Supplementary Online Content

Karn T, Jiang T, Hatzis C, et al. Immune sculpting of the triple-negative breast cancer genome. *JAMA Oncol.* Published online July 27, 2017. doi:10.1001/jamaoncol.2017.2140

eMethods. Supplementary Methods
eFigure 1. Strategy of RNA-Seq and Whole-Exome-Seq Analyses for TNBC Classification
eFigure 2. ER, PR, and HER2 Expression Assessed by RNA-Seq and Agilent Arrays
eFigure 3. Dependency of Platform Correlation on Gene Expression Level
eFigure 4. Correlation Between RNA-Seq and Affymetrix for Metagene Clusters
eFigure 5. Classification of TNBC (n = 208) Based on RNA-Seq Data
eFigure 6. Correlation of MHC2 Metagene Expression and Histological Quantification of TILs in TCGA Samples
eFigure 7. Classification Algorithm of the Prognostic Immune Signature
eFigure 8. Validation of Improved Prognosis of TNBC Patients With "Good Prognosis" Signature in RNA-Seq Data
eFigure 9. Mutational Count Distribution in 186 TNBC
eFigure 10. Relationship Between SCNA Levels and MATH in TNBC From TCGA
eFigure 11. Validation of Inverse Relationships Between Measures of Genomic Complexity and Immune Cell Infiltration in TNBC Using Different Immune Metagenes
eFigure 12. Prognostic Value of Histologically Quantified TILs in the TCGA TNBC Data Set
eFigure 13. Validation of Inverse Relationship of Genomic Heterogeneity and Immune Cell Infiltration Using Histologically Quantified TILs in the TCGA TNBC Data Set
eFigure 14. Differences in Mutation Count by Immune Cell Infiltration Metagenes and IL8/VEGF Metagene Expression Categories
eFigure 15. Correlation of the Number of Predicted Neoantigens and Mutational Load
eFigure 16. Independence of MATH Score and Total Mutation Counts
eFigure 17. Validation Analyses in METABRIC Data Set
eFigure 18. Differences in Mutation Count, Neoantigen Count, and CYT by Molecular Subtype in Breast Cancer
eFigure 19. Confounding of Molecular Breast Cancer Subtypes on Predicted Neoantigen Count and CYT
eFigure 20. High Intercorrelation of TIL Metagenes
eFigure 21. Association Between Clonal Heterogeneity and Immune Metagene Expression
eFigure 22. Association of MATH and SCNA With Prognostic Groups in TNBC
eTable 1. Annotated Cancer Genes Mutated in ≥3 Samples
eTable 2. TCGA Samples Included in the Study
eTable 3. "Cancer Genes" Curated by Vogelstein and colleagues
eTable 4. Individual Genes Constituting TNBC Metagenes and Their Correlation With Affymetrix Microarray

eMETHODS

Contents

eMETHODS 1

Data sources 2

Transfer of immune and stromal prognostic metagenes from the Affymetrix to RNA-seq data 3

Selection of a gene expression based TNBC cohort from TCGA 3

Gene filtering in RNA-Seq data 3

Metagene construction 3

Transfer of prognostic signature 4

Analysis of somatic mutations 4

Overall mutation counts 4

Differences in mutational load according to prognostic signature 4

Mutation frequencies of known cancer genes in prognostic groups 5

Analysis of tumor genomic heterogeneity 5

MATH scores and SCNA levels in TNBC subgroups by prognostic immune signature 6

Deconvolution of the effects of immune cell infiltration and tumor cellularity on genomic heterogeneity of TNBC 6

Validation by histological quantification of immune infiltration 6

Validation in independent METABRIC dataset 7

Relationship to other recent studies 7

References 8

eFIGURES 11

eTABLES 20

R-MarkDown document 26

All analyses were performed according to the "*REporting recommendations for tumour MARKer prognostic studies*" (REMARK) [16,17] . A diagram of the complete analytical strategy and the flow of patients through the study, including the number of patients included in each stage of the analysis, is given in **eFigure 1**. The R software environment (http://www.r-project.org/) using RStudio (www.rstudio.com) and IBM SPSS version 22.0 (http://www.ibm.com/) were used for all analyses. Chi square test was applied to assess associations between categorical parameters. All reported P values are two sided and $P < 0.05$ was considered significant. Detailed R-code and results of analyses are given in the accompanying R-MarkDown document (*R-MarkDown-document.pdf*). Code and data are also available at https://github.com/tkarn/TNBC-TIL .

Data sources

- Level 3 RNA-Seq V2 data for 20530 genes from 1215 treatment naïve BRCA samples processed on 2015-01-28 were downloaded from TCGA (https://tcga-data.nci.nih.gov/tcgafiles/ftp_auth/distro_ftpusers/anonymous/tumor/brca/cgcc/unc.edu/illuminahiseq_rnaseqv2/rnaseqv2/) and $\log_2(x+1)$ transformed (filename < *TCGA_BRCA_exp_HiSeqV2-2015-02-24.tgz* >). A current version (2016-08-16) of these data is available from UCSC Xena browser (http://xena.ucsc.edu/) TCGA hub (https://xenabrowser.net/datapages/?dataset=TCGA.BRCA.sampleMap/HiSeqV2&host=https://tcga.xenahubs.net) and can be dowloaded at (https://tcga.xenahubs.net/download/TCGA.BRCA.sampleMap/HiSeqV2 ; filename < *HiSeqV2* >) [18,19]. **eTable 2** catalog the TCGA samples included in the study.

- Agilent 244K custom gene expression G4502A_07_3 microarrays by the University of North Carolina TCGA genomic characterization center for 597 samples were obtained from UCSC cancer genome browser ((https://genome-cancer.ucsc.edu/proj/site/hgHeatmap ; filename < *TCGA_BRCA_G4502A_07_3-2015-02-24.tgz* >). A current copy is available at UCSC Xena (https://xenabrowser.net/datapages/?dataset=TCGA.BRCA.sampleMap/AgilentG4502A_07_3&host=https://tcga.xenahubs.net).

- Gene-level mutation data (nonsilent somatic mutation; wustl curated) for 982 samples were obtained from UCSC cancer genome browser processed on 2015-01-27 ((https://genome-cancer.ucsc.edu/proj/site/hgHeatmap ; filename < *TCGA_BRCA_mutation_curated_wustl_gene-2015-02-24.tgz* >) and can be downloaded from https://github.com/tkarn/TNBC-TIL. A current copy is available at UCSC Xena (https://xenabrowser.net/datapages/?dataset=TCGA.BRCA.sampleMap/mutation_curated_wustl_gene&host=https://tcga.xenahubs.net).

- Updated clinical data and sample information for TCGA samples were obtained from cBIO portal (www.cbioportal.org) using the R library *cgdsr* [20].

- Mutant variant allele frequencies (vaf) of all genes for MATH score calculation were obtained from file < genome.wustl.edu_BRCA.IlluminaGA_DNASeq.Level_2.1.1.0.curated.somatic.maf.txt > of 2014-Feb-08 from the TCGA Data Portal (https://tcga-data.nci.nih.gov/tcgafiles/ftp_auth/distro_ftpusers/anonymous/). A copy of the file is available from GDC (https://portal.gdc.cancer.gov/legacy-archive/files/50d6fb1d-5bb1-4a30-9e91-6d45bd9b1c3f). The vaf data are also available at https://github.com/tkarn/TNBC-TIL.

- The numbers of predicted neo-epitopes based on tumor-specific HLA typing for each patient [21] were obtained for 760 BRCA samples from Supplementary Table S4 from Rooney et al. [5] available at http://www.ncbi.nlm.nih.gov/pmc/articles/PMC4856474/bin/NIHMS717941-supplement-Table_S4.xlsx

- Somatic copy number alteration (SCNA) levels were obtained for 941 BRCA samples from Supplementary Table S7 from Davoli et al.[8] available at http://science.sciencemag.org/highwire/filestream/689461/field_highwire_adjunct_files/7/aaf8399-Davoli-SM-table-S7.xlsx

- Histological quantification of mononuclear cells from sections of 180 BRCA samples from TCGA were obtained from Supplementary Table S1 from Lehmann et al. [22] available at

http://journals.plos.org/plosone/article/file?type=supplementary&id=info:doi/10.1371/journal.pone.0157368.s006

- Spearman correlation values between RNA-Seq and Affymetrix microarray for 16,097 Jetset probes for 57 paired frozen breast cancer samples was obtained from Supplementary Table S2 of Fumagalli et al.[23] and are available at https://github.com/tkarn/TNBC-TIL

- Mutational profiles and patterns of immune infiltration of the METABRIC dataset from Pereira et al.[24] and Ali et al.[25] were obtained from http://github.com/cclab-brca

- R-code and respective data files are available at https://github.com/tkarn/TNBC-TIL

Transfer of immune and stromal prognostic metagenes from the Affymetrix to RNA-seq data

We transferred triple negative breast cancer (TNBC) prognostic and immune gene expression classification based on DNA microarray data to RNA-Seq data from the TCGA using the following steps (**eFigure 1**):

Selection of a gene expression based TNBC cohort from TCGA

We first compared RNA-Seq and Agilent array expression results for estrogen (ESR1) and progesterone (PR) receptors and HER2 mRNA to ensure that we can use RNA-Seq to identify TNBC from TCGA (**eFigure 2 A-C**). We identified n=208 TNBC from the TCGA based on bimodal distribution of RNA-Seq for ESR1 and HER2 (**eFigure 2 D-F**), full details in Section-1 in *R-MarkDown-document.pdf*.

Gene filtering in RNA-Seq data

To reduce noise from the RNA-Seq data we applied a gene filtering step. A recent report [23] provides Spearman correlation coefficients for expression of 16097 genes measured both on Affymetrix microarray and RNA-Seq platform in a series of 57 breast cancers. We analyzed these correlation scores based on median expression of the genes among our sample set of 208 TNBC from TCGA and found a poor correlation for genes with median expression in the lower quartile (<2 in $\log_2$ expression space) (**eFigure 3**), (see Section-2 in *R-MarkDown-document.pdf*). Based on that criterium we selected 15876 of 20530 genes from RNA-Seq data with $\log_2$ expression values above 2 in at least 50% of the 208 samples.

Metagene construction

We performed unsupervised hierarchical clustering of the 15876 genes based on RNA-Seq data from the 208 TNBC samples (with single linkage and Pearson correlation as distance metric) using GenePattern [26] (http://genepattern.broadinstitute.org) and confirmed co-expression clusters of 15 metagenes including a total of 304 distinct genes (**eTable 4**). Median correlation between RNA-Seq and Affymetrix for the 304 genes was 0.88 (**eFigure 4**) (Section-3.1 in *R-MarkDown-document.pdf*). These co-expression clusters correspond to the previously reported Affymetrix expression data based TNBC phenotypes and immune and stromal metagenes, including Basal-like, Molecular-Apocrine, Claudin, Proliferation, B-cell, T-cell, MHC1, MHC2, IFN, Collagen/Stroma, Endothel, Histone, HOXA metagenes [9]. The expression of inflammation (IL8, CXCL1) and angiogenesis markers (VEGFA, adrenomeddulin, ANGPTL4 a.o.), were highly correlated in this dataset and therefore were combined into a single metagene IL8-VEGF. Metagene values were calculated as the average expression of all member genes (Section-3.2 in *R-MarkDown-document.pdf*). The metagenes allowed to reproduce the previously described classification of TNBC into several subtypes (**eFigure 5**) [9]. As an alternative method to calculate immune cell activity based on gene expression, we also used the

CYT gene expression score score (geometric mean of GZMA and PRF1 mRNA expression) as described by Rooney and colleagues[5].

Transfer of prognostic signature

Different immune metagenes (e.g. T-Cell, B-Cell, MHC2, CYT) are highly correlated in breast cancer gene expression data[3,5,14,27], and because of that may all be similarily applicable as surrogates for a T_H1 type immune response[4]. In contrast, inflammation markers [4,9,10] and myeloid cell markers[28] may capture the signal of tumor promoting inflammation and immunosuppressive factors.

We have previously shown that extensive immune infiltration reflected by high expression of either B-cell, or MHC2, or T-cell metagenes together with low expression of inflammatory metagenes (IL8, and VEGF) define a group of good prognosis TNBC [9,10,29]. We tested if these metagenes retained prognostic function in the RNA-Seq data from TCGA. Since distributions of the immune metagenes did not show bimodality which would have allowed clear cutoff selection[30] we stayed with previously used cutoffs values to avoid overfitting of the data (Section-5 in *R-MarkDown-document.pdf*). We used two approaches to transfer the signatures from microarray to RNA-Seq. First, we considered the highest quartile of MHC2 metagene expression to represent high immune infiltration and high MHC2 expression[27] in combination with low (below the median) IL8/VEGF metagene expression[9] to define the good prognosis category (**eFigure 7** for signature definition), full details in Section-5.1 in *R-MarkDown-document.pdf*. The upper quartile cutoff for MHC2 metagene has been used in one of our previous papers based on earlier histological TIL data[27]. A robust median split of the IL8/VEGF metagene has also been applied before[9]. This method assigned 25 cases to good prognosis and these patients had significantly better survival compared to the poor prognosis group (P=0.019, **eFigure 8B**). As a second strategy, we used B-cell metagene expression above the lowest quartile and IL8-VEGF metagene expression below the median to categorize a case as good prognosis (in this case we applied the original cutoffs from our previous description of this signature in the microarray dataset[9]) (**eFigure 8C**) (Section-5.2 in *R-MarkDown-document.pdf*). We only observed a trend for increased survival in the respective group of 76 samples in the TCGA in Kaplan-Meier analysis (P=0.22 log rank test, **eFigure 8C**). Many studies have demonstrated that different immune metagenes are highly correlated as shown for the MHC2 and T-cell metagenes in **eFigure 20** contributing redundant information (R^2=0.807). Since the prognostic value of these immune metagenes has been repeatedly and independently shown in many datasets, the modest effect of our second strategy involving the B-cell metagene in the TCGA data may be related to the limited power of the TCGA data set due to short median follow up of 24 months and only 29 events in 193 cases [31]. In our previous microarray study 139 events occurred in 402 TNBC with a median follow up of 60 months. To assure robustness in our analysis, we used both prognostic classifications methods when we performed correlation with genomic metrics and report results for both.

Analysis of somatic mutations

Overall mutation counts

Matching whole exome sequencing (WES) data was available for 186 of the 208 TNBC samples. For these we obtained curated somatic mutation data from UCSC cancer browser (see above section *Data Sources* for details). The median number of non-silent mutations was 53 per tumor (range 1-1138) (**eFigure 9**). The most frequently mutated gene was TP53 (74.7%). Six samples displayed a hypermutated phenotype with more than 300 mutated genes. Inclusion or exclusion of these six cases did not change the results, indicating that our results are not driven by these rare outliers with high mutation rate.

Differences in mutational load according to prognostic signature

When we compared total mutation counts at gene level (in 39741 curated genes/pseudogenes from WES), good prognosis TNBC had significantly lower mutation count for the two signatures using MHC2 (P=0.021, **Figure 2D**) and B-cell (P=0.014), respectively**.** This difference was primarily driven

by the lower mutation count in samples with high immune infiltration (i.e high MHC2 (P=0.003) or B-cell (P=0.018) metagene expression (**eFigure 14).** There was no significant difference in mutation counts by IL8-VEGF metagene expression levels or by TNBC molecular subtype (i.e. apocrine, basal-like or claudin-low, P=0.7, not shown). We performed the same analyses using the predicted neo-antigen load obtained from [5] and obtained similar, significant inverse association between prognostic category and neo-antigen load (**Figure 2D**). We noted that mutation count and neoantigen load are highly correlated (**eFigure 15**). However, since the number of predicted neoantigens is nearly a magnitude smaller than mutation counts, the inverse association associations with immune cell metagens were no longer significant for neo-antigen load. This is likely due to the reduced power. **eFigure 18** presents mutation count, neoantigen count, and CYT metagene expression in different breast cancer molecular subtypes. The results illustrate that the previously reported association of higher neo-antigen load and higher CYT metagene expression in a combined analysis of all breast cancers is mainly driven by the differential distribution of neo-antigen load and immune cell infiltration between the different breast cancer molecular subtypes (P<0.001, Mann-Whitney U-Test, for all three parameters). When TNBC is analyzed separately, neo-antigen load and CYT metagene expression is significantly inversely correlated.

Mutation frequencies of known cancer genes in prognostic groups

We examined the frequency of mutations in 119 annotated "cancer genes" (**eTable 3**) curated by Vogelstein and colleagues [12,13] in the 186 TNBC. 45 genes were mutated in at least three samples (**eTable 1**). We analyzed whether we observed differences in mutation frequencies among these 45 genes between TNBC classified as "Good" or "Poor" prognosis by the MHC2/IL8-VEGF gene signature. The only differentially altered gene with nominal significance was CASP8 with 2 of 25 samples mutated in the "Good Prognosis" group and only 1 of 161 samples mutated in the "Poor Prognosis" group (P=0.007, Chi^2 test without adjustment for multiple testing). When we used the B-cell/IL8-VEGF signature to assign prognosis, all 3 cases with CASP8 mutation were in the "Good Prognosis" group (n=76) (P=0.037, Chi^2 test). In addition, another 3 cases with mutation in the TSC2 gene were also found in this subgroup (P=0.037, Chi^2 test). We next studied enrichment of mutations in 12 different "cancer pathways" to which Vogelstein and colleagues have assigned the 119 cancer genes. There was no significantly differentially mutated pathway between good and poor prognosis TNBC.

Analysis of tumor genomic heterogeneity

Intratumor clonal heterogeneity can be inferred from genomic sequence data by several different methods including PyClone, SciClone, or EXPANDS [32–35]. Unfortunately these methods are suboptimal for breast cancer data because they detect only one, or very few, subclonal populations in the majority of breast cancers [7]. We therefore adopted another method, MATH (mutant allele tumor heterogeneity), which uses the broadness of the distribution of mutant allele frequencies as a measure of mixed cell population of the tumor [11,36]. The MATH score is calculated as the median absolute deviation of each somatic mutation's allelic fraction from the median allelic fraction for all mutations in the tumor, divided by the median variant allelic fraction (vaf). The use of a ratio corrects for the confounding effect of normal tissue in the sample. A detailed description of the method can be found in the article [11]. MATH uses the overall variance of the VAF distribution to approximate clonal heterogeneity, however this metric is influenced by the combined effect of clonality and CNAs. We calculated MATH values for TCGA samples from mutant variant allele frequencies (vaf) of all genes (see above section *Data Sources* for details). Median vaf and MAD of all mutated genes were calculated for each tumor sample and MATH calculated as 100*MAD/median (Section-4 in *R-MarkDown-document.pdf*).

We also analyzed the level of somatic copy number alterations (SCNA) in samples stratified by the prognostic signature. TNBC harbor larger numbers of SCNAs than other breast cancer subtypes [37–39] and their ability to generate immunogenic epitopes has been suggested [40]. SCNA levels for TCGA

samples were obtained from a recent publication[8]. Both single nucleotide variants (SNV) and SCNA can effect MATH scores. However, MATH scores and mutational load did not correlate (**eFigure 16**), and the correlation between MATH and SCNA levels was only limited (**eFigure 10**). This suggests that all three parameters can provide independent information on tumor and clonal heterogeneity.

MATH scores and SCNA levels in TNBC subgroups by prognostic immune signature

MATH scores were significantly lower in good prognosis TNBC compared to poor prognosis (**eFigure 22B**, P=0.001, Mann-Whitney U-Test). This difference was driven by a strong inverse relationship between the MATH score and immune metagene expression (**Figure 2B**). There was no correlation between MATH and the IL8-VEGF metagene expression, but there was a significantly lower MATH score in MHC2 or B-cell or T-cell high TNBC (**eFigure 21**). Within particularly the good prognosis group, defined by high MHC2 and low IL8/VEGFR, there was a strong negative correlation between T cell infiltration measured by the T-cell metagene and clonal heterogeneity (R^2=0.479, **Figure 3B**, P<0.001). When MATH scores were plotted against the T-cell metagene in the poor prognosis group, we continued to observe a weak negative correlation (**Figure 3A**). Analogous results were obtained for SCNA levels (**eFigure 22C**, **Figure 2C, 3C, 3D,** respectively). The inverse relationship between immune infiltration and both MATH score and SCNA levels, respectively, was validated using different surrogate measures for immune cell infiltration as MHC2 or B-cell metagenes[9] and the CYT metric[5] (**eFigure 11**). In contrast no significant correlation with expression of the IL8-VEGF metagene was detected.

Deconvolution of the effects of immune cell infiltration and tumor cellularity on genomic heterogeneity of TNBC

Since tumor purity can affect the power of mutation calling it could be a confounding factor in our analysis. While MATH calculation inherently controls for the amount of normal cells, tumor cellularity may effect mutational and neoantigen count data. Even if a requirement for TCGA samples was tumor cellularity of ≥50% and most normal cell contamination may be non-TIL stroma we tried to control for read depth and for tumor purity in multivariate analysis of genomic metrics. We calculated both the median values of variant reads for all mutated genes for each tumor and obtained tumor purity estimates from ASCAT[37]. In multivariate regression of MATH only T-cell metagene expression was significant (P=0.003) as well as the interaction term (P=0.019), but not tumor purity (P=0.357). In a model including T-cell metagene and median read depth both parameters were significant (P<0.001) but the contribution of T-cell metagene expression to Chi square was four fold higher than for read depth. Similar to the results for MATH, we also obtained for SCNA levels independent significance of the T-cell metagene (P<0.001) in multivariate models with purity estimates from ASCAT or median read depth. For lower mutation and neoantigen load we were not able to fully deconvolute whether high immune cell infiltration or the consequently resulting lower tumor cellularity may be the major driving force, since ASCAT derived surrogate for tumor cellularity is also negatively correlated with mutation and neoantigen counts.

Validation by histological quantification of immune infiltration

TIL quantification by immune gene signatures from bulk tumor biopsies may be expected to be more reliable than histological quantification from single sections. Nevertheless, studies have demonstrated reproducibility of histological quantification[41] and good correlation with molecular methods[42]. Thus, we also analyzed data on histologically quantification of TILs in TCGA (see above section *Data Sources* for details). As shown in **eFigure 6** we observed a strong correlation of expression of the MHC2 metagene with histologically quantified TILs (R^2=0.367) and a trend for a better prognosis in samples with ≥50% TILs (P=0.127, **eFigure 12**). Moreover, the negative relationship between immune cell infiltration and tumor heterogeneity was also highly significant (**eFigure 13** for both MATH and SCNA levels, respectively).

Validation in independent METABRIC dataset

We also performed a validation of our results from the TCGA data in the independent METABRIC dataset[24,25] (see above section *Data Sources* for details). We calculated MATH scores based on only a small panel of 173 sequenced genes from the METABRIC dataset. Thus we expect a lower precision of the corresponding MATH score. Still we observed a highly significant (P=1e-6) negative correlation between MATH and the CYT metric for immune cell infiltration from gene expression (Spearman's rho= -0.286, **eFigure 17A**). Since no SCNA levels were available for this dataset we compared the fraction of the genome affected by CNAs[24] to the CYT metric and observed a trend for a negative correlation (rho=-0.104, P=0.138; **eFigure 17B**). Interestingly, we also detected significantly lower MATH scores (P<0.001) in samples from the "integrative cluster" IntClust4- (which contains most TNBC with TILs) as compared to IntClust10 (TNBC with no TILs, **eFigure 17C**).

Relationship to other recent studies

Two recent pan-cancer genome studies reported results in line with our hypothesis: One large pan-cancer study on lymphocyte infiltration[5] observed a lower than expected number of predicted neoantigens in some cancer types suggesting immune-mediated elimination. In our dataset this ratio between observed vs. predicted neoantigens was somewhat lower in the good prognosis TNBC group (0.84, SD 0.41, n=19) than in the poor prognosis group (1.03, SD 0.57, n=111), but this difference was not significant (P=0.15, T-test). Despite only such large studies allow sound statistical proof of associations in the long tailed distributions of mutations (e.g. CASP8 mutation), inclusion of different tumor types can lead to confounding effects. We show this in **eFigures 18 and 19** for comparisons across different breast cancer subtypes. Therefore we based our study only on a single subtype of breast cancer (TNBC). A second pan-cancer study observed enrichment of immune infiltration in tumors types that are characterized by lower intratumor heterogeneity (defined as number of subclonal populations from PyClone) and suggested that this may, in part, reflect results of immunoediting[7].

A recent Science paper[43] presents data that are all in perfect agreement with our model. However, the authors of this article follow a distinct hypothesis: They suggest that only clonal neoantigens elicit an immune response and put the "cause" of the association on the cancer cells, while our hypothesis suggests that the "source" of the observed association may be the immune systems' effect on the tumor. Similarily, a very recent pan-cancer study in Nature[8] observed less SCNA in tumors with immune infiltration measured by gene signatures. The authors of that study conclude that aneuploidy adversely affects immune cell action. We have also used data from that study in our analyses leading to similar correlations (**Figure 2 and 3**). Several hypothetical models may explain how aneuploidy might create an immune suppressing microenvironment. However, according to our model (**Figure 1**) the situation would be vice versa with the stage of immunoediting (equilibrium/escape) "responsible" for the clonal heterogeneity and the observed aneuploidy of the tumor. We think that it is important to view the immunological and clonal stage of a tumor not as an endpoint but as a snapshot in time of evolution of the tumor-immune interaction. This view would be highly important when interpreting such biomarkers and their relationship with response to new immunological therapies.

