

## **Hierarchical organization of the human cell from a cancer coessentiality network**

Eiru Kim<sup>1</sup>, Merve Dede<sup>1,2</sup>, Walter F. Lenoir<sup>1,2</sup>, Gang Wang<sup>1</sup>, Sanjana Srinivasan<sup>1,2</sup>, Medina Colic<sup>2</sup>, Traver Hart<sup>1,\*\*</sup>

<sup>1</sup> Department of Bioinformatics and Computational Biology, The University of Texas MD Anderson Cancer Center, Houston, TX USA

<sup>2</sup> UTHealth Graduate School of Biomedical Sciences, The University of Texas MD Anderson Cancer Center, Houston, TX USA

\*\*To whom correspondence should be addressed: [traver@hart-lab.org](mailto:traver@hart-lab.org)

### **Abstract**

Genetic interactions mediate the emergence of phenotype from genotype. Systematic survey of genetic interactions in yeast showed that genes operating in the same biological process have highly correlated genetic interaction profiles, and this observation has been exploited to infer gene function in model organisms. Systematic surveys of digenic perturbations in human cells are also highly informative, but are not scalable, even with CRISPR-mediated methods. As an alternative, we developed an indirect method of deriving functional interactions. We show that genes having correlated knockout fitness profiles across diverse, non-isogenic cell lines are analogous to genes having correlated genetic interaction profiles across isogenic query strains, and similarly implies shared biological function. We constructed a network of genes with correlated fitness profiles across 400 CRISPR knockout screens in cancer cell lines into a “coessentiality network,” with up to 500-fold enrichment for co-functional gene pairs, enabling strong inference of human gene function. Modules in the network are connected in a layered web that gives insight into the hierarchical organization of the cell.

## Introduction

Genetic interactions govern the translation of genotype to phenotype at every level, from the function of subcellular molecular machines to the emergence of complex organismal traits. In the budding yeast *Saccharomyces cerevisiae*, systematic genetic deletion studies showed that only ~1,100 of its ~6,000 genes (~20%) were required for growth under laboratory conditions (Giaever et al., 2002). A systematic survey of digenic knockouts, however, yielded hundreds of thousands of gene pairs whose double knockout induced a fitness phenotype significantly more severe (synergistic genetic interactions) or less severe (suppressor interactions) than expected from each gene's single mutant fitness (Costanzo et al., 2010, 2016; Tong et al., 2001), with triple-mutant screens adding yet another layer of complexity (Kuzmin et al., 2018). When trying to decipher the genetic contribution to as simple a phenotype as fitness, then, there are vastly more candidate explanations involving genetic interactions than monogenic fitness effects. Moreover, the impact of each gene variant not only depends on the sum of all other genetic variants in the cell, but also is strongly influenced by the cell's environment (Bandyopadhyay et al., 2010; Hillenmeyer et al., 2008).

Patterns of genetic interaction are deeply informative. Genetic interactions frequently occur either within members of the same pathway or process ("within pathway interactions") or between members of parallel pathways ("between pathway interactions") (Kelley and Ideker, 2005). When assayed systematically, the result is that genes that operate in the same biological process tend to interact genetically with the same sets of other genes in discrete, related pathways, culminating in highly correlated genetic interaction profiles across a diverse panel of genetic backgrounds or "query strains." This observation has been exploited extensively to infer gene function in model organisms and, on a smaller scale, in human cells based on similarity of genetic interaction profiles (Lehner et al., 2006; Horn et al., 2011; Bassik et al., 2013; Kampmann et al., 2013; Roguev et al., 2013; Costanzo et al., 2016). Therefore, beyond the specific interactions themselves, a gene's pattern of fitness phenotypes across a diverse set of backgrounds can inform our knowledge of that gene's function.

Translating these concepts into human cells has proved biologically and technically challenging. The *S. cerevisiae* genome has less than one-third the number of protein coding genes as humans, and, despite the quantum leap in technology that the CRISPR/Cas system offers to mammalian forward genetics, yeast remains far simpler to perturb reliably in the lab. Several groups have applied digenic perturbation technologies, using both shRNA and CRISPR, to find cancer genotype-specific synthetic lethals for drug targeting (Han et al., 2017; Najm et al., 2018; Sage et al., 2017; Shen et al., 2017; Wong et al., 2016); (Du et al., 2017) and to identify genetic interactions that enhance or suppress phenotypes related to drug and toxin resistance (Bassik et al., 2013; Jost et al., 2017; Roguev et al., 2013). The current state of the art in CRISPR-mediated gene perturbation relies on observations from three independent guide RNA (gRNA) targeting each gene, or nine pairwise perturbations for each gene pair, plus non-targeting or other negative controls. The largest such mapping to date puts the scale of the problem in stark terms: Han *et al.* use a library of 490,000 gRNA doublets – seven times larger than a latest generation whole-genome, single-gene knockout library – to query all pairs of 207 target genes, or ~0.01% of all gene pairs in the human genome (Han et al., 2017).

An additional dimension of the scale problem is that of backgrounds. Whereas one strain of yeast was systematically assayed in fixed media and environmental conditions to create a reference genetic interaction network, no such reference cell exists for humans. Indeed first-generation whole-genome CRISPR screens in cancer cell lines demonstrated that one of the features associated with the hugely increased sensitivity of CRISPR over shRNA (Hart et al., 2014, 2015) was the ability to resolve tissue- and genetic-driven differences in gene essentiality, as well as the unexpected variation in gene essentiality in cell lines with ostensibly similar genetic backgrounds (Hart et al., 2015; Wang et al., 2014).

Nevertheless, small-scale, targeted genetic interaction screens in human cells using both shRNA and CRISPR showed that the architecture of the genetic interaction network holds true across species. Positive and negative genetic interactions within pathways and between related biological processes yield a correlation network with the same properties:

genes with similar profiles of genetic interaction across different backgrounds are often in the same process or complex, providing a strong basis for inference of gene function (Bassik et al., 2013, 2013; Horn et al., 2011; Kampmann et al., 2013, 2014; Roguev et al., 2013). Since digenic perturbation screens are difficult to scale, we considered whether indirect methods of determining functional genomic information might be effective. Whole-genome CRISPR knockout screens have been performed in over 400 cancer and immortalized cell lines, with the bulk coming from Project Achilles using standardized protocols and reagents (Aguirre et al., 2016; Meyers et al., 2017; Tsherniak et al., 2017). We hypothesized that genes having correlated knockout fitness profiles across diverse cell lines would be analogous genes having correlated genetic interaction profiles across specified query backgrounds in the same cells, and would similarly imply shared biological function. We constructed a network of genes with correlated essentiality scores into a “coessentiality network,” from which we identified clusters of genes with high functional coherence. The network provides powerful insight into functional genomics, cancer targeting, and the capabilities and limitations of CRISPR-mediated genetic screening in human cell lines.

## **Results & Discussion**

We considered CRISPR and shRNA whole-genome screen data from multiple libraries and laboratories (Avena (Doench et al., 2014; Meyers et al., 2017), GeCKOv2 (Aguirre et al., 2016), TKO (Hart et al., 2015, 2017; Steinhart et al., 2017), Sabatini (Wang et al., 2014, 2017) the Moffat shRNA library (Koh et al., 2012; Marcotte et al., 2012, 2016; Medrano et al., 2017)) and other large data sets (McDonald et al., 2017; Tsherniak et al., 2017) (Figure 1a and Supplementary Table 1). From raw read count data, we used the BAGEL pipeline (described in (Hart and Moffat, 2016) and improved here; see Supplementary Methods) to generate Bayes Factors for each gene in each cell line. We removed nontargeting and nonhuman gene controls and quantile normalized each data set, yielding an essentiality score where a positive value indicates a strong knockout fitness defect and a negative value generally implies no phenotype (see Supplemental Methods for details). Each gene therefore has an “essentiality profile” of its scores across the screens in that data set.

