Artificially inserted G-quadruplex DNA secondary structures induce long-distance chromatin activation

Shuvra Shekhar Roy[1,2], Sulochana Bagri[1,2], Avik Sengupta[3], Claudia Regina Then[4], Rahul Kumar[3], Sriram Sridharan[4], Shantanu Chowdhury[1,2*]

[1]CSIR-Institute of Genomics & Integrative Biology, Mathura Road, New Delhi, 110025, India

[2]Academy of Scientific & Innovative Research (AcSIR), Ghaziabad, 201002, India

[3]Department of Biotechnology, Indian Institute of Technology Hyderabad, Kandi, Telangana, 502284, India

[4]Cancer Science Institute of Singapore, National University of Singapore, 117599, Singapore

* To whom correspondence should be addressed. Email: shantanuc@igib.in

Abstract

Although the role of G-quadruplex (G4) DNA structures has been suggested in chromosomal looping this was not tested directly. Here, to test causal function, an array of G4s, or control sequence that does not form G4s, were inserted within chromatin in cells. Compared to the control insert, we observed a remarkable increase in the number of 3D chromatin looping interactions from the inserted G4 array. This was evident within the immediate topologically associated domain (TAD) and throughout the genome. Locally, recruitment of enhancer histone marks and the transcriptional coactivator p300/Acetylated-p300 increased in the G4-array, but not in the control insertion. Resulting promoter-enhancer interactions and gene activation were clear up to 5 Mb away from the insertion site. Together, these show the causal role of G4s in enhancer function and long-range chromatin interactions. Mechanisms of 3D topology are primarily based on DNA-bound architectural proteins that induce/stabilize long-range interactions. Involvement of the underlying intrinsic DNA sequence/structure in 3D looping shown here therefore throws new light on how long-range chromosomal interactions might be induced or maintained.

Introduction

G-quadruplexes (G4s), non-canonical DNA secondary structures with quartets of Guanines bonded by Hoogsteen base pairing, are instrumental in regulating gene expression (Sengupta et al., 2020; Varshney et al., 2020). G4s were primarily observed to be able to regulate gene expression when present around transcription start sites (TSSs) (Huppert and Balasubramanian, 2007; Rawal et al., 2006; Verma et al., 2008). G4s can regulate gene expression by directly

regulating recruitment of transcription factors and RNA polymerase or via alteration of DNA accessibility by modulating the epigenetic state of the gene promoters (Hussain et al., 2017; Kumar et al., 2011; Lago et al., 2021; Mukherjee et al., 2019; Saha et al., 2017; Sharma et al., 2021; Varshney et al., 2020). Recent studies have implicated the role of G4s in long-distance gene regulation (Robinson et al., 2021).

High-throughput chromosome conformation capture techniques reveal that specific regions of the human genome interact in three dimensions (3D) via chromatin looping and formation of topologically associated domains (TADs) (Bonev Boyan and Cavalli Giacomo, 2016; Denker and De Laat, 2016; Roy et al., 2018). Interestingly, recent computational studies observed enrichment of G4s in TAD boundaries along with higher enrichment of architectural proteins like CTCF and cohesin (Hou et al., 2019). Further, multiple studies noted the presence of G4s correlated with enhancer histone marks like H3K27Ac and H3K4Me1, and predominantly open chromatin regions (Calo and Wysocka, 2013; Hou et al., 2021; Shlyueva et al., 2014)

Although these studies implicate the role of G4s in long-range interactions and/or enhancer function, this was not directly tested. Here we asked if G4 structures might directly alter 3D chromatin, and affect long-range interactions including the epigenetic state of chromatin. To address this, we inserted an array of G4s into an isolated locus devoid of G4-forming sequences using CRISPR-Cas9 genome editing. To evaluate the specific function of G4s, a similar sequence of identical length but devoid of G4-forming capability was introduced. Using these pair of cell lines, we observed insertion of G4s specifically led to the recruitment of enhancer histone marks and increased expression of genes in a 10 Mb window. 3C and Hi-C results showed induced long-range interactions throughout the genome affecting topologically associated domains (TADs) that were specifically due to the incorporated G4s, and not found in case of the control insertion.

Results

Insertion of an array of G4s in an isolated locus

First, we sought to insert an array of G4s in a relatively isolated locus. We looked into Hi-C data from Rao et al., 2014 and identified a region that was markedly isolated with little or no interaction with its surrounding regions (as shown by snapshots of Hi-C interaction matrices obtained using the 3D genome browser (Wang et al., 2018) in Figure S1). In addition, this region was devoid of any G4s in the vicinity (no G4 forming motifs in a ±2.5 kb window).

Thereafter we artificially inserted an array of G4 forming sequences (275 bp long) at this region (chr12:79,872,423-79,872,424, hg19 genome assembly) using CRISPR-Cas9 genome editing (Figure 1A, S2). To study specific effects due to G4s, if any, a control sequence of identical length was inserted in HEK293T cells at the same locus where specific G/Cs necessary for G4 formation were substituted so that G4s are not formed by this sequence (Figure 1A, S2); we also ensured that the GC content was minimally affected by the substitutions (72.4% from 76.73%). Homozygous insertion was confirmed by PCR using primers adjacent to the insertion site followed by Sanger sequencing (Figure 1B, S3). The array of G4-forming sequences used for insertion was previously reported to form stable G4s in human cells (Lim et al., 2010; Monsen et al., 2020; Palumbo et al., 2009).

Chromatin epigenetic landscape upon insertion of G4s

To understand how the introduction of G4s altered the local chromatin, chromatin immunoprecipitation (ChIP) of different chromatin-modifying histone marks was done followed by qRT-PCR using primers spanning the inserted locus. The PCR primers were designed such that none of the primers bind to any site of G/C alteration in the mutated control insert. We observed significant increase in H3K4Me1 and H3K27Ac enhancer marks in the G4-array when compared to the G4-mutated control (Figures 2A, B). However, there was no G4-specific change in the presence of chromatin compaction marks, H3K27Me3 and H3K9Me3 (Figures 2C, D). The G4-dependent recruitment of H3K4Me1 (associated with enhancers (Heintzman et al., 2009, 2007)) and H3K27Ac (associated with active enhancers and promoters (Creyghton, 2010; Heintzman et al., 2009, 2007)) indicated enhancer-like characteristics of the inserted G4s.

The inserted G4-array acted as an enhancer

We next asked how the insertion of the G4-array influenced the expression of surrounding genes. To understand the distance-dependent gene regulatory impacts of the inserted G4-array, the mRNA expression of the nearest three genes and then some arbitrarily chosen genes further away up to 5 megabases (Mb) both up and downstream from the insertion site was quantified. Notably, the expression of four of the tested genes (PAWR, PPP1R12A, NAV3, and SLC6A15) increased in the G4-array insert compared to the mutated insert control cells (Figure 3A). Based on this enhanced expression, we further tested and observed a concomitant increase in the recruitment of Ser5 phosphorylated RNA Pol II in the surrounding gene promoters (Figure 3B). Next, we tested if chromosomal looping interactions between the insertion site and the gene

promoters were involved in these long-distance effects by using chromosome conformation capture (3C). The 3C assay between the insertion locus and the gene promoters could only be performed till the NAV3 promoter 1.6 Mb away. Beyond this distance, there was not any significantly detectable PCR amplification of 3C interaction products. The 3C assays revealed that there was a G4-dependent increase in chromosomal looping interactions of the insertion locus with the gene promoters (Figure 3C). These results suggested that the insertion of the G4-array led to the formation of an enhancer.

To understand the mechanism behind the enhancer-like property of the inserted G4-array we analyzed the recruitment of transcriptional coactivator p300 (Kalkhoven, 2004). There was an increase in recruitment of both p300 and as well as the more functionally active acetylated p300/CBP to the G4-array when compared against the mutated control (Figures 3D, E). Together, these results supported the enhancer-like function of the inserted G4-array.

LNA-mediated disruption of the inserted G4s reverses enhancer phenotype

To further establish that the enhancer effects upon the G4 array insertion are due to the formation of G4s, we wanted to see if some of the effects observed could be reversed upon disrupting the inserted structures. Specific Locked Nucleic Acid (LNA) probes were designed to target and disrupt the G4 using a similar approach as shown by others (Cadoni et al., 2021; Chowdhury et al., 2022; Kumar et al., 2008). Three probes were designed with stretches of mostly cytosines (Cs) as LNAs which would hybridize with stretches of guanines (Gs) in the G4-array insert important for the structure formation (Figure 4A; see Methods). We observed that there was a significant decrease in the expression of PPP1R12A and NAV3, two of the genes initially observed to have G4-dependent enhanced expression (Figure 3A), when the G4 array inserted cells were treated with the G4 targeting LNAs (Figure 4B). As expected, a decrease in the H3K4Me1 and H3K27Ac enhancer histone modifications was observed within the insert upon the LNAs treatment (Figures 4C, D). This indicated that the disruption of the inserted G4s can reverse some of the enhancer functions observed upon G4 insertion, further supporting the role of the G4 structure in the observed enhancer functions.

Domain-wide increase in looping interactions by G4s

For in-depth analysis of the long-range changes in chromatin architecture upon G4 insertion, we performed genome-wide interaction by Hi-C. First, we compared all the Hi-C contacts originating within a ±10 kb window comprising the G4-array insert, or the G4-mutated control insert. Compared to the mutated control, the G4-inserted locus had more than twice as many

genome-wide Hi-C interactions (6390 vs 3133) (Figures 5A, B, Supplementary Table 1). To rule out the possibility of artifacts due to the insertion we independently analyzed Hi-C data in HEK293T cells reported earlier (taken from GSE44267, Zuin et al., 2014). After normalizing for sequencing depth, the number of Hi-C contacts from the same window in HEK293T was relatively similar to the G4-mutated insert control (3968 and 3133 respectively, Figure 5C, Supplementary Table 1). Together, these showed that a significant number of new long-range interactions were induced throughout the genome due to the inserted G4s, but not from the inserted control sequence.

For closer analysis, we focused on intrachromosomal Hi-C interaction matrices of the G4-array insert, or the mutated control insert. This was centered on the insertion locus on chromosome 12 (chr12:7,80,72,423-8,16,72,423; insertion site marked with arrows in Figures 6A, B). The number of Hi-C interactions in the G4-array insert was clearly enriched compared to the G4-mutated insert control, as expected from the global Hi-C contacts noted above. We noted that while the interactions from the G4-array insert were significantly more, the insertion *per se* did not affect the overall domain architecture, which was largely similar between G4 or G4-mutated inserts as clear from Figures 6A and B. Further, we asked if the domain architecture was retained from that seen in HEK293T cells (with no insertion): Comparison using reported HiC data for the same region from HEK293T cells showed this to be the case confirming that the chromatin domain architecture remained relatively unchanged on introducing the G-array or G4-mutated regions (Figure S4).

To evaluate the effect of G4s in more detail, we plotted a Hi-C heatmap to show the enhanced or reduced (differential) contacts in the G4-array insert compared to the G4-mutated insert control cells (Figure 6C; relatively enriched/reduced contacts in the G4-array insert w.r.t. the G4-mutated insert plotted in red or blue, respectively; using Juicebox for analysis). This clearly showed that the G4-array induced significantly more Hi-C interactions; interestingly this was particularly evident in the downstream regions. For a closer analysis, we mapped the interaction frequency in a ±100 kb window centered on the insertion site. This clearly showed the difference in the number of interactions between the upstream regions vis-a-vis the region downstream of the insertion (Figure 6D).

