

In the format provided by the authors and unedited.

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

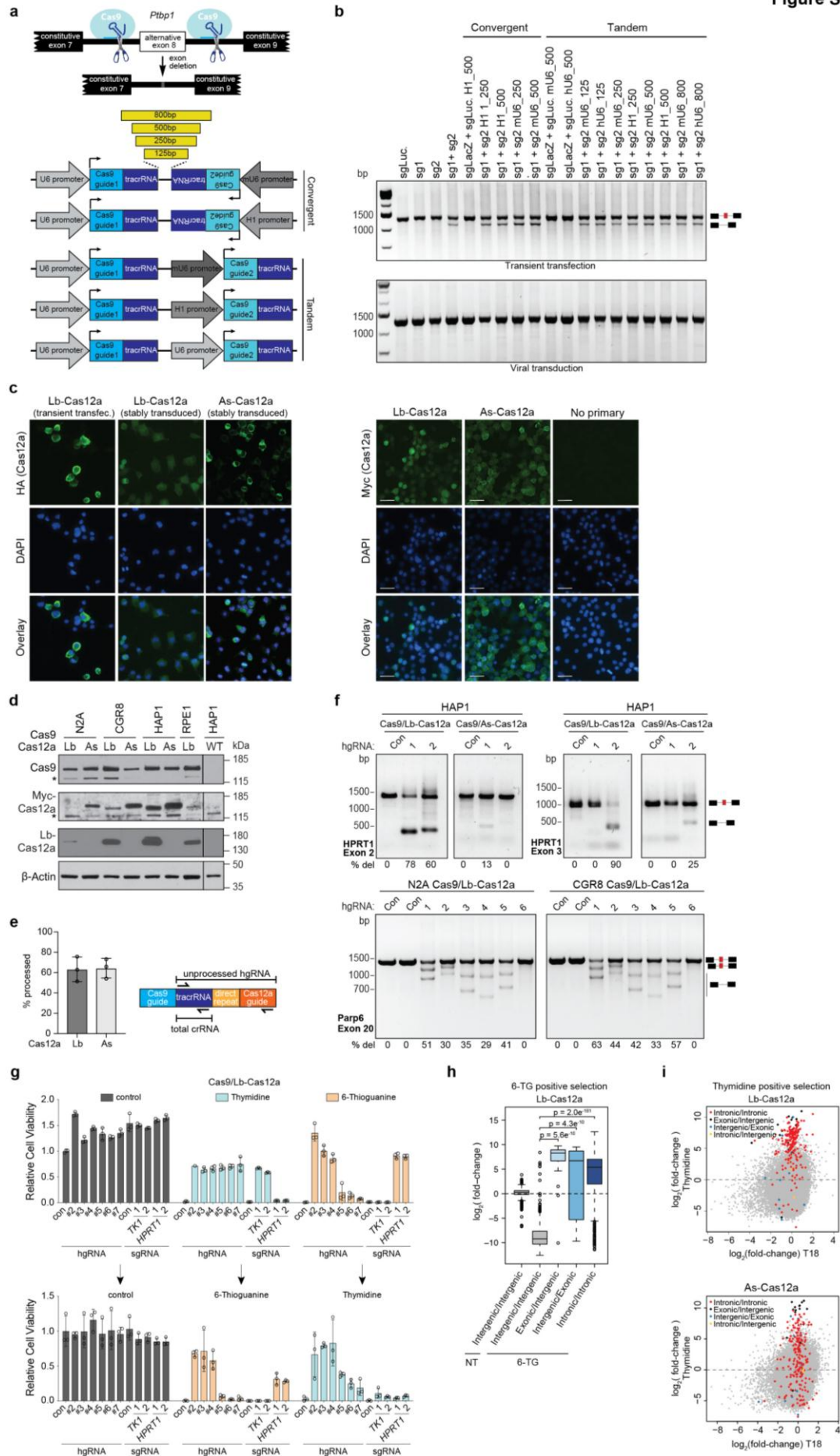
Thomas Gonatopoulos-Pournatzis ^{1,6} ✉, Michael Aregger ^{1,6}, Kevin R. Brown ^{1,6},
Shaghayegh Farhangmehr^{1,2}, Ulrich Braunschweig ¹, Henry N. Ward ³, Kevin C. H. Ha ¹,
Alexander Weiss¹, Maximilian Billmann ⁴, Tanja Durbic¹, Chad L. Myers ^{3,4}, Benjamin J. Blencowe^{1,2} ✉
and Jason Moffat ^{1,2,5} ✉

¹Donnelly Centre, University of Toronto, Toronto, Ontario, Canada. ²Department of Molecular Genetics, University of Toronto, Toronto, Ontario, Canada.

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

⁶These authors contributed equally: Thomas Gonatopoulos-Pournatzis, Michael Aregger, Kevin R. Brown. ✉e-mail: t.gonatopoulos@gmail.com; b.blencowe@utoronto.ca; j.moffat@utoronto.ca

