Widespread genetic epistasis among cancer genes

Xiaoyue Wang[1,2,*,†], Audrey Q. Fu[1,2,*], Megan E. McNerney[1,3] & Kevin P. White[1,2]

Quantitative genetic epistasis has been hypothesized to be an important factor in the development and progression of complex diseases. Cancers in particular are driven by the accumulation of mutations that may act epistatically during the course of the disease. However, as cancer mutations are uncovered at an unprecedented rate, determining which combinations of genetic alterations interact to produce cancer phenotypes remains a challenge. Here we show that by using combinatorial RNAi screening in cell culture, dense and often previously undetermined interactions among cancer genes were revealed by assessing gene pairs that are frequently co-altered in primary breast cancers. These interacting gene pairs are significantly associated with survival time when co-altered in patients, indicating that genetic interaction mapping may be leveraged to improve risk assessment. As many of these interacting gene pairs involve known drug targets, personalized treatment regimens may be improved by overlaying genetic interactions with mutational profiling.

[1] Institute of Genomics and Systems Biology, University of Chicago, Chicago, Illinois 60637, USA. [2] Department of Human Genetics, University of Chicago, Chicago, Illinois 60637, USA. [3] Department of Pathology, University of Chicago, Chicago, Illinois 60637, USA. * These authors contributed equally to this work. † Present address: State Key Laboratory of Medical Molecular Biology, Department of Biochemistry and Molecular Biology, Institute of Basic Medical Sciences, Chinese Academy of Medical Sciences, Peking Union Medical College, Beijing 100005, China. Correspondence and requests for materials should be addressed to K.P.W. (email: kpwhite@uchicago.edu).

Tumour cells may harbour many hundreds of alterations in their genomes, including amplifications, deletions, point mutations, loss of heterozygosity and epigenetic changes[1–3]. The most intensively studied genetic changes have been shown to interact during the complex process of tumour formation and progression[4]. For example, oncogenes such as *MYC* and *RAS* have long been recognized to cooperate in the processes of transformation and immortalization[5]. With the growing lists of genetic mutations generated by genome sequencing projects, it remains a significant challenge to determine which of these genetic alterations interact and whether those interactions are clinically relevant. Furthermore, it is unknown whether there are a few key commonly mutated factors such as *RAS* and *P53* that interact with large numbers of other factors (as hubs in oncogenic networks) or whether genetic interactions (GIs) are more widespread and common among even less frequently mutated genes.

The observation of an unexpected (for example, non-additive or non-multiplicative) phenotypic outcome when combining two genetic alleles is referred to as a GI or epistasis. Pairwise GIs for thousands of genes have been systematically mapped in yeast by using a large collection of deletion strains[6–10]. In higher organisms, combinatorial RNA interference (RNAi) methods have been used in cell culture to generate GI maps for genes involved in kinase signalling, chromatin regulation and ricin susceptibility[11–14]. Such analyses of large-scale GI maps have successfully identified functionally connected pathways and genetic networks, allowed functional prediction for uncharacterized genes and revealed changes in the interaction landscape upon stimulus[11–14].

Therefore, a comprehensive mapping of epistasis among cancer genes could be tremendously useful for understanding the genetic network that underlies tumorigenesis, and for identifying new tumour suppressors or oncogenes, thus facilitating design of cancer therapies. For example, screens for 'synthetic lethal' partners of the *RAS* oncogene have uncovered genes that could be used as drug targets to specifically eliminate cancer cells with *RAS* mutation[15,16]. Studies using PARP1 inhibitors have also demonstrated the potential for utilizing 'synthetic lethality' in treating BRCA-deficient patients[17].

In this study, using combinatorial RNAi in human breast epithelial cells, we produce a quantitative GI map for 67 genes that are frequently altered in breast cancer as well as in many other cancer types. We find hundreds of previously unreported GIs among these genes. Our analysis of the GI map in the context of gene expression and genome-wide transcriptional factor binding data also reveals potential mechanisms of expression regulation in tumorigenesis. Most strikingly, GIs are enriched in gene pairs that synergistically impact patient survival when mutated, indicating that the GI maps produced in cell lines can have biological relevance in tumours.

Results

Combinatorial RNAi for co-altered breast cancer genes. To study GIs among cancer genes, we began by using diverse genomic data in The Cancer Genome Atlas (TCGA)[18] to identify frequently altered genes in breast cancer. By combining somatic mutations, gene expression and copy number variation data from 849 invasive breast cancer samples, we identified 67 genes that are involved in these frequent alterations. These genes span a range of biological functions, including chromatin regulators, transcription factors, signal transduction molecules and metabolic enzymes (Supplementary Data 1).

To build a GI map from these 67 genes, we performed a combinatorial RNAi screen in MCF10A cells by using 66 of the

67 genes as template genes and 29 of the 67 genes as query genes (Fig. 1a,d and Supplementary Data 1). Among the 1,508 gene pairs we assayed, 248 gene pairs corresponded to significantly co-occurring genetic alterations in primary breast tumours (Supplementary Fig. 1). We chose the MCF10A breast epithelial cell line[19,20] because these cells are diploid and lack alterations in most cancer genes.

To minimize the influence of short interfering RNA (siRNA) off-target effects on GI detection, for each gene we tested multiple siRNAs and selected two independent siRNAs that yielded quantitatively similar phenotypes for the screen (Supplementary Fig. 2; see Methods section). To validate the on-target knockdown efficiency of the selected siRNAs, we measured target transcripts levels by quantitative reverse transcription–polymerase chain reaction (qRT–PCR) for 48 selected siRNAs targeting 24 of the 67 genes (Supplementary Fig. 3 and Supplementary Data 2). The siRNAs exhibited various degrees of knockdown efficiency, with the percentage reduction of targeted transcript levels ranging from 18% to ~100% with a median of 69% and an interquartile range of 45–85%. In addition, 28 of the 48 siRNAs were able to reduce the targeted transcript levels by >60%, including 10 that reduced the transcripts levels by >90%. These results on siRNA efficiency are consistent with previous reports[13]. For six genes where <40% silencing was observed at 24 h, an enhanced reduction in mRNA levels was observed at 72 h, the time point when the phenotypic assays were performed (Supplementary Fig. 3b and Supplementary Data 2). Ten of the 24 genes have two siRNAs that are different in knockdown efficiencies at 24 h (Student's *t*-test, $P < 0.05$; Supplementary Data 2). Although the siRNA knockdown efficiencies varied and sometimes resulted in partial knockdown, in our assay they generated consistent phenotypes in single knockdowns, as well as reproducible and statistically significant quantitative phenotypic changes in double knockdowns. Overall each gene pair was assayed by four different combinations of siRNA pairs (6,032 siRNA pairs in total; Fig. 1a). Two to three replicates were performed for each siRNA pair, yielding a total of 18,623 measurements.

We used automated fluorescence microscopy to record cell morphology in images and extracted 10 quantitative features from each image. The three most reproducible and non-redundant features were used for further analysis: the number of cells per image, the mean nucleus size and the mean cell size (Supplementary Fig. 4). Although the features we measured in our screen do not necessarily correspond to pathological phenotypes, they are complex traits influenced by many cancer-relevant pathways. For example, the number of cells is a reflection of the cell proliferation capacity that tends to change significantly in tumours. Furthermore, the quantitative nature of these features allows us to detect subtle deviations from the expected effects of each pair of genes. The multitude and quantitative nature of the phenotypes thus increase the likelihood of ascertaining whether the two genes can interact at all, regardless of the type of alteration (amplifications, deletions or point mutations) seen in patients. Our RNAi assay can still detect a loss-of-function interaction, even when the pair of gene alterations includes a gain in copy number or an activating mutation in primary tumours.

