Cancer Research

ÆÆℝ

Pathway Analysis of Breast Cancer Genome-Wide Association Study Highlights Three Pathways and One Canonical Signaling Cascade

Idan Menashe, Dennis Maeder, Montserrat Garcia-Closas, et al.

Updated version	Access the most recent version of this article at: doi:10.1158/0008-5472.CAN-09-4502
Supplementary Material	Access the most recent supplemental material at: http://cancerres.aacrjournals.org/content/suppl/2010/05/10/0008-5472.CAN-09-4502.DC1.html

Cited Articles	This article cites by 37 articles, 8 of which you can access for free at: http://cancerres.aacrjournals.org/content/70/11/4453.full.html#ref-list-1
Citing articles	This article has been cited by 12 HighWire-hosted articles. Access the articles at: http://cancerres.aacrjournals.org/content/70/11/4453.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.

Pathway Analysis of Breast Cancer Genome-Wide Association Study Highlights Three Pathways and One Canonical Signaling Cascade

Idan Menashe[1], Dennis Maeder[1], Montserrat Garcia-Closas[1], Jonine D. Figueroa[1], Samsiddhi Bhattacharjee[1], Melissa Rotunno[1], Peter Kraft[2], David J. Hunter[2], Stephen J. Chanock[1], Philip S. Rosenberg[1], and Nilanjan Chatterjee[1]

Abstract

Genome-wide association studies (GWAS) focus on relatively few highly significant loci, whereas less attention is given to other genotyped markers. Using pathway analysis to study existing GWAS data may shed light on relevant biological processes and illuminate new candidate genes. We applied a pathway-based approach to the breast cancer GWAS data of the National Cancer Institute (NCI) Cancer Genetic Markers of Susceptibility project that includes 1,145 cases and 1,142 controls. Pathways were retrieved from three databases: KEGG, BioCarta, and NCI Protein Interaction Database. Genes were represented by their most strongly associated SNP, and an enrichment score reflecting the overrepresentation of gene-based association signals in each pathway was calculated by using a weighted Kolmogorov-Smirnov procedure. Finally, hierarchical clustering was used to identify pathways with overlapping genes, and clusters with an excess of association signals were determined by the adaptive rank-truncated product (ARTP) method. A total of 421 pathways containing 3,962 genes was included in our study. Of these, three pathways (syndecan-1–mediated signaling, signaling of hepatocyte growth factor receptor, and growth hormone signaling) were highly enriched with association signals [$P_{ES} < 0.001$, false discovery rate (FDR) = 0.118]. Our clustering analysis revealed that pathways containing key components of the RAS/RAF/mitogen-activated protein kinase canonical signaling cascade were significantly more likely to have an excess of association signals than expected by chance ($P_{ARTP} = 0.0051$, FDR = 0.07). These results suggest that genetic alterations associated with these three pathways and one canonical signaling cascade may contribute to breast cancer susceptibility. *Cancer Res; 70(11); 4453–9. ©2010 AACR.*

Introduction

Breast cancer is a complex disease with a well-established genetic component (1–4). Different genetic methodologies have been used in the last 20 years to identify multiple susceptibility loci that vary in population frequency, relative risk, and potential functional role. For example, familial linkage and positional cloning studies found that rare mutations in the genes *BRCA1, BRCA2, TP53*, and *PTEN* are associated with a 10-fold to 20-fold increase in breast cancer risk (5–8). Variations in these genes are thought to contribute to breast cancer susceptibility through different cellular mechanisms. Whereas *BRCA1* and *BRCA2* belong to the DNA repair mechanism in cells (9, 10), *TP53* and *PTEN* are tumor suppressor genes that participate in processes related to cell cycle control and cell proliferation (11, 12). Further interrogation of candidate genes associated with these cellular processes has led to the discovery of additional rare genetic variants conferring moderate relative risks (2–3-fold) of breast cancer (13–16).

Recently, genome-wide association studies (GWAS) have become a key paradigm in genetic studies of complex diseases. These studies are particularly powerful in identifying common-low penetrance risk alleles. Indeed, several novel markers with low (<2) relative risk for breast cancer have been detected by this approach (17–21). However, these reported loci are only those that reached a stringent statistical "genome-wide" significance criterion, whereas less attention has been given to the other hundreds of thousands of genotyped markers. Using new methods to study existing GWAS data may provide additional biological insights and highlight new candidate loci. To this end, a pathway-based approach is particularly appealing. This method examines whether the cumulative contribution of genes with a common biological denominator is greater than expected by chance. This

approach has recently been applied to GWAS of several non-cancer complex diseases (22–25).

