Genome-scale analysis identifies paralog lethality as a vulnerability of chromosome 1p loss in cancer

Srinivas R. Viswanathan1,2,3, Marina F. Nogueira1,2,12, Colin G. Buss4,5,12, John M. Krill-Burger2, Mathias J. Wawer6, Edyta Malolepsza2,7, Ashton C. Berger1,2, Peter S. Choi1,2,3, Julianni Shih2, Alison M. Taylor1,2,3, Benjamin Tanenbaum5,2, Chandra Sekhar Pedamallu1, Andrew D. Cherniack2,4,5,12, John M. Krill-Burger2, William C. Hahn1,2,3⁎, Pablo Tamayo2,8, Craig A. Stratthdee2, Kasper Lage2,7, Steven A. Carr2, Monica Schenone2, Sangeeta N. Bhatia2,3,4,5,9,10,11, Francisca Vazquez2, Aviad Tsherniak2, William C. Hahn1,2,3 and Matthew Meyerson1,2,3⁎

Functional redundancy shared by paralog genes may afford protection against genetic perturbations, but it can also result in genetic vulnerabilities due to mutual interdependency1–5. Here, we surveyed genome-scale short hairpin RNA and CRISPR screening data on hundreds of cancer cell lines and identified MAGOH and MAGOHB, core members of the splicing-dependent exon junction complex, as top-ranked paralog dependencies. MAGOH is the top gene dependency in cells with hemizygous MAGOH deletion, a pervasive genetic event that frequently occurs due to chromosome 1p loss. Inhibition of MAGOH in a MAGOH-deleted context compromises viability by globally perturbing alternative splicing and RNA surveillance. Dependency on IPO13, an importin-β receptor that mediates nuclear export of the MAGOH/B-Y14 heterodimer6, is highly correlated with dependency on both MAGOH and MAGOHB. Both MAGOHB and IPO13 represent dependencies in murine xenografts with hemizygous MAGOH deletion. Our results identify MAGOH and MAGOHB as reciprocal paralog dependencies across cancer types and suggest a rationale for targeting the MAGOHB-IPO13 axis in cancers with chromosome 1p deletion.

The systematic integration of data from genomic characterization and genetic screening of cancer cell lines can identify gene dependencies induced by specific somatic alterations and inform the development of targeted therapeutics. For example, several studies have shown that inactivation of specific driver or passenger genes may confer dependency on functionally redundant paralogs2,3,10–13. Paralog dependencies have also emerged as important in recent genome-scale functional genomic screens2,3,10–13, underscoring the importance of further characterizing this class of cancer vulnerabilities.

To systematically identify paralog dependencies that may represent attractive cancer targets, we analyzed data from pooled, genome-scale short hairpin RNA (shRNA) screening of 501 cancer cell lines3,4,14. We determined the correlation between a dependency on a gene5 and loss of function of its paralog across 10,287 paralog pairs (Supplementary Fig. 1; Supplementary Note). We identified 167 genes for which dependency was significantly correlated with loss of a paralog (1.6% of paralog test pairs at q <0.05), including many previously reported paralog dependencies (for example, ARID1B dependency with ARID1A inactivation9, SMARCA2 dependency with SMARCA4 inactivation11, UBC dependency with UBD inactivation12, and FERMT1 dependency with FERMT2 inactivation13). However, of these 167 paralog dependency pairs, only 7 were ‘symmetric’, in which dependency for each of the genes in the pair was significantly correlated with inactivation of its partner paralog (Fig. 1a,b; Supplementary Table 1). A similar analysis of data from genome-scale CRISPR screening of 341 cell lines identified 125 significant paralog dependencies (1.4% of paralog test pairs at q <0.05), of which 7 pairs were symmetric (Supplementary Table 2; Supplementary Note). Paralog genes arise via ancestral duplication events and may functionally diverge over time14–16. Symmetric paralog pairs likely share complete functional redundancy, making them particularly attractive targets for ‘collateral lethality’ strategies. An enrichment for RNA-splicing related genes was noted among symmetric, but not asymmetric, paralog pairs in the shRNA and CRISPR screening datasets (Supplementary Table 3), suggesting that redundant essentiality may be exploited to target splicing-related pathways.

One symmetric paralog pair was shared between the shRNA and CRISPR datasets: MAGOH-MAGOHB; a second pair, FUBP1-KHSRP, was highly significant for symmetry in the shRNA data and borderline significant in the CRISPR dataset (q = 0.0547) (Fig. 1a,b; Supplementary Fig. 1; Supplementary Tables 1 and 2)15. We focus here on validation of the former pair. MAGOH and MAGOHB encode core members of the exon–junction complex (EJC), a multiprotein complex that is deposited on messenger RNAs at the time of splicing and that mediates diverse downstream processes including mRNA transport, stability, and nonsense-mediated decay (NMD)17,18.
Using both shRNA and CRISPR technologies, we individually validated MAGOH/B dependency in the setting of MAGOH loss, as well as MAGOH dependency in the setting of MAGOH loss. Furthermore, in a cell line without hemizygous deletion of either paralog, knockdown of either MAGOH or MAGOHB individually was tolerated, but the combination was lethal (Supplementary Fig. 2). We noted that MAGOH/B dependency in the setting of MAGOH inactivation was particularly pronounced based on (1) effect size (log-fold difference in MAGOH/B dependency between MAGOH-inactivated and non-MAGOH-inactivated cell lines) and (2) MAGOH/B scoring as a robust $\sigma$ differential dependency (having a dependency score in some cell lines greater than six standard deviations below its mean dependency score across all cell lines) in both the RNA interference (RNAi) and CRISPR screening data. We therefore sought to further characterize MAGOH/B dependency in the setting of MAGOH loss.

MAGOH/B was the top differential dependency in cells with hemizygous deletion of MAGOH (Fig. 1c; Supplementary Tables 4 and 5; Supplementary Note) and dependency on MAGOH/B was predicted by low expression of MAGOH, consistent with the notion that hemizygous deletion of MAGOH leads to its decreased expression (Supplementary Fig. 3). shRNA-mediated knockdown of MAGOH/B led to a decrease in cell viability and colony-forming capacity in three MAGOH-deleted cell lines, but not in control cell lines euploid for MAGOH (Fig. 1d,e; Supplementary Fig. 4). Ectopic expression of MAGOH/B in an MAGOH/B-deleted cell line fully rescued MAGOH/B dependency, indicating that MAGOH/B dependency in MAGOH-deleted cells is solely due to MAGOH loss, and consistent with complete functional redundancy between these paralogs (Fig. 1f; Supplementary Fig. 4). CRISPR/Cas9-mediated deletion of MAGOH/B in a cell line with two copies of MAGOH/B also conferred MAGOH/B dependency (Supplementary Figs. 5 and 6).

To assess the clinical contexts in which these dependencies might be exploited, we next surveyed the frequency of MAGOH and MAGOH/B loss in tumor cohorts from The Cancer Genome Atlas (TCGA). We observed pervasive hemizygous MAGOH loss across tumor types (frequency of 21% (1,675 of 8,009) in the entire TCGA dataset, and >50% in multiple tumor types). Moreover, MAGOH/B deletion most frequently occurs as a result of arm-level deletion of chromosome 1p across human tumors (Fig. 1g; Supplementary Table 6). We confirmed that chromosome 1p-deletion status correlates with MAGOH/B dependency in the genome-scale CRISPR screening data (Supplementary Fig. 7). In the context of neuroblastoma—a 1p deletion is a hallmark event in a subset of tumors—MAGOH/B knockdown was lethal in a 1p-deleted, but not a 1p-neutral, cell line (Supplementary Fig. 7). MAGOH/B is located on chromosome 12p, an arm also recurrently lost across tumor types, albeit with markedly lower frequency than chromosome 1p (Supplementary Fig. 8). Analysis of genome-scale CRISPR screening data confirmed a reciprocal dependency on MAGOH/B in the setting of chromosome 12p deletion. Interestingly, we also observed mutual exclusivity between chromosome 1p and chromosome 12p co-deletion in many tumor types, suggesting that concurrent loss of both MAGOH and MAGOH/B may be poorly tolerated (Supplementary Fig. 8). We conclude that MAGOH and MAGOH/B represent potential vulnerabilities in large, genetically defined subsets of tumors.