Figure S1

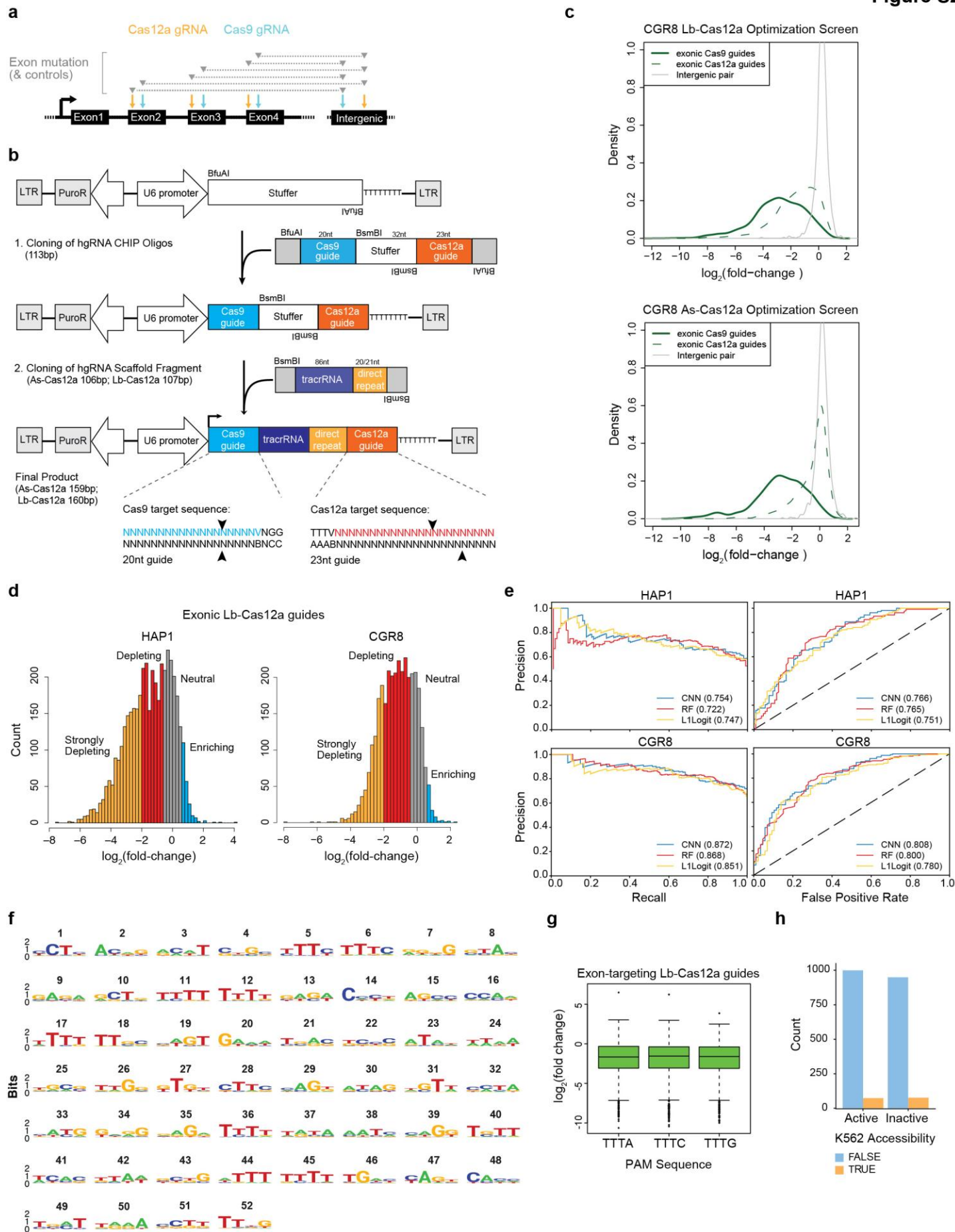


Supplementary Figure 1

Generation of dual Cas9 gRNA expression vectors for exon deletion.

- (a)** Schematic of targeting *Ptbp1* exon 8 for deletion (top panel) and of dual Cas9 gRNA expression cassettes (bottom panel).
- (b)** PCR monitoring of *Ptbp1* exon 8 deletion in CGR8 cells transiently transfected (top panel) or transduced (bottom panel) with dual Cas9 guides (see Figure S1A). Representative data from two independent experiments.
- (c)** Immunofluorescence analysis of N2A cells transiently transfected or stably transduced with lenti-Lb- or As-Cas12a containing a single nuclear localization signal (left panel). Immunofluorescence analysis of stably transduced N2A cells with lenti-Lb- or As-Cas12a containing two nuclear localization signals (right panel). Scale corresponds to 27 μm . Representative data from two independent experiments.
- (d)** Western blot analysis of Cas9 and Cas12a in stably transduced N2A, CGR8, HAP1 and RPE1 cells, as indicated. Asterisk indicates non-specific signal. Representative data from two independent experiments.
- (e)** hgRNA pre-crRNA processing based on qRT-PCR analysis. The strategy used for the quantification is indicated below the panel. All data are represented as mean $\pm$ SD ($n = 3$ replicates).
- (f)** PCR monitoring of exon deletion from *Parp6* and *HPRT1* genes in the indicated cell lines using CHyMERa. Independent pLCHKO constructs expressing Cas9 and Cas12a gRNAs targeting flanking intronic sites for exon deletions or controls were used as indicated. Representative data from two independent experiments.
- (g)** Relative cell viability following sequential drug treatments (thymidine and 6-thioguanine) of HAP1 cells transduced with pLCHKO vectors expressing hgRNAs targeting *TK1* and *HPRT1*, as indicated in the schematic on the left. As described in Fig. 1c, 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. After subjecting cells to the first drug treatment, cells were passaged at an equal ratio and challenged with the second drug treatment. Cell viability was assessed following both treatments using an AlamarBlue assay. Data represented as mean $\pm$ SD, $n = 3$ independent biological replicates.
- (h)** Enrichment of intergenic, exonic and intronic *HPRT1* targeting hgRNAs in non-treated (NT) or 6-TG treated HAP1 cells (pairwise two-tailed Mann-Whitney U test with Holm multiple testing correction). Data derived from $n = 3$ independent technical replicates. Boxes show the interquartile range (IQR - 25th to 75th percentile), with the median indicated by a line. The whiskers extend to the quartile $\pm 1.5 \times$ IQR.
- (i)** Enrichment of paired guides targeting an exon in *TK1* for deletion (red), or gene knockout (black = Cas9, blue = Cas12a) in double-thymidine block-treated (y-axis) vs. non-treated (x-axis) HAP1 cells at the end time point (T18). Intronic/intergenic control hgRNAs are indicated in yellow and non-*TK1* targeting constructs detectable in at least one Thymidine-treated replicate ($n=43,117$ sequences) are shown in grey. Screens performed with Lb-Cas12a and As-Cas12a are depicted in the top and bottom panel, respectively.

