Sensitive dissection of a genomic regulatory landscape using bulk and targeted single-cell activation

Dubravka Vučićević[§1], Che-Wei Hsu[1,2], Lorena Sofia Lopez Zepeda[1*], Martin Burkert[*1], Antje Hirsekorn[1], Ilija Bilić[1,3], Nicolai Kastelić[4], Markus Landthaler[4], Scott Allen Lacadie[§,x,1] , Uwe Ohler[§,x,1,5]

[1] Computational Regulatory Genomics, Berlin Institute for Medical Systems Biology of the Max Delbrück Center for Molecular Medicine in the Helmholtz Association, 10115 Berlin, Germany

[2] present address: Department of Biology, Duke University, Durham, NC 27708, USA and Howard Hughes Medical Institute, Duke University, Durham, NC 27708, USA

[3] present address: BioNTech Cell & Gene Therapies GmbH, 55131 Mainz, Germany

[4] RNA Biology and Posttranscriptional Regulation, Berlin Institute for Medical Systems Biology (BIMSB), Max Delbrück Center for Molecular Medicine, 10115 Berlin, Germany

[5] Department of Biology, Humboldt Universität Berlin,10117 Berlin, Germany

[*x] These authors contributed equally

[x] Lead contact

[§] Co-corresponding authors

vucicevic.dubravka@gmail.com

uwe.ohler@mdc-berlin.de

scott.lacadie@mdc-berlin.de

Abstract

Transcriptional enhancers are non-coding DNA elements that regulate gene transcription in a temporal and tissue-specific manner. Despite advances in computational and experimental methods, identifying enhancers and their target genes essential for specific biological processes remains challenging. Determining target genes for enhancers is also complex and often relies on indirect, low-resolution, and/or assumptive methodologies. To identify and functionally perturb enhancers at their endogenous sites without altering their sequence, we performed a pooled tiling CRISPR activation (CRISPRa) screen surrounding PHOX2B, a master regulator of neuronal cell fate and a key player in neuroblastoma development. This screen allowed the *de novo* identification of CRISPRa responsive elements (CaREs) that alter cellular growth within the 2 Mb genomic region. To determine CaRE target genes, we developed TESLA-seq (TargEted SingLe cell Activation), which combines CRISPRa screening with targeted single-cell RNA-sequencing and enables the parallel readout of the effect of hundreds of enhancers on all genes in the locus. While most TESLA-revealed CaRE-gene relationships involved neuroblastoma-related regulatory elements already active in the system, we found many CaREs and target connections normally active only in other tissue types or with no previous evidence and induced out of context by CRISPRa. This highlights the power of TESLA-seq to reveal gene regulatory networks, including edges active outside of a given experimental system.

Keywords: CRISPR screen, CRISPRa, targeted sc-RNA-seq, TESLA-seq, PHOX2B, SHISA3, enhancers, transcription, single-cell genomics.

Highlights

- Systematically perturbed regulatory landscape in a 2 Mb genomic region surrounding PHOX2B to identify hundreds of CRISPRa-responsive elements that affect cellular growth
- Developed TESLA-seq as a principled molecular approach to find gene targets of dozens of candidate regulatory elements
- Validated interactions between identified regulatory elements and target genes and characterized their genomic features
- Integrated a compendium of epigenomic datasets to identify regulatory relationships induced out of context or with no previous evidence

Introduction

One of the critical questions in molecular biology is how gene expression is regulated in a temporal and tissue-specific manner in both health and disease. The genome-wide mapping of open chromatin and histone modifications indicative of transcriptional regulatory mechanisms and their states have provided large compendia of cis-regulatory regions (CREs). However, identifying the target genes of regulatory elements is challenging: CREs do not necessarily regulate their closest gene, may regulate several genes, and may only show a functional effect in combination with other CREs. Approaches to determining CRE-target relationships include measurements of 3D proximity between CREs and candidate targets, CRE-gene co-activity correlations across many cell types or states, and CRE perturbation with transcriptional activity

readouts (Shlyueva et al., 2014, Kim and Shiekhattar, 2015, Field and Adelman, 2020, Preissl et al., 2023, Kim and Wysocka, 2023). The first two have the potential advantage of being high throughput and the disadvantage of being purely correlative without direct functional evidence, whereas traditional perturbation approaches provide such direct evidence but are comparatively low throughput. Despite numerous computational and experimental methods for predicting genetic cis-regulatory elements (CREs) based on sequence and chromatin features, identifying functional CREs and their targets has therefore remained challenging (Kim and Shiekhattar, 2015, Field and Adelman, 2020, Panigrahi and O'Malley, 2021, Preissl et al., 2023, Kim and Wysocka, 2023).

For a long time, functional characterization of CREs, such as enhancers, has been performed outside their genomic context using reporter assays (Shlyueva et al., 2014, Kim and Shiekhattar, 2015, Fulco et al., 2019, Field and Adelman, 2020, Preissl et al., 2023, Kim and Wysocka, 2023). The development of CRISPR/Cas9 technologies now allows the examination of CREs at their native locus. Specifically, engineered fusion proteins can endogenously activate (CRISPRa) or inhibit (CRISPRi) CREs, and we can perform large-scale assays to study thousands of CREs in a single experiment using pooled CRISPR/Cas9 screens (Bock et al., 2022, Schraivogel et al., 2023).

Transcriptome-wide measurements of RNA in up to hundreds of thousands of single-cells in a single study have revolutionized cell-type quantification from heterogeneous samples. Single-cell CRISPR screening approaches hold great promise for overcoming the low-throughput disadvantage of CRE perturbation experiments (Adamson et al., 2016, Dixit et al., 2016, Xie et al., 2017, Datlinger et al., 2017,

Gasperini et al., 2019, Alda-Catalinas et al., 2020, Replogle et al., 2020, Schraivogel et al., 2020, Morris et al., 2023), yet there are some limitations to these approaches. First, scRNA-seq screens have so far largely employed wild-type Cas9 and the dCas9-KRAB transcriptional repressor construct to perturb enhancers (Adamson et al., 2016, Dixit et al., 2016, Xie et al., 2017, Datlinger et al., 2017, Gasperini et al., 2019, Alda-Catalinas et al., 2020, Replogle et al., 2020, Schraivogel et al., 2020, Morris et al., 2023). Although they can provide valuable insight in enhancer biology, these Cas9 constructs are less effective at perturbing distal CREs than they are for proximal CREs (Yeo et al., 2018, Alerasool et al., 2020, Nuñez et al., 2021). Second, selecting CREs based on pre-existing features such as chromatin accessibility and the presence of histone marks (Gasperini et al., 2019) introduces potential biases from the quality and completeness of such annotations and our interpretation of them. Third, scRNA-seq approaches are costly and extremely sparse, leading to unreliable assessment of differential gene expression, especially for lowly expressed genes (Adamson et al., 2016, Dixit et al., 2016, Xie et al., 2017, Datlinger et al., 2017, Gasperini et al., 2019, Morris et al., 2023). Single-cell perturbations of CREs do not require full transcriptome quantification due to known spatial restrictions between CREs and their targets and, therefore, benefit greatly from targeted sequencing via the substantial reduction of drop-out rates for genes of interest (Adamson et al., 2016, Dixit et al., 2016, Xie et al., 2017, Datlinger et al., 2017, Gasperini et al., 2019, Alda-Catalinas et al., 2020, Replogle et al., 2020, Schraivogel et al., 2020).

To overcome current limitations, we propose a two-step strategy for identifying REs. In the first step, we assay tens of thousands of potential REs in a tiling CRISPR(a)

screen for their effect on the cellular phenotype. In the second step, we examine hundreds of elements that affect the phenotype upon activation on the dozens of genes in the vicinity with targeted scRNA-seq. With this two-step strategy, we combine the scalability and unbiased nature of tiling CRISPR screens with the high-resolution targeted scRNA-seq readout.

To showcase this approach, we use it to comprehensively dissect the transcriptional regulation of an entire genomic locus, the 2 Mb regulatory landscape of PHOX2B, a master regulator of autonomous nervous system development and a key player in the development of a variety of disorders such as the childhood cancer neuroblastoma (Bachetti and Ceccherini, 2020). Through an unbiased tiling screen using a robustly activating CRISPRa construct, we identified 619 CRISPRa responsive elements (CaREs) that influence cellular survival. To identify the targets of CaREs and study the molecular effect of their activation on genes across a larger genomic context, we developed a TargEted SingLe cell Activation screen (TESLA-seq). TESLA-seq combines CRISPRa screening with targeted single-cell RNA sequencing (scRNA-seq), enabling the detection of affected genes for thousands of perturbations in a single experiment. We applied TESLA-seq on the hits from the phenotypic screen to quantify their impact on the expression of transcripts within a 6 Mb space surrounding PHOX2B.

Many TESLA-seq-identified CaRE-gene interactions exhibit characteristics in agreement with current understanding of CREs, but previous evidence from neuroblastoma cell lines alone cannot fully explain the revealed functional regulatory relationships. We combined the TESLA-seq results with available epigenomic maps from 800 tissues to assign systems in which they are likely active and to define CaRE-

gene pairs for which there was no previous regulatory evidence. We validate that functional transactivation can originate from elements irrespective of whether they exhibit typical regulatory traits prior to activation.

Results

CRISPR activation screening reveals candidate elements in the PHOX2B regulatory landscape

PHOX2B is a master regulator of neurogenesis and a key player in the development of neuroblastoma, whose expression level is tightly associated with the growth rates of neuroblastoma cell lines (Ke et al., 2015, Naftali et al., 2016, Ponzoni et al., 2022, Windels et al., 2024). It has been suggested to be under the control of a large cluster of enhancers (super-enhancer) (Boeva et al., 2017), and it is located in the vicinity of other genes that play roles in a variety of disorders (Bachetti and Ceccherini, 2020), increasing the challenge of enhancer-target prediction. Therefore, we sought to perform an exhaustive search for regulatory elements affecting growth or viability in the 2 Mb genomic space surrounding PHOX2B using CRISPR activation (Figure 1A), without requiring any previous annotations or characteristics of CREs.

We tested several CRISPRa constructs for robust activation of known enhancers (Chavez et al., 2016) (Figure S1A) and selected dCas9-VPR for further experiments. We next tested whether we could successfully activate and repress PHOX2B expression and detected robust activation by targeting dCas9-VPR and repression by dCas9-KRAB at its promoter in the neuroblastoma-derived cell line SHSY-5Y (Figure

S1B). Both actions reduced cellular viability, aligning with the literature (Ke et al., 2015, Naftali et al., 2016) (Figure 1B, FigureS 1C).

We conducted a dense tiling viability screen in the PHOX2B locus, designing 2-3 gRNAs per 100 bp genomic bin within the 2 Mb space (46,722 gRNAs total). These were delivered via lentivirus to SH-SY5Y-VPR cells and collected at multiple time points in two replicates. Deep sequencing identified gRNAs that were depleted or enriched, indicating regions influencing cellular growth/viability upon activation (Figure 1A). This approach identified 758 (1.7%) depleted and 27 (0.1%) enriched gRNAs that respectively repress or promote cellular growth/viability upon dCas9-VPR activation (FDR <0.05) (Figure 1C, Table S3). Positive control gRNAs targeting the PHOX2B promoter were depleted, confirming that overexpression reduces viability (Figure S1D).

To define regulatory elements from individual hits, we applied a sliding window approach to identify neighboring significant gRNAs, which were required to have the same effect on the cellular phenotype to be merged into individual functional regions. In this way, we *de novo* identified CRISPR activation responsive elements (CaREs; Figure 1D) across the whole locus, with 536 CaREs showing gRNA depletion (genomic regions whose activation leads to a reduction of cellular viability) and 83 CaREs showing gRNA enrichment (genomic regions whose activation leads to an increase in cellular growth/proliferation; FDR <0.05; Figure 1E; Table S3). The average width of a CaRE is 447bp (Figure 1F), and most CaREs are promoter distal (Figure 1G). Although the hits are equally distributed across the whole 2 Mb space, the highest scoring CaREs are in the proximity of the PHOX2B promoter. They overlap a super-enhancer predicted to regulate PHOX2B expression in neuroblastoma ((Boeva et al., 2017); Figure 1H).

Besides PHOX2B, the promoters of several other genes coincide with CaREs that led to a reduction or activation of cellular growth and survival: NSUN7, APBB2, UCHL1, TMEM33, DCAF4L1, SLC30A9, BEND4, RP11-63A11.1, SHISA3, RP11-109E24.2, RP11-109E24.1. Some of them are known to affect cellular growth/proliferation: SHISA3 is a known tumor suppressor (Chen et al., 2014, Shahzad et al., 2020, Zhang et al., 2024); APBB2 plays a role in Alzheimer's disease (Li et al., 2005, Golanska et al., 2013) and cell cycle (Bruni et al., 2002, Zhou et al., 2021); UCHL1 promotes cellular proliferation in cancer (Kwan et al., 2020, Mondal et al., 2022). Therefore, CaREs are not limited to effects on PHOX2B; our tiling screen identified CaREs targeting any of these genes surrounding PHOX2B and, potentially, genes outside of the 2 Mb screening window.

Next, we sought to characterize the features of the CaREs we identified. Since CREs tend to reside in open chromatin regions flanked with post-translational modified histones and bound by transcription factors (Shlyueva et al., 2014, Kim and Shiekhattar, 2015, Kim and Wysocka, 2023), we generated ATAC-seq libraries and compiled available ChIP-seq-data of neuroblastoma transcriptional core regulators in a variety of cell lines (Henrich et al., 2016, Durbin et al., 2018, Wang et al., 2019, Boeva et al., 2017, Zhang et al., 2020). 28% of CaREs intersect with these typical enhancer features, and generalized linear models (see methods) could predict significant CaREs based on accessible chromatin, histone modifications H3K27ac and H3K4me1 (Figure 1I), as well as binding of key neuronal transcription factors MYCN, HAND2, ISL1, ASCL1, GATA3 and GATA2 (Figure 1J).

In summary, using CRISPRa screening in a neuroblastoma cell line in the 2 Mb genomic space on chromosome 4 surrounding the PHOX2B gene, we have identified 619 CaREs that play a role in cellular growth or survival. Some CaREs display typical enhancer features such as accessibility and H3K27ac modification, but a substantial fraction do not, highlighting the importance of unbiased tiling screens to understand the regulation of gene expression in greater depth. We performed an equivalent CRISPRi screen with the same 46 722 gRNAs, but here, we focus on the CRSIPRa screen and provide additional data in the Supplement (Figure S1E,F; Table S4).

Figure 1

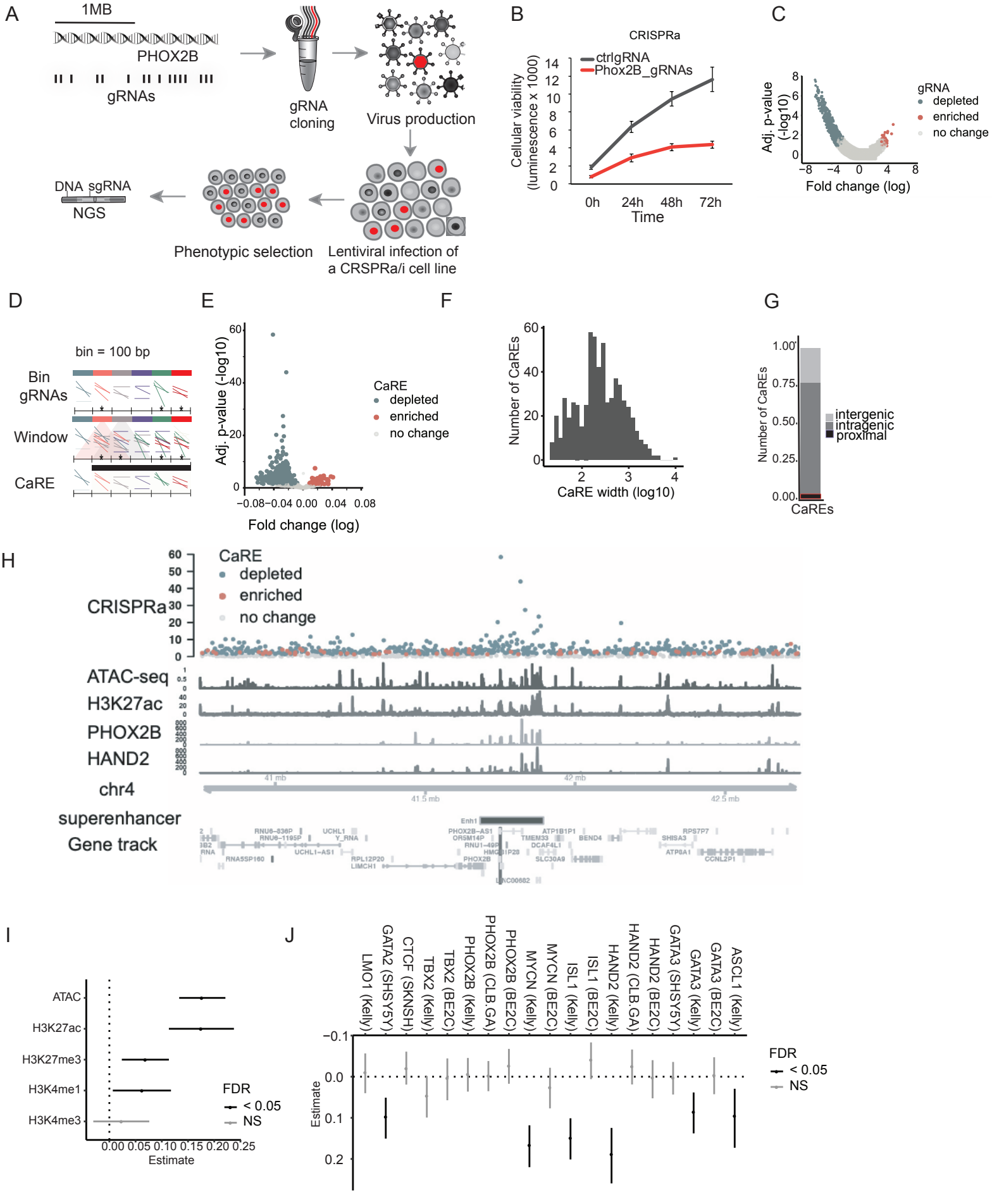


Figure 1. Bulk phenotypic CRISPR activation screen reveals genetic regulatory elements in the PHOX2B genomic landscape. A) Overview of the CRISPRa screen. **B)** MTT viability assay with control or gRNAs targeting the promoter of PHOX2B in an SH-SY5Y-VPR (CRISPRa) cell line. **C)** Volcano plot showing the log-fold change of gRNA representation between the first and last time-point of the experiment. Significantly enriched/depleted gRNAs (FDR<0.05) are shown in red and blue, respectively. *P*-values shown are adjusted for multiple testing (FDR). **D)** Graphical representation of grouping strategy for analysis. Asterisks denote significance. **E)** The volcano plot shows the results of CaRE analysis, each dot corresponds to a CaRE. The x-axis shows the slope calculated by our MLM. Significantly enriched/depleted CaREs (FDR<0.05) are shown in red and blue, respectively. **F)** Bar plot representing the number of CaREs with indicated CaRE width on a log10 scale. **G)** Number of significant intragenic, intergenic, and promoter-proximal CaREs. **H)** CaREs signal around the PHOX2B locus (+/- 1 MB). From bottom to top: annotation for PHOX2B and its location within the genome, ChIP-seq signal for HAND2, PHOX2B, H3K27ac, and ATAC-seq signal in SH-SY5Y cell line. At the top is the score and direction (blue for depletion, red for enrichment) of CaREs. *P*-values shown are adjusted for multiple testing (FDR). **I)** Coefficient estimates from generalized linear regression, predicting significant CaREs by ATAC-seq and histone modification ChiP-seq signal in neuroblastoma cell line SH-SY5Y. **J)** Coefficient estimates from individual generalized linear regressions, predicting significant CaREs by neuroblastoma core regulatory circuit transcription factor ChIP-seq signals. Whiskers in (I,J) indicate 95% confidence interval of estimate, FDR in (I,J) derived from p-values of Z-statistic; FDR: false discovery rate; NS: Not significant. See also Figure S1, Table S3 and Table S4.

