PLOS ONE

RESEARCH ARTICLE

A Systems Biology Comparison of Ovarian Cancers Implicates Putative Somatic Driver Mutations through Protein-Protein Interaction Models

Mary Qu Yang[1]*, **Laura Elnitski**[2]*

1 MidSouth Bioinformatics Center and Joint Bioinformatics Ph.D. Program, University of Arkansas at Little Rock and University of Arkansas for Medical Sciences, 2801 S. University Avenue, Little Rock, Arkansas, 72204, United States of America, **2** National Human Genome Research Institute, National Institutes of Health, Rockville, MD, 20852, United States of America

* mqyang@ualr.edu (MQY); elnitski@mail.nih.gov (LE)

Citation: Yang MQ, Elnitski L (2016) A Systems Biology Comparison of Ovarian Cancers Implicates Putative Somatic Driver Mutations through Protein-Protein Interaction Models. PLoS ONE 11(10): e0163353. doi:10.1371/journal.pone.0163353

Editor: Joel S. Bader, Johns Hopkins University, UNITED STATES

Received: January 4, 2016

Accepted: September 7, 2016

Published: October 27, 2016

Data Availability Statement: All relevant data are within the paper and its Supporting Information files.

Funding: This project was supported by the Intramural Research Program of the National Human Genome Research Institute (NHGRI) at the National Institutes of Health (NIH). In addition, we acknowledge the NIH Academic Research Enhancement Award 1R15GM114739 and Arkansas Science and Technology Authority (ASTA) Award # 15-B-23 to MQY. Funding for the publication of the article came from NIH/NHGRI/ Division of Intramural Research (DIR). The funders

Abstract

Ovarian carcinomas can be aggressive with a high mortality rate (e.g., high-grade serous ovarian carcinomas, or HGSOCs), or indolent with much better long-term outcomes (e.g., low-malignant-potential, or LMP, serous ovarian carcinomas). By comparing LMP and HGSOC tumors, we can gain insight into the mechanisms underlying malignant progression in ovarian cancer. However, previous studies of the two subtypes have been focused on gene expression analysis. Here, we applied a systems biology approach, integrating gene expression profiles derived from two independent data sets containing both LMP and HGSOC tumors with protein-protein interaction data. Genes and related networks implicated by both data sets involved both known and novel disease mechanisms and highlighted the different roles of *BRCA1* and *CREBBP* in the two tumor types. In addition, the incorporation of somatic mutation data revealed that amplification of *PAK4* is associated with poor survival in patients with HGSOC. Thus, perturbations in protein interaction networks demonstrate differential trafficking of network information between malignant and benign ovarian cancers. The novel network-based molecular signatures identified here may be used to identify new targets for intervention and to improve the treatment of invasive ovarian cancer as well as early diagnosis.

Introduction

Ovarian cancer is the most lethal gynecological malignancy, and serous carcinomas of the ovary account for the majority of ovarian cancer deaths[1]. Papillary serous ovarian cancer, the most common ovarian tumor subtype, comprises a spectrum of disease, ranging from invasive carcinomas to benign, low-malignant-potential (LMP) tumors. Invasive serous carcinomas have been further subdivided into low-grade and high-grade subtypes based on molecular

had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Competing Interests: The authors have declared that no competing interests exist.

Abbreviations: GEO, Gene Expression Omnibus; HGSOC, high-grade serous ovarian carcinoma; LMP, low-malignant-potential tumors; PPI, protein-protein interaction; TCGA, The Cancer Genome Atlas.; PCC, Pearson Correlation Coefficient; AUC, area under curve; GO, Gene Ontology; OMIM, Online Mendelian Inheritance in Man; COSMIC, Catalogue of Somatic Mutations in Cancer; SVM, Support Vector Machne.

characteristics, disrupted functional pathways, and patient outcomes[2]. Low-grade serous carcinomas have characteristics similar to those of LMP tumors; both differ substantially from high-grade serous carcinomas (HGSOCs) [2–6]. Currently, it is poorly understood why LMP tumors follow a benign clinical course, despite their malignant features and metastatic potential, whereas HGSOCs are very aggressive and can spread quickly to other organs.

Comparisons of LMP and HGSOC tumors may offer unique insights into malignant ovarian tumors by revealing the characteristics of aggressive tumors. However, the methods available to make such comparisons have significant limitations. Genome-wide microarray data have been used widely in ovarian cancer research [7–13] to provide information on the relative abundance of transcripts [14, 15]. But these expression profiles do not reveal the presence of somatic mutations. By contrast, whole genome sequencing and exome sequencing can be used to identify disease-associated mutations. For example, the Cancer Genome Atlas (TCGA) Consortium recently published a large-scale, comprehensive overview of HGSOCs that revealed recurrent *TP53* mutations in more than 90% of samples [16]. What is currently missing is an integrative means of analysis designed to identify somatic mutations that affect downstream gene expression levels and protein interaction networks. By integrating expression and mutation data, we can better identify driver mutations and dysregulated functional pathways present in cancer cells.

Models depicting protein-protein interaction (PPI) data provide a useful proxy for cellular communication lattices [17, 18] and, when integrated with transcript expression data, can reveal disruptions cascading from gene mutations or other functional alterations. The communication lattice consists of a grid-like matrix of multiple interleaved protein interaction networks, including smaller focal interaction units known as subnetworks. These individual components constitute single targeted outcomes within the larger, more dispersed framework of the network. Tying together the interactions of multiple subnetworks are centralized conduits, or hub proteins, such as MYC, P53, and EGFR, which play information-trafficking roles in the cell. Hub proteins are defined as the proteins with the highest numbers of interactions with other proteins in the proteome [19–21]. Due to their essential functions, the mutation, deletion, or functional alteration of hub proteins causes severe, though rare, phenotypic outcomes [22–24]. Thus, proteins severely affected by mutation are more likely to reside in less central network locations. A network-based method that incorporates transcriptome and human interactome enable the identification of genes acting as regulators by mediating expression of downstream genes and play essential roles in disease.

Here, we integrate the vast amount of expression data available from HGSOC and LMP microarrays with PPI data to identify disrupted downstream PPIs. Our method proceeds by objectively selecting the most strongly affected protein interactions from over 100,000 options. We also demonstrate our method's predictive ability by using our gene set to classify an independent set of ovarian tumor samples according to HGSOC or LMP subtype. This information may aid diagnosis of HGSOC, which is often asymptomatic in its early stages and thus detected late, contributing to the low survival rate associated with this subtype [7, 25, 26]. In addition to strengthening our ability to classify tumors at the molecular level, this type of systems biology approach aids in identifying molecular perturbations that are camouflaged in gene expression data, providing insight into the biological mechanisms underlying cancer.

Results

Comparison of methods to rank genes with respect to differential expression

Two independent microarray expression profiles, GSE17308 [9] and GSE9891 [12], were obtained from the Gene Expression Omnibus (GEO) to compare LMP and HGSOC samples

across independent patient groups. We first created lists of genes ranked by fold-change and statistical significance [27]. We then compared the reproducibility rate of lists created with fold change in mean values, fold change in median values, t-test *P*-values, and Wilcoxon rank-sum test *P*-values, selecting sets of genes at random to calculate the statistical significance of our findings (Fig 1). Here, reproducibility was defined as the percentage of differentially expressed genes from one expression profile's list also included in the other profile's list (i.e., those genes identified by both the GSE17308 and GSE9891 comparisons). The reproducibility rate of the lists generated with each of the four methods was significantly higher than the rate for the random gene list, as indicated by $P < 2 \times 10^{-12}$ for all four statistical tests. However, the lists

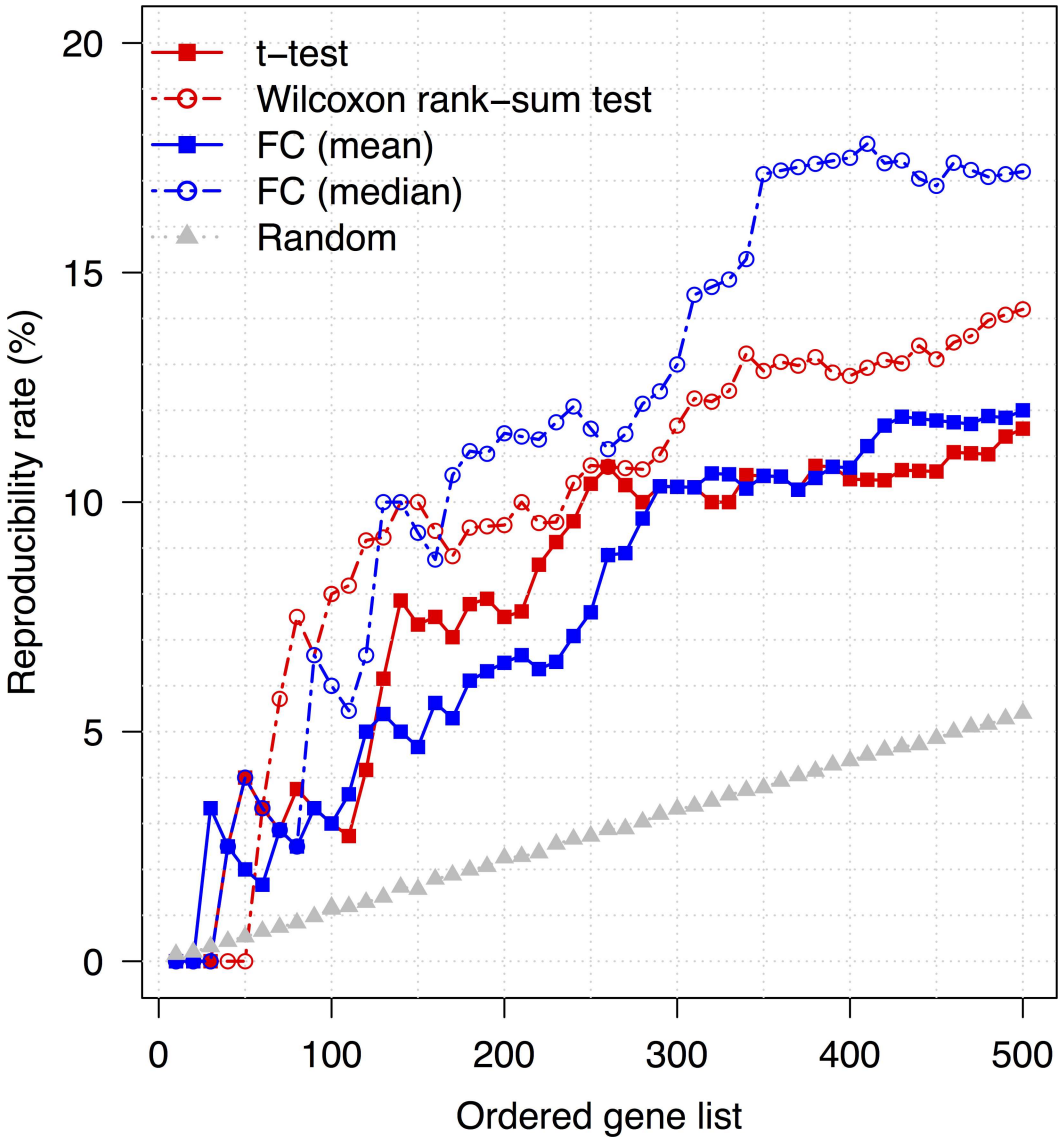


Fig 1. Reproducibility rates for various methods of identifying genes differentially expressed between low-malignant-potential (LMP) and high-grade serous carcinoma (HGSOC) samples in two data sets, GSE17308 (9) and GSE9891 (12). Genes were selected by: the *P*-values from t-tests (*solid red line*); the *P*-values from Wilcoxon rank-sum tests (*dotted red line*); mean fold change (FC; *solid blue line*); median FC (*dotted blue line*); and random selection (*dotted grey line*). Reproducibility was defined as the percentage of differentially expressed genes from one data set's list also included in the other data set's list.

doi:10.1371/journal.pone.0163353.g001

compiled on the basis of fold-change in median expression values demonstrated a higher rate of reproducibility than those compiled on the basis of fold-change in mean expression levels (Fig 1). Further, we evaluated the effectiveness of these gene lists in discriminating between different phenotypes using hierarchical cluster analysis. More homogeneous sample clusters were achieved with the top-ranked genes from the Wilcoxon rank-sum test (S1A Fig) than with those identified by median fold change (S1B Fig), with homogeneity being defined as how well the selected genes were able to separate disease types (i.e., LMP vs HGSOC). In other words, more homogeneous clusters contain more samples with the same phenotype. For instance, the bottoms of panels A and B in S1 Fig display the hierarchical clusters of samples from GSE17308. Five of seven HGSOC tissue samples were placed in the invasive clusters, indicating less homogeneous clusters, by median fold change, whereas all seven samples were correctly clustered together by the genes from the Wilcoxon rank-sum test.

Thus, to simultaneously maximize reproducibility and effectiveness at classifying tumor types, as well as ensure that differential expression was sufficiently large to be biological meaningful [28], we used both the Wilcoxon rank-sum test and median-fold change to identify differentially expressed genes in subsequent analyses.

