

**UNIVERSIDAD DE BUENOS AIRES**  
**Facultad de Farmacia y Bioquímica.**



***“ANÁLISIS DE POLIMORFISMOS EN LOS GENES DEL CENTRO DE INACTIVACIÓN E  
INACTIVACIÓN SESGADA DEL CROMOSOMA X EN HUMANOS”***

Tesis presentada para optar al título de Magister en Biología Molecular Médica.

EMILCE LILIANA MORA PANTOJA

**Director:** Dr. Carlos Daniel de Brasi (Inv. Independiente CONICET)

**Director Adjunto:** Dra. Pamela Radic (Inv. Asistente CONICET)

**Lugar de trabajo:** Instituto de Medicina Experimental (IMEX),  
CONICET-Academia Nacional de Medicina.

Ciudad de Buenos Aires, 2016.

## RESUMEN

La inactivación del cromosoma X (XCI) es un proceso normal por el cual uno de los dos cromosomas X de las mujeres es inactivado, para compensar la dosis génica con los varones XY. Este proceso de XCI ocurre temprano en el desarrollo y la elección del cromosoma X a inactivar se produce al azar, generando hembras mosaico con un 50% de células con el cromosoma X paterno activo y el otro 50% con el X materno activo. La inactivación del cromosoma X sesgada es una marcada desviación del 50% y entre otras causas puede asociarse a mutaciones sobre el gen *XIST* (*X-inactive specific transcript*) responsable de la iniciación y el mantenimiento de la inactivación. En el centro de inactivación del cromosoma X (XIC), en Xq13, se han descrito genes asociados a ARNs no codificantes (ARNnc) involucrados en el proceso de XCI: *XIST*, *JPX* (*just proximal to XIST*) y *FTX* (*five prime to XIST*). La presencia de variantes tipo SNP (*single nucleotide polymorphism*) en los genes del XIC podrían modificar su nivel de expresión, su función e impactar el balance alélico conduciendo a la XCI sesgada.

Se realizó un estudio molecular exploratorio de tipo Caso/Control (n=22) sobre la eventual asociación de variantes genéticas en los genes del XIC con el patrón de XCI, con 11 Casos con XCI sesgada extrema (XIP>90%) y 11 Controles con inactivación al azar (50%>XIP<55%). En el primer grupo, se incluyeron mujeres portadoras y no-portadoras, sintomáticas y no-sintomáticas de Hemofilia A (HA) y Distrofia Muscular de Duchenne (DMD). Se realizó un *screening* genético del *XIST*, *JPX* y *FTX* mediante la estrategia de *CSGE* (*conformation sensitive gel electrophoresis*) y secuenciación de Sanger, incluyendo las regiones con SNPs de frecuencias alélicas relevantes.

El estudio de un total de 990 amplímeros analizados no mostró deleciones ni mutaciones nuevas, sino sólo 18 variantes alélicas en los 161 SNPs anotados en las bases de datos para los tres genes. Once de los 18 muestran ORs de riesgo (e.g., OR>2,0) indicativos de asociación con XCI sesgada aunque no significativa ( $p>0,05$ ). En *XIST*, rs6527, rs16992443, rs16992436 y rs16992442 mostraron ORs en el rango 3-4; en *FTX*, rs174138, rs68124822, rs146632102 y rs187332478, ORs de 2-6 y en *JPX*, rs68178357, rs55824328 y rs67649459, ORs de 2-3. Los datos estadísticos junto a los estudios bio-informáticos asociados a modificación de los sitios exónicos *enhancer* de *splicing* (ESE) y predictores de sitios de *splicing* sobre los SNPs informativos no mostraron claros indicios de causalidad del fenotipo aunque la mayoría de ellos mapeaban en regiones conservadas filogenéticamente.

Nuestros resultados, aunque no concluyentes, nos estimulan a ampliar las series de Casos y más aún de Controles para aumentar la consistencia estadística del estudio confirmando o desestimando las tendencias observadas. Asimismo, la ausencia de SNPs heterocigotas en algunas regiones del XIC en ciertos Casos nos sugiere la posible asociación con deleciones que podrían asociarse a la abolición de la función de la XCI en *cis* del X afectado.

**Palabras clave:** Inactivación del cromosoma X (XCI), *XIST*, *JPX*, *FTX*, SNP.

## ABSTRACT

X chromosome inactivation (XCI) is a normal process by which one of the two X chromosomes of women is inactivated to compensate the gene dosage with XY males. XCI occurs early in the development and the choice of X to be inactivated is random resulting in mosaic females with 50% of cells with the paternal X active and the other 50% with the maternal X active. Skewed XCI is a marked deviation of 50% and, among other causes, may be associated with mutations on the *XIST* gene (X-inactive specific transcript) responsible for the initiation and maintenance of XCI. On the X chromosome inactivation center (XIC), on Xq13, has been described three genes associated with non-coding RNA (RNAnc) involved in XCI process: *XIST*, *JPX* (just proximal to *XIST*) and *FTX* (five prime to *XIST*). The presence of SNP (single nucleotide polymorphism) variants in the XIC genes may modify their expression level, their function and may impact in the allelic balance leading to skewed XCI.

A molecular exploratory study type "Case/Control" (n=22) was performed to investigate the possible association between genetic variants in XIC genes with the pattern of XCI, on 11 Cases with extreme skewed XCI (XIP > 90%) and 11 Controls with random inactivation (50% > XIP < 55%). In the first group, symptomatic and non-symptomatic carriers, and non-carriers, of Hemophilia A (HA) and Duchenne Muscular Dystrophy (DMD) were included. A genetic screening of *XIST*, *JPX* and *FTX* was performed using screening by CSGE (conformation sensitive gel electrophoresis) and Sanger sequencing, including those regions with SNPs associated with relevant allele frequencies.

The study of a total of 990 amplicons analyzed showed no deletions or new mutations, only 18 allelic variants out of the 161 SNPs annotated in databases for the three genes. Eleven of 18 showed ORs associated with XCI skewing risk (e.g., OR > 2.0) although no statistically significant ( $p > 0.05$ ). *XIST* variants rs6527, rs16992443, rs16992436 y rs16992442 showed ORs ranging 3-4; *FTX*, rs174138, rs68124822, rs146632102 y rs187332478 with ORs of 2-6 and *JPX* variants, rs68178357, rs55824328 y rs67649459, ORs of 2-3. These statistical data together with the bioinformatic studies performed *in silico* to investigate potential modifications in exonic splicing enhancer sites (ESE) and predictors of splice sites on the informative SNPs showed no clear evidence of phenotype causality although most of them mapped to phylogenetically conserved regions.

Our results, although still inconclusive, encourage us to extend the series of Cases and, even more, of controls to increase the statistical consistency of the study either to confirm or rule out the preliminary observed trends. Additionally, the absence of heterozygous SNPs in some regions of XIC genes in some Cases suggests the possible association with large deletions that may abolish their role of XCI of the affected X chromosome *in cis*.

