Comprehensive Genomic Analysis Identifies Novel Subtypes and Targets of Triple-Negative Breast Cancer

Matthew D. Burstein[1], Anna Tsimelzon[2], Graham M. Poage[3], Kyle R. Covington[2], Alejandro Contreras[2,4], Suzanne A.W. Fuqua[2], Michelle I Savage[3], C. Kent Osborne[2], Susan G. Hilsenbeck[2], Jenny C. Chang[5], Gordon B. Mills[6], Ching C. Lau[7], and Powel H. Brown[3]

Abstract

Purpose: Genomic profiling studies suggest that triple-negative breast cancer (TNBC) is a heterogeneous disease. In this study, we sought to define TNBC subtypes and identify subtype-specific markers and targets.

Experimental Design: RNA and DNA profiling analyses were conducted on 198 TNBC tumors [estrogen receptor (ER) negativity defined as Allred scale value ≤ 2] with >50% cellularity (discovery set: $n = 84$; validation set: $n = 114$) collected at Baylor College of Medicine (Houston, TX). An external dataset of seven publically accessible TNBC studies was used to confirm results. DNA copy number, disease-free survival (DFS), and disease-specific survival (DSS) were analyzed independently using these datasets.

Results: We identified and confirmed four distinct TNBC subtypes: (i) luminal androgen receptor (AR; LAR), (ii) mesenchymal (MES), (iii) basal-like immunosuppressed (BLIS), and (iv) basal-like immune-activated (BLIA). Of these, prognosis is worst for BLIS tumors and best for BLIA tumors for both DFS (log-rank test: $P = 0.042$ and 0.041, respectively) and DSS (log-rank test: $P = 0.039$ and 0.029, respectively). DNA copy number analysis produced two major groups (LAR and MES/BLIS/BLIA) and suggested that gene amplification drives gene expression in some cases [FGFR2 (BLIS)]. Putative subtype-specific targets were identified: (i) LAR: androgen receptor and the cell surface mucin MUC1, (ii) MES: growth factor receptors [platelet-derived growth factor (PDGF) receptor A; c-Kit], (iii) BLIS: an immunosuppressing molecule (VTCN1), and (iv) BLIA: Stat signal transduction molecules and cytokines.

Conclusion: There are four stable TNBC subtypes characterized by the expression of distinct molecular profiles that have distinct prognoses. These studies identify novel subtype-specific targets that can be targeted in the future for the effective treatment of TNBCs. *Clin Cancer Res; 21(7); 1688–98. ©2014 AACR.*

See related commentary by Vidula and Rugo, p. 1511

Introduction

Recent studies have demonstrated that breast cancer heterogeneity extends beyond the classic immunohistochemistry (IHC)-based divisions of estrogen receptor (ER), progesterone receptor (PR), and human epidermal growth factor receptor 2 (Her2; ref. 1).

Nearly 10% to 20% of primary breast cancers are triple-negative breast cancers (TNBC; ref. 2), which lack expression of ER, PR, and Her2, present with higher grade, often contain mutations in TP53 (3), and have a poor prognosis (4). Molecularly targeted therapy has shown limited benefit so far in TNBCs, and although PARP inhibitors in the BRCA-mutant setting are promising (5, 6), new strategies for classifying and treating women affected by this aggressive disease are urgently needed.

The intrinsic subtyping of breast cancer by gene expression analyses (7) was recently supported by The Cancer Genome Atlas (TCGA) Program through mRNA, miRNA, DNA, and epigenetic analyses (8). The basal-like subtype, traditionally defined by RNA profiling or cytokeratin expression (9), accounts for 10% to 25% of all invasive breast cancers (10). In addition, basal-like breast cancers account for 47% to 88% of all TNBCs (8, 11, 12). Tumors of the "claudin-low" (CL) subtype (13, 14) have particularly poor prognoses than hormone-sensitive tumors (15). The results from an aggregate analysis of publically available expression datasets, in a study performed by Lehmann and colleagues. (12), suggested that TNBCs are more heterogeneous than previously described and identified 6 subtypes: (i) androgen receptor (AR)-positive, (ii) claudin-low–enriched mesenchymal, (iii) mesenchymal stem–like, (iv) immune response and 2 cell-cycle–disrupted basal subtypes: (v) BL-1 and (vi) BL-2. However, IHC detection of ER, PR, and Her2 protein is the clinical standard used to define TNBC.

[1]Structural and Computational Biology & Molecular Biophysics Graduate Program and Medical Scientist Training Program, Baylor College of Medicine, Houston, Texas. [2]Lester and Sue Smith Breast Center and Dan L. Duncan Cancer Center, Baylor College of Medicine, Houston, Texas. [3]Department of Clinical Cancer Prevention, MD Anderson Cancer Center, Houston, Texas. [4]Department of Pathology, Baylor College of Medicine, Houston, Texas. [5]The Methodist Hospital Research Institute, Houston, Texas. [6]Department of Systems Biology, MD Anderson Cancer Center, Houston, Texas. [7]Department of Pediatrics, Texas Children's Cancer and Hematology Centers, Baylor College of Medicine, Houston, Texas.

Corresponding Authors: Powel H. Brown, The University of Texas MD Anderson Cancer Center, Houston, TX 77030. Phone: 713-792-4509; Fax: 713-794-4679; E-mail: phbrown@mdanderson.org; and Ching C. Lau (pertaining to DNA analysis), Texas Children's Cancer and Hematology Centers, Baylor College of Medicine, Houston, TX 77030. E-mail: cclau@txch.org

AACR

Translational Relevance

This study describes the results of RNA and DNA genomic profiling of a large set of triple-negative breast cancers (TNBC). We identified four stable TNBC subgroups with distinct clinical outcomes defined by specific overexpressed or amplified genes. The four subgroups have been named the luminal androgen receptor (LAR), mesenchymal (MES), basal-like immunosuppressed (BLIS), and basal-like immune-activated (BLIA) groups. We also identified specific molecules that define each subgroup, serving as subgroup-specific biomarkers, as well as potential targets for the treatment of these aggressive breast cancers. Specific biomarkers and targets include the androgen receptor, MUC1, and several estrogen-regulated genes for the LAR subgroup; IGF1, prostaglandin F receptor for the MES subgroup; SOX transcription factors and the immunoregulatory molecule VTCN1 for the BLIS subgroup; and STAT transcription factors for the BLIA group. Thus, these studies form the basis to develop molecularly targeted therapy for TNBCs.

In the study by Lehmann and colleagues, when tumors with IHC-confirmed ER, PR, and Her2 protein expression were analyzed, only 5 of the 6 described subtypes were observed (see Supplementary Figs. S4 and S5 in the study by Lehmann and colleagues; ref. 12). Therefore, while previous genomic studies have advanced our understanding of TNBCs, stable subtypes as well as subtype-specific molecular targets still need to be identified.

In this study, we investigated 198 previously uncharacterized TNBCs using mRNA expression and DNA profiling and identified 4 stable TNBC subtypes: (i) luminal AR (LAR), (ii) mesenchymal (MES), (iii) basal-like immunosuppressed (BLIS), and (iv) basal-like immune-activated (BLIA). Using independent TNBC datasets, we show that BLIS and BLIA tumors have the worst and best prognoses, respectively (independently of other known prognostic factors), compared with the other subtypes. Our DNA studies demonstrate unique subtype-specific gene amplification, with CCND1, EGFR, FGFR2, and CDK1 amplified in the LAR, MES, BLIS, and BLIA subtypes, respectively. Collectively, our RNA and DNA genomic results identify stable, reproducible TNBC subtypes characterized by specific RNA and DNA markers and identify potential targets for the more effective treatment of TNBCs.