MAGOH and MAGOH/B constitute core components of the EJC; EJC deposition at exon–exon junctions allows transcripts containing premature termination codons to be identified and targeted for degradation via NMD.3,4 We therefore hypothesized that MAGOH/B inhibition in the setting of decreased MAGOH dosage may compromise cell viability by perturbing RNA splicing and RNA surveillance. To evaluate the global transcriptomic consequences of MAGOH/B inhibition, we performed RNA sequencing on hemizygous MAGOH/B-deleted ChagoK1 cells in the presence or absence of MAGOH/B knockdown, with or without ectopic re-expression of MAGOH. We observed an increased expression of NMD biotype transcripts on MAGOH/B knockdown in ChagoK1 cells (Fig. 2a, left). In contrast, MAGOH/B knockdown in MAGOH-reconstituted ChagoK1 cells was well tolerated without a notable shift in NMD biotype transcript distribution (Fig. 2a, right). We next sought to determine whether the upregulation of NMD isoforms on MAGOH/B knockdown in ChagoK1 cells was occurring at the expense of other transcript biotypes. Among genes that had significantly upregulated NMD isoform(s) on MAGOH/B knockdown, we observed a significant proportional decrease in coding isoform expression in ChagoK1 cells but not MAGOH/B-reconstituted ChagoK1 cells (Fig. 2b, compare left and right). To investigate whether particular splice event classes were driving this redistribution of isoform types, we quantified the proportion of differentially spliced events of each class that were more common in either the absence (Fig. 2c, red) or presence (Fig. 2c, blue) of MAGOH/B knockdown in either ChagoK1 cells or MAGOH/B-reconstituted ChagoK1 cells. As compared with MAGOH/B knockdown in MAGOH/B-reconstituted ChagoK1 cells, MAGOH/B knockdown in ChagoK1 cells resulted in reduced cassette exon inclusion and increased intron retention (Fig. 2c).

Therefore, many global transcriptomic effects of MAGOH/B insufficiency appear attributable to alterations in these two splice event types, indicative of a defect in exon definition/recognition.

We identified 22 instances in which there was both a significant absolute upregulation of an NMD isoform (beta > 1 in differential expression analysis using Kallisto5) and corresponding downregulation of at least one protein coding isoform (beta < -1) (Supplementary Table 7). These genes were significantly enriched for pathways involved in mRNA splicing and mRNA processing (Supplementary Table 8; Fig. 2d). Notably, among the seven splicing-related genes driving this enrichment were four genes (SRSF2, SRSF7, HNRNPD, HNRNPH1) reported to auto-regulate their expression via alternative splicing-NMD (AS-NMD) loops6,7. Such AS-NMD loops, many of which are in splicing-related genes, involve ultraconserved, regulated alternative splicing events that induce NMD substrates, thus maintaining homeostatic control of gene expression8,9,10. We observed perturbations in isoform distributions of HNRNPD and other splicing-related genes on MAGOH/B knockdown in ChagoK1 cells, but not in MAGOH/B-reconstituted ChagoK1 cells (Fig. 2e; Supplementary Figs. 9 and 10; Supplementary Note). Altered RNA isoform abundance accompanied by changes in the levels of functional protein, either via disruption of AS-NMD loops or through other mechanisms, could have deleterious direct and indirect consequences on cellular splicing.

Given these transcriptomic consequences of MAGOH/MAGOH/B insufficiency, we next sought to determine whether MAGOH loss unveils a broader dependency on splicing/NMD-related complexes. We performed immunoprecipitation and mass spectrometry to identify MAGOH- and MAGOH/B-associated binding partners and found that these interactors, which include many splicing-related genes, were enriched for gene dependencies correlated with both MAGOH and MAGOH/B dependencies. However, these dependencies were weaker than the reciprocal MAGOH/B paralog dependencies driven by redundant essentiality (Supplementary Fig. 11; Supplementary Tables 9–11; Supplementary Note).

MAGOH and MAGOH/B share near-identity at the protein level and functional and crystallographic studies do not necessarily show domains easily amenable to targeting by small molecules. To identify other more tractable targets that might indirectly affect MAGOH/MAGOH/B function, we interrogated the genome-scale shRNA screening data for gene dependencies highly correlated with either MAGOH or MAGOH/B dependency. IPO13 emerged as the top, outlier-correlated gene dependency to both MAGOH and MAGOH/B (Fig. 3a; Supplementary Table 12). IPO13 is a bidirectional karyopherin responsible for nuclear import of the...
Fig. 1 | Hemizygous MAGOH deletion confers MAGOHβ dependency. a. Analysis of paralog dependencies in genome-scale screening of cancer cell lines (shRNA, 501 lines; CRISPR-Cas9, 341 lines). b. q-value/q-value plot showing significance of pairwise correlation between a gene's dependency score and inactivation of its paralog. q-value 1, significance for dependency on the paralog labeled first with inactivation of the paralog labeled second. q-value 2, significance for dependency on the paralog labeled second with inactivation of the paralog labeled first. ‘Symmetric’ paralogs are in the upper right quadrant (q1 < 0.05 and q2 < 0.05). Plots show n = 1,970 paralog pairs for shRNA data and n = 1,593 pairs for CRISPR data. One-sided P value from two-class comparison was calculated via moderated t-statistic and adjusted for multiple comparisons using the Benjamini–Hochberg FDR. c. PARIS analysis to identify gene dependencies correlated with hemizygous MAGOH loss. Mutual information metric (RNMI) is plotted against FDR for gene dependencies identified in genome-scale screening of cancer cell lines. d. Cell viability in cell lines with (left) and without (right) hemizygous MAGOH loss upon MAGOH suppression using a doxycycline-inducible shRNA against MAGOHβ. Error bars show mean ± s.d., n = 3 replicates from a representative experiment repeated at least twice in each cell line; P value by two-tailed, two-sample t-test. Dox, doxycycline; NS, not significant. e. Colony formation in cell lines with (H1437, H460) or without (H1373) hemizygous MAGOH loss upon MAGOH suppression using a doxycycline-inducible shRNA against MAGOHβ. Photographs show representative wells from an experiment conducted in triplicate (quantification in Supplementary Fig. 4); experiment was repeated at least twice in each cell line. f. Cell viability measured on MAGOHβ knockdown in ChangK1 cells with or without reconstitution of MAGOH-VS. Error bars show mean ± s.d., n = 5 replicates from a representative experiment repeated at least twice; P value by two-tailed, two-sample t-test. g. Frequency of hemizygous MAGOH deletion across TCGA cohorts. Total frequency of MAGOH loss is indicated by a light blue bar; frequency of MAGOH loss occurring as a result of chromosome 1p deletion is indicated by a dark blue bar. Top panel shows total number of arm-level copy number events in each tumor type.
**RNA splicing is globally altered upon MAGOH loss, leading to the upregulation of NMD substrates.**

**a,** Differentially expressed transcripts in ChagoK1 cells (left) and in MAGOH-V5 reconstituted ChagoK1 cells (right) on MAGOH knockdown (KD). Transcripts annotated as NMD substrates shown in red. Significance determined by a Wald test and adjusted for multiple comparisons using the Benjamini-Hochberg FDR. \( n = 3 \) replicates were used in all conditions. **b,** Density distribution of proportional expression levels among coding isoforms corresponding to genes whose NMD isoforms are upregulated upon MAGOH knockdown in ChagoK1 cells (left) or MAGOH-V5 reconstituted ChagoK1 cells (right). X axis shows expression level of coding isoform(s) proportional to all expressed transcripts for a given gene. Y axis shows density. **c,** Global changes in patterns of splice site usage upon MAGOH knockdown in ChagoK1 cells (left) or MAGOH-V5 reconstituted ChagoK1 cells (right). Splice event classes are shown in the schematic at top; solid lines denote ‘inclusion’ event and dotted lines denote the ‘alternative’ event for each class. Bar graphs denote the proportion of significant events.

**d,** Significantly enriched gene ontology (GO) classes for genes (\( n = 17 \)) that show upregulation of NMD isoform(s) and concomitant downregulation of coding isoform(s). Significance was determined by a binomial test and adjusted for multiple comparisons using the Bonferroni correction. **e,** Left panel shows Sashimi plots around an activated exon within the 3'UTR of the HNRNPDL gene (inclusion of which creates an NMD-substrate transcript) in either the absence or presence of MAGOH knockdown in either ChagoK1 cells or MAGOH-V5 reconstituted ChagoK1 cells. Numbers reflect junction spanning reads averaged over three replicates for each condition. Right panel, top left: transcript abundances for various isoforms of the HNRNPDL gene (labeled by the last six digits of the ENSEMBL ID) in either the absence (gray) or presence (red) of MAGOH knockdown in either ChagoK1 cells or MAGOH-V5 reconstituted ChagoK1 cells. Right panel, top right: isoform abundances grouped by predicted coding protein length in each condition. aa, amino acid; tpm, transcripts per million. Right panel, bottom: western blot showing increased HNRNPDL protein levels on MAGOH knockdown in ChagoK1 cells but not MAGOH-V5 reconstituted ChagoK1 cells. Representative of a similar experiment repeated three times.
MAGOHB-Y14 heterodimer, a function critical for recycling of the EJC; it is also located on chromosome 1p in proximity to MAGOH, and the two genes are frequently codeleted23 (Fig. 3b; Supplementary Table 13).

