Mutual exclusivity analysis identifies oncogenic network modules

Giovanni Ciriello,[1,3,4] Ethan Cerami,[1,2,3] Chris Sander,[1] and Nikolaus Schultz[1]

[1] *Computational Biology Center, Memorial Sloan-Kettering Cancer Center, New York, New York 10065, USA;* [2] *Tri-Institutional Training Program in Computational Biology and Medicine, New York, New York 10065, USA*

Although individual tumors of the same clinical type have surprisingly diverse genomic alterations, these events tend to occur in a limited number of pathways, and alterations that affect the same pathway tend to not co-occur in the same patient. While pathway analysis has been a powerful tool in cancer genomics, our knowledge of oncogenic pathway modules is incomplete. To systematically identify such modules, we have developed a novel method, Mutual Exclusivity Modules in cancer (MEMo). The method uses correlation analysis and statistical tests to identify network modules by three criteria: (1) Member genes are recurrently altered across a set of tumor samples; (2) member genes are known to or are likely to participate in the same biological process; and (3) alteration events within the modules are mutually exclusive. Applied to data from the Cancer Genome Atlas (TCGA), the method identifies the principal known altered modules in glioblastoma (GBM) and highlights the striking mutual exclusivity of genomic alterations in the PI(3)K, p53, and Rb pathways. In serous ovarian cancer, we make the novel observation that inactivation of *BRCA1* and *BRCA2* is mutually exclusive of amplification of *CCNE1* and inactivation of *RB1*, suggesting distinct alternative causes of genomic instability in this cancer type; and, we identify *RBBP8* as a candidate oncogene involved in Rb-mediated cell cycle control. When applied to any cancer genomics data set, the algorithm can nominate oncogenic alterations that have a particularly strong selective effect and may also be useful in the design of therapeutic combinations in cases where mutual exclusivity reflects synthetic lethality.

[Supplemental material is available for this article.]

Large-scale cancer genomics projects, such as the Cancer Genome Atlas (TCGA) and the International Cancer Genome Consortium (ICGC), are providing an unprecedented and high-resolution view of the molecular defects in dozens of cancer types (Stratton et al. 2009). A key challenge in all of these projects is to distinguish "driver" mutations, which contribute to tumorigenesis, from "passenger" mutations, which are functionally neutral and do not contribute to tumor development (Greenman et al. 2007). A second key challenge is to identify biological pathways, which are frequently perturbed within tumor cells, and lead to the acquisition of tumorigenic properties, such as cell proliferation, angiogenesis, or metastasis (Hanahan and Weinberg 2000, 2011).

A number of approaches have been developed to address both of these challenges. For example, several methods identify recurrently altered driver mutations in cancer, by comparing alteration rates in individual genes or regions of copy number alteration against an empirically derived background alteration rate (Beroukhim et al. 2007; Getz et al. 2007). Other methods explicitly do not take recurrence into account—for example, machine learning methods based on prior known cancer-causing mutations have been successfully trained to classify and predict the functional consequences of somatic missense mutations (Kaminker et al. 2007; Carter et al. 2009).

Other recent methods have attempted to use integrative network analysis to both identify candidate driver genes and candidate pathways. For example, Torkamani and Schork inferred regulatory networks from gene expression data and identified network modules enriched for mutations and potential rare cancer driver mutations (Torkamani and Schork 2009). Vandin and colleagues have used a network diffusion algorithm to identify subnetworks enriched for mutations within a large gene interaction network (Vandin et al. 2010). Vaske and colleagues have used a factor graph belief propagation algorithm to integrate copy number and expression data to score curated pathways (Vaske et al. 2010). We and others have also investigated the network properties of cancer alterations, and have noted that cancer alterations tend to cluster within closely knit network modules or communities, and that altered modules are closely linked to specific biological pathways (Cerami et al. 2010; Wu et al. 2010).

Properties of perturbed cancer pathways

Rather than developing a new broad-based algorithmic framework for identifying all driver genes and altered pathways, we have chosen to develop an algorithm, Mutual Exclusivity Modules (MEMo), for identifying a specific class of connected gene sets. These gene sets have three properties: First, member genes are altered (either via somatic mutation of copy number alteration) more frequently than expected by chance; second, member genes are likely to participate in the same biological pathway or process, as determined from background pathway and network knowledge; and third, genomic events within the network exhibit a statistically significant level of mutual exclusivity.

The rationale for finding such connected modules begins with several fundamental observations from recent cancer genomics studies. First, although individual tumors exhibit a diversity of somatic mutations and copy number alterations, many of these events tend to affect a limited number of biological pathways. For example, in glioblastoma multiforme (GBM), TCGA has identified alterations in the p53 pathway in up to 87% of patients, but the exact mechanism of alteration varies by patient, and includes mutation or

[3] **These authors contributed equally to this work.**
[4] **Corresponding author.**
E-mail memo@cbio.mskcc.org.
Article published online before print. Article, supplemental material, and publication date are at http://www.genome.org/cgi/doi/10.1101/gr.125567.111.
Freely available online through the *Genome Research* Open Access option.

22:398–406 © 2012 by Cold Spring Harbor Laboratory Press; ISSN 1088-9051/12; www.genome.org

homozygous deletion of *TP53* or p16/ARF (*CDKN2A*), or amplification of *MDM2*/*MDM4* (The Cancer Genome Atlas Research Network 2008). This and other examples from recent sequencing studies have provided increased evidence that cancer genes tend to cluster within a limited set of essential biological pathways, and that diversity and complexity at the gene level can be substantially reduced at the pathway level (Velculescu 2008; Stratton et al. 2009).

Second, many tumor profiling projects have observed mutually exclusive genomic alterations across many patients—for example, *TP53* is mutated and *MDM2* is copy number amplified, but only very few patients harbor both genetic lesions (The Cancer Genome Atlas Research Network 2008). Additional examples in other cancer types include mutual exclusivity between *APC* and *CTNNB1* mutations (both involved in the beta-catenin signaling pathway) (Sparks et al. 1998), and *BRAF* and *KRAS* mutations (both involved in the common RAS/RAF signaling pathway) in colorectal cancer (Rajagopalan et al. 2002); and mutual exclusivity between *BRCA1/2* mutations and *BRCA1* epigenetic silencing in serous ovarian cancer (The Cancer Genome Atlas Research Network 2011).

As these diverse examples demonstrate, mutually exclusive genomic events provide strong genetic evidence that the altered genes are functionally linked in a common biological pathway. Alteration to these pathways enables tumors to bypass or activate a specific set of cellular processes, also known as the hallmarks of cancer (Hanahan and Weinberg 2000, 2011). Once a gene that is involved in one of these processes is altered, the tumor cell acquires a selective advantage, e.g., increased proliferation, which promotes clonal expansion. Observations indicate that a second hit, leading to the same downstream effect, is less likely to occur.

Two biologically plausible scenarios may explain the resulting pattern of mutually exclusive genomic alterations within a cancer study. In the first scenario, alteration to a second gene within the same pathway offers no further selective advantage. This hypothesis would, for example, explain the observed mutual exclusivity between *MDM2* amplification and *TP53* inactivation in the p53 signaling pathway in GBM: Once either of the two genes is altered, the pathway is compromised and apoptosis evaded. Additional alterations to the pathway do not change the effect on the apoptosis process and are not selected for.

In the second scenario, alteration to the second gene within the same pathway actually leads to a disadvantage for the cell, in the extreme case, to cell death. This scenario is referred to as synthetic lethality.

As evidenced in the examples below, several of the networks identified by MEMo show significant mutual exclusivity between functionally redundant genomic alterations, but multiple alterations in the same tumor are also occasionally present. This evidence supports the first hypothesis as the more plausible of the scenarios, but we cannot systematically distinguish between the two hypotheses based on genomic data alone.

