Blood Cells, Molecules and Diseases 54 (2015) 210–216

Contents lists available at ScienceDirect

Blood Cells, Molecules and Diseases

journal homepage: www.elsevier.com/locate/bcmd

Identification of novel informative loci for DNA-based X-inactivation analysis

Dita Musalkova, Jakub Minks, Gabriela Storkanova, Lenka Dvorakova, Martin Hrebicek *

Institute of Inherited Metabolic Disorders, First Faculty of Medicine, Charles University in Prague and General University Hospital in Prague, Prague, Czech Republic

ARTICLE INFO

Article history:
Revised 23 February 2014
Accepted 23 April 2014
Available online 20 November 2014

(Communicated by J.T. Prchal, M.D.,
23 April 2014)

Keywords:
X-chromosome inactivation
Skewing
DNA methylation
Androgen receptor
Trinucleotide repeats
Blood cells

ABSTRACT

The HUMARA assay, the most common method for evaluation of X-inactivation skewing in blood cells, has been reported to be usable in only about 80% of females, emphasizing the need for alternative methods for testing of HUMARA-uninformative individuals. We conducted an in silico search for potentially polymorphic tri-to-hexanucleotide repeats in the proximity of CpG islands located in 5′ regions of X-chromosome genes to design five candidate assays (numbered I, II, III, IV, and V) combining methylation-specific restriction digest with PCR amplification in a manner similar to the HUMARA assay. The results obtained by these assays in 100 healthy females were compared to X-inactivation skewing measured by the AR-MSP method which is based on methylation-specific PCR amplification of the first exon of the *AR* gene. On the basis of statistical evidence, three of the novel assays (II, IV, and V), which were informative in 18%, 61%, and 55% of females in the cohort, respectively, may be used as alternatives or conjointly with the HUMARA assay to improve its reliability. The three new assays were combined with the HUMARA assay into a novel X-inactivation test leading to the increase of informative females in the cohort from 67% to 96%.

Introduction

In placental mammals, the dosage compensation that equalizes the expression of X-linked genes between sexes is achieved by inactivation of one of the X chromosomes in female cells [1]. This process occurs early in the embryonic development and is normally random with the paternal and the maternal X chromosomes having the same chance of being inactivated. The inactive state of the X chromosome is maintained by epigenetic regulation and is stably inherited through the subsequent cell divisions. As a result, females are mosaics of two cell populations with different active X chromosomes. Their size is usually equal but in some women the ratio may significantly differ from the 50:50 average and this deviation is referred to as skewing of X-inactivation. The reason for this phenomenon may be solely stochastic processes or genetic differences that influence the course of X-inactivation. However, the most common basis of significant skewing is the secondary cell selection. X-chromosome inactivation (XCI) patterns are most often tested in DNA isolated from nucleated blood cells, apparently because of the ease with which they can be obtained. Determination of the XCI pattern is important in a number of disorders affecting blood cells because skewing is known to affect the clinical manifestation of these diseases, including hemophilia A [2–8] and B [9, 10], Wiskott–Aldrich syndrome [11–13], G6PD deficiency [14], sideroblastic anemia [15, 16], etc. X-inactivation studies are also used to determine the clonality of hematological malignancies [17, 18] and they can be helpful when identifying carrier females of e.g. Fanconi anemia B or X-SCID because of extremely skewed X-inactivation as a result of selective pressure against cells carrying the mutation on the active X chromosome [19–21]. While X-inactivation skewing in affected solid organs and tissues is frequently just extrapolated from blood cells, for disorders affecting hematopoiesis the analysis in blood cells is directly relevant.

