

Survival Outcomes are Associated with Genomic Instability in Luminal Breast Cancers: Supplementary Material

Lydia King, Andrew Flaus and Aaron Golden

Contents

1	Supplementary Results	3
1.1	Luminal A Cases	3
1.2	Luminal B Cases	4
2	Supplementary Figures	4
2.1	All Luminal Cases	4
2.2	Luminal A Cases	6
2.3	Luminal B Cases	9
2.4	Luminal A Cases by Grade	11
2.4.1	Grade 1	11
2.4.2	Grade 2	13
2.4.3	Grade 3	14
2.5	Luminal B Cases by Grade	16
2.5.1	Grade 1	16
2.5.2	Grade 2	17
2.5.3	Grade 3	19
2.6	Tishchenko Quantiles	20
3	Supplementary Tables	21
3.1	Clinical Variables	21
3.2	Tishchenko Genes	22
3.3	All Luminal Cases	22
3.4	Luminal A Cases	25
3.5	Luminal B Cases	28
3.6	Luminal A Grade 3	30

List of Figures

1	CNA quartiles within PAM50 subtypes in METABRIC luminal cases	5
2	Assessing OS proportional hazard assumptions in METABRIC luminal cases	5
3	Assessing DSS proportional hazard assumptions in METABRIC luminal cases	6
4	Density plot of CNA score distribution in METABRIC lumA cases	6
5	OS according to CNA score quartiles in METABRIC lumA cases.	7
6	DSS according to CNA score quartiles in METABRIC lumA cases.	7
7	Assessing OS proportional hazard assumptions in METABRIC lumA cases.	8
8	Assessing DSS proportional hazard assumptions in METABRIC lumA cases.	8
9	Density plot of CNA score distribution in METABRIC lumB cases	9
10	OS according to CNA score quartiles in METABRIC lumB cases.	9
11	DSS according to CNA score quartiles in METABRIC lumB cases.	10
12	Assessing OS proportional hazard assumptions in METABRIC lumB cases.	10

13	Assessing DSS proportional hazard assumptions in METABRIC lumB cases	11
14	Density plot of CNA score distribution in METABRIC lumA grade 1 cases.	11
15	OS according to CNA score quartiles in METABRIC lumA grade 1 cases.	12
16	DSS according to CNA score quartiles in METABRIC lumA grade 1 cases.	12
17	Density plot of CNA score distribution in METABRIC lumA grade 2 cases.	13
18	OS according to CNA score quartiles in METABRIC lumA grade 2 cases.	13
19	DSS according to CNA score quartiles in METABRIC lumA grade 2 cases	14
20	Density plot of CNA score distribution in METABRIC lumA grade 3 cases.	14
21	Assessing OS proportional hazard assumptions in METABRIC lumA grade 3 cases	15
22	Assessing DSS proportional hazard assumptions in METABRIC lumA grade 3 cases	15
23	Density plot of CNA score distribution in METABRIC lumB grade 1 cases.	16
24	OS according to CNA score quartiles in METABRIC lumB grade 1 cases.	16
25	DSS according to CNA score quartiles in METABRIC lumB grade 1 cases.	17
26	Density plot of CNA score distribution in METABRIC lumB grade 2 cases.	17
27	OS according to CNA score quartiles in METABRIC lumB grade 2 cases.	18
28	DSS according to CNA score quartiles in METABRIC lumB grade 2 cases.	18
29	Density plot of CNA score distribution in METABRIC lumB grade 3 cases	19
30	OS according to CNA score quartiles in METABRIC lumB grade 3 cases.	19
31	DSS according to CNA score quartiles in METABRIC lumB grade 3 cases.	20
32	Relationship of lumA grade 3 breast cancers in Q4 based on quantiles of Tishchenko et al. (2016).	20

List of Tables

1	List and description of the clinical variables recorded within the METABRIC cohort	21
2	List of top ten most up-regulated genes in the Tishchenko et al. [1] study.	22
3	Univariate Cox regression on clinical variables and CNA quartiles associated with METABRIC luminal cases.	22
4	Assessment of association between CNA quartiles and clinical variables that showed a significant association with either OS, DSS or both in METABRIC luminal cases.	24
5	Multivariate Cox regression on CNA quartiles and top three associated variables in METABRIC luminal cases.	24
6	Univariate Cox regression on clinical variables and CNA quartiles associated with METABRIC lumA cases.	25
7	Assessment of association between CNA quartiles and clinical variables that showed a significant association with either OS, DSS or both in METABRIC lumA cases.	26
8	Multivariate Cox regression on CNA quartiles and top four associated variables in METABRIC lumA cases.	27
9	Multivariate Cox regression on CNA quartiles and top three associated variables in METABRIC lumA cases.	27
10	Multivariate Cox regression on CNA quartiles and top two associated variables in METABRIC lumA cases.	27
11	Multivariate Cox regression on CNA quartiles and top associated variable in METABRIC lumA cases.	28
12	Univariate Cox regression on clinical variables and CNA quartiles associated with METABRIC lumB cases.	28
13	Assessment of association between CNA quartiles and clinical variables that showed a significant association with either OS, DSS or both in METABRIC lumB cases.	29
14	Multivariate Cox regression on CNA quartiles and top four associated variables in METABRIC lumB cases.	30

15	Multivariate Cox regression on CNA quartiles and top three associated variables in METABRIC lumB cases.	30
16	Univariate Cox regression on clinical variables and CNA quartiles associated with METABRIC lumA grade 3 cases.	30
17	Assessment of association between CNA quartiles and clinical variables that showed a significant association with either OS, DSS or both in METABRIC lumA grade 3 cases.	32
18	Multivariate Cox regression on CNA quartiles and associated clinical variables in METABRIC lumA grade 3 cases.	32
19	DSS Multivariate Cox regression on CNA quartiles in METABRIC lumA grade 3 cases. . . .	32

1 Supplementary Results

1.1 Luminal A Cases

For luminal A (lumA) cases the patients were again split into CNA quartiles based on absolute CNA score (Supplementary Figure 4). The CNA quartiles are significantly associated with both OS and DSS (Supplementary Figures 5-6). LumA patients within quartile 4 (Q4) have higher CNA scores indicative of higher levels of genomic instability and significantly worse survival outcomes than patients in the lower quartiles with less genomic instability (Supplementary Figures 5-6, both p -values < 0.05).

A number of steps were taken to determine whether the association between survival outcomes and CNA quartiles in the lumA METABRIC cohort was the result of confounding variables, which are additional factors influencing survival outcomes that are correlated with CNA quartiles.

First, a survival analysis using Kaplan-Meier (KM) plots and univariate Cox models was carried out to identify whether any of 23 clinical variables (Supplementary Table 1) and CNA quartiles in the lumA data were associated with survival outcome - 18 of the former and the CNA quartiles were subsequently shown to be associated with either OS or DSS or both (Supplementary Table 6) and were taken forward for examination using statistical tests. The χ^2 test was used to test the association between two categorical variables with sufficient cell sizes in the two-way table of categorical variables. Fisher's exact test was used in the case where any cell size was sufficiently small. The non-parametric Kruskal-Wallis test was used to determine if there were statistically significant differences between CNA quartiles and continuous clinical variables associated with survival outcomes. These tests indicated that the CNA quartiles are significantly associated with a number of clinical variables (Supplementary Table 7). Therefore, the observed association between survival outcomes and CNA quartiles in lumA cases could be the result of confounding variables.

Nottingham prognostic index (NPI), three gene classification, integrative cluster and grade were the most significant confounding factors (Supplementary Table 7). Multivariate Cox regression models were used to examine the survival association and assumptions of these factors. Several reduced multivariate Cox regression models were produced (Supplementary Tables 8-11). The results from the final multivariate Cox regression model indicated that NPI is strongly significantly associated with OS and DSS. Quartiles also appear to be associated with survival outcome, with patient survival outcomes decreasing as CNA quartile increases. Overall p -value suggests that the final model is statistically significant (Supplementary Table 11). The assumptions of the final multivariate Cox regression model for OS and DSS were tested and showed that the test was not statistically significant for any of the covariates, and the global test is also not statistically significant (Supplementary Figures 7-8). Therefore, we can assume proportional hazards and conclude that NPI significantly affects survival outcome and cannot be adjusted for in the model. This result confirms that the association between CNA quartiles and survival outcomes in lumA patients is a consequence of a confounding variable and therefore further stratification is required.

1.2 Luminal B Cases

For luminal B (lumB) cases the patients were also split into CNA quartiles based on absolute CNA score (Supplementary Figure 9). The CNA quartiles are significantly associated with both OS and DSS (Supplementary Figures 10-11).

A number of steps were taken to determine whether the association between survival outcomes and CNA quartiles in the lumB METABRIC cohort was the result of confounding variables, which are additional factors influencing survival outcomes that are correlated with CNA quartiles.

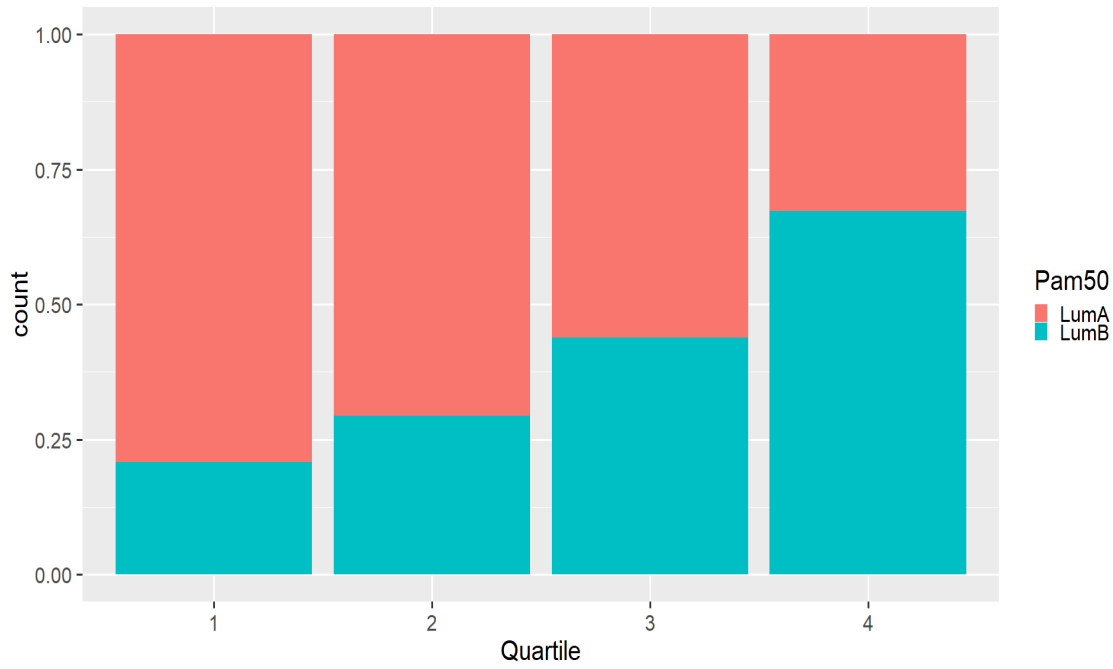
First, a survival analysis using Kaplan-Meier (KM) plots and univariate Cox models was carried out to identify whether any of 23 clinical variables (Supplementary Table 1) and CNA quartiles in the lumB data were associated with survival outcome - 12 of the former and the CNA quartiles were subsequently shown to be associated with either OS or DSS or both (Supplementary Table 12) and were taken forward for examination using statistical tests. The χ^2 test was used to test the association between two categorical variables with sufficient cell sizes in the two-way table of categorical variables. Fisher's exact test was used in the case where any cell size was sufficiently small. The non-parametric Kruskal-Wallis test was used to determine if there were statistically significant differences between CNA quartiles and continuous clinical variables associated with survival outcomes. These tests indicated that the CNA quartiles are significantly associated with a number of clinical variables (Supplementary Table 13). Therefore, the observed association between survival outcome and CNA quartiles in lumB cases could be the result of confounding variables.

NPI, three gene classification, integrative cluster and human epidermal growth factor receptor 2 (HER2) status were the most significant confounding factors (Supplementary Table 13). Multivariate Cox regression models were used to examine the survival association and assumptions of these factors. Several multivariate Cox regression models were produced (Supplementary Tables 14-15). The results from the final multivariate Cox regression model suggested that CNA quartile, NPI, three gene classification and integrative cluster were associated with survival outcomes (Supplementary Table 15). The overall p -value suggests that the model is statistically significant (Supplementary Table 15). The assumptions of the multivariate Cox regression model for OS and DSS were tested and showed that the test was statistically significant for some of the covariates, and the global test is also statistically significant for OS but not DSS (Supplementary Figure 12-13). Therefore, we assume proportional hazards for DSS but not OS. We conclude that NPI, three gene classification, integrative clusters and CNA quartiles significantly affect survival outcome. This result confirms that the association between CNA quartiles and survival outcome in lumB patients is a consequence of a confounding variable and therefore further stratification is required.

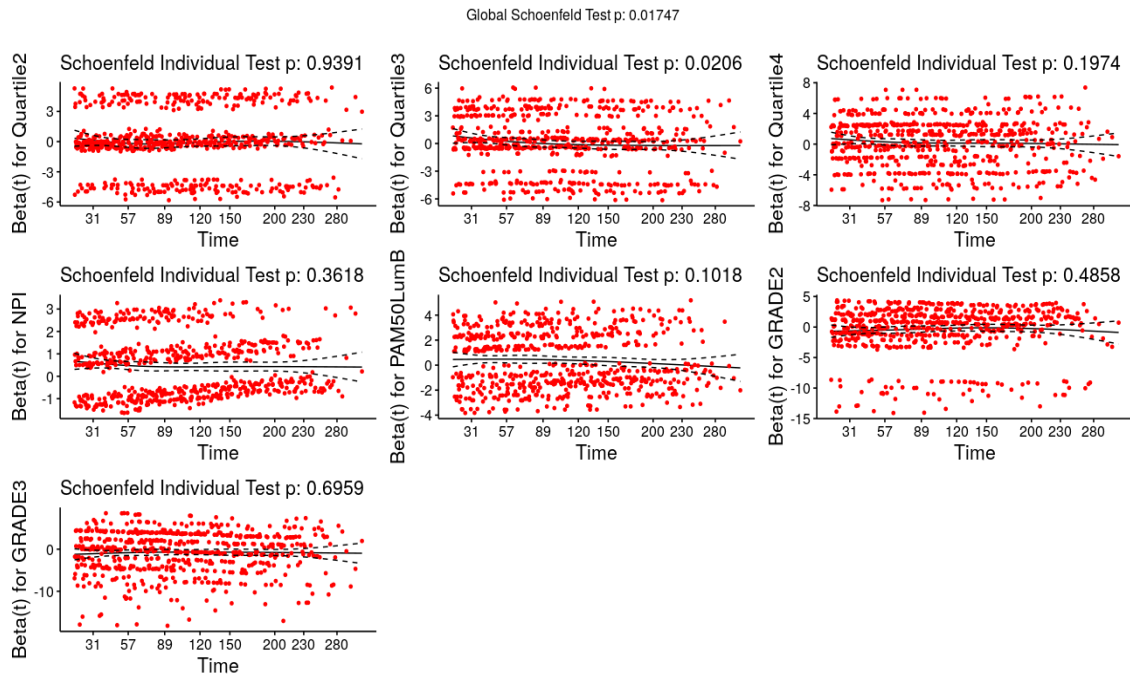
2 Supplementary Figures

2.1 All Luminal Cases

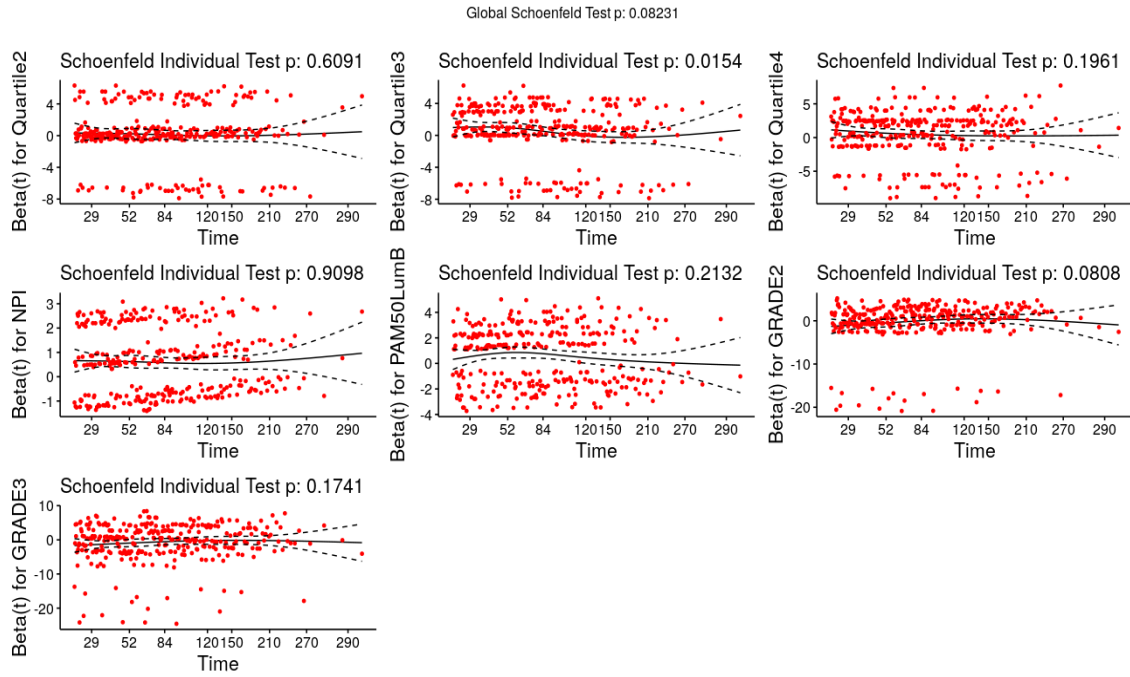
Note: For KM plots for OS and DSS according to CNA score quartiles in METABRIC luminal cases see Figure 3 in manuscript.



Supplementary Figure 1: CNA quartiles within PAM50 subtypes in METABRIC luminal cases. Shows the proportion of lumA and lumB cases present in each CNA quartile.

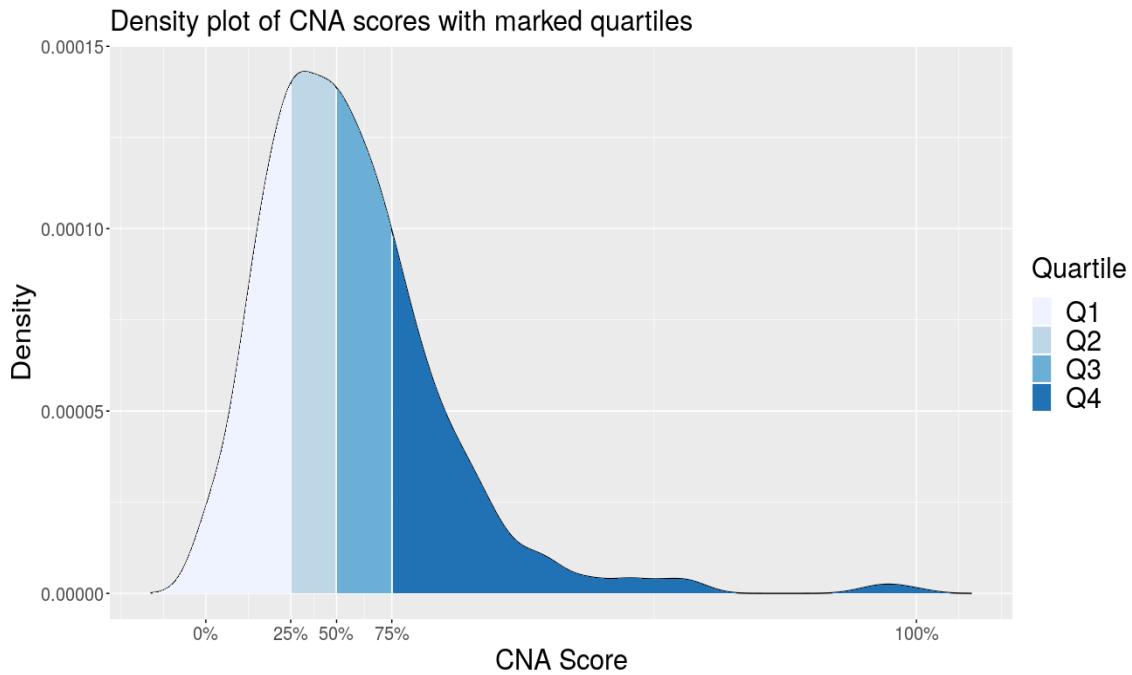


Supplementary Figure 2: Assessing OS proportional hazard assumptions in METABRIC luminal cases. The multivariate Cox model includes CNA quartiles, NPI, PAM50 subtype and grade. Carried out using graphs of the scaled Schoenfeld residuals against the transformed time.