Analysis of differentially expressed genes concordant between expression profiles

Differential gene expression. We identified 195 differentially expressed genes ($P < 0.01$, FC > 0.6) for GSE17308 and 230 ($P < 10^{-6}$ and FC > 1.5) for GSE9891. To attain similar numbers of top-ranked differentially expressed genes, different thresholds were adopted. The distributions of the median fold-changes and the *P*-values from the Wilcoxon rank-sum test differed between the two GEO data sets (S2 Fig) were varied, which may cause by different platforms used in experiments for attaining expression profiles and other technical variations [27]. The list of genes differentially expressed between tumor types in the two independently generated GEO data sets converged on a common set of 23 genes (8 upregulated and 15 downregulated in HGSOC) that were directionally consistent and statistically significant in both expression profiles (Fig 2A). To measure the reliability of these data, we used expression levels of these 23 genes to perform hierarchical clustering of a third, independent patient expression profile (GSE27651) [29] containing LMP and HGSOC samples (S3 Fig). Not a single one of the tissue samples in this third data set was clustered erroneously.

Data interpretation using pathway analysis. Six of the 23 differentially expressed genes have already been implicated in the ovarian cancer literature (*STAT1*, *MYBL2*, *SPRY2*, *NR2F1*, *DUSP4*, and *RPS23*) [30–35]. However, most of these genes represent potentially novel disease biomarkers. An Ingenuity pathway analysis showed that the list of 23 genes is enriched in tumor-related processes such as cellular movement, cancer, and cellular development (12 genes, Fig 2B) and cell cycle, drug metabolism, and molecular transport (11 genes, S4 Fig). Importantly, *DUSP4*, a known cancer risk gene [36], is represented in both of these pathways, suggesting pleiotropy with multiple functional associations. In addition, the analysis identified canonical networks whose members relay signals from the plasma membrane to the nucleus. Pathways enriched for this gene list included aryl hydrocarbon receptor (AhR) signaling, which is linked to homologous recombination [37]; chromosomal instability, as reported in HGSOC [16]; and retinoic acid receptor (RAR) activation. Thus, LMP and HGSOC tumors exhibit differences in cellular pathways, as revealed by altered expression of the genes involved.

The static mapping of differentially expressed genes onto network and pathway illustrations showed that known disease processes are affected by differential gene expression

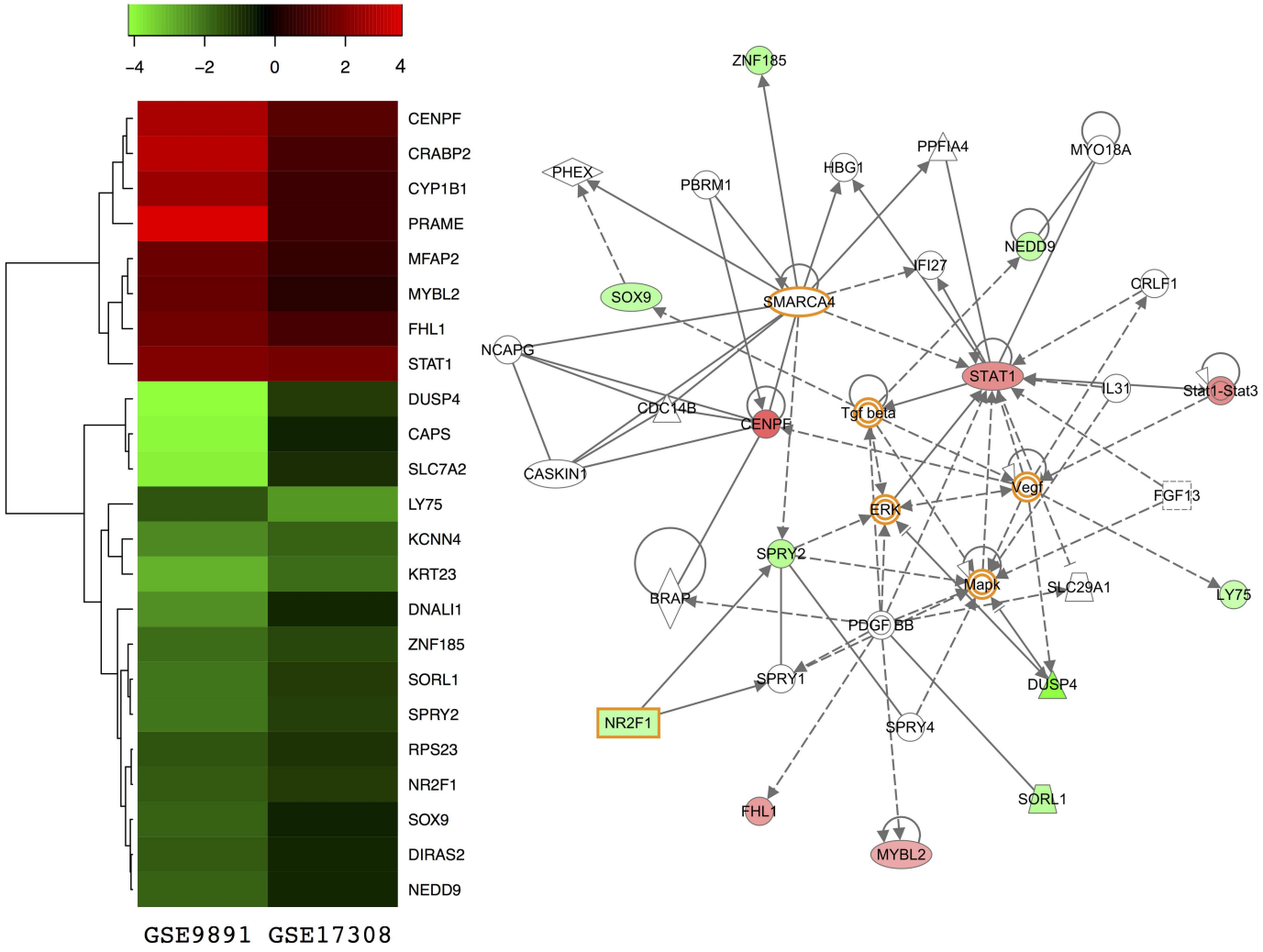


Fig 2. Differentially expressed genes concordant between two data sets containing low-malignant-potential (LMP) and high-grade serous carcinoma (HGSOC) samples. (A) Heatmap of 23 genes differentially expressed between LMP and HGSOC tumors in two independent GEO expression profiles, GSE17308 (9) and GSE9891 (12), organized by median fold change. (B) The "cellular movement, cancer, and cellular development" network is enriched for genes differentially expressed between LMP and HGSOC, as determined by pathway analysis. In both panels, red represents overexpression and green represents underexpression in HGSOC. Genes highlighted in orange are well-known ovarian cancer genes that have been used as biomarkers and targets for drug design.

doi:10.1371/journal.pone.0163353.g002

(Fig 2B **and** S4 Fig). However, contrary to our expectation, the majority of affected genes were localized to the periphery of the pathway/network maps rather than to interior positions, suggesting limited impact on information dissemination. By contrast, some well known ovarian cancer genes whose expression was not altered, such as *VEGF*, *ERK*, *MAPK*, *TGF-β*, and *SMARCA4* [38–45], displayed centralized placement in these same network maps, indicating high connectivity within the gene networks. We postulated that genes encoding proteins with important regulatory functions may not be differentially expressed themselves but could mediate downstream effects on genes whose expression levels were altered in our microarray meta-analysis.

Subnetwork Analysis

To identify characteristics that can be used to discriminate between HGSOC and LMP samples, we employed a subnetwork analysis method adapted from Chuang et al. [46] and projected pairwise PPI data onto a larger combinatorial lattice informed by gene expression values. We overlaid the expression value measured for each gene onto its corresponding protein, as we built an increasingly large, interconnected protein framework. We searched for subnetwork conduits whose differential activities scored as being highly discriminative for tumor type, as assessed by amutual information score and statistical thresholds (see Materials and Methods). Starting from a single gene as a subnetwork "seed," neighboring genes were added to expand the interaction framework by prioritizing choices that maximized the differential scores between tumor types; a computationally intensive approach was used to check all options before proceeding to the next addition. To ascertain the statistical significance of the differential networks, we compared our finished network data to a null distribution, which we generated by randomizing tumor type assignments or gene assignments within each subnetwork. We then sampled the randomized sets 10,000 times to evaluate the statistical significance of the results returned for each subnetwork. In this way, we identified 175 subnetworks from the GSE9891 data set and 179 from the GSE17308 data set that were differentially affected in HGSOC and LMP tumors (each with $P \leq 10^{-4}$, discrimination score, or DS ≥ 0.66; **S1 Table**). Among these subnetworks, we found 75 genes including 9 seed genes implicated in both data sets.

Differential expression and subnetwork participation. Six genes present in the significantly impacted subnetworks also showed differential expression in both data sets: *STAT1*, *MYBL2*, *RPS23*, *NR2F1*, *SOX9*, and *SPRY2*. Each gene participated in from one to eight subnetworks, the components of which showed enrichment in gene ontology (GO) terms for the cell cycle, apoptosis, regulation of transcription, signal processing, cell communication, and receptor protein tyrosine kinase signaling pathways.

Differential subnetwork participants associated with tumor type. We postulated that the 75 subnetwork genes that were implicated in both data sets and could implicate events that disrupt the protein interaction lattice. In particular, we hypothesized that mutations may have significantly impacted the expression of protein interaction partners and their downstream targets or network affiliations, without affecting the expression of the mutant genes themselves [7, 47]. The majority of the subnetworks we identified consisted of a mixture of genes with and without significantly altered expression levels. For example, although *TP53* displays driver mutations in HGSOC samples[48–50], we did not find significant changes in *TP53* expression between HGSOC and LMP samples. Nevertheless *TP53* was present in subnetworks that distinguished HGSOC from LMP (**Fig 3A**, $P < 10^{-4}$). Similarly, *BRCA1* [51] participated in the significant differential subnetworks ($P < 10^{-4}$) but was not differentially expressed (**Fig 3B**). Other examples of driver genes identified in the subnetwork analysis that were not differentially expressed included *ERBB2* [52] (**Fig 3C**), *MYC* (**Fig 3D**) [53–55], and *CTNNB1* (**Fig 3G**) [56] ($P < 10^{-4}$ for each). Although we did not have the original tumor samples with which to assess the presence of somatic mutations, the disrupted pathway interactions detected by the subnetwork analysis implicate consistent alteration of these gene functions between tumor types. Thus, this network-based method allowed us to assess mutations in the context of networks, enhancing our ability to identify driver mutations.

Data interpretation using pathway analysis. To address the biological roles of the subnetwork genes, we performed a GO analysis on all significant subnetworks. A large proportion of GSE9891- and GSE17308-generated subnetworks (59.7% and 52.8%, respectively) were enriched for biological process terms related to cancer, including proliferation, apoptosis, cell

