in this panel. Mitotic score was determined by histological examination by F. Baehner. All mRNA abundance data were normalized and scaled to log base 2. IHC staining for HER2 abundance (%3+ cells) were extracted from the clinical record, as previously described. **Abbreviations: % IHC 3+ = percent of tumor cells with membrane scoring 3+ by clinical pathology, Ki67=mRNA encoded by *MKI67***

† Mann-Whitney U test, two sided p-value for comparison to Basal-like tumors.

Table 3. HER2+ tumors with Basal-like signatures have different outcomes than HER2+ tumors with Non-Basal signature*

Panel A: RFS as a function of subtype (Basal vs Non-Basal) by treatment arm						
Arm	Subtype	N	Events(%)	HR	95%CI	p-value†
A	Basal-like	38	13 (34.2)	1.0		
A	Non-Basal-like	446	123 (27.5)	0.76	0.43, 1.34	0.34
B/C	Basal-like	59	20 (33.8)	1.0		
B/C	Non-Basal-like	849	161 (19.0)	0.47	0.29, 0.74	0.001
Panel B: RFS as a function of treatment (Arm A vs Arm B/C) by subtype						
Arm	Subtype	N	Events(%)	HR	95%CI	p-value†
A	Basal-like	38	13 (34.2)	1.0		
B/C	Basal-like	59	20 (33.8)	1.06	0.53, 2.13	0.87
A	Non-Basal-like	446	123 (27.5)	1.0		
B/C	Non-Basal-like	849	161 (19.0)	0.65	0.51, 0.82	<0.001
Panel C: RFS by ER/PR status as a function of Arm within Non-Basal tumors						
Arm	Hormone Receptor Status	N	Events(%)	HR	95%CI	p-value
A	ER/PR-negative	192	54 (28.1)	1.0		
A	ER/PR-positive	254	69 (27.2)	0.90	0.63, 1.28	0.55
B/C	ER/PR-negative	381	76 (19.9)	1.0		
B/C	ER/PR-positive	468	85 (18.2)	0.84	0.62, 1.14	0.27
Panel D: RFS by treatment Arm and ER/PR status with Non-Basal tumors						
Arm	Hormone Receptor Status	N	Events(%)	HR	95%CI	p-value†
A	ER/PR-negative	192	54 (28.1)	1.0		
B/C	ER/PR-negative	381	76 (19.9)	0.67	0.47, 0.95	0.02
A	ER/PR-positive	254	69 (27.2)	1.0		
B/C	ER/PR-positive	468	85 (18.2)	0.62	0.45, 0.85	0.003

***** Cox model analysis was carried out to compare RFS between Basal-like and Non-Basal tumors in Arm A or in Arm C (Panel A). In addition, the Cox model was used to compare outcome as a function of treatment (Arm A vs Arm B/C) as a function of Basal-like or Non-Basal subtype (Panel B). Test for interaction of Arm and Basal Subtype Status p=0.20. A Cox model analysis was used to define hazard ratio (HR) for recurrence as a function of ER/PR status and treatment (Panel C). Wald Chi-square p-values were calculated from the Cox model data. The effect of treatment (Arm A vs Arm B/C) is shown as a function of ER/PR status in Panel D. Test for interaction of Arm and HR Status p=0.79.

FIGURE LEGENDS

Figure 1: Risk of Recurrence (ROR) scores and RISK categories among HER2+ tumors. The Prosigna algorithm reports two scores, ROR and RISK. The association between ROR scores and RFS after chemotherapy (**Panel A**) or chemotherapy plus trastuzumab (**Panel B**) are shown. The majority of the HER2+ tumors were assigned as High Risk. The association between RISK category and RFS after chemotherapy (**Panel C**) or chemotherapy plus trastuzumab (**Panel D**) are given. Kaplan-Meier log rank survival analysis (two-sided) was used to calculate p-values.

