



# Genetic interaction mapping and exon-resolution functional genomics with a hybrid Cas9–Cas12a platform

Thomas Gonatopoulos-Pournatzis<sup>1,6</sup>  , Michael Aregger<sup>1,6</sup> , Kevin R. Brown<sup>1,6</sup> , Shaghayegh Farhangmehr<sup>1,2</sup>, Ulrich Braunschweig<sup>1</sup> , Henry N. Ward<sup>3</sup> , Kevin C. H. Ha<sup>1</sup> , Alexander Weiss<sup>1</sup>, Maximilian Billmann<sup>4</sup> , Tanja Durbic<sup>1</sup>, Chad L. Myers<sup>3,4</sup> , Benjamin J. Blencowe<sup>1,2</sup>  and Jason Moffat<sup>1,2,5</sup>  

**Systematic mapping of genetic interactions (GIs) and interrogation of the functions of sizable genomic segments in mammalian cells represent important goals of biomedical research. To advance these goals, we present a CRISPR (clustered regularly interspaced short palindromic repeats)-based screening system for combinatorial genetic manipulation that employs coexpression of CRISPR-associated nucleases 9 and 12a (Cas9 and Cas12a) and machine-learning-optimized libraries of hybrid Cas9–Cas12a guide RNAs. This system, named Cas Hybrid for Multiplexed Editing and screening Applications (CHyMERa), outperforms genetic screens using Cas9 or Cas12a editing alone. Application of CHyMERa to the ablation of mammalian paralog gene pairs reveals extensive GIs and uncovers phenotypes normally masked by functional redundancy. Application of CHyMERa in a chemogenetic interaction screen identifies genes that impact cell growth in response to mTOR pathway inhibition. Moreover, by systematically targeting thousands of alternative splicing events, CHyMERa identifies exons underlying human cell line fitness. CHyMERa thus represents an effective screening approach for GI mapping and the functional analysis of sizable genomic regions, such as alternative exons.**

Recent breakthroughs in gene editing technologies have transformed mammalian cell genetics and disease research<sup>1–4</sup>. In particular, genome-scale screens employing CRISPR–Cas nucleases have already begun to deliver unprecedented insight into genotype–phenotype relationships<sup>5</sup>. For example, such screens have defined genes required for human cell line proliferation that share functional, evolutionary and physiological properties with essential genes in other organisms<sup>6–9</sup>. These studies have ushered in a new era of functional genomics by enabling the systematic perturbation and characterization of genes that underlie biological processes and phenotypes<sup>10–16</sup>.

Despite these advances, current major challenges in genomics include the development of efficient tools for the comprehensive mapping of GIs—that is, deviations from expected phenotypes when multiple genetic perturbations are combined—as well as the functional interrogation of sizable genomic fragments such as alternative exons. For example, an important question is the extent to which combinations of paralogous mammalian genes are important for phenotypic robustness. In particular, despite the widespread emergence of paralogous genes in higher organisms as a consequence of small-scale and whole-genome duplication events during vertebrate evolution<sup>17</sup>, it is unclear to what extent paralogs have redundant or distinct functions in human cells for a given phenotype. Similarly, it is also not known to what extent annotated alternative exons contribute to critical cell functions.

Key to addressing these questions is the generation of tools for combinatorial genetic perturbation. Although screening systems employing expression of two or more Cas9 guides have been described<sup>18–22</sup>, these approaches have limitations that impact their efficiency, including recombination between duplicated promoters and expression cassettes<sup>18,23–26</sup>. Cas12a (formerly Cpf1) nuclease possesses intrinsic RNase activity and can generate multiple guide RNAs from a single concatemeric gRNA transcript<sup>27–29</sup>, making this an attractive option for combinatorial gene targeting. However, the previously reported efficiency of generation of multiple indels in the same cell with Cas12a is less than 15%<sup>28,30</sup>. Nevertheless, Cas12a has been exploited in a positive-selection screen to identify pairwise GIs between tumor suppressor genes that, when ablated, accelerated tumor growth in a mouse model of metastasis<sup>30</sup>. With strong selection pressure, rare editing events can lead to small numbers of clones with substantial positive growth potential. In contrast, detection of negative growth effects presents a greater challenge as it requires highly efficient genetic perturbation systems.

To address the limitations of current screening approaches we describe CHyMERa, a system that uses coexpression of *Streptococcus pyogenes* (Sp)-Cas9 and *Lachnospiraceae* bacterium (Lb)-Cas12a nucleases, together with ‘hybrid guide’ (hg)RNAs generated from fusions of Cas9 and Cas12a gRNAs expressed from a single promoter. Through iterative rounds of pooled hgRNA library construction, screening, and the use of deep learning, we describe

<sup>1</sup>Donnelly Centre, University of Toronto, Toronto, Ontario, Canada. <sup>2</sup>Department of Molecular Genetics, University of Toronto, Toronto, Ontario, Canada.

<sup>3</sup>Bioinformatics and Computational Biology Graduate Program, University of Minnesota, Minneapolis, MN, USA. <sup>4</sup>Department of Computer Science and Engineering, University of Minnesota, Minneapolis, MN, USA. <sup>5</sup>Institute for Biomedical Engineering, University of Toronto, Toronto, Ontario, Canada.

<sup>6</sup>These authors contributed equally: Thomas Gonatopoulos-Pournatzis, Michael Aregger, Kevin R. Brown. ✉e-mail: [t.gonatopoulos@gmail.com](mailto:t.gonatopoulos@gmail.com); [b.blencowe@utoronto.ca](mailto:b.blencowe@utoronto.ca); [j.moffat@utoronto.ca](mailto:j.moffat@utoronto.ca)

optimized hgRNA designs for the human and mouse genomes and demonstrate that the resulting constituent Cas12a gRNA efficiencies are comparable to those of efficient Cas9 gRNAs. Using CHyMERa, we perform screens with optimized hgRNA libraries targeting 672 human paralog pairs (representing >90% of predicted duplicate paralog genes in the human genome), explore chemogenetic interactions in the mTOR pathway and interrogate the functions of 2,157 alternative cassette exons in cell fitness. These screens demonstrate a previously unappreciated degree of complexity of GIs among paralogous genes, reveal new chemical GIs, and identify numerous alternative exons that impact cell growth.

## Results

**Development of a hybrid CRISPR–Cas system for programmable multisite genome editing.** To develop a multitargeting CRISPR editing platform, we compared different lentiviral-based approaches employing gRNAs designed to direct deletion of exon 8 of the mouse *Ptbp1* gene by targeting flanking intronic sequences (see Methods). Employing Cas9 with constructs expressing two gRNAs results in poor exon deletion efficiency (Supplementary Fig. 1a,b and data not shown). We reasoned that combination of the relatively efficient Cas9 system with Cas12a, which affords combinatorial targeting through gRNA multiplexing<sup>28,29</sup>, would generate a more effective system. Accordingly, we generated cell lines coexpressing Sp-Cas9 and either Lb-Cas12a or *Acidaminococcus* sp. *BV3L6* (As)-Cas12a, together with hgRNAs fusing Cas9 and Cas12a guides (Fig. 1a and Supplementary Fig. 1c,d). These hgRNAs are processed by Cas12a RNase activity (Supplementary Fig. 1e)<sup>28,29</sup>, liberating individual Cas9 and Cas12a gRNAs for loading into their respective nucleases (Fig. 1a).

Cas9 and Cas12a hgRNA pairs targeting sequences flanking *Ptbp1* exon 8 yield deletion efficiencies of 10–43% in mouse embryonic stem cells (Fig. 1b). These efficiencies are substantially higher than for any other tested combination of Cas nucleases (Fig. 1b and data not shown). Increased editing efficiency was also observed for hgRNA pairs targeting other alternative exons for deletion in mouse and human cells (Supplementary Fig. 1f). Next, we tested combinations of Cas9 and Cas12a hgRNAs targeting the *HPRT1* and *TK1* genes, which, when knocked out, result in cells becoming resistant to 6-thioguanine (6-TG) or thymidine block, respectively. Lentiviral transduction of CHyMERa hgRNA constructs targeting *HPRT1* and *TK1*, with Cas12a and Cas9 respectively, results in strong resistance to both treatments (Fig. 1c). Importantly, sequential drug treatment leads to generation of dual-resistant cell populations, confirming combinatorial editing within the same cells (Supplementary Fig. 1g). Furthermore, multiplexing of up to three Cas12a guides results

in robust editing after the addition of intergenic (that is, negative control) guide sequences at internal positions while keeping an *HPRT1*-targeting guide at the last position of a multitargeting hgRNA construct (Fig. 1c).

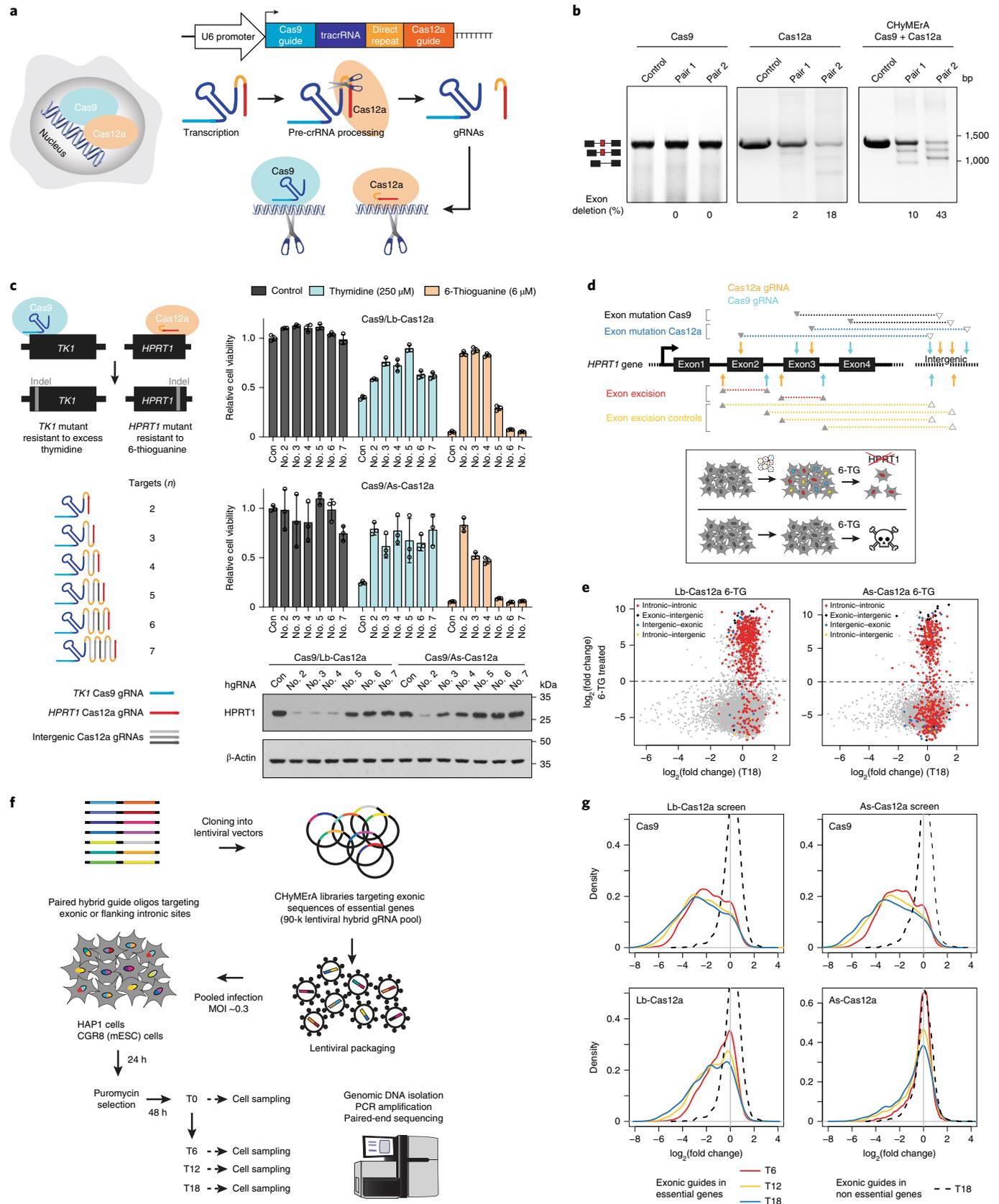
We next tested the efficiency of CHyMERa in a pooled screen format. Lentiviral-based, positive-selection screens were performed in human HAP1 cells with pools of hgRNAs targeting ~1,000 human genes, including *HPRT1* or *TK1*, using hgRNA pairs where one gRNA is directed to a constitutive exon sequence that, when disrupted, is expected to result in loss of gene function, and the other gRNA is directed to a control intergenic sequence (Supplementary Table 1). Additional hgRNAs were tested that target intronic sites flanking the same exons and were expected to result in exon deletion (Fig. 1d). Following treatment with 6-TG, 95.8% of all library constructs were undetectable, indicating strong negative selection driven by the drug treatment. Importantly, we also observed strong enrichment of hgRNAs targeting *HPRT1* exonic sequences, and hgRNAs comprising Cas9–Cas12a pairs targeting *HPRT1* exons for deletion (Wilcoxon rank-sum test,  $P < 2.2 \times 10^{-16}$ ; Fig. 1e and Supplementary Fig. 1h). Furthermore, 94 and 67% of hgRNAs that directly target exon sequences are also enriched using the same criteria. Similar results were obtained for guides targeting *TK1* after double-thymidine block in HAP1 cells (Supplementary Fig. 1i). These experiments also reveal that Lb-Cas12a is more efficient at editing compared to As-Cas12a (Fig. 1c,e and Supplementary Fig. 1i; see also below). Collectively, these data demonstrate that coexpression of Cas9, Cas12a and hgRNAs represents an effective system for combinatorial genetic perturbation, including deletion of sizable genetic elements.

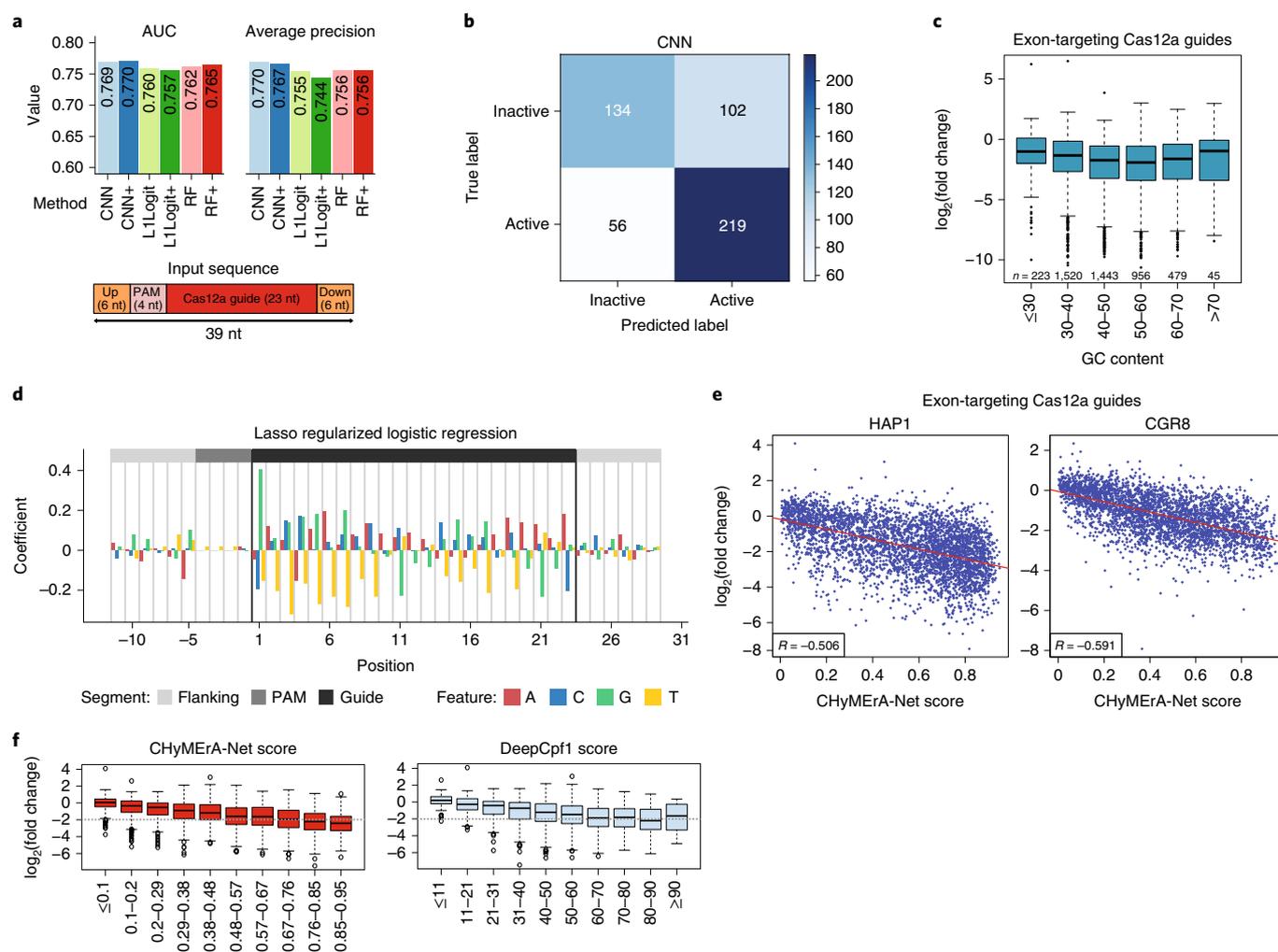
**Optimization of Cas12a gRNAs employed by CHyMERa.** While the rules for the design of efficient Cas9 gRNAs are well established<sup>31–34</sup>, the parameters governing the editing efficiency of Cas12a guides are less well understood. Accordingly, we generated human and mouse hgRNA ‘optimization’ libraries targeting core essential genes for inactivation and exons for deletion. These optimization libraries target more than 450 core essential genes<sup>33</sup> with over 6,000 Cas9 and Cas12a exon-targeting guides and over 35,000 exon-flanking guides. They also contain 1,000 control constructs targeting intergenic regions, which control for toxicity induced by double-stranded (ds)DNA breaks (see Methods, Supplementary Fig. 2a and Supplementary Tables 1 and 2). To construct these hgRNA libraries, we developed a two-step cloning strategy (Supplementary Fig. 2b; see also Methods), generated high-titer lentiviral stocks and transduced them at a low multiplicity of infection (MOI) into HAP1 and CGR8 embryonic stem cells (Fig. 1f). Following puromycin selection

