nature methods

BRIEF COMMUNICATION

https://doi.org/10.1038/s41592-018-0007-1

Discovery of proteins associated with a predefined genomic locus via dCas9–APEX-mediated proximity labeling

Samuel A. Myers[1*], Jason Wright[1], Ryan Peckner[1], Brian T. Kalish[2,3], Feng Zhang[1] and Steven A. Carr[1*]

Regulation of gene expression is primarily controlled by changes in the proteins that occupy genes' regulatory elements. We developed genomic locus proteomics (GLoPro), in which we combine CRISPR-based genome targeting, proximity labeling, and quantitative proteomics to discover proteins associated with a specific genomic locus in native cellular contexts.

Transcriptional regulation is a highly coordinated process that is largely controlled by changes in protein occupancy at regulatory elements of the modulated genes. Chromatin immunoprecipitation (ChIP) has been invaluable for furthering understanding of transcriptional regulation and chromatin structure both at individual loci and genome-wide[1,2]. However, because ChIP requires highly specific antibodies, its utility often is limited. Previously developed 'inverse ChIP' methods have had limited utility in mammalian systems owing to several drawbacks, including loss of cellular and/or chromatin context, extensive engineering and locus disruption, reliance on repetitive DNA sequences, and chemical cross-linking[3–8], which often requires extensive optimization for mass-spectrometry-based applications[9]. We developed GLoPro, a method to identify proteins associated with a specific, nonrepetitive genomic locus in the native cellular context without the need for cross-linking or genomic alterations.

We fused catalytically dead RNA-guided nuclease Cas9 (dCas9)[10] to the engineered peroxidase APEX2[11] and targeted specific genomic loci with single guide RNAs (sgRNAs)[12] (Fig. 1a and Supplementary Fig. 1). APEX2, in the presence of hydrogen peroxide, oxidizes the phenol moiety of biotin–phenol compounds to phenoxyl radicals that react with surface exposed tyrosine residues, thus labeling nearby proteins with biotin derivatives[13]. We chose APEX2 for its small labeling radius and short reaction time[14,15]. We cloned the dCas9–APEX2 gene (*Caspex*) in-frame with the self-cleaving T2A peptide and *Gfp* under the control of a tetracycline response element into a puromycin-selectable piggyBac plasmid (Supplementary Fig. 2). The inducibility of *Caspex* expression provides temporal control to minimize the amount of time that CASPEX occupies the targeted locus and the accumulation of excess CASPEX, which leads to higher background biotinylation, common with proximity labeling[14–16].

To determine whether the CASPEX protein correctly localized to the genomic site of interest, we created a single-colony HEK293T line with stably integrated *Caspex* plasmid (293T-Caspex) and stably expressed sgRNAs tiling the human *TERT* promoter (*hTERT*) (Fig. 1b). Only one sgRNA was expressed per line. We focused on *hTERT* because *TERT* expression is a hallmark of cancer, and recurrent promoter mutations in *hTERT* have been shown to reactivate *TERT* expression[17]. We induced biotinylation in each sgRNA-expressing 293T-Caspex line, and then carried out anti-Flag or anti-biotin ChIP-qPCR with probes tiling *hTERT*. ChIP-qPCR showed proper localization of CASPEX with the peak of the anti-Flag signal overlapping the destination of the sgRNA. The anti-biotin ChIP-qPCR signal showed a similar trend of enrichment, indicating that CASPEX biotinylated proteins within approximately 400 bp of either side of its target locus (Supplementary Fig. 3). As expected, we did not observe any enrichment at the targeted locus for the no-sgRNA controls. To assess the off-target binding of CASPEX, we carried out anti-Flag ChIP-qPCR to probe the top predicted off-target site for each respective sgRNA. No two off-target sites were within 5 Mb of each other (Supplementary Table 1). Each on-target site showed CASPEX occupation that was 3- to 40-fold higher than that of the predicted off-target site, and the cumulative enrichment of the sgRNA-expressing 293T-Caspex lines with overlapping labeling radii (430T, 107T, T092, and T266) at *hTERT* was 50-fold (Supplementary Fig. 4). Western blotting analysis showed that biotinylation was CASPEX dependent, but we did not observe any difference in biotin patterns between sgRNA lines (Supplementary Fig. 5).