Figure 2: Intrinsic subtype and relapse free survival among patients in the N9831 trial. The relationship between RFS and time, evaluated as a function of Prosigna subtype in all arms of the N9831 trial is shown in **Panel A**. The relationship between RFS and subtype in patients enrolled in Arm A, chemotherapy alone, in **Panel B**, whereas RFS as a function of subtype in patients enrolled in Arm B/C, chemotherapy plus trastuzumab, is given in **Panel C**.Kaplan-Meier log rank p-values (two sided) were calculated. The Cox model (two sided) was used to estimate hazard ratios.

Figure 3: HER2+ tumors with Basal-like expression profiles compared to tumors with HER2-enriched or Luminal-type expression profiles. Relapse-free survival, plotted as Kaplan-Meier curves was compared for Arm A (chemotherapy alone) versus Arm B/C (chemotherapy plus trastuzumab) for N9831 tumors with Basal-like (**Panel A**) or HER2-enriched (**Panel B**) Prosignal subtype profiles. Log rank statistics were used to calculate p-values for RFS as a continuous variable. HRs on the basis of % RFS 10yr following randomization. Survival analysis was also carried out for Luminal A and Luminal B subtypes, individually (**Panel C**) and combined into a Luminal-type category (**Panel D**). All statistical tests were two-sided.

Supplemental Information

Supplemental Table 1: Patient and Tumor Characteristics: Comparison of Nanostring Cohort vs Eligible Patients Not Included in the Cohort

Supplemental Table 2: Patient and Tumor Characteristics Comparison by Arm

Supplemental Table 1: Patient and Tumor Characteristics: Comparison of Nanostring Cohort vs Eligible Patients Not Included in the Cohort.

Characteristic	Eligible But Not Included (n=1740)	Nano-String Cohort (n=1392)	p-value*
Arm			
A	603 (34.6%)	484 (34.8%)	
B	602 (34.6%)	494 (35.5%)	0.80
C	535 (30.7%)	414 (29.7%)	
Age at Randomization (years)			
18-39	304 (17.5%)	239 (17.2%)	
40-49	609 (35.0%)	428 (30.7%)	0.03
50-59	553 (31.8%)	467 (33.5%)	
≥60	274 (15.7%)	258 (18.5%)	
Race			
Caucasian	1440 (82.8%)	1206 (86.6%)	0.003
Other	300 (17%)	186 (13%)	
Menopausal Status			
Premenopausal	965 (55.4%)	720 (51.7%)	0.04
Postmenopausal	775 (44.5%)	672 (48.3%)	
Extent of Surgery			
Breast sparing	671 (38.6%)	540 (38.8%)	0.90
Mastectomy	1069 (61.4%)	852 (61.2%)	
Extent of Nodal Examination			
Axillary node dissection	1556 (89.4%)	1266 (90.9%)	0.16
Sentinel biopsy	184 (10.6%)	126 (9.1%)	
Histologically Positive Nodes			
0	233 (13.4%)	183 (13.1%)	
1-3	847 (48.7%)	638 (45.8%)	0.15
4-9	440 (25.3%)	378 (27.2%)	
≥10	220 (12.6%)	193 (13.9%)	
Tumor Size (cm)			
≤ 2.0	722 (41.5%)	528 (37.9%)	
2.1 - 4.9	893 (51.3%)	723 (51.9%)	0.004
≥5	125 (7.2%)	141 (10.1%)	
Tumor Grade			
Low/Intermediate	473 (27.2%)	380 (27.3%)	
High	1244 (71.5%)	994 (71.4%)	0.95
Unknown	23 (1.6%)	18 (1.3%)	
Hormone Receptor Status			
ER or PR Positive	954 (54.8%)	735 (52.8%)	0.26
Other	786 (45.2%)	657 (47.2%)	
Histology			
Ductal	1641 (94.3%)	1321 (94.9%)	0.50
Other	97 (5.6%)	70 (5.0%)	

Characteristic	Eligible But Not Included (n=1740)	Nano-String Cohort (n=1392)	p-value*
Missing	2 (0.1%)	1 (0.1%)	
Hormonal Therapy			
Yes	922 (53.0%)	705 (51%)	0.16
No	804 (46.2%)	681 (49%)	
Missing	14 (0.8%)	6 (0.4%)	
Agent			
Tamoxifen	627 (36.0%)	471 (33.8%)	0.32
Other	295 (17.0%)	234 (16.8%)	
None	804 (46.2%)	681 (48.8%)	
Missing	14 (0.8%)	6 (0.4%)	

* p-value from Chi-square test for nominal categories or Cochran-Mantel-Haenszel test for ordered categories.