TESLA-seq sensitively and reproducibly detects functional regulatory relationships

To determine the precise regulatory targets of the CaREs identified in the phenotypic screen, we developed a **t**arg**e**ted **s**ing**l**e-cell **a**ctivation screen followed by sequencing (TESLA-seq). TESLA-seq allows the detection of differential expression of selected transcripts for thousands of perturbations in a single experiment by combining CRISPRa screening with a targeted scRNA-seq readout. To perform TESLA-seq, we

selected 1046 gRNA corresponding to the top 222 CaREs from our phenotypic CRISPR viability screen and 52 gRNA controls (Figure 1; see methods). We synthesized and cloned them into a CROP-seq vector (Figure 2A) that enables a direct readout of the gRNA from a polyadenylated transcript in scRNA-seq experiments (Datlinger et al., 2017). Next, we infected SHSY-5Y-VPR cells at a low MOI to ensure that each cell would only receive one gRNA. Four days post-transduction, we assayed 20,000 cells in each of two experiments on the BD RhapsodyTM Express Single-Cell Analysis system. In this microwell-based system, the polyadenylated transcriptome of each cell is captured using barcoded magnetic beads and subsequently amplified by two sets of primers (universal and gene-specific) for each transcript in semi-nested multiplex PCRs. This is followed by sequencing upon the addition of Illumina indices (Figure 2A; (Mair et al., 2020)). Upon retrieval of barcoded complementary DNA, we enriched for the gRNA transcripts and transcripts of interest, in our case, 151 transcript isoforms annotated for all 78 annotated genes in an expanded 6 Mb space surrounding PHOX2B to detect possible effects on genes outside of the tiled screening window (see methods). In our experiments, 93% of the reads are mapped to the enriched targets.

We applied stringent quality filtering, after which 16,535 cells were retained from two BD Rhapsody runs (see Methods, Figure S2A). The scRNA-seq readouts of the two TESLA-seq experiments are highly correlated (Pearson correlation R^2= 0.886; Figure 2B), indicating excellent reproducibility. We capture gRNA in 90% of cells with a minimum of 2 molecules/cell, and 59 out of 78 genes within the 6 Mb space surrounding PHOX2B are captured. Non-captured genes are expressed at relatively low levels in a bulk chromatin RNA-seq assay (Figure 2C). Each cell has an average of 650 gene

molecules and 158 gRNA counts, as well as a median of 1 unique gRNA as aimed for by the experimental design (Figure S2B). On average, each gRNA or CaRE is assayed in 17 or 68 cells, respectively (see Methods) (Figure 2D). These results demonstrate that TESLA-seq efficiently captures both the gRNA transcripts and the transcripts of interest with high sensitivity and reproducibility.

We utilize the targeted single-cell transcriptome replicates upon activation of each CaRE for differential gene expression analysis by comparing the expression of transcripts in the cells carrying the gRNA of interest to all other cells containing either control or distal gRNAs (at least 4 Kb from the examined gRNA). As proof of principle, we focus on the expression of PHOX2B, BEND4 and SHISA3. We observe that cells in which the gRNAs fall within their promoter exhibit higher expression than the cells carrying control gRNAs (Figure 2E, Figure S2C), demonstrating that TESLA-seq can be used to sensitively detect differential gene expression.

Next, we examined the effect of all CaREs and identified 60 CaREs that cause a significant expression change of 33 of the captured transcripts, for a total of 92 CaRE-gene pairs (adjusted p-value < 0.05) (Figure 2F-H; Table S5). On average, each CaRE causes a differential expression of 1.5 genes, with 22 of them affecting more than one gene (Figure 2I). In turn, each examined gene is significantly affected by an average of 2.8 assayed CaREs (Figure 2H, Figure S2D). Of note, we only examined a subset of CaREs in the 2 Mb screening window, and their effects on genes located within a 6 Mb window. This leaves the possibility of additional CaREs regulating assayed genes within this window and beyond. Notably, most of the TESLA-assayed CaRE target genes whose promoters were hits in the bulk growth screen are targets in the TESLA-seq, and

these interactions are especially strong and significant (Figure 2H - highlighted genes, Figure 2J). This is expected since the TESLA-seq guides were chosen based on their phenotypic effect in the tiling bulk screen (Figure 1), and these results can, therefore be considered as validation for CaREs affecting growth/viability. Overall, we observe comparable results when we perform similar analyses at the gRNA rather than the CaRE level (Figure S2E-G; Table S6). Some of the CaREs coinciding with promoters also alter the expression of other genes, potentially indicating that these promoters can also act as enhancers (Diao et al., 2017, Fulco et al., 2016, Panigrahi and O'Malley, 2021). In summary, our findings demonstrate that TESLA-seq can sensitively detect potential transcriptional regulatory elements and their targets.

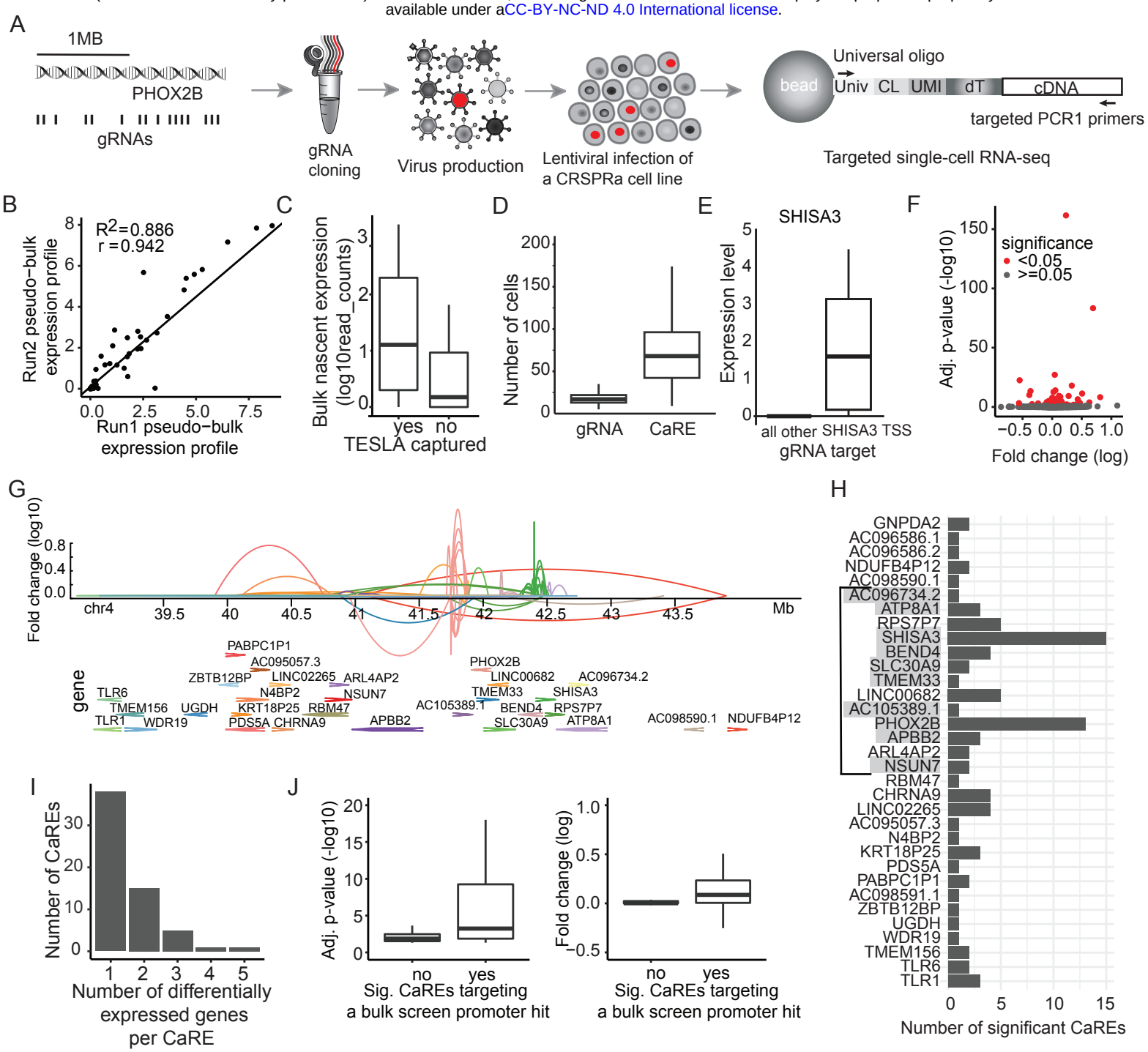


Figure 2. TESLA-seq captures gene expression change sensitively, and reproducibly. A) Schematic representation of the TESLA-seq. **B)** Pseudobulk expression profiles of two TESLA-seq experiments. **C)** Expression of genes detected by bulk chromatin RNA-seq that were captured by TESLA-seq (yes) versus the ones not captured by TESLA-seq (no). **D)** Number of cells assayed in TESLA-seq per gRNA and CaRE. **E)** TESLA-seq normalized gene expression level, comparing cells having gRNA targeting SHISA3 promoter versus all other cells. **F)** Adjusted p-values (-log10) relative to fold change (log2) of each CaRE-gene regulatory pair. **G)** Genome browser snapshot representing genomic distance and regulatory relationship between a CaRE and a gene determined by TESLA-seq. Each link represents an effect of a CaRE on the gene. Color reflects the target gene. **H)** Number of significant CaREs that cause a differential gene expression of an indicated number of genes. Highlighted genes are the ones whose promoters were significant hits in the bulk screen. Genes are displayed in the order that matches their order in the genome in decreasing genomic coordinates. The line indicates the screening window of the phenotypic CRSIPRa screen from Figure 1. **I)** Histogram showing a number of genes differentially expressed by a perturbation of an indicated number of CaREs. **J)** Boxplots of adjusted p-values (-log10) - left and fold change (log2) - right of CaRE-gene regulation pairs. Grouped by whether the CaRE belongs to bulk screen promoter hit or not. See also Figure S2 and Table S5-6.

TESLA-seq CaRE-target interactions exhibit properties consistent with current models of CREs

To gain further insight into the identified CaREs and their targets, we explored the distance between them in both linear and 3D genomic space. Significant CaRE-gene relationships are closer in the linear genome (median distance of 284kb; Figure 3A, Figure S3A) than non-significant CaRE-gene relationships. In 71% (65) of significant cases, there is at least one gene between a CaRE and its target (Figure 3B), with an average of 7.7 skipped genes. Of these, the strongest hits tend to be closer to their CaREs with up to 5 skipped genes (Figure S3B). However, the expression of these skipped genes is unaffected by the CaREs (Figure 3C), demonstrating the precise

targeting of the CRISPRa construct and the ability of TESLA-seq to characterize distal regulatory relationships.

To assess the proximity of CaREs to their target genes in 3D space prior to CRISPR perturbation, we captured all 3D promoter interactions in the 6Mb PHOX2B genomic space in the SH-SY5Y cell line using Nuclear Capture C (nuCaptureC; (Downes et al., 2022); Figure 3D, PHOX2B promoter interaction example in Figure S3C). NuCaptureC interaction scores for significant CaRE-gene pairs are higher than for non-significant CaRE-gene pairs (Figure 3E). Significant CaRE-gene pairs displaying prior 3D interaction result in greater changes in expression levels compared to significant CaRE-gene pairs that are not in prior 3D proximity (Figure 3D; Figure 3F, right panel). Furthermore, TESLA-seq contains data for over 1000 CaRE-gene relationships that show prior interaction according to nuCaptureC data but do not affect the interacting gene's expression when targeted by CRISPRa (Figure 3F, right panel). These observations suggest that prior 3D proximity allows for stronger transactivation but is neither required nor sufficient.

Our observations are further supported by published databases of regulatory associations determined either via correlations of histone modifications across cell types (Boix et al., 2021); Figure S3D,F) or from Hi-C catalogs (Nasser et al., 2021); Figure S3E,F). Of note, only a minority of identified CaRE-gene pairs have reliable evidence in these 3C- or correlation-based regulatory associations, highlighting a potential advantage of our functional relationships over other indirect approaches.

CaRE-gene pairs for which the target gene promoters displayed high signals for ATAC and H3K4me3 showed greater changes in target gene expression upon CaRE

activation, compared to pairs with low levels of these chromatin features at the target promoter (Figure 3F, Figure S3G). We trained a simple yet highly predictive random forest model for significant CaRE-gene pairs based on the genomic features mentioned above. We find that linear distance, the presence of active histone marks (H3K4me1 and H3K27ac), and accessibility are the most predictive features for regulatory relationships, with perfect separation between the top and bottom 50 CaRE-gene pairs (see methods, Figure 3G).

In summary, many TESLA-seq identified CaRE-target interactions exhibit characteristics consistent with current models of CREs, but as expected for the gain-of-function approach, evidence from SH-SY5Y prior to CRISPR activation cannot fully explain the revealed functional regulatory relationships.

Figure 3

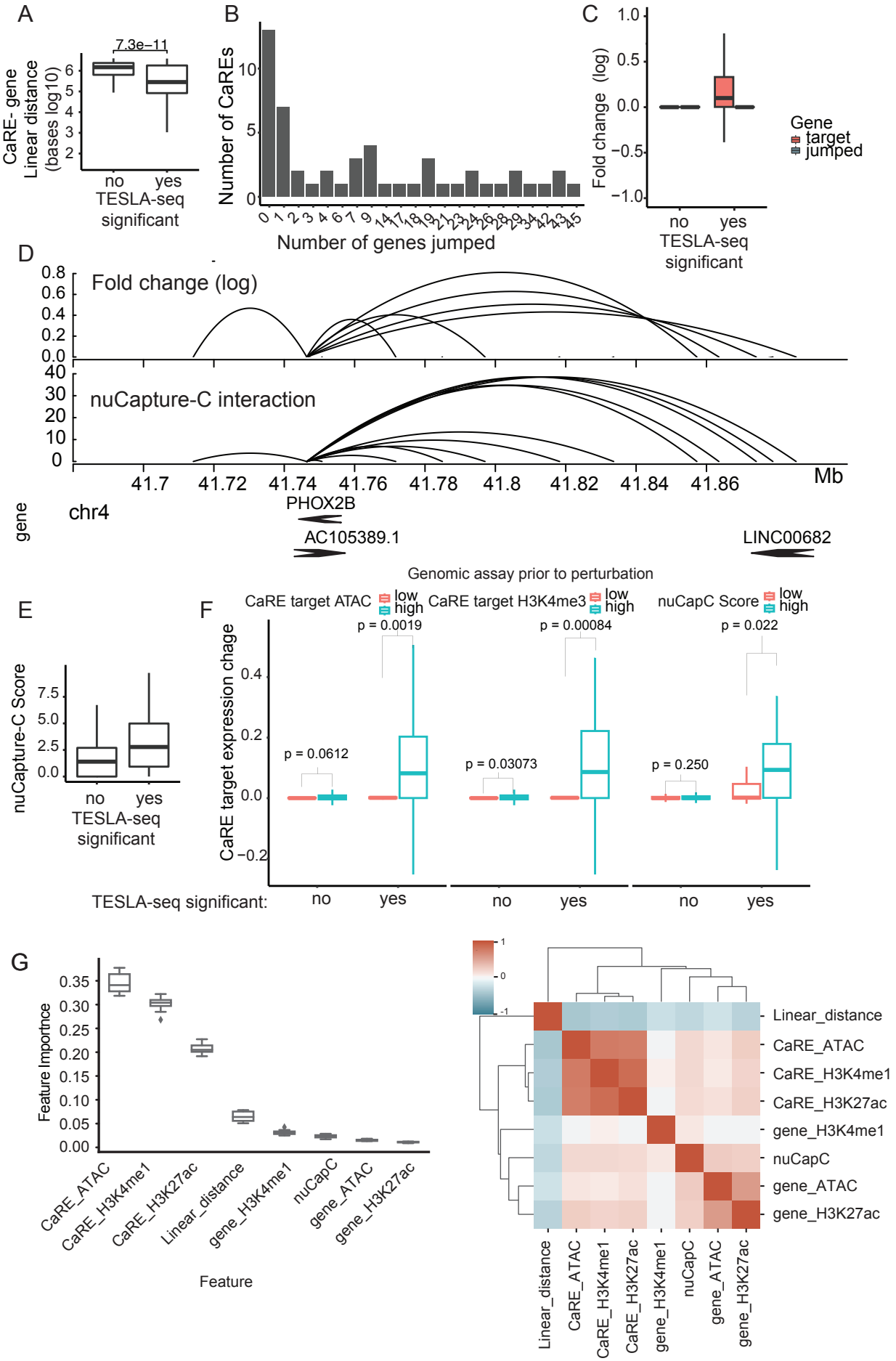


Figure 3. Properties of CaREs and their targets identified by TESLA-seq. **A)** Linear distance between CaREs and their targets (bases in a log10 scale) stratified by their significance in the TESLA- screen (p-adjusted value < 0.05 is yes). **B)** Histogram showing the number of genes located between a CaRE and a gene in the genome (jumped genes) for each CaRE-gene regulatory pair. **C)** Comparison of the fold change (log2) of targeted genes and jumped genes between significant and non-significant TESLA-seq CaRE-gene regulation pairs. **D)** Genome browser snapshot of CaREs affecting PHOX2B expression determined by TESLA-seq (top) and PHOX2B 3D interactions determined by nuCaptureC (bottom). **E)** Comparison of nuCaptureC score between significant and non-significant TESLA-seq CaRE-gene regulation pairs. **F)** Comparison of the CaRE target gene expression (average fold change at log 2 scale) between TESLA-seq significant and non-significant CaRE-gene regulation pairs that have a low or high: ATAC signal at the promoter of the target gene (left), H3K4me3 signal at the promoter of the target gene (middle) and nuCapC score (right). **G)** Features that contribute the most to the predictability of a random forest model. The model predicts whether a CaRE-gene regulation pair is in top 50 or bottom 50. Panel on the right: hierarchical clustering of features based on correlation. See also Figure S3 and Table S5-6.

TESLA-seq identifies regulatory network edges normally active inside and outside of the experimental system

Utilizing CRISPRa allows for the discovery of not only CREs active in the assayed cell line but, in principle, also for the discovery of genetic regulatory elements that may be active in any biological context (Wu et al., 2023). Therefore, we examined EpiMap data, in which a compendium comprising 10,000 epigenomic maps across 800 samples is used to define chromatin states, accessible regions, promoters, enhancers, and target genes (Boix et al., 2021).

Out of 60 CaREs participating in significant TESLA-seq interactions, 52 are defined by EpiMap as either enhancers (30), promoters (7) or accessible regions (15) (Figure 4A, Figure S4A). For all of these, we provide the target gene as identified by TESLA-seq as well as the tissue in which the genetic regulatory element is annotated

as endogenously active in EpiMap (Table S5). In 31 cases, the EpiMap annotated target gene is in agreement with the TESLA-seq, validating both the TESLA-seq experiment and the EpiMap computational prediction.

