Genome-wide annotation of gene regulatory elements linked to cell fitness

Authors: Tyler S. Klann[1,2,3]*, Alejandro Barrera[2,3]*, Adarsh R. Ettyreddy[1,2,3], Ryan A. Rickels[1,2,3], Julien Bryois[4], Simon Jiang[1,2,3], Shaunak S. Adkar[2,3,5], Nahid Iglesias[1,2,3], Patrick F. Sullivan[4,6,7], Timothy E. Reddy[1,2,3,8#], Gregory E. Crawford[2,3,9#], Charles A. Gersbach[1,2,3,5,10#]

Affiliations:

[1]Department of Biomedical Engineering, Duke University, Durham, NC, USA.

[2]Center for Genomic and Computational Biology, Duke University, Durham, NC, USA.

[3]Center for Advanced Genomic Technologies, Duke University, Durham, NC, USA

[4]Department of Medical Epidemiology and Biostatistics, Karolinska Institutet, Stockholm, Sweden.

[5]Department of Cell Biology, Duke University Medical Center, Durham, NC, USA

[6]Department of Genetics, University of North Carolina at Chapel Hill, Chapel Hill, NC, USA

[7]Department of Psychiatry, University of North Carolina at Chapel Hill, Chapel Hill, NC, USA.

[8]Department of Biostatistics and Bioinformatics, Duke University Medical Center, Durham, NC, USA.

[9]Department of Pediatrics, Division of Medical Genetics, Duke University Medical Center, Durham, NC, USA.

[10]Department of Surgery, Duke University Medical Center, Durham, NC, USA.

*These authors contributed equally

[#]**Correspondence**: Timothy E. Reddy (tim.reddy@duke.edu), Gregory E. Crawford (greg.crawford@duke.edu), and Charles A. Gersbach (charles.gersbach@duke.edu)

Keywords: Regulatory element screen, CRISPR epigenome editing, cell fitness, cell-type specificity, cancer, single cell RNA-seq

Abstract

Noncoding regulatory elements control gene expression and govern all biological processes. Epigenomic profiling assays have identified millions of putative regulatory elements, but systematically determining the function of each of those regulatory elements remains a substantial challenge. Here we adapt CRISPR-dCas9-based epigenomic regulatory element screening (CERES) technology to screen all >100,000 putative non-coding regulatory elements defined by open chromatin sites in human K562 leukemia cells for their role in regulating essential cellular processes. In an initial screen containing more than 1 million gRNAs, we discovered approximately 12,000 regulatory elements with evidence of impact on cell fitness. We validated many of the screen hits in K562 cells, evaluated cell-type specificity in a second cancer cell line, and identified target genes of regulatory elements using CERES perturbations combined with single cell RNA-seq. This comprehensive and quantitative genome-wide map of essential regulatory elements represents a framework for extensive characterization of noncoding regulatory elements that drive complex cell phenotypes and for prioritizing non-coding genetic variants that likely contribute to common traits and disease risk.

Introduction

Human gene regulatory elements control gene expression and orchestrate many biological processes including cell differentiation (*1*), proliferation (*2*), and environmental responses (*3*). Genetic and epigenetic variation that alters gene regulatory element function is a primary contributor to human traits and susceptibility to common disease (*4*). Studies of chromatin state and transcription factor occupancy have identified millions of putative human gene regulatory elements (*5*). The biological importance and large number of putative human gene regulatory elements have motivated the development of high-throughput technologies to measure regulatory element activity genome-wide. Examples include genome-wide assays that measure putative regulatory element activity on reporter gene expression (*6*), and targeted

CRISPR-based methods to measure the effects of genetic or epigenetic perturbation of up to thousands of regulatory elements in their native chromosomal context (*7*).

A key measure of gene or regulatory element function is its contribution to overall cell fitness, comprising the balance of cell survival and proliferation. Genome-wide technologies, such as RNAi and CRISPR-based screens, have identified genes involved in diverse cellular processes (*8–14*). CRISPR-based genetic or epigenetic perturbation of noncoding regulatory elements within specific genomic loci have identified target genes and downstream effects on cell phenotypes (*15–20*). However, these perturbation screens of distal regulatory elements have generally been limited to small regions of the genome or loci encoding oncogenes (*21*). Consequently, functional understanding of the millions of predicted human gene regulatory elements remains sparse, making it difficult to routinely establish gene regulatory contributions to human traits and disease.

Here, we identify thousands of human gene regulatory elements that functionally contribute to cell fitness. To do so, we used a genome-wide CRISPR-based epigenome editing screen that individually targeted each of the >100,000 putative gene regulatory elements in human K562 leukemia cells. We further characterize the properties, distribution, cell-type specificity, and target genes of the identified regulatory elements. The identified regulatory elements and target genes confirm and complement results from gene-based screens, and suggest new pathways and molecular processes that contribute to cell fitness. Comprehensive annotations of regulatory element function, such as the study presented here, are critical for creating new reference datasets to prioritize regulatory elements and noncoding variants that most likely contribute to human traits and diseases.

Results

Genome-wide screen of all regulatory elements in K562 cells

We used whole-genome CERES (wgCERES) to measure the effect of epigenetically silencing 111,756 putative regulatory elements, defined by DNase-I hypersensitive sites (DHS), on cell fitness in K562 cells (**Fig. 1, A and B, fig. S1**) (*22*). We assayed K562 cells because they are one of the most extensively characterized cell models in terms of chromatin accessibility, histone marks, transcription factor binding, and gene expression (*23*). Our library contained 1,092,706 unique gRNAs averaging ~10 gRNAs per DHS (**tables S1 to S4**), and we transduced this library into a clonal K562 cell line stably expressing the dCas9^{KRAB} transcriptional repressor (*24*, *25*). After ~14 population doublings, we identified a significant depletion of gRNAs for 7,696 DHSs, indicating that repressing those DHSs impaired cell viability or proliferation (**Fig. 1, C and D, fig. S2A**). We also found 4,566 DHSs with gRNAs or combinations of gRNAs that were significantly enriched, indicating that repressing these elements increased cell fitness (**Fig. 1D, fig. S2B, table S5**). A relatively small number of DHS hits (n=228) contained both enriched and depleted gRNAs (**Fig. 1D, fig. S2C**).

Attributes of gRNAs that impact cell fitness

Effect sizes for gRNAs that reduced cell fitness were overall greater (average $\log_2$(fold-change) = -0.91) than those that increased cell fitness (average $\log_2$(fold-change) = 0.48, **Fig. 1E**). That result is consistent with a model that it is easier to reduce fitness than increase fitness of the rapidly growing K562 cell line.

To better understand the characteristics that distinguish the significantly enriched or depleted gRNAs, we annotated each gRNA in the library with a selection of features (**fig. S3**). The gRNAs with significantly changed abundance were enriched for GC content in the protospacer, G4 quadruplex motifs

(*26*, *27*), nearby genes that were more highly expressed, higher accessibility, higher H3K27ac marks, and higher Hi-C contact frequency (**fig. S3**). Additionally, genes nearest to significant gRNAs were enriched for genes previously identified as essential (*11*, *12*) (**fig. S3**). These features have been used previously to predict enhancer-gene interactions (*17*, *28*), and support the power of this genome-wide screen to identify active regulatory elements associated with our selection criteria.

Attributes of significant DHS hits

While significant DHS hits are called at a continuum of distances from the nearest gene, the strongest observed signals centered on DHSs that overlapped with transcriptional start sites (TSSs, **Fig. 1F**). This is consistent with previous studies showing that repressing promoters with dCas9^{KRAB} has a larger effect on gene expression than repressing distal regulatory elements (*19*, *25*). Although overall scores decrease away from TSSs, some distal DHSs have particularly strong signals, similar to TSS DHS hits. For example, several DHS hits 10 kb - 1 Mb upstream of their putative target genes scored similarly to gRNAs that target the promoter of the same gene (**Fig. 1G, fig. S4**). Some of the distal elements were previously validated in mice to control genes such as the oncogene *Lmo2* in erythroid cell lineages (*29*). Together, this indicates that wgCERES can identify regulatory elements distal from target genes, and quantify the relative impact of those regulatory elements on cell proliferation.

To identify epigenetic characteristics of DHS that control cell fitness, we used dimensionality reduction analysis and compared DHS hits to K562 ChIP-seq data for several histone modifications and epigenome-modifying proteins from the ENCODE project (**fig. S5**) (*23*). We then used ChromHMM genome annotations (*30*) to identify classes of regulatory elements that were overrepresented in the enriched or depleted DHS hits (**Fig. 1H**). We observed hits in almost every class of annotation, including regions classified as polycomb-repressed (**fig. S6**). Relative to all DHS sites, depleted DHS hits were overrepresented at active promoters and underrepresented at enhancers and CTCF sites (**Fig. 1I**). In contrast, enriched DHS hits have similar genomic location characteristics as all DHS (**Fig. 1I**). Together,

these results indicate that promoters, enhancers, insulators, and polycomb-repressed regions can all contribute to cell fitness.

Clustering of Significant DHS Hits

Clusters of individual regulatory elements can function together as larger ensembles to coordinate gene expression, as seen with the β-globin locus control region (*31*, *32*). To determine if DHS hits from this screen cluster together, we compared the distances between adjacent DHS hits. We separately measured distances for proximal DHS hits (TSSs, **fig. S7A**) and distal DHS hits (>3 kb away from TSS, **fig. S7B**). We observed that DHS hits are significantly closer to each other than expected by chance using permutation analysis (**fig. S7, A and B**). Dividing the data into deciles (**fig. S7, C and D**), DHS hits were significantly closer to each other in all but the most distant decile (t-test, $p < 0.0001$). Some clusters included up to 7 significant DHS hits, such as around the HDAC7/VDR locus that is known to be involved in cancer cell proliferation (*33*, *34*) (**fig. S7E**). Together, these results indicate that regulatory elements that influence cell fitness tend to cluster. That may be due to coordinated effects on key cell fitness genes and/or the presence of clustered genes that contribute to cell fitness.

Validation of distal regulatory DHS hits using a secondary screen

We validated primary screen hits using both comparisons to previously identified essential genes and a secondary screen targeted to positive hits. For promoter hits, we compared our results to other studies of promoter inactivation (*35*) or gene disruption (*9*, *12*) in K562 cells. The observed promoter hits positively correlated (Pearson $\rho = 0.62$, Spearman $\rho = 0.18$) with the promoter CRISPRi screen (*35*) (**fig. S8A, table S5**). Overall, ~400 genes were hits in all four studies (**fig. S8B**). That was the most common configuration

for overlapping hits, indicating substantial concordance between gene- and promoter-based screens for effects on cell fitness.

Our screen was distinct from previous efforts in that most of our gRNAs targeted putative distal regulatory elements. To validate and characterize the effects of individual gRNA and DHS hits, we completed a validation screen of 234,593 gRNAs that collectively target 8,850 DHSs, of which 7,188 were hits called at a FDR < 0.1 in the initial discovery screen (**fig. S1, tables S5 to 13**). Individual gRNA effects in the validation screen had a similar distribution as the original screen, both in terms of effects sizes and that most hits corresponded to a decrease in cell fitness (**Fig. 2A, table S5**).

To evaluate performance for single gRNAs, we characterized 50,021 individual gRNAs assayed in both the discovery screen and the validation screen. Of those, 4,087 gRNAs were individually significant hits in the discovery screen at an FDR < 0.1, and 1,829 were also significant in the validation screen at an FDR < 0.1 (**Fig. 2B, fig. S9**). The remaining 45,934 gRNAs were included because they were in DHSs that had a significant effect but the individual gRNA did not. Of the 45,934 gRNAs that were negative in the discovery screen, 43,131 gRNAs were also negative in the validation screen. Together, the validation screen indicated gRNA-level sensitivity of 40% with 45% precision and 95% specificity. We also evaluated performance of the DHS-level analysis. Of the 8,850 DHS targeted, 7,188 had significant effects at FDR < 0.1 in the discovery screen, and 3,532 (49%) positively validated at a more stringent FDR < 0.05 in the validation screen.

The validation screen had more significant gRNA hits per DHS (**Fig. 2, C and D**), suggesting that increasing the density of gRNAs tested per DHS from 10 to 26 improved detection of regulatory elements

that impact cell fitness. That improved detection may be in part due to variation in the effects of gRNAs targeting the same DHS (*25*).

Functional validations of target genes for DHS hits by expression and competition assays.

To test the effects of distal regulatory elements on target gene expression, we measured the effects of individually targeting dCas9^{KRAB} via 23 gRNAs on the expression of 22 predicted target genes using qRT-PCR (**Fig. 3A, fig. S10, and tables S14 and S15**). Predicted DHS-gene links were obtained using the Activity-by-Contact (ABC) model (*28*). There were significant changes in predicted target gene expression for 15 of the 23 gRNAs targeting 7 of 13 DHSs. The genes altered by targeting these DHSs include *LMO2*, *GATA1*, and *GMPR*. Previous studies have shown *LMO2* and *GATA1* are essential for K562 cell growth (*9*, *36–38*).