We observed a selective IPO13 dependency in MAGOH-deleted cell lines compared to non-deleted cell lines (Fig. 3c) and found that dependency on IPO13 in MAGOH-deleted H460 and H1437 cells was partially attenuated by MAGOH re-expression (Fig. 3d; Supplementary Fig. 12; Supplementary Note). Knockdown of IPO13 in MAGOH-deleted cells led to cytoplasmic accumulation of MAGOH/MAGOHB and subsequent upregulation of the NMD substrates SC1.6 and SC1.7, an effect that was rescued by MAGOH re-expression (Fig. 3e; Supplementary Fig. 13). This suggests that IPO13 dependency in MAGOH- and MAGOHB-deleted cells is mediated in part by defective shuttling of MAGOH/B, resulting in mis-splicing and impaired RNA surveillance. Haploinsufficiency of IPO13, as occurs when MAGOH and IPO13 are codeleted, may also contribute to IPO13 dependency in some contexts.

Finally, we sought to validate MAGOHB as a target in vivo. We formed xenografts from H1437 cells (which carry a hemizygous deletion in MAGOH/B) transduced with a lentiviral vector encoding a doxycycline-inducible shRNA against MAGOH/B. MAGOHB knockdown was significantly impaired on intratumoral injection of si-MAGOHB and si-IPO13-containing TPNCs, but not TPNCs containing control siRNA against GFP (Fig. 4d,e; Supplementary Fig. 14). This finding was confirmed using a second TNPC system using a distinct tumor-homing peptide, Lyp-1 (Supplementary Fig. 14). Additionally, tumors treated with si-MAGOHB-containing TPNCs displayed higher levels of cleaved caspase-3, indicating that targeting MAGOH/B in a MAGOH-hemizygous context triggers apoptotic cell death (Supplementary Fig. 14). Thus, MAGOHB and IPO13 represent potential in vivo targets in a MAGOH-deleted context, and this paralog vulnerability may be exploited by antisense or RNAi-based approaches.

Hemizygous chromosome arm loss is one of the commonest features of cancer genomes24,25 and rational therapeutic targeting of this class of somatic events would therefore be attractive. Prior studies have identified several candidate targets unmasked by genomic loss26–28. Here, we integrate genomic characterization and genome-scale functional screening of cancer cell lines to systematically extend such studies. We identify a set of robust paralog dependencies that may provide the foundation for future target validation efforts and show that hemizygous loss of the MAGOH gene on chromosome 1p confers novel vulnerabilities on MAGOHB and IPO13, perhaps due to decreased nuclear reserve of MAGOH/MAGOHB (Supplementary Fig. 15). Insufficient MAGOH/MAGOHB dosage perturbs splicing and RNA surveillance and adds to growing evidence implicating splicing as a cancer dependency27,29. Therapeutic approaches to targeting MAGOH-deleted cells may involve either direct MAGOHB

Fig. 3 | IPO13 dependency is correlated with MAGOH and MAGOHB dependencies and is rescued by MAGOH reconstitution. a, Plot of gene dependencies (n = 6,300) correlated with MAGOH dependency versus those correlated with MAGOHB dependency. Axes reflect Z-scored Pearson correlation of each dependency with either MAGOH dependency (x axis) or MAGOHB dependency (y axis). MAGOH and MAGOHB self-correlations are not shown. b, Heatmap of IPO13 dependency scores across 243 screened cell lines. Black bars denote cell lines that share dependency on IPO13 and either MAGOH, MAGOHB, or both (1); cell lines that carry a hemizygous deletion in IPO13 (2); MAGOH (3), and MAGOHB (4). c, Cell viability measured on shRNA-mediated IPO13 knockdown in cell lines without (HCC1359, left) and with (H1437, right) MAGOH loss. Error bars show mean ± s.d., n = 4 replicates per cell line; P value by two-tailed, two-sample t-test. d, Colony formation in MAGOH-deleted H460 cells on IPO13 knockdown in either the absence (top) or presence (bottom) of MAGOH-V5 reconstitution. Photographs show representative wells from an experiment conducted in triplicate (quantification in Supplementary Fig. 12); experiment was repeated three times. e, Fold change in expression of the NMD substrates SC1.6 and SC1.7 of the SRSF2 gene in H460 cells on IPO13 knockdown (IPO3-sh2) in either the presence (red) or absence (blue) of MAGOH-V5 reconstitution. Data normalized to expression in the shGFP condition. Error bars show mean ± s.d., n = 3 technical replicates per condition.
transcript suppression (such as through antisense/RNAi approaches), targeted MAGOH protein degradation, or indirect suppression of MAGOH/MAGOHB activity via inhibition of IPO13. Antisense/RNAi-based approaches may be well suited to the exploitation of paralog dependencies, as they may allow for selective targeting of paralogs that show greater variability on the nucleotide level than on the protein level. Targeted protein degradation approaches have also recently proven to be a promising means to target conventionally ‘undruggable’ genes, including RNA splicing factors. In the case of IPO13 dependency, small-molecule inhibitors of other importin family members have been described, raising the possibility that IPO13 can be selectively targeted using a similar strategy. As hemizygous loss of chromosome 1p is extremely common across multiple tumor types, these or other approaches to targeting this pathway may have future biomarker-driven therapeutic applications. More broadly, our work can be generalized to cancers with other chromosome arm deletions and underscores the power of intersecting comprehensive molecular characterization and functional genomic studies of cancer cell lines.

URLs. Broad RNAi Consortium, http://www.broadinstitute.org/rnaipublic; LentiCRISPRv2 Cloning Protocol, http://genome-engineering.org/gecko/wp-content/uploads/2013/12/lentiCRISPRv2-and-lentiGuide-oligo-cloning-protocol.pdf; Lentiviral Production Protocol, http://portals.broadinstitute.org/ggp/public/resources/protocols; Bash script for expectation maximization algorithm, http://www.lagelab.org/resources; HUGO Gene Nomenclature Committee, http://www.genenames.org/cgi-bin/genefamilies; rMATS2Sashimiplot, https://github.com/Xinglab/rmats2sashimiplot; ENSEMBL Biomart, http://www.ensembl.org/biomart DepMap Portal, http://depmap.org; MassIVE, http://massive.ucsd.edu; NCBI GEO, https://www.ncbi.nlm.nih.gov/geo/; GO website, www.geneontology.org.

Methods

Methods, including statements of data availability and any associated accession codes and references, are available at https://doi.org/10.1038/s41588-018-0155-3.

Received: 10 October 2017; Accepted: 10 May 2018; Published online: 28 June 2018

References

1. Diss, G. et al. Gene duplication can impart fragility, not robustness, in the yeast protein interaction network. Science 355, 630–634 (2017).
2. Muller, F. L., Aquilanti, E. A. & DePinho, R. A. Collateral lethality: a new therapeutic strategy in oncology. Trends Cancer 1, 161–173 (2015).
3. Frei, E. Gene deletion: a new target for cancer chemotherapy. Lancet 342, 662–664 (1993).
4. McDonald, E. R. 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).
5. Tsherniak, A. et al. Defining a cancer dependency map. Cell 170, 564–576.e16 (2017).
6. Boehm, V. & Gehring, N. H. Exon junction complexes: supervising the gene expression assembly line. Trends Genet. 32, 724–735 (2016).
7. Chang, Y.-F., Imam, J. S. & Wilkinson, M. F. The nonsense-mediated decay RNA surveillance pathway. Annu. Rev. Biochem. 76, 51–74 (2007).
8. Singh, K. K., Wachsmuth, L., Kulozik, A. E. & Gehring, N. H. Two mammalian MAGOH genes contribute to exon junction complex composition and nonsense-mediated decay. RNA Biol. 10, 1291–1298 (2013).

9. Mingot, J.-M., Kostka, S., Kraft, R., Hartmann, E. & Gorlich, D. Importin 13: a novel mediator of nuclear import and export. EMBO J. 20, 3685–3694 (2001).

10. Helming, K. C. et al. ARID1B is a specific vulnerability in ARID1A-mutant cancers. Nat. Med. 20, 251–254 (2014).

11. Hoffman, G. R. et al. Functional epigenetics approach identifies BRM/SMARC2A as a critical synthetic lethal target in BRG1-deficient cancers. Proc. Natl Acad. Sci. USA 111, 3128–3133 (2014).

12. D’Antonio, M. et al. Recessive cancer genes engage in negative genetic interactions with their functional paralogs. Cell Rep. 5, 1519–1526 (2013).

13. Dey, P. et al. Genomic deletion of malic enzyme 2 confers collateral lethality in pancreatic cancer. Nature 542, 119–123 (2017).

14. Cowley, G. S. et al. Parallel genome-scale loss of function screens in 216 cancer cell lines for the identification of context-specific genetic dependencies. Sci. Data 1, 140035 (2014).

15. Meyers, R. Computational correction of copy-number effect improves specificity of CRISPR-Cas9 essentiality screens in cancer cells. Nat. Genet. 49, 1779–1784 (2017).

16. Veitia, R. A. Gene duplicates: agents of robustness or fragility? Trends Genet. 33, 377–379 (2017).

17. Bono, F. & Gehring, N. H. Assembly, disassembly and recycling. Science 346, 525–527 (2016).

18. Caron, H. et al. Allelic loss of chromosome 1p as a predictor of survival in patients with glioblastoma. J. Clin. Oncol. 34, 3128–3133 (2016).

19. Bray, N. L., Pimentel, H., Melsted, P. & Pachter, L. Near-optimal probabilistic explorative meta-analysis of chromosomal CGH data. Genome Biol. 14, 232 (2013).

