Published in final edited form as:
Pac Symp Biocomput. 2015 ; 20: 84–95.

IDENTIFYING MUTATION SPECIFIC CANCER PATHWAYS USING A STRUCTURALLY RESOLVED PROTEIN INTERACTION NETWORK

H. BILLUR ENGIN,
School of Medicine, University of California San Diego, 9500 Gilman Dr. San Diego, CA 92093, USA

MATAN HOFREE, and
Department of Computer Science and Engineering, University of California San Diego, 9500 Gilman Dr. San Diego, CA 92093, USA

HANNAH CARTER*
School of Medicine, University of California San Diego, 9500 Gilman Dr. San Diego, CA 92093, USA

H. BILLUR ENGIN: hengin@ucsd.edu; MATAN HOFREE: mhofree@ucsd.edu; HANNAH CARTER: hkcarter@ucsd.edu

Abstract

Here we present a method for extracting candidate cancer pathways from tumor 'omics data while explicitly accounting for diverse consequences of mutations for protein interactions. Disease-causing mutations are frequently observed at either core or interface residues mediating protein interactions. Mutations at core residues frequently destabilize protein structure while mutations at interface residues can specifically affect the binding energies of protein-protein interactions. As a result, mutations in a protein may result in distinct interaction profiles and thus have different phenotypic consequences. We describe a protein structure-guided pipeline for extracting interacting protein sets specific to a particular mutation. Of 59 cancer genes with 3D co-complexed structures in the Protein Data Bank, 43 showed evidence of mutations with different functional consequences. Literature survey reciprocated functional predictions specific to distinct mutations on APC, ATRX, BRCA1, CBL and HRAS. Our analysis suggests that accounting for mutation-specific perturbations to cancer pathways will be essential for personalized cancer therapy.

1. Introduction

Cancer is a complex genetic disease in which the genomes of normal cells accumulate somatic mutations. A subset of these mutations confer neoplastic behaviors to cells through disregulation of a small number of common pathways[1]. Identifying the genes that participate

*This work is supported by NIH grant DP5 OD017937-01.

6. Supplementary Material
http://chianti.ucsd.edu/perm/hcarter/psb2015/Supplementary_Material.docx

in these pathways is an important objective in cancer genomics. However, linking somatically altered genes to perturbed pathways remains an open problem[2].

Individual proteins rarely mediate cellular behaviors; instead molecular machines comprising multiple proteins arbitrate various intracellular processes. As a result, proteins that interact physically within the cell are frequently involved in the same biological activities. This phenomenon, sometimes called guilt-by-association, has motivated the development of a variety of computational methods to identify disease-specific regions on the human Protein-Protein Interaction (PPI) network from molecular measurement data. Ideker *et al.*[3] integrate PPIs with mRNA expression data to detect differentially expressed sub-networks of genes, while, NIMMI[4] combines PPI networks with GWAS data to produce sub-networks that are functionally related and enriched for genetic variants linked with a trait. HotNet[5] maps mutation data onto PPI networks to identify sub-networks significantly enriched with cancer causing mutations, and NetBox[6] detects oncogenic network modules from DNA Copy Number Variants (CNVs), PPIs and signaling pathways. Additionally, Hofree *et al*.[7] combine patient mutation profiles, gene interaction networks and PPIs to find network regions that are specific to subtypes of cancer.

Missense mutations, a class of non-synonymous single nucleotide variants (nsSNVs), cause an amino acid substitution, resulting in a subtly different protein sequence. These sequence changes can alter protein structure. The resulting consequences for protein activity span the spectrum from neutral to completely disruptive. To date, methods for automated pathway extraction have treated missense mutations either as disruptive or neutral to protein activity, however it is well established that distinct amino acid sites within a protein mediate different functions. Simply modeling proteins as active or not may detract from the biological relevance of extracted pathways.

Recently, several groups have published high-resolution three-dimensional (3D) PPI networks[8–10] that include the molecular details of binding interfaces. Applications of these to investigate inherited disease mutations[11–15] have suggested that a) nsSNVs located at protein interfaces result in distinct phenotypes from those located in the protein core[9,10], b) known disease associated variants outside of the core are enriched at residues participating in protein interaction interfaces[10] c) in particular, in-frame disease mutations are enriched at interface regions of interacting proteins[9] and d) disease mutations at distinct interfaces of the same protein can be associated with distinct disease phenotypes[9]. In cancer, 3D location of mutations at an interface has served as evidence that protein interactions may be important for metastasis site determination.[11] These observations suggest that distinct changes to the network topology of protein interaction networks will result in different phenotypes. Thus efforts to identify disease-associated pathways may need to account for mutation-specific effects to the PPI network.

Here, we investigate the extent to which distinct somatic mutations observed in known cancer genes have distinct phenotypic consequences. We present a structure-guided sub-network extraction pipeline (Figure 1) that identifies protein sets associated with specific missense mutations. We divide mutations observed in tumor exome sequencing data from The Cancer Genome Atlas (TCGA) into two categories: core and interface, then use

structurally resolved protein interaction data to model the effects of mutations on PPI network topology. Then using a diffusion-based approach, we identify distinct sets of interacting proteins in the global interaction network associated with different residue alterations of the same cancer gene and show that in many cases, these protein sets are implicated in distinct biological activities.

