Title: Straightforward protocol for Allele-Specific Chromatin Conformation Capture.

Authors: Acemel RD[1+], Tena JJ[1+], Gomez-Skarmeta JL[1], Fibla J [2] & Royo JL[3*]

Affiliations:

1. Andalusian Centre for Developmental Biology, University Pablo de Olavide-CSIC-Junta de Andalucia. Sevilla, Spain.

2. Institute of Biomedical Research of Lleida, University of Lleida, Lleida, Spain.

3. Department of Surgery, Biochemistry, and Immunology, School of Medicine, University of Malaga, 29071, Malaga, Spain.

+ These authors contributed equally to this work.

* To whom correspondence should be addressed: Jose Luis Royo (jlroyo@uma.es)

University of Malaga

School of Medicine

Department of Biochemistry, Molecular Biology and Immunology 29071, Malaga, Spain

Abstract

A key advance in our understanding of gene regulation came with the finding that the genome undergoes three-dimensional nuclear folding in a genetically determined process. This 3D conformation directly influences the association between enhancers and their target promoters. This complex interplay has been proven to be essential for gene regulation, and genetic variants affecting this process have been associated to human diseases. The development of new technologies that quantify these DNA interactions represented a revolution in the field. High throughput techniques like HiC provide a general picture of chromatin topology. However, they often lack resolution to evidence subtle effects that single nucleotide polymorphisms exert over the contacts between cis-regulatory regions and target promoters. Here we propose a cost-efficient approach to perform allele-specific chromatin conformation analysis. As a proof of concept, we analyzed the impact of a common deletion mapping between *SIRPB1* promoter and one of its downstream enhancers.

Keywords: chromatin conformation; allele-specific 3C; eQTL; enhancer; promoter

1. Introduction

Over the past decades, genome-wide association studies (GWAS) have linked thousands of common genetic variants to a large spectrum of human traits. Fine-mapping of causal variants at GWAS *loci* is usually difficult because the majority (~88%) of these variants lie in noncoding genomic regions [1]. To fill the gap between genetic association and functionality, expression quantitative trait *loci* scans (eQTL) have been conducted, connecting germline genetic variation to somatic-cell biology [2, 3]. These studies have successfully identified tens of thousands of eQTLs in a variety of human tissues [4]. Many of such non-coding variants lay on open chromatin regions that function as transcriptional enhancers and control the expression of target genes that can be located hundreds of kilobases away [5, 6]. Point mutations in enhancers can increase or reduce the transcription of target genes, and they have been linked to several human traits [7-9]. However, since affected genes are often far away from regulatory mutations, it is not trivial to establish such associations.

Fortunately, key advances in the understanding of how chromatin folds shed light on long range regulation. Chromatin conformation capture techniques (C-techniques, e.g. 3C, 4C-seq and HiC) showed that chromatin folding facilitates the spatial interaction between enhancers and their target promoters [10]. HiC experiments, that profile chromatin folding genome-wide, unveiled that the mammalian chromosomes are partitioned in self insulated domains called TADs [11]. Importantly, enhancers and their target promoters often lay within the same TAD in order to interact and promote transcription. Indeed, structural variants leading to the disruption of TAD structures (e.g. deletions, duplications or inversions) have been linked to diseases caused by the spurious activation of genes [12]. However, TADs often comprise several genes and not all of them have coordinated expression patterns. Higher resolution HiC maps [13] showed that the activation of genes is sometimes preceded by cell type specific changes in intra-TAD structure, often involving the looping between the promoter of the activated gene and their enhancers. Those fine grain 3D changes are critical to link regulatory mutations with target genes, and were previously anticipated by studies based on *locus* specific C-techniques. Those techniques, like 4C-seq and 3C, easily achieve higher resolutions in exchange for their limited scope [6, 14]. For that reason, genome-wide studies linking SNPs to putative target genes use modifications of the HiC protocol that increase resolution in regulatory hotspots (e.g. Capture-C or HiChIP) [10].