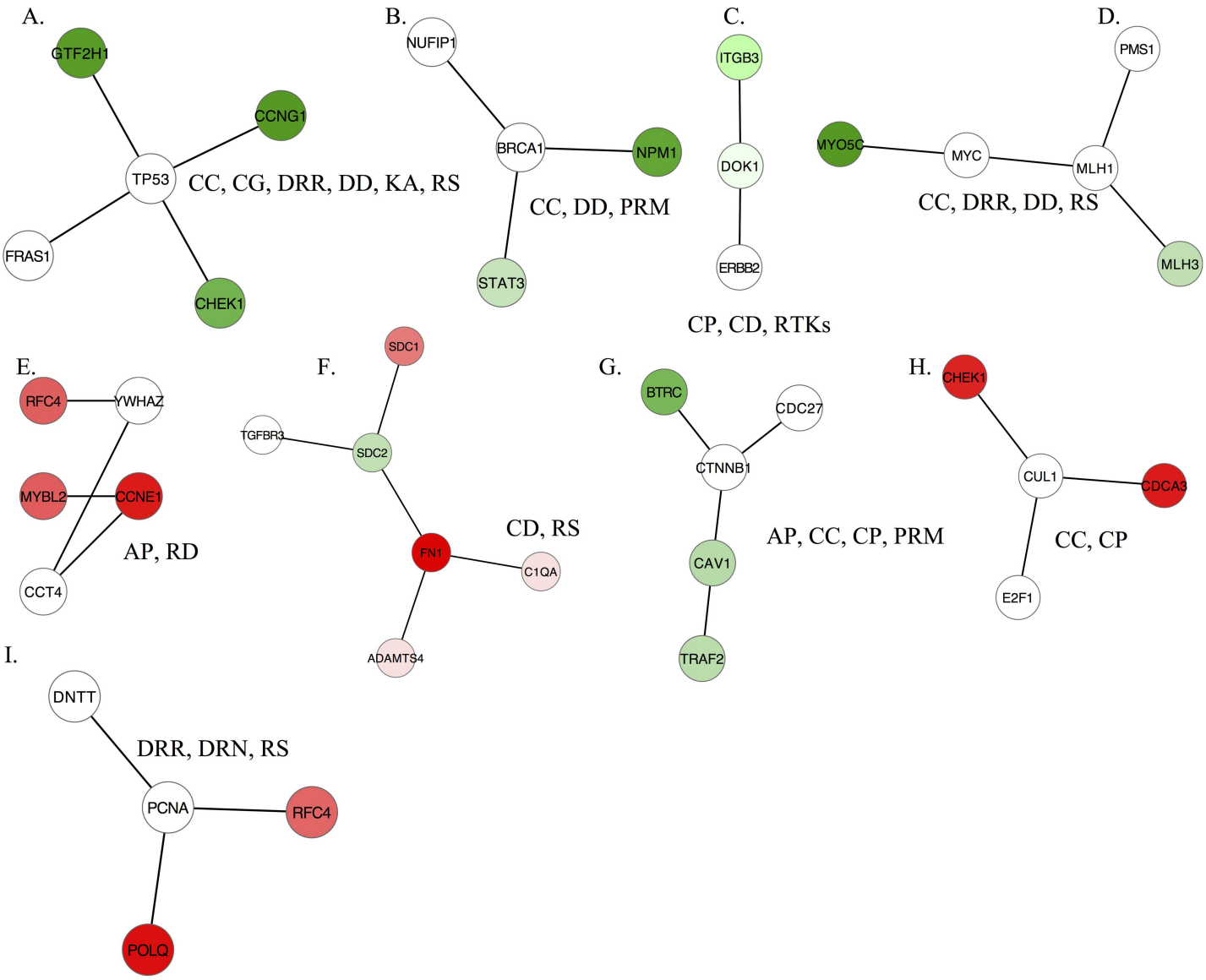


Fig 3. Representative subnetworks that discriminate between low-malignant-potential (LMP) and high-grade serous carcinoma (HGSOC) samples. (A-I) Subnetworks include genes such as *TP53*, *BRCA1*, and *MYC*, which are mutated in ovarian carcinomas although their expression level is not significantly altered (*white nodes*). Red nodes indicate overexpression of genes in HGSOC subnetworks, and green nodes indicate underexpression. Abbreviations: cell cycle (CC), cell growth (CG), cell proliferation (CP), cell differentiation (CD), DNA damage (DD), DNA repair (DDR), DNA replication (DRN), regulation of kinase activity (RKA), receptor protein tyrosine kinase signaling pathway (RTKs), positive regulation of metabolic process (PRM), response to drug (RD), response to stress (RS).

doi:10.1371/journal.pone.0163353.g003

cycle, differentiation, kinase activity, stress response, DNA replication, and DNA damage repair. By contrast, only 8.6% and 8.3% of randomly generated networks were associated with these biological processes ($P < 2.2^{*}10^{-16}$, Fisher's exact test) Additionally, a KEGG pathway analysis showed that 41.5% of the GSE9891 and 28.3% of the GSE17308 significantly altered subnetworks were enriched for *TP53*, *ERBB2*, *MAPK*, B-cell receptor, cell cycle, and focal adhesion pathways, compared with 0.5% and 1% of random networks ($P < 2.2^{*}10^{-16}$, Fisher's exact test). Importantly, disruption of *TP53*, *ERBB*, and *MAPK* pathways has been implicated in ovarian tumorigenesis [57–60], focal adhesion is the most deregulated pathway in ovarian

cancer [61], and B-cell infiltration of ovarian carcinoma effusions is associated with worse outcomes [62]. Thus, the reproducibility of our findings across data sets and their alignment with the existing literature show that an *in silico* approach can be used to implicate subnetworks that play important functional roles in tumorigenesis and identify known processes involved in tumor progression.

Significantly Modified Hub Protein Interactions

The connectivity of disease-relevant proteins and their interaction partners was further investigated by assessing the involvement of the most highly connected proteins in the cell: the hub proteins. Because these proteins play critical roles in development and reproduction, deleterious mutations are typically lethal [24, 63, 64]. Compared with mutations at the network periphery, loss-of-function mutations in these genes are rare and devastate network functions in cells. However, tumor cells can adapt to a partial loss of hub protein function by rewiring their networks to establish workarounds. To assess altered hub protein function, we applied differential network mapping [65] to our HGSOC and LMP comparison. To do so, we calculated the average Pearson correlation coefficient (PCC) between gene expression values of all hub proteins and their interaction partners, directly comparing LMP and HGSOC samples. We analyzed 3,128 hub proteins that interact with at least 5 other proteins and identified 178 hub proteins in GSE17308 and 220 in GSE9891 whose PCCs differed across tumor types ($P < 0.05$); each was connected to genes with measurable expression changes between the tumor types. Of these hub proteins, 34 were implicated in both data sets (**S2 Table**).

One of the hub proteins was *BRCA1*. *BRCA1* expression was less correlated with *TP53* and *BRCA2* expression in HGSOC than in LMP samples. The value of the PCC decreased significantly from 0.704 and 0.551 in LMP to 0.096 and 0.065 in HGSOC for *TP53* and *BRCA2*, respectively (P = 0.0023 and P = 0.022, **Fig 4**, **Table 1**). Several additional genes also had PCCs with *BRCA1* that differed significantly across tumor types, including *AKT1*, which encodes a multifunctional serine-threonine protein kinase; *XRCC1*, a DNA repair gene; and *RBBP7*, *CDS1*, and *SMARCC2*. All of the genes mentioned exhibited decreased expression correlations with *BRCA1* in HGSOC (**Fig 4**, **Table 1**), suggesting disrupted function of the BRCA1 protein.

Similar to the *BRCA1* results, the vast majority of significant hub proteins in each GEO set were not differentially expressed themselves [205/220 (93.2%) in GSE9891 and 168/178 (94.4%) in GSE17308]. Instead, hub proteins were implicated through differential expression PCCs among interaction partners between tumor types.

Data interpretation using pathway analysis. We traced most of the significant hub proteins detected in GSE17308 (109/178) and GSE9891 (167/220) to an interconnected network enriched for cancer pathways, cell cycle, signaling, and growth factor binding (as defined by IPA: $P < 0.05$ for Fisher's exact test; **Fig 5A and 5B**). Of the 34 hub proteins found in both data sets, 19 are associated with cancer ($P < 0.05$), 23 have been linked to genetic disorders ($P < 0.05$), and 31 have been associated with the molecular function GO term "protein binding" ($P < 5.3 \times 10^{-5}$). Moreover, 10 of the 34 hub proteins interacted with each other (**Fig 5C**, middle panel; five interacting hub protein pairs), and 13 directly participated in the same interconnected network we derived from known protein interactions, whereas 11 proteins lacked any central ties (**Fig 5C**). Notably, *CREBBP*, whose product plays an essential role in the cell cycle, was frequently interconnected with hubs identified here and also directly interacted with three of the genes differentially expressed in both data sets: *STAT1*, *MYBL2*, and *SOX9*. The published TCGA data have documented five nonsynonymous and two frameshift somatic mutations in *CREBBP* in 316 HGSOC cases [16]. Our data implicate *CREBBP* interactions as a source of differential information trafficking between LMP and HGSOC tumors.

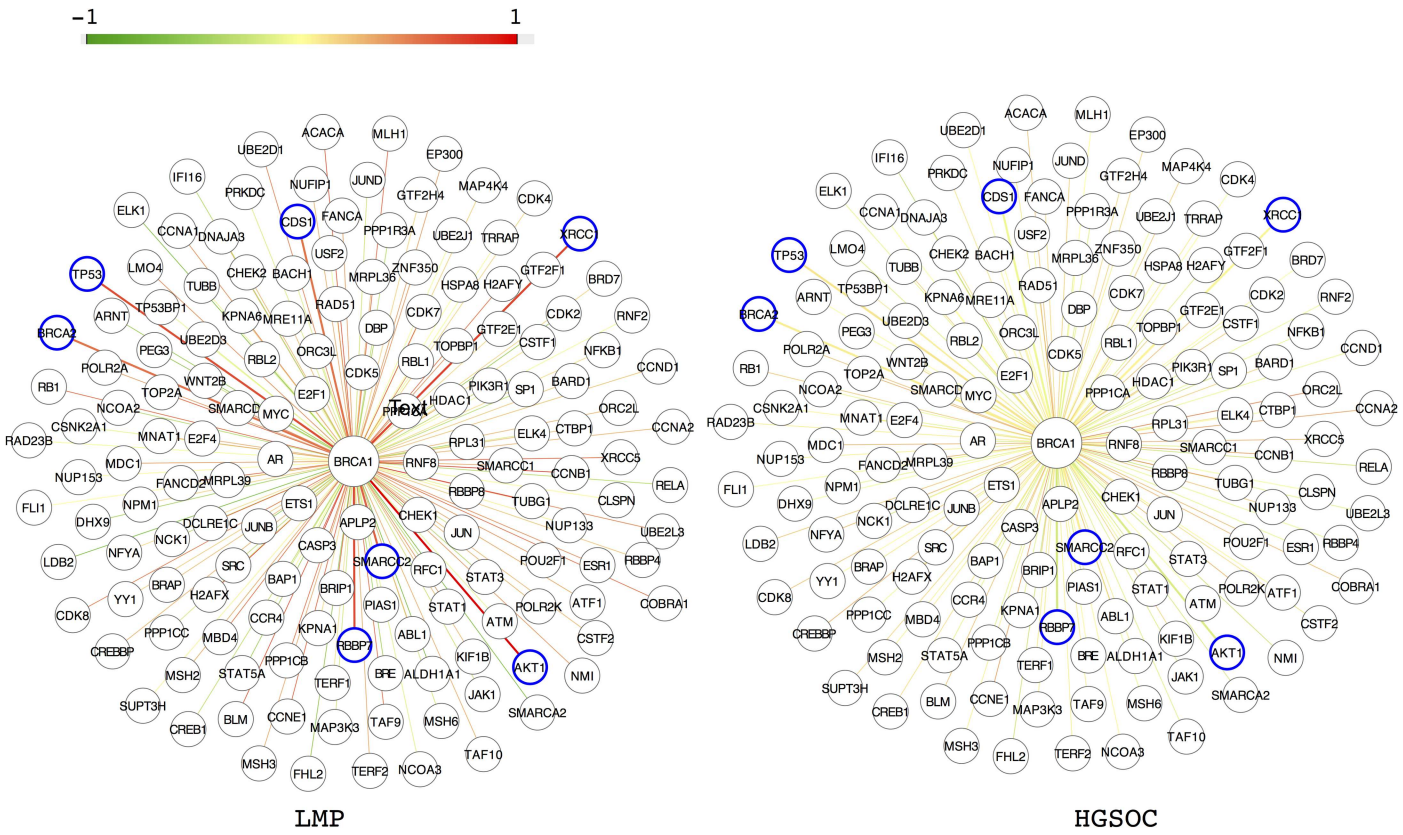


Fig 4. Pearson correlation coefficients (PCCs) for expression levels of *BRCA1* and interaction partners in low-malignant-potential (LMP) and high-grade serous carcinoma (HGSOC) samples. The color of an edge represents the value of the PCC for the genes that it connects. Blue circles indicate genes whose expression levels are highly correlated with *BRCA1* expression levels in LMP, but not in HGSOC.

doi:10.1371/journal.pone.0163353.g004

Analysis of missing links in network data. To further explore the extent of hub protein disruption, we searched for "bridge" or "broker" proteins [66] that connect isolated hub proteins with mainstream core networks. This analysis revealed that *TP53* connects the largest number of orphan hubs ($n = 6$), with *SRC*, *EGFR*, and *TP73* providing additional connections. The central position of these four proteins in a larger network implicates them in disrupting network lattice interactions (Fig 5D). As is true for *TP53*, the documented involvement of these genes in ovarian cancer or tumors in general supports their inclusion as genes of interest [48, 67–69].

Table 1. The correlation coefficients of expression levels of *BRCA1* and interaction partners in low-malignant-potential (LMP) and high-grade serous carcinoma (HGSOC) data sets.