To calculate the GI scores, we adopted the regression framework, comparing the observed and predicted phenotypes after double knockdowns in a single linear model, similar to the π score method[11,12] (see Methods section). However, since the three phenotypes in our assay are correlated (Supplementary Fig. 5), we developed a seemingly unrelated regression (SUR) model[21], which computes the GI scores of a gene pair for each phenotype while accounting for the correlation among the phenotypes. Our model also accounts for batch effects (see Methods section). The statistical significance of a GI score, which
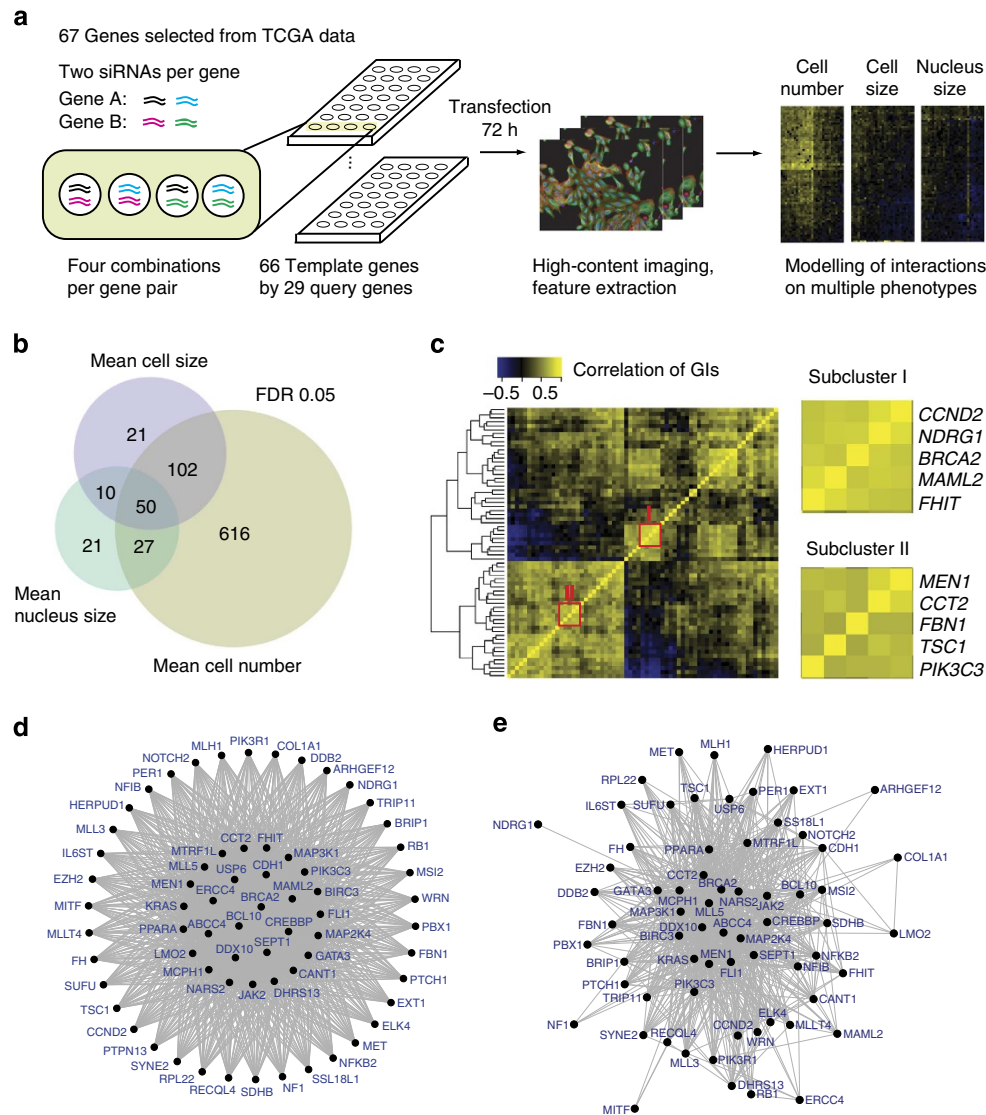
Figure 1 | GI mapping of 67 breast cancer genes. (**a**) A schematic representation of the combinatorial RNAi screen. Heat maps of GI scores for three phenotypes are shown. (**b**) Overlap of significant GIs within the three phenotypes. (**c**) Unsupervised clustering of the Pearson correlations of GI profiles across all three phenotypes between the 66 genes (left), and clusters of genes in cell cycle checkpoint (right; module I) and cytoskeleton regulation (right; module II). (**d**) The design map showing the 67 genes (nodes with yellow labels) and 1,508 interactions (grey lines) assayed in the RNAi screen. Twenty-nine genes (at the centre) were each assayed for interaction with all the other 66 genes, and the other 38 genes (in the outer circle) each for interaction with the 29 genes at the centre. (**e**) The inferred GI map of 847 significant interactions (FDR: 0.05) across 66 genes in all three phenotypes. One of the 67 genes did not show significant interaction with any other gene, and is therefore absent from the map. A gene with fewer significant interactions than the assayed interactions tends to move away from the centre in comparison with the design graph in **d**.

is the coefficient of the interaction term in the SUR model, is determined by a *t*-test following the standard approach for assessing the significance of a coefficient in a linear model. We further adjusted the *P* values of the GI scores in order to control the false discovery rate (FDR; see Methods section).

Topology of the inferred GI map. In total, 847 significant gene pair interactions were identified at an FDR of 0.05 (Fig. 1b and Supplementary Data 3, heat maps in Fig. 1a and Supplementary Figs 6–8), demonstrating widespread epistasis among cancer genes. Out of the 67 genes assayed, 66 have at least one significant interaction; *PTPN13* being the only gene for which our data do not support interaction with any other gene (Supplementary Data 3). This result is robust even with an FDR of 0.01 (579 GIs identified across 65 genes).

The inferred GI map has a high median connectivity of 62% for each gene (accounting for different numbers of probed interactions for each gene, Fig. 1e; Supplementary Data 5 and Supplementary Figs 9 and 10), and exhibits small-worldness[22] ($S = 1.5$, higher than 1.1, which is the small-worldness of all the probed interactions; see Methods section).

GI profiles identify functional clusters of genes. We compiled the interaction profiles for each gene with respect to all three phenotypes and calculated the pairwise correlation matrix of interaction profiles (Fig. 1c). As expected, if our approach is identifying biologically relevant interactions[12,13], unsupervised clustering of genes based on the correlation of their GI profiles revealed genes with previously described functional connections. For example, four genes (*CCND2*, *BRCA2*, *NDRG1* and *FHIT*)

shown in subcluster I (Fig. 1c) are involved in DNA damage checkpoint pathways for cell cycle progression[23–26]. Subcluster II (Fig. 1c) highlights genes (*MEN1*, *CCT2*, *FBN1* and *TSC1*) with roles in the regulation of focal adhesion and cell migration[27–31]. Overall, this clustering reveals that there are distinct subsets of correlated genes, and that at least some of these correlated gene sets represent functional relationships.

The majority of GIs exhibit epistatic masking effects. The magnitude of the interaction effect size may have functional implications, especially when compared with the effect sizes of single knockdowns. We classified epistasis into three broad classes: epistatic masking, synthetic suppression and synthetic enhancement (Fig. 2 and Methods section). Surprisingly, over half of significantly interacting gene pairs exhibited masking effects (epistasis as first described by Bateson and Mendel[32]), where the effects of double knockdown are indistinguishable from the effects of one of the two single knockdowns (Fig. 2). Further examination of these pairs may generate useful information for personalized cancer therapy. For example, if inhibiting gene *A* masks the phenotype of gene *B*, gene *A* might be an alternate

drug target for treating cancers caused by gene *B* activation. Indeed, largely because we focused the study on known cancer genes, ∼20% of gene pairs we selected included a potential drug target.

Epistatic masking is not an artefact due to differences in transcript abundance or knockdown efficiencies. We could identify which gene is masked if the two single-gene effects are different and the double-knockdown effect is equal to one of them. In 188 of the 403 epistatic masking gene pairs identified for the cell number, the masked gene in these pairs had lower transcript abundance (measured by FPKM (fragments per kilobase per million mapped reads) in RNA-sequencing (RNA-seq) data), which gives an insignificant P value of 0.18 by a χ^2-test. In addition, in 40 of the 88 epistatic masking gene pairs, for which we have qRT–PCR knockdown efficiency data, the masked gene had a lower remaining transcript level (measured as the percentage of untreated), which also gives an insignificant P value of 0.66 by a χ^2-test.

Most inferred GIs are novel. Using the STRING 9.0 database, we identified significant enrichment of gene pairs previously shown to be involved in the same pathways or complexes (Fisher's exact test $P = 0.05$; Supplementary Fig. 11a). These interactions included direct protein–protein interactions (for example, *BRCA2* and *MCPH1*; Fig. 3a)[33], interactions between genes in kinase signalling pathways (for example, *MAP2K4* and *MAP3K1*; Fig. 3b)[34], interactions between genes encoding factors that compete for substrate binding (for example, *CREBBP* and *MEN1*; Fig. 3c)[35] and interactions that are likely explained by transcriptional regulatory relationships (for example, *BIRC3* and *NFKB2*; Fig. 3d)[36].