20. Neri, J. Z. et al. Ultraconserved elements are associated with homeostatic control of splicing regulators by alternative splicing and nonsense-mediated decay. Genes Dev. 21, 708–718 (2007).

21. Lawrence, L. F., Inada, M., Green, R. E., Wengrod, J. C. & Brenner, S. E. Unproductive splicing of SR genes associated with highly conserved and ultraconserved DNA elements. Nature 446, 926–929 (2007).

22. Lau, C.-K., Diem, M. D., Dreyfuss, G. & Van Duyne, G. D. Structure of the Y14-Magoh core of the exon junction complex. Curr. Biol. 13, 933–941 (2003).

23. Ren, Y. et al. Targeted tumor-penetrating siRNA nanocomplexes for specific inhibition of Importin β1-mediated nuclear import by confocal on-bead screening of tagged one-bead one-compound libraries. ACS Chem. Biol. 5, 967–979 (2010).

24. Wagstaff, K. M., Sivakumar, H., Heaton, S. M., Harrich, D. & Jans, D. A. Ivermectin is a specific inhibitor of importin α/β-mediated nuclear import able to inhibit replication of HIV-1 and dengue virus. Biochem. J. 443, 851–856 (2012).

Acknowledgements
S.R.V. was supported by a Young Investigator Award from the American Society of Clinical Oncology. This work was supported by a National Cancer Institute grant 1R35CA197568 and an American Cancer Society Research Professorship to M.M. P.T. was supported by NIH grants U01CA217785 and R01HG009285. W.C.H. was supported by U01CA176058. C.G.B. and S.N.B. were supported by a Koch Institute Support Grant (P30-CA14051) from the National Cancer Institute (Swanson Biotechnology Center) and a Core Center Grant (P30-ES0502109) from the National Institute of Environmental Health Sciences, and the Ludwig Center for Molecular Oncology. C.G.B. was supported by the National Science Foundation Graduate Research Fellowship Program. S.N.B. is a Howard Hughes Medical Institute Investigator. P.S.C. was supported by an NIH Pathway to Independence Award (K99 CA208028). E.M. would like to thank H. Horn for help with the expectation maximization algorithm. The authors thank the Koch Institute Swanson Biotechnology Center for technical support, specifically K. Cormier in the Hope Babette Tang Histology Facility.

Author contributions
S.R.V. and M.M. conceived of the research and wrote the manuscript. S.R.V. and M.M. performed experiments. C.G.B. performed in vivo xenograft experiments. M.I.W., S.R.V., and P.S.C. performed data analysis on RNA sequencing data. J.M.K.-B. and S.R.V. performed data analysis on shRNA and CRISPR screening data. A.C.B., A.M.T., and J.S. performed copy number analysis on TCGA and cell line data. C.A.S. assisted in generation of shRNA reagents. P.T., A.D.C., and C.S.P. performed or oversaw data analysis. B.T., K.L., S.A.C., E.M., and M.S. performed mass spectrometry or were involved in downstream data analysis. S.N.B. oversaw in vivo xenograft experiments. F.V., A.T., and W.C.H. directed shRNA and CRISPR screening efforts. M.M. directed the overall project.

Competing interests
A.C.B., A.D.C., C.A.S., and M.M. receive research support from Bayer Pharmaceuticals. M.M. is a scientific advisory board member of, consultant for, and holds equity in OrgiMed. The content of this manuscript is the subject of a pending patent application (S.V., M.M.).

Additional information
Supplementary information is available for this paper at https://doi.org/10.1038/s41588-018-0155-3.

Reprints and permissions information is available at www.nature.com/reprints.

Correspondence and requests for materials should be addressed to M.M.

Publisher’s note: Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.
Methods

Cell culture. Cell line stocks used for validation experiments were obtained either from the Cancer Cell Line Encyclopedia (CCLE) repository at the Broad Institute or from M.M.'s laboratory, with original sources being either the American Type Culture Collection, the European Collection of Authenticated Cell Cultures, the Health Science Research Resources Bank, the Korean Cell Line Bank, or academic laboratories. Cell line identity was verified by either short tandem repeat profiling or Affymetrix SNP profiling. Cultures were maintained in media specified by the source and supplemented with 10% fetal bovine serum. Cell cultures, the Health Science Research Resources Bank, the Korean Cell Line Bank, or academic laboratories. Cell line identity was verified by either short tandem repeat profiling or Affymetrix SNP profiling. Cultures were maintained in media specified by the source and supplemented with 10% fetal bovine serum. For lentiviral transduction and colony formation assays, cells were seeded in 12-well plates at a density of 10,000 cells per well and cultured for 10–20 days. Colonies were fixed in 4% formaldehyde and stained with 0.5% crystal violet. Colonies were photographed using a Leica microscope. Colonies were then destained using 10% acetic acid and crystal violet staining was quantified by measuring absorbance at 595 nm using a Spectramax M5 instrument (Molecular Devices).

Immunofluorescence. For immunofluorescence assays, 200,000 cells were plated on SecureSlip silicone supported coverslips (Sigma Aldrich) in 6-well plates that had been precoated for 60 minutes with 0.1 ml−1 human fibronectin (Collaborative) in PBS. The following day, cells were fixed in 4% paraformaldehyde diluted in PBS for 15 minutes at room temperature. Cells were permeabilized with 0.2% Triton X-100 in PBS for 10 minutes. Blocking was performed in 1% nonfat dry milk, 1% penicillin, 100 μg ml−1 streptomycin, 2 mM L-glutamine, and 100 μg ml−1 Normocin (Invivogen). Mycoplasma testing was performed in source repository before creation of frozen stocks and repeated periodically if lines were persistently maintained in culture.

Lentiviral constructs and transduction of cell lines. For overexpression experiments, ORFs were expressed from within the pLX304-eGFP (Addgene no. 25890, Blastidin resistance) or pPLX304-eGFP as an overexpression negative control. Ectopic expression of untagged MAGOH was performed using either pLX304 (with stop codon introduced before V5 tag) or Gateway-compatible, hygromycin-resistant, doxycycline-inducible overexpression vector with complementary RNA expression driven from Tet-regulated cytomegalovirus promoter, created by modification of pLX304. MAGOH-reconstitution experiments were performed with V5-tagged MAGOH expressed from within the pLX304-Blast-V5 vector38 using a plentiCRISPRv2. Following 72 hours of selection with puromycin, the bulk resistant population was sorted at single-cell density into 96-well plates using a CellTracker Orange CMFDA (Thermo Fisher Scientific). Lentivirus was produced in HEK293T cells as per the ‘low throughput viral production’ protocol on the RNAi Consortium Portal (see URLs).

Cell viability and colony formation assays. For cell viability assays, cells were seeded in 96-well plates in 100 μl medium after lentiviral transduction and completion of antibiotic selection. For inducible hairpin experiments, equal numbers of cells were seeded for both ‘−Dox’ and ‘+Dox’ conditions and medium was supplemented with 100 ng ml−1 doxycycline in the ‘+Dox’ condition. Cells were seeded in 96-well plate format (range 1,000–8,000 per well, depending on the cell line). At 7–10 days after cell seeding, cell viability was assessed using the Cell Titer-Glo luminescent cell viability assay (Promega) using either an EnVision Multilabel Reader (PerkinElmer) or a Spectramax M5 plate reader (Molecular Devices).

For colony formation assays, cells were seeded in 12-well plates at a density of 10,000 cells per well and cultured for 10–20 days. Colonies were fixed in 4% formaldehyde and stained with 0.5% crystal violet. Colonies were photographed using a Leica microscope. Colonies were then destained using 10% acetic acid and crystal violet staining was quantified by measuring absorbance at 595 nm using a Spectramax M5 instrument (Molecular Devices).

Immunoprecipitation and mass spectrometry. Immunoprecipitation. For immunoprecipitation experiments, 293 T cells were either untransduced (control) or transduced with pLX304-Blast-V5 (Addgene no. 25890) expressing MAGOH-V5 or MAGOH-B-V5. Following antibiotic selection to derive stably

NATURE GENETICS | www.nature.com/naturegenetics

© 2018 Nature America Inc., part of Springer Nature. All rights reserved.
transduced cell populations, immunoprecipitation was carried out using the Pierce Class Magnetic IP Kit (no. 88804) and anti-V5 magnetic beads (MBL no. M167-11) using a starting amount of 2 mg protein and 50 μl beads. Lysis buffer was pH 8.0, 0.025 M Tris, 0.15 M NaCl, 0.001 M EDTA, 1% NP40, and 5% glycerol. Immunoprecipitation was carried out overnight at 4°C. Samples were washed twice in sample buffer, followed by twice in PBS, before mass spectrometry. Efficient immunoprecipitation was confirmed by western blotting before proceeding with mass spectrometry.

