Science
AAAS

Supplementary Materials for

Interpretation of cancer mutations using a multi-scale map of protein systems

Fan Zheng[1,2,*], Marcus R. Kelly [1,2,*], Dana J. Ramms[2,3,4], Marissa L. Heintschel[5], Kai Tao[6,7], Beril Tutuncuoglu[2,8,9,10], John J. Lee[1], Keiichiro Ono[1], Helene Foussard[8,9,10], Michael Chen[1], Kari A. Herrington[11], Erica Silva[1], Sophie Liu[1], Jing Chen[1], Christopher Churas[1], Nicholas Wilson[1], Anton Kratz[1,2], Rudolf T. Pillich[1,2], Devin N. Patel[1,2], Jisoo Park[1,2], Brent Kuenzi[1,2], Michael K. Yu[1], Katherine Licon[1,2], Dexter Pratt[1], Jason F. Kreisberg[1,2], Minkyu Kim[2,8,9,10], Danielle L. Swaney[2,8,9,10], Xiaolin Nan[6,7,12], Stephanie I. Fraley[5], J. Silvio Gutkind[2,3,4], Nevan J. Krogan[2, 8,9,10, †], and Trey Ideker[1,2, †]

1 Division of Genetics, Department of Medicine, University of California, San Diego, La Jolla, CA 92093, USA; 2 Cancer Cell Map Initiative (CCMI), La Jolla and San Francisco, CA, USA; 3 Moores Cancer Center, University of California, San Diego, La Jolla, CA 92093, USA; 4 Department of Pharmacology, University of California, San Diego, La Jolla, CA 92093, USA; 5 Department of Bioengineering, University of California, San Diego, La Jolla, CA 92093, USA; 6 Department of Biomedical Engineering, Oregon Health and Science University, Portland, OR, 97239, USA; 7 Center for Spatial Systems Biomedicine, Oregon Health and Science University, Portland, OR, 97201, USA; 8 Department of Cellular and Molecular Pharmacology, University of California, San Francisco, CA 94158, USA; 9 The J. David Gladstone Institutes, San Francisco, CA 94158, USA; 10 Quantitative Biosciences Institute, University of California, San Francisco, San Francisco, CA, 94158, USA; 11 Department of Biochemistry and Biophysics Center for Advanced Light Microscopy at UCSF, University of California, San Francisco, San Francisco, CA, 94158, USA; 12 Knight Cancer Early Detection Advanced Research Center, Oregon Health and Science University, Portland, OR, 97201, USA

[*] Equal Contribution

[†] Correspondence: nevan.krogan@ucsf.edu; tideker@health.ucsd.edu

This PDF file includes:

Supplementary Text

Figs. S1 to S8

Descriptions of Data S1 to S3

References (60, 63, 104, 108, 113, 132-143)

Other Supplementary Materials for this manuscript include the following:

Table S1 to S8 (Excel format)

Data S1 to S3

Supplementary Text

Gene Ontology

The Gene Ontology (GO) used in all analyses throughout this study was obtained from http://geneontology.org/ in Nov. 2018, via an API provided by the DDOT python package (*132*). For each GO term, we defined the associated gene sets to be all genes annotated with either the same term itself or any of its descendant terms in the GO hierarchical structure.

Relative Contribution of Evidence Types

The relative contribution of an evidence type e at an IAS threshold x (**Fig. S1C**) was calculated as $\frac{\log R_e}{\sum_{e' \in E} \log R_{e'}}$, where R_e is the odds-ratio of observing n protein pairs scored highest by the evidence type e among the n protein pairs scored highest by IAS, versus the rest of protein pairs, while n equals to the number of protein pairs with IAS > x.

Benchmark of CliXO v1.0 Against the Previous Version

We performed simulation to confirm the improvement of CliXO v1.0 compared to the previous version (v0.3) (**Fig. S2A**). A weighted network, in which the protein pairs were scored by their Resnik semantic similarity (*108*) based on the GO term DNA repair (GO:0006281) and all its children terms (scores were scaled to 0-1), was used as the input of the algorithms. We evaluated the accuracy of recovering the ground truth gene sets associated with GO terms, a task described in the original CliXO paper (*133*). The accuracy was evaluated by the average F1-score (*134*) between the ground truth gene sets and gene sets built by algorithms. Each codebase was tested under different parameters; we used all combinations of α = 0.02, 0.03, 0.05, 0.07, and β = 0.4, 0.5, 0.6, 0.7 for both codebases. For CliXO v1.0 we tried different m values, which did not alter the results, since the built-in communities in this test have high modularity and were not significantly affected by this filter. The reported results were generated at m = 0.005. To test the robustness of the conclusion, we also compared two codebases with perturbed inputs, in which a random Gaussian noise N (0, 0.005) was added to the input Resnik scores.

Coarse-Grained Structure of the IAS Network

For computational efficiency, we used HiDeF, another algorithm we recently developed (*113*), to identify protein systems at lower thresholds of IAS score (IAS < 0.3). HiDeF is available at https://github.com/fanzheng10/HiDeF (*136*). HiDeF identifies robust network communities of multiple scales by integrating the Louvain community detection algorithm (*135*) at continuously scanned resolution parameters. We applied HiDeF to instances of the integrated network filtered at a series of permissive thresholds of the IAS score (0.18, 0.21, 0.24, 0.27, 0.30). Highly persistent protein communities (systems) were retained, with a cutoff of persistence selected by maximizing the average F1-score between the detected communities and the gene sets in the GO cellular component branch (*60*). CliXO-identified systems smaller than 4 genes and HiDeF-identified systems smaller than 20 genes were discarded.

All systems (generated by CliXO and HiDeF) were subsequently integrated into a directed acyclic graph (DAG) by inferring the containment relationships between systems (i.e., one system being part of another) with the weaver module in HiDeF with default parameters.

Our method is designed to identify protein communities in a weighted network at multiple size resolutions (multiscale) and degrees of overlap, including those that are contained within multiple others (pleiotropy). To our best knowledge, despite numerous computational methods of identifying modules in networks, there is not another off-the-shelf method that can simultaneously satisfy all the requirements specified in this study.