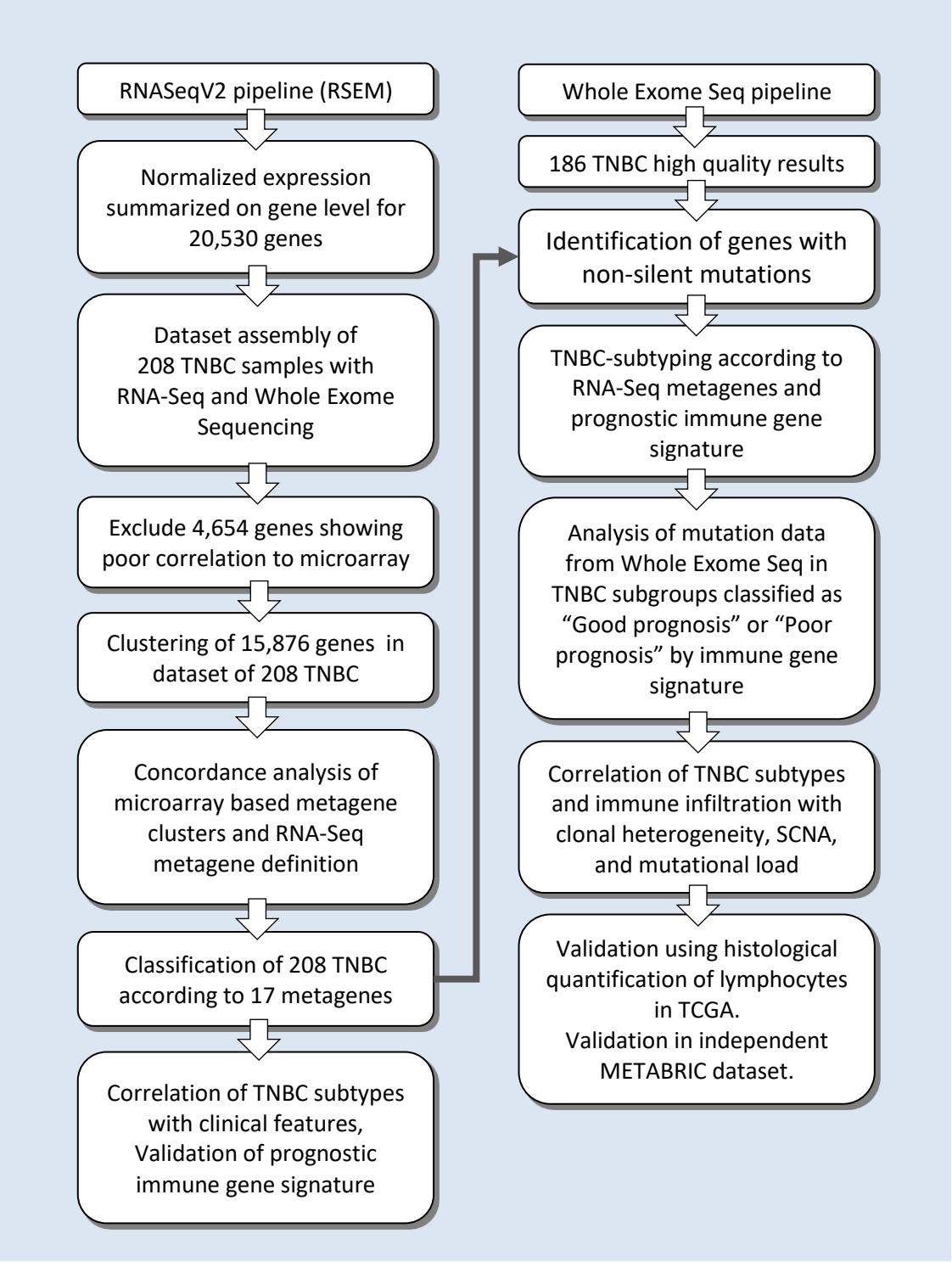
References

1. Pusztai L, Karn T, Safonov A, Abu-Khalaf MM, Bianchini G. New Strategies in Breast Cancer: Immunotherapy. *Clin. Cancer Res.* 2016. doi:10.1158/1078-0432.CCR-15-1315.
2. Schreiber RD, Old LJ, Smyth MJ. Cancer immunoediting: integrating immunity's roles in cancer suppression and promotion. *Science*. 2011;331(6024):1565-1570. doi:10.1126/science.1203486.
3. Bianchini G, Qi Y, Alvarez RH, et al. Molecular Anatomy of Breast Cancer Stroma and Its Prognostic Value in Estrogen Receptor-Positive and -Negative Cancers. *Journal of Clinical Oncology*. 2010;28(28):4316-4323. doi:10.1200/JCO.2009.27.2419.
4. Karn T, Pusztai L, Rody A, Holtrich U, Becker S. The Influence of Host Factors on the Prognosis of Breast Cancer: Stroma and Immune Cell Components as Cancer Biomarkers. *Current Cancer Drug Targets*. 2015;15(8):652-664.
5. Rooney MS, Shukla SA, Wu CJ, Getz G, Hacohen N. Molecular and genetic properties of tumors associated with local immune cytolytic activity. *Cell*. 2015;160(1-2):48-61. doi:10.1016/j.cell.2014.12.033.
6. Brown SD, Warren RL, Gibb EA, et al. Neo-antigens predicted by tumor genome meta-analysis correlate with increased patient survival. *Genome Res.* 2014;24(5):743-750. doi:10.1101/gr.165985.113.
7. Morris LGT, Riaz N, Desrichard A, et al. Pan-cancer analysis of intratumor heterogeneity as a prognostic determinant of survival. *Oncotarget*. 2016;7(9):10051-10063. doi:10.18632/oncotarget.7067.
8. Davoli T, Uno H, Wooten EC, Elledge SJ. Tumor aneuploidy correlates with markers of immune evasion and with reduced response to immunotherapy. *Science*. 2017;355(6322). doi:10.1126/science.aaf8399.
9. Rody A, Karn T, Liedtke C, et al. A clinically relevant gene signature in triple negative and basal-like breast cancer. *Breast Cancer Res*. 2011;13(5):R97. doi:10.1186/bcr3035.
10. Karn T, Pusztai L, Holtrich U, et al. Homogeneous Datasets of Triple Negative Breast Cancers Enable the Identification of Novel Prognostic and Predictive Signatures. *PLoS ONE*. 2011;6(12):e28403. doi:10.1371/journal.pone.0028403.
11. Mroz EA, Rocco JW. MATH, a novel measure of intratumor genetic heterogeneity, is high in poor-outcome classes of head and neck squamous cell carcinoma. *Oral Oncol*. 2013;49(3):211-215. doi:10.1016/j.oraloncology.2012.09.007.
12. Kandoth C, McLellan MD, Vandin F, et al. Mutational landscape and significance across 12 major cancer types. *Nature*. 2013;502(7471):333-339. doi:10.1038/nature12634.
13. Vogelstein B, Papadopoulos N, Velculescu VE, Zhou S, Diaz LA, Kinzler KW. Cancer genome landscapes. *Science*. 2013;339(6127):1546-1558. doi:10.1126/science.1235122.
14. Safonov A, Jiang T, Bianchini G, et al. Immune gene expression is associated with genomic aberrations in breast cancer. *Cancer Res*. 2017. doi:10.1158/0008-5472.CAN-16-3478.
15. Jiang T, Shi W, Wali VB, et al. Predictors of Chemosensitivity in Triple Negative Breast Cancer: An Integrated Genomic Analysis. *PLoS Med*. 2016;13(12):e1002193. doi:10.1371/journal.pmed.1002193.
16. McShane LM, Altman DG, Sauerbrei W, et al. Reporting recommendations for tumor marker prognostic studies. *J Clin Oncol*. 2005;23:9067-9072.
17. Simon RM, Paik S, Hayes DF. Use of archived specimens in evaluation of prognostic and predictive biomarkers. *J. Natl. Cancer Inst.* 2009;101(21):1446-1452. doi:10.1093/jnci/djp335.
18. Zhu J, Sanborn JZ, Benz S, et al. The UCSC Cancer Genomics Browser. *Nat Methods*. 2009;6(4):239-240. doi:10.1038/nmeth0409-239.

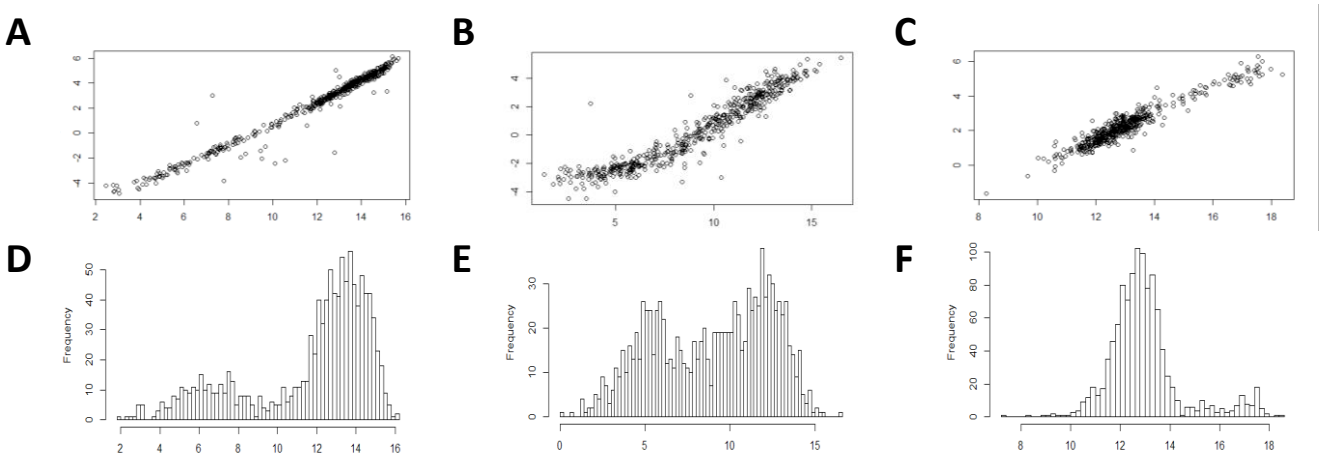
19. Cline MS, Craft B, Swatloski T, et al. Exploring TCGA Pan-Cancer data at the UCSC Cancer Genomics Browser. *Sci Rep*. 2013;3:2652. doi:10.1038/srep02652.
20. Cerami E, Gao J, Dogrusoz U, et al. The cBio cancer genomics portal: an open platform for exploring multidimensional cancer genomics data. *Cancer Discov*. 2012;2(5):401-404. doi:10.1158/2159-8290.CD-12-0095.
21. Rajasagi M, Shukla SA, Fritsch EF, et al. Systematic identification of personal tumor-specific neoantigens in chronic lymphocytic leukemia. *Blood*. 2014;124(3):453-462. doi:10.1182/blood-2014-04-567933.
22. Lehmann BD, Jovanovic B, Chen X, et al. Refinement of Triple-Negative Breast Cancer Molecular Subtypes: Implications for Neoadjuvant Chemotherapy Selection. *PLoS ONE*. 2016;11(6):e0157368. doi:10.1371/journal.pone.0157368.
23. Fumagalli D, Blanchet-Cohen A, Brown D, et al. Transfer of clinically relevant gene expression signatures in breast cancer: from Affymetrix microarray to Illumina RNA-Sequencing technology. *BMC Genomics*. 2014;15:1008. doi:10.1186/1471-2164-15-1008.
24. Pereira B, Chin S-F, Rueda OM, et al. The somatic mutation profiles of 2,433 breast cancers refines their genomic and transcriptomic landscapes. *Nat Commun*. 2016;7:11479. doi:10.1038/ncomms11479.
25. Ali HR, Chlon L, Pharoah PDP, Markowetz F, Caldas C. Patterns of Immune Infiltration in Breast Cancer and Their Clinical Implications: A Gene-Expression-Based Retrospective Study. *PLoS Med*. 2016;13(12):e1002194. doi:10.1371/journal.pmed.1002194.
26. Reich M, Liefeld T, Gould J, Lerner J, Tamayo P, Mesirov JP. GenePattern 2.0. *Nat Genet*. 2006;38(5):500-501. doi:10.1038/ng0506-500.
27. Rody A, Holtrich U, Pusztai L, et al. T-cell metagene predicts a favorable prognosis in estrogen receptor-negative and HER2-positive breast cancers. *Breast Cancer Res*. 2009;11(2):R15. doi:10.1186/bcr2234.
28. Ruffell B, Au A, Rugo HS, Esserman LJ, Hwang ES, Coussens LM. Leukocyte composition of human breast cancer. *Proc. Natl. Acad. Sci. U.S.A.* 2012;109(8):2796-2801. doi:10.1073/pnas.1104303108.
29. Hanker LC, Rody A, Holtrich U, et al. Prognostic evaluation of the B cell/IL-8 metagene in different intrinsic breast cancer subtypes. *Breast Cancer Res. Treat.* 2013;137(2):407-416. doi:10.1007/s10549-012-2356-2.
30. Karn T, Metzler D, Ruckhäberle E, et al. Data driven derivation of cutoffs from a pool of 3,030 Affymetrix arrays to stratify distinct clinical types of breast cancer. *Breast Cancer Res Treat*. 2010;120(3):567-579. doi:10.1007/s10549-009-0416-z.
31. Iglesia MD, Vincent BG, Parker JS, et al. Prognostic B-cell Signatures Using mRNA-Seq in Patients with Subtype-Specific Breast and Ovarian Cancer. *Clin. Cancer Res.* 2014. doi:10.1158/1078-0432.CCR-13-3368.
32. Roth A, Khattra J, Yap D, et al. PyClone: statistical inference of clonal population structure in cancer. *Nat Methods*. 2014;11(4):396-398. doi:10.1038/nmeth.2883.
33. Shah SP, Roth A, Goya R, et al. The clonal and mutational evolution spectrum of primary triple-negative breast cancers. *Nature*. 2012. doi:10.1038/nature10933.
34. Miller CA, White BS, Dees ND, et al. SciClone: inferring clonal architecture and tracking the spatial and temporal patterns of tumor evolution. *PLoS Comput Biol*. 2014;10(8):e1003665. doi:10.1371/journal.pcbi.1003665.
35. Andor N, Graham TA, Jansen M, et al. Pan-cancer analysis of the extent and consequences of intratumor heterogeneity. *Nat. Med.* 2016;22(1):105-113. doi:10.1038/nm.3984.

36. Mroz EA, Tward AD, Tward AM, Hammon RJ, Ren Y, Rocco JW. Intra-tumor genetic heterogeneity and mortality in head and neck cancer: analysis of data from the Cancer Genome Atlas. *PLoS Med*. 2015;12(2):e1001786. doi:10.1371/journal.pmed.1001786.
37. van Loo P, Nordgard SH, Lingjærde OC, et al. Allele-specific copy number analysis of tumors. *Proceedings of the National Academy of Sciences of the United States of America*. 2010;107(39):16910-16915. doi:10.1073/pnas.1009843107.
38. Curtis C, Shah SP, Chin S-F, et al. The genomic and transcriptomic architecture of 2,000 breast tumours reveals novel subgroups. *Nature*. 2012;486(7403):346-352. doi:10.1038/nature10983.
39. Ciriello G, Miller ML, Aksoy BA, Senbabaoglu Y, Schultz N, Sander C. Emerging landscape of oncogenic signatures across human cancers. *Nat. Genet.* 2013;45(10):1127-1133. doi:10.1038/ng.2762.
40. Schumacher TN, Schreiber RD. Neoantigens in cancer immunotherapy. *Science*. 2015;348(6230):69-74. doi:10.1126/science.aaa4971.
41. Salgado R, Denkert C, Demaria S, et al. Harmonization of the evaluation of tumor infiltrating lymphocytes (TILs) in breast cancer: recommendations by an international TILs-working group 2014. *Ann. Oncol.* 2014. doi:10.1093/annonc/mdu450.
42. Denkert C, Minckwitz G von, Brase JC, et al. Tumor-Infiltrating Lymphocytes and Response to Neoadjuvant Chemotherapy With or Without Carboplatin in Human Epidermal Growth Factor Receptor 2-Positive and Triple-Negative Primary Breast Cancers. *J. Clin. Oncol.* 2014. doi:10.1200/JCO.2014.58.1967.
43. McGranahan N, Furness AJS, Rosenthal R, et al. Clonal neoantigens elicit T cell immunoreactivity and sensitivity to immune checkpoint blockade. *Science*. 2016;351(6280):1463-1469. doi:10.1126/science.aaf1490.

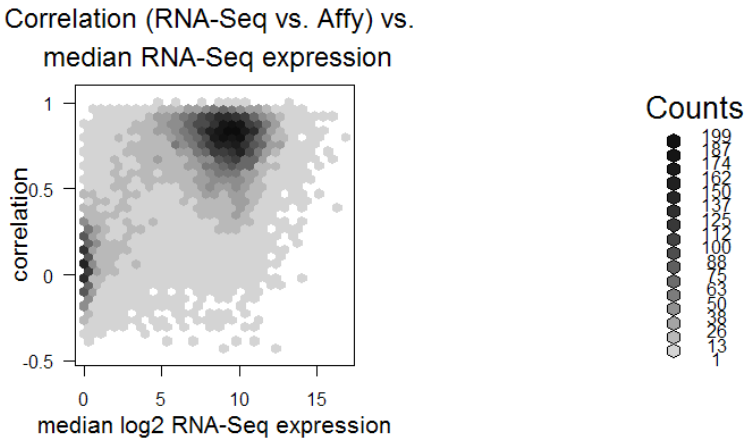
eFigures



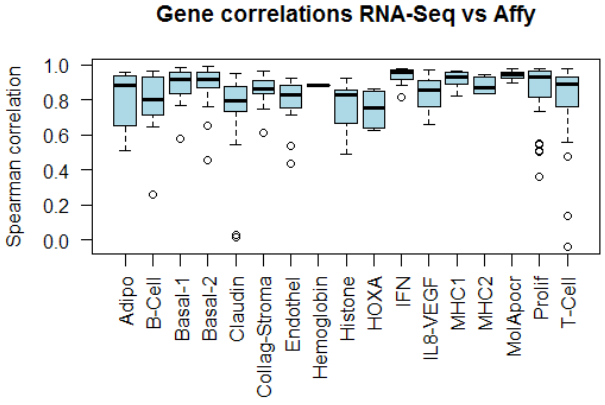
eFigure 1: **Strategy of RNA-Seq and Whole-Exome-Seq analyses for TNBC classification.**
The analytical strategy and the flow of patients through the study is presented as recommended by the REMARK criteria ("REporting recommendations for tumour MARKer prognostic studies" McShane et al. J Clin Oncol. 2005;23:9067).



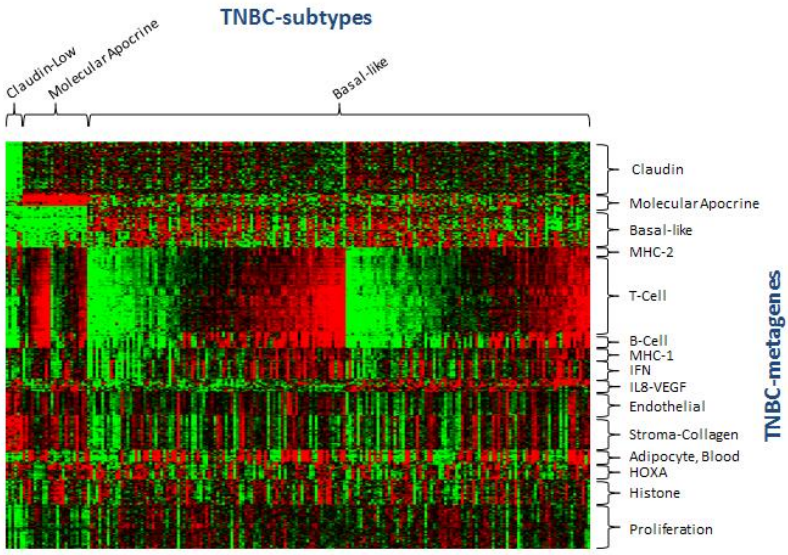
eFigure 2: **ER, PR and HER2 expression assessed by RNA-Seq and Agilent arrays.**
Correlation of Agilent-microarray (y-axis) and RNA-Seq data (x-axis) on mRNA expression of ESR1(A), PgR(B), and HER2(C) in 529 breast cancers. Expression distribution of RNA-Seq results for the same three genes (D, E, F).



eFigure 3: **Dependency of platform correlation on gene expression level.**
Spearman correlation coefficients of 16097 genes measured both on Affymetrix microarray and RNA-Seq platform in a series of 57 breast cancers were compared to the median RNA-Seq expression of the genes in the 208 TNBC samples. Poor correlation is frequent for genes with median log2 RNA-Seq expression < 2.

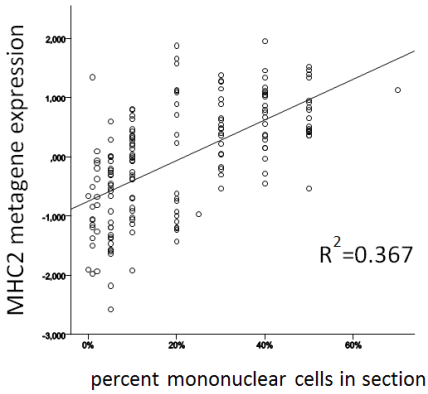


e Figure 4: **Correlation between RNA-Seq and Affymetrix for Metagene clusters.**
Boxplots of Spearman correlation coefficients between Affymetrix microarray and RNA-Seq platform for 304 genes from Metagene clusters.



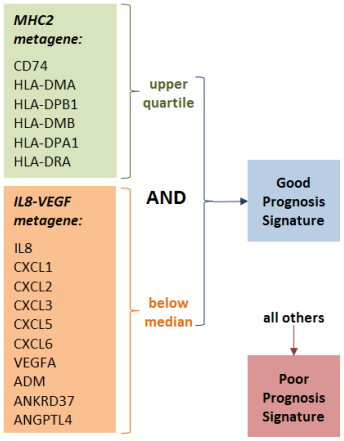
eFigure 5: Classification of TNBC (n=208) based on RNA-Seq data.

Heatmap of RNA-Seq data for 208 TNBC (in colums) and 304 genes (in rows) representing the metagenes given on the right. Samples were classified and sorted according the previously described TNBC subtypes (given above heatmap), a median split according to IL8-VEGF metagene, and increasing order according to immune cell infiltration as measured by T-Cell metagene expression (Rody et al. 2011).



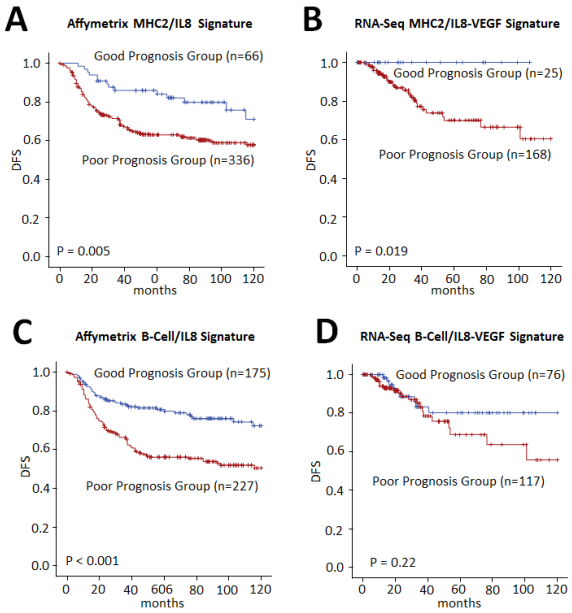
eFigure 6: Correlation of MHC2 metagene expression and histological quantification of TILs in TCGA samples.

Expression of the MHC2 metagene in TNBC samples from TCGA was compared to the quantification of mononuclear cells from Lehmann et al. 2016 from tissue slides of TCGA.

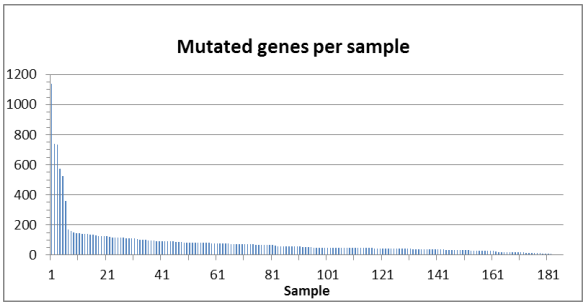


eFigure 7: Classification algorithm of the prognostic immune signature.

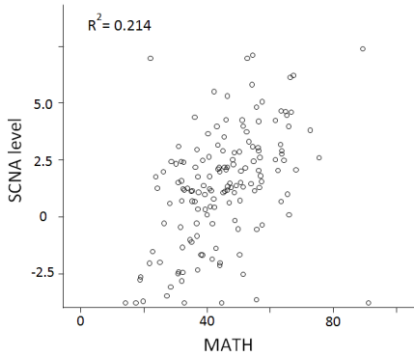
TNBC with high lymphocyte infiltration measured by MHC2 metagene expression (upper quartile) in combination with low IL8/VEGF-metagene expression (below median) were classified as “good prognosis” while all other samples were classified as “poor prognosis”. Metagene values represent the mean of log_2-transformed RNA-Seq data for the given genes.



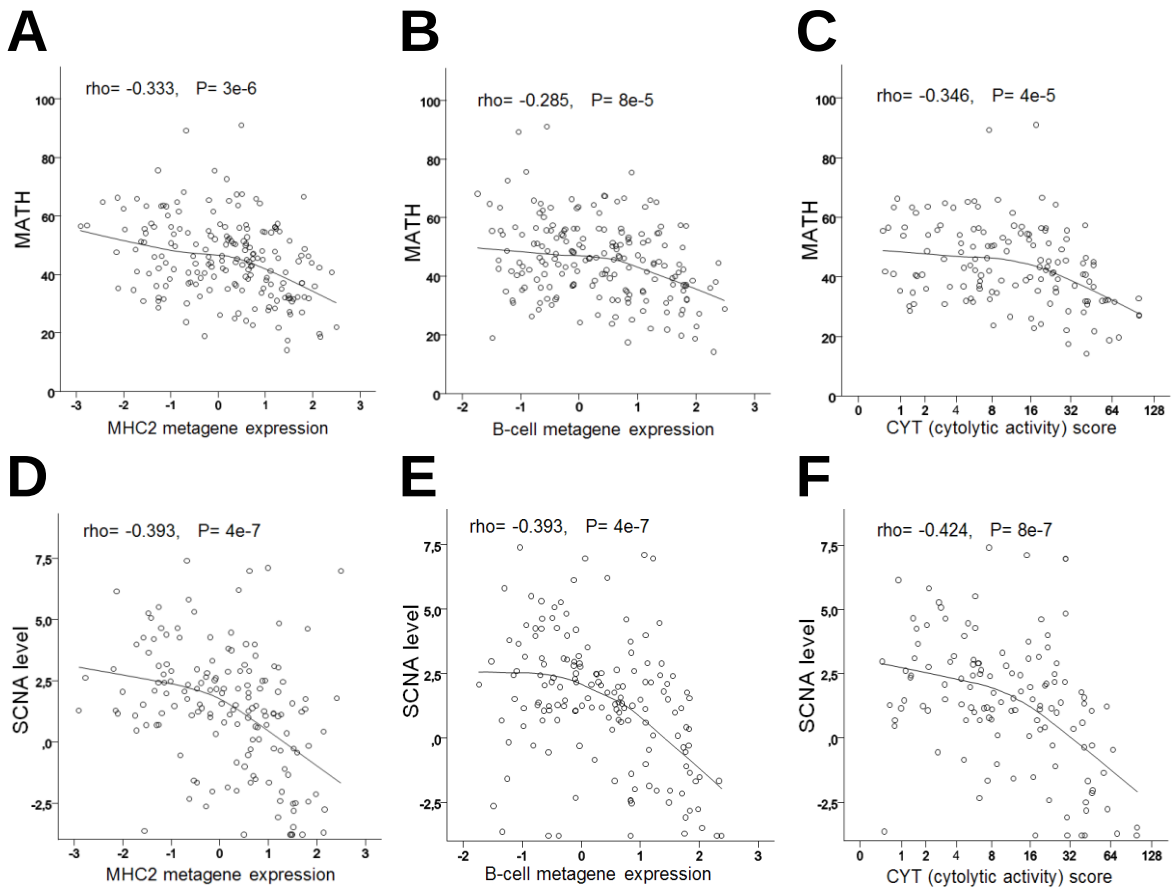
eFigure 8: **Validation of improved prognosis of TNBC patients with "Good prognosis" signature in RNA-Seq data.**
Survival of 402 previously published TNBC patients from Affymetrix microarray dataset (A,C) (Rody et al. 2011) and 193 TNBC new patients from TCGA RNA-Seq data (B,D) according to prognostic signatures.
In Panels A and B the patients classified as "Good Prognosis Group" are defined by highest quartile of the MHC2 metagene as immune infiltration marker AND low IL8-VEGF metagene (below median of cohort). In Panels C and D samples were stratified using high B-Cell metagene expression (above lowest quartile) IL8-metagene expression (below median).



eFigure 9: **Mutational count distribution in 186 TNBC.**
The number of curated mutations for each of the 186 TNBC samples is given (sorted in decreasing order).

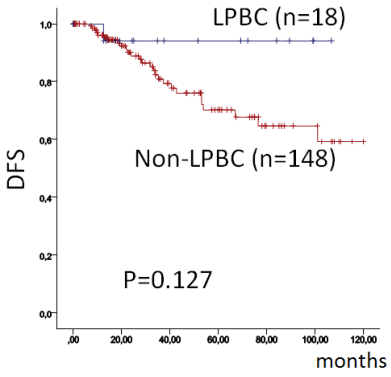


eFigure 10: **Relationship between SCNA levels and MATH in TNBC from TCGA.**
SCNA (somatic copy number alteration) levels are compared to the MATH score of clonal intratumor heterogeneity for 186 TNBC samples from TCGA . Despite some correlation of the two parameters (R^2=0.214) both measures mainly provide independent information.



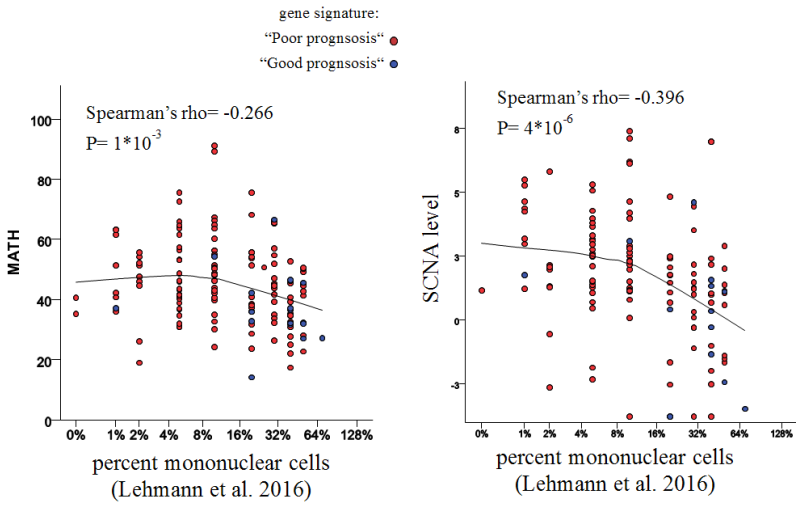
eFigure 11: Validation of inverse relationships between measures of genomic complexity and immune cell infiltration in TNBC using different immune metagenes.

The inverse relationship between immune infiltration and both MATH score (A-C) and SCNA levels (D-E), respectively, is validated using different surrogate measures for immune cell infiltration in TNBC from TCGA. The following metagene expression values are plotted on the x-axes: MHC2 metagene (A, E), B-cell metagene (B, E), and the CYT score from Rooney et al. 2015 (C, F). Spearman's rho correlation coefficient and corresponding P-values are given.



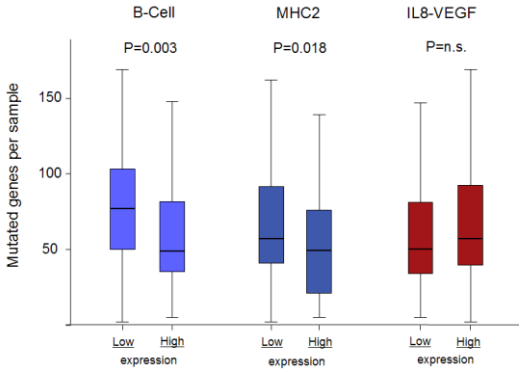
eFigure 12: Prognostic value of histologically quantified TILs in the TCGA TNBC dataset.

Kaplan-Meier analysis of disease free survival in TNBC from TCGA according to lymphocyte infiltration. Lymphycte predominant breast cancer (LPBC) was defined as ≥50% mononuclear cells according to the quantification by Lehmann et al. 2016 from tissue slides of TNBC samples from the TCGA.



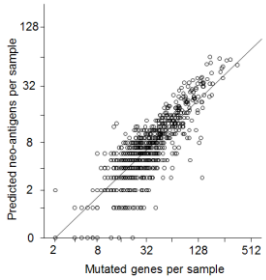
eFigure 13: Validation of inverse relationship of genomic heterogeneity and immune cell infiltration using histologically quantified TILs in the TCGA TNBC dataset.

