nature genetics

Network modeling links breast cancer susceptibility and centrosome dysfunction

Miguel Angel Pujana[1,2,16,17], Jing-Dong J Han[1,2,16,17], Lea M Starita[3,16,17], Kristen N Stevens[4,17], Muneesh Tewari[1,2,16], Jin Sook Ahn[1,2], Gad Rennert[5], Víctor Moreno[6,7], Tomas Kirchhoff[8], Bert Gold[9], Volker Assmann[10], Wael M ElShamy[2], Jean-François Rual[1,2], Douglas Levine[8], Laura S Rozek[6], Rebecca S Gelman[11], Kristin C Gunsalus[12], Roger A Greenberg[2], Bijan Sobhian[2], Nicolas Bertin[1,2], Kavitha Venkatesan[1,2], Nono Ayivi-Guedehoussou[1,2,16], Xavier Solé[7], Pilar Hernández[13], Conxi Lázaro[13], Katherine L Nathanson[14], Barbara L Weber[14], Michael E Cusick[1,2], David E Hill[1,2], Kenneth Offit[8], David M Livingston[2], Stephen B Gruber[4,6,15], Jeffrey D Parvin[3,16] & Marc Vidal[1,2]

Many cancer-associated genes remain to be identified to clarify the underlying molecular mechanisms of cancer susceptibility and progression. Better understanding is also required of how mutations in cancer genes affect their products in the context of complex cellular networks. Here we have used a network modeling strategy to identify genes potentially associated with higher risk of breast cancer. Starting with four known genes encoding tumor suppressors of breast cancer, we combined gene expression profiling with functional genomic and proteomic (or 'omic') data from various species to generate a network containing 118 genes linked by 866 potential functional associations. This network shows higher connectivity than expected by chance, suggesting that its components function in biologically related pathways. One of the components of the network is *HMMR*, encoding a centrosome subunit, for which we demonstrate previously unknown functional associations with the breast cancer–associated gene *BRCA1*. Two case-control studies of incident breast cancer indicate that the *HMMR* locus is associated with higher risk of breast cancer in humans. Our network modeling strategy should be useful for the discovery of additional cancer-associated genes.

Combinations of mutated and/or aberrantly expressed tumor suppressor genes and oncogenes, or 'cancer genes', are thought to be responsible for most steps of cancer progression. Although fundamental principles have emerged from the study of known cancer genes and their products, many questions remain unanswered. Notably, most cancer genes remain to be identified[1]. In addition, it is becoming increasingly clear that most genes and their products interact in complex cellular networks, the properties of which might be altered in cancer cells as compared with their unaffected counterparts[2]. Achieving a deeper understanding of cancer molecular mechanisms

[1]Center for Cancer Systems Biology (CCSB) and [2]Department of Cancer Biology, Dana-Farber Cancer Institute and Department of Genetics, Harvard Medical School, 44 Binney St., Boston, Massachusetts 02115, USA. [3]Department of Pathology, Brigham and Women's Hospital and Harvard Medical School, 77 Louis Pasteur Ave., Boston, Massachusetts 02115, USA. [4]Department of Epidemiology, University of Michigan, 109 Zina Pitcher Pl., Ann Arbor, Michigan 48109, USA. [5]CHS National Cancer Control Center, Department of Community Medicine and Epidemiology, Carmel Medical Center and Bruce Rappaport Faculty of Medicine, Technion, Haifa 34362, Israel. [6]Department of Internal Medicine, University of Michigan, 109 Zina Pitcher Pl., Ann Arbor, Michigan 48109, USA. [7]Department of Epidemiology and Cancer Registry, and Translational Research Laboratory, Catalan Institute of Oncology, IDIBELL, Gran Vía km 2.7, L'Hospitalet, Barcelona 08907, Spain. [8]Clinical Genetics Service, Department of Medicine, Memorial Sloan-Kettering Cancer Center, 1275 York Ave., New York, New York 10021, USA. [9]National Cancer Institute, Human Genetics Section, Laboratory of Genomic Diversity, Frederick, Maryland 21702, USA. [10]Center for Experimental Medicine, Institute of Tumor Biology, University Hospital Hamburg–Eppendorf, Martinistrasse 52, Hamburg 20246, Germany. [11]Department of Biostatistics and Computational Biology, Dana-Farber Cancer Institute and Department of Biostatistics, Harvard School of Public Health, 44 Binney St., Boston, Massachusetts 02115, USA. [12]Center for Comparative Functional Genomics, Department of Biology, New York University, 100 Washington Square East, New York, New York 10003, USA. [13]Translational Research Laboratory, Catalan Institute of Oncology, IDIBELL, Gran Vía km 2.7, L'Hospitalet, Barcelona 08907, Spain. [14]Abramson Family Cancer Research Institute, University of Pennsylvania School of Medicine, 421 Curie Blvd., Philadelphia, Pennsylvania 19104, USA. [15]Department of Human Genetics, University of Michigan, 109 Zina Pitcher Pl., Ann Arbor, Michigan 48109, USA. [16]Present addresses: Bioinformatics and Biostatistics Unit, Translational Research Laboratory, Catalan Institute of Oncology, IDIBELL, Gran Vía km 2.7, L'Hospitalet, Barcelona 08907, Spain (M.A.P.); Institute of Genetics and Developmental Biology, Chinese Academy of Sciences, Datun Rd., Beijing 100101, China (J.-D.J.H.); Department of Genome Sciences, University of Washington, 1705 NE Pacific St., Seattle, Washington 98195, USA (L.M.S.); Human Biology Division, Fred Hutchinson Cancer Research Center, 1100 Fairview Ave. North, Seattle, Washington 98109, USA (M.T.); Harvard School of Public Health, Boston, Massachusetts 02115, USA (N.A.-G.); Department of Biomedical Informatics, Ohio State University Medical Center, 460 West 12th Ave., Columbus, Ohio 43210, USA (J.D.P.). [17]These authors contributed equally to this work. Correspondence should be addressed to M.V. (marc_vidal@dfci.harvard.edu), J.D.P. (jeffrey.parvin@osumc.edu) or S.B.G. (sgruber@med.umich.edu).

Received 31 March; accepted 2 August; published online 7 October 2007; doi:10.1038/ng.2007.2

may therefore require global strategies aimed at modeling the functional interrelationships between genes and/or proteins (genes/proteins) as complex interdependent networks[3].

Here we propose a complementary approach to systematic resequencing efforts[4–6] for identifying cancer genes and/or proteins. This approach is based on global network modeling of functional associations between potential cancer genes and their products.

RESULTS

A network modeling strategy

Macromolecular networks can be modeled on the basis of both global correlations observed among transcriptional profiling compendia, protein-protein interaction or 'interactome' networks, and genome-wide phenotypic profiling data sets[7], and comparisons of 'interolog' data sets from different organisms[8]. We combined these two strategies to generate models of macromolecular networks that are possibly perturbed in human cancer, starting from four known breast cancer–associated genes and their products—*BRCA1* and *BRCA2* (both identified by high-penetrance mutations)[9–11] and *ATM* and *CHEK2* (both identified by low-penetrance mutations)[12,13]—referred to hereafter as the 'reference genes/proteins' (**Fig. 1**). The strategy first integrates coexpression profiles in human tissues and then integrates functional associations derived from various functional genomic and proteomic, or 'omic', data sets obtained in both humans and model organisms. This integrated network modeling strategy provides a ranking system to classify potential network components from low to high likelihood; the components are then functionally and genetically tested.

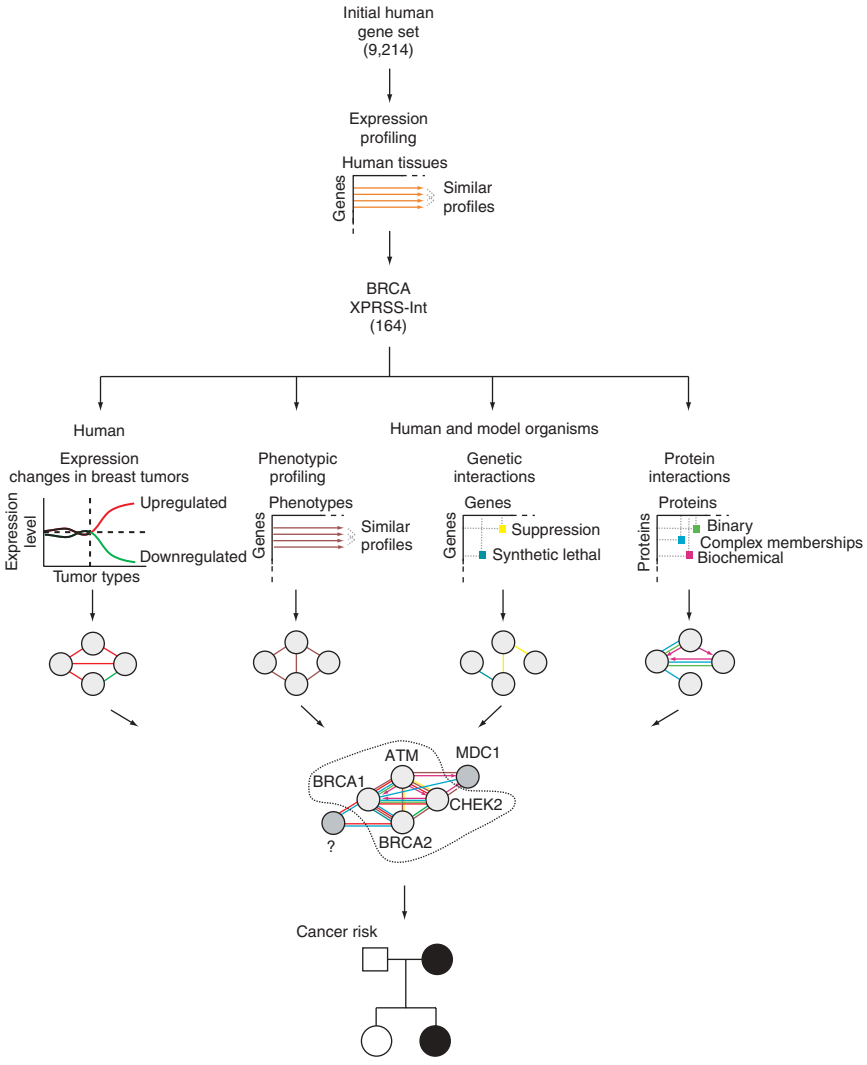

Figure 1 Outline for the generation of a BCN model.