To test whether CASPEX could be used to identify proteins associated with *hTERT*, we enriched biotinylated proteins from *hTERT*-targeted 293T-Caspex lines and analyzed the proteins by quantitative LC-MS/MS. We initiated biotinylation in the five individual *hTERT*-targeting 293T-Caspex lines that tiled the genomic locus of interest, as well as in the no-guide control 293T-Caspex line. The no-guide control was chosen for its negligible interaction and residence time with irrelevant genomic loci when assayed by ChIP-seq[18,19]. Tiling is an important feature of this method, as 'noise' from off-target binding of dCas9 from each individual line will be diluted, and only reproducibly enriched proteins from on-target occupancy contribute to the 'signal'[18,19]. Tiling may also circumvent the loss of native protein identification if dCas9 binding precludes protein occupancy. We incubated whole-cell lysates from each line with streptavidin-coated beads, washed the beads stringently, and carried out on-bead trypsin digestion. We labeled digests of the enriched proteins with isobaric tandem mass tags for relative quantitation, then mixed the digests and analyzed them by LC-MS/MS (Supplementary Fig. 1). We used a ratiometric approach for each individual sgRNA 293T-Caspex line compared with the no-guide control line. Enrichment from the four overlapping *hTERT* 293T-Caspex lines showed high correlation of protein enrichment (Supplementary Fig. 6). The T959 293T-Caspex line, which lies ≥2

[1]Broad Institute of MIT and Harvard, Cambridge, MA, USA. [2]Department of Neurobiology, Harvard Medical School, Boston, MA, USA. [3]Division of Newborn Medicine, Department of Medicine, Boston Children's Hospital, Boston, MA, USA. *e-mail: sammyers@broadinstitute.org; scarr@broad.mit.edu

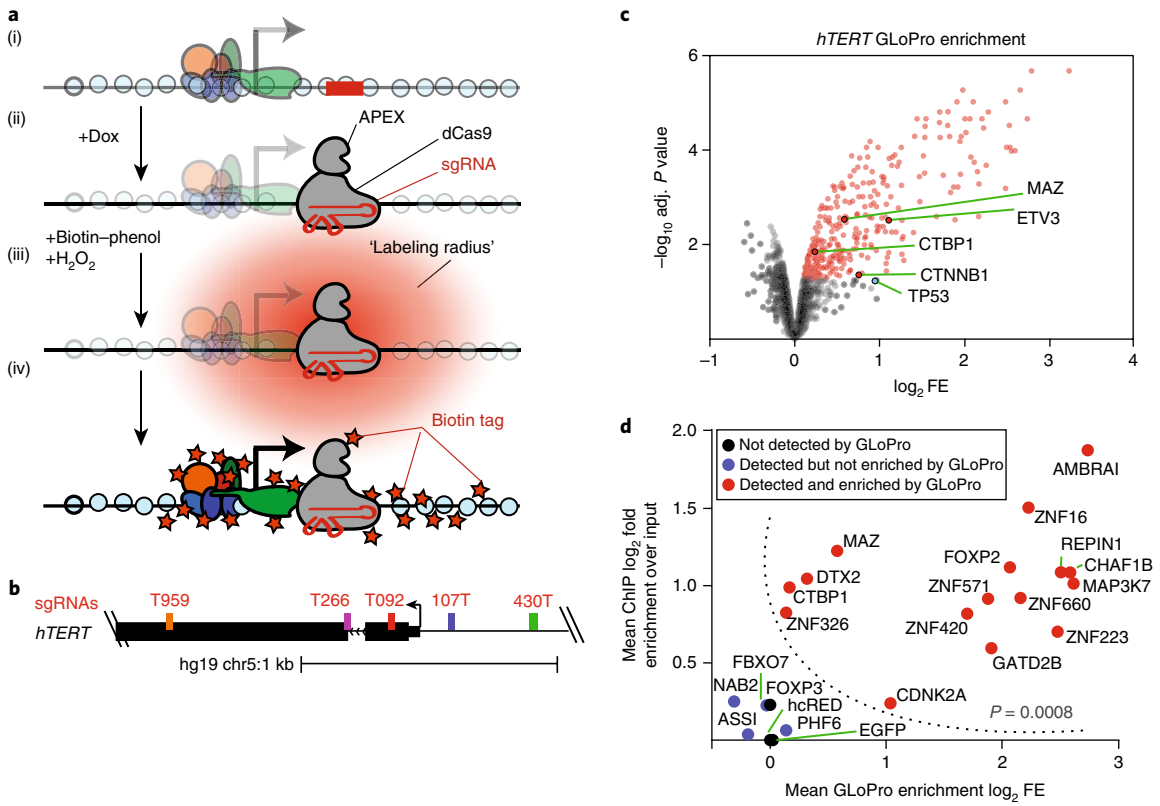

Fig. 1 | Genomic locus proteomics of *hTERT*. **a**, Illustration of CASPEX targeting and affinity-labeling reaction. (i) A specific genomic locus is identified (red bar) and targeted with an sgRNA. (ii) CASPEX expression is induced with doxycycline (dox), and the protein binds the genomic region of interest. (iii) The 'labeling radius' of the reactive biotin–phenol is represented by a 'cloud' of red shading. (iv) Proteins proximal to CASPEX are labeled with biotin (orange stars) for subsequent enrichment. **b**, UCSC Genome Browser representation of *hTERT* (hg19). sgRNAs (colored bars) are shown relative to the transcription start site (black arrow). **c**, Volcano plot of proteins quantified across the four overlapping *hTERT* 293T-Caspex cell lines ($n = 4$ independent sgRNA lines) compared with protein levels in the no-sgRNA control. Data points representing proteins enriched with an adjusted (adj.) *P* value of <0.05 and $\log_2$ fold enrichment (FE) > 0 are red. Proteins known to associate with *hTERT* and identified as enriched by GLoPro are labeled and indicated by green lines. TP53, a known *hTERT* binder, had an adjusted *P* value of 0.058 and is represented by a blue data point. **d**, Correlation between ChIP-qPCR mean $\log_2$ fold enrichment over input of the four primer pairs spanning the sgRNA targets (biological quadruplicates, measurement singlicate) and GLoPro enrichment of the four overlapping sgRNAs at *hTERT*. The dotted line separates ChIP-qPCR data tested for statistical significance via Mann–Whitney test; the *P* value is shown in the lower right.