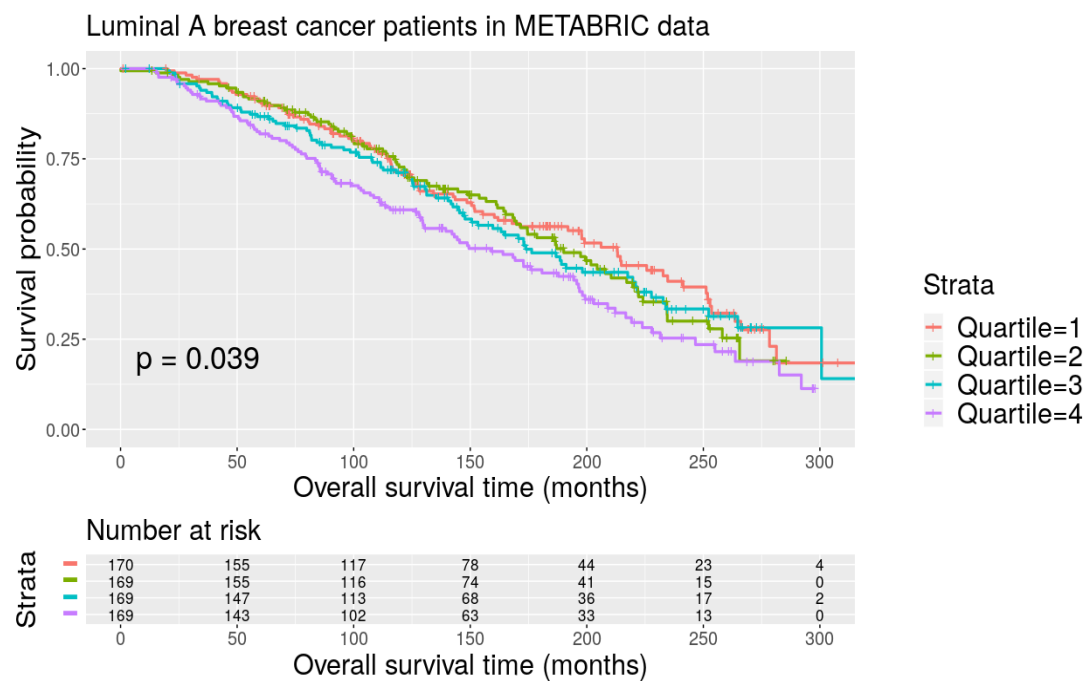


Supplementary Figure 3: Assessing DSS proportional hazard assumptions in METABRIC luminal cases. The multivariate Cox model includes CNA quartiles, NPI, PAM50 subtype and grade. Carried out using graphs of the scaled Schoenfeld residuals against the transformed time.

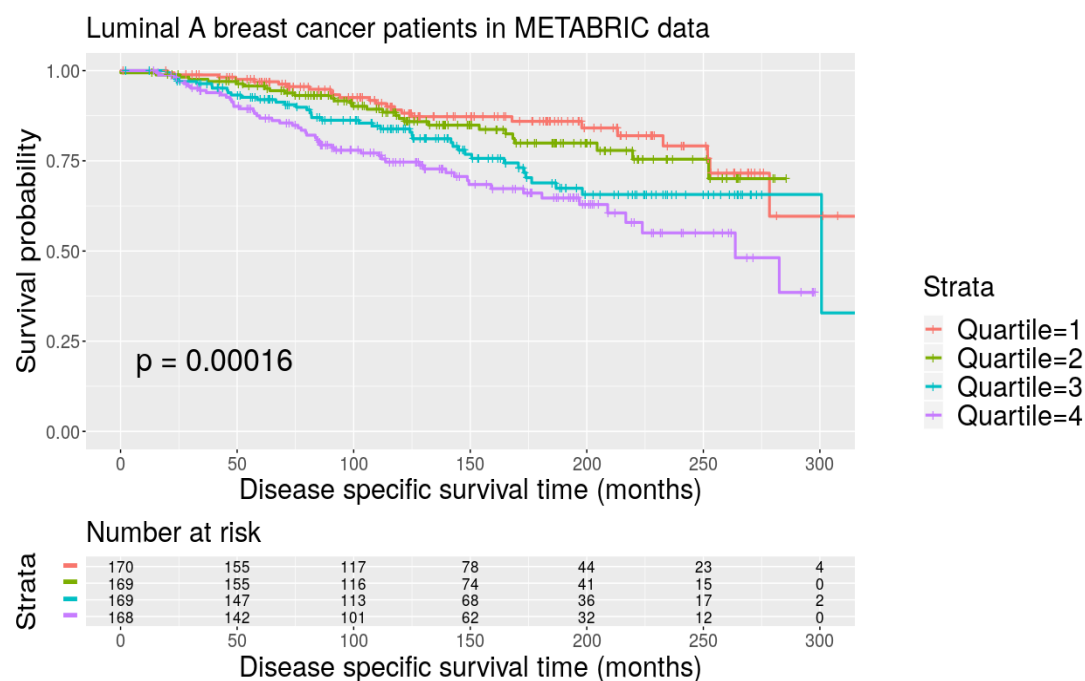
2.2 Luminal A Cases



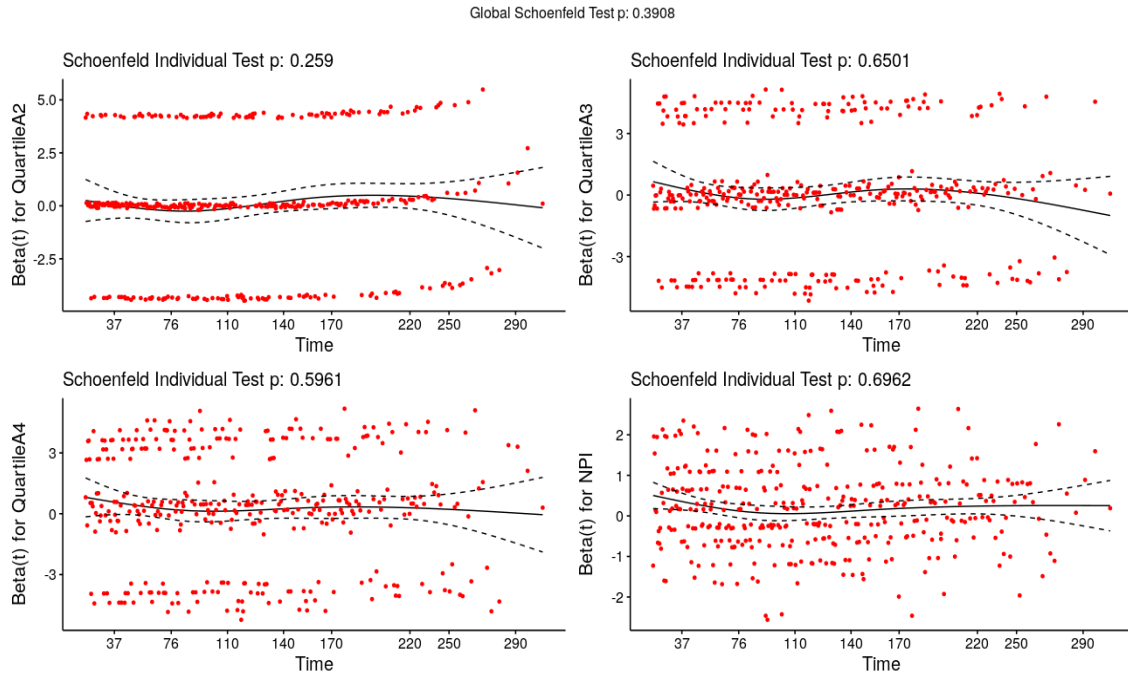
Supplementary Figure 4: Density plot of CNA score distribution in METABRIC lumA cases. The CNA quartiles are marked out based on differing shades of blue displayed in the legend.



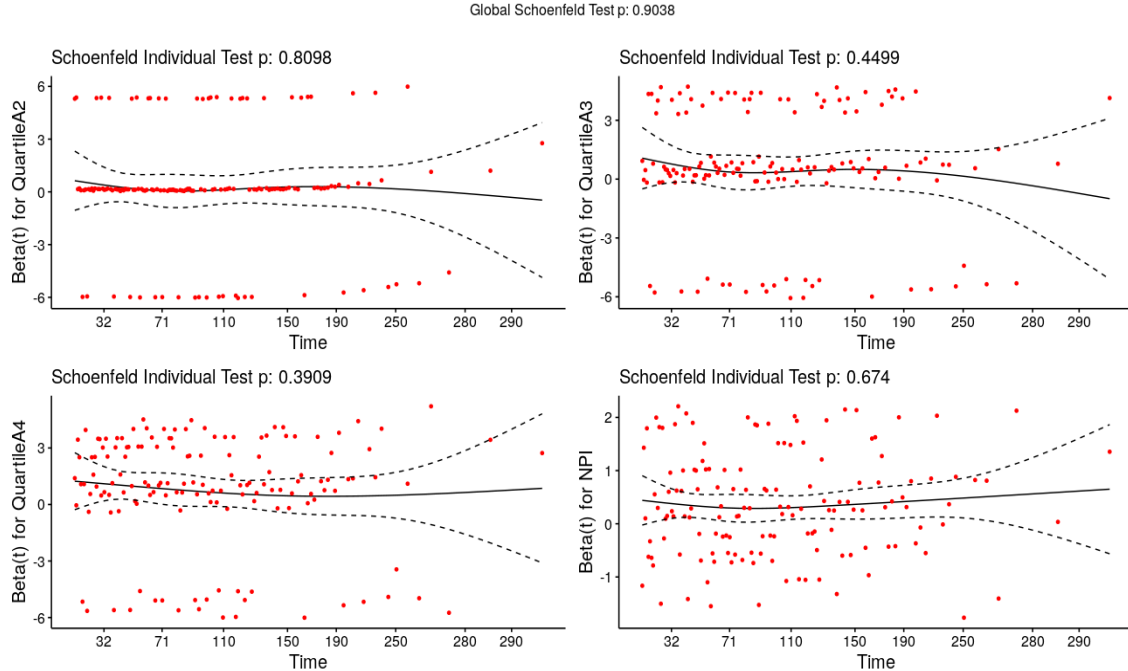
Supplementary Figure 5: KM plot for OS according to CNA score quartiles in METABRIC lumA cases. The p-value associated with the log-rank test and a risk table displaying the number of patients at risk at each time interval is displayed.



Supplementary Figure 6: KM plot for DSS according to CNA score quartiles in METABRIC lumA cases. The p-value associated with the log-rank test and a risk table displaying the number of patients at risk at each time interval is displayed.

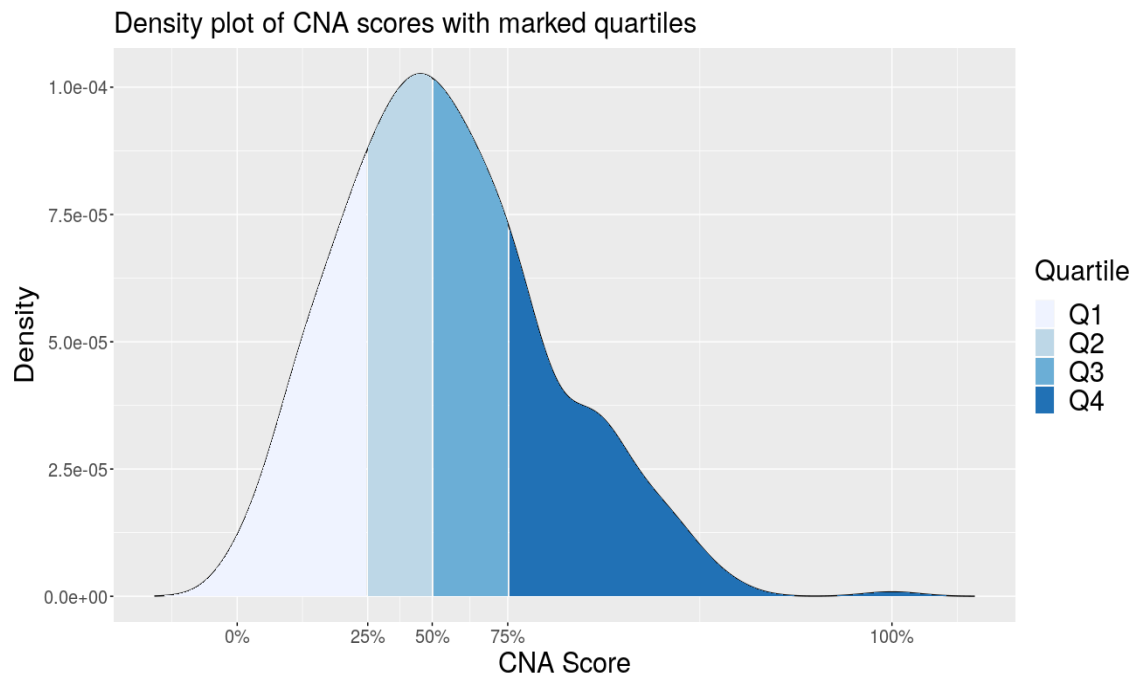


Supplementary Figure 7: Assessing OS proportional hazard assumptions of the multivariate Cox model including CNA quartiles and NPI in METABRIC lumA patients. Carried out using graphs of the scaled Schoenfeld residuals against the transformed time.

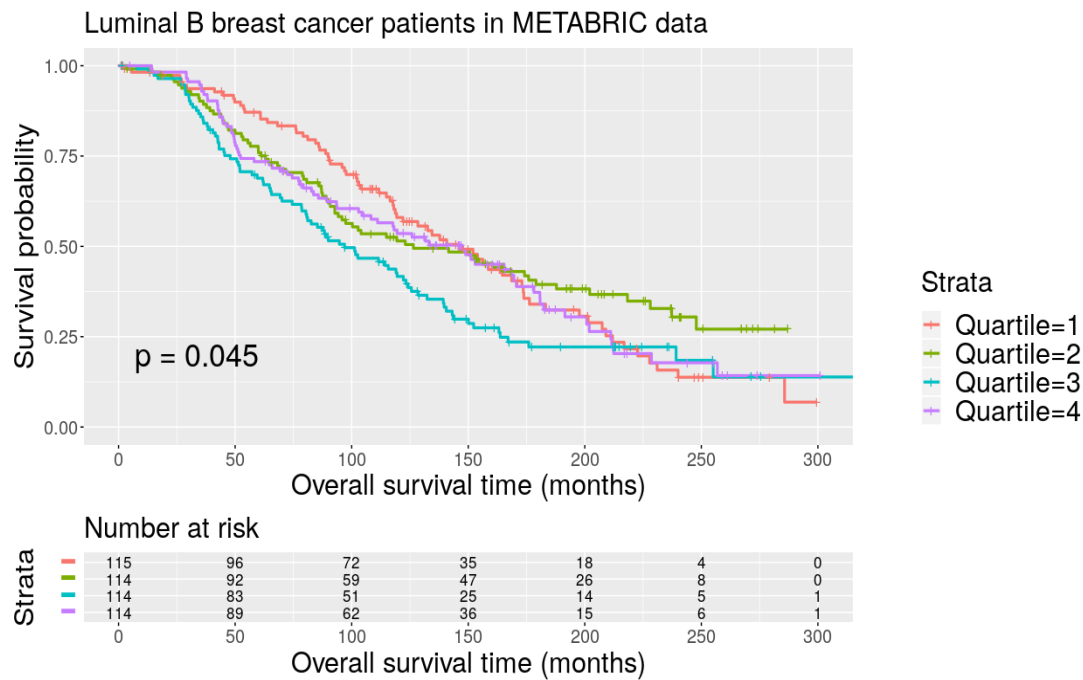


Supplementary Figure 8: Assessing DSS proportional hazard assumptions of the multivariate Cox model including CNA quartiles and NPI in METABRIC lumA patients. Carried out using graphs of the scaled Schoenfeld residuals against the transformed time.

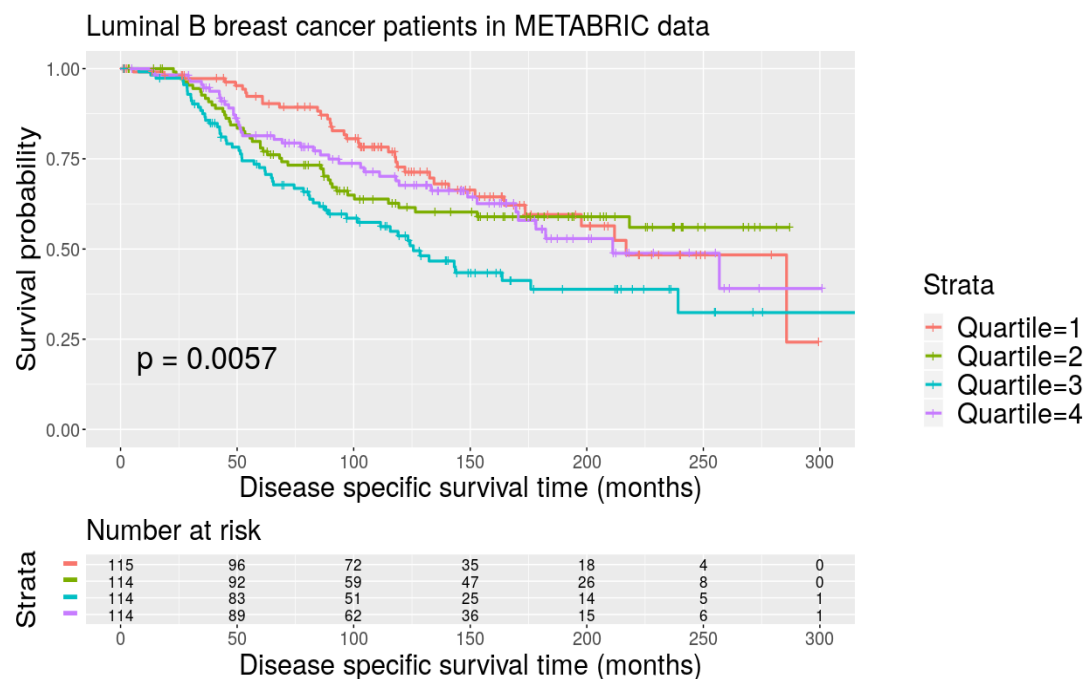
2.3 Luminal B Cases



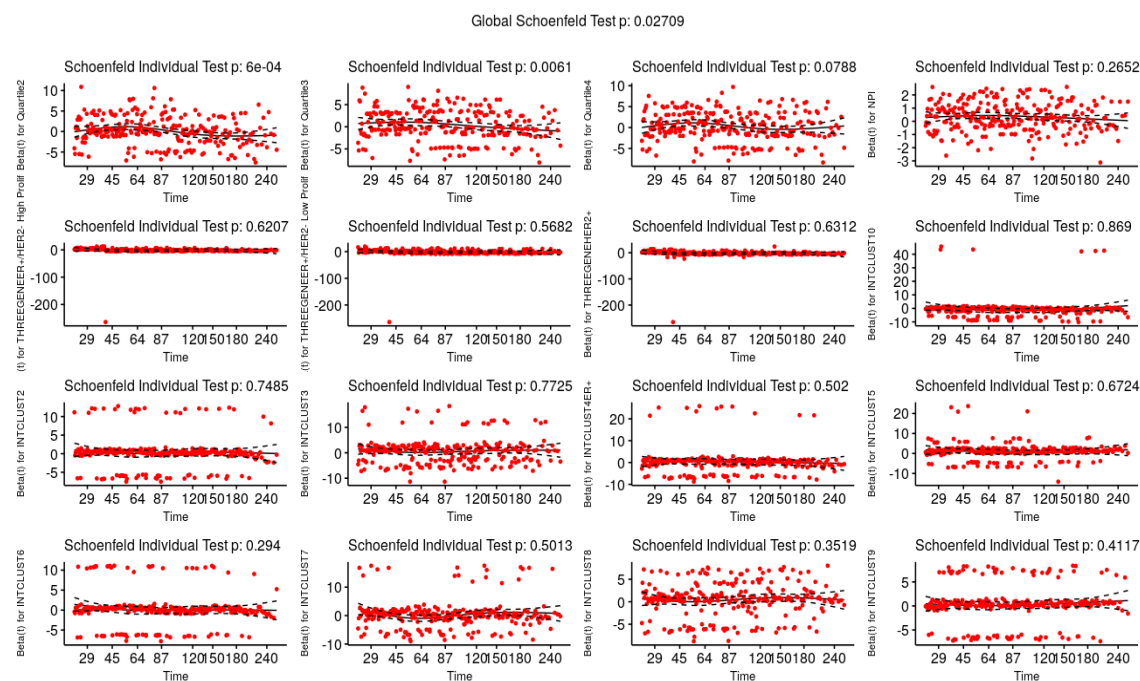
Supplementary Figure 9: Density plot of CNA score distribution in METABRIC lumB cases. The CNA quartiles are marked out based on differing shades of blue displayed in the legend.



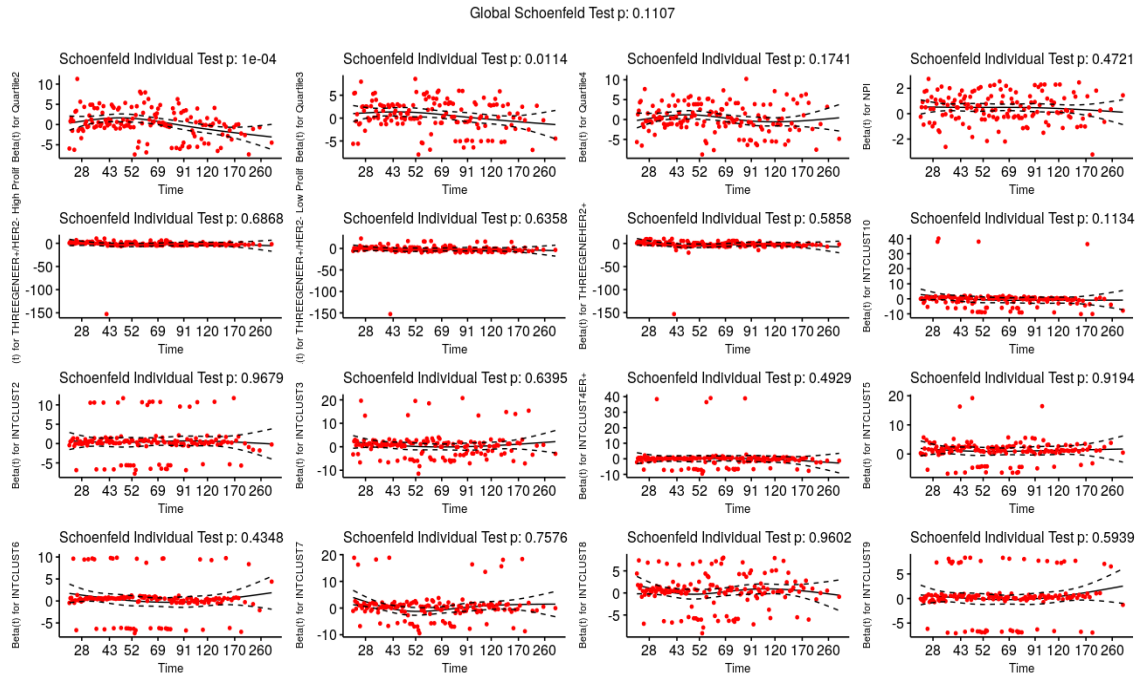
Supplementary Figure 10: KM plot for OS according to CNA score quartiles in METABRIC lumB cases. The p-value associated with the log-rank test and a risk table displaying the number of patients at risk at each time interval is displayed.