**Keywords:** X chromosome inactivation (XCI), *XIST*, *JPX*, *FTX*, SNP.

## **AGRADECIMIENTOS**

A *DIOS* gracias a quien todo fue posible.....

Al Dr. Carlos y Dra. Pamela, gracias por aceptar ser mi director y directora adjunta, nunca tendré palabras para manifestar mi enorme agradecimiento por su predisposición, paciencia y colaboración, a ustedes y todo el personal del Laboratorio de Genética y Hemofilia, agradezco por su excelencia, sencillez y humildad aun contando con tantos conocimientos, porque además de ser un laboratorio son una familia, gracias por su comprensión y por no negarse a enseñar a quienes tocamos sus puertas, su apoyo y confianza en mi trabajo fueron un aporte invaluable en el desarrollo de esta tesis.

Al Instituto de Medicina Experimental (IMEX), del Consejo Nacional de Investigaciones Científicas y Técnicas, CONICET, y la Academia Nacional de Medicina de Buenos Aires por brindar el lugar físico y los subsidios de investigación para motivar el crecimiento profesional.

A la Cátedra de Genética y Biología Molecular de la UBA por compartir el conocimiento con quienes deseen aprender.... Dra Florencia Giliberto, gracias por facilitar algunas muestras para este estudio.

A mis amistades por acompañar y alentar constantemente el avance de la tesis. A mis compañeros de trabajo sinceros que aún sin interesarles mucho los temas moleculares, siempre preguntaban, escuchaban y me animaban a avanzar....

Finalmente a quienes están en mi corazón y en mi vida: A Johnny; por brindarme su amor y ayuda, gracias por motivar y alentar mi crecimiento profesional....

Un agradecimiento y reconocimiento especial a mi familia: padres, hermanos y sobrinos, por estar y acompañarme siempre en las decisiones, aciertos y desaciertos, gracias por su luz y tenacidad. Por tener paciencia, ustedes con inteligencia y generosidad dan un ejemplo de valentía, capacidad y superación constante....

.....*Por ellos y para ellos....*

## TABLA DE CONTENIDO

1. INTRODUCCIÓN .....	1
2. OBJETIVOS .....	6
3. MATERIALES Y MÉTODOS .....	7
3.1. <i>Poblaciones y muestras.</i> .....	7
3.2. <i>Extracción de muestras de ADN</i> .....	7
3.3. <i>Estimación del patrón de inactivación del cromosoma X (XIP)</i> .....	7
3.4. <i>Amplificación y análisis de genes XIST, FTX y JPX.</i> .....	8
3.5. <i>Screening mediante electroforesis CSGE en búsqueda de polimorfismos.</i> .....	13
3.6. <i>Secuenciación de ADN.</i> .....	14
3.7. <i>Bioinformática.</i> .....	14
3.8. <i>Métodos estadísticos y gráficos</i> .....	15
4. RESULTADOS .....	16
4.1. <i>Estudio de los patrones de inactivación del cromosoma X (XIP). Clasificación de Casos y Controles.</i> .....	16
4.2. <i>Estudio exhaustivo del gen XIST</i> .....	17
4.3. <i>Estudio exhaustivo del gen FTX</i> .....	27
4.4. <i>Estudio exhaustivo del gen JPX</i> .....	33
5. DISCUSIÓN .....	40
6. CONCLUSIÓN .....	46
7. BIBLIOGRAFÍA .....	48

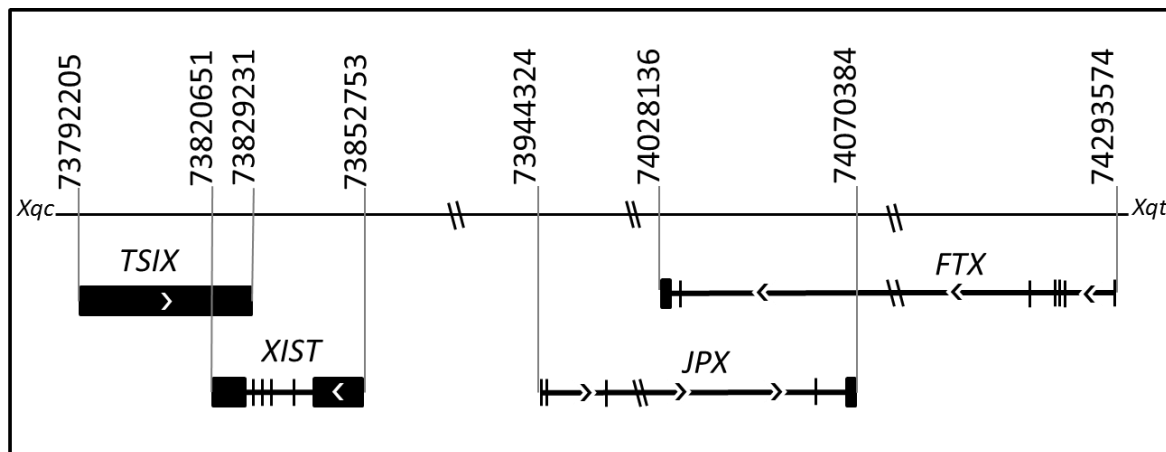
## 1. INTRODUCCIÓN

En el desarrollo embrionario de las hembras de mamíferos, la inactivación del cromosoma X (XCI) es el mecanismo molecular que asegura el silenciamiento masivo de un cromosoma X (X) y, en consecuencia, dosis génicas equivalentes entre los genes ligados al X de hembras (XX) y machos (XY) (Lyon, 1961). El proceso de inactivación se inicia muy temprano en la embriogénesis generando hembras que son mosaicos con dos líneas celulares, una con el cromosoma X materno ( $X_m$ ) inactivo y otra con el X paterno ( $X_p$ ) inactivo, con una distribución teórica promedio del 50% de cada línea en cada órgano o tejido. Una vez ocurrida, en cada célula somática (exceptuando las células germinales), esta inactivación se hereda clonalmente durante toda la vida adulta (Brown y Willard, 1990).

Esta inactivación de base estocástica ocurre independientemente en cada célula del macizo celular interno del estadio *Blastocisto* mediante una serie de eventos que se desencadenan sobre el *locus* XIC (*X-chromosome inactivation centre*) y que determinan con probabilidades generalmente equivalentes ( $P=1/2$ ) cuál X (el  $X_p$  o el  $X_m$ ) será silenciado y cuál permanecerá activo (Ng et al, 2007). Aunque menos frecuentemente, también puede observarse una marcada desviación en la inactivación equilibrada (50%), dando lugar a una inactivación del cromosoma X sesgada. En este caso todas (o casi todas) las células presentarán el mismo X inactivo (sea  $X_m$  o  $X_p$ ). La inactivación sesgada puede estar asociada a distintas causas: (1) aún por azar existen casos donde la inactivación se ubica en los extremos de la distribución binomial, con escasos eventos, ya que número de células progenitoras de cada órgano o tejido, es pequeño; (2) por eventual desventaja del clon del X activo portador en *cis* de una mutación que afecta (total o parcialmente) su viabilidad celular; y (3) por un defecto en el proceso de inactivación en sí mismo, por ejemplo, mutaciones en el gen *XIST* (*X-inactive specific transcript*), que produce un RNA no-codificante de expresión diferencial sobre el X inactivo y es el responsable primario de la iniciación y desarrollo del proceso de inactivación del X (Puck y Willard, 1998).

