HHS Public Access

Author manuscript

*Nat Genet*. Author manuscript; available in PMC 2022 January 17.

Published in final edited form as:

*Nat Genet*. 2021 May ; 53(5): 638–649. doi:10.1038/s41588-021-00840-z.

# A genome-wide atlas of co-essential modules assigns function to uncharacterized genes

**Michael Wainberg**[1,2,5], **Roarke A. Kamber**[1,5], **Akshay Balsubramani**[1,5], **Robin M. Meyers**[1], **Nasa Sinnott-Armstrong**[1], **Daniel Hornburg**[1], **Lihua Jiang**[1], **Joanne Chan**[1], **Ruiqi Jian**[1], **Mingxin Gu**[1], **Anna Shcherbina**[1], **Michael M. Dubreuil**[1], **Kaitlyn Spees**[1], **Wouter Meuleman**[3], **Michael P. Snyder**[1], **Michael C. Bassik**[1,4,✉], **Anshul Kundaje**[1,2,✉]

[1]Department of Genetics, Stanford University, Stanford, CA, USA

[2]Department of Computer Science, Stanford University, Stanford, CA, USA

[3]Altius Institute for Biomedical Sciences, Seattle, WA, USA

[4]Chemistry, Engineering, and Medicine for Human Health, Stanford University, Stanford, CA, USA

[5]These authors contributed equally: Michael Wainberg, Roarke A. Kamber, Akshay Balsubramani

## Abstract

A central question in the post-genomic era is how genes interact to form biological pathways. Measurements of gene dependency across hundreds of cell lines have been used to cluster genes into 'co-essential' pathways, but this approach has been limited by ubiquitous false positives. In the present study, we develop a statistical method that enables robust identification of gene co-essentiality and yields a genome-wide set of functional modules. This atlas recapitulates diverse pathways and protein complexes, and predicts the functions of 108 uncharacterized genes. Validating top predictions, we show that *TMEM189* encodes plasmanylethanolamine desaturase, a key enzyme for plasmalogen synthesis. We also show that *C15orf57* encodes a protein that binds



**Correspondence and requests for materials** should be addressed to M.C.B. or A.K. bassik@stanford.edu; akundaje@stanford.edu.

Author contributions

M.W., R.A.K., A.B., A.K. and M.C.B. contributed to conceptualizing the project. M.W. developed the method for identifying co-essential gene pairs using GLS and performed benchmarking analyses. R.A.K. annotated clusters according to biological pathways. M.W. and R.M.M. generated the co-essential modules using ClusterONE, with guidance from R.A.K. R.A.K. performed the experiments with help from M.G., K.S. and M.M.D. A.B. created the 2D visualization and the webtool with help from A.S. and W.M., and guidance from R.A.K. N.S.-A. contributed to analysis of tissue-selective module dependencies. L.J., J.C. and R.J. performed the proteomic analysis. D.H. performed the lipidomic analysis. M.W., R.A.K., A.B., A.K. and M.C.B. wrote the original draft. M.P.S., A.K. and M.C.B. supervised the work. All authors edited and reviewed the paper. A.K. and M.C.B. acquired the funding.

Online content

Any methods, additional references, Nature Research reporting summaries, source data, extended data, supplementary information, acknowledgements, peer review information; details of author contributions and competing interests; and statements of data and code availability are available at https://doi.org/10.1038/s41588-021-00840-z.

Code availability

Code to generate co-essential gene pairs, co-essential modules, modules with cancer-type-specific dependencies and the 2D layout is available at https://github.com/kundajelab/coessentiality.

Competing interests

The authors declare no competing interests.

Extended data is available for this paper at https://doi.org/10.1038/s41588-021-00840-z.

**Supplementary information** The online version contains supplementary material available at https://doi.org/10.1038/s41588-021-00840-z.

the AP2 complex, localizes to clathrin-coated pits and enables efficient transferrin uptake. Finally, we provide an interactive webtool for the community to explore our results, which establish co-essentiality profiling as a powerful resource for biological pathway identification and discovery of new gene functions.

---

A fundamental and still largely unresolved question in biology is how finite numbers of genes generate the vast phenotypic complexity of cells and organisms[1,2]. Modules of interacting genes represent a key layer of biological organization, and the complete identification of such functional modules and their constituent genes has emerged as a central goal of systems biology[3–6]. However, efforts to map genetic interactions and biological modules at the genome scale have been hindered by the enormous number of possible gene–gene interactions: assaying all pairs of genetic interactions among the ~20,000 human genes[7] would require 200 million distinct readouts. Furthermore, despite substantial progress in elucidating the functions of individual genes in recent decades through both targeted studies and unbiased approaches[8–11], hundreds of human genes remain functionally uncharacterized.

Pioneering high-throughput work in yeast measured pairwise genetic interactions by quantifying the fitness of double-knockout strains[12,13]; this work has since been extended into a genome-wide map of yeast genetic interactions and modules[4,14]. In human cells, which unlike yeast cells cannot be crossed to generate double-knockout mutants, a key advance toward genetic interaction mapping has been the development of genome-scale clustered regularly interspaced short palindromic repeats (CRISPR)–Cas9 and RNA interference screens[10,11] and their repurposing to perform pairs of perturbations[5,15–20]. Yet despite considerable successes, double-perturbation genetic interaction mapping is inherently limited by the combinatorial explosion of gene pairs: the largest human genetic interaction map to date[5] assayed only 222,784 gene pairs, ~0.1% of all possible genetic interactions.

A complementary approach that circumvents this limitation is co-essentiality mapping, which measures the fitness of single-gene perturbations across multiple conditions, and maps putative functional interactions by correlating the resulting phenotypic profiles (Extended Data Fig. 1a). Both co-essentiality and genetic interaction mapping measure gene essentiality across many different genetic backgrounds, but, whereas the background for genetic interaction mapping is the knockout of a single partner gene, for co-essentiality mapping it is the entire set of genetic and phenotypic characteristics of a given cell line. Furthermore, the two have distinct meanings: genetic interaction mapping yields cell-type-specific functional interactions, whereas co-essentiality yields pan-cell-type interactions. Co-essentiality mapping across diverse cancer cell lines has recently been used to group genes into pathways and in some cases has identified new gene functions[21–26].

Co-essentiality mapping has, however, its own fundamental limitation: unlike double-perturbation mapping, where each pair of gene knockouts is independent, measurements in two different cell lines may be strongly related, for example, because some pairs of cell lines are derived from the same tissue or lineage. A secondary problem is systemic variation in single guide (sg)RNA toxicity or sensitivity to perturbation across cell lines,

although this problem has largely been addressed[27]. Existing approaches fail to account for violations of independence, leading to inflated *P* values, incorrect determinations of statistical significance and an inability to identify gene co-essentiality relationships in a robust, systematic manner (Extended Data Fig. 1b). In the present study, we address this limitation with a statistical method that explicitly accounts for cell-line nonindependence. We apply the method to genome-wide CRISPR screens across 485 diverse cancer cell lines[28] and find substantially improved enrichment for known pathway interactions and protein complexes.

These analytical advances greatly improve our ability to detect bona fide functional modules. We generate a genome-wide atlas of co-essential modules, which both recapitulates diverse known pathways and protein complexes and nominates putative functions for 108 poorly characterized genes. We experimentally validate two such genes: we identify *TMEM189* as the long-sought gene encoding the plasmanylethanolamine desaturase (PEDS) orphan enzyme required for synthesis of plasmalogen lipids, one of the most abundant lipid classes in the human body, and we discover a role for *C15orf57* in regulating clathrin-mediated endocytosis. Finally, to accelerate further biological discovery, we present an interactive webtool to visualize and analyze co-essential gene pairs and modules.

## Results

### A genome-wide map of co-essential interactions.

To map co-essential interactions across genome-wide screens while accounting for cell-line nonindependence, we devised an approach based on generalized least squares (GLS), a classic statistical technique[29] (Methods). We applied the approach (Fig. 1a) to a dataset of CRISPR screens in 485 cell lines from the Achilles project[28], with gene-level essentiality scores corrected for copy number and guide efficacy using the CERES algorithm[27]. Under the assumption that only a small fraction of gene pairs is expected to functionally interact or participate in the same biological pathway or process[5], the median *P* value across gene pairs ought to be very close to 0.5 for a well-calibrated method. Indeed, we found that the median GLS *P* value was 0.48, indicating almost perfect statistical calibration, whereas the median Pearson's correlation *P* value on the same dataset was 0.21, indicating substantial *P*-value inflation[30] and false-positive co-essential gene pairs (Fig. 1b); this contrast is also apparent from a quantile–quantile plot of the GLS and Pearson's correlation *P* values (Extended Data Fig. 2). We provide each gene's significant co-essential interactors at a false discovery rate (FDR) of 10% (Supplementary Data 1).

Even while correcting for *P*-value inflation, GLS still has substantial power to detect co-essential interactions. Approximately 80% of genes have at least one co-essential partner at 10% FDR (Extended Data Fig. 3), and ~40% of genes have at least ten partners. In all, we detect 93,575 significant co-essential gene pairs, of which 99.4% are positively correlated and 0.6% negatively correlated. We noted that, in many cases, negative correlations occur when one gene negatively regulates the other: for example, *TP53* negatively correlates with *MDM2* ($P = 1 \times 10^{-12}$), which encodes a protein that ubiquitinates p53 to mark it for degradation[31]; *HER2* negatively correlates with *PHLDA2* ($P=5 \times 10^{-6}$), which was recently

shown to inhibit *HER2* signaling[32]; and *MAPK1* negatively correlates with *DUSP6* ($P$=2 × $10^{-6}$), encoding a phosphatase that inactivates several mitogen-activated protein (MAP) kinases including MAPK1 (ref. [33]). A second class of negative correlation arises from genes with similar functions but in mutually exclusive cell types, such as *MYC* and *MYCN* ($P$=3 × $10^{-11}$)[34].

A common finding from CRISPR–Cas9 viability screens is that most genes do not affect cell viability on knockout, which intuitively should result in a less informative essentiality signal across lines. However, although more essential genes (defined by average essentiality across lines) tend to have more partners, 70% of the 10% least essential genes have at least one partner at 10% FDR, and nearly half of these least essential genes have at least one partner at 1% FDR (Extended Data Fig. 3). Thus, rather than being limited to detecting interactions among only strongly essential genes, co-essentiality can be a genome-wide tool for pathway mapping.

We developed a method to visualize the co-essentiality network by placing more strongly co-essential gene pairs closer together, inspired by similar visualizations of yeast genetic interaction maps[4,14]. We found that naive application of dimensionality reduction techniques like principal component analysis (PCA) and Uniform Manifold Approximation and Projection (UMAP)[35] had difficulty modeling the multi-scale nature of the co-essentiality network; previous co-essentiality network visualizations (for example, McDonald et al.[23]) also lack discernible structure. Instead, we first applied diffusion maps[36], a technique from spectral graph theory, to separate coarse- and fine-scale components before applying UMAP (Methods). To further improve the layout, we incorporated module membership (defined below) into the diffusion map in addition to pairwise co-essentiality. To showcase the power of this approach, we manually annotated 39 ‘neighborhoods’ within the interaction map highly enriched for a particular pathway or complex (Fig. 1c,d); collectively, these pathways and complexes encompass many of the major aspects of cell biology.

## Co-essentiality complements co-expression in mapping biological pathways.

We next investigated whether GLS’s improved calibration translated into improved co-functional gene partner prioritization. Using an established benchmarking strategy[24], we measured how accurately GLS could recall each gene’s top 1–10 interaction partners compared with Pearson’s correlation. The ability of GLS and Pearson’s correlation to recall known interactions was measured using three distinct types of databases of co-functional genes previously benchmarked in Pan et al.[24]: CORUM, a manually curated protein complex database[37]; hu.MAP, a mass spectrometry (MS) protein–protein interaction database[38]; and STRING, a gene–gene interaction database integrating multiple sources of direct and indirect evidence[39]—experimental evidence, other pathway/complex databases, co-expression, literature text mining, genomic colocalization across species, co-occurrence across species and the existence of a gene–gene fusion in any species. As several of these lines of evidence do not relate directly to gene function or might introduce circularity into our analyses, we restricted our analysis to gene pairs supported by experimental evidence.

We found that GLS consistently prioritized genes more effectively than previous co-essentiality detection methods, including Pearson’s correlation[21,24,26] and Pearson’s

correlation bias corrected with PCA using olfactory receptor genes as a gold-standard negative set[25], across all three databases and a wide variety of rank thresholds (Fig. 2a and Extended Data Fig. 4). For example, the top-ranked partners for each gene are approximately 160-fold enriched for CORUM interactions for GLS compared with 120-fold for bias-corrected Pearson's correlation, for hu.MAP, 130-fold versus 90-fold enriched, and, for STRING, 7.5-fold versus 5.5-fold enriched. Remarkably, failing to perform PCA-based bias correction substantially degrades the performance of Pearson's correlation but not GLS, suggesting that GLS automatically performs bias correction without requiring a putatively nonessential gene set like olfactory receptors.

We also compared co-essentiality with co-expression, a complementary approach to assessing co-functionality, using the COXPRESdb database[40]. We observed that co-essentiality substantially outperformed co-expression at recalling protein complexes and physical interactions from CORUM and hu.MAP, but performance was more equivocal for STRING, with co-essentiality outperforming co-expression only for top-ranked partner genes. To further explore the relative merits of co-expression and co-essentiality, we also benchmarked on DoRothEA, a transcription factor-regulon database[39,41], and found that co-expression substantially outperformed co-essentiality (Fig. 2a). We conducted an analysis of key cancer drivers, and found that co-essentiality outperformed co-expression in detecting interactions between known oncogenes and tumor suppressors (Supplementary Table 1 and Supplementary Note). Collectively, these results suggest that co-essentiality and co-expression may have complementary roles in biological pathway mapping, with co-essentiality better suited to protein complexes and direct physical interactions and co-expression better suited to transcriptional regulatory relationships.

## Co-essential modules recapitulate known pathways and nominate new members.

To group genes into modules based on their GLS co-essentiality profiles, we used ClusterONE[42], a commonly used algorithm originally developed for de novo discovery of protein complexes from protein–protein interaction data (Methods). Crucially, ClusterONE allows overlapping modules, enabling pleiotropic genes to be constituents of multiple modules. One major parameter affecting ClusterONE module detection quality is the module density *d*, which determines how strong the internal connections within a module must be relative to the connections on the edge of the module between members and nonmembers. By applying ClusterONE with a range of values of *d* (Extended Data Fig. 5 and Supplementary Note), we generated a total of 5,218 modules of various sizes.

These co-essential modules, containing between 4 and 741 genes, correspond to a wide range of biological pathways (Supplementary Data 2). To estimate the fraction of the genome our modules assign a putative function, we counted how many genes were present in modules that were highly (at least 100-fold) and significantly (Bonferroni's corrected $P < 0.05$) enriched for some gene ontology (GO) term. We excluded syntenic modules (that is, with all genes on the same chromosome), because, although many probably represent bona fide co-functional gene sets, others may be confounded by residual copy number artifacts or other factors. Indeed, we found that syntenic co-essential gene pairs are generally less enriched for known functional relationships (Methods and Extended Data Fig. 6). By this

metric, our co-essential modules assign putative functions to 9,891 genes, a much larger fraction of the genome compared with previous approaches used to cluster genes based on co-essentiality profiles (Fig. 2b and Extended Data Fig. 7). Additional benchmarking strategies revealed that our co-essential modules substantially improved predictions of gene co-functionality compared with predictions based on Pearson's correlations between individual gene pairs (Fig. 2c, Extended Data Fig. 8 and Supplementary Note).

Among the 1,269 modules with >100-fold enrichments are modules highly enriched for genes involved in growth regulation (Fig. 3a,b), autophagy (Fig. 3c), cell–cell signaling (Fig. 3d), the DNA-damage response (Fig. 3e), innate immunity (Fig. 3f), glycolysis (Fig. 3g), transcriptional regulation (Fig. 3h,i), the cell cycle (Fig. 3j) and mitochondrial respiration (Fig. 3k), among many others (Supplementary Data 2).