	LMP	HGSCO	Z-value*	P-value*
TP53	0.704	0.096	2.84	0.0023
BRCA2	0.551	0.065	2.02	0.022
AKT1	0.577	-0.163	3	0.0013
XRCC1	0.746	0.043	3.35	0.0004
RBBP7	0.72	0.34	2.02	0.022
CDS1	0.55	-0.66	5.14	0
SMARCC2	0.623	0.067	2.4	0.008

* Fisher's z transformation was appled to obtain Z-value and P-value.

doi:10.1371/journal.pone.0163353.t001

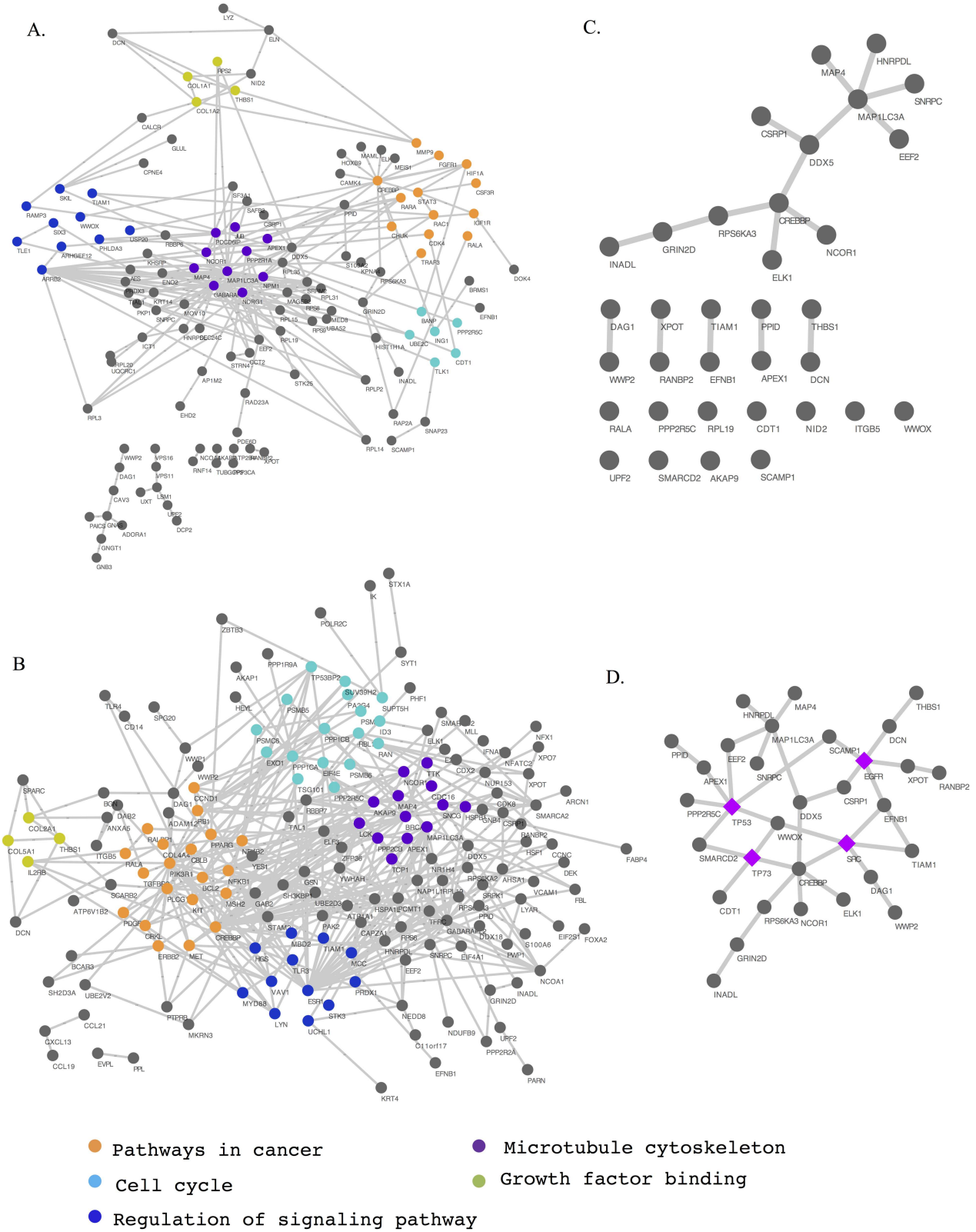


Fig 5. An overarching network connecting hub proteins differentially involved in low-malignant-potential (LMP) and high-grade serous carcinoma (HGSOC) samples. Nodes and edges illustrate the network diagrams of GEO data sets (A) GSE17308 and (B) GSE9891. Each functional complex is represented by a different color. (C) There are 34 hubs that overlap between the two data sets.

(D) Additional gene products, highlighted in purple, have the largest number of connections to the 34 concordant hubs, interconnecting all of the hubs except for six orphans.

doi:10.1371/journal.pone.0163353.g005

Comparison with Indexed Literature

To explore alternative and complementary approaches to our network analysis, we used the GeneIndexer tool, which reconstructs functional relationships among genes based on extensive literature searches and semantic indexing. We fed our list of 34 hub proteins to GeneIndexer, along with three genes that contain causal mutations in HGSOC genes, either hereditary (*BRCA1* and *BRCA2*) or somatic (*TP53*). Using a functional hierarchy tree built using our list of 34 hub proteins and three disease-associated genes, GeneIndexer indicated that *BRCA1*, *BRCA2*, and *TP53* were the genes from the list with the strongest functional connection to each other (**S5 Fig**). The gene most closely related to these three genes was *APEX1*, a base excision repair gene with elevated or altered expression in breast, cervical, and germ cell tumors; gliomas; rhabdomyosarcomas; and non–small cell lung cancer [70, 71]. *CREBBP*, the fifth most connected gene using the GeneIndexer approach (**S5 Fig**), clustered with *NCOR1* (nuclear receptor corepressor 1) in the functional relationship tree. The close functional relationship between *CREBBP* and *NCOR1* was also demonstrated in our hub protein network (**Fig 5C and 5D**). A number of studies have suggested that *NCOR1* plays an important role in human cancers [16, 72–75].

Known Genetic Mutations in Concordant Gene Lists

Using TCGA data, which contain information on somatic mutations and copy number variation for HGSOC patient samples, we searched for genetic mutations occurring in our concordant differentially expressed genes, significant subnetwork genes, and genes encoding significant hub proteins. The somatic mutations were identified using a combination of three algorithms, VarScan 2, SomaticSniper and GATK, applied to 316 high-grade serous carcinomas (HGSOC) samples and 236 normal tissue samples in the TCGA project[16]. Variants were annotated as somatic mutations if they were not observed in the normal samples [16]. We found that 17 of those genes were somatically mutated in HGSOC cases, and that each of these mutant genes was present in 5 to 95% (or 16 to 300 of 316) of patient samples (**Fig 6A**). For example, *TP53* somatic mutations were found in 95% (300/316) of HGSOC samples. When *TP53* was removed from consideration, 67% (210/316) of the samples contained one or more somatic mutations from the 16 remaining concordant genes. We also assessed the presence of amplification and deletion events among the genes implicated in our analyses. Collectively, homozygous deletions occurred in *PTEN*, *CREBBP*, and *WWOX* in 43/316 (13.7%) of HGSOC samples, whereas amplifications of *AP2M1*, *RYR1*, *DNAJB1*, *YWHAZ*, *PAK4*, *RGS19*, *ARRB1*, *STAT1*, *APEX1*, *UPF2*, *DNALI1*, *NEDD9*, and *SORL1* were found in 179/316 (56.7%) of HGSOC samples (**Fig 6A and 6B**). Notably, *RYR1* and *APEX1* are the targets of two FDA-approved cancer drugs (caffeine and lucanthone, respectively). We also found that amplification of *PAK4* and *RGS19* tended to be mutually exclusive, although this trend did not reach statistical significance, ($P = 0.07$, Fisher's exact test, **Fig 6C**), whereas amplification of *RYR1* and *PAK4* tended to co-occur, an association that did reach statistical significance ($P < 0.001$, Fisher's exact test, **Fig 6D**).

The known roles of these proteins in tumorigenesis support our predictions of their relevance. For example, *PAK4* mutations promote oncogenic transformation, and *PAK4* deletions inhibit tumorigenesis [76, 77], whereas *RGS19* deregulates cell proliferation through multiple pathways [78]. In addition, we performed a univariate survival analysis for 17 significant genes

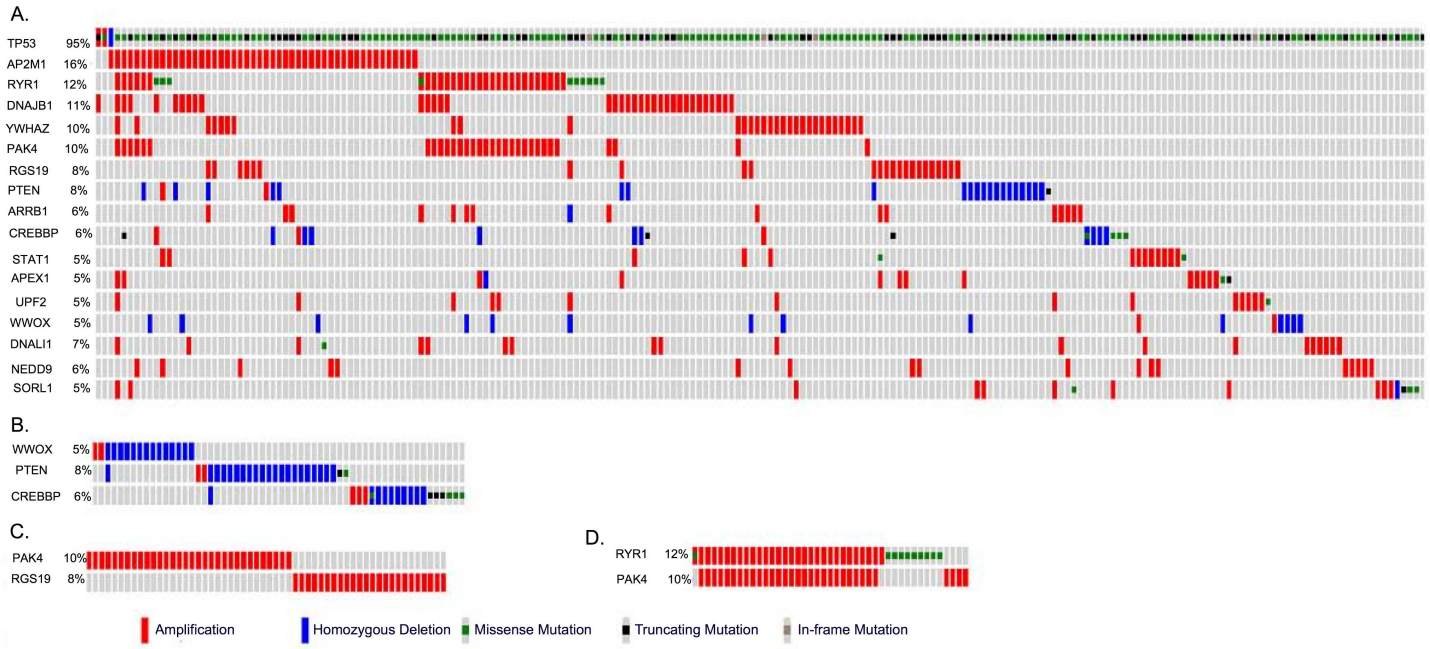


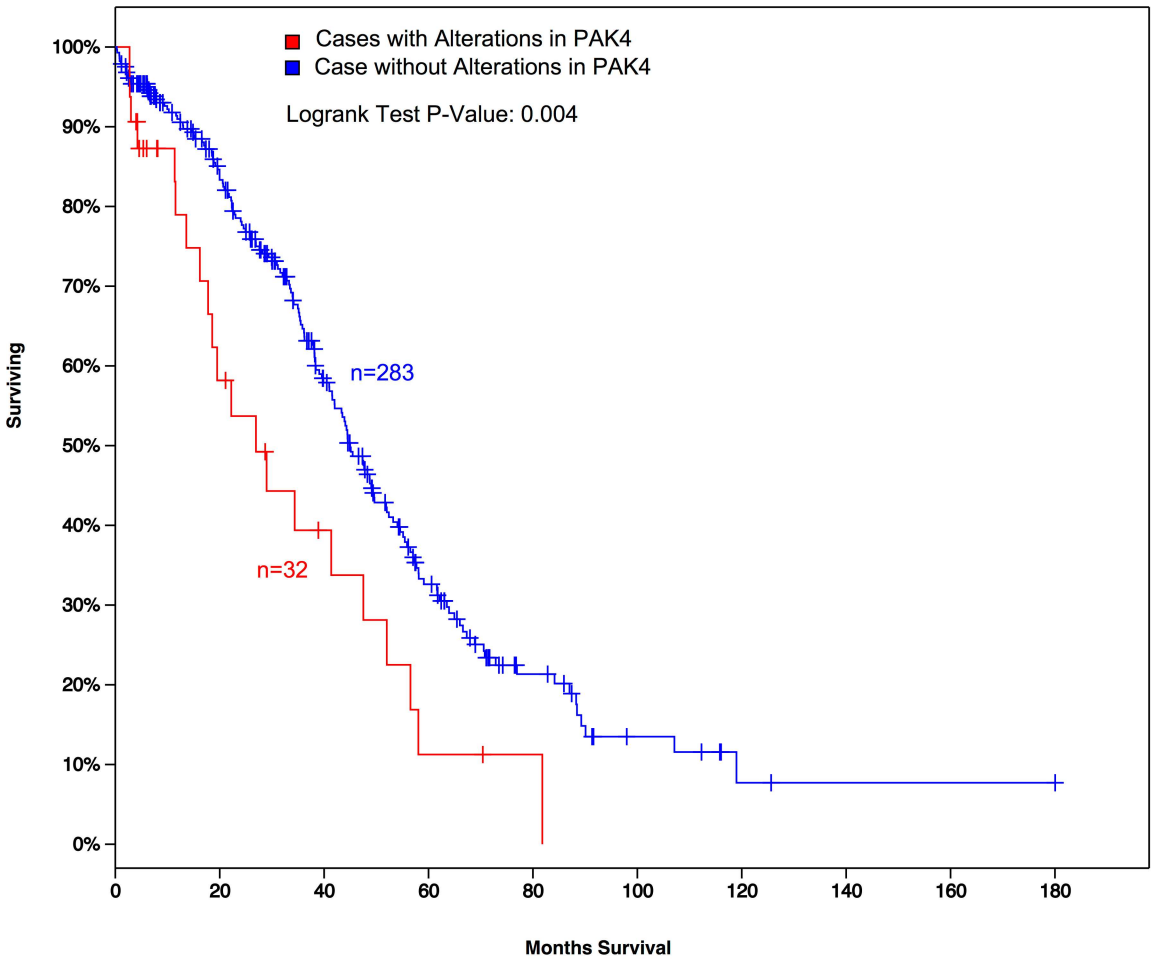
Fig 6. Mutations in genes that significantly differed between low-malignant-potential (LMP) and high-grade serous carcinoma (HGSOC) samples in two data sets, GSE17308 (13) and GSE9891 (16), using gene expression, network, or hub protein analyses. (A) At least 5% (16/316) of Cancer Genome Atlas HGSOC samples contained one or more mutant versions of 17 genes identified from differential expression, network, or hub protein analyses. (B) The majority of genetic alterations in *WWOX*, *PTEN*, and *CREBBP* in HGSOC samples were homozygous deletions. (C) Amplifications in *PAK4* and *RGS19* tended to be mutually exclusive, although this association was not statistically significant (P <0.07). (D) Amplifications in *PAK4* and *RYR1* frequently co-occurred (P <0.001).

doi:10.1371/journal.pone.0163353.g006

identified by our study and also displayed somatic mutations reported by TCGA project in more than 5% HGSOC patients. The *PAK4* has lowest P-value in the survival analysis (P = 0.004, P-adjust = 0.068) (Fig 7). Despite amplification of *RYR1* and *PAK4* tended to co-occur, survival times in patients with *RYR1* alterations alone were not significantly shorter (P = 0.18). Both *RYR1* and *PAK4* participate in diverse functional pathways in the cell. *RYR1* is present in several KEGG pathways, including calcium signaling, circadian entrainment, long-term depression, and oxytocin signaling. *PAK4* is present in pathways for renal cell carcinoma, ErbB signaling (S6 Fig), focal adhesion, T-cell receptor signaling, and regulation of the actin cytoskeleton. Although *RYR1* and *PAK4* are not known to act in the same pathways, their observed co-occurrence in our PPI data suggests that these two genes could interact indirectly through other proteins (S7 Fig).

