Spectrum of *CREBBP* mutations in Indian patients with Rubinstein–Taybi syndrome

Neeti Sharma, Avinash M Mali and Sharmila A Bapat*
National Centre for Cell Science, NCCS Complex, Pune University Complex, Ganeshkhind, Pune 411 007, India
**Corresponding author (Fax, 91-020-25692259; Email, sabapat@nccs.res.in)*

Rubinstein–Taybi syndrome (RSTS), a developmental disorder comprising abnormalities that include mental retardation, an unusual facial appearance, broad thumbs and big toes is frequently associated with molecular lesions in the CREB-binding protein gene, *CREBBP*. The objective of the present study was to identify and analyse *CREBBP* mutations in Indian RSTS patients on which there are no data. Direct sequencing of *CREBBP* performed in 13 RSTS patients identified the three zinc fingers (CH1, CH2, CH3) and HAT domain as mutational hotspots in which ten novel pathogenic mutations were localized. Functional analysis revealed that three of these mutations affecting amino acids Glu1459, Leu1668 and Glu1724 were critical for histone acetyltransferase activity. Twenty-eight novel *CREBBP* single-nucleotide polymorphisms (SNPs) were also identified in the Indian population. Linkage disequilibrium studies revealed associations between (i) SNP (*rs129974*/c.3836-206G>C) and mutation (p.Asp1340Ala); (ii) (*rs130002*) with mutation (p.Asn435Lys) and (iii) SNPs *rs129974, rs130002* and SNP (*c.3836*-206G>C) signifying a disease affection status. In conclusion, the present study reports the highest detection rate of *CREBBP* mutations (76.9%) in RSTS patients to date, of which ten are predicted to be pathogenic and three critical for histone acetyltransferase activity. Moreover, identification of the association of *CREBBP* polymorphisms with disease susceptibility could be an important risk factor for the pathogenesis of RSTS.

[Sharma N, Mali A M and Bapat S A 2010 Spectrum of *CREBBP* mutations in Indian patients with Rubinstein–Taybi syndrome; *J. Biosci.* **35** 187–202] **DOI** 10.1007/s12038-010-0023-5

1. Introduction

Rubinstein–Taybi syndrome (RSTS) is a congenital disorder. The associated symptoms described earlier by Michail *et al.* (1957) were precisely delineated later as a syndrome by Rubinstein and Taybi (1963). Broad, angulated thumbs and halluces are considered to be the hallmark of the syndrome. Other specific symptoms include growth and mental retardation, and a wide range of dysmorphic features and skeletal abnormalities. Mutations and chromosomal rearrangements in the gene encoding the CREB-binding protein (*CREBBP/CBP*) located on 16p13.3 are reported to be the causative lesions in several cases of RSTS (Petrij *et al.* 1995, 2000b; Giles *et al.* 1997), although a total association with pathogenic mutations in *CREBBP* has not been reported.

Human CREBBP is a nuclear protein consisting of 2442 amino acids; predicted domains include the nuclear receptor-binding and receptor-interacting domain (NHRD, aa 1–170), the amino-terminal transactivation domain

Keywords. Acetyltransferase activity; CREB-binding protein (CREBBP/CBP); molecular lesions; Rubinstein–Taybi syndrome (RSTS); single-nucleotide polymorphisms (SNPs) in the Indian population

Nucleotide sequence data reported are available in the GenBank databases under the accession numbers(s) EU790714 to EU790785

Abbreviations used: BROMO/BRD, bromo domain; CH, Cys/His-rich region; CI, conservation index; CRE, cyclic AMP response element; CREBBP/CBP, CREB-binding protein; CREP, CRE-binding protein; CTAD, C-terminal transactivation domain; EBV, Epstein–Barr virus; EP300, E1A binding protein p300; GC, guanine–cytosine; GERBIL, genotype resolution and block identification using likelihood; HAT, histone acetyltransferase domain; HW, Hardy–Weinberg; I/E, intron/exon; KIX, CREB-binding domain; NHRD, nuclear hormone receptor-binding domain; NHRD, nuclear receptor-binding and receptor-interacting domain; NTAD, amino-terminal transactivation domain; NTAD, N-terminal transactivation domain; PCR, polymerase chain reaction; PVDF, polyvinylidene fluoride; RSTS, Rubinstein–Taybi syndrome; SIFT, Sorting Intolerant From Tolerant; SNP, single-nucleotide polymorphism

Supplementary tables and figures pertaining to this article are available on the *Journal of Biosciences* Website at *http://www.ias.ac.in/jbiosci/June2010/pp187-202/suppl.pdf*

(NTAD, aa 228–461), the Cys/His-rich region 1 (CH1, aa 363–496), the CREB-binding domain (KIX, aa 452–682), the bromo domain (BROMO/BRD, aa 1108–1170), the histone acetyltransferase domain (HAT, aa 1173–1849), Cys/His-rich region 2 (CH2, aa 1232–1487), Cys/His-rich region 3 (CH3, aa 1690–1860) and the C-terminal transactivation domain (CTAD, aa 1960–2162) (Bentivegna *et al.* 2006). CREBBP was first identified as a transcriptional co-activator for the cyclic AMP response element (CRE)-binding protein CREB (Chrivia *et al.* 1993) that mediates the genomic effects of cAMP by binding to CRE elements on DNA. CREBBP functions by transactivating CREB-regulated transcription when bound to CREB via its KIX domain (Murata *et al.* 2001). The nuclear hormone receptor-binding domain (NHRD) is the region for interaction with the nuclear hormone receptors that represent ligand-dependent transcription factors and regulate target genes involved in cell growth and differentiation. NTAD and CTAD are the N- and C-terminal transactivation domains of the protein, while the BROMO domain interacts with acetylated lysine and other proteins. The HAT domain is essential for acetyltransferase activity which catalyses the acetylation of all four core histones and of several non-histone nuclear proteins, while the three CH domains in the protein serve as docking sites for several transcription factors and also mediate protein–protein interactions (Goodman and Smolik 2000). The CREBBP protein thus plays a critical role in complex biological processes during development. Effectively, mutations in this gene and abnormal functioning of the protein can lead to aberrant patterns in lineage development during embryonic development. Some symptoms of the RSTS syndrome are believed to emerge from such defects in development (Roelfsema *et al.* 2005).

The aim of the present study was to identify the rate of point mutations and other molecular lesions in *CREBBP* among true RSTS patients in the Indian population. To the best of our knowledge, no such analysis has been done, nor is any information available on the occurrence of single-nucleotide polymorphisms (SNPs) in the *CREBBP* gene. A salient feature of this study is the highest detection rate of *CREBBP* mutations (76.9%) in RSTS patients to date, of which 10 are predicted to be pathogenic and three were identified to be critical for histone acetyltransferase activity. Linkage disequilibrium studies revealed an association of *CREBBP* polymorphisms with disease susceptibility.

2. Materials and methods

2.1 *Clinical material*

The Institutional Bio-Ethics Committee at NCCS, Pune approved the present study and informed written consent was obtained from all patients. In case of minor patients (age less than 18 years), written consent was obtained from their parents. Patients with RSTS (*N*=13) were identified by clinical geneticists, based on the presentation of the classical symptoms known to be associated with this syndrome (supplementary table 1). The main facial features observed in the patients included heavy and arched eyebrows, long eyelashes, strabismus, nasolacrimal duct obstruction, myopia/cataracts/glaucoma/coloboma, downward slanting palpebral fissures, large nose, convex/beaked profile of the nose, narrow palate, micrognathia, and low-set ears, which are considered to be classical symptoms (Dai *et al.* 1999). Accordingly, we classified 11/13 (84.6%) patients included in our study as classical/typical RSTS patients. Within this patient cohort, the sex ratio (female: male) was 5:8.

All patients also showed mild to severe psychomotor/mental retardation. Microcephaly was present in 12/13 patients. Skeletal abnormalities such as the typical large thumb were evident in 11/13 (84.6%); radially deviated thumbs were also observed at the same frequency. Patients 4 and 6 (RSTS4a and RSTS6a) who did not have broad thumbs had radially deviated thumbs. Thus, thumb abnormality was present in the entire patient cohort.

Peripheral blood samples were collected from the above patients as well as their parents (whenever available) and 100 randomly selected healthy controls. All samples were used for genomic DNA isolation using the Qiagen FlexiGene DNA kit (Qiagen, USA).

2.2 *Polymerase chain reaction amplification, DNA sequencing and mutational analysis*

The 31 exons of the *CREBBP* gene were amplified from genomic DNA by polymerase chain reaction (PCR). Primers were designed to amplify all the coding exons and intron/exon (I/E) boundaries (supplementary table 2) except exon 2 where, due to a high guanine–cytosine (GC)-rich region, the primers were designed to cover only the exon region. The amplified fragments were sequenced using an ABI 3700 DNA sequencer (Applied Biosystems Inc. USA). Sequences thus obtained were subjected to alignment against a consensus sequence of the *CREBBP* (NM_004380) using the ABI Seqscape (version 2.0) software; this identified mismatches in the nucleotide bases. Re-sequencing was done twice in the opposite directions to rule out sequencing errors. Likely polymorphic sites were then recorded as putative nucleotide changes. In addition, to minimize genotyping errors, the measure of non-conformity in the samples to Hardy–Weinberg proportion for each nucleotide change was calculated, and a deviation from the Hardy–Weinberg proportion was considered as being indicative of sequencing errors.

2.3 *Database mining from existing datasets*

Previously identified DNA variants, which include point mutations, chromosomal translocations such as in-del mutations as well as microdeletions of *CREBBP* reported in public databases including the National Center for Biotechnology Information (NCBI; *http://www.ncbi.nlm.nih.gov/sites/entrez?db=OMIM*), SNP Consortium Limited (TSC) (*http://snp.cshl.org*), Institute of Medical Science – Japanese Science and Technology Japanese SNPs (IMS-JST JSNP) database (*http://snp.ims.u-tokyo.ac.jp)*, the Ensembl Genome Browser (*http://www.ensemle.org*), GeneSNPs Public Internet (*http://www.genome.utah.edu/genesnps*), and the 1000 genomes database (*http://browser.1000genomes.org/index.html*) were used as reference databases to classify the nucleotide changes in our study as being either known/novel.

2.4 *Splice site analysis*

Splice site analysis was carried out using the NetGene2 Server, a service producing neural network prediction through splice sites in *H. sapiens*, *C. elegans* and *A. thaliana* DNA (Hebsgaard *et al.* 1996).

2.5 *Conservation index (CI) analysis*

Conservation of mutated amino acids was screened across 11 species including *Caenorhabditis elegans, Melanospora aegyptica, Danio rerio, Xenopus tropicalis, Gallus gallus, Mus musculus, Rattus norvegicus, Canis familiaris, Felis catus, Pan troglodytes and Homo sapiens* using ClustalW (*http://www.ebi.ac.uk/clustalw/*). The conservation index (CI) was calculated using the following formula:

$$CI = \frac{\text{Frequency of conservation} * 100}{\text{Total no. of species}}$$

where frequency of conservation denotes the number of species in which an amino acid is conserved.

Values of CI <0.5 were construed to be low; values ranging between 0.5 and 0.75 were considered as medium, and CI values >0.75 were considered to be high.

2.6 *SIFT (Sorting Intolerant From Tolerant) analysis*

Prediction of mutation effects was carried out using a web program SIFT (*http://blocks.fhcrc.org/sift/SIFT.html*), which utilizes evolutionary information from homologous proteins (Ng *et al.* 2001, 2003).

2.7 *Protein estimation and western blotting*

Western blotting was performed using standard procedures. Briefly, equal amounts of nuclear proteins extracted using a HAT assay kit (#17-289; Upstate Cell Signaling solutions. NY, USA) were resolved on 10% SDS-PAGE gels and the resolved proteins were transferred to a polyvinylidene fluoride (PVDF) membrane. Blots were incubated with 1:1000 dilution of CBP antibody (A22; Santa Cruz Biotechnology Inc., Santa Cruz, CA, USA) and the immunoreactive bands were visualized by using the Pierce Super Signal West Pico Chemiluminescence system (Thermo Fisher Scientific, USA). Blots were subsequently stripped and E1A binding protein p300 (EP300) was detected with a 1:800 dilution of EP300 antibody (C-20; Santa Cruz Biotechnology Inc., Santa Cruz, CA, USA) as described above. Finally, the blots were stripped and developed using a 1:800 dilution of histone (H1) antibody (FL-219; Santa Cruz Biotechnology Inc., Santa Cruz, CA, USA), which was used as the loading control.

2.8 *Acetyltransferase assays*

To test endogenous CREBBP acetyltransferase activity from Epstein–Barr virus (EBV)-transformed lymphoblastoid cells *in vitro*, a HAT assay kit (#17-289, Upstate Biotechnology) was used according to the manufacturer's instructions. Briefly, 100 ng of biotinylated histone H4 peptide was added per well in streptavidin-coated 96-well microplates. After blocking with 3% bovine serum albumin, tested compounds diluted in DMSO were added to the wells at the indicated concentrations. Acetyl-coenzyme A (100 μmol/l) and nuclear extract (50 ng) were then added to a total assay volume of 50 μl. The plates were incubated at 30°C for ~8 min, then washed five times with TBS, and incubated with anti-acetyl-lysine antibodies for 1.5 h at room temperature. After incubation with horseradish peroxidase-conjugated secondary antibodies, 100 μl of the substrate mixture was added to each well and incubated for 10 min at room temperature. The reaction was terminated with 50 μl of fresh 1 mol/l sulphuric acid and the plates were read at 450 nm.

2.9 *Phylogenetic analysis*

CREBBP variant sequences obtained from RSTS patients were applied for constructing phylogenetic trees as described earlier (Wani *et al.* 2006).

2.10 *Haplotype determination*

Gevalt 2.0 (GEnotype Visualization and ALgorithmic Tool) (Davidovich *et al.* 2007) was used to determine haplotypes as

well as haplotype blocks. Gevalt constructs haplotypes using an accelerated expectation maximization algorithm similar to the partition/ligation method (Qin *et al.* 2002), which creates highly accurate population frequency estimates of the phased haplotypes based on the maximum likelihood derived from the unphased input genotypes (genotype resolution and block identification using likelihood [GERBIL]). Gevalt employs the Li–Stephens model (Li and Stephens 2003), which simulates case and control individuals conditional upon a set of known haplotype data using an estimate of the fine-scale recombination rate across a region. This analysis is preferable to a direct resampling approach (de Bakker *et al.* 2005) which produces new haplotypes that are copies of the original HapMap haplotypes. Even if there is a modest amount of overfitting in the reported power levels, because of the same simulated datasets the qualitative differences between the performances of the two methods remain the same.