To further confirm we used an independent HiC analysis method, HOMER (Hypergeometric Optimization of Motif EnRichment, Heinz et al., 2018) to compute the enhanced/reduced long-range interactions in the G4-array insert, compared to the control G4-mutated

insert. Differential analysis using HOMER showed that the inserted locus induced significantly higher number of interactions in the case of G4-array insert relative to the control G4-mutated case (Figure 6E). When we plotted the significantly different chromosomal interactions with minimum 20 interaction reads, it was again clear that the number of interactions with the G4-array insertion region was significantly enhanced in the downstream region relative to the upstream (Figure 6F).

Together these show a clear role of G4s in inducing long-range interactions. A similar sequence devoid of G4-forming capability did not induce such interactions. Furthermore, the overall nature of the TAD was not disturbed, and largely consistent with what is noted in cells with no insertion. Overall, these support that the insertion of G4s induced long-range interactions with minimal organizational changes in the 3D chromatin domain, underlining the molecular role of G4s in the arrangement of 3D chromatin.

A second significant feature was notable at the insertion locus. The number of induced long-range interactions was more significant downstream of the insertion site, compared to the upstream region (Figures 6C, D, F). A close look at the Hi-C contact matrices indicated that the site of insertion was very close and downstream to the TAD boundary (Figures 6A-C). We reasoned that the G4-dependent long-range interactions were largely within the TAD, and limited in the upstream region due to the TAD boundary. This is clearly seen in Figure 6C, akin to an 'architectural stripe' displaying that the inserted G4 array had enhanced Hi-C interactions across the domain, thus prominently featured in the downstream regions.

Discussion

To directly test if G4s affect long-range chromatin organization we artificially inserted an array of G4s in the chromatin. Hi-C experiments clearly showed an enhanced number of cis- and trans-chromosomal long-range interactions emanating from the introduced G4s. This was G4-specific because a similar sequence devoid of G4-forming capability introduced at the same site did not result in enhanced interactions. Furthermore, interestingly, most new long-range interactions following G4 incorporation were downstream from the site of insertion. This is likely because the G4 insertion locus was proximal to the upstream TAD boundary thereby restricting most new interactions to the downstream regions within the TAD (Figures 5, 6).

The insertion of the G4 array led to enhanced expression of genes up to 5 Mb away compared to cells with the G4 mutated control insertion. Furthermore, there was enrichment in the

H3K4Me1 and H3K27Ac enhancer histone marks, along with recruitment of transcriptional coactivator p300 and the more functionally active acetylated p300/CBP. This was clearly due to the introduction of G4s and not found upon the introduction of the G4-mutated control sequence. Moreover, the enhancer marks were reduced when the inserted G4s were specifically disrupted (Figure 4). Together, these directly support the function of G4s as enhancer elements and as factors that enhance long-range chromatin interactions. It is possible that such interactions are contextually dependent on the type of G4 structure, and/or the adjacent sequence context, and further studies will be necessary to elucidate these.

To ensure that the observed effects were from intracellular G4 formation we accounted for the following while designing the experiments. First, we introduced an array of G4s that were reported earlier to form stable G4s inside human cells; the sequence was from the hTERT promoter region with multiple arrayed G4s that have been well studied (Lim et al., 2010; Monsen et al., 2020; Palumbo et al., 2009). Second, we selected an insertion locus that was otherwise devoid of intrinsic G4s in a ±2.5 kb window. Third, the selected insertion locus was relatively sparse in long-range interactions. Fourth, we independently inserted a sequence of identical length (and similar GC%) which does not form G4s at the same locus (G4-mutated control). All results were compared to the G4-mutated insertion.

Existing literature shows promoter G4s are involved in regulating gene expression (Huppert and Balasubramanian, 2007; Rawal et al., 2006; Verma et al., 2008). Additionally, G4s have been reported to regulate chromatin epigenetics through both cytosine methylation and histone modifications (Halder et al., 2010; Mao et al., 2018; Sengupta et al., 2020). Previous studies by us further show that promoter G4s regulate gene expression by recruiting histone-modifying regulatory complexes (Hussain et al., 2017; Mukherjee et al., 2019; Saha et al., 2017; Sharma et al., 2021). Here we aimed to study how G4s affect the expression of genes far from their location, and if this was through G4-induced modifications in long-range 3D chromatin interactions.

Multiple studies have correlated the presence of G4s with long-range associations. CTCF, an architectural protein primarily involved in TAD boundary formation, was observed to bind to G4s and G4 stabilization was noted to enhance CTCF occupancy (Tikhonova et al., 2021). In addition, G4s were noted to be enriched in TAD boundaries and associated with the formation of chromatin loops (Hou et al., 2019). G4s were also found to coincide with open chromatin regions and H3K27Ac and H3K4Me1 ChIP-Seq peaks, which are markers for transcriptional

enhancers (Calo and Wysocka, 2013; Hou et al., 2021; Shlyueva et al., 2014). Most of these regions were observed to overlap with annotated enhancers, and promoters regulated by such enhancers were enriched in G4s (Williams et al., 2020). Furthermore, it was proposed that inter-molecular G4 formation between distant stretches of Gs may lead to DNA looping (Hegyi, 2015; Liano et al., 2022). The YY1 transcription factor was found to bind to G4s and dimerization of G4-bound YY1 led to chromatin looping interactions and consequent regulation of target gene expression (Li et al., 2020). A recent study also showed that R-loop-dependent G4 formation led to stronger CTCF binding (Wulfridge et al., 2023). In addition, it was shown that G4s assist in RNA polymerase II-associated chromatin looping (Yuan et al., 2023). Our findings here provide a direct molecular basis for the above observations by demonstrating a causal role of G4s in inducing both long-range associations and enhancer function. Although from one particular G4-forming stretch, and at one specific genomic locus, this illustrates the function of G4s in 3D gene regulation. Together these shed new mechanistic light on how DNA secondary structure motifs directly control the state of 3D chromatin and thereby biological function.

Materials and Methods

Cell lines and Cell Culture Conditions

HEK 293T cells were cultured in Dulbecco's Modified Eagle's Medium- High Glucose (DMEM-HG) supplemented with 10% FBS and 1XAnti-Anti (Gibco).

Primary Antibodies

Histone H3 rabbit polyclonal (Abcam ab1791), H3K4Me1 rabbit polyclonal (Abcam ab8895), H3K27Ac rabbit polyclonal (Abcam ab4729), H3K4Me3 mouse monoclonal (Abcam ab1012), H3K27Me3 mouse monoclonal (Abcam ab6002), H3K9Me3 rabbit polyclonal (Abcam ab8848), p300 rabbit monoclonal (CST 54062), Ac-p300/CBP rabbit polyclonal (CST 4771), TRF2 rabbit polyclonal (Novus NB110-57130).

Genomic Insertions using CRISPR-Cas9 genome editing

For the genomic insertions CRISPR-Cas9 genome editing technique was used (Ran et al., 2013). For the G4 array insertion, 275 bp long *hTERT* promoter region was PCR amplified from HEK 293T genomic DNA. For the insertion of the mutated G4s, a synthetic DNA template was synthesized and cloned into pUC57 vector by Genscript Biotech Corp, where 12 Gs were substituted with Ts (see supplementary methods for detailed sequences). Both the G4

array and the G4 mutated insertion templates were PCR amplified using longer primers where the short homology arms were introduced as overhangs of the primer for the accurate insertion at the 79M locus via homologous recombination (see supplementary methods for primer sequences) (Paix et al., 2017). For cleavage at the 79M locus (chr12:79,478,643-79,478,644 (hg38)), the gRNA sequence, 5'-ACTATGTATGTACATCCAGG-3', was cloned into the pX459 v2.0, a gift from Feng Zhang, that co-expresses cas9 protein and the gRNA. Guide RNAs (gRNAs) were designed using the CRISPOR tool (Haeussler et al., 2016). Once the gRNA cloned vector and the insertion donor templates were ready, they were transfected into HEK293T cells and the transfected cells were selected using puromycin, whose resistance gene was present in the pX459 vector. Then these selected cells were serially diluted to isolate clones originating from single cells. Many such clones were screened to detect cells with homozygous/heterozygous insertion of the G4 array or mutated G4 insert by performing locus-specific PCR. Either primers adjacent to the insertion site or cross primers, i.e., one primer within the insert and another from the adjacent region, were used to screen and identify insertions. While using adjacent primers, a shift in PCR product with an increase in amplicon size by 275 bp (size of the insert) indicated successful insertion (see supplementary methods for primer sequences).

ChIP (Chromatin Immunoprecipitation)

ChIP assays were performed as per the protocol previously reported in (Mukherjee et al., 2018). Immunoprecipitation was done using relevant primary antibodies. IgG was used for isotype control in all ChIP experiments. Three million cells were harvested and crosslinked with ~1% formaldehyde for 10 min and lysed. Chromatin was sheared to an average size of ∼250-500 bp using Biorupter (Diagenode). 10% of sonicated fraction was processed as input using phenol–chloroform and ethanol precipitation. ChIP was performed using 3 μg of the respective antibody incubated overnight at 4°C. Immune complexes were collected using herring sperm DNA-saturated magnetic protein G Dynabeads and washed extensively using a series of low salt, high salt and LiCl Buffers. The Dynabeads were then resuspended in TE (Tris- EDTA pH 8.1) buffer and treated with proteinase K at 65º C for ~5 hrs. Then, phenol-chloroform-isoamyl alcohol was utilized to extract DNA. Extracted DNA was precipitated by centrifugation after incubating overnight at -20 º C with isopropanol, 0.3M sodium acetate and glycogen. The precipitated pellet was washed with freshly prepared 70% ethanol and resuspended in TE buffer. ChIP DNA was analyzed by qRT-PCR method (see supplementary methods for primer sequences).

Real-time PCR for Gene (mRNA) expression:

Total RNA was isolated using TRIzol® Reagent (Invitrogen, Life Technologies) according to the manufacturer's instructions. RNA was quantified and cDNA was synthesized using iScript cDNA Synthesis Kits. A relative transcript expression level for genes was measured by quantitative real-time PCR using a SYBR Green based method (see supplementary methods for primer sequences). Average fold change was calculated by the difference in threshold cycles (Ct) between test and control samples. GAPDH gene was used as internal control for normalizing the cDNA concentration of each sample.

Chromosome Conformation Capture (3C)

Chromosome Conformation Capture (3C) assay was done as per the protocol reported in (Cope and Fraser, 2009) with certain modifications. Briefly, about 5-6 million cells were crosslinked using 1% formaldehyde for 10 minutes and then lysed to isolate the nuclei. Nuclei were digested overnight by HindIII and then ligated in a diluted reaction so that intramolecular ligation is favored. After ligation, the reaction mixture was treated with proteinase K at 65º C to de-crosslink the DNA, followed by RNase A treatment. Then, phenol-chloroform-isoamyl alcohol was utilized to extract DNA. Extracted DNA was precipitated by centrifugation after incubating overnight at -80 º C with 70% ethanol, 0.1M sodium acetate and glycogen. The precipitated pellet was washed with freshly prepared 70% ethanol and resuspended in TE buffer. 3C looping interactions were analyzed by TaqMan qRT-PCR method. For comparison, each interaction frequency was normalized to the interaction between exons 2 and 8 of the human α-actin (ACTA2)(Hadjur et al., 2009). See supplementary methods for primer sequences.

