Unique for human centromeric regions of interphase chromatin homing (CENTRICH) facilitate clustering of long-range chromosome interaction sites and govern dynamic features of chromatin fractal globules

Gennadi V. Glinsky

Sanford-Burnham Medical Research Institute, 10901 North Torrey Pines Road, La Jolla, CA 92037

Abstract

The fractal globule geometry is the most efficient means to fold interphase chromatin chains in the small restricted nuclear space without entanglements to facilitate the unfolding and rapid expansion of chromatin chains. Current fractal globule models of the interphase chromatin geometry lack considerations of the interactions of chromatin chains folded into fractal globules with major external to globules nuclear structures. These interactions are critically important because they define the 3D boundaries of the fractals in the nucleus and restrict the volume of nucleoplasmic space accessible for chromatin polymers. Here we report the results of the external point of contact analysis of chromatin fractal globules based on a genome-wide alignment of inter- and intra-chromosomal chromatin interactions within the context of interphase chromatin binding to nuclear lamina and nucleolus. For all human chromosomes, a significant correlation exists between binding of chromosomal loci to nuclear lamina and segregation into spatially-defined distinct compartments of genome-wide chromatin interactions identified by Hi-C method. We report identification of near-centromeric intergenic regions on human chromosomes (chr2; chr10; chr17; chr1), which are highly enriched for interphase chromatin homing sites and function as attractors of long-range physical interactions (Centromeric Regions of Interphase Chromatin Homing, CENTRICH). CENTRICH are engaged in 397-1526 pair-wise interactions per 1 Kb distance, which represents 199 - 716-fold enrichment of interphase chromatin homing sites compared to genome-wide average (2-tail Fisher's exact test p values range $2.10E^{-101}$ - $1.08E^{-292}$). CENTRICH represent unique for human highly homologous DNA sequences of 3.9 – 22.4 Kb in size which are: 1) associated with nucleolus; 2)

exhibit remarkably diverse regulatory protein contexts of chromatin state maps; 3) bind multiple intergenic disease-associated genomic loci (IDAGL) with documented long-range enhancer activities and established links to increased risk of developing epithelial malignancies and other common human disorders. For cancer, coronary artery disease, and type 2 diabetes, there is a statistically significant inverse correlation between the genome-wide association studies (GWAS)-defined odds ratios of increased risk of a disease and distances of SNP loci homing sites to the middle-point genomic coordinates of CENTRICH. We conclude that interactions with nucleolus and nuclear lamina may have a physiologically and pathologically significant global impact on 3D genome architecture by governing a dynamic transition of large segments of interphase human chromosomes from the open loop to folded fractal globule conformations. We propose that emergence of CENTRICH in genome's architecture during the evolution contributes to distinct phenotypic features of *H.sapiens*.

Introduction

Most of regulatory events affecting genomic architecture and functions of eukaryotic cells during physiological and pathological conditions occur during the interphase. However, experimental observations highlighting detailed mechanisms of structural-functional alignment and compartmentalization of genomic functions in interphase nuclei remain highly fragmentary. Pioneering genome-wide analyses were focused on a limited subset of experimental variables selected from nuclear, genetic, and chromatin components with already established or suspected structural-functional significance. In interphase nucleus, the three-dimensional conformation of chromosomes is defined by both chromatin folding and chromatin polymer interactions with major external to chromatin globules nuclear structures such as nuclear lamina and nucleoli (1-4). These interactions are likely to influence the compartmentalization of the nuclear space into functionally-related compartments which brings widely separated genetic loci into close, spatially-defined proximity. Understanding how chromosomes align within the three-dimensional space of interphase nucleus can provide insight into the complex relationships between chromatin architecture, regulation of gene expression, and the functional state of the cell.

Recently, a genome-wide, high-resolution analysis of DNA sequences that have a high probability to associate endogenously and copurify with nucleoli (nucleolar-associated domains [NADs]) in human cells, has been carried-out using a combination of fluorescence comparative genome hybridization (CGH), deep DNA sequencing and fluorescence photoactivation (1, 2). It has been noted that the size distribution (0.1–10 Mb) and median sequence length (749 kb) of NADs were very similar to lamina-associated domains, LADs (0.1–10 Mb, 553 kb) suggesting that the major architectural units of chromosome organization within the mammalian interphase nucleus are about 0.5–1 Mb in length (1). Unexpectedly, careful analysis of deep sequencing data detected a significant overlap between the NADs and loci previously reported to associate with the nuclear envelope (LADs; ref. 3). It has been suggested that because both the LAD and NAD profiles were generated using the

high-throughput analysis of large cell populations, the apparent overlap in loci may indicate that specific regions could alternate (shuttle) between binding to the nucleolus and the nuclear lamina either in different cells, or at different times within the same cells (1). Genome-wide analysis of interactions of interphase chromatin with nuclear lamina in mouse embryonic stem cells revealed that nuclear lamina-genome interactions are broadly involved in the control of gene expression programs during lineage commitment and terminal differentiation of embryonic stem cells (4). Therefore, recent experiments demonstrate that interactions of interphase chromosomes with nuclear lamina and nucleoli are critically important in defining a 3D architecture of chromatin organization and functions. Expression of genes located in both lamina-associated and nucleoli-associated chromatin compartments is predominantly repressed. Transcription at genetic loci that are located in chromatin compartments that are not associated with either nuclear lamina or nucleoli is often activated and expression of corresponding genes is increased.

Long-range interactions between specific pairs of loci have been evaluated using multiple methods. Chromosome Conformation Capture (3C) utilizes spatially constrained ligation followed by locus-specific PCR. Adaptations of 3C have extended the process with the use of inverse PCR (4C) or multiplexed ligation-mediated amplification (5C). A genome-wide method of the analysis of long-range chromatin interactions named Hi-C was reported that extends the above approaches to enable purification of ligation products followed by massively parallel sequencing (5). Therefore, Hi-C allows unbiased identification of chromatin interactions across an entire genome and provides strong support for the fractal globule model of interphase chromatin folding.

Hi-C method revealed the fractal globule conformation of the interphase chromatin folding and the existence of two independent compartments of genome-wide chromatin interactions along the chains of chromatin polymers which were designated compartment A and compartment B. The compartments were defined based on the enrichment of contacts between DNA sequences within each compartment and depletion of contacts between sequences in different compartments.

Compartment B showed a higher interaction frequency, lower overall DNaseI accessibility, and lower gene expression, indicating that compartment B represents more densely packed heterochromatin. In contrast, compartment A showed a relatively lower interaction frequency, higher overall DNaseI accessibility, and increased gene expression, indicating that compartment A represents less densely packed, transcriptionally active chromatin. The Hi-C-defined chromatin compartments corresponded to spatially separated regions of chromosomes as confirmed by fluorescence in situ hybridization (FISH). Plotting of these compartments linearly along the length of each chromosome using the eigenvector index values revealed the intermitting chromosome-specific patterns with enrichment in compartment A as positive values and enrichment in compartment B as negative values (5). We noted the striking similarities between independently defined genome-wide patterns of Hi-C interaction maps and lamin B1 and nucleoli binding profiles both in terms of the overall resemblance of patterns and in functional designations of the associated chromatin loci defined as predominantly silent heterochromatin (Hi-C compartment B; LAD and NAD domains) and active chromatin (Hi-C compartment A; chromatin loci not associated with nuclear lamina and/or nucleoli). These observations suggest that interphase chromatin compartments which were independently defined using distinct methodological approaches may be structurally and, perhaps, functionally related. However, analysis of the fractal globule model of interphase chromatin geometry lacks systematic considerations of the impact of nuclear matrix and nucleolus.

Here we report the results of the external points of contact analysis of chromatin fractal globules based on a genome-wide alignment of inter- and intra-chromosomal chromatin interactions within the context of interphase chromatin binding to nuclear lamina and nucleolus. This analysis revealed near-centromeric intergenic regions on human chromosomes which are: (i) highly enriched for interphase chromatin homing sites; (ii) function as attractors of long-range genome-wide physical interactions; (iii) represent unique for human DNA sequences of 3.9 – 22.4 Kb in size. We propose to define these regions as Centromeric Regions of Interphase Chromatin Homing (CENTRICH) and postulate that CENTRICH play a critical role in 3D genome architecture and function by governing a

dynamic transition of large segments of interphase human chromosomes from the open loop to folded fractal globule conformations. We propose that emergence during the evolution of these unique to human arrangements of genomic DNA sequences contributes to distinct phenotypic features of *H.sapiens.*

Results and Discussion

Genome-wide maps of the interphase chromatin binding to nuclear lamina and Hi-C-defined chromatin folding compartments reveal highly correlated profiles inversely aligned along human chromosomes

Plotting the chromatin compartments defined by various methods linearly along the length of each chromosome revealed the intermitting chromosome-specific patterns of regions with positive and negative values (Figure 1). Regions with positive values reflect the enrichment in Hi-C compartment A and chromatin domains bound to nuclear lamina and/or nucleoli and regions with negative values reflect the enrichment in Hi-C compartment B and chromatin domains not associated with either nuclear lamina or nucleoli (Figure 1). Alignment of Hi-C interaction maps and binding profiles of Lamin B1 and nucleoli identified the striking similarities between independently defined genome-wide interphase chromatin patterns both in terms of the overall resemblance of patterns and in functional designations of the associated chromatin loci defined as predominantly silent heterochromatin (Hi-C compartment B; LAD and NAD domains) and active chromatin (Hi-C compartment A; chromatin loci not associated with either nuclear lamina or nucleoli). To quantify these relationships genome-wide, we correlated Hi-C model data (eigenvector values) for each chromosome with Lamin B1 and nucleoli binding data. We found that for all human chromosomes, a significant correlation exists between binding of chromosomal loci to nuclear lamina and segregation into spatially-defined distinct compartments of genome-wide chromatin interactions identified by Hi-C method (Figure 1; Table 1; Supplemental Figure 1). Therefore, loci in compartment B, which corresponds to the chromatin domains bound to the nuclear lamina, exhibit a stronger tendency for close spatial localization and

represent close, less accessible chromatin state, whereas loci in compartment A, which corresponds to the chromatin domains not associated with the nuclear lamina, are more closely associated with open, accessible, actively transcribed chromatin. This interpretation is in a good agreement with the observations that loci in compartment B showed a consistently higher interaction frequency at a given genomic distance than pairs of loci in compartment A indicating that compartment B is more densely packed and FISH data are consistent with this observation (5).

Careful analysis of deep sequencing data by van Koningsbruggen et al. (1) detected a significant overlap between the NADs and loci previously reported to associate with the nuclear envelope (LADs; ref. 3). Across the genome, analysis of Lamin B1 binding and association with nucleolus consistently identifies regions of heterochromatin with overlapping genomic coordinates, including multiple transcriptionally-competent intergenic disease-associated genomic loci (IDAGL; Figure 2) which produce biologically active trans-regulatory snpRNAs (6). These observations suggest that chromatin binding to Lamin B1 and association with nucleoli may impact genome-wide chromatin interactions by shuttling heterochromatin loci defined by overlapping locations and affecting dynamic features of a fractal globule. Periodic directional movements of specific regions of the chromatin fractal globules would likely to depend on relative strengths of association to nucleoli and may contribute to different availability of distinct chromatin chains for engagements in interchromosomal interactions. Consistent with this prediction, we found that there is a significant direct correlation between the relative strengths of the association to nucleoli of centromeric and q-telomeric regions of interphase chromatin homing on human chromosomes and increased likelihood of interchromosomal versus intrachromosomal interactions (Figure 2), that is human chromosomes with higher nucleoli/genomic ratios of chromatin binding exhibit increased ratios of genome-wide interchromosomal to intrachromosomal interactions. This model is in agreement with the idea that the apparent overlap in LAD and NAD loci may indicate that specific regions could alternate between binding to the nucleolus and the nuclear lamina either in different cells, or at different times within the same cells (1).

Proposed model of 3D architecture of interphase chromatin dynamics appears highly consistent with previously discovered strong correlation between replication timing and spatial proximity of chromatin as measured by Hi-C analysis that indicates that early and late initiation of replication occurs in spatially separate nuclear compartments, but rarely within the intervening chromatin (7). As shown in Figure 2 for chromosome 17, early replication timing chromatin compartments correspond to loci in Hi-C compartment A, which corresponds to the chromatin domains not bound to the nuclear lamina and is more closely associated with open, accessible, actively transcribed chromatins. Late replication timing chromatin compartments correspond to loci in Hi-C compartment B, which corresponds to the chromatin domains bound to the nuclear lamina and nucleolus, exhibits a stronger tendency for close spatial localization and represents close, less accessible, more densely packed and less actively transcribed heterochromatin state (Figure 2).

