Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 2 November 2018 doi:10.20944/preprints201811.0066.v1

Low RNA Binding Strength of Human X Chromosome May Explain X Chromosome Inactivation

Ning Ji，Lifang Yan, Zhixue Song, Shufeng Liu, Chao Liu, Shuangshuang Yang, Hongyi Cao, Xiufang Wang* and Zhanjun Lv*

Department of Genetics, Hebei Medical University, Hebei Key Lab of Laboratory Animal, Shijiazhuang 050017, Hebei Province, China; 854869776@qq.com (N.J.); lifangyiyaa@163.com (L.Y.); songzhixue123@163.com (Z.S.); liushf@126.com (S.L.); 5595758@163.com (C.L.); 17803211887@163.com (S.Y.); 3253503763@qq.com (H.C.); wangxiufang@live.cn (X.W.); zhanjun_lv@hotmail.com (Z.L.)

* Correspondence: wangxiufang@live.cn (X. W.) and zhanjun_lv@hotmail.com (Z. L.); Tel: +86-311-86266844.

© 2018 by the author(s). Distributed under a Creative Commons CC BY license.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 2 November 2018 doi:10.20944/preprints201811.0066.v1

Abstract Two X chromosomes of female mammals randomly inactivate one of paternal or maternal X chromosome in early embryonic development and all the daughter cells produced from these cells retain the same feature of X chromosome inactivation, which is called X chromosome inactivation (XCI). Studying the mechanisms of XCI is important for understanding epigenetic that plays an important role in age-associated diseases. The previous studies have demonstrated that binding of RNAs and DNAs may play a role in activating gene expression. In this paper, our study aims to explore whether the mechanisms of XCI involve the RNA binding strength to X chromosome DNAs. The bioinformatics analyses based on big data were used to analyze the simulated binding strength of RNAs (RNA binding strength) to 23 chromosomes (including X chromosome and 22 human autosomes) and the characteristics of repetitive sequences in the X-inactivation centre. The results revealed that RNA binding strength of the long arm of the X chromosome that is almost entirely inactivated in XCI was significantly lower than that of all autosomes and the short arm of X chromosome, meanwhile the RNA binding strengths of inactivation regions in X chromosome were significantly lower than that of regions escaping from XCI. Different repetitive sequence clusters within the center of XCI presented a cross distribution characteristic. To further prove whether the repetitive sequences in human X chromosome involve in XCI, we cloned long interspersed element (LINE-1, L1) and short interspersed element (Alu) from human Xq13, the center of XCI, and constructed expression vectors carrying sense-antisense combination repetitive sequences (L1s or Alus). Effects of combined L1 or combined

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 2 November 2018 doi:10.20944/preprints201811.0066.v1

Alu sequences on expression of EGFP reporter gene were examined in stably transfected HeLa cells, which simulates the effects of repetitive sequences located on chromosomes. The results of experiments revealed transcribed L1 repetitive sequences activated *EGFP* reporter gene expression, so did the Alus. The experiment results suggested repetitive sequences activated genes by interaction of transcribed RNAs and DNAs. Since the binding of RNAs and DNAs can activate gene, so the low RNA binding strength of human X chromosome may be one of reasons of XCI. The cross distribution characteristics of different repetitive sequence clusters leading to a cascade of gene activation or gene inactivation may be the reason of transcriptional silencing one of the X chromosomes in female mammals.

Keywords: Alu; LINE-1 (L1); X-chromosome inactivation (XCI); repetitive sequences; RNA binding strength

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 2 November 2018 doi:10.20944/preprints201811.0066.v1

1.Introduction

Aging is the most important risk factor inducing human aged diseases and death and epigenetic plays an important role in aging [1,2]. Two X chromosomes of female mammals randomly inactivate one of paternal or maternal X chromosome in early embryonic development and all the daughter cells produced from these cells retain the same feature of X chromosome inactivation, which is called X chromosome inactivation (XCI) [3]. XCI is the important phenomenon of epigenetic and the model of studying the mechanisms of epigenetic [4]. The dosage compensation of X-linked gene products between the sexes is crucial for growth and development [5]. In *Drosophila* and *Caenorhabditis elegans* (*C. elegans*), the activity of female X chromosomes is half of that of male X chromosomes to maintain a balance[6-9]. X chromosome inactivation (XCI) occurs early in mammalian development [10] and serves to equalize the dosage of X chromosome gene products between females and males by transcriptional silencing one of the X chromosomes in female mammals [11]. Although previous studies have indicated that two X-linked genes (Xist and Tsix) [12-14] and differential DNA methylation [15] have been certified to play an important role in XCI, the definite mechanism of XCI remains exclusive.

Repetitive sequences are ubiquitous in multicellular eukaryotes, which constitute more than 60% of the whole human genome [16,17], and are considered to be crucial for maintaining life activities of organism. Higher abundance of repetitive sequences have been detected in mammalian sex chromosomes than autosomes [18-20] and accumulated evidences have indicated that repetitive sequences play crucial roles in

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 2 November 2018 doi:10.20944/preprints201811.0066.v1

XCI [21-24]. Researching the association between repetitive sequences and XCI may provide new clues for the mechanism of XCI which contributes to the normal growth, development and aging of mammals.

Long interspersed element, LINE-1 (L1) and short interspersed element (Alu) are two of most key repetitive sequences, which account for 17% and 10% of human genome respectively. L1 is the most overrepresented type of repetitive sequences on the X chromosome in eutherians and the copy number of L1 in X chromosome was higher than that in the autosomes [25]. Although XCI results in the silence of most genes on one X chromosome, approximately 30% of genes located on the short arm of the X chromosome (Xp) escape from XCI [26]. Moreover, XCI can spread to a distance of 21.4 kb in neighboring regions, and may play a role in cis X inactivation [27]. High density of L1 sequence in X chromosomes was reported to promote heterochromatin to spread throughout the X chromosome [28-30]. Alu was reported to be involved in XCI as well [23], while the precise role of Alu needs to be uncovered. Since XCI is a mechanism of gene dosage compensation by regulating gene expression [31-34] and our previous studies proved that L1s and Alus played a role in regulating gene expression [35-36], we speculated that these two repetitive sequences (L1 and Alu) may be involved in the process of XCI by regulating gene expression.

In human cells, interaction of RNAs and DNAs can activate gene expression [37-40] and this phenomenon is called RNA activation [41,42]. Our previous studies showed that RNAs specifically activate gene expression in a position, length and sequence dependent manner, providing evidence of RNA regulatory networks [43, 44].

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 2 November 2018 doi:10.20944/preprints201811.0066.v1

As a regulatory mechanism of gene expression, RNA activation may be associated with XCI. We supposed that RNAs may activate gene expression by binding to DNAs to prevent chromatin over-packing. Based on it, the strength of RNA binding to DNA (RNA binding strength) can be used to assess chromatin activity.

In this paper, to explore whether RNA binding strength involves the XCI we used bioinformatics methods to analyze the RNA binding strength to X chromosome and autosomes. The transient and stable transfections were used to observe the effects of one repetitive sequence on another repetitive sequence located in same expression vector, which simulates the effects between repetitive sequences located on chromosomes. Based on our bioinformatics analyses and experimental results, we proved that the low RNA binding strength of human X chromosome was associated with XCI.

2.Results

2.1. The results of bioinformatics analyses proved that the low RNA binding strength of human X chromosome was associated with XCI

2.1.1.The comparison of RNA binding strengths of X chromosome and autosomes

We analyzed the RNA binding strength of human autosomes 1-22 and the X chromosome. The chromosome DNAs were divided into 50-kb fragments that were

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 2 November 2018 doi:10.20944/preprints201811.0066.v1

recorded as a set of numbers of DNA 7-nt strings via bioinformatics methods. RNA 7-nt strings were looked up for RNAs transcribed from tonsil germinal center B cells. The DNA 7-nt strings were multiplied by strings of RNAs, and the sum of products represents the strength of binding the 50-kb DNA sequences to RNAs (RNA binding strength algorithm is written in Materials and Methods). This analysis revealed that the binding strength of RNAs to X chromosome is significantly lower than that to all autosomes other than chromosomes 8, 11, 18, 20, 21 and 22 (Table 1). We used chr19 as a representative of autosomes since the genes in chr19 are highly expressed. As shown in Figure 1, RNA binding strength and its dispersion degree of X chromosome (Figure 1A) are lower than in chr19 (Figure 1B).

Table 1. Comparison of RNA binding strengths of autosomes with X chromosome

Chromosomes	Number of fragments (each fragment is 50 kb)	Mean ± S.D. of RNA binding strengths	*P* values *vs.* X	*P* values *vs.* Xq	*P* values of fragments with high RNA binding strength *vs.* that of Xq
1	4438	6333574.22 ± 691245.81	<0.01*	<0.01**	<0.01***
2	4758	6314539.81 ± 565702.45	<0.01	<0.01	<0.01
3	3884	6337212.06 ± 552823.21	<0.01	<0.01	<0.01
4	3747	6370787.96 ± 476562.48	<0.01	<0.01	<0.01
5	3556	6320322.79 ± 560634.19	<0.01	<0.01	<0.01
6	3346	6351251.86 ± 512578.84	<0.01	<0.01	<0.01
7	3102	6422567.89 ± 635270.25	<0.01	<0.01	<0.01
8	2845	6273460.86 ±		<0.01	<0.01

		550807.81			
9	2334	6306166.84 ± 605680.41	<0.05	<0.01	<0.01
10	2633	6304370.16 ± 672467.68	<0.05	<0.01	<0.01
11	*2620*	*6189916.53 ± 624444.70*			*<0.01*
12	2595	6415714.81 ± 666539.13	<0.01	<0.01	<0.01
13	1915	6373274.15 ± 478290.45	<0.01	<0.01	<0.01
14	1745	6364194.37 ± 662510.32	<0.01	<0.01	<0.01
15	1632	6346786.11 ± 665506.10	<0.01	<0.01	<0.01
16	1602	6367224.26 ± 740683.39	<0.01	<0.01	<0.01
17	1559	6524509.02 ± 831046.94	<0.01	<0.01	<0.01
18	1495	6281989.52 ± 518609.70		<0.01	<0.01
19	1117	6721733.43 ± 807647.61	<0.01	<0.01	<0.01
20	*1192*	*6193198.50 ± 753727.48*			<0.01
21	682	6279488.50 ± 578744.81		<0.01	<0.01
22	695	6242274.77 ± 831045.21			<0.01
Xp	1142	6345337.19 ± 562236.71			<0.01
X	2994	6267628.33 ± 526047.84			
Xq	1852	6219616.19 ± 496520.70			