En la región XIC de humanos se identificaron dos pares de RNAs no codificantes (ncRNA) cuyas secuencias se solapan, en orientación *sense-antisense*, *XIST-TSIX* y *JPX-FTX*

(Kolesnikov y Elisaphenko, 2010) que estarían involucrados en el proceso de inactivación del cromosoma X (Figura 1).



**Figura 1. Esquema de los genes ubicados en el centro de inactivación del cromosoma X de humanos.** *FTX*: Five prime To *XIST*, *JPX*: Just Proximal To *XIST*, *XIST*: X-inactive specific transcript y *TSIX*: RNA Xist antisentido. Xqc: Cromosoma X brazo q centrómero. Xqt: Cromosoma X brazo q telómero. La numeración se refiere a las coordenadas nucleotídicas del cromosoma X versión NC\_000023.11.

Los ncRNA representan una clase diversa de moléculas, que no se traducen a proteínas, aunque son funcionalmente activos y están implicados en la regulación de muchos procesos celulares. Estudios recientes sugieren que la mayor parte del genoma de mamíferos se transcribe a ncRNA, y que la mayoría de ellos genera una alta diversidad de expresión por *splicing* alternativo (Esteller, 2011).

El XIC en humanos, se localiza en el brazo largo del cromosoma X en la banda Xq13.2, expandiéndose en aproximadamente 2.300 Kb, y se ha observado mediante comparación con las secuencias de ratón, una alta homología entre los genes de ambas especies. En el XIC humano se localizan los genes: *XPCT*, *TSX*, *CHIC1*, *CDX4*, *NAPII2*, *CNBP2*; el pseudogen *TSIX*, así como también los ncRNA: *XIST*, *FTX* y *JPX* (Chureau *et al*, 2002; Kolesnikov y Elisaphenko, 2010).

El gen *XIST* fue el primer ncRNA identificado dentro de la región XIC y su función fue ampliamente estudiada. El *XIST* se transcribe orientado hacia el centrómero (Lafreniere *et al*, 1993), se localiza en la posición 73.820.651-73.852.753 (complementaria), se expande

en 32,1 kilo bases (Kb) generando un producto de 19,3 Kb que incluye 6 exones (exones del 1 al 6: 11372, 64, 137, 209, 164 y 7325 pb, respectivamente) (Hong *et al*, 2000). La iniciación del proceso de inactivación del cromosoma X requiere la acumulación en *cis* del ncRNA *XIST*, seguido de cambios epigenéticos en el cromosoma X inactivo asociado (Starmer *et al*, 2009).

La contraparte del *XIST* humano en ratón se denomina *Xist*. En ratones se encuentra descrito un RNA antisentido al *Xist*, el gen *Tsix* que codifica un ncRNA que incluye dentro de su secuencia todo del gen *Xist* en orientación antisentido. *Tsix* fue asociado funcionalmente al X activo, mostrando así una respuesta antagonista a lo observado por *Xist* (Lee, 2000; Sado *et al*, 2001). En humanos, también se ha identificado un transcrito antisentido al *XIST*, *TSIX* (73.792.205-73.829.231) que se expande 37 Kb e incluye parte de la secuencia 3' del gen *XIST* (exones 5 y 6) en orientación antisentido (Figura 1). Migeon y colaboradores (2001) observaron, en humanos, la ausencia de este transcrito en células somáticas adultas y la falta de solapamiento total con el gen *XIST*, sugiriendo la ausencia de función antagonista en la expresión del gen *XIST* en humanos.

*JPX* (*Just Proximal To XIST*, nombre alternativo *ENOX*) se localiza en la posición 73.944.324-74.070.384, transcribe orientado hacia el telómero Xq, se expande 126 Kb y genera un producto de 1,6 Kb que incluye 5 exones (142, 79, 122, 113 y 1216 pb respectivamente). Se ubica 871 Kb río arriba del *XIST* (Figura 1). Se postula que *JPX* participaría en el proceso de inactivación del cromosoma X basados en los resultados observados en su contraparte, u ortólogo, en ratones (*Jpx*) donde se observó que *Jpx* escapa a la inactivación y ejercería su efecto sobre el gen *Xist* tanto en *cis* como en *trans*. La delección total del *Jpx* en heterocigosis disminuye la expresión del alelo no delecionado en más del 50% y observando un menor número de focos del *Xist*, en distintos experimentos coincidentes; mientras que el aumento en la expresión de *Jpx* estaría asociado al aumento en la expresión de *Xist*, por lo que se postula que el gen *Jpx* sería un activador del *Xist* en murinos (Chow *et al*, 2003; Tian *et al*, 2010).

El transcrito *FTX* (*Five prime To XIST*) (NC\_000023.10: 74.028.136-74.293.574, complementaria) está superpuesto y en orientación opuesta al *JPX*, pues transcribe



orientado hacia el centrómero del X, abarca 265,4 Kb y genera un producto de 2,3 Kb que incluye 7 exones (360, 350, 376, 397, 2114, 430 y 611 pb respectivamente). *FTX* presenta regiones de homología entre las especies humana y murina, específicamente en la región 5' del gen, e incluye un grupo de cuatro microRNA (MIR421, MIR374A, MIR374B y MIR545). A pesar de la falta de conocimiento del rol específico de los microRNA (MIR) y *FTX* en el proceso de inactivación, se ha descrito un mayor potencial regulador en humanos que en ratones (Kolesnikov y Elisaphenko, 2010).

Las mujeres bajo la condición de heterocigotas de una enfermedad recesiva ligada al X resultan generalmente no afectadas y por ello son llamadas portadoras. La evidencia fenotípica de enfermedad en desordenes recesivos ligados al cromosoma X son fenómenos infrecuentes. Sin embargo la literatura muestra algunos reportes de mujeres sintomáticas. Por ejemplo, tanto en Hemofilia A (HA) como en Distrofia Muscular de Duchenne (DMD) se han reportado que cierto número importante de las mujeres sintomáticas encontradas resultan heterocigotas para las mutaciones en el gen respectivo, del *F8* o *distrofina* según el caso, asociada a sesgo completo en la inactivación del cromosoma X (Nissen *et al*, 1989; Kling *et al*, 1991; Orstavik *et al*, 1999; Bicocchi *et al*, 2005; Hoffman *et al*, 1992; Brown *et al*, 2000; Giliberto *et al*, 2014).