**Fig. 1 | Development of CHyMERa, a screening platform for combinatorial genetic perturbations.** **a**, Schematic overview of CHyMERa. A hgRNA consisting of a fusion of Cas9 and Cas12a gRNAs is expressed under a single U6 promoter. Cas12a RNA processing activity cleaves hgRNA to generate functional Cas9 and Cas12a gRNAs. **b**, PCR assay monitoring of *Ptbp1* exon 8 deletion efficiency using paired Cas9 intronic guides (left), paired Cas12a intronic guides (middle) or CHyMERa (right). Representative data from two to four independent experiments. **c**, HAP1 cells expressing Cas9 and Cas12a (Lb or As) were transduced with lentiviral expression cassettes for multiplexed hgRNAs encoding an increasing number of targets as indicated. For all hgRNA constructs, the first and last positions encode a *TK1*-targeting Cas9 and *HPRT1*-targeting Cas12a gRNA, respectively, while the intervening positions encode intergenic Cas12a gRNAs (left). To assay resistance to thymidine block and 6-TG treatment, cells were either control (Con)-treated or challenged with 250  $\mu$ M thymidine or 6  $\mu$ M 6-TG. Cell viability was measured by alamarBlue staining 4 d post treatment, relative to the nontargeting control. Bars indicate mean  $\pm$  s.d. of three independent biological replicates. Immunoblot was performed to detect HPRT1 levels, and  $\beta$ -actin was used as a loading control (right). Representative data from two independent experiments. **d**, Schematic of hgRNA constructs designed to mutate or delete exons by targeting exonic or flanking intronic sequences, respectively (top). Cas9- and Cas12a-mediated exon mutation guide pairs are indicated in black and blue, while exon excision and the respective control guide pairs are indicated in red and yellow, respectively. Overview of positive-selection screens in which cells were treated with 6-TG (bottom). **e**, Enrichment of guide pairs targeting exons in *HPRT1* for deletion (red) or gene knockout (black, Cas9; blue, Cas12a) in 6-TG-treated (6  $\mu$ M; y axis) versus nontreated (x axis) HAP1 cells. Intronic/intergenic control hgRNAs are indicated in yellow, and non-*HPRT1*-targeting constructs detectable in at least one 6-TG-treated replicate ( $n = 9,659$ ) are shown in gray. Screens performed with Lb-Cas12a (left) and As-Cas12a (right). **f**, Overview of CHyMERa library generation and experimental setup for genetic screens. **g**, LFC distribution of Cas9 gRNAs (upper) or Cas12a gRNAs (lower) targeting essential (solid lines) and nonessential (dotted lines) genes for the indicated time points (T6, T12, T18) in HAP1 cells. Left: Lb-Cas12a screen; right: As-Cas12a screen. crRNA, CRISPR RNA.

for 2 d, cells were collected for the reference T0 time point. The remaining cells were split into three parallel replicates and passed independently every 3 d for 18 d (T18) while maintaining ~250-fold

library coverage. Genomic DNA was isolated at the T0, T6, T12 and T18 time points, and hgRNA barcode sequences were quantified by paired-end sequencing (Fig. 1f; see also Methods).



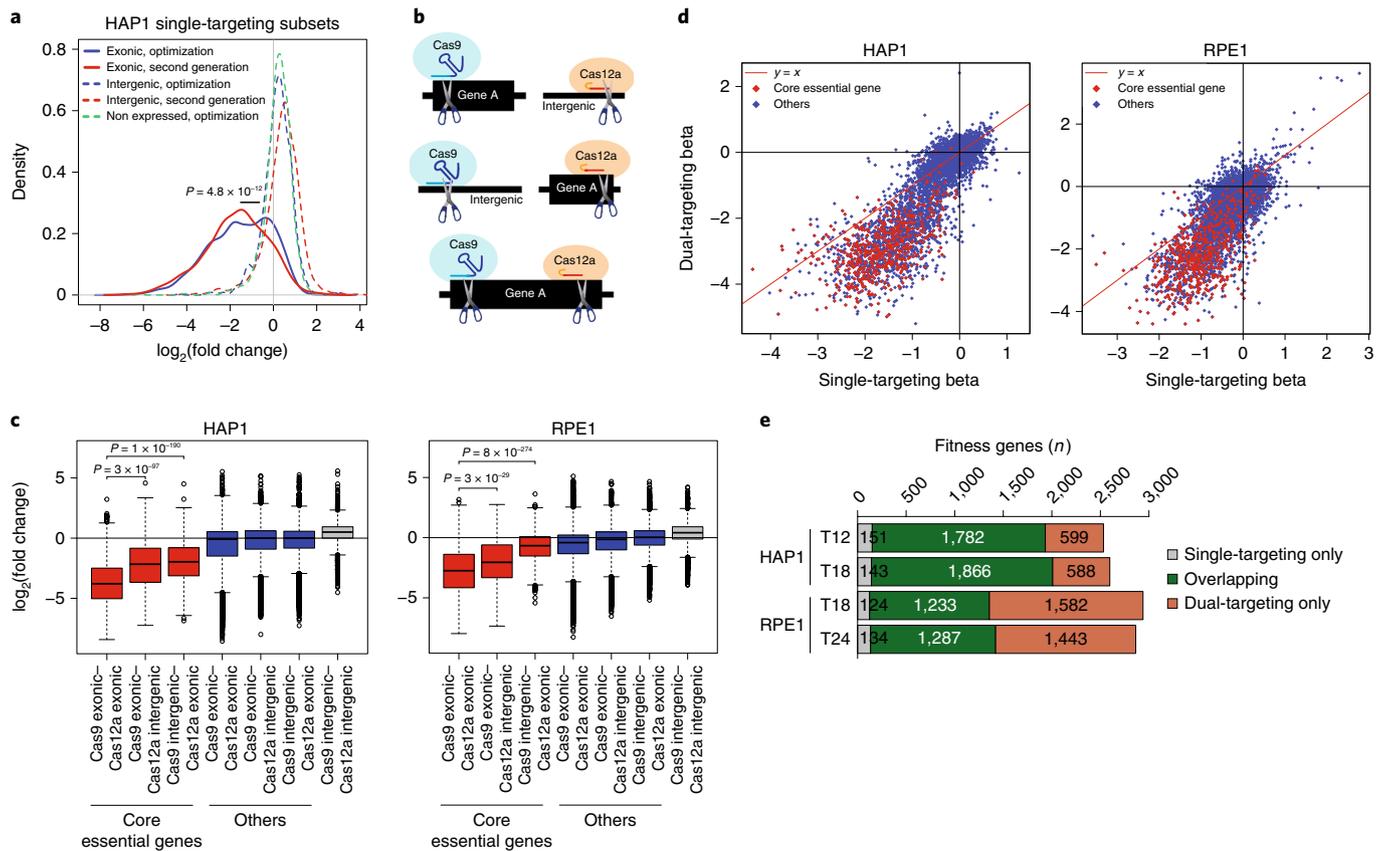


**Fig. 2 | Machine learning-based prediction of efficient Lb-Cas12a guides.** **a**, Evaluation of different machine learning algorithms for prediction of active Lb-Cas12a guides using AUC (left) or average precision (right). Active guides are defined as those with LFC < -1 at T18 compared to T0 (likelihood-ratio test, FDR < 0.05 with Benjamini-Hochberg multiple testing correction), and were chosen from three independent screens with three biological replicates each. Inactive guides defined by LFC between -0.5 and 0.5. Machine learning classifiers were trained using only the Cas12a gRNA target ( $n = 5,097$  unique sequences) and flanking sequence (39 nt), or with the addition of secondary structure and melting temperature (+). **b**, Assessment of predictive performance of the CNN classifier via cross-validation. **c**, Activity of exonic Lb-Cas12a guides binned by GC content. Boxes show the interquartile range (IQR, 25th to 75th percentile), with the median indicated by a horizontal line. Whiskers extend to the quartile  $\pm 1.5 \times$  IQR. **d**, Sequence composition of active exonic Lb-Cas12a guides from human and mouse optimization screens as determined by a logistic regression model. **e**, Pearson correlation coefficients between LFC and ChyMERa-Net score for Lb-Cas12a exonic guides in HAP1 (left,  $n = 4,268$  guides) and CGR8 (right,  $n = 3,338$  guides) cells. **f**, LFC distributions of 4,268 guides as a function of ChyMERa-Net (left) and DeepCpf1 scores (right). Boxes show IQR 25th to 75th percentile, with the median indicated by a horizontal line. Whiskers extend to the quartile  $\pm 1.5 \times$  IQR.

Analysis of  $\log_2$  fold-change (LFC) distributions for each time point shows strong depletion of hgRNAs from both Lb- and As-Cas12a libraries where Cas9 guides target core fitness genes and Cas12a guides target nonfunctional intergenic sequences, in both HAP1 and mouse CGR8 cells (Fig. 1g, Supplementary Fig. 2c and Supplementary Tables 3 and 4). LFC distributions also show strong depletion of hgRNAs where Cas12a guides target essential genes and Cas9 guides target nonfunctional intergenic sequences, an effect that is significantly stronger with Lb-Cas12a compared to As-Cas12a nuclease (Wilcoxon rank-sum test,  $P = 1.5 \times 10^{-150}$ ; Fig. 1g and Supplementary Fig. 2c). These results demonstrate the potential of Lb-Cas12a and hgRNA libraries in performing multi-site-targeting negative selection screens. Since Lb-Cas12a outperformed As-Cas12a, the former was used for the remainder of the study and, for simplicity, is referred to below as Cas12a.

**Deep learning prediction of efficient Cas12a guides.** Using data from human and mouse libraries targeting essential genes, we developed machine learning models based on convolutional neural networks (CNNs), Lasso regression and random forests (RF), to predict features of the most active Cas12a guides (see Methods and Supplementary Fig. 2d). The CNN algorithm, named ChyMERa-Net, achieved an area under the receiver operating characteristic curve (AUC) of 77% for both human and mouse cells (Fig. 2a,b and Supplementary Fig. 2e). The other approaches performed similarly, although with slightly reduced predictive power (Fig. 2a,b and Supplementary Fig. 2e).

The weights assigned to each feature in the Lasso classifier were used to determine parameters underlying optimal Cas12a guides and target sites. Active guides are neutral with respect to GC content, although there is a preference for G at the first



**Fig. 3 | Dual Cas9-Cas12a gene targeting outperforms single Cas9 or Cas12a editing.** **a**, LFC distributions of Lb-Cas12a exonic guides from optimized second-generation CHyMERa libraries at the endpoint (T18). Guides targeting intergenic regions or nonexpressed genes are included as negative controls ( $n = 4,268$  optimization guides,  $n = 2,026$  second-generation guides;  $P = 4.8 \times 10^{-12}$ , two-tailed Mann-Whitney  $U$ -test). **b**, Schematic of single- versus dual-gene targeting. **c**, Depletion of single- versus dual-targeting hgRNAs in HAP1 (T18, left) or RPE1 cells (T24, right). Boxes show IQR, 25th to 75th percentile, with the median indicated by a horizontal line. Whiskers extend to the quartile  $\pm 1.5 \times$  IQR. Subsets were compared using two-tailed Mann-Whitney  $U$ -tests. Tests were performed only between groups with indicated  $P$  values. hgRNA guides per group: 3,310 (Cas9 exonic-Cas12a exonic), 1,148 (Cas9 exonic-Cas12a intergenic) and 1,676 (Cas9 intergenic-Cas12a exonic) targeting core essential genes; 25,578 (Cas9 exonic-Cas12a exonic), 8,753 (Cas9 exonic-Cas12a intergenic) and 12,874 (Cas9 intergenic-Cas12a exonic) targeting other protein-coding genes; and 4,993 (Cas9 intergenic-Cas12a intergenic) controls. **d**, Correlation between gene-level dropout beta scores, as determined by the MAGeCK algorithm, for genes targeted by dual- (y axis) or single-targeting (x axis) hgRNAs in HAP1 (T18, left) or RPE1 cells (T24, right). **e**, Number of essential genes identified by the MAGeCK algorithm analyzing single- and dual-targeting hgRNAs at the indicated time points (T12 and T18).

position proximal to the PAM sequence, depletion of T at the first nine positions, and for a C at the PAM-distal 23rd nucleotide (Fig. 2c,d). Similar nucleotide preferences were observed in the filters learned by the CNN classifier (Supplementary Fig. 2f). Little predictive information is attributed to secondary structure, melting temperature, 6-nt regions flanking the target site or the 4-nt PAM sequence (Fig. 2d, Supplementary Fig. 2g and data not shown). In contrast to previous studies employing As-Cas12a<sup>35,36</sup>, we did not detect enrichment of active guides in regions with chromatin accessibility (Supplementary Fig. 2h). Supporting the accuracy of the CNN predictions, we observed a strong negative correlation between the CHyMERa-Net score for hgRNAs targeting essential genes and LFC guide scores between T0 and T18 (Fig. 2e). A comparison of CHyMERa-Net scores with those of DeepCpf1, a recently described deep learning algorithm that predicts Cas12a guide activity<sup>36</sup>, revealed comparable performance with the exception that the best CHyMERa-Net-scored guides outperformed those determined by DeepCpf1 (Fig. 2f). These analyses support CHyMERa-Net's scoring approach as a robust and quantitative method for predicting Cas12a guide activity at endogenous loci.

**Gene inactivation using CHyMERa dual-targeting outperforms conventional single-targeting.** Using the Cas12a guide design principles inferred by CHyMERa-Net, we designed an optimized hgRNA library that comprised 58,332 hgRNAs targeting 4,993 genes having the highest expression across five commonly used human cell lines (see Methods); 30,848 combinatorial and single-targeting hgRNAs directed at 1,344 human paralogs and 22 hand-selected gene-gene pairs of interest; and 3,566 control hgRNAs targeting intergenic or exogenous sequences for the assessment of single- versus dual-cutting effects (Supplementary Table 5).

Fitness screens were performed in HAP1 and hTERT-immortalized retinal pigment epithelial (RPE1) cells constitutively expressing Cas9 and Cas12a (Supplementary Fig. 1d). These cell lines were selected based on their differential dependency on *TP53*; while RPE1 cells harbor a wild-type (WT) *TP53* gene, HAP1 cells have a loss-of-function mutation in *TP53* (ref. 37). Quantification of hgRNA abundance showed correlated depletion of hgRNAs targeting core fitness genes compared to controls in both cell lines (Supplementary Fig. 3a and Supplementary Table 6; Spearman's  $\rho = 0.62$ ). Notably, CNN-optimized Cas12a guides (individual Cas12a guides paired with intergenic control guides) were more

efficiently depleted than Cas12a guides tested in the optimization screen (Fig. 3a;  $P=1.4\times 10^{-28}$ , Wilcoxon rank-sum test), confirming that our CNN algorithm identifies more effective guides.

We next assessed whether the combination of Cas9 and Cas12a guides in hgRNAs would result in increased signal in fitness screens (Fig. 3b). We considered that the probability of loss-of-function indel frequencies caused by a single Cas9 or Cas12a gRNA targeting a given gene would be enhanced if a second indel event were introduced in the same gene in the same cell. As controls, we compared LFC distributions for nontargeting and intergenic-targeting hgRNAs. On average, hgRNA constructs targeting intergenic regions show no net LFC (Fig. 3c) although there is a modest correlation between the number of genomic cuts and reduced fitness (Supplementary Fig. 3b), consistent with previous reports<sup>6,38,39</sup>. With these observations in mind, when comparing single- versus dual-targeting of genes in our screens, single-targeting constructs were always paired with an intergenic-targeting guide to control for this effect.

