HHS Public Access

Author manuscript

Science. Author manuscript; available in PMC 2017 March 30.

Published in final edited form as:

Science. 2016 September 30; 353(6307): 1545–1549. doi:10.1126/science.aaf7613.

High-resolution interrogation of functional elements in the noncoding genome

Neville E. Sanjana[1,2,*,†,‡], **Jason Wright**[1,2,*], **Kaijie Zheng**[1,2], **Ophir Shalem**[1,2], **Pierre Fontanillas**[1], **Julia Joung**[1,2], **Christine Cheng**[1,3], **Aviv Regev**[1,3], and **Feng Zhang**[1,2,†]

[1]Broad Institute of MIT and Harvard, 7 Cambridge Center, Cambridge, MA 02142, USA

[2]McGovern Institute for Brain Research, Department of Brain and Cognitive Sciences, Department of Biological Engineering, Massachusetts Institute of Technology, Cambridge, MA 02139, USA

[3]Howard Hughes Medical Institute, David H. Koch Institute of Integrative Cancer Biology, Department of Biology, Massachusetts Institute of Technology, Cambridge, MA 02142, USA

Abstract

The noncoding genome affects gene regulation and disease, yet we lack tools for rapid identification and manipulation of noncoding elements. We develop a CRISPR screen employing ~18,000 sgRNAs targeting >700 kb surrounding the genes *NF1*, *NF2*, and *CUL3*, which are involved in BRAF inhibitor resistance in melanoma. We find that noncoding locations that modulate drug resistance also harbor predictive hallmarks of noncoding function. With a subset of regions at the *CUL3* locus, we demonstrate that engineered mutations alter transcription factor occupancy and long-range and local epigenetic environments, implicating these sites in gene regulation and chemotherapeutic resistance. Though our expansion of the potential of pooled CRISPR screens we provide tools for genomic discovery and for elucidating biologically relevant mechanisms of gene regulation.

Pooled CRISPR mutagenesis identifies functional elements in the noncoding genome.

More than 98% of the human genome does not code for proteins, however, unlike the coding genome there exists no overarching framework to translate noncoding genomic sequence into functional elements (1, 2). Evidence from genome-wide association studies (GWAS) suggests many noncoding regions are critical for human health (3, 4). The significance of these associations, however, has been difficult to assess, in part because we lack the tools to determine which variants alter functional elements. Molecular hallmarks, such as epigenetic

[†]Correspondence should be addressed to: nsanjana@nygenome.org (N.S.) and zhang@broadinstitute.org (F.Z.).
[*]These authors contributed equally to this work.
[‡]Present address: New York Genome Center, New York, NY 10013, Department of Biology, New York University, New York, NY 10012.

Supplementary Materials

Materials and Methods
Supplementary Text
Figures S1 to S13
Tables S1 to S7
References (*26–42*)

state, chromatin accessibility, transcription factor binding, and evolutionary conservation, correlate with putative functional elements in the noncoding genome and can predict regulatory function (2, 5). However, these predictions largely bypass regions lacking hallmarks, and it is difficult to ascertain what hallmarks play a correlative or truly causal role in function or phenotype (6, 7). Efforts to determine causality have employed preselected DNA fragments with expression serving as a proxy for function (8) but these methods lack the local chromatin context and broader regulatory interactions. Thus there is a need for systematic approaches to sift through noncoding variants and determine if and how they affect phenotypes within a native biological context.

For this purpose, we designed a high-throughput method using pooled CRISPR (Clustered regularly-interspaced short palindromic repeat)-Cas9 single guide RNA (sgRNA) libraries to screen noncoding genomic loci to identify functional regions related to phenotype and gene regulation. Previous applications of CRISPR screens within the noncoding genome have focused on specific functional elements (e.g. miRNAs, transcription factor binding sites) or required fluorescent reporters (9–12). Here, we comprehensively assay a total of 715 kb of sequence surrounding three different genes by performing unbiased mutagenesis to identify functional elements relevant to cancer drug resistance.