El laboratorio de Genética Molecular de la Hemofilia (GMH) del IMEX (CONICET-Academia Nacional de Medicina) cuenta con un valioso grupo de muestras de mujeres portadoras de HA (y DMD) todas ellas caracterizadas en cuanto a la mutación causal y confirmada su condición heterocigota, y en la mayoría de los casos, también la medición del grado de inactivación del cromosoma X. Estas mediciones del grado (%) de inactivación del cromosoma X se realizan mediante el sistema HUMARA (*human androgen receptor*, nombre oficial asignado por la *Human Genome Variation Society Nomenclature Committee*, HGVS AR) basado en la metilación del primer exón de AR. Las mujeres portadoras de HA y DMD que resultaban sintomáticas mostraron consistentemente un sesgo extremo en la inactivación del cromosoma X junto a la condición de portadora heterocigota de una mutación en el *F8* o *la distrofina*, respectivamente, lo que permitió estimar una causa razonable para bloquear la producción cuali-cuantitativa de la proteína involucrada en cada caso y, por lo tanto, expresar la enfermedad. Sin embargo, la causa

de la inactivación sesgada del cromosoma X en este grupo de pacientes se plantea como una incógnita. La literatura tampoco revela la causa de la inactivación sesgada del X, salvo aquellos casos donde un rearrreglo cromosómico, del tipo translocación X-autosoma donde el cromosoma X derivado genera severos desbalances en la actividad de genes autosómicos si se inactivara, y por lo tanto no sería viable, resultando en sesgo completo a favor de la inactivación del X normal no involucrado en la translocación.

En este escenario, este trabajo se focaliza en estudiar la región del cromosoma X involucrada en las fases iniciales del proceso de inactivación, específicamente el gen *XIST* y los genes recientemente descritos como reguladores de su expresión, *JPX* y *FTX* (Kolesnikov y Elisaphenko, 2010; Tian *et al*, 2010), en búsqueda de las causas de la inactivación sesgada del X a partir de eventuales deleciones, mutaciones deletéreas, polimorfismos o variantes generadoras de *splicing* alternativo que nos ayuden a entender el proceso molecular asociado al sesgo extremo en la inactivación del cromosoma X en estos grupos de mujeres afectadas.

## 2. OBJETIVOS

El presente proyecto tiene dos tipos de objetivos, el primero, de naturaleza práctica, que consiste en un ajuste técnico experimental, y el segundo, para ayudar a responder la pregunta científica, incluye la aplicación de las técnicas ajustadas y validadas para explorar las causas del sesgo en la inactivación del cromosoma X que condiciona la expresión de HA/DMD en mujeres portadoras heterocigotas analizando exhaustivamente el XIC como región clave para la iniciación, propagación y mantenimiento del estado inactivo.

1. Estimar el valor XIP (X-inactivation pattern, porcentaje de células del linaje mayoritario,  $X_{mi}$  o  $X_{pi}$ ) ( $50\% \leq XIP \leq 100\%$ ) mediante el sistema HUMARA en las mujeres portadoras y no-portadoras para identificar casos ( $XIP > 85\%$ , XIP sesgado) y controles ( $XIP \leq 85\%$ , XIP no sesgado) en muestras de ADN hasta el momento no determinadas para incluirlos en la serie de casos y controles a ser estudiadas.

2. Análisis de mutaciones y polimorfismos (variantes genéticas) sobre regiones exónicas e intrónicas vecinas de los genes *XIST*, *JPX* y *FTX* sobre muestras de DNA de casos y controles, mediante la estrategia de *screening* de mutaciones por CSGE (*conformation sensitive gel electrophoresis*) de alta resolución y secuenciación automática de ADN de Sanger.

3. Análisis estadístico de los resultados obtenidos indicando las probabilidades de asociación entre las variantes encontradas en casos y controles y el estado de inactivación del cromosoma X (sesgado o no sesgado).

### 3. MATERIALES Y MÉTODOS

#### 3.1. Poblaciones y muestras.

La población inicial incluye 61 muestras de mujeres previamente genotipificadas para dos patologías distintas: 54 mujeres relacionadas a pacientes con HA severa (tres portadoras sintomáticas, 35 portadoras y 16 no portadoras) y 7 mujeres relacionadas a pacientes con DMD (todas portadoras sintomáticas). Asimismo, en las muestras que hasta el momento no han sido analizadas se estimará el patrón de inactivación del cromosoma X (XIP).

La población para el estudio molecular de las variantes presentes en los genes *XIST*, *FTX* y *JPX* incluyó todos los casos observados con sesgo extremo en el XIP (Grupo 1) y el mismo número de casos con ausencia de sesgo (Grupo 2).

Grupo 1: Mujeres con sesgo extremo en la inactivación del cromosoma X (XIP>90%) (n=11).

Grupo 2: Mujeres con inactivación al azar del cromosoma X (XIP<60%) (n=11).

#### 3.2. Extracción de muestras de ADN

Las muestras de ADN se obtuvieron a partir de leucocitos de sangre periférica por repetidos lavados con solución de lisis de glóbulos rojos, por extracción por el método de fenol-cloroformo o de *salting out* y precipitación alcohólica (Sambrook *et al*, 1989; Lahiri y Nuremberg, 1991). La calidad y concentración de las muestras de ácidos nucleicos se estimaron por electroforesis en gel de agarosa y espectrofotometría UV (260/280 nm).

#### 3.3. Estimación del patrón de inactivación del cromosoma X (XIP)

El sistema HUMARA es un método que permite analizar el balance en la inactivación del cromosoma X (Allen *et al*, 1992). El ensayo está basado en la digestión con la enzima sensible a metilación *HpaII* (en paralelo con la digestión simulada (*mock*), donde el ADN genómico se trata en iguales condiciones de incubación pero sin agregado de la enzima *HpaII*, seguido análisis por amplificación PCR para investigar el alelo polimórfico asociado al X-inactivo (metilado). La estimación cuantitativa del XIP incluyó el análisis del VNTR

(*variable number of tandem repeats*) (CAG) del exón 1 del gen del receptor de andrógenos humano (*AR*) mediante el análisis de fragmentos por el método de electroforesis capilar con marcación fluorescente del producto obtenido con los *primers* del sistema HUMARA (Tabla 1), a través del servicio de secuenciación de la FCEyN-UBA del que nuestro Laboratorio es parte como consorcista PME.

La evaluación de las secuencias se realizó mediante el programa GeneMarker (V2.2.0). El XIP es calculado mediante la fórmula  $XIP = 100\% - 50\% (A'/A) (A+B) / (A'+B')$ , donde A' y B' son las señales obtenidas con la digestión enzimática *HpaII* (A' es la señal correspondiente al área bajo la curva de mayor densidad fluorescente), y A y B son las señales correspondientes del experimento *mock*.