In either of the above scenarios, the observed mutual exclusivity provides evidence that the altered genes are functionally linked, and most likely linked in a common pathway or biological process. These patterns have not been adequately exploited by algorithms to automatically identify altered pathways in cancer.

Results

Overview of MEMo algorithm

The goal of MEMo is to identify sets of connected genes that are recurrently altered, likely to belong to the same pathway or biological process, and exhibit patterns of mutually exclusive genetic alteration across multiple patients. Modules that exhibit these three properties are very likely to drive cancer progression, and we refer to such modules as candidate "driver networks." As outlined in Figure 1, the algorithm proceeds in four steps.

Step 1: Build binary event matrix of significantly altered genes

In Step 1, the algorithm uses the full set of somatic mutations and copy number events across all observed samples, applies multiple gene filters for recurrence and concordant mRNA expression, and generates a binary event matrix of all target genes in all samples.

Three gene filters are used, with the goal of identifying those genes most likely involved in tumor initiation or progression. The first filter identifies genes that are mutated significantly above the background mutation rate (BMR). Specifically, input is restricted to significantly mutated genes, as determined by the Standard test of

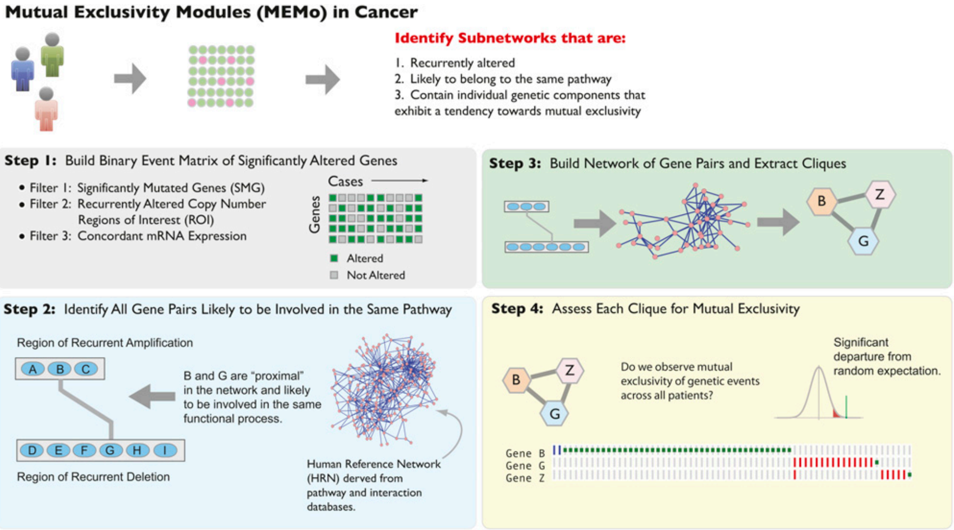
Mutual Exclusivity Modules (MEMo) in Cancer
Identify Subnetworks that are:
1. Recurrently altered
2. Likely to belong to the same pathway
3. Contain individual genetic components that exhibit a tendency towards mutual exclusivity
Step 1: Build Binary Event Matrix of Significantly Altered Genes
• Filter 1: Significantly Mutated Genes (SMG)
• Filter 2: Recurrently Altered Copy Number Regions of Interest (ROI)
• Filter 3: Concordant mRNA Expression
Cases
Genes
Altered
Not Altered
Step 3: Build Network of Gene Pairs and Extract Cliques
B
Z
G
Step 2: Identify All Gene Pairs Likely to be Involved in the Same Pathway
Region of Recurrent Amplification
A
B
C
B and G are "proximal" in the network and likely to be involved in the same functional process.
D
E
F
G
H
I
Region of Recurrent Deletion
Human Reference Network (HRN) derved from pathway and interaction databases.
Step 4: Assess Each Clique for Mutual Exclusivity
B
Z
G
Do we observe mutual exclusivity of genetic events across all patients?
Significant departure from random expectation.
Gene B
Gene G
Gene Z

Figure 1. Identifying mutual exclusivity modules (MEMo) in cancer. Overview of the algorithm.

the MutSig algorithm (Getz et al. 2007). The Standard MutSig test takes as input the number of bases successfully sequenced for each gene, the number of observed mutations, and the empirically derived BMR, and applies a standard binomial test to determine if the number of observed mutations is greater than expected by chance (Getz et al. 2007). Restricting the initial search space to significantly mutated genes is a critical filter, as it removes genes which may be subject to frequent mutation, owing to large gene size, rather than tumorigenic advantage.

The second filter identifies genes that are targets of recurrent copy number amplification or deletion. Specifically, MEMo restricts its input to genes within statistically significant Regions of Interest (ROI), as determined by the GISTIC or RAE algorithms (Beroukhim et al. 2007; Taylor et al. 2008). GISTIC and RAE combine copy number data for multiple samples and use a permutation test to identify regions of the genome that are altered more frequently than expected by chance (Beroukhim et al. 2007; Taylor et al. 2008). Both methods also identify ROI, each of which can contain between one and hundreds of target genes. For each ROI, MEMo uses an additional filter to retain only the genes that are altered by high-level amplification, homozygous deletion, or mutation in at least 3% of samples.

The third filter identifies copy number altered genes that have concordant mRNA expression: Genes that are not significantly up-regulated when amplified (or not significantly down-regulated when deleted) are unlikely to be drivers (Cowin et al. 2010). MEMo uses a two-class comparison to identify genes with correlated expression: For amplified regions, the mRNA levels of all diploid cases are compared to the mRNA levels of all amplified cases, and a statistically significant increase in the amplified cases marks the gene as concordant (a similar test for deleted genes is performed, see Methods).

Step 1 results in a set of recurrently altered entities, where an entity can be a significantly mutated gene or a copy number ROI containing multiple genes, all of which have passed the mRNA concordance filter. In Step 1, MEMo also builds a binary matrix M, where each entry m_{ij} refers to the status of gene i in the sample j, and whose value is determined as follows:

$$m_{ij} = \begin{cases} 1 & \text{if gene } i \text{ is altered by a non-synonymous somatic mutation in sample } j; \\ 1 & \text{if gene } i \text{ is homozygously deleted in sample } j \text{ and is in a deleted ROI;} \\ 1 & \text{if gene } i \text{ is amplified in sample } j \text{ and is in an amplified ROI;} \\ 0 & \text{otherwise.} \end{cases}$$

Note that, under the first condition, the binary event matrix does not account for multiple mutations within the same gene/case pair, nor does it account for varying allelic frequency.

Step 2: Identify all gene pairs likely to be involved in the same pathway

In Step 2, MEMo performs a global gene comparison test to determine all pairs of genes that are functionally connected to one another, based on prior pathway and network knowledge. This step is based on the observation that many biological processes and pathways are implemented by modules of interacting proteins, and proteins involved in the same process have a high propensity to interact with one another, or to interact in the same local clusters (Hartwell et al. 1999; Oti et al. 2006; Barabási et al. 2011). To determine if two genes are functionally connected to one another, the algorithm requires a background reference network, referred to here as the Human Reference Network (HRN). We used two different background networks to serve as the HRN. The first network is derived from manually curated interactions only. The second contains manually curated interactions plus additional inferred interactions derived from non-curated sources of information, including high-throughput derived protein–protein interactions, gene coexpression, protein domain interaction, Gene Ontology (GO) annotations, and text-mined protein interactions (Wu et al. 2010) (see Methods).

Multiple metrics have been developed for assessing network proximity of two nodes in a graph (for review, see Liben-Nowell and Kleinberg 2003). These metric quantify proximity by assessing the number and optionally the edge degree of common neighbors shared by the two target nodes. In the context of the MEMo HRN, two genes can be assessed as proximal even if they are not directly connected but share a large number of common neighbors, and are therefore likely to belong to the same functional module.