The test for association is usually based on the difference in allele frequency between case and control individuals. For a single SNP, a common test suggested by Olson and Wijsman (1994) is based on building a contingency table of alleles compared with disease phenotypes (i.e. case/control) and then calculating a -χ^2 distributed statistic. When multiple markers in a chromosomal region are to be tested, several studies suggest the use of generalized linear models (Schaid *et al.* 2002; Zaykin *et al.* 2002; Lin 2006). Such methods must assume a specific distribution of the trait, given the SNPs, and this assumption does not always hold true. Typically, a Bonferroni correction for the *P* value is employed to account for multiple testing. However, this correction does not take into account the dependence of strongly linked marker loci and may lead to overconservative conclusions. This problem worsens when the number of sites increases. Thus, to address these difficulties, Zhang *et al.* (2002) suggested a Monte Carlo procedure (Kalos 1986) to evaluate the overall *P* value of the association between the SNP data and the disease: the χ^2 value of each marker is calculated, and the maximum value over all markers, denoted by CC_{max}, is used as the test statistic. The same statistic is calculated for many datasets with the same genotypes and with randomly permuted labels of case and control individuals. The fraction of permutations for which this value exceeds the original CC_{max} is used as the *P* value. A clear advantage of this test is that no specific distribution function is assumed. Additionally, the test handles multiple testing directly and avoids correction bias. Gevalt uses the same principle and hence *P* values were not additionally corrected using Bonferroni correction.

Accordingly, linkage disequilibrium (the D' statistic) was calculated between each pair-wise combination of all the mutations and SNPs within the *CREBBP* gene. Gevalt is able to determine haplotype blocks using different block partition algorithms (Barrett *et al.* 2005). In our dataset, the solid spine method of the linkage disequilibrium algorithm was used to determine the haplotype blocks within the *CREBBP* gene. Haplotypes within each block were assigned to individual subjects only when the assignment could be made with a greater than 95% certainty. Markers with D' values = 1 were selected as these were in complete linkage disequilibrium and were not separated as a result of recombination events. This resulted in the identification of eight groups showing markers of independent association or interdependency (supplementary table 3). These groups of markers were further analysed to identify an increased level of stringency of correlation with disease susceptibility, i.e. those markers were selected whose r^2 values were equal to one. This finally led to the identification of the two groups as discussed in detail in the Results section.

2.11 *Statistical analysis*

RSTS patient and control data were analysed using a χ^2 test. The gene counting method estimated allele frequencies for all the genotyped polymorphisms and the χ^2 test was used to identify deviation from Hardy–Weinberg (HW) proportions.

3. Results

3.1 *Nucleotide variations in* CREBBP *in RSTS patients*

Screening for nucleotide changes by direct sequencing of 31 PCR genomic fragments encompassing the complete coding sequence and splice sites of the *CREBBP* gene in the 13 Indian RSTS patients led to the identification of 45 sequence alterations (figure 1; supplementary tables 4 and 5). Of these, 25 nucleotide changes were located in the exon regions, while the remaining 20 were either in intron or I/E boundary regions (supplementary figure 1; supplementary tables 4, 5). All the exon changes were novel and have not been reported previously; 18 of these were missense while seven did not result in an altered amino acid (i.e. were sense/silent) (figure 2A). Further, seven of the 18 missense mutations were construed to be *de novo* mutations since they were not identified either in the parents or the controls. The remaining 11 mutations, although novel, could not be established as being either *de novo* or inherited, as parent samples of the patients carrying these mutations were unavailable.

Of the seven silent mutations, changes in amino acid 424, i.e. p.Arg424Arg identified in our study, may be critical in RSTS as a nonsense mutation in the same amino acid generating a stop codon (p.Arg424X) has been reported earlier and considered to be pathogenic due to the formation of a truncated CREBBP protein (Bartsch *et al.* 2002; Bartsch *et al.* 2005). The remaining six silent mutations are novel; three of these were identified as being *de novo.* Among the 20 intron or I/E boundary nucleotide changes, g.18617C>G (c.3982+93C>G; *rs129974)* and g.51363T>G (c.2158+81T>G; *rs130002)* are

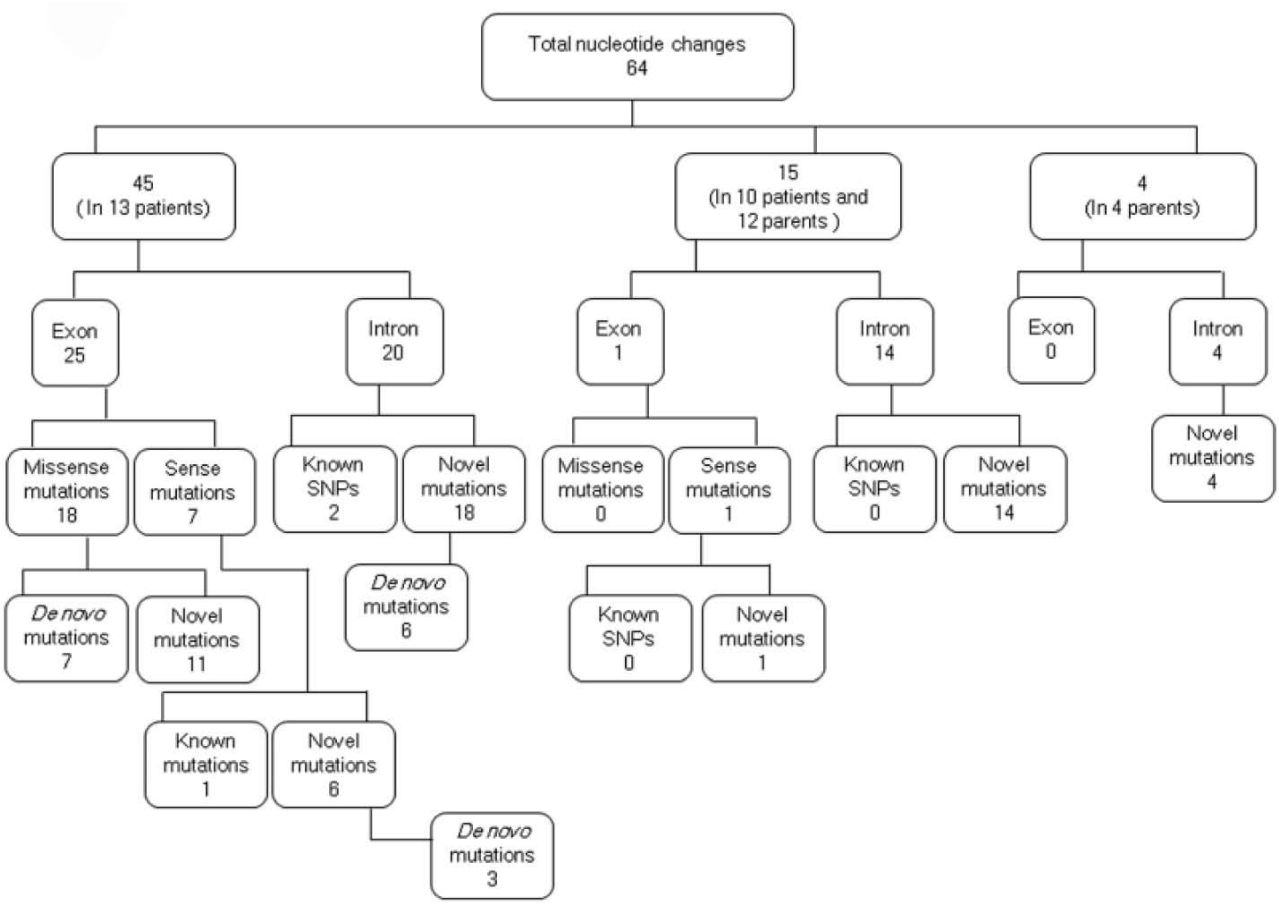


Figure 1. Representation of the distribution of nucleotide changes in the *CREBBP* gene of RSTS patients and their parents. Numbers in each block represent the number of mutations in the same group. As is the convention, exon mutations/SNPs causing amino acid changes were classified as missense/non-synonymous, while those that translated into the same amino acid were termed silent/synonymous. Those variants which were present exclusively in the patients and not in the parents were considered to have arisen *de novo;* those which were not found either in the parents, controls or in the dbSNP database were referred to as novel. Nucleotide changes found in controls alone or listed in the dbSNP were classified as SNPs.

known SNPs, six arose *de novo*, while the remaining were novel changes that were not detected in the control samples and may contribute to the syndrome.

Fifteen nucleotide changes common to patients as well as parents were identified (supplementary table 6). One of these was present in the exon region and is novel but silent (p.His1235His); the remaining 14 were located at I/E boundaries and were neither present in the controls nor have they been reported earlier. Besides the patient-exclusive and patient–parent common nucleotide changes, four novel variants were also identified in the intron regions only in the parents (supplementary table 7).

3.2 *Domain-wise distribution of mutations*

The overall pattern of distribution of nucleotide changes in the *CREBBP* gene identified in our Indian RSTS patients initially appeared to be similar to that reported in the existing database (Bartsch *et al.* 2002, 2005; Coupry *et al.* 2002; Kalkhoven et *al.* 2003; Roelfsema *et al.* 2005; Udaka *et al.* 2005; Bentivegna *et al.* 2006) (supplementary table 10). Further analysis revealed a distinctive signature of distribution in our study. A majority of the exon and I/E variants identified in previous studies (figure 2B) are located in the domains of HAT (44.7%), CH2 (15%), CH1 (13.1%), and NTAD (8.1 %), while variants in the other domains, viz. NHRD, CREB, BROMO, CH3 and CTAD (together 16.4%) were very few. Our study in Indian patients identified a maximum number of exon- and I/E-localized nucleotide changes (figure 2C) in the HAT (42.2%) followed by CREB (15.5%), CH1 (13.3%), CH2 (13.3%) and CH3 (6.6%) domains (figure 2C), while the frequency of variants in the NHRD, NTAD, BROMO and CTAD regions was very low (together 8.8%). Distinct differences were seen in the frequency of variants lying in the CREB domain (15.5% vs

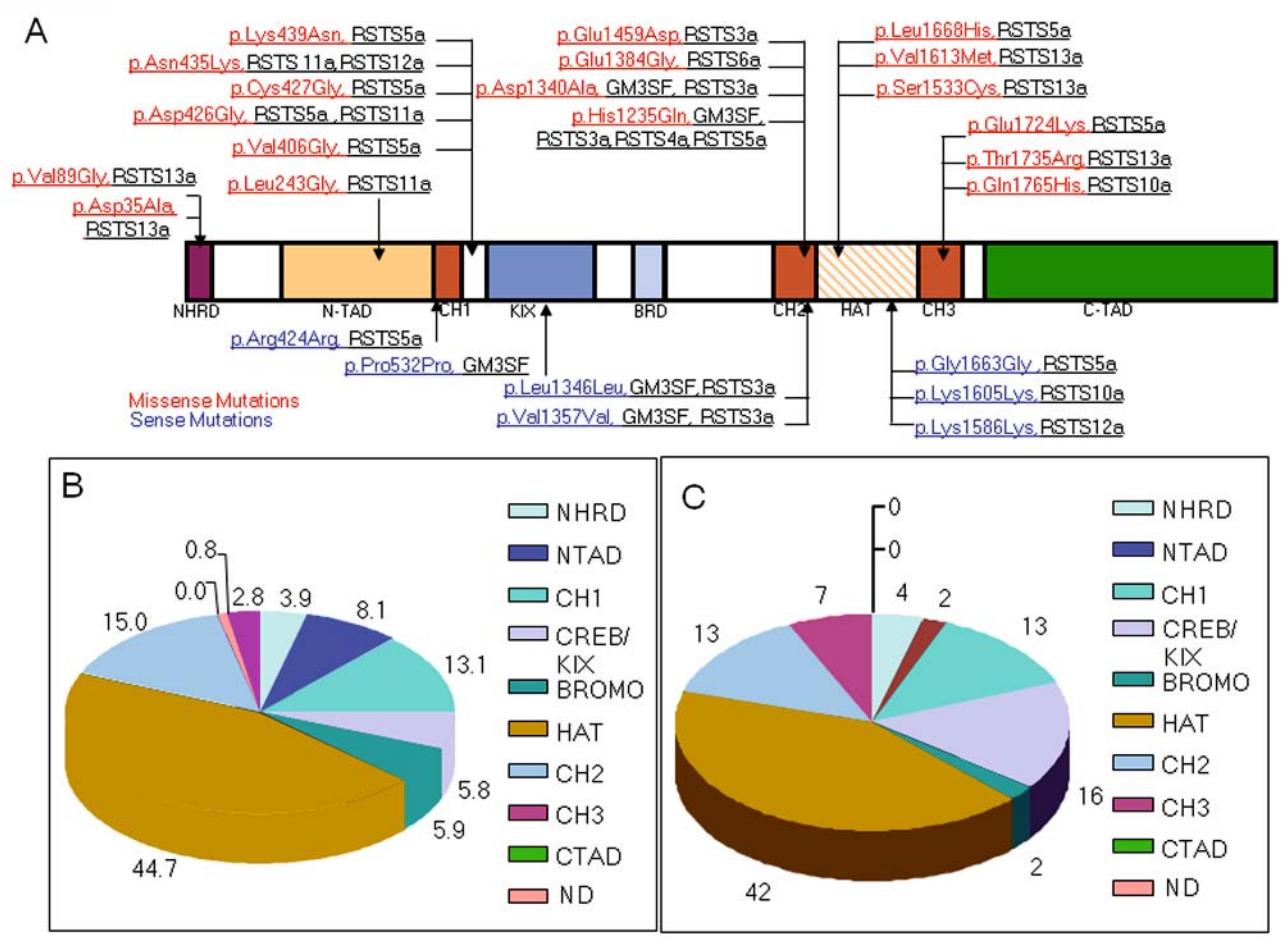


Figure 2. **(A)** Location of the exon *CREBBP* mutations found in this study. Only exons are drawn to scale. Missense mutations are in red colour, sense (silent) mutations in blue colour. The domains indicated in the figure include the nuclear receptor-binding and receptor-interacting domain (NHRD, aa 1–170), the amino-terminal transactivation domain (NTAD, aa 228–461), the Cys/His-rich region 1 (CH1, aa 363–496), the CREB-binding domain (CREB /KIX, aa 452–682), the bromo domain (BROMO/BRD, aa 1108–1170), the histone acetyltransferase domain (HAT, aa 1173–1849), Cys/His-rich region 2 (CH2, aa 1232–1487), Cys/His-rich region 3 (CH3, aa 1690–1860) and the C-terminal transactivation domain (CTAD, aa 1960–2162). **(B)** Domain-wise distribution of nucleotide changes in the existing database. **(C)** Domain-wise distribution of nucleotide changes in this study. The domains indicated in the figure include NHRD, the nuclear receptor-binding and receptor interacting domain; NTAD, the amino-terminal transactivation domain; CH1, first zinc finger; CREB/KIX, the CREB-binding domain; BROMO, the bromo domain; HAT, the histone acetyltransferase domain;CH2 and CH3, the second and third zinc fingers; CTAD, the C-terminal transactivation domain; ND, regions of the gene not included in any domain. The numbers indicate the percentage of mutations identified in the respective domains. A resampling-based approach was used to determine the frequencies of distribution so that the frequency of mutations in each domain in our study and the database are from the same number of individuals.

