Journal of Clinical and Experimental Genetics
Volume 1 | Issue 1

Research Article Open Access

Low RNA Binding Strength of Human X Chromosome may contribute to X Chromosome Inactivation

Ji N[#1], Lv B[#2,3], Song Z[1], Wu P[1], Wang XD[1], Liu X[1], Shah S[1], Khan M[1], Wu C[1], Wang X[*1] and Lv Z[*1]

[1]Department of Genetics, Hebei Medical University, Hebei Key Lab of Laboratory Animal, Shijiazhuang 050017, Hebei Province, People's Republic of China

[2]Department of Ultrasound, Union Hospital, Tongji Medical College, Huazhong University of Science and Technology, Wuhan 430022, Hubei Province, People's Republic of China

[3]Hubei Province Key Laboratory of Molecular Imaging, Wuhan 430022, Hubei Province, People's Republic of China

[#]Ji N, and Lv B contributed equally to this work

[*]**Corresponding author:** Wang X and Lv Z, Department of Genetics, Hebei Medical University, Hebei Key Lab of Laboratory Animal, 361 Zhongshan East Road, Shijiazhuang 050017, Hebei Province, People's Republic of China, Tel: +8631186266844, +86-15076110502, E-mail: wangxf1966@hebmu.edu.cn, zhanjun_lv@hotmail.com

Citation: Ji N, Lv B, Song Z, Wu P, Wang XD, et al. (2021) Low RNA Binding Strength of Human X Chromosome may contribute to X Chromosome Inactivation. J Clin Exp Gen 1(1): 102

Abstract

During early embryonic development in female mammals, one copy of the X chromosome is randomly inactivated in a process known as X chromosome inactivation. In X chromosome inactivation, approximately 70% of genes on the short arm and nearly all of the genes on the long arm of the designated chromosome are inactivated. RNA activation is known to RNAs activating gene expression; however its roles in X chromosome inactivation have not been determined. Here, we used bioinformatic analyses to simulate the strength of binding between RNA and human 23 chromosomes (X chromosome and 22 autosomes) via a metric we refer to as RNA binding strength. We found that the RNA binding strength of the long arm of the X chromosome was significantly weaker than that of the autosomes and the short arm of X chromosome. Moreover, the RNA binding strengths of inactivated regions on the X chromosome were significantly lower than those of regions escaping X chromosome inactivation. The top 50% RNA binding strength value of the long arm of the X chromosome was significantly lower than those of all autosomes. Moreover, the RNA binding strengths of the chromosome 19 and X chromosome centromeres, which consist of constitutive heterochromatin, were weaker than those of the flanking sequences. These results suggest that RNA binding strength is associated with transcription. Because RNA binding to DNA can activate genes, the low RNA binding strength of the human X chromosome may be one explanation for X chromosome inactivation.

Keywords: X Chromosome Inactivation; Autosome; RNA Binding Strength; Centromere DNA; RNA Activation

List of abbreviations: chrX: X chromosome; XCI: X chromosome inactivation; *C. elegans*: *Caenorhabditis elegans*; nt: Nucleotide; Xq: Long arm of the chrX; Xp: Short arm of the chrX; SD: Standard deviation; chr19: Chromosome 19; ER: Escaping region (from XCI); IR: Inactivation region; ncRNA: Non-coding RNA; LINE-1: Long interspersed nuclear element-1

Introduction

During early embryonic development in female mammals, one of two copies of the X chromosome (chrX) is randomly inactivated in a process known as X chromosome inactivation (XCI). The pattern of XCI is retained through cell divisions and inherited by all daughter cells [1,2]. XCI represents a great model system with which to study a broad range of developmental and epigenetic processes-those involving stable gene expression without changes to the underlying DNA sequence [3-6]. Therefore, it is of great significance to study XCI for understanding epigenetics. In addition, dosage compensation of X-linked gene products between the sexes is crucial for mammalian growth and development [7-10]. In Drosophila and *Caenorhabditis elegans* (*C. elegans*), dosage compensation is accomplished by reducing the expression of genes located on the female chrX to half of that of those located on the male chrX [11-13]. Although previous studies have indicated that two X-linked genes (*Xist* and *Tsix*) [14,15] and differential DNA methylation [16] play important roles in XCI, a definitive mechanism of XCI remains elusive.

In human cells, interactions between RNA and DNA can activate gene expression [17-20] in a process known as RNA activation [21,22]. Our previous studies showed that RNAs complementary to DNA sequences play important roles in activating genes [23,24]. RNA activation may also regulate gene expression by binding to DNAs to prevent chromatin over-packing [25]. Previous studies have reported that the strength of RNA binding to DNA (the RNA binding strength) may influence chromatin activity. For example, the RNA binding strength of centromere DNA, which consists of constitutive heterochromatin, from C group chromosomes was significantly lower than that of flanking sequences, which suggests that the centromere is not easily affected by RNAs produced from other transcribed regions [22].

In this study, we explored whether RNA binding strength affects XCI. For this analysis, we used bioinformatics methods to examine RNA binding to the chrX and to the autosomes. Based on our findings, we conclude that the low RNA binding strength of human chrX is associated with XCI.

Materials and Methods

Sequence Data

Nucleotide sequences of the X chromosome and 22 autosomes were obtained from NCBI (build 33; http://www.ncbi.nlm.nih.gov/genome/guide). A total of 1,000 genes highly expressed in tonsil germinal center B cells were selected for analysis based on the results of the Digital Differential Display (NCBI UniGene Lib.289 -NCI_CGAP_GCB1). Germinal center B cells were selected for this analysis because there are many proliferating cells in tonsil germinal center B cells, and the set of expressed genes effectively reflects the interactions between RNAs and chromosomal DNA. Normally, there are about 1,000 highly expressed genes in differentiated cells [25,26], so we selected 1,000 highly expressed genes to calculate RNA binding strengths. These 1,000 highly expressed genes are located on different chromosomes, including chrX and reflect the set of RNAs in the B cells (excluding tRNAs and rRNAs). RNAs produced from these 1,000 genes were used to simulate the interactions between RNA and each chromosome.

Software

Gene-Analyser 2.0 software was used to analyze the number of 7-nucleotide (7-nt) strings. The software was written in-house and can be found at http://dx.doi.org/10.1016/j.biocel.2016.08.004 [22]. The number of 7-nt strings in each DNA fragment were calculated by using the method of stepping into a base. The possible combination of 7-nt strings for 4 bases is 4^7=16,384. Genomic DNA is packaged with nucleosomes, and the DNA strands are very long; therefore, DNAs will not form single strands, even with a negative superhelix. The binding between RNA and RNAs is stronger than that bewteen RNA and DNA, so RNAs preferentially form stem-loop structure or bind with RNAs. Therefore RNA binding to DNA occurs in short fragments of about 5 nt-9 nt in length. Because 7 nucletodies is the median length of RNA-DNA binding fragments, we chose 7-nt strings for our statistical analysis of RNA binding strength.

Although our data were designed to analyze 7-nt strings, it actually reflects strings of any length of and any complementary base number. If complementation exists in the long fragments, there will be an even greater number of complementary short fragments.

RNA binding strength algorithm

In this paper, to describe the possibility of all RNAs binding to DNA sequences of a certain length in cells, we developed a metric we call RNA binding strength. The stronger the RNA binding strength, the stronger the binding strength of all the RNAs in the cells to the DNA sequence, and vice versa [22]. The RNA binding strength algorithm is based on the principle that more complementarity between RNA and DNA results in more binding between RNAs and DNAs [22]. For example, when there is one 5'-TTTTTTT DNA molecule and ten 5'-AAAAAAA RNA molecules in a certain volume solution, the likelihood of DNA binding with RNAs is 10 (10×1=10). If there are ten 5'-TTTTTTT DNA molecules, the likelihood of DNA binding with RNAs is 100 (10×10=100). The binding of single-strand RNA and double-strand DNA accounts for competition between RNA and DNA for binding. Although G-C binding is strong, G-C melting is weak, so the affinity of A-T binding is not lower than that of G-C. Therefore, we did not introduce a coefficient to correct for G-C binding.