To measure transcriptome-wide effects of a subset of the above-described perturbations, we used RNA-seq. Our analyses revealed that epigenetic perturbations of individual DHS resulted in many differentially expressed genes, and sometimes the predicted target gene was most affected (**Fig. 3B, fig. S11**). As one example, we evaluated the effect of perturbing four different distal DHS hits around the *SLC4A1* gene, which is a gene involved in differentiation (*39*) and when mutated causes hereditary spherocytosis and erythrocyte fragility (*40*). After perturbing each of these regions, expression of the *SLC4A1* gene was the most significantly reduced, and there was also a high correspondence of gene ontology similarities for other significant differentially expressed genes (**fig. S11**). There were also instances that the ABC predicted target gene was not the most differentially expressed gene (**Fig. 3, C and D, figs. S12 and S13**). For example, targeting two DHS hits in an intron of the GMPR gene did not impact GMPR expression, but did impact sets of histone genes 8 Mb away, and overall displayed similar gene

ontologies (**fig. S12**). Indeed, that result may explain why repressing those DHS impacted cell fitness even though GMPR has not been shown to be essential previously.

To further functionally characterize the targeted group of DHS hits, we used a cell growth competition assay to validate whether silencing each distal regulatory element reduces cellular fitness (**Fig. 4A**). Seven of 10 gRNAs that were depleted in the secondary screen also reduced cell fitness in the competition assay (**Fig. 4B**). Similarly, all 10 gRNAs that were enriched in the secondary screen also increased cell fitness in the competition assay (**Fig. 4C**). Therefore the effect of the epigenetic perturbations on the selected phenotype was robust and reproducible, even if the target gene of the regulatory element was not immediately apparent.

Identification of cell-type specific essential gene regulatory elements

We used chromatin accessibility data from 53 different cell types (https://personal.broadinstitute.org/meuleman/reg2map/) to characterize cell type specificity of DHS hits involved in cell fitness in K562 cells. For the significant DHS hits in our screen, most of the regions only overlapped open chromatin in K562 cells, while fewer regions overlapped open chromatin shared across many cell types (**Fig. 5A**). This suggests that many of the DHS hits we identified affect fitness in a cell-type specific manner.

To functionally assess the generalizability of essential regulatory elements across cell types, we re-purposed the validation gRNA library used on the chronic myeloid leukemia (CML) K562 cell line (**Fig. 2**) to perform an additional screen in the acute myeloid leukemia (AML) cell line OCI-AML2 (**tables S10 to S13**). Similar to the results in K562 cells, we also detected depleted gRNAs with larger effect sizes in OCI-AML2 cells (**Fig. 5B**). When comparing individual gRNAs between OCI-AML2 and K562 cells, we detected 5,088 gRNAs that are significantly depleted in both cell types, indicating these gRNAs lie in DHSs that are essential across different cancer cell lines (**Fig. 5C**). We also detected 1,855 gRNAs that are significantly depleted only in K562 cells and 15,670 gRNAs significantly depleted only in OCI-AML2

cells, indicating that OCI-AML2 cells may be more sensitive to regulatory element perturbation. We also detected small numbers of non-targeting guides in shared and cell type specific hits, representing around 2-5% of the pool of significant gRNAs, which supports our estimates of FDR of 5% (**tables S6 to S13**). A number of DHS hits overlap with chromatin accessibility that are shared between cell types (**Fig. 5D**) or are cell type-specific (**Fig. 5E**), indicating that chromatin accessibility may define essential regulatory element activity.

Identification of regulatory element target genes using single cell expression

To empirically identify the target genes for the distal regulatory elements detected in these screens, we adapted a method that combines single cell RNA-seq readout with CRISPR screens (*20*, *41–44*), which we refer to as single-cell CERES (scCERES). This allows the capture and quantification of all mRNA and gRNA identity on a per-cell basis, enabling the identification of genes that change in response to regulatory element perturbations. For this screen, polyclonal K562 cells constitutively expressing dCas9^{KRAB} were transduced with a library of 3,201 gRNAs (**fig. S1, tables S16 and S17**) cloned into the opti-CROP-seq plasmid (*20*, *43*) in order to capture gRNA sequence identity on the 10x Genomics platform.

We transduced cells at an MOI of ~7 to increase overall library coverage, found each cell contained an average of 8 gRNAs, and each gRNA was represented by an average of 111 cells (**fig. S14, A and B**). After library preparation and sequencing, differential expression analysis was performed by grouping cells that expressed the same gRNA (**fig. S14C**). To increase statistical power to detect changes in gene expression, differential expression tests were limited to genes in a 2 megabase window centered on the DHS (**Fig. 6A**).

Collectively, we identified 992 genes that were affected by perturbing 815 unique regulatory elements. While most genes (N = 932) had only a single link to a regulatory element, 52 genes were linked to 2 regulatory elements, and 8 genes were connected to 3 regulatory elements (**Fig. 6B**). The majority of the regulatory elements (N = 638) only affected a single gene. However, perturbation of 177 regulatory

elements altered expression of 2 or more genes within the 2 Mb window, including one element that affected 7 genes (**Fig. 6C and D**). Interestingly, this multi-gene affector overlaps a CTCF site, suggesting it may impact a TAD domain (**Fig. 6D**). We also found DHS hits in polycomb regions that affect genes outside of the polycomb repressed region (**fig. S6**).

Several gene-regulatory element links were corroborated by validating changes in gene expression by RT-qPCR following delivery of a single gRNA, including the ATF7IP (**fig. S10**), GMPR (**Fig. 3A**), and LMO2 loci (**Fig. 6, E to G**). For LMO2, we were able to identify gene-enhancer connections for two regulatory elements ~60kb upstream of the LMO2 gene that were among the strongest hits from the initial wgCERES screen (Target 1 and 2, **Fig. 6E**). Repression of either element by dCas9^{KRAB} led to significantly reduced expression of the LMO2 gene (**Fig. 6, E and F**). A third regulatory element (Target 3) in the same cluster did not show statistically significant changes in LMO2 expression by scCERES (**Fig. 6, E and F**), but did show comparatively modest repression by RT-qPCR (**Fig. 6G**). This may represent the need for increased sequencing depth to achieve better sensitivity. Regardless, this single cell readout identifies a substantial number of regulatory elements that are simultaneously linked to both target genes and cell fitness.

Discussion

Virtually all human phenotypes and disease susceptibilities are linked to the function of gene regulatory elements and associated genetic variation in the non-coding genome (*4*, *45*). There is strong rationale for cancer susceptibility and progression to be similarly dictated by the non-coding genome, and a few discrete examples of this mechanism have recently emerged (*46–49*). However, cancer genetics and the discovery of oncogenic driver mutations has historically been limited to analysis of protein coding sequences because (i) whole-genome sequencing of primary tumors is costly, and (ii) our functional understanding of noncoding genetic variation is still in its infancy. The first limitation is being addressed by lower costs of whole genome sequencing and growing databases of thousands of whole genomes from

tumor and healthy control tissue (*50*, *51*). However, these advances also emphasize the urgency in addressing the second limitation, which has recently become a tractable challenge due to the advent of CRISPR-based screens of non-coding regulatory elements in their endogenous chromosomal context (*13*–*15*, *17*–*19*, *25*, *52*). Knowing the location of regulatory elements that affects cell growth is a critical first step for cancer geneticists to find additional noncoding cancer driver mutations (*50*, *53*).

This study is a significant step towards addressing these limitations and realizing the potential of whole genome sequencing for cancer biology. We describe a systematic genome-wide screen of all putative regulatory elements in a commonly used cancer cell line and describe their role in cell fitness. We identified >12,000 regulatory elements that have negative or positive impacts on cellular viability and/or proliferation, and report ~1,000 element-gene links that drive this phenotype. These data provide a rich resource of regulatory element function and connection to target genes that will be broadly useful for understanding gene network regulation and the mechanisms of non-coding element control on gene expression. We expect these characterizations that relate the non-coding genome to cell fitness will identify functional noncoding sequence variants that contribute to cancer phenotypes. These functional annotations also complement the growing body of chromatin conformation maps that provide structural relationships between regulatory elements and genes (*54*). Moreover, this work provides a blueprint for executing similar studies in other cell types, genetic backgrounds, environmental conditions, or pharmacologic treatments. In the future, we hope that this approach will facilitate the development of methods to predict element-gene relationships and inform efforts to learn the quantitative rules of gene regulation (*17*, *55*).

A challenge to implementing genome-wide screens of the non-coding genome is the sheer scale of the experiment, which is dictated by the number of putative elements in any cell type and the required numbers of gRNAs per element and cells per gRNA. As the field of CRISPR-based screens is still in its relative infancy, an important area of future focus is the design of more efficient and sensitive screening methods. For example, this dataset may be used to define the properties for effective gRNA design in distal regulatory elements, similar to what has been done for designing optimal gRNA libraries for genes and promoters (*35*, *52*, *56*). This work depended on extensive characterization of gRNA libraries targeting these

classes of elements. In contrast, relatively little is known about which key gRNA attributes contribute to effective perturbation of distal regulatory elements. We expect that the knowledge gained from thousands of gRNAs that impact cellular growth from distal regulatory elements will facilitate the design of more compact and robust libraries, and enable similar genome-wide screens in cell lines or primary cells that are more difficult to culture at scale.

Many epigenetic modifying drugs used as potential cancer treatments cause widespread changes throughout the genome (*57*, *58*). However, it is currently unclear what subset of gene regulatory elements drive drug response. Using maps of essential regulatory elements in conjunction with the epigenetic profiles of cells after drug treatment could help identify modifications to specific gene regulatory elements necessary and sufficient for drug response. This may ultimately inform the development of safer and more specific cancer therapies.

Interestingly, one of the loci with the strongest effect on cellular proliferation was the *LMO2* locus. This locus is also the location of retroviral insertions in gene therapy patients which lead to increased expression of LMO2 via viral enhancer elements and ultimately led to leukemia (*59*). Better understanding the regulatory landscape of these and other types of regions will help elucidate mechanisms of aberrant gene expression and tumorigenesis that will ultimately also inform design, safety monitoring, and regulation of emerging classes of genetic medicines such as gene therapy and genome editing. Therefore we anticipate this approach will be a valuable resource to diverse fields of the biomedical research community.

Materials and methods

Plasmids

The lentiviral dCas9-KRAB plasmid (Addgene #83890) was generated by cloning in a P2A-HygroR (APH) cassette after dCas9-KRAB using Gibson assembly (NEB, E2611L). The lentiviral gRNA

expression plasmid was cloned by combining a U6-gRNA cassette containing the gRNA-(F+E)-combined scaffold sequence (*60*) with an EGFP-P2A-PAC or mCherry-P2A-PAC cassette into a lentiviral expression backbone (Addgene #83925) using Gibson assembly. Individual gRNAs were ordered as oligonucleotides (IDT-DNA), phosphorylated, hybridized, and ligated into the EGFP gRNA plasmid or the mCherry gRNA plasmid using BsmBI sites.

Cell Culture

K562 and HEK293T (for lentiviral packaging) cells were obtained from the American Tissue Collection Center (ATCC) via the Duke University Cancer Center Facilities. OCI-AML2 cells were gifted from Anthony Letai at Dana Farber Cancer Institute. K562 and OCIAML2 cells were maintained in RPMI 1640 media supplemented with 10% FBS and 1% penicillin-streptomycin. HEK293T cells were maintained in DMEM High Glucose supplemented with 10% FBS and 1% penicillin-streptomycin. All cell lines were cultured at 37 °C and 5% CO2.

For the genome-wide discovery screen, a clonal K562-dCas9^{KRAB} cell line used, and was generated by transduction of dCas9-KRAB-P2A-HygroR lentivirus with polybrene at a concentration of 8 µg/mL. Cells were selected 2 days post-transduction with Hygromycin B (600 µg/mL, ThermoFisher, 10687010) for 10 days followed by sorting single-cells into 96-well plates with a SH800 sorter (Sony Biotechnology). Individual clones were grown and stained for dCas9^{KRAB} with a Cas9 antibody (Mouse mAb IgG1 clone 7A9-3A3 Alexa Fluor 647 Conjugate, Cell Signaling Technologies, 48796) to assess protein expression. Briefly, 1x10^6 cells were harvested and washed once with 1X FACS buffer (1% BSA in PBS). The cells were then fixed and permeabilized for 30 minutes at room temperature with 500 µL of fixation and permeabilization buffer (eBioscience Foxp3/TF/nuclear staining kit, ThermoFisher, 00-5523-00). Next, 1 mL of permeabilization buffer was added and cells were pelleted (600 RCF for 5 min) and washed again in 1 mL of permeabilization buffer. Cells were pelleted again and resuspended in 50 µL of permeabilization buffer with 2% mouse serum (Millipore Sigma, M5905) to block for 10 minutes at room temperature. Following blocking, 50 µL of permeabilization buffer with 2% mouse serum and 1 µL of Cas9 antibody

was added and allowed to incubate for 30 minutes at room temperature. Following incubation, 1 mL of permeabilization buffer was added, cells were pelleted and washed once more with 1 mL of permeabilization buffer. Finally, cells were resuspended in 1X FACS buffer for analysis. Each clone was analyzed using an Accuri C6 flow cytometer (BD Biosciences). A clone was selected based on high and uniform expression of dCas9^{KRAB} and expanded for further use.

