

MIT Open Access Articles

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

The MIT Faculty has made this article openly available. **Please share** how this access benefits you. Your story matters.

Citation: Meyers, Samuel A. et al. "Discovery of proteins associated with a predefined genomic locus via dCas9–APEX-mediated proximity labeling." *Nature Methods* 15, 6 (May 2018): 437–439
© 2018 The Author(s)

As Published: <http://dx.doi.org/10.1038/s41592-018-0007-1>

Publisher: Springer Science and Business Media LLC

Persistent URL: <https://hdl.handle.net/1721.1/126374>

Version: Author's final manuscript: final author's manuscript post peer review, without publisher's formatting or copy editing

Terms of use: Creative Commons Attribution-Noncommercial-Share Alike





Published in final edited form as:

Nat Methods. 2018 June ; 15(6): 437–439. doi:10.1038/s41592-018-0007-1.

Discovery of proteins associated with a predefined genomic locus in living cells via dCAS9-APEX-mediated proximity labeling

Samuel A Myers^{1,*}, Jason Wright¹, Ryan Peckner¹, Brian T Kalish^{2,3}, Feng Zhang¹, and Steven A Carr^{1,*}

¹Broad Institute of MIT and Harvard, Cambridge, Massachusetts 02142

²Department of Neurobiology, Harvard Medical School, Boston, Massachusetts 02115

³Division of Newborn Medicine, Department of Medicine, Boston Children's Hospital, Boston, Massachusetts 02115

Abstract

Regulation of gene expression is primarily controlled by changes in the proteins that occupy their regulatory elements. Chromatin immunoprecipitation can confirm a protein's occupancy at a genomic locus but requires specific, high-quality, IP-competent antibodies against nominated proteins, limiting its utility. Here, we combine, genome targeting, proximity labeling, and quantitative proteomics to develop genomic locus proteomics, a method able to identify proteins associated a specific genomic locus in native cellular contexts.

Transcriptional regulation is a highly-coordinated process largely controlled by changes in protein occupancy at regulatory elements of the modulated genes. Chromatin immunoprecipitation (ChIP) has been invaluable for our understanding of transcriptional regulation and chromatin structure at both the individual locus and genome-wide levels^{1–3}. However, because ChIP requires the use of antibodies, its utility can often be limited by the presupposition of a suspected protein's occupancy, and lack of highly specific and high affinity reagents. Previously developed “inverse ChIP” methods have had limited utility in mammalian systems due to one of several drawbacks including loss of cellular and/or chromatin context, extensive engineering and locus disruption, reliance on repetitive DNA sequences, or chemical crosslinking^{4–10} which often requires extensive optimization for mass spectrometry-based applications¹¹. We sought to develop a method to identify proteins

Users may view, print, copy, and download text and data-mine the content in such documents, for the purposes of academic research, subject always to the full Conditions of use: http://www.nature.com/authors/editorial_policies/license.html#terms

*Correspondence: sammyers@broadinstitute.org and scarr@broad.mit.edu.

Conflict of Interest Statement

The authors declare no conflicting financial interests. A patent application has been filed by The Broad Institute relating to this work.

Author Contributions

S.A.M. and J.W. conceived the idea. S.A.M., J.W. and S.A.C. conceived the experimental set-up and designed the research. S.A.M. created the constructs and stable lines, performed the labeling, Western blots, enrichments, mass spectrometry, and the ChIP-qPCRs for the *c-MYC* lines. J.W. performed the ChIP-qPCRs 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. All authors interpreted the data. S.A.M., J.W. and S.A.C. wrote the manuscript.

associated with a specific, non-repetitive genomic locus in the native cellular context without the need for crosslinking or genomic alterations. Here, we utilized recent advances in sequence-specific DNA targeting and affinity labeling in cells to develop genomic locus proteomics (GLoPro) to characterize proteins associated with a particular genomic locus.

We fused the catalytically-dead RNA-guided nuclease Cas9 (dCas9)^{12, 13} to the engineered peroxidase APEX2¹⁴ to biotinylate proteins at a defined genomic loci for subsequent enrichment and identification by liquid chromatography-mass spectrometry (LC-MS/MS) (Figure 1A and Supplemental Figure 1). dCAS9 was employed due to the reprogrammable nature of the single guide RNAs (sgRNA)¹⁵. 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^{16, 17}, labeling nearby proteins with biotin derivatives^{14, 17-19}. APEX2 was chosen for its small labeling radius and short reaction time²⁰⁻²². The dCas9-APEX2 (*Caspex*) gene was cloned in-frame with the self-cleaving T2A peptide and *Gfp* under the control of a tetracycline response element into a puromycin-selectable piggybac plasmid²³ (Supplemental Figure 2). The inducibility of Caspex expression provides temporal control to minimize the time CASPEX occupies the targeted locus, and the accumulation of excess CASPEX which leads to higher background biotinylation, common with proximity labeling²⁴⁻²⁸.

To test whether the CASPEX protein correctly localized to the genomic site of interest, we created a single-colony HEK293T line, stably integrated with the *Caspex* plasmid (293T-Caspex), and expressed a sgRNA targeting 92 base pairs (bp) 3' of the transcription start site (TSS) (denoted as T092) of the *TERT* gene. We focused on the *TERT* promoter (*hTERT*) as *TERT* expression is a hallmark of cancer and recurrent promoter mutations in *hTERT* have been shown to re-activate *TERT* expression²⁹. Biotinylation in T092 sgRNA-expressing 293T-Caspex cells was induced, followed by anti-FLAG or anti-biotin ChIP- quantitative PCR (qPCR) with probes tiling *hTERT*. ChIP-qPCR showed proper localization of CASPEX with the peak of the anti-FLAG signal overlapping with the destination of the sgRNA. The anti-biotin ChIP-qPCR signal showed a similar trend of enrichment, indicating that CASPEX biotinylates proteins within approximately 400 base pairs on either side of its target locus. As expected, no enrichment was observed around T092 for the non-spatially constrained no sgRNA control (Figure 1B).