Materials and Methods

Patients and study recruitment

A total of 278 anonymized tissues collected from multiple U.S. and European sites were obtained from the Lester and Sue Smith Breast Cancer Tumor Bank at Baylor College of Medicine (BCM; Houston, TX), diagnosis-confirmed, and flash-frozen. BCM purchased these tumors [with clinical information, including age, menopausal status, histology, American Joint Committee on Cancer (AJCC) stage, tumor grade] from Asterand. No treatment or outcome data were available for these tumors. Tissues were managed by the BCM Breast Center's Human Tissue Acquisition and Pathology (HTAP) shared resource. Cellularity, histology, and IHC ER, PR, and Her2 status in discovery and validation

samples were assessed by Breast Center pathologists. Only tumors exhibiting >50% tumor cellularity were used. ER negativity is defined as Allred scale ≤ 2.

RNA/DNA extraction and array experiments

For extraction and quality control details, see Supplementary Material. Briefly, tumors were profiled using the Affymetrix U133 Plus 2.0 gene expression array and *affy* (16) package in R (17). Discovery and validation set SNP experiments were performed on Illumina 610K and 660K platforms, respectively. Common SNPs were analyzed after independent processing in Illumina Genome Studio v2011 Genotyping Module 1.9.4.

PAM50, TNBCType, and ERSig

TNBCs were assigned to previously described subtypes using the TNBCType tool (18). Intrinsic subtypes were established with the PAM50 Breast Cancer Intrinsic Classifier (19) and compared with 67 non-TNBC randomly sampled tumors representing 80% of the assigned sample (confirmed by Pearson correlation). This comparison was used to create a 32-gene centroid signature [derived from Williams and colleagues' estrogen receptor 1 (ESR1) downstream targets gene list (ref. 20)] and was accessed via the Molecular Signatures Database (MSigDB; ref. 21) to correlate TNBCs with ER activation ("ERSig").

Gene selection, non-negative matrix factorization clustering, differential expression, and centroid signatures

Genes were sorted by aggregate rank of median absolute deviations (MAD) across all samples and the MAD across each of the 2 most predominant clusters (approximating basal-like vs. the remaining intrinsic subtypes) for the discovery set using R package differential expression via distance summary (DEDS; ref. 22). The top 1,000 median-centered genes were used for clustering and split into 2,000 positive input features (23). The ideal rank basis and factorization algorithm was determined using the R package non-negative matrix factorization (NMF; ref. 24) before taking the 1,000-iteration consensus for a final clustering basis of 4.

Genes were sorted by DEDS using (i) Goeman's global test (GGT; ref. 25) applied to each set individually for all 18,209 genes, using a Benjamini–Hochberg false discovery rate (FDR) multitest correction and (ii) computed $\log_2$ (fold change) (FC) values. The top 20 unique genes by P value and $\log_2$(FC) became a classifier comprising 80 genes and representing the median quantiles of all 80 genes for each discover set cluster, with cases assigned by minimum average Euclidean distances of quantile gene expression data. Nonsignificant P values ($P > 0.05$ by 10,000 permutations) or deviations from any centroid >0.25 were left unclassified.

Preprocessing and assignment of expression data for publically accessible cases

Normalization and quality control procedures identical to the primary study sets [but using the Partek Genomics Suite program (ref. 26) to perform ANOVA-based batch correction across the 221 arrays before summarization of probe set data] were performed on 7 publically accessible studies in Gene Expression Omnibus (GEO) with TNBCs (by IHC) profiled on the Affymetrix U133 Plus 2.0 array ("external set"). Series GEO matrices and accompanying TNBC tumor clinical data from Sabatier and colleagues (ref. 27; ref. also included in external set) and Curtis and

colleagues (11) studies were assigned using gene-centric representation of array data.

Ingenuity Pathway Analysis

Significant genes (Benjamini–Hochberg correction: $P < 0.001$ from GGT) for each dataset group were uploaded independently into Ingenuity Systems' Interactive Pathway Analysis (IPA) software (www.ingenuity.com). A 0.05 significance threshold was used for pathway enrichment. Molecules, chemicals, or groups with regulatory function(s) were analyzed by IPA to produce final gene lists.

Copy number segmentation and analysis

Allele-specific piecewise constant fitting (ASPCF) analysis and allele-specific copy number (CN) analysis of tumors (ASCAT, default values; ref. 28) of 84 discovery and 58 validation set tumors yielded 62 and 46 samples, respectively, with assigned reliable DNA ploidy- and tumor percentage-corrected integer copy numbers. These segments were uploaded collectively and individually by assigned expression-based subtypes to Genomic Identification of Significant Targets in Cancer (GISTIC) 2.0 (ref. 29; default settings, with a 0.5 linear margin for gains and losses).

Survival analyses

Survival curves were constructed using the Kaplan–Meier product limit method and compared between subtypes with the log-rank test using publically available datasets for which disease-free survival (DFS) and disease-specific survival (DSS) results are available; however, no treatment information was available for these datasets. Cox proportional hazard regression model adjusted for available prognostic clinical covariates was performed to calculate subtype-specific HRs, 95% confidence intervals (CI), and disease-free survival and overall survival (OS). Survival analyses were performed using the R package *survival*.

Results

Patient population

A total of 198 TNBCs were assigned to discovery ($n = 84$) or validation ($n = 114$) sets based on chronological acquisition of tissue. Subjects were predominantly postmenopausal, Caucasian, and of mean and median age of 53 years (Table 1). About 95% of TNBCs were invasive ductal carcinomas, predominantly stages I–III (1% were metastatic breast cancers), and >75% of tumors were >2 cm at diagnosis.

Table 1. Clinical characteristics of the patients and tumor samples used in study

Characteristic	Both sets n (%)	Discovery set n (%)	Validation set n (%)	P
Number of tumors	198	84 (42)	114 (38)	
Age	192	84	108	0.02
<50 y	81 (42)	43 (51)	38 (35)	
≥50 y	111 (58)	41 (49)	70 (65)	
Missing	6	0	6	
Race	194	80	114	0.10
Caucasian	187 (96)	75 (94)	112 (98)	
Asian/Pacific Islander	7 (4)	5 (6)	2 (2)	
Missing	4	4	0	
Menopausal status	167	71	96	0.24
Premenopausal	62 (37)	31 (44)	31 (32)	
Menopausal	11 (7)	3 (4)	8 (8)	
Postmenopausal	94 (56)	37 (52)	57 (59)	
Missing	31	13	18	
Body mass index	166	65	101	0.65
Underweight (<18.5)	3 (2)	2 (3)	1 (1)	
Normal (18.5–24.9)	46 (28)	17 (26)	29 (29)	
Overweight (25–29.9)	61 (37)	26 (40)	35 (35)	
Obese (≥30)	56 (33)	20 (31)	36 (35)	
Missing	32	19	13	
Tumor size	195	83	112	0.01
<2 cm	35 (18)	10 (12)	25 (22)	
2–5 cm	139 (71)	60 (72)	79 (71)	
>5 cm	12 (6)	10 (12)	2 (2)	
Any size with direct extension	9 (5)	3 (4)	6 (5)	
Cannot be assessed	3	1	2	
Positive lymph nodes	150	66	84	0.14
0	74 (49)	29 (44)	45 (54)	
1–3	49 (33)	28 (42)	21 (25)	
4–9	17 (11)	6 (9)	11 (13)	
>10	10 (7)	3 (5)	7 (8)	
Unknown	48	18	30	
Metastases	146	64	82	0.86
No metastases	144 (99)	63 (98)	81 (99)	
Metastases found	2 (1)	1 (2)	1 (1)	
Unknown	52	20	32	
Histology	198	84	114	
Infiltrating ductal carcinoma	188 (95)	82 (98)	106 (93)	0.41
Infiltrative lobular carcinoma	1 (0.5)	0 (0)	1 (1)	
Adenocarcinoma/carcinoma, not otherwise specified	7 (3.5)	2 (2)	5 (4)	
Medullary carcinoma	2 (1)	0 (0)	2 (2)	