For the secondary sub-library screens, we used polyclonal K562 and OCI-AML2 cell lines that express the dCas9^{KRAB} repressor. Polyclonal lines were used to account for possible hits in the first screen that could be specific to the clonal line used. K562 and OCI-AML2 cells were transduced with dCas9-KRAB-P2A-HygroR lentivirus with polybrene at a concentration of 8 µg/mL. At two days post-transduction, cells were selected for 10 d in Hygromycin B (600 µg/mL). Following selection, polyclonal cells were stained to detect expression of dCas9^{KRAB} protein as described above.

gRNA Library Design

DNase I hypersensitive sites (DHSs) for the K562 cell line were downloaded from www.encodeproject.org (ENCFF001UWQ) and used to extract genomic sequences as input for gRNA identification. We used the gt-scan algorithm to identify gRNA protospacers within each DHS region and identify possible alignments to other regions of the genome (*61*). The result is a database containing all possible gRNAs targeting all targetable DHSs in K562 cells and each gRNA's possible off-target locations. gRNAs were selected based on minimizing the number of off-target alignments. For the initial genome-wide library, 1,092,706 gRNAs were selected (**tables S1 to S4**), targeting 111,756 DHSs (269 DHSs contained no NGG SpCas9 PAM), limited to a maximum of 10 gRNAs per DHS (mean, 9.77 gRNAs per DHS).

For the second sub-library targeting distal non-promoter hits (>3kb from TSS) identified in the first screen, 234,593 gRNAs were selected (**tables S6 to S13**), targeting 8,850 distal DHSs identified as significant (FDR-adjusted p-value < 0.1) from the first screen. For each DHS, gRNAs were chosen to be spread evenly across the region by dividing each DHS into bins of 100 bp and selecting up to 7 gRNAs per

bin. The gRNAs for each bin were selected in order by the fewest number of off-target alignments calculated by gt-scan. 15,407 non-targeting gRNAs were designed as previously described (*35*). A larger number of gRNAs per DHS were designed in the second screen (~24 per DHS) compared to the first screen (10 per DHS).

All libraries were synthesized by Twist Biosciences and the oligo pools were cloned into the lentiviral gRNA expression plasmid using Gibson assembly as previously described (*21*). Briefly, oligo pools were amplified across 16 PCRs (100 ng oligo per PCR) with the following primers for 10 cycles using Q5 2X master mix and the following primers:

Fwd: 5′- TAACTTGAAAGTATTTCGATTTCTTGGCTTTATATATCTTGTGGAAAGGACGAAACACCG

Rev: 5′- GTTGATAACGGACTAGCCTTATTTAAACTTGCTATGCTGTTTCCAGCATAGCTCTTAAAC

Pools were gel purified (Qiagen, 28704) and used to assemble plasmid pools with Gibson assembly (NEB, E2611L). Pools were assembled across 16 Gibson assembly reactions (~900 ng backbone, 1:3 backbone to insert) for the first screen, and 4 reactions for the second sub-library screen.

Lentivirus Production

The lentivirus encoding gRNA libraries or dCas9^{KRAB} was produced by transfecting 5x10^6 HEK293T cells with the lentiviral gRNA expression plasmid pool or dCas9^{KRAB} plasmid (20 µg), psPAX2 (Addgene, 12260, 15 µg), and pMD2.G (Addgene, 12259, 6 µg) using calcium phosphate precipitation (*62*). After 14-20 hours, the transfection media was exchanged with fresh media. Media containing lentivirus was collected 24 and 48 hours later. Lentiviral supernatant was filtered with a 0.45 µm CA filter (Corning, 430627). The dCas9^{KRAB} lentivirus was concentrated 20X the initial media volume using Lenti-X concentrator (Clontech, 631232), following manufacturer's instructions. The lentivirus encoding gRNA libraries was used unconcentrated.

The titer of the lentivirus containing either the genome-wide library or distal sub-library of gRNAs was determined by transducing 5x10^5 cells with varying dilutions of lentivirus and measuring the percentage of GFP-positive cells 4 days later using the Accuri C6 flow cytometer (BD Biosciences).

To produce lentivirus for individual gRNA validations, $8x10^5$ cells were transfected with gRNA plasmid (2440 ng), psPAX2 (1830 ng), and pMD2.G (730 ng) using Lipofectamine 3000 following the manufacturer's instructions. After 14 to 20 hours, transfection media was exchanged with fresh media. Media containing produced lentivirus was harvested 24 and 48 hours later, centrifuged for 10 minutes at 800xg, and directly used to transduce cells.

Lentiviral gRNA Screens

For the first genome-wide screen, $1.7x10^9$ cells were transduced with the gRNA library during seeding in 3 L spinner flasks across 4 replicates for controls (K562 cells without dCas9^{KRAB}) and 4 replicates for dCas9^{KRAB}-expressing cells. For sub-library screens, $4.17x10^8$ cells were transduced during seeding in 500 mL spinner flasks across 4 replicates for both controls and dCas9^{KRAB}-expressing cells. Cells were transduced at a multiplicity of infection (MOI) of 0.4 to generate a cell population with >80% of cells harboring only 1 gRNA and 500-fold coverage of each gRNA library. After 2 days, cells were treated with puromycin (Millipore Sigma, P8833) at a concentration of 2 μg/mL. Cells (control and dCas9^{KRAB}-expressing) were selected for 7 days and allowed to grow for a total of 16 days (including 7 days of selection, or ~14 doublings). Cells were passaged to ensure at least 500X fold coverage of the gRNA library to maintain representation. After culturing, for the genome-wide screen, $5.5x10^8$ K562 cells were harvested for genomic DNA isolation. For the sub-library distal screens in K562 cells or OCI-AML2 cells, $1.5x10^8$ cells were harvested. Genomic DNA was harvested from cells as described by Chen and Sanjana et al.(*63*).

Single-Cell RNA-seq Screen

For the single-cell RNA-seq screen, cells constitutively expressing dCas9^{KRAB} were transduced with a library of gRNAs cloned into the CROP-seq-opti vector (Addgene #106280) in order to capture gRNA information on the 10X platform. The library contains 3,201 total gRNAs consisting of the most significant gRNA for all 3,051 distal DHS hits identified in the second K562 distal sub-library screen, as well as the most significant gRNA for a subset of TSS DHSs as positive controls, and 150 non-targeting

gRNAs as negative controls (**tables S16-S17**). Cells were transduced at an MOI of ~7 to achieve multiple integrations of gRNAs per cell, as done previously (*20*). Cells were grown for 5 days after transfection of the gRNA library and 56,882 cells were collected and barcoded with the 10X 3' v3 chemistry. gRNAs were amplified from barcoded cDNA as described in previously (*20*). Total transcriptome libraries were sequenced on a NovaSeq S4 flow cell and gRNA-enriched libraries were sequenced on a NextSeq 550 flow cell.

Genomic DNA Sequencing

To amplify the genome-wide gRNA libraries from each sample, 5.25 mg of genomic DNA (gDNA) was used as template across 525 x 100 µL PCR reactions using Q5 2X Master Mix (NEB, M0492L). For the distal sub-library screens, 1.2 mg of gDNA was used as template across 120 PCR reactions using Q5 2X Master Mix. Amplification was carried out following the manufacturer's instructions using 25 cycles at an annealing temperature of 60 °C using the following primers:

`Fwd` `5'-AATGATACGGCGACCACCGAGATCTACACAATTTCTTGGGTAGTTTGCAGTT`

`Rev` `5'-CAAGCAGAAGACGGCATACGAGAT` (6 bp index sequence) `GACTCGGTGCCACTTTTTCAA`

Amplified libraries were purified using Agencourt AMPure XP beads (Beckman Coulter, A63881) using double size selection of 0.65X and then to 1X the original volume. Each sample was quantified after purification using the Qubit dsDNA High Sensitivity Assay Kit (ThermoFisher, Q32854). Samples were pooled and sequenced on a HiSeq 4000 or NovaSeq 6000 (Illumina) at the Duke GCB sequencing core, with 21 bp single read sequencing using the following custom read and index primers:

`Read1 5'-GATTTCTTGGCTTTATATATCTTGTGGAAAGGACGAAACACCG`

`Index 5'-GCTAGTCCGTTATCAACTTGAAAAAGTGGCACCGAGTC`

Data Processing and Differential Expression Analysis of gRNA libraries

To identify and quantify the effects of regulatory element perturbation on cell fitness, we compared gRNA abundance before and after cell growth. Since library size constraints limited the number of gRNAs

per DHS and as the effect of any individual gRNA may be subtle, we characterized the effects of perturbing each DHS by four levels of gRNA analyses: 1) individual gRNAs, 2) a sliding window across each DHS in bins of two gRNAs, 3) a sliding window across each DHS in bins of three gRNAs, and 4) grouping all gRNAs in a DHS together (**Fig. 1B**).

FASTQ files were aligned to custom indexes (generated from the bowtie2-build function) using Bowtie2 (*64*) (options -p 24 --no-unal --end-to-end --trim3 6 -D 20 -R 3 -N 0 -L 20 -a). Counts for each gRNA were extracted and used for further analysis. All gRNA enrichment analysis was performed using R. For differential expression analysis, the DESeq2 package was used to compare between dCas9^{KRAB} and control (no dCas9^{KRAB}) conditions for each screen.

To summarize enrichment or depletion across a DHS in the first screen, we generated composite scores (wgCERES-top3 score) where the list of gRNAs, bins of 2 gRNAs, or bins of 3 gRNAs for each DHS were sorted by adjusted p-value (ascending order, calculated from DESeq2) and the average of the top three $\log_2$(fold-change) values in each category was calculated. The $\log_2$(fold-change) averages (or single value for the DHS group) of each analysis category (gRNA/bin2/bin3/DHS) were then summed to calculate the wgCERES-top3 score. For the distal screens, the same procedure was performed except instead of the top 3 gRNAs/bins, the top 5 were averaged since gRNAs were more densely tiled for each DHS (wgCERES-top5 score).

Data Processing and Differential Expression Analysis of Single-Cell RNA-seq Screen

Sequencing data from transcriptome and gRNA libraries generally used distinct pre-processing pipelines, as detailed below. However, for both types of libraries, reads were first demultiplexed using the mkfastq command from 10X Genomics Cell Ranger 3.1.0 with the default configuration and BAM files with transcript counts were generated using the count command and the hg19 reference dataset included in Cell Ranger 3.1.0. At that point, the preprocessing of the transcriptomic data finished.

Custom processing of the gRNA library sequencing data

For the gRNA libraries, we filter out properly aligned reads, since usable reads should not map against the hg19 transcriptome. We converted BAM files containing unaligned reads into FASTQ files using the bam2fq command in samtools. Next, we used the custom bowtie2 index from the wgCERES library described above to align the reads again using bowtie2. We trimmed 23 and 48 bp from the 5` and 3` ends respectively of the reads to remove scaffolding sequences. The full set of bowtie2 params were --trim5 23 --trim3 48 --no-unal --end-to-end -D 15 -R 2 -N 1 -L 18 -i S,1,0 --score-min G,0,0 --ignore-quals. Because of this extra step, we lost the corrected cell barcodes and UMI tags assigned by the cellranger software. We recover those by extracting into FASTQ files the optional fields CB for cell barcodes and UB for UMI barcodes, and reassigning these to the BAM files created with the custom bowtie index using the AnnotateBamWithUmis function in the fgbio package (v.0.8.1). Finally, a custom script (scRNAseq.extract_umi_counts_from_grna_bam.py) was written to extract unique UMI counts per gRNA per cell. The resulting sparse matrix was saved in MarketMatrix format, compatible with existing single-cell RNA-seq software.

Differential expression analysis of single-cell CERES

We loaded both the gRNA library and transcriptomic data in R using the Read10X function from Seurat v3.1.2 and merged the information in a single Seurat object. We assigned gRNAs to cells by requiring gRNAs to have $\geq$ 5 UMI counts and $\geq$ 0.5% of the total UMI counts in a cell (library size). Cells with >20% of mitochondrial UMI counts or <10,000 transcript UMIs were filtered out. Transcriptomic UMI counts were normalized using the NormalizeData function with default parameters. Cells with no gRNA assigned were discarded.