We then tested four additional sgRNAs tiling *hTERT* (Figure 1C). After labeling in stable sgRNA-expressing 293T-Caspex lines, Western blot analysis showed that biotinylation was CASPEX-dependent but no difference in biotin patterns were observed between sgRNA lines (Supplemental Figure 3). ChIP-qPCR against FLAG or biotin showed all constructs correctly targeted and labeled the region of interest, where the peak of enrichment resided at the sgRNA site (Supplemental Figure 4). To assess off-target binding of CASPEX, we performed anti-FLAG ChIP-qPCR probing the top predicted off-target site for each respective sgRNA. No two off-target sites were within 5 Mb from each other (Supplemental Table 1). Each on-target site was occupied by CASPEX 3- to 40-fold more than 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 (Supplemental Figure 5). These data demonstrate that CASPEX targeting can be

accurately reprogrammed by substitution of the sgRNAs and that proximal protein biotinylation is CASPEX mediated.

To test whether CASPEX could be used to identify proteins associated with *hTERT*, we enriched biotinylated proteins from *hTERT*-targeted 293T-Caspex lines, followed by analysis by quantitative LC-MS/MS. Biotinylation was initiated in the five individual *hTERT*-targeting 293T-Caspex lines that tiled the genomic locus of interest, along with the no-guide control 293T-Caspex line. 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”^{30, 31}. Tiling may also circumvent the loss of native protein identification if CAS9 binding precludes their occupancy, though evidence suggests minor affects³⁰. Whole cell lysates from each line were incubated with streptavidin-coated beads, stringently washed, and subjected to on-bead trypsin digestion. Digests of the enriched proteins were labeled with isobaric tandem mass tags (TMT)³² for relative quantitation, multiplexed, and analyzed by LC-MS/MS (Supplemental Figure 1). We used a ratiometric approach for each individual sgRNA 293T-Caspex line compared to the no-guide control line²⁸. The no-guide control^{13, 30, 33} was chosen over a random genomic site to improve chances of finding “housekeeping” chromatin proteins that may reside in many places in the genome. Enrichment from the four overlapping *hTERT* Caspex lines showed high correlation of protein enrichment (Figure 1D). The T959 Caspex line, which lies 2 labeling radii from its closest neighbor, showed decreased correlation of protein enrichment. We performed a moderated one-sample T-test by treating the four overlapping sgRNA lines as replicates, using the non-spatially constrained no sgRNA 293T-CasPEX line as the background biotinylation and enrichment control. 387 of the 3,199 proteins identified with at least two peptides were significantly enriched (adj. p value < 0.05, FE > 0) at *hTERT* over the no sgRNA control, including five proteins known to occupy *hTERT* in various cell types (MAZ;^{34, 35} CTNNB1;^{36–38} ETV3;³⁹ CTBP1;⁴⁰ and TP53 [adj. p = 0.056]^{41, 42}) (Figure 1E, Supplemental Table 2). Histones and subunits of RNA polymerase II, while detected, were not significantly enriched, likely due to the inverse relationship between a protein’s abundance and the ability to determine that it is enriched over background (Supplemental Figure 6)^{43, 44}. These results demonstrate GLoPro is able to enrich and identify proteins at a particular genomic locus from the native cellular context.

In addition to several known proteins, we identified a number of candidate proteins associated with *hTERT*. To corroborate the GLoPro results, we performed ChIP-qPCR for candidates spanning the GLoPro enrichment range (Figure 1F). Many of these proteins do not have ChIP grade antibodies so we turned to V5-tagged ORF expression in unmodified HEK293T cells⁴⁵. We individually transfected 16 V5-tagged candidate ORFs, four V5-tagged ORFs for proteins not significantly enriched at *hTERT*, and three proteins that were not detected as negative controls. Comparing anti-V5 ChIP-qPCR signals from each individual ORF to their 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) (Figure 1G). 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 described to bind *hTERT*, CTBP1 and

MAZ^{34, 35, 40}, were found in a regime of high ChIP enrichment and low GLoPro enrichment, suggesting ChIP-qPCR provides orthogonal information to GLoPro for protein occupancy at a genomic locus. These data show that GLoPro can identify known and novel proteins that can be corroborated by ChIP-qPCR that associate with *hTERT*.

To explore the generalizability of GLoPro, we created 293T-Caspex cells expressing individual sgRNAs tiling the *c-MYC* promoter (Figure 2A). ChIP-qPCR against CASPEX verified the proper localization of each *c-MYC* 293T-Caspex line (Supplemental Figure 7) and off-target binding analysis showed a cumulative 32-fold enrichment at the *c-MYC* promoter over any predicted off-target site (Supplemental Figure 8). GLoPro analysis of the *c-MYC* promoter identified 66 proteins as significantly enriched (adj. p val < 0.05) compared to the no-guide control (Figure 2B, Supplemental Table 3). We applied a machine learning algorithm to identify association of GLoPro-enriched proteins with canonical pathways from the Molecular Signature Database^{46, 47}, <http://apps.broadinstitute.org/genets>). We identified 21 statistically enriched networks (adj. p val. < 0.01), including the “MYC_Active_Pathway”, a gene set of validated targets responsible for activating *c-MYC* transcription⁴⁷ (Figure 2C). ChIP-qPCR confirmed the presence of pathway components at the *c-MYC* promoter, including HUWE1, RUVBL1, and ENO1 for *MYC_active_pathway*; RBMX for *mRNA_splicing_pathway*; and MAPK14 for the *Lymph_angiogenesis_pathway* (Figure 2D). Taken together, these results illustrate that GLoPro enriches and identifies proteins associated in multiple pathways that are known to activate *c-MYC* expression, while directly implicating specific proteins potentially involved in regulating *c-MYC* transcription through association with its promoter.