mRNA profiling of TNBCs reveals four stable molecular phenotypes

Using RNA gene expression profiling, we explored TNBC molecular phenotypes. NMF was performed on 1,000 discovery set genes selected to maximize separation across and within conventional intrinsic subtypes. These tumors were most stably divided into 4 clusters by cophenetic, dispersion, silhouette, and statistical significance of clustering (SigClust; ref. 30) metrics, in addition to visual inspection of the consensus heatmap (Fig. 1A and B and Supplementary Fig. S1). This quadrilateral division of data was also observed in the validation set tumors using the same input features (Fig. 1D and E and Supplementary Fig. S2). ER$^-$, PR$^-$, and Her2$^-$ were IHC-confirmed by our participating pathologist Dr. Contreras (Supplementary Fig. S3). Differentially expressed genes (Benjamini–Hochberg adjusted $P < 0.001$ from GGT) were significantly enriched only within corresponding discovery and validation set clusters (Fisher exact test: $P = 4.01E-30$, $3.47E-17$, $2.88E-46$, and $3.61E-10$, respectively; Supplementary Tables S1–S5), independently confirming the 4 molecular phenotypes observed. In addition, significant enrichment of discovery set IPA results in the validation set also support the 4 cluster separation (Supplementary Tables S6–S10).

Comparison of our NMF results to Perou's "PAM50" TNBC molecular classification (luminal A, luminal B, Her2$^+$, basal-like, and normal-like subtypes; ref. 9) shows clusters 3 and 4 to be entirely basal-like, containing 86% and 74% of all PAM50 basal-like tumors in the discovery and validation sets, respectively (Fig. 1C). Conversely, cluster 1 contains all luminal A, luminal B, and Her2$^+$ PAM50 tumors and cluster 2 contains basal-like and normal-like PAM50 tumors.

We then compared our NMF results with the Lehmann/Pietenpol "TNBC Type" molecular classification (basal-like 1, basal-like 2, immunomodulatory, LAR, mesenchymal, and mesenchymal stem–like subtypes; ref. 12), in which "claudin-low" tumors are split between the mesenchymal and mesenchymal stem–like subtypes. Our results show that cluster 1 contains all of Lehmann's LAR tumors and cluster 2 contains most of Lehmann's mesenchymal stem–like and some claudin-low mesenchymal tumors (Fig. 1F and Supplementary Figs. S4B and S5). Conversely, our TNBC clustering did not separate Lehmann's (12) "basal-like 1" and "basal-like 2" types even when using all 6 subtype signatures described in Lehmann and colleagues (12) in a semisupervised NMF (2,188 genes; Supplementary Fig. S4). Instead, Lehmann's basal-like 1 and basal-like 2 tumors are split between clusters 3 and 4 (Supplementary Fig. S4). Finally, Lehmann's remaining claudin-low mesenchymal tumors reside in cluster 3, whereas the immunomodulatory tumors are distributed across clusters 2 and 4, which express common signaling pathways (Supplementary Figs. S4 and S5).

Gene signatures define four prognostically distinct TNBC subtypes

Using the discovery and validation sets, we developed and confirmed an 80-gene signature for these clusters (Fig. 2A and Supplementary Tables S11–S16). This analysis was repeated using an independent set of 221 publically accessible TNBCs with IHC data (external set, Fig. 2B and Supplementary Tables S17 and S18) and other publically accessible datasets with available clinical data (Supplementary Tables S19 and S20).

Comparisons of group assignment against existing NMF clusters demonstrated strong reproducibility, with Rand indices of 0.94 ($P < 0.0001$) and 0.82 ($P < 0.0001$), respectively (Supplementary Tables S21 and S22).

Clinical outcome data were available for this publically available "external set" of TNBCs. However, treatment information for the "external set" data is not available. Analysis of DFS and DSS showed that subtype 3 has the worst prognosis of all 4 subtypes, whereas subtype 4 has a relatively good prognosis for DFS (log-rank test: $P = 0.042$ and 0.041, respectively) and DSS (log-rank test: $P = 0.039$ and 0.029, respectively; Fig. 2C and Supplementary Tables S23 and S24). The associations between subtypes 3 and 4 and DFS and DSS remained significant in multivariate models adjusted for available prognostic clinical covariates.

TNBC subtype–specific enrichment of molecular pathways

Differentially expressed genes from each subtype (Benjamini–Hochberg adjusted $P < 0.001$ from GGT) were analyzed for pathway enrichment. Results from the validation and external sets significantly overlapped the discovery set, with predicted regulator activation and inhibition patterns stable across the 3 datasets but distinct between subtypes (Fig. 3 and Supplementary Tables S25–S29).

Subtype 1 tumors exhibit AR, ER, prolactin, and ErbB4 signaling (Fig. 3) but ERα$^-$ IHC staining. Gene expression profiling demonstrates expression of ESR1 (the gene encoding ERα; Supplementary Fig. S6) and other estrogen-regulated genes (PGR, FOXA, XBP1, GATA3). Thus, these "ER-negative" tumors demonstrate molecular evidence of ER activation. This may be because 1% of these tumor cells express low levels of ER protein, defining them as "ER-negative" by IHC analysis. These observations suggest that subtype 1 tumors may respond to traditional anti-estrogen therapies as well as to anti-androgens, as previously suggested (12). To be consistent with previous studies (12), we termed subtype 1 the LAR subtype.

Subtype 2 is characterized by pathways known to be regulated in breast cancer, including cell cycle, mismatch repair, and DNA damage networks, and hereditary breast cancer signaling pathways (Fig. 3). In addition, genes normally exclusive to osteocytes (OGN) and adipocytes (ADIPOQ, PLIN1) and important growth factors (IGF1) are highly expressed in this subtype, previously described as "mesenchymal stem–like" or "claudin-low" (Supplementary Fig. S7). Therefore, we named Subtype 2 the MES subtype.

Subtype 3 is 1 of 2 basal-like clusters and exhibits downregulation of B cell, T cell, and natural killer cell immune-regulating pathways, and cytokine pathways (Fig. 3). This subtype has the worst DFS and DSS and low expression of molecules controlling antigen presentation, immune cell differentiation, and innate and adaptive immune cell communication. However, this cluster uniquely expresses multiple SOX family transcription factors. We termed subtype 3 the BLIS subtype.