Comparison of Differentially Expressed Genes, Subnetworks, and Significant Hubs

To evaluate the effectiveness of the various approaches to identifying disease genes used in this study, we compiled a list of 18 known ovarian cancer susceptibility genes present in the GEO expression data sets; we identified these genes by searching for known somatic and germline mutations in ovarian cancer from the Online Mendelian Inheritance in Man (OMIM) database. We identified 195 and 230 differentially expressed genes, 179 and 175 differential subnetworks that consist of 502 and 397 genes, and 178 and 220 significant hubs, for GSE17308 and GSE9891, respectively (S8 Fig). The comparison of our findings with this known list showed that the subnetwork analysis identified nine of these known genes (50%), the hub protein analysis identified five (27.8%), and the differential expression analysis identified one (5.6%). The



	# Total Cases	# Cases Deceased	Median Months Survival
Cases with Alterations in *PAK4*	**32**	**21**	**26.94**
Cases without Alterations in *PAK4*	**283**	**160**	**44.98**

Fig 7. A Kaplan-Meier survival analysis of HGSOC patients with *PAK4* amplification. HGSOC patients ($n = 32$) with *PAK4* amplification have a significantly shorter survival time ($P = 0.004$) than patients without *PAK4* amplification ($n = 283$).

doi:10.1371/journal.pone.0163353.g007

difference in sensitivity between these methods was not statistically significant ($P = 0.18$ for the subnetwork vs. differentially expressed genes comparison, $P = 0.1$ for the hub vs. differentially expressed genes comparison, Fisher exact test). However, *TP53*, *BRCA1*, *ERBB2*, *MLH1*, *PIK3CA*, and *RAD51C* were detectable only with the two network-based approaches. When we examined the methods' ability to detect a list of 288 genes from the Catalogue of Somatic Mutations in Cancer (COSMIC) database that display somatic and germline mutations in general cancer, our subnetwork (21.1%; $P = 0.01$, Fisher exact test) and hub approaches (12.5%, $P = 0.07$) identified significantly or nearly significantly more genes than did the differential expression analysis (5.2%; S9 Fig). This result indicates that network-based approaches may be more useful in terms of identifying important cancer-associated genes and aiding in hypothesis generation than differential expression analyses.

Blind Classification Using a Support Vector Machine Classifier

Finally we constructed support vector machine (SVM) classifiers, which adopted different types molecular signatures (i.e., subnetworks, hub proteins, and differential gene expression, details in the Materials and Methods section) as input features, to assess and compare their effectiveness in classifying blinded sets of HGSOC and LMP tumor samples. Performance was evaluated using threefold cross-validation with bootstrap sampling. Regardless of the molecular signature used in the classification, we achieved robust separation of the samples, as shown by area under the ROC curve (AUC) values (**S10 Fig**, **S3 Table**). AUCs reached 0.98 for both intra-data set classification, in which the molecular signatures used for training and classification were derived from the same data set, and inter-data set classification, in which the molecular signatures used for training and classification were derived from different data sets. The successful classification of HGSOC and LMP samples suggests that these signatures capture the molecular perturbations and alterations occurring in the tumors and, moreover, that the pathways that underlie the initiation and progression of LMP and HGSOC are different.

Discussion

We have extended traditional gene expression analyses by incorporating a systems biology approach to dissect the molecular differences between LMP and HGSOC. Although the differentially expressed genes that we identified characterized distinct properties of the tumors, the cause-and-effect relationships underlying transcriptional disruption could not be explained with gene expression data alone. Our subnetwork analysis identified a series of significantly altered, interacting complexes that coordinate higher-level functions in important biological pathways, demonstrating that differential gene expression reflects multiple processes in tumor cells. In addition, our hub protein analysis identified consistent alterations in the information-dissemination centers that perpetuated the largest numbers of downstream events. Many hub proteins involve in multiple signaling pathways such as P53 and BRCA1. Mutation of the genes encoding these hub proteins could affect multiple pathways. Thus, targeting these genes could simultaneously activate or inhibit disease pathways.

In this analysis, 6 of 23 differentially expressed between LMP and HGSOC samples in two independent patient data sets also participated in subnetworks that were altered between tumor types in both data sets. Hence, these six genes (*STAT1*, *MYBL2*, *RPS23*, *NR2F1*, *SOX9*, and *SPRY2*) merit the highest priority for further study. Furthermore, our results suggest several hypotheses regarding disease mechanisms. For instance, *CREBBP* (*CBP*) interactions with *STAT1* and *MYBL2* appeared in three altered subnetworks, suggesting that each of these subnetworks is disrupted through *CREBBP* dysfunction. *CREBBP* mutations are documented in HGSOC (11), and our results suggest that *CREBBP* may contains causative mutations or homozygous deletions in the expression data sets we studied.

On the basis of a static GO analysis of differentially expressed genes concordant in both GEO data sets, we infer that RAR activation and AhR signaling pathways may be disrupted in HGSOC. However, our dynamic subnetwork analysis suggests that other pathways are differentially affected as well, including the P53, ERBB, chemokine, MAPK, and B-cell receptor signaling pathways. Consistent with our results, the P53, ERBB, and MAPK signaling pathways have documented regulatory roles in ovarian carcinomas (63–66). In addition, several genes mutated in HGSOC, such as *TP53*, *BRCA1*, and *MYC*, were identified as participants in significant subnetworks. Our results suggested that the network method, which combines expression profile and protein-protein interactions, can infer disease genes carried causative mutations and regulated the expression levels of the downstream genes.

In addition, by combining our network analysis results with HGSOC genotype data from TCGA, we found co-occurring mutations in *RYR1* (which encodes the ryanodine receptor) and *PAK4*, a result with potential clinical applications. *RYR1* alteration has been implicated in various types of disease, including cancer [79]. Presently, *RYR1* is a target of four FDA-approved drugs, including procaine, dantrolene, suramin, and caffeine (**S7 Fig**). Our survival analysis suggests that HGSOC patients with *PAK4* mutations have poor survival rates, but currently there is no drug that specifically targets *PAK4*, and it is often very expensive and time-consuming to develop a new therapeutic. As mutations of *PAK4* and *RYR1* tend to co-occur in HGSOC, and the two genes might interact, we hypothesize that caffeine, a cancer drug that targets *RYR1* [80] (**S7 Fig**), may represent a useful intervention strategy to treat HGSOC patients with *PAK4* mutations. Caffeine has been reported to impact cell cycle function, trigger apoptosis or programmed cell death or and disturbe key cell cycle regulatory proteins[81]. In addition, *PAK4* lies downstream of EGFR in the ERBB signaling pathway; hence, the nine FDA-approved cancer drugs targeting EGFR should be assessed to see if they ameliorate the effects of *PAK4* mutations and could be used to treat HGSOC patients.

A previous study reported that LMP affects women at a younger age than invasive ovarian cancer [82]. Consistently, the age distribution of patients of the two datasets (GSE17308 and GSE9891) showed a similar trend. For GSE17308, LMP patients ranged in age from 25 to 76 years (mean = 52 years), whereas HGSOC patients ranged in age from 25 to 80 years (mean = 62 years). For GSE9891, LMP patients ranged in age from 22 to 79 years (median = 50 years), whereas HGSOC patients ranged in age from 23 to 80 years (median = 59 years). In the original study that analyzed GSE9891[12], the author found that the age difference between the subtypes was significantly different (Kruskal-Wallis test for age as continuous variable, $P = 0.003$). However, the original study that analyzed GSE17308 showed that the difference did not quite reach significance ($P = 0.08$, Kruskal-Wallist test) and had no observable effect on gene expression profiles in unsupervised clustering analysis or in supervised ANOVA analyses [9]. The age difference might impact the gene expression analysis in some datasets. However, in this study our goal is to identify the common differentially expressed genes. We found that the common differentially expressed genes of the two datasets (GSE17308 and GSE9891) successfully clustered LMP and HGSOC patients in the third independent data (GSE27651, **S3 Fig**), suggesting the age influence on gene expression analysis is minimized in our approach.

In this study, network methods implicated a larger number of known ovarian cancer susceptibility genes than differential expression analysis. Given that ovarian cancer is a highly heterogeneous disease, the insights gained through network analyses may improve our understanding of the biological mechanisms involved in disease development and progression. Comparisons between LMP and HGSOC tumors using network approaches indicate that cellular regulatory pathways are wired differently between these tumor types. By considering LMP samples as the reference group, our results provide insight into the mechanisms responsible for the formation and progression of malignant ovarian cancer. More broadly, our work demonstrates that differential gene expression translates into altered network communication at the protein level. Hence, network models, which integrate multilayer information, foster the identification of genomic mutations and aberrant pathways, while facilitating the development of strategies for disease detection and points of intervention.

Conclusions

A comprehensive catalog of biomarkers is critical for improving our understanding and treatment of HGSOC. Our systematic, systems biology comparison of LMP and HGSOC tumors provides new insights into probable mechanisms underlying malignancy. Integrating gene

expression profiles from two independent patient data sets with PPI data allowed us to identify a set of biomarkers that can be used to distinguish between the two tumor types. We then applied these molecular signatures to a third, independent data set to demonstrate their utility. In addition, by combining information on networks altered between the two types of tumors with mutation data, we were able to prioritize mutations for further examination. Our results underscore the strength of systems biology approaches in implicating novel disease mechanisms, and our network approach is especially valuable when a single type of tumor displays recurrent alterations to the same pathways but varies in terms of the individual mutations responsible.

Materials and Methods

Data Availability and Online Tools

Subnetwork data have been deposited on the website http://mqyang.net/CancerResearch/HGSOC_Biomarker2.cgi. A web tool that allows users to stratify ovarian cancer samples on the basis of expression data, are available on our website http://mqyang.net/CancerResearch/ClusterTissues.cgi. Supplemental data including our gene lists and related tools are available at http://mqyang.net/CancerResearch/HGSOC_Biomarker2.cgi.

Disease-related subnetwork and hub proteins are searchable on the website, and data from external databases, such as OMIM and TCGA, are linked to the query genes as well. In addition, each gene is linked to external databases, including the UCSC genome browser, KEGG pathways, and the genetic mutations identified in TCGA.

Expression profiles

Two ovarian tumor expression profiles, GSE17308 [9] and GSE9891 [12], were obtained from the GEO database. In GSE17308, microarray expression profiling was conducted using the PC human Operon 21k v2 platform; 7 LMP and 22 HGSOC samples were collected from patients who were diagnosed with ovarian cancer and treated at the Royal Brisbane and Women's Hospital. LMP patients ranged in age from 25 to 76 years (mean = 52 years), whereas HGSOC patients ranged in age from 25 to 80 years for GSE17308. Experienced pathologists independently reviewed all tumor tissues. GSE9891 contains 18 LMP and 118 HGSOC samples from the AOCS (Australian Ovarian Cancer Study); here, profiling was carried out on well-characterized ovarian cancer tissues from patients with the Affymetrix U133_plus2.0 platform. LMP patients ranged in age from 22 to 79 years (median = 50 years), whereas HGSOC patients ranged in age from 23 to 80 years (median = 59 years) for GSE9891. Because two distinct array platforms were compared, official gene symbols were used to identify genes present in both data sets. Duplicate genes and low-quality data were removed from the analysis, leaving a total of 9,016 genes present in both profiles for evaluation. A third expression profile, GSE27651 [29], was used for validation. It contained 8 LMP and 24 HGSOC samples from the archives of the Department of Pathology at The University of Texas MD Anderson Cancer Center (Houston, Texas). This profile was generated with the commercial GeneChip Human Genome U133 Plus 2.0 Array [29].