However, the majority of GIs are not annotated by protein interaction databases, and yet extensive cross talk between pathways was revealed. We define cross talk as the interactions between one or more components of two pathways defined in the literature. For example, as shown in Fig. 3e, we found 12 GIs between the DNA damage repair pathway and the mitogen-activated protein kinase (MAPK) pathway. In this case, *MAP2K4* is upstream of the Jun N-terminal kinase (JNK) pathway, which is also activated by DNA damage through JNK phosphatase[37,38], perhaps explaining the observed cross talk. In total, 45 of the inter-pathway GIs are supported by published results (Fig. 3f and Supplementary Data 3). We classify another 78 GIs as potential inter-pathway interactions, because the two genes in each pair each interact either with another gene that is known to have a role in pathway cross talk or with other genes that have inter-pathway interactions (Supplementary Data 3). An overview of the cross talk between different biological processes (Supplementary Data 4) is plotted in Supplementary Fig. 11b.

Figure 2 | Definition of the three types of GIs. (**a**) Illustration of the definition of three types of GIs for genes with single-knockdown effect of same direction. (**b**) Illustration of definition of three types of GIs for genes with single-knockdown effects of different directions. ΔA and ΔB represent single-knockdown effects of genes *A* and *B*, respectively. ΔA + ΔB represents expected effects of double knockdown of genes *A* and *B*. Shaded bars represent the observed double-knockdown effect. (**c**). Counts of the three types of interactions for each phenotype.

Integrative network analysis reveals transcriptional modules. We also identified several potential transcriptional regulatory modules by overlaying the GI network with a co-expression network built from TCGA RNA-seq data (see Methods section). One such module is centred on the transcription factor *GATA3* (Fig. 4a). GATA3 is an important transcriptional regulator in both normal mammary gland development and breast cancer[39–41], and low expression levels of *GATA3* are associated with a poor prognosis[42]. Three genes, *PTCH1*, *PPARA* and *NFIB*, exhibit epistatic interactions with *GATA3* and also display negatively correlated expression levels with *GATA3* (Fig. 4a). According to the published ENCODE data[43], GATA3 binds to gene regions near *NFIB*, *PTCH1* and *PPARA* in T47-D cells, a human ductal breast epithelial tumour cell line (Fig. 4b–d). These peaks of GATA3 binding also overlap with DNaseI
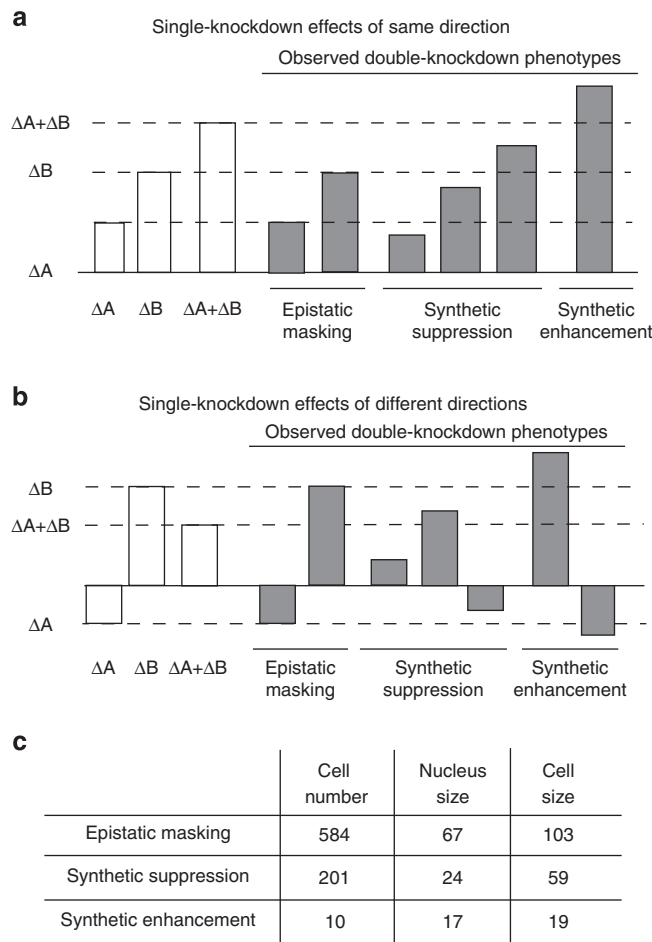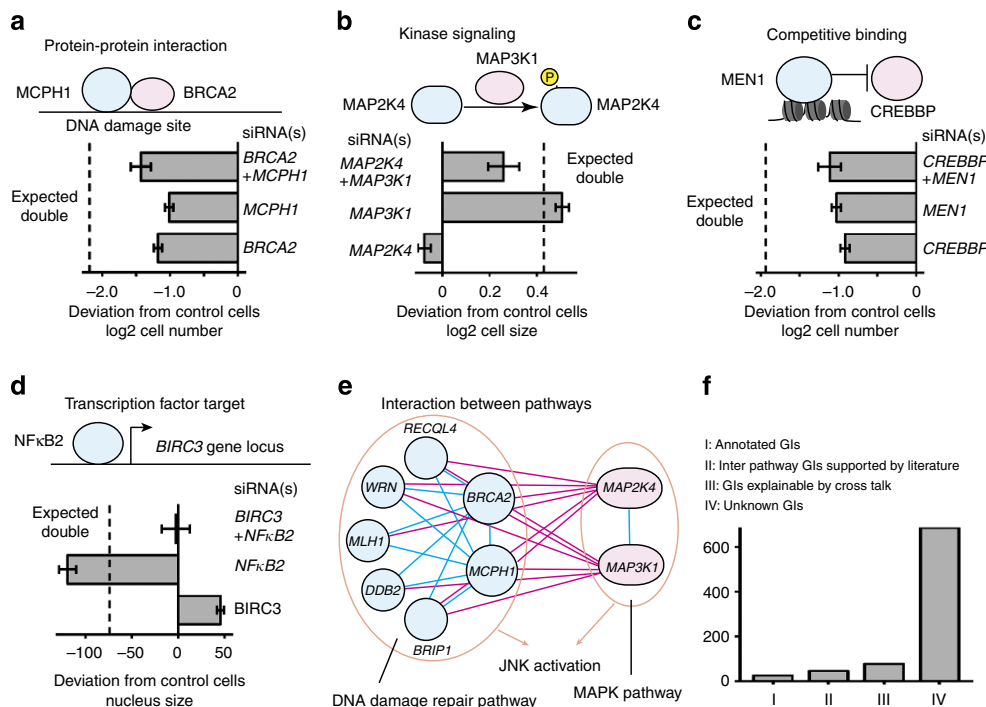
Figure 3 | Combinatorial RNAi screening reveals both known and novel molecular interactions. (**a**) MCPH1 recruits BRCA1 to DNA damage sites through physical protein–protein interaction. (**b**) MAP3K1 activates MAP2K4 by phosphorylation. (**c**) MEN1, a histone methyltransferase complex member, interferes with binding of CREBBP, a histone acetyltransferase, at target gene promoters. (**d**) NFKB2 regulates BIRC3 expression through transcriptional activity. In **a**–**d**, shaded bars indicate observed single- or double-knockdown effect, and error bars represent s.e.'s. (**e**) Interactions may exist between the DNA damage response pathway and MAPK pathway, possibly at the JNK activation step. The edges are GIs identified from the RNAi screen. (**f**) Bar plot of the count in each category of GIs.

hypersensitivity clusters, indicating that *GATA3* may actively regulate the expression of these three genes in breast cancer cells.