Benchmark Test of HiSig, Lasso, and Gene Set Enrichment Analysis (GSEA)

To benchmark the performance of HiSig and alternative approaches, we used a generative procedure to simulate hierarchical structures, in which some systems were set to be "positive systems" and produced signals accruing to a simulated mutational profile (**Fig. S2C-E**). Specifically, a simulated hierarchical structure is a height-balanced tree, where each non-leaf node represents a system, and each leaf node represents a "gene". An ensemble of structures was created by varying three structural parameters (note these are different from the CliXO parameters discussed above). The resultant tree consists of n levels of systems. Each system at the 1st level (the bottom level) contains m genes, and each system at the i-th level merges μ systems from one level lower (i-1). Thus, each system at the n-th level contains μ^n genes, and the entire tree contains

μ^{n+1} genes. *Positive systems* are randomly drawn from all systems with a probability *p*, and all positive systems are denoted as S_P. For each positive system $s \in S_P$, we generated a vector of integers v_s that follows:

```
for g in (1, 2, ... len(y)) {
      if s contains g:
            vs(g) ~ Log1p[Poisson(1)]
      else:
            vs(g) ~ 0
}
```

Poisson distribution makes the output vector have a "long-tail" distribution, mimicking the mutation frequency profiles in real cancer data. The response vector, *y*, was then calculated as:

```
for g in (1, 2, ... len(y)) {
      y(g) = Σ vs(g)
}
```

The response vector *y* and the "gene sets" defined by the simulated systems were used as the inputs of the HiSig, GSEA (*63*) and regular Lasso with a single λ parameter (*136*). We tested all the combinations of the following parameters: $N = 2$, 3, 4; $\mu = 4, 5, 6$; $p = 0.1, 0.2, 0.3, 0.4$. The systems were ranked by the statistic of each method (see below), and the performances were measured by the ability of each method to favorably rank positive systems, i.e., the area under receiver operating characteristic curve (AUROC).

For HiSig, we performed 1000 permutations and ranked systems by their p-values. For Lasso, we performed 5-fold cross-validation to determine the optimal value of λ and used this λ to make predictions; systems were ranked by their coefficients. Gene set enrichment analysis (GSEA) was conducted in its "preranked" form. A small non-negative noise was added to each element of y, as GSEA does not allow tied ranks. As described in the original GSEA paper (*63*), for each system an enrichment score was calculated, and then a nominal p-value was calculated with a distribution of enrichment score of 1000 randomly selected gene sets of equal size. Systems were then ranked by their nominal p-values.

Analysis of Copy Number Alterations (CNA)

CNA calls of genes, derived from GISTIC 2.0 (*137*) , were obtained from cBioPortal (*104*) for the 13 analyzed cancer cohorts in TCGA. We considered values ±2, representing high-level amplification and deep deletion, respectively, as CNA events in patients, and calculated the CNA frequencies of genes in cancer cohorts. We noted that an CNA event can impact a chromosomal segment encompassing many genes, and thus many gene-level CNA events are likely passenger events.

Efforts to address this confounding effect have not yet reached a consensus. Here, we adjusted the CNA frequencies by the support of mRNA expression levels. Specifically, for each tumor sample, a gene is defined as up- or down-regulated if its Z-score relative to all tumor samples in cohort is more extreme than ±2. A CNA event in a sample was counted if an amplified gene was also up-regulated in this sample, or a deep deleted gene was also down-regulated. With the same approach taken for the analysis of mutations, these adjusted frequencies were log-transformed and used as the inputs of HiSig (**Fig. 3A**). For two new CNA systems, all members of the system were in the same chromosomal locus and had little apparent functional relationship; concerned about potential confounding effects of mRNA co-expression on the identification of this system, both systems have been removed.

Enrichment of AP-MS data in NeST Systems

Following multi-scale community detection (see above), the enrichment of each system for the new AP-MS PPIs (**Fig. S6**, **Table S6**) was computed as the probability $Pr(X > k)$ under a hypergeometric distribution $Hypergeometric(N, K, n)$, where K is the total number of AP-MS interactions reported in the entire dataset (here 1722), N is the total number of possible interactions with baits (here *No.baits* * (*No.proteins* – 1); *No.baits* = 61; *No.proteins* = 19,035), k is the actual number of new AP-MS interactions within the system, and n is the total number of possible interactions with baits within the system. Systems with $P < 0.05$ (Bonferroni correction) were labeled as enriching for new AP-MS PPIs (**Fig. S6**).

Germline Collagen Mutations in Tu To Cells

Tu To Cells were analyzed by whole-exome sequencing (see Materials and Methods), and found to contain the following non-silent mutations in fibrillar collagen coding sequences: COL3A1: H1353Q; COL5A1: G530S, N951S; COL6A1: S890L; COL6A2: S399N, R680H; COL12A1: I574T, G1894S. Underlined mutations are changes to glycine codons. TuTo fibroblast cells (ATCC® CRL-1298™) are derived from a patient with osteogenesis imperfecta, a disease that is caused by mutations in collagens matching some of those identified by HiSig. They therefore made a good candidate to generate a matrix resembling those found in collagen-mutant tumors.