G4 disruption using LNA probes

Probes were designed to specifically bind to regions of genomic DNA containing G repeats which would form the G stems of the G4 structure. The probes containing LNA nucleotides should hybridize with the target with higher stability than the stability of the G4 structure thus destabilizing the G4. The probes used to target the G4 array insert were: 5'-**C*CC**GA**CCCC**<u>T**CC*C**</u>-3', 5'-**C*C**AG**CC**<u>**CCC**T</u>CC***G**-3', 5'-**C***<u>**CCC**T</u>**CCCC**TTC*C-3'. Stretches of three or more Cs are shown in bold, LNA nucleotides within the probes are underlined, the ends of the probes were protected using phosphorothioate bonds, shown as *. Approximately 0.8 µg of LNA probes (all 3 mixed in equimolar amounts) were transfected per million cells. Cells were treated with the LNA probes for 108 hours by transfecting thrice with

a gap of 36 hours in between. The schematic below shows the LNA probes designed to disrupt the inserted G4 structures along with the inserted G4 array sequence to show the specific sites of hybridization by the LNA probes.

CCAGGCC**GGG**CT**CCC**AGTGGATTCGC**GGG**CACAGACG**CCC**AGGACCGCGCTT**CCC**ACGTGGCG
GA**GGG**ACT**GGGG**A**CCCGGG**CA**CCC**GTCCTG**CCCC**TTCACCTTCCAGCTCCGCCTCCTCCGCGCG
GA**CCCC**G**CCCC**GT**CCC**GA**CCCC**T**CCC**GGGT**CCCC**GG**CC**AG**CC**CCCTCC**G**GGCCCT**CCC**AG**CC**
CCTCCCCTTCCTTTCCGCGG**CCCC**G**CCC**TCTCCTCGCGGCGCGAGTTTCAGGCAGCGCTGCGTC
CTGCTGCGCACGT**GGG**AAGCC
C*CCGACCCCTCC*C
C*CAGCCCCCTCC*G
C*CCCTCCCCTTC*C

Hi-C

Hi-C was performed using the Arima-HiC Kit as per the manufacturer's protocol. After the proximally-ligated Hi-C templates were generated, sequencing libraries were prepared using NEBNext Ultra II DNA Library Prep Kit as per the Arima-HiC Kit's protocol. The quality of the sequencing libraries was cross-checked using TapeStation (Agilent Technologies) and the KAPA Library Quantification Kit (Roche) before proceeding with sequencing using NovaSeq 6000 (Illumina).

Hi-C data analysis

Hi-C reads were mapped to the hg19 human genome and processed using default parameters using Juicer (https://github.com/ aidenlab/juicer). Hi-C count matrices were generated at 5kb, 10kb, 25kb, 50kb, 100kb, and 250kb using Juicer. Hi-C heatmap figures were rendered using Juicebox (https://github. com/aidenlab/Juicebox/wiki/Download). Hi-C contacts originating in the loci flanking the G4 insertion site were generated using bedtools (https://bedtools.readthedocs.io/en/latest/). The circos plots were rendered using Circos (http://circos.ca). To identify significant interaction the data was processed using homer (http://homer.ucsd.edu/homer/) using analyzeHiC function. The bins showing 2-fold enrichment in G4 WT over G4 Mut and vice-versa were retained for filtering contacts for representation on circos plots.

Data Availability

The sequencing data underlying this article are available in the NCBI Sequence Read Archive, scheduled to be released upon publication and would be accessible using the following link- https://www.ncbi.nlm.nih.gov/sra/?term=PRJNA1048044. The rest of the data are available in

the article and its online supplementary material. Further inquiries can be directed to the corresponding author.

Funding

This work was supported by The Wellcome Trust DBT India Alliance (IA/S/18/2/504021).

References

Bonev Boyan, Cavalli Giacomo. 2016. Organization and function of the 3D genome. *Nature Review Genetics* **17**:661–678. doi:10.1038/nrg.2016.112

Cadoni E, De Paepe L, Manicardi A, Madder A. 2021. Beyond small molecules: targeting G-quadruplex structures with oligonucleotides and their analogues. *Nucleic Acids Res* **49**:6638–6659. doi:10.1093/NAR/GKAB334

Calo E, Wysocka J. 2013. Modification of Enhancer Chromatin: What, How, and Why? *Mol Cell* **49**:825–837. doi:10.1016/j.molcel.2013.01.038

Chowdhury S, Wang J, Nuccio SP, Mao H, Di Antonio M. 2022. Short LNA-modified oligonucleotide probes as efficient disruptors of DNA G-quadruplexes. *Nucleic Acids Res* **50**:7247–7259. doi:10.1093/NAR/GKAC569

Cope NF, Fraser P. 2009. Chromosome Conformation Capture. *Cold Spring Harb Protoc* **2009**:pdb.prot5137. doi:10.1101/PDB.PROT5137

Creyghton MP. 2010. Histone H3K27ac separates active from poised enhancers and predicts developmental state. *Proc Natl Acad Sci USA* **107**:21931–21936.

Denker A, De Laat W. 2016. The second decade of 3C technologies: Detailed insights into nuclear organization. *Genes Dev* **30**:1357–1382. doi:10.1101/gad.281964.116

Hadjur S, Williams LM, Ryan NK, Cobb BS, Sexton T, Fraser P, Fisher AG, Merkenschlager M. 2009. Cohesins form chromosomal cis-interactions at the developmentally regulated IFNG locus. *Nature* **460**:410–413. doi:10.1038/nature08079

Haeussler M, Schönig K, Eckert H, Eschstruth A, Mianné J, Renaud JB, Schneider-Maunoury S, Shkumatava A, Teboul L, Kent J, Joly JS, Concordet JP. 2016. Evaluation of off-target and on-target scoring algorithms and integration into the guide RNA selection tool CRISPOR. *Genome Biol* **17**:1–12. doi:10.1186/s13059-016-1012-2

Halder R, Halder K, Sharma P, Garg G, Sengupta S, Chowdhury S. 2010. Guanine quadruplex DNA structure restricts methylation of CpG dinucleotides genome-wide. *Mol Biosyst* **6**:2439–2447. doi:10.1039/c0mb00009d

Hegyi H. 2015. Enhancer-promoter interaction facilitated by transiently forming G-quadruplexes. *Sci Rep* **5**:1–6. doi:10.1038/srep09165

Heintzman ND, Hon GC, Hawkins RD, Kheradpour P, Stark A, Harp LF, Ye Z, Lee LK, Stuart RK, Ching CW, Ching KA, Antosiewicz-Bourget JE, Liu H, Zhang X, Green RD, Lobanenkov V V., Stewart R, Thomson JA, Crawford GE, Kellis M, Ren B. 2009. Histone modifications at human enhancers

reflect global cell-type-specific gene expression. *Nature 2009 459:7243* **459**:108–112. doi:10.1038/nature07829

Heintzman ND, Stuart RK, Hon G, Fu Y, Ching CW, Hawkins RD, Barrera LO, Van Calcar S, Qu C, Ching KA, Wang W, Weng Z, Green RD, Crawford GE, Ren B. 2007. Distinct and predictive chromatin signatures of transcriptional promoters and enhancers in the human genome. *Nature Genetics 2007 39:3* **39**:311–318. doi:10.1038/ng1966

Heinz S, Texari L, Hayes MGB, Urbanowski M, Chang MW, Givarkes N, Rialdi A, White KM, Albrecht RA, Pache L, Marazzi I, García-Sastre A, Shaw ML, Benner C. 2018. Transcription Elongation Can Affect Genome 3D Structure. *Cell* **174**:1522-1536.e22. doi:10.1016/J.CELL.2018.07.047

Hou Y, Guo Y, Dong S, Yang T. 2021. Novel Roles of G-quadruplexes on Enhancers in human chromatin. *bioRxiv* 2021.07.12.451993. doi:10.1101/2021.07.12.451993

Hou Y, Li F, Zhang R, Li S, Liu H, Qin ZS, Sun X. 2019. Integrative characterization of G-Quadruplexes in the three-dimensional chromatin structure. *Epigenetics*. doi:10.1080/15592294.2019.1621140

Huppert JL, Balasubramanian S. 2007. G-quadruplexes in promoters throughout the human genome. *Nucleic Acids Res* **35**:406–413. doi:10.1093/nar/gkl1057

Hussain T, Saha D, Purohit G, Kar A, Kishore Mukherjee A, Sharma S, Sengupta S, Dhapola P, Maji B, Vedagopuram S, Horikoshi NT, Horikoshi N, Pandita RK, Bhattacharya S, Bajaj A, Riou JF, Pandita TK, Chowdhury S. 2017. Transcription regulation of CDKN1A (p21/CIP1/WAF1) by TRF2 is epigenetically controlled through the REST repressor complex. *Sci Rep*. doi:10.1038/s41598-017-11177-1

Kalkhoven E. 2004. CBP and p300: HATs for different occasions. *Biochem Pharmacol* **68**:1145–1155. doi:10.1016/j.bcp.2004.03.045

Kumar N, Patowary A, Sivasubbu S, Petersen M, Maiti S. 2008. Silencing c-MYC expression by targeting quadruplex in P1 promoter using locked nucleic acid trap. *Biochemistry* **47**:13179–13188. doi:10.1021/BI801064J/SUPPL_FILE/BI801064J_SI_001.PDF

Kumar Pankaj, Yadav VK, Baral A, Kumar Parveen, Saha D, Chowdhury S. 2011. Zinc-finger transcription factors are associated with guanine quadruplex motifs in human, chimpanzee, mouse and rat promoters genome-wide. *Nucleic Acids Res*. doi:10.1093/nar/gkr536

Lago S, Nadai M, Cernilogar FM, Kazerani M, Domíniguez Moreno H, Schotta G, Richter SN. 2021. Promoter G-quadruplexes and transcription factors cooperate to shape the cell type-specific transcriptome. *Nature Communications 2021 12:1* **12**:1–13. doi:10.1038/s41467-021-24198-2

Li L, Williams P, Ren W, Wang MY, Gao Z, Miao W, Huang M, Song J, Wang Y. 2020. YY1 interacts with guanine quadruplexes to regulate DNA looping and gene expression. *Nat Chem Biol*. doi:10.1038/s41589-020-00695-1

Liano D, Monti L, Chowdhury S, Raguseo F, Di Antonio M. 2022. Long-range DNA interactions: inter-molecular G-quadruplexes and their potential biological relevance. *Chemical Communications* **58**:12753–12762. doi:10.1039/D2CC04872H

Lim KW, Lacroix L, Yue DJE, Lim JKC, Lim JMW, Phan AT. 2010. Coexistence of two distinct G-quadruplex conformations in the hTERT promoter. *J Am Chem Soc* **132**:12331–12342. doi:10.1021/ja101252n

Mao SQ, Ghanbarian AT, Spiegel J, Martínez Cuesta S, Beraldi D, Di Antonio M, Marsico G, Hänsel-Hertsch R, Tannahill D, Balasubramanian S. 2018. DNA G-quadruplex structures mold the DNA methylome. *Nat Struct Mol Biol*. doi:10.1038/s41594-018-0131-8

Monsen RC, DeLeeuw L, Dean WL, Gray RD, Sabo TM, Chakravarthy S, Chaires JB, Trent JO. 2020. The hTERT core promoter forms three parallel G-quadruplexes. *Nucleic Acids Res* **48**:5720–5734. doi:10.1093/NAR/GKAA107