Immunoregulation pathways are upregulated in subtype 4, the other basal-like cluster (Fig. 3). Contrary to BLIS, subtype 4 tumors display upregulation of genes controlling B cell, T cell, and natural killer cell functions. This subtype has the best prognosis, exhibits activation of STAT transcription factor–mediated pathways, and has high expression of STAT genes. To contrast BLIS tumors, we termed subtype 4 the BLIA subtype.
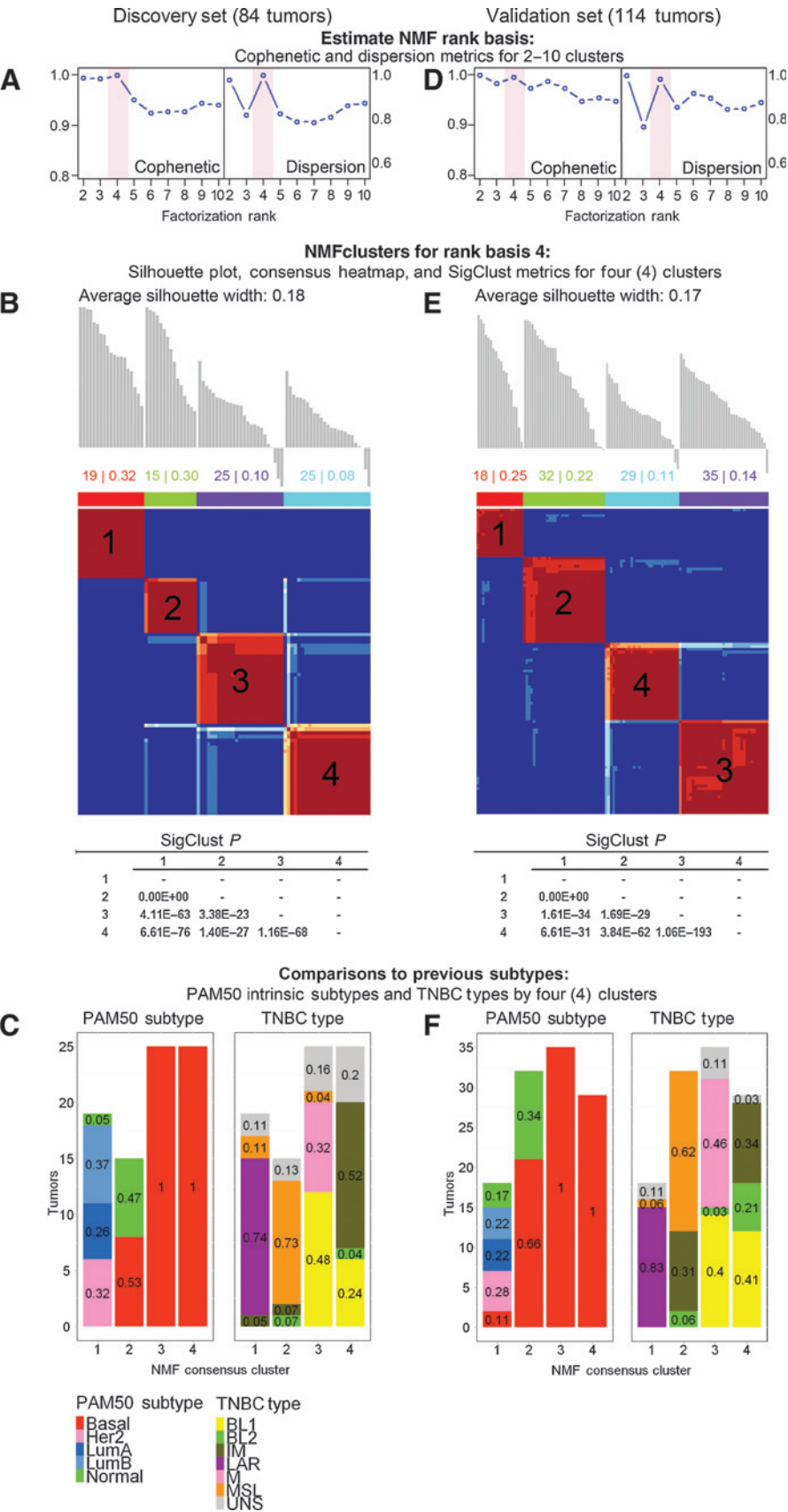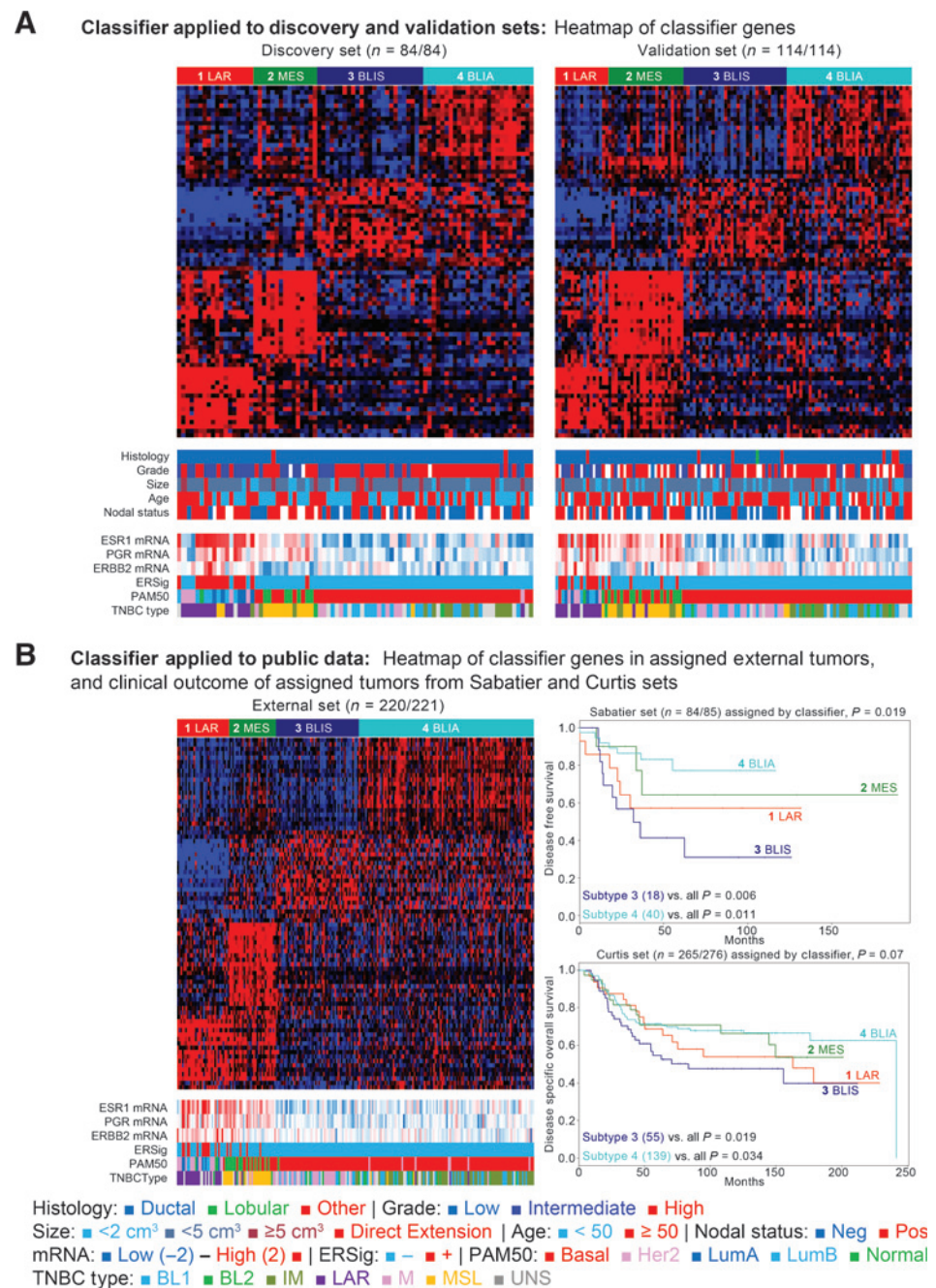
Figure 1.
Classification of TNBCs by mRNA profiling reveals 4 stable molecular phenotypes. Both 84 (discovery set) and 114 triple-negative breast tumors (validation set) demonstrate 4 stable clusters by NMF of mRNA expression across the top 1,000 genes [interquartile range (IQR) summarized] selected by DEDS aggregate rank of MADs (see Materials and Methods) of the discovery set. A and D, cophenetic and dispersion metrics for NMF across 2 to 10 clusters with 50 runs suggest 4 stable clusters. Full metrics are available for each set in Supplementary Figs. S1 and S2. B and E, silhouette analyses and consensus plots for rank basis 4 NMF clusters (1,000 runs, nsNMF factorization). Average silhouette widths worsened with increasing clusters beyond the 4 shown. SigClust was significant for all pairwise comparisons with this feature set. C and F, PAM50 intrinsic subtypes and TNBC Type distributions by 4 NMF clusters.