Identification of Centromeric Regions of Interphase Chromatin Homing (CENTRICH).

In human genome many gene deserts are significantly enriched for predicted long-range enhancers and/or insulators genomic coordinates of which are in close proximity or overlap with disease-associated SNPs (6, 8). Several groups experimentally validated the predicted enhancers and transcriptional activities of intergenic disease-associated genetic loci (IDAGL) with documented increased risk of developing prostate cancer (6; 9-13), autoimmune disorders (14), coronary artery diseases and type 2 diabetes (8). We performed a genome-wide survey of Hi-C-defined long-range physical chromatin interactions utilizing chromosomal positions of 99 intergenic disease-associated SNPs as probes to identify genomic coordinates of interacting pairs of enhancers and targeted genetic loci. Unexpectedly, our analysis revealed that all IDAGL home to specific discrete near-centrometic regions on human chromosomes 2 and 10. We designated these regions as the Cetromeric Regions of Interphase Chromatin Homing (CENTRICH). To determine whether CENTRICH are unique for human chromosomes 2 and 10, we performed genome-wide analysis of the frequencies of long-range interactions for all human chromosomes. To this end, we computed all

interactions along each chromosome consecutively segregated into 1Mb windows, plotted results of the analysis and look for regions that manifest statistically higher frequencies of long-range interactions. To evaluate statistical significance of the differences in the frequencies of long-range chromatin interactions, we compared the candidate CENTRICH frequency values to genome-wide frequency (test 1), individual chromosomal frequency (test 2), and interaction frequencies of neighboring regions (test 3; neighboring regions were 5 Mb segments on both sides of the candidate CENTRICH). Using these criteria, we identified ten CENTRICH on human chromosomes 2; 10; 17; 1; 7; 4; 21; 19; 16; 12 (Figure 1). Four CENTRICH on chr2; chr10; chr17; and chr1 represent most prominent interphase chromatin homing sites of human genome: they engaged in 397-1526 pair-wise interactions per 1 Kb distance, which represents 199 - 716-fold enrichment of interphase chromatin homing sites compared to genome-wide average (2-tail Fisher's exact test p values range $2.10E^{-101}$ - $1.08E^{-292}$). Comparisons of CENTRICH sequences reveals that CENTRICH on different chromosomes are highly homologous (Tables 2 & 3) with maximum sequence identities defined by BLAST search ranging from 76% (query coverage 99%; E value = 0; total score = 2.99E+05) to 88% (query coverage 100%; E value = 0; total score = 2.78E+05). Query coverage range in all pair-wise searches was from 99% to 100% and total score range was from 6.57E+04 to 1.43E+06 (E value = 0 in all pair-wise comparisons). Systematic BLAST searches did not identify significantly similar CENTRICH homologues in genomes of other mammals and vertebrates, including Primates (Callithrix jacchus, Macaca mulatta, Nomascus leucogenys, Pan troglodytes, Pongo abelii) and Rodents (Cavia porcellus, Cricetulus griseus, Mus musculus, Rattus norvegicus). These data suggest that CENTRICH are unique for human genomic regulatory sequences.

Notably, many protein-coding genes that were experimentally validated as regulatory targets of long-range enhancers and/or snpRNAs have multiple homing sites within the chr2 and chr10 CENTRICH (Figure 3). These data indicate that both long-range enhancers and regulatory target loci are bound to the chr2 and chr10 CENTRICH. Binding of both long-range enhancers and regulatory target loci to the CENTRICH would place them in close spatial proximity within the nuclear space and

facilitate physical interactions and clustering thus increasing the probability of phenotypically relevant regulatory events. Consistent with this model, for cancer, coronary artery disease, and type 2 diabetes, there is a statistically significant inverse correlation between the genome-wide association studies (GWAS)-defined odds ratios of increased risk of a disease and distances of SNP loci homing sites to the middle-point genomic coordinates of CENTRICH on chromosome 2 (Figure 3).

Interactions with nuclear lamina and nucleoli govern dynamic features of the chromatin fractal globules.

Across the genome, applications of three independent methods of genome-wide interphase chromatin analysis (Hi-C chromosome conformational capture, Lamin B1 binding, and association with nucleolus) consistently identify regions of heterochromatin with overlapping genomic coordinates, including multiple loci harboring disease-linked SNPs. These observations suggest that chromatin binding to Lamin B1 and association with nucleoli may impact genome-wide chromatin interactions by shuttling overlapping heterochromatin loci and affecting dynamic features of a fractal globule. According to this model, chromatin binding to either nuclear lamina or nucleolus initiates a polymer collapse and folding, promotes spatial segregation of genomic and chromosome territories, and contributes to stability of a fractal globule conformation (Figure 2). Shuttling between nuclear lamina and nucleoli would facilitate chromatin unfolding and loop opening, which would enable a genome-wide cross-talk due to chromatin loop invasion within and across territories and enhance the probability of long-range inter- and intra-chromosomal interactions (Figure 2). A fractal globule conformation would allow a rapid chromatin loop expansion (5, 15), crossing of boundaries of genomic and chromosome territories, enhancing a likelihood of long-range enhancer/promoter interactions and contributing to formation of transcription factories. Biomechanical forces created during a two-fold expansion of the nuclear volume in G1, changes of nuclear shape, and periodic nuclear rotations would impact the equilibrium of chromatin binding to nuclear lamina and nucleoli and influence initiation, duration, and frequency of a fractal globule loop folding/opening cycles. We

provide initial experimental evidence indicating that dynamic transition cycles of the fractals in the nucleus are affected by interactions of long-range intergenic enhancers with active promoters, concomitant activation of transcription at the enhancer's loci, and effects of enhancer's RNAs on the equilibrium state of fractal globule binding to the external nuclear structures. Systemic effects on fractal's dynamics of biomechanical forces during the interphase such as expansion of nuclear volume, nuclear rotation, fragmentation and assembly of nucleoli, periodic directional movements of sub-nuclear structures are likely to contribute to remarkable diversity of transcriptional landscapes and phenotypic features within populations of human cells.

CENTRICH exhibit remarkably diverse regulatory context of chromatin state maps, appear engaged in self-folding, intrachromosomal and interchromosomal interactions, and attract binding of multiple IDAGL with established association to increased risk of developing epithelial malignancies and other common human disorders. This high-complexity molecular assortment of CENTRICH structural features seems uniquely suitable to facilitate the extraordinary high density of long-range chromatin interactions.

Based on the analysis of the protein compositions of chromatin state maps of the CENTRICH and experimental evidence of homing to CENTRICH of protein coding genes with common patterns of transcriptional regulations, we propose the following biological functions of CENTRICH (Figures 3 & 4):

1) Monitoring of DNA integrity and repair of double-strand DNA breaks;

2) Matchmaking of interchromosomal pairs for long-range interactions; building of chromosome and genomic territories and transcription factories;

3) Reassembly of protein composition of chromatin and chromatin remodeling:

4) Change of transcriptional competence of interacting genomic regions;

5) Effect on stability and transition from condensed to open states of chromatin fractals and regulation of dynamics of chromatin fractal globules (folding, unfolding, and loop opening).

Recent experimental evidence and theoretical considerations strongly support the validity of the fractal globule model of a 3D architecture of interphase chromatin (5; 15-17). Our analysis reveals novel unique structural-functional features of interphase chromatin fractals which may have a significant impact on understanding of fundamental principles of dynamic behavior of chromatin chains folded into fractal globules. Our analysis argues that CENTRICH manifest certain levels of specificity with respect to the binding of regulatory proteins, chromatin chains of interphase human chromosomes, and protein-coding genes (Figure 4). Strikingly, chromatin of human mitochondrion manifest 682-fold and 907-fold greater propensity to interact with CENTRICH on chrr10 and chr2, respectively, compared to other human chromosomes (Figure 4). We demonstrate that near-centromeric gene deserts on human chromosomes contain unique to human homo- and heterotypic chromatin binding domains which function as powerful attractors of genome-wide long-range interactions and, therefore, contribute to definition of the 3D architecture of human genome. We propose that assortments of proteins bound to CENTRICH domains are designed to facilitate chromatin reprogramming, transcriptional competence, and 3D-alignment of interphase chromatin folding by performing continuous genome-wide scans of the interphase chromatin status. It will be of interest to determine how these unique to human arrangements of genomic DNA sequences contribute to distinct phenotypic features of H.sapiens.

Methods

Genome-wide analysis of long-range interphase chromatin interactions. In our computational analyses we considered all 59,309,839 of sequencing read pairs reported in Lieberman-Aiden et al. (2009). All Hi-C data were obtained from NCBI GEO (http://www.ncbi.nlm.nih.gov/geo/; GSE189199) and used in their original alignment summaries. Micosoft Access database containing 59,309,839 records of individual interactions was designed using published Hi-C data (Lieberman-Aiden et al.,

2009). For integenic snpRNA/enhancers, a match is called when of the interacting pair coordinates were found within an IDAGL region defined 10 kb up- or downstream of the corresponding disease-associated SNP position. For protein-coding genes, a match is called when at least one of the interacting pair coordinates was found within a gene boundaries or < 1 kb up- or downstream of the corresponding target gene.

Comparison to Hi-C Data. Hi-C data from Dekker and colleagues [Lieberman-Aiden et al., 2009] were obtained from NCBI GEO (http://www.ncbi.nlm.nih.gov/geo/; GSE189199); we used the data in their alignment summaries. We compared their list of interacting pairs to our regulator–target gene pairs defined as the snpRNA/enhancer loci and protein coding genes.

Comparison to lamin B1 binding data. Lamin B1 binding data were originally reported in Guelen et al. (2008) and are available from the UCSC genome browser at http://genome.ucsc.edu/ .

Comparison to nucleoli binding data. Nucleoli binding data were originally reported in van Koningsbruggen et al. (2010) and were obtained from Dr. Lamond (angus@lifesci.dundee.ac.uk).

To quantify these relationships genome-wide, we correlated Hi-C model data (eigenvector data) for each chromosome with lamin B1 and nucleoli binding data. Comparison to the Hi-C model was made using 100-kb window eigenvector data corresponding to the checkerboard patterns in Lieberman-Aiden et al. (2009). The correlation analyses of the 5 Mb moving averages data sets for each comparison are reported. To perform systematic genome-wide analysis of the frequencies of long-range interactions for all human chromosomes, we computed all interactions along each chromosome consecutively segregated into 1Mb windows, plotted results of the analysis and look for regions that manifest statistically higher frequencies of long-range interactions. To evaluate statistical significance of the differences in the frequencies of long-range chromatin interactions, we compared the candidate CENTRICH frequency values to genome-wide frequency (test 1), individual chromosomal frequency (test 2), and interaction frequencies of neighboring regions (test 3; neighboring regions were 5 Mb

segments on both sides of the candidate CENTRICH). Candidate CENTRICH that passes all three statistical tests were considered in subsequent analyses and reported in this paper.

References

1. van Koningsbruggen S, Gierlinski M, Schofield P, Martin D, Barton GJ, Ariyurek Y, den Dunnen JT, Lamond AI. High-resolution whole-genome sequencing reveals that specific chromatin domains from most human chromosomes associate with nucleoli. Mol Biol Cell. 2010; 21:3735-48.

2. Németh A, Conesa A, Santoyo-Lopez J, Medina I, Montaner D, Péterfia B, Solovei I, Cremer T, Dopazo J, Längst G. Initial genomics of the human nucleolus. PLoS Genet. 2010; 6:e1000889.

3. Guelen L, Pagie L, Brasset E, Meuleman W, Faza MB, Talhout W, Eussen BH, de Klein A, Wessels L, de Laat W, van Steensel B. Domain organization of human chromosomes revealed by mapping of nuclear lamina interactions. Nature 2008; 453:948-51.

4. Peric-Hupkes D, Meuleman W, Pagie L, Bruggeman SW, Solovei I, Brugman W, Gräf S, Flicek P, Kerkhoven RM, van Lohuizen M, Reinders M, Wessels L, van Steensel B. Molecular maps of the reorganization of genome-nuclear lamina interactions during differentiation. Mol Cell. 2010; 38:603-13.

5. Lieberman-Aiden E, van Berkum NL, Williams L, Imakaev M, Ragoczy T, et al. (2009) Comprehensive mapping of long-range interactions reveals folding principles of the human genome. Science 2009; 326:289–293.

6. Glinskii AB, Ma S, Ma J, Grant D, Lim CU, Guest I, Sell S, Buttyan R, Glinsky GV. Networks of intergenic long-range enhancers and snpRNAs drive castration-resistant phenotype of prostate cancer and contribute to pathogenesis of multiple common human disorders. Cell Cycle 2011; 10:3571-97.