### **3.4. Amplificación y análisis de genes *XIST*, *FTX* y *JPX*.**

Sobre muestras de ADN genómico de las mujeres con medición del XIP, se estudiaron las secuencias relevantes de los genes *XIST*, *FTX* y *JPX* (promotor, regiones exónicas con SNPs anotados en las bases de datos: dbSNP (URL: [www.ncbi.nlm.nih.gov/SNP/](http://www.ncbi.nlm.nih.gov/SNP/)) y ENSEMBL (URL: [www.ensembl.org/index.html](http://www.ensembl.org/index.html)) y las secuencias asociadas a la maduración del ARN o splicing). Las secuencias fueron amplificadas en reacciones PCR cuyos tamaños moleculares, características y diseño se ajustaban a las características del método de *screening* de mutaciones elegido, CSGE (Hink et al, 1999; Mitchell et al, 2005). En aquellas regiones genómicas donde no fue posible amplificar dentro de los tamaños óptimos para aplicar la técnica de CSGE (250-450 pb) se realizaron PCR-anidadas dentro de cada producto, o fueron tratados con enzimas de restricción reduciendo el tamaño molecular hasta entrar en el rango de detección óptimo de la técnica de *screening*.

El gen *XIST* fue amplificado en 21 amplímeros: el promotor fue estudiado en dos amplímeros; el exón 1, en nueve; el exón 2, en dos; el exón 6, en 4 amplímeros y los exones 3, 4 y 5, en un único amplímero cada uno.

El gen *JPX* fue amplificado en 8 amplímeros: el promotor junto al primer exón se amplificó en un único amplímero, al igual que los exones 2, 3 y 4; y el exón 5 en cambio fue amplificado en cuatro amplímeros.

El gen *FTX* fue amplificado en 17 amplímeros. El promotor junto al exón 1 se amplificó en dos amplímeros; el exón 1 alternativo, los sitios MIR, el exón 2, 5 y 6 en un único

amplímero cada uno; el exón 3 y 4 fueron amplificados juntos en tres reacciones PCR, y el exón 7 fue amplificado en cinco amplímeros.

La Tabla 1 resume la secuencia de los oligonucleótidos *primers* utilizados para estas amplificaciones las temperaturas de *annealing*, los tamaños de cada amplímero y la posición o coordinación de bases de cada oligonucleótido sobre el archivo de secuencias nucleotídicas correspondiente al cromosoma X humano (GenBank NC\_000023.10).

**Tabla 1.** Oligonucleótidos *primers* para análisis de los genes *HUMARA*, *XIST*, *FTX*, y *JPX*.

Nombre	Secuencia Primer 5'>3'	Temperatura Anneling [°C]	Tamaño del producto [bp]	NC_000023.10 ChrX coordenadas nucleotídicas
<b>Sistema AR para la determinación del patrón de inactivación del cromosoma X *</b>				
HUM-F:	FAM-TTCAGAATCTGTTCCAGAGCGTGC	67/65	255-324	66765057-081
HUM-R:	GCTGTGAAGGTTGCTGTTCCCTCAT			66765321-345
<b>XIST</b>				
Pa-F	TGAAGCCAAGGACAGACTC	50	595	73075740-759
Pa-R	TTCAATTCACCTTACTTCCTCCTC			73075164-188
Pb-F	GTCTCAAACCTCCTCACCTC	50	392	73074120-139
Pb-R	TTCCAGAAAAATAATAAAGTA			73073747-768
E1a-F	TTTATTATGCAGTTTATGGAGGAT	50	413	73072847-871
E1a-R	GTGTTGGGGGACTAGAAAATGT			73072458-480
E1b-F	GAACCAACCAAATCACAAAG	50	529	73072649-669
E1b-R	AATATAGAAAAGAGAGTGGAAG			73072222-244
E1c-F	ATTTTTGGTTGACATCTCTG	50	498*	73072268-288
E1c-R	TGCAACAACCCACAAAAC			73071790-808
E1d-F	TCCCCTCTGAACCCCAACAC	55	415*	73072032-053
E1d-R	TGATTCTTACTGCCTCCCGATACA			73071463-487
E1e-F	TATCGGGAGGCAGTAAGAATCATC	55	502*	73071461-485
E1e-R	ATCCAGCACTGTCCATCCCACCTT			73070983-1007
E1f-F	TGAAGGTCTTGCCGCAGTGAAAA	55	561*	73071185-209
E1f-R	AATGGAGGGAGGTTTCAGAC			73070648-667
E1g-F	TGCGTAACAGAGAAGACC	50	378	73068753-771
E1g-R	TAGCGTGCAAATAGGATAC			73068393-412
E1h-F	GCGCATAATTAGCAGTAGT	52	422	73067103-122
E1h-R	ATTGAAGGACCATTGACA			73066746-765
E1i-F	GCTGCCTAGTGGAATGAATAATA	53	371	73061441-465
E1i-R	TGCAGATGAGGTGGGAAGAAG			73061094-115
E2a-F	CCAGTTCCTTAGGTTCCATTTT	51	1494	73057621-642
E2a-R	ATTTAGGGCAGAAGGGTCATT			73056148-169
E2b F	CACACCCAGTTTTTCCAGT	49	329	73057467-487
E2b R	GTCATGAGAAGAACATCAGAC			73057158-179
E3-F	GATACCTATGTCTAATGCTTCAG	50	329	73053309-332
E3-R	TTAAAAAGGAGAGTTGTGG			73053003-023

Nombre	Secuencia Primer 5'>3'	Temperatura Anneling [°C]	Tamaño del producto [bp]	NC_000023.10 ChrX coordenadas nucleotídicas
E4-F E4-R	GAAAAGGGTAAAAATATGGTTGAA ATGCAGAATAATGATGAATACGAC	53	422	73051189-213 73050791-815
E5-F E5-R	AACATTTTCTATTGCTTATTA GGTCATCTTTCCATCTCA	50	396	73049145-166 73048770-788
E6a-F E6a-R	GTAAGAAATCCCTAAGGTAAAAAGT AGAAATAGCAACAAAATGAAAAAT	50	444	73048047-071 73047627-651
E6b-F E6b-R	GGACACATGGAAGGCCTAAGAAGT CTAAAGGTAACCGCAACATCAAA	56	370	73046702-726 73046356-380
E6c-F E6c-R	GATTCTGGGGTTTGGGATTTCTA ATTGGTTTTTCTCTTATGTTTTAC	52	304	73045920-943 73045639-663
E6d-F E6d-R	AGATAACAAGTGGGGGCACAA GGCATTTATTTTTACTCCTTACT	52	493	73043870-891 73043398-422
E6e-F E6e-R	AAAAATCATGACGAAAAGACA AGCAAGCTCACTACAAAACC	51	337	73042417-438 73042101-121
<b>JPX</b>				
E1-F E1-R	CCGGCGGAGGCATTTAG TCCGATATTTTGAAGAGACTGG	56	360	73164078-095 73164415-437
E2-F E2-R	ATAGTATTGTGGTGAGTGATGAG AAAGTTAGGCCCAAGAGTTAGG	56	350	73164291-314 73164461-640
E3-F E3-R	TCATTCTTTACCTTGGTTACTTTG TATTTTACTGCAGTCTATTTCTCCTA	56	376	73214525-549 73214874-900
E4-F E4-R	ATCACCTCCTTTATTGGCTTCTCA GCTTTTGTCTGTTTGGGGTAG	56	397	73281698-721 73282070-094
E5La-F E5La-R	CTAGAGTCTGCACCCTTGTTTAT CTCCACTCCCAGTCTACCCAG	53	2114	73287249-272 73289341-362
E5a-F E5La-R	TTCATCAGTGTTTTATAGTTTTCA CTCCACTCCCAGTCTACCCAG	53	476	73288887-911 73289341-362
E5b-F E5b-R:	CAAGTCCTGGTCTGGGTAG GATTTTATTCCTTTCTCTTCTC	56	430	73289331-350 73289736-760
E5c-F E5c-F:	AGCAATCTCCGTGACTT GAGCAGATTCAGGTTGTTT	56	611	73281698-721 73282070-094
<b>FTX</b>				
E1a-F E1a-R	TATGGGCTGAGGGAGGGATGATGG GCCGCCCTCCATGTTCTGTTCTTC	60	472	73513593-618 73513145-168
E1b-F E1b-R	CTAAGAATTCCTGCTACGACTG AGTACCCAGGATTCACATAAGAC	55	418	73512808-832 73512831-855
E1a1-F E1a1-R	ACTGTATACAAATGGGAGCAAGAA CAAGATGGCCAAGCAAAGAC	55	453	73512584-608 73512135-159
MIR374-F MIR374-R	GTGGGTAGTAGGGTGGGGAGGTTA AGAGACAGGCATATACACAGACAA	60	381	73507466-490 73507109-133

