Epigenetic profile of the euchromatic region of human Y chromosome

Narendra Pratap Singh[1], Sri Ranganayaki Madabhushi[1], Surabhi Srivastava[1], R. Senthilkumar[1,2], C. Neeraja[3], Sanjeev Khosla[3,*] and Rakesh K. Mishra[1,*]

[1]Centre for Cellular and Molecular Biology, Council for Scientific and Industrial Research, Uppal Road, Hyderabad 500007, [2]Centre for Computational Natural Sciences and Bioinformatics, International Institute of Information Technology, Gachibowli, Hyderabad 500032 and [3]Centre for DNA Fingerprinting and Diagnostics, Nampally, Hyderabad 500001, India

ABSTRACT

The genome of a multi-cellular organism acquires various functional capabilities in different cell types by means of distinct chromatin modifications and packaging states. Acquired during early development, the cell type-specific epigenotype is maintained by cellular memory mechanisms that involve epigenetic modifications. Here we present the epigenetic status of the euchromatic region of the human Y chromosome that has mostly been ignored in earlier whole genome epigenetic mapping studies. Using ChIP-on-chip approach, we mapped H3K9ac, H3K9me3, H3K27me3 modifications and CTCF binding sites while DNA methylation analysis of selected CpG islands was done using bisulfite sequencing. The global pattern of histone modifications observed on the Y chromosome reflects the functional state and evolutionary history of the sequences that constitute it. The combination of histone and DNA modifications, along with CTCF association in some cases, reveals the transcriptional potential of all protein coding genes including the sex-determining gene *SRY* and the oncogene *TSPY*. We also observe preferential association of histone marks with different tandem repeats, suggesting their importance in genome organization and gene regulation. Our results present the first large scale epigenetic analysis of the human Y chromosome and link a number of cis-elements to epigenetic regulatory mechanisms, enabling an understanding of such mechanisms in Y chromosome linked disorders.

INTRODUCTION

The human genome generates about 220 cell types and thousands of functional states. This multiplicity of phenotypic output is possible by epigenetic regulatory mechanisms that involve DNA methylation and a rich variety of histone modifications (1,2). In general, histone acetylation correlates with active state of chromatin whereas DNA methylation correlates with inactive chromatin. Histone methylation leads to active or inactive chromatin depending on the site of methylation of the histone tail (3). Tri-methylation of H3K4, H3K36 and H3K79 mark active chromatin whereas tri-methylation of H3K9, H3K27 and H3K56 mark inactive chromatin. Combinations of histone modifications provide further diversity in marking specific epigenetic signatures of a cell and are a crucial determinant of its fate (4,5). Cytosine methylation of DNA is the most widely studied epigenetic modification and has been found to be associated with tissue, germ line and cancer specific genes (1,6,7). CpG islands, the target of DNA methylation, are part of ~60% of all promoters in the human genome (8). Histone hypoacetylation and hypermethylation at specific residues of core histone proteins are also correlated with DNA methylation, showing that there is an association between histone modifications and DNA methylation which finally determines the state of chromatin. Histone modifications, together with DNA methylation, have a vital role in organizing nuclear architecture and are involved in regulating transcription and other nuclear processes (1,9). These epigenetic marks not only reflect the cell type-specific functional state of the genome but also play an important role in maintaining that functional state during cell cycle and development (10–12).

The Y chromosome is one of the smaller chromosomes (~57 Mb) in the human genome (13,14) and has been

*To whom correspondence should be addressed. Tel: +91 40 2719 2658/+91 40 2716 0031; Fax: +91 40 2716 0951; Email: mishra@ccmb.res.in
Correspondence may also be addressed to Sanjeev Khosla. Tel: +91 40 2474 9410; Fax: +91 40 2474 9448; Email: sanjuk@www.cdfd.org.in

fascinating to researchers due to its peculiar genetics, role in sex determination and evolutionary history (15–17). Except for the small pseudoautosomal regions on both ends, no other part of the Y chromosome participates in meiotic recombination. The absence of recombination makes it very special not only for medical and forensic reasons but also to study human evolution (18). Non-recombining region of Y chromosome (NRY) or Male specific Y chromosome (MSY) has ~40% (~23 Mb) euchromatin and 60% (~40 Mb) heterochromatin. It has poor gene density and has often been considered as a collection of junk DNA except for the sex-determining region. This view is gradually changing and now, many phenotypes have been attributed to the Y chromosome, including spermatogenesis, cerebral asymmetry, stature, tooth size, handedness, testoblastoma and male specific cancers (16,19–24). The pseudoautosomal regions present at the short (p) and long (q) arm contain 16 and 3 coding genes, respectively, whereas the euchromatic region of MSY contains 156 transcriptional units, half of which code for proteins. Because of internal duplications, these 78 genes code for only 27 distinct proteins (13,16). Of the 27 genes, 12 are expressed ubiquitously, 11 are expressed exclusively or predominantly in testis and the remaining 4 are expressed only in teeth and brain (16). Repeat elements are widespread in the Y chromosome and it has been postulated that some of these repeat elements may have a role in genome packaging and coordinated expression of genes (25).

These observations imply that though the Y chromosome has very few genes, it employs a variety of regulatory mechanisms for its function. Because of these features, we chose the Y chromosome for epigenetic analysis, generally neglected in whole genome epigenetic mapping studies (1,2,26,27) presumably because of the presence of many repeats and the instability of the chromosome in various cultured cell lines. To overcome both these limitations, we decided to analyze chromatin obtained from white blood cells (WBCs) of individuals and perform ChIP-on-chip using Y chromosome specific probes from both unique and duplicated regions. In this study we have mapped H3K9ac, H3K9me3, H3K27me3 histone modifications and binding sites of CTCF, a conserved chromatin protein with multiple functions including a role in chromatin domain boundary (28–31), across the euchromatic region of human Y chromosome using ChIP-on-chip technique. DNA methylation analysis of some CpG islands associated with Y-chromosome genes was also done using bisulfite sequencing method. Our results provide for the first time an epigenetic view of the MSY euchromatin. The distinct patterns of histone marks associated with different types of sequences on the chromosome are also indicative of the epigenetic mechanisms that regulate most of the genes and tandem repeats.

MATERIALS AND METHODS

Sample collection and isolation of white blood cells

Blood samples of male individuals, one each from the two distinct populations, South Indian (Tamil Nadu) and North Indian (Jammu and Kashmir) populations were collected (32). The sample collection was done as per institutional ethical guidelines. The blood samples were diluted 4 times in ELB (erythrocyte lysis buffer: 155 mM NH_4Cl, 10 mM $KHCO_3$, 0.1 mM EDTA pH 7.4) and incubated on ice for 30 min. After centrifugation at 2500 rpm for 10 min at room temperature, the supernatant was carefully removed and the pellet was re-suspended thoroughly in ELB. Again, centrifugation at 2500 rpm for 10 min at room temperature was done to pellet down white blood cells (WBCs) and the pellet was finally resuspended in 1× PBS. Cells were cross-linked immediately by incubation in 1% formaldehyde for 10 min at room temperature. Cross-linking was stopped by the addition of 125 mM glycine. For future use, cells were snap frozen and stored at −80°C.

Chromatin immuno-precipitation

Chromatin immuno-precipitation (ChIP) was carried out as per the Upstate kit (catalog 17–295) protocol. In brief, about 10^8 cross-linked cells were resuspended in 2 ml SDS lysis buffer containing protease inhibitors. After 10 min of incubation on ice, 200 μl aliquots were sonicated using Bioruptor (Diagenode). These aliquots were pooled, mixed well and again split into 200 μl aliquots. Then ChIP was done using H3K9Me3 (Upstate), H3K9ac (Upstate), H3K27Me3 (Upstate) and CTCF (Abcam) antibodies on WBCs of two different individuals, 8 experiments in total. To rule out experimental errors, we performed two more ChIPs with the South Indian sample, using H3K9Me3 and H3K9ac antibody. So, ten ChIP experiments were done using four antibodies for two samples and keeping two duplicates with the South Indian sample. After ChIP, DNA recovered from the bound fraction was used for the microarray experiment.