labeling radii from its closest neighbor, showed decreased correlation of protein enrichment. We conducted a moderated one-sample *t*-test by treating the four overlapping sgRNA lines as replicates, using the nonspatially constrained no-sgRNA 293T-Caspex line as the control for background biotinylation and enrichment. We found that 387 of the 3,199 proteins identified with at least two peptides were significantly enriched (adjusted $P < 0.05$; fold enrichment > 0) at *hTERT* relative to the no-sgRNA control, including 5 proteins known to occupy *hTERT* in various cell types (Fig. 1c, Supplementary Table 2). Histones and subunits of RNA polymerase II, although detected, were not significantly enriched in comparison with the control, probably because of the inverse relationship between a protein's abundance and the ability to determine that it is enriched relative to the background level (Supplementary Fig. 7).

To validate a number of candidate proteins associated with *hTERT*, we carried out ChIP-qPCR against candidates spanning the GLoPro enrichment range. ChIP-grade antibodies are not available for many of these proteins, so we turned to V5-tagged open reading frame (ORF) expression in unmodified HEK293T cells. We individually transfected 16 V5-tagged candidate ORFs, 4 V5-tagged ORFs for proteins not significantly enriched at *hTERT*, and 3 proteins that were not detected (negative controls). When we compared anti-V5 ChIP-qPCR signals from each individual ORF to the respective GLoPro enrichment values, we found that all proteins enriched in the GLoPro analysis were, as a group, statistically enriched by ChIP-qPCR (Mann–Whitney test, $P = 0.0008$) (Fig. 1d). Most candidates deemed statistically enriched according to the GLoPro analysis were separated in the ChIP enrichment space from those not enriched or not detected. Two proteins previously reported to bind *hTERT*, CTBP1 and MAZ, were found in a regime of high ChIP enrichment and low GLoPro enrichment, which suggests that ChIP-qPCR provides information orthogonal to that obtained by GLoPro for protein occupancy at a genomic locus.

To explore the generalizability of GLoPro, we created 293T-Caspex cells that express individual sgRNAs tiling the *MYC* promoter (Fig. 2a). ChIP-qPCR against CASPEX verified the proper localization of each *MYC* 293T-Caspex line (Supplementary Fig. 8), and off-target binding analysis showed cumulative 32-fold enrichment at the *MYC* promoter compared with levels at any predicted off-target site (Supplementary Fig. 9). GLoPro analysis of the *MYC* promoter identified 66 proteins as significantly enriched (adjusted $P < 0.05$) compared with amounts in the no-guide control (Fig. 2b, Supplementary Table 3). We applied a machine learning algorithm to identify the association of GLoPro-enriched proteins with canonical pathways from the Molecular Signature Database (MSigDB). We identified 21 statistically enriched networks (adjusted $P < 0.01$),