In this study, we applied pathway analysis to the breast cancer GWAS of the Cancer Genetic Markers of Susceptibility (CGEMS) project of the National Cancer Institute (NCI). This study has recently identified a significant association of single nucleotide polymorphisms (SNPs) in the *FGFR2* gene with breast cancer susceptibility (18). We used the modified gene-set enrichment analysis of Wang and colleagues (22) to identify an excess of genotype-phenotype association signals in pathways from different resources. Finally, we examined whether the excess of association signals in different pathways is driven by the same subset of genes comprising a common biological module. These analyses illuminated three pathways and one canonical cascade that are possibly involved in genetic susceptibility to breast cancer.

Materials and Methods

Pathway data construction

We collected pathway data from three widely used resources: the BioCarta pathway database (26), the Kyoto Encyclopedia of Genes and Genomes (KEGG; 27), and the NCI Protein Interaction Database (PID; ref. 28). Genes belonging to these pathways were associated with SNPs included in the Illumina Sentrix HumanHap550 genotyping BeadChip that was used in the CGEMS GWAS. SNPs were mapped to genes if they were located within a genomic region encompassing 20 kb 5′ upstream and 10 kb 3′ downstream of the gene coding region (National Center for Biotechnology Information human genome build 36.3). Because *FGFR2* is highly associated with breast cancer in CGEMS (18), we excluded all SNPs mapped to this gene from our analysis. Finally, we restricted our analysis to pathways with 10 to 100 genes to alleviate the multiple testing problem by avoiding testing too narrowly or too broadly defined functional categories. A complete list of the studied pathways is available in Supplementary Table S1.

Gene-set enrichment analysis

The CGEMS breast cancer GWAS includes 1,145 postmenopausal women of European ancestry with invasive breast cancer and 1,142 controls. Associations between SNPs and breast cancer were determined using the Cochran-Armitage test for trend (29). Each gene G_j (j = 1, …, N, where N is the total number of gene in our data set) was assigned the χ^2 trend test statistic of the SNP with the highest statistical significance among all genotyped SNPs that mapped to its region (r_j). Next, for each pathway (S), we ordered its gene test statistics ($r_j \in S$) from the highest to lowest, and used a weighted Kolmogorov-Smirnov procedure to calculate the enrichment score (ES):

$$\text{ES}_S = \max_{1 \le j \le N} \left\{ \sum G_{j* \epsilon Sj* \le j} \frac{|r_{j*}|}{W_S} - \sum G_{j* \notin Sj* \le j} \frac{1}{N - N_H} \right\} \quad (1)$$

where $W_S = \sum G_{j* \epsilon S} |r_{j*}|$ and N_H is the number of genes in a pathway. This statistics reflects the relative overrepresentation of genotype-phenotype association signals in a particular set of genes (30). We used a permutation procedure (10,000 permutations, permuting the case-control status) to assess the significance of the pathway-based ES. This procedure generates a null distribution for the ES of each pathway based on its real genotype data and therefore accounts for the variable SNP count and the intrinsic correlation between SNPs within the different genes. Finally, a normalized enrichment score (NES) was calculated for each pathway in the observed and permutated data to allow a direct comparison of pathways of different sizes as suggested by Wang and colleagues (22). The calculation of the NES for each pathway, NES_S, relied on the ES_S, and the mean and SD of ES_S^{per} in all permutations for a given pathway (S):

$$\text{NES}_S = \frac{\text{ES}_S - \text{mean}(\text{ES}_S^{\text{per}})}{\text{SD}(\text{ES}_S^{\text{per}})} \quad (2)$$

These data were then used to calculate a false discovery rate (FDR; ref. 31) that controls for the proportion of expected false-positive findings to be under a certain threshold. For a pathway with a NES value of NES_S^*, the FDR is calculated as:

$$\text{FDR} = \frac{\% \text{ of all } (S, \text{per}) \text{ with } \text{NES}_S^{\text{per}} \ge \text{NES}_S^*}{\% \text{ of all observed } S \text{ with } \text{NES}_S \ge \text{NES}_S^*} \quad (3)$$

Pathways with FDR < 0.2 were considered noteworthy.

Pathway overlap

To explore the relationships between pathways in our database, we defined the overlap between the set of genes in pathway A and the set of genes in pathway B as the percentage of genes common to both pathways (A and B) and calculated it as follows:

$$\text{Overlap}(\%) = \frac{\#\{A \cap B\}}{\#\{A \cup B\}} \times 100\% \quad (4)$$

Next, we used a hierarchical clustering algorithm to assemble pathways with similar gene content. This algorithm iterates over all pathways, finds the most similar pair of pathways, and groups them together. This process was repeated until all pathways are grouped in one cluster in a hierarchical manner. To determine clusters of significantly related pathways, we used the same guidelines of the confidence interval algorithm originally developed for linkage disequilibrium block determination in the genome described by Gabriel and colleagues (32). In short, clusters were defined as containing significantly related pathways if they included ≥5 pathways, and if ≥95% of all the pairwise overlap scores were greater than 9.6% (the 95th percentile of all pairwise overlap scores in our database).