Supplementary Figures

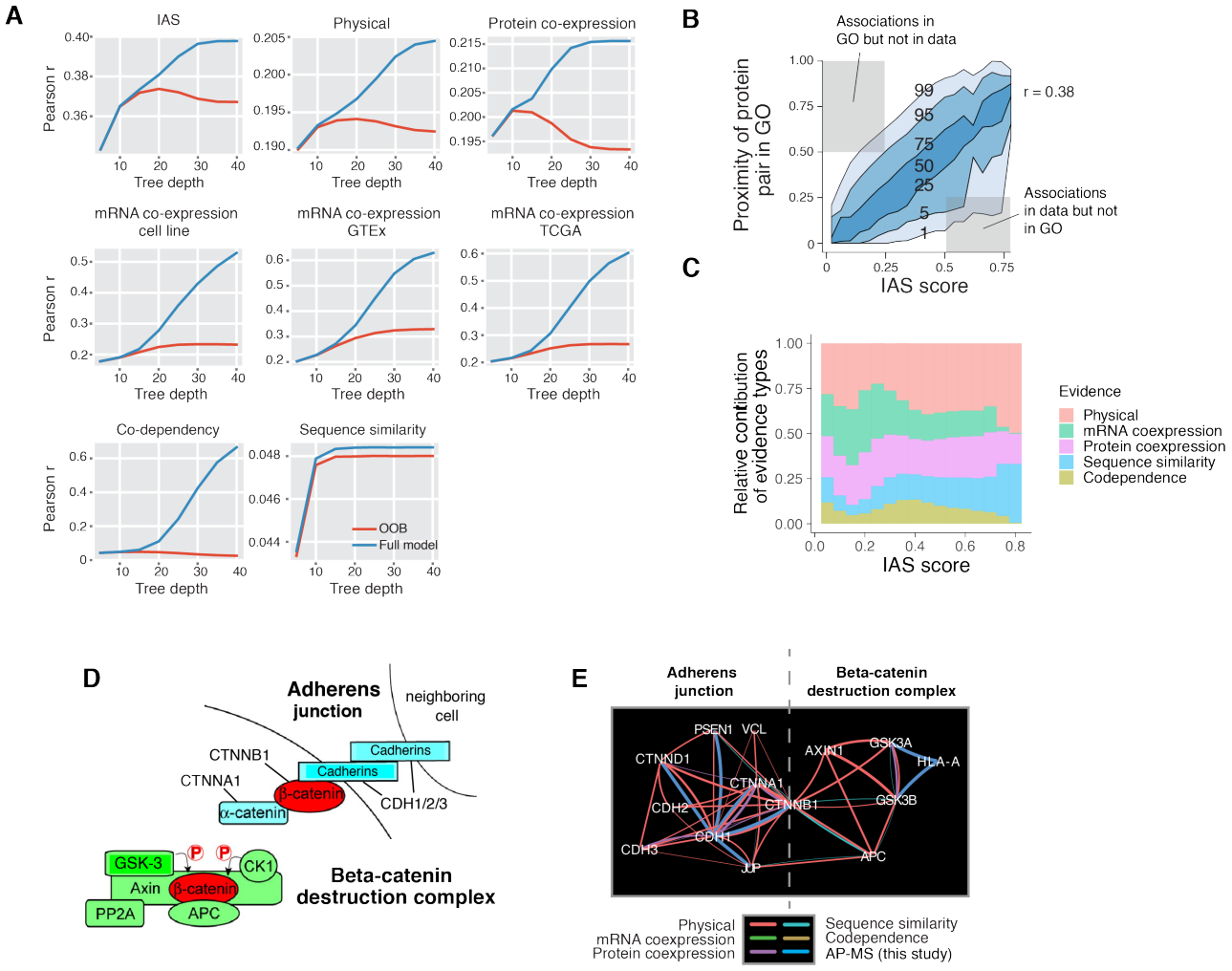

Fig. S1. Mapping protein systems from integrated association stringencies. (A) Each plot shows the Pearson correlation coefficient (*r;* y-axis) of the regression model output with the target function (protein-protein semantic similarity in GO), as a function of random forest tree depth (x-axis). A separate plot is given for each regression model in stage 1 (one model per individual evidence type, labeled above plots) as well as the final integrated regression model in stage 2 (plot labeled as "Integrated"). The correlation performance is reported both in cross-validation mode ("out-of-bag" or "OOB", red) and without cross-validation ("Full model", blue). See details in Materials and Methods. **(B)** For all protein pairs, proximity in the reference cell component hierarchy (GO Resnik semantic similarity (*108*); y-axis) is plotted against the IAS score (x-axis). Curves indicate different percentiles (1, 5, 25, 50, 75, 95, 99) of GO proximity scores at a given IAS score. **(C)**

Relative contribution of evidence types as a function of the IAS score (x-axis). See Supplementary Text. **(D)** A diagram showing the β-catenin protein (CTNNB1) participating the adherens junction the β-catenin destruction complex, which both have important functions in cancer pathways. **(E)** We identified overlapping protein systems closely recapitulating known protein members of adherens junction and β-catenin destruction complex. Note CTNNB1 was assigned as a member of both systems (i.e., pleiotropy).

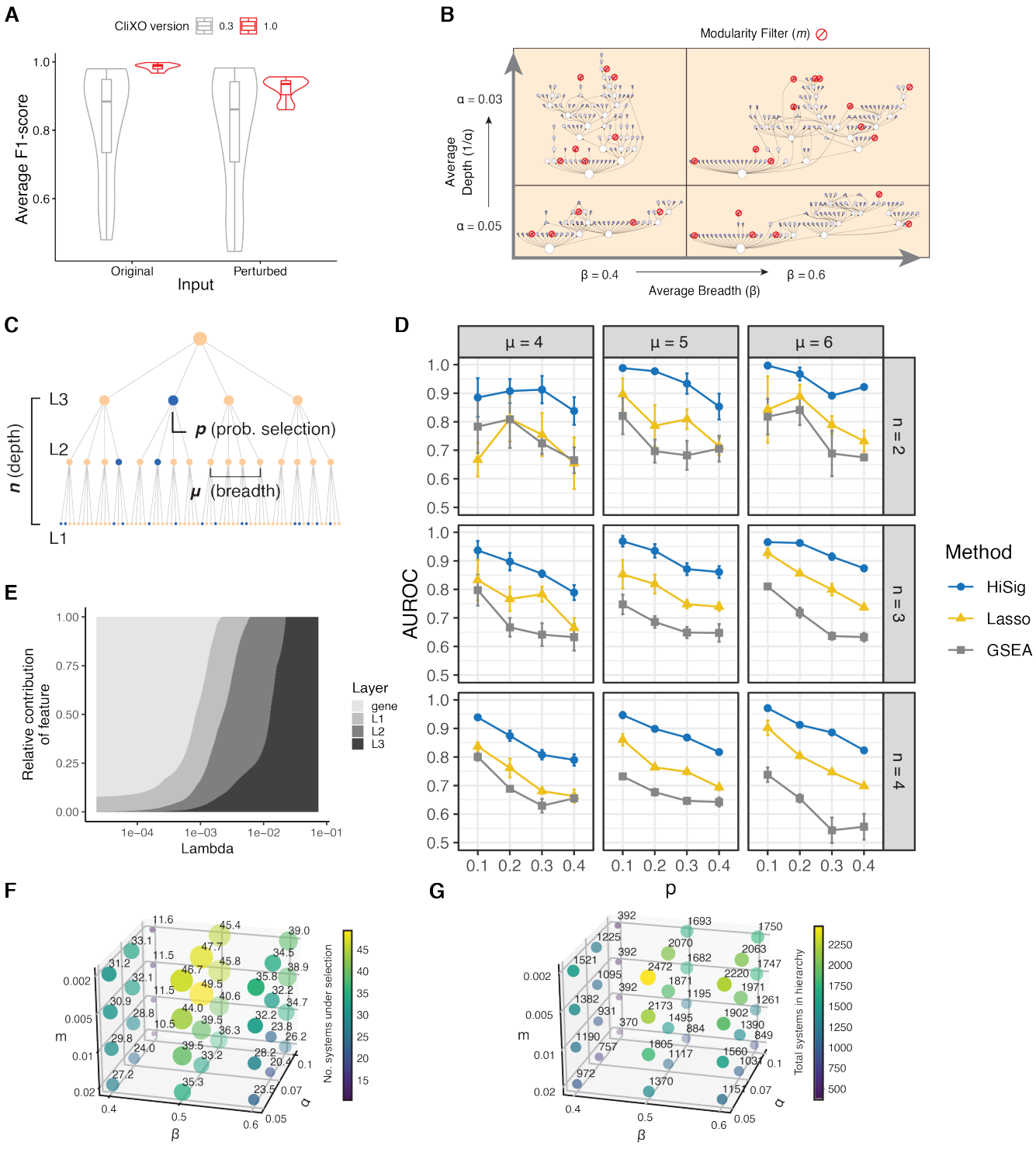
A
CliXO version
0.3
1.0
Average F1-score
Original
Perturbed
Input
B
Modularity Filter (m)
α = 0.03
α = 0.05
Average Depth (1/α)
β = 0.4
β = 0.6
Average Breadth (β)
C
L3
L2
L1
n (depth)
p (prob. selection)
μ (breadth)
D
μ = 4
μ = 5
μ = 6
n = 2
n = 3
n = 4
AUROC
p
Method
HiSig
Lasso
GSEA
E
Relative contribution of feature
Lambda
Layer
gene
L1
L2
L3
F
No. systems under selection
m
α
β
G
Total systems in hierarchy
m
α
β

