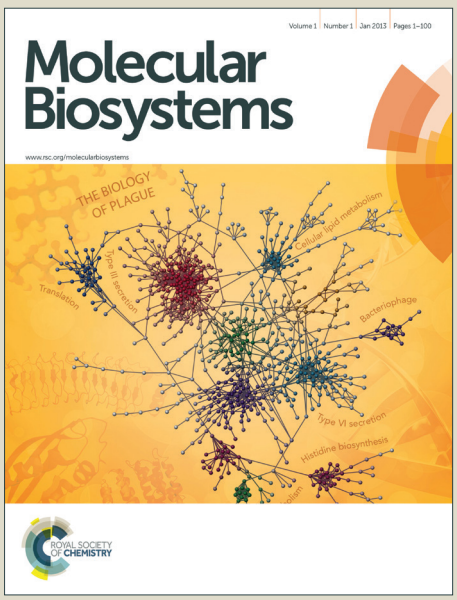
Molecular BioSystems

Accepted Manuscript



ROYAL SOCIETY OF CHEMISTRY

www.rsc.org/molecularbiosystems

Cite this: DOI: 10.1039/c0xx00000x

www.rsc.org/xxxxxx

ARTICLE TYPE

IndividualizedPath: identifying genetic alterations contributing to the dysfunctional pathways in glioblastoma individuals

Yanyan Ping, Hongyi Zhang, Yulan Deng, Li Wang, Hongying Zhao, Lin Pang, Huihui Fan, Chaohan Xu, Feng Li, Yong Zhang, Yonghui Gong, Yun Xiao*, Xia Li*

Received (in XXX, XXX) Xth XXXXXXXXX 20XX, Accepted Xth XXXXXXXXX 20XX
DOI: 10.1039/b000000x

Due to the extensive complexity and high genetic heterogeneity of genetic alterations in cancer, comprehensively depicting the molecular mechanisms of cancer remains difficult. Characterizing personalized pathogenesis in cancer individuals can help to reveal new details of the complex mechanisms. In this study, we proposed an integrative method called IndividualizedPath to identify genetic alterations and their downstream risk pathways from the perspective of individuals through combining DNA copy number, gene expression data and topological structures of biological pathways. By applying the method to TCGA glioblastoma (GBM) samples, we identified 394 gene-pathway pairs in 252 GBM individuals. We found that genes with copy number alterations showed high heterogeneity across GBM individuals, whereas they affected relatively consistent biological pathways. A global landscape of gene-pathway pairs showed that *EGFR* linked with multiple cancer-related biological pathways confers the highest risk of GBM. GBM individuals with *MET*-pathway pairs showed significantly shorter survival times than those with only MET amplification. Importantly, we found that the same risk pathways were affected by different genes in distinct groups of GBM individuals with a significant pattern of mutual exclusivity. Similarly, GBM subtype analysis revealed some subtype-specific gene-pathway pairs. In addition, we found that some rare copy number alterations had a large effect on contribution to numerous cancer-related pathways. In summary, our method offers the possibility to identify personalized cancer mechanisms, which can be applied to other types of cancer through the web server (http://bioinfo.hrbmu.edu.cn/IndividualizedPath/).

Introduction

Glioblastoma multiforme (GBM) is a common primary brain tumor in adults, with a median survival rate of 12–15 months[1]. A large number of studies have demonstrated that somatic genetic alterations, such as copy number alterations (CNAs), were the fundamental events driving the initiation and progression of cancer[2-4]. However, comprehensive genomic characterization of cancer genomes reveals the extensive complexity[5-9] and highly genetic heterogeneity in human cancer[10, 11], posing a challenge in identifying the causal genetic alterations conferring cancer initiation and progression.

With the accumulation of whole-genome measurements of cancer genomes[5, 12, 13], many studies were dedicated to developing computational methods to discover the casual genetic alterations in cancerogenesis[14-18]. Some computational methods were proposed to detect driver genetic alterations just based on the alteration frequencies of genes in cancer populations[17, 18]. For example, GISTIC was developed to predict the genomic regions harboring driver genes by calculating the significance of gene amplification or deletion based on copy number variations across cancer samples[18]. Recently, several computation methods by integrating additional information, such as gene expression, were proposed[19-22]. Based on the postulation that the "genomic footprint" in gene expression reflects the functional impact of driver alterations, Akavia et al. [19] detected the driver genes whose CNAs were frequent, and influenced the expression of a group of genes by regulating their own expression. DriverNet was proposed to identify the minimum number of driver alterations that can account for most transcriptional changes across cancer samples[22]. These integrative genome analyses identified some well-known cancer genes based on the cancer population[23]. However, due to the highly genetic heterogeneity, the similar phenotypes of cancer individuals may be driven by the different combinations of genetic alterations[24], which imply that the population-based methods cannot capture the pathogenesis of cancer individuals. Thus, exploring the driver genetic alterations and their downstream effects at the individual level will provide new insights into the mechanisms of cancer.

Gene expression can characterize the activity of biological pathways that underlie the cancer phenotype[25]. Differential expression of genes in key pathways can thus reflect the abnormal states of the functional mechanism[26]. Previous studies have reported that CNAs have direct roles on the global deregulation of gene expression[27, 28]. The CNAs of key genes can

Cite this: DOI: 10.1039/c0xx00000x

www.rsc.org/xxxxxx

ARTICLE TYPE

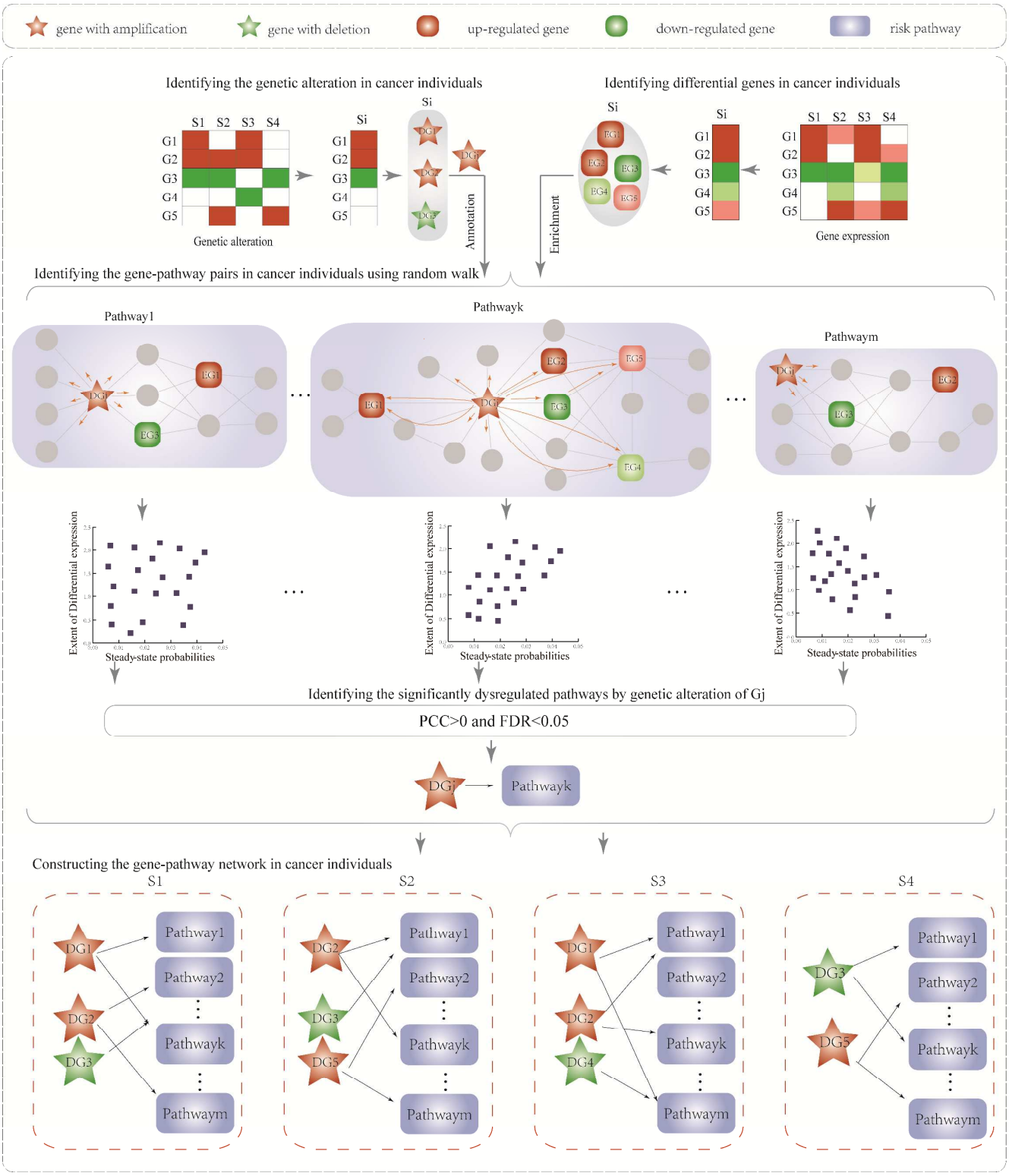


Figure 1. The workflow of the integrated method IndividualizedPath for identifying cancer-related genes and their affecting risk pathways in cancer individuals.