Finally, we applied the adaptive rank-truncated product (ARTP) method (33) using the 10,000 permutated data sets to assess whether pathways within the defined clusters had higher ES than expected by chance. This method assesses the null hypothesis that a cluster of pathways does not contain excess of pathways with high ES. Denote the ordered P values of the ES of pathways belonging to a particular cluster as ($p_1 \le \ldots \le p_L$), with p_1 being the smallest P value. To test

for the global null hypothesis, the ARTP procedure calculate the RTP statistics ($W_k = \Pi_{i=1}^{K} p_i$) over all possible truncation points (k), and assess the significance of the best RTP using the permutation data.

Results

Overall, 421 pathways and 3,962 genes (164/996, 155/3,088, and 102/1,317 pathways/genes from BioCarta, KEGG, and PID, respectively) were included in our study. These were represented by 69,525 of the 528,173 SNPs (13.2%) that were originally genotyped in this GWAS. A significant variation was seen between pathway resources. More than 90% of the pathways from BioCarta in our database had <30 genes, whereas pathways from KEGG and PID were generally larger and more variable in size (mean gene count, 44 and 36; SD, 24 and 16, respectively). This variation between databases was evident even in pathways nominally representing the same biological processes. For example, the mammalian target of rapamycin signaling pathway included 19, 24, and 49 genes in BioCarta, PID, and KEGG, respectively. Interestingly, pathways from BioCarta tended to be ranked higher (higher ES) than KEGG pathways but not than PID pathways (Kruskal-Wallis test, $P = 0.013$ and $P = 0.151$, respectively). No rank differences were seen between pathways from PID and KEGG.

Of the 421 examined pathways, 21 were significantly enriched with association signals at the $P < 0.05$ level (Table 1). Of these, "syndecan-1–mediated signaling events" from the PID database (Fig. 1A; $P_{ES} = 0.00053$), "signaling of hepatocyte growth factor receptor" from BioCarta (Fig. 1B; $P_{ES} = 0.00063$), and "growth hormone signaling" from BioCarta (Fig. 1C; $P_{ES} = 0.00084$) had distinctly smaller P values than the rest of the pathways, with a noteworthy FDR value of 0.118. For brevity, we will further refer to these three pathways as "syndecan-1 signaling," "c-Met signaling," and "growth hormone signaling," respectively.

Further examination of the gene content of these three pathways revealed some overlap. For example, the gene mitogen-activated protein kinase 3 (*MAPK3*) is involved in all three pathways but does not seem to contribute to their significant ES due to its nonsignificant association with breast cancer risk in this study ($P = 0.35$). In contrast, the gene hepatocyte growth factor (*HGF*) was the strongest associated gene in the syndecan-1 signaling and c-Met signaling pathways ($P_{trend} = 0.0007$, for rs975360) and therefore a major contributor for their high ES. Another two genes [*MAPK1* and met proto-oncogene (hepatocyte growth factor receptor; *MET*)] were shared by these two pathways. Similarly, six other genes [growth factor receptor-bound protein 2 (*GRB2*); v-Ha-ras Harvey rat sarcoma viral oncogene homologue (*HRAS1*); mitogen-activated protein kinase kinase 1 (*MAP2K1*); phosphoinositide-3-kinase, catalytic,

Table 1. A list of pathways associated with breast cancer susceptibility at the $P \leq 0.05$ level