Supplemental Table 2: Patient and Tumor Characteristics Comparison by Arm.

Characteristic	Arm A (n=484)	Arm B/C (n=908)	p-value*
Age at Randomization (years)			
18-39	78 (16.1%)	161 (17.7%)	
40-49	153 (31.6%)	275 (30.3%)	0.89
50-59	167 (34.5%)	300 (33.0%)	
≥60	86 (18%)	172 (18.9%)	
Race			
Caucasian	426 (88.6%)	780 (85.9%)	0.27
Other	58 (12%)	128 (14.1%)	
Menopausal Status			
Premenopausal	247 (51.0%)	473 (52.1%)	0.71
Postmenopausal	237 (49.0%)	435 (48.2%)	
Extent of Surgery			
Breast sparing	188 (38.8%)	352 (38.8%)	0.98
Mastectomy	296 (61.1%)	556 (61.2%)	
Extent of Nodal Examination			
Axillary node dissection	439 (90.7%)	827 (91.1%)	0.82
Sentinel biopsy	45 (9.2%)	81 (8.9%)	
Histologically Positive Nodes			
0	58 (12.0%)	125 (13.8%)	
1-3	210 (43.4%)	428 (47.1%)	0.04
4-9	140 (28.9%)	238 (26.2%)	
≥10	76 (15.7%)	117 (12.9%)	
Tumor Size (cm)			
≤ 2.0	192 (39.7%)	336 (37.0%)	
2.1 - 4.9	247 (51.0%)	476 (52.4%)	0.27
≥5	45 (9.3%)	96 (10.6%)	
Tumor Grade			
Low/Intermediate	131 (27.1%)	249 (27.4%)	
High	345 (71.3%)	649 (71.4%)	0.93
Unknown	8	10	
Hormone Receptor Status			
ER or PR Positive	260 (53.7%)	475 (52.3%)	0.62
Other	224 (46.3%)	433 (47.7%)	
Histology			
Ductal	459 (94.8%)	862 (94.9%)	
Other	25 (5.2%)	45 (4.9%)	0.87
Missing	0 (0.0%)	1 (0.1%)	
Hormonal Therapy			
Yes	247 (51.0%)	458 (51%)	0.75
No	233 (48.1%)	448 (49%)	
Missing	4 (0.8%)	2 (0.2%)	

Characteristic	Arm A (n=484)	Arm B/C (n=908)	p-value*
Agent			
Tamoxifen	166 (34.3%)	305 (33.6%)	0.94
Other	81 (16.7%)	153 (16.8%)	
None	233 (48.1%)	448 (49.3%)	
Missing	4 (0.8%)	2 (0.2%)	

* p-value from Chi-square test for nominal categories or Cochran-Mantel-Haenszel test for ordered categories.