GLoPro relies on the localization of the affinity labeling enzyme APEX2 directed by the catalytically-dead CRISPR/Cas9 system to biotinylate proteins in close proximity to a specific site in the genome. Other than the expression of *Caspex* and its associated sgRNA, no genome engineering or cell disruption is required to capture a snapshot of proteins associated with the locus of interest. This advantage, in combination with 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. Beyond circumventing the need for antibodies for discovery, LC-MS/MS analysis using isobaric peptide labeling allows for sample multiplexing, enabling multiple sgRNA lines and/or replicates to be measured 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 the genomic region 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. While GLoPro in this initial work only identifies association with a locus and not locus-specificity (a protein bound at the queried site likely binds elsewhere) or functional relevance, we expect that analyzing promoters or enhancer elements during relevant perturbations may provide novel functional insights into transcriptional regulation. In addition, we envision 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.

In summary, we describe genomic locus proteomics, a novel approach to identify proteins associated with a particular genomic loci in live cells without modifying the site of interest. This method provides an orthogonal and highly complementary approach to ChIP for the unbiased discovery of proteins that may regulate gene expression and chromatin structure. We applied GLoPro to identify proteins associated with the *hTERT* and *c-MYC* promoters. Both well-established and previously unreported interactors of the respective promoter regions identified by GLoPro were corroborated using ChIP-qPCR, demonstrating that this method enables the discovery of proteins and pathways that potentially regulate a gene of interest.

Methods

Plasmid construction

The *Caspex* construct (dox inducible dCas9-APEX2-T2-GFP) was created 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 (Church lab) via ligation independent cloning (InFusion, Clontech). Guide sequences were selected and cloned as previously described^{49, 50}. Guides were designed irrespective of sense/anti-sense strands but spacing 100–200 bp between guides and on-target scores were used. All V5 ORF constructs were purchased through the Broad Genetics Perturbation Platform and were expressed from the pLX-TRC_317 backbone. V5 ORFs were only selected for validation if the construct was available, had protein homology >99%, and an in-frame V5 tag. All constructs were individually transfected into unmodified HEK293T cells for anti-V5 ChIP experiments and 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% fetal bovine serum, glutamine and non-essential amino acids (Gibco). All constructs were prepared using Zippy Maxi Prep kits (Zymo Research) and transfected with Lipofectamine 2000 (Thermo). After *Caspex* transfection, puromycin was added to a final concentration of 4 ug/ml and selected for two weeks. Single colonies were picked, expanded and tested for doxycycline inducibility of the *Caspex* construct monitored by GFP detection and by anti-FLAG Western blotting. The HEK293T cell line with the best inducibility (now referred to as 293-Caspex cells) was expanded and used for all subsequent experiments. For stable sgRNA expression, single sgRNA constructs were transfected into 293-Caspex cells and were selected for stable incorporation by hygromycin treatment at 200 ug/ml for two weeks. CASPEX binding was tested using ChIP followed by digital droplet PCR (ddPCR) or Sybr qPCR described below.

APEX-mediated labeling

Prior to labeling, doxycycline dissolved in 70% ethanol was added to cell culture media to a final concentration of either 500 ng/mL for 18–24 hours (*hTERT*) or 12 hours at 1 ug/mL (*c-MYC*) for proteomic experiments. For off-target CASPEX binding analysis, cells were treated with 1 ug/mL dox for 20 hours. Biotin tyramide phenol (Iris Biotech) in DMSO, stock