5.8% in the database and Indian RSTS patients, respectively). The CH-rich regions also revealed nucleotide changes between the database and Indian RSTS patients (28% vs. 33.2%, respectively). The highest degree of nucleotide change was evident in the CH3 region between the database and Indian RSTS patients ((0 vs 13.3%, respectively; figure 2 B, 2C).

A similar pattern was also evident in the exclusive distribution of exon mutations in RSTS patients (figure 3A). The existing database (Bartsch *et al.* 2002, 2005; Coupry *et al.* 2002; Kalkhoven *et al.* 2003; Roelfsema *et al.* 2005; Udaka *et al.* 2005; Bentivegna *et al.* 2006) indicated a distribution of exon mutations maximally in the HAT (38.5%) followed by the NTAD (12.8%) and CH2 (11.4%) domains, whereas in the Indian RSTS patients, we noted that, while a large number of mutations were indeed located in the HAT region (24%), a majority of the remaining mutations clustered around the three zinc finger regions, with their frequencies being CH1 (24%), CH2 (24%) and CH3 (12%). A significant difference in the number of mutations in the NTAD domain

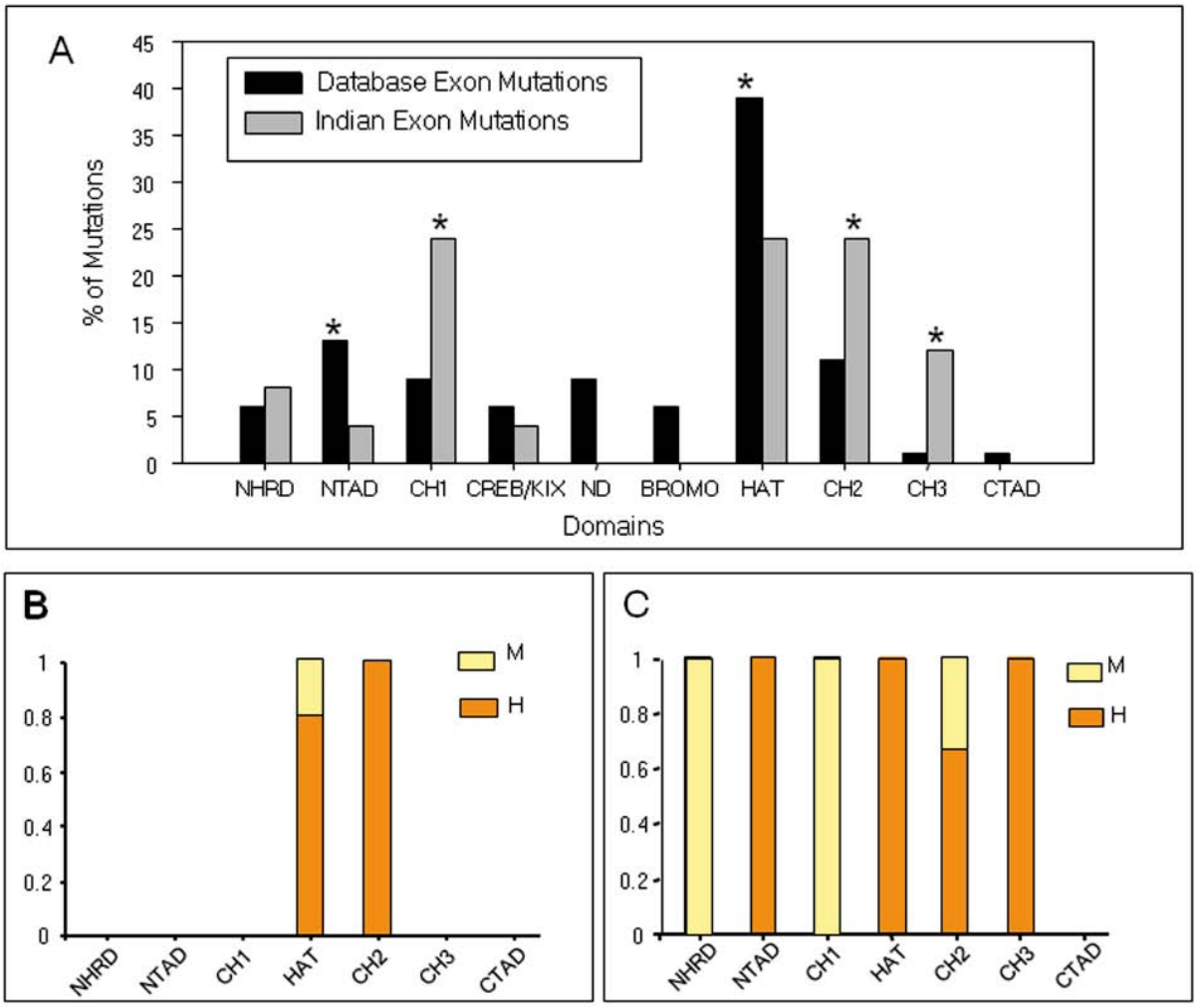


Figure 3. **(A)** Comparative distribution of exon mutations between the existing database and our study. The exon mutations in the database are primarily localized in the HAT domain, whereas in our study they are concentrated in the CH1, CH2, CH3 and HAT domains. *****Denotes the domains having significant difference (>5%) between database exon mutations and those identified in this study. **(B)** Distribution of missense mutations only in the existing database according to their conservation indices. **(C)** Distribution of missense mutations in the Indian population according to their conservation indices (*see also* footnotes # and $ of table 2). CREBBP domains are plotted on the X axis, CI on the Y axis. M stands for medium conservation whereas H stands for high conservation index.

was also evident, with those in the database being reported at a higher frequency (12.8% as compared to 2% in the Indian patients). Thus, the distribution of nucleotide changes in the I/E regions as well as exon mutations in the *CREBBP* gene in Indian RSTS patients exhibited a specific pattern, with the three zinc fingers and HAT domain being putative hot spots.

3.3 *Assessment of the pathogenicity of missense mutations in our study*

Missense mutations are considered pathogenic when at least one of the following features is associated with the mutation (Downing *et al.* 1996; Faivre *et al.* 2009): (i) *de novo* missense mutation, (ii) missense mutation substituting or creating a cysteine, (iii) missense mutation involving a consensus calcium-binding residue, (iv) substitution of glycines that are implicated in correct domain–domain packing, (v) intrafamilial segregation of a missense mutation involving a conserved amino acid. These parameters were applied towards assessing the pathogenicity of the missense mutations in the present study. Thus, the following analyses were performed.

3.3.1 *Conservation index (CI) analysis:* All the missense mutations identified in our study as well as those documented earlier in the database were analysed on the basis of their distribution in the various functional domains of the protein,

as described earlier, and their conservation indices (CIs) across 11 species (table 1). Applying a classification of low (CI <0.5), medium (CI 0.5–0.75) or high CI (CI >0.75) to the *CREBBP* missense mutations in the existing database, some of them were found to be associated with a high CI (figure 3B), and have been reported to be pathogenic (Murata *et al.* 2001; Coupry *et al.* 2002; Kalkhoven *et al.* 2003; Bartsch *et al.* 2005; Roelfsema *et al.* 2005; Udaka *et al.* 2005; Bentivegna *et al.* 2006); exceptions, however include p.Asn1978Ser, p.Ala981Thr, p.Met2221Leu (table 2). Applying the same analysis to our study, we observed that missense mutations in the NHRD and CH1 domains showed a medium CI, whereas those localized in the CH2, CH3 and HAT domains were highly conserved (figure 3C). In conclusion, the 10 missense mutations identified in our study with a CI >0.75 (table 1) could be potentially pathogenic.

3.3.2 *SIFT (Sorting Intolerant From Tolerant) analysis:* The SIFT algorithm predicts the effects of amino acid substitution on protein function, based on the position and type of amino acid change (Ng *et al.* 2001). On subjecting the missense mutations identified in our study to SIFT analysis, we predicted that 13 out of the 18 missense mutations were potentially deleterious and may alter protein function (table 1).

3.3.3 *Structural changes in the protein structure predicted from domain-specific mutations:* The structure of proteins permits a more detailed understanding of mutations observed in a number of human disease states. We thus

Table 1. Distribution of missense mutations in Indian RSTS patients based on their conservation index

No.	Mutation	CREBBP domain	Conservation across species (number)#	Conservation index (CI)	SIFT analysis prediction	SIFT analysis prediction level (high/ low)	Altered HAT activity	Pathogenic
1	p.Asp35Ala	NHRD	7	0.64	Affect protein function	low	ND	No
2	p.Val89Gly	NHRD	7	0.64	Tolerated	high	ND	No
3	p.Leu243Gly	NTAD	9	0.81	Affect protein function	low	ND	No
4	p.Val406Gly	CH1	6	0.54	Affect protein function	high	ND	Yes
5	p.Asp426Gly	CH1	6	0.54	Tolerated	high	ND	No
6	p.Cys427Gly	CH1	6	0.54	Affect protein function	high	ND	Yes
7	p.Asn435Lys	CH1	6	0.54	Affect protein function	high	ND	Yes
8	p.Lys439Asn	CH1	6	0.54	Affect protein function	low	ND	No
9	p.His1235Gln	CH2	8	0.72	Affect protein function	low	No	No
10	p.Asp1340Ala	CH2	10	0.9	Affect protein function	low	No	No
11	p.Glu1384Gly	CH2	10	0.9	Tolerated	high	Yes*	Yes
12	p.Glu1459Asp	CH2	10	0.9	Tolerated	high	Yes	Yes
13	p.Ser1533Cys	HAT	10	0.9	Affect protein function	high	ND	Yes
14	p.Val1613Met	HAT	9	0.81	Affect protein function	low	ND	Yes
15	p.Leu1668His	HAT	11	1	Affect protein function	low	Yes	Yes
16	p.Glu1724Lys	CH3	11	1	Affect protein function	low	Yes	Yes
17	p.Thr1735Arg	CH3	10	0.9	Tolerated	high	ND	Yes
18	p.Gln1765His	CH3	10	0.9	Affect protein function	low	ND	No

*not statistically significant

Table 2. Distribution of missense mutations reported in the existing RSTS database based on their conservation index

No.	Mutation	Domain	Conservation across species (number)#	Conservation index (CI)$	References
1	p.Ala981Thr	No Domain	6	0.54	Coupry *et al*. 2002
2	p.Tyr1175Cys	HAT	10	0.9	Bartsch *et al.* 2002
3	p.Glu1278Lys	HAT	10	0.9	Kalkhoven *et al*. 2003; Roelfsema *et al*. 2005
4	p.Glu1278Gly	HAT	10	0.9	Udaka *et al.* 2005
5	p.Arg1378Pro	CH2	7	0.6	Murata *et al.* 2001
6	p.His1413Pro	CH2	11	1	Udaka *et al.* 2005
7	p.Asp1435Val	CH2	11	1	Bartsch *et al.* 2005
8	p.Thr1447Ile	CH2	11	1	Roelfsema *et al.* 2005
9	p.Tyr1450His	CH2	11	1	Roelfsema *et al.* 2005
10	p.His1470Arg	CH2	11	1	Roelfsema *et al.* 2005
11	p.Tyr1482Cys	CH2	11	1	Bentivegna *et al*. 2006
12	p.Asp1651His	HAT	11	1	Murata *et al.* 2001
13	p.Arg1664His	HAT	11	1	Roelfsema *et al.* 2005
14	p.Asn1978Ser	HAT	8	0.72	Bartsch *et al.* 2005

#Conservation was screened across 11 species including *C. elegans, M. aegyptica, D. rerio, X. tropicalis, G. gallus, M. musculus, R. norvegicus, C. familiaris, F. catus, P. troglodytes.*

$Conservation index was calculated by the following formula=Frequency of conservation*100/total number of species

searched the literature and databases to access established structures of the normal human CREBBP protein. A recent publication (Liu *et al.* 2008) describes the EP300 HAT domain using high-resolution X-ray crystal structure. CREBBP has a high homology with EP300 (88% in the HAT region – supplementary figure 2). Thereby, using the EP300 HAT domain as a reference structure, we could analyse the effects of five missense mutations identified in our study located in the HAT region – p.Asp1340Ala, p.Leu1668His, p.Glu1384Gly, p.Ser1533Cys and p.Glu1459Asp. The analysis revealed that the first two of these mutations were present in the helix region, p.Glu1384Gly and p.Ser1533Cys in the β-sheet region, while p.Glu1459Asp was located in the L1 loop. The latter region has been identified as the substrate-binding loop of the CREBBP protein essential for HAT activity (Liu *et al.* 2008). By this derivation, only the p.Glu1459Asp associated mutation can be considered to have the maximum probability of being pathogenic (figure 4A).

Similarly, we accessed a structure in the Protein Data Bank (1U2N) (*http://www.rcsb.org/pdb/home*) describing the CREBBP amino acid 340–442 regions that encompass a part of the NTAD (CH1) domain. Using this as a reference structure, we could analyse the effects of five missense mutations identified in our study which were located in this region and affected five amino acids – p.Val406Gly, p.Asp426Gly, p.Cys427Gly, p.Lys439Asn and p.Asn435Lys. The former four were localized to a part of the loop region while p.Asn435Lys was localized to the helix region (figure 4B). The exact implications of these mutations cannot be assessed since the significance of these regions with respect to the functionality of the protein have not been identified as yet.

Analysing the 18 missense mutations on the basis of the three parameters outlined above, we thus predict that 10 of these could be potentially pathogenic (table 1). It may be noted that certain mutations which were predicted to be tolerated through SIFT analysis are considered by us as being pathogenic, since they have been earlier reported to be critical for HAT activity (Bannister and Kouzarides 1996; Ogryzko *et al.* 1996; Giles *et al.* 1997; Kalkhoven *et al.* 2003).