Figure S2



Supplementary Figure 2

Feature analysis of Cas12a guides.

(a) Schematic of exon-targeting hgRNA approach with CHyMERa.

(b) HgRNA screening libraries were generated by performing two rounds of Golden Gate assembly. During the first step the synthesized 113-nt oligos containing both Cas9 and Cas12a guides were introduced into pLHCKO. During the second step, the spacer sequence between the two oligos was replaced with a hybrid scaffold consisting of the Cas9 tracrRNA followed by the Lb- or As-Cas12a direct repeat (DR). Schematic of Cas9 and Cas12a guide length, PAM sequence and double-stranded DNA cutting pattern is indicated at the bottom.

(c) Dropout of Cas9 or Cas12a guides targeting essential genes in CGR8 cells.

(d) Exonic Lb-Cas12a guides were grouped based on LFC cut-offs in the HAP1 and CGR8 optimization screens. Strongly depleting guides were used as positive, and neutral guides as negative cases.

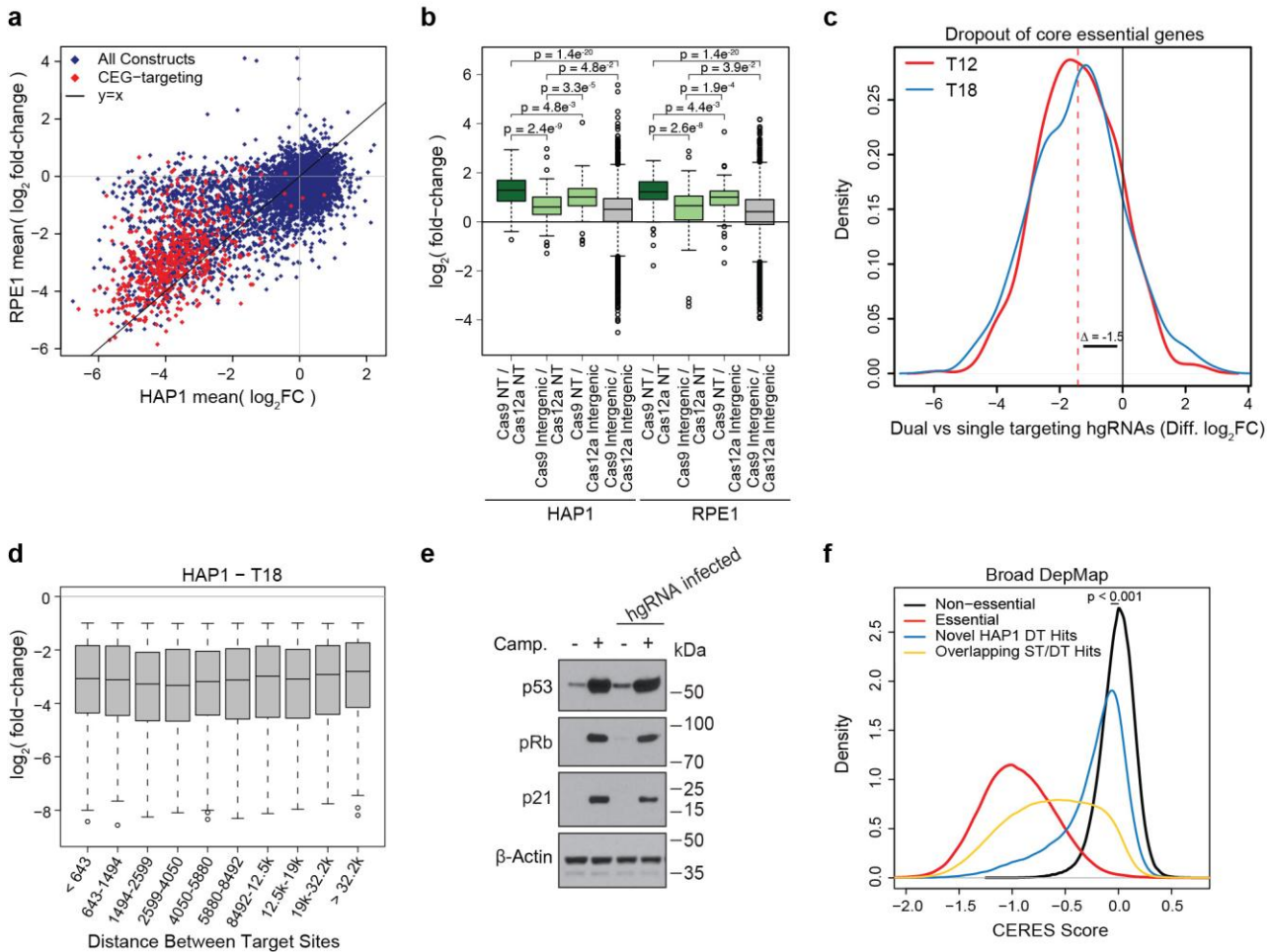
(e) Precision-recall (left panel) and receiver operating characteristic (right panel) curves of different machine-learning approaches for predicting Cas12a guide performance in HAP1 and CGR8 cells. CNN: convolutional neural networks; L1 Logit: lasso regularized logistic regression; RF: random forest.

(f) Weblogo of filters learned by CHyMERa-Net in the convolutional layer.

(g) Performance of exonic Lb-Cas12a hgRNAs grouped according to their PAM sequence. (n = 1,237 TTTA, 1,882 TTTC, 1,547 TTTG). Data derived from n = 3 independent technical replicates. Boxes show the interquartile range (IQR - 25th to 75th percentile), with the median indicated by a line. The whiskers extend to the quartile $\pm 1.5 \times$ IQR.