Rank	Source	Pathway name	No. of genes	P_{ES}*	FDR[†]
1	PID	Syndecan-1-mediated signaling events	13	0.00053	0.118
2	BioCarta	Signaling of hepatocyte growth factor receptor	33	0.00063	0.118
3	BioCarta	Growth hormone signaling	22	0.00084	0.118
4	BioCarta	Insulin signaling	19	0.00642	0.442
5	KEGG	Basal cell carcinoma	53	0.00674	0.487
6	BioCarta	Eicosanoid metabolism	19	0.00695	0.515
7	PID	Signaling events mediated by stem cell factor receptor (c-Kit)	50	0.00779	0.515
8	KEGG	Tryptophan metabolism	57	0.01516	0.664
9	BioCarta	ADP-ribosylation factor	16	0.01800	0.842
10	PID	Fibroblast growth factor signaling pathway	52	0.02189	0.871
11	KEGG	ErbB signaling	86	0.02379	0.871
12	BioCarta	ALK in cardiac myocytes	33	0.02747	0.871
13	BioCarta	Interleukin-2 receptor β-chain in T-cell activation	34	0.03389	0.881
14	BioCarta	Intrinsic prothrombin activation	16	0.03600	0.881
15	KEGG	Streptomycin biosynthesis	10	0.03600	0.881
16	BioCarta	Angiotensin II–mediated activation of JNK pathway via Pyk2-dependent signaling	27	0.03663	0.881
17	BioCarta	Sprouty regulation of tyrosine kinase signals	15	0.04095	0.881
18	KEGG	Maturity onset diabetes of the young	24	0.04547	0.881
19	BioCarta	Regulation of hematopoiesis by cytokines	15	0.04568	0.881
20	KEGG	Phosphatidylinositol signaling system	75	0.04674	0.881
21	KEGG	Melanogenesis	98	0.04747	0.881

*P values are computed for the pathway ES using 10,000 permutations.
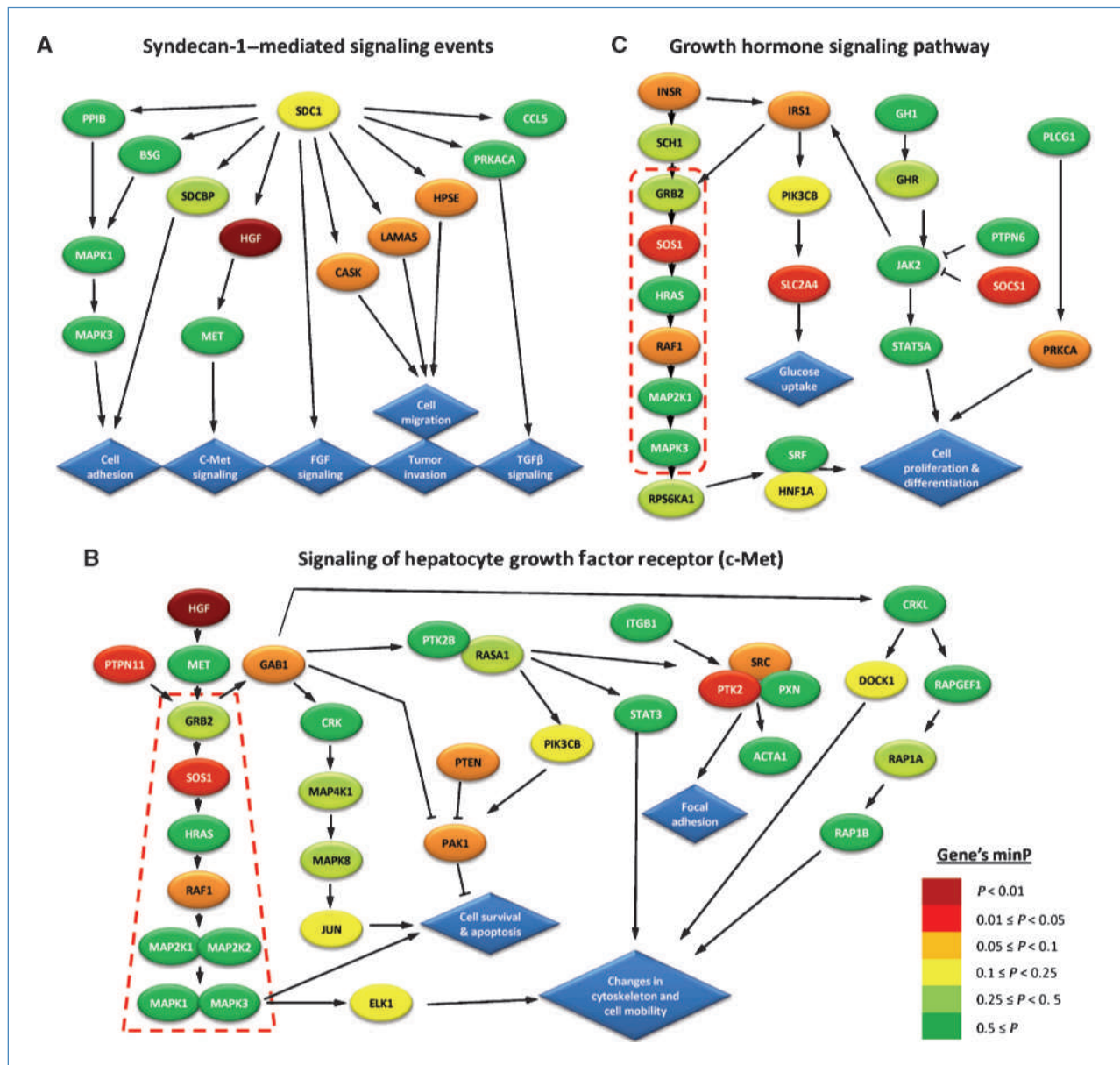[†]FDR for the 421 pathways examined in this study.