Another transcriptional module consists of *FLI1*, *JAK2* and *CCND2* (Fig. 4e). The expression levels of the three genes are all positively correlated (red dashed lines in Fig. 4e). In addition, *JAK2* and *FLI1* have correlated interaction profiles (Fig. 1c and Supplementary Fig. 6). We found that knockdown of *FLI1* in MCF10A cells reduced the expression levels of *JAK2* (Supplementary Fig. 12). It has also been reported that FLI1 modulates CCND2 expression[44]. While FLI1 expression levels have been associated with the clinical outcome in acute myeloid leukaemia[45], to our knowledge, this module has not been previously associated with breast cancer patient outcome. To test this association, we examined TCGA breast cancer patient survival data and found that low expression or mutation in one or more members of the *FLI1-JAK2-CCND2* module is associated with reduced overall survival time in all patients (Fig. 4f) as well as in estrogen receptor-positive patients (Supplementary Fig. 13). These results indicate that this module may have a previously unappreciated role in breast cancer.

GIs are associated with patient survival. To more broadly investigate the role of GIs in cancer progression, we performed patient survival analysis to assess the potential impact of co-alterations in gene pairs. We used TCGA genomic data to define the alteration of a gene in a given sample, similar to the approach we took to selecting significantly co-altered gene pairs (see Methods section). Each gene has one of the three states: 'gain' if the gene in the sample is amplified and expressed at a high level relative to the average level of the gene in samples that are diploid for the gene, or if the gene has previously identified gain-of-function point mutations; 'loss' if the gene is deleted and

expressed at a low level, or if the gene has (putative) loss-of-function mutations; or 'no alteration' (see Methods section). For each gene pair, we examined each of the four co-alteration scenarios, namely, 'gain–gain', 'loss–loss', 'gain–loss' and 'loss–gain'.

By comparing the survival time of patients with both genes altered to survival time of patients with only one gene altered, we found 51 gene pairs, spanning 43 genes, whose co-alterations were associated with worse outcome for breast cancer patients (Fig. 5a and Supplementary Data 6). Interestingly, GIs are significantly enriched in those gene pairs ($P = 0.023$, Fisher's exact test), indicating that GIs indeed have effects on patient survival. For example, in the case of *SS18L1* gain and *MAP2K4* loss (Fig. 5b), patients with both alterations have shorter survival time than those with only one alteration (Fig. 5c). This property is independent of the ER status since the association was also discovered in ER-positive patients (Fig. 5d). In addition, this property is independent of *TP53* status, which is located in the same chromosome region as *MAP2K4* (Supplementary Fig. 14). For the overall association between GIs and patient survival, the instability of tumour genomes was not a confounding factor ($P > 0.1$ for both survival time and survival status; see Methods section). These data indicate that information on GIs can be useful for risk assessment and outcome prediction.

Discussion

Although GIs are thought to be pervasive in cancer, quantitative large-scale mapping of GIs have not been performed among cancer genes. In this study, we systematically profiled GIs among 67 cancer-related genes in human cells by using combinatorial RNAi and found over 800 significant GIs across 66 genes. Our results strongly indicate that epistatic effects of somatic mutations
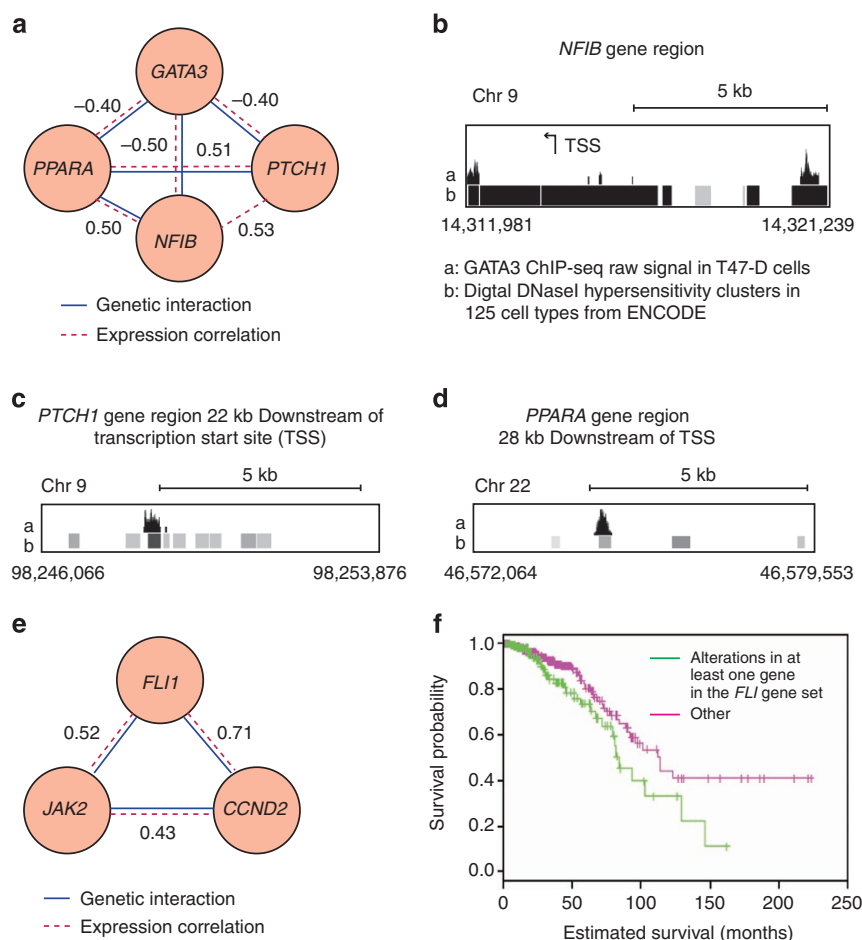
Figure 4 | Transcriptional modules were identified by overlaying the GI network with co-expression networks. Solid blue lines represent GIs identified from the RNAi screen; dashed red lines represent correlated expression levels (Pearson correlation coefficients are indicated) based on TCGA RNA-seq data of breast cancer patient samples. (**a**) *GATA3* module. (**b**–**d**) ENCODE data support the direct binding of GATA3 in the gene regions of *NFIB*, *PPARA* and *PTCH1*. Black bars represent chromatin immunoprecipitation-sequencing (ChIP-seq) raw signal of GATA3 binding in T47-D cells. Grey bars represent DNaseI hypersensitivity clusters in 125 cell types from ENCODE. (**e**) A module composed of *FLI1*, *JAK2* and *CCND2*. (**f**) Kaplan–Meier curves of survival probabilities in patients with and without alterations in the *FLI1* module. Each tick mark indicates a patient with censored survival time. Lower expression of at least one of the three genes in *FLI1* module is associated with shorter average survival time in breast cancer patients ($P = 0.03$; two-sided log rank test).

are widespread, revealing a highly connected and complicated genetic architecture associated with breast cancer.

Our GI mapping strategy is similar to those developed in flies and mammalian systems[11–13], using combinatorial RNAi and high-content imaging. In all of these studies, epistasis is commonly defined as deviation from the expected phenotype of combining two alleles. However, there are several choices for quantitative definitions of GI scores[46–49]. For example, additive models examine the difference from the sum of the two single-gene phenotypes, whereas multiplicative models examine the fold change over the product of the two single-gene phenotypes (see Methods section). Here, we adopted a regression framework, following the definition in classical quantitative genetics[50] and similar to previous approaches[11,12]. In this case, we used SUR to provide a unified framework for both additive and multiplicative models. This unified SUR framework is particularly relevant for our analysis as we considered three related but non-redundant features. It enables us to used multiplicative models and additive models for different features based on the scales and distributions of the raw data, while accounting for any correlation among the phenotypes and for confounding factors as covariates (for example, the batches in our model).

Our results systematically specify, and for the first time quantify, the genetic architecture that underlie cancer gene interactions. It is very likely that our GI mapping results will generalize to other cancer types, since the genes we studied are mutated in many different types of tumours and are involved in known cancer-related pathways. The high connectivity for our gene set is most likely owing to the fact that cancer genes often have pleiotropic effects, and pleiotropy is characteristic for hub genes in networks[8,9]. Therefore, the genes we selected based on tumour mutation data are more likely to be hubs in the overall genetic network and mediate cross talk between multiple pathways. Consistent with our results, in a recent study of cancer driver genes, genes significantly altered in cancer patients are found to have a high connectivity in the BioGRID human protein–protein interaction network[51], as well as more likely to be in protein complexes than a typical protein in the CORUM database of experimentally validated human protein complexes. Interestingly, haploinsufficiency appears to be common in cancer[51], which is consistent with our observation that a partial knockdown of 40–50% by siRNAs was sufficient to generate reproducible phenotypic changes in cells.