MATH scores and SCNA levels are compared to lymphocyte infiltration as percent mononuclear cells according to the quantification by Lehmann et al. 2016 from tissue slides of TNBC samples from the TCGA (Spearman's rho = -0.266 and -0.396, respectively).



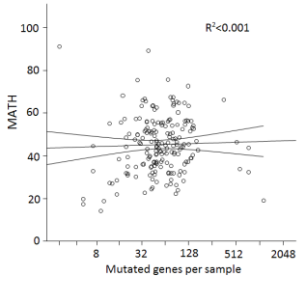
eFigure 14: Differences in mutation count by immune cell infiltration metagenes and IL8/VEGF metagene expression categories.

Box plots show the numbers of mutated genes in 186 TNBC stratified by B-cell metagene (q1 vs. q2-q4, n=47 vs. 139, P=0.003), MHC2 metagene (q1-3 vs. q4, n=141 vs. 45, P=0.018), and IL8/VEGF metagene expression (below versus above the median, n=95 vs. 91, n.s.)**.** Note, that the y-axis has been croped at 170 mutated genes per sample for this presentation, which will exclude six individual hypermutated samples with 300-1200 (see **eFigure 9**). P-values are from Mann-Whitney U-Tests.



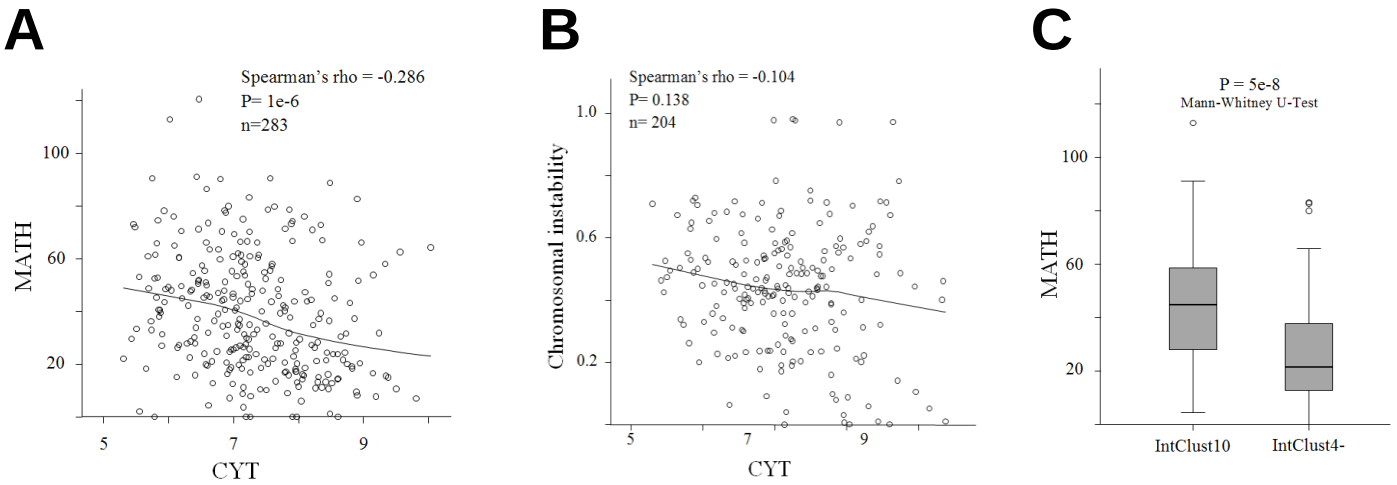
eFigure 15: Correlation of the number of predicted neo-antigens and mutational load.

Scatter plot showing the high correlation of total mutation count and the numbers of predicted neo-antigens in 760 breast cancers (R^2=0.68)**.**



eFigure 16: Independence of MATH score and total mutation counts.

Scatter plot showing lack of correlation between MATH score and the total number of mutated genes per sample (R^2<0.001).



eFigure 17: Validation analyses in METABRIC dataset.

A) Negative relationship of clonal heterogeneity (MATH score) and immune cell infiltration (CYT metric from gene expression).

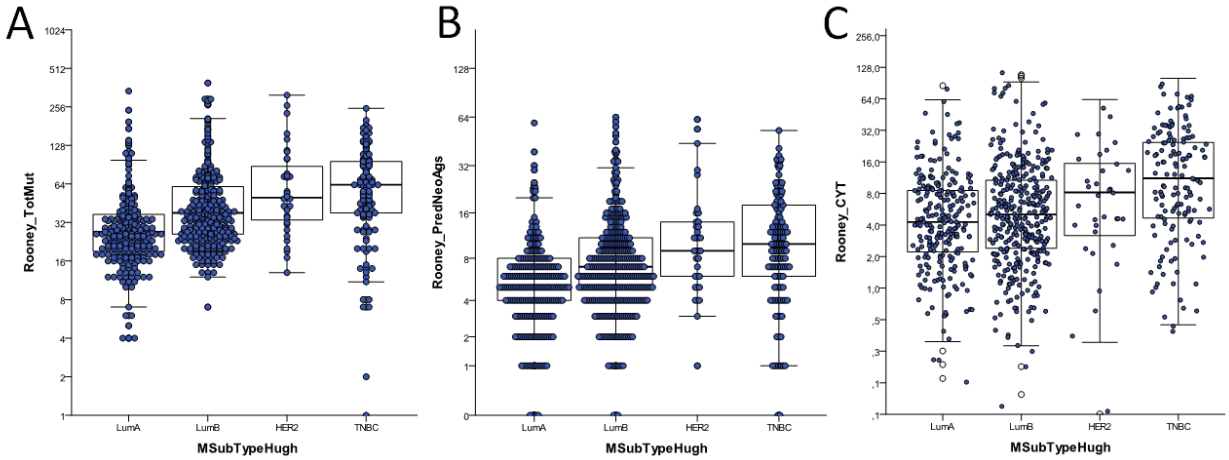
We calculated MATH scores based on only a small panel of 173 sequenced genes available from the METABRIC dataset. Thus we expected only a very low precision of the corresponding MATH score. Still we observed a highly significant (P=1e-6) negative correlation between this MATH score and the CYT metric for immune cell infiltration from gene expression data (Spearman's rho = -0.286).

B) Relationship of chromosomal instability and immune cell infiltration.

Chromosomal instability (defined as the fraction of the genome affected by CNAs by Pereira et al 2016) is compared to immune cell infiltration (measured by the CYT score according to Rooney et al. 2015) in 204 ER-/HER2- breast cancers from the METABRIC dataset (Spearman's rho= -0.104).

C) Difference of clonal heterogeneity between Integrative Clusters.

Boxplots of MATH scores from 222 TNBC samples from IntClust10 (characterized by low amounts of TILs, n=163) and IntClust4- (showing high TILs, n=59).

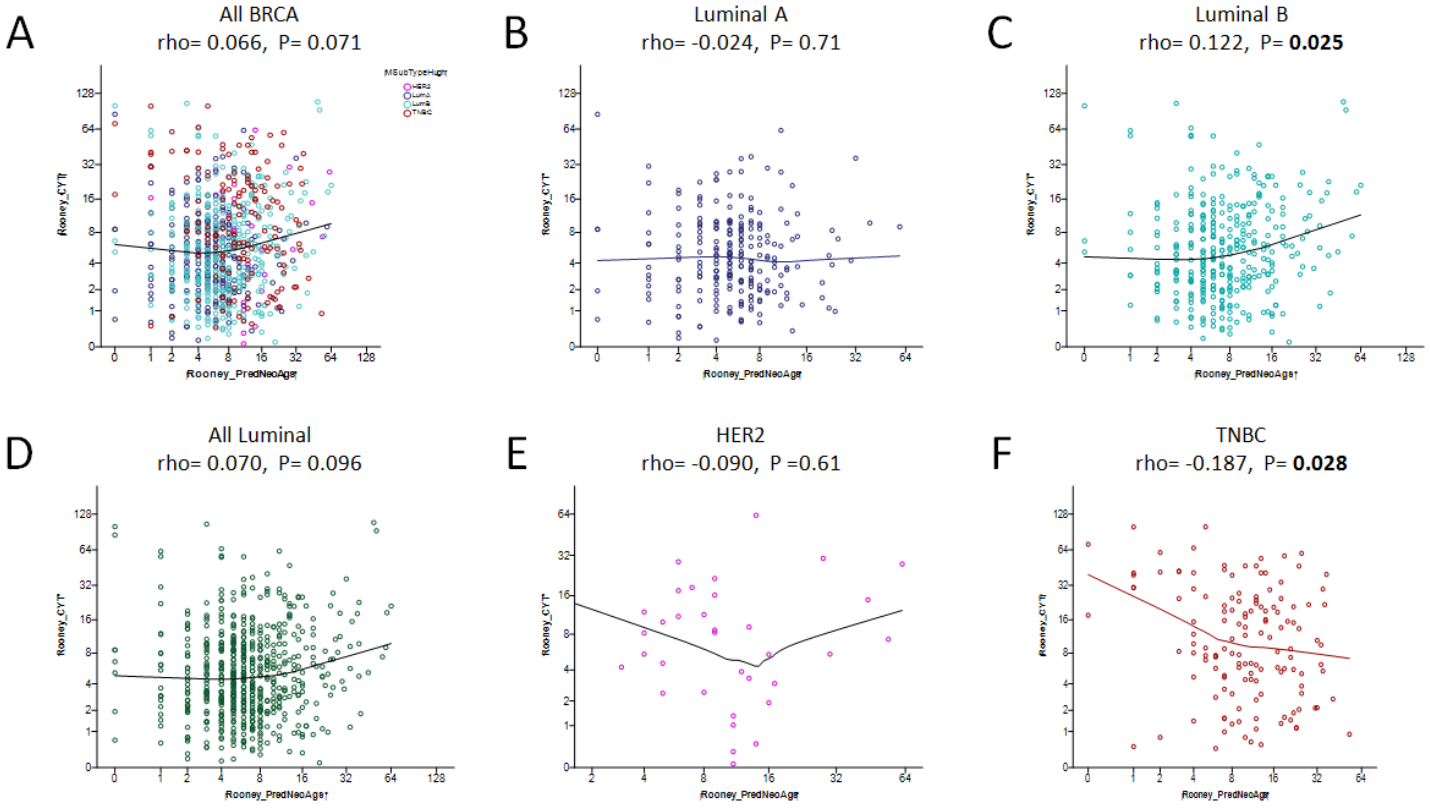


eFigure 18: Differences in mutation count, neoantigen count, and CYT by molecular subtype in breast cancer.

A) Increased total mutation counts (Rooney et al. 2015) in TNBC compared to other breast cancer subtypes (Luminal A, n=239; Luminal B, n=345; HER2, n=35; TNBC, n=141).

B) Increased predicted neoantigen load (Rooney et al. 2015) compared to other breast cancer subtypes.

C) Increased immune cytolytic activity (CYT) according to Rooney et al. 2015 in TNBC subtype.

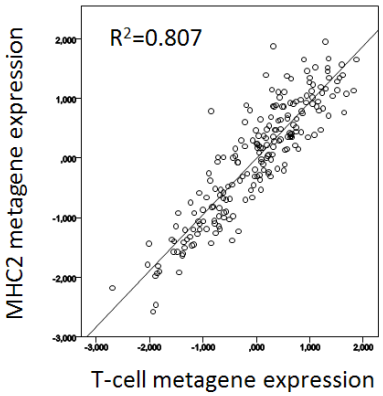


eFigure 19: Confounding of molecular breast cancer subtypes on predicted neoantigen count and CYT.

Relationship of CYT (y-axis) and predicted neoantigen load (x-axis) over all breast cancer subtypes (panel A), and in subgroups of Luminal A (panel B), Luminal B (panel C), all luminal (panel D), HER2 (panel E), and TNBC (panel F). Lines represent lowess fit, rho-values from Spearman rank correlation and corresponding P-values are given. Significance was only observed for a small positive correlation in Luminal B (rho=0.122, panel C) and a negative correlation in TNBC (rho= -0.187, panel F).

MHC2 metagene

CD74
HLA-DMA
HLA-DPB1
HLA-DMB
HLA-DPA1
HLA-DRA

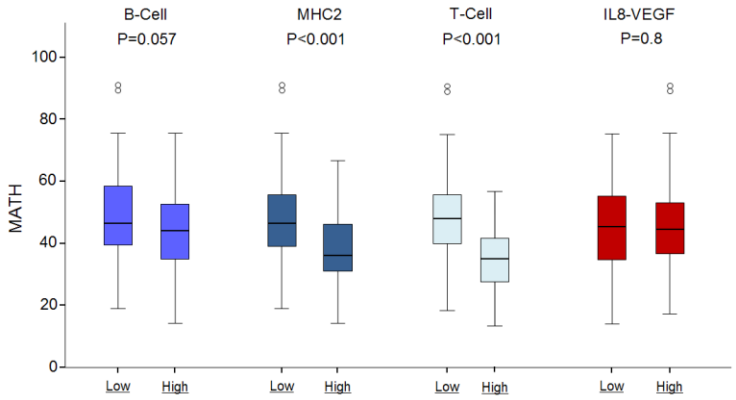


T-cell metagene

MYO1F	DOCK2	PTPRCAP	ITGAL	GZMK
FERMT3	NCKAP1L	TBC1D10C	SP140	SH2D1A
SPI1	WDFY4	SEP1	CD27	CXCR6
NCF4	EVI2B	S1PR4	LY9	ICOS
WAS	PTPRC	GVIN1	SLAMF1	TIGIT
TYROBP	LCP2	BTLA	SLAMF6	
LILRB1	PLEK	CD3G	CXCR3	
CD4	SNX20	THEMIS	CD2	
LAPTM5	SAMSN1	ITK	CD3E	
IL10RA	CD37	TRAT1	CD3D	
BTK	ARHGAP9	P2RY10	SIRPG	
CD53	RASAL3	ZNF831	SIT1	
SASH3	ACAP1	CD48	SLA2	

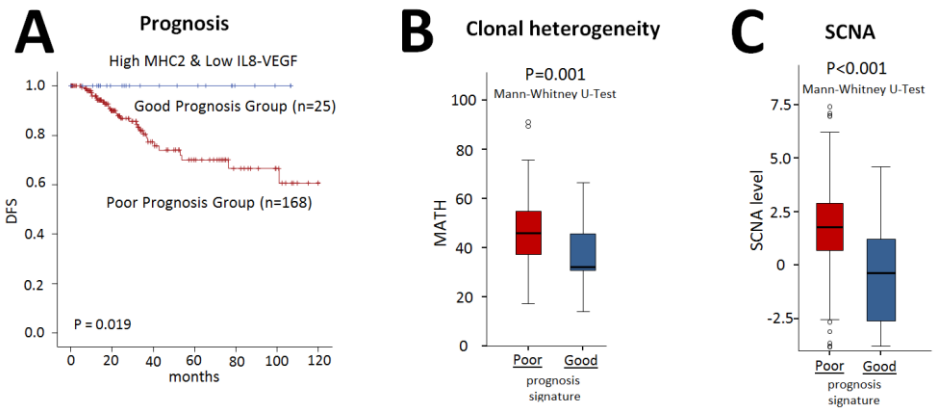
eFigure 20: High inter-correlation of TIL-metagenes

Scatter plot showing the high correlation of MHC2- and T-cell metagene expression in 208 TNBC from TCGA (R^2=0.807). The sets of individual genes represented by the two metagenes are also listed left and right from the scatter plot.



eFigure 21: Association between clonal heterogeneity and immune metagene expression.

MATH score difference between individual components of the prognostic signatures: Box plots of TNBC cohorts with low or high expression of either B-cell (q1 vs. q2-q4, n=47 vs. 139), MHC2 (q1-3 vs. q4, n=141 vs. 45), T-Cell (q1-3 vs. q4, n=140 vs. 46), or IL8-VEGF (below versus above the median, n=95 vs. 91) metagenes (P-values from Mann-Whitney U-Tests).



eFigure 22: Association of MATH and SCNA With Prognostic Groups in TNBC.

A) Kaplan-Meier analysis of disease-free survival of 193 TNBC samples with follow-up data from TCGA. Patients classified as "Good Prognosis Group" are defined by highest quartile of the MHC2 metagene as immune infiltration marker AND low IL8-VEGF metagene (below median of cohort) as in Figure 2 (P=0.019, log-rank test).

B) Box plot showing differences in MATH value between prognostic groups in A (P=0.001, Mann Whitney test).

C) Box plot showing differences in SCNA levels between prognostic groups in A (P=0.001, Mann Whitney test).

eTables

eTable 1: Annotated cancer-genes mutated in ≥3 samples.

Gene	TNBC with mutations (186 total)	Proportion of samples	Cancer pathway
TP53	139	74.7%	Cell Cycle/Apoptosis; DNA Damage Control
PIK3CA	16	8.6%	PI3K
MLL3	10	5.4%	Chromatin Modification
PTEN	9	4.8%	PI3K
MLL2	9	4.8%	Chromatin Modification
FANCD2	7	3.8%	DNA Damage Control
BRCA1	7	3.8%	DNA Damage Control
NF1	7	3.8%	RAS
NCOR1	7	3.8%	Chromatin Modification
FBXW7	7	3.8%	NOTCH
ARID1B	7	3.8%	Chromatin Modification
CREBBP	6	3.2%	Chromatin Modification; Transcriptional Regulation
APC	6	3.2%	APC
NOTCH2	6	3.2%	NOTCH
RB1	6	3.2%	Cell Cycle/Apoptosis
BRCA2	5	2.7%	DNA Damage Control
ATM	4	2.2%	DNA Damage Control
PALB2	4	2.2%	DNA Damage Control
PMS1	4	2.2%	DNA Damage Control
BAP1	4	2.2%	DNA Damage Control
KDM6A	4	2.2%	Chromatin Modification
SMARCA4	4	2.2%	Chromatin Modification
MED12	4	2.2%	Cell Cycle/Apoptosis; TGF-b
BCOR	4	2.2%	Transcriptional Regulation
ARID2	4	2.2%	Chromatin Modification
CARD11	4	2.2%	Cell Cycle/Apoptosis
NOTCH1	4	2.2%	NOTCH
ASXL1	4	2.2%	Chromatin Modification
MSH6	3	1.6%	DNA Damage Control
STAG2	3	1.6%	DNA Damage Control
PIK3R1	3	1.6%	PI3K
EP300	3	1.6%	Chromatin Modification; APC; TGF-b; NOTCH
TET2	3	1.6%	Chromatin Modification
ACVR1B	3	1.6%	TGF-b
JAK2	3	1.6%	STAT
ALK	3	1.6%	PI3K; RAS
ARID1A	3	1.6%	Chromatin Modification
CASP8	3	1.6%	Cell Cycle/Apoptosis
TRAF7	3	1.6%	Apoptosis
FAM123B	3	1.6%	APC
MAP2K1	3	1.6%	RAS
TSC2	3	1.6%	PI3K
SETD2	3	1.6%	Chromatin Modification
CDH1	3	1.6%	APC
EXT1	3	1.6%	Hedgehog

eTable 2: TCGA samples included in the study.

TCGA Sample	Patient Barcode
TCGA-3C-AAAU-01	TCGA-3C-AAAU
TCGA-3C-AALI-01	TCGA-3C-AALI
TCGA-3C-AALJ-01	TCGA-3C-AALJ
TCGA-3C-AALK-01	TCGA-3C-AALK
TCGA-4H-AAAK-01	TCGA-4H-AAAK
TCGA-5L-AAT0-01	TCGA-5L-AAT0
TCGA-5L-AAT1-01	TCGA-5L-AAT1
TCGA-5T-A9QA-01	TCGA-5T-A9QA
TCGA-A1-A0SB-01	TCGA-A1-A0SB
TCGA-A1-A0SD-01	TCGA-A1-A0SD
TCGA-A1-A0SE-01	TCGA-A1-A0SE
TCGA-A1-A0SF-01	TCGA-A1-A0SF
TCGA-A1-A0SG-01	TCGA-A1-A0SG
TCGA-A1-A0SH-01	TCGA-A1-A0SH
TCGA-A1-A0SI-01	TCGA-A1-A0SI
TCGA-A1-A0SJ-01	TCGA-A1-A0SJ
TCGA-A1-A0SK-01	TCGA-A1-A0SK
TCGA-A1-A0SM-01	TCGA-A1-A0SM
TCGA-A1-A0SN-01	TCGA-A1-A0SN
TCGA-A1-A0SO-01	TCGA-A1-A0SO
TCGA-A1-A0SP-01	TCGA-A1-A0SP
TCGA-A1-A0SQ-01	TCGA-A1-A0SQ
TCGA-A2-A04N-01	TCGA-A2-A04N
TCGA-A2-A04P-01	TCGA-A2-A04P
TCGA-A2-A04Q-01	TCGA-A2-A04Q
TCGA-A2-A04R-01	TCGA-A2-A04R
TCGA-A2-A04T-01	TCGA-A2-A04T
TCGA-A2-A04U-01	TCGA-A2-A04U
TCGA-A2-A04V-01	TCGA-A2-A04V
TCGA-A2-A04W-01	TCGA-A2-A04W
TCGA-A2-A04X-01	TCGA-A2-A04X
TCGA-A2-A04Y-01	TCGA-A2-A04Y
TCGA-A2-A0CK-01	TCGA-A2-A0CK
TCGA-A2-A0CL-01	TCGA-A2-A0CL
TCGA-A2-A0CM-01	TCGA-A2-A0CM
TCGA-A2-A0CO-01	TCGA-A2-A0CO
TCGA-A2-A0CP-01	TCGA-A2-A0CP
TCGA-A2-A0CQ-01	TCGA-A2-A0CQ
TCGA-A2-A0CR-01	TCGA-A2-A0CR
TCGA-A2-A0CS-01	TCGA-A2-A0CS
TCGA-A2-A0CT-01	TCGA-A2-A0CT
TCGA-A2-A0CU-01	TCGA-A2-A0CU
TCGA-A2-A0CV-01	TCGA-A2-A0CV
TCGA-A2-A0CW-01	TCGA-A2-A0CW
TCGA-A2-A0CX-01	TCGA-A2-A0CX
TCGA-A2-A0CY-01	TCGA-A2-A0CY
TCGA-A2-A0CZ-01	TCGA-A2-A0CZ
TCGA-A2-A0D0-01	TCGA-A2-A0D0
TCGA-A2-A0D1-01	TCGA-A2-A0D1
TCGA-A2-A0D2-01	TCGA-A2-A0D2
TCGA-A2-A0D3-01	TCGA-A2-A0D3
TCGA-A2-A0D4-01	TCGA-A2-A0D4
TCGA-A2-A0EM-01	TCGA-A2-A0EM
TCGA-A2-A0EN-01	TCGA-A2-A0EN
TCGA-A2-A0EO-01	TCGA-A2-A0EO
TCGA-A2-A0EP-01	TCGA-A2-A0EP
TCGA-A2-A0EQ-01	TCGA-A2-A0EQ
TCGA-A2-A0ER-01	TCGA-A2-A0ER
TCGA-A2-A0ES-01	TCGA-A2-A0ES
TCGA-A2-A0ET-01	TCGA-A2-A0ET
TCGA-A2-A0EU-01	TCGA-A2-A0EU
TCGA-A2-A0EV-01	TCGA-A2-A0EV
TCGA-A2-A0EW-01	TCGA-A2-A0EW
TCGA-A2-A0EX-01	TCGA-A2-A0EX
TCGA-A2-A0EY-01	TCGA-A2-A0EY
TCGA-A2-A0ST-01	TCGA-A2-A0ST
TCGA-A2-A0SU-01	TCGA-A2-A0SU
TCGA-A2-A0SV-01	TCGA-A2-A0SV
TCGA-A2-A0SW-01	TCGA-A2-A0SW
TCGA-A2-A0SX-01	TCGA-A2-A0SX
TCGA-A2-A0SY-01	TCGA-A2-A0SY
TCGA-A2-A0T0-01	TCGA-A2-A0T0
TCGA-A2-A0T1-01	TCGA-A2-A0T1
TCGA-A2-A0T2-01	TCGA-A2-A0T2
TCGA-A2-A0T3-01	TCGA-A2-A0T3
TCGA-A2-A0T4-01	TCGA-A2-A0T4
TCGA-A2-A0T5-01	TCGA-A2-A0T5
TCGA-A2-A0T6-01	TCGA-A2-A0T6
TCGA-A2-A0T7-01	TCGA-A2-A0T7
TCGA-A2-A0YC-01	TCGA-A2-A0YC
TCGA-A2-A0YD-01	TCGA-A2-A0YD
TCGA-A2-A0YE-01	TCGA-A2-A0YE
TCGA-A2-A0YF-01	TCGA-A2-A0YF
TCGA-A2-A0YG-01	TCGA-A2-A0YG
TCGA-A2-A0YH-01	TCGA-A2-A0YH
TCGA-A2-A0YI-01	TCGA-A2-A0YI
TCGA-A2-A0YJ-01	TCGA-A2-A0YJ
TCGA-A2-A0YK-01	TCGA-A2-A0YK
TCGA-A2-A0YL-01	TCGA-A2-A0YL
TCGA-A2-A0YM-01	TCGA-A2-A0YM
TCGA-A2-A0YT-01	TCGA-A2-A0YT
TCGA-A2-A1FV-01	TCGA-A2-A1FV
TCGA-A2-A1FW-01	TCGA-A2-A1FW
TCGA-A2-A1FX-01	TCGA-A2-A1FX
TCGA-A2-A1FZ-01	TCGA-A2-A1FZ
TCGA-A2-A1G0-01	TCGA-A2-A1G0
TCGA-A2-A1G1-01	TCGA-A2-A1G1
TCGA-A2-A1G4-01	TCGA-A2-A1G4
TCGA-A2-A1G6-01	TCGA-A2-A1G6
TCGA-A2-A259-01	TCGA-A2-A259
TCGA-A2-A25A-01	TCGA-A2-A25A
TCGA-A2-A25B-01	TCGA-A2-A25B
TCGA-A2-A25C-01	TCGA-A2-A25C
TCGA-A2-A25D-01	TCGA-A2-A25D
TCGA-A2-A25E-01	TCGA-A2-A25E
TCGA-A2-A25F-01	TCGA-A2-A25F
TCGA-A2-A3KC-01	TCGA-A2-A3KC
TCGA-A2-A3KD-01	TCGA-A2-A3KD
TCGA-A2-A3XS-01	TCGA-A2-A3XS
TCGA-A2-A3XT-01	TCGA-A2-A3XT
TCGA-A2-A3XU-01	TCGA-A2-A3XU
TCGA-A2-A3XV-01	TCGA-A2-A3XV
TCGA-A2-A3XW-01	TCGA-A2-A3XW
TCGA-A2-A3XX-01	TCGA-A2-A3XX
TCGA-A2-A3XY-01	TCGA-A2-A3XY
TCGA-A2-A3XZ-01	TCGA-A2-A3XZ
TCGA-A2-A3Y0-01	TCGA-A2-A3Y0
TCGA-A2-A4RW-01	TCGA-A2-A4RW
TCGA-A2-A4RX-01	TCGA-A2-A4RX
TCGA-A2-A4RY-01	TCGA-A2-A4RY
TCGA-A2-A4S0-01	TCGA-A2-A4S0
TCGA-A2-A4S1-01	TCGA-A2-A4S1
TCGA-A2-A4S2-01	TCGA-A2-A4S2
TCGA-A2-A4S3-01	TCGA-A2-A4S3
TCGA-A7-A0CD-01	TCGA-A7-A0CD
TCGA-A7-A0CE-01	TCGA-A7-A0CE
TCGA-A7-A0CE-11	TCGA-A7-A0CE
TCGA-A7-A0CG-01	TCGA-A7-A0CG
TCGA-A7-A0CH-01	TCGA-A7-A0CH
TCGA-A7-A0CH-11	TCGA-A7-A0CH
TCGA-A7-A0CJ-01	TCGA-A7-A0CJ
TCGA-A7-A0D9-01	TCGA-A7-A0D9
TCGA-A7-A0D9-11	TCGA-A7-A0D9
TCGA-A7-A0DA-01	TCGA-A7-A0DA
TCGA-A7-A0DB-01	TCGA-A7-A0DB
TCGA-A7-A0DB-11	TCGA-A7-A0DB
TCGA-A7-A0DC-01	TCGA-A7-A0DC
TCGA-A7-A0DC-11	TCGA-A7-A0DC
TCGA-A7-A13D-01	TCGA-A7-A13D
TCGA-A7-A13E-01	TCGA-A7-A13E
TCGA-A7-A13E-11	TCGA-A7-A13E
TCGA-A7-A13F-01	TCGA-A7-A13F
TCGA-A7-A13F-11	TCGA-A7-A13F
TCGA-A7-A13G-01	TCGA-A7-A13G
TCGA-A7-A13G-11	TCGA-A7-A13G
TCGA-A7-A13H-01	TCGA-A7-A13H
TCGA-A7-A26E-01	TCGA-A7-A26E
TCGA-A7-A26F-01	TCGA-A7-A26F
TCGA-A7-A26G-01	TCGA-A7-A26G
TCGA-A7-A26H-01	TCGA-A7-A26H
TCGA-A7-A26I-01	TCGA-A7-A26I
TCGA-A7-A26J-01	TCGA-A7-A26J
TCGA-A7-A2KD-01	TCGA-A7-A2KD