Targeting essential genes with both Cas9 and Cas12 guides via hgRNAs results in significantly higher depletion in both HAP1 (2.8×) and RPE1 (2.6×) cells, compared to targeting with single Cas9 or Cas12 guides in the context of hgRNAs (Welch's two-sample *t*-test,  $P<2.2\times 10^{-16}$ ; Fig. 3c and Supplementary Fig. 3c). It is noteworthy that, despite RPE1 cells harboring a WT *TP53* gene (Supplementary Fig. 3e), the efficiency of targeting core essential genes between these lines is comparable (Fig. 3c). The enhanced dropout of dual-targeted genes appears to be independent of the distance between paired sites (Supplementary Fig. 3d). Importantly, the number of fitness genes identified by dual-targeting substantially exceeds that detected through single-targeting, yielding ~600 and 1,500 additional genes for HAP1 and RPE1 cells, respectively (Fig. 3d,e and Supplementary Table 7).

To independently assess whether these additional genes are important for cell growth, we examined their fitness profiles in the Cancer Dependency Map dataset (DepMap<sup>40</sup>, Supplementary Fig. 3f), which comprises data from CRISPR loss-of-function screens in 558 cancer cell lines performed using a different single-targeting Cas9 gRNA library. The distribution of CERES scores (a measure of essentiality that corrects for copy number alterations at target sites) for core essential genes has a median of -0.96, while genes detected by both single- and dual-targeting have a median CERES score of -0.61. Interestingly, fitness genes identified only by dual-targeting hgRNAs have a distribution with a median of -0.16, which is significantly lower than that of nonessential genes ( $P<0.001$ , Wilcoxon rank-sum test; Supplementary Fig. 3f). Thus, dual-targeted-only genes appear to have distinct, less pronounced fitness defects across a broad range of cell lines, which fail to be detected when targeted by single gRNAs. Collectively, these results indicate that targeting the same gene twice in the same cell using CHyMERa significantly increases the efficiency of gene disruption, and thus improves the sensitivity of detection of genes with relatively modest effects on cell fitness.

**CHyMERa detects digenic interactions.** To assess the efficacy of the multisite-targeting capacity of CHyMERa for mapping GIs, we used CNN-optimized hgRNAs to target pairs of genes involved in a small number of known digenic interactions (Supplementary Tables 5 and 6). These gene pairs were targeted either individually or in combination by Cas9 and Cas12a gRNAs, and LFC values for double-knockouts were compared with the sums of LFCs for single-knockouts (Fig. 4a and see Methods). This screen detected expected GIs between *TP53* and its negative regulators *MDM4* and *MDM2* in RPE1 cells, which express WT *TP53* (Fig. 4b and Supplementary Fig. 4a). However, these interactions were not detected in HAP1 cells, which harbor an expressed but inactive mutant version of *TP53* (*TP53-S215G*)<sup>41</sup> (Fig. 4b and Supplementary Fig. 4a). CHyMERa also

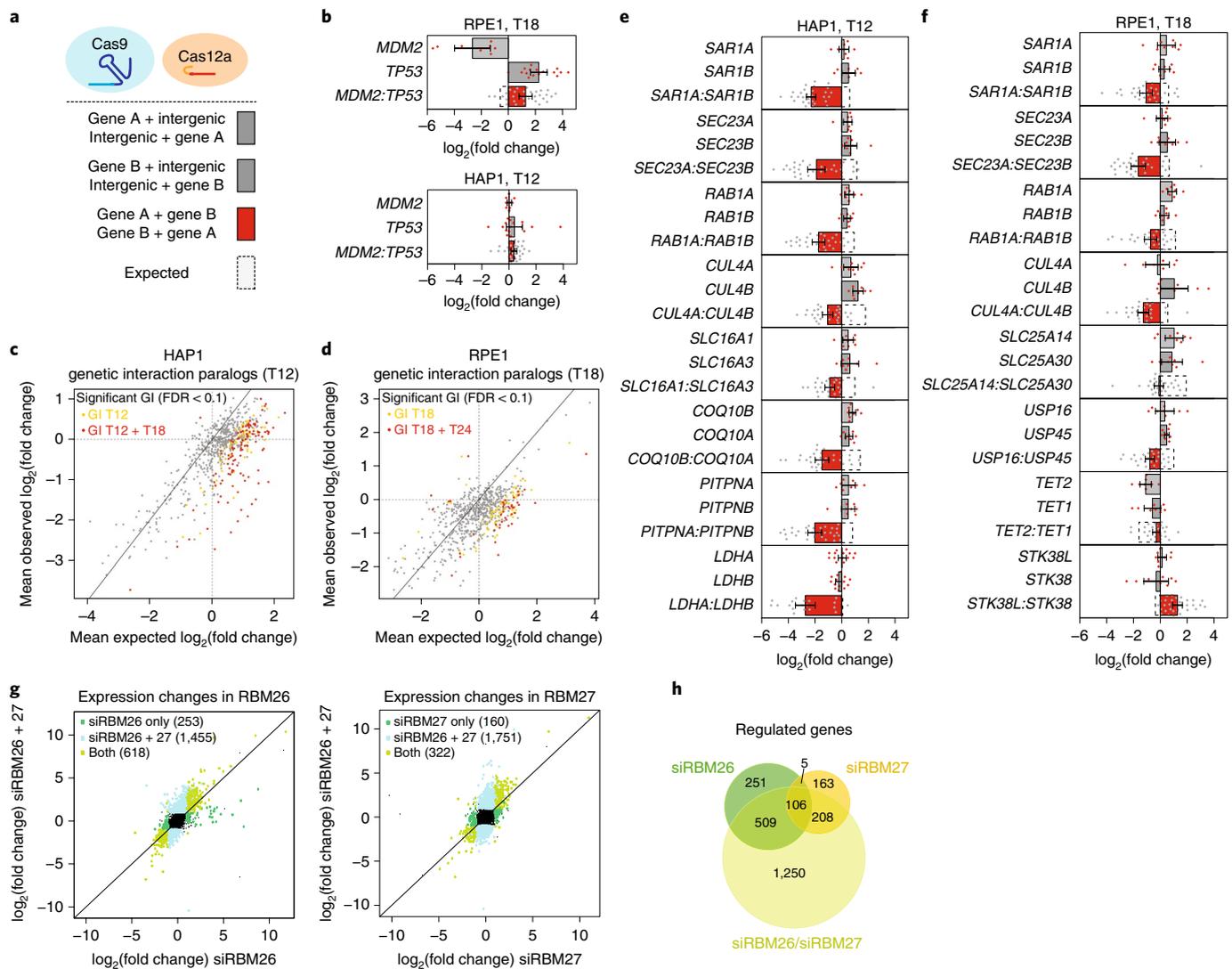
captured known negative GIs between *MCL1* and *BCL2L1* (ref. 18,19), and between *KDM6B* and *BRD4* (ref. 21) (Supplementary Fig. 4b), effectively demonstrating its utility in mapping digenic interactions in mammalian cells.

**CHyMERa screens uncover functional relationships between paralogous genes.** Although genetic redundancy helps ensure phenotypic robustness<sup>42</sup>, it also limits the detection of gene functions using single-gene loss-of-function approaches<sup>43</sup>. We therefore used CHyMERa to systematically target 1,344 genes that represent paralog pairs (excluding gene families with more than two paralogs), and targeted these either individually or in combination using the CNN-optimized hgRNA library described above (Supplementary Table 5). This set of paralogs represents genes that encode proteins involved in a broad range of biological processes (for example, cell cycle, protein trafficking, splicing, protein turnover and modification, and metabolism).

Following the strategy for scoring of hgRNAs targeting genes involved in known digenic interactions described above, we examined the effect of targeting paralogs (single versus combinatorial targeting) in HAP1 and RPE1 cells. Notably, 33% (219 pairs) of tested paralog pairs in HAP1 cells and 18% (122 pairs) in RPE1 cells display a nonadditive fitness phenotype when targeted in combination, as compared to expected phenotypes from targeting each paralog individually (Fig. 4c,d, Supplementary Fig. 4c–f and Supplementary Table 8). As in yeast<sup>44</sup>, the majority of these effects represent negative GIs, although positive interactions are also detected (Fig. 4e,f and Supplementary Fig. 4g,h). Negative GIs include several paralog pairs known to exhibit functional redundancy (for example, *SEC23A–SEC23B*, *ARID1A–ARID1B* and *TIA1–TIAL1* (refs. 45–47), as well as strong yet previously uncharacterized negative interactions (*SARIA–SARIB*, *RAB1A–RAB1B*, *LDHA–LDHB*, *RBM26–RBM27* and *hnRNPf–hnRNPH3*), whereas positive GIs were detected between *STK38–STK38L* and *TET1–TET2* (Fig. 4e,f, Supplementary Fig. 4g,h and Supplementary Table 8). Importantly, GIs between paralogous genes (*LDHA–LDHB*, *SLC16A1–SLC16A3*, *ROCK1–ROCK2* and *SPI1–SPI3*) were also independently validated in HAP1 clonal knockout cell lines, with a clear fitness defect observed in double-knockouts compared to the corresponding single-knockouts (Supplementary Fig. 4i).

A relatively strong GI observed in both HAP1 and RPE1 cells involves paralogous genes encoding the RNA recognition motif-containing proteins RBM26 and RBM27 which, to our knowledge have not previously been characterized. To validate and further investigate the functional interaction between RBM26 and RBM27, we performed single and double small interfering (si)RNA knockdowns (Supplementary Fig. 4j). Depletion of RBM27 has little effect on the proliferation of HAP1 or RPE1 cells, whereas their combined depletion results in a more than additive effect on cell viability (Supplementary Fig. 4k). Moreover, RNA-sequencing (RNA-seq) profiling of HAP1 cells following siRNA knockdown of RBM26 and RBM27 reveals that their co-depletion results in a 72% increase in the number of genes with altered expression compared to that of both single-knockdowns (2,073 versus 1,204 genes,  $P<2.2\times 10^{-16}$ , Fisher's exact test; Fig. 4g,h and Supplementary Table 9). Interestingly, genes downregulated following RBM26 and/or RBM27 co-depletion are enriched in terms related to the cell cycle (Supplementary Fig. 4l and Supplementary Table 10). Collectively, these analyses demonstrate the efficacy of CHyMERa in detecting known and new GIs between pairs of paralogous genes, including a previously unknown interaction between *RBM26* and *RBM27* that shapes the human transcriptome.

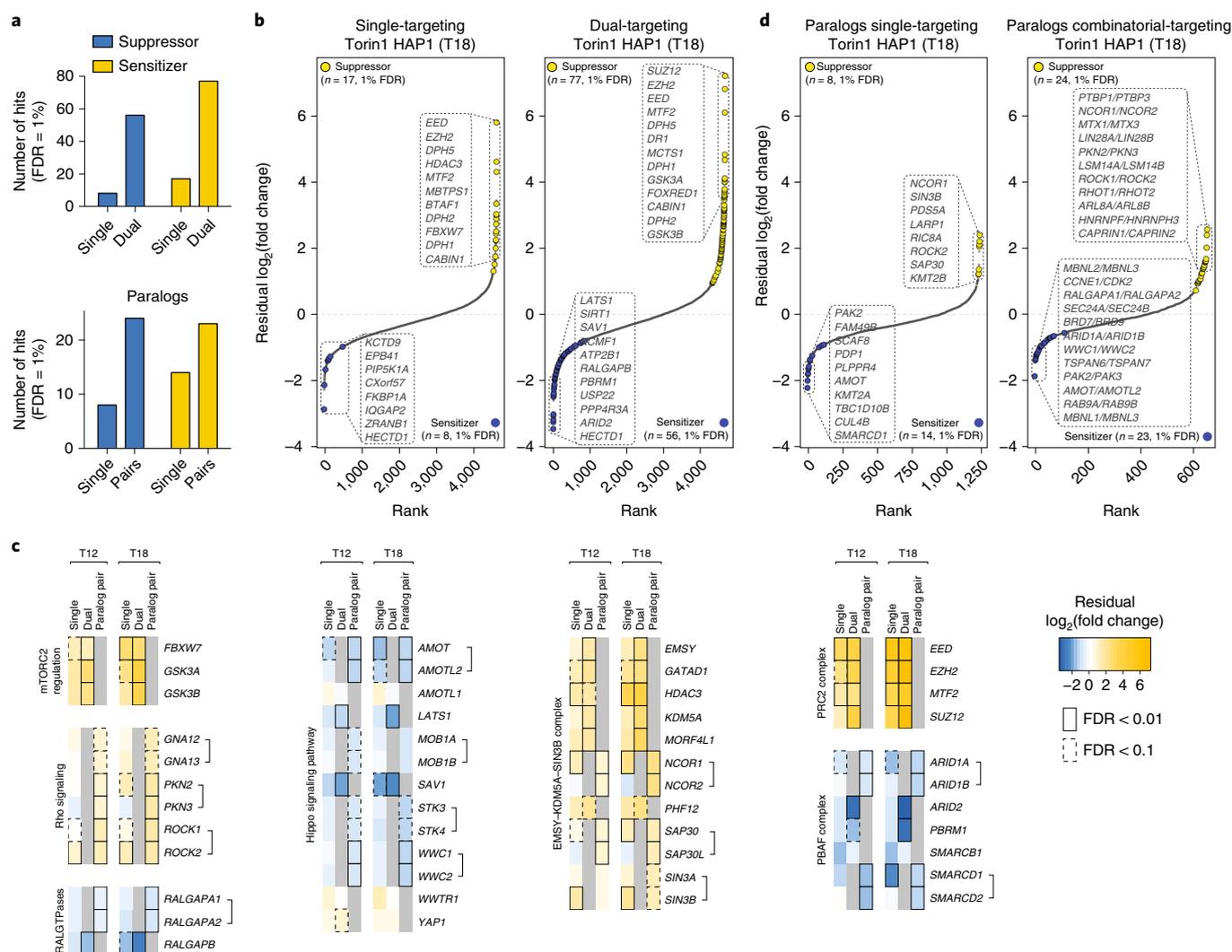
**CHyMERa increases the sensitivity of chemogenetic screens.** Chemical–genetic interactions can uncover molecular mechanisms and targets of drug action. For example, the identification of



**Fig. 4 | Mapping GIs among gene paralogs in human cells.** **a**, Schematic outlining dual-hgRNA designs for interrogation of digenic interactions. **b**, LFC of single- or dual-gene perturbation, as indicated. The expected combinatorial effect size based on single perturbation is indicated with dotted bars. All data are represented as mean  $\pm 2 \times$  s.e.m. derived from three independent experiments with eight hgRNAs targeting *MDM2*, 16 targeting *TP53* and 30 targeting both genes together. **c,d**, Expected versus observed LFC of paralog pairs in HAP1 (**c**) or RPE1 (**d**) cells at the indicated time points. Two-tailed Wilcoxon rank-sum test, Benjamini–Hochberg multiple testing correction,  $n=3$  independent technical replicates. **e,f**, LFC of single or combinatorial gene ablations of paralog pairs in HAP1 (**e**) or RPE1 (**f**) cells at the indicated time points. Bars show mean  $\pm 2 \times$  s.e.m. derived from three independent experiments. Each gene was targeted by eight hgRNA constructs (except *LDHA* and *LDHB*, which were targeted by 16 and 12 hgRNAs, respectively), while the gene pair was targeted with 30 hgRNA constructs (20 for *LDHA:LDHB*). **g**, Expression changes following siRNA-mediated depletion of RBM26 (left) or RBM27 (right) versus RBM26/RBM27 co-depletion in HAP1 cells, as assessed by RNA-seq. Differentially expressed genes were identified using exactTest from the Bioconductor package edgeR, and were defined as those with RPKM > 5, a twofold change compared to control treatment and FDR < 0.05, and are highlighted.  $n=2$  independent biological replicates. **h**, Venn diagram of the number of genes regulated in response to depletion of RBM26, RBM27 or both, as defined above.

genes that impact small-molecule modulation of the mTOR pathway, which plays a central role in regulation of protein synthesis, autophagy and cell growth, is of considerable interest<sup>48,49</sup>. To test the utility of CHyMERa in the detection of small-molecule GIs, we treated HAP1 cells transduced by the dual-gene and paralog-targeting hgRNA library with the catalytic mTOR inhibitor Torin1, which targets mTORC1 and mTORC2 kinase complexes<sup>50</sup>. To identify genes whose depletion alters the response to Torin1, we compared hgRNA LFC distributions with or without drug treatment. This analysis identifies 17 and 8 single-guide-targeted genes as Torin1 suppressors and sensitizers, respectively, whereas dual-targeting identifies 77 suppressors and 56 sensitizers at the same false discov-

ery rate (FDR) (Fig. 5a,b, Supplementary Fig. 5a and Supplementary Table 11). The PRC2 complex member encoded by the *EED* gene scores amongst the top three positive chemical GIs for both single- and dual-targeting hgRNAs (Fig. 5b,c). This finding was confirmed by treating HAP1 WT and *EED* knockout cells with Torin1, where an increased tolerance of mTOR inhibition was observed in *EED*-deficient cells (Supplementary Fig. 5b). We also identify 20 positive and 20 negative chemical GIs amongst paralog pairs (suppressors and sensitizers, respectively), including genes not identified when individually targeted (FDR < 0.01; Fig. 5a,d, Supplementary Fig. 5c and Supplementary Table 11). As examples, the Torin1 sensitivity screen identifies previously described regulators and downstream