Figure 1. Schematic illustration of the three top ranked pathways in this study. A, syndecan-1–mediated signaling events from PID. B, signaling of hepatocyte growth factor (c-Met) from BioCarta. C, growth hormone signaling pathways from BioCarta. Genes (ellipses) are color-coded according to their *minP* value, which assesses the significance of the most strongly associated SNP in each gene, using a permutation-based resampling procedure (10,000 permutations) that takes into account the number of SNPs in a gene and their underlying linkage disequilibrium structure. Arrows indicate interaction or positive regulation, and perpendicular lines indicate negative regulation. Blue diamonds represent the cellular processes induced by these pathways. Red broken boxes outline the genes of the RAS/RAF/MAPK canonical signaling cascade.

β polypeptide (*PIK3CB*); v-raf-1 murine leukemia viral onco-gene homologue (*RAF1*); and son of sevenless homologue 1 (*SOS1*)] were shared by the c-Met signaling and growth hormone signaling pathways.

In light of the considerable overlap between the three most significant pathways in our study, we hypothesized that common biological modules may underlie the enrichment signals in multiple biological pathways. To explore this hypothesis, we used hierarchical clustering to identify pathways with overlapping genes likely to constitute common biological

process. Consequently, 16 clusters of highly related pathways were characterized (Fig. 2A). Of these, cluster #1 (Fig. 2B) was significantly more likely to include pathways with high ES than expected by chance (ARTP $P = 0.0045$, FDR = 0.07). Further examination of this cluster revealed that it contained two of three of the top three pathways (signaling of hepatocyte growth factor receptor and growth hormone signaling) and was dominated by the genes *GRB2, SOS1, HRAS, RAF1, MAP2K1*, and *MAPK3*, which are major components of the RAS/RAF/MAPK canonical signaling cascade (Fig. 1B and C).

To assess the power of our analysis to identify true susceptibility pathways, we constructed an artificial positive control pathway from 11 genes that were previously associated with breast cancer risk in different GWAS of postmenopausal women of European ancestry (see details in Supplemental Table S2). Applying the pathway analysis to this pseudopathway yielded a $P_{ES} = 0.0126$, which ranked it 8th out of all pathways in the study. It is noteworthy that most of the individual genes in the pseudo-pathway had unremarkable significance (Supplemental Table S2, last column) in the CGEMS study itself. Thus, the high rank of this artificial positive control pathway shows the ability of the gene-set enrichment analysis to detect pathways containing multiple genes that individually have weak association signals and may be missed by standard single-SNP analysis of GWAS data.

Discussion

Three pathways and one signaling cascade are highlighted in this study. The top ranked pathway in this study, syndecan-1 signaling, contains 13 genes involved in different cellular processes that are mediated by syndecan-1 (*SDC1*). This gene encodes a transmembrane heparan sulfate proteoglycan that mediates signal transduction cascades leading to cell proliferation, cell migration, and cell adhesion processes following interactions with extracellular matrix proteins. There are multiple lines of evidence for a potential role of syndecan-1 in breast cancer development. For example, altered syndecan-1 expression has been detected in several different tumor types and has been linked with unfavorable breast cancer prognosis (34). Additionally, expression of

Figure 2. Clustering analysis of pathway overlap. Pairwise pathway overlap scores are depicted in a matrix color coded in grayscale. A, clusters of highly significantly related pathways are delineated by black triangles. B, details of the pathways included in cluster #1.