Nombre	Secuencia Primer 5'>3'	Temperatura Anneling [°C]	Tamaño del producto [bp]	NC_000023.10 ChrX coordenadas nucleotídicas
MIR545-F	CATCGGCCATTATAATACAACCTG	55	439	73507168-192 73506753-777
MIR545-R	TATCACTCAAGTCCCACCAAGAAA			
E2-F	ATATAATTGTGATGCTAAGTGT	55	415	73501830-852 73501437-459
E2-R	TAATGTGAAAATATCCAGTAAT			
E3/4a-F	TTGCCCAGGCTGTTCTTGA	53	478	73501158-182 73500704-728
E3/4a-R	GAGACAACACTGTTTTATTTCAC			
E3/4b-F	TGGCTTTGGTGGGGTTCC	55	498	73500737-755 73500257-278
E3/4b-R	TTTGCTGCATGGTCACTCACA			
E3/4c-F	CAATGCAATGGAGGCTGACC	60	279	73500398-418 73500139-162
E3/4c-R	GCTCTGGAAATGGCATCTCTAAG			
E5-F	GTACGAATCTTCTTCCAGA	50	329	73494371-389 73494060-079
E5-R	AGATTGCACCACTACACTCC			
MIR2-F	ACCCAGCACGTGACAAAAA	55	439	73438533-552 73438113-137
MIR2-R	TCAACAAAAGAATTAGAAGACAGG			
E6-F	ACAGCATTAGTCATTCAAGTATC	55	439	73251736-758 73251319-339
E6-R	GGAACCTTCACTAACCAGAT			
E7a-F	GTTTTTGTATGGCTGTTTTGTATT	55	443	73249651-675 73249232-256
E7a-R	GGAGATGACCTGAACCTTATGTGTA			
E7b-F	CACATAAGTTCAGGTCATCTCCAC	55	275	73249230-254 73248979-9003
E7b-R	AAGCTGTCAACCAATTAGGCATAG			
E7c-F	GTAAGGCTAGAACATCCCGAACCT	55	444	73249141-165 73248721-741
E7c-R	CTGAAGAGACCAATAACAAG			
E7d-F	TTTACTTGTATTGGTCTCTTCAG	55	497	73248721-745 73248248-272
E7d-R	GTCTAACAGCGGCACAATCTACAT			
E7e-F	ATGTAGATTGTGCCGCTGTT	55	409	73248252-272 73247863-882
E7e-R	CTGAAAGTAATGGGATGAA			

\*Oligonucleótidos del Sistema *HUMARA* fueron diseñados por Allen et al (1992). Los Oligonucleótidos de *XIST*, *JPX*, *FTX* fueron diseñados para este trabajo.

Las reacciones PCR incluyen 3 µl de ADN genómico (100 ng/µl) en un volumen total de reacción de 25 µl, con 200 µM de cada dNTP (desoxiribonucleótido trifosfato, dATP, dTTP, dGTP, dCTP), 0,5 µM de cada primer, 0,5 U de Taq DNA Polymerase (Promega, Argentina) y buffer Taq DNA Polymerase (Tris-HCl 50 mM pH 9,0, NaCl 50 mM, BSA, seroalbúmina bovina, 0,1 mg/ml).



**Tabla 2.** Detalle de reacciones multiplex para cada gen.

Gen	Multiplex	Exones	Tamaño (Pb)	Temperatura de annealing (°C)
XIST	1	E1h	422	54
		E6e	337	
	2	E5	396	50
		E6c	304	
	3	E3	329	54
		E4	422	
E6d		493		
FTX	1	E1b	418	55
		E3/4b	498	
		MIR2	439	
	2	E7d	497	55
		E1a1	473	
		MIR545	439	
	3	E6	439	55
		E7b	275	
4	E7e	409	55	
	E2	415		
JPX	1	E1	360	56
		E4	397	
		E5c	611	
	2	E2	350	56
		E5b	430	

La identidad y el tamaño de cada amplímero fue analizada por electroforesis en gel de agarosa (1-2%). En este contexto las delecciones grandes (mayores 100 pb) fueron definidas como la ausencia repetida y consistente de un exón o un grupo de amplicones exón-específico vecino.

El termo-ciclado estándar óptimo consta de treinta y cinco ciclos de 94° C, 45 seg.; la temperatura de *annealing* utilizada en cada caso corresponde específicamente a cada par de *primers* (Tabla 1, Tabla 2), 40 seg.; 72° C, 1 min.; precedidos por 94° C, 5 min. y terminados por 5 min. a 72° C.

Para la reacción de PCR Múltiplex de dos y tres amplicómeros se utilizó 4 µl de ADN genómico (100 ng/µl), en las mismas condiciones que la reacción simple (Tabla 2).

Las regiones repetitivas encontradas cerca de los exones 2 del *XIST* y 5 del *JPX* obligaron a incluir una reacción PCR-anidada sobre la reacción de PCR estándar. La primera ronda de PCR genera un producto de 1.494 pb para el exón 2 y 2.114 pb para el exón 5, el termo-ciclado incluyó treinta ciclos de 94° C, 15 seg.; 54° C, 50 seg.; 72°C, 2 min.;

precedidos por 94° C, 3 min. y seguidos por 5 min. a 75° C. La segunda ronda muestra un producto de 329 y 472 pb (exones 2 y 5, respectivamente) con el termociclado estándar descrito previamente.