Coexpression profiling

To initiate our search for genes that are likely to be functionally associated with the reference genes/proteins, we used a data set containing transcript abundance measurements for 9,214 human genes in 101 samples originating from 43 healthy human tissues or organs and three cell lines[14]. We determined Pearson correlation coefficient (PCC) values between each of the four reference genes and all of the genes tested on the array. To determine the likelihood of predicting functional associations by this coexpression approach, we curated published data from the scientific literature (until 1 October 2004) on protein interactions involving any of the four reference human proteins. The resulting 'literature interaction' (LIT-Int) network contains 103 proteins and 129 functional associations (**Fig. 2a** and **Supplementary Table 1** online). We determined that a PCC value of > 0.4 captures 36% of the LIT-Int functional associations (**Fig. 2b**). Beyond this threshold, the LIT-Int pairs are enriched between two- and tenfold (depending on the control set used) in coexpressed pairs as compared with 10,000 gene pairs generated randomly from the same expression data set.

To identify potential functional associations involving all four reference genes, we focused on those transcripts found in the 'expression intersection' (XPRSS-Int) of the four coexpression sets. The XPRSS-Int contains 164 genes, of which 15 are also present in the LIT-Int data set (**Fig. 2c** and **Supplementary Table 2** online). To evaluate the significance of the XPRSS-Int set, we generated 200 randomly chosen sets of four genes, calculated their transcriptional PCCs with each gene in the array and measured their coexpression intersection by using PCC > 0.4 as the cutoff. This simulation showed that 88% (176/200) of the randomly generated sets do not overlap in coexpression at any level and that, at most, their intersection contains 20 genes (**Fig. 2d**). These results indicate that the identification of 164 genes in the XPRSS-Int set could not be expected by chance (empirical $P < 0.005$).

We demonstrated that XPRSS-Int genes are functionally related to the four reference genes by examining three types of shared characteristics. First, the XPRSS-Int set showed an enrichment of Gene Ontology (GO) terms also present in the annotations for the reference genes (**Supplementary Table 3** online). Second, evolutionary conservation of coexpression patterns was observed between orthologs of XPRSS-Int and reference genes, corresponding to 52 XPRSS-Int genes

(**Supplementary Fig. 1** online). Third, using an expression data set for the analysis of breast tumor cell lines treated with various agonists or antagonists of mammary cell growth and differentiation[15], we observed significant coexpression among 33 XPRSS-Int genes (**Supplementary Table 2**).

Expression changes in breast tumors

To evaluate further the functional significance of the XPRSS-Int set, we reasoned that many XPRSS-Int genes might show expression changes in breast tumors arising from mutations in one of the reference genes. We compared the expression of each XPRSS-Int gene in breast tumors from individuals with *BRCA1* germline mutations (*BRCA1*^{mut}) with that in 'sporadic' breast tumors[16] (typically germline wild type (*BRCA1*^{wt})). Of the 132 XPRSS-Int single-probe genes (see Methods), 50 were upregulated and 9 were downregulated, as compared with an average of 18 upregulated and 22 downregulated genes in randomly generated sets ($P < 0.01$; **Fig. 3**, top right, and **Supplementary Table 4** online).

A total of 66 XPRSS-Int genes showing expression changes in *BRCA1*^{mut} tumors are shown as nodes in a coexpression network in which the length of the links, or 'edges', is inversely proportional to the expression correlation (PCC value) for each reference gene (**Fig. 3**,

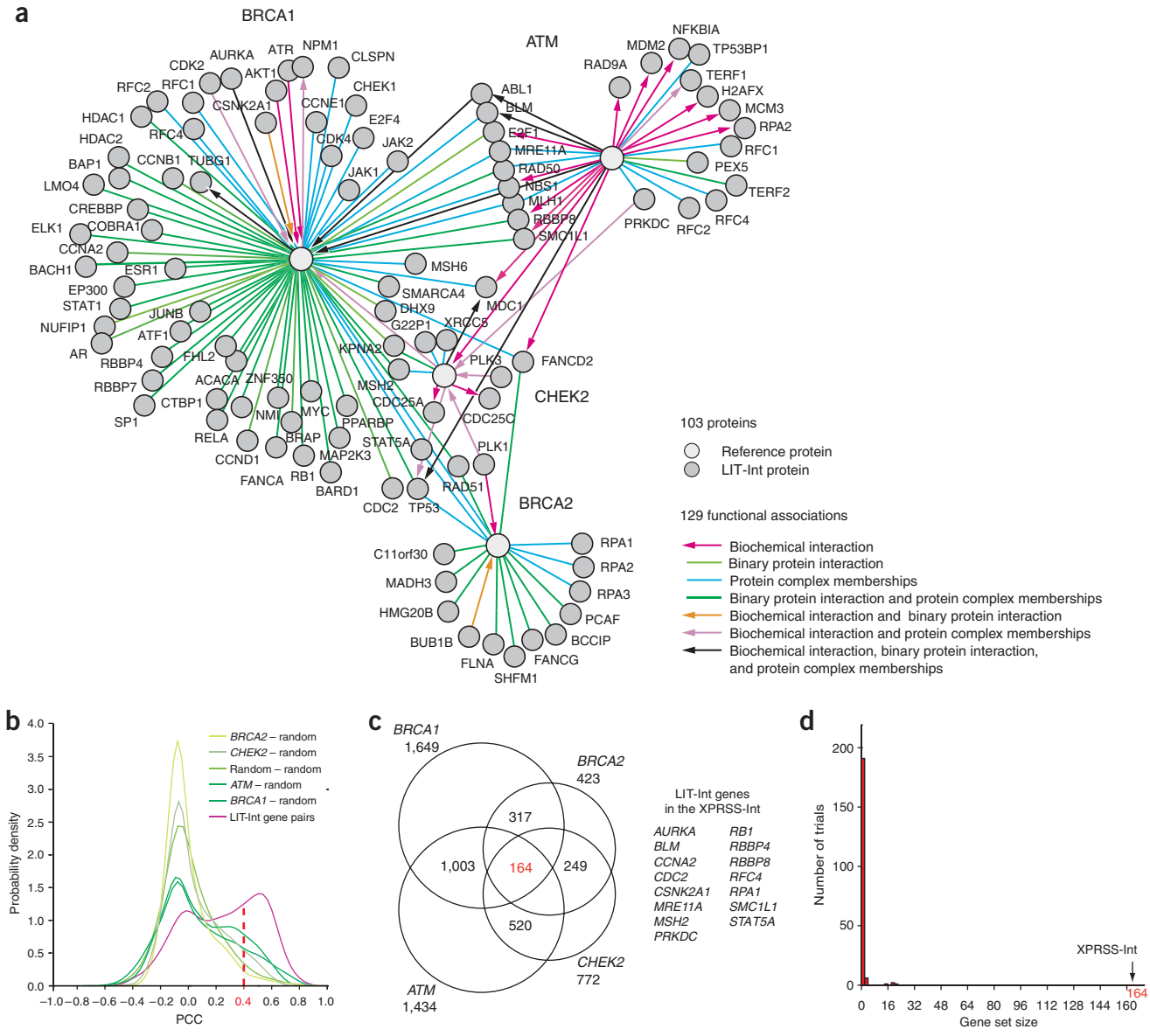

Figure 2 Generation of the XPRSS-Int data set. (**a**) LIT-Int network for ATM, BRCA1, BRCA2 and CHEK2. Proteins are represented by nodes and functional associations by edges, as indicated in the insets (**Supplementary Table 1** and **Supplementary Methods**). Multiple functional associations are indicated by blended line colors. (**b**) Probability density distributions of transcriptional PCC values between gene pairs for each of the four reference genes and randomly selected genes through 10,000 iterations (*BRCA* gene – random), for randomly selected gene pairs through 10,000 iterations (random – random), and for gene pairs from the LIT-Int data set. (**c**) Expression intersection (XPRSS-Int) of the four reference coexpression sets using PCC > 0.4. LIT-Int genes included in the XPRSS-Int set are shown. (**d**) Distribution of the coexpression intersection for randomly chosen sets of four genes and comparison with the XPRSS-Int set using PCC > 0.4.

