OPEN ACCESS Freely available online

PLOS ONE

5^{me}CpG Epigenetic Marks Neighboring a Primate-Conserved Core Promoter Short Tandem Repeat Indicate X-Chromosome Inactivation

Filipe Brum Machado[1], Fabricio Brum Machado[2], Milena Amendro Faria[2], Viviane Lamim Lovatel[2], Antonio Francisco Alves da Silva[2,7], Claudia Pamela Radic[3], Carlos Daniel De Brasi[3], Álvaro Fabricio Lopes Rios[2], Susana Marina Chuva de Sousa Lopes[4], Leonardo Serafim da Silveira[5], Carlos Ramon Ruiz-Miranda[6], Ester Silveira Ramos[1]*, Enrique Medina-Acosta[2,7]*

1 Department of Genetics, School of Medicine, University of São Paulo, Ribeirão Preto, São Paulo, Brazil, **2** Laboratory of Biotechnology, Universidade Estadual do Norte Fluminense Darcy Ribeiro, Campos do Goytacazes, Rio de Janeiro, Brazil, **3** Laboratory of Molecular Genetics of Hemophilia, Institute of Experimental Medicine, National Academy of Medicine, Buenos Aires, Argentina, **4** Department of Anatomy and Embryology, Leiden University Medical Center, Leiden, South Holland, the Netherlands, **5** Laboratory of Animal Morphology and Pathology, Center for Studies and Research in Wildlife, Universidade Estadual do Norte Fluminense Darcy Ribeiro, Campos do Goytacazes, Rio de Janeiro, Brazil, **6** Laboratory of Environmental Sciences, Sector of Studies of Ethology, Reintroduction and Conservation of Wild Animals, Universidade Estadual do Norte Fluminense Darcy Ribeiro, Campos do Goytacazes, Rio de Janeiro, Brazil, **7** Molecular Identification and Diagnostics Unit, Hospital Escola Álvaro Alvim, Campos dos Goytacazes, Rio de Janeiro, Brazil

Abstract

X-chromosome inactivation (XCI) is the epigenetic transcriptional silencing of an X-chromosome during the early stages of embryonic development in female eutherian mammals. XCI assures monoallelic expression in each cell and compensation for dosage-sensitive X-linked genes between females (XX) and males (XY). DNA methylation at the carbon-5 position of the cytosine pyrimidine ring in the context of a CpG dinucleotide sequence (5^{me}CpG) in promoter regions is a key epigenetic marker for transcriptional gene silencing. Using computational analysis, we revealed an extragenic tandem GAAA repeat 230-bp from the landmark CpG island of the human X-linked retinitis pigmentosa 2 *RP2* promoter whose 5^{me}CpG status correlates with XCI. We used this *RP2* onshore tandem GAAA repeat to develop an allele-specific 5^{me}CpG-based PCR assay that is highly concordant with the human androgen receptor (*AR*) exonic tandem CAG repeat-based standard HUMARA assay in discriminating active (Xa) from inactive (Xi) X-chromosomes. The *RP2* onshore tandem GAAA repeat contains neutral features that are lacking in the *AR* disease-linked tandem CAG repeat, is highly polymorphic (heterozygosity rates approximately 0.8) and shows minimal variation in the Xa/Xi ratio. The combined informativeness of *RP2/AR* is approximately 0.97, and this assay excels at determining the 5^{me}CpG status of alleles at the Xp (*RP2*) and Xq (*AR*) chromosome arms in a single reaction. These findings are relevant and directly translatable to nonhuman primate models of XCI in which the *AR* CAG-repeat is monomorphic. We conducted the *RP2* onshore tandem GAAA repeat assay in the naturally occurring chimeric New World monkey marmoset (*Callitrichidae*) and found it to be informative. The *RP2* onshore tandem GAAA repeat will facilitate studies on the variable phenotypic expression of dominant and recessive X-linked diseases, epigenetic changes in twins, the physiology of aging hematopoiesis, the pathogenesis of age-related hematopoietic malignancies and the clonality of cancers in human and nonhuman primates.

Citation: Machado FB, Machado FB, Faria MA, Lovatel VL, Alves da Silva AF, et al. (2014) 5^{me}CpG Epigenetic Marks Neighboring a Primate-Conserved Core Promoter Short Tandem Repeat Indicate X-Chromosome Inactivation. PLoS ONE 9(7): e103714. doi:10.1371/journal.pone.0103714

Editor: Osman El-Maarri, University of Bonn, Institut of experimental hematology and transfusion medicine, Germany

Received January 22, 2014; **Accepted** July 4, 2014; **Published** July 31, 2014

Funding: The research leading to these results has received funding from FAPERJ (http://www.faperj.br), CNPq (http://www.cnpq.br), FAPESP (http://www.fapesp.br/en/), CAPES-NUFFIC (http://www.capes.gov.br/cooperacao-internacional/holanda/programa-capesnuffic), under grant agreements 302731/2009-1, E-26/111.450/2012, E- 111.398/2013, E-26/111.788/2013, E-26/110.035/201, and graduate fellowships by FAPESP(FiBM), FAPERJ (FaBM and MAF), CAPES (VLL) and CNPq (AFAS). The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Competing Interests: The authors have declared that no competing interests exist.

* Email: esramos@fmrp.usp.br (ESR); quique@uenf.br (EM-A)

Introduction

In eukaryotes, the CpG dinucleotide sequence is distributed sparsely but genome-wide, except in distinct regions termed CpG islands (CGI), in which its density is increased approximately five-fold; these regions generally correspond to promoters [1]. Depending on the methylation state of the carbon-5 position of the cytosine residue, the self-complimentary CpG dinucleotide functions as a genomic signaling sequence for the recruitment of either repressive or permissive histone modification marks, which modulate the chromatin structure into mutually exclusive transcriptionally inactive (silenced) or active configurations, respectively [2]. With the exception of the sites in active promoter regions, nearly 80% of CpG sites in the mammalian genome are in the 5^{me}CpG state in somatic cells [2]. Thus, transcriptional silencing correlates positively with the maintenance (in frequency and breadth) of 5^{me}CpG in promoter regions.

Gene silencing based on 5^{me}CpG marks underlies key cellular processes such as cellular differentiation, cell-, tissue- and

embryonic developmental stage-specific gene expression, preservation of chromatin structure and chromosomal integrity, aging of the hematopoietic system, carcinogenesis, random autosomal monoallelic gene expression, parent-of-origin-dependent monoallelic gene expression (genomic imprinting) and X-chromosome inactivation (XCI) [3].

XCI is the stable, (nearly) chromosome-wide transcriptional silencing of either the maternal (^{M}X) or the paternal (^{P}X) X-chromosome in the inner cell mass of female eutherian mammals [4]. XCI entails selecting (normally at random), targeting and driving either ^{M}X or ^{P}X in each early stage embryonic female cell into a facultative heterochromatin configuration of sustained transcriptional gene suppression [5,6].

Overall, XCI ensures monoallelic gene expression in each cell and compensation for dosage-sensitive X-linked genes between females (XX) and males (XY) [7]. In human females, there is extensive variability in X-linked gene expression, with approximately 15% of genes resisting XCI and being expressed from both active X (Xa) and inactive X (Xi) chromosomes and an additional 10% being expressed to varying degrees from some Xi chromosomes [8]. Thus, while most genes on Xi are stably silenced, a discrete yet significant subset of genes escape transcriptional suppression by being excluded from the condensed heterochromatic body of Xi [9]. Escape genes (e.g., active genes on Xi) may exhibit tissue-specific differences in the escape from inactivation [10]. Escape genes have distinct evolutionary implications for sex differences in specific phenotypes [10,11].

The 5^{me}CpG-sensitive restriction endonuclease-based PCR assay targeting the polymorphic trinucleotide tandem CAG repeat (microsatellite, short tandem repeat - STR) in exon 1 of the human androgen receptor (*AR*) gene (MIM 313700) in the Xq12 region, known as the HUMARA assay, is a standard readout method for determining the methylation statuses of alleles on Xa and Xi and is widely used as a marker of X-chromosome activity [12]. The *AR* tandem CAG repeat yields heterozygosity rates of approximately 0.85 worldwide, and it is therefore uninformative in a significant proportion of females. The *AR* tandem CAG repeat genotype is not neutral, with threshold numbers of repeat units being positive and negatively correlated with Kennedy disease (KD [MIM 313200]) [13] and prostate cancer [14,15], respectively. Moreover, the *AR* CAG-repeat locus is monomorphic in the small nonhuman primate species used in biomedical research [16], which precludes its use in studies of XCI in these important experimental models.

We sought to identify X-linked repeats that are conserved in primates and consist of neutral features to accurately assess the methylation statuses of alleles in Xa and Xi. We aimed to develop a method that is highly concordant with the *AR* disease-linked tandem CAG repeat assay, but with minimal ^{M}X/^{P}X variation due to lesser *in vitro* replication slippage by Taq polymerase across repeat units greater than triplets. This goal has not been realized to date in either humans or nonhuman primate species.

Materials and Methods

Ethics Statement

Samples from human subjects were collected with written informed consent for projects approved by the Ethics Committee of the Faculdade de Medicina de Campos, Brazil (approval code FR-278769); Leiden University Medical Center, the Netherlands (P08.087); Faculdade de Medicina de Ribeirão Preto, Brazil (HCRP 5810/2009); and Institutos de la Academia Nacional de Medicina, Argentina (14/08/2008). The capture of individual marmosets (wild hybrids of *Callithrix jacchus* and *Callithrix penicillata*), confinement in a captive colony, management, care, drawing of biological samples and necropsies were all carried out under authorizations from the Brazilian Chico Mendes Institute for the Conservation of Biodiversity – ICMBio (URL: http://www.icmbio.gov.br/portal/) with license #33965-2 and the Brazilian Institute of the Environment and Renewable Natural Resources - IBAMA (URL: http://www.ibama.gov.br/) with license CGEF AM3301.8101/2013-RJ. The marmoset specimens were taken into captivity in strict accordance with the recommendations of ICMBio as part of a control program for these invasive species. They were previously introduced into an industrial zone belonging to the Brazilian Oil company TRANSPETRO, located in the State of Rio de Janeiro, inhabited by the endangered, native golden lion tamarins (*Leontopithecus rosalia*). The program was licensed by ICMBio and IBAMA because the presence of the marmosets increases the risk of extinction of golden lion tamarins by exposing them to transmissible infectious diseases, predation or limiting-resource competition. The captive colony was founded in the Sector of Studies on the Ethology, Reintroduction and Conservation of Wild Animals (SERCAS, website URL: http://uenf.br/cbb/sercas/) of the Universidade Estadual do Norte Fluminense Darcy Ribeiro, Brazil, as a model for management. Animal management activities were supervised by an IBAMA-licensed, expert investigator (CRRM). The capture, clinical and laboratory examinations and handling of animals were conducted essentially as previously reported [17]. No marmoset specimen was euthanized to obtain tissue for this study. Marmoset peripheral blood samples (50 μL) were drawn into EDTA during routine examination of confined animals. Samples (3–5 mm^3) of muscle, liver, brain and skin/hair tissues were strictly taken from the frozen remains of necropsies carried out by a licensed veterinarian (LSS) that were exclusively performed on specimens that died of natural causes during the process of adapting to confinement including failure to thrive, wasting syndrome and/or nematode infestation. Care was taken to alleviate suffering, and measures were implemented according to IBAMA guidelines for the well-being of wildlife and the recommendations of the Guide for the Care and Use of Laboratory Animals of the Universidade Estadual do Norte Fluminense Darcy Ribeiro, Brazil.