In our study, DNA sequences from the indicated chromosomes were divided into 50-kb segments and recoded as the number of 7-nt strings using all possible 7-nt strings (4^7=16,384). The 50 kb fragment size was chosen because a transcription unit contains 10-50kb sequences (based on DNase I digestion of the hemoglobin and ovoalbumin genes). To determine the RNA binding strength of a given 7-nt string, we first multiplied the number of times each 7-nt string appears in RNA and DNA fragments. One-thousand genes highly expressed in human tonsil germinal center B cells were selected (as described above), and the 7-nt string numbers for these genes were calculated from the sense strand (including introns and exons). The 7-nt string numbers for each gene multiply by the expression frequency of the gene (Lib.5601; http://www.ncbi.nlm.nih.gov/UniGene/), which results in the calculated numbers of the 7-nt string for the gene. Table 1 illustrates the calculation of RNA binding strengths in a 50-kb DNA fragment. The sum of the numbers for the 7-nt strings of 1,000 genes was regarded as the simulated total RNAs (total RNAs) in cells (Table 1, C column). The simulated RNA binding strength (Table 1, E column and E16385) was defined as the sum of the products of the RNAs and the corresponding 7-nt strings numbers within each 50-kb fragment. Therefore, RNA binding strength represents a measure of the amount of RNA that can bind to each 50-kb region.

A Number	B 7nt strings*	C Total RNAs**	D The amount of 7-nt string in the 50 kb DNA fragment #	E RNA binding strengths ##
1	AAAAAAA	2191.14455	78	170909.2749
2	AAAAAAC	197.00345	7	1379.02415
3	AAAAAAG	318.66205	12	3823.9446
.	.	.	.	.
.	.	.	.	.
.	.	.	.	.
16384	TTTTTTT	3324.5065	203	674874.8195
16385				6230540.035###

*: All possible 7-nt strings (the number of possible 7-nt strings for 4 bases is 4^7=16,384) (B1: B16384). **: The number of 7-nt strings for 1,000 highly expressed genes was calculated. Because the expression frequency of each gene is different, the number of 7-nt strings of each gene must be multiplied by the expression frequency of the gene. The sum of each 7-nt string of the one-thousand highly expressed genes is the number of each 7-nt string; C1:C16384 are the total RNAs. #: The number of each 7-nt string in a 50kb DNA fragment (D1:D16384). ##: The product of the number of each 7-nt string in the total RNAs and DNA fragment is the RNA binding strength of this 7-nt string. For example, E1=C1×D1, E2= C2×D2, etc. ###: E16385 is the sum of RNA binding strengths of all 7-nt strings (ΣE1:E16384), i.e. the RNA binding strength of the 50-kb DNA fragment.

Table 1: The calculation method of RNA binding strengths in a 50 kb DNA fragment

The RNA binding strength value of each 50-kb DNA fragment is regarded as a sample. T-tests were calculated using MS Excel and used to compare the mean of RNA binding strength values of all 50 kb fragments across different chromosomes. RNA binding strengths of different 50-kb fragments on the autosomes and the Xq were ranked in descending order, and the fragments having the top 50% of RNA binding strength values on Xq were compared to those having the top 50% RNA binding strength values on autosomes using a rank sum test. Both the standard deviation and variance reflect the degree of dispersion of data. To determine whether there was a significant difference in the degree of dispersion of RNA binding strengths between Xq and chromosomes 11, 20, and 22, we performed homogeneity of variance test using SPSS software.

Results

The comparison of RNA binding strengths of X chromosome (chrX) and autosomes

We first analyzed the RNA binding strengths of human autosomes 1-22 and compared them to those of the X chromosome. For this purpose, the chromosomal DNA was computationally divided into 50-kb fragments that were recorded as a set of 7-nucleotide (nt) DNA strings. A separate set of 7-nt strings were calculated from the RNAs transcribed from tonsil germinal center B cells. The number of DNA 7-nt strings was multiplied by the number of RNA strings. The sum of all such products represents the total strength of binding to the 50-kb DNA sequence (see Materials and Metho**ds)**. This analysis revealed that the mean binding strength of RNAs to the chrX is significantly lower than to most autosomes, with only chromosomes 8, 11, 18, 20, 21, and 22 having lower binding strengths than chrX (Table 2, Figure 1).

Chr	Number of 50-kb fragments	Mean ± S.D. of RNA binding strengths	*P* values *vs.* X using t-test*	*P* values *vs.* Xq using t-test**
1	4438	6333574.22± 691245.81	<0.01	<0.01
2	4758	6314539.81± 565702.45	<0.01	<0.01
3	3884	6337212.06± 552823.21	<0.01	<0.01
4	3747	6370787.96± 476562.48	<0.01	<0.01
5	3556	6320322.79 ± 560634.19	<0.01	<0.01
6	3346	6351251.86 ± 512578.84	<0.01	<0.01
7	3102	6422567.89 ± 635270.25	<0.01	<0.01
8	2845	6273460.86 ± 550807.81		<0.01
9	2334	6306166.84 ± 605680.41	<0.05	<0.01
10	2633	6304370.16 ± 672467.68	<0.05	<0.01
11	2620	6189916.53 ± 624444.70		
12	2595	6415714.81 ± 666539.13	<0.01	<0.01
13	1915	6373274.15 ± 478290.45	<0.01	<0.01
14	1745	6364194.37 ± 662510.32	<0.01	<0.01
15	1632	6346786.11 ± 665506.10	<0.01	<0.01
16	1602	6367224.26 ± 740683.39	<0.01	<0.01
17	1559	6524509.02 ± 831046.94	<0.01	<0.01
18	1495	6281989.52 ± 518609.70		<0.01
19	1117	6721733.43 ± 807647.61	<0.01	<0.01
20	1192	6193198.50 ± 753727.48		
21	682	6279488.50 ± 578744.81		<0.01
22	695	6242274.77 ± 831045.21		
X	2994	6267628.33 ± 526047.84		
Xp	1142	6345337.19 ± 562236.71		<0.01
Xq	1852	6219616.19 ± 496520.70		

*: Mean ± S.D. of RNA binding strengths of autosomes compared to those of the chrX using *t*-tests. **: Mean ± S.D. of RNA binding strengths of autosomes compared to that of Xq using *t*-tests. Xp: the short arm of chrX. Xq: the long arm of chrX.