We classified all CaREs having a significant effect on at least one gene into three groups according to regulatory evidence in published chromatin accessibility or ChIP-seq data: 31 CaREs (involved in 51 interactions) have regulatory evidence in examined neuroblastoma (nb) cell line data, 21 CaREs (involved in 29 interactions) have evidence in EpiMAP data only (other), and 8 CaREs have no previous evidence (12 interactions; noEF) (Figure 4A,B, Figure S4A). We then compared these three classes of CaREs to determine if there are any striking differences. Laying out the results per gene (Figure 4A), we observed that: 1) some genes have no regulators despite being surrounded by genes with altered expression in response to CRISPRa (as described in Figure 3B-C), 2) many interactions in the “nb” group are closer in linear or 3D space (quantified in Figure 4E-F, Figure S4B,D-F), and 3) most of the genes targeted by CaREs in the “other” or “noEF” classes are also targeted by CaREs in the “nb” class (quantified in Figure S4C). Of note, the “noEF” and “other” class can not be explained purely based on prior 3D proximity, as there are a number of functional interactions despite a lack of prior 3D interaction. Considering the most significant target per CaRE, we find differences between the groups for their TESLA-seq significance (adj. p-val) and distance to target gene, but no difference between the groups for their effect size (log-fold-change) on their target genes (Figure 4C-E, Figure S4B), arguing against borderline false positives due purely to the chosen statistical cutoff (adj. P-val < 0.05;).

Together, these results suggest that TESLA-seq can activate and identify targets for regulatory elements active both within and outside of the assayed cellular context and that specific and functional transactivation can occur from elements with or without evidence of known regulatory characteristics, with or without prior 3D proximity. Furthermore, combining epigenomic databases with TESLA-seq results can associate CaRE-gene relationships with the tissues in which they are endogenously active.

Figure 4

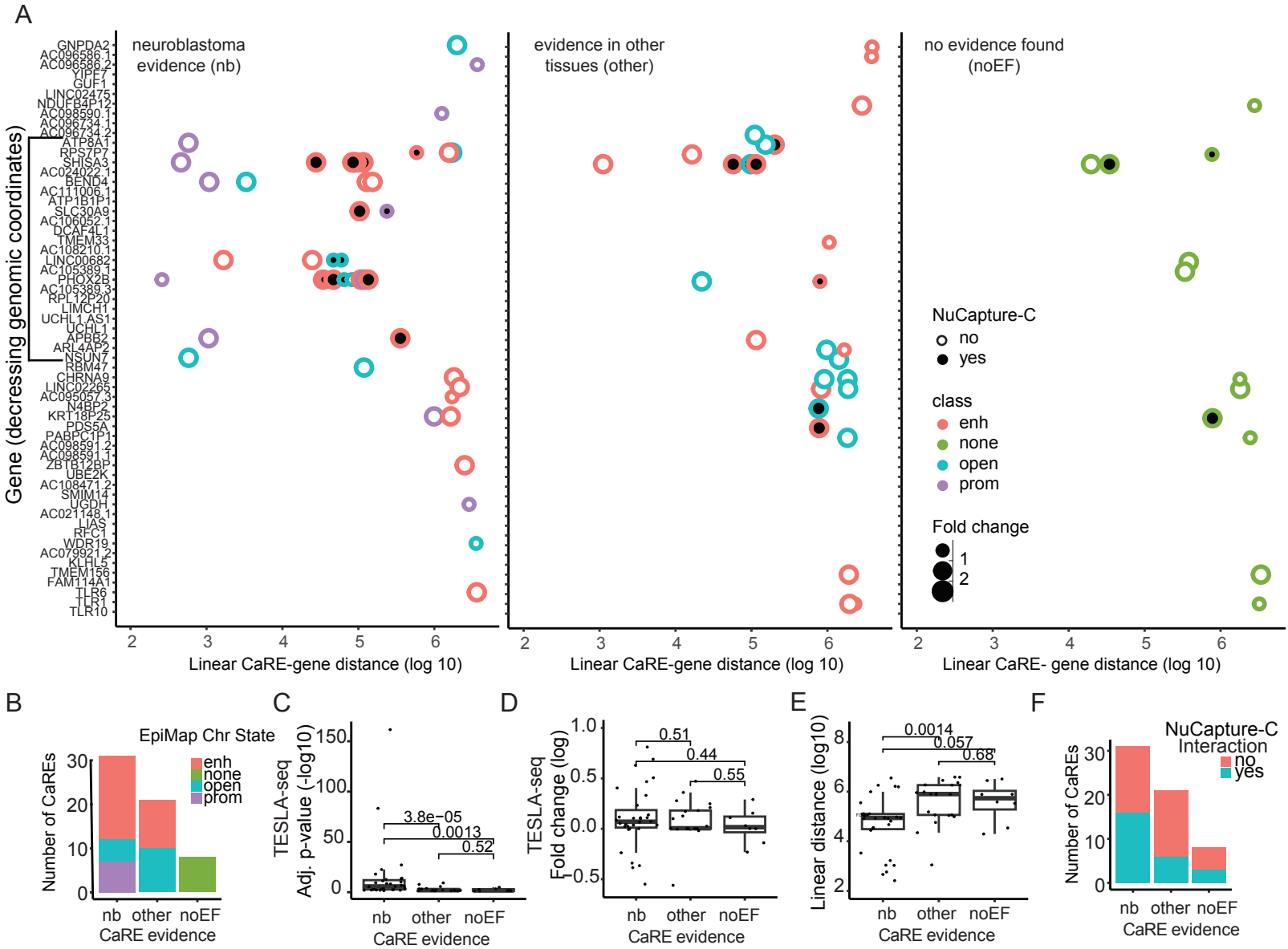


Figure 4. Integration of TESLA-seq results with epigenomic data.
A) For each gene target (displayed in the order that matches their order in the genome in decreasing genomic coordinates) linear distance to their CaRE determined by TESLA-seq is shown. The line indicates the screening window in the bulk screen from which CaREs were selected for TESLA-seq. The data is further stratified by: nuCapture-C signal displayed via circle fill (yes - 3D interaction is detected or no - no interaction detected); class defined by EpiMap indicated via circle colour; and fold changed determined via TESLA-seq depicted by circle size. CaREs are classified into three groups according to regulatory evidence in published ATAC or ChIP-seq data: left - CaREs with evidence in neuroblastoma (nb n=31 involved in 51 interactions), middle - evidence in tissue other than neuroblastoma (other n=21 involved in 29 interactions) and right - CaREs for which no evidence was found (noEF n= 12). **B)** Number of CaREs in each category, nb, other, noEF grouped by the EpiMap classification indicated with colours. **C-E)** Comparison between different CaRE evidence categories (nb, other, noEF) in CaRE-gene pair adj. P-value (-log10 scale) **(C),** fold change (log2 scale) **(D)** and linear distance **(E). F)** Number of CaREs in each category (nb, other, noEF) grouped by NuCaptureC detected interaction (yes - 3D interaction is detected or no - no interaction detected). See also Figure S4 and Table S5-6.

TESLA-seq identified regulators of APBB2 and SHISA3

In addition to PHOX2B, our initial focus for identifying CREs in the CRISPRa screen, TESLA-seq has identified CREs of several other genes that play important roles in both health and disease. We examined two examples in more detail.

APBB2 plays a role in Alzheimer's disease (Golanska et al., 2013, Li et al., 2005) and cell cycle (Bruni et al., 2002, Zhou et al., 2021). TESLA-seq identified 3 CaREs for this gene (Figure 5A). CaRE_233 is the promoter of APBB2 gene (highlighted). For CaRE_52 there is prior evidence in neuroblastoma epigenomics data (nb), and we validated it in SH-SY5Y cells using individual gRNAs (Figure 5A middle). For CaRE_174 we found no evidence in neuroblastoma data, but there is evidence that it acts as a regulatory element in cardiac tissue (Boix et al., 2021).

By promoting the degradation of β-catenin, SHISA3 contributes to the suppression of tumorigenesis, invasion and metastasis (Chen et al., 2014, Shahzad et al., 2020, Zhang et al., 2024). The silencing of this gene has been observed in a variety of different cancer types, such as colorectal cancer (Tsai et al., 2015, Tang et al., 2021), nasopharyngeal carcinoma (Zhang et al., 2019), lung adenocarcinoma (Si et al., 2019), and breast cancer (Shahzad et al., 2020). TESLA-seq identified 15 CaREs that regulate SHISA3 expression (Figure 5B). For 2 of them, we do not find any evidence in either neuroblastoma or in other tissues (Table S5). We validated 5 of them, including one without other evidence (Fig 5Bf); two with evidence in other contexts (Fig 5Bk,n); as well as two annotated neuroblastoma CREs with high (Fig 5Bm) and low fold change in the TESLA-seq (Fig 5Bb). Of note, the fold change effects detected via TESLA-seq resemble the fold changes detected via RT- qPCR, highlighting the sensitivity of the TESLA-seq.

Two of the CaREs identified to regulate SHISA3 by TESLA-seq (SHISA3 promoter CaRE_790 – Fig 5Bh and CaRE_816 – Fig5Bk) intersect with the somatic risk variants found in cancer (Hoadley et al., 2018). In EpiMAP, CaRE_816 is annotated to be active in renal tissue and cancer. This is another example of linking disease-relevant CREs to their target genes in tissues other than the one used for the screen.
In addition to regulating SHISA3, the promoter CaRE of SHISA3 (CaRE_790) is also identified to regulate the UGDH and TLR6 genes. This could be an example of a promoter acting as an enhancer, but we cannot rule out a secondary effect of modulating SHISA3 expression.

Figure 5

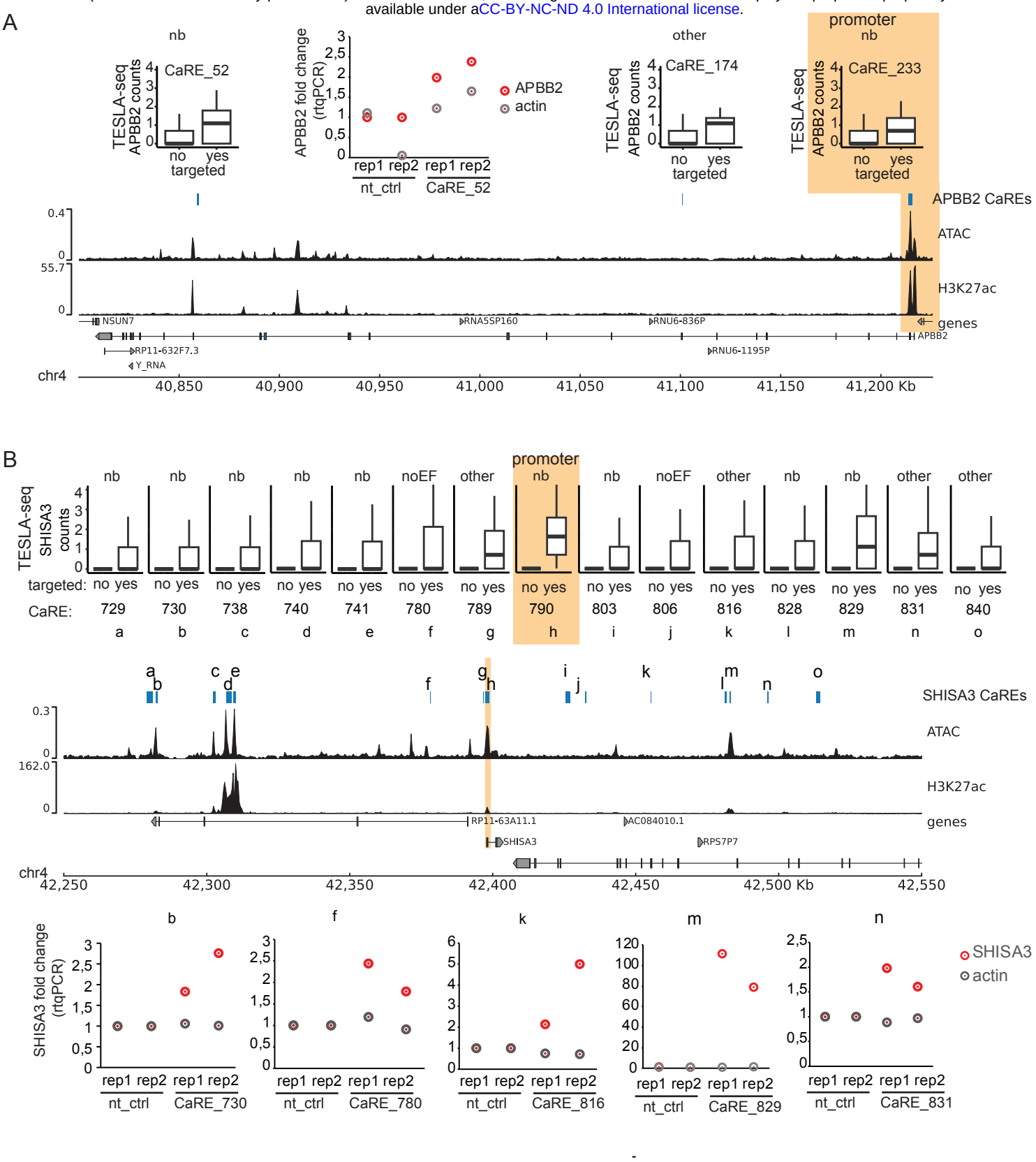


Figure 5. Selected examples of CaRE - gene pairs. A) Top left and right: TESLA-seq normalized gene expression level of APBP2 gene, comparing cells having gRNA targeting individual indicated CaREs versus all other cells. Top middle: APBP2 expression change upon activation with dCas9-VPR with control or gRNA targeting CaRE_52 in SH-SY5Y cell line. Relative expression of target genes is determined by RT-qPCR and normalized to GAPDH. Shown are two replicates and actin as a control gene. Bottom: Browser shot of the APBP2 genomic region, from top to bottom, the following tracks are displayed: TESLA-seq identified CaREs affecting APBP2 expression, ATAC-seq andH3K27ac ChIP-seq from SH-SY5Y cell line, and gene track. **B)** Top: TESLA-seq normalized gene expression level of SHISA3 gene, comparing cells having gRNA targeting individual indicated CaREs versus all other cells. Middle: Browser shot of the SHISA3 genomic region. From top to bottom the following tracks are displayed: TESLA-seq identified CaREs affecting SHISA3 expression, ATAC-seq and H3K27ac ChIP-seq from SH-SY5Y cell line, and gene track. Bottom: SHISA3 expression change upon activation with dCas9-VPR with control or gRNA targeting indicated CaRE in SH-SY5Y cell line. The following gRNAs were used: X886 for CaRE_730, X327 for CaRE_780, X341 for CaRE_816, X351 for CaRE_829, X355 for CaRE_831. Relative expression of target genes was determined by RT- qPCR and normalized to GAPDH. Shown are two replicates and actin as a control gene.

Discussion

A wide variety of resources have cataloged CREs and their potential involvement in disease based on their biochemical features (Boix et al., 2021, Hoadley et al., 2018, Zhang et al., 2020). CRE validation, functional characterization, and identification of their target genes represent a major challenge in the field. Here we developed a strategy that allows the identification of CREs and their targets at a large scale and at high sensitivity, by combining an unbiased tiled phenotypic screen with TESLA-seq - a CRISPRa screen with a targeted scRNA-seq readout of guides and their effects.

In our application on dissecting the regulatory landscape of the PHOX2B locus, we set up a tiling CRISPRa screen in a 2 Mb window and identified 619 CaREs that affect cellular growth/proliferation. For a subset of 222 CaREs, we examined the effect on the expression of genes within an extended 6 Mb window via TESLA-seq, resulting in 92 functional CaRE-target gene pairs. Since we selected CaREs based on a phenotypic effect in the tiled screen window, there are likely many functional CaREs in this window that we did not assay, affecting genes within or outside of this window. Due to the novelty of the approach, we employed very stringent and conservative filtering criteria at every step of the TESLA-seq analysis. Additional data from further TESLA-seqs will allow us to improve on computational tools for the analysis of this type of data.

On average, there are 7.7 genes skipped between a CaRE and its target, meaning that the commonly used strategy of associating an enhancer to its closest active gene is likely suboptimal (Gasperini et al., 2020, Kim and Wysocka, 2023), although this will strongly depend on the specific locus. These observations underline the difficult task of computationally predicting enhancers and their targets and the need for scaling up the functional characterization of CREs. Since the expression of skipped genes is not affected, these results also demonstrate the sensitivity of TESLA-seq to characterize distal regulatory relationships. However, due to the time period elapsing between transduction and single-cell readout, we cannot rule out additional secondary/indirect regulatory effects. While further studies are needed to confidently distinguish between the two, the observed evidence for prior 3D interaction between significant CaRE-gene pairs as well as the overlap with known regulatory elements and targets suggest that we are primarily identifying direct effects.

Given current understanding about CREs (Gasperini et al., 2020, Kim and Wysocka, 2023), we find that the majority of significant enhancer-gene pairs interact in 3D genomic space before CRISPRa. Prior 3D proximity allows for stronger transactivation, however it is neither sufficient nor required. This observation further exemplifies that, although 3D proximity can be an indicator of a functional relationship between a CRE and a gene, functional characterization is necessary to reliably assign a regulatory effect of a CRE on its target.

A major advantage of an activation screen is that it includes CREs that are not necessarily active in the assayed cell line. By combining the TESLA-seq results with available epigenomic maps from 800 tissues, we were able to assign the likely relevant tissue for 29 CaRE-gene relationships. Using a CRISPR activator to assay possible CREs regardless of their endogenous activity is especially important for studying CREs that are only active in rare cell types and cell types that we cannot easily culture and study.

Finally, the epigenetic landscape at the time of the experiment will influence the efficiency of CRISPR perturbation. In general, CRISPRi/a tends to be more efficient in open chromatin regions (Horlbeck et al., 2016), and although gene activation by dCas9-VPR is successful in most genomic contexts, including bivalent chromatin, constitutive heterochromatin is less responsive (Wu et al., 2023). Our findings are in line with this: the most responsive CaREs are the ones that are in a permissive chromatin state in the assayed cell line. These results may reflect a combination of dCas9 requiring physical accessibility to its DNA target and/or dCas9-VPR requiring other CRE features, such as well positioned nucleosomes and divergently-oriented core promoter sequences (Duttke

et al., 2015, Lacadie et al., 2016, Ibrahim et al., 2018).To bypass these limitations, future studies will benefit from the use of other activators and repressors as well as screening in a handful of different cell types (Kearns et al., 2015, Gasperini et al., 2020, Bock et al., 2022).

TESLA-seq enables detailed, sensitive investigation of gene expression regulation. Its high-throughput approach offers the potential to uncover transcriptional regulatory logic in its natural environment at a competitive scale, and it provides a significant step forward for uncovering the effects of non-coding sequence variation on disease.

Acknowledgments

We thank Marieke Oudelaar for advice on nuCaptureC and for sharing unpublished work with us. This work was supported by International Research Training Group 2403 of the Deutsche Forschungsgemeinschaft and the MDC/BIMSB-NYU exchange program.

Author Contributions

DV performed all experiments with technical assistance from AH and IB, performed some bioinformatic analysis and made the figures. LSZ, MB, and SAL performed the bioinformatic analysis of the bulk CRISPR screens. CW and SAL performed the bioinformatic analysis of the TESLA-seq data. NC and ML cloned the dcas9-VPR and dCas9-KRAB piggy back vector. DV, SAL and UO wrote the manuscript. CW, LSZ, and

MB gave comments and suggestions on the manuscript and performed minor editing.

DV, SAL, and UO designed the study.

Declaration of interests

The authors declare no competing interests.