7. Ryba T, Hiratani I, Lu J, Itoh M, Kulik M, Zhang J, Schulz TC, Robins AJ, Dalton S, Gilbert DM. Evolutionarily conserved replication timing profiles predict long-range chromatin interactions and distinguish closely related cell types. Genome Research 2010; 20:761–770.

8. Harismendy O, Notani D, Song X, Rahim NG, Tanasa B, Heintzman N, Ren B, Fu XD, Topol EJ, Rosenfeld MG, Frazer KA. 9p21 DNA variants associated with coronary artery disease impair interferon-γ signalling response. Nature 2011; 470: 264–268.

9. Pomerantz MM., et al. The 8q24 cancer risk variant rs6983267 demonstrates long-range interaction with MYC in colorectal cancer. Nat Genet. 2009;41: 882–884.

10. Jia L, Landan G, Pomerantz M, Jaschek R, Herman P, et al. (2009). Functional Enhancers at the Gene-Poor 8q24 Cancer-Linked Locus. PLoS Genet 2009;5(8): e1000597.

11. Sotelo J., Esposito D, Duhagon MA., et al. Long-range enhancers on 8q24 regulate c-Myc. PNAS 2010;107:3001-3005.

12. Wright JB, Brown SJ, Cole MD. Upregulation of c-MYC in cis through a Large Chromatin Loop Linked to a Cancer Risk-Associated Single-Nucleotide Polymorphism in Colorectal Cancer Cells. Mol Cell Biol 2010;30: 1411–1420.

13. Wasserman NF, Ivy Aneas I, Nobrega MA. An 8q24 gene desert variant associated with prostate cancer risk confers differential in vivo activity to a MYC enhancer. Genome Res 2010;20: 1191-1197.

14. Glinskii, AB, Ma, J, Ma, S, Grant, D, Lim, C, Sell, S, Glinsky, GV. Identification of intergenic trans-regulatory RNAs containing a disease-linked SNP sequence and targeting cell cycle progression/differentiation pathways in multiple common human disorders. Cell Cycle 2009; 8:3925-42.

15. Mirny LA. The fractal globule as a model of chromatin architecture in the cell. Chromosome Res 2011; 19:37-51.

16. McNally JG, Mazza D. Fractal geometry in the nucleus. The EMBO Journal 2010; 29, 2-3.

17. Bancaud A, Huet S, Daigle N, Mozziconacci J, Beaudouin J, Ellenberg. Molecular crowding affects diffusion and binding of nuclear proteins in heterochromatin and reveals the fractal organization of chromatin. EMBO J 2009; 28:3785-98.

Table 1. Correlation of interphase chromatin folding patterns of human interphase chromosomes independently defined by the Hi-C compartments Eigenvector values and Lamin B1 binding scores

	GM06990 cells	K562 cells
Chromosome 1	-0.475	-0.485
Chromosome 2	-0.521	-0.457
Chromosome 3	-0.412	-0.366
Chromosome 4	-0.385	-0.388
Chromosome 5	-0.477	-0.428
Chromosome 6	-0.592	-0.432
Chromosome 7	-0.343	-0.149
Chromosome 8	-0.545	-0.506
Chromosome 9	-0.578	-0.574
Chromosome 10	-0.369	-0.302
Chromosome 11	-0.571	-0.582
Chromosome 12	-0.396	-0.337
Chromosome 13	-0.544	-0.633
Chromosome 14	-0.71	-0.463
Chromosome 15	-0.732	-0.586
Chromosome 16	-0.765	-0.691
Chromosome 17	-0.573	-0.585
Chromosome 18	-0.507	-0.426
Chromosome 19	-0.867	-0.757
Chromosome 20	-0.838	-0.739
Chromosome 21	-0.662	-0.559
Chromosome 22	-0.636	-0.588
Chromosome X	-0.448	-0.452

Table 2. Sequences producing significant alignments and 99% - 100% query coverage in CENTRICH sequence homology analysis.

chr 2 attrcator (20799 bp) vs chr10 attractor (22389 bp)

Accession	Max score	Total score	Query coverage	E value	Max ident
21883	**3048**	**3.49E+05**	**100%**	**0**	**79%**

chr17 attractor (19134 bp) vs chr2 attractor (20799 bp)

Accession	Max score	Total score	Query coverage	E value	Max ident
52035	**1.18E+04**	**1.42E+06**	**100%**	**0**	**77%**

chr17 attractor (19134 bp) vs chr10 attractor (22389 bp)

Accession	Max score	Total score	Query coverage	E value	Max ident
55233	**2841**	**2.99E+05**	**99%**	**0**	**76%**

chr 2 attrcator (20799 bp) vs chr1 attractor (3860 bp)

Accession	Max score	Total score	Query coverage	E value	Max ident
4731	**1701**	**2.77E+05**	**99%**	**0**	**88%**

chr 2 attrcator (20799 bp) vs chr17 attractor (19134 bp)

Accession	Max score	Total score	Query coverage	E value	Max ident
52979	**1.18E+04**	**1.42E+06**	**99%**	**0**	**77%**

chr17 attractor (19134 bp) vs chr1 attractor (3860 bp)

Accession	Max score	Total score	Query coverage	E value	Max ident
51573	**2365**	**2.66E+05**	**100%**	**0**	**85%**

chr 1 attrcator (3860 bp) vs chr10 attractor (22389 bp)

Accession	Max score	Total score	Query coverage	E value	Max ident
49495	**1624**	**6.57E+04**	**100%**	**0**	**88%**

chr 1 attrcator (3860 bp) vs chr2 attractor (20799 bp)

Accession	Max score	Total score	Query coverage	E value	Max ident
65391	**1701**	**2.78E+05**	**100%**	**0**	**88%**

chr 1 attrcator (3860 bp) vs chr17 attractor (19134 bp)

Accession	Max score	Total score	Query coverage	E value	Max ident
37109	**2365**	**2.66E+05**	**100%**	**0**	**85%**

Legend: BLAST Reference: Zheng Zhang, Scott Schwartz, Lukas Wagner, and Webb Miller . A greedy algorithm for aligning DNA sequences. J Comput Biol 2000; 7(1-2):203-14.

Figure legends

Figure 1. Alignments of the Hi-C Eigenvector profiles, Lamin B1 binding, and association with nucleolus of the human chromosome 17 interphase chromatin chain.

A. Inverse correlation of the Hi-C Eigenvector profile and Lamin B1 binding pattern of the human chromosome 17.

B. Hi-C map of long-range interphase chromatin interactions of the human chromosome 17 at 1 Mb resolution.

C. Alignment of the chromosome 17 CENTRICH region to the centromeric DNA attachment site to the nucleolus.

Figure 2. Attachment to nucleolus and nuclear lamina governs dynamic features and functions of interphase chromatic chains.

A. Direct correlation of the strength of association with nucleoli of the centromeric regions of interphase chromatin homing (defined by the values of the nucleoli/genomic ratio indices measured for each chromosome) and the propensity to engage in the long-range inter-chromosomal interactions (defined by the ratios of the inter- to intra-chromosomal interactions measured for each chromosome).

B, C. Intergenic disease-associated genomic loci (IDAGL) manifest the overlapping pattern of binding to both nucleolus and nuclear lamina which indicate that these specific genomic regions could alternate (shuttle) between binding to the nucleolus and the nuclear lamina either in different cells within cell population, or at different times within the same cells.

D. Correlations of binding to nuclear lamina (Lamin B1 binding, LB1), association with nucleoli (nucleoli/genomic ratio, N), and replication timing (RT) profiles measured for the human chromosome 17 interphase chromatin chain.

E. Same as in D with the addition of the Hi-C eigenvector data (Hi-C) defining intrephase chromatin compartments A (Hi-C positive values) and B (Hi-C negative values).

Figure 3. Initial characterization of chromatin-binding features and potential relevance to human diseases of three main attractors of long-range interphase chromatin homing sites in human genome.

A. Enrichment of the interphase chromatin homing sites in defined intergenic regions of human genome on chromosome 2 and chromosome 10.

B. Genomic coordinates (human genome build 36.3) of binding sites of the 8q24 gene desert IDAGL within 20 Kb chromosome 2 CENTRICH (top figure), correlation plot (bottom left figure) and linear regression analysis (bottom right figure) showing the inverse correlation between cancer risk odds ratios (OR) defined by the cancer susceptibility risk loci of the 8q24 gene desert and distances between homing sites of the 8q24-locus cancer susceptibility SNPs within chr2 CENTRICH and a middle point of the chr2 CENTRICH.

C. Genomic coordinates (human genome build 36.3) of the 9p21 gene desert SNPs and intergenic enhancers associated with increased risk of coronary artery disease (CAD) and type 2 diabetes (T2D) and homing sites of the 9p21 locus within 20 Kb chr2 CENTRICH (top figure), correlation plot (bottom left figure) and linear regression analysis (bottom right figure) showing the inverse correlation between CAD and T2D risk odds ratios (OR) defined by the CAD and T2D susceptibility risk loci of the 9p21 gene desert and distances to the nearest homing sites of the 9p21-locus disease susceptibility SNPs within the chr2 CENTRICH.

D. Genomic coordinates of type 2 diabetes (T2D) risk loci defined by the genome-wide association studies (GWAS).

E. Correlation plot (left figure) and linear regression analysis (right figure) showing the inverse correlation between T2D risk odds ratios (OR) defined by the intronic (6 SNPs) and intergenic (5 SNPs) T2D susceptibility risk loci listed in D and distances between their homing sites within chr2 CENTRICH and a middle point of the chr2 CENTRICH.

F. Genomic coordinates (human genome build 36.3) of binding sites of the chromatin chains within the boundaries of the six protein-coding genes listed in D (these genes carry intronic T2D susceptibility SNPs) within 20 Kb chromosome 2 CENTRICH.

Figure 4. CENTRICH manifest distinct binding features with respect to regulatory proteins (A, B), interphase chromatin chains of human chromosomes (C, D), and protein-coding genes (E).

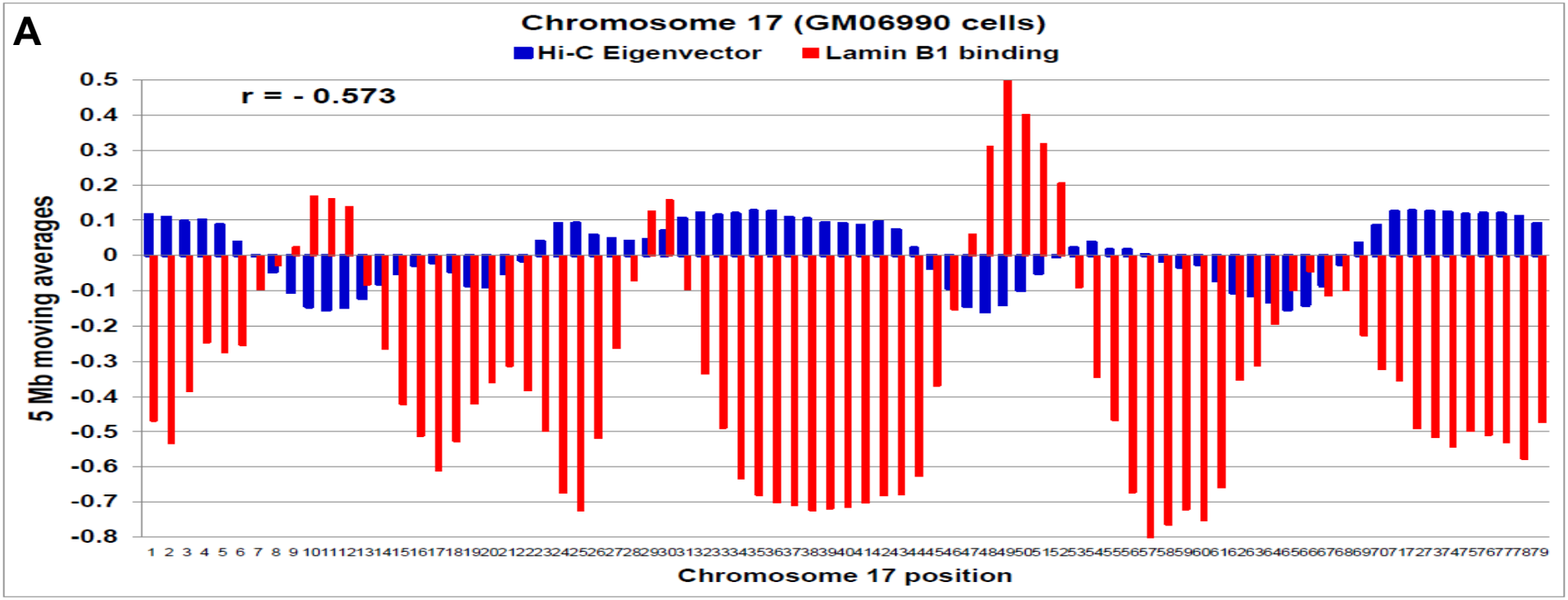
A. Summary of the analysis of the ENCODE chromatin state maps for the 99 IDAGL homing sites within chr2 CENTRICH. Homing sites for each individual IDAGL were mapped and segregated in distinct sets of homing sites based on disease risk loci designation. Regulatory proteins bound to each sites were recorded using publicly available ENCODE chromatin state maps data. Chromatin state maps of CENTRICH were visualized using the custom tracks of the UCSC Genome Browser (http://genome.ucsc.edu/ENCODE/).