(h) Enrichment analysis of active and inactive Lb-Cas12a guides based on chromatin accessibility from K562 cells.

Figure S3



Supplementary Figure 3

Second-generation CHyMERa libraries provide enhanced gene targeting.

(a) Correlation between the mean LFC of hgRNA-targeted genes in HAP1 and RPE1 cells at the T18 timepoint. HgRNAs targeting core essential genes (CEG2) are indicated in red and all other hgRNAs are indicated in blue.

(b) LFC distribution of hgRNAs targeting intergenic regions and/or non-targeting (NT) regions in HAP1 and RPE1 cells. Each box comprises 92 hgRNAs except the intergenic:intergenic controls, which comprise 4,993 constructs. Distributions compared by two-tailed Wilcoxon rank-sum tests with Benjamini-Hochberg multiple testing correction. Data derived from $n = 3$ independent technical replicates. Boxes show the interquartile range (IQR - 25th to 75th percentile), with the median indicated by a line. The whiskers extend to the quartile $\pm 1.5 \times$ IQR.

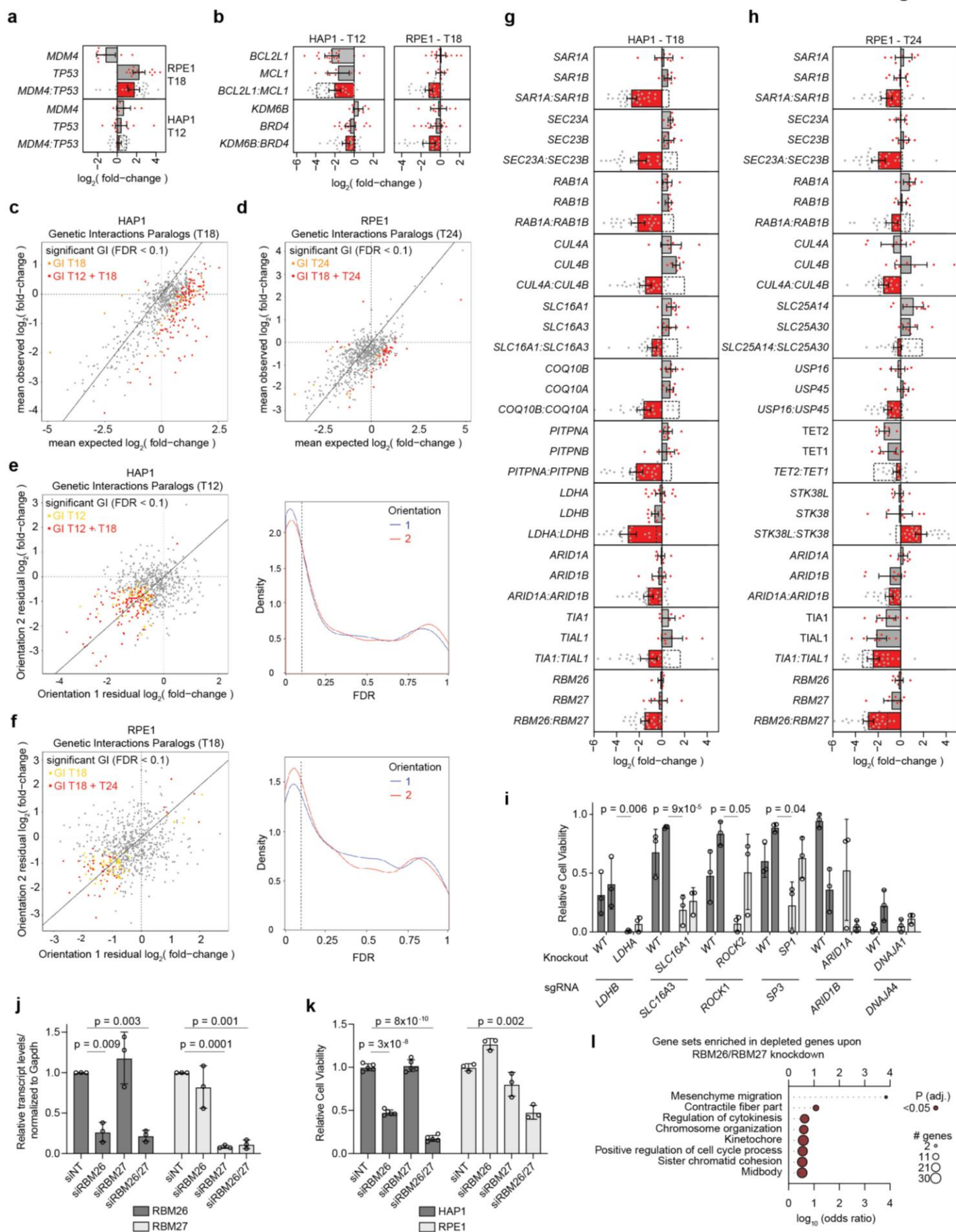
(c) Enhanced depletion of genes using the dual-targeting hgRNAs compared to single-targeting Cas9 guides.

(d) Dropout profiles of dual-targeting hgRNAs, as measured by the LFC at T18 in the HAP1 cell line, were binned into ten equal sized bins ($n = 1,093 - 1,097$) according to the distance between Cas9 and Cas12a target sites. Data derived from $n = 3$ independent technical replicates. Boxes show the interquartile range (IQR - 25th to 75th percentile), with the median indicated by a line. The whiskers extend to the quartile $\pm 1.5 \times$ IQR.

(e) Western blot depicting p53, pRb and p21 protein levels following camptothecin treatment in RPE1 CHyMERa cells transduced or not with hgRNA constructs. Representative data of two independent experiments.