Protein digestion and labeling with tandem mass tag (TMT) isobaric mass tags. The beads from immunopurification were washed once with IP lysis buffer, then three times with PBS. The four different lysates of each replicate were resuspended in 90 μl digestion buffer (2 M urea, 50 mM Tris HCl) and then 2 μg sequencing grade trypsin was added, followed by 1 hour of shaking at 70 rpm. The supernatant was removed and placed in a fresh tube. The tubes were acidified twice with 50 μl digestion buffer and combined with the supernatant. The combined supernatants were reduced (2 μl 1 M dithiothreitol, 30 minutes, room temperature) and alkylated (4 μl 500 mM iodoacetamide, 45 minutes, dark), and a longer overnight digestion was performed: 2 μl (4 μl) trypsin, shaken overnight. The samples were then quenched with 10% formic acid on 10 mg Oasis cartridges.

Desalted peptides were labeled with TMT6 reagents lot Q218427 (Thermo Fisher Scientific) according to the following: 126, NoBaitCntRep1; 127, NoBaitCntRep2; 128, MAGOHRep1; 129, MAGOHRep2; 130, MAGOHRep3; 131, MAGOHRep4. Peptides were dissolved in 25 μl fresh 100 μM HEPES buffer. The labeling reagent was resuspended in 42 μl acetonitrile and 10 μl added to each sample as described below. After 1 hour incubation the reaction was stopped with 8 μl 5 mM hydroxylamine.

Protein identification with a nanoLC-MS system. Reconstituted peptides were separated on an online nanoflow EASY-nLC 1000 UHPLC system (Thermo Fisher Scientific) and analyzed on a benchtop Orbitrap Q Exactive Plus mass spectrometer (Thermo Fisher Scientific). The peptide samples were injected onto a capillary column (Picolfrt with 10 μm tip opening/75 μm diameter, New Objective, PF360-75-10-N-5) packed in-house with 20 cm C18 silica material (1.9 μm ReproSil-Pur C18-AQ medium, Dr. Maisch GmbH) and heated to 50°C in column heater sleeves (Phoenix ST) to reduce backpressure during UHPLC separation. Injected peptides were separated at a flow rate of 200 nl/min with a linear 230 min gradient from 100% solvent A (3% acetonitrile, 0.1% formic acid) to 50% solvent B (90% acetonitrile, 0.1% formic acid), followed by a linear 9 min gradient from 50% solvent B to 60% solvent B and a 1 min ramp to 90% B. Each sample was run for 260 min, including sample loading and column equilibration times. The Q Exactive ion trap was used in the data-dependent mode acquiring higher-energy collisional dissociation (HCD) tandem mass spectrometry (MS/MS) scans (R = 17,500) after each MS1 scan (R = 70,000) on the 12 top most abundant ions using an MS1 ion target of 3 × 10^6 ions and an MS2 target of 5 × 10^6 ions. The maximum ion utilization time for the MS/MS scans was 120 ms; the HCD-normalized collision energy was set to 27%; the dynamic exclusion time was set to 20 s; and the peptide match and isotope exclusion functions were enabled.

Database search and data processing. All mass spectra were processed using the Spectrum Mill software package v6.0 prerelease (Agilent Technologies), which included modules for database searching. Initially, results were filtered using the Scaffold program (Proteome Software) and searched against the human Uniprot database to which a set of common laboratory contaminant proteins was appended. Search parameters included ESI-QEXACTIVE-HCD scoring parameters, trypsin enzyme specificity with a maximum of two missed cleavages, 40% minimum matched peak intensity, ± 20 ppm precursor mass tolerance, ± 20 ppm product mass tolerance, and carbamidomethylation of cysteines and TMT5 labeling of lysines and peptide N termini as fixed modifications. Allowed variable modifications were oxidation of methionine, N-terminal acetylation, pyroglutamic acid (N-termQ), deamidated (N), pyro carboxamidomethyl Cys (N-termC), with a precursor MH+ shift range of −18–64 Da. Identities interpreted for individual spectra were automatically designated as significant using scoring and delta rank thresholds separately for each precursor charge state in each liquid chromatography-MS/MS while allowing a maximum target-decoy-based false-discovery rate (FDR) of 1.0% at the spectrum level.

Analysis. The expectation maximization algorithm was applied to the results of the peptide report (the in-house written bash script is available on the Lage Lab Resources Site (see URLs) and the peptide report can be found in the supplementary material). The list of most likely observed proteins was generated for each channel of the mass spectrometry experiment based on Swiss-Prot and TREMBLE databases of protein sequences. Next, ratios of intensities between channels were calculated and median normalized. Resulting data were analyzed and visualized using R. Statistical analyses were performed via moderated t-test from R package limma to estimate P values for each protein and the FDR corrections were applied to account for multiple hypothesis testing. Plots were created using in-house written R scripts and ggplot2. RNA-binding and S ribosomal protein families were taken from the HUGO Gene Nomenclature Committee (see URLs). Proteins previously reported to be EJC/NMD complex members were annotated as such.

Analysis of cancer cell lines. Copy number analysis. Details regarding arm-level copy number calling are as described by Taylor and colleagues. Briefly, to determine arm-level events that is, 1p or 12p deletion status in TCGA and CCLE samples, the ABSOLUTE algorithm was used to determine the likeliest ploidy and absolute total copy number of each genomic segment. Segments were called as amplified, deleted, or copy neutral based on copy number with reference to integer-rounded ploidy. Arm- or chromosome-level amplification/deletion status was then determined from segment data as described by Taylor and colleagues. CCLE cell lines were fit to clusters within their corresponding TCGA tumor type to generate cell line–specific, arm-specific calls. For CCLE data, ABSOLUTE algorithm was run on the CCLE Affymetrix SNP6.0 array data as previously reported. For all datasets, copy number was calculated with the TSVkit software tool. Copy number was determined as described above. Hemizygous MAGOH deletion was defined as the loss of one or more copies of the MAGOH gene (for example, ploidy − MAGOH copy number ≤ 1) using rounded tumor ploidy and MAGOH copy number calculated from the ABSOLUTE algorithm.

Genome-scale shRNA and CRISPR screening data analysis. Genome-wide shRNA screening on 501 cell lines was performed as described. The DEMETER method, which summarizes multiple shRNAs targeting a gene into a gene-level dependency score, was used to quantify gene dependency in 17,998 unique genes. The differential dependency data set of 769 genes, as defined previously, were used for all downstream analyses. These sets represent the genes with the most significant differential dependency across cell lines and were selected based on the following criteria: (1) for each gene, there is at least one cell line with a dependency score that is two (differential) or six (set 68) standard deviations from the mean of scores from all genes and all cell lines, and (2) the gene effect is not explained by any known toxicity for the gene in the most dependent cell line is above −2 log2 reads per kilobase million.

To identify synthetic lethal relationships linked to loss of a paralog, a query was performed for each of 17,670 genes using EnsemblCompara via the R interface to BioMart to obtain a list of paralogs and their pairwise sequence identity. Pearson correlation of RNA sequencing (RNA-seq) expression values between genes in each paralog pair indicate that co-expression is limited until DNA sequence identity exceeds 25% (Supplementary Fig. 1). To increase the likelihood that the gene pairs function redundantly, pairs with less than 25% sequence identity were removed. An additional 35 genes were removed for having duplicate DEMETER scores (caused by non-unique hairpins), resulting in 3,483 genes in the DEMETER dataset with at least one paralog. Differential dependency for each of these genes was tested by grouping cell lines based on loss of the gene’s paralogs and performing a two-class comparison of the DEMETER scores using empirical Bayes moderated t-statistics implemented by the R package limma. The binary classification of paralog loss used to group the cell lines was determined by a logic combination of RNA-seq gene expression, proten TCGA correlation (RPPA), relative copy number, methylation fraction (RRBS), and mutation status (whole-exome sequencing, whole-genome sequencing, RNA-seq). The gene expression, RPPA, copy number, and RRBS datasets are z-scored per gene so loss of a gene is defined as having a 60 decrease in gene expression or RPPA, or no gene expression at all (below 0.05 on a transcripset). 28% of paralogs were removed for increase in RRBS, or a deleterious mutation (predicted by frameshift indel or nonsense single-nucleotide variant). Genes are labeled ‘symmetric’ if loss of either gene in a pair is significantly associated with a selective dependency on its respective paralog gene with FDR < 0.05.

The synthetic lethal paralog analysis was repeated using the Achilles CRISPR dataset consisting of 341 whole genome CRISPR/cas9 knockout screens corrected for copy number effects (one cell line, PK59, was removed from the prior set of 342 as it failed fingerprinting). Genes with variance in essentiality below 0.01 across the 341 cell lines were removed to reduce false positives, leaving 6,535 genes for paralog dependency analysis. The definition of gene loss as well as method for determining paralog is defined previously. The 6.305 gene differential dependency set of 6,305 genes, and the 6.305 gene differential dependency of each paralog pair is identical to the analysis using DEMETER data.

For analysis of gene dependencies correlated with MAGOH deletion (Fig. 1), the Probability Analysis by Ranked Information Score (PARIS) algorithm was run as a GenePattern module (see URLs). MAGOH-deletion status was determined by the ABSOLUTE algorithm as described above. Cell lines for which MAGOH absolute copy number was less than the cell line’s ploidy were considered deleted. Based on available ABSOLUTE calls, 191 lines were considered deleted and 807 lines were considered non-deleted. In total, both absolute copy number and filtered DEMETER gene-score data were available for 445 overlap cell lines.