Mukherjee AK, Sharma S, Chowdhury S. 2019. Non-duplex G-Quadruplex Structures Emerge as Mediators of Epigenetic Modifications. *Trends Genet* **35**:129–144. doi:10.1016/j.tig.2018.11.001

Mukherjee AK, Sharma S, Sengupta S, Saha D, Kumar P, Hussain T, Srivastava V, Roy SD, Shay JW, Chowdhury S. 2018. Telomere length-dependent transcription and epigenetic modifications in promoters remote from telomere ends. *PLoS Genet* **14**. doi:10.1371/journal.pgen.1007782

Paix A, Folkmann A, Goldman DH, Kulaga H, Grzelak MJ, Rasoloson D, Paidemarry S, Green R, Reed RR, Seydoux G. 2017. Precision genome editing using synthesis-dependent repair of Cas9-induced DNA breaks. *Proceedings of the National Academy of Sciences* 201711979. doi:10.1073/pnas.1711979114

Palumbo SML, Ebbinghaus SW, Hurley LH. 2009. Formation of a unique end-to-end stacked pair of G-quadruplexes in the hTERT core promoter with implications for inhibition of telomerase by G-quadruplex-interactive ligands. *J Am Chem Soc*. doi:10.1021/ja902281d

Ran FA, Hsu PD, Wright J, Agarwala V, Scott DA, Zhang F. 2013. Genome engineering using the CRISPR-Cas9 system. *Nat Protoc* **8**:2281–2308. doi:10.1038/nprot.2013.143

Rao SSP, Huntley MH, Durand NC, Stamenova EK, Bochkov ID, Robinson JT, Sanborn AL, Machol I, Omer AD, Lander ES, Aiden EL. 2014. A 3D map of the human genome at kilobase resolution reveals principles of chromatin looping. *Cell* **159**:1665–1680. doi:10.1016/j.cell.2014.11.021

Rawal P, Kummarasetti VBR, Ravindran J, Kumar N, Halder K, Sharma R, Mukerji M, Das SK, Chowdhury S. 2006. Genome-wide prediction of G4 DNA as regulatory motifs: Role in Escherichia coli global regulation. *Genome Res* **16**:644–655. doi:10.1101/gr.4508806

Robinson J, Raguseo F, Nuccio SP, Liano D, Di Antonio M. 2021. DNA G-quadruplex structures: more than simple roadblocks to transcription? *Nucleic Acids Res* **49**:8419–8431. doi:10.1093/NAR/GKAB609

Roy SSSS, Mukherjee AKAK, Chowdhury S. 2018. Insights about genome function from spatial organization of the genome. *Hum Genomics* **12**:8. doi:10.1186/s40246-018-0140-z

Saha D, Singh A, Hussain T, Srivastava V, Sengupta S, Kar A, Dhapola P, Dhople V, Ummanni R, Chowdhury S. 2017. Epigenetic suppression of human telomerase (hTERT) is mediated by the metastasis suppressor NME2 in a G-quadruplex– dependent fashion. *Journal of Biological Chemistry* **292**:15205–15215. doi:10.1074/jbc.M117.792077

Sengupta A, Roy SS, Chowdhury S. 2020. Non-duplex G-Quadruplex DNA Structure: A Developing Story from Predicted Sequences to DNA Structure-Dependent Epigenetics and Beyond. *Acc Chem Res* **54**:46–56. doi:10.1021/ACS.ACCOUNTS.0C00431

Sharma S, Mukherjee AK, Roy SS, Bagri S, Lier S, Verma M, Sengupta A, Kumar M, Nesse G, Pandey DP, Chowdhury S. 2021. Human telomerase is directly regulated by non-telomeric TRF2-G-quadruplex interaction. *Cell Rep* **35**:109154. doi:10.1016/J.CELREP.2021.109154

Shlyueva D, Stampfel G, Stark A. 2014. Transcriptional enhancers: From properties to genome-wide predictions. *Nat Rev Genet* **15**:272–286. doi:10.1038/nrg3682

Tikhonova P, Pavlova I, Isaakova E, Tsvetkov V, Bogomazova A, Vedekhina T, Luzhin A V., Sultanov R, Severov V, Klimina K, Kantidze OL, Pozmogova G, Lagarkova M, Varizhuk A. 2021. Dna g-quadruplexes contribute to ctcf recruitment. *Int J Mol Sci* **22**:7090. doi:10.3390/IJMS22137090/S1

Varshney D, Spiegel J, Zyner K, Tannahill D, Balasubramanian S. 2020. The regulation and functions of DNA and RNA G-quadruplexes. *Nat Rev Mol Cell Biol* **21**:459–474. doi:10.1038/s41580-020-0236-x

Verma A, Halder K, Halder R, Yadav VK, Rawal P, Thakur RK, Mohd F, Sharma A, Chowdhury S. 2008. Genome-wide computational and expression analyses reveal G-quadruplex DNA motifs as conserved cis-regulatory elements in human and related species. *J Med Chem* **51**:5641–5649. doi:10.1021/jm800448a

Wang Y, Song F, Zhang B, Zhang L, Xu J, Kuang D, Li D, Choudhary MNK, Li Y, Hu M, Hardison R, Wang T, Yue F. 2018. The 3D Genome Browser: A web-based browser for visualizing 3D genome organization and long-range chromatin interactions. *Genome Biol* **19**:1–12. doi:10.1186/S13059-018-1519-9/FIGURES/5

Williams JD, Houserova D, Johnson BR, Dyniewski B, Berroyer A, French H, Barchie AA, Bilbrey DD, Demeis JD, Ghee KR, Hughes AG, Kreitz NW, McInnis CH, Pudner SC, Reeves MN, Stahly AN, Turcu A, Watters BC, Daly GT, Langley RJ, Gillespie MN, Prakash A, Larson ED, Kasukurthi M V., Huang J, Jinks-Robertson S, Borchert GM. 2020. Characterization of long G4-rich enhancer-associated genomic regions engaging in a novel loop:loop “G4 Kissing” interaction. *Nucleic Acids Res* **48**:5907–5925. doi:10.1093/nar/gkaa357

Wulfridge P, Yan Q, Rell N, Doherty J, Jacobson S, Offley S, Deliard S, Feng K, Phillips-Cremins JE, Gardini A, Sarma K. 2023. G-quadruplexes associated with R-loops promote CTCF binding. *Mol Cell* **83**:3064-3079.e5. doi:10.1016/J.MOLCEL.2023.07.009

Yuan J, He X, Wang Y. 2023. G-quadruplex DNA contributes to RNA polymerase II-mediated 3D chromatin architecture. *Nucleic Acids Res* **51**:8434–8446. doi:10.1093/NAR/GKAD588

Zuin J, Dixon JR, Van Der Reijden MIJA, Ye Z, Kolovos P, Brouwer RWW, Van De Corput MPC, Van De Werken HJG, Knoch TA, Van Ijcken WFJ, Grosveld FG, Ren B, Wendt KS. 2014. Cohesin and CTCF differentially affect chromatin architecture and gene expression in human cells. *Proc Natl Acad Sci U S A* **111**:996–1001. doi:10.1073/PNAS.1317788111/-/DCSUPPLEMENTAL/SAPP.PDF

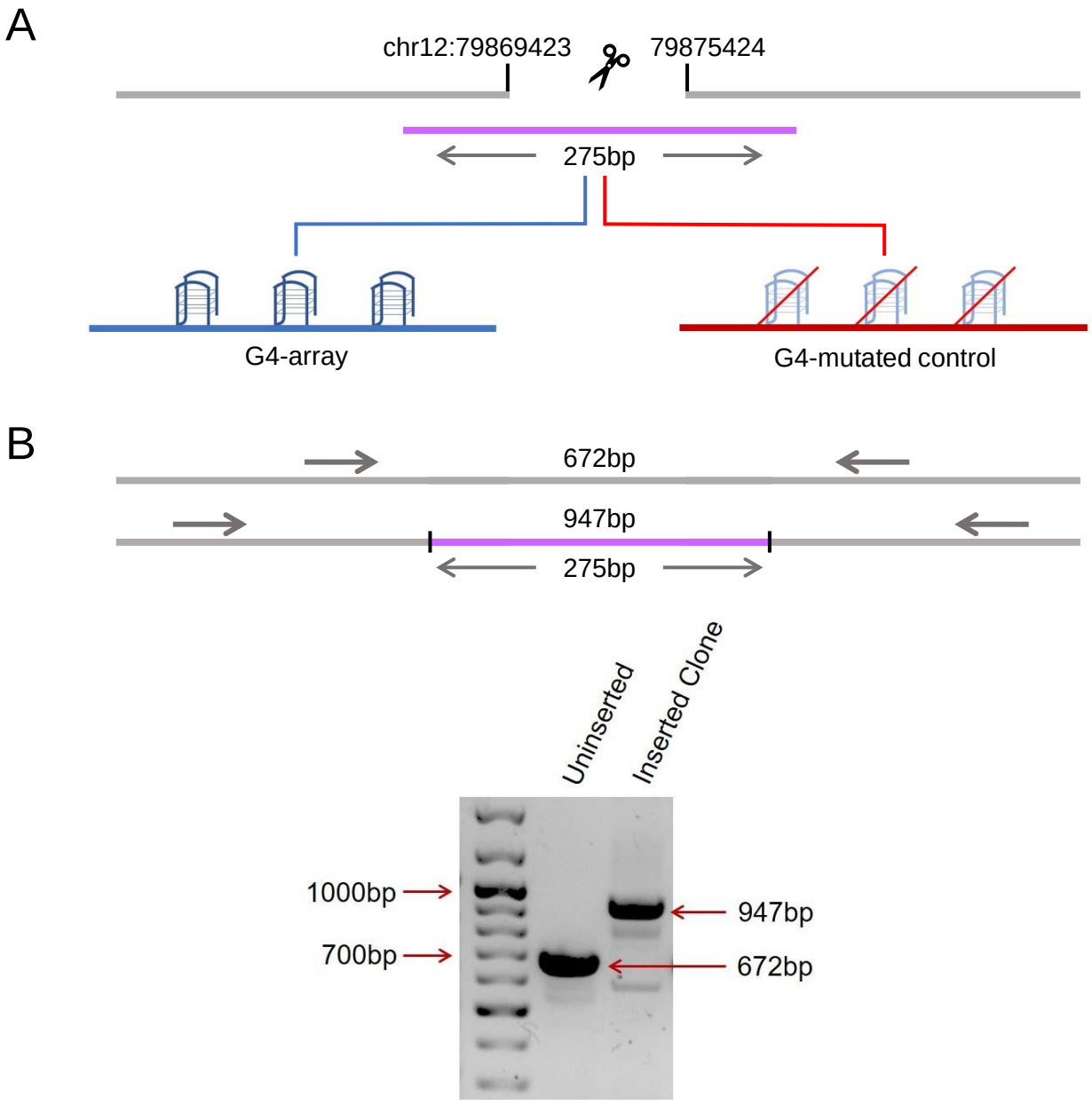


Figure 1: Insertion of an array of G4s in an isolated locus. (A) Schematic showing the insertion of the G4-array and the G4-mutated control at chr12:79,872,423-79,872,424 (hg19). (B) PCR of the insertion locus showing the successful insertion of the 275bp long insert sequence.

