January 2016

Molecular And Genetic Properties Of Breast Cancer Associated With Local Immune Activity

Anton Safonov
Yale University, anton.safonov@yale.edu

Molecular and genetic properties of breast cancer associated with local immune activity

A Thesis Submitted to the Yale University School of Medicine
in Partial Fulfillment of the Requirements for the
Degree of Doctor of Medicine

by

Anton Safonov

2016

Abstract

The causes of high versus low, or absent, immune cell infiltration in breast cancer remain unknown. The goals of this analysis were to examine if total mutation load, neoantigen load, copy number variations (CNV), gene-level or pathway level somatic mutations or germline polymorphisms (SNP) are associated with the level of immune infiltration measured by immune metagene expression levels. We used RNA-Seq, DNA copy number, mutation and germline SNP data from the TCGA representing n=627 ER+, n=207 HER2+ and n=191 TNBC cancers. 13 published immune metagenes were used in correlation and multivariate linear regression analyses performed separately for the 3 major clinical subtypes. P-values were adjusted for multiple comparisons and permutation testing was used to assess false discovery rates. Overall mutation, neoantigen and amplification, or deletion loads did not correlate strongly with any of the immune metagenes in any subtype (Spearman coefficient 0.2). In ER+ cancers, mutations in MAP2K4 and TP53 were associated with lower and higher levels of immune infiltration, respectively. In TNBC, mutations in MYH9 and HERC2 were associated with lower immune infiltration. None were found in HER2+ cancers. Three SNPs (rs425757, rs410232, rs470797) in the exonic regions of the FHPR1 and MLP genes were associated with low immune infiltration in ER+ cancers, none in the other subtypes. Two amplicons in TNBC and 3 amplicons in HER2+ cancers were associated with lower immune infiltration. We also identified alterations in several biological pathways that were associated with immune infiltration in different breast cancer subtypes. At the individual patient level, each pathway was affected at different genes through distinct genomic mechanisms. Our results suggest that immune infiltration in breast cancer is not driven by a single global metric of genomic aberrations such as mutation, neoantigen or CNV loads, but by multiple different gene and pathway level associations that each affect small subsets of patients within each subtype.

Acknowledgements:

This work would not have been possible without the patient support and valuable mentorship of Dr. Lajos Pusztai, Dr. Christos Hatzis, and Dr. Giampaolo Bianchini. In addition, the entire Pusztai laboratory has been instrumental in my study of the bioinformatics methods within this thesis. In no particular order, Tingting Jiang, Weiwei Shi, Xiatong Li, James Platt, Vikram Wali, Tomoko Kurita, Bilge Aktas, Marie-Kristin Walde have contributed wonderful thoughts and ideas to this project.

Contents

Introduction

The development of effective new immunotherapies against cancer has rekindled interest in the tumor immune microenvironment. Antibodies directed against CTLA4, PD-1 and PD-L1 immune checkpoint molecules which are expressed on the surface of tumor cells and immune cells have induced durable responses in several different cancer types including metastatic melanoma, lung, renal, head and neck, bladder and breast cancers. [1-5]. However, objective tumor response and durable clinical benefit is only observed in 10-35% of patients and therefore, there is intense interest in identifying molecular predictors of response to these novel immune-targeted drugs. Despite the highly targeted nature of these therapeutic antibodies, surprisingly, all studies have shown objective responses in cancers that have very little or no detectable expression of the targets.

The presence of tumor infiltrating lymphocytes (TIL) in the breast cancer microenvironment has long been recognized as a favorable prognostic marker [6]. More recently it has also become clear that the prognostic value of immune infiltration depends on breast cancer subtype. In triple negative breast cancer (TNBC), characterized by the lack of human epidermal growth factor receptor 2 (HER2), estrogen receptor (ER), and progesterone receptor (PR) expression, high levels of immune infiltration either measured as tumor infiltrating lymphocyte (TIL) count or through the expression of immune cell related genes, predicts for markedly improved survival, even in patients not receiving systemic adjuvant therapy [7] [8]. In TNBC, high levels of immune infiltration are also associated with higher pathologic complete response

rates (i.e. complete eradication of invasive cancer from the breast and lymph nodes) after preoperative (neoadjuvant) chemotherapy [9, 10]. However, only 5-15% of TNBCs represent lymphocyte predominant cancers (defined as lymphocytes constituting >50% of stromal cells), another 15-20% contain no lymphocytes at all, and the rest contain low to moderate immune lymphocytic infiltration. In HER2 positive breast cancer, high TIL count and immune gene expression are also associated with better prognosis, both in the presence and absence of adjuvant chemotherapy with or without trastuzmab [11, 12]. Among ER positive breast cancers with high proliferation rates, immune infiltration also predicts for better prognosis compared to immune cell poor cancers. In contrast, in ER-positive cancers with low proliferation and generally better prognosis, immune cell presence appears to have no clinically meaningful prognostic value [13]. Only two Phase I clinical trials reported results using immune checkpoint inhibitors (anti-PD1 pembroluzimab and anti-PD-L1 MPDL3280A, respectively) in metastatic TNBC. Objective tumor responses were seen in about 20% of cases in both studies but no biomarker results were presented and therefore the relationship between tumor immune infiltration and response to immune checkpoint inhibitor therapy is unknown. An emerging hypothesis, yet to be proven, suggests that immunotherapies may work best in cancers with intermediate to high levels of immune infiltration [14]

An important feature of the tumor immune microenvironment is the diversity of immune cells present. T-lymphocytes are the most abundant cell type in breast cancer tissues (70-80% of total immune cells), followed by B-cells (10-20%), macrophages (5-10%), natural killer and

dendritic cells (~5%). Each of these main cell types consist of multiple functionally distinct subtypes (e.g. CD8+ effector T-cells, CD4+ T helper cells, CD4+ regulatory T cells, memory T-cells, etc..) and the functional activity of these cells changes dynamically in response to various stimuli in the tumor microenvironment. This anatomical feature of the immune milieu explains the highly correlated nature of immune gene expression signatures. For example, the expression of immunoglobulin genes is highly correlated with the expression of T-cell genes, the expression of genes involved in T-cell activation are co-expressed with interferon and NK-cell derived genes, etc. These immune gene signatures represent convenient mRNA expression based surrogates for histological assessment of immune infiltration.

The observation that some breast cancers contain a large number of lymphocytes and have better prognosis and others contain few, or none at all, is not well understood. It is well established in the immunology literature [15] that cytotoxic T cells can recognize single amino acid changes in antigens [16, 17]. In melanoma and lung cancer, it has been suggested that immunogenicity of a tumor, and the extent of lymphocytic infiltration, may be driven by total number of somatic mutations (i.e. mutation load) or the number of new antigen epitopes due to somatic mutations in the cancer (i.e. neoantigen load) [18]. Computational tools were developed to assign antigenicity to non-synonymous mutations in protein coding genes, and these can be used to calculate the number of putative neoantigens in a given cancer [19]. However, the association between these genome wide mutational metrics and immune infiltration of the tumor may be cancer type specific.

One previous study that examined many different types of cancers showed no correlation between overall mutation load and immune infiltration measured by immune gene expression. This study also revealed only a weak association between neoantigen load and immune infiltration in breast cancer. Other types of genetic aberrations may also influence the immunogenicity of breast cancer. For instance, copy number alterations are thought to account for the vast majority of the variation in gene expression observed in cancers [20]. Indeed, other investigators have demonstrated that certain copy number alterations are associated with immunogenicity of hepatocellular carcinoma [21], lung squamous cell carcinoma [22], and colorectal cancer [23]

One of the most comprehensive undertakings to date involved analysis of 18 tumor types with an expression-based metagene representing cytolytic activity ("CTL"), by cleverly implementing the geometric mean of the genes encoding perforin (PRF1) and granzyme A (GZMA), direct cytotoxic agents employed by CD8+ T cells [15]. In addition to the neoantigen calculations stated above, the authors utilized a regression-based approach to identify several potential somatic mutations and sites of copy number alterations significantly associated with cytolytic activity metagene expression in breast cancer. However, the method of encountering breast cancer-specific drivers of immune response by the precedent of pan-cancer significance may miss specific genomic events which are specific to breast cancer.

Statement of Purpose

The goal of this study was to systematically examine if DNA level genomic alterations are associated with immune cell infiltration measured by immune metagene expression in the three distinct clinically relevant breast cancer subtypes. We tested the hypotheses that either (i) total mutation load, (ii) neoantigen load, (iii) copy number variations (CNV), (iv) gene-level or (v) biological pathway level somatic mutations, or (vi) germline polymorphisms (SNP) are associated with the levels of immune gene expression. These analyses were performed separately for TNBC, HER2 positive and ER positive breast cancers. While the total mutation and neoantigen loads have previously been correlated with immune cell activity in all breast cancers combined, comprehensive analysis of gene-level mutations, biological pathway level alterations and DNA copy number changes in breast cancer subtypes have not yet been reported. Additionally, we extend our analysis to examine whether somatic mutation, germline SNP, and copy number variation in specific genomic locations and pathways are associated with immune response.

Methods

Data

We obtained gene-level RNA-Seq expression data and corresponding clinical information from The Cancer Genome Atlas (TCGA) obtained through the public access web portal (tcga-data.nci.nih.gov/tcga/dataAccessMatrix.htm). All cancer specimens included in our analysis were collected before any systemic therapy. RNA-Seq reads were mapped to exonic splice

junctions using MapSlice [24], and quantitation based on read mappings was performed by RNA-Seq by Expectation Maximization (RSEM) as previously reported [25]. The output was formatted as the estimated fraction of transcripts using the transcripts per kilobase million (TPM) values as the measures of gene expression, which were log2 normalized for further analysis.

We obtained individual gene-level variant data (MAF file) (n=817) from a recent publication by Ciriello et al . The data set represents high confidence somatic single nucleotide variants (SNVs), insertions and deletions (indels) from the TCGA after the following processing steps: (i) inclusion of variants previously filtered out with dbSNP-based filter at the time of first release of the TCGA data, (ii) point mutations called by integration of RNA-Seq data and DNA sequencing results [26], (iii) indels called by an improved assembly-based realignment tool [27, 28], (iv) calls with fewer than 8 reads in tumor or normal were removed, (v) calls with high normal variant allele fraction (VAF), (vi) calls with minimal reads supporting the variant (< 2 DNA and RNA reads combined), and (vi) calls with low combined DNA and RNA VAF (and therefore, low precision ([29], [30],[31]]) were also removed. In our downstream analysis we only considered somatic mutations which affected genes expressed in the given cancer (greater than > 1 TPM) and were classified as non-silent.

We obtained GISTIC2 [32] "Level 4" copy number calls for each patient from the TCGA public access portal. Data preprocessing was previously published by the TCGA Breast Cancer Workgroup [33]. Briefly, copy number value was estimated by fitting each sample to discrete

copy number classes by Gaussian mixture modeling, and a SNP locus is iteratively informed by an expectation-maximization algorithm Circular Binary Segmentation is subsequently performed to assign these locus values into biologically relevant contiguous segments; the TCGA data uses a Circular Binary Segmentation algorithm to recursively identify segments with probe distribution different from neighbors [34]. We assigned DNA segments to copy number alteration categories based on GISTIC threshold scores. We considered GISTIC scores -2 and +2 (i.e. log2-transformed either less than -1 or greater than 1, respectively) as evidence for define deletion or amplification, respectively. Values of -1 and +1 were considered as possible deletions or amplifications (log2-transformed less than -0.3 or greater than 0.3, corresponding to noise-level) and values of 0 as low likelihood of copy number event.

We also obtained germline single nucleotide polymorphisms (SNP) data for the same cohort through the TCGA (dbGaP) authorized access portal. The TCGA cohort included 18,585,361 germline variants. We filtered out variants in duplicated genes reported in the Duplicated Genes Database (DGD) [35] and also rare variants with frequency <2%, which resulted in 5,853,796 SNPs for further analysis. Next we calculated allelic imbalance and variants that deviated from the Hardy-Weinberg Equilibrium were removed. The remaining 1,646,930 SNPs were classified as moderate (i.e. inframe indels or missense variants in coding genes or regulatory region ablation) or high functional impact (predicted transcript ablations, splice variant, stop-gain mutations, frameshift mutations, stop/start-loss). SNPs that were not predicted to be moderate or

high functional impact were not included in further downstream analysis, resulting in 8,861 studied SNPs.

Breast Cancer Clinical Subtype Assignment

Three clinically relevant breast cancer subtypes were defined as (i) ER positive and HER2 negative (ii) HER2 positive (with any ER) and (iii) ER and HER2 negative (broadly corresponding to triple negative cancers). This method of classification was chosen over PAM50 molecular subtypes because of its ready clinical applicability and to maintain consistency during analysis of gene expression signatures with previous studies [13]. RNAseq data of 1066 breast cancers were downloaded from the TCGA. Of these cases, 892 had routine pathology HER2 results available (160 positive, 545 negative, 11 indeterminate, 176 equivocal, 174 not evaluated). Histological ER status, determined by immunohistochemistry, was available for 1003 samples. To maximize sample size, we assigned HER2 and ER status to cases with missing or equivocal clinical information using HER2 and ER mRNA levels [Gong]. Receiver operating characteristic curves were constructed with the aid of the 'pROC' package in R [36] for varying thresholds of ERBB2 and ESR1 gene expression, using the histological result as measure of true receptor status. For ERBB2 mRNA expression, the 95% confidence interval of optimal cutpoints was calculated from 10,000 bootstrap replicates [37], and when ERBB2 expression was < 7.86 TPM the case was considered negative, when it was > 8.52 it was considered positive. Samples between 7.86 and 8.53 were considered indeterminate, and were excluded from further analysis (n = 51). At the end

197 samples were defined as HER2 positive and 818 samples as HER2 negative. The ESR1 mRNA expression showed a strong bimodal distribution, allowing calculation of a single optimal threshold of the ROC curve. This was calculated using the Youden's J statistic (which identifies the maximum of the distance from the identity line) that yielded 4.35 TPM as the threshold to define ER positive versus negative cases based on mRNA expression. This resulted in 243 ER negative cases and 823 ER positive cases. The final sample size for this study is N=627 ER positive/HER2 negative cases, N=207 HER2 positive cases (including 47 ER- and 105 ER+ cancers) and N=191 ER and HER2 negative cases (largely corresponding to triple negative breast cancer).

Assignment of Immune Groups

For each sample, 13 previously reported immune gene expression metagenes were calculated[13, 15, 38] .The metagene values are the arithmetic mean of log2-transformed expression values of the genes in each metagene (**Table 1).** The genes included in each metagene represent genes which are highly co-expressed and correspond to various immune cell subtypes and immune functions. The prognostic and chemotherapy response predictive values of T-cell related immune signatures have been explored [13, 15, 38]. We, as well as multiple other investigators, have noted the very highly correlated co-expression of most immune metagenes. For correlation with DNA-level genomic variables, we used the expression level of the "most

representative immune metagene" as the measure of the extent of immune infiltration. The most representative metagene was defined as the metagene that showed the highest average correlation with each of the other metagenes based on a correlation matrix, aided by the 'corr' function from the R package 'Hmisc' [39]. Using this continuous variable, we also created three categories of cancers including (i) immune-poor, (ii) immune-intermediate and (iii) immune-rich corresponding to the tertiles of the metagene expression distribution.

Association between overall mutation load, neoantigen load, DNA copy number alterations and immune infiltration

Overall mutation load was defined using the number of somatic mutations in a sample $N_{somatic}$, and the length of Refseq exonic region with adequate read coverage, $L_{cov>=16}$

$$\text{Mutation Load} = \text{ML} = \frac{N_{somatic}}{L_{cov \geq 16}} \times 10^6$$

Neoantigen load information was taken from a previously published paper by Rooney et al that has calculated neoantigen load for each of the samples and is publically available (need reference). This data was generated using a neoantigen prediction pipeline [Rajasag], specific for each individual's HLA type (as predicted by the POLYSOLVER pipeline [Shukla].