Supplementary Figure 11: KM plot for DSS according to CNA score quartiles in METABRIC lumB cases. The p-value associated with the log-rank test and a risk table displaying the number of patients at risk at each time interval is displayed.



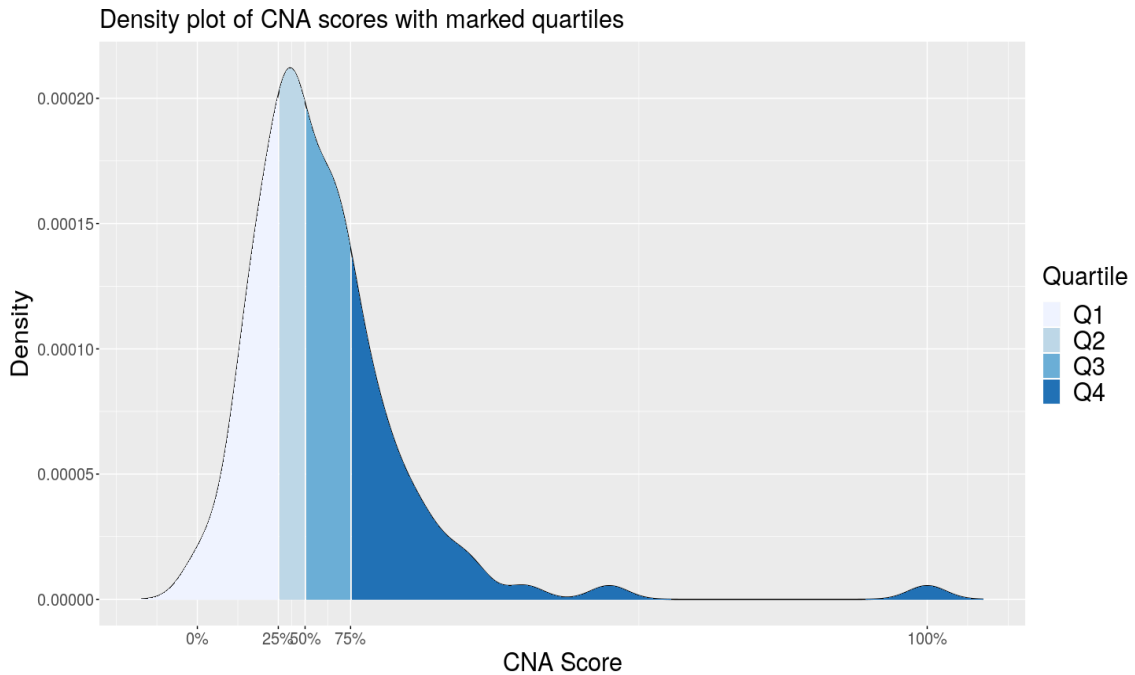
Supplementary Figure 12: Assessing OS proportional hazard assumptions of the multivariate Cox model including CNA quartiles, NPI, three gene classification and integrative cluster in METABRIC lumB patients. Carried out using graphs of the scaled Schoenfeld residuals against the transformed time.



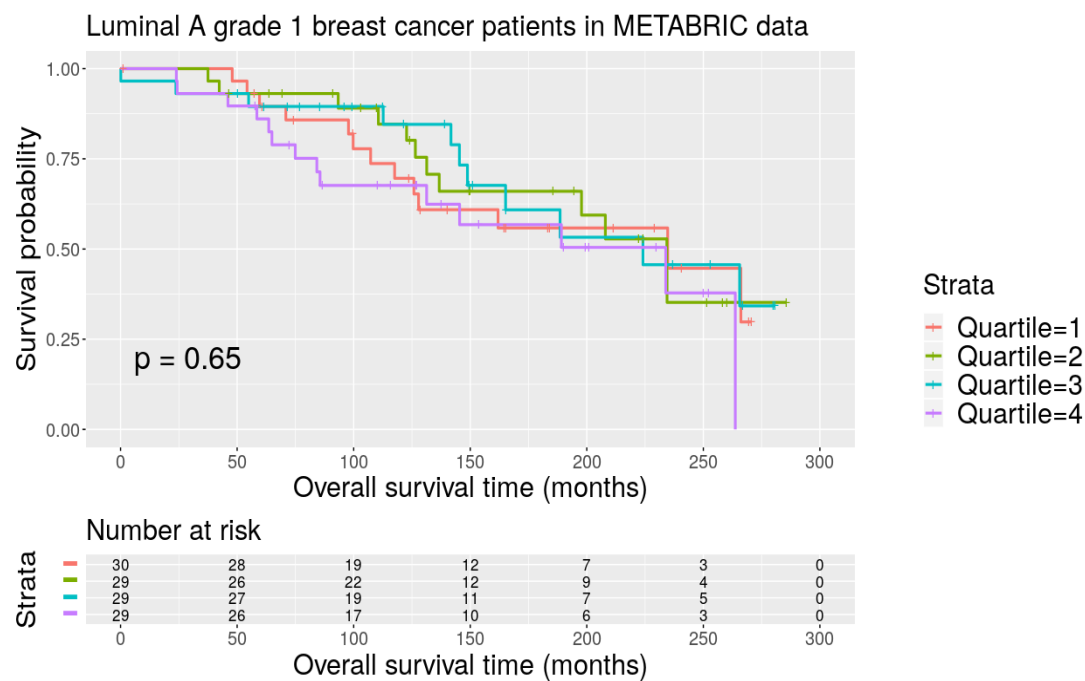
Supplementary Figure 13: Assessing DSS proportional hazard assumptions of the multivariate Cox model including CNA quartiles, NPI, three gene classification and integrative cluster in METABRIC lumB patients. Carried out using graphs of the scaled Schoenfeld residuals against the transformed time.

2.4 Luminal A Cases by Grade

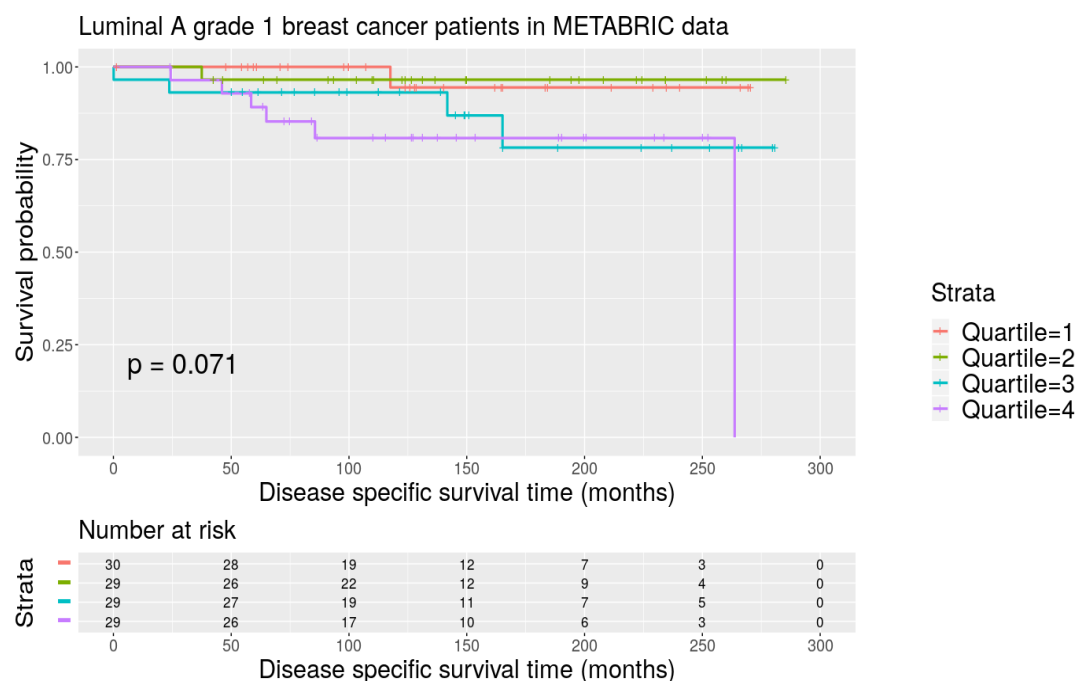
2.4.1 Grade 1



Supplementary Figure 14: Density plot of CNA score distribution in METABRIC lumA grade 1 cases. The quartiles are marked out based on differing shades of blue displayed in the legend.

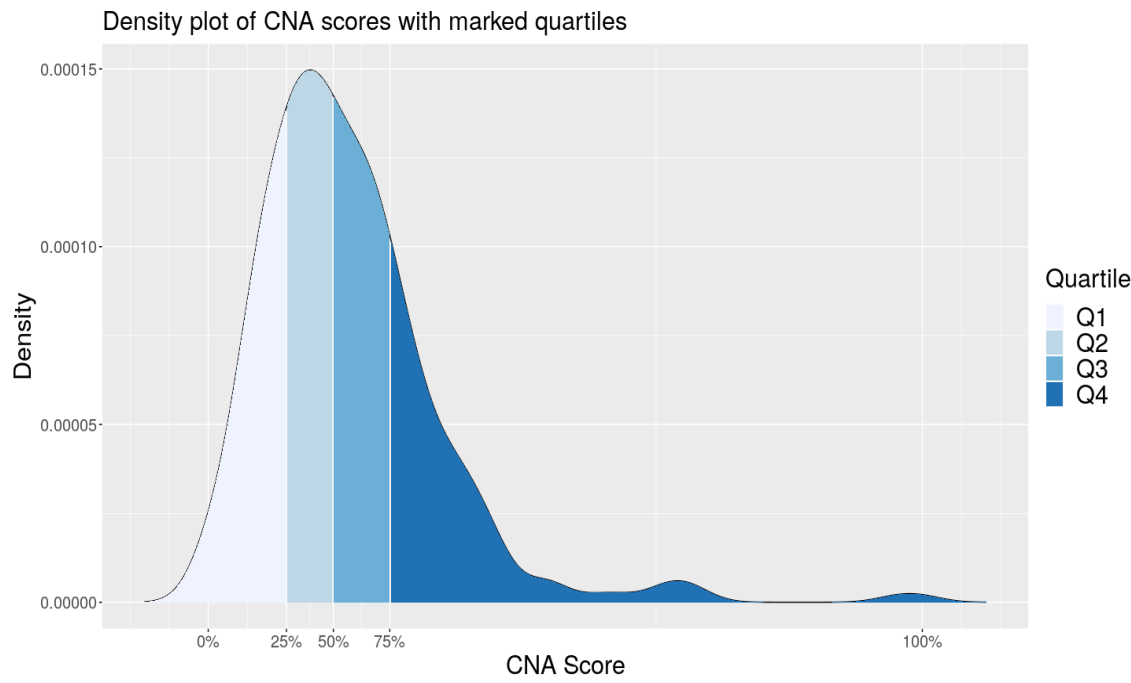


Supplementary Figure 15: KM plot for OS according to CNA score quartiles in METABRIC lumA grade 1 cases. The p-value associated with the log-rank test and a risk table displaying the number of patients at risk at each time interval is displayed.

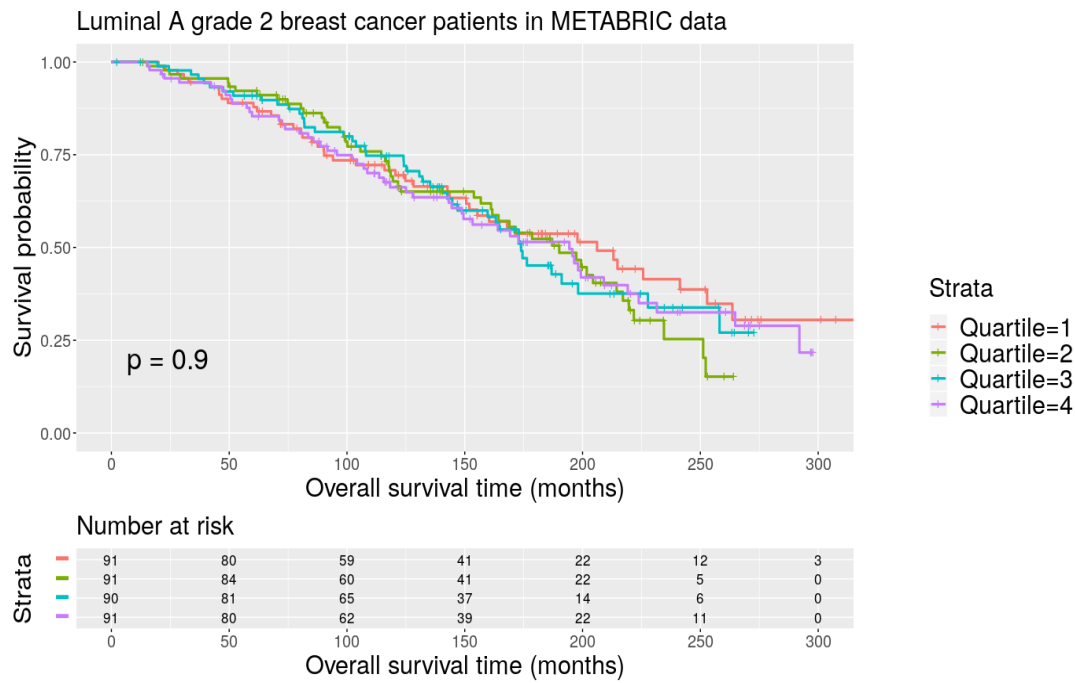


Supplementary Figure 16: KM plot for DSS according to CNA score quartiles in METABRIC lumA grade 1 cases. The p-value associated with the log-rank test and a risk table displaying the number of patients at risk at each time interval is displayed.

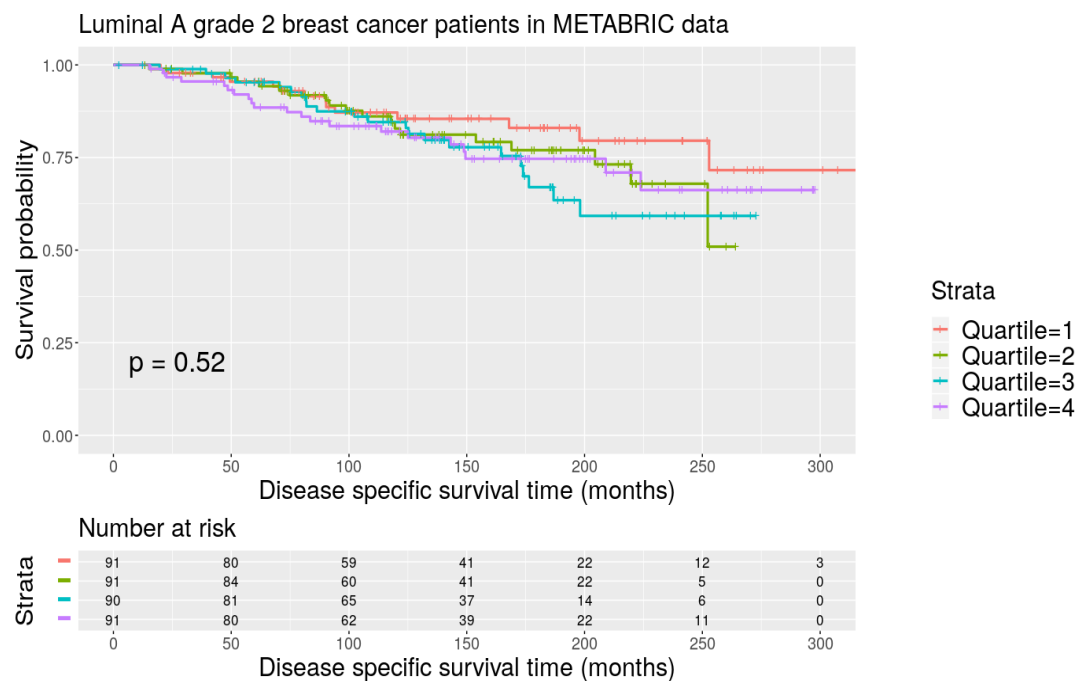
2.4.2 Grade 2



Supplementary Figure 17: Density plot of CNA score distribution in METABRIC lumA grade 2 cases. The CNA quartiles are marked out based on differing shades of blue displayed in the legend.

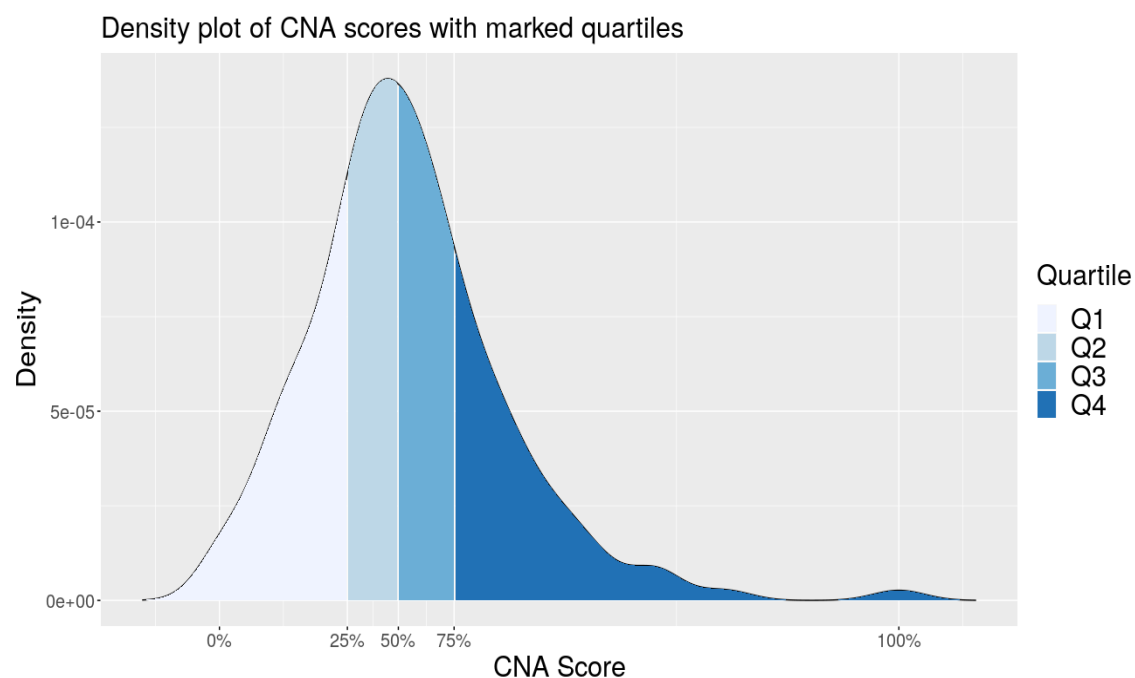


Supplementary Figure 18: KM plot for OS according to CNA score quartiles in METABRIC lumA grade 2 cases. The p-value associated with the log-rank test and a risk table displaying the number of patients at risk at each time interval is displayed.



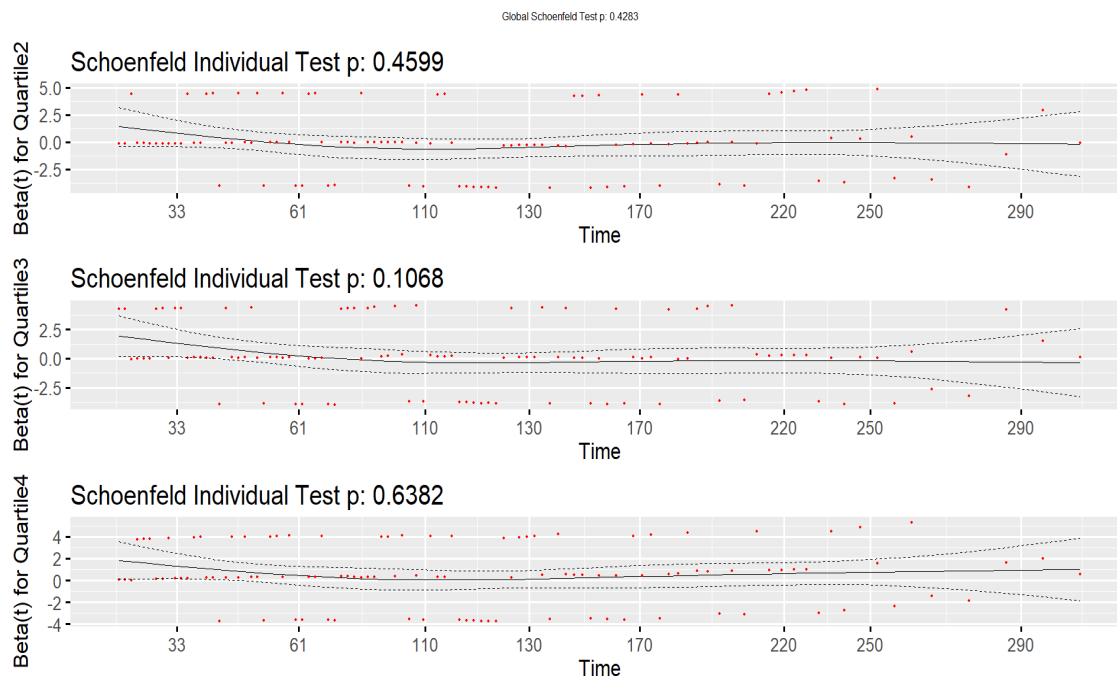
Supplementary Figure 19: KM plot for DSS according to CNA score quartiles in METABRIC lumA grade 2 cases. The p-value associated with the log-rank test and a risk table displaying the number of patients at risk at each time interval is displayed.