2. Materials and Methods

2.1 Sources of Protein Interaction Data

We assembled a highly reliable set of human PPIs from STRING[16], HINT[17] and the Protein Data Bank (PDB)[18] (Figure 1a). The respective contributions were 57985 interactions among 10211 proteins from STRING v.9.1 with experimental support and a confidence score higher than 0.4, 24523 interactions involving 8126 proteins from HINT, and 11583 physical interactions between 1653 proteins from the PDB. After eliminating self-interactions, our network comprises of 74699 interactions among 11951 proteins.

2.2 Cancer Genes

We investigated mutations in 125 genes implicated by Vogelstein *et al*.'s[1] as driving tumorigenesis. Of these genes, 123 were present in our human PPI, participating in 7503 interactions. Ninety-seven of the encoded proteins had structural entries in the PDB, and 59 had PDB structures in complex with one or more binding partners, resulting in a total of 169 structurally resolved interaction interfaces (Figure 1a).

2.3 Source of Mutation Data

The TCGA mutation data (merged results of MutSig v2.0 and MutSigCV v0.9) was downloaded from the 01/15/2014 Firehose release (http://gdac.broadinstitue.org). Only missense mutations were considered in this analysis.

2.4 Mapping Mutations to Protein Structure

2.4.1 From DNA Sequence Position to Structural Position—In order to determine the three-dimensional location of mutated residues, chromosomal coordinates of nsSNVs were mapped onto to PDB coordinates. Chromosomal coordinates were mapped to transcripts in Gencode19[19] using psl format files downloaded from the UCSC Genome Browser[20]. UniProt proteins were aligned to transcripts in Gencode19 using tblastn[21] software. Lastly, we performed the Uniprot to PDB mapping with the PDBSWS[22] server.

2.4.2 Designating Interface and Core Residues—We classified nsSNVs into two groups depending on their structural location: core or interface. To designate a residue as participating in a protein interaction interface, we used the consensus of interface predictions made by the HotPoint[23] and KFC2[24] servers to identify residues in physical contact. We removed incomplete interfaces by discarding interactions with fewer than 5 residues in at least one of the interacting chains. We used NACCESS[25] to calculate the accessible surface area (ASA) of all protein residues. Residues with an ASA of 0 were classified as core.

2.4.3 Positioning Mutation Data on Protein Structure—We obtained core residue positions for 97 of the proteins encoded by cancer genes and positions at interfaces for 169 interactions between these proteins and their partners. In total there were 398 mutations located at either core or interface residues of cancer genes (Figure 1b). We made the simplifying assumption that mutations mapping to the same interface are likely to affect it in the same way. Thus we select a single representative mutated residue for each interface. Because we are interested in differential functional consequences of mutations in the same protein, we focused on genes with at least 2 mutations in different locations (i.e. different interfaces or core and interface mutations). This reduced our list to 137 distinct events in 43 cancer genes (Supplementary Table 1).

2.5 Network Perturbation

If a mutation occurs in the protein core we assume that all of the protein's interactions are affected (Figure 2b), and if it occurs at an interface, only the interactions mediated by the interface are affected (Figure 2c). To implement these perturbations, we removed edges from our structurally resolved PPI network corresponding to the affected interactions (Figure 1c).

2.6 Network Propagation Algorithm

We used network propagation[26] to implicate protein sets most likely to be affected by each mutation (Figure 1d). This method has been applied to the related problem of clustering of patients based on somatic mutation profiles by Hofree *et al.*[7], and uses a random walk (with restarts) according to the function in Eq.(1).

$$F_{t+1}=\alpha F_t A+(1-\alpha)F_0 \quad (1)$$

$\boldsymbol{F_0}$ is a binary vector with size equal to the number of proteins in the network. Mutated cancer genes are set to 1, representing 'heat sources', while other proteins are initialized to 0. The $\boldsymbol{A}$ matrix is the degree normalized adjacency matrix of the PPI network. The $\boldsymbol{\alpha}$ parameter affects the distance that the heat signal propagates during the diffusion. The distribution of the propagated values was similar for different $\boldsymbol{\alpha}$ values and the choice of this parameter had limited impact on the results within the range of [0.4–0.7], as was previously reported.[26] We used 0.4 as the $\boldsymbol{\alpha}$ parameter.

In order to avoid numerical inaccuracy issues, the propagation algorithm is solved by iterative use of equation (1) until convergence (i.e. the sum of absolute differences between elements of $\boldsymbol{F_{t+1}}$ and $\boldsymbol{F_t}$ is smaller than $\mathbf{10}^{-6}$). The algorithm returns $\boldsymbol{F_t}$, which contains a value for each node in the network proportional to the expected number of times the node is visited during a random walk originating from the heat source, and restarting at the heat source with probability $\boldsymbol{\alpha}$.