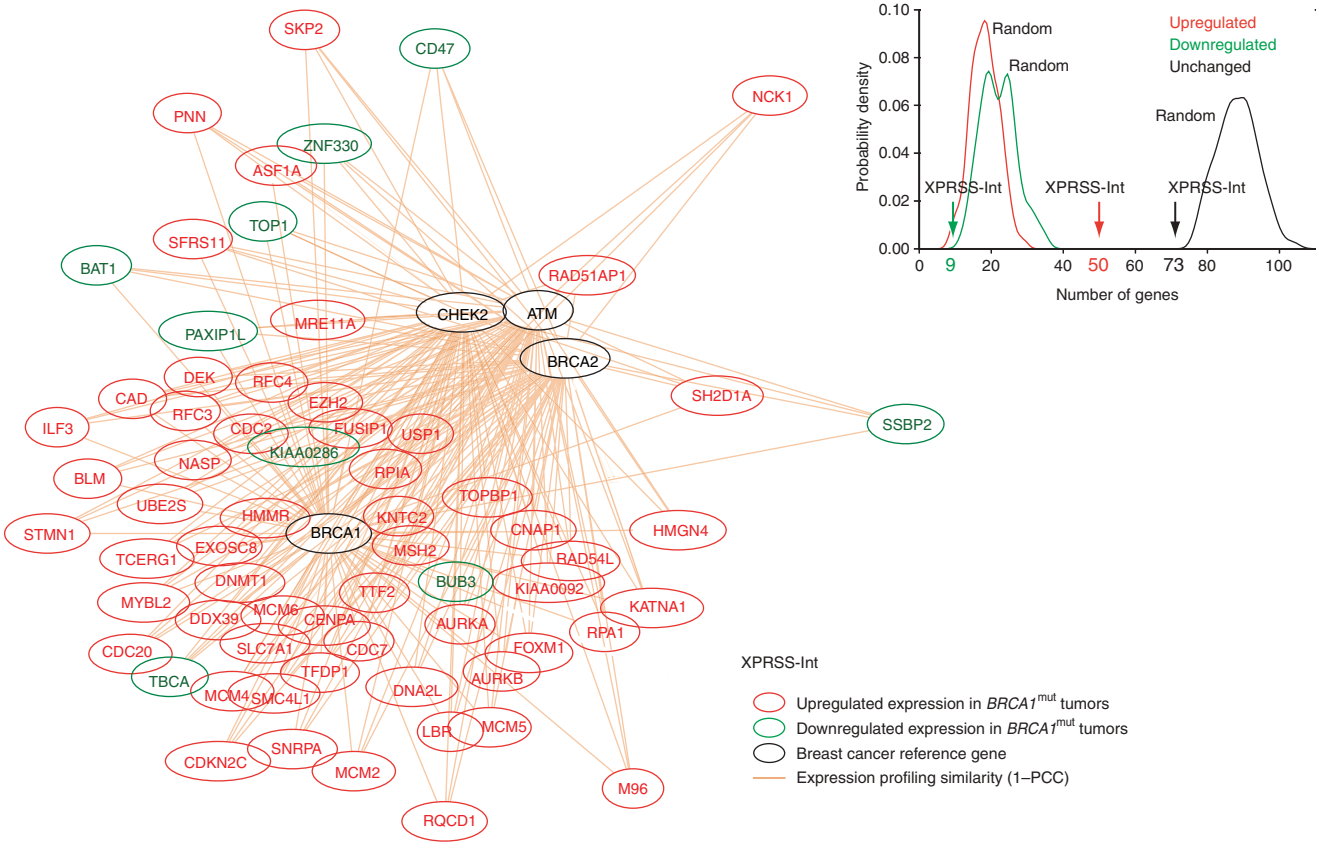

Figure 3 Expression analysis of the XPRSS-Int data set in breast tumors. Top right, comparison of gene expression in *BRCA1*^{mut} breast tumors relative to sporadic breast tumors both for genes in the XPRSS-Int set (vertical lines) and for 100 randomly generated equivalent sets of genes (curves; see Methods). Left, network representation of genes in the XPRSS-Int set that show expression changes in *BRCA1*^{mut} breast tumors relative to sporadic breast tumors. Edge distances to the four reference genes are optimized to be inversely proportional to their average PCCs across normal tissues (in other words, shorter edges indicate higher coexpression values).

left). We observed that the XPRSS-Int genes that were upregulated in *BRCA1*^{mut} breast tumors showed expression profiles more similar to that of *BRCA1* in healthy tissues than did those that did not show expression level changes between *BRCA1*^{mut} and sporadic tumors (two-tailed Student's *t*-test, $P = 0.039$). This observation suggests that XPRSS-Int genes that show expression changes in *BRCA1*^{mut} breast tumors are more likely to be functionally related to *BRCA1*.

A BRCA-centered network model

We investigated functional associations between each of the 164 XPRSS-Int genes/proteins and the reference genes/proteins. We systematically integrated functional associations found in largely non-overlapping omic data sets by using interologous functional relationships from four species (see Methods). The resulting BRCA-centered network (BCN) model consists of 118 genes/proteins (114 XPRSS-Int genes/proteins plus the four reference genes/proteins) and 866 potential functional associations (321 direct 'one-hop', and 545 indirect 'two-hop' associations), represented as nodes and edges, respectively (**Fig. 4a** and **Supplementary Table 5** online).

Interactome network models that are currently available for model organisms show relatively high clustering coefficients and particularly high connectivity between the members of protein complexes or 'molecular machines' involved in specific biological processes[17]. Evaluation of BCN connectivity with 1,000 randomly generated networks of sets of 164 genes from the same expression data set used for the identification of the XPRSS-Int showed that there are significantly more connected nodes and more edges in the BCN ($P < 0.001$; **Fig. 4b** and **Supplementary Fig. 2** online). This finding indicates that BCN components may function in biologically related pathways. Of all the potential functional BCN associations, those that are only in the literature-curated set constitute a limited and clustered portion. In addition, each omic approach generates its own characteristic functional clusters (**Fig. 4c**), further illustrating the need for data integration to generate higher quality network models.

To rank XPRSS-Int genes/proteins according to their possible functional association with the four reference genes/proteins, we combined the five criteria described above (enrichment in GO terms, conserved coexpression across species, coexpression in breast tumor–derived cell lines, changes in expression in *BRCA1*^{mut} breast tumors, and omic functional association in human or model organisms) in a matrix format, in which the number of matched criteria is indicated for each of the XPRSS-Int genes/proteins (**Fig. 4d**). We color-coded each of the five criteria to indicate the strength of coexpression with *BRCA1*. The percentages of LIT-Int genes within each group in the ranking correlate with the number of criteria (66, 38, 21, 0, 4 and 0% for matching all five to matching none of the criteria, respectively), which suggests that the likelihood that newly defined BCN components are functionally associated with any of the reference genes/proteins increases with the number of criteria matched. Observations made available in the literature since the generation of

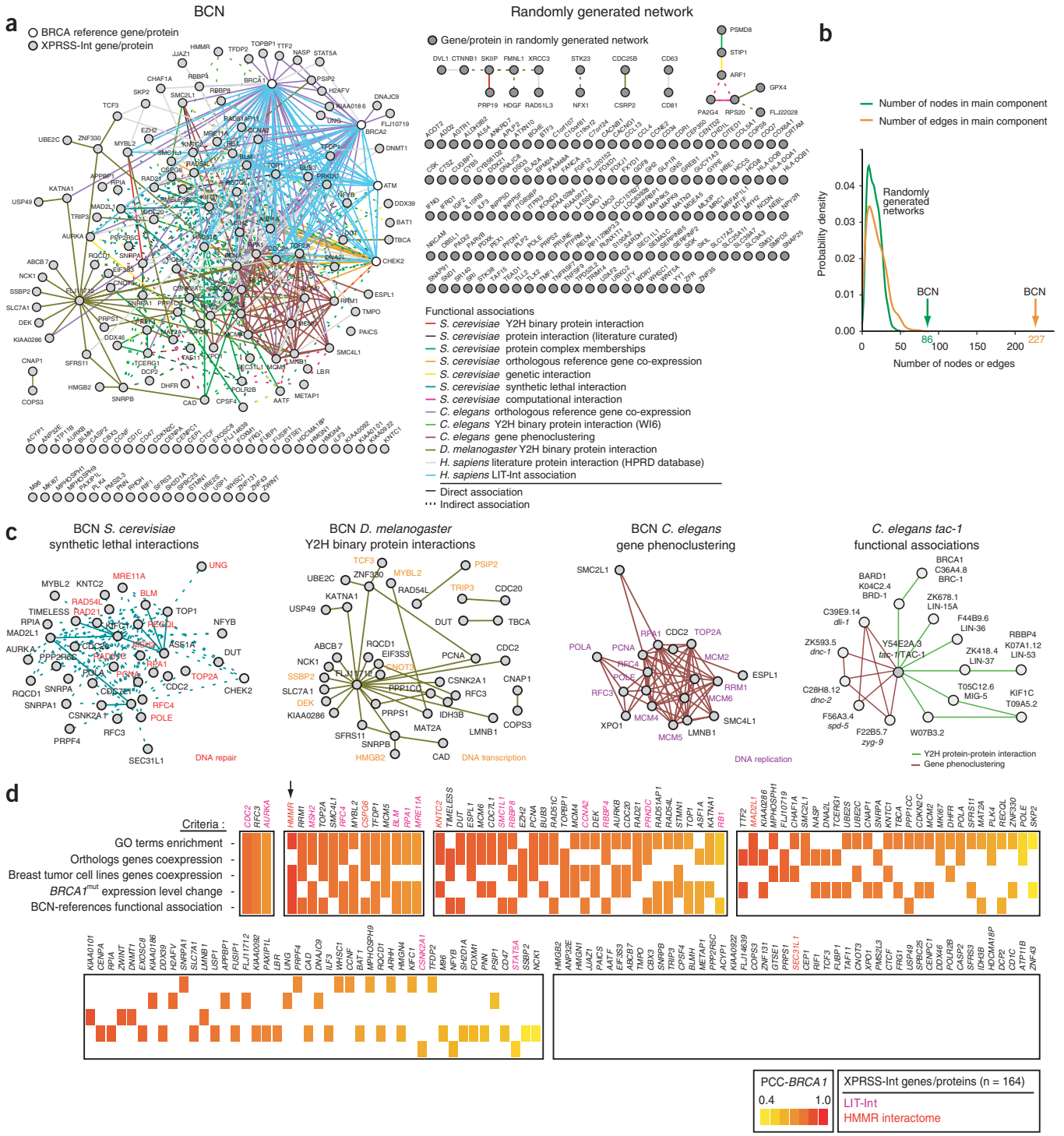

Figure 4 Generation of the BCN and ranking of XPRSS-Int genes/proteins. (**a**) Left, 13 omic data sets were integrated into the XPRSS-Int data set to generate the BCN. Edges represent direct 'one-hop' (unbroken lines) or indirect 'two-hop' (corresponding to two XPRSS-Int genes/proteins connected through a non–XPRSS-Int gene/protein; broken lines) functional associations. Right, example of a network that was randomly generated with the algorithm in **b**. Y2H, yeast two-hybrid. (**b**) Number of nodes (genes/proteins) and edges (functional associations) included in the main giant component of networks randomly generated through 1,000 iterations (curves) and in the BCN (vertical lanes). To avoid the biases of an a priori selection of references and subjective data compilation, the four human reference genes/proteins and all edges connecting them in the BCN, in addition to the *C. elegans* phenotypic profiling data, were excluded from this analysis. (**c**) Three left panels, GO terms annotations reveal functional clusters contained in distinct omic data sets used to generate the BCN. Right panel, functional associations of the *C. elegans tac-1* gene and TAC-1 protein with connections to BCN genes/proteins. (**d**) Five criteria were integrated to estimate the overall functional significance of XPRSS-Int genes/proteins relative to breast cancer reference genes/proteins. XPRSS-Int genes are clustered according to the number of criteria they match (from 5 to 0) and then ordered within clusters according to their average PCC value for *BRCA1* (PCC-*BRCA1*). LIT-Int genes are shown in purple. Genes shown in red encode proteins that are also present in the HMMR-centered interactome (**Fig. 5a**).

