RSC Advances

Methylation of HTR3A promoter variant alters binding of transcription factor CTCF

Journal:	*RSC Advances*
Manuscript ID:	RA-ART-03-2015-004455.R1
Article Type:	Paper
Date Submitted by the Author:	11-May-2015

Complete List of Authors:	Jajodia, Ajay; Genomics and Molecular Medicine Unit, CSIR-Institute of Genomics and Integrative Biology, Singh, Kh.; Alagappa University, Department of Bioinformatics Singhal, Anshika; CSIR-IGIB, Genomic and Molecular Medicine Unit Vig, Saurabh; CSIR-IGIB, Genomic and Molecular Medicine Unit Datta, Malabika; CSIR-IGIB, Genomic and Molecular Medicine Unit Singh, Yogendra; CSIR-IGIB, Genomic and Molecular Medicine Unit Muthusamy, Karthikeyan; Alagappa University, Department of Bioinformatics Kukreti, Ritushree; Genomics and Molecular Medicine Unit, CSIR-Institute of Genomics and Integrative Biology, Genomic and Molecular Medicine Unit; igib,

Methylation of HTR3A promoter variant alters binding of transcription factor CTCF

Ajay Jajodia[1], Kh. Dhanachandra Singh[2], Anshika Singhal[1], Saurabh Vig[1], Malabika Datta[1], Yogendra Singh[1], Muthusamy Karthikeyan[2], and Ritushree Kukreti[1,*]

[1]Genomics and Molecular Medicine, CSIR-Institute of Genomics and Integrative Biology, Mall Road, Delhi-110007, India;

[2]Department of Bioinformatics, Alagappa University, Karaikudi, Tamil Nadu 630 004, India

[*] Correspondence should be addressed to: Ritushree Kukreti ,Genomics and Molecular Medicine, CSIR-Institute of Genomics and Integrative Biology, Mall Road, Delhi-110 007, India, Phone : 91-11-27662202, 27666156, Fax : 91-11-27667471,Email : ritus@igib.res.in

ABSTRACT

Genetic studies pertaining to effector molecules have been pivotal in schizophrenia research. Serotonin receptor HTR3A is an effector that plays a key role in schizophrenia development. Previously, we identified a promoter variant of *HTR3A*, rs1062613, to be associated with the disease in Indian population. The present study was undertaken to dissect the possible functional role of rs1062613. Using *in silico* simulation and *in vitro* gel shift assays, CCCTC-binding factor (CTCF) was found to bind at this variation site. A C/T polymorphism was found to affect the DNA binding of CTCF where CTCF binds less proficiently to the T-allele (alternate allele). Moreover, the binding was found to be dependent on methylation at the C-allele (reference allele). The CTCF was found to bind with a greater strength to methylated cytosine. Molecular dynamics simulation suggested the possible role of CTCF N-terminal in providing the binding flexibility. Our results suggest the role of epigenetic mechanisms in the development of schizophrenia by modulating transcription factor binding.

Keywords: CTCF; CpG methylation; rs1062613; Molecular dynamics simulation;

Introduction

Schizophrenia is a common debilitating neuropsychiatry disorder with lifetime prevalence of approximately 1% worldwide.[1] The vital characteristics include psychotic symptoms such as hallucinations, delusions, social withdrawal, and cognitive impairment.[2] Various family-based studies have unequivocally supported a strong predisposition to schizophrenia, with an estimated heritability of around 80%.[3] Still, these underlying genetic risk factors are yet to be recognized. Tremendous work involving genome wide association studies (GWASs), linkage studies, and candidate gene association studies have projected several candidate genes.[3] Many of these studies have successfully replicated loci considered to be associated with schizophrenia across different ancestral populations. Functional validation of applicable genetic variants is of utmost importance in describing the disease pathophysiology.

Previously, we identified a promoter variant of *HTR3A*, rs1062613, to be significantly associated with schizophrenia in Indian population.[4] *HTR3A* encodes for subunit A of type 3, 5-hydroxytryptamine (serotonin) receptors positioned on chromosome 11q23.1. It is a well recognized schizophrenia susceptibility gene and a potential response predictor to

antipsychotics.[5] Thus, *HTR3A* can be a potential risk gene for schizophrenia, especially in the context of Indian population. Studies utilising animal models have shown the prospective role of HTR3A in performing cognitive tasks.[6] This particular single nucleotide polymorphism (SNP) has been associated with various other disorders including, irritable bowel syndrome and nicotine dependence.[7,8]

SNP rs1062613 is a regulatory variant located 42 base pairs upstream of *HTR3A* gene.[9] Polymorphisms occurring in the promoter regions of the genes are known to affect gene expression and/or transcription factor (TF) binding affinities. Regulatory sequences of several genes contain various CpG sites and differentially methylated regions. Studies have focused on the role of DNA methylation in governing genome regulation during normal physiological processes[10,11] as well as during disease development.[12] DNA methylation is the most stable and appreciated epigenetic modification, altering the TF binding, which determines the expression of various developmentally important genes. Studies have implicated epigenetic process to be a contributing factor in defining etiology of psychotic disorders, including schizophrenia.[13] Therefore, we attempted to investigate the functional impact of the promoter SNP (rs1062613) using *in silico* and DNA-binding experiments. The alternate and the methylated forms of the variant have been analyzed and their effect on DNA binding activity of TF has been assessed.

Experimental

In silico analysis

To investigate the CpG site methylation at the rs1062613 (variation site), MEDIP-seq data was obtained from ENCODE data set of the postmortem human frontal cortex gray matter.[14]

Chromatin Immuno-Precipitation (ChIP)-seq co-ordinates for the putative TFs binding at *HTR3A* promoter were obtained from ENCODE data set deposited in UCSC Genome Browser.[15]

LASAGNA-Search[16] was employed to select TF that binds at variation site under investigation, utilizing a 41 nt sequence (5'-CTGGCCCTTGGTGGGCCTCG[C]CCTGAGCACTCGGAGGCACT-3') representing 20 nt flanking the variation site. The ORegAnno model of transcription factor binding sites was used for this analysis.