Figure 2.
Gene signature defines 4 subtypes of TNBC with prognostic differences. Discovery, validation, and external sets tumors with intermediate grade, high ESR1, PGR, and ERBB2 expression, activated ER downstream targets, and luminal A/B subtypes are enriched in subtype 1. A, the 4 assigned subtypes in both the discovery (84 of 84) and validation sets (114 of 114). B, gene signature applied successfully to 220 of 221 external set TNBCs. Clinical outcomes from independent sets classified by the discovery set–based signature. Subtype 4 has a better prognosis for both DFS and DSS.

DNA copy number analysis identifies TNBC subtype–specific focal changes

We next investigated TNBC subtype–defined copy number variation (CNV) by ploidy- and tumor percentage-correcting 62 discovery and 46 validation set TNBCs, before analyzing them together in GISTIC 2.0. Overall, genomes were very unstable and exhibited common TNBC chromosomal arm gains and deletions (Fig. 4A and Supplementary Figs. S7 and S8 and Supplementary Tables S30–S35). Focal variations present in all 4 TNBC subtypes include: (i) focal gains on 8q23.3 (CSMD3), 3q26.1 (BCHE), and 1q31.2 (FAM5C), which are

the greatest gains and characterize >84% of all tumors and (ii) focal losses on 9p21.3 (CDKN2A/B), 10q23.31 (PTEN), and 8p23.2 (CSMD1; Fig. 4B).

Subtype-specific variation is greatest between LAR and the remaining 3 subtypes (Fig. 4). LAR tumors have focal gains twice as frequently on 11q13.3 (CCND1, FGF family) and 14q21.3 (MDGA2), but one third as frequently on 12p13.2 (MAGOHB, KLR subfamilies) and 6p22.3 (E2F3, CDKAL1) compared with MES, BLIS, and BLIA tumors (Fig. 4). The LAR subtype also has more frequent deletions of 6q, lacks armwide deletions across 5q, 14q, and 15q, and has significantly fewer
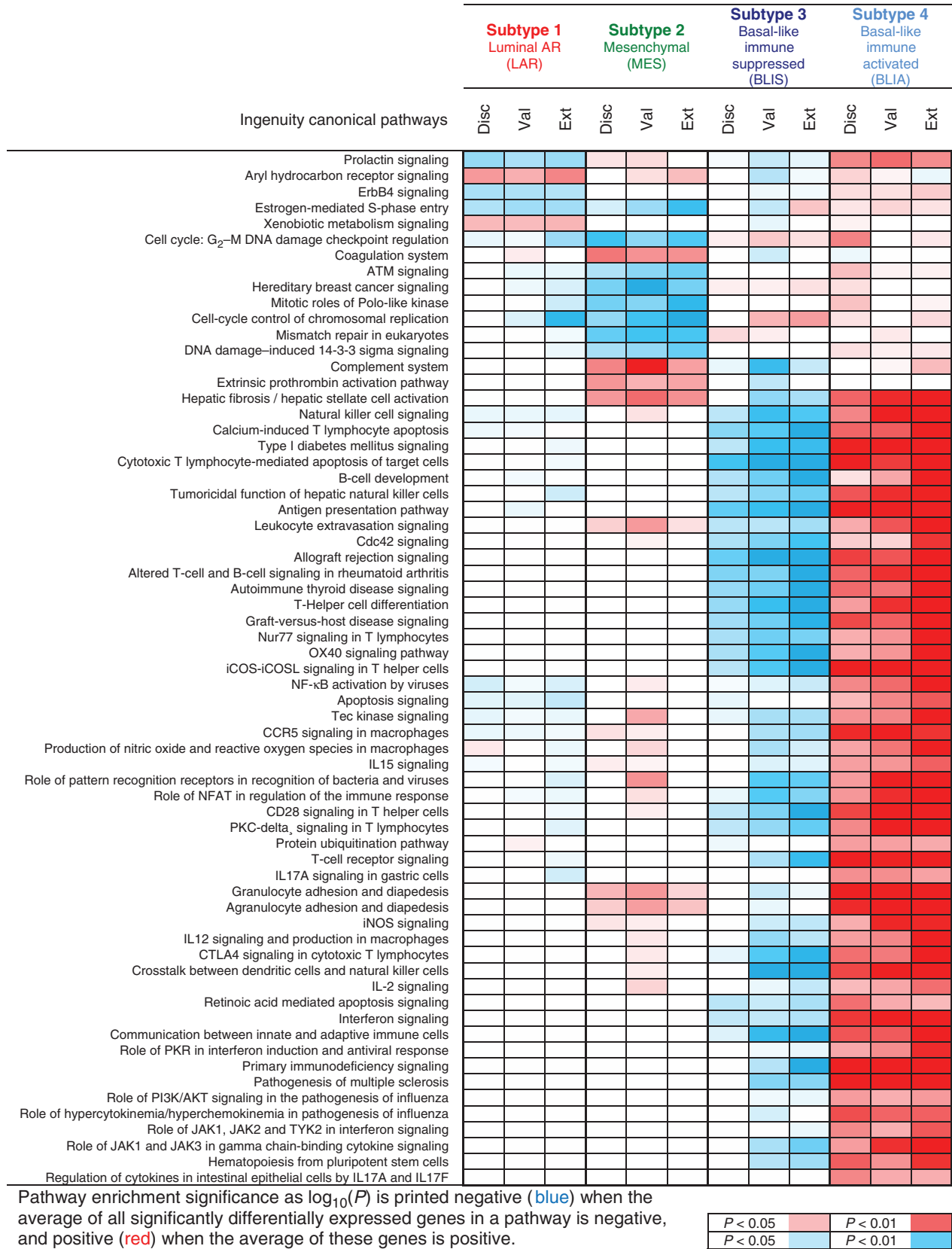
Pathway enrichment significance as log$_{10}$(P) is printed negative (blue) when the average of all significantly differentially expressed genes in a pathway is negative, and positive (red) when the average of these genes is positive.