For geneset enrichment analysis on gene dependencies correlated with MAGOH deletion, the PARIS algorithm was first run using continuous copy number data on CCLE cell lines generated using SNP arrays, as previously reported, to generate a ranked list of gene dependencies correlated with MAGOH copy number. RNMI metric score for each gene was then used as input for prerranked geneset enrichment analysis, which was run as a GenePattern module using default parameters against the following genesets:
For analysis of correlated gene dependency profiles, Pearson correlations of DEMETER gene dependency scores were computed across cell lines (N = 501) for all pairs of genes that share overlap in cell lines (N = 6,300). Correlation coefficients were converted to standard scores across the full correlation matrix before evaluating the specific MAGOH and MAGOHB correlation profiles.

RNA-seq analysis. RNA-seq libraries were prepared using the Illumina strand-specific mRNA-seq Library Prep Kit (Illumina) followed by paired-end 75 bp sequencing on a NextSeq (>400 M reads per run; >33 M reads per sample). Transcript levels were quantified with kallisto19 (version 0.43.0; options: --rf 400 M reads per run; RNA-seq analysis. and evaluating the specific MAGOH correlation profiles. were converted to standard scores across the full correlation matrix before obtaining from the ENSEMBL database (release 87)41. Differential expression analysis was performed with sleuth42 (version 0.28.1). Differential expression was quantified based on the beta value, a bias estimator used by kallisto43 analogous to fold-change. Significant upregulation cut-offs were b > 1, q < 0.01; downregulation b < -1. Gene Ontology44 term enrichment analysis was carried out using PANTHER45; Overrepresentation Test (release 20160715) using the Gene Ontology database (release 2017-01-26), accessible via the Gene Ontology website (see URLs, last accessed 2017-01-31). Transcript biotypes were obtained from the ENSEMBL database (release 87)41. For analysis of differential alternative splicing events, reads were aligned with HiSat2 (v2.0.4, “rna-strandness RF option”) using the prebuilt index Ensembl GRCh38 genome (human), and splicing events were quantified using rMATs v3.2.55. For increased stringency, rMATs output was filtered based on read support (sum of inclusion/exclusion reads ≥ 10 in both samples), FDR (< 0.05), and inclusion level difference (|ILD| > 0.1). Sashimi plots were plotted using rMatssashimplot (see URLs) in grouping mode.

For estimation of protein-level effects from RNA-seq data, peptide sequences for each transcript were obtained using ENSEMBL biomart (see URLs), accessed through the R package biomaRt (version 2.26.1)52. For each gene, expected peptide expression was then estimated by summing over transcript per million values for all transcripts that encode peptides of the same length.

REVEALER analysis. To identify associations between MAGOH/B dependency and copy number/expression features of EJC/splicing-related genes, MAGOH or MAGOHB dependency scores across screened cell lines were correlated with copy number or expression features53 in EJC/splicing-related genes using the previously described method based on estimating the information coefficient46. For radial plots, the top-scoring 50 features (for copy number) or top-scoring 16 features (for expression) were plotted. A list of EJC/splicing-related genes used for this analysis was compiled by combining EJC/NMD genes in MSigDB (Reactome geneset no. 501) for = 8 in both samples), FDR (< 0.05), and inclusion level difference (|ILD| > 0.1). Gene Ontology56 term enrichment analysis was performed using PANTHER57; Overrepresentation Test (release 20160715) using the Gene Ontology database (release 2017-01-26), accessible via the Gene Ontology website (see URLs, last accessed 2017-01-31). Transcript biotypes were obtained from the ENSEMBL database (release 87)41. For analysis of differential alternative splicing events, reads were aligned with HiSat2 (v2.0.4, “rna-strandness RF option”) using the prebuilt index Ensembl GRCh38 genome (human), and splicing events were quantified using rMATs v3.2.55. For increased stringency, rMATs output was filtered based on read support (sum of inclusion/exclusion reads ≥ 10 in both samples), FDR (< 0.05), and inclusion level difference (|ILD| > 0.1). Sashimi plots were plotted using rMatssashimplot (see URLs) in grouping mode.

For estimation of protein-level effects from RNA-seq data, peptide sequences for each transcript were obtained using ENSEMBL biomart (see URLs), accessed through the R package biomaRt (version 2.26.1)52. For each gene, expected peptide expression was then estimated by summing over transcript per million values for all transcripts that encode peptides of the same length.

Statistics. No statistical methods were used to predetermine sample size. Investigators were not blinded to allocation for experiments. Statistical tests applied, P values, and sample size are as listed in figure captions. For in vitro experiments, number of biologically independent replicates is as listed in figure captions. When two-sample Student’s t-tests were applied to assess significance of experimental data, unequal variance parameter was used and P values were calculated using Microsoft Excel (function t.test; heteroscedastic). Other statistical tests were performed using R (v. 3.4.1) or GraphPad Prism 7 software.

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

Data availability. The original mass spectra may be downloaded from MassIVE (see URLs) using the identifier MSV000082292. RNA-seq data can be accessed at NCBI Gene Expression Omnibus (GSE113848) (see URLs). Code for analysis of IP-MS data is deposited in the Lage Lab website (see URLs). The authors declare that other data supporting the findings of this study are available within the paper and its supplementary information files. Other source data are available from the corresponding author on reasonable request.

References
38. Yang, X. et al. A public genome-scale lentiviral expression library of human ORFs. Nat. Methods 8, 659–661 (2011).
39. Lippa, M. S. et al. Expression of anti-apoptotic factors modulates Apo2L/TRAIL resistance in colon carcinoma cells. Apoptosis 12, 1465–1478 (2007).
40. Brown, C. Y. et al. Robust, reversible gene knockdown using a single lentiviral short hairpin RNA vector. Hum. Gene Ther. 21, 1005–1017 (2010).
41. Root, D. E., Haconen, N., Hahn, W. C., Lander, E. S. & Sabatini, D. M. Genome-scale loss-of-function screening with a lentiviral RNAi library. Nat. Methods 3, 715–719 (2006).
42. Kuw, E. I., Dudani, J. S. & Bhatia, S. N. Ultrasensitive tumour-penetrating nanosensors of protease activity. Nat. Biomed. Eng. 1, 0054 (2017).
43. McLachlan, G. J. & Krishman, T. The EM Algorithm and Extensions 2nd edn (Wiley-Interscience, Hoboken, NJ, 2008).
44. The UniProt Consortium. UniProt: the universal protein knowledgebase. Nucleic Acids Res. 45, D158–D169 (2017).
45. Ritchie, M. E. et al. limma powers differential expression analyses for RNA-sequencing and microarray studies. Nucleic Acids Res. 43, e47 (2015).
46. Wickham, H. ggplot2 https://doi.org/10.1007/978-0-387-98141-3 (Springer, New York, 2009).
47. Taylor, A. M. et al. Genomic and functional approaches to understanding cancer aneuploidy. Cancer Cell 33, 676–689 (2018).
48. Carter, S. L. et al. Absolute quantification of somatic DNA alterations in human cancer. Nat. Biotechnol. 30, 413–421 (2012).
49. Pedregosa, F. et al. Scikit-learn: machine learning in Python. JMLR 12, 2825–2830 (2011).
50. Barretina, J. et al. The Cancer Cell Line Encyclopedia enables predictive modelling of anticancer drug sensitivity. Nature 483, 603–607 (2012).
51. Villéla, A. J. et al. EnsembCompara GeneTrees: complete, duplication-aware phylogenetic trees in vertebrates. Genome Res. 19, 527–335 (2008).
52. Durinck, S., Spellman, P. T., Birney, E. & Huber, W. Mapping identifiers for the integration of genomic datasets with the R/Bioconductor package biomaRt. Nat. Protoc. 4, 1184–1191 (2009).
53. Subramanian, A. et al. Gene set enrichment analysis: A knowledge-based approach for interpreting genome-wide expression profiles. Proc. Natl Acad. Sci. USA 102, 15545–15550 (2005).
54. Aken, B. L. et al. Ensembl 2017. Nucleic Acids Res. 45, D635–D642 (2017).
55. Pimentel, H. J., Bray, N., Puente, S., Melsted, P. & Pachter, L. Differential analysis of RNA-seq incorporating quantification uncertainty. Nat. Methods 14, 687–690 (2016).
56. The Gene Ontology Consortium. Gene Ontology Consortium: going forward. Nat. Protoc. 8, 1551–1566 (2013).
57. Mi, H., Muruganujan, A., Casagrande, J. T. & Thomas, P. D. Large-scale gene function analysis with the PANTHER classification system. Nat. Protoc. 8, 357–360 (2013).
58. Shen, S. et al. rMATs: robust and flexible detection of differential alternative splicing from replicate RNA-seq data. Proc. Natl Acad. Sci. USA 111, ES593–ES601 (2014).
59. Sven, A., Kilpinen, S., Ruusulehto, A., Lothe, R. A., & Skotheim, R. I. Aberrent RNA splicing in cancer; expression changes and driver mutations of splicing factor genes. Oncogene 35, 2413–2427 (2015).
Reporting Summary

Nature Research wishes to improve the reproducibility of the work that we publish. This form provides structure for consistency and transparency in reporting. For further information on Nature Research policies, see Authors & Referees and the Editorial Policy Checklist.