concentration 500 mM, was added directly to cell culture media, which was swirled until the precipitate dissolved, to a final concentration of 500 μ M. After 30 minutes 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 seconds of very gently swirling the media was decanted as quickly as possible and the cells were washed three times with 15 ml 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 500g for 3 minutes, 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 single cell suspension and fresh formaldehyde was added to a final concentration of 1% and incubated at 37°C for 10 minutes, being inverted several times every two minutes. Formaldehyde was quenched with 5% glycine for 37°C for 5 minutes and the samples were aliquoted into 3e6 cell aliquots, spun down and flash frozen in 0.5 mL Axygen tubes. Chromatin was sheared using a QSonica Q800R2 Sonicator at an amplitude of 50 for 30 seconds on/30 off, for 7.5 minutes, until 60% of fragments were between 150 and 700 bp, with an average size of \sim 350 bp. Lysis buffer was comprised of 1% SDS, 10 mM EDTA and Tris HCl, pH 8.0. For off-target CASPEX analysis, one approximately 80% confluent 15 cm² plate was washed once with 20 mls ice cold PBS and then incubated with room temperature PBS with 1% formaldehyde (freshly prepared from 16% stock, Thermo) at room temperature for ten minutes with gentle rocking. After crosslinking, 1.5 mL 2M glycine in PBS was added to each dish and rocked at room temperature for 5 minutes. Cells were then washed twice with ice cold PBS containing protease inhibitors (Roche), where the second wash was allowed to sit at 4°C for 5 minutes. Cells were then scraped in 4 mL PBS plus protease inhibitors, spun at 4°C for 5 minutes at 500g, 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, plus protease inhibitors) for ten minutes, and spun at 5,000g for 5 minutes 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% NP40, 0.5% sodium deoxycholate, 0.1% SDS, plus protease inhibitors) for sonication. Sonication was performed on a Branson microtip sonicator for 8 minutes 0.3 seconds on, 1.7 off at 4°C. All sonicated chromatin was assessed for fragment size, requiring more than 60% of the size distribution to be between 150 to 700 bp long, with the average around 350 bp (Agilent TapeStation). Off-target sites were predicted via crispr.mit.edu. For CHIP, streptavidin (SA)-conjugated to magnetic beads (Thermo), M2 anti-FLAG antibody (Sigma) or anti-V5 antibodies (MBL Life Sciences) was conjugated to a 50:50 mix of Protein A: Protein G Dynabeads (Invitrogen) was incubated with sheared chromatin at 4°C overnight. qPCR was performed with either Roche 2 \times Sybr mix (biological duplicates or triplicates, measurement triplicates) on a Lightcycler (Agilent) or via digital droplet PCR (biological quadruplicates, measurement singlicate) (BioRad). All primers were synthesized by Integrative DNA Technologies (IDT).

Western blot analysis

sgRNA-293-Caspex cells were labeled as described above. 40 ug of whole cell lysate was separated by SDS-PAGE, transferred to nitrocellulose, and blotted against FLAG (Sigma, 1:2000 dilution) or biotin (Li-Cor IRdye 800 CW Streptavidin and IRdye 680RD anti-Mouse IgG, both at 1:10,000 dilution).

Enrichment of biotinylated proteins for proteomic analysis

Eight 15 cm² plates of each sgRNA-293-Caspex line (~3e8 cells per line), or no-guide as a negative control, were used for proteomic experiments. Labeled whole cell pellets were lysed with RIPA (50 mM TRIS pH 8.0, 150 mM NaCl, 1% NP-40 and 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 14,000g for 30 minutes at 4°C and protein concentration was determined by Bradford. 500 uL SA magnetic bead slurry (Thermo) was used for each sgRNA line (between 60–90 mgs of protein/state). Lysates of equal protein concentrations were incubated with SA for 120 minutes at room temperature, washed twice with cold lysis buffer, once with cold 1M KCl, once with cold 100 mM Na₃CO₂, and twice with cold 2 M urea in 50 mM ammonium bicarbonate (ABC). Beads were resuspended in 50 mM ABC and 300 ng trypsin and digested at 37°C overnight. 10 mM TCEP and 10 mM iodoacetamide was added after digestion and allowed to incubate at room temperature for 30 minutes in the dark. The inherent sensitivity limits of current mass spectrometers and the unavoidable sample losses at each sample handling step, require that a large amount of input material be used per guide. Fortunately, these input requirements are readily attainable with many cell culture systems but may prove more challenging with recalcitrant or limited passaging cells.

Isobaric labeling and liquid chromatography tandem mass spectrometry

On-bead digests were desalted via Stage tip⁵¹ and labeled with TMT (Thermo) using an on-column protocol. For on-column TMT labeling, Stage tips were packed with one punch C18 mesh (Empore), washed with 50 uL methanol, 50 uL 50% acetonitrile (ACN)/0.1% formic acid (FA), and equilibrated with 75 uL 0.1% FA twice. The digest was loaded by spinning at 3,500g until the entire digest passed through. The bound peptides were washed twice with 75 uL 0.1% FA. One uL of TMT reagent in 100% ACN was added to 100 uL freshly made HEPES, pH 8, and passed over the C18 resin at 2,000g for 2 minutes. The HEPES and residual TMT was washed away with 75 uL 0.1% FA twice and peptides were eluted with 50 uL 50% ACN/0.1% FA followed by a second elution with 50% ACN/20 mM ammonium hydroxide, pH 10. Peptide concentrations were estimated using an absorbance reading at 280 nm and mixed at equal ratios. Mixed TMT labeled peptides were step fractionated by basic reverse phase on a sulfonated divinylbenzene (SDB-RPS, Empore) packed Stage tip into 6 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% formic acid for subsequent LC-MS/MS analysis.

Chromatography was performed using a Proxeon EASY-nLC at a flow rate of 200 nl/min. Peptides were separated at 50°C using a 75 micron i.d. PicoFit (New Objective) column packed with 1.9 um 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 (*c-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 HCD 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-10plex 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 ppm. 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 removed from the figures. The moderated T-test (<http://software.broadinstitute.org/cancer/software/genepattern/>) was used to determine proteins statistically enriched in the sgRNA-293-Caspex lines compared to the no sgRNA control. After correcting for multiple comparisons (Benjamini-Hochberg procedure), any proteins with an adjusted p-value of less than 0.05 were considered statistically enriched.

Pathway analysis was performed using the Quack algorithm incorporated into Genets (<http://apps.broadinstitute.org/genets>) to test for enrichment of canonical pathways in the Molecular Signature Database (MSigDB). Proteins identified as significantly enriched (adj. p-val. < 0.05) by GLoPro were input into Genets and were queried against MSigDB. Pathways enriched (FDR < 0.05) were investigated manually for specific proteins for follow-up.