*: Mean ± S.D. of RNA binding strengths of chromosome 1 compared with that of the X chromosome using *t*-test. **: Mean ± S.D. of RNA binding strengths of chromosome 1 compared with that of Xq using *t*-test. ***: RNA binding strengths of different 50-kb fragments on chromosome 1 and Xq were ordered descendingly, then the former 1/2 fragments of the high RNA binding strengths were chosen and compared using *t*-test.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 2 November 2018 doi:10.20944/preprints201811.0066.v1

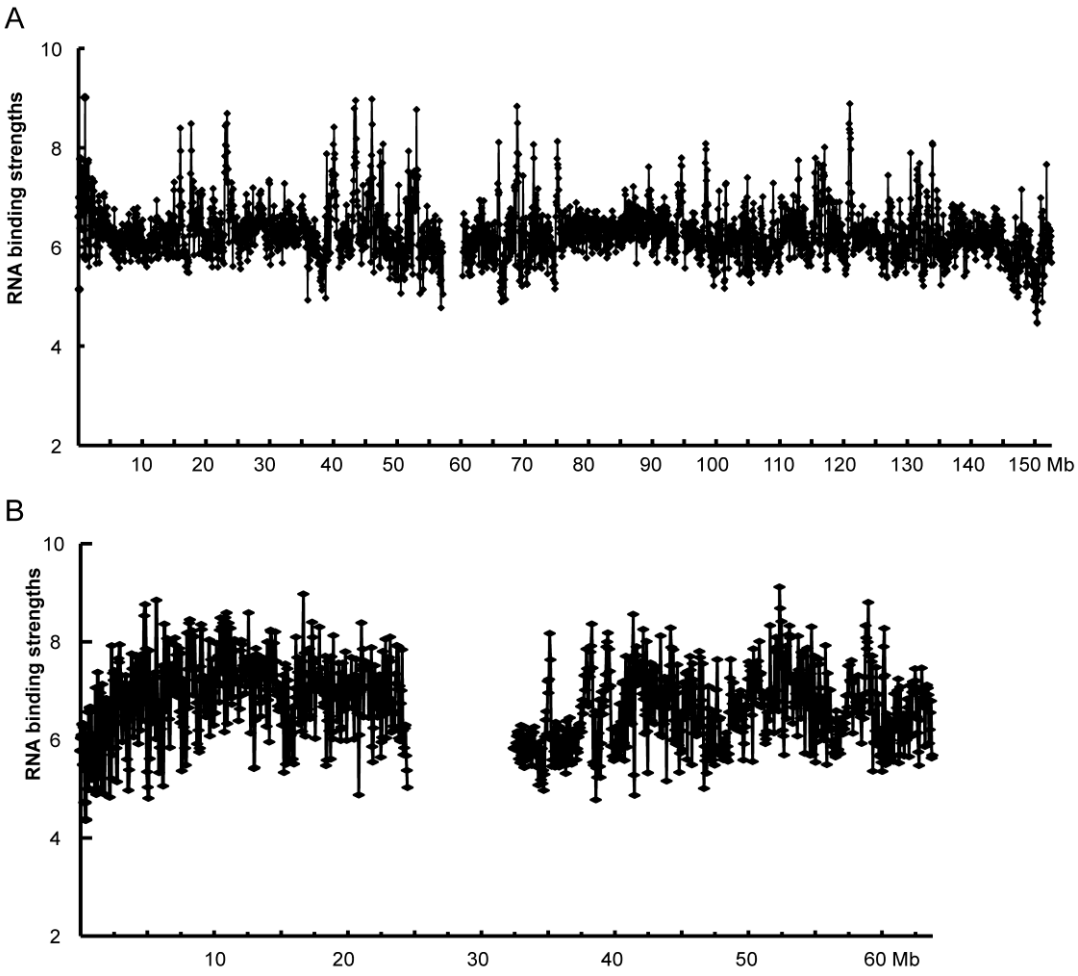


Figure 1. RNA binding strengths to chromosomes 19 and X. (**A**) Chromosome 19; (**B**) X chromosome. The y-axis represents the RNA binding strengths; numbers 2, 4, 6, 8 and 10 represent 2×10^6, 4×10^6, 6×10^6, 8×10^6 and 10×10^6. The x-axis represents the gene position.

2.1.2.The comparison of RNA binding strengths of Xq and autosomes

Since long arm of X chromosome (Xq) is almost entirely inactivated, the RNA binding strength of Xq was further compared with that of autosomes; and the results

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 2 November 2018 doi:10.20944/preprints201811.0066.v1

revealed that the RNA binding strength of Xq was significantly lower than that of autosomes, except chromosomes 11, 20, and 22 (Table 1).

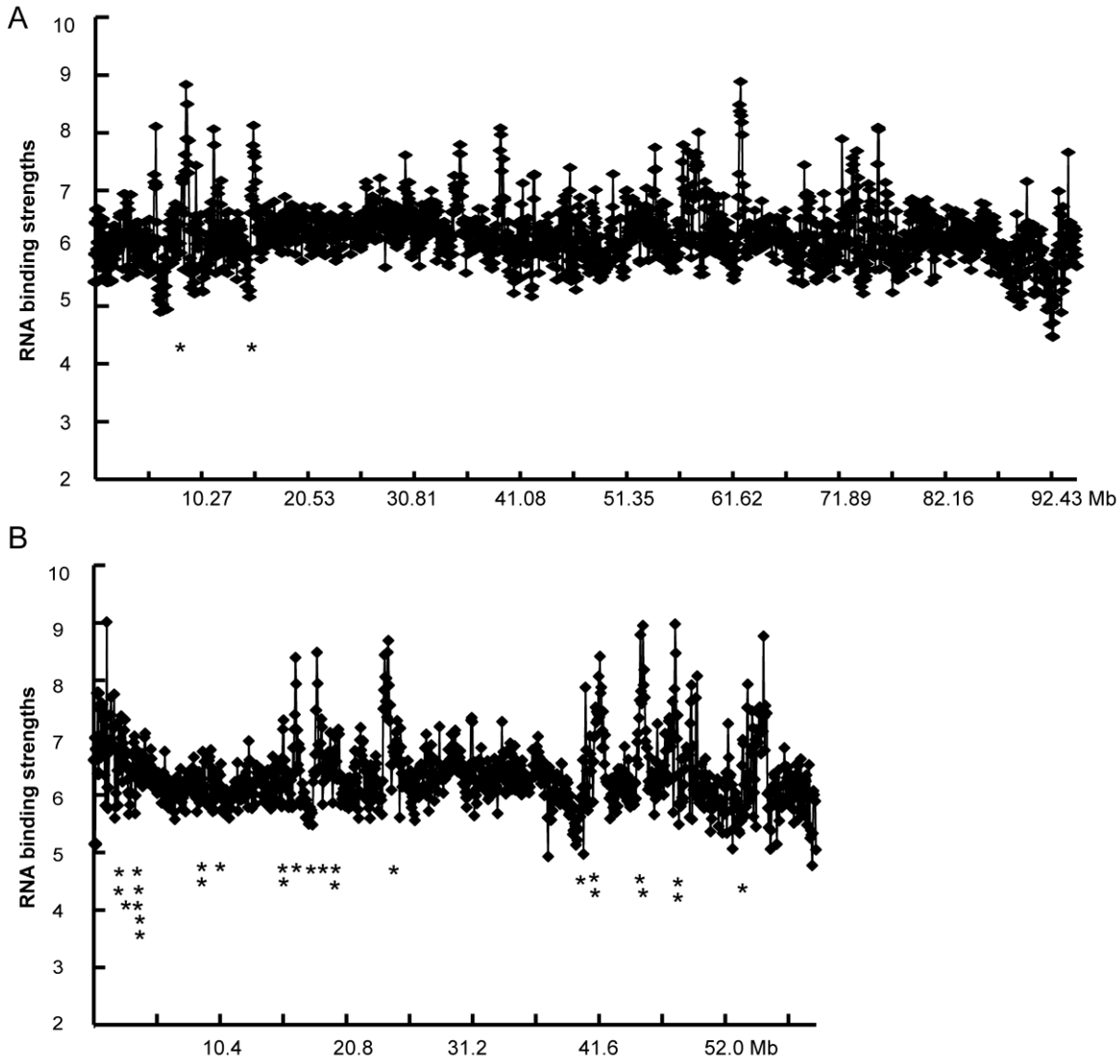
Then we found that the standard deviation of RNA binding strength in Xq was lower than that in 11, 20, and 22 chromosomes (Table 1), which illustrates that RNA binding strengths between different 50 kb DNA fragments in Xq have higher uniformity, i.e. lower dispersion degree. Since the fragments possessing high RNA binding strength easily are activated, the RNA binding strengths of the fragments with the upper half values in Xq were found to be significantly lower than that in all autosomes including 11, 20, and 22 (Table 1), which illustrates Xq lacks the fragments possessing high RNA binding strengths. Although the mean value of RNA binding strengths is a indicator that decides XCI, high RNA binding strength value of a fragment is also a indicator since activation of fragments with high RNA binding strengths does not entirely affected by the flanking sequences if the RNA binding strengths are beyond a certain threshold.

2.1.3.The comparison of RNA binding strengths of Xq with Xp

Approximately 30% genes located on short arm of X chromosome (Xp) escape from XCI while Xq is almost entirely inactivated. Carrel et al [44] reported that 29 of 224 X-linked genes escaped from XCI and 2 of these genes escaping from XCI locate on Xq and other 27 genes locate on Xp (Figure 2). The results of bioinformatic methods revealed that RNA binding strength of Xq was significantly lower than that

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 2 November 2018 doi:10.20944/preprints201811.0066.v1

of Xp ($p<0.01$) (Figure 2, Table 1).



Fi

fr[illegible] escape from XCI in the Xp shown as below: 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. The y-axis represents the RNA binding strengths; numbers 2, 4, 6, 8 and 10 represent 2×10^6, 4×10^6, 6×10^6, 8×10^6 and 10×10^6. The x-axis represents the gene position.

2.1.4.The comparison of RNA binding strengths of inactivation regions with regions escaping from XCI on Xp

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 2 November 2018 doi:10.20944/preprints201811.0066.v1

Since approximately 30% genes located on Xp escape from XCI, we compared the RNA binding strengths of inactivation regions with regions escaping from XCI on Xp and found that RNA binding strength of inactivation regions was significantly lower than that of the regions escaping from XCI (Table 2).