For each data set, we ranked gene pairs by correlated essentiality profiles and measured the enrichment for co-functional pairs (see Methods). Data from Meyers et al, where CRISPR knockout screens were conducted using the Avana library in 342 cancer cell lines, showed the strongest enrichment for co-functional gene pairs (Figure 1b), likely due to the relatively high quality of the screens (Supplementary Figure 1) as well as the lineage and genetic diversity of the cells being screened. In contrast, screens from Wang et al. (Wang et al., 2014, 2017) were equally high quality but were performed only in 17 AML cell lines with correspondingly limited diversity. To further increase the co-functionality signal, we removed screens with poor performance and only considered genes that were hits in at least 3 of the remaining screens; filtering resulted in an additional twofold enrichment for co-functional gene pairs (Figure 1b and Supplementary Figure 1). The filtered data from Meyers et al. (Meyers et al., 2017) (n=276 cell lines; 5,387 genes; hereafter “Avana data”) was used for all subsequent analysis. We selected gene pairs with a Bonferroni-corrected P-value < 0.05 and combined them into a network, the Cancer Coessentiality Network, containing 3,327 genes connected by 68,641 edges (Figure 1c). The network is highly modular, with clusters showing strong functional coherence, similar to the networks directly inferred from correlated yeast genetic interaction profiles.(Costanzo et al., 2010, 2016)

### ***Essential genes specific to oncogenic contexts***

The data underlying the Cancer Coessentiality Network is derived from well-characterized cancer cell lines from 30+ lineages, representing the major oncogenic mutation profiles common to those cancers. Many clusters in the network can therefore be associated with specific tissues and cancer-relevant genotypes. By testing cluster-level essentiality profiles for tissue specificity (see Supplementary Methods), we identified a number of clusters that correspond to tissue specific cancers (Figure 2a), which in turn contain the characteristic oncogenes. For example, Cluster 14 (Figure 2b) consists of BRAF and related genes that are highly specific to BRAF-mutated melanoma cells ( $P < 10^{-12}$ ; Figure 2c). The cluster contains other elements of the MAP kinase pathway (MAP2K1, MAPK1, DUSP4), indicating their

essentiality in BRAF-mutant cells, supporting efforts to incorporate ERK inhibitors into combinatorial therapies to overcome resistance to targeted BRAF treatments (Smalley and Smalley, 2018). This example highlights the utility of this indirect approach to identify synthetic lethal interactions: genes co-essential with oncogenes are synthetic lethals. Beyond the downstream elements of the MAPK pathway itself, the BRAF/melanoma cluster also contains the transcription factors (TFs) Melanogenesis Associated Transcription Factor (MITF); two developmental Sry-related box (Sox) genes, SOX9 and SOX10; and mesenchymal marker ZEB2, indicative of the non-epithelial origin of melanocyte cells and providing insight into the genetic requirements for tissue differentiation in this lineage.

Similar observations hold for other tissue-specific oncogenes. Cluster 17, essential in lymphoid cell lines ( $P < 10^{-7}$ ), contains oncogene FLI1 and tissue-specific transcription factor MYB (Figure 2d-e), and cluster 38 is enriched for ovarian cancer cells ( $P < 10^{-7}$ ) and carries lineage-specific TF PAX8, previously shown to be essential in these cells (Cheung et al., 2011) (Figure 2f-g). Cluster 75, essential in colorectal cancer cells ( $P < 10^{-9}$ ), contains  $\beta$ -catenin (CTNNB1) and transcription factor partner TCF7L2 (Figure 2h-i); both are linked to E2 ubiquitin ligase UBE2Z, which mediates UBA6-specific suppression of EMT (Liu et al., 2017), indicating a functional linkage with  $\beta$ -catenin signaling. Additional tissue- and oncogene-driven clusters delineating breast cancer subtypes, joint cyclin/cyclin-dependent kinase (CDK) dependencies, candidate KRAS synthetic lethals, and glioblastoma-specific essential genes are shown in Supplementary Figure 2.

Neuroblastoma cells require MYCN, the neuroblastoma-specific paralog of the MYC oncogene (Huang and Weiss, 2013), as well as nervous system developmental transcription factor SOX11 (Pötzner et al., 2010) (Figure 2j). Interestingly, MYC is highly essential in virtually all non-neuroblastoma cell lines, resulting in a relatively uncommon anticorrelation in essentiality profiles ( $r = -0.49$ ;  $P < 10^{-17}$ ; Figure 2k). While this negative correlation is driven by mutual exclusivity in tissue essentiality, we also observe anticorrelation in tumor suppressors and their repressors in the same cells. CRISPR knockout of tumor suppressors in cells carrying wildtype alleles frequently results in increased growth rate, which manifests as extreme negative BF scores with our updated

BAGEL algorithm (see Supplementary Methods). TP53 has these extreme negative values in melanoma cells and others with unmutated genes, resulting in strong anticorrelation with TP53 suppressors MDM2 ( $r=-0.86$ ,  $P<10^{-81}$ ), MDM4 ( $r=-0.61$ ,  $P<10^{-28}$ ), and PPM1D ( $r=-0.72$ ,  $P<10^{-44}$ ) (Figure 2l-m).

While TP53 shows the characteristic extreme negative essentiality score of a tumor suppressor gene in wildtype backgrounds, surprisingly, it causes a growth defect when knocked out in three cell lines: HCT1143 breast cancer, PC14 lung cancer, and NB4 AML cells (Supplementary Figure 3a). All three carry the R248Q oncogenic mutation; in fact, R248Q is weakly predictive of TP53 essentiality generally, and strongly predictive when it is the only P53 mutation detected (Supplementary Figure 3b). Nor is this the only case where a tumor suppressor in one background is an essential gene in another: the von Hippel-Landau tumor suppressor gene VHL shows no phenotype in renal cancer cells, where the gene is nearly universally deleted, but is essential specifically in BTFC-909 renal carcinoma cells which lack the characteristic Chr3 copy loss (Sinha et al., 2017). In contrast, VHL shows a fitness defect when knockout out in most other backgrounds (Supplementary Figure 3c). The essentiality profile for VHL is strongly correlated with EGLN1 (commonly called PHD2), an oxygen sensor that hydroxylates hypoxia response genes HIF1A and HIF2A, marking them for degradation by the VHL complex in normoxic environments (Berra et al., 2003). EGLN1 essentiality is overrepresented in melanoma cells ( $P<10^{-4}$ , rank-sum test; essential in 14 of 22 skin cancer cell lines).

### ***A high-precision functional interaction map of human genes***

These examples indicate the breadth and precision of the coessentiality network, but represent results from hypothesis-guided queries. In an effort to learn novel associations from the data, we tested each cluster for its correlation with cell lineage as well as correlation with gene expression, mutation, and copy number amplification of all genes both inside and outside the cluster to identify underlying molecular genetic drivers of modular, emergent essentiality. We identified 270 genes in 30 clusters whose essentiality profiles strongly correlated with their own copy number profiles but not their expression

profiles (Figure 3a). As copy number amplification is a known source of false positives in CRISPR screens, we labeled these clusters as amplification artifacts. An additional 56 genes in 11 clusters showed significant association with both copy number and expression profiles. These clusters notably include KRAS amplifications in pancreatic and colorectal cancer (Cluster 276), ERBB2 amplifications in breast and other cancers (Cluster 52), and CCNE1 overexpression/RB1 mutation. (Cluster 101), consistent with well-studied patterns of oncogenesis.

Given the underlying data, it is perhaps not surprising that oncogenic signatures are clearly evident in the coessentiality network. However, the vast majority of the network structure does not appear to be driven by tissue specificity or mutational signatures. The network contains information complementary to existing functional (Figure 3b) and physical (Figure 3c) interaction networks, and the network derived from Avena data exhibits far greater coverage than equivalent networks from the GeCKOv2 subset of Project Achilles (Aguirre et al., 2016) or Wang (Wang et al., 2017) AML-specific data (Figure 3d). Nevertheless, the remaining network modules show strong functional coherence (Figure 3a). Coessentiality often proves a stronger predictor of complex membership than coexpression (Figure 3e), and this signature is reflected in the network clusters we identified. Indeed 53 clusters, comprised of 1,422 genes, show enrichment for CORUM-annotated protein complexes at  $P$ -value  $< 10^{-6}$ , and fitness profiles have been used to implicate additional members of protein complexes (Pan et al.). However, this holds only for genes whose knockout fitness defects vary across cell lines; coessentiality of core essential genes is poorly predictive of co-complex membership (Supplementary Figure 4). All 53 CORUM-annotated clusters, plus an additional 44 clusters containing 413 genes (totaling 97 network modules with 1,835 genes), show enrichment for GO biological process, cellular component, KEGG pathway, or Reactome pathway annotations at a similarly strict threshold. (All network cluster annotations can be found in the master annotation file, Supplementary Table 7). In addition, we evaluated the relative performance of the coessentiality network by measuring its ability to recover cancer gene sets using DisGeNET (Huang et al., 2018). The coessentiality network ranks comparably with other large functional networks (Figure 3g), while starting from a much smaller data set,



suggesting that the coessentiality network explains not only protein complexes but also cancer pathways including interactions between protein complexes and signaling transduction.

Epistatic interactions frequently underlie covariation in fitness profiles (Phillips, 2008). Cluster 2 is highly enriched for genes involved in the mitochondrial electron transport chain, including 30 of 48 genes encoding subunits of NADH dehydrogenase complex (ETC Complex I;  $P < 10^{-42}$ ) plus additional subunits of all other ETC complexes. The cluster also contains 49 of 51 subunits of the mitochondrial large ribosomal subunit ( $P < 10^{-87}$ ), 23 of 25 members of the small subunit ( $P < 10^{-39}$ ), plus 20 mitochondrion-specific tRNA synthases ( $P < 10^{-20}$ ). This mitochondrial translation machinery is required for the synthesis of proteins in the ETC complexes. These genes' inclusion in this cluster, where their essentiality profiles are correlated with those of the complexes they support, reflects a fundamental feature of saturating genetic screens: the essentiality of a given enzyme or biological process is matched by the essentiality of the cellular components required for the biogenesis and maintenance of that process.

