Genome-Wide Functional Synergy between Amplified and Mutated Genes in Human Breast Cancer

Yuri Nikolsky,1 Evgeny Sviridov,2 Jun Yao,3,5 Damir Dosymbekov,2 Vadim Ustyansky,1 Valery Kaznacheev,7 Zoltan Dezso,5 Laura Mulvey,3,6 Laura E. Macconaill,4,5 Wendy Winckler,4,7 Tatiana Serebryiskaya,2 Tatiana Nikolskaya,1,2 and Kornelia Polyak3,4,7

Abstract

A single cancer cell contains large numbers of genetic alterations that in combination create the malignant phenotype. However, whether amplified and mutated genes form functional and physical interaction networks that could explain the selection for cells with combined alterations is unknown. To investigate this issue, we characterized copy number alterations in 191 breast tumors using dense single nucleotide polymorphism arrays and identified 1,747 genes with copy number gain organized into 30 amplicons. Amplicons were distributed unequally throughout the genome. Each amplicon had distinct enrichment pattern in pathways, networks, and molecular functions, but genes within individual amplicons did not form coherent functional units. Genes in amplicons included all major tumorigenic pathways and were highly enriched in breast cancer–causative genes. In contrast, 1,188 genes with somatic mutations in breast cancer were distributed randomly over the genome, did not represent a functionally cohesive gene set, and were relatively less enriched in breast cancer marker genes. Mutated and gained genes did not show statistically significant overlap but were highly synergistic in populating key tumorigenic pathways including transforming growth factor β, WNT, fibroblast growth factor, and PIP3 signaling. In general, mutated genes were more frequently upstream of gained genes in transcription regulation signaling than vice versa, suggesting that mutated genes are mainly regulators, whereas gained genes are mostly regulated. ESR1 was the major transcription factor regulating amplified but not mutated genes. Our results support the hypothesis that multiple genetic events, including copy number gains and somatic mutations, are necessary for establishing the malignant cell phenotype. [Cancer Res 2008;68(22):9532–40]

Introduction

Gene copy number gain and somatic mutations are two major molecular mechanisms underlying the activation of oncopgenes during tumor development. Gene amplifications occur recurrently at specific chromosomal locations and most tumors display multiple regions of copy number gains indicating the combined activation of several oncogenes (1). In human breast cancer, the most frequent and well-characterized amplicons are located on chromosomes 1q, 8p12, 8q24, 11q13, 12p13, 16p13, 17q11.2-q21, 17q22, 2q3, and 20q13 (2–7). Genes targeted by these amplification events have been identified in some cases, but emerging views suggest a selection for a combination of coamplified genes rather than single amplicon targets (8). In fact, in breast cancer, the majority of amplified or gained genes show increased levels of gene expression (9). In addition, possible cooperation within amplicons, functional interactions among genes in distinct amplicons may exist and could explain the worse clinical outcome of tumors with multiple chromosomal regions of copy number gain (10).

Cancer genomes were screened for somatic mutations and genetically altered pathways in several large-scale studies using unbiased genome-wide technologies (3, 10, 11). Recently, 1,142 genes were identified with somatic alterations based on exon resequencing of over 18,000 human genes in 11 breast cancer samples (12). One hundred forty of these mutated genes were defined as “CAN” (candidate cancer) driver genes. A study focused on protein kinases identified 79 genes with somatic mutations in breast cancer (9). Both studies concluded that the cancer genome is extremely complex, a large number of genes are mutated in each tumor, and very few genes are mutated in a high fraction of tumors. This high degree of complexity raised new challenges with data interpretation and for the identification of key genetically altered pathways amenable to therapeutic targeting.

To help the functional interpretation of complex data sets, several systems analysis methods have been developed and applied for breast cancer including gene set ontology enrichment, interactome, network and pathway analysis, and network modeling (11, 13–15). We have previously used the data analysis MetaCore platform for elucidating pathways specifically activated in breast cancer stem cells (16), for the identification of pathways differentiating normal mammary luminal epithelial and myoepithelial cells (17), and for the comparative analysis of DNA methylation and gene expression profiles of four distinct normal mammary epithelial cell types (18).

