Published online 4 February 2015

Nucleic Acids Research, 2015, Vol. 43, No. 4 ***1997–2007***
doi: 10.1093/nar/gkv074

Identifying core gene modules in glioblastoma based on multilayer factor-mediated dysfunctional regulatory networks through integrating multi-dimensional genomic data

Yanyan Ping†, Yulan Deng†, Li Wang, Hongyi Zhang, Yong Zhang, Chaohan Xu, Hongying Zhao, Huihui Fan, Fulong Yu, Yun Xiao* and Xia Li*

College of Bioinformatics Science and Technology, Harbin Medical University, Harbin, Heilongjiang 150086, China

Received December 23, 2013; Revised January 05, 2015; Accepted January 20, 2015

ABSTRACT

The driver genetic aberrations collectively regulate core cellular processes underlying cancer development. However, identifying the modules of driver genetic alterations and characterizing their functional mechanisms are still major challenges for cancer studies. Here, we developed an integrative multi-omics method CMDD to identify the driver modules and their affecting dysregulated genes through characterizing genetic alteration-induced dysregulated networks. Applied to glioblastoma (GBM), the CMDD identified a core gene module of 17 genes, including seven known GBM drivers, and their dysregulated genes. The module showed significant association with shorter survival of GBM. When classifying driver genes in the module into two gene sets according to their genetic alteration patterns, we found that one gene set directly participated in the glioma pathway, while the other indirectly regulated the glioma pathway, mostly, via their dysregulated genes. Both of the two gene sets were significant contributors to survival and helpful for classifying GBM subtypes, suggesting their critical roles in GBM pathogenesis. Also, by applying the CMDD to other six cancers, we identified some novel core modules associated with overall survival of patients. Together, these results demonstrate integrative multi-omics data can identify driver modules and uncover their dysregulated genes, which is useful for interpreting cancer genome.

INTRODUCTION

Genetic alteration was a major mechanism underlying the transformation of normal cells to cancerous cells (1,2). Driver alterations can cooperatively disrupt a cascade of crucial biological pathways and thus provide cells with significant growth advantages (3,4). However, due to the extensive complexity of genetic alterations in cancer genome, how to identify the driver modules and characterize their functional mechanism during carcinogenesis is still a major challenge.

Initially, multiple computational methods were proposed to identify driver genes based on somatic mutation rates in cancer patient populations (5–7). With the whole-genome measurements of somatic mutations and copy number variants in the mass of cancer samples, several cancer studies revealed that driver genetic alterations involved in the same biological pathways showed mutual exclusivity (8,9). Based on the observation of mutual exclusivity, some computational methods were developed to identify causal gene modules in which genomic alterations of their member genes cover the majority of cancer samples and exhibit significant mutual exclusivity (10–13). Similarly, there are also some computational methods based on co-occurrence alteration events for identifying driver genes (14,15). Recently, combination of DNA copy number/mutation profiles and gene expression profiles were used to find driver genes by quantifying the effect of genetic alterations on the transcriptional changes (16,17). For instance, Masica *et al.* (18) proposed a fisher-based method to identify driver genes by calculating the correlations between somatic mutations and expression of outlying genes. DriverNet was developed to find the minimum number of driver genes whose genetic alterations can explain transcriptional changes of all genes across cancer samples, to the largest extent, by integrating the matched genome and transcriptome data (19). Although many com-

*To whom correspondence should be addressed. Tel: +86 451 86615922; Fax: +86 451 86615922; Email: lixia@hrbmu.edu.cn
Correspondence may also be addressed to Yun Xiao. Tel: +86 451 86615922; Fax: +86 451 86615922; Email: xiaoyun@ems.hrbmu.edu.cn
†These authors contributed equally to the paper as first authors.

putational methods have been proposed to identify potential driver genetic alterations, it is very limited to understand how driver genes result in extensive transcriptional changes and in turn contribute to malignant phenotypes (20,21).

In general, gene expression can be regulated by multilayer factors such as copy number, methylation, transcription factors (TFs) and miRNAs (22–24). Recurrent copy number alterations of ∼76% genes exhibited significant correlations with gene expression in glioblastoma (24). The promoter methylation of genes can repress their own gene expression, such as the methylation of O-6-methylguanine-DNA methyltransferase (MGMT) in glioblastoma (25). In addition, TFs like oncogenic v-myc avian myelocytomatosis viral oncogene homolog (MYC) were capable of shaping tumor-specific gene expression profile (26). Also, most miRNAs predominantly decrease the mRNA levels of their targets by binding to 3′ untranslated region (UTR) (27). These factors constitute a regulatory system for maintaining homeostasis of normal cells (28–31). Diverse genetic alterations may coordinately disrupt the delicate balance by perturbing different types of regulatory relationships (32), which can finally lead to the transition from normal cell to cancer cell (33,34). Based on the principle, Jornsten *et al.* combined mRNA transcription regulation and turnover in a steady-state condition and constructed a CNA-driven network using lasso regression to identify driver copy number alterations (CNAs) and detect their effect on transcription in glioblastoma (GBM) (35). Subsequently, an integrative regression model further combined regulatory sequence information of TFs and miRNAs to explain the expression changes in each tumor sample, aiming to inferring transcriptional and miRNA-mediated regulatory program in GBM (36). However, little is known about the effect of driver alterations on the complex regulatory mechanisms of gene expression. Therefore, characterizing dysregulated network induced by genetic alterations in cancer will not only help us discover causal genes but also provide the details for understanding the molecular mechanism of causal genes.

In this work, we proposed a systematic method, called CMDD (Core Modules Driving Dysregulation in cancer) to identify driver modules and their dysregulated genes based on multilayer factor-mediated dysregulated networks through integrating multi-dimensional genomic data including DNA mutation, copy number, methylation, gene expression and miRNA expression profiles, as well as regulatory networks and protein interaction network. We applied our method to GBM and identified a core module consisting of 17 genes and their dysregulated genes. Seven well-known GBM-associated genes were included in the module. Moreover, we found that the module was significantly involved in many cancer-related pathways and was associated with survival. We further explored the dysregulated genes associated with the module, and revealed that two genes subsets participated in GBM-related important pathways with different ways. Also, we applied CMDD to other types of cancer and identified some novel cancer-associated modules.

MATERIALS AND METHODS

Multi-dimensional genomic data

The multi-dimensional GBM-associated data sets containing DNA mutation, copy number, DNA methylation, mRNA and miRNA expression data were collected from TCGA data portal (https://tcga-data.nci.nih.gov/tcga/). We extracted a somatic mutation data (level 2) involving 291 GBM samples directly from TCGA. Through removing silent mutations, a total of 8299 mutation genes were obtained. For a CNV array data (SNP6.0) referring to 530 GBM samples, the segmentations were identified using the circular binary segmentation method-based copy numbers estimated by calculation of normalized log2-tranformed ratios (level 3) (7). The level 3 data of DNA methylation data of 296 samples detected by JHU-USC HumanMenthylation27 were downloaded and the average methylation value for a given gene was calculated. The gene expression profile of 12042 genes in 538 samples and the miRNA expression profile of 470 human miRNAs in 500 samples were obtained. We also obtained clinical information of 465 GBM samples. Finally, 121 GBM samples having all of these genomic data were determined.

MiRNA target interactions between 1909 miRNAs and 16708 genes were downloaded from the database miRecords (37), which was an integrated miRNA-target resource. We obtained experimentally validated TF-gene regulatory relationships from the TRANSFAC (38) and predicted novel regulatory relationships using conserved TF binding sites from UCSC genome browser. The protein interaction network was obtained from the Human Protein Reference Database (HPRD) (39).

Overview of the method CMDD

We proposed a systematic method CMDD to identify cancer core gene modules and their affecting dysregulated genes through multilayer factor-mediated dysregulated networks (Figure 1). The detailed description of the method is as follows.

Construction of genetic alteration profiles of candidate genes

We combined the DNA copy number data and DNA mutation data to construct the genetic alteration profile. The GISTIC (version 2) was applied to the DNA copy number data, and five types of discretized copy number calls (homozygous deletion, heterozygous deletion, diploid, gain and amplification) for each gene in all cancer samples were determined. Only the amplification and homozygous deletions were considered as copy number alterations. Combining with DNA mutation data, we generated a binary profile D of genetic alteration, in which d_{ij} was scored as 1 if the ith gene in the jth sample shows genetic alteration (i.e. amplification/homozygous deletion/mutation), and 0 otherwise.