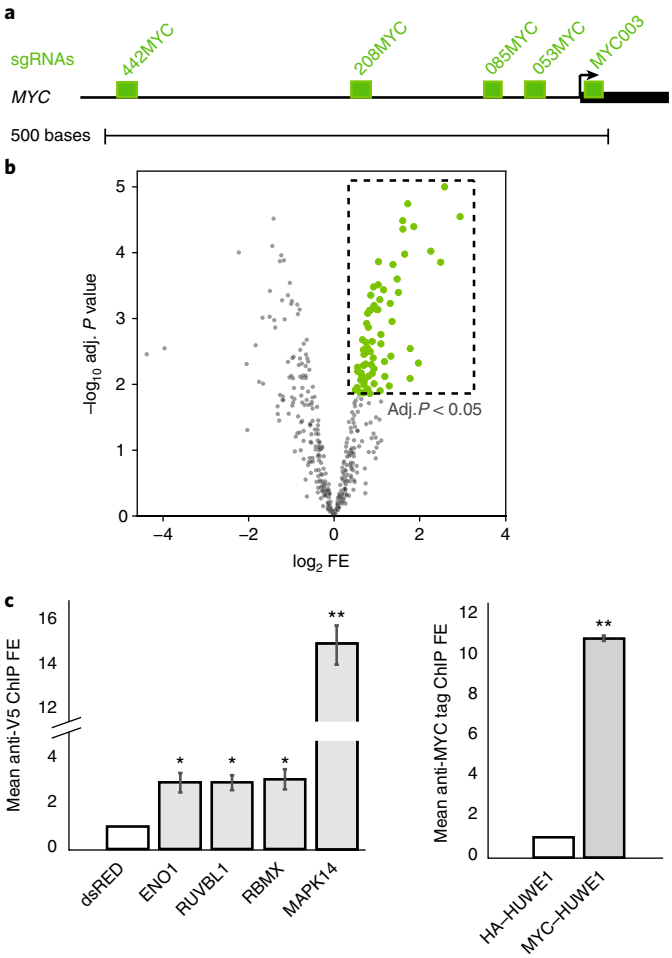

Fig. 2 | Genomic locus proteomic analysis of *MYC* promoter. a, UCSC Genome Browser representation (hg19) of the *MYC* promoter and the location of sgRNA sites relative to the transcription start site (black arrow). **b**, Volcano plot of proteins quantified across the five *MYC* 293T-Caspex cell lines compared with levels in the no-sgRNA control line ($n = 5$ independent sgRNA lines). Data points representing positively enriched proteins with an adjusted (adj.) P value of <0.05 and $\log_2$ fold enrichment (FE) > 0 are green and highlighted by a dashed rectangle. **c**, ChIP-qPCR of candidate proteins identified by GLoPro at the *MYC* promoter. V5-tagged dsRED served as the negative control for V5-tagged proteins ENO1, RUVBL1, RBMX, and MAPK14, whereas hemagglutinin (HA)-tagged HUWE1 was used for comparison with MYC-tagged HUWE1. $^*P < 0.05$, $^{**}P < 0.01$, t-test. Data are shown as mean and s.e. (transfection duplicates, measurement triplicates).

including 'MYC_active_pathway', a gene set of validated targets responsible for activating *MYC* transcription (Supplementary Fig. 10). ChIP-qPCR confirmed the presence of pathway components at the *MYC* promoter, including HUWE1, RUVBL1, and ENO1 for MYC_active_pathway; RBMX for mRNA_splicing_pathway; and MAPK14 for Lymph_angiogenesis_pathway (Fig. 2c).

The generalizability of dCAS9 and APEX2 suggests that GLoPro can be used in a wide variety of cell types and at any dCAS9-targetable genomic element. LC-MS/MS analysis with isobaric peptide labeling allows for sample multiplexing, and thus the measurement of multiple sgRNA lines and/or replicates in a single experiment with few or no missing values for relative quantitation of enrichment. GLoPro-derived candidate proteins can and should be validated for association with genomic regions by ChIP or another orthogonal assay, as the current version of the method is not yet comprehensive and is still subject to false positives/negatives. Although in this initial work GLoPro identifies only association with a locus, and not locus specificity (i.e., it is likely that a protein bound at the queried site binds elsewhere) or functional relevance, we expect that analysis of promoters or enhancer elements during relevant perturbations may provide novel functional insights into transcriptional regulation. In addition, we envision that CASPEX can be used for enrichment of genomic locus entities such as locus-associated RNAs or DNA elements in close three-dimensional space within the nucleus (i.e., enhancers). Further work will be needed to assess the extended capabilities of CASPEX.

Methods

Methods, including statements of data availability and any associated accession codes and references, are available at https://doi.org/10.1038/s41592-018-0007-1.

Received: 6 December 2017; Accepted: 12 March 2018;
Published online: 7 May 2018

References

1. Bernstein, B. E. et al. *Cell* **125**, 315–326 (2006).
2. Liber, D. et al. *Cell Stem Cell* **7**, 114–126 (2010).
3. Mittler, G., Butter, F. & Mann, M. *Genome Res.* **19**, 284–293 (2009).
4. Déjardin, J. & Kingston, R. E. *Cell* **136**, 175–186 (2009).
5. Pourfarzad, F. et al. *Cell Rep.* **4**, 589–600 (2013).
6. Schmidtmann, E., Anton, T., Rombaut, P., Herzog, F. & Leonhardt, H. *Nucleus* **7**, 476–484 (2016).
7. Liu, X. et al. *Cell* **170**, 1028–1043 (2017).
8. Tsui, C. et al. *Proc. Natl. Acad. Sci. USA* **115**, E2734–E2741 (2018).
9. Ly, T. et al. *eLife* **6**, e27574 (2017).
10. Qi, L. S. et al. *Cell* **152**, 1173–1183 (2013).
11. Lam, S. S. et al. *Nat. Methods* **12**, 51–54 (2015).
12. Cong, L. et al. *Science* **339**, 819–823 (2013).
13. Rhee, H. W. et al. *Science* **339**, 1328–1331 (2013).
14. Paek, J. et al. *Cell* **169**, 338–349 (2017).
15. Lobingier, B. T. et al. *Cell* **169**, 350–360 (2017).
16. Lambert, J. P., Tucholska, M., Go, C., Knight, J. D. & Gingras, A. C. *J. Proteomics* **118**, 81–94 (2015).
17. Huang, F. W. et al. *Science* **339**, 957–959 (2013).
18. Thakore, P. I. et al. *Nat. Methods* **12**, 1143–1149 (2015).
19. Wu, X. et al. *Nat. Biotechnol.* **32**, 670–676 (2014).