Cite this: DOI: 10.1039/c0xx00000x

www.rsc.org/xxxxxx

ARTICLE TYPE

cause the abnormal activity of pathways by disturbing expression of genes in the pathways[29, 30]. For example, the amplification of *N-myc* can inhibit the mRNA expression of *CDC42* directly or indirectly through nm23-mediated inhibition and, in turn, prevent the *CDC42*-mediated differentiation in neuroblastoma[31]. *MET* amplification can maintain the phosphorylation of *ERBB3* and *Akt* in the presence of gefitinib, and further activate PI3K/Akt signaling pathway[32]. Significant expression differences were observed between samples with distinct CNA patterns in ewing's sarcoma[33]. Thus, identifying the CNAs that can explain the abnormal expression of genes in the pathways will further enhance the understanding of roles of CNAs underlying cancer mechanisms.

Based on the hypothesis that genetic alterations contribute to the carcinogenesis of individuals by dysregulating gene expression in some key pathways, we developed a novel computational method IndividualizedPath to identify the genetic alterations and their affecting risk pathways in disease individuals. IndividualizedPath integrated DNA copy number and gene expression data, as well as the topological information of biological pathways. Applying our method to GBM individuals, we identified casual genes and their associated pathways (gene-pathway pairs) for each GBM individual. We found that genetic alterations with inconsistency linked with relatively consistent pathways across GBM samples and that some well-known GBM-associated genes (e.g. *EGFR*) frequently contributed to multiple risk pathways in GBM individuals. We also found that different genetic alterations mutually exclusively affected the same pathways. In addition, we further explored the subtype specificity of gene-pathway pairs, and elucidated the important roles of some rare genetic alterations in GBM development.

Material and methods

Datasets

DNA copy number and gene expression data

We obtained DNA copy number data and gene expression profile of GBM individuals from the TCGA data portal (https://tcga-data.nci.nih.gov/tcga/). The DNA copy number data (level 3, Affymetrix SNP array 6.0) segmented by the circular binary segmentation method[34] contained 540 GBM samples. The gene expression profile referring to 12042 genes (HU_HG_U133A) involved 528 GBM and 10 normal samples. The 476 common GBM samples in these two datasets (together with 10 normal samples) were used for subsequent analyses.

KEGG pathways

The 300 pathways containing 150 regulatory pathways and 150 metabolic pathways were downloaded from the Kyoto Encyclopedia of Genes and Genomes (KEGG, http://www.kegg.jp/kegg/)[35]. We got the corresponding undirected graphs of metabolic pathways and regulatory pathways using getMetabolicGEGEUEMGraph and getNonMetabolicGEGEUEMGraph in the R package SubpathwayMiner (version 3.0) [36].

Methods

We hypothesized that starting from the copy number alterations (CNAs) of some driver genes, the dysfunctional information can be propagated to downstream genes and then influence their expression. That is, cancer-related genes with CNAs contributed to the carcinogenesis in individuals by dysregulating gene expression in some members of some key pathways. Thus, for a specific cancer individual, if a gene with CNA can explain most expression changes of some key pathways, the gene and the linked pathways can be identified as important molecular events contributing to the carcinogenesis of this individual. Based on this hypothesis, we developed an integrated method called IndividualizedPath to identify cancer-related genes and their affecting risk pathways in cancer individuals by combining DNA copy number and gene expression data as well as the topological information of biological pathways (Figure 1).

Identifying the genes with CNAs in GBM individuals

We identified the genes with CNAs by applying GISTIC (version 2)[37] to the segmentation data of DNA copy number in 476 GBM samples using default parameters. For a GBM individual S_i, the copy number calls calculated by GISTIC was used to determine CNA events including homozygous deletion (-2), heterozygous deletion (-1), diploid (0), gain (1), and high-level amplification (2). Genes with high-level amplification or homozygous deletions were identified as genes with CNAs (labeled as DG) for subsequent analyses.

Identifying differentially expressed genes in GBM individuals

For a GBM individual S_i, we identified differentially expressed genes by comparing gene expression levels between S_i and 10 normal samples. For a gene G_j in the expression profile, we calculated a Z-score as the normalized gene expression value of G_j[20, 38]:

$$Z_{ij} = \frac{\left| E_{ij} - \mu_j \right|}{\sigma_j}$$

where E_{ij} is the expression level of G_j in S_i, and μ_j and σ_j are the average expression level and standard deviation of G_j in normal individuals, respectively. If Z_{ij} is greater than or equal to 4, at which the significance of differential expression is 6.3e-05 using a Z-test, G_j is considered a differentially expressed gene (labeled as *EG*) in S_i. The difference $|Z_{ij}$ -4| represents the extent of the differential expression of G_j in S_i.

Identifying the risk pathways in GBM individuals

For a GBM individual S_i, the risk pathways were defined as the ones in which most of the genes were likely to be differentially expressed. We identified the risk pathways that were significantly enriched by the differentially expressed genes in S_i using SubpathwayMiner (P<0.01).

Extracting candidate gene-pathway pairs in GBM individuals

Cite this: DOI: 10.1039/c0xx00000x

www.rsc.org/xxxxxx

ARTICLE TYPE

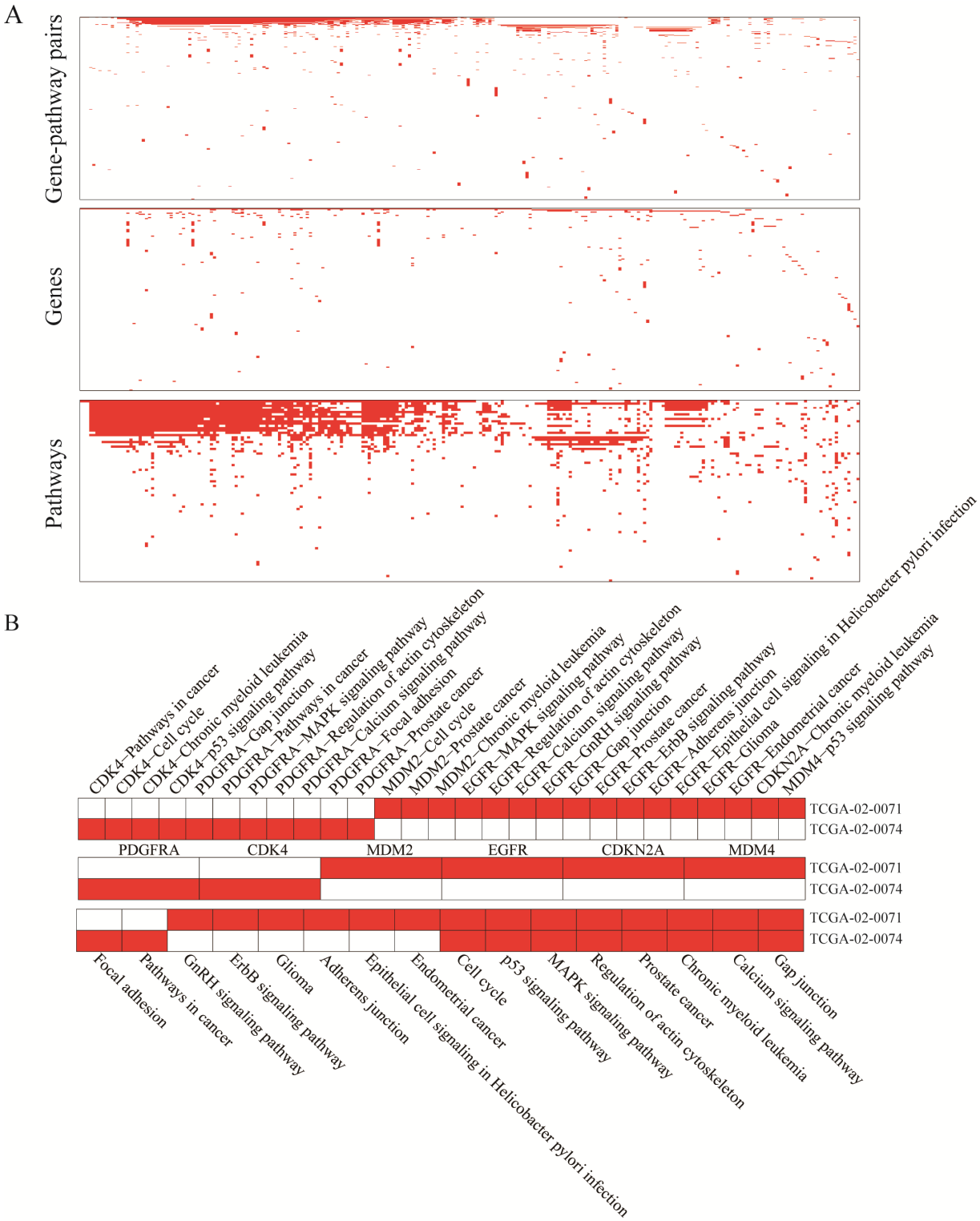


Figure 2. The gene-pathway pairs across GBM individuals. (A) The heatmaps of gene-pathway pairs (top), cancer-related genes (middle) and risk pathways (bottom) in 252 GBM samples. Rows represent gene-pathway pairs, genes and pathways, respectively, and columns are samples. (B) The gene-pathway pairs, cancer-related genes and risk pathways in two GBM individuals including TCGA-02-0071 and TCGA-02-0074.