For each target gRNA, we used the FindMarkers function to test genes in the ±1Mb window around the gRNA midpoint. MAST was the test used to recover significant differences in the expression of transcripts from cells containing the gRNA versus all other cells. The union set of all genes tested at least once was used to run the same analysis for non-targeting guides.

Finally, for each target gRNA-gene pair, an empirical p-value was calculated by counting the number of instances in which the observed p-value was larger than those in the non-targeting gRNA-gene pairs.

Permutation analysis

To test whether significant DHS hits clustered at distances that were closer than random chance, we generated 1000 permutations of non-significant DHS sets. Each permutation had the same number of non-significant DHSs as the significant DHS set. We then measured the distance for each significant DHS to any non-significant DHS from each permuted set.

Individual gRNA Validations using qRT-PCR, RNA-seq and competition assays

Validation of individual gRNAs in distal (non-promoter) putative regulatory elements were chosen from a list of 81 element-gene connections predicted by the ABC model (*17*, *28*) (**tables S14 and S15**). We focused these validations on distal DHS hits that also had a corresponding promoter DHS hit of the predicted ABC target gene. From the list of ABC-predicted element-gene connections, we also included several gRNAs corresponding to nearby DHSs that were significant in the wgCERES screen but did not have a predicted gene by ABC.

The protospacers from the top enriched gRNAs found in each screen (**tables S1 to S13**) were ordered as oligonucleotides from IDT and cloned into a lentiviral gRNA expression vector as described earlier. The same modified cell lines used in the corresponding screen were used for the individual gRNA validations. The cells were transduced with individual gRNAs and after 2 days were selected with puromycin (2 µg/mL) for 7 days for the four distal gRNAs not connected with an ABC connection or 4 days for the gRNAs targeting DHSs connected to genes by ABC model predictions.

For all screen validations by qRT-PCR and RNA-seq, mRNA expression analysis was done in triplicate. Total mRNA was harvested from cells and cDNA was generated using the TaqMan Fast Advanced Cells-to-CT kit (ThermoFisher, A35377). qRT-PCR was performed using the TaqMan Fast

Advanced Cells-to-CT kit with the FX96 Real-Time PCR Detection System (Bio-Rad) with the TaqMan probes listed in **table S15**. The results are expressed as fold-increase mRNA expression of the gene of interest normalized to TBP expression by the ΔΔCt method.

RNA-seq analysis was performed as follows. Raw reads were trimmed to remove adapters and bases with average quality score (Q) (Phred33) of < 20 using a 4bp sliding window (SLIDINGWINDOW:4:20) with Trimmomatic v0.32 (*65*). Trimmed reads were subsequently aligned to the primary assembly of the GRCh37 human genome using STAR v2.4.1a (*66*). Aligned reads were assigned to genes in the GENCODE v19 comprehensive gene annotation (*67*) using the featureCounts command in the subread package v1.4.6-p4 with default settings (*68*). Differential expression analysis was performed using DESeq2 v1.22.0 (*69*) in R (v3.5.1). Briefly, raw counts were imported and filtered to remove genes with low or no expression (i.e., keeping genes having ≥2 counts per million (CPMs) in ≥2 samples). Filtered counts were then normalized using the DESeq function, which internally uses estimated size factors accounting for library size, estimated gene and global dispersion. To find significantly differentially expressed genes, the nbinomWaldTest was used to test the coefficients in the fitted Negative Binomial GLM using the previously calculated size factors and dispersion estimates. Genes having a Benjamini-Hochberg false discovery rate (FDR) less than 0.05 were considered significant (unless otherwise indicated). Log2 fold-change values were shrunk towards zero using the adaptive shrinkage estimator from the 'ashr' R package (*70*). For estimating transcript abundance, transcripts per million (TPMs) were computed using the rsem-calculate-expression function in the RSEM v1.2.21 package (*71*).

For growth competition assays, $1x10^6$ cells were transduced with lentivirus encoding a single gRNA into polyclonal K562 dCas9KRAB cells. Cells were transduced with either 1) an individual targeting gRNA and GFP or 2) a non-targeting gRNA and mCherry. After 2 days, cells were selected with puromycin (2 µg/mL) for 5 days. After selection, for each validation gRNA, $5x10^4$ GFP-positive cells were seeded with $5x10^4$ mCherry-positive cells expressing the non-targeting gRNA. The percent of GFP- and mCherry-positive cells in each well was assayed 1, 7, and 14 days later using a FACSCanto II flow cytometer (BD Biosciences).

Acknowledgements

This work was supported by the US National Institutes of Health (NIH) grants to GEC, TER, and CAG (DP2OD008586, R01DA036865, UM1HG009428, R01HG010741, RM1HG011123, and U01AI146356), an Allen Distinguished Investigator Award from the Paul G. Allen Frontiers Group, the Open Philanthropy Project, and a National Science Foundation (NSF) grant (EFMA-1830957). T.S.K. was supported by a NIH Biotechnology Training Grant (T32GM008555). We thank Alejandro Berrio Escobar and Greg Wray for help with comparative chromatin analysis, and the DCI Flow Core for using the FACSCanto. We thank the Duke University School of Medicine for the use of the Sequencing and Genomic Technologies Shared Resource for sequencing support, and the Computational Solutions Shared Resource for providing the HARDAC computational resources.

Declaration of Interests

T.S.K., G.E.C., T.E.R., and C.A.G. are co-founders of Element Genomics. C.A.G. is a co-founder of Tune Therapeutics and Locus Biosciences and is an advisor to Tune Therapeutics, Sarepta Therapeutics, Levo Therapeutics, and Iveric Bio. T.S.K. is a co-founder and employee of Tune Therapeutics. T.S.K., S.S.A., N.I., T.E.R., G.E.C., and C.A.G. are inventors on patents or patent applications related to CRISPR epigenome editing and screening technologies.

List of Supplementary Tables

Table S1	original wgCERES discovery screen gRNA info and raw counts and DESeq2 results
Table S2	original wgCERES discovery screen bin2 raw counts and DESeq2 results
Table S3	original wgCERES discovery screen bin3 raw counts and DESeq2 results
Table S4	original wgCERES discovery screen DHS raw counts and DESeq2 results
Table S5	DHS summary results table

Table S6	K562 validation screen gRNA info and raw counts and DESeq2 results
Table S7	K562 validation screen bin2 raw counts and DESeq2 results
Table S8	K562 validation screen bin3 raw counts and DESeq2 results
Table S9	K562 validation screen DHS raw counts and DESeq2 results
Table S10	OCIAML2 validation screen gRNA info and raw counts and DESeq2 results
Table S11	OCIAML2 validation screen bin2 raw counts and DESeq2 results
Table S12	OCIAML2 validation screen bin3 raw counts and DESeq2 results
Table S13	OCIAML2 validation screen DHS raw counts and DESeq2 results
Table S14	23 individual validation gRNAs used for qRT-PCR, RNA-seq, and competition assays
Table S15	Taqman probes for qRT-PCR validation
Table S16	gRNA list used for single cell experiment
Table S17	scCERES MAST analysis results table

References

1. M. L. T. Nguyen, S. A. Jones, J. E. Prier, B. E. Russ, Transcriptional Enhancers in the Regulation of T Cell Differentiation. *Front. Immunol.* **6**, 462 (2015).

2. I. Sur, J. Taipale, The role of enhancers in cancer. *Nat. Rev. Cancer*. **16**, 483–493 (2016).

3. S. Ghisletti, I. Barozzi, F. Mietton, S. Polletti, F. De Santa, E. Venturini, L. Gregory, L. Lonie, A. Chew, C.-L. Wei, J. Ragoussis, G. Natoli, Identification and characterization of enhancers

controlling the inflammatory gene expression program in macrophages. *Immunity*. **32**, 317–328 (2010).

4. M. T. Maurano, R. Humbert, E. Rynes, R. E. Thurman, E. Haugen, H. Wang, A. P. Reynolds, R. Sandstrom, H. Qu, J. Brody, A. Shafer, F. Neri, K. Lee, T. Kutyavin, S. Stehling-Sun, A. K. Johnson, T. K. Canfield, E. Giste, M. Diegel, D. Bates, R. S. Hansen, S. Neph, P. J. Sabo, S. Heimfeld, A. Raubitschek, S. Ziegler, C. Cotsapas, N. Sotoodehnia, I. Glass, S. R. Sunyaev, R. Kaul, J. A. Stamatoyannopoulos, Systematic localization of common disease-associated variation in regulatory DNA. *Science*. **337**, 1190–1195 (2012).

5. R. E. Thurman, E. Rynes, R. Humbert, J. Vierstra, M. T. Maurano, E. Haugen, N. C. Sheffield, A. B. Stergachis, H. Wang, B. Vernot, K. Garg, S. John, R. Sandstrom, D. Bates, L. Boatman, T. K. Canfield, M. Diegel, D. Dunn, A. K. Ebersol, T. Frum, E. Giste, A. K. Johnson, E. M. Johnson, T. Kutyavin, B. Lajoie, B.-K. Lee, K. Lee, D. London, D. Lotakis, S. Neph, F. Neri, E. D. Nguyen, H. Qu, A. P. Reynolds, V. Roach, A. Safi, M. E. Sanchez, A. Sanyal, A. Shafer, J. M. Simon, L. Song, S. Vong, M. Weaver, Y. Yan, Z. Zhang, Z. Zhang, B. Lenhard, M. Tewari, M. O. Dorschner, R. S. Hansen, P. A. Navas, G. Stamatoyannopoulos, V. R. Iyer, J. D. Lieb, S. R. Sunyaev, J. M. Akey, P. J. Sabo, R. Kaul, T. S. Furey, J. Dekker, G. E. Crawford, J. A. Stamatoyannopoulos, The accessible chromatin landscape of the human genome. *Nature.* **489**, 75–82 (2012).

6. C. D. Arnold, D. Gerlach, C. Stelzer, Ł. M. Boryń, M. Rath, A. Stark, Genome-wide quantitative enhancer activity maps identified by STARR-seq. *Science*. **339**, 1074–1077 (2013).

7. A. Montalbano, M. C. Canver, N. E. Sanjana, High-Throughput Approaches to Pinpoint Function within the Noncoding Genome. *Mol. Cell*. **68**, 44–59 (2017).

8. B. Rauscher, F. Heigwer, M. Breinig, J. Winter, M. Boutros, GenomeCRISPR - a database for high-throughput CRISPR/Cas9 screens. *Nucleic Acids Res.* **45**, D679–D686 (2017).

9. W. F. Lenoir, T. L. Lim, T. Hart, PICKLES: the database of pooled in-vitro CRISPR knockout library essentiality screens. *Nucleic Acids Res.* **46**, D776–D780 (2018).

10. E. E. Schmidt, O. Pelz, S. Buhlmann, G. Kerr, T. Horn, M. Boutros, GenomeRNAi: a database for cell-based and in vivo RNAi phenotypes, 2013 update. *Nucleic Acids Res.* **41**, D1021–6 (2013).

11. T. Hart, M. Chandrashekhar, M. Aregger, Z. Steinhart, K. R. Brown, G. MacLeod, M. Mis, M. Zimmermann, A. Fradet-Turcotte, S. Sun, P. Mero, P. Dirks, S. Sidhu, F. P. Roth, O. S. Rissland, D. Durocher, S. Angers, J. Moffat, High-Resolution CRISPR Screens Reveal Fitness Genes and Genotype-Specific Cancer Liabilities. *Cell*. **163**, 1515–1526 (2015).

12. T. Wang, H. Yu, N. W. Hughes, B. Liu, A. Kendirli, K. Klein, W. W. Chen, E. S. Lander, D. M. Sabatini, Gene Essentiality Profiling Reveals Gene Networks and Synthetic Lethal Interactions with Oncogenic Ras. *Cell*. **168**, 890–903.e15 (2017).

13. T. Wang, J. J. Wei, D. M. Sabatini, E. S. Lander, Genetic screens in human cells using the CRISPR-Cas9 system. *Science*. **343**, 80–84 (2014).

14. O. Shalem, N. E. Sanjana, E. Hartenian, X. Shi, D. A. Scott, T. Mikkelson, D. Heckl, B. L. Ebert, D. E. Root, J. G. Doench, F. Zhang, Genome-scale CRISPR-Cas9 knockout screening in human cells. *Science*. **343**, 84–87 (2014).

15. M. C. Canver, E. C. Smith, F. Sher, L. Pinello, N. E. Sanjana, O. Shalem, D. D. Chen, P. G. Schupp, D. S. Vinjamur, S. P. Garcia, S. Luc, R. Kurita, Y. Nakamura, Y. Fujiwara, T. Maeda, G.-C. Yuan, F. Zhang, S. H. Orkin, D. E. Bauer, BCL11A enhancer dissection by Cas9-mediated in situ saturating mutagenesis. *Nature*. **527**, 192–197 (2015).