Here, we describe a large-scale study to characterize copy number alterations in primary breast carcinomas and to define functional interactions among genetically altered genes in breast cancer. Using high-density single nucleotide polymorphism (SNP) arrays, we analyzed 191 breast cancer samples and identified 1,747 genes with copy number gain (called the breast cancer ampicome). We compared these amplified genes to genes mutated in breast cancer (called the breast cancer mutome) and identified interactions and networks reflecting potential cooperation among genes genetically altered in breast tumors.
Materials and Methods

Detailed description of the procedures is in Supplementary Materials and Methods.

Breast tumor samples. Breast cancer cell lines were obtained from American Type Culture Collection or were generous gifts of Drs. Steve Ethier (University of Michigan, Ann Arbor, MI), Fred Miller (Karmanos Cancer Institute, Detroit, MI), and Arthur Pardee (Dana-Farber Cancer Institute, Boston, MA). All cell lines were grown in medium recommended by the provider. Fresh or frozen tumor specimens were obtained from the Brigham and Women’s Hospital, Beth-Israel Deaconess Medical Center, Massachusetts General Hospital, University of Zagreb (Croatia), Duke University, and National Disease Research Interchange. All human tissue was collected using protocols approved by the Institutional Review Boards. Only immunomagnetic bead purified or microdissected tumor samples were used for the analysis. Tissue was snap frozen on dry ice and stored at −80°C until use or were processed for purification as described previously (16).

SNP array analysis. SNP array analysis was performed by the Dana-Farber Microarray Core and by the Broad Institute using Affymetrix 250K Sty1 SNP arrays as described (16). The raw .CEL files were normalized, and the copy number of each SNP were generated using the GenePattern software.8 The procedure was optimized to exclude normal references that were contaminated with tumor tissues. Normal references (matched or virtual) were used as controls to exclude copy number polymorphisms. Raw SNP copy numbers were then smoothed by a segmentation algorithm using the DNAcopy package.9 To facilitate detailed analysis, copy number of each gene was estimated by averaging copy numbers from all SNPs found within the gene structure and flanking 100-kbp regions.

Results and Discussion

Definition of the breast cancer amplicome. To define the breast cancer amplicome, we analyzed copy number alterations in 191 human breast cancer samples (154 primary tumors and 37 breast cancer cell lines) using 250K Affymetrix SNP arrays. To delineate amplicons, we used a modification of a method developed for the extraction of minimal common regions from array comparative genomic hybridization data (4, 19). Briefly, genes with predicted copy number above two were categorized as gained. X chromosome was not included in the analysis due to difficulties with accurately predicting copy numbers for this chromosome based on our SNP data. One hundred seventy-three of 191 samples contained at least one gained chromosomal region or gene (Fig. 1). Previous studies have shown that the number and type of amplicons correlates with breast tumor subtypes (10). In our data set, we observed an enrichment of specific amplicons in luminal subtypes (17q12 amplicon (1)). However, in our data set, genes within individual amplicons generally did seem to form concise functional entities such as pathways, cellular processes, and networks. We conducted enrichment analysis of genes in all 30 amplicons in 5 functional ontologies (GO processes, GO molecular functions, canonical pathway maps, cellular process networks, and disease biomarkers). Genes within individual amplicons were enriched in cellular processes and pathways relevant to tumorigenesis including cell cycle, cell adhesion, DNA repair, development, innate response, and cytoskeleton (Supplementary Excel File S3). However, in individual amplicons, cellular pathways were sparsely populated with amplified genes (only one or two genes per pathway) and the p values of hypergeometric distribution for individual amplicons were high in comparison with the complete amplicome, suggesting that none of the processes and pathways were encoded by a single amplicon. All large amplicons encoded multiple pathways and processes without significant dominance of individual functional entities. Thus, amplicons did not show distinct specialization for pathways and processes. However, individual amplicons were highly synergistic in encoding for genes with functional roles in tumorigenesis, as evident from the enrichment analysis of the entire amplicome. This synergy was clearly visible in some tumorigenic pathway maps assembled from genes located in different amplicons, including the IGF-RI pathway map composed of the genes derived from 11 amplicons (Fig. 2).