Fig. S2. Towards assembly of the NeST model: CliXO and HiSig methods (A) Benchmark of the CliXO program against its previous version (v1.0 and v0.3), measured by the recovery of ground truth in GO when using the semantic similarities of gene pairs as the input. The performance was evaluated by average F1-score (y-axis) and different combinations of parameters were tested. See details in Supplementary Text. **(B)** Illustration of hierarchy assembly (CliXO 1.0) parameters: α, β, m. Given a weighted input network, α tends to control the depth of the hierarchy (i.e., average distance from hierarchy root to leaves); β tends to control the breadth (i.e., the average number of children per parent system); and m is used to remove systems with lower modularity, e.g., lower ratios of within-system interactions versus out-of-system interactions at a given IAS. **(C-E)** Benchmark analysis of the HiSig program for identifying selection signals among nested systems. **(C)** Generative hierarchical models, controlled by three structural parameters (n, μ, p). **(D)** At different λ parameter (x-axis), the Lasso regression selected systems of different scales. The percentage chart shows the relative contribution (y-axis; measured by the sum of feature coefficients) from different layers of the simulated hierarchy. The figure was created with a simulation with n=3, μ= 3, p=0.3. **(E)** Evaluating the performance of prioritizing positive systems (AUROC, y-axis) by HiSig and baseline methods (regular Lasso and GSEA) under different structural parameters. See details in Supplementary Text. **(F)** The average number of significantly mutated systems (output of HiSig; FDR < 0.25) per tumor cohort (point size and color) when scanning values of CliXO parameters α, β, m (three axes of the plot), for the IAS score input. **(G)** The total number of systems in the hierarchical models generated with different parameters.

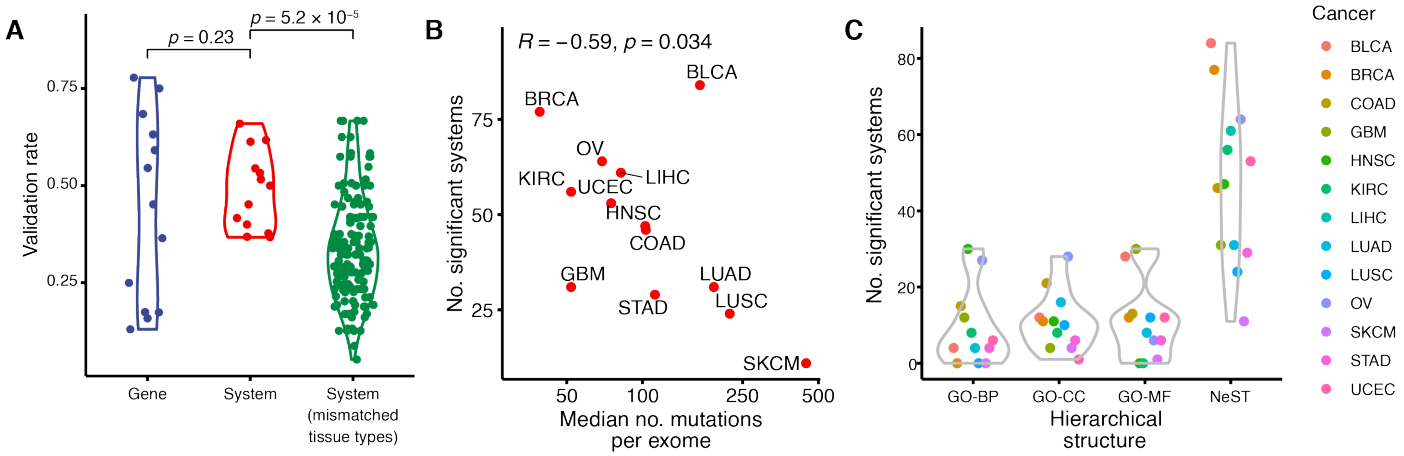

Fig. S3. Global properties of the NeST model. (A) The cross-cohort validation rates are evaluated for significantly mutated genes (defined by FDR < 0.1 in MutSigCV) with matched tissue types (left, $N = 13$), NeST systems with matched tissue types (middle, by HiSig; $N = 13$) and NeST systems with mismatched tissue types (right, by HiSig; $N = 104$). For each validation test (a pair of two tumor cohorts), genes or systems with FDR < 0.1 in the validation cohort were defined as validated and the percentage of validated entities was calculated (y-axis). *P*-values were computed via one-tailed Wilcoxon rank sum test. **(B)** The relationship between the number of significantly mutated systems (y-axis; selected by HiSig; FDR < 0.25) and the mutation burden of each cancer type (x-axis). **(C)** HiSig identified higher numbers of significantly mutated systems (y-axis) using the IAS data-driven hierarchical structure than using the hierarchical structures of GO (existing multi-scale models of cell biology).