We observe numerous additional instances of such epistatic interactions that highlight functional relationships. For example, glutathione peroxidase gene GPX4 shows highly variable essentiality across cell lines (Figure 4a). GPX4 is a selenoprotein that contains the cysteine analog selenocysteine (Sec), the “21<sup>st</sup> amino acid,” at its active site. Coessential with GPX4 are all the genes required for conversion of serine-conjugated tRNA<sup>Ser</sup> to selenocysteine-conjugated tRNA<sup>Sec</sup> (PSTK, SEPHS2, SEPSECS), as well as selenocysteine-specific elongation factor EEFSEC, which guides Sec-tRNA<sup>Sec</sup> to specific UGA codons. (Figure 4b) (Schoenmakers et al., 2016). Cellular dependence on GPX4 was recently shown to be associated with mesenchymal state (Viswanathan et al., 2017). Our analysis corroborates this finding: we show that GPX4 essentiality is higher in cells expressing mesenchymal marker ZEB1 ( $P < 10^{-5}$ ; Figure 4c). In our global analysis, however, GPX4 sensitivity is more highly correlated with the expression level of GPX2, another member of the glutathione peroxidase family (Figure 4d).

Similarly, a pair of genes, ACOX1 and HSD17B4, which encode three of the four enzymatic steps in peroxisomal fatty acid  $\beta$ -oxidation (FAO), are found in a cluster with ten PEX genes involved in peroxisome biogenesis, maintenance, and membrane transport (Figure 5a-b). The cluster shows a discrete pattern of essentiality, preferentially in lung cells (essential in 6/42 lung cancer lines in the Avana data; Figure 5c) but also appearing intermittently in other lineages. The cluster is not otherwise significantly associated with mutational, copy number, or lineage features. The emergent dependence on peroxisomal FAO is associated with overexpression of 18 genes at P-value  $< 10^{-4}$  (See Methods and Supp Table X); two of these include ACOXL, a largely uncharacterized gene similar to ACOX1, and ELOVL7, which catalyzes the rate-limiting step involved in long-chain fatty acid elongation. Notably, this cluster is intact in the network generated from Aguirre et al. (Aguirre et al., 2016)(Figure 5d), though it preferentially arises in pancreatic cells rather than lung cells.

### ***A network of interactions between biological processes***

While individual clusters show high functional coherence, the network of connections between clusters offers a unique window into process-level interactions in human cells. The peroxisomal FAO cluster is strongly connected to another functionally coherent module containing 12 genes, ten of which are tightly connected to other members of the cluster (Figure 5a). Those ten include seven genes whose proteins reside in the ER, five of which regulate cholesterol biosynthesis via posttranslational modification of sterol regulator element binding proteins (SREBPs). The remaining three genes, DHRS7B, TMEM41A, and C12orf49, are largely or completely uncharacterized; their strong association with other genes in this cluster implicates a role in the SREBP maturation pathway. Both the peroxisomal FAO cluster and the SREBP maturation cluster are strongly linked with a module containing RAB18, a RAS-related GTPase involved in Golgi-to-ER retrograde transport, as well as its associated GAP (RAB3GAP1, RAB3GAP2) and GEF (TBC1D120) (Feldmann et al., 2017).

A similar network of modules describes the regulation of the mechanistic Target of Rapamycin (mTOR), in particular its detection of amino acid levels. Figure 6 shows the

relationships between a series of network modules describing the core mTOR pathway and several regulatory modules. The mTOR cluster includes mTORC1/2 subunits MTOR, MLST8, MAPKAP1, and RICTOR (mTORC1-specific subunit RAPTOR is never essential and therefore absent from the network); canonical mTORC1/mTORC2 regulatory and signaling components PDK1, AKT1, and PIK3CB; plus G-protein subunit GNB2, previously shown to physically interact with mTOR in response to serum stimulation (Robles-Molina et al., 2014). Canonical inhibition of mTOR by the TSC1/TSC2 heterodimer – the TSC1-TSC2 link is the top-ranked correlation in the entire dataset, with  $\rho=0.93$  ( $P<10^{-117}$ ) – is reflected in the anticorrelation of fitness profiles connecting the TSC1/2 cluster and the mTOR cluster.

MTOR response to cellular amino acid levels is modulated by an alternative pathway that functions at the lysosomal membrane (Bar-Peled and Sabatini, 2014). We identify a large cluster containing several genes involved in lysosomal protein and transport, including the HOPS complex (Balderhaar and Ungermann, 2013; Jiang et al., 2014) and the VPS26/29/35 retromer complex (Hierro et al., 2007; Seaman, 2012). This strongly connected cluster also contains the Rag GTPases RagA (RRAGA) and RagC (RRAGC) that transmit information on amino acid abundance to mTORC1 (Bar-Peled and Sabatini, 2014). The Rag GTPases are in turn activated by the Ragulator complex (Bar-Peled et al., 2012; Sancak et al., 2010) and folliculin (FLCN) (Mu et al., 2017), also members of the cluster. The GATOR-1 complex is a nonessential suppressor of essential Rag GTPase activity (Bar-Peled et al., 2013) and is therefore absent from our network, but essential suppression of GATOR-1 by GATOR-2 (Bar-Peled et al., 2013; Wei et al., 2014) is reflected by the strong linkage of the GATOR-2 complex to both the Ragulator and mTOR complexes.

Within the MTOR meta-cluster, we further identify a complex containing three regulators of protein phosphatase 2A (LCMT1, TIPRL, PTPA), whose strong connectivity to the TSC1/2 complex may suggest a regulatory role for PP2A in MTOR signaling. PP2A has previously been posited to be an activator of TSC1/2 upstream of MTOR (Vereshchagina et al., 2008); the coessentiality network suggests specific PP2A regulators that may mediate this regulation.

A third example of the process-level interactions in cells demonstrates the hierarchy of operations required for posttranslational maturation of cell surface receptors. Several clusters in our network describe the ER-associated glycosylation pathways (Figure 7a-b), including synthesis of lipid-linked sugars via the dolichol-phosphate-mannose (DPM) pathway (Ashida et al., 2006; Maeda and Kinoshita, 2008) and extension via the mannosyltransferase family. Glycan chains are transferred to asparagine residues of target proteins via the N-oligosaccharyltransferase (OST) complex. Nascent polypeptide chains are glycosylated as they are cotranslationally translocated into the ER, a process facilitated by signal sequence receptor dimer SSR1/SSR2, and ER-specific Hsp90 chaperone HSP90B1 facilitates proper folding. The OST complex and its functional partners are represented in a single large complex (Figure 7a). Both DPM and OST are highly connected to the large complex encoding GPI anchor synthesis; DPM is required for GPI anchor production (Kinoshita and Inoue, 2000; Watanabe et al., 1998) before transfer to target proteins.

The variety of oncogenic drivers among the cell lines underlying this network give rise to background-specific dependencies, including a variety of mutated and/or amplified receptor tyrosine kinases (RTKs) with specific, and mutually exclusive, essentiality profiles. Insulin-like growth factor receptor IGF1R is one such RTK, which appears in a cluster with receptor-specific downstream signaling proteins insulin-receptor substrate 1 and 2 (IRS1, IRS2). IGF1R is a highly N-glycosylated RTK and the IGF1R complex is tightly connected to the OST complex in our network. EGFR, also highly glycosylated (Kaszuba et al., 2015), appears in its own cluster with signaling adapter protein SHC1 and is also linked to the OST complex (Figure 7a) despite being mutually exclusive with IGF1R (Supp Figure 2). Interestingly, EGFR is more strongly connected with a separate complex involved in glycosphingolipid biosynthesis (that is itself linked to the OST complex). Prior work suggests that membrane glycolipid composition can strongly influence EGFR autophosphorylation and signaling (Coskun et al., 2011). In contrast, fibroblast growth factor receptor FGFR1 is absent from this meta-network but is strongly associated with heparin sulfate biosynthesis (Supp Fig 2); HS is a known mediator of FGF receptor-ligand interaction (Wu et al., 2003).

## Conclusions

Systematic genetic interaction screens in yeast revealed that most genetic interactions occur either within a biological pathway or between related pathways. We demonstrate that single-gene fitness profiles across screens in genetically diverse human cell lines are analogous to genetic interaction screens across defined isogenic query strains. Importantly, as with model organisms, human genes with correlated fitness profiles are highly likely to participate in the same biological process. We take advantage of this fundamental architectural feature of genetic networks to create a functional interaction map of bioprocesses that demonstrates information flow through a human cell. The network predicts gene function and provides a view of process-level interactions in human cells, allowing a level of abstraction beyond the gene-centric approach frequently employed.

The network is derived from the emergent essentiality of defined biological processes and the genes required to execute them. We show that this approach significantly expands our knowledge beyond current networks of comparable design (e.g. STRING, HumanNet). A critical next step will be to understand the underlying context that drives the emergent essentiality of specific bioprocesses in specific backgrounds. The health implications of this question are profound. In cancer, to understand the causal basis of modular emergent essentiality is to identify matched pairs of biomarkers (the causal basis) and precision targets (the essential pathway) for personalized chemotherapeutic treatment. Additionally, lineage-specific essential processes could provide explanatory power for germline mutations causing tissue-specific disease presentation, in cancer as well as other diseases.