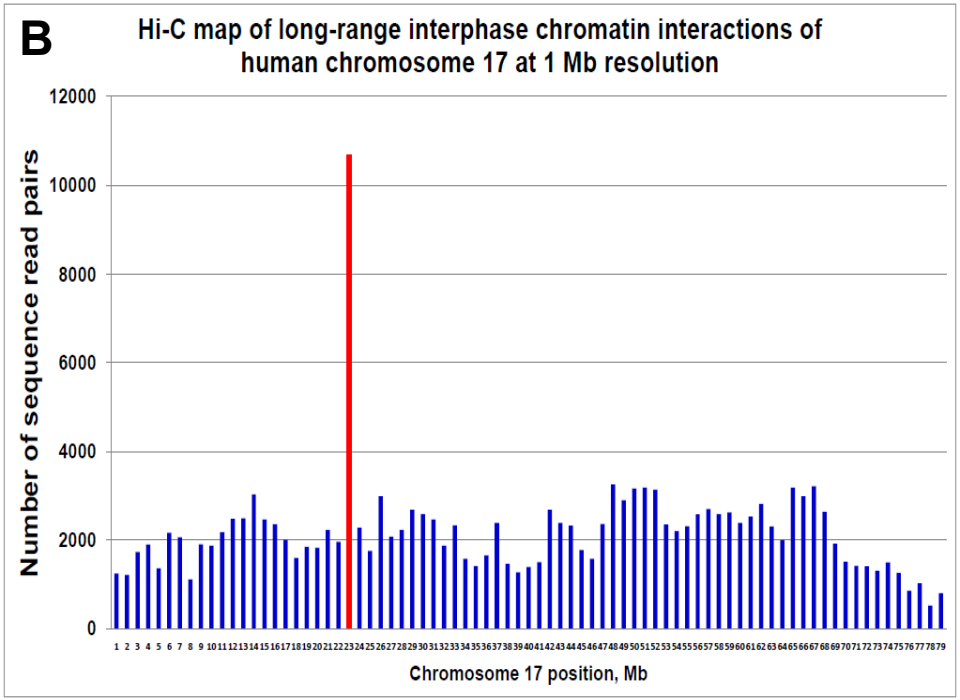
B. Consensus and cumulative signatures of regulatory proteins bound to chr2 CENTRICH.

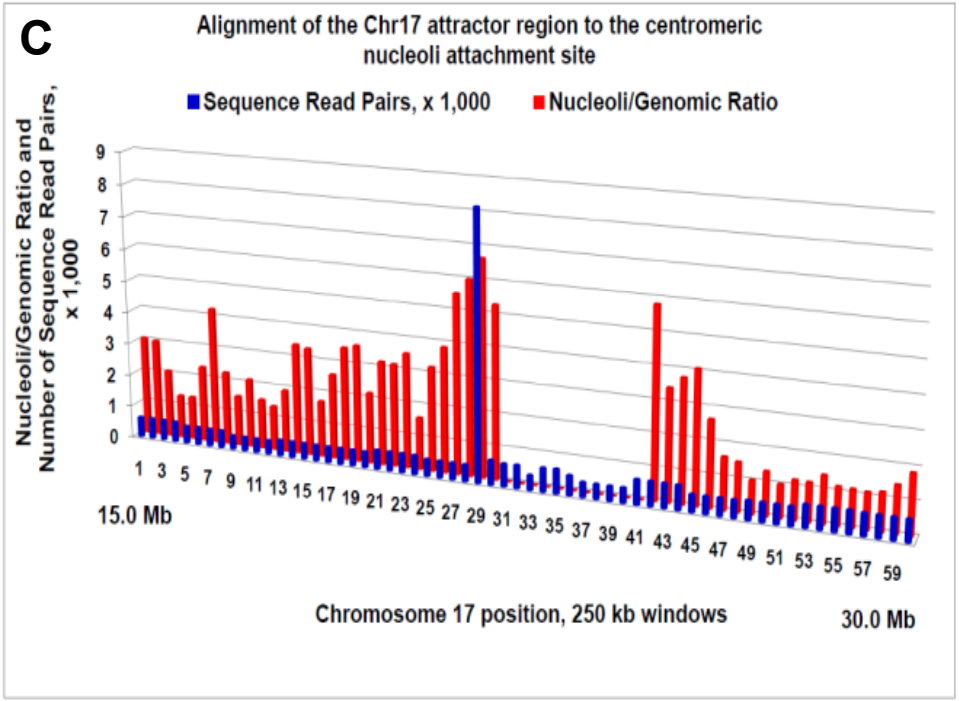
C. Chromatin of human mitochondrion manifest 682-fold and 907-fold greater propensity to interact with CENTRICH on chrr10 and chr2, respectively. All interactions within defined CENTRICH for individual chromosomes were recorded, quantified, and reported as values normalized per 1 Kb distance.

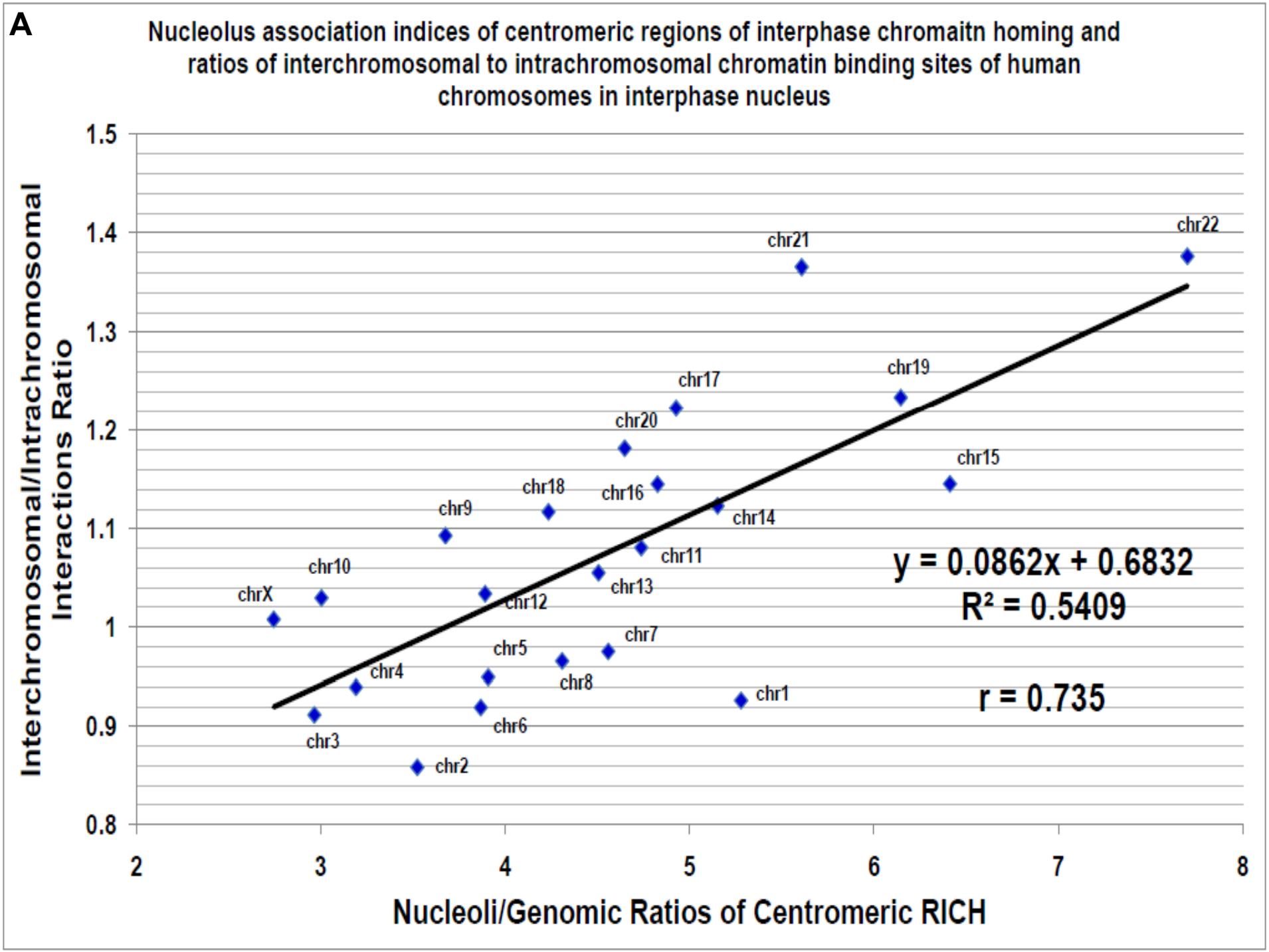
D. Most human chromosomes manifest ~2-fold preference to engage in long-range interactions with chr2 CENTRICH. Note the 150-fold and 750-fold maximum value scale differences in Figure D and inset, respectively, comapred to the Figure C.

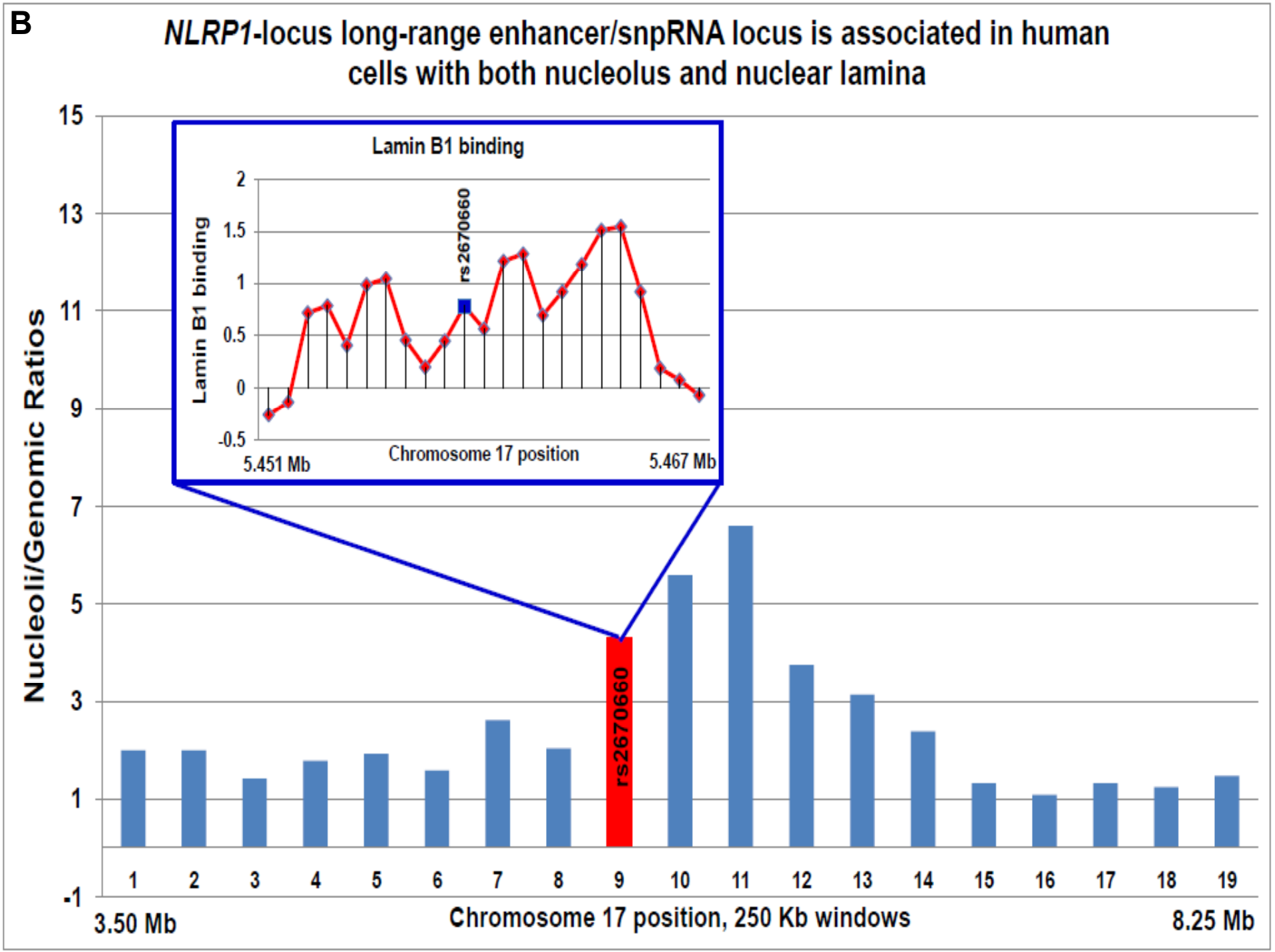
E. Protein-coding genes manifest distinct patterns of binding to the CENTRICH. 33% of the surveyed 192 protein-coding genes had homing sites only in single CENTRICH, while 29% and 38% had homing sites in two and three CENTRICH, respectively. Note statistically significant difference in gene size (left figure) and number of homing sites (right figure) per gene for protein-coding genes with distinct patterns of binding to CENTRICH.