Gene manipulation

Total Human RNA was isolated from human hepatoma cell line (HepG2) and used for cDNA preparation by High capacity reverse transcription cDNA synthesis kit (Applied Biosystem) as per manufacturer's instructions. The cDNA was PCR amplified using CTCF specific primers (FP: CGCGGATCCCTTTGCAGCCACGGAGAG, RP: CCGCTCGAGAACACAGCCCAGAGAAGTCC) and Phusion High-Fidelity DNA Polymerase (ThermoScientific). The PCR product and the vector (pET28-His_{10}-Smt3) were double-digested with BamHI and XhoI restriction enzymes followed by ligation. The ligated product was chemically transformed into *E. coli* DH5α cells. The clones were confirmed by restriction digestion and DNA sequencing.

Recombinant protein purification

For over-expression of His_{10}-tagged CTCF, pET28-His_{10}-Smt3-CTCF was transformed in BL21 (DE3) cells and transformants were grown in LB medium containing kanamycin at 37°C until OD_{600} = ~0.6. The cultures were induced with 1 mM isopropyl thiogalactopyranoside (IPTG) for 12-14 hr at 16°C followed by cell harvesting. All the purification procedures were performed at 4°C.

For purification, Ni-NTA affinity chromatography was performed using an earlier protocol.[17] Briefly, the harvested cells were dissolved and sonicated in lysis buffer (50 mM Tris-Cl [pH 8.5], 300 mM NaCl, 5 mM β-mercaptoethanol, 1 mM PMSF and 1× protease inhibitor cocktail [Roche]). After removing the cytosolic fraction by centrifugation, the protein from inclusion bodies was recovered using solubilization buffer (50 mM Tris-Cl [pH 8.5], 300 mM NaCl, 1.5% N-lauryl sarcosine, 25 mM triethanolamine, 1% Triton X-100, 5 mM β-mercaptoethanol, 1 mM PMSF and 1× protease inhibitor cocktail). After centrifugation (16000 rpm, 30 min, 4°C), the supernatant was incubated with Ni^{2+}-NTA affinity resin (Qiagen) pre-equilibrated with lysis buffer. The column was washed extensively with wash buffer (50 mM Tris-Cl [pH 8.5], 1 M NaCl, 5 mM β-mercaptoethanol, 20 mM imidazole, 10% glycerol and 1 mM PMSF). Desired His_{10}-tagged CTCF was obtained using elution buffer (50 mM Tris-Cl [pH 8.5], 150 mM NaCl, 10% glycerol, 1 mM PMSF and 200 mM imidazole). Purified protein was run on 10% SDS-PAGE and analyzed by Coomassie staining.

Gel shift assay

DNA binding activity of CTCF was assessed by gel shift assay using a 21-nucleotide long probe encompassing the variant site at the centre. This sequence represents the promoter region of *HTR3A*, 42 base pair upstream of the transcription start site. The sequences of the oligonucleotides used as templates were as follows:

C-allele: 5'-GTGGGCCTCG[C]CCTGAGCACT-3',

T-allele: 5'- GTGGGCCTCG[T]CCTGAGCACT-3',

Methylated-allele: 5'- GTGGGCCTCG[C-CH_3]CCTGAGCACT-3'.

Similar protocol was followed as described earlier.[18] For annealing, sense and antisense oligonucleotides were mixed at equimolar concentration in annealing buffer (100 mM Tris-Cl [pH 7.5], 1 M NaCl and 10 mM EDTA). The mixture was heated for 5 min at 95°C and subsequently cooled to 25°C in step-down manner at a rate of -1°C/30s. [γ-^{32}P] labeled DNA probe was prepared by end labeling the annealed oligonucleotides using Polynucleotide kinase (Roche) as per manufacturer's guidelines. The reaction was stopped using 1 mM EDTA. Labeled probe was purified using nucleotide removal kit (Qiagen) and eluted in Tris-EDTA buffer. Equal amounts of the three labeled probes were incubated with 5-15 µg of CTCF at 25°C for 20 min in EMSA binding buffer (25mM HEPES [pH 7.0], 50 mM KCl, 6.25 mM $MgCl_2$, and 5% glycerol) in a total volume of 20 µl according to previously published protocol.[19] Control reactions were setup in the absence of CTCF. After the incubation, 6× non-denaturing gel loading dye were added to the samples. The DNA:Protein complex and the free DNA probes were resolved by 5% non-denaturing polyacrylamide gel in running buffer (0.5× TBE). Gels were then dried and subjected to autoradiography using Personal Molecular Imager (Bio-Rad).

Molecular docking of CTCF bound to DNA

The 3D-structure of CTCF protein (PDB ID: 2CT1) was retrieved from PDB (www.rcsb.org/). Water molecules were deleted manually. The structure was finally energy minimized using NOMAD-Ref server (http://lorentz.immstr.pasteur.fr/gromacs/minimization_submission. php). The starting structures for the docking were a B-form model of the double helix DNA fragments. The DNA sequences used for docking were as follows:

C-allele: 5'-GTGGGCCTCG[C]CCTGAGCACT-3',

T-allele: 5'- GTGGGCCTCG[T]CCTGAGCACT-3',

Methylated-allele: 5'- GTGGGCCTCG[C-CH_3]CCTGAGCACT-3'.