Nevertheless, to establish if genetic variants are related to changes in chromatin folding, either genetic models or allele-specific C-techniques are in need. Allele-resolved HiC maps have been obtained phasing HiC interactions using known SNPs in deeply sequenced libraries [15]. However, even in such dense datasets, phased HiC maps are sparse and the resolution is

insufficient to reliably establish enhancer-promoter interactions. Allele specific 4C-seq strategies have been also successfully developed [16, 17], reaching the resolutions required to discriminate changes in enhancer-promoter interactions. However, the experimental design is very sequence-dependent and often difficult to extrapolate to different *loci*. On top of this, the impact of a SNP over the enhancer-promoter interaction may require a relative large number of independent biological samples, what might turn into an unaffordable increase in the budget of massive sequencing. Here we describe a straightforward, inexpensive approach to perform allele-specific 3C using conventional Taqman probes that can be used to assay allele specific contacts in up to 90% of the promoters of the human genome.

2. Materials and methods

2.1. Bioinformatics

Human and mouse DNA sequence, Single nucleotide polymorphisms and TSS positions were directly downloaded from UCSC Genome Browser using hg38 and mm10 assemblies. Restriction blocks were calculated using custom scripts in R (Supplementary methods). As potentially functional SNPs we selected every common variant from human hg38 and mouse mm10 with a heterozygosity >0.15, mapping a restriction enzyme fragment together with a transcriptional start gene site (TSS), and within a 100-1000 bp range from one of the restriction ends. The SNPs and the TSS were downloaded using the Table tool from the UCSC genome browser with the appropriate filters. TSS positions, valid SNPs and restriction fragments were overlapped using Bedtools software suite [18].

2.2. *SIRPB1* allele-specific 3C

Peripheral blood mononuclear cells were obtained from two different healthy donors heterozygous for the *SIRPB1* Copy-Number Variant, purified using Ficoll and rapidly fixed using paraformaldehyde (2% final concentration) for 10’ at room temperature. Fixation was then quenched with glycine (0.15 M final concentration). Isolated cells were lysed (10 mM Tris-HCl pH 8, 10 mM NaCl, 0.3% Igepal CA-630 (Sigma-Aldrich, I8896) and 1× protease inhibitor cocktail (Complete, Roche, 11697498001)), and the DNA was digested with HindIII (New England BioLabs, R0543M). T4 DNA ligase (Promega, M1804) was used for the subsequent ligation step. Specific primers were designed in the HindIII site harboring *SIRPB1* TSS (anchor) and in the 80Kb downstream from rs2209313 (Sensors, primers design in Table 1). PCR reactions using the Anchor, each of the Sensor primers and the 3C ligated DNA as template were set up (35 cycles of 95ºC for 30’’, 58ºC for 30’’ and 72ºC for 120’’). Every PCR product contained a common fragment of 1.8 Kb comprising the rs2200313 SNP, in linkage

disequilibrium with the deletions of the two downstream CTCF sites. Next, in order to determine the C/T ratio for each Anchor-Sensor pair, we used Taqman assay C1911298 (Thermofisher, CA, USA) over the generated amplicons according to manufacturer's instructions. As a C/T unbiased reference, a control primer within the same HindIII region of the SIRPB1 transcriptional start site (TSS) was used to amplify the adjacent region of the rs2209313. Those ratios reflect preferential Anchor-Sensor interactions in either genetic background.

3. Results

We propose a novel and cost-effective approach to profile allele specific interactions between specific pairs of loci (i.e between promoters and enhancers) with high resolution. The procedure begins by performing conventional 3C experiments from heterozygous donor cells to ensure equal amounts of both the mutant and the wild type alleles. Importantly, the 3C design must guarantee that the restriction fragment that contains the anchor primer also contains the SNP under evaluation that should be close enough to the restriction site to be amplifiable by PCR. Here, we proposed a maximum of 1Kb from the RE site, although this parameter can be adjusted depending on the complexity of the region. Next, a first PCR is performed using the anchor oligo (in the fragment containing, for instance, a promoter of interest) and a sensor oligo designed within the restriction fragment of interest (i.e. containing a known or putative enhancer). Crucially, this amplicon must contain the SNP of interest so that the PCR product can be further genotyped using conventional Taqman probes to determine potential bias in allele distributions. The allelic bias in this context can be interpreted as preferential interactions between anchor and sensor *loci* in either of the genetic backgrounds. Untreated genomic DNA from heterozygous donors and reverse primers mapping of the same restriction fragments are potential controls to monitor bias attributed to the Taqman assay (Figure 1).