Vemurafenib inhibits BRAF proteins carrying the V600E mutation, found in 50–70% of melanomas (13). Resistance to vemurafenib arises within months in almost all melanoma patients (14) and surviving tumor cells display increased malignancy that rapidly leads to lethality (15). A genome-scale CRISPR screen found that loss-of-function mutations in *NF1*, *NF2*, and *CUL3* result in vemurafenib resistance (16). To explore if mutations in the noncoding regions around these three genes could similarly impact drug resistance, we designed three sgRNA libraries tiling across 100 kb regions 5′ and 3′ of each gene's major isoforms (Fig. 1A). For each library, we synthesized the sgRNAs as a pool (6,682 for *NF1*, 6,934 for *NF2*, and 4,699 for *CUL3;* 18,315 sgRNAs total) and cloned them into a lentiviral vector (fig. S1). We transduced A375 human melanoma cells, which carry the BRAF mutation, with the sgRNA libraries at a low multiplicity of infection and cultured them in 2uM vemurafenib or control (DMSO) for 14 days. Using deep sequencing, we counted the representation of sgRNAs in the library in both conditions (Fig. 1B–D) and identified vemurafenib-enriched sgRNAs as those enriched >4 standard deviations from the control distribution (fig. S2).

Overall, most sgRNAs were depleted after treatment with vemurafenib, which is expected since vemurafenib targets the oncogene addiction that drives A375 growth (Fig. 1E). However, in all three libraries, we found a small group of sgRNAs that were enriched after vemurafenib treatment ($\log_2$ ratio of Vemu/Control > 0), with the *CUL3* library having the largest percentage of enriched sgRNAs. As we also included a small number of sgRNAs targeting the coding region of each gene, most sgRNAs targeting coding regions (70 – 80%) were enriched, as expected (fig. S3A). However, amongst the sgRNAs targeting noncoding regions, approximately 4-fold more sgRNAs were enriched in the *CUL3* library than in the *NF1* or *NF2* libraries (7.2% in *CUL3*, 1.7% in *NF1*, and 2.1% in *NF2*), suggesting the presence of more gene regulatory elements in the noncoding regions flanking the gene (fig. S3A). To determine if this increase in putative gene regulatory elements in the 200 kb region

surrounding *CUL3* is also reflected in human gene expression and genotyping data, we queried the Genotype-Tissue Expression (GTEx) database (7,051 tissue samples from 449 donors). Indeed, we found that *CUL3* had the largest number of *cis*-expression quantitative trait loci (eQTL) (n = 161 eQTLs, mean effect size = −0.21), and the region targeted by the sgRNA library overlaps with a large number of these eQTLs (fig. S3B) (17). We thus chose to focus our downstream analysis and validation efforts on *CUL3*.

We visualized the enriched sgRNAs in a genome browser-style view (Figs. 1F, S4A, B). We found that a higher percentage of sgRNAs targeting gene-proximal elements were enriched compared to other noncoding regions (Fig. 1G) and greater enrichment for sgRNAs targeting noncoding elements on the 5′ side of the gene than for those on the 3′ side (fig. S4C).

To test if regions targeted by enriched sgRNAs from the screen physically interact with the *CUL3* promoter via chromatin looping (18), we created three independent chromosome conformation capture (3C) libraries (Fig. 2A) (19). We quantified the interaction frequency for each site across the ~200 kb region (43) and found that regions on the 5′ side of *CUL3* tend to interact more strongly with the promoter. Regions with higher 3C interaction contain, on average, more vemurafenib-enriched sgRNAs (Fig. 2B).

Since chromatin accessibility can identify regulatory regions (20, 21), we performed Assay of Transposase-Accessible Chromatin sequencing (ATAC-seq) in A375 melanoma, MCF7 breast adenocarcinoma and U87 glioblastoma cells (Fig. 2C). Overall, we found higher sgRNA enrichment near A375-specific ATAC peaks than near those from other cell types, which was replicated with DNAseI hypersensitivity data (Figs. 2D–E, S5). These regions suggest the presence of cell type-specific enhancers (22, 23). Even though the accessible peaks overlap with enriched sgRNA sites, the chromatin accessibility data by itself only predicts a small fraction of the total number of enriched sgRNA sites (table S1).

Since evolutionary conservation varies widely across the noncoding genome, we sought to test whether regions exhibiting higher levels of conservation harbor more enriched sgRNAs. We examined phastCons conservation scores among primates, placental mammals, and vertebrates over the *CUL3* locus (Fig. 2F) (24). Overall, enriched sgRNAs are ~1.8-fold more likely to be found near peaks of primate conservation and are ~1.7-fold less likely to be found near conservation peaks among mammals and vertebrates (Figs. 2G, S5). In contrast, the genomic sites of sgRNAs targeting coding regions of *CUL3* do not demonstrate differential conservation (phastCons probability ~ 0.95 in primates, mammals and vertebrates). This observation supports recent findings that enhancers evolve rapidly in a lineage- or species-specific manner and conserved enhancers between mammals tend to be rare (25).