2.4.3 Grade 3

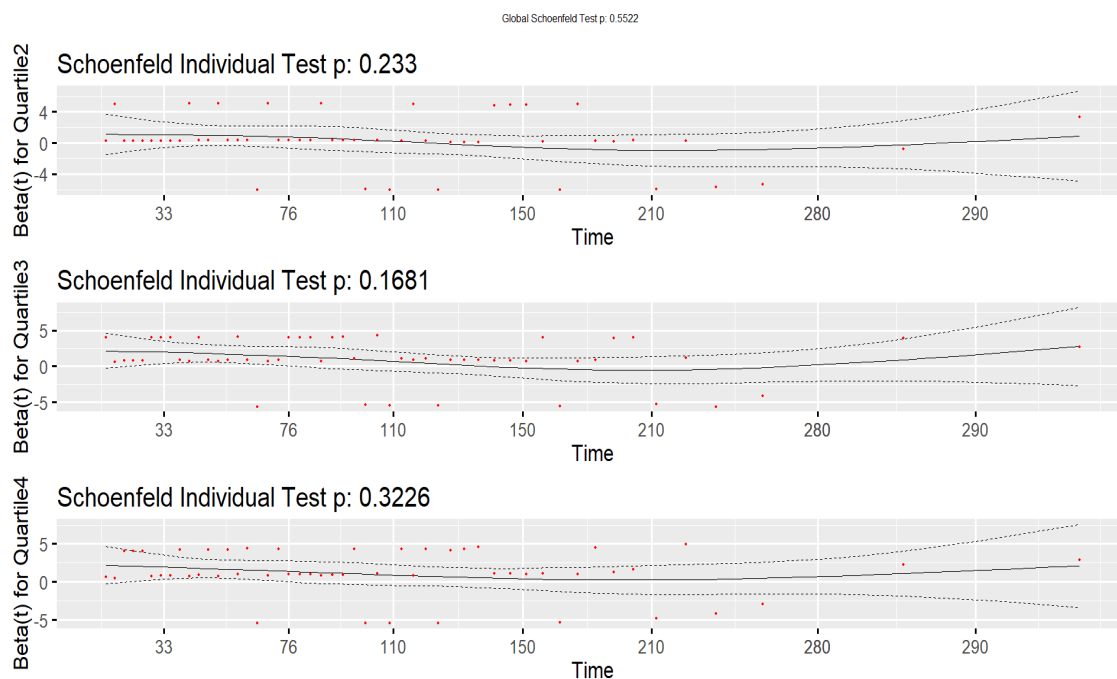


Supplementary Figure 20: Density plot of CNA score distribution in METABRIC lumA grade 3 cases. The CNA quartiles are marked out based on differing shades of blue displayed in the legend.

Note: For KM plots for OS and DSS according to CNA score quartiles in METABRIC luminal A grade 3 cases see Figure 4 in manuscript.



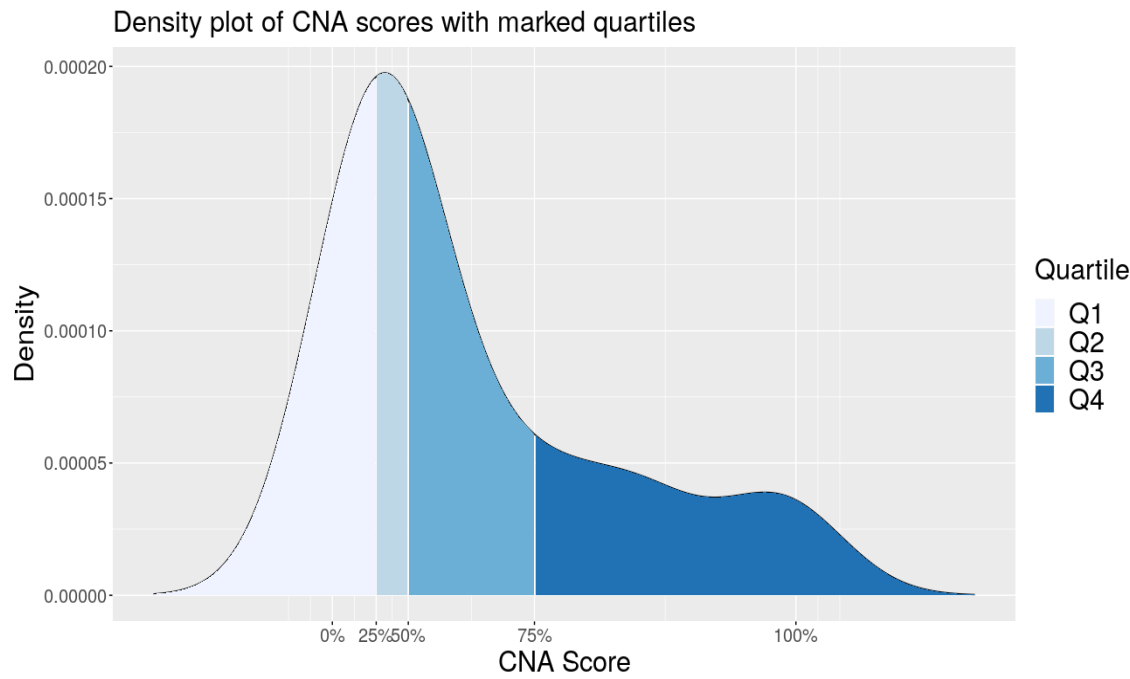
Supplementary Figure 21: Assessing OS proportional hazard assumptions of the multivariate Cox model including CNA quartiles in METABRIC lumA grade 3 cases. Carried out using graphs of the scaled Schoenfeld residuals against the transformed time.



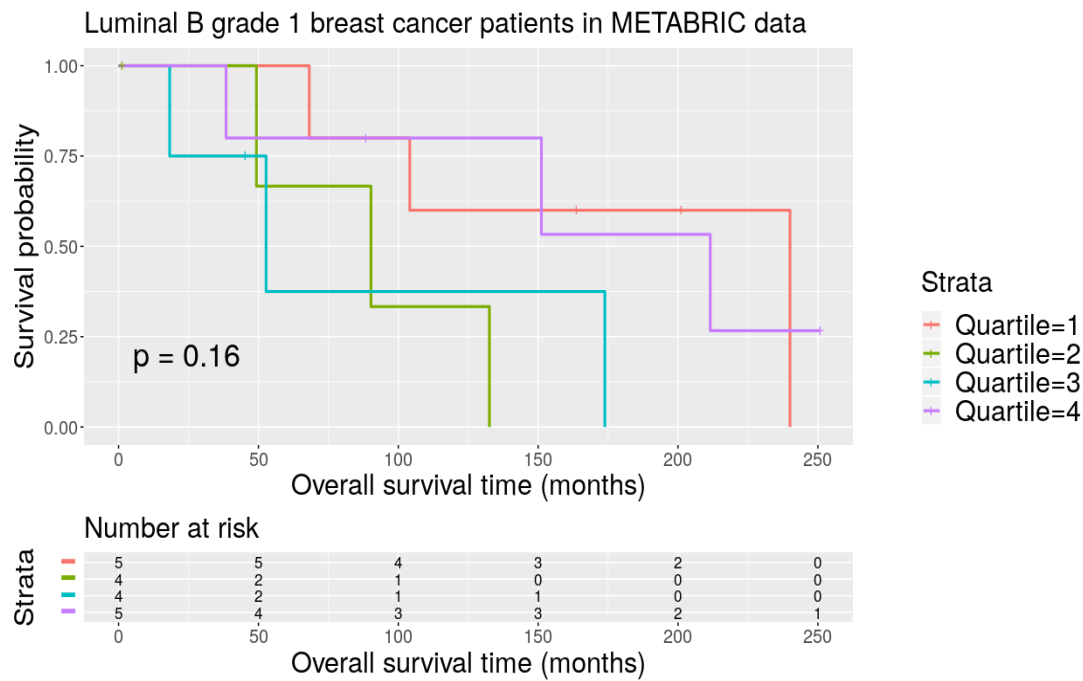
Supplementary Figure 22: Assessing DSS proportional hazard assumptions of the multivariate Cox model including CNA quartiles in METABRIC lumA grade 3 cases. Carried out using graphs of the scaled Schoenfeld residuals against the transformed time.

2.5 Luminal B Cases by Grade

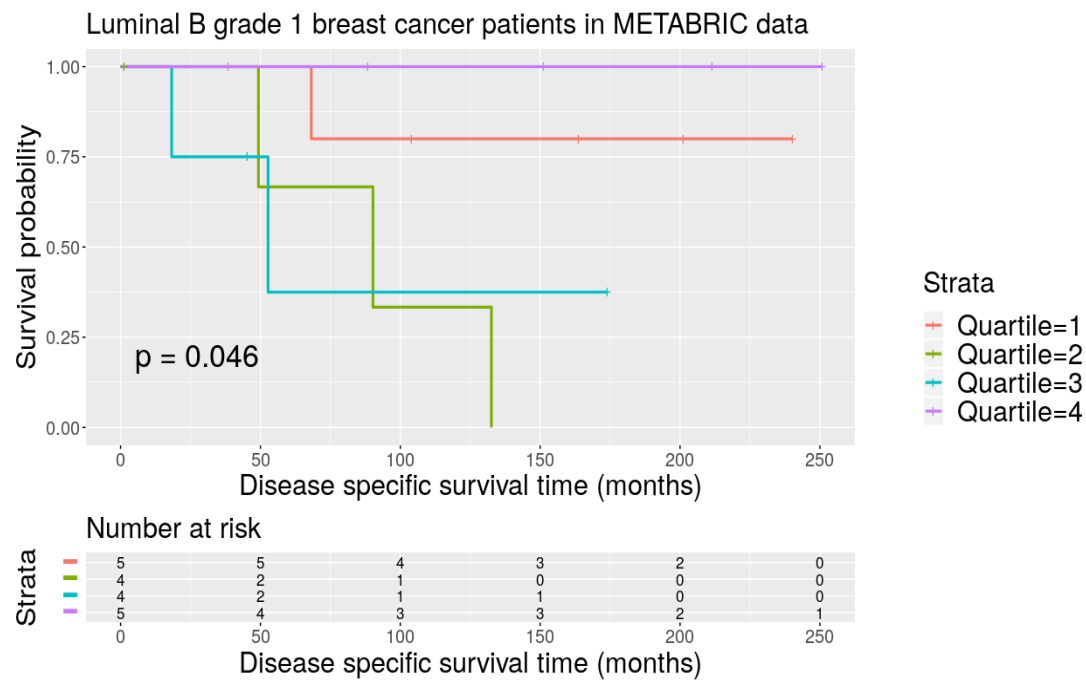
2.5.1 Grade 1



Supplementary Figure 23: Density plot of CNA score distribution in METABRIC lumB grade 1 cases. The CNA quartiles are marked out based on differing shades of blue displayed in the legend.

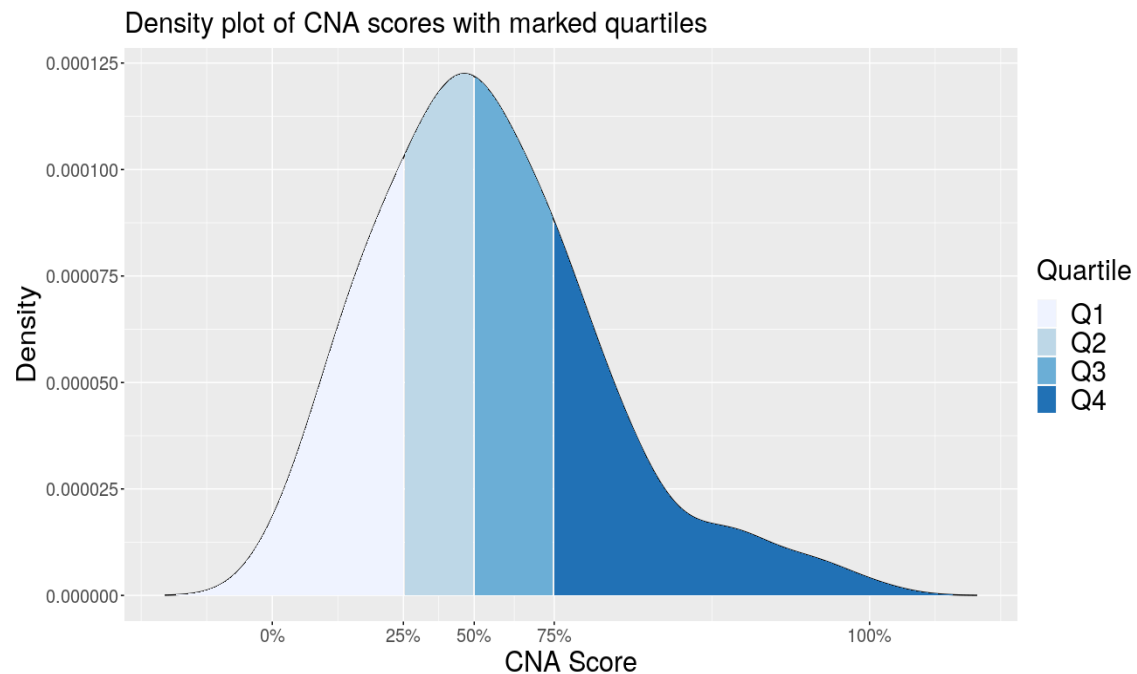


Supplementary Figure 24: KM plot for OS according to CNA score quartiles in METABRIC lumB grade 1 cases. The p-value associated with the log-rank test and a risk table displaying the number of patients at risk at each time interval is displayed.

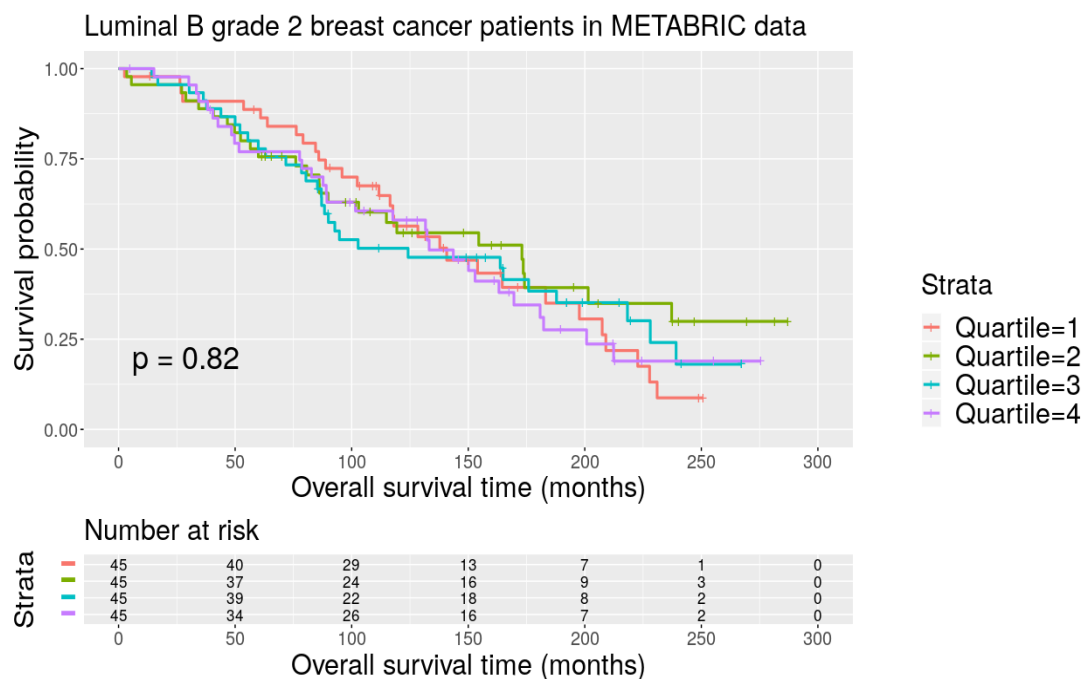


Supplementary Figure 25: KM plot for DSS according to CNA score quartiles in METABRIC lumB grade 1 cases. The p-value associated with the log-rank test and a risk table displaying the number of patients at risk at each time interval is displayed.

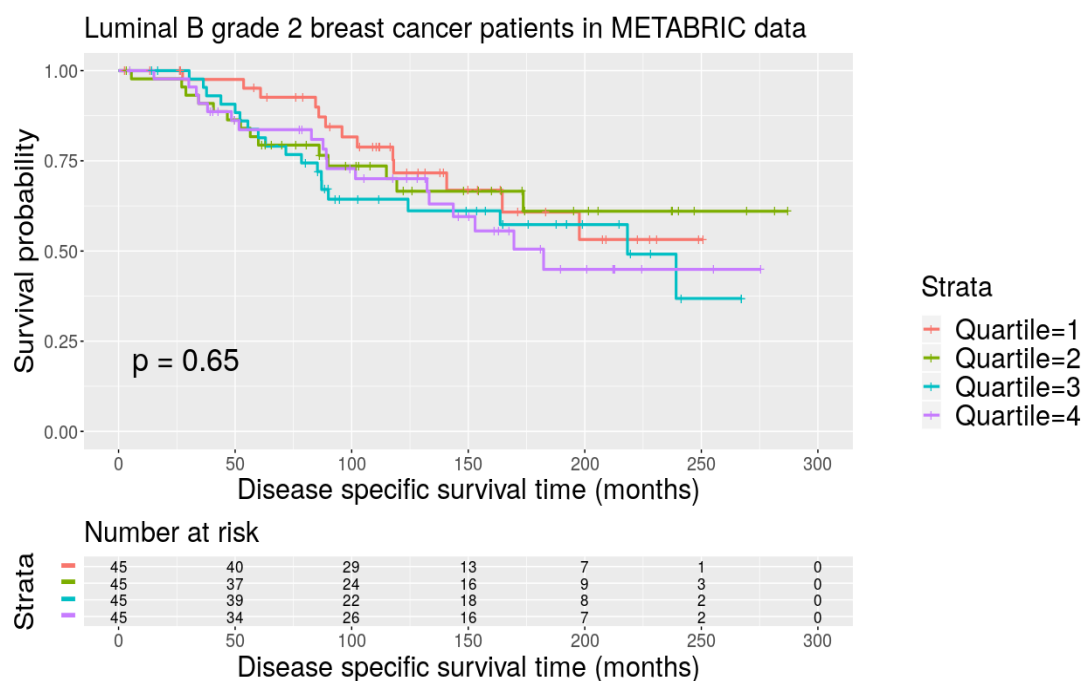
2.5.2 Grade 2



Supplementary Figure 26: Density plot of CNA score distribution in METABRIC lumB grade 2 cases. The CNA quartiles are marked out based on differing shades of blue displayed in the legend.

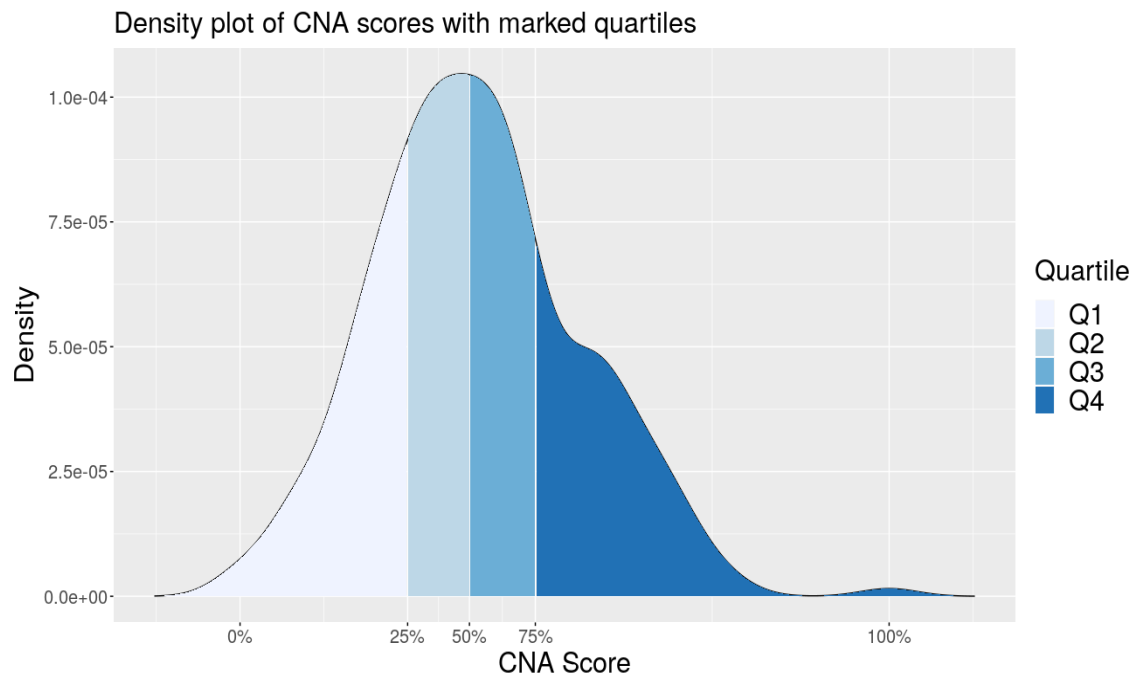


Supplementary Figure 27: KM plot for OS according to CNA score quartiles in METABRIC lumB grade 2 cases. The p-value associated with the log-rank test and a risk table displaying the number of patients at risk at each time interval is displayed.