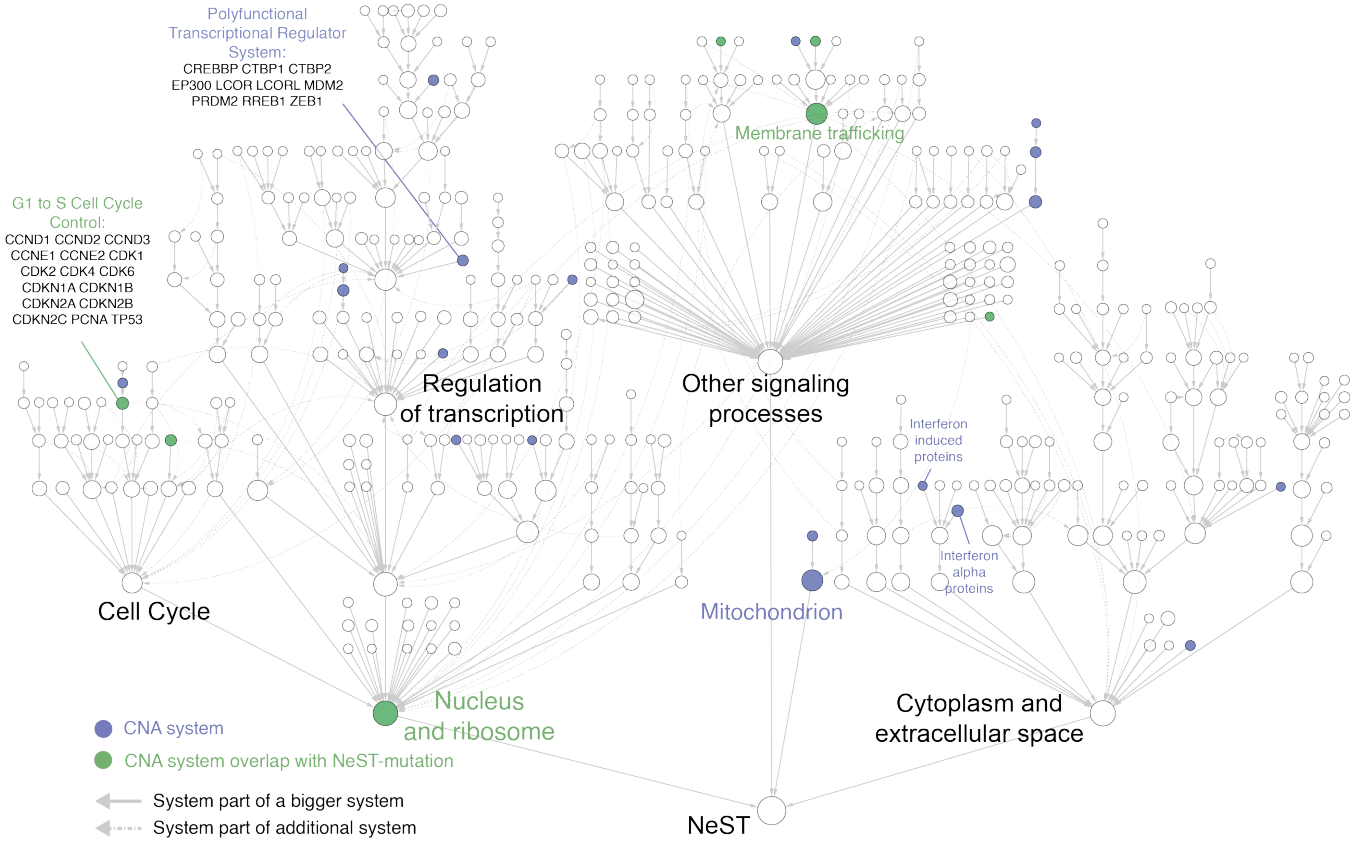

Fig. S4. Additional cancer systems defined by HiSig analysis of copy number alterations. Applying the HiSig method to copy number alterations (CNAs), both existing NeST systems (included because of significant mutation signal; green) and additional systems (blue) were found to have enrichment for CNAs (see Supplementary Text). These systems are described in **Table S4**.

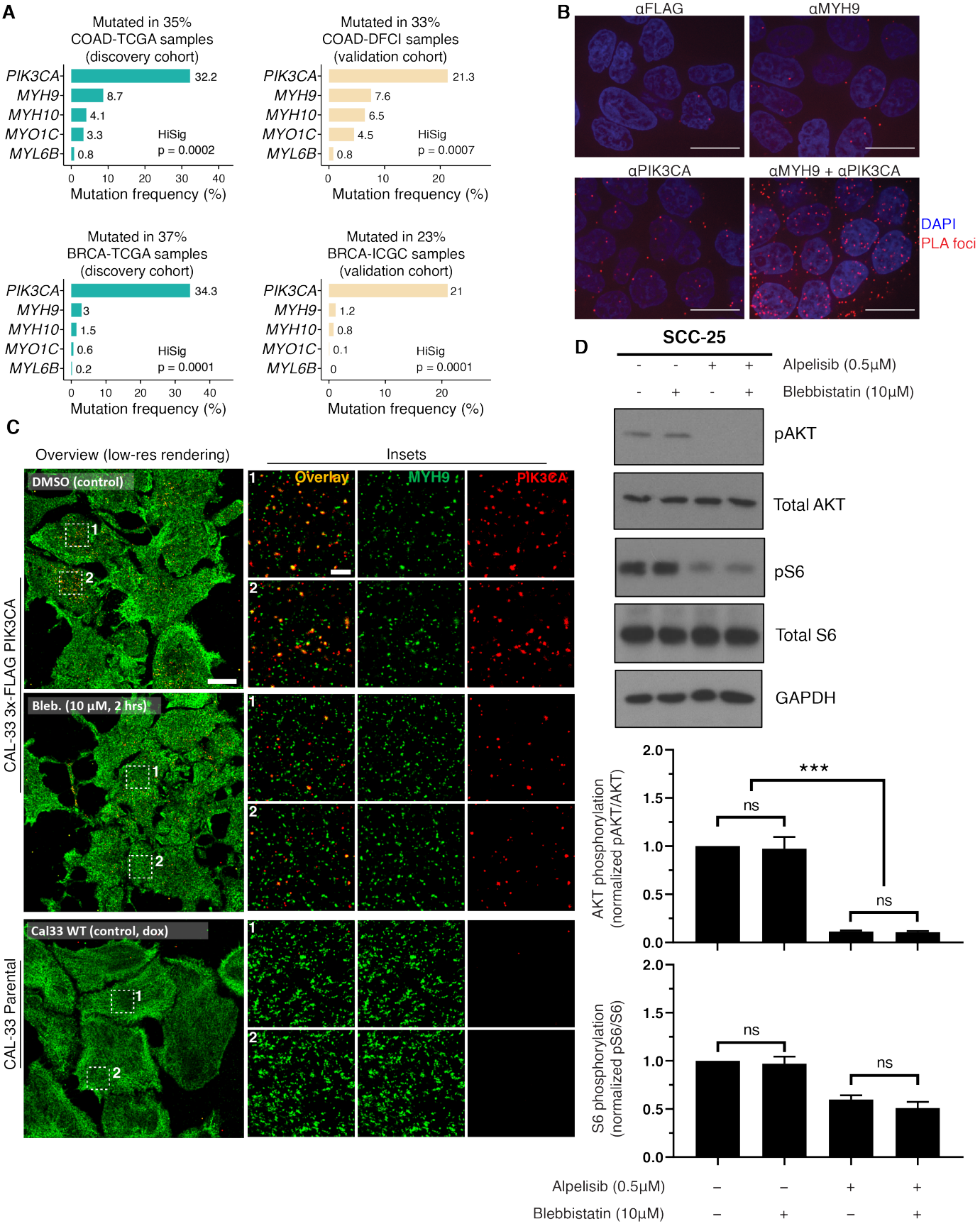
A
Mutated in 35%
COAD-TCGA samples
(discovery cohort)
PIK3CA 32.2
MYH9 8.7
MYH10 4.1
MYO1C 3.3
MYL6B 0.8
HiSig
p = 0.0002
Mutation frequency (%)
Mutated in 33%
COAD-DFCI samples
(validation cohort)
PIK3CA 21.3
MYH9 7.6
MYH10 6.5
MYO1C 4.5
MYL6B 0.8
HiSig
p = 0.0007
Mutation frequency (%)
Mutated in 37%
BRCA-TCGA samples
(discovery cohort)
PIK3CA 34.3
MYH9 3
MYH10 1.5
MYO1C 0.6
MYL6B 0.2
HiSig
p = 0.0001
Mutation frequency (%)
Mutated in 23%
BRCA-ICGC samples
(validation cohort)
PIK3CA 21
MYH9 1.2
MYH10 0.8
MYO1C 0.1
MYL6B 0
HiSig
p = 0.0001
Mutation frequency (%)
B
αFLAG
αMYH9
αPIK3CA
αMYH9 + αPIK3CA
DAPI
PLA foci
C
Overview (low-res rendering)
Insets
Overlay
MYH9
PIK3CA
DMSO (control)
Bleb. (10 μM, 2 hrs)
Cal33 WT (control, dox)
CAL-33 3x-FLAG PIK3CA
CAL-33 Parental
D
SCC-25
Alpelisib (0.5μM)
Blebbistatin (10μM)
pAKT
Total AKT
pS6
Total S6
GAPDH
AKT phosphorylation
(normalized pAKT/AKT)
S6 phosphorylation
(normalized pS6/S6)
***
ns
Alpelisib (0.5μM)
Blebbistatin (10μM)