The extensive interactions among cancer genes could be a critical component for explaining nonlinear cumulative effects of
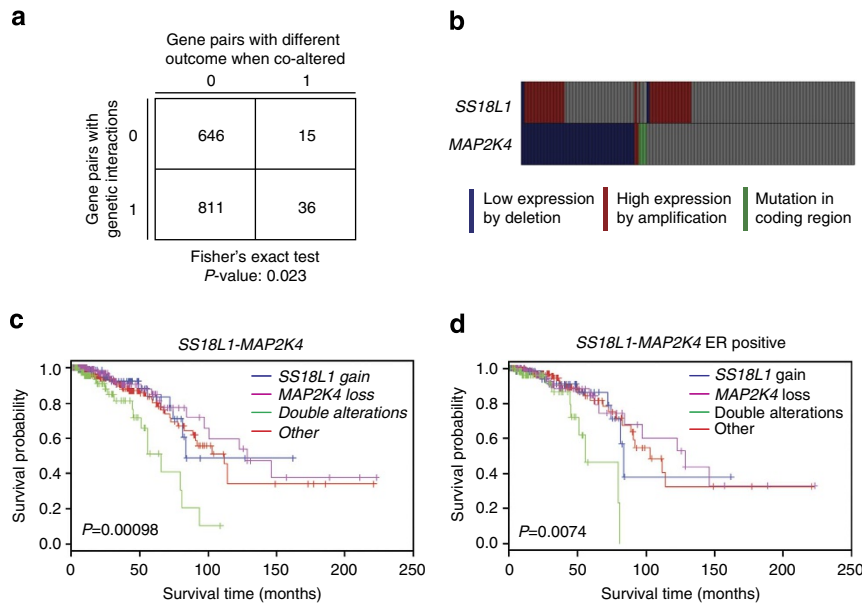
Figure 5 | Co-alterations of GI gene pairs are significantly associated with clinical outcome in breast cancer patients. (**a**). GIs are enriched within gene pairs that impact overall survival when co-altered in breast patient samples. One-sided Fisher's exact test was used. (**b**). Frequent co-occurrence of *MA2K4* loss and *SS18L1* gain in 822 breast cancer patients. (**c**) Kaplan–Meier curves of overall survival in patients with *MAP2K4* loss, *SS18L1* gain or both alterations ($P = 0.00098$; two-sided log rank test). (**d**) Kaplan–Meier curves of overall survival in ER-positive patients with *MAP2K4* loss, *SS18L1* gain or both changes ($P = 0.0074$; two-sided log rank test).

mutations in cancer. Our results strongly indicate that epistatic effects of somatic mutations are likely critical for breast cancer development and progression. By combining GI information with publically available genomic and survival data for breast cancer patients, we found that the identified GIs are dominantly present in gene pairs whose co-alteration is associated with reduced survival time of breast cancer patients. Current models of cancer progression usually do not account for GIs. Therefore, more comprehensive quantitative mapping of GIs among cancer genes will facilitate the modelling of complex effects that result from multiple mutations. Finally, our results indicate that patients with certain combinations of mutations have poorer prognosis as a result of epistasis, suggesting that therapeutic interventions countering these epistatic interactions may hold promise.

Methods

TCGA data acquisition. TCGA breast cancer data was downloaded through cBio Cancer Genomics Portal[52] (http://cbioportal.org) on 27th December, 2012. These data include copy number variation (GISTIC score), somatic mutation, mRNA expression (RNA-seq z score) and survival data.

Detection of significantly co-altered gene pairs. To identify genetic aberrations that co-occur non-randomly, we used data sets generated by TCGA[18]. By combining somatic mutations, gene expression and copy number variation data from 849 invasive breast cancer samples, we generated a gene-by-sample matrix. In this matrix, each gene-sample pair has one of the three states: (i) no alteration, denoted by 0; (ii) gain, denoted by 1, which is defined by either a higher copy number (GISTIC score >0) with high gene expression (z score >0.5), or gain-of-function mutations (from manually curated literature); or (iii) loss, denoted by -1, which is defined either as a lower copy number (GISTIC score <0) and low gene expression (z score < -0.5), or loss-of-function mutations. For a pair of genes A and B, co-occurrence of alterations corresponds to the number of tumour samples containing both genes being altered (this number is denoted C_{11}; Table 1). Within the tumour samples, these counts each follow a binomial distribution, with the 'success' probability computed under the null hypothesis that the two genes are altered independently (Table 1). We conducted three one-sided binomial tests for each gene pair, corresponding to the three possible scenarios: both genes being 'gain', both being 'loss', and one being 'loss' and the other 'gain'). Binomial tests are exact tests, thus alleviating the problems with small counts. We further computed the Q values for the binomial P values by using the R function Q value from package Q value.

Table 1 | Observed count and expected frequency of the mutation pattern of two genes.

	Gene B	
	0 (WT)	**1 or −1 (altered)**
Gene A		
0 (WT)	C_{OO}	C_{O1}
1 or −1 (altered)	C_{10}	C_{11}
Gene A		
0 (WT)	$(1 - p_A)(1 - p_B)$	$(1 - p_A)p_B$
1 or −1 (altered)	$(1 - p_B)p_A$	$p_A p_B$

WT, wild type.
Top: the observed count of cases in each combination of the mutation pattern. Bottom: the expected frequencies under the null hypothesis of two genes being altered independently.

Human tissue culture. Human MCF10A cells were obtained from American Type Culture Collection (number: CRL-10317, LOT: 7635052). The cells were maintained in DMEM/F12 medium containing 5% horse serum and supplemented with epidermal growth factor (EGF) (20 ng ml^{-1}), hydrocortisone (0.5 ng ml^{-1}), cholera toxin (100 ng ml^{-1}), penicillin (100 U ml^{-1}) and streptomycin (100 µg ml^{-1}). Cells were cultured at 37 °C and in 5% CO_2.

RNA-seq of MCF10A cells. A quantity of 2 µg of RNA was extracted from MCF10A cells by using the AllPrep RNA mini kit (Qiagen). The RNA integrity number of the sample was over 7.5 (Bioanalyzer, Agilent). A single-end library was prepared following the Illumina TruSeq RNA sample preparation protocol. The library was indexed with one of the Illumina barcodes and sequenced in 1/4 lane on an Illumina Hiseq 2000. Single-end reads (37,386,273 in total) of length 44 bp were generated. Sequence files were parsed and low-quality reads (average Phred-scaled quality score <15) were removed by using custom perl scripts. The remaining reads were then mapped to hg18 by using TopHat and output in BAM format. Unmapped reads were removed from further analysis. Cufflinks was used to estimate transcript abundance using RefSeq gene models and output as FPKM. Genes (12,332) with FPKM >1 were considered as expressed in MCF10A cells.

Transfection. Transfection was performed in 384-well plates. Each well contains 0.25 pmol of both siRNAs in 5 µl Opti-MEM (Life Technologies) and 0.03 µl

RNAiMAX (Life Technologies) diluted in 5 μl Opti-MEM. For assaying single siRNA knockdown phenotypes, negative control siRNAs were combined with the targeting siRNAs. Five-hundred cells were seeded in 40 μl MCF10A growth medium with no antibiotics by using a Multidrop Nano reagent dispenser (Thermo Scientific). Cells were incubated at 37 °C and in 5% CO_2 for 72 h before fixation and staining.

Cell staining. All steps of staining and imaging were performed at room temperature. Cells were fixed in 5% paraformaldehyde (Electron Microscopy Sciences) in PBS for 20 min., followed by permeabilization with 0.2% Triton X-100 in PBS. Cells were then washed with 0.05% Triton X-100 and blocked in 3% BSA PBS for 1 h. Cells were incubated in anti-alpha-tubulin (Clone B-5-1-2, Sigma, 1:1,000) for 2 h followed by three washes with PBS. Secondary antibody staining was performed with Alexa 488 goat anti-mouse immunoglobulin G (Invitrogen, 1:500), Hoechst 33342 (Invitrogen, 13 ng ml^{-1}) and rhodamine phalloidin (Invitrogen, 1:50) for 45 min, followed by three PBS washes. The plates were then stored in PBS at 4 °C before imaging.