Cite this: DOI: 10.1039/c0xx00000x

www.rsc.org/xxxxxx

ARTICLE TYPE

For each gene with CNA DG_j in a given GBM individual S_i, we obtained the annotated pathways of DG_j based on KEGG. Then, the common pathways between the annotated pathways of DG_j and risk pathways in S_i were determined. These common pathways were considered as the candidate pathways affected by DG_j, and candidate gene-pathway pairs formed by DG_j and the candidate pathways were identified.

Identifying gene-pathway pairs in GBM individuals

The problem of information flow within pathways can be addressed by utilizing the 'random walk with restart' (RWR). RWR has been used to characterize functional similarity between genes in the network for prioritizing the disease-associated genes by integrating the global topology of the functional network and local functional links of disease genes[39, 40].

For each candidate gene-pathway pair (DG_j and $path_k$) in an individual S_i, we utilized RWR to calculate the impacts of DG_j on $path_k$, which were used to explain the differential expression of the genes in $path_k$. We extracted the connected sub-network containing DG_j in pathway $path_k$. In this sub-network, the gene DG_j was considered as the seed node. The dysfunctional information as a random walker started from DG_j, and the flow of information randomly transited from current nodes to all neighbor nodes with equal probability. Meanwhile, the random walk can restart from the seed node with the probability of r in each step of information flowing. The RWR [39] can be described as follows:

$$P_{t+1} = (1-r)\times \mathrm{W}\times P_t + r\times P_0$$

where P_0 is the vector of initial probabilities of genes in the sub-network, in which the probability of seed node was 1 and others 0; P_t and P_{t+1} are the probabilities of random walker in nodes at the t_{th} and $(t+1)_{th}$ steps, respectively; W is the transfer matrix in which the columns are normalized according to gene degrees; r represents the restart probability. The restart probability r was set to 0.3. If the maximum difference between P_{t+1} and P_t is less than 10^{-8}, the random walk reaches the steady-state. The probabilities of genes in the sub-network under the steady state were obtained, which characterized the influence of DG_j on genes in the risk pathway (Supplementary Figure S1).

Finally, we extracted the probabilities of differentially expressed genes in the sub-network and calculated the Pearson correlation coefficient (PCC_{jk}) between the probabilities and the extents of differential expression. The gene DG_j was considered contributing to the risk pathway $path_k$ in the individual S_i if PCC_{jk} was positive and significant (Pearson's correlation coefficient test, FDR<0.05), and then the gene DG_j and the pathway $path_k$ formed a gene-pathway pair in the individual S_i.

Results

The gene-pathway pairs across GBM individuals

We proposed a novel method IndividualizedPath to identify cancer-related genes and their downstream risk pathways in individuals through integrating the DNA copy number data, gene expression and KEGG pathways. The method was applied to the 476 GBM samples, which had an average of about 113 genes with CNAs and 88 risk pathways. Finally, among 6625 candidate gene-pathway pairs, 394 gene-pathway pairs involving 167 genes and 79 risk pathways were identified. Of these pairs, 67.5% occurred in only one GBM individual. We found that 252 GBM individuals presented at least one gene-pathway pair. The number of gene-pathway pairs in different individuals ranged from 1 to 34.

Comparison with the candidate gene-pathway pairs showed that a number of candidate gene-pathway pairs (with an average of about 57) were eliminated in cancer individuals (Supplementary Figure S2A). The ratio of cancer genes recorded in the Cancer Gene census (CGC) database in the identified gene-pathway pairs was elevated to 19.2% (5.7% in the candidate gene-pathway pairs). In the HPRD protein interaction network[41], the mean degree of genes in the identified gene-pathway pairs across GBM individuals were significantly higher than those eliminated (P<0.01, t-test, Supplementary Figure S2B). As an example, among 23 candidate pathways of AKT, 20 were identified. Out of the three eliminated candidate pathways, two pathways were not significantly enriched by differential genes using the AKT knockdown data (GSE31534) (Supplementary Figure S2C). These results suggest that our method can improve the accuracy of dysfunctional gene-pathway discovery.

We observed obvious inconsistency of gene-pathway pairs across GBM individuals (Figure 2A, top). Notably, we found that the distribution of cancer-related genes in gene-pathway pairs across 252 GBM individuals showed inconsistency (Figure 2A, middle), while the distribution of pathways showed relatively higher consistency (P=2.026e-10, Kolmogorov-Smirnov Test, Figure 2A, bottom, and Supplementary Figure S3). These results suggested that different genes in different GBM individuals can disturb the same pathways. For example, in one GBM patient, 16 gene-pathway pairs involving 4 genes (including *MDM4*, *EGFR*, *CDKN2A,* and *MDM2*) and 14 pathways (Figure 2B) were found, while in another patient, we found 2 genes (including *PDGFRA,* and *CDK4*) that contributed to the dysfunction of 10 pathways. There were 8 common deregulated pathways but without common genes.

The landscape of gene-pathway from individual contribution

To globally characterize the landscape of gene-pathway, we constructed a weighted network by integrating the gene-pathway pairs from GBM individuals (Figure 3A). In this network, the weights of nodes (genes and pathways) and edges (gene-pathway pairs) were calculated as their frequencies in GBM individuals. For each gene, we calculated the sum of weights of its connected edges as a metric (termed as contribution degree) to characterize the extent to which this gene contributed to GBM tumorigenesis.

Cite this: DOI: 10.1039/c0xx00000x

www.rsc.org/xxxxxx

ARTICLE TYPE

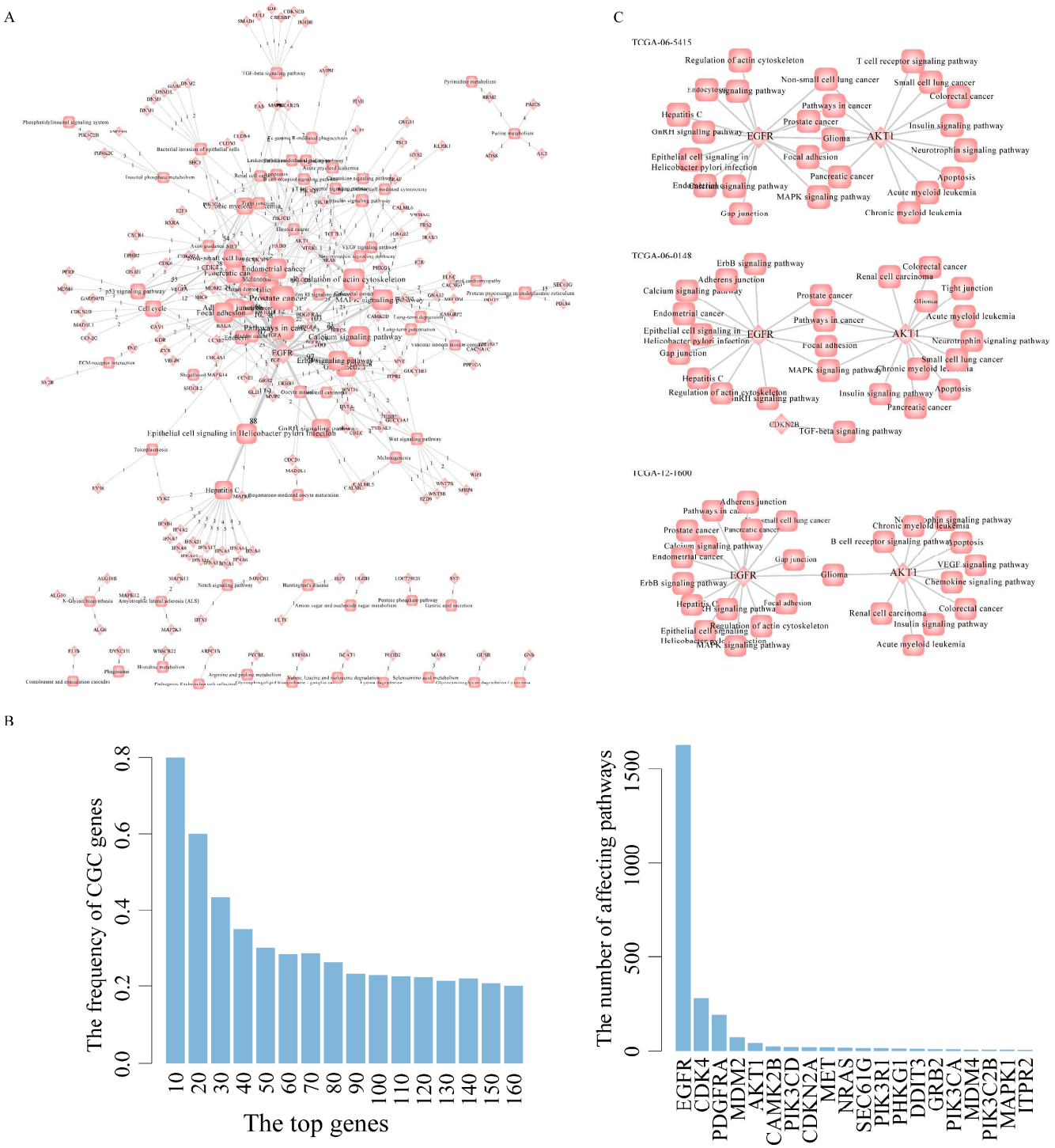


Figure 3. The landscape of gene-pathway from individual contribution. (A) The weighted gene-pathway network of GBM. The diamond and round rectangle nodes are genes and pathways, respectively. The node sizes are proportional to the frequency of gene or pathway. The edges represent the gene-pathway pairs, the widths of which are proportional to the frequencies of gene-pathway pairs. (B) Left: The proportion of CGC genes in the top ranked genes according to the contribution degree of genes; Right: The top 20 genes with high contribution degree. (C) The gene-pathway networks for three GBM individuals with AKT-pathway pairs (TCGA-06-5415, TCGA-06-0148 and TCGA-12-1600).

