

Supplementary Information for
**Pan-Cancer Network Analysis Identifies Combinations of Rare Somatic
Mutations across Pathways and Protein Complexes**

Mark D.M. Leiserson^{1,2,*}, Fabio Vandin^{1,2,*}, Hsin-Ta Wu^{1,2}, Jason R. Dobson^{1,2,3}, Jonathan V. Eldridge¹, Jacob L. Thomas¹, Alexandra Papoutsaki¹, Younhun Kim¹, Beifang Niu⁴, Michael McLellan⁴, Michael S. Lawrence⁵, Abel Gonzalez-Perez⁶, David Tamborero⁶, Yuwei Cheng⁷, Gregory Ryslik⁸, Nuria Lopez-Bigas^{6,9}, Gad Getz^{5,10}, Li Ding^{4,11,12}, Benjamin J. Raphael^{1,2}

¹*Department of Computer Science, Brown University, Providence, RI, USA*

²*Center for Computational Molecular Biology, Brown University, Providence, RI, USA*

³*Department of Molecular Biology, Cell Biology, and Biochemistry, Brown University, Providence, RI, USA*

⁴*The Genome Institute and Department of Medicine, Siteman Cancer Center, Washington University in St. Louis, MO 63108, USA*

⁵*The Broad Institute of MIT and Harvard, Cambridge, Massachusetts 02141, USA*

⁶*Research Unit on Biomedical Informatics, Department of Experimental and Health Sciences, University Pompeu Fabra, Barcelona, Spain*

⁷*Program of Computational Biology and Bioinformatics, Yale University, New Haven, CT, USA*

⁸*Department of Biostatistics, Yale School of Public Health, New Haven, CT, USA*

⁹*Institució Catalana de Recerca i Estudis Avançats (ICREA), Barcelona, Spain.*

¹⁰*Massachusetts General Hospital, Boston, Massachusetts 02114, USA*

¹¹*Department of Medicine, Washington University in St. Louis, MO 63108, USA.*

¹²*Siteman Cancer Center, Washington University in St. Louis, MO 63108, USA.*

*Equal contribution.

Correspondence: braphael@brown.edu

Supplementary Note

1 HotNet2 Algorithm

1.1 Motivation

We developed the HotNet2 (HotNet diffusion oriented subnetworks) algorithm to identify subnetworks of a genome-scale interaction network that are mutated more than expected by chance (Supplementary Figure 2). Standard computational approaches to analyze mutations on pathways and protein complexes are severely limited by the statistical requirement of defining *a priori* a reasonable number of gene sets, or combinations of genes to evaluate. This imposes multiple undesirable constraints. First, gene sets often contain dozens of genes, reducing the power to identify a smaller subset containing a few directly interacting proteins. Second, there is typically extensive overlap between annotated gene sets, complicating the interpretation of analyses that report tens to hundreds of “significant,” but overlapping, pathways. Third, by ignoring the topology of interactions, gene-set analyses have reduced ability to identify crosstalk between pathways. Finally, *a priori* definition of gene sets prevents the discovery of novel combinations of mutations.

While interaction networks have proven useful in analyzing various types of genomic data [1, 2], statistically robust identification of significantly mutated subnetworks is a difficult problem with three major challenges. First, the number of subnetworks is too large to test exhaustively in a computationally efficient and statistically rigorous manner; e.g. $> 10^{14}$ subnetworks of ≥ 8 proteins in a medium-size human interaction network. Second, the topology of biological interaction networks is heterogeneous; many proteins,

and in particular many cancer-related proteins, have a large number of neighbors. Third, both the frequency of somatic mutations in individual genes/proteins and the topology of the interactions between proteins determine the significance of a subnetwork. While approaches have been introduced to find network modules [1, 2], rank gene sets, or prioritize disease-related genes [3] these approaches consider only network topology and not also the scores of individual genes.

1.2 Insulated heat diffusion as a random walk

HotNet2 uses an insulated heat process (See *Online Methods*) that can also be described in terms of a random walk with restart. The random walk starts from a protein g , and at each time step t moves to one of the neighbors (chosen uniformly at random) of the current node g_t with probability $1 - \beta$ ($0 \leq \beta \leq 1$), while the walk restarts from g with probability β . This process is defined by a transition matrix W :

$$W_{ij} = \begin{cases} \frac{1}{\deg(j)} & \text{if node } i \text{ interacts with node } j, \\ 0 & \text{otherwise.} \end{cases} \quad (1)$$

where $\deg(i)$ is the number of neighbors (i.e., the degree) of protein g_i in the interaction network. The locality of the walk – and therefore the insulated heat diffusion process – is governed by the parameter β , representing the probability with which the walk starting at g_i is forced to restart from g_i . Assuming the graph is connected (in practice we restrict attention to the largest connected component) the Ergodic Theorem guarantees that the random walk starting at protein g_i reaches a stationary distribution described by the vector \vec{s}_i :

$$\vec{s}_i = \beta(I - (1 - \beta)W)^{-1}\vec{e}_i, \quad (2)$$

where \vec{e}_i is the vector with a 1 in position i and 0's in the remaining positions. The j -th entry $\vec{s}_i(j)$ of \vec{s}_i gives the probability that, in the limit, the random walk starting at node g_i is at node g_j . We define the diffusion matrix F for a graph G as

$$F = \beta(I - (1 - \beta)W)^{-1}. \quad (3)$$

Note that \vec{s}_i is the i^{th} column of F .

1.2.1 Asymmetry of heat diffusion

Any clustering algorithm – whether on a graph or not – depends on a notion of distance, or oppositely similarity, between points. Distances are by definition symmetric; however similarity might not be symmetric. For example, some models of protein sequence similarity are non-symmetric [4]. In the case of HotNet2, similarity is non-symmetric for two reasons: first, the local topology of a pair of nodes u and v in the network which is encoded in the heat diffusion process – is not symmetric. A simple example is shown in Supplementary Figure 4: a node u of degree 1 sends all its heat to its neighbor v , but v sends heat to many nodes, including u . Thus, u might be “closer” to v than v is to u . Second, the score on node u and the score on node v are typically different.

1.3 Statistical significance

We evaluated the statistical significance of HotNet2 subnetworks and the HotNet2 consensus using the two-stage statistical test introduced in the original HotNet algorithm [1, 2]. In brief, the first stage is to compute a P -value for the statistic X_k , the number of subnetworks of size $\geq k$ reported by HotNet2. We compute the empirical distribution of X_k by running HotNet2 on random data where we permute the heat

scores on genes, restricting the permutation to the genes that are in the network and not removed by the expression filter. Although this permutation does not preserve any correlation between a gene’s heat and its network topology, we found that a more restricted permutation test did not substantially change the empirical distribution (Supplementary Note Section 5.3 and Supplementary Figure 19). For the HotNet2 consensus, we permuted data preserving the relationship between the mutation frequency and MutSigCV score (See *Online Methods: Statistical Significance*).

The second stage computes the False Discovery Rate for the set of significant subnetworks identified, as described in HotNet [1, 2]. In this analysis we used

$$\gamma_i = \frac{1}{2^{i-1}},$$

for the set of subnetworks of size i , for $i = 2, \dots, 8$, and

$$\gamma_9 = 1 - \sum_{i=2}^8 \frac{1}{2^{i-1}},$$

since we considered only subnetworks of size between 2 and 9. (We note that the parameter γ was called β in the HotNet publication [1, 2].)

1.4 Parameter selection

1.4.1 Insulated heat-diffusion: β

The HotNet2 parameter β is chosen for a given protein-protein interaction network, and remains fixed for different heat datasets. We chose β to balance the amount of heat that diffuses from a protein to its immediate neighbors and to the rest of the network. This was done by computing the amount of heat retained in the neighbors of vertices (“source proteins”) with different network centrality. In particular, we computed the *betweenness centrality* for each protein v , the number of shortest paths between all pairs of proteins that pass through v . We then picked five source proteins from the interaction network; those with the minimum, 25% quantile, median, 75% quantile, and maximum betweenness centrality. We examine vertices with different network centrality in order to choose diffusion parameter β such that all proteins retain most of their heat in their immediate neighbors. To determine whether the heat from each source protein v was retained in v ’s neighbors $N(v)$ or spread throughout the network for a given β , we compared the influence each source protein had on three sets of proteins: $N(v)$, all nodes distance 2 from v , and all other nodes in the network. We compared the three sets by counting the number of proteins in each set on which v had at least θ influence. We considered the 20 distributions with $\beta = 0.05, 0.10, \dots, 1$. From these distributions we chose β such that each source protein had most of its heat concentrated in its neighbors. As β decreases from 1 (the maximum restart probability), the amount of heat on $N(v)$ increases as less and less heat is retained in v . However, there is an inflection point at which the amount of heat on $N(v)$ will decrease, as more and more heat will reach the neighbors of $N(v)$. We choose β as the minimum β before this inflection point. For example, in Supplementary Figure 24, we can see the inflection point near $\beta = 0.4$ for HINT+HI2012. We used $\beta = 0.4$ for the HINT+HI2012 network, $\beta = 0.45$ for the iRefIndex network, and $\beta = 0.50$ for the Multinet network (see below).

We evaluated the sensitivity of the HotNet2 results to the value of parameter β by comparing subnetworks reported by HotNet2 on the HINT+HI2012 network and mutation frequency scores (See below), varying $\beta \pm 10\%$ from the value 0.40 determined by the automated procedure. We found that changing β has only a minor effect on the results, with at most 7 genes (3.8% of total) added/removed from the subnetworks (See Supplementary Table 29).

1.4.2 Minimum edge weight: δ

We choose the edge weight parameter δ such that HotNet2 will not find large subnetworks using random data. Specifically, for each mutation dataset and interaction network, we generated 100 random networks with the same degree distribution as the original network by performing $Q \times |E|$ connected edge swaps (E is the set of edges in the interaction network), ensuring that each node retains the same degree as in the original network and that the resulting network is connected, setting $Q = 100$ [5]. For each permuted network, we identified the minimum δ such that all strongly connected components found by HotNet2 have size $\leq L_{\max}$, for $L_{\max} = 5, 10, 15, 20$. For each L_{\max} , we report the median of these values of δ across the 100 permuted networks (Supplementary Figure 25). For each run, we selected the smallest δ with the most significant ($P < 0.05$) subnetwork sizes k .

We evaluated the sensitivity of the HotNet2 results to the value of the parameter δ by comparing subnetworks reported by HotNet2 on the HINT+HI2012 network and mutation frequency scores (See below) varying $\delta \pm 5\%$ from the value 0.000496 determined by the automated procedure. We found that varying δ changed at most 35 genes (12.3% of total) in the subnetworks (See Supplementary Table 30).

1.5 Consensus over multiple networks and gene scores

We derive “consensus” subnetworks from HotNet2 runs on different networks and/or gene scores using the following iterative procedure on a weighted graph. We built a complete, weighted graph G with genes on the vertices, and an edge between a pair of genes (g, g') weighted by the number of networks in which HotNet2 reports g, g' in the same subnetwork (independent of the scores used). We then identified the consensus subnetworks from G as follows. First, we initialized the consensus subnetworks C as the connected components of G when restricting attention to edges with weight = 3 (i.e. pairs of genes found in the same subnetwork by HotNet2 in all three interaction networks). Then, we extended each consensus subnetwork $S \in C$ by adding each gene g such that all g ’s weight two edges end in S . We then extended the consensus subnetworks once more using the same procedure but restricting to only weight one edges. We call genes “linkers” if they have weight two edges in multiple consensus subnetworks.

2 Data

2.1 Somatic aberration data

The somatic aberration data downloaded from Synapse and Firehose consists of SNVs in 20,472 genes from 3,281 samples, and CNAs in 720 genes from 4,334 samples. We describe the sources and our pipeline for processing somatic mutation data in *Online Methods: Somatic aberration data* and Supplementary Figure 1. The dataset after processing consists of somatic aberrations in 19,459 genes from 3,110 samples, and after applying the gene expression filter includes somatic aberrations in 11,565 genes in 3,110 samples.

2.2 Gene scores

Many factors determine the significance of mutations in individual genes, including the frequency of recurrence across samples, gene length, mutation context, regional variation in mutation rates, etc. Several approaches have been introduced to account for these factors [6, 7, 8, 9]. We performed HotNet2 analysis using two approaches to assign heat to individual genes according to recurrence and predicted functional impact.

Mutation frequency Also used by Vandin *et al.* in HotNet [1]. We computed the mutation frequency for each gene as the number of samples with at least one SNV or CNA in that gene. After processing

the dataset contains somatic aberrations in 19,459 genes from 3110 samples. See Supplementary Figure 1.

MutSigCV *q*-values MutSigCV [6] computes the statistical significance of mutations in genes across a cohort of samples. MutSigCV first estimates a per-gene BMR, and then computes the probability of observing the given number of mutations in each gene given its BMR. We used the $-\log_{10}$ of the MutSigCV *q*-values for all 18,371 genes.

We use these two scoring schemes because they are at opposite extremes in terms of stringency. The mutation frequency scores provide an opportunity for higher sensitivity, particularly for rarely mutated genes or genes primarily mutated by copy number aberrations. MutSigCV assesses the statistical significance of individual genes, and identifies only 175 genes with $q < 0.05$ in the dataset, which we expect to result in high specificity.

2.3 Expression filtering

We downloaded a list of expressed genes from [syn1734155](#). This list includes 12,081 genes with at least 3 RNA-Seq reads per sample in at least 70% of samples, as used in [10]. We restricted HotNet2 scores to these expressed genes plus a list of 18 well-known cancer genes (*AR*, *CDH4*, *EGFR*, *EPHA3*, *ERBB4*, *FGFR2*, *FLT3*, *FOXA1*, *FOXA2*, *MECOM*, *MIR142*, *MSH4*, *PDGFRA*, *SOX1*, *SOX9*, *SOX17*, *TBX3*, *WT1*) that have low transcript detection levels. After expression filtering, the input datasets to HotNet2 consisted of the following:

- somatic aberrations in 11,565 genes from 3,110 samples;
- MutSigCV *q*-values for 11,215 genes.

When we restrict these genes to those also in one of the three protein interaction networks, the somatic aberration data includes 10,208 genes and the MutSigCV *q*-values include 10,215 genes.

2.4 Copy number calling and target selection

We determined targets of copy number alterations and performed copy number calling by the following two steps. First, we downloaded GISTIC2 results for individual cancer types and Pan-Cancer (Pan12) from Firehose. For the Pan-Cancer GISTIC2 dataset, we selected target genes in each maxpeak by checking whether the peak contains genes from [11]. For the individual cancer type GISTIC2 datasets, we selected target genes in each maxpeak by checking whether the peak contains cancer genes defined by The Sanger Institute Gene Census [12]. For max peaks that did not have target genes, we picked a max peak as a target gene if the peak contains only one gene. Second, to determine if each sample had a copy number aberration in a given target gene, we extracted discretized copy number amplifications and deletions from focal copy number binary table provided by GISTIC2 with \log_2 ratio value above 0.9 and below -0.9 , respectively (Supplementary Figure 1).

2.5 Protein-protein interaction networks

To date there is no single database of human protein-protein interactions with high sensitivity and specificity. We used three interaction networks with varying numbers of interactions to allow for different false positive and false negative interactions in the analysis: (1) HINT+HI2012, a combination of high-quality protein-protein interactions from HINT [13] and the recent HI-2012 [14] (See URLs) set of protein-protein interactions; (2) MultiNet [15] a network that integrates multiple types of interactions from different databases;

(3) iRefIndex [16], an integrated network from multiple data sources. These three interaction networks have different trade-offs in sensitivity vs. specificity that would not be represented by simply merging the three networks. Supplementary Figure 29 shows the overlap between these networks. Unfortunately, a simple merge of networks will create a highly-connected network of 16,843 nodes and 180,271 edges. While such a network might have higher sensitivity for pairwise protein-protein interactions than individual networks, it will have very low specificity: the merged network will contain *all* false positive interactions from individual networks. Such a network is not expected to be a reasonable representative of the underlying biology. Moreover, a merged interaction network will have different network properties than the individual interaction networks (e.g. the average pairwise distances between genes are 3.302 on the merged network compared to 4.087, 3.644, and 3.394 on HINT+HI2012, iRefIndex, and Multinet, respectively). Changes in topology may have large effects for HotNet2 which analyzes network topology (not just nearest neighbors of a protein). Rather than creating a merged network, we run HotNet2 on the individual network individually and then form a consensus across the three networks from the HotNet2 runs (See *Online Methods: Finding consensus subnetworks and linkers*). This procedure is in effect giving greater weight to the interactions that are common among the three networks, while also accounting for the topology of each individual network.

We provide details of our processing of each of the three networks below.

iRefIndex. We downloaded iRefIndex version 9.0¹, and constructed an interaction network using all interactions except colocalizations and genetic interactions, including disulfide bond, ubiquitination, palmitoylation, deacetylation, sumoylation, physical interaction, methylation, hydroxylation, phosphorylation, aggregation, adp ribosylation, physical association, enzymatic reaction, covalent binding, phosphotransfer, deubiquitination, cleavage, acetylation, deneddylation, genetic inequality, demethylation, protein cleavage, glycosylation, dephosphorylation, neddylation, and direct interaction. The iRefIndex network consists of 91,872 interactions among 12,338 proteins.

HINT+HI2012. We created HINT+HI2012 protein interaction network by considering two interactome databases: HI-2012 prepublication data in human HI2 Interactome database² (HI2012) and high-quality interactomes database³ (HINT). We merged HI-2012 and the binary and co-complex interactomes in HINT to create the HINT+HI2012 protein-protein interaction network. The HINT+HI2012 network consists of 40,783 interactions among 10,008 proteins.

MultiNet. We downloaded MultiNet⁴, and constructed a interaction network using the entire set of interactions, including protein-protein, phosphorylation, metabolic, signaling, genetic, and regulatory interactions from multiple databases. The MultiNet network consists of 109,597 interactions among 14,445 proteins.

We removed self-loop interactions and multiple edges between interactions in all three networks. We restricted our HotNet2 analysis to the largest connected component consisting of the vast majority of genes and interactions in each network: 91,808 interactions among 12,128 proteins in iRefIndex; 40,704 interactions among 9,858 proteins in HINT+HI2012; and 109,569 interactions among 14,398 proteins in MultiNet.

2.6 Germline variants

To identify potential germline mutations, we compared the mutations used in our HotNet2 analysis to the ones identified using more stringent germline filtering (See [syn1729383](#)). This more stringent germline

¹http://irefindex.uio.no/wiki/Sources_iRefIndex_9.0

²<http://interactome.dfci.harvard.edu>

³<http://hint.yulab.org/batch.html>

⁴<http://homes.gersteinlab.org/>

filtering removes variants that are common in NHLBI exomes or 1000 Genomes data (global minor allele frequency > 0.1%). However, this more stringent filtering removes some *bonafide* somatic mutations.

We ran HotNet2 on this strictly filtered set of SNVs. We found 13 subnetworks in the strictly filtered consensus, 12 of which were also found in the original HotNet2 consensus (Supplementary Table 39). The only subnetworks that were missing were the RTK subnetwork and the *BOC-CDON* subnetworks. The strictly filtered consensus included an additional subnetwork not included in the HotNet2 consensus that consists of the *ANK1*, *SPTB*, and *ITPR3* genes. We note that we still identify the MHC Class I subnetwork using the strictly filtered dataset, emphasizing that there are *bonafide* somatic mutations driving the subnetwork even though many of its genes are known to have germline mutations.

3 Statistical Test to Evaluate HotNet2 Pan-Cancer results

3.1 Finding positional and structural clusterings: NMC and iPAC algorithms

We tested whether genes in subnetworks identified by HotNet2 had either significant clusters of missense mutations or at least 20% inactivating mutations (defined as nonsense, nonstop, splice-site, and frameshift insertions and deletions). These two mutational signals are the “20/20 Rule” that is hypothesized to distinguish cancer genes [17]. We use the NMC [18] algorithm to test for non-random clustering of missense mutations in the protein sequence, and the iPAC [19] algorithm to test for non-random clustering in the protein structure. We ran the NMC Algorithm [18] on genes with missense mutations. We matched the ENSEMBL transcript ID to the corresponding protein sequence. We used only those mutations for which there was concordance between the missense mutation and the amino acid at the appropriate position in the FASTA-format protein sequence. We applied the NMC algorithm with Bonferroni correction and reported the genes with at least one mutated amino acid cluster that had a P -value smaller than the significance threshold of $\alpha < 0.05$. Similarly, we applied the iPAC [19] algorithm on a subset of the above data to identify non-random somatic mutational clusters while taking into account the protein tertiary structure via Multidimensional Scaling (MDS). We only used genes with a defined protein structure in the PDB.