16. N. Rajagopal, S. Srinivasan, K. Kooshesh, Y. Guo, M. D. Edwards, B. Banerjee, T. Syed, B. J. M. Emons, D. K. Gifford, R. I. Sherwood, High-throughput mapping of regulatory DNA. *Nat. Biotechnol.* **34**, 167–174 (2016).

17. C. P. Fulco, M. Munschauer, R. Anyoha, G. Munson, S. R. Grossman, E. M. Perez, M. Kane, B. Cleary, E. S. Lander, J. M. Engreitz, Systematic mapping of functional enhancer–promoter connections with CRISPR interference. *Science*. **354** (2016), pp. 769–773.

18. N. E. Sanjana, J. Wright, K. Zheng, O. Shalem, P. Fontanillas, J. Joung, C. Cheng, A. Regev, F. Zhang, High-resolution interrogation of functional elements in the noncoding genome. *Science*. **353**, 1545–1549 (2016).

19. T. S. Klann, J. B. Black, M. Chellappan, A. Safi, L. Song, I. B. Hilton, G. E. Crawford, T. E. Reddy, C. A. Gersbach, CRISPR-Cas9 epigenome editing enables high-throughput screening for functional regulatory elements in the human genome. *Nat. Biotechnol.* **35**, 561–568 (2017).

20. M. Gasperini, A. J. Hill, J. L. McFaline-Figueroa, B. Martin, S. Kim, M. D. Zhang, D. Jackson, A. Leith, J. Schreiber, W. S. Noble, C. Trapnell, N. Ahituv, J. Shendure, A Genome-wide Framework for Mapping Gene Regulation via Cellular Genetic Screens. *Cell* (2018), doi:10.1016/j.cell.2018.11.029.

21. T. S. Klann, J. B. Black, C. A. Gersbach, CRISPR-based methods for high-throughput annotation of regulatory DNA. *Curr. Opin. Biotechnol.* **52**, 32–41 (2018).

22. ENCODE Project Consortium, An integrated encyclopedia of DNA elements in the human genome. *Nature*. **489**, 57–74 (2012).

23. ENCODE Project Consortium, J. E. Moore, M. J. Purcaro, H. E. Pratt, C. B. Epstein, N. Shoresh, J. Adrian, T. Kawli, C. A. Davis, A. Dobin, R. Kaul, J. Halow, E. L. Van Nostrand, P. Freese, D. U. Gorkin, Y. Shen, Y. He, M. Mackiewicz, F. Pauli-Behn, B. A. Williams, A. Mortazavi, C. A. Keller, X.-O. Zhang, S. I. Elhajjajy, J. Huey, D. E. Dickel, V. Snetkova, X. Wei, X. Wang, J. C. Rivera-Mulia, J. Rozowsky, J. Zhang, S. B. Chhetri, J. Zhang, A. Victorsen, K. P. White, A. Visel, G. W. Yeo, C. B. Burge, E. Lécuyer, D. M. Gilbert, J. Dekker, J. Rinn, E. M. Mendenhall, J. R. Ecker, M.

Kellis, R. J. Klein, W. S. Noble, A. Kundaje, R. Guigó, P. J. Farnham, J. M. Cherry, R. M. Myers, B. Ren, B. R. Graveley, M. B. Gerstein, L. A. Pennacchio, M. P. Snyder, B. E. Bernstein, B. Wold, R. C. Hardison, T. R. Gingeras, J. A. Stamatoyannopoulos, Z. Weng, Expanded encyclopaedias of DNA elements in the human and mouse genomes. *Nature*. **583**, 699–710 (2020).

24. L. A. Gilbert, M. H. Larson, L. Morsut, Z. Liu, G. A. Brar, S. E. Torres, N. Stern-Ginossar, O. Brandman, E. H. Whitehead, J. A. Doudna, W. A. Lim, J. S. Weissman, L. S. Qi, CRISPR-mediated modular RNA-guided regulation of transcription in eukaryotes. *Cell*. **154**, 442–451 (2013).

25. P. I. Thakore, A. M. D'Ippolito, L. Song, A. Safi, N. K. Shivakumar, A. M. Kabadi, T. E. Reddy, G. E. Crawford, C. A. Gersbach, Highly specific epigenome editing by CRISPR-Cas9 repressors for silencing of distal regulatory elements. *Nat. Methods*. **12**, 1143–1149 (2015).

26. D. Rhodes, H. J. Lipps, G-quadruplexes and their regulatory roles in biology. *Nucleic Acids Res.* **43**, 8627–8637 (2015).

27. L. T. Gray, A. C. Vallur, J. Eddy, N. Maizels, G quadruplexes are genomewide targets of transcriptional helicases XPB and XPD. *Nat. Chem. Biol.* **10**, 313–318 (2014).

28. C. P. Fulco, J. Nasser, T. R. Jones, G. Munson, D. T. Bergman, V. Subramanian, S. R. Grossman, R. Anyoha, B. R. Doughty, T. A. Patwardhan, T. H. Nguyen, M. Kane, E. M. Perez, N. C. Durand, C. A. Lareau, E. K. Stamenova, E. L. Aiden, E. S. Lander, J. M. Engreitz, Activity-by-contact model of enhancer–promoter regulation from thousands of CRISPR perturbations. *Nature Genetics*. **51** (2019), pp. 1664–1669.

29. J.-R. Landry, N. Bonadies, S. Kinston, K. Knezevic, N. K. Wilson, S. H. Oram, M. Janes, S. Piltz, M. Hammett, J. Carter, T. Hamilton, I. J. Donaldson, G. Lacaud, J. Frampton, G. Follows, V. Kouskoff, B. Göttgens, Expression of the leukemia oncogene Lmo2 is controlled by an array of tissue-specific elements dispersed over 100 kb and bound by Tal1/Lmo2, Ets, and Gata factors.

Blood. **113**, 5783–5792 (2009).

30. J. Ernst, M. Kellis, ChromHMM: automating chromatin-state discovery and characterization. *Nat. Methods*. **9**, 215–216 (2012).

31. G. Jimenez, S. D. Griffiths, A. M. Ford, M. F. Greaves, T. Enver, Activation of the beta-globin locus control region precedes commitment to the erythroid lineage. *Proceedings of the National Academy of Sciences*. **89**, 10618–10622 (1992).

32. S. Sengupta, R. E. George, Super-Enhancer-Driven Transcriptional Dependencies in Cancer. *Trends Cancer Res.* **3**, 269–281 (2017).

33. C. Zhu, Q. Chen, Z. Xie, J. Ai, L. Tong, J. Ding, M. Geng, The role of histone deacetylase 7 (HDAC7) in cancer cell proliferation: regulation on c-Myc. *J. Mol. Med.* . **89**, 279–289 (2011).

34. Y. Chen, X. Liu, F. Zhang, S. Liao, X. He, D. Zhuo, H. Huang, Y. Wu, Vitamin D receptor suppresses proliferation and metastasis in renal cell carcinoma cell lines via regulating the expression of the epithelial Ca2+ channel TRPV5. *PLoS One*. **13**, e0195844 (2018).

35. M. A. Horlbeck, L. A. Gilbert, J. E. Villalta, B. Adamson, R. A. Pak, Y. Chen, A. P. Fields, C. Y. Park, J. E. Corn, M. Kampmann, J. S. Weissman, Compact and highly active next-generation libraries for CRISPR-mediated gene repression and activation. *eLife*. **5** (2016), , doi:10.7554/elife.19760.

36. T. Wang, K. Birsoy, N. W. Hughes, K. M. Krupczak, Y. Post, J. J. Wei, E. S. Lander, D. M. Sabatini, Identification and characterization of essential genes in the human genome. *Science*. **350**, 1096–1101 (2015).

37. R. M. Meyers, J. G. Bryan, J. M. McFarland, B. A. Weir, A. E. Sizemore, H. Xu, N. V. Dharia, P. G. Montgomery, G. S. Cowley, S. Pantel, A. Goodale, Y. Lee, L. D. Ali, G. Jiang, R. Lubonja, W. F.

Harrington, M. Strickland, T. Wu, D. C. Hawes, V. A. Zhivich, M. R. Wyatt, Z. Kalani, J. J. Chang, M. Okamoto, K. Stegmaier, T. R. Golub, J. S. Boehm, F. Vazquez, D. E. Root, W. C. Hahn, A. Tsherniak, Computational correction of copy number effect improves specificity of CRISPR–Cas9 essentiality screens in cancer cells. *Nat. Genet.* **49**, 1779–1784 (2017).

38. J. M. Dempster, J. Rossen, M. Kazachkova, J. Pan, G. Kugener, D. E. Root, A. Tsherniak, Extracting Biological Insights from the Project Achilles Genome-Scale CRISPR Screens in Cancer Cell Lines. *Cold Spring Harbor Laboratory* (2019), p. 720243.

39. J. Wu, Y.-C. Zhang, W.-H. Suo, X.-B. Liu, W.-W. Shen, H. Tian, G.-H. Fu, Induction of anion exchanger-1 translation and its opposite roles in the carcinogenesis of gastric cancer cells and differentiation of K562 cells. *Oncogene*. **29**, 1987–1996 (2010).

40. B.-J. He, L. Liao, Z.-F. Deng, Y.-F. Tao, Y.-C. Xu, F.-Q. Lin, Molecular Genetic Mechanisms of Hereditary Spherocytosis: Current Perspectives. *Acta Haematol.* **139**, 60–66 (2018).

41. B. Adamson, T. M. Norman, M. Jost, M. Y. Cho, J. K. Nuñez, Y. Chen, J. E. Villalta, L. A. Gilbert, M. A. Horlbeck, M. Y. Hein, R. A. Pak, A. N. Gray, C. A. Gross, A. Dixit, O. Parnas, A. Regev, J. S. Weissman, A Multiplexed Single-Cell CRISPR Screening Platform Enables Systematic Dissection of the Unfolded Protein Response. *Cell*. **167**, 1867–1882.e21 (2016).

42. A. Dixit, O. Parnas, B. Li, J. Chen, C. P. Fulco, L. Jerby-Arnon, N. D. Marjanovic, D. Dionne, T. Burks, R. Raychowdhury, B. Adamson, T. M. Norman, E. S. Lander, J. S. Weissman, N. Friedman, A. Regev, Perturb-Seq: Dissecting Molecular Circuits with Scalable Single-Cell RNA Profiling of Pooled Genetic Screens. *Cell*. **167**, 1853–1866.e17 (2016).

43. P. Datlinger, A. F. Rendeiro, C. Schmidl, T. Krausgruber, P. Traxler, J. Klughammer, L. C. Schuster, A. Kuchler, D. Alpar, C. Bock, Pooled CRISPR screening with single-cell transcriptome readout. *Nat. Methods*. **14**, 297–301 (2017).

44. S. Xie, J. Duan, B. Li, P. Zhou, G. C. Hon, Multiplexed Engineering and Analysis of Combinatorial Enhancer Activity in Single Cells. *Mol. Cell.* **66**, 285–299.e5 (2017).

45. V. Tam, N. Patel, M. Turcotte, Y. Bossé, G. Paré, D. Meyre, Benefits and limitations of genome-wide association studies. *Nat. Rev. Genet.* **20**, 467–484 (2019).

46. E. Khurana, Y. Fu, D. Chakravarty, F. Demichelis, M. A. Rubin, M. Gerstein, Role of non-coding sequence variants in cancer. *Nat. Rev. Genet.* **17**, 93–108 (2016).

47. C. M. Laumont, K. Vincent, L. Hesnard, É. Audemard, É. Bonneil, J.-P. Laverdure, P. Gendron, M. Courcelles, M.-P. Hardy, C. Côté, C. Durette, C. St-Pierre, M. Benhammadi, J. Lanoix, S. Vobecky, E. Haddad, S. Lemieux, P. Thibault, C. Perreault, Noncoding regions are the main source of targetable tumor-specific antigens. *Sci. Transl. Med.* **10** (2018), doi:10.1126/scitranslmed.aau5516.

48. G. Orlando, P. J. Law, A. J. Cornish, S. E. Dobbins, D. Chubb, P. Broderick, K. Litchfield, F. Hariri, T. Pastinen, C. S. Osborne, J. Taipale, R. S. Houlston, Promoter capture Hi-C-based identification of recurrent noncoding mutations in colorectal cancer. *Nat. Genet.* **50**, 1375–1380 (2018).