Table 2: Comparison of RNA binding strengths of the autosomes and chrX

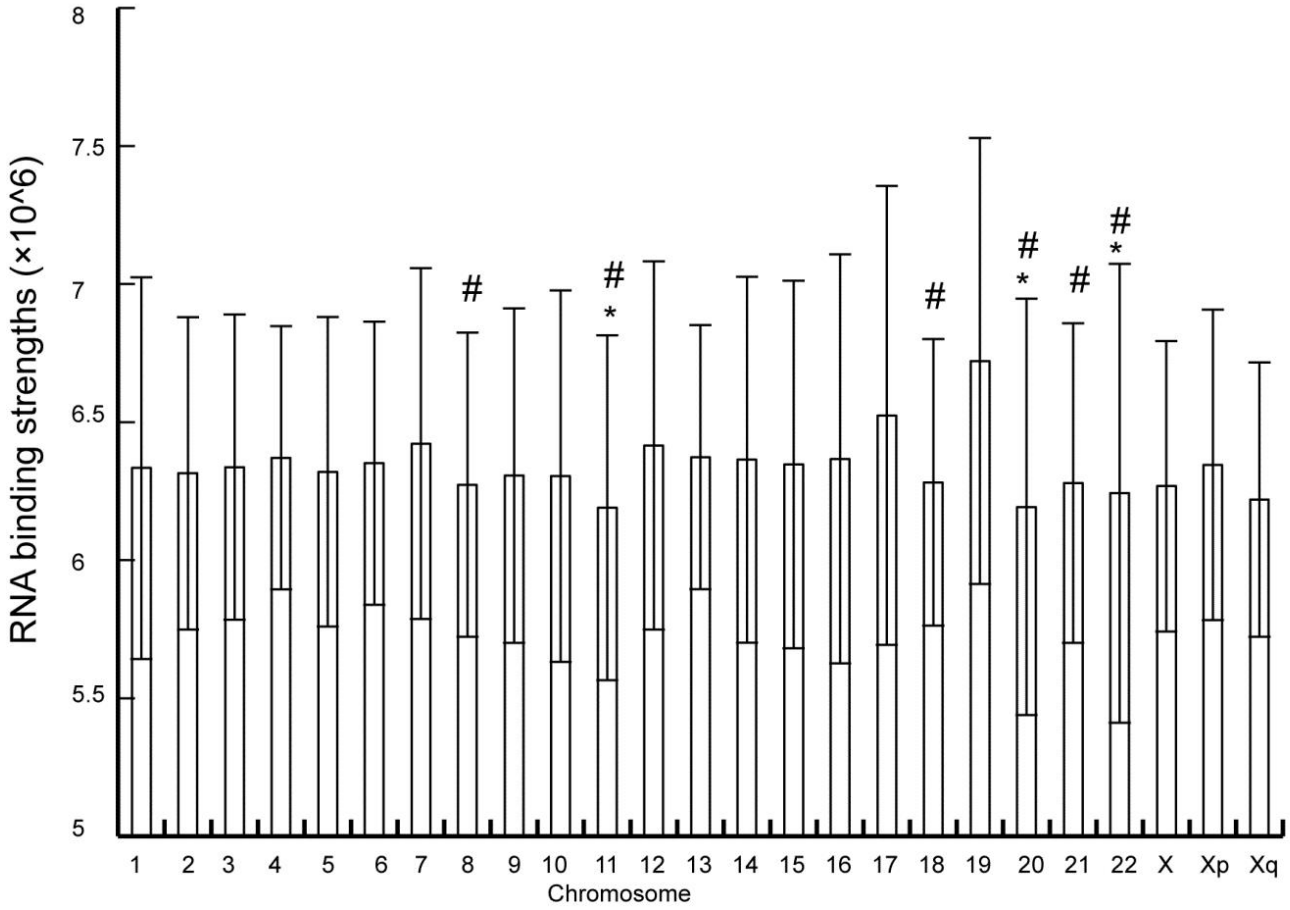

Figure 1: Comparison of mean RNA binding strengths of autosomes and chrX: The y-axis displays the mean RNA binding strength of each chromosome. The RNA binding strengths of autosomes without # symbol were significantly higher than those of X chromosomes. The RNA binding strength of autosome without * symbol were significantly higher than that of Xq

Comparison of RNA binding strengths of Xq and autosomes

Because the long arm of the chrX (Xq) is almost entirely inactivated, we further compared the RNA binding strength of Xq to that of the autosomes. The RNA binding strength of Xq was significantly lower than that of the autosomes, except chromosomes 11, 20, and 22 (Figure 1, Table 2).

We then examined whether there were significant differences between the fragments with high RNA binding strengths between Xq and chromosomes 11, 20, and 22. The fragments with high RNA binding strengths are indicated by two factors: the standard deviation (SD) of RNA binding strength and an upper 50% value of RNA binding strengths.

We found that the SD of RNA binding strength in Xq was lower than that for chromosomes 11, 20, and 22 (Table 2). The SD reflects the degree of dispersion of data. To observe whether there was a significant difference in the degree of dispersion of RNA binding strengths between Xq and chromosomes 11, 20 and 22, we performed homogeneity of variance tests (see Materials and Methods). The variance of RNA binding strengths of Xq was significantly lower than that of chromosomes 11, 20 and 22 ($p<0.01$). These results suggest that there is greater uniformity in RNA binding strengths of different 50 kb DNA fragments on Xq than in those on chromosomes 11, 20 and 22. Although the mean RNA binding strength value of chromosomes 11, 20, and 22 was not significantly higher than that of Xq, the variance of chromosomes 11, 20, and 22 was significantly greater than that of Xq, which suggests that some fragments of chromosomes 11, 20, and 22 have higher RNA binding strengths, and some fragments have lower RNA binding strengths. We therefore compared the top 50% RNA binding strength values on Xq to those of the autosomes. Because the distribution of the top 50% RNA binding strength values does not follow a quasi-Gaussian distribution, we used SPSS software

to do rank sum tests, which are suitable for the two independent samples test of non-normal distribution data. The top 50% RNA binding strength values of Xq were significantly lower than those of all autosomes ($p<0.01$, Table 3, Figure 2) and suggest that Xq lacks DNA fragments possessing high RNA binding strengths. These results indicate that lower mean RNA binding strengths and lower top 50% RNA binding strengths are two characteristics of Xq. The adjacent sequences with low RNA binding strengths cannot lead to complete inactivation of the fragments having very high RNA binding strengths.

Chr	Number of fragments (each fragment is 50 kb)	Mean ± S.D. of RNA binding strengths	*P* values *vs.* Xq using rank sum test*
1	2219	6852557.91±523160.54	<0.01
2	2379	6731999.18±429290.51	<0.01
3	1942	6736464.99±438393.58	<0.01
4	1874	6703132.77±356352.10	<0.01
5	1778	6727999.61±442933.87	<0.01
6	1673	6726310.07±390574.13	<0.01
7	1551	6882650.85±494280.01	<0.01
8	1423	6665717.93±396150.80	<0.01
9	1167	6751692.66±448601.14	<0.01
10	1317	6811085.63±482917.99	<0.01
11	1310	6661089.76±453199.60	<0.01
12	1298	6903220.43±533706.22	<0.01
13	958	6718085.01±333040.20	<0.01
14	873	6851483.23±514543.32	<0.01
15	816	6849173.39±526660.95	<0.01
16	801	6925111.51±538637.01	<0.01
17	780	7374808.36±569105.61	<0.01
18	748	6667064.16±375926.80	<0.01
19	559	7389657.41±459417.20	<0.01
20	596	6791394.78±610571.35	<0.01
21	341	6714836.54±346958.46	<0.01
22	348	6889563.32±612669.24	<0.01
Xp	571	6745827.53±499507.51	<0.01
Xq	926	6578083.40±399766.62	

Chr: chromosome. *: RNA binding strengths of different 50-kb fragments on the autosomes, Xp and Xq were placed in descending order, and the fragments with the upper 50% values of RNA binding strengths on Xq were compared with those of the autosomes and Xp using rank sum tests. Xp: the short arm of the chrX. Xq: the long arm of the chrX.