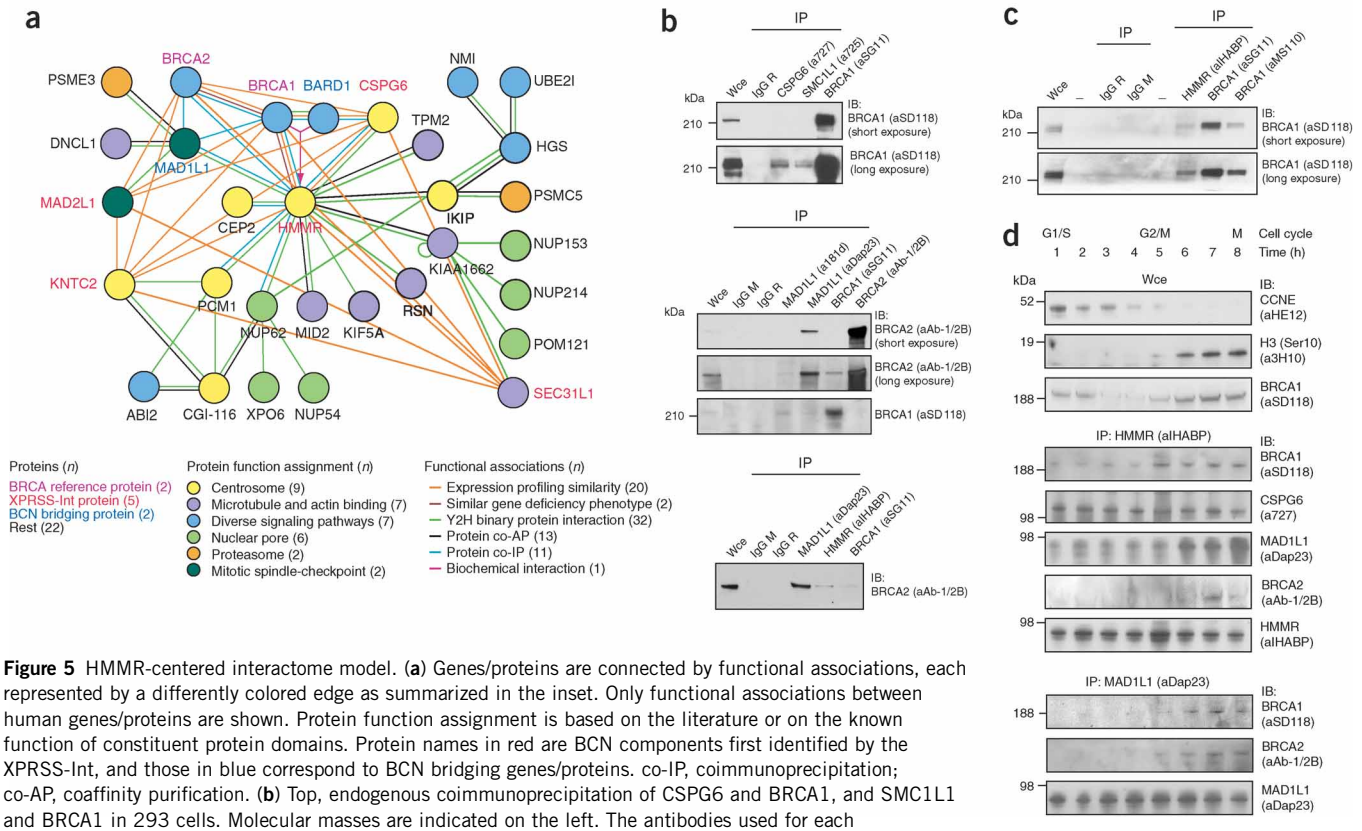

Figure 5 HMMR-centered interactome model. (**a**) Genes/proteins are connected by functional associations, each represented by a differently colored edge as summarized in the inset. Only functional associations between human genes/proteins are shown. Protein function assignment is based on the literature or on the known function of constituent protein domains. Protein names in red are BCN components first identified by the XPRSS-Int, and those in blue correspond to BCN bridging genes/proteins. co-IP, coimmunoprecipitation; co-AP, coaffinity purification. (**b**) Top, endogenous coimmunoprecipitation of CSPG6 and BRCA1, and SMC1L1 and BRCA1 in 293 cells. Molecular masses are indicated on the left. The antibodies used for each coimmunoprecipitation experiment are given in parentheses. IB, immunoblot; Wce, whole-cell extract; IgG, normal purified immunoglobulin negative controls (G, goat; M, mouse; R, rabbit). Bottom, endogenous coimmunoprecipitation of MAD1L1 and BRCA1 or BRCA2, and HMMR and BRCA1 or BRCA2 in U2OS cells. (**c**) Endogenous coimmunoprecipitation of HMMR and BRCA1 in non-synchronized HeLa S3 cells. (**d**) Cell cycle synchronization and coimmunoprecipitation assays in HeLa S3 cells.

the LIT-Int data set (1 October 2004) support the hypothesis that the BCN points to numerous functional associations with known breast cancer genes/proteins. Since we generated LIT-Int, 20 proteins have been newly described to have a functional relationship with at least one of the reference genes/proteins, and 4 of these proteins (20%; detailed in **Supplementary Table 2**) are present in the XPRSS-Int (~10-fold increase relative to chance); this percentage is similar to that determined above for LIT-Int proteins in the XPRSS-Int data set.

To investigate predictions based on the BCN model, we looked for new potential connections between different biological processes. The highest-ranking connection involves a *TACC* family member[18], *HMMR*, which encodes the hyaluronan-mediated motility receptor[19] (HMMR, also known as RHAMM). *HMMR* has the highest PCC coexpression value relative to *BRCA1* (0.90; **Fig. 3**). In itself, the high PCC of *HMMR* with *BRCA1* might be a substantial indicator of a functional association between the two genes or gene products; however, it is the integration of other types of relationships that points to *HMMR* beyond the possible false-positive discovery rate obtained by profiling similarity analysis. The HMMR gene product connects BRCA1 and RPPB4 (also known as RBAP48, pRb and BRCA1-interacting protein) by binary physical interactions in the BCN (**Fig. 4c**, right), and it matches four of the five functional significance criteria (**Fig. 4d**). A protein-protein physical interaction identified by yeast two-hybrid screening predicts an association between the TAC-1 and BRD-1 *Caenorhabditis elegans* proteins, orthologs of TACCs and BARD1, respectively (refs. 20,21). Clues to the function of TACCs also come from the phenotypic profiling of *C. elegans* genes[17], which identifies a phenocluster of genes (**Fig. 4c**, right) whose inactivation is linked to microtubule-instability phenotypes, particularly centrosome dysfunction[22–24]. Observations in human cells also indicate that HMMR may have a potential role in centrosomal functions[18,25]. Considered together, these observations suggest that *HMMR* plays a role in centrosome function in conjunction with *BRCA1*.

New BRCA functional associations

Generation of an HMMR-centered interactome map through yeast two-hybrid screens identified a network of 31 proteins and interactions between them (**Fig. 5a**). Several of these interactions were validated through coaffinity purification of exogenously expressed fusion proteins and coimmunoprecipitation of endogenous proteins (**Supplementary Figs. 3–5** online). Notably, among interactors of HMMR, CSPG6 cohesin and MAD1L1 mitotic spindle-assembly checkpoint protein are also present in the BCN, further illustrating how the BCN is a source of testable associations with breast reference genes/proteins, and the endogenous CSPG6 and BRCA1 proteins form a complex in 293 cells (**Fig. 5b**). In addition to the direct association between MAD1L1 and HMMR, we also identified MAD1L1-BRCA1 endogenous protein complexes (**Fig. 5b**). The identification of several interactors of HMMR as members of BRCA1 protein complexes suggests that HMMR is a genuine component of a BRCA1 functional module involved in centrosomal function.

Consistent with the indirect evidence described above, HMMR and BRCA1 were found to associate in endogenous protein complexes in

several cell lines (293, HeLa (**Fig. 5c**) and U2OS cell lines) tested. Reciprocal coimmunoprecipitation experiments were hampered by unspecific binding of HMMR to secondary antibodies. Because the HMMR-MAD1L1 protein interaction was detectable most strongly during M phase, and because the HMMR-BRCA1 protein complex association seemed to be greater during the G2/M and M phases, we examined whether a MAD1L1-BRCA1 association also occurs in M phase. MAD1L1 and BRCA1 were found together in endogenous protein complexes as cells entered mitosis (**Fig. 5d**). Consistent with the participation of BRCA1 and BRCA2 in the regulation of mitotic progression[26,27], we found that BRCA2 was also present in HMMR and MAD1L1 mitotic complexes (**Fig. 5b**,**c**). Notably, *HMMR* was previously identified among other proliferation-associated genes to be tightly coexpressed with *BRCA2* in breast carcinomas[28]. We did not observe such cell cycle–dependent protein associations for CSPG6, which suggests that the HMMR-BRCA1 and CSPG6-BRCA1 complexes have different cellular outcomes. In summary, by experimentally testing BCN predictions, we identified three previously unknown protein associations with breast cancer tumor suppressors, thereby deriving a centrosomal interactome model that has several other candidates, some of which are also found in the BCN (KNTC2 (also known as HEC1), MAD2L1 and SEC31L1).