Reproducibility of Differentially Expressed Genes

Four methods of ranking differentially expressed genes in the GSE17308 and GSE9891 data sets were compared: (i) fold change calculated with sample means, (ii) fold change calculated with sample medians, (iii) t-test *P*-values, and (iv) Wilcoxon rank-sum test *P*-values. After genes were ranked either by *P*-value or fold change, the reproducibility rate was calculated as

the percent overlap of top-ranked genes in the independent GEO profiles. To estimate the null reproducibility rate, we randomly extracted genes from both profiles for the overlap calculation 1,000 times.

Human Protein-Protein Interaction Data

We combined human PPI data from five public databases: IntAct, MINT, BioGrid, DIP, and HPRD [82–86]. Each database contains PPIs curated by experts. After removing redundant entries, we obtained a total of 122,403 unique human PPIs.

Construction of Distinguishing Subnetworks

To construct a subnetwork, expression levels of genes in the profile were first normalized so that the mean and variance across samples were 0 and 1, respectively:

$$Z_{i,j} = \frac{x_{i,j} - \mu_i}{\sigma_i} \quad (1)$$

where $Z_{i,j}$ represents the normalized expression value of the *ith* gene for the *jth* sample. The expression level Z_j of a network of n interacting genes was obtained by averaging expression values over all n genes in the *jth* sample, as follows:

$$\bar{Z}_j = \frac{1}{n-1} \sum_{i=1}^{1} Z_{i,j} \quad (2)$$

We then used mutual information to estimate the ability of each subnetwork to identify distinct phenotypes. Mutual information quantifies the degree to which two random variables are independent. When a random variable, X, is independent of another random variable, Y, $I(X; Y) = 0$. When applied to subnetworks, the larger the mutual information value, MI, the greater the discrimination power of the subnetwork. Thus, the DS of a subnetwork was defined as MI, given by

$$MI(X;Y) = \sum_{x \in X} \sum_{y \in Y} P\{X = x,\ Y = y\} \log_2 \left(\frac{P\{X = x, Y = y\}}{P\{X = x\} P\{Y = y\}} \right) \quad (3)$$

where X refers to the average normalized expression level of the subnetwork and Y refers to the tissue phenotype. In the equation above, X is assumed to take discrete values, but the average normalized expression level, Z, defined by (2), is not a discrete variable. We therefore discretized Z by dividing its range into equally spaced bins defined by split points, s_k, resulting in the following expression for the DS:

$$DS(\bar{Z};Y) = \sum_{k=1}^{m} \sum_{y \in Y} P\{s_k < \bar{Z} \le s_{k=1}, Y = y\} \log_2 \left(\frac{P\{s_k < \bar{Z} \le s_{k=1}, Y = y\}}{P\{s_k < \bar{Z} \le s_{k+1}\} P\{Y = y\}} \right) \quad (4)$$

To cover all values, we took $s_1 = min(Z) - \delta$, and $s_{m+1} = max(Z) + \delta$. The number of bins was $m = \log_2(\#\text{ samples}) + 1$.

The growth of each subnetwork was guided by a greedy algorithm in an iterative procedure. At each iteration, genes that neighbored at least one gene in the network were candidates for addition.

For each candidate gene, the DS was evaluated using the average expression of that gene and the genes in the current subnetwork. Among all candidate genes, the one that generated the largest DS was selected and added to the current network. The search procedure was terminated when the improvement rate, defined as the ratio of the winning DS for successive

iterations, either increased or was <0.1, or if the distance between the winning gene and the seed of the network was >2.

To test the hypothesis that a subnetwork could distinguish between different phenotypes, we obtained two DS null distributions by selecting two groups of genes with the same number of members as the given subnetwork, one with the same seed protein and the other with a different seed protein; we repeated this process 10,000 times. To test the hypothesis that genes in a subnetwork were associated with a particular phenotype, we constructed a DS null distribution by randomly permuting the phenotypes of tissues 10,000 times. Subnetworks with *P*-values $<10^{-4}$ in both tests and DS >0.66 were selected.

Identification of Significant Hub Proteins

Hub proteins were required to have at least five interactions. These hubs represented approximately the top 20% of all proteins in terms of number of interactions. For each hub (*H*), the difference in the PCC between LMP and HGSOC samples for an individual interaction (*I*) was calculated as follows:

$$\nabla r_{H,I} = \frac{\sum_j (I_{j,L} - \bar{I}_L)(H_{j,L} - \bar{H}_L)}{(n_L - 1) S_{I_L} S_{H_L}} - \frac{\sum_j (I_{j,C} - \bar{I}_C)(H_{j,C} - \bar{H}_C)}{(n_C - 1) S_{I_C} S_{H_C}} \quad (5)$$

,where *L* and *C* denote LMP and HGSOC, respectively, and *S* represents the standard deviation. The average of these PCC differences is given by

$$\text{AvgPCC} = \frac{1}{m-1} \sum_{i=1}^{m} |\nabla r_{H,I_i}| \quad (6)$$

where *m* is the total number of interactions of the hub. To test the hypothesis that hub protein modularity is significantly altered by disease type, we constructed a null distribution by randomly shuffling the tissue phenotypes 1,000 times. Hubs with *P*-values <0.05 were selected for our study. In addition, by searching known PPIs, the proteins that had the most frequent interactions with isolated hub proteins or protein clusters were identified and used to bridge these isolated instances into a larger network.

Cross-check with Literature, Pathway Databases and Know Disease Genes

We employed the GeneIndexer webtool (http://geneindexer.com/) to examine and validate functional relationships among the significant hub proteins implicated in both data sets, as well as three well-known cancer genes (*TP53*, *BRCA1 and BRCA2*). On the basis of the scientific literature, GeneIndexer generates a tree by clustering functionally related genes together. In addition, we performed a hypergeometric test to identify the KEGG pathways enriched in gene clusters from the subnetwork analysis. The hypergeometric test was also used to find GO terms that were significantly associated with each discriminative subnetwork. Canonical pathways enriched by differentially expressed genes were detected using Ingenuity's IPA software.

We identified our list of 18 known ovarian cancer susceptibility genes by downloading gene map file from the OMIM database and searching on the phrase "ovarian cancer susceptibility". The 288 cancer genes were obtained directly from the COSMIC database.

Genetic Mutations in Differentially Expressed Genes and Significant Hub Proteins

Genetic mutations in differentially expressed genes and significant hub proteins, which were concordant between expression data sets, were detected using mutation data from a total of 316 sequenced HGSOC tissue samples from TCGA [16]. The statistical significance of the mutual exclusivity and co-occurrence of genetic mutations in gene pairs was assessed using Fisher's exact test. We used the cBioPortal [87, 88] tool to perform these analyses.

Prediction of Tumor Types

To evaluate the prediction performance of the molecular signatures identified, we randomly selected two-thirds of the data for training and used the remaining one-third for testing; we repeated this procedure 50 times. The ROC curve and the AUC were plotted and analyzed using the R package ROCR [89]. The features used for classification for the three different types of molecular signatures were calculated as followings: (1) the expression levels of differentially expressed genes, (2) the average expression level of genes comprising discriminative subnetworks, (3) the expression correlation changes of significant hub proteins with their interacting genes.

Statistical Tests and Network Visualization

We used the R software package to perform statistical tests. Cytoscape 2.8.1 [90] was used to visualize networks.

Supporting Information

S1 Fig. Low-malignant-potential (LMP) and high-grade serous carcinoma (HGSOC) samples, clustered based on the expression levels of top-ranked, differentially expressed genes identified using Wilcox rank-sum test *P*-values or median fold-change. (A) represents tissue clusters for expression data sets GSE9891 (*top*) and GSE17308 (*bottom*) obtained using expression levels of top genes ranked by the Wilcox rank-sum test. (B) represents tissue clusters for GSE9891 (*top*) and GSE17308 (*bottom*) obtained using expression levels of top genes ranked by median fold-change.
(PDF)

S2 Fig. The *P*-values and median fold changes for gene expression in low-malignant-potential (LMP) and high-grade serous carcinoma (HGSOC) samples in the GEO data sets GSE17308 (*left*) and GSE9891 (*right*). The red circle represent the top differentially expressed genes selected by P-value and fold change.
(PDF)

S3 Fig. Hierarchical cluster analysis on an independent group of samples. Based on expression levels of 23 genes differentially expressed in low-malignant-potentihujujiijnal (LMP) and high-grade serous carcinoma (HGSOC) samples in both the GSE9891 and GSE17308 data sets, samples from a third, independent patient data set (GSE27651) were separated into two homogenous sets. (A) represents a heatmap of tissue clusters for the GSE27651 data set, and (B) represents a hierarchical tree of the same data.
(PDF)

S4 Fig. The cell cycle, drug metabolism, and molecular transport network is enriched with genes differentially expressed between low-malignant-potential (LMP) and high-grade serous carcinoma (HGSOC) samples in both the GSE9891 and GSE17308 data sets. Red

indicates overexpression in HGSOC and green represents underexpression. Genes highlighted in yellow participate in RAR activation and AhR signaling.
(PDF)

S5 Fig. The functional relationships among genes that encode common protein hubs differentially expressed between low-malignant-potential (LMP) and high-grade serous carcinoma (HGSOC) samples in both the GSE9891 and GSE17308 data sets. Using GeneIndexer, which can search over one million Entrez Gene abstracts to identify mechanistic functional relationships among genes, a functional hierarchy tree was constructed. It suggests that the known ovarian cancer genes *BRCA1*, *BRCA2*, and *TP53* have the strongest functional relationships with each other, followed by *APEX1*, a protein hub gene identified in this study.
(PDF)

S6 Fig. ERBB signal pathway. *PAK4* is located downstream of *EGFR* in the ERBB signalling pathway.
(PDF)

S7 Fig. *PAK4* and *RYR1* protein interaction networks suggest that the two proteins interact indirectly. The yellow octagons respresent FDA approved drugs target the corresponding gene.
(PDF)

S8 Fig. The total number of molecular signatures by differential methods.
(PDF)

S9 Fig. Comparison of the ability of differentially expressed genes, subnetwork connections, and significant hubs to identify known cancer genes. We compiled a list of ovarian cancer susceptibility genes affected by somatic and germline mutations from the OMIM database, and also a list of general cancer genes that carry somatic and germline mutations from the COSMIC database. Subnetwork and hub protein analyses did not reveal significantly more known cancer genes than differential expression analysis in the ovarian cancer gene set, but they did reveal significantly more of the general cancer genes.
(PDF)

S10 Fig. ROC curves of support vector machine (SVM)-based classifiers. The purple line represents the classification performance ROC curve for expression profile data set GSE9891, whereas the green line represents that for expression profile data set GSE17308.
(PDF)

S1 Table. The significant subnetworks that differentiate HGSOC from LMP.
(DOCX)

S2 Table. The common significant hub proteins.
(DOCX)

S3 Table. The performance of classification using different type molecular signature as features.
(DOCX)

Acknowledgments

We thank the NIH Fellows Editorial Board and Kristin Harper for editorial assistance.

Author Contributions

Conceptualization: MQY.

Data curation: MQY.

Formal analysis: MQY.

Funding acquisition: MQY LE.

Investigation: MQY LE.

Methodology: MQY LE.

Project administration: MQY LE.

Resources: MQY LE.

Software: MQY.

Supervision: MQY LE.

Validation: MQY LE.

Visualization: MQY LE.

Writing – original draft: MQY.

Writing – review & editing: MQY LE.