3.2 Cancer type specificity test

We annotated the subnetworks output by HotNet2 on each interaction network by determining which subnetworks are enriched for mutations from a specific cancer type. For a subnetwork S and cancer type τ , we calculated the following statistic. Let N_τ be the number of samples of type τ , and let $\Gamma(S)$ be the set of samples with at least one mutation in a gene in S . Let $C_\tau(S) = |\{p|p \in \Gamma(S), \text{type}(p) = \tau\}|$, i.e. the number of mutated samples from S of type τ . We calculate the enrichment of subnetwork S for mutations of cancer type τ conditioning on the number of mutations in the subnetwork, and the number of samples of type τ . The enrichment statistic is calculated using Fisher’s exact test of the 2×2 contingency table having as categories the mutation status of a sample (‘Mutated’ if the sample has a mutation in the subnetworks, ‘Not mutated’ otherwise) and the cancer type of the sample (‘In τ ’ if the sample is in type τ , ‘Not in τ ’ otherwise). The entries of the contingency table are (in row-wise order): $C_\tau(S)$; $N_\tau - C_\tau(S)$; $\sum_{\tau' \neq \tau} C_{\tau'}(S)$; $\sum_{\tau' \neq \tau} N_{\tau'} - C_{\tau'}(S)$. We correct the P -value from Fisher’s exact test using the Bonferroni correction, where the number of hypotheses is the product of the number of cancer types and the number of genes in the consensus, and only report the corrected P -values.

We also tested each gene in each subnetwork identified by HotNet2 for enrichment for mutations from a particular cancer type. We test this by conditioning on the number of observed mutations for the gene, regardless of the number of mutations in the subnetwork, by using Fisher’s exact test for the 2×2 contingency table with categories the mutations status of the (‘Mutated’ if the sample has a mutation in the gene, ‘Not mutated’ otherwise) and the cancer type of the sample (‘In τ ’ if the sample is in type τ , ‘Not in τ

otherwise). Again, we correct the P -value from Fisher's exact test using the Bonferroni-correction, where the number of hypotheses is the number of cancer types, and only report the corrected P -values.

3.3 Statistical approach to evaluate exclusivity and co-occurrence of mutations

We assessed the statistical significance of the mutual exclusivity or co-occurrence of mutations in subnetworks identified by Hotnet2 using a one-sided Fisher's exact test on a 2×2 contingency table. In particular, let x be a 2×2 contingency table, each cell $x_{i_1 i_2}$ denotes the count of samples corresponding to the mutation status (not mutated (0) or mutated (1)) of the pair of subnetworks, e.g. x_{10} is the number of samples with mutations in one subnetwork($i_1 = 1$) and without mutations in another subnetwork($i_2 = 0$). To test the exclusivity of a contingency table, we use as a test statistic $T_x = x_{10} + x_{01}$.

Since P -values from exact tests like Fisher's exact test are typically overconservative, we use the mid P -value instead [20]. The mid P -value is the average of the probability of a value at least as extreme as the observed value and the probability of a value more extreme than observed. In our case, the mid P -value is defined as

$$\frac{1}{2}(P(T \geq T_x) + P(T > T_x)) = \frac{1}{2}P(T = T_x) + P(T > T_x).$$

We assess the statistical significance of the mutual exclusivity or co-occurrence on each cancer type using the above method. To assess on the whole Pan-Cancer data, we performed the Cochran-Mantel-Haenszel test to test for independence across the contingency table of each cancer type.

4 Website

By default, the HotNet2 software package outputs an interactive website that includes visualizations of each component identified by HotNet2. These visualizations incorporate protein-protein interaction network and mutation data (See example in Supplementary Figure 6).

The Pan-Cancer HotNet2 results are available online at <http://compbio.cs.brown.edu/pancancer/hotnet2/>. This website includes visualizations for each subnetwork which incorporate protein-protein interaction network, mutation, gene transcript, and protein domain data (See example in Supplementary Figure 6).

5 HotNet2 Pan-Cancer results

We ran HotNet2 for each combination of interaction network and score. HotNet2 identified a number of significant subnetworks (Supplementary Tables 1-2) in all six runs. Below we describe some of the subnetworks identified using the consensus procedure (Supplementary Table 3-5), as well as other subnetworks (Supplementary Table 6-18). Supplementary Tables 6-18 report the cancer type enrichments P -value and the NMC and iPAC clustering P -values for the subnetworks in the main text. Most of the subnetworks identified by HotNet2 are not identified by standard pathways enrichment analysis (Supplementary Table 35) and also contain genes not reported as significantly or recurrently mutated by MuSiC [9] or MutSigCV [6] or Oncodrive [7, 8] (Supplementary Table 20). The subnetworks identified by HotNet2 show cancer type enrichments in mutations and genes in the subnetworks show enrichments for mutations in different cancer types (Supplementary Tables 6-18). Genes in the subnetworks also show significant position or structure clustering.

5.1 Genes with positional and structural clusters

We found that high scoring genes in our data were enriched for both positional and structural clusters of mutations ($P < 10^{-20}$, details below). We find sequence and structural clustering for 100 and 16 genes,

respectively, in the subnetworks (Supplementary Table 6-18). In addition, we find 25 genes mutated in $\geq 1\%$ of samples that have at least 20% inactivating mutations. Overall, genes in the HotNet2 consensus subnetworks are enriched for NMC clusters ($P < 0.0001$, Supplementary Note Section 5.1.1) or 20% inactivating mutations ($P < 0.0001$, Supplementary Note Section 5.1.2) compared to genes not in subnetworks (Supplementary Note Section 6.1).

5.1.1 Contingency tests for genes with NMC clusters

One method to distinguish driver mutations from passenger mutations is to look for positional clustering of mutations in genes across patient samples [17]. We used an implementation of the NMC algorithm to look for clusters of missense mutations [18] in genes that have mutation heat in any of the three PPI networks used for Hotnet2 runs. We then classified these 10,478 genes based on whether they have a significant clustering of missense mutations (NMC corrected $P < 0.05$), are found in a consensus subnetwork, are neither, or are both. Using a χ^2 test with Yate's correction, we find that genes with significant clusters of missense mutations were significantly ($P < 0.0001$, $\chi^2 = 42.23$) overrepresented in consensus subnetworks for Hotnet2 (See Supplementary Table 21). Specifically, we found: there were 5,467 genes that do not have either an NMC cluster or were in a Hotnet2 consensus subnetwork; there were 4,886 genes that have an NMC cluster and were not in a consensus subnetwork; there were 29 genes which were in a Hotnet2 consensus subnetwork and did not have an NMC cluster; and there were 96 genes that had an NMC cluster and were found in a Hotnet2 consensus subnetwork.

5.1.2 Contingency tests for genes with inactivating mutations

One method for discovering driver mutations in genes is to look at the frequency at which inactivating mutations occur in a gene across patient samples. Inactivating mutations are those which would cause aberrant transcription (splice site mutations), translational frameshift (frameshift indels), or premature translational stop (nonsense mutations). A proposed threshold for distinguishing driver mutations from passenger mutations is to look for genes whose inactivation frequency across patients samples is $\geq 20\%$ [17]. To classify a gene as being inactivated by the proposed 20/20 rule, we only considered genes that were mutated in $\geq 1\%$ of patients to reduce false positives from the high number of genes that have an inactivating mutation in only 1 or 2 samples. Using a χ^2 test with Yate's correction, we find that genes with inactivating mutations were significantly ($P < 0.0001$, $\chi^2 = 280.86$) overrepresented in consensus subnetworks for Hotnet2 (See Supplementary Table 22). Specifically, we found: there were 10,221 genes that were not inactivated in $\geq 20\%$ of samples or were in a Hotnet2 consensus subnetwork; there were 132 genes that were inactivated in $\geq 20\%$ of patient samples and were not in a consensus subnetwork; there were 100 genes which were in a Hotnet2 consensus subnetwork and were not inactivated in $\geq 20\%$ of patient samples; and there were 25 genes that were inactivated in $\geq 20\%$ of patient samples and were found in a Hotnet2 consensus subnetwork.

5.1.3 Correlation between heat and mutation clusters and inactivating mutations

We evaluated whether hotter genes had more clustering of missense mutations or $\geq 20\%$ of mutations as inactivating. We indeed find that hotter genes (according to mutation frequency) are enriched for these two properties ($P < 10^{-20}$ by Mann-Whitney U test). We performed a similar analysis for MutSigCV scores finding $P = 3.09 \times 10^{-7}$. See Supplementary Figure 13.

5.2 Sample mutation rates in novel genes

We compared the mutation rates of samples that have a mutation in at least one of the 80 novel genes found by HotNet2 but no mutations in the 139 known cancer genes (reported in [10, 11, 17], or by MutSigCV

or Oncodrive) to the mutation rates of samples with at least one mutation in at least one of the 139 known cancer genes, but no mutations in the 80 novel genes. We found no significant difference in these rates (mean of 32.26 mutations per sample for novel genes, versus 32.30 mutations per sample for known genes, P -value = 0.4 by Mann-Whitney U test), showing that mutations in the novel genes are not restricted to those samples with large numbers of passenger mutations (See Supplementary Figure 14).

5.3 Heat scores in the network

The null hypothesis tested by HotNet2 is that the heat scores on genes are independent of the network topology. Two issues that confound this analysis are:

1. The distribution of scores for genes in the network is different from the distribution of heat scores not in the network. We find that this is the case. We performed a two-sample z-test to determine if the heat scores in genes in or out of the PPI network were significant. We repeated this experiment for the mutation frequency and MutSigCV datasets for each of the HINT+HI2012, iRefIndex, and Multinet PPI networks. For each dataset and network combination, the distributions were statistically significantly different, with genes in the network having higher overall heat. We show the CDFs of each distribution, including Z-scores and P -values, in Supplementary Figure 21.

We address this concern by using the score distribution for genes in the network; i.e. our permutation test is restricted to genes in the network.

2. There is a correlation between gene score and degree in the network. We computed the correlation between degree and heat scores. We find the correlation ranges between $\rho = 0.09$ and $\rho = 0.15$ across the three interaction networks and two heat scores (MutSigCV and mutation frequency). However, we found that this correlation was primarily due to the top 100 highest scoring genes in the network. Supplementary Figure 22 shows a plot of the correlation as we remove the N highest-scoring genes from each network. After $N = 100$, the correlation has dropped below $\rho = 0.05$ for each network-score pair except Multinet-MutSigCV.

To assess the effect of the degree-heat correlation on the HotNet2 permutation test, we computed the number of subnetworks found by HotNet2 on the HINT+HI2012 network, using a permutation test that preserves the correlation between degree and heat. Specifically, we permute the scores of 100 highest scoring genes amongst themselves, and permute the scores of the remaining genes amongst themselves. We generated 1000 such datasets. This permutation test maintains the correlation between degree and heat (Supplementary Figure 20c). The sizes of subnetworks found by HotNet2 were very consistent with those from the original permutation test that permuted heat scores across the nodes uniformly at random (See Supplementary Figure 20a-b). We compared the distributions of the number of components of size at least 3 for both permutation methods, and found that while the permutation method that does not preserve degree produces significantly more components using a t-test ($P = 0.02$), the difference between the distributions is tiny (permuted mean: 25.375, degree permuted mean: 25.294).

5.4 TP53, PI(3)K, and NOTCH subnetworks

We show mutation matrices for the PI(3)K, and NOTCH subnetworks in Supplementary Figures 16 and 17. The mutation matrix for the *TP53* subnetwork is too large to fit in one page, but is viewable in the online browser of the HotNet2 Pan-Cancer results (Supplementary Note Section 4). These subnetworks include genes that were *not* marked as significant by individual gene scores. For example, in the *TP53* subnetwork, mutations in *EPHA3* have been reported in several cancers; we find a cluster of 6 mutations ($P = 0.001$) in a domain that has been shown to significantly reduce the catalytic activity of *EPHA3* [21].

5.5 RTK subnetwork

HotNet2 Pan-Cancer analysis found a subnetwork containing several receptor tyrosine kinases (RTK) that have important roles in the development of multiple cancer types, which includes *EGFR*, *ERBB2* and *ERBB4* (Supplementary Figure 7 and Supplementary Table 11). The RTK subnetwork also contains *ELF3*, a gene involved in cell growth that was recently identified by [22] as a promising candidate cancer gene. *ELF3* is mutated in 19 samples, 11 (58%) of which have inactivating mutations, and is enriched for mutations in BLCA ($P < 3 \times 10^{-5}$) and COADREAD ($P < 0.02$).

5.6 ASCOM complex and interactors

The fifth most mutated consensus subnetwork (16.9% of all samples) contained the ASCOM complex (*MLL2* and *MLL3*), the putative ASCOM-interacting protein encoded by *KDM6A*, as well as *E2F3*, *N4BP2* and *PROSER1* (Supplemental Figure 11 and Supplementary Table 16). The interaction between these latter two genes and *KDM6A* was reported in a mass spectrometry screen [23], and thus is less characterized than the other interactions. The ASCOM complex and *KDM6A* are involved in coordinate deposition of histone modifications that promote transcription. This subnetwork is mutated in at least 5 samples from each cancer type, and is enriched for mutations in BLCA, HNSC, and LUSC. At the individual gene level, we identified a number of enrichment for mutations in different cancer types, including the association of *KDM6A* mutations with BLCA ($P < 10^{-15}$), and the association of *MLL3* mutations with BLCA ($P = 5 \times 10^{-5}$). Both *MLL3* and *KDM6A* have been previously reported to be associated with transitional cell carcinoma of the bladder[24]. *MLL2* is enriched for mutations in BLCA ($P = 7.5 \times 10^{-7}$), HNSC ($P = 5 \times 10^{-14}$), and LUSC ($P = 1.8 \times 10^{-9}$). *MLL3* and *PROSER1* show significant ($P < 10^{-14}$ and $P < 3 \times 10^{-4}$, respectively) clustering of missense mutations. *MLL3* mutations are clustered in the zinc-finger domain. Thus HotNet2 identifies novel cancer type specificities for mutations in the ASCOM complex and interactors, as well as some of the recently reported associations for their mutations. *KDM6A* showed significant clustering of mutations ($FDR < 0.1$ from iPAC) in its protein structure.

5.7 SWI/SNF complex

The SWI/SNF subnetworks consist of two consensus subnetworks. One includes the core members *SMARCA4* and *SMARCB1* of SWI/SNF complex, while the other includes SWI/SNF members *ARID1A*, *ARID1B*, *ARID2*, *PBRM1* plus *ADNP*. We jointly analyze these subnetworks as many individual HotNet2 runs report them together and they are members of the same protein complex. Supplementary Table 12 shows the enrichments for mutations for the subnetworks as a whole as well as for the single genes in the SWI/SNF complex subnetwork. See main text for discussion of this subnetwork.

5.8 BAP1 complex

Supplementary Table 13 show the enrichments for mutations for the subnetworks as a whole as well as for the single genes in the BAP1 complex subnetwork (Figure 3b). The subnetwork is enriched for mutations in BLCA ($P = 0.01$), KIRC ($P = 0.00023$), and LUSC ($P = 0.006$). KIRC enrichment was driven mostly by *BAP1*, while LUSC enrichment was due mostly to mutations in *ASXL1* and *ASXL2*. *BAP1* showed significant ($P < 9 \times 10^{-6}$) clustering of missense mutations in the N-terminal region where mutations have been observed to be associated with mesothelioma [25] and have been shown to inhibit the nuclear localization and deubiquitinase functions of BAP1 [26]). *FOXK1*, *FOXK2*, and *KDM1B* were the genes in the subnetwork with lowest mutation frequencies. *FOXK2*, a putative member of the BAP1 core complex [27], is a transcription factor whose phosphorylation results in apoptosis [28]. None of the mutations in *FOXK2*, were within these critical phosphorylation residues [28], suggesting that they are not perturbing

this interaction; however most of the missense mutations in *FOXK2* (6/13) were found in the forkhead transcription factor domain and forkhead associated domain, which may be inactivating the DNA-binding properties of *FOXK2*. With a few exceptions [29, 30, 31], many of these associations are novel.

5.8.1 Mutual exclusivity between BAP1 and SWI/SNF complexes

We find that the mutations in the BAP1 and SWI/SNF subnetworks display strong mutual exclusivity ($P = 9.4 \times 10^{-5}$) due to 12 mutually exclusive mutations in the BAP1 complex and 21 mutations (20/21 mutually exclusive) in SWI/SNF complex. These include rare mutations in *ASXL1*, *ASXL2*, *ANKRD17*, *FOXK1*, and *FOXK2* in BAP1 complex (2 of which are inactivating) and *ARID1A*, *SMARCA4*, *ARID2*, *ARID1B* and *SMARCB1* in the SWI/SNF complex (8 of which are inactivating), consistent with the inactivation of *BAP1* or *PBRM1*. We observe that 26 of the 42 KIRC samples with *BAP1* mutations have inactivating mutations. Moreover, 12 of the 16 KIRC samples with missense mutations have these mutations in the Peptidase_C12 domain ($P < 3 \times 10^{-4}$), and thus we classify such missense mutations as inactivating. Using this classification, we observe that 5/10 of mutated samples in BRCA, GBM, HNSC and LUAD are inactivating, suggesting that BAP1 inactivation also occurs in these samples. *ASXL1* mutations are common in LAML [32] and we observe that 5/5 mutations in LAML are inactivating mutations. In addition 6/9 *ASXL1* mutations in HNSC are inactivating, while 3/7 *ASXL2* mutations in BRCA are inactivating. These *ASXL1* and *ASXL2* mutations are exclusive of *BAP1* inactivation, demonstrating alternative strategies for inactivation of the BAP1 complex.

5.9 Core-binding factors

HotNet2 identifies a consensus subnetwork containing *RUNX1*, *CBFB*, and *ELF4* mutated in 2.8% of samples (Supplementary Figure 10 and Supplementary Table 14). The subnetwork is enriched for mutations in LAML ($P = 4.6 \times 10^{-6}$) and BRCA ($P = 9.9 \times 10^{-5}$), which is largely driven by *RUNX1* and *CBFB* mutations that are enriched in BRCA (respectively $P = 2 \times 10^{-4}$, $P = 9 \times 10^{-7}$). We observed 16/17 of missense mutations clustered in the Runt-homology domain (RHD) of *RUNX1* ($P = 2 \times 10^{-6}$), which is required for DNA binding as well as interaction with *CBFB* [33], and with the exception of one missense mutation outside of the RHD, all other mutations in *RUNX1* were inactivating. RHD mutations have been associated with development of LAML [34] and our analysis suggests that abrogation of DNA-binding capabilities of *RUNX1* through mutation of *RUNX1*-RHD or *CBFB* may be a critical step in breast cancer progression [35]. Further, the mutations in *RUNX1* that occur in BLCA, COADREAD and HNSC patients were located in the RHD. In addition, each of the mutations in *CBFB* that occur in BLCA, LAML, and UCEC samples were located in the domain that is critical for interaction between *CBFB* and *RUNX1* or were inactivating. These observations suggest that members of the core-binding complex are inactivated by somatic mutations in cancers other than BRCA and LAML, although perhaps only in rare cases. Thus, the HotNet2 Pan-Cancer analysis further characterizes the pattern of mutation in core binding factors across BRCA, AML, and other cancer types.

5.10 Cohesin complex

Supplementary Table 15 show the enrichments for mutations for the subnetworks as a whole as well as for the single genes in the cohesin complex subnetwork. We identified gene-specific enrichments for *STAG2* in BLCA ($P = 0.005$). In STAG1, we found three significant clusters of missense mutations that together encompass 28 total mutations. All of the genes in the subnetwork have $> 79\%$ of their mutations as missense, with the exception of *STAG2* and *RAD21* (57% and 69% respectively). Interestingly, all of the 8 mutations in these two genes in LAML were not missense (3 nonsense for *STAG2*; 2 nonsense, 2 frame shift insertions,

and 1 frame shift deletion for *RAD21*), suggesting that these genes are inactivated by mutation in LAML and other cancer types.