8 http://www.broad.mit.edu/cancer/software/genepattern/index.html
9 http://www.bioconductor.org
Figure 1. Definition of amplicons and gained genes data distribution. Heatmap depicting the relatedness of tumor samples based on the clustering of 2,791 amplified (>5 copies) genes. Red signal, genes with copy number gains. Depicted chromosome length is proportionate to the number of genes contained in each of the indicated amplicons. Colored rectangles, tumor subtype (red, basal; pink, HER2+; blue, luminal), ER, and HER2 status (red, positive; blue, negative; white, unknown). Breast tumors were classified into luminal, basal-like (triple negative), and HER2 subtypes based on the expression of ER and HER2.
Genes within individual amplicons did not form concise networks. One measure of functional connectivity among a set of genes and proteins is their ability to self-assemble into biological networks with physical protein interactions as links between them (22). We applied direct interactions network building algorithm to each of the 30 amplicons. Genes in individual amplicons did not assemble into significant direct interaction networks. Only a few genes were directly linked with other genes in two of the largest (17q21-q25 and 20q12-q13) amplicons (Supplementary Fig. S2A and B).

The density of interactions within individual amplicons was lower than that in the entire human interactome. Network topology analysis evaluates the connectivity among genes within a set measured as the average number of interactions per network object (degree) divided into incoming interactions (degree IN) and outgoing interactions (degree OUT; ref. 23). The global human interactome of 16,000 proteins with >200,000 interactions (MetaCore database) was used as a control for evaluating the relative connectivity within amplicons. Only 6 of 30 amplicons (7p15, 8q12-q22, 8q23-q24, 17q11-q21, 17q21-q25, and 20q12-q13) displayed any sensible degree of intraamplicon connectivity. Moreover, connectivity within these amplicons was significantly lower than that in the global human interactome (Supplementary Excel File S4). The largest hubs within individual amplicons were GRB2, c-MYC, 14-3-3β/α, PKC-α, MAP3K3, NDPKB, and CAAT/enhancer binding protein (C/EBP)-β (Supplementary Fig. S3). Conversely, some amplicons had significantly higher number of interactions with genes outside of the amplicon than the global interactome had (Supplementary Excel File S4). Interestingly, half of the amplicons were enriched with degrees IN, but only four (8q23-24, 15q26, 17p11, and 20q12-13) were enriched with degrees OUT. This indicates that amplicons in general are "regulated" rather than "regulating" data sets, and they were enriched for "effector" rather than signaling pathways and processes.

Network analysis of the breast cancer amplicome. To further evaluate functional interactions among genes within amplicons, we analyzed signaling and metabolic networks among 1,360 amplified genes encoding known function using Analyze Networks (AN) algorithms. This series of algorithms reveal and visualize the most relevant multistep subnetworks of a selected size. The AN Transcription factors (ANTF) algorithm (24) connects all genes via directed shortest path and prioritizes subnetworks based on the following variables: (a) relative enrichment of the network with the seed genes (in this case, gained genes); (b) relative enrichment of the network with canonical pathways; and (c) relative completeness of noncanonical pathways from ligands to TF targets.