Data availability

The original mass spectra may be downloaded from MassIVE (<http://massive.ucsd.edu>), PXD009187. The data are directly accessible via <ftp://massive.ucsd.edu/MSV000082154>.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

We are grateful to Seth Shipman and George Church for the tetracycline-inducible/puro^R plasmid, and to Albert Edge for the *HUWE1* plasmids. We would like to thank Alice Ting, John Doench, Namrata Udeshi, Aviv Regev, Pratiksha Thakore, B. Hamilton, and Egle Kvedaraitė for useful discussion; Elizabeth M. Perez, John Ray, and Robbyn Issner for technical assistance; Malvina Papanastasiou, Eric Kuhn, Jennifer Abelin for critical assessment of the manuscript. This work is funded by NCI CPTAC PCC 1U24CA210986-01 and a Broad Institute SPARC grant to S.A.M. B.T.K. is funded through the Pediatric Scientist Development Program and March of Dimes Birth Defects Foundation. The *Caspex* plasmid is available through Addgene (plasmid # 97421).

References

1. Bernstein BE, et al. A bivalent chromatin structure marks key developmental genes in embryonic stem cells. *Cell*. 2006; 125:315–326. [PubMed: 16630819]

2. Kagey MH, et al. Mediator and cohesin connect gene expression and chromatin architecture. *Nature*. 2010; 467:430–435. [PubMed: 20720539]
3. Liber D, et al. Epigenetic priming of a pre-B cell-specific enhancer through binding of Sox2 and Foxd3 at the ESC stage. *Cell Stem Cell*. 2010; 7:114–126. [PubMed: 20621055]
4. Mittler G, Butter F, Mann M. A SILAC-based DNA protein interaction screen that identifies candidate binding proteins to functional DNA elements. *Genome Res*. 2009; 19:284–293. [PubMed: 19015324]
5. Dejardin J, Kingston RE. Purification of proteins associated with specific genomic Loci. *Cell*. 2009; 136:175–186. [PubMed: 19135898]
6. Fujita T, Fujii H. Efficient isolation of specific genomic regions and identification of associated proteins by engineered DNA-binding molecule-mediated chromatin immunoprecipitation (enChIP) using CRISPR. *Biochem Biophys Res Commun*. 2013; 439:132–136. [PubMed: 23942116]
7. Pourfarzad F, et al. Locus-specific proteomics by TChP: targeted chromatin purification. *Cell Rep*. 2013; 4:589–600. [PubMed: 23911284]
8. Schmidtman E, Anton T, Rombaut P, Herzog F, Leonhardt H. Determination of local chromatin composition by CasID. *Nucleus*. 2016; 7:476–484. [PubMed: 27676121]
9. Liu X, et al. In Situ Capture of Chromatin Interactions by Biotinylated dCas9. *Cell*. 2017; 170:1028–1043 e1019. [PubMed: 28841410]
10. Tsui C, et al. dCas9-targeted locus-specific protein isolation method identifies histone gene regulators. *Proc Natl Acad Sci U S A*. 2018
11. Ly T, et al. Proteomic analysis of cell cycle progression in asynchronous cultures, including mitotic subphases, using PRIMMUS. *Elife*. 2017; 6
12. Jinek M, et al. A programmable dual-RNA-guided DNA endonuclease in adaptive bacterial immunity. *Science*. 2012; 337:816–821. [PubMed: 22745249]
13. Qi LS, et al. Repurposing CRISPR as an RNA-guided platform for sequence-specific control of gene expression. *Cell*. 2013; 152:1173–1183. [PubMed: 23452860]
14. Lam SS, et al. Directed evolution of APEX2 for electron microscopy and proximity labeling. *Nat Methods*. 2015; 12:51–54. [PubMed: 25419960]
15. Cong L, et al. Multiplex Genome Engineering Using CRISPR/Cas Systems. *Science*. 2013; 339:819–823. [PubMed: 23287718]
16. Schonhuber W, Fuchs B, Juretschko S, Amann R. Improved sensitivity of whole-cell hybridization by the combination of horseradish peroxidase-labeled oligonucleotides and tyramide signal amplification. *Appl Environ Microbiol*. 1997; 63:3268–3273. [PubMed: 9251215]
17. Rhee HW, et al. Proteomic Mapping of Mitochondria in Living Cells via Spatially Restricted Enzymatic Tagging. *Science*. 2013; 339:1328–1331. [PubMed: 23371551]
18. Hung V, et al. Proteomic mapping of cytosol-facing outer mitochondrial and ER membranes in living human cells by proximity biotinylation. *Elife*. 2017; 6
19. Udeshi ND, et al. Antibodies to biotin enable large-scale detection of biotinylation sites on proteins. *Nat Methods*. 2017
20. Roux KJ, Kim DI, Raida M, Burke B. A promiscuous biotin ligase fusion protein identifies proximal and interacting proteins in mammalian cells. *J Cell Biol*. 2012; 196:801–810. [PubMed: 22412018]
21. Paek J, et al. Multidimensional Tracking of GPCR Signaling via Peroxidase-Catalyzed Proximity Labeling. *Cell*. 2017; 169:338–349. [PubMed: 28388415]
22. Lobingier BT, et al. An Approach to Spatiotemporally Resolve Protein Interaction Networks in Living Cells. *Cell*. 2017; 169:350–360. [PubMed: 28388416]
23. Wang G, et al. Modeling the mitochondrial cardiomyopathy of Barth syndrome with induced pluripotent stem cell and heart-on-chip technologies. *Nat Med*. 2014; 20:616–623. [PubMed: 24813252]
24. Lobingier BT, et al. An Approach to Spatiotemporally Resolve Protein Interaction Networks in Living Cells. *Cell*. 2017; 169:350–360 e312. [PubMed: 28388416]
25. Paek J, et al. Multidimensional Tracking of GPCR Signaling via Peroxidase-Catalyzed Proximity Labeling. *Cell*. 2017; 169:338–349 e311. [PubMed: 28388415]