Table 2. Comparison of RNA binding strengths of inactivation regions with the regions escaping from XCI on Xp

	Genome position		Total fragment number	RNA binding strengths (Means ± SD)
	bp	The number of 50 kb fragment		
Inactivation regions	7200001-12850000*	145-257**	801***	2542266 ± 217373.3****, #
	16050001-39300000	322-786		
	45900001-57200000	919-1144		
Regions escaping from XCI	1-7200000	1-144	332	2629104 ± 242311.4
	12850001-16050000	258-321		
	39300001-45900000	787-918		

*7200001-12850000 represents that this inactivation region is from 7200001 bp to 12850000 bp at genome position.

** 145-257 represents that this inactivation region is from the 145th 50 kb DNA fragment to 257th 50kb DNA fragment.

***801 represents that the three inactivation regions totally include 801 50 kb DNA fragments.

****2542266 ± 217373.3 represents the mean value of RNA binding strengths of the three inactivation regions.

represents that RNA binding strength of the three inactivation regions is significantly lower than that of the three regions escaping from XCI ($p<0.01$).

2.1.5. Distribution of repetitive sequences within 1 Mb in the X-inactivation centre

In order to analyze the roles of repetitive sequences in XCI, the distribution of repetitive sequences within the 1-Mb downstream of the RPS4X gene in the X-inactivation centre was compared. As shown in Figure 3, L1s were the most abundant repetitive sequences within 1 Mb. Repetitive sequences (L1, Alu, MIR, L2,

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 2 November 2018 doi:10.20944/preprints201811.0066.v1

L3, LTR and DNA elements) are distributed in clusters, and there are alternative distributions of different repetitive sequence clusters in the 1 Mb (Figure 3). There are peaks of other repetitive sequences at the low ebb regions of L1 (Table 3).

Table 3. The cross distribution of repetitive sequences within 1 Mb at the Xq13 region

Repetitive sequences	The low ebb regions (kb) of L1 sequence distribution and the other repetitive sequences with a peak at these low ebbs															
L1	38	85	122	280	305	380	440	515	578	603	670	732	776	812	865	990
Alu		+		+	+	+	+		+	+	+				+	
MIR		+	+	+	+		+		+	+						
L2	+	+	+	+	+	+	+	+		+	+	+		+		
L3								+								
LTR	+	+	+					+	+	+	+		+		+	+
DNA		+	+	+	+		+		+	+					+	

+: peaks of other repetitive sequences at the low ebb regions of L1.
L1: long interspersed nuclear element-1.
L2: long interspersed nuclear element-2.
L3: long interspersed nuclear element-3.
MIR: mammalian-wide interspersed repeat.
LTR: long terminal repeat.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 2 November 2018 doi:10.20944/preprints201811.0066.v1

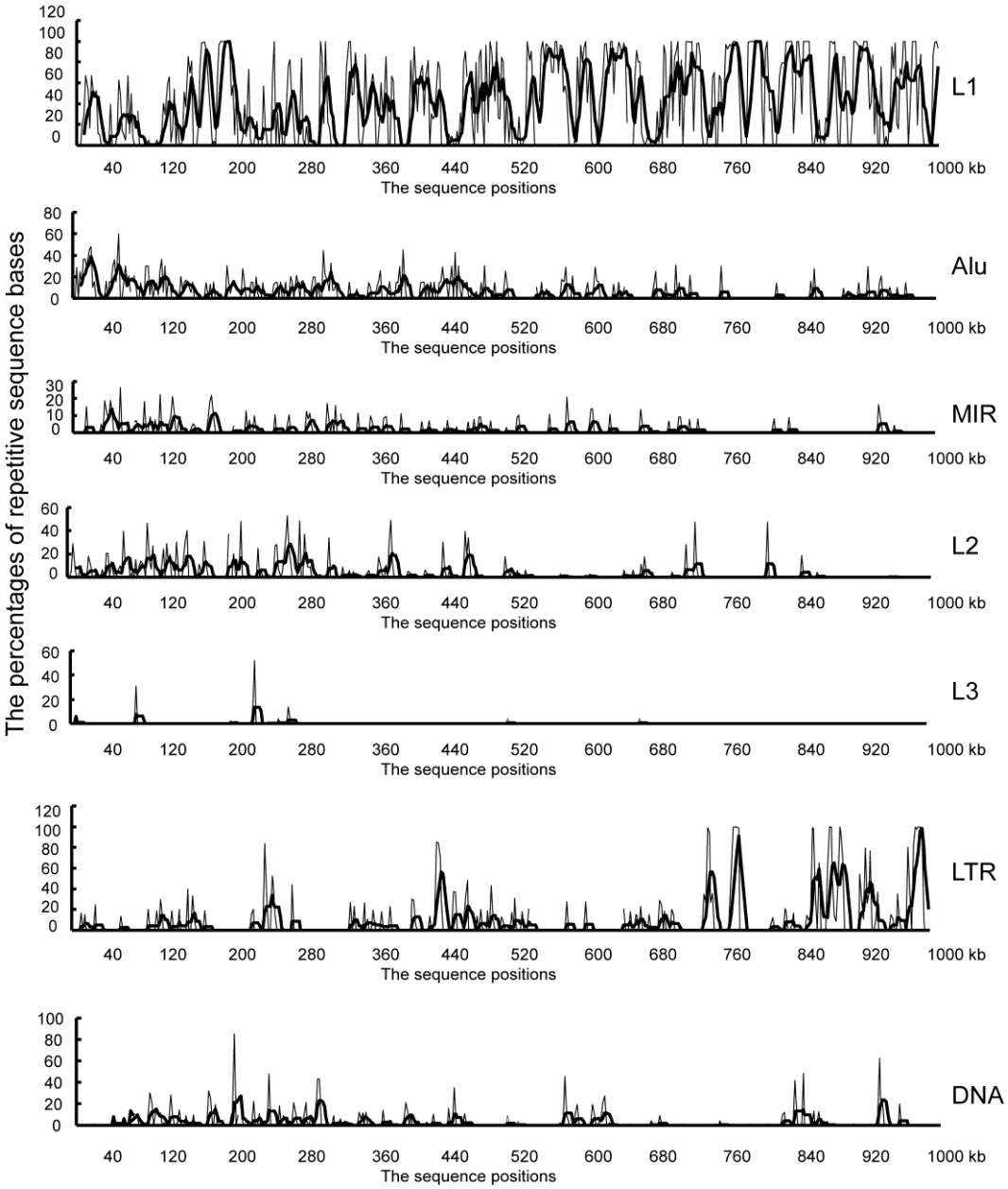


Figure 3. The distribution of repetitive sequences within the 1-Mb downstream of the RPS4X gene in Xq13. A 1-Mb fragment within Xq13 was divided into 2-kb fragments. The distribution of L1, Alu, MIR, L2, L3, LTR and DNA elements was analyzed. The x-axis represents the sequence position, while the y-axis represents the percentage of repetitive sequence bases.

2.2. The results of experiments revealed transcribed repetitive sequences activated EGFP reporter gene expression

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 2 November 2018 doi:10.20944/preprints201811.0066.v1

2.2.1. The transcribed L1s activated EGFP reporter gene expression

The results from bioinformatics analysis have demonstrated that RNA binding strengths of X chromosome are low and there are alternative distributions of different repetitive sequence clusters. To further prove whether the repetitive sequences in human X chromosome involve in XCI, we cloned L1 from human Xq13, the X-inactivation centre, and constructed different expression vectors carrying combined sense and antisense repetitive sequences. Effects of sense or antisense L1 sequences on expression of EGFP reporter were examined in transiently and stably transfected HeLa cells, simulating the effects of repetitive sequences located on chromosomes.

Firstly, the effects of repetitive sequences locating in different plasmids on EGFP gene expression using co-transfection methods were observed. HeLa cells were co-transfected transiently with modulator plasmids carrying sense or antisense repetitive sequences (derived from pcDNA3.1), and with reporter plasmids carrying sense repetitive sequences (derived from pEGFP-C1). The EGFP labeled amount of cells co-transfected with C1-280-1 and pc280-1as (Table 4, Figure 4B) was 0.94 fold that of cells co-transfected with C1-280-1 and pc280-1 (Table 4, Figure 4A). These results showed a trend in antisense inhibition (no significant difference, Figure 4H).

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 2 November 2018 doi:10.20944/preprints201811.0066.v1

Table 4. Expression vectors used in this study

Expression vector annotation	Expression vectors	Fragments inserted into pEGFP-C1 or pcDNA3.1 and annotation
C1-Alu	C1-4TMI×2-Alu×14	Two copies of 4TMI×2 were inserted upstream of Alu×14 in the C1-Alu×14 vector [42].
C1-Alu-sense-sense	C1-4TMI×2-Alu×14-CMV-Alu×2	CMV promoter was inserted downstream of Alu×14 in the C1-4TMI×2-Alu×14 vector to construct C1-4TMI×2-Alu×14-CMV, then two copies of Alu were inserted downstream of CMV in the C1-4TMI×2-Alu×14-CMV vector.
C1-Alu-sense-antisense	C1-4TMI×2-Alu×14-CMV-Alu×2as	Two copies of Alu in antisense orientation were inserted downstream of CMV in the C1-4TMI×2-Alu×14-CMV vector.
C1-Alu-sense-control	C1-4TMI×2-Alu×14-CMV- LacZ2F2R×2	Two copies of LacZ2F2R were inserted downstream of CMV in the C1-4TMI×2-Alu×14-CMV vector. LacZ2F2R is the second 280-bp fragment of LacZ gene of Escherichia coli.
C1-280-1	C1-4TMI×2-280-1×14	Two copies of 4TMI×2 were inserted upstream of the 280-1×14 in the C1-280-1×14 vector [43].
C1-280-1-sense-sense	C1-4TMI×2-280-1×14-CMV-280-1×2	CMV promoter was inserted downstream of 280-1×14 in the C1-4TMI×2-280-1×14 vector to construct C1-4TMI×2-280-1×14-CMV, and two copies of 280-1 were inserted downstream of CMV in the C1-4TMI×2-280-1×14-CMV vector.
C1-280-sense-antisense	C1-4TMI×2-280-1×14-CMV-280-1×2as	Two copies of 280-1 in antisense orientation were inserted downstream of CMV in the C1-4TMI×2-280-1×14-CMV vector.
C1-280-sense-control	C1-4TMI×2-280-1×14-CMV- LacZ2F2R×2	Two copies of LacZ2F2R were inserted downstream of CMV in the C1-4TMI×2-280-1×14-CMV vector.
C1-280-1-antisense-sense	C1-4TMI×2-280-1×14as-CMV-280-1×2	CMV promoter was inserted downstream of 280-1×14as in the C1-4TMI×2-280-1×14as vector to construct C1-4TMI×2-280-1as×14-CMV, and two copies of 280-1 were inserted downstream of CMV in the C1-4TMI×2-280-1×14as-CMV vector.
C1-280-1-antisense-antisense	C1-4TMI×2-280-1×14as-CMV-280-1×2as	Two copies of 280-1as were inserted downstream of CMV in the C1-4TMI×2-280-1×14as-CMV vector.
C1-280-1-antisense-control	C1-4TMI×2-280-1×14as-CMV-LacZ2F2R×2	Two copies of LacZ2F2R were inserted downstream of CMV in the C1-4TMI×2-280-1×14as-CMV vector.
pc280-1	pcDNA-280-1×2	Two copies of 280-1 fragment were inserted in MCS in pcDNA3.1.
pc280-1as	pcDNA-280-1×2as	Two copies of 280-1 fragment in antisense orientation were inserted in MCS in pcDNA3.1.
pcAlu	pcDNA-Alu×2	Two copies of Alu fragment were inserted in MCS in pcDNA3.1.
pcAluas	pcDNA-Alu×2as	Two copies of Alu fragment in antisense orientation were inserted in MCS in pcDNA3.1.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 2 November 2018 doi:10.20944/preprints201811.0066.v1