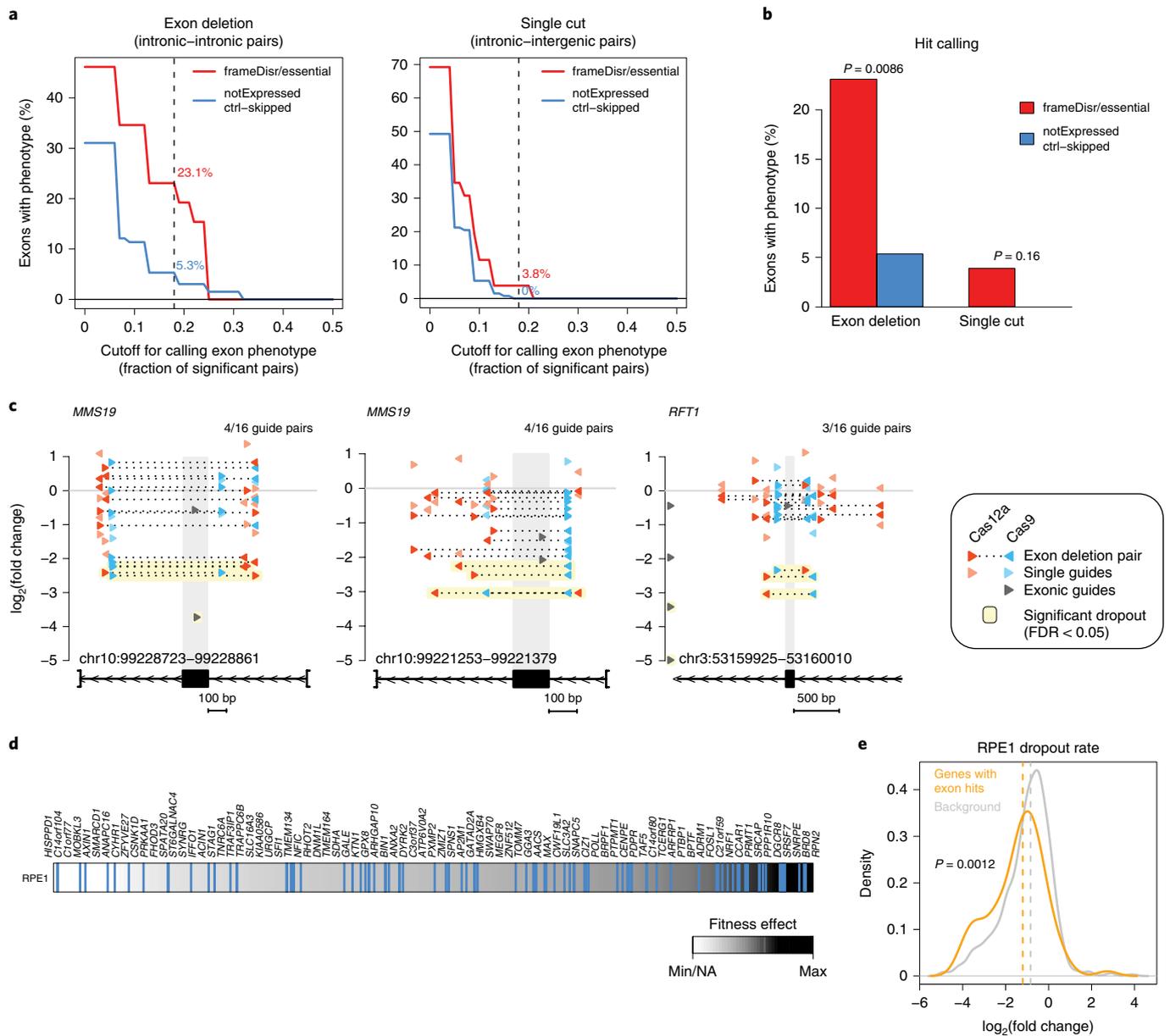
**Fig. 5 | Dual-gene targeting and combinatorial perturbation of paralogs identifies chemogenetic interactions in response to inhibition of mTOR.**

**a**, Numbers of Torin1 sensitizer and suppressor gene hits detected by single- or dual-targeting (top) or using single- or combinatorial targeting of paralogous genes (lower) at the end time point (T18) (FDR < 0.01, two-tailed Wilcoxon rank-sum test with Benjamini-Hochberg multiple testing correction,  $n = 3$  independent technical replicates). **b**, Differential LFC of genes perturbed by single- (left) and dual-targeting (right) hgRNAs following Torin1 treatment in HAP1 cells at the late time point (T18). Sensitizer (blue) and suppressor gene hits (yellow) are highlighted (FDR < 0.01, two-tailed Wilcoxon rank-sum test, Benjamini-Hochberg multiple testing correction,  $n = 3$  independent technical replicates); the top 10, in addition to selected genes from the top 20 significant hits, are listed. **c**, Differential LFC of selected complex members perturbed by single- or dual-targeting hgRNAs, or perturbed in a combinatorial manner as a paralog pair as indicated at the early (T12) and late (T18) time points. Statistical analysis using a two-tailed Wilcoxon rank-sum test with Benjamini-Hochberg multiple testing correction,  $n = 3$  independent technical replicates. **d**, Differential LFC of paralogs perturbed by single- (left) and combinatorial targeting (right) hgRNAs following Torin1 treatment in HAP1 cells at the late time point (T18). Sensitizer (blue) and suppressor gene hits (yellow) are highlighted (FDR < 0.01, Wilcoxon rank-sum test with Benjamini-Hochberg multiple testing correction,  $n = 3$  independent technical replicates); the top 10, and selected genes from the top 20 significant hits, are listed.

effectors of mTOR signaling, including *FBXW7* (ref. <sup>51</sup>), *GSK3A* and *GSK3B* (ref. <sup>52</sup>), Rho signaling components such as *ROCK1* and *ROCK2* (ref. <sup>53</sup>) and RAL GTPases<sup>54</sup> (Fig. 5c). Previously unseen Torin-GIs include genes in the Hippo signaling and diphthamide modification pathways, as well as those in multiple transcriptional and chromatin regulators associated with the EMSY-KDM5A-SIN3B complex, the Polycomb repressive complex 2 (PRC2) and the pBAF complex (Fig. 5c and Supplementary Fig. 5d,e). These data highlight the utility of CHyMERa as an effective tool in the discovery of new chemical GI relationships.

**CHyMERa as a tool for exon-resolution functional genomics.** We next applied CHyMERa to the large-scale screening of exon function.

To this end, we designed a CNN-optimized hgRNA library targeting 2,157 alternative cassette exons for deletion in RPE1 cells, using multiple hgRNAs directed at flanking intronic sequences. Furthermore, each intronic Cas9 and Cas12a gRNA was also paired with intergenic gRNAs to control for nonspecific toxicity. The library further included Cas9 gRNAs directly targeting sequences within constitutive exons, to assess the phenotypic impact of inactivation of genes harboring targeted alternative exons (see Methods and Supplementary Table 12). The analyzed alternative exons are detected in transcripts expressed across a panel of human cell lines and belong to functionally diverse genes with a range of fitness profiles and conservation levels (Supplementary Table 12, and see Methods). Among the targeted exons, 132 are frame-altering and



**Fig. 6 | Identification of fitness exons in RPE1 cells using an exon-targeting ChyMERa library.** **a**, Percentage of interrogated alternative exons having a fitness phenotype across increasing fractions of significant exon deletion intronic-intronic (left,  $n = 4,196$ ) or intronic-intergenic (right,  $n = 4,180$ ) hgRNA pairs targeting each exon (refer to Methods for details of significance scoring). **b**, Percentage of exons with a significant phenotype, defined as those with at least 18% of targeting hgRNA constructs showing significant depletion, in essential or nonessential genes (exon deletion,  $P = 0.02$ ,  $n = 26$ ; single cut,  $P = 0.16$ ,  $n = 132$ ; both, two-sided Fisher's exact test). **c**, All hgRNA constructs targeting frame-disruptive exons in *MMS19* or *RFT1* are shown above the gene model (x axis) with the observed LFC value for each hgRNA (y axis). Exon deletion (intronic-intronic), single-targeting control (intronic-intergenic) and exon-targeting (exonic-intergenic) hgRNAs are indicated in different colours, while significantly depleted hgRNAs are highlighted and numbers of significant and total exon deletion pairs are indicated in the top right corner. Significant depletion was scored against the empirical null distribution of 1,647 intergenic-intergenic control pairs (refer to Methods for details). **d**, Visualization of frame-preserving alternative exons with a fitness phenotype. All exons targeted in the library were ranked based on the mean LFC of exonic guides targeting the corresponding genes, and genes that contain fitness exons are indicated. NA, not available. **e**, Mean LFC of hgRNAs targeting exonic regions in genes containing alternative exons interrogated in the library. Genes with exons identified as significant screen hits are indicated (two-tailed Mann-Whitney  $U$ -test,  $P = 0.00012$ ;  $n = 77$  for hits and  $n = 1,019$  for background).

predicted to result in gene inactivation via truncation of the coding sequence, or else introduction of a premature termination codon with potential to elicit nonsense-mediated messenger RNA decay. The frame-altering category includes exons in both fitness and nonfitness genes, affording a comparative measure of the efficiency of hgRNAs in directing exon deletion and guide depletion in cell fitness screens.

To assess the efficiency of exon deletion, we first determined which guide pairs display significant dropout or enrichment compared to intergenic-intergenic control guide pairs. We then scored the percentage of targeted frame-disrupting exons in fitness and nonfitness genes based on the fraction of significantly depleted guide pairs. As expected, among the guide pairs displaying a significant dropout phenotype there is a strong enrichment for

frame-disruptive exons residing in fitness genes compared to exons in nonfitness genes (Fig. 6a–c and Supplementary Fig. 6a;  $P < 0.001$ , Fisher's exact test). This enrichment is not detected for single-cutting intronic–intergenic control guide pairs (Fig. 6a–b). The strongest separation (~4.5-fold) between fitness and nonfitness genes is observed with exons for which there is a significant dropout of at least 18% of tested hgRNA exon-deletion pairs (Fig. 6a–c). These results demonstrate that CHyMERa is capable of interrogating the phenotypic consequences of exon deletion in the context of large-scale dropout screens.

**CHyMERa reveals alternative exons that impact cell fitness.** We next used CHyMERa to investigate the consequences of deleting frame-preserving cassette exons on cell fitness in RPE1 cells. Of 2,025 frame-preserving cassette exons targeted for deletion, 124 resulted in significant depletion of guides (Fig. 6d and Supplementary Table 13). These 'fitness exons' are significantly enriched in essential genes (Fig. 6e;  $P < 0.00012$ , Mann–Whitney  $U$ -test) and, according to RNA-seq analysis, display a significantly skewed distribution towards increased inclusion levels (Supplementary Fig. 6b;  $P = 0.012$ , Mann–Whitney  $U$ -test). However, we do not detect apparent differences between exons impacting fitness versus those that do not when comparing exon length or overlap with annotated functional domains (Supplementary Fig. 6c,d). Validating the specificity of CHyMERa for exon deletion, several hgRNAs that produce strong phenotypes in the screen are also found to have greater editing efficiency relative to hgRNAs targeting the same exons but with marginal LFC values (Supplementary Fig. 6e).

Among the frame-preserving exons impacting cellular fitness is exon 12A of the *BIN1* gene (Fig. 6d and Supplementary Fig. 6f). *BIN1* is a tumor suppressor that interacts with MYC and inhibits MYC-dependent transformation<sup>55</sup>. Our observation is in agreement with a previous study linking exon 12A to cell proliferation by abolishing a *BIN1*–MYC interaction<sup>56</sup>, and aberrant splicing of this exon has been observed in melanoma cells<sup>57</sup>. In addition to capturing alternative exons with known roles in proliferation, our screen reveals alternative exons with unexpected roles in cell fitness (Fig. 6d and Supplementary Fig. 6f) and provides a resource for prioritization of alternative splicing events for future investigation. Overall, the results demonstrate the utility of CHyMERa as an effective method for systematic investigation of the function of alternative exons when coupled with an appropriate biological assay.

## Discussion

Identifying GIs and the roles of gene segments is crucial in the advancement of knowledge of gene function and how genome alterations contribute to diseases and disorders<sup>58</sup>. Pioneering studies using budding yeast have generated global GI networks and wiring diagrams of cellular function<sup>44,59</sup>. However, systematic mapping of GIs in mammalian cells has been hampered by the lack of efficient and readily scalable targeting systems<sup>18–22,45,60</sup>. Similarly, while it is well established that disruption of splicing regulators and individual exons can significantly impact development and lead to diseases and disorders<sup>61–63</sup>, there is a lack of efficient gene targeting systems for the systematic functional interrogation of exons, or other gene regions. The results in the present study support the efficacy of CHyMERa as a system for addressing these timely challenges.

Previously described CRISPR-based multitargeting strategies coexpress pairs of Cas9 guides<sup>18,20,22,60,64,65</sup>. However, these approaches have limited efficiency, especially when attempting synchronous targeting (for example, for gene segment deletion) (Supplementary Fig. 1a,b). More recent systems employing orthologous *Staphylococcus aureus* Cas9 and Sp-Cas9 enzymes<sup>19,66</sup>, like CHyMERa, have increased editing efficiency, possibly due to reduced recombination through the use of different transactivating CRISPR RNAs. However, CHyMERa is on average more efficient in

regard to gene segment deletion (Supplementary Fig. 7a,b), probably due to its unique feature of affording coexpression of multiple (up to four) machine learning-optimized gRNAs from a single hgRNA, which allows synchronous dsDNA breaks. Consistently, recent studies employing Cas12a-processed polycistronic guides for multitargeting displayed increased editing efficiency as compared to the use of single Cas12a gRNAs<sup>67,68</sup>. Moreover, the combination of Cas9 and Cas12a systems leverages twice the number of possible targeting sites in comparison to the use of either enzyme alone—that is, the human genome contains 227.3 M Cas9 (NGG) and 207.6 M Cas12a (TTTV) candidate PAM target sites (Supplementary Fig. 7c). Engineering Cas9, Cas12a or other Cas enzymes for increased editing efficiency<sup>69,70</sup> and target specificity is expected to further expand the targeting landscape of the CHyMERa system. Furthermore, CHyMERa can be combined with multiple effector domains<sup>5,71</sup> to afford more complex assays, such as interrogating the effects of simultaneous inactivation and activation of different sets of genes in the same cells.

In conclusion, CHyMERa represents an efficient and versatile system for the combinatorial perturbation of genetic elements in mammalian cells. As such, we anticipate its future use in charting GIs and the functions of genome segments, such as the myriad of previously uncharacterized alternative splicing events and noncoding RNA sequences linked to development and disease.

## 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/s41587-020-0437-z>.

Received: 26 May 2019; Accepted: 27 January 2020;

Published online: 16 March 2020

## References

- Cong, L. et al. Multiplex genome engineering using CRISPR/Cas systems. *Science* **339**, 819–823 (2013).
- Jinek, M. et al. A programmable dual-RNA-guided DNA endonuclease in adaptive bacterial immunity. *Science* **337**, 816–821 (2012).
- Mali, P. et al. RNA-guided human genome engineering via Cas9. *Science* **339**, 823–826 (2013).
- Wright, A. V., Nuñez, J. K. & Doudna, J. A. Biology and applications of CRISPR systems: harnessing nature's toolbox for genome engineering. *Cell* **164**, 29–44 (2016).
- Doench, J. G. Am I ready for CRISPR? A user's guide to genetic screens. *Nat. Rev. Genet.* **19**, 67–80 (2018).
- Hart, T. et al. High-resolution CRISPR screens reveal fitness genes and genotype-specific cancer liabilities. *Cell* **163**, 1515–1526 (2015).
- Shalem, O. et al. Genome-scale CRISPR-Cas9 knockout screening in human cells. *Science* **343**, 84–87 (2014).
- Wang, T. et al. Identification and characterization of essential genes in the human genome. *Science* **350**, 1096–1101 (2015).
- Wang, T., Wei, J. J., Sabatini, D. M. & Lander, E. S. Genetic screens in human cells using the CRISPR-Cas9 system. *Science* **343**, 80–84 (2014).
- Liu, G. et al. Data independent acquisition analysis in ProHits 4.0. *J. Proteomics* **149**, 64–68 (2016).
- Patel, S. J. et al. Identification of essential genes for cancer immunotherapy. *Nature* **548**, 537–542 (2017).
- Park, R. J. et al. A genome-wide CRISPR screen identifies a restricted set of HIV host dependency factors. *Nat. Genet.* **49**, 193–203 (2016).
- Haney, M. S. et al. Identification of phagocytosis regulators using magnetic genome-wide CRISPR screens. *Nat. Genet.* **50**, 1716–1727 (2018).
- Gonatopoulos-Pournatzis, T. et al. Genome-wide CRISPR-Cas9 interrogation of splicing networks reveals a mechanism for recognition of autism-misregulated neuronal microexons. *Mol. Cell* **72**, 510–524.e12 (2018).
- 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).
- Mair, B. et al. Essential gene profiles for human pluripotent stem cells identify uncharacterized genes and substrate dependencies. *Cell Rep.* **27**, 599–615.e12 (2019).