TCGA Sample	Patient Barcode
TCGA-A7-A3IY-01	TCGA-A7-A3IY
TCGA-A7-A3IZ-01	TCGA-A7-A3IZ
TCGA-A7-A3J0-01	TCGA-A7-A3J0
TCGA-A7-A3J1-01	TCGA-A7-A3J1
TCGA-A7-A3RF-01	TCGA-A7-A3RF
TCGA-A7-A425-01	TCGA-A7-A425
TCGA-A7-A426-01	TCGA-A7-A426
TCGA-A7-A4SA-01	TCGA-A7-A4SA
TCGA-A7-A4SB-01	TCGA-A7-A4SB
TCGA-A7-A4SC-01	TCGA-A7-A4SC
TCGA-A7-A4SD-01	TCGA-A7-A4SD
TCGA-A7-A4SE-01	TCGA-A7-A4SE
TCGA-A7-A4SF-01	TCGA-A7-A4SF
TCGA-A7-A56D-01	TCGA-A7-A56D
TCGA-A7-A5ZV-01	TCGA-A7-A5ZV
TCGA-A7-A5ZW-01	TCGA-A7-A5ZW
TCGA-A7-A5ZX-01	TCGA-A7-A5ZX
TCGA-A7-A6VV-01	TCGA-A7-A6VV
TCGA-A7-A6VW-01	TCGA-A7-A6VW
TCGA-A7-A6VX-01	TCGA-A7-A6VX
TCGA-A7-A6VY-01	TCGA-A7-A6VY
TCGA-A8-A06N-01	TCGA-A8-A06N
TCGA-A8-A06O-01	TCGA-A8-A06O
TCGA-A8-A06P-01	TCGA-A8-A06P
TCGA-A8-A06Q-01	TCGA-A8-A06Q
TCGA-A8-A06R-01	TCGA-A8-A06R
TCGA-A8-A06T-01	TCGA-A8-A06T
TCGA-A8-A06U-01	TCGA-A8-A06U
TCGA-A8-A06X-01	TCGA-A8-A06X
TCGA-A8-A06Y-01	TCGA-A8-A06Y
TCGA-A8-A06Z-01	TCGA-A8-A06Z
TCGA-A8-A075-01	TCGA-A8-A075
TCGA-A8-A076-01	TCGA-A8-A076
TCGA-A8-A079-01	TCGA-A8-A079
TCGA-A8-A07B-01	TCGA-A8-A07B
TCGA-A8-A07C-01	TCGA-A8-A07C
TCGA-A8-A07E-01	TCGA-A8-A07E
TCGA-A8-A07F-01	TCGA-A8-A07F
TCGA-A8-A07G-01	TCGA-A8-A07G
TCGA-A8-A07I-01	TCGA-A8-A07I
TCGA-A8-A07J-01	TCGA-A8-A07J
TCGA-A8-A07L-01	TCGA-A8-A07L
TCGA-A8-A07O-01	TCGA-A8-A07O
TCGA-A8-A07P-01	TCGA-A8-A07P
TCGA-A8-A07R-01	TCGA-A8-A07R
TCGA-A8-A07S-01	TCGA-A8-A07S
TCGA-A8-A07U-01	TCGA-A8-A07U
TCGA-A8-A07W-01	TCGA-A8-A07W
TCGA-A8-A07Z-01	TCGA-A8-A07Z
TCGA-A8-A081-01	TCGA-A8-A081
TCGA-A8-A082-01	TCGA-A8-A082
TCGA-A8-A083-01	TCGA-A8-A083
TCGA-A8-A084-01	TCGA-A8-A084
TCGA-A8-A085-01	TCGA-A8-A085
TCGA-A8-A086-01	TCGA-A8-A086
TCGA-A8-A08A-01	TCGA-A8-A08A
TCGA-A8-A08B-01	TCGA-A8-A08B
TCGA-A8-A08C-01	TCGA-A8-A08C
TCGA-A8-A08F-01	TCGA-A8-A08F
TCGA-A8-A08G-01	TCGA-A8-A08G
TCGA-A8-A08H-01	TCGA-A8-A08H
TCGA-A8-A08I-01	TCGA-A8-A08I
TCGA-A8-A08J-01	TCGA-A8-A08J
TCGA-A8-A08L-01	TCGA-A8-A08L
TCGA-A8-A08O-01	TCGA-A8-A08O
TCGA-A8-A08P-01	TCGA-A8-A08P
TCGA-A8-A08R-01	TCGA-A8-A08R
TCGA-A8-A08S-01	TCGA-A8-A08S
TCGA-A8-A08T-01	TCGA-A8-A08T
TCGA-A8-A08X-01	TCGA-A8-A08X
TCGA-A8-A08Z-01	TCGA-A8-A08Z
TCGA-A8-A090-01	TCGA-A8-A090
TCGA-A8-A091-01	TCGA-A8-A091
TCGA-A8-A092-01	TCGA-A8-A092
TCGA-A8-A093-01	TCGA-A8-A093
TCGA-A8-A094-01	TCGA-A8-A094
TCGA-A8-A095-01	TCGA-A8-A095
TCGA-A8-A096-01	TCGA-A8-A096
TCGA-A8-A097-01	TCGA-A8-A097
TCGA-A8-A099-01	TCGA-A8-A099
TCGA-A8-A09A-01	TCGA-A8-A09A
TCGA-A8-A09B-01	TCGA-A8-A09B
TCGA-A8-A09C-01	TCGA-A8-A09C
TCGA-A8-A09D-01	TCGA-A8-A09D
TCGA-A8-A09E-01	TCGA-A8-A09E
TCGA-A8-A09G-01	TCGA-A8-A09G
TCGA-A8-A09I-01	TCGA-A8-A09I
TCGA-A8-A09K-01	TCGA-A8-A09K
TCGA-A8-A09M-01	TCGA-A8-A09M
TCGA-A8-A09N-01	TCGA-A8-A09N
TCGA-A8-A09Q-01	TCGA-A8-A09Q
TCGA-A8-A09R-01	TCGA-A8-A09R
TCGA-A8-A09T-01	TCGA-A8-A09T
TCGA-A8-A09V-01	TCGA-A8-A09V
TCGA-A8-A09W-01	TCGA-A8-A09W
TCGA-A8-A09X-01	TCGA-A8-A09X
TCGA-A8-A09Z-01	TCGA-A8-A09Z
TCGA-A8-A0A1-01	TCGA-A8-A0A1
TCGA-A8-A0A2-01	TCGA-A8-A0A2
TCGA-A8-A0A4-01	TCGA-A8-A0A4
TCGA-A8-A0A6-01	TCGA-A8-A0A6
TCGA-A8-A0A7-01	TCGA-A8-A0A7
TCGA-A8-A0A9-01	TCGA-A8-A0A9
TCGA-A8-A0AB-01	TCGA-A8-A0AB
TCGA-A8-A0AD-01	TCGA-A8-A0AD
TCGA-AC-A23C-01	TCGA-AC-A23C
TCGA-AC-A23E-01	TCGA-AC-A23E
TCGA-AC-A23G-01	TCGA-AC-A23G
TCGA-AC-A23H-01	TCGA-AC-A23H
TCGA-AC-A23H-11	TCGA-AC-A23H
TCGA-AC-A2B8-01	TCGA-AC-A2B8
TCGA-AC-A2BK-01	TCGA-AC-A2BK
TCGA-AC-A2BM-01	TCGA-AC-A2BM
TCGA-AC-A2FB-01	TCGA-AC-A2FB
TCGA-AC-A2FB-11	TCGA-AC-A2FB
TCGA-AC-A2FE-01	TCGA-AC-A2FE
TCGA-AC-A2FF-01	TCGA-AC-A2FF
TCGA-AC-A2FF-11	TCGA-AC-A2FF
TCGA-AC-A2FG-01	TCGA-AC-A2FG
TCGA-AC-A2FK-01	TCGA-AC-A2FK
TCGA-AC-A2FM-01	TCGA-AC-A2FM
TCGA-AC-A2FM-11	TCGA-AC-A2FM
TCGA-AC-A2FO-01	TCGA-AC-A2FO
TCGA-AC-A2QH-01	TCGA-AC-A2QH
TCGA-AC-A2QI-01	TCGA-AC-A2QI
TCGA-AC-A2QJ-01	TCGA-AC-A2QJ
TCGA-AC-A3BB-01	TCGA-AC-A3BB
TCGA-AC-A3EH-01	TCGA-AC-A3EH
TCGA-AC-A3HN-01	TCGA-AC-A3HN
TCGA-AC-A3OD-01	TCGA-AC-A3OD
TCGA-AC-A3QP-01	TCGA-AC-A3QP
TCGA-AC-A3QQ-01	TCGA-AC-A3QQ
TCGA-AC-A3TM-01	TCGA-AC-A3TM
TCGA-AC-A3TN-01	TCGA-AC-A3TN
TCGA-AC-A3W5-01	TCGA-AC-A3W5
TCGA-AC-A3W6-01	TCGA-AC-A3W6
TCGA-AC-A3W7-01	TCGA-AC-A3W7
TCGA-AC-A3YI-01	TCGA-AC-A3YI
TCGA-AC-A3YJ-01	TCGA-AC-A3YJ
TCGA-AC-A4ZE-01	TCGA-AC-A4ZE
TCGA-AC-A5EH-01	TCGA-AC-A5EH
TCGA-AC-A5XS-01	TCGA-AC-A5XS
TCGA-AC-A5XU-01	TCGA-AC-A5XU
TCGA-AC-A62V-01	TCGA-AC-A62V
TCGA-AC-A62X-01	TCGA-AC-A62X
TCGA-AC-A62Y-01	TCGA-AC-A62Y
TCGA-AC-A6IV-01	TCGA-AC-A6IV
TCGA-AC-A6IW-01	TCGA-AC-A6IW
TCGA-AC-A6IX-01	TCGA-AC-A6IX
TCGA-AC-A6IX-06	TCGA-AC-A6IX
TCGA-AC-A6NO-01	TCGA-AC-A6NO
TCGA-AC-A7VB-01	TCGA-AC-A7VB
TCGA-AC-A7VC-01	TCGA-AC-A7VC

TCGA Sample	Patient Barcode
TCGA-AC-A8OP-01	TCGA-AC-A8OP
TCGA-AC-A8OQ-01	TCGA-AC-A8OQ
TCGA-AC-A8OR-01	TCGA-AC-A8OR
TCGA-AC-A8OS-01	TCGA-AC-A8OS
TCGA-AN-A03X-01	TCGA-AN-A03X
TCGA-AN-A03Y-01	TCGA-AN-A03Y
TCGA-AN-A041-01	TCGA-AN-A041
TCGA-AN-A046-01	TCGA-AN-A046
TCGA-AN-A049-01	TCGA-AN-A049
TCGA-AN-A04A-01	TCGA-AN-A04A
TCGA-AN-A04C-01	TCGA-AN-A04C
TCGA-AN-A04D-01	TCGA-AN-A04D
TCGA-AN-A0AJ-01	TCGA-AN-A0AJ
TCGA-AN-A0AK-01	TCGA-AN-A0AK
TCGA-AN-A0AL-01	TCGA-AN-A0AL
TCGA-AN-A0AM-01	TCGA-AN-A0AM
TCGA-AN-A0AR-01	TCGA-AN-A0AR
TCGA-AN-A0AS-01	TCGA-AN-A0AS
TCGA-AN-A0AT-01	TCGA-AN-A0AT
TCGA-AN-A0FD-01	TCGA-AN-A0FD
TCGA-AN-A0FF-01	TCGA-AN-A0FF
TCGA-AN-A0FJ-01	TCGA-AN-A0FJ
TCGA-AN-A0FK-01	TCGA-AN-A0FK
TCGA-AN-A0FL-01	TCGA-AN-A0FL
TCGA-AN-A0FN-01	TCGA-AN-A0FN
TCGA-AN-A0FS-01	TCGA-AN-A0FS
TCGA-AN-A0FT-01	TCGA-AN-A0FT
TCGA-AN-A0FV-01	TCGA-AN-A0FV
TCGA-AN-A0FW-01	TCGA-AN-A0FW
TCGA-AN-A0FX-01	TCGA-AN-A0FX
TCGA-AN-A0FY-01	TCGA-AN-A0FY
TCGA-AN-A0FZ-01	TCGA-AN-A0FZ
TCGA-AN-A0G0-01	TCGA-AN-A0G0
TCGA-AN-A0XL-01	TCGA-AN-A0XL
TCGA-AN-A0XN-01	TCGA-AN-A0XN
TCGA-AN-A0XO-01	TCGA-AN-A0XO
TCGA-AN-A0XP-01	TCGA-AN-A0XP
TCGA-AN-A0XR-01	TCGA-AN-A0XR
TCGA-AN-A0XS-01	TCGA-AN-A0XS
TCGA-AN-A0XT-01	TCGA-AN-A0XT
TCGA-AN-A0XU-01	TCGA-AN-A0XU
TCGA-AN-A0XV-01	TCGA-AN-A0XV
TCGA-AN-A0XW-01	TCGA-AN-A0XW
TCGA-AO-A03L-01	TCGA-AO-A03L
TCGA-AO-A03M-01	TCGA-AO-A03M
TCGA-AO-A03N-01	TCGA-AO-A03N
TCGA-AO-A03O-01	TCGA-AO-A03O
TCGA-AO-A03P-01	TCGA-AO-A03P
TCGA-AO-A03R-01	TCGA-AO-A03R
TCGA-AO-A03T-01	TCGA-AO-A03T
TCGA-AO-A03U-01	TCGA-AO-A03U
TCGA-AO-A03V-01	TCGA-AO-A03V
TCGA-AO-A0J2-01	TCGA-AO-A0J2
TCGA-AO-A0J3-01	TCGA-AO-A0J3
TCGA-AO-A0J4-01	TCGA-AO-A0J4
TCGA-AO-A0J5-01	TCGA-AO-A0J5
TCGA-AO-A0J6-01	TCGA-AO-A0J6
TCGA-AO-A0J7-01	TCGA-AO-A0J7
TCGA-AO-A0J8-01	TCGA-AO-A0J8
TCGA-AO-A0J9-01	TCGA-AO-A0J9
TCGA-AO-A0JA-01	TCGA-AO-A0JA
TCGA-AO-A0JB-01	TCGA-AO-A0JB
TCGA-AO-A0JC-01	TCGA-AO-A0JC
TCGA-AO-A0JD-01	TCGA-AO-A0JD
TCGA-AO-A0JE-01	TCGA-AO-A0JE
TCGA-AO-A0JF-01	TCGA-AO-A0JF
TCGA-AO-A0JG-01	TCGA-AO-A0JG
TCGA-AO-A0JI-01	TCGA-AO-A0JI
TCGA-AO-A0JJ-01	TCGA-AO-A0JJ
TCGA-AO-A0JL-01	TCGA-AO-A0JL
TCGA-AO-A0JM-01	TCGA-AO-A0JM
TCGA-AO-A124-01	TCGA-AO-A124
TCGA-AO-A125-01	TCGA-AO-A125
TCGA-AO-A126-01	TCGA-AO-A126
TCGA-AO-A128-01	TCGA-AO-A128
TCGA-AO-A129-01	TCGA-AO-A129
TCGA-AO-A12A-01	TCGA-AO-A12A
TCGA-AO-A12B-01	TCGA-AO-A12B
TCGA-AO-A12C-01	TCGA-AO-A12C
TCGA-AO-A12D-01	TCGA-AO-A12D
TCGA-AO-A12E-01	TCGA-AO-A12E
TCGA-AO-A12F-01	TCGA-AO-A12F
TCGA-AO-A12G-01	TCGA-AO-A12G
TCGA-AO-A12H-01	TCGA-AO-A12H
TCGA-AO-A1KO-01	TCGA-AO-A1KO
TCGA-AO-A1KP-01	TCGA-AO-A1KP
TCGA-AO-A1KQ-01	TCGA-AO-A1KQ
TCGA-AO-A1KR-01	TCGA-AO-A1KR
TCGA-AO-A1KS-01	TCGA-AO-A1KS
TCGA-AO-A1KT-01	TCGA-AO-A1KT
TCGA-AQ-A04H-01	TCGA-AQ-A04H
TCGA-AQ-A04J-01	TCGA-AQ-A04J
TCGA-AQ-A04L-01	TCGA-AQ-A04L
TCGA-AQ-A0Y5-01	TCGA-AQ-A0Y5
TCGA-AQ-A1H2-01	TCGA-AQ-A1H2
TCGA-AQ-A1H3-01	TCGA-AQ-A1H3
TCGA-AQ-A54N-01	TCGA-AQ-A54N
TCGA-AQ-A54O-01	TCGA-AQ-A54O
TCGA-AQ-A7U7-01	TCGA-AQ-A7U7
TCGA-AR-A0TP-01	TCGA-AR-A0TP
TCGA-AR-A0TQ-01	TCGA-AR-A0TQ
TCGA-AR-A0TR-01	TCGA-AR-A0TR
TCGA-AR-A0TS-01	TCGA-AR-A0TS
TCGA-AR-A0TT-01	TCGA-AR-A0TT
TCGA-AR-A0TU-01	TCGA-AR-A0TU
TCGA-AR-A0TV-01	TCGA-AR-A0TV
TCGA-AR-A0TW-01	TCGA-AR-A0TW
TCGA-AR-A0TX-01	TCGA-AR-A0TX
TCGA-AR-A0TY-01	TCGA-AR-A0TY
TCGA-AR-A0TZ-01	TCGA-AR-A0TZ
TCGA-AR-A0U0-01	TCGA-AR-A0U0
TCGA-AR-A0U1-01	TCGA-AR-A0U1
TCGA-AR-A0U2-01	TCGA-AR-A0U2
TCGA-AR-A0U3-01	TCGA-AR-A0U3
TCGA-AR-A0U4-01	TCGA-AR-A0U4
TCGA-AR-A1AH-01	TCGA-AR-A1AH
TCGA-AR-A1AI-01	TCGA-AR-A1AI
TCGA-AR-A1AJ-01	TCGA-AR-A1AJ
TCGA-AR-A1AK-01	TCGA-AR-A1AK
TCGA-AR-A1AL-01	TCGA-AR-A1AL
TCGA-AR-A1AM-01	TCGA-AR-A1AM
TCGA-AR-A1AN-01	TCGA-AR-A1AN
TCGA-AR-A1AO-01	TCGA-AR-A1AO
TCGA-AR-A1AP-01	TCGA-AR-A1AP
TCGA-AR-A1AQ-01	TCGA-AR-A1AQ
TCGA-AR-A1AR-01	TCGA-AR-A1AR
TCGA-AR-A1AS-01	TCGA-AR-A1AS
TCGA-AR-A1AT-01	TCGA-AR-A1AT
TCGA-AR-A1AU-01	TCGA-AR-A1AU
TCGA-AR-A1AV-01	TCGA-AR-A1AV
TCGA-AR-A1AW-01	TCGA-AR-A1AW
TCGA-AR-A1AX-01	TCGA-AR-A1AX
TCGA-AR-A1AY-01	TCGA-AR-A1AY
TCGA-AR-A24H-01	TCGA-AR-A24H
TCGA-AR-A24K-01	TCGA-AR-A24K
TCGA-AR-A24L-01	TCGA-AR-A24L
TCGA-AR-A24M-01	TCGA-AR-A24M
TCGA-AR-A24N-01	TCGA-AR-A24N
TCGA-AR-A24O-01	TCGA-AR-A24O
TCGA-AR-A24P-01	TCGA-AR-A24P
TCGA-AR-A24Q-01	TCGA-AR-A24Q
TCGA-AR-A24R-01	TCGA-AR-A24R
TCGA-AR-A24S-01	TCGA-AR-A24S
TCGA-AR-A24T-01	TCGA-AR-A24T
TCGA-AR-A24U-01	TCGA-AR-A24U
TCGA-AR-A24V-01	TCGA-AR-A24V
TCGA-AR-A24W-01	TCGA-AR-A24W
TCGA-AR-A24X-01	TCGA-AR-A24X
TCGA-AR-A24Z-01	TCGA-AR-A24Z
TCGA-AR-A250-01	TCGA-AR-A250
TCGA-AR-A251-01	TCGA-AR-A251
TCGA-AR-A252-01	TCGA-AR-A252
TCGA-AR-A254-01	TCGA-AR-A254

TCGA Sample	Patient Barcode
TCGA-AR-A255-01	TCGA-AR-A255
TCGA-AR-A256-01	TCGA-AR-A256
TCGA-AR-A2LE-01	TCGA-AR-A2LE
TCGA-AR-A2LH-01	TCGA-AR-A2LH
TCGA-AR-A2LJ-01	TCGA-AR-A2LJ
TCGA-AR-A2LK-01	TCGA-AR-A2LK
TCGA-AR-A2LL-01	TCGA-AR-A2LL
TCGA-AR-A2LM-01	TCGA-AR-A2LM
TCGA-AR-A2LN-01	TCGA-AR-A2LN
TCGA-AR-A2LO-01	TCGA-AR-A2LO
TCGA-AR-A2LQ-01	TCGA-AR-A2LQ
TCGA-AR-A2LR-01	TCGA-AR-A2LR
TCGA-AR-A5QM-01	TCGA-AR-A5QM
TCGA-AR-A5QN-01	TCGA-AR-A5QN
TCGA-AR-A5QP-01	TCGA-AR-A5QP
TCGA-AR-A5QQ-01	TCGA-AR-A5QQ
TCGA-B6-A0I1-01	TCGA-B6-A0I1
TCGA-B6-A0I2-01	TCGA-B6-A0I2
TCGA-B6-A0I5-01	TCGA-B6-A0I5
TCGA-B6-A0I6-01	TCGA-B6-A0I6
TCGA-B6-A0I8-01	TCGA-B6-A0I8
TCGA-B6-A0I9-01	TCGA-B6-A0I9
TCGA-B6-A0IA-01	TCGA-B6-A0IA
TCGA-B6-A0IB-01	TCGA-B6-A0IB
TCGA-B6-A0IC-01	TCGA-B6-A0IC
TCGA-B6-A0IE-01	TCGA-B6-A0IE
TCGA-B6-A0IG-01	TCGA-B6-A0IG
TCGA-B6-A0IH-01	TCGA-B6-A0IH
TCGA-B6-A0IJ-01	TCGA-B6-A0IJ
TCGA-B6-A0IK-01	TCGA-B6-A0IK
TCGA-B6-A0IM-01	TCGA-B6-A0IM
TCGA-B6-A0IN-01	TCGA-B6-A0IN
TCGA-B6-A0IO-01	TCGA-B6-A0IO
TCGA-B6-A0IP-01	TCGA-B6-A0IP
TCGA-B6-A0IQ-01	TCGA-B6-A0IQ
TCGA-B6-A0RE-01	TCGA-B6-A0RE
TCGA-B6-A0RG-01	TCGA-B6-A0RG
TCGA-B6-A0RH-01	TCGA-B6-A0RH
TCGA-B6-A0RI-01	TCGA-B6-A0RI
TCGA-B6-A0RL-01	TCGA-B6-A0RL
TCGA-B6-A0RM-01	TCGA-B6-A0RM
TCGA-B6-A0RN-01	TCGA-B6-A0RN
TCGA-B6-A0RO-01	TCGA-B6-A0RO
TCGA-B6-A0RP-01	TCGA-B6-A0RP
TCGA-B6-A0RQ-01	TCGA-B6-A0RQ
TCGA-B6-A0RS-01	TCGA-B6-A0RS
TCGA-B6-A0RT-01	TCGA-B6-A0RT
TCGA-B6-A0RU-01	TCGA-B6-A0RU
TCGA-B6-A0RV-01	TCGA-B6-A0RV
TCGA-B6-A0WS-01	TCGA-B6-A0WS
TCGA-B6-A0WT-01	TCGA-B6-A0WT
TCGA-B6-A0WV-01	TCGA-B6-A0WV
TCGA-B6-A0WW-01	TCGA-B6-A0WW
TCGA-B6-A0WX-01	TCGA-B6-A0WX
TCGA-B6-A0WY-01	TCGA-B6-A0WY
TCGA-B6-A0WZ-01	TCGA-B6-A0WZ
TCGA-B6-A0X0-01	TCGA-B6-A0X0
TCGA-B6-A0X1-01	TCGA-B6-A0X1
TCGA-B6-A0X4-01	TCGA-B6-A0X4
TCGA-B6-A0X5-01	TCGA-B6-A0X5
TCGA-B6-A0X7-01	TCGA-B6-A0X7
TCGA-B6-A1KC-01	TCGA-B6-A1KC
TCGA-B6-A1KF-01	TCGA-B6-A1KF
TCGA-B6-A1KI-01	TCGA-B6-A1KI
TCGA-B6-A1KN-01	TCGA-B6-A1KN
TCGA-B6-A2IU-01	TCGA-B6-A2IU
TCGA-B6-A3ZX-01	TCGA-B6-A3ZX
TCGA-B6-A400-01	TCGA-B6-A400
TCGA-B6-A401-01	TCGA-B6-A401
TCGA-B6-A402-01	TCGA-B6-A402
TCGA-B6-A408-01	TCGA-B6-A408
TCGA-B6-A409-01	TCGA-B6-A409
TCGA-B6-A40B-01	TCGA-B6-A40B
TCGA-B6-A40C-01	TCGA-B6-A40C
TCGA-BH-A0AU-01	TCGA-BH-A0AU
TCGA-BH-A0AU-11	TCGA-BH-A0AU
TCGA-BH-A0AV-01	TCGA-BH-A0AV
TCGA-BH-A0AW-01	TCGA-BH-A0AW
TCGA-BH-A0AY-01	TCGA-BH-A0AY
TCGA-BH-A0AY-11	TCGA-BH-A0AY
TCGA-BH-A0AZ-01	TCGA-BH-A0AZ
TCGA-BH-A0AZ-11	TCGA-BH-A0AZ
TCGA-BH-A0B0-01	TCGA-BH-A0B0
TCGA-BH-A0B1-01	TCGA-BH-A0B1
TCGA-BH-A0B2-01	TCGA-BH-A0B2
TCGA-BH-A0B2-11	TCGA-BH-A0B2
TCGA-BH-A0B3-01	TCGA-BH-A0B3
TCGA-BH-A0B3-11	TCGA-BH-A0B3
TCGA-BH-A0B4-01	TCGA-BH-A0B4
TCGA-BH-A0B5-01	TCGA-BH-A0B5
TCGA-BH-A0B5-11	TCGA-BH-A0B5
TCGA-BH-A0B6-01	TCGA-BH-A0B6
TCGA-BH-A0B7-01	TCGA-BH-A0B7
TCGA-BH-A0B7-11	TCGA-BH-A0B7
TCGA-BH-A0B8-01	TCGA-BH-A0B8
TCGA-BH-A0B8-11	TCGA-BH-A0B8
TCGA-BH-A0B9-01	TCGA-BH-A0B9
TCGA-BH-A0BA-01	TCGA-BH-A0BA
TCGA-BH-A0BA-11	TCGA-BH-A0BA
TCGA-BH-A0BC-01	TCGA-BH-A0BC
TCGA-BH-A0BC-11	TCGA-BH-A0BC
TCGA-BH-A0BD-01	TCGA-BH-A0BD
TCGA-BH-A0BF-01	TCGA-BH-A0BF
TCGA-BH-A0BG-01	TCGA-BH-A0BG
TCGA-BH-A0BJ-01	TCGA-BH-A0BJ
TCGA-BH-A0BJ-11	TCGA-BH-A0BJ
TCGA-BH-A0BL-01	TCGA-BH-A0BL
TCGA-BH-A0BM-01	TCGA-BH-A0BM
TCGA-BH-A0BM-11	TCGA-BH-A0BM
TCGA-BH-A0BO-01	TCGA-BH-A0BO
TCGA-BH-A0BP-01	TCGA-BH-A0BP
TCGA-BH-A0BQ-01	TCGA-BH-A0BQ
TCGA-BH-A0BQ-11	TCGA-BH-A0BQ
TCGA-BH-A0BR-01	TCGA-BH-A0BR
TCGA-BH-A0BS-01	TCGA-BH-A0BS
TCGA-BH-A0BS-11	TCGA-BH-A0BS
TCGA-BH-A0BT-01	TCGA-BH-A0BT
TCGA-BH-A0BT-11	TCGA-BH-A0BT
TCGA-BH-A0BV-01	TCGA-BH-A0BV
TCGA-BH-A0BV-11	TCGA-BH-A0BV
TCGA-BH-A0BW-01	TCGA-BH-A0BW
TCGA-BH-A0BW-11	TCGA-BH-A0BW
TCGA-BH-A0BZ-01	TCGA-BH-A0BZ
TCGA-BH-A0BZ-11	TCGA-BH-A0BZ
TCGA-BH-A0C0-01	TCGA-BH-A0C0
TCGA-BH-A0C0-11	TCGA-BH-A0C0
TCGA-BH-A0C1-01	TCGA-BH-A0C1
TCGA-BH-A0C3-01	TCGA-BH-A0C3
TCGA-BH-A0C3-11	TCGA-BH-A0C3
TCGA-BH-A0C7-01	TCGA-BH-A0C7
TCGA-BH-A0DD-01	TCGA-BH-A0DD
TCGA-BH-A0DD-11	TCGA-BH-A0DD
TCGA-BH-A0DE-01	TCGA-BH-A0DE
TCGA-BH-A0DG-01	TCGA-BH-A0DG
TCGA-BH-A0DG-11	TCGA-BH-A0DG
TCGA-BH-A0DH-01	TCGA-BH-A0DH
TCGA-BH-A0DH-11	TCGA-BH-A0DH
TCGA-BH-A0DI-01	TCGA-BH-A0DI
TCGA-BH-A0DK-01	TCGA-BH-A0DK
TCGA-BH-A0DK-11	TCGA-BH-A0DK
TCGA-BH-A0DL-01	TCGA-BH-A0DL
TCGA-BH-A0DL-11	TCGA-BH-A0DL
TCGA-BH-A0DO-01	TCGA-BH-A0DO
TCGA-BH-A0DO-11	TCGA-BH-A0DO
TCGA-BH-A0DP-01	TCGA-BH-A0DP
TCGA-BH-A0DP-11	TCGA-BH-A0DP
TCGA-BH-A0DQ-01	TCGA-BH-A0DQ
TCGA-BH-A0DQ-11	TCGA-BH-A0DQ
TCGA-BH-A0DS-01	TCGA-BH-A0DS
TCGA-BH-A0DT-01	TCGA-BH-A0DT
TCGA-BH-A0DT-11	TCGA-BH-A0DT
TCGA-BH-A0DV-01	TCGA-BH-A0DV
TCGA-BH-A0DV-11	TCGA-BH-A0DV