5.11 Condensin complex

HotNet2 Pan-Cancer analyses on the MultiNet and iRefIndex networks identified two subnetworks containing six proteins in the condensin complex, which together were mutated in 4.2% of samples. Supplementary Table 6 show the enrichments for mutations for the subnetworks as a whole as well as for the single genes in the condensin complex subnetwork. See main text for discussion of this subnetwork.

5.12 KEAP1, NFE2L2, and interactors

The seventh most mutated consensus subnetwork (8.5% of all samples) includes *KEAP1*, *NFE2L2*, *CHD6*, *WAC*, and *PTMA* (Supplementary Figure 8 and Supplementary Table 18). This subnetwork is mutated in at least 10 samples of each cancer type with the exception of GBM (3 mutated samples) and LAML (3 mutated samples). *NFE2L2* is a transcription factor that modulates the response to oxidative stress and is normally sequestered in the cytoplasm by *KEAP1* [36]. The subnetwork was enriched for mutations in BLCA ($P = 0.04$), HNSC ($P = 0.0004$), LUAD ($P = 4 \times 10^{-8}$), and LUSC ($P < 10^{-15}$). We find enrichment for mutations in *KEAP1* in LUAD ($P < 10^{-15}$), and in LUSC ($P = 8 \times 10^{-6}$), and enrichment for *NFE2L2* mutations in HNSC ($P = 2 \times 10^{-5}$) and LUSC ($P = 9 \times 10^{-13}$), consistent with previous reports [37, 38, 39]. Missense mutations in *KEAP1* and *NFE2L2* were clustered in their respective protein structures ($P < 4 \times 10^{-5}$ and $P < 10^{-15}$ respectively from NMC analyses, and $Q < 0.025$ for *KEAP1* from iPAC analysis). The 4 clustered mutations for *KEAP1* were all in the position of the *NFE2L2* binding Kelch domain [40], and were in different cancer types (2 in LUSC, 1 in LUAD, and 1 in HNSC), suggesting mutation of this domain is important as it is targeted in multiple types. The 36 significantly clustered mutations of *NFE2L2* are in the N-terminal region of the protein.

5.13 MHC Class I proteins

HotNet2 identified a significantly mutated subnetwork (3.1% of samples) containing five genes – *HLA-A*, *CD1D*, *HLA-B*, *B2M*, and *PH4TM* – with *HLA-A* and *HLA-B* of which are from the major histocompatibility complex (MHC) Class I (Supplementary Figure 12 and Supplementary Table 17). Two of these genes were previously reported to harbor somatic mutations: *HLA-A* in squamous lung cancer [37] and *B2M* in colon cancer [41]. However, we find that the mutations in LUSC are distributed throughout the subnetwork, while BLCA mutations are concentrated in *HLA-A* ($P = 0.001$). We find that the genes in this subnetwork harbor many inactivating mutations (41/97 samples), with *B2M* (7/17), *HLA-A* (19/34), and *HLA-B* (10/20) harboring the most. We also find a significant cluster of missense mutations in *B2M*'s first amino acid residue ($P = 1.53 \times 10^{-9}$), suggesting that different cancers utilize different mechanisms for inactivation of *B2M*. Finally, since *HLA-A* and *HLA-B* are highly polymorphic [41], we used more stringent filtering (See [syn1729383](#)) to remove potential germline variants. This procedure classified 14/34 samples with *HLA-A* mutations and 11/21 samples with *HLA-B* mutations as probable germline variants. These results demonstrate that different somatic aberrations, and possibly a combination of germline and somatic aberrations, perturb this subnetwork in different cancer types.

5.14 Telomerase complex and interactors

HotNet2 identified a significantly mutated subnetwork (8.5% of samples) containing two members of the telomerase complex, telomerase reverse transcriptase (*TERT*) and telomerase-associated protein 1 (*TEP1*),

as well as *SMG1*, *SMG5*, *SMG6*, and *SMG7* (Supplementary Figure 30 and Supplementary Table 38). HotNet2 identified this subnetwork in the consensus of the runs where no expression filter was applied to the mutation frequency or MutSigCV genes (Supplementary Note Section 2.3). The subnetwork is enriched for mutations in LUAD ($P = 0.004$) and LUSC ($P = 0.0009$), and *SMG7* is itself enriched for mutations in LUSC ($P = 0.03$).

5.15 CLASP and CLIP proteins

HotNet2 identified a significantly mutated subnetwork (2% of samples) containing the linker protein CLIP2 and the CLIP associated proteins CLASP1 and CLASP2 (Supplementary Figure 5 and Supplementary Table 7). HotNet2 identified this subnetwork on the Multinet network using mutation frequency gene scores. The CLASP1/2 proteins have been shown to be involved in cellular migration [42], and LOH in the chromosomal region of CLASP2 is associated with non-small cell lung cancer [43]. The association with lung cancer is supported by the mutations in this subnetwork, as the subnetwork ($P = 0.0001$) and both *CLASP1* ($P = 0.01$) and *CLASP2* ($P = 0.008$) are enriched for mutations in LUSC. Furthermore, 7/12 LUSC samples with a mutation in this subnetwork have an inactivating mutation.

6 Mutation Validation

To provide further support for a subset of novel cancer genes identified in the HotNet2 subnetworks, we examined RNA-Seq and whole-genome sequencing (WGS) data from the same TCGA samples, data that was generated independently from the whole-exome sequencing data that was used to call the somatic mutations. We focused our analysis on rare mutations in the condensin complex, which was one of the novel discoveries of our analysis (Figure 4 of manuscript), as well as genes that were not identified by other methods (marked by ‘**’ in the figures).

Examination of reference and variant allele counts in the RNA-Seq and WGS data validated 39 somatic mutations. These include a total of 12 non-silent mutations in condensin genes using RNA-Seq, 9 non-silent mutations using WGS, and one mutation using both data types. These mutations occur in condensin genes *SMC2*, *SMC4*, *NCAPD2*, *NCAPD3*, *NCAPH2*, and *NCAPG2*. In the larger set of novel genes, we have validated 8 non-silent mutations using RNA-Seq, 14 non-silent mutations using WGS, and 3 non-silent mutations using both types of data. These include mutations in the cohesin subunit *STAG1*, and the *BAP1* complex subunit *ASXL2*, among others.

Supplementary Table 23 provides the read counts of validated mutations (green rows for RNA-Seq, yellow for WGS, and orange for both), with missing data indicated by -1. This analysis demonstrates that several of the novel genes identified by the HotNet2 analysis contain *bonafide* somatic mutations, and provide promising targets for further functional characterization.

7 Comparison of HotNet2 to HotNet

7.1 Stars and spider graphs in Pan-Cancer data

We compared the proportion of subnetworks returned by HotNet and HotNet2 on the Pan-Cancer mutation data that are star/spider graphs dominated by a hot, central node. We say that a subnetwork S is a hot spider/star graph if the center (root) node $u \in S$ contains $> 50\%$ of the heat in S . A star graph of $n > 2$ nodes is a tree where the root has degree $n - 1$ and all other nodes have degree one. A spider graph of $n > 3$ nodes is a tree where the root has degree of at least three and all other nodes have degree of at most 2. Note that any star graph of at least 4 nodes is also a spider graph, and that we restrict our analysis to subnetworks of at least 3 nodes as star/spider graphs are defined for subnetworks of at least 3 nodes.

Because HotNet2 uses non-symmetric edge weights to identify strongly connected components, we hypothesized that HotNet2 would identify fewer hot star/spider subnetworks than HotNet. Indeed, HotNet2 returned only 12/90 hot spider/star graphs (13%) compared to 67/94 hot spider/star graphs (71%) returned by HotNet across the three PPI networks and two gene scores (Supplementary Table 31). Thus, HotNet2 returns $> 80\%$ fewer hot stars/spiders than HotNet. This is a major difference between the algorithms and is one of the reasons why HotNet fails to find statistically significant results ($P \leq 0.01$ for any subnetwork size k) on three of six runs (Supplementary Table 32,33), while HotNet2 finds statistically significant results on all six runs.

The difference is explained by the undirected vs. directed heat similarity measures used in HotNet vs. HotNet2. In HotNet (undirected heat), placing a high heat score on a node with relatively low-degree is likely to result in a hot star/spider. Thus, we find hot stars/spiders frequently in randomly permuted data, meaning that the number and size of subnetworks found in real data will not be significant. In contrast, in HotNet2 (directed heat), a single high heat score placed on a random node will not result in a hot star spider. Consequently, by accounting for the direction of heat flow, HotNet2 returns fewer hot stars/spiders on both real and random data, and thus the number and size of subnetworks found in real data is significant.

We note that the goal of HotNet2 is not to eliminate hot stars/spiders, but rather to reduce the number of such subnetworks that are false positives. Due to the incomplete and noisy PPI networks, there are true positive subnetworks that are stars/spiders; e.g. the three most mutated members of the BAP1 complex – *BAP1*, *ASXL1*, *ASXL2* – are a hot star in the iRefIndex PPI network. Restricting attention to the statistically significant subnetworks identified by each algorithm: only 2/36 (5%) of the significant subnetworks reported by HotNet2 are hot stars/spiders, compared to 15/35 (43%) of the significant subnetworks reported by HotNet. Furthermore, only 4/399 (1%) of protein-protein interactions within the HotNet2 subnetworks occur in hot spiders/stars, while 45/245 (18%) of interactions within the HotNet subnetworks occur in hot spiders/stars. It is difficult to determine which, if any, of these stars/spiders are false positives. However, this analysis demonstrates that hot stars/spiders are a much smaller proportion of HotNet2s subnetworks. This results in HotNet2 having higher statistical power and returning subnetworks that have a higher fraction of interactions with proteins other than a hot central node.

7.2 Simulated data

We performed three different simulations comparing HotNet2 and HotNet with: (1) randomized datasets; (2) highly connected subnetworks; (3) implanting highly connected groups of known cancer genes from [17]. The results of (1) show that neither algorithm identifies significantly mutated subnetworks in randomized datasets. The results of (2) and (3) show that HotNet2 achieves high sensitivity and specificity, and outperforms HotNet.

7.2.1 Randomized datasets

We evaluated whether HotNet2 and HotNet identify significant subnetworks on randomized datasets by permuting the MutSigCV heat scores 50 times and running both algorithms on the HINT+HI2012 network. For each run, we chose the smallest δ value for the run with lowest P -value across any subnetwork size k . On these simulations, we find that the P -values for different subnetwork size k are approximately uniformly distributed (Supplementary Figure 23(a) and Supplementary Table 28), and even appear to be overly conservative for larger subnetwork sizes. On 450 total subnetwork sizes (9 sizes \times 50 runs), HotNet2 found only 3 subnetwork sizes with $P < 0.01$ (compared to 4 for HotNet). The P -values for HotNet and HotNet2 are deflated (more conservative) only at larger P -values and for $k \geq 6$, and thus would not result in false negatives. Moreover, we note a larger deflation for HotNet (Supplementary Figure 23(a) bottom) compared to HotNet2 (Supplementary Figure 23(a) top), demonstrating one advantage of the newer algorithm.

The observed deflation of larger P -values is a result of the discreteness of the distribution of our test statistic X_k (= number of subnetworks of size at least k) for large values of k (Supplementary Figure 23(b-c)). In particular, on the 50 randomized datasets, we have that for large values of k (e.g., $k \geq 6$) the test statistic X_k assumes an integer value x^* within a limited range (e.g. for HotNet2, X_k takes values between 0 and 3 for $k = 7$ (Supplementary Figure 23(c)) and $k = 8$, and between 0 and 2 for $k = 9$ and $k = 10$ - see Supplementary Table 37). This implies that if x^* is the number of subnetworks of size $\geq k$ on a randomly selected dataset, then for large k , $x^* = 0$ with high probability; e.g. for $k = 7$ we have $Pr[x^* = 0] \approx 0.75$ for HotNet. On such a dataset, it follows that the tail probability $Pr[X_k \geq x^*] = 1$, resulting in the deflated p -values seen on the Q-Q plot for larger P -values and large k .

7.2.2 Highly connected subnetworks

We evaluated the performance of HotNet2 and HotNet in the ideal case where the highest heat scores are placed on genes that each algorithm considers “most connected” in the HINT+HI2012 interaction network. We identified the components to implant by running HotNet and HotNet2 with uniform heat scores on all genes in the network that passed the gene expression filter. We then selected the resulting components of size at least 10, giving us 16 components to implant for each algorithm. These components consisted of 203 genes for HotNet and 214 genes for HotNet2. We then randomly assigned the top mutation frequency scores to the genes in these selected components and randomly assigned the remaining scores to the rest of the genes in the network that had mutation frequency scores and passed expression filtering. HotNet2 returns 16 components of size at least 10 with a $P < 0.01$. These 16 components consist of 223 genes and contain all 214 implanted genes, giving a sensitivity of 100% and a specificity of 96.0%. Furthermore, all genes that are in the same implanted component are in the same recovered component. HotNet, by contrast, performs poorly. It recovers only 12 components of size at least 10 with $P < 0.01$. These 12 components consist of 206 genes of which only 151 are implanted genes, giving a sensitivity of 74.4% and a specificity of 73.3%. HotNet does recover some additional implanted genes in smaller components, but many of these components are of a size that does not achieve statistical significance. Even if we consider the best case for HotNet without regard to statistical significance, it only achieves sensitivity of 92.6% and specificity of 76.7%.

7.2.3 Highly connected cancer genes

We compared HotNet2 and HotNet on simulated mutation data where subnetworks of known cancer genes from [17] were mutated randomly in samples according to a specific driver mutation probability and other genes were mutated randomly in samples according to a passenger mutation probability. The passenger mutation probability was lower than the driver mutation probability, but these probabilities were chosen so that the distributions of the number of mutations in cancer genes and passenger genes overlapped (Supplementary Figure 26(a)). We selected subnetworks to implant by greedily adding the highest weight edges from the diffusion matrix F of the HINT+HI2012 interaction network to identify all non-overlapping subnetworks of size at most 15 that include multiple cancer genes from [17]. We used the subnetworks with at least four genes, resulting in implanted networks of 50 genes in six subnetworks. We also chose 5 known cancer genes that were not part of the implanted networks to have scores from the driver distribution with the hypothesis that HotNet would identify these subnetworks as hot stars/spiders. We then generated 30 simulated datasets by assigning heat scores from overlapping normal distributions for the 50 implanted (mean: 0.05, std: 0.01) and 2945 passenger (mean: 0.01, std: 0.01) genes (Supplementary Figure 26(a)). We then compared HotNet2 and HotNet on these datasets.⁵

⁵Running HotNet2 and HotNet on simulated datasets requires an automated procedure for choosing which subnetworks to report, which is dependent both on the minimum edge weight δ and the minimum component size k . For each run, we chose

Across the 30 datasets, both HotNet2 and HotNet identify statistically significant subnetworks that overlap the implanted subnetworks with high sensitivity and specificity (Supplementary Figure 26(b-c)). HotNet2 achieves a better balance in the tradeoff of sensitivity vs. specificity, as summarized by the Youden's index [44] which has a mean value of 0.86 for HotNet2 versus 0.76 for HotNet (difference is significant: $P = 4.2 \times 10^{-11}$ by t-test).

In order to determine if the relative advantage of HotNet2 compared to HotNet came from its selection of the parameter δ , we created ROC curves for each of these datasets varying δ . The ROC curves show that the performance of HotNet2 dominates HotNet with average area under the curve (AUC) = 0.998 for HotNet2 and 0.947 for HotNet. These results demonstrate that HotNet2 has consistently better performance than HotNet over a range of δ values and hence p-value thresholds.

We note that in practice HotNet2 chooses the value of the δ parameter automatically using randomized networks (See *Online Methods: HotNet2* section and Supplementary Note Section 1.4.2). This is to avoid testing numerous δ values which can result in overfitting to the data. The ROC curves in Supplementary Figure 26(d) show that the δ values automatically chosen by HotNet2 (dots) provide a reasonable balance between sensitivity and specificity.

7.3 Cross-validation

We compared the performance and stability of HotNet2 and HotNet using two-fold cross-validation. We repeatedly split the Pan-Cancer mutation dataset in two halves, by randomly selecting half of the samples from each cancer type. We ran HotNet2 (respectively HotNet) on the first half, finding the value of the parameter δ and the subnetwork size k that gave the smallest P -value under the HotNet2 (respectively HotNet) statistical test. HotNet2 returned significant subnetworks ($P < 0.05$, after multi-hypothesis correction for δ and k values) for all 15 cross-validation datasets, while HotNet had returned significant subnetworks for only 3 out of the 15 datasets. This result demonstrates that HotNet2 has higher power to detect mutated subnetworks with fewer samples.

Next, we compared the performance of HotNet2 and HotNet in detecting putative cancer genes (genes satisfying the “20/20 Rule” (See Supplementary Note Section 8.1.4)) using only 50% of the samples. We summarized the tradeoff in sensitivity vs. specificity of each method using the diagnostic odds ratio (DOR):

$$DOR = \frac{TP/FN}{FP/TN} = \frac{sensitivity \times specificity}{(1 - sensitivity) \times (1 - specificity)}$$

We found that the DORs for the subnetworks found by HotNet2 were significantly higher than those found by HotNet: 14.93 ± 2.81 for HotNet2 vs. 7.09 ± 1.08 for HotNet (Supplementary Figure 28(a)), demonstrating that HotNet2 maintains a performance advantage using only 50% of the samples. Moreover, for HotNet2 the DOR was typically much lower when using 50% of the samples than using 100% of samples, demonstrating HotNet2’s improved performance with more data. In contrast, the DOR’s for HotNet on 50% of samples (including those DORs from runs that were not significant according to HotNet’s statistical test) fluctuate around the DOR for 100% of samples, showing that HotNet does not show consistent gains in performance with more data. We suspect that the reason is because the benefits of additional data are outweighed in HotNet (See *Online Methods: Comparison of HotNet2 to other algorithms*).

Finally, we performed two-fold cross validation on the DOR. We repeatedly computed the values of the parameters δ and the subnetwork size k from one half of the data (training set) and then used these values of δ and k to compute subnetworks in the other half of the data (test set). We found that the relative change in DOR (Supplementary Figure 28(b)) was much smaller for HotNet2 (-0.0078 ± 0.257) compared to HotNet

the delta with the largest number of k with $P < 0.05$, and chose the k with $P < 0.05$ with the maximum Youden’s index [44] ($J = sensitivity + specificity - 1$).

(-0.481 ± 0.814) demonstrating that HotNet2 is more stable than HotNet.

8 Comparison of HotNet2 to other approaches

8.1 Pathway and gene set analysis

We performed a comparison of the HotNet2 results with standard pathway-based enrichment analyses. Specifically, we used the DAVID [45, 46] and GSEA [47, 48] tools to analyze the genes input to HotNet2 for enrichment in known pathways and gene sets.

These comparisons show that pathway and gene set enrichment approaches produce results that are both qualitatively and quantitatively very different from HotNet2. HotNet2 reports a small number of subnetworks of interacting genes, while gene set methods report large numbers of overlapping and redundant gene sets. The resulting gene sets are difficult to interpret and also do not contain some of the well-known and novel protein complexes identified by HotNet2. The pathway permutation test results also show that many of the subnetworks identified by HotNet2 are significantly enriched for overlap with known pathways/complexes more often than expected by chance. This suggests that HotNet2 provides new insights and is a useful complement (or arguably a replacement for) other pathway tests. We provide details of the pathway and known gene set analysis below.

8.1.1 DAVID analysis

We first analyzed all 11,565 mutated genes in the mutation frequency dataset to HotNet2 using DAVID to search for functional enrichment in pathways from the BBID⁶, BioCARTA⁷, and KEGG [49, 50] databases (Supplementary Table 35(a)). DAVID reported 31 of the tested pathways were enriched with FDR < 0.05, 5 of which include “cancer” in their name. These pathways overlapped considerably: the 31 pathways contained 1484 distinct genes, with each gene being contained in an average of 1.6 pathways. We repeated this analysis with the top 200 most mutated genes and DAVID reported 20 of the tested pathways were enriched. These 20 pathways contained only 46 distinct genes, with each gene being contained in an average of 6.4 pathways. Furthermore, none of the pathways found with either gene list overlap some of the subnetworks (complexes and pathways) discovered by HotNet2 with known (or posited) roles in cancer. For example, the pathways include no genes from the BAP1, condensin, or SWI/SNF complexes.