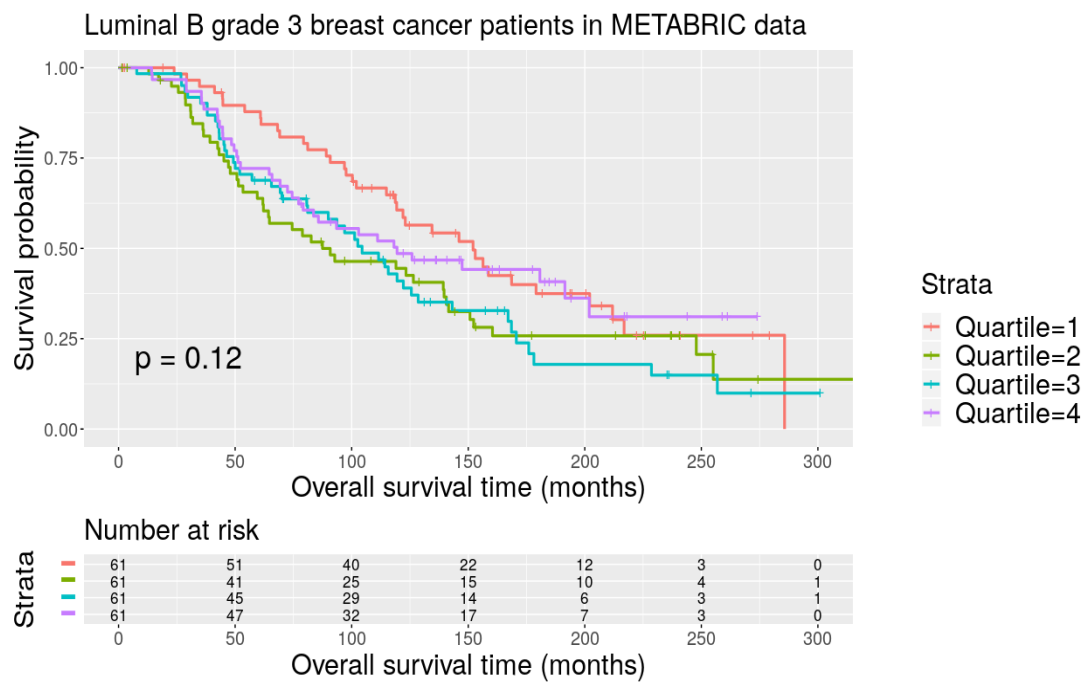


Supplementary Figure 28: KM plot for DSS according to CNA score quartiles in METABRIC lumB grade 2 cases. The p-value associated with the log-rank test and a risk table displaying the number of patients at risk at each time interval is displayed.

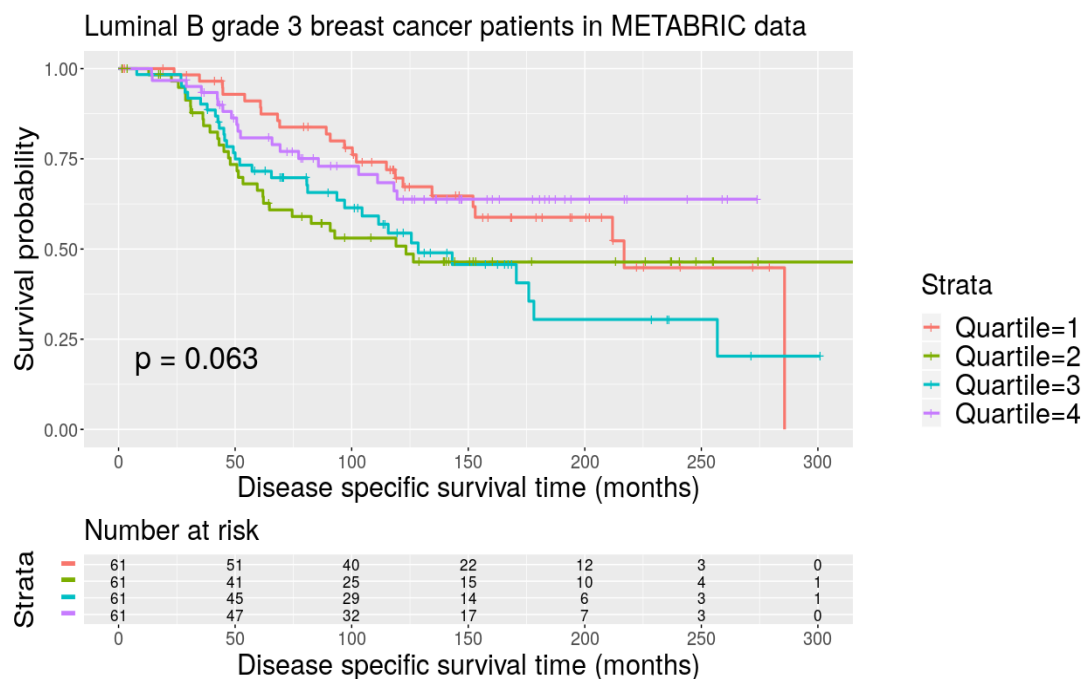
2.5.3 Grade 3



Supplementary Figure 29: Density plot of CNA score distribution in METABRIC lumB grade 3 cases. The CNA quartiles are marked out based on differing shades of blue displayed in the legend.

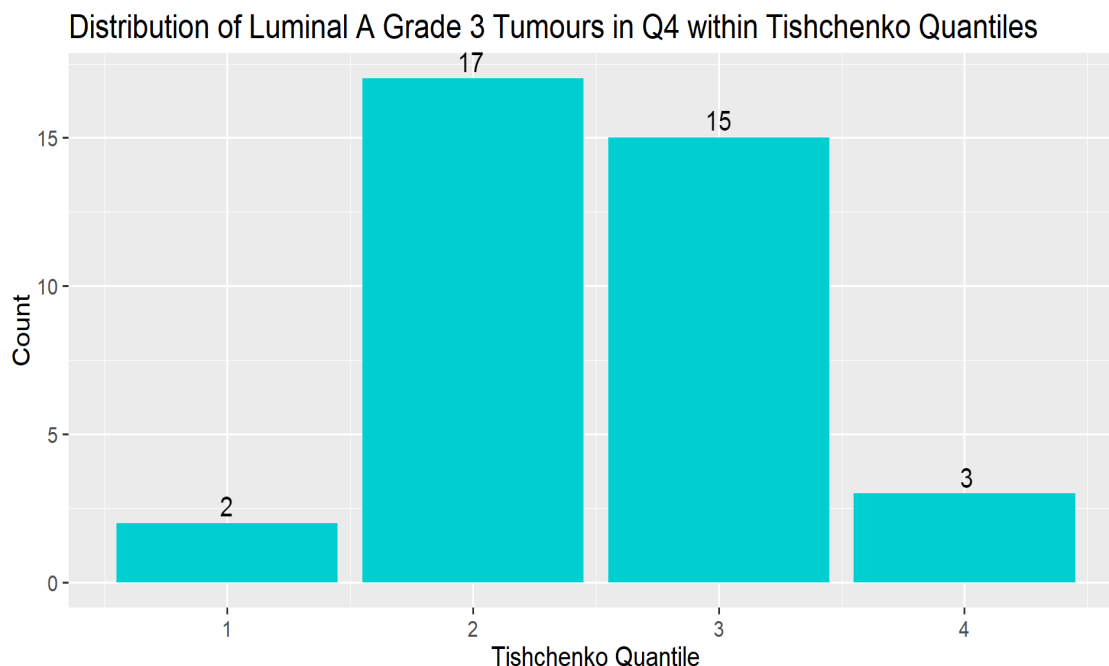


Supplementary Figure 30: KM plot for OS according to CNA score quartiles in METABRIC lumB grade 3 cases. The p-value associated with the log-rank test and a risk table displaying the number of patients at risk at each time interval is displayed.



Supplementary Figure 31: KM plot for DSS according to CNA score F quartiles in METABRIC lumB grade 3 cases. The p-value associated with the log-rank test and a risk table displaying the number of patients at risk at each time interval is displayed.

2.6 Tishchenko Quantiles



Supplementary Figure 32: Relationship of lumA grade 3 breast cancers in Q4 based on quantiles of Tishchenko et al. [1]. Barplot shows the number of lumA grade 3 patients corresponding to quantiles one to four.

3 Supplementary Tables

3.1 Clinical Variables

Supplementary Table 1: List and description of the clinical variables recorded within the METABRIC cohort

Clinical Variables	Description
Positive Lymph Nodes	Provides information on the number of positive lymph nodes a patient is found to have. Ranges from 0-45 in the METABRIC luminal cohort. Lymph node involvement is a prognostic factor in breast cancer, where the presence of positive lymph nodes acts as a predictor of increased risk of local and distant recurrence, which directly affects patient outcome [2].
Nottingham Prognostic Index	The Nottingham Prognostic Index (NPI) is a prognostication tool which aids in the management of breast cancers. It is a prognostic indicator that considers the size of the tumour, the number of lymph nodes involved, and tumour grade. It is a continuous score [3]. NPI ranges from 1-6.5 in the METABRIC luminal cohort.
Cellularity	Tumour cellularity is the relative proportion of tumour and normal cells in a sample. Tumour cellularity can affect the sensitivity of processes such as mutation detection and copy number analysis [4]. In the METABRIC study tumour cellularity was scored visually in a semi-quantitative fashion. Cellularity values were binned into low cellularity (<40% tumour DNA), moderate cellularity (40% - 70% tumour DNA), and high cellularity (>70% tumour DNA) [5].
ER Immunohistochemistry	The presence of estrogen receptors (ERs) as detected by immunohistochemistry (IHC). ERs are hormone receptors found on breast cells that detect hormone signals. This results in cell growth [6]. ER IHC is a weak prognostic marker of clinical outcome in breast cancer, but a strong predictive marker for response [7]. Breast tumours are categorised as either ER positive or ER negative and the METABRIC luminal cohort contained both positive and negative cases.
HER2 Status	Human epidermal growth factor 2 (HER2) is on chromosome 17q12. This gene is amplified in approximately 15-25% of breast cancers and can act as an important prognostic indicator [8]. HER2 status informs whether the HER2 gene is amplified in breast cancer [6]. METABRIC luminal patients were categorised as being HER2 negative (no amplification) or HER2 positive (HER2 amplified).
HER2 SNP6	HER2 status determined through the use of the Affymetrix SNP6 array [8]. Categorised as gain, loss, neutral or undefined in the METABRIC luminal cohort.
Inferred Menopausal State	Indicates whether women were pre or post-menopausal at diagnosis of breast cancer. The METABRIC luminal cohort contains both pre- and post-menopausal women.
PAM50 Subtype	The PAM50 classification is based on a 50-gene predictor (PAM50) which categorises breast tumours into lumA, lumB, HER2-enriched, and basal-like subtypes. The PAM50 classification can provide outcome prediction in both ER-positive and ER-negative tumors [9]. For the purpose of this study only luminal patients (lumA and lumB) within the METABRIC cohort will be analysed.
Three Gene Classification	Samples were classified into the four Gene Expression prognostic Index Using Subtypes (GENIUS) subgroups which included HER2+, ER-/HER2-, ER+/HER2- high proliferation and ER+/HER2- low proliferation. This was carried out by Curtis et al. [5] using the geneFu package in Bioconductor. The METABRIC luminal cohort contained all subgroups.
Laterality	This is the location of primary tumour. Primary tumour laterality was classified as left side, right side, or bilateral [10]. In the METABRIC luminal cohort patients presented with either left or right side laterality.
ER Status	ERs are hormone receptors found on breast cells which receive signals from estrogen which promote its growth [6]. Breast tumours are categorised as ER-positive (ER+) if they have receptors for the hormone estrogen and ER-negative (ER-) if they do not. The METABRIC luminal cohort contains both positive and negative cases.
Cancer Type Detailed	Indicates whether the breast cancers begin in the cells that line the ducts (ductal cancers), in the cells that line the lobules (lobular cancers) or in other tissues [11]. Within the METABRIC luminal cohort patients presented with Breast, Breast Invasive Ductal Carcinoma, Breast Invasive Lobular Carcinoma, Breast Invasive Mixed Mucinous Carcinoma and Breast Mixed Ductal and Lobular Carcinoma.
Age at Diagnosis	The age at which a patient is diagnosed with a disease or condition. In this case the age at which the patient was diagnosed with breast cancer. Within the METABRIC luminal cohort patients ranged from 25-95 years of age.
Tumour Grade	Tumour grade is based on the degree of differentiation of the tumour tissue. Can be a well-differentiated tumour (grade 1), a moderately differentiated tumour (grade 2) or a poorly differentiated tumour (grade 3) [12]. Patients within the METABRIC luminal cohort presented with tumour grades 1-3.
PR Status	Progesterone receptors (PRs) are hormone receptors found on breast cells that detect hormone signals. This results in cell growth. The cancer is PR-positive (PR+) if it has receptors for the hormone progesterone and PR-negative (PR-) if it does not [6]. Patients within the METABRIC luminal cohort are PR positive or negative.
Integrative Clusters	Dawson et al. [13] used molecular information on the genomic and transcriptomic landscapes of breast cancer to define 10 integrative clusters (IntClust). The IntClust 1-10 each have distinct CNAs and gene expression profiles. These clusters were also associated with distinct clinical features and outcomes [13]. Within the METABRIC cohort there are 11 integrative clusters due to the splitting of IntClust4 into IntClust4ER- and IntClust4ER+.
Tumour Stage	Tumour stages are defined based on the size of the tumour, regional nodal involvement, and distant metastasis. Stages are generally denoted by Roman numerals from I to IV where higher numbers denote an increase in the extent of disease and corresponding overall prognosis. The term Stage 0 describes carcinoma in situ and is considered to not have metastatic potential [14]. All tumour stages were seen in the METABRIC luminal cohort.
Tumour Size	The size of the breast tumour measured at its widest point. The tumour size within the METABRIC luminal cohort was measured in millimetres (mm) and ranges from 0-150mm.
Histological Subtype	Breast cancer classification and sub-classification based on histopathologic characteristics, including cell morphology, architecture, and growth patterns, into 21 distinct subtypes [15]. The METABRIC luminal cohort contains the following histological subtypes: Ductal/No Special Type, Lobular, Medullary, Mixed, Mucinous, Other and Tubular/Cribiform.
Chemotherapy	Cancer chemotherapy refers to the administration of cytotoxic chemicals with the aim to kill or at least shrink the tumour. Cytotoxic drugs are mostly given by the intravenous route [16]. This variable indicates whether a patient received chemotherapy as part of their treatment regime (Yes or No).
Radiotherapy	Radiation therapy is the application of high energy X-rays to shrink a tumour. Radiation therapy aims to deprive tumour cells of their multiplication potential. High-energy radiation does this by inducing genetic damage, leading to cell death. Radiation therapy is an important part of cancer treatment with about 50% of all cancer patients receiving radiation therapy [17]. This clinical variable indicates whether a patient received radiotherapy as part of their treatment regime (Yes or No).
Hormone Therapy	Endocrine therapy, also known as hormone therapy, lowers estrogen levels and inhibits the growth of the cancer. Hormone therapy is a systemic treatment for hormone receptor-positive (ER+ and PR+) breast cancer in the adjuvant, metastatic, and sometimes neoadjuvant setting [18]. This variable records whether a patient received hormone therapy as part of their treatment regime (Yes or No).
Breast Surgery	Whether the patient had breast-conserving surgery (BCS) (also known as lumpectomy) or a mastectomy. A mastectomy involves removal of the entire breast, while in BCS, the cancerous tumour is removed along with a margin of non-cancerous breast tissue [19]. Patients within the METABRIC luminal cohort were thus classified as having had BCS or a mastectomy.

3.2 Tishchenko Genes

Supplementary Table 2: List of top ten most up-regulated genes in the Tishchenko et al. [1] study.

Gene Symbol	Gene Name
POLQ	DNA Polymerase Theta
CKAP2L	Cytoskeleton Associated Protein 2 Like
KIFC1	Kinesin Family Member C1
FOXM1	Forkhead Box M1
TROAP	Trophinin Associated Protein
UBE2C	Ubiquitin Conjugating Enzyme E2C
AURKB	Aurora Kinase B
NCAPG	Non-SMC Condensin I Complex Subunit G
HJURP	Holliday Junction Recognition Protein
MCM10	Minichromosome Maintenance 10 Replication Initiation Factor

3.3 All Luminal Cases

Supplementary Table 3: Univariate Cox regression on clinical variables and CNA quartiles associated with METABRIC luminal cases. Rows where univariate Cox regression had significant p-value for either OS or DSS are highlighted. Note that both the Wald test p-values and likelihood ratio test p-values were used to determine significance. This was due to some factor level sample sizes being sufficiently small.

Clinical Variables	Overall Survival				Disease-Specific Survival			
	Beta	Hazard Ratio	Hazard Ratio CI	P-Values Wald Test / LRT	Beta	Hazard Ratio	Hazard Ratio CI	P-Values Wald Test / LRT
Positive Lymph Nodes	0.057	1.059	1.046 - 1.072	<2e-16 / 4.07e-13	0.070	1.072	1.057 - 1.087	<2e-16 / 3.01e-13
NPI	0.285	1.329	1.238 - 1.427	4.2e-15 / 6e-15	0.488	1.629	1.474 - 1.801	<2e-16 / <2.2e-16
Cellularity:	-	-	-	0.409 / 0.397	-	-	-	0.105 / 0.077
Low	Ref	Ref	Ref	Ref	Ref	Ref	Ref	Ref
Moderate	0.211	1.156	0.815 - 1.638	0.416	0.464	1.591	0.880 - 2.877	0.124
High	0.145	1.234	0.876 - 1.739	0.229	0.589	1.803	1.005 - 3.233	0.048
ER IHC:	-	-	-	0.035 / 0.052	-	-	-	6.27e-04 / 0.003
Negative	Ref	Ref	Ref	Ref	Ref	Ref	Ref	Ref
Positive	-0.517	0.596	0.368 - 0.965	0.035	-0.969	0.379	0.218 - 0.661	6.27e-04
HER2 SNP6:	-	-	-	0.058 / 0.070	-	-	-	0.033 / 0.044
Neutral	Ref	Ref	Ref	Ref	Ref	Ref	Ref	Ref
Loss	-0.088	0.916	0.644 - 1.303	0.626	0.145	1.156	0.724 - 1.844	0.545
Gain	0.275	1.316	1.072 - 1.615	0.009	0.419	1.520	1.150 - 2.008	0.003
Undefined	0.094	1.099	0.274 - 4.409	0.894	0.170	1.185	0.166 - 8.454	0.865
Inferred Menopausal State:	-	-	-	1.36e-12 / 3.61e-16	-	-	-	0.0282 / 0.022
Pre	Ref	Ref	Ref	Ref	Ref	Ref	Ref	Ref
Post	1.009	2.743	2.075 - 3.627	1.36e-12	0.357	1.429	1.039 - 1.966	0.0282
PAM50 Subtype:	-	-	-	3.81e-08 / 5.25e-08	-	-	-	2.24e-12 / 2.16e-12
Luminal A	Ref	Ref	Ref	Ref	Ref	Ref	Ref	Ref
Luminal B	0.431	1.539	1.320 - 1.794	3.81e-08	0.787	2.197	1.763 - 2.737	2.24e-12
Three Gene Classification:	-	-	-	6.35e-06 / 1.92e-05	-	-	-	5.22e-10 / 1.53e-09
ER-/HER2-	Ref	Ref	Ref	Ref	Ref	Ref	Ref	Ref
ER+/HER2- Low Prolif	-2.576	0.076	0.011 - 0.547	0.011	-3.213	0.040	0.006 - 0.293	0.002
ER+/HER2- High Prolif	-2.209	0.110	0.015 - 0.789	0.028	-2.512	0.081	0.011 - 0.586	0.013
HER2+	-2.058	0.128	0.017 - 0.945	0.044	-2.086	0.124	0.017 - 0.932	0.043
Integrative Clusters:	-	-	-	4.05e-06 / 1.36e-05	-	-	-	7.21e-11 / 5.03e-10
1	Ref	Ref	Ref	Ref	Ref	Ref	Ref	Ref
2	0.245	1.277	0.850 - 1.918	0.239	0.372	1.451	0.878 - 2.399	0.147
3	-0.243	0.784	0.567 - 1.085	0.142	-0.857	0.424	0.267 - 0.676	3e-04
4ER-	-12.120	5.45e-06	0.000 - Inf	0.988	-12.467	3.85e-06	0.000 - Inf	0.991
4ER+	-0.393	0.675	0.457 - 0.996	0.048	-0.705	0.494	0.287 - 0.851	0.011
5	0.674	1.963	1.287 - 2.994	0.002	0.742	2.101	1.251 - 3.526	0.005
6	0.088	1.092	0.727 - 1.641	0.672	0.121	1.129	0.672 - 1.896	0.646
7	-0.146	0.864	0.614 - 1.215	0.400	-0.482	0.617	0.386 - 0.988	0.044
8	-0.128	0.880	0.644 - 1.203	0.424	-0.404	0.668	0.441 - 1.012	0.057
9	0.236	1.266	0.868 - 1.848	0.220	0.126	1.134	0.694 - 1.851	0.616
10	-0.769	0.463	0.197 - 1.088	0.077	-0.577	0.562	0.198 - 1.593	0.278
Tumour Grade:	-	-	-	3.34e-04 / 2.95e-04	-	-	-	1.18e-07 / 2.12e-08
1	Ref	Ref	Ref	Ref	Ref	Ref	Ref	Ref
2	0.234	1.264	0.962 - 1.661	0.093	0.680	1.973	1.208 - 3.224	0.007
3	0.489	1.631	1.236 - 2.151	5.44e-04	1.164	3.203	1.965 - 5.220	3.01e-06
Laterality:	-	-	-	0.423 / 0.423	-	-	-	0.304 / 0.303
Left	Ref	Ref	Ref	Ref	Ref	Ref	Ref	Ref
Right	-0.064	0.938	0.802 - 1.097	0.423	-0.119	0.888	0.708 - 1.113	0.304