(f) CERES scores from the DepMap CRISPR screens are shown for CEG2 (red) and non-essential (black) genes, genes discovered by both single- (ST) and dual-targeting (DT) (yellow), or genes discovered only through dual-targeting by CHyMERa (blue). Lower CERES scores correspond to greater depletion through the screens. CERES scores for each gene set across all 558 screens were aggregated together for plotting: red – 367,164 scores corresponding to 658 genes, yellow – 990,450 scores from 1,775 genes, blue – 313,038 scores from 561 genes, black – 435,798 scores from 781 genes. CERES score distributions for CHyMERa DT-only genes (n = 313,038) and non-essential genes (n = 435,798) were compared using a two-tailed Wilcoxon rank-sum test.

Figure S4



Supplementary Figure 4

CHyMERa reveals widespread non-additive fitness phenotypes upon combinatorial perturbation of paralogous genes.

(a-b) Enrichment or depletion of single or combinatorial gene ablations as indicated. The expected combinatorial effect size based on single perturbation is indicated with dotted bars. All data are represented as mean $\pm$ 2*SEM. Number of hgRNAs: *MDM4*, *MCL1*, *KDM6B* – 8; *TP53*, *BCL2L1*, *BRD4* – 16; paralog pairs – 30.

(c-d) Expected vs observed LFC of paralog pairs in HAP1 (c) or RPE1 (d) cells. Paralogs displaying significant genetic interaction at both time points, or only at the late time point, are highlighted in red and orange, respectively. Two-tailed Wilcoxon rank-sum test with Benjamini–Hochberg multiple testing correction, n = 3 independent technical replicates.

(e-f) Difference (residual) between the mean observed and expected LFCs for orientation 1 against orientation 2 of paralog pairs in HAP1 (e) or RPE1 (f) cells (left panel). Paralogs with significant genetic interactions at the early or both time points are highlighted in yellow and red, respectively. Two-tailed Wilcoxon rank-sum test, Benjamini–Hochberg multiple testing correction, n = 3 independent technical replicates. The density of FDR values for all gene pairs in both orientations are also displayed, and the significance threshold of 0.1 is indicated as a dashed line (right panel).

(g-h) Enrichment or depletion of single or combinatorial gene ablations in HAP1 (g) or RPE1 (h), as indicated. LFC represented as mean $\pm$ 2*SEM from n = 3 independent technical replicates.

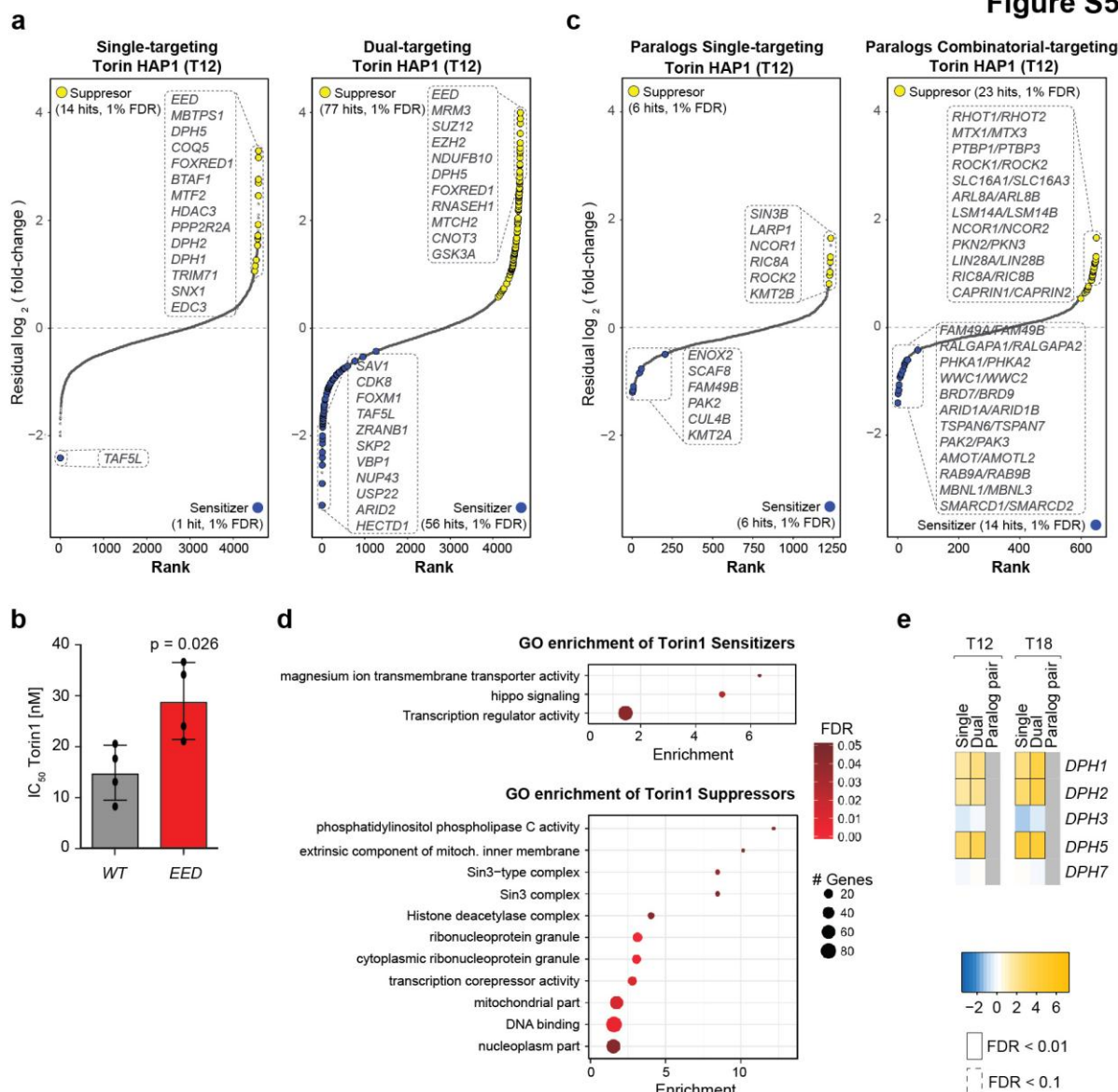
(i) Cell viability of WT and single-knockout HAP1 clones was measured by AlamarBlue staining 6 days post-transduction of two independent lentiCRISPRv2 gRNA expression cassettes targeting the indicated genes. Cell viability was normalized to intergenic-targeting control gRNAs. All data are represented as mean $\pm$ SD (n = 3 independent biological replicates; two-tailed t test).