Several important features of the co-essential modules are highlighted in Fig. 3. First, ClusterONE's ability to include genes in multiple modules enabled identification of pleiotropic gene functions, as illustrated by the identification of two modules containing *MTOR* that closely correspond to the two mTOR (mechanistic target of rapamycin)-containing complexes, mTORC1 (Fig. 3a) and mTORC2 (Fig. 3b)[43]. Second, co-essential modules are not limited to physical complexes, as illustrated by the almost complete identification of the glycolysis pathway (Fig. 3g), or even to cell-autonomous pathways, as illustrated by the identification of the jagged-notch intercellular signaling pathway (Fig. 3d). Third, by examining modules identified at different values of $d$, we could detect multiple scales of biological organization, as illustrated by multiple modules corresponding to mitochondrial respiration (Fig. 3$k_{i–v}$). Module no. 256, a 164-member module identified at $d = 0.2$, includes most nuclear-encoded subunits of the four respiratory chain complexes required for mitochondrial ATP synthesis, as well as numerous mitochondrial transfer (t)RNA synthases, elongation factors, and components of the mitoribosome required for synthesis of the mitochondrial subunits of the mitochondrial respiratory complexes (Fig. 3k). Several modules identified with $d = 0.9$, by contrast, correspond to smaller units of functional organization, such as module no. 4,250, a 13-member module containing 12 subunits of the ATP synthase complex (Fig. 3$k_{iv}$ and Supplementary Data 2), and module no. 2,072, a 99-member module comprising 61 subunits of the mitochondrial ribosome and many of its associated factors (Fig. 3$k_{v}$ and Supplementary Data 2). Fourth, although several modules are almost complete representations of a biological pathway, such as module no. 520, which exclusively comprises many of the genes identified in recent targeted screens for autophagy regulators[44] (Fig. 3c), many modules highly enriched for a particular pathway also contain one or more uncharacterized genes (red boxes, Fig. 3e,f,h–k). The set of modules containing members of the endoplasmic reticulum membrane complex provides an additional example of the ability of co-essential modules to capture multiple levels of biological organization (Supplementary Note).

## Using co-essential modules to systematically predict the functions of uncharacterized genes.

Co-essential modules are often highly enriched for functionally related genes, and thus enable unbiased, genome-wide prediction of uncharacterized gene function. This has

recently been used to assign uncharacterized genes to pathways based on Pearson's correlations with known pathway members[21,24,26]. However, it has remained unclear how broadly useful co-essentiality information is in predicting the functions of the hundreds of genes that remain uncharacterized, which probably span diverse biological processes.

To generate functional predictions for uncharacterized genes using our modules, we first enumerated 1,321 uncharacterized genes from the UniProt database with a UniProt annotation score (a heuristic measure of protein annotation content) of ≤2. We then restricted to genes in modules at least 100-fold enriched for one or more GO terms, excluding terms with <5 genes.

The 108 uncharacterized genes assigned putative functions by this method are included, on average, in ~2 co-essential modules, yielding a list of 232 functional predictions (Supplementary Data 3), excluding those in syntenic modules. Each functional prediction consists of an uncharacterized gene paired with a candidate module. Notably, several of these predictions are consistent with recent experiments not yet incorporated into the UniProt database, including *C16orf59* in centriole function[45] and *PTAR1* in Golgi body function[46], as well as with several of the results of recent yeast and human genetic interaction mapping approaches (Supplementary Note). To prioritize functional predictions for experimental validation, we ranked candidate modules by their maximal enrichment for a given GO term, because these predictions yield the most readily testable predictions. The top uncharacterized gene predictions (ranked by GO term enrichment) span a wide range of biological processes, including mitochondrial respiration, transcription, DNA replication, Golgi body function, lipid synthesis and endocytosis (Supplementary Data 3).

### *TMEM189* encodes the enzyme PEDS required for plasmalogen synthesis.

We selected two genes, *TMEM189* (ranked no. 3) and *C15orf57* (ranked no. 18), for experimental validation. *TMEM189*, also known as *KUA*, encodes a transmembrane protein of 270 amino acids, the function of which was largely unexplored before our work. The top-ranked co-essential module containing *TMEM189*, module no. 2,213, is highly enriched for genes required for synthesis of ether lipids (Fig. 4a), a broad class of structural and signaling lipids involved in regulation of membrane fluidity and sensitivity to oxidative stress, and which constitute ~20% of phospholipids in human cells[47]. We noted that genes in this module were particularly essential in cell lines derived from hematological cancers (Fig. 4b). Although several module genes (for example, *AGPS, FAR1* and *GNPAT*) are specifically involved in ether lipid synthesis, others (for example, *PCYT2* and *EPT1*) are also required to synthesize other ethanolamine-containing phospholipids[48,49]. Based on this prediction, we hypothesized that *TMEM189* was involved in lipid biosynthesis, particularly of ether lipids.

To interrogate *TMEM189*'s functional role in lipid biosynthesis in an unbiased manner, we extended a targeted lipidomic method[50,51] to measure the absolute concentrations of several hundred lipid species. We compared lipid concentrations in cell extracts derived from HeLa–Cas9 cells stably expressing sgRNAs targeting either *TMEM189* or a control genomic locus; most lipid species were similarly abundant. Strikingly, however, cells expressing *TMEM189*-targeting sgRNAs contained dramatically lower concentrations of

the set of 37 lipid species belonging to the ether lipid subclass plasmenylethanolamines (Fig. 4c,d and Supplementary Data 4–7), also known as ethanolamine plasmalogens. At the same time, *TMEM189* knockout cells had elevated levels of the set of 30 lipid species belonging to the ether lipid subclass plasmanylethanolamines (Fig. 4c,e), which differ from plasmenylethanolamines by a single double bond in the *sn*-1 chain (part of the plasmalogen-defining vinyl ether bond). Plasmanylethanolamines and plasmenylethanolamines form a known precursor–product relationship, with plasmanylethanolamines rapidly converted into plasmenylethanolamines in the endoplasmic reticulum by the orphan enzyme PEDS, first reported in mammalian cell extracts >40 years ago[52].

The accumulation of the precursors, and loss of the product, of the reaction catalyzed by PEDS in cells expressing *TMEM189*-targeting sgRNAs strongly implicates *TMEM189* as the gene responsible for orphan PEDS activity. Two orthogonal lines of evidence strongly support this conclusion. First, we examined a cell line, RAW.12, that was evolved to lack plasmalogens and shown to exhibit a specific defect in PEDS activity[53]. By western blotting for TMEM189 in cell extracts prepared from RAW.12 cells or its parent, unmutated cell line RAW264.7, we confirmed that TMEM189 levels were decreased in PEDS-deficient RAW.12 cell extracts (Fig. 4f). Second, *TMEM189* contains a histidine-rich domain conserved in most lipid desaturase enzymes, and is distantly related to the fatty acid desaturase *FAD4* in *Arabidopsis* sp.[54], which introduces an unusual double bond in the *sn*-2 fatty acid[54]. Taken together, our results provide strong evidence for a primary role for *TMEM189* as the orphan desaturase required for the final step of plasmalogen biosynthesis, although we cannot exclude the possibility that *TMEM189* has additional functions (Extended Data Fig. 9 and Supplementary Note). Overall, these findings provide a striking example of the power of co-essential modules to predict gene function.

### *C15orf57* is a regulator of clathrin-mediated endocytosis.

*C15orf57* (also known as coiled-coil domain containing 32 (*CCDC32*)) encodes a 185-residue protein with, to our knowledge, no annotated function. *C15orf57* is present in several overlapping co-essential modules (Supplementary Data 2), including a module (no. 2,067) highly enriched for genes required for clathrin-mediated endocytosis, in particular subunits of the adapter protein 2 (AP2) complex (Fig. 5a,b). One of AP2's best-described functions is mediating endocytosis of transferrin bound to the transferrin receptor[55]; thus, we hypothesized that *C15orf57* might be required for cellular transferrin uptake. To test this, we monitored uptake of transferrin that was labeled with a pH-sensitive fluorescent dye, pHrodo, by HeLa–Cas9 cells expressing sgRNAs targeting *C15orf57*, the transferrin receptor (*TFRC*) or a control locus. Cells expressing sgRNAs targeting either *C15orf57* or *TRFC* exhibited reduced transferrin uptake compared with cells expressing control sgRNAs, consistent with a role for *C15orf57* in transferrin uptake (Fig. 5c).

To gain further insight into the mechanism by which *C15orf57* functions in clathrin-mediated endocytosis, we immunoprecipitated C15orf57–GFP (green fluorescent protein) complexes and analyzed them by MS. C15orf57–GFP immunoprecipitates (IPs) were strongly enriched for all five members of the AP2 clathrin adapter complex: AP2S1, AP2A1, AP2A2, AP2M1 and AP2B1 (Fig. 5d and Supplementary Data 8). These findings

were consistent with the identification of C15orf57 as a physical interactor of multiple AP2 subunits by MS in the BioPlex[56,57] dataset. In reciprocal co-immunoprecipitation experiments, we confirmed that C15orf57–GFP physically interacts with AP2S1–mCherry (Fig. 5e). We additionally confirmed through confocal microscopy that C15orf57–GFP colocalizes with AP2S1–mCherry in small puncta at the cell surface, which probably correspond to clathrin-coated pits, the sites of clathrin-mediated endocytosis (Fig. 5f). The identification of the members of the AP2 complex as physical interactors of C15orf57, and their colocalization in cells, suggests that C15orf57 may regulate clathrin-mediated endocytosis of transferrin (and possibly other cargoes) by directly modulating AP2 function.

## Identification of cancer-type-specific pathway dependencies.

A major motivation for high-throughput cancer cell-line screening efforts, such as the Achilles project, is the possibility of identifying therapeutically targetable cancer-type-specific vulnerabilities[21,28]. These efforts have shown promise in identifying individual genes selectively essential in specific cancer types[21,58]. Some cancers even harbor selective dependencies on entire gene pathways. We asked whether our co-essential modules could identify such cancer-type-specific pathway dependencies.

To systematically identify differentially essential modules across tissue types, we first calculated *P* values for each gene using GLS and then aggregated them across the genes in each module (Methods). We identified 444 modules that are differentially essential at 10% FDR in cancers derived from 16 distinct tissue types (Fig. 6a and Supplementary Data 9).

Several of the differentially essential modules correspond to canonical tissue-specific cancer drivers, demonstrating the power of this approach to uncover bona fide selective pathway dependencies. For example, the most significantly breast cancer–specific module dependency contains *ESR1*, the estrogen receptor (ER), which is overexpressed in >70% of breast cancers and enables hormone-dependent growth[59]. This module (and its neighborhood in the two-dimensional (2D) network representation; Fig. 6b) also contains several genes that functionally interact with *ESR1*, including: *SPDEF, FOXA1* and *GATA3*, three master transcriptional regulators in ER-dependent breast cancer[60]; retinoic acid receptor α (*RARA*), a target of *ESR1*-dependent transcriptional activity[61]; and *TOB1*, a gene required for estrogen-independent growth of ER-positive breast cancers[62].

As a second example, the neighborhood of the most differentially essential module in skin cancer (Fig. 6c) includes several components of the *BRAF–MAPK* pathway, consistent with *BRAF* being mutated in ~50% of melanomas[63], as well as *MITF*, a melanoma-specific oncogene[64] activated downstream of *BRAF*. Module members *NFATC2, SOX9* and *SOX10* have well-established roles in melanoma[65,66]. In both examples, our differentially essential modules contain sets of lineage-specific cancer drivers that are known to functionally interact. The additional 442 modules we identify as selectively essential in 16 cancer types (Supplementary Data 9) represent a resource for identifying new pathway targets in specific cancer types.

## Discussion

Building a global map of biological pathways in human cells and assigning function to the thousands of poorly characterized genes remain key challenges in cell biology. In this work, we demonstrate that mapping co-essentiality across diverse cancer cell lines enables substantial progress toward both objectives. To facilitate the use of this resource, we developed an interactive webtool, http://coessentiality.net, that enables exploration of the co-essentiality network (Extended Data Fig. 10, Supplementary Note and Supplementary Video 1).

The co-essential network developed in the present study represents a comprehensive, and statistically robust, genome-wide perturbational pathway map of human cells. Unlike double-perturbation approaches, ours can be applied genome wide; unlike prior co-essentiality methods, it is statistically well calibrated despite the lack of independence among the screens from which it was derived. The gene–gene relationships evidenced by these different datasets may be complementary, with co-essentiality better at detecting protein complexes and co-expression better at detecting transcriptional regulatory relationships (Fig. 2a). Our global interaction map and webtool showcase the high resolution and versatility of co-essentiality for new pathway mapping.

Our validations of the role of *TMEM189* in plasmalogen biosynthesis and *C15orf57* in clathrin-mediated endocytosis highlight the utility of hypothesis generation from co-essential modules. Of note, during preparation of this manuscript, an entirely orthogonal approach found that the bacterial *TMEM189* homolog CarF was responsible for PEDS activity in bacterial cells, and this activity was shown to be conserved in human cells[67]. A second group also confirmed that *TMEM189* encodes PEDS in human cells[68]. These complementary validations will potentiate dissection of plasmalogen lipid function. The specific function of the plasmalogen-defining vinyl ether bond, proposed to be critical for antioxidant and oxygen-sensing activity, has remained difficult to assess experimentally. With the identity of PEDS now in hand, these and other basic questions about plasmalogen function can be addressed. Furthermore, plasmalogens are highly upregulated in many cancers, and inhibitors of this pathway have been explored as anti-cancer agents[69]; *TMEM189* represents a potential therapeutically targetable node in this pathway.

Our identification of *C15orf57* as a regulator of clathrin-mediated endocytosis adds another key player to this pathway; further work is required to uncover its precise mechanistic function. None the less, our discovery that *C15orf57* binds the AP2 complex and regulates endocytosis may advance understanding of the significance of recurrent *C15orf57–CBX3* gene fusions in hepatocellular carcinoma[70].

Several additional functional predictions generated by our method are supported by evidence from other unbiased, high-throughput approaches. For example, *C7orf26*, which we predict is involved in the function of the integrator complex (required for small noncoding RNA transcription[71]), was observed to interact with several subunits of the integrator complex in IP–MS experiments[72]; its expression is also highly correlated with several integrator subunits[39,40]. As a second example, the functionally uncharacterized gene *TMEM242*, for

which we predict a function in mitochondrial respiration, was reported to interact with the gene product of *NDUFA3*, a subunit of mitochondrial complex I (ref. [73]). More broadly, the co-essential modules we report may be used to predict not only functions of uncharacterized genes but also new functions for partially characterized genes (Supplementary Note).

Overall, our experimental validation of two uncharacterized genes and functional predictions for 106 additional uncharacterized genes constitutes an immediately useful resource for the broader cell biology community. A key future direction in expanding the capabilities of this resource to detect functional genetic relationships is to measure phenotypes beyond cancer cell-line growth under standard conditions. Although the Achilles project plans to ultimately screen several thousand cell lines, our subsampling analysis suggests that some aspects of performance have already started to saturate at 485 lines (Supplementary Fig. 1). Nonetheless, our approach may benefit greatly from screens performed in primary tissues[74], across individuals, under non-ambient conditions, such as in the presence of a drug or cellular stress, or with readouts besides cellular fitness, such as cell morphology, gene expression or cellular activity. Such screens may uncover an even broader spectrum of functional interactions and could enable a dynamic map of pathway rewiring across conditions. Overall, our genome-wide mapping of the human co-essential network comprises a powerful resource for biological hypothesis generation and discovery.

## Methods

### Dataset.

The dataset used to determine co-essential interactions consists of the 485 genome-wide CRISPR screens from the Achilles project 18Q3 release[28]. Specifically, 17,634 genes were screened in 485 cell lines from 27 distinct lineages using the Avana CRISPR library[78], and gene-level effects were quantified using the CERES algorithm to account for variability in guide effectiveness and copy number across lines[27], resulting in a 17,634 × 485 matrix of normalized gene-level effects. Intuitively, gene-level effects represent the number of times fewer cells with the knockout doubled during the screen, compared with control cells. This dataset is publicly available at https://ndownloader.figshare.com/files/12704099 or https://depmap.org/portal/download/all under release ‘DepMap Public 18Q3’ and file ‘gene_effect.csv’.

### Bias correction.

Bias correction was applied as described in Boyle et al.[25]. Specifically, the first four principal components of the gene-by-cell-line essentiality matrix across all olfactory receptor genes, defined in the present study as those with the ‘olfactory receptor activity’ GO term[79,80], were subtracted from the original CERES score matrix, resulting in a new bias-corrected matrix. To avoid multicollinearity and allow inversion of the covariance matrix for GLS (see below), subtraction of the first 4 principal components was followed by removal of 4 cell lines (arbitrarily chosen to be the last 4), resulting in a 17,634 × 481 matrix of bias-corrected CERES scores.

### Statistics.

Data are generally plotted as mean ± s.d. unless otherwise indicated. No statistical methods were used to pre-compute sample size. Statistical significance was determined using two-tailed Student's *t*-tests performed using Microsoft Excel 2016 or GraphPad Prism (v.9) software unless otherwise indicated.