As a proof of concept, we focused on SIRPB1 promoter and its known 3' enhancer. The SNP rs2209313 maps 1.5 Kb downstream the SIRPB1 promoter, and has been described to be in linkage disequilibrium with a functional deletion of two CTCF sites placed between the promoter and a 3' enhancer. According to previous results, rs2209313 allele-C was in *cis* with the deletion allele [19]. In order to perform an allele-specific 3C we obtained PBMCs from two independent rs2209313 heterozygous donors and conducted a conventional 3C using HindIII. The Anchor primer was designed to point towards the closest HindIII site, leaving the SNP rs2209313 in between. Four different reverse primers were designed to point towards the HindIII sites, in fragments that mapped 80, 57, 52 and 46 Kb away from the anchor primer of the *SIRPB1* promoter (Table 1). Once PCR fragments were obtained, these were subjected to a

nested rs2209313 quantitative genotyping using a conventional Taqman probe. As a control, a reverse primer was designed containing rs2209313 but within the same HindIII fragment than the anchor primer. Importantly, the Deletion/Insertion ratio (C/T) was not significantly biased when interrogating interactions between the Anchor and Sensors -70, -57 and -45: those that did not contain enhancer elements (U-Mann Whitney p-value > 0.05). However, Sensor-52 contacts with the Anchor were 3.17 times more frequent in the allele associated with the deletion of the CTCF sites ($2^{\Delta Ct}$=2.38 vs 0.75; U-Mann Whitney p=0.042) (Figure 2).

Once we demonstrated that the method is able to identify allele-specific differences in chromatin folding in the *SIRPB1* locus we wondered how well does it generalize to other *loci*. To interrogate that we performed *in silico* restriction analysis of the both the human (hg38) and mouse (mm10) genomes, using four conventional restriction enzymes commonly used in 3C. According to UCSC genome browser refseq annotation, Hg38 contains 36,672 TSS, from which 35,418 (96.58%) were contained within an EcoRI restriction fragment with at least one SNP potentially informative (Figure 3). Nevertheless, if the gene of interest is not captured by this initial approach, BamHI could be used instead. The alternative use of two enzymes increased the percentage of TSS captured to 98.63%. Upon the use of four alternative RE, 99.05% of all TSS could be included in our allele-specific 3C assay, since they are all contained in a fragment with at least one candidate SNP which could be used to perform the allele-specific assay. The remaining 0.95% (347 TSS) were not contained in a RE fragment with a candidate SNP. When the same approach was followed using mouse genomic data, the percentage of captured TSS were slightly lower. An initial analysis with EcoRI could capture 80.2% of the 26,607 mouse annotated TSSs, and using any of the four candidate RE this percentage raised up to 93.39%.

Finally, we wondered how often potential Anchor restriction fragments contain unique TSSs. From the potentially studied TSS from the human genome, 57.65% mapped alone in an EcoRI fragment. From those that shared the restriction fragment with another promoter, an additional 8.41% could be assayed alone in a BamHI restriction fragment using (reaching 66.07% of all TSS). This percentage increased again when using alternatively any of the previous enzymes and HindIII (up to 69.46%), and when BglII was also included in the design, 71.57% of all TSSs could be assayed alone in their restriction fragment. When the same approach was followed using mouse genomic data, the percentage of captured TSS alone in a RE fragment were again slightly lower. An initial analysis with EcoRI could capture 64.97% of the mouse TSS, and using any of the four candidate RE this percentage raised up to 84.35%.

4. Discussion

Here we describe an easy and inexpensive procedure to study the impact of point mutations in chromatin architecture and enhancer-promoter interactions. According to our data, allele-specific interactions of > 90% of the human promoters can be analyzed with this approach, once the appropriate restriction enzyme and SNP of interest are selected. The rationale underlying this protocol is simple: the restriction fragment used as 3C Anchor must contain a SNP susceptible to be genotyped, that can be either the functional mutation or simply a SNP in linkage disequilibrium with that functional mutation. The latest scenario is of special interest since SNPs can be in linkage disequilibrium with structural variants that have been implicated in modifications of chromatin folding when they are located in enhancer regions and TAD boundaries. This is indeed the case of our proof of concept example, in which the SNP rs2209313 is in linkage disequilibrium with the deletion of two CTCF sites and associated with an increased expression of the nearby *SIRPB1* gene [19, 20]. Using our approach, we showed that the linked deletion is likely facilitating the interaction between the *SIRPB1* promoter and a downstream enhancer.