Next, we used two criteria to select the candidate genes (CGs) through integrating the binary profile of genetic alteration and gene expression profile across cancer samples: (i) CGs should alter in at least 10 samples; ii) for a given CG, it should be differentially expressed between patients

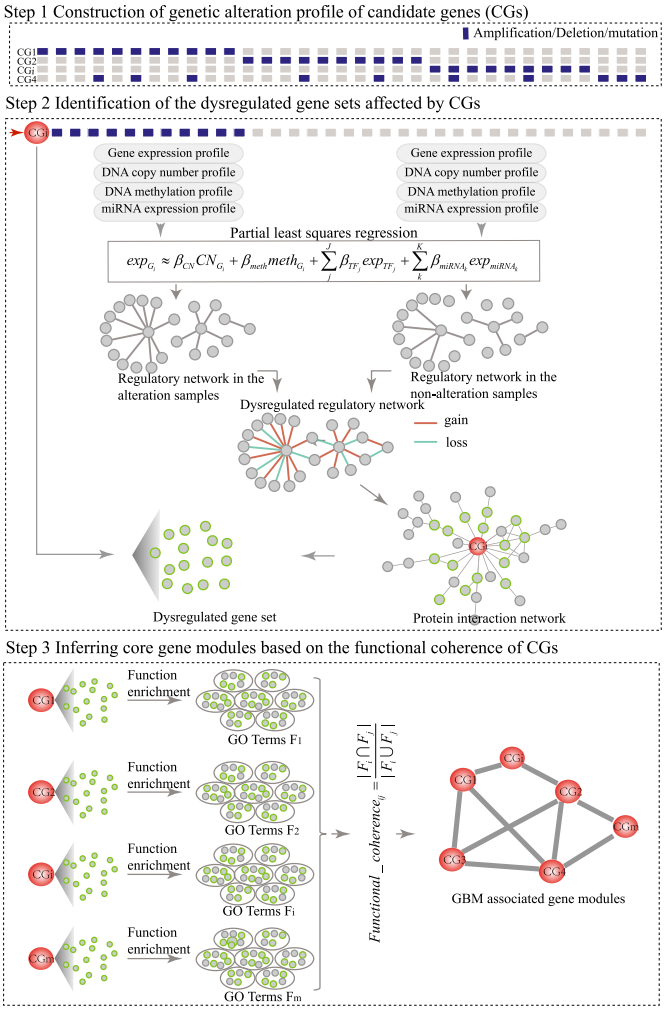

Figure 1. The workflow of the systematic method for identifying cancer core gene modules through integrating multi-dimensional genomic data.

with and without genetic alterations of this CG using *t*-test at false discovery rate (FDR) ≤ 0.05. Then, the binary profile of the selected CGs was obtained.

Identification of dysregulated gene sets affected by CGs

To identify the dysregulated gene sets of CGs, we built CG-associated regulatory networks under the conditions with/without genetic alterations of CGs and constructed corresponding dysregulated networks by network comparisons.

Construction of regulatory networks under the conditions with and without genetic alterations. For every CG, we grouped the cancer samples into two groups according to its alteration status (one group with genetic alteration and the other without). Then, we constructed linear regression models for each gene G_i ($i = 1,2,\ldots,n$) separately in the two groups of samples to explain gene expression changes using multi-layer regulatory factors including DNA copy number, methylation, TF and miRNA regulations as covariates.

Given a gene i (G_i) in a specific condition containing N samples, there are J TFs ($\mathrm{TF}_1, \mathrm{TF}_2, \ldots, \mathrm{TF}_j, \ldots, \mathrm{TF}_J$) and K miRNAs ($\mathrm{miRNA}_1, \mathrm{miRNA}_2, \cdots, \mathrm{miRNA}_k, \cdots, \mathrm{miRNA}_K$) binding G_i. A linear regression model was trained as

$$\exp_{G_i} \approx \beta_{\mathrm{CN}}\mathrm{CN}_{G_i} + \beta_{\mathrm{meth}}\mathrm{meth}_{G_i} + \sum_{j}^{J} \beta_{\mathrm{TF}_j}\exp_{\mathrm{TF}_j} + \sum_{k}^{K} \beta_{\mathrm{miRNA}_k}\exp_{\mathrm{miRNA}_k},$$

where the $\exp_{G_i}$ is the expression levels of G_i in N samples, CN_{G_i} is the copy number calls of G_i, meth_{G_i} is the methylation values of G_i, $\exp_{\mathrm{TF}_j}$ is the expression levels of the jth TF regulating G_i and $\exp_{\mathrm{miRNA}_k}$ is the expression levels of the kth miRNA targeting G_i. β_{CN},β_{meth}, β_{TF_j} and β_{miRNA_k} represent regression coefficients of CN_{G_i}, meth_{G_i}, $\exp_{TF_j}$ and $\exp_{\mathrm{miRNA}_j}$, respectively. Only the genes that were differentially expressed between cancer samples and normal samples and showed high expression variability (within the top 70%) across cancer samples (see Supplementary Materials for details) were used to train regression models. As the large number of variables and their possible high collinearity, we used the partial least squares (40) model to train the linear regression model (see Supplementary Materials). Then, using the functions ('plsr', 'RMSEP' and 'jack.test') in R package 'pls', the significance of the effect of input variables on gene expression variation was estimated using 10-fold cross-validation. Subsequently, we determined the regulatory factors (including copy number, methylation, TFs and miRNAs) with significant coefficients (FDR ≤ 0.001) for each gene and two regulatory networks corresponding to different cancer groups were formed.

Construction of dysregulated networks affected by CGs. For each CG, we compared the two regulatory networks in the groups with/without genetic alterations. The relationships that were only presented in one of the two regulatory networks were identified and defined as the dysregulated regulatory relationships. To determine the significance of the dysregulated relationships, the genetic alteration profile was permuted 1000 times, preserving alteration frequency of CGs. We re-built 1000 random dysregulated networks and calculated the frequencies of the real dysregulated relationships in the 1000 random networks. Only the dysregulated relationships with less than 10% of frequency were used to form the dysregulated network.

Extraction of dysregulated genes affected by CGs. Next, in order to obtain dysregulated genes affected by CGs, the genes in the dysregulated network of a given CG were mapped onto the protein–protein interaction network. The genes within two-step distance from this CG in the protein interaction network were selected as the dysregulated genes of the CG.

Inferring core gene modules based on the functional coherence of CGs

We assumed that different cancer-associated genes disturbed the same or similar functions during tumorigenesis.

Based on the hypothesis, for each CG, we determined the GO biological functions significantly enriched by its dysregulated genes using hypergeometric distribution test ($P < 0.001$). Then, we calculated a functional coherence score for each pair of CGs as follows:

$$\text{Functional\_coherence}_{ij} = \frac{|F_i \cap F_j|}{|F_i \cup F_j|},$$

where F_i and F_j represent significantly enriched function sets associated with the ith and jth CGs, respectively, $|F_i \cap F_j|$ represents the number of their common functions and $|F_i \cup F_j|$ represents the total number of unique functions associated with them. Finally, the CG pairs with coherence scores larger than a threshold that was determined through analyzing the distribution of coherence scores were selected to constitute cancer core gene modules.

Survival analysis

We used the survival analysis to test the association of the genetic alteration of single genes or modules with cancer overall survivals. Survival analysis was performed using the R package 'survival'.

Survival analysis for single genes. For a given single gene, we divided the cancer samples into two groups according to the genetic alteration status of the gene: the alteration group containing the samples with genetic alteration of this gene and the non-alteration group without. The significance of survival difference between the alteration and non-alteration groups was estimated using Kaplan–Meier analysis and log-rank test.

Survival analysis for core modules. For a given module, we divided the cancer samples into two groups according to the genetic alteration status of all genes in the module: the alteration group containing the samples with genetic alteration in at least one gene of the module and the non-alteration group without any alterations in the genes of the module. Likewise, the significance of survival difference between alteration and non-alteration groups was estimated using Kaplan–Meier analysis and log-rank test.