49. N. K. Hayward, J. S. Wilmott, N. Waddell, P. A. Johansson, M. A. Field, K. Nones, A.-M. Patch, H. Kakavand, L. B. Alexandrov, H. Burke, V. Jakrot, S. Kazakoff, O. Holmes, C. Leonard, R. Sabarinathan, L. Mularoni, S. Wood, Q. Xu, N. Waddell, V. Tembe, G. M. Pupo, R. De Paoli-Iseppi, R. E. Vilain, P. Shang, L. M. S. Lau, R. A. Dagg, S.-J. Schramm, A. Pritchard, K. Dutton-Regester, F. Newell, A. Fitzgerald, C. A. Shang, S. M. Grimmond, H. A. Pickett, J. Y. Yang, J. R. Stretch, A. Behren, R. F. Kefford, P. Hersey, G. V. Long, J. Cebon, M. Shackleton, A. J. Spillane, R. P. M. Saw, N. López-Bigas, J. V. Pearson, J. F. Thompson, R. A. Scolyer, G. J. Mann, Whole-genome landscapes of major melanoma subtypes. *Nature.* **545**, 175–180 (2017).

50. E. Rheinbay, M. M. Nielsen, F. Abascal, J. A. Wala, O. Shapira, G. Tiao, H. Hornshøj, J. M. Hess, R. I. Juul, Z. Lin, L. Feuerbach, R. Sabarinathan, T. Madsen, J. Kim, L. Mularoni, S. Shuai, A.

Lanzós, C. Herrmann, Y. E. Maruvka, C. Shen, S. B. Amin, P. Bandopadhayay, J. Bertl, K. A. Boroevich, J. Busanovich, J. Carlevaro-Fita, D. Chakravarty, C. W. Y. Chan, D. Craft, P. Dhingra, K. Diamanti, N. A. Fonseca, A. Gonzalez-Perez, Q. Guo, M. P. Hamilton, N. J. Haradhvala, C. Hong, K. Isaev, T. A. Johnson, M. Juul, A. Kahles, A. Kahraman, Y. Kim, J. Komorowski, K. Kumar, S. Kumar, D. Lee, K.-V. Lehmann, Y. Li, E. M. Liu, L. Lochovsky, K. Park, O. Pich, N. D. Roberts, G. Saksena, S. E. Schumacher, N. Sidiropoulos, L. Sieverling, N. Sinnott-Armstrong, C. Stewart, D. Tamborero, J. M. C. Tubio, H. M. Umer, L. Uusküla-Reimand, C. Wadelius, L. Wadi, X. Yao, C.-Z. Zhang, J. Zhang, J. E. Haber, A. Hobolth, M. Imielinski, M. Kellis, M. S. Lawrence, C. von Mering, H. Nakagawa, B. J. Raphael, M. A. Rubin, C. Sander, L. D. Stein, J. M. Stuart, T. Tsunoda, D. A. Wheeler, R. Johnson, J. Reimand, M. Gerstein, E. Khurana, P. J. Campbell, N. López-Bigas, PCAWG Drivers and Functional Interpretation Working Group, PCAWG Structural Variation Working Group, J. Weischenfeldt, R. Beroukhim, I. Martincorena, J. S. Pedersen, G. Getz, PCAWG Consortium, Analyses of non-coding somatic drivers in 2,658 cancer whole genomes. *Nature*. **578**, 102–111 (2020).

51. M. R. Corces, J. M. Granja, S. Shams, B. H. Louie, J. A. Seoane, W. Zhou, T. C. Silva, C. Groeneveld, C. K. Wong, S. W. Cho, A. T. Satpathy, M. R. Mumbach, K. A. Hoadley, A. G. Robertson, N. C. Sheffield, I. Felau, M. A. A. Castro, B. P. Berman, L. M. Staudt, J. C. Zenklusen, P. W. Laird, C. Curtis, Cancer Genome Atlas Analysis Network, W. J. Greenleaf, H. Y. Chang, The chromatin accessibility landscape of primary human cancers. *Science*. **362** (2018), doi:10.1126/science.aav1898.

52. L. A. Gilbert, M. A. Horlbeck, B. Adamson, J. E. Villalta, Y. Chen, E. H. Whitehead, C. Guimaraes, B. Panning, H. L. Ploegh, M. C. Bassik, L. S. Qi, M. Kampmann, J. S. Weissman, Genome-Scale CRISPR-Mediated Control of Gene Repression and Activation. *Cell*. **159**, 647–661 (2014).

53. M. S. Lawrence, P. Stojanov, C. H. Mermel, J. T. Robinson, L. A. Garraway, T. R. Golub, M.

Meyerson, S. B. Gabriel, E. S. Lander, G. Getz, Discovery and saturation analysis of cancer genes across 21 tumour types. *Nature.* **505**, 495–501 (2014).

54. R. Kempfer, A. Pombo, Methods for mapping 3D chromosome architecture. *Nat. Rev. Genet.* **21**, 207–226 (2020).

55. G. Eraslan, Ž. Avsec, J. Gagneur, F. J. Theis, Deep learning: new computational modelling techniques for genomics. *Nat. Rev. Genet.* **20**, 389–403 (2019).

56. S. Konermann, M. D. Brigham, A. E. Trevino, J. Joung, O. O. Abudayyeh, C. Barcena, P. D. Hsu, N. Habib, J. S. Gootenberg, H. Nishimasu, O. Nureki, F. Zhang, Genome-scale transcriptional activation by an engineered CRISPR-Cas9 complex. *Nature*. **517**, 583–588 (2015).

57. C. L. Frank, D. Manandhar, R. Gordân, G. E. Crawford, HDAC inhibitors cause site-specific chromatin remodeling at PU.1-bound enhancers in K562 cells. *Epigenetics Chromatin*. **9**, 15 (2016).

58. M. Szyf, Epigenetics, DNA methylation, and chromatin modifying drugs. *Annu. Rev. Pharmacol. Toxicol.* **49**, 243–263 (2009).

59. S. Hacein-Bey-Abina, C. Von Kalle, M. Schmidt, M. P. McCormack, N. Wulffraat, P. Leboulch, A. Lim, C. S. Osborne, R. Pawliuk, E. Morillon, R. Sorensen, A. Forster, P. Fraser, J. I. Cohen, G. de Saint Basile, I. Alexander, U. Wintergerst, T. Frebourg, A. Aurias, D. Stoppa-Lyonnet, S. Romana, I. Radford-Weiss, F. Gross, F. Valensi, E. Delabesse, E. Macintyre, F. Sigaux, J. Soulier, L. E. Leiva, M. Wissler, C. Prinz, T. H. Rabbitts, F. Le Deist, A. Fischer, M. Cavazzana-Calvo, LMO2-associated clonal T cell proliferation in two patients after gene therapy for SCID-X1. *Science*. **302**, 415–419 (2003).

60. B. Chen, L. A. Gilbert, B. A. Cimini, J. Schnitzbauer, W. Zhang, G.-W. Li, J. Park, E. H. Blackburn, J. S. Weissman, L. S. Qi, B. Huang, Dynamic imaging of genomic loci in living human cells by an optimized CRISPR/Cas system. *Cell*. **155**, 1479–1491 (2013).

61. A. O'Brien, T. L. Bailey, GT-Scan: identifying unique genomic targets. *Bioinformatics*. **30**, 2673–2675 (2014).

62. P. Salmon, D. Trono, *Curr. Protoc. Neurosci.*, in press.

63. S. Chen, N. E. Sanjana, K. Zheng, O. Shalem, K. Lee, X. Shi, D. A. Scott, J. Song, J. Q. Pan, R. Weissleder, H. Lee, F. Zhang, P. A. Sharp, Genome-wide CRISPR screen in a mouse model of tumor growth and metastasis. *Cell*. **160**, 1246–1260 (2015).

64. B. Langmead, S. L. Salzberg, Fast gapped-read alignment with Bowtie 2. *Nat. Methods*. **9**, 357–359 (2012).

65. A. M. Bolger, M. Lohse, B. Usadel, Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics*. **30**, 2114–2120 (2014).

66. A. Dobin, C. A. Davis, F. Schlesinger, J. Drenkow, C. Zaleski, S. Jha, P. Batut, M. Chaisson, T. R. Gingeras, STAR: ultrafast universal RNA-seq aligner. *Bioinformatics*. **29**, 15–21 (2013).

67. J. Harrow, A. Frankish, J. M. Gonzalez, E. Tapanari, M. Diekhans, F. Kokocinski, B. L. Aken, D. Barrell, A. Zadissa, S. Searle, I. Barnes, A. Bignell, V. Boychenko, T. Hunt, M. Kay, G. Mukherjee, J. Rajan, G. Despacio-Reyes, G. Saunders, C. Steward, R. Harte, M. Lin, C. Howald, A. Tanzer, T. Derrien, J. Chrast, N. Walters, S. Balasubramanian, B. Pei, M. Tress, J. M. Rodriguez, I. Ezkurdia, J. van Baren, M. Brent, D. Haussler, M. Kellis, A. Valencia, A. Reymond, M. Gerstein, R. Guigó, T. J. Hubbard, GENCODE: the reference human genome annotation for The ENCODE Project. *Genome Res.* **22**, 1760–1774 (2012).

68. Y. Liao, G. K. Smyth, W. Shi, The Subread aligner: fast, accurate and scalable read mapping by seed-and-vote. *Nucleic Acids Res.* **41**, e108 (2013).

69. M. I. Love, W. Huber, S. Anders, Moderated estimation of fold change and dispersion for RNA-seq

data with DESeq2. *Genome Biol.* **15**, 550 (2014).

70. M. Stephens, False discovery rates: a new deal. *Biostatistics*. **18**, 275–294 (2017).

71. B. Li, C. N. Dewey, RSEM: accurate transcript quantification from RNA-Seq data with or without a reference genome. *BMC Bioinformatics*. **12**, 323 (2011).

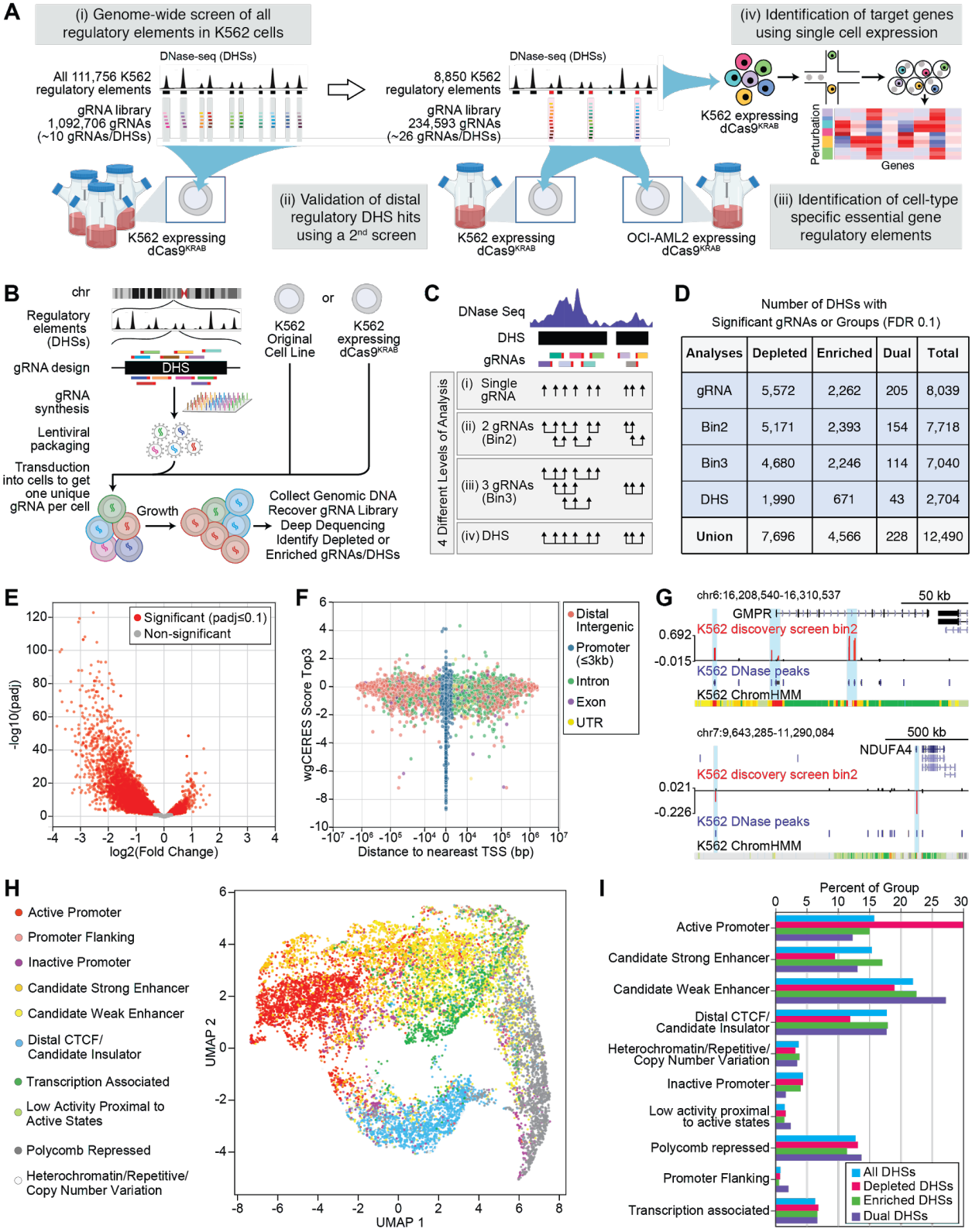

Fig. 1. wgCERES identifies thousands of regulatory elements essential for K562 proliferation. (A) Overall schematic of (i) discovery wgCERES screen, (ii) secondary validation screen of regulatory elements, (iii) cell-type specificity, and (iv) single-cell (scCERES) readout to connect cell fitness-associated regulatory elements to target genes. (B) A schematic of wgCERES approach. gRNAs are