To confirm that mutations in these specific noncoding regions were mediated by *CUL3* and lead to altered drug resistance, we transduced cells with individual sgRNAs having at least one other enriched sgRNA within 500 bp (Fig. 3A). We validated that the sgRNA created mutations at the intended target sites (fig. S6) and found that 24 out of the 25 sgRNAs resulted in decreased *CUL3* expression relative to non-targeting sgRNAs (Fig. 3B). As expected, there is a negative correlation between *CUL3* gene expression and vemurafenib

resistance ($r = -0.54$, $p = 0.005$) (Fig. 3C) and increased vemurafenib resistance can be reversed by restoring *CUL3* expression (fig. S7).

Next, we surveyed changes in post-translational histone modifications at each target site (fig. S8A) (43). With target sites near the promoter, we found a 56% average decrease of H3K4me3 after editing ($p = 7\times10^{-4}$, $n = 9$ sites) (Fig. 3D), consistent with the reduced gene expression. At distal sites, we found a 41% average decrease in H3K27ac ($p = 0.02$, $n = 7$ sites) after editing and no significant change in H3K4me2 ($p = 0.82$, $n = 7$ sites) (Fig. 3D), although a subset of H3K4me2 levels decreased after editing (fig. S8B). We also found that mutagenesis of a ~22 kb distal histone acetyl-transferase (p300) binding site that has a strong 3C promoter interaction results in a 75% decline of promoter H3K27ac and a 50% decrease in *CUL3* expression (fig. S9, 43).

By examining regions targeted by enriched sgRNAs, we found individual loci containing the canonical transcription factor binding motifs for Yin Yang 1 (YY1), Zinc Finger Protein 263 (ZNF263), CCCTC-binding factor (CTCF) and activation protein 1 (AP-1) complex which were disrupted after editing (Figs. 4A–D, S10). We found that mutations within these binding sites abrogate transcription factor recruitment leading to loss of *CUL3* expression (Figs. 4E–H). For example, specific sgRNAs that target near a YY1 ChIP peak (Fig. 4B) disrupt the YY1 motif (fig. S11) and vemurafenib treatment selects for mutations that are more deleterious to the binding site (fig. S12). Although both sgRNAs targeting near the site decrease YY1 binding, the sgRNA whose cut site overlaps the motif more efficiently disrupts YY1 binding (67% vs. 26%) (Fig. 4E). In addition, mutagenesis by either sgRNA significantly decreases *CUL3* expression. Similarly, 2 sgRNAs in the first intron of *CUL3* spaced 30 bp apart overlap a ZNF263 ChIP-seq peak (Fig. 4B). Both result in a significant decrease in ZNF263 occupancy and in *CUL3* expression (Fig. 4F).

Although we observed a bias in the presence of regulatory elements 5′ of the transcription start site, we did find several enriched sgRNAs downstream of *CUL3* (Fig. 4C, D) (43). One sgRNA cuts inside the core motif of CTCF (Fig. 4C). After editing, CTCF occupancy is decreased by 45% with a concurrent 30% decrease in *CUL3* expression (Fig. 4G). For AP-1, a heterodimer of FOS and JUN, editing at either of two nearby sites decreases FOS and JUN binding compared with control cells and decreases *CUL3* expression by ~25% (Fig. 4H). Overall, as in the pooled screen, transcription factor binding sites located on the 3′ side exhibit weaker effects on gene expression than those located on the 5′ side.

Thus we show how a Cas9-mediated systematic dissection of noncoding loci can identify functional elements involved in gene regulation and cancer drug resistance. In combination with other genome-wide assays, we demonstrate high-throughput identification of regions where changes in chromatin context and transcription factor binding are causally linked to loss of gene expression and a disease-relevant phenotype. This approach is generalizable, and we anticipate that the extension of pooled CRISPR screens into the noncoding genome will provide further insights and methods for unbiased interrogation of the genome.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

We would like to thank R. Macrae, D. Scott and the entire Zhang laboratory. N.S. is supported by the NIH through a NHGRI Pathway to Independence Award (R00-HG008171) and a postdoctoral fellowship from the Simons Center for the Social Brain at MIT. J.W. is supported by the NIH through a Ruth L. Kirschstein National Research Service Award (F32-DK096822). O.S. is a Klarman Fellow of the Broad-Israel Partnership. F.Z. is a New York Stem Cell Foundation-Robertson Investigator and is supported by the NIH through NIMH (5DP1-MH100706 and 1R01-MH110049), NSF, the New York Stem Cell, Simons, Paul G. Allen Family, and Vallee Foundations; and James and Patricia Poitras, Robert Metcalfe, and David Cheng. A.R. is on the SAB for Syros Pharmaceuticals and Thermo Fisher and a consultant for Driver group. N.S., J.W., O.S., and F.Z. are listed as inventors on a related patent application. Plasmids are available from Addgene subject to a material transfer agreement. Deep sequencing data are available at Sequence Read Archive under BioProject accession number PRJNA324504 and software tools are available on GitHub (https://github.com/nsanjana/BashRegion).