One of the top scored ANTF networks of the breast cancer amplicome included the JUN oncogene as central hub with 32

![Figure 2. Development of IGF-RI signaling canonical pathway map. Gained genes on the pathway map are marked with red thermometers. Names of corresponding amplicons are displayed in yellow boxes.](image-url)
targets (Supplementary Fig. S4). Although JUN itself was not amplified, it is regulated by the amplified MEF2 TF via calmodulin signaling and its targets include amplified ligands (IL24, IFN-γ, galanin, and endostatin) and receptors (HER2, GRP30, and SSTR2). We found 30 additional highly scored ANTf networks, many of them including tumor suppressors (TP53 and SMAD3) and oncogenes (C/EBPB and STAT3) encoding for TFs as central hubs highlighting a key role for transcriptional regulation in tumorigenesis (Supplementary Table S2).

Next, we applied the AN Receptors (ANR) algorithm that adds selectivity of the entire pathway from the most likely ligand to one membrane receptor as additional prioritization criteria (24). The highest scored ANR subnetwork was centered around the ESR1 nuclear hormone receptor (Supplementary Fig. S5; Fig. 3). ESR1 was not amplified in our data set, although it was previously reported to have copy number gain in a subset of breast tumors (25), but it had significantly more direct targets among amplified genes than expected by chance highlighting its importance in breast cancer. The algorithm selected the amplified IGF1R as the most likely receptor connected with ESR1 via SP1 signaling path between the two. Other important nodes of the ESR1 network included BRCA1 (downstream of ATM and upstream of ESR1) and amplified MYC (downstream of BRCA1 and CDK4; Fig. 3).

Cross-connections and trans-regulation among amplicons. In an arbitrary data set, each protein has interactions with other proteins within the same set (intraconnections) and with proteins that do not belong to the set (interconnections). Both types of connectivity can be evaluated quantitatively in terms of overconnectivity and underconnectivity compared with the expected number of links based on the size of the data set. We developed a new statistical procedure (Supplementary Fig. S6A and B) and applied it to define overconnectivity and underconnectivity for all 30 amplicons in 140 pair-wise combinations. We found that the number of direct interactions in pairs varied substantially, demonstrating a complex regulatory pattern among amplicons (Supplementary Excel File S5). Twenty-six and 114 amplicon pairs were overconnected and underconnected, respectively, relative to the expected number of links. The largest number of interactions was observed between the overconnected pair 8q12-q22 and 17q11-q21 (15 links) and the underconnected pair 8q23-q24 and 17q21-q25 (39 links). Amplicon 7p15 (HOXA gene cluster) had intraconnections with the lowest P value, indicating cross-regulation among HOX genes in this amplicon. In general, interconnectivity among amplicons was higher than intraconnectivity within individual amplicons. Three amplicons (17q21-q25, 20q12-q13, and 8q23-q24) had the highest number of interconnections with other amplicons and nearly half of all interconnections in the amplicome involved these three amplicons.
Among 17 interaction mechanisms tested, interactions between amplicons were enriched in transcription regulation (TR) links, defined as physical binding of TFs to the transcription regulatory regions of their target genes (Supplementary Excel File S5). Out of 65 amplicon pairs featuring TR interactions, 49 were over-connected and only 16 under-connected. Two amplicons were particularly enriched with outgoing TR interactions: 8q23-q24 and 20q12-q13 that transcriptionally regulated 18 (102 TR links) and 14 (52 TR links) amplicons, respectively. The most prominent TF in amplicon 8q23-q24 was c-Myc as it had the highest number of outgoing TR links. In contrast, in amplicon 20q12-q13, several TFs in amplicon 8q23-q24 were equally prominent. These data were consistent with the above-described functional analysis demonstrating that amplicon 20q12-q13 was enriched in TFs (Supplementary Excel File S2), whereas amplicon 8q23-q24 has the highest number of outgoing interactions (Supplementary Excel File S4).

Comparative analysis of breast cancer amplicome and mutome. Next, we investigated functional interactions among genes genetically altered due to amplification (amplicome) or somatic mutations (mutome) in breast cancer. The mutome gene set was defined as a nonredundant union of 1,188 somatically mutated genes from two large-scale genome sequencing studies (3, 12). Nine hundred sixty-five of these 1,188 genes had interactions in MetaCore database (Supplementary Excel File S6). We also analyzed the set of 140 CAN (candidate cancer) genes that were likely to play a casual role in tumorigenesis (11).