RESULTS

Identifying the core gene module in GBM

Identifying cancer-associated core module of genetic alterations was a key step for understanding fundamental mechanisms of carcinogenesis. We developed a systematic method CMDD based on dysregulated genes affected by genetic alterations. By applying CMDD to the multi-dimensional data sets of 121 GBM samples including DNA mutation, copy number, methylation, gene expression and miRNA expression profiles, we identified the dysregulated genes for each CG (see the Materials and Methods section for details) and determined the GO biological processes significantly enriched by the dysregulated genes of each CG ($P = 0.001$, hypergeometric distribution test). Notably, their functional coherence scores showed a bimodal distribution. We set the threshold of 0.4 that can distinguish the two peaks to find CGs with highly consistent functions (Figure 2A). Then, a core gene module composed of 17 genes including *EGFR*, *PDGFRA*, *RB1*, *CDKN2A*, *PTEN*, *MDM4*, *CDK4*, *PIK3C2B*, *KIT*, *PHKG1*, *DDIT3*, *DCTN2*, *KIF5A*, *OS9*, *METTL1*, *AVIL* and *CTDSP2* was identified (Figure 2B).

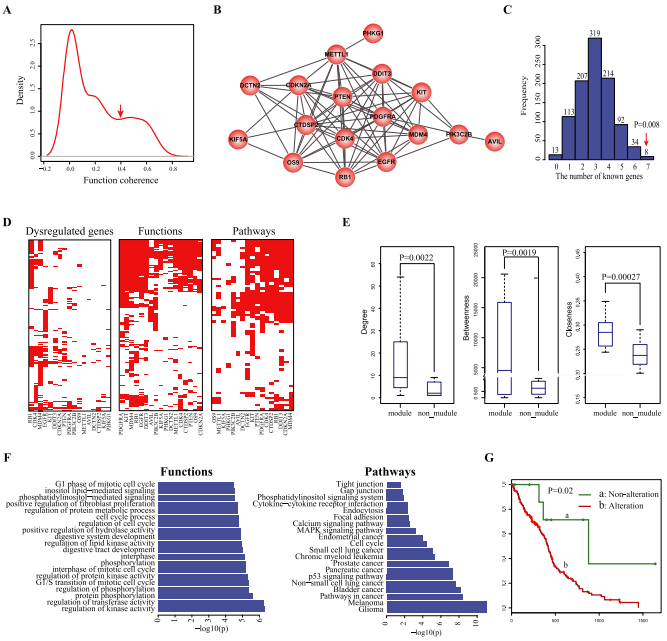

Figure 2. Identification of the core module in GBM. (**A**) The distribution of functional coherence scores of pairwise candidate genes.The arrow represents the position of the selected threshold for determining core functional modules. (**B**) The GBM core gene module composed of 17 highly interconnected genes. (**C**) The distribution of the numbers of known GBM associated genes in randomly selected gene sets. The arrow represents the real number of known GBM associated in the module. (**D**) The heatmaps of dysregulated genes (left), and their enriched functions (middle) and pathways (right) across the 17 genes in the module. (**E**) Comparisons of three topological features including degree (right), betweenness (middle) and closeness (right) of the 17 genes in module with the rest candidate genes not in the module. (**F**) The biological functions (left) and pathways (right) significantly enriched by the 17 genes in the module. (**G**) Kaplan–Meier estimates of overall survival for GBM patients in TCGA data set, with patients stratified into two groups according to the alteration status of member genes in the module.

Among the 17 genes, seven (including *EGFR*, *PDGFRA*, *RB1*, *CDKN2A*, *PTEN*, *MDM4* and *CDK4*) have been reported to be associated with GBM ($P = 0.0014$, hypergeometric distribution test). *PDGFRA* and *EGFR* are the gene signatures for proneural and classical GBM subclass, respectively (41). The dysfunctions of *EGFR* and *PDGFRA* could lead to glioma formation in mouse model. (42,43). Mutation of *RB1* disturbs its interaction with *E2F*, thus leading to cell cycle disorder and affecting therapeutic efficacy in GBM (44,45). Meanwhile, we randomly selected 17 genes from the original CGs 1000 times and found that the core module identified has more known GBM genes ($P = 0.008$; Figure 2C). In addition to known GBM-associated genes, two novel genes including *METTL1* and *CTDSP2* showing the highest degree in the core module were identified (Supplementary Figure S1). We found that the dysregulated genes of both *METTL1* and *CTDSP2* were signifi-

cantly enriched in many cancer-related functions and pathways. Moreover, the survival analyses based on their genetic alteration profiles and expression profiles consistently showed that *METTL1* and *CTDSP2* were associated with shorter survival in GBM (Supplementary Figures S2 and S3; see Supplementary Materials for details). These findings suggested that *METTL1* and *CTDSP2* were the potential novel genes whose genetic alterations may contribute to the development of GBM.

In addition, we found that three topological features (including degree, betweenness and closeness) of the 17 genes in the protein interaction network were significantly higher than those of the rest CGs (Wilcoxon–Wilcox test, $P = 0.002$ for degree, $P = 0.0019$ for betweenness and $P = 0.00027$ for closeness; Figure 2E), highlighting their close functional associations. Then, we performed function and pathway enrichment analyses directly using the 17 genes in the module. We found that the module participated in cancer-associated functions and pathways such as 'Glioma', 'G1 phase of mitotic cell cycle' and 'regulation of protein kinase activity' (Figure 2F). Also, the prognostic effect of the module on GBM clinical outcomes was analyzed. We classified GBM samples into two groups: the alteration group contained patients harboring the alteration of at least one member gene in the module and the non-alteration group containing patients without any alterations of member genes. We found the alteration of the module was associated with shorter survival ($P = 0.02$, log-rank test; Figure 2G). Together, the GBM-associated core gene module identified by our method not only identified the known and novel GBM-associated genes but also determined their functional relationships in GBM, which could be an effective prognostic indicator of GBM.

Characterization of dysregulated genes of the core module

Next, we characterized different property of the dysregulated genes of the core module. Through the expression correlation analysis of dysregulated genes, we found that dysregulated genes induced by member genes in the module tend to show significantly higher expression correlations compared to randomly selected genes ($P < 0.05$, Kolmogorov–Smirnov test, Supplementary Figure S4; see Supplementary Materials for details). Enrichment analyses showed that the dysregulated genes can significantly capture the functions of their corresponding member genes ($P < 0.05$, hypergeometric test; Supplementary Figure S5). Moreover, after knockdown of *CDK4* (GSE8866) (46) and *RB1* (GSE31534) (47), both of their dysregulated genes showed significant enrichments in the differentially expressed genes ($P = 0.089$ for *CDK4* and $P = 0.02$ for *RB1*). Notably, despite the high overlapping of functions and pathways between these 17 genes, only a few common dysregulated genes were observed (Figure 2D), suggesting that these genes in the module influenced similar biological processes through different dysregulated genes. In summary, our results showed that dysregulated genes could effectively reflect the abnormal molecular events associated with genetic alteration of CGs.

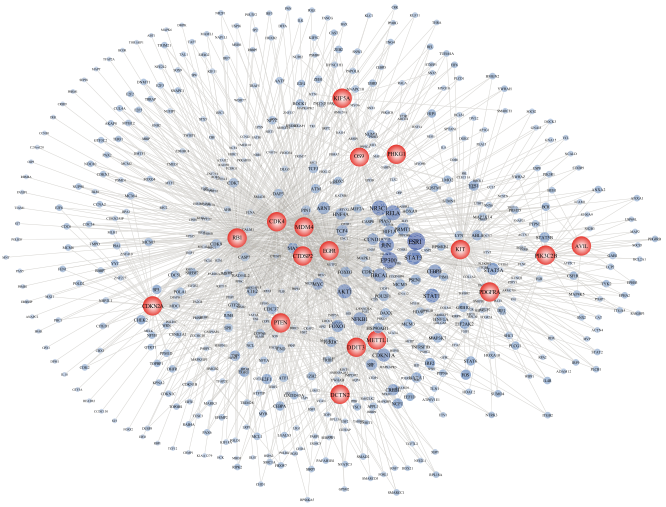

Figure 3. Comprehensive dysregulated network induced by the core module. Two types of nodes represent the member genes and dysregulated gens, respectively. The node size for member genes is proportion to the number of member genes connecting to the dysregulated genes.