Cite this: DOI: 10.1039/c0xx00000x

www.rsc.org/xxxxxx

ARTICLE TYPE

We ranked all genes by contribution degrees in decreasing order and compared the top-ranked genes with cancer genes recorded in the Cancer Gene census (CGC) database. The 60% of the top 20 genes were known cancer genes (Figure 3B). In the top 5 genes, *EGFR*, *CDK4*, *PDGFRA,* and *MDM2* have been reported to be associated with GBM, and their CNAs affected the pathway of Glioma.

In addition, some genes that altered with low frequency but had high contribution degree were also identified, such as *AKT1*, *PIK3CD*, and *NRAS*, suggesting their crucial roles in some GBM individuals. For example, the amplification of *AKT1* was found in only 5 of 476 GBM individuals but contributed to 15, 15, and 11 cancer-associated pathways such as Glioma, apoptosis and VEGF signalling pathway in 3 GBM individuals, respectively (Figure 3C). Interestingly, amplification of EGFR was also found in these 3 GBM individuals and contributed to the dysfunction of multiple pathways, most of which were different from those of *AKT1*; thus, combined treatment based on *AKT1* and *EGFR* may be effective for personalized therapy.

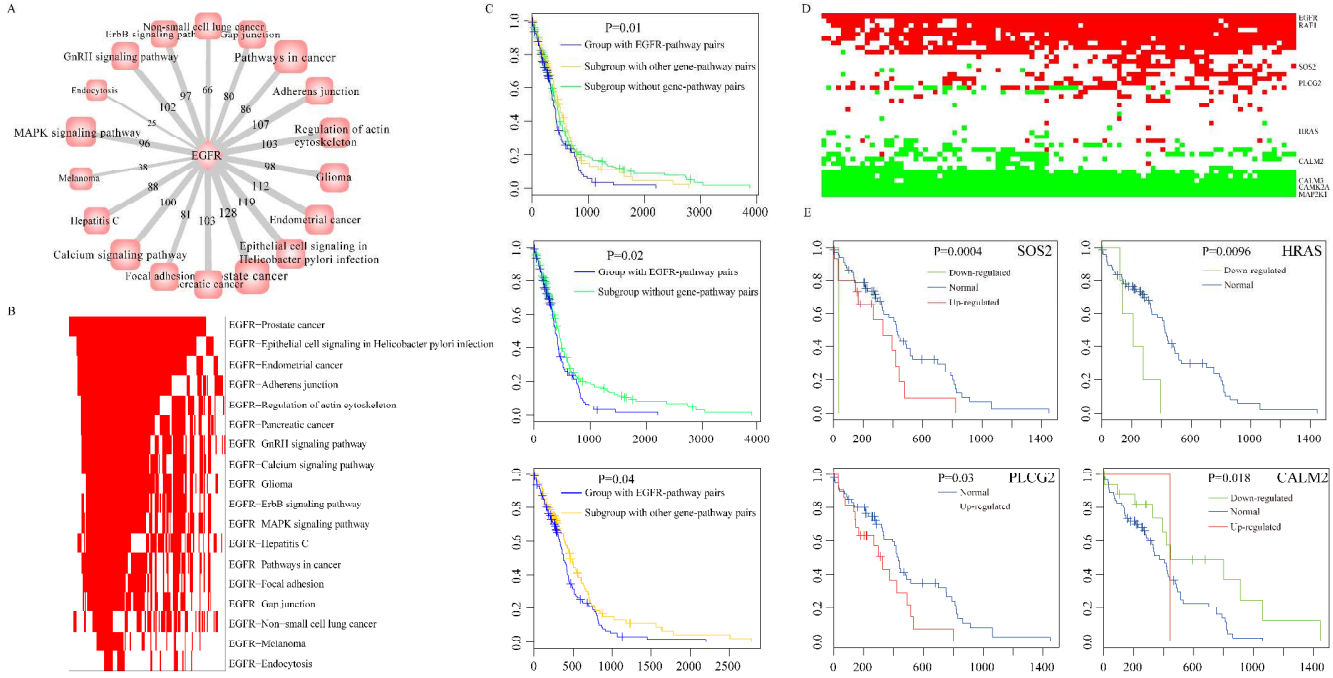


Figure 4. The EGFR-pathway pairs in GBM individuals. (A) The network of EGFR-pathway pairs. (B) The heatmap of EGFR-pathway pairs in 146 GBM samples. (C) Kaplan–Meier estimates of overall survival for three GBM groups (I, II and III). (D) The heatmap of differential expression patterns of genes in the Glioma pathway in 98 GBM samples with EGFR-Glioma pairs. (E) Kaplan–Meier estimates of overall survival for SOS2, HRAS, PLCG2 and CAML2.

EGFR and MET drive the cancer pathways in GBM individuals

Among the genes contributing to dysregulated pathways, *EGFR* was frequently altered and had the highest contribution degree in GBM individuals. *EGFR* was found to connect with 18 pathways, most of which were related to cancer, such as Glioma, ErbB signalling pathway, and MAPK signalling pathway. All *EGFR*-pathway pairs showed higher frequencies in GBM individuals than other gene-pathway pairs (Figure 4A). Among the 252 GBM individuals with gene-pathway pairs, 146 GBM individuals were identified to be linked with *EGFR*-pathway pairs. Moreover, we found that *EGFR* could induce multiple cancer-related pathways in each GBM individual (Figure 4B), consistent with its versatile roles in GBM. Then, we divided the GBM individuals into three groups: group I with *EGFR*-pathway pairs, group II with other gene-pathway pairs, and group III without any gene-pathway pairs. The overall survival of the three groups showed significant difference (P=0.01, log-rank test). Group I with *EGFR*-pathway pairs had significantly shorter survival time than the other groups (P=0.02 for group III, and P=0.04 for group II, log-rank test) (Figure 4C), consistent with the previous studies[42, 43].

To further investigate how *EGFR* contributes to dysregulated pathways, we took the *EGFR*-Glioma pathway pair as an example, which was identified in 98 GBM individuals. The Glioma pathway contains two important signalling cascades: PI3K/Akt/mTOR and Ras/MEK/MAPK. *EGFR* is an upstream member of the Glioma pathway and the amplification of *EGFR* can destroy the downstream signalling cascades of PI3K/Akt/mTOR and Ras/MEK/MAPK and, in turn, promote cell survival, proliferation, and growth[44]. Expression analysis showed that many genes in the Glioma pathway were consistently up-regulated (such as *EGFR* and *RAF1*) or down-regulated (such as

MAP2K1, *CALM3*, and *CAMK2A*) across the 98 GBM individuals (Figure 4D). Nonetheless, some genes showed inconsistent expression patterns across GBM individuals, some of which were associated with the survival of GBM (Figure 4E). For example, *SOS2* and *HRAS* were two members of the signalling cascade Ras/MEK/MAPK. The overexpression of *SOS2* and underexpression of *HRAS* were significantly associated with shorter survival (P=0.0004 for *SOS2* and P=0.0096 for *HRAS*, log-rank test), which could be attributed to the activation of the signalling cascade Ras/MEK/MAPK. *PLCG2* and *CALM2* located in an alternative path were involved in the signalling cascade Ras/MEK/MAPK. The overexpression of *PLCG2* was significantly associated with shorter survival (P=0.03 for *PLCG2*, log-rank test), and the underexpression of *CAML2* was associated with better survival (P=0.018 for *CAML2*, log-rank test). These results may provide an explanation why GBM patients with *EGFR* amplification exhibited different responses to *EGFR* inhibitor[45, 46].