References

1. Siegel R, Ma J, Zou Z, Jemal A: Cancer statistics, 2014. CA Cancer J Clin 2014, 64(1):9–29. doi: 10.3322/caac.21208 PMID: 24399786
2. Malpica A, Deavers MT, Lu K, Bodurka DC, Atkinson EN, Gershenson DM, Silva EG: Grading ovarian serous carcinoma using a two-tier system. The American journal of surgical pathology 2004, 28 (4):496–504. PMID: 15087669
3. Ayhan A, Kurman RJ, Yemelyanova A, Vang R, Logani S, Seidman JD, Shih Ie M: Defining the cut point between low-grade and high-grade ovarian serous carcinomas: a clinicopathologic and molecular genetic analysis. The American journal of surgical pathology 2009, 33(8):1220–1224. doi: 10.1097/PAS.0b013e3181a24354 PMID: 19461510
4. Dehari R, Kurman RJ, Logani S, Shih Ie M: The development of high-grade serous carcinoma from atypical proliferative (borderline) serous tumors and low-grade micropapillary serous carcinoma: a morphologic and molecular genetic analysis. The American journal of surgical pathology 2007, 31 (7):1007–1012. doi: 10.1097/PAS.0b013e31802cbbe9 PMID: 17592266
5. Malpica A, Deavers MT, Tornos C, Kurman RJ, Soslow R, Seidman JD, Munsell MF, Gaertner E, Frishberg D, Silva EG: Interobserver and intraobserver variability of a two-tier system for grading ovarian serous carcinoma. The American journal of surgical pathology 2007, 31(8):1168–1174. doi: 10.1097/PAS.0b013e31803199b0 PMID: 17667538
6. Santillan A, Kim YW, Zahurak ML, Gardner GJ, Giuntoli RL 2nd, Shih IM, Bristow RE: Differences of chemoresistance assay between invasive micropapillary/low-grade serous ovarian carcinoma and high-grade serous ovarian carcinoma. International journal of gynecological cancer: official journal of the International Gynecological Cancer Society 2007, 17(3):601–606.
7. Bonome T, Lee JY, Park DC, Radonovich M, Pise-Masison C, Brady J, Gardner GJ, Hao K, Wong WH, Barrett JC et al: Expression profiling of serous low malignant potential, low-grade, and high-grade tumors of the ovary. Cancer research 2005, 65(22):10602–10612. doi: 10.1158/0008-5472.CAN-05-2240 PMID: 16288054
8. Gilks CB, Vanderhyden BC, Zhu S, van de Rijn M, Longacre TA: Distinction between serous tumors of low malignant potential and serous carcinomas based on global mRNA expression profiling. Gynecologic oncology 2005, 96(3):684–694. doi: 10.1016/j.ygyno.2004.11.039 PMID: 15721412

9. Merritt MA, Parsons PG, Newton TR, Martyn AC, Webb PM, Green AC, Papadimos DJ, Boyle GM: Expression profiling identifies genes involved in neoplastic transformation of serous ovarian cancer. BMC cancer 2009, 9:378. doi: 10.1186/1471-2407-9-378 PMID: 19849863
10. Santin AD, Zhan F, Bellone S, Palmieri M, Cane S, Bignotti E, Anfossi S, Gokden M, Dunn D, Roman JJ et al: Gene expression profiles in primary ovarian serous papillary tumors and normal ovarian epithelium: identification of candidate molecular markers for ovarian cancer diagnosis and therapy. International journal of cancer Journal international du cancer 2004, 112(1):14–25. doi: 10.1002/ijc.20408 PMID: 15305371
11. Spentzos D, Levine DA, Ramoni MF, Joseph M, Gu X, Boyd J, Libermann TA, Cannistra SA: Gene expression signature with independent prognostic significance in epithelial ovarian cancer. J Clin Oncol 2004, 22(23):4700–4710. doi: 10.1200/JCO.2004.04.070 PMID: 15505275
12. Tothill RW, Tinker AV, George J, Brown R, Fox SB, Lade S, Johnson DS, Trivett MK, Etemadmoghadam D, Locandro B et al: Novel molecular subtypes of serous and endometrioid ovarian cancer linked to clinical outcome. Clinical cancer research: an official journal of the American Association for Cancer Research 2008, 14(16):5198–5208.
13. Welsh JB, Zarrinkar PP, Sapinoso LM, Kern SG, Behling CA, Monk BJ, Lockhart DJ, Burger RA, Hampton GM: Analysis of gene expression profiles in normal and neoplastic ovarian tissue samples identifies candidate molecular markers of epithelial ovarian cancer. Proceedings of the National Academy of Sciences of the United States of America 2001, 98(3):1176–1181. doi: 10.1073/pnas.98.3.1176 PMID: 11158614
14. Shlomi T, Cabili MN, Herrgard MJ, Palsson BO, Ruppin E: Network-based prediction of human tissue-specific metabolism. Nature biotechnology 2008, 26(9):1003–1010. doi: 10.1038/nbt.1487 PMID: 18711341
15. Bossi A, Lehner B: Tissue specificity and the human protein interaction network. Molecular systems biology 2009, 5:260.
16. Integrated genomic analyses of ovarian carcinoma. Nature 2011, 474(7353):609–615. doi: 10.1038/nature10166 PMID: 21720365
17. Barabasi AL, Oltvai ZN: Network biology: understanding the cell's functional organization. Nature reviews Genetics 2004, 5(2):101–113. doi: 10.1038/nrg1272 PMID: 14735121
18. Papin JA, Hunter T, Palsson BO, Subramaniam S: Reconstruction of cellular signalling networks and analysis of their properties. Nature reviews Molecular cell biology 2005, 6(2):99–111. doi: 10.1038/nrm1570 PMID: 15654321
19. Simeone P, Trerotola M, Urbanella A, Lattanzio R, Ciavardelli D, Di Giuseppe F, Eleuterio E, Sulpizio M, Eusebi V, Pession A et al: A unique four-hub protein cluster associates to glioblastoma progression. PloS one 2014, 9(7):e103030. doi: 10.1371/journal.pone.0103030 PMID: 25050814
20. Timofeev O, Schlereth K, Wanzel M, Braun A, Nieswandt B, Pagenstecher A, Rosenwald A, Elsasser HP, Stiewe T: p53 DNA binding cooperativity is essential for apoptosis and tumor suppression in vivo. Cell reports 2013, 3(5):1512–1525. doi: 10.1016/j.celrep.2013.04.008 PMID: 23665223
21. McCole DF, Truong A, Bunz M, Barrett KE: Consequences of direct versus indirect activation of epidermal growth factor receptor in intestinal epithelial cells are dictated by protein-tyrosine phosphatase 1B. The Journal of biological chemistry 2007, 282(18):13303–13315. doi: 10.1074/jbc.M700424200 PMID: 17339316
22. Kurant M, Thiran P, Hagmann P: Error and attack tolerance of layered complex networks. Physical review E, Statistical, nonlinear, and soft matter physics 2007, 76(2 Pt 2):026103.
23. Han JD, Bertin N, Hao T, Goldberg DS, Berriz GF, Zhang LV, Dupuy D, Walhout AJ, Cusick ME, Roth FP et al: Evidence for dynamically organized modularity in the yeast protein-protein interaction network. Nature 2004, 430(6995):88–93. doi: 10.1038/nature02555 PMID: 15190252
24. Jeong H, Mason SP, Barabasi AL, Oltvai ZN: Lethality and centrality in protein networks. Nature 2001, 411(6833):41–42. doi: 10.1038/35075138 PMID: 11333967
25. Shvartsman HS, Sun CC, Bodurka DC, Mahajan V, Crispens M, Lu KH, Deavers MT, Malpica A, Silva EG, Gershenson DM: Comparison of the clinical behavior of newly diagnosed stages II-IV low-grade serous carcinoma of the ovary with that of serous ovarian tumors of low malignant potential that recur as low-grade serous carcinoma. Gynecologic oncology 2007, 105(3):625–629. doi: 10.1016/j.ygyno.2007.01.030 PMID: 17320156
26. Smith Sehdev AE, Sehdev PS, Kurman RJ: Noninvasive and invasive micropapillary (low-grade) serous carcinoma of the ovary: a clinicopathologic analysis of 135 cases. The American journal of surgical pathology 2003, 27(6):725–736. PMID: 12766576
27. Guo L, Lobenhofer EK, Wang C, Shippy R, Harris SC, Zhang L, Mei N, Chen T, Herman D, Goodsaid FM et al: Rat toxicogenomic study reveals analytical consistency across microarray platforms. Nature biotechnology 2006, 24(9):1162–1169. PMID: 17061323

28. Patterson TA, Lobenhofer EK, Fulmer-Smentek SB, Collins PJ, Chu TM, Bao W, Fang H, Kawasaki ES, Hager J, Tikhonova IR et al: Performance comparison of one-color and two-color platforms within the MicroArray Quality Control (MAQC) project. Nature biotechnology 2006, 24(9):1140–1150. doi: 10.1038/nbt1242 PMID: 16964228

29. King ER, Tung CS, Tsang YT, Zu Z, Lok GT, Deavers MT, Malpica A, Wolf JK, Lu KH, Birrer MJ et al: The anterior gradient homolog 3 (AGR3) gene is associated with differentiation and survival in ovarian cancer. The American journal of surgical pathology 2011, 35(6):904–912. doi: 10.1097/PAS.0b013e318212ae22 PMID: 21451362

30. Sosa MS, Bragado P, Aguirre-Ghiso JA: Mechanisms of disseminated cancer cell dormancy: an awakening field. Nature reviews Cancer 2014, 14(9):611–622. doi: 10.1038/nrc3793 PMID: 25118602

31. Moghaddam SM, Amini A, Wei AQ, Pourgholami MH, Morris DL: Initial report on differential expression of sprouty proteins 1 and 2 in human epithelial ovarian cancer cell lines. Journal of oncology 2012, 2012:373826. doi: 10.1155/2012/373826 PMID: 23251157

32. Kramer OH, Heinzel T: Phosphorylation-acetylation switch in the regulation of STAT1 signaling. Molecular and cellular endocrinology 2010, 315(1–2):40–48. doi: 10.1016/j.mce.2009.10.007 PMID: 19879327

33. Kramer OH, Knauer SK, Greiner G, Jandt E, Reichardt S, Guhrs KH, Stauber RH, Bohmer FD, Heinzel T: A phosphorylation-acetylation switch regulates STAT1 signaling. Genes & development 2009, 23 (2):223–235.

34. Sieben NL, Oosting J, Flanagan AM, Prat J, Roemen GM, Kolkman-Uljee SM, van Eijk R, Cornelisse CJ, Fleuren GJ, van Engeland M: Differential gene expression in ovarian tumors reveals Dusp 4 and Serpina 5 as key regulators for benign behavior of serous borderline tumors. J Clin Oncol 2005, 23 (29):7257–7264.

35. Tanner MM, Grenman S, Koul A, Johannsson O, Meltzer P, Pejovic T, Borg A, Isola JJ: Frequent amplification of chromosomal region 20q12-q13 in ovarian cancer. Clinical cancer research: an official journal of the American Association for Cancer Research 2000, 6(5):1833–1839.

36. Chitale D, Gong Y, Taylor BS, Broderick S, Brennan C, Somwar R, Golas B, Wang L, Motoi N, Szoke J et al: An integrated genomic analysis of lung cancer reveals loss of DUSP4 in EGFR-mutant tumors. Oncogene 2009, 28(31):2773–2783. doi: 10.1038/onc.2009.135 PMID: 19525976

37. Chan CY, Kim PM, Winn LM: TCDD-induced homologous recombination: the role of the Ah receptor versus oxidative DNA damage. Mutation research 2004, 563(1):71–79. doi: 10.1016/j.mrgentox.2004.05.015 PMID: 15324750

38. Witkowski L, Carrot-Zhang J, Albrecht S, Fahiminiya S, Hamel N, Tomiak E, Grynspan D, Saloustros E, Nadaf J, Rivera B et al: Germline and somatic SMARCA4 mutations characterize small cell carcinoma of the ovary, hypercalcemic type. Nature genetics 2014, 46(5):438–443. doi: 10.1038/ng.2931 PMID: 24658002

39. Yeung TL, Leung CS, Wong KK, Samimi G, Thompson MS, Liu J, Zaid TM, Ghosh S, Birrer MJ, Mok SC: TGF-beta modulates ovarian cancer invasion by upregulating CAF-derived versican in the tumor microenvironment. Cancer research 2013, 73(16):5016–5028. doi: 10.1158/0008-5472.CAN-13-0023 PMID: 23824740

40. Cheng JC, Auersperg N, Leung PC: TGF-beta induces serous borderline ovarian tumor cell invasion by activating EMT but triggers apoptosis in low-grade serous ovarian carcinoma cells. PloS one 2012, 7(8):e42436. doi: 10.1371/journal.pone.0042436 PMID: 22905131

41. Lok GT, Chan DW, Liu VW, Hui WW, Leung TH, Yao KM, Ngan HY: Aberrant activation of ERK/FOXM1 signaling cascade triggers the cell migration/invasion in ovarian cancer cells. PloS one 2011, 6(8):e23790. doi: 10.1371/journal.pone.0023790 PMID: 21858223

42. Luo H, Jiang BH, King SM, Chen YC: Inhibition of cell growth and VEGF expression in ovarian cancer cells by flavonoids. Nutrition and cancer 2008, 60(6):800–809. doi: 10.1080/01635580802100851 PMID: 19005980

43. Wang J, Zhou JY, Wu GS: ERK-dependent MKP-1-mediated cisplatin resistance in human ovarian cancer cells. Cancer research 2007, 67(24):11933–11941. doi: 10.1158/0008-5472.CAN-07-5185 PMID: 18089824

44. Dhillon AS, Hagan S, Rath O, Kolch W: MAP kinase signalling pathways in cancer. Oncogene 2007, 26(22):3279–3290. doi: 10.1038/sj.onc.1210421 PMID: 17496922

45. Santin AD, Hermonat PL, Ravaggi A, Cannon MJ, Pecorelli S, Parham GP: Secretion of vascular endothelial growth factor in ovarian cancer. European journal of gynaecological oncology 1999, 20 (3):177–181. PMID: 10410879

46. Chuang HY, Lee E, Liu YT, Lee D, Ideker T: Network-based classification of breast cancer metastasis. Molecular systems biology 2007, 3:140. doi: 10.1038/msb4100180 PMID: 17940530

47. Schadt EE, Lamb J, Yang X, Zhu J, Edwards S, Guhathakurta D, Sieberts SK, Monks S, Reitman M, Zhang C et al: An integrative genomics approach to infer causal associations between gene expression and disease. Nature genetics 2005, 37(7):710–717. doi: 10.1038/ng1589 PMID: 15965475

48. Ahmed AA, Etemadmoghadam D, Temple J, Lynch AG, Riad M, Sharma R, Stewart C, Fereday S, Caldas C, Defazio A et al: Driver mutations in TP53 are ubiquitous in high grade serous carcinoma of the ovary. The Journal of pathology 2010, 221(1):49–56. doi: 10.1002/path.2696 PMID: 20229506

49. Rivlin N, Brosh R, Oren M, Rotter V: Mutations in the p53 Tumor Suppressor Gene: Important Milestones at the Various Steps of Tumorigenesis. Genes & cancer 2011, 2(4):466–474.