8.1.2 GSEA analysis

We also used the GSEA algorithm [47, 48] to identify known gene sets with an overrepresentation of highly mutated genes. We applied GSEA to the list of 11,565 genes in the mutation frequency dataset to HotNet2 ranked by their mutation frequency, and searched the MSigDB curated gene sets [48]. We restricted the analysis to the 9,270 gene sets of sizes 5-200 that included genes in the input list. GSEA identified 190 enriched gene sets with FDR < 0.05, 17 of which included “cancer” in their name (Supplementary Table 35(b)). The significant gene sets overlap considerably, as they include 2028 distinct genes with an average of 2.5 pathways per gene. While the significant gene sets are more comprehensive than the pathways identified by DAVID in terms of the subnetworks (complexes and pathways) identified by HotNet2, they still do not include well-known cancer genes such as *ARID1A*, *MDM4*, or *MLL2*, and completely miss HotNet2 subnetworks such as the ASCOM complex.

⁶<http://bbid.grc.nia.nih.gov/>

⁷<http://www.biocarta.com/genes/index.asp>

8.1.3 Pathway permutation test

We compared the functional enrichment of HotNet2 subnetworks to that of random subnetworks as measured by overlap with known pathways and protein complexes. We computed pathway/complex enrichment using KEGG pathways and PINdb [51] complexes. We call a subnetwork “enriched” if it has a corrected hypergeometric P -value less than 0.05 for any pathway/complex, where we Bonferroni-correct by the product of the number of tested pathways and number of tested subnetworks. We tested each subnetwork identified by HotNet2 for enrichment against each complex and pathway. We find that 15/30 subnetworks are enriched (Supplementary Table 36). The 15 subnetworks without enrichment were mostly small subnetworks (13/15 had size ≤ 7) where enrichment with large pathways is difficult. We then computed an empirical P -value for each of these enriched subnetworks by comparing the minimum corrected P -value (over all tested pathways) for a subnetwork against the minimum corrected P -value (over all tested pathways) of 1000 random subnetworks of the same size drawn from the same protein-protein interaction network in which HotNet2 identified the subnetwork. We find that 14/15 of the enriched subnetworks have an empirical P -value of less than 0.05.

8.1.4 Identifying putative cancer genes

We compared the performance of HotNet2 in identifying putative cancer genes to the performance of HotNet and two pathway-based approaches, GSEA and DAVID. Because there is not a comprehensive list of cancer genes to use as a gold-standard, we constructed a list of putative cancer genes following the “20/20 Rule” recently proposed by Vogelstein, *et al.* [17]. The 20/20 rule posits that cancer genes contain either a surprising cluster of missense mutations or enrichment for inactivating mutations. We constructed a list of 278 such “20/20” genes from the genes in the mutation frequency dataset that were mutated in $\geq 1\%$ of TCGA samples. The list was constructed in two steps. First, we identified 155 genes with significant ($P < 5 \times 10^{-5}$) clusters of missense mutations using the NMC algorithm [18]. Next, we identified 123 additional genes that have $\geq 20\%$ of their mutations as inactivating (nonsense, frame shift indels, nonstop, or splice site mutations).

We compared the sensitivity and specificity of HotNet2, HotNet, DAVID, and GSEA in identifying these 278 20/20 genes. Note that we did not include gene-centric methods such as MuSiC, MutSigCV, or Oncodrive in this comparison, because these methods use clustering of missense mutations and/or enrichment of inactivating mutations in deriving their gene scores. In contrast, HotNet2, HotNet, and pathway-based approaches do not consider these signals, and thus the 20/20 gene list provides an independent evaluation of these methods.

We compared the output of HotNet2 and HotNet on the mutation frequency dataset and HINT+HI2012 network to the results of the GSEA and DAVID algorithms applied to the 11,565 genes in that dataset. We found that HotNet2 had a false positive rate (FPR) that was 52%, 61%, 83% lower than HotNet, GSEA, and DAVID, respectively, at equal sensitivity (Supplementary Table 34(a)). To evaluate the performance of each method over a range of sensitivities/specificities, we computed the receiver operator characteristic (ROC) curves for each method by varying the parameter δ for HotNet and HotNet2, and the FDR for DAVID and GSEA (Supplementary Figure 27(a)). The ROC curves demonstrate that HotNet2 dominates the other approaches in the range of low FPR ($FPR < 0.1$) that is typically used in the identification of cancer genes [52]. At higher FPR, the HotNet algorithm begins to outperform the other approaches (Supplementary Figure 27(b)). However, this result should be viewed with caution for two reasons: (1) $FPR > 0.1$ corresponds to more than 1100 false positive genes; (2) these results will not be significant according to HotNet’s statistical test.

The ROC curves provide a comprehensive comparison of the tradeoff in sensitivity vs. specificity of the different methods; summarizing this tradeoff in a single number entails some loss of information [53].

Nevertheless, to provide such a number, we computed the partial area under the curve (pAUC), a standard measure for summarizing the ROC curve in a reduced FPR range. We computed the pAUC up to $FPR = 0.1$ and standardized the pAUC so that it is 1 for a perfect predictor and 0.5 for a random predictor as in [54]. HotNet2 has an 4%, 4%, and 13% higher standardized pAUC (Supplementary Table 34(b)) than HotNet, GSEA, and DAVID, respectively, demonstrating the advantages of HotNet2 over HotNet and pathway based methods for predicting cancer genes *and* their combinations.

We also test whether HotNet2’s advantage over DAVID and GSEA was due solely to HotNet2’s (and HotNet’s) restriction to genes only in the HINT+HI2012 PPI network. We compared HotNet2, HotNet, DAVID, and GSEA on the reduced set of 6930 genes in HINT+HI2012 that mutated in at least one sample in the mutation dataset. HotNet2 again outperformed the other algorithms, with a false positive rate (FPR) that was 52%, 46%, and 55% lower than HotNet, GSEA, and DAVID, respectively, at equal sensitivity (Supplementary Table 38 (a)). The ROC curve for HotNet2 dominated other methods, both at $FPR < 0.1$ (Supplementary Figure 27(c)) and $FPR < 0.3$ (Supplementary Figure 27(d)). For $FPR < 0.1$, HotNet2’s pAUC was 4%, 2%, and 8% higher than HotNet, GSEA, and DAVID, respectively (Supplementary Table 34 (b)).

8.2 Comparison to MEMo

We attempted to run MEMo [55] on the 3110 samples in the Pan-Cancer dataset using SNVs and CNAs from the mutation frequency dataset. However, MEMo was not able to run on a dataset this large. In particular, MEMo reported that it was using 11,179 genes after filtering, but terminated with a heap error after 6 minutes on a large memory machine. We note that the challenge of running MEMo on large datasets is reported by the authors who recommend “that the list of recurrently altered genes be kept below 100 [for optimal results]” [56].

8.3 Comparision to Ciriello *et al.* [57] and Lawrence *et al.* [22]

We compare our analysis to the Pan-Cancer study from Ciriello *et al.* [57]. We emphasize that our study and Ciriello *et al.* have very different goals. Ciriello *et al.* focus on the clustering of samples in the Pan-Cancer dataset according to the presence/absence of individually significant mutations. While Ciriello *et al.* manually annotate a few known pathways, they do not assess the clustering of mutations in these pathways, nor do they attempt to identify novel combinations. Specifically, Ciriello *et al.* limit attention to 479 significant mutation events. Only 209 of these events are single gene mutations (or promoter methylation) with the remainder being copy number aberrations that affect many genes. In addition, 36 of the 479 events do not occur in any of the 3299 samples analyzed in Ciriello *et al.*, according to their mutation matrix. Finally, the clustering of samples further restricts attention to only those events that appear in $\geq 1\%$ of the samples; this implies that only 164 events are used in the first stage of the clustering when all samples are considered. While the mutations considered in Ciriello *et al.* are useful for clustering samples, Ciriello *et al.* do not claim that they form a complete list of driver mutations in these samples.

We also note that Ciriello *et al.* analyze a slightly different dataset than the dataset analyzed by us and several other TCGA Pan-Cancer projects. In particular, compared to the official TCGA Pan-Cancer data freeze [58], Ciriello *et al.* analyze:

- 284 fewer BRCA samples;
- 268 more COADREAD samples; and,
- 131 more OV samples

with other cancers having smaller discrepancies. 2639 samples are common to the datasets.

We also compare our study to Lawrence *et al.* [22], who also recently performed a Pan-Cancer study. Lawrence *et al.* predict cancer genes from somatic SNVs – not including any CNAs – in a larger cohort of samples from 21 cancer types, containing more than 2,000 tumors outside of the TCGA Pan-Cancer cohort. Similar to our analysis, Lawrence *et al.* identify both known and novel cancer genes, and there is overlap between the novel genes identified by both approaches (e.g. *ELF3*, *HLA-A*, *ARID2*, and *ASXL2*). Notably, since HotNet2 analyzes combinations of mutations, we identify these genes as significant using fewer samples. We also identify additional candidate cancer genes not reported in Lawrence *et al.*, but with supporting evidence of their function in cancer from protein interactions and mutation clustering.

References

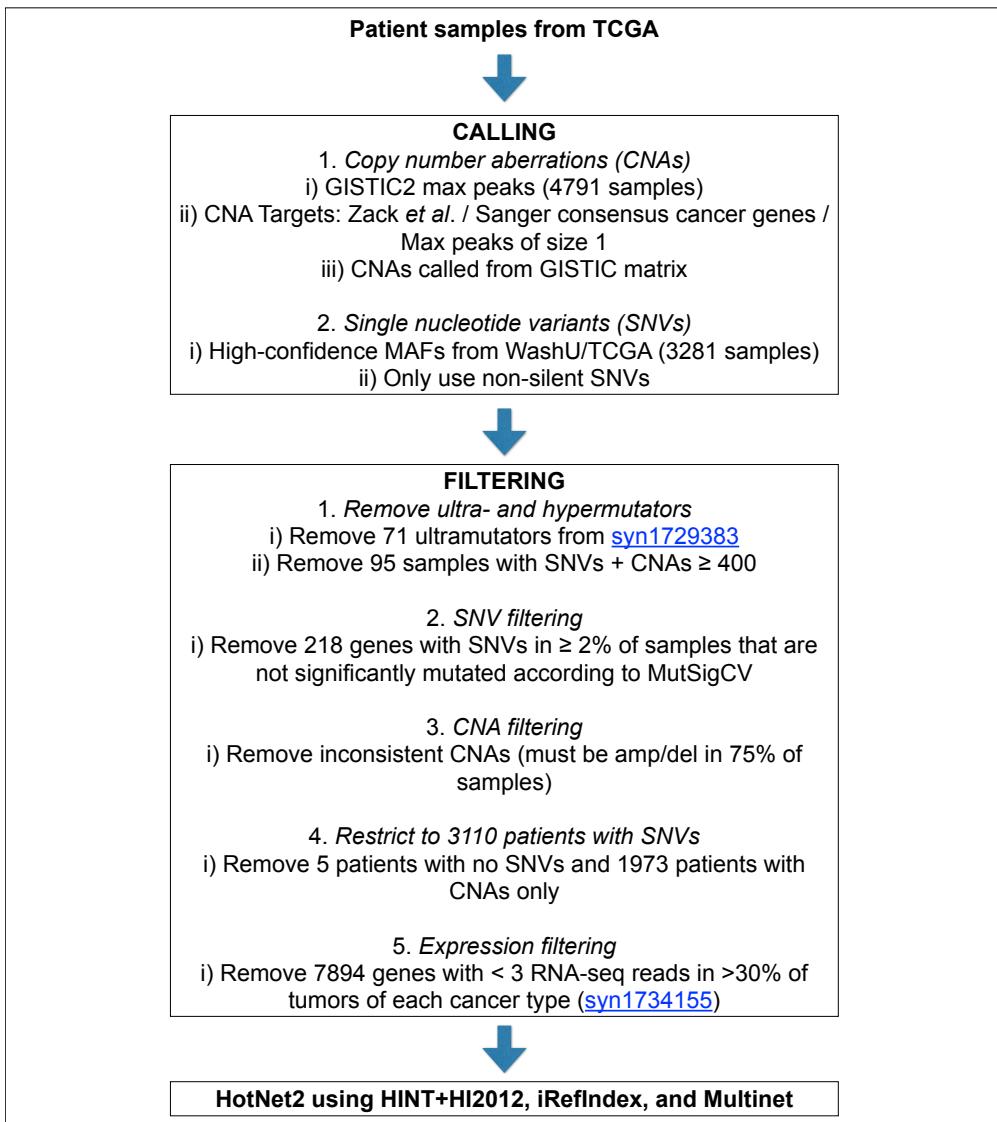
1. Vandin F, Upfal E, Raphael B (2011) Algorithms for detecting significantly mutated pathways in cancer. *Journal of Computational Biology* 18: 507–22.
2. Vandin F, Clay P, Upfal ELI, Raphael BJ (2012) Discovery of Mutated Subnetworks Associated with Clinical Data in Cancer. In: Pacific Symposium on Biocomputing. Kohala Coast, Hawaii, pp. 55–66.
3. Köhler S, Bauer S, Horn D, Robinson PN (2008) Walking the interactome for prioritization of candidate disease genes. *American journal of human genetics* 82: 949–58.
4. Müller T, Rahmann S, Rehmsmeier M (2001) Non-symmetric score matrices and the detection of homologous transmembrane proteins. *Bioinformatics* (Oxford, England) 17 Suppl 1: S182–9.
5. Milo R, Kashtan N, Itzkovitz S, Newman MEJ, Alon U (2003) On the uniform generation of random graphs with prescribed degree sequences. at <http://arxiv.org/abs/condmat/0312028> .
6. Lawrence MS, Stojanov P, Polak P, Kryukov GV, Cibulskis K, et al. (2013) Mutational heterogeneity in cancer and the search for new cancer-associated genes. *Nature* .
7. Gonzalez-Perez A, Lopez-Bigas N (2012) Functional impact bias reveals cancer drivers. *Nucleic Acids Res* 40: e169.
8. Tamborero D, Lopez-Bigas N, Gonzalez-Perez A (2013) Oncodrive-cis: a method to reveal likely driver genes based on the impact of their copy number changes on expression. *PLoS One* 8: e55489.
9. Dees ND, Zhang Q, Kandoth C, Wendl MC, Schierding W, et al. (2012) Music: identifying mutational significance in cancer genomes 22: 1589–98.
10. Kandoth C, McLellan MD, Vandin F, Ye K, Niu B, et al. (2013) Mutational landscape and significance across 12 major cancer types. *Nature* 502: 333–9.
11. Zack TI, Schumacher SE, Carter SL, Cherniack AD, Saksena G, et al. (2013) Pan-cancer patterns of somatic copy number alteration. *Nature Genetics* 45: 1134–1140.
12. Forbes SA, Bindal N, Bamford S, Cole C, Kok CY, et al. (2011) Cosmic: mining complete cancer genomes in the catalogue of somatic mutations in cancer. *Nucleic acids research* 39: D945–D950.
13. Das J, Yu H (2012) Hint: High-quality protein interactomes and their applications in understanding human disease. *BMC systems biology* 6: 92.

14. Yu H, Tardivo L, Tam S, Weiner E, Gebreab F, et al. (2011) Next-generation sequencing to generate interactome datasets. *Nature methods* 8: 478–480.
15. Khurana E, Fu Y, Chen J, Gerstein M (2013) Interpretation of genomic variants using a unified biological network approach. *PLoS computational biology* 9: e1002886.
16. Razick S, Magklaras G, Donaldson IM (2008) iRefIndex: a consolidated protein interaction database with provenance. *BMC bioinformatics* 9: 405.
17. Vogelstein B, Papadopoulos N, Velculescu VE, Zhou S, Diaz LA, et al. (2013) Cancer genome landscapes. *Science (New York, NY)* 339: 1546–58.
18. Ye J, Pavlicek A, Lunney EA, Rejto PA, Teng CH (2010) Statistical method on nonrandom clustering with application to somatic mutations in cancer. *BMC bioinformatics* 11: 11.
19. Ryslik GA, Cheng Y, Cheung KH, Modis Y, Zhao H (2013) Utilizing protein structure to identify non-random somatic mutations. *BMC Bioinformatics* 14: 190.
20. Lancaster H (1961) Significance tests in discrete distributions. *Journal of the American Statistical Association* 56: 223–234.
21. Davis TL, Walker JR, Loppnau P, Butler-Cole C, Allali-Hassani A, et al. (2008) Autoregulation by the juxtamembrane region of the human ephrin receptor tyrosine kinase A3 (EphA3). *Structure (London, England : 1993)* 16: 873–84.
22. Lawrence MS, Stojanov P, Mermel CH, Robinson JT, Garraway La, et al. (2014) Discovery and saturation analysis of cancer genes across 21 tumour types. *Nature* 505: 495–501.
23. Zada AA, Pulikkan JA, Bararia D, Geletu M, Trivedi AK, et al. (2006) Proteomic discovery of max as a novel interacting partner of c/ebpalpha: a myc/max/mad link. *Leukemia* 20: 2137-46.
24. Gui Y, Guo G, Huang Y, Hu X, Tang A, et al. (2011) Frequent mutations of chromatin remodeling genes in transitional cell carcinoma of the bladder. *Nat Genet* 43: 875-8.
25. Bott M, Brevet M, Taylor BS, Shimizu S, Ito T, et al. (2011) The nuclear deubiquitinase bap1 is commonly inactivated by somatic mutations and 3p21.1 losses in malignant pleural mesothelioma. *Nat Genet* 43: 668-72.
26. Ventii KH, Devi NS, Friedrich KL, Chernova TA, Tighiouart M, et al. (2008) Brca1-associated protein-1 is a tumor suppressor that requires deubiquitinating activity and nuclear localization. *Cancer Res* 68: 6953-62.
27. Carbone M, Yang H, Pass HI, Krausz T, Testa JR, et al. (2013) BAP1 and cancer. *Nature reviews Cancer* 13: 153–9.
28. Marais A, Ji Z, Child ES, Krause E, Mann DJ, et al. (2010) Cell cycle-dependent regulation of the forkhead transcription factor FOXK2 by CDKcyclin complexes. *The Journal of biological chemistry* 285: 35728–39.
29. Kato M (2013) Functional and cancer genomics of asxl family members. *Br J Cancer* .
30. Sahu N, Grandis JR (2011) New advances in molecular approaches to head and neck squamous cell carcinoma. *Anticancer Drugs* 22: 656-64.