The DNA templates were constructed using 3D-DART (http://haddock.Science.uu.nl/dna/dna.php).[20]

CTCF protein (DNA binding domain) and DNA residues were defined as those that both underwent significant changes in chemical shift (Figure S1). Since CCCTC is the binding site for CTCF factor, DNA residues 11-14 were defined as active sites for protein binding. For CTCF, C18, C21, H34, H39, C50, C53, H66 and C70 were defined as active sites as defined in its crystal structure. Passive residues were defined automatically which are around the active residues for both protein and DNA. Additional restraints to maintain base planarity and Watson Crick bonds were introduced for the DNA. During the rigid body energy minimization, 1000 structures were calculated, and the 200 best solutions based on the intermolecular energy were used for the semi-flexible, simulated annealing followed by an explicit water refinement. The solutions were clustered using a cutoff of 3.5 Å based on the pair wise backbone RMSD matrix. The semi-flexible annealing and the water refinement steps of HADDOCK were re-run with the best five structures of the lowest energy clusters (cutoffs 0.9 Å). The final 120 structures were clustered, resulting in a single low energy cluster of 23 structures. The best 10 structures (RMSD 0.7 Å over backbone atoms) of this cluster were analyzed using standard HADDOCK protocols and were used to represent a model of the complex.

Molecular Dynamics Simulation

Molecular Dynamics (MD) simulations of CTCF protein with the DNA representing C-, T- and me-C- alleles in explicit water at 300 K were performed using the Gromacs 4.5.6[21] and the Amber force field.[22] All complex molecules were solvated by cubic boxes of TIP3P water molecules[23] having dimensions 9.60 × 9.60 × 9.60 for C-allele, 9.74 × 9.74 × 9.74 for T-allele and 10.73 × 10.73 × 10.73 for me-C allele. A total of 28797, 29870, and 40236 solvent molecules were added respectively to the protein-ligand complexes. The systems were equilibrated by 5000 steps of energy minimization, followed by a 250 ps MD simulation in the NVT ensemble, with harmonic restraints (20 kcal mol^{-1} $Å^{-2}$) applied to the backbone atoms of the biomolecules.

The temperature was maintained at 300K by Langevin dynamics with damping factor equal to 5 ps. Periodic boundary conditions were applied, and the pressure was kept at 1 atm by the Nose-Hover Langevin method, with an oscillation period of 200 fs and a damping time of 100 fs.[24, 25]

A smoothed cut-off (10-12 Å) was used for the van der Waals interactions. Electrostatic forces were computed by the Particle Mesh Ewald algorithm with a maximum grid spacing of 1.0 Å. Bonds with hydrogen atoms were restrained by the SETTLE algorithm[26,27] to use a time step of 2 fs.

Results

In silico analysis: cytosine methylation at rs1062613

For analyzing the functional significance of rs1062613, we used the methylated DNA immunoprecipitation-sequencing (MeDIP-seq) data of the human brain sample present in the ENOCDE dataset.[15] The data shows that rs1062613 is a site for cytosine methylation (Figure 1). Presence of methylation suggests the possibility of epigenetic mechanisms for the regulation of *HTR3A*. The change from C (reference allele) to T (alternate allele) in rs1062613 (variation site) may affect its methylation and thus, may influence the binding affinity of transcription factors and promoter activity. To detect the possible transcription factors binding at this site, we utilized two independent approaches. In the first approach, the available Chromatin Immuno-Precipitation (ChIP) data from ENCODE was utilised.[15] This analysis showed the presence of putative binding sites for 6 transcription factors (TFs)- E2F6, ZNF263, CTCF, Max, YY1_(C-20), and TAF1. For the second approach, we used LASAGNA-Search.[16] A 41-nucleotide DNA sequence, 20 nucleotides flanking the variation site from both sides, was used to predict TF-binding. The analysis predicted significant binding (p value<0.005) of only CCCTC-binding factor (CTCF) exactly at the variation site, i.e. rs1062613. Therefore, the binding of CTCF to this DNA sequence was further analyzed in detail.

DNA binding assay of CTCF

To validate the binding of CTCF to the *HTR3A* promoter region, DNA binding assays were carried out using gel shift assay. For this, a 21-nucleotide DNA region encompassing rs1062613 variant was used as probe. The C and T DNA sequences represented reference- and alternate-alleles, respectively. A third DNA sequence containing methyl-cytosine (methylated-allele) at the variation site was generated. CTCF protein was over-expressed and purified from *E. coli* as a recombinant His-tagged fusion protein. Increasing concentration of CTCF protein was used to form complexes with radiolabelled DNA probes and the complexes are analyzed by

autoradiography. The assay (Figure 2) shows significant binding of CTCF to the C-allele. The mutation of cytosine to thymine results in compromised binding of CTCF protein to the DNA. Interestingly, presence of a methyl group at this cytosine enhanced DNA binding of CTCF. The intensity of DNA-protein complex increased in a concentration-dependent manner when the cytosine is methylated.

Molecular docking of DNA:protein complex

To gain insight into the likely DNA and CTCF binding mode, we performed docking of the three DNA sequences (C, T and me-C) in complex with CTCF, using HADDOCK program. Following a two-stage docking, simulated annealing, and water refinement protocol, the ten lowest energy structures were selected to represent a model of the complex structure (Figure 3, S2-S4 and S6). Docking shows that CTCF binds to the DNA sequences representing C- and T- alleles, overlapping the variation site (Figure 3A and 3B). On the other hand, CTCF binds to the methylated DNA with a completely different orientation. CTCF inserts into the major groove of the DNA which is opposite to the methyl-Cytosine (Figure 3C). In order to understand a clearer picture of CTCF binding with DNA, we further performed molecular dynamics (MD) simulation for the three different complexes namely C-, T-, and me-C-, allele:CTCF complexes.

Molecular dynamics simulation

To obtain a dynamic picture of the conformational changes occurring in an aqueous solution, and to explore the binding affinities of the three DNA sequences with the CTCF protein, MD simulation was performed (Supplementary Video 1-3). Backbone root mean square deviations (RMSD), principal component analysis (PCA), and number of hydrogen bonds (H-bonds) were analyzed during 50 ns simulation time. During the simulation, it was found that CTCF N-terminal region is flexible throughout the simulation (Figure S5).