50. Supek F, Minana B, Valcarcel J, Gabaldon T, Lehner B: Synonymous mutations frequently act as driver mutations in human cancers. Cell 2014, 156(6):1324–1335. doi: 10.1016/j.cell.2014.01.051 PMID: 24630730

51. Alsop K, Fereday S, Meldrum C, deFazio A, Emmanuel C, George J, Dobrovic A, Birrer MJ, Webb PM, Stewart C et al: BRCA mutation frequency and patterns of treatment response in BRCA mutation-positive women with ovarian cancer: a report from the Australian Ovarian Cancer Study Group. J Clin Oncol 2012, 30(21):2654–2663. doi: 10.1200/JCO.2011.39.8545 PMID: 22711857

52. Anglesio MS, Arnold JM, George J, Tinker AV, Tothill R, Waddell N, Simms L, Locandro B, Fereday S, Traficante N et al: Mutation of ERBB2 provides a novel alternative mechanism for the ubiquitous activation of RAS-MAPK in ovarian serous low malignant potential tumors. Molecular cancer research: MCR 2008, 6(11):1678–1690. doi: 10.1158/1541-7786.MCR-08-0193 PMID: 19010816

53. Haverty PM, Hon LS, Kaminker JS, Chant J, Zhang Z: High-resolution analysis of copy number alterations and associated expression changes in ovarian tumors. BMC medical genomics 2009, 2:21. doi: 10.1186/1755-8794-2-21 PMID: 19419571

54. Hoffman B, Liebermann DA: Apoptotic signaling by c-MYC. Oncogene 2008, 27(50):6462–6472. doi: 10.1038/onc.2008.312 PMID: 18955973

55. Dang CV: c-Myc target genes involved in cell growth, apoptosis, and metabolism. Molecular and cellular biology 1999, 19(1):1–11. PMID: 9858526

56. Karst AM, Levanon K, Drapkin R: Modeling high-grade serous ovarian carcinogenesis from the fallopian tube. Proceedings of the National Academy of Sciences of the United States of America 2011, 108(18):7547–7552. doi: 10.1073/pnas.1017300108 PMID: 21502498

57. de Graeff P, Crijns AP, Ten Hoor KA, Klip HG, Hollema H, Oien K, Bartlett JM, Wisman GB, de Bock GH, de Vries EG et al: The ErbB signalling pathway: protein expression and prognostic value in epithelial ovarian cancer. British journal of cancer 2008, 99(2):341–349. doi: 10.1038/sj.bjc.6604471 PMID: 18628764

58. Pohl G, Ho CL, Kurman RJ, Bristow R, Wang TL, Shih Ie M: Inactivation of the mitogen-activated protein kinase pathway as a potential target-based therapy in ovarian serous tumors with KRAS or BRAF mutations. Cancer research 2005, 65(5):1994–2000. doi: 10.1158/0008-5472.CAN-04-3625 PMID: 15753399

59. Hashiguchi Y, Tsuda H, Yamamoto K, Inoue T, Ishiko O, Ogita S: Combined analysis of p53 and RB pathways in epithelial ovarian cancer. Human pathology 2001, 32(9):988–996. doi: 10.1053/hupa.2001.27115 PMID: 11567230

60. Green JA, Berns EM, Coens C, van Luijk I, Thompson-Hehir J, van Diest P, Verheijen RH, van de Vijver M, van Dam P, Kenter GG et al: Alterations in the p53 pathway and prognosis in advanced ovarian cancer: a multi-factorial analysis of the EORTC Gynaecological Cancer group (study 55865). European journal of cancer (Oxford, England: 1990) 2006, 42(15):2539–2548.

61. Pils D, Hager G, Tong D, Aust S, Heinze G, Kohl M, Schuster E, Wolf A, Sehouli J, Braicu I et al: Validating the impact of a molecular subtype in ovarian cancer on outcomes: a study of the OVCAD Consortium. Cancer science 2012, 103(7):1334–1341. doi: 10.1111/j.1349-7006.2012.02306.x PMID: 22497737

62. Dong HP, Elstrand MB, Holth A, Silins I, Berner A, Trope CG, Davidson B, Risberg B: NK- and B-cell infiltration correlates with worse outcome in metastatic ovarian carcinoma. American journal of clinical pathology 2006, 125(3):451–458. PMID: 16613351

63. Fraser HB, Plotkin JB: Using protein complexes to predict phenotypic effects of gene mutation. Genome biology 2007, 8(11):R252. doi: 10.1186/gb-2007-8-11-r252 PMID: 18042286

64. Makino T, Gojobori T: The evolutionary rate of a protein is influenced by features of the interacting partners. Molecular biology and evolution 2006, 23(4):784–789. doi: 10.1093/molbev/msj090 PMID: 16407461

65. Taylor IW, Linding R, Warde-Farley D, Liu Y, Pesquita C, Faria D, Bull S, Pawson T, Morris Q, Wrana JL: Dynamic modularity in protein interaction networks predicts breast cancer outcome. Nature biotechnology 2009, 27(2):199–204. doi: 10.1038/nbt.1522 PMID: 19182785

66. Cai JJ, Borenstein E, Petrov DA: Broker genes in human disease. Genome biology and evolution 2010, 2:815–825. doi: 10.1093/gbe/evq064 PMID: 20937604

67. Armaiz-Pena GN, Allen JK, Cruz A, Stone RL, Nick AM, Lin YG, Han LY, Mangala LS, Villares GJ, Vivas-Mejia P et al: Src activation by beta-adrenoreceptors is a key switch for tumour metastasis. Nature communications 2013, 4:1403. doi: 10.1038/ncomms2413 PMID: 23360994

68. Colomiere M, Ward AC, Riley C, Trenerry MK, Cameron-Smith D, Findlay J, Ackland L, Ahmed N: Cross talk of signals between EGFR and IL-6R through JAK2/STAT3 mediate epithelial-mesenchymal transition in ovarian carcinomas. British journal of cancer 2009, 100(1):134–144. doi: 10.1038/sj.bjc.6604794 PMID: 19088723

69. Concin N, Hofstetter G, Berger A, Gehmacher A, Reimer D, Watrowski R, Tong D, Schuster E, Hefler L, Heim K et al: Clinical relevance of dominant-negative p73 isoforms for responsiveness to chemotherapy and survival in ovarian cancer: evidence for a crucial p53-p73 cross-talk in vivo. Clinical cancer research: an official journal of the American Association for Cancer Research 2005, 11(23):8372–8383.

70. Kelley MR, Luo M, Reed A, Su D, Delaplane S, Borch RF, Nyland RL 2nd, Gross ML, Georgiadis MM: Functional analysis of novel analogues of E3330 that block the redox signaling activity of the multifunctional AP endonuclease/redox signaling enzyme APE1/Ref-1. Antioxidants & redox signaling 2011, 14(8):1387–1401.

71. Fishel ML, Kelley MR: The DNA base excision repair protein Ape1/Ref-1 as a therapeutic and chemopreventive target. Molecular aspects of medicine 2007, 28(3–4):375–395. doi: 10.1016/j.mam.2007.04.005 PMID: 17560642

72. Comprehensive molecular characterization of human colon and rectal cancer. Nature 2012, 487(7407):330–337. doi: 10.1038/nature11252 PMID: 22810696

73. Taylor BS, Schultz N, Hieronymus H, Gopalan A, Xiao Y, Carver BS, Arora VK, Kaushik P, Cerami E, Reva B et al: Integrative genomic profiling of human prostate cancer. Cancer cell 2010, 18(1):11–22. doi: 10.1016/j.ccr.2010.05.026 PMID: 20579941

74. Abedin SA, Thorne JL, Battaglia S, Maguire O, Hornung LB, Doherty AP, Mills IG, Campbell MJ: Elevated NCOR1 disrupts a network of dietary-sensing nuclear receptors in bladder cancer cells. Carcinogenesis 2009, 30(3):449–456. doi: 10.1093/carcin/bgp005 PMID: 19126649

75. Zhang Z, Yamashita H, Toyama T, Sugiura H, Ando Y, Mita K, Hamaguchi M, Hara Y, Kobayashi S, Iwase H: NCOR1 mRNA is an independent prognostic factor for breast cancer. Cancer letters 2006, 237(1):123–129. doi: 10.1016/j.canlet.2005.05.046 PMID: 16019133

76. Liu Y, Xiao H, Tian Y, Nekrasova T, Hao X, Lee HJ, Suh N, Yang CS, Minden A: The pak4 protein kinase plays a key role in cell survival and tumorigenesis in athymic mice. Molecular cancer research: MCR 2008, 6(7):1215–1224. doi: 10.1158/1541-7786.MCR-08-0087 PMID: 18644984

77. Qu J, Cammarano MS, Shi Q, Ha KC, de Lanerolle P, Minden A: Activated PAK4 regulates cell adhesion and anchorage-independent growth. Molecular and cellular biology 2001, 21(10):3523–3533. doi: 10.1128/MCB.21.10.3523-3533.2001 PMID: 11313478

78. Tso PH, Yung LY, Wang Y, Wong YH: RGS19 stimulates cell proliferation by deregulating cell cycle control and enhancing Akt signaling. Cancer letters 2011, 309(2):199–208. doi: 10.1016/j.canlet.2011.06.002 PMID: 21705135

79. Monnier N, Ferreiro A, Marty I, Labarre-Vila A, Mezin P, Lunardi J: A homozygous splicing mutation causing a depletion of skeletal muscle RYR1 is associated with multi-minicore disease congenital myopathy with ophthalmoplegia. Human molecular genetics 2003, 12(10):1171–1178. PMID: 12719381

80. Rask-Andersen M, Almen MS, Schioth HB: Trends in the exploitation of novel drug targets. Nature reviews Drug discovery 2011, 10(8):579–590. doi: 10.1038/nrd3478 PMID: 21804595

81. Bode AM, Dong Z: The enigmatic effects of caffeine in cell cycle and cancer. Cancer letters 2007, 247(1):26–39. doi: 10.1016/j.canlet.2006.03.032 PMID: 16709440

82. Licata L, Briganti L, Peluso D, Perfetto L, Iannuccelli M, Galeota E, Sacco F, Palma A, Nardozza AP, Santonico E et al: MINT, the molecular interaction database: 2012 update. Nucleic acids research 2012, 40(Database issue):D857–861. doi: 10.1093/nar/gkr930 PMID: 22096227

83. Orchard S, Ammari M, Aranda B, Breuza L, Briganti L, Broackes-Carter F, Campbell NH, Chavali G, Chen C, del-Toro N et al: The MIntAct project—IntAct as a common curation platform for 11 molecular interaction databases. Nucleic acids research 2014, 42(Database issue):D358–363. doi: 10.1093/nar/gkt1115 PMID: 24234451

84. Stark C, Breitkreutz BJ, Reguly T, Boucher L, Breitkreutz A, Tyers M: BioGRID: a general repository for interaction datasets. Nucleic acids research 2006, 34(Database issue):D535–539. doi: 10.1093/nar/gkj109 PMID: 16381927

85. Xenarios I, Salwinski L, Duan XJ, Higney P, Kim SM, Eisenberg D: DIP, the Database of Interacting Proteins: a research tool for studying cellular networks of protein interactions. Nucleic acids research 2002, 30(1):303–305. PMID: 11752321

86. Keshava Prasad TS, Goel R, Kandasamy K, Keerthikumar S, Kumar S, Mathivanan S, Telikicherla D, Raju R, Shafreen B, Venugopal A et al: Human Protein Reference Database—2009 update. Nucleic acids research 2009, 37(Database issue):D767–772. doi: 10.1093/nar/gkn892 PMID: 18988627

87. Gao J, Aksoy BA, Dogrusoz U, Dresdner G, Gross B, Sumer SO, Sun Y, Jacobsen A, Sinha R, Larsson E et al: Integrative analysis of complex cancer genomics and clinical profiles using the cBioPortal. Science signaling 2013, 6(269):pl1. doi: 10.1126/scisignal.2004088 PMID: 23550210

88. Cerami E, Gao J, Dogrusoz U, Gross BE, Sumer SO, Aksoy BA, Jacobsen A, Byrne CJ, Heuer ML, Larsson E et al: The cBio cancer genomics portal: an open platform for exploring multidimensional cancer genomics data. Cancer discovery 2012, 2(5):401–404. doi: 10.1158/2159-8290.CD-12-0095 PMID: 22588877

89. Sing T, Sander O, Beerenwinkel N, Lengauer T: ROCR: visualizing classifier performance in R. Bioinformatics (Oxford, England) 2005, 21(20):3940–3941.

90. Saito R, Smoot ME, Ono K, Ruscheinski J, Wang PL, Lotia S, Pico AR, Bader GD, Ideker T: A travel guide to Cytoscape plugins. Nature methods 2012, 9(11):1069–1076. doi: 10.1038/nmeth.2212 PMID: 23132118