26. Teo G, et al. SAINTexpress: improvements and additional features in Significance Analysis of INteractome software. *J Proteomics*. 2014; 100:37–43. [PubMed: 24513533]
27. Lambert JP, Tucholska M, Go C, Knight JD, Gingras AC. Proximity biotinylation and affinity purification are complementary approaches for the interactome mapping of chromatin-associated protein complexes. *J Proteomics*. 2015; 118:81–94. [PubMed: 25281560]
28. Hung V, et al. Proteomic Mapping of the Human Mitochondrial Intermembrane Space in Live Cells via Ratiometric APEX Tagging. *Mol Cell*. 2014; 55:332–341. [PubMed: 25002142]
29. Huang FW, et al. Highly Recurrent TERT Promoter Mutations in Human Melanoma. *Science*. 2013; 339:957–959. [PubMed: 23348506]
30. Thakore PI, et al. Highly specific epigenome editing by CRISPR-Cas9 repressors for silencing of distal regulatory elements. *Nature Methods*. 2015; 12:1143+. [PubMed: 26501517]
31. Wu XB, et al. Genome-wide binding of the CRISPR endonuclease Cas9 in mammalian cells. *Nat Biotechnol*. 2014; 32:670+. [PubMed: 24752079]
32. Thompson A, et al. Tandem mass tags: A novel quantification strategy for comparative analysis of complex protein mixtures by MS/MS. *Anal Chem*. 2003; 75:1895–1904. [PubMed: 12713048]
33. Perez-Pinera P, et al. RNA-guided gene activation by CRISPR-Cas9-based transcription factors. *Nat Methods*. 2013; 10:973–976. [PubMed: 23892895]
34. Su JM, et al. X protein of hepatitis B virus functions as a transcriptional corepressor on the human telomerase promoter. *Hepatology*. 2007; 46:402–413. [PubMed: 17559154]
35. Xu M, Katzenellenbogen RA, Grandori C, Galloway DA. An unbiased in vivo screen reveals multiple transcription factors that control HPV E6-regulated hTERT in keratinocytes. *Virology*. 2013; 446:17–24. [PubMed: 24074563]
36. Hoffmeyer K, et al. Wnt/beta-Catenin Signaling Regulates Telomerase in Stem Cells and Cancer Cells. *Science*. 2012; 336:1549–1554. [PubMed: 22723415]
37. Jaitner S, et al. Human telomerase reverse transcriptase (hTERT) is a target gene of beta-catenin in human colorectal tumors. *Cell Cycle*. 2012; 11:3331–3338. [PubMed: 22894902]
38. Zhang Y, Toh L, Lau P, Wang XY. Human Telomerase Reverse Transcriptase (hTERT) Is a Novel Target of the Wnt/beta-Catenin Pathway in Human Cancer. *J Biol Chem*. 2012; 287:32494–32511. [PubMed: 22854964]
39. Bell RJA, et al. The transcription factor GABP selectively binds and activates the mutant TERT promoter in cancer. *Science*. 2015; 348:1036–1039. [PubMed: 25977370]
40. Glasspool RM, Burns S, Hoare SF, Svensson C, Keith WN. The hTERT and hTERC telomerase gene promoters are activated by the second exon of the adenoviral protein, E1A, identifying the transcriptional corepressor CtBP as a potential repressor of both genes. *Neoplasia*. 2005; 7:614–622. [PubMed: 16036112]
41. Xu DW, et al. Downregulation of telomerase reverse transcriptase mRNA expression by wild type p53 in human tumor cells. *Oncogene*. 2000; 19:5123–5133. [PubMed: 11064449]
42. Kanaya T, et al. Adenoviral expression of p53 represses telomerase activity through down-regulation of human telomerase reverse transcriptase transcription. *Clin Cancer Res*. 2000; 6:1239–1247. [PubMed: 10778946]
43. Mellacheruvu D, et al. The CRAPome: a contaminant repository for affinity purification-mass spectrometry data. *Nat Methods*. 2013; 10:730–736. [PubMed: 23921808]
44. Keilhauer EC, Hein MY, Mann M. Accurate protein complex retrieval by affinity enrichment mass spectrometry (AE-MS) rather than affinity purification mass spectrometry (AP-MS). *Mol Cell Proteomics*. 2015; 14:120–135. [PubMed: 25363814]
45. Yang XP, et al. A public genome-scale lentiviral expression library of human ORFs. *Nature Methods*. 2011; 8:659–U680. [PubMed: 21706014]
46. Li TB, et al. A scored human protein-protein interaction network to catalyze genomic interpretation. *Nature Methods*. 2017; 14:61–64. [PubMed: 27892958]
47. Subramanian A, et al. Gene set enrichment analysis: A knowledge-based approach for interpreting genome-wide expression profiles. *P Natl Acad Sci USA*. 2005; 102:15545–15550.
48. Livak KJ, Schmittgen TD. Analysis of relative gene expression data using real-time quantitative PCR and the 2(-Delta Delta C(T)) Method. *Methods*. 2001; 25:402–408. [PubMed: 11846609]