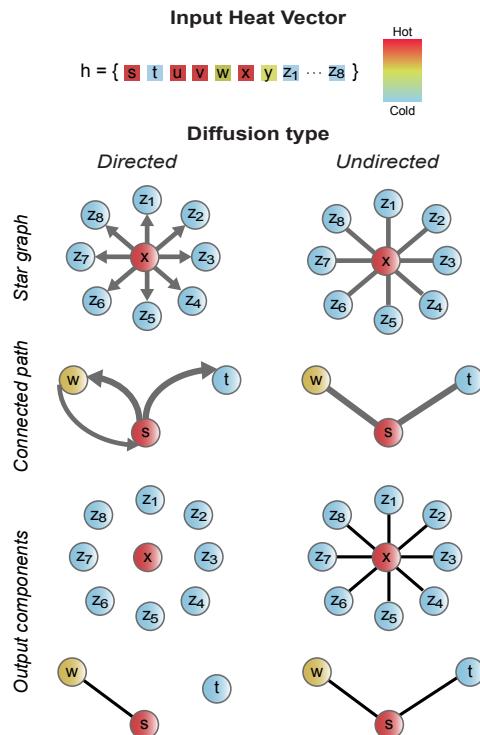
31. Mo Q, Wang S, Seshan VE, Olshen AB, Schultz N, et al. (2013) Pattern discovery and cancer gene identification in integrated cancer genomic data. *Proceedings of the National Academy of Sciences of the United States of America* 110: 4245–50.
32. Abdel-Wahab O, Adli M, LaFave LM, Gao J, Hricik T, et al. (2012) ASXL1 mutations promote myeloid transformation through loss of PRC2-mediated gene repression. *Cancer cell* 22: 180–93.
33. Bravo J, Li Z, Speck NA, Warren AJ (2001) The leukemia-associated AML1 (Runx1)-CBF beta complex functions as a DNA-induced molecular clamp. *Nature structural biology* 8: 371–8.
34. Cammenga J, Niebuhr B, Horn S, Bergholz U, Putz G, et al. (2007) RUNX1 DNA-binding mutants, associated with minimally differentiated acute myelogenous leukemia, disrupt myeloid differentiation. *Cancer research* 67: 537–45.
35. Janes KA (2011) RUNX1 and its understudied role in breast cancer. *Cell cycle (Georgetown, Tex)* 10: 3461–5.
36. Huang HC, Nguyen T, Pickett CB (2000) Regulation of the antioxidant response element by protein kinase c-mediated phosphorylation of nf-e2-related factor 2. *Proc Natl Acad Sci U S A* 97: 12475–80.
37. Cancer Genome Atlas Research Network (2012) Comprehensive genomic characterization of squamous cell lung cancers. *Nature* 489: 519–25.
38. Sporn MB, Liby KT (2012) Nrf2 and cancer: the good, the bad and the importance of context. *Nat Rev Cancer* 12: 564–71.
39. Shibata T, Ohta T, Tong KI, Kokubu A, Odogawa R, et al. (2008) Cancer related mutations in nrf2 impair its recognition by keap1-cul3 e3 ligase and promote malignancy. *Proc Natl Acad Sci U S A* 105: 13568–73.
40. Itoh K, Wakabayashi N, Katoh Y, Ishii T, Igarashi K, et al. (1999) Keap1 represses nuclear activation of antioxidant responsive elements by nrf2 through binding to the amino-terminal neh2 domain. *Genes Dev* 13: 76–86.
41. Dawkins R, Leelayuwat C, Gaudieri S, Tay G, Hui J, et al. (1999) Genomics of the major histocompatibility complex: haplotypes, duplication, retroviruses and disease. *Immunol Rev* 167: 275–304.
42. Drabek K, van Ham M, Stepanova T, Draegestein K, van Horssen R, et al. (2006) Role of CLASP2 in microtubule stabilization and the regulation of persistent motility. *Current biology : CB* 16: 2259–64.
43. Maitra a, Wistuba II, Washington C, Virmani aK, Ashfaq R, et al. (2001) High-resolution chromosome 3p allelotyping of breast carcinomas and precursor lesions demonstrates frequent loss of heterozygosity and a discontinuous pattern of allele loss. *The American journal of pathology* 159: 119–30.
44. Youden W (1950) Index for rating diagnostic tests. *Cancer* 3: 32–35.
45. Huang DW, Sherman BT, Lempicki RA (2009) Bioinformatics enrichment tools: paths toward the comprehensive functional analysis of large gene lists. *Nucleic Acids Res* 37: 1–13.
46. Huang DW, Sherman BT, Lempicki RA (2009) Systematic and integrative analysis of large gene lists using david bioinformatics resources. *Nat Protoc* 4: 44–57.

47. Mootha VK, Lindgren CM, Eriksson KF, Subramanian A, Sihag S, et al. (2003) PGC-1alpha-responsive genes involved in oxidative phosphorylation are coordinately downregulated in human diabetes. *Nature genetics* 34: 267–73.
48. Subramanian A, Tamayo P, Mootha VK, Mukherjee S, Ebert BL, et al. (2005) Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. *Proceedings of the National Academy of Sciences of the United States of America* 102: 15545–50.
49. Kanehisa M, Goto S (2000) Kegg: kyoto encyclopedia of genes and genomes. *Nucleic Acids Res* 28: 27-30.
50. Kanehisa M (2013) Molecular network analysis of diseases and drugs in kegg. *Methods Mol Biol* 939: 263-75.
51. Luc PV, Tempst P (2004) PINdb: a database of nuclear protein complexes from human and yeast. *Bioinformatics (Oxford, England)* 20: 1413–5.
52. Hanahan D, Weinberg RA (2011) Hallmarks of cancer: the next generation. *Cell* 144: 646–74.
53. JM L, A JV, R R (2008) AUC: a misleading measure of the performance of predictive distribution models. *Global Ecology and Biogeography* 17: 141-145.
54. DK M (1989) Analyzing a portion of the ROC curve. *Medical Decision Making* 9: 190-195.
55. Ciriello G, Cerami EG, Sander C, Schultz N (2011) Mutual exclusivity analysis identifies oncogenic network modules. *Genome Research* 22: 3980496.
56. Ciriello G, Cerami E, Aksoy BA, Sander C, Schultz N (2013) Using MEMo to discover mutual exclusivity modules in cancer. *Current protocols in bioinformatics / editorial board, Andreas D Baxevanis [et al]* 41: 8.17.2.
57. Ciriello G, Miller ML, Aksoy BA, Senbabaoglu Y, Schultz N, et al. (2013) Emerging landscape of oncogenic signatures across human cancers. *Nature Genetics* 45: 1127–1133.
58. Weinstein JN, Collisson Ea, Mills GB, Shaw KRM, Ozenberger Ba, et al. (2013) The Cancer Genome Atlas Pan-Cancer analysis project. *Nature genetics* 45: 1113–20.

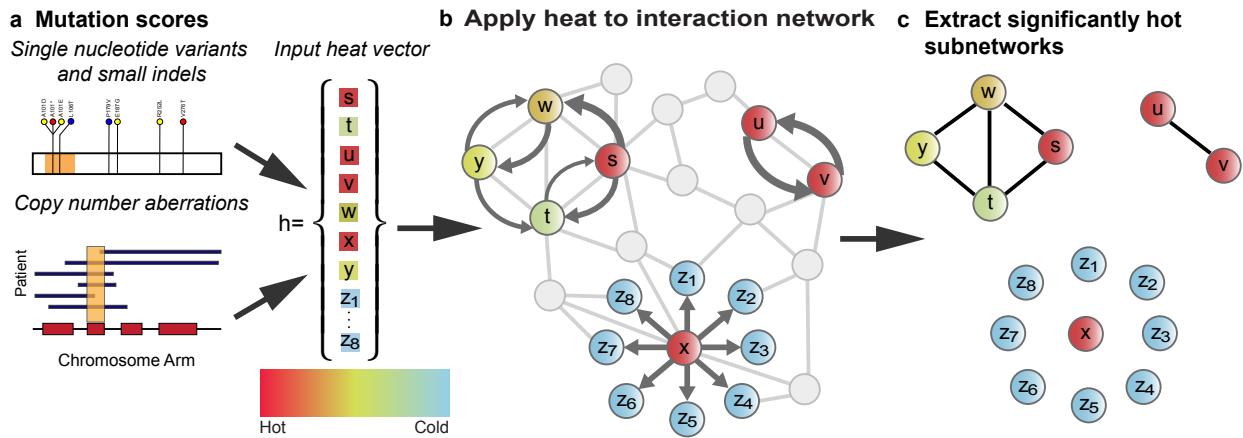
Supplementary Figures



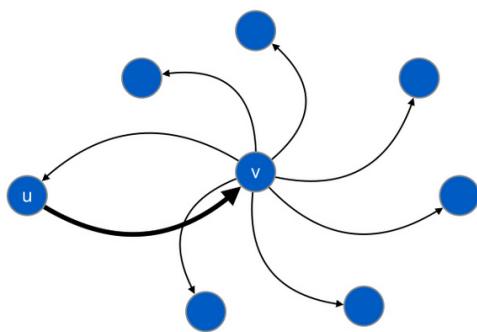
Supplementary Figure 1: HotNet2 mutation data processing.



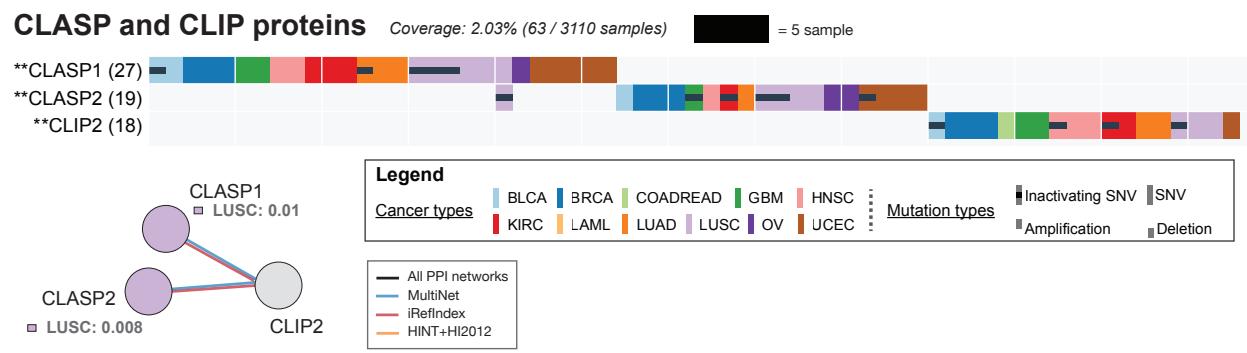
Supplementary Figure 2: Comparison of directed and undirected diffusion kernels. (Left) HotNet2 identifies strongly connected components in a directed graph with edges weighted by a directed diffusion kernel. By using directed edges, HotNet2 will not output star graphs around a single "hot" node (*star graph*), nor will it include cold nodes in larger components (*connected path*). (Right) HotNet identifies connected components in an undirected graph with edges weighted by an undirected diffusion kernel, and therefore identifies more components comprised of more cold nodes.



Supplementary Figure 3: Overview of HotNet2 algorithm and Pan-Cancer analysis. HotNet2 assigns heat to each gene (node) in an interaction network according to a gene score encoding the frequency and/or predicted functional impact of mutations in the gene. This heat spreads to neighboring nodes using an insulated heat diffusion process. At the equilibrium heat distribution, the network is partitioned into subnetworks according to the amount and direction of heat exchange between pairs of nodes. Thus, the partition depends on both the individual genes scores and the local topology of protein interactions. The statistical significance (p -value and FDR) for the resulting subnetworks is computed using the same procedure on random data. In our TCGA Pan-Cancer analysis, gene scores are computed according to single nucleotide variants, small indels, and splice site mutations (from exome sequencing data), copy number aberrations (from SNP array data), and gene expression (from RNA-seq data).



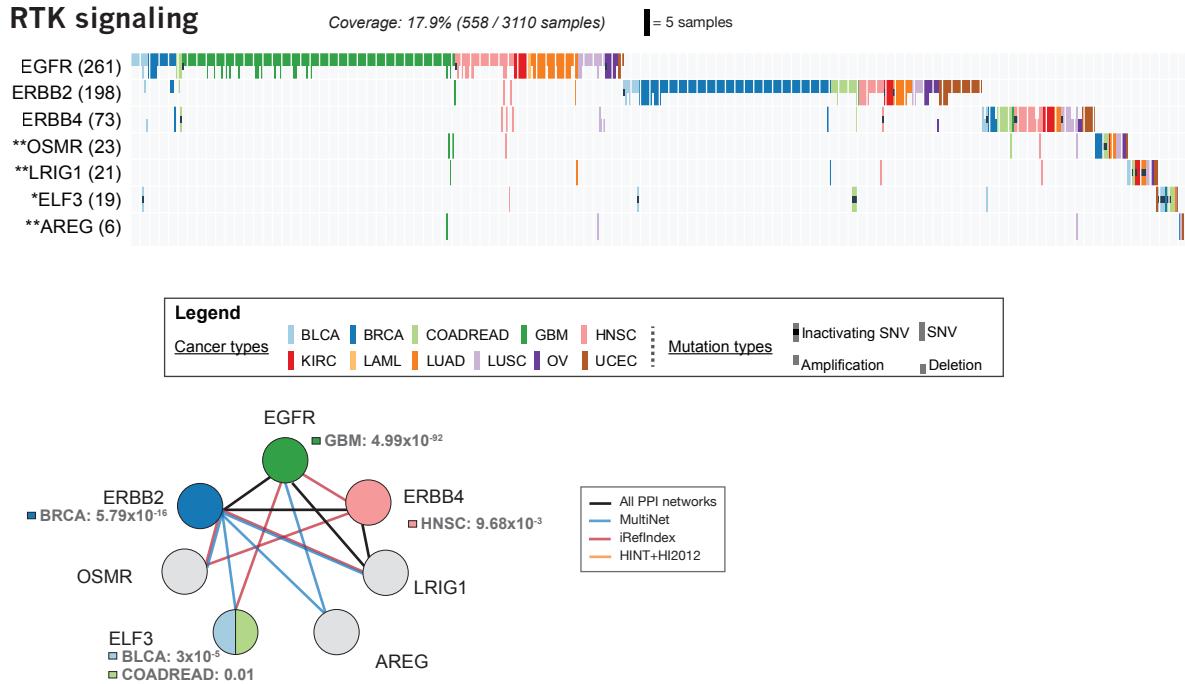
Supplementary Figure 4: HotNet2 similarity between neighbors in a small graph. Node u has degree one, so sends most of its heat to its one neighbor v . Node v has multiple neighbors, and therefore sends less of its heat to each of its neighbors.



Supplementary Figure 5: The CLASP and CLIP proteins subnetwork and its mutation matrix. The mutation matrix shows the mutations in the genes in the subnetwork, represented on the left, across the samples with mutations in the subnetwork. The number of samples with mutations in a gene is in parenthesis. The number of samples with mutations in any of the members of the subnetwork is on the top. Upticks and downticks represent amplifications and deletions, respectively. Full ticks represent SNVs, indels, and splicesite mutations. Colors reflect the cancer type of the samples. Genes with * were significant by exactly one of MuSiC, MutSigCV, or Oncodrive, while genes with ** were not significant by any of these methods. Colors of interactions in the subnetwork represent the interaction network where it was found. *P*-values for cancer type enrichment of mutations in the genes are shown.



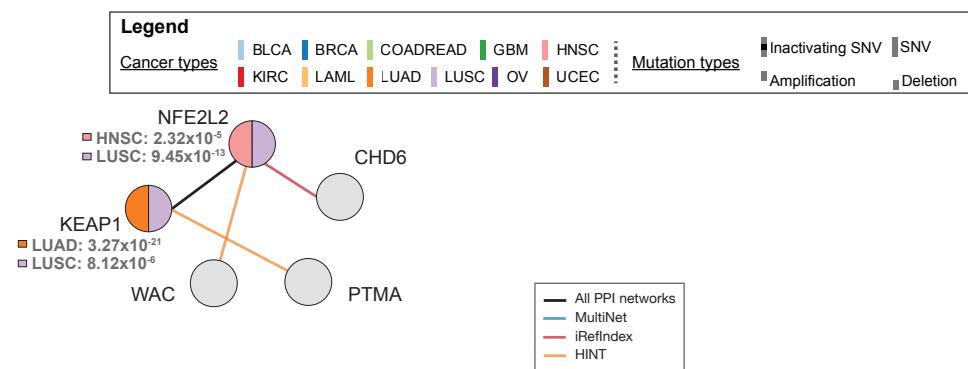
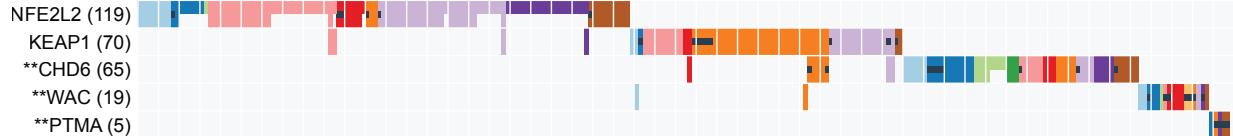
Supplementary Figure 6: Example subnetwork visualization from the HotNet2 web output of the SWI/SNF subnetwork.



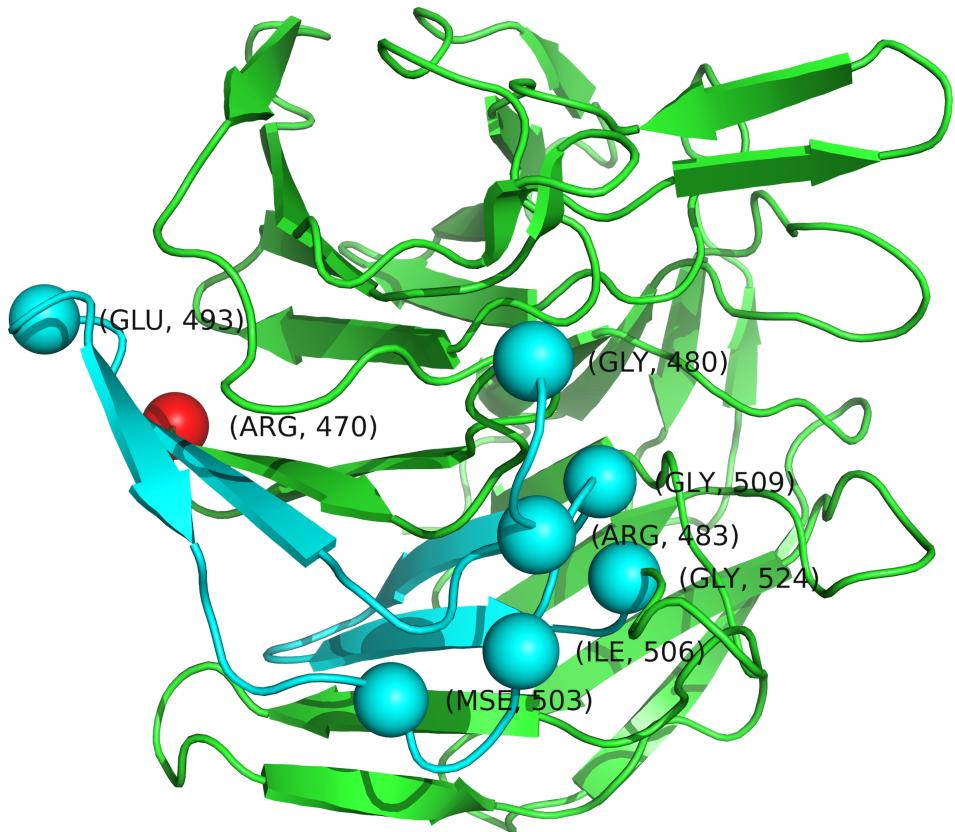
Supplementary Figure 7: The RTK subnetwork and its mutation matrix. The mutation matrix shows the mutations in the genes in the subnetwork, represented on the left, across the samples with mutations in the subnetwork. The number of samples with mutations in a gene is in parenthesis. The number of samples with mutations in any of the members of the subnetwork is on the top. Upticks and downticks represent amplifications and deletions, respectively. Full ticks represent SNVs, indels, and splicesite mutations. Colors reflect the cancer type of the samples. Black dots corresponds to inactivating genes, that is genes that contain at least one of the following mutations: frame shift indels, nonsense, nonstop, and splice sites. Genes with * were significant by exactly one of MuSiC, MutSigCV, or Oncodrive, while genes with ** were not significant by any of these methods. Colors of interactions in the subnetwork represent the interaction network where it was found. P -values for cancer type enrichment of mutations in the genes are shown.

KEAP1, NFE2L2 and interactors

Coverage: 8.4% (264 / 3110 samples)



Supplementary Figure 8: The KEAP-NFE2L2 subnetwork and its mutation matrix. The mutation matrix shows the mutations in the genes in the subnetwork, represented on the left, across the samples with mutations in the subnetwork. The number of samples with mutations in a gene is in parenthesis. The number of samples with mutations in any of the members of the subnetwork is on the top. Upticks and downticks represent amplifications and deletions, respectively. Full ticks represent SNVs, indels, and splicesite mutations. Colors reflect the cancer type of the samples. Black dots corresponds to inactivating genes, that is genes that contain at least one of the following mutations: frame shift indels, nonsense, nonstop, and splice sites. Genes with * were significant by exactly one of MuSiC, MutSigCV, or Oncodrive, while genes with ** were not significant by any of these methods. Colors of interactions in the subnetwork represent the interaction network where it was found. P -values for cancer type enrichment of mutations in the genes are shown.