Subjects

To determine heterozygosity rates and allele frequencies, we genotyped two population subsets, each consisting of sixty healthy, unrelated women from Brazil and the Netherlands. To analyze the correlations between random or non-random X-inactivation patterns and the *RP2*-extragenic GAAA repeat or the *AR* exonic CAG repeat (HUMARA assay), we genotyped a third subset of fifty unrelated women who had known HUMARA-based methylation profiles (e.g., Xa/Xi ratios). We genotyped four healthy male donors as a control for methylation-sensitive restriction enzyme activity. To demonstrate the power of *RP2*-extragenic GAAA repeats in discriminating Xa from Xi in heterozygous female carriers of an X-linked recessive defect that manifests due to non-random (skewed) X-inactivation, we genotyped four confirmed heterozygous carriers of hemophilia A. Two of these individuals were conventional, non-symptomatic carriers who screened positive for *F8* intron 22 inversions via inverse shifting-PCR [18] and for random X-inactivation via the *AR* CAG repeat assay [12]. The other two were heterozygous carriers of missense and frameshift mutations in factor VIII domains A1 and B, respectively. They were screened through conformational sensitive gel electrophoresis [19] and direct sequencing and presented with a severe hemophilia A phenotype due to extremely skewed XCI. For the assessment of marmosets, we genotyped necropsy tissues

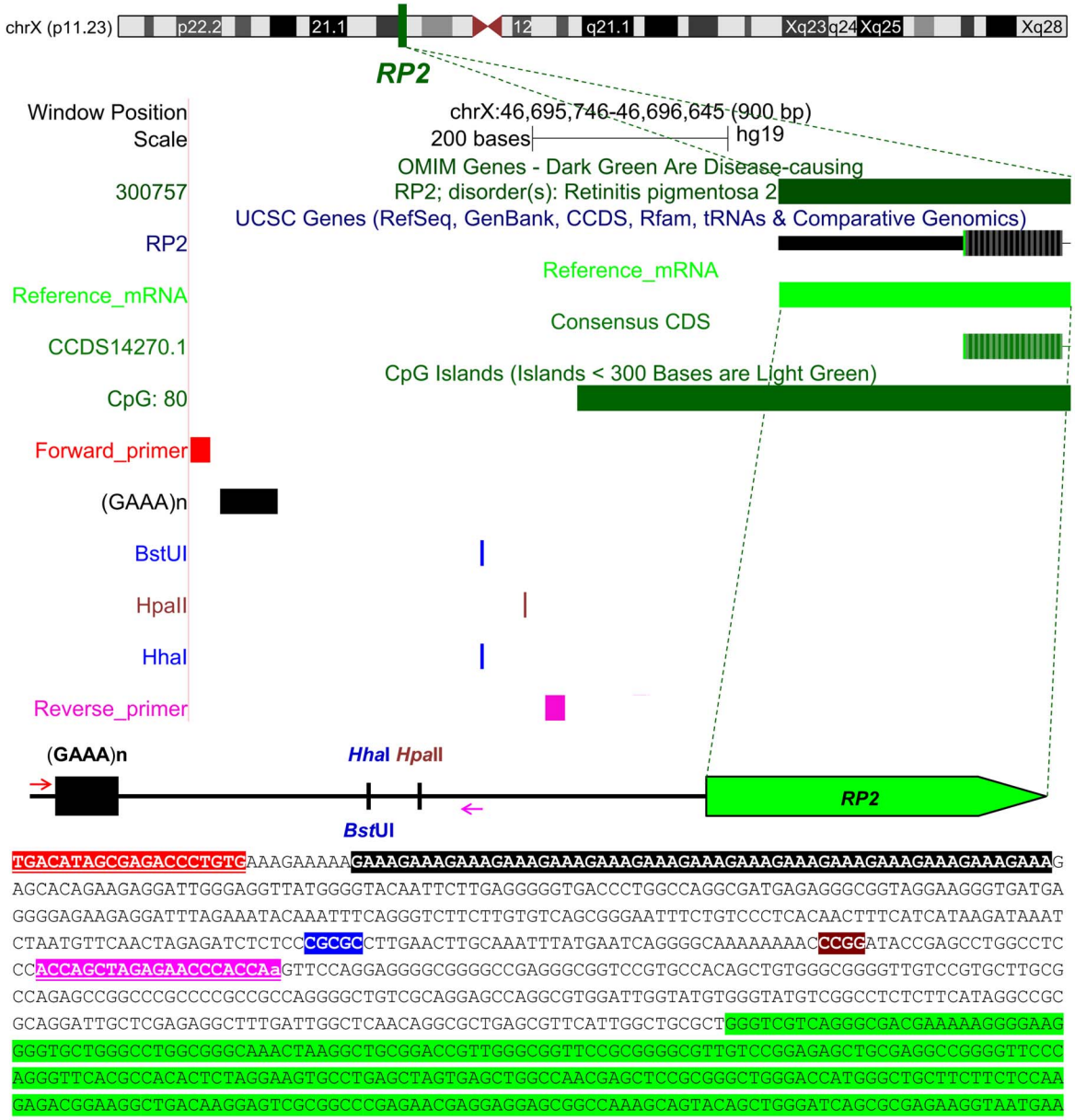


Figure 1. Chromosomal and physical map positions and sequence features of the locus encompassing the *RP2* onshore tandem GAAA repeat. The composite image above the DNA sequence is based on screenshots generated using the UCSC Genome Browser (http://genome.ucsc.edu) [49], with the *RP2* onshore tandem GAAA repeat-containing region viewing coordinates chrX:46,695,746-46,696,645 (GRCh37.p5/hg19 primary reference assembly of human X-chromosome; NC_000023.10), centered on the landmark CpG island of the *RP2* promoter. The GAAA repeat element maps within the Xp11.3. The presented features (from top to bottom) are annotated tracks for OMIM genes, UCSC Genes (RefSeq, GenBank, CCDS, Rfam, tRNAs & Comparative Genomics), reference mRNA, CpG and the tandem (GAAA)n repeat. The line drawing above the DNA sequence represents the physical map of the target locus, with the *RP2* 5′coding region highlighted in light green. The locations of the forward and reverse primer sequences used for genotyping the *RP2* onshore tandem GAAA repeat are highlighted in red and pink, respectively. The tandem GAAA repeat sequence is highlighted in black with white symbols. The 5meC-sensitive restriction endonuclease recognition sites analyzed in the XCI experiments are highlighted in blue and brown in white symbols.
doi:10.1371/journal.pone.0103714.g001

from twenty-two adult subjects (fourteen females and eight males from different social groups).

Cells

The THP-1 cell line was cultured in RPMI-1640 medium, 10% fetal bovine serum, penicillin/streptomycin, 10 mM HEPES, 1 mM sodium pyruvate and 50 μM 2-mercaptoethanol [20].

DNA and RNA extraction

Human genomic DNA was extracted from either peripheral blood or mouth epithelial cells (swabs) utilizing a commercial Illustra blood genomic Prep Mini Spin kit (GE Healthcare, Little Chalfont, UK) [21]. Genomic DNA from blood samples from female carriers of the *F8* defect, the Dutch population subset and the marmoset necropsy tissues (blood, muscle, liver, brain and skin) was extracted via phenol-chloroform and ethanol precipita-