References

1. Visel A, Rubin EM, Pennacchio LA. Genomic views of distant-acting enhancers. Nature. 2009; 461:199–205. [PubMed: 19741700]
2. ENCODE Project Consortium. An integrated encyclopedia of DNA elements in the human genome. Nature. 2012; 489:57–74. [PubMed: 22955616]
3. Hindorff LA, et al. Potential etiologic and functional implications of genome-wide association loci for human diseases and traits. Proc Natl Acad Sci U S A. 2009; 106:9362–9367. [PubMed: 19474294]
4. Maurano MT, et al. Systematic localization of common disease-associated variation in regulatory DNA. Science. 2012; 337:1190–1195. [PubMed: 22955828]
5. Roadmap Epigenomics Consortium. et al. Integrative analysis of 111 reference human epigenomes. Nature. 2015; 518:317–330. [PubMed: 25693563]
6. Kwasnieski JC, Fiore C, Chaudhari HG, Cohen BA. High-throughput functional testing of ENCODE segmentation predictions. Genome Res. 2014; 24:1595–1602. [PubMed: 25035418]
7. Mundade R, Ozer HG, Wei H, Prabhu L, Lu T. Role of ChIP-seq in the discovery of transcription factor binding sites, differential gene regulation mechanism, epigenetic marks and beyond. Cell Cycle. 2014; 13:2847–2852. [PubMed: 25486472]
8. Melnikov A, et al. Systematic dissection and optimization of inducible enhancers in human cells using a massively parallel reporter assay. Nat Biotechnol. 2012; 30:271–277. [PubMed: 22371084]
9. Chen S, et al. Genome-wide CRISPR screen in a mouse model of tumor growth and metastasis. Cell. 2015; 160:1246–1260. [PubMed: 25748654]
10. Canver MC, et al. BCL11A enhancer dissection by Cas9-mediated in situ saturating mutagenesis. Nature. 2015; 527:192–197. [PubMed: 26375006]
11. Korkmaz G, et al. Functional genetic screens for enhancer elements in the human genome using CRISPR-Cas9. Nat Biotechnol. 2016; 34:192–198. [PubMed: 26751173]
12. Rajagopal N, et al. High-throughput mapping of regulatory DNA. Nat Biotechnol. 2016; 34:167–174. [PubMed: 26807528]
13. Cancer Genome Atlas Network. Genomic Classification of Cutaneous Melanoma. Cell. 2015; 161:1681–1696. [PubMed: 26091043]
14. Sosman JA, et al. Survival in BRAF V600-mutant advanced melanoma treated with vemurafenib. N Engl J Med. 2012; 366:707–714. [PubMed: 22356324]
15. Zubrilov I, et al. Vemurafenib resistance selects for highly malignant brain and lung-metastasizing melanoma cells. Cancer Lett. 2015; 361:86–96. [PubMed: 25725450]
16. Shalem O, et al. Genome-scale CRISPR-Cas9 knockout screening in human cells. Science. 2014; 343:84–87. [PubMed: 24336571]