**Quantifying co-essential gene pairs.—**The co-essentiality between each pair of genes was quantified using GLS[29]. In a departure from previous approaches to co-essentiality profiling, GLS automatically and flexibly accounts for the nonindependence of cell lines by incorporating information about the covariation between every pair of screens. When all screens are independent and have the same variance in effect sizes across genes, the GLS effect size becomes exactly equivalent to Pearson's correlation coefficient. GLS is closely related to the linear mixed models used for population structure correction in genome-wide association studies[81], an analogous problem to ours.

Specifically, GLS estimates the vector of parameters $\boldsymbol{\beta}$ of the linear regression model $\mathbf{Y} = X\boldsymbol{\beta} + \boldsymbol{\varepsilon}$, where $\mathbf{Y}$ is a vector of observations, $X$ is a matrix of features corresponding to those observations and $\boldsymbol{\varepsilon}$ is the error or residual, under the assumption that the mean of the errors is 0 and their variance is $\Sigma$, where $\Sigma$ is a covariance matrix specified by the practitioner. The only difference from ordinary least squares (OLS) is the value of $\Sigma$; OLS assumes that it is the identity matrix, whereas GLS allows it to be any user-specified value. In the present study, we set $\Sigma$ to be the covariance matrix of the data themselves, that is, $\Sigma_{i,j}$ is the covariance of cell lines $i$ and $j$ across all genes in the CRISPR screen.

In practice, GLS is solved by: (1) inverting $\Sigma$, in our implementation (statsmodels.regression.linear_model.GLS from the *statsmodels* Python package), by using the Moore–Penrose pseudoinverse instead of the true inverse as a computational optimization; (2) taking Cholesky's decomposition of this inverse covariance matrix, $\text{chol}(\Sigma^{-1})$; (3) transforming both $Y$ and $X$ by $\text{chol}(\Sigma^{-1})$ to obtain the transformed observations $\mathbf{Y}' = \text{chol}(\Sigma^{-1})\mathbf{Y}$ and transformed features $\mathbf{X}' = \text{chol}(\Sigma^{-1})\mathbf{X}$; and (4) running OLS on $\mathbf{Y}'$ and $\mathbf{X}'$. (When $\Sigma$ is the identity matrix, $\text{chol}(\Sigma^{-1})$ is as well, so $\mathbf{Y}' = \mathbf{Y}$ and $\mathbf{X}' = \mathbf{X}$ and GLS reduces to OLS.)

GLS was run separately on each gene pair, resulting in a 17,634 × 17,634 matrix of GLS $P$ values. Specifically, the *endog* argument of statsmodels.regression.linear_model.GLS (the output) was set to the length-481 vector of bias-corrected CERES scores for one of the two genes, the *exog* argument (input) set to a 481 × 2 matrix, where the first column is the other gene's bias-corrected CERES scores and the second column is a constant vector of all ones (that is, the intercept), and the $\Sigma$ argument set to the 481 × 481 covariance matrix of the bias-corrected CERES scores. Given these three pieces of data, the GLS outputs a $P$ value indicating the statistical significance of the degree of co-essentiality between the pair of genes. Note that, although the GLS $P$ value is consistent regardless of which of the two genes is chosen as *endog* and which as *exog*, the GLS effect size is not consistent with respect to this choice, and as a result is not reported. For benchmarking, GLS was also run on the non-bias-corrected data using the exact same procedure, but using the full 485 cell

lines. Benjamini–Hochberg FDR correction[82] was performed for each gene across its 17,633 partners.

As a computational optimization, the rate-limiting step of the GLS calculation (inverting the covariance matrix and then taking Cholesky's decomposition) was cached and reused for each pair of genes, because all gene pairs use the same covariance matrix. With this optimization, the amortized time complexity of GLS is equivalent to linear regression. The same GLS implementation was used to calculate Pearson's correlation (with and without bias correction) between each pair of genes, by setting the covariance matrix to the identity matrix.

**Identification of cancer-type-specific pathway dependencies.—**Cancer-type-specific pathway dependency *P* values for each module and cancer type (Supplementary Data 9) were obtained by (1) computing *P* values for each gene and cancer type, and then (2) aggregating *P* values across genes in each module. In step (1), GLS was run separately for each gene with the same covariance matrix and output/*endog* argument (bias-corrected essentiality for a particular gene) as before (see Quantifying co-essential gene pairs). However, unlike before, the *exog* argument (input) was set to a 481 × 21 matrix of binary indicator variables for the 20 cancer types listed in Fig. 6a (1 if a cell line is from that cancer type, 0 otherwise) plus an all-ones intercept column. The two other cancer types with CRISPR screen data from DepMap, cervical and biliary, were excluded due to having only a single cell line each. This multiple regression yielded 20 *P* values for the gene, one per cancer type. We note that this approach is equivalent to an analysis of variance, except using GLS instead of OLS.

In step (2), *P*-value aggregation was performed separately for each module and cancer type using the Cauchy Combination Test/Aggregated Cauchy Association Test[83,84] with equal weights on all genes. In Python, this step can be expressed straightforwardly as 'module_P = cauchy.sf(np.tan((0.5 - gene_ps) * np.pi).mean())', where gene_ps is a (number of module genes)-length vector of gene *P* values for a particular cancer type, and module_P is the combined *P* value for the module. Crucially, given that our gene-level *P* values are highly correlated among genes in a module, the test is able to accommodate *P* values from correlated tests (unlike the more commonly used Fisher's combined *P* test, which uses a $\chi^2$ instead of a Cauchy distribution to perform *P*-value aggregation), and we verified that the combined *P* values were not inflated (median *P* value = 0.56).

## Benchmarking on CORUM, hu.MAP, STRING and DoRothEA.

For the benchmarking, we compared five methods: co-essentiality with GLS or Pearson's and with or without bias correction, and co-expression with COXPRESdb. We used the same versions of COXPRESdb benchmarked in Pan et al.[24], downloaded from the supplementary data to that paper at https://ndownloader.figshare.com/files/10975364 and remapped from Entrez IDs to gene names using the mapping at https://ndownloader.figshare.com/files/9120082. When benchmarking, we considered only the $N = 15{,}552$ genes present in both the Avana library and COXPRESdb.

For STRING, we used all the gene pairs in v.10.5 restricted to *Homo sapiens* (https://stringdb-static.org/download/protein.links.detailed.v10.5/9606.protein.links.detailed.v10.5.txt.gz). To avoid circularity, we removed gene pairs supported only by co-expression, that is, for which the only non-zero score was for co-expression.

Following the strategy of Pan et al.[24], we compared methods by considering their rankings on a per-gene basis. Specifically, we considered only the top $N$ partners for each gene for $N$ from 1 to 10, and looked at how enrichment varied as a function of $N$. We used the same versions of CORUM and hu.MAP benchmarked in Pan et al.[24] (that is, CORUM Core complexes 3.5.2017 release and hu.MAP v.1).

Enrichments were calculated as the percentage of the top $N$ gene pairs in the pathway or complex database, divided by the percentage of gene pairs found in the database. For instance, to calculate the enrichment of COXPRESdb in CORUM for $N = 2$, we found the top two co-expressed partners per gene according to COXPRESdb ($N = 2 \times 15{,}552$ gene pairs), computed the percentage of these pairs that were part of the same CORUM complex, and divided by the percentage of the $15{,}552 \times 15{,}552$ gene pairs that were part of the same CORUM complex.

Note that Boyle et al.[25] perform an additional transformation of $P$ values after PCA correction based on the empirical null distribution of $P$ values for olfactory genes, but as this transformation is monotonic it does not affect the rankings of partner genes used in our benchmarking.

### Co-essential modules.

Co-essential modules were ascertained with the ClusterONE algorithm[42]. Briefly, ClusterONE generates modules by greedily adding nodes (genes), starting from a randomly selected seed node, so long as the sum of the edge weights within the module is sufficiently high relative to the sum of the boundary edge weights between genes in the module and their neighbors. It then merges sufficiently overlapping modules as a post-processing step, while allowing genes to be members of multiple modules (protein complexes or pathways).

ClusterONE was run on the $17{,}634 \times 17{,}634$ matrix of GLS FDRs, with edge weights set to 1 minus the FDR $q$ value[82]. Default settings were used for ClusterONE, except for changing the module density parameter -d (also known as --min-density) from its default of 0.3, as discussed in the main text. For the list of 5,229 modules in Supplementary Data 2, all modules generated with values of $d$ set to 0.2, 0.5 and 0.9 were merged into a single list. Eleven modules that were identical at different values of $d$ were retained in this list but were excluded from the reported count of total modules (5,218).

We noted that the resulting list of co-essential modules contained many modules that are highly enriched for genes that localize close to each other in the genome. In several cases, these modules correspond to clusters of functionally related genes that are known to colocalize in the genome, such as histone- and protocadherin-encoding genes, although in most cases it remains unclear whether the presence of colocalized genes in a module

reflects their shared function in a biological pathway or relates to vulnerabilities of CRISPR screening to copy-number artifacts that are difficult to account for perfectly[27]. Supporting the idea that co-essentiality for colocalized genes may represent a mix of true- and false-positive signals, we find substantial enrichment of syntenic gene pairs (both genes on the same chromosome) in CORUM, hu.MAP and STRING, but less enrichment than for nonsyntenic gene pairs (Extended Data Fig. 6). To enable full utilization of the dataset as well as easy discernment of syntenic and nonsyntenic gene pairs and modules, we report all co-essential gene pairs and modules in Supplementary Data 1 (co-essential pairs), Supplementary Data 2 (co-essential modules) and Supplementary Data 3 (uncharacterized gene predictions), and annotate each as syntenic or nonsyntenic.

### Global structure of the co-essential network.

The 2D interaction map visualization was constructed to have two properties: (1) genes in many of the same ClusterONE modules are close together; and (2) gene pairs with high GLS co-essentiality are close together. This was done by forming a graph $G_{CO}$ from the ClusterONE modules (as above) and another $G_{GLS}$ from the co-essentiality data, mixing the two with proportion $\alpha$ to form the mixed graph:

$$G = \alpha G_{CO} + (1 - \alpha) G_{GLS}$$

(We set $\alpha = 0.99$ to rely on the relatively specific and dense ClusterONE modules where possible, while falling back on pairwise GLS analysis to link genes not in any module to the rest of the network.)

The graph $G_{GLS}$ was constructed by computing, for each pair of genes, $-\log(P)$ given by GLS between the two genes. This was denoised and compressed by keeping each gene's edges to its ten nearest neighbors and zeroing the other edges, resulting in each gene having a minimum of ten neighbors in the graph. (We found our analyses fairly stable to varying the number of nearest neighbors between 4 and 100.) The graph $G_{CO}$ was constructed using the same procedure, but with each pairwise similarity computed using Jaccard's similarity between the sets of ClusterONE modules to which the respective genes belonged (for sets $A$ and $B$, this is $J(A,B) = |A \cap B|/|A \cup B|$).

To visualize the network $G$ efficiently on a global scale, we relied on the framework of diffusion maps[36], which basically decompose the variation in essentiality profiles over the network into short- and long-range pathway components, resulting in an embedded space for genes in the network. The genes' positions in the present study are relatively accurate for genes in well-separated pathways, and less so for finer distinctions—this embedded space (the 'diffusion map') is a smoothed version of the network, with each gene being represented in low-dimension $d = 40$. The embedded space was constructed from $G$ as follows.

$G$ was first normalized to remove the disproportionate influence of high-degree 'hub' genes in the layout, resulting in a matrix $G_2$. With this gene-wise degree expanded as a matrix $D_G = \mathrm{diag}(\Sigma_j G_{ij})$, the normalization operation is:

$$G_2 = D_G^{-1} G D_G^{-1}$$

This density normalization further corrects for biased sampling of the network by the data[36,85], as analyses on $G_2$ consider the gene network corrected for the variable density of characterized genes.

The diffusion map embeds $G_2$, and takes the properties of random walks on it to reveal a multi-scale pathway structure. The transition probabilities of such a random walk on $G_2$ are the row-sum-normalized $T = D_2^{-1} G_2$, where $D_2 = \mathrm{diag}(\Sigma_j [G_2]_{ij})$.

This transition matrix $T$ describes the evolution of any random walk, and its right eigenvectors $\mathbf{e}_1, \ldots, \mathbf{e}_n$ give a diffusion map embedding when appropriately scaled. The embedding requires a parameter $t$, which controls the overall scale of the pathways modeled by the embedding. If the corresponding eigenvalues are $\lambda_1 \geq \lambda_2 \geq, \ldots$, then for any $t > 0$, the embedded coordinates of the genes $[\boldsymbol{\Phi}_t]_1$, $[\boldsymbol{\Phi}_t]_2$, …, $[\boldsymbol{\Phi}_t]_{40}$ are:

$$[\boldsymbol{\Phi}_t]_i = \lambda_i^t \mathbf{e}_i$$

A crucial choice is that of the scale parameter $t$. As the current co-essentiality data are somewhat noisy for inferring fine-grained gene–gene relationships, we found it necessary to smooth them by increasing the value of $t$ in constructing the embedding. We increased $t$ to the minimum such that $d = 40$ dimensions captured 90% of the variance in the embedded space $\boldsymbol{\Phi}_t$, and computed the resulting diffusion map $\boldsymbol{\Phi}$. This simultaneous optimization of $t$ and $\boldsymbol{\Phi}_t$ made the procedure adapt to and preserve large-scale global structure in a fully data-driven way, without substantive parameter tuning and using only a few matrix multiplications and one singular value decomposition.

We applied UMAP[35] to this diffusion map embedding as in scanpy for the final global layout. Our diffusion map implementation is in Python using the numpy and scipy packages, and includes other choices of normalization as well. The entire process ran in less than 4 min on the GLS- and ClusterONE-derived matrices on an Intel i7 Core CPU.

## Lipidomics.

HeLa cells expressing sgRNAs targeting either safe loci or the *TMEM189* or *SPTLC2* loci were cultured in quadruplicate and harvested by centrifugation after washing with phosphate-buffered saline. Lipids were extracted from 60 mg of cell pellets using a biphasic separation with cold methyl *tert*-butyl ether, methanol and water, as described previously[51]. The solvent mixture contained labeled standard lipids stock (SCIEX, catalog no. 5040156) to control for extraction efficiency and facilitate quantification relative to the known concentrations.

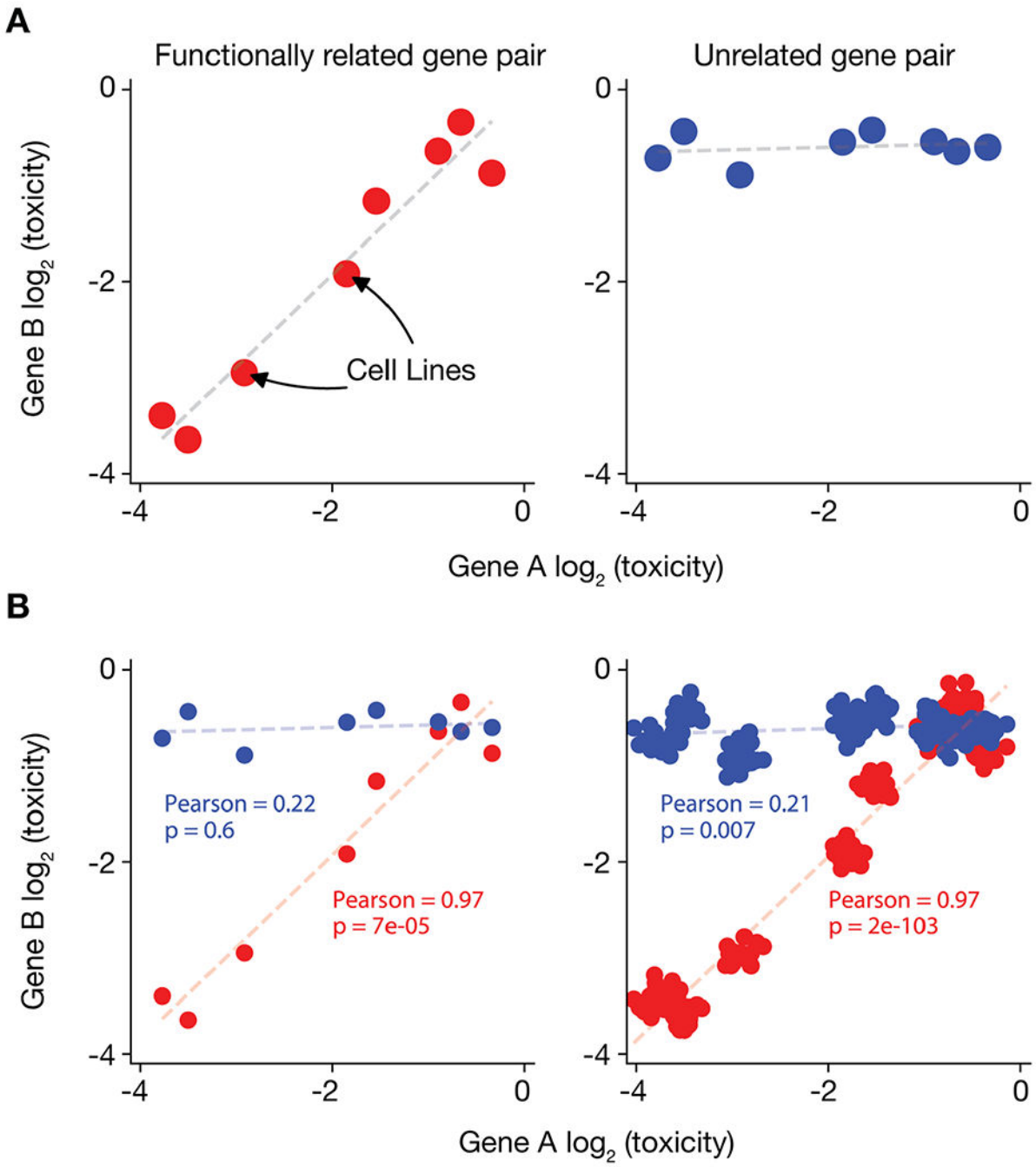
Lipid extracts were analyzed by MS using the Lipidyzer platform[50], comprising a 5500 QTRAP mass spectrometer equipped with a differential mobility scan (DMS) interface (SCIEX) and high-flow LC-30AD delivery unit (Shimadzu), as described previously[51].