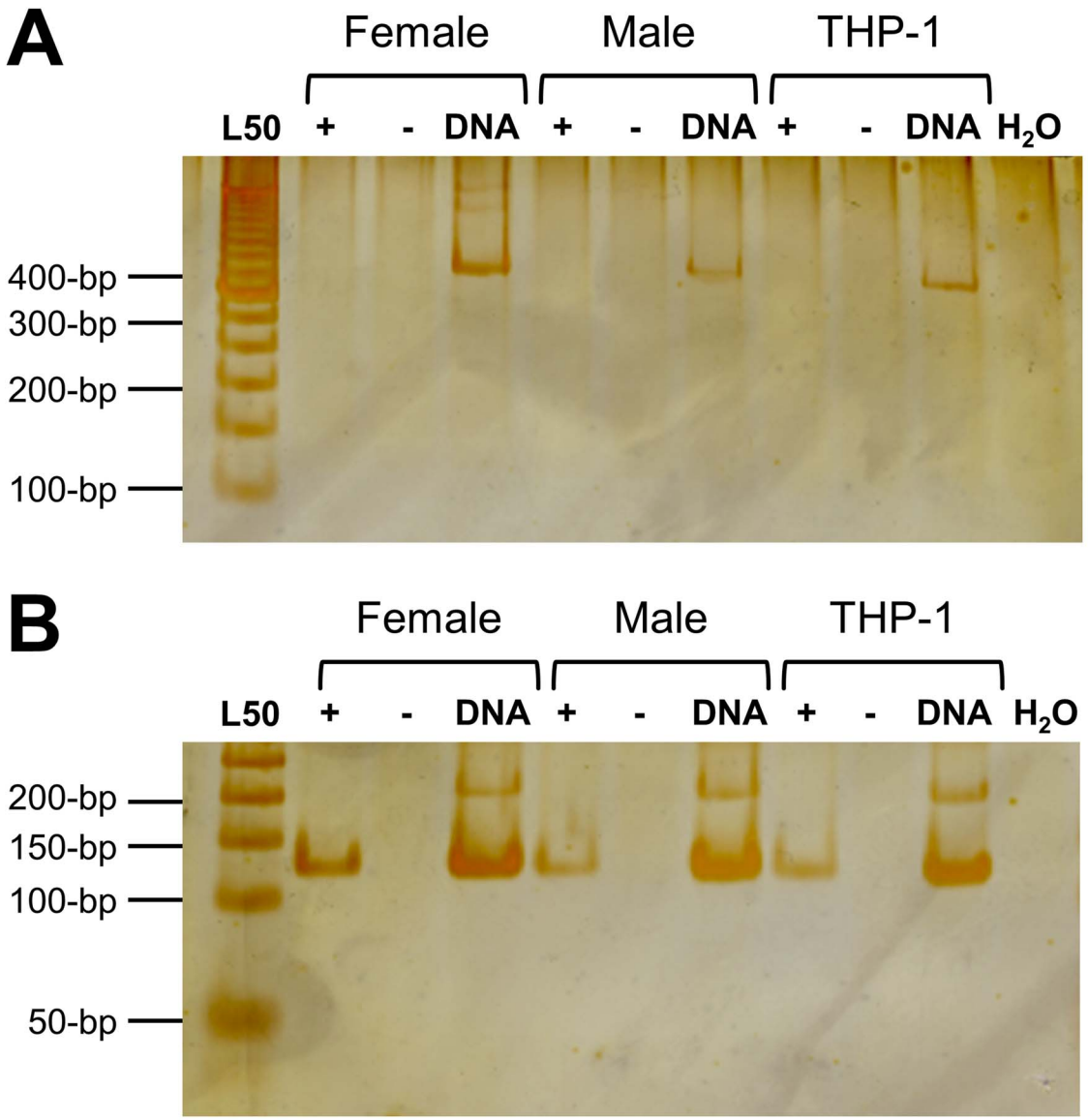


Figure 2. Reverse transcription-PCR across the GAAA repeat-containing region. *RP2* onshore tandem GAAA repeat-specific steady-state RNA is not detected in mononucleated blood cells from two healthy female donors (21 years old) or a male donor (33 years old) or from the THP-1 male cell line. RNA samples were either reverse (+) or mock (−) transcribed (RT) prior to PCR amplification across the *RP2* onshore tandem GAAA repeat-specific region (**A**) or the *GAPDH*-specific region (**B**). Corresponding samples of genomic DNA were used as positive controls for the PCR assays. The amplification products were separated via electrophoresis in an 8% acrylamide: bis-acrylamide gel and silver-stained for detection. Lane L50 shows a standard 50-bp ladder (Invitrogen); lane H_2O is the negative PCR amplification control. The range of the *RP2* onshore tandem GAAA repeat-specific DNA amplimer is 350 to 391-bp. The *GAPDH*-specific DNA amplimers are as follows: 130-bp for *GAPDHP63* (6:80663360-80663489) and *GAPDHP1* (X:39647022-39647151) and 220-bp for *GAPDH* (12:6646089-6646308). The processed (mature) *GAPDH*-specific cDNA-derived product is 130- bp.
doi:10.1371/journal.pone.0103714.g002

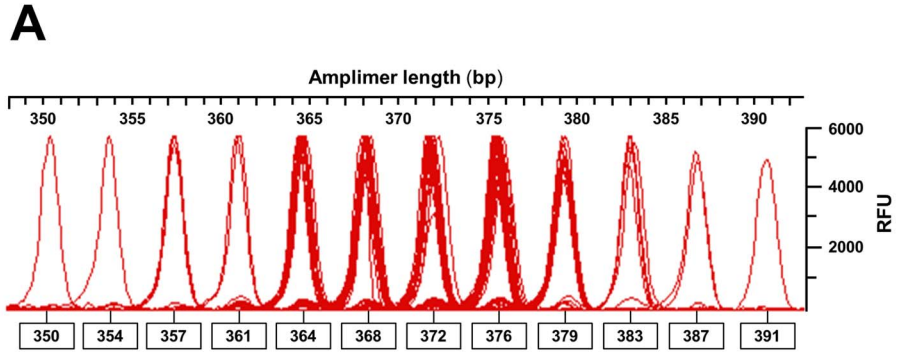
tion [22]. Total cellular RNA from human nucleated blood cells and the THP-1 cell line was extracted using TRIzol reagent (Invitrogen, Carlsbad, CA, USA).

Digestion with methylation-sensitive restriction enzymes

Genomic DNA (500 ng) was digested with *Hpa*II (Invitrogen, Carlsbad, CA, USA), *Bst*UI and *Hha*I (New England Biolabs, Ipswich, MA, USA) for 6 h at 37°C (*Hpa*II and *Hha*I) or 60°C (*Bst*UI), or was mock-digested without the restriction enzymes. The final volume of the reaction mixture was 10 µL. Throughout the methylation-based PCR assays, 5^{me}CpG-sensitive restriction endonuclease activity was assessed by genotyping DNA from four healthy males (not shown).

Analysis of allele-specific methylation

DNA genotyping was carried out in quantitative fluorescence polymerase chain biplex reactions (QF-PCR) in approximately 50 ng of digested or undigested DNA using 0.8 µM (*AR*) and 1.2 µM (*RP2*) of each primer pair (Table S1). The thermal cycling conditions were as follows: 95°C for 11 minutes (1 cycle); 94°C×1 min, 59°C×1 min and 72°C×1 min (28 cycles); and 60°C×60 min (1 cycle) in a Gene Amp PCR system 9700 (Applied



B

Allele (bp)	Allele frequencies	
	Brazilian	Dutch
350	0.008	0.008
354	0.008	
357	0.033	0.025
361	0.025	0.033
364	0.117	0.117
368	0.217	0.183
372	0.158	0.208
376	0.258	0.233
379	0.108	0.117
383	0.042	0.075
387	0.017	
391	0.008	
H_E	0.850	0.733

Figure 3. Allelic distribution of the *RP2* onshore tandem GAAA repeat. (**A**) Electropherogram of alleles observed in 60 unrelated Brazilian females genotyped via quantitative fluorescent PCR. The intensity of the red line tracing is related to the allele frequency. Smaller peaks preceding the designated allele peaks represent Taq polymerase stutter products corresponding to a mean of 2.6% of the amount of the true allele. In contrast, the mean stuttering for the *AR* disease-linked CAG repeat was 17.6% (not shown). Allele names are the lengths in base pairs of each fluorescence peak and the intensity of each peak is in relative fluorescence units (RFU). The *RP2* onshore tandem GAAA repeat locus exhibited an allelic span (the difference in length between the longest and the shortest allele per locus) of 41-bp in this population subset. (**B**) *RP2* onshore tandem GAAA repeat-containing allele frequencies and heterozygosity (H_E) rates observed in the population subsets consisting of Brazilian and Dutch women.
doi:10.1371/journal.pone.0103714.g003

Biosystems, Foster City, CA, USA). The allele profiles and areas under the curves for each allele were determined in an ABI 310 Prism Genetic Analyzer (Applied Biosystems). The data were analyzed with GeneScan Analysis 3.7 and Genotyper 3.7 software (Applied Biosystems). Fluorescent peak areas representing true alleles were normalized for the occurrence of stutter products using the approach outlined in the literature [23]. The degree of association between the percentages of the Xi/Xa referred by the methylation statuses at the *RP2* GAAA onshore and *AR* CAG repeat loci across women with varying extents of random and non-random XCI was determined by calculating the Spearman correlation coefficient, CI95% and *p* value and visualized with a scatterplot using Graph Pad Prism 5.0.

Reverse transcription-PCR (RT-PCR)

Samples of 500 ng of total RNA were digested using 1 U of DNAse I (Invitrogen) at room temperature for 15 min and then inactivated by the addition of 1 µl of EDTA (25 mM) and incubation at 65°C for 5 min in a final volume of 10 µL. The DNase I-treated RNA was reverse transcribed to single-stranded cDNA using a High Capacity cDNA Reverse Transcription Kit (Applied Biosystems) according to the manufacturer's protocol. To test for possible transcription spanning the *RP2* GAAA repeat, the primer pair used for QF-PCR typing was employed on target cDNA samples (diluted 10-fold) from nucleated blood cells and the THP-1 cell line. As a positive control, cDNA samples were tested for *GAPDH* expression using the primer sequences shown in Table S1. These primers align to three different locations in reference genomic sequences: *GAPDH* (chr12:6646089-6646308) and two pseudogenes, *GAPDHP63* (chr6:80663360-80663489) and *GAPDHP1* (chrX:39647022-39647151). In *GAPDH*, the primers anneal to exons 5 and 6 (the RNA-specific cDNA product is 130-bp in length). In all experiments, mock RT-PCR assays (without Reverse Transcriptase) were included.

Conservation of the *RP2*-extragenic GAAA repeat in nonhuman primates

The extent of conservation of the GAAA repeat-containing locus in nonhuman primates was investigated computationally using the MegaBLAST search algorithm [24] with the *in silico*-generated human PCR amplimer as the query reference sequence, followed by multiple sequence alignment of the target regions in the Molecular Evolutionary Genetics Analysis (MEGA) stand-alone program [25].

Results

Experimental strategy

To ensure success in the identification of highly polymorphic candidate repeat loci, we applied a combined comprehensive computational and empirical strategy consisting of mining the *Homo sapiens* chromosome X GRCh37.p5/hg19 primary reference genome assembly [26] for repeats that fulfill all of the following criteria: (*i*) tetranucleotides or pentanucleotides with at least twelve repeat units and a match percentage >90 according to Tandem Repeat Finder [27] (alignment parameters of 2, 7 and 7 for matches, mismatches and indels, respectively); (*ii*) mapping outside of exons and pseudoautosomal regions [24]; (*iii*) mapping <300-bp from or residing within landmark CpG islands [1] relevant to genes expressed only from Xa (e.g., escape genes excluded) [8]; and (*iv*) the occurrence of at least one 5^{me}CpG-sensitive restriction endonuclease site within 300-bp of the tandem repeat. Matching these criteria should improve the base-calling precision of templates and the measurement of true alleles by effectively limiting Taq polymerase stuttering (the magnitude of stuttering decreases as the repeat unit length increases [28,29]), and allow to achieve the power of informativeness of the *AR* disease-linked CAG repeat assay regarding the methylation statuses of X-chromosomes [12] (*AR* does not escape XCI [8], and the informativeness of repeats on X correlates with the number of perfect tandem repeat units [29,30]). The real power of this combined approach for predicting highly polymorphic STR loci in promoter regions is its direct applicability to available X-chromosome sequences of any mammalian species.