Step 3: Build graph of gene pairs and extract cliques

In Step 3, MEMo builds a graph of all similar gene pairs, by creating an edge between two genes if they are found similar by the network proximity metric in Step 2. MEMo then extracts from this graph all "maximal cliques," i.e., all fully connected subgraphs such that each subgraph cannot be contained by another fully connected subgraph. These cliques represent local clusters, containing proteins of likely similar biological function.

Step 4: Assess each clique for mutual exclusivity

In Step 4, MEMo determines whether each clique identified in Step 3 exhibits a pattern of mutually exclusive genomic alterations, and whether this pattern is unlikely to be observed by chance. We propose a null model generated by randomly permuting the set of genomic events, while preserving the overall distribution of observed alterations across both genes and samples. This is crucial to preserve both tumor specific alterations, and heterogeneity in mutation and copy number alteration rate across patients. To do so, we introduce a Markov chain Monte Carlo permutation strategy based on random network generation models, which we refer to as "switching permutation" (see Methods). An empirically derived P-value is generated to estimate the significance of the observed alteration frequency of each module, compared to those expected for the same module after randomly permuting the set of observed genomic alterations (see Methods).

Recall that, in Step 3, MEMo identifies only maximal cliques within the graph. These maximal cliques may not exhibit patterns of mutual exclusivity that pass the permutation test outlined above; however, sub-cliques may exhibit these patterns. To assess sub-cliques in a manner that limits the number of statistical tests, MEMo performs a conditional trace through the cliques, selecting at each step only those sub-cliques that are more likely to be significant based on the overall extent of alteration (see Methods).

In searching for mutually exclusive gene sets, a brute force approach could be used—for example, one could examine all gene triplets or quadruplets and assess for frequency of alteration and mutual exclusivity. However, the combinatorics of such an approach would result in a extremely large number of hypotheses, making it difficult to achieve statistical significance. MEMo addresses this issue by reducing the number of statistical tests to a list of high-value candidate modules, enabling one to not only achieve statistical significance, but also to make biological hypothesis about the modules identified.

Application to glioblastoma multiforme (GBM)

We analyzed 138 GBM cases from TCGA, all of which have targeted sequencing data for ~1200 genes, genome-wide copy number profiling, and mRNA expression data. This includes the 91 cases originally analyzed in the 2008 TCGA paper (The Cancer Genome Atlas Research Network 2008), plus 47 newly sequenced samples. Ten significantly mutated genes and 158 ROI of amplification or deletion were used as input. We analyzed GBM data using two different HRNs, and complete results are provided in Supplemental Table 1 (Tabs 1 and 2). Very similar modules were identified for both networks, and here we describe in detail the results with HRN1, while providing comments on modules found specifically with HRN2. Below, P^* indicates a P-value that has been adjusted for multiple testing.

Within HRN1, eight modules with $P^* < 0.05$ are identified. The two highest scoring modules contain four genes in total: The first includes *CDKN2A*, *CDK4*, and *RB1*, and the second *CDKN2B*, *CDK4*, and *RB1* (both have $P^* < 1.0 \times 10^{-2}$) (Fig. 2A). The modules are altered in 68% and 73% of cases, respectively, and all genes are core members of the Rb pathway (Sherr and McCormick 2002). These modules recapitulate findings from TCGA and earlier studies that glioblastomas nearly universally circumvent cell cycle inhibition through genetic alterations to the Rb pathway (Ohgaki and Kleihues 2007; The Cancer Genome Atlas Research Network 2008).

Two other high-scoring modules include one involving *CDNK2A*, *MDM2*, and *TP53*, and a second involving *TP53*, *MDM2*, and *MDM4* (Fig. 2B). The modules are altered in 75% and 48% of patients, respectively. Because all genes in these two modules are members of the p53 signaling pathway (Sherr and McCormick 2002), we merged them and tested the resulting set of four genes for mutual exclusivity. The union of the two modules still shows significant mutual exclusivity ($P < 1.0 \times 10^{-4}$).

The final pair of high-scoring modules contains core members of the RTK/RAS/PI(3)K signaling pathway. Specifically, one module contains *EGFR*, *PDGFRA*, and *PTEN* (altered in 74% of cases) and a second module contains *EGFR*, *PTEN*, and *PIK3R1* (altered in 73% of cases). Similar to the p53 signaling case, we merged the two modules and, again, the combined gene set shows alterations in a statistically significant mutually exclusive pattern ($P = 0.0018$). Major downstream effects of RTK/RAS/PI(3)K activation include cell growth, proliferation, survival, and motility, all factors that drive tumor progression, and these pathway components have all been previously identified in glioblastoma (Ohgaki and Kleihues 2007; The Cancer Genome Atlas Research Network 2008).

Using HRN2, MEMo confirms the findings of HRN1, while adding two new modules. First, it finds a module involving *EGFR*, *PDGFRA*, and *NF1* ($P^* < 1.0 \times 10^{-2}$), thus correctly including *NF1* in the set of alterations affecting RTK/RAS/PI(3)K signaling pathway. Second, MEMo identifies as highly significant the triplet including *TP53*, *CDKN2A*, and *GLI1*. This glioma-associated oncogene (*GLI1*) has been shown to repress *TP53* activity by forming an inhibitory loop (Stecca and i Altalba 2009), and was not reported in the original p53 pathway analysis reported by the TCGA project