(j) Real-time RT-PCR was performed to quantify *RBM26* and *RBM27* knock-down efficiency in HAP1 cells. All data are represented as mean $\pm$ SD (n = 3 independent biological replicates; two-tailed t test).

(k) Cell viability of HAP1 and RPE1 cells was measured by AlamarBlue staining 3 days post-transfection of siRNAs targeting *RBM26*, *RBM27* or both. All data are represented as mean $\pm$ SD (n = 5 and 3 independent biological replicates for HAP1 and RPE1, respectively; two-tailed t test).

(l) Gene ontology enrichment analysis for genes with significantly decreased expression upon co-depletion of *RBM26* and *RBM27* following siRNA treatment. (n = 2 independent biological replicates. FDR was calculated using FuncAssociate (Berriz et al., *Bioinformatics*, 2003).

Figure S5



Supplementary Figure 5

CHyMERa outperforms single Cas9 targeting chemo-genetic screens.

(a) Differential LFC of genes perturbed by single- (left panel) and dual-targeting (right panel) hgRNAs upon Torin1 treatment in HAP1 cells at the early time point (T12). Sensitizer (blue) and suppressor gene hits (yellow) are highlighted (FDR < 0.01, two-tailed Wilcoxon rank-sum test with Benjamini–Hochberg multiple testing correction, $n = 3$ independent technical replicates) and the top 10 as well as selected genes from the top 20 significant hits are listed.

(b) Torin1 IC₅₀ values (drug concentration resulting in 50% reduction of cell viability) in HAP1 WT and *EED* knockout cell clones. IC₅₀ values were calculated based on dose response curves in the respective HAP1 cell lines. Data represented as mean $\pm$ SD ($n = 4$ independent biological replicates; $p = 0.026$, two-tailed t test).

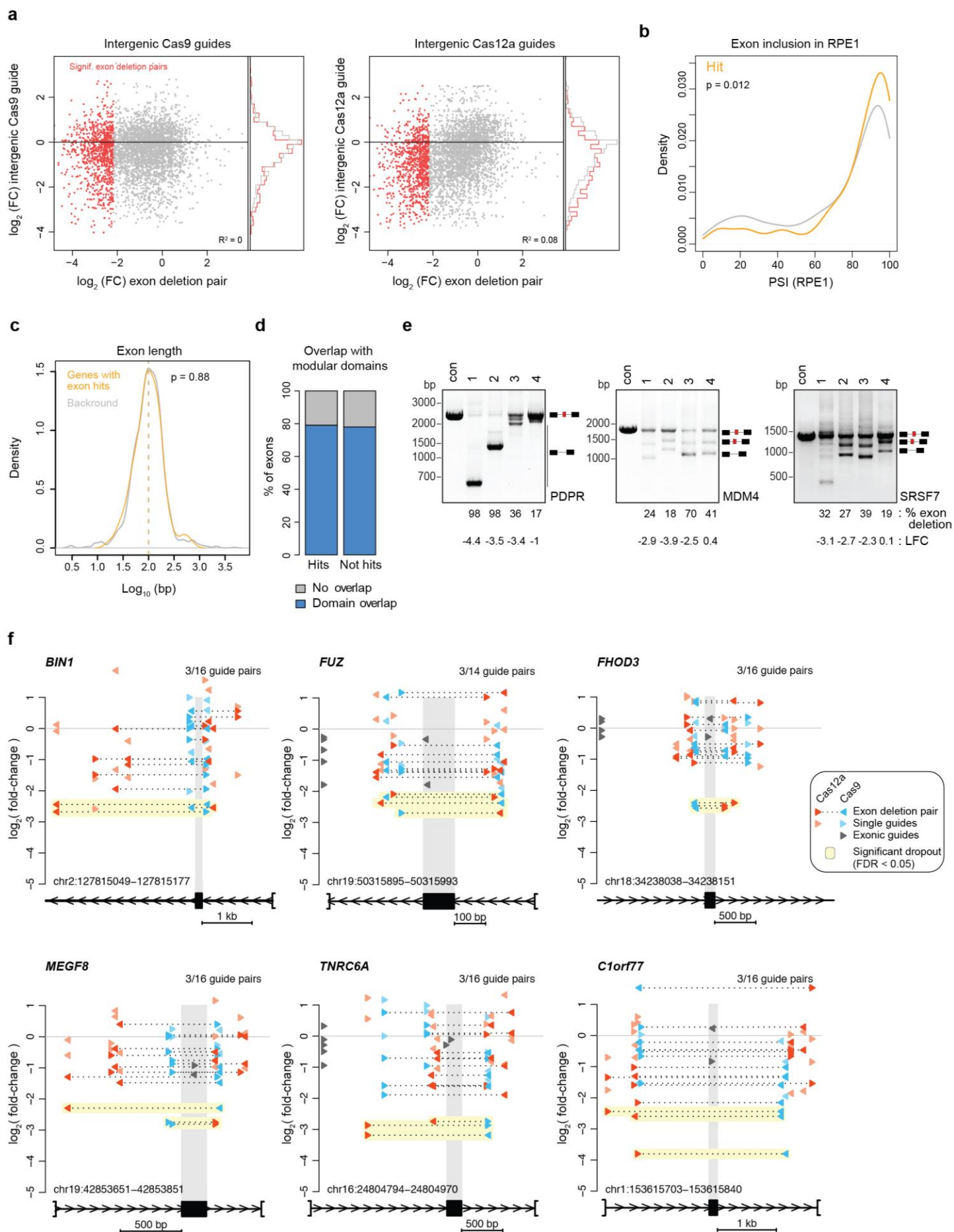
(c) Differential LFC of paralogs perturbed by single- (left panel) and combinatorial-targeting (right panel) hgRNAs upon Torin1 treatment in HAP1 cells at the early time point (T12). Sensitizer (blue) and suppressor gene hits (yellow) are highlighted (FDR < 0.01, two-tailed

Wilcoxon rank-sum test with Benjamini–Hochberg multiple testing correction, $n = 3$ independent samples) and the top 10 as well as selected genes from the top 20 significant hits are listed.