P value	Pathway
0.03663	Angiotensin II–mediated activation of JNK pathway via Pyk2 dependent signaling
0.08432	BCR signaling pathway
0.79126	Fc epsilon receptor I signaling in mast cells
0.53758	EGF signaling pathway
0.42737	PDGF signaling pathway
0.66232	TPO signaling pathway
0.71716	EPO signaling pathway
0.47916	Nerve growth factor pathway (NGF)
0.33232	IGF-1 signaling pathway
0.00642	Insulin signaling pathway
0.17274	IL-6 signaling pathway
0.17400	IL-2 signaling pathway
0.48053	Inhibition of cellular proliferation by Gleevec
0.00084	Growth hormone signaling pathway
0.88768	Role of ERBB2 in signal transduction and oncology
0.51842	Fc-epsilon receptor I signaling in mast cells
0.60389	Signaling pathway from G-protein families
0.03389	IL-2 receptor β chain in T-cell activation
0.10442	Erk and PI-3 kinase are necessary for collagen binding in corneal epithelia
0.42453	Integrin signaling pathway
0.00063	Signaling of hepatocyte growth factor receptor
0.18842	Signaling events activated by hepatocyte growth factor receptor (c-Met)
0.68411	IL-2–mediated signaling events
0.04095	Sprouty regulation of tyrosine kinase signals
0.49126	Bioactive peptide induced signaling pathway
0.09168	Transcription factor CREB and its extracellular signals
0.26063	Multiple antiapoptotic pathways from IGF-IR signaling lead to BAD phosphorylation
0.79674	Role of ErkS in neuronal survival
0.63937	Trka receptor signaling pathway
0.96337	PDGFR-α signaling pathway

syndecan-1 by stromal fibroblasts has been shown to promote breast carcinoma growth *in vivo* and stimulates tumor angiogenesis (35). In this GWAS, *SDC1* was only moderately associated with breast cancer susceptibility (P_{trend} = 0.019 for rs7563245); however, it is a key mediator of different pathways illuminated in this study (e.g., c-Met signaling and fibroblast growth factor signaling; see discussion below). Therefore, it can modulate breast cancer susceptibility through different biological mechanisms.

The second highest ranked pathway in our analysis, c-Met signaling, consists of 33 genes participating in signal transduction mechanisms induced by the tyrosine kinase proto-oncogene c-Met (*MET*). Stimulation of the c-Met pathway can lead to several different cellular processes related to tumor growth and progression, such as proliferation, enhanced cell motility, invasion, and apoptosis (36). Moreover, both c-Met and its ligand, hepatocyte growth factor, have been shown to be dysregulated and correlated with poor prognosis in a number of human malignancies, including breast cancer (37). Consequently, this pathway has served as an important therapeutic target for human cancers, particularly through the development of a small molecule that inhibit the c-Met/hepatocyte growth factor–dependent signaling (37). Notably, coexpression of *SDC1* and *MET*, the key players in the top two ranked pathways in our study, have been established as a marker signature associated with poor prognostic factors in ductal carcinoma *in situ* of the breast (38).

The third ranked pathway in our study, growth hormone signaling, contains 22 genes participating in cellular mechanisms induced by either growth hormone or insulin receptors. These two receptors as well as the insulin-like growth factor (IGF) receptor are all transmembrane tyrosine kinase receptors that induce cell growth and proliferation. Alteration in the activity of these receptors or their related pathways may lead to hyperplasia and eventually to the development of a tumor (39). Naturally, the growth hormone signaling pathway had considerable overlap with both the IGF-I signaling and insulin signaling pathways from BioCarta (32% and 41%, respectively); however, only the latter had a significant ES in our study (P_{ES} = 0.0064). Examining the genes of these three closely related pathways revealed that what differentiates their respective ES is a SNP in the insulin receptor gene (*INSR*) with a small *P* value (P_{trend} = 0.0019 for rs12460755) that is absent from the IGF-I signaling pathway.

Interestingly, the fibroblast growth factor signaling pathway was ranked 10th in our study (Table 1; P_{ES} = 0.0219) despite the exclusion of the *FGFR2* SNPs from our analyses (see Materials and Methods). Adding the *FGFR2* SNPs to this pathway improved its enrichment signal (P_{ES} = 0.0053) and ranked it 4th in our study. This finding, in combination with the *FGFR2* signal in CEGEMS (18) and elsewhere (17), suggests that variations in other genes involved in *FGFR2* signaling may modulate breast cancer susceptibility.

An important extension to the gene-set enrichment analysis is the clustering analysis that highlighted the RAS/RAF/MAPK canonical signaling cascade as the common denominator of pathways associated with breast cancer risk in this study. This cascade plays an essential role in transmitting

extracellular signals from growth factors to promote the growth, proliferation, differentiation, and survival of cells, and modification in its activity has been linked to multiple human malignancies (40). Notably, this cascade plays an important role in all three top pathways in this study. It is an integral component in the signaling of hepatocyte growth factor receptor and growth hormone signaling pathways, and a transducer for signals initiated by the syndecan-1 pathway.