Statistical parameters

When statistical analyses are reported, confirm that the following items are present in the relevant location (e.g. figure legend, table legend, main text, or Methods section).

- **n/a**
- **Confirmed**
  - The exact sample size (n) for each experimental group/condition, given as a discrete number and unit of measurement
  - An indication of whether measurements were taken from distinct samples or whether the same sample was measured repeatedly
  - The statistical test(s) used AND whether they are one- or two-sided
    - Only common tests should be described solely by name; describe more complex techniques in the Methods section.
  - A description of all covariates tested
  - A description of any assumptions or corrections, such as tests of normality and adjustment for multiple comparisons
  - A full description of the statistics including central tendency (e.g. means) or other basic estimates (e.g. regression coefficient) AND variation (e.g. standard deviation) or associated estimates of uncertainty (e.g. confidence intervals)
  - For null hypothesis testing, the test statistic (e.g. F, t, r) with confidence intervals, effect sizes, degrees of freedom and P value noted
    - Give P values as exact values whenever suitable.
  - For Bayesian analysis, information on the choice of priors and Markov chain Monte Carlo settings
  - For hierarchical and complex designs, identification of the appropriate level for tests and full reporting of outcomes
  - Estimates of effect sizes (e.g. Cohen's d, Pearson's r), indicating how they were calculated
  - Clearly defined error bars
    - State explicitly what error bars represent (e.g. SD, SE, CI)

Our web collection on statistics for biologists may be useful.

Software and code

Policy information about availability of computer code

Data collection

1) RNA-Seq: Software used is publicly available and as described in Methods ("Analysis of Cancer Cell Lines" section; "RNA-Seq Analysis" sub-section)
2) Data analysis of CCLE Cell Lines: Data was processed and analyzed as described in Methods ("Analysis of Cancer Cell Lines")
3) IP-MS Data: Data was processed and analyzed as described in Methods ("Immunoprecipitation and Mass Spectrometry" section; "Database search and data processing" and "Analysis" sub-section).

Data analysis

Code used to apply expectation maximization algorithm to the IP-MS data is available through the Lage Lab website (http://www.lagelab.org/resources) and described in Methods ("Immunoprecipitation and Mass Spectrometry" section; "Analysis" sub-section).

For manuscripts utilizing custom algorithms or software that are central to the research but not yet described in published literature, software must be made available to editors/reviewers upon request. We strongly encourage code deposition in a community repository (e.g. GitHub). See the Nature Research guidelines for submitting code & software for further information.
Policy information about availability of data
All manuscripts must include a data availability statement. This statement should provide the following information, where applicable:
- Accession codes, unique identifiers, or web links for publicly available datasets
- A list of figures that have associated raw data
- A description of any restrictions on data availability

Accession codes for RNA-Seq and MS data are as specified in the Methods ("Data Availability Statement").

Field-specific reporting
Please select the best fit for your research. If you are not sure, read the appropriate sections before making your selection.

☑ Life sciences  ☐ Behavioural & social sciences  ☐ Ecological, evolutionary & environmental sciences

For a reference copy of the document with all sections, see nature.com/authors/policies/ReportingSummary-flat.pdf

Life sciences study design
All studies must disclose on these points even when the disclosure is negative.

Sample size
Statistical methods were not used to predetermine sample size. There was no pre-specified effect size. Sample size for in vitro experiments was comparable to those used in the field. Sample size for in vivo experiments was felt to be reasonable on the basis of the phenotype observed in vitro and also comparable to those used in the field for xenograft experiments.

Data exclusions
For analysis of Project Achilles data, paralog genes were filtered to genes sharing > 25% DNA sequence similarity. In analysis of dependency data, “differentially dependent” genes were filtered to those showing 2sigma differential dependence (shRNA) or using a variance cutoff of 0.01 (CRISPR) as described in Methods ("Analysis of Cancer Cell Lines" section).

Replication
The experiments were reproducible. Additionally, the underlying findings were replicated across multiple cell lines and experimental contexts (i.e. see Fig. 1a-b, Supplemental Tables 1-2).

Randomization
For doxycycline-inducible shRNA xenograft experiments, mice were randomized at 7 days post-injection to match tumor size between groups and one group was started on doxycycline diet.

Blinding
Investigators were not blinded to group allocation.

Behavioural & social sciences study design
All studies must disclose on these points even when the disclosure is negative.

Study description
Briefly describe the study type including whether data are quantitative, qualitative, or mixed-methods (e.g. qualitative cross-sectional, quantitative experimental, mixed-methods case study).

Research sample
State the research sample (e.g. Harvard university undergraduates, villagers in rural India) and provide relevant demographic information (e.g. age, sex) and indicate whether the sample is representative. Provide a rationale for the study sample chosen. For studies involving existing datasets, please describe the dataset and source.

Sampling strategy
Describe the sampling procedure (e.g. random, snowball, stratified, convenience). Describe the statistical methods that were used to predetermine sample size OR if no sample-size calculation was performed, describe how sample sizes were chosen and provide a rationale for why these sample sizes are sufficient. For qualitative data, please indicate whether data saturation was considered, and what criteria were used to decide that no further sampling was needed.

Data collection
Provide details about the data collection procedure, including the instruments or devices used to record the data (e.g. pen and paper, computer, eye tracker, video or audio equipment) whether anyone was present besides the participant(s) and the researcher, and whether the researcher was blind to experimental condition and/or the study hypothesis during data collection.

Timing
Indicate the start and stop dates of data collection. If there is a gap between collection periods, state the dates for each sample cohort.

Data exclusions
If no data were excluded from the analyses, state so OR if data were excluded, provide the exact number of exclusions and the rationale behind them, indicating whether exclusion criteria were pre-established.

Non-participation
State how many participants dropped out/declined participation and the reason(s) given OR provide response rate OR state that no participants dropped out/declined participation.
Eco logical, ev olutionary & environmental sciences study design

All studies must disclose on these points even when the disclosure is negative.

Study description
Briefly describe the study. For quantitative data include treatment factors and interactions, design structure (e.g. factorial, nested, hierarchical), nature and number of experimental units and replicates.

Research sample
Describe the research sample (e.g. a group of tagged Passer domesticus, all Stenocereus thurberi within Organ Pipe Cactus National Monument), and provide a rationale for the sample choice. When relevant, describe the organism taxo, source, sex, age range and any manipulations. State what population the sample is meant to represent when applicable. For studies involving existing datasets, describe the data and its source.

Sampling strategy
Note the sampling procedure. Describe the statistical methods that were used to predetermine sample size OR if no sample-size calculation was performed, describe how sample sizes were chosen and provide a rationale for why these sample sizes are sufficient.

Data collection
Describe the data collection procedure, including who recorded the data and how.

Timing and spatial scale
Indicate the start and stop dates of data collection, noting the frequency and periodicity of sampling and providing a rationale for these choices. If there is a gap between collection periods, state the dates for each sample cohort. Specify the spatial scale from which the data are taken.

Data exclusions
If no data were excluded from the analyses, state so OR if data were excluded, describe the exclusions and the rationale behind them, indicating whether exclusion criteria were pre-established.

Reproducibility
Describe the measures taken to verify the reproducibility of experimental findings. For each experiment, note whether any attempts to repeat the experiment failed OR state that all attempts to repeat the experiment were successful.

Randomization
Describe how samples/organisms/participants were allocated into groups. If allocation was not random, describe how covariates were controlled. If this is not relevant to your study, explain why.

Blinding
Describe the extent of blinding used during data acquisition and analysis. If blinding was not possible, describe why OR explain why blinding was not relevant to your study.

Field work, collection and transport

Field conditions
Describe the study conditions for field work, providing relevant parameters (e.g. temperature, rainfall).

Location
State the location of the sampling or experiment, providing relevant parameters (e.g. latitude and longitude, elevation, water depth).

Access and import/export
Describe the efforts you have made to access habitats and to collect and import/export your samples in a responsible manner and in compliance with local, national and international laws, noting any permits that were obtained (give the name of the issuing authority, the date of issue, and any identifying information).

Disturbance
Describe any disturbance caused by the study and how it was minimized.

Reporting for specific materials, systems and methods

Materials & experimental systems

| n/a | Involved in the study |
|-----|-----------------------|
| ✗   | Unique biological materials |
| ✗   | Antibodies |
| ✗   | Eukaryotic cell lines |
| ✗   | Palaeontology |
| ✗   | Animals and other organisms |
| ✗   | Human research participants |

Methods

| n/a | Involved in the study |
|-----|-----------------------|
| ✗   | ChIP-seq |
| ✗   | Flow cytometry |
| ✗   | MRI-based neuroimaging |
Unique biological materials

Policy information about availability of materials

Obtaining unique materials

Describe any restrictions on the availability of unique materials OR confirm that all unique materials used are readily available from the authors or from standard commercial sources (and specify these sources).