3.4 *Reduced CREBBP HAT activity in RSTS patient-derived cells*

An assessment of functionality serves as an essential validation of the predicted pathogenicity of missense mutations. We thus determined whether the missense mutations identified in our study (figure 2A) disrupt the HAT activity of CREBBP in cells derived from RSTS patients. To this end, we used EBV-transformed lymphoblastoid cell lines established in our laboratory from patients GM3SF, RSTS3a, RSTS5a and RSTS6a. Equal amounts of nuclear extracts from patient cell lines were used in the acetylation assay and compared with those from cell lines derived from four healthy control individuals (figure 4C[i], 4C[ii]). To further ascertain that the altered HAT activity

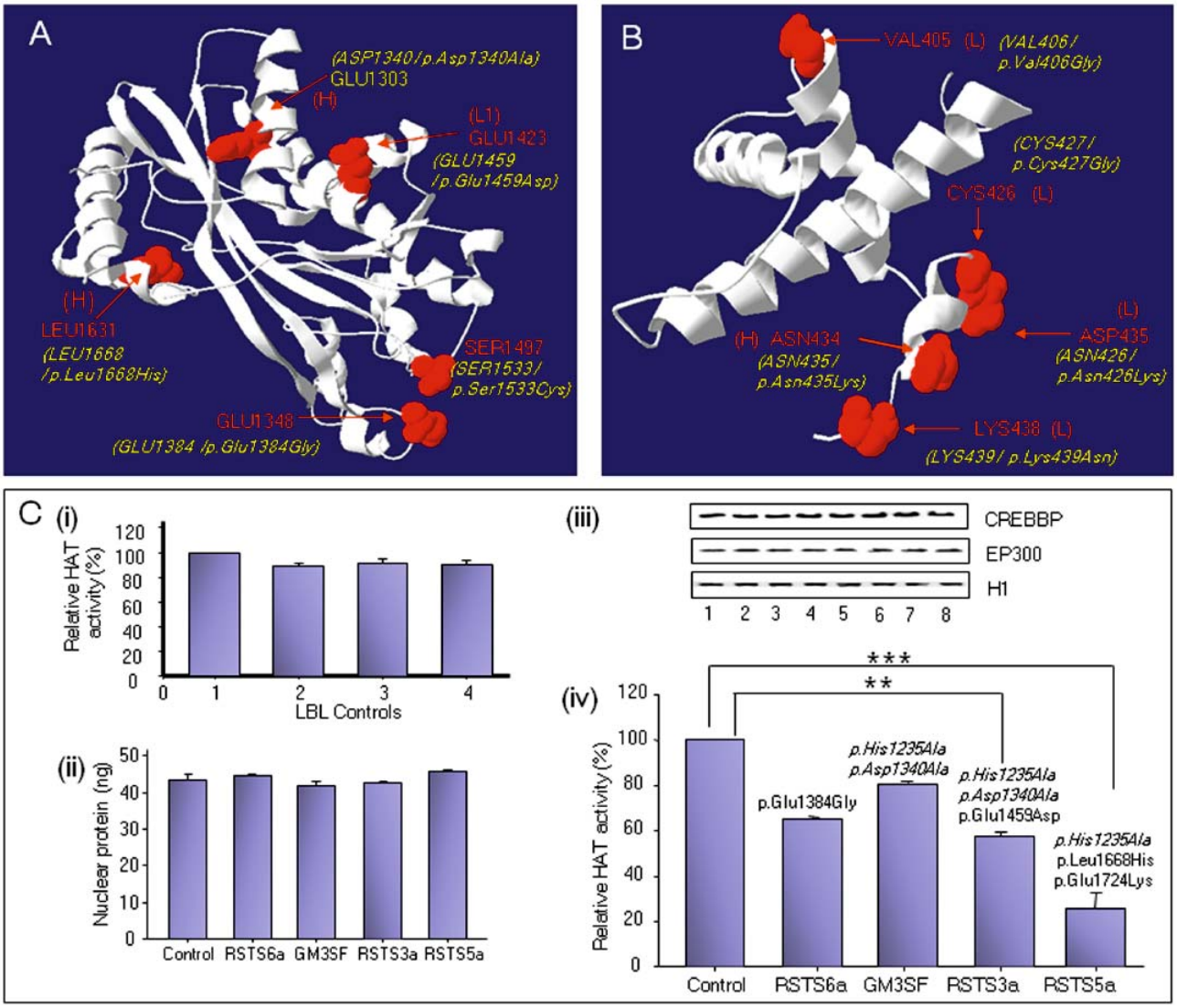


Figure 4. Depiction of amino acid residues in the HAT region **(A)** and CH1 region **(B)** of *CREBBP* mutated in our study. **(A)** Structure of the EP300 HAT region (adopted from Liu *et al*. 2008). **(B)** Structure of CH1 region of murine *CREBBP*. The letters 'L' and 'H' represent the loop and helix region of the protein. Actual amino acid residues in human CREBBP are represented in parentheses/missense mutations at the same amino acids identified in the study. **(C) (i)** Comparable HAT activities of lymphoblastoid cell lysates from four healthy individuals used as control. **(ii)** Graphical representation of the protein estimation of the HAT reaction cocktail used for each assay. It may be noted that comparable protein levels were used for each assay. **(iii)** Western blotting showing expression levels of CREBBP and EP300 in controls and RSTS patients. Numbers 1–4 are the controls used, numbers 5, 6, 7, 8 are RSTS patients RSTS6a, GM3SF, RSTS3a and RSTS5a, respectively. **(iv)** Reduced CREBBP acetyltransferase activity in cells derived from RSTS patients. Epstein–Barr virus (EBV)-transformed lymphoblastoid cell lysate from healthy control and RSTS patients were used to acetylate histones *in vitro*. The data shown are representative of HAT assays done in triplicate and depicted as mean ± SEM. **P*<0.05, ***P*<0.01, ****P*<0.001. The level of acetylation for the control lane was set to 100%. Patient RSTS5a showed ~87% less activity whereas GM3SF, RSTS3a, RSTS5a and RSTS6a showed ~17%, 27% and 20% less activity as compared to the control individual. The mutations in *italics* are the common mutations found in the patients and those in **bold** are the mutations specific to the patient.

would be reflected by changes in the CREBBP protein, we checked the expression levels of a similar HAT molecule, EP300. Thus, western blotting using nuclear extracts from the lymphoblastoid cell lines from these patients as well as control individuals revealed that both CREBBP as well as EP300 protein expression levels in all the controls as well as the patients – GM3SF, RSTS3a, RSTS5a and RSTS6a – were similar (figure 4C[iii]). Thereby, alterations in

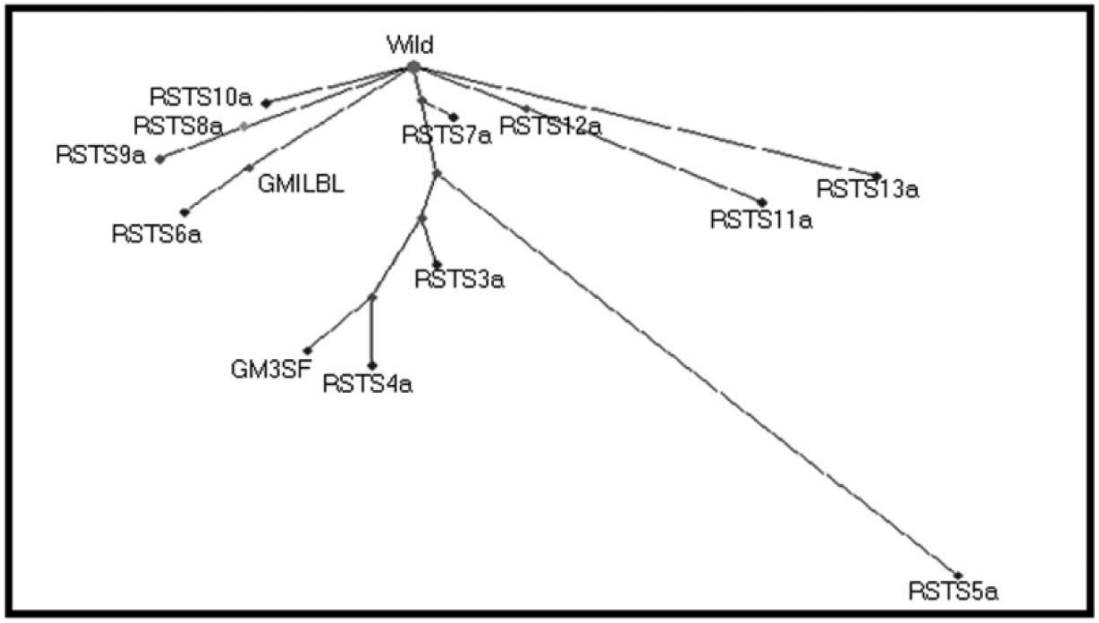


Figure 5. Phylogenetic tree showing distribution coexistence of the Indian RSTS patients in the study.

HAT activity (if any) were likely to result from the specific *CREBBP* mutations observed in the corresponding patients.

As seen in figure 4C(iv), the overall acetyltransferase activity of RSTS-derived cell lines was identified to be lower than those derived from healthy individuals; although a significant reduction was associated with only two samples – RSTS3a and RSTS5a. RSTS3a is associated with the missense mutations p.His1235Glu, p.Asp1340Ala and p.Glu1459Asp, and any/all of these could be critical modulators of HAT activity. However, p.His1235Glu and p.Asp1340Ala were also present in patient GM3SF, wherein a significant decrease in HAT activity was not identified. From these observations, we deduce that the reduced HAT activity in RSTS3a may possibly be attributed to the missense mutation p.Glu1459Asp. Along similar lines, the reduction in HAT activity in patient RSTS5a may involve either/or two missense mutations p.Leu1668His and p.Glu1724Lys but not p.His1235Glu, which was present in GM3SF cells. The functional implications of missense mutations identified in other patients could not be established, since lymphoblastoid cell lines were not established from these samples.

In conclusion, at least 3 of the missense mutations identified – p.Glu1459Asp, p.Leu1668His and p.Glu1724Lys – show a significant influence on the HAT function of CREBBP.

3.5 *Phylogenetic similarity of the missense mutations*

Phylogenetic analysis reveals that individuals with different nucleotide variation(s) form separate branches on the phylogenetic tree. The further a branch is from the wild type, the more the number of sequence variations that can be expected. Hence, branch length is considered as an additional parameter of divergence (figure 5). We thus analysed the RSTS patients in our study to assess their phylogenetic proximity. Patients RSTS10a and RSTS13a showed exclusive mutations and hence formed separate branches. The pairs RSTS8a and RSTS9a, GM1LBL and RSTS6a, RSTS11a and RSTS12a occupy the same branch since the members within each pair share some common nucleotide changes. GM3SF, RSTS3a, RSTS4a, RST5a and RSTS7a lie on one branch as these samples showed some common nucleotide variations; divergence emerging due to individual mutations is indicated through the sub-branches that arise subsequently. Based on the length of its sub-branch, RSTS5a especially may be deduced to be drastically deviant from the wild type, a fact that is in conformation with the identification of a maximum number of nucleotide changes in the *CREBBP* gene in this patient (supplementary tables 4 and 5).

3.6 *Analysis of SNPs in* CREBBP *in the Indian population*

Amplification and sequencing of the exon, intron and I/E boundary regions of the *CREBBP* gene in 100 control individuals to ascertain polymorphisms in the Indian population led to the identification of a total of 33 SNPs (supplementary table 8). Of these, seven SNPs were located in the exon regions, two in the intron and 24 in the I/E

boundaries. The distribution of these SNPs in the population was as follows: 24 SNPs were identified exclusively in the controls; five in controls, RSTS patients and parents; one in controls and RSTS patients; and three in controls and parents. Among the former, the p.Ile1300Ile polymorphism has been reported earlier (*rs129974*-dbSNP), while the rest are novel. Four of the intron SNPs – c.3836+77T>C (*rs129982*), c.3836-173T>C (*rs129981*), c.2158+81T>C (*rs130002*), c.4394+42G>A (*rs129667*) – have been reported. Although at present the 28 novel SNPs may be considered to be specific to the Indian population, their presence is not reported in the 1000 genome database that includes a large number of Gujarati individuals among many individuals from other populations. This warrants further investigations.

3.7 *Allelic association of the nucleotide changes between cases and controls and linkage disequilibrium studies*

The exact molecular–genetic aetiology of RSTS is still not clearly understood; most groups report the *de novo* occurrence of these mutations. A rare case study has been reported in a family wherein the mother expressed partial symptoms, while the child expressed most RSTS symptoms along with more pronounced clinical features (Petrij *et al.* 2000a). Such findings are suggestive of co-segregation of genetic markers. Further, one cannot rule out the possibility of somatic mosaicism, which has been documented in conditions such as tuberous sclerosis (Verhoef *et al*. 1999), wherein the parents may be unaffected but the offspring may be affected. Earlier reports have also speculated on its occurrence in RSTS but the frequency of somatic and germ-line mosaicism in the RSTS families described were found to be low (Hennekam *et al.* 1990; Stevens *et al.* 1990; Chiang *et al.* 2009). With this rationale, we applied linkage disequilibrium analyses in our study, which would give an insight of co-segregation or association between a genetic marker (variant) identified in the *CREBBP* gene in our study and RSTS.

Initial haplotype analysis (haplotype frequency estimation and pair-wise linkage disequilibrium between the SNPs and mutations) using Gevalt was followed with single locus and multi-marker haplotype association tests and chi-square and *P* value (uncorrected) estimations for the allele frequency computation in RSTS cases vs controls (supplementary table 9). Ten thousand permutations for association results of individual markers within our study were carried out towards obtaining a measure of significance corrected for multiple testing biases. A P value of <0.05 was considered statistically significant. The entire gene was defined as a single haplotype block of 120 kb within which markers with D' values ≤ 1 were identified. Of these, only markers with D' values = 1 were selected as such a state is indicative of significantly high linkage disequilibrium in which the likelihood of marker separation is low during recombination events.