2.7 The Differential Heat Profiles

We performed network propagation for each of the 137 representative mutations separately. For each mutation, we calculated $\boldsymbol{F_t}$ vectors for the unaltered network and the perturbed

network. For subsequent analysis steps (protein module detection and functional annotation), we used the differential heat profiles, obtained by subtracting the $\mathbf{F_t}$ values for each gene in the unaltered and perturbed networks.

As methods used in this analysis are sensitive to differences in scale differential heat profiles were aggregated into a mutation x gene matrix and quantile normalized using the "preprocessCore" package of Bioconductor[27] for R[28].

2.8 Sub-network Extraction

We used an approach similar to that used by the HotNet[5] method to identify altered sub-networks in our global PPI from the differential heat profiles for the 137 mutations (Figure 1e). First, each edge was assigned the minimum heat value of the corresponding protein pair. Edges were then sorted by heat value and the top 10^{th} percentile of edges were extracted. Next, we executed our pipeline for 1000 random mutations with similar consequences to those observed in the TCGA data (390 core and 610 interface affecting 1–10 edges). We removed edges that had differential heat scores in the top 10^{th} percentile in over 5% of the random runs as these edges likely resulted from the underlying topology of our PPI network rather than the perturbation of interest. This procedure resulted in a set of connected components for each of the 137 mutations, representing mutation-specific candidate cancer pathway genes.

2.9 Functional Annotation

We used David[29] to annotate the gene sets in the mutation-specific connected components from the GO Biological Process data set[30]. For each cancer gene, functional annotations were divided into those common to all mutations and those specific to particular mutations (Figure 1f).

3. Results and Discussions

3.1 A Pipeline to Extract Mutation-Specific Pathways

We constructed a pipeline (Figure 1) for mining and annotating cancer related protein sets from somatic mutation data while accounting for mutation-specific network perturbations. We applied this pipeline to analyze mutations observed in 125 frequently mutated cancer genes, where the vast majority of observed mutations are likely to be cancer causing driver mutations. Our pipeline can be applied to mutations in any gene, however for genes not known to drive tumorigenesis, efforts should be made to discriminate between causal driver and non-causal passenger events.

3.2 Mutational Distribution in Cancer Genes

We investigated the spatial distribution of somatic missense mutations on 125 cancer genes using individual crystal structures and co-crystallization of the encoded proteins with interaction partners. We then incorporated these structural data into a larger network of experimentally validated PPIs. We note that because these 125 genes are well studied, there is a positive bias towards data availability relative to other genes in the human interactome. Nonetheless, only 59 of the 125 proteins had structural complexes in the PDB that included

an interaction partner. Despite this, co-crystallization of the 59 driver genes with interaction partners covered fully 6% of the experimentally validated protein interactions (Supplementary Table 2).

In order to assess the extent of structural diversity of missense mutations observed across known cancer genes, we mapped mutations to core and interface regions. We observed that cancer gene encoded proteins for which co-crystallization structures are available harbor mutations at an average of 2.5 distinct sites (Supplementary Table 2). Of core and interface sites observed to harbor mutations, 21% demonstrated tissue specificity (Fisher's Exact test, Bonferroni corrected p –value < 0.05) (Supplementary Table 3). In particular, 4 cancer genes showed significant differences in mutation counts at distinct sites in different cancer types. These observations suggest that the physical location of mutations in known cancer genes may have functional significance.

3.3 Determining the Altered Sub-networks and Their Functionality

To identify perturbed network modules in the global network, we applied a HotNet-like method (section 2.8) to the differential heat profiles obtained for missense mutations at protein core or interface residues. We focused on cancer genes with mutations mapping to multiple distinct locations likely to have different functional consequences. Filtering redundant mutations (those occurring at residues in the same protein core or interface), we retained 137 mutations for 43 cancer genes. These 137 events returned an average of 56 altered sub-networks derived from an average of 686 proteins (Supplementary Table 4). We annotated the resulting sub-networks from the GO Biological Process database, and found that all 43 cancer genes harbored events that implicated specific functional consequences. Published events were consistent with our functional annotations for sites in APC, ATRX, BRCA1, CBL and HRAS via literature search (**Section 4** and Supplementary Table 5). For this purpose we assumed mutagenesis experiments reported for other residues at the same interface or core would be equivalent to the events we modeled.

3.4 HRAS Case Study: Implicated Sub-Networks and Functions

The RAS family oncogenes, KRAS, HRAS and NRAS were among the first discovered oncogenes, and are frequently mutated across a variety of human cancers. These genes regulate cell proliferation, differentiation and survival[31] via interaction with a number of different protein targets. Amino acid substitutions occurring on these 3 genes disturb signaling through these pathways and lead to tumorigenesis.

In our current network, we have structurally resolved interfaces for HRAS binding to RASA1 and SOS1, but not for KRAS and NRAS. Given the high degree of similarity among RAS proteins, and various experimental findings that support similar functional capabilities[32], the model we present here for HRAS likely generalizes to KRAS and NRAS as well.