Methods

Resource availability

- Lead contact: Further information and requests for resources and reagents should be directed to and will be fulfilled by the lead contact, Uwe Ohler (Uwe.Ohler@mdc-berlin.de).
- Technical contact: Technical questions on executing this protocol should be directed to and will be answered by the technical contact, Dubravka Vučićević (vucicevic.dubravka@gmail.com).
- Materials availability: Correspondence and requests for materials should be addressed to the technical contact. The unique identifiers of all biological materials are listed in the methods. The newly generated plasmids can be obtained from the technical contact.
- Data and code availability: Data generated for this study are accessible at GEO under accession numbers: GSE274254 for ATAC-seq, GSE274255 for bulk CRISPR screen, GSE274256 for TESLA-

seq, GSE274257 for nuCapture-C and GSE274258 for chromatin RNA-seq. All codes for data analysis and visualization have been deposited at https://github.com/ohlerlab/Vucicevic_et_al and are publicly available. Any additional information necessary to re-analyze the data reported in this paper is available from the technical contact upon request.

Cell culture

SHSY-5Y cells (DSMZ #ACC 209) were cultured in DMEM/F12 (Thermo Fisher Scientific #31330038) supplemented with 20% FBS (Thermo Fisher Scientific #16000044) and 1% penicillin/streptomycin (Thermo Fisher Scientific #15070063) at 37°C with 5% CO_2. When 90% confluent, cells were washed with PBS (Thermo Fisher Scientific #10010015), trypsinized using 0.05% Trypsin-EDTA (Thermo Fisher Scientific #25300096) for 2 min, spun down in complete growth medium at 400g for 3 min and split 1:6 into clean TC dishes.

Comparison of CRISPR activators

We targeted MyoD enhancers: distal regulatory region (DRR) and core enhancer (CE) as well as the MyoD promoter through co-transfection of gRNAs targeting each region and either dCas9-p300 (Hilton et al., 2015), dCas9-SunTag10 (Tanenbaum et al., 2014), dCas9-SunTag24 (Tanenbaum et al., 2014) or dCas9-VPR (Chavez et al., 2015). $2x10^5$ HEK293 cells were co-transfected with 20ng of either equimolar pool of gRNAs or 20ng of undivudual gRNAs and 50ng of either one of the dCas9 fusion constructs with lipofectamine2000 (Chavez et al., 2015). Cells were collected 48h or

96h later in the case of dCas9-p300. Relative expression was determined by qRT-PCR and normalized to gapdh, actin is shown as a control gene.

Cloning and transduction of individual gRNAs

Ordered oligonucleotides contained sticky-end overhangs matching the Esp3I overhangs of the CROPseq-Puro digested with Esp3I (Thermo Fisher Scientific #FD0454). Sense and antisense oligonucleotides were annealed and ligated into the digested backbone using T4 DNA ligase (NEB). They were transformed in Stbl3 bacteria and the clones were validated with Sanger sequencing. Lentiviral production was done as described in (Datlinger et al., 2017). The lentiviral prep was concentrated using Lenti-X Concentrator (Takara Bio #631232). Cells were transduced at a low multiplicity of infection (<0.3 MOI), selected 48h post-transduction using 2ug/ml of puromycin (Sigma Aldrich #P9620) and treated as indicated in different assays.

Cloning of dCas9 constructs

dCas9-VPR and KRAB-dCas9-P2A-mCherry constructs were designed to be flanked by NheI and NotI restriction sites and synthesized into the pEX-A2 backbone (Eurofins). To generate dCas9 piggybac transposons, pEX-A2 plasmids containing the dCas9 constructs were digested with NheI and NotI and cloned into the backbone of pPB-CAG-3xFLAG-empty-pgk-hph (addgene #48754) between XbaI and NotI sites, rendering the XbaI site destroyed. These constructs will be deposited to addgene.

Generation of stable cell lines

Stable SH-SY5Y-dCas9VPR and SH-SY5Y-dCas9KRAB cell lines were generated using the PiggyBacTM Transposon Vector System. According to the manufacturer's

instructions, $0.5x10^6$ SHSY-5Y cells (passage number p10) per well of a 6-well plate were transfected using Lipofectamine 2000 (Thermo Fisher Scientific #11668019) in a 1:5 ratio of PiggyBac transposase (BioCat #PB210PA-1-SBI) to PiggyBac transposon carrying either dCas9-VPR or KRAB-dCas9-P2A-mCherry constructs, respectively. The following day, cells were selected for successful integration using 300 ug/ul of hygromycin (Invivogen #ant-hm-1) until the control cells were dead (4 days). They were re-selected before each experiment to ensure the expression of the constructs.

Viability assays

The viability of SH-SY5Y-dCas9VPR and SH-SY5Y-dCas9KRAB cells treated with either control, six gRNAs targeting the promoter of PHOX2B, or any other indicated treatment was measured using the RealTime-Glo MT Cell Viability Assay (Promega #G9711) according to the manufacturer's instructions. Luminescence was measured 24, 48, and 72 hours after the cells were selected for gRNA expression.

qPCR measurement of CRISPRa/i effects

Total RNA was extracted from cells following the manufacturer's instructions with Trizol (Thermo Fisher Scientific #15596018). Reverse transcription was carried out using the High Capacity RNA-to-cDNA Kit (Thermo Fisher Scientific #4387406), and RTqPCR was performed using Fast SYBR Green Master Mix (Thermo Fisher Scientific #4385616) according to the manufacturer's instructions on a Real-Time PCR System (Roche lightcycler 480 II). The relative expression was calculated by normalizing to GAPDH expression levels as a control housekeeping gene. Primers used for RT-qPCR can be found in Table S1.

Design and selection of gRNAs for the bulk screen

100bp bins of the 2MB region surrounding PHOX2B were used as input for guidescan software (Perez et al., 2017) with parameters: guidescan_guidequery (--target within -o . --output_format csv --select score -n 3) to select the top 3 guides for each bin from a pre-computed Cas9 guide database for hg38. We included 100 control gRNAs from the GeCO V2 library (Sanjana et al., 2014). Suitable overhangs for the Gibson assembly were added (Joung et al., 2017).

CRISPRa/i viability screen

Oligo pools containing 46,722 oligos (Table S3) ordered from Twist Bioscience were cloned into the CROPseq-Guide-Puro plasmid (Addgene #86708) digested with Esp3I (Thermo Fisher Scientific #FD0454) according to (Joung et al., 2017). The library was amplified using primers and according to the protocol from (Datlinger et al., 2017) and sequenced on a NextSeq 500/550 machine according to the Illumina user manual. Lentiviral production was done as described in (Datlinger et al., 2017). The lentiviral prep was concentrated using Lenti-X Concentrator (Takara Bio #631232) and the viral titer was determined using a Crystal violet viability assay (Vučićević et al., 2016). The screen was performed as described in (Datlinger et al., 2017). For each gRNA at least 1000 cells per gRNA were seeded and they were transduced at a low multiplicity of infection (<0.3 MOI) in either SH-SY5Y-dCas9VPR and SH-SY5Y-dCas9KRAB cell line. The cells were selected 48h post-transduction for the ones that received the gRNA using 2ug/ml of puromycin (Sigma Aldrich #P9620) for three days. They were kept under constant antibiotic selection with 1ug/ml of puromycin and 100ug/ml of hygromycin and collected at several time points (day 5, day 20, day 29 and day 33).

DNA was extracted using the Quick-DNA Midiprep Plus kit (Zymo Research #D4075). The library preparation was done as described in (Datlinger et al., 2017) and sequenced on a NextSeq 500/550 according to the Illumina protocol.

Bulk CRISPRa/i data analysis

Sequencing reads were assigned to sgRNA using a modified version of a script in (Sanjana, 2016). Mean guide-counts were obtained at each time-point and all time-points were combined in a single table. Normalized counts c_j were obtained using the following equation: $c_j = (q_j/(q)_k) \times 10^6$ (where q_j is reads mapped to a specific guide and $(q)_k$ is total number of reads mapped to all guides). We employed a sliding window strategy to score each bin based on the information from the sgRNAs targeting it and the adjacent bins upstream and downstream. Each bin was analyzed using guides targeting a 300bp window centered on it. To assess guide enrichment or depletion within these windows, we utilized a mixed linear model (MLM) with random intercepts:

$\mathrm{Counts}i = \mathrm{j}[\mathrm{i}] + [1](\mathrm{Time})$, for bin i

$\sim \mathrm{N}(\mathrm{j}, 2\mathrm{j})$, for sgRNA $\mathrm{j} = 1, \ldots, \mathrm{J}$

Where $\mathrm{Counts}i$ represents the log counts of bin$i$; $\mathrm{j}[\mathrm{i}]$ is the intercept for guide j of bin i; $[1]$ is the overall estimate for Time.

The model was fitted using the 'lme4'(v. 1.1.28) *R* package (REML = FALSE) (Bates *et al.*, 2015).

Hypothesis testing was done by comparing our full model to a reduced version that does not account for time using a maximum likelihood ratio via ANOVA (Kaufmann and Schering, 2014) as follows:

$\mathrm{Counts}i = \mathrm{j}[\mathrm{i}]$, for bin i

$\sim N(j, 2j)$, for sgRNA $j = 1, \ldots, J$

Adjustment of the p-value for correction of multiple tests was performed using the Benjamini-Hochberg method (Benjamini and Hochberg, 1995). After identifying significant windows, bedtools (v 2.30.0) (Quinlan and Hall, 2010) was used to merge significant bins within 500bp of each other into CaREs. If a significant bin was not within 500 bp of another significant bin, it was classified as a distinct CaRE. CaREs were scored by MLM using information from all guides targeting the genomic area defined by the merged CaRE. Scores were calculated as: -log10 (adjusted p-value).

ATAC-seq

ATAC-seq experiments were performed in the SH-SY5Y, CLB-Ga, IMR-5, Kelly, NGP and SK-N-SH neuroblastoma cell lines using 100,000 cells according to the protocol (Buenrostro et al., 2015) with the following modifications: transposition time was increased from 30 min to 1 h and the cell pellets were taken directly to the transposition reaction omitting the lysis step as described in (Karabacak Calviello et al., 2019). For all samples, 12 PCR cycles were performed, and the libraries were sequenced (2x75nt) on a NextSeq 500/550 using a HighOutput v2 Kit for 150 cycles (Illumina #FC-404-2002, discontinued).

ATAC-seq processing

ATAC-seq reads were trimmed for adapter content using flexbar (-f i1.8 -u 10 -ae RIGHT -at 1.0; (Dodt et al., 2012)), mapped to hg19 using bowtie2 (-X 1500 --no-discordant; reads eventually lifted over to hg38; see below; (Langmead and Salzberg, 2012), filtered for unique mapping reads (grep -v "XS:i:"), and collapsed for PCR

duplicates using Picard Tools MarkDuplicates (http://broadinstitute.github.io/picard). Finally, the 5'ends of reads were selected and extended to account for the estimated footprint size of Tn5 on the DNA using bedtools slop (-l 15 -r 22 -s; (Quinlan and Hall, 2010, Adey et al., 2010); 11(12): R119). Peaks were called using JAMM (-e auto -b 100; (Ibrahim et al., 2015)) on fragment extended reads.

Chromatin RNA-seq

Cellular fractionation was performed in SH-SY5Y cell line according to (Conrad and Ørom, 2017). Chromatin RNA were extracted using Trizol and Direct-zol RNA MiniPrep Kit (Zymo Research #R2052) according to the manufacturer's instructions. The library was prepared using NEXTflex Rapid Directional qRNA-Seq Kit (BiooScientific #NOVA-5130-01D) according to the manufacturer's instructions and paired-end sequencing (2x75nt) was performed on a NextSeq 500/550 using a HighOutput v2 Kit for 150 cycles.

RNA-seq processing and analysis

Unique molecular identifiers (UMIs) were extracted from .fastq files using UMI-tools (Smith et al., 2017), and reads were trimmed using fastx_trimmer from the FASTX-toolkit (http://hannonlab.cshl.edu/fastx_toolkit/). Reads were then filtered for ERCC spike-in reads and rRNA by mapping to a custom index with Bowtie 1 (Langmead et al., 2009). Trimmed and filtered reads were then mapped using STAR (Dobin et al., 2013). Mapped .bam files were subjected to PCR deduplication using UMI-tools (Smith et al., 2017), followed by conversion to .fastq and remapping with STAR to generate final mapped files and normalized coverage tracks.

Analysis of Ca/iRE features in neuroblastoma data

Raw sequencing data from ChIP-seq analyses of histone modifications and transcription factors in neuroblastoma cell lines (Henrich et al., 2016, Durbin et al., 2018, Wang et al., 2019, Boeva et al., 2017, Zhang et al., 2020) was downloaded from the Sequence Read Archive (https://www.ncbi.nlm.nih.gov/sra) under accessions SRR3363255, SRR3363256, SRR3363257, SRR3363258, SRR3363259, SRR5249434, SRR5249436, SRR5249437, SRR5249438, SRR5249439, SRR5249440, SRR5249442, SRR5249443, SRR5249446, SRR5249447, SRR5249451, SRR5675976, SRR5675978, SRR5676027, SRR5676028, SRR5676029, SRR6451360, SRR6451361, SRR6451362, SRR7101491, SRR7101492, SRR7865946, SRR7865947 and SRR8169718 and from the ENCODE database (https://www.encodeproject.org/) under accessions ENCFF000ZPW, ENCFF000ZPZ, ENCFF000ZQF, ENCFF000ZQI, ENCFF000ZQL, ENCFF000ZQX, ENCFF072EIX, ENCFF443LZC, ENCFF458ASE, ENCFF557DAH and ENCFF615LEB. Reads were mapped to the hg38 reference genome using bowtie2 (version 2.3.4.3). Fragment lengths were estimated for each sample (and input/background) alignment by macs2 (version 2.1.1.20160309) using parameter -g 1.6e+9. From the list of estimated fragment lengths per sample, the length closest to 200 bp was selected and read mapping coordinates of that sample were extended to the respective fragment length. Read coverage of processed ChIP-seq and ATAC read coordinates were summarized in bins and regions of the CRISPRa phenotypic screen analysis respectively, using the function summarizeOverlaps from the R/Bioconductor package GenomicAlignments (version 1.22.1) with parameters mode="Union" and inter.feature=FALSE. To capture

signals of histones adjacent to CaREs the respective summary features of histone ChIP-seq samples were resized to a minimum of 1000 bp centering on the CaRE. ATAC-seq read coordinates were translated to hg38 using the function liftOver from the R/Bioconductor package rtracklayer (version 1.46.0). Read coverage per region was normalized to reads per kilobase (RPK). Peaks were called using JAMM (-e auto -b 100; (Ibrahim et al., 2015)) on fragment extended reads, and peaks for transcription factors were further filtered to have a peak score > 515.

Significant CaREs were defined as regions with FDR < 0.05 from the respective CRISPR screen analysis. ChIP-seq and ATAC-seq read coverages were normalized to z-scores for downstream analysis. A unified generalized linear model (binomial link function) was applied to predict significant CaREs by ChIP-seq z-scores of H3K27ac, H3K4me1, H3K4me3 and H3K27me3 and ATAC-seq z-scores in neuroblastoma cell line SH-SY5Y. Additionally, significant CaREs were predicted by individual generalized linear models (binomial link function) from transcription factor ChIP-seq experiments in neuroblastoma cell lines, where each model was controlled for signal z-scores of the more general epigenetic features H3K27ac, H3K4me1 and H3K4me3 from ChIP-seq and ATAC-seq in cell line SH-SY5Y.

TESLA-seq

We selected 222 top-scoring significant CaREs from the bulk phenotypic screen. For each of them, we selected gRNAs with the highest fold change. These gRNAs, together with 52 control gRNAs (total of 1098 in Table S5), were ordered as oligos from Twist Bioscience, cloned, and sequenced as described for the phenotypic screen. The lentiviral production was done as in the viability screen. SHSY-5Y-VPR line was infected

at <0.3 MOI and the selection with 2ug/ul of puromycin started 24h post transduction. As soon as the antibiotic control non-infected cells were dead,on day 4 after transduction, the cells were collected. The viability of the cells was > 85% in both experiments as determined by the BD Rhapsody scanner after staining with Calcein AM (Thermo Fisher Scientific #C1430) and Draq7 (Thermo Fisher Scientific #564904) according to the manufacturer's protocol. Single-cell capture and cDNA synthesis were performed using the BD RhapsodyTM Single-Cell Analysis System according to the manufacturer's instructions. Capture probes for 146 transcripts corresponding to 78 genes in the 6MB genomic space (+/- 3MB from the PHOX2B TSS) on chr4 were designed by BD. The targets were enriched and the library was prepared according to the BD mRNA Targeted Library Preparation protocol. Paired-end sequencing (2x75nt) was performed on a NextSeq 500/550 using a HighOutput v2 Kit for 150 cycles with a 20% PhiX spike-in.

Computational analysis of TESLA-seq data

The raw sequencing output files were preprocessed using BD's Rhapsody™ Targeted Analysis pipeline (https://scomix.bd.com/hc/en-us/articles/360019763251-Bioinformatics-Guides) on the Seven Bridges Platform (https://www.sevenbridges.com/bdgenomics/), which generated a cell-by-gene count matrix corrected for base-calling errors in cell barcodes, followed by quality-filtering on cells. Processed count matrices of two sample runs contain 12,597 (run1) and 19,406 (run2) cells, respectively, with features consisting of 1,046 guide RNAs (gRNA) targeting the 2Mb region around the PHOX2B gene, 52 control gRNAs, and 78 genes within the 6Mb region around the PHOX2B gene. The count matrices were loaded into

the R environment for further preprocessing with Seurat v4 (Stuart et al., 2019). Logarithm base 10 was applied to both the gene count matrices and gRNA count matrices. The gRNA count values display a clear bimodal pattern, suggesting that the peak of lower values could be noise. After setting the lower values in gRNA matrices to zero, gRNAs with no count values across cells and genes with no expression were removed. Cells with either low gRNA values or low gene expression were filtered. After filtering, run1 has 6,196 cells with 1,090 gRNAs and 62 genes, while run2 has 10,363 cells with 1,097 gRNAs and 64 genes.

The quality-filtered matrices of two runs were normalized using SCTransform (Hafemeister and Satija, 2019) and integrated with batch correction done by linear regression between runs (SCTransform argument 'var.to.regress'). Cells that had less than two normalized and batch-corrected guide counts were removed and genes that have no expression following the cell removal were filtered. gRNA capture efficiency was computed using a generative model that takes into account the multiplicity of infection (Dixit et al., 2016, Hill et al., 2018).

To test the effect of each gRNA and avoid potential confounding caused by nearby gRNAs, differential expression analysis of each targeted gene was performed between cells that contain the considered gRNA and all other cells that contain other gRNAs targeting at least 400 kb away from the considered gRNA. MAST.cov test was chosen to perform differential expression analysis as recommended for targeted single-cell data (Schraivogel et al., 2020). A significance level of 0.05 adjusted p-value was used to identify regulation pairs of gRNA/CaRE and gene. Plots were generated using R packages ggplot2 v3.3.2 (Wickham, 2016) and Sushi v1.7.1 (Phanstiel et al., 2014).

Nuclear Capture C

nuCapture C was performed according to (Downes et al., 2022) in SH-SY5Y cell line. Single-stranded DNA probes for genes in the 6MB space surrounding PHOX2B were obtained from IDT as an xGen Lockdown Pool and are listed in table S2. Paired-end sequencing (2x150nt) was performed on a NextSeq 500/550 using a HighOutput v2 Kit for 300 cycles (Illumina #FC-420-1004, discontinued).

Nuclear Capture C data processing

Nuclear Capture C reads were initially processed using the HiCUP pipeline (Wingett et al., 2015) with Bowtie2 mapping to hg38. Mapped reads were then converted to .chinput format using chicagoTools and used as input to chicagoPipeline (Cairns et al., 2016) along with DpnII genomic fragment regions targeted for capture during library preparation.