17. Lynch, M. & Conery, J. S. The evolutionary fate and consequences of duplicate genes. *Science* **290**, 1151–1155 (2000).
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).
19. Najm, F. J. et al. Orthologous CRISPR–Cas9 enzymes for combinatorial genetic screens. *Nat. Biotechnol.* **36**, 179–189 (2017).
20. Zhu, S. et al. Genome-scale deletion screening of human long non-coding RNAs using a paired-guide RNA CRISPR–Cas9 library. *Nat. Biotechnol.* **34**, 1279–1286 (2016).
21. Wong, A. S. L. et al. Multiplexed barcoded CRISPR–Cas9 screening enabled by CombiGEM. *Proc. Natl Acad. Sci. USA* **113**, 2544–2549 (2016).
22. Shen, J. P. et al. Combinatorial CRISPR–Cas9 screens for de novo mapping of genetic interactions. *Nat. Methods* **14**, 573–576 (2017).
23. Brake, Oter et al. Lentiviral vector design for multiple shRNA expression and durable HIV-1 inhibition. *Mol. Ther.* **16**, 557–564 (2008).
24. Vidigal, J. A. & Ventura, A. Rapid and efficient one-step generation of paired gRNA CRISPR–Cas9 libraries. *Nat. Commun.* **6**, 8083 (2015).
25. Sack, L. M., Davoli, T., Xu, Q., Li, M. Z. & Elledge, S. J. Sources of error in mammalian genetic screens. *G3 (Bethesda)* **6**, 2781–2790 (2016).
26. Adamson, B. et al. A multiplexed single-cell CRISPR screening platform enables systematic dissection of the unfolded protein response. *Cell* **167**, 1867–1882.e21 (2016).
27. Zetsche, B. et al. Cpf1 is a single RNA-guided endonuclease of a class 2 CRISPR–cas system. *Cell* **163**, 759–771 (2015).
28. Zetsche, B. et al. Multiplex gene editing by CRISPR–Cpf1 using a single crRNA array. *Nat. Biotechnol.* **35**, 31–34 (2016).
29. Fonfara, I., Richter, H., Bratovič, M., Le Rhun, A. & Charpentier, E. The CRISPR-associated DNA-cleaving enzyme Cpf1 also processes precursor CRISPR RNA. *Nature* **532**, 517–521 (2016).
30. Chow, R. D. et al. In vivo profiling of metastatic double knockouts through CRISPR–Cpf1 screens. *Nat. Methods* **16**, 405 (2019).
31. Xu, H. et al. Sequence determinants of improved CRISPR sgRNA design. *Genome Res.* **25**, 1147–1157 (2015).
32. Doench, J. G. et al. Optimized sgRNA design to maximize activity and minimize off-target effects of CRISPR–Cas9. *Nat. Biotechnol.* **34**, 184–191 (2016).
33. Hart, T. et al. Evaluation and design of genome-wide CRISPR/SpCas9 knockout screens. *G3 (Bethesda)* **7**, 2719–2727 (2017).
34. Listgarten, J. et al. Prediction of off-target activities for the end-to-end design of CRISPR guide RNAs. *Nat. Biomed. Eng.* **2**, 38–47 (2018).
35. Kim, H. K. et al. In vivo high-throughput profiling of CRISPR–Cpf1 activity. *Nat. Methods* **14**, 153–159 (2017).
36. Kim, H. K. et al. Deep learning improves prediction of CRISPR–Cpf1 guide RNA activity. *Nat. Biotechnol.* **36**, 239–241 (2018).
37. Moder, M. et al. Parallel genome-wide screens identify synthetic viable interactions between the BLM helicase complex and Fanconi anemia. *Nat. Commun.* **8**, 1238 (2017).
38. Bertomeu, T. et al. A high-resolution genome-wide CRISPR/Cas9 viability screen reveals structural features and contextual diversity of the human cell-essential proteome. *Mol. Cell Biol.* **38**, e00302–17 (2017).
39. Morgens, D. W. et al. Genome-scale measurement of off-target activity using Cas9 toxicity in high-throughput screens. *Nat. Commun.* **8**, 15178 (2017).
40. Meyers, R. M. et al. Computational correction of copy number effect improves specificity of CRISPR–Cas9 essentiality screens in cancer cells. *Nat. Genet.* **49**, 1779–1784 (2017).
41. Slovackova, J., Smarda, J. & Smardova, J. Roscovitine-induced apoptosis of H1299 cells depends on functional status of p53. *Neoplasma* **59**, 606–612 (2012).
42. Gu, Z. et al. Role of duplicate genes in genetic robustness against null mutations. *Nature* **421**, 63–66 (2003).
43. Ewen-Campen, B., Mohr, S. E., Hu, Y. & Perrimon, N. Accessing the phenotype gap: enabling systematic investigation of paralog functional complexity with CRISPR. *Dev. Cell* **43**, 6–9 (2017).
44. Costanzo, M. et al. A global genetic interaction network maps a wiring diagram of cellular function. *Science* **353**, aaf1420 (2016).
45. Bassik, M. C. et al. A systematic mammalian genetic interaction map reveals pathways underlying ricin susceptibility. *Cell* **152**, 909–922 (2013).
46. Viswanathan, S. R. et al. Genome-scale analysis identifies paralog lethality as a vulnerability of chromosome 1p loss in cancer. *Nat. Genet.* **50**, 937–943 (2018).
47. Meyer, C. et al. The TIA1 RNA-binding protein family regulates EIF2AK2-mediated stress response and cell cycle progression. *Mol. Cell* **69**, 622–635.e6 (2018).
48. Saxton, R. A. & Sabatini, D. M. mTOR signaling in growth, metabolism, and disease. *Cell* **168**, 960–976 (2017).
49. Valvezan, A. J. & Manning, B. D. Molecular logic of mTORC1 signalling as a metabolic rheostat. *Nat. Metab.* **1**, 321–333 (2019).
50. Thoreen, C. C. et al. An ATP-competitive mammalian target of rapamycin inhibitor reveals rapamycin-resistant functions of mTORC1. *J. Biol. Chem.* **284**, 8023–8032 (2009).
51. Koo, J., Yue, P., Deng, X., Khuri, F. R. & Sun, S.-Y. mTOR complex 2 stabilizes Mcl-1 protein by suppressing its glycogen synthase kinase 3-dependent and SCF-FBXW7-mediated degradation. *Mol. Cell Biol.* **35**, 2344–2355 (2015).
52. Koo, J., Yue, P., Gal, A. A., Khuri, F. R. & Sun, S.-Y. Maintaining glycogen synthase kinase-3 activity is critical for mTOR kinase inhibitors to inhibit cancer cell growth. *Cancer Res.* **74**, 2555–2568 (2014).
53. Shu, L. & Houghton, P. J. The mTORC2 complex regulates terminal differentiation of C2C12 myoblasts. *Mol. Cell Biol.* **29**, 4691–4700 (2009).
54. Martin, T. D. et al. Ral and Rheb GTPase activating proteins integrate mTOR and GTPase signaling in aging, autophagy, and tumor cell invasion. *Mol. Cell* **53**, 209–220 (2014).
55. Sakamuro, D., Elliott, K. J., Wechsler-Reya, R. & Prendergast, G. C. BIN1 is a novel MYC-interacting protein with features of a tumour suppressor. *Nat. Genet.* **14**, 69–77 (1996).
56. Pineda-Lucena, A. et al. A structure-based model of the c-Myc/Bin1 protein interaction shows alternative splicing of Bin1 and c-Myc phosphorylation are key binding determinants. *J. Mol. Biol.* **351**, 182–194 (2005).
57. Ge, K. et al. Mechanism for elimination of a tumor suppressor: aberrant splicing of a brain-specific exon causes loss of function of Bin1 in melanoma. *Proc. Natl Acad. Sci. USA* **96**, 9689–9694 (1999).
58. Ashworth, A., Lord, C. J. & Reis-Filho, J. S. Genetic interactions in cancer progression and treatment. *Cell* **145**, 30–38 (2011).
59. Costanzo, M. et al. Global genetic networks and the genotype-to-phenotype relationship. *Cell* **177**, 85–100 (2019).
60. Horlbeck, M. A. et al. Mapping the genetic landscape of human cells. *Cell* **174**, 953–967.e22 (2018).
61. Dvinge, H., Kim, E., Abdel-Wahab, O. & Bradley, R. K. RNA splicing factors as oncoproteins and tumour suppressors. *Nat. Rev. Cancer* **16**, 413–430 (2016).
62. Quesnel-Vallières, M., Weatheritt, R. J., Cordes, S. P. & Blencowe, B. J. Autism spectrum disorder: insights into convergent mechanisms from transcriptomics. *Nat. Rev. Genet.* **20**, 51–63 (2019).
63. Scotti, M. M. & Swanson, M. S. RNA mis-splicing in disease. *Nat. Rev. Genet.* **17**, 19–32 (2016).
64. Zhao, D. et al. Combinatorial CRISPR–Cas9 metabolic screens reveal critical redox control points dependent on the KEAP1–NRF2 regulatory axis. *Mol. Cell* **69**, 699–708.e7 (2018).
65. Gasperini, M. et al. CRISPR/Cas9-mediated scanning for regulatory elements required for HPRT1 expression via thousands of large, programmed genomic deletions. *Am. J. Hum. Genet.* **101**, 192–205 (2017).
66. Boettcher, M. et al. Dual gene activation and knockout screen reveals directional dependencies in genetic networks. *Nat. Biotechnol.* **36**, 170–178 (2018).
67. Liu, J. et al. Pooled library screening with multiplexed Cpf1 library. *Nat. Commun.* **10**, 3144 (2019).
68. Campa, C. C., Weisbach, N. R., Santinha, A. J., Incarnato, D. & Platt, R. J. Multiplexed genome engineering by Cas12a and CRISPR arrays encoded on single transcripts. *Nat. Methods* **16**, 887–893 (2019).
69. Sanson, K. R. et al. Optimization of AsCas12a for combinatorial genetic screens in human cells. Preprint at *bioRxiv* <https://doi.org/10.1101/747170> (2019).
70. Kleinstiver, B. P. et al. Engineered CRISPR–Cas12a variants with increased activities and improved targeting ranges for gene, epigenetic and base editing. *Nat. Biotechnol.* **37**, 276–282 (2019).
71. Joung, J. et al. Genome-scale CRISPR–Cas9 knockout and transcriptional activation screening. *Nat. Protoc.* **12**, 828–863 (2017).

**Publisher's note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

© The Author(s), under exclusive licence to Springer Nature America, Inc. 2020

## Methods

**Cell culture methods.** HAP1 cells were obtained from Horizon Genomics (clone C631; sex: male with lost Y chromosome; RRID: CVCL\_Y019). hTERT-RPE1 or RPE1 cells were obtained from ATCC (catalog. no. CRL-4000). Neuro-2A (N2A) cells were obtained from ATCC (catalog. no. CCL-131). Mouse CGR8 embryonic stem cells were obtained from the European Collection of Authenticated Cell Cultures. Human HAP1 cells were maintained in low-glucose (10 mM), low-glutamine (1 mM) DMEM (Wisent, no. 319-162-CL) supplemented with 10% FBS (Life Technologies) and 1% penicillin/streptomycin (Life Technologies). Human hTERT-RPE1 cells were maintained in DMEM with high glucose and pyruvate (Life Technologies) supplemented with 10% FBS (Life Technologies) and 1% penicillin/streptomycin (Life Technologies). Mouse neuroblastoma Neuro-2A (N2A) cells were grown in DMEM (high-glucose; Sigma-Aldrich) supplemented with 10% FBS, sodium pyruvate, nonessential amino acids and penicillin/streptomycin. CGR8 mouse embryonic stem cells were grown in gelatin-coated plates in GMEM supplemented with 100  $\mu$ M  $\beta$ -mercaptoethanol, 0.1 mM nonessential amino acids, 2 mM sodium pyruvate, 2 mM L-glutamine, 5,000 units ml<sup>-1</sup> of penicillin/streptomycin, 1,000 units ml<sup>-1</sup> of recombinant mouse LIF (all Life Technologies) and 15% ES fetal calf serum (ATCC). Cells were maintained at subconfluent conditions. Cells were dissociated using Trypsin (Life Technologies), and all were maintained at 37 °C and 5% CO<sub>2</sub>. Cells were regularly monitored for mycoplasma infection.

**Lenti-Cas12a vector construction.** A nucleoplasmic nuclear localization signal (NLS) was added at the C terminus of an N-terminal SV40 NLS-tagged Cas12a followed by a Myc tag using conventional restriction enzyme cloning, to generate As- or Lb-Cas12a-NLS-MYC-2A-NeoR lentiviral-based expression vectors named plenti-As-Cas12a-2xNLS and plenti-Lb-Cas12a-2xNLS, respectively.

**pLCHKO hgRNA vector construction.** The pLCHKO vector for hgRNA expression was derived from the pLCKO vector (Addgene, no. 73311) by inverting the U6 expression cassette consisting of a stuffer sequence containing BfuAI/BvuI sites followed by a RNA polymerase III transcription termination signal (AAAAAAA). Cloning of hgRNAs into the vector was performed as follows. First, the pLCHKO vector was digested with BsmBI, and oligos containing the Cas9 and Cas12a guides separated by a 32-nt spacer containing BsmBI/Esp31 sites were ligated. Separately, the tracrRNA-Direct Repeat (DR) fragment was cloned into a TOPO vector by annealing and ligating oligos encoding BsmBI-tracrRNA-DR-BsmBI.

BsmBI-tracrRNA-Lb-Cas12a\_DR-BsmBI:  
 5'-cgctctGTTTCAGAGCTATGCTGGAAACAGCATAG-  
 CAAGTTGAAATAAGGCTAG  
 TCCGTTATCAACTGAAAAAGTGGCACCAGTCCGGTGCTTAATTTCT-  
 TACTAAGTGTAGATagagacg-3'  
 BsmBI-tracrRNA-As-Cas12a\_DR-BsmBI:  
 5'-cgctctGTTTCAGAGCTATGCTGGAAACAGCATAG-  
 CAAGTTGAAATAAGGCTAG  
 TCCGTTATCAACTGAAAAAGTGGCACCAGTCCGGTGCTTAATTTCT  
 ACTCTGTAGATagagacg-3'

BsmBI restriction recognition sites are in lower case and underlined, optimized tracrRNA<sup>72</sup> are in upper case and Cas12 DR is in upper case and bold.

In the second cloning step, the 32-nt spacer was excised from pLCHKO vectors containing the dual guides using BsmBI, and the Cas9 tracrRNA-Cas12a DR fragment was ligated (see also Supplementary Fig. 2a).

pPapi constructs were cloned using oligos (generated by Twist Biosciences) as described previously<sup>19</sup>.

**Cell line generation.** Previously generated HAP1 and hTERT-RPE1 clonal cell lines expressing Cas9<sup>6,33</sup> were transduced with lentivirus carrying the As- or Lb-Cas12a-2A-NeoR expression cassette, and transduced cells were selected with G418 (500  $\mu$ g ml<sup>-1</sup>) for 2 weeks. HAP1 and RPE1 Cas9-Cas12a cells were not subjected to single-cell isolation but were used as pools in CHyMERa screens. HAP1 Cas9-Cas12a cells became diploid during the selection process, as determined by ploidy analysis using flow cytometry.

Neuro-2A and CGR8 cells were transduced with lentivirus carrying the Cas9-2A-BlasticidinR-expressing cassette (Addgene, no. 73310) and selected with blasticidin (10  $\mu$ g ml<sup>-1</sup> for N2A and 6  $\mu$ g ml<sup>-1</sup> for CGR8) for 10 d. Cas9-expressing cell lines were then transduced with lentivirus carrying the As- or Lb-Cas12a-2A-NeoR expression cassette and selected with G418 (500  $\mu$ g ml<sup>-1</sup>). After 14 d of selection, N2A single cells were sorted by manual seeding of a single-cell suspension at 0.6 cells per well in 96-well plates. A cell clone with high editing efficiency was selected for subsequent CHyMERa screens. CGR8 Cas9-Cas12a cells were not subjected to single-cell isolation but instead were used as pools in CHyMERa screens.