The RMSD of the trajectory for C-allele:CTCF complex with respect to the initial structure (black line in Figure 4A) shows that the RMSD increases up to approximately 10 Å and then plateaued near 45 ns. This suggests that a relatively stable conformation of protein is achieved through the MD simulation. The fluctuations observed during the initial period could be due to high flexibility of CTCF N-terminal during the course of MD simulation (Figure 4A and S1A). For T-allele:CTCF complex, RMSD plateaued at about 6 Å which indicates that CTCF protein

has less conformational change when it binds to the DNA sequence containing alternate allele (Figure 4A and S1B). In case of me-C-allele:CTCF complex, the RMSD stabilizes near 5 Å, suggesting that CTCF protein has least conformational change when it binds with the methylated DNA (Figure 4A and S1C). We next calculated the binding free energy of all these complexes where, me-C-allele:CTCF complex was found to have the highest free energy followed by C-allele:CTCF complex and T-allele:CTCF complex (Table 1). This observation again suggested a more stable binding of CTCF with the methylated DNA.

Effect of cytosine methylation on the H-bond between DNA and CTCF

Two atoms are considered to form a hydrogen bond (H-bond) if they are closer than 3.0 Å and if the donor-hydrogen-acceptor angle is lower than 30°. The average number of H-bonds was found to be 10 for both C-allele:CTCF complex and T-allele:CTCF complex; and 15 for meC-allele:CTCF complex. Interestingly, after 18 ns of the simulation, number of H-bonds increases in meC-allele:CTCF complex, while there is a decrease in the C-allele:CTCF and T-allele:CTCF complexes (Figure 4B).

The initial docking experiments had suggested that the C-allele:CTCF and T-allele:CTCF complexes have a higher number of H-bonds than the meC-allele:CTCF complex (Figure 3). But during the course of MD simulation, a reduction in the number of H-bonds was observed in the C-allele:CTCF and T-allele:CTCF complexes. This observation suggests that the interactions observed for the C- and T-alleles in the docking experiments were probably weak. Thus, it can be concluded that the binding of CTCF to the methylated allele is stronger and more stable. Such strong binding may be attributed to the N-terminal region of CTCF.

Principal component analysis

Principal component analysis (PCA) was used to compare the global motions of atoms of the CTCF protein when it is bound to the three DNA sequences. Projections of the MD trajectories obtained at 298 K on to the principal components (PC1 and PC2) mapping motion of the three complexes are shown in Figure 5. The result shows that the first two eigenvectors account for more than 90% of the collective motions of backbone protein atoms. A comparison of data on the distribution of the point clusters shows that CTCF bound with C-allele occupy more space, followed by CTCF bound to the T-allele. CTCF bound to the methylated DNA occupy the

minimum space. The result also reveals that CTCF bound with C-allele has increased the sampling of phase space coverage. Thus, the result indicates that the C-allele:CTCF complex has a higher degree of flexibility than the T-allele:CTCF complex and meC-allele:CTCF complex.

Discussion

Serotonin system of the brain has a critical role in the development of psychiatric disorders.[28] HTR3 is a serotonin ion-gated channel composed of two subunits (HTR3A and HTR3B) and mediates fast depolarization events of excitatory response. HTR3A localizes in limbic region that comprises amygdala and hippocampus, and has been suggested to have an involvement in cognition, anxiety, and functionality of the hypothalamic-pituitary-adrenal (HPA) axis. HTR3A can form homomeric ion channels[29] and it variant has been associated with neuropsychiatry disorders like bipolar affective disorder[9] and personality traits[30], and also modulates amygdaloidal activity in normal human subjects.[31]

Our previous study reported a significant association of rs1062613 (p=0.019) variant with schizophrenia.[4] The minor allele (T-allele) of the variant rs1062613 is associated with the disease susceptibility as found in the combined analysis of case-control and familial based association study in South Indian population. The same direction of effect was found for the minor allele in a meta-analysis of two independent North and South Indian populations.[4] To understand the mechanism by which the T-allele is associated with the disease, the present study was undertaken and it was found that the T-allele results in a compromised binding of CTCF.

CTCF is an essential transcription factor that is ubiquitously expressed[32] and highly conserved in higher eukaryotic organisms. It is involved in a multitude of cellular processes, including regulation of gene expression, organization of genome topology, genome imprinting, hormone-responsive silencing, enhancer-promoter interactions, and chromatin interactions and its subnuclear localization[33]. It has a diverse regulatory mechanism and is implicated in both activation and impediment of gene expression.[34, 35]

CTCF binding at promoter region is ubiquitous, whereas binding at the enhancer region is tissue-specific.[36] CTCF consensus binding sequence contains CpG dinucleotide and that is subjected to methylation at carbon 5 of the nitrogen base to form 5-methylcytosine (5mc).[37] DNA methylation has been hypothesized to regulate the tissue-specific binding of CTCF.[38] Additionally, it has been reported that 41% of variable CTCF binding sites are linked to DNA

methylation.[38] It shows a preferential binding to unmethylated sequences[39]; however, it can also bind to methylated DNA sequences in some cases.[40] Such methylation-specific DNA binding has been previously observed for the tumor suppressor protein p53.[41]

In the present study, our functional assay suggests that rs10623613 can influence the affinity of CTCF binding to the promoter region of *HTR3A*. Presence of C-allele and its subsequent methylation may accompany a greater binding of CTCF. The T-allele has been previously shown to be associated with increased promoter activity than C-allele.[9] Therefore, it can be hypothesized that CTCF binds to HTR3A promoter and is responsible for its repression. Since HTR3A receptors have been shown to be associated with the levels of adrenocorticotropic hormone (ACTH), rs10623613 may potentially modulate the expression of this stress hormone.[42] Interestingly, there are reports of elevated ACTH levels in depression and anxiety disorder cases.[43]

In conclusion, we report the functional importance of a well-known genetic variant rs10623613, in schizophrenia and changes in the specific binding affinities of CTCF transcription factor. The methylation of CpG could be of utmost importance for gene regulatory events in this regards. Abnormal methylation is established in many compromised human conditions, and treating them pharmacologically is a progressive research direction in the field of epigenetic therapy.