Chromosomal and physical map positions and sequence features of the novel locus

The endeavor rendered only one, albeit suitable, repeat: a tetranucleotide repeat element (physical location chrX:46695765-46695834) near *RP2* (MIM 300757) (Figure 1), the gene corresponding to X-linked retinitis pigmentosa 2 (MIM 312600), which maps to Xp11.3 [31] and does not escape XCI [8,31]. Using the alignment parameters 2, 7 and 7 for matches, mismatches and indels, respectively, Tandem Repeats Finder marks the repeat unit

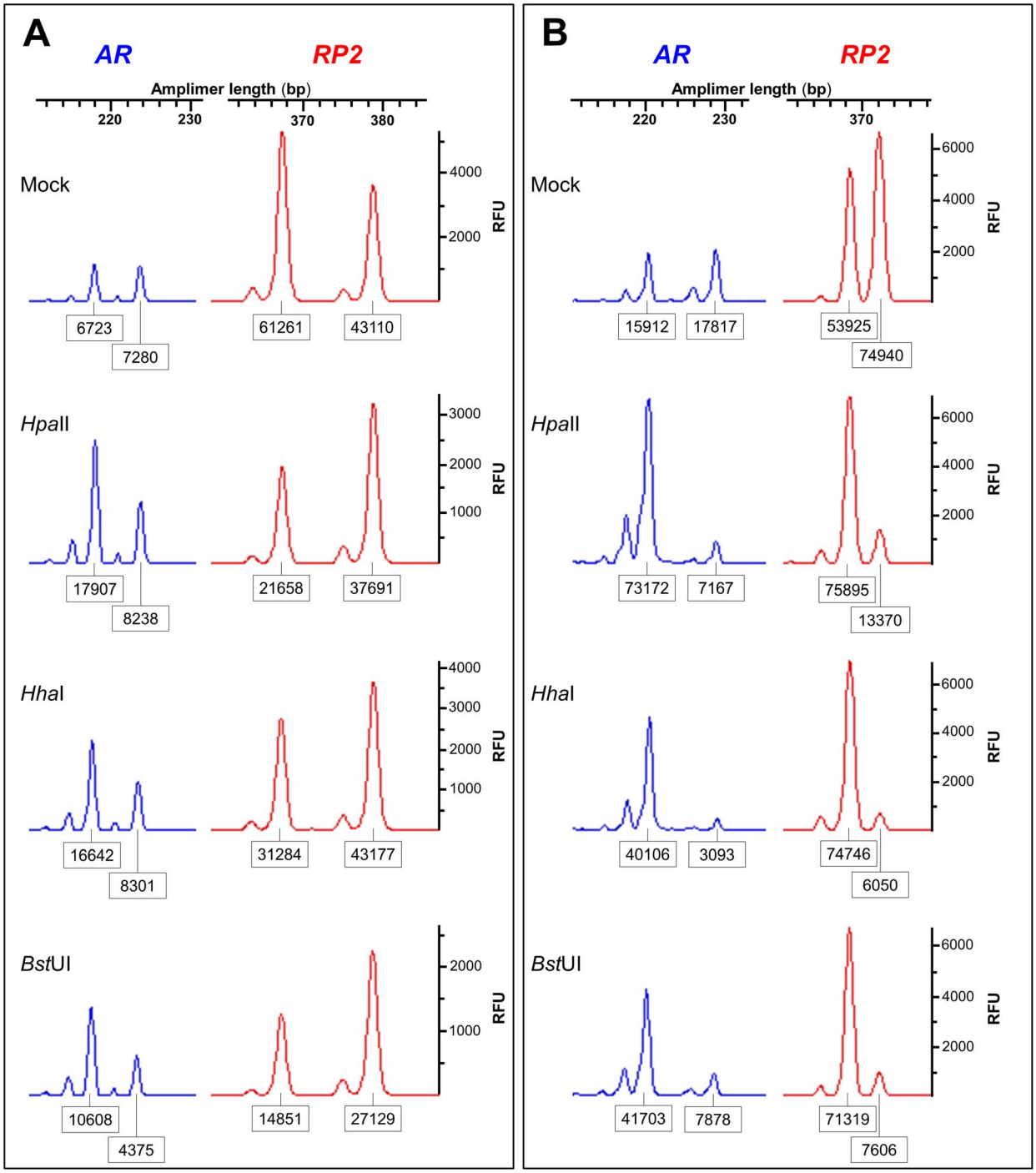


Figure 4. Methylation statuses at CpG sites near the *RP2* onshore tandem GAAA repeat. Random (**A**) and non-random (**B**) X-inactivation patterns generated for different CpG-containing 5meCpG-sensitive restriction endonuclease sites obtained using the 5meCpG-based PCR *RP2*/*AR* biplex assay across the restriction sites. Electropherograms of alleles observed in either undigested genomic DNA or DNA digested with *Hpa*II, *Hha*I or *Bst*UI from females genotyped via quantitative fluorescent PCR are shown. The boxed numbers correspond to the areas under the allele peaks and the intensity of each peak is in relative fluorescence units (RFU).
doi:10.1371/journal.pone.0103714.g004

as AAAG. However, comparison of three public reference genomic sequences showed that the alleles consist of multiple copies of the GAAA repeat unit (Figure S1). Henceforth, we refer to this repeat element as GAAA to indicate the physical location of the GAAA repeat-containing allele in the GRCh37.p5/hg19 primary reference assembly of the human X-chromosome.

The GAAA repeat is positioned -582, -598 or -630-bp (upstream) of known transcription start sites of *RP2* (Figure S2). The element maps on shore, 230-bp upstream of the *RP2* CpG island (Genomic coordinates NC_000023.10 Reference GRCh37.p5 Primary Assembly X:46695995-46696984), a landmark that exhibits differential methylation [1], displaying

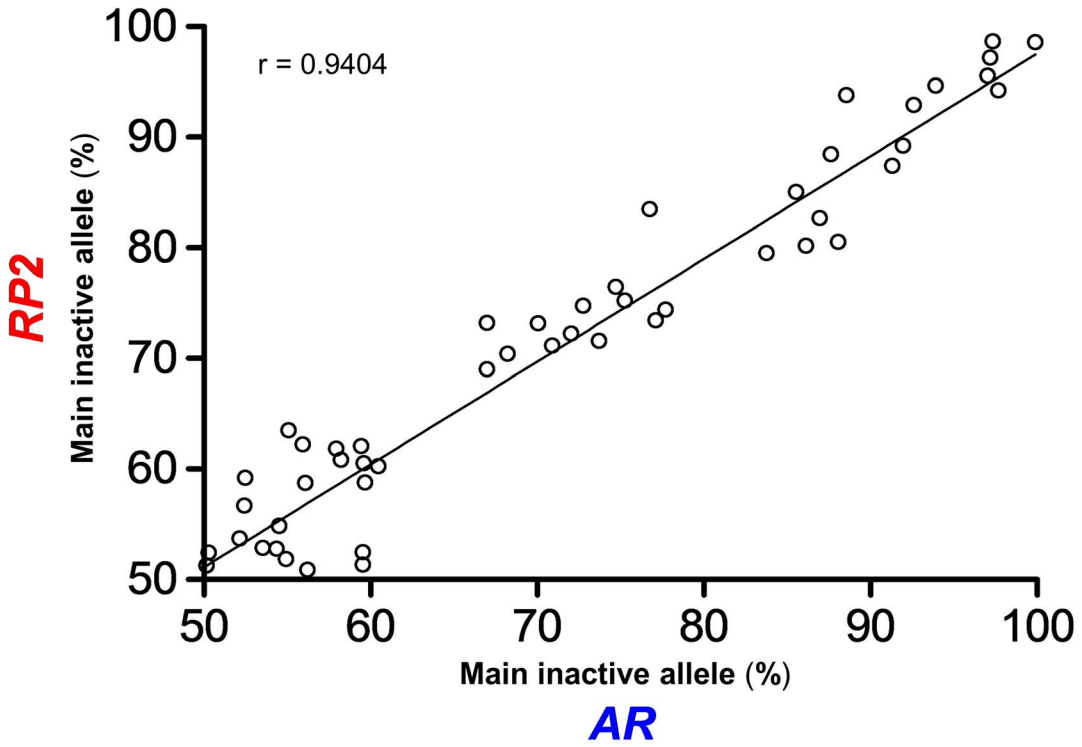


Figure 5. *RP2* and *AR* repeat-based methylation results are highly concordant. Scatterplot visual assessment of the strength of association between the percentages of the main inactive allele referred by the methylation statuses at the *RP2* onshore tandem GAAA repeat (*y*-axis) and the *AR* CAG repeat (*x*-axis) loci. The methylation statuses are highly concordant (Spearman r = 0.9404, CI95% = 0.8950 to 0.9665; $p<0.0001$) across varying degrees of random (50–80%) and non-random (>80%) XCI. The regression line superimposed on the plot provides the best-fitting straight line for the scattered data.
doi:10.1371/journal.pone.0103714.g005

increased methylation on Xi in 46, XX and reduced methylation in 45, X females [32]. The *RP2* onshore tandem GAAA repeat is therefore positioned approximately 20 Mb upstream of the *AR* disease-linked, exonic CAG repeat, which maps to Xq12.

The *RP2* onshore tandem GAAA repeat does not overlap with *RP2* cDNAs, known transcription factor binding sites (Figure S2), cap analysis gene expression promoters (Figure S3) or microRNA precursors (Figure S2) that are predicted or annotated in public repositories (see Web Resources) [24,33].

Reverse transcription-PCR across the *RP2* onshore tandem GAAA repeat

We performed reverse transcription-PCR experiments on total RNA from peripheral blood (normal women and men) and from the FANTOM-DB [33] human acute monocytic leukemia THP-1 reference cell line and found no detectable GAAA repeat-specific steady-state RNA (Figure 2). *In silico* PCR analyses using public RNA-Seq expression databases revealed no significant transcription activity across (or within) the *RP2* onshore tandem GAAA repeat locus in many different cell types and lines (Figure S4). However, the evidence does support the prediction of long RNA-Seq junctions based on ENCODE/CSHL, pooled from GM12878 whole-cell polyA (hg19 coordinates chrX:46545885-46727348). These long RNA-Seq junctions encompass multiple genes.