Overall DNA deletion load was defined as the number of definitely deleted genes (GISTIC threshold value of -2), and amplification load was defined as the number of definitely amplified genes (GISTIC threshold value of 2) per sample [Mermel]. Correlation coefficients of mutation

load, neoantigen load, amplification and deletion loads were calculated for each immune metagene as continuous variables. We also performed a linear regression analysis of each immune metagene on each global DNA alteration metrics. Significance was tested by an ANOVA Chi-squared test on each respective variable.

To identify patterns of variation in the dataset of somatic mutation and copy number reads, we performed a principal component analysis within ER+/HER2-, HER2+, and ER-/HER2- breast cancers. The principal component analysis was conducted on the highest and lowest most representative immune metagene tertiles, using the built-in "prcomp" function in R and median, maximum absolute difference (MAD) scaled values.

Gene-level Mutational Analysis

To assess correlation between immune metagene expression and somatic mutations in individual genes, we performed gene-specific linear regression, with mutation load and histological diagnosis as patient-level covariates. Multiple hypothesis correction was performed using an empirical false-discovery rate cutoff of 10%.

Gene-Level Copy Number Analysis

We performed linear regression of log-normalized immune metagene expression level with GISTIC threshold score as the independent variable. This analysis was performed separately for amplified and deleted regions. The amplification-centric analysis excluded regions with negative

GISTIC scores and deletion-centric analysis excluded regions with positive scores. In the former, the overall amplification load was included as a covariate, and for the deletion-centric analysis, overall deletion load was used as covariate. We included histological diagnosis (i.e. invasive lobular versus ductal histology) as a covariate in both types of analyses.

The extensive auto-correlation of associations, due to the tendency of copy number alterations to affect many genes, precluded standard multiple-hypothesis correction for individual gene level analysis. We estimated significance through permutation testing. We first identified continuous segments of genes with nominal p values < 0.01, we refer to these as gene peaks, and describe each gene peak by minimum p-value of genes included in the peak as the significance score. To obtain the null distribution of gene peak significance scores, the most representative metagene value was permuted 500 times. Peak scores were generated for each permutation and the quantile of each true peak within this peak score null distribution was assigned as the adjusted p-value. These adjusted p-values were adjusted based on an empirical false-discovery rate.

Germline SNP Analysis

Association between SNPs and log-normalized immune metagene expression levels were examined using linear regression analysis including histological diagnosis as a covariate. Significance was tested with the likelihood-ratio test between the "null" model including histological diagnosis alone and the "alternative" model of the SNP status variable, with

histological diagnosis as the covariate. False discovery rate was calculated by the Benjamini-Hochberg method.

Combined effect of genomic alterations on immune metagene expression levels

To examine the relative contribution of the different genomic alterations as explanatory variables for immune infiltration, we constructed a multivariable regression model of the most representative immune metagene (LCK metagene) expression and deletion, amplification, and somatic mutation status as the independent variables. Histological diagnosis was included as a categorical covariate. This regression was performed for each gene and covariates included expression, copy number alteration, and somatic mutation data. Null distributions of each regression coefficient were computed by permuting the LCK metagene and refitting the regression model over 5000 random iterations. The observed coefficient was compared with the corresponding null distribution and was considered significant if it was > 95% percentile of the null distribution. For each breast cancer subtype, a Venn diagram was constructed to describe the number of genes with putatively significant associations of expression, deletion, amplification, and mutation with immune activity.

Pathway-Level Analysis

Since the functional output of a biological pathway, defined as a set of expressed genes that collectively contribute to a given function, may be altered by genomic changes in any of its member genes, and the genomic changes may include somatic mutations, gene amplification or

deletion, we also examined the association between immune metagene expression and pathway level alterations. We have assembled 714 canonical biological pathways from the NCI Pathway Interaction and BioCarta Pathway databases that correspond to most known biological functions [Pusztai NeoALTTO citation]. For each pathway, we defined an "aberration ratio score" calculated as the number of genes affected by mutation or copy number change (i.e. GISTIC score of 2 or -2), divided by the total number of genes in the pathway. We examined the association between immune metagene levels and pathway aberration score using linear regression including the histological diagnosis as covariate. The same gene is often included in multiple different pathways and therefore pathway-specific p-values cannot be adjusted for multiple comparisons assuming from independent observations. To calculate significance, we constructed random gene sets with the same number of genes as a given pathway, and tested the observed aberrations in these random sets with immune metagene expression in 1000 iterations. The coefficients were compiled into a null distribution for each pathway and the observed coefficient was compared with this null distribution. A coefficient was considered significant if it was > 95% percentile of the null distribution.

Results

Immune metagene expression distribution and co-expression correlations

The expression distribution of each of the 13 immune metagenes in the 3 breast cancer subtypes is shown on **Figures 1A-C[38]**. Metagenes describing IF1, macrophage, MHC1, MHC2, STAT1, T follicular cells, T cell inhibitory and stimulatory activity, as well as lymphocyte-specific kinase (LCK), cytolytic activity (CTL), and consensus T-cell metagene (CTM/Giam) had unimodal normal distributions. Metagenes describing natural killer (NK) cells, and regulatory T-cells (T-regs) showed bimodal distributions. Most immune metagenes showed high levels of co-expression in all 3 breast cancer subtypes except the Treg signature which had low correlation with all others (**Figures 2A-C)**, despite no overlap in member genes. The LCK metagene showed the highest degree of auto-correlation with other metagenes in all subtypes. This metagene was also shown to strongly correlate with the presence of T-lymphocytes on histologic evaluation in breast cancer samples [38]. For these reasons, we selected the LCK metagene as the most representative immune metagene and as our measure of immune infiltration. We used the highest expression tertile of the LCK metagene within each subtype to define immune-rich and the lowest tertile to define immune-poor cancers. For metagenes initially derived using microarray analysis,

correlation between RNA-Seq and Affymetrix H133A expression was demonstrated to be high, for each member gene in an available TCGA lung carcinoma dataset (data not shown).

Correlation between immune metagene expression levels and overall mutation and neoantigen loads and DNA copy number variations

Next, we examined correlation between the expression levels of each of the 13 immune metagenes and four different measures of global genomic aberrations including (i) overall mutation load, (ii) neoantigen load, (iii) DNA segment amplification load and (iv) DNA segment deletion load. We used the Spearman rank correlation because it is less sensitive than the Pearson correlation to strong outliers. **Figures 3A-C** show the results for each of the 3 breast cancer subtypes. We only detected very weak correlations between any of the immune metagenes and the 4 global genomic aberrations metrics. The highest Spearman correlation was 0.21, between the STAT metagene and mutation load in ER+/HER2- samples. ANOVA Chi-Square significance testing revealed no significant linear regression coefficients between the metagenes and the 4 different types of global genomic aberration metrics. We also performed principal component analysis **(Figures 4A-I),** which also failed to reveal any separation by global genomic measures between cases assigned to high vs low LCK metagene tertiles. For HER2+ cancers, the first principal component was responsible for explaining 40% of the variation in the data, and aligned most closely with overall deletion and mutation loads. For the ER-/HER2- cancers, 48% of the variation was explained by overall amplifications and deletion loads. For ER+/HER2- cancers,

41% of the variation was attributed to overall amplifications and deletion loads. Overall these results indicate that immune infiltration in breast cancer is not driven by a single global metric of genomic aberrations such as mutation or neoantigen load or copy number variation. Even when considered together, these anomalies account for a minority of the variation in immune metagene expression across cases within clinically important subtypes.

Association between single gene-level somatic mutations, germline SNPs and LCK metagene expression

Somatic mutations were detected for 379 ER+/HER2- , 119 HER2+, 107 ER-/HER2- cases including a total number of 14,440 mutated genes; it should be noted that these totals represent available somatic data from Ciriello et al; mutation analysis was performed for fewer samples than the RNASeq expression data directly available from TCGA.

After filtering somatic mutations for > 3% frequency in each patient cohort, we had 188 and 104 mutated genes in the ER+/HER2- and HER2+ cohorts, respectively and 37 mutated genes in the ER-/HER2- cohort. Among the ER+/HER2- cancers mutations in 6 genes had nominally significant ($p<0.05$) association with the LCK metagene expression but only 2 genes, MAP2K4 and TP53 remained significant after correction for multiple hypothesis testing. Mutations in MAP2K4 were associated with lower levels of immune infiltration and mutations in TP53 with higher levels (**Figure 5A**). Among the ER-/HER2- cancers, mutations in 7 genes had nominally significant association with immune metagene expression and 2, MYH9 and HERC2 remained

significant after correction for multiple hypothesis testing (**Figure 5B**). Mutations in both genes were associated with lower levels of immune infiltration in these cancers. Among HER2+ cancers, mutations 11 genes had nominal association with immune metagenes but none remained significant after multiple hypothesis correction **(Figure 5C).**

Germline SNP data were available for 239 ER+/HER2-, 80 HER2+, 68 ER-/HER2- cases. It should be noted that these totals represent available germline data accessed from the original breast cancer TCGA cohort, which includes fewer samples than the expression data processed by RNASeq or the somatic mutation calls. In the ER+/HER2- and HER2+ cohorts, 446 and 460 SNPs had nominally significant p-values but none remained significant after correction for multiple hypotheses testing**.** Linear regression yielded 361 nominally significant associations in the ER-HER2- cohort, with only 3 SNPs (rs425757, rs410232, rs470797) meeting a false discovery rate threshold of 10%. These 3 SNPs are located in the coding regions of FHPR1 (rs425757 and rs410232) and MLP genes (rs470797), and all 3 were associated with low LCK metagene expression (**Figure 6).**

Association between DNA copy number deletions and amplifications and LCK metagene expression

In ER-/HER2- cancers, the deletion-centric analysis resulted in 96 peaks with nominally significant p-values and 4 remained significant after comparison to null gene peak distribution. After further adjustment for multiple hypotheses testing none remained significant. The

amplification-centric analysis resulted in 49 initial peaks, with 7 significant peaks after comparison to null peak distribution. Two of these, amplifications of 5p12-14.3 and 17q11-24, remained significant after adjustment for multiple hypotheses testing. Both amplifications were associated with lower immune infiltration measured by LCK metagene expression (**Figure 7A**). 0.8-23.7% of ER-/HER2- patients had an amplification in the former region, while 0-47.6% of ER-/HER2- patients had an amplification in the latter. This range reflects the uncertainty of copy number variations defined by GISTIC scores of -1/1. The lower bound of the range corresponds to the proportion of patients that had definite copy number alterations (GISTIC scores of -2 or 2), while the upper bound represents the proportion GISTIC scores above noise level (GISTIC scores of -1 and -2 or 1 and 2)

In HER2+ cancers, the deletion-centric analysis yielded 91 initial peaks, with 4 remaining significant after comparison to null peak distribution but none after adjustment for multiple hypothesis testing. The amplification-centric analysis yielded 43 initial peaks, with 4 significant after comparison to null peak distribution. Three of these remained significant after adjustment for multiple hypothesis testing (1q21-23.1, 1q24-32.1, 17q21.2-32) and each were associated with lower immune infiltration (**Figure 7B**).

For ER+/HER2- cancers, the deletion-centric analysis yielded 77 initial peaks, with 10 significant after comparison to null peak distribution. The amplification-centric analysis yielded 58 initial peaks, with 4 significant peaks after comparison to null peak distribution. **(Figure 7C)**

Multivariate assessment of the contribution of mutations, copy number alterations and expressions at gene level with LCK metagene expression

Next we combined gene level mutation data, gene level copy number alteration, and gene level expression data which resulted in 11,440 common genes between the datasets. **Figures 8A-F**, ER+/HER2- cases demonstrate more frequent association (positive and negative) of expression with immune response. Interestingly, in HER2+ and ER-/HER2- cases, higher immune infiltration is more closely associated with expression, while decreased immune infiltration is more frequently associated with deletions and amplifications.

Association between biological pathway level alterations and LCK metagene expression

We considered a pathway affected by genomic alterations if any of its member genes had a somatic mutation or copy number change as described in the methods section. After merging the available copy number and mutation data we had 474 ER+/HER2-, 149 HER2+, 145 ER-/HER2- cases.

In ER+/HER2- samples, aberrations in 77 pathways exhibited putative significance as compared to the null distribution of regression coefficients. Of these, 12 met the < 10% false discovery rate criteria **(Figure 9A).** Alterations in 11 pathways were associated with decreased LCK metagene expression, and one was associated with higher expression. Eight pathways had

shared aberrations in several members of the MAP-kinase family (MAP3K1, MAPK8, MAP2K4, MAPK1, MAPK3, MAP2K1, MAPK14, MAP2K3) **(Figure 9D)**.

In ER-/HER2- samples, aberrations in 44 pathways exhibited putative significant as compared to null distribution. Of these, 6 met the 10% false discovery rate criteria ("Regulation of beta-catenin", "Calcium signaling in the CD4 TCR pathway", "IL1R". "Validated Transcriptional Targets of Fra1 and Fra2", "Stabizilation of the E-Cadherens Junction", and "FGF signaling pathway") (**Figure 9B)**. These pathways were all associated with a decreased immune infiltration and shared aberrations in common genes including JUN, IL6, IL8. **Figure 9E** represents the degree of overlap of member genes in each significant pathway.

In HER2+ cancers, aberrations in 18 pathways were putatively significant (**Figure 9C**). Only one "Visual transduction-Rods" satisfied the 10% false discovery rate criteria and was associated with lower immune infiltration.

Figures 10A-C summarizes the proportions of patient-level alterations in the various affected pathways for the three breast cancer subtypes. The proportion of patients with a certain pathway affected by a particular aberration or combination of aberrations is depicted for the immune low vs. immune high tertile. Thus, this figure compares the types of aberrations present in each pathway, but does not describe relative quantity of a particular aberration. At the individual patient level, each pathway appears to contain a unique combination of genomic aberrations. This analysis is supplemented by **Figure 11A-C,** which describes the aberrations in the member genes

from the most significant pathways for each subtype. At the individual patient level, each gene appears to demonstrate association of immune response with a unique combination of genomic aberrations. Copy number deletion and mutation are associated with low immune response in specific genes; amplification appears more ubiquitous in each patient, and a clear association pattern is difficult to ascertain.

Discussion

While causative mechanisms cannot be determined with certainty from an association study, our observations raise several biologically interesting hypotheses about what processes may influence immune infiltration in breast cancer. Our study suggests, that in some ER positive cancers mutations in members of the MAP-kinase family (MAP3K1, MAPK8, MAP2K4, MAPK1, MAPK3, MAP2K1, MAPK14, MAP2K3) are associated with, and perhaps cause, decreased immunogenicity. Since many of these enzymes activate other kinases MAPK8/JNK1, MAPK9/ JNK2, and MAPK14/p38 involved in stress-response and response to environmental stimuli, they have broad downstream effects, including generation of inflammatory cytokines [40], inducing proliferation of cytotoxic T-cells [41]. A connection between this and MHC expression was also recently described in TNBC, where genomic and transcriptomic alterations in the Ras-MAPK pathway upregulated MHC expression and had an overall negative effect on immune infiltration [42]. In ER negative cancer, MYH9 (non-muscle myosin IIa) mutations, seen in 4% of these cancers, was one of the two single gene events that were significantly associated with low

immune presence. This protein has been shown to function as a tumor suppressor in squamous cell carcinoma by leading to post-transcriptional p53 destabilization [43]. How these molecular defects could influence immunogenicity of a tumor remains to be investigated. Additionally, HERC2 mutations were also shown to be significant in ER negative cancers; HERC2 is an E3 ubiquitin ligase. This family of genes has been implicated in development, activation, differentiation of lymphocytes, antigen presentation [44], although a specific mechanism involving HERC2 mutations has not been demonstrated in the literature.