Acknowledgments

We thank Prof. Stewart Shuman and Dr. K M Sinha for pET28-His_{10}-Smt3 plasmid.

Competing Interests

The authors declare no competing interests.

Contributors

AJ and RK conceived and designed the experiments. AJ, AS and SV performed the gel shift assays. KDS performed the molecular dynamics simulations. AJ, KDS, AS, and SV analyzed the data. AJ, KDS and AS wrote the manuscript. RK, MK, YS and MD provided the financial support and resources. All authors read and approved the manuscript.

Funding

The work was supported by grants from the Council of Scientific and Industrial Research (BSC0123) and Department of Biotechnology, Ministry of Science and Technology

(GAP0040). AJ, KDS and AS are thankful to Council of Scientific and Industrial Research (CSIR) and SV acknowledge Indian Council of Medical Research (ICMR) for providing senior research fellowship.

References

1. A. Jablensky, N. Sartorius, A. Korten, G. Ernberg, M. Anker, J. E. Cooper and R. Day, *The British journal of psychiatry : the journal of mental science*, 1987, 151, 408-409.
2. N. C. Andreasen, *Lancet*, 1995, 346, 477-481.
3. R. Tandon, M. S. Keshavan and H. A. Nasrallah, *Schizophrenia research*, 2008, 102, 1-18.
4. A. Jajodia, H. Kaur, K. Kumari, M. Gupta, R. Baghel, A. Srivastava, M. Sood, R. K. Chadda, S. Jain and R. Kukreti, *Schizophrenia research*, 2015, DOI: 10.1016/j.schres.2014.12.031.
5. A. P. Rajkumar, B. Poonkuzhali, A. Kuruvilla, A. Srivastava, M. Jacob and K. S. Jacob, *Psychopharmacology*, 2012, 224, 441-449.
6. N. M. Barnes and T. Sharp, *Neuropharmacology*, 1999, 38, 1083-1152.
7. L. A. Kilpatrick, J. S. Labus, K. Coveleskie, C. Hammer, G. Rappold, K. Tillisch, J. A. Bueller, B. Suyenobu, J. M. Jarcho, J. A. McRoberts, B. Niesler and E. A. Mayer, *Gastroenterology*, 2011, 140, 1943-1951.
8. Z. Yang, C. Seneviratne, S. Wang, J. Z. Ma, T. J. Payne, J. Wang and M. D. Li, *Drug and alcohol dependence*, 2013, 129, 217-225.
9. B. Niesler, T. Flohr, M. M. Nothen, C. Fischer, M. Rietschel, E. Franzek, M. Albus, P. Propping and G. A. Rappold, *Pharmacogenetics*, 2001, 11, 471-475.
10. Z. D. Smith and A. Meissner, *Nature reviews. Genetics*, 2013, 14, 204-220.
11. M. J. Ziller, H. Gu, F. Muller, J. Donaghey, L. T. Tsai, O. Kohlbacher, P. L. De Jager, E. D. Rosen, D. A. Bennett, B. E. Bernstein, A. Gnirke and A. Meissner, *Nature*, 2013, 500, 477-481.
12. Y. Bergman and H. Cedar, *Nature structural & molecular biology*, 2013, 20, 274-281.
13. V. Labrie, S. Pai and A. Petronis, *Trends in genetics : TIG*, 2012, 28, 427-435.

14. A. K. Maunakea, R. P. Nagarajan, M. Bilenky, T. J. Ballinger, C. D'Souza, S. D. Fouse, B. E. Johnson, C. Hong, C. Nielsen, Y. Zhao, G. Turecki, A. Delaney, R. Varhol, N. Thiessen, K. Shchors, V. M. Heine, D. H. Rowitch, X. Xing, C. Fiore, M. Schillebeeckx, S. J. Jones, D. Haussler, M. A. Marra, M. Hirst, T. Wang and J. F. Costello, *Nature*, 2010, 466, 253-257.
15. E. P. Consortium, *Nature*, 2012, 489, 57-74.
16. C. Lee and C. H. Huang, *BioTechniques*, 2013, 54, 141-153.
17. A. Singhal, G. Arora, A. Sajid, A. Maji, A. Bhat, R. Virmani, S. Upadhyay, V. K. Nandicoori, S. Sengupta and Y. Singh, *Scientific reports*, 2013, 3, 2264.
18. S. Vig, A. K. Pandey, G. Verma and M. Datta, *The international journal of biochemistry & cell biology*, 2012, 44, 113-122.
19. M. Renda, I. Baglivo, B. Burgess-Beusse, S. Esposito, R. Fattorusso, G. Felsenfeld and P. V. Pedone, *The Journal of biological chemistry*, 2007, 282, 33336-33345.
20. M. van Dijk and A. M. Bonvin, *Nucleic acids research*, 2009, 37, W235-239.
21. D. Van Der Spoel, E. Lindahl, B. Hess, G. Groenhof, A. E. Mark and H. J. Berendsen, *J Comput Chem*, 2005, 26, 1701-1718.
22. E. Carletti, H. Li, B. Li, F. Ekstrom, Y. Nicolet, M. Loiodice, E. Gillon, M. T. Froment, O. Lockridge, L. M. Schopfer, P. Masson and F. Nachon, *Journal of the American Chemical Society*, 2008, 130, 16011-16020.
23. J. C. William L Jorgensen, Jeffry D Madura, Roger W Impey, Michael L Klein, *The Journal of chemical physics*, 1983, 79, 926-935.
24. W. G. Hoover, *Physical review. A*, 1985, 31, 1695-1697.
25. S. Nosé, *Molecular Physics: An International Journal at the Interface Between Chemistry and Physics*, 1984, Volume 52, 255-268.
26. H. B. Berk Hess, Herman J. C. Berendsen and Johannes G. E. M. Fraaije, *Journal of Computational Chemistry*, 1997, Volume 18, 1463–1472.
27. M. N. Kawata, Umpei, *Chemical Physics Letters*, 2001, Volume 340.
28. A. Caspi, K. Sugden, T. E. Moffitt, A. Taylor, I. W. Craig, H. Harrington, J. McClay, J. Mill, J. Martin, A. Braithwaite and R. Poulton, *Science*, 2003, 301, 386-389.
29. P. A. Davies, M. Pistis, M. C. Hanna, J. A. Peters, J. J. Lambert, T. G. Hales and E. F. Kirkness, *Nature*, 1999, 397, 359-363.