GBM-associated dysregulated network mediated by multi-factors

To further characterize how the genetic alterations of member genes in the core module affected the regulatory relationships of their dysregulated genes, we constructed a network by connecting the 17 member genes in the module to their dysregulated genes (Figure 3). In this network, we found that most dysregulated genes linked with only one or two member genes. Among the dysregulated genes, genes affected by multiple member genes were more likely to be cancer-associated genes reported in OMIM (48), CGC (49) or GAD (50) such as *ESR1*, *AKT1* and *BRCA1* (Supplementary Figure S6). Especially, *ESR1* was found to be connected by the largest number of member genes (15/17 member genes; Figure 3). Previous studies have demonstrated that the encoded protein of *ESR1* was involved in pathological processes of multiple cancers (51,52). And Uhlmann *et al.* discovered a putative association between *ESR1* and gliomagenesis (53). Furthermore, we also found high expression of ESR1 was associated with favorable survival in TCGA GBM data set ($P = 0.007$, log-rank test; Supplementary Figure S7). These findings further support the importance of the member genes in GBM.

In addition, we observed that the dysregulated genes of the module were consistently implicated in many cancer-related biological processes (such as 'epidermal growth factor receptor signaling pathway', 'glial cell differentiation', 'apoptotic signaling pathway' and 'MAPK cascade') and pathways (such as Glioma, 'p53 signaling pathway' and 'Notch signaling pathway') (Figure 4A). For each member gene, the regulatory factors of its dysregulated genes were analyzed. We found that TFs accounted for most dysregulated genes (Figure 4B), suggesting that disturbance of these cancer-related pathways is primarily dependent on dysfunction of TF-mediated regulations. We further analyzed a well-known GBM-associated pathway 'epidermal growth factor receptor signaling pathway'. Nine of the 17

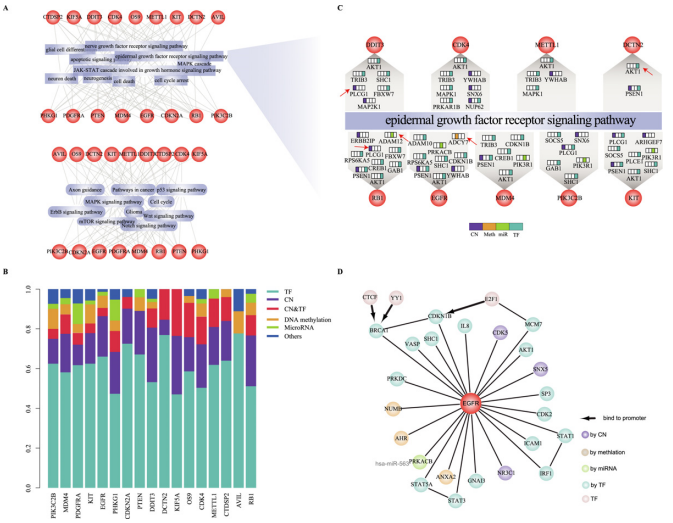

Figure 4. Estimation of multi-factor mediated dysregulations affected by the module. (**A**) The cancer associated biological processes (top) and pathways (bottom) enriched by the dysregulated genes of 17 member genes. (**B**) The transcription factors account for most of dysfunctions of dysregulated genes. (**C**) Nine member genes affect 'epidermal growth factor receptor signaling pathway' through distinct dysregulation of their dysregulated genes. The dysregulation of dysregulated genes is arranged in order: copy number, methylation, miRNA, TF. The genes mentioned in the text are labeled by arrows. (**D**) The literature-curated dysregulated relationships linked with EGFR.

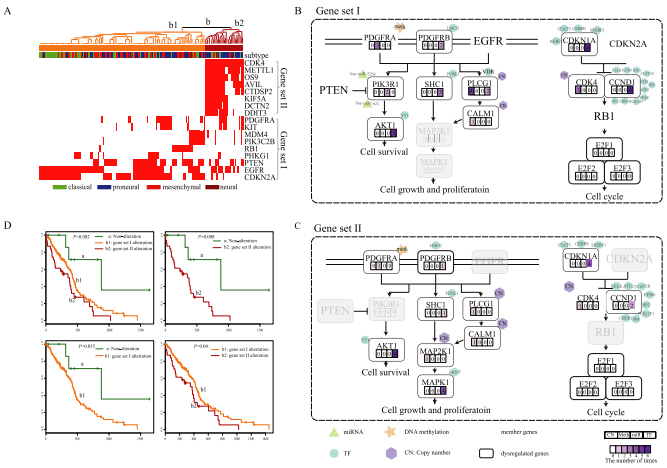

Figure 5. The disturbance of Glioma pathway affected by two gene sets in the core module contributing to GBM survival. (**A**) Two gene sets (I and II) in the core module characterized two subgroups using unsupervised hierarchical clustering of GBM patients with genomic alterations in the module. The 'subtype' bar represents four subtypes (including classical, proneural, mesenchymal and neural) identified using expression of genes determined in a previous study. (**B**) The gene set I including EGFR, PTEN, PDGFRA, CDKN2A, MDM4, KIT, PIK3C2B, RB1 and PHKG1 directly participates in Glioma pathway. (**C**) The gene set II including CDK4, METTL1, OS9, AVIL, CTDSP2, KIF5A, DCTN2 and DDIT3 indirectly regulates the Glioma pathway through dysregulated genes. The digits within the four rectangles represent the frequency of copy number, methylation, miRNA and TF-mediated dysregulations of dysregulated genes. The different regulatory factors are affiliated to the dysregulated genes. (**D**) Both of the two gene sets (I and II) contribute to the GBM survival.

genes influenced dysregulated genes by distinct regulatory relationships and in turn regulated this pathway. For example, the TF-mediated dysregulation of *AKT* was consistently observed in dysregulated networks of eight member genes. The copy number-mediated dysregulation of *PLCG1* was observed for *DDIT3* and *RB1*, while additional TF-mediated dysregulation occurred in *PIK3C2B* and *KIT*. MiRNA-mediated dysregulation of *ADAM12* and DNA methylation-mediated dysregulation of *ADCY7* appeared in *RB1* and *EGFR*, respectively (Figure 4C). Another three cancer-associated biological processes including 'glial cell differentiation', 'apoptotic signaling pathway' and 'MAPK cascade' showed similar multi-factor-mediated dysregulation (Supplementary Figure S8).

Furthermore, by literature searching, we confirmed that many dysregulated relationships linked with *EGFR* had been demonstrated (Figure 4D). For example, genetic alteration of *EGFR* disturbed the expression of *STAT1* and *IRF1* (54). Also, the TF *STAT1*-mediated dysregulation of *IRF1* has been confirmed in the proliferation- and apoptosis-associated signaling pathway (55). *SHC1*, regulated by TF *FOSL1*, was reported to specifically co-purified with mutated *EGFR*, and inhibition of *SHC1* showed significant association with drug resistance in *EGFR* mutant cells (56). The overexpression of Numb induced by genetic alteration of *EGFR* may depend on methylation dysfunction (57). In addition, in the condition of *EGFR* mutation, miR-563-mediated dysfunction of *PRKACB*, a subunit of PKA, was found to promote glioma cell growth and invasion (58). *CDK5* with copy number-mediated dysregulation is a downstream gene in the EGFR-family signaling (59). Taken together, the genetic alterations of member genes in the core modules could induce the multifactor-mediated regulation of dysregulated genes, especially TF-mediated regulation, and in turn disturb the cancer associated functions and pathways.