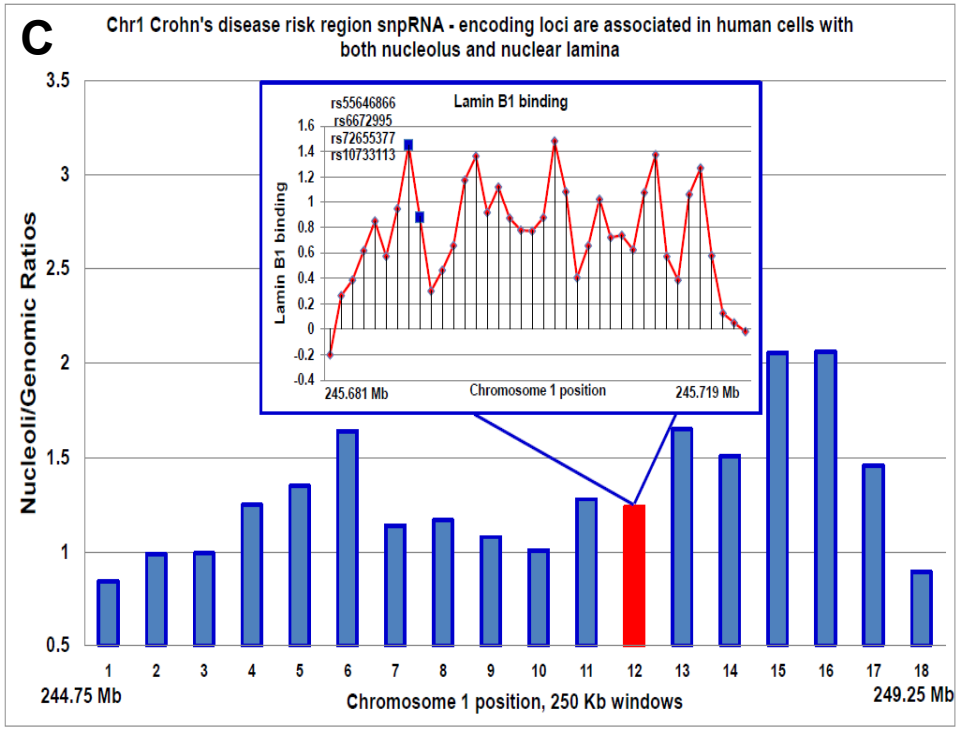


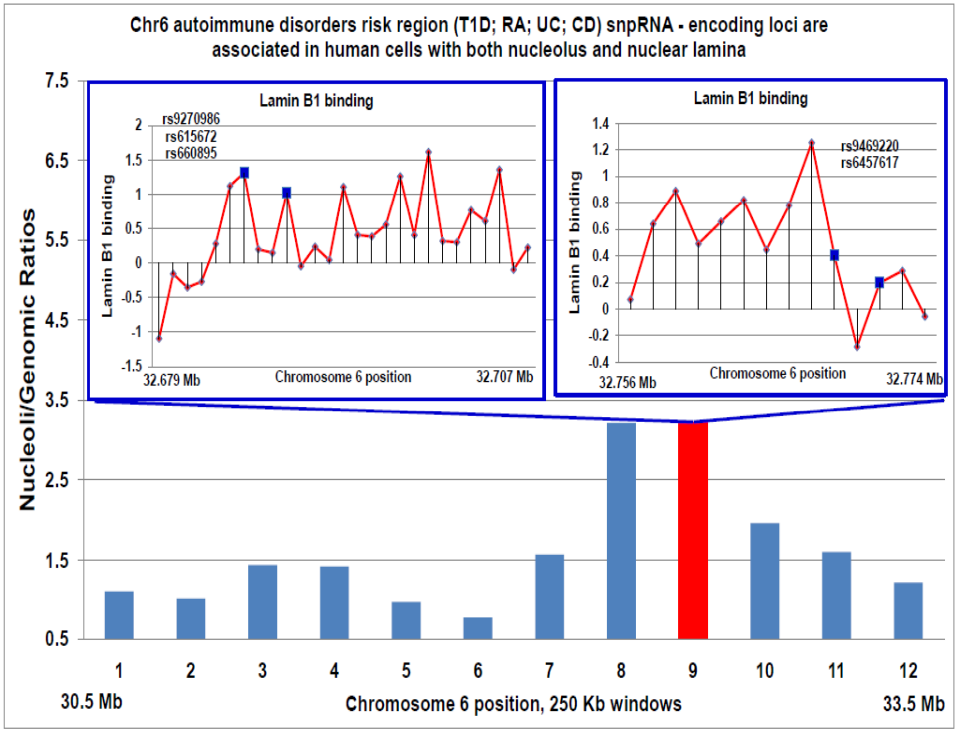


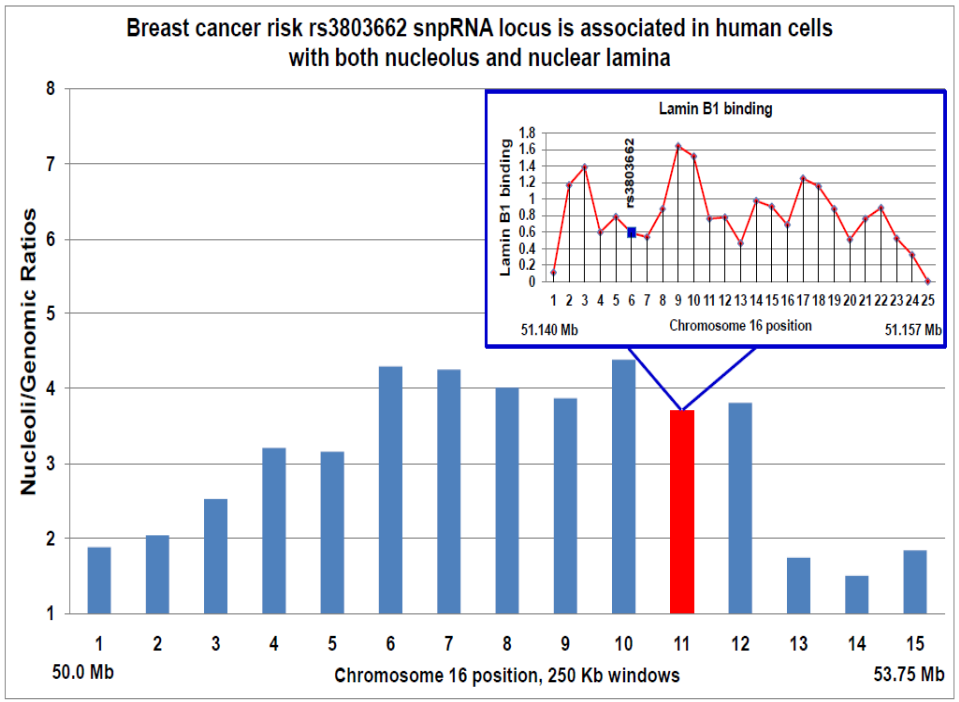


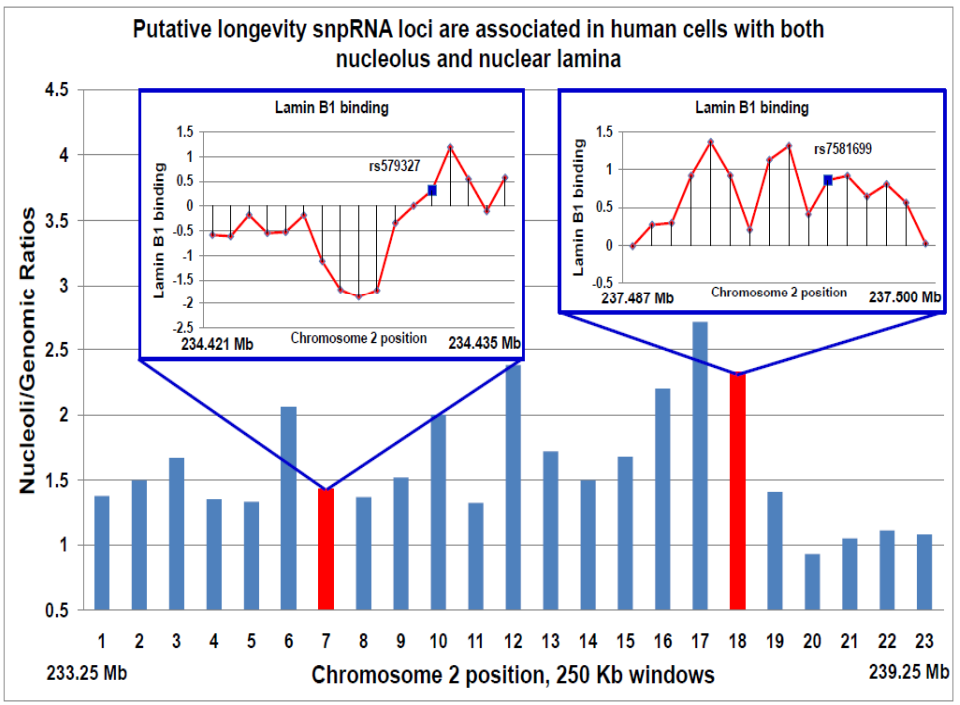


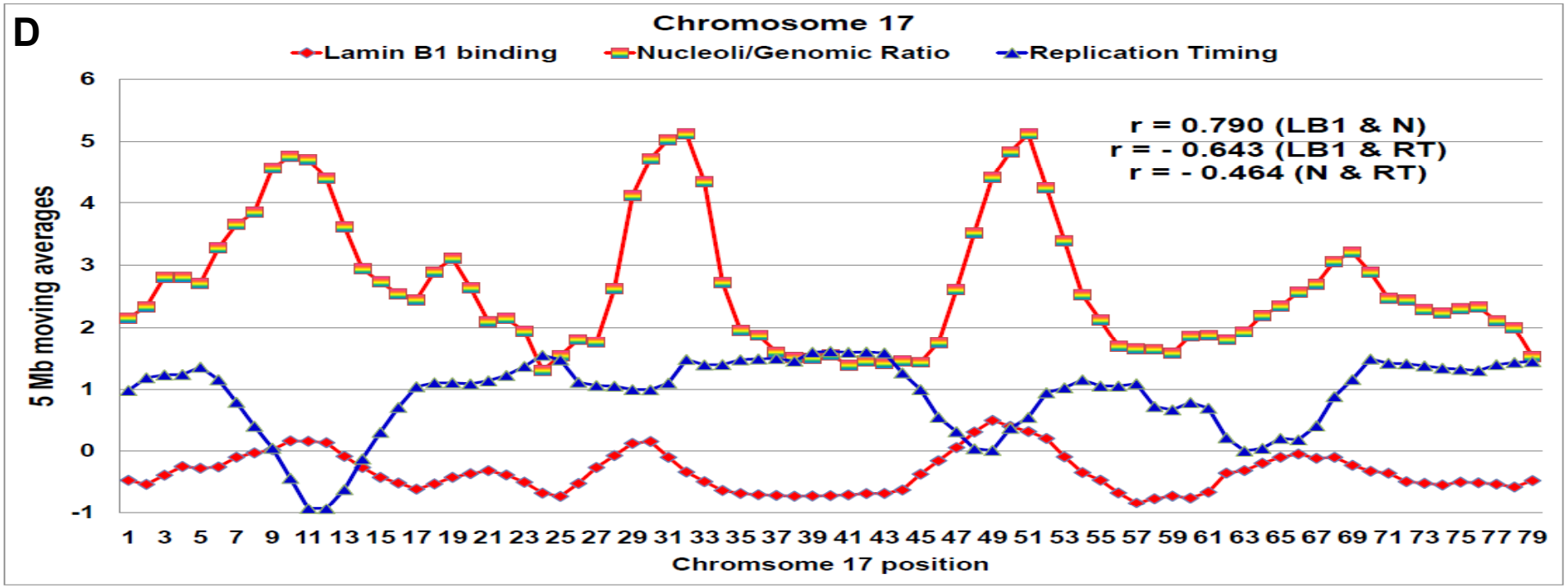


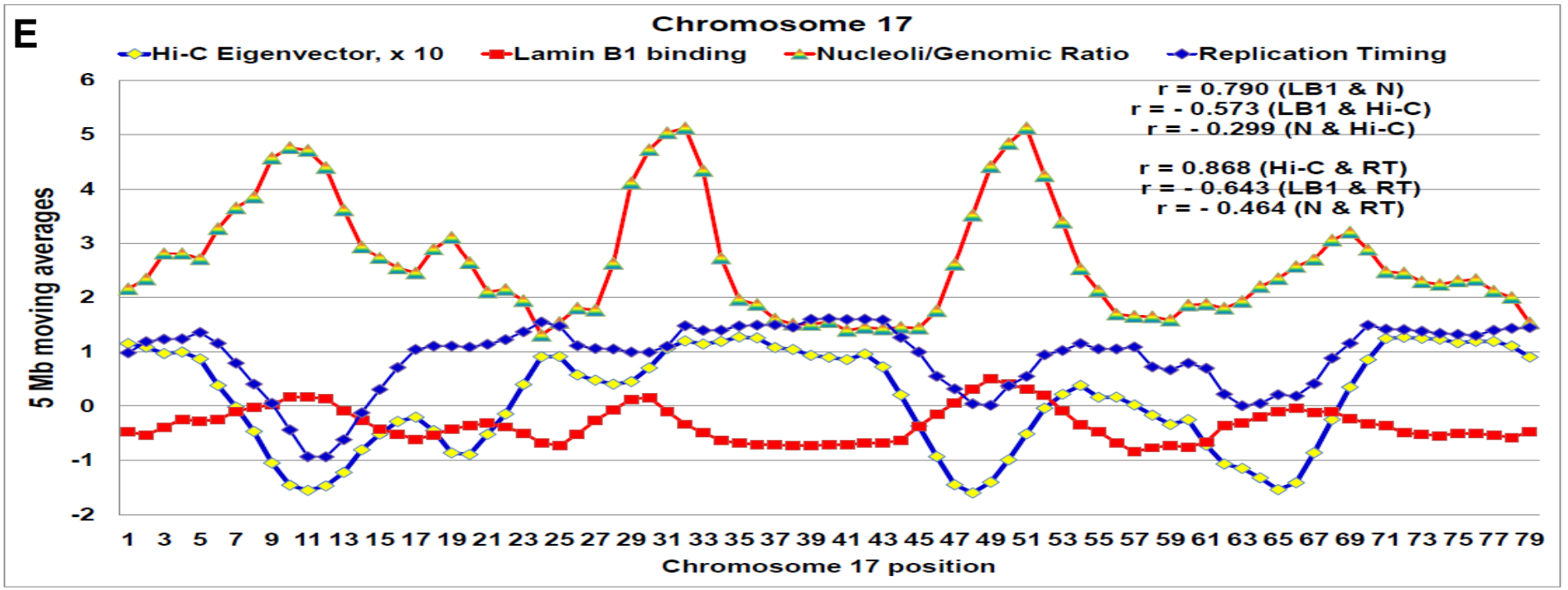


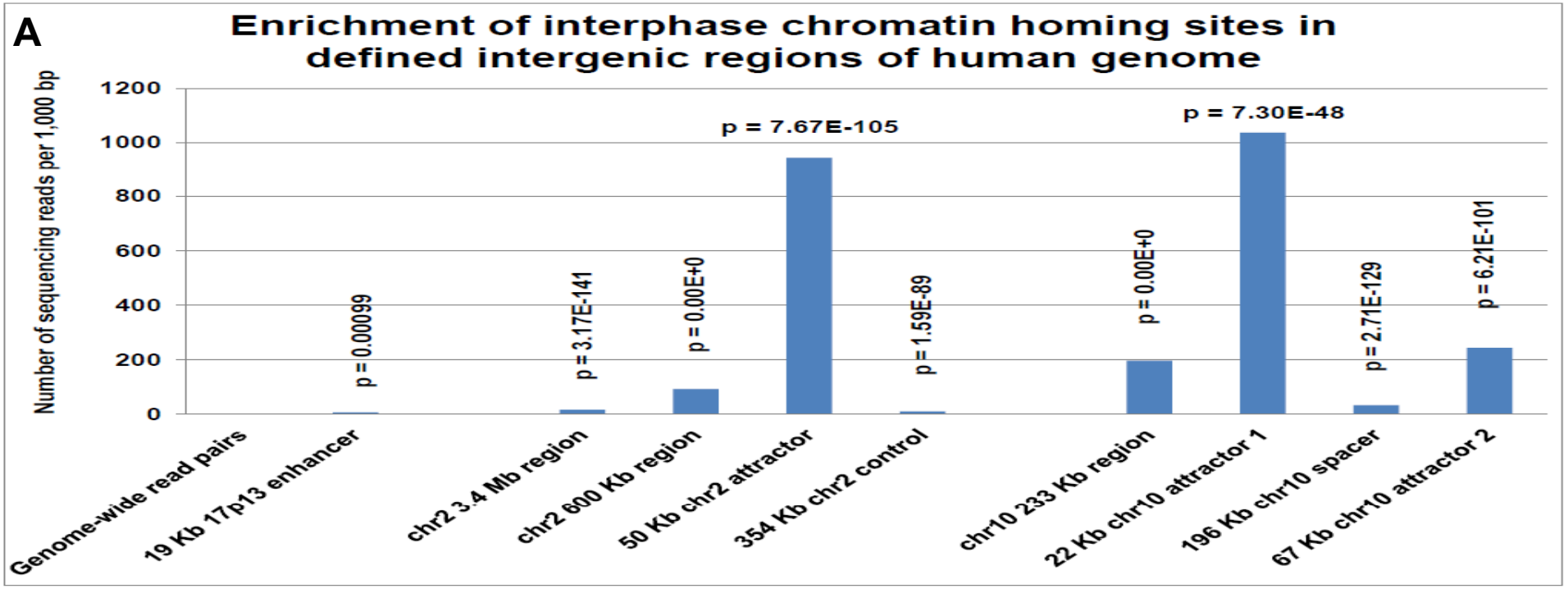


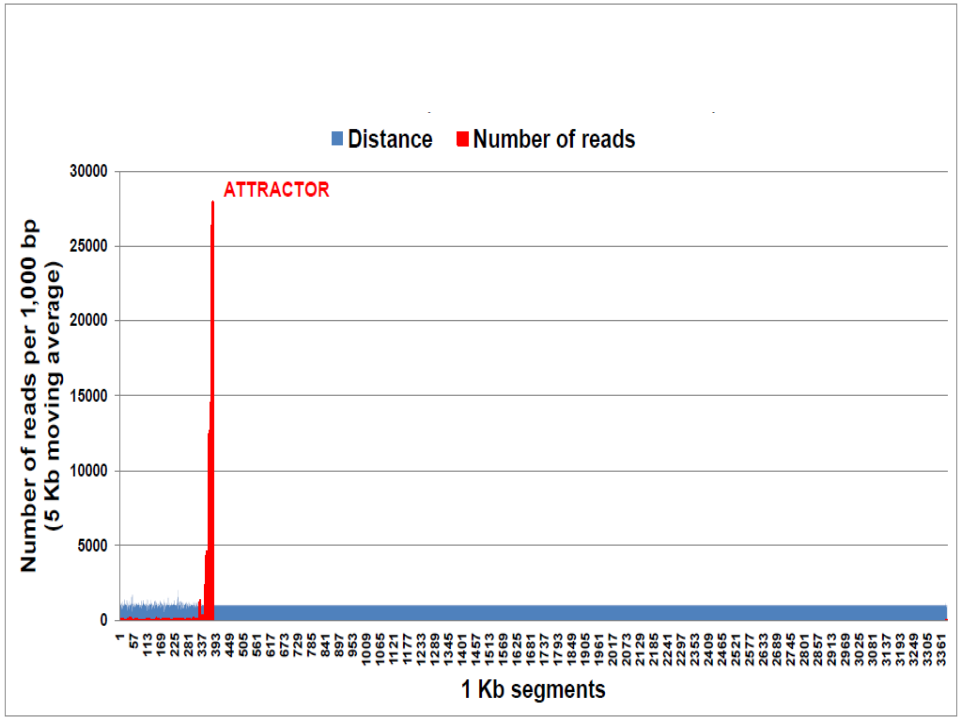


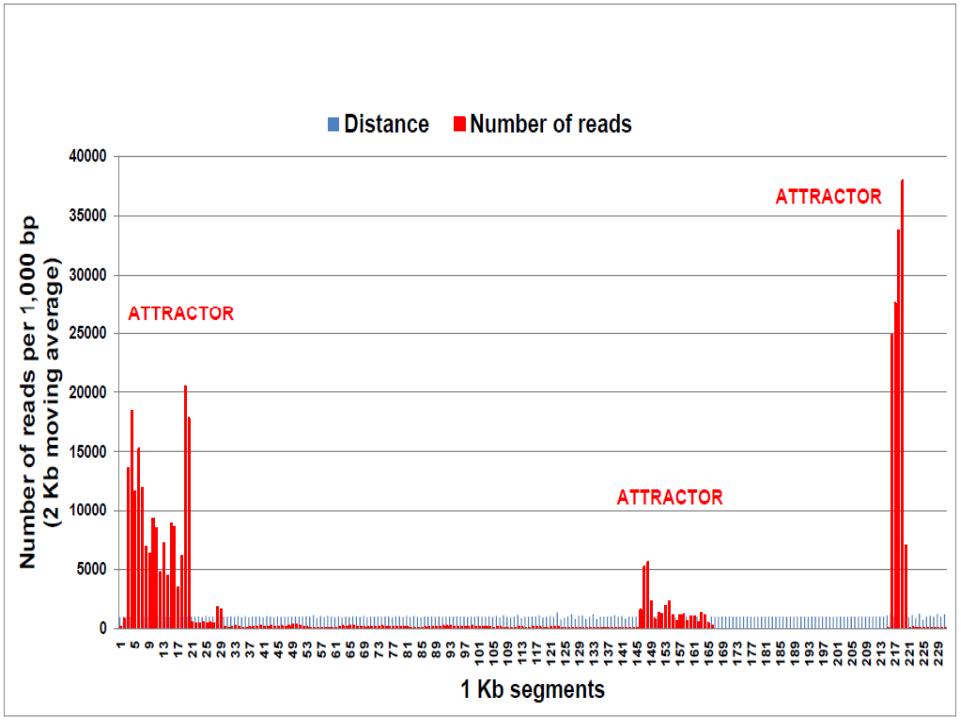