Several germline SNPs have been linked to autoimmune disorders and immune functions in general, therefore we also examined if germ line SNPs may be associated with immunological features of breast cancer. In ER negative cancers, two co-localized variants rs425757 and rs410232 corresponding to a missense aberration in the CFHR1 (Complement factor H-related protein) gene were associated with lower levels of immune infiltration. The CFHR1 protein binds to and inhibit the C3b components of the C5 convertase, and thus inhibit the terminal complement complex (TCC) [45]. CFHR1 also competes with factor H, a more potent inhibitor of the complement cascade and certain isoforms are associated with immunoprotection against IgA nephropathy and age-related macular degeneration [46, 47]. Deletion of CFHR1 is associated with autoimmune diseases such as hemolytic uremic syndrome and lupus. The rs470797 variant that corresponds to a possible stop-gain in the MBP (myelin basic protein) gene was also associated with lower immunogenicity. MBP is a putative cancer antigen; previous authors have demonstrated T-lymphocytes react with MBP in leukocyte adherence inhibition (LAI) assays [48]. Furthermore,

mutated MBP peptide ligands have been shown to be capable of switching immune responses from Th1 cells (mediating pro-inflammatory responses) to Th2 cells (anti-inflammatory) [49].

We also linked several copy number alterations to lower immune infiltration. Amplification of the 5p12-14.3 region was associated with decreased immune cell presence in ER negative cancers. A potential candidate gene mapping to this region is IL7R (interleukin 7 receptor). IL7 and IL7R over-expression has been correlated with an undifferentiated pathology and poor treatment response in breast cancers. Another proposed explanation involves the observation that IL-7 induces lymphatic endothelial cell growth [50]. Interestingly, lymphangiogenesis has increasingly been explored as a mechanism of immune escape [51]. Tumors potentially take advantage of self-tolerance functions in the lymph nodes.

Amplifications at 17q11-25 region are also associated with decreased immune presence in ER negative cancers. This amplicon is also associated with poorer prognosis in breast cancers [52]. A potential candidate gene in this region is CCR7 (Chemokine (C-C Motif) Receptor 7). High expression of CCR7 was shown to lead to decreased T-cell presence in the melanoma [53]; the putative mechanism for this effect is thought to be related to the establishment of a lymphangiogenesis-associated immune escape, as detailed above.

Several interesting aberrations were associated with immune response on the pathway level. For instance, ceramide signaling was related to decreased immune metagene expression in ER+/HER2- cancers. Ceramide is generated by hydrolysis of plasma membrane phospholipid

sphingomyelin, and plays an important role in a ubiquitous apoptosis pathway [54].The inhibition of ceramide action has been elucidated as a method of immune escape by preventing cytochrome c release and caspase activation [55].Additionally, our findings of poor immune response in patients with eicosanoid pathway aberrations may be explained by the demonstrated effect of cancers to control prostaglandins and leukotrienes, in order to evade attack from the immune system [56].

Further studies may benefit from validating high interest aberrations in a dataset with recorded measures of clinical outcomes and histological validation of tumor infiltrating lymphocytes.For the purposes of clinical utility, it is imperative to focus on mutations and SNV's which are mutated in a relatively high proportion of patients. It would additionally be interesting to explore the immunogenicity of epigenetic variations in certain areas of interest, as methylation of HLA genes is a well known mechanism for immune modulation [57].

In conclusion, our analysis shows that the extent of immune cell infiltration in breast cancer is not driven by an all-encompassing global metric of genomic instability such as overall mutation, copy number or neoantigen loads, or by a few, highly recurrent, single gene mutations or amplifications/deletions. We identified many different individually rare mutations and interesting copy number alterations at single gene level, and at the level of biological pathways, that were significantly associated with lower (or less commonly higher) immune gene expression in different breast cancer subtypes. These results suggest that a broad range of genomic aberrations, as well as SNPs can influence the immunogenicity of a given breast cancer subtype.

Figure 1: The expression distribution of each of the 13 immune metagenes is shown for the ER+/HER2- (**A**), ER-/HER2- **(B**), and HER2+ (**C**) cohorts. Metagenes describing IF1, macrophage, MHC1, MHC2, STAT1, T follicular cells, T cell inhibitory and stimulatory activity, as well as lymphocyte-specific kinase (LCK), cytolytic activity (CTL), and consensus T-cell metagene (CTM) unimodal normal distributions. Metagenes describing natural killer (NK) cells, and regulatory T-cells (T-regs) showed bimodal distributions.

Figure 2: The correlation matrix of each of the 13 immune metagenes is shown for the ER+/HER2- (**A**), ER-/HER2- **(B**), and HER2+ (**C**) cohorts. Most immune metagenes showed high levels of co-expression in all 3 breast cancer subtypes. Notably, the regulatory T-cell signature had low correlation in all subtypes. Metagenes marked with (*) indicate previous evidence of association to clinical outcome. Lymphocyte-specific kinase (LCK) was chosen as most representative metagene due to its high degree of correlation with prognostically favorable metagenes and evidence in the literature demonstrating correlation with histological T-leukocyte presence.

Figure 3: The correlation between each of the 13 immune metagenes and mutation, neoantigen copy number amplification and deletion load is shown for the ER+/HER2- (**A**), ER-/HER2- **(B**), and HER2+ (**C**) cohorts. Spearman correlation values demonstrate low levels of correlation. Additionally, ANOVA Chi-Square significance testing revealed no significant linear regression coefficients between the metagenes and the 4 different types of global genomic aberration metrics.

Figure 4: Principal component analysis was performed on a combined dataset featuring mutation load and copy number amplification and deletion loads for each patient, according to ER+/HER2- **(A-C)** subtype, ER-/HER2- **(D-F)** cohort, and HER2+ **(G-I)** subtype. Biplots are also divided by PC1 vs PC2 **(A,D,G),** PC2 vs PC3 (**B,E,H**) and PC1 vs PC3 (**C,F,I**) Generally, no clear pattern of separation by between high and low LCK metagene tertiles was apparent for any combination of aberrational type. .

Figure 5: The subtype-specific significance of association between immune activity and mutation for all nominally significant genes. (**A)** represents the ER+/HER2- cohort, (**B)** depicts the ER-/HER2- cohort, and (**C)** represents the HER2+ cohort. Rightward bars represent genes in which a mutation is positively associated with LCK, and leftward bars show genes in which mutation was negatively associated with LCK. The mutation frequency of each gene is labeled as a percentile value; vertical red lines represent the significance cutoffs yielding 10% FDR.

Figure 6: The subtype-specific significance of association between immune activity and single nucleotide variation (SNV) for all nominally significant genes in ER-/HER2- cohort. ER+/HER2- and HER2+ not shown, as there were no SNV's significant after correction for multiple hypothesis testing in these subtypes. Rightward bars represent genes in which a mutation is positively associated with LCK, and leftward bars show genes in which mutation was negatively associated with LCK. The mutation frequency of each gene is labeled as a percentile value; vertical red lines represent the significance cutoffs yielding 10% FDR. Only the 50 SNV's with the smallest nominal p-value are depicted.

Figure 7: The subtype-specific significance of association between immune activity and copy number amplification/ deletion for all genic loci. (**A)** represents the ER+/HER2- cohort, (**B)** depicts the ER-/HER2- cohort, and (**C)** represents the HER2+ cohort. Rightward lines show nominal p-values for instances in which the lesion was positively associated with LCK, and leftward lines show nominal p-values for instances in which the lesion was negatively associated with LCK. Labeled peaks represent cytobands which were significant after adjustment for multiple-hypothesis testing.

Figure 8: Venn diagrams specific for each subtype represent the number of genes undergoing a nominally significant aberration type (or combination of aberration types) after multivariate regression of the mutations, copy number alterations, and expression at gene level with immune function. In HER2+ and ER-/HER2- cases, higher immune infiltration is more closely associated with expression, while decreased immune infiltration is more frequently associated with deletions and amplifications.

Figure 9: The subtype-specific significance of association between immune activity and pathway aberration frequency for all nominally significant pathways. (**A)** represents the ER+/HER2- cohort, (**B)** depicts the ER-/HER2- cohort, and (**C)** represents the HER2+ cohort. Rightward bars represent genes in which a mutation is positively associated with LCK, and leftward bars show genes in which mutation was negatively associated with LCK. The mutation frequency of each gene is labeled as a percentile value; vertical red lines represent the significance cutoffs yielding 10% FDR.(**D)** and (**E)** represent membership of genes in each pathway found to be significant.

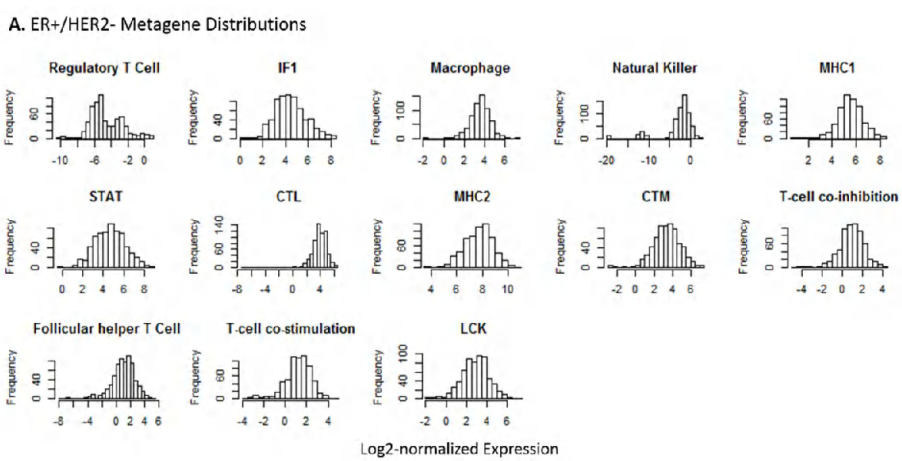
Figures 10A-C summarizes the patient level alterations in the various affected pathways for the three breast cancer subtypes. The proportion of patients with a certain pathway affected by a particular aberration or combination of aberrations is depicted for the immune low vs. immune high tertile; the number of each aberration in a particular pathway is thus not reflected in this

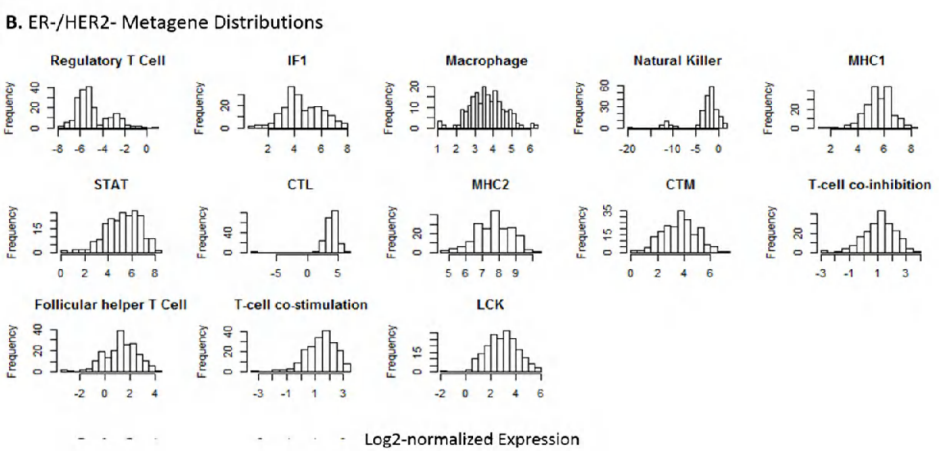
figure. At the individual patient level, each pathway appears to contain a unique combination of genomic aberrations.

Figures 11A-B summarizes the patient level alterations in member genes from the most significant pathways for ER+/HER2- and ER-/HER2-. At the individual patient level, each gene appears to demonstrate association of immune response with a unique combination of genomic aberrations. Copy number deletion and mutation are associated with low immune response in specific genes; amplification appears more ubiquitous in each patient, and a clear association pattern is difficult to ascertain.

Figures

Figure 1





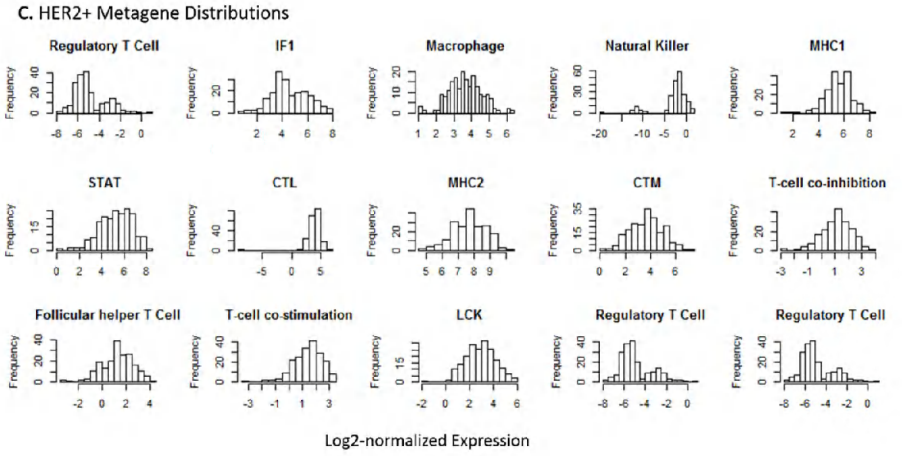


Figure 2:

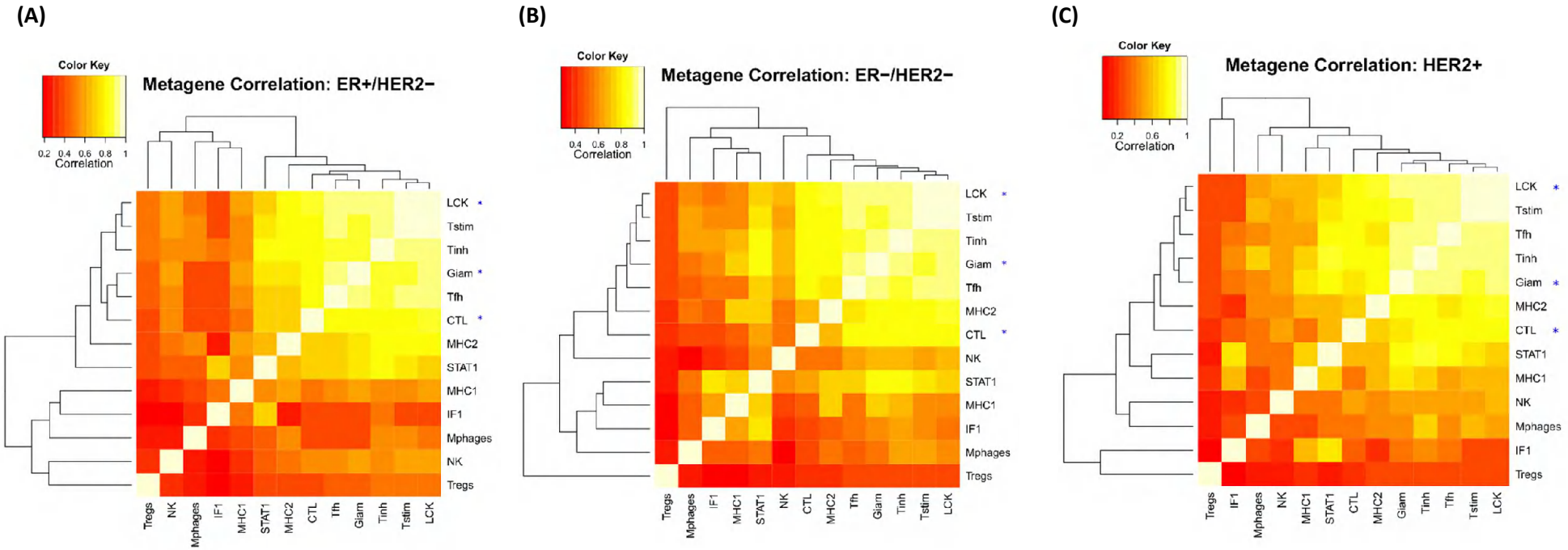
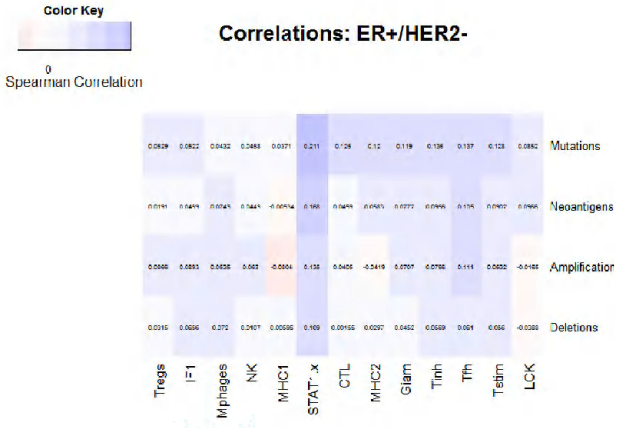
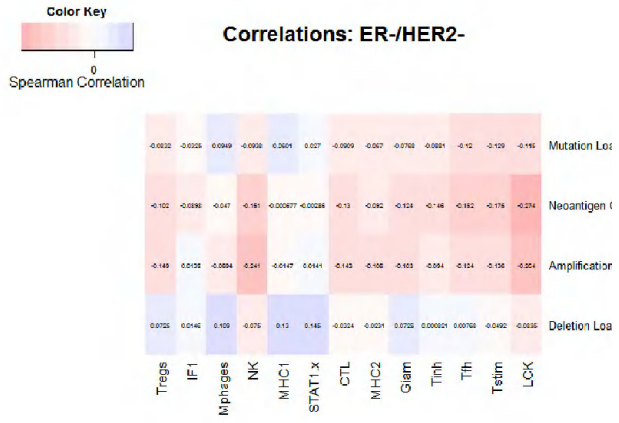