17. GTEx Consortium. Human genomics. The Genotype-Tissue Expression (GTEx) pilot analysis: multitissue gene regulation in humans. Science. 2015; 348:648–660. [PubMed: 25954001]
18. Lieberman-Aiden E, et al. Comprehensive mapping of long-range interactions reveals folding principles of the human genome. Science. 2009; 326:289–293. [PubMed: 19815776]
19. Dekker J, Rippe K, Dekker M, Kleckner N. Capturing chromosome conformation. Science. 2002; 295:1306–1311. [PubMed: 11847345]
20. Crawford GE, et al. Identifying gene regulatory elements by genome-wide recovery of DNase hypersensitive sites. Proc Natl Acad Sci U S A. 2004; 101:992–997. [PubMed: 14732688]
21. Buenrostro JD, Giresi PG, Zaba LC, Chang HY, Greenleaf WJ. Transposition of native chromatin for fast and sensitive epigenomic profiling of open chromatin, DNA-binding proteins and nucleosome position. Nat Methods. 2013; 10:1213–1218. [PubMed: 24097267]
22. Heintzman ND, et al. Distinct and predictive chromatin signatures of transcriptional promoters and enhancers in the human genome. Nat Genet. 2007; 39:311–318. [PubMed: 17277777]
23. Sheffield NC, et al. Patterns of regulatory activity across diverse human cell types predict tissue identity, transcription factor binding, and long-range interactions. Genome Res. 2013; 23:777–788. [PubMed: 23482648]
24. Felsenstein J, Churchill GA. A Hidden Markov Model approach to variation among sites in rate of evolution. Mol Biol Evol. 1996; 13:93–104. [PubMed: 8583911]
25. Villar D, et al. Enhancer evolution across 20 mammalian species. Cell. 2015; 160:554–566. [PubMed: 25635462]
26. Sanjana NE, Shalem O, Zhang F. Improved vectors and genome-wide libraries for CRISPR screening. Nat Methods. 2014; 11:783–784. [PubMed: 25075903]
27. Langmead B, Trapnell C, Pop M, Salzberg SL. Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. Genome Biol. 2009; 10:R25. [PubMed: 19261174]
28. Miele A, Gheldof N, Tabuchi TM, Dostie J, Dekker J. Mapping chromatin interactions by chromosome conformation capture. Curr Protoc Mol Biol. 2006 **Chapter 21**, Unit 21.11.
29. Van der Auwera GA, et al. From FastQ data to high confidence variant calls: the Genome Analysis Toolkit best practices pipeline. Curr Protoc Bioinformatics. 2013; 11:11.10.1–11.10.33.
30. Quinlan AR, Hall IM. BEDTools: a flexible suite of utilities for comparing genomic features. Bioinformatics. 2010; 26:841–842. [PubMed: 20110278]
31. Joung J, et al. Protocol: Genome-scale CRISPR-Cas9 Knockout and Transcriptional Activation Screening. 2016; doi: 10.1101/059626
32. Mathelier A, et al. JASPAR 2016: a major expansion and update of the open-access database of transcription factor binding profiles. Nucleic Acids Res. 2016; 44:D110–5. [PubMed: 26531826]
33. Tan G, Lenhard B. TFBSTools: an R/Bioconductor package for transcription factor binding site analysis. Bioinformatics. 2016; doi: 10.1093/bioinformatics/btw024
34. Wasserman WW, Sandelin A. Applied bioinformatics for the identification of regulatory elements. Nat Rev Genet. 2004; 5:276–287. [PubMed: 15131651]
35. Bhandaru M, et al. A combination of p300 and Braf expression in the diagnosis and prognosis of melanoma. BMC Cancer. 2014; 14:398. [PubMed: 24893747]
36. Lai F, et al. Cotargeting histone deacetylases and oncogenic BRAF synergistically kills human melanoma cells by necrosis independently of RIPK1 and RIPK3. Cell Death Dis. 2013; 4:e655. [PubMed: 23744355]
37. Bushmeyer S, Park K, Atchison ML. Characterization of functional domains within the multifunctional transcription factor, YY1. J Biol Chem. 1995; 270:30213–30220. [PubMed: 8530432]
38. Zhang Q, Stovall DB, Inoue K, Sui G. The oncogenic role of Yin Yang 1. Crit Rev Oncog. 2011; 16:163–197. [PubMed: 22248053]
39. Katainen R, et al. CTCF/cohesin-binding sites are frequently mutated in cancer. Nat Genet. 2015; 47:818–821. [PubMed: 26053496]
40. Sanborn AL, et al. Chromatin extrusion explains key features of loop and domain formation in wild-type and engineered genomes. Proc Natl Acad Sci U S A. 2015; 112:E6456–65. [PubMed: 26499245]

41. Ding X, et al. Epigenetic activation of AP1 promotes squamous cell carcinoma metastasis. Sci Signal. 2013; 6 ra28.1–13–S0–15.

42. Alipanahi B, Delong A, Weirauch MT, Frey BJ. Predicting the sequence specificities of DNA- and RNA-binding proteins by deep learning. Nat Biotechnol. 2015; 33:831–838. [PubMed: 26213851]