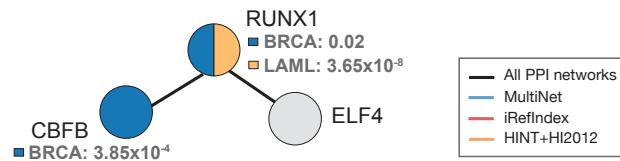
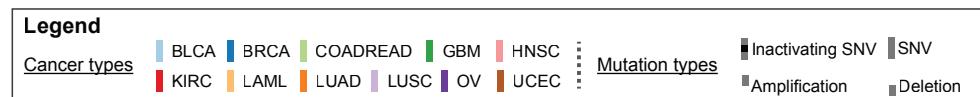
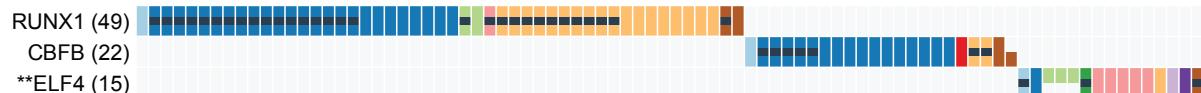


Supplementary Figure 9: The KEAP1 structure (PDB ID: 1ZGK) color coded by segment. Residues that are only in the iPAC most significant cluster (amino acids 480 – 524) are colored in light blue. The residue in the NMC most significant cluster (amino acid 470) is shown in red.

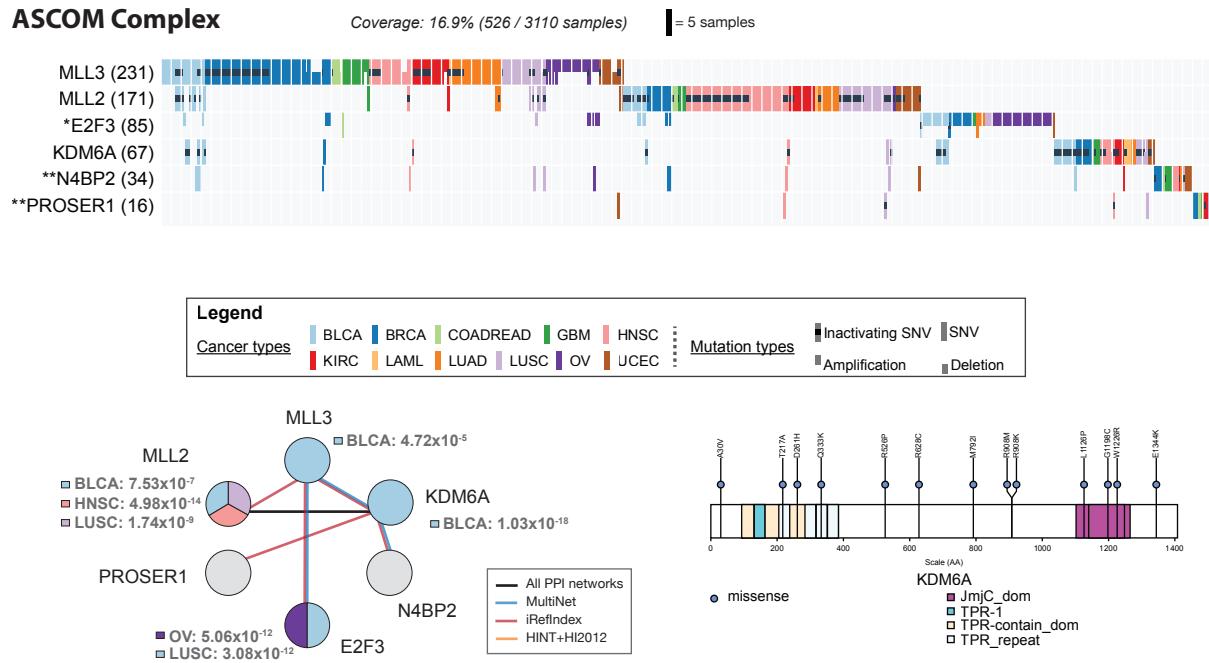
Core-binding factors

Coverage: 2.7% (86/ 3110 samples)

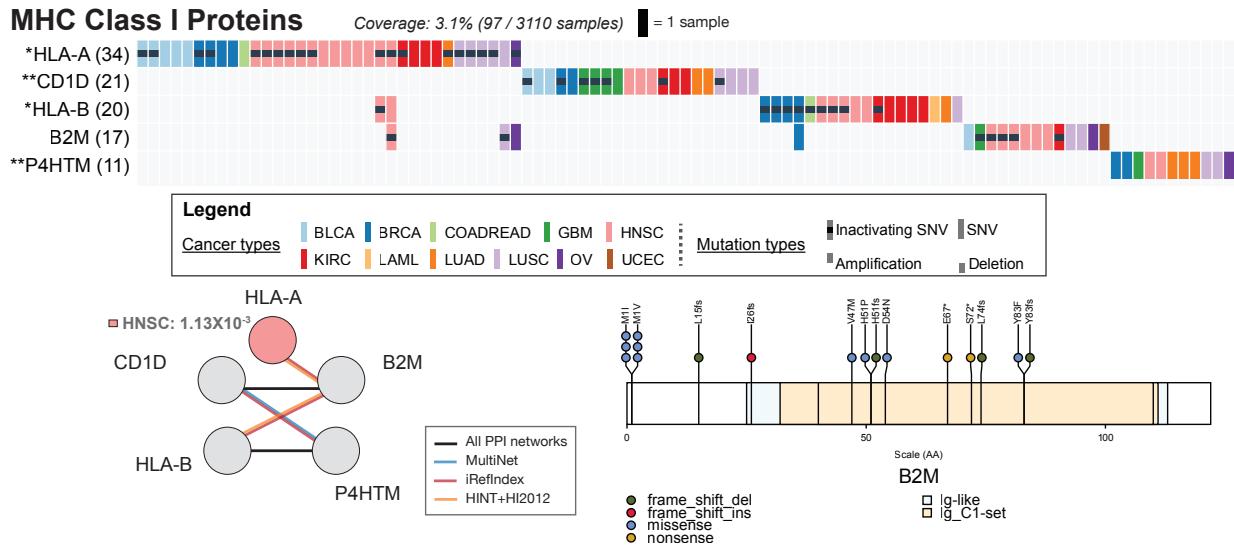
= 1 sample



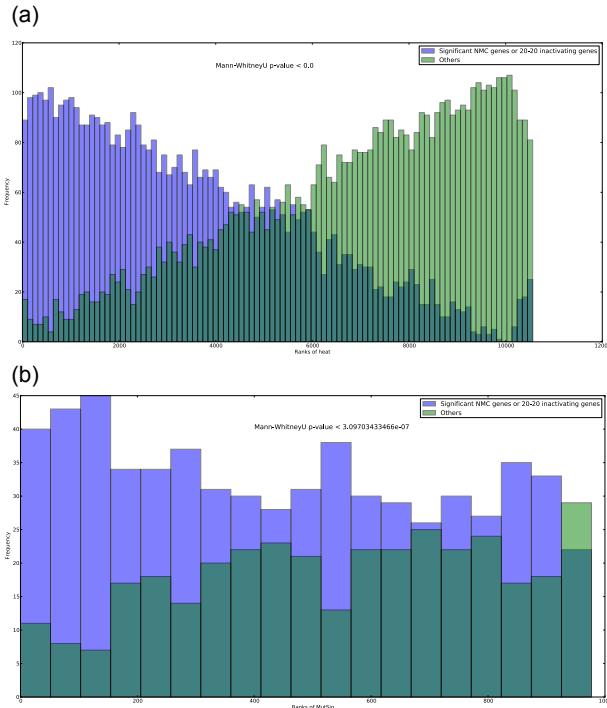
Supplementary Figure 10: The RUNX1-CBFB-ELF3 subnetwork and its mutation matrix. The mutation matrix shows the mutations in the genes in the subnetworks, represented on the left, across the samples with mutations in the subnetwork. The number of samples with mutations in a gene is in parenthesis. The number of samples with mutations in any of the members of the subnetwork is on the top. Upticks and downticks represent amplifications and deletions, respectively. Full ticks represent SNVs, indels, and splicesite mutations. Colors reflect the cancer type of the samples. Black dots corresponds to inactivating genes, that is genes that contain at least one of the following mutations: frame shift indels, nonsense, nonstop, and splice sites. Genes with * were significant by exactly one of MuSiC, MutSigCV, or Oncodrive, while genes with ** were not significant by any of these methods. Colors of interactions in the subnetwork represent the interaction network where it was found. *P*-values for cancer type enrichment of mutations in the genes are shown.



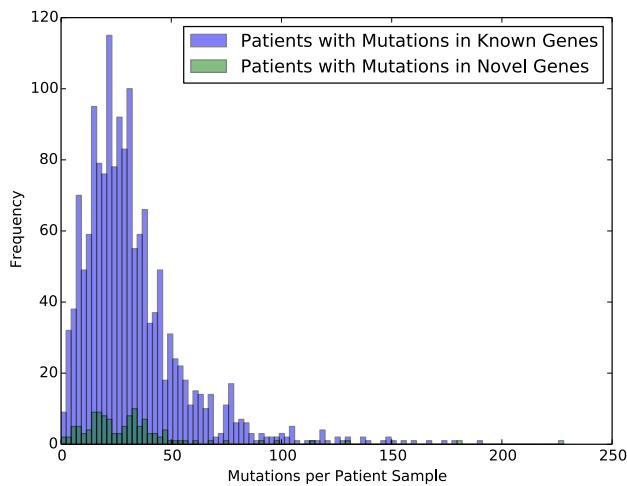
Supplementary Figure 11: The ASCOM subnetwork and its mutation matrix. The mutation matrix shows the mutations in the genes in the subnetworks, represented on the left, across the samples with mutations in the subnetwork. The number of samples with mutations in a gene is in parenthesis. The number of samples with mutations in any of the members of the subnetwork is on the top. Upticks and downticks represent amplifications and deletions, respectively. Full ticks represent SNVs, indels, and splicesite mutations. Colors reflect the cancer type of the samples. Black dots corresponds to inactivating genes, that is genes that contain at least one of the following mutations: frame shift indels, nonsense, nonstop, and splice sites. Genes with * were significant by exactly one of MuSiC, MutSigCV, or Oncodrive, while genes with ** were not significant by any of these methods. Colors of interactions in the subnetwork represent the interaction network where it was found. P -values for cancer type enrichment of mutations in the genes are shown.



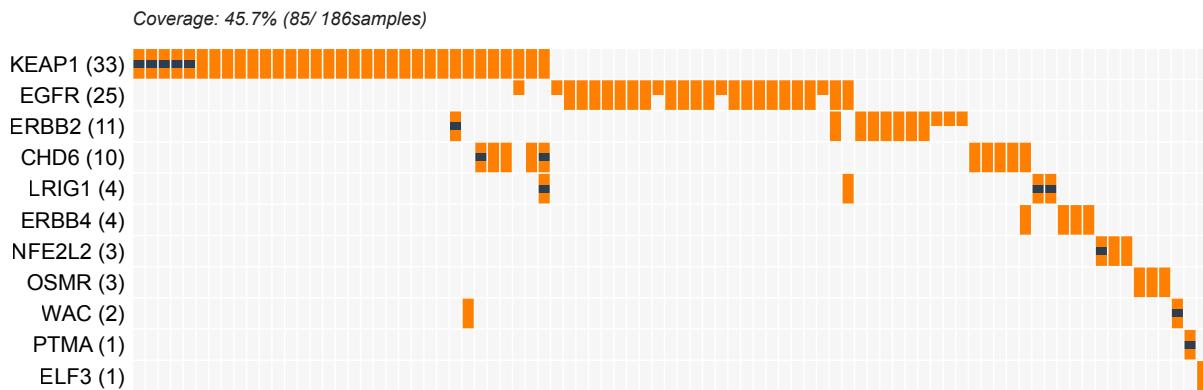
Supplementary Figure 12: The MHC Class I subnetwork and its mutation matrix. The mutation matrix shows the mutations in the genes in the subnetworks, represented on the left, across the samples with mutations in the subnetwork. The number of samples with mutations in a gene is in parenthesis. The number of samples with mutations in any of the members of the subnetwork is on the top. Upticks and downticks represent amplifications and deletions, respectively. Full ticks represent SNVs, indels, and splicesite mutations. Colors reflect the cancer type of the samples. Genes with * were significant by exactly one of MuSiC, MutSigCV, or Oncodrive, while genes with ** were not significant by any of these methods. Colors of interactions in the subnetwork represent the interaction network where it was found. *P*-values for cancer type enrichment of mutations in the genes are shown. The distribution of mutations in *B2M* is shown.



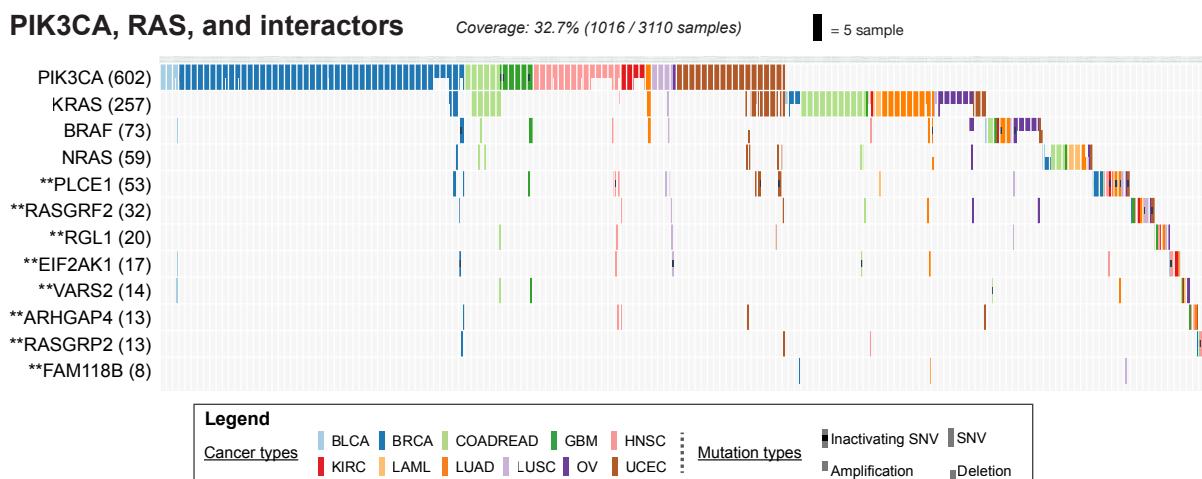
Supplementary Figure 13: Histograms of two groups of genes ranked by heat score: genes in blue have either a cluster of missense mutations (NMC cluster $P < 0.05$) or $\geq 20\%$ of mutations inactivating; genes in green have mutations that meet neither criteria. (a) **Mutation frequency as heat**. The two distributions are significantly different ($P < 10^{-20}$, Mann Whitney U test), with higher-scoring genes showing enrichment for mutation clustering or high percentages of inactivating mutations. (b) **MutSigCV scores as heat**. The two distributions are significantly different ($P = 3.09 \times 10^{-7}$, Mann Whitney U test), with higher-scoring genes showing enrichment for mutation clustering or high percentages of inactivating mutations.



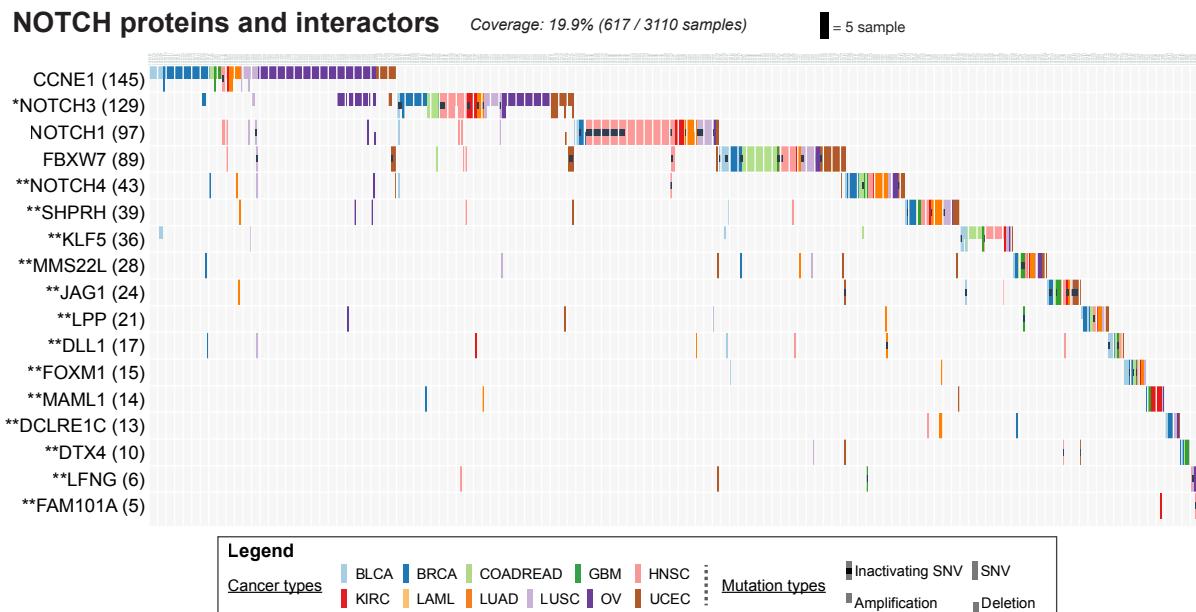
Supplementary Figure 14: Histograms of the number of mutations per patient for two gene groups: novel genes identified by HotNet2, and a list of known cancer or significantly mutated cancer genes. The two distributions are not significantly different ($P = 0.4$, Mann Whitney U test).



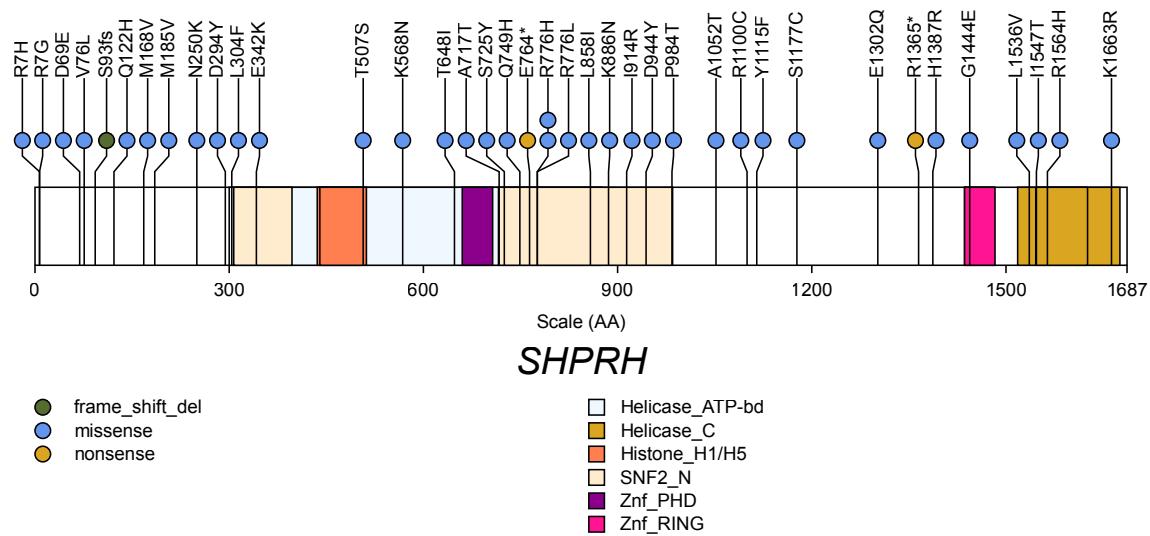
Supplementary Figure 15: A mutation matrix in LUAD using genes in subnetworks, ErbB signaling and *KEAP1*, *NFE2L2* and interactors.