(A)
Color Key
0.2 0.4 0.6 0.8 1
Correlation
Metagene Correlation: ER+/HER2−
LCK *
Tstim
Tinh
Giam *
Tfh
CTL *
MHC2
STAT1
MHC1
IF1
Mphages
NK
Tregs
Tregs NK Mphages IF1 MHC1 STAT1 MHC2 CTL Tfh Giam Tinh Tstim LCK
(B)
Color Key
0.4 0.6 0.8 1
Correlation
Metagene Correlation: ER−/HER2−
LCK *
Tstim
Tinh
Giam *
Tfh
MHC2
CTL *
NK
STAT1
MHC1
IF1
Mphages
Tregs
Tregs Mphages IF1 MHC1 STAT1 NK CTL MHC2 Tfh Giam Tinh Tstim LCK
(C)
Color Key
0.2 0.6 1
Correlation
Metagene Correlation: HER2+
LCK *
Tstim
Tfh
Tinh
Giam *
MHC2
CTL *
STAT1
MHC1
NK
Mphages
IF1
Tregs
Tregs IF1 Mphages NK MHC1 STAT1 CTL MHC2 Giam Tinh Tfh Tstim LCK

Figure 3:

(A)


Color Key
0
Spearman Correlation
Correlations: ER+/HER2-
Mutations
Neoantigens
Amplification
Deletions
Tregs
IF1
Mphages
NK
MHC1
STAT1.x
CTL
MHC2
Glam
Tinh
Tfh
Tstim
LCK

(B)


Color Key
0
Spearman Correlation
Correlations: ER-/HER2-
Mutation Loa
Neoantigen (
Amplificatior
Deletion Loa
Tregs
IF1
Mphages
NK
MHC1
STAT1.x
CTL
MHC2
Glam
Tinh
Tfh
Tstim
LCK

(C)

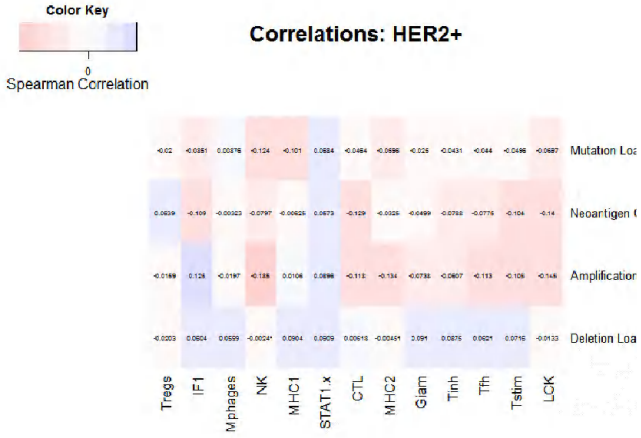
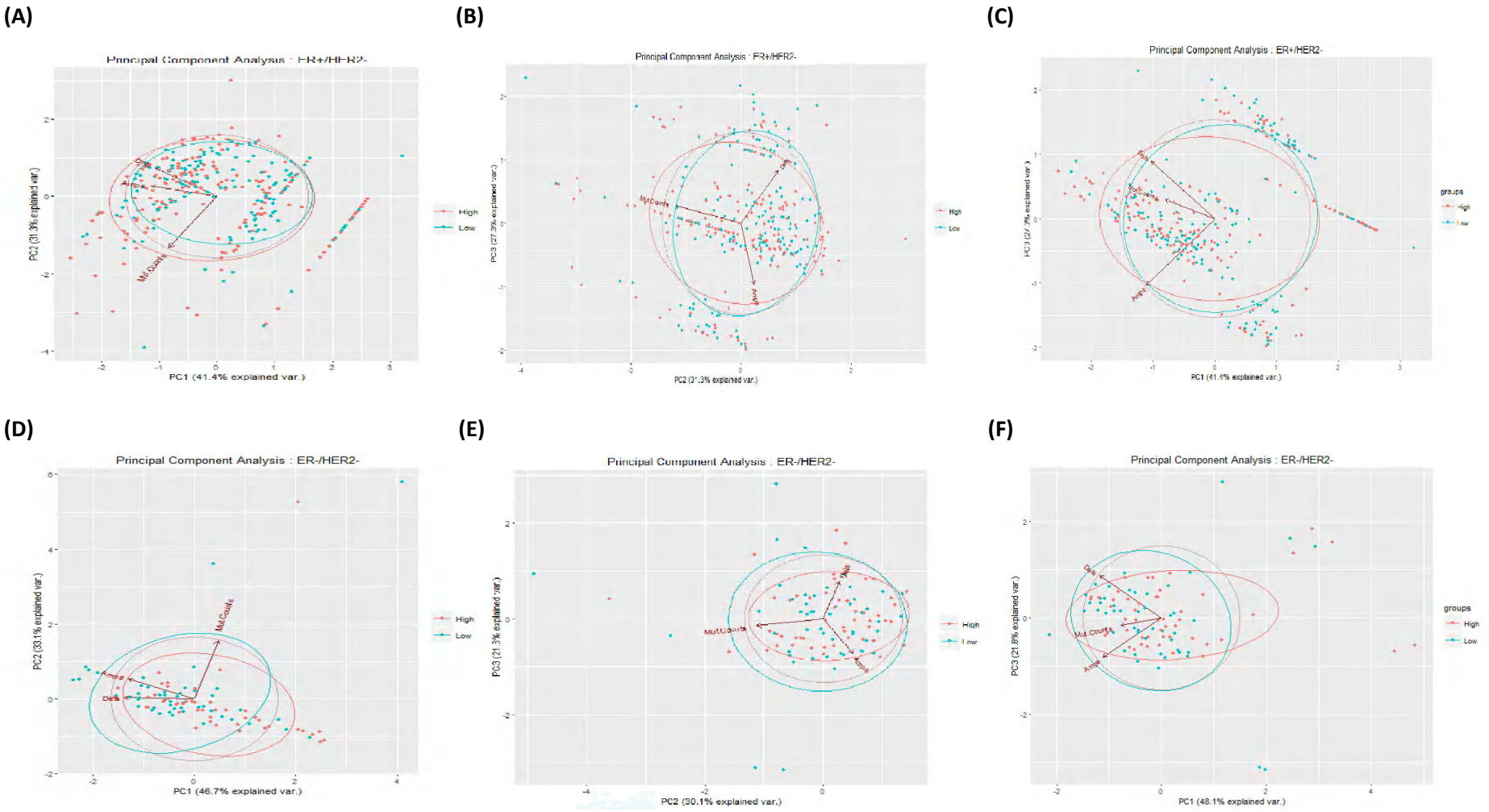
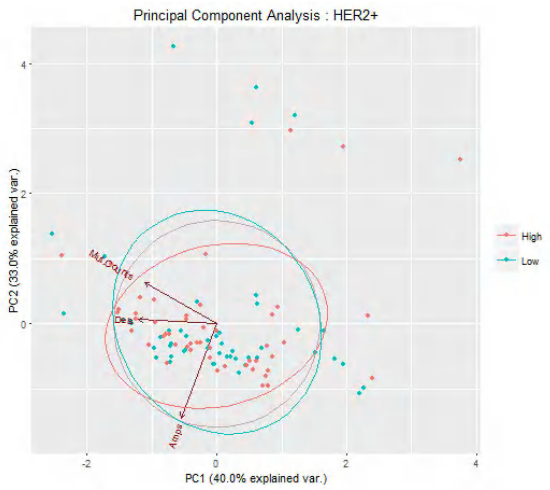

Color Key
0
Spearman Correlation
Correlations: HER2+
Mutation Loa
Neoantigen (
Amplificatior
Deletion Loa
Tregs
IF1
Mphages
NK
MHC1
STAT1.x
CTL
MHC2
Glam
Tinh
Tfh
Tstim
LCK

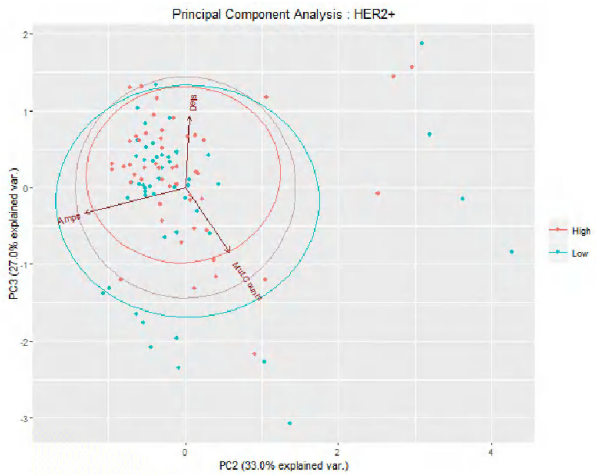
Figure 4:


(A)
Principal Component Analysis : ER+/HER2-
PC1 (41.4% explained var.)
PC2 (31.3% explained var.)
Del
Amp
MutCounts
High
Low
(B)
Principal Component Analysis : ER+/HER2-
PC2 (31.3% explained var.)
PC3 (27.9% explained var.)
Del
MutCounts
Amp
High
Low
(C)
Principal Component Analysis : ER+/HER2-
PC1 (41.4% explained var.)
PC3 (27.9% explained var.)
Del
MutCounts
Amp
groups
High
Low
(D)
Principal Component Analysis : ER-/HER2-
PC1 (46.7% explained var.)
PC2 (33.1% explained var.)
MutCounts
Amp
Del
High
Low
(E)
Principal Component Analysis : ER-/HER2-
PC2 (30.1% explained var.)
PC3 (21.3% explained var.)
Del
MutCounts
Amp
High
Low
(F)
Principal Component Analysis : ER-/HER2-
PC1 (48.1% explained var.)
PC3 (21.8% explained var.)
Del
MutCounts
Amp
groups
High
Low

(G)


Principal Component Analysis : HER2+
PC2 (33.0% explained var.)
PC1 (40.0% explained var.)
MutCounts
Del
Amps
High
Low

(H)


Principal Component Analysis : HER2+
PC3 (27.0% explained var.)
PC2 (33.0% explained var.)
DRB
Amps
MutCounts
High
Low

(I)

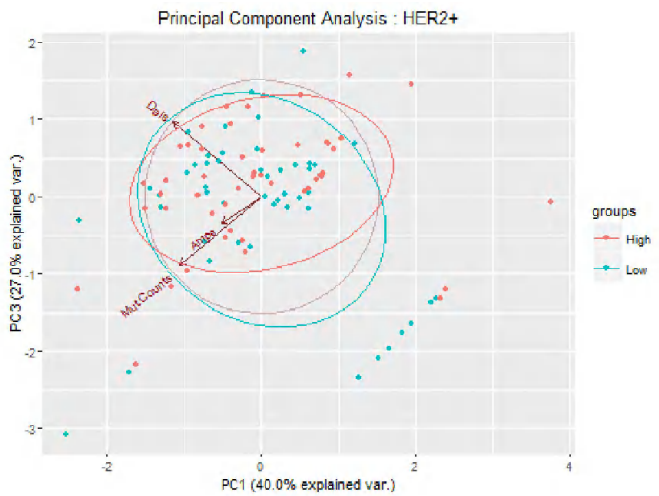
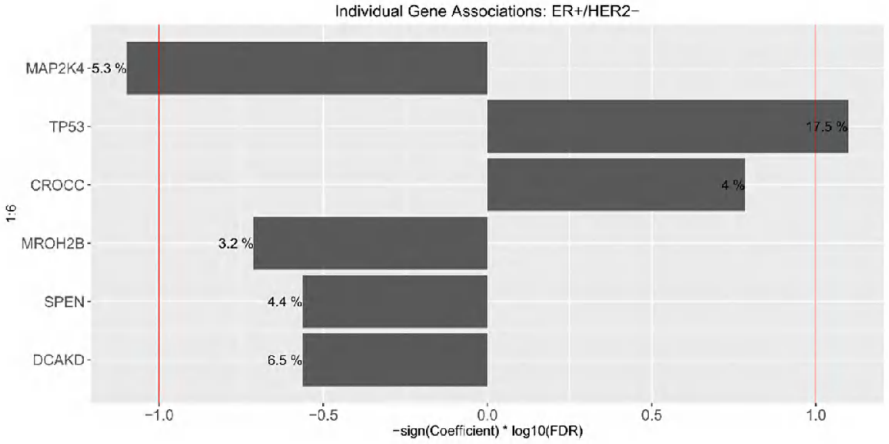
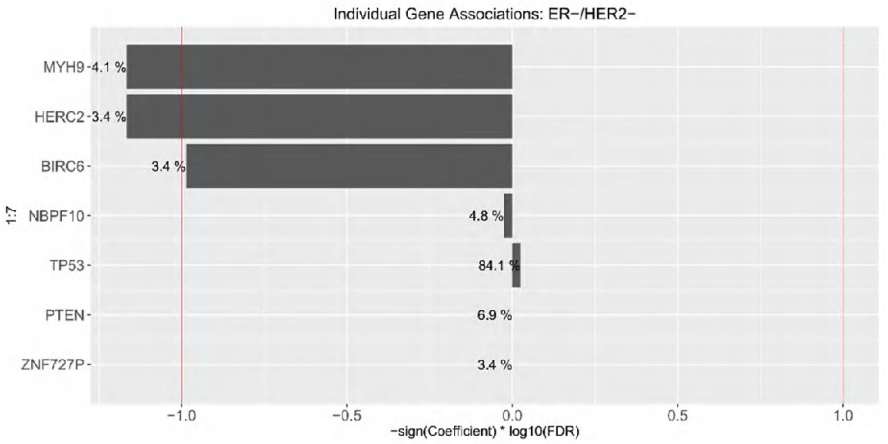

Principal Component Analysis : HER2+
PC3 (27.0% explained var.)
PC1 (40.0% explained var.)
Del
Amps
MutCounts
groups
High
Low

Figure 5:

(A)


Individual Gene Associations: ER+/HER2−
MAP2K4
TP53
CROCC
MROH2B
SPEN
DCAKD
5.3 %
17.5 %
4 %
3.2 %
4.4 %
6.5 %
1:6
−1.0
−0.5
0.0
0.5
1.0
−sign(Coefficient) * log10(FDR)