TCGA Sample	Patient Barcode
TCGA-BH-A0DX-01	TCGA-BH-A0DX
TCGA-BH-A0DZ-01	TCGA-BH-A0DZ
TCGA-BH-A0DZ-11	TCGA-BH-A0DZ
TCGA-BH-A0E0-01	TCGA-BH-A0E0
TCGA-BH-A0E0-11	TCGA-BH-A0E0
TCGA-BH-A0E1-01	TCGA-BH-A0E1
TCGA-BH-A0E1-11	TCGA-BH-A0E1
TCGA-BH-A0E2-01	TCGA-BH-A0E2
TCGA-BH-A0E6-01	TCGA-BH-A0E6
TCGA-BH-A0E7-01	TCGA-BH-A0E7
TCGA-BH-A0E9-01	TCGA-BH-A0E9
TCGA-BH-A0EA-01	TCGA-BH-A0EA
TCGA-BH-A0EB-01	TCGA-BH-A0EB
TCGA-BH-A0EE-01	TCGA-BH-A0EE
TCGA-BH-A0EI-01	TCGA-BH-A0EI
TCGA-BH-A0GY-01	TCGA-BH-A0GY
TCGA-BH-A0GZ-01	TCGA-BH-A0GZ
TCGA-BH-A0H0-01	TCGA-BH-A0H0
TCGA-BH-A0H3-01	TCGA-BH-A0H3
TCGA-BH-A0H5-01	TCGA-BH-A0H5
TCGA-BH-A0H5-11	TCGA-BH-A0H5
TCGA-BH-A0H6-01	TCGA-BH-A0H6
TCGA-BH-A0H7-01	TCGA-BH-A0H7
TCGA-BH-A0H7-11	TCGA-BH-A0H7
TCGA-BH-A0H9-01	TCGA-BH-A0H9
TCGA-BH-A0HA-01	TCGA-BH-A0HA
TCGA-BH-A0HA-11	TCGA-BH-A0HA
TCGA-BH-A0HB-01	TCGA-BH-A0HB
TCGA-BH-A0HF-01	TCGA-BH-A0HF
TCGA-BH-A0HI-01	TCGA-BH-A0HI
TCGA-BH-A0HK-01	TCGA-BH-A0HK
TCGA-BH-A0HK-11	TCGA-BH-A0HK
TCGA-BH-A0HL-01	TCGA-BH-A0HL
TCGA-BH-A0HN-01	TCGA-BH-A0HN
TCGA-BH-A0HO-01	TCGA-BH-A0HO
TCGA-BH-A0HP-01	TCGA-BH-A0HP
TCGA-BH-A0HQ-01	TCGA-BH-A0HQ
TCGA-BH-A0HU-01	TCGA-BH-A0HU
TCGA-BH-A0HW-01	TCGA-BH-A0HW
TCGA-BH-A0HX-01	TCGA-BH-A0HX
TCGA-BH-A0HY-01	TCGA-BH-A0HY
TCGA-BH-A0RX-01	TCGA-BH-A0RX
TCGA-BH-A0W3-01	TCGA-BH-A0W3
TCGA-BH-A0W4-01	TCGA-BH-A0W4
TCGA-BH-A0W5-01	TCGA-BH-A0W5
TCGA-BH-A0W7-01	TCGA-BH-A0W7
TCGA-BH-A0WA-01	TCGA-BH-A0WA
TCGA-BH-A18F-01	TCGA-BH-A18F
TCGA-BH-A18G-01	TCGA-BH-A18G
TCGA-BH-A18H-01	TCGA-BH-A18H
TCGA-BH-A18I-01	TCGA-BH-A18I
TCGA-BH-A18J-01	TCGA-BH-A18J
TCGA-BH-A18J-11	TCGA-BH-A18J
TCGA-BH-A18K-01	TCGA-BH-A18K
TCGA-BH-A18K-11	TCGA-BH-A18K
TCGA-BH-A18L-01	TCGA-BH-A18L
TCGA-BH-A18L-11	TCGA-BH-A18L
TCGA-BH-A18M-01	TCGA-BH-A18M
TCGA-BH-A18M-11	TCGA-BH-A18M
TCGA-BH-A18N-01	TCGA-BH-A18N
TCGA-BH-A18N-11	TCGA-BH-A18N
TCGA-BH-A18P-01	TCGA-BH-A18P
TCGA-BH-A18P-11	TCGA-BH-A18P
TCGA-BH-A18Q-01	TCGA-BH-A18Q
TCGA-BH-A18Q-11	TCGA-BH-A18Q
TCGA-BH-A18R-01	TCGA-BH-A18R
TCGA-BH-A18R-11	TCGA-BH-A18R
TCGA-BH-A18S-01	TCGA-BH-A18S
TCGA-BH-A18S-11	TCGA-BH-A18S
TCGA-BH-A18T-01	TCGA-BH-A18T
TCGA-BH-A18U-01	TCGA-BH-A18U
TCGA-BH-A18U-11	TCGA-BH-A18U
TCGA-BH-A18V-01	TCGA-BH-A18V
TCGA-BH-A18V-06	TCGA-BH-A18V
TCGA-BH-A18V-11	TCGA-BH-A18V
TCGA-BH-A1EN-01	TCGA-BH-A1EN
TCGA-BH-A1EN-11	TCGA-BH-A1EN
TCGA-BH-A1EO-01	TCGA-BH-A1EO
TCGA-BH-A1EO-11	TCGA-BH-A1EO
TCGA-BH-A1ES-01	TCGA-BH-A1ES
TCGA-BH-A1ES-06	TCGA-BH-A1ES
TCGA-BH-A1ET-01	TCGA-BH-A1ET
TCGA-BH-A1ET-11	TCGA-BH-A1ET
TCGA-BH-A1EU-01	TCGA-BH-A1EU
TCGA-BH-A1EU-11	TCGA-BH-A1EU
TCGA-BH-A1EV-01	TCGA-BH-A1EV
TCGA-BH-A1EV-11	TCGA-BH-A1EV
TCGA-BH-A1EW-01	TCGA-BH-A1EW
TCGA-BH-A1EW-11	TCGA-BH-A1EW
TCGA-BH-A1EX-01	TCGA-BH-A1EX
TCGA-BH-A1EY-01	TCGA-BH-A1EY
TCGA-BH-A1F0-01	TCGA-BH-A1F0
TCGA-BH-A1F0-11	TCGA-BH-A1F0
TCGA-BH-A1F2-01	TCGA-BH-A1F2
TCGA-BH-A1F2-11	TCGA-BH-A1F2
TCGA-BH-A1F5-01	TCGA-BH-A1F5
TCGA-BH-A1F6-01	TCGA-BH-A1F6
TCGA-BH-A1F6-11	TCGA-BH-A1F6
TCGA-BH-A1F8-01	TCGA-BH-A1F8
TCGA-BH-A1F8-11	TCGA-BH-A1F8
TCGA-BH-A1FB-01	TCGA-BH-A1FB
TCGA-BH-A1FB-11	TCGA-BH-A1FB
TCGA-BH-A1FC-01	TCGA-BH-A1FC
TCGA-BH-A1FC-11	TCGA-BH-A1FC
TCGA-BH-A1FD-01	TCGA-BH-A1FD
TCGA-BH-A1FD-11	TCGA-BH-A1FD
TCGA-BH-A1FE-01	TCGA-BH-A1FE
TCGA-BH-A1FE-06	TCGA-BH-A1FE
TCGA-BH-A1FE-11	TCGA-BH-A1FE
TCGA-BH-A1FG-01	TCGA-BH-A1FG
TCGA-BH-A1FG-11	TCGA-BH-A1FG
TCGA-BH-A1FH-01	TCGA-BH-A1FH
TCGA-BH-A1FH-11	TCGA-BH-A1FH
TCGA-BH-A1FJ-01	TCGA-BH-A1FJ
TCGA-BH-A1FJ-11	TCGA-BH-A1FJ
TCGA-BH-A1FL-01	TCGA-BH-A1FL
TCGA-BH-A1FM-01	TCGA-BH-A1FM
TCGA-BH-A1FM-11	TCGA-BH-A1FM
TCGA-BH-A1FN-01	TCGA-BH-A1FN
TCGA-BH-A1FN-11	TCGA-BH-A1FN
TCGA-BH-A1FR-01	TCGA-BH-A1FR
TCGA-BH-A1FR-11	TCGA-BH-A1FR
TCGA-BH-A1FU-01	TCGA-BH-A1FU
TCGA-BH-A1FU-11	TCGA-BH-A1FU
TCGA-BH-A201-01	TCGA-BH-A201
TCGA-BH-A202-01	TCGA-BH-A202
TCGA-BH-A203-01	TCGA-BH-A203
TCGA-BH-A203-11	TCGA-BH-A203
TCGA-BH-A204-01	TCGA-BH-A204
TCGA-BH-A204-11	TCGA-BH-A204
TCGA-BH-A208-01	TCGA-BH-A208
TCGA-BH-A208-11	TCGA-BH-A208
TCGA-BH-A209-01	TCGA-BH-A209
TCGA-BH-A209-11	TCGA-BH-A209
TCGA-BH-A28O-01	TCGA-BH-A28O
TCGA-BH-A28Q-01	TCGA-BH-A28Q
TCGA-BH-A2L8-01	TCGA-BH-A2L8
TCGA-BH-A42T-01	TCGA-BH-A42T
TCGA-BH-A42U-01	TCGA-BH-A42U
TCGA-BH-A42V-01	TCGA-BH-A42V
TCGA-BH-A5IZ-01	TCGA-BH-A5IZ
TCGA-BH-A5J0-01	TCGA-BH-A5J0
TCGA-BH-A6R8-01	TCGA-BH-A6R8
TCGA-BH-A6R9-01	TCGA-BH-A6R9
TCGA-BH-A8FY-01	TCGA-BH-A8FY
TCGA-BH-A8FZ-01	TCGA-BH-A8FZ
TCGA-BH-A8G0-01	TCGA-BH-A8G0
TCGA-BH-AB28-01	TCGA-BH-AB28
TCGA-C8-A12K-01	TCGA-C8-A12K
TCGA-C8-A12L-01	TCGA-C8-A12L
TCGA-C8-A12M-01	TCGA-C8-A12M
TCGA-C8-A12N-01	TCGA-C8-A12N
TCGA-C8-A12O-01	TCGA-C8-A12O

TCGA Sample	Patient Barcode
TCGA-C8-A12P-01	TCGA-C8-A12P
TCGA-C8-A12Q-01	TCGA-C8-A12Q
TCGA-C8-A12T-01	TCGA-C8-A12T
TCGA-C8-A12U-01	TCGA-C8-A12U
TCGA-C8-A12V-01	TCGA-C8-A12V
TCGA-C8-A12W-01	TCGA-C8-A12W
TCGA-C8-A12X-01	TCGA-C8-A12X
TCGA-C8-A12Y-01	TCGA-C8-A12Y
TCGA-C8-A12Z-01	TCGA-C8-A12Z
TCGA-C8-A130-01	TCGA-C8-A130
TCGA-C8-A131-01	TCGA-C8-A131
TCGA-C8-A132-01	TCGA-C8-A132
TCGA-C8-A133-01	TCGA-C8-A133
TCGA-C8-A134-01	TCGA-C8-A134
TCGA-C8-A135-01	TCGA-C8-A135
TCGA-C8-A137-01	TCGA-C8-A137
TCGA-C8-A138-01	TCGA-C8-A138
TCGA-C8-A1HE-01	TCGA-C8-A1HE
TCGA-C8-A1HF-01	TCGA-C8-A1HF
TCGA-C8-A1HG-01	TCGA-C8-A1HG
TCGA-C8-A1HI-01	TCGA-C8-A1HI
TCGA-C8-A1HJ-01	TCGA-C8-A1HJ
TCGA-C8-A1HK-01	TCGA-C8-A1HK
TCGA-C8-A1HL-01	TCGA-C8-A1HL
TCGA-C8-A1HM-01	TCGA-C8-A1HM
TCGA-C8-A1HN-01	TCGA-C8-A1HN
TCGA-C8-A1HO-01	TCGA-C8-A1HO
TCGA-C8-A26V-01	TCGA-C8-A26V
TCGA-C8-A26W-01	TCGA-C8-A26W
TCGA-C8-A26X-01	TCGA-C8-A26X
TCGA-C8-A26Y-01	TCGA-C8-A26Y
TCGA-C8-A26Z-01	TCGA-C8-A26Z
TCGA-C8-A273-01	TCGA-C8-A273
TCGA-C8-A274-01	TCGA-C8-A274
TCGA-C8-A275-01	TCGA-C8-A275
TCGA-C8-A278-01	TCGA-C8-A278
TCGA-C8-A27A-01	TCGA-C8-A27A
TCGA-C8-A27B-01	TCGA-C8-A27B
TCGA-C8-A3M7-01	TCGA-C8-A3M7
TCGA-C8-A3M8-01	TCGA-C8-A3M8
TCGA-C8-A8HP-01	TCGA-C8-A8HP
TCGA-C8-A8HQ-01	TCGA-C8-A8HQ
TCGA-C8-A8HR-01	TCGA-C8-A8HR
TCGA-D8-A13Y-01	TCGA-D8-A13Y
TCGA-D8-A13Z-01	TCGA-D8-A13Z
TCGA-D8-A140-01	TCGA-D8-A140
TCGA-D8-A141-01	TCGA-D8-A141
TCGA-D8-A142-01	TCGA-D8-A142
TCGA-D8-A143-01	TCGA-D8-A143
TCGA-D8-A145-01	TCGA-D8-A145
TCGA-D8-A146-01	TCGA-D8-A146
TCGA-D8-A147-01	TCGA-D8-A147
TCGA-D8-A1J8-01	TCGA-D8-A1J8
TCGA-D8-A1J9-01	TCGA-D8-A1J9
TCGA-D8-A1JA-01	TCGA-D8-A1JA
TCGA-D8-A1JB-01	TCGA-D8-A1JB
TCGA-D8-A1JC-01	TCGA-D8-A1JC
TCGA-D8-A1JD-01	TCGA-D8-A1JD
TCGA-D8-A1JE-01	TCGA-D8-A1JE
TCGA-D8-A1JF-01	TCGA-D8-A1JF
TCGA-D8-A1JG-01	TCGA-D8-A1JG
TCGA-D8-A1JH-01	TCGA-D8-A1JH
TCGA-D8-A1JI-01	TCGA-D8-A1JI
TCGA-D8-A1JJ-01	TCGA-D8-A1JJ
TCGA-D8-A1JK-01	TCGA-D8-A1JK
TCGA-D8-A1JL-01	TCGA-D8-A1JL
TCGA-D8-A1JM-01	TCGA-D8-A1JM
TCGA-D8-A1JN-01	TCGA-D8-A1JN
TCGA-D8-A1JP-01	TCGA-D8-A1JP
TCGA-D8-A1JS-01	TCGA-D8-A1JS
TCGA-D8-A1JT-01	TCGA-D8-A1JT
TCGA-D8-A1JU-01	TCGA-D8-A1JU
TCGA-D8-A1X5-01	TCGA-D8-A1X5
TCGA-D8-A1X6-01	TCGA-D8-A1X6
TCGA-D8-A1X7-01	TCGA-D8-A1X7
TCGA-D8-A1X8-01	TCGA-D8-A1X8
TCGA-D8-A1X9-01	TCGA-D8-A1X9
TCGA-D8-A1XA-01	TCGA-D8-A1XA
TCGA-D8-A1XB-01	TCGA-D8-A1XB
TCGA-D8-A1XC-01	TCGA-D8-A1XC
TCGA-D8-A1XD-01	TCGA-D8-A1XD
TCGA-D8-A1XF-01	TCGA-D8-A1XF
TCGA-D8-A1XG-01	TCGA-D8-A1XG
TCGA-D8-A1XJ-01	TCGA-D8-A1XJ
TCGA-D8-A1XK-01	TCGA-D8-A1XK
TCGA-D8-A1XL-01	TCGA-D8-A1XL
TCGA-D8-A1XM-01	TCGA-D8-A1XM
TCGA-D8-A1XO-01	TCGA-D8-A1XO
TCGA-D8-A1XQ-01	TCGA-D8-A1XQ
TCGA-D8-A1XR-01	TCGA-D8-A1XR
TCGA-D8-A1XS-01	TCGA-D8-A1XS
TCGA-D8-A1XT-01	TCGA-D8-A1XT
TCGA-D8-A1XU-01	TCGA-D8-A1XU
TCGA-D8-A1XV-01	TCGA-D8-A1XV
TCGA-D8-A1XW-01	TCGA-D8-A1XW
TCGA-D8-A1XY-01	TCGA-D8-A1XY
TCGA-D8-A1XZ-01	TCGA-D8-A1XZ
TCGA-D8-A1Y0-01	TCGA-D8-A1Y0
TCGA-D8-A1Y1-01	TCGA-D8-A1Y1
TCGA-D8-A1Y2-01	TCGA-D8-A1Y2
TCGA-D8-A1Y3-01	TCGA-D8-A1Y3
TCGA-D8-A27E-01	TCGA-D8-A27E
TCGA-D8-A27F-01	TCGA-D8-A27F
TCGA-D8-A27G-01	TCGA-D8-A27G
TCGA-D8-A27H-01	TCGA-D8-A27H
TCGA-D8-A27I-01	TCGA-D8-A27I
TCGA-D8-A27K-01	TCGA-D8-A27K
TCGA-D8-A27L-01	TCGA-D8-A27L
TCGA-D8-A27M-01	TCGA-D8-A27M
TCGA-D8-A27N-01	TCGA-D8-A27N
TCGA-D8-A27P-01	TCGA-D8-A27P
TCGA-D8-A27R-01	TCGA-D8-A27R
TCGA-D8-A27T-01	TCGA-D8-A27T
TCGA-D8-A27V-01	TCGA-D8-A27V
TCGA-D8-A27W-01	TCGA-D8-A27W
TCGA-D8-A3Z5-01	TCGA-D8-A3Z5
TCGA-D8-A3Z6-01	TCGA-D8-A3Z6
TCGA-D8-A4Z1-01	TCGA-D8-A4Z1
TCGA-D8-A73U-01	TCGA-D8-A73U
TCGA-D8-A73W-01	TCGA-D8-A73W
TCGA-D8-A73X-01	TCGA-D8-A73X
TCGA-E2-A105-01	TCGA-E2-A105
TCGA-E2-A106-01	TCGA-E2-A106
TCGA-E2-A107-01	TCGA-E2-A107
TCGA-E2-A108-01	TCGA-E2-A108
TCGA-E2-A109-01	TCGA-E2-A109
TCGA-E2-A10A-01	TCGA-E2-A10A
TCGA-E2-A10B-01	TCGA-E2-A10B
TCGA-E2-A10C-01	TCGA-E2-A10C
TCGA-E2-A10E-01	TCGA-E2-A10E
TCGA-E2-A10F-01	TCGA-E2-A10F
TCGA-E2-A14N-01	TCGA-E2-A14N
TCGA-E2-A14O-01	TCGA-E2-A14O
TCGA-E2-A14P-01	TCGA-E2-A14P
TCGA-E2-A14Q-01	TCGA-E2-A14Q
TCGA-E2-A14R-01	TCGA-E2-A14R
TCGA-E2-A14S-01	TCGA-E2-A14S
TCGA-E2-A14T-01	TCGA-E2-A14T
TCGA-E2-A14U-01	TCGA-E2-A14U
TCGA-E2-A14V-01	TCGA-E2-A14V
TCGA-E2-A14W-01	TCGA-E2-A14W
TCGA-E2-A14X-01	TCGA-E2-A14X
TCGA-E2-A14Y-01	TCGA-E2-A14Y
TCGA-E2-A14Z-01	TCGA-E2-A14Z
TCGA-E2-A150-01	TCGA-E2-A150
TCGA-E2-A152-01	TCGA-E2-A152
TCGA-E2-A153-01	TCGA-E2-A153
TCGA-E2-A153-11	TCGA-E2-A153
TCGA-E2-A154-01	TCGA-E2-A154
TCGA-E2-A155-01	TCGA-E2-A155
TCGA-E2-A156-01	TCGA-E2-A156
TCGA-E2-A158-01	TCGA-E2-A158
TCGA-E2-A158-11	TCGA-E2-A158

TCGA Sample	Patient Barcode
TCGA-E2-A159-01	TCGA-E2-A159
TCGA-E2-A15A-01	TCGA-E2-A15A
TCGA-E2-A15A-06	TCGA-E2-A15A
TCGA-E2-A15C-01	TCGA-E2-A15C
TCGA-E2-A15D-01	TCGA-E2-A15D
TCGA-E2-A15E-01	TCGA-E2-A15E
TCGA-E2-A15E-06	TCGA-E2-A15E
TCGA-E2-A15F-01	TCGA-E2-A15F
TCGA-E2-A15G-01	TCGA-E2-A15G
TCGA-E2-A15H-01	TCGA-E2-A15H
TCGA-E2-A15I-01	TCGA-E2-A15I
TCGA-E2-A15I-11	TCGA-E2-A15I
TCGA-E2-A15J-01	TCGA-E2-A15J
TCGA-E2-A15K-01	TCGA-E2-A15K
TCGA-E2-A15K-06	TCGA-E2-A15K
TCGA-E2-A15K-11	TCGA-E2-A15K
TCGA-E2-A15L-01	TCGA-E2-A15L
TCGA-E2-A15M-01	TCGA-E2-A15M
TCGA-E2-A15M-11	TCGA-E2-A15M
TCGA-E2-A15O-01	TCGA-E2-A15O
TCGA-E2-A15P-01	TCGA-E2-A15P
TCGA-E2-A15R-01	TCGA-E2-A15R
TCGA-E2-A15S-01	TCGA-E2-A15S
TCGA-E2-A15T-01	TCGA-E2-A15T
TCGA-E2-A1AZ-01	TCGA-E2-A1AZ
TCGA-E2-A1B0-01	TCGA-E2-A1B0
TCGA-E2-A1B1-01	TCGA-E2-A1B1
TCGA-E2-A1B4-01	TCGA-E2-A1B4
TCGA-E2-A1B5-01	TCGA-E2-A1B5
TCGA-E2-A1B6-01	TCGA-E2-A1B6
TCGA-E2-A1BC-01	TCGA-E2-A1BC
TCGA-E2-A1BC-11	TCGA-E2-A1BC
TCGA-E2-A1BD-01	TCGA-E2-A1BD
TCGA-E2-A1IE-01	TCGA-E2-A1IE
TCGA-E2-A1IF-01	TCGA-E2-A1IF
TCGA-E2-A1IG-01	TCGA-E2-A1IG
TCGA-E2-A1IG-11	TCGA-E2-A1IG
TCGA-E2-A1IH-01	TCGA-E2-A1IH
TCGA-E2-A1II-01	TCGA-E2-A1II
TCGA-E2-A1IJ-01	TCGA-E2-A1IJ
TCGA-E2-A1IK-01	TCGA-E2-A1IK
TCGA-E2-A1IL-01	TCGA-E2-A1IL
TCGA-E2-A1IN-01	TCGA-E2-A1IN
TCGA-E2-A1IO-01	TCGA-E2-A1IO
TCGA-E2-A1IP-01	TCGA-E2-A1IP
TCGA-E2-A1IU-01	TCGA-E2-A1IU
TCGA-E2-A1L6-01	TCGA-E2-A1L6
TCGA-E2-A1L7-01	TCGA-E2-A1L7
TCGA-E2-A1L7-11	TCGA-E2-A1L7
TCGA-E2-A1L8-01	TCGA-E2-A1L8
TCGA-E2-A1L9-01	TCGA-E2-A1L9
TCGA-E2-A1LA-01	TCGA-E2-A1LA
TCGA-E2-A1LB-01	TCGA-E2-A1LB
TCGA-E2-A1LB-11	TCGA-E2-A1LB
TCGA-E2-A1LE-01	TCGA-E2-A1LE
TCGA-E2-A1LG-01	TCGA-E2-A1LG
TCGA-E2-A1LH-01	TCGA-E2-A1LH
TCGA-E2-A1LH-11	TCGA-E2-A1LH
TCGA-E2-A1LI-01	TCGA-E2-A1LI
TCGA-E2-A1LK-01	TCGA-E2-A1LK
TCGA-E2-A1LL-01	TCGA-E2-A1LL
TCGA-E2-A1LS-01	TCGA-E2-A1LS
TCGA-E2-A1LS-11	TCGA-E2-A1LS
TCGA-E2-A2P5-01	TCGA-E2-A2P5
TCGA-E2-A2P6-01	TCGA-E2-A2P6
TCGA-E2-A3DX-01	TCGA-E2-A3DX
TCGA-E2-A56Z-01	TCGA-E2-A56Z
TCGA-E2-A570-01	TCGA-E2-A570
TCGA-E2-A572-01	TCGA-E2-A572
TCGA-E2-A573-01	TCGA-E2-A573
TCGA-E2-A574-01	TCGA-E2-A574
TCGA-E2-A576-01	TCGA-E2-A576
TCGA-E2-A9RU-01	TCGA-E2-A9RU
TCGA-E9-A1N3-01	TCGA-E9-A1N3
TCGA-E9-A1N4-01	TCGA-E9-A1N4
TCGA-E9-A1N4-11	TCGA-E9-A1N4
TCGA-E9-A1N5-01	TCGA-E9-A1N5
TCGA-E9-A1N5-11	TCGA-E9-A1N5
TCGA-E9-A1N6-01	TCGA-E9-A1N6
TCGA-E9-A1N6-11	TCGA-E9-A1N6
TCGA-E9-A1N8-01	TCGA-E9-A1N8
TCGA-E9-A1N9-01	TCGA-E9-A1N9
TCGA-E9-A1N9-11	TCGA-E9-A1N9
TCGA-E9-A1NA-01	TCGA-E9-A1NA
TCGA-E9-A1NA-11	TCGA-E9-A1NA
TCGA-E9-A1NC-01	TCGA-E9-A1NC
TCGA-E9-A1ND-01	TCGA-E9-A1ND
TCGA-E9-A1ND-11	TCGA-E9-A1ND
TCGA-E9-A1NE-01	TCGA-E9-A1NE
TCGA-E9-A1NF-01	TCGA-E9-A1NF
TCGA-E9-A1NF-11	TCGA-E9-A1NF
TCGA-E9-A1NG-01	TCGA-E9-A1NG
TCGA-E9-A1NG-11	TCGA-E9-A1NG
TCGA-E9-A1NH-01	TCGA-E9-A1NH
TCGA-E9-A1NI-01	TCGA-E9-A1NI
TCGA-E9-A1QZ-01	TCGA-E9-A1QZ
TCGA-E9-A1R0-01	TCGA-E9-A1R0
TCGA-E9-A1R2-01	TCGA-E9-A1R2
TCGA-E9-A1R3-01	TCGA-E9-A1R3
TCGA-E9-A1R4-01	TCGA-E9-A1R4
TCGA-E9-A1R5-01	TCGA-E9-A1R5
TCGA-E9-A1R6-01	TCGA-E9-A1R6
TCGA-E9-A1R7-01	TCGA-E9-A1R7
TCGA-E9-A1R7-11	TCGA-E9-A1R7
TCGA-E9-A1RA-01	TCGA-E9-A1RA
TCGA-E9-A1RB-01	TCGA-E9-A1RB
TCGA-E9-A1RB-11	TCGA-E9-A1RB
TCGA-E9-A1RC-01	TCGA-E9-A1RC
TCGA-E9-A1RC-11	TCGA-E9-A1RC
TCGA-E9-A1RD-01	TCGA-E9-A1RD
TCGA-E9-A1RD-11	TCGA-E9-A1RD
TCGA-E9-A1RE-01	TCGA-E9-A1RE
TCGA-E9-A1RF-01	TCGA-E9-A1RF
TCGA-E9-A1RF-11	TCGA-E9-A1RF
TCGA-E9-A1RG-01	TCGA-E9-A1RG
TCGA-E9-A1RH-01	TCGA-E9-A1RH
TCGA-E9-A1RH-11	TCGA-E9-A1RH
TCGA-E9-A1RI-01	TCGA-E9-A1RI
TCGA-E9-A1RI-11	TCGA-E9-A1RI
TCGA-E9-A226-01	TCGA-E9-A226
TCGA-E9-A227-01	TCGA-E9-A227
TCGA-E9-A228-01	TCGA-E9-A228
TCGA-E9-A229-01	TCGA-E9-A229
TCGA-E9-A22A-01	TCGA-E9-A22A
TCGA-E9-A22B-01	TCGA-E9-A22B
TCGA-E9-A22D-01	TCGA-E9-A22D
TCGA-E9-A22E-01	TCGA-E9-A22E
TCGA-E9-A22G-01	TCGA-E9-A22G
TCGA-E9-A22H-01	TCGA-E9-A22H
TCGA-E9-A243-01	TCGA-E9-A243
TCGA-E9-A244-01	TCGA-E9-A244
TCGA-E9-A245-01	TCGA-E9-A245
TCGA-E9-A247-01	TCGA-E9-A247
TCGA-E9-A248-01	TCGA-E9-A248
TCGA-E9-A249-01	TCGA-E9-A249
TCGA-E9-A24A-01	TCGA-E9-A24A
TCGA-E9-A295-01	TCGA-E9-A295
TCGA-E9-A2JS-01	TCGA-E9-A2JS
TCGA-E9-A2JT-01	TCGA-E9-A2JT
TCGA-E9-A3HO-01	TCGA-E9-A3HO
TCGA-E9-A3Q9-01	TCGA-E9-A3Q9
TCGA-E9-A3QA-01	TCGA-E9-A3QA
TCGA-E9-A3X8-01	TCGA-E9-A3X8
TCGA-E9-A54X-01	TCGA-E9-A54X
TCGA-E9-A5FK-01	TCGA-E9-A5FK
TCGA-E9-A5FL-01	TCGA-E9-A5FL
TCGA-E9-A5UO-01	TCGA-E9-A5UO
TCGA-E9-A5UP-01	TCGA-E9-A5UP
TCGA-E9-A6HE-01	TCGA-E9-A6HE
TCGA-EW-A1IW-01	TCGA-EW-A1IW
TCGA-EW-A1IX-01	TCGA-EW-A1IX
TCGA-EW-A1IY-01	TCGA-EW-A1IY
TCGA-EW-A1IZ-01	TCGA-EW-A1IZ

