

Cell Genomics, Volume 5

Supplemental information

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

Dubravka Vučićević, Che-Wei Hsu, Lorena Sofia Lopez Zepeda, Martin Burkert, Antje Hirsekorn, Ilija Bilić, Nicolai Kastelić, Markus Landthaler, Scott Allen Lacadie, and Uwe Ohler

Figure S1

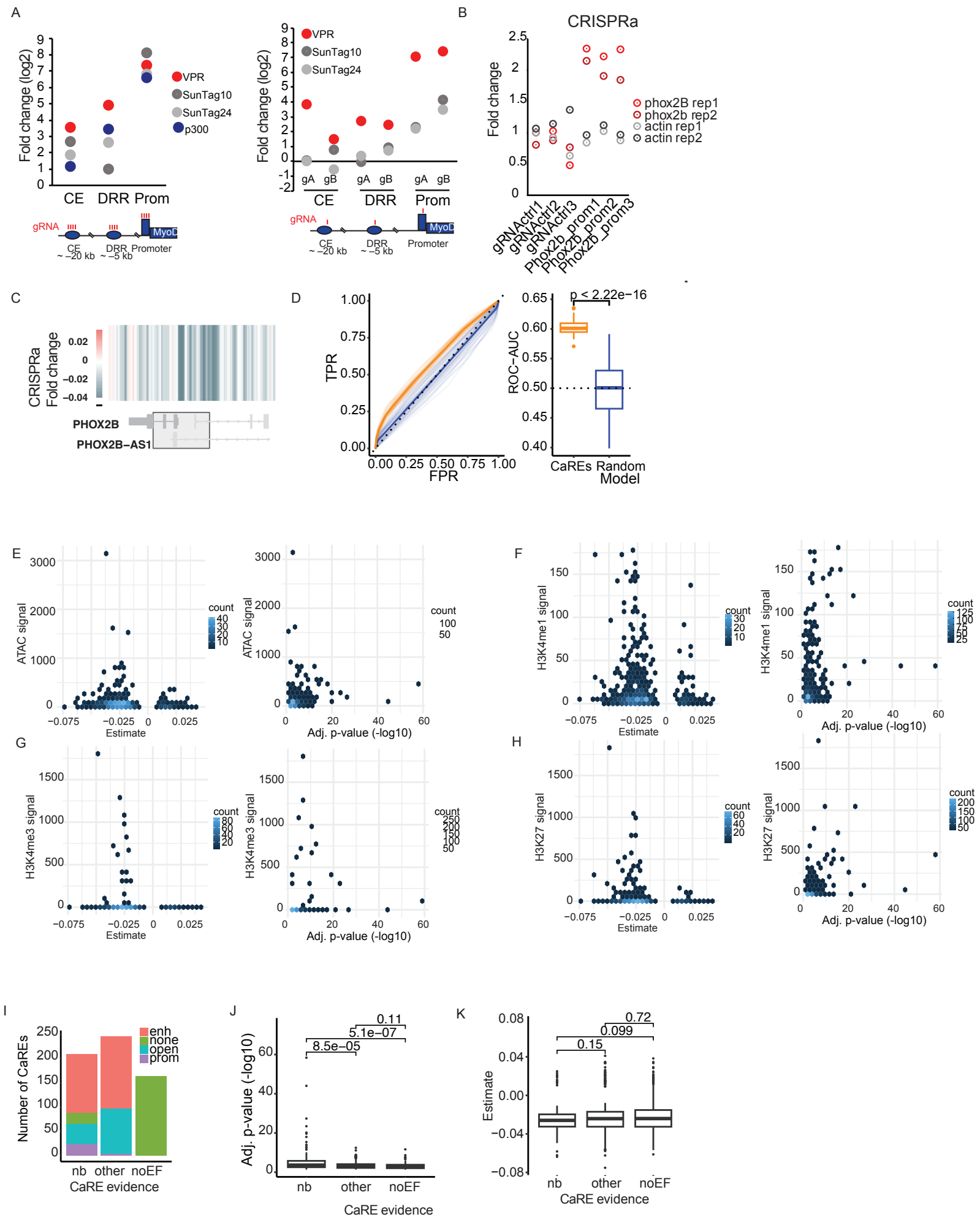
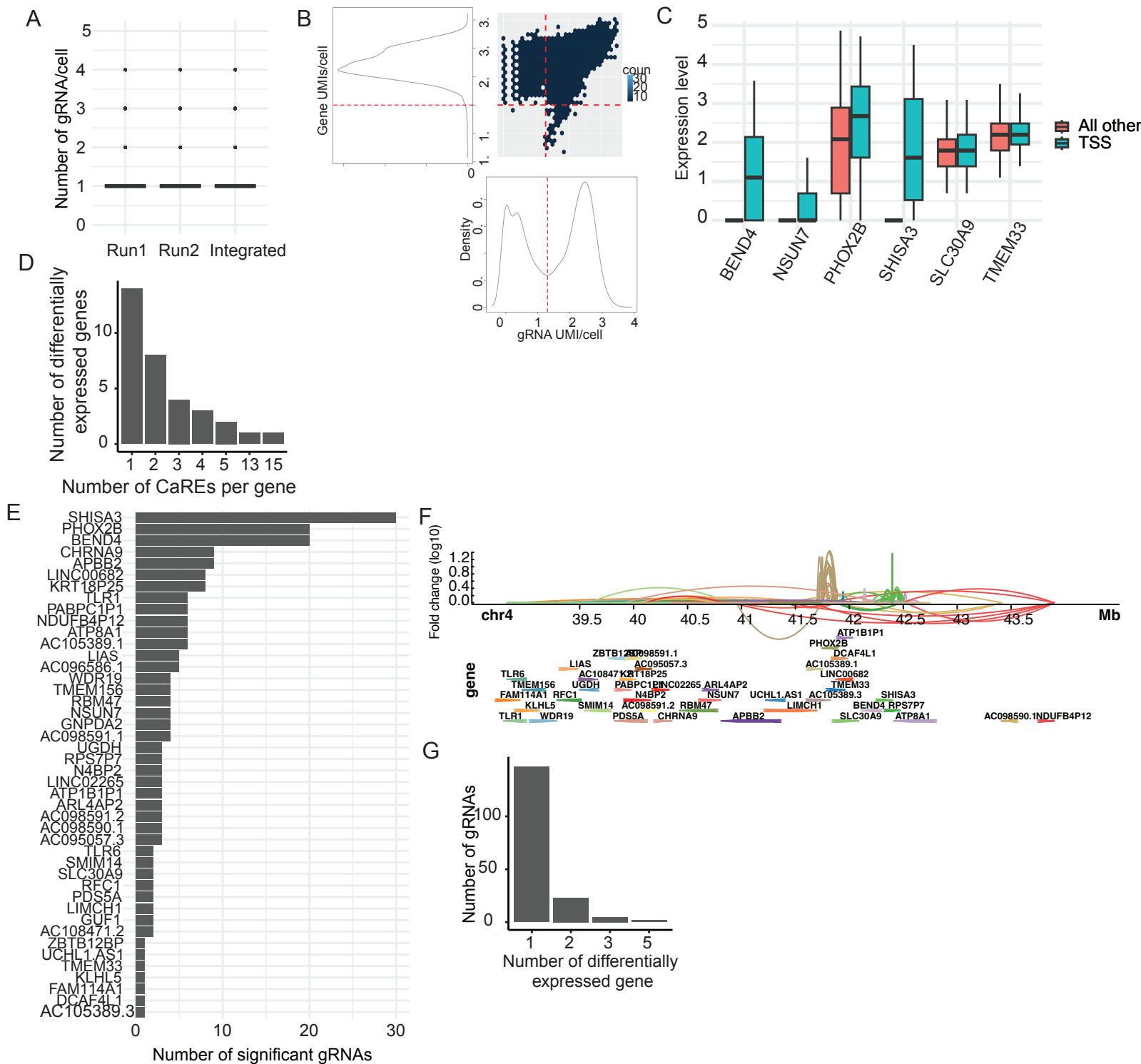