HER2 Status:	-	-	-	0.001 / 0.002	-	-	-	6.04e-06 / 5.41e-05
Negative	Ref	Ref	Ref	Ref	Ref	Ref	Ref	Ref
Positive	0.525	1.691	1.236 - 2.313	0.001	0.870	2.386	1.637 - 3.478	6.04e-06
ER Status:	-	-	-	0.880 / 0.877	-	-	-	0.672 / 0.693
Negative	Ref	Ref	Ref	Ref	Ref	Ref	Ref	Ref
Positive	0.151	1.163	0.164 - 8.272	0.880	-0.424	0.655	0.092 - 4.664	0.672
PR Status:	-	-	-	2.28e-05 / 3.51e-05	-	-	-	5.72e-06 / 1.04e-05
Negative	Ref	Ref	Ref	Ref	Ref	Ref	Ref	Ref
Positive	-0.355	0.701	0.595 - 0.826	2.28e-05	-0.526	0.591	0.471 - 0.742	5.72e-06
Cancer Type Detailed:	-	-	-	0.113 / 0.022	-	-	-	0.562 / 0.340
Breast	Ref	Ref	Ref	Ref	Ref	Ref	Ref	Ref
BI Ductal Carcinoma	1.518	4.562	1.137 - 18.30	0.032	1.418	4.131	0.579 - 29.453	0.157
BI Lobular Carcinoma	1.604	4.974	1.210 - 20.450	0.026	1.458	4.299	0.581 - 31.790	0.153
BI Mixed Mucinous Carcinoma	1.012	2.752	0.572 - 13.250	0.207	0.861	2.365	0.246 - 22.747	0.456
Mixed Ductal and Lobular Carcinoma	1.605	4.976	1.228 - 20.170	0.025	1.392	4.023	0.555 - 29.182	0.169
Chemotherapy:	-	-	-	0.099 / 0.112	-	-	-	9.75e-06 / 5.12e-05
No	Ref	Ref	Ref	Ref	Ref	Ref	Ref	Ref
Yes	0.247	1.280	0.955 - 1.717	0.099	0.738	2.092	1.508 - 2.902	9.75e-06
Hormone Therapy:	-	-	-	1.33e-06 / 6.17e-07	-	-	-	7.67e-04 / 4.93e-04
No	Ref	Ref	Ref	Ref	Ref	Ref	Ref	Ref
Yes	0.435	1.546	1.295 - 1.844	1.33e-06	0.445	1.560	1.204 - 2.021	7.67e-04
Radiotherapy:	-	-	-	0.008 / 0.008	-	-	-	0.725 / 0.725
No	Ref	Ref	Ref	Ref	Ref	Ref	Ref	Ref
Yes	-0.208	0.813	0.697 - 0.946	0.008	-0.040	0.961	0.772 - 1.198	0.725
Breast Surgery:	-	-	-	3.95e-10 / 1.45e-10	-	-	-	6.53e-05 / 4.43e-05
Lumpectomy	Ref	Ref	Ref	Ref	Ref	Ref	Ref	Ref
Mastectomy	0.520	1.682	1.429 - 1.979	3.95e-10	0.475	1.608	1.274 - 2.03	6.53e-05
Histological Subtype:	-	-	-	0.033 / 0.002	-	-	-	0.723 / 0.005
Ductal/No Special Type	Ref	Ref	Ref	Ref	Ref	Ref	Ref	Ref
Lobular	0.066	1.068	0.798 - 1.429	0.659	0.014	1.014	0.667 - 1.542	0.949
Medullary	0.096	1.101	0.155 - 7.838	0.923	0.690	1.994	0.280 - 14.223	0.491
Mixed	0.066	1.068	0.864 - 1.322	0.543	-0.053	0.949	0.692 - 1.300	0.743
Mucinous	-0.529	0.589	0.279 - 1.244	0.166	-0.586	0.557	0.178 - 1.740	0.314
Other	-1.540	0.214	0.053 - 0.860	0.030	-1.447	0.235	0.033 - 1.678	0.149
Tubular/cribiform	-1.128	0.324	0.134 - 0.782	0.012	-16.07	1.052e-07	0.000 - Inf	0.989
Tumour Stage:	-	-	-	2.64e-13 / 6.34e-13	-	-	-	3.63e-12 / 1.19e-10
0	-9.651	6.435e-05	0.000 - Inf	0.99	-10.02	4.464e-05	0.000 - Inf	0.993
1	Ref	Ref	Ref	Ref	Ref	Ref	Ref	Ref
2	0.613	1.847	1.509 - 2.260	2.74e-09	0.646	1.908	1.424 - 2.557	1.52e-05
3	1.161	3.192	2.263 - 4.503	3.74e-11	1.440	4.222	2.731 - 6.529	9.30e-11
4	1.702	5.483	2.412 - 12.462	4.86e-05	2.233	9.326	4.022 - 21.625	1.96e-07
Tumour Size	0.019	1.019	1.015 - 1.023	<2e-16 / 4.43e-15	0.021	1.021	1.017 - 1.026	<2e-16 / 8.74e-13
Age at Diagnosis	0.054	1.055	1.048 - 1.063	<2e-16 / <2.2e-16	0.022	1.022	1.012 - 1.032	9.44e-06 / 7.18e-06
Quartiles:	-	-	-	2.13e-05 / 3.08e-05	-	-	-	8.53e-09 / 3.89e-09
1	Ref	Ref	Ref	Ref	Ref	Ref	Ref	Ref
2	0.088	1.092	0.871 - 1.369	0.445	0.283	1.327	0.9224 - 1.908	0.127
3	0.220	1.246	0.999 - 1.555	0.051	0.688	1.991	1.4182 - 2.794	6.91e-05
4	0.496	1.642	1.325 - 2.035	6.08e-06	0.962	2.616	1.8810 - 3.638	1.10e-08
Ref: Factor level used as baseline; CI: Confidence interval; LRT: Likelihood Ratio Test; NPI: Nottingham Prognostic Index; HER2: Human epidermal growth factor receptor 2; ER: Estrogen receptor; ER IHC: Estrogen receptor immunohistochemistry; PR: Progesterone receptor; BI: Breast Invasive								

Supplementary Table 4: Assessment of association between CNA quartiles and clinical variables that showed a significant association with either OS, DSS or both in METABRIC luminal cases. Rows with significant p-values are highlighted.

Clinical Variables	Chi-Squared Test	Fisher's Exact Test	Kruskal-Wallis Test
Positive Lymph Nodes	-	-	6.79e-04
NPI	-	-	<2.2e-16
ER IHC	0.309	0.262	-
HER2 SNP6	8.002e-12*	4.998e-04'	-
Inferred Menopausal State	0.033	0.029	-
PAM50 Subtype	<2.2e-16	<2.2e-16	-
Three Gene Classification	<2.2e-16*	4.998e-04'	-
Integrative Clusters	<2.2e-16*	4.998e-04'	-
Tumour Grade	<2.2e-16	4.998e-04'	-
HER2 Status	2.497e-06	1.354e-06	-
PR Status	1.323e-09	2.15e-10	-
Cancer Type Detailed	4.29e-04*	0.001'	-
Chemotherapy	0.001	3.32e-04	-
Hormone Therapy	4.511e-05	5.048e-05	-
Radiotherapy	0.0098	0.010	-
Breast Surgery	0.339	0.342	-
Histological Subtype	1.13e-07*	4.998e-04'	-
Tumour Stage	0.003*	9.995e-04'	-
Tumour Size	-	-	2.958e-12
Age at Diagnosis	-	-	0.020
* Chi-Squared approximation may be incorrect ' Simulated P-value for Fisher's Exact Test NPI: Nottingham Prognostic Index; ER IHC: Estrogen receptor immunohistochemistry; HER2: Human epidermal growth factor receptor 2; PR: Progesterone receptor			

Supplementary Table 5: Multivariate Cox regression on CNA quartiles and top three associated variables in METABRIC luminal cases. Rows with significant results for OS and DSS are highlighted.

Variables	Overall Survival				Disease-Specific Survival			
	Beta	Hazard Ratio	Hazard Ratio CI	P-value	Beta	Hazard Ratio	Hazard Ratio CI	P-value
Quartile2	-0.053	0.948	0.749 - 1.1999	0.657	0.147	1.158	0.791 - 1.694	0.450
Quartile3	0.074	1.076	0.850 - 1.363	0.541	0.507	1.660	1.154 - 2.388	0.006
Quartile4	0.231	1.260	0.987 - 1.610	0.064	0.545	1.725	1.185 - 2.512	0.004
NPI	0.468	1.597	1.436 - 1.776	<2e-16	0.607	1.835	1.588 - 2.121	<2e-16
PAM50LumB	0.348	1.417	1.191 - 1.686	8.65e-05	0.564	1.757	1.376 - 2.243	6.14e-06
Grade2	-0.376	0.687	0.508 - 0.930	0.015	-0.222	0.801	0.474 - 1.353	0.406
Grade3	-0.819	0.441	0.299 - 0.649	3.38e-05	-0.719	0.487	0.263 - 0.903	0.022
	Likelihood Ratio Test P-value = <2e-16				Likelihood Ratio Test P-value = <2e-16			
	Wald Test P-value = <2e-16				Wald Test P-value = <2e-16			
	Score (logrank) Test P-value = <2e-16				Score (logrank) Test P-value = <2e-16			

3.4 Luminal A Cases

Supplementary Table 6: Univariate Cox regression on clinical variables and CNA quartiles associated with METABRIC lumA cases. Rows where univariate Cox regression had a significant p-value for OS or DSS are highlighted. Note that both the Wald test p-values and likelihood ratio test p-values were used to determine significance. This was due to some factor level sample sizes being sufficiently small.

Clinical Variables	Overall Survival				Disease-Specific Survival			
	Beta	Hazard Ratio	Hazard Ratio CI	P-Values Wald Test / LRT	Beta	Hazard Ratio	Hazard Ratio CI	P-Values Wald Test / LRT
Positive Lymph Nodes	0.114	1.120	1.09 - 1.151	3.52e-16 / 9.58e-12	0.142	1.153	1.114 - 1.193	2.84e-16 / 6.41e-11
NPI	0.206	1.229	1.116 - 1.353	2.6e-05 / 3.15e-05	0.412	1.510	1.305 - 1.748	3.37e-08 / 5.43e-08
Cellularity:	-	-	-	0.376 / 0.351	-	-	-	0.279 / 0.209
Low	Ref	Ref	Ref	Ref	Ref	Ref	Ref	Ref
Moderate	0.246	1.279	0.826 - 1.982	0.270	0.680	1.974	0.857 - 4.547	0.110
High	0.306	1.358	0.879 - 2.098	0.168	0.638	1.892	0.821 - 4.362	0.134
ER IHC:	-	-	-	0.103 / 0.136	-	-	-	0.004 / 0.015
Negative	Ref	Ref	Ref	Ref	Ref	Ref	Ref	Ref
Positive	-0.583	0.558	0.277 - 1.126	0.103	-1.201	0.301	0.133 - 0.682	0.004
HER2 SNP6:	-	-	-	0.038 / 0.023	-	-	-	0.124 / 0.133
Neutral	Ref	Ref	Ref	Ref	Ref	Ref	Ref	Ref
Loss	-0.367	0.693	0.405 - 1.185	0.180	-0.016	0.985	0.459 - 2.112	0.968
Gain	0.413	1.512	1.089 - 2.097	0.013	0.580	1.785	1.109 - 2.875	0.017
Undefined	-13.99	8.362e-07	0.000 - Inf	0.991	-13.97	8.584e-07	0.000 - Inf	0.995
Inferred Menopausal State:	-	-	-	2.55e-11 / 1.07e-15	-	-	-	0.034 / 0.025
Pre	Ref	Ref	Ref	Ref	Ref	Ref	Ref	Ref
Post	1.315	3.724	2.531 - 5.481	2.55e-11	0.492	1.635	1.037 - 2.578	0.034
PAM50 Subtype	-	-	-	-	-	-	-	-
Three Gene Classification:	-	-	-	0.015 / 0.018	-	-	-	0.002 / 0.002
ER-/HER2-	Ref	Ref	Ref	Ref	Ref	Ref	Ref	Ref
ER+/HER2- Low Prolif	0.310	1.363	1.076 - 1.726	0.010	0.563	1.756	1.236 - 2.495	0.002
ER+/HER2- High Prolif	0.533	1.704	0.904 - 3.211	0.100	0.946	2.574	1.124 - 5.894	0.025
HER2+	-	-	-	0.062 / 0.043	-	-	-	0.004 / 0.009
Integrative Clusters:	Ref	Ref	Ref	Ref	Ref	Ref	Ref	Ref
1	0.724	2.063	0.689 - 6.176	0.196	0.985	2.677	0.593 - 12.090	0.201
2	0.165	1.179	0.434 - 3.206	0.747	-0.396	0.673	0.160 - 2.824	0.589
3	-14.83	3.635e-07	0.000 - Inf	0.997	-14.21	6.77e-07	0.000 - Inf	0.997
4ER-	-0.050	0.951	0.339 - 2.667	0.924	-0.184	0.832	0.192 - 3.611	0.806
4ER+	1.007	2.738	0.882 - 8.499	0.081	0.923	2.516	0.507 - 12.483	0.259
5	0.513	1.671	0.545 - 5.128	0.370	0.543	1.721	0.357 - 8.291	0.499
6	0.284	1.328	0.485 - 3.639	0.581	0.070	1.072	0.255 - 4.504	0.924
7	0.126	1.134	0.417 - 3.087	0.805	-0.129	0.879	0.212 - 3.649	0.859
8	0.488	1.629	0.518 - 5.121	0.403	0.658	1.930	0.401 - 9.302	0.412
9	-14.85	3.565e-07	0.000 - Inf	0.992	-14.20	6.77e-07	0.000 - Inf	0.994
10	-	-	-	0.046 / 0.045	-	-	-	9.35e-04 / 4.1e-04
Tumour Grade:	Ref	Ref	Ref	Ref	Ref	Ref	Ref	Ref
1	0.211	1.235	0.904 - 1.688	0.186	0.716	2.047	1.113 - 3.764	0.021
2	0.415	1.515	1.079 - 2.126	0.016	1.137	3.117	1.661 - 5.851	0.001
3	-	-	-	0.225 / 0.224	-	-	-	0.135 / 0.133
Laterality:	Ref	Ref	Ref	Ref	Ref	Ref	Ref	Ref
Left	-0.130	0.878	0.712 - 1.083	0.225	-0.255	0.775	0.555 - 1.083	0.135
Right	-	-	-	0.061 / 0.083	-	-	-	0.043 / 0.070
HER2 Status:	Ref	Ref	Ref	Ref	Ref	Ref	Ref	Ref
Negative	0.531	1.700	0.977 - 2.96	0.061	0.786	2.195	1.026 - 4.695	0.043
Positive	-	-	-	0.925 / 0.927	-	-	-	0.389 / 0.451
ER Status:	Ref	Ref	Ref	Ref	Ref	Ref	Ref	Ref
Negative	-0.094	0.910	0.128 - 6.49	0.925	-0.865	0.421	0.059 - 3.016	0.389
Positive	-	-	-	3.24e-05 / 6.47e-05	-	-	-	2.27e-04 / 4.28e-04
PR Status:	Ref	Ref	Ref	Ref	Ref	Ref	Ref	Ref
Negative	-0.492	0.612	0.485 - 0.771	3.24e-05	-0.660	0.517	0.364 - 0.734	2.27e-04
Positive	-	-	-	0.203 / 0.077	-	-	-	0.680 / 0.318
Cancer Type Detailed:	Ref	Ref	Ref	Ref	Ref	Ref	Ref	Ref
Breast	1.549	4.707	0.660 - 33.56	0.122	0.548	1.729	0.241 - 12.42	0.586
BI Ductal Carcinoma	1.569	4.803	0.656 - 35.17	0.122	0.414	1.513	0.195 - 11.74	0.692
BI Lobular Carcinoma	0.500	1.649	0.103 - 26.38	0.724	-14.44	5.338e-07	0.000 - Inf	0.993
BI Mixed Mucinous Carcinoma	1.751	5.762	0.800 - 41.50	0.082	0.809	2.245	0.306 - 16.49	0.427
Mixed Ductal and Lobular Carcinoma	-	-	-	0.621 / 0.627	-	-	-	0.002 / 0.006
Chemotherapy:	Ref	Ref	Ref	Ref	Ref	Ref	Ref	Ref
No	0.115	1.122	0.711 - 1.768	0.621	0.792	2.208	1.322 - 3.686	0.002
Yes	-	-	-	2.59e-05 / 1.57e-05	-	-	-	0.012 / 0.009
Hormone Therapy:	Ref	Ref	Ref	Ref	Ref	Ref	Ref	Ref
No	0.488	1.630	1.298 - 2.046	2.59e-05	0.469	1.599	1.111 - 2.302	0.012
Yes	-	-	-	1.27e-04 / 1.19e-04	-	-	-	0.014 / 0.013
Radiotherapy:	Ref	Ref	Ref	Ref	Ref	Ref	Ref	Ref
No	-0.406	0.666	0.541 - 0.820	1.27e-04	-0.412	0.662	0.477 - 0.920	0.014
Yes	-	-	-	1.46e-09 / 2.58e-10	-	-	-	5.08e-05 / 1.94e-05
Breast Surgery:	Ref	Ref	Ref	Ref	Ref	Ref	Ref	Ref
Lumpectomy	0.713	2.040	1.619 - 2.571	1.46e-09	0.774	2.169	1.492 - 3.155	5.08e-05
Mastectomy	-	-	-	0.066 / 0.011	-	-	-	0.852 / 0.013
Histological Subtype:	Ref	Ref	Ref	Ref	Ref	Ref	Ref	Ref
Ductal/No Special Type	-0.012	0.988	0.684 - 1.428	0.949	-0.181	0.834	0.448 - 1.555	0.568
Lobular	-	-	-	-	-	-	-	-
Medullary	0.170	1.185	0.909 - 1.546	0.209	0.213	1.238	0.823 - 1.861	0.306
Mixed	-1.081	0.339	0.048 - 2.419	0.281	-17.06	3.91e-08	0.000 - Inf	0.997
Mucinous	-1.583	0.205	0.029 - 1.465	0.114	-0.598	0.549	0.077 - 3.948	0.552
Other	-0.971	0.379	0.156 - 0.919	0.032	-17.06	3.90e-08	0.000 - Inf	0.994
Tubular/ciribriform	-	-	-	-	-	-	-	-

Tumour Stage:	-	-	-	1.40e-05 / 1.75e-05	-	-	-	1.87e-05 / 1.07e-04
0	Ref	Ref	Ref	Ref	Ref	Ref	Ref	Ref
1	0.540	1.716	1.324 - 2.224	4.56e-05	0.607	1.835	1.215 - 2.770	0.004
2	0.949	2.584	1.536 - 4.349	3.5e-04	1.380	3.975	2.008 - 7.869	7.474e-05
3	1.616	5.033	1.234 - 20.523	0.024	2.412	11.157	2.663 - 46.752	9.682e-04
4	0.022	1.022	1.014 - 1.03	8.86e-09 / 2.54e-07	0.030	1.030	1.021 - 1.04	9.57e-11 / 1.29e-07
Tumour Size	0.070	1.072	1.061 - 1.083	<2e-16 / <2e-16	0.029	1.029	1.014 - 1.044	9.68e-05 / 7.89e-05
Age at Diagnosis	-	-	-	0.040 / 0.045	-	-	-	2.53e-04 / 2.14e-04
Quartiles:	Ref	Ref	Ref	Ref	Ref	Ref	Ref	Ref
1	0.094	1.098	0.812 - 1.485	0.542	0.211	1.235	0.715 - 2.135	0.450
2	0.135	1.145	0.847 - 1.547	0.379	0.609	1.839	1.108 - 3.053	0.018
3	0.395	1.485	1.114 - 1.980	0.007	0.954	2.597	1.600 - 4.216	1.13e-04
4								

Ref: Factor level used as baseline; CI: Confidence interval; LRT: Likelihood Ratio Test; NPI: Nottingham Prognostic Index;
HER2: Human epidermal growth factor receptor 2; ER: Estrogen receptor; ER IHC: Estrogen receptor immunohistochemistry; PR: Progesterone receptor;
BI: Breast Invasive

Supplementary Table 7: Assessment of association between CNA quartiles and clinical variables that showed a significant association with either OS, DSS or both in METABRIC lumA cases. Rows where p-value was significant are highlighted.