Whenever this design is possible (> 90% of the cases if considering only promoters as Anchor) this methodology provides an affordable way to find differences in chromatin folding between alleles with single fragment resolution. Reaching fragment resolution is challenging and expensive when using genome wide techniques such as HiC, and the resolution is even more compromised if contact frequencies need to be phased in different alleles. The high resolution of our approach, together with its ease and moderate cost, facilitate the use of several biological and technical replicates resulting in an increased statistical power and the possibility to detect subtler changes in interaction frequencies. Notably, throughout this study we focused on the possibility of using restriction fragments containing promoters of interest as Anchor. However, enhancers are also susceptible to be used as Anchor. This fact is noteworthy: first because it increases the number of enhancer-promoter interactions susceptible to be assayed, but also because it allows to investigate the contacts of enhancers containing SNPs linked to the binding of specific activator or repressor transcription factors. Altogether, we consider that this approach will be a useful tool for those researchers that set to understand the role of chromatin architecture in transcriptional regulation and, more generally, to those trying to decipher the regulatory implications of non-coding mutations.

Aknowledgements

JLG-S acknowledges the Spanish Ministerio de Economía y Competitividad (Grant BFU2016-74961-P), the European Research Council (ERC, grant agreement No 740041) and the institutional grant Unidad de Excelencia María de Maeztu (MDM-2016-0687 to the Department

of Gene regulation and morphogenesis of Centro Andaluz de Biología del Desarrollo. Project supported by a 2019 Leonardo Grant for Researchers and Cultural Creators, BBVA Foundation, granted to JJT (The BBVA Foundation accepts no responsibility for the opinions, statements and contents included in the project and/or the results thereof, which are entirely the responsibility of the authors).

References

1. Edwards, S.L., Beesley, J., French, J.D., & Dunning, A.M. (2013). Beyond GWASs: illuminating the dark road from association to function. *Am. J. Hum. Genet.* 93, 779-797.
2. Lappalainen, T., Sammeth, M., Friedländer, M.R., 't Hoen, P.A., Monlong, J., Rivas, M.A., et al. (2013). Transcriptome and genome sequencing uncovers functional variation in humans. *Nature*. 501, 506-511.
3. GTEx Consortium. (2015) Human genomics. The Genotype-Tissue Expression (GTEx) pilot analysis: multitissue gene regulation in humans. *Science*. 348, 648-660.
4. Chiang, C., Scott, A.J., Davis, J.R., Tsang, E.K., Li, X., Kim, Y., Hadzic, T., Damani, F.N., Ganel, L., GTEx Consortium, Montgomery, S.B., Battle, A., Conrad, D.F.,& Hall, I.M. (2017). The impact of structural variation on human gene expression. *Nat. Genet.* 49, 692-699.
5. de la Calle-Mustienes, E., Feijóo, C.G., Manzanares, M., Tena, J.J., Rodríguez-Seguel, E., Letizia, A., Allende, M.L., & Gómez-Skarmeta, J.L. (2005). A functional survey of the enhancer activity of conserved non-coding sequences from vertebrate Iroquois cluster gene deserts. *Genome Res.* 15, 1061-1072.
6. Montavon, T., Soshnikova, N., Mascrez, B., Joye, E., Thevenet, L., Splinter, E., de Laat, W., Spitz, F., & Duboule, D. (2011). A regulatory archipelago controls Hox genes transcription in digits. *Cell*. 47, 1132-1145.
7. Smemo, S., Tena, J.J., Kim, K.H., Gamazon, E.R., Sakabe, N.J., Gómez-Marín, C., et al. (2014). Obesity-associated variants within FTO form long-range functional connections with IRX3. *Nature*. 507, 371-375.
8. Pasquali, L., Gaulton, K.J., Rodríguez-Seguí, S.A., Mularoni, L., Miguel-Escalada, I., Akerman, İ., et al. (2014). Pancreatic islet enhancer clusters enriched in type 2 diabetes risk-associated variants. *Nat. Genet.* 46, 136-143.
9. Spieler, D., Kaffe, M., Knauf, F., Bessa, J., Tena, J.J., Giesert, F., et al. (2014). Restless legs syndrome-associated intronic common variant in Meis1 alters enhancer function in the developing telencephalon. *Genome Res.* 24, 592-603.
10. Kempfer, R. & Pombo, A. (2020). Methods for mapping 3D chromosome architecture. *Nat. Rev. Genet.* 21, 207-226.
11. Nora, E.P., Lajoie, B.R., Schulz, E.G., Giorgetti, L., Okamoto, I., Servant, N., et al. (2012). Spatial partitioning of the regulatory landscape of the X-inactivation centre. *Nature.* 485, 381-385.