Expanding the coverage of the network will require different screening approaches. Fitness screens in cancer cell lines in rich media will miss cellular dependencies that are present only under stress conditions. In yeast (Hillenmeyer et al., 2008) and nematodes (Ramani et al., 2012), these context-dependent fitness effects comprise the majority of genes in the genome. Increasing the coverage of the genetic interaction network beyond the ~3,000 genes whose fitness profiles covary across human cancer cell lines will require screening in

different nutrients and perturbagens, as well as sampling the effects of genetic mutations outside common cancer genotypes. Nevertheless, the indirect approach to identifying genetic interactions from monogenic perturbation studies is demonstrably effective, and offers a powerful tool for navigating the network of connections between cellular bioprocesses. The coessentiality network used in this study can be viewed interactively at [https://pickles-hartlab.shinyapps.io/cyto\\_app/](https://pickles-hartlab.shinyapps.io/cyto_app/) and downloaded at the NDEx project.

## **Methods**

See Supplementary Data for a complete description of methods used in this study.

## **Acknowledgments**

The authors would like to acknowledge the scientists, administrators, and funding agents behind Project Achilles. Without their commitment to rapid release of open access data, none of this work would have been possible.

Funding was supplied by the MD Anderson Cancer Center Support Grant P30 CA016672 (the Bioinformatics Shared Resource) and the Cancer Prevention Research Institute of Texas (CPRIT) grant RR160032.

## References

- Aguirre, A.J., Meyers, R.M., Weir, B.A., Vazquez, F., Zhang, C.-Z., Ben-David, U., Cook, A., Ha, G., Harrington, W.F., Doshi, M.B., et al. (2016). Genomic Copy Number Dictates a Gene-Independent Cell Response to CRISPR/Cas9 Targeting. *Cancer Discov.* 6, 914–929.
- Ashida, H., Maeda, Y., and Kinoshita, T. (2006). DPM1, the Catalytic Subunit of Dolichol-phosphate Mannose Synthase, Is Tethered to and Stabilized on the Endoplasmic Reticulum Membrane by DPM3. *J. Biol. Chem.* 281, 896–904.
- Balderhaar, H.J. kleine, and Ungermann, C. (2013). CORVET and HOPS tethering complexes - coordinators of endosome and lysosome fusion. *J. Cell Sci.* 126, 1307–1316.
- Bandyopadhyay, S., Mehta, M., Kuo, D., Sung, M.-K., Chuang, R., Jaehnig, E.J., Bodenmiller, B., Licon, K., Copeland, W., Shales, M., et al. (2010). Rewiring of genetic networks in response to DNA damage. *Science* 330, 1385–1389.
- Bar-Peled, L., and Sabatini, D.M. (2014). Regulation of mTORC1 by amino acids. *Trends Cell Biol.* 24, 400–406.
- Bar-Peled, L., Schweitzer, L.D., Zoncu, R., and Sabatini, D.M. (2012). Ragulator is a GEF for the rag GTPases that signal amino acid levels to mTORC1. *Cell* 150, 1196–1208.
- Bar-Peled, L., Chantranupong, L., Cherniack, A.D., Chen, W.W., Ottina, K.A., Grabiner, B.C., Spear, E.D., Carter, S.L., Meyerson, M., and Sabatini, D.M. (2013). A Tumor suppressor complex with GAP activity for the Rag GTPases that signal amino acid sufficiency to mTORC1. *Science* 340, 1100–1106.
- Bassik, M.C., Kampmann, M., Lebbink, R.J., Wang, S., Hein, M.Y., Poser, I., Weibezahn, J., Horlbeck, M.A., Chen, S., Mann, M., et al. (2013). A systematic mammalian genetic interaction map reveals pathways underlying ricin susceptibility. *Cell* 152, 909–922.
- Berra, E., Benizri, E., Ginouvès, A., Volmat, V., Roux, D., and Pouyssegur, J. (2003). HIF prolyl-hydroxylase 2 is the key oxygen sensor setting low steady-state levels of HIF-1 $\alpha$  in normoxia. *EMBO J.* 22, 4082–4090.
- Cheung, H.W., Cowley, G.S., Weir, B.A., Boehm, J.S., Rusin, S., Scott, J.A., East, A., Ali, L.D., Lizotte, P.H., Wong, T.C., et al. (2011). Systematic investigation of genetic vulnerabilities across cancer cell lines reveals lineage-specific dependencies in ovarian cancer. *Proc. Natl. Acad. Sci. U. S. A.* 108, 12372–12377.
- Coskun, Ü., Grzybek, M., Drechsel, D., and Simons, K. (2011). Regulation of human EGF receptor by lipids. *Proc. Natl. Acad. Sci. U. S. A.* 108, 9044–9048.
- Costanzo, M., Baryshnikova, A., Bellay, J., Kim, Y., Spear, E.D., Sevier, C.S., Ding, H., Koh, J.L.Y., Toufighi, K., Mostafavi, S., et al. (2010). The genetic landscape of a cell. *Science* 327, 425–431.



Costanzo, M., VanderSluis, B., Koch, E.N., Baryshnikova, A., Pons, C., Tan, G., Wang, W., Usaj, M., Hanchard, J., Lee, S.D., et al. (2016). A global genetic interaction network maps a wiring diagram of cellular function. *Science* 353.

Doench, J.G., Hartenian, E., Graham, D.B., Tothova, Z., Hegde, M., Smith, I., Sullender, M., Ebert, B.L., Xavier, R.J., and Root, D.E. (2014). Rational design of highly active sgRNAs for CRISPR-Cas9-mediated gene inactivation. *Nat. Biotechnol.* 32, 1262–1267.

Du, D., Roguev, A., Gordon, D.E., Chen, M., Chen, S.-H., Shales, M., Shen, J.P., Ideker, T., Mali, P., Qi, L.S., et al. (2017). Genetic interaction mapping in mammalian cells using CRISPR interference. *Nat. Methods* 14, 577–580.

Feldmann, A., Bekbulat, F., Huesmann, H., Ulbrich, S., Tatzelt, J., Behl, C., and Kern, A. (2017). The RAB GTPase RAB18 modulates macroautophagy and proteostasis. *Biochem. Biophys. Res. Commun.* 486, 738–743.

Giaever, G., Chu, A.M., Ni, L., Connelly, C., Riles, L., Véronneau, S., Dow, S., Lucau-Danila, A., Anderson, K., André, B., et al. (2002). Functional profiling of the *Saccharomyces cerevisiae* genome. *Nature* 418, 387–391.

Han, K., Jeng, E.E., Hess, G.T., Morgens, D.W., Li, A., and Bassik, M.C. (2017). Synergistic drug combinations for cancer identified in a CRISPR screen for pairwise genetic interactions. *Nat. Biotechnol.* 35, 463–474.

Hart, T., and Moffat, J. (2016). BAGEL: a computational framework for identifying essential genes from pooled library screens. *BMC Bioinformatics* 17, 164.

Hart, T., Brown, K.R., Sircoulomb, F., Rottapel, R., and Moffat, J. (2014). Measuring error rates in genomic perturbation screens: gold standards for human functional genomics. *Mol. Syst. Biol.* 10, 733.

Hart, T., Chandrashekhara, M., Aregger, M., Steinhart, Z., Brown, K.R., MacLeod, G., Mis, M., Zimmermann, M., Fradet-Turcotte, A., Sun, S., et al. (2015). High-Resolution CRISPR Screens Reveal Fitness Genes and Genotype-Specific Cancer Liabilities. *Cell* 163, 1515–1526.

Hart, T., Tong, A.H.Y., Chan, K., Van Leeuwen, J., Seetharaman, A., Aregger, M., Chandrashekhara, M., Hustedt, N., Seth, S., Noonan, A., et al. (2017). Evaluation and Design of Genome-Wide CRISPR/SpCas9 Knockout Screens. *G3 Bethesda Md* 7, 2719–2727.

Hierro, A., Rojas, A.L., Rojas, R., Murthy, N., Effantin, G., Kajava, A.V., Steven, A.C., Bonifacino, J.S., and Hurley, J.H. (2007). Functional architecture of the retromer cargo-recognition complex. *Nature* 449, 1063–1067.

Hillenmeyer, M.E., Fung, E., Wildenhain, J., Pierce, S.E., Hoon, S., Lee, W., Proctor, M., St Onge, R.P., Tyers, M., Koller, D., et al. (2008). The chemical genomic portrait of yeast: uncovering a phenotype for all genes. *Science* 320, 362–365.

Horn, T., Sandmann, T., Fischer, B., Axelsson, E., Huber, W., and Boutros, M. (2011). Mapping of signaling networks through synthetic genetic interaction analysis by RNAi. *Nat. Methods* 8, 341–346.

Huang, M., and Weiss, W.A. (2013). Neuroblastoma and MYCN. *Cold Spring Harb. Perspect. Med.* 3.

Huang, J.K., Carlin, D.E., Yu, M.K., Zhang, W., Kreisberg, J.F., Tamayo, P., and Ideker, T. (2018). Systematic Evaluation of Molecular Networks for Discovery of Disease Genes. *Cell Syst.* 6, 484-495.e5.