Contrary to amplified genes, mutated genes were randomly distributed in the genome (Supplementary Excel File S6). Only 94 genes were part of both the mutome and the amplicome, and only 12 from the CAN genes were amplified, which is a statistically significantly smaller set than expected (P = 0.048 and P = 0.456, respectively; Supplementary Excel File S7; Table 1). Enrichment analysis of 94 mutated and gained genes did not show significantly enriched pathways, processes, and cellular processes (data not shown). One possible reason for the small overlap between two data sets could be that the mutational analysis was conducted on a different and much smaller set of tumors than what we used for copy number analysis. Alternatively, some functional gene categories are preferentially affected by copy number gain versus mutation.

On the other hand, the mutome was closely interlinked with individual amplicons via transcriptional regulation interactions (Supplementary Table S3; Supplementary Fig. S7). The most prominent and statistically significant (P < 0.007) relationship was between mutome and amplicon 15q26 that included eight direct interactions between TFs in the mutome and their targets present in this amplicon (Supplementary Excel File S8). Importantly, TR interactions between mutome and amplicome were largely unidirectional. The absolute number of outgoing interactions was higher in the direction from mutome to amplicome than vice versa (Supplementary Excel File S8). This finding suggested that the mutome was enriched for TFs, whereas the amplicome was enriched for the targets of these genes. To further investigate this, we applied Transcription Target Modeling (TTM) algorithm to the mutome data set and mapped amplified genes on it. TFM algorithm reconstructs canonical pathways from membrane receptors to TFs most densely populated with the seed genes. The top-scored TTM network connected cytokine IL12 with the TF ERM and IL18 with NF-kB in 6 and 7 steps, correspondingly (Supplementary Fig. S7A). Both pathways were highly enriched in mutated genes. The IL12 subnetwork resulted in amplification of IFN γ, the direct target of ERM. The second scored network enriched with DNA exchange genes, linked the mutated ATM, ATR, BRCA1, and BRCA2 genes with the amplified RAD9, BRIP1, and EMSY (Supplementary Fig. S8B). Interestingly, mutated genes were always upstream of the amplified ones on these networks.

In total, 23 mutated TFs were overconnected with amplicome genes with the highest overconnection scores observed for TFs ATF2 and NFYC (Supplementary Table S4). Seven of CAN TFs had direct targets among amplified genes: TP53 (52 targets), HNF1A (25 targets), GLI1 (4 targets), BRCA1 (3 targets), PLU1 (2 targets), RELX2, and TAF1 (1 target each; Supplementary Excel File S9). In addition to TFs, several other CAN genes of different function had a large number of downstream interactions with amplicome genes including NOTCH1 (11 interactions), HDAC4 (5 interactions), and FLNA (4 interactions). Among amplified genes, the gene encoding for cytoskeletal actin had the highest number (5)