designed to all DHSs in the K562 cell line and synthesized as a pool for lentiviral delivery. K562 cells either constitutively expressing or not expressing dCas9^{KRAB} are treated with the lentiviral gRNA library at a low MOI and cultured for 14 population doublings. Genomic DNA is then harvested and the gRNA abundance is quantified by Illumina sequencing. (C) A schematic describing four levels of gRNA grouping analyses, including individual gRNA, sliding windows of 2 or 3 gRNAs, and averaging all gRNAs within a DHS. (D) Summary of DHS hits identified by significant changes to individual gRNAs or grouped gRNAs. (E) Volcano plot of significance of gRNA changes relative to $\log_2$(fold-change). (F) Distribution of significant gRNAs relative to transcriptional start sites of nearest genes. (G) Representative examples of significant distal DHS hits (blue boxes) that also have a significant DHS hit at a TSS of the nearest gene. ChromHMM tracks indicate promoters (red), putative enhancers (yellow), and polycomb repressed regions (gray). (H) UMAP dimensionality reduction plot showing different ChromHMM chromatin state informed classes of significant (FDR < 0.1) DHS hits. Histone modifications as well as several epigenetic modifying proteins were included as input for dimensionality reduction. (I) Relative abundances of significantly depleted or enriched gRNAs relative to ChromHMM classes of genome annotations**.**

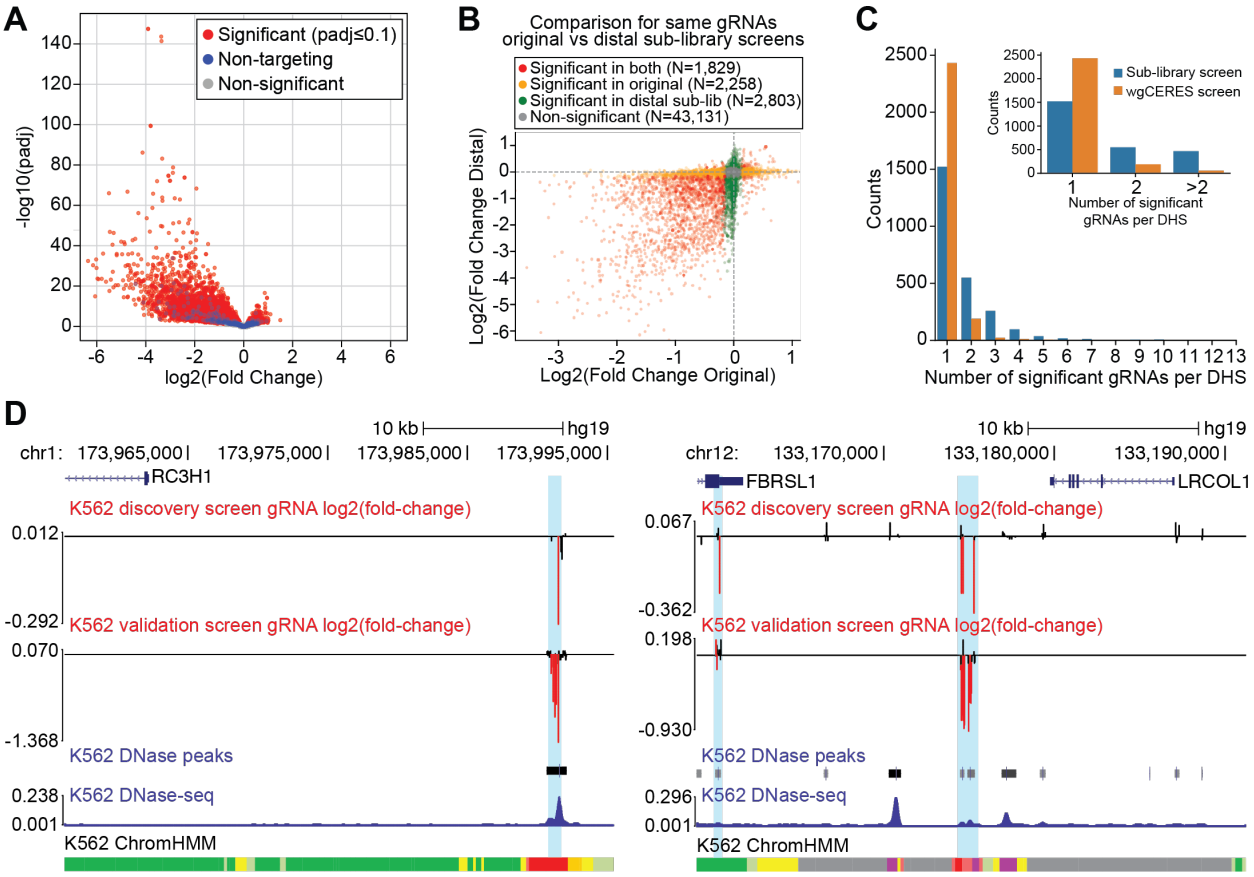

Fig. 2. Sub-library screen targeting all distal significant DHSs from the first wgCERES screen. (A) Volcano plot showing 9,833 significantly depleted or enriched gRNAs. Depleted gRNAs were more abundant and show larger effect size than enriched gRNAs. (B) Comparison of $\log_2$(fold-change) between the first wgCERES screen and the second sub-library screen. Only gRNAs that overlap both studies are represented (N=50,021 common gRNAs). (C) Distribution of number of significant gRNAs per DHS between the first wgCERES screen and the second sub-library screen. Inset shows combined counts for more than 1 gRNA per DHS. (D) Screenshot examples of DHS hits (blue boxes) that displayed a smaller number significant gRNA hits (red lines) in the wgCERES discovery screen, and additional significant gRNAs in the more densely tiled distal sub-library validation screen.

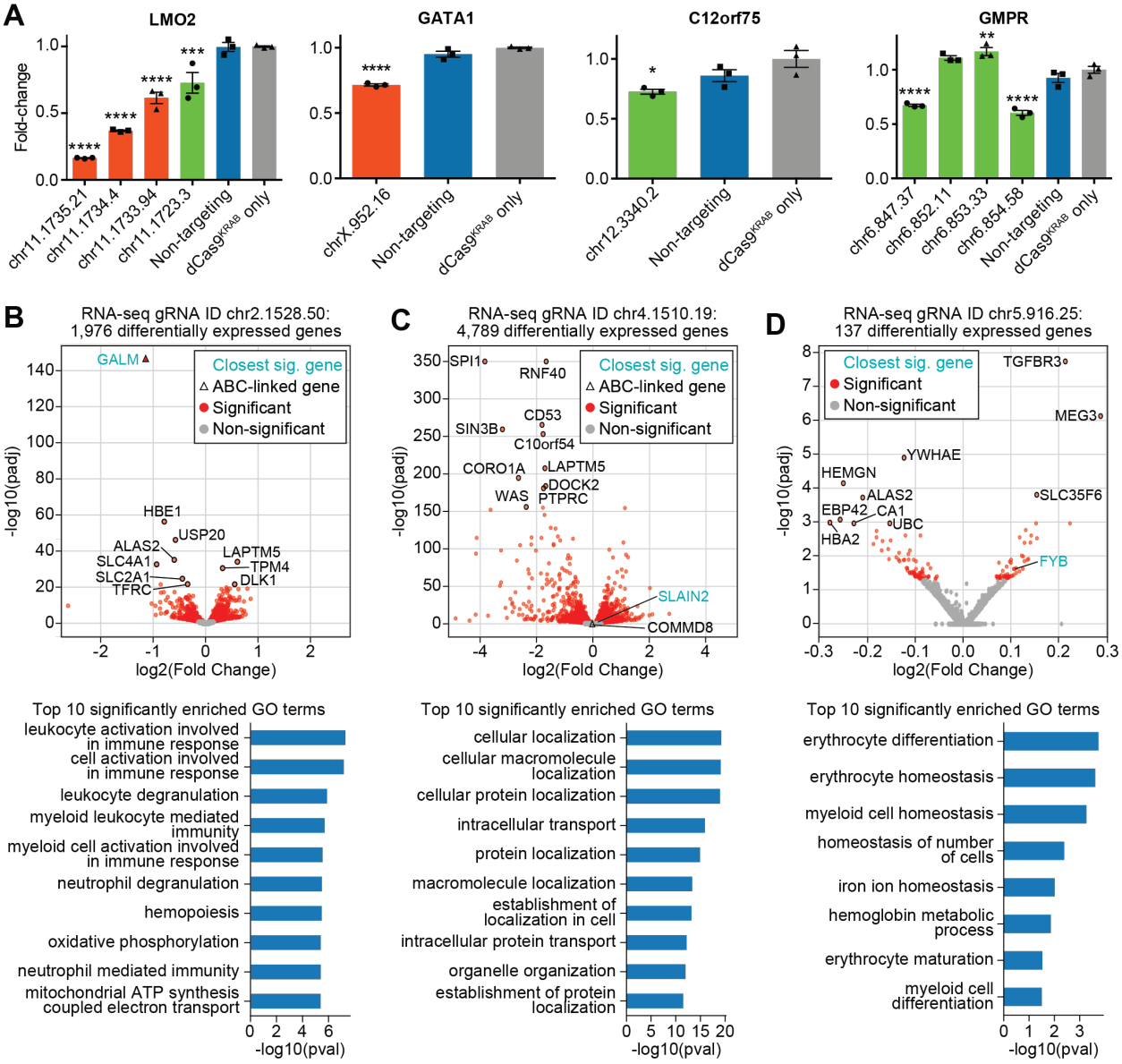

Fig. 3. Validation of wgCERES-seq hits with individual gRNAs. (A) Individual gRNAs were tested in K562 cells constitutively expressing dCas9^{KRAB}, and mRNA changes were detected through qRT-PCR. Red and green bars indicate validation gRNAs that were depleted or enriched, respectively, in the screen. Blue bars indicate non-targeting gRNA control. Gray bars indicate cells only expressing dCas9^{KRAB}, without gRNA transduction. Significance was determined by one-way analysis of variance followed by Dunnett's test (n = 3 biological replicates, mean ± s.e.m.): ****P < 0.0001, ***P < 0.001, **P < 0.01, *P < 0.05 versus dCas9^{KRAB} only control. All fold enrichments are relative to the transduction of a control gRNA and normalized to *TBP*. (B) Validation of a gRNA by RNA-seq shows that the largest effect is the nearest gene (GALM) and the "ABC linked gene" indicates GALM as the predicted target gene (Fulco et al. 2016). (C) Validation of a gRNA by RNA-seq that shows the nearest gene (COMMD8) is not differentially expressed, and therefore is not the likely target gene. (D) Validation of a gRNA that has no predicted ABC target gene, but displays many differentially expressed genes. Genes with significant differences in gene expression are shown in red (padj ≤ 0.05). Gene ontologies show top 10 enriched categories for each RNA-seq analysis.