Eight groups of markers were thus identified (supplementary table 3). These groups were further screened to identify two subgroups in which an association existed between the identified mutation and SNPs. The first of these two groups had nine markers including one known SNP (*rs129974*), and eight novel variants specific to the Indian population including one missense mutation (table 3). Marker M32 was dropped during further analysis as a χ^2 analysis revealed its association with the other markers in the group to be statistically insignificant (P=0.1874). Markers M19, M28, M31, M60, M61 are located in the intron regions or I/E boundaries and hence may either be silent or have no obvious effect on biological function. Interestingly, the remaining markers M26, M33, and M41 suggested to be

Table 3. First group of markers identified in significant association with RSTS in Indian patients using linkage disequilibrium analysis

No.	Markers	Alleles	Exact position	*dbSNP/Novel*	Samples
1	M19	A/G	IVS27+238A>G	Novel	GM1LBL
2	M26	T/G	p.Asp1340Ala	Novel	GM3SF, RSTS3a
3	M28	T/G	IVS23-277T>G	Novel	GM1LBL
4	M31	C/G	IVS23+103C>G	Novel	GM1LBL
5	M32	G/T	IVS23+93G>T	*rs129974* *	C21, C17, GM1LBL
6	M33	C/T	p.Ile1300Ile	*rs3025684*	C1, C23, GM1LBL, GM3SF, RSTS3a, RSTS3c, RSTS4a, RSTS4b, RSTS6a, RSTS6b
7	M41	G/C	IVS21-206G>C	Novel	C8, GM3SF, RSTS3a, RSTS3b, RSTS4a, RSTS4b, RSTS4c, RSTS5a, RSTS5c, RSTS6c, RSTS7a
8	M60	C/T	IVS8-235C>T	Novel	RSTS3a
9	M61	A/G	IVS8-237A>G	Novel	RSTS3a

*Association not significant

in complete linkage disequilibrium (D'=1), were identified in association within the family of the patient RSTS3a. In this case, all the three markers were present in the patient whereas the father possessed M33 and M41 (SNPs *rs129974 and c.3836*-206G>C), but lacked M26 (missense mutation p.Asp1340Ala). This suggests the *de novo* emergence of the mutation, though associated with the two SNP markers that were inherited from the father. This is further strengthened by the r^2 values of these markers, which were found to be <1, indicating a possibility of recombination at these regions. Similarly, all the three markers also coexisted in GM3SF; unfortunately, due to a lack of samples from the parents, a similar pattern of marker inheritance could not be mapped. Based on the significant *P* values of the three markers M26, M33 and M41 in addition to the analyses, we conclude that they may be highly associated with RSTS; however, the possibility of some recombination events cannot be totally ruled out.

The second group (table 4) included 11 markers of which M48 is a known SNP (*rs130002*). Six markers within this group – M30, M46, M49, M67, M68 and M69 – were identified in the intron or I/E boundary regions and may be silent. Further analysis revealed that markers M48 (*rs130002*) and M55 (missense mutation, p.Asn435Lys) were in complete linkage disequilibrium, i.e. they expressed D'=1, yet the possibility of some recombination around these regions was indicated, as values for r^2 were <1. These two markers were identified in patients RSTS11a and RSTS12a; unfortunately, due to a lack of samples from their parents, marker inheritance could not be mapped. The remaining marker pairs M64 and M67, M65 and M67, and M66 and M67, in addition to having D'=1, also depicted $r^2 = 1$, suggesting that they were inseparable during recombination events and inheritance. All these three marker pairs were present in RSTS13a; however, inheritance could not be identified due to unavailability of parent samples.

4. Discussion

The present study is a first report on the molecular investigations of the *CREBBP* gene in RSTS patients in the Indian population, together with SNP profiling of 100 control individuals. All thirteen RSTS patients exhibited nucleotide changes in the *CREBBP* gene. Ten of the 13 patients had exon as well as intron mutations, whereas three patients (RSTS7a, RSTS8a, and RSTS9a) showed exclusive intron mutations. Previous reports on RSTS patients (Bartsch *et al.* 2005; Roelfsema *et al.* 2005) suggest an even distribution of mutations in *CREBBP*; mutations identified in our study were primarily localized in the CH1, CH2, CH3 and HAT domains. Unlike previous reports, we did not identify any truncating mutations that could lead to a premature formation of a stop codon and protein truncation (Petrij *et al.* 2000a,b; Bartsch *et al.* 2002; Coupry *et al.* 2002; Bentivegna *et al.* 2006). Sixteen of the nucleotide changes arose *de novo* (figure 1; supplementary tables 3 and 4); 29 others could not be ascertained due to lack of corresponding data from parent samples. Fourteen variants identified in this study in the patients and parents, and four in the parents alone, were classified as mutations. However, one cannot rule out the possibility of these being polymorphisms specific to the particular ethnic subgroup of the patients.

All the 18 missense mutations identified in our study were subjected to pathogenicity assessment based on their CI, SIFT analysis and structural modifications in the protein structure due to these mutations. The application of these analyses led to the identification of 10 novel pathogenic mutations. Since it has been earlier reported that inactivation of HAT function is sufficient to cause RSTS (Kalkhoven *et al.* 2003), through functional assays we identified that three mutations located in the HAT domain – p.Glu1459Asp, p.Leu1668His and p.Glu1724Lys – brought about significant reduction in acetyltransferase activity. This, therefore,

Table 4. Second group of markers identified in significant association with RSTS in Indian patients

No.	Markers	Alleles	Exact position	*dbSNP/Novel*	Samples
1	M30	C/A	IVS23+164C>A	Novel	RSTS12a
2	M46	T/C	IVS18+3T>C	Novel	RSTS12a
3	M48	T/C	IVS11+81T>C	*rs130002*	C2,C17,C63, RSTS4a, RSTS12a
4	M49	C/A	IVS10+26C>A	Novel	RSTS12a
5	M55	A/C	p.Asn435Lys¥	Novel	RSTS11a,RSTS12a
6	M64	T/A	p.Ser1533Cys¥	Novel	RSTS13a
7	M65	T/G	p.Val89Gly¥	Novel	RSTS13a
8	M66	A/C	p.Asp35Ala¥	Novel	RSTS13a
9	M67	G/C	IVS21-28G>C	Novel	RSTS13a
10	M68	C/A	IVS30-178C>A	Novel	RSTS13a
11	M69	C/A	IVS30-178C>A	Novel	RSTS13a

¥Missense mutations

clearly underscores the *in vivo* relevance of these amino acids in CREBBP HAT function. The consequence of four other missense mutations – p.Ser1533Cys, p.Val1613Met, p.Thr1735Arg and p.Gln1765His – identified in the other patients could not be determined. However, we suggest that these could lead to a reduced HAT activity due to (i) a high CI (table 1), and (ii) their location in a region previously identified to be a critical determinant of HAT functionality (aa 1098–1757 in human CREBBP corresponding to mouse region aa 1099–1758) (Bannister and Kouzarides 1996; Ogryzko *et al.* 1996; Giles *et al.* 1997; Kalkhoven *et al.* 2003). (iii) Ser1533 is a part of the conserved coenzyme A (CoA) binding site (Kalkhoven *et al.* 2003) and mutations would possibly affect HAT activity and could be putative pathogenic mutations. We cannot negate the significance of the missense mutations identified outside the HAT domain, which might, for example, disrupt HAT–substrate interactions or destabilize the CREBBP protein, giving rise to the same phenotype as RSTS.

The effects of silent mutations were difficult to predict since they were always present along with missense mutations, except in patient GM1LBL, who lacked missense mutations. However, such mutations may be pathogenic as reported earlier in Marfan syndrome and ataxia telangiectasia (Liu *et al.* 1997; Teraoka *et al.* 1999) or may complement the effects of missense mutations in the patients. Similarly, it is difficult to assess the exact implication of intron mutations in causing the disease; however, silent mutations and intron variants may directly/indirectly affect the specific regulatory events involving CREBBP. Phylogenetic analysis indicated that while the nucleotide changes in *CREBBP* are highly diverse in RSTS, a high CI of the specific amino acid residues implicated in missense mutations highlights the importance of these mutated specific residues in the normal functioning of the protein and thus may possibly be *bona fide* causative lesions leading to manifestation of the syndrome. Identification of several novel polymorphisms that may be specific to the Indian population is an important feature of this study.

Linkage disequilibrium analysis revealed the following associations: (i) p.Asp1340Ala (missense mutation) with *rs129974 and* c.3836-206G>C (a novel SNP), and (ii) p.Asn435Lys (missense mutation) with *rs130002* and (iii) SNPs *rs129974, rs130002* and a novel SNP (*c.3836-*206G>C). Similar associations have been reported in other genetic conditions including Wolfram syndrome (Sandhu *et al.* 2007), schizophrenia (Chowdari *et al.* 2002) and systemic lupus erythematous (Graham *et al.* 2006). Thus, one may predict polymorphisms as one of the risk factors in the pathogenesis of RSTS; however; a larger sample size may be required to further validate this finding.

CREBBP is a protein for which a plethora of functions have been identified, including chromatin remodelling, as a co-activator for several transcription factors through its capacity to provide a scaffold for binding of several proteins, or providing a bridge for proteins in DNA binding (Ogryzko *et al.* 1996). Mutations in the conserved CH domains of CREBBP could lead to functional aberrations not only in basal transactivation, but also in the cascade of events that follow protein–protein interactions. The CH1 domain is also implicated in the activation of appropriate transcription of genes during development of the skeletal framework (Tsuda *et al.* 2003). During haematopoiesis, myogenesis and chondrocyte differentiation, binding of the CH3 domain of CREBBP to the transcription factors GATA-1, MyoD and Sox9 are required for lineage differentiation (Polesskaya *et al.* 2000; Blobel 2002; Tsuda *et al.* 2003). Thereby, mutations in CH3 could be involved in aberrant haematopoiesis and myogenesis, while mutations in the CH1 and/or CH3 domains could contribute to the skeletal abnormalities observed in RSTS patients.

The HAT domain, besides mediating acetyltransferase activity, is also known to interact with other transcription factors including Twist, *GLI3* (Dai *et al.* 1999) or have a high degree of homology in structure and functions with *CREBBP*, e.g. *EP300* with a 63% homology (Udaka *et al.* 2005). Mutations in such genes may also contribute to RSTS; altered Twist signalling is reported to complement the symptomatology of RSTS (Goodman and Smolik 2000). Overlapping phenotypes in individuals in whom partial symptoms of RSTS may be observed also correlate with other genes that have a high degree of homology with CREBBP/similar HAT function, e.g. *EP300* mutations have recently been reported to play a minor role in the aetiology of RSTS (Roelfsema *et al.* 2005), but could lead to phenotypes other than classical RSTS (Zimmermann *et al.* 2007). Thus, one cannot rule out the possibility of involvement of mutations in other genes besides *CREBBP* in the manifestations of RSTS. This is further highlighted as a large number of patients do not harbour mutations in the implicated genes – *CREBBP/EP300* (Bentivegna *et al.* 2006). Another possibility is that epigenetic regulation of chromatin packaging and accessibility to specific transcription factors may be altered/disrupted during development; the finding that treatment with histone deacetylase (HDAC) inhibitors can alleviate some of the symptoms of RSTS patients supports this notion (Hallam and Bourtchouladze 2006).

In conclusion, the present study demonstrated a high detection rate (10 out of 13 patients. [76.9%] showed the presence of causative mutations) in typical Indian RSTS patients. Further, 10 novel missense mutations that may be pathogenic were identified in the *CREBBP* gene in RSTS patients. The association of SNPs revealed in this study with RSTS suggests their additional contribution to the pathogenesis of the syndrome, which can be studied further in patient–parent cohorts. Together, our findings may

enhance the understanding of this syndrome at the molecular level.

Acknowledgements

We extend our sincere thanks to Dr G C Mishra for encouragement and support. Ms Neeti Sharma receives a research fellowship from the Council of Scientific and Industrial Research, New Delhi. We also thank Mr Satoor for DNA sequencing. The involvement of Dr P Gambhir, Pune; Dr G Oswal, Pune; Dr K Prabhakara, CDFD, Hyderabad, Dr Jyoti Sharma, Bharti Vidyapeeth Medical College, Pune, Dr S Phadke, SGPGIMS, Lucknow and Dr I C Verma, Sir Ganga Ram Hospital, New Delhi in the clinical identification of Rubinstein–Taybi syndrome in individuals, collection of samples and provision of clinical data is gratefully acknowledged.

References

Bannister A J and Kouzarides T 1996 The CBP co-activator is a histone acetyltransferase; *Nature (London)* **384** 641–643

Barrett J C, Fry B, Maller J and Daly M J 2005 Haploview: analysis and visualization of LD and haplotype maps; *Bioinformatics* **21** 263–265

Bartsch O, Locher K, Meinecke P, Kress W, Seemanova E, Wagner A, Ostermann K and Rodel G 2002 Molecular studies in 10 cases of Rubinstein–Taybi syndrome, including a mild variant showing a missense mutation in codon 1175 of CREBBP; *J. Med. Genet.* **39** 496–501

Bartsch O, Schmidt S, Richter M, Morlot S, Seemanova E, Wiebe G and Rasi S 2005 DNA sequencing of CREBBP demonstrates mutations in 56% of patients with Rubinstein–Taybi syndrome (RSTS) and in another patient with incomplete RSTS; *Hum. Genet.* **117** 485–493

Bentivegna A, Milani D, Gervasini C, Castronovo P, Mottadelli F, Manzini S, Colapietro P, Giordano L, *et al.* 2006 Rubinstein–Taybi Syndrome: spectrum of CREBBP mutations in Italian patients; *BMC. Med. Genet.* **7** 77

Blobel G A 2002 CBP and p300: versatile coregulators with important roles in hematopoietic gene expression; *J. Leukoc. Biol.* **71** 545–556

Chiang P W, Lee N C, Chien N, Hwu W L, Spector E and Tsai A C 2009 Somatic and germ-line mosaicism in Rubinstein–Taybi syndrome; *Am. J. Med. Genet. A* **149A** 1463–1467

Chowdari K V, Mirnics K, Semwal P, Wood J, Lawrence E, Bhatia T, Deshpande S N, B K Thelma, *et al.* 2002 Association and linkage analyses of RGS4 polymorphisms in schizophrenia; *Hum. Mol. Genet.* **11** 1373–1380

Chrivia J C, Kwok R P, Lamb N, Hagiwara M, Montminy M R and Goodman R H 1993 Phosphorylated CREB binds specifically to the nuclear protein CBP; *Nature (London)* **365** 855–859

Coupry I, Roudaut C, Stef M, Delrue M A, Marche M, Burgelin I, Taine L, Cruaud C, *et al.* 2002 Molecular analysis of the CBP gene in 60 patients with Rubinstein–Taybi syndrome; *J. Med. Genet.* **39** 415–421

Dai P, Akimaru H, Tanaka Y, Maekawa T, Nakafuku M and Ishii S 1999 Sonic Hedgehog-induced activation of the Gli1 promoter is mediated by GLI3; *J. Biol. Chem.* **274** 8143–8152

Davidovich O, Kimmel G and Shamir R 2007 GEVALT: an integrated software tool for genotype analysis; *BMC. Bioinformatics* **8** 36

de Bakker P I, Yelensky R, Pe'er I, Gabriel S B, Daly M J and Altshuler D 2005 Efficiency and power in genetic association studies; *Nat. Genet.* **37** 1217–1223