Acknowledgements

We are grateful to S. Shipman and G. Church (Harvard University, Cambridge, MA, USA) for the tetracycline-inducible/puro^R plasmid, and to A. Edge (Mass Eye and Ear, Boston, MA, USA) for the *HUWE1* plasmids. We thank A. Ting, J. Doench, N. Udeshi, A. Regev, P. Thakore, B. Hamilton, and E. Kvedaraite for useful discussion; E.M. Perez, J. Ray, and R. Issner for technical assistance; and M. Papanastasiou, E. Kuhn, and J. Abelin for critical assessment of the manuscript. This work was funded by the NCI CPTAC PCC (grant 1U24CA210986-01 to S.A.C.), the Broad Institute (SPARC grant to S.A.M.), the Pediatric Scientist Development Program (to B.T.K.), and the March of Dimes Birth Defects Foundation (to B.T.K.).

Author contributions

S.A.M. and J.W. conceived the idea. S.A.M., J.W., and S.A.C. conceived the experimental setup and designed the research. S.A.M. created the constructs and stable lines, and performed labeling, western blotting, enrichments, mass spectrometry, and ChIP-qPCR for the c-MYC lines. J.W. performed ChIP-qPCR for all *hTERT*-related experiments. R.P. and S.A.M. performed the computational and statistical analyses. B.T.K. and S.A.M. performed the off-target analyses. S.A.M., J.W., R.P., B.T.K., F.Z., and S.A.C. interpreted the data. S.A.M., J.W., and S.A.C. wrote the manuscript.

Competing interests

A patent application related to this work has been filed by The Broad Institute.

Additional information

Supplementary information is available for this paper at https://doi.org/10.1038/s41592-018-0007-1.

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

Correspondence and requests for materials should be addressed to S.A.M. or S.A.C.

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

Methods

Plasmid construction. We created the *Caspex* construct (doxycycline-inducible dCas9-APEX2-T2-GFP) by subcloning 3×Flag-*dCas9* and T2A-*Gfp* from pLV-hUBC-dCas9-VP64-T2A-GFP (Addgene #53192) and V5-APEX2-NLS from mito-V5-APEX2 (Addgene #42607) into an all-in-one piggyBac, TREG/Tet-3G plasmid[20] (Church lab), via ligation-independent cloning (InFusion; Clontech). Guide sequences were selected and cloned as previously described[21,22]. Guides were designed irrespective of sense/anti-sense strands but with spacing of 100–200 bp between guides, and on-target scores were used. All V5 ORF constructs[23] were purchased through the Broad Genetics Perturbation Platform and were expressed from the pLX-TRC_317 backbone. V5 ORFs were selected for validation only if the construct was available, had protein homology >99%, and had an in-frame V5 tag. Only two of the five known *hTERT* interactors (MAZ[24,25], CTNNB1[26–28], ETV3[29], CTBP1[30], and TP53[31,32]) were available. All constructs were individually transfected into unmodified HEK293T cells for anti-V5 ChIP experiments at one-fourth the recommended DNA amount to mitigate gross overexpression. The *Caspex* plasmid is available through Addgene (plasmid #97421).

Cell line construction and culture. HEK293T cells were grown in DMEM supplemented with heat-inactivated 10% FBS, glutamine, and non-essential amino acids (Gibco). All constructs were prepared using Zyppy Maxi prep kits (Zymo Research) and transfected with Lipofectamine 2000 (Thermo). After *Caspex* transfection, puromycin was added to a final concentration of 4 µg/ml and selection was carried out for 2 weeks. Single colonies were picked, expanded, and tested for doxycycline-inducibility of the Caspex construct, monitored by GFP detection and anti-Flag western blotting. The HEK293T cell line with the best inducibility (referred to as 293T-Caspex) was expanded and used for all subsequent experiments. For stable sgRNA expression, single sgRNA constructs were transfected into 293T-Caspex cells and selected for stable incorporation by hygromycin treatment at 200 µg/ml for 2 weeks. CASPEX binding was tested by ChIP followed by digital droplet PCR or Sybr qPCR as described below.

APEX-mediated labeling. Prior to labeling, doxycycline dissolved in 70% ethanol was added to the cell culture media to a final concentration of either 500 ng/mL for 18–24 h (*hTERT*) or 1 µg/mL for 12 h (*MYC*) for proteomic experiments. For off-target CASPEX-binding analysis, cells were treated with 1 µg/mL doxycycline for 20 h. Biotin tyramide phenol (Iris Biotech) in DMSO (stock concentration: 500 mM) was added directly to cell culture media to a final concentration of 500 µM, and the media was swirled until the precipitate dissolved. After 30 min at 37 °C, hydrogen peroxide was diluted in media to 100 mM before being added to the cell culture media to a final concentration of 1 mM to induce biotinylation. After 60 s of very gentle swirling, the media was decanted as quickly as possible, and the cells were washed three times with 15 ml of ice-cold PBS containing 100 mM sodium azide, 100 mM sodium ascorbate, and 50 mM TROLOX (6-hydroxy-2,5,7,8-tetramethylchroman-2-carboxylic acid). Cells were scraped and transferred to 15-ml Falcon tubes with ice-cold PBS, spun at 500*g* for 3 min, flash-frozen in liquid nitrogen, and stored at −80 °C.