Fig. S5. A PIK3CA-actomyosin assembly regulating the PI3K/AKT pathway (A) Mutation frequencies of the "PIK3CA-actomyosin complex part" in the TCGA discovery cohorts and those in the validation cohorts, respectively. **(B)** Representative images of the proximity ligation assay (PLA) data (quantified in **Fig. 5D**). Scale bars are 20μm. **(C)** Whole-field DNA-PAINT images of CAL-33-FLAG-PIK3CA cells (top), CAL-33-FLAG-PIK3CA cells treated with blebbistatin (middle), and parental CAL-33 cells (top). Scale bars are 10 μm (left) and 500 nm (right). **(D)** SCC-25 cells were treated with DMSO, blebbistatin (10 μM) and/or alpelisib (0.5 μM) and harvested for immunoblotting with indicated antibodies (N = 3; a representative image is displayed). Signals were normalized relative to the total protein and control samples. Error bars indicate mean ± standard error *** : $p < 0.001$ by one-way ANOVA.

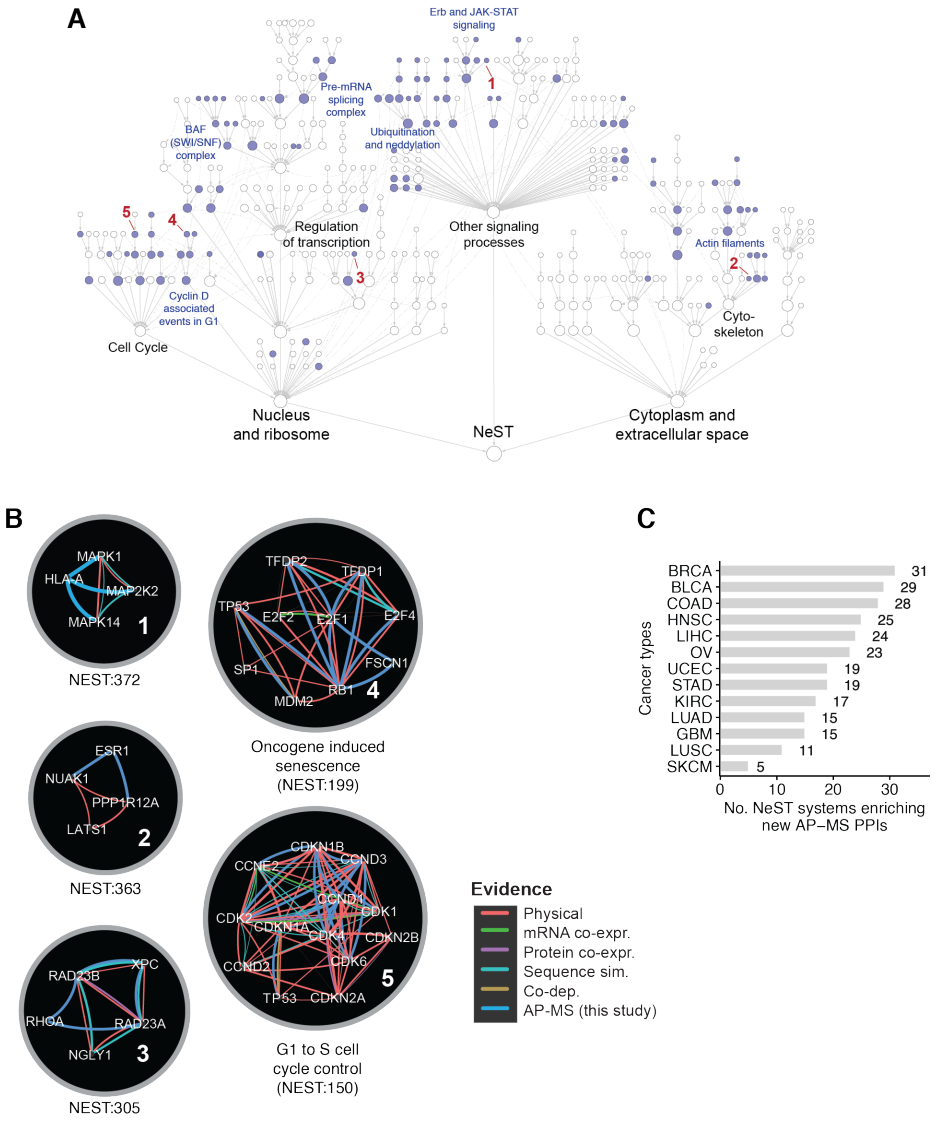

Fig. S6. NeST systems supported by new AP-MS interactions. (A) The distribution of systems enriching new AP-MS interactions in the NeST hierarchy; Node colors indicate the systems enriching the new AP-MS interactions (blue: FDR < 0.05; hypergeometric test). **(B)** Examples of systems supported by new AP-MS interactions (No. 1-5), showing together with different interaction evidence types from public databases (colors). Numbered systems in (B) are also indicated in (A). **(C)** Number of new AP-MS enriching systems per cancer types.