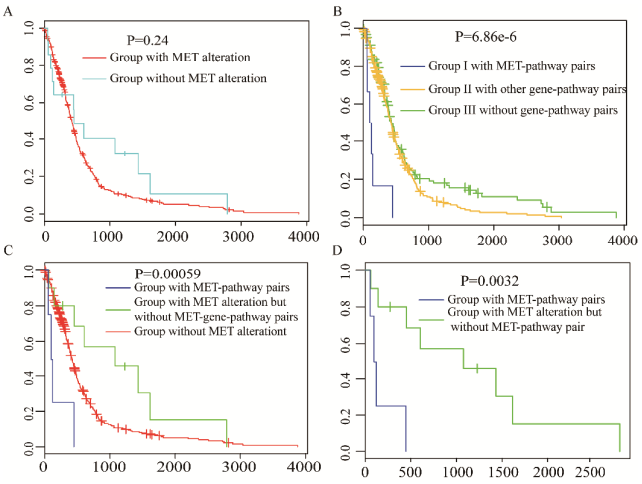


Figure 5. The role of *MET*-pathway pairs in GBM survival. (A) Kaplan–Meier estimates of overall survival for *MET* alteration. (B) Kaplan–Meier estimates of overall survival for three GBM groups (I, II and III) according to distribution of *MET*-pathway pairs and other gene-pathway pairs. (C) Kaplan–Meier estimates of overall survival according to *MET* alteration and *MET*-pathway pairs. (D) Kaplan–Meier estimates of overall survival for *MET*-pathway pairs in GBM samples with *MET* alteration.

In addition to genes with frequent alteration like *EGFR*, *MET*, a functional marker of glioblastoma stem cells[46], with low-frequency alteration (3.5%, 17/476) was ranked in the top 10 of genes according to their degree. We observed that the copy number of *MET* was not an indicator of survival in GBM (P=0.23, log-rank test, Figure 5A). Interestingly, when we divided GBM samples into three groups like *EFGF* (group I with *MET*-pathway pairs, group II with other gene-pathway pairs and group III without any gene-pathway pairs), the overall survival of the three groups showed significant difference (P=6.86e-5, log-rank test, Figure 5B). Group I with *MET*-pathway pairs had significantly shorter survival times than the other groups (P=3.87e10-5 for group III, and P=8.13e-5 for group II, log rank test). Furthermore, the GBM samples with *MET*-pathway pairs showed significantly shorter survival times than the samples with *MET* amplification but without *MET*-pathway pairs (P=0.0032, log rank test, Figure 5C and 5D). Together, these results proved that our method can be effectively used in identifying key gene-pathway pairs in cancer.

GBM pathways driven by distinct copy number alterations in GBM individuals

To explore how CNAs of genes influence the same pathways, we examined the distributions of gene-pathway pairs across GBM individuals. We found that the same pathway could be affected by different genes. Notably, these genes linked with the same pathways showed obvious mutual exclusivity in GBM individuals (Figure 6). To determine the statistical significance of mutual exclusivity, the distributions of genes across the samples with a specific risk pathway were randomly permuted 1000 times. We calculated the number of samples covered by genes in each permutation, and calculated the significance as the fraction of permutations that produced an equal number of covered samples with that observed in real data. For instance, the Glioma pathway was affected by 12 genes, including *EGFR*, *PDGFRA*, *CAMK2B*, *AKT1*, *CDK4*, *MDM2*, *NRAS*, *PIK3CA*, *TGFA*, *SHC4*, *CDKN2A*, and *PDGFA*, in 119 GBM samples. *EGFR* and *PDGFRA* were identified to contribute to the Glioma pathway in 74% of these GBM samples. The dysfunction of the Glioma pathway in the rest of the GBM samples was dependent on the other 10 genes. These 12 genes showed a significant pattern of mutual exclusivity (P<0.001). The similar mutual exclusivity phenomenon was also observed for other pathways such as MAPK signalling pathway (P<0.001), apoptosis (P<0.001), Wnt signalling pathway (P=0.003), TGF-beta signalling pathway (P=0.005) and neurotrophin signalling pathway (P<0.001), which was in line with previous studies[47]. The property of mutual exclusivity suggests complex heterogeneity in cancer.

Specific gene-pathway pairs in GBM subtypes.

To investigate whether gene-pathway pairs were specific in different GBM subtypes, we clustered the 252 GBM individuals into four subtypes—classical (92), proneural (67), mesenchymal (49) and neural (42)[6, 48]—according to the expression of signature genes identified in[5, 6]. In the classical subtype, *EGFR*-pathway pairs were found to be significantly enriched in classical GBM individuals (P<2.2e-16, OR=16.4, Fisher's Exact Test, Figure 7A), which was consistent with previous reports that the classical subtype was characterized by the amplification of *EGFR*[6]. Similarly, for the signature gene *PDGFRA* of the proneural subtype[6], our results showed that *PDGFRA* and its affected pathways were presented in 29 proneural GBM individuals (P=2.4e-16, OR=45.4, Fisher's Exact Test, Figure 7B). Notably, *CDK4*-pathway pairs were presented in 30 proneural GBM individuals (P<2.2e-16, OR=72.5, Fisher's Exact Test, Figure 7B). *PDGFRA* and *CDK4* showed a complement tendency—both of these pairs covered approximately 73.1% of proneural GBM individuals. These results indicated that *CDK4* may be another significant gene of the proneural subtype. In addition, it should be noted that the individuals in the same GBM subtype showed different gene-pathway pairs. For example, completely different sets of gene-pathway pairs were identified in two classical GBM individuals. In one individual, *CDK4* and *CCND2* influenced the dysfunction of 9 pathways such as Glioma, cell cycle, and Wnt signalling pathway. In the other individual, *PLA2G5* and *GNG12*

Cite this: DOI: 10.1039/c0xx00000x

www.rsc.org/xxxxxx

ARTICLE TYPE

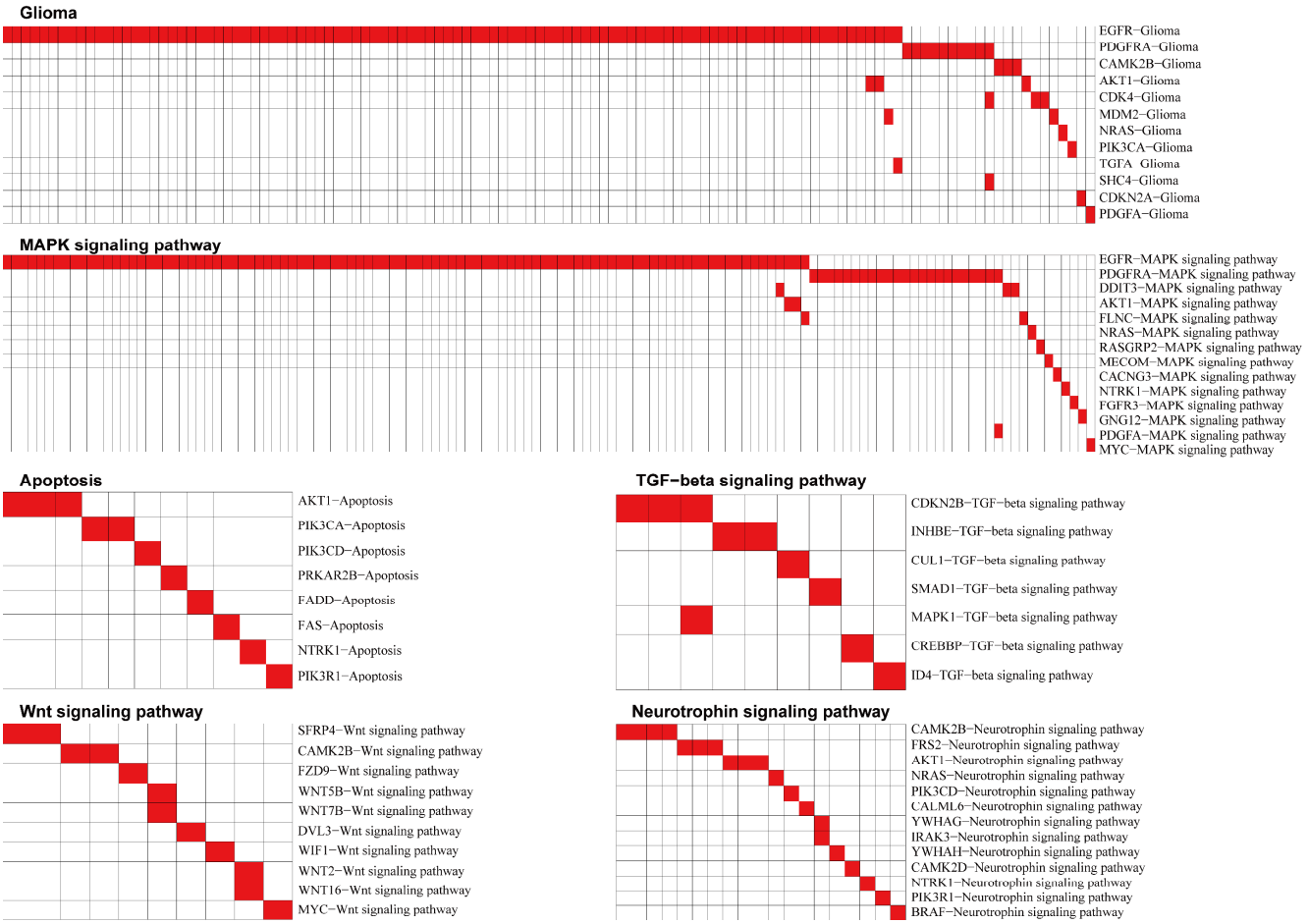


Figure 6. Distinct genes with CNAs contributed to the same pathways with mutual exclusivity.

affected 7 different pathways such as MAPK signaling pathway, VEGF signaling pathway, and chemokine signaling pathway (Supplementary Figure S4).