(d) Gene ontology enrichment of sensitizer (upper panel) or suppressor hits (lower panel) identified at an FDR < 0.1 across both time points ($n = 3$ independent technical replicates). FDR was calculated using GOrilla (Eden et al., *BMC Bioinformatics*, 2009).

(e) Differential LFC of diphthamide biosynthesis genes perturbed by single- or dual-targeting hgRNAs as indicated. Two-tailed Wilcoxon rank-sum test with Benjamini–Hochberg multiple testing correction, $n = 3$ independent technical replicates.

Figure S6



Supplementary Figure 6

CHyMERa is suitable for exon-deletion phenotypic screens.

(a) The LFC of exon-deletion hgRNAs (intronic/intronic) vs. control hgRNAs in which only the Cas9 (left) or Cas12a guide (right) is targeting an intronic region, while the other nuclease is targeting an intergenic region. The red dots represent exon-deletion hgRNAs that are significantly depleted, while grey dots represent all other exon-deletion hgRNAs. Significant depletion was scored against the empirical null distribution of 1,647 intergenic-intergenic control pairs (refer to Methods for details). Marginal histograms indicate the density distribution of control guide pairs corresponding to significant and non-significant exon-deletion pairs, respectively.

(b) The density of exonic “hits” (orange) compared to all other exons (grey) from the exon-deletion screen as a function of PSI (percent spliced in). p-value is from a two-tailed Mann-Whitney U test ($n = 91$ for hits, 1,514 for background).

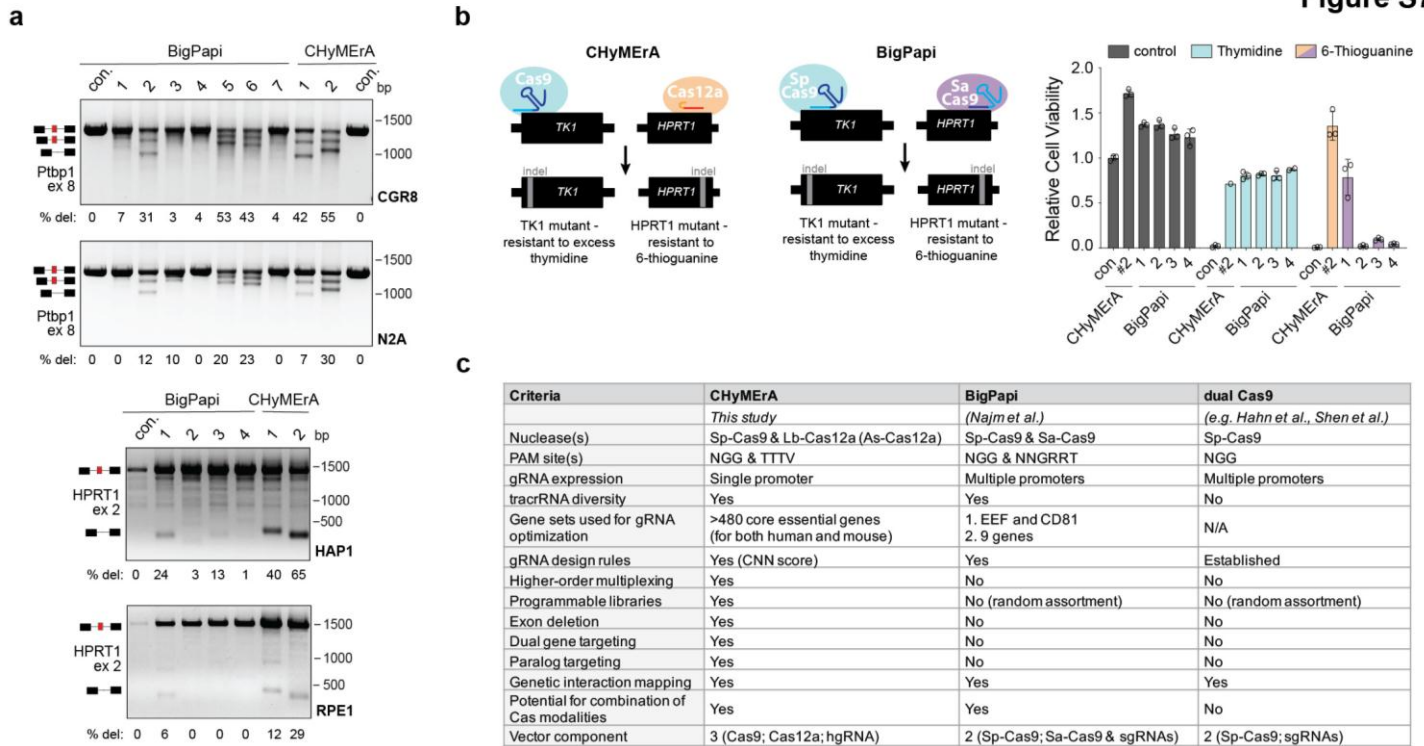
(c) Length distribution of the alternative exons targeted by CHyMERa exon-deletion library. p-value is from a two-tailed Mann-Whitney U test ($n = 91$ for hits, 1,514 for background).