Microarray

A custom made Human Y chromosome specific microarray chip manufactured by Agilent Technologies was used. About 95 049 non-overlapping probes varying in size from 44 to 60 nt from the euchromatic region were spotted on microarray chip. The array contained 50 220 unique probes, 34 501 probes for multi-copy loci spotted once on the microarray chip and 4735 probes for multi-copy loci spotted multiple times (total 10 328). About 44 829 probes designed for multi-copy loci covered sequences like pseudoautosomal region, tandem repeats, palindromes and multi-copy *TSPY* gene. ChIP DNA was used for the hybridization as per the Mammalian ChIP-on-chip Protocol of Agilent Technologies. Data was extracted from scanned microarray slides using Agilent Feature Extraction Software 9. Output files from the Feature Extraction Software were used as input files for Agilent ChIP Analytics software for data extraction and visualization. Microarray data was normalized using default settings of Intra-array Lowess (intensity-dependent) normalization, Whitehead Error Model v1.0 in the 'ChIP analytics' software (Supplementary Figure S1). For data analysis and visualization in graphical form, human genome build May 2004

was used as per designed microarray chip probes, in custom track of UCSC genome browser (http://genome .ucsc.edu).

Bisulfite sequencing

The sodium bisulfite modification method was used to analyze DNA methylation of selected CpG island as described previously (33). In brief, DNA was digested with the appropriate enzyme and then denatured before sodium bisulfite treatment. The DNA sample was mixed with two volumes of low melting point agarose and kept at 50°C. The bisulfite modification solution (1.9 g of sodium bisulfite dissolved in 2.5 ml of water and 750 μl of 2 M NaOH) and hydroquinone solution (0.055 g dissolved in 500 μl of water) were made separately at 50°C and then combined and filtered in the dark. Beads from 10 μl aliquots of each DNA-agarose mix were made in mineral water, which was then transferred to the bisulfite modification solution. Once the beads were in the lower bisulfite solution, the tubes were incubated on ice for 30 min and then at 50°C for 3.5 h. After the incubation, the solution was removed and beads were washed before using for PCR. Two rounds of PCR were used to amplify the targets, which were then cloned in a T-overhang vector for sequencing. To obtain the methylation profile, a few clones were analyzed for each sample. The efficiency of bisulfite conversion was analyzed by calculating the percentage conversion of cytosines in non-CpG context. In our study, the non-CpG 'C to T conversion rate' was 97.1 ± 0.62 (standard error of mean).

Quantitative PCR

The chromatin prepared for the ChIP-on-chip study was also used to perform subsequent ChIP-qPCR validation on a panel of 13 loci selected from the array data. ChIP was performed on chromatin from 25 000 cells using LowCell ChIP Kit (Diagenode, #kch-maglow-A16) following the manufacturer's protocol. Briefly, sonicated chromatin was incubated without any antibody (no antibody) or with 1 μg of H3K9me3 (abcam, #ab8898) or H3K27me3 (Millipore, #07-449) antibody or 2 μg of H3K9ac (abcam, #ab4441) or non-specific IgG (Diagenode, #kch-803-015) antibody bound to protein A coated paramagnetic beads. An aliquot of the same cross-linked and sonicated chromatin was kept aside as input DNA and purified in parallel. Following elution and purification of the immunoprecipitated DNA, relative abundance at target loci was estimated using Power SYBR Green qPCR Master mix (Applied Biosystems) on an ABI7900HT Fast Real-Time PCR System (2 min at 50°C; 10 min at 95°C; 40 cycles of 15 s at 94°C, 30 s at 60°C and 30 s at 68°C, followed by dissociation curve analysis). Enrichment in the ChIP DNA was determined relative to the no-antibody control using the dC_t method. Enrichment was determined from three independent sets of ChIP assays performed on chromatin from the South Indian sample. Statistical significance of the enrichment was calculated using the Wilcoxon paired *t*-test on the raw data (34). Primer sequences for all the loci tested are provided in Supplementary Table S1.

RESULTS

Coverage of epigenetic analysis and validation

We mapped H3K9me3, H3K9ac, H3K27me3 histone modifications and CTCF binding sites on the Y chromosome of two individuals from two distinct Indian populations—from North and South India—as biological duplicates. As a technical replicate we performed anti H3K9me3 and anti H3K9ac ChIP-on-chip analysis in duplicate with the South Indian sample. Considering the experimental and biological duplicates, we found that on an average, 10 330 of H3K9me3, 12 542 of H3K9ac, 9952 of H3K27me3 and 10 497 of CTCF probes were significantly enriched, at enrichment log ratio ≥1. The global profile of epigenetic marks was found to be very similar and differences were statistically insignificant between the two individuals representing the two different populations, although a large population size analysis is needed to look for individual specific variations (Supplementary Figure S2). Here we used a biological duplicate to find out if the global epigenetic profile to see the degree of similarity between the two distantly related individuals. Considering the overall similarity in experimental and biological duplicates, we have discussed further only the data obtained from the South Indian sample.

To further validate the ChIP-on-chip experiment we performed independent ChIP assays followed by quantitative real time PCR analysis for different histone antibodies using WBCs from the blood sample used for the ChIP-on-chip study. We selected a panel of 13 loci spanning 7 unique genes (*PLCXD, SRY, E1F1, DDX3Y, SMCY, UTY* and *PRKY*) and 3 duplicated genes (*TSPY, DAZ* and *CDY*) that showed distinct epigenetic patterns in the array data and validated the trend of enrichment with histone antibodies across these loci. To validate the sharply changing pattern in *PLCXD, TSPY* and *DAZ* loci (as discussed below), we selected two probes in each case. We observed striking similarity in the two data sets between the relative enrichments of the chromatin marks across the Y-chromosome genes (Supplementary Figure S3). More than 90% of the tested unique loci showed good correlation of ChIP-on-chip results with ChIP-qPCR. The duplicated loci also showed good correlation and, more interestingly, as in case of ChIP-on-chip results, ChIP-qPCR also showed low enrichment with these antibodies. These observations not only support our results but also suggest that duplicated loci can be studied for epigenetic modifications using this approach. ChIP-qPCR results confirmed that H3K9ac mark was most abundant around the promoters of E1F1, SMCY, PRKY and DDX3Y, which are expressed in blood cells while silenced SRY gene appeared to be mostly associated with the repressive marks of hetero-chromatinization, with high methylation at H3K9 as well as H3K27. These results effectively validate the observed ChIP-on-chip patterns of the histone marks and conclusions drawn about the epigenetic patterns across the Y-chromosome from the array data.