Among TCGA patients we observe four amino acid substitutions localizing to protein-binding interfaces on RAS proteins (G12 in 6 patients, G13 in 8 patients, A59 in 1 patient and Q61 in 21 patients) and no mutations affecting the protein core (Figure 3). By

superimposing the RASA1 – HRAS (PDB ID: 1WQ1) and SOS1 – HRAS (PDB ID: 1BKD) complexes, we observed that these two interactors utilize the same binding site on HRAS. Physical distance between residues in co-crystalized structures implicated three residues in interactions with RASA1 (residues 12, 13 and 61) and SOS1 (residues 13, 59 and 61) respectively. This suggests that neighboring residues 12–13 and 59–61 participate in different interactions. Mutations at residues 12 and 13 have been observed to have prognostic and therapeutic differences. For example, KRAS G13 mutated colorectal cancers show some response to cetuximab, while G12 mutated cancers do not respond or may even progress more rapidly[33]. We applied our pipeline to each mutated interface, resulting in predictions for mutations altering signaling through RASA1(HRAS G12), signaling through SOS1 (HRAS A59) or both simultaneously(HRAS G13/Q61).

3.4.1 HRAS G12 alterations—G12 mutations in HRAS returned protein modules involved in GTPase activity, cytokine production, vasculogenesis, blood coagulation, endothelial cell differentiation/proliferation and smooth muscle cell proliferation/migration (Supplementary Table 6). Evidence suggests that these functions may be linked. Inappropriate blood coagulation is frequently observed in cancer patients and is closely related to tumor growth[34]. Vasculogenesis, which involves proliferation, migration and remodeling of endothelial cells, have been found to be related with tumor recurrence[35]. Chemokines play an important role in the behavior of endothelial cells during vessel formation[36]. Vascular smooth muscle cells provide homeostatic control and protect newly formed vessels against rupture and regression via inhibition of endothelial cell proliferation and migration[37]. Oncogenic HRAS G12 is known to stimulate chemokine secretion[38], cause VGEF activation and endothelial cell apoptosis[39], and is thought to be essential for solid tumor maintanence[37]. Furthermore, VEGF (Entrez Gene ID: 7422), one of the prominent molecules that control vasculogenesis, is present in the protein set (Supplementary Table 7) implicated by the HRAS G12 mutation. Amino acid changes at HRAS G12 have been shown to affect the strength of GTPase activity and binding to GTP[40].

3.4.2 HRAS G13/Q61 alterations—Residues G13 and Q61, which participate in a common binding site on HRAS, mediate interactions with both RASA1 and SOS1 and are frequently mutated in cancer. GO analysis of the protein modules implicated by these residues found functional enrichment for response to UV light (Supplementary Table 8–9). Even though there is an extensive body of evidence supporting the connection between UV radiation and melanoma[41–43], the exact mechanism remains unclear. Yang *et al*[44] proposed a possible mechanism that drives melanoma photocarcinogenesis through KRAS Q61 mutagenesis. Besides, its shown that UV-radiation has a bias towards targeting pyrimidine dimers that more frequently lead to RAS Q61 mutations[45].

3.4.3 HRAS A59 alterations—Mutations at HRAS A59 exclusively affect the SOS1 interaction. When analyzed with our model, the implicated proteins were involved in GPCR signaling. (Supplementary Table 10–11). RAS activation is catalyzed by guanine nucleotide exchange factors (GEFs) which include GPCRs. On the other hand SOS1[46] acts as a GEF or RAS. SOS1 forms a complex with GRB2 that is obligatory for GPCR-mediated RAS activation. Since these proteins are tightly bound to each other, when the interaction between

SOS1 and HRAS is hindered, it is not unexpected to see GPCR signaling as an altered pathway.

4. Additional Validation

We assessed two mutated residues affecting interactions between APC and KHDRBS1 (res640) and CTNNB1 (res1527) respectively. Consistent with the published finding that the R640G of APC causes exon 14 skipping by disrupting ASF/SF2 binding[47], our unctional annotations included mRNA splice related activities and the implicated protein set included ASF/SF2. In contrast, functional annotations specific to the res1527 perturbation included a number of nervous system related activities, such as “neuron apoptosis” and “negative regulation of neuron differentiation”. In the literature, mutations at codon 1495 which is also at the CTNNB1 interface have been observed in Medulloblastoma[48].

Mutated residues at positions 220 and 263 map to the core of ATRX and an interface that mediates binding to members of the Histone H3 family, H3F3A and HIST1H3A, respectively. ATRX has been implicated in chromatin remodeling and regulation of gene expression. The ATRX ADD domain and HP1 are required for ATRX localization to heterochromatin. Mutation E218A reduces pericentromeric localization of ATRX without disturbing the stability of the ADD domain[49]. H3 tails bearing tri-methylated Lys9 (H3-K9) are also required for ATRX localization via the ADD domain[50]. Annotations for residue 263 were specific to histone H4-K acetylation, while the ATRX core mutations, which presumably destabilize the protein, returned histone H3-K9 methylation. This is consistent with the ADD domain being unaffected by the interface mutations.