Jiang, P., Nishimura, T., Sakamaki, Y., Itakura, E., Hatta, T., Natsume, T., and Mizushima, N. (2014). The HOPS complex mediates autophagosome-lysosome fusion through interaction with syntaxin 17. *Mol. Biol. Cell* 25, 1327–1337.

Jost, M., Chen, Y., Gilbert, L.A., Horlbeck, M.A., Krenning, L., Menchon, G., Rai, A., Cho, M.Y., Stern, J.J., Protá, A.E., et al. (2017). Combined CRISPRi/a-Based Chemical Genetic Screens Reveal that Rigosertib Is a Microtubule-Destabilizing Agent. *Mol. Cell* 68, 210-223.e6.

Kampmann, M., Bassik, M.C., and Weissman, J.S. (2013). Integrated platform for genome-wide screening and construction of high-density genetic interaction maps in mammalian cells. *Proc. Natl. Acad. Sci. U. S. A.* 110, E2317-2326.

Kampmann, M., Bassik, M.C., and Weissman, J.S. (2014). Functional genomics platform for pooled screening and generation of mammalian genetic interaction maps. *Nat. Protoc.* 9, 1825–1847.

Kaszuba, K., Grzybek, M., Orłowski, A., Danne, R., Róg, T., Simons, K., Coskun, Ü., and Vattulainen, I. (2015). N-Glycosylation as determinant of epidermal growth factor receptor conformation in membranes. *Proc. Natl. Acad. Sci. U. S. A.* 112, 4334–4339.

Kelley, R., and Ideker, T. (2005). Systematic interpretation of genetic interactions using protein networks. *Nat. Biotechnol.* 23, 561–566.

Kinoshita, T., and Inoue, N. (2000). Dissecting and manipulating the pathway for glycosylphosphatidylinositol-anchor biosynthesis. *Curr. Opin. Chem. Biol.* 4, 632–638.

Koh, J.L.Y., Brown, K.R., Sayad, A., Kasimer, D., Ketela, T., and Moffat, J. (2012). COLT-Cancer: functional genetic screening resource for essential genes in human cancer cell lines. *Nucleic Acids Res.* 40, D957-963.

Kuzmin, E., VanderSluis, B., Wang, W., Tan, G., Deshpande, R., Chen, Y., Usaj, M., Balint, A., Mattiazzi Usaj, M., van Leeuwen, J., et al. (2018). Systematic analysis of complex genetic interactions. *Science* 360.

Lehner, B., Crombie, C., Tischler, J., Fortunato, A., and Fraser, A.G. (2006). Systematic mapping of genetic interactions in *Caenorhabditis elegans* identifies common modifiers of diverse signaling pathways. *Nat. Genet.* *38*, 896–903.

Liu, X., Sun, L., Gursel, D.B., Cheng, C., Huang, S., Rademaker, A.W., Khan, S.A., Yin, J., and Kiyokawa, H. (2017). The non-canonical ubiquitin activating enzyme UBA6 suppresses epithelial-mesenchymal transition of mammary epithelial cells. *Oncotarget* *8*, 87480–87493.

Maeda, Y., and Kinoshita, T. (2008). Dolichol-phosphate mannosyltransferase: structure, function and regulation. *Biochim. Biophys. Acta* *1780*, 861–868.

Marcotte, R., Brown, K.R., Suarez, F., Sayad, A., Karamboulas, K., Krzyzanowski, P.M., Sircoulomb, F., Medrano, M., Fedyszyn, Y., Koh, J.L.Y., et al. (2012). Essential gene profiles in breast, pancreatic, and ovarian cancer cells. *Cancer Discov.* *2*, 172–189.

Marcotte, R., Sayad, A., Brown, K.R., Sanchez-Garcia, F., Reimand, J., Haider, M., Virtanen, C., Bradner, J.E., Bader, G.D., Mills, G.B., et al. (2016). Functional Genomic Landscape of Human Breast Cancer Drivers, Vulnerabilities, and Resistance. *Cell* *164*, 293–309.

McDonald, E.R., de Weck, A., Schlabach, M.R., Billy, E., Mavrakis, K.J., Hoffman, G.R., Belur, D., Castelletti, D., Frias, E., Gampa, K., et al. (2017). Project DRIVE: A Compendium of Cancer Dependencies and Synthetic Lethal Relationships Uncovered by Large-Scale, Deep RNAi Screening. *Cell* *170*, 577-592.e10.

Medrano, M., Communal, L., Brown, K.R., Iwanicki, M., Normand, J., Paterson, J., Sircoulomb, F., Krzyzanowski, P., Novak, M., Doodnauth, S.A., et al. (2017). Interrogation of Functional Cell-Surface Markers Identifies CD151 Dependency in High-Grade Serous Ovarian Cancer. *Cell Rep.* *18*, 2343–2358.

Meyers, R.M., Bryan, J.G., McFarland, J.M., Weir, B.A., Sizemore, A.E., Xu, H., Dharia, N.V., Montgomery, P.G., Cowley, G.S., Pantel, S., et al. (2017). Computational correction of copy number effect improves specificity of CRISPR-Cas9 essentiality screens in cancer cells. *Nat. Genet.* *49*, 1779–1784.

Mu, Z., Wang, L., Deng, W., Wang, J., and Wu, G. (2017). Structural insight into the Ragulator complex which anchors mTORC1 to the lysosomal membrane. *Cell Discov.* *3*, 17049.

Najm, F.J., Strand, C., Donovan, K.F., Hegde, M., Sanson, K.R., Vaimberg, E.W., Sullender, M.E., Hartenian, E., Kalani, Z., Fusi, N., et al. (2018). Orthologous CRISPR-Cas9 enzymes for combinatorial genetic screens. *Nat. Biotechnol.* *36*, 179–189.

Pan, J., Meyers, R.M., Michel, B.C., Mashtalir, N., Sizemore, A.E., Wells, J.N., Cassel, S.H., Vazquez, F., Weir, B.A., Hahn, W.C., et al. Interrogation of Mammalian Protein Complex Structure, Function, and Membership Using Genome-Scale Fitness Screens. *Cell Syst.*

Phillips, P.C. (2008). Epistasis—the essential role of gene interactions in the structure and evolution of genetic systems. *Nat. Rev. Genet.* 9, 855–867.

Potzner, M.R., Tsarovina, K., Binder, E., Penzo-Méndez, A., Lefebvre, V., Rohrer, H., Wegner, M., and Sock, E. (2010). Sequential requirement of Sox4 and Sox11 during development of the sympathetic nervous system. *Dev. Camb. Engl.* 137, 775–784.

Ramani, A.K., Chuluunbaatar, T., Verster, A.J., Na, H., Vu, V., Pelte, N., Wannissorn, N., Jiao, A., and Fraser, A.G. (2012). The majority of animal genes are required for wild-type fitness. *Cell* 148, 792–802.

Robles-Molina, E., Dionisio-Vicuña, M., Guzmán-Hernández, M.L., Reyes-Cruz, G., and Vázquez-Prado, J. (2014). G $\beta$ y interacts with mTOR and promotes its activation. *Biochem. Biophys. Res. Commun.* 444, 218–223.

Roguev, A., Talbot, D., Negri, G.L., Shales, M., Cagney, G., Bandyopadhyay, S., Panning, B., and Krogan, N.J. (2013). Quantitative genetic-interaction mapping in mammalian cells. *Nat. Methods* 10, 432–437.

Sage, C., Lawo, S., Panicker, P., Scales, T.M.E., Rahman, S.A., Little, A.S., McCarthy, N.J., Moore, J.D., and Cross, B.C.S. (2017). Dual direction CRISPR transcriptional regulation screening uncovers gene networks driving drug resistance. *Sci. Rep.* 7, 17693.

Sancak, Y., Bar-Peled, L., Zoncu, R., Markhard, A.L., Nada, S., and Sabatini, D.M. (2010). Regulator-Rag complex targets mTORC1 to the lysosomal surface and is necessary for its activation by amino acids. *Cell* 141, 290–303.

Schoenmakers, E., Carlson, B., Agostini, M., Moran, C., Rajanayagam, O., Bochukova, E., Tobe, R., Peat, R., Gevers, E., Muntoni, F., et al. (2016). Mutation in human selenocysteine transfer RNA selectively disrupts selenoprotein synthesis. *J. Clin. Invest.* 126, 992–996.

Seaman, M.N.J. (2012). The retromer complex - endosomal protein recycling and beyond. *J. Cell Sci.* 125, 4693–4702.

Shen, J.P., Zhao, D., Sasik, R., Luebeck, J., Birmingham, A., Bojorquez-Gomez, A., Licon, K., Klepper, K., Pekin, D., Beckett, A.N., et al. (2017). Combinatorial CRISPR-Cas9 screens for de novo mapping of genetic interactions. *Nat. Methods* 14, 573–576.

Sinha, R., Winer, A.G., Chevinsky, M., Jakubowski, C., Chen, Y.-B., Dong, Y., Tickoo, S.K., Reuter, V.E., Russo, P., Coleman, J.A., et al. (2017). Analysis of renal cancer cell lines from two major resources enables genomics-guided cell line selection. *Nat. Commun.* 8, 15165.