Table 3: Upper 50% RNA binding strength values of the autosomes, Xp, and Xq

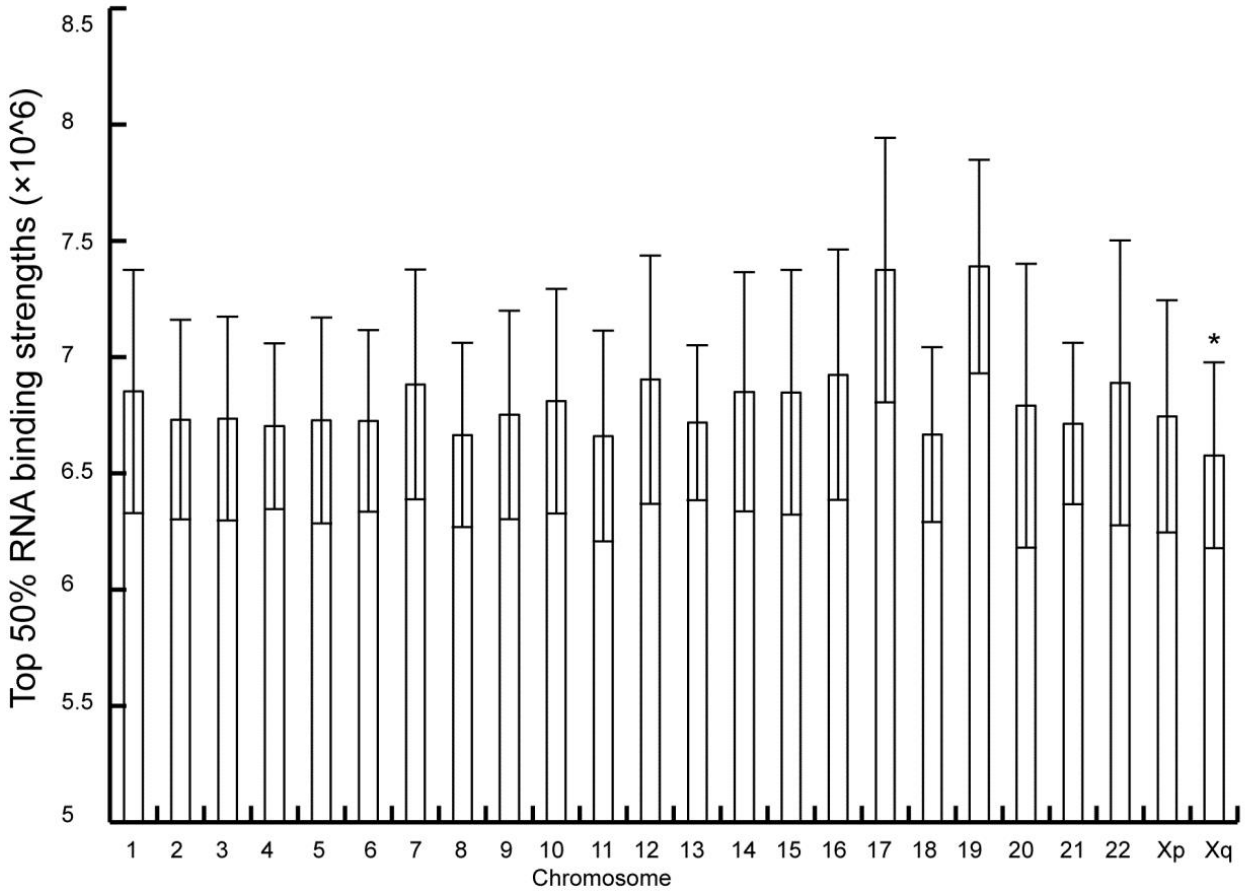

Figure 2: Comparison of top 50% RNA binding strengths of Xq with autosomes and Xp: A * symbol indicates that the top 50% RNA binding strengths of Xq is significantly lower than that of autosomes and Xp

Based on our analyses, chromosome 19 (chr19) has the highest mean RNA binding strength, and chr11 has the lowest mean RNA binding strength. So we used Figure 3A and Figure 3B to show the RNA binding strengths of the two chromosomes. Both the mean of RNA binding strength of chr19 (6721733.43 ± 807647.61 vs. 6219616.19 ± 496520.70, Table 2) and the upper 50% RNA binding strength values of chr19 (7389657.41 ± 459417.20 vs. 6578083.40 ± 399766.62, Table 3) were significantly higher than those of Xq (Figure 3A). Although the mean RNA binding strength value of chr11 was lower than that of Xq, its top 50% RNA binding strength values (6661089.76±453199.60 vs. 6578083.40±399766.62, Table 3) are significantly higher than those of Xq (Figure 3B).

The RNA binding strength of Xq was significantly lower than that of Xp

Approximately 30% of the genes located on the Xp escape XCI, whereas Xq is almost entirely inactivated. Carrel et al [27] reported that 29 of 224 X-linked genes escape XCI. Two of these genes escaping XCI are located on Xq (Figure 3C); the other 27 genes are located on Xp (Figure 3D). Our evaluation using bioinformatic methods revealed that the RNA binding strength of Xq was significantly lower than that of Xp ($p<0.01$) (Table 2). The upper 50% RNA binding strength values of Xq were significantly lower than those of Xp ($p<0.01$, Table 3).

From Figure 3C, we can see that the RNA binding strength of Xq at different genome positions was relatively uniform and Xq possess fewer fragments with high RNA binding strengths. These analyses proved that only Xq simultaneously has the both characteristics of low RNA binding strengths and low SD (Table 2). The RNA binding strength of Xp varied greatly at different genome positions, and Xp possess more fragments with high RNA binding strengths (Figure 3D).

RNA binding strengths of inactivation regions (IRs) are significantly lower than those of the escaping regions from XCI (ERs)

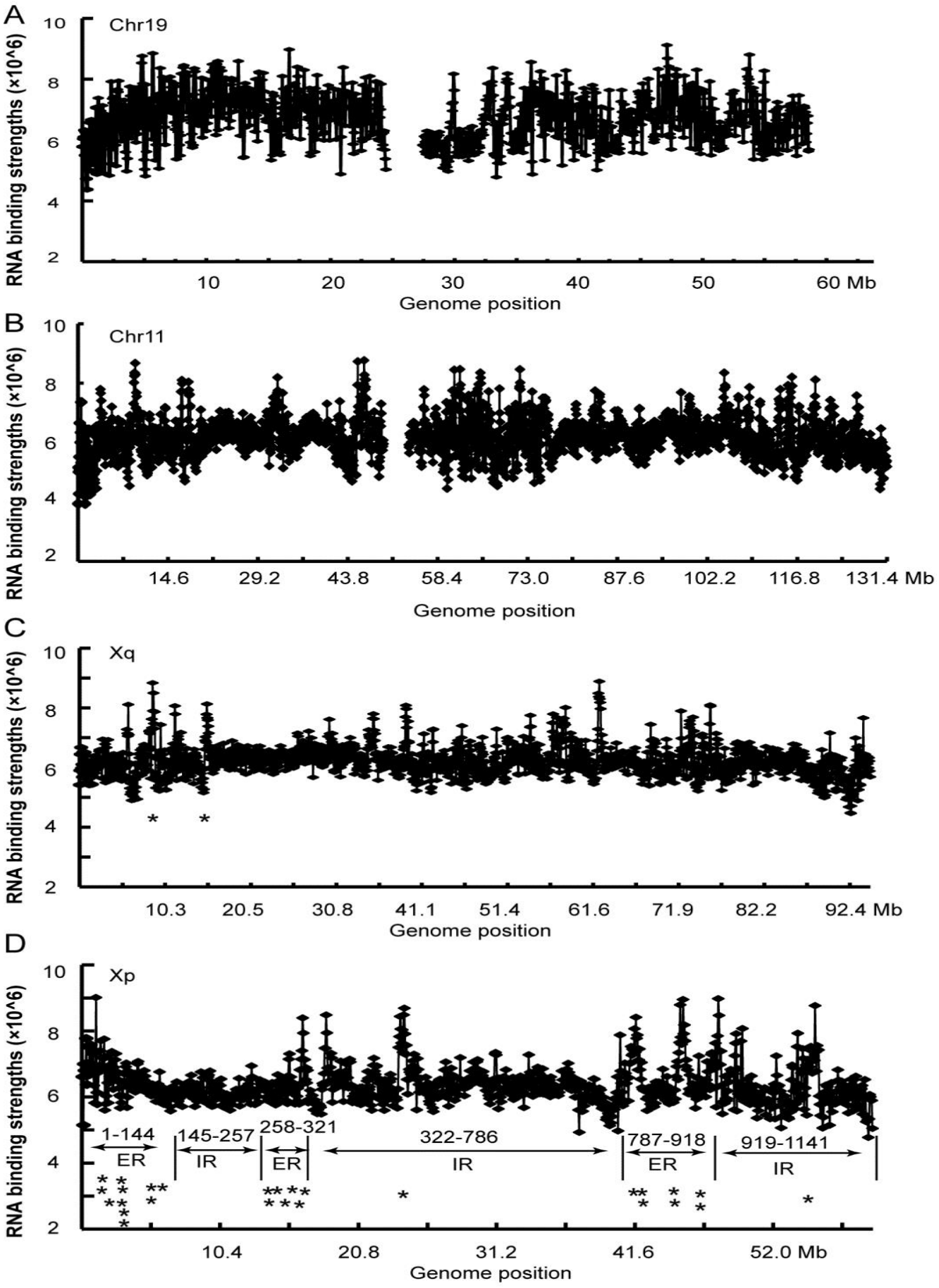
A
Chr19
RNA binding strengths (×10^6)
10
8
6
4
2
10
20
30
40
50
60 Mb
Genome position
B
Chr11
RNA binding strengths (×10^6)
14.6
29.2
43.8
58.4
73.0
87.6
102.2
116.8
131.4 Mb
Genome position
C
Xq
RNA binding strengths (×10^6)
10.3
20.5
30.8
41.1
51.4
61.6
71.9
82.2
92.4 Mb
Genome position
D
Xp
RNA binding strengths (×10^6)
1-144
ER
145-257
IR
258-321
ER
322-786
IR
787-918
ER
919-1141
IR
10.4
20.8
31.2
41.6
52.0 Mb
Genome position

Figure 3: RNA binding strengths of chr19, chr11, Xq, and Xp: (**A**) RNA binding strengths of chr19. Chr19 is the chromosome with the highest mean RNA binding strength value and has many fragments with relatively high RNA binding strength. (**B**) RNA binding strengths of chr11. Chr11 is the chromosome with the lowest mean RNA binding strength value and has some fragments with relatively high RNA binding strength. (**C**) RNA binding strengths of Xq. Each * symbol represents a gene that escapes from XCI on Xq: RPS4X and WI-12. (**D**) RNA binding strengths of Xp. Each * symbol represents a gene that escapes from XCI on Xp: SLC25A6(ANT3), DXYS155E, ALTE(TRAMP), stSG15779, MIC2, StSG9723, StSG1369, ARSD, GS1(Hs.78991), Hs.79876, GS2, SEDT, CXORF5, INE2, PIR, GRPR, StSG4551, RbAp46, eIF-2 gamma, CRSP150, DFFRX, DDX3, INE1, UTX, UBE1, PCTK1, SMCX. Region 1-144 contains 50-kb fragments numbered 1-144 on Xp; regions 145-257, 258-321, 322-786, 787-918 and 919-1141 are named for the number of 50kb fragments they contain. ER: escaping region (from XCI); IR: inactivation region. For the sake of brevity, we classify region 322-786 as an IR, in fact the position shown with * is the ER region. The x-axis represents the genome position (Mb). The genomic position of Xq is calculated from the origin of Xq. The genome positions of other chromosomes are the chromosomal sequence map positions

Because approximately 30% of the genes located on Xp escape XCI, w**e** compared the RNA binding strengths of IRs on Xp with those of ERs. The RNA binding strengths of the IRs were significantly lower than those of the ERs (Table 4). Xq has one region where the genes escape from XCI (Figure 3C). In contrast, Xp has three ERs, and the ER region possess the characteristic of higher RNA binding strengths and the RNA binding strength of IR region is lower (Figure 3D).

	Genome position		Total fragment number	RNA binding strengths (Means ± SD)
	Mb	The number of 50 kb fragment		
IRs	7.20-12.85*	145-257**	801***	6282853.28± 537207.57 ****, #
	16.05-39.30	322-786		
	45.90-57.20	919-1144		
ERs	1-7.20	1-144	332	6497461.20±598838.59
	12.85-16.05	258-321		
	39.30-45.90	787-918		

IRs: inactivation regions; ERs: escaping regions (from XCI). *7.20-12.85 indicates that this IR is from genome position 7.20 Mb to 12.85 Mb. ** 145-257 indicates that this IR is from the 145th 50 kb DNA fragment to the 257th 50kb DNA fragment. ***801 indicates that the three IR regions (7.20-12.85, 16.05-39.30 and 45.90-57.20) include 801 50 kb DNA fragments. **** 6282853.28 ± 537207.57 indicates the mean value and SD of RNA binding strengths of the three IRs. A # symbol indicates that the mean RNA binding strength of the three IRs is significantly lower than that of the three ERs ($p<0.01$)

Table 4: RNA binding strengths of IRs and ERs on Xp

RNA binding strengths of centromeres is significantly lower than that of their flanking sequences

The bioinformatics analyses above suggested that XCI is associated with low RNA binding strength of chrX. To further investigate this hypothesis, we analyzed the RNA binding strength of the centromere regions and their flanking sequences, because the centromere regions are structural heterochromatin that is almost non-transcribed. The lengths of the centromere sequences with low RNA binding strengths were 2.69 Mb in chr19 and 3.81 Mb in chrX (Table 5). Figure 4 and Table 5 show that the RNA binding strengths of the chr19 and chrX centromere sequences were significantly lower than those of their flanking sequences. These results suggest that non-transcribed sequences (structural heterochromatin) have lower RNA binding strengths than do transcribed sequences. These low RNA-binding strength regions do not allow their flanking sequences to become structurally heterochromatin (that is, they can still be transcribed under certain conditions), which is consistent with the fact that regions with low RNA-binding strengths on Xp did not lead to the complete inactivation of genes from their flanking regions with high RNA binding strengths.

Chr	Centromere length (Mb)	Centromere sequences Mean ± S.D.	Upstream sequences Mean ± S.D.	Downstream sequences Mean ± S.D.	T values vs. up stream*	p values vs. up stream**	T values vs. down stream#	p values vs. down stream##
19	2.69	5276187.71±191896.57	6795727.93±720967.15	5973147.56±526793.11	17.828	P<0.001	10.708	P<0.001
X	3.81	4767978.51±65818.50	5970283.78±448036.54	6087100.85±374574.22	23.741	P<0.001	31.011	P<0.001

Chr: chromosome. * t tests comparing the mean RNA binding strength values of centromere sequences to those of the upstream sequences.

**p values for mean RNA binding strength values of centromere sequences compared to those of the upstream sequences.

t tests comparing the mean RNA binding strength values of centromere sequences to those of the downstream sequences.

p values for mean RNA binding strength values of centromere sequences compared to those of the downstream sequences.

Table 5: Mean RNA binding strengths of the chr19 and chrX centromeres and flanking sequences

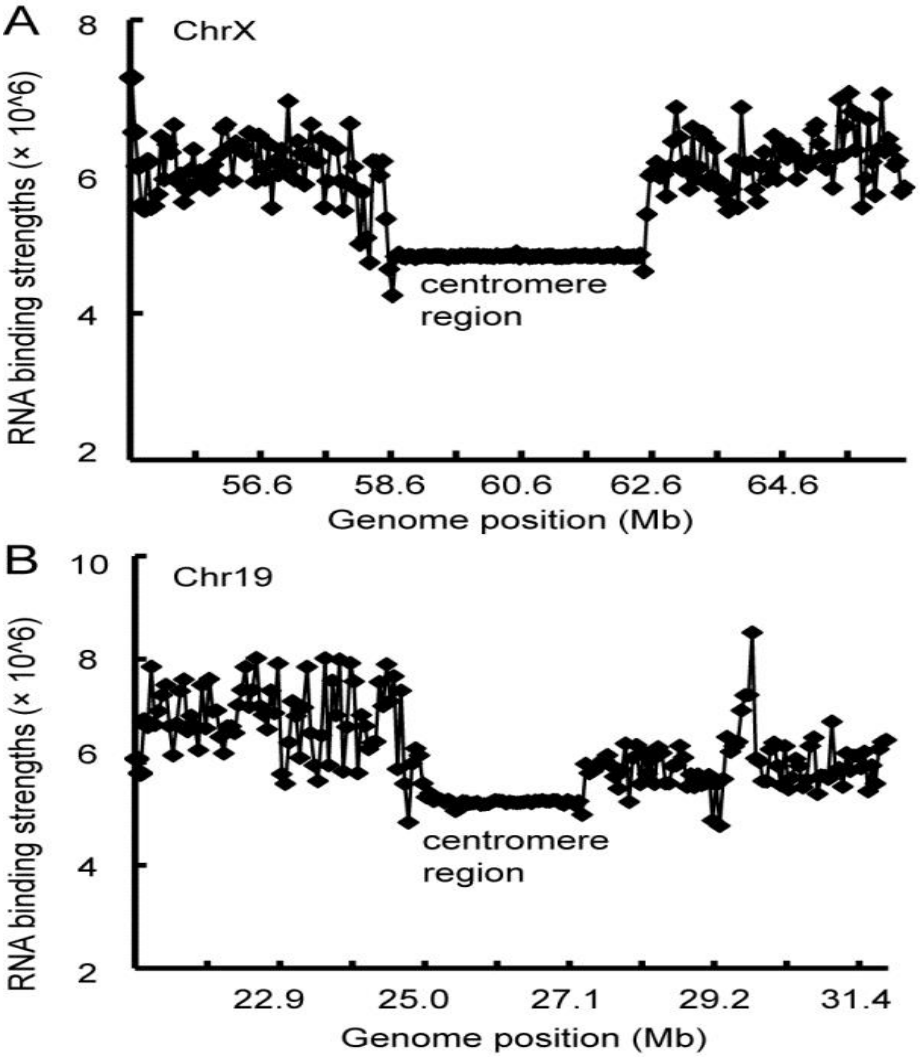

Figure 4: The RNA binding strengths of centromere regions are significantly lower than those of the corresponding flanking sequences: (**A**) RNA binding strengths of the ChrX centromere region and flanking sequences. (**B**) RNA binding strengths of the Chr19 centromere region and flanking sequences. The centromere regions and 4 Mb of upstream and downstream sequence were divided into 50-kb fragments. RNA binding strength was calculated for each fragment as described in the Materials and methods section. Each point represents the RNA binding strength of a 50-kb DNA sequence (i.e., the total number of RNAs that bind to that 50-kb DNA sequence). The y-axis represents the levels of RNA binding strengths. The x-axis represents genome position that is the chromosomal sequence map positions

Discussion

Previous studies have demonstrated that binding of RNAs to DNA may play a crucial role in activating gene expression, in a process known as RNA activation. In human cells, small regulatory RNAs can target promoters to activate gene expression [28,29]. Increasing evidence suggests that non-coding RNA (ncRNA) transcripts play a fundamental role in activating ncRNA-targeted locus via targeting epigenetic modifications [30]. Our previous studies have shown that RNAs specifically activate gene expression in a length-, position-, and sequence-dependent manner, binding of RNA with DNA affected chromatin packaging and activated genes [22,31]. However, the roles of RNA activation in XCI has not been investigated.

We introduced a computational metric for RNA binding strength to evaluate the binding ability of RNA to DNA. Our previous studies found that the RNA binding strength of centromere DNA, which consists of constitutive heterochromatin, from C group chromosomes was significantly lower than that of flanking sequences, which suggests that low RNA binding strengths were related to centromere lacking transcription [22]. Here, to explore whether low RNA binding strength may be one reason for XCI, we analyzed the simulated RNA binding strengths to DNA segments for autosomes and the X chromosome. We found that the RNA binding strength of the chrX was significantly lower than that of most autosomes (except chromosomes 8, 11, 18, 20, 21, and 22) (Table 2, Figure 1). Although the mean of RNA binding strengths of chromosomes 8, 11, 18, 20, 21, and 22 are not higher than that of chrX, they contain some regions with high RNA binding strength, which is reason that maintain them activation (Table 3).

Because Xq is almost completely inactivated, we then compared the RNA binding strength of Xq with that of the autosomes, which revealed that the RNA binding strength of Xq was significantly lower than that of autosomes (except chromosomes 11, 20, and 22) (Table 2). Although the RNA binding strength of chromosomes 11, 20, and 22 is less than or close to that of Xq, the standard deviation of RNA binding strengths in these three chromosomes were significantly higher than that of Xq, suggesting that chromosomes 11, 20 and 22 have some fragments with low RNA binding and other fragments with high RNA binding strengths. After analyzing the fragments possessing upper 50% of RNA binding strength values of Xq and the autosomes, we found that the mean upper 50% value of RNA binding strengths of Xq were significantly lower than those in chromosomes 11, 20 and 22 (Table 3). This is consistent with the results that the standard deviations of RNA binding strengths of chromosome 11, 20, and 22 were larger than that of Xq. In addition, the mean upper 50% value of RNA binding strengths of Xq was significantly lower than other autosomes. We therefore speculated that mean upper 50% value of RNA binding strengths of a autosome being higher than that of Xq was a condition of leading to its activation.

Approximately 30% of the genes located on Xp escape XCI, while the genes located on Xq are almost entirely inactivated. Using our bioinformatic approach, we found that the RNA binding strength of Xq was significantly lower than that of Xp ($p<0.01$) (Table 2) and that the RNA binding strength of the IRs on Xp was significantly lower than that of the activated regions ($p<0.01$) (Table 4). These findings provide new clues for uncovering the mechanism by which genes escape from a XCI.

Based on these results, we propose that XCI involves the entire sequence structure of the X chromosome (i.e., the base composition) rather than individual segments and factors. We further suggest that the low RNA binding strength of the human chrX may be one reason for XCI. Because Xq lacks the fragments with high RNA binding strengths, the Xq may be affected first when activators (for example: transcription factor, activating RNA) reduce and inhibitors (for example: RNA interference, DNA modification) increase.

The X chromosome and chr19 have distinctive features. Long interspersed nuclear element-1 (LINE-1) comprise 17% of the human genome[32], but accounts for 30% of chrX [33]; Alu elements belonging to the short interspersed nuclear elements of repetitive elements are present in more than one million copies which altogether represent 10% of the whole human genome [34], but accounts for 29% of chr19 [35]. In this paper, we compared the RNA binding strength of the centromeres and flanking sequences of chrX and chr19. The RNA binding strengths of the centromere regions were significantly lower than those of their flanking sequences on both chromosomes (Figure 4, Table 5). Human centromeric DNA is mainly composed of α-satellite DNA,

β-satellite DNA and γ-satellite DNA [36,37]. These satellite DNA are also scattered and transcribed in small amounts in other parts of the genome [38,39]. The low RNA binding strength of the centromeres was detected in this paper, illustrating that the small amount transcription of these satellite DNA in other parts of the genome does not affect the RNA binding strength of the centromeres.

Prior studies have indicated that autosomes translocated to the chrX could not be completely inactivated [40,41], suggesting that the sequence characteristics is important factor for determining whether inactivation. The transcription activity is reduced when autosomes translocats to the chrX but not completely inactivated, suggesting that XCI has cis action.The results of IR and ER showed that the regions with low RNA binding strength could not completely silence the activity of regions with the high RNA binding strengths, which was consistent with the results of autosomal translocation to chrX. Xq lacks fragments with high RNA binding strengths (the mean of upper 50% RNA binding strength is the lowest, and the variance of RNA binding strength of Xq is the lowest).The sequence composition of Xq is more susceptible to cis action (inactivation of one region in Xq can inactivate its adjacent regions, and vice versa), which results in inactivation or actvation cascade reactions.We proposed that the nature of cis reactions includes RNA activation (e.g., LINE-1 activation) and heterochromatin formation. These results, in combination with the analysis results of centromeres and their flanking sequences suggested that regions with low RNA binding strengths cannot completely lead to the inactivation of their flanking sequences with high RNA binding strength, and vice versa.

The activation of the X chromosome is dependent upon X-linked RNAs and upon RNAs produced from autosomes [42,43]. Both X chromosomes are active during the embryonic period, and one chrX is randomly inactivated during cell differentiation, which illustrates that XCI is associated with the genome environment of the chrX [44,45]. The observation that female diploid cells support the activation of one chrX, and female tetraploid cells support the activation of two X chromosomes also provides evidence for a relationship between XCI and the genome environment of the chrX [46-48]. After the early embryonic stage, activating factors (RNAs, transcription factors, etc.) decrease, and inhibitors (chromatin packaging, DNA modification, etc) increase, so that the cells can only support the activation of one chrX. This transition may also explain why diploid cells support the activation of one chrX and why tetraploid cells support the activation of two X chromosomes. Because the two X chromosomes in a diploid cell compete for activating factors and there are transcriptional activation effects (i.e., transcribed genes can activate themselves) [49,50], once activating factors trigger the activation of one chrX, the other chrX will be inactivated.

Previous studies have reported several key mechanisms of XCI, such as XIST coating of the chrX, the recruitment of DNA-, RNA-, and histone modification enzymes, and compaction and compartmentalization of the inactive X [51-55]. Xist represents a paradigm for the function of long non-coding RNA in epigenetic regulation. ChrX deletion studies and X; autosome translocation studies demonstrated that Xist sequences are not completely related to XCI and its maintenance, indicating that the XCI induced by Xist requires the specific sequence structure of chrX [56-58]. We proposed that once the inactivation center of chrX is triggered, the inactivation cascade requires the special structure characteristic of chrX, i.e. low and uniform RNA binding strength. The specific sequence structure characteristic of chrX is the important reason of inducing the cascade of inactivation or activation, which means that inactivation of one region leads to inactivation of surrounding sequences, or activation of one region leads to activation of surrounding sequences in chrX.

Conclusion

The top 50% RNA binding strength values of Xq were significantly lower than those of all autosomes and Xp. The RNA binding strength of the inactivation regions on Xp was significantly lower than that of the escaping regions from XCI, and the RNA binding strengths of the centromere region (structural heterochromatin, non-transcribed regions) were significantly lower than those of their flanking sequences. Taken together, these bioinformatic analyses suggest that the low RNA binding strength and the lack of fragments with high RNA binding strengths may be one of reasons for XCI.

Availability of data and material

Data and materials are available from the authors on reasonable request. The datasets generated and/or analysed during the current study are available in the Science Data Bank, http://www.doi.org/10.11922/sciencedb.00890.

Competing interests

The authors declare no conflict of interest.

Acknowledgements

We thank Dr. Angelika Hofmann of SciWri Services for providing editing support. This work was supported by the National Natural Science Foundation of China (grant number 81771499), the Hebei Province Natural Science Foundation of China (grant numbers H2018206099 and H2019206535).

References

1. Disteche CM, Berletch JB (2015) X-chromosome inactivation and escape. J Genet 94: 591-9.

2. Brenes AJ, Yoshikawa H, Bensaddek D, Mirauta B, Seaton D, et al. (2021) Erosion of human X chromosome inactivation causes major remodeling of the iPSC proteome. Cell Rep 35: 109032.

3. Patrat C, Ouimette JF, Rougeulle C (2020) X chromosome inactivation in human development. Development 147: 10.1242/dev.183095.

4. Hui L (2018) Quantifying the effects of aging and urbanization on major gastrointestinal diseases to guide preventative strategies. BMC Gastroenterol 18: 145.

5. Brunet A, Berger SL (2014) Epigenetics of aging and aging-related disease. J Gerontol A Biol Sci Med Sci 69: S17-20.

6. Yin X, Latif R, Tomer Y, Davies TF (2007) Thyroid epigenetics: X chromosome inactivation in patients with autoimmune thyroid disease. Ann N Y Acad Sci 1110: 193-200.

7. Rayner JG, Hitchcock TJ, Bailey NW (2021) Variable dosage compensation is associated with female consequences of an X-linked, male-beneficial mutation. Proc Biol Sci 288: 20210355.

8. Li C, Tian H, Chiu-Ho W, Heather K, Sun S, et al. (2016) A self-enhanced transport mechanism through long noncoding RNAs for X chromosome inactivation. Sci Rep 6: 31517.

9. Yue M, Richard JLC, Ogawa Y (2016) Dynamic interplay and function of multiple noncoding genes governing X chromosome inactivation. Biochim Biophys Acta 1859: 112-20.

10. Chow J, Heard E (2009) X inactivation and the complexities of silencing a sex chromosome. Curr Opin Cell Biol 21: 359-66.

11. Ercan S, Dick LL, Lieb JD, The C (2009) elegans dosage compensation complex propagates dynamically and independently of X chromosome sequence. Curr Biol 19: 1777-87.

12. Lucchesi JC, Kelly WG, Panning B (2005) Chromatin remodeling in dosage compensation. Annu Rev Genet 39: 615-51.

13. Anderson EC, Frankino PA, Higuchi-Sanabria R, Yang Q, Bian Q, et al. (2019) X chromosome domain architecture regulates caenorhabditis elegans lifespan but not dosage compensation. Dev Cell 51: 10.1016/j.devcel.2019.08.004.

14. Del Rosario BC, Del Rosario AM, Anselmo A, Wang PI, Sadreyev RI, et al. (2017) Genetic intersection of Tsix and Hedgehog signaling during the initiation of X-chromosome inactivation. Dev Cell 43: 359-71.

15. Loos F, Maduro C, Loda A, Lehmann J, Kremers GJ, et al. (2016) Xist and Tsix transcription dynamics is regulated by the X-to-autosome ratio and semistable transcriptional states. Mol Cell Biol 36: 2656-67.

16. Escamilla-Del-Arenal M, Rocha STD, Heard E (2011) Evolutionary diversity and developmental regulation of X-chromosome inactivation. Hum Genet 130: 307-27.

17. Huang V, Qin Y, Wang J, Wang X, Place RF, et al. (2010) RNAa is conserved in mammalian cells. PloS ONE 5: e8848.

18. Janowski BA, Younger ST, Hardy DB, Ram R, Huffman KE, et al. (2007) Activating gene expression in mammalian cells with promoter-targeted duplex RNAs. Nat Chem Biol 3: 166-73.

19. Li LC, Okino ST, Zhao H, Pookot D, Place RF, et al. (2006) Small dsRNAs induce transcriptional activation in human cells. Proc Natl Acad Sci U S A 103: 17337-42.

20. Matsui M, Sakurai F, Elbashir S, Foster DJ, Manoharan M, et al. (2010) Activation of LDL receptor expression by small RNAs complementary to a noncoding transcript that overlaps the LDLR promoter. Chem Biol 17: 1344-55.

21. Zheng L, Wang L, Gan J, Zhang H (2014) RNA activation: promise as a new weapon against cancer. Cancer Lett 355: 18-24.

22. Wang X, Ma Z, Kong X, Lv Z (2016) Effects of RNAs on chromatin accessibility and gene expression suggest RNA-mediated activation. Int J Biochem Cell Biol 79: 24-32.

23. Cheng J, Wang X, Cai N, Ma Z, Zhang L, Lv Z. RNAs specifically affect gene expression in a length, position and sequence dependent manner. Int J Clin Exp Pathol 7: 948-58.

24. Ma Z, Kong X, Liu S, Yin S, Zhao Y, et al. (2017) Combined sense-antisense Alu elements activate the EGFP reporter gene when stable transfection. Mol Genet Genomics 292: 833-46.

25. Obata-Onai A, Hashimoto S, Onai N, Kurachi M, Nagai S, et al. (2002) Comprehensive gene expression analysis of human NK cells and CD8(+) T lymphocytes. Int Immunol 14: 1085-98.

26. Swanson KM, Stelwagen K, Dobson J, Henderson HV, Davis SR, et al. (2009) Transcriptome profiling of Streptococcus uberis-induced mastitis reveals fundamental differences between immune gene expression in the mammary gland and in a primary cell culture model. J Dairy Sci 92: 117-29.

27. Carrel L, Cottle AA, Goglin KC, Willard HF (1999) A first-generation X-inactivation profile of the human X chromosome. Proc Natl Acad Sci U S A. 96: 14440-4.

28. Morris KV, Santoso S, Turner AM, Pastori C, Hawkins PG (2008) Bidirectional transcription directs both transcriptional gene activation and suppression in human cells. PLoS Genet 4: e1000258.

29. Morris KV (2009) RNA-directed transcriptional gene silencing and activation in human cells. Oligonucleotides 19: 299-306.

30. Damski C, Morris KV (2014) Targeted small noncoding RNA-directed gene activation in human cells. Methods Mol Biol 1173: 1-10.

31. Cheng J, Wang X, Cai N, Ma Z, Zhang L, et al. RNAs specifically affect gene expression in a length, position and sequence dependent manner. Int J Clin Exp Pathol 7: 948-58.

32. Suarez NA, Macia A, Muotri AR (2018) LINE-1 retrotransposons in healthy and diseased human brain. Dev Neurobiol 78: 434-55.

33. Gendrel AV, Heard E (2014) Noncoding RNAs and epigenetic mechanisms during X-chromosome inactivation. Annu Rev Cell Dev Biol 30: 561-80.

34. Stenz L (2021) The L1-dependant and Pol III transcribed Alu retrotransposon, from its discovery to innate immunity. Mol Biol Rep 48: 2775-89.

35. Leem SH, Kouprina N, Grimwood J, Kim JH, Mullokandov M, et al. (2004) Closing the gaps on human chromosome 19 revealed genes with a high density of repetitive tandemly arrayed elements. Genome Res 14: 239-46.

36. Jain M, Olsen HE, Turner DJ, Stoddart D, Bulazel KV, et al. (2018) Linear assembly of a human centromere on the Y chromosome. Nat Biotechnol 36: 321-3.

37. Logsdon GA, Gambogi CW, Liskovykh MA, Barrey EJ, Larionov V, et al. (2019) Human artificial chromosomes that bypass centromeric DNA. Cell 178: 624-39.e19.

38. McNulty SM, Sullivan BA (2018) Alpha satellite DNA biology: finding function in the recesses of the genome. Chromosome Res 26: 115-38.

39. Kim JH, Ebersole T, Kouprina N, Noskov VN, Ohzeki J, et al. (2009) Human gamma-satellite DNA maintains open chromatin structure and protects a transgene from epigenetic silencing. Genome Res 19: 533-44.

40. Favilla BP, Meloni VA, Perez AB, Moretti-Ferreira D, de Souza DH, et al. (2021) Spread of X-chromosome inactivation into autosomal regions in patients with unbalanced X-autosome translocations and its phenotypic effects. Am J Med Genet A Online ahead of print. 10.1002/ajmg.a.62228.

41. Di-Battista A, Moysés-Oliveira M, Melaragno MI (2020) Genetics of premature ovarian insufficiency and the association with X-autosome translocations. Reproduction 160: R55-64.

42. Migeon BR (2019) The non-random location of autosomal genes that participate in X inactivation. Front Cell Dev Biol 7: 144.

43. Migeon BR (2021) Stochastic gene expression and chromosome interactions in protecting the human active X from silencing by XIST. Nucleus 12: 1-5.

44. Zito A, Davies MN, Tsai PC, Roberts S, Andres-Ejarque R, et al. (2019) Heritability of skewed X-inactivation in female twins is tissue-specific and associated with age. Nat Commun 10: 5339.

45. Cotton AM, Price EM, Jones MJ, Balaton BP, Kobor MS, et al. (2015) Landscape of DNA methylation on the X chromosome reflects CpG density, functional chromatin state and X-chromosome inactivation. Hum Mol Genet 24: 1528-39.

46. Liu W, Sun X (2009) Skewed X chromosome inactivation in diploid and triploid female human embryonic stem cells. Hum Reprod 24: 1834-43.

47. Lu Z, Carter AC, Chang HY (2017) Mechanistic insights in X-chromosome inactivation. Philos Trans R Soc Lond B Biol Sci 372: 1733.

48. Jeon Y, Lee JT (2011) YY1 tethers Xist RNA to the inactive X nucleation center. Cell 146: 119-33.

49. Mikhaylova LM, Nurminsky DI (2011) Lack of global meiotic sex chromosome inactivation, and paucity of tissue-specific gene expression on the Drosophila X chromosome. BMC Biol 9: 29.

50. Mutzel V, Schulz EG (2020) Dosage sensing, threshold responses, and epigenetic memory: a systems biology perspective on random X-chromosome inactivation. Bioessays 42: e1900163.

51. Cerase A, Young AN, Ruiz NB, Buness A, Sant GM, et al. (2021) Chd8 regulates X chromosome inactivation in mouse through fine-tuning control of Xist expression. Commun Biol 4: 485.

52. Yu B, Qi Y, Li R, Shi Q, Satpathy AT, et al. (2021) B cell-specific XIST complex enforces X-inactivation and restrains atypical B cells. Cell 184: 1790-803.e17.

53. Kolpa HJ, Fackelmayer FO, Lawrence JB (2016) SAF-A requirement in anchoring XIST RNA to chromatin varies in transformed and primary cells. Dev Cell 39: 9-10.

54. Sakaguchi T, Hasegawa Y, Brockdorff N, Tsutsui K, Tsutsui KM, et al. (2016) Control of chromosomal localization of Xist by hnRNP U family molecules. Dev Cell 39: 11-2.

55. Makhlouf M, Ouimette JF, Oldfield A, Navarro P, Neuillet D, et al. (2014) A prominent and conserved role for YY1 in Xist transcriptional activation. Nat Commun 5: 4878.

56. Hall LL, Clemson CM, Byron M, Wydner K, Lawrence JB (2002) Unbalanced X;autosome translocations provide evidence for sequence specificity in the association of XIST RNA with chromatin. Hum Mol Genet 11: 3157-65.

57. Popova BC, Tada T, Takagi N, Brockdorff N, Nesterova TB (2006) Attenuated spread of X-inactivation in an X;autosome translocation. Proc Natl Acad Sci U S A 103: 7706-11.

58. Brown CJ, Willard HF (1994) The human X-inactivation centre is not required for maintenance of X-chromosome inactivation. Nature 368: 154-6.