Dysregulated Glioma pathway induced by the core module contributing to survival

We investigated how the core module and their dysregulated genes participated in Glioma pathway, and contributed to survival. The patients with at least one altered member gene were divided into two subgroups using unsupervised hierarchical clustering based on the genetic alteration profiles of the core gene module (Figure 5A). We found that one subgroup (b1) showed frequent genetic alterations of one gene set (I) including nine genes (*EGFR*, *PTEN*, *PDGFRA*, *CDKN2A*, *MDM4*, *KIT*, *PIK3C2B*, *RB1* and PHKG1), while the other (b2) had the frequent genetic alterations of another gene set (II) including eight genes (*CDK4*, *METTL1*, *OS9*, *AVIL*, *CTDSP2*, *KIF5A*, *DCTN2* and *DDIT3*). Investigation of the dysregulated genes showed that both of these two gene sets were associated with the 'Glioma' pathway. Notably, among the gene set I, five genes including *PDGFRA*, *RB1*, *PTEN*, *EGFR* and *CDKN2A* were directly involved in the 'Glioma' pathway (Figure 5B). In contrast, the gene set II seemed to indirectly regulate the pathway through influencing their dysregulated genes (Figure 5C). Meanwhile, these two gene sets showed local differences in the pathway. *MAPK1* dysregulation was only observed in the pathway indirectly influenced by the gene

set II. *MAPK1* is the convergence point in pathway cross-talk (60), which has been widely validated to be involved in cell cycle (61), angiogenesis (62), invasiveness (63) and other cancer-associated behaviors (64). The dysfunction of *MAPK1* can lead to acquired drug resistance (65) and worse survival (66).

In addition, through survival analysis, we found the two subgroups with distinct alteration patterns (subgroup b1 and subgroup b2 in Figure 5A) showed shorter survival compared to the group without any alterations of member genes (group a in Figure 2G). Interestingly, these two subgroups also had different survival risks (Figure 5D), and the subgroup b2 had the shortest survival. A possible explanation is the local differences in Glioma pathway between the two subgroups, such as the dysregulation of *MAPK1*. Furthermore, this poor-prognosis subgroup was not correlated with any GBM subtypes determined by the expression of 840 genes reported in (41), which suggested a potential new subtype of GBM. These findings indicated that these two mutual exclusive gene sets can induce similar changes of common pathways in different ways and the dysregulated genes identified can provide novel clues for explaining malignant phenotypes in GBM.

Applying CMDD to other cancer types

To reveal gene modules in other cancer types, we applied the CMDD to other six types of cancer, including Ovarian serous cystadenocarcinoma (OV), Head and Neck squamous cell carcinoma (HNSC), Lung adenocarcinoma (LUAD), Cervical squamous cell carcinoma and endocervical adenocarcinoma (CESC), Breast invasive carcinoma (BRCA) and Prostate adenocarcinoma (PRAD) (Supplementary Table S2). Like GBM, we performed similar analyses for each type of cancer (Supplementary Figures S9–S14 and Supplementary Tables S3–S8).

For example, in OV, 31 core modules including 212 genes were identified. The function and pathway enrichment analysis showed that the genes in these core modules were significantly involved in cancer-associated functions and pathways, such as 'mitotic cell cycle', 'p53 signaling pathway' and 'Apoptosis' (Supplementary Figure S9B). A global dysregulated network connecting the genes in modules with their affecting dysregulated genes showed that most dysregulated genes linked with only one or two member genes (Supplementary Figure S9C) and that most genes in modules dysregulated TF-mediated regulations (Supplementary Figure S9D). Also, we found five modules showing significant associations with OV survival. For example, a core module including five completely connected genes (*PUF60*, *SOX18*, *CCDC88A*, *MED4* and *THOC7*) was significantly associated with longer survival ($P = 0.0059$, log-rank test; Supplementary Figure S15A), suggesting that the module functioned as a protective role in the OV survival. Further, we found that the module showed better prognostic effect than any single gene in the module. By analyzing the alteration profiles of these five genes across OV samples, we found that the genetic alterations of these genes exhibited mutually exclusive patterns (Supplementary Figure S15B), indicating that different genes in the module can dysregulate the same or similar important functions and pathways in cancer. Similarly, another core module including *LDHB*, *MRPS22* and *ATP5C1* was also observed. The alterations of the three genes were mutually exclusive, and none of these genes showed significantly association with OV survival, while the prognostic effect of the whole module was significant ($P = 0.007$, log-rank test; Supplementary Figure S15B). We also found the survival-associated modules in other cancers, such as a clique module including *NCL*, *DVL3*, *TRRAP*, *KDM5A*, *KLF5*, *HES1* and *BARD1* ($P = 0.007$, log-rank test; Supplementary Figure S10E) in HNSC and a module including *CSPG4*, *NDUFS6*, *SATB2* and *TRIO* in LUAD ($P = 0.037$, log-rank test; Supplementary Figure S11E).

In addition, we found that genes in some modules showed different roles in prognosis. For instance, using a clique module in HNSC including *ARNT*, *NFE2L2*, *BARD1*, *SUB1*, *ATF2*, *PTEN* and *TOPBP1*, the HNSC samples with the alteration of the module were divided into two subgroups. We found that these two subgroups showed significantly different prognosis ($P = 0.0004$, log-rank test; Supplementary Figure S16). Compared with samples without the alteration of the clique module, the subgroup with frequent alterations of *ATF2*, *BARD1* and *NFE2L2* showed significantly shorter survival time, while the other subgroup harboring frequent alterations of *ARNT*, *SUB1*, *PTEN* and *TOPBP1* exhibited the significantly longer survival times. Similar results were also found for other clique modules, such as a module including *TAF9*, *HOXB7*, *EEF1D* and *RXRB* in OV (Supplementary Figure S17) and a module including *MED21*, *MYC*, *ZHX1*, *KPNA2* and *USF2* in LUAD (Supplementary Figure S18).

Finally, we compared the genes in the core modules among the seven types of cancer and observed that the genes in different types of cancer were obviously different. Also, the dysregulated genes affected by the core modules were mostly different, while the functions and pathways significantly enriched by the dysregulated genes exhibited high coherence (Supplementary Figure S19).

Performance evaluation of the method

We compared our method with two other methods: a frequency-based and a fisher-based method (18). Based on the frequency method, only five of the nine known GBM-associated genes were found (Figure 6A). We also applied the fisher-based method to obtain outlying genes associated with each CG based on the correlation between genetic alteration of CGs and expression of differential genes ($P < 0.05$). Then we ranked the CGs according to the number of outlying genes. In the top 17 genes, only two known GBM-associated genes were identified (Figure 6B). Furthermore, using differentially expressed genes between altered samples and non-altered samples (SAM, fold change ≥ 1.5 and FDR < 0.05) to replace dysregulated genes, we obtained three modules using the same threshold according to functional coherence of differential genes. These modules derived from distinct chromosome regions did not exhibit functional connection and contained only four known GBM-associated genes (Figure 6C), suggesting that multi-factor-mediated dysregulated networks can capture intrinsic functional associations among distinct genomic alterations.

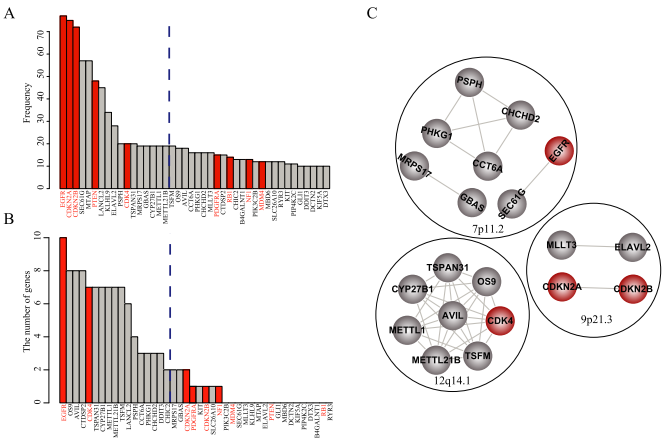

Figure 6. Comparison with other methods. (**A**) Gene list identified by the frequency-based method, which was ranked based on alteration frequency. (**B**) Gene list identified by the fisher-based method, which was ranked by the number of related genes. Dashed lines indicate the threshold for top 17 genes. (**C**) The gene modules identified using the differentially expressed genes. Genes in the same black circle are located in the same chromosomal region. *EGFR*, *CDKN2A*, *CDKN2B*, *PTEN*, *CDK4*, *PDGFRA*, *RB1*, *NF1*, *MDM4* were known GBM associated genes.