P < 0.05		P < 0.01	
P < 0.05		P < 0.01	

Figure 3.
Molecular pathways enriched in the 4 identified subtypes of TNBCs. Significant pathways from the discovery set also found in validation and external sets are listed for the LAR, MES, BLIS, and BLIA subtypes.
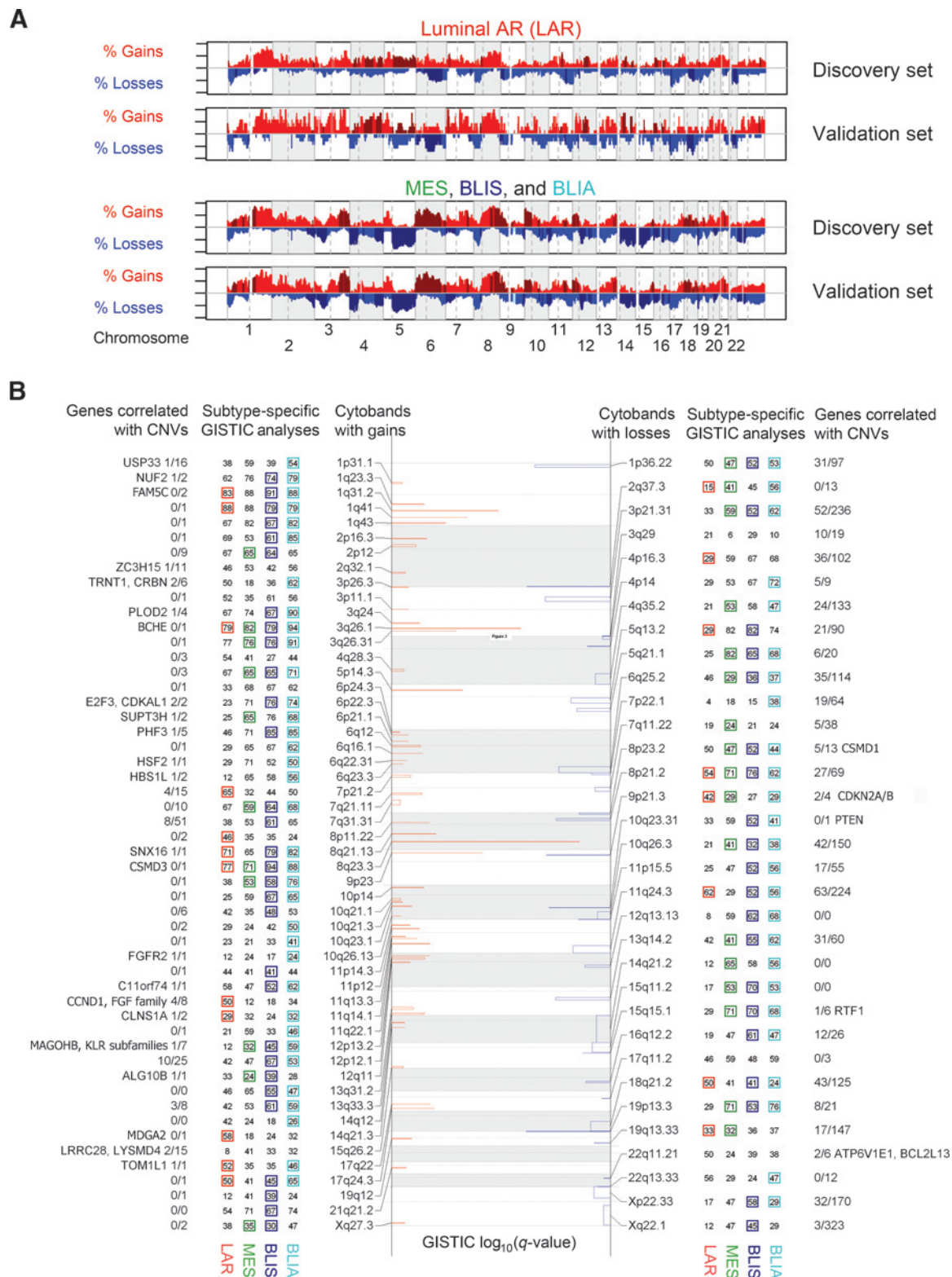
Figure 4.
DNA copy number analysis identifies focal changes in TNBC subtypes. DNA copy number changes observed in each subtype are listed. A, focal gains (red) and losses (blue) detected by GISTIC 2.0 are plotted by $\log_{10}(q\text{-value})$ and reported by cytoband. Adjacent numbers are percentages of subtype-specific cases ($n = 24, 17, 33, 34$, respectively) with this focal aberration. Presence of a colored square demonstrates this region was detected by subtype-specific GISTIC 2.0 analysis as well. All structural events for each subtype and set are available in Supplementary Material. B, broad copy number events distinguish the LAR subtype from all others. Gains (red) and losses (blue) are plotted along the genome, with darker colors representing a region enriched to the displayed subtype by Fisher exact test.

Table 2. Selected genes from pathway analysis with significant relative overexpression (>2-fold, Benjamini–Hochberg $P \leq 0.05$) in discovery and validation sets

TNBC subtype	Symbol	Description	Discovery FC	Druggable	CNV seen
1. LAR	DHRS2	Dehydrogenase/reductase (SDR family) member 2	68.6		
	PIP	Prolactin-induced protein	21.1		
	AGR2	Anterior gradient 2 homolog (*Xenopus laevis*)	17.1		Yes
	FOXA1	Forkhead box A1	17.1		Yes
	ESR1	Estrogen receptor 1	13.9	Yes	
	ERBB4	v-erb-a erythroblastic leukemia viral oncogene homolog 4 (avian)	11.3	Yes	
	CA12	Carbonic anhydrase XII	11.3	Yes	
	AR	Androgen receptor	9.8	Yes	
	TOX3	TOX high mobility group box family member 3	7.5		Yes
	KRT18	Keratin 18	4.3		Yes
	MUC1	Mucin 1, cell surface associated	4.3	Yes	
	PGR	Progesterone receptor	3.5	Yes	
	ERBB3	v-erb-b2 erythroblastic leukemia viral oncogene homolog 3 (avian)	3		Yes
	RET	Ret proto-oncogene	2.5	Yes	
	ITGB5	Integrin, beta 5	2.1	Yes	
2. MES	ADH1B	Alcohol dehydrogenase 1B (class I), beta polypeptide	42.2	Yes	
	ADIPOQ	Adiponectin, C1Q and collagen domain containing	32		
	OGN	Osteoglycin	16		
	FABP4	Fatty acid–binding protein 4, adipocyte	14.9		
	CD36	CD36 molecule (thrombospondin receptor)	14.9		
	NTRK2	Neurotrophic tyrosine kinase, receptor, type 2	6.1	Yes	
	EDNRB	Endothelin receptor type B	5.7	Yes	
	GHR	Growth hormone receptor	4.9	Yes	
	ADRA2A	Adrenoceptor alpha 2A	4.6	Yes	
	PLA2G2A	Phospholipase A2, group IIA (platelets, synovial fluid)	4.6	Yes	
	PPARG	Peroxisome proliferator–activated receptor gamma	4	Yes	
	ADRB2	Adrenoceptor beta 2, surface	3.5	Yes	
	PTGER3	Prostaglandin E receptor 3 (subtype EP3)	3.2	Yes	
	IL1R1	Interleukin 1 receptor, type I	3	Yes	
	TEK	TEK tyrosine kinase, endothelial	2.8	Yes	
3. BLIS	ELF5	E74-like factor 5 (ets domain transcription factor)	7		
	HORMAD1	HORMA domain containing 1	5.7		Yes
	SOX10	SRY (sex-determining region Y)-box 10	4.9		Yes
	SERPINB5	Serpin peptidase inhibitor, clade B (ovalbumin), member 5	4.6		
	FOXC1	Forkhead box C1	4.6		
	SOX8	SRY (sex-determining region Y)-box 8	4.3		
	TUBB2B	Tubulin, beta 2B class IIb	3.2	Yes	
	VTCN1	V-set domain containing T-cell activation inhibitor 1	3		
	SOX6	SRY (sex-determining region Y)-box 6	3		
	KIT	v-kit Hardy-Zuckerman 4 feline sarcoma viral oncogene homolog	2.5	Yes	
	FGFR2	Fibroblast growth factor receptor 2	2	Yes	Yes
4. BLIA	CXCL9	Chemokine (C-X-C motif) ligand 9	5.3		
	IDO1	Indoleamine 2,3-dioxygenase 1	4.9		
	CXCL11	Chemokine (C-X-C motif) ligand 11	4.9		
	RARRES1	Retinoic acid receptor responder (tazarotene induced) 1	4		Yes
	GBP5	Guanylate-binding protein 5	4.3		Yes
	CXCL10	Chemokine (C-X-C motif) ligand 10	4.3	Yes	
	CXCL13	Chemokine (C-X-C motif) ligand 13	4.3		
	LAMP3	Lysosomal-associated membrane protein 3	3.7		Yes
	STAT1	Signal transducer and activator of transcription 1, 91kDa	3		
	PSMB9	Proteasome (prosome, macropain) subunit, beta type, 9	2.8	Yes	
	CD2	CD2 molecule	2.5	Yes	
	CTLA4	Cytotoxic T lymphocyte–associated protein 4	2.5	Yes	
	TOP2A	Topoisomerase (DNA) II alpha 170kDa	2.1	Yes	Yes
	LCK	Lymphocyte-specific protein tyrosine kinase	2.1	Yes	