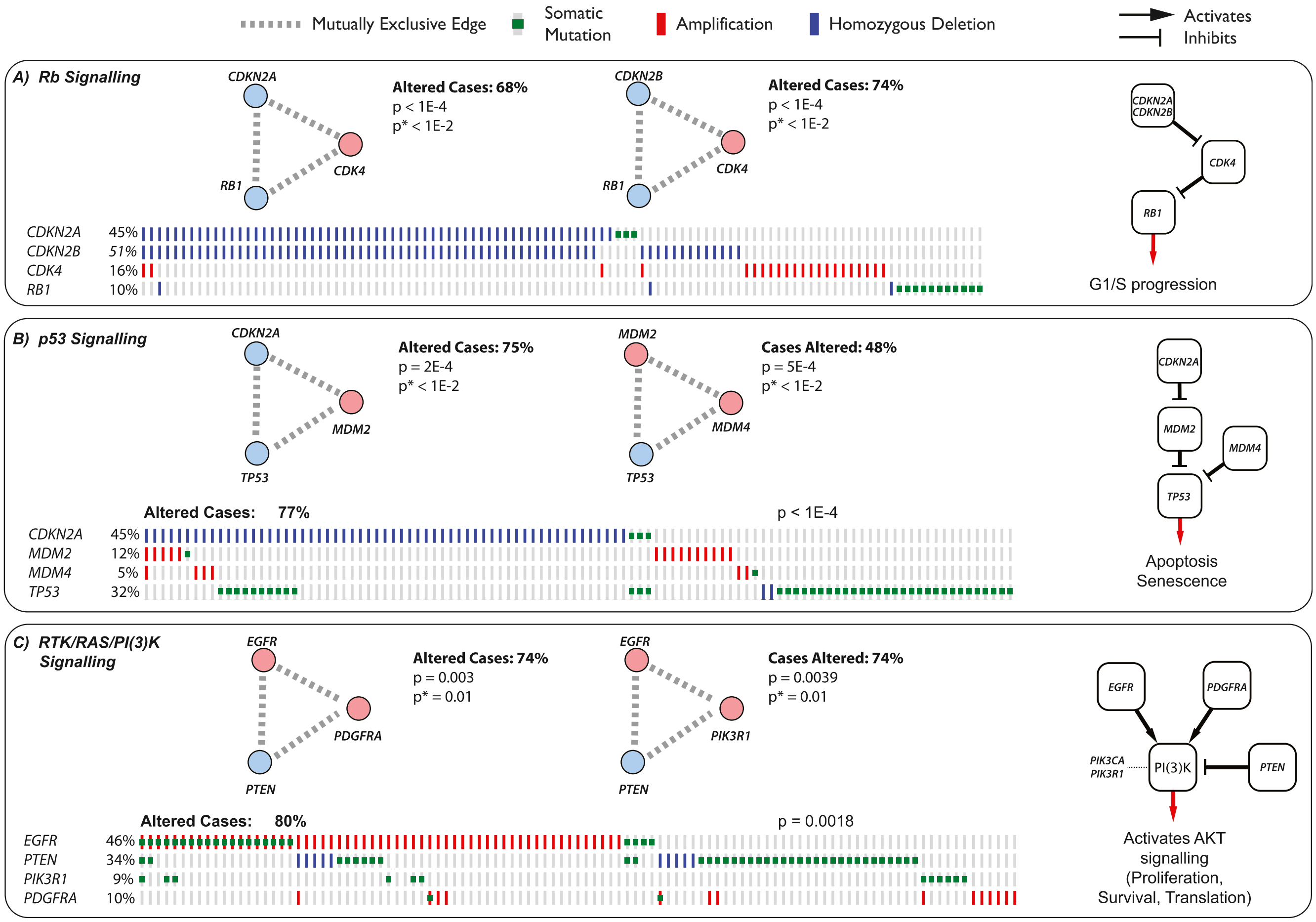

Figure 2. Top-scoring modules in the TCGA GBM data set. The top-scoring mutually exclusive modules correspond closely to core signaling pathways including Rb signaling (*A*), p53 signaling (*B*), and RTK/RAS/PI(3)K signaling (*C*).

(The Cancer Genome Atlas Research Network 2008). The repression function of the protein may thus explain the observed mutual exclusivity and further increases the alteration extent of the p53 signaling pathway observed in GBM.

In summary, MEMo is able to automatically recapitulate previously manually identified altered pathways in glioblastoma (The Cancer Genome Atlas Research Network 2008). Furthermore it highlights the striking mutual exclusivity of genomic alterations in the PI(3)K, p53, and Rb pathways.

Application to serous ovarian cancer

We analyzed 316 serous ovarian cancer cases from TCGA, all of which have whole-exome sequencing data, DNA methylation, copy number profiling, and mRNA expression data. Sixteen significantly mutated genes and 118 ROI of amplification or deletion were used as input. We analyzed ovarian cancer data using both HRNs (see Methods), and complete results are provided in Supplemental Table 1 (Tabs 3 and 4). Within HRN1, we found three statistically significant modules as shown in Figures 3 and 4.

The top scoring module ($P^* < 1.0 \times 10^{-2}$) contains five genes: *BRCA1*, *BRCA2*, *CCNE1*, *RB1*, and *RBBP8* (Fig. 3A). All five genes were identified as proximal by Step 3 of MEMo, and all five share a network neighborhood consisting of genes involved in cell cycle regulation. For example, *BRCA1* and *CCNE1* share 12 common interactors, 10 of which are involved in cell cycle regulation, including *CDK1*, *CDK2*, *CCNA2*, and *BARD1*. Likewise, *BRCA2* and *RB1* share six common interactors, including *CDK1*, *CDK2*, and *CCNA2*.

The identified module contains members of two distinct pathways. Specifically, *BRCA1* and *BRCA2* are components of the homologous recombination DNA repair pathway and *RB1* and *CCNE1* are components of the Rb cell cycle regulation pathway. Notably, tumors can acquire genomic instability via alterations to either pathway. Specifically, mutations in *BRCA1* and *BRCA2* lead to defects in the homologous recombination pathway, and an inability to repair double strand breaks (DSBs) (Turner et al. 2004). In contrast, *RB1* deletions and *CCNE1* amplifications accelerate the cell cycle, resulting in defective S-phase progression, increased chromosome breakage, and increased genomic instability (Hwang and Clurman 2005). This dual path to genomic instability may account for the observed mutual exclusivity in the identified module.

There are two known functions of *RBBP8*. It is known to complex with *BRCA1* in the G_2 to M transition and acts as a tumor suppressor (Yu and Chen 2004; Chen et al. 2005). *RBBP8* has also been shown to facilitate the G_1 to S transition by activating a series of S-phase genes, including *CCND1* (Liu and Lee 2006). Specifically, *RBBP8* activates its own transcription by displacing Rb from the promoter and thus releasing its repressing activity. Higher levels of the protein come in tandem with decreasing levels of Rb and increased expression of *CCND1* (Liu and Lee 2006). The observed mutual exclusivity of *RBBP8* amplifications with other alterations in the Rb pathway suggests that the gene has an oncogenic role in ovarian cancer, and its amplification is yet another way to disable the Rb pathway (Fig. 3B).

The identified MEMo module also sheds new light on previous reports concerning *CCNE1* amplification as a marker of poor prognosis in ovarian cancer (Etemadmoghadam et al. 2009, 2010; Nakayama et al. 2010). As previously reported (The Cancer Genome Atlas Research Network 2011), the better survival of *BRCA1/2* mutated cases and the observed tendency toward mutual exclusivity between *BRCA1/2* mutation and *CCNE1* amplification prompted us to reevaluate the survival characteristics associated with *CCNE1* amplification. To do so, we first evaluated the full set of

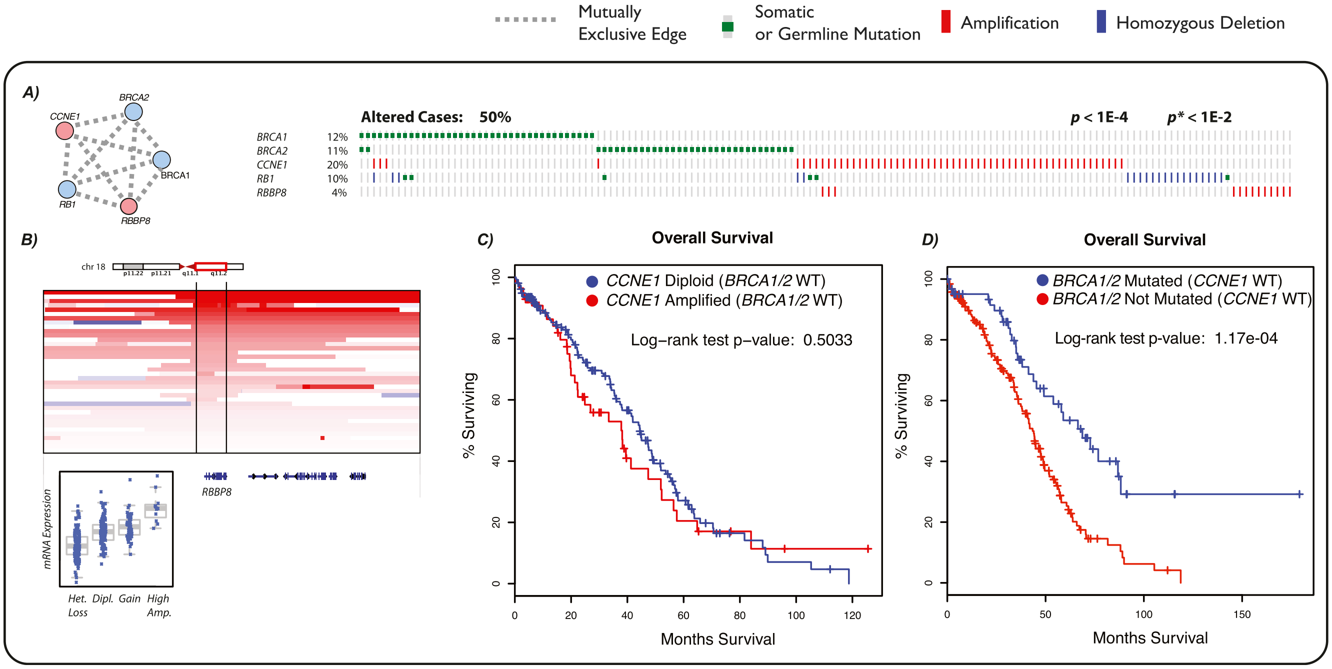

Figure 3. Top-scoring module in the TCGA serous ovarian cancer data set. (*A*) The top-scoring module contains five genes: *BRCA1*, *BRCA2*, *CCNE1*, *RB1*, and *RBBP8*. (*B*) RBBP8 is focally amplified in the TCGA ovarian cancer data set. First, overall survival for *CCNE1* amplified cases is compared to *CCNE1* wild-type cases among BRCA wild-type cases only (*C*). Then, overall survival for BRCA mutated cases is compared to BRCA wild-type cases among *CCNE1* diploid cases only (*D*). The worst outcome associated with *CCNE1* amplification is no longer detectable once BRCA mutated cases are removed. In contrast, patients with *BRCA1* or *BRCA2* mutated still show a better outcome even when compared among *CCNE1* wild type only.