DISCUSSION

In this paper, we proposed a systematic method called CMDD to identify cancer-related core gene modules and their affecting dysregulated genes through characterizing genetic alteration-induced dysregulated networks that are constructed using multi-dimensional genomic data including DNA copy number, methylation, gene expression and miRNA expression profiles, as well as protein interaction network. When applied CMDD to GBM, a core module composed of 17 genes was identified. The core module includes several well-known oncogenes and tumor suppressors associated with GBM, such as *EGFR*, *PTEN*, *PDGFRA* and *CDKN2A*, which partially provided an important validation of our method. These genes were found to influence numerous regulatory relationships, such as loss (or gain) of gene regulations by transcriptional factors, which in turn affect downstream key genes with important roles in tumorigenesis. Dysregulated genes induced by the genes in the module show significantly functional connections and are commonly involved in some important cancer-related pathways. In addition, we found novel CGs *METTL1* and *CTDSP2* that showed strong functional associations with known GBM-related genes. Dysregulated genes induced by *METTL1* or *CTDSP2* share similar functions as some known GBM-associated genes, and the amplification/high expression of *METTL1* or *CTDSP2* is associated with poor prognosis in GBM, suggesting that *METTL1* and *CTDSP2* may play an important role in the tumorigenesis and progression of GBM.

Recent studies focusing on somatic tumor genome and exome sequences revealed that most human cancers are resulted from two to eight sequential alterations (2). An interesting problem is how the small number of genetic alterations can destroy key signaling pathways associated with genome maintenance, cell fate and cell survival to cause selective growth advantage and in turn initiate cancer cell transformations. A general explanation is that genes harboring these genetic alterations are directly involved in these key signaling pathways. For example, *EGFR* and *PDGFRA* identified in the core module of GBM are the key members of RTK/Ras/PI3K/AKT signaling pathway, and their alterations can directly prompt activation of this pathway contributing to cell growth, proliferation and motility (67). Our method combined multiple layers of information (CNA, methylation, TFs and miRNAs) to infer dysregulated networks relevant to specific genetic alterations. It not only can identify the core module of genetic alterations with the direct effect on the key pathways but also can find the indirect destroy effect of these genetic alterations, which were propagated to downstream key signaling molecules via the dysregulated network. Genes in the dysregulated network of a given genetic alternation can comprehensively characterize the direct and indirect influence of this alteration. Indeed, our results showed that some key pathways, whose important member genes do not exhibit recurrent genetic alterations in certain cancer patient groups, are disturbed by indirect effects of genes harboring genetic alterations. For example, the majority of member genes in the DNA replication pathway and nod-like receptor signaling pathway were the dysregulated genes without obvious recurrent genetic alterations. Such indirect relationships between genetic alteration and downstream dysregulated genes are hard to be captured only with mutation data.

Interestingly, the core module contained two gene sets (I and II), both of which can characterize a GBM subgroup identified using unsupervised hierarchical clustering, respectively. The gene set I (including *EGFR*, *PTEN*, *PDGFRA*, *CDKN2A*, *MDM4*, KIT, PIK3C2B, RB1 and PHKG1) exerts important roles in the glioma signaling pathway, such as *PTEN*, *RB1* and *EGFR* (9), whereas most of the gene set II (including *CDK4*, *METTL1*, *OS9*, *AVIL*, *CTDSP2*, *KIF5A*, *DCTN2*, *DDIT3*) is not the members of this pathway. Notably, our results of dysregulated genes showed that both of these two gene sets can target the glioma signaling pathway by their dysregulated genes. Furthermore, the genetic alterations of both of two gene sets were significantly associated with survival risk in GBM patients. The GBM subgroup characterized by the gene set II showed significantly shorter survival, without obvious overlap with the four expression-based subtypes of GBM previously reported. These findings suggest that the core module identified by our method is predictive for survival in GBM patients, probably characterizing new GBM subtypes.

In addition, we investigated whether integration of multi-omics data could provide more comprehensive information than using only two or three types of data. We applied our method to GBM by integrating only two or three types of data, including nine different combinations. We found that integration of multi-omics data could capture more regulatory relationships than using two or three types of data, such as regulatory relationships constructed under *EGFR* alteration (Supplementary Figure S20). Almost all of the regulated genes identified by different combinations were included in those identified by multi-omics data (Supplementary Figure S20C, CT: $P = 1.95e\text{-}21$; CMT: $P = 1.4e\text{-}19$; MT: $P = 4.0e\text{-}18$ CHT: $P = 0$; CH: $P = 0$, hypergeometric

test). We further identified the core modules for different combinations. The results showed that no modules were identified in five combinations. In the rest four combinations, we identified only small modules with two to four genes, which were completely contained in the core modules identified by multi-omics data (Supplementary Figure S21A). For example, by integrating copy number and TF, we identified a module including *RB1*, *CDK4*, *PTEN* and *DDIT3* (Supplementary Figure S21B), all of which formed a completely connected graph in the core module identified by integrating multi-omics data. It is worth noting that integration of multi-omics data could identify more known GBM-associated genes (such as *PDGFRA*, *MDM4* and *CDKN2A*) and some potentially novel genes (such as *METTL1* and *CTDSP2*) when compared with those identified by integrating two or three types of data. These results suggested that integrating multiple omics data can capture more useful information which was crucial for identifying core modules and further characterizing their downstream mechanisms.

In summary, we offered an integrative method to identify cancer-related core gene modules and their affecting dysregulated genes through inferring dysregulated regulatory networks induced by genetic alterations. Also, our approach also was applied to other cancer types for identifying causal gene modules and providing the causal gene-induced dysregulated genes, which will give a better interpretation for molecular mechanisms of cancer.

SUPPLEMENTARY DATA

Supplementary Data are available at NAR Online.

FUNDING

The National High Technology Research and Development Program of China (863 Program) [2014AA021102, in part]; the National Program on Key Basic Research Project (973 Program) [2014CB910504]; the National Natural Science Foundation of China [91129710, 61073136, 31200997, 61170154, 81070946]; the National Science Foundation of Heilongjiang Province [C201207, H0906]; Key Laboratory of Cardiovascular Medicine Research (Harbin Medical University), Ministry of Education and the Undergraduate Innovation Funds of Harbin Medical University [YJSCX2012-210HLJ]. Funding for open access charge: the National High Technology Research and Development Program of China (863 Program) [2014AA021102, in part]; the National Program on Key Basic Research Project (973 Program) [2014CB910504]; the National Natural Science Foundation of China [91129710, 61073136, 31200997, 61170154, 81070946]; the National Science Foundation of Heilongjiang Province [C201207, H0906]; Key Laboratory of Cardiovascular Medicine Research (Harbin Medical University), Ministry of Education and the Undergraduate Innovation Funds of Harbin Medical University [YJSCX2012-210HLJ].

Conflict of interest statement. None declared.