focal deletions on 5q13.2 (RAD17, ERBB2IP), 12q13.13 (CCNT1, ERBB3), 14q21.2 (FOXA1), and 15q11.2 (HERC2; ref. Fig. 4 and Supplementary Fig. S8). MES and BLIA tumors, which exhibit increased normal (diploid) immune cell infiltration, are characterized by lower aberrant cell fractions than LAR and BLIS tumors (Supplementary Fig. S9). Additional subtype-specific gene overexpression includes: (i) LAR: AR and MUC1; (ii) MES: IGF1, ADRB2, EDNRB, PTGER3/4, PTGFR, and PTGFRA; (iii) BLIS: VTCN1; and (iv) BLIA: CTLA4 (Table 2 and Supplementary Tables S36–S39).

Discussion

Using RNA and DNA profiling, we identified 4 stable, molecularly defined TNBC subtypes, LAR, MES, BLIS, and BLIA, characterized by distinct clinical prognoses, with **BLIS** tumors having the worst and BLIA tumors having the best outcome. DNA analysis demonstrated subtype-specific gene amplifications, suggesting the possibility of using *in situ* hybridization techniques to identify these TNBC subsets. Our results also demonstrate subtype-specific molecular expression, thereby enabling TNBC subtype classification based on molecules they do express as opposed to molecules they do not express.

Many highly expressed molecules in specific TNBC subtypes can be targeted using available drugs (Table 2 and Supplementary Tables S36–S39). Our results suggest that AR antagonists (12) and MUC1 vaccines may prove effective for the treatment of AR- and MUC1-overexpressing LAR tumors, whereas β-blockers, IGF inhibitors, or PDGFR inhibitors may be useful therapies for MES tumors. Conversely, immune-based strategies (e.g., PD1 or VTCN1 antibodies) may be useful treatments for BLIS tumors, whereas STAT inhibitors, cytokine, or cytokine receptor antibodies, or the recently FDA-approved CTLA4 inhibitor, ipilumimab (31), may be effective treatments for BLIA tumors. Thus, these studies have identified novel TNBC subtype-specific markers that distinguish prognostically distinct TNBC subtypes and may be targeted for the more effective treatment of TNBCs.

Lehmann's TNBC subtyping study identified 6 TNBC subtypes through the combined analysis of 14 RNA profiling datasets ("discovery dataset"; ref. 12). Assignment to these subtypes was confirmed using a second dataset composed of 7 other publically available datasets; however, all 6 subtypes were not detected when subtyping was limited to only those tumors with ER, PR, and Her2 IHC data. In addition, basal-like 1 and basal-like 2 tumors are not readily distinguishable by hierarchical clustering of public TNBC datasets using Lehmann's gene signatures (32), despite demonstration of molecular heterogeneity beyond the classic intrinsic subtypes. In Lehmann's study, TNBCs strongly segregated into stromal, immune, and basal gene modules, partially supporting our model. Additional studies have also demonstrated that an immune signature is an important clinical predictor for ER-negative tumors (27, 33, 34). The large set of ER-, PR-, and Her2-characterized tumors used in our study enabled us to further separate TNBCs into LAR, MES (including "claudin-low"), BLIS, and BLIA subtypes and define the clinical outcome of each subtype.

Previous genomic profiling studies have not demonstrated this degree of heterogeneity in basal-like breast tumors. Profiling of TCGA data across miRNA, DNA, and methylation data supported the intrinsic subtypes of breast cancer and grouped all basal-like tumors (8). In the Curtis dataset (11), unsupervised clustering by CNV-driven gene expression did not identify multiple basal-like subtypes, confirming that CNV alone does not distinguish these tumor subtypes. However, our integrated DNA and mRNA data demonstrate that gene amplification drives several subtype-specific genes. The CCND1 and FGFR2 genes are amplified in LAR tumors, whereas MAGOHB is more commonly amplified in MES, BLIS, and BLIA tumors. Conversely, CDK1 is amplified in all 4 TNBC subtypes (most highly in BLIA tumors) and thus represents a potential target. While broad and focal copy numbers differentiate LAR tumors from the remaining subtypes, they cannot dissociate BLIS and BLIA tumors.

All LARs and most mesenchymal stem–like tumors identified by the Pietenpol group (12) fall within our LAR and MES subtypes. However, our study splits the remaining proposed subtypes, including Lehmann's basal-like 1 and basal-like 2 tumors into distinct BLIS and BLIA subtypes based on immune signaling. Furthermore, stratification of our subtypes is based on a few broad biologic functions. LAR and MES tumors downregulate cell-cycle regulators and DNA repair genes, whereas MES and BLIA tumors upregulate immune signaling and immune-related death pathways (Supplementary Tables S36–S39). Conversely, our BLIS and BLIA subtypes show a relative lack of p53-dependent gene activation (p53 mutations characterize most TNBC tumors), and BLIA tumors highly express and activate STAT genes. Both our current study and the study by Lehmann and colleagues used RNA-based gene profiling to subtype TNBCs. Until more TNBC datasets are analyzed, it will not be clear which specific subgrouping will ultimately be most clinically useful. The study by Lehmann and colleagues subdivided TNBCs into 6 subtypes, whereas this article describes subgrouping of TNBCs into 4 distinct subtypes, 2 of which overlap with Lehmann and colleagues (LAR and MES), whereas our other 2 subtypes (BLIS and BLIA) contain mixtures of the other 4 Lehmann subgroups (see Fig. 1C and F). Our attempt at reproducing the 6 Lehmann and colleagues subgroups by clustering our data using their gene signatures was unsuccessful ($n = 198$, Supplementary Fig. S5). The exact subdivision of these TNBC subtypes, while important, is less important than the clinical prognosis defined by each subtype, and most importantly, the specific molecular targets identified within the subtypes. To this point, the identification of specific targets that modulate the immune system in the BLIA and BLIS subtypes is one of the most important and unique findings in this study.