Because ubiquitination has an important role in the regulation of centrosome duplication[29], both BRCA1 and HMMR have been localized to centrosomes and mitotic spindle poles[18,26], and HMMR-BRCA1 protein complexes are most prominent during G2/M, we examined the possibility that HMMR might be a substrate of the BRCA1-BARD1 E3 ubiquitin ligase activity[30]. HMMR was found to be efficiently polyubiquitinated by the BRCA1-BARD1 heterodimer *in vitro* (**Fig. 6a**, left). *In vivo* ubiquitination of HMMR was then observed after co-transfection of glutathione *S*-transferase (GST)-conjugated HMMR and hemagglutinin (HA)-conjugated ubiquitin into 293 cells (**Fig. 6a**, right). These observations suggest that HMMR has a role in centrosomal function mediated by BRCA1-BARD1 polyubiquitination.

HMMR-BRCA1 and centrosome dysfunction

To understand better the functional association between HMMR and BRCA1 *in vivo*, we examined phenotypes mediated by short interfering RNA (siRNA). In agreement with changes observed in *HMMR* expression in *BRCA1*^{mut} tumors relative to sporadic breast tumors, depletion of BRCA1 resulted in an upregulation of *HMMR* and its product in immortalized human mammary epithelial cells (HME/TERT cells; **Supplementary Fig. 6** online). Centrosome hypertrophy and amplification were observed in cells depleted of either BRCA1 or HMMR (**Fig. 6b** and **Supplementary Figs. 7** and **8** online). These effects on centrosome numbers are similar to those previously observed on functional inhibition or depletion of BRCA1 (ref. 31). The breast tumor–derived cell line Hs578T and the HME/TERT cells showed higher levels of centrosome amplification on depletion of HMMR or BRCA1 than either the HeLa or U2OS cell lines. Notably, a genetic interaction between *HMMR* and *BRCA1* was observed in HME/TERT and Hs578T cells. In these cells, simultaneous depletion of *HMMR* and *BRCA1* suppressed the centrosome abnormality phenotypes observed on depletion of either single transcript (**Fig. 6c**).

To investigate further the relationship between HMMR, BRCA1 and centrosome dysfunction, we examined the role of aurora-A kinase (AURKA, ranked third in the XPRSS-Int) because, first, its *C. elegans* ortholog is required for centrosomal localization of TAC-1 (refs. 22–24); second, it associates with BRCA1 in the regulation of the G2/M transition[32]; and third, when overexpressed in the mammary epithelium, it induces centrosome amplification and genetic instability before tumorigenesis[33]. Consistent with an association in mitosis, endogenous AURKA-HMMR protein complexes were identified in greater amounts in nocodazole-arrested cells (**Fig. 6d**). Immunofluorescence analysis of HMMR in *BRCA1*^{mut} cells showed mislocalization relative to γ-tubulin (TUBG1); this mislocalization coincided with that of AURKA, as revealed by multiple foci, some showing detectable HMMR and others not (**Fig. 6e** and **Supplementary Fig. 9** online). No mislocalization of HMMR or AURKA on depletion of BRCA1 by siRNA was observed in the cell lines studied, a finding that could be due to partial BRCA1 depletion or, alternatively, to the fact that HMMR mislocalization results from an indirect effect mediated by overexpression of AURKA. As shown above, *AURKA* and *HMMR* expression is upregulated in *BRCA1*^{mut} breast tumors relative to sporadic breast tumors (**Fig. 3**) and AURKA and HMMR seem to be overexpressed and mislocalized in *BRCA1*^{mut} cell lines (**Supplementary Fig. 9**). Overexpression of *AURKA* in HME/TERT (*BRCA1*^{wt}) cells was sufficient to cause the overexpression and mislocalization of HMMR (**Fig. 6e**). These observations, together with the genetic interaction identified between *HMMR* and *BRCA1*, suggest that overexpression of HMMR and/or its biochemical modification resulting in centrosome amplification, which have been previously associated in myeloma[25], are early somatic molecular events that contribute to breast tumorigenesis.

HMMR and breast cancer risk

The genotyping of three *HMMR* haplotype-tagging SNPs (htSNPs) in 923 individually matched case-control pairs from a population-based study of incident breast cancer in northern Israel identified statistically significant associations for each htSNP (**Supplementary Table 6** online). All three individual htSNPs were associated with risk of breast cancer and were consistent with an additive model showing increasing risk with each additional risk-tagging allele. For example, each copy of the A allele of rs10515860 was associated with a 32% increase in risk (odds ratio (OR) = 1.32, 95% confidence interval (CI) 1.09–1.60; $P = 0.004$). Each copy of the A-C-A haplotype (rs7712023-rs299290-rs10515860) estimated by the expectation maximization method increased a woman's risk by 33% (unadjusted OR = 1.33 for an additive model, 95% CI 1.08–1.63; $P = 0.007$). The single intronic htSNP rs10515860 essentially captured all of the variation associated with risk, which suggests that the A allele of rs10515860 is likely to be in linkage disequilibrium with a variant that confers risk of breast cancer.

Further study of this allele showed that the risk of breast cancer associated with the *HMMR* locus was not accounted for by the presence of *BRCA1* or *BRCA2* mutations (data not shown). Consistent with models of inherited susceptibility to breast cancer[34], rs10515860 A allele carriers of age 40 years and over were 1.22 times as likely to develop breast cancer than were controls (OR = 1.22, 95% CI 1.02–1.46; $P = 0.026$), whereas carriers younger than 40 years were 2.7 times as likely to develop breast cancer (OR = 2.73, 95% CI 1.25–5.97; $P = 0.012$). Risk was not restricted to Ashkenazi Jews and was similar in magnitude and frequency for Sephardi Jews and Christian and Muslim Arabs, in addition to Bedouin, Druze and Cherkazi populations.

We investigated this association in an independent Ashkenazi Jewish cohort, ascertained at Memorial Hospital in New York, comprising 247 individuals with breast cancer and a family history of three or more cases of breast cancer but no identified *BRCA1* or *BRCA2* mutation and 298 women with no personal or family history of breast cancer[35]. Genotyping SNPs across the complete 5q34 region

identified associations with P values lower than 10^{-3} for the rs1458338 and rs12658549 SNPs located proximally to the *HMMR* locus (**Supplementary Table 7** online). Association with these SNPs was also identified in the Ashkenazi subgroup of the Israeli case-control study (rs1458338: OR = 2.87, 95% CI 1.06–7.79; rs12658549: OR = 2.73, 95% CI 1.09–6.85; P = 0.030 and 0.032, respectively), which suggested that these SNPs are in linkage disequilibrium with genetic variation 5′ upstream of the *HMMR* coding region in a 'gene desert' region spanning ~1 Mb proximally. The G-C-A-T-G haplotype was also significantly associated with risk of breast cancer (**Supplementary Table 8** online), which suggests that there may be two discrete risk variants that can be ascertained on different haplotypes.

To confirm the association between genetic variation in *HMMR* and early-onset breast cancer, we genotyped rs10515860 in a third independent sample of 997 cases and 464 controls (212 controls with age data available) derived from the New York Cancer Project[35]. We found a significant association of a 12-month-earlier age of onset in homozygous cases as compared with controls in this cohort (OR = 1.56, 95% CI 1.02–1.95; P = 0.009), consistent with the