Automated image acquisition. Images were acquired on a Beckman ImageXpress Micro by using a × 10 objective lens in a 12-bit tiff format. Images from four different sites were collected for each well of a 384-well plate. Each site typically contains 200–600 cells. Channels for Hoechst 33342 (DNA), Alexa 488 (tubulin) and rhodamine were imaged. The total number of images acquired was 58,212.

Image analysis. The image analysis pipeline was built upon the EBImage package[53] in R. Images of 12-bit tiff were first converted to 16-bit tiff and then to JPEG format for display and storage. The illumination non-uniformity across the field of view was corrected by using background subtraction before analysis. Artefacts were removed from all channels by a thresholding method. Connected regions that are larger than 4,000 pixels in the nucleus channel were removed from subsequent processing in all three channels.

Nuclei and cell segmentations were performed following the protocols described in Fuchs et al.[54] The images were first normalized for better segmentation while the raw images were kept for raw intensity calculation. Nuclei were segmented by adaptive thresholding of the DNA channel with the threshold set to 0.01. Objects that were too dark (average intensity < 0.01), too large (> 2,000 pixels) or too small (< 75 pixels) were treated as artefacts and removed. Cell masks were calculated by using the summation of the actin, tubulin and DNA channel signals. Cell boundaries were then separated by location of the nucleus by using the Voronoi segmentation algorithm through the 'propagate' function incorporated in EBImage[54]. Cells that were too large (> 145,000 pixels), too small (< 150 pixels), too dark (average intensity < 0.1) or too close to the border (edge/peripheral length > 0.3) were removed as artefacts. Each cell was characterized by eight morphological features for further analysis: cell number, cell size, nucleus size, cell eccentricity, nucleus eccentricity, nucleus fluorescence intensity, actin intensity and tubulin intensity.

siRNA selection. 366 Silencer Select siRNA targeting 122 genes were obtained from Ambion Applied Biosystems (Life Technologies). To identify high-quality siRNAs, we screened three siRNAs for each of the 122 genes and selected the two siRNAs with the most consistent phenotype. Each siRNA was arrayed twice in 384-well plates and cells were transfected by using RNAiMAX (Life Technologies) following the manufacturer's instructions on reverse transfection. After staining, images were obtained and ten cell features were extracted. Euclidean distances between two feature vectors were calculated. Two siRNAs with distances not larger than the mean distance between technical replicates were selected for 67 genes. The sequences of the 134 siRNAs used in combinatorial screen are listed in Supplementary Data 1.

Combinatorial RNAi screen. For the combinatorial RNAi screen, a template-query design was used similar to the one described in Horn et al.[11] A quantity of 100 nM of the 132 template siRNAs stocks were arrayed twice in a 384-well plate (Beckman), and aliquoted to 63 plates. Each plate also contained 16 empty wells, five positive controls (PLK2) and eight negative controls (non-targeted cells), all randomly arrayed on the plate. Each template plate was then combined with one of the 29 query siRNAs or Opti-MEM medium. In this way, each pair of genes was assayed four times through four different combinations of siRNA within one plate. Liquid handling was performed by the automated liquid handlers Evolution Precision Pipetting Platform (PerkinElmer) and Freedom EVO100 (Tecan). Cells were then plated and stained as described above.

Validation of RNAi knockdown using qRT–PCR. siRNA transfections were carried out in the same conditions as in the primary screen. Twenty-four hours later, siRNA-transfected cells were lysed by using the Cells-to-ct Kit (Life Technologies), followed by reverse transcription and real-time PCRs according to the manufacturer's instructions. PCRs were performed on a Step One Plus (Applied Biosystems). Percent mRNA remaining was calculated by using the Step One software.

Expression data were first normalized to a house keeping gene (GAPDH), and expression fold changes were calculated by comparison with cells treated with non-targeting control siRNA. For selected genes, remaining mRNA levels after 48 or 72 h were assayed. The low abundance of templates in our assay may result in high variances between technical replicates for some of the genes. The primers used for qRT–PCRs are listed in Supplementary Data 2.

SUR for interaction analysis. We focused on three phenotypes in the RNAi screen: the cell number, cell size and nucleus size, which are correlated among each other, with the correlation between cell size and nucleus size being stronger. To identify significant GIs for each phenotype while accounting for the correlation among the phenotypes, we adopted a SUR framework[21]. That is,

$$y_{tkijr} = \beta_0 + \beta_{1k} + \beta_2 G_i + \beta_3 G_j + \beta_4 (G_i \times G_j) + \varepsilon_{tkijr} \tag{1}$$

$$\varepsilon_{tkijr} \sim \text{MVN}(0, \Omega), \tag{2}$$

where t represents indexes phenotypes, k batches, i and j genes, and r replicates. G_i (or G_j) takes value 1 if the i-th (or j-th) gene is knocked down, and ε_{tkijr} denotes the error term, which follows a multivariate normal distribution with mean 0 and covariance matrix Ω. Furthermore, β_0 indicates the baseline level, β_{1k} the impact of the k-th batch, β_2 and β_3 the single-gene effect size, and β_4 the interaction effect size (or score). The SUR accounts for dependence among phenotypes through the covariance matrix Ω of the error term ε_{tkijr}. Specifically, if Σ is the covariance matrix of 3 by 3 for the phenotypes, the covariance matrix Ω of the error term in the above regression is,

$$\Omega = \Sigma \otimes I_N \tag{3}$$

where I_N is an N by N matrix with 1 along the diagonal and 0 off the diagonal, N is the total number of observations for each phenotype and $\otimes$ represents the Kronecker product.

Since the cell number grows exponentially, we log$_2$-transformed this phenotype in regression. We also log$_2$-transformed the cell size, as the transformed data were closer to a normal distribution (Supplementary Fig. 5). The data on the nucleus size followed roughly a normal distribution (Supplementary Fig. 5), so we did not transform these data. In terms of their implications on the interaction, these strategies mean that we examine a multiplicative model of GIs on the cell number and nucleus size, and an additive mode on the cell size (see the description of the additive and multiplicative models in next section).

We carried out the linear regression for each phenotype by using the R function lm and performed SUR with the function systemfit in the R package systemfit[55,56]. To assess the statistical significance of the P values for the estimated coefficients of the interactions, we computed the Q values for the P values of the estimated coefficients of the interactions for each phenotype, using the function Q value in the R package Q value.

In addition, we pooled the interaction effect sizes across the three phenotypes that generated the overall GI profile for each gene. We computed the Pearson correlation for all gene pairs and applied hierarchical clustering to the resulting correlation matrix.

Notes on additive and multiplicative model for epistasis. Ignoring the noise in the data and assuming that the single-gene and double-knockdown phenotypes can be estimated perfectly, we can describe epistasis through the models below:

Additive model:

$$\beta = x_{ij} - (x_i + x_j); \tag{4}$$

Multiplicative model (raw scale):

$$\beta = x_{ij} / (x_i x_j); \tag{5}$$

Multiplicative model (log transformed):

$$\log \beta = \log x_{ij} - (\log x_i + \log x_j). \tag{6}$$

In the models above, β is the GI score, x_{ij} is the observed phenotype of double knockdown, x_i and x_j each are the phenotype of single knockdowns. Both $x_i + x_j$ and $x_i x_j$ ($\log x_i + \log x_j$ on the log scale) are the expected phenotype from the double knockdown. In the additive model, the GI score β is the (mathematical) difference between observed and the expected phenotypes. In the multiplicative model, β is the ratio between observed and the expected phenotypes.

Both multiplicative and additive models are commonly used when assessing epistasis[46–49]. The practical choice of model often depends on the particular phenotype and scale of measurement[57]. Whereas additive models are suitable for phenotypes that change linearly, multiplicative models are intuitively suitable for growth-related phenotypes, such as yeast fitness or cell growth, because the cells are known to grow nonlinearly.

Network analysis of the inferred GI map. For a graph, g, with n nodes and m edges, the metric of the small-worldness, denoted S, is largely based on the

definition proposed in Humphries and Gurney[22]:

$$S = \left(\frac{C_g}{C_{rand}}\right) \Big/ \left(\frac{L_g}{L_{rand}}\right) \qquad (7)$$

where C_g and C_{rand} are the clustering coefficient of the graph g and that of a random graph with the same number of nodes and edges, and L_g and L_{rand} are the mean shortest path length of g and that of a random graph. To estimate C_{rand} and L_{rand}, however, we generated a large number of random graphs with n nodes and m edges under the Erdos–Renyi model[58], such that each edge is created with the same probability. We computed C and L for each random graph, and used the average as the estimate for C_{rand} and L_{rand}. We used functions erdos.renyi.game, transitivity and average.path.length in the R package igraph[3] for the computation.