We evaluated residues affecting BRCA1’s interactions with BRIP1 (res1813,res1699) and BARD1 (res96). The BRCA1–BRIP1–TOPBP1 complex is associated with DNA repair during replication and is essential for the S-phase checkpoint in response to collapsed replication forks.[51] The mutated residues affecting BRCA1 and BRIP1 interactions returned related terms including “DNA repair”, “maintenance of fidelity during DNA-dependent DNA replication”, and “DNA replication checkpoint and replication fork protection”. Weakening of the BRCA1–BARD1 interaction due to ionizing radiation leads to the induction of p21 and initiation of the G1/S checkpoint[52]. Our annotations for this included “DNA damage response, signal transduction by p53 class mediator resulting in transcription of p21 class mediator”. In addition, of 10 BRCA1 mutations found to be functional by Carvalho *et al.*[53], 7 mapped to interface or core residues with our pipeline.

CBL’s interfaces with UBE2D2 (res418, res417, res384) and EGFR (res322) were observed to harbor mutations in TCGA samples. CBL residue R420 is involved in ubiquitination[54] and the G298E mutation was shown to abolish NFAT activation[55]. Consistent with these findings, our annotations specific to the UBE2D2 interface included terms related to unbiquitination, while annotations specific to the EGFR interaction included T-cell activation and, NFAT activation molecule 1 was present in the protein set.

5. Conclusions

We describe here our efforts to incorporate information about the differential consequences of somatic mutations in the same protein for extracting cancer pathways from large tumor 'omics data sets. Although there is limited structural knowledge available for PPIs, what exists provides strong evidence for specific functional consequences of mutations at distinct sites within the same protein. Among 59 proteins with sufficient structural data, 43 had mutations with specific functional annotations. Despite the paucity of functionally characterized missense mutations in databases and literature, we were able to find supporting evidence in the literature for mutated sites on 6 genes of the 43 genes. A case study investigating the mutation consequences for distinct interactions of HRAS further highlights that biological processes associated with each event can be specific and may have phenotypic relevance to the patient. For more systematic validation, experimental assays could be designed to validate predictions of our method, guided by the implicated protein sub-networks and the associated annotations. In aggregate, our findings suggest that perturbation to cancer pathways may in fact be mutation-specific and point to the need for analysis methods aware of tumor-specific network topologies.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

References

1. Vogelstein B, et al. Cancer genome landscapes. Science. 2013; 339:1546–1558.10.1126/science.1235122 [PubMed: 23539594]
2. Garraway LA, Lander ES. Lessons from the cancer genome. Cell. 2013; 153:17–37.10.1016/j.cell.2013.03.002 [PubMed: 23540688]
3. Ideker T, Ozier O, Schwikowski B, Siegel AF. Discovering regulatory and signalling circuits in molecular interaction networks. Bioinformatics. 2002; 18(Suppl 1):S233–240. [PubMed: 12169552]
4. Akula N, et al. A network-based approach to prioritize results from genome-wide association studies. PloS one. 2011; 6:e24220.10.1371/journal.pone.0024220 [PubMed: 21915301]
5. Vandin F, Upfal E, Raphael BJ. Algorithms for detecting significantly mutated pathways in cancer. Journal of computational biology: a journal of computational molecular cell biology. 2011; 18:507–522.10.1089/cmb.2010.0265 [PubMed: 21385051]
6. Cerami E, Demir E, Schultz N, Taylor BS, Sander C. Automated network analysis identifies core pathways in glioblastoma. PloS one. 2010; 5:e8918.10.1371/journal.pone.0008918 [PubMed: 20169195]
7. Hofree M, Shen JP, Carter H, Gross A, Ideker T. Network-based stratification of tumor mutations. Nature methods. 2013; 10:1108–1115.10.1038/nmeth.2651 [PubMed: 24037242]
8. Mosca R, Ceol A, Aloy P. Interactome3D: adding structural details to protein networks. Nature methods. 2013; 10:47–53.10.1038/nmeth.2289 [PubMed: 23399932]
9. Wang X, et al. Three-dimensional reconstruction of protein networks provides insight into human genetic disease. Nature biotechnology. 2012; 30:159–164.10.1038/nbt.2106
10. David A, Razali R, Wass MN, Sternberg MJ. Protein-protein interaction sites are hot spots for disease-associated nonsynonymous SNPs. Human mutation. 2012; 33:359–363.10.1002/humu.21656 [PubMed: 22072597]
11. Engin HB, Guney E, Keskin O, Oliva B, Gursoy A. Integrating structure to proteinprotein interaction networks that drive metastasis to brain and lung in breast cancer. PloS one. 2013; 8:e81035.10.1371/journal.pone.0081035 [PubMed: 24278371]