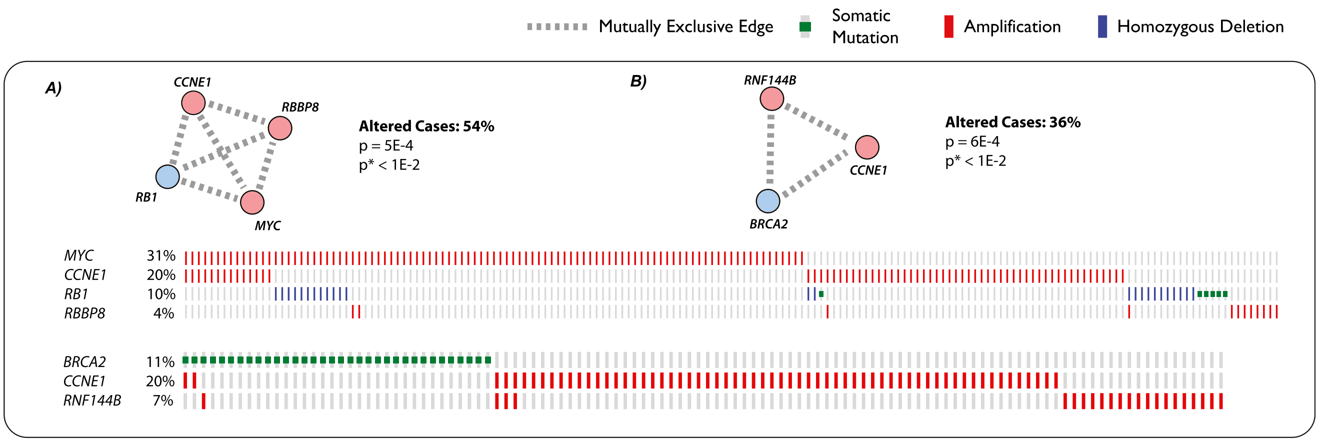

Figure 4. Significantly mutually exclusive modules in the TCGA serous ovarian cancer data set. (*A*) *CCNE1-RBBP8-RB1-MYC*; and (*B*) *RNF144B, CCNE1, BRCA2.*

316 ovarian cancer cases and observed worse outcome for *CCNE1* amplified cases ($P = 0.0718$, log-rank test).

However, when examining survival differences in *BRCA1/2* wild-type cases only, CCNE1 amplification is no longer associated with worse outcome ($P = 0.5$, log-rank test) (Fig. 3C).

To verify whether the previously reported worse outcome for *CCNE1* amplified cases is due to mutual exclusivity with *BRCA1/2* mutated cases or vice-versa, we performed survival analysis on the set of *CCNE1* diploid cases only, which shows that *BRCA1/2* mutated cases maintain longer overall survival ($P = 1.17 \times 10^{-4}$, log-rank test) (Fig. 3D). These results suggest that the previously reported survival difference in *CCNE1* amplified cases can be explained by the better survival of *BRCA1/2*-mutated cases, and that *BRCA1/2*-mutated cases remain a marker for good prognosis, independent of *CCNE1* alteration. The ability of MEMo to automatically identify mutually exclusive components may therefore provide a means of identifying other genes with such dependencies and survival characteristics in other cancer types in the future.

The second most significant module includes *RBBP8*, *CCNE1*, *RB1*, and *MYC* ($P^* < 1.0 \times 10^{-2}$) (Fig. 4A). This confirms some of the observations on correlated events originally reported by TCGA (The Cancer Genome Atlas Research Network 2011). In that earlier analysis, we compared all *BRCA1/2* alterations (including *BRCA1/2* germline and somatic mutations and *BRCA1* epigenetic silencing) to all GISTIC regions of interest and identified 19q12 (containing *CCNE1*) as the most significant mutually exclusive region ($P^* = 0.009$), and 8q24.21 (containing *MYC*) as the most significantly co-occurrent region ($P^* = 0.002$). Given the significant co-occurrence of mutations in *BRCA1* and *BRCA2* with amplification of *MYC*, it is not surprising that both the events are found to be mutually exclusive with the same set of genes (*CCNE1*, *RB1*, and *RBBP8*).

Finally, MEMo identifies a module including *CCNE1*, *BRCA2*, and *RNF144B* ($P^* < 1.0 \times 10^{-2}$) (Fig. 4B). *RNF144B* encodes for a ring finger protein and frequently amplified in the 6p22.3 locus. It negatively regulates p21 (*CDKN1A*) which itself directly inhibits several cyclin-dependent kinases, including *CCNE1* (Ng et al. 2003). This mechanism is consistent with the observed mutual exclusivity between *RNF144B* and *CCNE1* amplification.

Together, our findings in GBM and ovarian cancer show how MEMo is able to highlight the significant mutual exclusivity in alterations between functionally related genes, but furthermore it proved useful in formulating a new hypothesis and pointing to so far partially unexplored, yet potentially interesting, genes selectively altered in tumors.

Discussion

MEMo attempts to identify candidate driver networks in cancer by focusing on modules whose member genes exhibit a pattern of mutually exclusive genomic alterations across a set of patients. We have successfully applied the method to two cancer types with comprehensive genomic profiling data generated by TCGA. In the case of glioblastoma, the method successfully identifies several core modules involving p53, Rb, and PI(3)K signaling, providing an important validation of the method. In the case of ovarian cancer, the method identifies mutual exclusivity between BRCA genes and members of the Rb pathway, providing a new hypothesis for the pronounced genomic instability in ovarian tumors, and shedding new light on *CCNE1* as a marker for poor prognosis in ovarian cancer. Notably, mutual exclusivity within the BRCA/Rb module extends to *RBBP8*, shedding new light on the possible dual role of this protein in different phases of the cell cycle.

Other recent studies in cancer genomics have sought to identify co-occurring or mutually exclusive pairs of copy number events in ovarian cancer and GBM (Bredel et al. 2009; Gorringe et al. 2010). In contrast to these approaches, however, MEMo is unique in its ability to integrate copy number and mutation data, to assess sets of genes, rather than just pairs, and to automatically connect genomic alterations to prior known biological knowledge. Much like these previous studies, however, MEMo is reliant on large sample sizes to achieve statistical significance, and it is therefore most useful for large-scale cancer genomic projects, such as those at TCGA and ICGC. MEMo is also constrained by the prior biological network knowledge used to connect gene pairs. Importantly, it is able to connect genes which are proximal to one another in the network, but are not directly linked to one another, thereby enabling the discovery of biologically plausible novel interactions. Finally, we introduce here a simple but effective method to evaluate the statistical significance of correlations between genomic events, that concurrently preserves both tumor selectivity and tumor heterogeneity.