**Toxicity assays using thymidine or 6-TG.** To determine Cas9 and Cas12a editing efficiency, HAP1 and RPE1 cells expressing Cas9 and Cas12a were transduced with hgRNAs targeting *TK1* (by Cas9) and *HPRT1* (by Cas12a). After selection for transduced cells using 1  $\mu$ g ml<sup>-1</sup> of puromycin for 2 d, cells were reseeded and, after

18 h, treated with either 2.5 mM thymidine or 6  $\mu$ M 6-TG, or mock treated for 4 d. 6-TG results in cell death whereas thymidine block causes cell cycle arrest. As such, both drugs strongly affect cell fitness. Cell viability was assessed by alamarBlue staining.

**siRNA transfections and cell viability assays for RBM26-RBM27 interaction.** HAP1 and RPE1 cell lines were transfected with 10 nM of siGENOME siRNA pools targeting *RBM26* and *RBM27* (Dharmacon) using RNAiMax (Life Technologies). A nontargeting siRNA pool was used as control. Cells were harvested 48 h post-transfection for RNA extraction. For cell viability assays, knockdown was performed for 72 h and viability was assessed by alamarBlue.

**Validation of Torin1-EED chemogenetic interaction.** For validation of a Torin1 suppressor, HAP1 WT and *EED* knockout cells were treated with Torin1 at concentrations ranging from 0 to 100 nM for 4 d. Cell viability was measured post treatment using alamarBlue, and half-maximal inhibitory concentration values were calculated using GraphPad Prism software.

**Validation of GIs between paralog pairs.** HAP1 parental and knockout clones were transduced with lentiviruses derived from lentiCRISPRv2 Cas9 and gRNA expression cassettes targeting either an intergenic site in the AAVS1 locus or the corresponding paralog pair. Each gene was targeted with two independent gRNAs. Twenty-four hours after transduction, cells were selected with 1  $\mu$ g ml<sup>-1</sup> of puromycin for 48 h and seeded for proliferation assay. Cell viability was measured by alamarBlue after 6 d. Average viability of cells transduced with the two gene-targeting gRNAs was calculated and normalized to the intergenic control gRNAs.

**Assessment of Cas9-Cas12a editing by PCR.** To determine Cas9 and Cas12a editing efficiency, cells expressing Cas9 and Cas12a were transduced with lentiviruses derived from dual pLCKO (see Supplementary Fig. 1a), pLCHKO or pPapi constructs targeting intronic regions flanking exons. Transduced cells were selected with 1  $\mu$ g ml<sup>-1</sup> of puromycin for 48 h, and gDNA was extracted using the PureLink Genomic DNA Kit (Thermo Fisher Scientific). Successful editing was assessed by PCR using primers flanking the targeted regions, and PCR products were resolved by agarose gel electrophoresis.

Percentage exon deletion was calculated using ImageJ software. Exon-included and -excluded band intensities were corrected by subtracting the background, and values were normalized by product size. Intensity of the exon-included band was divided by the sum of the exon-included and -excluded bands; the result was then multiplied by 100 to obtain percentage exon deletion, which was rounded to the nearest integer.

**Immunofluorescence.** Cells were seeded on coverslips and fixed with 4% paraformaldehyde in PBS for 10 min at room temperature. Cells were then permeabilized with 1% NP-40 in antibody dilution solution (PBS, 0.2% BSA, 0.02% sodium azide) for 10 min and blocked with 1% goat serum for 45 min. Cells were incubated with anti-HA (1:1,000, Sigma) and anti-Myc antibodies (1:1,000, Sigma, no. M4439) for 1 h at room temperature. Subsequently, cells were incubated with Alexa Fluor488 goat anti-rabbit antibodies (1:500, Invitrogen, no. A-1108) and counterstained with 1  $\mu$ g ml<sup>-1</sup> of DAPI (Cell Signaling, no. 4083S) for 45 min in the dark. Finally, cells were visualized by confocal microscopy (WaveFX confocal microscope, Quorum Technologies).

**Immunoblotting.** Cells were lysed in buffer F (10 mM Tris pH 7.05, 50 mM NaCl, 30 mM Na pyrophosphate, 50 mM NaF, 10% glycerol, 0.5% Triton X-100) and centrifuged at 14,000 r.p.m. for 10 min. Supernatant was collected and protein concentration was determined using Bradford reagent (BioRad). Protein (10–25  $\mu$ g) was resolved on 4–12% Bis-Tris gels (Life Technologies) by electrophoresis and then transferred to Immobilon-P nitrocellulose membrane (Millipore) at 66 V for 90 min. Subsequently, proteins were detected using the following antibodies: anti- $\beta$ -actin (1:10,000, Abcam, no. ab8226), anti-Cas9 (1:4,000, Diagenode, no. C15200229), anti-Cpf1 (1:1,000, Sigma, no. SAB4200777), anti-P53 (1:2,000, Life Technologies, no. AHO0152), anti-pRb S807/811 (1:500, Cell Signaling, no. 9308), anti-p21 (1:500, Cell Signaling, no. 2946) or anti-Myc (1:1,000, Sigma, no. M4439). After binding with HRP-conjugated secondary antibodies (1:5,000, anti-Mouse Jackson ImmunoResearch, no. 715-035-151; anti-Rabbit, Cell Signaling Technology, no. 7074), proteins were visualized on X-ray film using Super Signal chemiluminescence reagent (Thermo Fisher Scientific).

**Cas12a RNA processing activity.** HAP1 cells expressing both Cas9 and Cas12a or Cas9 alone were transduced with a lentiviral hgRNA expression cassette. RNA was extracted using TRIzol (Thermo Fisher Scientific). Subsequently, RNA was converted to complementary DNA using a Maxima H cDNA synthesis kit (Thermo Fisher Scientific) and random primers. Total and unprocessed Cas9 and Cas12a guides were amplified and quantified by quantitative PCR using a SensiFAST real-time PCR kit (Bioline). Full-length (unprocessed) hgRNA was quantified using primers annealing to the beginning of the tracrRNA and to the end of the Cas12a guide. To quantify total levels of the Cas9 guide (processed and unprocessed), primers annealing to the beginning and end of tracrRNA were used. Cas12a

processing activity was estimated by normalizing the levels of unprocessed hgRNA to total levels of Cas9 gRNA.

**Surveyor assays.** On-target genomic editing efficiency was estimated using the Surveyor assay, as previously described<sup>33</sup>. Briefly, N2A cells were transduced with multiple independent Cas9- and gRNA-expressing viruses targeting *Ptbp1* intronic regions. Cells were selected in puromycin (2.5 µg ml<sup>-1</sup>) for 48 h and, 4 d post-selection, genomic DNA was extracted using the PureLink Genomic DNA Kit (Thermo Fisher Scientific). After amplification of targeted loci by PCR (Supplementary Table 14), PCR products were denatured and reannealed to form heteroduplexes. Reannealed PCR products were incubated with T7 endonuclease (NEB) for 20 min at 37°C, and cleavage efficiency was determined by agarose gel electrophoresis.

**Lentiviral hgRNA library construction.** For construction of CHyMERa libraries, Cas9 and Cas12a gRNA sequences were cloned into a lentiviral vector via two rounds of Golden Gate assembly. Oligo pools (113 nt) were designed carrying 20-nt Cas9 and 23-nt Cas12a guide sequences separated by a 32-nt stuffer sequence flanked by BsmBI restriction sites, all flanked by short sequences containing BfuAI restriction sites (see Supplementary Fig. 2b). The oligo pools were synthesized on 90k microarray chips by CustomArray Inc. (a member of GenScript), each with a density of ~94,000 sequences. Oligos were amplified by PCR over ten cycles using Q5 polymerase ((1) 98°C, 30 s; (2) 98°C, 10 s; (3) 53°C, 30 s; (4) 72°C, 10 s; (5) 72°C, 2 min; steps 2–4 repeated for nine cycles). Amplified oligos were purified on a PCR purification column, and an aliquot was run on 2% agarose gel to check purity. The pLCHKO hgRNA vector backbone was digested with BfuAI (NEB) overnight at 37°C and BspMI (NEB) for 2 h. The digested backbone was dephosphorylated with rSAP (NEB) for 1 h at 37°C and gel purified using the GeneJet gel extraction kit (Thermo Fisher Scientific). Amplified oligos were digested with BveI (Thermo Fisher Scientific, FastDigest) and ligated into the digested pLCHKO backbone using T4 ligase (NEB) in a combined reaction overnight over 12 cycles ((1) 37°C, 30 min; (2) 16°C, 30 min; (3) 24°C, 60 min; (4) 37°C, 15 min; (5) 65°C, 10 min; steps 1–3 were repeated for 11 cycles) using an empirically determined vector/insert ratio. The ligation mix was precipitated using sodium acetate and ethanol. The purified ligation reaction was used to transform Endura competent cells (Lucigen) by electroporation (1-mm cuvette, 25 µF, 200 V, 1,600 V), and a sufficient number of cells were plated on 15-cm ampicillin Luria–Bertani (LB) agar plates to reach a library coverage of 500–1,000-fold. Bacterial colonies were scraped from the plates, pooled and bacterial pellets were collected. The Ligation 1 library plasmid was extracted using a Mega-prep plasmid purification kit (Qiagen).

In the second step, Cas9 tracrRNA and the Cas12a direct repeat were inserted into the pooled library. The Ligation 1 plasmid library was digested overnight using Esp3I (Thermo Fisher Scientific, FastDigest) and BsmBI (2 h, 55°C), dephosphorylated using rSAP (1 h, 37°C) and purified on a PCR purification column. A TOPO vector carrying Cas9 tracrRNA and the Cas12a direct repeat was digested using Esp3I and subsequently ligated into the digested pLCHKO-Ligation 1 vector overnight over 12 cycles ((1) 37°C, 30 min; (2) 16°C, 30 min; (3) 24°C, 60 min; (4) 37°C, 15 min; (5) 65°C, 10 min; steps 1–3 were repeated for 11 cycles) using a vector/insert ratio of 1:25. The ligation mix was precipitated using sodium acetate and ethanol. The purified ligation reaction was used to transform Endura competent cells (Lucigen) by electroporation (1-mm cuvette, 25 µF, 200 V, 1,600 V), and a sufficient number of cells were plated on 15-cm ampicillin LB agar plates to reach a library coverage of 500–1,000-fold. Bacterial colonies were scraped from the plates, pooled and bacterial pellets were collected. The Ligation 2 library plasmid was extracted using a Mega-prep plasmid purification kit (Qiagen).

**Library virus production and MOI determination.** For library virus production, 8 million HEK293T cells were seeded per 15-cm plate in high-glucose, pyruvate DMEM medium + 10% FBS. Twenty-four hours after seeding, cells were transfected with a mix of 6 µg of lentiviral pLCHKO vector containing the hgRNA library, 6.5 µg of packaging vector psPAX2, 4 µg of envelope vector pMD2.G, 48 µl of X-treme Gene transfection reagent (Roche) and 1.4 ml of Opti-MEM medium (Life Technologies). Twenty-four hours after transfection, the medium was replaced with serum-free, high-BSA growth media (DMEM, 1.1 g 100 ml<sup>-1</sup> BSA, 1% penicillin/streptomycin). Virus-containing medium was harvested 48 h after transfection, centrifuged at 1,500 r.p.m. for 5 min, aliquoted and frozen at –80°C.

For determination of viral titers, cells were transduced by titration of the lentiviral hgRNA library along with polybrene (8 µg ml<sup>-1</sup>). After 24 h, virus-containing medium was replaced with fresh medium containing puromycin (1–2 µg ml<sup>-1</sup>) and cells were incubated for an additional 48 h. MOI of the titrated virus was determined 72 h post-infection by comparing percentage survival of puromycin-selected cells to infected but nonselected control cells. Due to pre-existing puromycin resistance, RPE1 cells were lifted and reseeded in media containing puromycin (20 µg ml<sup>-1</sup>) to achieve efficient selection of cells transduced with the lentiviral hgRNA library.

**Pooled hgRNA dropout screens.** For pooled screens, 3 × 10<sup>6</sup> cells were seeded in 15-cm plates. A total of 9 × 10<sup>7</sup> cells was transduced with lentiviral libraries

at MOI = ~0.3 such that each hgRNA is represented in about 250–300 cells. Twenty-four hours after infection, transduced cells were selected with 1–2 µg ml<sup>-1</sup> of puromycin for 48 h. Cells were then harvested and pooled, and 3 × 10<sup>7</sup> cells were collected for subsequent gDNA extraction and determination of hgRNA at T0. Pooled cells were then seeded into three replicate plates, each containing 2.1 × 10<sup>7</sup> cells (>200-fold library coverage), which were passaged every 3 d and maintained at >200-fold library coverage until T18. gDNA pellets from each replicate were collected on each day of cell passage.

**Pooled positive-selection hgRNA screens for resistance to 6-TG and thymidine block.** For positive-selection screens, 2 × 10<sup>7</sup> (1 × 10<sup>7</sup> cells per 15-cm plate) HAP1 and CGR8 cells transduced with the human or mouse hgRNA optimization libraries (see Supplementary Tables 1 and 2), respectively, were seeded at T6 in triplicate, and 24 h later were treated with 2.5 mM thymidine or 6 µM 6-TG. After 16 h, thymidine-treated cells were washed and released into normal media, and 10 h later were treated with thymidine for a second time. Cells were maintained in media containing thymidine or 6-TG for the remainder of the screen. At T18, 1.5 × 10<sup>7</sup> cells were collected for gDNA extraction, and hgRNA expression cassettes were amplified and subjected to high-throughput sequencing as described below.

**Torin1 CHyMERa chemogenetic screen.** To identify genes mediating sensitivity or resistance to mTOR inhibition, HAP1 cells were transduced with the CHyMERa library and, following selection, the population was continuously treated with Torin1 (Selleckchem, no. S2827) at a concentration resulting in 60% reduction in cell growth from d 3–18 (that is, the assay endpoint).

**Preparation of CRISPR sequencing libraries and Illumina sequencing.** Genomic DNA was extracted using the Wizard Genomic DNA Purification Kit (Promega). The gDNA pellets were resuspended in TE buffer and concentration was estimated by Qubit using dsDNA Broad Range Assay reagents (Invitrogen). Sequencing libraries were prepared from the extracted gDNA (55 µg for HAP1, RPE1 and CGR8; 87.5 µg for N2A cells) in two PCR steps, the first to enrich gRNA regions from the genome and the second to amplify gRNA and attach Illumina TruSeq adapters with indices i5 and i7. Barcoded libraries were gel purified, fragment size distribution was assessed using a BioAnalyzer and final concentrations were estimated by quantitative PCR with reverse transcription. Sequencing libraries were sequenced on an Illumina NextSeq500 or NovaSeq using paired-end sequencing. The first read included 29 dark cycles, followed by 31 cycles for reading the Cas12a guide and an index read of 8 cycles. For the paired read, 20 dark cycles were followed by 30 cycles for reading the Cas9 guide and an index read of 8 cycles.

**Dual-guide mapping and quantification.** FASTQ files from paired-end sequencing were first processed to trim off flanking sequences up- and downstream of the guide sequence using a custom Perl script. Reads that did not contain the expected 3' sequence, allowing up to two mismatches, were discarded. Preprocessed paired reads were then aligned to a FASTA file containing the library sequences using Bowtie (v.0.12.7) with the following parameters: -v 3 -1 18 -chunkmbs 256 -t <library\_name>. The number of mapped read pairs for each dual-guide construct was then counted and merged, along with annotations, into a matrix.

**Human and mouse hgRNA optimization library design.** Human and mouse hgRNA libraries were designed in which exonic regions of the reference core essential genes (CEG2)<sup>33</sup> and nonessential genes were targeted either with Cas9 (paired with an intergenic-targeting Cas12a guide) or Cas12a (paired with an intergenic-targeting Cas9 guide). To target constitutive exons of mouse core essential genes, we first identified all one-to-one orthologs of the CEG2 set. From all possible 23-nt Cas12a guides targeting these constitutive exons and adjacent to a TTTV 5'-end PAM sequence, we randomly selected up to 15 Cas12a guides per target exon. Cas9 gRNAs (20-nt) were selected based on previously defined rules<sup>33</sup>. Collectively, our optimization libraries target over 450 CEG2 essential genes and include up to 5 Cas12a and 3 Cas9 exon-targeting guides per exon, up to 15 Cas12a and 2 Cas9 exon-flanking guides per exon, as well as 1,000 control constructs targeting intergenic regions with spacing between target sites similar to that in the exon-targeting guide pairs (Supplementary Tables 1 and 2). To control for toxicity induced by hgRNA-directed dsDNA breaks, each gRNA sequence was paired with a gRNA targeting a noncoding intergenic sequence. In addition, the *TK1* and *HPRT1* genes were also targeted in the same manner. Furthermore, exon-deletion constructs targeting *TK1* and *HPRT1* were designed by pairing guides targeting intronic regions up- and downstream of selected exons with target sites located at least 100 nucleotides away from splice sites. The full contents of human and mouse optimization libraries can be found in Supplementary Tables 1 and 2, respectively.