12. Lupiáñez, D.G., Kraft, K., Heinrich, V., Krawitz, P., Brancati, F., Klopocki, E., et al. (2015). Disruptions of topological chromatin domains cause pathogenic rewiring of gene-enhancer interactions. *Cell*. 161, 1012-1025.
13. Bonev, B., Mendelson Cohen, N., Szabo, Q., Fritsch, L., Papadopoulos, G.L., et al. (2017). Multiscale 3D Genome Rewiring during Mouse Neural Development. *Cell*. 171, 557-572.
14. Palstra, R.J., Tolhuis, B., Splinter, E., Nijmeijer, R., Grosveld, F., de Laat, W. (2003). The beta-globin nuclear compartment in development and erythroid differentiation. *Nat. Genet*. 35, 190-194.
15. Rao, S.S., Huntley, M.H., Durand, N.C., Stamenova, E.K., Bochkov, I.D., Robinson, J.T., et al. (2014). A 3D map of the human genome at kilobase resolution reveals principles of chromatin looping. *Cell*. 159, 1665-1680.
16. Simonis, M., Klous, P., Splinter, E., Moshkin, Y., Willemsen, R., de Wit, E., van Steensel, B., de Laat, W. (2006) Nuclear organization of active and inactive chromatin domains uncovered by chromosome conformation capture-on-chip (4C). *Nat. Genet*. 38, 1348-1354.
17. Llères, D., Moindrot, B., Pathak, R., Piras, V., Matelot, M., Pignard, B., Marchand, A., Poncelet, M., Perrin, A., Tellier, V., Feil, R., Noordermeer, D. (2019) CTCF modulates allele-specific sub-TAD organization and imprinted gene activity at the mouse Dlk1-Dio3 and Igf2-H19 domains. *Genome Biol.* 20, 272.
18. Quinlan AR, IM Hall. BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics*. 26, 841.
19. Royo, J.L., Valls, J., Acemel, R.D., Gómez-Marin, C., Pascual-Pons, M., Lupiañez, A., Gomez-Skarmeta, J.L., Fibla, J. (2018) A common copy-number variant within SIRPB1 correlates with human Out-of-Africa migration after genetic drift correction. *PLoS One*. 13, e0193614.
20. Laplana, M., Royo, J.L., García, L.F., Aluja, A., Gomez-Skarmeta, J.L., Fibla, J. (2014) SIRPB1 copy-number polymorphism as candidate quantitative trait locus for impulsive-disinhibited personality. *Genes Brain Behav.* 13, 653-662.

Table I. Primers used for the SIRPB1 allele-specific 3C.

	Sequence 5'-3'	Hg38	Distance from Anchor	Distance to the closest HindIII site
Anchor	CAATTGAGCTCTTCCTACCATGT	1,618,528	-	1,828 bp
Control	CAGCTGCTCCACTGACTGAG	1,618,368	160 bp	-
Sensor-79	GACATCCCAGCTAAGATGCAG	1,538,727	79,801 bp	65 bp
Sensor-57	TTCTCTTCCAGCATCCCATC	1,561,441	57,087 bp	62 bp
Sensor-52	CCTCCACTTACCCCATCTGA	1,566,328	52,200 bp	52 bp
Sensor-45	CATTCCGTATTTTTCTTTTCACAA	1,572,710	45,818 bp	70 bp

Figure captions.

Figure 1. Graphical representation of the allele-specific 3C determination strategy in a heterozygous subject. A) An anchor primer shall map the restriction fragment containing the minimal promoter, pointing to the closest restriction site (RE) while leaving the candidate SNP in between. A control "sensor" primer shall be designed to amplify the SNP while allowing a subsequent Taqman nested genotyping PCR. A series of additional sensor primers, mapping distant RE fragments will be individually assayed together with the anchor primer and the resulting 3C-amplified fragments. These fragments will be further genotyped in a nested qPCR. B) Box plots illustrating a deviation in the T/C ratio upon genotyping the anchor-"sensor 3" fragment from panel A). This bias can be interpreted as a more stable enhancer-promoter interaction associated to the T-allele (solid line) than the C-allele (dashed line).