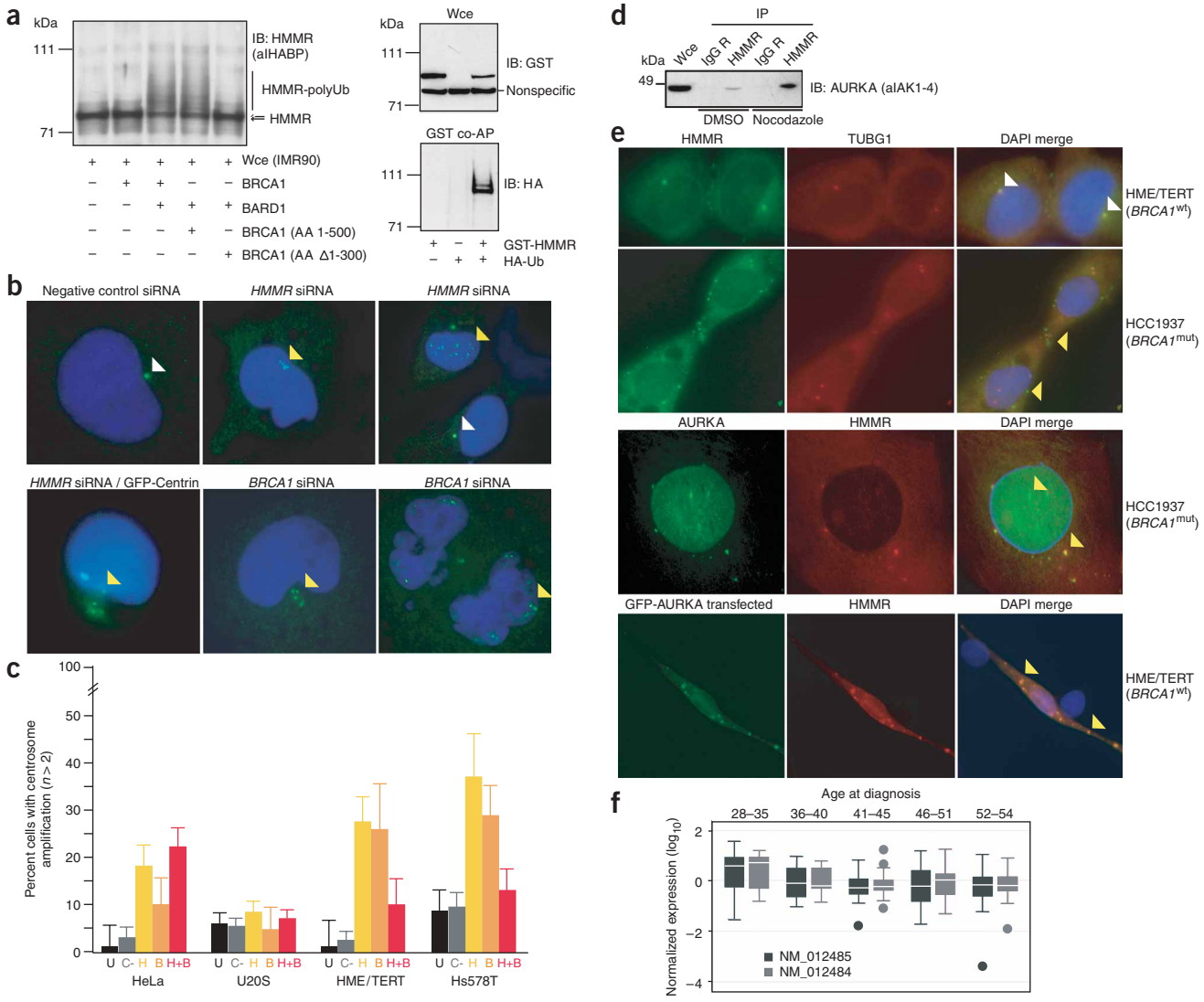

Figure 6 BRCA1-BARD1–mediated polyubiquitination, HMMR-BRCA1 and HMMR-AURKA interactions and centrosome dysfunction. (**a**) Left, HMMR *in vitro* polyubiquitination analysis. Right, detection of polyubiquitinated HMMR by co-transfection of GST-HMMR and HA-ubiquitin into 293 cells. (**b**) Immunofluorescence microscopy results for γ-tubulin (TUBG1) and GFP-centrin are shown in Hs578T cells transfected with different siRNAs. White arrowheads indicate normal centrosomes; yellows arrowheads indicate centrosome amplification or hypertrophy. (**c**) Percentage of cells showing centrosome amplification in HeLa, U2OS, HME/TERT and Hs578T cells. U, untreated cells; C-, negative control siRNA; H, *HMMR* siRNA; B, *BRCA1* siRNA; H+B, *HMMR* plus *BRCA1* siRNAs. Immunofluorescence microscopy results and evaluation of siRNA-mediated depletion are provided in **Supplementary Figures 7** and **8**. (**d**) Endogenous coimmunoprecipitation of HMMR and AURKA in U2OS cells treated with DMSO or nocodazole. (**e**) Immunofluorescence microscopy results for HMMR, TUBG1, AURKA, GFP-AURKA and DAPI counterstaining in HME/TERT (*BRCA1*^{wt}) and HCC1937 (*BRCA1*^{mut}) cells. Yellow arrowheads indicate protein cellular mislocalization. A cell overexpressing GFP-AURKA shows more endogenously expressed HMMR than neighbor cells that do not express GFP-AURKA, identified by DAPI staining. Normal colocalization of HMMR and AURKA in HME/TERT cells and analysis of additional *BRCA1*^{mut} cell lines are shown in **Supplementary Figure 9**. (**f**) Early age at diagnosis is associated with higher *HMMR* expression in primary breast tumors of individuals without *BRCA1* or *BRCA2* mutations, as assessed with a published data set[16] (P = 0.02 and 0.07 for microarray probes NM_012484 and NM_012485, respectively).

Table 1 Association results for rs10515860 and risk of breast cancer

Israel and New York data

	Israel			New York			Total data combined			
Genotypes	Controls	Cases	OR (95% CI)	Controls	Cases	OR (95% CI)	Controls	Cases	OR (95% CI)	*P* value[a]
G/G	1,211	1,173	1.00	375	794	1.00	1,586	1,967	1.00	
G/A	230	288	1.29 (1.07–1.46)	85	189	1.06 (0.80–1.41)	315	477	1.23 (1.05–1.43)	
A/A	12	17	1.46 (0.70–3.08)	5	14	1.28 (0.47–3.57)	17	31	1.41 (0.78–2.50)	0.007

Combined Israel and New York data by age cohort

	Age ≤ 40				Age > 40			
Genotypes	Controls	Cases	OR (95% CI)	*P* value[a]	Controls	Cases	OR (95% CI)	*P* value[a]
G/G	101	177	1.00		1,276	1,790	1.00	
G/A	15	50	2.01 (1.08–3.75)		257	427	1.18 (1.01–1.40)	
A/A	0	3	—	0.026	16	28	1.21 (0.65–2.24)	0.044

[a]*P* values for the additive model (trend test).

findings in the case-control study from Israel. Combined analysis for all 2,475 cases and 1,918 controls gave an OR of 1.23 (95% CI 1.05–1.43; $P = 0.007$) for each copy of the rs10515860 risk allele (**Table 1**). Combined analysis by age cohort showed greater risk for breast cancer for this htSNP in women younger than 40 years (OR = 2.01, 95% CI 1.08–3.75) as compared with older women (OR = 1.18, 95% CI 1.01–1.40), although both associations were significant (**Table 1**).

Given the association that extended proximally to the *HMMR* coding region in these three studies, we investigated differences in germline expression by using HapMap genotypes and paired expression data. Consistent with the results outlined above for risk of breast cancer, the rs10515860 SNP showed an association with *HMMR* overexpression under an additive model ($\log_2$ units difference = 0.35, 95% CI 0.18–0.53; $P = 0.0001$) for European descendents. The A-C-A haplotype showed an association with an increase in 0.32 $\log_2$ units in *HMMR* expression for European descendents and for all HapMap individuals included in the data set GSE6534 (**Table 2**). The same trend was observed for European descendents when an independent expression data set (GSE5859; $\log_2$ units difference = 0.21, 95% CI −0.03 to 0.44; $P = 0.085$) was used. These differences in expression for genotypes and haplotypes agree with a model in which *HMMR* overexpression is both an early molecular mechanism promoting breast tumorigenesis and a risk factor for breast cancer. The G-C-A-T-G haplotype that is also associated with risk of breast cancer correlates with downregulation of *HMMR* (**Table 2**), reinforcing the hypothesis that there is tight germline regulation of *HMMR* expression and suggesting that any alteration in HMMR protein quantities could promote tumorigenesis, as previously discussed for TACC proteins[36].

To examine further this model of *HMMR* expression changes and risk of breast cancer, we investigated how *HMMR* expression in primary breast tumors from individuals without *BRCA1* or *BRCA2* mutations is associated with age at diagnosis. Linear regression analysis showed that higher *HMMR* expression is associated with early age at diagnosis (**Fig. 6f**), in agreement with the idea that overexpression of *HMMR* is a risk factor for breast tumorigenesis.

Table 2 Haplotypes association with *HMMR* germline expression level

HapMap, European ancestry ($n = 60$)[a]

Haplotypes	rs1458338	rs12658549	rs7712023	rs299290	rs10515860	Frequency	Difference (95% CI)	*P* value
1	A	T	T	T	G	0.525	—	—
2	A	T	A	T	G	0.233	0.19 (−0.03/0.41)	0.090
3	A	T	A	C	G	0.133	0.04 (−0.21/0.30)	0.722
4	A	T	A	C	A	0.109	0.32 (0.02/0.62)	0.037

HapMap, all populations ($n = 210$)[b]

1	A	T	T	T	G	0.277	—	—
2	A	T	A	T	G	0.251	0.01 (−0.10/0.13)	0.822
3	A	T	A	C	G	0.206	−0.05 (−0.17/0.07)	0.399
4	G	C	A	T	G	0.067	−0.39 (−0.57/−0.21)	3×10^{-5}
5	G	C	A	C	G	0.055	0.10 (−0.08/0.29)	0.271
6	A	T	A	C	A	0.053	0.32 (0.12/0.51)	2×10^{-3}
7	G	T	A	C	G	0.039	−0.16 (−0.42/0.09)	0.217
8	A	C	A	C	G	0.018	0.06 (−0.22/0.33)	0.689
9	G	T	A	T	G	0.015	−0.29 (−1.05/0.47)	0.452

[a]Global haplotype association $P = 0.04$. [b]Global haplotype association $P = 0.92$. Rare estimated haplotypes ($n = 7$, cumulative frequency = 0.019) not shown.

We cannot exclude the possibility that linkage disequilibrium of genetic variation independent of *HMMR* might potentially explain the observed associations. However, the strength of the genetic associations, coupled with the functional associations that we have found, indicates that *HMMR* may be a previously unknown susceptibility gene for breast cancer within diverse human populations.

DISCUSSION

We have shown that analysis of gene expression, combined with integration of omic data sets from various species, can be used to generate a network of hundreds of potential functional associations with cancer genes/proteins. Many of the hypotheses provided by the BCN have been tested and validated either here or in the literature. Additional omic data sets can be integrated as they are produced to generate a larger and more comprehensive BCN.

Among the newly defined associations predicted in the BCN, we focused on those of HMMR on the basis of a combination of different functional criteria. Network modeling also points to functional associations for other genes/proteins with properties similar to those of HMMR, such as CSPG6. HMMR and two of its interactors (CSPG6 and MAD1L1) associate in protein complexes with BRCA1 and BRCA2. We established that HMMR is an *in vitro* substrate for BRCA1-BARD1–mediated polyubiquitination and that *BRCA1* and *HMMR* genetically interact to control centrosome number in breast tumor– and mammary epithelium–derived cell lines. In addition, we identified an association in breast tumorigenesis between BRCA1 and HMMR with AURKA, which is also in the BCN. Collectively, these findings describe a role for HMMR together with BRCA1 and AURKA in microtubule-based processes essential for proper chromosome segregation. Observations made in *Xenopus laevis* have associated the HMMR ortholog XRHAMM with the control of spindle-pole assembly mediated by the BRCA1-BARD1 heterodimer[37]. This observation further supports the idea that the HMMR-centered interactome network described here has a role in genomic stability and breast tumorigenesis.