Briefly, flow injection analysis was performed at 8 μl $min^{-1}$ in 10 mM ammonium acetate in 50:50 dichloromethane:methanol running solution, with 1-propanol included in curtain gas. DMS parameter settings were set as follows: temperature = low, separation voltage = 3.5 kV and DMS resolution = low. Phosphatidylcholine and phosphatidylethanolamine were quantified with DMS on and in negative ionization mode; sphingomyelin was quantified with DMS on and in positive ionization mode; free fatty acids were quantified with DMS off and in negative ionization mode; triacylgycerol, diacylglycerol, and ceramides were quantified with DMS off and in positive ionization mode. DMS compensation voltages were tuned using a set of lipid standards (SCIEX, catalog no. 5040141), and a quick system suitability test (SCIEX, catalog no. 50407) was performed to ensure an acceptable limit of detection for each lipid class. Lipid molecular species were quantified with the Lipidyzer Workflow Manager using 54 deuterated IS developed with Avanti Polar Lipids covering 10 lipid classes (SCIEX, catalog no. 5040156). Some 17 plasmenylethanolamine species with fully saturated, 18-carbon chains at the *sn*-1 position were excluded from analyses, because they cannot be reliably differentiated from plasmanylethanolamine species containing unsaturated 18-carbon chains at the *sn*-1 position with the Lipidyzer platform (M. Pearson, SCIEX, personal communication).

## Reporting Summary.

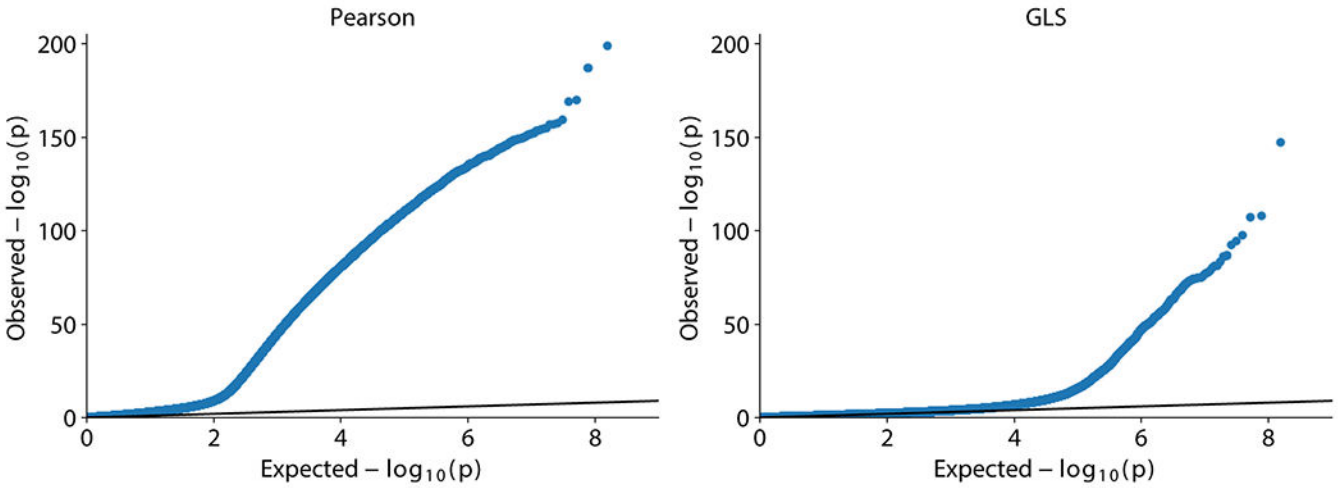
Further information on research design is available in the Nature Research Reporting Summary linked to this article.

## Extended Data

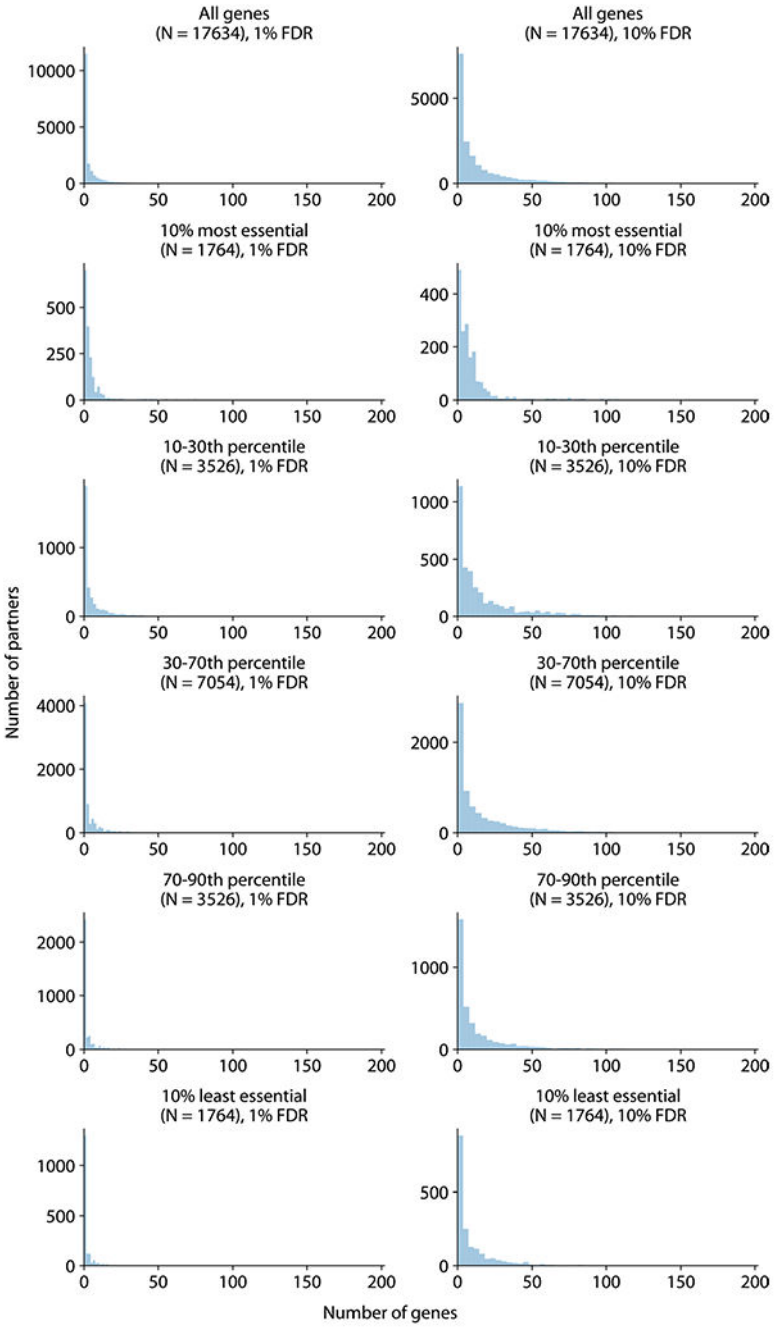



**Extended Data Fig. 1|. Co-essentiality profiling and the limitations of Pearson's correlation.**
**a**. The concept of co-essentiality: (left) a pair of functionally related genes are both essential in some cell lines and both non-essential in other lines. Essentiality can be quantified from CRISPR screens as the logarithm of the growth effect of the gene's knockout (intuitively, the number of times fewer cells with the knockout doubled during the screen, compared to control cells). (Right) a pair of unrelated genes have uncorrelated essentiality across cell

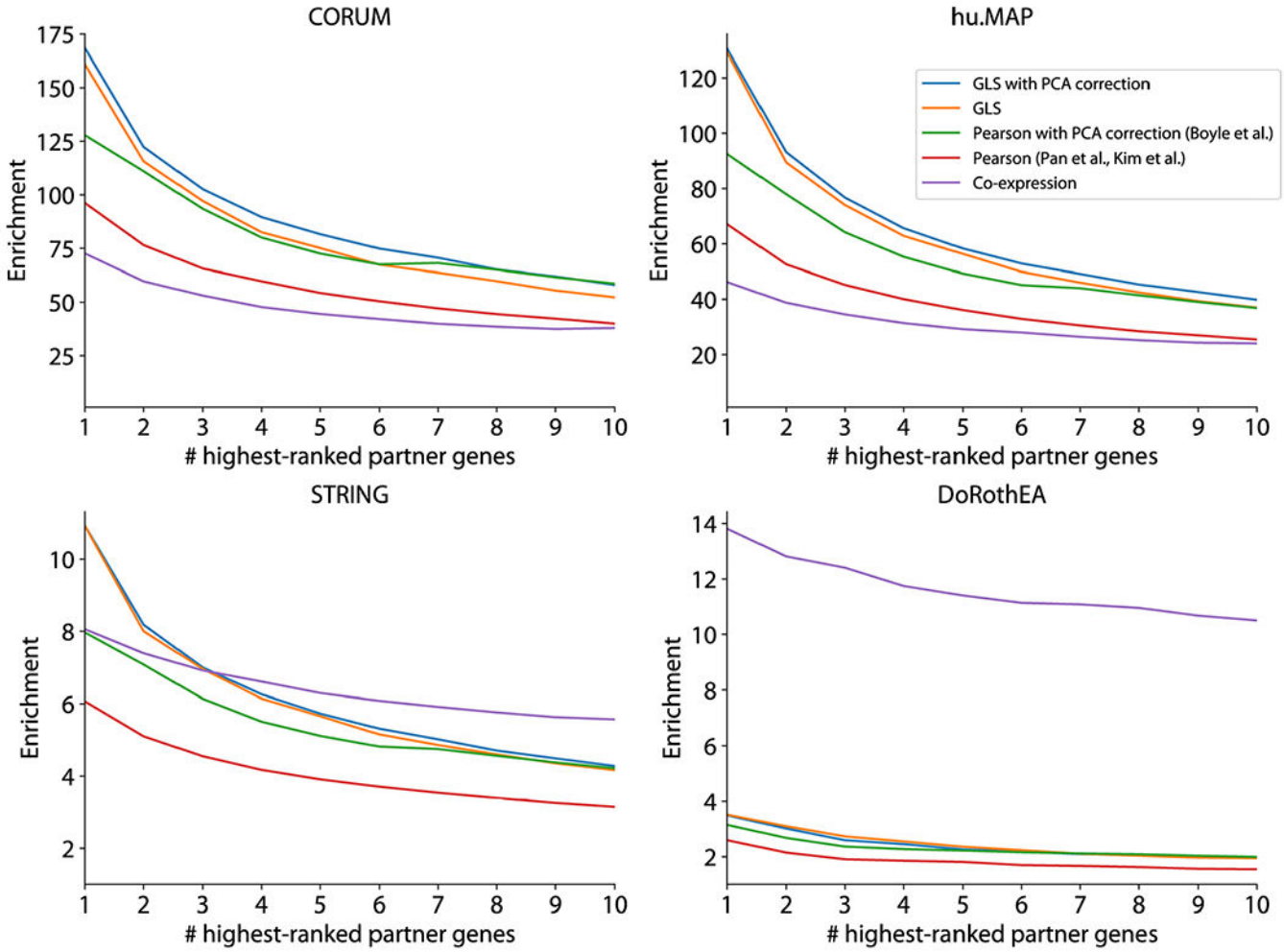
lines. **b**. Simulation of how biological relatedness between cell lines inflates Pearson's correlation *p*-values. Duplicating each point 10 times with slight noise (analogous to duplicating each screen in 10 related lines) makes the previously non-significant (p = 0.6) blue correlation highly significant (p = 0.007) and the significant red correlation ($p = 7 \times 10^{-5}$) substantially more so ($p = 2 \times 10^{-103}$), despite similar correlation magnitudes.




**Extended Data Fig. 2|. Quantile-quantile plots for Pearson's and GLS.**
Quantile-quantile plots for Pearson's correlation and GLS *p*-values (an alternate visualization of the *p*-value histograms in Fig. 1b). The observed *p*-values (y), sorted from largest to smallest, are plotted against the uniform distribution of *p*-values (x) expected under the null hypothesis.

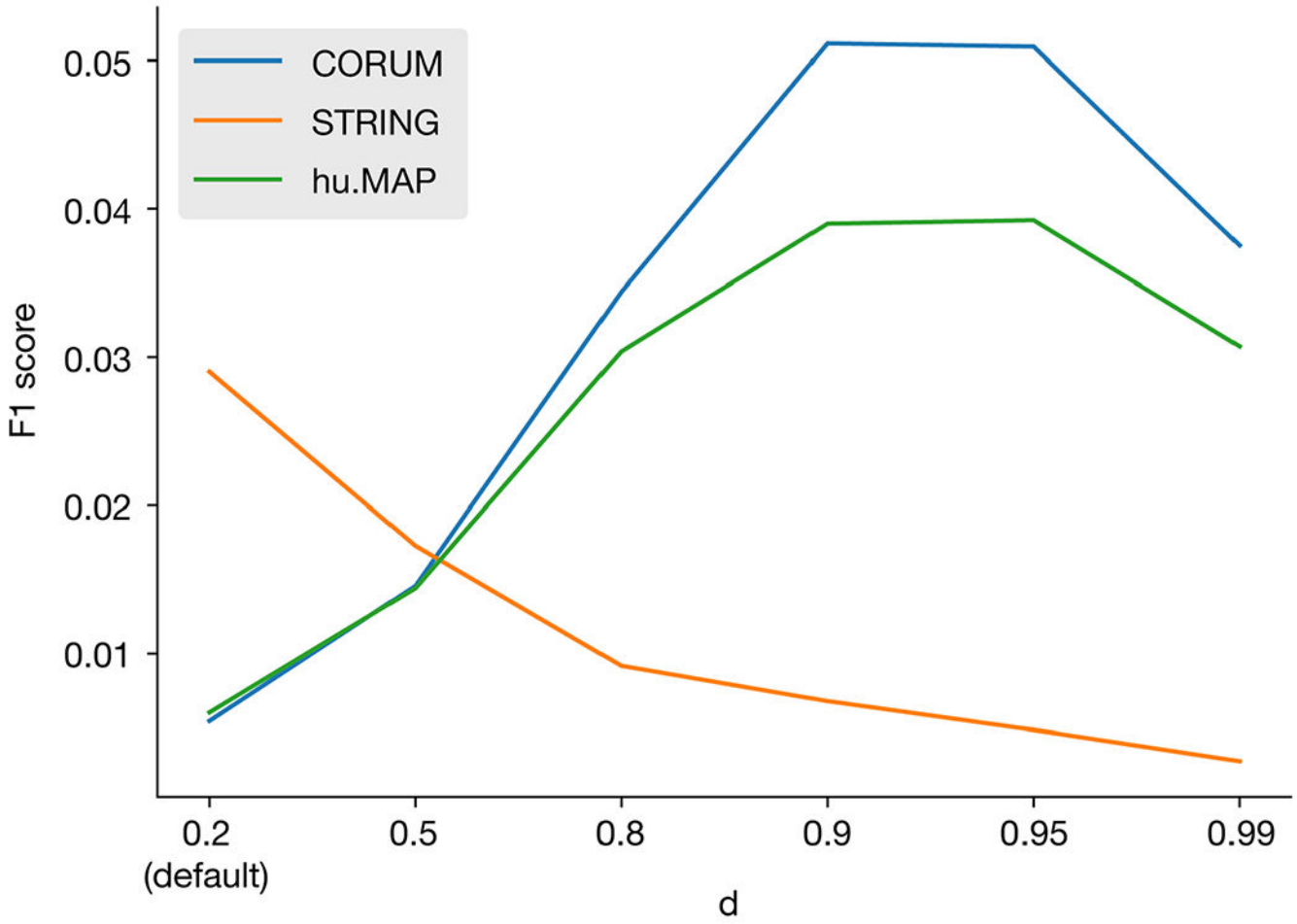



**Extended Data Fig. 3|. Number of co-essential partners per gene by average gene essentiality.** Histograms of genes' number of co-essential partners at 1% and 10% FDR as a function of the gene's average essentiality (pre-bias-correction CERES score) across lines.