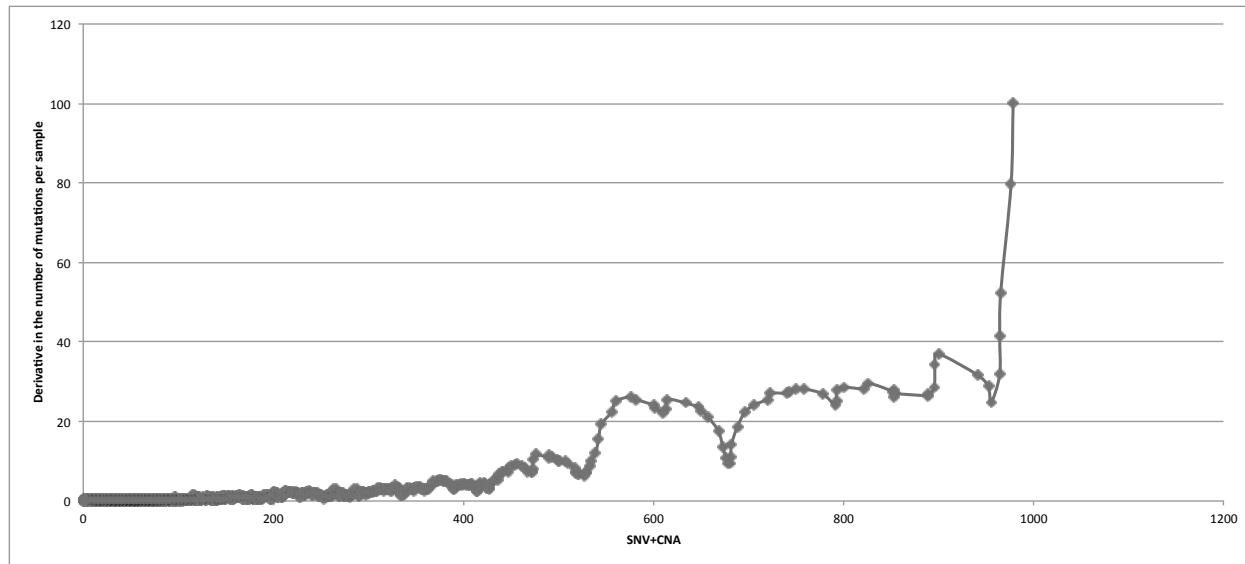
Supplementary Figure 16: The PI(3)K/RAS subnetwork and its mutation matrix. The mutation matrix shows the mutations in the genes in the subnetworks, represented on the left, across the samples with mutations in the subnetwork. The number of samples with mutations in a gene is in parenthesis. The number of samples with mutations in any of the members of the subnetwork is on the top. Upticks and downticks represent amplifications and deletions, respectively. Full ticks represent SNVs, indels, and splicesite mutations. Colors reflect the cancer type of the samples. Black dots corresponds to inactivating genes, that is genes that contain at least one of the following mutations: frame shift indels, nonsense, nonstop, and splice sites. Genes with * were significant by exactly one of MuSiC, MutSigCV, or Oncodrive, while genes with ** were not significant by any of these methods. Colors of interactions in the subnetwork represent the interaction network where it was found. *P*-values for cancer type enrichment of mutations in the genes are shown.



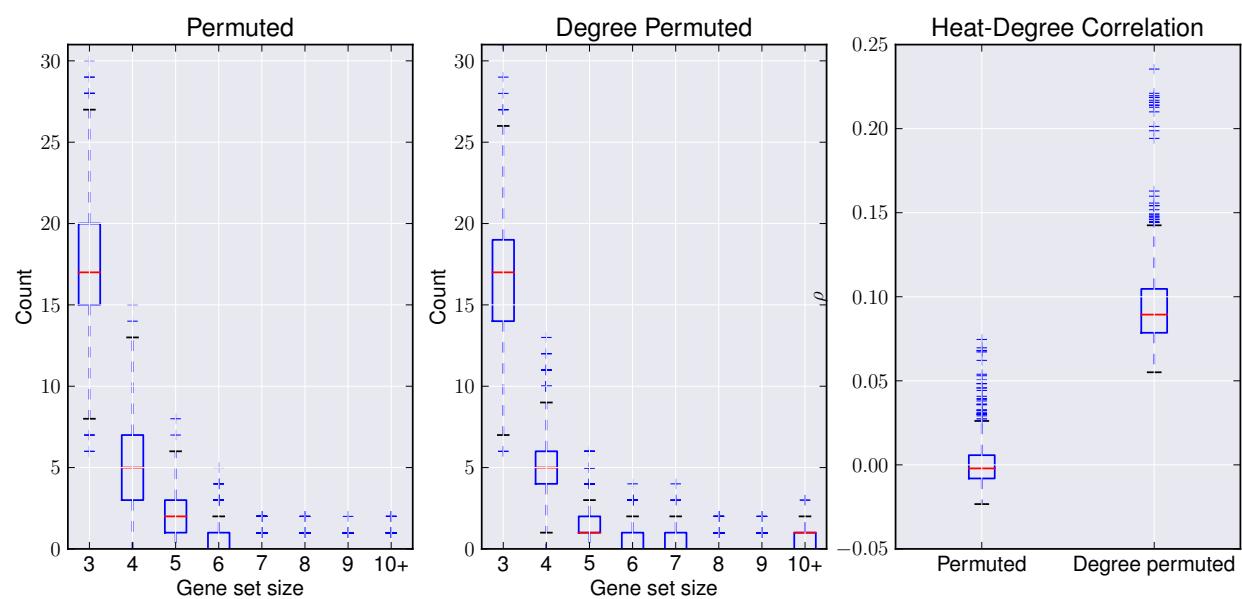
Supplementary Figure 17: The NOTCH subnetwork and its mutation matrix. The mutation matrix shows the mutations in the genes in the subnetwork, represented on the left, across the samples with mutations in the subnetwork. The number of samples with mutations in a gene is in parenthesis. The number of samples with mutations in any of the members of the subnetwork is on the top. Upticks and downticks represent amplifications and deletions, respectively. Full ticks represent SNVs, indels, and splicesite mutations. Colors reflect the cancer type of the samples. Black dots corresponds to inactivating genes, that is genes that contain at least one of the following mutations: frame shift indels, nonsense, nonstop, and splice sites. Genes with * were significant by exactly one of MuSiC, MutSigCV, or Oncodrive, while genes with ** were not significant by any of these methods. Colors of interactions in the subnetwork represent the interaction network where it was found. *P*-values for cancer type enrichment of mutations in the genes are shown.



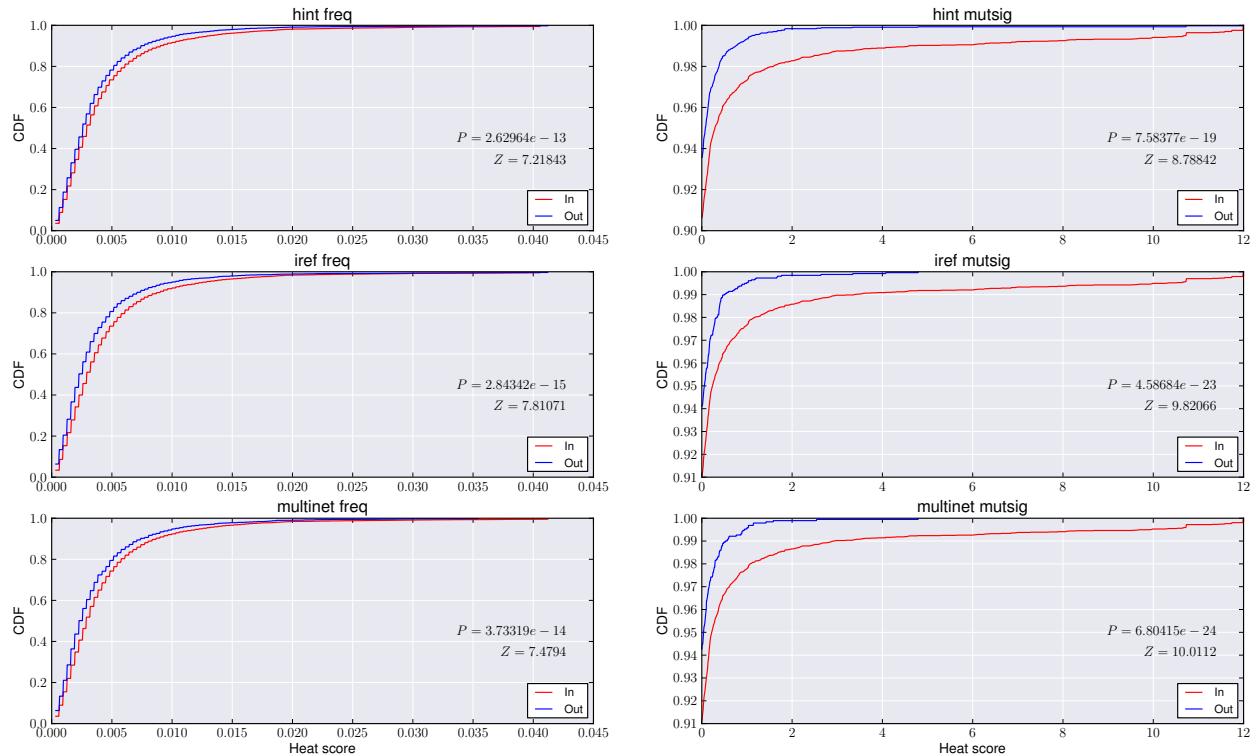
Supplementary Figure 18: Pan-Cancer non-silent mutations in the *SHPRH* gene (ENSEMBL transcript ENST00000367503). *SHPRH* had a significant cluster of three mutations in position 776 ($P < 8 \times 10^{-5}$). In LUAD, GBM, and UCEC, these mutations are primarily in the helicase domain, indicating that they are likely inactivating.



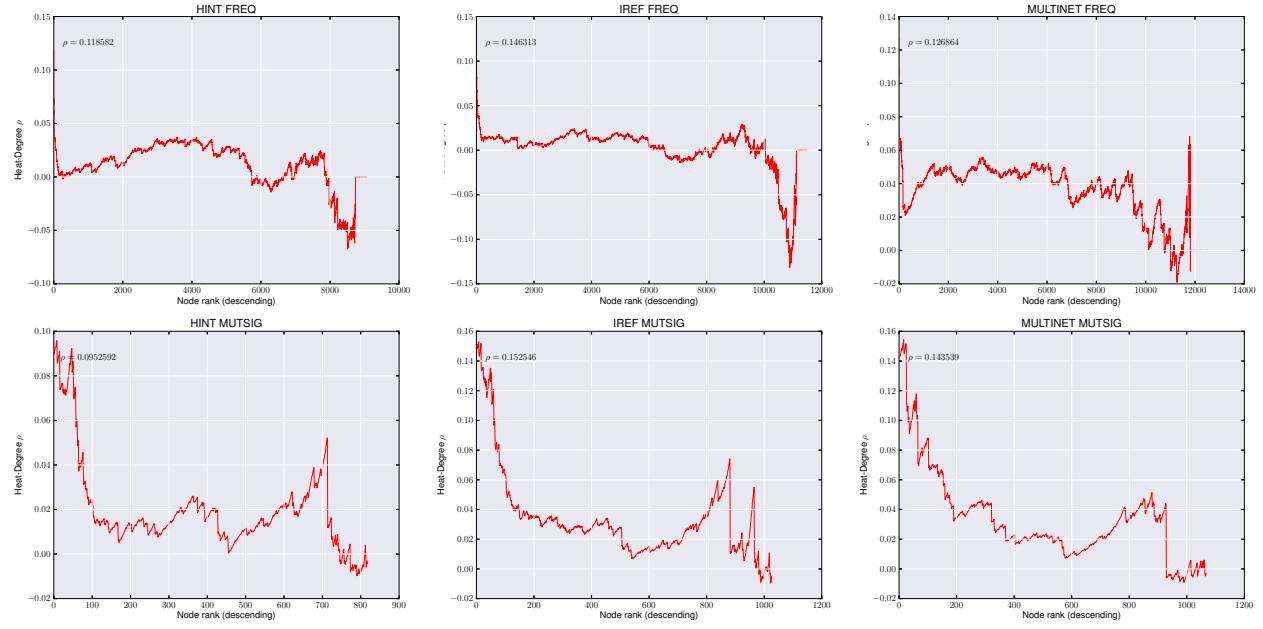
Supplementary Figure 19: The number of aberrations (SNV and CNA) in each sample. Samples with > 400 aberrations were removed.



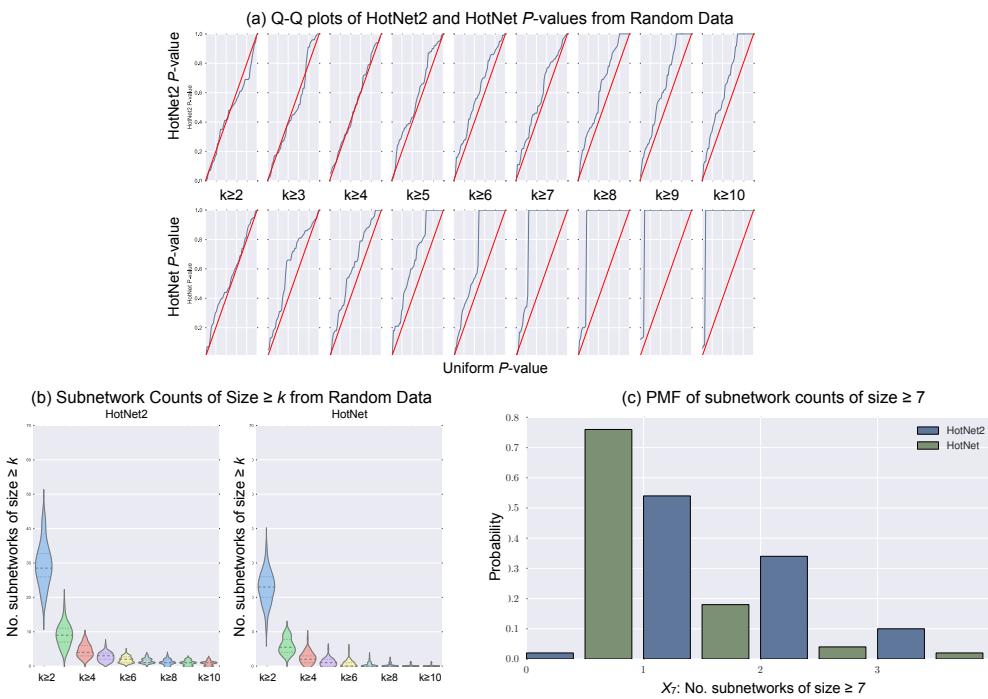
Supplementary Figure 20: (a-b) The distribution of sizes of subnetworks identified by HotNet2 on 1000 permuted datasets. (a) Datasets were generated permuting heat scores uniformly at random. (b) Datasets were generated by permuting the top 100 highest heat scores amongst themselves, and permuting the remaining heat scores amongst themselves. (c) The distribution of heat-degree correlations from the datasets in (a) and (b).



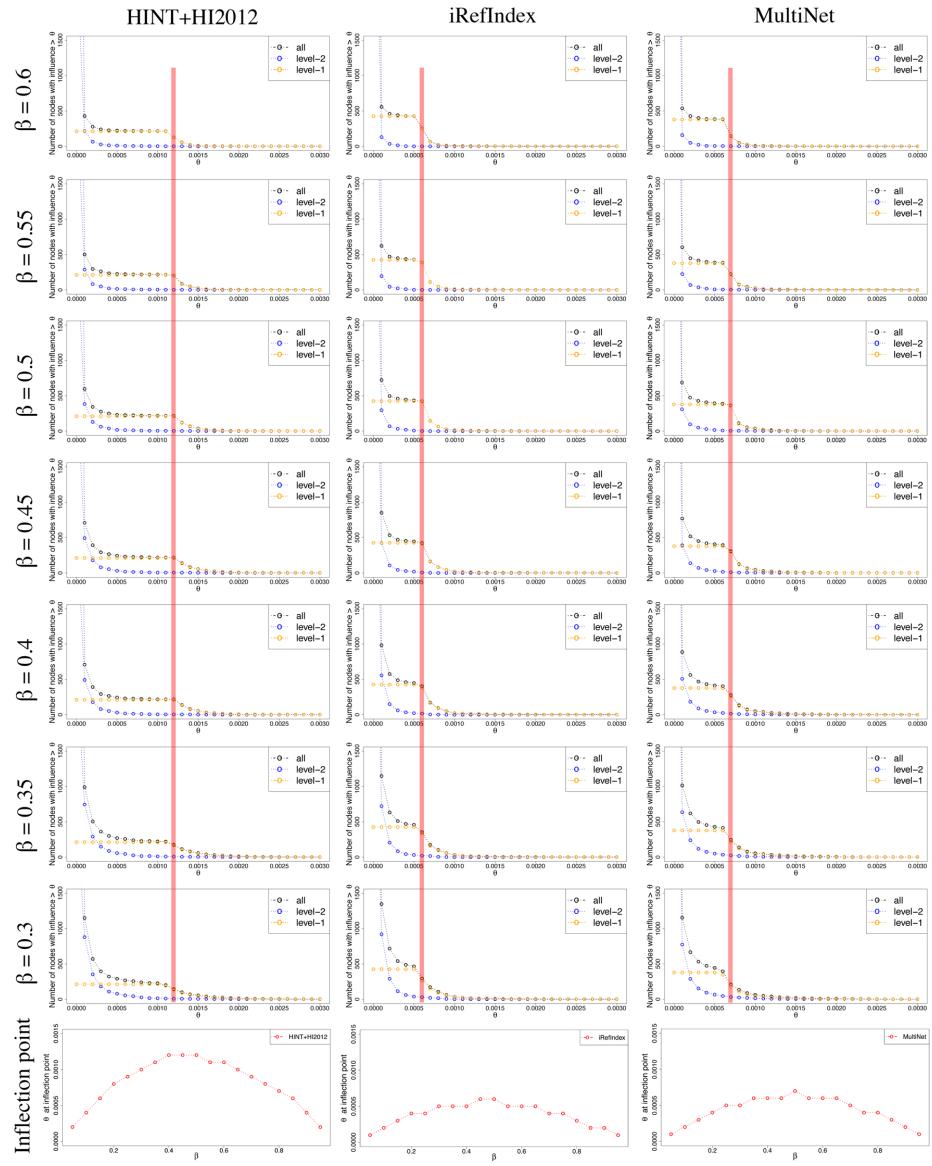
Supplementary Figure 21: The CDF of the distribution of heat (mutation frequency or MutSigCV) in each of the three interaction networks. Using a two-sample z-test, the heat scores of genes in the network are significantly higher than the heat scores of genes not in the network (P -values and Z -scores are shown on each plot). The fifty highest heat scores are removed from the plots to demonstrate the difference visually.



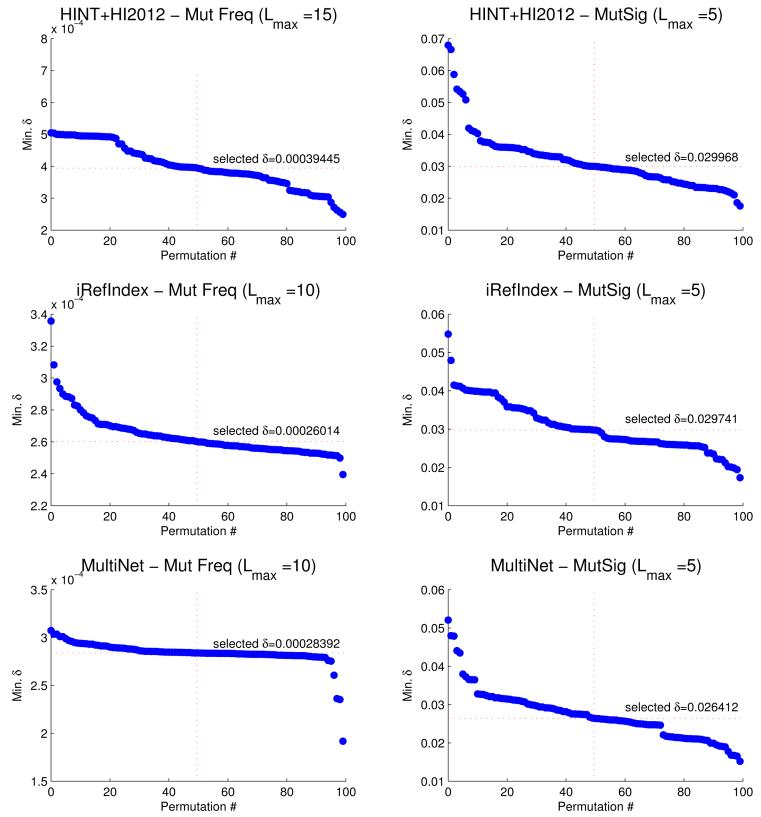
Supplementary Figure 22: The correlation between heat and degree (y-axis, Spearman's ρ) after removing the top N hottest nodes (x-axis) for each combination of network and heat score. After removing $N = 100$ nodes, the correlation between heat and degree has dropped to below $\rho = 0.05$ for each pair except Multinet-MutSigCV. The correlation statistic before removing any nodes is shown on each plot.