Hi-C data processing

Raw reads from published IMR5/75 cells Hi-C experiments (Helmsauer et al., 2020) were downloaded from SRA (PRJNA622577). Processing steps were implemented within the Snakemake framework (Mölder et al., 2021). Hi-C reads were initially processed using the Juicer pipeline (Durand et al., 2016). TADs were called using the insulation method (Crane et al., 2015) with default settings (–is500000–nt0–ids250000–ss0–immean) after dumping valid interactions at 25 kb binned resolution using juicer-tools and converting formats using HiTC (Servant et al., 2012). For scoring Hi-C interactions, valid interactions were dumped by juicer-tools at fragment resolution, filtered to remove interactions less than 20 kb apart, format converted using custom

scripts, and then subjected to shuffling and scoring using the SHAMAN method (Cohen et al., 2017).

Random forest model for predicting significant CaRE-gene pairs from genomic features

Random forest models were trained with scikit-learn tools v1.2.1 (Varoquaux et al., 2015) in Python. The model was trained to learn how well genomic features, linear distance between a perturbed region and gene, CaptureC score, ATAC signal, and epigenetic markers can predict the regulatory relationships between CaREs and genes. The top 50 and bottom 50 regulatory pairs, ranked by adjusted p-values, were selected as the target of prediction (1 for top 50, 0 for bottom 50), and the genomic features were treated as the explanatory features.

TESLA-seq multimodal integration analysis

Nuclear capture-c scores were assigned to CaRE-gene pairs by intersecting 25kb regions surrounding CaRE midpoints and gene TSSs with fragment-target pair coordinates and scores as output by the Chicago pipeline using bedtools pairtopair (-slop 12500; (Quinlan and Hall, 2010)) and taking the maximum Chicago score per pair within this 25kb interaction square. Pairs were classified as interacting if this maximum Chicago score was greater than 3.

Hi-C scores were assigned to CaRE-gene pairs by taking the highest Shaman score within a 25kb square centered on the CaRE midpoint and gene TSS using Shaman scripts. A CaRE-gene pair was classified as Hi-C interacting if the highest Shaman score in this 25kb square was greater than or equal to 25.

Epimap (Boix et al., 2021) link coordinates and scores were downloaded by tissue group and concatenated. Epimap region coordinates were shifted to midpoints and gene TSS coordinates were assigned via Ensembl gene ID. These pair coordinates were intersected with CaRE-midpoint-TSS pairs using bedtools pairtopair (-slop 2500) and the maximum score within this 5kb square was assigned to CaRE-gene pairs.

Abc file AllPredictions.AvgHiC.ABC0.015.minus150.ForABCPaperV3.txt was downloaded, converted to bed format with region coordinates, gene name, and ABC.score, lifted over to hg38, target gene TSS coordinates retrieved via gene name, and the maximum ABC.score across tissues was taken for each region-geneTSS pair. Regions coordinates were then shifted to midpoints and these pairs were intersected with CaRE-midpoint-geneTSS coordinates using bedtools pairtopair (-slop 2500) and the maximum ABC.score in this 5kb square was assigned to each CaRE-gene pair.

Histone modification ChIP-seq and ATAC-seq signal was assigned to CaREs and gene TSSs using 1kb windows centered on CaRE midpoints or gene TSSs and summarizing the signal as described above in section “Analysis of Ca/iRE features in neuroblastoma data”.

Chromatin RNA-seq signal was assigned to genes as the mean counts between two replicates as output by STAR.

In order to classify CaREs by previous evidence, we used our own ATAC-seq and published ChIP-seq from neuroblastoma cells (except for H3K27me3), called peaks as described above, and concatenated and merged all resulting regions as neuroblastoma regulatory evidence. Epimap DHS, dyad, promoter, and enhancer regions for hg38 were downloaded from /www.meuleman.org/ and

personal.broadinstitute.org/cboix/epimap/mark_matrices/ and DHS coordinates were matched to Epimap annotations using identifiers. The resulting file and the merged nueroblastoma regulatory evidence were each separately used to find the closest element to each CaRE using bedtools closest (-d -t "first"; (Quinlan and Hall, 2010)) and a true assignment was made if the distance of the closest element was less than or equal to 250bp. Dyads were considered promoters. If a CaRE was within 250bp of a neuroplastoma reulatory evidence element, it was classified as “neuroblastoma” (“nb”), otherwise if it was close to an Epimap element, it was classified as “other”, otherwise it was classified as “no evidence found” (“noEF”).

Supplemental information

Document S1:

Figures S1–S4, supplementary figure legends, Table S1- primers and individual gRNA oligos used in the study and Table S2 - oligos used for nuCaptureC experiment.

Table S3. File containing the results of the bulk CRISPRa screen, related to Figure 1.

Table S4. File containing the results of the bulk CRISPRi screen, related to Figure 1.

Table S5. File containing the results of the TESLA-seq for CaREs, related to Figures 2-5.

Table S6. File containing the results of the TESLA-seq for individual gRNAs, related to Figures 2-5.

Supplementary Figure legends:

Figure S1. Details of the bulk CRISPR activation screen related to Figure 1. A) Upregulation of MyoD in HEK293 cells via different indicated CRISPRa constructs at three different non-coding loci via pools of four gRNAs (left) or individual gRNAs (right). Relative expression of target genes is determined by RT- qPCR and normalized to GAPDH. Shown are fold changes relative to the control at a log2 scale. **B)** PHOX2B expression changes upon repression with dCas9-KRAB (left) or activation with dCas9-VPR (right) with control or gRNAs targeting the promoter of Phox2B in SH-SY5Y cell line. Relative expression of target genes is determined by RT- qPCR and normalised to GAPDH. Shown are fold changes relative to the control. Expression of actin is shown as a control. **C)** MTT viability assay with control or gRNAs targeting the promoter of Phox2B in a SH-SY5Y-KRAB (CRISPRi) cell line. **D)** CaREs signal at the PHOX2B promoter. Significantly enriched/depleted CaREs (FDR<0.05) are shown in red and blue respectively. **E)** Volcano plot showing the log-fold change of gRNA representation between the first and last time-point of the CRISPRi experiment. Significantly enriched/depleted gRNAs (FDR<0.05) are shown in red and blue respectively. **F)** Volcano plot showing the results for our CRISPRi screen, each dot corresponds to a CaRE defined by the CRISPRa screen. The x axis shows the slope calculated by MLM. Significantly enriched/depleted CaREs (FDR<0.05) are shown in red and blue respectively.

Figure S2. Details of the TESLA-seq related to Figure 2. A) Number of gRNA detected per cell in each TESLA-seq run and both together (integrated). **B)** Cell density plot. The red line represents thresholds used for cell filtering. Cells having either gene UMI count per cell or gRNA UMI counts per cell below thresholds are filtered. **C)** TESLA-seq normalized gene expression level, comparing cells having gRNA targeting PHOX2B and BEND4 promoter versus all other cells. **D)** Histogram showing the number of CaRE that induce differential expression of an indicated number ofgene. **E)** Number of gRNA that significantly cause differential gene expression of an indicated gene. **F)**. Genome browser snapshot representing genomic distance and regulatory relationship between a gRNA and a gene determined by TESLA-seq. Each link represents an effect of a gRNA on the gene. Color reflects the target gene. **G)** Histogram showing the number of genes that are differentially expressed due to perturbation of an indicated number of gRNA.

Figure S3. Additional details of properties of CaREs and their targets identified by TESLA-seq related to Figure 3. A) Relationship between adjusted p-values (-log10) of each CaRE-gene regulatory pair and genomic distance between the CaRE and the gene's transcriptional start site (in bases at log10 scale). **B)** Relationship between adjusted p-values (-log10) of each CaRE-gene regulatory pair and number of genes

located between the CaRE and the gene (jumped genes). Color depicts the average fold change. **C)** Number of contacts detected for PHOX2B by nuCaptureC centered on the PHOX2B TSS (red: highly significant interactions (pLevel = 5, blue: significant interactions, pLevel = 3, black: background interactions**). D)** Comparison of EpiMap link score between TESLA-seq significant and non-significant CaRE-gene regulatory pairs**. E)** Comparison of Abc score between significant and TESLA-seq non-significant CaRE-gene regulation pairs**. F)** Genome browser snapshot of PHOX2B regulatory regions determined by EpiMap (top) and Abc (bottom). **G)** Comparison of the CaRE target gene expression (average fold change at log 2 scale) between TESLA-seq significant and non-significant CaRE-gene regulation pairs that have a low or high: H3K27 signal at the promoter of the target gene (left) and chromatin RNA signal (right) (non-significant).

Figure S4. Additional details of the integration of TESLA-seq results with epigenomic data related to Figure 4. A) Number of CaREs defined by EpiMap in each indicated tissue. The color indicates the EpiMap chromatin state. **B)** Comparison between CaRE evidence categories (nb, other, noEF) relationship between TESLA-seq adjusted p-value (-log10 scale) and fold change (log2 scale). The data is further stratified by: nuCapture-C interaction displayed via circle fill and class defined by EpiMap indicated via circle colour. **C)** Venn diagram displaying an overlap between genes targeted by CaREs in nb, other or noEF classes. **D)** CaRE-gene nuCapture-C score stratified by CaRE evidence class. **E)** Number of CaREs in each evidence categories (nb, other, noEF) grouped by Hi-C detected interaction (yes - 3D interaction is detected or no - no interaction detected). **F)** Comparison between different CaRE evidence categories (nb, other, noEF) in Hi-C score.

References

ADAMSON, B., NORMAN, T. M., JOST, M., CHO, M. Y., NUñEZ, J. K., CHEN, Y., VILLALTA, J. E., GILBERT, L. A., HORLBECK, M. A., HEIN, M. Y., PAK, R. A., GRAY, A. N., GROSS, C. A., DIXIT, A., PARNAS, O., REGEV, A. & WEISSMAN, J. S. 2016. A Multiplexed Single-Cell CRISPR Screening Platform Enables Systematic Dissection of the Unfolded Protein Response. *Cell,* 167**,** 1867-1882.e21.

ADEY, A., MORRISON, H. G., ASAN, XUN, X., KITZMAN, J. O., TURNER, E. H., STACKHOUSE, B., MACKENZIE, A. P., CARUCCIO, N. C., ZHANG, X. & SHENDURE, J. 2010. Rapid, low-input, low-bias construction of shotgun fragment libraries by high-density in vitro transposition. *Genome Biol.,* 11**,** R119.

ALDA-CATALINAS, C., BREDIKHIN, D., HERNANDO-HERRAEZ, I., SANTOS, F., KUBINYECZ, O., ECKERSLEY-MASLIN, M. A., STEGLE, O. & REIK, W. 2020.

A Single-Cell Transcriptomics CRISPR-Activation Screen Identifies Epigenetic Regulators of the Zygotic Genome Activation Program. *Cell Syst,* 11**,** 25-41.e9.

ALERASOOL, N., SEGAL, D., LEE, H. & TAIPALE, M. 2020. An efficient KRAB domain for CRISPRi applications in human cells. *Nat. Methods,* 17**,** 1093-1096.

BACHETTI, T. & CECCHERINI, I. 2020. Causative and common PHOX2B variants define a broad phenotypic spectrum. *Clin. Genet.,* 97**,** 103-113.

BATES, D., MäCHLER, M., BOLKER, B. & WALKER, S. 2015. Fitting Linear Mixed-Effects Models Using lme4. *J. Stat. Softw.,* 67**,** 1-48.

BENJAMINI, Y. & HOCHBERG, Y. 1995. Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing. *J. R. Stat. Soc. Series B Stat. Methodol.,* 57**,** 289-300.

BOCK, C., DATLINGER, P., CHARDON, F., COELHO, M. A., DONG, M. B., LAWSON, K. A., LU, T., MAROC, L., NORMAN, T. M., SONG, B., STANLEY, G., CHEN, S., GARNETT, M., LI, W., MOFFAT, J., QI, L. S., SHAPIRO, R. S., SHENDURE, J., WEISSMAN, J. S. & ZHUANG, X. 2022. High-content CRISPR screening. *Nat Rev Methods Primers,* 2.

BOEVA, V., LOUIS-BRENNETOT, C., PELTIER, A., DURAND, S., PIERRE-EUGèNE, C., RAYNAL, V., ETCHEVERS, H. C., THOMAS, S., LERMINE, A., DAUDIGEOS-DUBUS, E., GEOERGER, B., ORTH, M. F., GRüNEWALD, T. G. P., DIAZ, E., DUCOS, B., SURDEZ, D., CARCABOSO, A. M., MEDVEDEVA, I., DELLER, T., COMBARET, V., LAPOUBLE, E., PIERRON, G., GROSSETêTE-LALAMI, S., BAULANDE, S., SCHLEIERMACHER, G., BARILLOT, E., ROHRER, H., DELATTRE, O. & JANOUEIX-LEROSEY, I. 2017. Heterogeneity of neuroblastoma cell identity defined by transcriptional circuitries. *Nat. Genet.,* 49**,** 1408-1413.

BOIX, C. A., JAMES, B. T., PARK, Y. P., MEULEMAN, W. & KELLIS, M. 2021. Regulatory genomic circuitry of human disease loci by integrative epigenomics. *Nature,* 590**,** 300-307.

BRUNI, P., MINOPOLI, G., BRANCACCIO, T., NAPOLITANO, M. & OTHERS 2002. Fe65, a ligand of the Alzheimer's β-amyloid precursor protein, blocks cell cycle progression by down-regulating thymidylate synthase expression. *Journal of Biological*.

BUENROSTRO, J. D., WU, B., CHANG, H. Y. & GREENLEAF, W. J. 2015. ATAC-seq: A method for assaying chromatin accessibility genome-wide. *Curr. Protoc. Mol. Biol.,* 109**,** 21.29.1-21.29.9.

CAIRNS, J., FREIRE-PRITCHETT, P., WINGETT, S. W., VáRNAI, C., DIMOND, A., PLAGNOL, V., ZERBINO, D., SCHOENFELDER, S., JAVIERRE, B.-M., OSBORNE, C., FRASER, P. & SPIVAKOV, M. 2016. CHiCAGO: robust detection of DNA looping interactions in Capture Hi-C data. *Genome Biol.,* 17**,** 127.

CHAVEZ, A., SCHEIMAN, J., VORA, S., PRUITT, B. W., TUTTLE, M., P R IYER, E., LIN, S., KIANI, S., GUZMAN, C. D., WIEGAND, D. J., TER-OVANESYAN, D., BRAFF, J. L., DAVIDSOHN, N., HOUSDEN, B. E., PERRIMON, N., WEISS, R., AACH, J., COLLINS, J. J. & CHURCH, G. M. 2015. Highly efficient Cas9-mediated transcriptional programming. *Nat. Methods,* 12**,** 326-328.

CHAVEZ, A., TUTTLE, M., PRUITT, B. W., EWEN-CAMPEN, B., CHARI, R., TER-OVANESYAN, D., HAQUE, S. J., CECCHI, R. J., KOWAL, E. J. K., BUCHTHAL,

J., HOUSDEN, B. E., PERRIMON, N., COLLINS, J. J. & CHURCH, G. 2016. Comparison of Cas9 activators in multiple species. *Nat. Methods,* 13**,** 563-567.

CHEN, C.-C., CHEN, H.-Y., SU, K.-Y., HONG, Q.-S., YAN, B.-S., CHEN, C.-H., PAN, S.-H., CHANG, Y.-L., WANG, C.-J., HUNG, P.-F., YUAN, S., CHANG, G.-C., CHEN, J. J. W., YANG, P.-C., YANG, Y.-C. & YU, S.-L. 2014. Shisa3 is associated with prolonged survival through promoting β-catenin degradation in lung cancer. *Am. J. Respir. Crit. Care Med.,* 190**,** 433-444.

COHEN, N. M., OLIVARES-CHAUVET, P., LUBLING, Y., BARAN, Y., LIFSHITZ, A., HOICHMAN, M. & TANAY, A. 2017. SHAMAN: bin-free randomization, normalization and screening of Hi-C matrices. *bioRxiv*.

CONRAD, T. & ØROM, U. A. 2017. Cellular Fractionation and Isolation of Chromatin-Associated RNA. *Methods Mol. Biol.,* 1468**,** 1-9.

CRANE, E., BIAN, Q., MCCORD, R. P., LAJOIE, B. R., WHEELER, B. S., RALSTON, E. J., UZAWA, S., DEKKER, J. & MEYER, B. J. 2015. Condensin-driven remodelling of X chromosome topology during dosage compensation. *Nature,* 523**,** 240-244.

DATLINGER, P., RENDEIRO, A. F., SCHMIDL, C., KRAUSGRUBER, T., TRAXLER, P., KLUGHAMMER, J., SCHUSTER, L. C., KUCHLER, A., ALPAR, D. & BOCK, C. 2017. Pooled CRISPR screening with single-cell transcriptome readout. *Nat. Methods,* 14**,** 297-301.

DIAO, Y., FANG, R., LI, B., MENG, Z., YU, J., QIU, Y., LIN, K. C., HUANG, H., LIU, T., MARINA, R. J., JUNG, I., SHEN, Y., GUAN, K.-L. & REN, B. 2017. A tiling-deletion-based genetic screen for cis-regulatory element identification in mammalian cells. *Nat. Methods,* 14**,** 629-635.

DIXIT, A., PARNAS, O., LI, B., CHEN, J., FULCO, C. P., JERBY-ARNON, L., MARJANOVIC, N. D., DIONNE, D., BURKS, T., RAYCHOWDHURY, R., ADAMSON, B., NORMAN, T. M., LANDER, E. S., WEISSMAN, J. S., FRIEDMAN, N. & REGEV, A. 2016. Perturb-Seq: Dissecting Molecular Circuits with Scalable Single-Cell RNA Profiling of Pooled Genetic Screens. *Cell,* 167**,** 1853-1866.e17.

DOBIN, A., DAVIS, C. A., SCHLESINGER, F., DRENKOW, J., ZALESKI, C., JHA, S., BATUT, P., CHAISSON, M. & GINGERAS, T. R. 2013. STAR: ultrafast universal RNA-seq aligner. *Bioinformatics,* 29**,** 15-21.

DODT, M., ROEHR, J. T., AHMED, R. & DIETERICH, C. 2012. FLEXBAR—Flexible Barcode and Adapter Processing for Next-Generation Sequencing Platforms. *Biology,* 1**,** 895-905.

DOWNES, D. J., SMITH, A. L., KARPINSKA, M. A., VELYCHKO, T., RUE-ALBRECHT, K., SIMS, D., MILNE, T. A., DAVIES, J. O. J., OUDELAAR, A. M. & HUGHES, J. R. 2022. Capture-C: a modular and flexible approach for high-resolution chromosome conformation capture. *Nat. Protoc.,* 17**,** 445-475.

DURAND, N. C., SHAMIM, M. S., MACHOL, I., RAO, S. S. P., HUNTLEY, M. H., LANDER, E. S. & AIDEN, E. L. 2016. Juicer Provides a One-Click System for Analyzing Loop-Resolution Hi-C Experiments. *Cell Syst,* 3**,** 95-98.

DURBIN, A. D., ZIMMERMAN, M. W., DHARIA, N. V., ABRAHAM, B. J., INIGUEZ, A. B., WEICHERT-LEAHEY, N., HE, S., KRILL-BURGER, J. M., ROOT, D. E., VAZQUEZ, F., TSHERNIAK, A., HAHN, W. C., GOLUB, T. R., YOUNG, R. A.,

LOOK, A. T. & STEGMAIER, K. 2018. Selective gene dependencies in MYCN-amplified neuroblastoma include the core transcriptional regulatory circuitry. *Nat. Genet.,* 50**,** 1240-1246.