As DNA methylation and histone methylation are linked processes we evaluated the DNA methylation status at CpG residues of specific loci to elucidate its
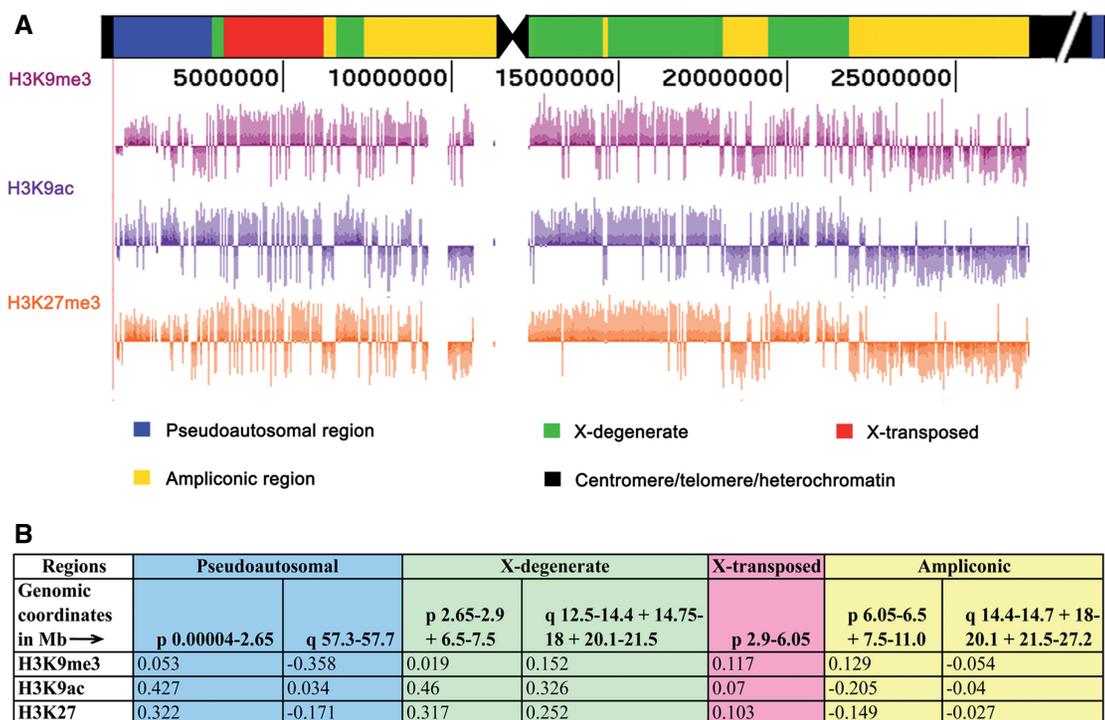
Regions	Pseudoautosomal		X-degenerate		X-transposed	Ampliconic	
Genomic coordinates in Mb ⟶	p 0.00004-2.65	q 57.3-57.7	p 2.65-2.9 + 6.5-7.5	q 12.5-14.4 + 14.75-18 + 20.1-21.5	p 2.9-6.05	p 6.05-6.5 + 7.5-11.0	q 14.4-14.7 + 18-20.1 + 21.5-27.2
H3K9me3	0.053	-0.358	0.019	0.152	0.117	0.129	-0.054
H3K9ac	0.427	0.034	0.46	0.326	0.07	-0.205	-0.04
H3K27	0.322	-0.171	0.317	0.252	0.103	-0.149	-0.027

Figure 1. Association of histone marks with different sequences of the Y chromosome. (**A**) Features of the Y chromosome are shown on the top, telomere, centromere and heterochromatin (black), and euchromatin. Different kinds of sequences in euchromatin are represented as colored blocks on the chromosome, blue (pseudoautosomal region), green (X-degenerate), red (X-transposed) and yellow (ampliconic). To visualize the data clearly, both enriched and not enriched regions have been shown for H3K9me3, H3K9ac and H3K27me3 histone modifications in dark red, blue and orange, respectively (the same convention has been followed in all the other figures also). We observe a specific global pattern of histone marks in different classes of sequences—the pseudoautosomal region is predominately enriched with H3K9ac, the X transposed is enriched with H3K9me3, the X-degenerate region shows almost equal enrichment with all marks and the ampliconic regions show a low enrichment with all histone marks, H3K9me3 being the most enriched. (**B**) Here we show the average enrichment of histone marks in different classes of sequences distributed along the chromosome. As visually observed in Figure 1A, different classes of sequences have a global pattern of histone marks. The pattern of histone marks in different class of sequences followed the visual observation except in the case of X-degenerate region. In visual analysis X-degenerate is enriched equally with H3K9me3 and H3K9ac but when we calculated average enrichment in different region we found that H3K9ac is more enriched than H3K9me3. In the given table, each class of sequences, from short (p) and long (q) arm are clubbed together to simplify the table.

role in gene regulation with histone modification. We performed DNA methylation analysis of 18 CpG islands present within the promoter of 9 protein-coding genes (*SRY*, *RPS4Y1*, *TGIF2LY*, *TSPY*, *TBL1Y*, *CYorf16*, *DDX3Y*, *TMSB4Y* and *VCY1B)* and two non-coding genes (*TTTY7* and *TTTY18*). These loci were selected on the basis of their biological importance and observed histone marks. Among these, 6 loci had H3K9ac coinciding with unmethylated CpG islands while 11 loci had H3K9me3 coinciding with methylated CpG islands (Supplementary Table S2). As expected all tested loci followed the conventional relation between H3K9 modification and DNA methylation and further strengthened our analysis.

Global pattern of histone marks on Y chromosome and gene expression

The euchromatin region of the human Y chromosome has been divided into distinct classes of sequences, viz., pseudoautosomal, X-transposed, X-degenerate and ampliconic, on the basis of their origin (13). Globally, we observed that distinct class of sequences, distributed in different regions of the chromosome, showed distinct histone marks whereas sequences of the common origin showed similar epigenetic pattern (Figure 1). To further confirm our visual observations, we calculated average enrichment value for the histone marks in distinct classes of sequences. As expected, different classes of sequences showed a pattern similar to that obtained by visual analysis (Figure 1B). The pseudoautosomal regions present on the ends of both the arms showed predominant enrichment of H3K9ac over H3K9me3 and H3K27me3. In contrast, the X-transposed sequences were more enriched with H3K9me3, while X-degenerate region followed a pattern similar to that of pseudoautosomal region. Most interestingly, ampliconic regions showed lower enrichment with all the three histone modification marks used in this study. Although all these histone marks showed lower enrichment in the ampliconic regions, H3K9me3 was still found to be the predominant mark. The pattern of histone marks shows such a clear pattern across the chromosome that different kind of sequences can be distinguished based on the epigenetic pattern (Figure 1A). These observations suggest that evolution of different sequences in the Y chromosome have

distinct epigenetic footprints. This also implies that on a global scale, during evolution, these sequences have not diverged enough to adapt different epigenetic patterns or functional states.

We also analyzed the association of histone marks and CTCF enrichment upstream to the transcription start site (up to ~1 kb upstream promoter sequence) of each gene in the different classes of sequences. The enrichment of histone marks and CTCF within 1 kb from transcriptional start site (TSS) was calculated using UCSC genome browser. Surprisingly, we observed that the genes present in a class of sequences showed a pattern of H3K9ac, H3K9me3 and H3K27me3 modifications similar to the global epigenetic pattern of that class. For example, genes present in the pseudoautosomal and X-degenerate regions were predominantly enriched with H3K9ac, genes in the X-transposed region were mostly enriched with H3K9me3 and genes in the ampliconic regions were found to be generally poor in all the examined histone marks at the TSS, with H3K9me3 showing predominant enrichment (Table 1). These observations suggest that there may be a global regulation operating to regulate one kind of sequences. Most surprisingly, a global regulation of similar kind of sequences is also borne out by the fact that genes present in same class of sequences follow same expression pattern.

To further understand epigenetic output in terms of gene expression, we compared our histone modification data with gene expression data of human blood cells from the UCSC genome browser (GNF Expression Atlas 2 Data from U133A and GNF1H) (35). Hyperacetylation of H3K9 was observed at the promoters of 44 genes, of which 28 genes are known to express in blood cells and for 2 genes expression pattern is unknown. The remaining 14 genes that do not express in blood cells indicate a different mechanism of repression that can tolerate the presence of H3K9ac. We also found that of the 47 genes associated with H3K9me3, 27 are not expressed in blood cells and for 18 genes expression data is not available. Only two genes (both the copies of *PRY*) are known to express in blood cells in spite of H3K9me3 enrichment. This analysis suggested that although at most of the genomic loci, H3K9me3 and H3K9ac regulate the loci as per conventional rule, there are a few exceptions. In our analysis, at many loci, the gene expression profile and H3K27me3 enrichment did not correlate well while H3K9me3 and H3K9ac showed better agreement with the expression status (Table 1). This observation suggests a mechanism whereby H3K9me3 and H3K9ac dominate over H3K27me3 modification to set the expression state of the chromatin, which can be further regulated by other sets of histone modifiers. Finally, there were instances when none of the tested modifications were enriched indicating the presence of other histone modifications and mechanisms to regulate the expression of those genes.

Tandem repeats and epigenetic modifications

The Y chromosome is full of repeats, which were initially thought to be junk, but now they are known to be important in maintaining the integrity of the Y chromosome in the absence of meiotic recombination (13). One such class of repeat is the tandem repeat which constitutes ~10% of human genome (36). These repeats are distributed in a non-random fashion and clustered at coding regions, near the centromere and telomeres (37). A change in the number of tandem repeats is not neutral, and can affect many basic cell processes like replication, transcription and translation because they recruit several factors which can change the status of the chromatin (38,39).