(B)


Individual Gene Associations: ER−/HER2−
MYH9
HERC2
BIRC6
NBPF10
TP53
PTEN
ZNF727P
4.1 %
3.4 %
3.4 %
4.8 %
84.1 %
6.9 %
3.4 %
1:7
−1.0
−0.5
0.0
0.5
1.0
−sign(Coefficient) * log10(FDR)

(C)

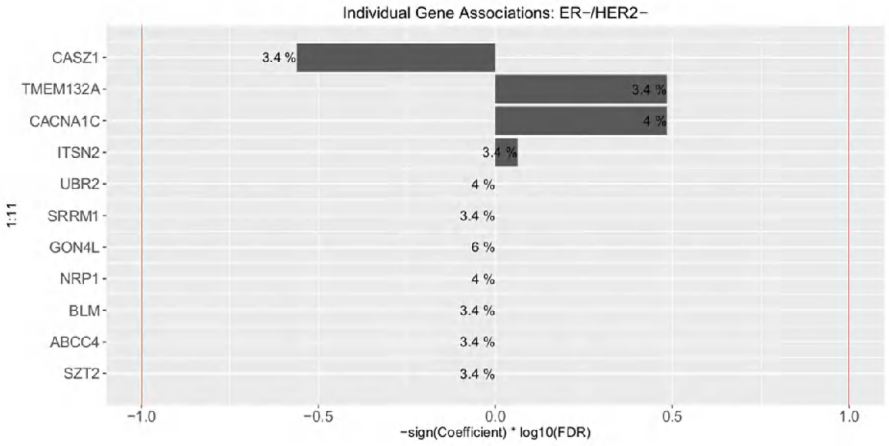
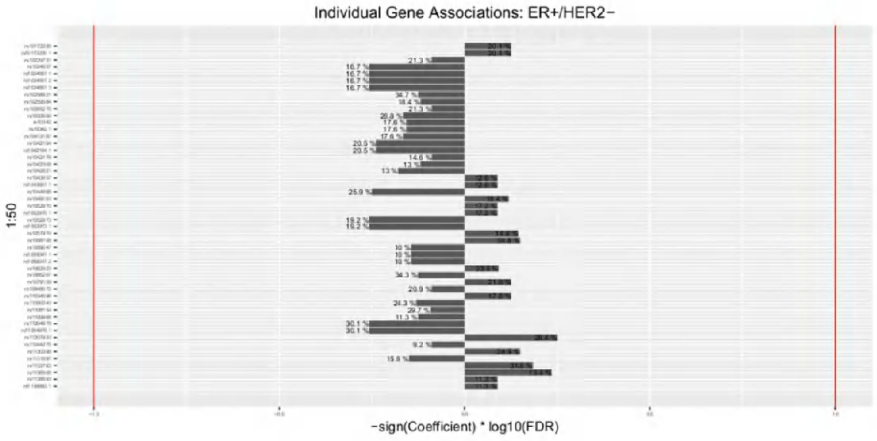
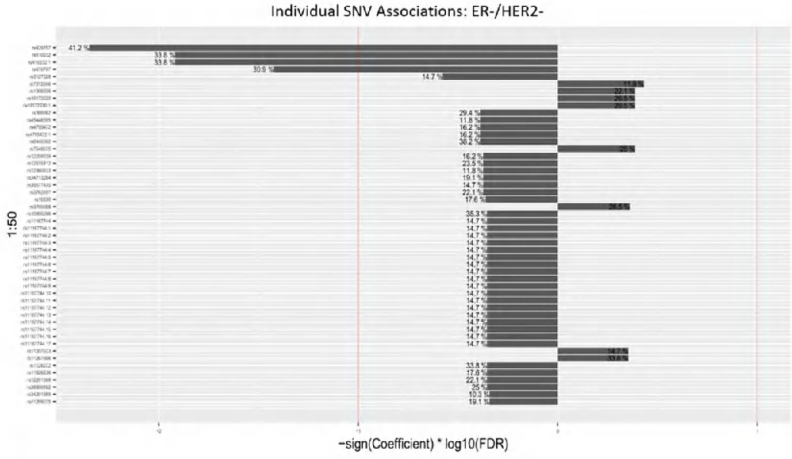

Individual Gene Associations: ER−/HER2−
CASZ1
TMEM132A
CACNA1C
ITSN2
UBR2
SRRM1
GON4L
NRP1
BLM
ABCC4
SZT2
3.4 %
3.4 %
4 %
3.4 %
4 %
3.4 %
6 %
4 %
3.4 %
3.4 %
3.4 %
1:11
−1.0
−0.5
0.0
0.5
1.0
−sign(Coefficient) * log10(FDR)

Figure 6:

(A)


Individual Gene Associations: ER+/HER2−
1:50
−sign(Coefficient) * log10(FDR)

(B)


Individual SNV Associations: ER-/HER2-
1:50
−sign(Coefficient) * log10(FDR)

(C)

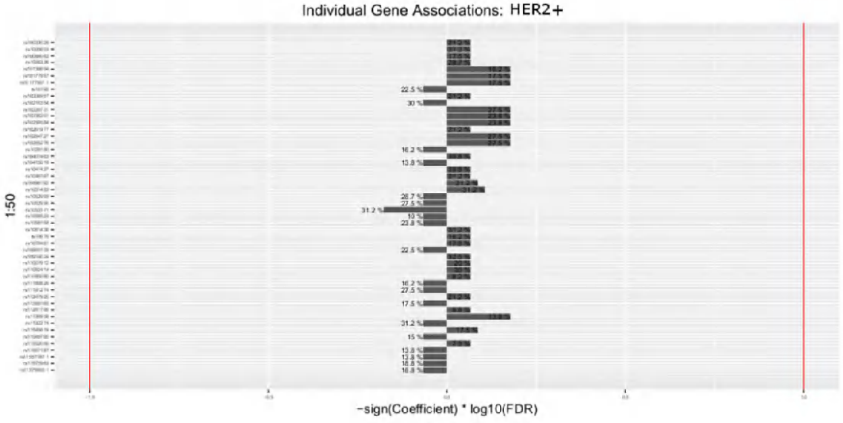

Individual Gene Associations: HER2+
1:50
−sign(Coefficient) * log10(FDR)

Figure 7.A

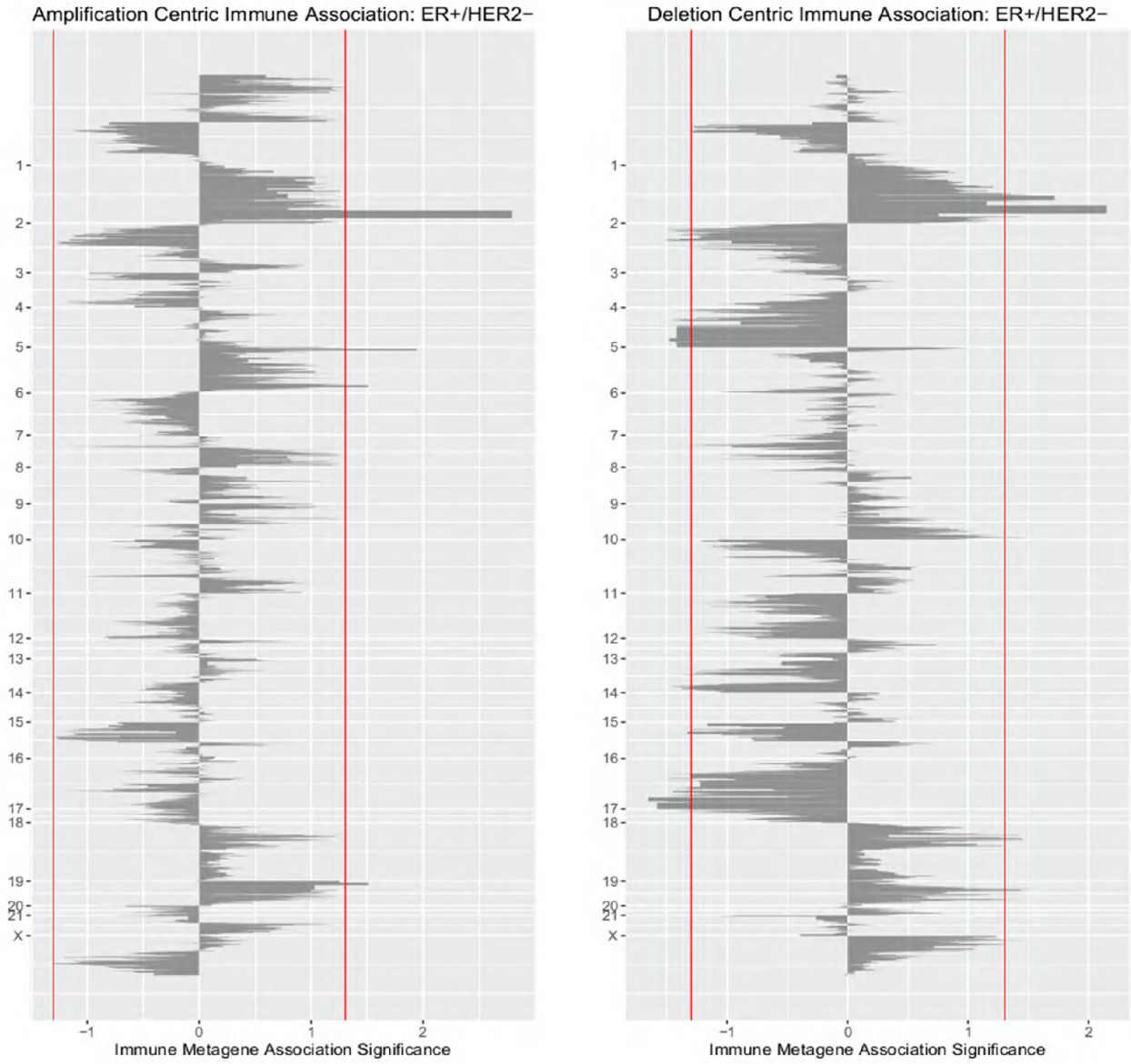

Amplification Centric Immune Association: ER+/HER2−
Deletion Centric Immune Association: ER+/HER2−
1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
X
−1
0
1
2
Immune Metagene Association Significance

Figure 7B

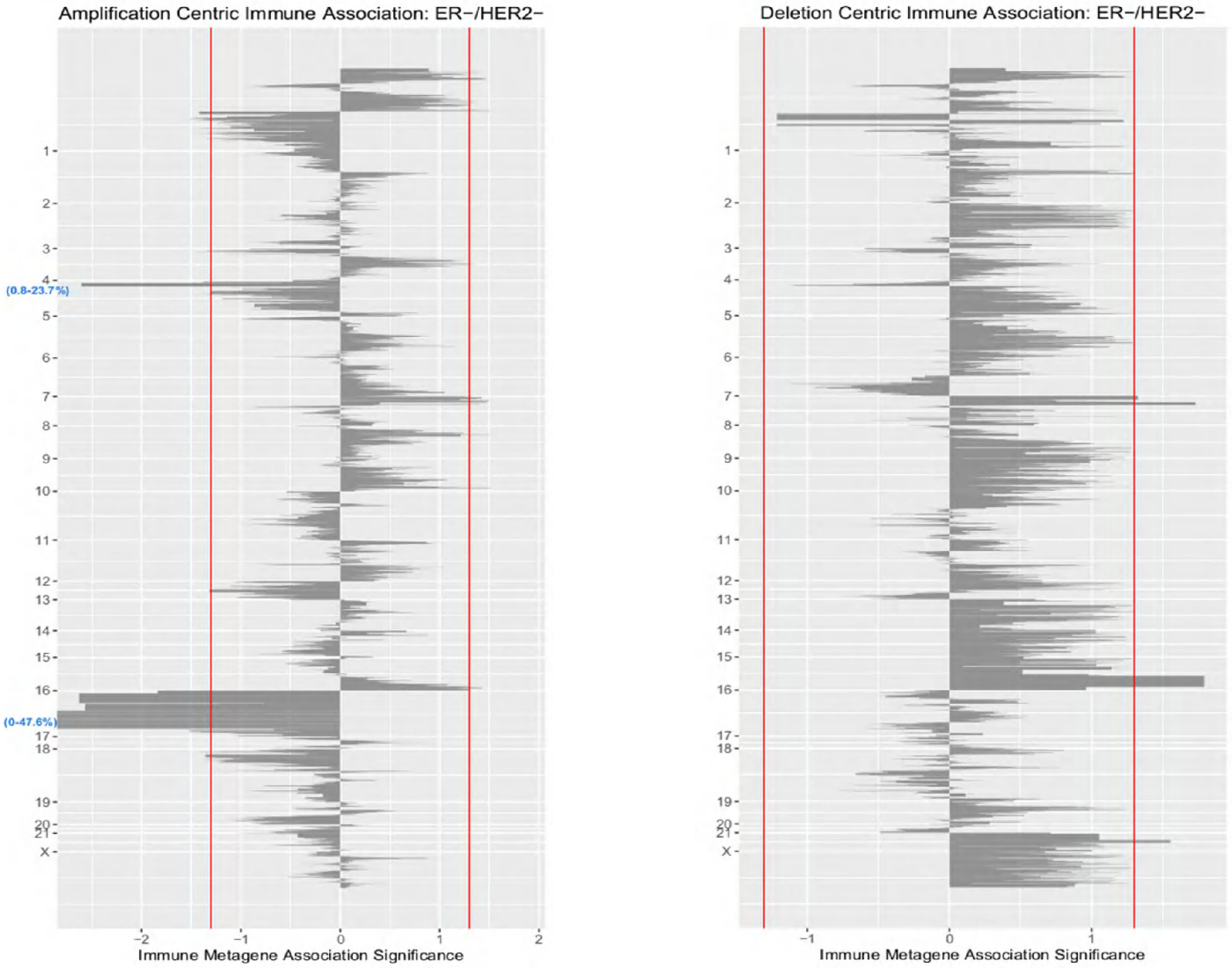

Amplification Centric Immune Association: ER−/HER2−
1
2
3
4
(0.8-23.7%)
5
6
7
8
9
10
11
12
13
14
15
16
(0-47.6%)
17
18
19
20
21
X
−2
−1
0
1
2
Immune Metagene Association Significance
Deletion Centric Immune Association: ER−/HER2−
1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
X
−1
0
1
Immune Metagene Association Significance