REFERENCES

1. Ricke,R.M. and van Deursen,J.M. (2013) Aneuploidy in health, disease, and aging. *J. Cell Biol.*, **201**, 11–21.
2. Vogelstein,B., Papadopoulos,N., Velculescu,V.E., Zhou,S., Diaz,L.A. Jr and Kinzler,K.W. (2013) Cancer genome landscapes. *Science*, **339**, 1546–1558.
3. Jonsson,P.F. and Bates,P.A. (2006) Global topological features of cancer proteins in the human interactome. *Bioinformatics*, **22**, 2291–2297.
4. Qiu,Y.Q., Zhang,S., Zhang,X.S. and Chen,L. (2010) Detecting disease associated modules and prioritizing active genes based on high throughput data. *BMC Bioinformatics*, **11**, 26.
5. Youn,A. and Simon,R. (2011) Identifying cancer driver genes in tumor genome sequencing studies. *Bioinformatics*, **27**, 175–181.
6. Beroukhim,R., Getz,G., Nghiemphu,L., Barretina,J., Hsueh,T., Linhart,D., Vivanco,I., Lee,J.C., Huang,J.H., Alexander,S. *et al.* (2007) Assessing the significance of chromosomal aberrations in cancer: methodology and application to glioma. *Proc. Natl. Acad. Sci. U.S.A.*, **104**, 20007–20012.
7. Mermel,C.H., Schumacher,S.E., Hill,B., Meyerson,M.L., Beroukhim,R. and Getz,G. (2011) GISTIC2.0 facilitates sensitive and confident localization of the targets of focal somatic copy-number alteration in human cancers. *Genome Biol.*, **12**, R41.
8. Cancer Genome Atlas Research Network. (2011) Integrated genomic analyses of ovarian carcinoma. *Nature*, **474**, 609–615.
9. Cancer Genome Atlas Research Network. (2008) Comprehensive genomic characterization defines human glioblastoma genes and core pathways. *Nature*, **455**, 1061–1068.
10. Vandin,F., Upfal,E. and Raphael,B.J. (2012) De novo discovery of mutated driver pathways in cancer. *Genome Res.*, **22**, 375–385.
11. Leiserson,M.D., Blokh,D., Sharan,R. and Raphael,B.J. (2013) Simultaneous identification of multiple driver pathways in cancer. *PLoS Comput. Biol.*, **9**, e1003054.
12. Ciriello,G., Cerami,E., Sander,C. and Schultz,N. (2012) Mutual exclusivity analysis identifies oncogenic network modules. *Genome Res.*, **22**, 398–406.
13. Ping,Y., Zhang,H., Deng,Y., Wang,L., Zhao,H., Pang,L., Fan,H., Xu,C., Li,F., Zhang,Y. *et al.* (2014) IndividualizedPath: identifying genetic alterations contributing to the dysfunctional pathways in glioblastoma individuals. *Mol. Biosyst.*, **10**, 2031–2042.
14. Bredel,M., Scholtens,D.M., Harsh,G.R., Bredel,C., Chandler,J.P., Renfrow,J.J., Yadav,A.K., Vogel,H., Scheck,A.C., Tibshirani,R. *et al.* (2009) A network model of a cooperative genetic landscape in brain tumors. *JAMA*, **302**, 261–275.
15. Klijn,C., Bot,J., Adams,D.J., Reinders,M., Wessels,L. and Jonkers,J. (2010) Identification of networks of co-occurring, tumor-related DNA copy number changes using a genome-wide scoring approach. *PLoS Comput. Biol.*, **6**, e1000631.
16. Akavia,U.D., Litvin,O., Kim,J., Sanchez-Garcia,F., Kotliar,D., Causton,H.C., Pochanard,P., Mozes,E., Garraway,L.A. and Pe'er,D. (2010) An integrated approach to uncover drivers of cancer. *Cell*, **143**, 1005–1017.
17. Kim,Y.A., Wuchty,S. and Przytycka,T.M. (2011) Identifying causal genes and dysregulated pathways in complex diseases. *PLoS Comput. Biol.*, **7**, e1001095.
18. Masica,D.L. and Karchin,R. (2011) Correlation of somatic mutation and expression identifies genes important in human glioblastoma progression and survival. *Cancer Res.*, **71**, 4550–4561.
19. Bashashati,A., Haffari,G., Ding,J., Ha,G., Lui,K., Rosner,J., Huntsman,D.G., Caldas,C., Aparicio,S.A. and Shah,S.P. (2012) DriverNet: uncovering the impact of somatic driver mutations on transcriptional networks in cancer. *Genome Biol.*, **13**, R124.
20. Mehnert,J.M. and Kluger,H.M. (2012) Driver mutations in melanoma: lessons learned from bench-to-bedside studies. *Curr. Oncol. Rep.*, **14**, 449–457.
21. Kalari,S. and Pfeifer,G.P. (2010) Identification of driver and passenger DNA methylation in cancer by epigenomic analysis. *Adv. Genet.*, **70**, 277–308.
22. Cancer Genome Atlas Network. (2012) Comprehensive molecular portraits of human breast tumours. *Nature*, **490**, 61–70.
23. Cancer Genome Atlas Research Network, Kandoth,C., Schultz,N., Cherniack,A.D., Akbani,R., Liu,Y., Shen,H., Robertson,A.G.,