We used this method to compute the small-worldness for the 'prior' graph of all 67 genes and 1,508 edges assayed in the RNAi screen, and for the GI map of 66 genes and 847 edges (significant GIs at FDR of 0.05). The small-worldness is 1.1 for the 'prior' graph and 1.5 for the GI map, suggesting that the GI map has a strong level of small-worldness, not an artefact of the experimental design.

To compute the connectivity of each gene in the GI map, we first computed the unadjusted connectivity, which is the number of edges (or signification GIs) a gene has, and then divided it by the number of interactions probed (66 or 29 in our experimental design) to obtain the proportion of significant GIs as the adjusted connectivity.

We computed the betweenness[59] under the canonical definition (using function betweenness in the R package igraph[59]), rescaled it for the nodes with 29 'prior' edges and normalized for all nodes to obtain a value between 0 and 1. Specifically, there exists a log-linear relationship between the canonical betweenness and adjusted connectivity in each node group. For nodes with 29 'prior' edges, log (betweenness) $= -3.5 + 5.2 \times$ (adj connectivity). For nodes with 66 'prior' edges, log (betweenness) $= -0.7 + 6.3 \times$ (adj connectivity). Note that the slope is roughly the same, whereas the intercept differs by 2.8. We therefore multiplied the canonical betweenness for nodes with 29 'prior' edges by $e^{2.8}$, such that the values are comparable to that of the other node group. We then divided all the betweenness values by the maximum (Supplementary Data 4).

Modes of interaction. We compared the effect size of a double knockdown (ΔAB) with that of single knockdowns (ΔA or ΔB) and the expected effect size of the double knockdown ($\Delta A + \Delta B$), and defined three broad types of epistasis: (i) epistatic masking, when $\Delta AB = \Delta A$ or $\Delta AB = \Delta B$. In other words, the effect of one gene is completely masked (or dominated) by the other gene; (ii) synthetic suppression, when $\Delta AB < (\Delta A + \Delta B)$; (iii) synthetic enhancement, when $\Delta AB > (\Delta A + \Delta B)$. To infer these types of epistasis from the RNAi data, we used the estimated effect sizes of single and double knockdowns from the linear regression, and allowed deviation of 1 s.e. in the comparisons in order to account for the estimation uncertainty. For example, if the estimated effect size of the double knockdown is within 1 s.e. of the estimated effect size of the single knockdown of gene A, then we categorize genes A and B as in epistatic masking.

Co-expression network from TCGA data. We calculated the pairwise Pearson correlation of gene expression z scores of the 67 genes in samples where both genes had a neutral copy number.

Survival analysis. Gene alterations used in the survival analysis here are the same as defined in 'Detection of significantly co-altered gene pairs' section. For a pair of genes, we examined the following three questions: (i) Are single-gene alterations (alteration of either gene) and double gene alterations associated with different survival prospects? (ii) Does this association hold after accounting for commonly mutated genes in cancers? (iii) Does this association hold after accounting for tumour subtypes? We further examined whether genome instability is a confounding factor in the above analysis.

At the core of the analysis to address these questions is survival analysis, which estimates the survival curves (that is, the survival probability as a function of the survival time) for groups with different alteration patterns (such as single alteration versus double alteration), and compares the statistical significance between the estimated survival curves. A log rank test based on a χ^2-distribution was used for the comparison. We used functions survfit and survdiff in the R package survival for this analysis. Similar to the detection of significant interactions, we computed the Q values by using the R function Q value from package Q value. Q values < 0.05 were considered significant. In each of the three analyses, we further considered three cases of alteration patterns: gain–gain, gain–loss and loss–loss, and required that there are at least five cases with double alteration for a gene pair in any of the three cases.

To address question (ii), we extracted patient samples with TP53 deletion, and performed survival analysis in these samples.

To address question (iii), we extracted breast cancer patients diagnosed as ER-positive or ER-negative and compared single and double alterations in each of the two cases.

To examine the impact of genomic instability, we measured genomic instability in two ways: one is the metric defined by TCGA, which is the fraction of the genome being altered and is based on copy number changes, and the other is our own metric, which is the total number of alterations in the genome. We performed beta regression[60] for the fraction, and quasi Poisson[61,62] regression for our own metric, using the overall survival months and survival status as covariates. We used the R function glm for parameter estimation and P value calculation.

Data availability. The analyses in this paper are implemented in the R package cancerGI, which will be available on CRAN (http://cran.r-project.org).

References

1. Hanahan, D. & Weinberg, R. A. Hallmarks of cancer: the next generation. *Cell* **144,** 646–674 (2011).
2. Vogelstein, B. & Kinzler, K. W. Cancer genes and the pathways they control. *Nat. Med.* **10,** 789–799 (2004).
3. Luo, J., Solimini, N. L. & Elledge, S. J. Principles of cancer therapy: oncogene and non-oncogene addiction. *Cell* **136,** 823–837 (2009).
4. Ashworth, A., Lord, C. J. & Reis-Filho, J. S. Genetic interactions in cancer progression and treatment. *Cell* **145,** 30–38 (2011).
5. Land, H., Parada, L. F. & Weinberg, R. A. Tumorigenic conversion of primary embryo fibroblasts requires at least two cooperating oncogenes. *Nature* **304,** 596–602 (1983).
6. Tong, A. H. *et al.* Global mapping of the yeast genetic interaction network. *Science* **303,** 808–813 (2004).
7. Collins, S. R. *et al.* Functional dissection of protein complexes involved in yeast chromosome biology using a genetic interaction map. *Nature* **446,** 806–810 (2007).
8. Costanzo, M. *et al.* The genetic landscape of a cell. *Science* **327,** 425–431 (2010).
9. Dixon, S. J., Costanzo, M., Baryshnikova, A., Andrews, B. & Boone, C. Systematic mapping of genetic interaction networks. *Annu. Rev. Genet.* **43,** 601–625 (2009).
10. Bandyopadhyay, S. *et al.* Rewiring of genetic networks in response to DNA damage. *Science* **330,** 1385–1389 (2010).
11. Horn, T. *et al.* Mapping of signaling networks through synthetic genetic interaction analysis by RNAi. *Nat. Methods* **8,** 341–346 (2011).
12. Laufer, C., Fischer, B., Billmann, M., Huber, W. & Boutros, M. Mapping genetic interactions in human cancer cells with RNAi and multiparametric phenotyping. *Nat. Methods* **10,** 427–431 (2013).
13. Roguev, A. *et al.* Quantitative genetic-interaction mapping in mammalian cells. *Nat. Methods* **10,** 432–437 (2013).
14. Bassik, M. C. *et al.* A systematic mammalian genetic interaction map reveals pathways underlying ricin susceptibility. *Cell* **152,** 909–922 (2013).
15. Luo, J. *et al.* A genome-wide RNAi screen identifies multiple synthetic lethal interactions with the Ras oncogene. *Cell* **137,** 835–848 (2009).
16. Barbie, D. A. *et al.* Systematic RNA interference reveals that oncogenic KRAS-driven cancers require TBK1. *Nature* **462,** 108–112 (2009).
17. Fong, P. C. *et al.* Inhibition of poly(ADP-ribose) polymerase in tumors from BRCA mutation carriers. *New Engl. J. Med.* **361,** 123–134 (2009).
18. TCGA. Comprehensive molecular portraits of human breast tumours. *Nature* **490,** 61–70 (2012).
19. Soule, H. D. *et al.* Isolation and characterization of a spontaneously immortalized human breast epithelial cell line, MCF-10. *Cancer Res.* **50,** 6075–6086 (1990).
20. Kadota, M. *et al.* Delineating genetic alterations for tumor progression in the MCF10A series of breast cancer cell lines. *PLoS ONE* **5,** e9201 (2010).
21. Zellner, A. An efficient method of estimating seemingly unrelated regressions and tests for aggregation bias. *J. Am. Statist. Assoc.* **57,** 348–368 (1962).
22. Humphries, M. D. & Gurney, K. Network 'small-world-ness': a quantitative method for determining canonical network equivalence. *PLoS ONE* **3,** e0002051 (2008).
23. Sherr, C. J. D-type cyclins. *Trends Biochem. Sci.* **20,** 187–190 (1995).
24. Patel, K. J. *et al.* Involvement of Brca2 in DNA repair. *Mol. Cell* **1,** 347–357 (1998).
25. Kurdistani, S. K. *et al.* Inhibition of tumor cell growth by RTP/rit42 and its responsiveness to p53 and DNA damage. *Cancer Res.* **58,** 4439–4444 (1998).
26. Ishii, H., Wang, Y. & Huebner, K. A Fhit-ing role in the DNA damage checkpoint response. *Cell Cycle* **6,** 1044–1048 (2007).
27. Yan, J. *et al.* Menin interacts with IQGAP1 to enhance intercellular adhesion of beta-cells. *Oncogene* **28,** 973–982 (2009).
28. Feng, Z. J. *et al.* Lung cancer cell migration is regulated via repressing growth factor PTN/RPTP beta/zeta signaling by menin. *Oncogene* **29,** 5416–5426 (2010).
29. Mariko, B. *et al.* Microfibrils and fibrillin-1 induce integrin-mediated signaling, proliferation and migration in human endothelial cells. *Am. J. Physiol. Cell Physiol.* **299,** C977–C987 (2010).
30. Lamb, R. F. *et al.* The TSC1 tumour suppressor hamartin regulates cell adhesion through ERM proteins and the GTPase Rho. *Nat. Cell Biol.* **2,** 281–287 (2000).
31. Sternlicht, H. *et al.* The t-complex polypeptide 1 complex is a chaperonin for tubulin and actin in vivo. *Proc. Natl Acad. Sci. USA* **90,** 9422–9426 (1993).