Figure 7C

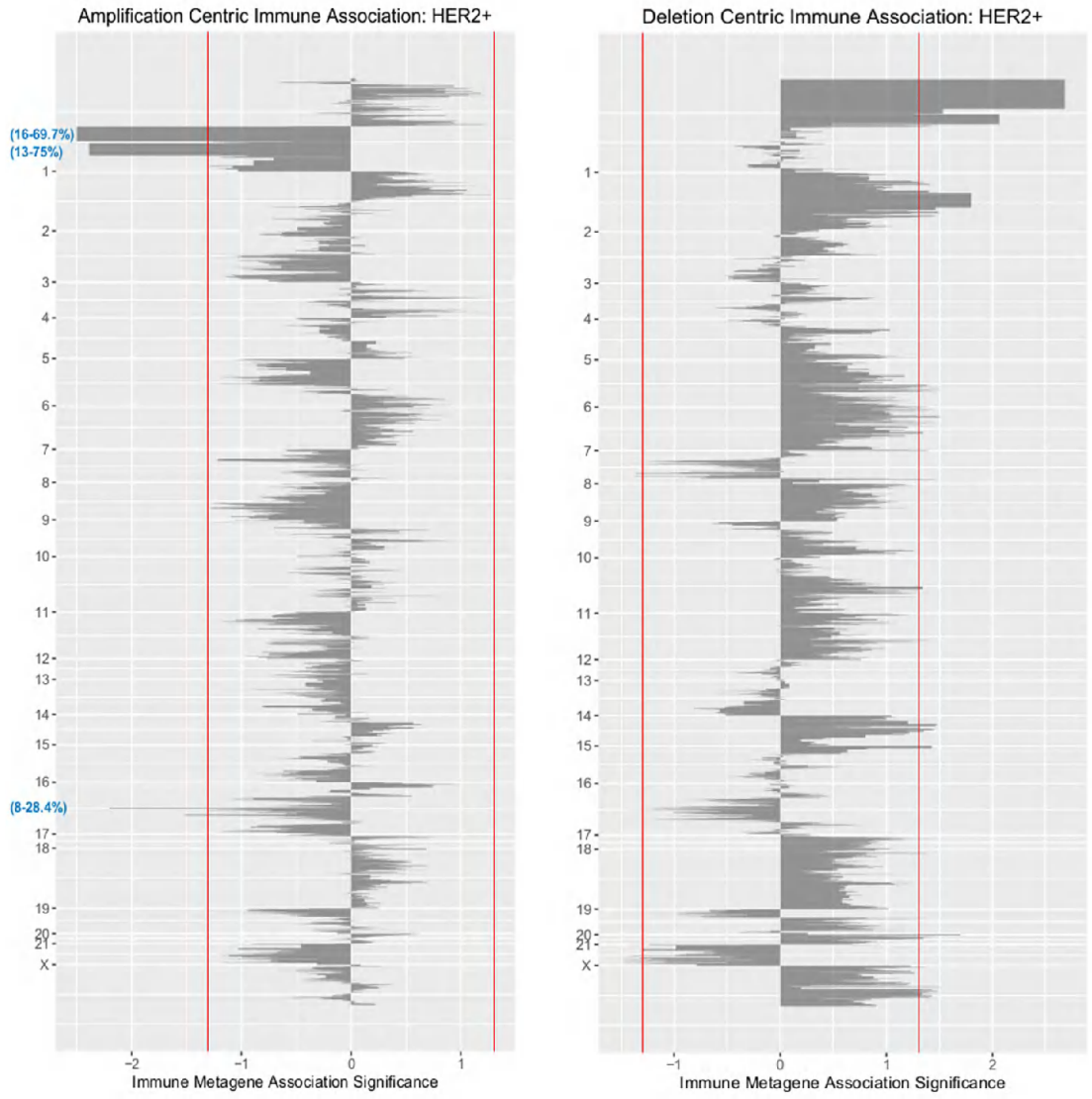
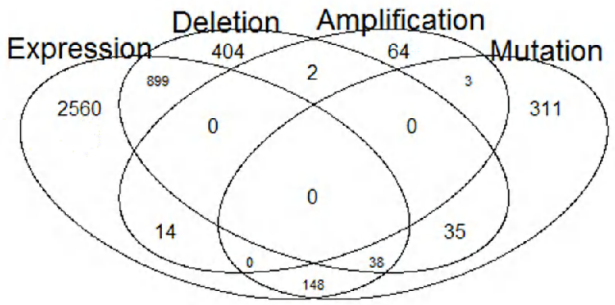

Amplification Centric Immune Association: HER2+
(16-69.7%)
(13-75%)
(8-28.4%)
1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
X
−2
−1
0
1
Immune Metagene Association Significance
Deletion Centric Immune Association: HER2+
−1
0
1
2
Immune Metagene Association Significance

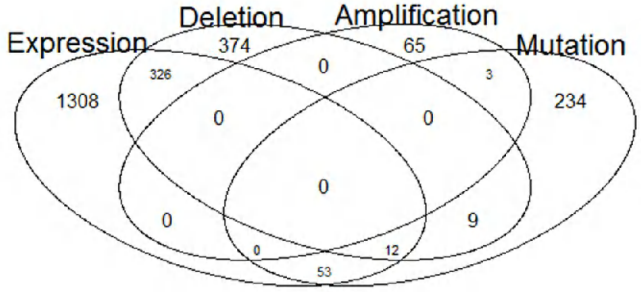
Figure 8

(A)

ER+/HER2- Aberrations Positively Associated with Imm. Response

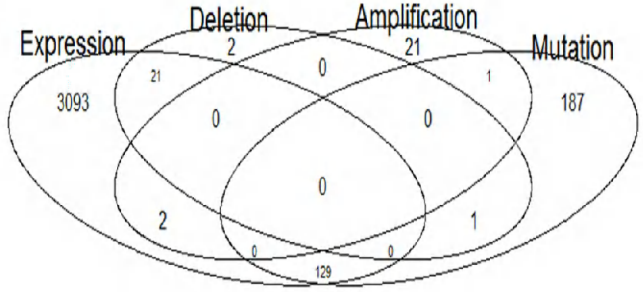


ER+/HER2- Aberrations Negatively Associated with Imm. Response

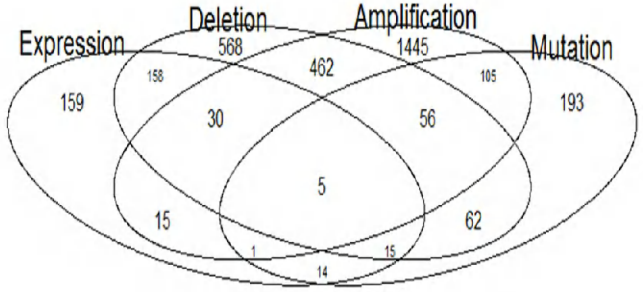


(B)

ER-/HER2- Aberrations Positively Associated with Imm. Response

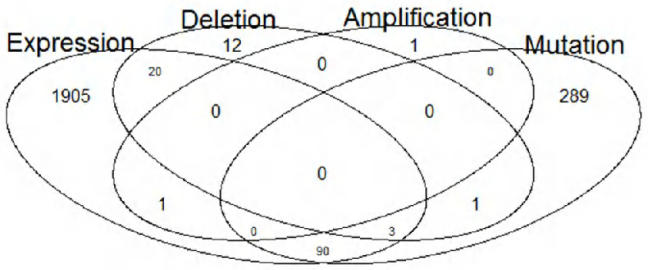


ER-/HER2- Aberrations Negatively Associated with Imm. Response



(C)

HER2+ Aberrations Positively Associated with Imm. Response



HER2+ Aberrations Negatively Associated with Imm. Response

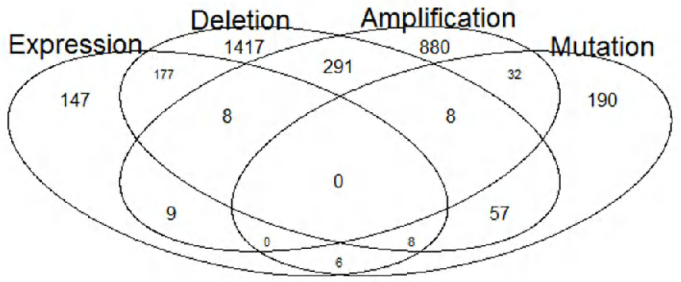


Figure 9A.

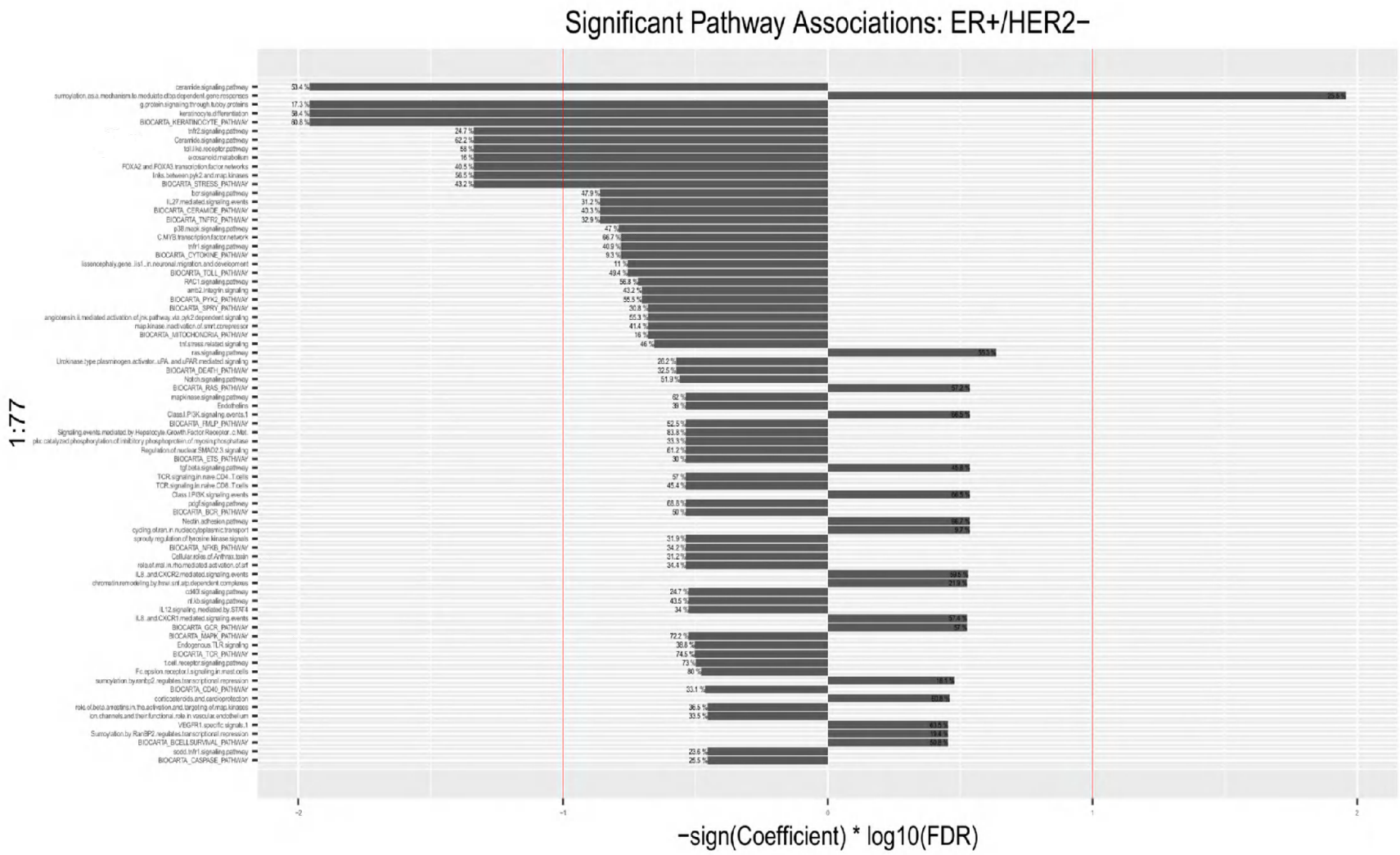

Significant Pathway Associations: ER+/HER2−
1:77
−sign(Coefficient) * log10(FDR)

Figure 9B.

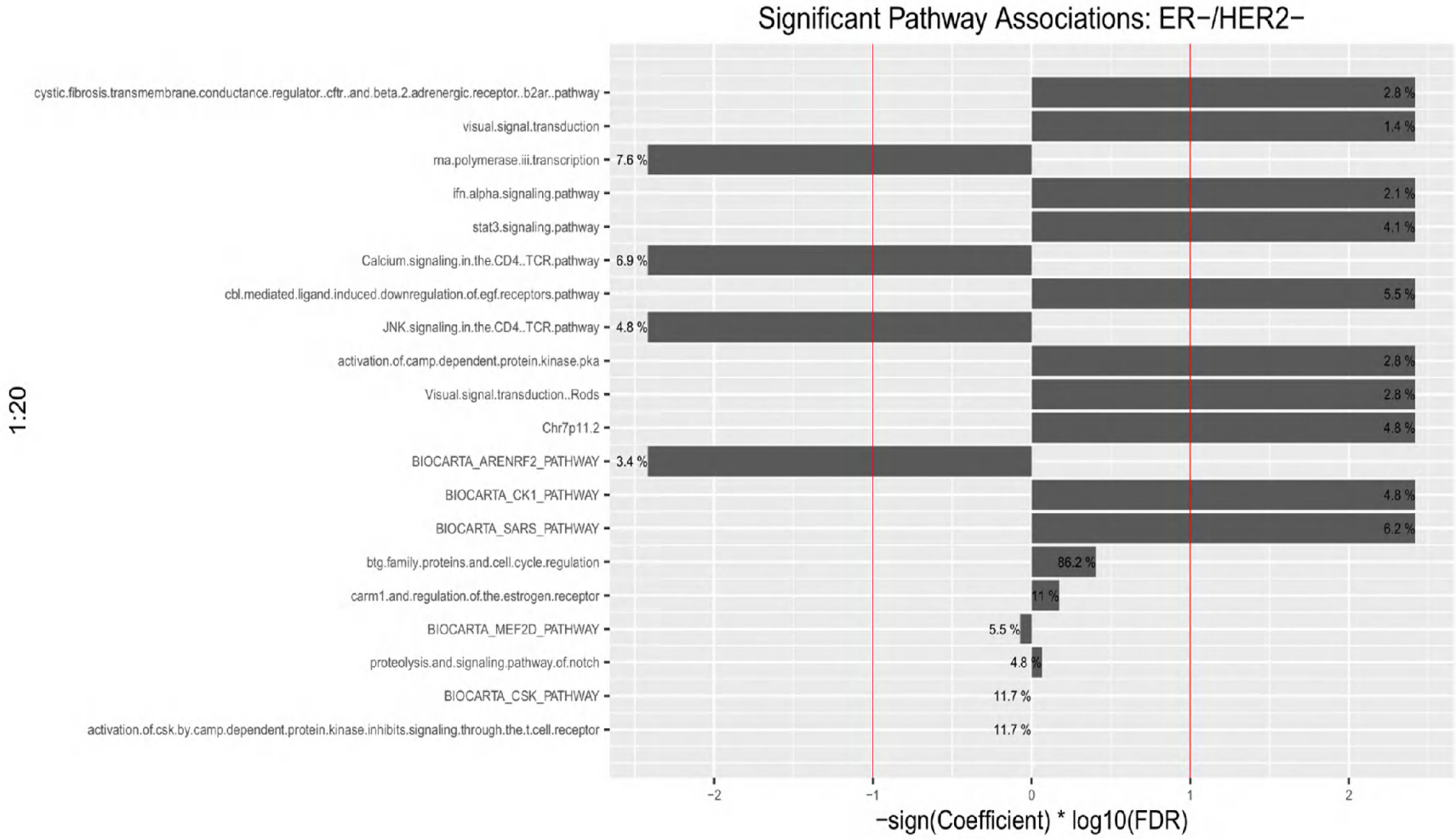

Significant Pathway Associations: ER−/HER2−
1:20
cystic.fibrosis.transmembrane.conductance.regulator..cftr..and.beta.2.adrenergic.receptor..b2ar..pathway
visual.signal.transduction
rna.polymerase.iii.transcription
ifn.alpha.signaling.pathway
stat3.signaling.pathway
Calcium.signaling.in.the.CD4..TCR.pathway
cbl.mediated.ligand.induced.downregulation.of.egf.receptors.pathway
JNK.signaling.in.the.CD4..TCR.pathway
activation.of.camp.dependent.protein.kinase.pka
Visual.signal.transduction..Rods
Chr7p11.2
BIOCARTA_ARENRF2_PATHWAY
BIOCARTA_CK1_PATHWAY
BIOCARTA_SARS_PATHWAY
btg.family.proteins.and.cell.cycle.regulation
carm1.and.regulation.of.the.estrogen.receptor
BIOCARTA_MEF2D_PATHWAY
proteolysis.and.signaling.pathway.of.notch
BIOCARTA_CSK_PATHWAY
activation.of.csk.by.camp.dependent.protein.kinase.inhibits.signaling.through.the.t.cell.receptor
2.8 %
1.4 %
7.6 %
2.1 %
4.1 %
6.9 %
5.5 %
4.8 %
2.8 %
2.8 %
4.8 %
3.4 %
4.8 %
6.2 %
86.2 %
11 %
5.5 %
4.8 %
11.7 %
11.7 %
−2
−1
0
1
2
−sign(Coefficient) * log10(FDR)

Fig 9C.