TCGA Sample	Patient Barcode
TCGA-EW-A1J1-01	TCGA-EW-A1J1
TCGA-EW-A1J2-01	TCGA-EW-A1J2
TCGA-EW-A1J3-01	TCGA-EW-A1J3
TCGA-EW-A1J5-01	TCGA-EW-A1J5
TCGA-EW-A1J6-01	TCGA-EW-A1J6
TCGA-EW-A1OV-01	TCGA-EW-A1OV
TCGA-EW-A1OW-01	TCGA-EW-A1OW
TCGA-EW-A1OX-01	TCGA-EW-A1OX
TCGA-EW-A1OY-01	TCGA-EW-A1OY
TCGA-EW-A1OZ-01	TCGA-EW-A1OZ
TCGA-EW-A1P0-01	TCGA-EW-A1P0
TCGA-EW-A1P1-01	TCGA-EW-A1P1
TCGA-EW-A1P3-01	TCGA-EW-A1P3
TCGA-EW-A1P4-01	TCGA-EW-A1P4
TCGA-EW-A1P5-01	TCGA-EW-A1P5
TCGA-EW-A1P6-01	TCGA-EW-A1P6
TCGA-EW-A1P7-01	TCGA-EW-A1P7
TCGA-EW-A1P8-01	TCGA-EW-A1P8
TCGA-EW-A1PA-01	TCGA-EW-A1PA
TCGA-EW-A1PB-01	TCGA-EW-A1PB
TCGA-EW-A1PC-01	TCGA-EW-A1PC
TCGA-EW-A1PD-01	TCGA-EW-A1PD
TCGA-EW-A1PE-01	TCGA-EW-A1PE
TCGA-EW-A1PF-01	TCGA-EW-A1PF
TCGA-EW-A1PG-01	TCGA-EW-A1PG
TCGA-EW-A1PH-01	TCGA-EW-A1PH
TCGA-EW-A2FR-01	TCGA-EW-A2FR
TCGA-EW-A2FS-01	TCGA-EW-A2FS
TCGA-EW-A2FV-01	TCGA-EW-A2FV
TCGA-EW-A2FW-01	TCGA-EW-A2FW
TCGA-EW-A3E8-01	TCGA-EW-A3E8
TCGA-EW-A3U0-01	TCGA-EW-A3U0
TCGA-EW-A423-01	TCGA-EW-A423
TCGA-EW-A424-01	TCGA-EW-A424
TCGA-EW-A6S9-01	TCGA-EW-A6S9
TCGA-EW-A6SA-01	TCGA-EW-A6SA
TCGA-EW-A6SB-01	TCGA-EW-A6SB
TCGA-EW-A6SC-01	TCGA-EW-A6SC
TCGA-EW-A6SD-01	TCGA-EW-A6SD
TCGA-GI-A2C8-01	TCGA-GI-A2C8
TCGA-GI-A2C8-11	TCGA-GI-A2C8
TCGA-GI-A2C9-01	TCGA-GI-A2C9
TCGA-GI-A2C9-11	TCGA-GI-A2C9
TCGA-GM-A2D9-01	TCGA-GM-A2D9
TCGA-GM-A2DA-01	TCGA-GM-A2DA
TCGA-GM-A2DB-01	TCGA-GM-A2DB
TCGA-GM-A2DC-01	TCGA-GM-A2DC
TCGA-GM-A2DD-01	TCGA-GM-A2DD
TCGA-GM-A2DF-01	TCGA-GM-A2DF
TCGA-GM-A2DH-01	TCGA-GM-A2DH
TCGA-GM-A2DI-01	TCGA-GM-A2DI
TCGA-GM-A2DK-01	TCGA-GM-A2DK
TCGA-GM-A2DL-01	TCGA-GM-A2DL
TCGA-GM-A2DM-01	TCGA-GM-A2DM
TCGA-GM-A2DN-01	TCGA-GM-A2DN
TCGA-GM-A2DO-01	TCGA-GM-A2DO
TCGA-GM-A3NW-01	TCGA-GM-A3NW
TCGA-GM-A3NY-01	TCGA-GM-A3NY
TCGA-GM-A3XG-01	TCGA-GM-A3XG
TCGA-GM-A3XL-01	TCGA-GM-A3XL
TCGA-GM-A3XN-01	TCGA-GM-A3XN
TCGA-GM-A4E0-01	TCGA-GM-A4E0
TCGA-GM-A5PV-01	TCGA-GM-A5PV
TCGA-GM-A5PX-01	TCGA-GM-A5PX
TCGA-HN-A2NL-01	TCGA-HN-A2NL
TCGA-HN-A2OB-01	TCGA-HN-A2OB
TCGA-JL-A3YW-01	TCGA-JL-A3YW
TCGA-JL-A3YX-01	TCGA-JL-A3YX
TCGA-LD-A66U-01	TCGA-LD-A66U
TCGA-LD-A74U-01	TCGA-LD-A74U
TCGA-LD-A7W5-01	TCGA-LD-A7W5
TCGA-LD-A7W6-01	TCGA-LD-A7W6
TCGA-LD-A9QF-01	TCGA-LD-A9QF
TCGA-LL-A440-01	TCGA-LL-A440
TCGA-LL-A441-01	TCGA-LL-A441
TCGA-LL-A442-01	TCGA-LL-A442
TCGA-LL-A50Y-01	TCGA-LL-A50Y
TCGA-LL-A5YL-01	TCGA-LL-A5YL
TCGA-LL-A5YM-01	TCGA-LL-A5YM
TCGA-LL-A5YN-01	TCGA-LL-A5YN
TCGA-LL-A5YO-01	TCGA-LL-A5YO
TCGA-LL-A5YP-01	TCGA-LL-A5YP
TCGA-LL-A6FP-01	TCGA-LL-A6FP
TCGA-LL-A6FQ-01	TCGA-LL-A6FQ
TCGA-LL-A6FR-01	TCGA-LL-A6FR
TCGA-LL-A73Y-01	TCGA-LL-A73Y
TCGA-LL-A73Z-01	TCGA-LL-A73Z
TCGA-LL-A740-01	TCGA-LL-A740
TCGA-LL-A7SZ-01	TCGA-LL-A7SZ
TCGA-LL-A7T0-01	TCGA-LL-A7T0
TCGA-LL-A8F5-01	TCGA-LL-A8F5
TCGA-LL-A9Q3-01	TCGA-LL-A9Q3
TCGA-LQ-A4E4-01	TCGA-LQ-A4E4
TCGA-MS-A51U-01	TCGA-MS-A51U
TCGA-OK-A5Q2-01	TCGA-OK-A5Q2
TCGA-OL-A5D6-01	TCGA-OL-A5D6
TCGA-OL-A5D7-01	TCGA-OL-A5D7
TCGA-OL-A5D8-01	TCGA-OL-A5D8
TCGA-OL-A5DA-01	TCGA-OL-A5DA
TCGA-OL-A5RU-01	TCGA-OL-A5RU
TCGA-OL-A5RV-01	TCGA-OL-A5RV
TCGA-OL-A5RW-01	TCGA-OL-A5RW
TCGA-OL-A5RX-01	TCGA-OL-A5RX
TCGA-OL-A5RY-01	TCGA-OL-A5RY
TCGA-OL-A5RZ-01	TCGA-OL-A5RZ
TCGA-OL-A5S0-01	TCGA-OL-A5S0
TCGA-OL-A66H-01	TCGA-OL-A66H
TCGA-OL-A66I-01	TCGA-OL-A66I
TCGA-OL-A66J-01	TCGA-OL-A66J
TCGA-OL-A66K-01	TCGA-OL-A66K
TCGA-OL-A66L-01	TCGA-OL-A66L
TCGA-OL-A66N-01	TCGA-OL-A66N
TCGA-OL-A66O-01	TCGA-OL-A66O
TCGA-OL-A66P-01	TCGA-OL-A66P
TCGA-OL-A6VO-01	TCGA-OL-A6VO
TCGA-OL-A6VR-01	TCGA-OL-A6VR
TCGA-OL-A97C-01	TCGA-OL-A97C
TCGA-PE-A5DC-01	TCGA-PE-A5DC
TCGA-PE-A5DD-01	TCGA-PE-A5DD
TCGA-PE-A5DE-01	TCGA-PE-A5DE
TCGA-PL-A8LV-01	TCGA-PL-A8LV
TCGA-PL-A8LX-01	TCGA-PL-A8LX
TCGA-PL-A8LY-01	TCGA-PL-A8LY
TCGA-PL-A8LZ-01	TCGA-PL-A8LZ
TCGA-S3-A6ZF-01	TCGA-S3-A6ZF
TCGA-S3-A6ZG-01	TCGA-S3-A6ZG
TCGA-S3-A6ZH-01	TCGA-S3-A6ZH
TCGA-S3-AA0Z-01	TCGA-S3-AA0Z
TCGA-S3-AA10-01	TCGA-S3-AA10
TCGA-S3-AA11-01	TCGA-S3-AA11
TCGA-S3-AA12-01	TCGA-S3-AA12
TCGA-S3-AA14-01	TCGA-S3-AA14
TCGA-S3-AA15-01	TCGA-S3-AA15
TCGA-S3-AA17-01	TCGA-S3-AA17
TCGA-UL-AAZ6-01	TCGA-UL-AAZ6
TCGA-UU-A93S-01	TCGA-UU-A93S
TCGA-V7-A7HQ-01	TCGA-V7-A7HQ
TCGA-W8-A86G-01	TCGA-W8-A86G
TCGA-WT-AB41-01	TCGA-WT-AB41
TCGA-WT-AB44-01	TCGA-WT-AB44
TCGA-XX-A899-01	TCGA-XX-A899
TCGA-XX-A89A-01	TCGA-XX-A89A
TCGA-Z7-A8R5-01	TCGA-Z7-A8R5
TCGA-Z7-A8R6-01	TCGA-Z7-A8R6

eTable 3: "Cancer genes" curated by Vogelstein and colleagues (Kandoth et al. 2013, PMID 24132290; Vogelstein et al. 2013, PMID 23539594)

APC	Apoptosis	Cell Cycle/Ap	Chromatin M	DNA Damage	HH	NOTCH	PI3K	PI3K; RAS	RAS	STAT	TGF-b	Transcription	PI3K_All	RAS_All
APC	TRAF7	CASP8	ARID1B	ATM	PTCH1	NOTCH2	TSC2	ERBB2	CIC	JAK3	ACVR1B	AR	TSC2	CIC
NF2		CARD11	DNMT3A	RECQL4	EXT2	NOTCH1	PIK3R1	MET	NF1	MPL	FOXL2	IKZF1	PIK3R1	NF1
AXIN1		MYC	SMARCB1	FANCA	EXT1	FBXW7	PIK3CA	PDGFRA	NRAS	JAK2		PHOX2B	PIK3CA	NRAS
RNF43		SKP2	WT1	FANCF		GATA1	FLCN	CBL	HRAS			RUNX1	FLCN	HRAS
CDH1		CYLD	ARID1A	XPA		GATA2	PTEN	SDHD	BRAF			BCOR	PTEN	BRAF
FAM123B		ABL1	SETD2	STAG2			TSHR	ALK	KRAS			DICER1	TSHR	KRAS
HNF1A		CHEK2	NCOR1	MSH6				EGFR	MAP2K1				ERBB2	MAP2K1
		FUBP1	ATRX	ERCC4				FGFR3	MAP3K1				MET	ERBB2
		MDM4	H3F3A	FANCG				FGFR2					PDGFRA	MET
		PPP2R1A	KDM5C	MLH1				B2M					CBL	PDGFRA
		RB1	MEN1	PMS2				GNA11					SDHD	CBL
		MDM2	PRDM1	BRCA1				KIT					ALK	SDHD
		TP53	ASXL1	FANCD2				GNAQ					EGFR	ALK
		TNFAIP3	MLL2	PMS1				RET					FGFR3	EGFR
		MED12	SMARCA4	PALB2									FGFR2	FGFR3
			ARID2	BAP1									B2M	FGFR2
			KDM6A	BLM									GNA11	B2M
			TET2	BUB1B									KIT	GNA11
			PBRM1	ERCC5									GNAQ	KIT
			DNMT1	BRCA2									RET	GNAQ
			EZH2											RET

eTable 4:
Individual genes constituting TNBC metagenes and their correlation with Affymetrix microarray

TNBCmetagene_RNA-Seq	EntrezID	Gene	Affymetrix probeset	TNBCmetagene_Affymetrix (PMID_21978456)	Correlation RNA-Seq vs Affymetrix (PMID_25412710)	Corelation Class (PMID_25412710)
Claudin	1365	CLDN3	203954_x_at	Claudin-CD24	0.844374	high.cor
Claudin	23779	ARHGAP8				
Claudin	553158	PRR5-ARHGAP8				
Claudin	29841	GRHL1			0.896681	high.cor
Claudin	10256	CNKSR1	204740_at		0.594828	low.cor
Claudin	10053	AP1M2	218261_at		0.782798	high.cor
Claudin	2886	GRB7	210761_s_at		0.814012	high.cor
Claudin	54836	BSPRY	218792_s_at		0.891690	high.cor
Claudin	4072	EPCAM	201839_s_at	Claudin-CD24	0.926497	high.cor
Claudin	2065	ERBB3	202454_s_at		0.939461	high.cor
Claudin	54845	ESRP1	219121_s_at		0.921312	high.cor
Claudin	79977	GRHL2	219388_at		0.738981	high.cor
Claudin	1364	CLDN4	201428_at	Claudin-CD24	0.817215	high.cor
Claudin	1999	ELF3	201510_at	Claudin-CD24	0.792715	high.cor
Claudin	126695	C1orf172			0.736777	high.cor
Claudin	92359	CRB3			0.562095	low.cor
Claudin	57111	RAB25	218186_at	Claudin-CD24	0.792520	high.cor
Claudin	58495	OVOL2	211778_s_at		0.894478	high.cor
Claudin	114569	MAL2				
Claudin	149466	C1orf210			0.731073	high.cor
Claudin	128218	TMEM125			0.891755	high.cor
Claudin	999	CDH1	201131_s_at		0.946655	high.cor
Claudin	5652	PRSS8	202525_at		0.821947	high.cor
Claudin	91862	MARVELD3			0.778001	high.cor
Claudin	2041	EPHA1	205977_s_at		0.540381	low.cor
Claudin	140893	C20orf151			0.011991	low.cor
Claudin	57662	KIAA1543			0.588605	low.cor
Claudin	11187	PKP3	209873_s_at		0.855263	high.cor
Claudin	3898	LAD1	203287_at		0.864078	high.cor
Claudin	146439	CCDC64B			0.782149	high.cor
Claudin	81607	PVRL4			0.794659	high.cor
Claudin	3855	KRT7	209016_s_at	Claudin-CD24	0.887283	high.cor
Claudin	10045	SH2D3A	219513_s_at		0.674034	low.cor
Claudin	64787	EPS8L2	218180_s_at		0.747796	high.cor
Claudin	64063	PRSS22	205847_at		0.024242	low.cor
Claudin	7163	TPD52	201691_s_at		0.714221	high.cor
Claudin	51361	HOOK1	219976_at		0.843013	high.cor
Claudin	10207	INADL	214493_s_at		0.785196	high.cor
MolApocr	148327	CREB3L4			0.971610	high.cor
MolApocr	57535	KIAA1324	221874_at		0.943803	high.cor
MolApocr	10551	AGR2	209173_at	Apocrine	0.979907	high.cor
MolApocr	3169	FOXA1	204667_at	Apocrine	0.921506	high.cor
MolApocr	401546	C9orf152			0.954109	high.cor
MolApocr	25803	SPDEF	213441_x_at	Apocrine	0.895191	high.cor
MolApocr	367	AR	211110_s_at	Apocrine		
MolApocr	116844	LRG1			0.932460	high.cor
MolApocr	79838	TMC5	219580_s_at		0.914312	high.cor
MolApocr	79083	MLPH	218211_s_at	Apocrine	0.953980	high.cor
Basal-2	2568	GABRP	205044_at	Basal-like	0.988138	high.cor
Basal-2	58473	PLEKHB1	209504_s_at	Basal-like	0.958387	high.cor
Basal-2	2001	ELF5	220625_s_at	Basal-like	0.975629	high.cor
Basal-2	7368	UGT8	208358_s_at		0.930127	high.cor
Basal-2	2296	FOXC1	213260_at	Basal-like	0.954045	high.cor
Basal-2	8645	KCNK5	219615_s_at	Basal-like	0.910552	high.cor
Basal-2	56963	RGMA			0.869199	high.cor
Basal-2	25984	KRT23	218963_s_at	Basal-like	0.963638	high.cor
Basal-2	57447	NDRG2	206453_s_at		0.927340	high.cor
Basal-2	8190	MIA	206560_s_at	Basal-like		
Basal-2	54763	ROPN1				
Basal-2	152015	ROPN1B	220425_x_at	Basal-like	0.885987	high.cor
Basal-2	6663	SOX10	209842_at	Basal-like	0.903033	high.cor
Basal-2	6271	S100A1	205334_at		0.783899	high.cor
Basal-2	399694	SHC4			0.934794	high.cor
Basal-2	57348	TTYH1	219415_at		0.649728	low.cor
Basal-2	30812	SOX8			0.756741	high.cor
Basal-2	260429	PRSS33			0.457610	low.cor
Basal-2	6422	SFRP1	202037_s_at	Basal-like	0.916710	high.cor
Basal-1	3868	KRT16				
Basal-1	3861	KRT14	209351_at	Basal-like	0.873736	high.cor
Basal-1	3872	KRT17	205157_s_at	Basal-like	0.954628	high.cor
Basal-1	3857	KRT9	208188_at		0.579466	low.cor
Basal-1	23650	TRIM29	202504_at	Basal-like	0.895904	high.cor
Basal-1	1830	DSG3	205595_at		0.835364	high.cor
Basal-1	5268	SERPINB5	204855_at	Basal-like	0.946785	high.cor
Basal-1	3853	KRT6A				
Basal-1	3854	KRT6B	213680_at	Basal-like	0.982305	high.cor
Basal-1	286887	KRT6C	209125_at	Basal-like	0.763093	high.cor
Basal-1	3852	KRT5	201820_at	Basal-like	0.929155	high.cor
Basal-1	6273	S100A2	204268_at	Basal-like	0.956313	high.cor
MHC2	972	CD74	209619_at	MHC-2	0.845152	high.cor
MHC2	3108	HLA-DMA	217478_s_at	MHC-2	0.925719	high.cor
MHC2	3115	HLA-DPB1	201137_s_at	MHC-2	0.831086	high.cor
MHC2	3109	HLA-DMB	203932_at	MHC-2	0.944257	high.cor
MHC2	3113	HLA-DPA1	211990_at		0.835429	high.cor
MHC2	3122	HLA-DRA	210982_s_at	MHC-2	0.888644	high.cor
T-Cell	4542	MYO1F	213733_at		0.779816	high.cor
T-Cell	83706	FERMT3			0.885468	high.cor
T-Cell	6688	SPI1	205312_at		0.139227	low.cor
T-Cell	4689	NCF4	207677_s_at		0.947693	high.cor
T-Cell	7454	WAS	38964_r_at			
T-Cell	7305	TYROBP	204122_at		0.895385	high.cor
T-Cell	10859	LILRB1	211336_x_at		0.724916	high.cor
T-Cell	920	CD4	203547_at		0.553020	low.cor
T-Cell	7805	LAPTM5	201721_s_at	T-Cell	0.931471	high.cor
T-Cell	3587	IL10RA	204912_at	T-Cell	0.953526	high.cor
T-Cell	695	BTK	205504_at			
T-Cell	963	CD53	203416_at	T-Cell	0.974008	high.cor
T-Cell	54440	SASH3	204923_at			
T-Cell	1794	DOCK2	213160_at		0.871921	high.cor
T-Cell	3071	NCKAP1L	209734_at		0.850532	high.cor
T-Cell	57705	WDFY4			0.777223	high.cor

TNBCmetagene_RNA-Seq	EntrezID	Gene	Affymetrix probeset	TNBCmetagene_Affymetrix (PMID_21978456)	Correlation RNA-Seq vs Affymetrix (PMID_25412710)	Corelation Class (PMID_25412710)
T-Cell	2124	EVI2B	211742_s_at	T-Cell	0.927470	high.cor
T-Cell	5788	PTPRC	212588_at	T-Cell	0.893052	high.cor
T-Cell	3937	LCP2	205269_at		0.893959	high.cor
T-Cell	5341	PLEK	203471_s_at		0.947239	high.cor
T-Cell	124460	SNX20			0.653228	low.cor
T-Cell	64092	SAMSN1	220330_s_at	T-Cell	0.913663	high.cor
T-Cell	951	CD37	204192_at		0.845865	high.cor
T-Cell	64333	ARHGAP9			0.667164	low.cor
T-Cell	64926	RASAL3			0.605976	low.cor
T-Cell	9744	ACAP1	205212_s_at		-0.041094	low.cor
T-Cell	5790	PTPRCAP	204960_at		0.859087	high.cor
T-Cell	374403	TBC1D10C			0.820197	high.cor
T-Cell	1731	Sep 01			0.587892	low.cor
T-Cell	8698	S1PR4	206437_at		0.553150	low.cor
T-Cell	387751	GVIN1	220577_at			
T-Cell	151888	BTLA			0.892209	high.cor
T-Cell	917	CD3G	206804_at		0.906274	high.cor
T-Cell	387357	THEMIS			0.900830	high.cor
T-Cell	3702	ITK	211339_s_at	T-Cell	0.926757	high.cor
T-Cell	50852	TRAT1	217147_s_at		0.951906	high.cor
T-Cell	27334	P2RY10	214615_at			
T-Cell	128611	ZNF831			0.761926	high.cor
T-Cell	962	CD48	204118_at	T-Cell	0.966425	high.cor
T-Cell	3683	ITGAL	213475_s_at		0.759917	high.cor
T-Cell	11262	SP140	207777_s_at		0.864143	high.cor
T-Cell	939	CD27	206150_at	T-Cell	0.928377	high.cor
T-Cell	4063	LY9	215967_s_at		0.789344	high.cor
T-Cell	6504	SLAMF1	206181_at		0.910487	high.cor
T-Cell	114836	SLAMF6			0.822855	high.cor
T-Cell	2833	CXCR3	207681_at			
T-Cell	914	CD2	205831_at	T-Cell	0.974268	high.cor
T-Cell	916	CD3E	205456_at		0.477508	low.cor
T-Cell	915	CD3D	213539_at	T-Cell	0.956832	high.cor
T-Cell	55423	SIRPG	220485_s_at		0.899144	high.cor
T-Cell	27240	SIT1	205484_at		0.744620	high.cor
T-Cell	84174	SLA2			0.855847	high.cor
T-Cell	3003	GZMK	206666_at	T-Cell	0.964480	high.cor
T-Cell	4068	SH2D1A	210116_at			
T-Cell	10663	CXCR6	206974_at		0.909515	high.cor
T-Cell	29851	ICOS	210439_at		0.924099	high.cor
T-Cell	201633	TIGIT			0.932720	high.cor
B-Cell	643	CXCR5	206126_at		0.258556	low.cor
B-Cell	640	BLK	206255_at		0.714091	high.cor
B-Cell	115350	FCRL1			0.748898	high.cor
B-Cell	931	MS4A1	210356_x_at		0.931553	high.cor
B-Cell	79368	FCRL2	221239_s_at		0.852606	high.cor
B-Cell	930	CD19	206398_s_at		0.829660	high.cor
B-Cell	973	CD79A	205049_s_at		0.641626	low.cor
B-Cell	8755	ADAM6				
B-Cell	51237	MGC29506	221286_s_at		0.937127	high.cor
B-Cell	83416	FCRL5			0.771325	high.cor
B-Cell	608	TNFRSF17	206641_at		0.962665	high.cor
B-Cell	96610	LOC96610	217179_x_at	B-Cell		
MHC1	3107	HLA-C	216526_x_at	MHC-1	0.850207	high.cor
MHC1	3105	HLA-A	215313_x_at	MHC-1		
MHC1	3136	HLA-H				
MHC1	3106	HLA-B	209140_x_at	MHC-1	0.929076	high.cor
MHC1	3134	HLA-F	221875_x_at	MHC-1	0.927016	high.cor
MHC1	5696	PSMB8	209040_s_at		0.961499	high.cor
MHC1	5698	PSMB9	204279_at		0.954628	high.cor
MHC1	6890	TAP1	202307_s_at		0.952295	high.cor
MHC1	6891	TAP2	208428_at		0.821040	high.cor
IFN	2537	IFI6	204415_at		0.961887	high.cor
IFN	4599	MX1	202086_at	IFN		
IFN	10561	IFI44	214453_s_at	IFN	0.917812	high.cor
IFN	10964	IFI44L	204439_at	IFN	0.917553	high.cor
IFN	54739	XAF1	206133_at		0.938100	high.cor
IFN	3434	IFIT1	203153_at	IFN	0.969795	high.cor
IFN	3433	IFIT2	217502_at		0.968758	high.cor
IFN	129607	CMPK2			0.974851	high.cor
IFN	91543	RSAD2	213797_at	IFN	0.975888	high.cor
IFN	3437	IFIT3	204747_at	IFN	0.973814	high.cor
IFN	4939	OAS2	204972_at	IFN	0.816243	high.cor
IFN	4938	OAS1	202869_at		0.891172	high.cor
IFN	8638	OASL	205660_at	IFN	0.959360	high.cor
IFN	4940	OAS3	218400_at	IFN	0.884625	high.cor
IL8-VEGF	2920	CXCL2	209774_x_at	IL-8	0.967915	high.cor
IL8-VEGF	2921	CXCL3	207850_at		0.759204	high.cor
IL8-VEGF	2919	CXCL1	204470_at	IL-8	0.885209	high.cor
IL8-VEGF	6372	CXCL6	206336_at		0.716730	high.cor
IL8-VEGF	6374	CXCL5	214974_x_at		0.657895	low.cor
IL8-VEGF	3576	IL8	202859_x_at	IL-8	0.909321	high.cor
IL8-VEGF	133	ADM	202912_at	VEGF	0.963832	high.cor
IL8-VEGF	353322	ANKRD37			0.823114	high.cor
IL8-VEGF	51129	ANGPTL4	221009_s_at	VEGF	0.825966	high.cor
IL8-VEGF	7422	VEGFA	210512_s_at	VEGF	0.898626	high.cor
Endothel	2828	GPR4	206236_at		0.792326	high.cor
Endothel	90952	ESAM			0.433368	low.cor
Endothel	51294	PCDH12	219656_at		0.746305	high.cor
Endothel	54538	ROBO4	220758_s_at		0.864208	high.cor
Endothel	161198	CLEC14A			0.858893	high.cor
Endothel	1003	CDH5	204677_at		0.886116	high.cor
Endothel	22899	ARHGEF15	205507_at		0.538177	low.cor
Endothel	7075	TIE1	204468_s_at		0.760695	high.cor
Endothel	947	CD34	209543_s_at		0.710721	high.cor
Endothel	79742	CXorf36	219652_s_at			
Endothel	80177	MYCT1	220471_s_at		0.914701	high.cor
Endothel	79812	MMRN2	219091_s_at		0.920210	high.cor
Endothel	22918	CD93	202878_s_at		0.833809	high.cor
Endothel	221395	GPR116	212950_at		0.897394	high.cor
Endothel	64123	ELTD1	219134_at		0.815595	high.cor
Endothel	5787	PTPRB	217177_s_at		0.767047	high.cor
Endothel	7010	TEK	206702_at		0.883135	high.cor
Collag-Stroma	165	AEBP1	201792_at	Stroma	0.938424	high.cor
Collag-Stroma	1303	COL12A1			0.965440	high.cor