**Second-generation human dual-cutting and paralog hgRNA library design.** A second-generation hgRNA library was designed in which the ~5,000 highest-expressed genes across a panel of human cell lines (HAP1, RPE1, HEK293T, HCT116, HeLa and A375) were targeted either with Cas9 (paired with an intergenic-targeting Lb-Cas12a guide), Lb-Cas12a (paired with an intergenic-targeting Cas9 guide) or simultaneously with both Cas9 and Lb-Cas12a guides

(dual-targeting). Target sites for the dual-targeting constructs were spaced between 107 base pairs (bp) and >946 kb (median distance, 6,863 bp). In addition, hgRNAs targeting intergenic and nontargeting sites were included as controls. This portion of the library included 61,888 hgRNA constructs.

In the second part of the library, we targeted paralog gene pairs<sup>74</sup> from gene families where both genes are expressed across a panel of human cell lines (HAP1, RPE1, HEK293T, HCT116, HeLa and A375). Of 1,381 strict human ortholog families that have arisen from whole-genome duplications of vertebrate genomes<sup>74</sup>, we selected 1,344 paralogs (avoiding gene families with more than two paralogs). In addition we also targeted selected gene pairs of interest, some of which have previously been reported to interact genetically. All gene pairs were either targeted individually by Cas9 (paired with an intergenic-targeting Lb-Cas12a guide) or Lb-Cas12a (paired with an intergenic-targeting Cas9 guide), or with both Cas9 and Lb-Cas12a guides paired in both possible orientations (dual-targeting). This portion of the library comprised 30,848 hgRNA constructs. The full contents of the human single-gene dual-targeting and paralog-targeting library can be found in Supplementary Table 5.

**Exon-deletion hgRNA library design.** For the first-generation exon-deletion hybrid guide library, murine exons in genes with a minimum expression level in N2A cells  $\geq 5$  corrected reads per kilobase of transcript per million mapped reads (cRPKM) and that are alternatively spliced in neural cells were selected according to any of the following criteria: (1) inclusion level >10 percentage spliced in (PSI) in N2A and dynamically regulated during neuronal differentiation<sup>75</sup>; (2) more highly included in neural compared to non-neural cells and tissues by an average of 10 PSI, and also more highly included in N2A versus non-neural cells by an average of 10 PSI<sup>76</sup>; or (3) microexons up to 27 nt in length with >10 PSI in N2A and differentially spliced between neural and non-neural cells by an average of 10 PSI.

In regard to the exon-targeting library for use in human cells, alternative exons were selected as follows. Alternative splicing and host gene expression in HAP1 cells was first quantified from RNA-seq data (unpublished) using vast-tools 1.2.0<sup>77</sup>. Exons were selected with a PSI range >30 across 108 diverse tissues and cell types in VASTDB (<http://vastdb.crg.eu>), and that were at least moderately included (PSI  $\geq 15$ ) in either HAP1, HeLa, 293T or MCF7 cells with expression level cRPKM > 5 in the same cell line. In total, 4,290 candidate exons from stream 1 and 466 from stream 2 were combined, and events were prioritized according to their requirement for cellular fitness in HAP1 cells<sup>6,33</sup> and whether they preserved the open reading frame. After guide selection, this resulted in 324 frame-preserving events in essential genes, 2,942 frame-preserving exons not in essential genes, 118 frame-disrupting events in essential genes and 40 events that were neither frame-preserving nor within essential genes. A group of control exons was also selected that were skipped in HAP1 cells (PSI < 3) but were included in at least one other cell type or tissue at PSI > 20, and whose host genes were clearly expressed in HAP1 cells (cRPKM > 5), irrespective of gene essentiality. For all exons, hgRNAs targeting intronic sites flanking the exon of interest were designed to introduce dsDNA breaks at intronic sites at least 100 bp distal from splice sites flanking the target exons. Each exon was targeted by multiple Cas9-Cas12a hgRNAs. Where possible (that is, depending on the availability of target sites), two individual Cas9 guides were paired with up to four Cas12a guides targeting both up- and downstream flanking intronic sequences, resulting in a total of 16 pairs of deletion-targeting hgRNA constructs for each exon. Furthermore, each intronic Cas9 and Cas12a gRNA was also paired with two intergenic gRNAs to control for nonspecific toxicity, adding 24 control hgRNA pairs per exon. Furthermore, each gene targeted by exon deletion hgRNAs was also targeted by exon-targeting Cas9 guides. Full contents of the human exon-targeting library can be found in Supplementary Table 2.

**RNA-seq.** RNA was extracted from HAP1 cells transfected with nontargeting siRNA, siRBM26 and/or siRBM27, as described above, using the RNeasy extraction kit (Qiagen) following the manufacturer's recommendations. Two independent biological samples for each condition were generated, resulting in a total of eight samples. DNase-treated RNA samples were submitted for RNA-seq at the Donnelly Sequencing Center at the University of Toronto. Total RNA was quantified using Qubit RNA BR (catalog. no. Q10211, Thermo Fisher Scientific) fluorescent chemistry, and 1 ng was used to obtain RNA integrity number (RIN) using the Bioanalyzer RNA 6000 Pico kit (catalog. no. 5067-1513, Agilent). The lowest RIN was 8.7, and median was 9.6.

Total RNA (2.5  $\mu$ g) per sample was processed using the MGIEasy Directional RNA Library Prep Set v.2.0 (protocol v. A0, catalog. no. 1000006385, Shenzhen) including mRNA enrichment with the Dynabeads mRNA Purification Kit (catalog. no. 61006, Thermo Fisher Scientific). RNA was fragmented at 87 °C for 6 min following the addition of 75% of the recommended volume of fragmentation buffer, to produce longer fragments. Libraries were amplified with 12 cycles of PCR.

The top stock (1  $\mu$ l) of each purified final library was run on an Agilent Bioanalyzer dsDNA High Sensitivity chip (catalog. no. 5067-4626, Agilent) to determine an average library size of 581 bp, and to confirm the absence of dimers. Libraries were quantified using the Quant-iT dsDNA High Sensitivity fluorometry kit (catalog. no. Q33120, Thermo Fisher), pooled equimolarly and libraries in each of four replicate pools were then circularized using the MGIEasy Circularization Module (catalog. no. 1000005260, Shenzhen).

From each of the four pools, 40 fmol of circularized library was sequenced  $2 \times 150$  bp on a single lane of an FCL flowcell on the MGISEQ-2000 platform (also known as the DNBSEQ-G400 platform, Shenzhen), for a total of four lanes of sequencing.

**Analysis of CHyMERa screens.** Depletion of dual-guide constructs was assessed with the Bioconductor package edgeR (v.3.18.1). Following depth normalization, only constructs with more than one count per million (CPM) in at least two samples were retained. Exon-targeting constructs resulting in significant depletion over time (active guides) were identified from T18 triplicate samples using the likelihood-ratio test, with LFC < 0 and FDR < 0.05. We found 1,073 guide constructs that were significantly active at this threshold in the HAP1 screen. In addition, 1,026 inactive (neutral) guides were identified where LFC = -0.5 to 0.5. These active and inactive sets were used to train the machine learning classifiers.

Of note, 4–6% of reads from plasmid pool samples map to recombinated guide constructs. We noticed that the level of recombination strongly increased following lentiviral transduction of cell lines (to >19%). This suggests that the predominant source of recombination occurs as a result of template switching by viral reverse transcriptase during production of the lentiviral library or viral transduction, and not as the result of template switching during PCR amplification as recently suggested<sup>25,78,79</sup>.

**Analysis of nucleotide composition of active Cas12a guides.** The physical properties of Cas12a guides targeting exons of the 'gold-standard essential' genes were examined to optimize guide design. LFC at the screen endpoint was used as the measure of 'activity'. Single-, di- and trinucleotide composition, GC content, PAM sequence and up- and downstream sequences were examined for the full set of exon-targeting guides, and also specifically for significantly depleted guides. Significantly depleted guides were defined as those with LFC < 0 and FDR < 0.05 (HAP1,  $n = 1,073$ ; CGR8,  $n = 1,749$ ; N2A,  $n = 1,063$ ).

**Training classifiers to predict Cas12a guide activity.** To better understand differences between Cas12a active and inactive guide sequences, and to help identify highly effective guides, we applied machine learning approaches to data from the pilot screen. We trained models using three different approaches: L1-regularized logistic regression (L1Logit), RF and CNN. For our training set, we combined Cas12a guide sequences from Cas9 intergenic-Cas12a exonic hgRNAs from optimization screens performed in human and mouse cell lines (2,096 HAP1 sequences, 2,401 CGR8 and 600 N2A), totaling 5,097 unique sequences. Each 23-bp guide sequence was extended by adding the upstream PAM sequence (4 bp) and flanking up- and downstream sequences (6 bp each), resulting in a total sequence length of 39 bp. Next, we assigned discrete labels to each guide according to its guide activity from the initial screen: active (FDR < 0.05, LFC < -1) and inactive (FDR > 0.05, LFC = -0.5 to 0.5). To construct the features for model training, we transformed each sequence into a set of numerical features using one-hot encoding, resulting in a  $4 \times 39$  binary matrix  $E$  such that element  $e_{ij}$  represents an indicator variable for nucleotide  $i$  (A, T, C and G) at position  $j$ . This representation serves as the main input to the CNN. To be amenable for L1Logit and RF, we converted this binary matrix into 156 individual nucleotide- and position-specific binary features. Additionally, we included binary features representing the 2-mer occurrences at every position (16 features per position), adding another 608 binary features for a total of 764 sequence-based features.

In addition to one-hot encoding of guide sequences, additional hand-crafted features were created: the predicted minimum free energy resulting from the secondary structure of the guide sequence, and melting temperatures for various segments of the guide sequence. For secondary structure prediction, we used RNAfold<sup>80</sup> to calculate minimum free energy values for each 23-bp guide sequence. We used the MeltingTemp.Tm\_NN() function in Biopython<sup>81</sup> to calculate the melting temperature for the guide sequence, seed (positions 1–6), trunk (7–18) and promiscuous region (19–23). In total, we generated an additional five hand-crafted features. Together these features were used to augment the sequence-based features.

**Predicting with chromatin accessibility information.** To investigate the influence of chromatin accessibility on measured Cas12a guide activity, we used DNase hypersensitive sites from K562 cells<sup>82</sup>. We first identified the chromatin status of each guide target site in our dataset, and found that 92% of the guides were targeting 'inaccessible' sites. The highly unbalanced nature of this feature suggested to us that chromatin accessibility would not be an informative feature in our model, and thus it was excluded from further consideration.

**CNN architecture for prediction of efficient Cas12a guides.** To identify features associated with the most active Cas12a guides, we applied machine learning algorithms to predict efficient Cas12a guides as follows: Cas12a guides targeting exons of core fitness genes were first binned into active or inactive categories based on their observed relative depletion levels, as determined by LFC scores in HAP1 and CGR8 cells (Supplementary Fig. 2d). For each guide, we assembled features based on single, di- and trinucleotide composition, PAM sequence, up- and downstream sequences as well as genomic accessibility at the target site. The CNN consist of three main components: convolutional-pool layers, fully connected

layers, and an output layer. First, we begin by passing matrix  $E$  into a convolutional layer consisting of 52 filters of length 4. Each filter is a  $4 \times 4$  matrix representing a motif to be learned from the data: in other words, a filter is a position weight matrix. During training, each filter scans along the input sequence and computes a score for each 4-mer, followed by a rectified linear unit activation. These activated scores are then passed through a pooling layer, where the average score is computed over a sliding window of 3. Next, to prevent overfitting of the model, the scores proceed through a dropout layer with a dropout rate of 0.22. At this stage, the convolution step has produced a set of summarized feature scores representing the input sequence. Before proceeding to the next fully connected layer, we extend the features set by concatenating the hand-crafted features described above. This new feature set is then passed to a single, fully connected hidden layer with 12 units, followed by another dropout layer. Finally, the scores proceed through an output layer consisting of a sigmoid function. Training was carried out using the Adam optimizer with a learning rate of 0.0001 and minimization of binary cross-entropy loss function. By the end of training, the filters in the convolutional layer will have learned a set of motifs that are predictive of guide activity. All hyperparameters were chosen through cross-validation as described below, with the exception of pooling size for pooling layers, which was fixed.

**Deep learning model selection.** To implement conventional algorithms, the scikit-learn framework<sup>43</sup> was used. To implement CNN, Keras (<https://keras.io>) with TensorFlow<sup>44</sup> backend was used. We randomly selected approximately 90% of the data for training, while the remaining 10% was withheld for testing. The sampling was stratified such that the relative proportions of each cell line were maintained:

Sample	Train	Test
HAP1	1,886	210
CGR8	2,160	241
N2A	540	60

To determine the optimal hyperparameters, we performed fivefold cross-validation on the training data. For the conventional methods, we performed a grid search for the following parameters:

- L1Logit: alpha
- RF: number of trees

For CNN, we performed a random sampling search<sup>45</sup> for the number of filters, filter size and batch size.

**Evaluation of deep learning models.** Performance of the classifiers was evaluated by prediction on heldout test data. For each algorithm, we compared models with and without the additional secondary structure and melting temperature features. Performance was measured based on AUC and average precision using the scikit-learn functions `auc()` and `average_precision_score()`.

To compare CHyMERa-Net scores with DeepCpf1<sup>36</sup>, we calculated the scores of Cas12a guides in our libraries using DeepCpf1 and compared LFC trends by binning CHyMERa-Net and DeepCpf1 scores into ten bins of approximately equal size. Although our CNN predictions and DeepCpf1 were trained using different readouts (proliferation versus indel frequencies), nucleases (Lb- versus As-Cas12a) and with different amounts of data (5,097 training sequences versus 15,000 sequences for DeepCpf1), we observed strong negative slopes for scores from both classifiers.

**Scoring of GIs in the optimized library.** Data were scored for GIs by comparison of observed LFC values for dual-targeting constructs to a null model derived from exonic–intergenic guides. We assumed an additive model of GIs for genes A and B (Equation 1), where GIs are called when the observed LFC values for a double-knockout (Equation 2) significantly differed from the sum of single-knockout LFCs (Equation 3). We compared the set of double-knockout LFCs for each gene pair to the set of all sums of single-knockout LFCs using the Wilcoxon rank-sum test followed by Benjamini–Hochberg FDR correction. We performed significance testing only on expected and observed sets with matching orientations, where Cas9 targets gene A and Cas12a targets gene B or vice versa, resulting in two  $P$  values per gene pair. Most Cas9 guides had three replicates and most Cas12a guides had five, but the number of replicates varied slightly across gene pairs (Supplementary Table 5). To avoid false positives, we called only significant GIs for a gene pair if both tested orientations were significant at  $FDR < 0.1$  with the same LFC sign. If, for example, both orientations for a specific gene pair were significant but with opposing signs on the LFC, that gene pair was not called as a significant GI. All scored data are shown in Supplementary Table 8.

$$LFC_{AB} = LFC_A + LFC_B + GI_{AB} \quad (1)$$

Equation (1) is an additive model of GIs for genes A and B.

$$\begin{aligned} \text{Observed}_1 &= \{A_{\text{Cas9}}, B_{\text{Cas12a}} \mid i \in 1 \dots 3 \wedge j \in 1 \dots 5\} \\ \text{Observed}_2 &= \{B_{\text{Cas9}}, A_{\text{Cas12a}} \mid i \in 1 \dots 3 \wedge j \in 1 \dots 5\} \end{aligned} \quad (2)$$

Equation (2) gives the gene-pair-specific set of observed LFCs for testing GIs. The set of all exonic–exonic LFCs where the Cas9 gRNA targets gene A and the Cas12a gRNA targets gene B for orientation 1, and vice versa for orientation 2.

$$\begin{aligned} \text{Expected}_1 &= \{A_{\text{Cas9}} + B_{\text{Cas12a}} \mid i \in 1 \dots 3 \wedge j \in 1 \dots 5\} \\ \text{Expected}_2 &= \{B_{\text{Cas9}} + A_{\text{Cas12a}} \mid i \in 1 \dots 3 \wedge j \in 1 \dots 5\} \end{aligned} \quad (3)$$

Equation (3) gives the gene-pair-specific set of expected LFCs for testing GIs. The set of all sums of exonic–intergenic LFC values where the Cas9 gRNA targets gene A and the Cas12a gRNA targets gene B for orientation 1, and vice versa for orientation 2.

**MAGECK scoring of dual-targeting library.** Because the dual-targeting library lacked the gold-standard negative genes required by the BAGEL algorithm, we employed model-based analysis of genome-wide CRISPR–Cas9 knockout (MAGECK) to score these data. Input matrices were prepared using a bespoke R script. A matrix of read counts was prepared separately for each single- and dual-targeting subset, along with a design matrix. Single-targeting constructs were identified as having one exon-targeting guide (either Cas9 or Cas12a) paired with an intergenic-targeting guide, while dual-targeting constructs comprise two exon-targeting guides. Each extracted matrix was filtered to remove guide constructs that had zero reads in all samples. MAGECK was run using the following command line: `mageck mle --count-table <count_file> --design-matrix <design-matrix> --norm-method median --output-prefix <sampleName>.mle`. Significantly depleted genes were called where beta score  $< 0$  and  $FDR < 0.05$ .