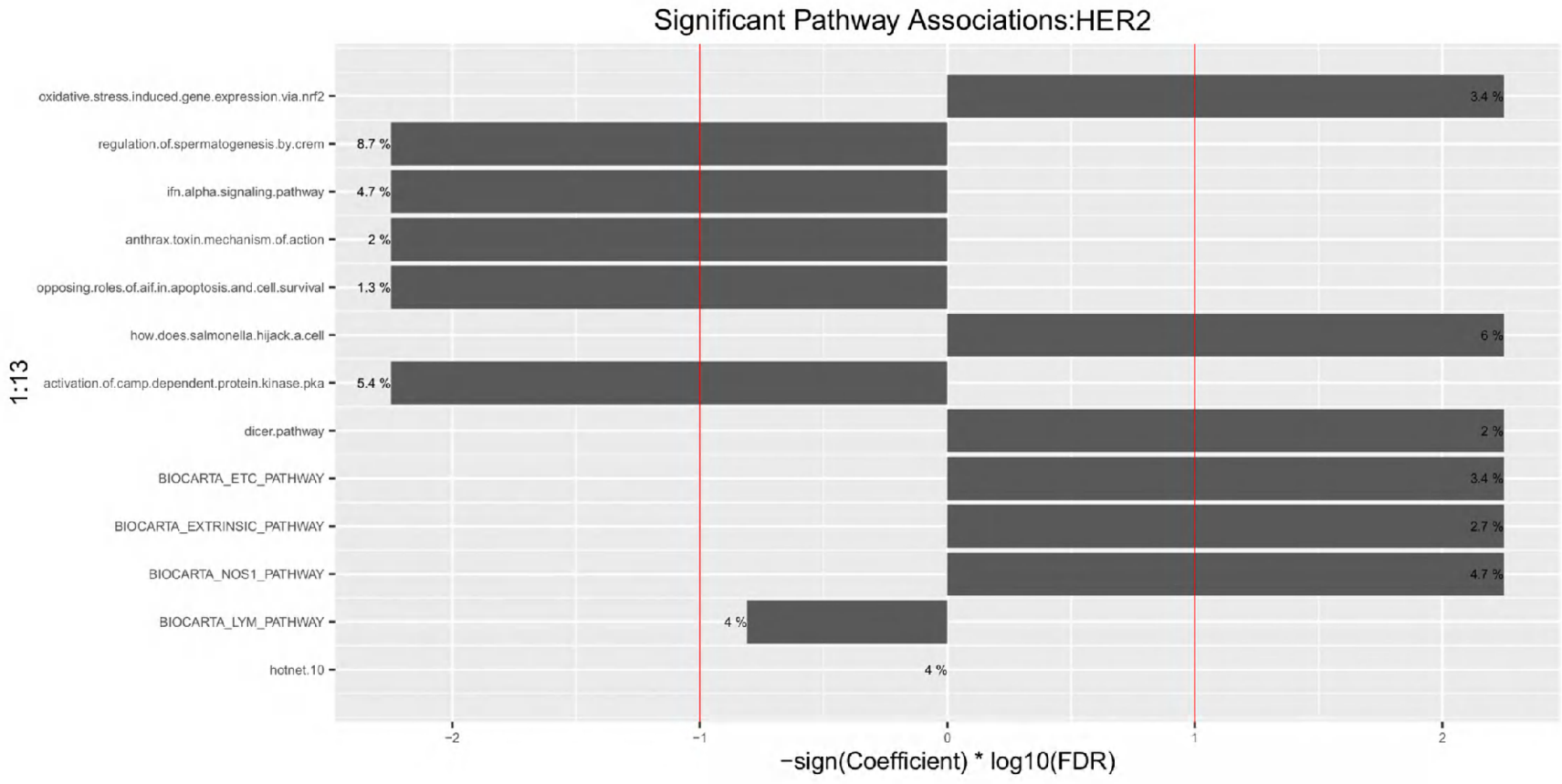
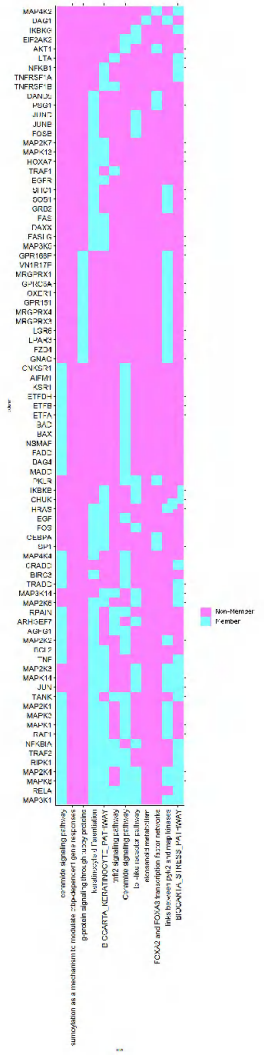

Significant Pathway Associations:HER2
oxidative.stress.induced.gene.expression.via.nrf2
regulation.of.spermatogenesis.by.crem
ifn.alpha.signaling.pathway
anthrax.toxin.mechanism.of.action
opposing.roles.of.aif.in.apoptosis.and.cell.survival
how.does.salmonella.hijack.a.cell
activation.of.camp.dependent.protein.kinase.pka
dicer.pathway
BIOCARTA_ETC_PATHWAY
BIOCARTA_EXTRINSIC_PATHWAY
BIOCARTA_NOS1_PATHWAY
BIOCARTA_LYM_PATHWAY
hotnet.10
3.4 %
8.7 %
4.7 %
2 %
1.3 %
6 %
5.4 %
2 %
3.4 %
2.7 %
4.7 %
4 %
4 %
1:13
−2
−1
0
1
2
−sign(Coefficient) * log10(FDR)

Fig9

D.


Non-Member
Member

E.

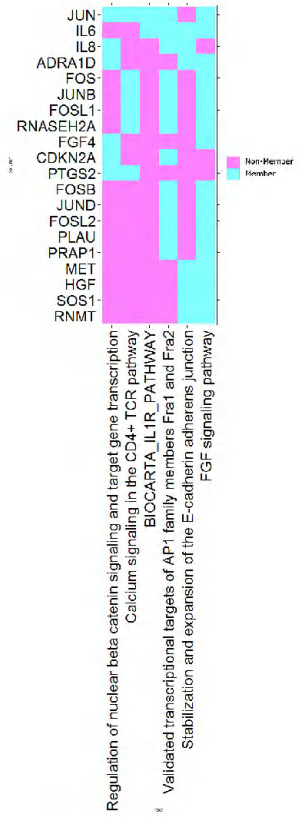
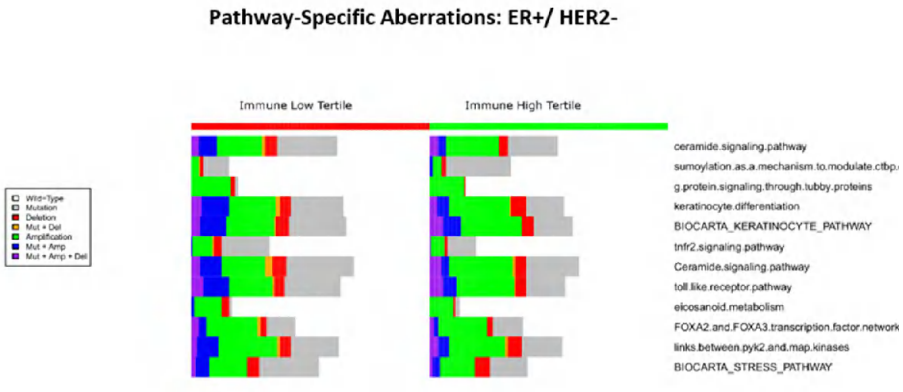

JUN
IL6
IL8
ADRA1D
FOS
JUNB
FOSL1
RNASEH2A
FGF4
CDKN2A
PTGS2
FOSB
JUND
FOSL2
PLAU
PRAP1
MET
HGF
SOS1
RNMT
Non-Member
Member
Regulation of nuclear beta catenin signaling and target gene transcription
Calcium signaling in the CD4+ TCR pathway
BIOCARTA_IL1R_PATHWAY
Validated transcriptional targets of AP1 family members Fra1 and Fra2
Stabilization and expansion of the E-cadherin adherens junction
FGF signaling pathway

Fig 10

(A)


Pathway-Specific Aberrations: ER+/ HER2-
Immune Low Tertile
Immune High Tertile
Wild-Type
Mutation
Deletion
Mut + Del
Amplification
Mut + Amp
Mut + Amp + Del
ceramide.signaling.pathway
sumoylation.as.a.mechanism.to.modulate.ctbp
g.protein.signaling.through.tubby.proteins
keratinocyte.differentiation
BIOCARTA_KERATINOCYTE_PATHWAY
tnfr2.signaling.pathway
Ceramide.signaling.pathway
toll.like.receptor.pathway
eicosanoid.metabolism
FOXA2.and.FOXA3.transcription.factor.network
links.between.pyk2.and.map.kinases
BIOCARTA_STRESS_PATHWAY

(B)

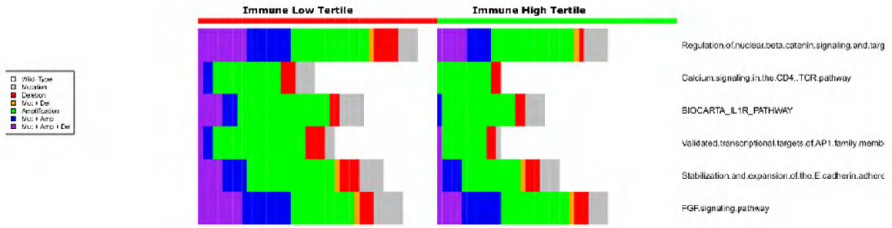
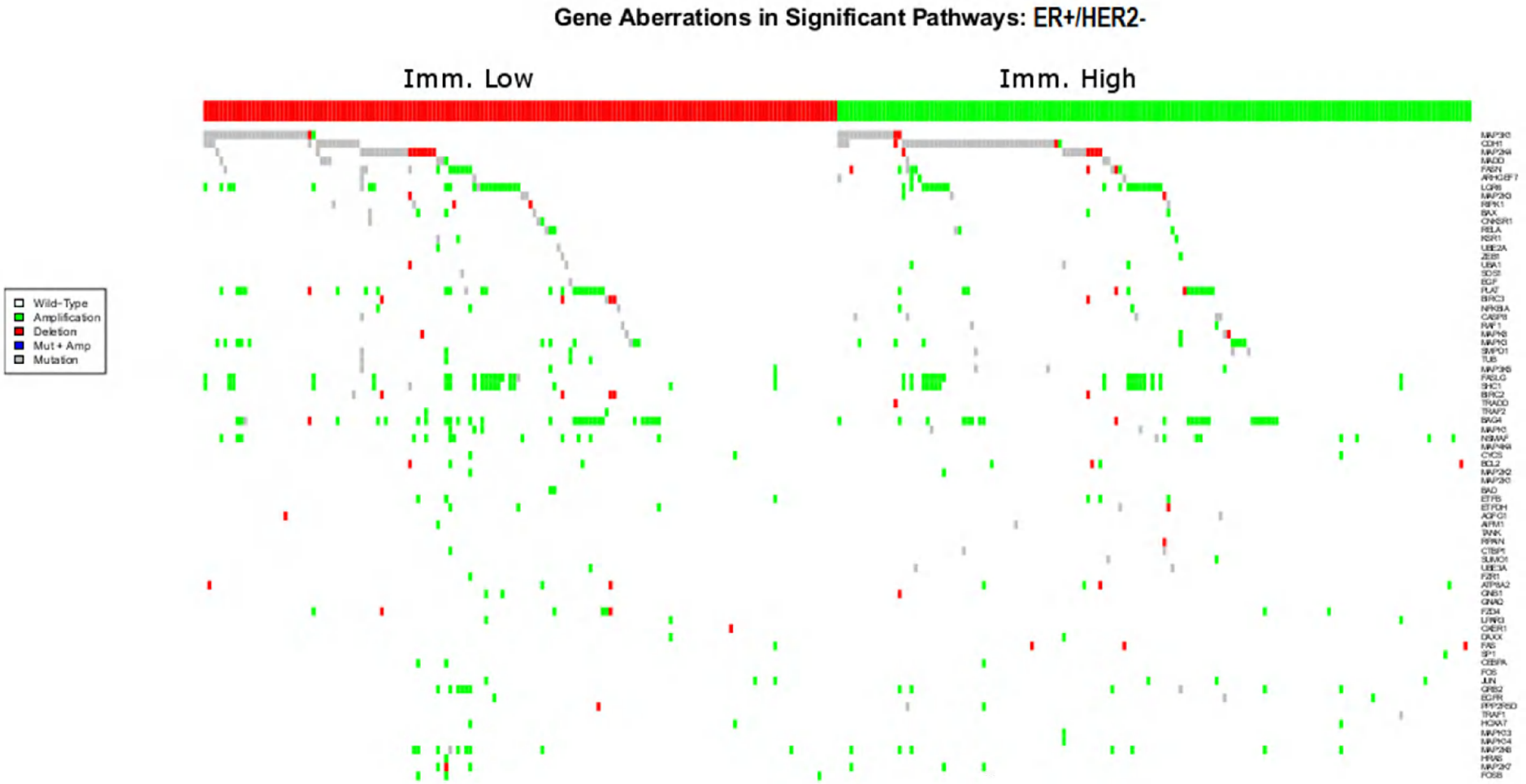

Immune Low Tertile
Immune High Tertile
Regulation.of.nuclear.beta.catenin.signaling.and.targ
Calcium.signaling.in.the.CD4..TCR.pathway
BIOCARTA_IL1R_PATHWAY
Validated.transcriptional.targets.of.AP1.family.memb
Stabilization.and.expansion.of.the.E.cadherin.adhere
FGF.signaling.pathway

Figure 11

(A)


Gene Aberrations in Significant Pathways: ER+/HER2-
Imm. Low
Imm. High
Wild-Type
Amplification
Deletion
Mut + Amp
Mutation

(B)