Aquellos amplímeros cuyos tamaños excedieron los límites de detección de mutaciones de la técnica CSGE (productos >500 pb) fueron sometidos a digestión por enzimas de restricción específicas para cada amplímero. En el gen *XIST* el amplímero 'Pb' de 595 pb fue digerido con la enzima de restricción *TaqI* generando dos segmentos de 226 y 369 pb, y el amplímero 'E1d' de 590 pb fue digerido con la enzima de restricción *BclI* generando dos fragmentos de 290 y 300 pb. Los amplímeros 'E1b' y 'E1c' de 447 y 498 pb, respectivamente, fueron digeridos con la enzima de restricción *HinfI*, generando también en ambos casos dos productos de digestión ('E1b': 67 y 380 pb y 'E1c': 267 y 231 pb). Los amplímeros 'E1e' y 'E1f' de 502 y 561 pb, respectivamente, fueron digeridos con la *TruI* generando cinco productos de digestión en el amplímero 'E1e' (47, 38, 137, 210 y 70 pb) y dos productos en el amplímero 'E1f' (156 y 343 pb). En todos los casos se emplearon 5 µl del producto de la PCR en un volumen total de reacción de 10 µl, con 10 U de enzima. Las digestiones *TaqI* y *TruI* son incubadas a 65°C por 3 y 1 hora, respectivamente. La enzima *HinfI* fue incubada a 37°C toda la noche y las digestiones con *BclI* fueron incubadas a 55°C por dos horas. La identidad y el tamaño de cada amplímero fue evaluada por electroforesis en gel de agarosa (1-1.5%).

### **3.5. Screening mediante electroforesis CSGE en búsqueda de polimorfismos.**

Todas las muestras fueron sometidos a electroforesis por CSGE en forma similar (descrito por Williams *et al*, 1998). Cinco µl de los amplímeros de las muestras de los pacientes y controles fueron procesados para formar los heterodúplex de ADN (con el fin de liberar las cadenas del ADN de los alelos paternos y maternos, y re-hibridarlos para obtener una mezcla combinatoria de cadenas de los alelos presentes generando especies homodúplex y heterodúplex), para lo cual fueron mantenidos a 96° C por 5 minutos seguido de 30 minutos a 64° C, para ser luego analizados electroforéticamente en corridas *CSGE-PAGE* (*polyacrylamide gel electrophoresis*) por 16-18 hs a una diferencia de potencial constante de 400 V, en geles de poliacrilamida al 12% (con entrecruzamiento de 99:1 acrilamida:bisacrilamida) en condiciones medianamente desnaturizantes (10% etilen

glicol, 15% formamida deionizada) en cubas de alta resolución (41 x 33 x 0,08 cm). La lectura de resultados del monitoreo por CSGE se realizó por coloración con plata coloidal. En la tinción con plata, el gel es lavado con agua destilada, sumergido por 10 minutos en fijador (10% de etanol absoluto y 0,5% ácido acético), mantenido en una solución de plata al 0,1% por 15 minutos en agitación, lavado con agua destilada, y revelado hasta la aparición de las señales con la solución de revelado (1,5% hidróxido de sodio, 0,28% formaldehído).

### **3.6. Secuenciación de ADN.**

El análisis por CSGE de las muestras se realiza en paralelo con un control normal (muestra amplificada de un control varón 46,XY que presenta un único alelo, hemicigota), para tomarlo como patrón normal de corrida y sobre la base de la comparación minuciosa de los patrones CSGE de muestras y controles, se seleccionan las muestras cuyos amplímeros mostraron patrones con movilidad CSGE diferencial para ser enviados a caracterizar por secuenciación de Sanger.

Los productos PCR de las muestras a secuenciar fueron amplificados y resueltos por electroforesis en gel de agarosa preparativa, purificados usando cromatografía en minicolumnas GFX<sup>TM</sup> *Spin columns* (Amersham, Argentina), cuantificados en geles de agarosa (1,5%) y secuenciados por el método de Sanger automático con marcación fluorescente en el Servicio de Secuenciación de ADN de la FCEyN-UBA. El análisis de las lecturas de las secuencias obtenidas se realizó por comparación (alineamiento) con las secuencias de referencia.

### **3.7. Bioinformática.**

La anotación de SNPs en las secuencias genómicas en los genes *XIST*, *FTX* y *JPX*, los mapas de restricción, alineamientos y manipulación *in silico* de las secuencias, fueron realizados mediante el paquete de programas *DNA Star Lasergene* (SeqBuilder y MegAlign).

El diseño de *primers* se realizó bajo la asistencia del programa PrimerSelect (DNA Star Lasergene). El estudio de los sitios consenso de *splicing* se realizó en-línea usando Splice-

Site Prediction by Neural Network (NNpredict) y NetGene2

([www.fruitfly.org/seq\\_tools/splice.html](http://www.fruitfly.org/seq_tools/splice.html); [www.cbs.dtu.dk/services/NetGene2](http://www.cbs.dtu.dk/services/NetGene2)).

El estudio de los sitios *enhancer* exónicos, sitios ESEs (Exonic Splicing Enhancers/Silencers), se realizó en-línea usando ESE Finder ([rulai.cshl.edu/cgi-bin/tools/ESE3/esefinder.cgi?process=home](http://rulai.cshl.edu/cgi-bin/tools/ESE3/esefinder.cgi?process=home)) y RESCUE-ESE ([genes.mit.edu/burgelab/rescue-ese/](http://genes.mit.edu/burgelab/rescue-ese/)).

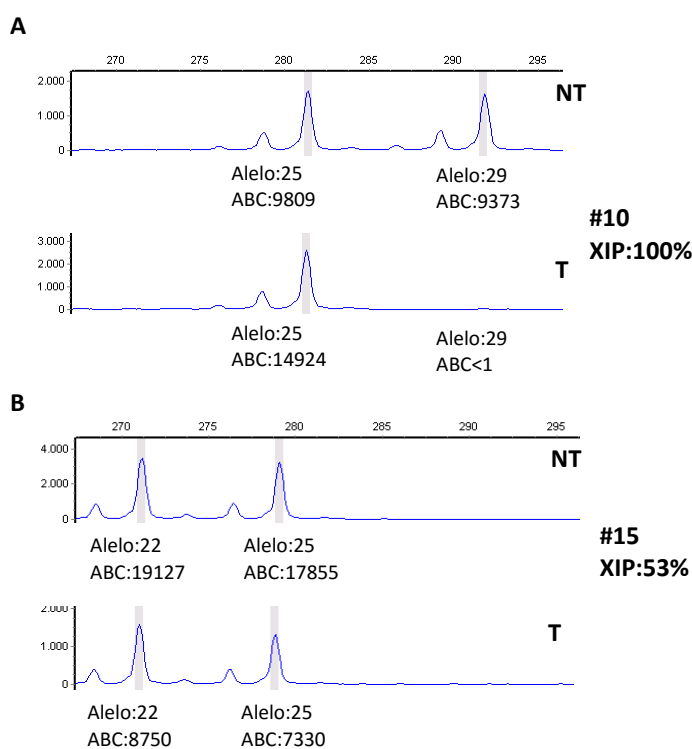
### **3.8. Métodos estadísticos y gráficos**

Fue utilizado el paquete estadístico y gráfico Graphpad Prism 5.0 y Excel (MS Office). Las tablas de contingencia de 2x2 (2 estados (caso/control) y 2 exposiciones (SNP+/SNP-)) fueron analizadas por la prueba exacta de Fisher, calculando cada OR (*odds ratio*) con su intervalo de confianza (95%).