Genetic analysis supports *HMMR* as a newly defined breast cancer susceptibility gene, thereby delineating a genetic link between risk of breast cancer and centrosome dysfunction. Notably, centrosome hypertrophy and amplification are often evident in early stage proliferative breast lesions[38,39], and loss of mouse *Brca1* or *Brca2* leads to centrosome amplification and genomic instability[40,41]. Our results suggest that *HMMR* can act as a breast cancer susceptibility gene when expressed in relatively large amounts; such overexpression probably leads to centrosome amplification and genomic instability, similar to what has been proposed in studies of an *Aurka* transgenic model[33] and in work involving transient *Mad2* overexpression[42]. Our network modeling strategy is applicable to other types of cancer; it will help to discover more cancer-associated genes and to generate a 'wiring diagram' of functional interactions between their products.

METHODS

Bioinformatic analyses. We obtained GO annotations from NetAffx (Affymetrix) and based enrichment calculations on the total number of distinct genes and their annotations in the U95A platform; *P* values were then determined by Fisher's exact test. Only grade 3 (poorly differentiated) tumors were used to study gene expression in $BRCA1^{mut}$ tumors relative to sporadic breast tumors; *P* values for differential expression were then determined by two-tailed Student's *t*-test ($P < 0.10$). To avoid probe discrepancies and differences in the number of data points, only single-probe genes were simulated and compared with single-probe genes from the XPRSS-Int data set, corresponding to 132 genes from a total of 164; *P* values were then calculated empirically by using 100 random simulations. The $BRCA1^{mut}$ coexpression network was generated with the Graphviz graph visualization package. Owing to constraints of the visualization program, the relative network positions of the four reference genes are not proportional to the PCCs between them. Transcriptional PCCs of XPRSS-Int genes that were upregulated and unchanged in $BRCA1^{mut}$ tumors were compared by a two-tailed Student's *t*-test, taking into account all U95A probes for each gene. Orthologs were defined by reciprocal BLASTP best hit ($P < 10^{-6}$) or from the literature.

For the BCN modeling, we integrated, first, gene expression profiling similarity above a given threshold, as determined by the PCC, from a compendium of *Saccharomyces cerevisiae*[43] and *C. elegans*[44] microarray profiles (6,174 and 18,451 genes, respectively); second, phenotypic similarity for 661 early embryogenesis *C. elegans* genes above a specific threshold[17]; third, genetic interactions for 1,347 *S. cerevisiae* genes[43]; and fourth, protein physical interactions (binary interactions, complex co-memberships and biochemical interactions (protein modification)) for 3,458 *S. cerevisiae*[43], 4,588 *C. elegans* (WI6 data set[21]), 7,198 *Drosophila melanogaster*[45,46] and 10,305 *Homo sapiens* proteins[47]. To rank XPRSS-Int genes/proteins, PCC average values for *BRCA1* were used and BCN links corresponding to LIT-Int interactions were not considered. Gene and protein names used throughout the text are detailed in **Supplementary Table 9** online.

Protein and biochemical analyses. Yeast two-hybrid methods were carried out as described[48]. We performed BRCA1-BARD1 heterodimer purification and ubiquitination reactions as described[31], using whole-cell extracts of IMR90 human fibroblasts. We used the following antibodies: polyclonal IHABP and E-19 (Santa Cruz Biotechnologies) for HMMR; monoclonal SD118, MS110 and SG11 for BRCA1; monoclonal Ab-1 (Oncogene) for BRCA2; monoclonal clone-53 (BD Transduction Laboratories) for NUP62; polyclonal anti-AIK (Cell Signaling) and monoclonal anti-IAK1 (BD Transduction Laboratories) for AURKA; and polyclonal 727 and 725 for CSPG6 and SMC1L1, respectively, polyclonal Dap23 and polyclonal 81d for MAD1L1, anti-PCM1 and anti-CEP2 (see Acknowledgments).

Cell culture and immunofluorescence microscopy. The negative control siRNA was Silencer Negative Control 1 siRNA (Ambion). The *HMMR* 1 and 2 siRNAs were Silencer Pre-Designed siRNAs (Ambion; 5′-GGUGCUUAU GAUGUUAAAATT-3′ and 5′-GGACCAGUAUCCUUUCAGATT-3′, respectively). The *BRCA1-b* siRNA has been published[31]. Centrosome immunofluorescence counts correspond to three independent experiments, each scoring 200 cells with >80% protein reduction of HMMR and/or BRCA1.

Association studies. Individuals with breast cancer and control subjects were recruited through an institutional review board–approved study of breast cancer in northern Israel and through protocols at the Memorial Sloan-Kettering Cancer Center in New York. The case-control study in northern Israel is a population-based study of incident breast cancer identified through rapid case ascertainment between January 2000 and July 2006. Controls were identified by randomly selecting women from a comprehensive list of insurees with Israel's largest health provider, which covers ~70% of women at risk in northern Israel. For each case, an individually matched control without breast cancer was identified by randomly sampling all female insurees meeting the matching criteria of age (within 1 year), ancestry (Jewish versus non-Jewish) and geographical clinic area. All breast cancers were pathologically confirmed, and each participant signed written, informed consent. Cases and controls were classified by religion and ancestry on the basis of self-report.

Genotyping. Genomic DNA derived from blood lymphocytes was used for all genotyping. All case and control samples were genotyped for *BRCA1* and *BRCA2* Jewish-derived founder mutations by using a TaqMan platform (Applied Biosystems). Genotyping for SNPs in *HMMR* was performed with Pyrosequencing (Biotage). Haplotype-tagging SNPs were selected for the northern Israel study using Haploview[49] to determine linkage disequilibrium blocks by analyzing Centre d' Etude du Polymorphisme Humain (CEPH) SNP genotype data downloaded from the HapMap website. To be included in the analysis, SNPs had to meet the following criteria: the Hardy-Weinberg χ^2 goodness-of-fit test yielded a *P* value > 0.05; at least 75% of subjects were genotyped for the SNP; and the minor allele frequency of the SNP was at least

0.001. For the SNPs that met these criteria, Haploview was run in aggressive tagging mode using two- and three-marker haplotypes. No difference in results was seen when the pairwise tagging mode was used. For a SNP to be selected, all alleles to be captured were correlated at a value of $r^2 > 0.8$ with the marker in that set. Three SNPs were identified as htSNPs, which tagged variation within three estimated haplotype blocks in *HMMR*: rs7712023, rs299290 and rs10515860. We performed duplicate genotyping for all samples in the population-based study for rs10515860 with 100% concordance. Duplicate genotyping for rs299290 was done for 1,789 samples, and 1,737 (97.1%) were in agreement. Duplicate genotyping was not performed in the validation study in New York.

Statistical analysis. Conditional and unconditional logistic regressions were used to analyze data from the case-control study, as appropriate for a matched study. No important differences between matched and unmatched analyses were noted; therefore, the results are presented for the unmatched analysis to optimize the sample size for tests of interaction and for ease of presentation. Haplotypes were estimated with the EM algorithm, as implemented in the R packages haplo.stats and SNPStats[50]. Association study between HapMap genotypes and haplotypes, and *HMMR* germline expression levels from two independent data sets (normalized values of Gene Expression Omnibus records GSE5859 and GSE6536) were performed with the haplo.stats package by fitting linear equations and *P* values obtained based on the *F*-test.

Additional methods. Detailed experimental methods are described in the **Supplementary Methods** online.

Note: Supplementary information is available on the Nature Genetics website.

ACKNOWLEDGMENTS

We thank members of our laboratories for discussion and comments on the manuscript; A. Merdes (Wellcome Trust Centre for Cell Biology) for anti-PCM1; B. Koch (Research Institute of Molecular Pathology) for anti-CSPG6 and anti-SMC1L1; K.-T. Jeang (National Institute of Allergy and Infectious Disease) for anti-MAD1L1 (Dap23); D.-Y. Jin (University Hong Kong) for anti-MAD1L1 (81d); E.A. Nigg (Max Planck Institute of Biochemistry) for anti-CEP2; D.R. Scoles (University California Los Angeles) for providing constructs; V. Joulov for sharing results before publication; C. McCowan, T. Clingingsmith and C. You for administrative assistance; and K. Salehi-Ashtiani, D. Szeto, R. Murray and C. Lin for characterizing the genomic structure of *HMMR*. L.M.S. was supported by a Department of Defense Breast Cancer Research Program fellowship and a grant from the National Cancer Institute (CA90281to J.D.P.). M.T. was supported by an award from the National Institutes of Health (NIH; K08-AG21613). K.C.G. received support from the US Army Medical Research Acquisition Activity (W23RYX-3275-N605) and NYSTAR (C040066). This work was supported by an NIH/National Cancer Institute (NCI) R33 grant (to M.V.), an NIH/NCI U01 grant (to S. Korsmeyer, S. Orkin, G. Gilliland and M.V.), an NIH/NCI ICBP grant (to J. Nevins and M.V.), an 'interactome mapping' grant from the NIH/National Human Genome Research Institute and the NIH/National Institute of General Medical Sciences (to F. Roth and M.V.), an NIH/NCI P30 grant (CA046592 to the University of Michigan), a Spanish Ministry of Education and Science grant (PR2006-0474 to V.M.) and awards from the Breast Cancer Research Foundation (BCRF13740) and the Niehaus, Southworth, Weissenbach Foundation (to K.O.) and the Koodish Foundation (to T.K.).
We also acknowledge the role of the New York Cancer Project, supported by the Academic Medicine Development Company of New York.