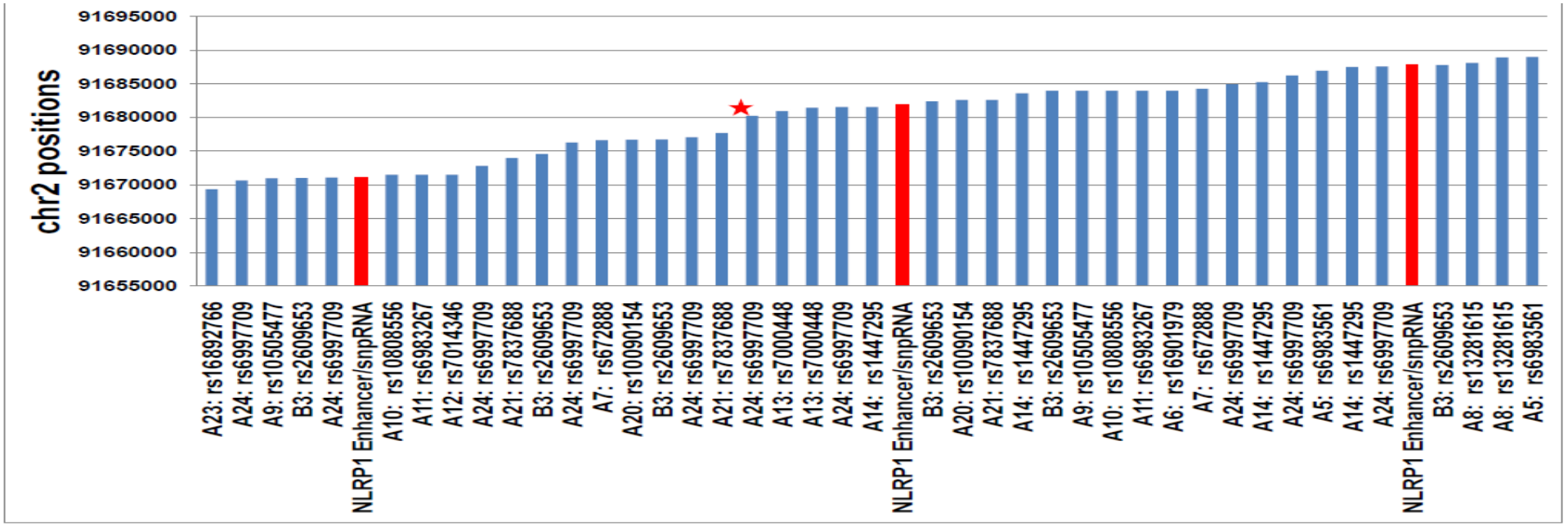


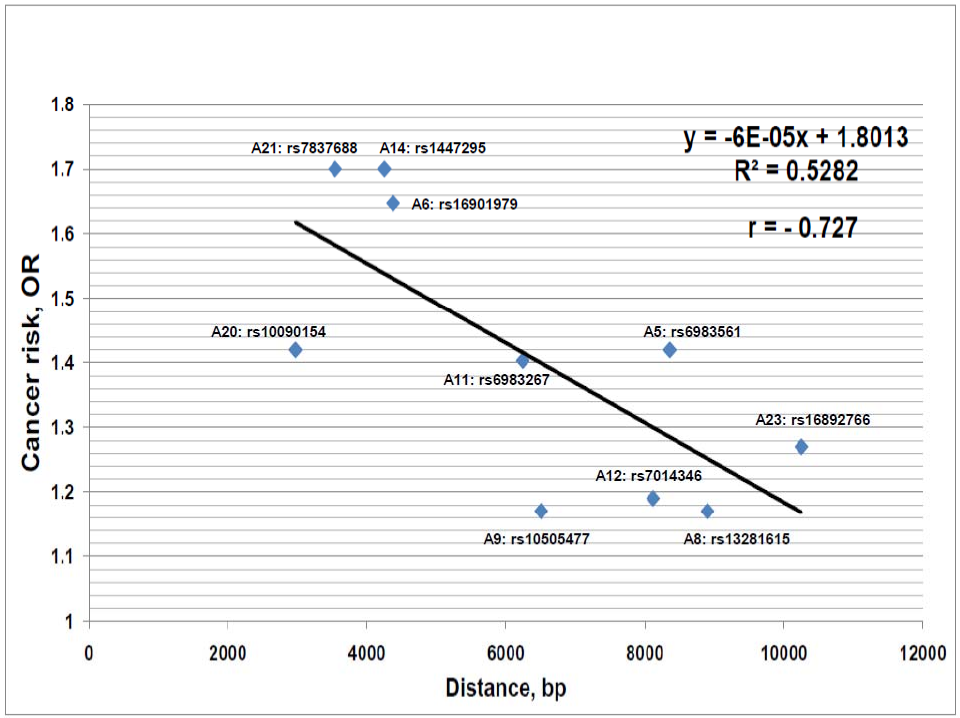


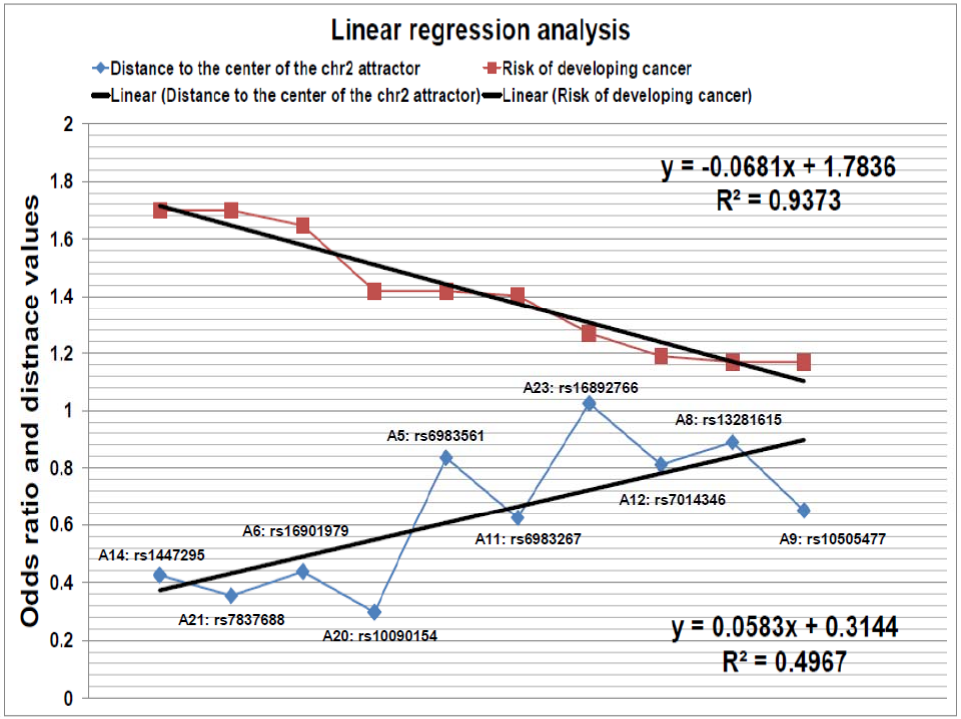


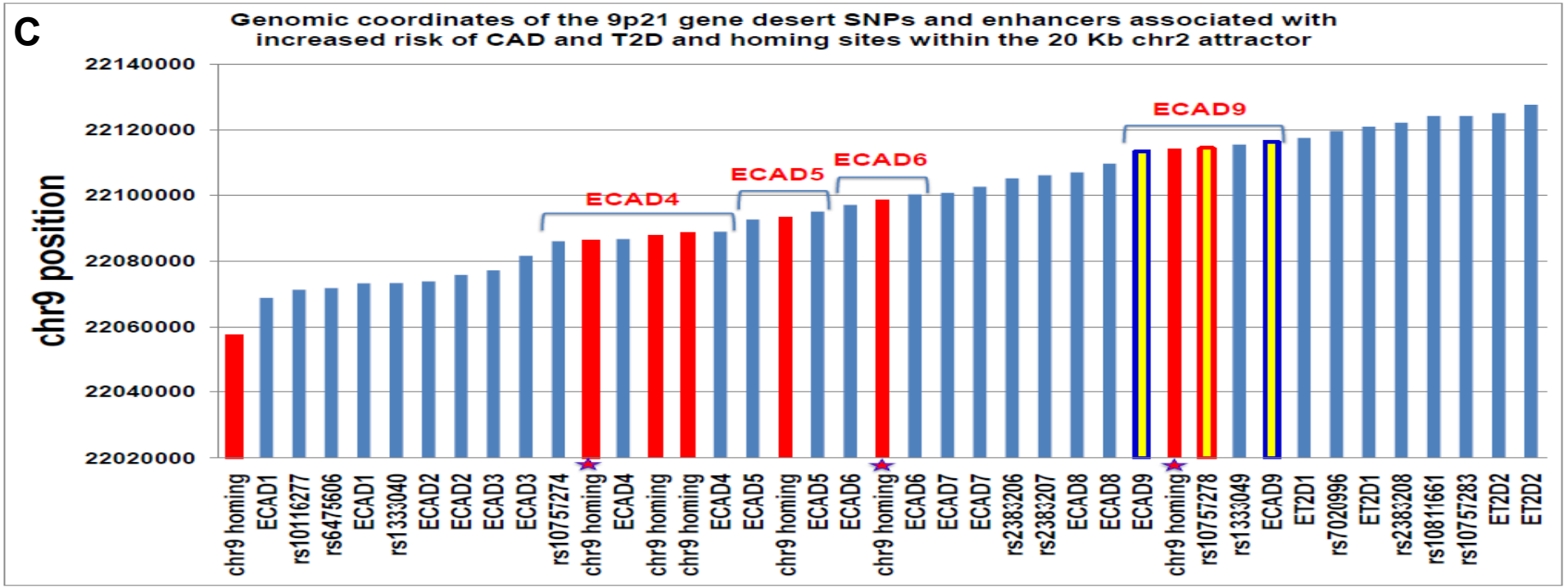


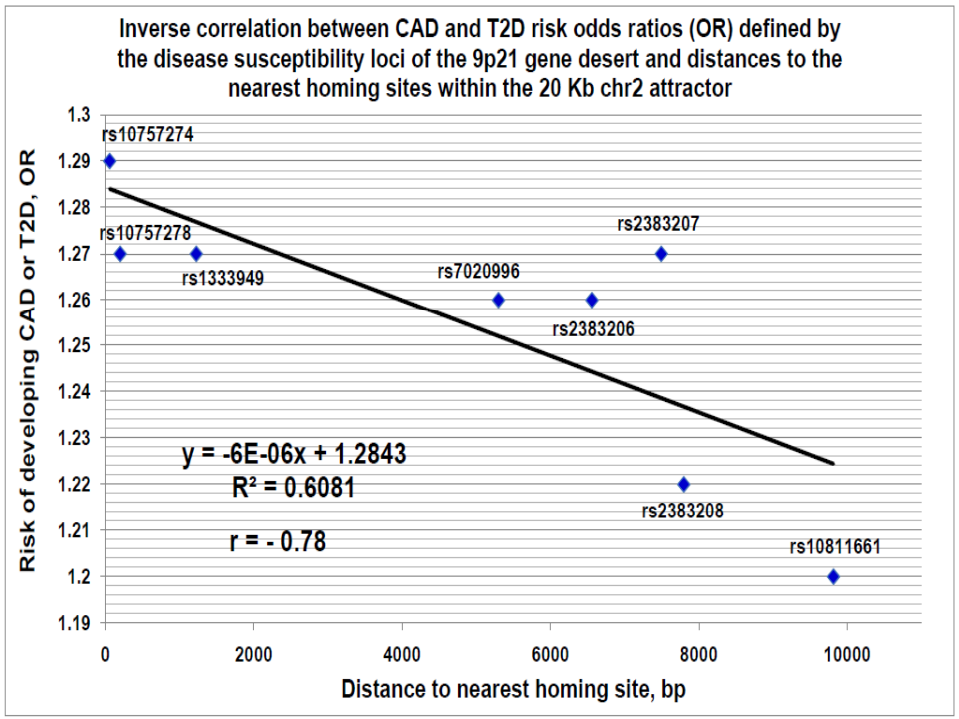
B

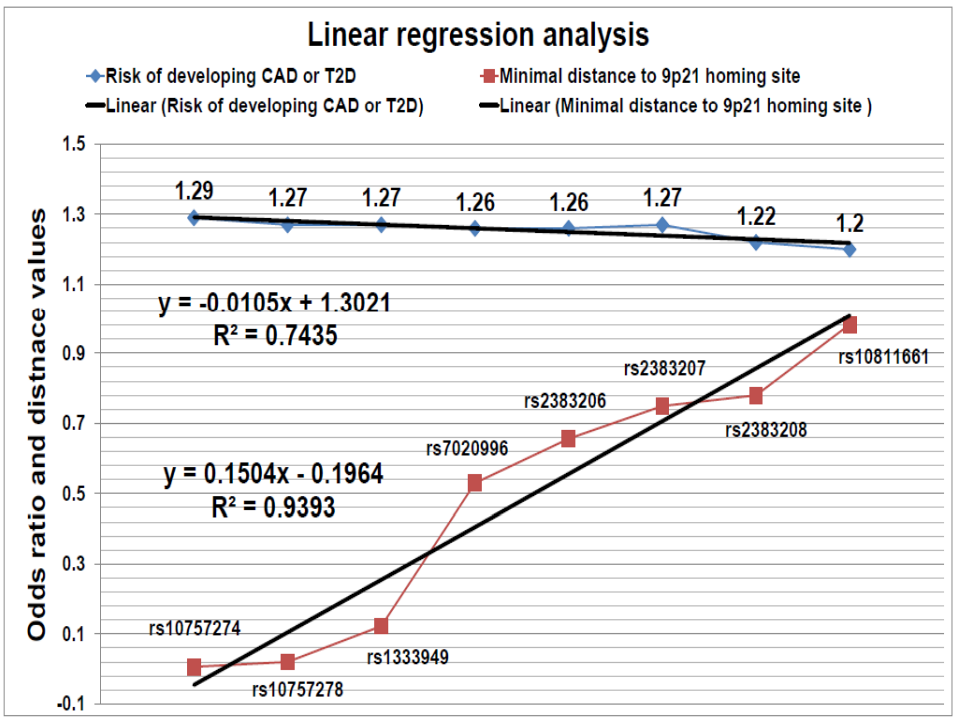






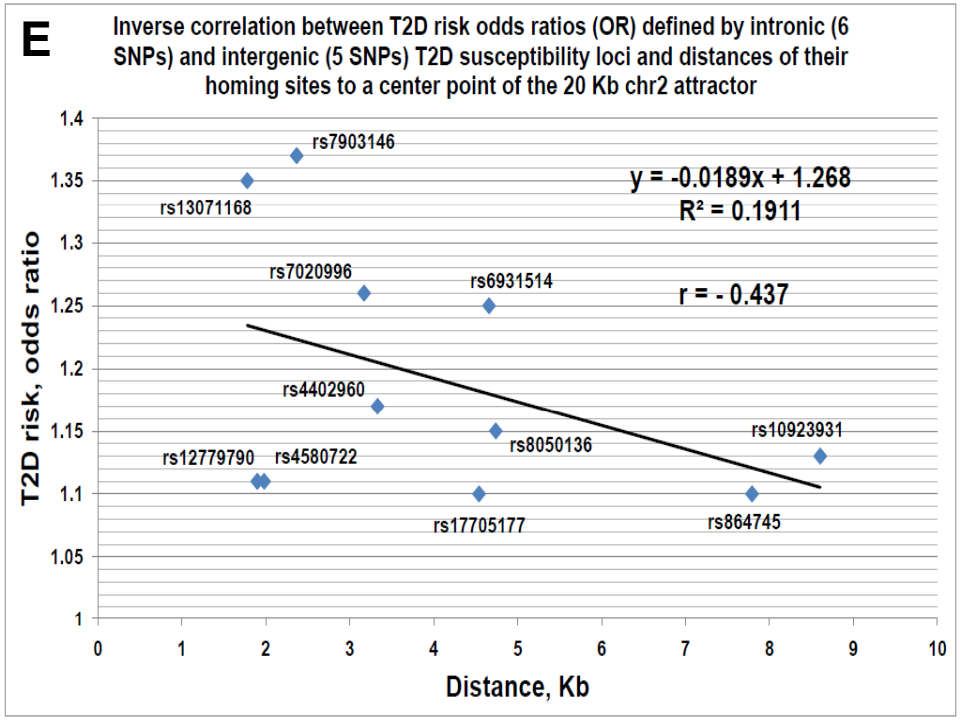


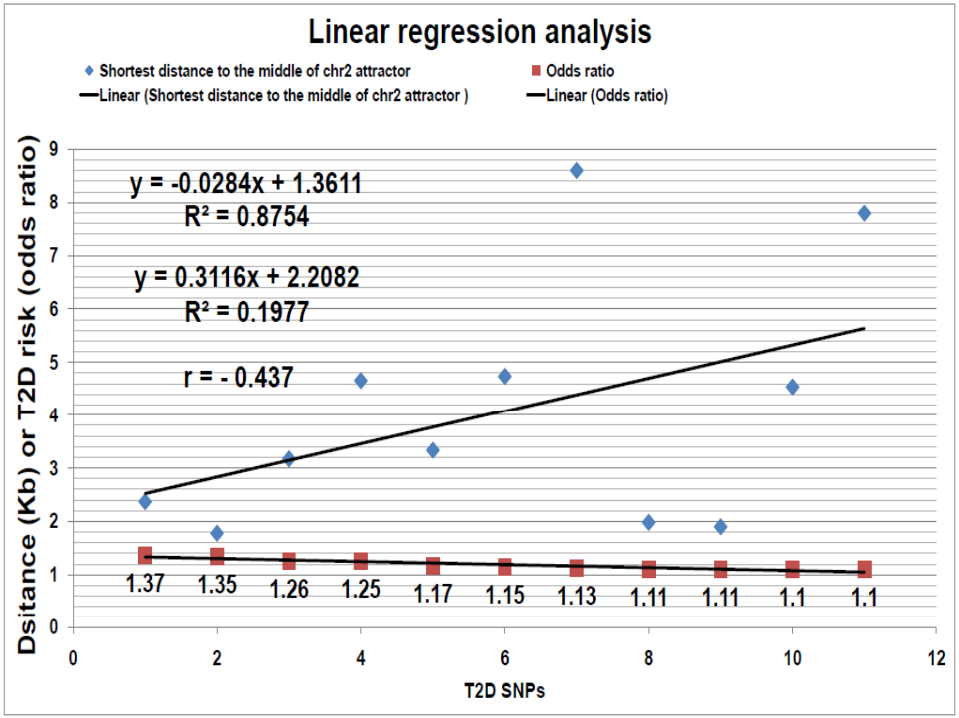