Figure 1

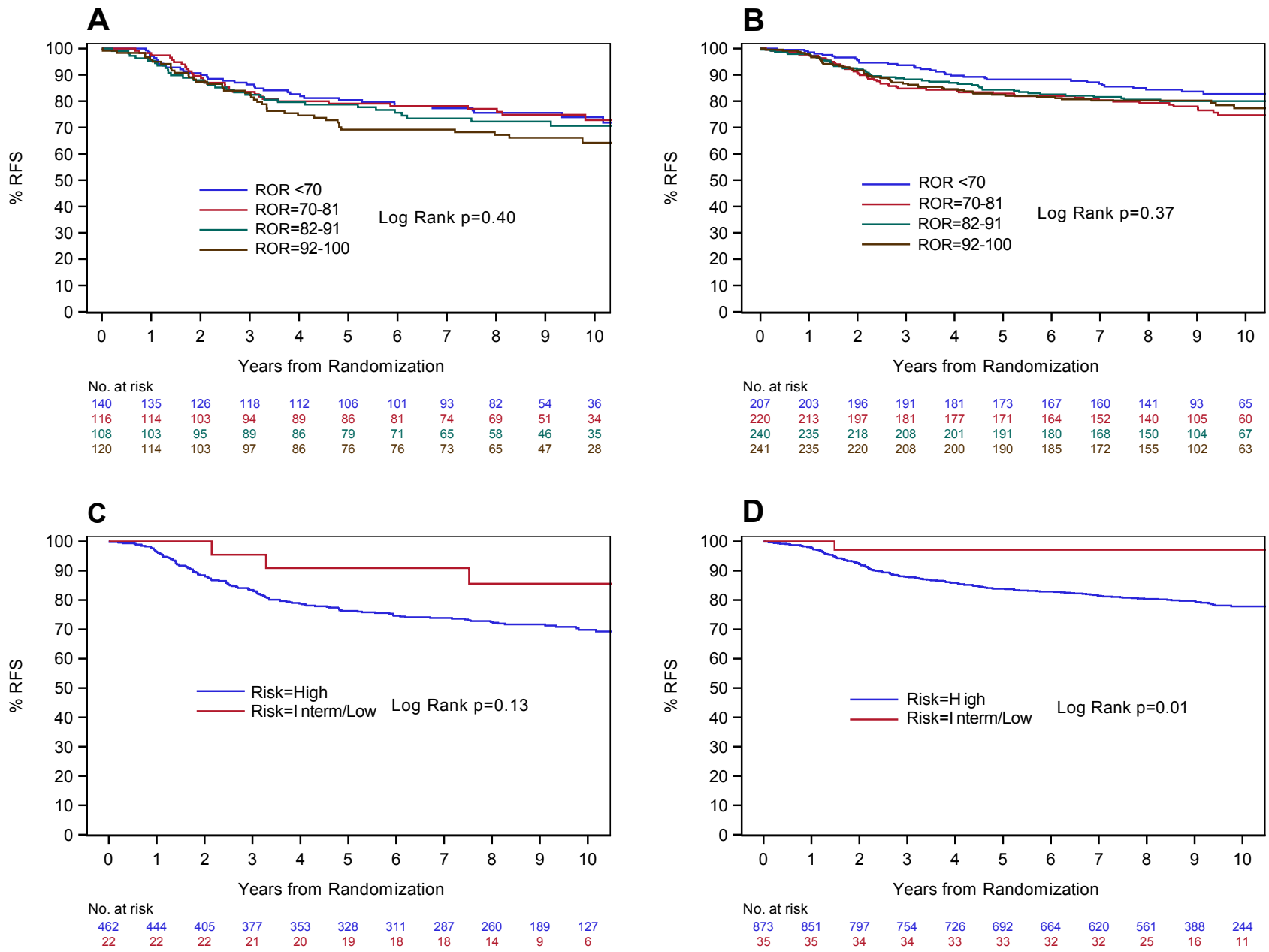
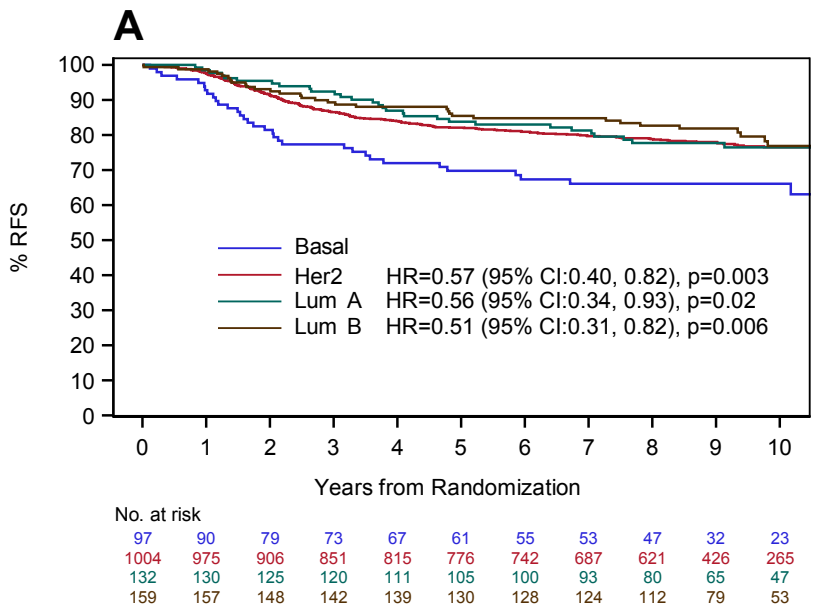
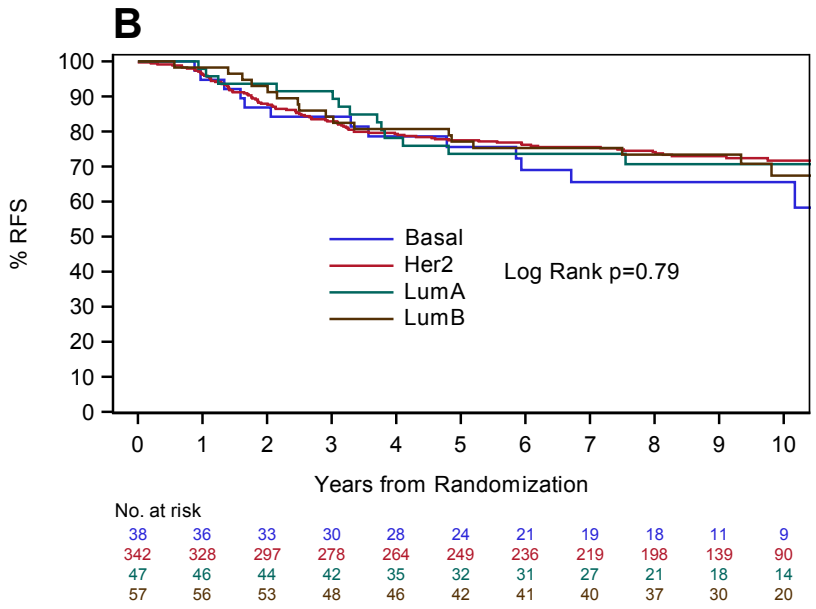