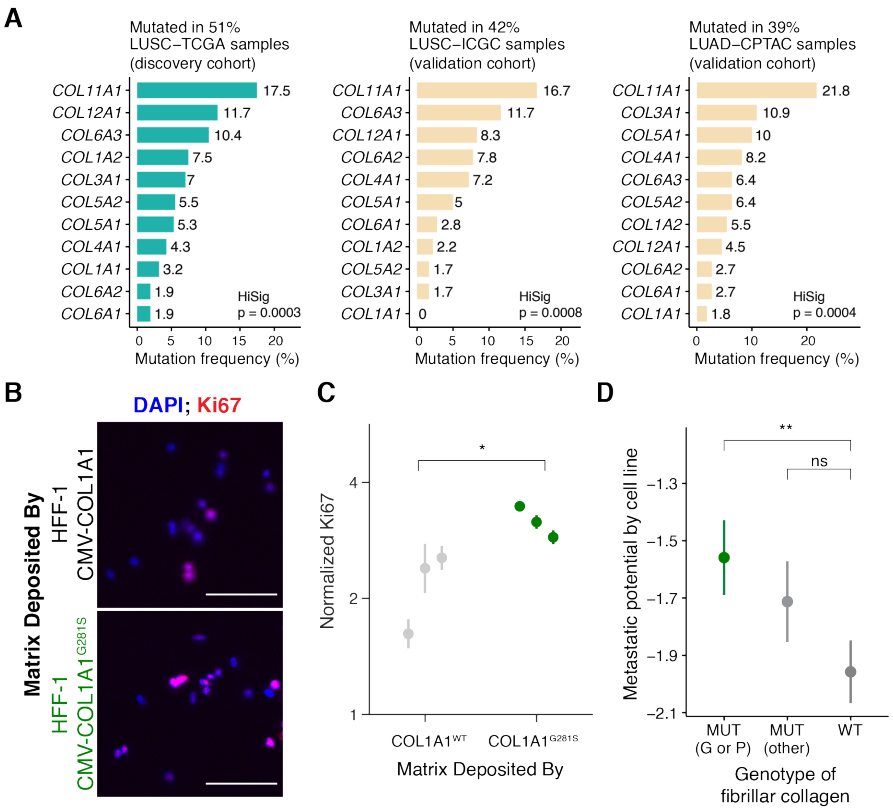

Fig. S7. Destabilizing mutations in collagen systems promote tumor progression. (A) Mutation frequencies of the "Fibrillar Collagen" system in the TCGA discovery cohorts and those in the validation cohorts, respectively. **(B)** Immunofluorescence of A549 cells seeded on matrix deposited by the indicated cell lines. Scale bars 100μm. **(C)** Quantification of immunofluorescence across three biological replicates (points) for each condition. Error bars indicate 95% confidence intervals. For $COL1A1^{WT}$, N=380, 182, 789; for $COL1A1^{G281S}$, $N = 4873, 2607, 2341$. *: $P<0.05$ by one-tailed Student's *t*-test. **(D)** Associations between fibrillar collagen mutation status and the metastatic potential of cancer cell lines (*143*) ($N = 146, 123, 219$), assessed by Wilcoxon rank sum test *: $P < 0.05$; **: $P < 0.01$; ns: not significant.

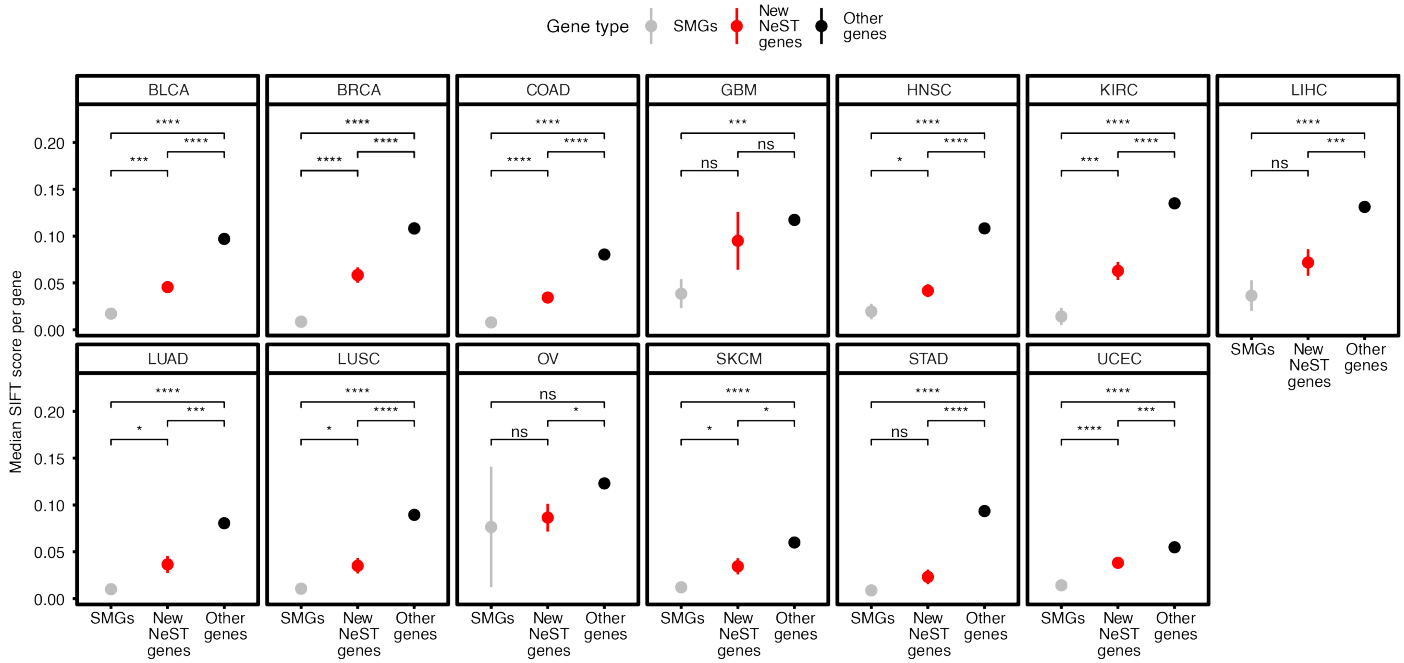

Fig. S8. Assessment of the functional impact of the mutations of new NeST cancer genes. For each analyzed cancer type, the functional impacts of the mutations (median SIFT scores of point mutations per gene; y-axis) of new NeST cancer genes (red) were compared to those of TCGA significantly mutated genes (grey; SMGs) and all other genes (black). Lower SIFT scores indicate predicted deleterious effects. Error bars show mean ± stderr. P-values assigned by Bonferroni corrected one-sided Welch's *t*-test. *: $P < 0.05$; **: $P < 0.01$; ***: $P < 0.001$; ****: $P < 0.0001$; ns: not significant.