30. J. Melke, L. Westberg, S. Nilsson, M. Landen, H. Soderstrom, F. Baghaei, R. Rosmond, G. Holm, P. Bjorntorp, L. G. Nilsson, R. Adolfsson and E. Eriksson, *Archives of general psychiatry*, 2003, 60, 1017-1023.
31. T. Iidaka, N. Ozaki, A. Matsumoto, J. Nogawa, Y. Kinoshita, T. Suzuki, N. Iwata, Y. Yamamoto, T. Okada and N. Sadato, *The Journal of neuroscience : the official journal of the Society for Neuroscience*, 2005, 25, 6460-6466.
32. H. Heath, C. Ribeiro de Almeida, F. Sleutels, G. Dingjan, S. van de Nobelen, I. Jonkers, K. W. Ling, J. Gribnau, R. Renkawitz, F. Grosveld, R. W. Hendriks and N. Galjart, *The EMBO journal*, 2008, 27, 2839-2850.
33. J. E. Phillips and V. G. Corces, *Cell*, 2009, 137, 1194-1211.
34. A. Baniahmad, C. Steiner, A. C. Kohne and R. Renkawitz, *Cell*, 1990, 61, 505-514.
35. V. V. Lobanenkov, R. H. Nicolas, V. V. Adler, H. Paterson, E. M. Klenova, A. V. Polotskaja and G. H. Goodwin, *Oncogene*, 1990, 5, 1743-1753.
36. Y. Shen, F. Yue, D. F. McCleary, Z. Ye, L. Edsall, S. Kuan, U. Wagner, J. Dixon, L. Lee, V. V. Lobanenkov and B. Ren, *Nature*, 2012, 488, 116-120.
37. C. T. Ong and V. G. Corces, *Nature reviews. Genetics*, 2014, 15, 234-246.
38. H. Wang, M. T. Maurano, H. Qu, K. E. Varley, J. Gertz, F. Pauli, K. Lee, T. Canfield, M. Weaver, R. Sandstrom, R. E. Thurman, R. Kaul, R. M. Myers and J. A. Stamatoyannopoulos, *Genome research*, 2012, 22, 1680-1688.
39. A. T. Hark, C. J. Schoenherr, D. J. Katz, R. S. Ingram, J. M. Levorse and S. M. Tilghman, *Nature*, 2000, 405, 486-489.
40. M. B. Stadler, R. Murr, L. Burger, R. Ivanek, F. Lienert, A. Scholer, E. van Nimwegen, C. Wirbelauer, E. J. Oakeley, D. Gaidatzis, V. K. Tiwari and D. Schubeler, *Nature*, 2011, 480, 490-495.
41. M. Petrovich and D. B. Veprintsev, *Journal of molecular biology*, 2009, 386, 72-80.
42. S. Bhatnagar, L. M. Sun, J. Raber, S. Maren, D. Julius and M. F. Dallman, *Physiology & behavior*, 2004, 81, 545-555.
43. E. A. Young, J. L. Abelson and O. G. Cameron, *Biological psychiatry*, 2004, 56, 113-120.

Table 1: Relative binding free energies between three different DNA: Protein complexes.

Energies (kJ/mol)	C-allele	T-allele	Me-C allele
Binding Energy*	-12372.71 ± 2811.97	-11281.29 ± 2648.94	-14063.32 ± 525.62
Van der Waal*	-188.82 ± 64.56	-167.10 ± 64.10	-308.22 ± 56.23
Electrostatic*	-12183.89 ± 2755.40	-11114.19 ± 2598.60	-13755.09 ± 500.39

*Values are expressed as Mean ± SD

Figure legends

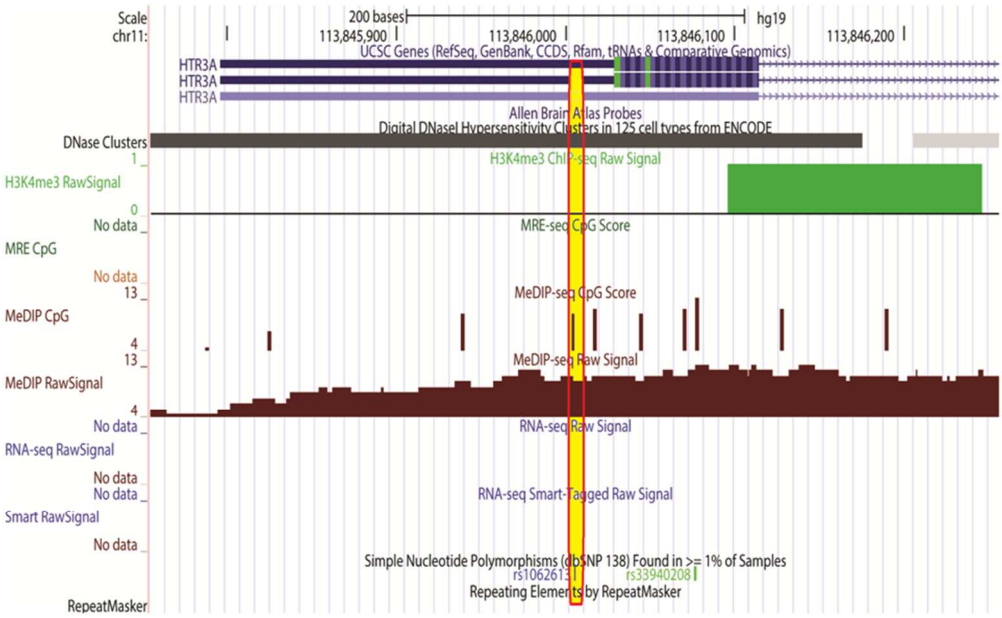
Figure 1: *In silico* analysis of the surrounding region of rs1062613. The SNP lies in the upstream regulatory region of *HTR3A* gene. A methylation mark was observed at this site according to the MeDIP-seq data of Human brain sample.