12. Acuner-Ozbabacan SE, et al. The structural network of Interleukin-10 and its implications in inflammation and cancer. BMC Genomics. 2014; 15:S2. [PubMed: 25056661]
13. Acuner Ozbabacan SE, Gursoy A, Nussinov R, Keskin O. The structural pathway of interleukin 1 (IL-1) initiated signaling reveals mechanisms of oncogenic mutations and SNPs in inflammation and cancer. PLoS computational biology. 2014; 10:e1003470.10.1371/journal.pcbi.1003470 [PubMed: 24550720]
14. Zhong Q, et al. Edgetic perturbation models of human inherited disorders. Molecular systems biology. 2009; 5:321.10.1038/msb.2009.80 [PubMed: 19888216]
15. Schuster-Bockler B, Bateman A. Protein interactions in human genetic diseases. Genome biology. 2008; 9:R9.10.1186/gb-2008-9-1-r9 [PubMed: 18199329]
16. Franceschini A, et al. STRING v9.1: protein-protein interaction networks, with increased coverage and integration. Nucleic acids research. 2013; 41:D808–815.10.1093/nar/gks1094 [PubMed: 23203871]
17. Das J, Yu H. HINT: High-quality protein interactomes and their applications in understanding human disease. BMC systems biology. 2012; 6:92.10.1186/1752-0509-6-92 [PubMed: 22846459]
18. Bernstein FC, et al. The Protein Data Bank. A computer-based archival file for macromolecular structures. European journal of biochemistry/FEBS. 1977; 80:319–324. [PubMed: 923582]
19. Harrow J, et al. GENCODE: the reference human genome annotation for The ENCODE Project. Genome research. 2012; 22:1760–1774.10.1101/gr.135350.111 [PubMed: 22955987]
20. Karolchik D, et al. The UCSC Genome Browser Database. Nucleic acids research. 2003; 31:51–54. [PubMed: 12519945]
21. Altschul SF, et al. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. Nucleic acids research. 1997; 25:3389–3402. [PubMed: 9254694]
22. Martin AC. Mapping PDB chains to UniProtKB entries. Bioinformatics. 2005; 21:4297–4301.10.1093/bioinformatics/bti694 [PubMed: 16188924]
23. Tuncbag N, Keskin O, Gursoy A. HotPoint: hot spot prediction server for protein interfaces. Nucleic acids research. 2010; 38:W402–406.10.1093/nar/gkq323 [PubMed: 20444871]
24. Zhu X, Mitchell JC. KFC2: a knowledge-based hot spot prediction method based on interface solvation, atomic density, and plasticity features. Proteins. 2011; 79:2671–2683.10.1002/prot.23094 [PubMed: 21735484]
25. Hubbard, SJ.; T, J. NACCESS. Department of Biochemistry and Molecular Biology, University College London; 1993.
26. Vanunu O, Magger O, Ruppin E, Shlomi T, Sharan R. Associating genes and protein complexes with disease via network propagation. PLoS computational biology. 2010; 6:e1000641.10.1371/journal.pcbi.1000641 [PubMed: 20090828]
27. Gentleman RC, et al. Bioconductor: open software development for computational biology and bioinformatics. Genome biology. 2004; 5:R80.10.1186/gb-2004-5-10-r80 [PubMed: 15461798]
28. R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing; Vienna, Austria: 2014.
29. Huang da W, Sherman BT, Lempicki RA. Bioinformatics enrichment tools: paths toward the comprehensive functional analysis of large gene lists. Nucleic acids research. 2009; 37:1–13.10.1093/nar/gkn923 [PubMed: 19033363]
30. Ashburner M, et al. Gene ontology: tool for the unification of biology. The Gene Ontology Consortium. Nature genetics. 2000; 25:25–29.10.1038/75556 [PubMed: 10802651]
31. Fernandez-Medarde A, Santos E. Ras in cancer and developmental diseases. Genes & cancer. 2011; 2:344–358.10.1177/1947601911411084 [PubMed: 21779504]
32. Potenza N, et al. Replacement of K-Ras with H-Ras supports normal embryonic development despite inducing cardiovascular pathology in adult mice. EMBO reports. 2005; 6:432–437.10.1038/sj.embor.7400397 [PubMed: 15864294]
33. De Roock W, et al. Effects of KRAS, BRAF, NRAS, and PIK3CA mutations on the efficacy of cetuximab plus chemotherapy in chemotherapy-refractory metastatic colorectal cancer: a retrospective consortium analysis. The lancet oncology. 2010; 11:753–762.10.1016/S1470-2045(10)70130-3 [PubMed: 20619739]