<table>
<thead>
<tr>
<th>Gene symbol</th>
<th>Entrez ID</th>
<th>Chr localization</th>
<th>Protein functional class</th>
<th>Protein name</th>
</tr>
</thead>
<tbody>
<tr>
<td>ARFGEF2</td>
<td>10564</td>
<td>20q13.13</td>
<td>Regulator</td>
<td>Brefeldin A-inhibited guanine nucleotide-exchange protein 2</td>
</tr>
<tr>
<td>CENTG1</td>
<td>116886</td>
<td>12q14.1</td>
<td>Regulator</td>
<td>Centaurin-γ 1</td>
</tr>
<tr>
<td>HOXA3</td>
<td>3200</td>
<td>7p15-p14</td>
<td>Transcription factor</td>
<td>Homeobox protein Hox-A3</td>
</tr>
<tr>
<td>NCOA6</td>
<td>23054</td>
<td>20q11</td>
<td>Binding protein</td>
<td>Nuclear receptor coactivator 6</td>
</tr>
<tr>
<td>PP1M1E</td>
<td>22843</td>
<td>1q22</td>
<td>Phosphatase</td>
<td>Protein phosphatase 1E</td>
</tr>
<tr>
<td>RBMS2</td>
<td>9699</td>
<td>8q22.3</td>
<td>Binding protein</td>
<td>Regulating synaptic membrane exocytosis protein 2</td>
</tr>
<tr>
<td>SLC6A3</td>
<td>6531</td>
<td>5p15.3</td>
<td>Transporter</td>
<td>Sodium-dependent dopamine transporter</td>
</tr>
<tr>
<td>SULF2</td>
<td>55959</td>
<td>20q12-q13.2</td>
<td>Metabolic enzyme</td>
<td>Extracellular sulfatase Sulf-2 precursor</td>
</tr>
<tr>
<td>TG</td>
<td>7038</td>
<td>8q24</td>
<td>Binding protein</td>
<td>Thyroglobulin precursor</td>
</tr>
<tr>
<td>THBS3</td>
<td>7059</td>
<td>1q21</td>
<td>Binding protein</td>
<td>Thrombospondin-3 precursor</td>
</tr>
<tr>
<td>ZFPM6</td>
<td>55734</td>
<td>20q13.2</td>
<td>Binding protein</td>
<td>Zinc finger protein 64, isoforms 3 and 4</td>
</tr>
<tr>
<td>ZNF569</td>
<td>148266</td>
<td>19q13.12</td>
<td>Binding protein</td>
<td>Zinc finger protein 64, isoforms 1 and 2</td>
</tr>
</tbody>
</table>

Table 1. Mutated and amplified genes in breast cancer. An intersection between CAN genes and amplified gene sets
of upstream links with mutated genes (DBN1, FLNB, GSN, TLN1, and TARA); MYC and IGF1R each had four links (BRCA1, HDAC4, NOTCH1, and TP53; and BRCA1, HNF1A, NOTCH1, and TP53), respectively; and Cyclin D1 had three links (BRCA1, GLI1, and HDAC4). Based on these data, BRCA1 and HDAC4 were the two mutated genes with the highest number of interactions with amplified genes.

Connectivity of mutome and amplicome proteins within sets and with the global proteome. Overall, both the mutome and the amplicome showed somewhat higher connectivity than the global human interactome. The degree was 11.75 and 9.579 for the mutome and the amplicome, respectively, compared with 9.133 for the entire human interactome (Supplementary Excel File S10). This may indicate that the mutome is slightly enriched.

Figure 4. Functional synergy between amplicome and mutome. The canonical pathway map “Cell adhesion–Chemokines and adhesion” was densely populated with amplified (red rectangles) and mutated (blue boxes) genes. Note that the genes were either amplified or mutated.
in hubs (highly connected proteins). Interestingly, the difference between the connectivity of the mutome and the amplicome was higher for outgoing links than for incoming links. On average, mutome and amplicome genes had 11.44 and 7.978 OUT interactions, respectively, compared with only 8.56 in the global interactome. This phenomenon was particularly striking in the case of the CAN genes that featured 22.1 OUT interactions (2.6-fold more than expected). These data were consistent with the observations above indicating that the mutome is generally a “regulating” data set. Intraconnectivity within data sets was comparable between mutome and amplicome (2.942 for mutome and 2.548 for amplicome; Supplementary Excel File S10).

Subsequently, we evaluated the relative distribution of genes encoding for proteins with different molecular function within the mutome and the amplicome (Supplementary Excel File S11). The mutome had fewer secreted ligands than expected but was enriched in kinases (2-fold over expected) and receptors. The amplicome contained slightly fewer receptors than expected. The fraction of TFs was similar in both data sets.