DUTTKE, S. H. C., LACADIE, S. A., IBRAHIM, M. M., GLASS, C. K., CORCORAN, D. L., BENNER, C., HEINZ, S., KADONAGA, J. T. & OHLER, U. 2015. Human promoters are intrinsically directional. *Mol. Cell,* 57**,** 674-684.

FIELD, A. & ADELMAN, K. 2020. Evaluating Enhancer Function and Transcription. *Annu. Rev. Biochem.,* 89**,** 213-234.

FULCO, C. P., MUNSCHAUER, M., ANYOHA, R., MUNSON, G., GROSSMAN, S. R., PEREZ, E. M., KANE, M., CLEARY, B., LANDER, E. S. & ENGREITZ, J. M. 2016. Systematic mapping of functional enhancer-promoter connections with CRISPR interference. *Science,* 354**,** 769-773.

FULCO, C. P., NASSER, J., JONES, T. R., MUNSON, G., BERGMAN, D. T., SUBRAMANIAN, V., GROSSMAN, S. R., ANYOHA, R., DOUGHTY, B. R., PATWARDHAN, T. A., NGUYEN, T. H., KANE, M., PEREZ, E. M., DURAND, N. C., LAREAU, C. A., STAMENOVA, E. K., AIDEN, E. L., LANDER, E. S. & ENGREITZ, J. M. 2019. Activity-by-contact model of enhancer–promoter regulation from thousands of CRISPR perturbations. *Nat. Genet.,* 51**,** 1664-1669.

GASPERINI, M., HILL, A. J., MCFALINE-FIGUEROA, J. L., MARTIN, B., KIM, S., ZHANG, M. D., JACKSON, D., LEITH, A., SCHREIBER, J., NOBLE, W. S., TRAPNELL, C., AHITUV, N. & SHENDURE, J. 2019. A Genome-wide Framework for Mapping Gene Regulation via Cellular Genetic Screens. *Cell,* 176**,** 377-390.e19.

GASPERINI, M., TOME, J. M. & SHENDURE, J. 2020. Towards a comprehensive catalogue of validated and target-linked human enhancers. *Nat. Rev. Genet.,* 21**,** 292-310.

GOLANSKA, E., SIERUTA, M., GRESNER, S. M., PFEFFER, A., CHODAKOWSKA-ZEBROWSKA, M., SOBOW, T. M., KLICH, I., MOSSAKOWSKA, M., SZYBINSKA, A., BARCIKOWSKA, M. & LIBERSKI, P. P. 2013. APBB2 genetic polymorphisms are associated with severe cognitive impairment in centenarians. *Exp. Gerontol.,* 48**,** 391-394.

HAFEMEISTER, C. & SATIJA, R. 2019. Normalization and variance stabilization of single-cell RNA-seq data using regularized negative binomial regression. *Genome Biol.,* 20**,** 296.

HELMSAUER, K., VALIEVA, M. E., ALI, S., CHAMORRO GONZáLEZ, R., SCHöPFLIN, R., RöEFZAAD, C., BEI, Y., DORADO GARCIA, H., RODRIGUEZ-FOS, E., PUIGGRòS, M., KASACK, K., HAASE, K., KESKENY, C., CHEN, C. Y., KUSCHEL, L. P., EUSKIRCHEN, P., HEINRICH, V., ROBSON, M. I., ROSSWOG, C., TOEDLING, J., SZYMANSKY, A., HERTWIG, F., FISCHER, M., TORRENTS, D., EGGERT, A., SCHULTE, J. H., MUNDLOS, S., HENSSEN, A. G. & KOCHE, R. P. 2020. Enhancer hijacking determines extrachromosomal circular MYCN amplicon architecture in neuroblastoma. *Nat. Commun.,* 11**,** 5823.

HENRICH, K.-O., BENDER, S., SAADATI, M., DREIDAX, D., GARTLGRUBER, M., SHAO, C., HERRMANN, C., WIESENFARTH, M., PARZONKA, M., WEHRMANN, L., FISCHER, M., DUFFY, D. J., BELL, E., TORKOV, A., SCHMEZER, P., PLASS, C., HöFER, T., BENNER, A., PFISTER, S. M. &

WESTERMANN, F. 2016. Integrative Genome-Scale Analysis Identifies Epigenetic Mechanisms of Transcriptional Deregulation in Unfavorable Neuroblastomas. *Cancer Res.,* 76**,** 5523-5537.

HILL, A. J., MCFALINE-FIGUEROA, J. L., STARITA, L. M., GASPERINI, M. J., MATREYEK, K. A., PACKER, J., JACKSON, D., SHENDURE, J. & TRAPNELL, C. 2018. On the design of CRISPR-based single-cell molecular screens. *Nat. Methods,* 15**,** 271-274.

HILTON, I. B., D'IPPOLITO, A. M., VOCKLEY, C. M., THAKORE, P. I., CRAWFORD, G. E., REDDY, T. E. & GERSBACH, C. A. 2015. Epigenome editing by a CRISPR-Cas9-based acetyltransferase activates genes from promoters and enhancers. *Nat. Biotechnol.,* 33**,** 510-517.

HOADLEY, K. A., YAU, C., HINOUE, T., WOLF, D. M., LAZAR, A. J., DRILL, E., SHEN, R., TAYLOR, A. M., CHERNIACK, A. D., THORSSON, V., AKBANI, R., BOWLBY, R., WONG, C. K., WIZNEROWICZ, M., SANCHEZ-VEGA, F., ROBERTSON, A. G., SCHNEIDER, B. G., LAWRENCE, M. S., NOUSHMEHR, H., MALTA, T. M., CANCER GENOME ATLAS, N., STUART, J. M., BENZ, C. C. & LAIRD, P. W. 2018. Cell-of-Origin Patterns Dominate the Molecular Classification of 10,000 Tumors from 33 Types of Cancer. *Cell,* 173**,** 291-304.e6.

HORLBECK, M. A., WITKOWSKY, L. B., GUGLIELMI, B., REPLOGLE, J. M., GILBERT, L. A., VILLALTA, J. E., TORIGOE, S. E., TJIAN, R. & WEISSMAN, J. S. 2016. Nucleosomes impede Cas9 access to DNA in vivo and in vitro. *Elife,* 5.

IBRAHIM, M. M., KARABACAK, A., GLAHS, A., KOLUNDZIC, E., HIRSEKORN, A., CARDA, A., TURSUN, B., ZINZEN, R. P., LACADIE, S. A. & OHLER, U. 2018. Determinants of promoter and enhancer transcription directionality in metazoans. *Nat. Commun.,* 9**,** 4472.

IBRAHIM, M. M., LACADIE, S. A. & OHLER, U. 2015. JAMM: a peak finder for joint analysis of NGS replicates. *Bioinformatics,* 31**,** 48-55.

JOUNG, J., KONERMANN, S., GOOTENBERG, J. S., ABUDAYYEH, O. O., PLATT, R. J., BRIGHAM, M. D., SANJANA, N. E. & ZHANG, F. 2017. Genome-scale CRISPR-Cas9 knockout and transcriptional activation screening. *Nat. Protoc.,* 12**,** 828-863.

KARABACAK CALVIELLO, A., HIRSEKORN, A., WURMUS, R., YUSUF, D. & OHLER, U. 2019. Reproducible inference of transcription factor footprints in ATAC-seq and DNase-seq datasets using protocol-specific bias modeling. *Genome Biol.,* 20**,** 42.

KAUFMANN, J. & SCHERING, A. G. 2014. Analysis of Variance ANOVA. *Wiley StatsRef: Statistics Reference Online.* Chichester, UK: John Wiley & Sons, Ltd.

KE, X.-X., ZHANG, D., ZHAO, H., HU, R., DONG, Z., YANG, R., ZHU, S., XIA, Q., DING, H.-F. & CUI, H. 2015. Phox2B correlates with MYCN and is a prognostic marker for neuroblastoma development. *Oncol. Lett.,* 9**,** 2507-2514.

KEARNS, N. A., PHAM, H., TABAK, B., GENGA, R. M., SILVERSTEIN, N. J., GARBER, M. & MAEHR, R. 2015. Functional annotation of native enhancers with a Cas9–histone demethylase fusion. *Nat. Methods,* 12**,** 401-403.

KIM, S. & WYSOCKA, J. 2023. Deciphering the multi-scale, quantitative cis-regulatory code. *Mol. Cell,* 83**,** 373-392.

KIM, T.-K. & SHIEKHATTAR, R. 2015. Architectural and Functional Commonalities between Enhancers and Promoters. *Cell,* 162**,** 948-959.

KWAN, S.-Y., AU-YEUNG, C.-L., YEUNG, T.-L., RYNNE-VIDAL, A., WONG, K.-K., RISINGER, J. I., LIN, H.-K., SCHMANDT, R. E., YATES, M. S., MOK, S. C. & LU, K. H. 2020. Ubiquitin Carboxyl-Terminal Hydrolase L1 (UCHL1) Promotes Uterine Serous Cancer Cell Proliferation and Cell Cycle Progression. *Cancers,* 12.

LACADIE, S. A., IBRAHIM, M. M., GOKHALE, S. A. & OHLER, U. 2016. Divergent transcription and epigenetic directionality of human promoters. *FEBS J.,* 283**,** 4214-4222.

LANGMEAD, B. & SALZBERG, S. L. 2012. Fast gapped-read alignment with Bowtie 2. *Nat. Methods,* 9**,** 357-359.

LANGMEAD, B., TRAPNELL, C., POP, M. & SALZBERG, S. L. 2009. Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. *Genome Biol.,* 10**,** R25.

LI, Y., HOLLINGWORTH, P., MOORE, P., FOY, C., ARCHER, N., POWELL, J., NOWOTNY, P., HOLMANS, P., O'DONOVAN, M., TACEY, K., DOIL, L., VAN LUCHENE, R., GARCIA, V., ROWLAND, C., LAU, K., CANTANESE, J., SNINSKY, J., HARDY, J., THAL, L., MORRIS, J. C., GOATE, A., LOVESTONE, S., OWEN, M., WILLIAMS, J. & GRUPE, A. 2005. Genetic association of the APP binding protein 2 gene (APBB2) with late onset Alzheimer disease. *Hum. Mutat.,* 25**,** 270-277.

MAIR, F., ERICKSON, J. R., VOILLET, V., SIMONI, Y., BI, T., TYZNIK, A. J., MARTIN, J., GOTTARDO, R., NEWELL, E. W. & PRLIC, M. 2020. A Targeted Multi-omic Analysis Approach Measures Protein Expression and Low-Abundance Transcripts on the Single-Cell Level. *Cell Rep.,* 31**,** 107499.

MöLDER, F., JABLONSKI, K. P., LETCHER, B., HALL, M. B., TOMKINS-TINCH, C. H., SOCHAT, V., FORSTER, J., LEE, S., TWARDZIOK, S. O., KANITZ, A., WILM, A., HOLTGREWE, M., RAHMANN, S., NAHNSEN, S. & KöSTER, J. 2021. Sustainable data analysis with Snakemake. *F1000Res.,* 10**,** 33.

MONDAL, M., CONOLE, D., NAUTIYAL, J. & TATE, E. W. 2022. UCHL1 as a novel target in breast cancer: emerging insights from cell and chemical biology. *Br. J. Cancer,* 126**,** 24-33.

MORRIS, J. A., CARAGINE, C., DANILOSKI, Z., DOMINGO, J., BARRY, T., LU, L., DAVIS, K., ZIOSI, M., GLINOS, D. A., HAO, S., MIMITOU, E. P., SMIBERT, P., ROEDER, K., KATSEVICH, E., LAPPALAINEN, T. & SANJANA, N. E. 2023. Discovery of target genes and pathways at GWAS loci by pooled single-cell CRISPR screens. *Science,* 380**,** eadh7699.

NAFTALI, O., MAMAN, S., MESHEL, T., SAGI-ASSIF, O., GINAT, R. & WITZ, I. P. 2016. PHOX2B is a suppressor of neuroblastoma metastasis. *Oncotarget,* 7**,** 10627-10637.

NASSER, J., BERGMAN, D. T., FULCO, C. P., GUCKELBERGER, P., DOUGHTY, B. R., PATWARDHAN, T. A., JONES, T. R., NGUYEN, T. H., ULIRSCH, J. C., LEKSCHAS, F., MUALIM, K., NATRI, H. M., WEEKS, E. M., MUNSON, G., KANE, M., KANG, H. Y., CUI, A., RAY, J. P., EISENHAURE, T. M., COLLINS, R. L., DEY, K., PFISTER, H., PRICE, A. L., EPSTEIN, C. B., KUNDAJE, A.,

XAVIER, R. J., DALY, M. J., HUANG, H., FINUCANE, H. K., HACOHEN, N., LANDER, E. S. & ENGREITZ, J. M. 2021. Genome-wide enhancer maps link risk variants to disease genes. *Nature,* 593**,** 238-243.

NUñEZ, J. K., CHEN, J., POMMIER, G. C., COGAN, J. Z., REPLOGLE, J. M., ADRIAENS, C., RAMADOSS, G. N., SHI, Q., HUNG, K. L., SAMELSON, A. J., POGSON, A. N., KIM, J. Y. S., CHUNG, A., LEONETTI, M. D., CHANG, H. Y., KAMPMANN, M., BERNSTEIN, B. E., HOVESTADT, V., GILBERT, L. A. & WEISSMAN, J. S. 2021. Genome-wide programmable transcriptional memory by CRISPR-based epigenome editing. *Cell,* 184**,** 2503-2519.e17.

PANIGRAHI, A. & O'MALLEY, B. W. 2021. Mechanisms of enhancer action: the known and the unknown. *Genome Biol.,* 22**,** 108.

PEREZ, A. R., PRITYKIN, Y., VIDIGAL, J. A., CHHANGAWALA, S., ZAMPARO, L., LESLIE, C. S. & VENTURA, A. 2017. GuideScan software for improved single and paired CRISPR guide RNA design. *Nature Biotechnology,* 35**,** 347-349.

PHANSTIEL, D. H., BOYLE, A. P., ARAYA, C. L. & SNYDER, M. P. 2014. Sushi.R: flexible, quantitative and integrative genomic visualizations for publication-quality multi-panel figures. *Bioinformatics,* 30**,** 2808-2810.

PONZONI, M., BACHETTI, T., CORRIAS, M. V., BRIGNOLE, C., PASTORINO, F., CALARCO, E., BENSA, V., GIUSTO, E., CECCHERINI, I. & PERRI, P. 2022. Recent advances in the developmental origin of neuroblastoma: an overview. *J. Exp. Clin. Cancer Res.,* 41**,** 92.

PREISSL, S., GAULTON, K. J. & REN, B. 2023. Characterizing cis-regulatory elements using single-cell epigenomics. *Nat. Rev. Genet.,* 24**,** 21-43.

QUINLAN, A. R. & HALL, I. M. 2010. BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics,* 26**,** 841-842.

REPLOGLE, J. M., NORMAN, T. M., XU, A., HUSSMANN, J. A., CHEN, J., COGAN, J. Z., MEER, E. J., TERRY, J. M., RIORDAN, D. P., SRINIVAS, N., FIDDES, I. T., ARTHUR, J. G., ALVARADO, L. J., PFEIFFER, K. A., MIKKELSEN, T. S., WEISSMAN, J. S. & ADAMSON, B. 2020. Combinatorial single-cell CRISPR screens by direct guide RNA capture and targeted sequencing. *Nat. Biotechnol.,* 38**,** 954-961.

SANJANA, N. E. 2016. Genome-scale CRISPR pooled screens. *Anal. Biochem.***,** 1-5.

SANJANA, N. E., SHALEM, O. & ZHANG, F. 2014. Improved vectors and genome-wide libraries for CRISPR screening. *Nat. Methods,* 11**,** 783-784.

SCHRAIVOGEL, D., GSCHWIND, A. R., MILBANK, J. H., LEONCE, D. R., JAKOB, P., MATHUR, L., KORBEL, J. O., MERTEN, C. A., VELTEN, L. & STEINMETZ, L. M. 2020. Targeted Perturb-seq enables genome-scale genetic screens in single cells. *Nat. Methods,* 17**,** 629-635.

SCHRAIVOGEL, D., STEINMETZ, L. M. & PARTS, L. 2023. Pooled Genome-Scale CRISPR Screens in Single Cells. *Annu. Rev. Genet.,* 57**,** 223-244.

SERVANT, N., LAJOIE, B. R., NORA, E. P., GIORGETTI, L., CHEN, C.-J., HEARD, E., DEKKER, J. & BARILLOT, E. 2012. HiTC: exploration of high-throughput 'C' experiments. *Bioinformatics,* 28**,** 2843-2844.

SHAHZAD, N., MUNIR, T., JAVED, M., TASNEEM, F., ASLAM, B., ALI, M., MUTAHIR, Z., AKHTAR ALI, M., UMER, M., AHMAD, M., FAROOQ, K., HASSAN, U., MUSTAFA, T., ANJUM, R. S. & SHAKOORI, A. R. 2020. SHISA3, an antagonist

of the Wnt/β-catenin signaling, is epigenetically silenced and its ectopic expression suppresses growth in breast cancer. *PLoS One,* 15**,** e0236192.

SHLYUEVA, D., STAMPFEL, G. & STARK, A. 2014. Transcriptional enhancers: from properties to genome-wide predictions. *Nat. Rev. Genet.,* 15**,** 272-286.

SI, J., MA, Y., BI, J. W., XIONG, Y., LV, C., LI, S., WU, N. & YANG, Y. 2019. Shisa3 brakes resistance to EGFR-TKIs in lung adenocarcinoma by suppressing cancer stem cell properties. *J. Exp. Clin. Cancer Res.,* 38**,** 481.

SMITH, T., HEGER, A. & SUDBERY, I. 2017. UMI-tools: modeling sequencing errors in Unique Molecular Identifiers to improve quantification accuracy. *Genome Res.,* 27**,** 491-499.

STUART, T., BUTLER, A., HOFFMAN, P., HAFEMEISTER, C., PAPALEXI, E., MAUCK, W. M., 3RD, HAO, Y., STOECKIUS, M., SMIBERT, P. & SATIJA, R. 2019. Comprehensive Integration of Single-Cell Data. *Cell,* 177**,** 1888-1902.e21.

TANENBAUM, M. E., GILBERT, L. A., QI, L. S., WEISSMAN, J. S. & VALE, R. D. 2014. A protein-tagging system for signal amplification in gene expression and fluorescence imaging. *Cell,* 159**,** 635-646.

TANG, S.-H., HSIAO, C.-W., CHEN, W.-L., WU, L.-W., CHANG, J.-B. & YANG, B.-H. 2021. Hypermethylation of SHISA3 DNA as a blood-based biomarker for colorectal cancer. *Chin. J. Physiol.,* 64**,** 51-56.

TSAI, M.-H., CHEN, W.-C., YU, S.-L., CHEN, C.-C., JAO, T.-M., HUANG, C.-Y., TZENG, S.-T., YEN, S.-J. & YANG, Y.-C. 2015. DNA Hypermethylation of SHISA3 in Colorectal Cancer: An Independent Predictor of Poor Prognosis. *Ann. Surg. Oncol.,* 22 Suppl 3**,** S1481-9.

VAROQUAUX, G., BUITINCK, L., LOUPPE, G., GRISEL, O., PEDREGOSA, F. & MUELLER, A. 2015. Scikit-learn. *GetMob. Mob. Comput. Commun.,* 19**,** 29-33.

VUčIćEVIć, D., GEHRE, M., DHAMIJA, S., FRIIS-HANSEN, L., MEIERHOFER, D., SAUER, S. & ØROM, U. A. 2016. The long non-coding RNA PARROT is an upstream regulator of c-Myc and affects proliferation and translation. *Oncotarget,* 7**,** 33934-33947.