Potential areas of future work for the algorithm include integration of discretized methylation and mRNA expression data, analysis of concordantly altered genes within modules, and evaluation of other background HRNs, including mRNA co-expression networks. Furthermore, we plan to limit the dependency of MEMo to prior biological knowledge to be able to find novel associations between genes that are either proximally distant, or which have never been described in prior scientific literature. As the TCGA and ICGC projects expand to 20 or more cancer types, we believe MEMo will be an important method for integrating mutation and copy number data, and automatically identifying new candidate driver networks in diverse cancer types.

Methods

Genomic data

For glioblastoma multiforme (GBM) and ovarian cancer, the following genomic data were used as input:

1. All observed somatic mutations across all sequenced cases (germline mutations were included only for *BRCA1* and *BRCA2* and only for ovarian cancer);
2. a statistical assessment of all mutated genes as determined by the standard test of MutSig (Q-value ≤ 0.1); for GBM, we used MutSig results generated on a larger set of TCGA samples ($N = 339$); for ovarian cancer, we used MutSig results generated for $N = 316$ samples;
3. recurrently altered copy number region of interest (ROI) as determined by GISTIC;
4. discretized copy number calls (homozygously deleted, heterozygously deleted, diploid, gained, or amplified) for all genes in all samples as determined by GISTIC;
5. normalized mRNA expression data, derived from Affymetrix U133 and Agilent expression platforms, as described in The Cancer Genome Atlas Research Network (2008, 2011).

For each ROI, MEMo uses an additional filter to retain only the genes that are altered by high-level amplification, homozygous deletion, or mutation in at least 3% of samples. A one-tailed Wilcoxon signed-rank test is used to assess if mRNA expression values are significantly higher or lower in amplified or homozygously deleted samples versus diploid. Detailed descriptions of MutSig and GISTIC are reported in Getz et al. (2007) and Beroukhim et al. (2007), respectively.

Human reference network (HRN)

We used and evaluated two HRNs. The first network, HRN1, is derived from Cerami et al. (2010), and consists of four curated data sources: Reactome (Matthews et al. 2009), the MSKCC Cancer Cell Map (Memorial Sloan-Kettering Cancer Center 2007), the NCI/Nature Pathway Interaction Database (Schaefer et al. 2009), and the Human Protein Reference Database (HPRD) (Keshava Prasad et al. 2009). Pathway data sets for Reactome (Release 35, Dec 2010), NCI/Nature Pathway Interaction Database (Sept 2010 Release), and the MSKCC Cancer Cell Map (May 2006 Release) were downloaded in the Simple Interaction Format (SIF) from Pathway Commons (Cerami et al. 2011) on April 4, 2011. HPRD interactions (Release 9, April 2010) were obtained from the HPRD website (http://www.hprd.org/) on April 4, 2011. The second network, HRN2, is derived from Wu et al. (2010) and consists of manually curated interactions (Reactome, Matthews et al. 2009; Panther, Mi et al. 2010; KEGG, Kanehisa et al. 2008; and INOH, Fukuda 2008), plus additional inferred interactions derived from non-curated sources of information, including high-throughput derived protein–protein interactions, gene coexpression, protein domain interaction, GO annotations, and text-mined protein interactions (Wu et al. 2010). Within each HRN, all redundant edges are collapsed into single edges, and all self-directed edges are pruned from the network. After edge pruning, the final network for HRN1 consists of 9566 genes and 46,608 edges, and HRN2 consists of 9115 genes and 175,585 edges.

Similarity network and gene modules

Step 2 of the algorithm seeks to identify all gene pairs that are likely to be involved in the same cellular process or pathway. The HRNs are used to infer such likelihood. The underlying hypothesis—referred to here as the "local hypothesis" (Barabási et al. 2011)—is that two genes that are functionally similar are connected to a common set of shared interactors within the HRN. The same hypothesis has been explored and tested in several other types of networks, including large-scale social networks. Metrics have also been developed to quantify the likelihood that two entities in the network, typically individuals, belong to the same community, based on shared interests or neighbors (Liben-Nowell and Kleinberg 2003).

A widely adopted metric for this purpose is the Jaccard (J) coefficient (Liben-Nowell and Kleinberg 2003; Pradines et al. 2004). Given two nodes within the network, this metric uses information within the local neighborhood only and, thus, scales efficiently with large networks. Formally, the "similarity" between two nodes, u and v, is evaluated from the set of their direct neighbors, $N(u)$ and $\mathrm{N}(v)$ respectively, as follows:

$$J(u,v) = \frac{|N(u) \cap N(v)|}{|N(u) \cup N(v)|}$$

To account for the presence of a direct link between u and v, we corrected the Jaccard coefficient by using the "closed" neighborhoods of u and v, i.e., $N^*(u) = N(u) \cup \{u\}$ and $N^*(v) = N(v) \cup \{v\}$.

As a confirmation of the local hypothesis on our networks, we selected two sets of gene pairs: the first including pairs of genes known to belong to at least one common pathway (CP), the second including random pairs (RP). To build the first set, we used two subsets of canonical pathways from MSigDB (Subramanian et al. 2005): the first from BIOCARTA, the second from KEGG (signaling pathways only). None of these sources were used to build HRN1, thus they represent two independent testbeds for this network. For HRN2, we only used the set derived from BIOCARTA, as HRN2 includes annotations from KEGG.

We found that the modified Jaccard coefficient defined above is able to well discriminate between CP and RP in both networks: in HRN1, pairs in CP have on average $J_{avg}^{CP} = 0.04$ versus $J_{avg}^{RP} = 0.005$ in RP for BIOCARTA, and $J_{avg}^{CP} = 0.05$ versus $J_{avg}^{RP} = 0.001$ for KEGG; in HRN2 we obtain $J_{avg}^{CP} = 0.07$ versus $J_{avg}^{RP} = 0.009$ for BIOCARTA.

Using the Jaccard coefficient as a functional similarity metric, MEMo compares all genes within all significantly altered entities, where an entity can be a significantly mutated gene or a copy number ROI containing multiple genes, all of which have passed the mRNA concordance filter. All gene pairs which pass a fixed threshold for network proximity are connected with a single edge in a new global graph G. Given the observed distributions for the Jaccard coefficient within the sets CP and RP, we experimented with thresholds in the interval $0.02 \leq J \leq 0.03$, and selected 0.02 as the threshold for the results in this paper.

In Step 3, MEMo attempts to extract all groups of functionally related genes from the graph G. This is done by extracting all fully connected subgraphs of maximal size, i.e., "maximal cliques," from the similarity network built in Step 2. Clique extraction is a complex problem in graph theory, although given the typically small size of our networks (order of 10 to 10^2 nodes and interactions), simple heuristics proved to be efficient. Specifically, we adopted the heuristic originally proposed by Wernicke (2006) to extract all subgraphs of a given size from a network, and applied it to a method we developed to extract all maximal cliques independently of a size threshold.

MEMo then refines the list of cliques by pruning non-informative nodes. A node is said to be informative if the number of times the corresponding gene is altered concurrently with other genes in the clique is smaller than the number of unique alterations. Thus, we first associate to each clique the set of cases that harbor an alteration in at least one of the genes in the clique, then we select the set of informative genes in a greedy fashion, starting from the most frequently altered. Non-informative genes are removed from the clique.

Mutual exclusivity test

To assess the significance of the observed mutual exclusivity within the extracted modules, we define a null model where genomic alterations are randomly permuted. To do so, MEMo uses a constrained permutation procedure, referred to as the "switching permutation" method, which preserves the overall distribution of observed alterations across both genes and samples. The switching permutation procedure automatically preserves tumor selectivity in altering specific genes, e.g., frequent *EGFR* amplification in GBM or frequent *BRCA1*/*BRCA2* mutations in ovarian cancer, while concurrently preserving patient heterogeneity in mutation or copy number alteration rates. Alteration rates can vary widely within a study—for example, the TCGA GBM project has identified a subset of patients with a hypermutator phenotype (The Cancer Genome Atlas Research Network 2008)—and maintaining the alteration distributions associated with individual patients can be critical to evaluating the significance of alterations within a gene set (Boca et al. 2010).