AUTHOR CONTRIBUTIONS

Experiments and data analyses were coordinated by M.A.P., J.-D.J.H., L.M.S and K.N.S. Computational analyses were performed by J.-D.J.H., K.C.G., N.B. and K.V. Yeast two-hybrid analysis screens were performed by M.A.P., J.S.A., J.-F.R and N.A.-G. Biochemical experiments were performed by M.A.P., L.M.S., W.M.E., R.A.G. and B.S. Cell culture and immunofluorescence experiments were performed by M.A.P. and L.M.S. The case-control study in Israel was conceived and executed by G.R. Genotyping and statistical analyses of the case-control studies were performed by K.N.S., L.S.R., G.R., V.M., T.K., B.G., D.L., K.O. and S.B.G. M.A.P., X.S. and P.H. performed the HapMap genotype-haplotype and gene expression association analysis. V.A. provided biochemical analysis support, R.S.G. provided statistical support and M.T., C.L., K.L.N., B.L.W., M.E.C., D.E.H. and D.M.L. helped with overall interpretation of the data. The manuscript was written by M.A.P., J.-D.J.H., L.M.S., D.E.H., M.E.C., S.B.G., J.D.P. and M.V. The project was conceived by M.V. and codirected by S.B.G., J.D.P. and M.V.

COMPETING INTERESTS STATEMENT

The authors declare competing financial interests: details accompany the full-text HTML version of the paper at http://www.nature.com/naturegenetics/.

Published online at http://www.nature.com/naturegenetics
Reprints and permissions information is available online at http://npg.nature.com/reprintsandpermissions

1. Futreal, P.A. *et al.* A census of human cancer genes. *Nat. Rev. Cancer* **4**, 177–183 (2004).
2. Kitano, H. Cancer as a robust system: implications for anticancer therapy. *Nat. Rev. Cancer* **4**, 227–235 (2004).
3. Khalil, I.G. & Hill, C. Systems biology for cancer. *Curr. Opin. Oncol.* **17**, 44–48 (2005).
4. Thomas, R.K. *et al.* High-throughput oncogene mutation profiling in human cancer. *Nat. Genet.* **39**, 347–351 (2007).
5. Sjoblom, T. *et al.* The consensus coding sequences of human breast and colorectal cancers. *Science* **314**, 268–274 (2006).
6. Greenman, C. *et al.* Patterns of somatic mutation in human cancer genomes. *Nature* **446**, 153–158 (2007).
7. Vidal, M. A biological atlas of functional maps. *Cell* **104**, 333–339 (2001).
8. Matthews, L.R. *et al.* Identification of potential interaction networks using sequence-based searches for conserved protein-protein interactions or 'interologs'. *Genome Res.* **11**, 2120–2126 (2001).
9. Futreal, P.A. *et al. BRCA1* mutations in primary breast and ovarian carcinomas. *Science* **266**, 120–122 (1994).
10. Miki, Y. *et al.* A strong candidate for the breast and ovarian cancer susceptibility gene *BRCA1*. *Science* **266**, 66–71 (1994).
11. Wooster, R. *et al.* Identification of the breast cancer susceptibility gene *BRCA2*. *Nature* **378**, 789–792 (1995).
12. Renwick, A. *et al. ATM* mutations that cause ataxia-telangiectasia are breast cancer susceptibility alleles. *Nat. Genet.* **38**, 873–875 (2006).
13. Meijers-Heijboer, H. *et al.* Low-penetrance susceptibility to breast cancer due to *CHEK2*(*)1100delC in noncarriers of *BRCA1* or *BRCA2* mutations. *Nat. Genet.* **31**, 55–59 (2002).
14. Su, A.I. *et al.* Large-scale analysis of the human and mouse transcriptomes. *Proc. Natl. Acad. Sci. USA* **99**, 4465–4470 (2002).
15. Cunliffe, H.E. *et al.* The gene expression response of breast cancer to growth regulators: patterns and correlation with tumor expression profiles. *Cancer Res.* **63**, 7158–7166 (2003).
16. van't Veer, L.J. *et al.* Gene expression profiling predicts clinical outcome of breast cancer. *Nature* **415**, 530–536 (2002).
17. Gunsalus, K.C. *et al.* Predictive models of molecular machines involved in *Caenorhabditis elegans* early embryogenesis. *Nature* **436**, 861–865 (2005).
18. Maxwell, C.A. *et al.* RHAMM is a centrosomal protein that interacts with dynein and maintains spindle pole stability. *Mol. Biol. Cell* **14**, 2262–2276 (2003).
19. Entwistle, J., Hall, C.L. & Turley, E.A. HA receptors: regulators of signalling to the cytoskeleton. *J. Cell. Biochem.* **61**, 569–577 (1996).
20. Boulton, S.J. *et al.* BRCA1/BARD1 orthologs required for DNA repair in *Caenorhabditis elegans*. *Curr. Biol.* **14**, 33–39 (2004).
21. Li, S. *et al.* A map of the interactome network of the metazoan *C. elegans*. *Science* **303**, 540–543 (2004).
22. Bellanger, J.M. & Gonczy, P. TAC-1 and ZYG-9 form a complex that promotes microtubule assembly in *C. elegans* embryos. *Curr. Biol.* **13**, 1488–1498 (2003).
23. Le Bot, N., Tsai, M.C., Andrews, R.K. & Ahringer, J. TAC-1, a regulator of microtubule length in the *C. elegans* embryo. *Curr. Biol.* **13**, 1499–1505 (2003).
24. Srayko, M., Quintin, S., Schwager, A. & Hyman, A.A. *Caenorhabditis elegans* TAC-1 and ZYG-9 form a complex that is essential for long astral and spindle microtubules. *Curr. Biol.* **13**, 1506–1511 (2003).
25. Maxwell, C.A., Keats, J.J., Belch, A.R., Pilarski, L.M. & Reiman, T. Receptor for hyaluronan-mediated motility correlates with centrosome abnormalities in multiple myeloma and maintains mitotic integrity. *Cancer Res.* **65**, 850–860 (2005).
26. Hsu, L.C. & White, R.L. BRCA1 is associated with the centrosome during mitosis. *Proc. Natl. Acad. Sci. USA* **95**, 12983–12988 (1998).
27. Marmorstein, L.Y. *et al.* A human BRCA2 complex containing a structural DNA binding component influences cell cycle progression. *Cell* **104**, 247–257 (2001).
28. Bieche, I., Tozlu, S., Girault, I. & Lidereau, R. Identification of a three-gene expression signature of poor-prognosis breast carcinoma. *Mol. Cancer* **3**, 37 (2004).
29. Freed, E. *et al.* Components of an SCF ubiquitin ligase localize to the centrosome and regulate the centrosome duplication cycle. *Genes Dev.* **13**, 2242–2257 (1999).
30. Hashizume, R. *et al.* The RING heterodimer BRCA1-BARD1 is a ubiquitin ligase inactivated by a breast cancer–derived mutation. *J. Biol. Chem.* **276**, 14537–14540 (2001).
31. Starita, L.M. *et al.* BRCA1-dependent ubiquitination of γ-tubulin regulates centrosome number. *Mol. Cell. Biol.* **24**, 8457–8466 (2004).
32. Ouchi, M. *et al.* BRCA1 phosphorylation by aurora-A in the regulation of G2 to M transition. *J. Biol. Chem.* **279**, 19643–19648 (2004).
33. Wang, X. *et al.* Overexpression of aurora kinase A in mouse mammary epithelium induces genetic instability preceding mammary tumor formation. *Oncogene* **25**, 7148–7158 (2006).

34. Claus, E.B., Risch, N.J. & Thompson, W.D. Using age of onset to distinguish between subforms of breast cancer. *Ann. Hum. Genet.* **54**, 169–177 (1990).
35. Mitchell, M.K., Gregersen, P.K., Johnson, S., Parsons, R. & Vlahov, D. The New York Cancer Project: rationale, organization, design, and baseline characteristics. *J. Urban Health* **81**, 301–310 (2004).
36. Raff, J.W. Centrosomes and cancer: lessons from a TACC. *Trends Cell Biol.* **12**, 222–225 (2002).
37. Joukov, V. *et al.* The BRCA1/BARD1 heterodimer modulates ran-dependent mitotic spindle assembly. *Cell* **127**, 539–552 (2006).
38. Lingle, W.L. *et al.* Centrosome amplification drives chromosomal instability in breast tumor development. *Proc. Natl. Acad. Sci. USA* **99**, 1978–1983 (2002).
39. Pihan, G.A., Wallace, J., Zhou, Y. & Doxsey, S.J. Centrosome abnormalities and chromosome instability occur together in pre-invasive carcinomas. *Cancer Res.* **63**, 1398–1404 (2003).
40. Tutt, A. *et al.* Absence of Brca2 causes genome instability by chromosome breakage and loss associated with centrosome amplification. *Curr. Biol.* **9**, 1107–1110 (1999).
41. Xu, X. *et al.* Centrosome amplification and a defective G2-M cell cycle checkpoint induce genetic instability in *BRCA1* exon 11 isoform–deficient cells. *Mol. Cell* **3**, 389–395 (1999).
42. Sotillo, R. *et al.* Mad2 overexpression promotes aneuploidy and tumorigenesis in mice. *Cancer Cell* **11**, 9–23 (2007).
43. Han, J.D. *et al.* Evidence for dynamically organized modularity in the yeast protein-protein interaction network. *Nature* **430**, 88–93 (2004).
44. Kim, S.K. *et al.* A gene expression map for *Caenorhabditis elegans*. *Science* **293**, 2087–2092 (2001).
45. Formstecher, E. *et al.* Protein interaction mapping: a *Drosophila* case study. *Genome Res.* **15**, 376–384 (2005).
46. Giot, L. *et al.* A protein interaction map of *Drosophila melanogaster*. *Science* **302**, 1727–1736 (2003).
47. Peri, S. *et al.* Human protein reference database as a discovery resource for proteomics. *Nucleic Acids Res.* **32**, D497–D501 (2004).
48. Walhout, A.J. & Vidal, M. High-throughput yeast two-hybrid assays for large-scale protein interaction mapping. *Methods* **24**, 297–306 (2001).
49. Barrett, J.C., Fry, B., Maller, J. & Daly, M.J. Haploview: analysis and visualization of LD and haplotype maps. *Bioinformatics* **21**, 263–265 (2005).
50. Solé, X., Guinó, E., Valls, J., Iniesta, R. & Moreno, V. SNPStats: a web tool for the analysis of association studies. *Bioinformatics* **22**, 1928–1929 (2006).