Smalley, I., and Smalley, K.S.M. (2018). ERK Inhibition: A New Front in the War against MAPK Pathway-Driven Cancers? *Cancer Discov.* 8, 140–142.

Steinhart, Z., Pavlovic, Z., Chandrashekar, M., Hart, T., Wang, X., Zhang, X., Robitaille, M., Brown, K.R., Jaksani, S., Overmeer, R., et al. (2017). Genome-wide CRISPR screens reveal a

Wnt-FZD5 signaling circuit as a druggable vulnerability of RNF43-mutant pancreatic tumors. *Nat. Med.* *23*, 60–68.

Tong, A.H., Evangelista, M., Parsons, A.B., Xu, H., Bader, G.D., Pagé, N., Robinson, M., Raghbizadeh, S., Hogue, C.W., Bussey, H., et al. (2001). Systematic genetic analysis with ordered arrays of yeast deletion mutants. *Science* *294*, 2364–2368.

Tsherniak, A., Vazquez, F., Montgomery, P.G., Weir, B.A., Kryukov, G., Cowley, G.S., Gill, S., Harrington, W.F., Pantel, S., Krill-Burger, J.M., et al. (2017). Defining a Cancer Dependency Map. *Cell* *170*, 564–576.e16.

Vereshchagina, N., Ramel, M.-C., Bitoun, E., and Wilson, C. (2008). The protein phosphatase PP2A-B' subunit Widerborst is a negative regulator of cytoplasmic activated Akt and lipid metabolism in *Drosophila*. *J. Cell Sci.* *121*, 3383–3392.

Viswanathan, V.S., Ryan, M.J., Dhruv, H.D., Gill, S., Eichhoff, O.M., Seashore-Ludlow, B., Kaffenberger, S.D., Eaton, J.K., Shimada, K., Aguirre, A.J., et al. (2017). Dependency of a therapy-resistant state of cancer cells on a lipid peroxidase pathway. *Nature* *547*, 453–457.

Wang, T., Wei, J.J., Sabatini, D.M., and Lander, E.S. (2014). Genetic screens in human cells using the CRISPR-Cas9 system. *Science* *343*, 80–84.

Wang, T., Yu, H., Hughes, N.W., Liu, B., Kendirli, A., Klein, K., Chen, W.W., Lander, E.S., and Sabatini, D.M. (2017). Gene Essentiality Profiling Reveals Gene Networks and Synthetic Lethal Interactions with Oncogenic Ras. *Cell* *168*, 890–903.e15.

Watanabe, R., Inoue, N., Westfall, B., Taron, C.H., Orlean, P., Takeda, J., and Kinoshita, T. (1998). The first step of glycosylphosphatidylinositol biosynthesis is mediated by a complex of PIG-A, PIG-H, PIG-C and GPI1. *EMBO J.* *17*, 877–885.

Wei, Y., Reveal, B., Reich, J., Laursen, W.J., Senger, S., Akbar, T., Iida-Jones, T., Cai, W., Jarnik, M., and Lilly, M.A. (2014). TORC1 regulators Iml1/GATOR1 and GATOR2 control meiotic entry and oocyte development in *Drosophila*. *Proc. Natl. Acad. Sci.* *111*, E5670–E5677.

Wong, A.S.L., Choi, G.C.G., Cui, C.H., Pregernig, G., Milani, P., Adam, M., Perli, S.D., Kazer, S.W., Gaillard, A., Hermann, M., et al. (2016). Multiplexed barcoded CRISPR-Cas9 screening enabled by CombiGEM. *Proc. Natl. Acad. Sci. U. S. A.* *113*, 2544–2549.

Wu, Z.L., Zhang, L., Yabe, T., Kuberan, B., Beeler, D.L., Love, A., and Rosenberg, R.D. (2003). The involvement of heparan sulfate (HS) in FGF1/HS/FGFR1 signaling complex. *J. Biol. Chem.* *278*, 17121–17129.

## Figure Legends

**Figure 1. The coessentiality network.** (A) CRISPR and shRNA screens analyzed for this study. (B) Measuring functional enrichment. For each data set, pairwise correlations of knockout/knockdown fitness profiles were ranked, binned ( $n=1000$ ), and measured for enrichment for shared KEGG terms. Data from Meyers et al (“Avana data”) carries significantly more functional information than other data sets. (C) The Cancer Coessentiality Network, derived from Avana data, contains 3,483 genes connected by 68,813 edges. Selected modules, derived by an unbiased clustering algorithm and color-coded, demonstrate the functional coherence of the network.

**Figure 2. Cancer-specific features of the network.** (A) Clusters of genes were evaluated for tissue specificity (size of circles) and differential mRNA expression of genes in the cluster (color of circles). (B) Cluster 14 (BRAF cluster); nodes are genes in cluster and edges reflect strength of correlation of fitness profile. (C) Heatmap of essentiality profiles of genes in BRAF cluster, ranked by median essentiality score. Gene essentiality in the cluster is associated with PBRAF mutation ( $P<10^{-23}$ ) and sensitivity to BRAF inhibitor PLX-4720 ( $P<10^{-7}$ ). (D,E) Network and heatmap of MYB-related cluster. (F,G) PAX8-associated cluster. (H,I) B-catenin cluster. (J) MYCN neuroblastoma cluster is anticorrelated with MYC. (K) MYC, MYCN essentiality is mutually exclusive. (L) MDM2 cluster heatmap is associated with TP53 mutation status ( $P<10^{-13}$ ) and sensitivity to Nutlin-3a ( $P<10^{-14}$ ). (M) MDM2 vs TP53 essentiality. TP53 essentiality scores  $< -50$  indicate tumor suppressor role.

**Figure 3. Beyond cancer: characterization of the coessentiality network.** (A) Number of clusters (top) and total genes in clusters (bottom) showing strong association with annotated protein complexes, biological function, tissue specificity, amplification-induced CRISPR artifacts, and differential expression of genes. (B,C,D) Comparing the Avana coessentiality network to other functional (B), protein-protein (C), and coessentiality networks (D) shows the unique information contained in our network. (F) For some protein complexes, coessentiality is a better predictor of co-complex membership than co-

expression. (G) The coessentiality network is a powerful predictor of cancer pathways (Huang et al.) compared to other databases and networks (lower rank is better).

**Figure 4. GPX4 cluster.** (A) Glutathione peroxidase GPX4, a selenoprotein, is strongly clustered with genes involved in the selenocysteine conversion pathway (B). (C) Entire GPX4 cluster shows marked differential essentiality in glioblastoma cell lines. (D) Cellular requirement for GPX4 is associated with ZEB1 expression, as previously reported, but (E) GPX2 expression is more strongly predictive.

**Figure 5. Peroxisomal beta-oxidation.** (A) A large group of peroxisome-associated genes (orange) is connected by high-correlation edges in the network (blue). This cluster is connected by less stringent edges (Benjamini adjusted P-value < 0.01; gray edges) to other clusters containing sterol regulatory genes (green nodes) and the RAB18 GTPase (purple nodes). (B) The PEX cluster contains 12 genes, including 2 enzymes involved in fatty acid oxidation and 10 peroxisome biogenesis and maintenance genes. (C) The PEX cluster is emergently essential in a subset of lung cancer cell lines in the Avana data, and (D) in a subset of pancreatic cancer cell lines in the GeCKO data.

**Figure 6. mTORC pathway regulation.** (A) The mTORC1/2 complexes are regulated by the canonical TSC1/2 pathway, but amino acid sensing is done via the Ragulator complex at the lysosome. (B) Clusters in the coessentiality network represent components involved in mTORC regulation, and edges between clusters are consistent with information flow through the regulatory network. (Red edges indicate negative correlation).

**Figure 7. Glycosylation of cell surface receptors.** (A) Pathways involved in protein glycosylation and GPI anchor biosynthesis in the ER. (B) A network of clusters around glycosylation tracks the biogenesis and elongation of carbohydrate trees (DPM synthase, mannosyltransferases, glucosyltransferases) to their transfer to target proteins via N-linked glycosylation by the OST complex. Cell surface glycoproteins EGFR and IGF1R are both strongly linked to the OST complex despite their essentiality being mutually exclusive in cell lines.