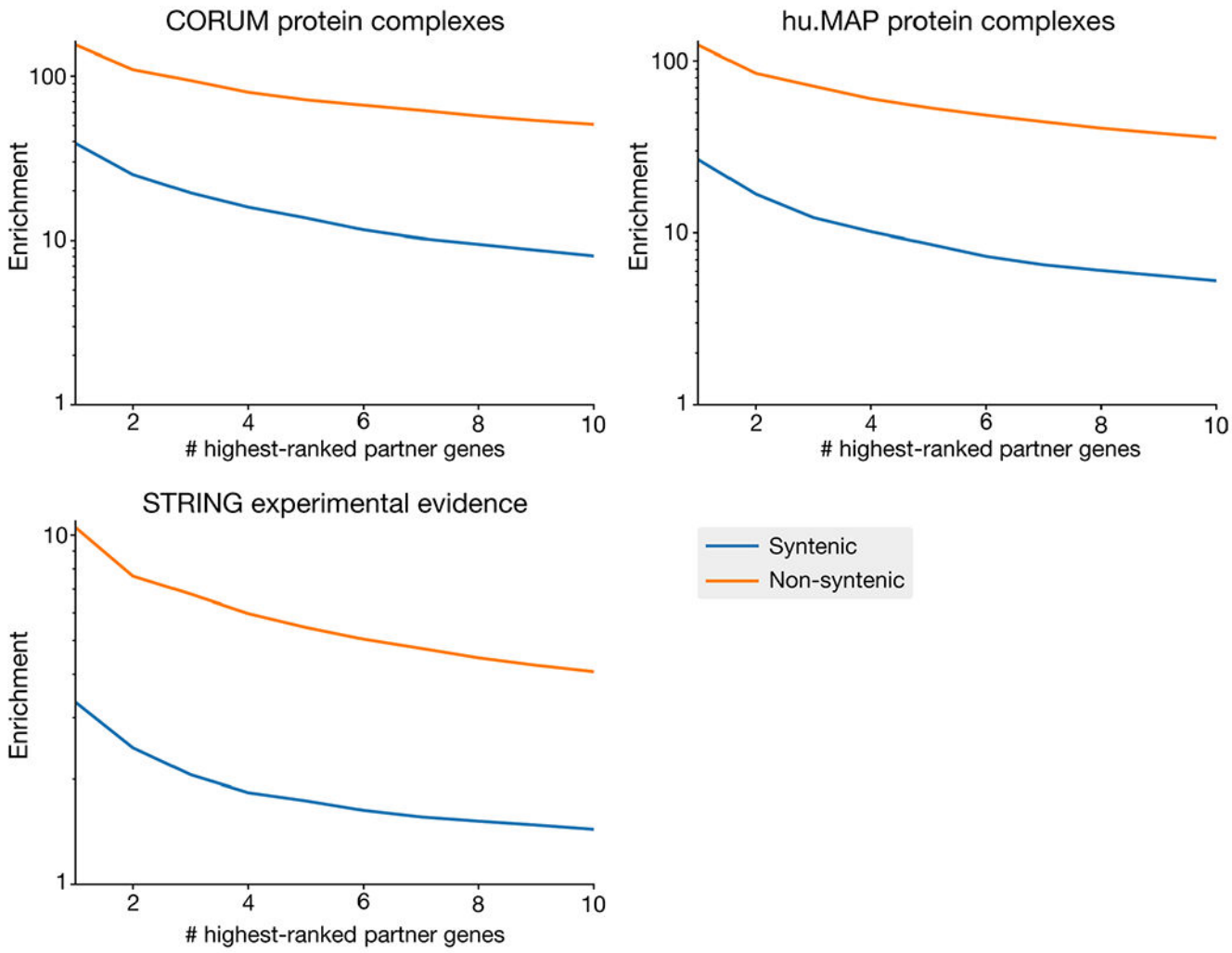

**Extended Data Fig. 4|. GLS improves recall of known functional interactions in co-essential gene pairs with and without PCA-based bias correction.**

Enrichment of interactions from GLS- and Pearson's-based co-essentiality using the DepMap dataset, as well as co-expression using the COXPRESdb dataset, in CORUM, hu.MAP and STRING, considering the top 1-10 partners per gene, similar to Fig. 2a but including GLS- and Pearson's-based co-essentiality done both with and without PCA-based bias correction.



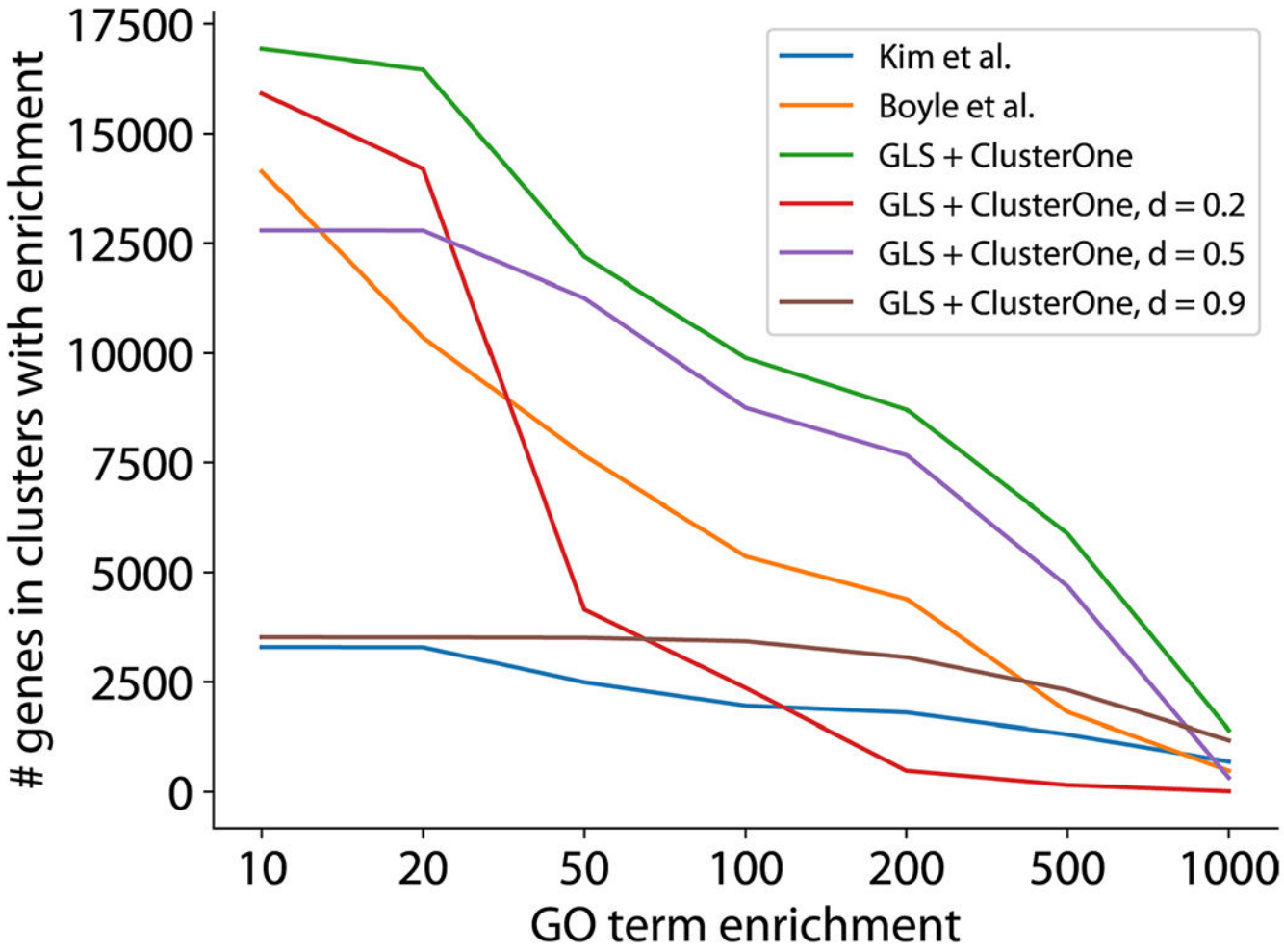

**Extended Data Fig. 5|. Benchmarking of cluster density *d*.**

F1 score (harmonic mean of precision and recall) for various values of the module density parameter *d* on CORUM, hu.MAP and STRING. F1 scores represent the performance of a binary network based on the modules (that is “are genes A and B in the same module?”) at predicting a binary network based on the benchmark dataset (that is “are genes A and B partners in the benchmark dataset?”).



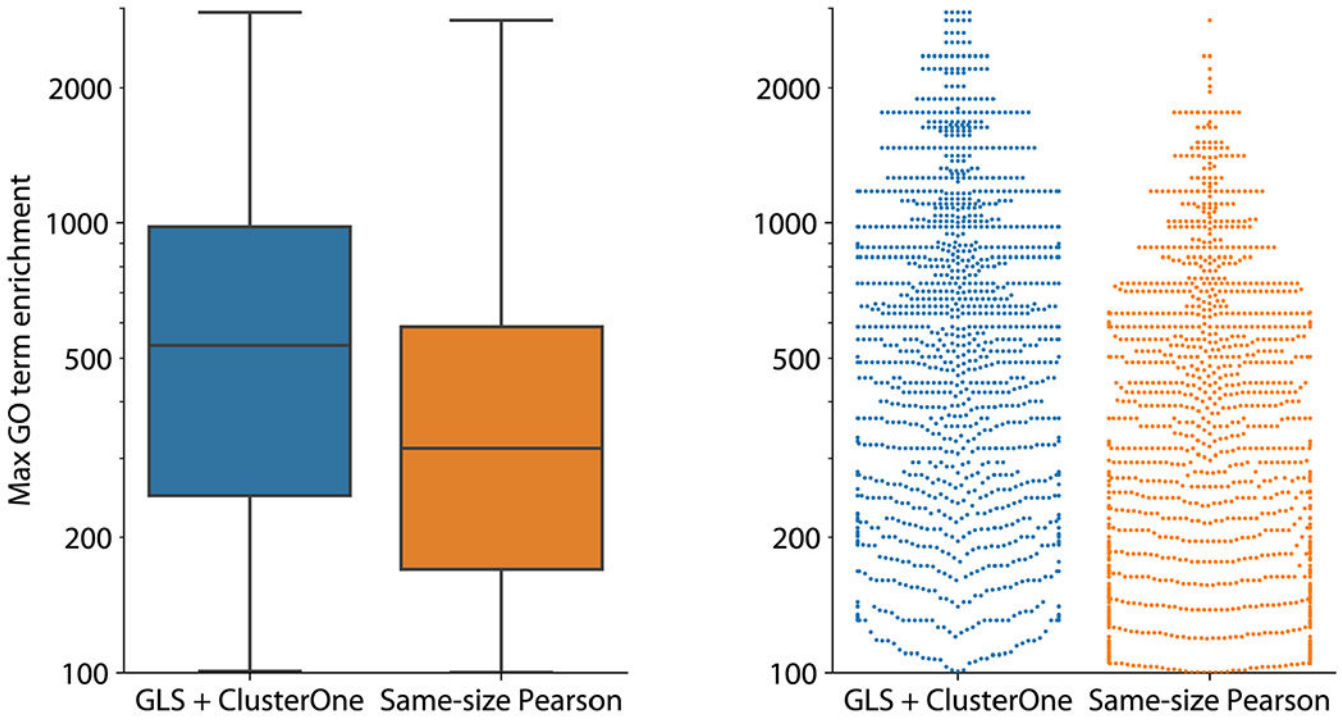

**Extended Data Fig. 6|. Benchmarking of syntenic versus non-syntenic genes.**

Enrichment of syntenic (both genes on same chromosome) and non-syntenic co-essential pairs for annotated interactions CORUM, hu.MAP and STRING databases, using the same benchmarking strategy as in Fig. 2a.



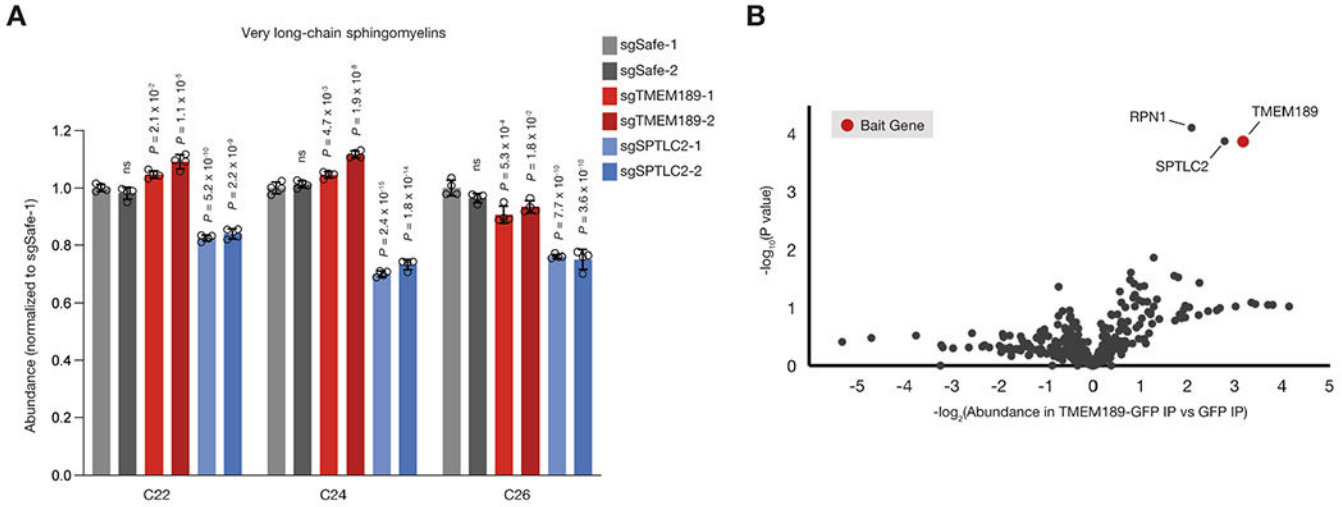

**Extended Data Fig. 7|. Number of genes assigned putative functions by various co-essentiality module detection methods, after excluding syntenic modules.**

Number of genes in non-syntenic clusters/modules at least N-fold enriched for some GO term with at least 5 total genes present across all clusters/modules, excluding the gene itself from the enrichment calculation, for various N from 10 to 1000.