Chromatin immunoprecipitation followed by quantitative PCR. For confirmation of CASPEX binding and labeling, cells were trypsinized to a single-cell suspension, fresh formaldehyde was added to a final concentration of 1%, and samples were incubated at 37 °C for 10 min, with inversion several times every 2 min. Formaldehyde was quenched with 5% glycine for 37 °C for 5 min, and the samples were aliquoted into 3×10^6 cell aliquots, spun down, and flash-frozen in 0.5-mL Axygen tubes. Chromatin was sheared with a QSonica Q800R2 sonicator at an amplitude of 50 with 30 s on/30 s off for 7.5 min, until 60% of fragments were between 150 and 700 bp, with an average size of ~350 bp. Lysis buffer comprised 1% SDS, 10 mM EDTA, and Tris-HCl, pH 8.0. For off-target CASPEX analysis, one approximately 80% confluent 15-cm^2 plate was washed once with 20 ml of ice-cold PBS and then incubated with room temperature PBS with 1% formaldehyde (freshly prepared from 16% stock; Thermo) at room temperature for 10 min with gentle rocking. After cross-linking, 1.5 mL of 2 M glycine in PBS was added to each dish and rocked at room temperature for 5 min. Cells were then washed twice with ice-cold PBS containing protease inhibitors (Roche), and the second wash was allowed to sit at 4 °C for 5 min. Cells were then scraped in 4 mL of PBS plus protease inhibitors and spun at 4 °C for 5 min at 500*g*, and the pellet was flash-frozen and stored at −80 °C. Cell pellets were allowed to thaw on ice, incubated with cell lysis buffer (20 mM Tris, pH 8.0, 85 mM KCl, 0.5% NP-40, and protease inhibitors) for 10 min, and spun at 5,000*g* for 5 min at 4 °C. The nuclear pellet was treated the same way a second time. The nuclear pellet was resuspended in nuclear lysis buffer (10 mM Tris, pH 7.8, 1% NP-40, 0.5% sodium deoxycholate, 0.1% SDS, and protease inhibitors) for sonication. Sonication was performed on a Branson microtip sonicator for 8 min with 0.3 s on, 1.7 s off at 4 °C. All sonicated chromatin was assessed for fragment size: we required more than 60% of the size distribution to be between 150 and 700 bp long, with the average around 350 bp (Agilent TapeStation). Off-target sites were predicted via the CRISPR design tool at http://crispr.mit.edu. For ChIP, streptavidin (SA) conjugated to magnetic beads (Thermo), M2 anti-Flag (Sigma), or anti-V5 (MBL Life Sciences) was conjugated to a 50:50 mix of Protein A:Protein G Dynabeads (Invitrogen) and incubated with sheared chromatin at 4 °C overnight. qPCR was performed with either Roche 2× Sybr mix (biological duplicates or triplicates, measurement triplicates) on a Lightcycler (Agilent) or via digital droplet PCR (biological quadruplicates, measurement singlicate) (Bio-Rad). All primers were synthesized by Integrative DNA Technologies.

Western blotting analysis. sgRNA-293T-Caspex cells were labeled as described above. 40 µg of whole-cell lysate was separated by SDS–PAGE, transferred to nitrocellulose, and blotted against Flag (Sigma; 1:2,000 dilution) or biotin (Li-Cor IRdye 800 CW SA and IRdye 680RD anti-mouse IgG, both at 1:10,000 dilution).

Enrichment of biotinylated proteins for proteomic analysis. Eight 15-cm^2 plates of each sgRNA-293T-Caspex line (~3×10^8 cells per line), or a no-guide cell line as a negative control, were used for proteomic experiments. Labeled whole-cell pellets were lysed with RIPA buffer (50 mM Tris, pH 8.0, 150 mM NaCl, 1% NP-40, 0.5% sodium deoxycholate, 0.1% sodium dodecyl sulfate) with protease inhibitors (Roche) and probe-sonicated to shear genomic DNA. Whole-cell lysates were clarified by centrifugation at 14,000*g* for 30 min at 4 °C, and protein concentration was determined by Bradford assay. 500 µL of SA magnetic bead slurry (Thermo) was used for each sgRNA line (between 60 and 90 mg of protein per state). Lysates of equal protein concentrations were incubated with SA for 120 min at room temperature, and then were washed twice with cold lysis buffer, once with cold 1 M KCl, once with cold 100 mM Na_2CO_3, and twice with cold 2 M urea in 50 mM ammonium bicarbonate. Beads were resuspended in 50 mM ammonium bicarbonate with 300 ng of trypsin and digested at 37 °C overnight. 10 mM TCEP and 10 mM iodoacetamide were added after digestion and allowed to incubate at room temperature for 30 min in the dark. The inherent sensitivity limits of current mass spectrometers and the unavoidable sample losses at each sample handling step required that a large amount of input material be used per guide. Fortunately, these input requirements are readily attainable with many cell culture systems, although meeting them might prove more challenging with recalcitrant or limited-passaging cells.