Figure S2



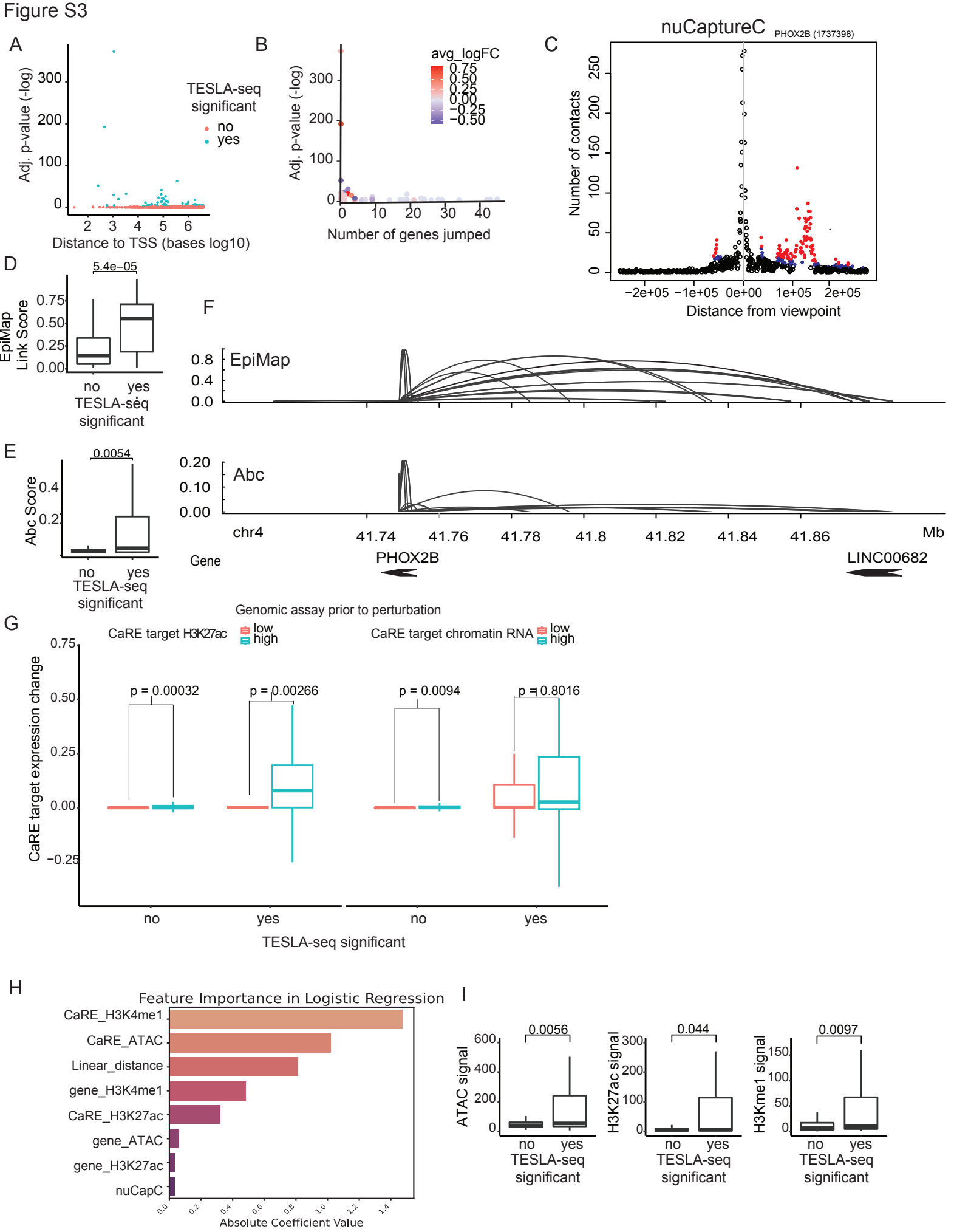


Figure S4