Pashtan,I., Shen,R. *et al.* (2013) Integrated genomic characterization of endometrial carcinoma. *Nature*, **497**, 67–73.
24. Cancer Genome Atlas Research Network. (2008) Comprehensive genomic characterization defines human glioblastoma genes and core pathways. *Nature*, **455**, 1061–1068.
25. Weller,M., Stupp,R., Reifenberger,G., Brandes,A.A., van den Bent,M.J., Wick,W. and Hegi,M.E. (2010) MGMT promoter methylation in malignant gliomas: ready for personalized medicine? *Nat. Rev. Neurol.*, **6**, 39–51.
26. Walz,S., Lorenzin,F., Morton,J., Wiese,K.E., von Eyss,B., Herold,S., Rycak,L., Dumay-Odelot,H., Karim,S., Bartkuhn,M. *et al.* (2014) Activation and repression by oncogenic MYC shape tumour-specific gene expression profiles. *Nature*, **511**, 483–487.
27. Guo,H., Ingolia,N.T., Weissman,J.S. and Bartel,D.P. (2010) Mammalian microRNAs predominantly act to decrease target mRNA levels. *Nature*, **466**, 835–840.
28. Li,W., Zhang,S., Liu,C.C. and Zhou,X.J. (2012) Identifying multi-layer gene regulatory modules from multi-dimensional genomic data. *Bioinformatics*, **28**, 2458–2466.
29. Xiao,Y., Ping,Y., Fan,H., Xu,C., Guan,J., Zhao,H., Li,Y., Lv,Y., Jin,Y., Wang,L. *et al.* (2013) Identifying dysfunctional miRNA-mRNA regulatory modules by inverse activation, cofunction, and high interconnection of target genes: a case study of glioblastoma. *Neuro-oncology*, **15**, 818–828.
30. Xiao,Y., Guan,J., Ping,Y., Xu,C., Huang,T., Zhao,H., Fan,H., Li,Y., Lv,Y., Zhao,T. *et al.* (2012) Prioritizing cancer-related key miRNA-target interactions by integrative genomics. *Nucleic Acids Res.*, **40**, 7653–7665.
31. Kristensen,V.N., Lingjaerde,O.C., Russnes,H.G., Vollan,H.K., Frigessi,A. and Borresen-Dale,A.L. (2014) Principles and methods of integrative genomic analyses in cancer. *Nat. Rev. Cancer*, **14**, 299–313.
32. de la Fuente,A. (2010) From 'differential expression' to 'differential networking'—identification of dysfunctional regulatory networks in diseases. *Trends Genet.*, **26**, 326–333.
33. Garnis,C., Buys,T.P. and Lam,W.L. (2004) Genetic alteration and gene expression modulation during cancer progression. *Mol. Cancer*, **3**, 9.
34. You,J.S. and Jones,P.A. (2012) Cancer genetics and epigenetics: two sides of the same coin? *Cancer Cell*, **22**, 9–20.
35. Jornsten,R., Abenius,T., Kling,T., Schmidt,L., Johansson,E., Nordling,T.E., Nordlander,B., Sander,C., Gennemark,P., Funa,K. *et al.* (2011) Network modeling of the transcriptional effects of copy number aberrations in glioblastoma. *Mol. Syst. Biol.*, **7**, 486.
36. Setty,M., Helmy,K., Khan,A.A., Silber,J., Arvey,A., Neezen,F., Agius,P., Huse,J.T., Holland,E.C. and Leslie,C.S. (2012) Inferring transcriptional and microRNA-mediated regulatory programs in glioblastoma. *Mol. Syst. Biol.*, **8**, 605.
37. Xiao,F., Zuo,Z., Cai,G., Kang,S., Gao,X. and Li,T. (2009) miRecords: an integrated resource for microRNA-target interactions. *Nucleic Acids Res.*, **37**, D105–D110.
38. Wingender,E., Chen,X., Hehl,R., Karas,H., Liebich,I., Matys,V., Meinhardt,T., Pruss,M., Reuter,I. and Schacherer,F. (2000) TRANSFAC: an integrated system for gene expression regulation. *Nucleic Acids Res.*, **28**, 316–319.
39. Keshava Prasad,T.S., Goel,R., Kandasamy,K., Keerthikumar,S., Kumar,S., Mathivanan,S., Telikicherla,D., Raju,R., Shafreen,B., Venugopal,A. *et al.* (2009) Human Protein Reference Database–2009 update. *Nucleic Acids Res.*, **37**, D767–D772.
40. Bjornstad,A., Westad,F. and Martens,H. (2004) Analysis of genetic marker-phenotype relationships by jack-knifed partial least squares regression (PLSR). *Hereditas*, **141**, 149–165.
41. Verhaak,R.G., Hoadley,K.A., Purdom,E., Wang,V., Qi,Y., Wilkerson,M.D., Miller,C.R., Ding,L., Golub,T., Mesirov,J.P. *et al.* (2010) Integrated genomic analysis identifies clinically relevant subtypes of glioblastoma characterized by abnormalities in PDGFRA, IDH1, EGFR, and NF1. *Cancer Cell*, **17**, 98–110.
42. Huse,J.T., Holland,E. and DeAngelis,L.M. (2013) Glioblastoma: molecular analysis and clinical implications. *Annu. Rev. Med.*, **64**, 59–70.
43. Ding,H., Shannon,P., Lau,N., Wu,X., Roncari,L., Baldwin,R.L., Takebayashi,H., Nagy,A., Gutmann,D.H. and Guha,A. (2003) Oligodendrogliomas result from the expression of an activated mutant epidermal growth factor receptor in a RAS transgenic mouse astrocytoma model. *Cancer Res.*, **63**, 1106–1113.
44. Di Fiore,R., D'Anneo,A., Tesoriere,G. and Vento,R. (2013) RB1 in cancer: different mechanisms of RB1 inactivation and alterations of pRb pathway in tumorigenesis. *J. Cell. Physiol.*, **228**, 1676–1687.
45. Goldhoff,P., Clarke,J., Smirnov,I., Berger,M.S., Prados,M.D., James,C.D., Perry,A. and Phillips,J.J. (2012) Clinical stratification of glioblastoma based on alterations in retinoblastoma tumor suppressor protein (RB1) and association with the proneural subtype. *J. Neuropathol. Exp. Neurol.*, **71**, 83–89.
46. Molenaar,J.J., Ebus,M.E., Koster,J., van Sluis,P., van Noesel,C.J., Versteeg,R. and Caron,H.N. (2008) Cyclin D1 and CDK4 activity contribute to the undifferentiated phenotype in neuroblastoma. *Cancer Res.*, **68**, 2599–2609.
47. Wang,L., Hurley,D.G., Watkins,W., Araki,H., Tamada,Y., Muthukaruppan,A., Ranjard,L., Derkac,E., Imoto,S., Miyano,S. *et al.* (2012) Cell cycle gene networks are associated with melanoma prognosis. *PLoS One*, **7**, e34247.
48. Hamosh,A., Scott,A.F., Amberger,J., Valle,D. and McKusick,V.A. (2000) Online Mendelian Inheritance in Man (OMIM). *Hum. Mutat.*, **15**, 57–61.
49. Futreal,P.A., Coin,L., Marshall,M., Down,T., Hubbard,T., Wooster,R., Rahman,N. and Stratton,M.R. (2004) A census of human cancer genes. *Nat. Rev. Cancer*, **4**, 177–183.
50. Becker,K.G., Barnes,K.C., Bright,T.J. and Wang,S.A. (2004) The genetic association database. *Nat. Genet.*, **36**, 431–432.
51. Thrane,S., Lykkesfeldt,A.E., Larsen,M.S., Sorensen,B.S. and Yde,C.W. (2013) Estrogen receptor alpha is the major driving factor for growth in tamoxifen-resistant breast cancer and supported by HER/ERK signaling. *Breast Cancer Res. Treat.*, **139**, 71–80.
52. Rahman,M.T., Nakayama,K., Rahman,M., Ishikawa,M., Katagiri,H., Katagiri,A., Ishibashi,T., Sato,E., Iida,K., Ishikawa,N. *et al.* (2013) ESR1 gene amplification in endometrial carcinomas: a clinicopathological analysis. *Anticancer Res.*, **33**, 3775–3781.
53. Uhlmann,K., Rohde,K., Zeller,C., Szymas,J., Vogel,S., Marczinek,K., Thiel,G., Nurnberg,P. and Laird,P.W. (2003) Distinct methylation profiles of glioma subtypes. *Int. J. Cancer*, **106**, 52–59.
54. Han,W., Carpenter,R.L., Cao,X. and Lo,H.W. (2013) STAT1 gene expression is enhanced by nuclear EGFR and HER2 via cooperation With STAT3. *Mol. Carcinog.*, **52**, 959–969.
55. Camicia,R., Bachmann,S.B., Winkler,H.C., Beer,M., Tinguely,M., Haralambieva,E. and Hassa,P.O. (2013) BAL1/ARTD9 represses the anti-proliferative and pro-apoptotic IFNgamma-STAT1-IRF1-p53 axis in diffuse large B-cell lymphoma. *J. Cell Sci.*, **126**, 1969–1980.
56. Li,J., Bennett,K., Stukalov,A., Fang,B., Zhang,G., Yoshida,T., Okamoto,I., Kim,J.Y., Song,L., Bai,Y. *et al.* (2013) Perturbation of the mutated EGFR interactome identifies vulnerabilities and resistance mechanisms. *Mol. Syst. Biol.*, **9**, 705.
57. Jiang,X., Xing,H., Kim,T.M., Jung,Y., Huang,W., Yang,H.W., Song,S., Park,P.J., Carroll,R.S. and Johnson,M.D. (2012) Numb regulates glioma stem cell fate and growth by altering epidermal growth factor receptor and Skp1-Cullin-F-box ubiquitin ligase activity. *Stem Cells*, **30**, 1313–1326.
58. Feng,H., Hu,B., Vuori,K., Sarkaria,J.N., Furnari,F.B., Cavenee,W.K. and Cheng,S.Y. (2013) EGFRvIII stimulates glioma growth and invasion through PKA-dependent serine phosphorylation of Dock180. *Oncogene*, **33**, 2504–2512.
59. Lockwood,W.W., Chari,R., Coe,B.P., Girard,L., Macaulay,C., Lam,S., Gazdar,A.F., Minna,J.D. and Lam,W.L. (2008) DNA amplification is a ubiquitous mechanism of oncogene activation in lung and other cancers. *Oncogene*, **27**, 4615–4624.
60. Olson,J.M. and Hallahan,A.R. (2004) p38 MAP kinase: a convergence point in cancer therapy. *Trends Mol. Med.*, **10**, 125–129.
61. Carduner,L., Picot,C.R., Leroy-Dudal,J., Blay,L., Kellouche,S. and Carreiras,F. (2013) Cell cycle arrest or survival signaling through alphav integrins, activation of PKC and ERK1/2 lead to anoikis resistance of ovarian cancer spheroids. *Exp. Cell Res*, **320**, 329–342.
62. Leelahavanichkul,K., Amornphimoltham,P., Molinolo,A.A., Basile,J.R., Koontongkaew,S. and Gutkind,J.S. (2013) A role for p38 MAPK in head and neck cancer cell growth and tumor-induced angiogenesis and lymphangiogenesis. *Mol. Oncol*, **8**, 105–108.
63. Donnelly,S.M., Paplomata,E., Peake,B.M., Sanabria,E., Chen,Z. and Nahta,R. (2013) P38 MAPK Contributes to Resistance and Invasiveness of HER2-Overexpressing Breast Cancer. *Curr. Med. Chem*, **21**, 501–510.

64. de la Cruz-Morcillo,M.A., Garcia-Cano,J., Arias-Gonzalez,L., Garcia-Gil,E., Artacho-Cordon,F., Rios-Arrabal,S., Valero,M.L., Cimas,F.J., Serrano-Oviedo,L., Villas,M.V. *et al.* (2013) Abrogation of the p38 MAPK alpha signaling pathway does not promote radioresistance but its activity is required for 5-Fluorouracil-associated radiosensitivity. *Cancer Lett.*, **335**, 66–74.
65. Hirose,Y., Katayama,M., Stokoe,D., Haas-Kogan,D.A., Berger,M.S. and Pieper,R.O. (2003) The p38 mitogen-activated protein kinase pathway links the DNA mismatch repair system to the G2 checkpoint and to resistance to chemotherapeutic DNA-methylating agents. *Mol. Cell. Biol.*, **23**, 8306–8315.
66. Ghosh,S., Kumar,A., Tripathi,R.P. and Chandna,S. (2013) Connexin-43 regulates p38-mediated cell migration and invasion induced selectively in tumour cells by low doses of gamma-radiation in an ERK-1/2-independent manner. *Carcinogenesis*, **35**, 383–395.
67. Ohgaki,H. and Kleihues,P. (2007) Genetic pathways to primary and secondary glioblastoma. *Am. J. Pathol.*, **170**, 1445–1453.