Downing A K, Knott V, Werner J M, Cardy C M, Campbell I D and Handford P A 1996 Solution structure of a pair of calcium-binding epidermal growth factor-like domains: implications for the Marfan syndrome and other genetic disorders; *Cell* **85** 597–605

Faivre L, Collod-Beroud G, Callewaert B, Child A, Binquet C, Gautier E, Loeys B L, Arbustini E, *et al.* 2009 Clinical and mutation-type analysis from an international series of 198 probands with a pathogenic FBN1 exons 24-32 mutation; *Eur. J. Hum. Genet.* **17** 491–501

Giles R H, Petrij F, Dauwerse H G, den Hollander A I, Lushnikova T, van Ommen G J, Goodman R H, Deaven L L, *et al.* 1997 Construction of a 1.2-Mb contig surrounding, and molecular analysis of, the human CREB-binding protein (CBP/CREBBP) gene on chromosome 16p13.3; *Genomics* **42** 96–114

Goodman R H and Smolik S 2000 CBP/p300 in cell growth, transformation, and development; *Genes Dev.* **14** 1553–1577

Graham R R, Kozyrev S V, Baechler E C, Reddy M V, Plenge R M, Bauer J W, Ortmann W A, Koeuth T, *et al.* 2006 A common haplotype of interferon regulatory factor 5 (IRF5) regulates splicing and expression and is associated with increased risk of systemic lupus erythematosus; *Nat. Genet.* **38** 550–555

Hallam T M and Bourtchouladze R 2006 Rubinstein–Taybi syndrome: molecular findings and therapeutic approaches to improve cognitive dysfunction; *Cell Mol. Life Sci.* **63** 1725–1735

Hebsgaard S M, Korning P G, Tolstrup N, Engelbrecht J, Rouze P and Brunak S 1996 Splice site prediction in *Arabidopsis thaliana* pre-mRNA by combining local and global sequence information; *Nucleic Acids Res.* **24** 3439–3452

Hennekam R C, Stevens C A and Van de Kamp J J 1990 Etiology and recurrence risk in Rubinstein–Taybi syndrome; *Am. J. Med. Genet. Suppl* **6** 56–64

Kalkhoven E, Roelfsema J H, Teunissen H, den Boer A, Ariyurek Y, Zantema A, Breuning M H, Hennekam R C, *et al*. 2003 Loss of CBP acetyltransferase activity by PHD finger mutations in Rubinstein–Taybi syndrome; *Hum. Mol. Genet.* **12** 441–450

Kalos MH 1986 *Monte Carlo methods* (New Jersey: JohnWiley and Sons)

Li N and Stephens M 2003 Modeling linkage disequilibrium and identifying recombination hotspots using single-nucleotide polymorphism data; *Genetics* **165** 2213–2233

Lin D Y 2006 Evaluating statistical significance in two-stage genomewide association studies; *Am. J. Hum. Genet.* **78** 505–509

Liu W, Qian C and Francke U 1997 Silent mutation induces exon skipping of fibrillin-1 gene in Marfan syndrome; *Nat. Genet.* **16** 328–329

Liu X, Wang L, Zhao K, Thompson P R, Hwang Y, Marmorstein R and Cole P A 2008 The structural basis of protein acetylation by the p300/CBP transcriptional coactivator; *Nature (London)* **451** 846–850

Michail J, Matsoukas J and Theodorou S 1957 [Arched, clubbed thumb in strong abduction-extension & other concomitant symptoms.]; *Rev. Chir. Orthop. Reparatr. Appar. Mot.* **43** 142–146

Murata T, Kurokawa R, Krones A, Tatsumi K, Ishii M, Taki T, Masuno M, Ohashi H, *et al*. 2001 Defect of histone acetyltransferase activity of the nuclear transcriptional coactivator CBP in Rubinstein–Taybi syndrome; *Hum. Mol. Genet.* **10** 1071–1076

Ng P C and Henikoff S 2001 Predicting deleterious amino acid substitutions; *Genome Res.* **11** 863–874

Ng P C and Henikoff S 2003 SIFT: predicting amino acid changes that affect protein function; *Nucleic Acids Res.* **31** 3812–3814

Ogryzko V V, Schiltz R L, Russanova V, Howard B H and Nakatani Y 1996 The transcriptional coactivators p300 and CBP are histone acetyltransferases; *Cell* **87** 953–959

Olson J M and Wijsman E M 1994 Design and sample-size considerations in the detection of linkage disequilibrium with a disease locus; *Am. J. Hum. Genet.* **55** 574–580

Petrij F, Dauwerse H G, Blough R I, Giles R H, van der Smagt J J, Wallerstein R, Maaswinkel-Mooy P D, van Karnebeek C D, *et al*. 2000a Diagnostic analysis of the Rubinstein–Taybi syndrome: five cosmids should be used for microdeletion detection and low number of protein truncating mutations; *J. Med. Genet.* **37** 168–176

Petrij F, Dorsman J C, Dauwerse H G, Giles R H, Peeters T, Hennekam R C, Breuning M H and Peters D J 2000b Rubinstein–Taybi syndrome caused by a de novo reciprocal translocation t(2;16)(q36.3;p13.3); *Am. J. Med. Genet.* **92** 47–52

Petrij F, Giles R H, Dauwerse H G, Saris J J, Hennekam R C, Masuno M, Tommerup N, van Ommen G J, *et al*. 1995 Rubinstein–Taybi syndrome caused by mutations in the transcriptional co-activator CBP; *Nature (London)* **376** 348–351

Polesskaya A, Duquet A, Naguibneva I, Weise C, Vervisch A, Bengal E, Hucho F, Robin P and Harel-Bellan A 2000 CREB-binding protein/p300 activates MyoD by acetylation; *J. Biol. Chem.* **275** 34359–34364

Qin Z S, Niu T and Liu J S 2002 Partition–ligation–expectation–maximization algorithm for haplotype inference with single-nucleotide polymorphisms; *Am. J. Hum. Genet.* **71** 1242–1247

Roelfsema J H, White S J, Ariyurek Y, Bartholdi D, Niedrist D, Papadia F, Bacino C A, den Dunnen J T, *et al.* 2005 Genetic heterogeneity in Rubinstein–Taybi syndrome: mutations in both the CBP and EP300 genes cause disease; *Am. J. Hum. Genet.* **76** 572–580

Rubinstein J H and Taybi H 1963 Broad thumbs and toes and facial abnormalities. A possible mental retardation syndrome; *Am. J. Dis. Child* **105** 588–608

Sandhu M S, Weedon M N, Fawcett K A, Wasson J, Debenham S L, Daly A, Lango H, Frayling T M, *et al.* 2007 Common variants in WFS1 confer risk of type 2 diabetes; *Nat. Genet.* **39** 951–953

Schaid D J, Rowland C M, Tines D E, Jacobson R M and Poland G A 2002 Score tests for association between traits and haplotypes when linkage phase is ambiguous; *Am. J. Hum. Genet.* **70** 425–434

Stevens C A, Carey J C and Blackburn B L 1990 Rubinstein–Taybi syndrome: a natural history study; *Am. J. Med. Genet. Suppl.* **6** 30–37

Teraoka S N, Telatar M, Becker-Catania S, Liang T, Onengut S, Tolun A, Chessa L, Sanal O, *et al*. 1999 Splicing defects in the ataxia-telangiectasia gene, ATM: underlying mutations and consequences; *Am. J. Hum. Genet.* **64** 1617–1631

Tsuda M, Takahashi S, Takahashi Y and Asahara H 2003 Transcriptional co-activators CREB-binding protein and p300 regulate chondrocyte-specific gene expression via association with Sox9; *J. Biol. Chem.* **278** 27224–27229

Udaka T, Samejima H, Kosaki R, Kurosawa K, Okamoto N, Mizuno S, Makita Y, Numabe H, *et al.* 2005 Comprehensive screening of CREB-binding protein gene mutations among patients with Rubinstein–Taybi syndrome using denaturing high-performance liquid chromatography; *Congenit. Anom. (Kyoto)* **45** 125–131

Verhoef S, Bakker L, Tempelaars A M, Hesseling-Janssen A L, Mazurczak T, Jozwiak S, Fois A, Bartalini G, *et al*. 1999 High rate of mosaicism in tuberous sclerosis complex; *Am. J. Hum. Genet.* **64** 1632–1637

Wani A A, Sharma N, Shouche Y S and Bapat S A 2006 Nuclear-mitochondrial genomic profiling reveals a pattern of evolution in epithelial ovarian tumor stem cells; *Oncogene* **25** 6336–6344

Zaykin D V, Westfall P H, Young S S, Karnoub M A, Wagner M J and Ehm M G 2002 Testing association of statistically inferred haplotypes with discrete and continuous traits in samples of unrelated individuals; *Hum. Hered.* **53** 79–91

Zhang K, Calabrese P, Nordborg M and Sun F 2002 Haplotype block structure and its applications to association studies: power and study designs; *Am. J. Hum. Genet.* **71** 1386–1394

Zimmermann N, Acosta A M, Kohlhase J and Bartsch O 2007 Confirmation of EP300 gene mutations as a rare cause of Rubinstein–Taybi syndrome; *Eur. J. Hum. Genet.* **15** 837–842

MS received 23 September 2009; accepted 1 April 2010

*e*Publication: 29 April 2010

Corresponding editor: Pragna I Patel

Spectrum of *CREBBP* mutations in Indian patients with Rubinstein–Taybi syndrome

Neeti Sharma, Avinash M Mali and Sharmila A Bapat

J. Biosci. **35(2), June 2010, 187–202 © Indian Academy of Sciences**

Supplementary tables and figures

Supplementary table 1. Details of the symptoms of Indian RSTS patients

	Patients	GM1LBL	GM3SF	RSTS3a	RSTS4a	RSTS5a	RSTS6a	RSTS7a	RSTS8a	RSTS9a	RSTS10a	RSTS11a	RSTS12a	RSTS13a
	Parents' sample available	No	No	Yes	Yes	Yes	Yes	No	only Mother	Yes	Yes	No	No	Unknown
1	Consanguinity	Unknown	Unknown	No	No	Yes	No	Unknown	Yes	No	Unknown	Unknown	Unknown	Unknown
2	Age	6 months	3 years	9 years	8 years	7 years	11 years	5 days	6 years	8 years	10 years	9 years	3 months	16 years
3	Sex	Male	Female	Male	Male	Male	Male	Female	Female	Male	male	Female	Female	Male
4	Growth retardation	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
5	Mental and psychometric retardation	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
6	Severity of mental and psychometric retardation	Severe	Severe	Severe	Mild	Severe	Mild	Severe	Severe	Severe	Severe	Severe	Severe	Severe
7	Heavy and arched eyebrows	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
8	Long eyelashes	Yes	Yes	Yes	Yes	Yes	Yes	Unknown	Yes	Yes	Yes	Yes	Unknown	Yes
9	Eye abnormality (myopia/ cataracts/ glaucoma/, coloboma)	Yes	Yes	Yes	Yes	Yes	Yes	Unknown	Yes	Yes	Yes	Unknown	Unknown	Yes
10	Prominent beaked nose	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
11	Low-set ears	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
12	Highly arched palate	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
13	Downward-slanting palpebral fissures	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
14	Radially deviated thumbs	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	No	Yes	Yes	Yes
15	Broad thumb	Yes	Yes	Yes	No	Yes	No	Yes	Yes	Yes	Yes	Yes	Yes	Yes
16	Hypotonic	Yes	Yes	Yes	No	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
17	Microcepahaly	Yes	Yes	Yes	No	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
18	Deviation from classical	Classical	Classical	Classical	-	Classical	-	Classical	Classical	Classical	Classical	Classical	Classical	Classical

Supplementary table 2. Primers used for CREBBP gene amplification

Primer	Exons	Forward	Reverse
CBPE2	Exon2	5′-CTG TTT TCG CGA GCA GGT-3′	5′-ATC GGT ATC CGC GAC CAC-3′
CBPE3a	Exon3	5′-AACGTGGCAGTTGGAGAGC-3′	5′-GTGCTGGCTGCCTGTTTAG-3′
CBPE3b	Exon3	5′-TAAACAGGCAGCCAGCACC-3′	5′-TACCTTGGCCATGCCTCCTG-3′
CBPE4	Exon4	5′-TCACCACAGCTGATGTTACCTC-3′	5′-TCCTGTTGCTAGCTCATCACAG-3′
CBPE5	Exon5	5′-GTGGTCGGTATTATCCATCAGC-3′	5′-GGCAAATTCTTCCTGACCTC-3′
CBPE6	Exon6	5′-TGTTTGTAAGTTGAGGACTGCC-3′	5′-TGTACCTTGGGCTGCTGTC-3′
CBPE7	Exon7	5′-TTCCTTTCTCCCTAGGCTTG-3′	5′-GAGAGTTCCTTCACCTACCCAG-3′
CBPE8	Exon8	5′-GGTGGCATGTTGTTATCGTC	5′-AAGTCAGGAGAGCCAAGTGATG-3′
CBPE9	Exon9	5′-GTGGTGGCAGAAGAACCTTAC-3′	5′-AGAGTGTGGGAATCTCTGGC-3′
CBPE10	Exon10	5′-AGGTGATTCTCCCGCCTCAG-3′	5′-ATCTGGGAAGTCTCCTTGGTC-3′
CBPE11	Exon11	5′-GCCAGTGTACTCAATGAGGAAG-3′	5′-ACATCAACAGCTTCTGCAGG-3′
CBPE12	Exon12	5′-CTC TCT TGG GAA GGC CCT TTA C-3′	5′-GAG GCA GGA GAA TGG TGT GAA C-3′
CBPE13	Exon13	5′-TGTCCTCGTCCTTCCCTGGTAC -3′	5′-TGTGAGAGGGAGGGCTATCTGC -3′
CBPE14	Exon14	5′-TGTGCTTGCCATCCTCTGGG -3′	5′-TCATGGGTCCTTCTCACTCGC -3′
CBPE15	Exon15	5′-CTC TGT CCA TTT CTG GTA GGG -3′	5′-GGA TGG TCT TGA ACT CAT GGG -3′
CBPE16	Exon16	5′-TCTGGCTGCTGTTACCTCCCTC-3′	5′-TACCCATGGCAGGCTCCAAG -3′
CBPE17	Exon17	5′-CTG ACT CTG TGC TGA AGC GAC-3′	5′-CTC TCA CTC CTG CCA TGA GC-3′
CBPE18	Exon18	5′-CCACCATGCCCAGCCTGAAAT G-3′	5′-TCCAAGGACCAGGGCTTCATG-3′
CBPE19	Exon19	5′-CCCAGCCGATTCTATGTAGCAG -3′	5′-CTCCCAGTATACAGGCGTGGTC -3′
CBPE20	Exon20	5′-CTGAAGCGATGTGATGGGTG-3′	5′-TGTCGCCAAGGCTGATAAC-3′
CBPE21	Exon21	5′-GATGATGGCCTGTGATGTGTGC -3′	5′-TGGTCAAGGCTGCAGTGAGTC -3′
CBPE22	Exon22	5′-GGCGTGGTGGTTTGTACCTG -3′	5′-CCCACAACCCACTCCATAAGG-3′
CBPE23-24	Exon23&24	5′-CCCAGAGAACAAGGTGTGGAC G-3′	5′-TTGACAACCGAACCTCAGAACC -3′
CBPE25	Exon25	5′-ATCCATCTGCTCCGCCCTAGAG-3′	5′-GCTACTGCACGCATTCGCTG -3′
CBPE26	Exon26	5′-CATTCACAGAGGTGCAGTTCCC -3′	5′-TAGGTAAGAGCGGGTCCCACAC-3′
CBPE27	Exon27	5′-GAGCCCATGAGGTTGAGGTTGC -3′	5′-TTCAGGTGATCCACCGCAGCAG -3′
CBPE28	Exon28	5′-CCCAAGGGAGAAGTGTTGCAG -3′	5′-TGCCCACTGATGGAAACAGC-3′
CBPE29	Exon29	5′-CTGTTTCCATCAGTGGGCATTG -3′	5′-TTTACATCAGCGGGTTGTGTCC -3′
CBPE30	Exon30	5′-TGGAATCTTCCGCATCTGGTTG -3′	5′-CCTGATGCCTTGGGATGGAAC-3′
CBPE31	Exon31	5′-ATCAGGGCCATCAGGCTCAG-3′	5′-GCGGGTGTTTCCAGCCATTTAG-3′
CBPE32	Exon32(i)	5′-GCTGTTGCCAGGAGTGAAGGTC-3′	5′-GCTGGCGGAGCTTGTGTTTG-3′
CBPE32	Exon32(ii)	5′-TCATCGCCCTCTGCTGCTAC -3′	5′-TGGGCCTGCATGGATATCAC-3′
CBPE32	Exon32(iii)	5′-CCTGTACCGGGTGAACATCAAC -3′	5′-GCTGCCTCCGTAACATTTCTCG -3′
CBPE32	Exon32(iv)	5′-CATGAACCCAGGACACAACCC -3′	5′-GGGAGCATTGCACTCTGTTCG -3′