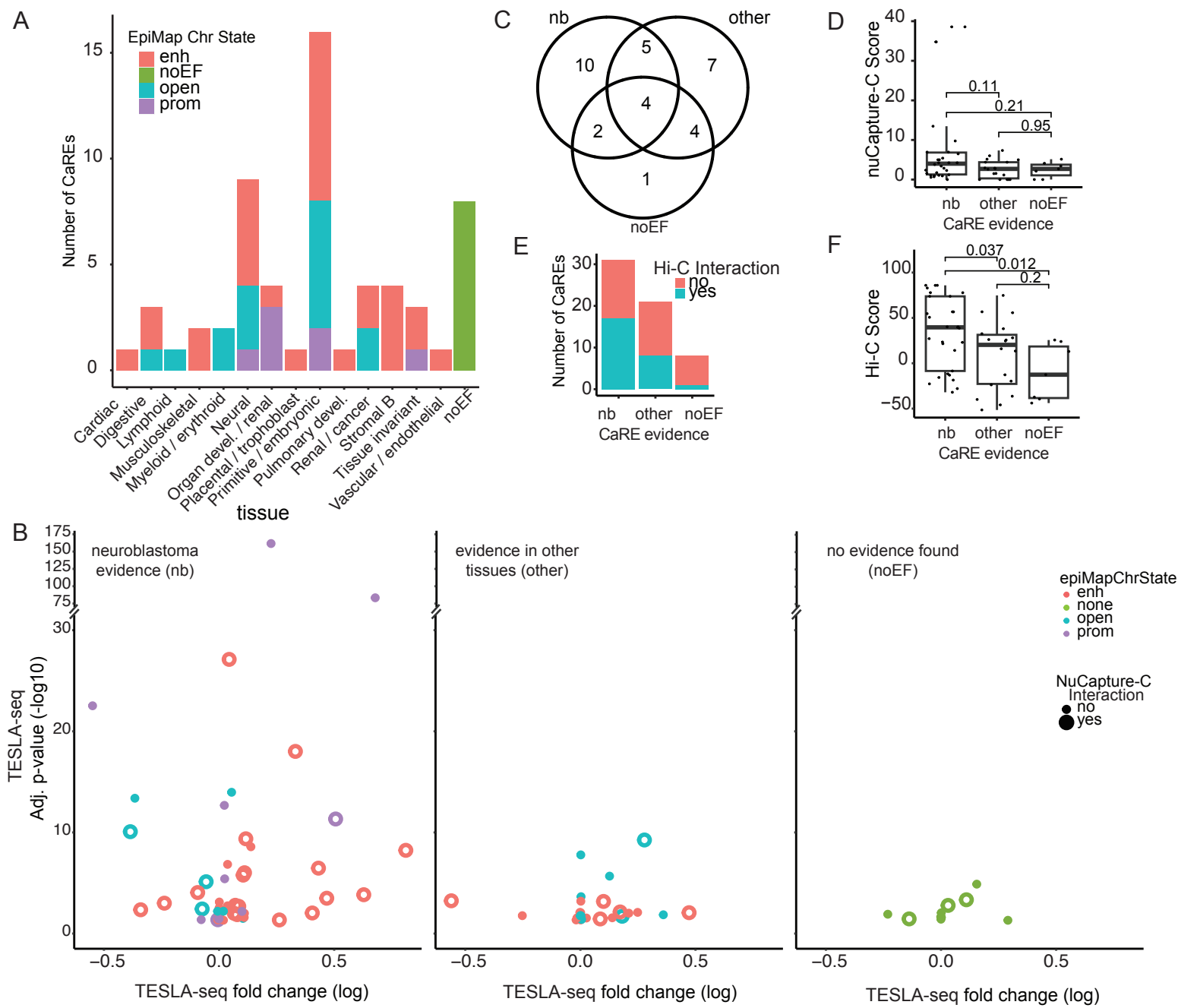


Figure S5

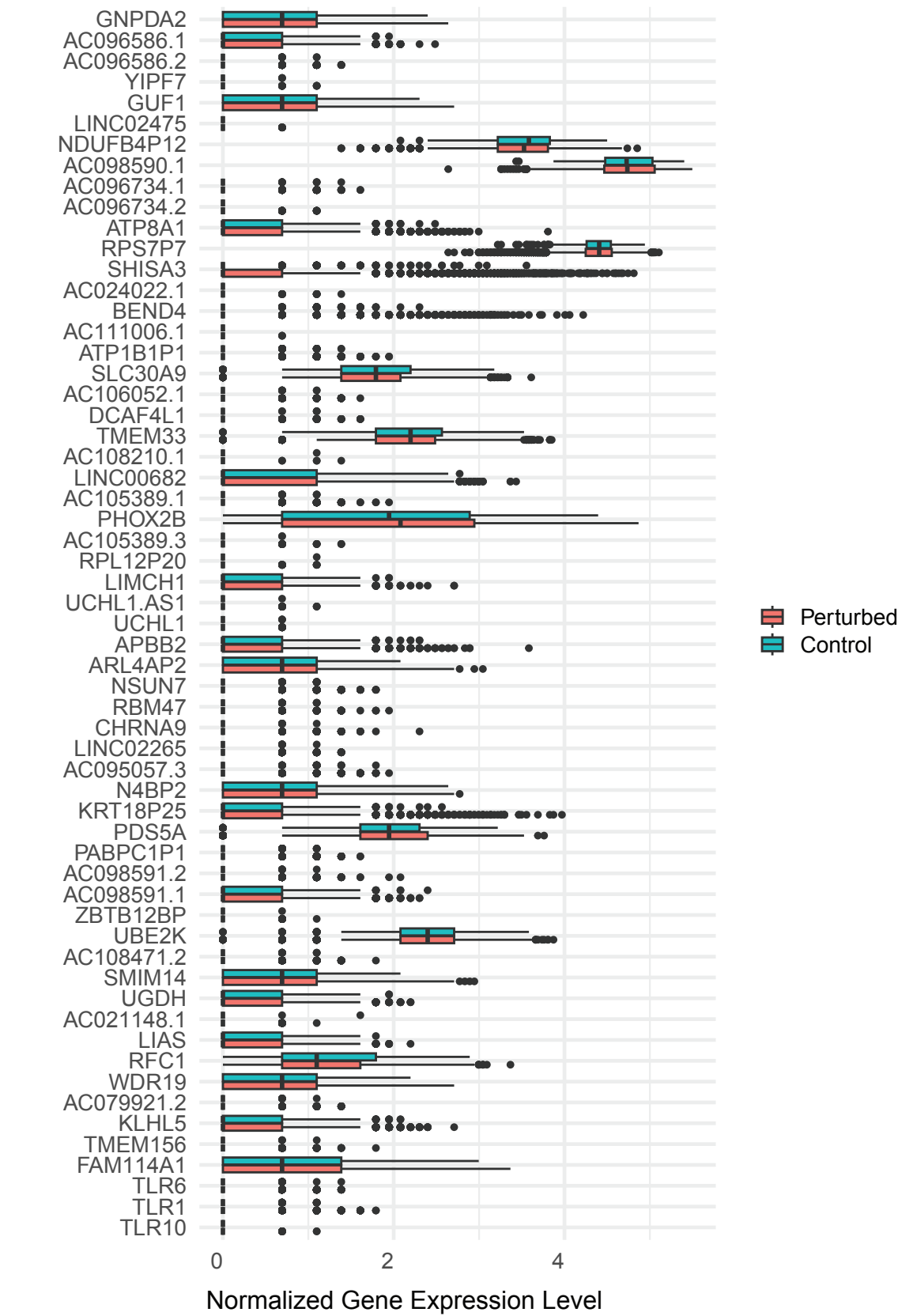


Figure S6

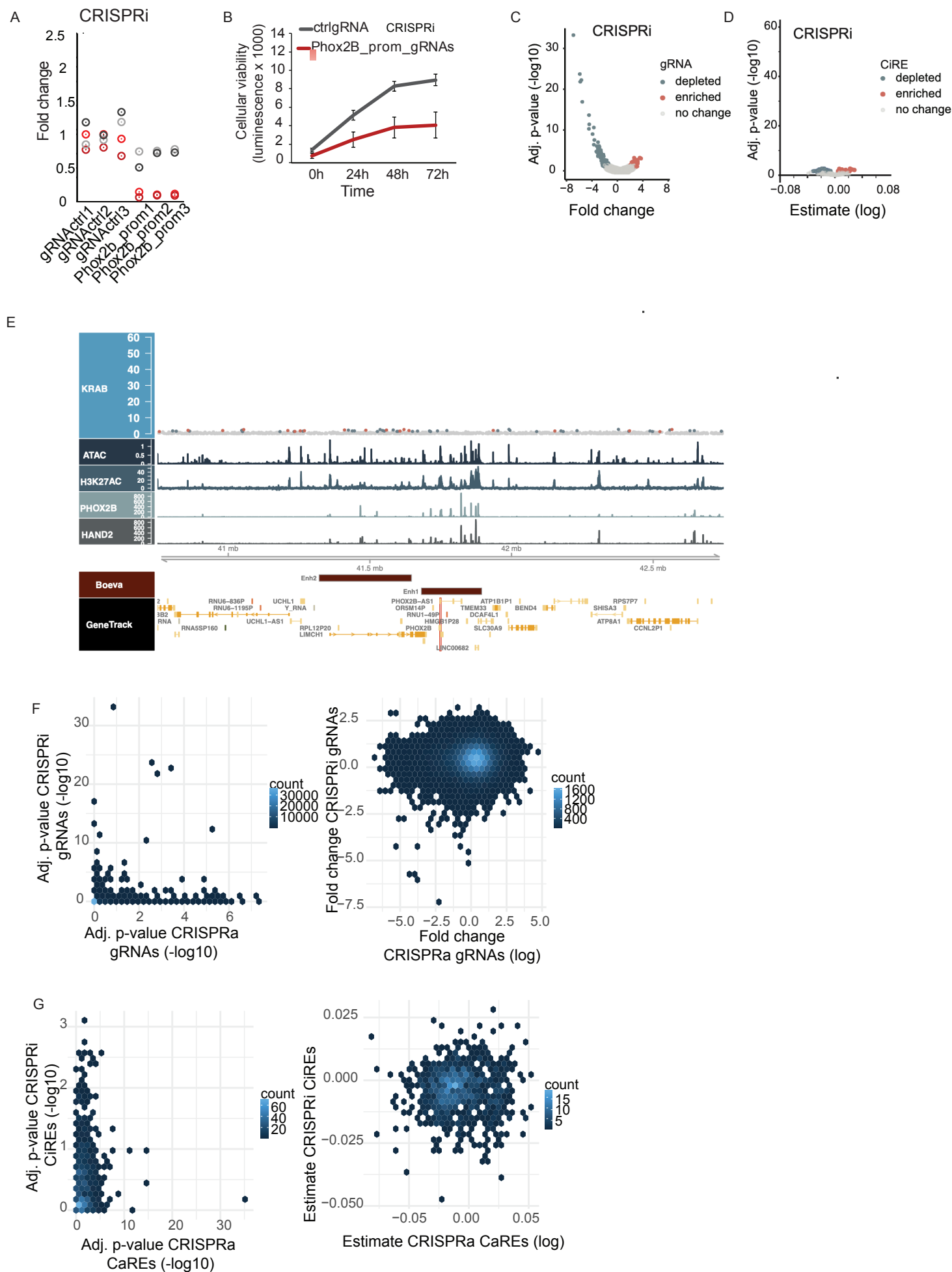
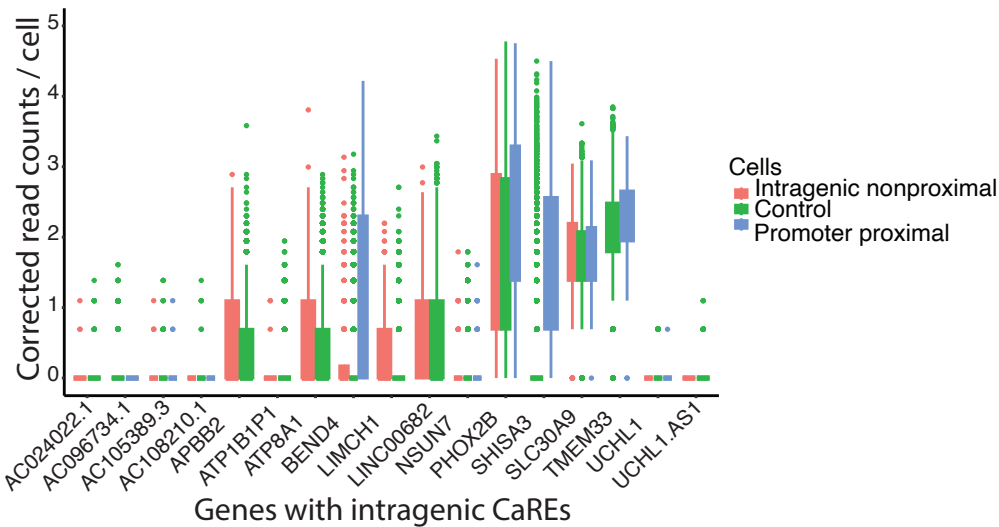


Figure S7

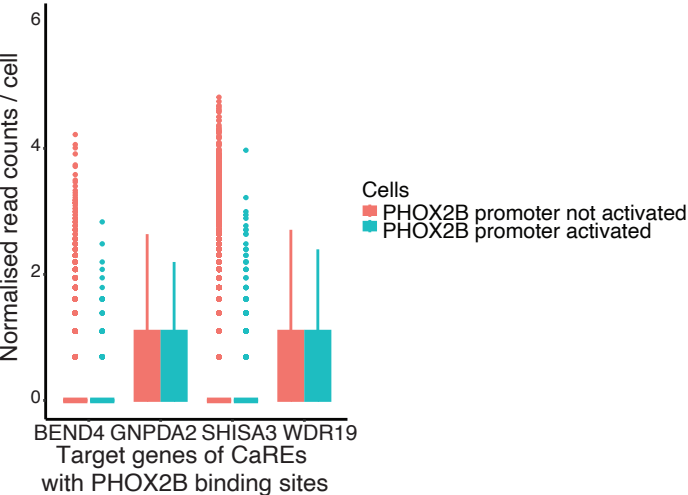
A

CaRE	Closest CaRE	distance (bp)	CaRE_target	closestCaRE_target
CaRE_1	CaRE_3	3671	RBM47,NSUN7,RPS7P7	PDS5A,TLR1
CaRE_171	CaRE_174	9631	NDUFB4P12	APBB2
CaRE_174	CaRE_179	9129	APBB2	TLR1
CaRE_553	CaRE_557	4752	PHOX2B,LINC00682	PHOX2B
CaRE_561	CaRE_562	1557	PHOX2B	PHOX2B
CaRE_668	CaRE_669	1133	CHRNA9	BEND4
CaRE_671	CaRE_669	2254	BEND4,AC098590.1	BEND4
CaRE_729	CaRE_730	1020	SHISA3,ZBTB12BP	BEND4,SHISA3
CaRE_740	CaRE_741	826	SHISA3,BEND4	SHISA3
CaRE_789	CaRE_790	654	SHISA3	SHISA3,TLR6,UGDH
CaRE_803	CaRE_806	5134	SHISA3	SHISA3,TMEM156,PABPC1P1
CaRE_828	CaRE_829	1266	SHISA3,TLR6,LINC02265	SHISA3

B



C



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 activation with dCas9-VPR 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)** CaREs signal at the PHOX2B promoter. Significantly enriched/depleted CaREs (FDR<0.05) are shown in red and blue respectively. **D)** Evaluation of epigenetic model for CaREs prediction. Left panel: Receiver operator characteristic (ROC) of trained CaRE models (light orange, n=50) and random models (light blue, n=50) from 10 repeats of a 5-fold cross validation each. Mean ROC curve for CaRE and random models in dark orange and blue respectively. TPR: True positive rate, FPR: False positive rate. Right panel: Distributions of ROC area under the curve (ROC-AUC) for CaRE and random models. Boxplot midlines mark median; upper and lower hinges extend to first and third quartile; upper and lower whiskers extend to the smallest and largest value max. $1.5 \times \text{IQR}$. Comparison p-value from two-sided Wilcoxon ranksum test. **E-H)** The association of CaREs to open chromatin - ATAC (E), H3K4me1 signal (F), H3K4me3 signal (G) and H3K27ac signal (H) relative to the CaRE fold change (estimate, left) or adjusted p-value (right). **I)** Number of CaREs identified in the phenotypic screen in each category, nb, other, noEF grouped by the EpiMap classification indicated with colours. **J-K)** Comparison between different CaRE evidence categories (nb, other, noEF) in CaRE-gene pair adj. P-value (J) and fold change (estimate, K). All indicated p-values were calculated using the two-sided Wilcoxon Rank Sum test.

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 indicated gene promoter versus all other cells. **D)** Histogram showing the number of CaRE that induce differential expression of an indicated number of genes. **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 ($-\log_{10}$) of each CaRE-gene regulatory pair and genomic distance between the CaRE and the gene's transcriptional start site (in bases at \log_{10} scale). **B)** Relationship between adjusted p-values ($-\log_{10}$) 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 ($p\text{Level} = 5$), blue: significant interactions, $p\text{Level} = 3$, 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). **H)** CaRE features that contribute the most to the predictability of a logistic regression model. **I)** Comparison of ATAC signal (left), H3K27ac signal (middle) and H3K4me1 signal (right) between significant and non-significant CaREs. All indicated p-values were calculated using the two-sided Wilcoxon Rank Sum test.

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 ($-\log_{10}$ scale) and fold change (\log_2 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. All indicated p-values were calculated using the two-sided Wilcoxon Rank Sum test.

Figure S5. Normalized expression level of all TESLA-seq captured genes related to Figure 4. Histogram showing normalized gene expression level of all TESLA-seq captured genes in perturbed and control cells.

Figure S6. CRISPRi phenotypic screen related to Supplementary Note 1.

A) PHOX2B expression changes upon repression with dCas9-KRAB 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. **B)** MTT viability assay with control or gRNAs targeting the promoter of Phox2B in a SH-SY5Y-KRAB (CRISPRi) cell line. **C)** 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. **D)** 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. **E)** CiREs 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 CiREs. *P*-values shown are adjusted for multiple testing (FDR). **F-G)** Comparison between CRISPRa and CRISPRi screen at the gRNA (F) and CaRE/CiRE level (G) in adjusted p-value (left) or fold change (estimate, right).

Figure S7. Characteristics of CRISPR activation related to Supplementary Note 2. **A)** Table showing the closest significant CaRE neighbors and their target genes identified by TESLA-seq. **B)** Genes expression comparison in cells in which the following CaREs are targeted: CaREs targeting the promoter of the gene depicted (promoter proximal), CaREs within genes that are not at the promoters of genes (intragenic nonproximal) and CaREs that are targeting genes other than the one depicted (control). **C)** Genes expression comparison of the genes targeted by CaREs that have PHOX2B binding sites in cells in which the PHOX2B promoter was activated (blue) and cells in which PHOX2B was not activated (red).

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	aaactgcccttaattcaatcacaC
phox2b_prom1FWDgRNA	CACCGtggtgattgaattaaagggca
phox2b_prom2REVgRNA	aaacccttctaaccagctccctgC
phox2b_prom2FWDgRNA	CACCGcaggagctggttagaaggg
phox2b_prom3REVgRNA	aaacctgatcctcccttctaaccaC
phox2b_prom3FWDgRNA	CACCGtggttagaagggaggatcag
phox2b_prom4REVgRNA	aaacCCCTATCATTGATTCTCTGCAC
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 S1. Primers and individual gRNAs oligoes, related to Figure 1.

name	sequence
HIC_chr4_91471	GATCAATAGCCCCAAGGTCATAATCATTAACTAGACAAAATAGTAAATACTAAAAATCTAAATAACCC
HIC_chr4_91683	GATCTCTTTTTCAGAAGAAGAGAAAATAACTTGTAATGAACGGAAATACAATTGAAAATGCTTGACAAA
HIC_chr4_91683	TACTGGTTGATACATATTTATGACTACATATTTAGCCATATGTACGAGCAATCATACCCACATTCTGATC
HIC_chr4_91712	GATCTGGCGAACCCCGAGACCCACCCGCCCTGGGCTGGAGAGGCGGCGCGCTCTGGCTTTCCGCGCTGG C
HIC_chr4_91712	ACACCCACACGCCCACACTCAGGGTCTGCCCCCTCGGCCTGCGTGAACCTCCGCGGAGCCTGCCTGGATC
HIC_chr4_92179	GGTGAATGTCAGTCATGTCTAGGTTGCATGCACATGACTCATTCACTTACTTATTAATGATTGTGATC
HIC_chr4_92220	GATCATAATTTTCGATGCCACAAACCCACCACCTAGAGCACACTAATTATCAGATTATTGCAAGGGGAATG
HIC_chr4_92220	TTCCCGAGACCCGCGCTGCCTGGTGAACTTTTGCACTTAGGCATTTATTTCACTGCATGCTCTGGGATC
HIC_chr4_92630	GATCCAAACTCTAAAGGAAAATATTGGAGCTCATTTTTCTGCCACGTTAGCTAGCACATTTTGTGA
HIC_chr4_92630	GACTGCAGACGCCGTAGAAGCGGTGCAGAAAGTGGAACCCCTCCCTGGCCGAAATGAGCGGACTGGAT C
HIC_chr4_93108	GATCCATTGCGCCAACAACCTTCTCCGCGAAGTGCAAGAAGGCGAAGACAGTGCGCGCGGTGATGACG
HIC_chr4_93108	GACATCGAGAGGAACCACGGTGCTTTGGTGTGAGCTCGGTTTTTGGCGGGGGCGCTAAAGTAGGGGGATC
HIC_chr4_93406	GATCATGTTCTGTACATGACTACAAATAGTCCGAACGGTAGCCAGTTCCTTTCTGTACCCACCATTTG
HIC_chr4_93406	AGCAGCTTGGCTGCTTGTATATAATGGAGCGACGTAATTCGACCTGTCTTTCCCGGGAGTTAGCGATC
HIC_chr4_93611	GATCTGCGGGTCGGTCTCGGCGGGCGGATTACTCCTCGCCACAACCTCGAGGCCCGGGGTCAACCGC G
HIC_chr4_93611	AAGAAAGTGATTAGTAATTTATAAAGGCACATTAATAATACCATATAAATTACGGTTTTCTTGATAGGATC
HIC_chr4_93953	GATCTGCCGGCTCCCGCGGGCGGCGGCGCAACAGATTGCAGCGCCTGGAGACTCCAGCTGCCCCGCCTG C
HIC_chr4_94155	GATCCGGGAGAGCGTTTTCTGCGCTAGACACGGCGTTTACGCTCCGGGTCCGGGTCTAGCTGAGTCAG
HIC_chr4_94155	GAGGGAGCAGGCCGAGCCTGAGAAAACCCGGAAGTGGGTTGGGGGAAGGGGAAAGGTGGTAAGTGGAT C
HIC_chr4_94380	GTAGCTCTGGGATAGAGAAAACCTCCCAAGGGATGCTGATGCTGCTGGTTGGATGAACAACACTGAGATC
HIC_chr4_95059	GATCAATTTTAAACAGCAGGAACACCAATGGCACTGTTAACTGCTTTCTGGGTAGCCTCTTTAGCTTGGT
HIC_chr4_95073	GATCGAACCCCGCCGGCCCCGCCAACCCGGCCCTGCCAGCCCCAGGCTCACGGGGCTCCTCCGACCAG C
HIC_chr4_95073	AGACGAAACTGGTTGGGAGGTAGTTAATTGCTCGGTGAAAATGAACTGATTTCTCCTCAGGGAGAGATC
HIC_chr4_95258	CTACGGCGCTCGGCCAGTCAGCAGCTCTGCCAGCATCTATGCAGGTGCCGGGGGCTCTGGTTCTGGATC
HIC_chr4_95387	GATCCACCTGGGCGACGTGGCAGCCATTGCCGGCGCCTCGAGACCCCTCCCCACCCGGCCGCCACCCG C
HIC_chr4_95387	AGCGGCGGGCGGCGCGGCGAATGAACCCCCAAGCCCTGAATGTGGGGCCCCGGCGGGCGCCCTCCGA TC
HIC_chr4_96122	GATCATTGTGACATTATAATGGTGAAAAATTAGAAATAACCTAAACATCTAGCATTTAGAGAATGATTAA
HIC_chr4_96122	GCTTTGCCACCCAGTTTTGGTCATGGGACTTGAGTTACTGTGACCTCAGAGGCCCTAGGGGCCTGATC
HIC_chr4_96184	GATCTCGGTCTGCATGCAATGCAAGCCTGAGCTCTCCCGCCATAAGGCTGCAGCGGTGTGGGCTCCTTG
HIC_chr4_96184	CTGCATGCAATGCAAGCCTGAGCTCTCCCGCCATAAGGCTGCAGCGGTGTGGGCTCCTTGTGCCAGATC

HIC_chr4_97125	GATCTCTGCGACCCCCGCGTGCCCGGGGAGAGCTCCCGGGGCGAGCTCCCGGGGCGTCCTTACCTGGG GC
HIC_chr4_97125	TAATCCCTTTCCATTAGTGAGGCCCTTTCTACTTTATGGAGCACTTTACCCTTCTTACTTTATGTGATC
HIC_chr4_97617	GATCAGTTACCAGGAGAAGTTCTAAAGCAAGAAGAGAAAAGCATTTCAATTTGGGACATTTATTTGCACC
HIC_chr4_97617	TTTGCAGCCTACCTGTGCAATCATAGGAGATGGCCTAAAGGAAGGACTTGAGAACTACATGATATGATC
HIC_chr4_98797	GATCCCCGGAACCCGGCCTGGCCACCCCGCTCTCTCTCGGGGGTCCGGGGTAGGTGAGCGGAGCCTGCC C
HIC_chr4_98797	ACGGCCACCAAGCGTTGCGCAGCTGCAGGAGGAAATCCCTTAATTATGAATTTACAGAGGGGACTGATC
HIC_chr4_98914	GATCTGTTTTCTCAAGTCTCCAATCGCCTGCCTTCTTTGTGTCTTGTATTACCCTCACATCCCCAGCTT
HIC_chr4_98914	TGTTTTTCGTCTTCCCTAGGCTATTTCTGCCGGGCGCTCCGCGAAGATGCAGCTCAAGCCGATGGAGATC
HIC_chr4_99200	GATCGTCACTTGGGCTGTAGTGCAAAACAGAAGGCATGCTCAAGTGGGAGTGGCCAAGGAGAGTTTAAAG
HIC_chr4_99200	AGAAGGCATGCTCAAGTGGGAGTGGCCAAGGAGAGTTTAAAGGAAGGGTTAGAAAGAAGTGAGCCGGATC
HIC_chr4_99266	ATCTCATATCCTGGTTCCTCTGACACCAGCTGCCTCTCCCATACCACCTAAGTTTGACCCAGTGAGATC
HIC_chr4_10016 0	GATCCCGGCGTGAGGGAAGGGCAGCCGGACGTGGCCCCAAAAGTGGTCCTTATCGGGTTATACTGGAAG C
HIC_chr4_10016 0	TCCCTATCATTGATTCCTGCATCTCTAATTAGAATTTAATACCACACCATTACGCACCGAGCCCCTGATC
HIC_chr4_10047 5	GATCCGTTTTCAAACGGCGCGGGGACGGCAGTGCCGGAGGCCGCGTCTCCTTAGTAATCGCGCGGGCAG G
HIC_chr4_10047 5	CGGATAGGCCGGGCTGCCGCCAAACAAAGAGATAATAAAAAATTAAGTATTTTAACATATATTACAGATC
HIC_chr4_10061 7	GATCATTTGTGAGCTGTATTTAATGCAAAAGTTGCTCCCCATCCTGATTTCTTAGCTCACTGGGCCAAT
HIC_chr4_10061 7	GTTGGCTCTTTAGGGCTTCACCCCGAAGCTCCACCTTCGCTCCCGTCTTTCTGAAACACCGCTTTGATC
HIC_chr4_10072 0	TCCCGTTTAGCCAACGAGCTGCGTGTGAGCTGCATGGAGCGGAAAAAGGTCCAAATTCGGAGCTTGGATC
HIC_chr4_10074 7	GATCAGGCTTGCCTAAAACGAGTTGAAACCAAAGCCATTTTAAGAATCCAAATATGAGATTAGTTTTGT
HIC_chr4_10074 7	TTTTCTTACTCTCCCACTTATTTTCTTAAATTTTCTAAAAGGAAGGAGGGGTGCTACTCACTACGGATC
HIC_chr4_10084 9	GATCCGCAGTGGCAGTGGTGTGTCTGTCTGCGGAGAGCCAGGCCAGAGACAATGAGCAACACCTCAGAG
HIC_chr4_10084 9	GAAGAAATCTTCTGGAACCTCAGAGAAGAAGGAGTTTTTAGGCAGGACTGGTGGCAGTTGGTTTAAGATC
HIC_chr4_10119 2	GATCTTGGCCCGGGTGGTCGCGCGGTGTTTACGGGGCTTTGGGGTCTGCTTCCCCGAGCATCGCGGC C
HIC_chr4_10119 2	ACTGCAACTGGCGCCATCCGGCGAGGGTCTGGAGGGTGCCCAATTTAGTAGCCGTTTGAATGAGAGATC
HIC_chr4_10190 4	GATCGCCACAGTTGGGACTCTGTGCGGCGCTGGAGTTGGCAGCCCCGGGCGCTATGGCTCCCTTGAGG A
HIC_chr4_10190 4	GGTCTGTGCGCCAGGGGGACGCGGCCGGGTGGGGGAATCAAGGGATGAGTGGGTGGTCGCGTTTCAGAT C
HIC_chr4_10206 1	GATCTTACTTGAATCCTTCTGAACTAAAATTTGGCAGGTTCAATTTATAAAATTTACATTGAGAAAAAAT
HIC_chr4_10206 1	AAAAGCCATACAAAAAATAAGCAAAAACATCCTAGGAGCTGTATATGACACAATTCTTGAGGACTTGATC
HIC_chr4_10249 2	GATCTCCGACACGGTCTCCGCATGGTGGGCATCGCGGCGGGCTGCAGGTGGGTCTCAGCCCCGGAC T
HIC_chr4_10249 2	GAGCCAGGGGGAGCCTCCTTCGCCCCGCCCCGCCCCGCTTCGGACCGTATCACGACTCAAAGGGGGGAT C

HIC_chr4_10524 4	GATCTTAAATGTTATATAATAGAAATATTATATATTTCTAAGGGCCTCAGAATCGTGCAGGCGCAATTGT
HIC_chr4_10524 4	CGGGTGCAAGCCGAGCGGTTGGCCATAAGAGCCCGGCTGAAACGAGAGTACCTGCTTCAGTACAATGATC
HIC_chr4_10549 5	GATCTGAAAAAATTCCTGTTGCTTAGTGATGTCTTAATGACCCTGTGTAGGCCCAGGCTAACAAGTGTGT
HIC_chr4_10549 5	ACTCAAAACCTGCAAATTGCTCAGAATTCAAGAAGTCAGATATTTCTGGAAATAGGAGAGATGGTGGATC
HIC_chr4_10694 6	GATCCTAAGAAATCCTCCAAATGAAGAACTTTCAAAGGCTCTGGAAACAATACGAATGATTCTGTCTTC
HIC_chr4_10694 6	TGATGTCACGACAGCGTGCGGCGTGCAGACGTCGGCAAGCTGCGCCGCCGCTTCGGGTTGCTTCGGAT C
HIC_chr4_10694 9	GATCCTACGGGGGGTACCTTCGAAAAAAACGGGCTATGCTGCTGTTGCGTGTGGGTACCCTCTCCTGAC
HIC_chr4_10694 9	AATTCAAGGTGACTGCCCCCTGGAATCTGATTTAGCCAAGTTTTCAAACGTTGGAGTGCCCCATGCGATC
HIC_chr4_10705 0	GATCCAACTATTGTGCCTTATGCAATTCTCTGGCATCACAAGACAGCATCAAAAAACAACAGAGCACC
HIC_chr4_10705 0	GCCTGAGGTGAGACGCGAGGAGTTCGGTCCGAGTGCGGTGCGGCTGGGCGCTCTTCTTGCCTGGAAGAT C

Table S2. Oligoes used for nuCaptureC, related to Figure 3.