The functional impact of rare CNAs in GBM individuals

By analyzing the frequency of CNAs among all GBM individuals, we found that 81.4% of genes in the gene-pathway pairs had the frequency of less than 5%. These genes with rare alterations were usually overlooked by population-based methods. For example, *NRAS*, neuroblastoma RAS viral (v-ras) oncogene homolog, showed somatic alterations in only one mesenchymal GBM individual. *NRAS* was reported to be associated with multiple types of cancers such as melanomas[49] and multiple myelomas[50]. Our results showed that the alteration of *NRAS* contributed to the dysfunction of many GBM-associated pathways such as Glioma, ErbB signalling pathway, MAPK signalling pathway, neurotrophin signalling pathway, gap junction and axon guidance (Supplementary Figure S5). Notably, a total of 27 genes with CNAs were determined in the mesenchymal GBM individual, whereas there were no any other genes contributing to these pathways except *NRAS*, suggesting that *NRAS* played key roles in this GBM individual. *NOTCH1*, altered in three GBM individuals, was found to be associated with the dysfunction of Notch signalling pathway in one proneural GBM individual. Consistently, recent studies reported that *NOTCH1* was associated with glioma cell differentiation[51], and the proneural subtype of GBM showed high Notch pathway activation[52]. The inhibition of *NOTCH1*[53, 54] or Notch signalling[55, 56] can block glioblastoma cell proliferation and tumor growth. These results implied that rare genetic alterations also have functional impact on the pathogenesis of GBM individuals.

Discussion

Different cancer individuals may exhibit different combinations of genetic alterations[11] that disturb the same or similar cellular biological pathways and, in turn, lead to the same or similar phenotypes[57]. Such genetic heterogeneity, as one of the most important hallmarks of tumors[58], poses a major challenge in diagnosing cancer and designing effective therapies[59-61]. In this

Cite this: DOI: 10.1039/c0xx00000x

www.rsc.org/xxxxxx

ARTICLE TYPE

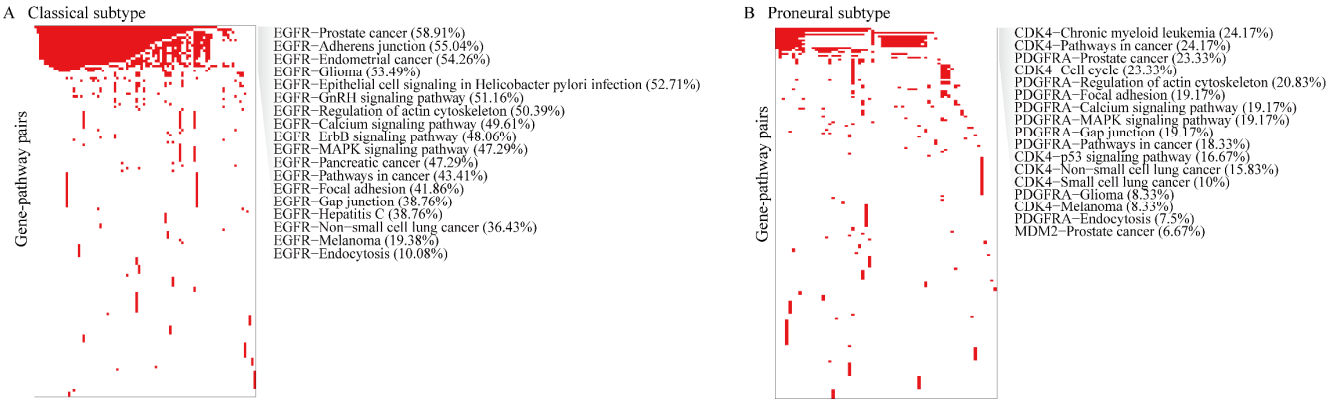


Figure 7. The gene-pathway pairs in the GBM subtypes. (A) The heatmaps of gene-pathway pairs in the classical GBM subtype. (B) The heatmaps of gene-pathway pairs in the proneural GBM subtype.

study, we developed an integrated method IndividualizedPath to identify cancer-related genetic alterations and their downstream risk pathways in cancer individuals. We applied the method to GBM and identified abnormal gene-pathway pairs for each GBM individual.

Based on the underlying assumption that genes with CNAs should account for most expression changes in a given dysfunctional pathway, our method allows us to identify potential driver genes and offers the ability to characterize their affected risk biological pathways at the individual lever. To characterize the links between genes with CNAs and downstream expression changes in risk pathways, our method not only combines DNA copy number and gene expression data but also takes into account the topological structures of KEGG pathways. Through scoring each gene in a risk pathway based on random walk started from a gene with CNAs, the method evaluates whether the CNA can capture most expression changes in the risk pathway.

More importantly, our method enables us to identify abnormal gene-pathway pairs from the individual perspective by analyzing CNAs and expression changes in cancer individuals. The results of our method not only give the global landscape of driver mechanisms across cancer populations but also provide insights into personalized cancer mechanisms. From the population perspective, the global landscape of gene-pathway showed that some GBM-associated genes contributed to multiple pathways in the majority of GBM individuals, while most present in a few samples. The dysfunction of the same pathways in different samples was affected by distinct genes showing significant mutual exclusivity, suggesting that different groups of cancer individuals depend on distinct genetic alterations in destroying the same pathways, consistent with high genetic heterogeneity of GBM[62]. This also highlights the importance of identifying personalized mechanisms of cancer. From the individual perspective, our results can partially explain the heterogeneity across cancer individuals. For example, *MET* has been reported to be associated with GBM stem cell, while the amplification of *MET* was not associated with GBM survival. We found that the samples with *MET*-associated risk pathways showed a significantly shorter survival time than the samples with only *MET* amplification. The risk pathways not only characterized the potential driver roles of *MET* in some GBM samples but also showed heterogeneity under the same copy number alteration. Also, our method can identify rare CNAs that have functional impact on the pathogenesis of GBM. For example, *NARS*, which has been reported to participate in tumorigenesis in other types of cancer, was altered with low frequency but affected various key cancer-associated pathways. This will facilitate understanding of the phenomenon that different cancer individuals exhibit remarkable differences in clinical drug response.

Furthermore, we compared the risk pathways identified by our method with those by other approaches (i.e., the pathways significantly enriched by genes by GISTIC, genes identified by both GISTIC and differential expression analysis, and genes whose CNAs significantly influenced their own expression levels). We found that the risk pathways identified by our method not only contained known cancer-related pathways identified by other methods but also contained additional important cancer-related pathways such as apoptosis, TGF-beta signalling pathway and Notch signalling pathway (P<0.01), (Supplementary Figure S6).

In addition, our approach can be applied to other types of cancer. We applied our method to 513 breast cancer and 149 colorectal adenocarcinoma samples derived from TCGA, separately. In breast cancer, a global gene-pathway network involving 166 genes and 61 pathways was constructed (Supplementary Figure S7). In colorectal adenocarcinoma, a global gene-pathway network involving 46 genes and 30 pathways was constructed (Supplementary Figure S8). Among the top-10 genes with the highest degree, some have been found to contribute to tumorigenesis, such as *PTEN*[63], *CCNE1*[64], *SMAD4*[65] and *PIK3CA*[66]. They showed extensive links with multiple cancer-associated pathways across cancer individuals. Then, we investigated gene-pathway pairs across these three cancer types (i.e., GBM, breast cancer and colorectal adenocarcinoma). For the

top-10 ranked pathways with the highest degree, we found four common pathways including Pathways in cancer, cell cycle, MAPK signalling pathway and p53 signalling pathway, all of which are crucial for tumorigenesis. Interestingly, no common genes linked with these pathways across the three cancer types were identified, implying that different types of cancer are dependent on distinct driver genes to destroy the same key pathways. Also, we developed an online webserver (available at http://bioinfo.hrbmu.edu.cn/IndividualizedPath/), where researchers can identify gene-pathway pairs in other cancer types using our method.

Notably, there were no gene-pathway pairs identified in about half of the samples. Three possible reasons for these results are: (1) A part of samples harbor no or few CNAs. We found that the samples with no gene-pathway pairs had significantly smaller numbers of genes with CNAs (Supplementary Figure S9); (2) Numerous biological pathways are yet incomplete. Among a total of 16526 genes with CNAs across GBM individuals, only 2528 were annotated in the pathways; (3) Identification of risk pathways depends on the threshold of Z-score. The higher Z-score threshold may lead to decreased numbers of differential genes and thus lower numbers of risk pathways. We used Z-score=4 as the threshold to identify risk pathways, more stringent than 2.3 used in a previous study[20]. If a more lenient threshold was used, more risk pathways may be identified, which might lead to identification of new gene-pathway pairs.

Conclusions

In summary, we proposed an integrative method to identify personalized genetic alterations and their affecting biological pathways, which will be helpful in better understanding the molecular mechanisms of cancer individuals and in promoting personalized therapy.

Author Contribution Statement

Xia Li and Yun Xiao participated in the design and coordination of the study. Yanyan Ping, Hongyi Zhang and Yulan Deng carried out the IndividualizedPath method. Yanyan Ping, Li Wang and Hongying Zhao analyzed the results. Lin Pang and Huihui Fan participated in biological annotation of the results. Yanyan Ping, Chaohan Xu, Feng Li and Yong Zhang wrote the manuscript. Yonghui Gong developed the web server for the method. All authors have read and approved the manuscript and its contents, and are aware of responsibilities connected to authorship.