Assays currently used for determining X-inactivation patterns are indirect and distinguish maternal X chromosomes from the paternal by leveraging either single nucleotide polymorphisms (SNP) or much more informative short tandem repeats (STR). The active chromosome (Xa) can be discriminated from the inactive (Xi) due to the differences in DNA methylation and also in gene expression [22]. Nevertheless, methylation-based assays are much more common because DNA is easily accessible and more stable than RNA. The most extensively used method (the HUMARA assay) takes advantage of a highly polymorphic $(CAG)_n$ repeat in the first exon of the human androgen receptor gene (*AR*) located near several methylation-sensitive restriction enzyme sites that are differentially methylated on the active and the inactive X chromosome [23]. A number of similar methods have been devised but their use is limited, mainly because they lack the HUMARA assay's ease of performance, robustness or a highly informative repeat. *MAOA* assay [24] requires an additional enzyme digestion step, *HPRT* assay

* Corresponding author at: Institute of Inherited Metabolic Disorders, First Faculty of Medicine, Charles University in Prague and General University Hospital in Prague, Ke Karlovu 2, 128 08 Prague 2, Czech Republic. Fax: +420 224967119.
E-mail address: martin.hrebicek@lf1.cuni.cz (M. Hrebicek).

http://dx.doi.org/10.1016/j.bcmd.2014.10.001
1079-9796/

[25] is based on the more laborious Southern blot hybridization, *HPRT* and *PGK* assays [26] are based on restriction fragment length polymorphism (RFLP) and thus are less informative, *FMR1* assay [27] can be complicated by differential amplification of alleles and though *M27β* assay [28] uses hypervariable locus, its methylation status is inconsistent. Recently, several new assays targeting *ZNF261* [29], *ZDHHC15*, *SLITRK4* and *PCSK1N* loci [30] emerged but they make use of dinucleotide repeats, which generally leads to higher amount of stutter peaks. The HUMARA method still remains the best option and the *AR* locus was also used to design a new method — AR-MSP [31] involving methylation-specific PCR (MSP) after chemical modification of DNA, which completely omits the use of restriction enzymes.

While the HUMARA assay is well validated and suitable for routine measurement of X-inactivation skewing, it cannot be used for all patients. Uninformative patients form a significant part of the population as the polymorphism in the *AR* locus is informative only in about 80% of females [32]. Obviously, there is a need for alternative methods that allow for an analysis of X inactivation in females when the HUMARA assay cannot be used.

Materials and methods

DNA samples

Genomic DNA was isolated according to standard procedures from lymphocytes separated on Ficoll-Paque cushions (GE Healthcare, Little Chalfont, UK) from peripheral blood of 100 healthy female volunteer donors aged 9 to 72 (mean age 36,5 years; median age 35 years), five adult male donors and eighteen members of five families. All donors were Caucasians from the Czech population. The study was approved by the local ethics committee and conducted in accordance with the Declaration of Helsinki. Informed consent was obtained from all subjects.

Bioinformatic study and primer design

The search for new loci was performed using the Galaxy platform [33–35]. The whole X chromosome (NCBI36/hg18 assembly) was screened for regions that contain a CpG island in close proximity (300 bp maximum) to a simple repeat (located by Tandem Repeats Finder programme) with the repetition length between 3 and 6 bp. Dinucleotide repeats were not included in the search. Regions located in the pseudoautosomal regions PAR1 and PAR2 that escape inactivation [36] were omitted from further processing. Genomic contexts of the other regions were checked using the UCSC genome browser [37] and only sequences lying in the presumed promoter regions were selected for evaluation because, unlike the active X, the CpG islands in the promoter regions of X-linked genes on the inactive X chromosome are generally hypermethylated [38]. As the data about the variability of the selected STR regions were not available in the literature and databases that were at our disposal, the variability was predicted using the SERV software package [39]. Only sequences with VARscore > 0.9 were chosen. Afterward, promoter regions of genes reported to escape X-inactivation [40–42] were excluded. Nineteen loci that fulfilled all the above requirements were selected for further investigation and primers were designed using Primer-BLAST [43]. Suitable methylation-sensitive and other restriction enzymes were found using Webcutter 2.0 (http://www.firstmarket.com/cutter/cut2.html).

Selection of the primers and determination of XCI pattern

Primers were used to amplify 24 female DNA samples to find out if there was variability in the length of the PCR products. Amplification with the five sets of primers (numbered I, II, III, IV, and V, Table 1) generated specific products with variable length as confirmed by amplification of an independent group of samples obtained from five families (examples in Supplementary Fig. S1).