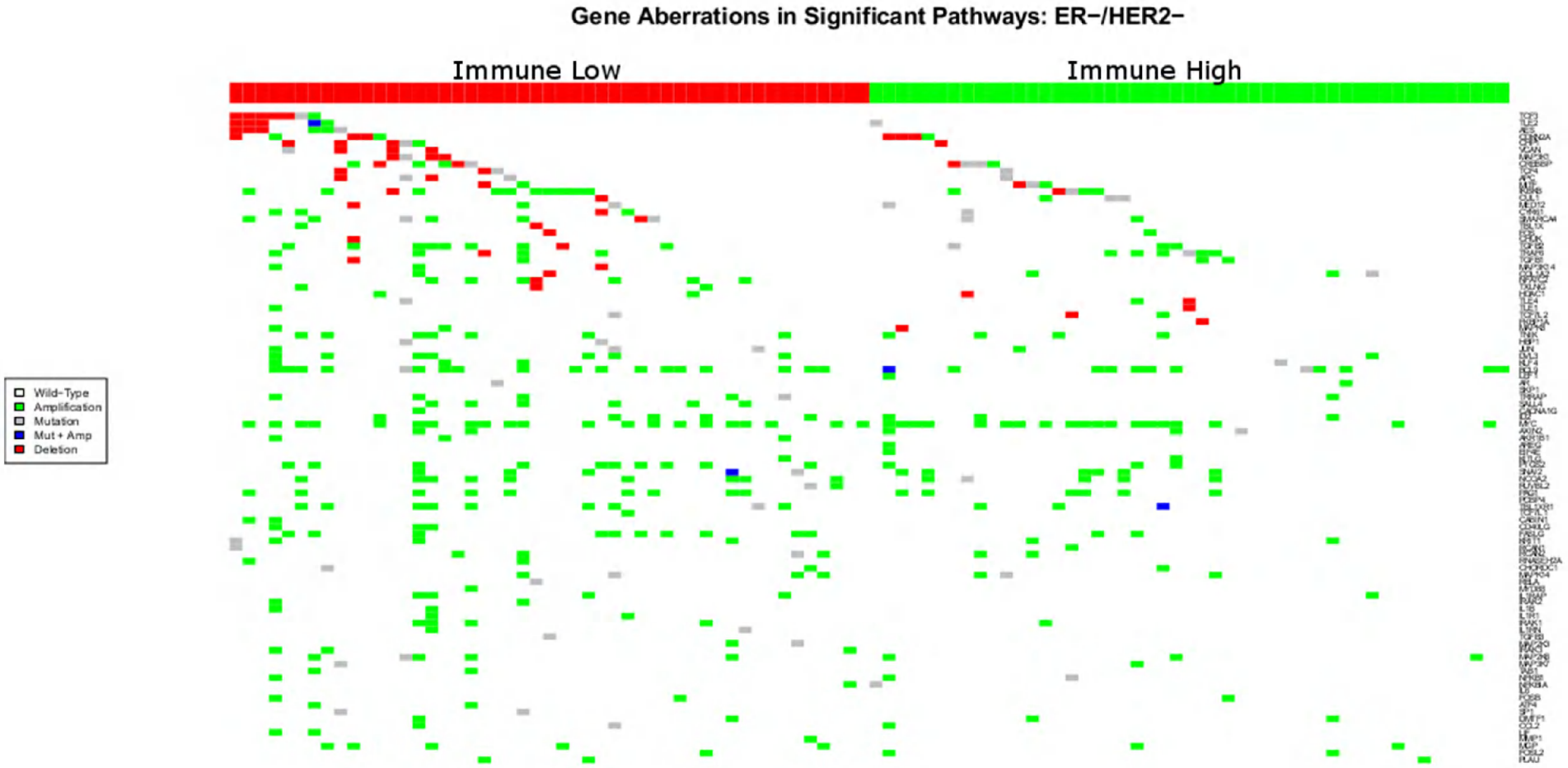

Gene Aberrations in Significant Pathways: ER−/HER2−
Immune Low
Immune High
Wild-Type
Amplification
Mutation
Mut + Amp
Deletion

Table 1

References	Callari	Rody	Rooney	Rooney	Rooney	NeoSphere (Bianchini)	Rooney	NeoSphere (Bianchini)	NeoSphere (Bianchini)	NeoSphere (Bianchini)		Rooney	Rooney
Immune-system component	T cell signatures				Activated CD8/NK cells	Antigen presenting cells	NK cells	Macrophages	Antigen Presentation	Interferon-related genes		T-cell regulation	
Name of the signature	ImmuneScore	LCK	Tfh	Tregs	Cytolytic	MHC2	NK	Macrophages	MHC1	STAT1	IF_I	Co_stimulation	Co_inhibition
	PRF1	ARHGAP15	CD200	FOXP3	GZMA	HLA-DMA	KLRF1	FUCA1	HLA-A	CXCL10	DDX58	CD2	BTLA
	GZMB	ARHGAP25	CXCL13	C15orf53	PRF1	HLA-DQB1	KLRC1	MMP9	HLA-B	CXCL11	HERC6	CD226	C10orf54
	CXCL13	CCL5	FBLN7	IL5		HLA-DRA		LGMN	HLA-C	GBP1	IFI44	CD27	CD160
	IRF1	CCR2	ICOS	CTLA4		HLA-DRB4		HS3ST2	HLA-F	STAT1	IFI44L	CD28	CD244
	IKZF1	CCR7	SGPP2	IL32				TM4SF19	HLA-G		IFIT1	CD40	CD274
	HLA-E	CD2	SH2D1A	GPR15				CLEC5A	HLA-J		IFIT2	CD40LG	CTLA4
		CD247	TIGIT	IL4				GPNMB			MX1	CD58	HAVCR2
		CD27 CD3D CD48	PDCD1					C11orf45 CD68 CYBB			OAS1 OAS3 RSAD2	CD70 ICOS ICOSLG	LAG3 LAIR1 LGALS9
		CD53										SLAMF1	PDCD1LG2
		CORO1A										TNFRSF18	PVRL3
		CSF2RB EVI2B FGL2 GIMAP4 GIMAP5 GMFG GZMA GZMK HCLS1										TNFRSF25 TNFRSF4 TNFRSF8 TNFRSF9 TNFSF14 TNFSF15 TNFSF18 TNFSF4 TNFSF8	TIGIT
		IL10RA										TNFSF9	
		IL2RG IL7R											
		INPP5D											
		IRF8											
		ITK											
		KLRK1											
		LCK											

References

1. Brahmer, J.R., et al., *Safety and activity of anti-PD-L1 antibody in patients with advanced cancer.* N Engl J Med, 2012. **366**(26): p. 2455-65.
2. Topalian, S.L., C.G. Drake, and D.M. Pardoll, *Immune checkpoint blockade: a common denominator approach to cancer therapy.* Cancer Cell, 2015. **27**(4): p. 450-61.
3. Drake, C.G., E.J. Lipson, and J.R. Brahmer, *Breathing new life into immunotherapy: review of melanoma, lung and kidney cancer.* Nat Rev Clin Oncol, 2014. **11**(1): p. 24-37.
4. Herbst, R.S., et al., *Predictive correlates of response to the anti-PD-L1 antibody MPDL3280A in cancer patients.* Nature, 2014. **515**(7528): p. 563-7.
5. Snyder, A., et al., *Genetic basis for clinical response to CTLA-4 blockade in melanoma.* N Engl J Med, 2014. **371**(23): p. 2189-99.
6. Hamlin, I.M., *Possible host resistance in carcinoma of the breast: a histological study.* Br J Cancer, 1968. **22**(3): p. 383-401.
7. Iwamoto, T., et al., *Gene pathways associated with prognosis and chemotherapy sensitivity in molecular subtypes of breast cancer.* J Natl Cancer Inst, 2011. **103**(3): p. 264-72.
8. Denkert, C., et al., *Tumor-associated lymphocytes as an independent predictor of response to neoadjuvant chemotherapy in breast cancer.* J Clin Oncol, 2010. **28**(1): p. 105-13.
9. Sabatier, R., et al., *Prognostic and predictive value of PDL1 expression in breast cancer.* Oncotarget, 2015. **6**(7): p. 5449-64.
10. Wimberly, H., et al., *PD-L1 Expression Correlates with Tumor-Infiltrating Lymphocytes and Response to Neoadjuvant Chemotherapy in Breast Cancer.* Cancer Immunol Res, 2015. **3**(4): p. 326-32.
11. Perez, E.A., et al., *Genomic analysis reveals that immune function genes are strongly linked to clinical outcome in the North Central Cancer Treatment Group n9831 Adjuvant Trastuzumab Trial.* J Clin Oncol, 2015. **33**(7): p. 701-8.
12. Loi, S., et al., *Prognostic and predictive value of tumor-infiltrating lymphocytes in a phase III randomized adjuvant breast cancer trial in node-positive breast cancer comparing the addition of docetaxel to doxorubicin with doxorubicin-based chemotherapy: BIG 02-98.* J Clin Oncol, 2013. **31**(7): p. 860-7.
13. Callari, M., et al., *Subtype-Specific Metagene-Based Prediction of Outcome after Neoadjuvant and Adjuvant Treatment in Breast Cancer.* Clin Cancer Res, 2015.
14. Sznol, M. and L. Chen, *Antagonist antibodies to PD-1 and B7-H1 (PD-L1) in the treatment of advanced human cancer--response.* Clin Cancer Res, 2013. **19**(19): p. 5542.
15. Rooney, M.S., et al., *Molecular and genetic properties of tumors associated with local immune cytolytic activity.* Cell, 2015. **160**(1-2): p. 48-61.
16. Lennerz, V., et al., *The response of autologous T cells to a human melanoma is dominated by mutated neoantigens.* Proc Natl Acad Sci U S A, 2005. **102**(44): p. 16013-8.
17. Heemskerk, B., P. Kvistborg, and T.N. Schumacher, *The cancer antigenome.* EMBO J, 2013. **32**(2): p. 194-203.
18. Brown, S.D., et al., *Neo-antigens predicted by tumor genome meta-analysis correlate with increased patient survival.* Genome Res, 2014. **24**(5): p. 743-50.
19. Segal, N.H., et al., *Epitope landscape in breast and colorectal cancer.* Cancer Res, 2008. **68**(3): p. 889-92.
20. Srihari, S., et al., *Understanding the functional impact of copy number alterations in breast cancer using a network modeling approach.* Mol Biosyst, 2016.
21. Hsiao, T.H., et al., *Identification of genomic functional hotspots with copy number alteration in liver cancer.* EURASIP J Bioinform Syst Biol, 2013. **2013**(1): p. 14.

22. Yang, Z., et al., *Integrated analyses of copy number variations and gene differential expression in lung squamous-cell carcinoma.* Biol Res, 2015. **48**: p. 47.
23. Madhavan, S., et al., *Genome-wide multi-omics profiling of colorectal cancer identifies immune determinants strongly associated with relapse.* Front Genet, 2013. **4**: p. 236.
24. Wang, K., et al., *MapSplice: accurate mapping of RNA-seq reads for splice junction discovery.* Nucleic Acids Res, 2010. **38**(18): p. e178.
25. Li, B. and C.N. Dewey, *RSEM: accurate transcript quantification from RNA-Seq data with or without a reference genome.* BMC Bioinformatics, 2011. **12**: p. 323.
26. Wilkerson, M.D., et al., *Integrated RNA and DNA sequencing improves mutation detection in low purity tumors.* Nucleic Acids Res, 2014. **42**(13): p. e107.
27. Mose, L.E., et al., *ABRA: improved coding indel detection via assembly-based realignment.* Bioinformatics, 2014. **30**(19): p. 2813-5.
28. Saunders, C.T., et al., *Strelka: accurate somatic small-variant calling from sequenced tumor-normal sample pairs.* Bioinformatics, 2012. **28**(14): p. 1811-7.
29. Radenbaugh, A.J., et al., *RADIA: RNA and DNA integrated analysis for somatic mutation detection.* PLoS One, 2014. **9**(11): p. e111516.
30. Larson, D.E., et al., *SomaticSniper: identification of somatic point mutations in whole genome sequencing data.* Bioinformatics, 2012. **28**(3): p. 311-7.
31. Koboldt, D.C., et al., *VarScan 2: somatic mutation and copy number alteration discovery in cancer by exome sequencing.* Genome Res, 2012. **22**(3): p. 568-76.
32. Mermel, C.H., et al., *GISTIC2.0 facilitates sensitive and confident localization of the targets of focal somatic copy-number alteration in human cancers.* Genome Biol, 2011. **12**(4): p. R41.
33. Cancer Genome Atlas, N., *Comprehensive molecular portraits of human breast tumours.* Nature, 2012. **490**(7418): p. 61-70.
34. Olshen, A.B., et al., *Circular binary segmentation for the analysis of array-based DNA copy number data.* Biostatistics, 2004. **5**(4): p. 557-72.
35. Ouedraogo, M., et al., *The duplicated genes database: identification and functional annotation of co-localised duplicated genes across genomes.* PLoS One, 2012. **7**(11): p. e50653.
36. Robin, X., et al., *pROC: an open-source package for R and S+ to analyze and compare ROC curves.* BMC Bioinformatics, 2011. **12**: p. 77.
37. Carpenter, J. and J. Bithell, *Bootstrap confidence intervals: when, which, what? A practical guide for medical statisticians.* Stat Med, 2000. **19**(9): p. 1141-64.
38. Rody, A., et al., *T-cell metagene predicts a favorable prognosis in estrogen receptor-negative and HER2-positive breast cancers.* Breast Cancer Res, 2009. **11**(2): p. R15.
39. Harrell, F.E. and C. Dupont, *Hmisc: Harrell miscellaneous. R package version 3.9-3.* Available at ht tp://CR AN. R-project. org/package= Hmisc, 2012.
40. Alter, J., D. Rozentzweig, and E. Bengal, *Inhibition of myoblast differentiation by tumor necrosis factor alpha is mediated by c-Jun N-terminal kinase 1 and leukemia inhibitory factor.* J Biol Chem, 2008. **283**(34): p. 23224-34.
41. Conze, D., et al., *c-Jun NH(2)-terminal kinase (JNK)1 and JNK2 have distinct roles in CD8(+) T cell activation.* J Exp Med, 2002. **195**(7): p. 811-23.
42. Loibl, S., et al., *PIK3CA mutations are associated with lower rates of pathologic complete response to anti–human epidermal growth factor receptor 2 (HER2) therapy in primary HER2-overexpressing breast cancer.* Journal of clinical oncology, 2014. **32**(29): p. 3212-3220.
43. Schramek, D., et al., *Direct in vivo RNAi screen unveils myosin IIa as a tumor suppressor of squamous cell carcinomas.* Science, 2014. **343**(6168): p. 309-13.
44. Liu, Y.C., *Ubiquitin ligases and the immune response.* Annu Rev Immunol, 2004. **22**: p. 81-127.

45. Heinen, S., et al., *Factor H-related protein 1 (CFHR-1) inhibits complement C5 convertase activity and terminal complex formation.* Blood, 2009. **114**(12): p. 2439-47.
46. Fritsche, L.G., et al., *An imbalance of human complement regulatory proteins CFHR1, CFHR3 and factor H influences risk for age-related macular degeneration (AMD).* Hum Mol Genet, 2010. **19**(23): p. 4694-704.
47. Goicoechea de Jorge, E., et al., *Dimerization of complement factor H-related proteins modulates complement activation in vivo.* Proc Natl Acad Sci U S A, 2013. **110**(12): p. 4685-90.
48. Thomson, D.M., W.J. Halliday, and K. Phelan, *Leukocyte adherence inhibition to myelin basic protein by cancer patients' T-lymphocytes in association with class II major histocompatibility antigens on monocytes.* J Natl Cancer Inst, 1985. **75**(6): p. 995-1003.
49. Katsara, M., et al., *Altered peptide ligands of myelin basic protein (MBP87-99) conjugated to reduced mannan modulate immune responses in mice.* Immunology, 2009. **128**(4): p. 521-33.
50. Al-Rawi, M.A., et al., *Aberrant expression of interleukin-7 (IL-7) and its signalling complex in human breast cancer.* Eur J Cancer, 2004. **40**(4): p. 494-502.
51. Swartz, M.A., *Immunomodulatory roles of lymphatic vessels in cancer progression.* Cancer Immunol Res, 2014. **2**(8): p. 701-7.
52. Bilal, E., et al., *Amplified loci on chromosomes 8 and 17 predict early relapse in ER-positive breast cancers.* PLoS One, 2012. **7**(6): p. e38575.
53. Fang, L., et al., *CCR7 regulates B16 murine melanoma cell tumorigenesis in skin.* J Leukoc Biol, 2008. **84**(4): p. 965-72.
54. Haimovitz-Friedman, A., R.N. Kolesnick, and Z. Fuks, *Ceramide signaling in apoptosis.* Br Med Bull, 1997. **53**(3): p. 539-53.
55. Cuvillier, O., et al., *Sphingosine generation, cytochrome c release, and activation of caspase-7 in doxorubicin-induced apoptosis of MCF7 breast adenocarcinoma cells.* Cell Death Differ, 2001. **8**(2): p. 162-71.
56. Young, M.R., *Eicosanoids and the immunology of cancer.* Cancer Metastasis Rev, 1994. **13**(3-4): p. 337-48.
57. Majumder, P. and J.M. Boss, *DNA methylation dysregulates and silences the HLA-DQ locus by altering chromatin architecture.* Genes Immun, 2011. **12**(4): p. 291-9.