Next, we evaluated intraconnectivity and interconnectivity of individual proteins in the mutome and the amplicome, specifically searching for statistically significantly overconnected and underconnected proteins. We analyzed separately interconnections between each sets (mutome and amplicome), and the global proteome (Supplementary Excel File S12) and for intraconnections within mutome and amplicome (Supplementary Excel File S13). The largest overconnected hubs in the case of interconnections were proteins outside the sets (e.g., proteins encoded by neither gained nor mutated genes). ESRI was the most important transcription hub (overconnected TF) for the amplicome. ESRI had 102 direct targets among amplified genes (128 interactions total), which was statistically significantly \( P < 0.00003 \) higher than the expected 73; a 63% enrichment (Supplementary Excel File S12). In contrast, ESRI was statistically significantly \( P = 0.03 \) underconnected with the mutome (41 TR links with mutated genes compared with the expected 54). This could potentially be due to the different distribution of estrogen receptor (ER)-positive and ER-negative tumors within the amplicome and mutome, although ESRI remained a main transcriptional regulator of amplified genes even when we restricted the amplicome analysis to ER-negative tumors (data not shown). Interestingly, the amplicome was specifically enriched in protein-encoding ESRI target genes but not in ESRI binding sites in general. About 11% of all ERI binding sites, determined based on ChIP-on-chip studies in MCF-7 cells (26), are located in the amplicome, which constitutes ~10% of the genome (Supplementary Table S5). The size of individual amplicons correlated \( P = 0.94 \) with the number of ERI sites they contained.

Other notably statistical significant outside (not amplified) overconnected hubs for the amplicome included TFs ESRI2, RUNX2, PIT1, CSD4, and MBD1; ligands SPP1, LAMA4, CTGF, PDGF-A, and CCL24; EPHA receptors and PGE2RH, kinases PRKCD, PRKACB, MAP2K5, mammalian target of rapamycin, and PRKCK2 and proteases S Chen, ABHD7, AMPL, CTRC, and CLPP (Supplementary Excel File S12).

The top notability statistically significant overconnected TFs within the mutome were HIVEP1, IRF7, NRSF, NEUROD2, and PAX8; receptors ROS1, FCGR1A, IL1R83, IGF1R, and E-selectin; ligands IFN5, IL-12x, FGF10, Jagged2, and IFN-g; kinases Chk1, LKB1, ZIP-kinase, Fyn, and BUBR1; and proteases caspase-3, calpain 3, Granzyme B, presenelin 1, and caspase-7.

Within the amplicome, C/EPBp, HOXB4, HOXB1, cKrox, and c-Myc were the top-scored self-regulatory TFs. The highest scored overconnected receptors included ITGB4, FGFR1, PGD2R, ADRB3, and PIGR, whereas the most connected ligands included K12, MIP-1, Substance P, Neuroptide W, and IL-24. The most relevant intranconnected kinases were MAPKAPK2, CLK2, Fructosamine-3-kinase, Casein kinase i6, and PKC-α; proteases included CAAX prenyl protease 2, cathepsin F, ADAM-33, matrix metalloproteinase (MMP)-17, and MMP-25. (Supplementary Excel File S13).

Ontology enrichment analysis of the breast cancer mutome and amplicome. To identify gene functions and pathways enriched in the breast cancer mutome and amplicome, we performed enrichment analysis in five functional ontologies (canonical pathways, process networks, disease biomarkers, GO processes, and GO molecular functions). Both the mutome and amplicome were enriched for genes and processes involved in tumorigenesis (Supplementary Excel File S14). The mutome was enriched for cell adhesion, cell cycle, and DNA damage pathways (this later was especially pronounced for the CAN genes), whereas the amplicome was enriched for developmental pathways. With respect of disease biomarkers (e.g., genes linked to over 500 diseases based on genetic and biochemical data), the amplicome was highly enriched for genes associated with breast cancer \( P = 2.03 \times 10^{-28} \) and skin disease \( 3.48 \times 10^{-17} \). Surprisingly, the top 15 disease markers enriched in the mutome did not include cancer-associated biomarkers (with the exception of Leukemia), but markers for different nervous system disorders (Supplementary Excel File S14). This suggests that activation of tumorigenic pathways by gene amplification is a more significant mechanism driving breast tumorigenesis than somatic mutations.