The method starts from a simple observation: The binary event matrix M built in Step 1 can be thought of as the adjacency matrix of a bipartite graph, with two set of nodes, one representing genes (G) and the other representing samples (S). Each non-null entry of the matrix, $m_{ij} = 1$, represents an edge connecting gene $i \in G$ to sample $j \in S$. Let A denote the whole set of edges, i.e., of genomic alterations.

Given this network representation of observed genomic events, the distribution of alterations across genes (samples) is thus given by the degree distribution of nodes in G (S). The switching algorithm proposed in R Milo, N Kashtan, S Itzkovitz, MEJ Newman, and U Alon (http://arxiv.org/abs/cond-mat/0312028v2) uses a Markov chain and proceeds through a series of Monte Carlo switching steps to generate random networks starting from an observed network and preserving its degree distribution. The same method can be adapted to randomize bipartite graphs as follow:

1. Randomly select two edges (a,b) and (c,d), with $a,c \in G$ and $b,d \in S$.
2. If (a,b) and $(b,c) \notin A$, then remove the edges (a,b) and (c,d), and add (a,d) and (b,c); otherwise do not modify the network during this step.
3. Iterate Steps 1 and 2 for $Q|A|$ steps, where $|A|$ is the total number of edges, and Q is a constant.

Even though the exact value of Q that guarantees the mixing of the Markov chain is not known, in R Milo, N Kashtan, S Itzkovitz, MEJ Newman, and U Alon (http://arxiv.org/abs/cond-mat/0312028v2) the authors empirically show that $Q = 100$ is adequate.

MEMo applies the switching permutation method to M to generate a set of randomized alteration matrices $R_{1,\ldots,N}$. Given a module X, we calculate the total number of cases altered in at least one of the genes in X, first referring to observed alterations in M, and then to the those in R_i with $i \in \{1,\ldots,N\}$. An empirical P-value is derived for X as the fraction of permutations that lead to a greater or equal number of altered cases than those observed on real data, over a set of $N = 10{,}000$ random alteration matrices. A low P-value indicates that the observed alteration frequency is unlikely to occur by chance, and that genes within the set exhibit a greater than expected trend toward mutual exclusivity.

Recall from the previous section that we have defined modules in terms of maximal cliques. Even though a maximal clique is not significant, we cannot exclude that one of its sub-cliques actually is. Testing only maximal cliques may therefore result in missing potentially interesting mutually exclusive driver networks. To address this issue, the following heuristic is used to explore mutual exclusivity within sub-cliques, while simultaneously limiting the number of overall tests performed. Given a module X of size k genes, if we observe a significant trend in mutual exclusivity ($P <$ 0.05) exploration ends and sub-cliques are not tested. Otherwise, we select the sub-clique X' of size $k - 1$, contained in X, that is more likely to be significant among the set of sub-cliques of X with $k - 1$ genes. To select such a clique we refer once again to the definition of "informative genes": X' is obtained by removing from X the less informative gene, i.e., the one with the smallest number of unique alterations. The process is repeated recursively until either one of the two conditions is reached: X' is significantly mutually exclusive or k =2.

Finally, P-values for all modules are adjusted for multiple testing by applying the step down procedure proposed by Westfall and Young (1993). This step down procedure controls the false discovery rate (FDR) under general dependence of the data, as is the case for the set of possibly overlapping modules produced by MEMo.

Data access

MEMo software, source code, and sample data sets are available for download at: http://cbio.mskcc.org/memo. Core software is written in Java, and has been released under an LGPL Open Source license. Graph structures and network algorithms are implemented using the JUNG Graph Library 2.01 (http://jung.sourceforge.net/).

Acknowledgments

We thank the members of the TCGA Research Network and NIH staff for a generous collaborative enterprise and for profiling of the GBM and ovarian cancer data sets. We also wish to thank Emek Demir, Özgün Babur, Barry S. Taylor, Serena Bradde, and Douglas A. Levine for stimulating discussions regarding network analysis and cancer genomics. Finally, we wish to thank the anonymous reviewers for their useful insights and questions. Support for this work was provided by National Cancer Institute grants (1U24CA143840, MSKCC Center for Translational Cancer Genomic Analysis). Additional support was provided by the Tri-Institutional Training Program in Computational Biology and Medicine.

References

Barabási AL, Gulbahce N, Loscalzo J. 2011. Network medicine: A network-based approach to human disease. *Nat Rev Genet* **12:** 56–68.

Beroukhim R, Getz G, Nghiemphu L, Barretina J, Hsueh T, Linhart D, Vivanco I, Lee JC, Huang JH, Alexander S, et al. 2007. Assessing the significance of chromosomal aberrations in cancer: Methodology and application to glioma. *Proc Natl Acad Sci* **104:** 20007–20012.

Boca SM, Kinzler KW, Velculescu VE, Vogelstein B, Parmigiani G. 2010. Patient-oriented gene set analysis for cancer mutation data. *Genome Biol* **11:** R112. doi: 10.1186/gb-2010-11-11-r112.

Bredel M, Scholtens DM, Harsh GR, Bredel C, Chandler JP, Renfrow JJ, Yadav AK, Vogel H, Scheck AC, Tibshirani R, et al. 2009. A network model of a cooperative genetic landscape in brain tumors. *JAMA* **302:** 261–275.

The Cancer Genome Atlas Research Network. 2008. Comprehensive genomic characterization defines human glioblastoma genes and core pathways. *Nature* **455:** 1061–1068.

The Cancer Genome Atlas Research Network. 2011. Integrated genomic analyses of ovarian carcinoma. *Nature* **474:** 609–615.

Carter H, Chen S, Isik L, Tyekucheva S, Velculescu VE, Kinzler KW, Vogelstein B, Karchin R. 2009. Cancer-specific high-throughput annotation of somatic mutations: Computational prediction of driver missense mutations. *Cancer Res* **69:** 6660–6667.

Cerami E, Demir E, Schultz N, Taylor BS, Sander C. 2010. Automated network analysis identifies core pathways in glioblastoma. *PLoS One* **5:** e8918. doi: 10.1371/journal.pone.0008918.

Cerami EG, Gross BE, Demir E, Rodchenkov I, Babur O, Anwar N, Schultz N, Bader GD, Sander C. 2011. Pathway commons, a web resource for biological pathway data. *Nucleic Acids Res* **39:** D685–D690.

Chen PL, Liu F, Cai S, Lin X, Li A, Chen Y, Gu B, Lee EYHP, Lee WH. 2005. Inactivation of CtIP leads to early embryonic lethality mediated by G1 restraint and to tumorigenesis by haploid insufficiency. *Mol Cell Biol* **25:** 3535–3542.

Cowin PA, Anglesio M, Etemadmoghadam D, Bowtell DDL. 2010. Profiling the cancer genome. *Annu Rev Genomics Hum Genet* **11:** 133–159.

Etemadmoghadam D, deFazio A, Beroukhim R, Mermel C, George J, Getz G, Tothill R, Okamoto A, Raeder MB, Harnett P, et al. 2009. Integrated genome-wide DNA copy number and expression analysis identifies distinct mechanisms of primary chemoresistance in ovarian carcinomas. *Clin Cancer Res* **15:** 1417–1427.

Etemadmoghadam D, George J, Cowin PA, Cullinane C, Kansara M, Australian Ovarian Cancer Study Group, Gorringe KL, Smyth GK, Bowtell DDL. 2010. Amplicon-dependent *CCNE1* expression is critical for clonogenic survival after cisplatin treatment and is correlated with 20q11 gain in ovarian cancer. *PLoS ONE* **5:** e15498. doi: 10.1371/journal.pone.0015498.