Allelic distribution for the *RP2* onshore tandem GAAA repeat

The *RP2* GAAA onshore repeat-containing locus encompasses the reference upstream gene deletion/insertion variations rs6151299, rs373239539, rs201864594, rs201168201 and rs71950018. No validation had been reported for these variants (dbSNP build 138). We employed both the *RP2* onshore tandem GAAA repeat and the *AR* disease-linked, exonic CAG repeat in developing a biplex 5meCpG-based quantitative fluorescent PCR surrogate assay of human X-chromosome activity. For the determination of heterozygosity rates and allele frequencies, we genotyped two population subsets of sixty healthy unrelated women from Brazil and the Netherlands. For the *RP2* onshore tandem GAAA repeat, we observed up to twelve alleles with virtually no stuttering (Figure 3) in either subset. In the Brazilian subset, the heterozygosity rate for the *RP2* onshore tandem GAAA repeat was 0.85, matching that of the *AR* disease-linked CAG repeat (Figure S5). For the Dutch subset, the rate was 0.73, which was lower than that observed for the *AR* marker (0.87) (Figure S6). When the two subsets were pooled, the combined informativeness (e.g., at least one informative marker) of the *RP2*/*AR* biplex assay was 0.97.

Methylation statuses of CpG sites near the human *RP2* onshore tandem GAAA repeat

Each *RP2* onshore tandem GAAA repeat-containing allele comprises eight CpG sites, corresponding to five 5meCpG-sensitive restriction endonucleases (*Aci*I, *Bst*UI, *Fau*I, *Hha*I and *Hpa*II) and is therefore liable to multipoint 5meCpG interrogation. We used *Hpa*II, *Bst*UI and *Hha*I in XCI experiments, applying the 5meCpG-based PCR assay targeting the polymorphic repeat. The random (Figure 4A) and non-random (Figure 4B) patterns of X-inactivation obtained using these restriction enzymes were similar. We note, however, that the Xa/Xi lyonization ratios obtained using the *Hha*I and *Bst*UI enzymes were not always

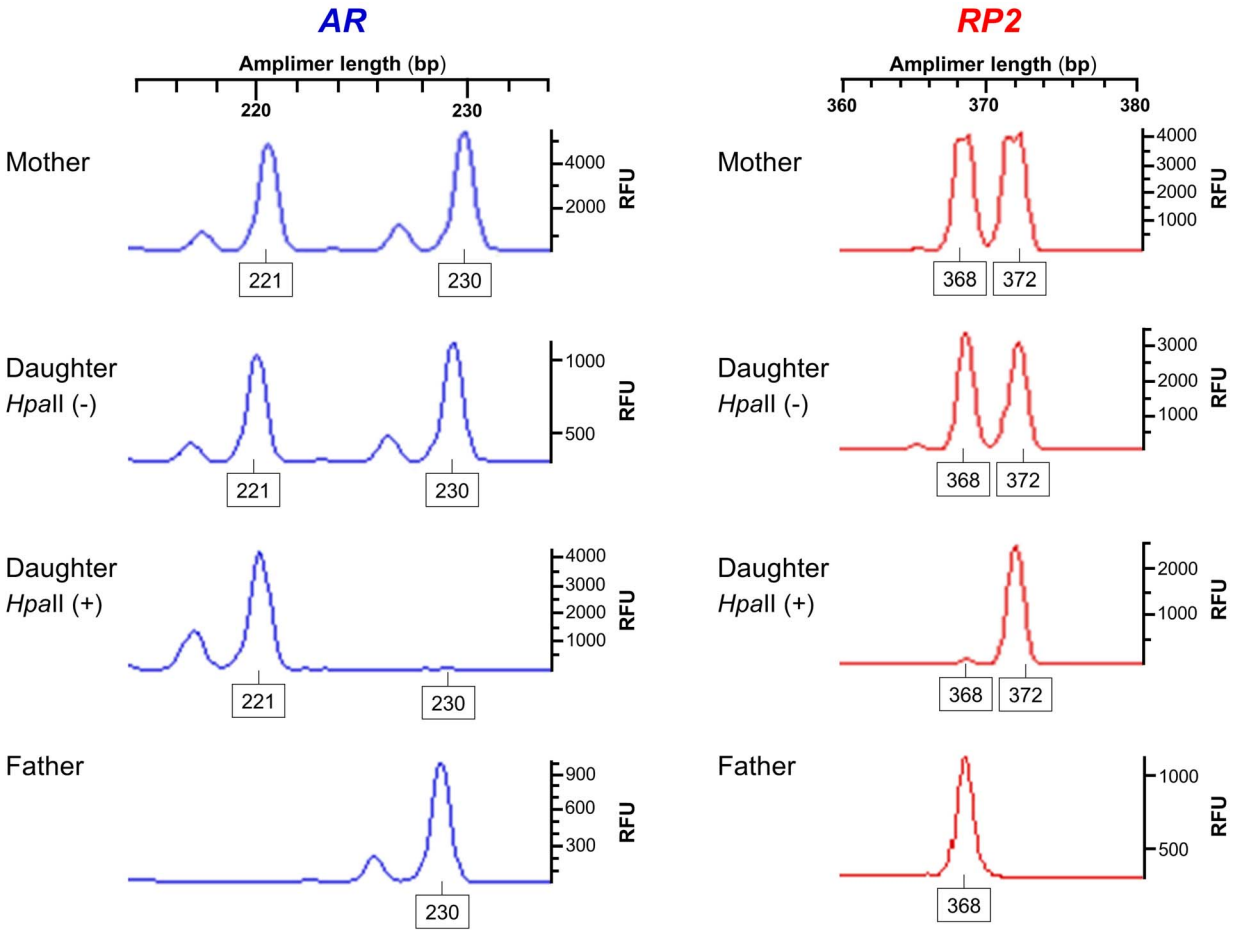


Figure 6. *AR* CAG and the *RP2* GAAA polymorphisms refer to the same X-chromosomes. Segregation analysis of either *AR* or *RP2* alleles distinguishes the maternal origin of the preferentially skewed Xi present in the daughter. Xi is identified based on the 230-bp *AR* allele and the 368-bp *RP2* allele. The allele names are the lengths in base pairs of each fluorescence peak and the intensity of each peak is in relative fluorescence units (RFU). Note that the magnitude of stuttering at the *RP2* onshore tandem GAAA repeat is minimal, in contrast with that at the *AR* CAG repeat.
doi:10.1371/journal.pone.0103714.g006

highly corresponding. This result may be related to the fact that in this particular target sequence the *Hha*I site overlaps the CpG within the *Bst*UI site and that overlapping CpG sites may block or impair cleavage if methylated (New England Biolabs usage guidelines). Therefore, this is a case where the overlapping CpG methylation cannot be predicted accurately.

RP2 and *AR* repeat-based methylation results are concordant

To correlate random and non-random X-inactivation patterns from the *RP2* onshore GAAA and *AR* CAG repeats, we genotyped a third subset of fifty unrelated women from Brazil and Argentina (Figure S7) and analyzed the CpG methylation statuses within the *Hpa*II sites. These women had known *AR* CAG repeat 5^{me}CpG allele-specific profiles and, hence, known XCI ratios. The patterns of X-inactivation obtained using the *RP2*/*AR* repeat biplex assay were highly concordant (Spearman $r = 0.9404$; $p < 0.0001$) (Figure 5).

To address the question of whether the *RP2* GAAA-containing alleles are located on the same Xa/Xi chromosomes identified based on the *AR* CAG-containing repeat, we determined the parent-of-origin of Xa and Xi in a nuclear family in which the normal daughter exhibited extremely skewed XCI in peripheral blood leukocytes (Figure 6). The segregation analysis demonstrated that the *AR* CAG and the *RP2* GAAA polymorphisms refer to the same X-chromosome based on correctly identifying the maternal origin (^{M}X) of the preferential Xi in this nuclear family.

To demonstrate the power of the *RP2* onshore tandem GAAA repeat in discriminating Xa from Xi in heterozygote carriers of an X-linked recessive defect that manifests through non-random XCI, we genotyped four confirmed heterozygous women affected by severe hemophilia A. Two of these individuals are conventional, non-symptomatic carriers who tested positive for *F8* intron 22 inversions via inverse shifting-PCR [34] and for random XCI based on the *AR* disease-linked CAG repeat assay; the other two are heterozygous carriers of missense and frameshift mutations in factor VIII domains A1 and B, respectively, and they present with symptoms of hemophilia A through non-random XCI. Again, the XCI patterns associated with the *RP2* onshore tandem GAAA repeat were highly concordant with those of the *AR* disease-linked CAG repeat, as exemplified in Figure 7 for a heterozygous female, hemophiliac due to highly skewed inactivation of the unaffected X-chromosome.

The *RP2* onshore tandem GAAA repeat locus is conserved in nonhuman primates

Although the *RP2* gene is conserved in mammals (data not shown), the *RP2* onshore tandem GAAA repeat locus is restricted