Figure 2. Allele specific 3C of *SIRPB1* and rs2209313 using HindIII sites. Upon a conventional 3C, five different fragments were amplified using the anchor primer and each sensor. The relative position of each primer is highlighted in the UCSC genome browser (hg19) in the upper panel, showing the different *SIRPB1* splice variants and the ENCODE Histone 3 lysine 27 (H3K27) acetylation. FAM (rs2209313-T insertion-linked allele) and HEX (rs2209313-C deletion-linked allele) respective C_t were determined. Y axis represents $2^{\Delta Ct}$ being $\Delta C_t = Ct_{ins(T)} - Ct_{Del(C)}$. Asterisk represents U-Mann Whitney p-value <0.05 when compared to control.

Figure 3. Analysis of human chromosome 1 for allele-specific 3C determination strategy. Panel A shows the probability of capturing a TSS in a restriction fragment using one enzyme such as EcoRI (1); EcoRI or BamHI (2); EcoRI or BamHI or HindIII (3), and EcoRI or BamHI or HindIII or BglII (4). Panel B shows the distance between end of the restriction fragment containing a TSS and the nearest informative SNP.

Figure 1

Click here to download high resolution image

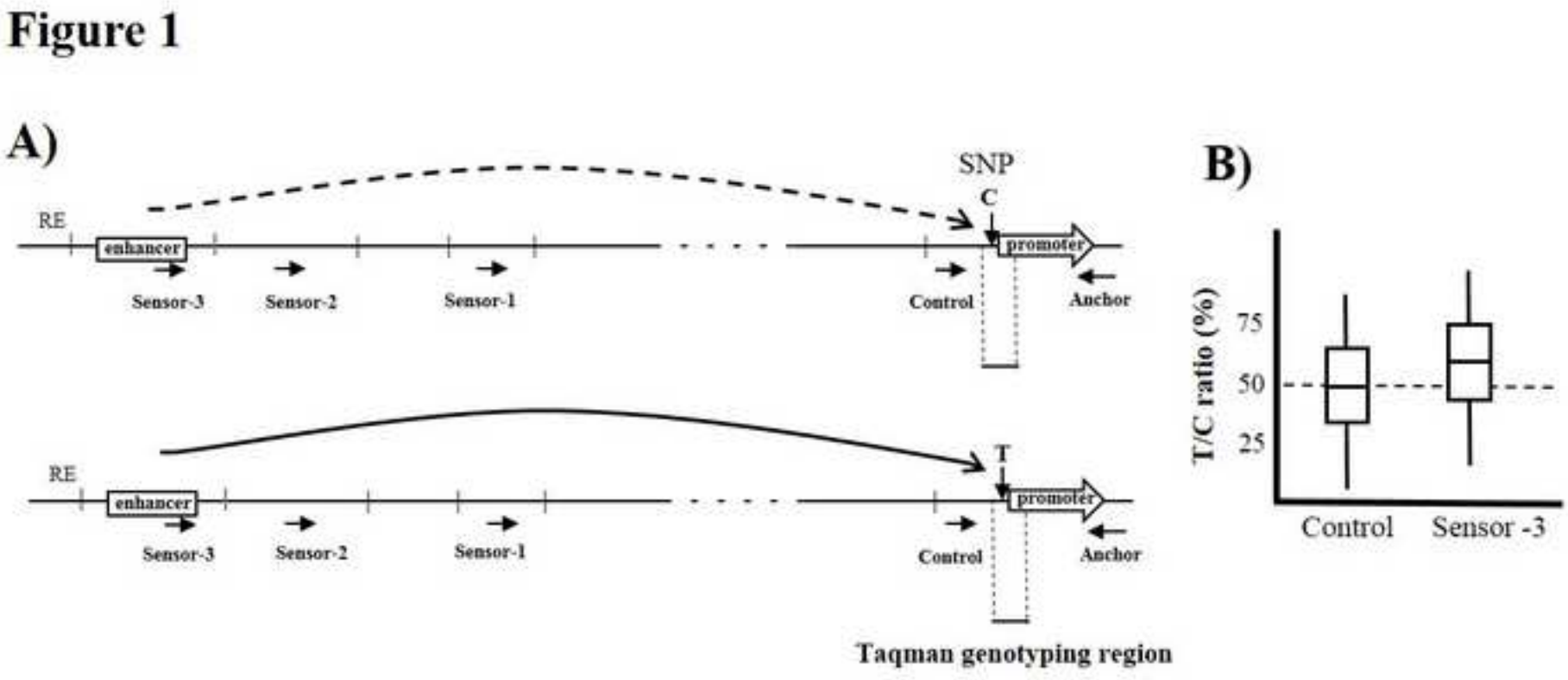

Figure 2

Click here to download high resolution image

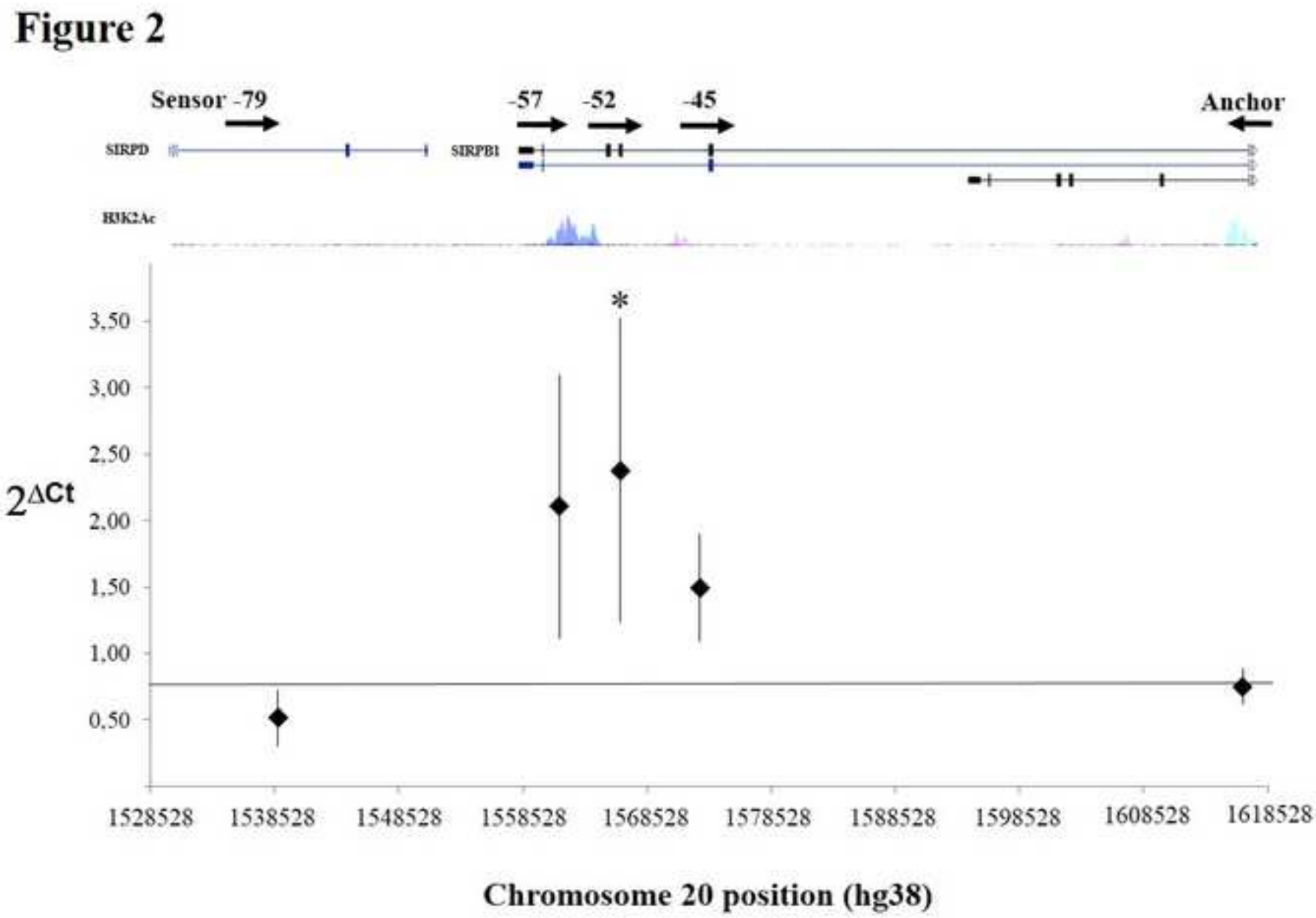

Figure 3

Click here to download high resolution image