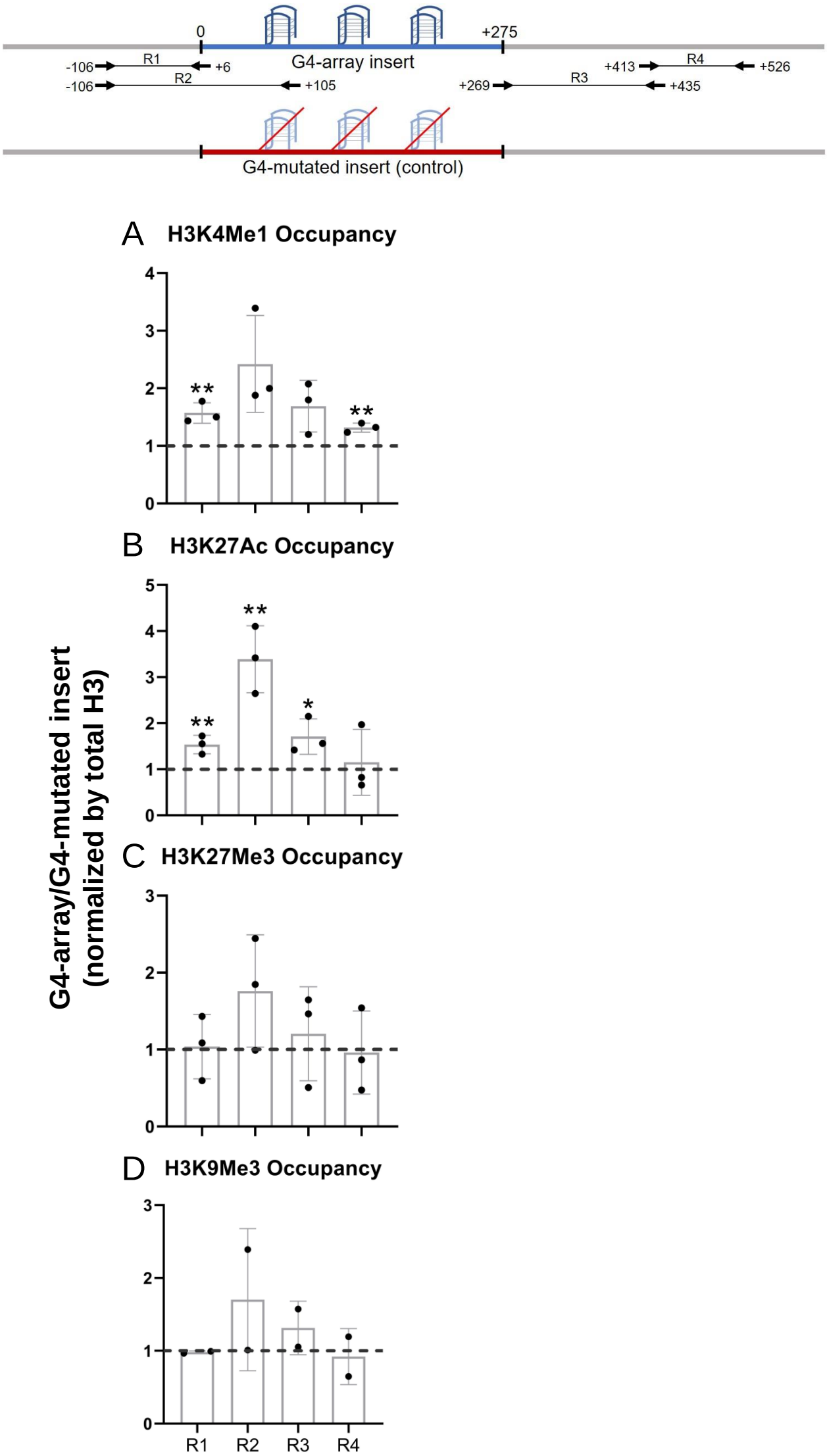


Figure 2: Changes in chromatin upon G4-array insertion. The top panel shows the positions of the PCR amplicons used in the Histone ChIP experiments. Changes in chromatin-modifying histone modifications in the insert region represented by calculating the ratio of occupancy of different histone marks in the G4-array insert cells over the G4-mutated insert (control) cells- enhancer mark, H3K4Me1 (A); active enhancer/promoter mark, H3K27Ac (B); facultative repressor mark, H3K27Me3 (C) and constitutive repressor mark, H3K9Me3 (D). Mean ± SD (n=2/3); unpaired, two-tailed *t*-test.

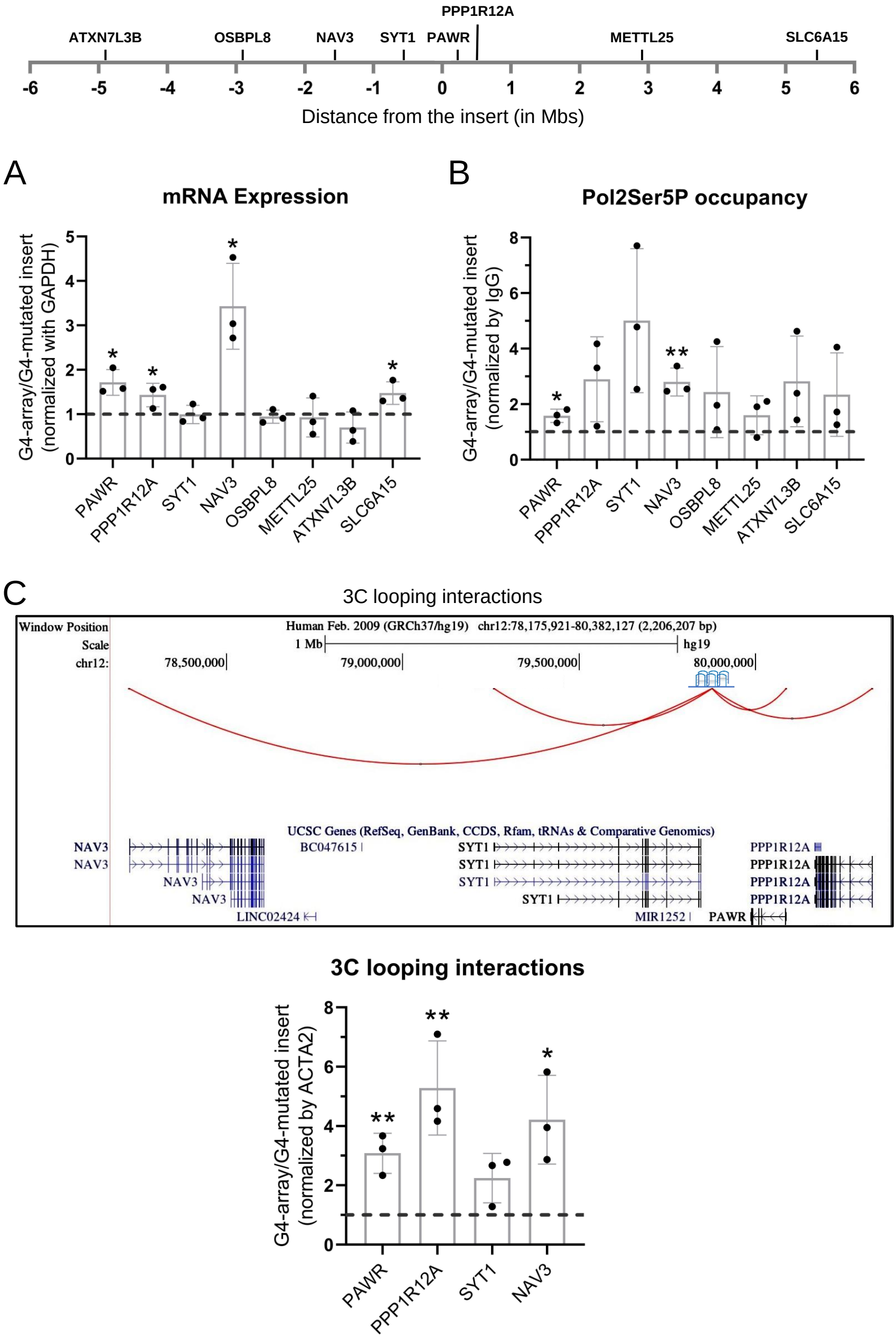
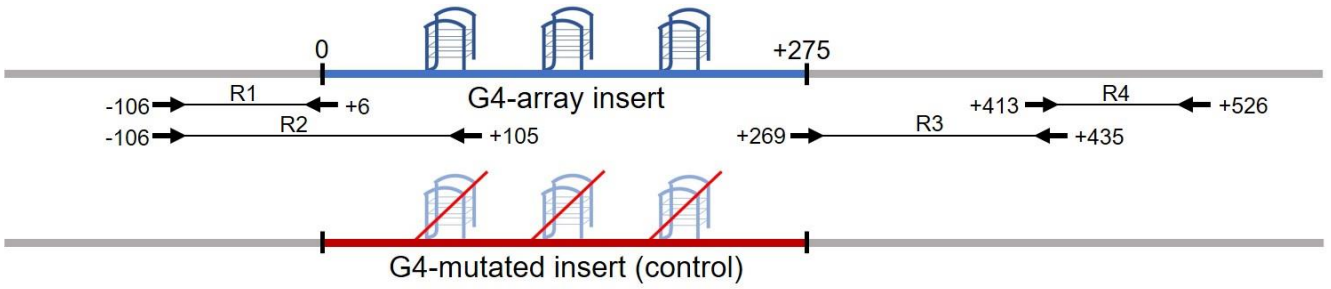


Figure 3: Insertion of the G4-array led to enhancer function. (A) Long-range G4-dependent changes in mRNA expression are represented by calculating the ratio of expression of surrounding genes in the G4-array insert cells over the G4-mutated insert (control) cells. Top panel shows the positions of the gene promoters with respect to the insertion site. (B) Ratio of Pol2 Phospho-Ser5 Occupancy at the promoters of the surrounding genes in the G4-array insert cells over the G4-mutated insert (control) cells. (C) A UCSC genome browser snapshot showing the 3C looping interactions between the insertion and the surrounding gene promoters. Fold change in 3C looping interactions between the insertion and the surrounding gene promoters in the G4-array insert cells over the G4-mutated insert (control) cells.