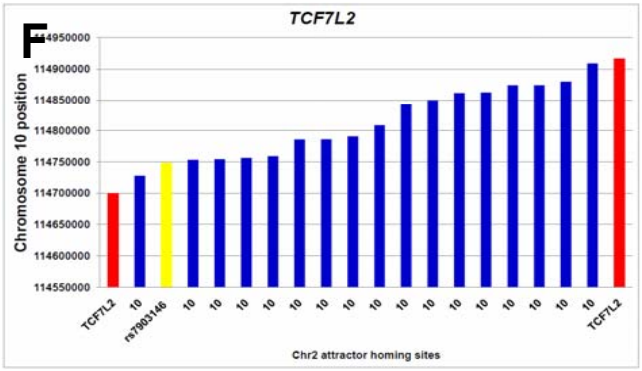


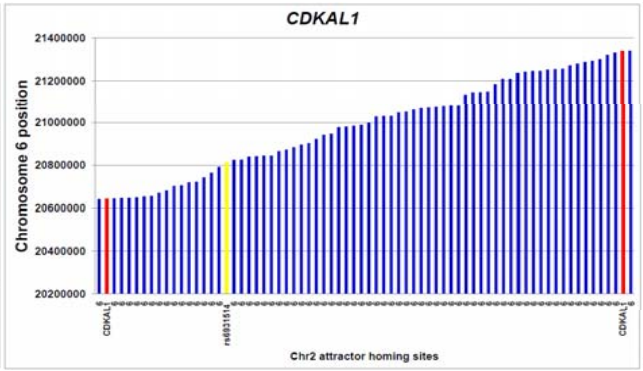
D

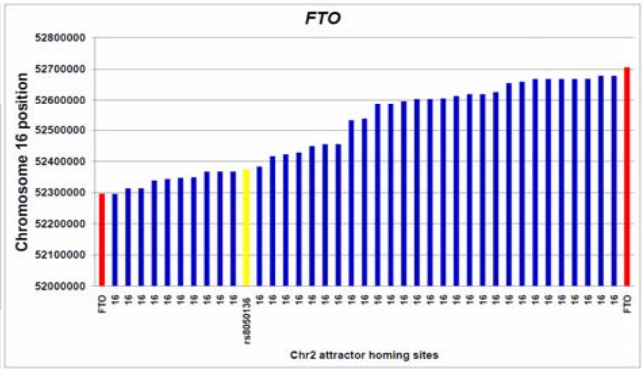
T2D SNP	Odds Ratio	Chromosome	Nearest gene
rs7903146	1.37	10	Intron of TCF7L2
rs13071168	1.35	3	Between SYN2 and PPARG
rs7020996	1.26	9	Near CDKN2A and CDKN2B
rs6931514	1.25	6	Intron of CDKAL1
rs4402960	1.17	3	Intron of IGF2BP2