34. Nash GF, Walsh DC, Kakkar AK. The role of the coagulation system in tumour angiogenesis. The lancet oncology. 2001; 2:608–613. [PubMed: 11902551]
35. Kioi M, et al. Inhibition of vasculogenesis, but not angiogenesis, prevents the recurrence of glioblastoma after irradiation in mice. The Journal of clinical investigation. 2010; 120:694–705.10.1172/JCI40283 [PubMed: 20179352]
36. Gupta SK, Lysko PG, Pillarisetti K, Ohlstein E, Stadel JM. Chemokine receptors in human endothelial cells. Functional expression of CXCR4 and its transcriptional regulation by inflammatory cytokines. The Journal of biological chemistry. 1998; 273:4282–4287. [PubMed: 9461627]
37. Carmeliet P. Mechanisms of angiogenesis and arteriogenesis. Nature medicine. 2000; 6:389–395.10.1038/74651
38. O'Hayer KM, Brady DC, Counter CM. ELR+ CXC chemokines and oncogenic Rasmediated tumorigenesis. Carcinogenesis. 2009; 30:1841–1847.10.1093/carcin/bgp198 [PubMed: 19805574]
39. Chin L, et al. Essential role for oncogenic Ras in tumour maintenance. Nature. 1999; 400:468–472.10.1038/22788 [PubMed: 10440378]
40. Pylayeva-Gupta Y, Grabocka E, Bar-Sagi D. RAS oncogenes: weaving a tumorigenic web. Nature reviews Cancer. 2011; 11:761–774.10.1038/nrc3106
41. Tronnier M, Smolle J, Wolff HH. Ultraviolet irradiation induces acute changes in melanocytic nevi. The Journal of investigative dermatology. 1995; 104:475–478. [PubMed: 7706761]
42. Zaidi MR, et al. Interferon-gamma links ultraviolet radiation to melanomagenesis in mice. Nature. 2011; 469:548–553.10.1038/nature09666 [PubMed: 21248750]
43. Bald T, et al. Ultraviolet-radiation-induced inflammation promotes angiotropism and metastasis in melanoma. Nature. 2014; 507:109–113.10.1038/nature13111 [PubMed: 24572365]
44. Yang G, Curley D, Bosenberg MW, Tsao H. Loss of xeroderma pigmentosum C (Xpc) enhances melanoma photocarcinogenesis in Ink4a-Arf-deficient mice. Cancer research. 2007; 67:5649–5657.10.1158/0008-5472.CAN-06-3806 [PubMed: 17575131]
45. Tormanen VT, Pfeifer GP. Mapping of UV photoproducts within ras proto-oncogenes in UV-irradiated cells: correlation with mutations in human skin cancer. Oncogene. 1992; 7:1729–1736. [PubMed: 1501884]
46. Buday L, Downward J. Many faces of Ras activation. Biochimica et biophysica acta. 2008; 1786:178–187.10.1016/j.bbcan.2008.05.001 [PubMed: 18541156]
47. Goncalves V, et al. A missense mutation in the APC tumor suppressor gene disrupts an ASF/SF2 splicing enhancer motif and causes pathogenic skipping of exon 14. Mutation research. 2009; 662:33–36.10.1016/j.mrfmmm.2008.12.001 [PubMed: 19111562]
48. Huang H, et al. APC mutations in sporadic medulloblastomas. The American journal of pathology. 2000; 156:433–437.10.1016/S0002-9440(10)64747-5 [PubMed: 10666372]
49. Eustermann S, et al. Combinatorial readout of histone H3 modifications specifies localization of ATRX to heterochromatin. Nature structural & molecular biology. 2011; 18:777–782.10.1038/nsmb.2070
50. Kourmouli N, Sun YM, van der Sar S, Singh PB, Brown JP. Epigenetic regulation of mammalian pericentric heterochromatin in vivo by HP1. Biochemical and biophysical research communications. 2005; 337:901–907.10.1016/j.bbrc.2005.09.132 [PubMed: 16213461]
51. Levran O, et al. The BRCA1-interacting helicase BRIP1 is deficient in Fanconi anemia. Nature genetics. 2005; 37:931–933.10.1038/ng1624 [PubMed: 16116424]
52. Roy R, Chun J, Powell SN. BRCA1 and BRCA2: different roles in a common pathway of genome protection. Nature reviews Cancer. 2012; 12:68–78.10.1038/nrc3181
53. Carvalho MA, et al. Determination of cancer risk associated with germ line BRCA1 missense variants by functional analysis. Cancer research. 2007; 67:1494–1501.10.1158/0008-5472.CAN-06-3297 [PubMed: 17308087]
54. Makishima H, et al. CBL mutation-related patterns of phosphorylation and sensitivity to tyrosine kinase inhibitors. Leukemia. 2012; 26:1547–1554.10.1038/leu.2012.7 [PubMed: 22246246]
55. Zhang Z, Elly C, Qiu L, Altman A, Liu YC. A direct interaction between the adaptor protein Cbl-b and the kinase zap-70 induces a positive signal in T cells. Current biology: CB. 1999; 9:203–206. [PubMed: 10074432]

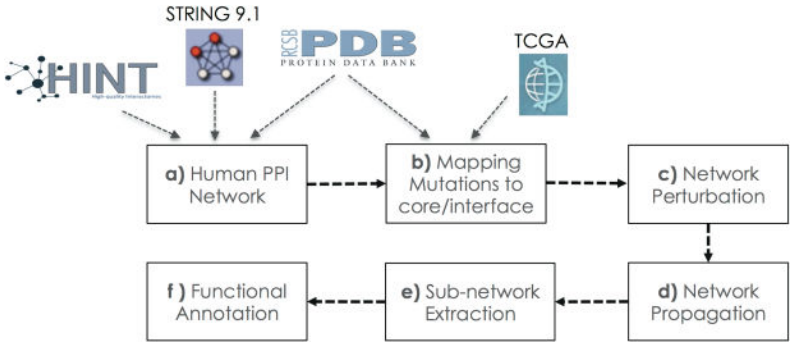

Figure 1.