Antibodies

Antibodies used

Catalog numbers for antibodies used in the manuscript are listed in Methods ("Western Blotting", "Immunofluorescence" and "Mouse Experiments" sections).

Validation

Primary antibodies were used at the dilutions specified in Methods ("Western Blotting", "Immunofluorescence" and "Mouse Experiments" sections), which were within the recommended usage range provided by the manufacturers. Antibodies were validated based on staining at the predicted molecular weight or in a pattern reported by manufacturer under standard conditions. For flow cytometry experiments, surface staining was quantified relative to an IgG negative control.

Eukaryotic cell lines

Policy information about cell lines

Cell line source(s)

Sources of cell lines were as described in Methods ("Cell Culture" section). Briefly, cell line stocks were obtained from the Biological Samples Platform at the Broad Institute and were originally obtained from validated source repositories or academic labs and authenticated prior to use in the Cancer Cell Line Encyclopedia. Original sources for cell lines used were as follows:

1. ChagoK1 - ECACC
2. H460 - ATCC
3. H1437 - ATCC
4. H1373 - ATCC
5. Kuramochi - HSRRB
6. HCC1359 - KCLB
7. HUH1 - HSRRB
8. Heya8 - Academic Lab
9. CHP-212 - ATCC
10. SK-N-DZ - ATCC

Authentication

Cell lines were authenticated in the Biological Samples Platform at the Broad Institute using fingerprint arrays prior to creation of frozen stocks. Additionally, the following cell lines were re-fingerprinted for confirmation using STR profiling after being maintained in culture over the course of the experiments in this manuscript: ChagoK1, H1437, H460.

Mycoplasma contamination

Cell lines were tested for mycoplasma prior to making frozen stocks in Biological Samples Platform. Cell lines were periodically tested for mycoplasma if maintained continuously in culture. No cell lines tested positive for mycoplasma during the course of experiments performed for this manuscript.

Commonly misidentified lines (See ICLAC register)

None

Palaeontology

Specimen provenance

Provide provenance information for specimens and describe permits that were obtained for the work (including the name of the issuing authority, the date of issue, and any identifying information).

Specimen deposition

Indicate where the specimens have been deposited to permit free access by other researchers.

Dating methods

If new dates are provided, describe how they were obtained (e.g. collection, storage, sample pretreatment and measurement), where they were obtained (i.e. lab name), the calibration program and the protocol for quality assurance OR state that no new dates are provided.

☐ Tick this box to confirm that the raw and calibrated dates are available in the paper or in Supplementary Information.

Animals and other organisms

Policy information about studies involving animals; ARRIVE guidelines recommended for reporting animal research

Laboratory animals

NCR-nude mice, female, 4-5 weeks of age. See also Methods ("Mouse Experiments" section).

Wild animals

Wild animals were not used in this study.
Field-collected samples: Field-collected samples were not used in this study.

Human research participants

Policy information about studies involving human research participants

Population characteristics: Describe the covariate-relevant population characteristics of the human research participants (e.g. age, gender, genotypic information, past and current diagnosis and treatment categories). If you filled out the behavioural & social sciences study design questions and have nothing to add here, write "See above."

Recruitment: Describe how participants were recruited. Outline any potential self-selection bias or other biases that may be present and how these are likely to impact results.

ChIP-seq

Data deposition

☐ Confirm that both raw and final processed data have been deposited in a public database such as GEO.

☐ Confirm that you have deposited or provided access to graph files (e.g. BED files) for the called peaks.

Data access links

For "Initial submission" or "Revised version" documents, provide reviewer access links. For your "Final submission" document, provide a link to the deposited data.

Files in database submission

Provide a list of all files available in the database submission.

Genome browser session

Provide a link to an anonymized genome browser session for "Initial submission" and "Revised version" documents only, to enable peer review. Write "no longer applicable" for "Final submission" documents.

Methodology

Replicates

Describe the experimental replicates, specifying number, type and replicate agreement.

Sequencing depth

Describe the sequencing depth for each experiment, providing the total number of reads, uniquely mapped reads, length of reads and whether they were paired- or single-end.

Antibodies

Describe the antibodies used for the ChIP-seq experiments; as applicable, provide supplier name, catalog number, clone name, and lot number.

Peak calling parameters

Specify the command line program and parameters used for read mapping and peak calling, including the CHIP, control and index files used.

Data quality

Describe the methods used to ensure data quality in full detail, including how many peaks are at FDR 5% and above 5-fold enrichment.

Software

Describe the software used to collect and analyze the ChIP-seq data. For custom code that has been deposited into a community repository, provide accession details.

Flow Cytometry

Plots

☐ The axis labels state the marker and fluorochrome used (e.g. CD4-FITC).

☐ The axis scales are clearly visible. Include numbers along axes only for bottom left plot of group (a 'group' is an analysis of identical markers).

☐ All plots are contour plots with outliers or pseudocolor plots.

☐ A numerical value for number of cells or percentage (with statistics) is provided.

Methodology

Sample preparation

Cancer cells were cultured using standard techniques and as described in Methods, then collected using enzyme-free dissociation buffer (Gibco, 13151-014), stained, and analyzed.

Instrument

BD LSR Fortessa (custom equipment)

Software

Data were collected using BD FACSDiva software and analyzed using FlowJo v10

Cell population abundance

Cells are a uniform population from cell line culture.
Dead cells were excluded based on viability stain, then cells were gated based on FSC and SSC to exclude debris. No other gating was necessary for the uniform cell populations analyzed in this manuscript.

Tick this box to confirm that a figure exemplifying the gating strategy is provided in the Supplementary Information.

Magnetic resonance imaging

**Experimental design**

**Design type**

Indicate task or resting state; event-related or block design.

**Design specifications**

Specify the number of blocks, trials or experimental units per session and/or subject, and specify the length of each trial or block (if trials are blocked) and interval between trials.

**Behavioral performance measures**

State number and/or type of variables recorded (e.g. correct button press, response time) and what statistics were used to establish that the subjects were performing the task as expected (e.g. mean, range, and/or standard deviation across subjects).

**Acquisition**

**Imaging type(s)**

Specify: functional, structural, diffusion, perfusion.

**Field strength**

Specify in Tesla

**Sequence & imaging parameters**

Specify the pulse sequence type (gradient echo, spin echo, etc.), imaging type (EPI, spiral, etc.), field of view, matrix size, slice thickness, orientation and TE/TR/flip angle.

**Area of acquisition**

State whether a whole brain scan was used OR define the area of acquisition, describing how the region was determined.

**Diffusion MRI**

- Used
- Not used

**Preprocessing**

**Preprocessing software**

Provide detail on software version and revision number and on specific parameters (model/functions, brain extraction, segmentation, smoothing kernel size, etc.).

**Normalization**

If data were normalized/standardized, describe the approach(es): specify linear or non-linear and define image types used for transformation OR indicate that data were not normalized and explain rationale for lack of normalization.

**Normalization template**

Describe the template used for normalization/ transformation, specifying subject space or group standardized space (e.g. original Talairach, MN1305, ICBM152) OR indicate that the data were not normalized.

**Noise and artifact removal**

Describe your procedure(s) for artifact and structured noise removal, specifying motion parameters, tissue signals and physiological signals (heart rate, respiration).

**Volume censoring**

Define your software and/or method and criteria for volume censoring, and state the extent of such censoring.

**Statistical modeling & inference**

**Model type and settings**

Specify type (mass univariate, multivariate, RSA, predictive, etc.) and describe essential details of the model at the first and second levels (e.g. fixed, random or mixed effects; drift or auto-correlation).

**Effect(s) tested**

Define precise effect in terms of the task or stimulus conditions instead of psychological concepts and indicate whether ANOVA or factorial designs were used.

**Specify type of analysis:**

- Whole brain
- ROI-based
- Both

**Statistic type for inference**

(See Eklund et al. 2016)

Specify voxel-wise or cluster-wise and report all relevant parameters for cluster-wise methods.

**Correction**

Describe the type of correction and how it is obtained for multiple comparisons (e.g. FWE, FDR, permutation or Monte Carlo).

**Models & analysis**

n/a

- Functional and/or effective connectivity
- Graph analysis
- Multivariate modeling or predictive analysis
| Section                                      | Description                                                                 |
|----------------------------------------------|-----------------------------------------------------------------------------|
| Functional and/or effective connectivity     | Report the measures of dependence used and the model details (e.g. Pearson correlation, partial correlation, mutual information). |
| Graph analysis                               | Report the dependent variable and connectivity measure, specifying weighted graph or binarized graph, subject- or group-level, and the global and/or node summaries used (e.g. clustering coefficient, efficiency, etc.). |
| Multivariate modeling and predictive analysis | Specify independent variables, features extraction and dimension reduction, model, training and evaluation metrics. |