rs8050136	1.15	16	Intron of FTO
rs10923931	1.13	1	Intron of NOTCH2
rs4580722	1.11	4	Near WFS1
rs12779790	1.11	10	Between CDC123 and CAMK1D
rs17705177	1.1	17	Near TCF2
rs864745	1.1	7	Intron of JAZF1

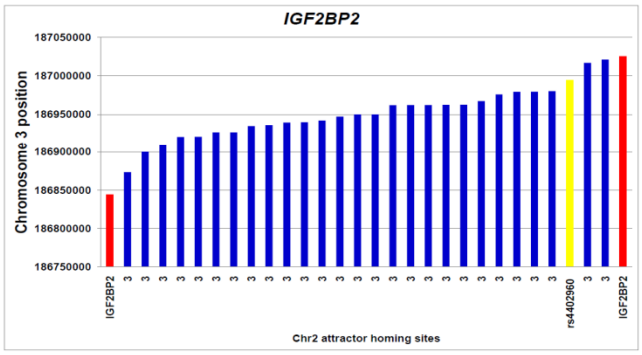


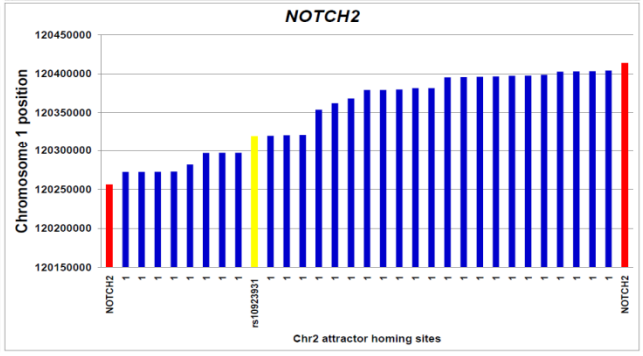


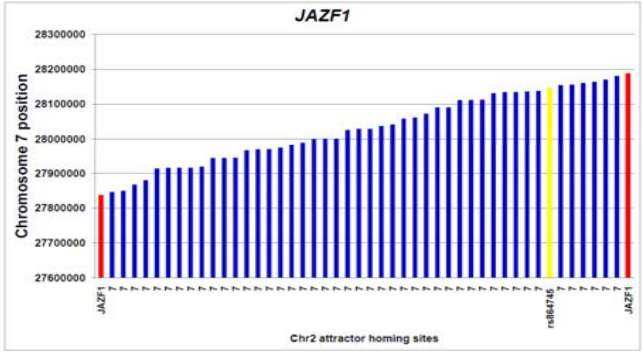




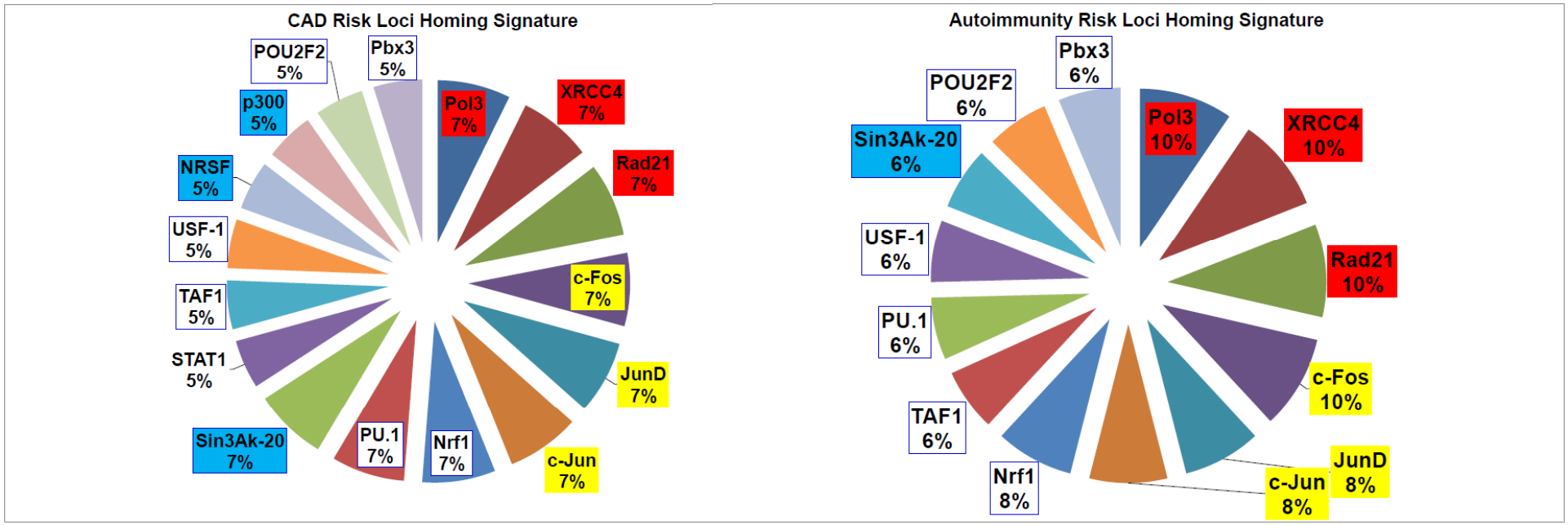








A



B