WANG, L., TAN, T. K., DURBIN, A. D., ZIMMERMAN, M. W., ABRAHAM, B. J., TAN, S. H., NGOC, P. C. T., WEICHERT-LEAHEY, N., AKAHANE, K., LAWTON, L. N., ROKITA, J. L., MARIS, J. M., YOUNG, R. A., LOOK, A. T. & SANDA, T. 2019. ASCL1 is a MYCN- and LMO1-dependent member of the adrenergic neuroblastoma core regulatory circuitry. *Nat. Commun.,* 10**,** 5622.

WICKHAM, H. 2016. ggplot2: Elegant Graphics for Data Analysis. Springer-Verlag, New York. ISBN 978-3-319-24277-4.

WINDELS, M.-L., CORDIER, F., VAN DORPE, J., FERDINANDE, L. & CREYTENS, D. 2024. PHOX2B: a diagnostic cornerstone in neurocristopathies and neuroblastomas. *J. Clin. Pathol.,* 77**,** 378-382.

WINGETT, S., EWELS, P., FURLAN-MAGARIL, M., NAGANO, T., SCHOENFELDER, S., FRASER, P. & ANDREWS, S. 2015. HiCUP: pipeline for mapping and processing Hi-C data. *F1000Res.,* 4**,** 1310.

WU, Q., WU, J., KARIM, K., CHEN, X., WANG, T., IWAMA, S., CAROBBIO, S., KEEN, P., VIDAL-PUIG, A., KOTTER, M. R. & BASSETT, A. 2023. Massively parallel characterization of CRISPR activator efficacy in human induced pluripotent stem cells and neurons. *Mol. Cell,* 83**,** 1125-1139.e8.

XIE, S., DUAN, J., LI, B., ZHOU, P. & HON, G. C. 2017. Multiplexed Engineering and Analysis of Combinatorial Enhancer Activity in Single Cells. *Mol. Cell,* 66**,** 285-299.e5.

YEO, N. C., CHAVEZ, A., LANCE-BYRNE, A., CHAN, Y., MENN, D., MILANOVA, D., KUO, C.-C., GUO, X., SHARMA, S., TUNG, A., CECCHI, R. J., TUTTLE, M., PRADHAN, S., LIM, E. T., DAVIDSOHN, N., EBRAHIMKHANI, M. R., COLLINS, J. J., LEWIS, N. E., KIANI, S. & CHURCH, G. M. 2018. An enhanced CRISPR repressor for targeted mammalian gene regulation. *Nat. Methods,* 15**,** 611-616.

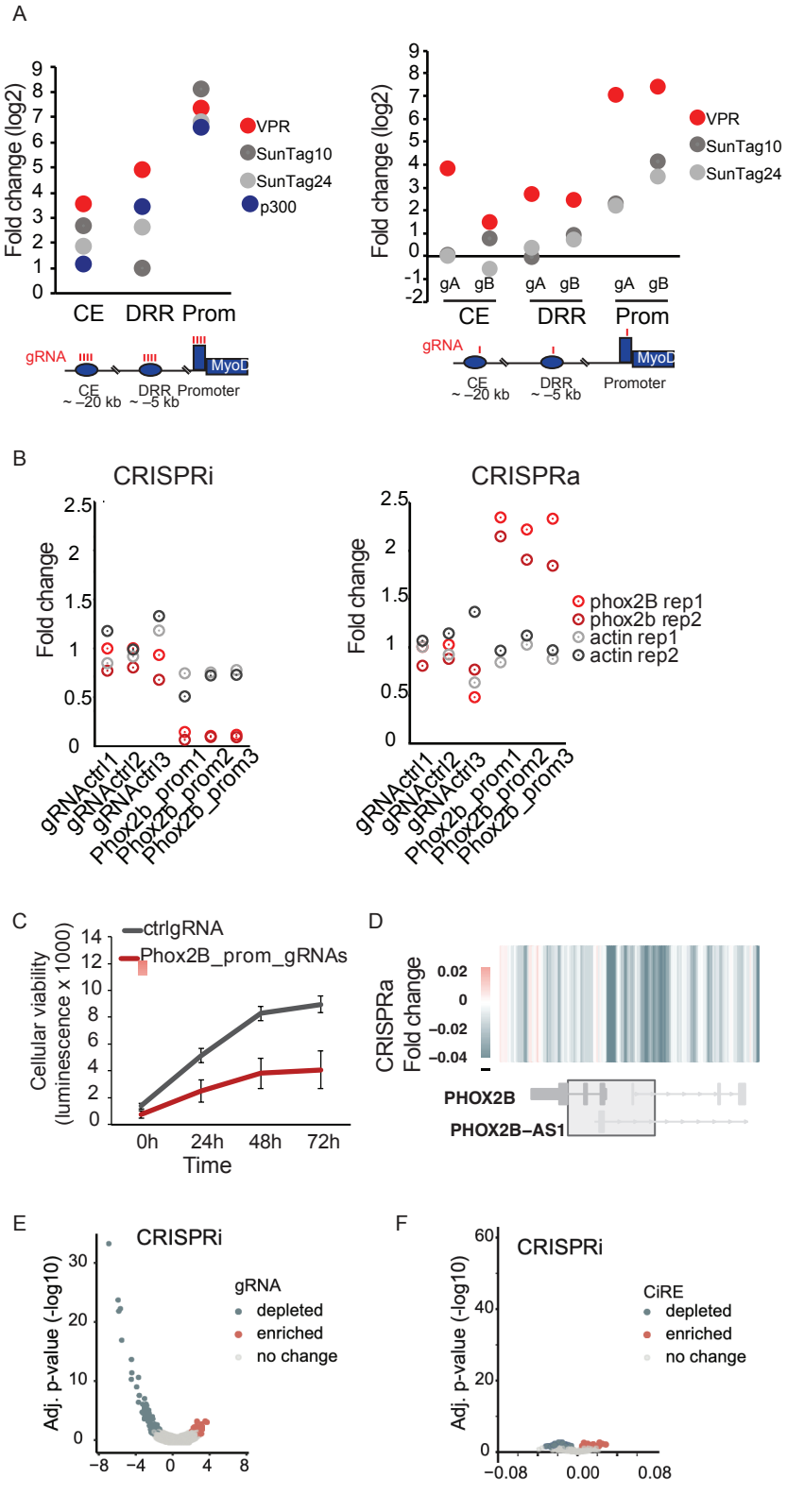
ZHANG, J., LEE, D., DHIMAN, V., JIANG, P., XU, J., MCGILLIVRAY, P., YANG, H., LIU, J., MEYERSON, W., CLARKE, D., GU, M., LI, S., LOU, S., XU, J., LOCHOVSKY, L., UNG, M., MA, L., YU, S., CAO, Q., HARMANCI, A., YAN, K.-K., SETHI, A., GüRSOY, G., SCHOENBERG, M. R., ROZOWSKY, J., WARRELL, J., EMANI, P., YANG, Y. T., GALEEV, T., KONG, X., LIU, S., LI, X., KRISHNAN, J., FENG, Y., RIVERA-MULIA, J. C., ADRIAN, J., BROACH, J. R., BOLT, M., MORAN, J., FITZGERALD, D., DILEEP, V., LIU, T., MEI, S., SASAKI, T., TREVILLA-GARCIA, C., WANG, S., WANG, Y., ZANG, C., WANG, D., KLEIN, R. J., SNYDER, M., GILBERT, D. M., YIP, K., CHENG, C., YUE, F., LIU, X. S., WHITE, K. P. & GERSTEIN, M. 2020. An integrative ENCODE resource for cancer genomics. *Nat. Commun.,* 11**,** 3696.

ZHANG, J., LI, Y.-Q., GUO, R., WANG, Y.-Q., ZHANG, P.-P., TANG, X.-R., WEN, X., HONG, X.-H., LEI, Y., HE, Q.-M., YANG, X.-J., SUN, Y., MA, J. & LIU, N. 2019. Hypermethylation of SHISA3 Promotes Nasopharyngeal Carcinoma Metastasis by Reducing SGSM1 Stability. *Cancer Res.,* 79**,** 747-759.

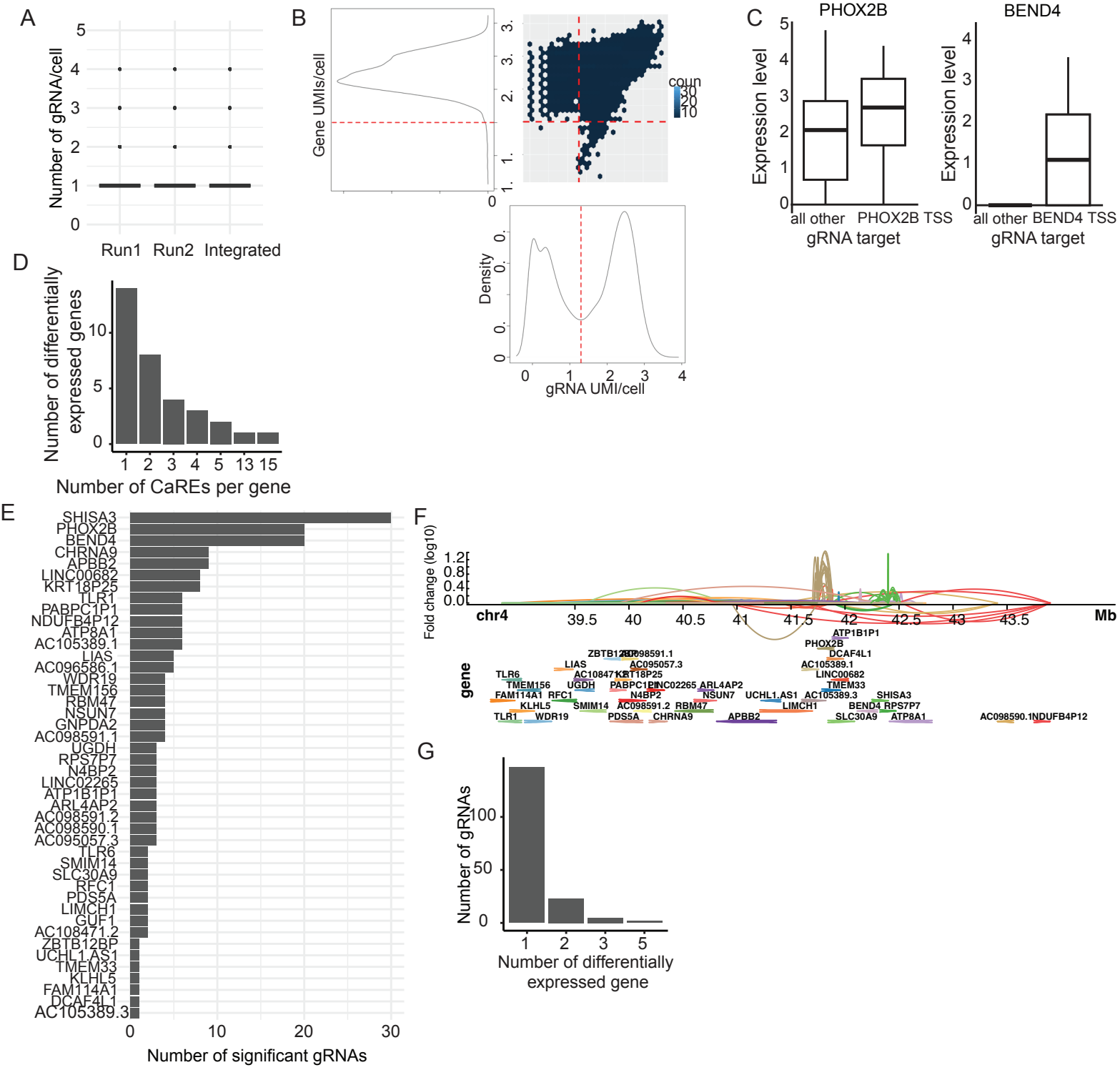
ZHANG, S., YU, B., SHENG, C., YAO, C., LIU, Y., WANG, J., ZENG, Q., MAO, Y., BEI, J., ZHU, B. & CHEN, S. 2024. SHISA3 reprograms tumor-associated macrophages toward an antitumoral phenotype and enhances cancer immunotherapy. *Adv. Sci.***,** e2403019.

ZHOU, K., CAO, D., WANG, Y., WANG, L. & MENG, X. 2021. Hsa-miR-30a-3p attenuates gastric adenocarcinoma proliferation and metastasis via APBB2. *Aging,* 13**,** 16763-16772.

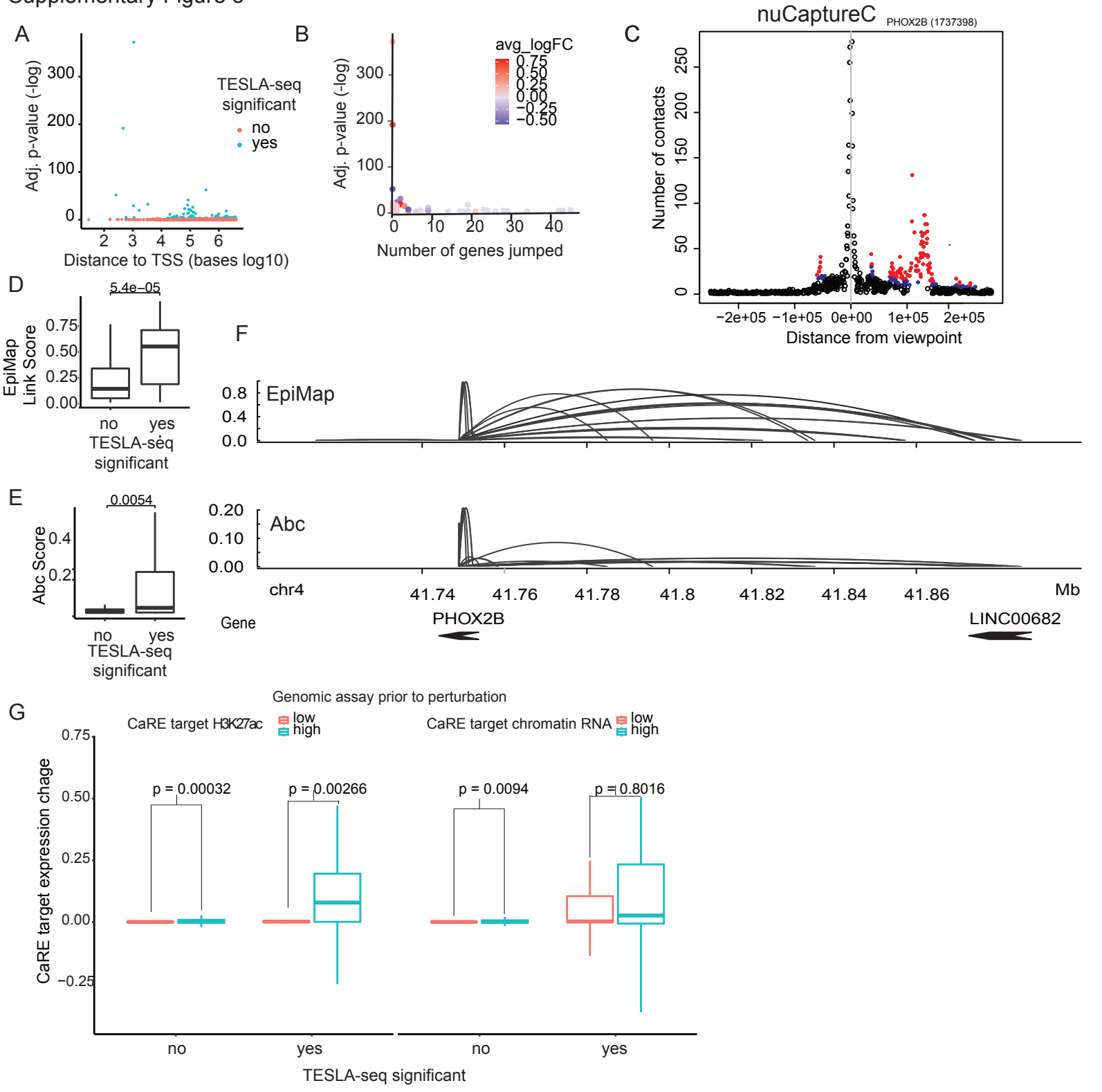
Supplementary figure 1



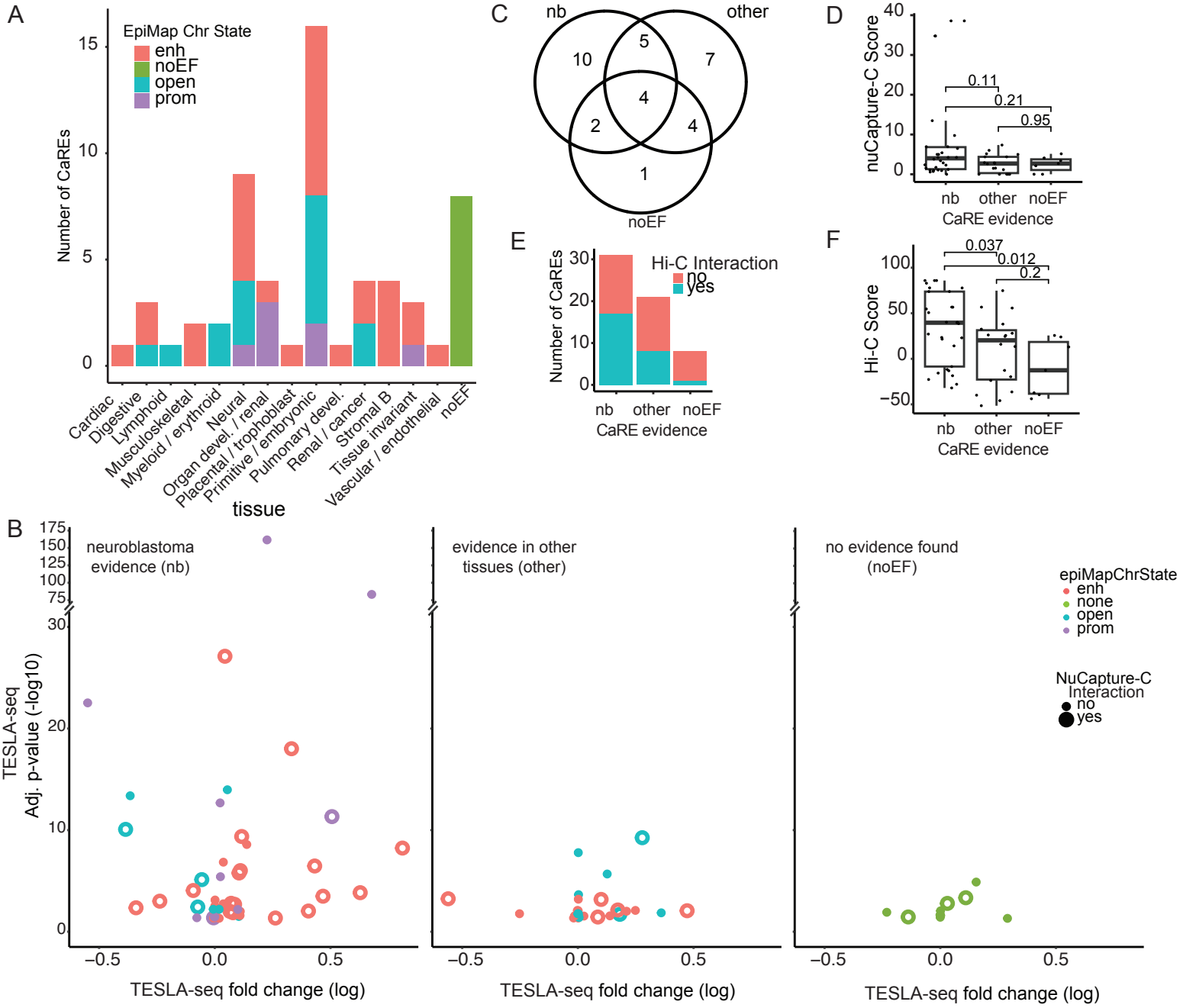
Supplementary Figure 2



Supplementary Figure 3



Supplementary Figure 4



Supplementary Figure legends.