Figure 1

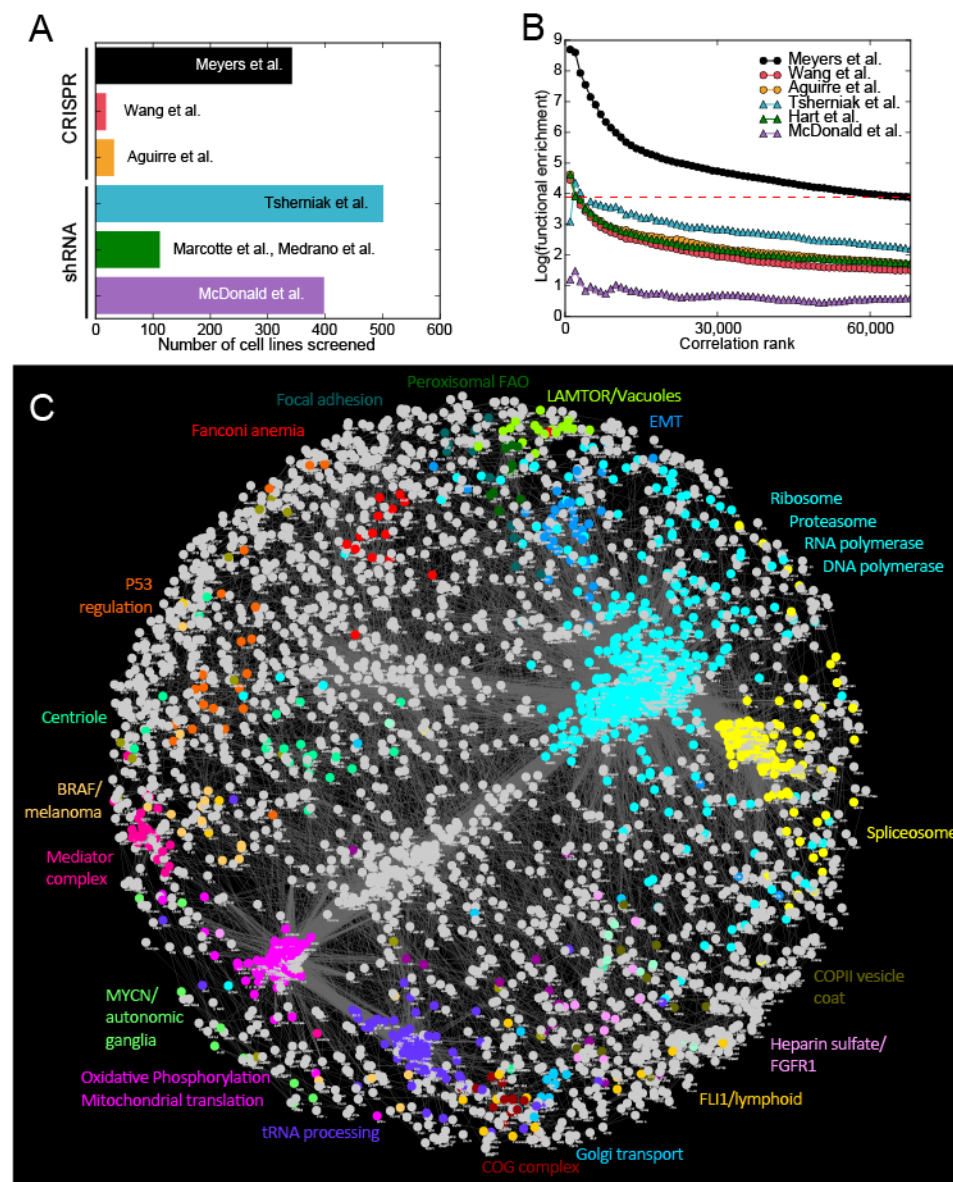




Figure 2

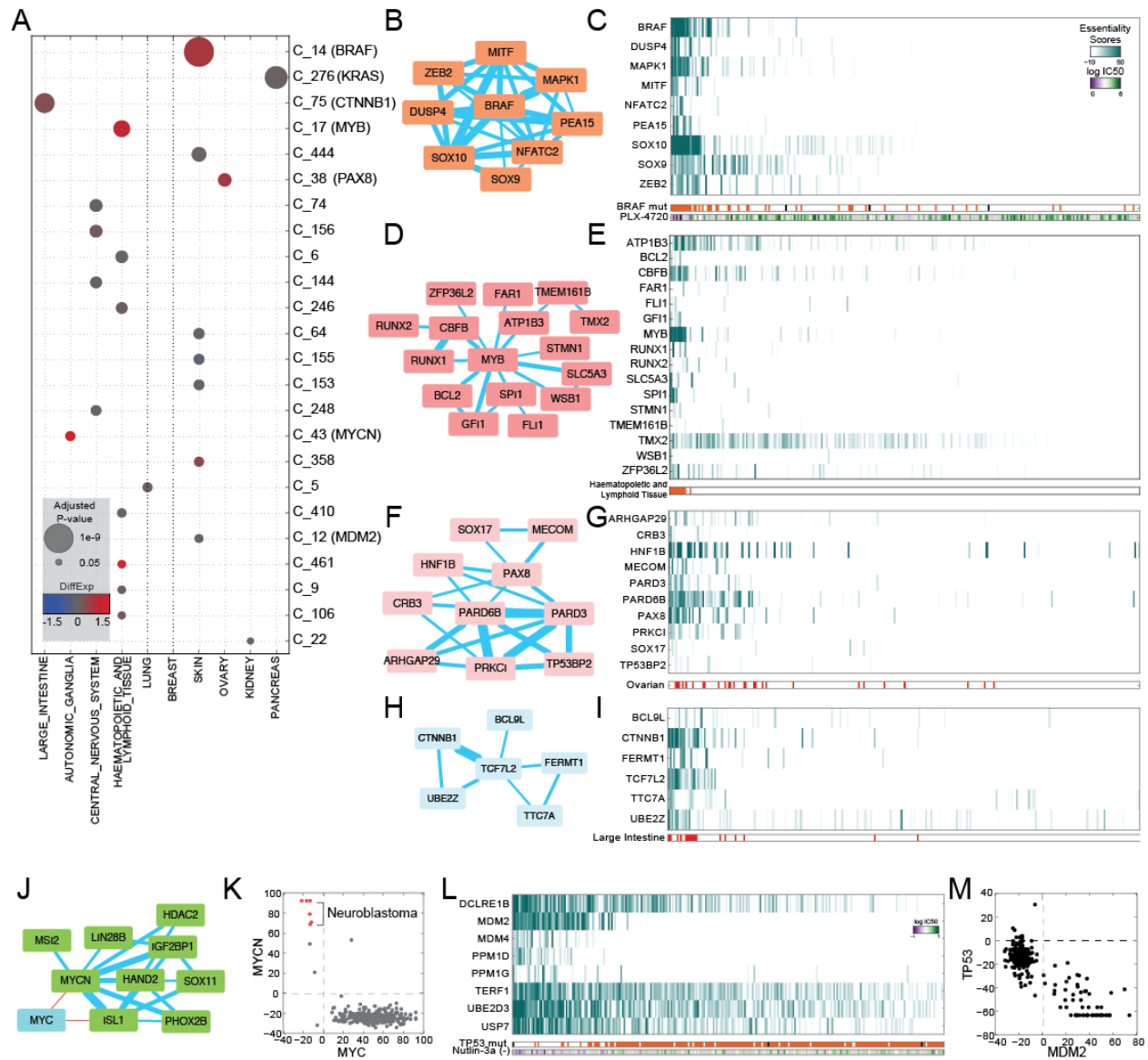


Figure 3

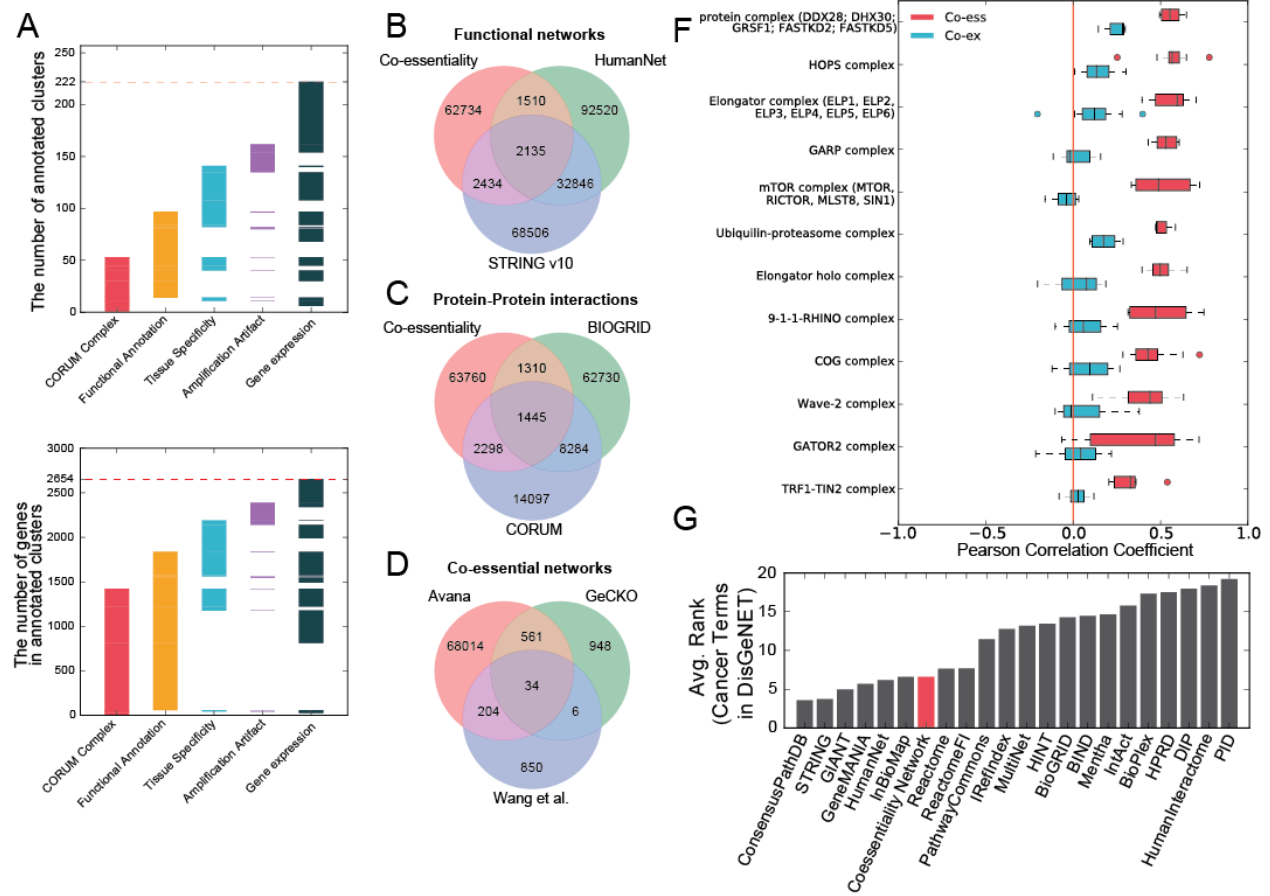