43. Materials, methods and supplementary text are available as supplementary materials on *Science* Online.

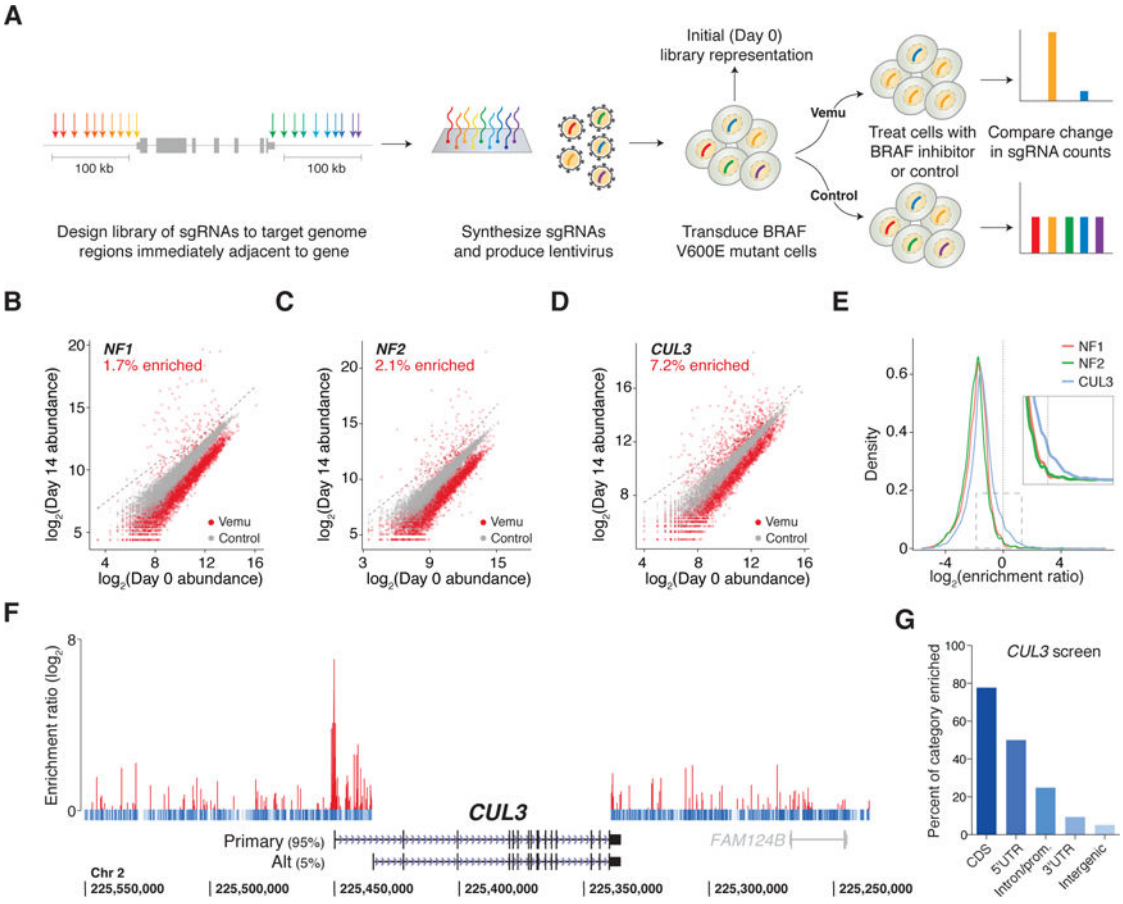

Figure 1. CRISPR mutagenesis of noncoding regions flanking three genes involved in BRAF inhibitor resistance

A) Design of sgRNA libraries targeting 100 kb 5` and 100 kb 3` of a gene. The sgRNAs are array synthesized and cloned into a lentiviral vector. BRAF mutant cells are transduced with the pooled lentivirus and treated with vemurafenib (Vemu) or DMSO (Control). A deep sequencing readout identifies sgRNAs enriched after treatment with vemurafenib.

Scatterplot of normalized read counts (average of the 2 infection replicates) for B: *NF1*, C: *NF2*, D: *CUL3* sgRNAs at Day 0 (*x* axis) and Day 14 (*y* axis) Read counts from control (gray) and vemurafenib-treated cells (red) are shown relative to 4 standard deviations of the control cell distribution (dotted line) with the percentage of enriched sgRNAs in vemurafenib (>4 s.d.).

E) Distribution of $\log_2$ ratio of the normalized read count for each sgRNA in vemurafenib to its normalized read count in control (minimum of the 2 infection replicates).

F) All *CUL3* sgRNAs plotted by *hg19* coordinates and the percent expression of the two most highly expressed *CUL3* isoforms (Primary, Alt.). For vemurafenib-enriched sgRNAs, the $\log_2$ enrichment over control (minimum value of 2 replicate screens) is plotted (red); non-enriched sgRNAs are indicated in blue.

G) Percent of enriched sgRNAs by genomic category.