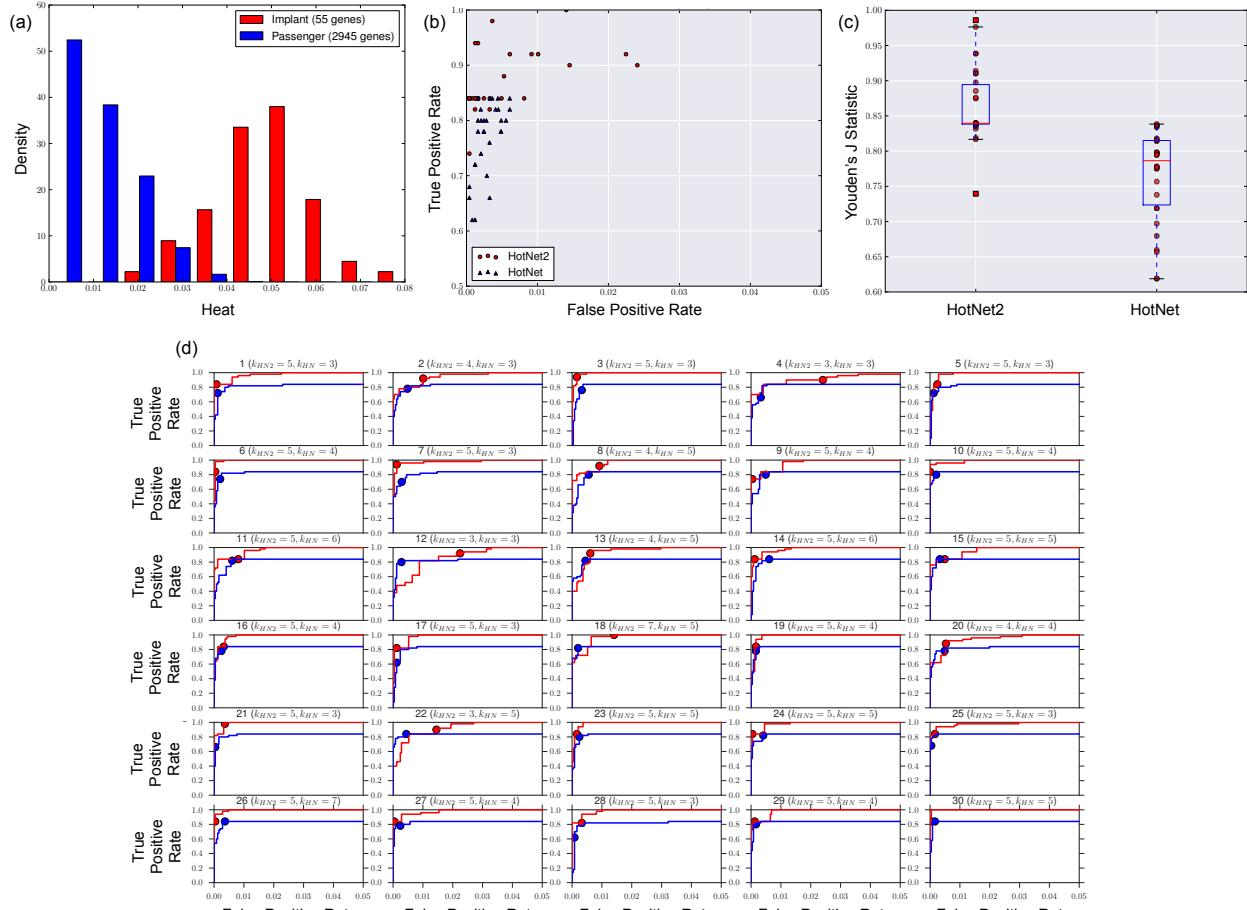
Supplementary Figure 23: Results of HotNet2 and HotNet when applied to the $N = 50$ randomized MutSigCV datasets. (a) Q-Q plots of the P -values from HotNet2 (top) and HotNet (bottom) for each $k = 2 - 10$. (b) Violin plots of the number of subnetworks of size at least k identified by HotNet2 (left) and HotNet (right). (c) Histogram of the number of subnetworks of size ≥ 7 identified by each algorithm.



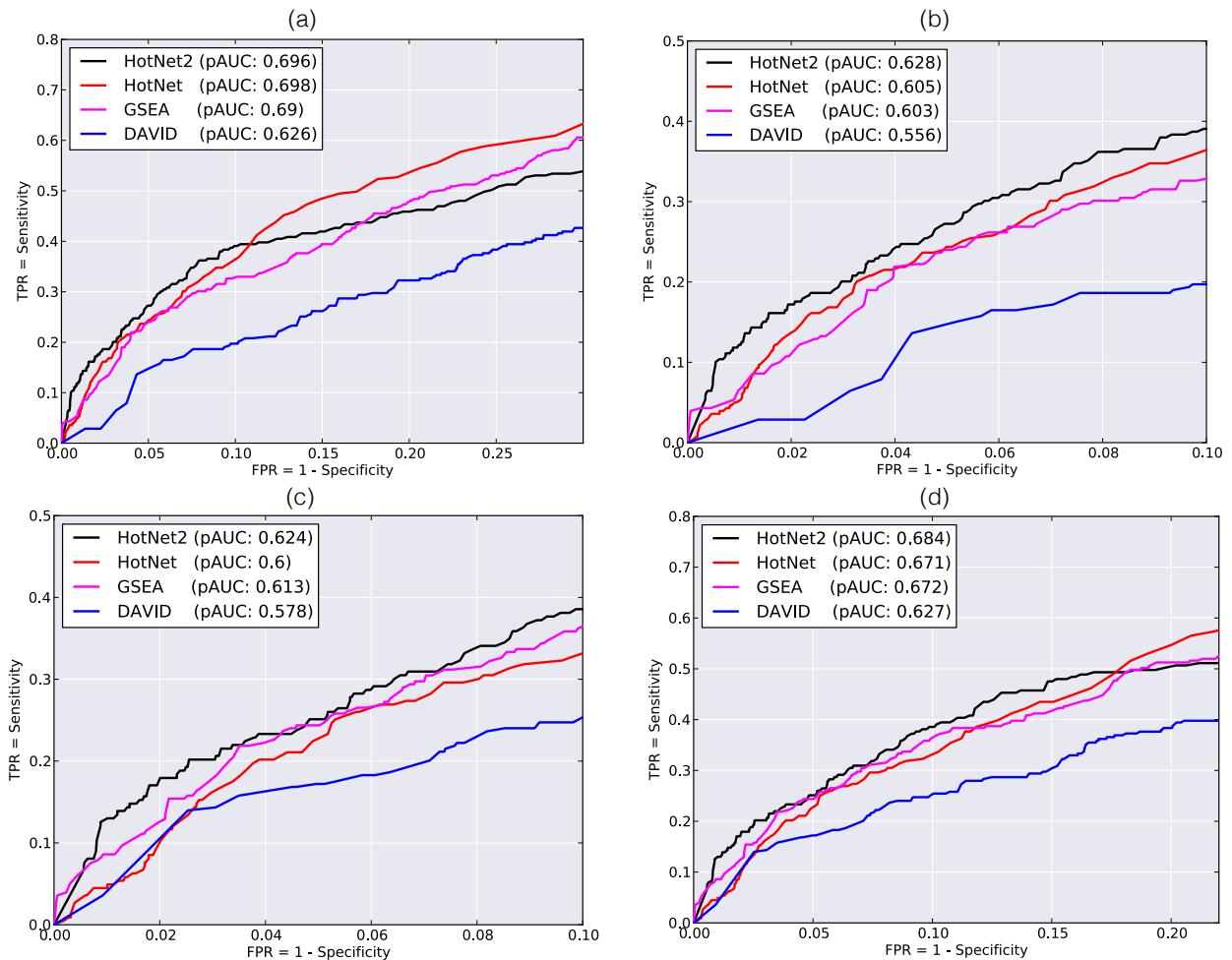
Supplementary Figure 24: Distributions used to set the threshold β . Figures in different columns and rows represent distributions on HINT+HI2012, iRefIndex, and MultiNet interaction networks across β from 0.6 to 0.3 for an example gene *TP53*, which has high betweenness centrality. The x -axis of each distribution represents θ , a cut off of influence. The y -axis of each distribution represents the number of nodes in the interaction network with influence larger than θ . Respectively, black, orange, and blue dotted circles represent the number of all nodes, level one nodes, and level two nodes with influence larger than θ across different θ . Three red vertical lines across all distributions in an interaction network represent the location of the inflection point in level one in the β we chose for different interaction networks, i.e. HINT+HI2012: $\beta = 0.4$, iRefIndex: $\beta = 0.45$, and MultiNet: $\beta = 0.5$. The bottom plot indicates the distribution of θ at the inflection point in β from 0.05 to 0.95.



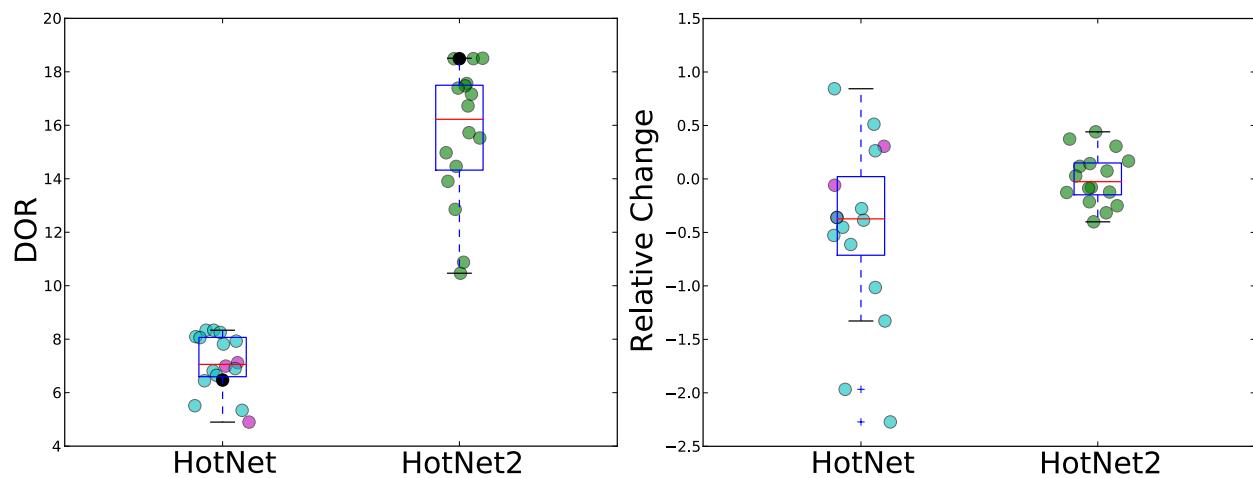
Supplementary Figure 25: Distributions used to set the threshold δ . For each combination of interaction network and score, 100 random networks were generated, and the minimum delta resulting in the maximum size of a connected component being $\leq L_{\max}$, for $L_{\max} = 5, 10, 15, 20$, were computed. The median δ is then used as threshold for HotNet2. For each combination, only the value of L_{\max} used to generate the HotNet2 results is shown. Values of δ are sorted in decreasing order in each plot. Dashed lines identify the median value of δ , reported in each distribution as well.



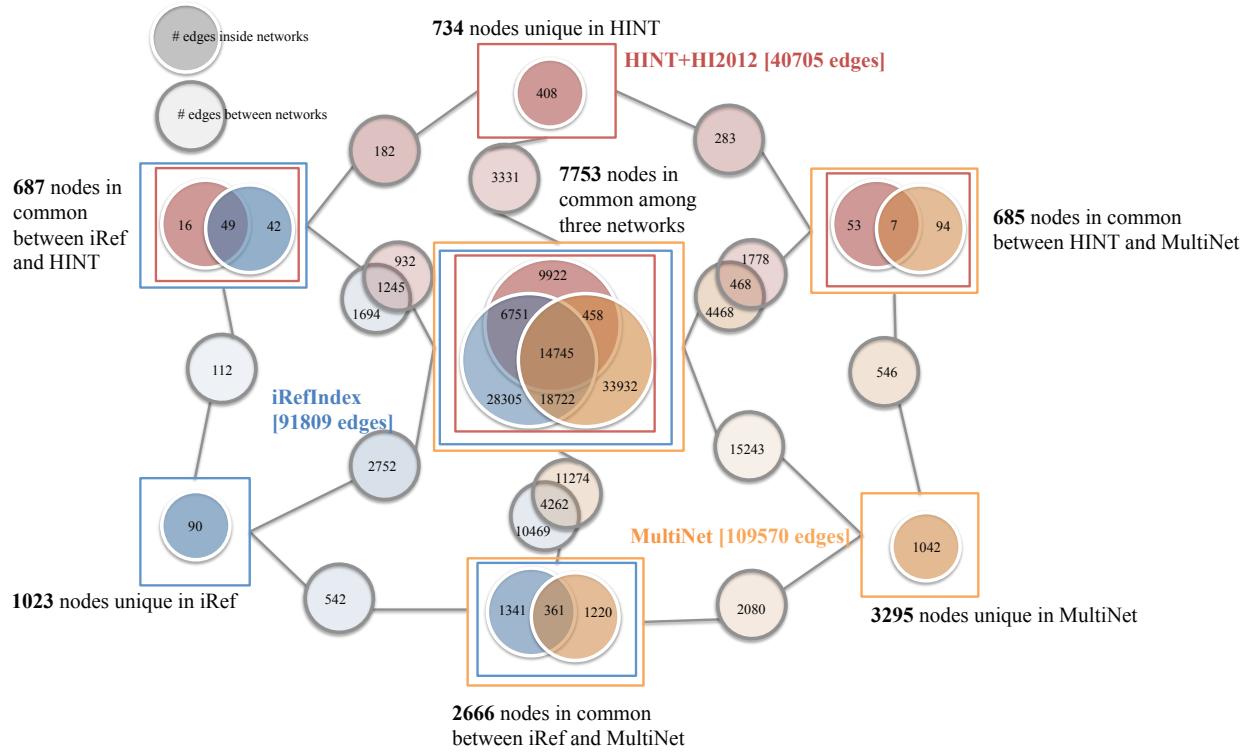
Supplementary Figure 26: (a) Gene score (heat) distribution for a simulated dataset. The distribution was generated using two overlapping normal distributions, where the implanted genes (red) had a higher mean (0.05) than the mean of the background (passenger; blue) genes (0.01). (b) Scatterplot of the false positive rate (x-axis) and true positive rate (y-axis) for HotNet2 (red circles) and HotNet (blue triangles) for identifying the implanted subnetworks in the 30 simulated datasets. (c) Boxplots of HotNet2 vs. HotNet's Youden's $J = \text{sensitivity} + \text{specificity} - 1$ for identifying the implanted subnetworks in the 30 simulated datasets. (d) ROC curves for HotNet2 (red) and HotNet (blue) on the 30 implanted pathway simulation datasets. The true positive and false positive rates were calculated as a function of the minimum edge weight parameter δ . Also shown in the plots are the δ values automatically selected for each dataset by the HotNet2 and HotNet algorithms.



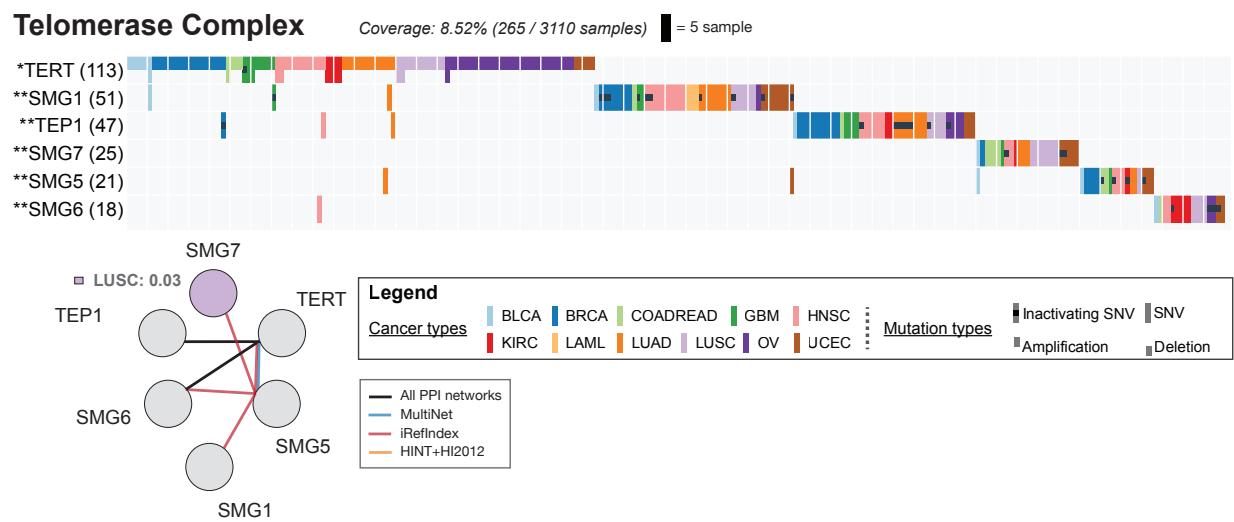
Supplementary Figure 27: (a-b) The receiver operator characteristic (ROC) curves for HotNet2, HotNet, GSEA, and DAVID in finding 20/20 genes. The ROC is computed using all 11,565 genes in the mutation frequency dataset as input to each algorithm. HotNet2 and HotNet were run on the HINT+HI2012 network. (a) ROC restricted to FPR 0.1 (corresponding to > 1,100 false positive predictions). (b) ROC restricted to FPR 0.3 (corresponding to > 3,300 false positive predictions). (c-d) The receiver operator characteristic (ROC) curves for HotNet2, HotNet, GSEA, and DAVID in finding 20/20 genes. The ROC is computed using the 6,930 genes in the mutation frequency dataset and the HINT+HI2012 interaction network as input to each algorithm. HotNet2 and HotNet were run on the HINT+HI2012 network (c) ROC restricted to FPR 0.1 (corresponding to around 700 false positive predictions). (d) ROC restricted to FPR 0.3 (corresponding to around 2100 false positive predictions).



Supplementary Figure 28: **Two-fold cross-validation comparison of HotNet2 and HotNet.** (a) The diagnostic odds ratio (DOR) in finding 20/20 genes is higher for HotNet2 than HotNet. Black points indicate DOR on 100% of mutation frequency samples; magenta/cyan points indicate DORs for HotNet results on 50% of mutation frequency samples that were significant/not significant; green dots indicate DORs for HotNet2 on 50% of mutation frequency samples. (b) Relative change in DOR across two halves of mutation frequency data, showing higher that HotNet2 has higher stability than HotNet.



Supplementary Figure 29: Overlap between nodes and edges in the HINT+HI2012 [13, 14], iRefIndex [16], and MultiNet [15] interaction networks. Each rectangle encloses the nodes (proteins) that are unique to an interaction networks, shared by two interaction networks, or shared by all three networks (middle). Inside each rectangle is a Venn diagram of the overlap in edges in each interaction network between nodes in the enclosing rectangle. Connections between two rectangles show the number of edges that join nodes in the two rectangles.



Supplementary Figure 30: The telomerase complex subnetwork and its mutation matrix. The mutation matrix shows the mutations in the genes in the subnetwork, represented on the left, across the samples with mutations in the subnetwork. The number of samples with mutations in a gene is in parenthesis. The number of samples with mutations in any of the members of the subnetwork is on the top. Upticks and downticks represent amplifications and deletions, respectively. Full ticks represent SNVs, indels, and splicesite mutations. Colors reflect the cancer type of the samples. Genes with * were significant by exactly one of MuSiC, MutSigCV, or Oncodrive, while genes with ** were not significant by any of these methods. Colors of interactions in the subnetwork represent the interaction network where it was found. *P*-values for cancer type enrichment of mutations in the genes are shown.