(d) Percentage of alternative exons that overlap a modular protein domain.

(e) PCR monitoring of exon deletion from *PDPR*, *MDM4* and *SRFS7* genes in RPE1 cells using hgRNAs guides with different phenotypic scores. Representative data of two independent experiments.

(f) Representative examples of hgRNA constructs targeting frame-disruptive exons in *BIN1*, *FUZ*, *FHOD3*, *MEGF8*, *TNRC6A* or *C1orf77* are shown above the gene model (x-axis), with the observed LFC value for each hgRNA (y-axis). Exon deletion (i.e. intronic-intronic) and single-targeting control (i.e. intronic-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).

Figure S7



Supplementary Figure 7

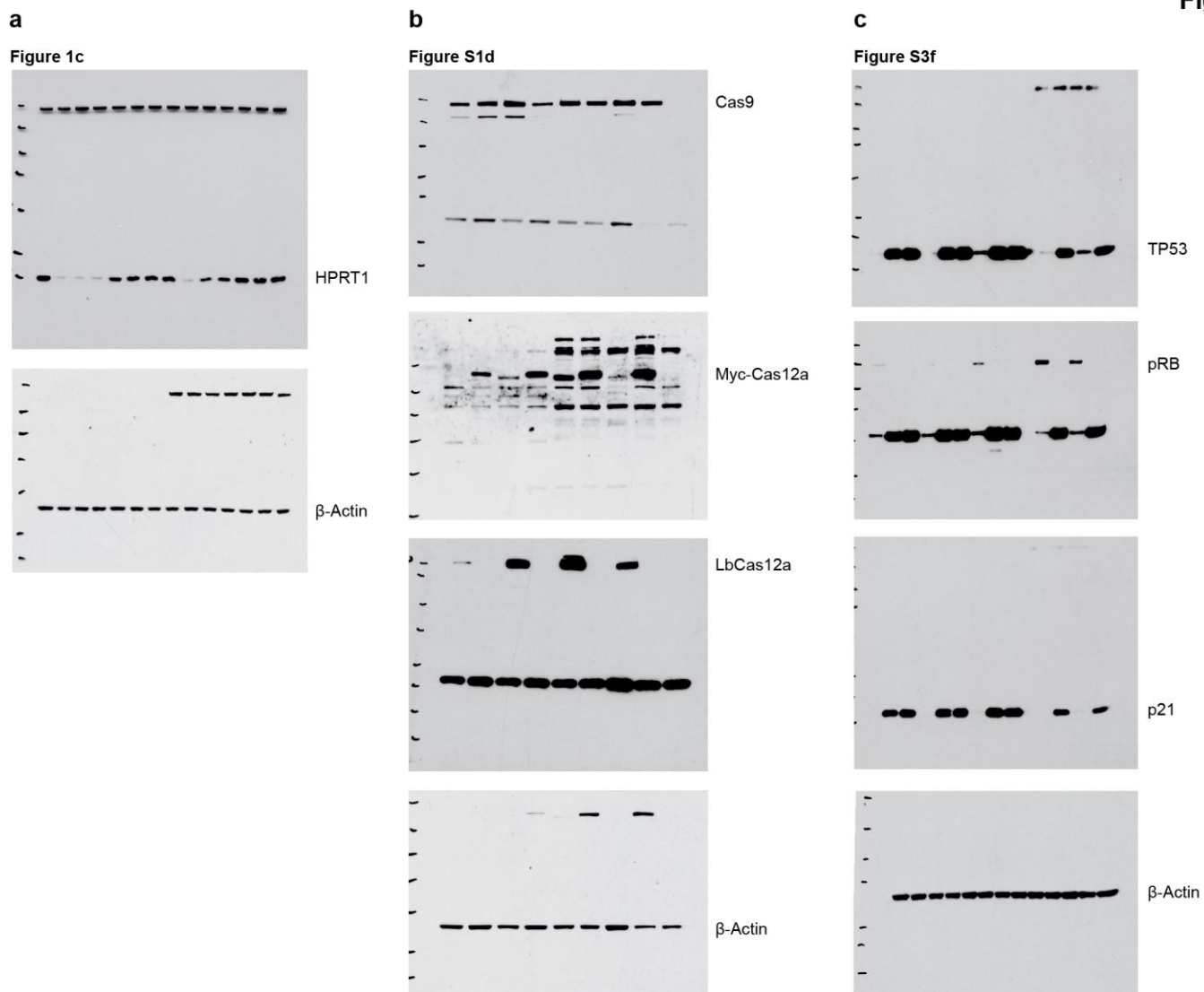
Comparison of CHyMERa with other dual-targeting screening systems.

(a) PCR monitoring of exon deletion from *Ptpb1* and *HPRT1* genes in the indicated cell lines using CHyMERa or BigPapi. Independent pLCHKO and pPapi constructs expressing Sp-Cas9 and Cas12a (CHyMERa) or Sa-Cas9 (BigPapi) gRNAs targeting flanking intronic sites for exon deletions or controls were used as indicated. Representative data of two independent experiments.

(b) Schematic of combinatorial gene targeting by CHyMERa (left panel) or BigPapi (middle panel). Comparison between CHyMERa and the BigPapi system for the combinatorial targeting of *TK1* and *HPRT1*, as determined by resistance to thymidine and 6-thioguanine treatments, respectively (right panel). The same Cas9 guide targeting *TK1* was used for CHyMERa and all BigPapi constructs. Data represented as mean $\pm$ SD, $n = 3$ independent biological replicates.

(c) Table summarizing the key characteristics and applications of dual-targeting CRISPR screening systems.

Figure S8



Supplementary Figure 8

Raw data for Western blots included in the figures.

(a-c) Uncropped images of western blots displayed in Fig. 1c (a), Supplementary Fig. 1d (b) and Supplementary Fig. 3f (c).