Isobaric labeling and liquid chromatography–tandem mass spectrometry. On-bead digests were desalted via StageTip[33] and labeled with tandem mass tags (TMTs; Thermo)[34] via an on-column protocol. For on-column TMT labeling, StageTips were packed with one-punch C18 mesh (Empore), washed with 50 µL of methanol, 50 µL of 50% acetonitrile (ACN)/0.1% formic acid (FA), and equilibrated with 75 µL of 0.1% FA twice. The digest was loaded by spinning at 3,500*g* until the entire digest passed through. The bound peptides were washed twice with 75 µL of 0.1% FA. One microliter of TMT reagent in 100% ACN was added to 100 µL of freshly made HEPES, pH 8, and passed over the C18 resin at 2,000*g* for 2 min. The HEPES and residual TMTs were washed away with 75 µL of 0.1% FA twice, and peptides were eluted with 50 µL of 50% ACN/0.1% FA and then with 50% ACN/20 mM ammonium hydroxide, pH 10. Peptide concentrations were estimated by absorbance reading at 280 nm and mixed at equal ratios. Mixed TMT-labeled peptides were step-fractionated by basic reverse-phase fractionation on a sulfonated divinylbenzene (SDB-RPS; Empore) packed StageTip into six fractions (5%, 10%, 15%, 20%, 30%, and 55% ACN in 20 mM ammonium hydroxide, pH 10). Each fraction was dried via vacuum centrifugation and resuspended in 0.1% FA for subsequent LC-MS/MS analysis.

Chromatography was carried out on a Proxeon EASY-nLC at a flow rate of 200 nl/min. Peptides were separated at 50 °C on a 75-µm-inner-diameter PicoFit (New Objective) column packed with 1.9-µm AQ-C18 material (Dr. Maisch) to 20 cm in length over an 84-min effective gradient. Mass spectrometry was performed on a Thermo Scientific Q Exactive Plus (*hTERT* data) or a Lumos (*MYC* data) mass spectrometer. After a precursor scan from 300 to 2,000 m/z at 70,000 resolution, the top 12 most intense multiply charged precursors were selected for higher-energy collisional dissociation at a resolution of 35,000. Data were searched with Spectrum Mill (Agilent) using the UniProt human database, in which the CASPEX protein was amended. A fixed modification of carbamidomethylation of cysteine and variable modifications of N-terminal protein acetylation, oxidation of methionine, and TMT-10-plex labels were searched. The enzyme specificity was set to trypsin, and a maximum of three missed cleavages was used for searching. The maximum precursor-ion charge state was set to 6. The precursor mass tolerance and MS/MS tolerance were set to 20 p.p.m. The peptide and protein false discovery rates were set to 0.01.

Data analysis. All non-human proteins and human proteins identified with only one peptide were excluded from downstream analyses. Human keratins were included in all analyses but were not included in the figures. The moderated *t*-test (http://software.broadinstitute.org/cancer/software/genepattern/) was used to determine which proteins were statistically enriched in the sgRNA-293T-Caspex lines compared with levels in the no-sgRNA control[10,18,35]. After correction for multiple comparisons (Benjamini–Hochberg procedure), any proteins with an adjusted P value of less than 0.05 were considered statistically enriched. To determine the relationship between protein abundance and the ability to determine statistical enrichment[36,37], we used the contribution of the 131 channel to the summed MS1 intensity of peptides for a particular protein.

Pathway analysis was done with the Quack algorithm incorporated into Genets (http://apps.broadinstitute.org/genets) to test for enrichment of canonical pathways in MSigDB[38,39]. Proteins identified as significantly enriched (adjusted $P < 0.05$) by GLoPro were input into Genets and queried against MSigDB. Enriched pathways (false discovery rate < 0.05) were investigated manually for specific proteins for follow-up.

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

Data availability. The original mass spectra can be downloaded from MassIVE (http://massive.ucsd.edu) under accesssion PXD009187. The data are also directly accessible at ftp://massive.ucsd.edu/MSV000082154.