Conflict of Interest

There is no competing financial interest in relation to the work.

Acknowledgments

This work was supported in part by the National High Technology Research and Development Program of China [863 Program, Grant Nos. SS2014AA021102], the National Program on Key Basic Research Project [973 Program, Grant Nos. 2014CB910504],the National Natural Science Foundation of China [Grant Nos. 91129710, 61073136, 31200997, 61170154, and 81070946], the National Science Foundation of Heilongjiang Province [Grant Nos.C201207 and H0906], Key Laboratory of Cardiovascular Medicine Research (Harbin Medical University), Ministry of Education and the Undergraduate Innovation Funds of Harbin Medical University [Grant Nos.YJSCX2012-210HLJ].

Notes and references

*Corresponding Authors: Xia Li, PhD, College of Bioinformatics Science and Technology, Harbin Medical University, Harbin 150081, China. Phone: 86-451-86615922; Fax: 86-451-86615922. Email:lixia@hrbmu.edu.cn

Yun Xiao, PhD, College of Bioinformatics Science and Technology, Harbin Medical University, Harbin 150081, China. Email: xiaoyun@ems.hrbmu.edu.cn

†Electronic Supplementary Information (ESI) available: [Supplementary Figures S1-S9]. See DOI: 10.1039/b000000x/

1. M. Nakada, D. Kita, T. Watanabe, Y. Hayashi, L. Teng, I. V. Pyko and J. Hamada, *Cancers*, 2011, 3, 3242-3278.
2. C. Lopez-Gines, R. Gil-Benso, R. Ferrer-Luna, R. Benito, E. Serna, J. Gonzalez-Darder, V. Quilis, D. Monleon, B. Celda and M. Cerda-Nicolas, *Modern pathology : an official journal of the United States and Canadian Academy of Pathology, Inc*, 2010, 23, 856-865.
3. M. C. Hollander, G. M. Blumenthal and P. A. Dennis, *Nature reviews. Cancer*, 2011, 11, 289-301.
4. S. Negrini, V. G. Gorgoulis and T. D. Halazonetis, *Nature reviews. Molecular cell biology*, 2010, 11, 220-228.
5. C. W. Brennan, R. G. Verhaak, A. McKenna, B. Campos, H. Noushmehr, S. R. Salama, S. Zheng, D. Chakravarty, J. Z. Sanborn, S. H. Berman, R. Beroukhim, B. Bernard, C. J. Wu, G. Genovese, I. Shmulevich, J. Barnholtz-Sloan, L. Zou, R. Vegesna, S. A. Shukla, G. Ciriello, W. K. Yung, W. Zhang, C. Sougnez, T. Mikkelsen, K. Aldape, D. D. Bigner, E. G. Van Meir, M. Prados, A. Sloan, K. L. Black, J. Eschbacher, G. Finocchiaro, W. Friedman, D. W. Andrews, A. Guha, M. Iacocca, B. P. O'Neill, G. Foltz, J. Myers, D. J. Weisenberger, R. Penny, R. Kucherlapati, C. M. Perou, D. N. Hayes, R. Gibbs, M. Marra, G. B. Mills, E. Lander, P. Spellman, R. Wilson, C. Sander, J. Weinstein, M. Meyerson, S. Gabriel, P. W. Laird, D. Haussler, G. Getz, L. Chin and T. R. Network, *Cell*, 2013, 155, 462-477.
6. R. G. Verhaak, K. A. Hoadley, E. Purdom, V. Wang, Y. Qi, M. D. Wilkerson, C. R. Miller, L. Ding, T. Golub, J. P. Mesirov, G. Alexe, M. Lawrence, M. O'Kelly, P. Tamayo, B. A. Weir, S. Gabriel, W. Winckler, S. Gupta, L. Jakkula, H. S. Feiler, J. G. Hodgson, C. D. James, J. N. Sarkaria, C. Brennan, A. Kahn, P. T. Spellman, R. K. Wilson, T. P. Speed, J. W. Gray, M. Meyerson, G. Getz, C. M. Perou, D. N. Hayes and N. Cancer Genome Atlas Research, *Cancer cell*, 2010, 17, 98-110.
7. K. Robison, *Briefings in bioinformatics*, 2010, 11, 524-534.
8. D. W. Parsons, S. Jones, X. Zhang, J. C. Lin, R. J. Leary, P. Angenendt, P. Mankoo, H. Carter, I. M. Siu, G. L. Gallia, A. Olivi, R. McLendon, B. A. Rasheed, S. Keir, T. Nikolskaya, Y. Nikolsky, D. A. Busam, H. Tekleab, L. A. Diaz, Jr., J. Hartigan, D. R. Smith, R. L. Strausberg, S. K. Marie, S. M. Shinjo, H. Yan, G. J. Riggins, D. D. Bigner, R. Karchin, N. Papadopoulos, G. Parmigiani, B. Vogelstein, V. E. Velculescu and K. W. Kinzler, *Science*, 2008, 321, 1807-1812.
9. N. Cancer Genome Atlas Research, *Nature*, 2008, 455, 1061-1068.
10. N. J. Szerlip, A. Pedraza, D. Chakravarty, M. Azim, J. McGuire, Y. Fang, T. Ozawa, E. C. Holland, J. T. Huse, S. Jhanwar, M. A. Leversha, T. Mikkelsen and C. W. Brennan, *Proceedings of the National Academy of Sciences of the United States of America*, 2012, 109, 3041-3046.
11. R. Martinez, V. Rohde and G. Schackert, *Journal of neuro-oncology*, 2010, 96, 321-329.
12. N. Cancer Genome Atlas, *Nature*, 2012, 490, 61-70.
13. N. Cancer Genome Atlas Research, *Nature*, 2011, 474, 609-615.
14. M. D. Leiserson, D. Blokh, R. Sharan and B. J. Raphael, *PLoS computational biology*, 2013, 9, e1003054.
15. F. Vandin, E. Upfal and B. J. Raphael, *Genome research*, 2012, 22, 375-385.