TNBCmetagene_RNA-Seq	EntrezID	Gene	Affymetrix probeset	TNBCmetagene_Affymetrix (PMID_21978456)	Correlation RNA-Seq vs Affymetrix (PMID_25412710)	Corelation Class (PMID_25412710)
Collag-Stroma	1289	COL5A1	212488_at	Stroma	0.890394	high.cor
Collag-Stroma	1277	COL1A1	202311_s_at	Stroma	0.916062	high.cor
Collag-Stroma	1278	COL1A2	202404_s_at	Stroma	0.945735	high.cor
Collag-Stroma	1281	COL3A1	211161_s_at	Stroma	0.905483	high.cor
Collag-Stroma	1290	COL5A2	221729_at	Stroma	0.899922	high.cor
Collag-Stroma	1293	COL6A3	201438_at	Stroma	0.944905	high.cor
Collag-Stroma	5159	PDGFRB	202273_at	Stroma	0.834975	high.cor
Collag-Stroma	83468	GLT8D2	221447_s_at		0.880931	high.cor
Collag-Stroma	4313	MMP2	201069_at	Stroma	0.927081	high.cor
Collag-Stroma	54796	BNC2	220272_at		0.836661	high.cor
Collag-Stroma	57616	TSHZ3			0.817475	high.cor
Collag-Stroma	1009	CDH11	207173_x_at	Stroma	0.855198	high.cor
Collag-Stroma	254228	FAM26E			0.610513	low.cor
Collag-Stroma	2200	FBN1	202766_s_at	Stroma	0.832123	high.cor
Collag-Stroma	51339	DACT1	219179_at		0.806585	high.cor
Collag-Stroma	2191	FAP	209955_s_at	Stroma	0.743194	high.cor
Collag-Stroma	1634	DCN	209335_at	Stroma	0.892339	high.cor
Collag-Stroma	4060	LUM	201744_s_at	Stroma	0.798613	high.cor
Collag-Stroma	283298	OLFML1	217525_at		0.848133	high.cor
Collag-Stroma	10631	POSTN	210809_s_at	Stroma	0.841716	high.cor
Collag-Stroma	22795	NID2	204114_at	Stroma	0.898367	high.cor
Collag-Stroma	6678	SPARC	212667_at	Stroma	0.863041	high.cor
Collag-Stroma	7070	THY1	213869_x_at	Stroma	0.855328	high.cor
Collag-Stroma	57125	PLXDC1	219700_at		0.747861	high.cor
Adipo	2167	FABP4	203980_at	Adipocyte	0.931747	high.cor
Adipo	729	C6	210168_at		0.650570	low.cor
Adipo	63924	CIDEC	219398_at		0.875292	high.cor
Adipo	364	AQP7				
Adipo	125	ADH1B	209613_s_at	Adipocyte	0.956378	high.cor
Adipo	9370	ADIPOQ	207175_at	Adipocyte	0.938553	high.cor
Adipo	5346	PLIN1	205913_at	Adipocyte	0.894089	high.cor
Adipo	286753	TUSC5			0.505769	low.cor
Adipo	729359	PLIN4			0.647459	low.cor
Hemoglobin	3040	HBA2				
Hemoglobin	3043	HBB	211696_x_at	Hemoglobin	0.884690	high.cor
HOXA	3206	HOXA10	213150_at	HOXA	0.823892	high.cor
HOXA	3207	HOXA11	213823_at	HOXA	0.638838	low.cor
HOXA	221883	HOXA11AS				
HOXA	3205	HOXA9				
HOXA	3198	HOXA1	214639_s_at		0.754019	high.cor
HOXA	3201	HOXA4	206289_at	HOXA	0.678636	low.cor
HOXA	3199	HOXA2	214457_at		0.622245	low.cor
HOXA	3200	HOXA3	208604_s_at		0.848587	high.cor
HOXA	3202	HOXA5	213844_at	HOXA	0.863430	high.cor
HOXA	3203	HOXA6	208557_at		0.638450	low.cor
HOXA	3204	HOXA7	206847_s_at	HOXA	0.864597	high.cor
Histone	8969	HIST1H2AG				
Histone	8349	HIST2H2BE	202708_s_at	Histone	0.924423	high.cor
Histone	440689	HIST2H2BF				
Histone	8343	HIST1H2BF				
Histone	3017	HIST1H2BD	209911_x_at	Histone	0.826484	high.cor
Histone	3012	HIST1H2AE	214469_at	Histone	0.854939	high.cor
Histone	8339	HIST1H2BG	215779_s_at	Histone	0.487101	low.cor
Histone	8351	HIST1H3D	214472_at			
Histone	8334	HIST1H2AC				
Histone	8347	HIST1H2BC				
Histone	8365	HIST1H4H	208180_s_at	Histone		
Histone	8344	HIST1H2BE				
Histone	85236	HIST1H2BK	209806_at	Histone	0.847874	high.cor
Histone	3006	HIST1H1C	209398_at	Histone	0.876718	high.cor
Histone	8337	HIST2H2AA3				
Histone	8336	HIST1H2AM	214481_at			
Histone	8970	HIST1H2BJ	214502_at		0.662886	low.cor
Histone	8348	HIST1H2BO	214540_at		0.520936	low.cor
Histone	26212	OR2B6	216522_at		0.663404	low.cor
Prolif	51514	DTL	218585_s_at	Proliferation	0.885144	high.cor
Prolif	4751	NEK2	204641_at	Proliferation	0.979258	high.cor
Prolif	25896	INTS7	218783_at		0.551854	low.cor
Prolif	79915	ATAD5	220223_at		0.730166	high.cor
Prolif	699	BUB1	209642_at	Proliferation	0.966684	high.cor
Prolif	150468	CKAP2L			0.895579	high.cor
Prolif	7153	TOP2A	201292_at	Proliferation	0.955022	high.cor
Prolif	1063	CENPF	209172_s_at	Proliferation	0.912626	high.cor
Prolif	259266	ASPM	219918_s_at	Proliferation	0.968693	high.cor
Prolif	9928	KIF14	206364_at	Proliferation	0.932136	high.cor
Prolif	6491	STIL	205339_at		0.904978	high.cor
Prolif	56992	KIF15	219306_at		0.945035	high.cor
Prolif	151648	SGOL1			0.772945	high.cor
Prolif	23397	NCAPH	212949_at		0.956702	high.cor
Prolif	2305	FOXM1	202580_x_at	Proliferation	0.971805	high.cor
Prolif	9918	NCAPD2	201774_s_at		0.778714	high.cor
Prolif	10635	RAD51AP1	204146_at		0.952489	high.cor
Prolif	171017	ZNF384	212369_at		0.500065	low.cor
Prolif	78995	C17orf53	219879_s_at		0.509269	low.cor
Prolif	80174	DBF4B			0.356819	low.cor
Prolif	146909	KIF18B	222039_at	Proliferation	0.939914	high.cor
Prolif	3833	KIFC1	209680_s_at		0.543039	low.cor
Prolif	113130	CDCA5			0.968175	high.cor
Prolif	991	CDC20	202870_s_at	Proliferation	0.971092	high.cor
Prolif	55143	CDCA8	221520_s_at	Proliferation	0.955211	high.cor
Prolif	11004	KIF2C	209408_at	Proliferation	0.979129	high.cor
Prolif	8438	RAD54L	204558_at		0.913858	high.cor
Prolif	4998	ORC1L	205085_at		0.834651	high.cor
Prolif	9212	AURKB	209464_at		0.934664	high.cor
Prolif	54478	FAM64A	221591_s_at		0.932460	high.cor
Prolif	10024	TROAP	204649_at		0.810539	high.cor
Prolif	11065	UBE2C	202954_at	Proliferation	0.964480	high.cor
Prolif	83461	CDCA3	221436_s_at	Proliferation	0.967786	high.cor

R-MarkDown-document: TNBC_TIL_analysis

Thomas Karn

May-18 2017

Table of Contents

SECTION-1 Selection of a gene expression based TNBC cohort from TCGA 1
1.1 Analysis of correlation of ESR1 gene expression from RNA-Seq and Agilent microarray platform 1
1.2 Analysis of correlation of PGR gene expression from RNA-Seq and Agilent microarray platform 3
1.3 Analysis of correlation of HER2 gene expression from RNA-Seq and Agilent microarray platform 4
1.4 Generate TNBC dataset 5
SECTION-2 Gene filtering in RNA-Seq data 8
SECTION-3 Metagene construction 10
3.1 Metagene genes: RNA-Seq vs. Affy correlation 10
3.2 Metagene calculation from RNA-Seq expression 12
SECTION-4 MATH analysis of dispersion in mutant allele frequencies 13
SECTION-5 Survival analysis 14
5.1 MHC2/IL8VEGF signature 16
5.2 B-Cell/IL8VEGF signature 18

SECTION-1 Selection of a gene expression based TNBC cohort from TCGA

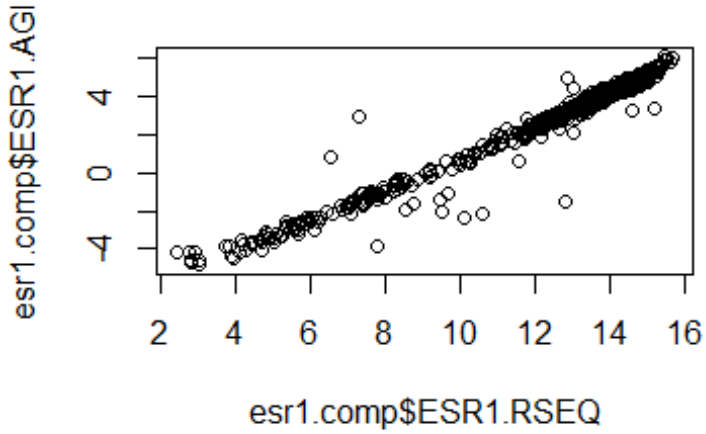
We use the cgdsr package to access data from the cBIO Portal.

```
library("cgdsr")
cbiop <- CGDS("http://www.cbioportal.org/public-portal/")
# getCancerStudies(cbiop)$cancer_study_id
clidat = getClinicalData(cbiop,"brca_tcga_all")
```

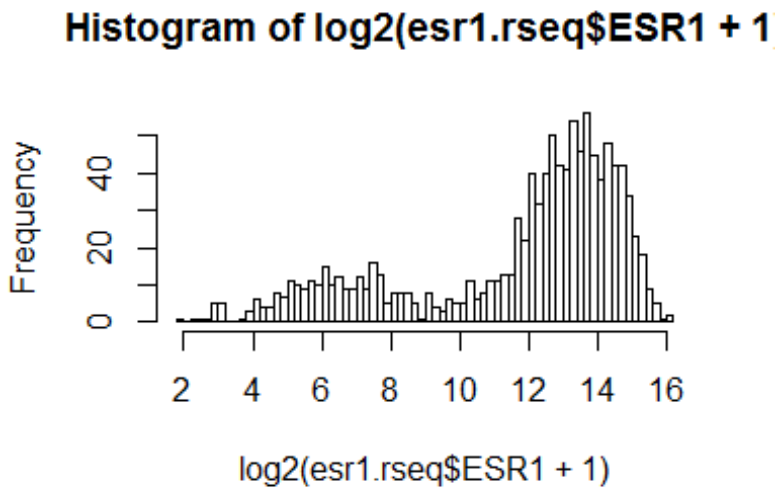
1.1 Analysis of correlation of ESR1 gene expression from RNA-Seq and Agilent microarray platform

```
esr1.rseq = getProfileData(cbiop,"ESR1","brca_tcga_rna_seq_v2_mrna",
"brca_tcga_all")
esr1.agi = getProfileData(cbiop,"ESR1","brca_tcga_mrna", "brca_tcga_all")
```

```r
# generate matrix of cases with both data for Agilent and RNA-Seq:
esr1.comp=as.data.frame(cbind(esr1.agi$ESR1, log2(esr1.rseq$ESR1+1))
        [(!is.nan(esr1.agi$ESR1)) & (!is.nan(esr1.rseq$ESR1)), ])
colnames(esr1.comp)=c("ESR1.AGI", "ESR1.RSEQ")

# correlation between Agilent and RNA-Seq:
plot(esr1.comp$ESR1.RSEQ, esr1.comp$ESR1.AGI)
```

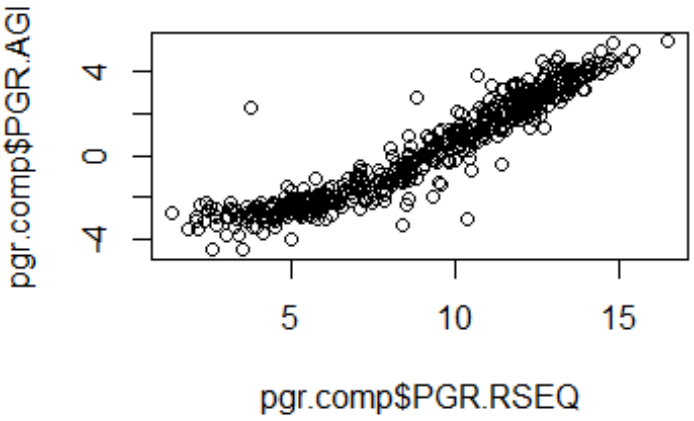



```r
cor(esr1.comp$ESR1.RSEQ, esr1.comp$ESR1.AGI)

## [1] 0.9821414

# bimodal distribution of RNA-Seq data
hist(log2(esr1.rseq$ESR1+1), breaks=80)
```

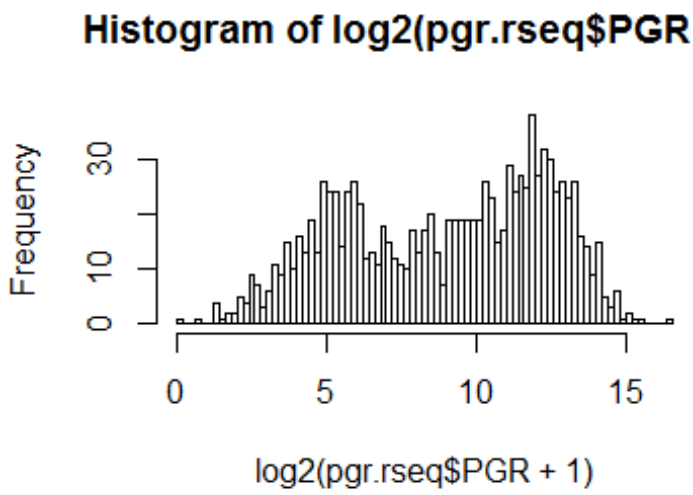


1.2 Analysis of correlation of PGR gene expression from RNA-Seq and Agilent microarray platform

```
pgr.rseq = getProfileData(cbiop,"PGR","brca_tcga_rna_seq_v2_mrna",
"brca_tcga_all")
pgr.agi = getProfileData(cbiop,"PGR","brca_tcga_mrna", "brca_tcga_all")

# generate matrix of cases with both data for Agilent and RNA-Seq:
pgr.comp=as.data.frame(cbind(pgr.agi$PGR, log2(pgr.rseq$PGR+1))
          [(!is.nan(pgr.agi$PGR)) & (!is.nan(pgr.rseq$PGR)), ])
colnames(pgr.comp)=c("PGR.AGI", "PGR.RSEQ")
# correlation between Agilent and RNA-Seq:
plot(pgr.comp$PGR.RSEQ, pgr.comp$PGR.AGI)
```

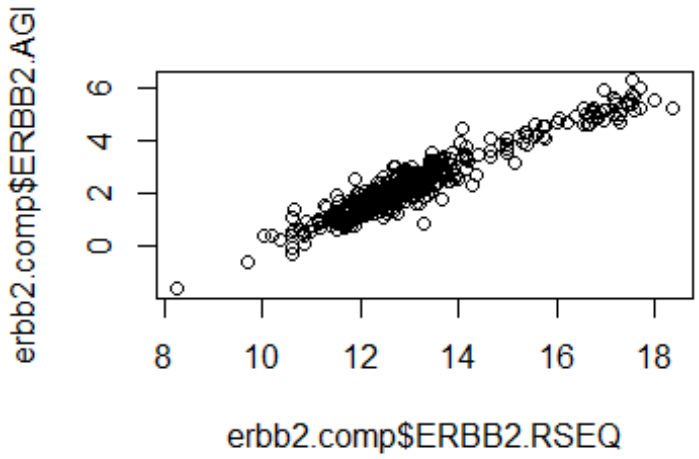



```
cor(pgr.comp$PGR.RSEQ, pgr.comp$PGR.AGI)

## [1] 0.9499931

# bimodal distribution of RNA-Seq data
hist(log2(pgr.rseq$PGR+1), breaks=80)
```



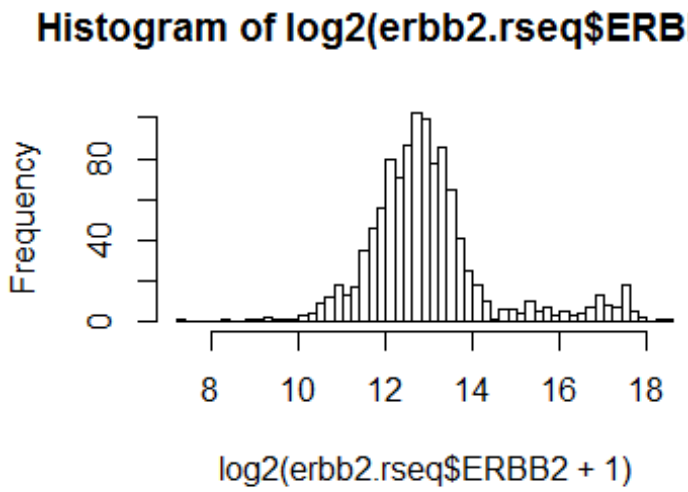
1.3 Analysis of correlation of HER2 gene expression from RNA-Seq and Agilent microarray platform

```
erbb2.rseq = getProfileData(cbiop,"ERBB2","brca_tcga_rna_seq_v2_mrna",
"brca_tcga_all")
erbb2.agi = getProfileData(cbiop,"ERBB2","brca_tcga_mrna", "brca_tcga_all")

# generate matrix of cases with both data for Agilent and RNA-Seq:
erbb2.comp=as.data.frame(cbind(erbb2.agi$ERBB2, log2(erbb2.rseq$ERBB2+1))
            [(!is.nan(erbb2.agi$ERBB2)) & (!is.nan(erbb2.rseq$ERBB2)), ])
colnames(erbb2.comp)=c("ERBB2.AGI", "ERBB2.RSEQ")
# correlation between Agilent and RNA-Seq:
plot(erbb2.comp$ERBB2.RSEQ, erbb2.comp$ERBB2.AGI)
```




```
cor(erbb2.comp$ERBB2.RSEQ, erbb2.comp$ERBB2.AGI)
```

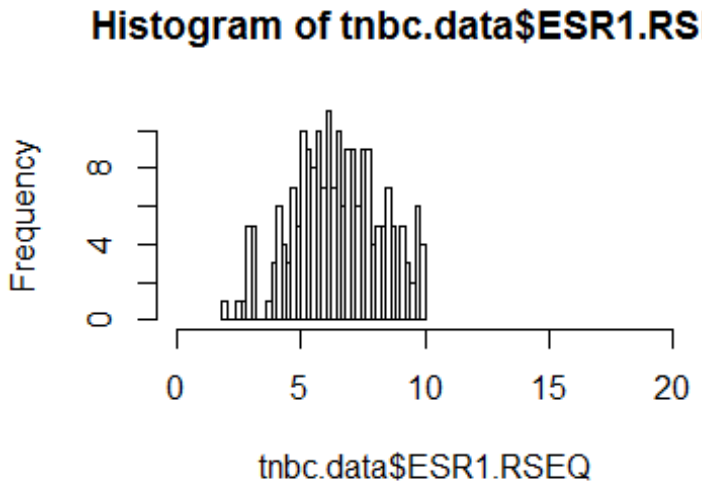
```
## [1] 0.9547622

# bimodal distribution of RNA-Seq data
hist(log2(erbb2.rseq$ERBB2+1), breaks=80)
```



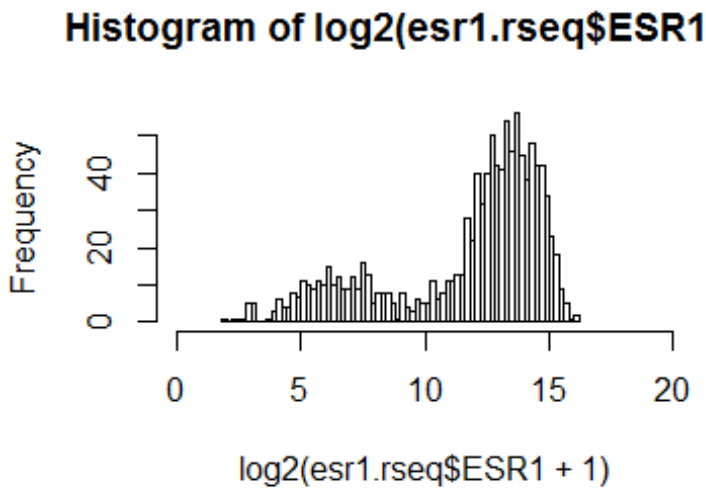
1.4 Generate TNBC dataset

```
# Select tnbc/dnbc based on cutoffs from distribution of RNA-Seq
# define a logical selection vector
tnbc.group= !is.na(esr1.rseq) & !is.na(erbb2.rseq) &
    (log2(esr1.rseq$ESR1+1)<10) &  (log2(erbb2.rseq$ERBB2+1)<14)
colnames(tnbc.group)="tnbc"
sum(na.omit(tnbc.group))

## [1] 208

# Generate tnbc dataset
tnbc.data= cbind(log2(esr1.rseq$ESR1+1)[tnbc.group],
                 log2(pgr.rseq$PGR+1)[tnbc.group],
                 log2(erbb2.rseq$ERBB2+1)[tnbc.group])
row.names(tnbc.data)= row.names(tnbc.group)[tnbc.group]
colnames(tnbc.data)=c("ESR1.RSEQ", "PGR.RSEQ", "ERBB2.RSEQ")

# Merge of Clinical data and tnbc dataset
# find subset in clidat corresponding to tnbc
clidat.sel=clidat[row.names(clidat)%in% row.names(tnbc.data),]
# merge tnbc.data and clinical data, left outer join:
tnbc.data= merge(tnbc.data, clidat.sel, by="row.names", all.x =TRUE)
#  "merge" creates resorted dataframe with the row.names
#      as a new first column "Row.names"
# rebuild structure (row.names):
row.names(tnbc.data)=tnbc.data$Row.names
tnbc.data=tnbc.data[,colnames(tnbc.data)!= "Row.names"]
```

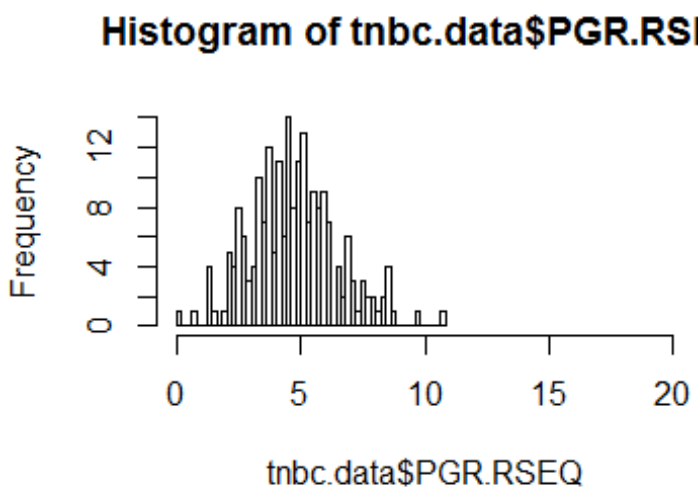
```r
# check residual receptor expression in tnbc dataset:
hist(tnbc.data$ESR1.RSEQ, xlim=c(0,20), breaks=40) # tnbc group
```



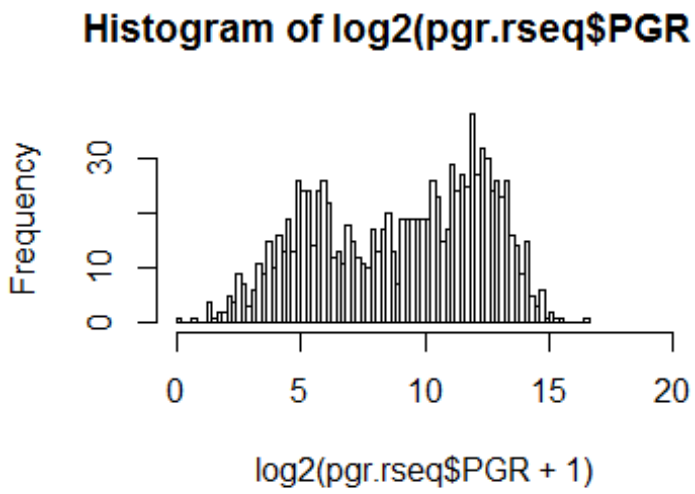

```r
hist(log2(esr1.rseq$ESR1+1),xlim=c(0,20), breaks=80) # all samples
```



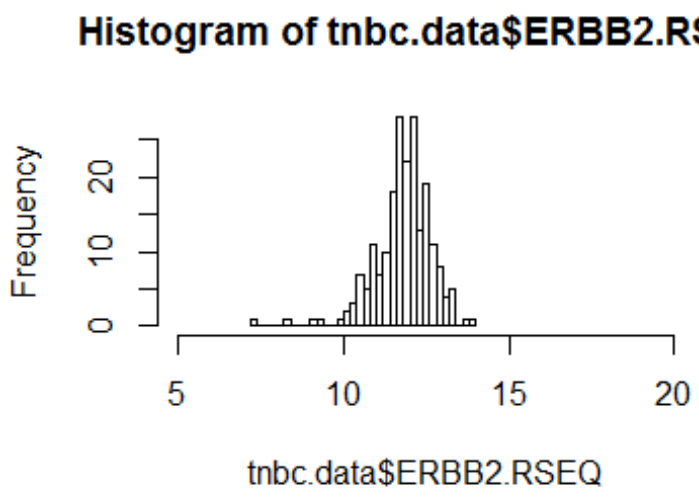

```r
hist(tnbc.data$PGR.RSEQ, xlim=c(0,20), breaks=40) # tnbc group
```



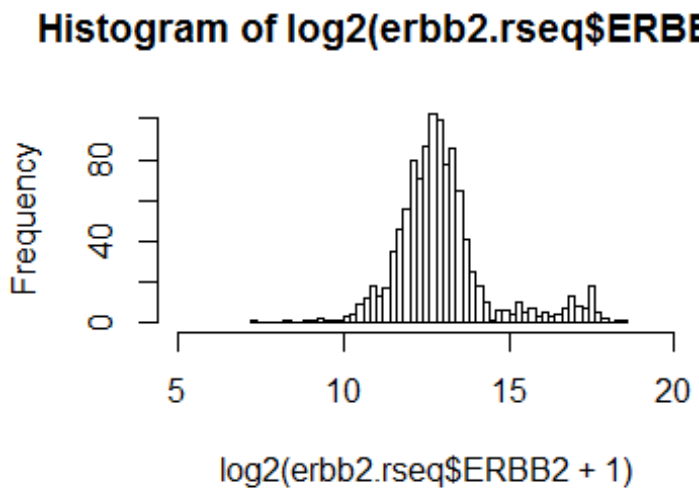

```r
hist(log2(pgr.rseq$PGR+1),xlim=c(0,20), breaks=80) # all samples
```




```r
hist(tnbc.data$ERBB2.RSEQ, xlim=c(5,20),  breaks=40) # tnbc group
```

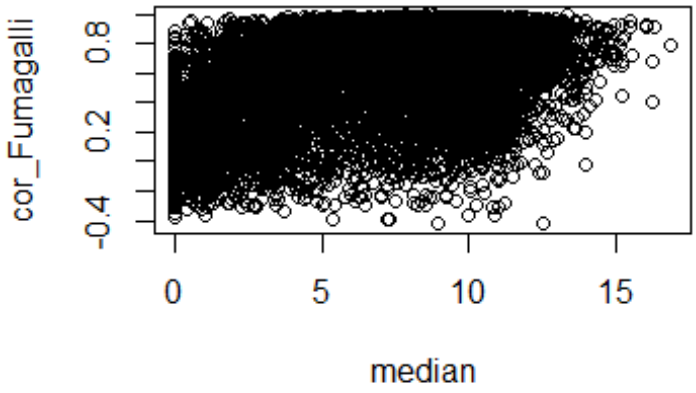



```
hist(log2(erbb2.rseq$ERBB2+1),xlim=c(5,20), breaks=80) # all samples
```