Supplementary Tables

Table S1. A full list of new AP-MS protein interactions used in this study.

Table S2. 127 numerical features used to construct the IAS network.

Table S3. NeST systems validated in independent cancer cohorts.

Table S4. NeST system enriched for copy number alterations (CNAs).

Table S5. NeST system annotations and associations to known cancer pathways.

Table S6. NeST systems with enrichment for AP-MS interactions.

Table S7. Prognostic values of NeST systems.

Table S8. Complete list of 548 NeST cancer genes and their functional supports.

Supplementary Data

Data S1. Integrated association stringency (IAS) network. The cancer protein network resource (containing multiple evidence types) stored in the Network Data Exchange (NDEx) database in Cytoscape (*138-140*) format is accessible from http://ccmi.org/nest/. An archived version of the same resource in the TSV (tab-separated values) format is available at (*141*).

Data S2. IAS catalog of multi-scale protein systems. A file containing 2337 protein systems of various sizes (multi-scale), derived from applying multiscale community detection methods on the IAS network, is available online(*142*).

Data S3. NeST 1.0 hierarchical cancer systems map. The interactive NeST 1.0 map, appearing as in **Fig. 4A**, is available at http:/ccmi.org/nest/. NeST is browsable in multiple formats, including a hierarchical model on NDEx with linkouts to Cytoscape (*138-140*) and as a set of nested modules in the HiView web explorer. This map and two other versions with altered visual styles for highlighting CCMI modules and CNA modules (**Figs. S4 and S6**) is available as a Cytoscape file (.cys) online (*143*).

References

60. M. Ashburner, C. A. Ball, J. A. Blake, D. Botstein, H. Butler, J. M. Cherry, A. P. Davis, K. Dolinski, S. S. Dwight, J. T. Eppig, M. A. Harris, D. P. Hill, L. Issel-Tarver, A. Kasarskis, S. Lewis, J. C. Matese, J. E. Richardson, M. Ringwald, G. M. Rubin, G. Sherlock, Gene Ontology Consortium, Gene ontology: Tool for the unification of biology. *Nat. Genet.* **25**, 25–29 (2000). doi:10.1038/75556 Medline

63. A. Subramanian, P. Tamayo, V. K. Mootha, S. Mukherjee, B. L. Ebert, M. A. Gillette, A. Paulovich, S. L. Pomeroy, T. R. Golub, E. S. Lander, J. P. Mesirov, Gene set enrichment analysis: A knowledge-based approach for interpreting genome-wide expression profiles. *Proc. Natl. Acad. Sci. U.S.A.* **102**, 15545–15550 (2005). doi:10.1073/pnas.0506580102 Medline

104. J. Gao, B. A. Aksoy, U. Dogrusoz, G. Dresdner, B. Gross, S. O. Sumer, Y. Sun, A. Jacobsen, R. Sinha, E. Larsson, E. Cerami, C. Sander, N. Schultz, Integrative analysis of complex cancer genomics and clinical profiles using the cBioPortal. *Sci. Signal.* **6**, pl1 (2013). doi:10.1126/scisignal.2004088 Medline

108. P. Resnik, Others, Semantic similarity in a taxonomy: An information-based measure and its application to problems of ambiguity in natural language. J. Artif. Intell. Res. 11, 95– 130 (1999). doi:10.1613/jair.514

113. F. Zheng, S. Zhang, C. Churas, D. Pratt, I. Bahar, T. Ideker, HiDeF: Identifying persistent structures in multiscale 'omics data. *Genome Biol.* **22**, 21 (2021). doi:10.1186/s13059- 020-02228-4 Medline

132. M. K. Yu, J. Ma, K. Ono, F. Zheng, S. H. Fong, A. Gary, J. Chen, B. Demchak, D. Pratt, T. Ideker, DDOT: A Swiss Army Knife for Investigating Data-Driven Biological Ontologies. *Cell Syst.* **8**, 267–273.e3 (2019). doi:10.1016/j.cels.2019.02.003

133. M. Kramer, J. Dutkowski, M. Yu, V. Bafna, T. Ideker, Inferring gene ontologies from pairwise similarity data. *Bioinformatics* **30**, i34–i42 (2014). doi:10.1093/bioinformatics/btu282

134. J. Yang, J. Leskovec, in *Proceedings of the 6th ACM International Conference on Web Search and Data Mining* (2013), pp. 587–596 .

135. V. D. Blondel, J.-L. Guillaume, R. Lambiotte, E. Lefebvre, Fast unfolding of communities in large networks. *J. Stat. Mech.* **2008**, P10008 (2008). doi:10.1088/1742- 5468/2008/10/P10008

136. R. Tibshirani, Regression shrinkage and selection via the lasso. *J. R. Stat. Soc. B* **58**, 267– 288 (1996). doi:10.1111/j.2517-6161.1996.tb02080.x

137. C. H. Mermel, S. E. Schumacher, B. Hill, M. L. Meyerson, R. Beroukhim, G. Getz, GISTIC2.0 facilitates sensitive and confident localization of the targets of focal somatic copy-number alteration in human cancers. *Genome Biol.* **12**, R41 (2011). doi:10.1186/gb- 2011-12-4-r41

138. D. Pratt, J. Chen, D. Welker, R. Rivas, R. Pillich, V. Rynkov, K. Ono, C. Miello, L. Hicks, S. Szalma, A. Stojmirovic, R. Dobrin, M. Braxenthaler, J. Kuentzer, B. Demchak, T. Ideker, NDEx, the Network Data Exchange. *Cell Syst.* **1**, 302–305 (2015). doi:10.1016/j.cels.2015.10.001

139. D. Pratt, J. Chen, R. Pillich, V. Rynkov, A. Gary, B. Demchak, T. Ideker, NDEx 2.0: A Clearinghouse for Research on Cancer Pathways. *Cancer Res.* **77**, e58–e61 (2017). doi:10.1158/0008-5472.CAN-17-0606

140. P. Shannon *et al*., Cytoscape: A software environment for integrated models of biomolecular interaction networks. *Genome Res.* **13**, 2498–2504 (2003). doi:10.1101/gr.1239303

141. F. Zheng, Integrated association stringency (IAS) network (2021); https://doi.org/10.5281/zenodo.4516939.

142. F. Zheng, IAS catalog of multi-scale protein systems (2021); http://doi.org/10.5281/zenodo.4654236.

143. F .Zheng, NeST (Nested Systems in Tumor) Cytoscape session (2021); http://doi.org/10.5281/zenodo.4726267.