Clinical Variables	Chi-Squared Test	Fisher's Exact Test	Kruskal-Wallis Test
Positive Lymph Nodes	-	-	0.114
NPI	-	-	8.37e-12
ER IHC	0.417*	0.344	-
HER2 SNP6	0.005*	0.001	-
Inferred Menopausal State	0.403	0.423	-
Three Gene Classification	5.6e-12*	4.998e-04'	-
Integrative Clusters	1.63e-15*	4.988e-04'	-
Tumour Grade	1.637e-11	4.988e-04'	-
HER2 Status	0.077	0.102	-
PR Status	1.69e-04	1.65e-04	-
Chemotherapy	0.005	0.003	-
Hormone Therapy	0.263	0.264	-
Radiotherapy	0.028	0.027	-
Breast Surgery	0.075	0.076	-
Histological Subtype	0.003*	0.005'	-
Tumour Stage	0.410*	0.323'	-
Tumour Size	-	-	1.52e-06
Age at Diagnosis	-	-	0.121

* Chi-Squared approximation may be incorrect
' Simulated P-value for Fisher's Exact Test
NPI: Nottingham Prognostic Index; ER IHC: Estrogen receptor immunohistochemistry;
HER2: Human epidermal growth factor receptor 2; PR: Progesterone receptor

Supplementary Table 8: Multivariate Cox regression on CNA quartiles and top four associated variables in METABRIC lumA cases. Significant results for OS and DSS are highlighted.

	Overall Survival				Disease-Specific Survival			
Variables	Beta	Hazard Ratio	Hazard Ratio CI	P-value	Beta	Hazard Ratio	Hazard Ratio CI	P-value
Quartile2	-0.058	0.944	0.672 - 1.326	0.739	0.148	1.160	0.628 - 2.143	0.637
Quartile3	0.007	1.007	0.716 - 1.416	0.968	0.531	1.701	0.958 - 3.020	0.070
Quartile4	0.155	1.168	0.827 - 1.649	0.377	0.767	2.153	1.223 - 3.789	0.008
NPI	0.296	1.345	1.150 - 1.572	2.03e-04	0.426	1.531	1.220 - 1.920	2.3e-04
ThreeGene ER+/HER2- High Prolif	0.234	1.263	0.964 - 1.656	0.090	0.181	1.198	0.809 - 1.773	0.367
ThreeGene HER2+	-0.485	0.616	0.092 - 4.101	0.616	-0.365	0.694	0.054 - 8.916	0.779
IntClust2	0.724	2.062	0.587 - 7.244	0.259	0.732	2.078	0.450 - 9.605	0.349
IntClust3	0.481	1.618	0.501 - 5.233	0.421	-0.228	0.796	0.183 - 3.459	0.761
IntClust4ER-	-14.306	6.12e-07	0.000 - Inf	0.997	-13.757	1.06e-06	0.000 - Inf	0.997
IntClust4ER+	0.234	1.264	0.375 - 4.256	0.706	0.080	1.083	0.237 - 4.944	0.918
IntClust5	1.643	5.170	0.570 - 46.914	0.144	1.260	3.525	0.195 - 63.709	0.394
IntClust6	0.743	2.101	0.590 - 7.487	0.252	0.417	1.517	0.310 - 7.426	0.607
IntClust7	0.560	1.750	0.540 - 5.671	0.351	0.282	1.325	0.308 - 5.706	0.705
IntClust8	0.420	1.522	0.474 - 4.890	0.481	-0.018	0.982	0.232 - 4.156	0.980
IntClust9	0.505	1.658	0.454 - 6.060	0.445	0.373	1.453	0.300 - 7.047	0.643
IntClust10	-14.611	4.51e-07	0.000 - Inf	0.993	-14.236	6.57e-07	0.000 - Inf	0.994
Grade2	-0.205	0.815	0.559 - 1.189	0.288	0.025	1.026	0.522 - 2.015	0.942
Grade3	-0.449	0.638	0.379 - 1.075	0.092	-0.248	0.780	0.334 - 1.822	0.566
	Likelihood Ratio Test P-value = 0.003				Likelihood Ratio Test P-value = 3e-06			
	Wald Test P-value = 0.005				Wald Test P-value = 4e-06			
	Score (logrank) Test P-value = 0.002				Score (logrank) Test P-value = 4e-07			

Supplementary Table 11: Multivariate Cox regression on CNA quartiles and top associated variable in METABRIC lumA cases. Significant results for OS and DSS are highlighted.

Variables	Overall Survival				Disease-Specific Survival			
	Beta	Hazard Ratio	Hazard Ratio CI	P-value	Beta	Hazard Ratio	Hazard Ratio CI	P-value
Quartile2	0.103	1.108	0.819 - 1.499	0.505	0.213	1.237	0.716 - 2.139	0.446
Quartile3	0.067	1.069	0.790 - 1.448	0.665	0.470	1.600	0.961 - 2.663	0.071
Quartile4	0.299	1.349	1.007 - 1.806	0.045	0.755	2.127	1.301 - 3.477	0.003
NPI	0.189	1.208	1.095 - 1.333	1.73e-04	0.358	1.430	1.230 - 1.663	3.39e-06
	Likelihood Ratio Test P-value = 2e-04 Wald Test P-value = 2e-04 Score (logrank) Test P-value = 2e-04				Likelihood Ratio Test P-value = 3e-08 Wald Test P-value = 3e-08 Score (logrank) Test P-value = 2e-08			

3.5 Luminal B Cases

Supplementary Table 12: Univariate Cox regression on clinical variables and CNA quartiles associated with METABRIC lumB cases. Rows where univariate Cox regression had significant p-value for OS or DSS are highlighted. Note that both the Wald test p-values and likelihood ratio test p-values were used to determine significance. This was due to some factor level sample sizes being sufficiently small.

Clinical Variables	Overall Survival				Disease-Specific Survival			
	Beta	Hazard Ratio	Hazard Ratio CI	P-Values Wald Test / LRT	Beta	Hazard Ratio	Hazard Ratio CI	P-Values Wald Test / LRT
Positive Lymph Nodes	0.039	1.040	1.023 - 1.056	1.43e-06 / 5.58e-05	0.048	1.049	1.031 - 1.068	1.37e-07 / 2.01e-05
NPI	0.319	1.375	1.228 - 1.540	3.7e-08 / 3.36e-08	0.455	1.576	1.362 - 1.824	1.06e-09 / 8.29e-10
Cellularity:	-	-	-	0.677 / 0.688	-	-	-	0.955 / 0.954
Low	Ref	Ref	Ref	Ref	Ref	Ref	Ref	Ref
Moderate	-0.164	0.849	0.476 - 1.514	0.579	0.078	1.082	0.466 - 2.511	0.855
High	-0.227	0.797	0.454 - 1.399	0.429	0.109	1.116	0.491 - 2.535	0.794
ER IHC:	-	-	-	0.239 / 0.268	-	-	-	0.073 / 0.105
Negative	Ref	Ref	Ref	Ref	Ref	Ref	Ref	Ref
Positive	-0.399	0.671	0.345 - 1.304	0.239	-0.692	0.501	0.235 - 1.067	0.073
HER2 SNP6:	-	-	-	0.915 / 0.925	-	-	-	0.965 / 0.967
Neutral	Ref	Ref	Ref	Ref	Ref	Ref	Ref	Ref
Loss	0.105	1.111	0.694 - 1.779	0.661	0.123	1.131	0.624 - 2.048	0.685
Gain	0.015	1.015	0.775 - 1.330	0.913	0.044	1.044	0.737 - 1.479	0.807
Undefined	0.413	1.511	0.375 - 6.091	0.561	0.290	1.337	0.187 - 9.577	0.773
Inferred Menopausal State:	-	-	-	0.080 / 0.066	-	-	-	0.723 / 0.726
Pre	Ref	Ref	Ref	Ref	Ref	Ref	Ref	Ref
Post	0.360	1.434	0.958 - 2.146	0.080	-0.081	0.922	0.590 - 1.443	0.723
PAM50 Subtype	-	-	-	-	-	-	-	-
Three Gene Classification:	-	-	-	0.190 / 0.384	-	-	-	0.037 / 0.120
ER-/HER2-	Ref	Ref	Ref	Ref	Ref	Ref	Ref	Ref
ER+/HER2- Low Prolif	-1.848	0.158	0.021 - 1.190	0.073	-2.569	0.077	0.009 - 0.624	0.016
ER+/HER2- High Prolif	-2.032	0.131	0.018 - 0.950	0.044	-2.336	0.097	0.013 - 0.706	0.021
HER2+	-1.922	0.146	0.019 - 1.107	0.063	-1.898	0.150	0.020 - 1.151	0.068
Integrative Clusters:	-	-	-	0.130 / 0.134	-	-	-	0.027 / 0.044
1	Ref	Ref	Ref	Ref	Ref	Ref	Ref	Ref
2	0.220	1.246	0.773 - 2.011	0.367	0.351	1.420	0.791 - 2.549	0.240
3	0.031	1.031	0.627 - 1.696	0.903	-0.257	0.773	0.387 - 1.545	0.466
4ER-	-	-	-	-	-	-	-	-
4ER+	0.005	1.005	0.534 - 1.892	0.988	-0.413	0.662	0.257 - 1.706	0.393
5	0.679	1.972	1.215 - 3.201	0.006	0.866	2.378	1.349 - 4.191	0.003
6	0.112	1.119	0.704 - 1.778	0.635	0.183	1.201	0.673 - 2.140	0.535
7	-0.103	0.902	0.535 - 1.521	0.699	-0.322	0.725	0.354 - 1.484	0.379
8	0.220	1.246	0.858 - 1.810	0.248	0.089	1.093	0.673 - 1.774	0.720
9	0.249	1.283	0.854 - 1.927	0.231	0.098	1.103	0.648 - 1.877	0.717
10	-0.561	0.571	0.244 - 1.335	0.196	-0.457	0.633	0.223 - 1.802	0.392
Tumour Grade:	-	-	-	0.594 / 0.592	-	-	-	0.260 / 0.256
1	Ref	Ref	Ref	Ref	Ref	Ref	Ref	Ref
2	-0.096	0.908	0.501 - 1.646	0.751	-0.012	0.989	0.428 - 2.283	0.978
3	0.028	1.029	0.572 - 1.850	0.925	0.244	1.276	0.560 - 2.908	0.562
Laterality:	-	-	-	0.693 / 0.693	-	-	-	0.797 / 0.797
Left	Ref	Ref	Ref	Ref	Ref	Ref	Ref	Ref
Right	0.048	1.049	0.827 - 1.330	0.693	0.040	1.041	0.765 - 1.417	0.797
HER2 Status:	-	-	-	0.093 / 0.108	-	-	-	0.006 / 0.011
Negative	Ref	Ref	Ref	Ref	Ref	Ref	Ref	Ref
Positive	0.329	1.389	0.947 - 2.039	0.093	0.614	1.847	1.191 - 2.865	0.006
ER Status:	-	-	-	-	-	-	-	-
Negative	-	-	-	-	-	-	-	-
Positive	-	-	-	-	-	-	-	-
PR Status:	-	-	-	0.413 / 0.415	-	-	-	0.149 / 0.152
Negative	Ref	Ref	Ref	Ref	Ref	Ref	Ref	Ref
Positive	-0.098	0.906	0.716 - 1.147	0.413	-0.221	0.801	0.594 - 1.082	0.149
Cancer Type Detailed:	-	-	-	0.067 / 0.038	-	-	-	0.132 / 0.022
Breast	Ref	Ref	Ref	Ref	Ref	Ref	Ref	Ref
BI Ductal Carcinoma	1.551	4.714	0.661 - 33.634	0.122	16.057	9405458	0.000 - Inf	0.993
BI Lobular Carcinoma	2.089	8.080	1.077 - 60.608	0.042	16.712	18102653	0.000 - Inf	0.993
BI Mixed Mucinous Carcinoma	1.083	2.955	0.355 - 24.565	0.316	15.546	5643865	0.000 - Inf	0.993
Mixed Ductal and Lobular Carcinoma	1.549	4.705	0.643 - 34.437	0.127	15.831	7503154	0.000 - Inf	0.993
Chemotherapy:	-	-	-	0.118 / 0.133	-	-	-	0.004 / 0.007
No	Ref	Ref	Ref	Ref	Ref	Ref	Ref	Ref
Yes	0.308	1.360	0.925 - 2.000	0.118	0.627	1.872	1.224 - 2.864	0.004

Hormone Therapy:	-	-	-	0.168 / 0.160	-	-	-	0.380 / 0.372
No	Ref	Ref	Ref	Ref	Ref	Ref	Ref	Ref
Yes	0.199	1.221	0.919 - 1.621	0.168	0.167	1.181	0.814 - 1.714	0.380
Radiotherapy:	-	-	-	0.658 / 0.658	-	-	-	0.326 / 0.322
No	Ref	Ref	Ref	Ref	Ref	Ref	Ref	Ref
Yes	-0.053	0.949	0.751 - 1.198	0.658	0.156	1.169	0.856 - 1.595	0.326
Breast Surgery:	-	-	-	0.003 / 0.003	-	-	-	0.039 / 0.038
Lumpectomy	Ref	Ref	Ref	Ref	Ref	Ref	Ref	Ref
Mastectomy	0.347	1.415	1.121 - 1.787	0.003	0.318	1.375	1.016 - 1.861	0.039
Histological Subtype:	-	-	-	0.117 / 0.069	-	-	-	0.211 / 0.041
Ductal/No Special Type	Ref	Ref	Ref	Ref	Ref	Ref	Ref	Ref
Lobular	0.538	1.712	1.059 - 2.768	0.028	0.656	1.928	1.092 - 3.405	0.024
Medullary	-0.190	0.827	0.116 - 5.902	0.850	0.262	1.299	0.182 - 9.296	0.794
Mixed	-0.003	0.997	0.693 - 1.436	0.988	-0.224	0.799	0.477 - 1.340	0.395
Mucinous	-0.468	0.626	0.277 - 1.414	0.260	-0.509	0.601	0.191 - 1.893	0.384
Other	-1.551	0.212	0.030 - 1.512	0.122	-16.055	1.065e-07	0.000 - Inf	0.993
Tubular/ciribriform	-	-	-	-	-	-	-	-
Tumour Stage:	-	-	-	7.99e-07 / 7.13e-07	-	-	-	2.94e-05 / 6.82e-05
0	-11.519	9.93e-06	0.000 - Inf	0.995	-10.679	2.30e-05	0.000 - Inf	0.995
1	Ref	Ref	Ref	Ref	Ref	Ref	Ref	Ref
2	0.697	2.007	1.450 - 2.779	2.66e-05	0.630	1.878	1.237 - 2.852	0.003
3	1.259	3.520	2.181 - 5.681	2.55e-07	1.298	3.663	2.055 - 6.529	1.08e-05
4	1.619	5.049	1.811 - 14.074	0.002	1.878	6.541	2.292 - 18.665	4.48e-04
Tumour Size	0.015	1.015	1.010 - 1.020	1.99e-10 / 2.73e-07	0.015	1.015	1.010 - 1.021	8.2e-08 / 1.76e-05
Age at Diagnosis	0.030	1.031	1.019 - 1.042	1.27e-07 / 5.49e-08	0.008	1.009	0.995 - 1.022	0.224 / 0.220
Quartiles:	-	-	-	0.047 / 0.054	-	-	-	0.007 / 0.009
1	Ref	Ref	Ref	Ref	Ref	Ref	Ref	Ref
2	-0.078	0.925	0.665 - 1.287	0.644	0.181	1.198	0.769 - 1.867	0.425
3	0.346	1.413	1.030 - 1.938	0.032	0.666	1.947	1.284 - 2.953	0.002
4	0.072	1.074	0.776 - 1.486	0.665	0.177	1.193	0.763 - 1.868	0.439

Ref: Factor level used as baseline; CI: Confidence interval; LRT: Likelihood Ratio Test; NPI: Nottingham Prognostic Index ;
HER2: Human epidermal growth factor receptor 2; ER: Estrogen receptor; ER IHC: Estrogen receptor immunohistochemistry; PR: Progesterone receptor;
BI: Breast Invasive

Supplementary Table 13: Assessment of association between CNA quartiles and clinical variables that showed a significant association with either OS, DSS or both in METABRIC lumB cases. Rows where p-value was significant are highlighted.

Clinical Variables	Chi-Squared Test	Fisher's Exact Test	Kruskal-Wallis Test
Positive Lymph Nodes	-	-	0.143
NPI	-	-	1.36e-04
Three Gene Classification	2.27e-10*	4.998e-04'	-
Integrative Clusters	<2.2e-16*	4.998e-04'	-
HER2 Status	9.45e-05	3.01e-05	-
Cancer Type Detailed	0.028*	0.020'	-
Chemotherapy	0.097	0.110	-
Breast Surgery	0.241	0.242	-
Histological Subtype	0.047*	0.015'	-
Tumour Stage	0.128*	0.071'	-
Tumour Size	-	-	0.028
Age at Diagnosis	-	-	0.316
* Chi-Squared approximation may be incorrect ' Simulated P-value for Fisher's Exact Test NPI: Nottingham Prognostic Index; ER IHC: Estrogen receptor immunohistochemistry; HER2: Human epidermal growth factor receptor 2; PR: Progesterone receptor			

Supplementary Table 14: Multivariate Cox regression on CNA quartiles and top four associated variables in METABRIC lumB cases. Significant results for OS and DSS are highlighted.

	Overall Survival				Disease-Specific Survival			
Variables	Beta	Hazard Ratio	Hazard Ratio CI	P-value	Beta	Hazard Ratio	Hazard Ratio CI	P-value
Quartile2	0.100	1.105	0.747 - 1.637	0.617	0.353	1.424	0.836 - 2.425	0.193
Quartile3	0.450	1.568	1.048 - 2.347	0.029	0.636	1.888	1.109 - 3.213	0.019
Quartile4	0.334	1.397	0.925 - 2.108	0.112	0.293	1.341	0.759 - 2.368	0.312
NPI	0.339	1.404	1.238 - 1.591	1.14e-07	0.466	1.593	1.347 - 1.884	5.48e-08
ThreeGeneER+/HER2- Low Prolif	-1.484	0.227	0.028 - 1.806	0.161	-2.005	0.135	0.015 - 1.200	0.072
ThreeGeneER+/HER2- High Prolif	-1.829	0.161	0.021 - 1.201	0.075	-2.013	0.134	0.017 - 1.028	0.053
ThreeGeneHER2+	-2.238	0.107	0.012 - 0.920	0.042	-2.389	0.092	0.010 - 0.854	0.036
IntClust2	0.310	1.363	0.809 - 2.298	0.245	0.487	1.628	0.837 - 3.169	0.151
IntClust3	0.616	1.851	1.025 - 3.340	0.041	0.579	1.785	0.787 - 4.047	0.165
IntClust4ER+	0.275	1.317	0.665 - 2.607	0.430	-0.049	0.952	0.323 - 2.804	0.929
IntClust5	1.235	3.440	1.237 - 9.563	0.018	1.335	3.800	1.153 - 12.522	0.028
IntClust6	0.167	1.182	0.713 - 1.960	0.517	0.346	1.414	0.742 - 2.695	0.292
IntClust7	0.297	1.346	0.748 - 2.421	0.322	0.382	1.465	0.651 - 3.298	0.356
IntClust8	0.443	1.557	0.999 - 2.429	0.051	0.441	1.555	0.856 - 2.825	0.147
IntClust9	0.314	1.369	0.874 - 2.145	0.170	0.240	1.271	0.696 - 2.322	0.435
IntClust10	-0.547	0.579	0.241 - 1.387	0.220	-0.178	0.837	0.281 - 2.492	0.749
HER2_StatusPositive	-0.325	0.723	0.245 - 2.130	0.556	-0.196	0.822	0.233 - 2.906	0.761
Likelihood Ratio Test P-value = 1e-05					Likelihood Ratio Test P-value = 4e-06			
Wald Test P-value = 5e-06					Wald Test P-value = 1e-06			
Score (logrank) Test P-value = 2e-06					Score (logrank) Test P-value = 1e-07			

Supplementary Table 15: Multivariate Cox regression on CNA quartiles and top three associated variables in METABRIC lumB cases. Significant results for OS and DSS are highlighted.