Figure 2: DNA binding assay of CTCF. Autoradiogram of gel shift assay performed using increasing amounts (0-15 µg) of CTCF and oligo probes containing C-allele (lanes 1-4), T-allele (lanes 5-8) and methyl-C allele (lanes 9-12). The bands corresponding to DNA:protein complex have been marked.

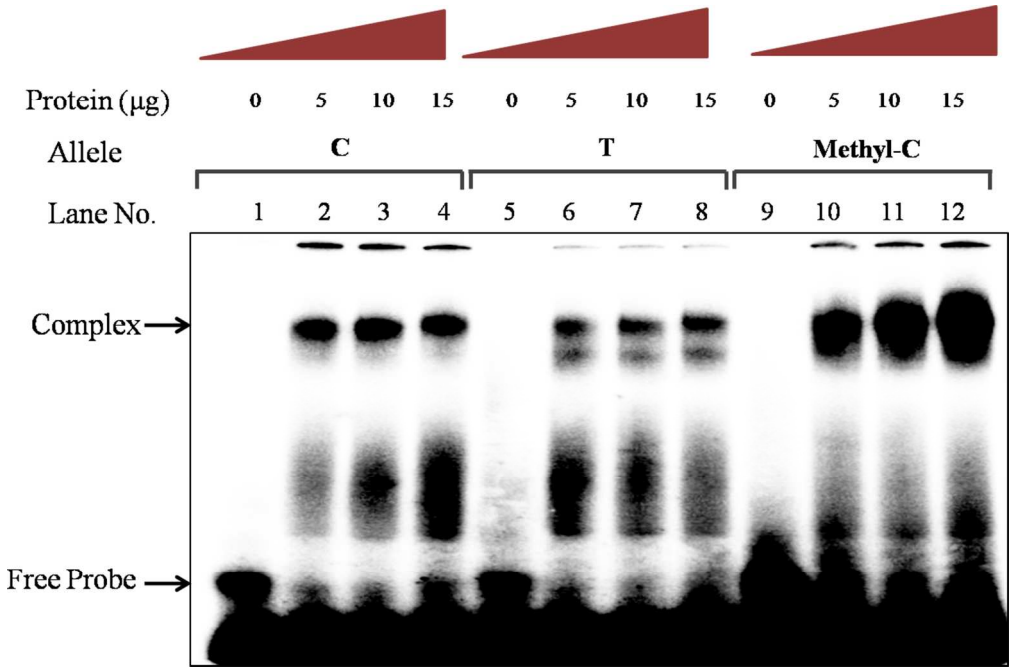
Figure 3: Molecular docked model of DNA and protein complex. (A) C-allele:CTCF complex. (B) T-allele:CTCF complex. (C) me-C:CTCF complex.

Figure 4: Comparative MD analysis of DNA-protein complex. (A) Backbone RMSD plot as a function of time (ns). (B) Hydrogen bonds between DNA:CTCF complex over simulation time of 50 ns. Color representations are: C-allele (Black), T-allele (Red) and me-C allele (Green).

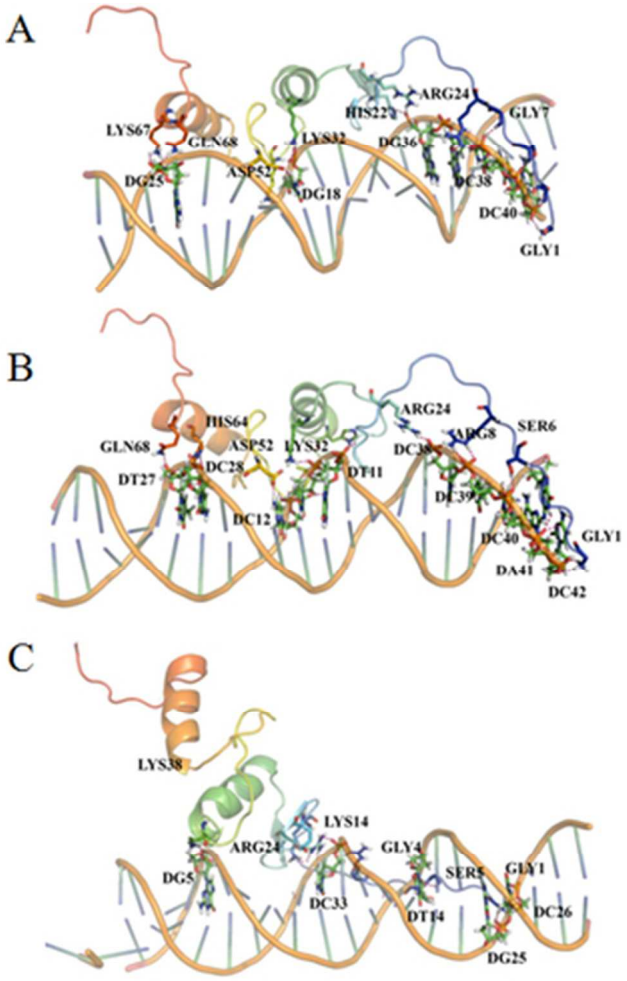
Figure 5: The principal component analysis. Projection of most significant principal components of motion of the atoms of CTCF. The trajectory projected to the two dimensional space. Color representations are: C-allele (Black), T-allele (Red) and me-C allele (Green).