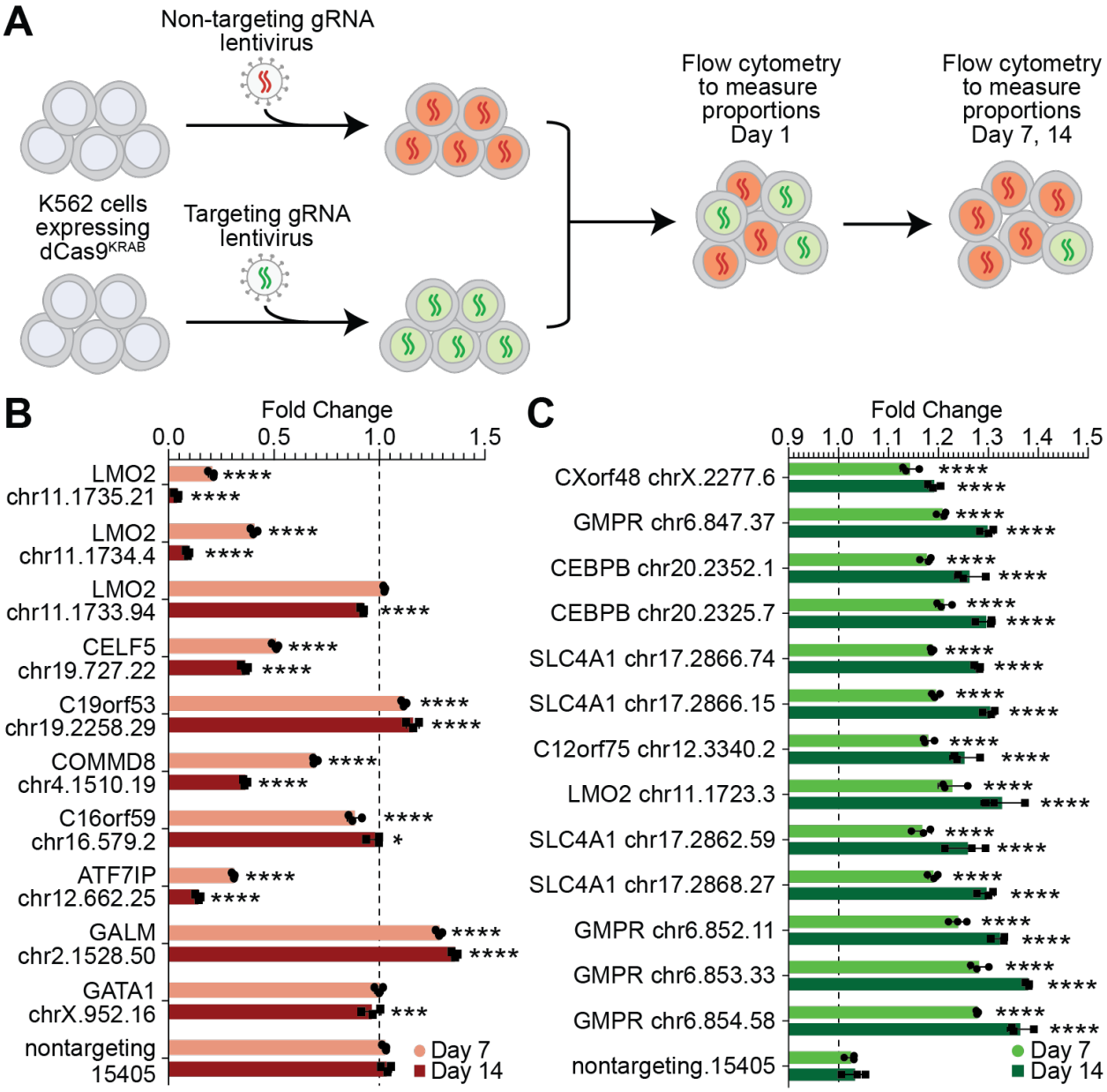

Fig. 4. Functional validation of gRNAs in competition growth assay. (A) Individual gRNAs were delivered to cells in a GFP expressing vector and co-seeded with equal numbers of cells transduced with a non-targeting gRNA in an mCherry expressing vector. Proportion of cells were assayed at day 1 post-seeding, and day 7 or day 14 post-seeding to determine the fold-change in the proportion of GFP vs mCherry positive cells. (B) Individual validations for gRNAs that were depleted in the second sub-library screen. (C) Individual validations for gRNAs that were enriched in the sub-library screen (n = 3 biological replicates, mean ± s.e.m.). Gene names represent putative target genes identified for each gRNA hit by ABC method (Fulco et al. 2016; Fulco et al. 2019). Statistics indicate significance by two-way ANOVA with Dunnett's multiple comparisons test relative to non-targeting gRNA at day 7 or day 14: *P < 0.05, ***P < 0.001, ****P < 0.0001.

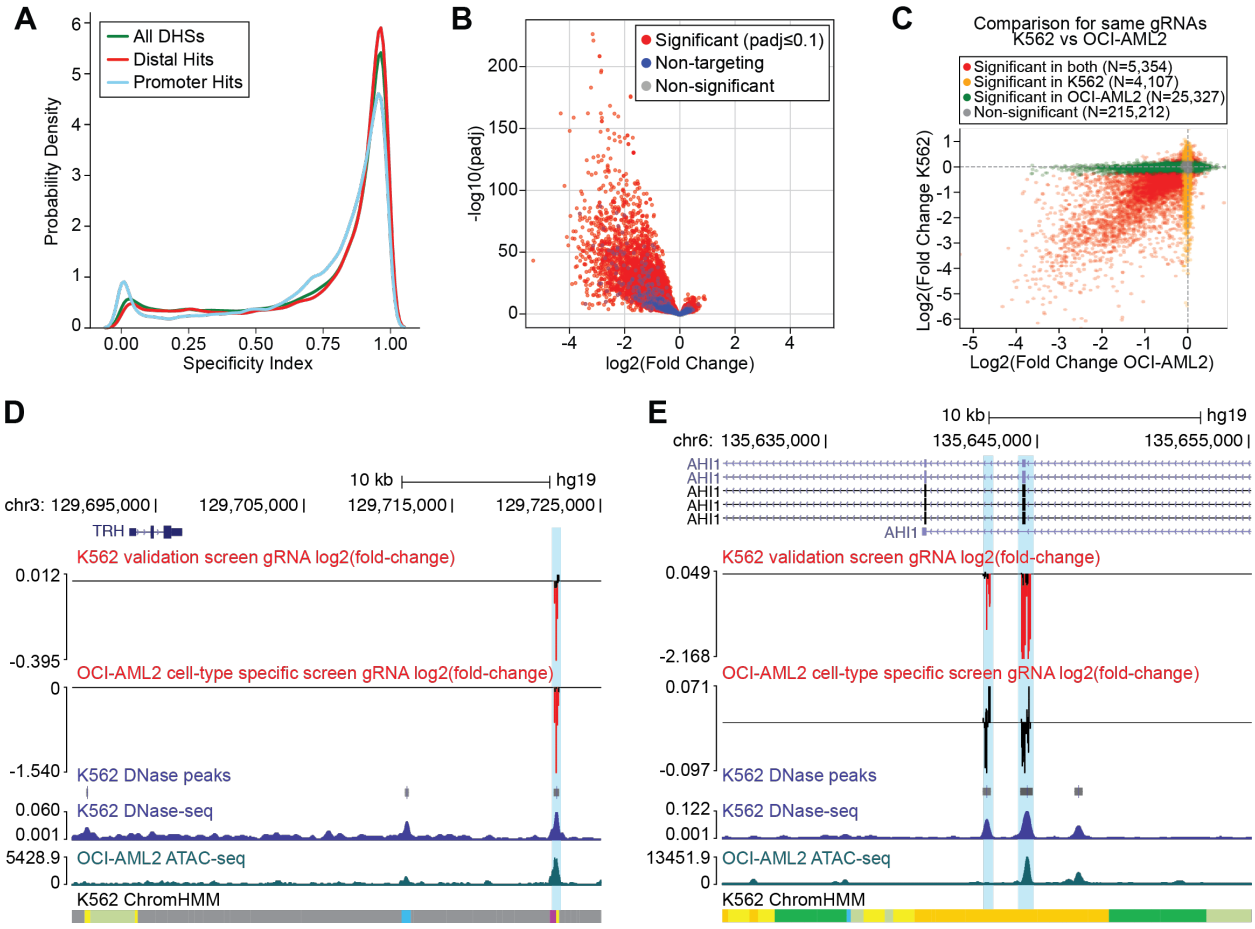

Fig. 5. Essential DHS hits show cell-type specificity. (A) Characterization of promoter and distal DHS hits relative to chromatin accessibility from 53 diverse cell types. Specificity index of 1 indicates DHS that is unique to K562, while specificity index of 0 represents ubiquitous DHS site across all cell types. All DHS sites identified in K562 cells are shown as a comparison. (B) Volcano plot showing 31,193 significant gRNAs either depleted or enriched following 14 population doublings of OCI-AML2 cells. Depleted gRNAs were more abundant and show larger effect size than enriched gRNAs. Red points indicate significant gRNAs (FDR < 0.1), blue points indicate non-targeting control gRNAs, and gray points indicate non-significant gRNAs. (C) Comparison of the sub-library screen in K562 cells versus OCI-AML2 cells. Log_2(fold-change) is plotted for every gRNA in each screen. Red points indicate gRNAs significant in both cell types. Green points indicate gRNAs significant only in the OCI-AML2 cells, while orange points indicate gRNAs only significant in K562 cells. Gray points indicate gRNAs not significant in either cell type. Legend shows the number of gRNAs that are significant in either direction. (D) Representative example of gRNA hits that are significantly depleted in both K562 and OCI-AML2 cell types. Blue box highlights the region of interest and red gRNAs indicate significant depletion. Note this region is marked by open chromatin in both cell types. (E) Representative example of gRNAs that are significantly depleted in K562 cells, but not significant (black gRNAs, note Y-axis difference) in OCI-AML2 cells. Blue boxes highlight regions of interest, and the left region represents a DHS that is uniquely accessible in K562 cells.

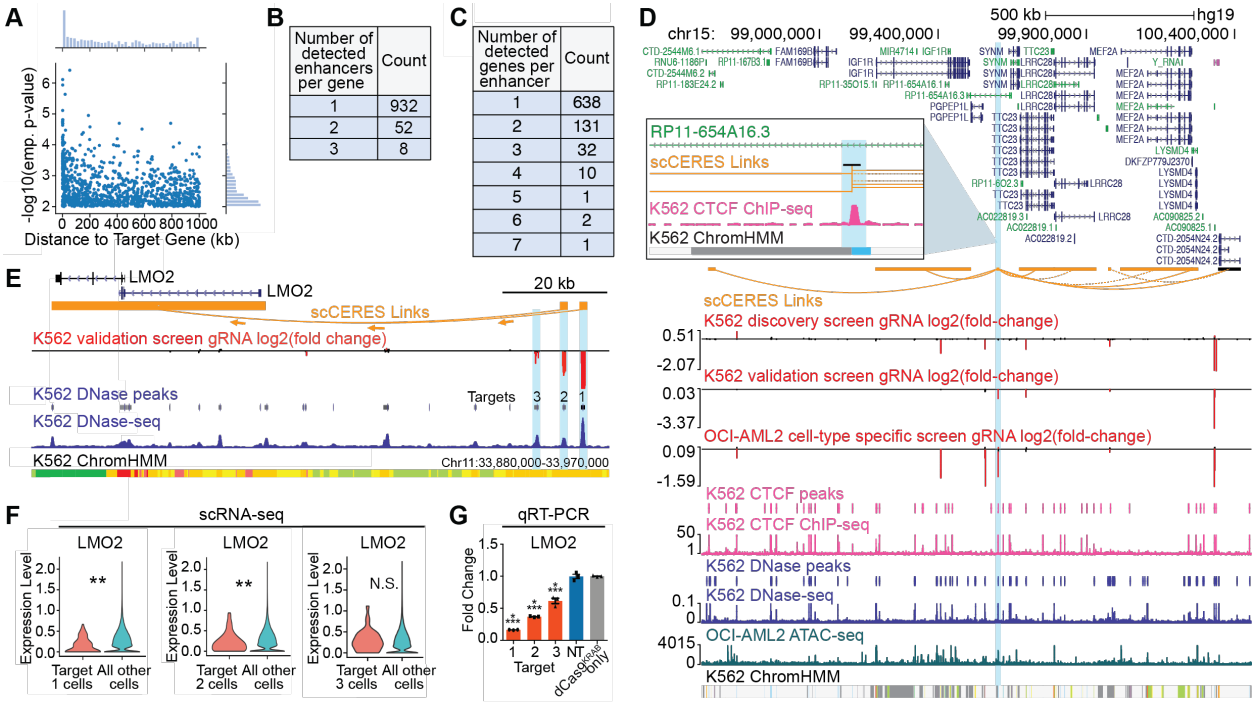

Fig. 6. scCERES to identify the target gene(s) for each perturbed gene regulatory element using single cell RNA-seq. (A) Relationship between distance of regulatory element-gene link and significance. Perturbations closer to the transcriptional start site of genes tend to be more significant overall. (B) Number of regulatory elements per individual gene detected. (C) Number of genes affected by individual regulatory elements. (D) DHS hit with seven gene connections listed in panel "C". Blue box represents DHS targeted for silencing, as well as gRNA $\log_2$(fold-change) depletion through CERES, and chromatin accessibility. The seven target genes are shown as yellow connectors. Inset shows this DHS hit directly overlaps a CTCF binding site (also marked by CTCF ChromHMM chromatin state). (E) Genome browser tracks of the LMO2 locus including links of enhancers to genes, and CERES depletion of three downstream regions listed at Target 1, 2, and 3. (F) Single cell expression analysis shows significant depletion of LMO2 expression in cells containing gRNAs for Target 1 and Target 2, but not Target 3. Asterisks indicate empirical p-values < 0.01. (G) LMO2 mRNA fold-change by qRT-PCR in response to individual gRNA perturbations for LMO2 Target 1, 2, and 3. ****p-value < 0.0001 versus control, one-way analysis of variance followed by Dunnett's test (n = 3 biological replicates; mean ± s.e.m.).