An important limitation of the pathway-based approach for GWAS analysis is the incomplete annotation of the human genome. At present, the function of many human genes is unknown and therefore they cannot be assigned to known pathways. Moreover, susceptibility loci in intergenic regions are also not included in a study of this kind. As a result, when using this approach, only a small portion of the human genome variation can be studied and therefore it should only be used as a supplementary method to the standard single-SNP analysis of GWAS. Additionally, there is no gold standard or pathway definition, and different databases have different guidelines for their pathway construction and curation. Consequently, the gene content of pathways representing the same biological process may vary substantially between different databases, and this may have a major impact on the sensitivity and specificity of this approach. We aimed at minimizing this effect by selecting pathways from three commonly used and manually curate resources. Still, considerable differences were observed between similar pathways from different resources. For example, the c-Met signaling pathway from BioCarta contained 33 genes and was ranked second in our analysis, whereas a pathway with the same name and presumably the same cellular function in the PID database included 52 genes and was ranked only 69th. Although 30 of the 33 genes in the BioCarta pathway were also included in the PID pathway, the remaining 22 genes likely attenuated the signal in the PID pathway. These differences emphasize the importance of further synthesizing the results of such pathway-based approach. The clustering analysis we applied in this study is one way to do so, as it helps in finding the common genes underlying enrichment signals in multiple pathways and allows one to focus on a limited number of candidate genes. Other methods aiming to improve the characterization of pathways and their overlap might be useful.

A second limitation of our study is that it does not include validation of the results using a completely independent data set. The experiment we conducted using a positive control pathway with known breast cancer susceptibility genes provides us validation that our approach has high power to detect pathways containing multiple weakly associated susceptibility genes. Therefore, the top ranked pathways in our study should be prime targets for future analyses in independent data sets.

In conclusion, our results suggest that genetic alterations associated with the top three pathways and one canonical signaling cascade in our study may contribute to breast cancer susceptibility. Ultimately, additional studies would be needed to confirm and further explore the genetic variations underlying the association of these pathways with

breast cancer. Moreover, this study highlights the potential insight that could be gained by using a pathway-based approach as a complimentary method to the primary single-SNP analysis of GWAS. Particularly, the organization of multiple association signals according to underlying biological processes may improve our understanding of the cellular mechanisms underlying this too common malignancy.

Disclosure of Potential Conflicts of Interest

No potential conflicts of interest were disclosed.

References

1. Cancer CGoHFiB. Familial breast cancer: collaborative reanalysis of individual data from 52 epidemiological studies including 58,209 women with breast cancer and 101,986 women without the disease. Lancet 2001;358:1389–99.
2. Ahlbom A, Lichtenstein P, Malmstrom H, Feychting M, Hemminki K, Pedersen NL. Cancer in twins: genetic and nongenetic familial risk factors. J Natl Cancer Inst 1997;89:287–93.
3. Lichtenstein P, Holm NV, Verkasalo PK, et al. Environmental and heritable factors in the causation of cancer-analyses of cohorts of twins from Sweden, Denmark, and Finland. N Engl J Med 2000;343:78–85.
4. Peto J, Mack TM. High constant incidence in twins and other relatives of women with breast cancer. Nat Genet 2000;26:411–4.
5. Miki Y, Swensen J, Shattuck-Eidens D, et al. A strong candidate for the breast and ovarian cancer susceptibility gene BRCA1. Science 1994;266:66–71.
6. Wooster R, Bignell G, Lancaster J, et al. Identification of the breast cancer susceptibility gene BRCA2. Nature 1995;378:789–92.
7. Malkin D, Li FP, Strong LC, et al. Germ line p53 mutations in a familial syndrome of breast cancer, sarcomas, and other neoplasms. Science 1990;250:1233–8.
8. Nelen MR, Padberg GW, Peeters EA, et al. Localization of the gene for Cowden disease to chromosome 10q22-23. Nat Genet 1996;13:114–6.
9. Jasin M. Homologous repair of DNA damage and tumorigenesis: the BRCA connection. Oncogene 2002;21:8981–93.
10. Venkitaraman AR. Cancer susceptibility and the functions of BRCA1 and BRCA2. Cell 2002;108:171–82.
11. Tokino T, Nakamura Y. The role of p53-target genes in human cancer. Crit Rev Oncol Hematol 2000;33:1–6.
12. Keniry M, Parsons R. The role of PTEN signaling perturbations in cancer and in targeted therapy. Oncogene 2008;27:5477–85.
13. Meijers-Heijboer H, van den Ouweland A, Klijn J, et al. Low-penetrance susceptibility to breast cancer due to CHEK2(*)1100delC in noncarriers of BRCA1 or BRCA2 mutations. Nat Genet 2002;31:55–9.
14. Rahman N, Seal S, Thompson D, et al. PALB2, which encodes a BRCA2-interacting protein, is a breast cancer susceptibility gene. Nat Genet 2007;39:165–7.
15. Renwick A, Thompson D, Seal S, et al. ATM mutations that cause ataxia-telangiectasia are breast cancer susceptibility alleles. Nat Genet 2006;38:873–5.
16. Seal S, Thompson D, Renwick A, et al. Truncating mutations in the Fanconi anemia J gene BRIP1 are low-penetrance breast cancer susceptibility alleles. Nat Genet 2006;38:1239–41.
17. Easton DF, Pooley KA, Dunning AM, et al. Genome-wide association study identifies novel breast cancer susceptibility loci. Nature 2007; 447:1087–93.
18. Hunter DJ, Kraft P, Jacobs KB, et al. A genome-wide association study identifies alleles in FGFR2 associated with risk of sporadic postmenopausal breast cancer. Nat Genet 2007;39:870–4.
19. Gold B, Kirchhoff T, Stefanov S, et al. Genome-wide association study provides evidence for a breast cancer risk locus at 6q22.33. Proc Natl Acad Sci U S A 2008;105:4340–5.
20. Zheng W, Long J, Gao YT, et al. Genome-wide association study