Figure 4

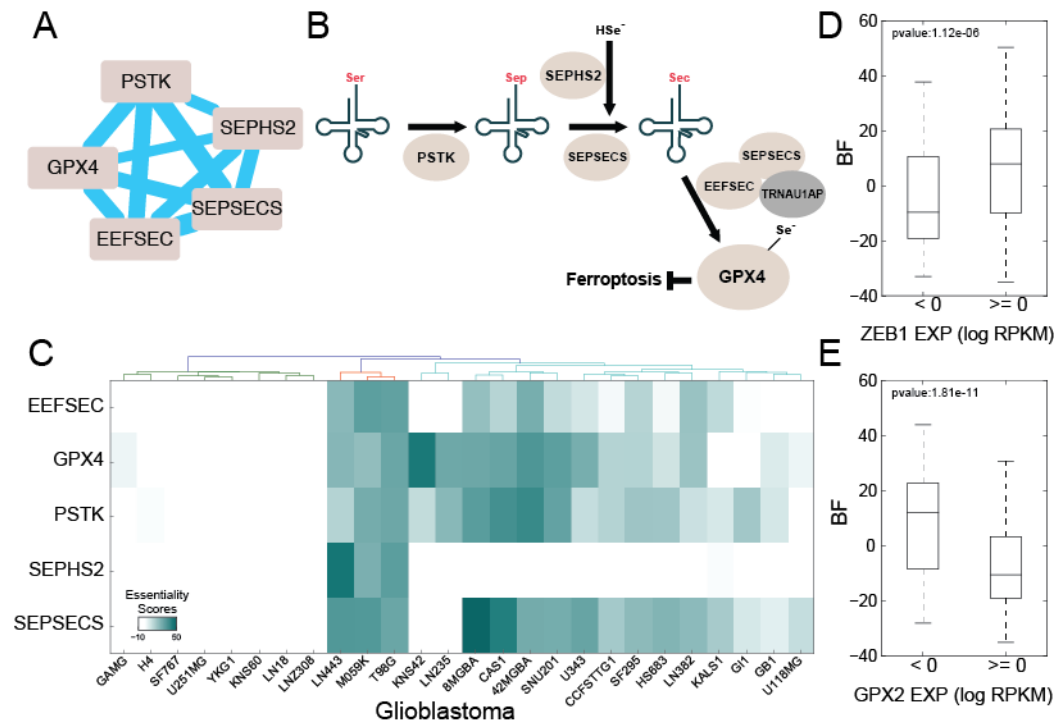


Figure 5

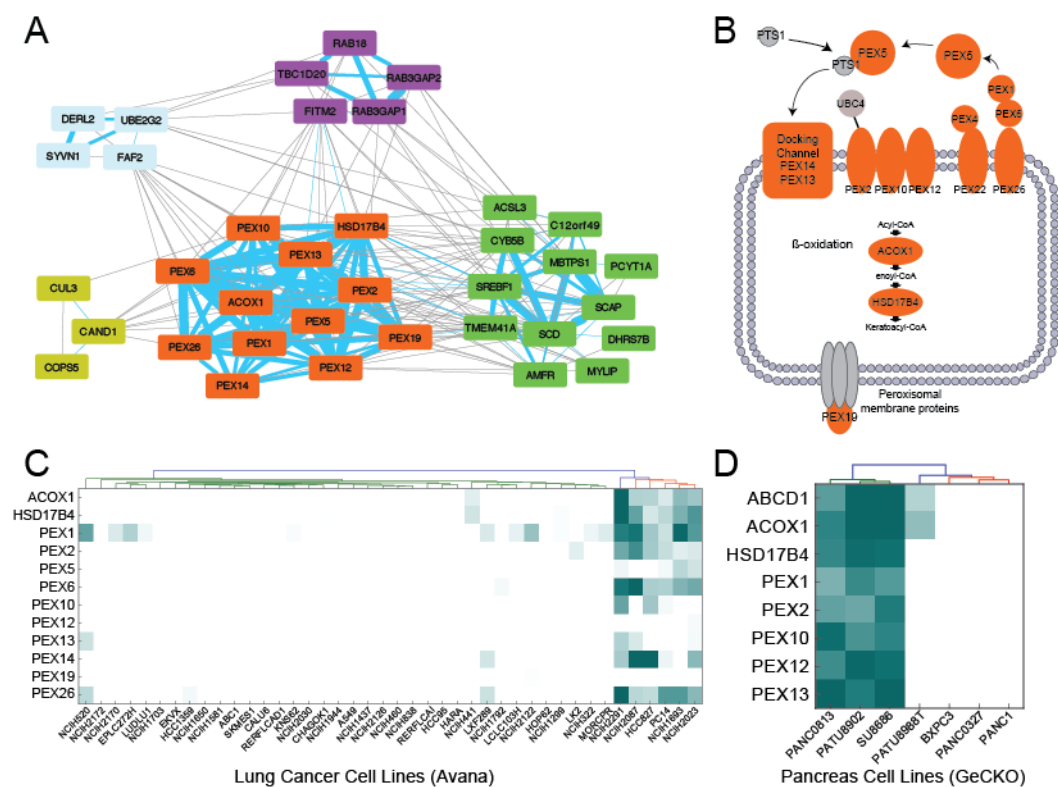


Figure 6

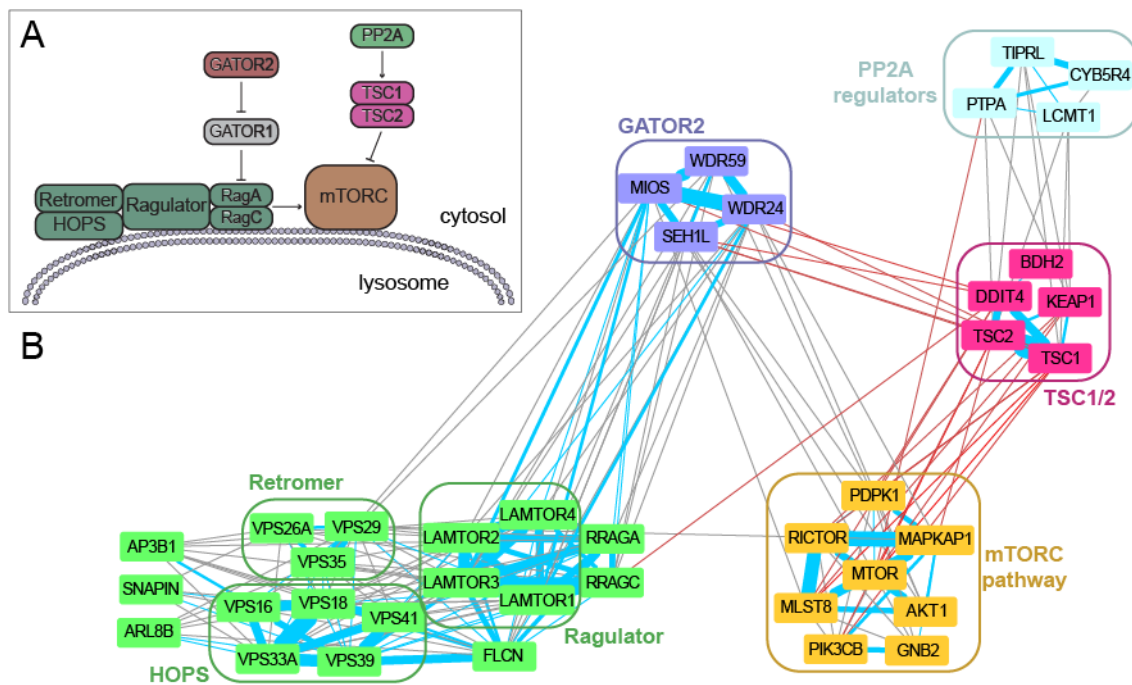


Figure 7