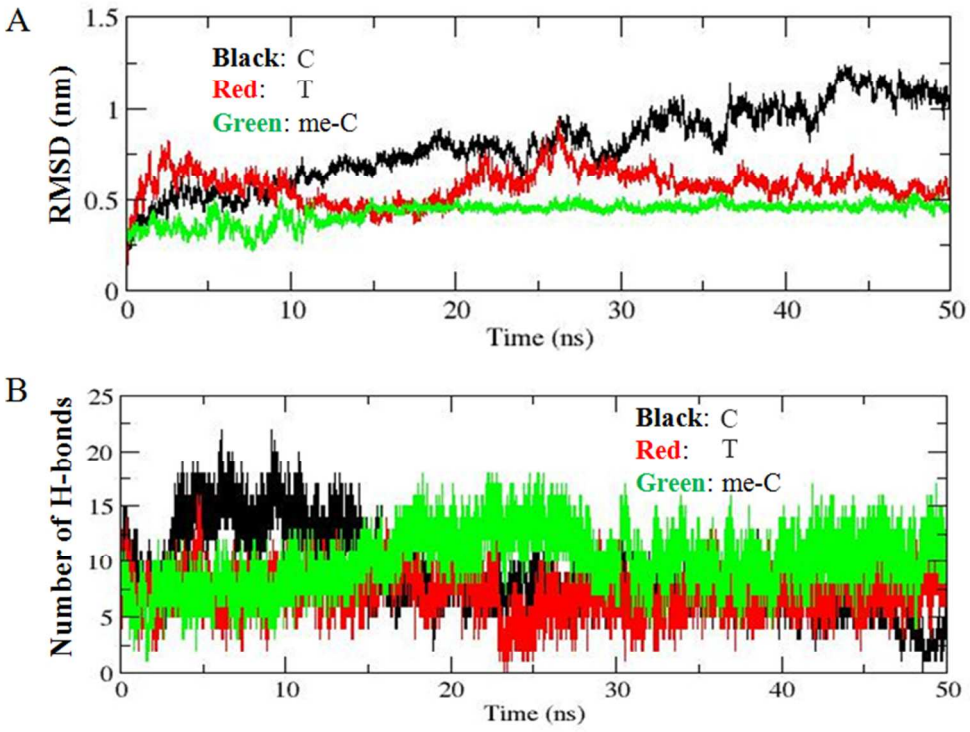
In silico analysis of the surrounding region of rs1062613. The SNP lies in the upstream regulatory region of HTR3A gene. A methylation mark was observed at this site according to the MeDIP-seq data of Human brain sample.
81x49mm (300 x 300 DPI)



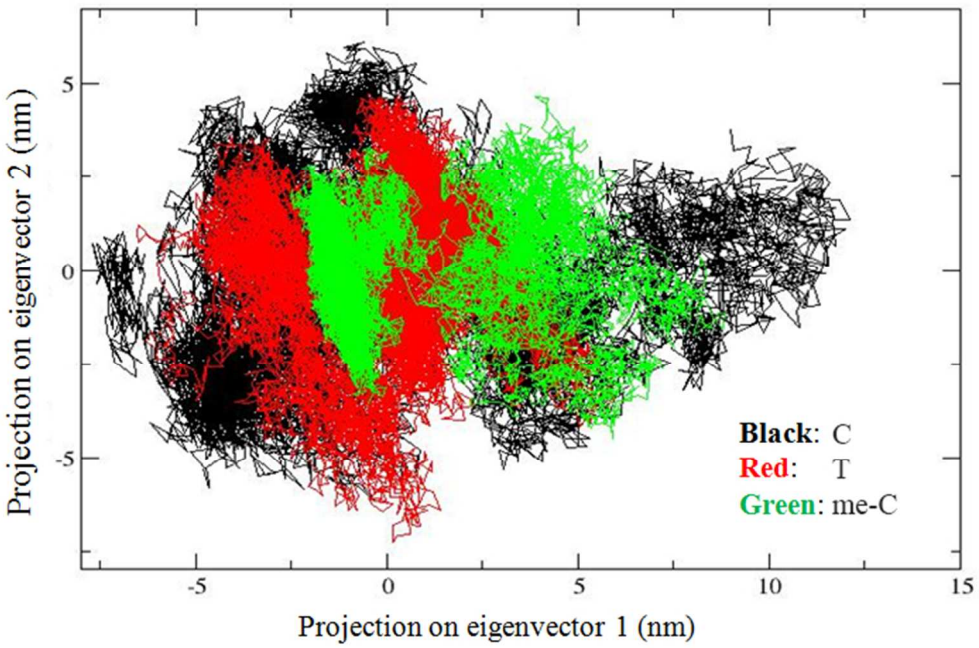
DNA binding assay of CTCF. Autoradiogram of gel shift assay performed using increasing amounts (0-15 μg) of CTCF and oligo probes containing C-allele (lanes 1-4), T-allele (lanes 5-8) and methyl-C allele (lanes 9-12). The bands corresponding to DNA:protein complex have been marked.
213x141mm (300 x 300 DPI)



Molecular docked model of DNA and protein complex. (A) C-allele:CTCF complex. (B) T-allele:CTCF complex. (C) me-C:CTCF complex.
27x41mm (300 x 300 DPI)



Comparative MD analysis of DNA-protein complex. (A) Backbone RMSD plot as a function of time (ns). (B) Hydrogen bonds between DNA:CTCF complex over simulation time of 50 ns. Color representations are: C-allele (Black), T-allele (Red) and me-C allele (Green).
65x49mm (300 x 300 DPI)



The principal component analysis. Projection of most significant principal components of motion of the atoms of CTCF. The trajectory projected to the two dimensional space. Color representations are: C-allele (Black), T-allele (Red) and me-C allele (Green).
59x40mm (300 x 300 DPI)