A
% RFS
ROR <70
ROR=70-81
ROR=82-91
ROR=92-100
Log Rank p=0.40
Years from Randomization
No. at risk
140 135 126 118 112 106 101 93 82 54 36
116 114 103 94 89 86 81 74 69 51 34
108 103 95 89 86 79 71 65 58 46 35
120 114 103 97 86 76 76 73 65 47 28
B
% RFS
ROR <70
ROR=70-81
ROR=82-91
ROR=92-100
Log Rank p=0.37
Years from Randomization
No. at risk
207 203 196 191 181 173 167 160 141 93 65
220 213 197 181 177 171 164 152 140 105 60
240 235 218 208 201 191 180 168 150 104 67
241 235 220 208 200 190 185 172 155 102 63
C
% RFS
Risk=High
Risk=Interm/Low
Log Rank p=0.13
Years from Randomization
No. at risk
462 444 405 377 353 328 311 287 260 189 127
22 22 22 21 20 19 18 18 14 9 6
D
% RFS
Risk=High
Risk=Interm/Low
Log Rank p=0.01
Years from Randomization
No. at risk
873 851 797 754 726 692 664 620 561 388 244
35 35 34 34 33 33 32 32 25 16 11

Figure 2


A
% RFS
Basal
Her2
Lum A
Lum B
HR=0.57 (95% CI:0.40, 0.82), p=0.003
HR=0.56 (95% CI:0.34, 0.93), p=0.02
HR=0.51 (95% CI:0.31, 0.82), p=0.006
Years from Randomization
No. at risk
97 90 79 73 67 61 55 53 47 32 23
1004 975 906 851 815 776 742 687 621 426 265
132 130 125 120 111 105 100 93 80 65 47
159 157 148 142 139 130 128 124 112 79 53


B
% RFS
Basal
Her2
LumA
LumB
Log Rank p=0.79
Years from Randomization
No. at risk
38 36 33 30 28 24 21 19 18 11 9
342 328 297 278 264 249 236 219 198 139 90
47 46 44 42 35 32 31 27 21 18 14
57 56 53 48 46 42 41 40 37 30 20

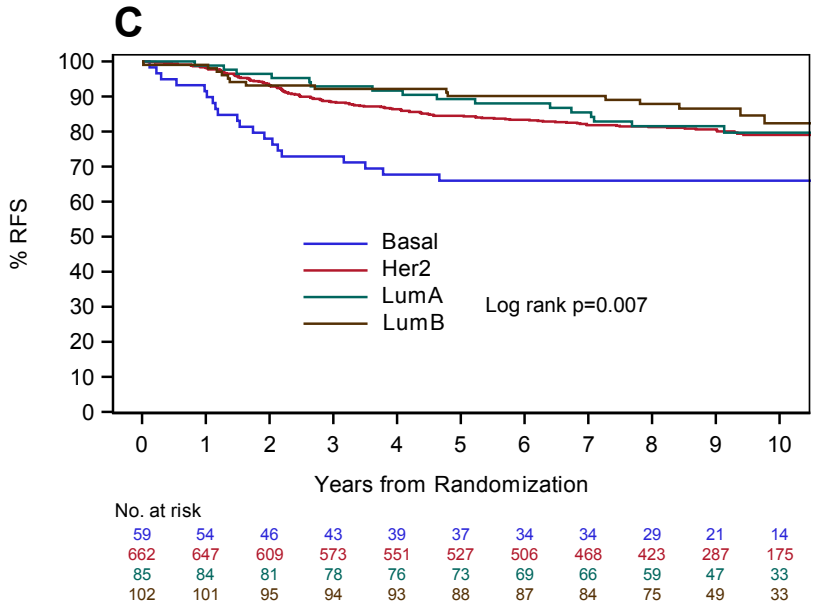
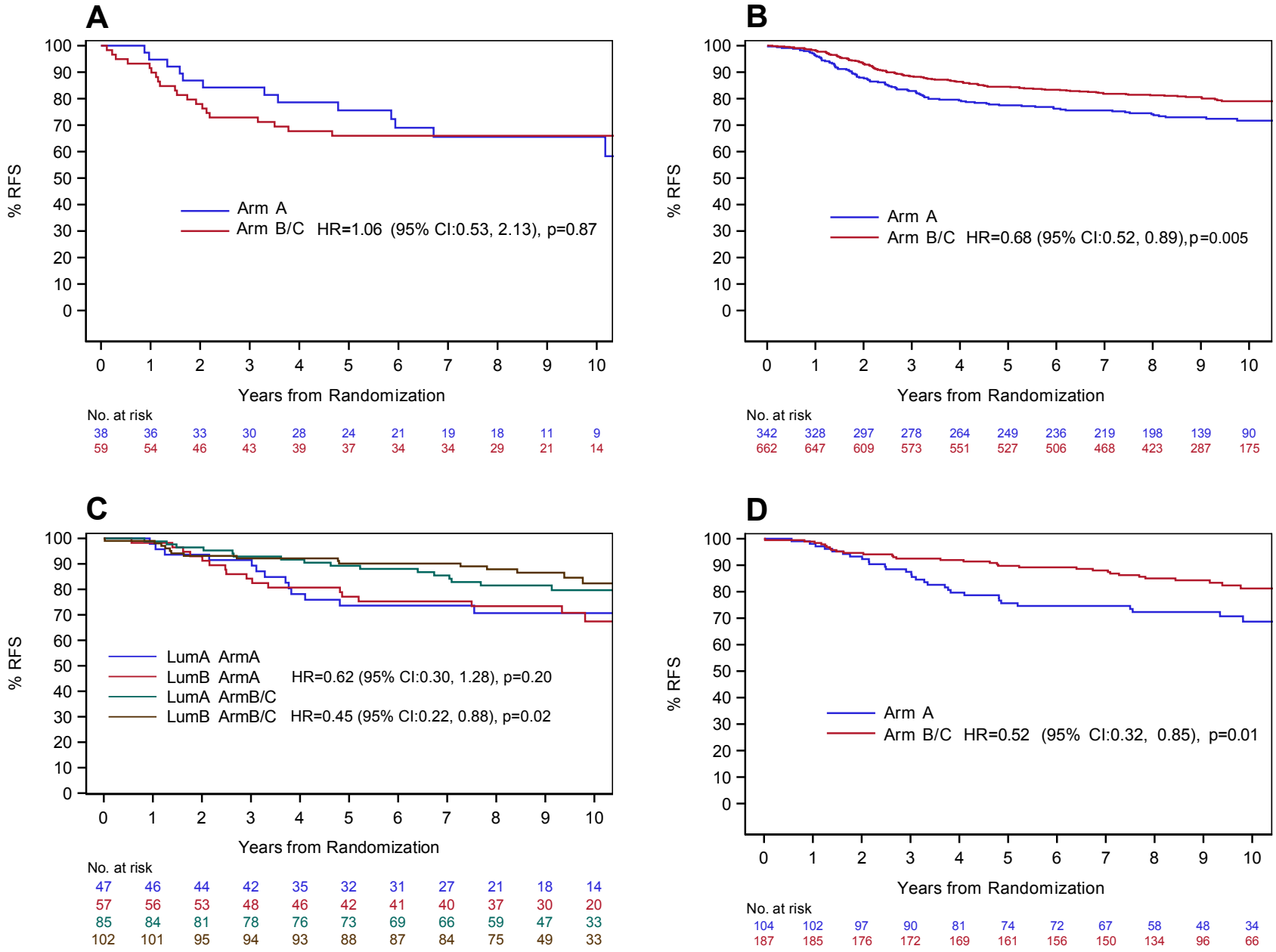