Fukuda K. 2008. INOH pathway database: Curation, annotation, integration. *InterOntology08* **1:** 47–50.

Getz G, Hofling H, Mesirov JP, Golub TR, Meyerson M, Tibshirani R, Lander ES. 2007. Comment on "The consensus coding sequences of human breast and colorectal cancers". *Science* **317:** 1500. doi: 10.1126/science.1138764.

Gorringe KL, George J, Anglesio MS, Ramakrishna M, Etemadmoghadam D, Cowin P, Sridhar A, Williams LH, Boyle SE, Yanaihara N, et al. 2010. Copy number analysis identifies novel interactions between genomic loci in ovarian cancer. *PLoS ONE* **5**. doi: 10.1371/journal.pone.0011408.

Greenman C, Stephens P, Smith R, Dalgliesh GL, Hunter C, Bignell G, Davies H, Teague J, Butler A, Stevens C, et al. 2007. Patterns of somatic mutation in human cancer genomes. *Nature* **446:** 153–158.

Hanahan D, Weinberg RA. 2000. The hallmarks of cancer. *Cell* **100:** 57–70.

Hanahan D, Weinberg RA. 2011. Hallmarks of cancer: The next generation. *Cell* **144:** 646–674.

Hartwell LH, Hopfield JJ, Leibler S, Murray AW. 1999. From molecular to modular cell biology. *Nature* (Suppl) **402:** C47–C52.

Hwang HC, Clurman BE. 2005. Cyclin E in normal and neoplastic cell cycles. *Oncogene* **24:** 2776–2786.

Kaminker JS, Zhang Y, Watanabe C, Zhang Z. 2007. CanPredict: A computational tool for predicting cancer-associated missense mutations. *Nucleic Acids Res* **35:** W595–W598.

Kanehisa M, Araki M, Goto S, Hattori M, Hirakawa M, Itoh M, Katayama T, Kawashima S, Okuda S, Tokimatsu T, et al. 2008. KEGG for linking genomes to life and the environment. *Nucleic Acids Res* **36:** D480–D484.

Keshava Prasad TS, 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.

Liben-Nowell D, Kleinberg J. 2003. The link prediction problem for social networks. In *Proceedings of the Twelfth International Conference on Information and Knowledge Management*, pp. 556–559. New York, NY.

Liu F, Lee WH. 2006. CtIP activates its own and cyclin D1 promoters via the E2F/Rb pathway during G_1/S progression. *Mol Cell Biol* **26:** 3124–3134.

Matthews L, Gopinath G, Gillespie M, Caudy M, Croft D, de Bono B, Garapati P, Hemish J, Hermjakob H, Jassal B, et al. 2009. Reactome knowledgebase of human biological pathways and processes. *Nucleic Acids Res* **37:** D619–D622.

Memorial Sloan-Kettering Cancer Center. 2007. Cancer cell map. http://cancer.cellmap.org/.

Mi H, Dong Q, Muruganujan A, Gaudet P, Lewis S, Thomas PD. 2010. PANTHER version 7: Improved phylogenetic trees, orthologs and collaboration with the Gene Ontology Consortium. *Nucleic Acids Res* **38:** D204–D210.

Nakayama N, Nakayama K, Shamima Y, Ishikawa M, Katagiri A, Iida K, Miyazaki K. 2010. Gene amplification *CCNE1* is related to poor survival and potential therapeutic target in ovarian cancer. *Cancer* **116:** 2621–2634.

Ng CC, Arakawa H, Fukuda S, Kondoh H, Nakamura Y. 2003. p53RFP, a p53-inducible RING-finger protein, regulates the stability of p21WAF1. *Oncogene* **22:** 4449–4458.

Ohgaki H, Kleihues P. 2007. Genetic pathways to primary and secondary glioblastoma. *Am J Pathol* **170:** 1445–1453.

Oti M, Snel B, Huynen MA, Brunner HG. 2006. Predicting disease genes using protein-protein interactions. *J Med Genet* **43:** 691–698.

Pradines J, Rudolph-Owen L, Hunter J, Leroy P, Cary M, Coopersmith R, Dancik V, Eltsefon Y, Farutin V, Leroy C, et al. 2004. Detection of activity centers in cellular pathways using transcript profiling. *J Biopharm Stat* **14:** 701–721.

Rajagopalan H, Bardelli A, Lengauer C, Kinzler KW, Vogelstein B, Velculescu VE. 2002. Tumorigenesis: *RAF/RAS* oncogenes and mismatch-repair status. *Nature* **418:** 934.

Schaefer CF, Anthony K, Krupa S, Buchoff J, Day M, Hannay T, Buetow KH. 2009. PID: The Pathway Interaction Database. *Nucleic Acids Res* **37:** D674–D679.

Sherr CJ, McCormick F. 2002. The Rb and p53 pathways in cancer. *Cancer Cell* **2:** 103–112.

Sparks AB, Morin PJ, Vogelstein B, Kinzler KW. 1998. Mutational analysis of the APC/β-catenin/Tcf pathway in colorectal cancer. *Cancer Res* **58:** 1130–1134.

Stecca B, i Altalba AR. 2009. A *GLI1*-p53 inhibitory loop controls neural stem cell and tumour cell numbers. *EMBO J* **28:** 663–676.

Stratton MR, Campbell PJ, Futreal PA. 2009. The cancer genome. *Nature* **458:** 719–724.

Subramanian A, Tamayo P, Mootha VK, Mukherjee S, Ebert BL, Gillette MA, Paulovich A, Pomeroy SL, Golub TR, Lander ES, et al. 2005.Gene set enrichment analysis: A knowledge-based approach for interpreting genome-wide expression profiles. *Proc Natl Acad Sci* **102:** 15545–15550.

Taylor BS, Barretina J, Socci ND, Decarolis P, Ladanyi M, Meyerson M, Singer S, Sander C. 2008. Functional copy-number alterations in cancer. *PLoS ONE* **3:** e3179. doi: 10.1371/journal.pone.0003179.

Torkamani A, Schork NJ. 2009. Identification of rare cancer driver mutations by network reconstruction. *Genome Res* **19:** 1570–1578.

Turner N, Tutt A, Ashworth A. 2004. Hallmarks of "BRCAness" in sporadic cancers. *Nat Rev Cancer* **4:** 814–819.

Vandin F, Upfal E, Raphael BJ. 2010. Algorithms for detecting significantly mutated pathways in cancer. In *Proceedings of the 14th Annual International Conference on Research in Computational Molecular Biology (RECOMB)*. Lisbon, Portugal.

Vaske CJ, Benz SC, Sanborn JZ, Earl D, Szeto C, Zhu J, Haussler D, Stuart JM. 2010. Inference of patient-specific pathway activities from multi-dimensional cancer genomics data using paradigm. *Bioinformatics* **26:** i237–i245.

Velculescu VE. 2008. Defining the blueprint of the cancer genome. *Carcinogenesis* **29:** 1087–1091.

Wernicke S. 2006. Efficient detection of network motifs. *IEEE/ACM Trans Comput Biol Bioinform* **3:** 347–359.

Westfall PH, Young S. 1993. *Resampling-based multiple testing: Examples and methods for P-value adjustment*. Wiley, New York.

Wu G, Feng X, Stein L. 2010. A human functional protein interaction network and its application to cancer data analysis. *Genome Biol* **11:** R53. doi: 10.1186/gb-2010-11-5-r53.

Yu X, Chen J. 2004. DNA damage-induced cell cycle checkpoint control requires CtIP, a phosphorylation-dependent binding partner of *BRCA1* C-terminal domains. *Mol Cell Biol* **24:** 9478–9486.

Received April 29, 2011; accepted in revised form August 18, 2011.