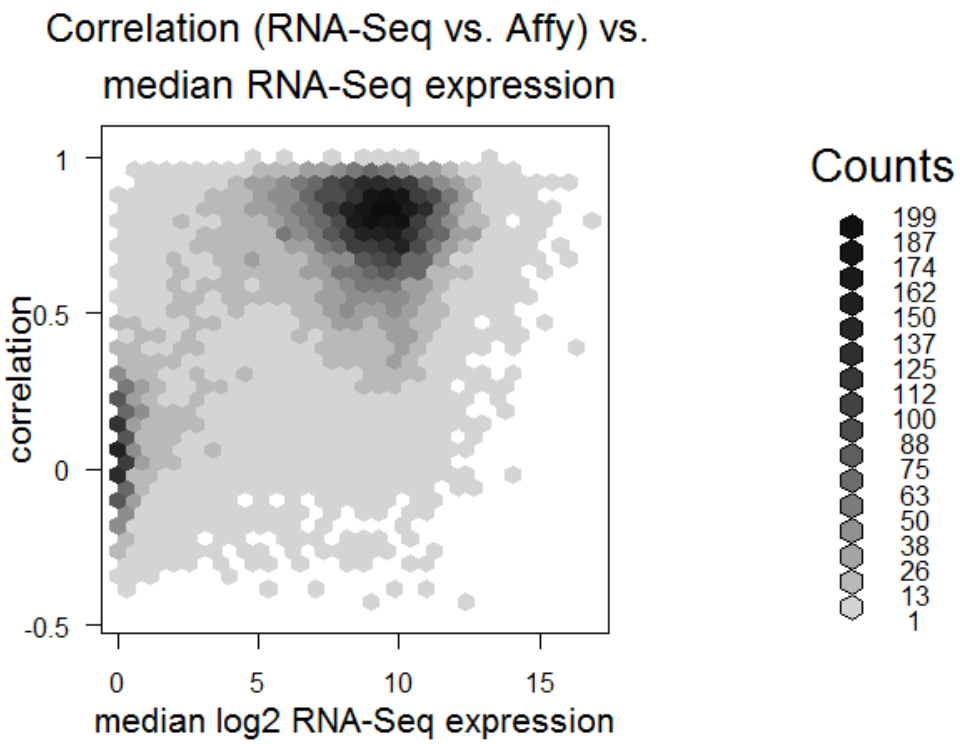


SECTION-2 Gene filtering in RNA-Seq data

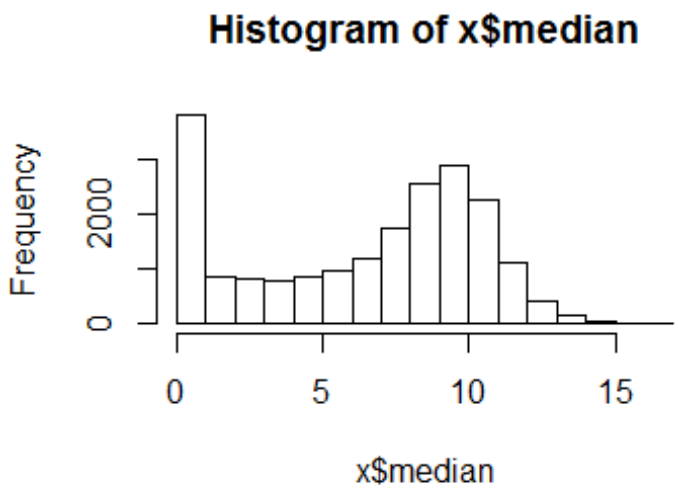
```
# Spearman correlation values between RNA-Seq and Affymetrix microarray
#  for 16,097 Jetset probes for 57 paired frozen breast cancer samples
#  can be obtained from:
#  Suppl.Tab.S2 of Fumagalli et al. 2014, PubmedID 25412710

n208.FumagCorrel <-
read.delim("2016_05_31_median_mean_n208RNASeq_vs_FumagalliCorrel.txt")

# Plot median expression vs Spearman correlation coefficient
x=n208.FumagCorrel[,c(1,3)]
plot(x)
```



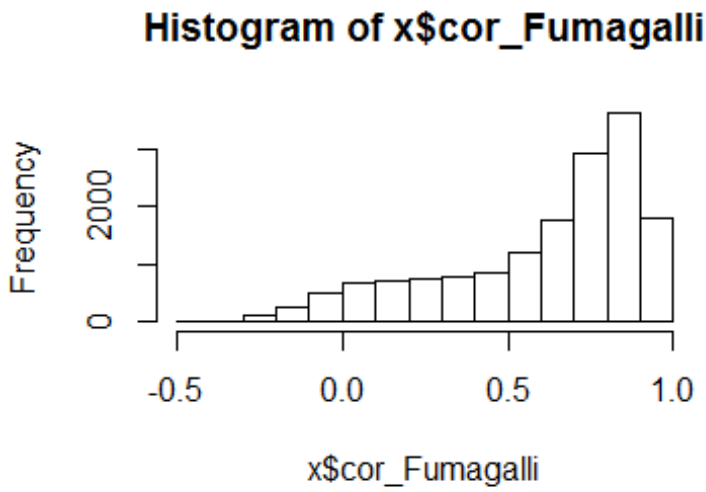

```r
# Use hexbin plot to display the density of the scatter
library(hexbin)
plot(hexbin(x$median, x$cor_Fumagalli, xbins=30),
     xlab="median log2 RNA-Seq expression", ylab="correlation",
     main="Correlation (RNA-Seq vs. Affy) vs. \n median RNA-Seq expression")
```




```r
# Distribution of median expression values
hist(x$median)
```




```
# Distribution of Spearman correlation coefficients
hist(x$cor_Fumagalli)
```

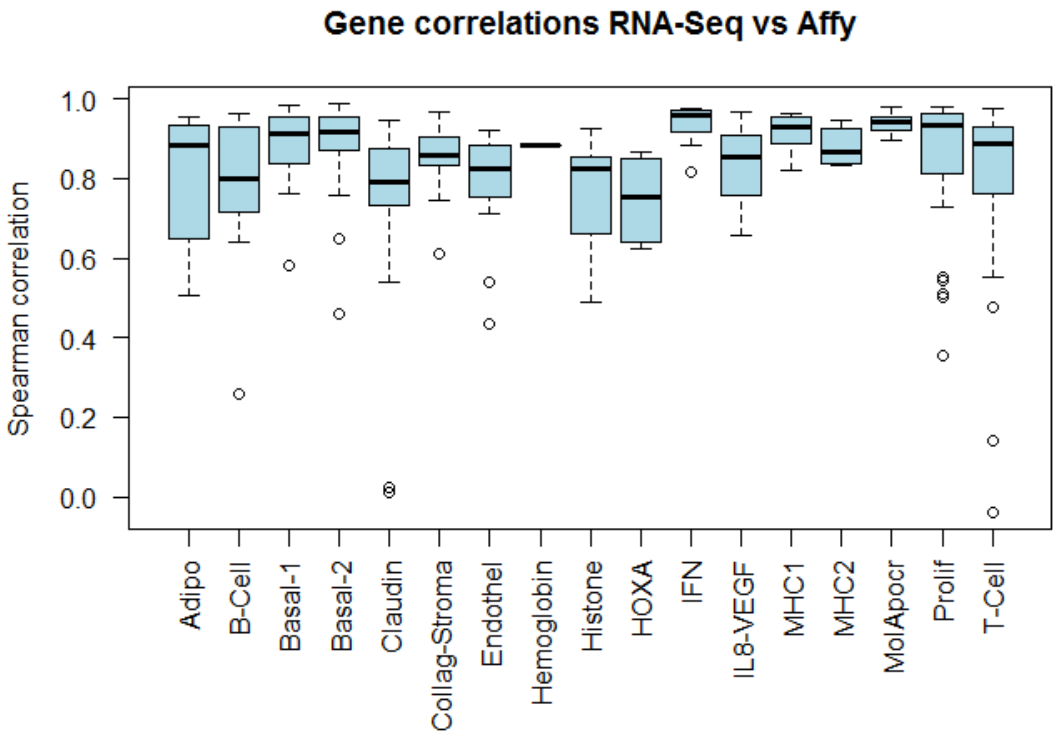



```
rm(x)
```

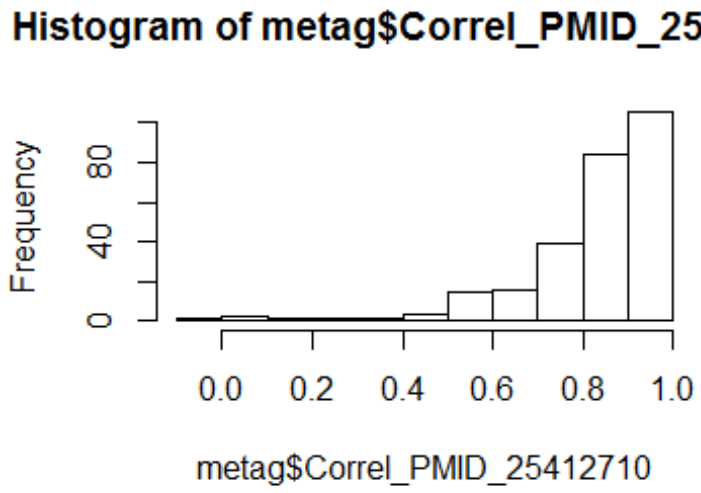
SECTION-3 Metagene construction

3.1 Metagene genes: RNA-Seq vs. Affy correlation

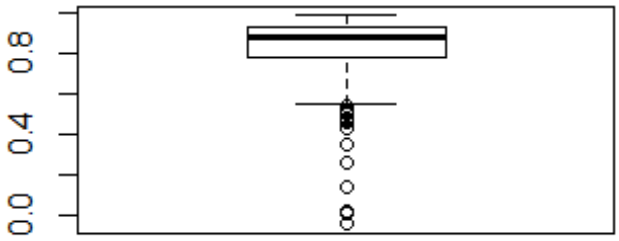
```
metag <- read.delim("2016_06_01_TNBC-metagenes_gene_list.txt")

par(las = 2) # labels always perpendicular to the axis
par(mar=c(7,4,4,2)+0.1) # increase bottom margin
boxplot(Correl_PMID_25412710~TNBCmetagene_RNA.Seq,
        data=metag, notch=F, col="lightblue",
        ylab="Spearman correlation",
        main="Gene correlations RNA-Seq vs Affy" )
```




```
par(mar=c(5.1, 4.1, 4.1, 2.1))
hist(metag$Correl_PMID_25412710)
```




```
boxplot(metag$Correl_PMID_25412710)
```




```
median(metag$Correl_PMID_25412710, na.rm=T)

## [1] 0.8831346

summary(metag$Correl_PMID_25412710, na.rm=T)

##     Min.  1st Qu.   Median     Mean  3rd Qu.     Max.     NA's
## -0.04109  0.77800  0.88310  0.82610  0.93210  0.98810       35
```

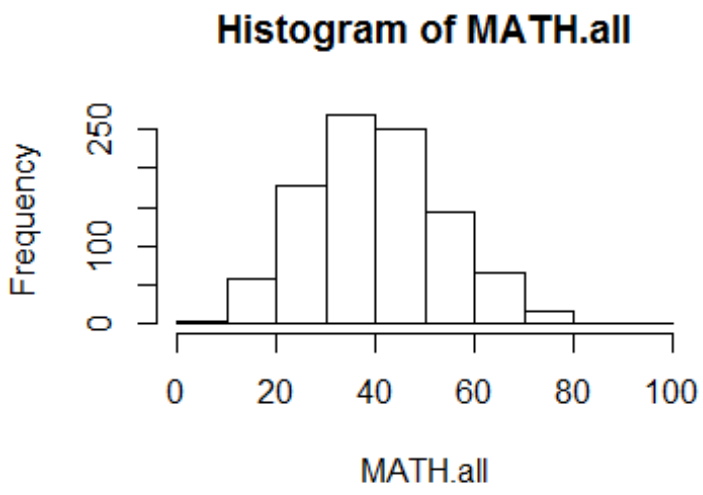
3.2 Metagene calculation from RNA-Seq expression

```
# load RNAseq data of 304 genes for 208 tnbc samples
# RNAseq data of 1218 TCGA BRCA can be downloaded from UCSC Xena browser
(https://tcga.xenahubs.net/download/TCGA.BRCA.sampleMap/HiSeqV2)

n304genes <- read.table("n208tnbc_n304genes_RNAseq.csv", header=TRUE,
sep=";")

# scale transposed expression data and re-transpose
n304.expr.sca= t(scale(t(n304genes[,5:212])))
colnames(n304.expr.sca)=colnames(n304genes[,5:212])

# calculate mean expression of each metag-cluster from scaled expression for
17 metagenes
metag17=array(NA,dim=c(0,17))
for (i in 1: ncol(n304.expr.sca)) {
  mdf= as.data.frame(as.list(by(n304.expr.sca[,i],
                                n304genes$MetagCluster17, mean)))
  rownames(mdf)=colnames(n304.expr.sca)[i]
  metag17=rbind(metag17, mdf)
}
rm(mdf)

# merge 17 metagene expression data with tnbc.data dataframe, left outer
join:
```

```
tnbc.data.meta17= merge(tnbc.data, metag17, by="row.names", all.x =TRUE)
# "merge" command results in resorting of dataframe and loss of row.names
#    but an additional new first column "Row.names"
# Assign new row.names from this additional column and then delete it
row.names(tnbc.data.meta17)=tnbc.data.meta17$Row.names
tnbc.data.meta17=tnbc.data.meta17[,colnames(tnbc.data.meta17)!= "Row.names"]
```

SECTION-4 MATH analysis of dispersion in mutant allele frequencies

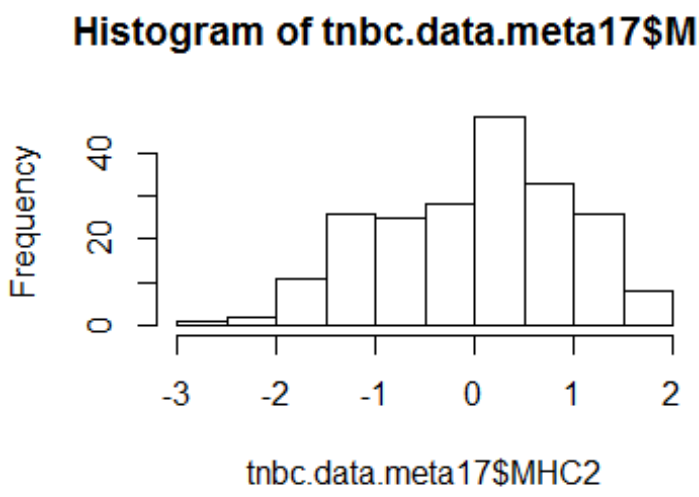
```
# Copy of maf file from TCGA
genome.wustl.edu_BRCA.IlluminaGA_DNASeq.Level_2.1.1.0.curated.somatic.maf.txt
(52MB) is available at https://portal.gdc.cancer.gov/legacy-
archive/files/50d6fb1d-5bb1-4a30-9e91-6d45bd9b1c3f

# The required variant allele frequencies have been extracted in the smaller
file used here: "VAF-
table_genome.wustl.edu_BRCA.IlluminaGA_DNASeq.Level_2.1.1.0.curated.somatic.m
af.txt""

maf.download <- read.delim(
    "VAF-
table_genome.wustl.edu_BRCA.IlluminaGA_DNASeq.Level_2.1.1.0.curated.somatic.m
af.txt")

all.maf = maf.download[,c("Hugo_Symbol", "Tumor_Sample_Barcode",
"tumor_vaf")]

TCGA_Sample=substr(all.maf$Tumor_Sample_Barcode, 1, 15)

all.maf = cbind(TCGA_Sample, all.maf)

# calculate for each sample the median of tumor_vaf values
med=by(all.maf$tumor_vaf, all.maf$TCGA_Sample, median)

# convert list to dataframe and transpose
med.df = t(as.data.frame(as.list(med)))
colnames(med.df)= "med.mut.AF"

# calculate MAD (Median Absolute Deviation) for each sample
MAD=by(all.maf$tumor_vaf, all.maf$TCGA_Sample, mad)

# convert list to dataframe and transpose
MAD.df= t(as.data.frame(as.list(MAD)))
colnames(MAD.df)= "MAD.mut.AF"

# calculate MATH (Mutant Allele Tumor Heterogeneity) as MATH=100*MAD/median
MATH.all =100 * MAD.df / med.df
colnames(MATH.all)= "MATH"

hist(MATH.all)
```



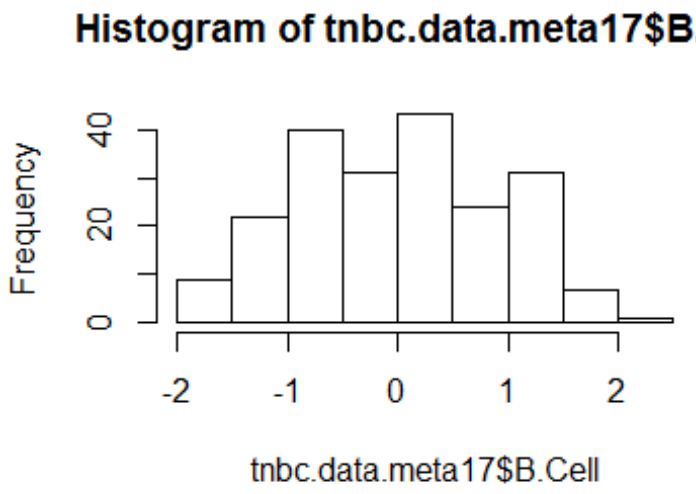

```
# Export MATH values:

# write.table(MATH.all, file="n982TCGA_MATH.txt",
#         row.names=TRUE, col.names = NA, quote=FALSE, sep="\t")
```

SECTION-5 Survival analysis

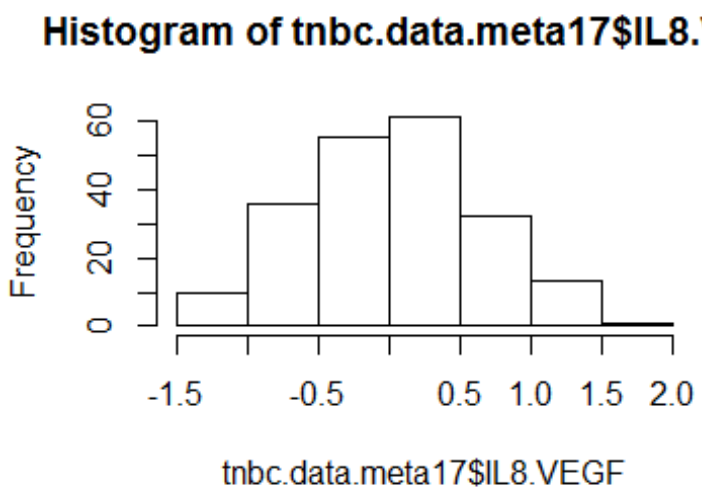
```
library("survival")

# Censor DFS at 120 months
dfs.120=tnbc.data.meta17$DFS_MONTHS
ev.120=tnbc.data.meta17$DFS_STATUS

for (i in 1:nrow(tnbc.data.meta17)) {
     if (is.na(tnbc.data.meta17$DFS_MONTHS[i]))
       {dfs.120[i]=NA ; ev.120[i]=NA}
       else
         { if (tnbc.data.meta17$DFS_MONTHS[i] > 120)
           {dfs.120[i]=120 ; ev.120[i]="DiseaseFree"}
           else {dfs.120[i]=tnbc.data.meta17$DFS_MONTHS[i] ;
ev.120=tnbc.data.meta17$DFS_STATUS}
         }
     }

# Add censored DFS to dataframe
tnbc.data.meta17=cbind(tnbc.data.meta17, dfs.120, ev.120)

# Distributions of MHC2 metagene, B-Cell metagen, and IL8VEGF metagene
hist(tnbc.data.meta17$MHC2)
```




```
hist(tnbc.data.meta17$B.Cell)
```




```
hist(tnbc.data.meta17$IL8.VEGF)
```



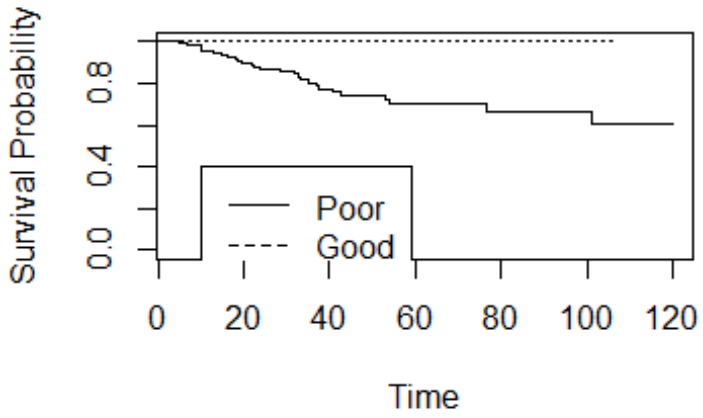

```
# Since no clear bimodality observed in distributions,
# we stay with previously established cutoffs for metagenes/signatures:
# MHC2 metagene: Upper quartile (Rody 2009, PMID 19272155)
# B-Cell metagene: Lower quartile (Rody 2011, PMID 21978456)
# IL8.VEGF metagene: Median split (Rody 2011, PMID 21978456)
```

5.1 MHC2/IL8VEGF signature

```
# Define upper quartile MHC2 metagene (based on Rody 2009, PMID 19272155)
MHC2.q4=tnbc.data.meta17$MHC2 > quantile(tnbc.data.meta17$MHC2, probs=0.75)
# Define below median IL8.VEGF metagene (cutoff from Rody 2011, PMID
21978456)
IL8.VEGF.q12=tnbc.data.meta17$IL8.VEGF < quantile(tnbc.data.meta17$IL8.VEGF,
probs=0.5)
# Define prognostic signature
MHC2.IL8.VEGF.sig = MHC2.q4 & IL8.VEGF.q12


## Check MHC2.IL8.VEGF.sig in Survival analysis
time=tnbc.data.meta17$dfs.120
censor= (tnbc.data.meta17$ev.120 =="Recurred/Progressed")
strata= MHC2.IL8.VEGF.sig
test=survfit(Surv(time, censor)~strata,conf.type="none")
summary(test)

## Call: survfit(formula = Surv(time, censor) ~ strata, conf.type = "none")
##
## 14 observations deleted due to missingness
##                 strata=FALSE
##    time n.risk n.event survival std.err
##    5.09    151       1    0.993 0.00660
##    6.80    149       1    0.987 0.00933
##    7.79    145       1    0.980 0.01149
##    9.89    138       1    0.973 0.01342
```

```
##    10.02    135       1    0.966 0.01513
##    10.28    134       1    0.958 0.01665
##    12.55    128       1    0.951 0.01812
##    12.71    126       1    0.943 0.01949
##    14.98    113       1    0.935 0.02103
##    16.10    109       1    0.926 0.02252
##    18.27    103       1    0.917 0.02403
##    18.50    102       1    0.908 0.02542
##    19.32     99       1    0.899 0.02677
##    21.91     89       1    0.889 0.02831
##    22.40     88       1    0.879 0.02974
##    23.95     82       1    0.868 0.03125
##    28.22     74       1    0.857 0.03295
##    31.90     69       1    0.844 0.03474
##    32.65     67       1    0.832 0.03643
##    33.31     63       1    0.818 0.03817
##    35.22     57       1    0.804 0.04011
##    36.79     53       1    0.789 0.04212
##    37.32     52       1    0.774 0.04396
##    40.70     47       1    0.757 0.04600
##    42.81     44       1    0.740 0.04807
##    53.02     37       1    0.720 0.05076
##    53.88     36       1    0.700 0.05315
##    76.54     21       1    0.667 0.06017
##   101.05     11       1    0.606 0.07957
##
##                 strata=TRUE
##      time n.risk n.event survival std.err
```

```r
plot(test, lty=c(1,3), xlab="Time", ylab="Survival Probability")
legend(10, 0.4, c("Poor", "Good") , lty=c(1,2))
```



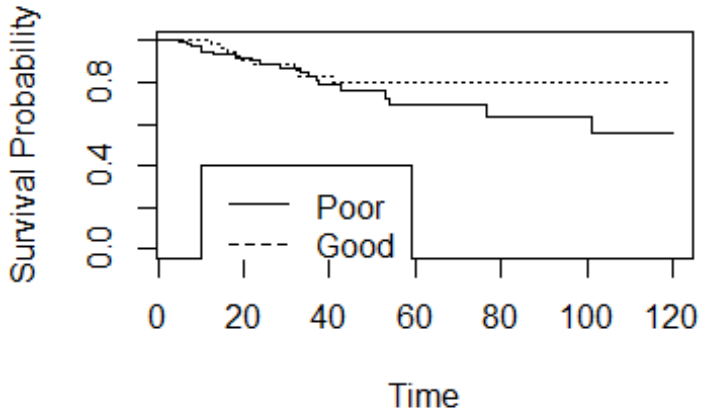
5.2 B-Cell/IL8VEGF signature

```
# Define B-Cell metagene above lowest quartile (cutoff from Rody 2011, PMID
21978456)
B.Cell.q234=tnbc.data.meta17$B.Cell > quantile(tnbc.data.meta17$B.Cell,
probs=0.25)
# Define below median IL8.VEGF metagene (cutoff from Rody 2011, PMID
21978456)
IL8.VEGF.q12=tnbc.data.meta17$IL8.VEGF < quantile(tnbc.data.meta17$IL8.VEGF,
probs=0.5)
# Define prognostic signature
B.Cell.IL8.VEGF.sig = B.Cell.q234 & IL8.VEGF.q12


## Check B.Cell.IL8.VEGF.sig in Survival analysis
time=tnbc.data.meta17$dfs.120
censor= (tnbc.data.meta17$ev.120 =="Recurred/Progressed")
strata= B.Cell.IL8.VEGF.sig
test=survfit(Surv(time, censor)~strata,conf.type="none")
summary(test)
```

```
## Call: survfit(formula = Surv(time, censor) ~ strata, conf.type = "none")
##
## 14 observations deleted due to missingness
##                 strata=FALSE
##    time n.risk n.event survival std.err
##    5.09    108       1    0.991 0.00922
##    6.80    106       1    0.981 0.01303
##    7.79    102       1    0.972 0.01607
##    9.89     97       1    0.962 0.01877
##   10.02     95       1    0.952 0.02113
##   10.28     94       1    0.942 0.02320
##   12.71     89       1    0.931 0.02524
##   18.27     71       1    0.918 0.02808
##   21.91     62       1    0.903 0.03129
##   23.95     58       1    0.887 0.03440
##   28.22     52       1    0.870 0.03774
##   33.31     45       1    0.851 0.04156
##   35.22     41       1    0.830 0.04544
##   36.79     37       1    0.808 0.04944
##   37.32     36       1    0.785 0.05292
##   42.81     29       1    0.758 0.05761
##   53.02     22       1    0.724 0.06448
##   53.88     21       1    0.689 0.07002
##   76.54     13       1    0.636 0.08230
##  101.05      8       1    0.557 0.10355
##
##                 strata=TRUE
##  time n.risk n.event survival std.err
##  12.6     61       1    0.984  0.0163
##  15.0     54       1    0.965  0.0241
##  16.1     52       1    0.947  0.0299
```

```
##  18.5     49       1    0.928  0.0350
##  19.3     45       1    0.907  0.0398
##  22.4     43       1    0.886  0.0441
##  31.9     34       1    0.860  0.0499
##  32.6     32       1    0.833  0.0551
##  40.7     27       1    0.802  0.0611
```

```
plot(test, lty=c(1,3), xlab="Time", ylab="Survival Probability")
legend(10, 0.4, c("Poor", "Good") , lty=c(1,2))
```




```
dir()
```

```
## [1] "2016_05_31_median_mean_n208RNASeq_vs_FumagalliCorrel.txt"
## [2] "2016_06_01_TNBC-metagenes_gene_list.txt"
## [3] "n208tnbc_n304genes_RNAseq.csv"
## [4] "TNBC_TIL_analysis_2017_05_18.Rmd"
## [5] "TNBC_TIL_analysis_2017_05_18_files"
## [6] "VAF-
table_genome.wustl.edu_BRCA.IlluminaGA_DNASeq.Level_2.1.1.0.curated.somatic.m
af.txt"
```

```
sessionInfo()
```

```
## R version 3.3.2 (2016-10-31)
## Platform: x86_64-w64-mingw32/x64 (64-bit)
## Running under: Windows 10 x64 (build 14393)
##
## locale:
## [1] LC_COLLATE=German_Germany.1252  LC_CTYPE=German_Germany.1252
## [3] LC_MONETARY=German_Germany.1252 LC_NUMERIC=C
## [5] LC_TIME=German_Germany.1252
##
## attached base packages:
## [1] stats     graphics  grDevices utils     datasets  methods   base
##
```

```
## other attached packages:
## [1] survival_2.40-1 hexbin_1.27.1   cgdsr_1.2.5
##
## loaded via a namespace (and not attached):
##  [1] Rcpp_0.12.9       lattice_0.20-34   digest_0.6.12
##  [4] rprojroot_1.2     R.methodsS3_1.7.1 grid_3.3.2
##  [7] backports_1.0.5   magrittr_1.5      evaluate_0.10
## [10] stringi_1.1.2     R.oo_1.21.0       Matrix_1.2-8
## [13] rmarkdown_1.3     splines_3.3.2     tools_3.3.2
## [16] stringr_1.2.0     yaml_2.1.14       htmltools_0.3.5
## [19] knitr_1.15.1
```