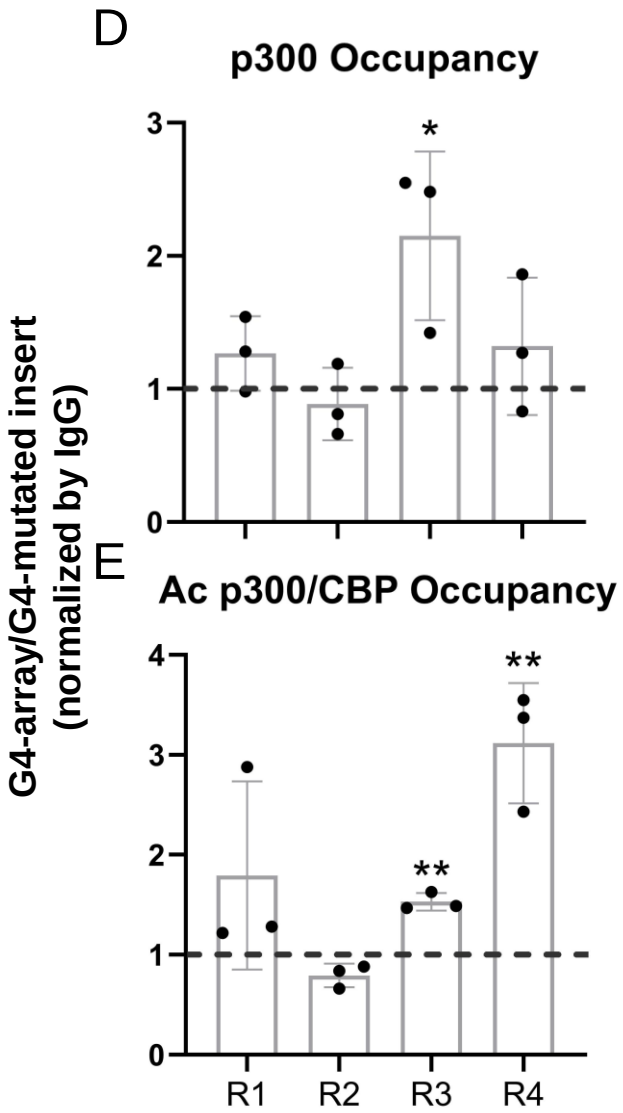


Figure 3: Insertion of the G4-array led to enhancer function. Top panel shows the positions of the PCR amplicons used in the ChIP experiments. The ratio of occupancy of p300 (D) and Ac p300/CBP (E) in the G4-array insert cells over the G4-mutated insert (control) cells. Mean ± SD (n=3); unpaired, two-tailed *t*-test.

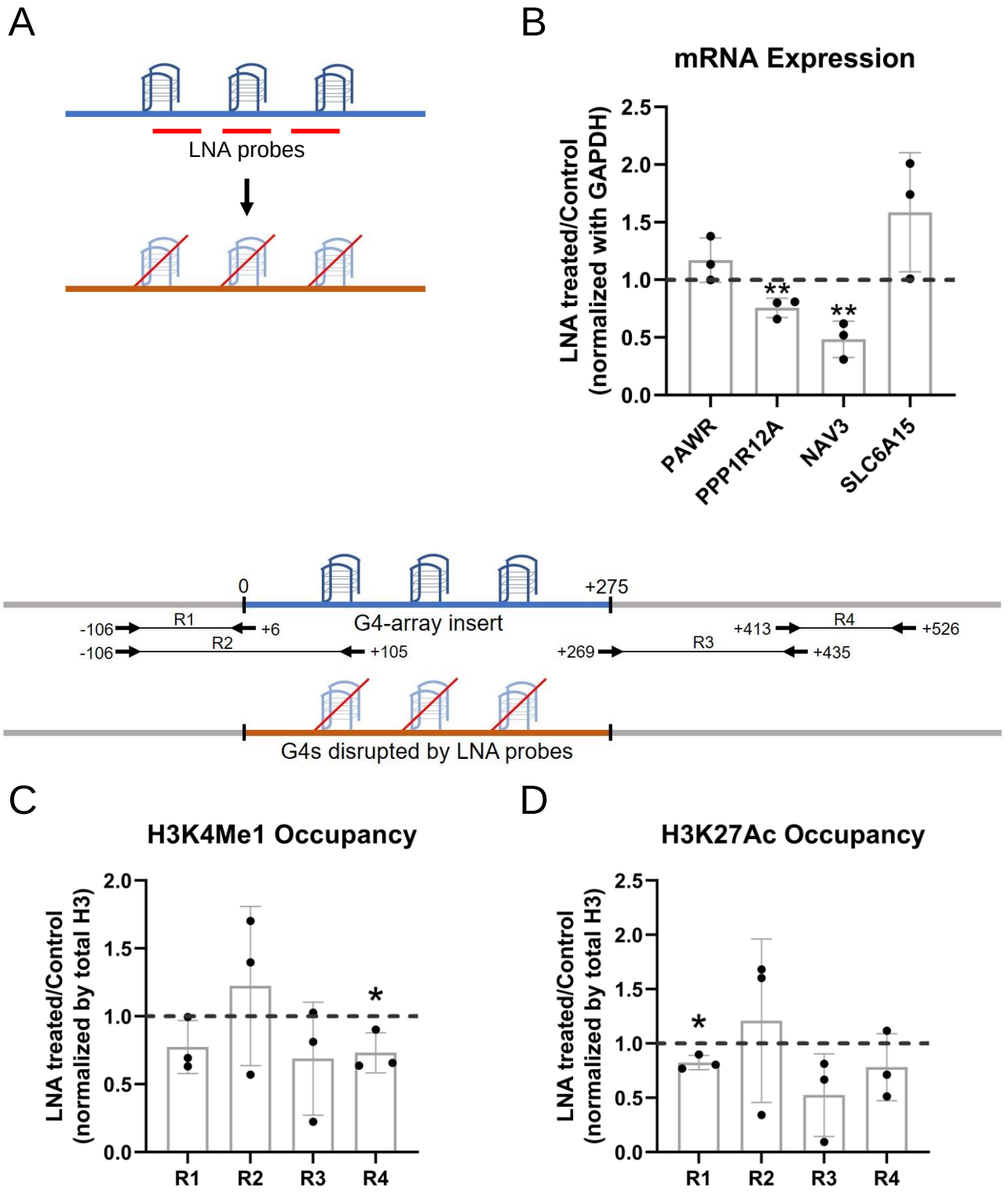
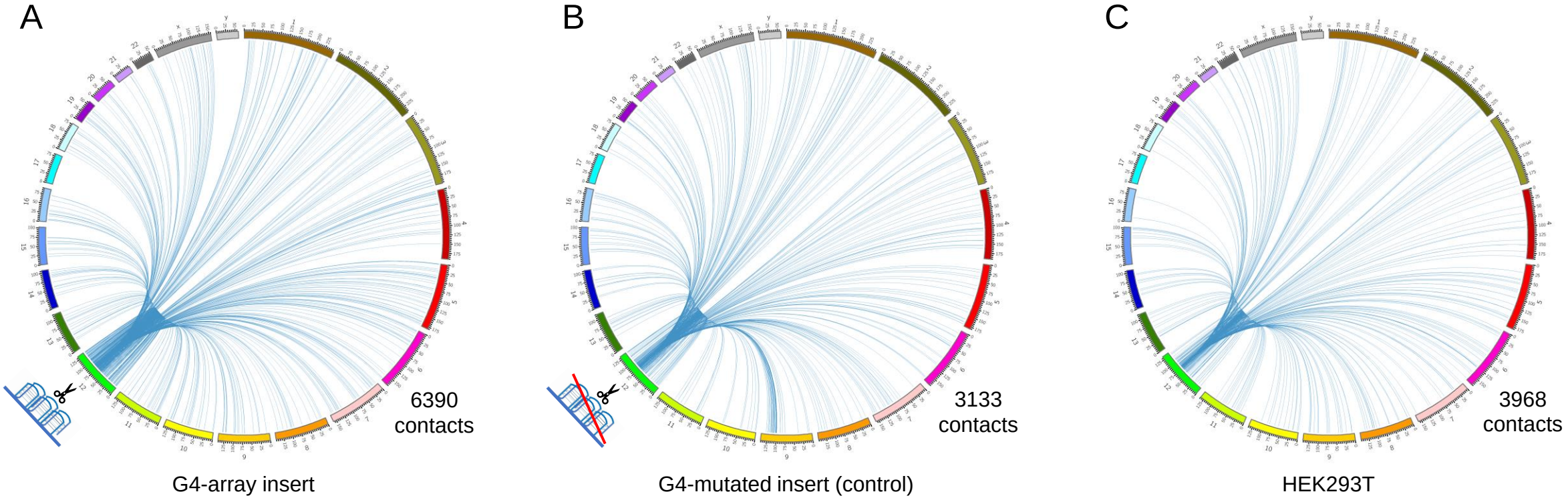


Figure 4: LNA-mediated disruption of the inserted G4s reverses enhancer phenotype. (A) Schematic showing the inserted G4 structures disrupted using LNA probes (details in methods). (B) Effects of LNA treatment in the G4-array insert cells on the expression of surrounding genes which showed enhanced expression when compared against the G4-mutated insert (control) cells in Figure 3A- represented by the ratio of expression of surrounding genes in the LNA-treated over the vehicle-treated (control) cells. Effects of LNA treatment in the G4-array insert cells on the levels of H3K4Me1 (C) and H3K27Ac (D) histone modifications at the insert locus represented by the ratio of occupancy of the histone marks in the LNA-treated over the vehicle-treated (control) cells. The top panel shows the positions of the PCR amplicons used in the Histone ChIP experiments. Mean ± SD (n=3); unpaired, two-tailed t-test.



D

	G4-array insert	G4-mutated insert (control)	HEK293T
Total Hi-C contacts	649,455,101	692,798,603	366,957,555
Normalized contacts from insertion site	6390	3133	3968

Figure 5: Insertion of the G4-array increased Hi-C interactions. Circos plots showing raw Hi-C contacts across the genome originating from a ±10 kb window with the insertion site at the middle across 3 samples- (A) G4-array insert cells, (B) G4-mutated insert (control) cells and (C) HEK293T control cells (taken from GSE44267). (D) Table showing the number of genome-wide raw Hi-C contacts and normalized contacts (normalized against the total raw Hi-C contacts to normalize for the sequencing depth) originating from the ±10 kb window with the insertion site at the middle across the 3 samples.