49. Doench JG, et al. Rational design of highly active sgRNAs for CRISPR-Cas9-mediated gene inactivation. *Nat Biotechnol.* 2014; 32:1262–1267. [PubMed: 25184501]
50. Doench JG, et al. Optimized sgRNA design to maximize activity and minimize off-target effects of CRISPR-Cas9. *Nat Biotechnol.* 2016; 34:184–191. [PubMed: 26780180]
51. Rappsilber J, Mann M, Ishihama Y. Protocol for micro-purification, enrichment, pre-fractionation and storage of peptides for proteomics using StageTips. *Nature Protocols.* 2007; 2:1896–1906. [PubMed: 17703201]

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

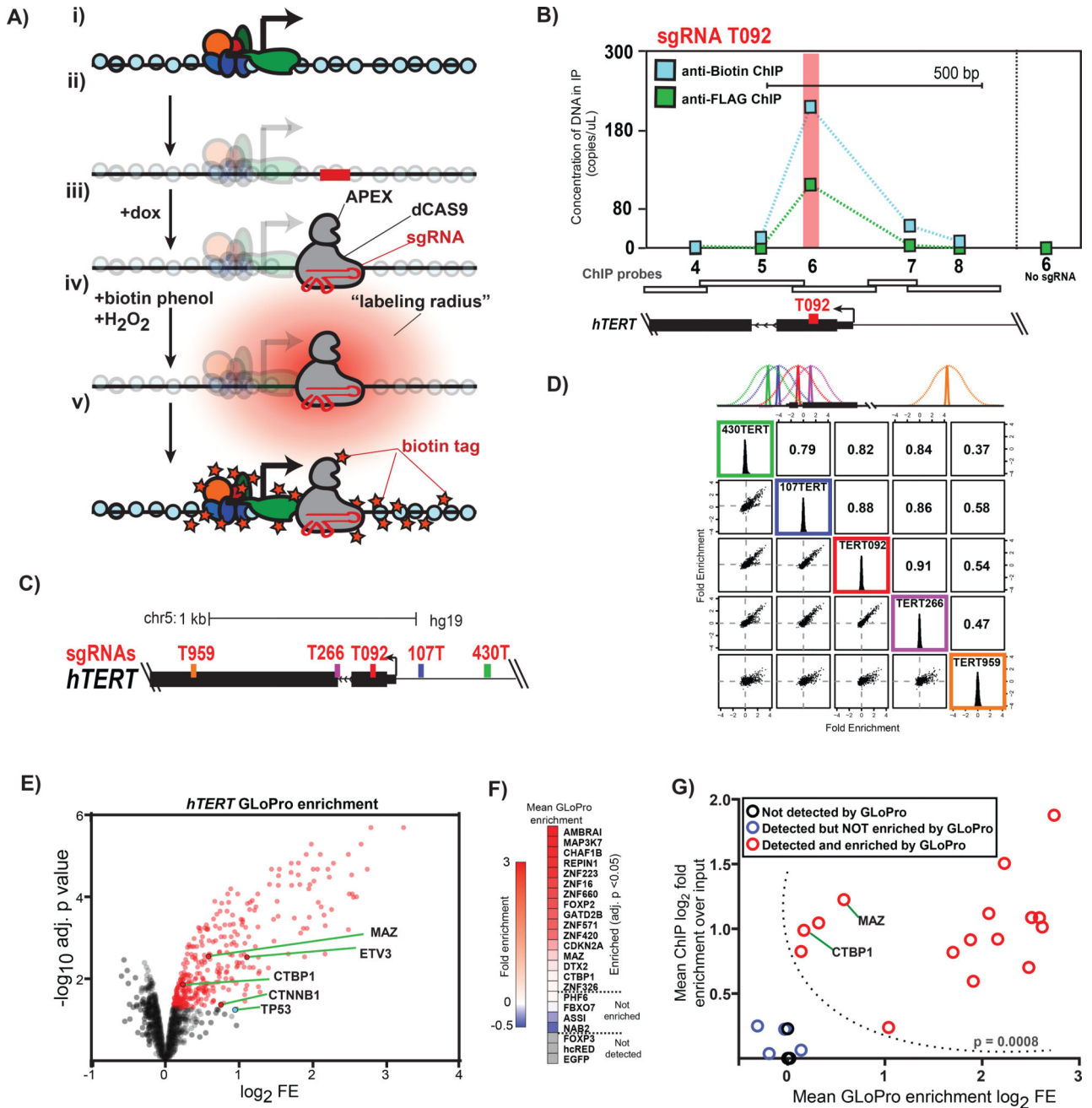


Figure 1. Genomic Locus Proteomics of *hTERT* **A)** Illustration of CASPEX targeting and affinity labeling reaction. **i)** A genomic locus of interest is identified. **ii)** A targeting sequence for the sgRNA is designed (red bar). **iii)** CASPEX expression is induced with doxycycline and, after association with sgRNA, binds region of interest. **iv)** After biotin-phenol incubation, H₂O₂ induces the CASPEX-mediated labeling of proximal proteins, where the “labeling radius” of the reactive biotin-phenol is represented by the red cloud. **v)** Proteins proximal to CASPEX are labeled with biotin (orange stars) for subsequent enrichment. **B)** ChIP-qPCR against biotin (blue boxes) and FLAG (green boxes) in 293T-CasPEX cells expressing either no