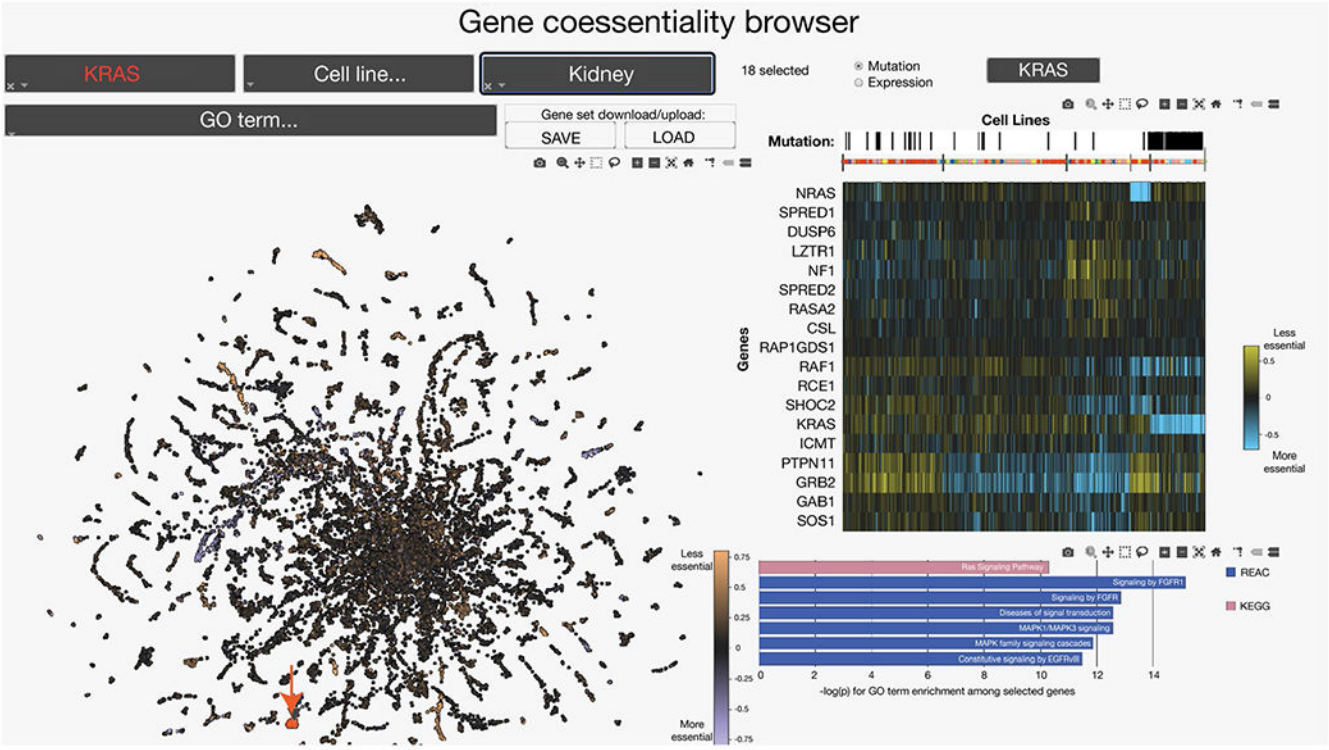

**Extended Data Fig. 8|. Strength of correct functional predictions of our modules versus same-size Pearson.**

Maximum GO term enrichment across all correctly predicted GO terms, for each of the n = 1407 genes correctly predicted by both our modules and same-size Pearson, shown as a boxplot (left) and swarmplot (right). Boxplot centre represents median, bounds of box represent 25th and 75th percentiles, and minima and maxima represent the minimum and maximum values, respectively.




**Extended Data Fig. 9|. Additional functional characterization of *TMEM189* suggests a secondary role in sphingolipid biosynthesis.**

**a**. Abundances (relative to Safe-targeting sgRNA control #1) of very long chain sphingomyelin species (with acyl chain length indicated on x-axis) in cell extracts prepared from HeLa cells transduced with indicated sgRNAs. sgSafe data and sgTMEM189 data are from same data set represented in Fig. 4c. n = 4 biologically independent cell extracts. Data are presented as mean+/− s.d. **b**. Volcano plot of mass spectrometric (TMT) analysis of TMEM189-GFP immunoprecipitates. Data are from same mass spectrometry analysis as data shown in Fig. 5d.




**Extended Data Fig. 10|. A web tool for interactive exploration of the co-essential network.** Example use case for the interactive web tool (http://coessentiality.net). A gene, *KRAS*, was selected using the dropdown menu at top left and is marked with a red arrow in the scatterplot below. Genes selected for analysis – KRAS and its gene neighborhood – are designated with red points in the main panel (left). The heatmap panel (top right) shows that KRAS-mutant lines (selected for display using the search bar above the heat map and indicated as black marks in the “Mutation” bar above the heatmap) are enriched in a cluster (far right) that is marked by increased essentiality of KRAS. The pathway enrichment panel (bottom right) shows strong enrichments for Ras signaling and related pathways. The points in the main panel have also been selected in the tissue search bar (top middle) to be colored according to the average essentialities of each gene in kidney-derived cell lines. Gene sets can also be either saved or uploaded as csv files using the respective buttons in the top center (under “Gene set download/upload”). Some web colors and font sizes were optimized for display in this figure.

## Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

## Acknowledgements

We thank R. Zoeller (Boston University Medical Center) for providing RAW.12 cells and the parent RAW264.7 cell line. We thank E. Boyle, J. Donnelly, M. Pearson, G. Anderson, S. Simpkins, T. Ideker and members of the Bassik and Kundaje laboratories for helpful discussions. This work was supported by a National Institute of Health (NIH) Director's New Innovator award (no. 1DP2HD084069-01 to M.C.B.), a grant from NIH/ENCODE (no. 5UM1HG009436-02 to A.K. and M.C.B.), a Stanford Bio-X Bowes Fellowship (to M.W.), and a Stanford School of Medicine Dean's Postdoctoral Fellowship and a Jane Coffin Childs Postdoctoral Fellowship (to R.A.K.).

## Data availability

The Achilles project 18C3 release is publicly available at https://ndownloader.figshare.com/files/12704099 or https://depmap.org/portal/download/all under release 'DepMap Public 18Q3' and file 'gene_effect.csv'. The HUGO Gene Nomenclature Committee Database is accessible at https://www.genenames.org. The STRING database is accessible at https://string-db.org. The CORUM database is accessible at https://mips.helmholtz-muenchen.de/corum. The hu.MAP database is accessible at http://proteincomplexes.org. The DoRothEA database is accessible at https://saezlab.github.io/dorothea. The COXPRESdb database is accessible at https://coxpresdb.jp. Data supporting the findings of the present study are available upon reasonable request. Lipidomic raw data, acquisition methods and quantitative results are available as Supplementary Data 5–7. The raw MS proteomic data have been deposited to the ProteomeXchange Consortium via the PRIDE[86] partner repository (http://www.ebi.ac.uk/pride) with the dataset identifier PXD023558. Source data are provided with this paper.

## References

1. Barabási A-L & Oltvai ZN Network biology: understanding the cell's functional organization. Nat. Rev. Genet 5, 101–113 (2004). [PubMed: 14735121]
2. Chuang H-Y, Hofree M & Ideker T A decade of systems biology. Annu. Rev. Cell Dev. Biol 26, 721–744 (2010). [PubMed: 20604711]
3. Stuart JM, Segal E, Koller D & Kim SK A gene-coexpression network for global discovery of conserved genetic modules. Science 302, 249–255 (2003). [PubMed: 12934013]
4. Costanzo M et al. A global genetic interaction network maps a wiring diagram of cellular function. Science 353, aaf1420 (2016). [PubMed: 27708008]
5. Horlbeck MA et al. Mapping the genetic landscape of human cells. Cell 174, 953–967.e22 (2018). [PubMed: 30033366]
6. Hartwell LH, Hopfield JJ, Leibler S & Murray AW From molecular to modular cell biology. Nature 402, C47–C52 (1999). [PubMed: 10591225]
7. Harrow J et al. GENCODE: the reference human genome annotation for the ENCODE Project. Genome Res. 22, 1760–1774 (2012). [PubMed: 22955987]
8. Carpenter AE & Sabatini DM Systematic genome-wide screens of gene function. Nat. Rev. Genet 5, 11–22 (2004). [PubMed: 14708012]
9. Alonso JM & Ecker JR Moving forward in reverse: genetic technologies to enable genome-wide phenomic screens in *Arabidopsis*. Nat. Rev. Genet 7, 524–536 (2006). [PubMed: 16755288]

10. Mohr SE, Smith JA, Shamu CE, Neumüller RA & Perrimon N RNAi screening comes of age: improved techniques and complementary approaches. Nat. Rev. Mol. Cell Biol 15, 591–600 (2014). [PubMed: 25145850]
11. Shalem O, Sanjana NE & Zhang F High-throughput functional genomics using CRISPR–Cas9. Nat. Rev. Genet 16, 299–311 (2015). [PubMed: 25854182]
12. Tong AH et al. Systematic genetic analysis with ordered arrays of yeast deletion mutants. Science 294, 2364–2368 (2001). [PubMed: 11743205]
13. Tong AHY Global mapping of the yeast genetic interaction network. Science 303, 808–813 (2004). [PubMed: 14764870]
14. Costanzo M et al. The genetic landscape of a cell. Science 327, 425–431 (2010). [PubMed: 20093466]
15. Bassik MC et al. A systematic mammalian genetic interaction map reveals pathways underlying ricin susceptibility. Cell 152, 909–922 (2013). [PubMed: 23394947]
16. Rosenbluh J et al. Genetic and proteomic interrogation of lower confidence candidate genes reveals signaling networks in β-catenin-active cancers. Cell Syst. 3, 302–316.e4 (2016). [PubMed: 27684187]
17. Shen JP et al. Combinatorial CRISPR–Cas9 screens for de novo mapping of genetic interactions. Nat. Methods 14, 573–576 (2017). [PubMed: 28319113]
18. Han K et al. Synergistic drug combinations for cancer identified in a CRISPR screen for pairwise genetic interactions. Nat. Biotechnol 35, 463–474 (2017). [PubMed: 28319085]
19. Du D et al. Genetic interaction mapping in mammalian cells using CRISPR interference. Nat. Methods 14, 577–580 (2017). [PubMed: 28481362]
20. Boettcher M et al. Dual gene activation and knockout screen reveals directional dependencies in genetic networks. Nat. Biotechnol 36, 170–178 (2018). [PubMed: 29334369]
21. Wang T et al. Gene essentiality profiling reveals gene networks and synthetic lethal interactions with oncogenic ras. Cell 168, 890–903.e15 (2017). [PubMed: 28162770]
22. Rauscher B et al. Toward an integrated map of genetic interactions in cancer cells. Mol. Syst. Biol 14, e7656 (2018). [PubMed: 29467179]
23. McDonald ER 3rd et al. Project DRIVE: a compendium of cancer dependencies and synthetic lethal relationships uncovered by large-scale, deep RNAi screening. Cell 170, 577–592.e10 (2017). [PubMed: 28753431]
24. Pan J et al. Interrogation of mammalian protein complex structure, function, and membership using genome-scale fitness screens. Cell Syst. 6, 555–568.e7 (2018). [PubMed: 29778836]
25. Boyle EA, Pritchard JK & Greenleaf WJ High-resolution mapping of cancer cell networks using co-functional interactions. Mol. Syst. Biol 14, e8594 (2018). [PubMed: 30573688]
26. Kim E et al. A network of human functional gene interactions from knockout fitness screens in cancer cells. Life Sci. Alliance 2, e201800278 (2019). [PubMed: 30979825]
27. Meyers RM et al. Computational correction of copy number effect improves specificity of CRISPR–Cas9 essentiality screens in cancer cells. Nat. Genet 49, 1779–1784 (2017). [PubMed: 29083409]
28. Tsherniak A et al. Defining a cancer dependency map. Cell 170, 564–576.e16 (2017). [PubMed: 28753430]
29. Aitkin AC On least squares and linear combination of observations. Proc. R. Soc. Edinb 55, 42–48 (1935).
30. Yang J et al. Genomic inflation factors under polygenic inheritance. Eur. J. Hum. Genet 19, 807–812 (2011). [PubMed: 21407268]
31. Moll UM & Petrenko O The MDM2-–p53 interaction. Mol. Cancer Res 1, 1001–1008 (2003). [PubMed: 14707283]
32. Wang X et al. PHLDA2 is a key oncogene-induced negative feedback inhibitor of EGFR/ErbB2 signaling via interference with AKT signaling. Oncotarget 9, 24914 (2018). [PubMed: 29861842]
33. Furukawa T, Tanji E, Xu S & Horii A Feedback regulation of DUSP6 transcription responding to MAPK1 via ETS2 in human cells. Biochem. Biophys. Res. Commun 377, 317–320 (2008). [PubMed: 18848526]

34. Rickman DS, Schulte JH & Eilers M The expanding world of N-MYC–driven tumors. Cancer Disco 8, 150–163 (2018).
35. McInnes L, Healy J, Saul N & Großberger L UMAP: Uniform Manifold Approximation and Projection. J. Open Source Softw 3, 861 (2018).
36. Coifman RR & Lafon S Diffusion maps. Appl. Comput. Harmon. Anal 21, 5–30 (2006).
37. Ruepp A et al. CORUM: the comprehensive resource of mammalian protein complexes. Nucleic Acids Res 36, D646–D650 (2008). [PubMed: 17965090]
38. Drew K et al. Integration of over 9,000 mass spectrometry experiments builds a global map of human protein complexes. Mol. Syst. Biol 13, 932 (2017). [PubMed: 28596423]
39. Szklarczyk D et al. The STRING database in 2017: quality-controlled protein–protein association networks, made broadly accessible. Nucleic Acids Res. 45, D362–D368 (2017). [PubMed: 27924014]
40. Okamura Y et al. COXPRESdb in 2015: coexpression database for animal species by DNA-microarray and RNAseq-based expression data with multiple quality assessment systems. Nucleic Acids Res. 43, D82–D86 (2015). [PubMed: 25392420]
41. Garcia-Alonso L, Holland CH, Ibrahim MM, Turei D & Saez-Rodriguez J Benchmark and integration of resources for the estimation of human transcription factor activities. Genome Res. 29, 1363–1375 (2019). [PubMed: 31340985]
42. Nepusz T, Yu H & Paccanaro A Detecting overlapping protein complexes in protein–protein interaction networks. Nat. Methods 9, 471–472 (2012). [PubMed: 22426491]
43. Saxton RA & Sabatini DM mTOR signaling in growth, metabolism, and disease. Cell 169, 361–371 (2017).
44. Shoemaker CJ et al. CRISPR screening using an expanded toolkit of autophagy reporters identifies TMEM41B as a novel autophagy factor. PLoS Biol. 17, e2007044 (2019). [PubMed: 30933966]
45. Breslow DK et al. A CRISPR-based screen for Hedgehog signaling provides insights into ciliary function and ciliopathies. Nat. Genet 50, 460–471 (2018). [PubMed: 29459677]
46. Blomen VA et al. Gene essentiality and synthetic lethality in haploid human cells. Science 350, 1092–1096 (2015). [PubMed: 26472760]
47. Nagan N & Zoeller RA Plasmalogens: biosynthesis and functions. Prog. Lipid Res 40, 199–229 (2001). [PubMed: 11275267]
48. Vaz FM et al. Mutations in PCYT2 disrupt etherlipid biosynthesis and cause a complex hereditary spastic paraplegia. Brain 142, 3382–3397 (2019). [PubMed: 31637422]
49. Horibata Y et al. EPT1 (selenoprotein I) is critical for the neural development and maintenance of plasmalogen in humans. J. Lipid Res 59, 1015–1026 (2018). [PubMed: 29500230]
50. Contrepois K et al. Cross-platform comparison of untargeted and targeted lipidomics approaches on aging mouse plasma. Sci. Rep 8, 17747 (2018). [PubMed: 30532037]
51. Schüssler-Fiorenza Rose SM et al. A longitudinal big data approach for precision health. Nat. Med 25, 792–804 (2019). [PubMed: 31068711]
52. Snyder F, Lee T-C & Wykle RL in The Enzymes of Biological Membranes, Vol. 2, Biosynthesis and Metabolism (ed. Martonosi AN) 1–58 (Springer US, 1985).
53. Zoeller RA et al. Mutants in a macrophage-like cell line are defective in plasmalogen biosynthesis, but contain functional peroxisomes. J. Biol. Chem 267, 8299–8306 (1992). [PubMed: 1569085]
54. Gao J et al. Fatty acid desaturase4 of *Arabidopsis* encodes a protein distinct from characterized fatty acid desaturases. Plant J. 60, 832–839 (2009). [PubMed: 19682287]
55. Motley A, Bright NA, Seaman MNJ & Robinson MS Clathrin-mediated endocytosis in AP-2-depleted cells. J. Cell Biol 162, 909–918 (2003). [PubMed: 12952941]
56. Huttlin EL et al. Architecture of the human interactome defines protein communities and disease networks. Nature 545, 505–509 (2017). [PubMed: 28514442]
57. Huttlin EL et al. The BioPlex Network: a systematic exploration of the human interactome. Cell 162, 425–440 (2015). [PubMed: 26186194]
58. Chan EM et al. WRN helicase is a synthetic lethal target in microsatellite unstable cancers. Nature 568, 551–556 (2019). [PubMed: 30971823]