In summary, using RNA profiling, we have defined 4 stable, clinically relevant subtypes of TNBC characterized by distinct molecular signatures. Our results uniquely define TNBCs by the molecules that are expressed in each subtype as opposed to molecules that are not expressed. Furthermore, these newly defined subtypes are biologically diverse, activate distinct molecular pathways, have unique DNA CNVs, and exhibit distinct clinical outcomes. By identifying molecules highly expressed in each TNBC subtype, this study provides the foundation for future TNBC subtype–specific molecularly targeted and/or immune-based strategies for more effective treatment of these aggressive tumors.

Acknowledgments

Grant Support

References

1. Brenton JD, Carey LA, Ahmed AA, Caldas C. Molecular classification and molecular forecasting of breast cancer: ready for clinical application? J Clin Oncol 2005;23:7350–60.
2. Morris GJ, Naidu S, Topham AK, Guiles F, Xu Y, McCue P, et al. Differences in breast carcinoma characteristics in newly diagnosed African-American and Caucasian patients: a single-institution compilation compared with the National Cancer Institute's Surveillance, Epidemiology, and End Results database. Cancer 2007;110:876–84.
3. Sørlie T, Perou CM, Tibshirani R, Aas T, Geisler S, Johnson H, et al. Gene expression patterns of breast carcinomas distinguish tumor subclasses with clinical implications. Proc Natl Acad Sci U S A 2001;98:10869–74.
4. Malorni L, Shetty PB, De Angelis C, Hilsenbeck S, Rimawi MF, Elledge R, et al. Clinical and biologic features of triple-negative breast cancers in a large cohort of patients with long-term follow-up. Breast Cancer Res Treat 2012;136:795–804.
5. Shastry M, Yardley DA. Updates in the treatment of basal/triple-negative breast cancer. Curr Opin Obstet Gynecol 2013;25:40–8.
6. Lee JM, Ledermann JA, Kohn EC. PARP Inhibitors for BRCA1/2 mutation-associated and BRCA-like malignancies. Ann Oncol 2014;25:32–40.
7. Perou CM, Sørlie T, Eisen MB, van de Rijn M, Jeffrey SS, Rees CA, et al. Molecular portraits of human breast tumours. Nature 2000;406:747–52.
8. Cancer Genome Atlas Network. Comprehensive molecular portraits of human breast tumours. Nature 2012;490:61–70.
9. Perou CM. Molecular stratification of triple-negative breast cancers. Oncologist 2011;16 Suppl 1:61–70.
10. Bertucci F, Finetti P, Birnbaum D. Basal breast cancer: a complex and deadly molecular subtype. Curr Mol Med 2012;12:96–110.
11. Curtis C, Shah SP, Chin SF, Turashvili G, Rueda OM, Dunning MJ, et al. The genomic and transcriptomic architecture of 2,000 breast tumours reveals novel subgroups. Nature 2012;486:346–52.
12. Lehmann BD, Bauer JA, Chen X, Sanders ME, Chakravarthy AB, Shyr Y, et al. Identification of human triple-negative breast cancer subtypes and preclinical models for selection of targeted therapies. J Clin Invest 2011;121:2750–67.
13. Herschkowitz JI, Simin K, Weigman VJ, Mikaelian I, Usary J, Hu Z, et al. Identification of conserved gene expression features between murine mammary carcinoma models and human breast tumors. Genome Biol 2007;8:R76.
14. Prat A, Parker JS, Karginova O, Fan C, Livasy C, Herschkowitz JI, et al. Phenotypic and molecular characterization of the claudin-low intrinsic subtype of breast cancer. Breast Cancer Res 2010;12:R68.
15. Prat A, Perou CM. Deconstructing the molecular portraits of breast cancer. Mol Oncol 2011;5:5–23.
16. Gautier L, Cope L, Bolstad BM, Irizarry RA. affy–analysis of Affymetrix GeneChip data at the probe level. Bioinformatics 2004;20:307–15.
17. R Core Team: R: A Language and Environment for Statistical Computing, (Ver.2.12.2). R Foundation for Statistical Computing; 2012.
18. Chen X, Li J, Gray WH, Lehmann BD, Bauer JA, Shyr Y, et al. TNBCtype: a subtyping tool for triple-negative breast cancer. Cancer Inform 2012;11: 147–56.
19. Parker JS, Mullins M, Cheang MC, Leung S, Voduc D, Vickery T, et al. Supervised risk predictor of breast cancer based on intrinsic subtypes. J Clin Oncol 2009;27:1160–7, 2009.
20. Williams C, Edvardsson K, Lewandowski SA, Ström A, Gustafsson JA. A genome-wide study of the repressive effects of estrogen receptor beta on estrogen receptor alpha signaling in breast cancer cells. Oncogene 2008;27:1019–32.
21. Liberzon A, Subramanian A, Pinchback R, Thorvaldsdóttir H, Tamayo P, Mesirov JP. Molecular signatures database (MSigDB) 3.0. Bioinformatics 2011;27:1739–40.
22. Yang YH, Xiao Y, Segal MR. Identifying differentially expressed genes from microarray experiments via statistic synthesis. Bioinformatics 2005;21: 1084–93.
23. Kim PM, Tidor B. Subsystem identification through dimensionality reduction of large-scale gene expression data. Genome Res 2003;13:1706–18.
24. Gaujoux R, Seoighe C. A flexible R package for nonnegative matrix factorization. BMC Bioinformatics 2010;11:367.
25. Goeman JJ, van de Geer SA, de Kort F, van Houwelingen HC. A global test for groups of genes: testing association with a clinical outcome. Bioinformatics 2004;20:93–9.
26. Partek Inc.: Partek® Discovery SuiteTM, (Ver. 6.3). St. Louis, MO: Partek Inc.; 2008.
27. Sabatier R, Finetti P, Mamessier E, Raynaud S, Cervera N, Lambaudie E, et al. Kinome expression profiling and prognosis of basal breast cancers. Mol Cancer 2011;10:86.
28. Van Loo P, Nordgard SH, Lingjærde OC, Russnes HG, Rye IH, Sun W, et al. Allele-specific copy number analysis of tumors. Proc Natl Acad Sci U S A 2010;107:16910–5.
29. Mermel CH, Schumacher SE, Hill B, Meyerson ML, Beroukhim R, Getz G. GISTIC2.0 facilitates sensitive and confident localization of the targets of focal somatic copy-number alteration in human cancers. Genome Biol 2011;12:R41.
30. Liu Y, Hayes DN, Nobel A, Marron J. Statistical significance of clustering for high-dimension, low–sample size data. J Am Stat Assoc 2008;103: 1281–93.
31. Stagg J, Allard B. Immunotherapeutic approaches in triple-negative breast cancer: latest research and clinical prospects. Ther Adv Med Oncol 2013;5:169–81.
32. Prat A, Adamo B, Cheang MC, Anders CK, Carey LA, Perou CM. Molecular characterization of basal-like and non-basal-like triple-negative breast cancer. Oncologist 2013;18:123–33.
33. Teschendorff AE, Miremadi A, Pinder SE, Ellis IO, Caldas C. An immune response gene expression module identifies a good prognosis subtype in estrogen receptor negative breast cancer. Genome Biol 2007;8:R157.
34. Rody A, Karn T, Liedtke C, Pusztai L, Ruckhaeberle E, Hanker L, et al. A clinically relevant gene signature in triple negative and basal-like breast cancer. Breast Cancer Res 2011;13:R97.