	Overall Survival				Disease-Specific Survival			
Variables	Beta	Hazard Ratio	Hazard Ratio CI	P-value	Beta	Hazard Ratio	Hazard Ratio CI	P-value
Quartile2	0.109	1.115	0.754 - 1.649	0.584	0.359	1.432	0.842 - 2.435	0.185
Quartile3	0.442	1.556	1.040 - 2.327	0.031	0.630	1.877	1.104 - 3.190	0.020
Quartile4	0.338	1.402	0.929 - 2.115	0.108	0.297	1.346	0.762 - 2.375	0.306
NPI	0.338	1.402	1.237 - 1.589	1.25e-07	0.464	1.590	1.345 - 1.880	5.71e-08
ThreeGeneER+/HER2- Low Prolif	-1.510	0.221	0.028 - 1.758	0.154	-2.025	0.132	0.015 - 1.173	0.069
ThreeGeneER+/HER2- High Prolif	-1.853	0.157	0.021 - 1.170	0.071	-2.032	0.131	0.017 - 1.007	0.051
ThreeGeneHER2+	-2.377	0.093	0.011 - 0.761	0.027	-2.472	0.084	0.010 - 0.734	0.025
IntClust2	0.319	1.376	0.817 - 2.317	0.231	0.492	1.636	0.841 - 3.182	0.147
IntClust3	0.630	1.877	1.042 - 3.382	0.036	0.590	1.804	0.797 - 4.079	0.157
IntClust4ER+	0.290	1.337	0.676 - 2.643	0.404	-0.038	0.963	0.328 - 2.830	0.945
IntClust5	1.004	2.729	1.439 - 5.172	0.002	1.197	3.309	1.534 - 7.139	0.002
IntClust6	0.187	1.205	0.730 - 1.991	0.466	0.360	1.433	0.756 - 2.717	0.270
IntClust7	0.309	1.362	0.757 - 2.447	0.302	0.390	1.477	0.657 - 3.320	0.345
IntClust8	0.457	1.579	1.014 - 2.458	0.043	0.452	1.571	0.868 - 2.845	0.136
IntClust9	0.319	1.376	0.878 - 2.155	0.164	0.244	1.277	0.699 - 2.332	0.426
IntClust10	-0.534	0.587	0.245 - 1.405	0.231	-0.168	0.845	0.284 - 2.513	0.762
Likelihood Ratio Test P-value = 7e-06					Likelihood Ratio Test P-value = 2e-06			
Wald Test P-value = 2e-06					Wald Test P-value = 5e-07			
Score (logrank) Test P-value = 9e-07					Score (logrank) Test P-value = 7e-08			

3.6 Luminal A Grade 3

Supplementary Table 16: Univariate Cox regression on clinical variables and CNA quartiles associated with METABRIC lumA grade 3 cases. Rows where univariate Cox regression had significant p-value for OS or DSS are highlighted. Note that both the Wald test p-values and likelihood ratio test p-values were used to determine significance. This was due to some factor level sample sizes being sufficiently small.

Clinical Variables	Overall Survival				Disease-Specific Survival			
	Beta	Hazard Ratio	Hazard Ratio CI	P-Values Wald Test / LRT	Beta	Hazard Ratio	Hazard Ratio CI	P-Values Wald Test / LRT
Positive Lymph Nodes	0.128	1.137	1.090 - 1.186	2.98e-09 / 2.96e-07	0.144	1.155	1.096 - 1.217	6.68e-08 / 4.40e-06
NPI	0.435	1.546	1.224 - 1.952	2.59e-04 / 3.53e-04	0.635	1.887	1.358 - 2.622	1.54e-04 / 1.76e-04
Cellularity:	-	-	-	0.492 / 0.495	-	-	-	0.779 / 0.731
Low	Ref	Ref	Ref	Ref	Ref	Ref	Ref	Ref
Moderate	-0.033	0.967	0.385 - 2.433	0.943	0.689	1.991	0.270 - 14.698	0.500
High	0.214	1.238	0.490 - 3.129	0.651	0.723	2.061	0.277 - 15.331	0.480
ER IHC:	-	-	-	0.002 / 0.014	-	-	-	1.2e-04 / 0.003
Negative	Ref	Ref	Ref	Ref	Ref	Ref	Ref	Ref
Positive	-1.894	0.150	0.046 - 0.495	0.002	-2.431	0.088	0.025 - 0.304	1.2e-04

HER2 SNP6:	-	-	-	0.944 / 0.572	-	-	-	0.732 / 0.560
Neutral	Ref	Ref	Ref	Ref	Ref	Ref	Ref	Ref
Loss	-0.025	0.975	0.357 - 2.665	0.961	0.444	1.558	0.374 - 6.484	0.542
Gain	0.172	1.188	0.683 - 2.066	0.542	0.377	1.459	0.703 - 3.024	0.310
Undefined	-14.996	3.07e-07	0.000 - Inf	0.994	-14.958	3.19e-07	0.000 - Inf	0.996
Inferred Menopausal State:	-	-	-	8.31e-04 / 1.1e-04	-	-	-	0.267 / 0.248
Pre	Ref	Ref	Ref	Ref	Ref	Ref	Ref	Ref
Post	1.115	3.050	1.586 - 5.864	8.31e-04	0.409	1.505	0.731 - 3.102	0.267
PAM50 Subtype	-	-	-	-	-	-	-	-
Three Gene Classification:	-	-	-	0.506 / 0.512	-	-	-	0.789 / 0.791
ER-/HER2-	-	-	-	-	-	-	-	-
ER+/HER2- Low Prolif	Ref	Ref	Ref	Ref	Ref	Ref	Ref	Ref
ER+/HER2- High Prolif	0.210	1.234	0.803 - 1.896	0.338	0.179	1.196	0.660 - 2.169	0.555
HER2+	0.379	1.461	0.623 - 3.428	0.383	0.291	1.338	0.403 - 4.447	0.634
Integrative Clusters:	-	-	-	0.270 / 0.232	-	-	-	0.615 / 0.625
1	Ref	Ref	Ref	Ref	Ref	Ref	Ref	Ref
2	-0.410	0.663	0.093 - 4.717	0.682	0.290	1.337	0.121 - 14.776	0.813
3	-0.067	0.936	0.223 - 3.934	0.928	-0.157	0.855	0.111 - 6.582	0.880
4ER-	-14.365	5.77e-07	0.000 - Inf	0.995	-14.301	6.15e-07	0.000 - Inf	0.996
4ER+	-0.799	0.450	0.098 - 2.056	0.303	-0.357	0.700	0.085 - 5.753	0.740
5	0.137	1.147	0.237 - 5.564	0.865	0.096	1.101	0.114 - 10.664	0.934
6	-0.078	0.925	0.191 - 4.487	0.923	0.205	1.227	0.136 - 11.055	0.855
7	-0.410	0.664	0.151 - 2.909	0.586	-0.231	0.794	0.098 - 6.430	0.829
8	-0.674	0.510	0.117 - 2.229	0.371	-0.738	0.478	0.058 - 3.933	0.493
9	0.366	1.441	0.299 - 6.956	0.649	0.719	2.053	0.239 - 17.649	0.512
10	-	-	-	-	-	-	-	-
Tumour Grade:	-	-	-	-	-	-	-	-
1	-	-	-	-	-	-	-	-
2	-	-	-	-	-	-	-	-
3	-	-	-	-	-	-	-	-
Laterality:	-	-	-	0.854 / 0.854	-	-	-	0.380 / 0.376
Left	Ref	Ref	Ref	Ref	Ref	Ref	Ref	Ref
Right	0.037	1.038	0.700 - 1.538	0.854	-0.254	0.775	0.440 - 1.368	0.380
HER2 Status:	-	-	-	0.545 / 0.558	-	-	-	0.890 / 0.891
Negative	Ref	Ref	Ref	Ref	Ref	Ref	Ref	Ref
Positive	0.238	1.268	0.587 - 2.739	0.545	0.083	1.086	0.337 - 3.498	0.890
ER Status:	-	-	-	0.995 / 0.472	-	-	-	0.996 / 0.588
Negative	Ref	Ref	Ref	Ref	Ref	Ref	Ref	Ref
Positive	14.011	1.216e+06	0.000 - Inf	0.995	14.011	1.216e+06	0.000 - Inf	0.996
PR Status:	-	-	-	0.009 / 0.011	-	-	-	0.008 / 0.011
Negative	Ref	Ref	Ref	Ref	Ref	Ref	Ref	Ref
Positive	-0.548	0.578	0.383 - 0.872	0.009	-0.762	0.467	0.265 - 0.823	0.008
Cancer Type Detailed:	-	-	-	0.098 / 0.114	-	-	-	0.253 / 0.278
Breast	Ref	Ref	Ref	Ref	Ref	Ref	Ref	Ref
BI Ductal Carcinoma	0.397	1.487	0.203 - 10.882	0.696	-0.519	0.595	0.078 - 4.553	0.617
BI Lobular Carcinoma	0.978	2.658	0.341 - 20.716	0.351	-0.002	0.998	0.115 - 8.670	0.999
BI Mixed Mucinous Carcinoma	-	-	-	-	-	-	-	-
Mixed Ductal and Lobular Carcinoma	0.865	2.376	0.315 - 17.911	0.401	0.103	1.108	0.139 - 8.821	0.923
Chemotherapy:	-	-	-	0.638 / 0.644	-	-	-	0.105 / 0.127
No	Ref	Ref	Ref	Ref	Ref	Ref	Ref	Ref
Yes	0.153	1.166	0.615 - 2.208	0.638	0.610	1.841	0.881 - 3.847	0.105
Hormone Therapy:	-	-	-	0.142 / 0.131	-	-	-	0.141 / 0.120
No	Ref	Ref	Ref	Ref	Ref	Ref	Ref	Ref
Yes	0.352	1.422	0.889 - 2.275	0.142	0.545	1.724	0.835 - 3.560	0.141
Radiotherapy:	-	-	-	0.020 / 0.021	-	-	-	0.139 / 0.142
No	Ref	Ref	Ref	Ref	Ref	Ref	Ref	Ref
Yes	-0.464	0.629	0.425 - 0.930	0.020	-0.421	0.657	0.376 - 1.146	0.139
Breast Surgery:	-	-	-	0.003 / 0.002	-	-	-	0.034 / 0.024
Lumpectomy	Ref	Ref	Ref	Ref	Ref	Ref	Ref	Ref
Mastectomy	0.687	1.988	1.256 - 3.148	0.003	0.725	2.066	1.057 - 4.037	0.034
Histological Subtype:	-	-	-	0.098 / 0.114	-	-	-	0.253 / 0.278
Ductal/No Special Type	Ref	Ref	Ref	Ref	Ref	Ref	Ref	Ref
Lobular	0.581	1.787	0.983 - 3.249	0.057	0.517	1.677	0.698 - 4.026	0.247
Medullary	-	-	-	-	-	-	-	-
Mixed	0.468	1.598	0.977 - 2.611	0.062	0.621	1.861	0.958 - 3.617	0.067
Mucinous	-	-	-	-	-	-	-	-
Other	-0.397	0.672	0.092 - 4.921	0.696	0.519	1.680	0.220 - 12.843	0.617
Tubular/ciribriform	-	-	-	-	-	-	-	-
Tumour Stage:	-	-	-	0.008 / 0.007	-	-	-	0.003 / 0.004
0	-	-	-	-	-	-	-	-
1	Ref	Ref	Ref	Ref	Ref	Ref	Ref	Ref
2	0.674	1.961	1.134 - 3.391	0.016	0.944	2.569	1.062 - 6.214	0.036
3	1.359	3.892	1.519 - 9.972	0.005	2.074	7.955	2.395 - 26.418	7.09e-04
4	-	-	-	-	-	-	-	-
Tumour Size	0.023	1.023	1.012 - 1.034	3.73e-05 / 5.56e-04	0.029	1.029	1.017 - 1.041	1.29e-06 / 1.56e-04
Age at Diagnosis	0.057	1.058	1.038 - 1.079	1.46e-08 / 3.20e-09	0.018	1.018	0.993 - 1.044	0.161 / 0.157
Quartiles:	-	-	-	0.095 / 0.111	-	-	-	0.025 / 0.019
1	Ref	Ref	Ref	Ref	Ref	Ref	Ref	Ref
2	-0.095	0.909	0.518 - 1.597	0.741	0.293	1.340	0.537 - 3.346	0.530
3	0.137	1.147	0.658 - 2.001	0.629	0.919	2.508	1.068 - 5.890	0.035
4	0.550	1.734	1.014 - 2.964	0.044	1.158	3.182	1.345 - 7.528	0.008
Ref: Factor level used as baseline; CI: Confidence interval; LRT: Likelihood Ratio Test; NPI: Nottingham Prognostic Index ; HER2: Human epidermal growth factor receptor 2; ER: Estrogen receptor; ER IHC: Estrogen receptor immunohistochemistry; PR: Progesterone receptor; BI: Breast Invasive								

Supplementary Table 17: Assessment of association between CNA quartiles and clinical variables that showed a significant association with either OS, DSS or both in METABRIC lumA grade cases. Rows where p-value was significant are highlighted.

Clinical Variables	Chi-Squared Test	Fisher's Exact Test	Kruskal-Wallis Test
Positive Lymph Nodes	-	-	0.358
NPI	-	-	0.223
ER IHC	1*	1	-
Inferred Menopausal State	0.176	0.181	-
PR Status	0.282	0.297	-
Radiotherapy	0.615	0.621	-
Breast Surgery	0.153	0.155	-
Tumour Stage	0.232*	0.238	-
Tumour Size	-	-	0.239
Age at Diagnosis	-	-	0.024
* Chi-Squared approximation may be incorrect ' Simulated P-value for Fisher's Exact Test NPI: Nottingham Prognostic Index; ER IHC: Estrogen receptor immunohistochemistry; PR: Progesterone receptor			

Supplementary Table 18: Multivariate Cox regression on CNA quartiles and associated clinical variables in METABRIC lumA grade 3 cases. Significant results for OS and DSS are highlighted.

Variables	Overall Survival				Disease-Specific Survival			
	Beta	Hazard Ratio	Hazard Ratio CI	P-value	Beta	Hazard Ratio	Hazard Ratio CI	P-value
Quartile2	0.378	1.459	0.805 - 2.643	0.213	0.435	1.544	0.606 - 3.937	0.363
Quartile3	0.609	1.839	1.025 - 3.300	0.041	1.065	2.902	1.207 - 6.977	0.017
Quartile4	0.670	1.954	1.135 - 3.365	0.016	1.168	3.215	1.357 - 7.619	0.008
Age at Diagnosis	0.063	1.065	1.042 - 1.088	1.54e-08	0.021	1.021	0.994 - 1.050	0.129
	Likelihood Ratio Test P-value = 2e-08 Wald Test P-value = 2e-07 Score (logrank) Test P-value = 7e-08				Likelihood Ratio Test P-value = 0.02 Wald Test P-value = 0.02 Score (logrank) Test P-value = 0.02			

Supplementary Table 19: DSS Multivariate Cox regression on CNA quartiles in METABRIC lumA grade 3 cases. Significant results for OS and DSS are highlighted.

Variables	Disease-Specific Survival			
	Beta	Hazard Ratio	Hazard Ratio CI	P-value
Quartile2	0.293	1.340	0.537 - 3.346	0.530
Quartile3	0.919	2.508	1.068 - 5.890	0.035
Quartile4	1.158	3.182	1.345 - 7.528	0.008
	Likelihood Ratio Test P-value = 0.02 Wald Test P-value = 0.02 Score (logrank) Test P-value = 0.02			

References

- [1] I. Tishchenko, H. H. Milioli, C. Riveros, and P. Moscato. Extensive Transcriptomic and Genomic Analysis Provides New Insights about Luminal Breast Cancers. *PLoS ONE*, 11:e0158259, 2016.
- [2] F. Tonello, A. Bergmann, K. de Souza Abrahao, S. S. de Aguiar, M. A. Bello, and L. C. S. Thuler. Impact of Number of Positive Lymph Nodes and Lymph Node Ratio on Survival of Women with Node-Positive Breast Cancer. *Eur J Breast Health*, 15(2):76–84, Apr 2019.
- [3] Y. Fong, J. Evans, D. Brook, J. Kenkre, P. Jarvis, and K. Gower-Thomas. The Nottingham Prognostic Index: five- and ten-year data for all-cause survival within a screened population. *Ann R Coll Surg Engl*, 97:137–139, 2015.
- [4] S. Song, K. Nones, D. Miller, I. Harliwong, K. S. Kassahn, M. Pinese, M. Pajic, A. J. Gill, A. L. Johns, M. Anderson, O. Holmes, C. Leonard, D. Taylor, S. Wood, Q. Xu, F. Newell, M. J. Cowley, J. Wu, P. Wilson, L. Fink, A. V. Biankin, N. Waddell, S. M. Grimmond, and J. V. Pearson. qpure: A tool to estimate tumor cellularity from genome-wide single-nucleotide polymorphism profiles. *PLoS ONE*, 7:e45835, 2012.
- [5] C. Curtis, S. P. Shah, S. F. Chin, G. Turashvili, O. M. Rueda, M. J. Dunning, D. Speed, A. G. Lynch, S. Samarajiwa, Y. Yuan, S. Graf, G. Ha, G. Haffari, A. Bashashati, R. Russell, S. McKinney, A. Langerod, A. Green, E. Provenzano, G. Wishart, S. Pinder, P. Watson, F. Markowitz, L. Murphy, I. Ellis, A. Purushotham, A. L. Borresen-Dale, J. D. Brenton, S. Tavaré, C. Caldas, S. Aparicio, C. Caldas, S. Aparicio, C. Curtis, S. P. Shah, C. Caldas, S. Aparicio, J. D. Brenton, I. Ellis, D. Huntsman, S. Pinder, A. Purushotham, L. Murphy, C. Caldas, S. Aparicio, C. Caldas, H. Bardwell, S. F. Chin, C. Curtis, Z. Ding, S. Graf, L. Jones, B. Liu, A. G. Lynch, I. Papatheodorou, S. J. Sammut, G. Wishart, S. Aparicio, S. , K. Gelmon, D. Huntsman, S. McKinney, C. Speers, G. Turashvili, P. Watson, I. Ellis, R. Blamey, A. Green, D. Macmillan, E. Rakha, A. Purushotham, C. Gillett, A. Grigoriadis, S. Pinder, E. de Rinaldis, A. Tutt, L. Murphy, M. Parisien, S. Troup, C. Caldas, S. F. Chin, D. Chan, C. Fielding, A. T. Maia, S. McGuire, M. Osborne, S. M. Sayalero, I. Spiteri, J. Hadfield, S. Aparicio, G. Turashvili, L. Bell, K. Chow, N. Gale, D. Huntsman, M. Kovalik, Y. Ng, L. Prentice, C. Caldas, S. Tavaré, C. Curtis, M. J. Dunning, S. Graf, A. G. Lynch, O. M. Rueda, R. Russell, S. Samarajiwa, D. Speed, F. Markowitz, Y. Yuan, J. D. Brenton, S. Aparicio, S. P. Shah, A. Bashashati, G. Ha, G. Haffari, and S. McKinney. The genomic and transcriptomic architecture of 2,000 breast tumours reveals novel subgroups. *Nature*, 486:346–352, 2012.
- [6] Banyameen. Iqbal and Archana. Buch. Hormone receptor (ER, PR, HER2/neu) status and proliferation index marker (Ki-67) in breast cancers: Their onco-pathological correlation, shortcomings and future trends. *Medical Journal of Dr. D.Y. Patil University*, 9:674–679, 2016.
- [7] A. M. Gown. Current issues in ER and HER2 testing by IHC in breast cancer. *Mod. Pathol.*, 21 Suppl 2:S8–S15, 2008.
- [8] T. V. Hansen, J. Vikesaa, S. S. Buhl, H. H. Rossing, V. Timmermans-Wielenga, and F. C. Nielsen. High-density SNP arrays improve detection of HER2 amplification and polyploidy in breast tumors. *BMC Cancer*, 15:35, 2015.
- [9] S. K. Chia, V. H. Bramwell, D. Tu, L. E. Shepherd, S. Jiang, T. Vickery, E. Mardis, S. Leung, K. Ung, K. I. Pritchard, J. S. Parker, P. S. Bernard, C. M. Perou, M. J. Ellis, and T. O. Nielsen. A 50-gene intrinsic subtype classifier for prognosis and prediction of benefit from adjuvant tamoxifen. *Clin. Cancer Res.*, 18(16):4465–4472, Aug 2012.
- [10] M. H. Amer. Genetic factors and breast cancer laterality. *Cancer Manag Res*, 6:191–203, 2014.

- [11] G. N. Sharma, R. Dave, J. Sanadya, P. Sharma, and K. K. Sharma. Various types and management of breast cancer: an overview. *J Adv Pharm Technol Res*, 1(2):109–126, Apr 2010.
- [12] E. A. Rakha, J. S. Reis-Filho, F. Baehner, D. J. Dabbs, T. Decker, V. Eusebi, S. B. Fox, S. Ichihara, J. Jacquemier, S. R. Lakhani, J. Palacios, A. L. Richardson, S. J. Schnitt, F. C. Schmitt, P. H. Tan, G. M. Tse, S. Badve, and I. O. Ellis. Breast cancer prognostic classification in the molecular era: the role of histological grade. *Breast Cancer Res.*, 12(4):207, 2010.
- [13] S. J. Dawson, O. M. Rueda, S. Aparicio, and C. Caldas. A new genome-driven integrated classification of breast cancer and its implications. *EMBO J.*, 32:617–628, 2013.
- [14] Mahul B Amin, Stephen B Edge, Frederick L Greene, David R Byrd, Robert K Brookland, Mary Kay Washington, Jeffrey E Gershenwald, Carolyn C Compton, Kenneth R Hess, Daniel C Sullivan, J Milburn Jessup, James D Brierley, Lauri E Gaspar, Richard L Schilsky, Charles M Balch, David P Winchester, Elliot A Asare, Martin Madera, Donna M Gress, and Laura R Meyer. *Principles of Cancer Staging*. Springer, New York ;, 7th ed. / edition, 2017.
- [15] N. L. Henry and L. A. Cannon-Albright. Breast cancer histologic subtypes show excess familial clustering. *Cancer*, 2019.
- [16] P. Nygren. What is cancer chemotherapy? *Acta Oncol*, 40(2-3):166–174, 2001.
- [17] R. Baskar, K. A. Lee, R. Yeo, and K. W. Yeoh. Cancer and radiation therapy: current advances and future directions. *Int J Med Sci*, 9(3):193–199, 2012.
- [18] A. Awan and K. Esfahani. Endocrine therapy for breast cancer in the primary care setting. *Curr Oncol*, 25(4):285–291, 08 2018.
- [19] F. McCrate, E. Dicks, E. Powell, J. Chafe, R. Roome, C. Simmonds, and H. Etchegary. Surgical treatment choices for breast cancer in Newfoundland and Labrador: a retrospective cohort study. *Can J Surg*, 61(6):377–384, 12 2018.