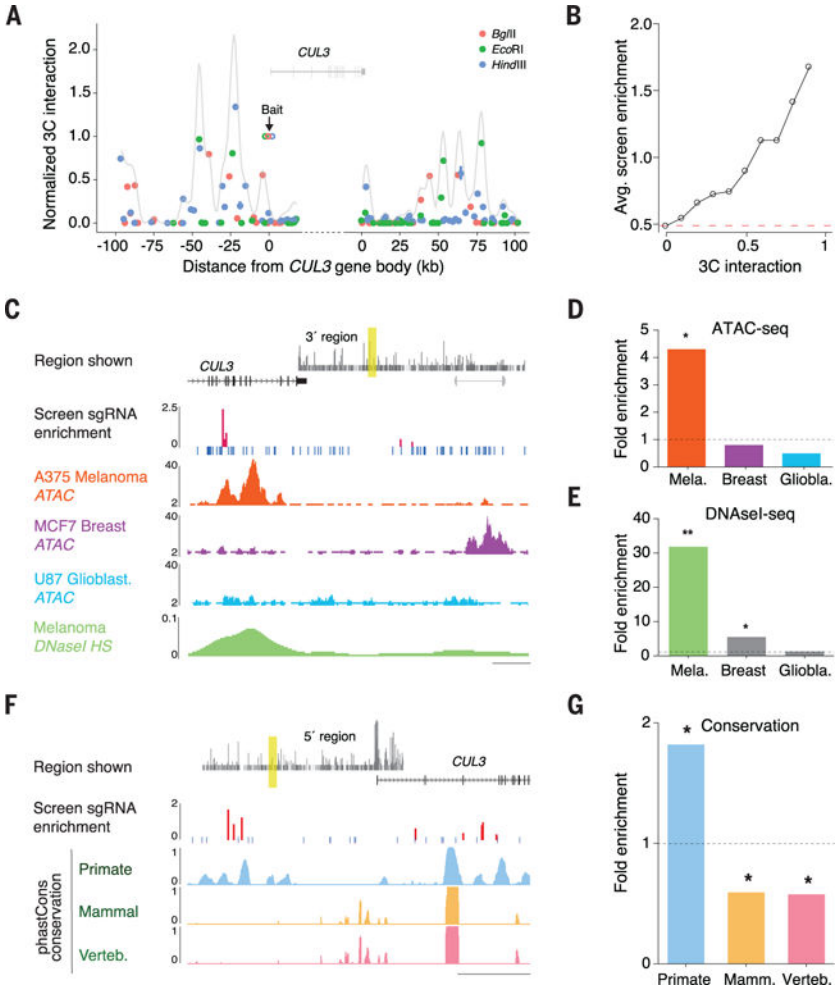

Figure 2. Characteristics of functional noncoding elements at the *CUL3* locus

A) Plot of 3C interaction frequencies with the *CUL3* promoter in A375 cells. Data points represent independent libraries generated with *BglII*, *EcoRI*, and *HindIII* restriction enzymes. The grey curve shows a smoothed estimate of interaction frequency.

B) Average enrichment of sgRNAs ($\log_2$ ratio of vemurafenib/DMSO reads) near 3C sites with specified minimum interaction frequency with the *CUL3* promoter (43).

C) An example of enriched sgRNAs (*red*) that overlap with a melanoma-specific region of open chromatin. ATAC-seq in A375 melanoma (*orange*), MCF-7 breast cancer (*purple*) and U-87 glioblastoma (*blue*) and Melanoma DNAse I hypersensitivity sequencing (*green*, ENCODE/Colo-829). Loci investigated relative to *CUL3* is shown at top (*yellow*). Scale bar: 500 bp.

D) Fold enrichment of enriched sgRNAs near ATAC-seq open chromatin peaks in melanoma, breast cancer and glioblastoma cell lines.

E) Fold enrichment of enriched sgRNAs near DNAse I HS-seq open chromatin peaks in melanoma, breast cancer and glioblastoma cell lines.

F) An example of enriched sgRNAs (*red*) that coincide with regions that show greater primate-specific conservation than placental mammal and vertebrate conservation. Loci investigated relative to *CUL3* is shown at top (*yellow*). Scale bar: 200 bp.

G) Fold enrichment of enriched sgRNAs near phastCons peaks in primates, placental mammals and vertebrates.