Accordingly, we also took care to design probes from tandem repeat regions unique to the Y chromosome with their counterparts on the X chromosome. We found that different tandem repeats associated with the Y chromosome are marked with a specific pattern of histone modifications. Most of the repeats showed a precise and clean pattern with the tested histone marks and CTCF binding. Such tandem repeats were distributed at both genic and intergenic regions. Figure 2A shows a locus of intergenic tandem repeats present near the telomeric region of the p arm. At this locus, tandem repeat-1 (each unit: 61 bp) was not enriched with any of the tested modifications, whereas another tandem repeat-2 (each unit: 26 bp) present just next to it was enriched preferentially with H3K9ac and CTCF. Epigenetic marks changed sharply with the change of the repeat, with different repeats showing their characteristic epigenetic marks and potential functional value. Interestingly, genic tandem repeats distributed near the TSS and introns of genes also showed precise and clean association with the tested histone marks and CTCF as observed for the intergenic tandem repeats. In the case of *P2RY8, PPP2R3B, CSF2R, SHOX, ASMTL, IL3RA* and *GTPBP6* genes, tandem repeats were clustered in the introns, while in the case of *PLCXD1* and *ASMTL* genes, they were clustered near the TSS. For example, many tandem repeats are clustered around TSS of *PLCXD1* a gene, which is expressed in blood cells. Each kind of tandem repeat has a pattern of histone modification and CTCF enrichment, which sharply changes with the change of repeat. As marked in Figure 2B, repeat-1 is highly enriched with H3K9ac and H3K27me3, whereas the upstream repeat-2 region is not enriched with any tested marks and repeat-3 is enriched only with H3K9ac and CTCF. The CpG island distributed around the TSS is enriched with H3K9ac, H3K27me3 and CTCF. Interestingly, none of the tandem repeats at this locus showed enrichment with H3K9me3. The observed pattern of H3K9ac, H3K27me3 and CTCF marks on different tandem repeats suggests a domain of active chromatin, which correlates well with the known expression of the *PLCXD1* gene in blood cells. Precise and specific association of tandem repeats with different histone modifications shows that the presence of such repeats can change the status of chromatin by interacting with different factors like CTCF that can have regulatory consequences. These repeats thus appear to act as regulatory elements to recruit factors for transcriptional control of the gene. As observed in the above mentioned cases, many tandem repeats were not enriched with tested histone modifications and CTCF, suggesting involvement of other histone modifications, histone variants and factors in the regulation of such repeats. Our observation of

Table 1. Analysis of histone marks and CTCF binding at promoter of the genes

S.No	Gene symbol	Known expression pattern in blood cells	H3K9me3 enrichment	H3K9ac enrichment	H3K27me3 enrichment	CTCF enrichment
1	PLCXD1	Y	0	●●● (green)	●● (orange)	● (blue)
2	PPP2R3B	Y	0	●●●●● (green)	● (orange)	● (blue)
3	SHOX	Y	0	● (green)	●●● (orange)	● (blue)
4	CSF2RA	Y	● (pink)	●●●● (green)	● (orange)	0
5	IL3RA	M	0	●●●● (green)	●● (orange)	●● (blue)
6	SLC25A6	Y	0	●●●● (green)	●● (orange)	● (blue)
7	CXYorf2	Y	0	● (green)	0	● (blue)
8	ASMTL	Y	0	●●● (green)	● (orange)	● (blue)
9	P2RY8	M	0	●●●●●● (green)	0	● (blue)
10	SFRS17A(Dxys)	Y	0	●● (green)	●● (orange)	0
11	ASMT	Y	0	●● (green)	0	● (blue)
12	ZBED1	Y	0	●● (green)	0	● (blue)
13	CD99	Y	0	●●●●● (green)	● (orange)	● (blue)
14	SRY	M	0	●●● (green)	●●● (orange)	● (blue)
15	RPS4Y1	Y	● (pink)	●●●●● (green)	● (orange)	● (blue)
16	ZFY	Y	0	●●●● (green)	● (orange)	●● (blue)
17	TGIF2LY	N	●●● (pink)	● (green)	● (orange)	●● (blue)
18	PCDH11Y	M	● (pink)	●● (green)	●●● (orange)	● (blue)
19	TTTY23	NA	●● (pink)	0	0	● (blue)
20	TSPY2	N	●● (pink)	0	0	● (blue)
21	TTTY1	NA	● (pink)	0	0	● (blue)
22	TTTY2	NA	● (pink)	0	0	● (blue)
23	TTTY21	NA	● (pink)	0	0	● (blue)
24	TTTY7	N	● (pink)	0	0	● (blue)
25	TTTY8	N	● (pink)	0	0	● (blue)
26	AMELY	N	● (pink)	●● (green)	● (orange)	●● (blue)
27	TBL1Y	N	●● (pink)	●●● (green)	●●●● (orange)	●● (blue)
28	PRKY	M	0	●●●● (green)	● (orange)	● (blue)
29	TTTY16	NA	● (pink)	0	● (orange)	● (blue)
30	TTTY12	NA	0	● (green)	● (orange)	● (blue)
31	TTTY18	NA	● (pink)	● (green)	● (orange)	● (blue)
32	TTTY19	NA	● (pink)	● (green)	● (orange)	● (blue)
33	TTTY11	N	● (pink)	●● (green)	● (orange)	0
34	TTTY20	NA	● (pink)	0	0	0
35	TSPY1	N	●● (pink)	0	0	0
36	AK093413	N	●● (pink)	0	0	0
37	RBMY3AP	NA	● (pink)	0	● (orange)	● (blue)
38	TTTY8	N	● (pink)	0	0	0
39	TTTY7	N	● (pink)	0	0	0

(continued)

Table 1. Continued

S.No	Gene symbol	Known expression pattern in blood cells	H3K9me3 enrichment	H3K9ac enrichment	H3K27me3 enrichment	CTCF enrichment
40	TTTY21	NA	●	0	0	0
41	TTTY2	N	●	0	0	●
42	TTTY1	M	●	0	0	●
43	TTTY22	NA	●●	0	●	●
44	TTTY23	NA	●●	0	0	0
45	TTTY15	Y	0	●●●●●●	●●	●
46	USP9Y	N	●●	●●●	●	●●
47	DDX3Y/DBY	Y	●	●●	●	●●
48	UTY	N	●	●●●●	●	●●●
49	TMSB4Y	Y	●	●●●	●●	●●
50	VCY1B	N	●●	0	0	0
51	VCY	N	●●	●	0	0
52	NLGN4Y	N	●	●●	●●●	●●
53	CDY2B	NA	●	0	0	0
54	CDY2A	NA	●	0	0	0
55	XKRY	NA	●	0	●	0
56	HSFY1	N	0	●	●	●
57	HSFY2	N	0	●	●●●	●●
58	CD24	M	0	●●●	●●	●
59	TTTY14	N	●	●●●●	●●	●●
60	BCORL2	NA	●●	●●●●●	●●	●
61	CYorf15A	M	0	●●●●	●	●●
62	SMCY	Y	0	●●●●	●	0
63	JARID1D	Y	●	●●●	●●	●
64	TTTY10	N	0	●	●	●
65	EIAF1AY	Y	0	●●●●●●	0	●
66	RPS4Y2	NA	●	●●	0	0
67	RBMY1A1	N	●●	0	●	●
68	RBMY1F1	N	●●	0	0	●
69	TTTY13	N	●	0	0	0
70	RBMY1D1	N	●●	0	●	●
71	RBMY1E	N	●●	0	●	●
72	TTTY6B	N	●	0	0	●
73	PRY	Y	●	0	0	0
74	RBMY1F1	N	●●	0	0	●
75	TTTY5	N	●	0	●	●
76	RBMY1J	N	●●	●	0	●
77	TTTY6B	N	●	0	●	●
78	PRY2	Y	●	0	0	●
79	TTTY17A	NA	●	0	0	0

(continued)

Table 1. Continued

S.No	Gene symbol	Known expression pattern in blood cells	H3K9me3 enrichment	H3K9ac enrichment	H3K27me3 enrichment	CTCF enrichment
80	DAZ1	N	●	●	●	●
81	DAZ2	N	●	●	●	●
82	CDY1	N	●	0	0	0
83	CSPG4LYP2	N	●	0	0	0
84	GOLGA2LY1	N	0	●	●	0
85	BPY2B	NA	0	0	0	0
86	DAZ3	N	●	●	●	●
87	DAZ4	N	●	●	●	0
88	TTTY17B	NA	●	0	0	0
89	BPY2C	NA	0	0	0	0
90	GOLGA2LY2	N	0	●	●	0
91	CSPG4LYP1	N	●	0	0	0
92	CDY1	N	●	0	0	0
93	TTTY3	NA	● ●	0	0	●
94	SPRY3	N	●	● ● ●	●	0
95	SYBL1	Y	●	● ● ● ●	●	●
96	IL9R	N	0	● ●	●	●
97	BC063682	Y	0	●	0	0

97 genes of the Y chromosome are listed as per their occurrence across the chromosome. Genes in rows of same color are present in the same kind of sequences in the chromosome (pseudoautosomal-blue, X-transposed-light red, X-degenerate-light green and ampliconic-yellow). Column 3 shows expression pattern of genes in whole blood and in different cell lineages of the blood. Expression status has been denoted as Y, Yes; N, No; M, Mix; NA, Not Available. Histone modifications and CTCF enrichment within 1.0 kb upstream of each gene has been listed in column 4 to 7. Number of circles in each row denotes the average enrichment of histone marks and CTCF as calculated by UCSC browser. Each colored circle shows an average enrichment value of 0.5 and 0 denotes average enrichment value less than 0.25. Globally we see that genes present in one class of sequence show similar histone modification pattern. Hyper acetylation at H3K9 was seen on promoters of genes present in pseudoautosomal and X-degenerate regions while hyper methylation at H3K9 was observed on promoters of genes present in X transposed and ampliconic regions

non-conflicting histone patterns also suggests that all the copies of the repeat carry the same histone marks. Our findings imply that tandem repeats at several places in the genome can have regulatory consequences. Enrichment of CTCF at the tandem repeats suggests a role for these sequences in chromatin-mediated regulation of genome organization and transcription.

Y chromosome linked oncogenesis and epigenetic regulation

In this study, we also analyzed a critical oncogenic region, the gonadoblastoma locus on the Y chromosome (GBY), which contains an array of the Testis-specific protein Y-encoded (*TSPY*) and Testis transcript Y (*TTTY*) genes. They have been linked to gonadoblastoma in intersex patients, testicular germ cell tumors, prostate cancer and other somatic cancers (20,24). *TSPY* shows homology with the *SET* oncogenes and interacts with Cyclin B protein, suggesting its role in the regulation of cell proliferation (20). Furthermore, high expression of *TSPY* in testicular and prostate cancer is implicated in tumorigenesis (40). *TSPY* is present in an array on the

short arm of the Y chromosome and its number varies from 20 to 40 (41,42). As these genes are present in an array and show a high degree of similarity in DNA sequence, we designed probes for nine copies of this gene. We designed a few unique probes and used other probes that could hybridize to all the copies of the duplicated region. Surprisingly, we found that the epigenetic marks at this region are set in a repeating pattern analogous to the repetitive unit of the TSPY genes. The repetitive pattern of H3K9me3 modification and CTCF binding at the *TSPY* gene cluster has a 20.7 kb epigenetic repeat unit (Figure 3). H3K9me3 is highly enriched around the start site of each *TSPY* gene and spreads over a 6.4 kb region. Enrichment of H3K9me3 starts about 4 kb upstream of the *TSPY* gene and peaks around the transcription start site, diminishing gradually towards the end of the gene. A small 1.6 kb region, present at the end of each 20.7 kb epigenetic repeat unit is also enriched with H3K9me3 and CTCF. Moreover, a ~7.2 kb region, upstream of the gene between the two H3K9me3 rich sequences, is devoid of any modifications tested, while a 5.5 kb region (marked in blue) downstream of the gene
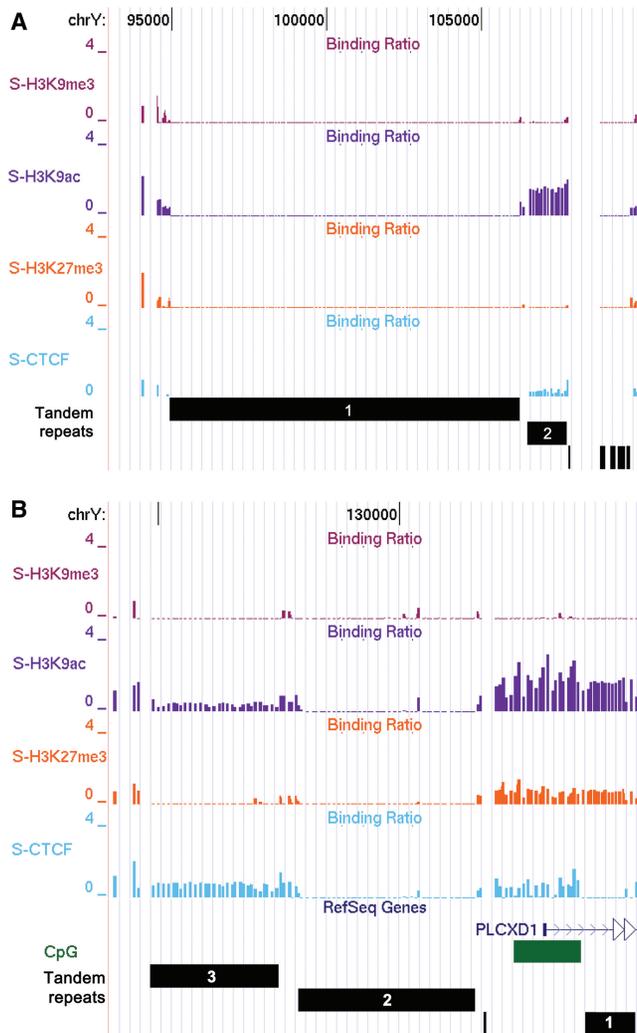
Figure 2. Profile of histone marks and CTCF with tandem repeats. Histone modification and CTCF binding profile are shown at different genomic coordinates with the tandem repeats. (**A**) Different kinds of tandem repeats present next to each other are shown in the Y chromosome. Two big tandem repeats are numbered as 1 and 2. Tandem repeat 1 does not show any enrichment with the tested antibodies whereas repeat 2 is enriched with H3K9ac and CTCF. There are few more tandem repeats on the right side of the tandem repeat 2, which does not show enrichment with any of the tested antibodies. (**B**) Multiple tandem repeats associated with the *PLCXD1* gene (blue line with arrow) have been shown with the histone modification and CTCF binding profile. Tandem repeats have been marked in the figure as 1, 2 and 3. Tandem repeat-3 enriches with anti-H3K9ac and anti-CTCF, tandem repeat-2 does not enrich with any of the tested antibodies whereas repeat-1 is enriched with anti-H3K9ac and anti-H3K27me3.

Figure 3. Epigenetic pattern and CTCF binding at *TSPY* gene arrays. Four genes of the *TSPY* array have been shown with their repetitive histone modification and CTCF distribution patterns. The profile of H3K9me3 was used to divide the whole region in the domains of 20.7 kb (written in pink) and H3K9me3 demarcates each domain. One domain has been shown below separately, divided in four parts using colored lines (black, red, green and blue). The start site of the gene is sharply enriched with H3K9me3, which shows highest peak and then further decrease towards both the sides (~6.4 kb). In the red-labeled (~7.2 kb) region none of the tested antibodies show any enrichment whereas in the blue labeled (~5.5 kb) region, there is little enrichment of H3K9me3 and CTCF. The black marked ~1.6 kb region upstream of the gene also shows H3K9me3 and CTCF enrichment.

has minimal H3K9me3 and CTCF enrichment. We also noticed overlapping CTCF binding and H3K9me3 enrichment, suggesting a link between the two, which might contribute to a locus-specific epigenetic regulatory combination. The specific and non-conflicting pattern of H3K9me3 and H3K9ac histone modifications observed suggests that all the copies of the gene are associated with a comparable epigenetic state. This implies that the regulatory sequences of *TSPY* genes have not diverged

after expansion and a common mechanism operates at this region. We also analyzed DNA methylation at two CpG islands upstream of the *TSPY* gene and found that they were hyper-methylated, correlating with the repressed state of *TSPY* in blood cells.

The oncogenic GBY locus also has an array of a non-protein-coding gene, *TTTY*. We found that the entire *TTTY* array showed high enrichment of H3K9me3 in comparison with other tested histone marks (Supplementary Figure S4). As expected, the promoters of all *TTTY* genes at this locus also showed predominant enrichment of the H3K9me3 mark (Table 1).
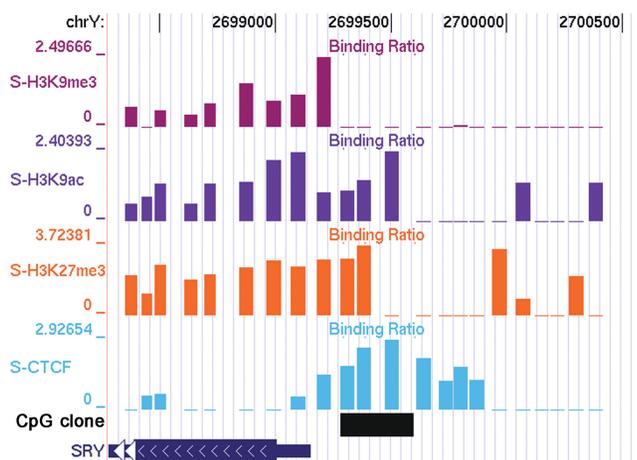
Figure 4. Epigenetic pattern of *SRY* gene. Co-ordinates of the *SRY* gene in the chromosome are given above in the figure while the gene body has been shown below. The genomic CpG island and the clone use to analyse the methylation status of the island are shown below the gene. We see a high enrichment of H3K9me3 at the start site and in the body of the gene, while H3K9ac and H3K27me3 are enriched from the body of the gene to upstream of the gene. A very clean enrichment of CTCF was seen at the CpG island upstream of the gene.

We analyzed the DNA methylation status in this region and found the CpGs present in the region to be methylated (Supplementary Table S2). These observations suggest that H3K9me3 and DNA methylation are necessary for repression of the *TSPY* and *TTTY* gene clusters at the GBY locus.

Epigenetic marks at sex determining gene *SRY*

SRY gene (Sex determining region, Y chromosome encoded), encodes a testis-determining factor essential for differentiation of Sertoli cells to induce masculinization of the embryonic gonad. Despite the key role played by *SRY* in genital development, its epigenetic profile has not been well studied. We found CTCF to be enriched in the region upstream of its TSS, with maximum binding within the CpG island (Figure 4). Interestingly, H3K9me3 was absent from the region of maximum CTCF binding, whereas H3K9ac and H3K27me3 were enriched. H3K9ac and H3K27me3 histone modifications taper down from the CTCF peak to the body of the gene, while enrichment of H3K9me3 starts from the transcription start site. As DNA methylation eliminates binding of CTCF to its recognition sites (43), it is interesting to note that bisulfite analysis found this CpG island to be unmethylated. High enrichment of H3K27me3 next to the CTCF binding supports the hypothesis that CTCF has the ability to recruit members of the PRC2 complex (44,45), and presents a potential mechanism for the repression of the *SRY* gene in blood cells. This interesting epigenetic pattern exemplifies a chromatin-mediated suppression mechanism based on recruitment of repressive factors by CTCF rather than by DNA methylation mediated gene silencing. Finally, we observed a conflicting pattern of H3K9ac and H3K9me3 marks within the gene body. This could be because *SRY* is expressed in the

Figure 5. Pattern of H3K9ac and H3K9me3 at *DAZ* gene loci. It shows *DAZ* gene, a duplicated gene with conflicting histone marks. The figure shows two copies of the DAZ gene, *DAZ3* and *DAZ4* present in opposite directions and the pattern of H3K9ac and H3K9me3 marks. Both alleles show almost equal enrichment of H3K9ac and H3K9me3 at their transcriptional start site.

B-cell lineage but not in all other blood cell lineages. These epigenetic marks reflected a novel mechanism that represses the transcription of this gene in some blood cell lineages.

Histone modifications at AZF locus

Almost all the ampliconic sequences on the long arm of the chromosome are collectively called the Azospermia factor (AZF). Deletion of the AZF has been observed in a number of the human male related infertility cases. This region has multiple palindromes, which go through intra-chromosomal recombination's, resulting in deletion of the genes required for male fertility (46,47). The AZF region contains *DDX3Y*, *DBY*, *XKRY, EIF1AY, RPS4Y2, RBMY*, *PRY*, *BPY, DAZ* and *CDY* genes. Our results showed a low enrichment of all tested histone modifications in the AZF region. *DDX3Y*, *EIF1AY* and *XKRY* genes were enriched with H3K9ac, *RBMY* and *CDY* genes were enriched with H3K9me3, *RPS4Y2* and *DAZ* genes showed a mixed pattern, whereas *BPY* did not show enrichment with any of the tested antibodies (Table 1). We found that most of the genes show a clean epigenetic pattern at different loci except the multi-copy *DAZ* (Figure 5) gene and the single copy *RPS4Y* gene. From this, we can reason that the mixed pattern at the single copy gene *RPS4Y* is because of the known differential expression in different blood cell lineages. Moreover, duplicated *DAZ* genes show enrichment with both H3K9me3 and H3K9ac, suggesting a diversion of epigenetic regulation at different loci after gene duplication, which may have led to changes in gene regulation (Figure 5). This indicates that after duplication, the regulatory elements of at least one of the copies have diverged to attain differential histone marks. The differential expression pattern of all the copies of *DAZ* has not been studied because of the high sequence similarity. In a preliminary study, however, polymorphic expression of *DAZ* gene copies was observed (48), supporting the idea that after duplication, regulatory elements of different copies of the *DAZ* gene have different histone modifications and therefore different regulation. As mutation in the *DAZ* gene region is correlated with azospermia, the function of this gene is non-redundant and essential; therefore, the cell has not

adopted to silence all copies of the gene. In contrast to the *TSPY* gene cluster (where all the duplicated copies have non-conflicting mark), only a few copies are preferentially silenced to balance the copy number in the case of duplicated *DAZ* gene.

DISCUSSION

Cell type-specific genome packaging defines the epigenome set during cellular differentiation. Epigenetic marks have been shown to specify the patterns of gene expression and a number of whole-genome epigenetic mapping studies of human cells have been done in the recent past to explore the functional states across genome (2,49,50). None of these, however, have explored the epigenetic signatures on the Y chromosome (1,2,26,27), presumably because of the presence of many repeats and the instability of the chromosome in various cultured cell lines. In this study, we have analyzed chromatin from WBCs and designed Y chromosome specific probes from both unique and duplicated regions to create an epigenetic map of the human Y chromosome. Unique probes were specific to unique regions of the Y chromosome, while probes for multi-copy loci were designed from sequences that are duplicated but are unique to Y chromosome and regions having homologues on X chromosome.

The evolutionary origin of the sequences that constitute the Y chromosome is known and it has been observed that genes distributed in sequences of the same origin have a similar expression pattern (13). We observed a global conservation of the epigenetic pattern associated with sequences of the same origin, which changed sharply at sequences of different origin. This implies that similar regulatory mechanisms might operate across genes that share a common origin and epigenetic profile, resulting in their similar expression pattern (35). This also insinuates that these sequences have not diverged enough to have different functions after they became part of the Y chromosome. Surprisingly, even regions of internal duplication within the Y chromosome showed non-conflicting epigenetic marks in our analysis, suggesting that after duplication, the associated loci have maintained similar regulatory environments. For instance, all copies of the *TSPY* and *TTTY* genes appear to be silenced in the WBCs by mechanisms that involve methylation of H3K9 around their promoters and methylation of DNA at associated CpG islands. This locus has been linked to oncogenesis and loss of methylation at both DNA and histone has been observed at the promoters of oncogenes in many cancers. It is possible that in the case of gonadoblastoma, the GBY locus may be hypomethylated, resulting in overexpression of *TSPY* and *TTTY* genes (51,52). Therefore, it would be interesting to test the epigenetic marks associated with this region in cancer tissues to explore the epigenetic regulation of *GBY* loci in different male-specific cancers.

Among the internally duplicated regions (palindromes) on the long arm, only *DAZ* genes at the AZF locus showed conflicting patterns of histone marks, suggesting that in these clusters, different copies of the genes may be regulated differently. It is possible that in the case of genes with non-redundant functions, mechanisms have evolved for silencing some of the copies to enable normal functioning of the cell, as an alternative to maintaining all the copies in a repressed or expressed state. The non-conflicting histone patterns observed for the rest of the duplicated genes suggest the ability of the cell to function normally with multiple copies of these genes being expressed after the duplication event. In the absence of meiotic recombination, palindromes may make the Y chromosome unstable, but our observation of very similar epigenetic marks on duplicated regions in contrast to expected results (53) indicates a regulatory mechanism at play that maintains the integrity of the Y chromosome in the absence of meiotic recombination and keeps global epigenetic regulation unchanged during evolution (16,46).

In this study, we have used profiles of histone modifications and CTCF binding to explore and correlate the transcriptional potential of most of the genes of the Y chromosome. The association of CTCF with different marks of histone and DNA methylation in the case of *SRY, TSPY* and tandem repeats suggests how epigenetic changes may lead to chromatin-mediated regulation of different coding and non-coding components of the genome. Interestingly, non-overlapping enrichment of H3K9me3 and DNA methylation at CpG islands of *SRY* and overlapping enrichment of the two at *TSPY* locus suggests that same epigenetic tools can be used differently at different loci. This also suggests the differential regulation of H3K9me3 modification and CTCF binding at each locus by other locus-specific factors. As observed in genic regions, we found distinct patterns of histone modifications and CTCF binding associated with different tandem repeats as well. Our data indicates that tandem repeats are capable of recruiting different kinds of chromatin modifiers, irrespective of their position, by virtue of their association with specific histone marks and this may play a role in setting the functional state of the chromatin. This suggests a non-neutral role for tandem repeats in recruiting various epigenetic regulatory factors. Since tandem repeats are known to be associated with almost every component of the genome, their epigenetic regulation can govern a large number of genomic loci by influencing many important cell processes like genome organization, replication and transcription. As CTCF is a transcriptional repressor and also binds with chromatin insulators (44), in these cases it may coordinate transcription of genes through higher order chromatin-mediated regulation.

This analysis also provides a way to study the epigenetic profile associated with multi-copy genes as a representation of the average pattern across all the copies. Analysis of ChIP data from probes representing repeated regions carries a risk of misinterpretation, however a cautious approach can yield new insights into the possible mechanisms of global epigenetic regulation. When duplicated loci carry the same epigenetic marks, the epigenetic pattern is observed to be non-conflicting with a clear association with specific marks, whereas in cases where the duplicated genes have diverged significantly and have distinct

epigenetic marks, conflicting histone modifications can be expected. This strategy is, therefore, useful in identifying the diversion in epigenetic regulation at duplicated loci. As our understanding of histone marks and their functional meaning increases, it is becoming more and more apparent that the functional status of a locus is decided by a complex histone code based on a combination of multiple histone modifications (3,7). Further mapping of other known histone marks, histone variants and chromatin modifiers would, therefore, be extremely useful in understanding various Y chromosome associated diseases.

SUPPLEMENTARY DATA

Supplementary Data are available at NAR Online.

ACKNOWLEDGEMENTS

We thank Prof. H. Sarat Chandra for initiating the discussion that lead to this analysis.

REFERENCES

1. Meissner,A., Mikkelsen,T.S., Gu,H., Wernig,M., Hanna,J., Sivachenko,A., Zhang,X., Bernstein,B.E., Nusbaum,C., Jaffe,D.B. *et al.* (2008) Genome-scale DNA methylation maps of pluripotent and differentiated cells. *Nature*, **454**, 766–770.
2. Barski,A., Cuddapah,S., Cui,K., Roh,T.Y., Schones,D.E., Wang,Z., Wei,G., Chepelev,I. and Zhao,K. (2007) High-resolution profiling of histone methylations in the human genome. *Cell*, **129**, 823–837.
3. Jenuwein,T. and Allis,C.D. (2001) Translating the histone code. *Science*, **293**, 1074–1080.
4. Ng,R.K. and Gurdon,J.B. (2008) Epigenetic inheritance of cell differentiation status. *Cell Cycle*, **7**, 1173–1177.
5. Turner,B.M. (2007) Defining an epigenetic code. *Nat. Cell Biol.*, **9**, 2–6.
6. Kondo,Y. and Issa,J.P. (2004) Epigenetic changes in colorectal cancer. *Cancer Metastasis Rev.*, **23**, 29–39.
7. Linggi,B.E., Brandt,S.J., Sun,Z.W. and Hiebert,S.W. (2005) Translating the histone code into leukemia. *J. Cell. Biochem.*, **96**, 938–950.
8. Cross,S.H., Clark,V.H., Simmen,M.W., Bickmore,W.A., Maroon,H., Langford,C.F., Carter,N.P. and Bird,A.P. (2000) CpG island libraries from human chromosomes 18 and 22: landmarks for novel genes. *Mamm. Genome*, **11**, 373–383.
9. Ioshikhes,I.P. and Zhang,M.Q. (2000) Large-scale human promoter mapping using CpG islands. *Nat. Genet.*, **26**, 61–63.
10. Francis,N.J., Follmer,N.E., Simon,M.D., Aghia,G. and Butler,J.D. (2009) Polycomb proteins remain bound to chromatin and DNA during DNA replication in vitro. *Cell*, **137**, 110–122.
11. Hansen,K.H., Bracken,A.P., Pasini,D., Dietrich,N., Gehani,S.S., Monrad,A., Rappsilber,J., Lerdrup,M. and Helin,K. (2008) A model for transmission of the H3K27me3 epigenetic mark. *Nat. Cell Biol.*, **10**, 1291–1300.
12. Aoto,T., Saitoh,N., Sakamoto,Y., Watanabe,S. and Nakao,M. (2008) Polycomb group protein-associated chromatin is reproduced in post-mitotic G1 phase and is required for S phase progression. *J. Biol. Chem.*, **283**, 18905–18915.
13. Skaletsky,H., Kuroda-Kawaguchi,T., Minx,P.J., Cordum,H.S., Hillier,L., Brown,L.G., Repping,S., Pyntikova,T., Ali,J., Bieri,T. *et al.* (2003) The male-specific region of the human Y chromosome is a mosaic of discrete sequence classes. *Nature*, **423**, 825–837.
14. Tilford,C.A., Kuroda-Kawaguchi,T., Skaletsky,H., Rozen,S., Brown,L.G., Rosenberg,M., McPherson,J.D., Wylie,K., Sekhon,M., Kucaba,T.A. *et al.* (2001) A physical map of the human Y chromosome. *Nature*, **409**, 943–945.
15. Garrigan,D. and Hammer,M.F. (2006) Reconstructing human origins in the genomic era. *Nat. Rev. Genet.*, **7**, 669–680.
16. Jobling,M.A. and Tyler-Smith,C. (2003) The human Y chromosome: an evolutionary marker comes of age. *Nat. Rev. Genet.*, **4**, 598–612.
17. Ellegren,H. and Parsch,J. (2007) The evolution of sex-biased genes and sex-biased gene expression. *Nat. Rev. Genet.*, **8**, 689–698.
18. Jobling,M.A. and Tyler-Smith,C. (2000) New uses for new haplotypes the human Y chromosome, disease and selection. *Trends Genet.*, **16**, 356–362.
19. Wilhelm,D. and Koopman,P. (2006) The makings of maleness: towards an integrated view of male sexual development. *Nat. Rev. Genet.*, **7**, 620–631.
20. Oram,S.W., Liu,X.X., Lee,T.L., Chan,W.Y. and Lau,Y.F. (2006) TSPY potentiates cell proliferation and tumorigenesis by promoting cell cycle progression in HeLa and NIH3T3 cells. *BMC Cancer*, **6**, 154.
21. Delbridge,M.L., Longepied,G., Depetris,D., Mattei,M.G., Disteche,C.M., Marshall Graves,J.A. and Mitchell,M.J. (2004) TSPY, the candidate gonadoblastoma gene on the human Y chromosome, has a widely expressed homologue on the X - implications for Y chromosome evolution. *Chromosome Res.*, **12**, 345–356.
22. Crow,T.J. (2002) Handedness, language lateralisation and anatomical asymmetry: relevance of protocadherin XY to hominid speciation and the aetiology of psychosis. Point of view. *Br. J. Psychiatry*, **181**, 295–297.
23. Lahn,B.T. and Page,D.C. (1997) Functional coherence of the human Y chromosome. *Science*, **278**, 675–680.
24. Lau,Y.F. (1999) Gonadoblastoma, testicular and prostate cancers, and the TSPY gene. *Am. J. Hum. Genet.*, **64**, 921–927.
25. Subramanian,S., Mishra,R.K. and Singh,L. (2003) Genome-wide analysis of Bkm sequences (GATA repeats): predominant association with sex chromosomes and potential role in higher order chromatin organization and function. *Bioinformatics*, **19**, 681–685.
26. Kim,T.H., Barrera,L.O., Zheng,M., Qu,C., Singer,M.A., Richmond,T.A., Wu,Y., Green,R.D. and Ren,B. (2005) A high-resolution map of active promoters in the human genome. *Nature*, **436**, 876–880.
27. Lister,R., Pelizzola,M., Dowen,R.H., Hawkins,R.D., Hon,G., Tonti-Filippini,J., Nery,J.R., Lee,L., Ye,Z., Ngo,Q.M. *et al.* (2009) Human DNA methylomes at base resolution show widespread epigenomic differences. *Nature*, **462**, 315–322.
28. Dunn,K.L. and Davie,J.R. (2003) The many roles of the transcriptional regulator CTCF. *Biochem. Cell Biol.*, **81**, 161–167.
29. Kim,T.H., Abdullaev,Z.K., Smith,A.D., Ching,K.A., Loukinov,D.I., Green,R.D., Zhang,M.Q., Lobanenkov,V.V. and Ren,B. (2007) Analysis of the vertebrate insulator protein CTCF-binding sites in the human genome. *Cell*, **128**, 1231–1245.
30. Lewis,A. and Murrell,A. (2004) Genomic imprinting: CTCF protects the boundaries. *Curr. Biol.*, **14**, R284–R286.
31. Holohan,E.E., Kwong,C., Adryan,B., Bartkuhn,M., Herold,M., Renkawitz,R., Russell,S. and White,R. (2007) CTCF genomic binding sites in Drosophila and the organisation of the bithorax complex. *PLoS Genet.*, **3**, e112.
32. Reich,D., Thangaraj,K., Patterson,N., Price,A.L. and Singh,L. (2009) Reconstructing Indian population history. *Nature*, **461**, 489–494.

33. Hajkova,P., el-Maarri,O., Engemann,S., Oswald,J., Olek,A. and Walter,J. (2002) DNA-methylation analysis by the bisulfite-assisted genomic sequencing method. *Methods Mol. Biol.*, **200**, 143–154.

34. Yuan,J.S., Reed,A., Chen,F. and Stewart,C.N.J. (2006) Statistical analysis of real-time PCR data. *BMC Bioinformatics*, **7**, 85.

35. Su,A.I., Wiltshire,T., Batalov,S., Lapp,H., Ching,K.A., Block,D., Zhang,J., Soden,R., Hayakawa,M., Kreiman,G. *et al.* (2004) A gene atlas of the mouse and human protein-encoding transcriptomes. *Proc. Natl Acad. Sci. USA*, **101**, 6062–6067.

36. Benson,G. (1999) Tandem repeats finder: a program to analyze DNA sequences. *Nucleic Acids Res.*, **27**, 573–580.

37. Verstrepen,K.J., Jansen,A., Lewitter,F. and Fink,G.R. (2005) Intragenic tandem repeats generate functional variability. *Nat. Genet.*, **37**, 986–990.

38. Pearson,C.E., Nichol Edamura,K. and Cleary,J.D. (2005) Repeat instability: mechanisms of dynamic mutations. *Nat. Rev. Genet.*, **6**, 729–742.

39. Usdin,K. (2008) The biological effects of simple tandem repeats: lessons from the repeat expansion diseases. *Genome Res.*, **18**, 1011–1019.

40. Su,M.T., Lee,I.W. and Kuo,P.L. (2006) Presence of TSPY transcript and absence of transcripts of other Y chromosomal genes in a case of microscopic gonadoblastoma. *Gynecol. Oncol.*, **103**, 357–360.

41. Vogel,T. and Schmidtke,J. (1998) Structure and function of TSPY, the Y-chromosome gene coding for the "testis-specific protein". *Cytogenet. Cell Genet.*, **80**, 209–213.

42. Krick,R., Jakubiczka,S. and Arnemann,J. (2003) Expression, alternative splicing and haplotype analysis of transcribed testis specific protein (TSPY) genes. *Gene*, **302**, 11–19.

43. Butcher,D.T., Mancini-DiNardo,D.N., Archer,T.K. and Rodenhiser,D.I. (2004) DNA binding sites for putative methylation boundaries in the unmethylated region of the BRCA1 promoter. *Int. J. Cancer*, **111**, 669–678.

44. Phillips,J.E. and Corces,V.G. (2009) CTCF: master weaver of the genome. *Cell*, **137**, 1194–1211.

45. Lutz,M., Burke,L.J., Barreto,G., Goeman,F., Greb,H., Arnold,R., Schultheiss,H., Brehm,A., Kouzarides,T., Lobanenkov,V. *et al.* (2000) Transcriptional repression by the insulator protein CTCF involves histone deacetylases. *Nucleic Acids Res.*, **28**, 1707–1713.

46. Lange,J., Skaletsky,H., van Daalen,S.K., Embry,S.L., Korver,C.M., Brown,L.G., Oates,R.D., Silber,S., Repping,S. and Page,D.C. (2009) Isodicentric Y chromosomes and sex disorders as byproducts of homologous recombination that maintains palindromes. *Cell*, **138**, 855–869.

47. Repping,S., Skaletsky,H., Lange,J., Silber,S., Van Der Veen,F., Oates,R.D., Page,D.C. and Rozen,S. (2002) Recombination between palindromes P5 and P1 on the human Y chromosome causes massive deletions and spermatogenic failure. *Am. J. Hum. Genet.*, **71**, 906–922.

48. Kim,B., Lee,Y., Kim,Y., Lee,K.H., Chun,S., Rhee,K., Seo,J.T., Kim,S.W. and Paick,J.S. (2009) Polymorphic expression of DAZ proteins in the human testis. *Hum. Reprod.*, **24**, 1507–1515.

49. Wang,Z., Zang,C., Rosenfeld,J.A., Schones,D.E., Barski,A., Cuddapah,S., Cui,K., Roh,T.Y., Peng,W., Zhang,M.Q. *et al.* (2008) Combinatorial patterns of histone acetylations and methylations in the human genome. *Nat. Genet.*, **40**, 897–903.

50. Hayashi,H., Nagae,G., Tsutsumi,S., Kaneshiro,K., Kozaki,T., Kaneda,A., Sugisaki,H. and Aburatani,H. (2007) High-resolution mapping of DNA methylation in human genome using oligonucleotide tiling array. *Hum. Genet.*, **120**, 701–711.

51. Irizarry,R.A., Ladd-Acosta,C., Wen,B., Wu,Z., Montano,C., Onyango,P., Cui,H., Gabo,K., Rongione,M., Webster,M. *et al.* (2009) The human colon cancer methylome shows similar hypo- and hypermethylation at conserved tissue-specific CpG island shores. *Nat. Genet.*, **41**, 178–186.

52. Jiang,L., Gonda,T.A., Gamble,M.V., Salas,M., Seshan,V., Tu,S., Twaddell,W.S., Hegyi,P., Lazar,G., Steele,I. *et al.* (2008) Global hypomethylation of genomic DNA in cancer-associated myofibroblasts. *Cancer Res.*, **68**, 9900–9908.

53. Graves,J.A. (2006) Sex chromosome specialization and degeneration in mammals. *Cell*, **124**, 901–914.