9

32. Bateson, W. & Mendel, G. *Mendel's Principles of Heredity* (Cambridge University Press, 1909).

33. Wu, X. *et al.* Microcephalin regulates BRCA2 and Rad51-associated DNA double-strand break repair. *Cancer Res.* **69,** 5531–5536 (2009).

34. Yan, M. *et al.* Activation of stress-activated protein kinase by MEKK1 phosphorylation of its activator SEK1. *Nature* **372,** 798–800 (1994).

35. Wu, T., Zhang, X., Huang, X., Yang, Y. & Hua, X. Regulation of cyclin B2 expression and cell cycle G2/m transition by menin. *J. Biol. Chem.* **285,** 18291–18300 (2010).

36. Wang, C. Y., Mayo, M. W., Korneluk, R. G., Goeddel, D. V. & Baldwin, Jr A. S. NF-kappaB antiapoptosis: induction of TRAF1 and TRAF2 and c-IAP1 and c-IAP2 to suppress caspase-8 activation. *Science* **281,** 1680–1683 (1998).

37. Roos, W. P. & Kaina, B. DNA damage-induced cell death by apoptosis. *Trends Mol. Med.* **12,** 440–450 (2006).

38. Derijard, B. *et al.* Independent human MAP-kinase signal transduction pathways defined by MEK and MKK isoforms. *Science* **267,** 682–685 (1995).

39. Kong, S. L., Li, G., Loh, S. L., Sung, W. K. & Liu, E. T. Cellular reprogramming by the conjoint action of ERalpha, FOXA1, and GATA3 to a ligand-inducible growth state. *Mol. Syst. Biol.* **7,** 526 (2011).

40. Theodorou, V., Stark, R., Menon, S. & Carroll, J. S. GATA3 acts upstream of FOXA1 in mediating ESR1 binding by shaping enhancer accessibility. *Genome Res.* **23,** 12–22 (2013).

41. Kouros-Mehr, H., Slorach, E. M., Sternlicht, M. D. & Werb, Z. GATA-3 maintains the differentiation of the luminal cell fate in the mammary gland. *Cell* **127,** 1041–1055 (2006).

42. Albergaria, A. *et al.* Expression of FOXA1 and GATA-3 in breast cancer: the prognostic significance in hormone receptor-negative tumours. *Breast Cancer Res.* **11,** R40 (2009).

43. Encode Project Consortium *et al.* An integrated encyclopedia of DNA elements in the human genome. *Nature* **489,** 57–74 (2012).

44. Pereira, R. *et al.* FLI-1 inhibits differentiation and induces proliferation of primary erythroblasts. *Oncogene* **18,** 1597–1608 (1999).

45. Kornblau, S. M. *et al.* Abnormal expression of FLI1 protein is an adverse prognostic factor in acute myeloid leukemia. *Blood* **118,** 5604–5612 (2011).

46. Mani, R., St Onge, R. P., Hartman, J. L. t., Giaever, G. & Roth, F. P. Defining genetic interaction. *Proc. Natl Acad. Sci. USA* **105,** 3461–3466 (2008).

47. Breslow, D. K. *et al.* A comprehensive strategy enabling high-resolution functional analysis of the yeast genome. *Nat. Methods* **5,** 711–718 (2008).

48. Baryshnikova, A. *et al.* Quantitative analysis of fitness and genetic interactions in yeast on a genome scale. *Nat. Methods* **7,** 1017–1024 (2010).

49. Axelsson, E. *et al.* Extracting quantitative genetic interaction phenotypes from matrix combinatorial RNAi. *BMC Bioinformatics* **12,** 342 (2011).

50. Cordell, H. J. Epistasis: what it means, what it doesn't mean, and statistical methods to detect it in humans. *Hum. Mol. Genet.* **11,** 2463–2468 (2002).

51. Davoli, T. *et al.* Cumulative haploinsufficiency and triplosensitivity drive aneuploidy patterns and shape the cancer genome. *Cell* **155,** 948–962 (2013).

52. Cerami, E. *et al.* The cBio cancer genomics portal: an open platform for exploring multidimensional cancer genomics data. *Cancer Discov.* **2,** 401–404 (2012).

53. Pau, G., Fuchs, F., Sklyar, O., Boutros, M. & Huber, W. EBImage–an R package for image essing with applications to cellular phenotypes. *Bioinformatics* **26,** 979–981 (2010).

54. Fuchs, F. *et al.* Clustering phenotype populations by genome-wide RNAi and multiparametric imaging. *Mol. Syst. Biol.* **6,** 370 (2010).

55. Storey, J. D., Taylor, J. E. & Siegmund, D. Strong control, conservative point estimation and simultaneous conservative consistency of false discovery rates: a unified approach. *J. R. Stat. Soc. B* **66,** 187–205 (2004).

56. Henningsen, A. & Hamann, J. D. Systemfit: A package for estimating systems of simultaneous equations in R. *J. Stat. Softw.* **23,** 1–40 (2007).

57. Walhout, M., Vidal, M. & Dekker, J. *Handbook of Systems Biology: Concepts and Insights* (Academic Press, 2012).

58. Erdos, P. & Renyi, A. On random graphs. *Publ. Math.* **6,** 290–297 (1959).

59. Freeman, L. C. Centrality in social networks I: conceptual clarification. *Soc. Networks* **1,** 215–239 (1979).

60. Ferrari, S. L. P. & Cribari-Neto, F. Beta regression for modelling rates and proportions. *J. Appl. Stat.* **31,** 799–815 (2004).

61. Jones, T. R., Carpenter, A. E. & Golland, P. in *Proceedings of the Workshop on Computer Vision for Biomedical Image Applications Lecture Notes in Computer Science 3765* 535–543 (Springer-Verlag, 2005).

62. Csardi, G. & Nepusz, T. The igraph software package for complex network research. *InterJournal.* Complex Systems **1695** (2006).

Author contributions

X.W. and K.P.W. designed the study. X.W. performed the experiments. X.W., A.Q.F. and M.E.M. analysed/interpreted data. A.Q.F. performed statistical analyses. X.W., A.Q.F., M.E.M. and K.P.W. wrote the manuscript. X.W. and A.Q.F. contributed equally to the study.

Additional information

Accession codes: RNA-seq data has been deposited in the Gene Expression Omnibus under the accession code GSE58834.

Supplementary Information accompanies this paper at http://www.nature.com/naturecommunications

Competing financial interests: The authors declare no competing financial interests.

Reprints and permission information is available online at http://npg.nature.com/reprintsandpermissions/

How to cite this article: Wang, X. *et al.* Widespread genetic epistasis among cancer genes. *Nat. Commun.* 5:4828 doi: 10.1038/ncomms5828 (2014).