identifies a new breast cancer susceptibility locus at 6q25.1. Nat Genet 2009;41:324–8.
21. Thomas G, Jacobs KB, Kraft P, et al. A multistage genome-wide association study in breast cancer identifies two new risk alleles at 1p11.2 and 14q24.1 (RAD51L1). Nat Genet 2009;41:579–84.
22. Wang K, Li M, Bucan M. Pathway-based approaches for analysis of genomewide association studies. Am J Hum Genet 2007;81:1278–83.
23. Wang K, Zhang H, Kugathasan S, et al. Diverse genome-wide association studies associate the IL12/IL23 pathway with Crohn disease. Am J Hum Genet 2009;84:399–405.
24. Baranzini SE, Galwey NW, Wang J, et al. Pathway and network-based analysis of genome-wide association studies in multiple sclerosis. Hum Mol Genet 2009;18:2078–90.
25. Elbers CC, van Eijk KR, Franke L, et al. Using genome-wide pathway analysis to unravel the etiology of complex diseases. Genet Epidemiol 2009;33:419–31.
26. BioCarta. [cited; available from: http://www.biocarta.com/genes/allpathways.asp].
27. KEGG. Kyoto encyclopedia of genes and genomes. Available from: http://www.genome.jp/kegg/.
28. PID. Pathway interaction database. Available from: http://pid.nci.nih.gov/.
29. Armitage P. Tests for linear trends in proportions and frequencies. Biometrics 1955;11:375–86.
30. Subramanian A, Tamayo P, Mootha VK, et al. Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. Proc Natl Acad Sci U S A 2005;102:15545–50.
31. Reiner A, Yekutieli D, Benjamini Y. Identifying differentially expressed genes using false discovery rate controlling procedures. Bioinformatics 2003;19:368–75.
32. Gabriel SB, Schaffner SF, Nguyen H, et al. The structure of haplotype blocks in the human genome. Science 2002;296:2225–9.
33. Yu K, Li Q, Bergen AW, et al. Pathway analysis by adaptive combination of P-values. Genet Epidemiol 2009;33:700–9.
34. Leivonen M, Lundin J, Nordling S, von Boguslawski K, Haglund C. Prognostic value of syndecan-1 expression in breast cancer. Oncology 2004;67:11–8.
35. Maeda T, Desouky J, Friedl A. Syndecan-1 expression by stromal fibroblasts promotes breast carcinoma growth in vivo and stimulates tumor angiogenesis. Oncogene 2006;25:1408–12.
36. Maulik G, Shrikhande A, Kijima T, Ma PC, Morrison PT, Salgia R. Role of the hepatocyte growth factor receptor, c-Met, in oncogenesis and potential for therapeutic inhibition. Cytokine Growth Factor Rev 2002;13:41–59.
37. Christensen JG, Burrows J, Salgia R. c-Met as a target for human cancer and characterization of inhibitors for therapeutic intervention. Cancer Lett 2005;225:1–26.
38. Gotte M, Kersting C, Radke I, Kiesel L, Wulfing P. An expression signature of syndecan-1 (CD138), E-cadherin and c-met is associated with factors of angiogenesis and lymphangiogenesis in ductal breast carcinoma in situ. Breast Cancer Res 2007;9:R8.
39. Wagner K, Hemminki K, Forsti A. The GH1/IGF-1 axis polymorphisms and their impact on breast cancer development. Breast Cancer Res Treat 2007;104:233–48.
40. Downward J. Targeting RAS signalling pathways in cancer therapy. Nat Rev Cancer 2003;3:11–22.