Next, the effects of repetitive sequences locating in the same plasmids on downstream EGFP gene expression in transient and stable transfected cells were compared. C1-280-1-sense-sense, C1-280-1-sense-antisense, C1-280-1-sense-control plasmids (Table 4) were transfected into HeLa cells. Then, expression of EGFP was determined using flow cytometry after transient or stable transfection. Following transient transfection, the EGFP labeled amount of C1-280-1-sense-antisense (Figure 4D) was 0.71 fold that of C1-280-1-sense-sense (Figure 4C), which reveals a significant antisense inhibition ($p<0.05$, Figure 4I). The results of stable transfection are different from that of transient transfection, the EGFP labeled amount of C1-280-1-sense-antisense (Figure 4F) was 2.76 fold that of C1-280-1-sense-sense (Figure 4E**)**, and 3.62 fold that of C1-280-1-sense-control (Figure 4G). These results show that combined sense-antisense L1 repetitive sequences located in the same plasmid activate the EGFP reporter gene in stable transfection (Figure 4J). There was no significant difference between the EGFP labeled amount of C1-280-1-sense-sense and C1-280-1-sense-control.

In order to further confirm the combined sense and antisense repetitive sequences activated the gene, 280-1×14as (14 copies of 280-1 in antisense orientation) was inserted into the pEGFP-C1 plasmid to construct C1-280-1-antisense-sense, C1-280-1-antisense-antisense and C1-280-1-antisense-control (Table 4). The expression of EGFP protein in stably transfected HeLa cells with these expression vectors were detected using a fluorescence microscope, and C1-280-1-antisense-sense induced the strongest EGFP expression (Figure 5A-5C).

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 2 November 2018 doi:10.20944/preprints201811.0066.v1

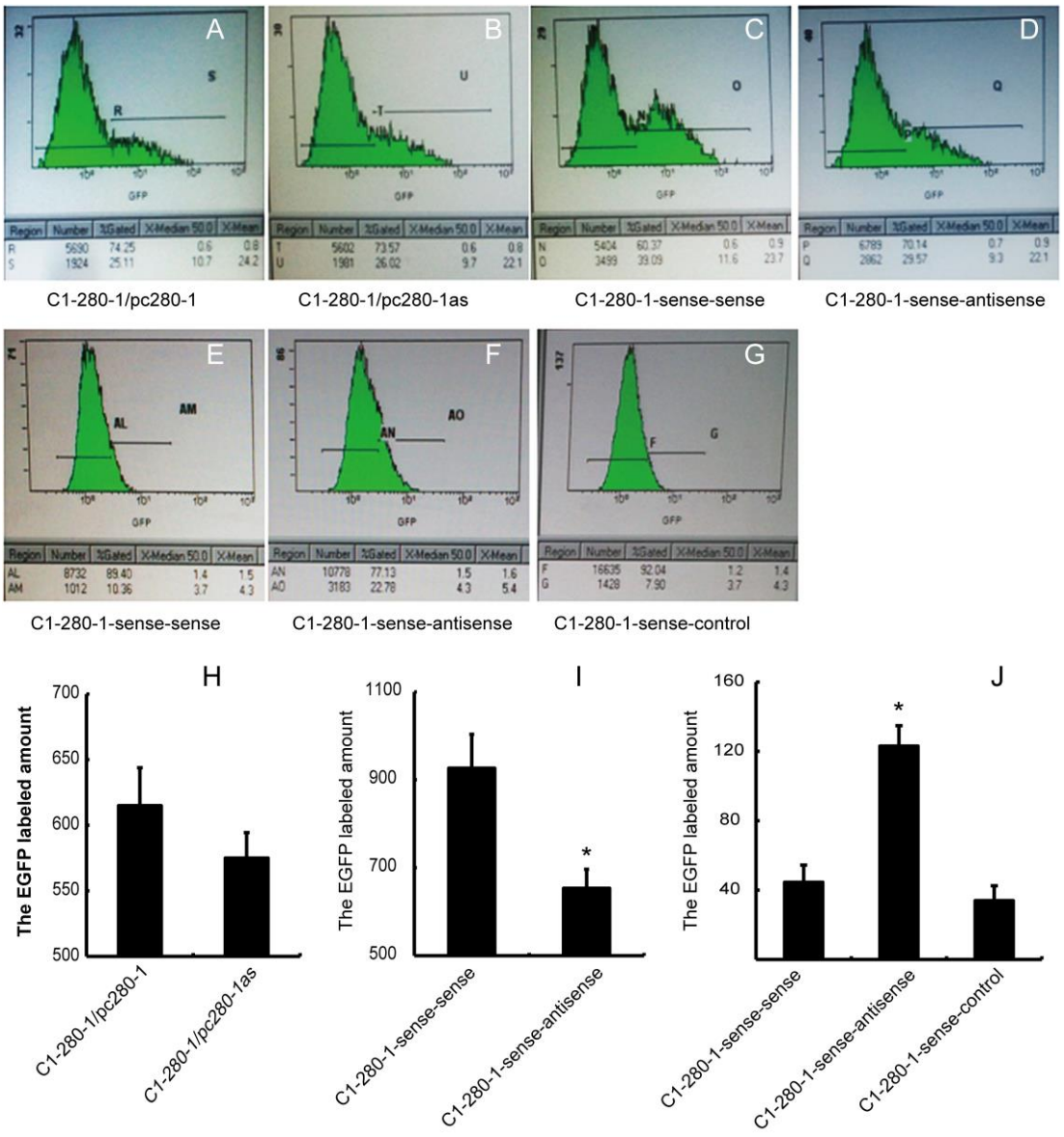


Figure 4. Combined sense-antisense 280-1 repetitive sequences activate the EGFP reporter gene with stable transfection, but not with transient transfection. (A-B) Representative EGFP gene expression induced by different 280-1 expression vectors were co-transfected into HeLa cells. EGFP expression was determined using a flow cytometer after co-transfection. (C-D) Representative EGFP gene expression induced by different 280-1 expression vectors were transiently transfected into HeLa cells. (E-G) Representative EGFP gene expression induced by different 280-1 expression vectors were stably transfected into HeLa cells. (H) The EGFP labeled amount induced by different expression vectors after co-transfection was plotted as mean ± SD values. (I) The EGFP labeled amount induced by different 280-1 expression vectors after transient transfection was plotted as mean ± SD values. The EGFP labeled amount of C1-280-1-sense-antisense was significant lower than that of C1-280-1-sense-sense (*$p<0.01$). (J) The EGFP labeled amount induced by different expression vectors after stable transfection was plotted as mean ± SD values. The EGFP labeled amount of C1-280-1-sense-antisense is significant higher than that of C1-280-1-sense-sense (*$p<0.01$), illustrating that combined sense-antisense 280-1 repetitive sequences activate the EGFP gene during stable transfection.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 2 November 2018 doi:10.20944/preprints201811.0066.v1

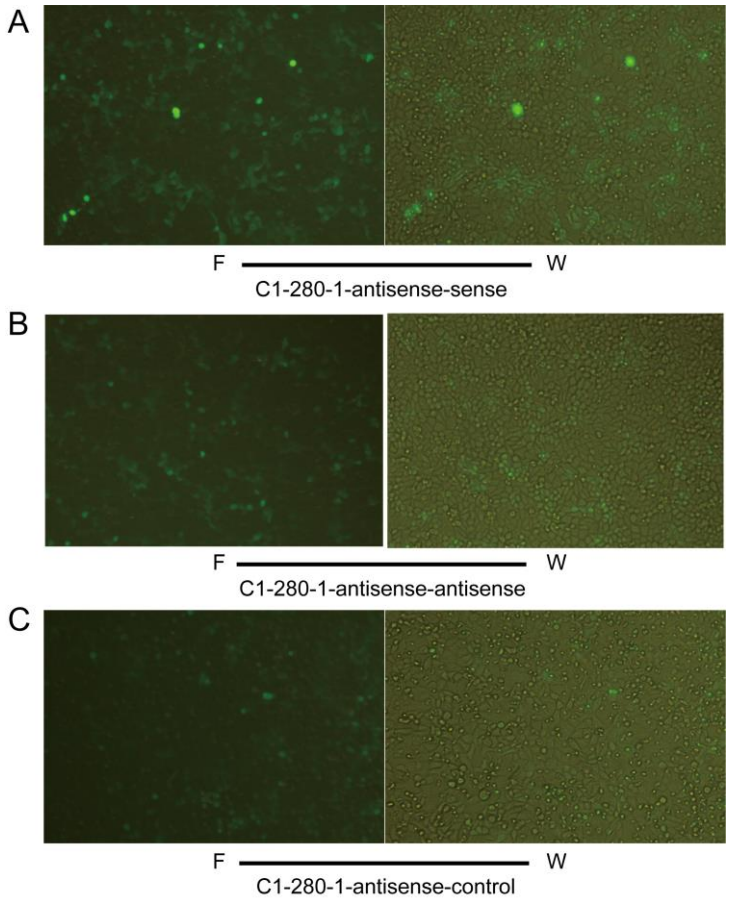


Figure 5. EGFP protein expression in HeLa cells stably transfected with 280-1 expression vectors were detected using a fluorescence microscope. (A-C) Representative images of EGFP protein expression in HeLa cells stably transfected with expression vectors. (W): white light; (F): fluorescent light.

2.2.2. Transcribed Alus activated EGFP reporter gene expression

To further prove whether the repetitive squences in human X chromosome involve in XCI, we also cloned Alu from human Xq13, the center of XCI, and constructed different expression vectors carrying combined sense and antisense repetitive sequences. Effects of sense or antisense Alu sequences on expression of EGFP reporter were examined in transiently and stably transfected HeLa cells, simulating the effects of repetitive sequences located on chromosomes.

Firstly, the effects of repetitive sequences locating in different plasmids on EGFP

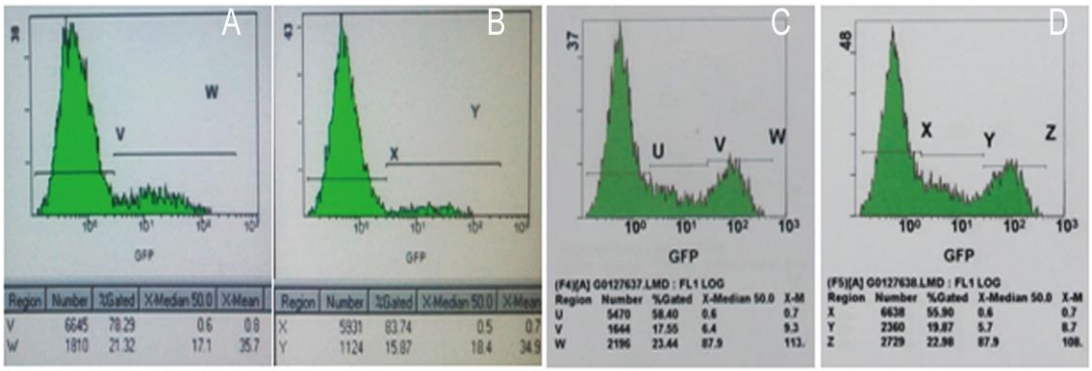
Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 2 November 2018 doi:10.20944/preprints201811.0066.v1

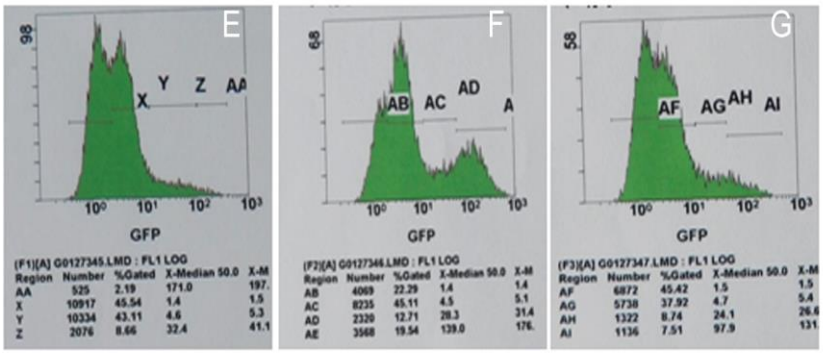
gene expression using co-transfection methods were observed. HeLa cells were co-transfected transiently with modulator plasmids carrying sense or antisense repetitive sequences (derived from pcDNA3.1), and with reporter plasmids carrying sense repetitive sequences (derived from pEGFP-C1). The EGFP labeled amount of cells co-transfected transiently with C1-Alu and pcAluas (Table 4, Figure 6B) was 0.73 fold that of co-transfected with C1-Alu and pcAlu (Table 4, Figure 6A), which revealed significant antisense inhibition ($P<0.05$, Figure 6H**)**.

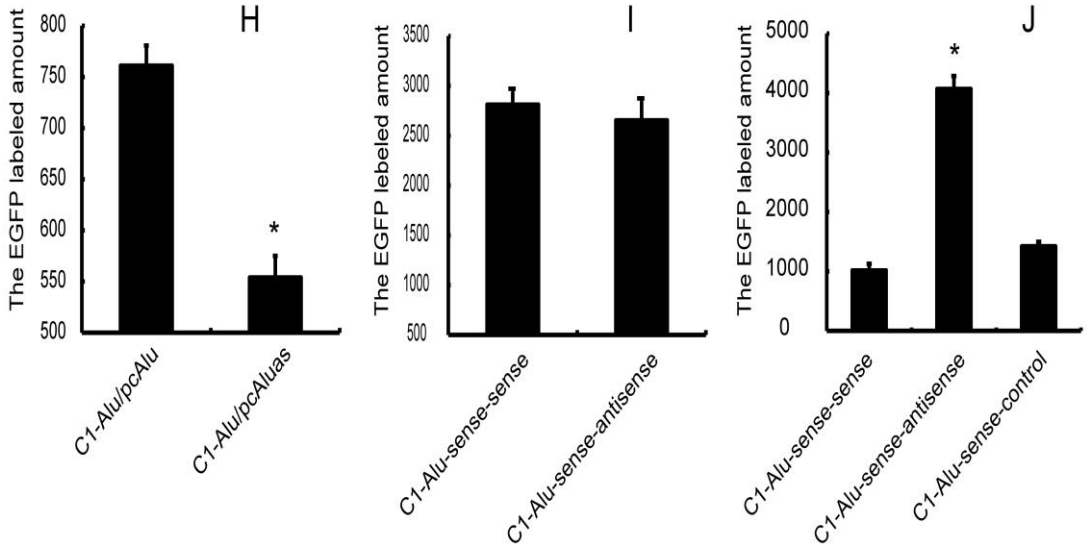
Next, the effects of repetitive sequences locating in the same plasmids on downstream EGFP gene expression in transient and stable transfected cells were compared. C1-Alu-sense-sense, C1-Alu-sense-antisense, C1-Alu-sense-control plasmids (Table 4) were transfected into HeLa cells. Then, expression of EGFP was determined using flow cytometry after transient or stable transfection. Following transient transfection, the EGFP-labeled amount of C1-Alu-sense-antisense (Figure 6D) was 0.94 fold that of C1-Alu-sense-sense (Figure 6C), which reveals a trend in antisense inhibition (no significant difference, Figure 6I).

The results of stable transfection are different from that of transient transfection, the EGFP labeled amount of C1-Alu-sense-antisense (Figure 6F) was 4.0 fold that of C1-Alu-sense-sense (Figure 6E) and 2.9 fold that of C1-Alu-sense-control (Figure 6G). These results show that combined sense-antisense repetitive sequences located in the same plasmid activate the EGFP reporter gene in stable transfection (Figure 6J). There was no significant difference between the EGFP labeled amount of C1-Alu-sense-sense and C1-Alu-sense-control.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 2 November 2018 doi:10.20944/preprints201811.0066.v1







Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 2 November 2018 doi:10.20944/preprints201811.0066.v1

Figure 6. Combined sense-antisense Alu repetitive sequences activate the EGFP reporter gene with stable transfection, but not with transient transfection. (A-B) Representative EGFP gene expression induced by different Alu expression vectors were co-transfected into HeLa cells. EGFP expression was determined using a flow cytometer after co-transfection. (C-D) Representative EGFP gene expression induced by different Alu expression vectors were transiently transfected into HeLa cells. (E-G) Representative EGFP gene expression induced by different Alu expression vectors were stably transfected into HeLa cells. (H) The EGFP labeled amount induced by different Alu expression vectors after co-transfection was plotted as mean ± SD values. The EGFP labeled amount of co-transfected with C1-Alu and pcAluas is significantly lower than that of C1-Alu and pcAlu (*$p<0.01$). (I) The EGFP labeled amount induced by different Alu expression vectors after transient transfection was plotted as mean ± SD values. The EGFP labeled amount of C1-Alu-sense-antisense was significant lower than that of C1-Alu-sense-sense (*$p<0.01$). (J) The EGFP labeled amount induced by different Alu expression vectors after stable transfection was plotted as mean ± SD values. The EGFP labeled amount of C1-Alu-sense-antisense is significant higher than that of C1-Alu-sense-sense (*$p<0.01$), illustrating that combined sense-antisense Alu repetitive sequences activate the EGFP gene during stable transfection.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 2 November 2018 doi:10.20944/preprints201811.0066.v1

3. Discussion

The previous studies have demonstrated that binding of RNAs and DNAs may play a role in activating gene expression, which is named RNA activation. We proposed that RNAs binding strength can be used to assess chromatin activity and the low RNA binding strength may be one of reasons of XCI. We analyzed the simulated RNA binding strengths to DNA segments within entire autosomes and X chromosome using bioinformatics methods. According to these analyses, we found that RNA binding strength to the X chromosome was significantly lower than that to autosomes (except chromosomes 8, 11, 18, 20, 21 and 22) (Table 1). Since Xq is almost completely inactive, we further compared RNA binding strength to Xq with that to autosomes, which revealed that the RNA binding strength of Xq is significantly lower than that to autosomes (except chromosomes 11, 20 and 22) (Table 1). Although the RNA binding strength to chromosomes 11, 20 and 22 is less than or near equal to that of Xq, the standard deviations of RNA binding strengths in these three chromosomes were significantly higher than that of Xq. After analyzing the fragments possessing upper half values of RNA binding strengths in Xq and autosomes, we found that the mean value of RNA binding strengths of these fragments in Xq was lower than that in all autosomes. Hence, we speculated that high mean values of RNA binding strengths and more fragments possessing high RNA binding strengths may always accompany with more active chromosome. Approximately 30% genes located on Xp escape from XCI while Xq almost entirely inactivated. The results of bioinformatic methods proved that RNA binding strength of Xq was significantly lower than that of Xp ($p<0.01$) (Figure 2, Table 1) and the RNA binding strength of the inactivation regions on Xp was significantly lower than that of activation regions ($p<0.01$) (Table 2). These finding provide new clues for uncovering the mechanism of genes escape from

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 2 November 2018 doi:10.20944/preprints201811.0066.v1

a roughly inactivated X chromosome. Based on our results, we proposed a hypothesis about XCI, i.e. the low RNA binding strength of human X chromosome may be one of reasons of XCI. Since the RNA binding strength of Xq is low, so the X chromosome may be first affected when the activating strength of the RNAs weakens. The low dispersion degree of the RNA binding strength of Xq implies that Xq lacks fragments with high RNA binding strength, which may easily causes silencing.

In this study, we constructed the modulator plasmids carrying L1s or Alus sequences and reporter plasmids carrying L1s or Alus, which were co-transfected into HeLa cells to study the effects of modulator plasmids on EGFP gene expression in reporter plasmids. Our findings revealed that antisense RNAs of repetitive sequences can inhibit the transcription of sense RNAs during transient transfection which was consistent with previous studies. Previous study has shown that antisense RNAs can inhibit the transcription of sense RNAs [45,46]. For instance, the transcriptions of the multidrug resistance genen (mdr1) [47], LacZ in Escherichia coli [48,49] and replicase in Agrobacterium tumefaciens were inhibited by their antisense RNAs [49]. Our results are consistent with the above reports.

The effects of sense or antisense repetitive sequences locating in same plasmids on gene expression of downstream EGFP were applied to simulate the effects of neighboring repetitive sequences in cis. Previous studies have indicated that more than 70% of mouse genome transcriptional units occur by way of antisense transcription and antisense knockdown can result in the reduction of sense transcripts [50,51] which implied that antisense RNAs can more commonly promote the transcription of sense RNAs within the genome. In our study, antisense RNAs were found to activate the transcription of sense RNAs in stably transfected cells which were consistent with previous studies [50, 51] . However, this phenomenon was not

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 2 November 2018 doi:10.20944/preprints201811.0066.v1

found in transiently transfected cells. Stable or transient transfection was depended on whether the exogenous DNA was integrated into the host genome. Nontranscribed repetitive sequences contribute to the silencing of genes, while transcribed repetitive sequences contribute to the activation of genes [52]. Hence, we concluded that gene activation induced by combined repetitive sequences may be associated with DNA replication. During DNA replication and transcription, the wrapping of DNA and histone was reduced [43], we hypothesize that naked DNA was likely to form a single stranded DNA that may bind with RNAs, and affect gene expression by RNA activation.

The dosage compensation strategies in female mammals are achieved by inactivating the entire chromosome and in *Drosphila* and *C. elegans* by decreasing chromosomal activity [6, 7, 9, 51] . The previous study has indicated that the inactivation of the X chromosome occurs through the cascade from the inactivation center [51]. X chromosome contains abundant repetitive sequences and accumulated evidences have indicated that repetitive sequences play crucial roles in XCI [22-24]. So we analyzed the distribution of repetitive sequences within 1 Mb sequence of the X-inactivation centre, we found a dominant distribution characteristic of repetitive sequences. The analyzed seven repetitive sequences (L1, Alu, MIR, L2, L3, LTR and DNA elements) were distributed in clusters, and there are alternative distributions of different repetitive sequence clusters in the 1 Mb (Figure 3). Peaks of other repetitive sequences were observed at the low ebb regions of L1 (Table 3). We propose that the gene activation induced by a repetitive sequence cluster needs the support of flanking repetitive sequences to form a cascade of gene activation. The inactivation of any one cluster can reduce the activity of flanking repetitive sequences, which lead to a cascade of gene inactivation. Based on above analyses, we proposed that the entire X

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 2 November 2018 doi:10.20944/preprints201811.0066.v1

chromosome inactivation or activation may be induced by the distribution characteristics of repetitive sequences in X chromosome.

In early embryonic cells, activation factors are abundant and inhibitory factors are less, which lead to both X chromosomes being activated in female cells. During cell differentiation, factors (i.e. RNA) that activate X chromosomes decreases; and thus, two X chromosomes compete for the same activation factors. As a result, a slight imbalance of role factors on one X chromosome may be amplified, and subsequently lead to the inactivation of the entire X chromosome due to cascade actions.

In this study, based on our experiments of gene activation induced by repetitive sequences and bioinformatics analysis results of RNA binding strengths, we proposed a model for XCI (illustrated in Figure 7). The main idea of this model is that the low RNA binding strength of human X chromosome may be the one of reasons of XCI and the cross distribution characteristics of different repetitive sequence clusters leading to a cascade of gene activation or gene inactivation may be the reason of transcriptional silencing one of the X chromosomes in female mammals. In Figure 7, “A”-“F” on the x-axis represent regions that contain different repetitive sequence clusters, while the y-axis represents RNA binding strength. The adjacent different repetitive sequence clusters appear cross distribution so that the activation of any cluster can lead to activation of its adjacent clusters and the inactivation of any cluster can lead to inactivation of its adjacent clusters, which forms the reason that leads to cascade of activation or inactivation of X chromosome. “F” represents a region possessing super-high RNA binding strength and “E” represents a region with low RNA binding strength. The inactivation of “E” region can not lead to the complete inactivation of “F” region so that the genes located in “F” region escape XCI, but the gene activation in “F” region is deceased by “E” region. Since existing the regions

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 2 November 2018 doi:10.20944/preprints201811.0066.v1

possessing super-high RNA binding strengths, this cascade activation or inactivation induced by repetitive sequences are not always the complete activation or inactivation, which may explain the escape from XCI, the activity of two X chromosomes in female Drosophila and C. elegans being half of that of the X chromosome in males and the XCI decrease following X chromosome and autosome translocation [23].

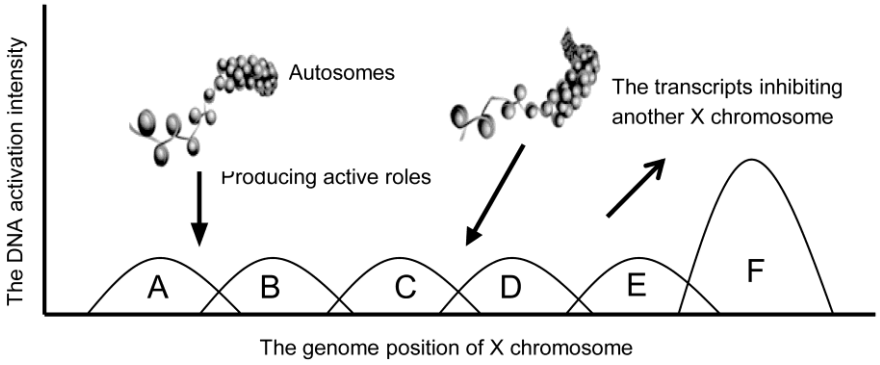


Figure 7. The repetitive sequence cascade model of XCI. The x-axis represents the genome position of X chromosome, while the y-axis represents DNA activation. "A", "B", "C", "D", "E" and "F" represent activation regions that are controlled by different repetitive sequences. For example, "A" represents the region of consisting of L1s, and "B" refers to the region of consisting of Alus. "A" and "B" partly present a cross distribution. The rest can be constituted in the same manner. The activation of region "A" can activate region "B", the activation of region "B" can activate region "A", and so on. The activation of each region needs support from neighboring regions, and the inactivation of one region would lead to cascade inactivation. "F" represents a region with a super-high activation, in which genes escape XCI. The neighboring region of "F" region, "E", display low DNA activation, so as to segregate cascade effects. If the activation of each region partly depends on neighboring regions, this would lead to a decreased activity of both X chromosomes. X activation need support from autosomes. The autosomes activate X chromosome when they locate in the same transcription sites with the X chromosome. The transcripts from one X chromosomes would inhibit another distant X chromosome.

X chromosome activation needs support from autosomes. The autosomes activate X chromosome when they locate in the same transcription sites with the X chromosome. The transcripts from one X chromosomes would inhibit another distant X chromosome. Both X chromosomes are active during the embryonic period, while one X chromosome randomly inactivates along with cell differentiation, which illustrates that XCI is associated with the genome environment of the X chromosome [11]. The observation that female diploid cells support the activation of one X

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 2 November 2018 doi:10.20944/preprints201811.0066.v1

chromosome, and female tetraploid cells support the activation of two X chromosomes also provides evidence for the relation between XCI and genome environment of the X chromosome[53,54]. The activation of the X chromosome needs RNAs produced from itself and RNAs produced from the genome. After early embryonic cells, 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 X chromosome. This may also explain why diploid cells support the activation of one X chromosome and why tetraploid cells support the activation of two X chromosomes.

Once the maternal or paternal X chromosomes are inactivated, they will remain inactivated in daughter cells, which is a well-known epigenetic process [55,56]. RNAs produced from repetitive sequence clusters on X chromosome activate the new duplicate DNAs of the X chromosome, which forms positive feedback that DNAs transcribe RNAs and RNAs in turn promote DNA transcription. This can explain the epigenetics of XCI. In addition, antisense RNAs produced from the active X chromosome may inhibit another X chromosome.

The model proposed in this study is reconcilable with other published mechanisms of XCI. Previous studies have reported some key mechanisms of XCI, such as XIST coating of the X-chromosome, recruitment of DNA, RNA and histone modification enzymes, and compaction and compartmentalization of the inactive X [57-61]. These studies suggest that increasing inhibitors and decreasing activators lead to XCI, and vice versa. In our model, RNAs and repetitive sequences are used as activators and inhibitors that participate in the activation or inactivation process of X chromosome.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 2 November 2018 doi:10.20944/preprints201811.0066.v1

4.Materials and Methods

4.1.Sequence data

Nucleotide sequences of 23 chromosomes were obtained from NCBI (the version of the genome is "build 33"; http://www.ncbi.nlm.nih.gov/genome/guide). A total of 1,000 genes which were 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 cells were involved DNA rapid replication, which could reflect the effects of epigenetic factors.

4.2.Software and algorithm

Gene-Analyzer 2.0 software was used to analyze the appearing numbers of 7-nucleotide (7-nt) strings which was written by our team (the software can be found at http://dx.doi.org/10.1016/j.biocel.2016.08.004) [42]. The numbers of 7-nt strings of a DNA fragments 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. The numbers of each 7-nt string within the DNA fragment were calculated. MS Excel software was used to perform t-test. Repetitive sequences were analyzed via RepeatMasker server (http://www.repeatmasker.org/).

4.3.RNA binding strength algorithm

In order to describe the possibility of RNAs transcribed in certain cells binding

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 2 November 2018 doi:10.20944/preprints201811.0066.v1

with DNAs, we introduced the concept of RNA binding strength. The RNA binding strength algorithm was calculated based on the principle that more complementary sequences of RNAs and DNAs result in greater binding between RNAs and DNAs [42]. For example, when there is one 5'-TTTTTTT DNA molecule and 10 5'-AAAAAAA RNA molecules in a certain volume solution, the possibility of DNA binding with RNAs is 10 (10×1=10). If there are 10 5'-TTTTTTT DNA molecules, the possibility of DNA binding with RNAs is 100 (10×10=100). DNA sequences from the indicated chromosomes were divided into many 50-kb segments and were re-coded as numbers of 7-nt strings using all possible 7-nt strings (4^7=16,384). We used the product of the number of the same 7-nt string in RNA and DNA fragment as the RNA binding strength of this 7-nt string. One-thousand genes highly expressed in human tonsil germinal center B cells were selected, and the 7-nt string numbers for these genes were calculated from the sense strand (including introns and exons). The 7-nt string numbers in one gene multiplied by the expression frequency of the gene (Lib.5601; http://www.ncbi.nlm.nih.gov/UniGene/) resulted in the calculated numbers of the 7-nt string for the gene. Table 6 illustrates the calculation method 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 5, C column). The simulated RNA binding strength (Table 5, E column and E16385) was defined as the sum of the products of the RNA and the corresponding 7-nt strings numbers within each 50-kb fragment. Therefore, the RNA binding strength represents a measure of the amount of RNA that can bind to each 50-kb region.

4.4.Analysis of the distribution of repetitive sequences in the center of XCI

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 2 November 2018 doi:10.20944/preprints201811.0066.v1

Xq13 is the X inactivation center. The RPS4X gene is located on Xq13. We analyzed repetitive sequences within 1 Mb (5'-3', 9810061 nt-10810060 nt) including the RPS4X gene and its 5' sequence, based on the direction of genomics. The 1-Mb sequence was divided into 2-kb fragments. These repetitive sequences were determined within each 2-kb fragment, and percentages of the base number of repetitive sequences that account for the 2-kb fragments were calculated. The Figure 3 displayed the distribution of repetitive sequences in the X inactivation center was drawn using sequence positions on the x-axis and percentages of the repetitive sequence bases on the y-axis.

4.5.Expression vector construction

The human Alu and L1 sequences were amplified by PCR from the RP11-29107 clone (Welcome Trust Sanger Institute, England), which is derived from human Xq13.1. The Alu or 280-1 (the first 280 bp of L1) sequences were inserted into pEGFP-C1 or pcDNA3.1 plasmids, as previously described [52] (Table 4). The LacZ2F2R sequence (the second 280 bp fragment of the LacZ gene) was amplified by PCR (forward primer: 5'-ATCGAAGCTTAATCTAGAACGGTTACGATGCGCCCATC; reverse primer: 5'-ATCGGGTACCATGCTAGCACCGCGAGGCGGTTTTCTCC) from Escherichia coli. The 4TMI (5'-GTGAAATAAATGCTTTTTTTGT) is a strong short enhancer that was identified in our previous studies [52,36]. The CMV promoter was amplified from pcDNA3.1+ (1-882 bp) and contained an intact CMV promoter.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 2 November 2018 doi:10.20944/preprints201811.0066.v1

4.6.Cell culture and cell transfection

HeLa cells were cultured in Dulbecco's modified Eagle medium (DMEM, Gibco, USA) with 10% fetal calf serum (FCS, Wisent Inc, Canada). Cells were placed in a 24-well plate at 1.5×10^5 cells/well and cultured for 24 hours at 37 °C with 5% CO_2. Cells were transfected with 0.4 μg of plasmid DNA using 2 μL of Lipofectamine™ 2000 (Invitrogen, USA), according to manufacturer's instructions. For transient transfection, cells were cultured for an additional 28 hours. For stable transfection, 10% of FCS-DMEM containing 1 mg/ml of G418 (Invitrogen, USA) was added to the cells, which were continually cultured for 28-35 days. For co-transfection, HeLa cells were transiently transfected with 0.4 μg of modulator plasmids (derived from pcDNA3.1) using 2 μl of Lipofectamine™ 2000. The cells were cultured for an additional 8-10 hours and were transfected with reporter plasmids (derived from pEGFP-C1). Following 23-25 hours after transfection with reporter plasmids, the transfected cells were used for determination of EGFP-positive cells with flow cytometry.

4.7.Determination of EGFP-positive cells with flow cytometry or fluorescence microscope

The transfected HeLa cells were collected in 1.5-ml Eppendorf tubes. Cells were centrifuged for 10 minutes at 2,000 rpm at 4°C, washed twice with PBS, and suspended in 0.7 ml of PBS. The expression of EGFP was determined using flow cytometry (Beckman Coulter, USA) [36] and displayed as an 'EGFP labeled amount'. EGFP labeled amount = (X mean of EGFP positive cells) × (percentage of EGFP

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 2 November 2018 doi:10.20944/preprints201811.0066.v1

positive cells), where X mean (mean value of fluorescence intensity in each cell) = ∑(X)/n [42]. At least three independent experiments were performed for all studies. Images are representative of at least three independent experiments. Group comparisons were made using mean ± S.D. and t-test. Statistical significances were defined as $p < 0.05$.

The expression of the EGFP protein was also assessed by inspecting HeLa cells stably transfected with expression vectors using a fluorescence microscope.

5.Conclusions

In conclusion, our study found that the repetitive sequences distribute in clusters in X chromosome; combination of sense and antisense repetitive sequences (L1 and Alu) can activate gene expression which may be affected by interactions between RNAs and DNAs; the RNA binding strength of Xq and its dispersion degree is the lowest in all human chromosomes and the regions with high RNA binding strengths in X chromosome are consistent with the regions that escape from XCI. Based on these findings, we proposed a new model of XCI that the low RNA binding strength of human X chromosome may be one of reasons of XCI and the cross distribution characteristics of different repetitive sequence clusters leading to a cascade of gene activation or gene inactivation may be the reason of transcriptional silencing one of the X chromosomes in female mammals. These findings provide new clues for the functions of repetitive sequences and the mechanism of XCI. However, further research was needed to explore the detailed functions of repetitive sequences in XCI.

Author Contributions: X.W. and Z.L conceived and designed the experiments. N.J., L.Y., Z.S., S.Y. and H.C. performed the experiments. N.J., X.W., S.L. and C.L.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 2 November 2018 doi:10.20944/preprints201811.0066.v1

analyzed the data. N.J. and Z.L. contributed to the analysis tools. X.W. and Z.L. written manuscript.

Acknowledgements: We thanked Linzhe Zhang for the technical help. This work was supported by the National Natural Sciences Foundation of China (81771499 to X.W.), the Hebei Province Natural Science Foundation of China (H2018206099 to X.W.).

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Hui, L. Quantifying the effects of aging and urbanization on major gastrointestinal diseases to guide preventative strategies. *BMC Gastroenterol*. 2018, 18, 145.
2. Brunet, A.; Berger, S.L. Epigenetics of aging and aging-related disease. *J. Gerontol. A Biol. Sci. Med. Sci*. 2014, 69, S17-S20.
3. Disteche, C.M.; Berletch, J.B. X-chromosome inactivation and escape. *J. Genet*. 2015, 94, 591-599.
4. Yin, X.; Latif, R.; Tomer, Y.; Davies, T.F. Thyroid epigenetics: X chromosome inactivation in patients with autoimmune thyroid disease. *Ann N Y Acad Sci*. 2007, 1110, 193-200.
5. Li, C.; Tian, H.; Chiu-Ho, W.; Heather, K.; Sun, S.; Nie, Q. A self-enhanced transport mechanism through long noncoding RNAs for X chromosome inactivation. *Sci. Rep.* 2016, 6, 31517.
6. Ercan, S.; Dick, L.L.; Lieb, J.D. The C. elegans Dosage Compensation Complex

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 2 November 2018 doi:10.20944/preprints201811.0066.v1

Propagates Dynamically and Independently of X Chromosome Sequence. *Curr. Biol*. 2009, 19,1777-1787.

7. Lucchesi, J.C.; Kelly, W.G.; Panning, B. Chromatin remodeling in dosage compensation. *Annu. Rev. Genet*. 2005, 39, 615-651.
8. Meyer, B.J. Sex in the wormcounting and compensating X-chromosome dose. *Trends Genet*. 2000, 16, 247-253.
9. Meyer, B.J.; Casson, L.P. Caenorhabditis elegans compensates for the difference in X chromosome dosage between the sexes by regulating transcript levels. *Cell* 1986, 47, 871-881.
10. Yue, M.; Richard, J.L.C.; Ogawa, Y. Dynamic interplay and function of multiple noncoding genes governing X chromosome inactivation. *Biochim. Biophys. Acta*. 2016, 1859, 112-120.
11. Chow, J.; Heard, E. X inactivation and the complexities of silencing a sex chromosome. *Curr. Opin. Cell Biol*. 2009, 21, 359-366.
12. Del Rosario, B.C.; Del Rosario, A.M.; Anselmo, A.; Wang, P.I.; Sadreyev, R.I.; Lee, J.T. Genetic Intersection of Tsix and Hedgehog Signaling during the Initiation of X-Chromosome Inactivation. *Dev. Cell* 2017, 43, 359-371.
13. Loos, F.; Maduro, C.; Loda, A.; Lehmann, J.; Kremers, G.J.; Ten Berge, D.; Grootegoed, J.A.; Gribnau, J. Xist and Tsix Transcription Dynamics Is Regulated by the X-to-Autosome Ratio and Semistable Transcriptional States. *Mol. Cell Biol*. 2016, 36, 2656-2667.
14. Penny, G.D.; Kay, G.F.; Sheardown, S.A.; Rastan, S.; Brockdorff, N. Requirement for Xist in X chromosome inactivation. *Nature* 1996, 379,131-137.
15. Escamilla-Del-Arenal, M.; Rocha, S.T.D.; Heard, E. Evolutionary diversity and developmental regulation of X-chromosome inactivation. *Hum. Genet*. 2011,

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 2 November 2018 doi:10.20944/preprints201811.0066.v1

130, 307-327.

16. Consortium, T.E.P. An integrated encyclopedia of DNA elements in the human genome. *Nature* 2012, 489, 57-74.
17. Venter, J.C.; Adams, M.D.; Myers, E.W.; Li, P.W.; Mural, R.J.; Sutton, G.G.; Smith, H.O.; Yandell, M.; Evans, C.A.; Holt, R.A.; et al. The sequence of the human genome. *Science* 2001, 291, 1304-1351.
18. Abrusán, G.; Giordano, J.; Warburton, P.E. Analysis of transposon interruptions suggests selection for L1 elements on the X chromosome. *PLoS Genet*. 2008, 4, e1000172.
19. Baker, R.J.; Wichman, H.A. Retrotransposon mys is concentrated on the sex chromosomes: implications for copy number containment. *Evolution* 1990, 44, 2083.
20. Waterston, R.H.; Lindblad-Toh, K.; Birney, E.; Rogers, J.; Abril, J.F.; Agarwal, P.; Agarwala, R.; Ainscough, R.; Alexandersson, M.; An, P.; et al. Initial sequencing and comparative analysis of the mouse genome. *Nature* 2002, 420, 520-562.
21. Chow, J.C.; Ciaudo, C.; Fazzari, M.J.; Mise, N.; Servant, N.; Glass, J.L.; Attreed, M.; Avner, P.; Wutz, A.; Barillot, E.; et al. LINE-1 Activity in Facultative Heterochromatin Formation during X Chromosome Inactivation. *Cell* 2016, 166, 782.
22. Bailey, J.A.; Carrel, L.; Chakravarti, A.; Eichler, E.E. Molecular evidence for a relationship between LINE-1 elements and X chromosome inactivation: the Lyon repeat hypothesis. *Proc. Natl. Acad. Sci. U S A*. 2000, 97, 6634-6639.
23. Cotton, A.M.; Chen, C.Y.; Lam, L.L.; Wasserman, W.W.; Kobor, M.S.; Brown, C.J. Spread of X-chromosome inactivation into autosomal sequences: role for

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 2 November 2018 doi:10.20944/preprints201811.0066.v1

DNA elements, chromatin features and chromosomal domains. *Hum. Mol. Genet*. 2014, 23,1211-1223.

24. Wang, Z.; Willard, H.F.; Sayan, M.; Furey, T.S. Evidence of influence of genomic DNA sequence on human X chromosome inactivation. *PLoS Comput. Biol.* 2006, 2, e113.
25. Parish, D.A.; Vise, P.; Wichman, H.A.; Bull, J.J.; Baker, R.J. Distribution of LINEs and other repetitive elements in the karyotype of the bat Carollia: implications for X-chromosome inactivation. *Cytogenet. Genome Res*. 2002, 96,191-197.
26. Carrel, L.; Willard, H.F. X-inactivation profile reveals extensive variability in X-linked gene expression in females. *Nature* 2005, 434,400-404.
27. Chong, S.; Kontaraki, J.; Bonifer, C.; Riggs, A.D. A Functional chromatin domain does not resist X chromosome inactivation: silencing of cLys correlates with methylation of a dual promoter-replication origin. *Mol. Cell. Biol*. 2002, 22, 4667-4676.
28. Bala, T.N.; Brahmachary, M.; Garg, P.; Borel, C.; Alnefaie, R.; Watson, C.T.; Thomas, N.S.; Sharp, A.J. DNA methylation profiling in X;autosome translocations supports a role for L1 repeats in the spread of X chromosome inactivation. *Hum. Mol. Genet.* 2014, 23, 1224-1236.
29. Chow, J.C.; Ciaudo, C.; Fazzari, M.J.; Mise, N.; Servant, N.; Glass, J.L.; Attreed, M.; Avner, P.; Wutz, A.; Barillot, E.; et al. LINE-1 activity in facultative heterochromatin formation during X chromosome inactivation. *Cell* 2010, 141, 956-969.
30. Kvikstad, E.M.; Makova, K.D. The (r)evolution of SINE versus LINE distributions in primate genomes: sex chromosomes are important. *Genome Res*.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 2 November 2018 doi:10.20944/preprints201811.0066.v1

2010, 20, 600-613.

31. Lee, J.T. Regulation of X-chromosome counting by Tsix and Xite sequences. *Science* 2005, 309, 768-771.
32. Lee, J.T.; Lum, N. Targeted mutagenesis of Tsix leads to nonrandom X inactivation. *Cell* 1999, 99, 47-57.
33. Marahrens, Y.; Loring, J.; Jaenisch, R. Role of the Xist gene in X chromosome choosing. *Cell* 1998, 92, 657-664.
34. Nicodemi, M.; Prisco, A. Symmetry-breaking model for X-chromosome inactivation. *Phys. Rev. Lett*. 2007, 98, 108104.
35. Wang, X.F.; Jin, X.; Wang, X.; Liu, J.; Feng, J.; Yang, Q.; Mu, W.; Shi, X.; Lu, Z. Effects of L1-ORF2 fragments on green fluorescent protein gene expression. *Genet Mol Biol*. 2009, 32, 688-696.
36. Wang, H.G.; Wang, X.F.; Jing, X.Y.; Li, Z.; Zhang, Y.; Lv, Z.J. Effect of mutations in a simian virus 40 PolyA signal enhancer on green fluorescent protein reporter gene expression. *Genet. Mol. Res*. 2011, 10, 1866-1883.
37. Huang, V.; Qin, Y.; Wang, J.; Wang, X.; Place, R.F.; Lin, G.; Lue, T.F.; Li, L.C. RNAa is conserved in mammalian cells. *PloS ONE* 2010, 5:e8848.
38. Janowski, B.A.; Younger, S.T.; Hardy, D.B.; Ram, R.; Huffman, K.E.; Corey, D.R. Activating gene expression in mammalian cells with promoter-targeted duplex RNAs. *Nat. Chem. Biol*. 2007, 3, 166-173.
39. LC, L.; Okino, S.T.; Zhao, H.; Pookot, D.; Place, R.F.; Urakami, S.; Enokida, H.; Dahiya, R. Small dsRNAs induce transcriptional activation in human cells. *Proc. Natl. Acad. Sci. U S A*. 2006,103, 17337-17342.
40. Matsui, M.; Sakurai, F.; Elbashir, S.; Foster, D.J.; Manoharan, M.; Corey, D.R. Activation of LDL receptor expression by small RNAs complementary to a

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 2 November 2018 doi:10.20944/preprints201811.0066.v1

noncoding transcript that overlaps the LDLR promoter. *Chem. Biol.* 2010,17, 1344-1355.

41. Zheng, L.; Wang, L.; Gan, J.; Zhang, H. RNA activation: promise as a new weapon against cancer. *Cancer Lett*. 2014, 355,18-24.
42. Wang, X.; Ma, Z.; Kong, X.; Lv, Z. Effects of RNAs on chromatin accessibility and gene expression suggest RNA-mediated activation. *Int. J. Biochem. Cell Biol*. 2016, 79, 24-32.
43. 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*. 2014, 7, 948-958.
44. Carrel, L.; Cottle, A.A.; Goglin, K.C.; Willard, H.F. A first-generation X-inactivation profile of the human X chromosome. *Proc. Natl. Acad. Sci. U S A.* 1999, 96, 14440-14444.
45. Khazenzon, N.M.; Ljubimov, A.V.; Lakhter, A.J.; Fujita, M.; Fujiwara, H.; Sekiguchi, K.; Sorokin, L.M.; Petäjäniemi, N.; Virtanen, I.; Black, K.L.; et al. Antisense inhibition of laminin-8 expression reduces invasion of human gliomas in vitro. *Mol. Cancer Ther*. 2003, 15, 985-994.
46. Yamanaka, Y.; Faghihi, M.A.; Magistri, M.; Alvarez-Garcia, O.; Lotz, M.; Wahlestedt, C. Antisense RNA controls LRP1 Sense transcript expression through interaction with a chromatin-associated protein, HMGB2. *Cell Rep*. 2015,11, 967-976.
47. Chan, Y.W.; Chu, C.Y.; Fung, K.P. Inhibition of P-glycoprotein expression and reversal of drug resistance of human hepatoma HepG2 cells by multidrug resistance gene (mdr1) antisense RNA. *Life Sci*. 2000, 67, 2117-2124.
48. Brantl, S.; Wagner, E.G. An antisense RNA-mediated transcriptional attenuation

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 2 November 2018 doi:10.20944/preprints201811.0066.v1

mechanism functions in Escherichia coli. *J. Bacteriol*. 2002, 184, 2740-2747.

49. Weaver, K.E. Emerging plasmid-encoded antisense RNA regulated systems. *Curr. Opin. Microbiol.* 2007, 10, 110-116.
50. Knowling, S.; Morris, K.V. Non-coding RNA and antisense RNA. Nature's trash or treasure? *Biochimie* 2011, 93, 1922-1927.
51. Tachibana, M.; Ma, H.; Sparman, M.L.; Lee, H.S.; Ramsey, C.M.; Woodward, J.S.; Sritanaudomchai, H.; Masterson, K.R.; Wolff, E.E.; Jia, Y.; et al. X-chromosome inactivation in monkey embryos and pluripotent stem cells. *Dev. Biol*. 2012, 371,146-155.
52. Ma, Z.; Kong, X.; Liu, S.; Yin, S.; Zhao, Y.; Liu, C.; Lv, Z.; Wang, X. Combined sense-antisense Alu elements activate the EGFP reporter gene when stable transfection. *Mol Genet Genomics*. 2017, 292, 833-846.
53. Monkhorst, K.; De, H.B.; Jonkers, I.; Mulugeta, A.E.; Monkhorst, W.; Hoogerbrugge, J.; Rentmeester, E.; Westerhoff, H.V.; Grosveld, F.; Grootegoed, J.A.; et al. The probability to initiate X chromosome inactivation is determined by the X to autosomal ratio and X chromosome specific allelic properties. *PloS ONE* 2009, 4, e5616.
54. van Bemmel, J.G.; Mira-Bontenbal, H.; Gribnau, J. Cis- and trans-regulation in X inactivation. *Chromosoma*. 2016,125, 41-50.
55. Gendrel, A.V.; Heard, E. Noncoding RNAs and epigenetic mechanisms during X-chromosome inactivation. *Annu. Rev. Cell Dev. Biol.* 2014, 30, 561-580.
56. Lu, Z.; Carter, A.C.; Chang, H.Y. Mechanistic insights in X-chromosome inactivation. *Philos Trans R Soc Lond B Biol Sci*. 2017, 372,1733.
57. Jeon, Y.; Lee, J.T. YY1 tethers Xist RNA to the inactive X nucleation center. Cell 2011, 146, 119-133.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 2 November 2018 doi:10.20944/preprints201811.0066.v1

58. Makhlouf, M.; Ouimette, J.F.; Oldfield, A.; Navarro, P.; Neuillet, D.; Rougeulle, C. A prominent and conserved role for YY1 in Xist transcriptional activation. *Nat. Commun*. 2014, 5, 4878.

59. Hasegawa, Y.; Brockdorff, N.; Kawano, S.; Tsutui, K.; Nakagawa, S. The matrix protein hnRNP U is required for chromosomal localization of Xist RNA. *Dev. Cell* 2010, 19, 469-476.

60. Kolpa, H.J.; Fackelmayer, F.O.; Lawrence, J.B. SAF-A requirement in anchoring XIST RNA to chromatin varies in transformed and primary cells. *Dev. Cell* 2016, 39, 9-10.

61. Sakaguchi, T.; Hasegawa, Y.; Brockdorff, N.; Tsutsui, K.; Tsutsui, K.M.; Sado, T.; Nakagawa, S. Control of chromosomal localization of Xist by hnRNP U family molecules. *Dev. Cell* 2016, 39, 11-12.