References

20. Wang, G. et al. *Nat. Med.* **20**, 616–623 (2014).
21. Doench, J. G. et al. *Nat. Biotechnol.* **32**, 1262–1267 (2014).
22. Doench, J. G. et al. *Nat. Biotechnol.* **34**, 184–191 (2016).
23. Yang, X. et al. *Nat. Methods* **8**, 659–661 (2011).
24. Su, J. M. et al. *Hepatology* **46**, 402–413 (2007).
25. Xu, M., Katzenellenbogen, R. A., Grandori, C. & Galloway, D. A. *Virology* **446**, 17–24 (2013).
26. Hoffmeyer, K. et al. *Science* **336**, 1549–1554 (2012).
27. Jaitner, S. et al. *Cell Cycle* **11**, 3331–3338 (2012).
28. Zhang, Y., Toh, L., Lau, P. & Wang, X. *J. Biol. Chem.* **287**, 32494–32511 (2012).
29. Bell, R. J. A. et al. *Science* **348**, 1036–1039 (2015).
30. Glasspool, R. M., Burns, S., Hoare, S. F., Svensson, C. & Keith, W. N. *Neoplasia.* **7**, 614–622 (2005).
31. Xu, D. et al. *Oncogene* **19**, 5123–5133 (2000).
32. Kanaya, T. et al. *Clin. Cancer Res.* **6**, 1239–1247 (2000).
33. Rappsilber, J., Mann, M. & Ishihama, Y. *Nat. Protoc.* **2**, 1896–1906 (2007).
34. Thompson, A. et al. *Anal. Chem.* **75**, 1895–1904 (2003).
35. Perez-Pinera, P. et al. *Nat. Methods* **10**, 973–976 (2013).
36. Mellacheruvu, D. et al. *Nat. Methods* **10**, 730–736 (2013).
37. Keilhauer, E. C., Hein, M. Y. & Mann, M. *Mol. Cell. Proteomics* **14**, 120–135 (2015).
38. Li, T. et al. *Nat. Methods* **14**, 61–64 (2017).
39. Subramanian, A. et al. *Proc. Natl. Acad. Sci. USA* **102**, 15545–15550 (2005).

nature research

Corresponding author(s): Samuel Myers and Steve Carr

Reporting Summary

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

Statistical parameters

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

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

Our web collection on statistics for biologists may be useful.

Software and code

Policy information about availability of computer code

Data collection	LCMS data was obtained on Thermo software; WB on Licor; Sybr on Roche Lightcycler; ddPCR on Biorad. Standard statistics were used for all analyses and should be easily repeated, including Pearson correlations, moderated T-tests, and p-value adjustments for multiple hypothesis testing.
Data analysis	Spectrum Mill was used for peptide ID and quantification. Moderated T-tests, p-value calculation and multiple hypothesis correcting, and Pearson correlation calculations were all performed with either Gene Pattern, readily available R scripts, or Genets (https://apps.broadinstitute.org/genets). All of which are publicly available.

For manuscripts utilizing custom algorithms or software that are central to the research but not yet described in published literature, software must be made available to editors/reviewers upon request. We strongly encourage code deposition in a community repository (e.g. GitHub). See the Nature Research guidelines for submitting code & software for further information.

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

The raw data can be accessed at ftp://massive.ucsd.edu/MSV000082154

Field-specific reporting

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

☒ Life sciences ☐ Behavioural & social sciences

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

Life sciences

Study design

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

Sample size	All sample/replicate sizes were determined from current literature standards. The LC-MS data, as an exception, used 4-5 replicates instead of 2 replicates, as they are considered quasi-replicates (querying the same locus via slightly different conditions but treated as replicates).
Data exclusions	No data used or discussed in the manuscript was excluded from this analysis. The line referenced about using the transposase on hard to transfect cells was removed as it is not necessary for using the method as described.
Replication	All attempts for replication were successful. Replicate recall is shown were applicable in the figures.
Randomization	The TMT channels, the only possible place for randomization, was not performed. There is little need for TMT randomization with high sample similarity.
Blinding	ChIP data was blinded from the experimentalist preparing the sample to the one performing the analyses where applicable.

Materials & experimental systems

Policy information about availability of materials

n/a	Involved in the study
☐	☒ Unique materials
☐	☒ Antibodies
☐	☒ Eukaryotic cell lines
☒	☐ Research animals
☒	☐ Human research participants

Unique materials

Obtaining unique materials	No restrictions on the new plasmid, which is available via Addgene.

Antibodies

Antibodies used	All antibodies are commercially available. V5, MBL Life Science; FLAG, M2 Sigma; Streptavidin IRdye 800-CW, Li-cor.
Validation	Validation was performed using alternative epitope tags, and proper controls based on the assay requirements.

Eukaryotic cell lines

Policy information about cell lines

Cell line source(s)	HEK293T cells were aqcuired from ATCC and modified in house. The modified lines will be available upon request.
Authentication	Mophology and proteomic confirmation authenticated the cells during experiments described in the manuscript.
Mycoplasma contamination	All lines were test for myco every 3 months or so. All were negative.

Commonly misidentified lines (See ICLAC register)	No commonly misidentified lines were used in this study.

Method-specific reporting

n/a	Involved in the study
☒	☐ ChIP-seq
☒	☐ Flow cytometry
☒	☐ Magnetic resonance imaging