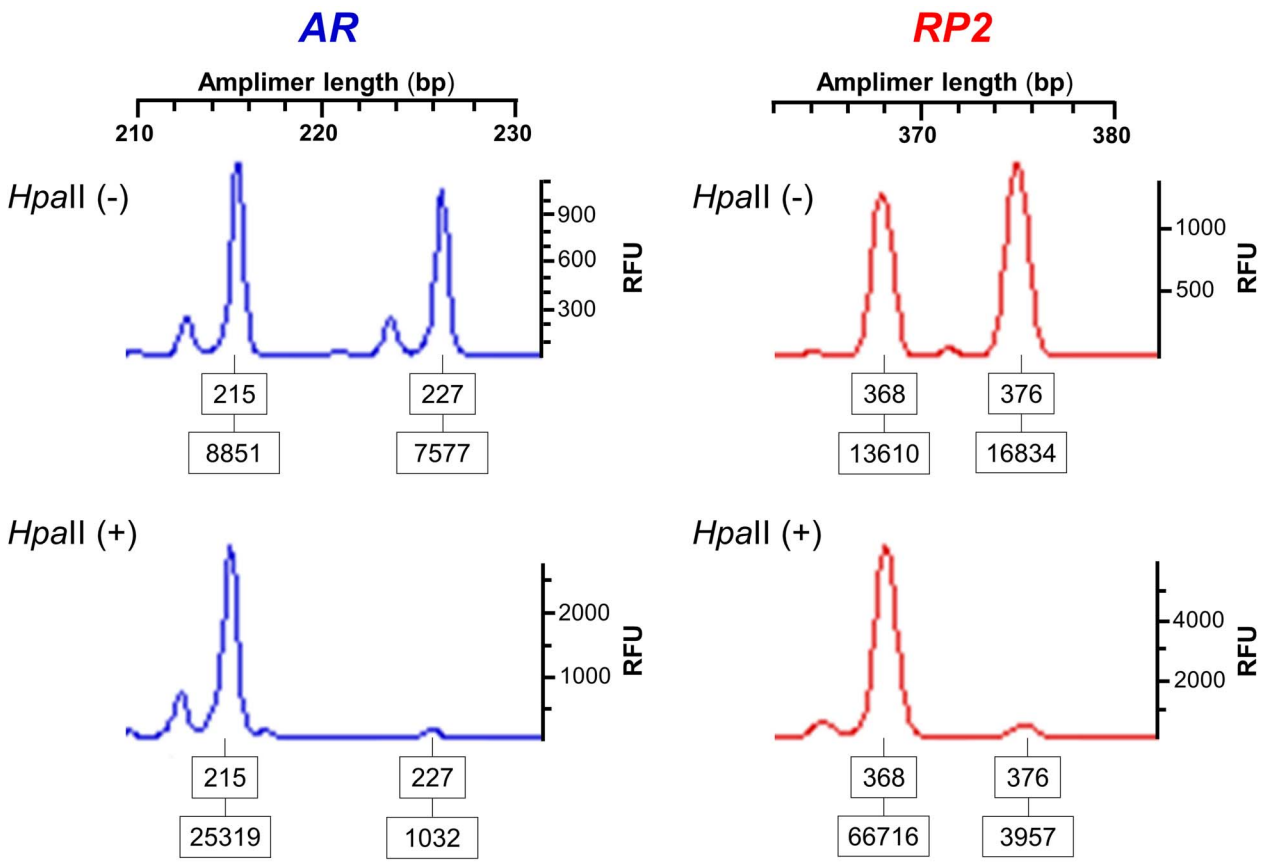


Figure 7. Hemophilia A occurs due to highly skewed XCI. Electropherograms of alleles obtained using the 5meCpG-based *RP2/AR* repeat biplex PCR assay across the *Hpa*II restriction site in a heterozygote female carrier of a one-base insertion, frameshift mutation in factor VIII domain B. The female is a hemophiliac due to highly skewed inactivation of the unaffected X-chromosome, represented by the *AR* 215-bp and *RP2* 368-bp alleles. The *RP2* and *AR* repeat-based 5meCpG readouts refer to the skewed X-inactivation state. The *F8* mutation was screened through conformational sensitive gel electrophoresis [19] and direct sequencing. Allele names (upper boxed numbers) are the lengths in base pairs of each fluorescence peak and the intensity of each peak is in relative fluorescence units (RFU). The lower boxed numbers correspond to the areas under the allele peaks.
doi:10.1371/journal.pone.0103714.g007

to primates, as judged based on comparative *in silico* analyses using genomic reference sequences from public databases (Figure S8). This observation indicates that the insertion of the GAAA repeat element was a very recent event. The number of uninterrupted (perfect tandem array) GAAA repeat units varied from 3 (squirrel monkey) to 16 (humans) (Table S2 and Figure S9). We used the human *RP2* GAAA onshore repeat amplimer reference sequence, without masking the repeat region, to computationally search public data for homologs in primates, and we conducted evolutionary analyses with unmasked, masked or exclusion of repeat regions to construct a phylogenic tree (Figure 8). We found no evidence of a linear increase in the number of uninterrupted GAAA repeat units proportional to the time of divergence between nonhuman primates and humans.

The *RP2* onshore tandem GAAA repeat is polymorphic in marmosets

We hypothesized that the *RP2* onshore tandem GAAA repeat locus may be useful in XCI studies in nonhuman primate species in which the *AR* CAG-repeat locus is not polymorphic [16]. We therefore tested this possibility in the naturally occurring, pervasive hematopoietic chimeric New World monkey marmoset (*Callitrichidae*) [35]. We observed only two alleles (318-bp and 327-bp) in 22 different animals (Figure 9A). All the males were monoallelic (hemizygous). The heterozygosity rate in females was 0.35. The *RP2* GAAA repeat-containing amplimer, as validated via *in silico* PCR, comprises five CpG sites, the methylation statuses of which can be determined with the restriction enzymes *Aci*I, *Bst*UI and *Fau*I. Here, we analyzed the 5meCpG-sensitive *Bst*UI recognition site (Figure 9B). For all heterozygote female marmosets tested, the pattern of methylation at the CpG site linked to the GAAA repeat of interest was random, with Xa/Xi ratios varying from 38 to 65%. Different tissues (blood, muscle, liver, brain and skin) from the same animal also yielded random, yet varying, Xa/Xi ratios (data not shown).

Discussion

Notwithstanding the remarkable advances in understanding human genome structural variation and rapidly evolving technologies, the *AR* disease-linked CAG repeat-based HUMARA assay has remained the mainstay of XCI diagnosis in the two decades since it was reported [12]. Despite the elevated heterozygosity observed worldwide, there are important drawbacks to genotyping with exonic rather than neutral repeats. CAG repeat-associated non-ATG translation (RAN translation) can occur across human genes, and CAG repeat expansions in transcripts without an ATG result in the accumulation of toxic homopolymeric proteins in all three reading frames [36]. There is also evidence of bidirectional

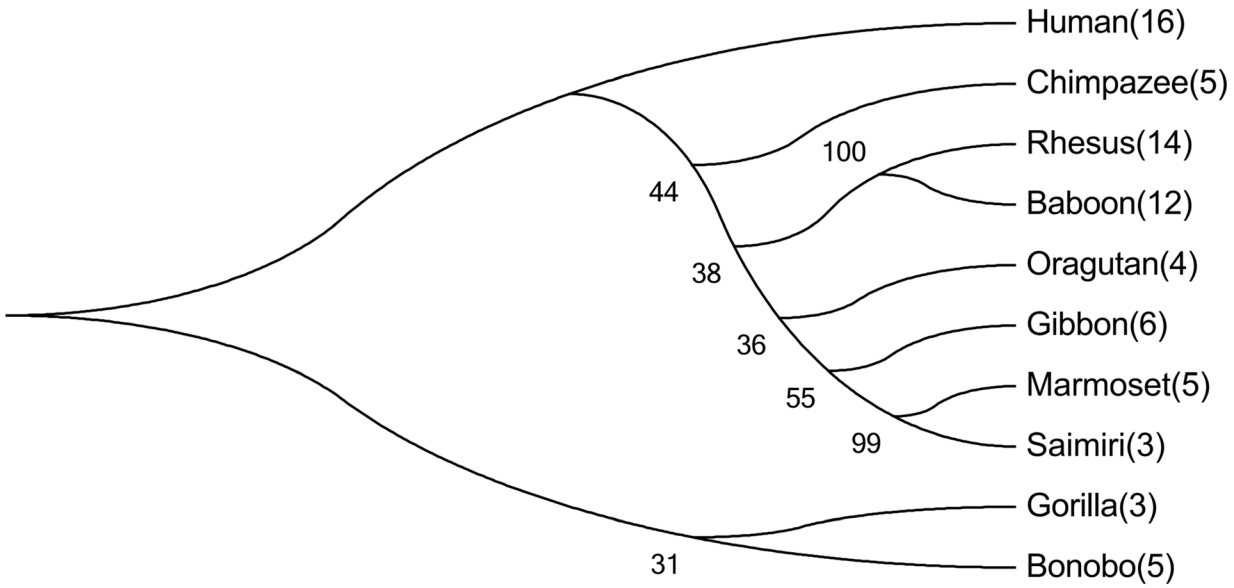


Figure 8. Molecular phylogenetic analysis using the maximum likelihood method. The evolutionary history was inferred using the maximum likelihood method based on the Tamura-Nei model [50]. The bootstrap consensus tree inferred from 1,000 replicates [51] is taken to represent the evolutionary history of the analyzed taxa [51]. The percentages of replicate trees in which the associated taxa clustered together in the bootstrap test (1,000 replicates) are shown next to the branches [51]. Initial tree(s) for the heuristic search were obtained automatically by applying the maximum parsimony method. The analysis involved 10 nucleotide sequences. The codon positions included were $1^{st} + 2^{nd} + 3^{rd}$+ Noncoding. In total, there were 410 positions in the final dataset. Evolutionary analyses were conducted in MEGA5 [25]. The numbers in parentheses correspond to the lengths of the uninterrupted tandem arrays in GAAA repeat units.
doi:10.1371/journal.pone.0103714.g008

transcription of triplet repeat disease genes [37]. Moreover, PCR genotyping involving trinucleotide repeats is prone to template errors due to *in vitro* replication slippage by Taq polymerase [38], resulting in unwanted n-3 stutter products consisting of multiples of the true template alleles [39] to varying magnitudes, in a repeat sequence-dependent manner. Although several dinucleotide repeat loci have been proposed as supplements or alternatives to the *AR* disease-linked CAG repeat assay [40–42], the greater magnitude of n-2 stutter products is an unfortunate shortcoming, which can considerably influence the results and confound the analysis, as discrepancies in Xa/Xi ratios relating to the *AR* disease-linked CAG repeat assay have been reported [41,42].

In contrast with the *AR* disease-linked CAG repeat (≥38 CAG repeat units are linked to KD [13]), the novel *RP2* onshore tandem GAAA repeat is endowed with neutral features. This observation suggests that expansions of the *RP2* onshore tandem GAAA repeat will not produce toxic RNAs that might otherwise influence cell viability, disease penetrance and pathological severity [43].

Data from a recent methylome study showed that the amplimer encompassing the human *RP2* onshore GAAA repeat spans eight CpG sites that are differentially hypomethylated in a tissue-dependent manner [44]. The same configuration occurs for the *AR* amplimer, but the levels of methylation are higher because the CpG sites are in the gene body. The observation that the Xa/Xi ratios inferred by determining the methylation statuses of CpG sites near the human *RP2* GAAA onshore repeat are highly concordant with the patterns of X-inactivation inferred from the HUMARA assay assuages the concerns related to typing the novel extragenic *RP2* onshore tandem GAAA repeat in XCI studies. We also showed that the extragenic *RP2* onshore tandem GAAA repeats and the neighboring CpG methylation statuses refer to exactly the same parental chromosomes identified based on the *AR* CAG repeat. Furthermore, it is known that the transcriptional XCI patterns generated by pyrosequencing correlate excellently (Pearson $r^2 = 0.96$) with the XCI ratios reported using the HUMARA assay [45]. Thus, we feel confident that the analysis using the methylation statuses surrounding the *RP2* onshore tandem GAAA repeat will be as accurate as those obtained using the *AR* CAG marker in discriminating Xa from Xi chromosomes in other tissues and population subsets.

Evolutionary analyses of the *RP2* onshore tandem GAAA repeat locus indicated that the tandem arrangement is well conserved in nonhuman primates. Although there is a trend of directional expansion of the repeat, we see no evidence for a linear continuous increase in the length of a perfect tandem array proportional to the time since divergence from the last common ancestor. This observation contrasts with findings related to the *AR* CAG exonic repeat, for which a linear increase in triplet repeat length proportional to the time since divergence has been reported twice [16,46].