sgRNA (far right) or T092 sgRNA. White boxes indicate ChIP probes for regions amplified and detected by qPCR, all of which are in Supplemental Figure 4. *hTERT* is below to show the gene structure with respect to the sgRNA target (red box). **C**) UCSC Genome Browser representation of *hTERT* (hg19). sgRNAs (colored bars) are shown to scale relative to the transcription start site (black arrow). **D**) Multi-scatter plots for log₂ fold enrichment values of proteins quantified, and the corresponding Pearson correlation coefficients between all pair-wise *hTERT*-293T-Caspex cells comparisons using the no sgRNA control line as the denominator (n = 5 independent sgRNA lines). **E**) Volcano plot of proteins quantified across the four overlapping *hTERT*-293T-Caspex cell lines (n = 4 independent sgRNA lines) compared to the no sgRNA control. Data points representing proteins enriched with an adjusted p-value of less than 0.05, FE > 0, are labeled in red. Proteins known to associate with *hTERT* and identified as enriched by GLoPro are highlighted. TP53, a known *hTERT* binder, had an adj. p val. = 0.058 and is highlighted blue. **F**) Mean GLoPro enrichment values for V5-tagged ORFs selected for ChIP-qPCR corroboration. Red indicates the protein was enriched at *hTERT*, blue that the protein was detected in the analysis but not statistically enriched. Grey proteins were not detected. **G**) Correlation between ChIP-qPCR mean log₂ 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*. Black, open circles indicate that the protein was not identified by GLoPro. Blue, open circles indicate the protein was identified but was not statistically enriched. Red open circles indicate proteins that are enriched according to the GLoPro analysis. Previously described *hTERT* binders are labeled. Dotted line separates ChIP-qPCR data tested for statistical significance via the Mann-Whitney test, and the p-value is shown.

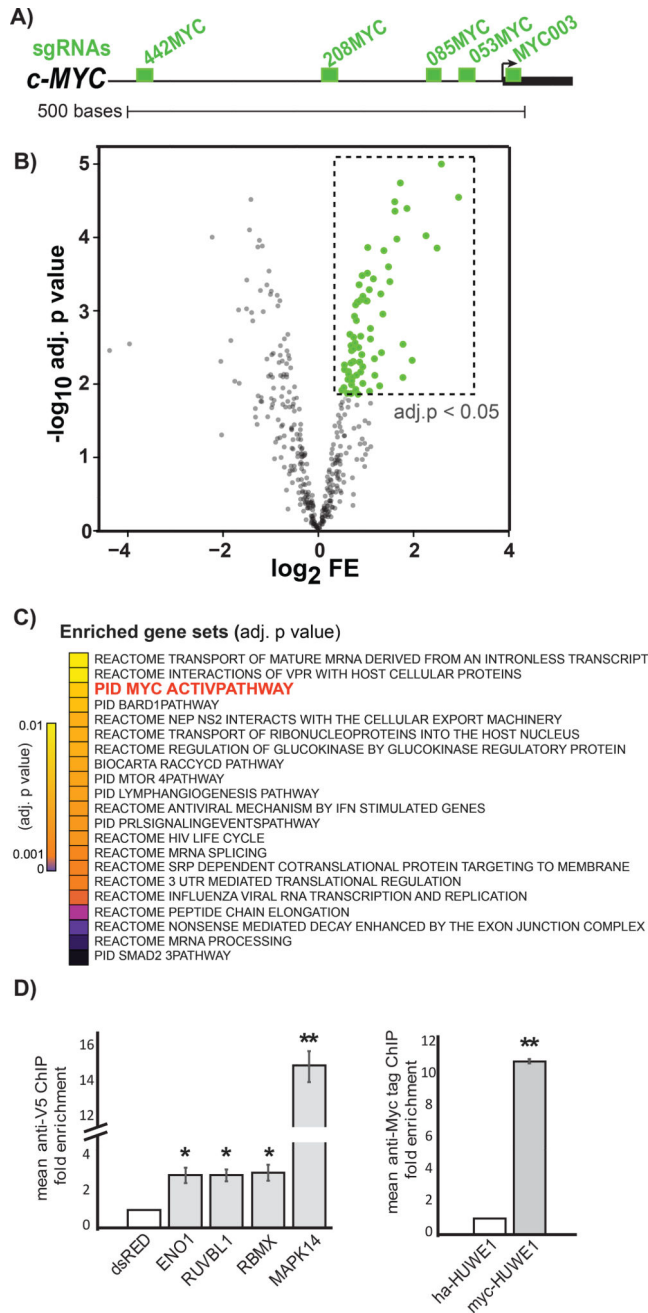


Figure 2. Genomic locus proteomic analysis of *c-MYC* promoter **A)** UCSC Genome Browser representation (hg19) of the *c-MYC* promoter and the location of sgRNA sites relative to the TSS. **B)** Volcano plot of proteins quantified across the five *c-MYC*-Caspex cell lines compared to the no sgRNA control Caspex line (n = 5 independent sgRNA lines). Data points representing positively enriched proteins with an adjusted p-value of less than 0.05 are labeled green. **C)** Significantly enriched gene sets from proteins identified to associate with the *c-MYC* promoter by GLoPro. Only gene sets with an adjusted p-value of less than 0.01 are shown. MYC_ACTIVE_PATHWAY is highlighted in red and discussed in the text

D) CHIP-qPCR of candidate proteins identified by GLoPro at the *c-MYC* promoter. V5 tagged dsRED served as the negative control for V5-tagged proteins ENO1, RBMX, RUVBL1 and MAPK14, whereas HA-tagged HUWE1 was used for MYC-tagged HUWE1. * indicates T-test p-value < 0.05, ** p < 0.01. Mean and standard error is plotted (transfection duplicates, measurement triplicates).

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript