Analysis of hundreds of cis-regulatory landscapes at high resolution in a single, high-throughput experiment.

Analysis of hundreds of *cis*-regulatory landscapes at high resolution in a single, high-throughput experiment.

Jim R Hughes[1], Nigel Roberts[1], Simon McGowan[2], Deborah Hay[1], Eleni Giannoulatou[2], Magnus Lynch[1], Marco De Gobbi[1], Stephen Taylor[2], Richard Gibbons[1] and Douglas R. Higgs[1]

[1] MRC Molecular Haematology Unit, Weatherall Institute of Molecular Medicine, Oxford University, Oxford, OX3 9DS, UK

[2] Computational Biology Research Group, Weatherall Institute of Molecular Medicine, Oxford University, Oxford, OX3 9DS, UK

Correspondence

Douglas R. Higgs doug.higgs@imm.ox.ac.uk

Jim R. Hughes jim.hughes@imm.ox.ac.uk

Abstract

Gene expression during development and differentiation is regulated in a cell- and stage-specific manner by complex networks of intergenic and intragenic *cis*-regulatory elements whose numbers and representation in the genome far exceed that of structural genes. Using chromosome conformation capture it is now possible to analyze, in detail, the interaction between enhancers, silencers, boundary elements, and promoters at individual loci, but these techniques are not readily scalable. Here we present a high throughput approach (Capture-C) to analyze *cis*-interactions, interrogating hundreds of specific interactions at high resolution in a single experiment. We show how this approach will facilitate detailed, genome-wide analysis to elucidate the general principles by which *cis*-acting sequences control gene expression. In addition, we show how Capture-C will expedite identification of the target genes and functional effects of SNPs associated with complex diseases, which most frequently lie in intergenic *cis*-acting regulatory elements.

Introduction

It is now possible to rapidly map the positions of many *cis*-regulatory sequences (promoters, enhancers, silencers and boundary elements) across the genome by analyzing chromatin structure, histone modifications, and the binding of transcription factors and co-factors[1-4]. Detailed studies of a relatively small number of individual genes have revealed unexpected levels of complexity in their interactions with *cis*-regulatory sequences. Expression of a single gene is often controlled by multiple regulatory elements which may lie 10s-100s kb upstream or downstream of their targets. In addition, the regulatory elements controlling one gene may lie within the introns of another, unrelated gene. Detailed studies of individual loci have revealed some mechanistic insights into how *cis*-acting elements regulate gene expression, but since only a few loci have been studied in detail, general principles have not yet emerged. Understanding the mechanisms underlying long-range regulation of gene expression is of critical importance in molecular medicine since it has recently been shown that most SNPs associated with complex diseases lie within distal *cis*-regulatory elements and presumably alter the timing or levels of expression of their target genes in specific cell types. Therefore, a major challenge in mammalian genetics is to develop high throughput techniques to link specific *cis*-regulatory elements to their cognate genes and to determine how these interactions and their associated variants influence gene expression during development and differentiation.

Where studied, it has been shown that when *cis*-acting sequences influence gene expression, they may physically interact with the promoter(s) they regulate. The resulting physical contacts can be detected by various protocols, generally referred to as chromosome conformation capture or 3C. These techniques involve digestion and re-ligation of fixed chromatin, followed by quantification of ligation junctions, which reflect the frequencies of interaction[5,6].

Several adaptations of the original 3C method (4C, 5C, HiC, ChIA-PET) have been developed to assay interactions across the genome, but are currently unsuitable for linking *cis*-acting sequences with gene promoters both at high resolution and in a high-throughput manner[7-10]. To map enhancer/promoter interactions in detail using chromosome conformation capture requires a resolution of ~1-2 kb since most *cis*-acting sequences are in the range of 100's of bp in length and may be closely clustered. Some versions of the 4C-seq method can map interactions at this resolution, but they only interrogate a single region of interest at a time[10]. However any given cell type contains thousands of active promoters and even greater numbers of potential *cis*-acting sequences[4]. Also, typically, the number of significant disease associated GWAS variants affecting these elements are

numbered in the thousands[11]. Hence, using 4C-seq to interrogate each, active element and associated variants individually is laborious and prohibitively expensive. Clearly this represents a major "bottleneck" in our ability to investigate both the normal regulation of genes and the effects of sequence variants.

Here we present a new approach to this problem. "Capture-C" combines oligonucleotide capture technology (OCT), chromosome conformation capture (3C) and high throughput sequencing (HTS) and enables researchers to interrogate *cis*-interactions at hundreds of selected loci at high resolution in a single assay. When combined with corresponding epigenetic data used to localize the major classes of *cis*-elements, this technique identifies all such elements interacting with chosen promoters or enhancers of interest. Here, we illustrate how this approach can be used to greatly increase the number of *cis*-interactions that can be simultaneously analyzed in various cell types. Furthermore, we show how Capture-C can be used to link SNPs in distal regulatory elements to the changes in expression of their cognate genes; the proposed mechanism underlying QTLs.

A multiplexed, high-resolution method of chromatin conformation analysis

There are four core design considerations for an assay suitable to address this problem. First, the resolution of a 3C assay depends both on the frequency with which the restriction enzyme used cuts within the genome, and also the size of the bioinformatic window used to analyze the resulting distribution of ligation junctions. For a given frequency of digestion, the size of the appropriate window to be used depends on the signal generated: the window size can be decreased as the density of the signal increases. Hence the method should be capable of producing sufficient enrichment of ligation junctions from a 3C library generated using a frequently cutting restriction enzyme. Second, the method should simultaneously enrich for ligation junctions at hundreds of chosen regions of interest within the genome and retain high levels of signal. Third, the chosen regions of interest should be enriched independently of chromatin structure, protein binding profiles or function, as these will not be constant across all regions of interest and in all cellular conditions. Finally, quantification of the ligation junctions should be as direct as possible, to avoid any bias introduced by extensive amplification or intermediate steps (see Supplementary note).

To address these design considerations we used oligo capture technology (OCT SureSelect platform)[12] combined with conventional 3C technology using a frequently cutting restriction enzyme (Dpn II, cutting at GATC, Supplementary figure 1a). The principle underlying this method is that after preparing a high-resolution 3C library[13], DNA is sonicated (300 bp) and sequencing adapters are

attached. The resultant library contains all Dpn II "junction fragments" which reflect the frequency of genomic interactions *in vivo* in the original nuclei[6,14]. To identify all interactions that occur with a particular Dpn II fragment (containing a specific element of interest) oligonucleotides located adjacent at each 5' and 3' flanking Dpn II restriction site are used to capture the sequences to which it is ligated in the 3C library (Supplementary figure 1b).

Using OCT, each targeted region (Supplementary figure 1a) included in the design, captured in the order of 10^3-10^5 reads. As a measure of background enrichment, we designed ~6,000 "control" targeted regions located around the promoters of all other active promoters in erythroid cells. To do this we used the same criteria as those for the targeted regions around promoters included on the SureSelect array. Relatively few reads from the library mapped non-specifically to these control regions although they are similar to the test regions in terms of sequence, chromatin accessibility and function. This means that at any one capture region, enrichment over this background ranges from 10^2 to 10^3 fold (Supplementary figure 1c(i)).

The targeted regions capture several different types of molecule. Some are simply interacting with the targeted region itself; others are interacting with local sequences (defined here as within 2kb). Some junction fragments will span two capture points and these are no longer quantitative. All such fragments are excluded from further analysis. This leaves junction fragments with one end in the target region and another in a unique distal fragment and these are the presumed *cis*-interacting sequences. However, we observed that some of these have precisely the same structure and could result from amplification of a single molecule. Therefore, to avoid over-estimating the effects, these reads were considered as a single interacting fragment. Using these criteria for analysis we still observed 100s-1000s of unique interactions with each targeted region (Supplementary figure 1c(ii)).

The ability to enrich for specific *cis*-interactions at these levels means that 3C libraries can be generated using 4 cutter restriction enzymes and analyzed using a small window size (2 kb) allowing for high resolution mapping of interactions. OCT oligos can be combined as pools to multiplex the analysis of hundreds of regions in a single assay and here we use a custom SureSelect platform (Agilent) to analyze the interactions of 455 pre-selected promoters (Supplementary table 1). Oligo capture is dependent only on suitable sequence content (Supplementary figure 2), therefore the regions are selected without needing to consider their function, chromatin structure or binding profiles. The captured junction fragments are small and of uniform size (150-300 bp) which minimizes bias in the relatively few rounds of amplification required in this protocol. Mapping of the paired end reads and analysis of the signal distribution is based on minor adaptations of our current bioinformatics methods used to generate

DNase-seq and ChIP-seq profiles and avoids complex mathematical modeling. In comparison to other "all vs all" 3C methods (Hi-C and Chia-PET) which typically generate very low numbers (<1-10s) of informative interactions from each restriction fragment, Capture-C typically identifies 100s-1000s of informative interactions from each targeted fragment (supplementary table 2). The high resolution interaction maps produced by the Capture-C method in combination with standard genomics data enabled us to accurately identify the *cis*-regulatory elements interacting with each of the chosen promoters in a single experiment (see online methods).

Validation of the Capture-C technique

To validate this multiplexed method against established techniques we analyzed the *cis*-interactions at the mouse α and β globin gene loci in erythroid (mTer119+) and non-erythroid cells (mES Cells). Interactions at the globin loci have been used to validate all chromosomal conformation techniques in the past and thus provide the ideal standard by which to assess newly developed techniques.

The mouse α globin cluster lies at an interstitial region on chromosome 11 (Figure 1). The cluster includes an embryonic gene (*Hba-x*), two fetal/adult genes (*Hba-a1* and *Hba-a2*) and two globin-like genes of unknown function (*Hba-q1* and *Hba-q2*) arranged in the order 5' *Hba-x-Hba-a1-Hba-q1-Hba-a2-Hba-q2* 3'. Expression of the α locus depends on a set of four multi-species conserved distal regulatory elements (MCS-R1-4) which correspond to erythroid-specific DNase1 hypersensitive sites (DHSs)[15], and a murine-specific element called HS-12[16]. Four of these regulatory elements (MCS-R1,2,3 and HS-12) lie within the introns of an adjacent widely expressed gene *Nprl3*.

Considering a large-scale view of the interactions identified by Capture-C (Figure 1a), it is clear that, as for many previously analyzed loci[17], and in keeping with previous 3C analyses of the human and mouse α globin cluster, most *cis*-interactions occur within a region of ~100 kb upstream of the most 5' α globin promoter and ~10 kb downstream of the locus (Figure 1b). However, analysis using large moving windows (of the size used in HI-C; 100kb-1Mb) reveals much weaker, diffuse interactions with other active regions of the genome (Supplementary figure 3). Closer inspection shows interactions between the α globin promoters and their distal regulatory elements are only seen when the α genes are active in erythroid cells and not when they are silent in non-erythroid cells (Supplementary figure 4a). The same pattern is seen in the β globin locus (Supplementary figure 4b) although no difference is seen at the *Hoxa7* locus

which is silent in both cell types (Supplementary figure 5). The predominant interaction of the α globin promoters occurs with MCS-R2, which is also known to exert the greatest effect on α globin expression *in-vivo*[18]. These interactions are consistently seen in duplicate experiments (Supplementary figure 4a and b) and are very similar to the interactions noted using the recently developed high-resolution 4C-seq assay (HR_4C-seq) in both the α (Figure 2a) and β globin loci (Supplementary figure 6a)[10].

Unlike the high-resolution 4C assays, Capture-C is able to simultaneously assay the same locus from a variety of perspectives in a single experiment (Figure 2b). Whereas the α genes interact with their distal regulatory elements in erythroid (Ter119+) cells, the closely linked, poorly expressed *Hbq* (theta) genes interact weakly. Similarly, the embryonic *Hba-x* (zeta) gene, which is silent in definitive erythroid cells interacts weakly even though it is physically closer to the distal elements than the α genes. Similar changes in developmental stage-specific interactions were also noted for the β globin cluster (Supplementary figure 6b). Of interest, the promoter of the adjacent *Nprl3* gene also interacts with the erythroid elements contained within its introns (Figure 2b) and, as noted previously, its expression is up-regulated in an erythroid-specific manner [1,19]. Finally, the promoter of the next gene upstream from the *Nprl3* gene (*Mpg*) interacts only minimally with these elements even though physically it lies just as close to these elements as the *Nprl3* promoter (Figure 2b).

These multiple, simultaneous analyses also give insight into the less dynamic interactions with putative structural elements including CTCF/Cohesin binding sites. The α globin domain is bounded at the 3' end by CTCF binding sites associated with the 3' end of the *Hbq* genes while the 5' end of the domain is characterized by interactions with several CTCF sites (at -29, -33, -40, -60 and -78 kb relative to the promoter of the *Hba-x* gene) in the bodies of upstream genes. In summary, Capture-C used in a single multiplexed analysis allowed quantification and statistical quantile ranking of the density of interactions of all potential binding sites within a locus. Analysis of the mouse α and β globin loci, as test cases, agrees with all previous 3C analyses, identifying all known regulatory elements of the globin genes and no additional elements[10,20-23].

Analysis of multiple loci in a single experiment

The α and β globin genes were included as performance controls for the Capture-C protocol. In addition, 71 gene promoters were selected either because of their proximity to the α globin locus or because of their known functional importance in erythropoiesis. The remainder of the gene promoters (384) where chosen

based on the increased expression of their associated genes in mouse erythroid Ter119+ cells relative to ES cells (determined by cufflink analysis of RNA-seq data). These criteria were used to select genes whose expression was likely regulated by enhancer-promoter interactions during erythropoiesis.

As the Capture-C experiment successfully identified all previously characterized interactions between the globin genes and their *cis*-regulatory elements, it seemed likely that similar interactions of the other chosen 455 promoters would also be identified, within the same experiment. Analysis of the capture efficiency across all of the targeted regions shows that by comparison to the α globin genes the majority (92%) were successfully captured. All such interactions were analyzed as for the globin loci. These interactions can also be visualized with the complementary DNase-seq, ChIP-seq and RNA-seq data in a custom GBrowse database (see URLs section).

The enhancer-promoter interactions of relatively few erythroid genes have been studied in detail. By contrast, analysis of the *Tal1* (*Scl*) gene, which encodes a transcription factor essential for normal erythroid development, has been extensively characterized and is a good example to provide additional evidence for the robustness of the Capture-C approach [24-26]. Both chromatin and functional studies have shown that, in erythroid cells, a single, intergenic erythroid enhancer (+40) lying 40 kb away from the *Tal1* gene up-regulates expression of both *Tal1* and another gene (*Pdzk1ip1*/*Map17*) lying adjacent to the *Tal1* gene[24]. Capture-C from the promoter of either gene clearly identified interactions with the +40 enhancer and also picked up a weaker interaction with the promoter of another nearby gene *Stil* and an associated CTCF binding site (CTCF -11.6, Figure 3a and Supplementary figure 7). Interestingly analysis of the expression of the *Stil* gene shows it is also upregulated in mouse erythropoiesis[27] and human bone marrow (GNF atlas) suggesting it may also be influenced by the +40 enhancer. Several additional examples of promoters sharing a common enhancer were seen in the data set (e.g. the Isca1 and Zcchc6 genes on chr13 and the Ypel4, Ube2l6 and Slc43a1 genes on chr2).

Given its performance in analyzing genes whose regulation is well characterized, it was also of interest to use Capture-C to analyze currently unknown promoter-*cis*-element interactions involving genes of known functional relevance. Mitoferrin 1 (*slc25a37*/*Mtrfn1*) is an essential mitochondrial iron transporter that plays an essential role in heme biosynthesis in erythroid cells[28]. Capture from the *Mtfrn1* promoter identified six elements (HS8, HS17 HS-20.4, HS-23, HS-34 HS-37.5), which are variously bound by erythroid transcription factors and bear the chromatin signatures of enhancers (Figure 3b and Supplementary figure 8). In addition Capture-C identified potential interactions with CTCF binding sites. Of interest, recent zebrafish and mouse transgenic assays strongly suggest that HS-20.4 and HS-37.5 elements are *bona fide* erythroid-specific

enhancers which physically and functionally resemble the β globin locus control region[29].

General principles of *cis*-regulation revealed by Capture-C

By analyzing large numbers of enhancer promoter interactions, Capture-C will be useful for establishing the principles by which such *cis*-regulatory sequences interact. For example, consistent with previous 5C studies[30], using Capture-C, we show that most interactions occur within ~300 kb of the promoters studied. Furthermore, the frequency of interactions with all classes of *cis*-elements decreases inversely with distance from the TSSs (supplementary figure 9). Capture-C also shows that the interactions between promoters and *cis*-elements are restricted to discrete genomic regions with a clear demarcation (or boundary) between an interaction 'domain' and the flanking regions of the genome (Figure 2, 3, 4a and supplementary figure 6). Close inspection of the *cis*-elements interacting with each promoter very often showed that in addition to interacting with specific elements (enhancers, other promoters and CTCF bound sites), presumably via looping, there is a general increase in interactions with the intervening chromatin. A very clear example is seen at a locus containing five genes (1810031K17Rik, Fam134a, Zfand2b, Atg9a and Dnajb2) that are upregulated in erythroid cells. Each of these promoters show a high density of interactions with the chromatin across a common 130kb domain, but with very low levels of interaction beyond the boundaries of this domain (supplementary figure 10). This pattern of interaction is seen at most of the loci analyzed here (supplementary figure 11).

Linking SNPs in distal regulatory elements to the genes they control

Given that most SNPs associated with complex traits lie in regulatory rather than coding regions of the genome[11], an important application of Capture-C will be to link regulatory SNPs to the genes whose expression they affect, in a high throughput manner. To illustrate this we re-analyzed our dataset of erythroid DNase-seq in 8 strains of mice[31] and asked if any of the *cis*-regulatory erythroid elements predicted by Capture-C in the current study, varied within these strains. In addition we asked what would be the consequences of such variation?

We found such an example in the *Pnpo* gene (Figure 4a) encoding pyridoxamine 5'-phosphate oxidase[32], which is highly upregulated in erythroid cells[27]. We found that the promoter of *Pnpo* is linked to a regulatory element with the signature of an enhancer and binds the erythroid transcription factors Gata1 and Scl1. The regulatory element lies 26 kb away from the *Pnpo* promoter and a non erythroid gene (*Prr15l*) lies between the *Pnpo* gene and its regulatory element. Furthermore, the regulatory element lies only ~1kb upstream of the *Cdk5rap3* gene promoter which is expressed but not up-regulated in erythroid tissues. Therefore, it appears that this regulatory element is not controlling the adjacent

genes (*Prr15l and Cdk5rap3*) but a more distant gene (*Pnpo*); a situation that is under-recognized but probably very common in the mammalian genome.

Using the very deep DNase-seq data[31] (~200 million reads per strain) we determined DNase1 footprints[33] across the putative regulatory element. The first footprint contains a GATA1/SCL1 binding site although no currently known erythroid binding site could be found underlying the second footprint (Figure 4b). The DHS sites and associated footprints in three strains (A/J, AKR/J, CBA/J) are essentially missing and each of these strains share 3 SNPs (haplotype A) that are not present in the other five strains (haplotype B) (Supplementary figure 12). Importantly, two of the three SNPs lie within the protein bound regions, while the third lies within the DNase 1 site but not within a footprint (Figure 4b). It thus seems most likely that one, or a combination of the SNPs affect binding of one or more transcription factors to this regulatory element. Finally, we found that in the three strains with SNPs in the distal regulatory element (and absent DNase1 footprints) there is drastically reduced expression of Poly(A)+ RNA (stable transcription, Figure 4c) or Poly(A)- RNA (nascent transcription, Supplementary figure 13) from the gene (*Pnpo*) to which the regulatory element was linked by Capture-C.

Discussion

Recent advances in genomics have identified a large proportion of all *cis*-regulatory elements throughout the genome in hundreds of cell types[4]. In parallel with this, an ever-increasing number of DNA variants are being identified in animal populations, including humans[34]. The next critical step in deciphering this information is to develop high throughput, methods to link *cis*-elements (enhancers, promoters and CTCF bound sequences) to the genes they control and determine how variants alter gene expression. In this way it will be possible to combine Capture-C with statistical genetic approaches [35] to analyze cell types in which the intersection of GWAS hits and polymorphisms in DNase1 hypersensitivity, chromatin modification and/or transcription indicates a functional change in a *cis*-acting element[11].

Here we show that the Capture-C approach provides a simple high throughput method that will be suitable for this approach. It overcomes many of the limitations of previous methods providing an unbiased, high-resolution map of *cis*-interactions for hundreds of genes in a single experiment. Paired-end sequences derived from this method are simply mapped to the genomic sequence to generate a genome-wide map of interaction density with the elements of interest without the need for complex bioinformatic analysis or mathematical modeling.

The technique was initially validated using the α and β globin gene loci in which the *cis*-acting sequences, their chromatin modifications and interactions, have been previously characterized in great detail. These two regions of the genome have been used as the standard by which to assess any new chromosome conformation assays. Capture-C reproducibly identified all previously described *cis*-interactions within the globin loci. Furthermore, using Capture-C, it was possible to view all of these interactions from complementary viewpoints in the same experiment. Similarly, we identified and extended previously described interactions at other, less well characterized genomic loci where, for example, two genes share a single enhancer [24] and where multiple enhancers regulate a single gene[29].

The analysis of all *cis*-interactions with the ~450 promoters (TSSs) studied here confirmed and extended previous 3C sequencing and 5C experiments. In agreement with analysis of other large datasets [30] our data show that promoter interactions with *cis*-elements are most likely to be found in a ~600 kb region around the promoter (Supplementary figure 9). Furthermore, the likelihood of an interaction between a *cis*-acting sequence and any specific promoter decreases with distance from the promoter.

We anticipate that a major use of this method will be to link SNPs found in regulatory elements to the genes whose expression they influence, thereby providing a possible mechanism by which such GWAS hits create QTLs. By combining Capture-C data with our previously published DNase1 analysis of eight strains of mice[31] we have illustrated how SNPs may be linked to changes in expression of a specific gene lying within a complex chromosomal landscape (Figure 4). By design, Capture-C will allow such analysis to be performed in a high throughput manner.

Viewing the interactions within a single domain, from multiple viewpoints, from a large number of promoters, also revealed an interesting pattern that is relevant to the current debate about the mechanism by which promoters and enhancers interact *in-vivo*[36,37]. Consistent with previous models, the results of Capture-C suggest that sequences interact transiently sampling (or scanning) a relatively large domain of chromatin, with preferred interactions at specific regulatory elements, rather than exclusively via the well-defined *cis*-acting sequences (promoters and enhancers) alone. Nevertheless, Capture-C suggests that the sampling is limited to a defined domain of chromatin. What sets the boundaries of such domains is not yet clear. This "sampling" effect is seen at the majority of loci studied here (Supplementary figure 10) but is most clearly illustrated by a group of five genes (spanning a domain of ~150 kb) that are upregulated in erythroid cells. Capture from each of the five promoters identifies significant interactions across the entire well demarcated 150 kb domain suggesting that

these genes may occupy a common nuclear substructure (Supplementary figure 11) with defined limits at either end.

Together these findings demonstrate how the Capture-C technique will further elucidate the mechanisms by which *cis*-acting regulatory elements interact to control gene expression in time and space and facilitate identification of the functional DNA variants underlying complex diseases.

Acknowledgements

We thank James Davies and Paolo Piazza for technical suggestions. We thank Maria Suciu, Bryony Graham and Tom Milne for suggestions and critically reading the manuscript. We thank Zong-Pei Han for computational support. We thank the High-Throughput Genomics Group at the Wellcome Trust Centre for Human Genetics (funded by Wellcome Trust grant reference 090532/Z/09/Z and MRC Hub grant G0900747 91070) for the generation of the Sequencing data. This work was supported by Medical Research Council (UK),

Author Contributions

J. Hughes, D. Higgs and R. Gibbons designed experiments. J. Hughes, N. Roberts, D. Hay, M. Lynch and M. De Gobbi performed experiments. J. Hughes performed bioinformatic analysis. E. Giannoulatou performed statistical analysis. S. McGowan and S. Taylor provided bioinformatic support. J. Hughes and D. Higgs wrote the manuscript.

Competing Interests

The authors declare no competing interests.

URLS

Capture-C mouse erythroid database.

https://gbrowse2.molbiol.ox.ac.uk/gbrowse2/MM9_CapC_erythroid/capc_full.html

Seqmonk software

www.bioinformatics.babraham.ac.uk/projects/seqmonk/

Accession numbers

Capture-C data GSE47758

Ter119+ cell ChIP-seq data GSE47492

Mouse ES cell RNA-seq data GSE47757

Figure Legends

Figure 1. Interaction profile viewed from the α globin gene promoters

a, Shows a 5 Mb region around the mouse α globin locus. The chromosome and scale in Mb are shown at the top and the position of RefSeq annotated genes are shown below. The Capture-C track shows the cumulative density of interactions using a sliding window of 2 kb and a moving increment of 200 bp and is shown relative to the position of the Capture-C probes (red boxes). This profile represents the combined data from both erythroid biological replicates and represents 3,296 unique interactions normalized for capture efficiency at the promoter region. Key chromatin marks from the same cell type are shown relative to this and are coded by color, DNase1 Hypersensitivity (DHS) in green, mono-methylation of histone H4 (H3K4me1) in blue and tri-methylation of histone H4 (H3K4me3) in red.

b, Shows a zoomed in view of the above data covering a region of 200 kb spanning the α globin locus. In addition to the Capture-C probes designed to the α globin gene promoters (red boxes) probes designed at the promoters of the surrounding genes are shown (blue boxes). Interactions between these regions and the α globin promoters have been bioinformatically removed as they are no longer quantitative (see main text). In addition to the chromatin marks described in Figure 1a, (DHS, H3K4me1 and H3K4me3) further ChIP-seq data for chromatin marks or chromatin associated proteins (H3K27ac in orange and CTCF in purple) and key transcription factors (GATA1, SCL, KLF1 and NFE2 all in black) are shown relative to the interaction profile. The position of the known α globin regulatory elements are highlighted by red dashed lines and labeled at the bottom of the figure. Black dashed lines highlight interactions with regions bound by the CTCF protein.

Figure 2. Validation of the Capture-C technique

a, Shows the comparison between conventional high-resolution 4C data (HR_4C_seq was plotted from the analyzed interaction data (median values within a 2kb window, centered on the 1kb genomic bin) supplied as part of the GEO submission from van de Werken et al[10]) from the *Hba* gene promoters in mouse erythroid cells. The DNase-seq, CTCF ChIP-seq data, gene annotation and capture probes are as described as in figure 1 over a 200 kb window around the locus. The previously characterized regulatory elements are highlighted with red dashed lines and labeled at the bottom of the panel.

b, Shows the comparison of the interactions of 5 gene promoters in the α globin locus including the housekeeping genes *Mpg* (829 unique distal interactions) and *Nprl3* (1561 unique distal interactions), the inactive embryonic globin gene *Hba-x* (822 unique distal interactions), the active adult globin genes *Hba* and the weakly expressed globin genes of unknown function *Hbq* (1811 unique distal interactions). In each track the positions of the capture probes used are shown as red boxes and the position of other captured regions (from which the data have been excluded in that track) are annotated with blue boxes. The previously characterized regulatory elements are highlighted with red dashed lines and labeled at the bottom of the panel.

Figure 3 Characterization of the *Tal1* and *Mtrfn1* gene loci.

a, Shows a comparison between the interactions of the promoter of the *Tal1* gene (1,839 unique distal interactions) and the neighboring *Pdzk1ip1* (*Map17*) gene (1,133 unique distal interactions). In both tracks the position of the capture probes used are shown as red boxes and the position of other captured regions, from which the data have been excluded in that track, are annotated with blue boxes. The previously characterized regulatory elements are highlighted with red dashed lines and labeled at the bottom of the panel. A black dashed line highlights a pronounced interaction with a CTCF bound region. Interactions with the neighboring promoter of the *Stil* gene are highlighted with a blue dashed line.

b, Shows the interaction profile of the promoter of the essential erythroid gene *Slc25a27* (*Mtrfn1*, 787 unique distal interactions). The positions of the capture probes used are shown as red boxes. Interactions with DHS bound by erythroid transcription factors (see Supplementary figure 8) are highlighted with red dashed lines and labeled depending on their distance in kb from the *Slc25a27* promoter (- in the transcriptional upstream direction and + in the downstream direction). Black dashed lines highlight pronounced interactions with CTCF bound regions.

Figure 4 Using Capture-C to determine the effect of distal SNPs on gene regulation.

a, Shows the interaction profile of the promoter of the erythroid upregulated gene *Pnpo* (1,905 unique distal interactions). The positions of the capture probes used are shown as red boxes. Interactions with DHS bound by erythroid transcription factors are highlighted with red dashed lines and labeled depending on their distance in kb from the *Pnpo* promoter (- in the transcriptional upstream direction and + in the downstream direction) Key chromatin marks from the same cell type are shown relative to this and are coded by color, DNase1 Hypersensitivity (DHS) in green, mono-methylation of histone H4 (H3K4me1) in blue and tri-methylation of histone H4 (H3K4me3) in red, and CTCF in purple and key transcription factors GATA1 and SCL in black.

b, Shows a 665 bp region around the variable enhancer. The DNase footprints derived from high read depth DNA-seq data (Hosseini et al Plos Genet.[31]) are shown for 2 mouse strains representative of Haplotype A (with SNPs, CBA/J strain) and Haplotype B (without SNPs, DBA/2J strain). The position of the 3 SNPs found in haplotype A are depicted as red ticks and their position within the DNase1 footprint within haplotype B are shown by red hatched lines and named at the bottom.

c(i) Shows the read-normalized signal (per million reads aligned) for erythroid PolyA+ RNA-seq (Hosseini et al Plos Genet.[31]) over a 722kb region surrounding the *Pnpo* locus in 8 mouse strains labeled as either haplotype A or B (see supplementary figure 12). The position of RefSeq annotated genes are shown at the bottom and transcription associated with the *Pnpo* gene is highlighted with a red-hatched box.

c(ii), Shows the transcription of the neighboring well-expressed *Kpnb1* gene as a control for RNA-seq comparability.

Online Methods

Primary cells and cell lines

Mature primary erythroid cells were obtained from phenylhydrazine-treated adult (6-9 weeks old) female mice as described [38] Mouse E14-TG2a.IV ES cells were grown as previously described [39]. Mice were maintained in specific-pathogen-free facilities at the Biomedical Services Unit, Oxford. Protocols were approved through the Oxford University Local Ethical Review process. Experimental procedures were performed in accordance with the European Union DIRECTIVE 2010/63/EU and/or the UK Animals (Scientific Procedures) Act, 1986 (Project License 30/2805).

Generation of 3C libraries

The protocol used for generation of the 3C libraries was similar to previous methods [22,40,41] with minor modifications. Briefly, 10^7 cells were crosslinked with 2% formaldehyde (10 minutes, room temperature); quenched with cold glycine; washed in phosphate buffered saline; resuspended in cold lysis buffer (tris 10mM, NaCl 10mM, NP40 0.2%, complete proteinase inhibitor (Roche) and snap frozen to -80^{0}C. Cells were thawed on ice, washed in dH_2O and Dounce homogenized on ice (x 40 strokes). Cells were then resuspended with 0.25% SDS and restriction enzyme buffer (1X NEBuffer DpnII) and incubated at 37^{o}C for 1h at 1400rpm on a Comfort Thermomixer (Eppendorf) followed by a further incubation of 1h following the addition of Triton X100 (source final concentration 1.67%). An overnight digestion was performed using Dpn II (NEB) at 500U /ml at 37^{0}C and 1400 rpm (Comfort Thermomixer Eppendorf).

The digested chromatin was ligated overnight (Fermentas HC Ligase final concentration 10U/ml) at 16 degrees at 1400 rpm on the Comfort Thermomixer. The samples were then decrosslinked overnight at 65^{o}C with Proteinase K (Roche) followed by a 30 minute incubation at 37^{o}C with RNAse (Roche). Phenol/Chloroform extraction was then performed followed by an Ethanol precipitation and a wash with 70% Ethanol. Digestion efficiencies were assessed by gel electrophoresis (1% agarose) and QPCR (Taqman Invitrogen), which showed digestion efficiencies in excess of 80%.

Performance of Capture C assay

DNA content of the Dpn II 3C libraries was quantitated using a Qubit fluorometer (Life technologies) and 5ug of each library was sheared in dH_2O using a Covaris S2.

Covaris settings used were: duty cycle 10%, Intensity 5, Cycles/burst 200, Time 6 cycles of 60seconds, Set Mode Frequency sweeping Temperature 4 to 7 degrees. Following shearing DNA was purified using AMPureXP beads (Agencourt) and DNA quality assessed on a Bioanalyser 2100 using a DNA 1000 Chip (Agilent).

DNA end repair and adapter ligation was performed using the NEB Next DNA sample preparation reagent kit using the standard protocol. Excess adapters were removed using AMPureXP beads.

500ng of each adapter ligated library were used as input for the SureSelect target enrichment kit using the recommended protocol and the custom designed Capture-C library. The quality of the resultant captured library was assessed by Bioanalyser using DNA1000 chip. The generated libraries fell within the expected guidelines for output of the SureSelect target enrichment kit and were submitted for Hi-Seq paired sequencing (50 bp read length).

DNase1 seq , Chip-Seq and RNA-seq

The DNase-seq and ChIP-seq data (H3K27ac, H3K4me1 and H3K4me3) for erythroid Ter119+ cells are from previous published work from the authors described in Kowalcyzk et al[1].

The DNase-seq, polyA+ and PolyA- RNA-seq from Ter119+ cells in 8 strains of mice are from Hosseini et al[31] and their alignment and analysis is described therein. The DNase1-seq alignments were converted to DNase1 footprints by using the uniquely aligned paired-end reads across the *Pnpo* locus to determine the DNase1 cut sites within the chromatin[33]. The density of these positions was calculated using a 3bp window and converted in to a UCSC bigwig file for viewing (Figure 4b and supplementary figure 12).

ChIP-seq data for SCL/TAL1[42] and Klf1[43] were from previously published datasets.

ChIP-seq for GATA1, NFE2 and CTCF were performed as described [16] using antibodies ab11852 (Abcam)[44], SC-477 (Santa Crus)[45] and 07-729 (Millipore)[46] respectively. ChIP-seq libraries were prepared and sequenced using the standard Illumina paired end protocol

RNA-seq from PolyA+ RNA from mouse E14-TG2a.IV ES cells was performed as previously described [1].

Peak calling and annotation of *cis* elements

The *cis*-element annotation based on DHS and relative quantitation of the H3K4me1 and H3K4me3 chromatin marks is as described in Kowalcyzk et al[1]. This annotation was supplemented with CTCF bound regions in Ter119+ cells. ChIP-seq data was aligned to the mm9 mouse genome build using bowtie (version 0.12.3, http://bowtie- bio.sourceforge.net/index.shtml) [47]. To prevent the exclusion of the duplicated globin genes bowtie was run with the –m reporting option set to 2 to allow reads to map twice to the genome. To exclude over-amplified products from these data sets, read pairs that map to exactly the same genomic position were collapsed into a single representative set of reads. Peaks of CTCF binding were called using Seqmonk (version 0.21.1, see URL section) with following parameters; Depth cut off=18; Minimum size=50 and Merge Distance=50. Problematic regions caused by copy number differences were removed as described in Kowalcyzk et al[1]. The CTCF peaks and DHS peaks were merged to produce a union file representing the *cis* elements in the genome of Ter119+ cells.

Design of the custom SureSelect capture library.

Gene promoters were included either as manual picks or due to their upregulation in Ter119+. Promoters were manually included either due to their known importance in erythropoiesis or if they formed part of the α globin control locus (supplementary table 1). Gene upregulation was determined by comparative analysis of PolyA+ RNA-seq data from Ter119+[1] and mES cells using the Cufflinks and CuffDiff programs[48]. From the determined statistically upregulated genes 360 genes were randomly added to the 95 manual picks to complete the design.

The position of the gene promoters were based on the transcriptional start sites (TSSs) annotated in RefSeq and UCSC "known gene" datasets for mouse genome build MM9. The annotated TSSs were then manually checked against the Ter119+ RNA-seq, DHS and chromatin data to refine the promoter annotation and to include alternate promoters where necessary.

Custom PERL scripts (FragExtract.pl and SureC2Oligo.pl) were then used to isolate the positions of the Dpn II fragments containing the target promoters and to generate an oligo walk of 140bp sequences from the flanking Dpn II sites. The oligos were then assessed for uniqueness in the genome using custom PERL scripts (DepthGauge.pl and MergeSimpleRepeats.pl) which analyze and annotate the outputs of BLAT[49] and Repeatmasker[50] respectively. The resulting metadata associated with each oligo was then used to generate a MIG database (https://mig.molbiol.ox.ac.uk/mig/) and filtered on repeat density score (Density <= 10 from DepthGuage.pl) and a minimum simple repeat content <= 30. The remaining designed sequences were submitted for production as a custom SureSelect library.

After peak calling the aligned erythroid data using Seqmonk (version 0.21.1) and analysis for other regions of enrichment other than capture regions in a MIG database[51] it was noted that off target capture predominated at sites of pseudogenes related to captured promoters, regions which had undergone segmental duplication or across highly related gene families. As a result of this a flag for duplicated regions has be incorporated into the design pipeline, as excess off capture is an important design consideration. Also, as with all hybridization and PCR based methods, capture probes with excess GC or CpG content performed poorly in the capture reaction (supplementary figure 2) and should excluded from a Capture-C design or another restriction enzyme should be considered to change the position of the ends of the target fragment.

Alignment of reads and generation of normalized density plots of Capture-C data.

Paired-end sequences, were mapped as single-end reads to overcome the built in assumptions of the relative positioning of paired-end sequences in the sequence alignment programs. The sequences were mapped to an *in silico* Dpn II digested version of MM9 (UCSC Genome browser NCBI build 37). Reads which did not align in the first pass were tested to see if they contained a Dpn II restriction site and if so were assumed to be a potential non-alignable junction fragment. These fragments where split at the Dpn II site and if the smallest resulting fragment was >= 20 bp the individual fragments were remapped.

Sample	Read Number	Aligned to MM9 build
Erythroid 1	97M	85%
Erythroid 2	340M	73%
mES	242M	80 %

The paired-end relationship between the reads were re-established and quantified using the custom script Frag2Arc.pl, excluding read pairs which map to the same Dpn II fragment and stored the resultant ligation events in a GFF3 file.

A pipeline of PERL scripts (ExtractLoci.pl) was created to isolate interactions originating from each of the targeted promoter fragments; generate a density map of the interaction across the relevant chromosome over a given window and increment of movement; to normalize this density map on the density of signal at the capture region and to generate wig and bigwig files for each locus in the design. The density of interactions with the merged DHS/CTCF track was

quantified in a 1kb window around each of the elements in *cis* to the target promoter. Read pairs which mapped to precisely the same genomic coordinates could be assumed to be PCR artifacts and so collapsed to a single representative interaction. Comparison between the profiles generated at the control loci showed little difference before and after collapsing the data indicating the library was suitably complex (supplementary figures 14 and 15). The resultant analysis of the pooled biological replicates was stored in a custom GBrowse 2.0 database [52] (see URL section). Control regions to assess background sequence and interaction enrichment were generated from 6,602 promoters which are not included in the present design but are associated with DNase1 hypersensitive sites in Ter119+ cells. Capture probes were designed using the same algorithms and criteria as the targeted regions included in the design. In this way the background enrichment could be assessed at regions similar in function, chromatin structure and sequence content to the regions enriched in the present Capture-C design. The background distribution was generated by randomly sampling x 100 times, the same number of control regions as the actual number of regions in the design (1,911). All scripts are available on request.

Identification of statistically significant interactions

The normalized density of all interactions with *cis*-elements for all of the target promoters were pooled to generate a background population. The limits of background variation were identified using boxplot-defined thresholds. Interactions of moderate significance (greater than 75th, but less than 95th percentile) were estimated as the values outside the lower and upper inner fence: Q1 ± 1.5*IQR and the upper inner fence: Q3 + 1.5*IQR, where IQR is the interquantile range, while Q1 and Q2 are the lower and upper quartiles respectively, defined as the 25th and 75th percentiles. Similarly, strongly significant interactions (greater than 95th percentile) were identified as the values outside of the lower outer fence: Q1 - 3*IQR and upper outer fence: Q3 + 3*IQR. As the distribution is right-skewed (positive values only), interactions correspond to the values above the upper inner and outer fences.

1. Kowalczyk, M.S. *et al.* Intragenic enhancers act as alternative promoters. *Mol Cell* **45**, 447-58 (2012).

2. Shen, Y. *et al.* A map of the cis-regulatory sequences in the mouse genome. *Nature* (2012).
3. Rada-Iglesias, A. *et al.* A unique chromatin signature uncovers early developmental enhancers in humans. *Nature* **470**, 279-83 (2011).
4. Dunham, I. *et al.* An integrated encyclopedia of DNA elements in the human genome. *Nature* **489**, 57-74 (2012).
5. de Laat, W. & Dekker, J. 3C-based technologies to study the shape of the genome. *Methods* **58**, 189-91 (2012).
6. Sajan, S.A. & Hawkins, R.D. Methods for identifying higher-order chromatin structure. *Annu Rev Genomics Hum Genet* **13**, 59-82 (2012).
7. Dixon, J.R. *et al.* Topological domains in mammalian genomes identified by analysis of chromatin interactions. *Nature* **485**, 376-80 (2012).
8. Lieberman-Aiden, E. *et al.* Comprehensive mapping of long-range interactions reveals folding principles of the human genome. *Science* **326**, 289-93 (2009).
9. Sexton, T. *et al.* Three-dimensional folding and functional organization principles of the Drosophila genome. *Cell* **148**, 458-72 (2012).
10. van de Werken, H.J. *et al.* Robust 4C-seq data analysis to screen for regulatory DNA interactions. *Nat Methods* **9**, 969-72 (2012).
11. Schaub, M.A., Boyle, A.P., Kundaje, A., Batzoglou, S. & Snyder, M. Linking disease associations with regulatory information in the human genome. *Genome Res* **22**, 1748-59 (2012).
12. Bainbridge, M.N. *et al.* Whole exome capture in solution with 3 Gbp of data. *Genome Biol* **11**, R62 (2010).
13. Gondor, A., Rougier, C. & Ohlsson, R. High-resolution circular chromosome conformation capture assay. *Nat Protoc* **3**, 303-13 (2008).
14. Dekker, J., Rippe, K., Dekker, M. & Kleckner, N. Capturing chromosome conformation. *Science* **295**, 1306-11 (2002).
15. Hughes, J.R. *et al.* Annotation of cis-regulatory elements by identification, subclassification, and functional assessment of multispecies conserved sequences. *Proc Natl Acad Sci U S A* **102**, 9830-5 (2005).
16. De Gobbi, M. *et al.* Tissue-specific histone modification and transcription factor binding in alpha globin gene expression. *Blood* **110**, 4503-10 (2007).
17. Vavouri, T., McEwen, G.K., Woolfe, A., Gilks, W.R. & Elgar, G. Defining a genomic radius for long-range enhancer action: duplicated conserved non-coding elements hold the key. *Trends Genet* **22**, 5-10 (2006).
18. Wallace, H.A. *et al.* Manipulating the mouse genome to engineer precise functional syntenic replacements with human sequence. *Cell* **128**, 197-209 (2007).
19. Lower, K.M. *et al.* Adventitious changes in long-range gene expression caused by polymorphic structural variation and promoter competition. *Proc Natl Acad Sci U S A* **106**, 21771-6 (2009).
20. Vernimmen, D. *et al.* Chromosome looping at the human alpha-globin locus is mediated via the major upstream regulatory element (HS -40). *Blood* **114**, 4253-60 (2009).
21. Bau, D. *et al.* The three-dimensional folding of the alpha-globin gene domain reveals formation of chromatin globules. *Nat Struct Mol Biol* **18**, 107-14 (2011).

22. Hughes, J.R. *et al.* High-resolution analysis of cis-acting regulatory networks at the alpha-globin locus. *Philos Trans R Soc Lond B Biol Sci* **368**, 20120361 (2013).
23. Vernimmen, D., De Gobbi, M., Sloane-Stanley, J.A., Wood, W.G. & Higgs, D.R. Long-range chromosomal interactions regulate the timing of the transition between poised and active gene expression. *EMBO J* **26**, 2041-51 (2007).
24. Ferreira, R. *et al.* Impaired in vitro erythropoiesis following deletion of the Scl (Tal1) +40 enhancer is largely compensated for in vivo despite a significant reduction in expression. *Mol Cell Biol* **33**, 1254-66 (2013).
25. Ogilvy, S. *et al.* The SCL +40 enhancer targets the midbrain together with primitive and definitive hematopoiesis and is regulated by SCL and GATA proteins. *Mol Cell Biol* **27**, 7206-19 (2007).
26. Gottgens, B. *et al.* The scl +18/19 stem cell enhancer is not required for hematopoiesis: identification of a 5' bifunctional hematopoietic-endothelial enhancer bound by Fli-1 and Elf-1. *Mol Cell Biol* **24**, 1870-83 (2004).
27. Flygare, J., Rayon Estrada, V., Shin, C., Gupta, S. & Lodish, H.F. HIF1alpha synergizes with glucocorticoids to promote BFU-E progenitor self-renewal. *Blood* **117**, 3435-44 (2011).
28. Shaw, G.C. *et al.* Mitoferrin is essential for erythroid iron assimilation. *Nature* **440**, 96-100 (2006).
29. Amigo, J.D. *et al.* Identification of distal cis-regulatory elements at mouse mitoferrin loci using zebrafish transgenesis. *Mol Cell Biol* **31**, 1344-56 (2011).
30. Sanyal, A., Lajoie, B.R., Jain, G. & Dekker, J. The long-range interaction landscape of gene promoters. *Nature* **489**, 109-13 (2012).
31. Hosseini, M. *et al.* Causes and consequences of chromatin variation between inbred mice. *PLoS Genet* **9**, e1003570 (2013).
32. Kang, J.H. *et al.* Genomic organization, tissue distribution and deletion mutation of human pyridoxine 5'-phosphate oxidase. *Eur J Biochem* **271**, 2452-61 (2004).
33. Neph, S. *et al.* An expansive human regulatory lexicon encoded in transcription factor footprints. *Nature* **489**, 83-90 (2012).
34. Donnelly, P. Progress and challenges in genome-wide association studies in humans. *Nature* **456**, 728-31 (2008).
35. Bush, W.S. & Moore, J.H. Chapter 11: Genome-wide association studies. *PLoS Comput Biol* **8**, e1002822 (2012).
36. Pennacchio, L.A., Bickmore, W., Dean, A., Nobrega, M.A. & Bejerano, G. Enhancers: five essential questions. *Nat Rev Genet* **14**, 288-95 (2013).
37. Kolovos, P., Knoch, T.A., Grosveld, F.G., Cook, P.R. & Papantonis, A. Enhancers and silencers: an integrated and simple model for their function. *Epigenetics Chromatin* **5**, 1 (2012).
38. Spivak, J.L., Toretti, D. & Dickerman, H.W. Effect of phenylhydrazine-induced hemolytic anemia on nuclear RNA polymerase activity of the mouse spleen. *Blood* **42**, 257-66 (1973).
39. Nichols, J., Evans, E.P. & Smith, A.G. Establishment of germ-line-competent embryonic stem (ES) cells using differentiation inhibiting activity. *Development* **110**, 1341-8 (1990).

40. Hagege, H. *et al.* Quantitative analysis of chromosome conformation capture assays (3C-qPCR). *Nat Protoc* **2**, 1722-33 (2007).
41. Stadhouders, R. *et al.* Multiplexed chromosome conformation capture sequencing for rapid genome-scale high-resolution detection of long-range chromatin interactions. *Nat Protoc* **8**, 509-24 (2013).
42. Kassouf, M.T. *et al.* Genome-wide identification of TAL1's functional targets: insights into its mechanisms of action in primary erythroid cells. *Genome Res* **20**, 1064-83 (2010).
43. Tallack, M.R. *et al.* A global role for KLF1 in erythropoiesis revealed by ChIP-seq in primary erythroid cells. *Genome Res* **20**, 1052-63 (2010).
44. Papadopoulos, G.L. *et al.* GATA-1 genome-wide occupancy associates with distinct epigenetic profiles in mouse fetal liver erythropoiesis. *Nucleic Acids Res* **41**, 4938-48 (2013).
45. Zhou, Z. *et al.* USF and NF-E2 cooperate to regulate the recruitment and activity of RNA polymerase II in the beta-globin gene locus. *J Biol Chem* **285**, 15894-905 (2010).
46. Horakova, A.H., Moseley, S.C., McLaughlin, C.R., Tremblay, D.C. & Chadwick, B.P. The macrosatellite DXZ4 mediates CTCF-dependent long-range intrachromosomal interactions on the human inactive X chromosome. *Hum Mol Genet* **21**, 4367-77 (2012).
47. Langmead, B., Trapnell, C., Pop, M. & Salzberg, S.L. Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. *Genome Biol* **10**, R25 (2009).
48. Trapnell, C. *et al.* Transcript assembly and quantification by RNA-Seq reveals unannotated transcripts and isoform switching during cell differentiation. *Nat Biotechnol* **28**, 511-5 (2010).
49. Kent, W.J. BLAT--the BLAST-like alignment tool. *Genome Res* **12**, 656-64 (2002).
50. Tarailo-Graovac, M. & Chen, N. Using RepeatMasker to identify repetitive elements in genomic sequences. *Curr Protoc Bioinformatics* **Chapter 4**, Unit 4 10 (2009).
51. McGowan, S.J., Hughes, J.R., Han, Z.P. & Taylor, S. MIG: Multi-Image Genome Viewer. *Bioinformatics* (2013).
52. Stein, L.D. Using GBrowse 2.0 to visualize and share next-generation sequence data. *Brief Bioinform* **14**, 162-71 (2013).

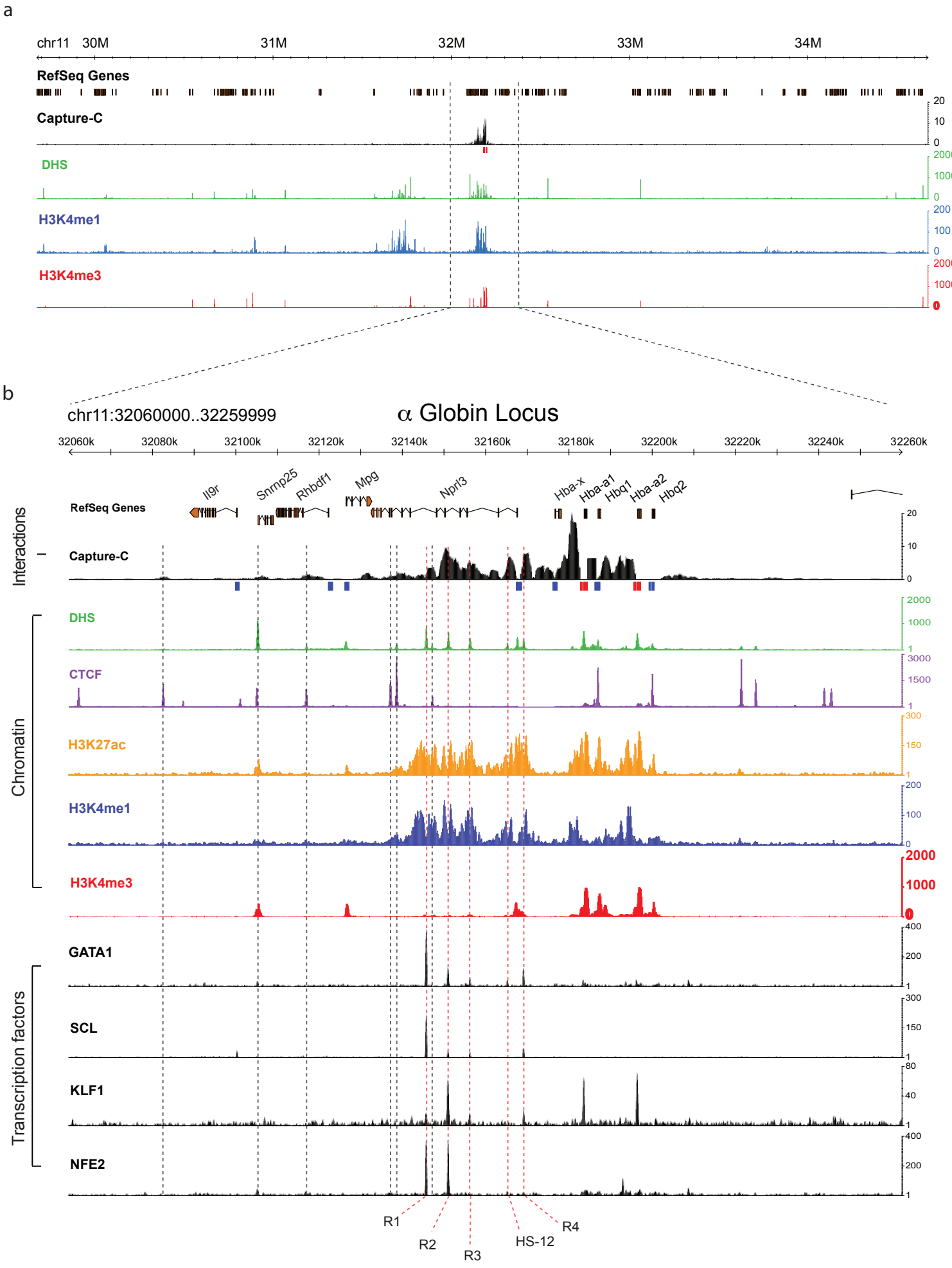
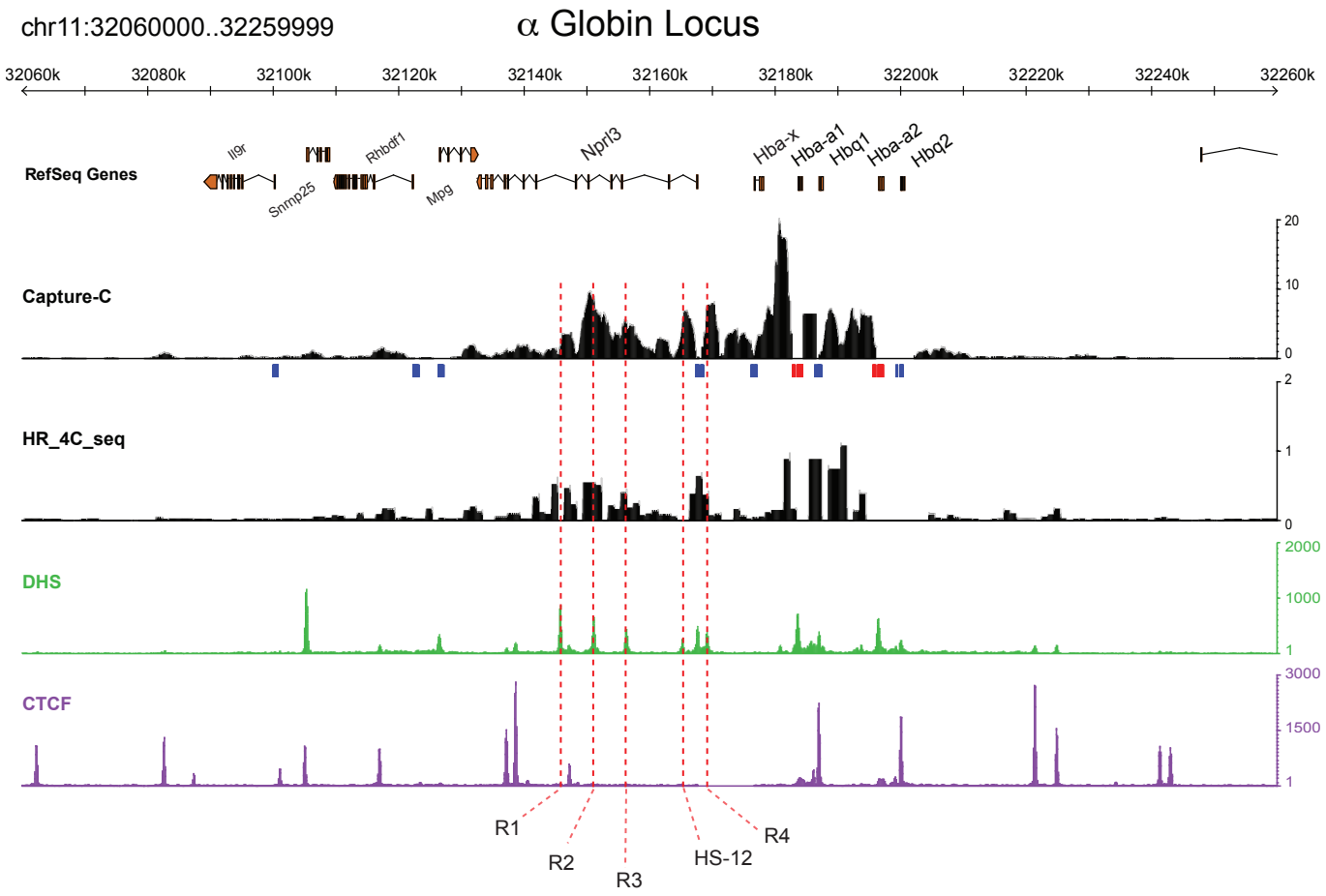

a
chr11 30M
31M
32M
33M
34M
RefSeq Genes
Capture-C
DHS
H3K4me1
H3K4me3
b
chr11:32060000..32259999
α Globin Locus
32060k
32080k
32100k
32120k
32140k
32160k
32180k
32200k
32220k
32240k
32260k
Il9r
Snrnp25
Rhbdf1
Mpg
Npri3
Hba-x
Hba-a1
Hbq1
Hba-a2
Hbq2
RefSeq Genes
Capture-C
Interactions
DHS
CTCF
H3K27ac
H3K4me1
H3K4me3
Chromatin
GATA1
SCL
KLF1
NFE2
Transcription factors
R1
R2
R3
HS-12
R4

Figure 1

a


chr11:32060000..32259999
α Globin Locus
32060k
32080k
32100k
32120k
32140k
32160k
32180k
32200k
32220k
32240k
32260k
RefSeq Genes
Il9r
Snrnp25
Rhbdf1
Mpg
Nprl3
Hba-x
Hba-a1
Hbq1
Hba-a2
Hbq2
Capture-C
HR_4C_seq
DHS
CTCF
R1
R2
R3
HS-12
R4

b

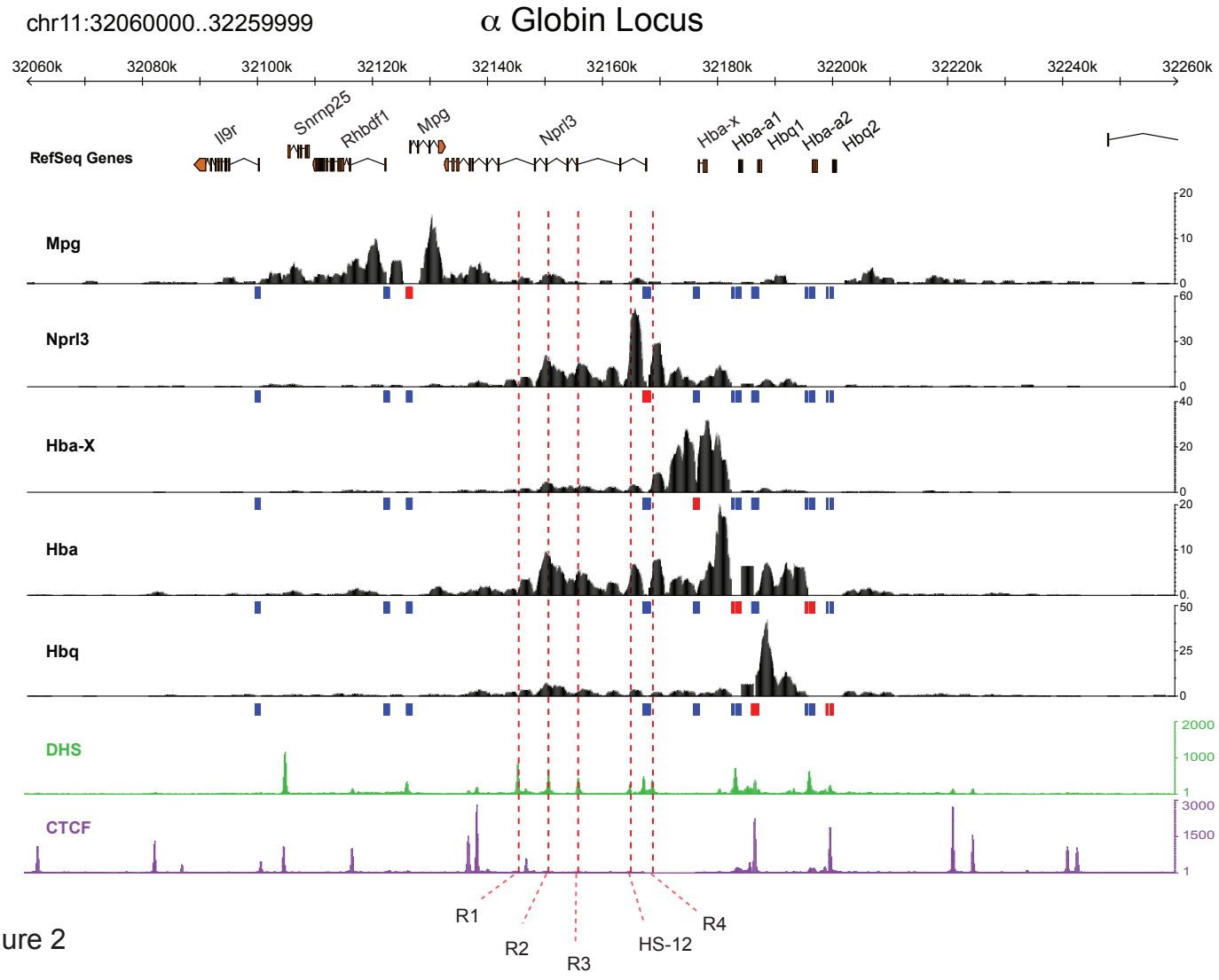
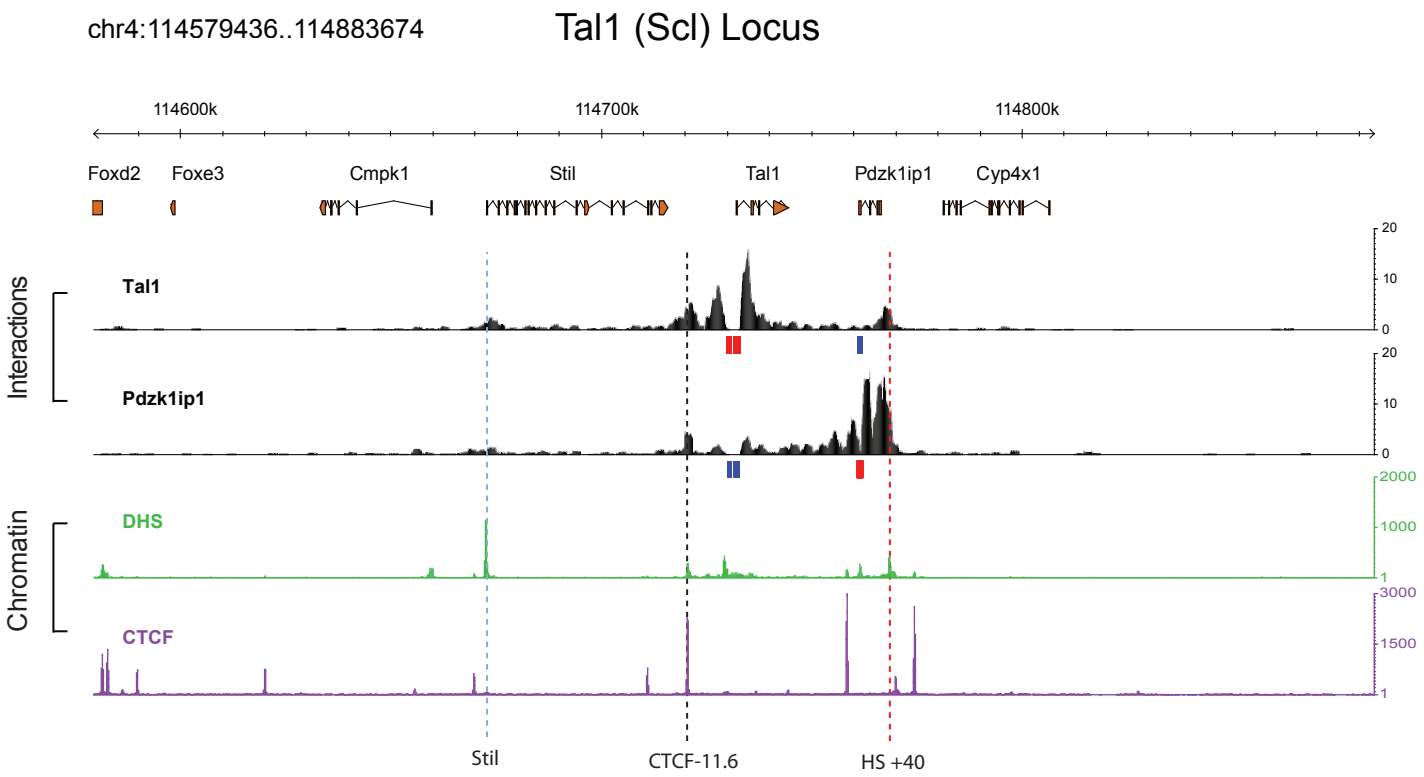

chr11:32060000..32259999
α Globin Locus
32060k
32080k
32100k
32120k
32140k
32160k
32180k
32200k
32220k
32240k
32260k
RefSeq Genes
Il9r
Snrnp25
Rhbdf1
Mpg
Nprl3
Hba-x
Hba-a1
Hbq1
Hba-a2
Hbq2
Mpg
Nprl3
Hba-X
Hba
Hbq
DHS
CTCF
R1
R2
R3
HS-12
R4

Figure 2

a


chr4:114579436..114883674
Tal1 (Scl) Locus
114600k
114700k
114800k
Foxd2
Foxe3
Cmpk1
Stil
Tal1
Pdzk1ip1
Cyp4x1
Interactions
Tal1
Pdzk1ip1
Chromatin
DHS
CTCF
Stil
CTCF-11.6
HS +40

b

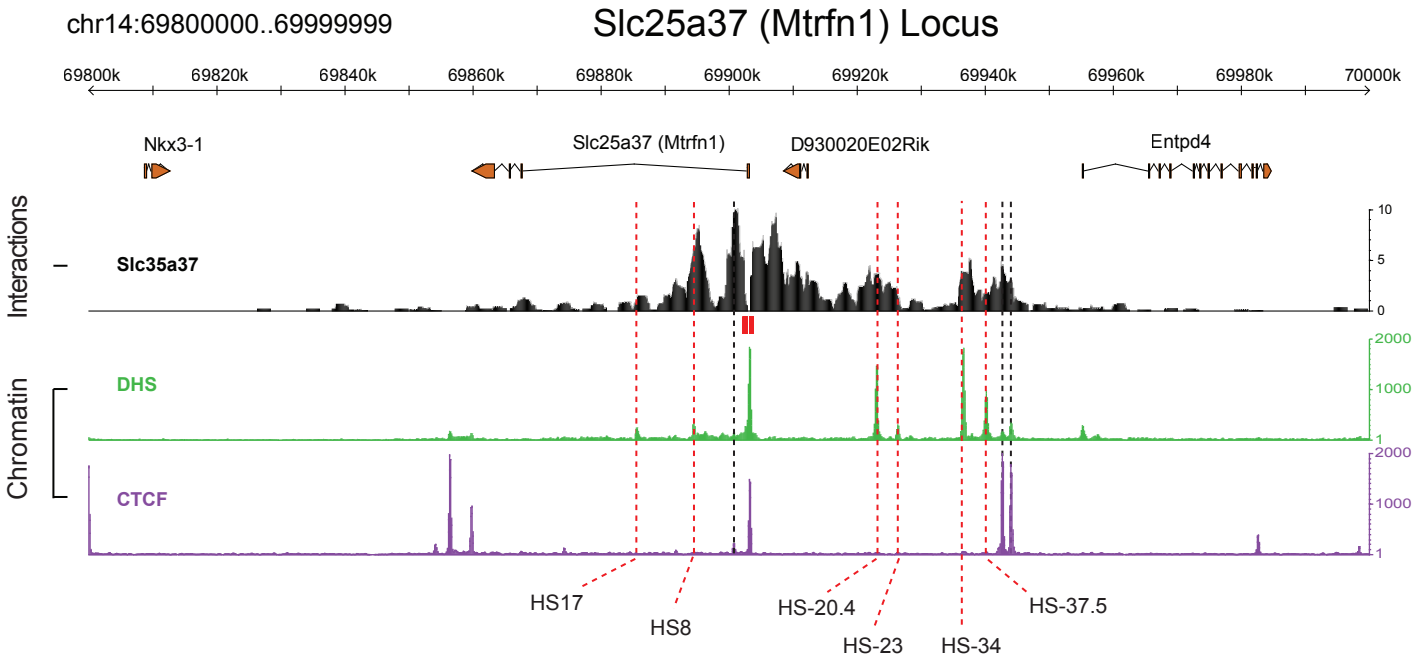

chr14:69800000..69999999
Slc25a37 (Mtrfn1) Locus
69800k
69820k
69840k
69860k
69880k
69900k
69920k
69940k
69960k
69980k
70000k
Nkx3-1
Slc25a37 (Mtrfn1)
D930020E02Rik
Entpd4
Interactions
Slc35a37
Chromatin
DHS
CTCF
HS17
HS8
HS-20.4
HS-23
HS-34
HS-37.5

Figure 3

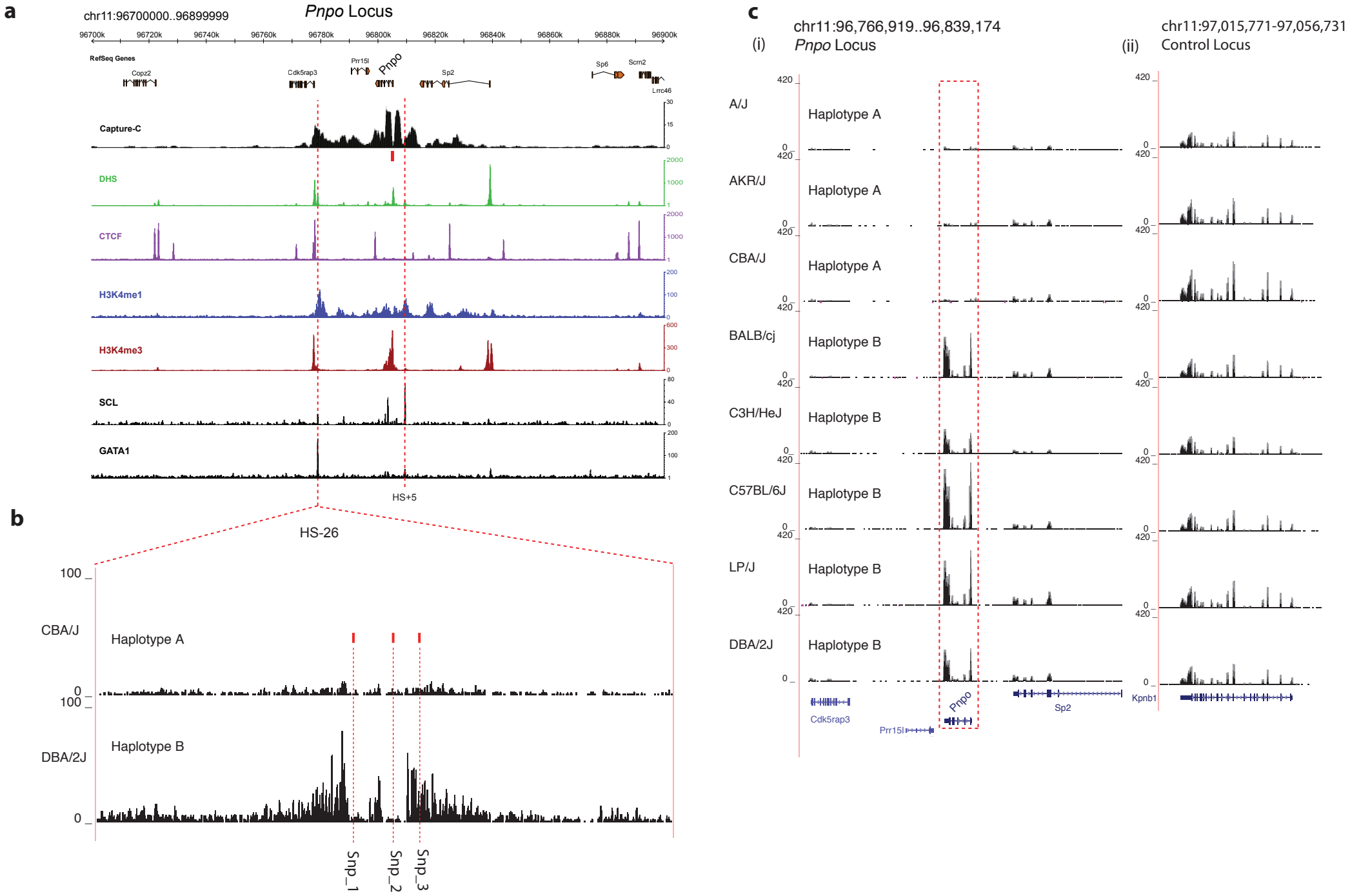
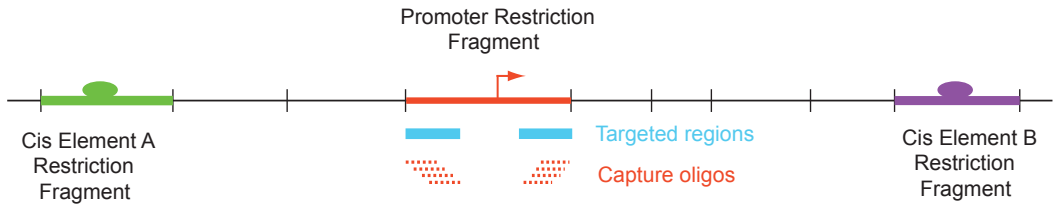

a
chr11:96700000..96899999
Pnpo Locus
96700k
96720k
96740k
96760k
96780k
96800k
96820k
96840k
96860k
96880k
96900k
RefSeq Genes
Copz2
Cdk5rap3
Prr15l
Pnpo
Sp2
Sp6
Scm2
Lrrc46
Capture-C
DHS
CTCF
H3K4me1
H3K4me3
SCL
GATA1
HS+5
b
HS-26
100
CBA/J
Haplotype A
0
100
DBA/2J
Haplotype B
0
Snp_1
Snp_2
Snp_3
c
(i)
chr11:96,766,919..96,839,174
Pnpo Locus
(ii)
chr11:97,015,771-97,056,731
Control Locus
A/J
AKR/J
CBA/J
BALB/cj
C3H/HeJ
C57BL/6J
LP/J
DBA/2J
Haplotype A
Haplotype B
420
0
Cdk5rap3
Prr15l
Pnpo
Sp2
Kpnb1

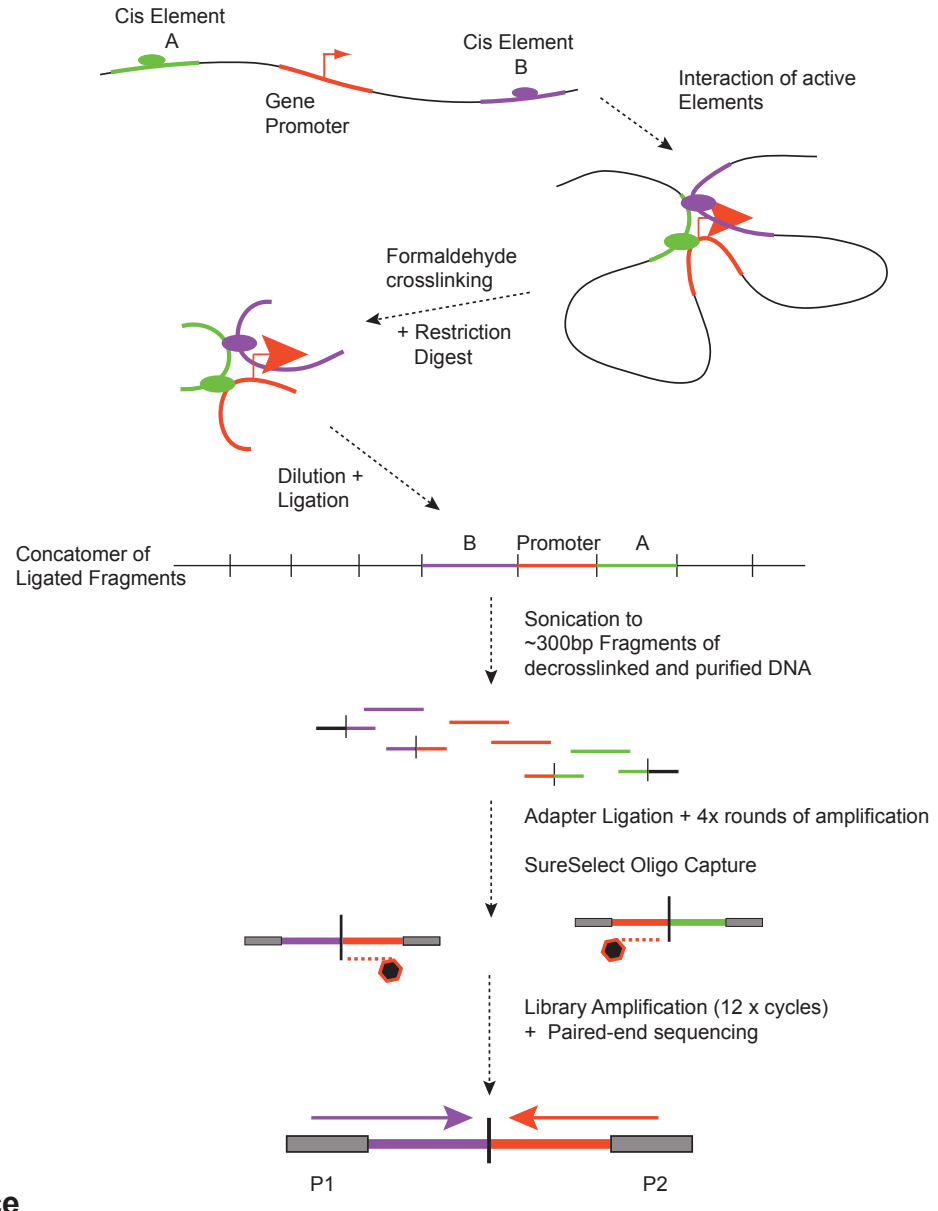
Figure 4

Design and performance of the Capture-C assay

a, Design

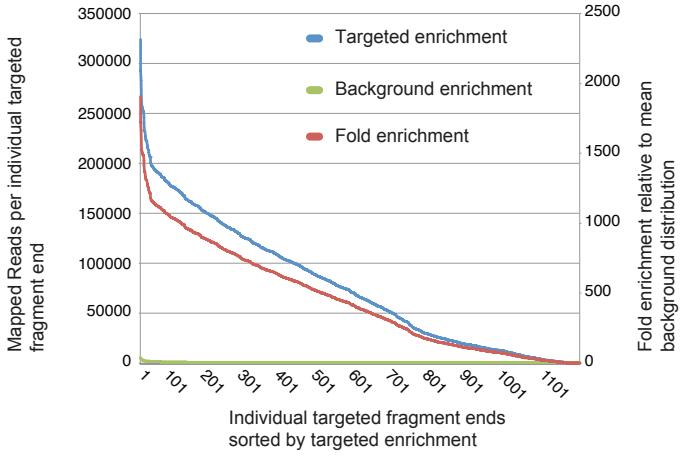


b, Assay

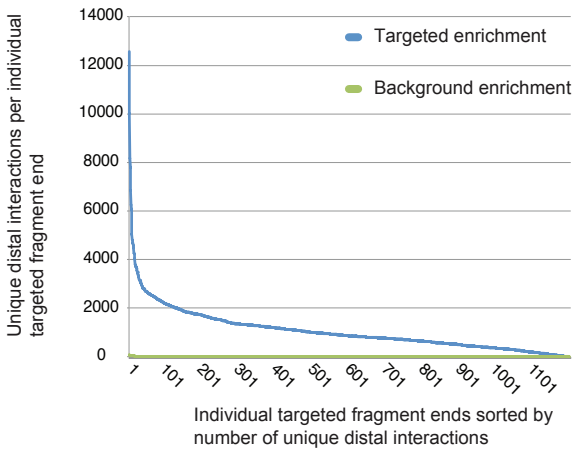


c, Performance

(i) Sequence Read Enrichment



(ii) Unique Interaction Enrichment



Supplementary figure 1.

Supplementary figure 1.

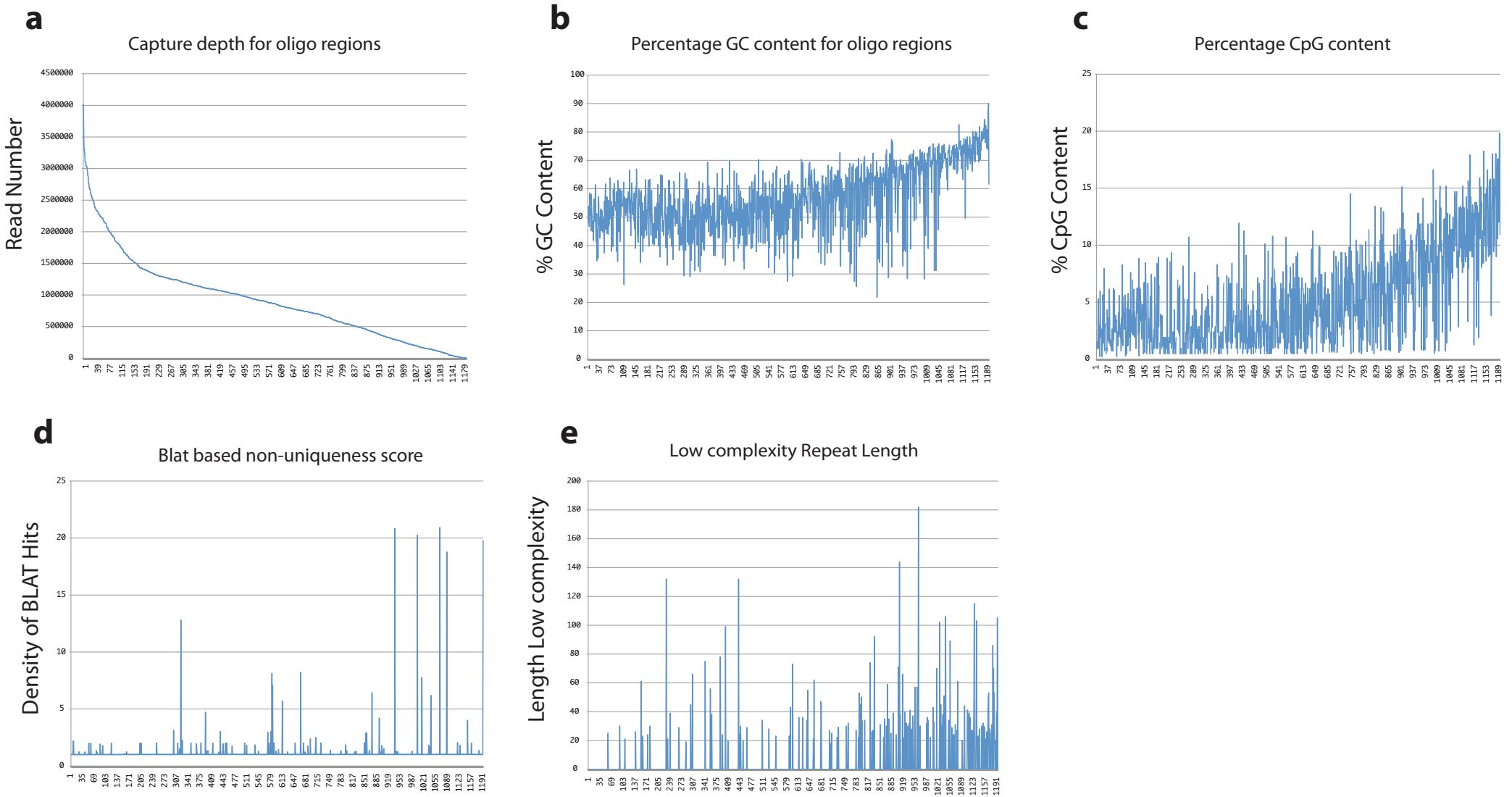
Panel a, Shows the positioning of biotinylated capture probes (red horizontal dash lines) to target specific genomic regions (light blue boxes) relative to the restriction enzyme recognition sites (vertical strokes) of a chromatin fragment containing a gene promoter (red arrow) which interacts with two distal elements (green and purple ovals).

Panel b, Shows an outline of the Capture-C assay. The interactions of the promoter and *cis* elements (as above) are stabilized by formaldehyde cross-linking. The cross-linked chromatin is then digested using a frequent cutting restriction enzyme (Dpn II) to generate free DNA ends capable of ligation. After dilution to minimise random ligation events the chromatin is ligated to form covalent links between the interaction fragments. After de-crosslinking of the chromatin and purification, the 3C ligated DNA exists as high-molecular weight concatemers in which the frequency of direct ligation of fragments to each other is dependant on their proximity in the nucleus. These concatemers are fragmented by sonication to ~300 bp to produce DNA suitable for oligo capture. After the ligation of Illumnia paired end adapters to these fragments the informative ends of the target restriction fragments are enriched by hybridisation to 120 bp biotinylated RNA SureSelect capture probes and streptavidin pull down. The captured DNA is amplified by 12x cycles of Illumina library amplification and paired-end sequenced on the HIseq Illumina platform.

Panel c,
(i) Shows the effectiveness of oligo capture technology for enriching specific sequences at targeted regions (blue boxes in Supplementary Figure 1a) from a 3C library. The sequence read enrichments (sorted from greatest to smallest enrichment) for all 1,191 targeted regions used in the current SureSelect design are shown as a blue line, calibrated on the left axis. The sorted background (control) sequence read enrichments (see online methods) are shown as a green line calibrated on the left axis. The degree of sequence enrichment, relative to the average background enrichment, is shown as a red line calibrated on the right axis.

(ii) Shows the effectiveness of oligo enrichment technology at enriching for unique distal interactions within a 3C library at the targeted regions in the genome (targeted regions in panel a). Unique distal interactions were identified after removal of all confounding interactions captured in each targeted region (see text). The number of unique interactions (sorted from the most to the fewest) detected from all 1,191 targeted regions is shown as a blue line. Unique interactions detected from the matched control regions (sorted from the most to the fewest) in the genome (see online methods) are shown as a green line.

Effect of sequence content on sequence capture efficiency



Supplementary figure 2

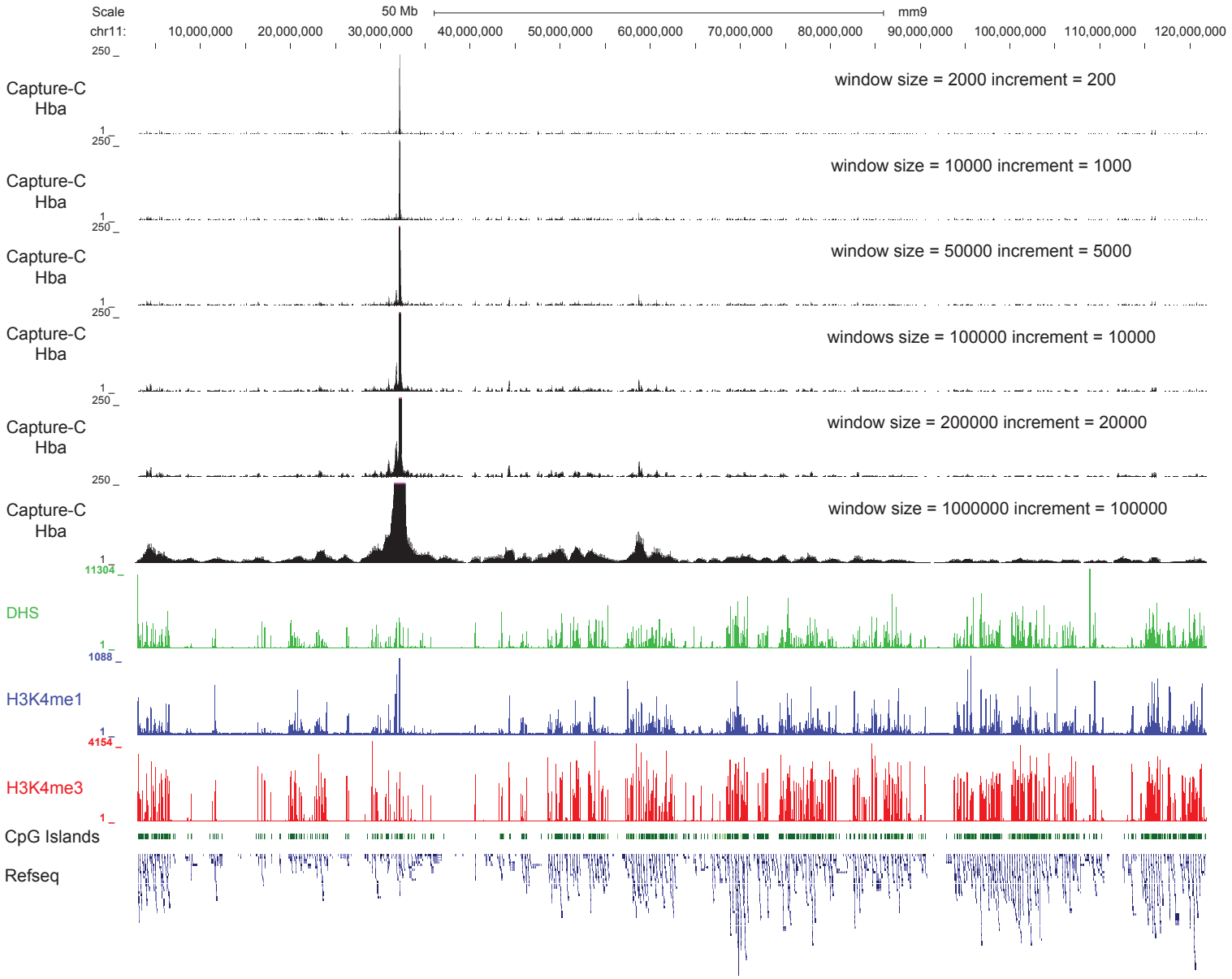
Supplementary figure 2.

Factors affecting efficiency of oligo capture.

Panel a, Shows a plot of the number of reads mapped to a region covered by oligo capture probes, sorted by the number of reads captured and sequenced.

Panels b-e Show plots of factors which may correlate with the efficiency of oligo capture in the same sort order as panel a (b, % GC content, c, % CpG content, d, repetitiveness based on per base BLAT scoring e, length of low complexity repeats).

Weak long-range interactions of the *Hba* gene promoters

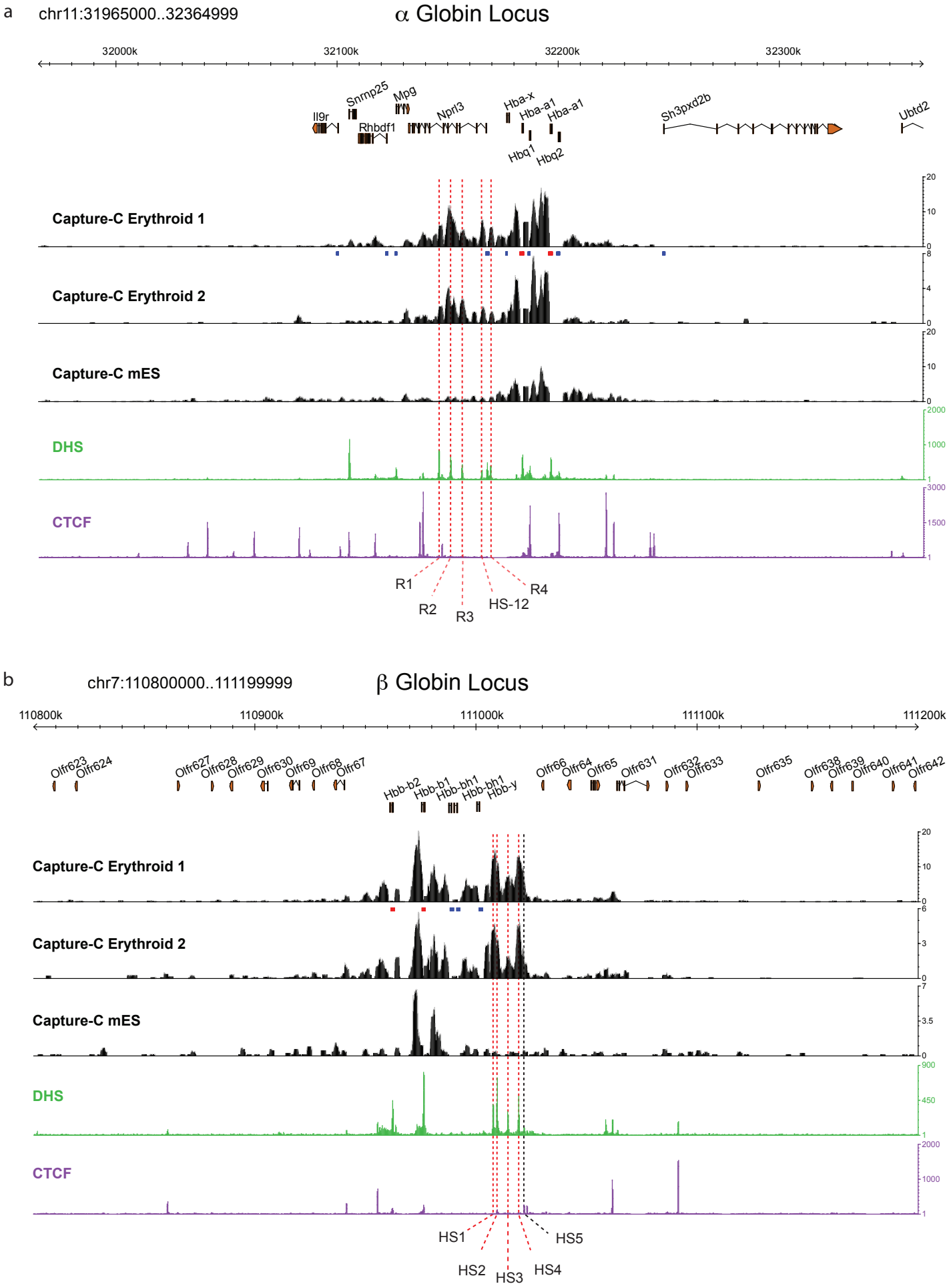


Supplementary figure 3

Supplementary figure 3.

Shows the effect of increasing window size on the interactions detected in the Capture-C data for the *Hba* gene promoters. The whole of mouse chromosome 11 is shown with the scale in kb at the top. In the top Capture–C panel the window size and moving increment are those used throughout the reported analysis (cumulative density in a window of 2 kb and increment 200 bp) in the subsequent 5 panels the window size and increment are increased stepwise to a window size of 1 Mb and increment of 100 kb. The window size and increment used are annotated on the right of each panel and the X-scale is fixed for ease of comparison. Below this active chromatin data (DHS in green, H3K4me1 in blue and H3K4me3 in red) are shown. The position of CpG island annotation (green boxes) for UCSC build mm9 and Refseq gene annotation are shown below this.

Reproducibility of the Capture-C assay



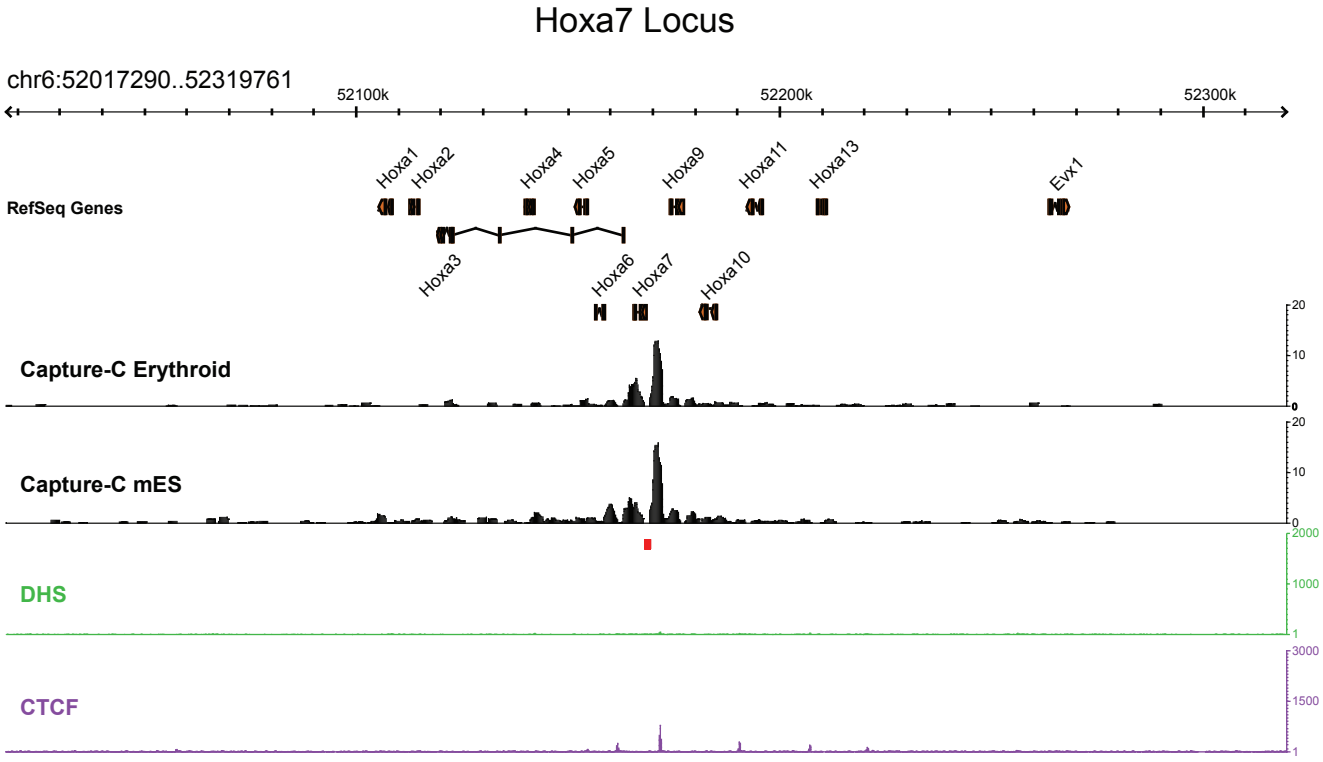
Supplementary figure 4

Supplementary figure 4.

Panel a, Genome scale and gene annotation are shown as described in Figure 1 over a 400 kb region of the α globin locus. The interaction profile of the *Hba* gene promoters in two erythroid biological replicates and a non-erythroid control (mouse ES cells) are plotted as previously described, relative to a DNase-seq (green) and CTCF (purple) ChIP-seq tracks from the same cell type. The position of the capture probes are shown as red boxes and the position of other captured promoters from which the data have been excluded are shown as blue boxes. The previously characterised regulatory elements are highlighted with red dashed lines and labelled at the bottom of the panel.

Panel b, Genome scale and gene annotation are shown over a 400 kb region of the β globin locus. The interaction profile of the *Hbb* gene promoters in two erythroid biological replicates and a non-erythroid control (mouse ES cells) are plotted as previously described, relative to a Dnase-seq (green) and CTCF (purple) ChIP-seq tracks from them the same cell type. The position of the capture probes are shown as red boxes and the position of other captured promoters from which the data have been exclude are shown as blue boxes. The previously characterised regulatory elements are highlighted with red dashed lines and labelled at the bottom of the panel.

Minimal interactions of the inactive *Hoxa7* gene promoter



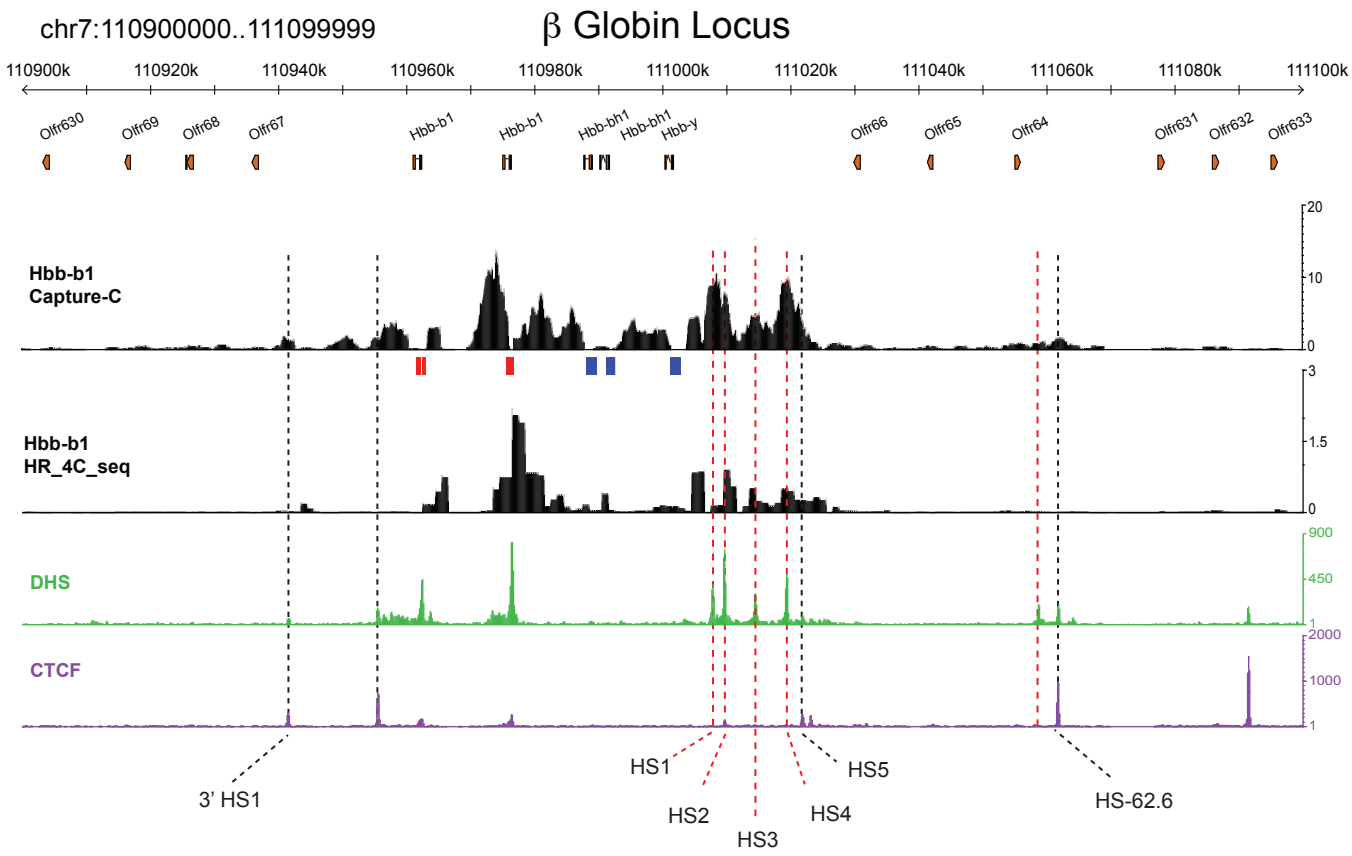
Supplementary figure 5

Supplementary figure 5.

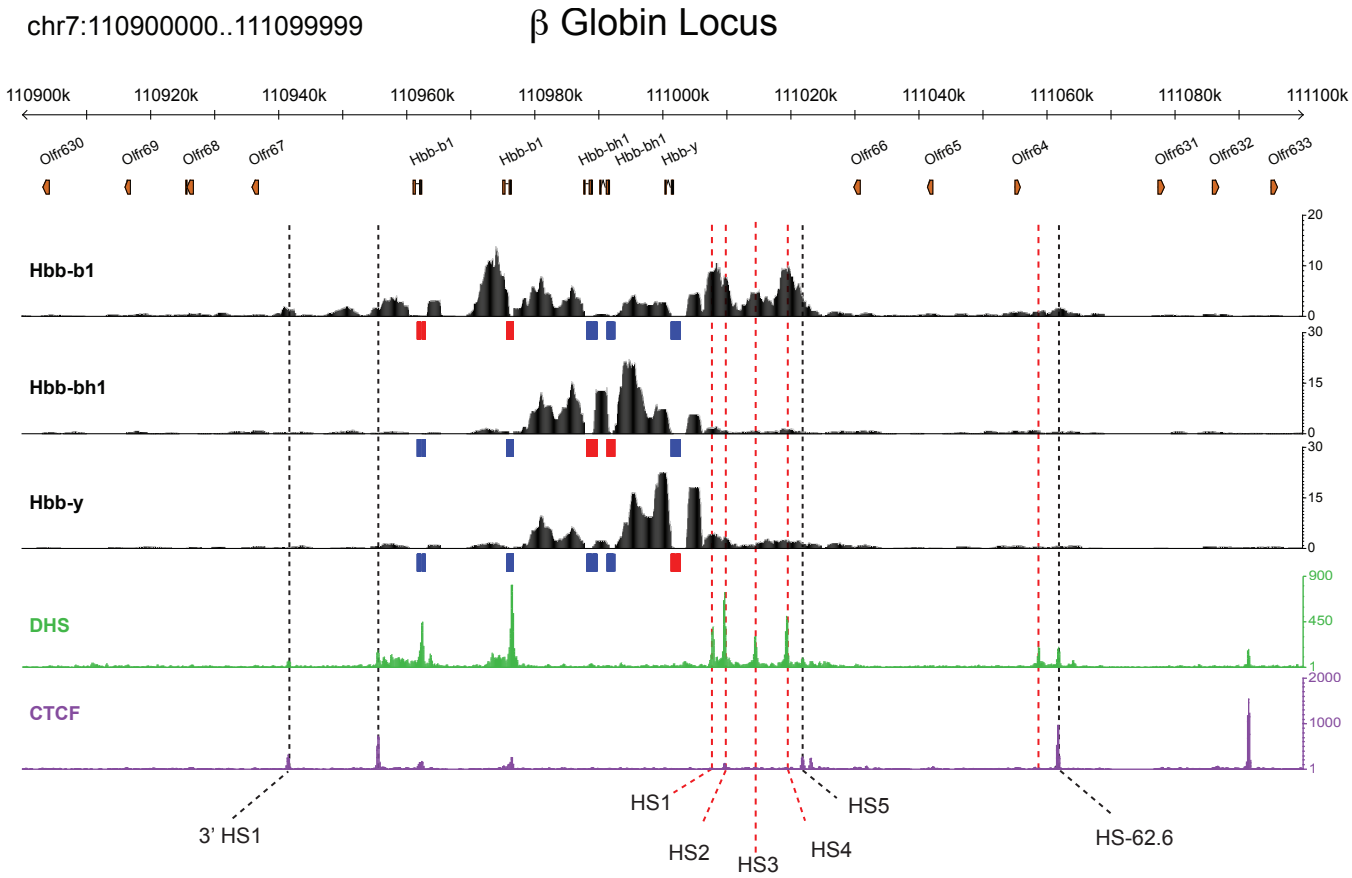
Genome scale and gene annotation are shown as described in Figure 1 over a 300 kb region of the *Hoxa7* locus. The interaction profiles of the *Hoxa7* gene promoter which is silent in both erythroid and in mouse ES cells are plotted as previously described, relative to DNase-seq (green) and CTCF (purple) ChIP-seq tracks from erythroid cells. The positions of the capture probes are shown as a red box.

The comparison of Capture-C and high-resolution 4C-seq

a



b



Supplementary figure 6

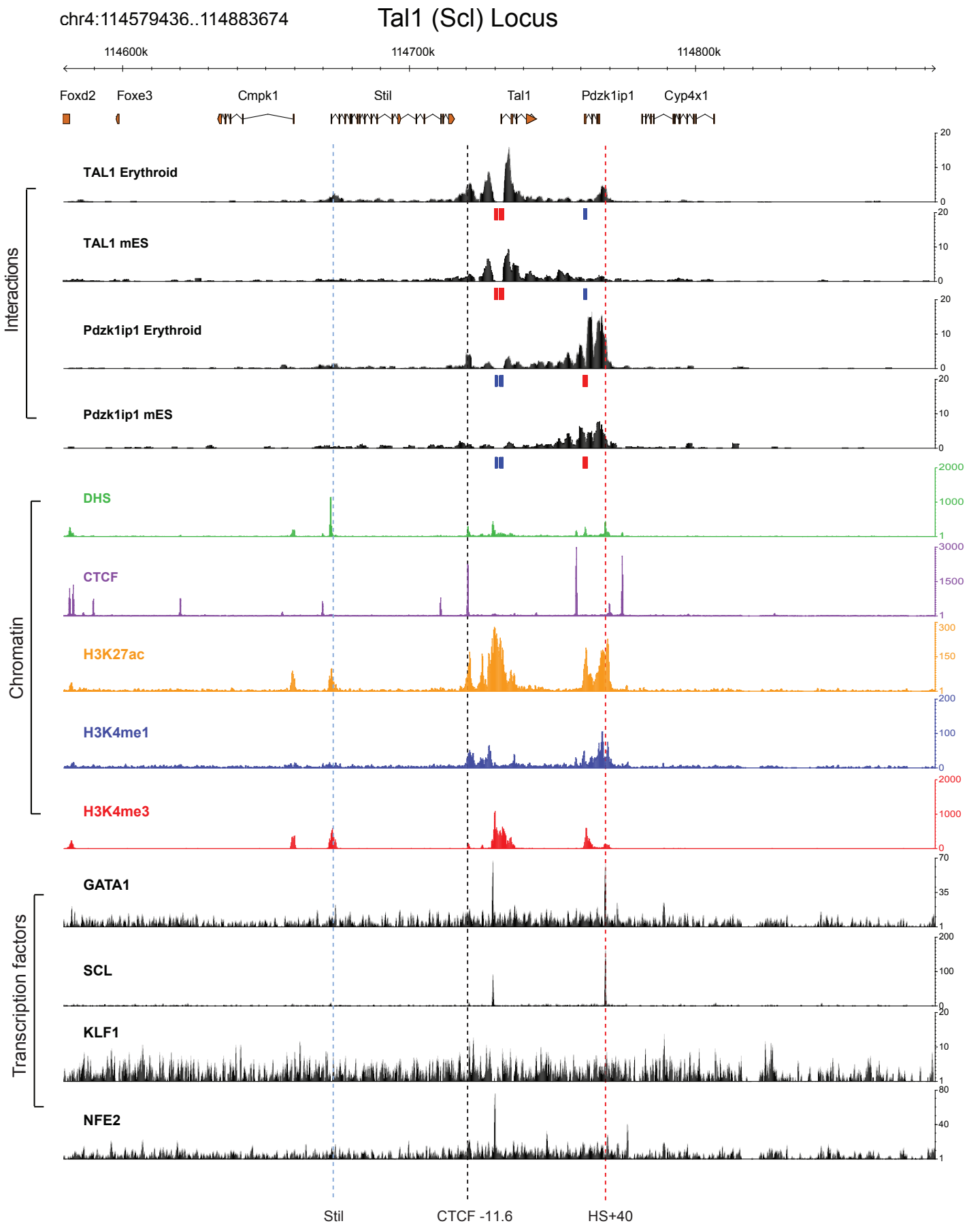
Supplementary figure 6.

Panel a, shows the comparison between conventional high-resolution 4C data (HR_4C_seq was plotted from the analysed interaction data (median values within a 2kb window, centred on the 1kb genomic bin) supplied as part of the GEO submission from van de Werken et al[1]) from the *Hbb* gene promoters in mouse erythroid cells. The chromatin data, gene annotation and capture probes are as described previously over a 200 kb region of the locus. The previously characterised regulatory elements are highlighted with red dashed lines and labelled at the bottom of the panel. Black dashed lines highlight interactions with regions bound by the CTCF protein

Panel b, Shows a comparison of the interactions of the active adult β globin gene promoters (*Hbb*) and the inactive fetal *(Hbb-bh1*) and inactive embryonic (*Hbb-y*) gene promoters. In each track the position of the capture probes used are shown as red boxes and the position of other captured regions from which the data have been excluded in that track are shown as blue boxes. The previously characterised regulatory elements are highlighted with red dashed lines and labelled at the bottom of the panel. Black dashed lines highlight interactions with regions bound by the CTCF protein.

1. van de Werken, H.J. *et al.* Robust 4C-seq data analysis to screen for regulatory DNA interactions. *Nat Methods* **9**, 969-72 (2012).

Interaction of the *Tal1* and *Pdzk1ip1* gene promoters with their chromatin landscape

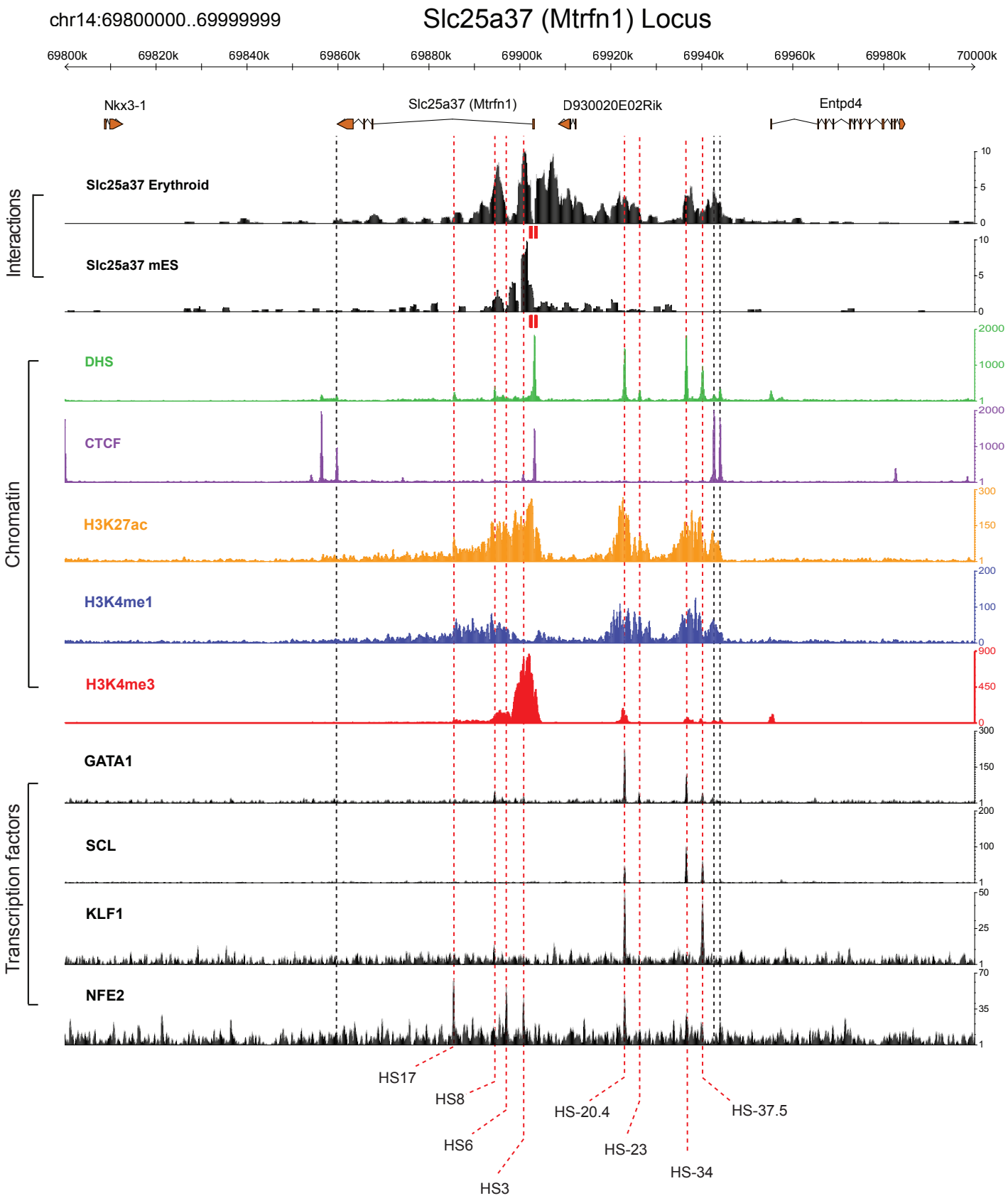


Supplementary figure 7

Supplementary figure 7.

Shows the comparison between the interactions of the promoter of the *Tal1* gene which, encodes an erythroid transcription factor, and the neighbouring *Pdzk1ip1* (*Map17*) gene in both erythroid and mES cells. In all tracks the position of the capture probes used are shown as red boxes and the position of other captured regions from which the data have been excluded in that track are shown as blue boxes. The previously characterised regulatory element is highlighted with a red dashed line and labelled at the bottom of the panel. Black dashed lines highlight interactions with regions bound by the CTCF protein. Interactions with the neighbouring promoter of the *Stil* gene are highlighted as a blue dashed line. In addition to DHS (green) and CTCF (Purple) ChIP-seq data for further chromatin marks or chromatin associated proteins (H3K27ac in orange, H3K4me1 in blue and H3K4me3 in red) and key transcription factors (GATA1, SCL, KLF1 and NFE2 all in black) are shown relative to the interaction profile

Interaction of the *Slc25a37* gene promoter with its chromatin landscape



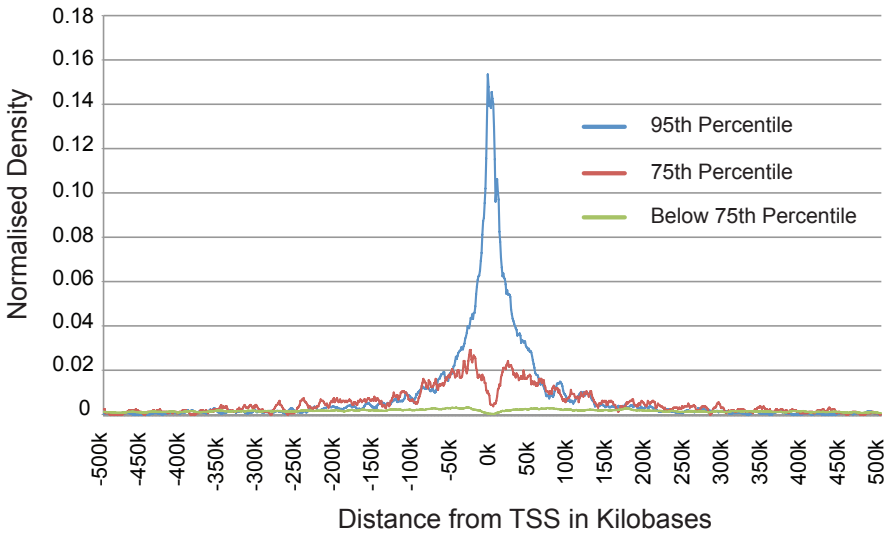
Supplementary figure 8

Supplementary figure 8.

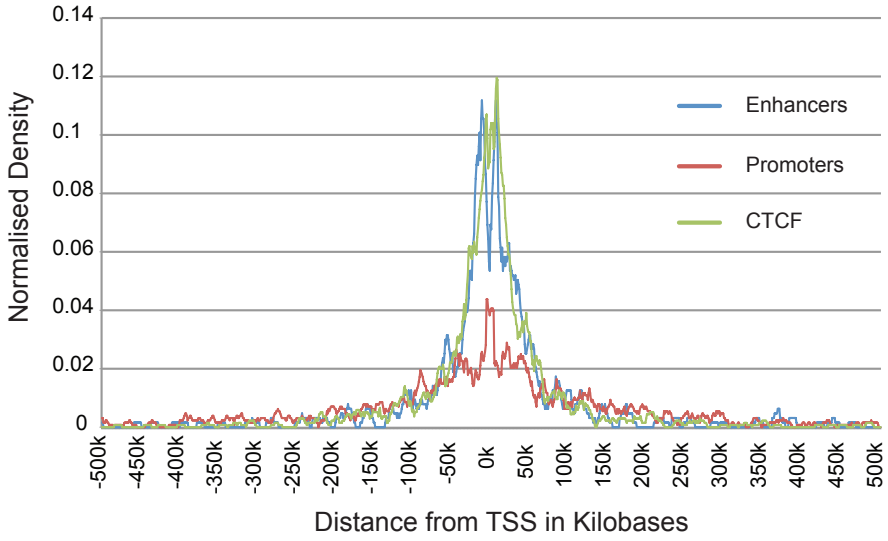
Shows the interaction profile of the promoter of the essential erythroid *Slc25a27* gene (*Mtrfn1*) in both erythroid and mES cells. The positions of the capture probes used are shown as red boxes. Interactions with DHS bound by erythroid transcription factors are highlighted with red dashed lines and labelled depending on their distance in kb from the *Slc25a27* promoter (- in the transcriptional upstream direction and + in the downstream direction). Black dashed lines highlight interactions with regions bound by the CTCF protein. In addition to DHS (green) and CTCF (Purple) further ChIP-seq data for chromatin marks (H3K27ac in orange, H3K4me1 in blue and H3K4me3 in red) and key transcription factors (GATA1, SCL, KLF1 and NFE2 all in black) are shown relative to the interaction profile.

The general distribution of interacting *cis* elements around gene promoters

a. Distribution of interacting *cis* elements around TSS



b. Distribution of interacting *cis* elements by class



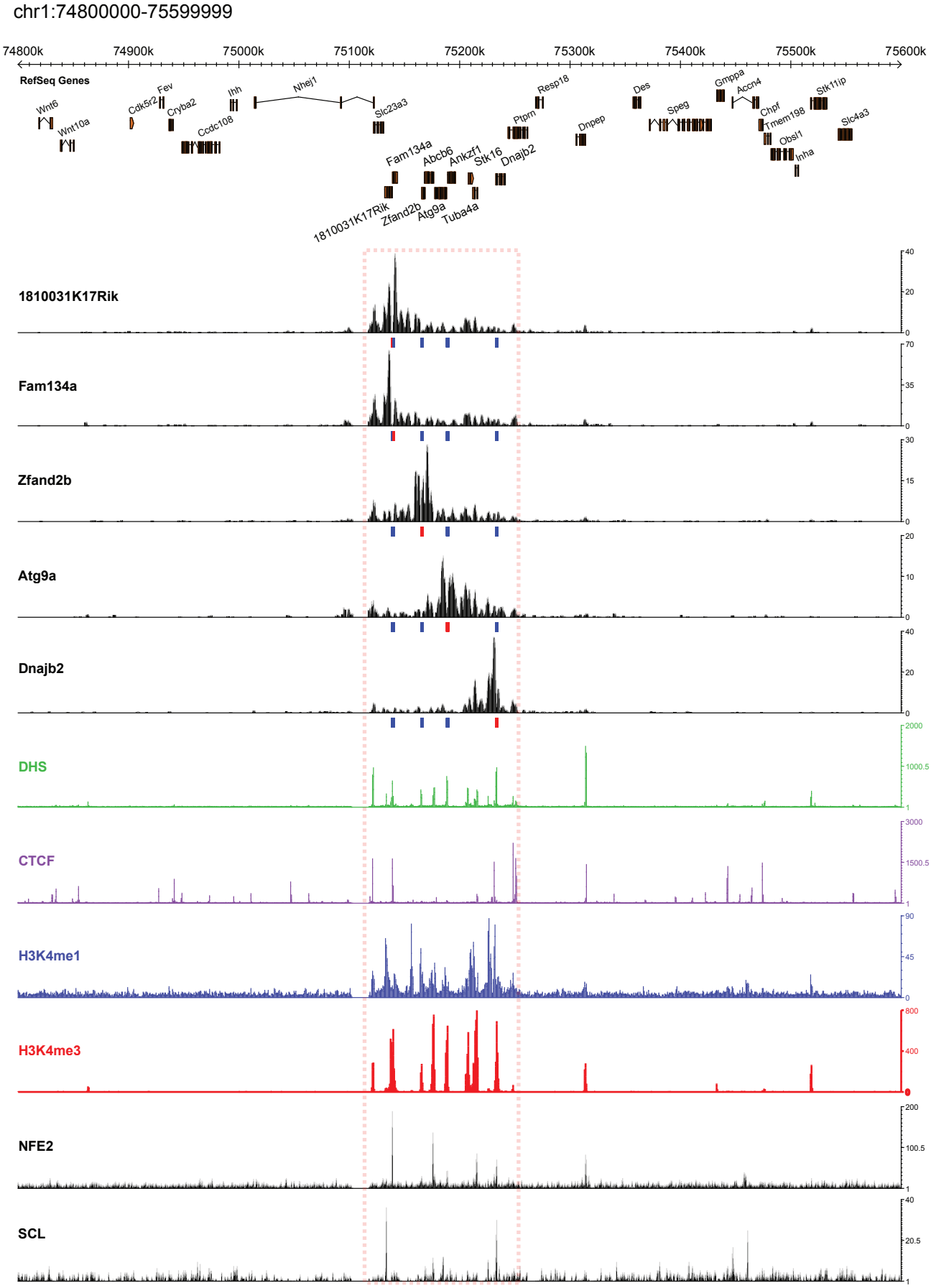
Supplementary figure 9

Supplementary figure 9.

Panel a, Shows the cumulative distribution of the three significance classes of *cis* element interactions centred on the TSS irrespective of transcriptional orientation or classification of element. The distribution was generated using a sliding window of 10 kb with an increment of 1 kb. The X-axis plots the distance upstream and downstream of the TSSs in 50 kb increments. The Y-axis shows the cumulative frequency normalized for the number of features in each class.

Panel b, Shows the cumulative distribution of the combined 95th and 75th significance classes of *cis* element interactions, centred on the TSS. Interacting *cis* elements are now sub-classified and split into three classes of element, CTCF (green), enhancer (blue) and promoter (red). The distribution was generated as described in Supplementary figure 9a.

A 150kb domain of interaction shared by multiple genes

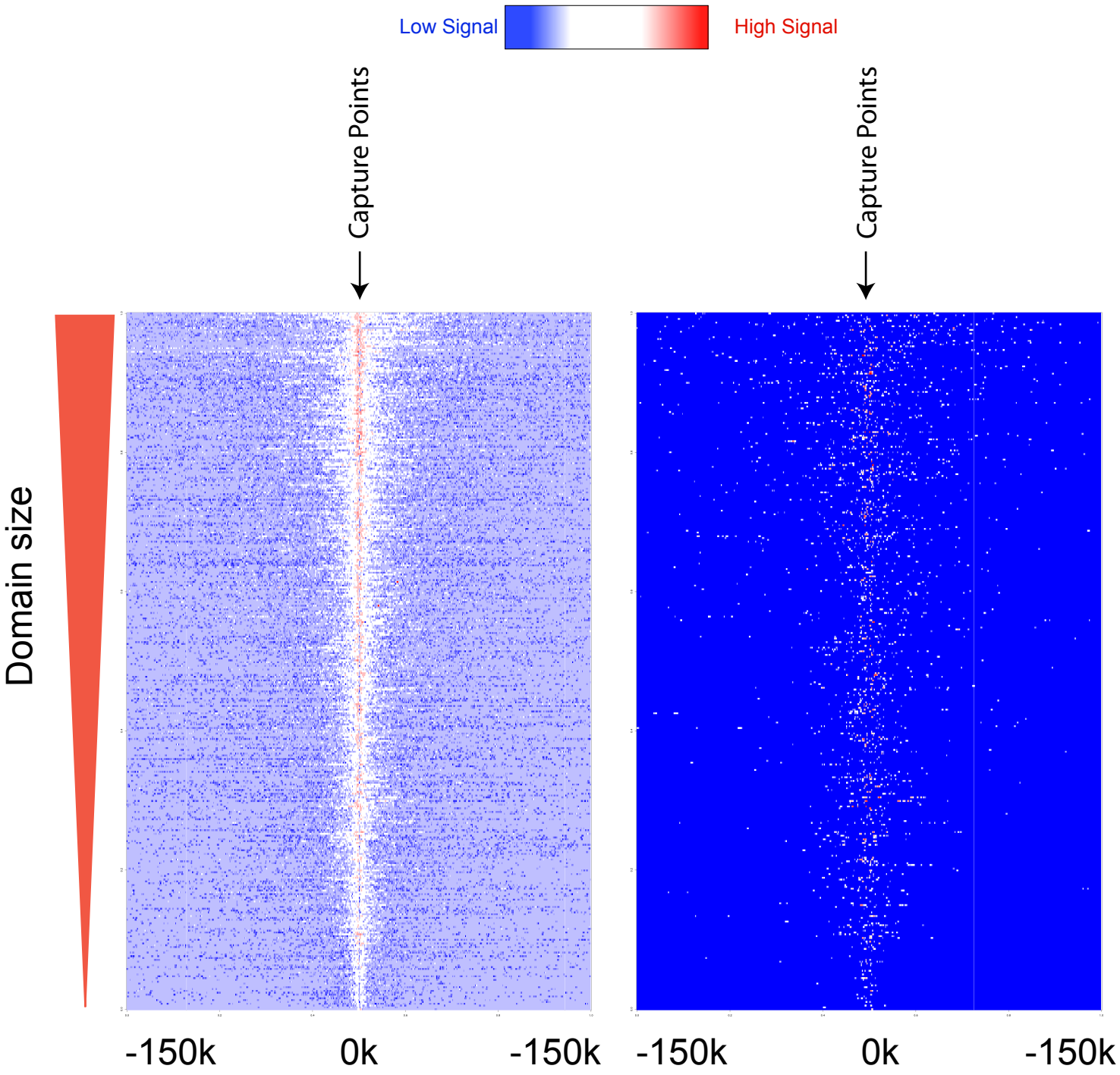


Supplementary figure 10

Supplementary figure 10.

Panel a, Shows a heat map of the distribution of Capture-C normalised interaction signal in 1 kb bins over a 300 kb window centred on the capture region (indicated by arrows). On the y-axis 404 captured regions sorted from top to bottom by size of interaction domain (number of positive bins) are shown. Signal strength is shown from blue, through white to red with increasing signal. The density plots were generated as using the custom PERL script DomainPlots.pl and then displayed in the R package using the GPlots library.

Panel b, Shows a heat map sorted as in a, of the signal associated with the highest statistically significant class of interacting elements (95^{th} percentile, see online methods). Signal strength colouring and axis are as in panel a, and allows the visualization of the interacting elements within the larger interaction domains in panel a.



a Normalised interaction density

The red signal shows the high signal proximal to the capture points while the white signal shows the extent of the domains of interaction

b Interactions with *cis*-elements

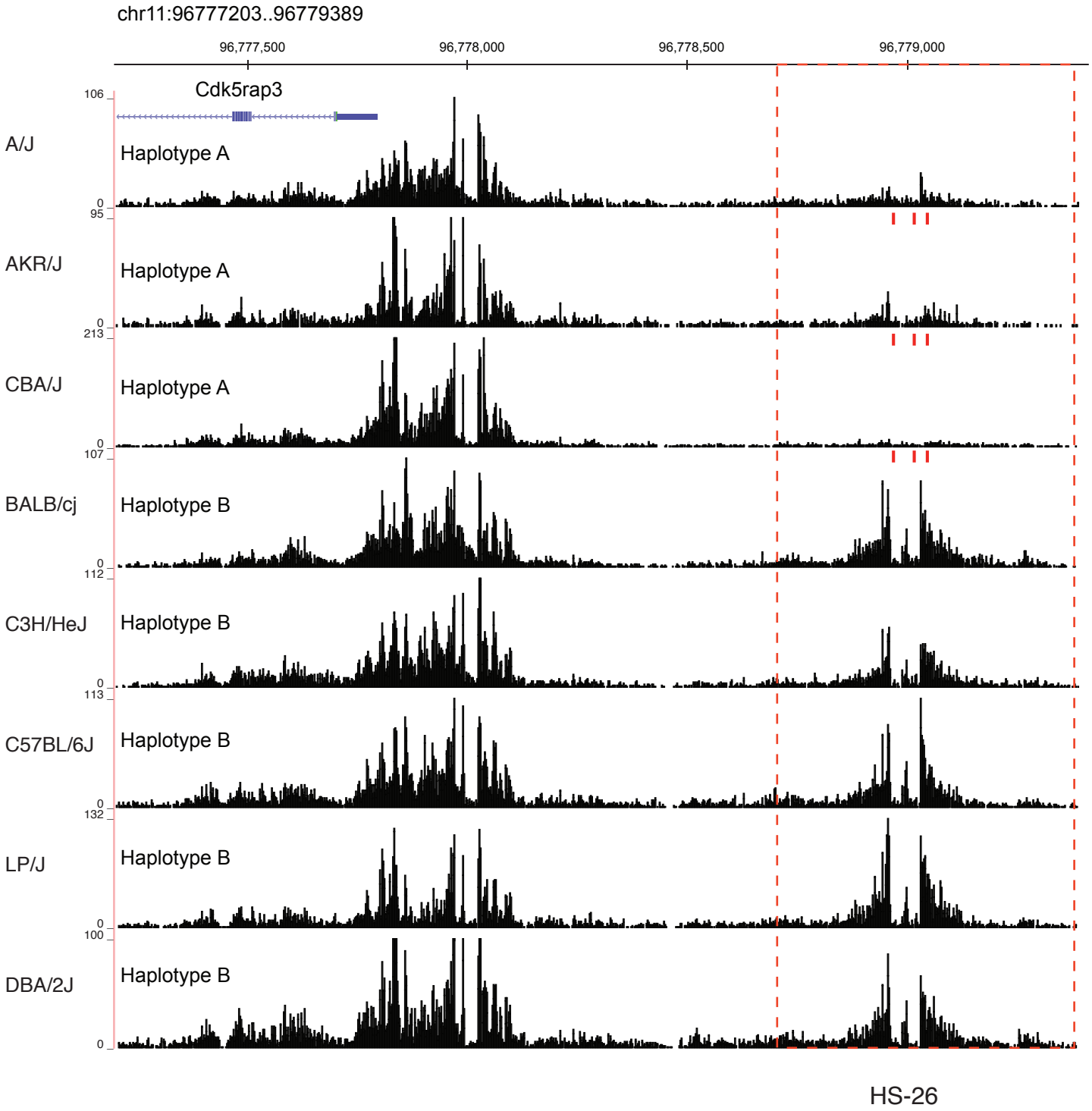
Shows the position of annotated *cis* elements (over the same genomic extent and in the same sort order as in panel a) colour coded by the degree of interaction. Comparision with panel a shows highly interacting elements (red) are often embedded in a general domain of interaction

Supplementary figure 11

Supplementary figure 11.

Shows the comparison of the interactions of a cluster of 5 gene promoters on chromosome 1qC3. In each track the position of the capture probes used are shown as red boxes and the position of other captured regions from which the data have been excluded in that track are shown as blue boxes. In addition to DHS (green) and CTCF (Purple) further ChIP-seq data for chromatin marks (H3K27ac in orange, H3K4me1 in blue and H3K4me3 in red) and key transcription factors (GATA1, SCL, KLF1 and NFE2 all in black) are shown relative to the interaction profile. A red dashed box highlights the extent of the shared interaction compartment.

SNP dependant loss of a *cis* element in 3 mouse strains



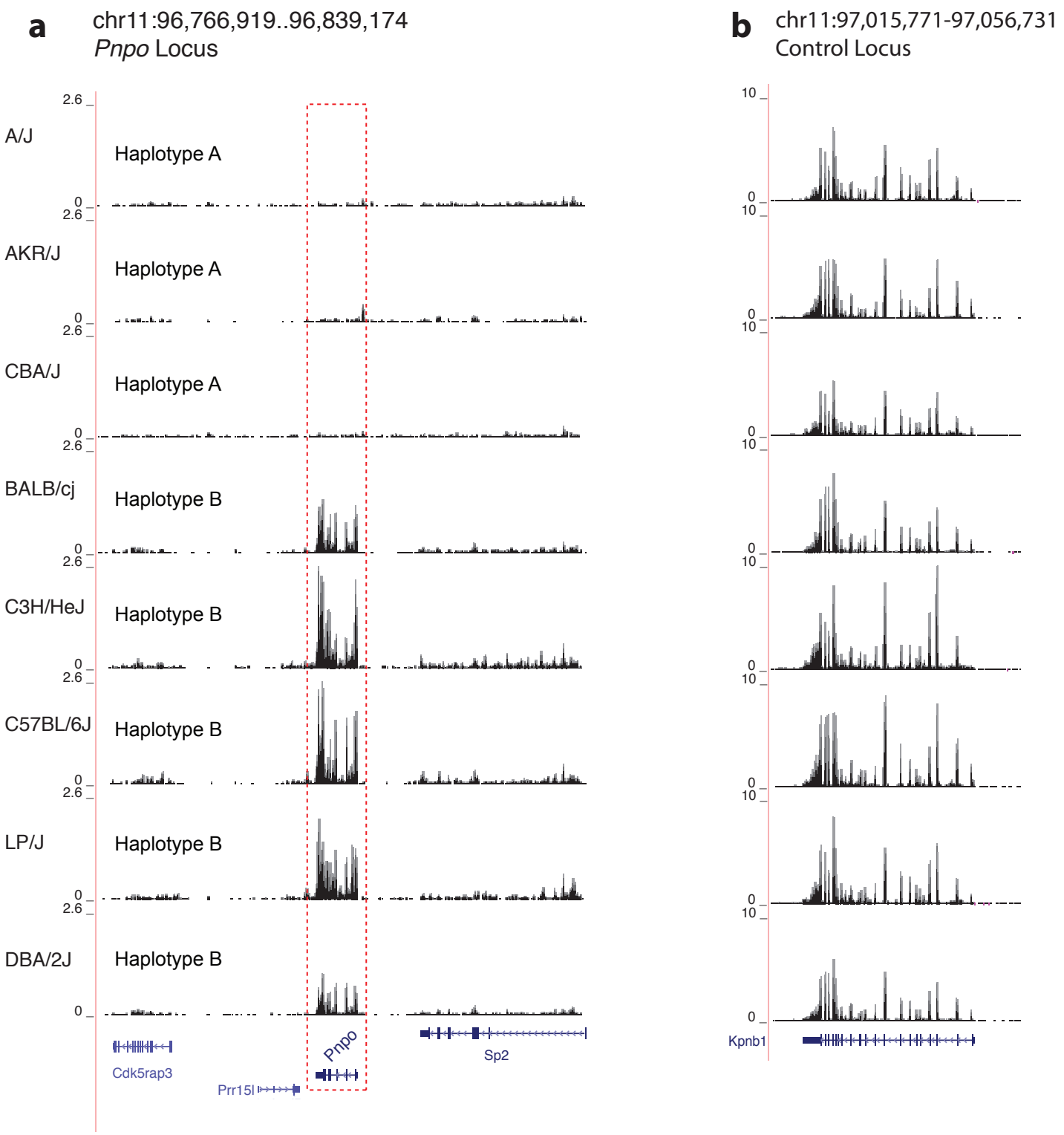
Supplementary figure 12

Supplementary figure 12.

Shows a ~ 2kb region around the promoter of the *Cdk5rap3* gene with position of RefSeq annotated genes in blue. The HS-26 enhancer region indicated in Figure 4a is within the red hatched box. The DNase1 footprints derived from high read depth DNA-seq data (Hosseini et al[1]) is shown for 8 mouse strains from either haplotype A or haplotype B. The position of the 3 SNPs found in haplotype A (A/J, AKR/J and CBA/J mouse strains) are depicted as red ticks.

1. Hosseini, M. *et al.* Causes and consequences of chromatin variation between inbred mice. *PLoS Genet* **9**, e1003570 (2013).

Effect of enhancer loss on “nascent” transcription of Capture-C linked gene promoter (PolyA- RNA-seq)



Supplementary figure 13

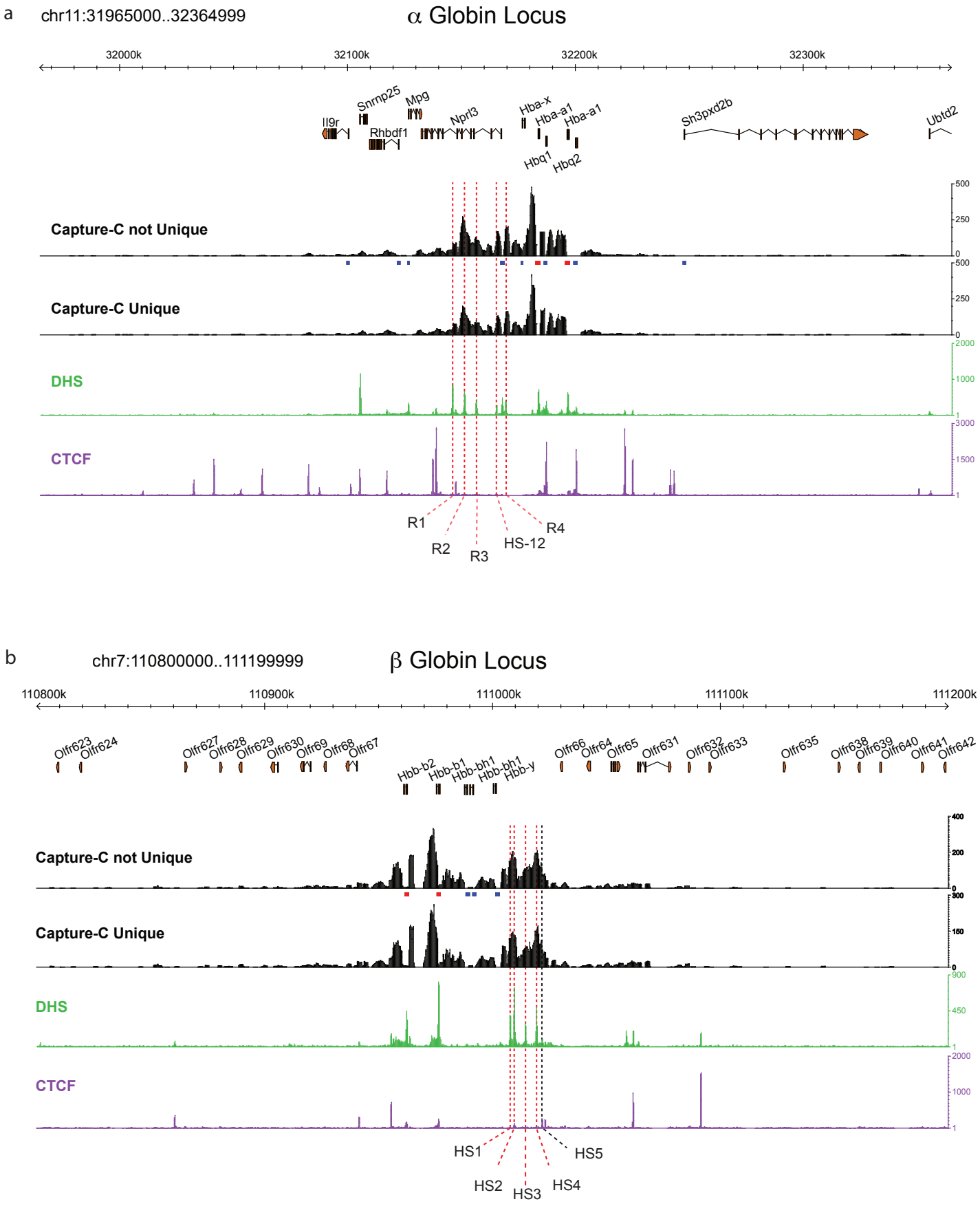
Supplementary figure 13.

Panel a, show the read normalised signal (normalised per million reads aligned) for erythroid PolyA- RNA-seq (Hosseini et al[1]) over a 722kb region surrounding the *Pnpo* locus in 8 mouse strains in the same order as Figure 4c and Supplementary figure 12. The position of RefSeq annotated genes are shown at the bottom and transcription associated with the *Pnpo* gene is shown within the red-hatched box.

Panel b, Shows the normalised transcription of the neighbouring well-expressed *Kpnb1* gene as a control for RNA-seq comparability.

1. Hosseini, M. *et al.* Causes and consequences of chromatin variation between inbred mice. *PLoS Genet* **9**, e1003570 (2013).

Capture-C profiles with and without PCR duplicates



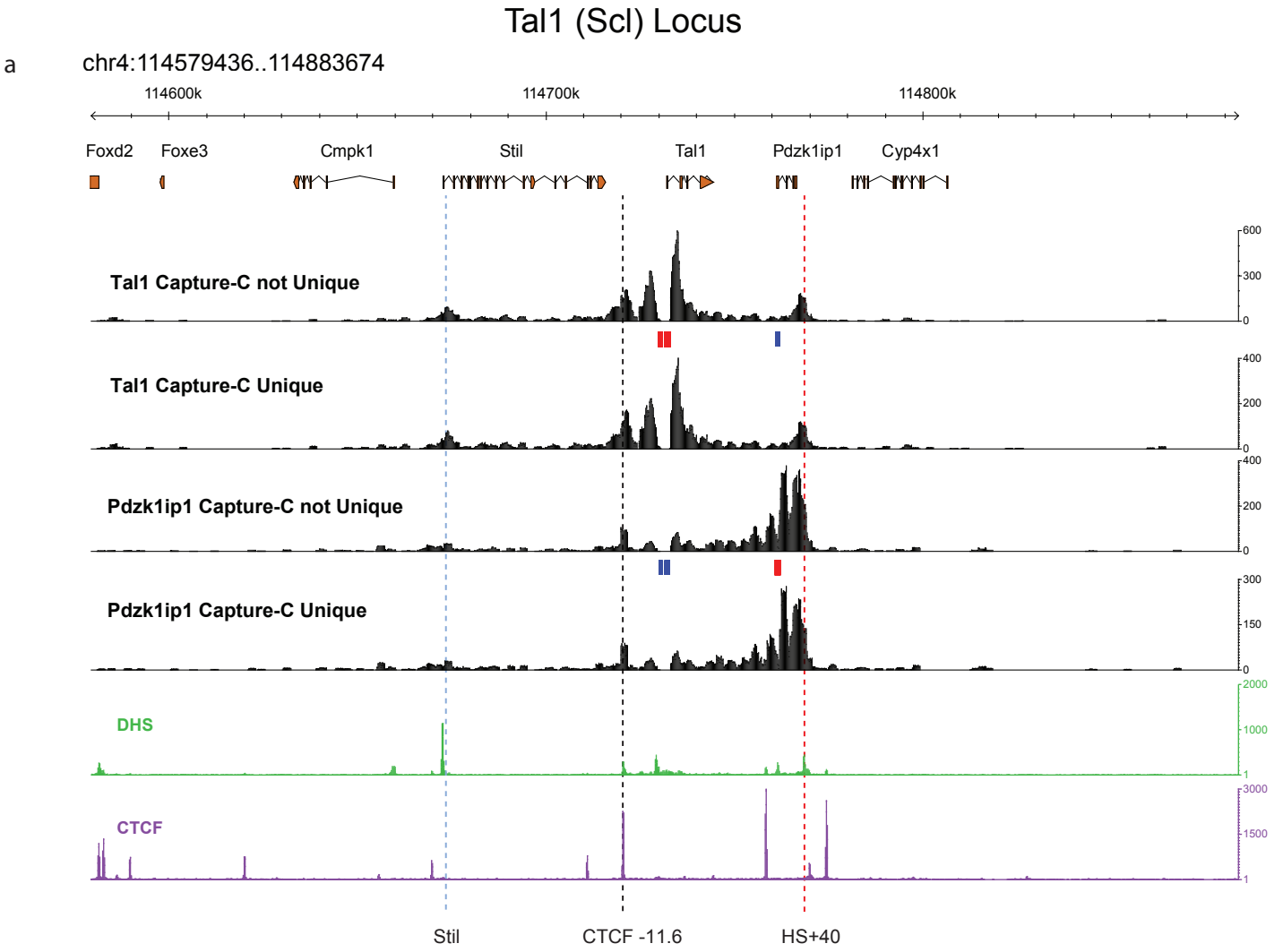
Supplementary figure 14

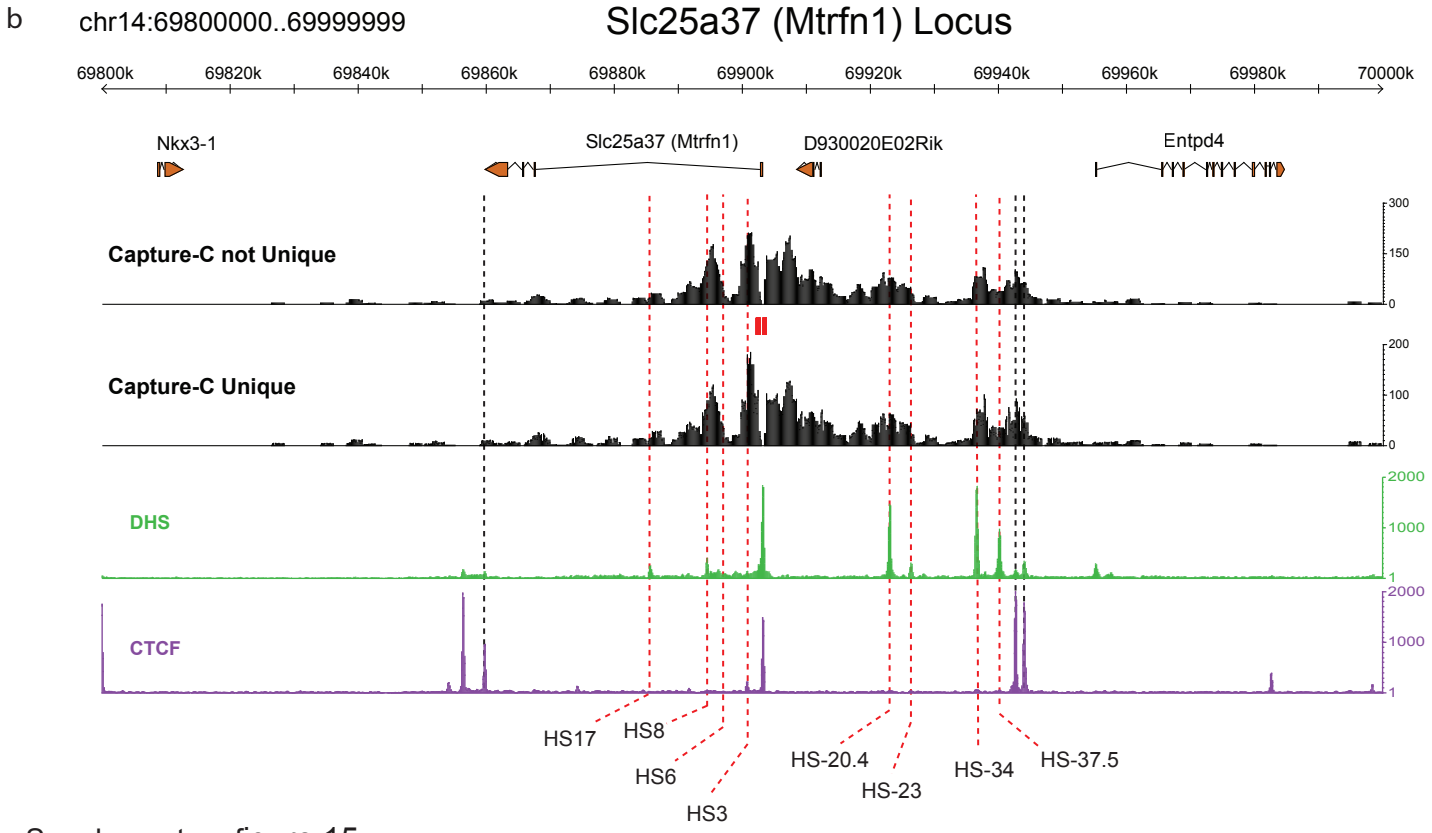
Supplementary figure 14.

Panel a, Shows the comparison between the interaction profiles of the *Hba* gene promoters in mouse erythroid cells, with interactions of precisely the same mapped genomic coordinates collapsed to a single representative interaction with PCR duplicates removed (Capture-C unique) or not removed (Capture-C not unique). The chromatin data, gene annotation and capture probes are as described previously over a 400 kb region of the locus. The previously characterised regulatory elements are highlighted with red dashed lines and labelled at the bottom of the panel. Black dashed lines highlight interactions with regions bound by the CTCF protein

Panel b, Shows the comparison between the interaction profiles of the *Hbb* gene promoters in mouse erythroid cells, with interactions of precisely the same mapped genomic coordinates collapsed to a single representative interaction with PCR duplicates removed (Capture-C unique) or not removed (Capture-C not unique). The chromatin data, gene annotation and capture probes are as described previously over a 400 kb region of the locus. The previously characterised regulatory elements are highlighted with red dashed lines and labelled at the bottom of the panel. Black dashed lines highlight interactions with regions bound by the CTCF protein

Capture-C profiles with and without PCR duplicates





Supplementary figure 15

Supplementary figure 15.

Panel a, Shows the comparison between the interaction profiles of the *Tal1* and *Pdzk1ip1* gene promoters in mouse erythroid cells, with interactions of precisely the same mapped genomic coordinates collapsed to a single representative interaction with PCR duplicates removed (Capture-C unique) or not removed (Capture-C not unique). The chromatin data, gene annotation and capture probes are as described previously over a 300 kb region of the locus. The previously characterised regulatory element is highlighted with a red dashed line and labelled at the bottom of the panel. Black dashed lines highlight interactions with a region bound by the CTCF protein. An interaction with the promoter of the neighbouring *Stil* promoter is highlighted by a blue dashed line.

Panel b, Shows the comparison between the interaction profiles of the *Slc25a37* (*Mtrfn1*) gene promoter in mouse erythroid cells, with interactions of precisely the same mapped genomic coordinates collapsed to a single representative interaction with PCR duplicates removed (Capture-C unique) or not removed (Capture-C not unique). The chromatin data, gene annotation and capture probes are as described previously over a 200 kb region of the locus. Interactions with DHS bound by erythroid transcription factors are highlighted with red dashed lines and labelled depending on their distance in kb from the *Slc25a27* promoter (- in the transcriptional upstream direction and + in the downstream direction). Black dashed lines highlight interactions with regions bound by the CTCF protein.

Supplementary note.

Chromatin conformation capture methods.

Capture-C provides a cost effective, high-resolution method to simultaneously identify and quantify regulatory interactions of multiple, selected regions of the genome. No other chromosome conformation capture method fulfils all of these requirements. The key features of Capture C are:

1. Regions of interest may be selected independently of chromatin structure, protein binding profiles or function, which inevitably vary across different regions of the genome.
2. The assay is of sufficient resolution to analyse small (100's bp) *cis*-acting elements even when clustered within the genome.
3. Despite the complexity of the ligation events to be analysed, Capture C generates sufficient signal to robustly score the known interactions in the well-studied globin test loci.
4. Quantification of the interactions is as direct as possible, largely avoiding potential sources of skewing of the data, such as PCR bias, which exist in other protocols.
5. It is possible to highly multiplex analyses

Of the current 3C methods only three protocols can analyse the interactions of multiple genomic regions in a single assay.

1. Hi-C enriches for all genomic ligation junctions within a 3C library. Due to the massively complex nature of the resultant library (all fragments against all fragments), the level of signal for any given point in the genome is very low. This demands the use of infrequently cutting restriction enzymes and large windows of analysis (~100's kb) to pool low-resolution ligation events and thereby achieve informative signals. To date, Hi-C has only been used for analysis of the large-scale (Mb) domain structure of the mammalian genome.

2. CHIA-PET uses Chromatin Immunoprecipitation (ChIP) to enrich for the interactions between regions bound by specific proteins or marked by specific chromatin modifications. After initial ChIP enrichment, biotinylated linkers are used to join interacting fragments that have been captured on beads. The initial ChIP only gives a low level of enrichment. Furthermore, all regions in the genome marked by a particular protein or chromatin modification (usually many 1000s) must be analysed. Together these factors give relatively low signal

density and poor resolution. In addition, there are many opportunities to artificially skew the signals in this assay[1].

3. Chromatin Conformation Capture Carbon Copy (5C) uses the ligation junctions within a 3C library as a hybridisation template for oligonucleotide probes designed to one end of each restriction fragment. When hybridised in close proximity on the DNA template, the probes can then be ligated. PCR amplification using adaptor sequences at the ends of the probes produces a "carbon copy" of the junctions within the original 3C library. The carbon copies can be sequenced and counted. Clearly, this method only detects fragments for which a probe has been designed but the process can be multiplexed by pooling probes from many different loci. In practice, this method is limited by the number of probes that can be made, purchased and analysed in any experiment. Hence, to date 5C experiments in mammalian cells have been limited to low resolution 3C libraries generated with 6bp cutter restriction enzymes. Other shortcomings are that interactions can only be detected at one end of a fragment and the method assumes that the orientation of ligation is random. Whatever, 5C only detects half of the potential ligation products. Quantitation of the ligation events may be biased since they are measured indirectly via the hybridisation of small probes of variable sequence.

Of the current 3C methods, only 4C has been adapted to work with high-resolution 3C libraries generated using 4bp cutter restriction enzymes.

Circularised chromatin conformation capture (4C) depends on the generation of circular DNA ligation products from a 3C library. The interactions of a chosen region can be determined using inverse PCR from outward pointing primers (within the target fragment) to amplify ligated fragments within each circular DNA molecule. Selecting specific target fragments and using PCR allows 4C to generate high signal levels and so can be adapted to the use of frequent cutting enzymes and high-resolution analysis[2]. Any fragment can be chosen for analysis in a 4C experiment as this method only depends on the ability to design suitable PCR primers. However, the circular amplification step contains inherent biases (in particular the length and GC content of the captured fragment), which skews the derived distribution, and requires extensive statistical modelling to correct for this bias. The 4C method cannot be multiplexed, although increased throughput can be achieved by multiplexing the sequencing step of the libraries.

1. de Wit, E. & de Laat, W. A decade of 3C technologies: insights into nuclear organization. *Genes Dev* **26**, 11-24 (2012).
2. van de Werken, H.J. *et al.* Robust 4C-seq data analysis to screen for regulatory DNA interactions. *Nat Methods* **9**, 969-72 (2012).

Supplementary Table 1

Capture Regions MM9 Genome Build

Cxcr4	chr1:130487551-130490371	b	yes
Tmcc2	chr1:134286762-134289689	b	yes
Atp2b4_1	chr1:135649423-135650884	b	yes
Atp2b4_2	chr1:135696735-135698467	b	yes
Btg2	chr1:135974255-135977270	b	yes
Adipor1	chr1:136311576-136312742	b	no
Rnasel	chr1:155595460-155597202	b	yes
Prrx1	chr1:165243179-165244864	m	yes
Ube2w	chr1:16608340-16610364	m	yes
Dcaf6	chr1:167389625-167391440	b	no
Creg1	chr1:167692812-167695788	b	yes
Mgst3	chr1:169322952-169325698	b	yes
Ppox	chr1:173210152-173213434	b	yes
Spna1	chr1:176100245-176104824	b	yes
Arid5a	chr1:36363177-36365656	m	yes
Trak2	chr1:59027148-59031794	b	yes
1810031K17Rik	chr1:75137189-75139542	b	yes
Fam134a	chr1:75138735-75140516	b	yes
Zfand2b	chr1:75163989-75166340	b	yes
Abcb6	chr1:75174963-75177940	b	yes
Atg9a	chr1:75187652-75189458	b	yes
Tuba4a	chr1:75215328-75217511	b	yes
Dnajb2	chr1:75232055-75234432	b	yes
Asb1	chr1:93435222-93438243	b	yes
Fbxo30	chr10:10998252-11002511	b	yes
Ddit3	chr10:126725417-126728935	m	yes
Cited2	chr10:17442499-17445193	m	yes
Slc16a10	chr10:39860124-39862842	b	yes
Cep57l1	chr10:41528703-41531103	b	yes
Foxo3	chr10:41995489-41997743	b	yes
Cd24a	chr10:43297922-43299924	b	yes
Hace1	chr10:45296323-45300451	b	yes
Serinc1	chr10:57249298-57255492	b	yes
1110038D17Rik	chr10:74978613-74981002	b	yes
Trappc10	chr10:77705596-77709440	b	yes
Bsg	chr10:79165603-79168673	b	yes
Mknk2	chr10:80137829-80140740	b	yes
Oaz1	chr10:80287997-80290284	b	yes
Sppl2b	chr10:80317413-80319362	b	no
Fzr1	chr10:80840029-80842618	b	yes
Tcp11l2	chr10:84037989-84040252	b	yes
Fbxo7	chr10:85483769-85488597	b	yes
Nr1h4_1	chr10:88968283-88970530	m	yes
Nr1h4_2	chr10:88994894-88997607	m	yes
Metap2	chr10:93353152-93354729	b	yes
Atp6v0a1	chr11:100869798-100871610	b	yes
Psme3	chr11:101174929-101178285	b	yes
Mpp2	chr11:101948899-101950826	b	yes

Slc4a1	chr11:102225355-102228282	b	yes
Slc25a39	chr11:102267448-102269966	b	yes
Kif18b	chr11:102785642-102787295	b	yes
Wipi1	chr11:109470360-109474606	b	yes
Ube2o	chr11:116440942-116445142	b	yes
Tmc6	chr11:117641125-117644894	b	yes
Birc5	chr11:117700802-117705086	b	no
Mafg	chr11:120493694-120495948	b	yes
Narf	chr11:121096579-121099741	b	yes
Fn3krp	chr11:121281661-121283777	b	yes
Fn3k	chr11:121295226-121298227	b	yes
Meis1	chr11:18918002-18921127	b	yes
Sertad2	chr11:20442022-20444356	b	yes
Bcl11a	chr11:23976168-23980333	m	yes
Cpeb4	chr11:31768593-31773009	m	yes
Nsg2	chr11:31899369-31901633	m	yes
Il9r	chr11:32099323-32101255	m	yes
Rhbdf1	chr11:32121593-32124003	m	yes
Mpg	chr11:32125669-32128921	m	yes
Nprl3	chr11:32166768-32169183	m	yes
Hba-x	chr11:32175705-32177430	m	yes
Hba-a2_1	chr11:32181947-32185207	m	yes
Hbq1b	chr11:32185405-32187866	m	yes
Hba-a2_2	chr11:32194916-32198024	m	yes
Hbq1a	chr11:32197731-32200971	m	yes
Sh3pxd2b	chr11:32246453-32248498	m	yes
Ubtd2	chr11:32353233-32357166	m	yes
Efcab9	chr11:32426039-32428721	m	yes
Stk10	chr11:32431062-32434188	m	yes
Fbxw11	chr11:32541910-32544625	m	yes
Rnf145	chr11:44331299-44333443	b	yes
Mtmr3	chr11:4493462-4496044	b	yes
Tcf7	chr11:52094595-52097941	m	yes
Sept8	chr11:53332526-53334206	b	yes
Slc22a4	chr11:53841061-53842597	b	no
Hist3h2ba	chr11:58761131-58763833	b	yes
Trim17	chr11:58767509-58769742	b	yes
Map2k3	chr11:60744592-60746334	b	yes
Specc1	chr11:61889616-61892561	b	yes
E130309D14Rik	chr11:74432469-74433897	b	no
Mir144	chr11:77884385-77886812	m	yes
Tlcd1	chr11:77989592-77993044	b	yes
AI662270	chr11:83036103-83039672	b	no
Dhrs11	chr11:84641128-84643500	b	yes
Usp32	chr11:84952036-84955047	b	yes
AK020764	chr11:87567970-87569787	m	yes
Samd14	chr11:94869855-94872231	b	yes
Fam117a	chr11:95197022-95200374	b	yes
Pnpo	chr11:96803865-96805833	b	yes
Sp2	chr11:96837475-96840161	b	yes
Arhgap23_1	chr11:97309412-97314133	b	yes
Ormdl3	chr11:98446773-98449205	b	yes
Asb2_1	chr12:104572413-104574626	m	yes
Asb2_2	chr12:104590874-104595194	m	yes
Otub2	chr12:104625331-104627731	b	yes

Ifi27l1	chr12:104669405-104673428	b	yes
Glrx5	chr12:106269451-106271709	b	yes
Tnfaip2	chr12:112679278-112681725	b	yes
Id2	chr12:25779861-25783726	b	yes
Rsad2	chr12:27139691-27144929	b	yes
Mfsd2b	chr12:4880376-4881885	b	yes
Nfkbia	chr12:56592933-56595258	b	yes
Spnb1	chr12:77810796-77813323	b	yes
Plek2	chr12:80007276-80009012	b	yes
1110018G07Rik	chr12:86310542-86313402	b	yes
Jdp2	chr12:86939306-86941925	m	yes
Taf9	chr13:101419558-101422507	b	yes
Pik3r1	chr13:102537098-102539312	b	yes
Hist1h4i	chr13:22131191-22134208	b	yes
Hist1h2ac	chr13:23774105-23776885	b	no
Hist1h1c	chr13:23829333-23832543	b	yes
Slc22a23	chr13:34436021-34438348	b	no
Mylip	chr13:45484097-45485927	b	yes
Mxd3	chr13:55429525-55433208	b	yes
Isca1	chr13:59869723-59872303	b	yes
Zcchc6	chr13:59922656-59926436	b	yes
Lpcat1	chr13:73603441-73606668	b	yes
AK043267	chr13:83662196-83666327	m	yes
Ankrd34b	chr13:93191166-93196550	m	yes
Ncoa4	chr14:32972468-32973628	b	no
5730469M10Rik	chr14:41819526-41822488	b	yes
Gch1	chr14:47807036-47810663	b	yes
Pnp2	chr14:51575183-51577049	b	no
Pck2	chr14:56158002-56160720	b	yes
Spata13	chr14:61252247-61255142	b	yes
Ctsb	chr14:63740255-63742060	b	yes
Bnip3l	chr14:67626250-67629334	b	yes
Slc25a37	chr14:69901732-69904243	b	yes
Epb4.9_1	chr14:71025987-71027781	b	yes
Epb4.9_2	chr14:71033839-71036895	b	yes
Xpo7	chr14:71165099-71167451	b	yes
Rb1	chr14:73723824-73726697	b	yes
Itm2b	chr14:73783358-73786473	b	yes
1190002H23Rik	chr14:79700554-79702772	b	yes
Mettl7a1	chr15:100134132-100137232	b	yes
Nfe2	chr15:103084003-103087544	m	yes
March11	chr15:26081543-26082693	b	no
Rnf19a	chr15:36211260-36213903	b	yes
Azin1	chr15:38446231-38449805	b	yes
Nov	chr15:54574751-54578962	m	yes
Rnf139	chr15:58718872-58721988	b	yes
St3gal1	chr15:67007509-67009366	b	yes
Ptp4a3_1	chr15:73554325-73556927	b	yes
Ptp4a3_2	chr15:73577540-73581237	b	yes
Grina	chr15:76076089-76079690	b	yes
Maf1	chr15:76180617-76182572	b	yes
Apol8	chr15:77584427-77587070	b	yes
Mpst	chr15:78234073-78238766	b	yes
Ankrd54	chr15:78892366-78895004	b	yes
Apobec3	chr15:79722300-79723792	b	yes

Fam109b	chr15:82170807-82172698	b	yes
Tspo	chr15:83392533-83394821	b	yes
Shank3	chr15:89328379-89331399	m	yes
1110020G09Rik	chr15:9000030-9001750	b	yes
Slc48a1	chr15:97614114-97616573	b	yes
Vdr	chr15:97738161-97739861	m	yes
Cebpd	chr16:15886391-15889248	m	yes
Arvcf	chr16:18346073-18350065	b	yes
Txnrd2	chr16:18425917-18429630	b	yes
Bcl6	chr16:23987936-23990543	m	yes
Pcyt1a	chr16:32428252-32431919	b	yes
Cd47	chr16:49854561-49857524	b	yes
Cpox	chr16:58669525-58671667	b	yes
Nrip1	chr16:76372481-76374197	m	no
Usp25	chr16:77013652-77016235	b	yes
Carhsp1	chr16:8670606-8673427	b	no
Ifnar2	chr16:91372445-91374160	b	yes
Runx1_1	chr16:92696555-92698536	m	yes
Runx1_2	chr16:92823080-92827339	m	yes
Hagh	chr17:24986287-24988802	b	yes
Spsb3	chr17:25022906-25024642	b	yes
Ift140	chr17:25152487-25154402	b	yes
Pigq	chr17:26077322-26079992	b	yes
D17Wsu92e	chr17:27956435-27959117	b	yes
Brpf3	chr17:28937547-28939705	b	yes
March2	chr17:33852499-33856525	b	yes
Hspa1a	chr17:35108514-35110086	b	no
Abhd16a	chr17:35225445-35227239	b	yes
H2-D1	chr17:35399043-35400754	b	no
Rhag	chr17:40946891-40949985	b	yes
Runx2	chr17:44872871-44875421	m	yes
Nfkbie	chr17:45691683-45693508	m	yes
Mad2l1bp	chr17:46289444-46291802	b	yes
Gtpbp2	chr17:46296182-46298947	b	yes
Mocs1	chr17:49566646-49568893	b	yes
Plcl2	chr17:50647197-50650125	b	yes
Rfx2	chr17:56968997-56971470	b	yes
Lpin2_1	chr17:71531666-71534016	b	yes
Lpin2_2	chr17:71551059-71557076	b	yes
Ypel5	chr17:73183425-73187416	b	yes
3110002H16Rik	chr18:12326340-12328179	b	yes
Gypc	chr18:32718578-32722082	b	yes
March3	chr18:57084595-57086657	b	yes
Adrb2	chr18:62338520-62342448	b	yes
Fech	chr18:64648218-64649891	b	yes
Cep76	chr18:67799255-67801556	b	yes
Fth1	chr19:10056452-10058162	b	yes
Rab3il1	chr19:10091508-10094125	b	yes
Ostf1	chr19:18705056-18707158	m	yes
Aldh1a1	chr19:20675477-20678699	b	yes
Ankrd22	chr19:34239607-34241523	m	yes
Btrc	chr19:45436425-45441041	b	yes
Ldb1_1	chr19:46111151-46115143	m	yes
Ldb1_2	chr19:46117834-46122858	m	yes
Nfkb2	chr19:46378484-46380755	b	yes

Mxi1	chr19:53382567-53387208	b	yes
Rnaseh2c	chr19:5600523-5603182	b	yes
Ehbp1l1	chr19:5725636-5726865	b	no
Mtvr2	chr19:5728820-5730856	b	yes
Neat1	chr19:5843460-5846423	b	yes
Snx15	chr19:6127502-6133650	b	yes
Pla2g16	chr19:7631125-7633129	b	yes
Slc3a2	chr19:8787619-8789162	b	yes
Snhg1	chr19:8797097-8800982	b	yes
Cat	chr2:103323660-103326195	b	yes
Lmo2_1	chr2:103808783-103811465	m	no
Chac1	chr2:119175255-119179085	b	yes
Oip5	chr2:119442669-119446950	b	yes
Ccndbp1	chr2:120833263-120836077	b	yes
1700037H04Rik	chr2:130984900-130986552	b	yes
Cdc25b	chr2:131010718-131013789	b	yes
Smox	chr2:131316001-131319532	b	yes
Gpcpd1	chr2:132401645-132404882	b	yes
Trib3	chr2:152167972-152171402	b	yes
Bcl2l1	chr2:152656140-152658916	b	yes
Serinc3	chr2:163470609-163472614	b	yes
Dnttip1	chr2:164570283-164572762	b	yes
2010011I20Rik	chr2:172169547-172172125	b	yes
Rbm38	chr2:172846576-172848808	b	yes
Ubac1	chr2:25876051-25878467	b	yes
Gfi1b	chr2:28475529-28478476	m	yes
Crat	chr2:30270243-30274365	b	yes
Stom	chr2:35191392-35194432	b	yes
Frmd4a	chr2:4073136-4075519	b	yes
Optn	chr2:4983609-4986046	b	yes
Ypel4	chr2:84573106-84575838	b	yes
Ube2l6	chr2:84638391-84641506	b	yes
Slc43a1	chr2:84678477-84680071	b	yes
Slc43a3	chr2:84776015-84780007	b	no
Sfpi1	chr2:90936070-90937553	b	yes
Cd82	chr2:93299985-93304193	b	yes
Fam46c	chr3:100292037-100293646	b	no
Trim33	chr3:103079869-103086036	m	yes
Hipk1	chr3:103594199-103596584	b	yes
Gclm	chr3:121947168-121949996	b	yes
Lmo4	chr3:143864664-143869324	m	yes
Car2	chr3:14885056-14887361	b	yes
Fubp1	chr3:151872142-151874625	b	yes
D930015E06Rik	chr3:83841670-83845480	b	yes
Fhdc1	chr3:84282979-84286119	b	yes
Lmna	chr3:88306787-88308042	b	yes
Cdkn2c	chr4:109337038-109339382	b	yes
Tal1	chr4:114729436-114733674	m	yes
Pdzk1ip1	chr4:114760467-114762542	b	yes
Urod	chr4:116666152-116668126	b	yes
Slc6a9	chr4:117506860-117511054	b	yes
Mpl	chr4:118128813-118131264	m	yes
Ermap	chr4:118861852-118865789	b	yes
Scmh1	chr4:120076147-120079484	m	yes
Cited4	chr4:120337767-120341208	m	yes

Bsdc1	chr4:129137911-129140874	b	yes
Tinagl1	chr4:129849721-129852859	b	yes
Epb4.1	chr4:131630511-131632317	b	yes
Rhd	chr4:134418892-134421378	b	yes
Runx3_1	chr4:134675801-134677986	m	yes
Runx3_2	chr4:134706782-134709676	m	yes
E2f2	chr4:135727018-135729011	m	yes
Pink1	chr4:137881396-137883130	b	yes
Tprgl	chr4:153534027-153536830	b	no
Tnfrsf14	chr4:154301460-154303640	b	yes
Isg15	chr4:155574138-155577740	b	yes
AK172315	chr4:155574923-155578872	b	yes
Dcaf12	chr4:41261247-41263785	b	yes
B230312A22Rik	chr4:43057778-43059673	b	yes
Mcart1	chr4:45420835-45422398	b	yes
Tmod1	chr4:46051468-46052855	b	no
Hemgn	chr4:46415817-46418430	b	yes
Alad	chr4:62180587-62182418	b	yes
Mllt3_1	chr4:87450692-87453128	m	yes
Mllt3_2	chr4:87675864-87684376	m	yes
Cdkn2b	chr4:88955909-88957893	m	yes
Gfi1	chr5:108151654-108155261	m	yes
Hscb	chr5:111267685-111270292	b	yes
Iscu	chr5:114221597-114223744	b	no
Rnf10	chr5:115721672-115724052	b	yes
Prkab1	chr5:116473841-116475345	b	yes
Iqcd	chr5:121038030-121040266	b	yes
Oas3	chr5:121226712-121229724	b	yes
Ubb	chr5:125869174-125870875	b	yes
Stx2	chr5:129512420-129516553	b	yes
Fis1_1	chr5:137427661-137429976	b	yes
Fis1_2	chr5:137436961-137438993	b	yes
Gpr146	chr5:139855624-139858383	b	yes
Mafk	chr5:140266971-140269233	b	yes
2810453I06Rik	chr5:144308385-144310220	b	yes
Cyth3	chr5:144382309-144384504	b	yes
Eif2ak1	chr5:144577907-144579989	b	yes
Fam126a	chr5:23534677-23537251	b	yes
Rnf4	chr5:34677603-34681197	m	yes
Ldb2	chr5:45188917-45192551	m	yes
Klf3	chr5:65193336-65197629	b	yes
Abcb4	chr5:8892729-8896355	b	yes
Dck	chr5:89193176-89194852	b	yes
Nup54	chr5:92863468-92864769	b	no
Bhlhe40	chr6:108608850-108611329	m	no
Pparg	chr6:115309649-115312175	m	yes
March8	chr6:116286738-116289964	b	yes
Cdkn1b	chr6:134869212-134872724	b	yes
Hebp1	chr6:135117642-135118964	b	yes
8430419L09Rik	chr6:135147113-135149307	b	yes
H2afj	chr6:136754913-136758153	b	yes
Art4	chr6:136805154-136806676	b	yes
Cmas	chr6:142703994-142707386	b	yes
Arf5	chr6:28370772-28374224	b	yes
Irf5	chr6:29475355-29480378	m	yes

Tspan33	chr6:29643315-29645851	b	yes
Ube2h	chr6:30252481-30256046	b	yes
Bpgm	chr6:34424942-34427818	b	yes
Wdr91	chr6:34858818-34862018	b	yes
1810058I24Rik	chr6:35200565-35203707	b	yes
Zc3hav1	chr6:38303717-38305469	b	yes
Hipk2	chr6:38825110-38827388	m	yes
Mkrn1	chr6:39368997-39371199	b	yes
Kel	chr6:41653200-41655178	b	yes
Hoxa7	chr6:52167290-52169761	m	yes
Hoxa9	chr6:52175925-52178462	m	yes
Aqp1	chr6:55284080-55288910	b	yes
Nt5c3	chr6:56873179-56874845	b	yes
Vopp1	chr6:57773645-57776301	b	yes
Gadd45a	chr6:66986110-66988617	b	yes
St3gal5	chr6:72046167-72049860	b	yes
Bola3	chr6:83297655-83300710	b	yes
Add2	chr6:86026790-86029485	b	yes
Mxd1	chr6:86617393-86620910	b	yes
Rab43	chr6:87761018-87762504	b	no
Gata2	chr6:88146537-88149425	m	yes
Abtb1	chr6:88791159-88793624	b	yes
Tpra1	chr6:88851221-88853198	b	yes
Arrb1	chr7:106682222-106685966	b	yes
Ucp2	chr7:107641210-107643288	b	yes
Rhog	chr7:109397880-109399842	b	yes
Hbb-b1	chr7:110961088-110963317	b	yes
Beta-s	chr7:110975097-110977640	b	yes
Gm5736	chr7:110987131-110990171	b	yes
Hbb-bh1	chr7:110990723-110993013	b	yes
Hbb-y	chr7:111000780-111003651	b	yes
Rnf141	chr7:117985710-117989068	m	no
Sox6_1	chr7:122986664-122991215	b	yes
Sox6_2	chr7:123173461-123176353	m	yes
Cdr2	chr7:128123999-128127289	b	yes
Gtf3c1	chr7:132849281-132851773	m	yes
Nupr1	chr7:133767785-133770162	m	yes
Ypel3	chr7:133919598-133922219	b	yes
Maz	chr7:134167950-134172194	m	yes
Fam53b	chr7:140003007-140007299	b	yes
Uros	chr7:140900208-140902579	b	yes
Mki67	chr7:142906671-142909854	b	yes
Ptdss2	chr7:148315923-148317995	b	yes
Irf7	chr7:148451704-148453465	b	yes
2700078K21Rik	chr7:149245333-149247654	b	yes
Dusp8	chr7:149280181-149283750	b	yes
Napa	chr7:16682949-16685957	b	yes
Slc1a5	chr7:17365892-17367398	b	yes
Relb	chr7:20213517-20215697	b	yes
Bcl3	chr7:20406886-20408968	m	yes
Kcnn4	chr7:25154029-25156467	b	yes
Dedd2	chr7:26004233-26005671	b	yes
Tgfb1	chr7:26470910-26472923	m	yes
Blvrb	chr7:28231907-28233654	b	yes
Cebpa	chr7:35903742-35905555	m	no

Josd2	chr7:51722545-51724293	b	yes
Atf5	chr7:52070509-52072661	m	no
Nup62	chr7:52070822-52072533	m	yes
Tbc1d17	chr7:52103351-52106002	b	yes
Ppp1r15a	chr7:52780094-52782567	b	yes
Akap13	chr7:82599514-82603008	b	yes
Klhl25	chr7:82992467-82993788	b	yes
Isg20	chr7:86057154-86060474	b	yes
Zfp710	chr7:87168048-87172716	m	yes
Mesdc1	chr7:91031121-91033671	b	yes
Ces2g	chr8:107484726-107486699	b	yes
Ranbp10	chr8:108349361-108352445	b	yes
St3gal2	chr8:113440006-113444610	b	yes
Map1lc3b	chr8:124113447-124114882	b	yes
Slc7a5	chr8:124430170-124432264	b	yes
Zfpm1	chr8:124804977-124807264	m	no
Cbfa2t3_1	chr8:125200822-125203623	m	yes
Cbfa2t3_2	chr8:125221585-125224790	m	yes
Abcb10	chr8:126505728-126508363	b	yes
Cul4a	chr8:13104860-13106982	b	yes
Cln8	chr8:14886149-14889106	b	yes
Ank1_1	chr8:24084385-24086619	m	no
Ank1_2	chr8:24144780-24148103	m	yes
Ank1_3	chr8:24166557-24170057	m	yes
Acsl1	chr8:47554747-47558165	b	yes
Irf2	chr8:47823962-47825987	b	yes
Clcn3	chr8:63460510-63464596	b	yes
Jund	chr8:73220763-73222453	b	yes
Ankle1	chr8:73929104-73931073	b	yes
Klf2	chr8:74840891-74844669	m	yes
Lyl1	chr8:87224459-87225982	b	yes
Rad23a	chr8:87364092-87365358	b	yes
Klf1	chr8:87424247-87427229	m	yes
Dnase2a	chr8:87431676-87433093	b	yes
Prdx2	chr8:87492765-87494325	b	yes
Mylk3	chr8:87888635-87890896	b	yes
Herpud1	chr8:96908762-96912223	b	yes
1300017J02Rik	chr9:103189124-103192079	b	yes
Cish	chr9:107197995-107200206	b	yes
Uba7	chr9:107876782-107878661	b	yes
Ip6k1	chr9:107903806-107906153	b	yes
Rnf123	chr9:107981035-107983000	b	yes
Apeh	chr9:107992516-107997687	b	yes
Gpx1	chr9:108238298-108242296	b	yes
Eomes	chr9:118385637-118389401	m	yes
Slc25a38_1	chr9:120018968-120021123	b	no
Slc25a38_2	chr9:120021727-120025190	b	yes
Limd1	chr9:123386649-123388849	b	yes
Slc6a20a	chr9:123586896-123588896	b	yes
Cdkn2d	chr9:123835521-123837497	b	yes
Icam4	chr9:20833137-20834654	b	yes
Atg4d	chr9:21068133-21071182	b	yes
Epor	chr9:21766745-21769451	b	yes
Tbcel	chr9:42279322-42280966	b	yes
Hmbs	chr9:44149863-44153146	b	yes

Fam55b	chr9:48148332-48150030	b	yes
Dnaja4	chr9:54546285-54548410	b	yes
Clk3	chr9:57612025-57614442	b	yes
Smad3	chr9:63604956-63606366	m	yes
Dennd4a	chr9:64657568-64661104	b	yes
Snx22	chr9:65916883-65919178	b	yes
Dapk2	chr9:66005331-66007851	b	yes
Herc1	chr9:66197455-66201012	b	yes
Rab8b	chr9:66766112-66768621	b	yes
Gclc	chr9:77601433-77604566	b	yes
Fbxo9	chr9:77955424-77959398	b	yes
Paqr9	chr9:95456922-95463249	b	yes
Tfdp2	chr9:96095398-96097673	b	no
Mid1ip1	chrX:10293229-10295726	b	yes
Bcor	chrX:11656536-11658198	m	no
Alas2	chrX:146981075-146983137	b	yes
Asb11	chrX:160873808-160876716	m	yes
Mpp1	chrX:72375520-72377351	b	yes
Wdr45	chrX:7298478-7299963	b	yes
Otud5	chrX:7417184-7419568	b	yes
Gata1	chrX:7543919-7545685	m	yes
Slc38a5	chrX:7847541-7850650	b	yes

Table S1

Column 1 contains the name of the targeted gene promoter. Genes with multiple promoter are denoted by as numbered suffix separated by an underscore. The coordinates of the captured fragment for each promoter in given in column 2 using UCSC notation based on genome build MM9. In the third column regions are annotated as either manually chosen (m) or included from bioinformatics analysis (b) of gene expression in mES and Ter119+ cells (see methods). The fourth columns shows whether capture efficiency at this region exceed the imposed cut off of 50000 reads.

Supplementary table S2. Number of unique distal interactions at captured gene promoters

Chromosome	Gene Name	Unique Distal Interactions
chr17	Hspa1a	17976
chr9	Cdkn2d	10757
chr5	Oas3	7941
chr14	Pnp2	6897
chr11	Hist3h2ba	6335
chr11	AI662270	5880
chr11	Fam117a	3548
chr7	Beta-s	3386
chr11	Hba-a2_1	3296
chr15	Mettl7a1	3064
chr11	Hba-a2_2	3045
chr17	H2-D1	2970
chr7	Hbb-b1	2718
chr7	Tbc1d17	2639
chr9	Rab8b	2548
chr1	Zfand2b	2534
chr6	Bpgm	2528
chr1	Trak2	2453
chr14	Spata13	2435
chr19	Snx15	2387
chr4	Bsdc1	2379
chr11	Tlcd1	2377
chr5	Stx2	2333
chr11	Trim17	2317
chr1	Creg1	2156
chr10	Slc16a10	2140
chrX	Asb11	2120
chr14	Gch1	2116
chr9	Dapk2	2090
chr11	Cpeb4	2088
chr2	Ypel4	2013
chr1	Btg2	2011
chr9	Uba7	1986
chr11	Mpp2	1984
chr4	Mllt3_1	1979
chr4	Hemgn	1978
chr2	1700037H04Rik	1968
chr7	Josd2	1964
chr8	Ranbp10	1933
chr11	Ubb	1932
chr7	Ppp1r15a	1928
chr7	Maz	1922
chr19	Neat1	1918
chr7	Hbb-y	1914
chr9	Gpx1	1910
chr11	Pnpo	1905
chr11	Samd14	1881
chr2	Gpcpd1	1860
chr19	Slc3a2	1843
chr6	Ube2h	1839
chr4	Tal1	1839
chr17	Lpin2_2	1831
chr6	Add2	1822
chr11	Hbq1b	1811
chr10	Hace1	1808
chr1	1810031K17Rik	1807
chr17	Plcl2	1794

chr7	Hbb-bh1	1787
chr9	Snx22	1778
chr17	Rhag	1775
chr7	Blvrb	1767
chr7	2700078K21Rik	1742
chr4	Tnfrsf14	1703
chr5	Gpr146	1694
chr12	Rsad2	1687
chr7	Mki67	1684
chr1	Rnasel	1683
chr1	Spna1	1680
chr5	Fam126a	1670
chr6	Mxd1	1669
chr11	Hbq1a	1668
chr4	B230312A22Rik	1656
chr4	Cdkn2c	1642
chr1	Tmcc2	1642
chr14	1190002H23Rik	1635
chr13	Taf9	1633
chr6	8430419L09Rik	1618
chr7	Slc1a5	1611
chr5	Eif2ak1	1609
chr2	Gfi1b	1605
chr8	Cbfa2t3_2	1584
chr9	Fbxo9	1581
chr9	Cish	1580
chr2	Oip5	1577
chr11	Nprl3	1561
chr1	Fam134a	1529
chr2	Ube2l6	1526
chr15	Ptp4a3_1	1519
chr9	Ip6k1	1499
chr13	Mylip	1492
chr11	Sp2	1491
chr7	Sox6_1	1486
chr13	Isca1	1475
chr16	Runx1_2	1474
chr2	Lmo2_1	1472
chr10	Mknk2	1464
chr9	Fam55b	1463
chr4	Epb4.1	1458
chr18	Gypc	1458
chr11	Ormdl3	1453
chr7	Ypel3	1448
chr12	Spnb1	1446
chr15	Azin1	1435
chr9	Paqr9	1422
chr7	Isg20	1419
chr18	Adrb2	1418
chr6	Cmas	1418
chr13	Lpcat1	1413
chr13	Hist1h1c	1409
chr19	Fth1	1408
chr4	Dcaf12	1408
chr11	Atp6v0a1	1402
chr7	Fam53b	1395
chr10	1110038D17Rik	1393
chr6	Abtb1	1385
chr13	Mxd3	1383
chr5	Dck	1367
chr17	Ift140	1365
chr4	Mpl	1363
chr11	Sertad2	1350

chr3	Gclm	1350
chr7	Gm5736	1349
chr2	Cdc25b	1348
chr10	Fbxo30	1341
chr14	5730469M10Rik	1340
chr15	Rnf19a	1335
chr17	Pigq	1329
chr1	Ppox	1327
chr5	Rnf4	1324
chr17	Brpf3	1320
chr17	Lpin2_1	1312
chr9	Slc6a20a	1311
chr6	Art4	1311
chr7	Dedd2	1301
chr16	Arvcf	1299
chr4	Isg15	1295
chr6	H2afj	1294
chr1	Dnajb2	1290
chr8	Klf1	1290
chr11	Slc25a39	1288
chr6	Cdkn1b	1286
chr10	Metap2	1260
chrX	Alas2	1256
chr2	Stom	1253
chr4	Ermap	1251
chr15	Ptp4a3_2	1235
chr12	1110018G07Rik	1234
chr16	Txnrd2	1232
chr10	Fbxo7	1230
chr1	Atg9a	1227
chr8	Lyl1	1227
chr8	Acsl1	1222
chr4	Mllt3_2	1208
chr9	Gclc	1208
chr17	Spsb3	1208
chr2	Slc43a1	1208
chr4	Tinagl1	1206
chr15	Mpst	1204
chr18	Mar-03	1194
chr11	Wipi1	1194
chr2	Bcl2l1	1189
chr11	Tcf7	1178
chr9	Hmbs	1177
chr14	Pck2	1177
chr8	St3gal2	1176
chr6	Tspan33	1175
chr11	Efcab9	1173
chr19	Snhg1	1172
chr2	Smox	1170
chr8	Herpud1	1165
chr11	AK020764	1165
chr4	Urod	1163
chr18	Cep76	1152
chr3	Trim33	1151
chr12	Ifi27l1	1148
chr5	Abcb4	1146
chr4	Pdzk1ip1	1133
chr19	Aldh1a1	1127
chr11	Fbxw11	1127
chr10	Serinc1	1114
chr1	Atp2b4_1	1112
chr6	Mkrn1	1109
chr17	Hagh	1109

chr19	Rnaseh2c	1107
chr7	Akap13	1106
chr7	Irf7	1103
chr12	Tnfaip2	1102
chr14	Epb4.9_1	1101
chr17	Rfx2	1100
chr16	Pcyt1a	1090
chr2	2010011I20Rik	1088
chr3	Fhdc1	1086
chr6	Kel	1084
chr9	Epor	1081
chr11	Slc4a1	1077
chr2	Dnttip1	1077
chr2	Serinc3	1076
chr2	Trib3	1067
chr15	Grina	1061
chr9	Clk3	1060
chr6	1810058I24Rik	1055
chr1	Asb1	1050
chr5	Cyth3	1046
chr11	Arhgap23_1	1037
chr11	Meis1	1033
chr14	Itm2b	1031
chr3	Hipk1	1031
chr11	Stk10	1020
chr7	Klhl25	1017
chr8	Jund	1003
chr6	Irf5	995
chr4	Scmh1	993
chr9	1300017J02Rik	987
chr15	Ankrd54	986
chr5	Hscb	984
chr7	Napa	982
chr4	Cited4	979
chr9	Atg4d	964
chr19	Ankrd22	961
chr7	Nupr1	960
chr14	Ncoa4	954
chr6	Mar-08	949
chr19	Pla2g16	948
chr5	Klf3	938
chr3	Lmo4	936
chr2	Optn	926
chr8	Prdx2	925
chr16	Cd47	924
chr4	Rhd	922
chr1	Ube2w	921
chr15	Rnf139	921
chr9	Limd1	917
chr11	Narf	916
chr7	Sox6_2	912
chr7	Cdr2	905
chr6	Gadd45a	904
chr6	Nt5c3	901
chr5	Iqcd	901
chr13	Hist1h4i	898
chr5	Fis1_2	896
chr11	Ubtd2	893
chr6	Hipk2	892
chr17	Mocs1	890
chr2	Ccndbp1	883
chr6	St3gal5	877
chr15	Maf1	874

chr4	Slc6a9	874
chr6	Bola3	872
chr2	Chac1	869
chr12	Id2	868
chrX	Slc38a5	868
chr6	Wdr91	867
chr14	Epb4.9_2	867
chr5	Ldb2	866
chr19	Btrc	865
chr11	Map2k3	861
chr10	Nr1h4_1	859
chr8	Rad23a	858
chr7	Ptdss2	857
chr11	Fn3k	854
chr16	Ifnar2	853
chr10	Trappc10	852
chr11	Dhrs11	850
chr17	Mad2l1bp	840
chr16	Bcl6	839
chr2	Frmd4a	835
chrX	Otud5	829
chr15	Apobec3	829
chr11	Mpg	829
chr18	3110002H16Rik	827
chr8	Ces2g	825
chr11	Hba-x	822
chr11	Mir144	822
chr9	Apeh	821
chr10	Oaz1	817
chrX	Mpp1	817
chr9	Rnf123	815
chr13	AK043267	798
chr1	Mgst3	796
chr3	D930015E06Rik	793
chr10	Foxo3	791
chr19	Mxi1	789
chr8	Ank1_3	789
chr8	Irf2	788
chr14	Slc25a37	787
chr14	Ctsb	787
chr12	Otub2	781
chr8	Ankle1	779
chr4	AK172315	776
chr9	Herc1	775
chr9	Tbcel	773
chr8	Clcn3	772
chr3	Lmna	771
chr7	Arrb1	760
chr1	Atp2b4_2	758
chr11	Fn3krp	748
chr8	Cbfa2t3_1	737
chr6	Hebp1	735
chr7	Gtf3c1	733
chr7	Zfp710	731
chr10	Tcp11l2	729
chr19	Mtvr2	720
chr14	Bnip3l	714
chr6	Aqp1	706
chr9	Icam4	699
chr10	Fzr1	698
chr12	Plek2	697
chr4	Pink1	696
chr17	Gtpbp2	690

chr15	Apol8	683
chr6	Zc3hav1	683
chr5	2810453I06Rik	682
chrX	Mid1ip1	678
chr6	Hoxa7	678
chrX	Gata1	675
chr7	Bcl3	674
chr9	Eomes	674
chr7	Relb	672
chr4	Mcart1	671
chr12	Mfsd2b	669
chr1	Prrx1	669
chr10	Bsg	650
chr10	Cep57l1	649
chr15	Fam109b	646
chr11	Psme3	631
chr12	Glrx5	630
chr16	Cpox	628
chr5	Fis1_1	623
chr5	Gfi1	620
chr12	Jdp2	611
chr10	Cited2	610
chr4	Runx3_2	600
chr12	Nfkbia	599
chr15	Slc48a1	588
chr13	Zcchc6	586
chr19	Ostf1	585
chr6	Hoxa9	582
chr16	Usp25	581
chr7	Ucp2	580
chr6	Vopp1	576
chr17	Ypel5	575
chr1	Arid5a	568
chr17	D17Wsu92e	566
chr11	Mtmr3	563
chr2	Crat	560
chr7	Rhog	555
chr2	Sfpi1	555
chr17	Nfkbie	552
chr9	Slc25a38_2	548
chr6	Gata2	546
chr19	Nfkb2	545
chr5	Mafk	542
chr5	Iscu	542
chr11	Il9r	541
chr6	Tpra1	534
chr11	Usp32	533
chr4	Alad	530
chr4	Cdkn2b	518
chr8	Cln8	504
chr10	Nr1h4_2	503
chr7	Uros	497
chr15	Nov	494
chr9	Dennd4a	487
chr8	Klf2	481
chr15	St3gal1	479
chr8	Map1lc3b	474
chr14	Rb1	472
chr7	Mesdc1	472
chr7	Atf5	460
chr5	Prkab1	460
chrX	Wdr45	459
chr5	Ubb	455

chr8	Ank1_2	453
chr1	Abcb6	452
chr2	Cd82	452
chr2	Lmo2_1	450
chr16	Cebpd	449
chr13	Ankrd34b	447
chr11	Mafg	447
chr1	Cxcr4	446
chr7	Kcnn4	443
chr9	Dnaja4	430
chr6	Rab43	421
chr3	Car2	421
chr8	Slc7a5	416
chr9	Smad3	412
chr9	Slc25a38_1	411
chr10	Cd24a	407
chr19	Ldb1_1	406
chr2	Cat	404
chr12	Asb2_2	402
chr11	Kif18b	402
chr6	Pparg	396
chr15	Vdr	391
chr2	Rbm38	382
chr11	Rnf145	371
chr15	Shank3	369
chr11	Arhgap23_1	358
chr6	Arf5	356
chr12	Asb2_1	353
chr15	Tspo	351
chr8	Cul4a	349
chr5	Nup54	346
chr1	Dcaf6	345
chr7	Tgfb1	325
chr11	Sep-08	322
chr17	Mar-02	307
chr19	Ldb1_2	305
chr11	Bcl11a	303
chr10	Ddit3	294
chr17	Runx2	294
chr1	Tuba4a	287
chr11	Sh3pxd2b	285
chr3	Fam46c	283
chr9	Tfdp2	275
chr11	Nsg2	271
chr10	Sppl2b	267
chr19	Rab3il1	266
chr11	Rhbdf1	261
chr11	Tmc6	258
chr5	Rnf10	245
chr7	Rnf141	243
chr7	Dusp8	239
chr8	Abcb10	235
chr13	Pik3r1	233
chr15	Nfe2	233
chr8	Dnase2a	232
chr4	Runx3_1	232
chr15	Mar-11	218
chr3	Fubp1	217
chr8	Mylk3	209
chr8	Zfpm1	202
chr2	Ubac1	199
chr14	Xpo7	198
chr4	E2f2	195

chr7	Nup62	190
chr17	Abhd16a	185
chr6	Bhlhe40	181
chr13	Slc22a23	181
chr11	Specc1	174
chr16	Runx1_1	172
chr4	Tmod1	172
chr18	Fech	164
chr15	1110020G09Rik	161
chr16	Nrip1	149
chr13	Hist1h2ac	143
chr11	E130309D14Rik	128
chr16	Carhsp1	111
chr11	Ube2o	100
chr19	Ehbp1l1	99
chr11	Slc22a4	97
chr11	Birc5	93
chr2	Slc43a3	91
chrX	Bcor	72
chr4	Tprgl	66
chr14	Ubb	65
chr8	Ank1_1	58
chr7	Cebpa	27
chr1	Adipor1	24

Table S2
Numbers of unique (non duplicated based on genomic position), distal (greater than 2kb away from Capture fragments) interactions. Quantitatively skewed interactions between capture points have also been removed.