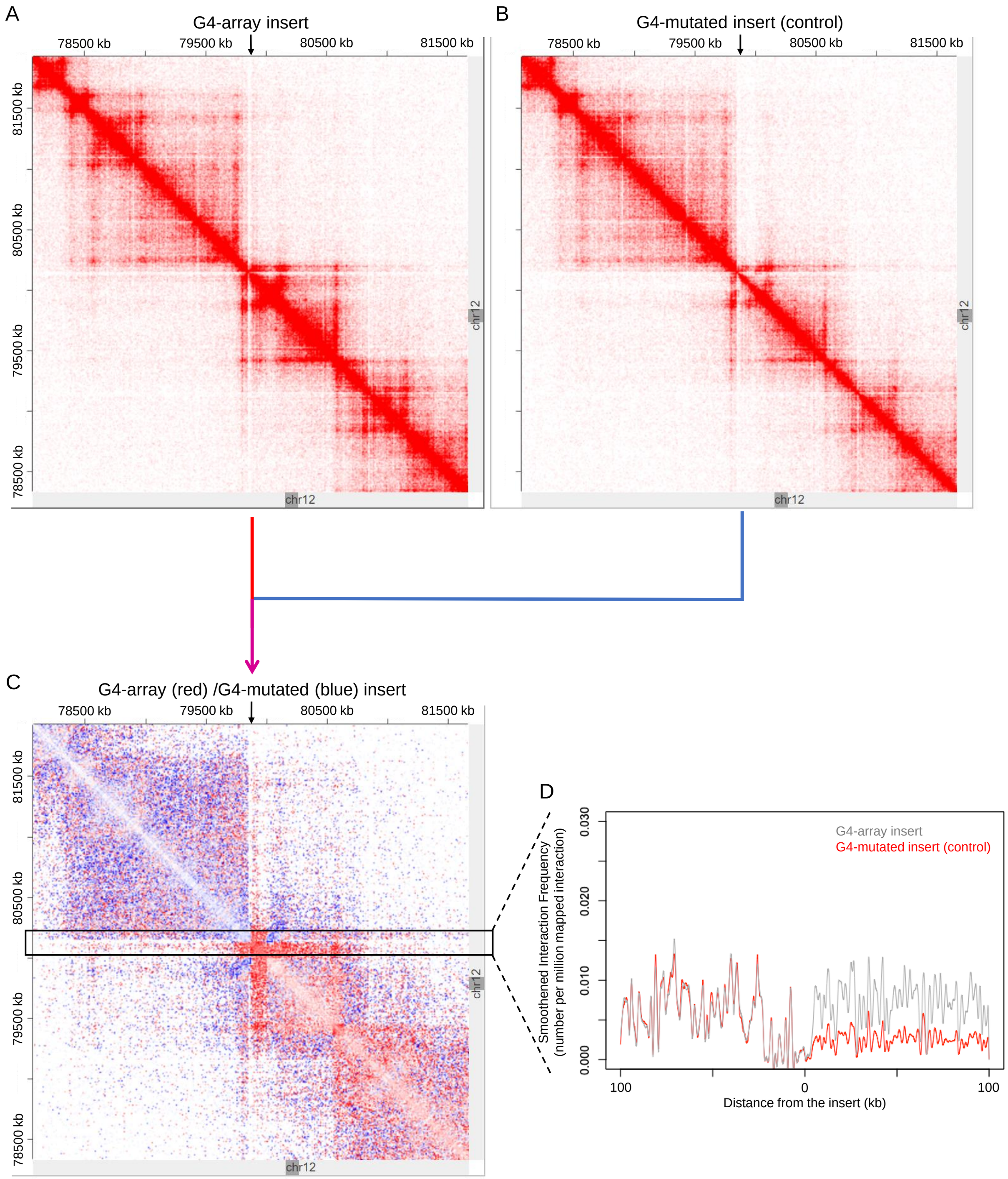
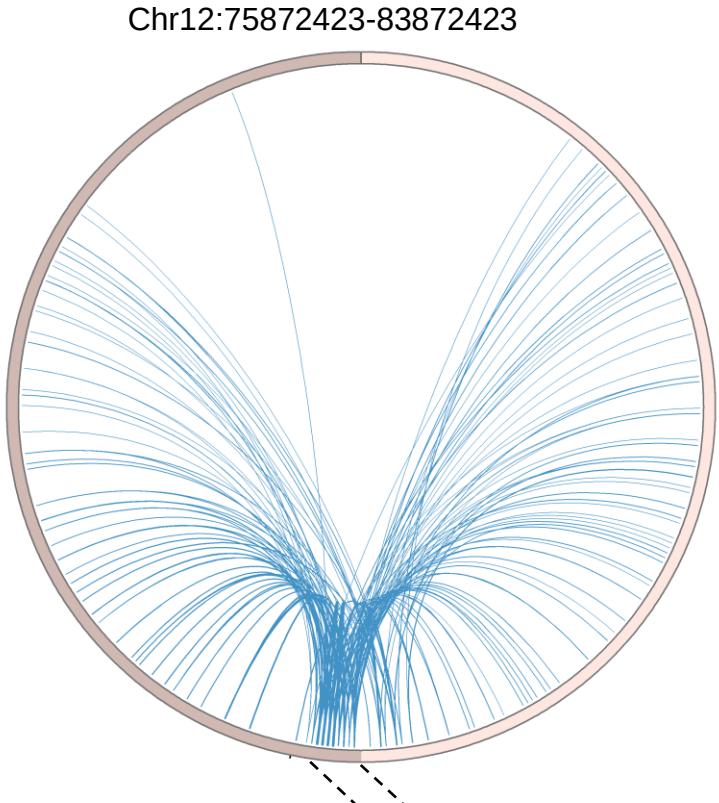


Figure 6: G4-dependent changes in local chromatin architecture. Juicebox Hi-C matrices showing Hi-C contacts in the (A) G4-array insert cells, (B) G4-mutated insert (control) cells in a 3.6Mb region of chromosome 12 with the insertion site at the middle of the matrices. The arrows at the top of the Hi-C matrices indicate the site of insertion. (C) Juicebox Hi-C matrix showing normalized Hi-C contacts in the G4-array insert cells over the G4-mutated insert (control) cells as a heatmap. The region of interest (i.e., interactions associated with the immediate vicinity of the insert) is marked with a box. The arrow at the top of the Hi-C matrix indicates the site of insertion. (D) A line histogram displaying the differences in interaction frequency across G4-array insert cells and G4-mutated insert (control) cells in regions up to 100 kb away from the insertion site. As seen interactions downstream of the insertion site are more enriched than upstream in the G4-array insert cells as compared to the G4-mutated control.

E



F

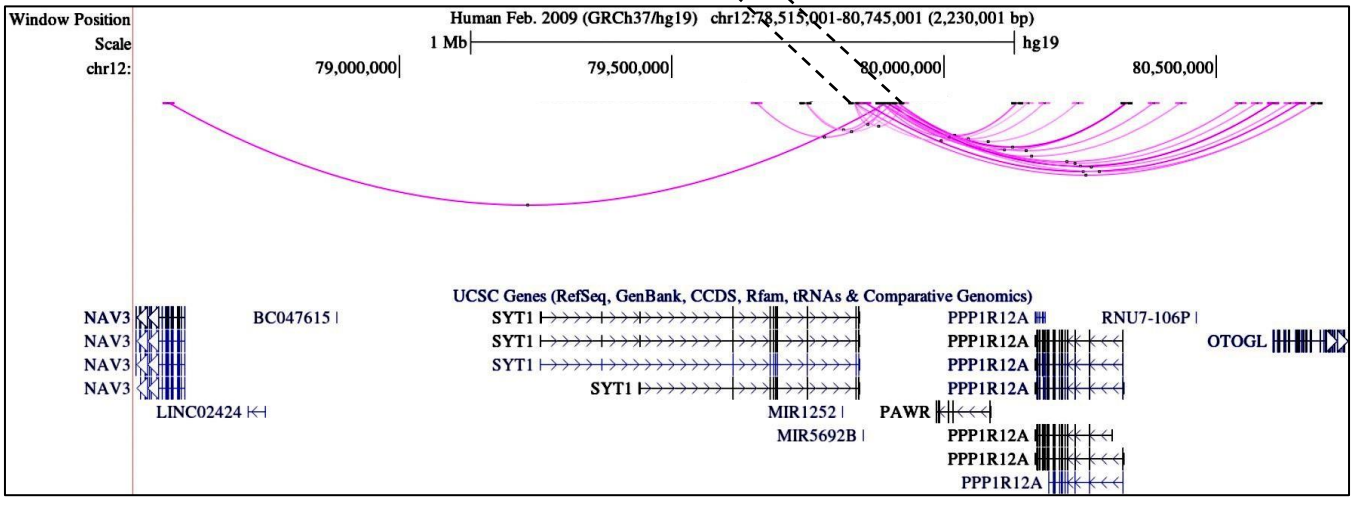


Figure 6: G4-dependent changes in local chromatin architecture. (E) Circos plot showing differential interactions (fold enrichment >=2) originating from a ± 100 kb window with the insertion site at the middle, in the G4-array insert cells over the G4-mutated insert (control) cells. (F) A UCSC genome browser snapshot showing the more significant differential interactions (fold enrichment >=2, interaction reads >20) originating from a ± 50 kb window with the insertion site at the middle, in the G4-array insert cells over the G4-mutated insert (control) cells. The color intensity of the arcs indicating the interacting bins is proportional to the fold enrichment.

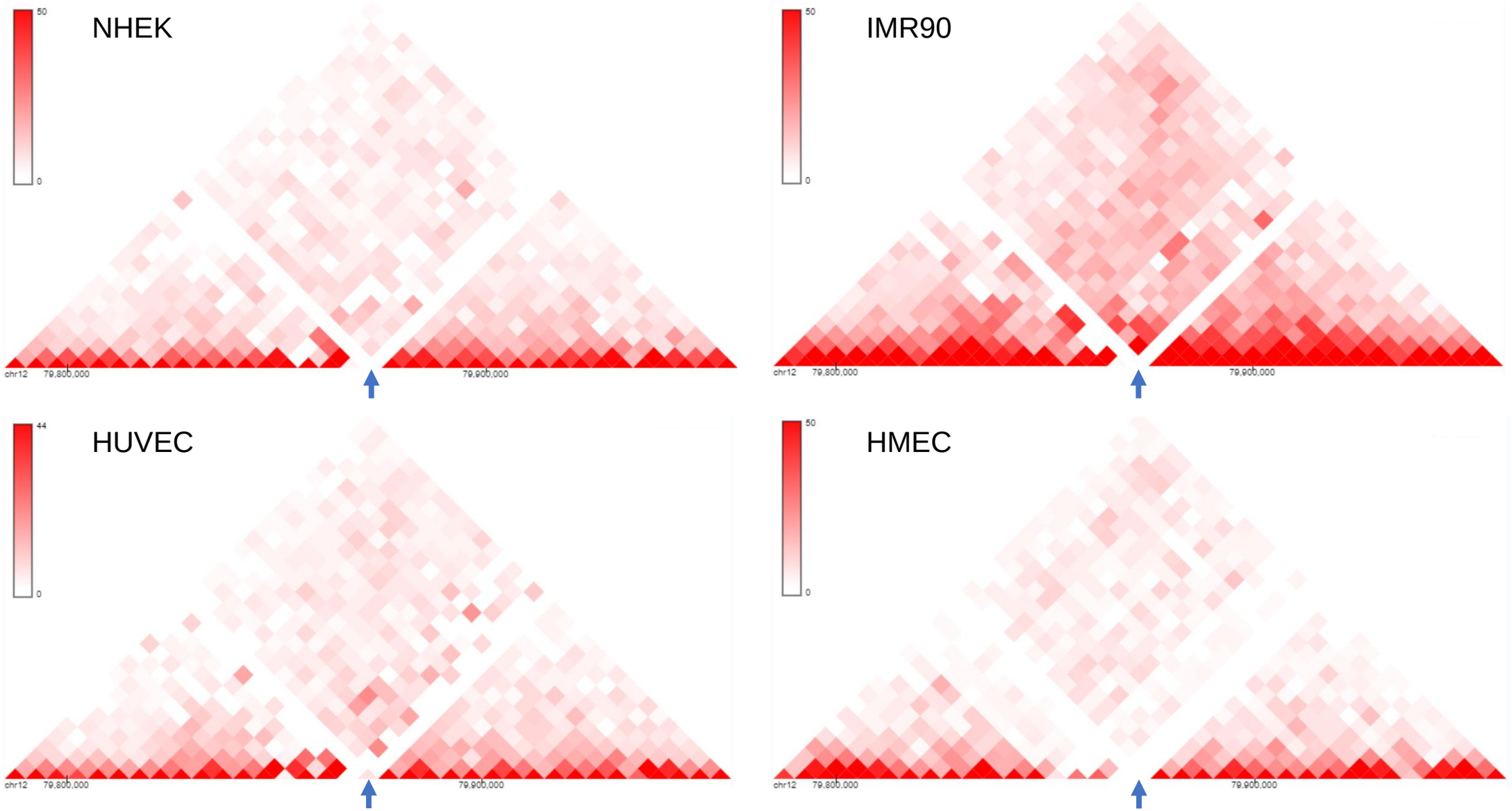


Figure S1: An isolated locus was chosen for insertion. (A) 3D Genome Browser (Wang et al., 2018) snapshots showing the Hi-C interaction matrices from 4 cell lines (NHEK, IMR90, HUVEC and HMEC) of a section of chromosome 12 with chr12:79,870,000-79,875,000 (hg19) locus in the middle (marked by arrows); the chr12:79,870,000-79,875,000 locus has very low interaction with its surrounding loci, indicated by the lack of red dots or very faint red dots.