Supplementary table 3. List of eight groups classified on the basis of markers showing independent association or interdependency

First group					
S. No.	Markers	Alleles	Exact position	Samples	*dbSNP/Novel*
1	M2	C/A	3185	RSTS10a	Novel
2	M6	A/T	4503	RSTS10a	Novel
3	M10	C/T	5478	RSTS10a	Novel
Second group					
1	M7	C/T	4628	RSTS5a	Novel
2	M8	T/A	4794	RSTS5a	Novel
3	M53	T/C	65411	RSTS5a	Novel
4	M54	C/G	65427	RSTS5a	Novel
5	M56	A/C	65465	RSTS5a	Novel
6	M57	T/C	65467	RSTS5a,RSTS11a	Novel
7	M58	A/C	65527	RSTS5a	Novel
8	M59	T/G	65648	RSTS5a	Novel
9	M62	A/T	123800	RSTS11a	Novel
10	M65	T/G	124424	RSTS13a	Novel
11	M66	A/C	124262	RSTS13a	Novel
Third group					
1	M12	G/T	5540	C75,C76,C86	Novel
2	M29	T/G	18478	C75,C83	Novel
3	M52	C/T	65348	C86	Novel
Fourth group					
1	M13	C/T	9919	C2,C13,C17	Novel
2	M36	T/C	25081	C1,C2,C3,C6,C9,C13,C17,C19,C67,C69,RSTS6b	Novel
3	M38	G/A	25230	C20,C69	Novel
Fifth group					
1	M14	A/T	10029	C5	Novel
2	M15	A/C	10054	C5,C9,C16	Novel
3	M16	T/G	10107	C9	Novel
4	M17	T/C	10152	C5,C9	Novel
Sixth group					
1	M19	A/G	11755	GM1LBL	Novel
2	M26	T/G	13946	GM3SF,RSTS3a	Novel
3	M31	T/G	18607	GM1LBL	Novel
4	M32	C/G	18617	C21,C17,GM1LBL	Novel
5	M33	G/T	18724	C1,C23,GM1LBL,GM3SF,RSTS3a,RSTS3c,RSTS4a ,RSTS4b,RSTS6a,RSTS6b	*rs129974*
6	M61	C/T	98804	RSTS3a	Novel
7	M41	G/C	25445	C8,GM3SF,RSTS3a,RSTS3b,RSTS4b,RSTS4a,RSTS 4c,RSTS5a,RSTS5c,RSTS6c,RSTS7a	Novel
8	M60	C/T	98802	RSTS3a	Novel
9	M28	A/G	18455	GM1LBL	Novel
Seventh group					

1	M30	C/A	18547	RSTS12a	Novel
2	M69	A/T	23282	RSTS13a	Novel
3	M46	T/C	32284	RSTS12a	Novel
4	M48	T/C	51363	C2,C17,C63,RSTS4a,RSTS12a	*rs130002*
5	M49	C/A	52107	RSTS12a	Novel
6	M55	A/C	65439	RSTS11a,RSTS12a	Novel
7	M65	T/G	124424	RSTS13a	Novel
8	M66	A/C	124262	RSTS13a	Novel
9	M64	T/A	9600	RSTS13a	Novel
10	M67	G/C	3334	RSTS13a	Novel
11	M68	C/A	5533	RSTS13a	Novel
Eighth group					
1	M43	G/A	30603	C2,C19,C52,C69	Novel
2	M44	A/T	31015	C18	Novel
3	M45	A/C	31609	C62	Novel
4	M50	C/T	54071	C62	Novel

Supplementary table 4. Exon mutations in RSTS patients in the Indian population

S.No	Gene Position	Sense / Missense	Patient	Exon	Amino acid change	Novel/*de novo*/SNP	Domain
1	g.3276G>C	missense	RSTS13a	32	p.Thr1735Arg	Novel	CH3
2	g.4503A>T	missense	RSTS10a	31	p.Gln1765His	*De novo;* novel	CH3
3	g.4627C>T	missense	RSTS5a	31	p.Glu1724Lys	*De novo;* novel	CH3
4	g.4794T>A	missense	RSTS5a	31	p.Leu1668His	*De novo;* novel	HAT
5	g.9600T>A	missense	RSTS13a	29	p.Ser1533Cys	Novel	HAT
6	g.5262C>T	missense	RSTS13a	30	p.Val1613Met	Novel	HAT
7	g.12009C>T	missense	RSTS3a	27	p.Glu1459Asp	Novel	CH2
8	g.13140T>C	missense	RSTS6a	26	p.Glu1384Gly	*De novo;* novel	CH2
9	g.13946T>G	missense	GM3SF, RSTS3a	25	p.Asp1340Ala	Novel	CH2
10	g.25233A>T	missense	GM3SF, RSTS3a, RSTS4a, RSTS5a	21	p.His1235Gln	Novel	CH2
11	g.65427C>G	missense	RSTS5a	6	p.Lys439Asn	*De novo;* novel	CH1
12	g.65439A>C	missense	RSTS11a, RSTS12a	6	p.Asn435Lys	Novel	CH1
13	g.65465A>C	missense	RSTS5a	6	p.Cys427Gly	*De novo;* novel	CH1
14	g.65467T>C	missense	RSTS5a, RSTS11a	6	p.Asp426Gly	Novel	CH1
15	g.65527A>C	missense	RSTS5a	6	p.Val406Gly	*De novo;* novel	CH1
16	g.123800A>T	missense	RSTS11a	3	p.Leu243Gly	Novel	NTAD
17	g.124262A>C	missense	RSTS13a	3	p.Val89Gly	Novel	NHRD
18	g.124424T>G	missense	RSTS13a	3	p.Asp35Ala	Novel	NHRD
19	g.4808C>T	sense(silent)	RSTS5a	31	p.Gly1663Gly	*De novo;* novel	HAT
20	g.5284C>T	sense(silent)	RSTS10a	31	p.Lys1605Lys	*De novo;* novel	HAT
21	g.5341C>T	sense(silent)	RSTS12a	30	p.Lys1586Lys	Novel	HAT
22	g.13894A>G	sense(silent)	GM3SF, RSTS3a	25	p.Val1357Val	Novel	CH2
23	g.13929A>G	sense(silent)	GM3SF, RSTS3a	25	p.Leu1346Leu	Novel	CH2
24	g.54717T>G	sense(silent)	GM3SF	8	p.Pro532Pro	Novel	KIX
25	g.65472T>C	sense(silent)	RSTS5a	6	p.Arg424Arg	*De novo;* novel	CH1

Supplementary table 5. Intron mutations in RSTS patients in the Indian population

S.No	Gene position	Exact position	Patient	Splice site	Novel/*De novo*/SNP	Domain
1	g.3334G>C	c.3379-28G>C	RST13a	No	Novel	HAT
2	g.3185C>A	c.3379+76C>A	RSTS10a	No	Novel	HAT
3	g.5478C>T	c.4728-108C>T	RSTS10a	No	*De novo; novel*	HAT
4	g.5533C>A	c.4890-178C>A	RSTS13a	No	Novel	HAT
5	g.11755A>G	c.4394+238A>G	GM1LBL	No	Novel	HAT
6	g.12857A>G	c.4113-154A>G	RSTS6a	No	*De novo; novel*	HAT
7	g..12833G>A	c.4113-178G>A	RSTS6a	No	*De novo; novel*	HAT
8	g.14259G>A	c.3982-277G>A	RSTS4a	No	*De novo; novel*	HAT
9	g.18455A>G	c.3914-277A>A	GM1LBL	No	Novel	HAT
10	g.18547C>A	c.3914+164C>A	RSTS12a	No	Novel	HAT
11	g.18617C>G	c.3914+93C>G	GM1LBL	No	Novel	HAT
12	g.18607T>G	c.3914+103T>G	GM1LBL	No	Novel	HAT
13	g.23282A>T	c.3836-67A>T	RST13a	No	Novel	HAT
14	g.32284T>C	c.3369+3T>C	RSTS12a	No	Novel	BROMO
15	g.51363T>C	c.2113+81T>C	RSTS4a	No	*rs130002*	KIX
16	g.52107C>A	c.1941+26C>A	RSTS12a	No	Novel	KIX
17	g.54546C>T	c.1676-235C>T	RSTS3a	No	Novel	KIX
18	g.54548C>T	c.1676-237C>T	RSTS3a	No	Novel	KIX
19	g.65411T>C	c.1330+3T>C	RSTS5a	Part of donor splice site	*De novo; novel*	KIX
20	g.65648T>G	c.1216-121T>G	RSTS5a	No	*De novo; novel*	KIX

Supplementary table 6. List of fifteen nucleotide changes both in the patients as well as parents

S.No.	Position	Intron/Exon	Samples	Exact position/ change	dbSNP/Novel
1	g.25233A>G	Ex21	RSTS3c,RSTS5a	p.His1235HiS	Novel
2	g.23474C>T	IVS21	RSTS8a,GM1LBL,RSTS4b,RSTS6a,RSTS6b,RSTS8b	c.3779-358C>T	Novel
3	g.23289T>C	IVS21	GM3SF,RSTS3a,RSTS3c,RSTS3b,RSTS4a,RSTS5a,RSTS5b,RSTS6a,RSTS6c	c.3779-173T>C	Novel
4	g.25445G>C	IVS21	GM3SF,RSTS3a,RSTS3b,RSTS4a,RSTS4b,RSTS4c,RSTS5a,RSTS5c,RSTS6c,RSTS7a	c.3779-206G>C	Novel
5	g.10446C>A	IVS27	RSTS3a,RSTS3b,RSTS3c,RSTS4c,RSTS5a,RSTS5b,RSTS6a,RSTS6b	c.4394-198C>A	Novel
6	g.46880A>T	IVS14	RSTS3b,RSTS4c,RSTS5b,RSTS6b,RSTS6c,RSTS8a,RSTS9a,RSTS9b,	c.2463+304A>T	Novel
7	g.46900A>T	IVS14	RSTS3b,RSTS4c,RSTS5b,RSTS6b,RSTS6c,RSTS8a,RSTS9a,RSTS9b	c.2463-284A>T	Novel
8	g.46908A>C	IVS14	RSTS3b,RSTS4c,RSTS5b,RSTS6b,RSTS6c,RSTS8a,RSTS9a,RSTS9b	c.2463+176A>C	Novel
9	g.46966A>C	IVS14	RSTS3b,RSTS4c,RSTS5b,RSTS6b,RSTS6c,RSTS8a,RSTS9a,RSTS9b	c.2463+218A>C	Novel
10	g.47122_ 47133insA	IVS14	RSTS3c,RSTS5a,RSTS5c,RSTS6a,RSTS9c	c.24634+62insA	Novel
11	g.12969G>A	IVS25	RSTS6a,RSTS6b,GM1LBL,RSTS4b	c.4113-42G>A	Novel
12	g.13739T>C	IVS25	RSTS3c,RSTS4a,RSTS4b	c.4280+93T>C	Novel
13	g.18795G>A	IVS23	GM1LBL,GM3SF,RSTS3a,RSTS3c,RSTS4a,RSTS4b, RSTS6b	c.3982+9G>A	Novel
14	g.13190A>G	IVS26	RSTS6a,RSTS6b	c.4394+43A>G	Novel
15	g.52289T>C	IVS9	RSTS10b,RSTS10c,RSTS13a	c.1941-139T>C	Novel

Supplementary table 7. List of variants identified in the parents only

S.No.	Position	Intron/Exon	Samples	Exact position	dbSNP/Novel
1	g.23322A>T	IVS21	RSTS3c	c.3836-206A>T	Novel
2	g.23339A>C	IVS21	RSTS3c	c.3836-223A>C	Novel
3	g.12796C>T	IVS25	RSTS6b	c.4280-215C>T	Novel
4	g.25045A>C	IVS21	RSTS6c	c.3836+114A>C	Novel

Supplementary table 8. Single-nucleotide polymorphisms (SNPs) identified in the Indian population SNPs located in exons

S. No.	SNP	Synonymous/ Non-synonymous	Location	Amino acid change	Known SNP (dbSNP)	Samples
1	g.25230G>A	Synonmous	Exon21	p.Phe1236Phe	Novel	C20,C69
2	g.10107T>G	Synonmous	Exon28	p.Ala1512Ala	Novel	C9
3	g.10152T>C	Synonmous	Exon28	p.Lys1497Lys	Novel	C5, C9
4	g.18724G>T	Synonmous	Exon23	p.Ile1300Ile	*rs129974*	C1, C23, GM1LBL, GM3SF, RSTS3a, RSTS3c, RSTS4a,RSTS6a
5	g.3227G>T	Non-synonmous	Exon32	p.Asp1751Glu	Novel	C81
6	g.54203A>G	Synonmous	Exon9	p.His595His	Novel	C57
7	g.123970T>C	Synonmous	Exon3	p.Pro186Pro	Novel	C51

SNPs located in introns and I/E boundaries

S. No.	SNP	Intron	Exact position	Part of splice site	Known SNP (dbSNP)	Samples
1	g.65348C>T	Intron6	c.1330-3C>T	No	Novel	C86
2	g.54071C>T	Intron9	c.1823+94C>T	No	Novel	C62
3	g.51363T>C	Intron11	c.2113+81T>C	No	*rs130002*	C2, C17, C63, RSTS12a
4	g.31609A>C	Intron18	c.3369-128A>C	No	Novel	C62
5	g.31015A>T	Intron20	c.3698-295A>T	No	Novel	C18
6	g.30603G>A	Intron20	c.3698+58G>A	No	Novel	C2, C19, C52, C69
7	g.25469G>T	Intron20	c.3698-230G>T	No	Novel	C17
8	g.25445G>C	Intron20	c.3698-206G>C	No	Novel	C8, GM3SF, RSTS3a, RSTS3b, RSTS4b, RSTS4a, RSTS4c, RSTS5a, RSTS5c, RSTS6c,RSTS7a
9	g.25355T>G	Intron20	c.3698-116T>G	No	Novel	C19
10	g.25081T>C	Intron21	c.3779+77T>C	No	*rs129982*	C1, C2,C3, C6, C9, C13, C17, C19, C67, C69, CRSTS6b
11	g.24982G>T	Intron21	c.3779+177G>T	No	Novel	C4, C6, C10, C22, RSTS3c, RSTS5b, RSTS5c, RSTS6b, RSTS6c
12	g.23289T>C	Intron21	c.3779-173T>C	No	*rs129981*	C1, C8, C21, C24,GM3SF, RSTS3a, RSTS3c, RSTS4a, RSTS5a, RSTS5b, RSTS6a, RSTS6c
13	g.18617C>G	Intron23	c.3914+93C>G	No	Novel	C21, C17, GM1LBL
14	g.18478T>G	Intron23	c.3914+232T>G	No	Novel	C75, C83
15	g.13013C>A	Intron25	c.4113-3C>A	No	Novel	C53
16	g.12969G>A	Intron26	c.4280+42G>A	No	*rs129667*	C56,RSTS6a,RSTS6b, GM1LBL,RSTS4b
17	g.11973T>A	Intron27	c.4394+19T>A	No	Novel	C10
18	g.9919C>T	Intron28	c.4560+264C>T	No	Novel	C2, C13, C17

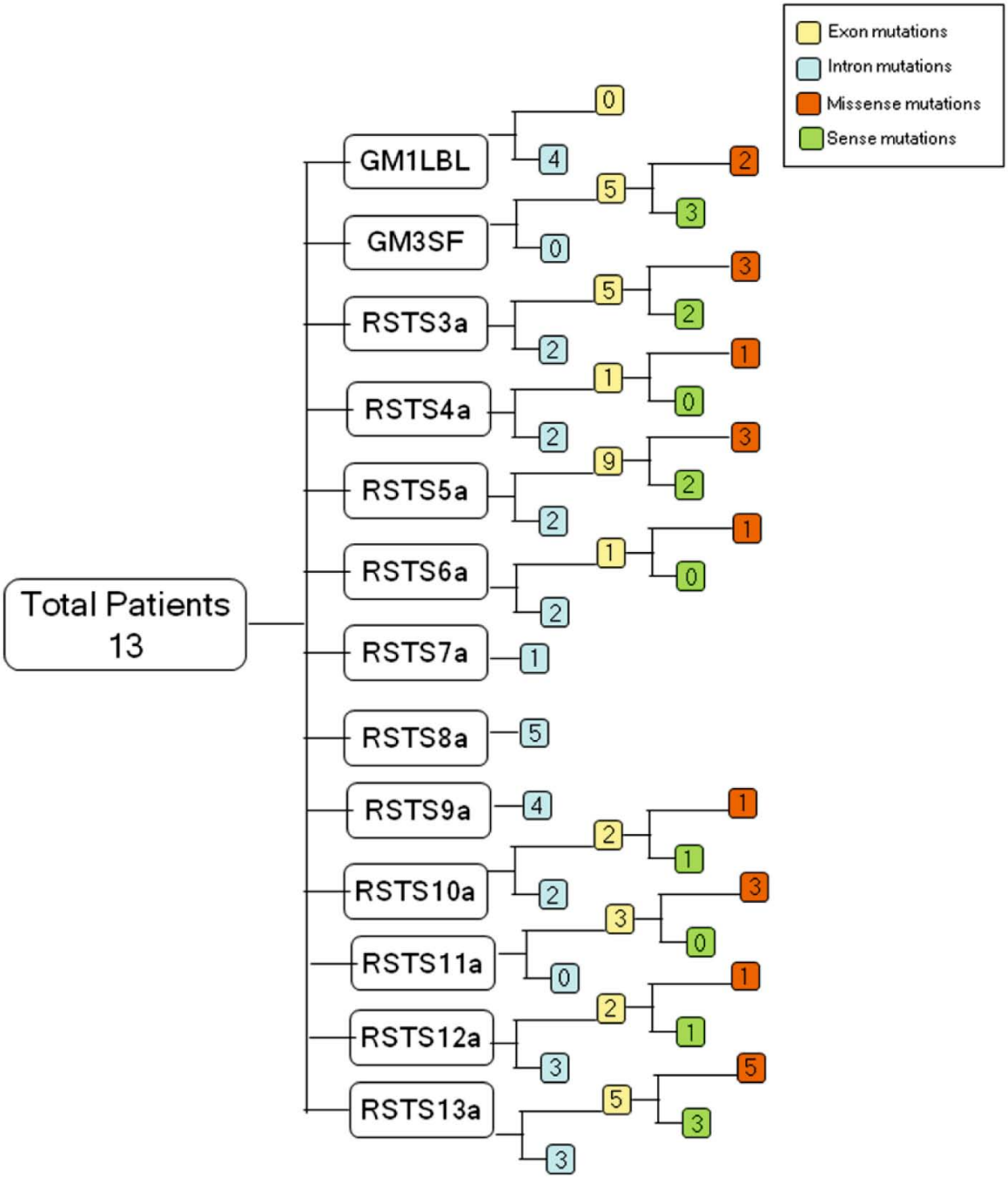
19	g.10029A>T	Intron28	c.4560+54A>T	No	Novel	C5
20	g.10054A>C	Intron28	c.4560+19A>C	No	Novel	C5, C9, C16
21	g.10268A>T	Intron28	c.4560-20A>T	Part of acceptor splice site*	Novel	C13,16
22	g.5540G>T	Intron30	c.4890-170G>T	No	Novel	C75, C76, C86
23	g.5522A>G	Intron30	c.4890-152A>G	No	Novel	C75, C76, C43, RSTS10b, C43
24	g.5148C>A	Intron31	c.5172+61C>A	No	Novel	C76
25	g.3425A>T	Intron31	c.5172-119A>T	No	Novel	C77, C78, C81, C82, C83, C85
26	g.3372C>T	Intron31	c.5172-65C>T	No	Novel	C75

Supplementary table 9. Calculated chi-square and P values (uncorrected) for the allele frequencies in RSTS cases vs controls

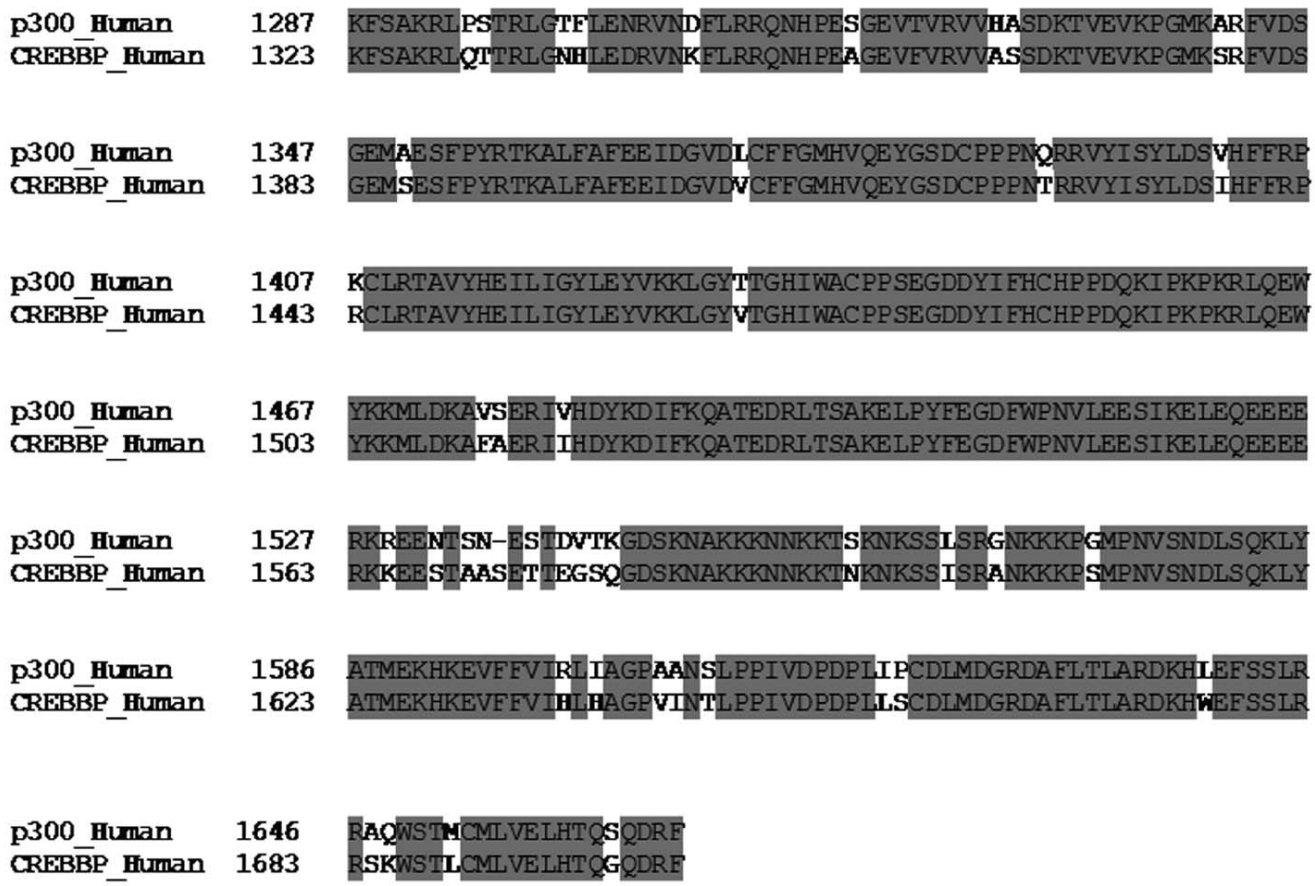
S.No.	Name	Chi-square values	Uncorrected *P* value
1	M1	0.116	0.734
2	M2	8.727	0.0031
3	M3	0.116	0.734
4	M4	0.116	0.734
5	M5	0.707	0.4004
6	M6	8.727	0.0031
7	M7	8.727	0.0031
8	M8	8.727	0.0031
9	M9	0.116	0.734
10	M10	8.727	0.0031
11	M11	0.468	0.4941
12	M12	0.349	0.5545
13	M13	0.349	0.5545
14	M14	0.116	0.734
15	M15	0.349	0.5545
16	M16	0.116	0.734
17	M17	0.232	0.6301
18	M18	0.232	0.6301
19	M19	8.727	0.0031
20	M20	0.116	0.734
21	M21	8.727	0.0031
22	M22	8.727	0.0031
23	M23	4.857	0.0275
24	M24	0.116	0.734
25	M25	8.727	0.0031
26	M26	17.524	2.84E-05
27	M27	8.727	0.0031
28	M28	8.727	0.0031
29	M29	0.116	0.734
30	M30	17.524	2.84E-05
31	M31	8.727	0.0031
32	M32	1.738	0.1874
33	M33	17.722	2.56E-05
34	M34	8.433	0.0037
35	M35	1.074	0.3001
36	M36	1.323	0.25
37	M37	0.116	0.734
38	M38	0.232	0.6301
39	M39	35.33	2.78E-09
40	M40	0.116	0.734
41	M41	8.433	0.0037
42	M42	0.468	0.4941
43	M43	0.116	0.734
44	M44	0.232	0.6301
45	M45	0.116	0.734
46	M46	17.524	2.84E-05
47	M47	0.116	0.734
48	M48	10.46	0.0012
49	M49	17.524	2.84E-05
50	M50	0.116	0.734
51	M51	0.116	0.734
52	M52	0.116	0.734
53	M53	8.727	0.0031
54	M54	8.727	0.0031
55	M55	26.391	2.79E-07
56	M56	8.727	0.0031
57	M57	17.524	2.84E-05
58	M58	8.727	0.0031
59	M59	8.727	0.0031
60	M60	8.727	0.0031
61	M61	8.727	0.0031
62	M62	8.727	0.0031
63	M63	0.116	0.734
64	M64	8.727	0.0031
65	M65	8.727	0.0031
66	M66	8.727	0.0031
67	M67	8.727	0.0031
68	M68	8.727	0.0031
69	M69	8.727	0.0031

Supplementary table 10. List of references with the number of patients reported in the study

Reference cited	Patient sample used in the study
Coupry *et al*. 2002	60
Bartsch *et al*. 2002	10
Kalkhoven *et al.* 2003	39
Bartsch *et al*. 2005	45
Udaka *et al*. 2005	21
Roelfsema *et al.* 2005	92
Bentivegna *et al*. 2006	31



Supplementary figure 1. Representation of the distribution of nucleotide changes in the *CREBBP* gene of RSTS patients. Numbers in the yellow block represent exon mutations and blue blocks represent intron mutations. Numbers in the red and green blocks represent the missense and the sense/silent mutations, respectively, in each of the patients.



Supplementary figure 2. Primary sequences corresponding to the minimal human p300 and CREBBP HAT domains are aligned with the same and conserved residues highlighted in black and white, respectively.