A pipeline for mining molecular cancer related sub-networks accounting for different effects of distinct mutations. Steps include network assembly (a), mapping of mutations to interface versus core residues of cancer genes (b), removal of affected edges (c), extraction of associated protein sets (d & e) and functional annotation (f).

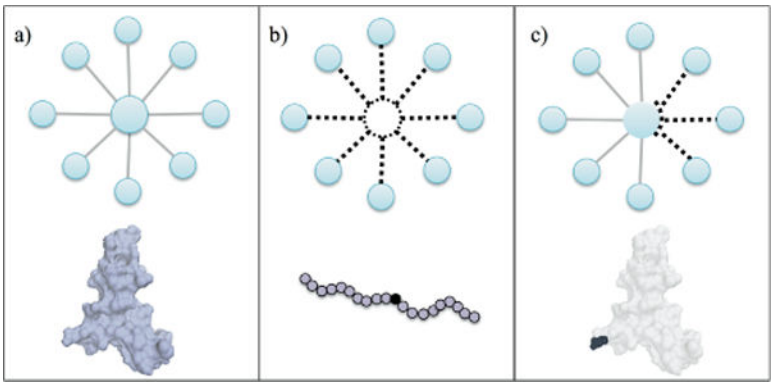

Figure 2.

Modeling mutations as network perturbations. a) The unaltered protein-protein interactions of a wild type protein, b) a core mutation has the tendency to destabilize the protein. We depict this phenomenon by removing all edges involving the protein c) an interface mutation may affect some of the interactions of a protein. In this case we remove the potentially affected edges of the protein from the network.

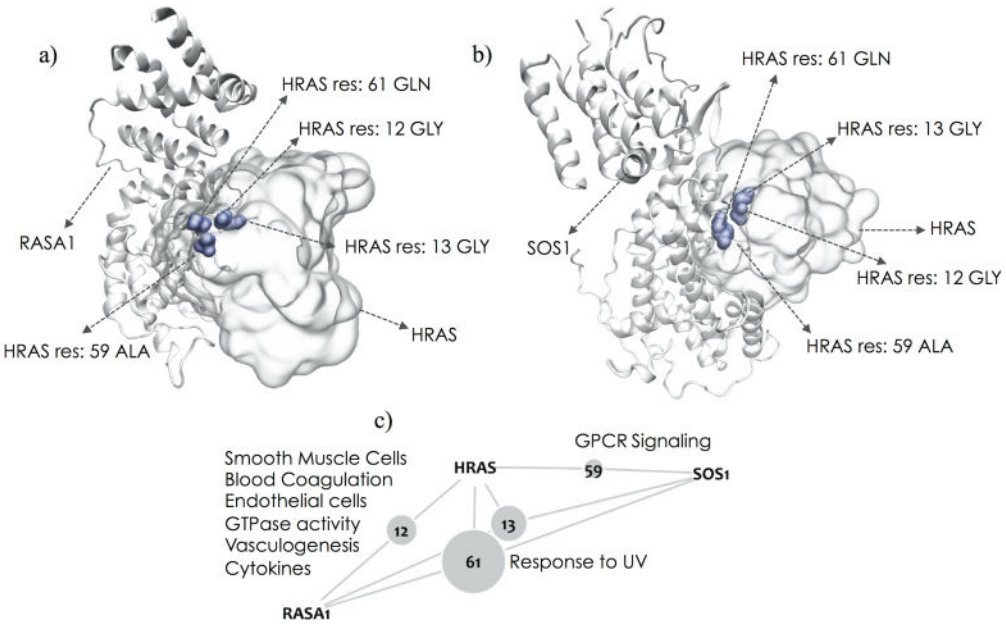

Figure 3.

The RASA1–HRAS complex (a) and the SOS1–HRAS complex (b) show that both interactors utilize the same binding site on HRAS. Residues 12, 13, 59 and 61 on HRAS that participate in the interface region of these interactions are highlighted in blue. Residue 12 mediates the HRAS-RASA1, residue 59 mediates the HRAS-SOS1 interaction, and 13 and 61 participate in both interactions. The small network (c) provides a schematic of the residues mediating particular interactions. The grey nodes represent the residues. Their sizes are proportional to the frequency with which they are mutated in the TCGA cohort. Functions predicted to be specifically altered by mutations at each interface are listed on the edges.