Figure S1. Details of the bulk CRISPR activation screen related to Figure 1. A) Upregulation of MyoD in HEK293 cells via different indicated CRISPRa constructs at three different non-coding loci via pools of four gRNAs (left) or individual gRNAs (right). Relative expression of target genes is determined by RT- qPCR and normalized to GAPDH. Shown are fold changes relative to the control at a log2 scale. **B)** PHOX2B expression changes upon repression with dCas9-KRAB (left) or activation with dCas9-VPR (right) with control or gRNAs targeting the promoter of Phox2B in SH-SY5Y cell line. Relative expression of target genes is determined by RT- qPCR and normalised to GAPDH. Shown are fold changes relative to the control. Expression of actin is shown as a control. **C)** MTT viability assay with control or gRNAs targeting the promoter of Phox2B in a SH-SY5Y-KRAB (CRISPRi) cell line. **D)** CaREs signal at the PHOX2B promoter. Significantly enriched/depleted CaREs (FDR<0.05) are shown in red and blue respectively. **E)** Volcano plot showing the log-fold change of gRNA representation between the first and last time-point of the CRISPRi experiment. Significantly enriched/depleted gRNAs (FDR<0.05) are shown in red and blue respectively. **F)** Volcano plot showing the results for our CRISPRi screen, each dot corresponds to a CaRE defined by the CRISPRa screen. The x axis shows the slope calculated by MLM. Significantly enriched/depleted CaREs (FDR<0.05) are shown in red and blue respectively.

Figure S2. Details of the TESLA-seq related to Figure 2. A) Number of gRNA detected per cell in each TESLA-seq run and both together (integrated). **B)** Cell density plot. The red line represents thresholds used for cell filtering. Cells having either gene UMI count per cell or gRNA UMI counts per cell below thresholds are filtered. **C)** TESLA-seq normalized gene expression level, comparing cells having gRNA targeting PHOX2B and BEND4 promoter versus all other cells. **D)** Histogram showing the number of CaRE that induce differential expression of an indicated number ofgene. **E)** Number of gRNA that significantly cause differential gene expression of an indicated gene. **F)**. Genome browser snapshot representing genomic distance and regulatory relationship between a gRNA and a gene determined by TESLA-seq. Each link represents an effect of a gRNA on the gene. Color reflects the target gene. **G)** Histogram showing the number of genes that are differentially expressed due to perturbation of an indicated number of gRNA.

Figure S3. Additional details of properties of CaREs and their targets identified by TESLA-seq related to Figure 3. A) Relationship between adjusted p-values (-log10) of each CaRE-gene regulatory pair and genomic distance between the CaRE and the gene's transcriptional start site (in bases at log10 scale). **B)** Relationship between adjusted p-values (-log10) of each CaRE-gene regulatory pair and number of genes located between the CaRE and the gene (jumped genes). Color depicts the average fold change. **C)** Number of contacts detected for PHOX2B by nuCaptureC centered on the PHOX2B TSS (red: highly significant interactions (pLevel = 5, blue: significant interactions, pLevel = 3, black: background interactions**). D)** Comparison of EpiMap link score between TESLA-seq significant and non-significant CaRE-gene regulatory pairs**. E)** Comparison of Abc score between significant and TESLA-seq non-significant CaRE-gene regulation pairs**. F)** Genome browser snapshot of PHOX2B regulatory regions determined by EpiMap (top) and Abc (bottom). **G)** Comparison of the CaRE target gene expression (average fold change at log 2

scale) between TESLA-seq significant and non-significant CaRE-gene regulation pairs that have a low or high: H3K27 signal at the promoter of the target gene (left) and chromatin RNA signal (right) (non-significant).

Figure S4. Additional details of the integration of TESLA-seq results with epigenomic data related to Figure 4. A) Number of CaREs defined by EpiMap in each indicated tissue. The color indicates the EpiMap chromatin state. **B)** Comparison between CaRE evidence categories (nb, other, noEF) relationship between TESLA-seq adjusted p-value (-log10 scale) and fold change (log2 scale). The data is further stratified by: nuCapture-C interaction displayed via circle fill and class defined by EpiMap indicated via circle colour. **C)** Venn diagram displaying an overlap between genes targeted by CaREs in nb, other or noEF classes. **D)** CaRE-gene nuCapture-C score stratified by CaRE evidence class. **E)** Number of CaREs in each evidence categories (nb, other, noEF) grouped by Hi-C detected interaction (yes - 3D interaction is detected or no - no interaction detected). **F)** Comparison between different CaRE evidence categories (nb, other, noEF) in Hi-C score.

name	sequence
phox2b_qPCR_FW	GGAGACTCACTACCCCGACA
phox2b_qPCR_RV	CTCCTGCTTGCGAAACTTG
myoD_qFW	AGCACTACAGCGGCGACT
myoD_qRV	GCGCCTTCGTAGCAGTTC
APBB2_qFW	TGCTGGTAACGTGTCTGAGG
APBB2_qRV	GGAGGTGGTCGAACTTTCTG
SHISA3_qFW	GAGCACCCAGGCATCACT
SHISA3_qRV	AACAGGTGCAACAATAAATAGCC
actin_FW	CGACAGGATGCAGAAGGAG
actin_RV	GTACTTGCGCTCAGGAGGAG
gapdh_FW	GCTCTCTGCTCCTCCTGTTC
gapdh_RV	ACGACCAAATCCGTTGACTC
phox2b_prom1REVgRNA	aaactgccctttaattcaatcacaC
phox2b_prom1FWDgRNA	CACCGtgtgattgaattaaagggca
phox2b_prom2REVgRNA	aaacccttctaaccagctccctgC
phox2b_prom2FWDgRNA	CACCGcagggagctggttagaaggg
phox2b_prom3REVgRNA	aaacctgatcctcccttctaaccaC
phox2b_prom3FWDgRNA	CACCGtggttagaagggaggatcag
phox2b_prom4REVgRNA	aaacCCCTATCATTGATTCCTGCAC
phox2b_prom4FWDgRNA	CACCGTGCAGGAATCAATGATAGGG
phox2b_prom5FWDgRNA	CACCGGAATCAATGATAGGGAGGT
phox2b_prom5REVgRNA	aaacACCTCCCTATCATTGATTCC
phox2b_prom6FWDgRNA	CACCGCGTCTATTGGGCTGGCACTG
phox2b_prom6REVgRNA	aaacCAGTGCCAGCCCAATAGACGC
Neg_ctrl_gRNA_FWgRNA	caccGTATTACTGATATTGGTGGG
Neg_ctrl_gRNA_RVgRNA	aaacCCCACCAATATCAGTAATAC

Table 1. Primers and individual gRNAs oligoes.

name	sequence
HIC_chr4_91471	GATCAATAGCCCCAAGGTCACTAATCATTAATCTAGACAAAATAGTAAATACTAAAAATCTAAATAACCC
HIC_chr4_91683	GATCTCTTTTTCAGAAGAAGAGAAAATAACTTGTAAATGAACGGAAATACAATTGAAAATGCTTGACAAA
HIC_chr4_91683	TACTGGTTGATACATATTTATGACTACATATTTAGCCATATGTACGAGCAATCATACCCACATTCTGATC
HIC_chr4_91712	GATCTGGCGAACCCCGAGACCCACCCGCCCTGGGCTGGAGAGGCGGCGCGCTCTGGCTTTCCGCGCTGGC
HIC_chr4_91712	ACACCCACACGCCCACACTCAGGGTCTGCCCCCTCGGCCTGCGTGAACCTCCGCGGAGCCTGCCTGGATC
HIC_chr4_92179	GGTGAATGTCAGTCATGTCTAGGTTGCATGCACATGACTCATTCATTTACTTATTAAATGATTGTCGATC
HIC_chr4_92220	GATCATAATTTCGATGCCACAAACCCACCACCTAGAGCACACTAATTATCAGATTATTGCAAGGGGAATG
HIC_chr4_92220	TTCCCGAGACCCGGCGTGCCTGGTGAAACTTTTGCATTTAGGCATTTATTTCAGTGCATGCTCTGGGATC
HIC_chr4_92630	GATCCAAACTCTAAAGGAAAACTATTGGAGCTCATTTTTCCTGCCCACGTTAGCTAGCACATTTTGTTGA
HIC_chr4_92630	GACTGCAGACGCCGTAGAAGCGGTGCAGAAAGTGGGAACCCCTCCCTGGCCGAAATGAGCGGACTGGATC
HIC_chr4_93108	GATCCATTCGCGCCAACAACTTCTCCCGCGAAGTGCAAGAAGGCGAAGACAGTGGCGCGCGGTGATGACG
HIC_chr4_93108	GACATCGAGAGGAACCACGGTGCTTTGGTGTGAGCTCGGTTTTTGGCGGGGGCGCTAAAGTAGGGGGATC
HIC_chr4_93406	GATCATGTTCTGTACATGACTACAAATAGTCCGAACGGTAGCCAGTTCCTTTCTGTTACCCCACCATTTG
HIC_chr4_93406	AGCAGCTTGGCTGCTTGTCATAAATGGAGCGACGTAATTTCGACCTGTCCTTTCCCGGGAGTTAGCGATC
HIC_chr4_93611	GATCTGCGGGTCGGTCTCGGCGGGCGGATTACTCCTCGCCCACAACCTCGAGGCCCCGGGGTCAACCGCG
HIC_chr4_93611	AAGAAAGTGATTAGTAATTTATAAAGGCACATTAAAATACCATATAAATTACGGTTTTCTTGATAGGATC
HIC_chr4_93953	GATCTGCCGGCTCCCGCGGGCGGCGGCGCAACAGATTGCAGCGCCTGGAGACTCCAGCTGCCCCGCCTGC
HIC_chr4_94155	GATCCGGGAGAGCGTTTTCCTGCGCTAGACACGGCGTTCAGCCTCCGGGTTCCGGGTCTAGCTGAGTCAG
HIC_chr4_94155	GAGGGAGCAGGCCGAGCCTGAGAAAACCCGGGAAGTGGGTTGGGGGAAGGGGAAAGGTGGTAACTGGATC
HIC_chr4_94380	GTAGCTCTGGGATAGAGAAAACTTCCCAAGGGATGCTGATGCTGCTGGTTGGATGAACAACACTGAGATC
HIC_chr4_95059	GATCAATTTTAAACAGCAGGAACACCAATGGCACTGTTAACTGCTTTCTGGGTAGCCTCTTTAGCTTGGT
HIC_chr4_95073	GATCGAACCCCGCCGGCCCCGCCAACCCGGCCCTGCCAGCCCCAGGCTCACGGGGCTCCTCCGCACCAGC
HIC_chr4_95073	AGACGAAACTGGTTGGGAGGTAGTTAATTGCTCGGTGAAAATGAACTGATTTCCTCCTCAGGGAGAGATC
HIC_chr4_95258	CTACGGCGCTCGGCCAGTCAGCAGCTCTGCCAGCATCTATGCAGGTGCCGGGGGCTCTGGTTCCTGGATC
HIC_chr4_95387	GATCCACCTGGGCGACGTGGCAGCCATTGCCGGCGCCTCGAGACCCCTCCCCACCCGGCCGCCCACCCGC
HIC_chr4_95387	AGCGGCGGGCGGCGCGGCGAATGAACCCCCAAGCCCTGAATGTGGGGCCCGGCGCGGGCCGCCTCCGATC
HIC_chr4_96122	GATCATTGTGACATTATAATGGTGAAAAATTAGAAATAACCTAAACATCTAGCATTTAGAGAATGATTAA
HIC_chr4_96122	GCTTTGCCACCCAGTTTTGGTCATGGGGACTTGAGTTACTGTGACCTCAGAGGCCCCTAGGGGCCTGATC
HIC_chr4_96184	GATCTCGGTCCTGCATGCAATGCAAGCCTGAGCTCTCCCGCCATAAGGCTGCAGCGGTGTGGGCTCCTTG
HIC_chr4_96184	CTGCATGCAATGCAAGCCTGAGCTCTCCCGCCATAAGGCTGCAGCGGTGTGGGCTCCTTGTGCCCAGATC

HIC_chr4_97125	GATCTCTGCGACCCCCGCGTGCCCGGGGAGAGCTCCCGGGGCGAGCTCCCGGGGCGTCCTTACCTGGGGC
HIC_chr4_97125	TAATCCCTTTCCATTAGTGAGGCCCTTTCTACTTTATGGAGCACTTTCACCCTTCTTACTTTATGTGATC
HIC_chr4_97617	GATCAGTTACCAGGAGAAGTTCTAAAGCAAGAAGAGAAAAGCATTTCAATTTGGGACATTTATTTGCACC
HIC_chr4_97617	TTTGCAGCCTACCTGTGCAATCATAGGAGATGGCCTAAAGGAAGGACTTGAGAAACTACATGATATGATC
HIC_chr4_98797	GATCCCCGGAACCCGGCCTGGCCACCCCGCTCTCTCTCGGGGGTCCGGGGTAGGTGAGCGGAGCCTGCCC
HIC_chr4_98797	ACGGCCACCAAGCGTTGCGCAGCTGCAGGAGGAAATTCCCTTAATTATGAATTTACAGAGGGGACTGATC
HIC_chr4_98914	GATCTGTTTTCTCAAGTCTCCAATCGCCTGCCTTCTTTGTGTCTTGTATTACCCTCACATCCCCCAGCTT
HIC_chr4_98914	TGTTTTTCGTCTTCCCTAGGCTATTTCTGCCGGGCGCTCCGCGAAGATGCAGCTCAAGCCGATGGAGATC
HIC_chr4_99200	GATCGTCACTTGGGCTGTAGTGCAAAACAGAAGGCATGCTCAAGTGGGAGTGGCCAAGGAGAGTTTAAAG
HIC_chr4_99200	AGAAGGCATGCTCAAGTGGGAGTGGCCAAGGAGAGTTTAAAGGAAGGGTTAGAAAGAAGTGAGCCGGATC
HIC_chr4_99266	ATCTCATATCCTGGTTCCTCTGACACCAGCTGCCTCTCCCATACCACCTAAGTTTGACCCCAGTGAGATC
HIC_chr4_100160	GATCCCGGCGTGAGGGAAGGGCAGCCGGACGTGGCCCCAAAAGTGGTCCTTATCGGGTTATACTGGAAGC
HIC_chr4_100160	TCCCTATCATTGATTCCTGCATCTCTAATTAGAATTTAATACCACACCATTACGCACCGAGCCCCTGATC
HIC_chr4_100475	GATCCGTTTTCAAACGGCGCGGGGACGGCAGTGCCGGAGGCCGCGTCTCCTTAGTAATCGCGCGGGCAGG
HIC_chr4_100475	CGGATAGGCCGGGCTGCCGCCAAACAAAGAGATAATAAAAAATTAACTATTTTAACATATATTACAGATC
HIC_chr4_100617	GATCATTTGTGAGCTGTATTTAATGCAAAAGTTGCTCCCCCATCCTGATTTCTTAGCTCACTGGGCCAAT
HIC_chr4_100617	GTTGGCTCTTTAGGGCTTCACCCCGAAGCTCCACCTTCGCTCCCGTCTTTCTGGAAACACCGCTTTGATC
HIC_chr4_100720	TCCCGTTTAGCCAACGAGCTGCGTGTGAGCTGCATGGAGCGGAAAAAGGTCCAAATTCGGAGCTTGGATC
HIC_chr4_100747	GATCAGGCTTGCCTAAAACGAGTTGAAACCAAAGCCATTTTAAGAATCCAAAATATGAGATTAGTTTTGT
HIC_chr4_100747	TTTTCTTACTCTCCCAACTTATTTTCTTAAATTTTCTAAAAGGAAGGAGGGGTGCTACTCACTACGGATC
HIC_chr4_100849	GATCCGCAGTGGCAGTGGTGTGTCCTGTCTGCGGAGAGCCAGGCCAGAGACAATGAGCAACACCTCAGAG
HIC_chr4_100849	GAAGAAATTCTTCTGGAACTCAGAGAAGAAGGAGTTTTTAGGCAGGACTGGTGGCAGTTGGTTTAAGATC
HIC_chr4_101192	GATCTTGGCCCGGGTGGTCGCGCGGTGTTTACGGGGCTTTGGGGTCCTGCTTTCCCCGAGCATCGCGGCC
HIC_chr4_101192	ACTGCAACTGGCGCCATCCGGCGAGGGTCTGGAGGGTGCCCAATTTCAGTAGCCGTTTGAATGAGAGATC
HIC_chr4_101904	GATCGCCACAGGTTGGGACTCTGTGCGGCGCTGGAGTTGGCAGCCCCGGGCGCTATGGCTCCCTTGAGGA
HIC_chr4_101904	GGTCTGTGCGCCAGGGGGACGCGGCCGGGTGGGGGAATCAAGGGATGAGTGGGTGGTCGCGTTTCAGATC
HIC_chr4_102061	GATCTTACTTGAATCCTTCTGAACTAAAATTTGGCAGGTTCATTTATAAAATTTCACATTGAGAAAAAAT
HIC_chr4_102061	AAAAGCCATACAAAAAATAAGCAAAAACATCCTAGGAGCTGTATATGACACAATTCTTGAGGACTTGATC
HIC_chr4_102492	GATCTCCGACACGGTCCTCCGCATGGTGGGCATCGCGGCGGCGGCTGCAGGTGGGTCCTCAGCCCGGACT
HIC_chr4_102492	GAGCCAGGGGGAGCCTCCTTCGCCCCGCCCCGCCCCGCTTCGGACCGTATCACGACTCAAAGGGGGGATC

HIC_chr4_105244	GATCTTAAATGTTATATAATAGAAATATTATATATTTCTAAGGGCCTCAGAATCGTGCAGGCGCAATTGT
HIC_chr4_105244	CGGGTGCAAGCCGAGCGGTTGGCCATAAGAGCCCGGCTGAAACGAGAGTACCTGCTTCAGTACAATGATC
HIC_chr4_105495	GATCTGAAAAAATTCCTGTTGCTTAGTGATGTCTTAATGACCCTGTGTAGGCCCAGGCTAACAAGTGTGT
HIC_chr4_105495	ACTCAAAACCTGCAAATTGCTCAGAATTCAAGAAGTCAGATATTTCTGGAAATAGGAGAGATGGTGGATC
HIC_chr4_106946	GATCCTAAGAAATCCTCCAAATGAAGAAACTTTCAAAGGCTCTGGAAACAATACGAATGATTCTGTCTTC
HIC_chr4_106946	TGATGTCACGACAGCGTGCGGCGTGCAGACGTCGGCAAGCTGCGCCGCCGCTTCGGGTTGCTTCCGGATC
HIC_chr4_106949	GATCCTACGGGGGGTACCTTCGAAAAAAAACGGGCTATGCTGCTGTTGCGTGTGGGTACCCTCTCCTGAC
HIC_chr4_106949	AATTCAAGGTGACTGCCCCCTGGAATCTGATTTAGCCAAGTTTTCAAACGTTGGAGTGCCCCATGCGATC
HIC_chr4_107050	GATCCAAACTATTGTCGCCTTATGCAATTCTCTGGCATCACAAGACAGCATCAAAAAACAACAGAGCACC
HIC_chr4_107050	GCCTGAGGTGAGACGCGAGGAGTTCGGTCCGAGTGCGGTGCGGCTGGGCGCTCTTCTTGCCTGGAAGATC

Table 2. Oligoes used for nuCaptureC.