Because of its proximity to known *RP2*+1 transcriptional start sites and its polymorphic nature, the *RP2* onshore tandem GAAA repeat could be regarded as a core promoter STR and may be a source of variation across species [47]. Whether the GAAA repeat expansion plays a role in *RP2* gene expression leading to inter-individual variation is currently unknown.

The *RP2* onshore tandem GAAA repeat was less polymorphic in marmosets than in humans, with only 2 alleles being observed in 22 animals. The marmoset reference genomic sequence bears only five uninterrupted GAAA repeat units, represented by the observed major (e.g., the most frequent and oldest) 327-bp allele. This result suggests that in marmosets, the *RP2*-extragenic GAAA locus may correspond to stable, fixed $(GAAA)_{5>3}$ deletion/insertion biallelic variation. Given that the highest possible heterozygosity rate for any biallelic system is 50%, the observed heterozygosity rate of 35% is highly significant. Alternatively, this result can be explained by reduced genetic diversity due to a

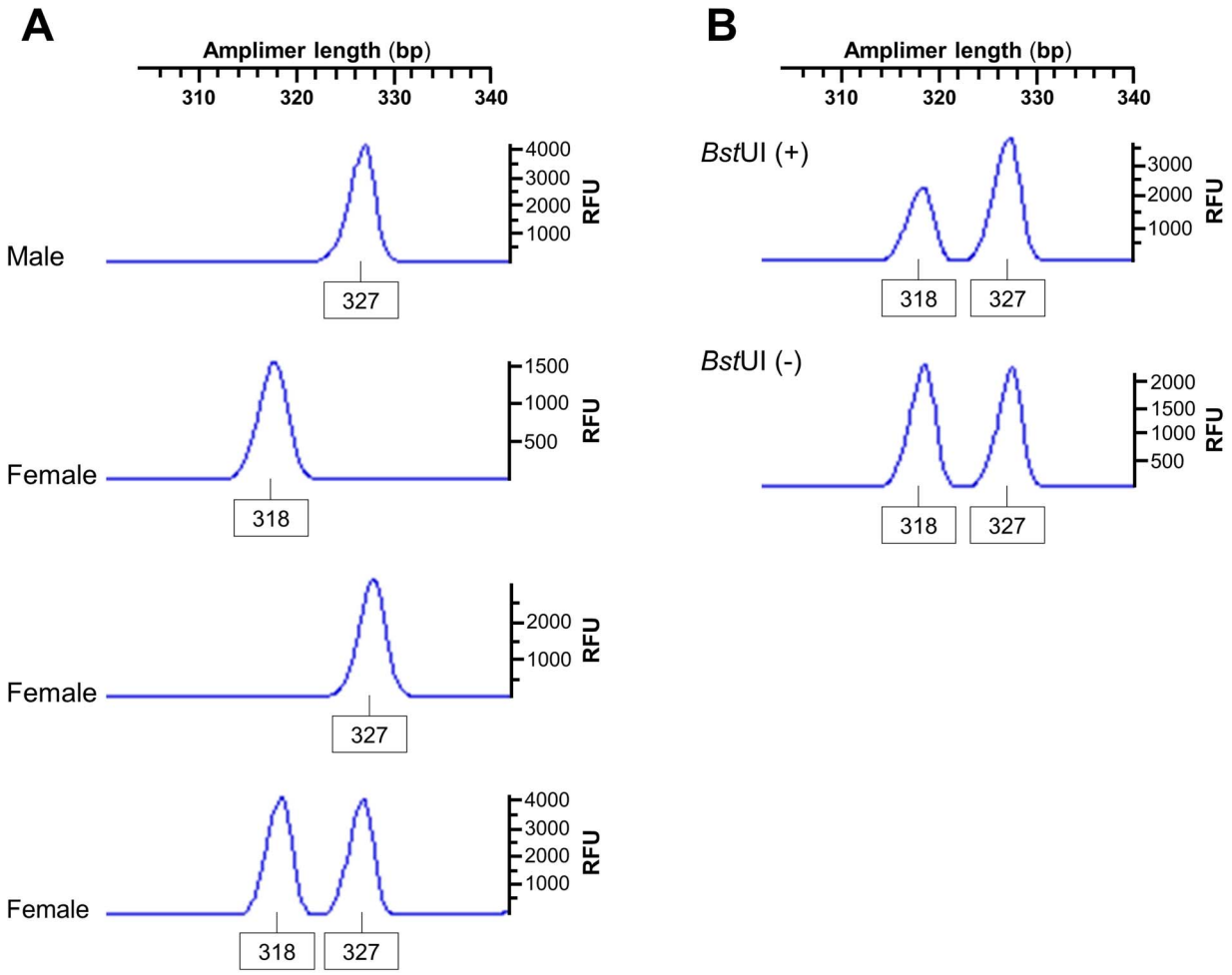


Figure 9. The *RP2* onshore tandem GAAA repeat is polymorphic in marmosets. Electropherograms of alleles observed in marmosets genotyped via quantitative fluorescent PCR. (**A**) Representative allele profiles from males, which exhibited only the major allele, and female animals with three distinct genotypes (homozygotes for either the minor or major allele or heterozygotes) are shown. (**B**) Representative random XCI pattern observed at the 5meCpG-sensitive *Bst*UI recognition site within the *RP2* GAAA-containing amplimer, with an Xa/Xi ratio of approximately 65%. The allele names are the lengths in base pairs of each fluorescence peak and the intensity of each peak is in relative fluorescence units (RFU).
doi:10.1371/journal.pone.0103714.g009

limited number of founder animals in the studied primate colony, as reported for the CAG *AR* repeat in nonhuman-primates [16] and/or functional restriction of the ability of the repeat to expand in these species. We are currently addressing the latter possibility. Nevertheless, the observed polymorphism in marmosets enabled us to develop a molecular genotyping assay to study XCI in a small nonhuman primate experimental model in which the *AR* disease-linked CAG repeat locus is known to be monomorphic [16].

Conclusions

The superior efficacy of the 5meCpG-based *RP2/AR* repeat biplex assay in differentiating the parental origins of Xa and Xi chromosomes in approximately 97% of human females constitutes a notable advance in the field of XCI, and this assay excels at determining the 5meCpG statuses of alleles on the Xp (*RP2*) and Xq (*AR*) chromosome arms in a single reaction. The *RP2* onshore tandem GAAA repeat will facilitate studies on the variable phenotypic expression of dominant and recessive X-linked diseases (e.g., Rett syndrome, hemophilia A and B, mental disability), epigenetic changes in twins, the physiology of aging hematopoiesis, the pathogenesis of age-related hematopoietic malignancies and the clonality of cancers in human and nonhuman primates [48].

Supporting Information

Figure S1 Computational validation of a polymorphism at the *RP2* onshore tandem GAAA repeat locus in reference genome sequences.
(DOC)

Figure S2 Physical positions of known transcription start sites, transcription factor binding sites and predicted microRNA precursors relative to the *RP2* onshore tandem GAAA repeat locus.
(DOC)

Figure S3 The *RP2* onshore tandem GAAA repeat locus does not overlap with *RP2* cDNAs or cap analysis gene expression promoters (CAGE).
(DOC)

Figure S4 RNA-Seq evidence across the human *RP2* onshore tandem GAAA repeat locus.
(DOC)

Figure S5 Distribution of distinct genotypes for the *RP2* onshore tandem GAAA repeat (A) and *AR* tandem CAG repeat (B) loci in the first population subset (n = 60 Brazilian females).
(DOC)

Figure S6 Distribution of distinct genotypes for the *RP2* onshore tandem GAAA repeat (A) and *AR* tandem CAG repeat (B) loci in the first population subset (n = 60 Dutch females).
(DOC)

Figure S7 Genotypes for the *RP2* onshore tandem GAAA repeat (A) and *AR* tandem CAG repeat (B) loci in the third population subset (n = 46 Brazilian females and n = 4 Argentinean females), consisting of women with known *AR* tandem CAG repeat 5^{me}C allele-specific profiles and, hence, known XCI ratios.
(DOC)

Figure S8 The *RP2* onshore tandem GAAA repeat locus is conserved in primates.
(DOC)

Figure S9 Multiple sequence alignment of the *RP2* onshore tandem GAAA repeat locus reveals high conservation in primates.
(DOC)

Table S1 PCR primer sequences used in this study.
(DOC)

Table S2 Structure of the *RP2* onshore tandem GAAA repeat region in primates.
(DOC)

Acknowledgments

The authors thank Adriane Araújo, Cristiana Libardi Miranda-Furtado, David van Bruggen and Liesbeth van Iperen for technical help in collecting and processing biological samples.

Web resources

The URLs for data presented herein are as follows:

CpG Island Searcher, http://cpgislands.usc.edu/

NCBI (National Center for Biotechnology), http://www.ncbi.nlm.nih.gov/guide/

UCSC Genome Bioinformatics tools, http://genome.ucsc.edu/

Ensembl, http://www.ensembl.org/index.html

TRF, http://tandem.bu.edu/trf/trf.html

ENCODE (Encyclopedia of DNA Elements), http://genome.ucsc.edu/ENCODE/

DBTSS: Database of Transcriptional Start Sites, http://dbtss.hgc.jp/

CID-miRNA, http://mirna.jnu.ac.in/cidmirna/

FANTOM, http://fantom.gsc.riken.jp/4/gev/gbrowse/hg18/

mirBase, http://www.mirbase.org/

REBASE, http://rebase.neb.com/cgi-bin/mslist

SwissRegulon, http://www.swissregulon.unibas.ch/cgi-bin/regulon?page = swissregulon

TRED: Transcriptional Regulatory Element Database, http://rulai.cshl.edu/TRED.

dbSNP, http://www.ncbi.nlm.nih.gov/snp/

dbSNP, http://www.ncbi.nlm.nih.gov/snp/

HEMApSTR, http://www.uenf.br/Uenf/Pages/CBB/LBT/HEMApSTR.html

Author Contributions

Conceived and designed the experiments: Filipe Machado EM-A. Performed the experiments: Filipe Machado Fabricio Machado MAF VLL AFAS AFLR. Analyzed the data: Filipe Machado Fabricio Machado MAF VLL CPR CDDB AFAS AFLR SMCSL ESR EM-A. Contributed reagents/materials/analysis tools: CPR CDDB SMCSL ESR LSS CRRM. Wrote the paper: EM-A.