Synergy in ontology enrichment distributions between breast cancer mutome and amplicome. To determine if mutated and amplified genes cooperatively alter certain pathways and networks, we analyzed synergy in ontology enrichment distributions between mutome and amplicome and found high degree of functional connectivity. Synergy was defined by a lower \( P \) value for an ontology entity in the distribution of the nonredundant union of amplicome and mutome gene content compared with amplicome and mutome individually. For example, pathway maps “Cell adhesion_Chemokines and adhesion” scored \( P = 3.2 \times 10^{-15} \) for the union set and \( 3.1 \times 10^{-9} \) for the mutome (Supplementary Excel File S14). The three orders of magnitude difference in \( P \) values reflects the enhancement of the mutant genes by amplified genes in this particular pathway and indicates functional connectivity between amplicome and mutome because a union of functionally disconnected data sets would yield higher \( P \) values in enrichment analysis due to its larger size.

Enrichment synergy between amplicome and mutome was evident for certain functional ontologies canonical pathway and processes networks (Fig. 4). The highest synergy in canonical pathways was found for “Cell adhesion_Chemokines and adhesion,” “Cytoskeleton remodeling_TGF, WNT and cytoskeletal remodeling,” “Cell cycle_ATM/ATR regulation of G1-S checkpoint,” and “Development_IGF-RI signaling” (three orders of magnitude). These synergies suggest close interactions between mutated and amplified genes in these selected
In summary, our results strongly support the conclusion that a combination of multiple genetic alterations is necessary for the aberrant self-sustaining activation of multiple tumorigenic signal pathways in breast cancer. Systemic genetic and bioinformatics analyses of human tumors, such as our report presented here, can lay the groundwork for future translational studies exploring the potential therapeutic targeting of key regulators of the malignant cellular phenotype.

References


Disclosure of Potential Conflicts of Interest

K. Polyak: research support and consultant, Novartis Pharmaceuticals Inc.; consultant, GeneGo, Inc. and Aveo Pharmaceuticals, Inc. The other authors disclosed no potential conflicts of interest.

Acknowledgments

Received 8/11/2008; revised 8/26/2008; accepted 9/2/2008.

Grant support: National Cancer Institute (CA89393 and CA94074) grants (K. Polyak), NCI (CA134175) to Y. Nikolsky, and NCI (CA112828) to T. Nikolskaya, and by GeneGo, Inc.

The costs of publication of this article were defrayed in part by the payment of page charges. This article must therefore be hereby marked advertisement in accordance with 18 U.S.C. Section 1734 solely to indicate this fact.

We thank Dr. Andrea Richardson for help with the acquisition of tumor samples and Drs. William Hahn and Matthew Meyerson for their critical reading of the manuscript.
Genome-Wide Functional Synergy between Amplified and Mutated Genes in Human Breast Cancer

Yuri Nikolsky, Evgeny Sviridov, Jun Yao, et al.


Updated version  Access the most recent version of this article at: http://cancerres.aacrjournals.org/content/68/22/9532

Supplementary Material  Access the most recent supplemental material at: http://cancerres.aacrjournals.org/content/suppl/2008/11/18/68.22.9532.DC1

Cited articles  This article cites 26 articles, 11 of which you can access for free at: http://cancerres.aacrjournals.org/content/68/22/9532.full.html#ref-list-1

Citing articles  This article has been cited by 11 HighWire-hosted articles. Access the articles at: /content/68/22/9532.full.html#related-urls

E-mail alerts  Sign up to receive free email-alerts related to this article or journal.

Reprints and Subscriptions  To order reprints of this article or to subscribe to the journal, contact the AACR Publications Department at pubs@aacr.org.

Permissions  To request permission to re-use all or part of this article, contact the AACR Publications Department at permissions@aacr.org.