59. Ariazi E, Ariazi J, Cordera F & Jordan V Estrogen receptors as therapeutic targets in breast cancer. Curr. Top. Med. Chem 6, 181–202 (2006). [PubMed: 16515478]
60. Fletcher MNC et al. Master regulators of FGFR2 signalling and breast cancer risk. Nat. Commun 4, 2464 (2013). [PubMed: 24043118]
61. Roman SD et al. Estradiol induction of retinoic acid receptors in human breast cancer cells. Cancer Res. 53, 5940–5945 (1993). [PubMed: 8261407]
62. Zhang Y-W et al. Acquisition of estrogen independence induces TOB1-related mechanisms supporting breast cancer cell proliferation. Oncogene 35, 1643–1656 (2016). [PubMed: 26165839]
63. Ascierto PA et al. The role of BRAF V600 mutation in melanoma. J. Transl. Med 10, 85 (2012). [PubMed: 22554099]
64. Garraway LA et al. Integrative genomic analyses identify MITF as a lineage survival oncogene amplified in malignant melanoma. Nature 436, 117–122 (2005). [PubMed: 16001072]
65. Perotti V et al. NFATc2 is an intrinsic regulator of melanoma dedifferentiation. Oncogene 35, 2862–2872 (2016). [PubMed: 26387540]
66. Harris ML, Baxter LL, Loftus SK & Pavan WJ Sox proteins in melanocyte development and melanoma. Pigment Cell Melanoma Res 23, 496–513 (2010). [PubMed: 20444197]
67. Gallego-García A et al. A bacterial light response reveals an orphan desaturase for human plasmalogen synthesis. Science 366, 128–132 (2019). [PubMed: 31604315]
68. Werner ER et al. The TMEM189 gene encodes plasmanylethanolamine desaturase which introduces the characteristic vinyl ether double bond into plasmalogens. Proc. Natl Acad. Sci. USA 117, 7792–7798 (2020). [PubMed: 32209662]
69. Piano V et al. Discovery of inhibitors for the ether lipid-generating enzyme AGPS as anti-cancer agents. ACS Chem. Biol 10, 2589–2597 (2015). [PubMed: 26322624]
70. Zhu C et al. The fusion landscape of hepatocellular carcinoma. Mol. Oncol 13, 1214–1225 (2019). [PubMed: 30903738]
71. Chen J & Wagner EJ snRNA 3′ end formation: the dawn of the integrator complex. Biochem. Soc. Trans 38, 1082–1087 (2010). [PubMed: 20659008]
72. Boeing S et al. Multiomic analysis of the UV-induced DNA damage response. Cell Rep 15, 1597–1610 (2016). [PubMed: 27184836]
73. Luck K et al. A reference map of the human binary protein interactome. Nature 580, 402–408 (2020). [PubMed: 32296183]
74. Shifrut E et al. Genome-wide CRISPR screens in primary human T cells reveal key regulators of immune function. Cell 175, 1958–1971.e15 (2018). [PubMed: 30449619]
75. Povey S et al. The HUGO gene nomenclature committee (HGNC). Hum. Genet 109, 678–680 (2001). [PubMed: 11810281]
76. Collard F et al. A conserved phosphatase destroys toxic glycolytic side products in mammals and yeast. Nat. Chem. Biol 12, 601–607 (2016). [PubMed: 27294321]
77. Braverman N et al. Human PEX7 encodes the peroxisomal PTS2 receptor and is responsible for rhizomelic chondrodysplasia punctata. Nat. Genet 15, 369–376 (1997). [PubMed: 9090381]
78. Doench JG et al. Optimized sgRNA design to maximize activity and minimize off-target effects of CRISPR-Cas9. Nat. Biotechnol 34, 184–191 (2016). [PubMed: 26780180]
79. Ashburner M et al. Gene ontology: tool for the unification of biology. The Gene Ontology Consortium. Nat. Genet 25, 25–29 (2000). [PubMed: 10802651]
80. The Gene Ontology Consortium. Expansion of the gene ontology knowledgebase and resources. Nucleic Acids Res. 45, D331–D338 (2017). [PubMed: 27899567]
81. Yu J et al. A unified mixed-model method for association mapping that accounts for multiple levels of relatedness. Nat. Genet 38, 203–208 (2006). [PubMed: 16380716]
82. Storey JD & Tibshirani R Statistical significance for genomewide studies. Proc. Natl Acad. Sci. USA 100, 9440–9445 (2003). [PubMed: 12883005]
83. Liu Y & Xie J Cauchy combination test: a powerful test with analytic p-value calculation under arbitrary dependency structures. J. Am. Statist. Assoc 10.1080/01621459.2018.1554485 (2019).
84. Liu Y et al. ACAT: a fast and powerful *p* value combination method for rare-variant analysis in sequencing studies. Am. J. Hum. Genet 104, 410–421 (2019). [PubMed: 30849328]

85. Haghverdi L, Buettner F & Theis FJ Diffusion maps for high-dimensional single-cell analysis of differentiation data. Bioinformatics 31, 2989–2998 (2015). [PubMed: 26002886]

86. Perez-Riverol Y et al. The PRIDE database and related tools and resources in 2019: improving support for quantification data. Nucleic Acids Res. 47, D442–D450 (2019). [PubMed: 30395289]

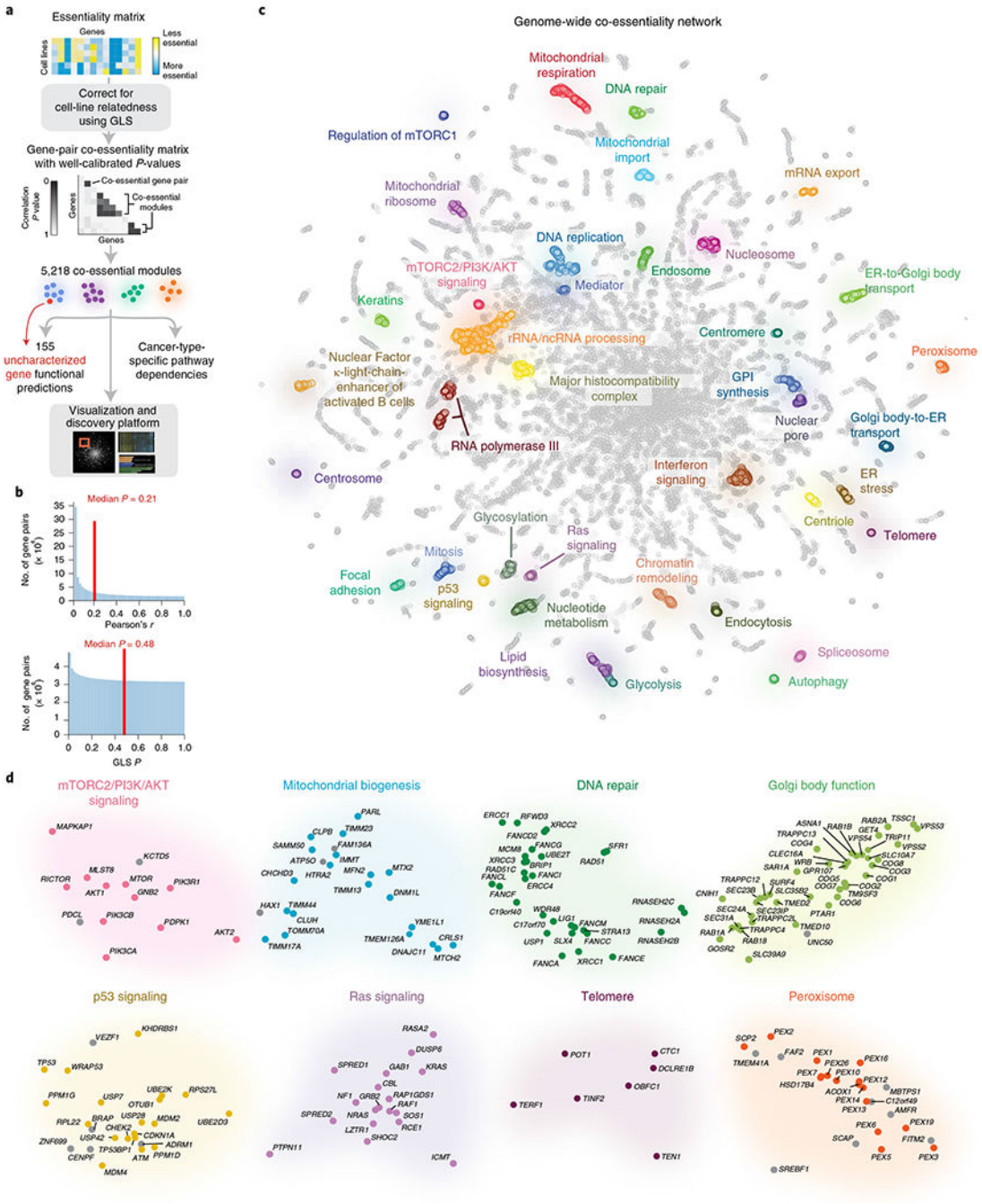



**Fig. 1 |. Construction of a genome-wide co-essentiality network.**

**a**, Overview of our approach. ER, endoplasmic reticulum; ncRNA, noncoding RNA; NFκB, nuclear factor κ-light-chain-enhancer of activated B cells; GPI, glycosylphosphatidylinositol. **b**, Histograms of GLS and Pearson's correlations across all pairs of genes. **c**, Global structure of the co-essentiality network, with manually annotated 'neighborhoods' highly enriched for particular pathways and complexes. **d**, Selected neighborhoods with manually defined known pathway members indicated in color and other genes in gray.

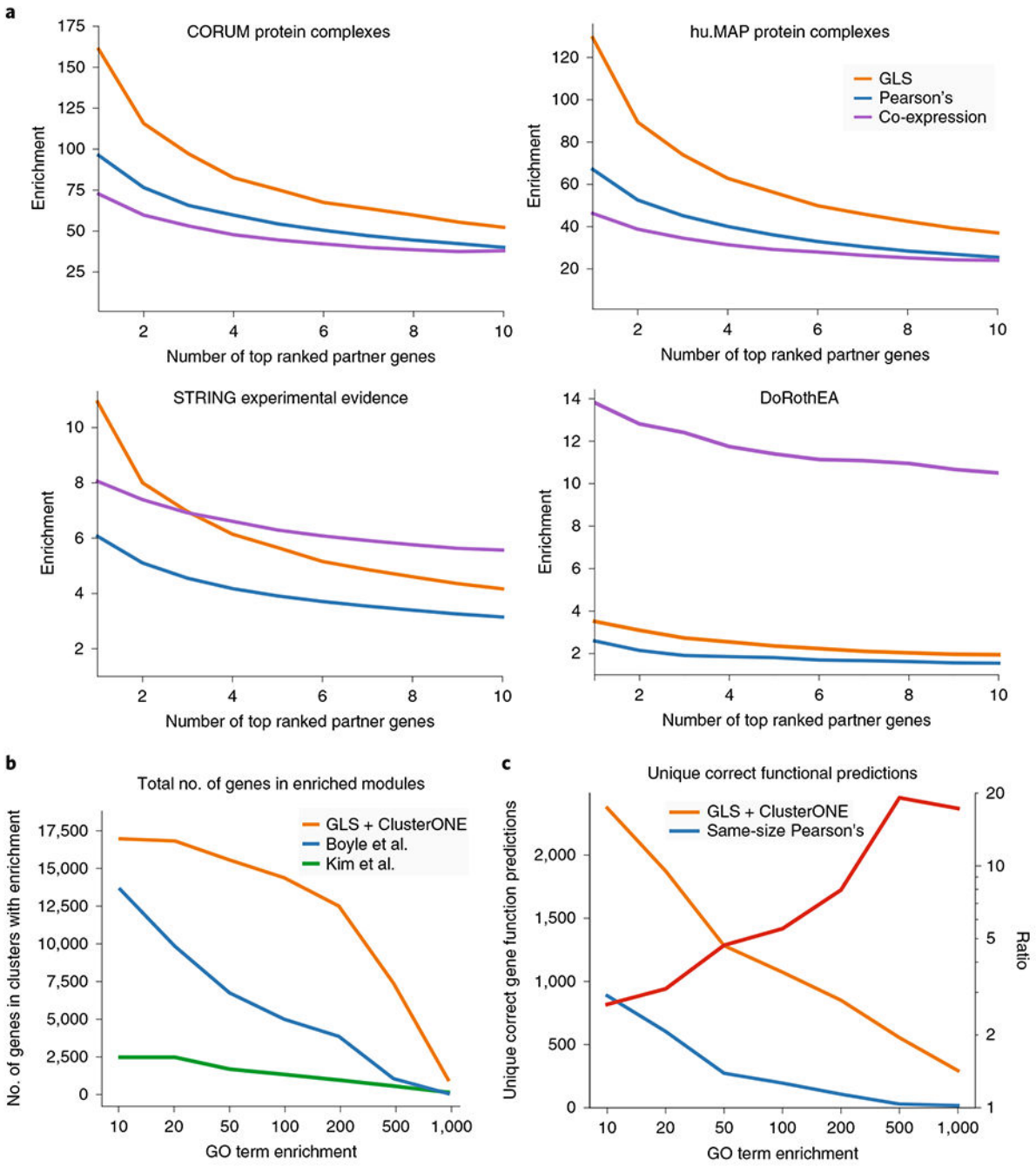



**Fig. 2 |. GLS improves recall of known functional interactions in co-essential gene pairs and modules.**

**a**, Enrichment of interactions from GLS- and Pearson's correlation-based co-essentiality using the DepMap dataset, as well as co-expression using the COXPRESdb dataset, in CORUM, hu.MAP, STRING and DoRothEA, considering the top 1–10 partners per gene. **b**, Number of genes in nonsyntenic clusters/modules at least *N*-fold enriched for some GO term with at least five total genes present across all clusters/modules, excluding the gene itself from the enrichment calculation, for various *N* values from 10 to 1,000. **c**, Number of

genes for which correct GO term-based functional predictions are made only by co-essential modules (‘GLS + ClusterONE’) or only by same-size Pearson’s modules across GO term enrichment thresholds, and the ratio (red line) of the number of genes uniquely correctly predicted by co-essential modules to the number of genes uniquely correctly predicted by same-size Pearson’s modules.

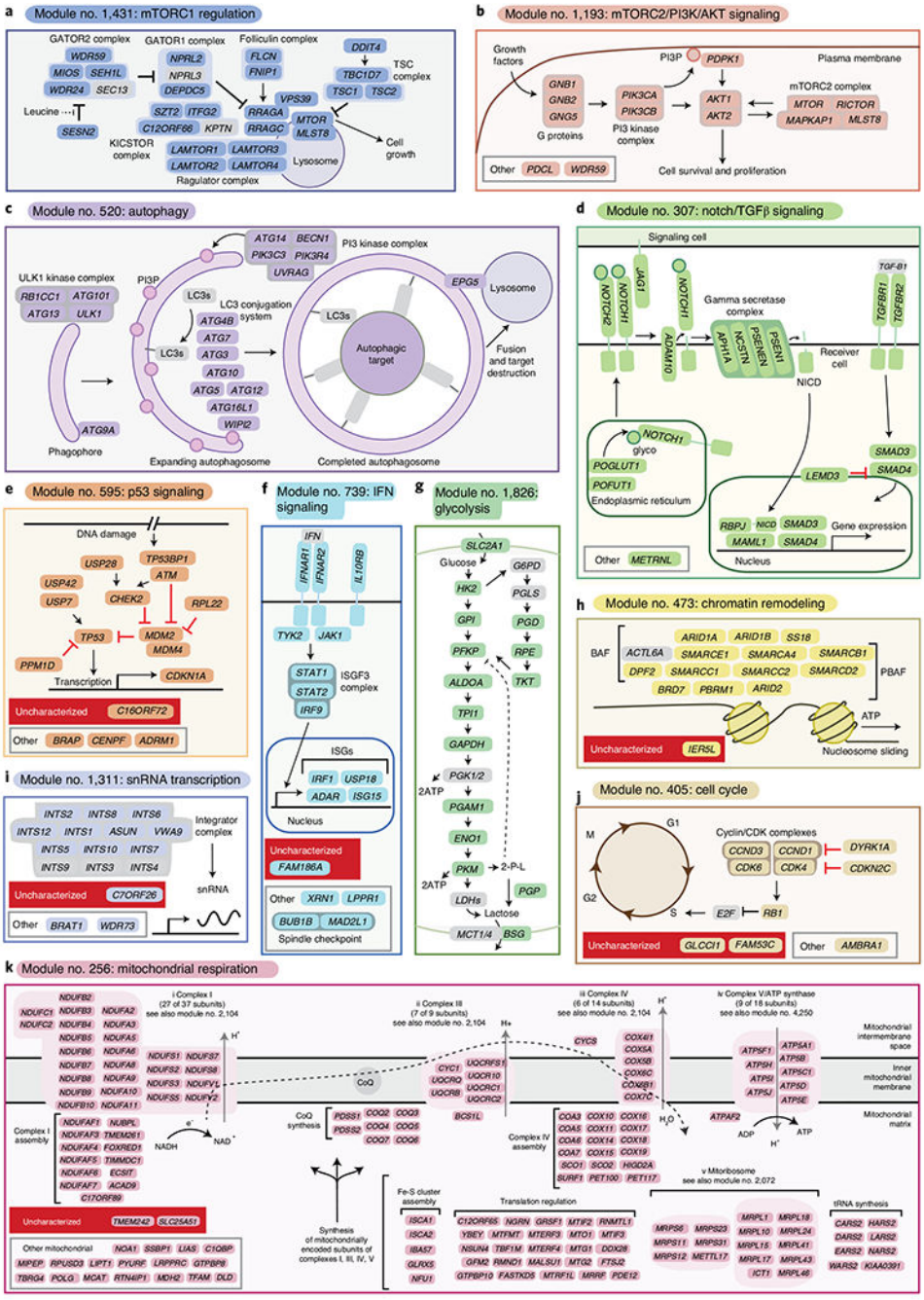



**Fig. 3 |. Co-essential modules recapitulate known pathways and nominate new pathway members.**

**a**–**k**, Ten examples of co-essential modules. All genes in each module are shown. Genes without previous evidence of pathway involvement are indicated as either ‘uncharacterized’ (UniProt annotation score <3) or ‘other’. Red inhibitory arrows between gene pairs indicate both negative regulation and negatively correlated essentiality profiles. In **a**, **c**, **g**, **i** and **j**, core pathway members not included in the module are shown in gray. Subunit counts for mitochondrial respiration complexes were based on HUGO Gene Nomenclature Committee

gene sets as of September 2020 (ref. [75]). **b**,**c**, PI3P, phosphatidyl-inositol-3-phosphate. **c**, LC3s, microtubule-associated 1A/1B-light chain (LC3) family members. **d**, glyco, fucose and glucose modifications transferred to NOTCH1 by POFUT1 and POGLUT1; NICD, notch intracellular domain; TGF-β1, transforming growth factor β1. **f**, IFN, interferon; ISGs, interferon-stimulated genes. **g**, 2-P-L, 2-phospholactate (toxic byproduct of pyruvate kinase M1/M2 (PKM))[76]. **h**, BAF, BRG- or HBRM-associated factors complex; PBAF, poly(bromo-BAF) complex. **k**, CoQ, coenzyme Q.