References

1. Illingworth RS, Gruenewald-Schneider U, Webb S, Kerr AR, James KD, et al. (2010) Orphan CpG islands identify numerous conserved promoters in the mammalian genome. PLoS Genet 6: e1001134.
2. Bird A (2011) The dinucleotide CG as a genomic signalling module. J Mol Biol 409: 47–53.
3. Kar S, Deb M, Sengupta D, Shilpi A, Parbin S, et al. (2012) An insight into the various regulatory mechanisms modulating human DNA methyltransferase 1 stability and function. Epigenetics 7: 994–1007.
4. Lyon MF (1961) Gene action in the X-chromosome of the mouse (Mus musculus L.). Nature 190: 372–373.
5. Augui S, Nora EP, Heard E (2011) Regulation of X-chromosome inactivation by the X-inactivation centre. Nat Rev Genet 12: 429–442.
6. Wutz A (2011) Gene silencing in X-chromosome inactivation: advances in understanding facultative heterochromatin formation. Nat Rev Genet 12: 542–553.
7. Pessia E, Makino T, Bailly-Bechet M, McLysaght A, Marais GA (2012) Mammalian X chromosome inactivation evolved as a dosage-compensation mechanism for dosage-sensitive genes on the X chromosome. Proc Natl Acad Sci U S A 109: 5346–5351.
8. Carrel L, Willard HF (2005) X-inactivation profile reveals extensive variability in X-linked gene expression in females. Nature 434: 400–404.
9. Berletch JB, Yang F, Disteche CM (2010) Escape from X inactivation in mice and humans. Genome Biol 11: 213.
10. Berletch JB, Yang F, Xu J, Carrel L, Disteche CM (2011) Genes that escape from X inactivation. Hum Genet 130: 237–245.
11. Sin HS, Namekawa SH (2013) The great escape: Active genes on inactive sex chromosomes and their evolutionary implications. Epigenetics 8: 887–892.
12. Allen RC, Zoghbi HY, Moseley AB, Rosenblatt HM, Belmont JW (1992) Methylation of HpaII and HhaI sites near the polymorphic CAG repeat in the human androgen-receptor gene correlates with X chromosome inactivation. Am J Hum Genet 51: 1229–1239.
13. La Spada AR, Wilson EM, Lubahn DB, Harding AE, Fischbeck KH (1991) Androgen receptor gene mutations in X-linked spinal and bulbar muscular atrophy. Nature 352: 77–79.
14. Giovannucci E, Stampfer MJ, Krithivas K, Brown M, Dahl D, et al. (1997) The CAG repeat within the androgen receptor gene and its relationship to prostate cancer. Proc Natl Acad Sci U S A 94: 3320–3323.
15. Gu M, Dong X, Zhang X, Niu W (2012) The CAG repeat polymorphism of androgen receptor gene and prostate cancer: a meta-analysis. Mol Biol Rep 39: 2615–2624.
16. Mubiru JN, Cavazos N, Hemmat P, Garcia-Forey M, Shade RE, et al. (2012) Androgen receptor CAG repeat polymorphism in males of six non-human primate species. J Med Primatol 41: 67–70.
17. dos Santos Sales I, Ruiz-Miranda CR, de Paula Santos C (2010) Helminths found in marmosets (*Callithrix penicillata* and *Callithrix jacchus*) introduced to the region of occurrence of golden lion tamarins (*Leontopithecus rosalia*) in Brazil. Vet Parasitol 171: 123–129.
18. Radic CP, Rossetti LC, Zuccoli JR, Abelleyro MM, Larripa IB, et al. (2009) Inverse shifting PCR based prenatal diagnosis of hemophilia-causative inversions involving int22h and int1h hotspots from chorionic villus samples. Prenat Diagn 29: 1183–1185.
19. Santacroce R, Acquila M, Belvini D, Castaldo G, Garagiola I, et al. (2008) Identification of 217 unreported mutations in the F8 gene in a group of 1,410 unselected Italian patients with hemophilia A. J Hum Genet 53: 275–284.
20. Kawaji H, Severin J, Lizio M, Waterhouse A, Katayama S, et al. (2009) The FANTOM web resource: from mammalian transcriptional landscape to its dynamic regulation. Genome Biol 10: R40.
21. Machado FB, Alves da Silva AF, Rossetti LC, De Brasi CD, Medina-Acosta E (2011) Informativeness of a novel multiallelic marker-set comprising an F8 intron 21 and three tightly linked loci for haemophilia A carriership analysis. Haemophilia 17: 257–266.
22. Sambrook J, Russell DW (2001) Molecular cloning: a laboratory manual. 3rd ed. Cold Spring Harbor: Cold Spring Harbor Laboratory Press. 999 p.
23. Busque L, Paquette Y, Provost S, Roy DC, Levine RL, et al. (2009) Skewing of X-inactivation ratios in blood cells of aging women is confirmed by independent methodologies. Blood 113: 3472–3474.
24. Sayers EW, Barrett T, Benson DA, Bolton E, Bryant SH, et al. (2011) Database resources of the National Center for Biotechnology Information. Nucleic Acids Res 39: D38–51.
25. Tamura K, Peterson D, Peterson N, Stecher G, Nei M, et al. (2011) MEGA5: molecular evolutionary genetics analysis using maximum likelihood, evolutionary distance, and maximum parsimony methods. Mol Biol Evol 28: 2731–2739.
26. Ross MT, Grafham DV, Coffey AJ, Scherer S, McLay K, et al. (2005) The DNA sequence of the human X chromosome. Nature 434: 325–337.

27. Benson G (1999) Tandem repeats finder: a program to analyze DNA sequences. Nucleic Acids Res 27: 573–580.
28. Bacher J, Schumm JW (1998) Development of highly polymorphic pentanucleotide tandem repeat loci with low stutter. Profiles DNA 2: 3–6.
29. Machado FB, Medina-Acosta E (2009) High-resolution combined linkage physical map of short tandem repeat loci on human chromosome band Xq28 for indirect haemophilia A carrier detection. Haemophilia 15: 297–308.
30. Machado FB, Duarte LP, Medina-Acosta E (2009) Improved criterion-referenced assessment in indirect tracking of haemophilia A using a 0.23 cM-resolution dense polymorphic marker set. Haemophilia 15: 1135–1142.
31. Schwahn U, Lenzner S, Dong J, Feil S, Hinzmann B, et al. (1998) Positional cloning of the gene for X-linked retinitis pigmentosa 2. Nat Genet 19: 327–332.
32. Sharp AJ, Stathaki E, Migliavacca E, Brahmachary M, Montgomery SB, et al. (2011) DNA methylation profiles of human active and inactive X chromosomes. Genome Res 21: 1592–1600.
33. Kawaji H, Severin J, Lizio M, Forrest AR, van Nimwegen E, et al. (2011) Update of the FANTOM web resource: from mammalian transcriptional landscape to its dynamic regulation. Nucleic Acids Res 39: D856–860.
34. Rossetti LC, Radic CP, Larripa IB, De Brasi CD (2008) Developing a new generation of tests for genotyping hemophilia-causative rearrangements involving int22h and int1h hotspots in the factor VIII gene. J Thromb Haemost 6: 830–836.
35. Sweeney CG, Curran E, Westmoreland SV, Mansfield KG, Vallender EJ (2012) Quantitative molecular assessment of chimerism across tissues in marmosets and tamarins. BMC Genomics 13: 98.
36. Zu T, Gibbens B, Doty NS, Gomes-Pereira M, Huguet A, et al. (2011) Non-ATG-initiated translation directed by microsatellite expansions. Proc Natl Acad Sci U S A 108: 260–265.
37. Ranum LP, Day JW (2002) Dominantly inherited, non-coding microsatellite expansion disorders. Curr Opin Genet Dev 12: 266–271.
38. Ji J, Clegg NJ, Peterson KR, Jackson AL, Laird CD, et al. (1996) In vitro expansion of GGC:GCC repeats: identification of the preferred strand of expansion. Nucleic Acids Res 24: 2835–2840.
39. Shinde D, Lai Y, Sun F, Arnheim N (2003) Taq DNA polymerase slippage mutation rates measured by PCR and quasi-likelihood analysis: (CA/GT)n and (A/T)n microsatellites. Nucleic Acids Res 31: 974–980.
40. Hendriks RW, Chen ZY, Hinds H, Schuurman RK, Craig IW (1992) An X chromosome inactivation assay based on differential methylation of a CpG island coupled to a VNTR polymorphism at the 5' end of the monoamine oxidase A gene. Hum Mol Genet 1: 187–194.
41. Beever C, Lai BP, Baldry SE, Penaherrera MS, Jiang R, et al. (2003) Methylation of ZNF261 as an assay for determining X chromosome inactivation patterns. Am J Med Genet A 120A: 439–441.
42. Bertelsen B, Tumer Z, Ravn K (2011) Three new loci for determining x chromosome inactivation patterns. J Mol Diagn 13: 537–540.
43. Batra R, Charizanis K, Swanson MS (2010) Partners in crime: bidirectional transcription in unstable microsatellite disease. Hum Mol Genet 19: R77–82.
44. Akalin A, Garrett-Bakelman FE, Kormaksson M, Busuttil J, Zhang L, et al. (2012) Base-pair resolution DNA methylation sequencing reveals profoundly divergent epigenetic landscapes in acute myeloid leukemia. PLoS Genet 8: e1002781.
45. Mossner M, Nolte F, Hutter G, Reins J, Klaumunzer M, et al. (2013) Skewed X-inactivation patterns in ageing healthy and myelodysplastic haematopoiesis determined by a pyrosequencing based transcriptional clonality assay. J Med Genet 50: 108–117.
46. Choong CS, Kemppainen JA, Wilson EM (1998) Evolution of the primate androgen receptor: a structural basis for disease. J Mol Evol 47: 334–342.
47. Ohadi M, Mohammadparast S, Darvish H (2012) Evolutionary trend of exceptionally long human core promoter short tandem repeats. Gene 507: 61–67.
48. Busque L, Mio R, Mattioli J, Brais E, Blais N, et al. (1996) Nonrandom X-inactivation patterns in normal females: lyonization ratios vary with age. Blood 88: 59–65.
49. Kent WJ, Sugnet CW, Furey TS, Roskin KM, Pringle TH, et al. (2002) The human genome browser at UCSC. Genome Res 12: 996–1006.
50. Tamura K, Nei M (1993) Estimation of the number of nucleotide substitutions in the control region of mitochondrial DNA in humans and chimpanzees. Mol Biol Evol 10: 512–526.
51. Felsenstein J (1985) Confidence limits on phylogenies: An approach using the bootstrap. Evolution 39: 783–791.