C
% RFS
Basal
Her2
LumA
LumB
Log rank p=0.007
Years from Randomization
No. at risk
59 54 46 43 39 37 34 34 29 21 14
662 647 609 573 551 527 506 468 423 287 175
85 84 81 78 76 73 69 66 59 47 33
102 101 95 94 93 88 87 84 75 49 33

Figure 3


A
% RFS
Arm A
Arm B/C HR=1.06 (95% CI:0.53, 2.13), p=0.87
Years from Randomization
No. at risk
38 36 33 30 28 24 21 19 18 11 9
59 54 46 43 39 37 34 34 29 21 14
B
% RFS
Arm A
Arm B/C HR=0.68 (95% CI:0.52, 0.89),p=0.005
Years from Randomization
No. at risk
342 328 297 278 264 249 236 219 198 139 90
662 647 609 573 551 527 506 468 423 287 175
C
% RFS
LumA ArmA
LumB ArmA HR=0.62 (95% CI:0.30, 1.28), p=0.20
LumA ArmB/C
LumB ArmB/C HR=0.45 (95% CI:0.22, 0.88), p=0.02
Years from Randomization
No. at risk
47 46 44 42 35 32 31 27 21 18 14
57 56 53 48 46 42 41 40 37 30 20
85 84 81 78 76 73 69 66 59 47 33
102 101 95 94 93 88 87 84 75 49 33
D
% RFS
Arm A
Arm B/C HR=0.52 (95% CI:0.32, 0.85), p=0.01
Years from Randomization
No. at risk
104 102 97 90 81 74 72 67 58 48 34
187 185 176 172 169 161 156 150 134 96 66