G4-array

CCAGGCC**GGG**CT**CCC**AGTGGATTCGC**GGG**CA
CAGACG**CCC**AGGACCGCGCTT**CCC**ACGTGGC
GGA**GGG**ACT**GGGG**A**CCCGGG**CA**CCC**GTCCT
G**CCCC**TTCACCTTCCAGCTCCGCCTCCTCCG
CGCGGA**CCCC**G**CCCC**GT**CCC**GA**CCCC**T**CC**
GGGT**CCC**GG**CCC**AG**CCCCC**TCC**GGGCCC**T
CCCAG**CCCC**T**CCCC**TTCCTTTCCGCGG**CCCC**
G**CCC**TCTCCTCGCGGCGCGAGTTTCAGGCAG
CGCTGCGTCCTGCTGCGCACGT**GGG**AAGCC

G4-mutated control

CCAGGCC**GGG**CT**CCC**AGTGGATTCGC**GTG**CA
CAGACG**CCC**AGGACCGCGCTT**CCC**ACGTGGC
GGA**GTG**ACT**GGTG**A**CCCGGG**CA**CAC**GTCCTG
CCCCTTCACCTTCCAGCTCCGCCTCCTCCGCG
CGGA**CCCC**G**CCCC**GT**CCC**GA**CACC**T**CACGTG**
T**CACC**GG**CAC**AG**CCACC**TCC**GGGCCC**T**CCC**A
G**CACC**T**CCAC**TTCCTTTCCGCGG**CCCC**G**CCC**
TCTCCTCGCGGCGCGAGTTTCAGGCAGCGCT
GCGTCCTGCTGCGCACGT**GGG**AAGCC

Figure S2: Insert sequences. Sequences of the G4 array and the G4-mutated array that were inserted; three or more runs of G/Cs (stem of a potential G4) are shown in bold and the G/Cs (marked in blue) that were changed to T/As (marked in red) in the G4-mutated array are marked.

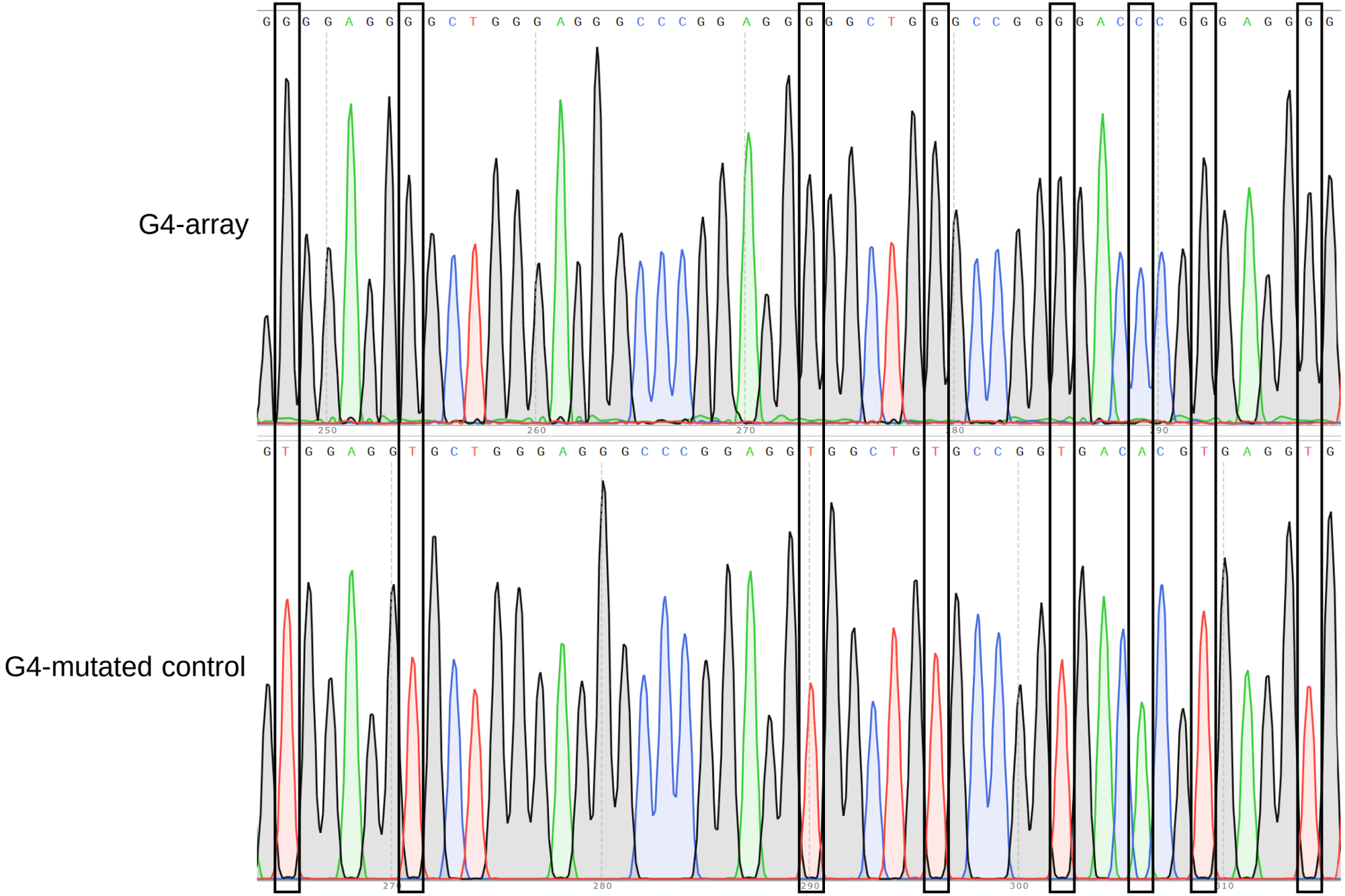


Figure S3: Insertion confirmed by Sanger sequencing. Representative snapshots of the Sanger sequencing chromatogram of the reverse strand of the insertion locus PCR product showing some of the specific G/C to T/A substitutions in the insert.

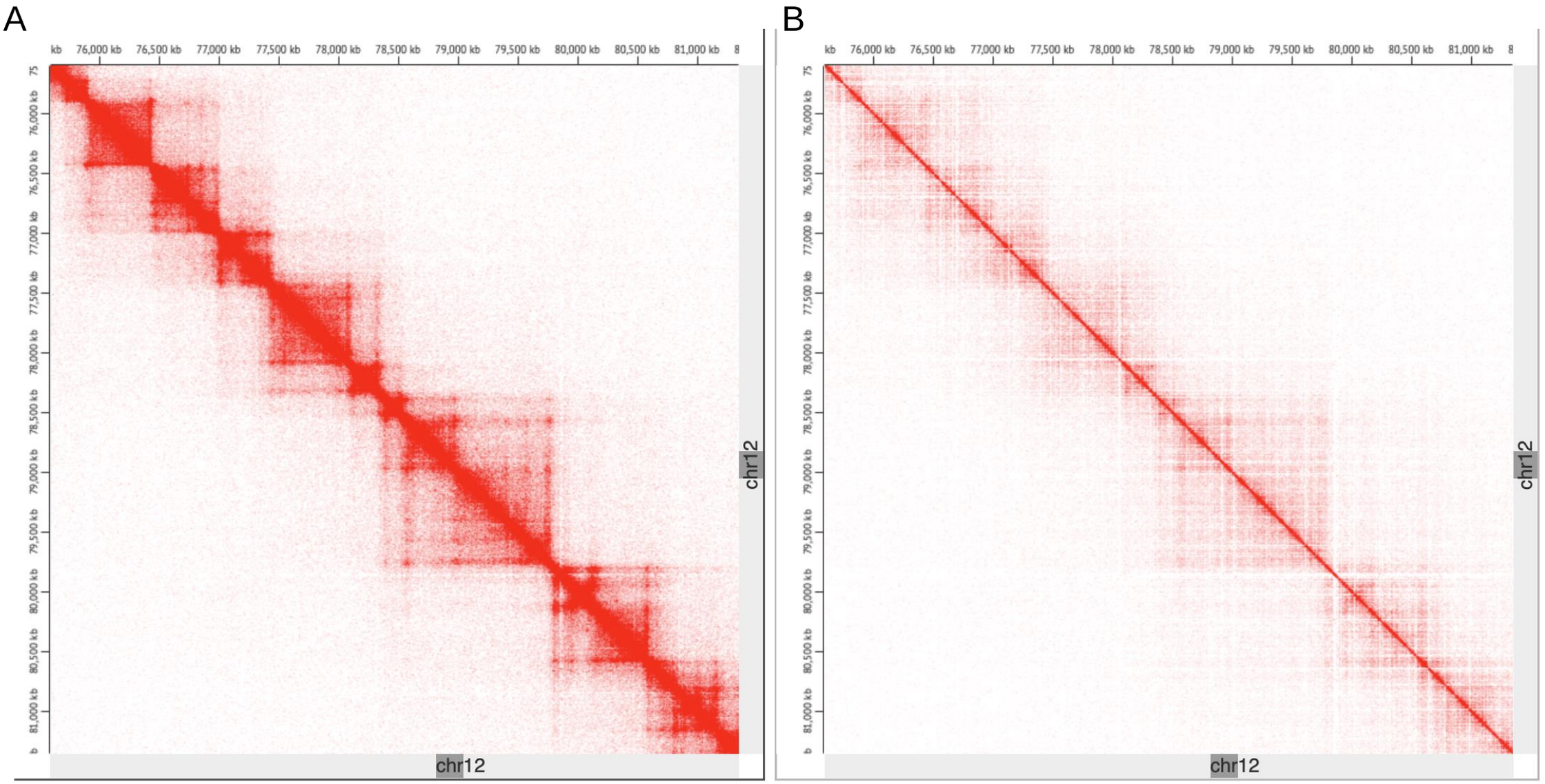


Figure S4: The chromosomal architecture of the insertion locus in the G4-array insert cells is broadly similar to uninserted cells except for the increase in looping interactions. Comparison of Hi-C contact matrices around the insertion site in the G4-array insert cells (A) and the HEK293T control cells (taken from GSE44267) (B) shows that the broad chromatin organization is conserved. The TADs appear to be otherwise unaltered upon insertion.

	G4-array insert	G4-mutated insert (control)	HEK293T
Total Hi-C contacts	649,455,101	692,798,603	366,957,555
Normalized contacts from insertion site	6390	3133	3968
Actual contacts from insertion site (±10 kb, i.e. 20kb window)	5990	3133	2102
Mean Hi-C contacts across 10,000 random 20kb windows across the genome	8910	9490	5021
Standard deviation of Hi-C contacts across 10,000 random 20kb windows across the genome	4078	4716	2806
Z-score for the actual contacts from the insertion site	-0.71	-1.34	-1.04

Supplementary Table 1: Comparative analysis of the G4-dependent increase in Hi-C interactions upon insertion. Table showing the number of genome-wide raw Hi-C contacts, actual number of contacts originating from the ±10 kb window with the insertion site at the middle and the comparative analysis (mean, standard deviation and z-score) of these number of contacts with contacts across 10,000 random 20kb windows across the genome across the 3 samples.