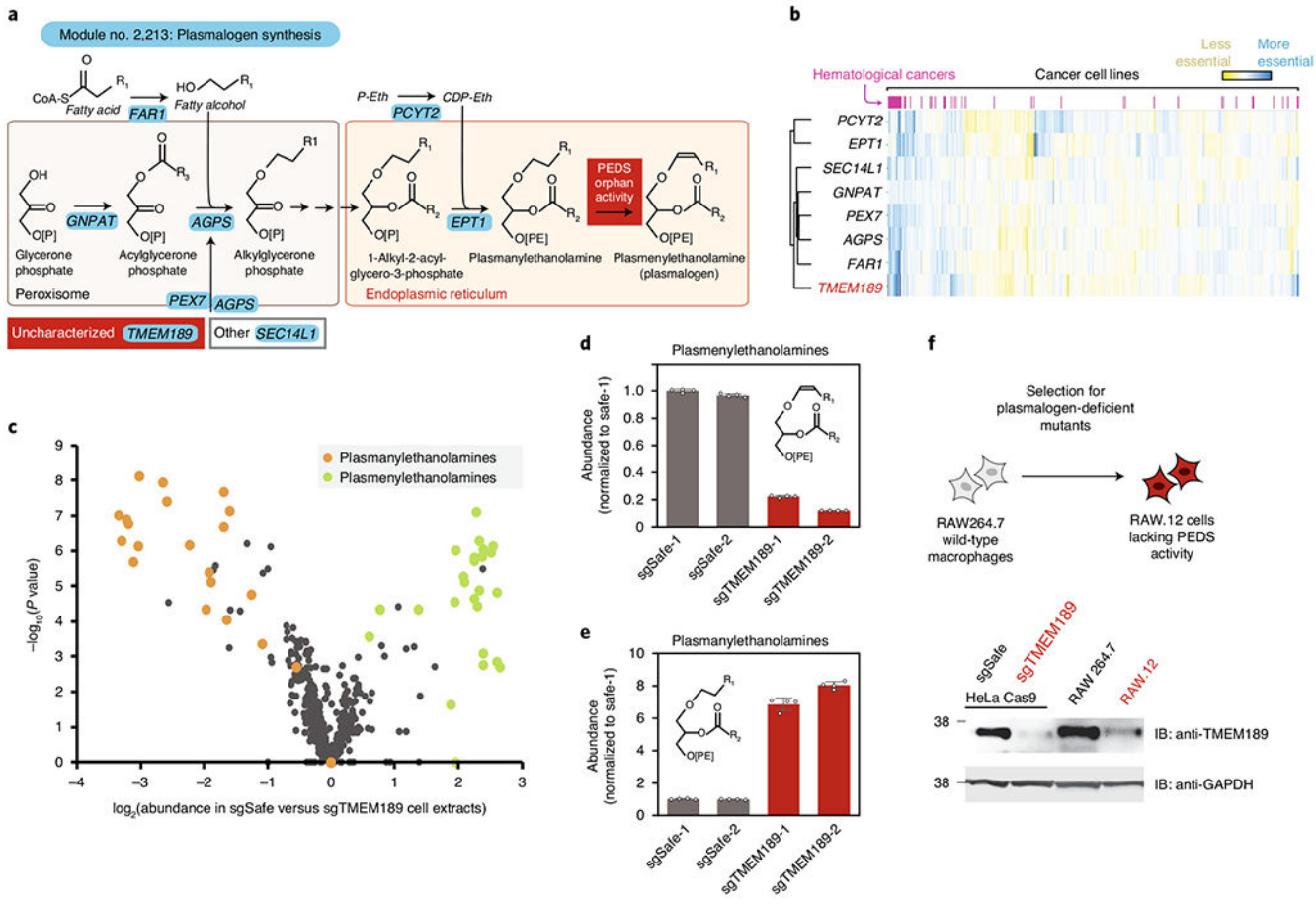



**Fig. 4 |. *TMEM189* encodes the enzyme PEDS required for synthesis of plasmalogen lipids.**

**a**, Schematic of module no. 2,213 with manual annotations of gene function. Uncharacterized gene selected for validation is shown in red box. PEX7 is shown importing cytosolic alkylglyceronephosphate synthase across the peroxisomal membrane into the peroxisome matrix[77]. PEDS enzymatic activity is indicated in red. CDP-Eth, cytidine diphosphate ethanolamine; P-Eth, phosphoethanolamine. **b**, Heatmap of bias-corrected essentiality scores of genes in module 2,213 in 485 cancer cell lines. **c**, Volcano plot of all lipid species detected in lipidomic experiment, with ratio of lipid abundance in extracts derived from sgSafe-1-expressing cells relative to sgTMEM189-1-expressing cells plotted on the *x* axis. **d**, Total abundance (relative to Safe-targeting sgRNA control no. 1) of 37 unambiguously identified plasmenylethanolamine species in cell extracts prepared from HeLa cells transduced with indicated sgRNAs. The error bars represent the s.d. ($n = 4$ cell extracts). Data are presented as mean ± s.d. **e**, Total abundance (relative to Safe-targeting sgRNA control no. 1) of 30 unambiguously identified plasmanylethanolamine species in cell extracts prepared from HeLa cells transduced with indicated sgRNAs. The error bars represent the s.d. ($n = 4$ cell extracts). Data are presented as mean ± s.d. **f**, Top: schematic of generation of RAW.12 derivative of RAW264.7 macrophage-like line with confirmed deficiency in PEDS activity, as reported in Zoeller et al.[53]. Bottom: western blotting (IB) with anti-TMEM189 antibodies of extracts derived from HeLa-Cas9 cells expressing sgSafe

or sgTMEM189, and from RAW264.7 parental line and RAW.12 (PEDS deficient) line. Western blots show representative data from experiments performed three times.

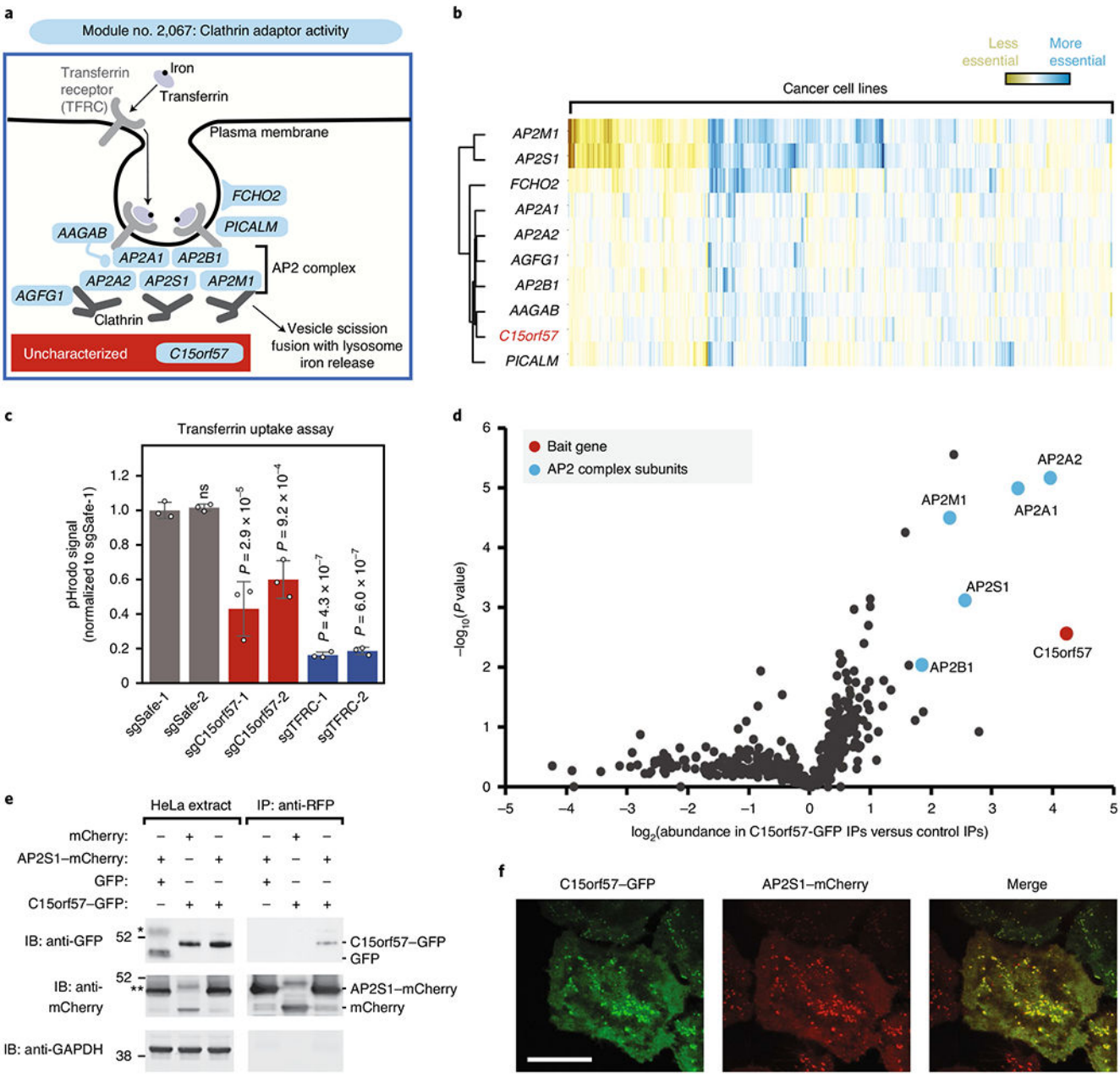



**Fig. 5 |. *C15orf57* is required for efficient clathrin-mediated endocytosis of transferrin.**

**a**, Schematic of module no. 2,067. Uncharacterized gene selected for validation is shown in red. **b**, Heatmap of bias-corrected essentiality scores of genes in module no. 2,067 in 485 cancer cell lines. **c**, Transferrin–pHrodo uptake assay for clathrin-mediated endocytosis (24-h timepoint). Data are presented as mean ± s.d. ($n = 3$ replicate wells, two-tailed Student's $t$-test). The data shown represent three independent experiments. **d**, Volcano plot of mass spectrometric (tandem mass tag) analysis of C15orf57–GFP IPs. **e**, Extracts prepared from indicated HeLa cell extracts were subjected to immunoprecipitation with anti-RFP magnetic resin. Extracts and IP samples were resolved by sodium dodecylsulfate–polyacrylamide gel electrophoresis followed by western blotting with indicated antibodies.

*GFP-specific species; **mCherry-specific species. Data represent two western blots from one experiment. **f**, Microscopy of HeLa cells transduced with C15orf57–GFP and AP2S1–mCherry constructs. Images show data representing two experiments. Scale bar, 20 μm.

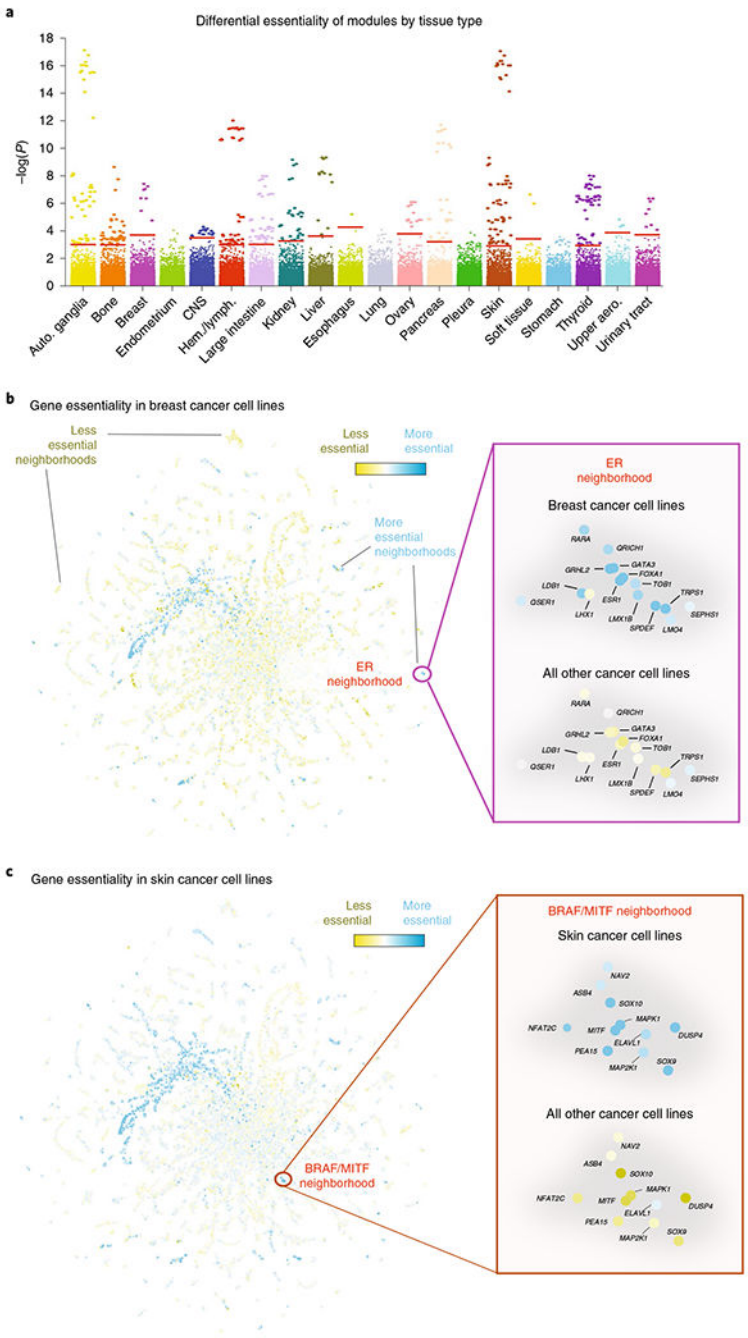



**Fig. 6 |. Identification of cancer-type-specific module dependencies.**

**a**, Differential essentiality of co-essential modules in cell lines derived from 20 tissue types. The $-\log_{10}(P \text{ values})$ for each module are plotted for each tissue (Methods). Red bars indicate FDR thresholds for each tissue type. aero., aerodigestive; Auto., autonomic; CNS, central nervous system; Hem., hematological; lymph., lymphoma. **b**, Average bias-corrected gene essentiality in breast cancer cell lines plotted on 2D co-essentiality network, with the gene neighborhood containing *ESR1* highlighted on the right. **c**, Average bias-corrected

gene essentiality in skin cancer cell lines plotted on a 2D co-essentiality network, with the gene neighborhood containing *BRAF/MITF*-pathway genes highlighted on the right.