**Analysis of DepMap data.** Data from the DepMap screening platform (DepMap Public 19Q1) were downloaded from <https://depmap.org/portal/download/>. The matrix consisted of CERES-adjusted, gene-level fitness scores for 558 screened cell lines. Gene annotations were parsed to gene symbols in R, and analyzed with no further adjustments. CERES scores for the four gene sets (CEG2, gold-standard negatives, dual-targeting only and single-targeting–dual-targeting overlap) were aggregated and plotted together.

**Scoring of differential response to mTOR inhibition.** Data were scored for differential response to mTOR inhibition by comparison of LFC values for the HAP1 screen with or without Torin1 drug treatment, across four different types of guide and two time points. The types of guide analyzed included (1) single-targeting guides targeting a single gene, (2) dual-targeting guides targeting a single gene, (3) single-targeting guides targeting a single paralogous gene and (4) dual-targeting guides targeting paralogous gene pairs in a combinatorial manner. All LFC values with or without Torin1 treatment were compared separately at T12 and T18 using Wilcoxon rank-sum tests between treated and untreated LFCs for each gene, followed by Benjamini–Hochberg FDR correction.

Data were processed as follows. (1) Each gene was targeted by three Cas12a and two Cas9 guides, with three replicates per guide. To measure Torin1 response for each gene, these guide LFCs were aggregated, including replicates, to test sets of 15 LFCs – Torin1 against corresponding sets of 15 LFCs + Torin1. (2) Each gene was dual-targeted by six guides with three replicates per guide. To ensure that the statistical power of this analysis was equivalent to the statistical power for (1), one of the six dual-targeting guides was randomly dropped for each contrast before comparing sets of 15 guides with replicates with or without Torin1 as in (1). (3) Each gene was targeted by five Cas12a guides and three Cas9 guides, with three replicates per guide. These guide LFCs were aggregated, including replicates, to test sets of 24 LFCs – Torin1 against corresponding sets of 24 LFCs + Torin1. (4) Each paralog pair was targeted by 15 guides in each orientation, with three replicates per guide. To ensure that the statistical power of this analysis was equivalent to the statistical power of (3), the mean of each replicate was taken and 6 of the remaining 30 guides across both orientations were randomly dropped before testing for differential Torin1 response.

Gene ontology analysis was performed using GOrilla<sup>46</sup> (<http://cbl-gorilla.cs.technion.ac.il>). Hits that were called at  $FDR < 0.1$  at the early and late time points were included in the target list, and all targeted genes were used as background. For data visualization we display terms with  $< 900$  members and enriched at  $FDR < 0.05$ .

**RNA-seq analysis of RBM26 and/or RBM27 knockdown experiments.** To quantify gene expression, pretrimmed reads were pseudoaligned to the GENCODE human gene annotation v.29. Transcript-level quantifications were aggregated per gene using the R package tximport, and differential expression between control non-targeting and RBM26 and/or RBM27 knockdown was assessed using the classic mode (exactTest) in edgeR. Genes changing more than two-fold and with  $FDR < 0.05$  were deemed significantly different. To compare overlaps in changes between treatments, only genes expressed at  $RPKM > 5$  in at least one treatment were considered.

Gene Ontology analysis of genes with  $LFC > 1$ ,  $FDR < 0.05$  and  $RPKM > 5$  was performed with FuncAssociate<sup>47</sup> (<http://llama.mshri.on.ca/funcassociate/>) using all detected genes ( $RPKM > 5$ ) as background. For plotting, overlapping categories

were removed when >70% of changing genes overlapped with another category with a more significant enrichment.

**Analysis of exon deletion screens.** Dropout rates were scored for significant exonic deletion events by comparison to a null distribution derived from intergenic–intergenic guides. We compared the LFC of each intronic–intronic hgRNA construct to the LFC distribution of all intergenic–intergenic guide pairs, and called intronic–intronic pairs as significant at a threshold of  $P < 0.05$  using a one-tailed test based on empirical null distribution.

A targeted exon was deemed a hit if >18% of intronic–intronic pairs targeting the exon produced a significant negative LFC, including at least one pair for which neither the Cas9 nor the Cas12a guide, in combination with an intergenic guide, resulted in significant dropout, measured similarly to that described for intronic–intronic pairs above. This threshold was chosen to maximize the difference in hit rates for frame-disrupting exons in expressed genes whose deletion is known to cause a growth defect, compared to exons that are skipped or within nonexpressed genes in the given cell line. Growth-related fitness in RPE1 cells was derived from a previous study<sup>6</sup>, and both gene expression and exon inclusion were scored from RNA-seq data<sup>6</sup> using vast-tools.

**Reporting Summary.** Further information on research design is available in the Life Sciences Reporting Summary.

### Data availability

The hgRNA sequences and contents of all libraries are included in supplementary tables. The datasets generated and analyzed in this study are included in the manuscript and deposited at the website <http://crispr.cabr.utoronto.ca/chymera>. Raw fastq files for all sequencing data are available upon request and at the Gene Expression Omnibus: RBM26 and RBM27 RNA-seq data, [GSE144078](https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE144078); CRISPR screens sequencing read archive, [GSE144281](https://www.ncbi.nlm.nih.gov/sra/acc.cgi?acc=GSE144281).

### Code availability

Descriptions of analyses, tools and algorithms are provided in Methods. Custom code for generation of hgRNA counts from fastq files, code for CHyMERa-Net and code for analysis of screens with the dual-targeting/paralog and exon-targeting libraries are available at Github (<https://github.com/BlencoweLab/chymeranet>) and at <http://crispr.cabr.utoronto.ca/chymera>.

### References

- Dang, Y. et al. Optimizing sgRNA structure to improve CRISPR-Cas9 knockout efficiency. *Genome Biol.* **16**, 280 (2015).
- Guschin, D. Y. et al. A rapid and general assay for monitoring endogenous gene modification. *Methods Mol. Biol.* **649**, 247–256 (2010).
- Singh, P. P., Arora, J. & Isambert, H. Identification of ohnolog genes originating from whole genome duplication in early vertebrates, based on synteny comparison across multiple genomes. *PLoS Comput. Biol.* **11**, e1004394 (2015).
- Hubbard, K. S., Gut, I. M., Lyman, M. E. & McNutt, P. M. Longitudinal RNA sequencing of the deep transcriptome during neurogenesis of cortical glutamatergic neurons from murine ESCs. *F1000Res.* **2**, 35 (2013).
- Raj, B. et al. A global regulatory mechanism for activating an exon network required for neurogenesis. *Mol. Cell* **56**, 90–103 (2014).
- Tapial, J. et al. An atlas of alternative splicing profiles and functional associations reveals new regulatory programs and genes that simultaneously express multiple major isoforms. *Genome Res.* **27**, 1759–1768 (2017).
- Hanna, R. E. & Doench, J. G. A case of mistaken identity. *Nat. Biotechnol.* **36**, 802–804 (2018).

- Feldman, D., Singh, A., Garrity, A. J. & Blainey, P. C. Lentiviral co-packaging mitigates the effects of intermolecular recombination and multiple integrations in pooled genetic screens. Preprint at *bioRxiv* <https://doi.org/10.1101/262121> (2018).
- Lorenz, R. et al. ViennaRNA Package 2.0. *Algorithms Mol. Biol.* **6**, 26 (2011).
- Cock, P. J. A. et al. Biopython: freely available Python tools for computational molecular biology and bioinformatics. *Bioinformatics* **25**, 1422–1423 (2009).
- ENCODE Project Consortium. An integrated encyclopedia of DNA elements in the human genome. *Nature* **489**, 57–74 (2012).
- Pedregosa, F. et al. Scikit-learn: machine learning in python. *J. Mach. Learn. Res.* **12**, 2825–2830 (2011).
- Abadi, M. et al. TensorFlow: large-scale machine learning on heterogeneous distributed systems. Preprint at *arXiv* <https://arxiv.org/abs/1603.04467> (2016).
- Bergstra, J. & Bengio, Y. Random search for hyper-parameter optimization. *J. Mach. Learn. Res.* **13**, 281–305 (2012).
- Eden, E. et al. GOrilla: a tool for discovery and visualization of enriched GO terms in ranked gene lists. *BMC Bioinformatics* **10**, 48 (2009).
- Berriz, G. F., King, O. D., Bryant, B., Sander, C. & Roth, F. P. Characterizing gene sets with FuncAssociate. *Bioinformatics* **19**, 2502–2504 (2003).

### Acknowledgements

We thank Q. Morris for advice on machine learning implementation, F. Zhang and B. Zetsche (Broad Institute, MIT) for providing Cas12a reagents and insights, and members of the Moffat, Blencowe and Myers laboratories for helpful discussions. G. O'Hanlon, Q. Huang, C. Sheene and S. Sidhu are gratefully acknowledged for assistance with sequencing and molecular biology experiments. T.G.-P. was supported by postdoctoral fellowships from the European Molecular Biology Organization, Ontario Institute of Regenerative Medicine (OIRM) and Canadian Institutes for Health Research. M.A. was supported by a Swiss National Science Foundation fellowship. H.N.W. was supported by a National Institutes of Health (NIH) Biotechnology Training Grant. This research was funded by grants from Canadian Institutes for Health Research (to B.J.B. and grant no. MOP-142375 to J.M.), Medicine-by-Design Canada First Research Excellence Fund (grant no. CITPA-2016-11 to B.J.B. and J.M.), OIRM (to B.J.B.), Genome Canada (grant no. OGI-157 to J.M.), NIH (to C.L.M.) and the National Science Foundation (to C.L.M.). J.M. is a Canadian Research Chair in Functional Genomics. B.J.B. holds the Banbury Chair of Medical Research at the University of Toronto.

### Author contributions

T.G.P., M.A., J.M. and B.J.B. conceptualized the study and study design. Initial conceptualization, development and demonstration of efficacy of the hybrid Cas9–Cas12a system were performed by T.G.-P. and S.F. Experimental investigation was carried out by T.G.-P., M.A., S.F. and A.W. Guide library design and data analysis were performed by K.R.B., U.B., H.N.W., M.A., T.G.-P., T.D., M.B. and J.M. Machine learning implementation was carried out by K.C.H. The original draft was written by T.G.-P., M.A., J.M. and B.J.B. with input from K.R.B. Writing, reviewing and editing were carried out by T.G.-P., M.A., K.R.B., J.M. and B.J.B. with input from the other authors. Supervision was done by B.J.B., J.M., C.L.M. and T.G.-P. Funding was acquired by B.J.B., J.M., C.L.M. and T.G.-P.

### Competing interests

A patent application (no. GB 1907733.8) describing the development and applications of CHyMERa, to the University of Toronto and T.G.-P., M.A., K.R.B., S.F., J.M. and B.J.B., is pending.

### Additional information

**Supplementary information** is available for this paper at <https://doi.org/10.1038/s41587-020-0437-z>.

**Correspondence and requests for materials** should be addressed to T.G.-P., B.J.B. or J.M.

**Reprints and permissions information** is available at [www.nature.com/reprints](http://www.nature.com/reprints).

## 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](#).

### Statistics

For all statistical analyses, confirm that the following items are present in the 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
- A statement on 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 statistical parameters 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

*Our web collection on [statistics for biologists](#) contains articles on many of the points above.*

### Software and code

Policy information about [availability of computer code](#)

Data collection

Illumina Nextseq500, Novaseq 6000 (<https://www.illumina.com/>)  
WaveFX confocal microscope, Quorum Technologies (<http://quorumtechnologies.com/>)

Data analysis

Bowtie 0.12.7 & 1.1.1 ([https://en.wikipedia.org/wiki/Bowtie\\_\(sequence\\_analysis\)](https://en.wikipedia.org/wiki/Bowtie_(sequence_analysis))),  
R 3.5 & 3.6 (various functions; <https://www.r-project.org/>)  
Bioconductor EdgeR 3.24.3 & 3.26.8 (<https://bioconductor.org/packages/release/bioc/html/edgeR.html>)  
Bioconductor tximport 1.12.3 (<https://git.bioconductor.org/packages/tximport>)  
Biopython 1.69 (<https://biopython.org/>)  
RNAfold 2.4.11 (<http://rna.tbi.univie.ac.at/cgi-bin/RNAWebSuite/RNAfold.cgi>)  
scikit-learn 0.21.1 (<https://scikit-learn.org/stable/index.html>)  
Keras 2.2.4 (<https://keras.io>)  
TensorFlow 1.13.1 (<https://www.tensorflow.org/>)  
GraphPad Prism 8.1.1 (<https://www.graphpad.com/scientific-software/prism/>)  
Vast-tools 2.2.2 (<https://github.com/vastgroup/vast-tools>)  
Perl 5.12.4 & 5.16.3 (<https://www.perl.org/>)  
Salmon 0.14.1 (<https://combine-lab.github.io/salmon/>)  
GORilla gene ontology tool (<http://cbl-gorilla.cs.technion.ac.il/>)  
ImageJ 1.46 (<https://imagej.nih.gov/ij/>)

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. We strongly encourage code deposition in a community repository (e.g. GitHub). See the Nature Research [guidelines for submitting code & software](#) for further information.

## Data

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

All raw data is available upon request. There is no restriction on data availability.

## Field-specific reporting

Please select the one below that is 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/documents/nr-reporting-summary-flat.pdf](https://www.nature.com/documents/nr-reporting-summary-flat.pdf)

## Life sciences study design

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

Sample size	No sample-size calculations were performed and sample sizes were arbitrarily chosen according to convention in the field. For small-scale experiments, the number of replicates exceeds at least 3 biological replicates (= independent experiments) and/or at least 3 technical replicates (= repeated measurements of the same original sample). For screens, the initial mutagenized cell pool was split into 3 replicates post-selection and processed independently in all downstream steps.
Data exclusions	No data were excluded from any experiments and figures shown.
Replication	We present no experimental results that were not reproducible.
Randomization	Screen samples were processed and sequenced in a randomized manner and labelled with numbers instead of sample names.
Blinding	No sample allocation to groups was performed, so blinding was not relevant to this study.

## Reporting for specific materials, systems and methods

We require information from authors about some types of materials, experimental systems and methods used in many studies. Here, indicate whether each material, system or method listed is relevant to your study. If you are not sure if a list item applies to your research, read the appropriate section before selecting a response.

### Materials & experimental systems

n/a	Involved in the study
<input type="checkbox"/>	<input checked="" type="checkbox"/> Antibodies
<input type="checkbox"/>	<input checked="" type="checkbox"/> Eukaryotic cell lines
<input checked="" type="checkbox"/>	<input type="checkbox"/> Palaeontology
<input checked="" type="checkbox"/>	<input type="checkbox"/> Animals and other organisms
<input checked="" type="checkbox"/>	<input type="checkbox"/> Human research participants
<input checked="" type="checkbox"/>	<input type="checkbox"/> Clinical data

### Methods

n/a	Involved in the study
<input checked="" type="checkbox"/>	<input type="checkbox"/> ChIP-seq
<input checked="" type="checkbox"/>	<input type="checkbox"/> Flow cytometry
<input checked="" type="checkbox"/>	<input type="checkbox"/> MRI-based neuroimaging

## Antibodies

Antibodies used	anti- $\beta$ -Actin (Abcam ab8226) anti-Cas9 (Diagenode C15200229) anti-LbCpf1 (Sigma SAB4200777) anti-P53 (Life Technologies AH00152) anti-pRb S807/811 (Cell Signalling 9308) anti-p21 (Cell Signalling 2946) anti-Myc (Sigma M4439) anti-Mouse (Jackson ImmunoResearch 715-035-151) anti-Rabbit (Cell Signaling Technology 7074)
Validation	All antibodies are commercially available, were validated by the manufacturers and are routinely used in scientific studies and publications. Furthermore, positive and negative controls built into our experimental design further confirmed specificity of the

Cas9, LbCpf1 and Myc antibodies. Finally the TP53, pRb and p21 antibodies confirmed well established expression responses to DNA damage.

## Eukaryotic cell lines

---

Policy information about [cell lines](#)

Cell line source(s)	HAP1 cells were obtained from Horizon (clone C631, sex: male with lost Y chromosome, RRID: CVCL_Y019). RPE1 (CRL-4000), N2A (CCL-131) and HEK 293T cells were obtained from ATCC. CGR8 cells were obtained from ECACC (07032901). HAP1 individual gene knock-out cell lines were obtained from Horizon: EED (HZGHC001200c001), LDHA (HZGHC004917c008), SLC16A1 (HZGHC002882c002), ROCK2 (HZGHC000061c017), SP1 (HZGHC001141c002), ARID1A (HZGHC000618c010), DNAJA1 (HZGHC004473c007).
Authentication	HAP1 and RPE1 cells were authenticated by STR profiling at the Centre for Applied Genomics (TCAG) at the Hospital for Sick Children (SickKids) in Toronto. HAP1 cells were also whole-genome sequenced.
Mycoplasma contamination	All cell lines were routinely tested and confirmed negative for mycoplasma contamination.
Commonly misidentified lines (See <a href="#">ICLAC</a> register)	None of the cell lines used in this study is listed as commonly misidentified.