16. C. Eifert and R. S. Powers, *Nature reviews. Cancer*, 2012, 12, 572-578.
17. A. Youn and R. Simon, *Bioinformatics*, 2011, 27, 175-181.
18. R. Beroukhim, G. Getz, L. Nghiemphu, J. Barretina, T. Hsueh, D. Linhart, I. Vivanco, J. C. Lee, J. H. Huang, S. Alexander, J. Du, T. Kau, R. K. Thomas, K. Shah, H. Soto, S. Perner, J. Prensner, R. M. Debiasi, F. Demichelis, C. Hatton, M. A. Rubin, L. A. Garraway, S. F. Nelson, L. Liau, P. S. Mischel, T. F. Cloughesy, M. Meyerson, T. A. Golub, E. S. Lander, I. K. Mellinghoff and W. R. Sellers, *Proceedings of the National Academy of Sciences of the United States of America*, 2007, 104, 20007-20012.
19. U. D. Akavia, O. Litvin, J. Kim, F. Sanchez-Garcia, D. Kotliar, H. C. Causton, P. Pochanard, E. Mozes, L. A. Garraway and D. Pe'er, *Cell*, 2010, 143, 1005-1017.
20. Y. A. Kim, S. Wuchty and T. M. Przytycka, *PLoS computational biology*, 2011, 7, e1001095.
21. D. L. Masica and R. Karchin, *Cancer research*, 2011, 71, 4550-4561.
22. A. Bashashati, G. Haffari, J. Ding, G. Ha, K. Lui, J. Rosner, D. G. Huntsman, C. Caldas, S. A. Aparicio and S. P. Shah, *Genome biology*, 2012, 13, R124.
23. M. G. McNamara, S. Sahebjam and W. P. Mason, *Cancers*, 2013, 5, 1103-1119.
24. B. Vogelstein and K. W. Kinzler, *Nature medicine*, 2004, 10, 789-799.
25. C. J. Vaske, S. C. Benz, J. Z. Sanborn, D. Earl, C. Szeto, J. Zhu, D. Haussler and J. M. Stuart, *Bioinformatics*, 2010, 26, i237-245.
26. D. K. Slonim, *Nature genetics*, 2002, 32 Suppl, 502-508.
27. J. R. Pollack, T. Sorlie, C. M. Perou, C. A. Rees, S. S. Jeffrey, P. E. Lonning, R. Tibshirani, D. Botstein, A. L. Borresen-Dale and P. O. Brown, *Proceedings of the National Academy of Sciences of the United States of America*, 2002, 99, 12963-12968.
28. C. N. Henrichsen, E. Chaignat and A. Reymond, *Human molecular genetics*, 2009, 18, R1-8.
29. D. G. Albertson, C. Collins, F. McCormick and J. W. Gray, *Nature genetics*, 2003, 34, 369-376.
30. D. Kita, Y. Yonekawa, M. Weller and H. Ohgaki, *Acta neuropathologica*, 2007, 113, 295-302.
31. L. J. Valentijn, A. Koppen, R. van Asperen, H. A. Root, F. Haneveld and R. Versteeg, *Cancer research*, 2005, 65, 3136-3145.
32. J. A. Engelman, K. Zejnullahu, T. Mitsudomi, Y. Song, C. Hyland, J. O. Park, N. Lindeman, C. M. Gale, X. Zhao, J. Christensen, T. Kosaka, A. J. Holmes, A. M. Rogers, F. Cappuzzo, T. Mok, C. Lee, B. E. Johnson, L. C. Cantley and P. A. Janne, *Science*, 2007, 316, 1039-1043.
33. B. I. Ferreira, J. Alonso, J. Carrillo, F. Acquadro, C. Largo, J. Suela, M. R. Teixeira, N. Cerveira, A. Molares, G. Gomez-Lopez, A. Pestana, A. Sastre, P. Garcia-Miguel and J. C. Cigudosa, *Oncogene*, 2008, 27, 2084-2090.
34. A. B. Olshen, E. S. Venkatraman, R. Lucito and M. Wigler, *Biostatistics*, 2004, 5, 557-572.
35. M. Kanehisa and S. Goto, *Nucleic acids research*, 2000, 28, 27-30.
36. C. Li, X. Li, Y. Miao, Q. Wang, W. Jiang, C. Xu, J. Li, J. Han, F. Zhang, B. Gong and L. Xu, *Nucleic acids research*, 2009, 37, e131.
37. C. H. Mermel, S. E. Schumacher, B. Hill, M. L. Meyerson, R. Beroukhim and G. Getz, *Genome biology*, 2011, 12, R41.
38. X. Shen, S. Li, L. Zhang, H. Li, G. Hong, X. Zhou, T. Zheng, W. Zhang, C. Hao, T. Shi, C. Liu and Z. Guo, *PloS one*, 2013, 8, e61214.
39. S. Kohler, S. Bauer, D. Horn and P. N. Robinson, *American journal of human genetics*, 2008, 82, 949-958.
40. Y. Li and J. C. Patra, *Bioinformatics*, 2010, 26, 1219-1224.
41. T. S. Keshava Prasad, R. Goel, K. Kandasamy, S. Keerthikumar, S. Kumar, S. Mathivanan, D. Telikicherla, R. Raju, B. Shafreen, A. Venugopal, L. Balakrishnan, A. Marimuthu, S. Banerjee, D. S. Somanathan, A. Sebastian, S. Rani, S. Ray, C. J. Harrys Kishore, S. Kanth, M. Ahmed, M. K. Kashyap, R. Mohmood, Y. L. Ramachandra, V. Krishna, B. A. Rahiman, S. Mohan, P. Ranganathan, S. Ramabadran, R. Chaerkady and A. Pandey, *Nucleic acids research*, 2009, 37, D767-772.
42. N. Shinojima, K. Tada, S. Shiraishi, T. Kamiryo, M. Kochi, H. Nakamura, K. Makino, H. Saya, H. Hirano, J. Kuratsu, K. Oka, Y. Ishimaru and Y. Ushio, *Cancer research*, 2003, 63, 6962-6970.
43. N. Montano, T. Cenci, M. Martini, Q. G. D'Alessandris, F. Pelacchi, L. Ricci-Vitiani, G. Maira, R. De Maria, L. M. Larocca and R. Pallini, *Neoplasia*, 2011, 13, 1113-1121.
44. M. Van den Eynde, J. F. Baurain, F. Mazzeo and J. P. Machiels, *Acta clinica Belgica*, 2011, 66, 10-17.
45. D. L. Wheeler, E. F. Dunn and P. M. Harari, *Nature reviews. Clinical oncology*, 2010, 7, 493-507.
46. I. K. Mellinghoff, M. Y. Wang, I. Vivanco, D. A. Haas-Kogan, S. Zhu, E. Q. Dia, K. V. Lu, K. Yoshimoto, J. H. Huang, D. J. Chute, B. L. Riggs, S. Horvath, L. M. Liau, W. K. Cavenee, P. N. Rao, R. Beroukhim, T. C. Peck, J. C. Lee, W. R. Sellers, D. Stokoe, M. Prados, T. F. Cloughesy, C. L. Sawyers and P. S. Mischel, *The New England journal of medicine*, 2005, 353, 2012-2024.
47. G. Ciriello, E. Cerami, C. Sander and N. Schultz, *Genome research*, 2012, 22, 398-406.
48. H. S. Phillips, S. Kharbanda, R. Chen, W. F. Forrest, R. H. Soriano, T. D. Wu, A. Misra, J. M. Nigro, H. Colman, L. Soroceanu, P. M. Williams, Z. Modrusan, B. G. Feuerstein and K. Aldape, *Cancer cell*, 2006, 9, 157-173.
49. J. A. Jakob, R. L. Bassett, Jr., C. S. Ng, J. L. Curry, R. W. Joseph, G. C. Alvarado, M. L. Rohlfs, J. Richard, J. E. Gershenwald, K. B. Kim, A. J. Lazar, P. Hwu and M. A. Davies, *Cancer*, 2012, 118, 4014-4023.
50. G. Mulligan, D. I. Lichter, A. Di Bacco, S. J. Blakemore, A. Berger, E. Koenig, H. Bernard, W. Trepicchio, B. Li, R. Neuwirth, N. Chattopadhyay, J. B. Bolen, A. J. Dorner, H. van de Velde, D. Ricci, S. Jagannath, J. R. Berenson, P. G. Richardson, E. A. Stadtmauer, R. Z. Orlowski, S. Lonial, K. C. Anderson, P. Sonneveld, J. F. San Miguel, D. L. Esseltine and M. Schu, *Blood*, 2014, 123, 632-639.
51. P. Dell'albani, M. Rodolico, R. Pellitteri, E. Tricarichi, S. A. Torrisi, S. D'Antoni, M. Zappia, V. Albanese, R. Caltabiano, N. Platania, E. Aronica and M. V. Catania, *Neuro-oncology*, 2014, 16, 204-216.
52. N. Saito, J. Fu, S. Zheng, J. Yao, S. Wang, D. D. Liu, Y. Yuan, E. P. Sulman, F. F. Lang, H. Colman, R. G. Verhaak, W. K. Yung and D. Koul, *Stem cells*, 2014, 32, 301-312.
53. W. B. Li, M. W. Ma, L. J. Dong, F. Wang, L. X. Chen and X. R. Li, *Cancer biology & therapy*, 2011, 12, 477-483.
54. Y. Xiao, J. Guan, Y. Ping, C. Xu, T. Huang, H. Zhao, H. Fan, Y. Li, Y. Lv, T. Zhao, Y. Dong, H. Ren and X. Li, *Nucleic acids research*, 2012, 40, 7653-7665.
55. J. Chen, S. Kesari, C. Rooney, P. R. Strack, J. Chen, H. Shen, L. Wu and J. D. Griffin, *Genes & cancer*, 2010, 1, 822-835.
56. Y. Xiao, Y. Ping, H. Fan, C. Xu, J. Guan, H. Zhao, Y. Li, Y. Lv, Y. Jin, L. Wang and X. Li, *Neuro-oncology*, 2013, 15, 818-828.
57. A. Ylipaa, O. Yli-Harja, W. Zhang and M. Nykter, *BMC systems biology*, 2013, 7 Suppl 1, S1.
58. J. G. Lohr, P. Stojanov, S. L. Carter, P. Cruz-Gordillo, M. S. Lawrence, D. Auclair, C. Sougnez, B. Knoechel, J. Gould, G. Saksena, K. Cibulskis, A. McKenna, M. A. Chapman, R. Straussman, J. Levy, L. M. Perkins, J. J. Keats, S. E. Schumacher, M. Rosenberg, C. Multiple Myeloma Research, G. Getz and T. R. Golub, *Cancer cell*, 2014, 25, 91-101.
59. S. Horswell, N. Matthews and C. Swanton, *Cancer letters*, 2013, 340, 220-226.
60. S. Bhatia, J. V. Frangioni, R. M. Hoffman, A. J. Iafrate and K. Polyak, *Nature biotechnology*, 2012, 30, 604-610.
61. A. Olar and K. D. Aldape, *The Journal of pathology*, 2014, 232, 165-177.
62. C. E. Meacham and S. J. Morrison, *Nature*, 2013, 501, 328-337.
63. D. Waniczek, M. Snietura, J. Mlynarczyk-Liszka, W. Piglowski, A. Kopec, D. Lange, M. Rudzki and J. Arendt, *Polish journal of pathology : official journal of the Polish Society of Pathologists*, 2013, 64, 15-20.
64. E. N. Gurzov and M. Izquierdo, *Neurological research*, 2006, 28, 493-499.
65. Y. Itatani, K. Kawada, T. Fujishita, F. Kakizaki, H. Hirai, T. Matsumoto, M. Iwamoto, S. Inamoto, E. Hatano, S. Hasegawa, T. Maekawa, S. Uemoto, Y. Sakai and M. M. Taketo, *Gastroenterology*, 2013, 145, 1064-1075 e1011.

66. L. Y. Huw, C. O'Brien, A. Pandita, S. Mohan, J. M. Spoerke, S. Lu, Y. Wang, G. M. Hampton, T. R. Wilson and M. R. Lackner, *Oncogenesis*, 2013, 2, e83.