Samples of genomic DNA isolated from 100 control individuals were PCR amplified using the five sets of primers to identify informative samples which were subsequently used for X-inactivation assays. XCI ratio was calculated as the ratio of the peak area of two alleles of the analyzed polymorphic repeat after the digestion with methylation-sensitive restriction enzyme *Hpa*II or *Hha*I. This intermediate result was corrected by the ratio of the peak area of two alleles in the undigested sample to avoid an error caused by preferential amplification of one of the alleles. The ratio was calculated for informative heterozygotes in which PCR products differed by two or more repeats. In addition, one control male sample was included in every batch of samples to confirm the complete digestion by methylation-sensitive restriction enzyme (Fig. 1, Supplementary Fig. S2).

Specifically, 200 ng of analyzed DNA was digested overnight at 37 °C with 10 U of methylation-sensitive enzyme together with 1–3 U of an auxiliary restrictase (New England Biolabs, Ipswich, MA) in the corresponding NEB buffer and 100 μg/ml BSA if needed (details in Table 1). The same amount of DNA was treated with the auxiliary enzyme alone. This enzyme, which was added to the reaction mix to reduce the complexity of genomic DNA, had restriction sites outside the PCR product and was not sensitive to methylation, digesting thus both active and inactive X alleles equally. The enzymes were inactivated at 95 °C for 10 min and 30 ng of digested DNA was PCR amplified with the designed primers (Applied Biosystems, Foster City, CA). One of the primers in each set was fluorescently labeled with 6-FAM or HEX dye (Table 1). PCR reactions were performed in 10 μl total reaction volume containing 1× PC2 buffer (50 mM Tris–HCl pH 9,1, 16 mM $(NH_4)_2SO_4$, 2,5 mM $MgCl_2$, 150 μg/ml BSA), 200 μM dNTP's, 0,3 μM primers, 10% DMSO, 240 mU of Klentaq1 (DNA Polymerase Technology, St. Louis, MO) and 2,2 mU of Deep Vent polymerase (New England Biolabs). The DNA was amplified in C1000 thermocycler (Bio-Rad Laboratories, Hercules, CA). Denaturation (94 °C for 2 min) was followed by 35 cycles consisting of 94 °C for 5 s and 68 °C for 20–25 s and by a final extension

Table 1
PCR primers and enzymes used in individual assays.

Set	Gene	Repeat	Primers	Restriction enzymes used	Length of the product (observed)
I	*RPS6KA3*	CGG/GCGGCA[a]	5′-6-FAM-CAC AGC CAT CTT CTG CCA CGG G-3′ 5′-AAG AAA GGG GCG AGA CCC GGT-3′	*Hpa*II + *Rsa*I	~305–311
II	*CNKSR2*	GCA	5′-6-FAM-CGA GCG GGC AAG TTG GCT GA-3′ 5′-TGT CGG GTC TCG CGG CTG TA-3′	*Hpa*II + *Rsa*I	~278–308
III	*RPS6KA6*	CCG	5′-HEX-AGG GCT CGT CCT GAG GAG CG-3′ 5′-CGA GCG GCT GTC GTT GTG GT-3′	*Hha*I + *Rsa*I	~306–324
IV	*TMEM185A*[b], *FAM11A*	CCG	5′-HEX-GGC CCC TCA GGT TCA TGG CG-3′ 5′-CCC TCG TAC GGA AGC CCG GA-3′	*Hpa*II + *Hpy*CH4III	~243–279
V	*HMGB3, DKFZp779G118*	CCG	5′-6-FAM-GTG GAG GCA GCT AGC GCG AG-3′ 5′-GCT CGG GGA ACG CGT TGG AA-3′	*Hpa*II + *Hpy*CH4III	~286–322

[a] The CGG repeat region includes a smaller GCGGCA repeat region.
[b] This gene is known for association with the CpG island of the fragile site FRAXF.