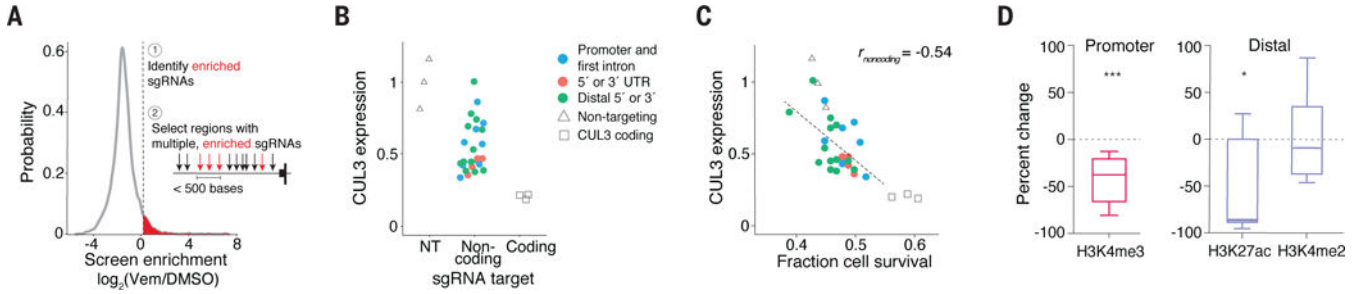

Figure 3. Noncoding mutations impact *CUL3* expression and histone modifications

A) Criteria selecting 25 sgRNAs targeting noncoding regions for validation.

B) *CUL3* RNA expression (normalized to non-targeting sgRNAs) after transduction with lentivirus expressing non-targeting (*triangles*), noncoding region-targeting (*colored circles*) and coding region-targeting (*squares*) sgRNAs.

C) Relationship between *CUL3* expression and cell survival after vemurafenib. Linear fit is to noncoding sgRNAs only ($r_{noncoding} = -0.54$, $p = 0.005$) and does not include coding region-targeting or non-targeting sgRNAs.

D) Percent change in average H3K4me3 chromatin immunoprecipitation (ChIP) for all validation sgRNAs within 1 kb of the transcription start site (TSS) of *CUL3* (*left*). Percent change in average H3K27ac and average H3K4me2 ChIP for all validation sgRNAs >1 kb from the TSS of *CUL3* (*right*).

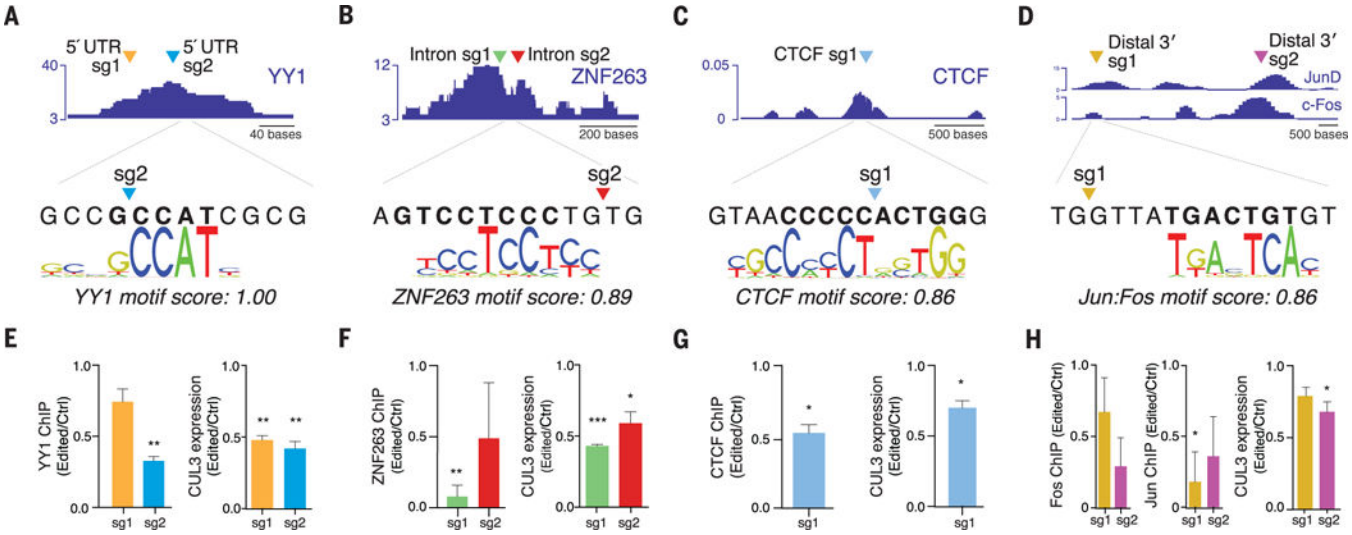

Figure 4. Cas9 mutagenesis disrupts predicted transcription factors and DNA binding proteins at target sites of vemurafenib-enriched sgRNAs

A–D) sgRNA target locations in relation to predicted binding sites.

E–H) Change in transcription factor/DNA binding protein occupancy around cleavage site and change in *CUL3* expression. Both measurements are normalized to cells transduced with non-targeting sgRNAs.