## 4. RESULTADOS

### 4.1. Estudio de los patrones de inactivación del cromosoma X (XIP). Clasificación de Casos y Controles

El análisis del patrón XIP en mujeres 46,XX mostró 11 casos con sesgo extremo en la inactivación del cromosoma X ( $XIP > 90\%$ ), 9 casos con sesgo moderado ( $90\% \geq XIP > 75\%$ ) y 41 casos sin sesgo ( $XIP \leq 75\%$ ) (Figura 2).



**Figura 2: Estudio de Inactivación del cromosoma X.** Electroferograma mostrando el análisis de dos casos. Se muestran las señales T y NT, obtenidas con el tratamiento (Tratado) con la enzima de restricción sensible a metilación *HpaII* y sin *HpaII* (No Tratado). (A) Caso #10, la desaparición de la señal de un alelo cuando la muestra es tratada con la enzima de restricción indica sesgo extremo en la inactivación del cromosoma X y XIP de 100%. (B) Caso #15, las señales invariables cuando se trata o no con la enzima de restricción indican ausencia de sesgo y XIP del 53%. Los números 22, 25, etc., asociados al Alelo indica el número de repeticiones (n) del motivo CAGn, mientras que ABC indica área bajo la curva en unidades arbitrarias (e.g., 7330).

La población para el estudio molecular de los polimorfismos presentes en los genes *XIST*, *FTX* y *JPX* incluirá todos los casos observados con sesgo extremo en el cromosoma X y el mismo número de casos con ausencia de sesgo.

**Grupo 1:** Once mujeres con sesgo extremo en la inactivación del cromosoma X ( $90\% < XIP \leq 100\%$ ). En este grupo se incluyen tres mujeres con síntomas de HA severa, casos #10, #384 y #677 con actividad del FVIII:C  $< 1$  IU/dL, dos portadoras de HA severa asintomáticas, casos #14 y #368 (FVIII:C  $> 20$  IU/dL), una mujer no portadora, caso #444 y cinco portadoras sintomáticas de Distrofia Muscular de Duchenne, casos #287, #334, #482, #499 y #514 (Tabla 3).

**Grupo 2:** Once mujeres con inactivación al azar, no sesgada, del cromosoma X ( $50\% \leq XIP \leq 55\%$ ). En este grupo se incluyen cinco portadoras asintomáticas de HA severa, casos #15, #265, #277, #446 y #601, cinco no portadoras, casos #233, #275, #373, #485 y #690, y una portadoras sintomáticas de Distrofia Muscular de Duchenne, caso #60 (Tabla 3).

**Tabla 3.** Características de las muestras incluidas en el trabajo.

Casos (Grupo 1)			Controles (Grupo 2)		
Caso #	XIP	Condición	Caso #	XIP	Condición
10	100,00	HA sintomática	15	52,79	HA portadora
14	91,10	HA portadora	D60	51,96	DMD sintomática
D287	94,44	DMD sintomática	233	50,51	No portadora
D334	100,00	DMD sintomática	265	55,05	HA portadora
368	100,00	HA portadora	275	52,52	No portadora
384	100,00	HA sintomática	277	52,92	HA portadora
444	99,9	No portadora	373	52,72	No portadora
D482	100,00	DMD sintomática	446	55,70	HA portadora
D499	97,03	DMD sintomática	485	54,82	No portadora
D514	89,36	DMD sintomática	601	55,09	HA portadora
677	97,26	HA sintomática	690	54,48	No portadora

#### 4.2. Estudio exhaustivo del gen XIST

Se realizó un *screening* genético exhaustivo del *XIST* por CSGE y secuenciación de Sanger. Se analizó la región regulatoria del *XIST* en dos amplímeros, los sitios de *splicing* asociados y los exones 2, 3, 4 y 5 completos (64, 137, 209 y 164 pb, respectivamente) cada uno mediante un amplímero, y finalmente se amplificaron parcialmente los exones 1

(11.372 pb) y 6 (7.325 pb) incluyendo sólo las regiones con SNPs de frecuencias alélicas relevantes mediante 9 y 5 amplímeros, respectivamente.

En primera instancia se determinó mediante el análisis de la base de datos de polimorfismos (SNP database) el número de SNPs anotados en las regiones amplificadas del *XIST* determinando 16 SNPs en la región promotora, 27 en el primer exón, 2 en el exón 3, 3 en el exón 4, 3 en el exón 5 y 29 en el exón 6. En la tabla 4 se detalla cada amplímero y los SNPs esperados en cada caso, así como la posición en el genoma y su secuencia.

**Tabla 4.** Detalle de los SNPs anotados en base de datos públicas y fueron estudiados.

Región	Amplímero	Nombre del SNP	NC_000023.11 g.	Secuencia
Promotora	Pa	rs187804319	73855831	TACAAGGTCA G TATCAGAAGA
		rs182678599	73855779	TGTATACAGG G CAGGACCAGA
		rs192169037	73855646	TATGTTTCATC T GTACATCCTC
		rs186482811	73855644	ACTATGTTCA T CTGTACATCC
		rs60717989	73855552	CTCTTCTTTG G GCTCCCAATC
		rs182891446	73855494	TCTGATGGGG A TATTCAGGC
		rs144257078	73855489	TGGGCTCTGA T GGGGATATTT
		rs190652440	73855431	GAAACTGAGA C ACTAGAAGGA
		rs35066732	73855421	CAGAAGTATG G AAAGTACAGC
		rs35360226	73855380	CTTCACTACT G CTTGGCTCCA
	Pb	rs189200366	73854265	TTTGGGAGGC C GAGGTGGCTG
		rs184714697	73854226	TTGGCCGGGC G CGGTGGCTCA
		rs142946283	73854179	GAAGTTTCTA C AAGTGTGGAA
		rs192681706	73854112	GTGACACCTA T GAAGCAGCTA
		rs188890828	73854099	GAAACTACCT A CAGTGACACC
		rs138444431	73854063	ATGGTACCTT T GAGAAAACAA
Exon 1	E1a	rs192366030	73852984	GGGACATTGA G TAAATCAAGC
	E1a/b	rs187705242	73852763	GGGGGGCTGA G GGCGGAGAGA
		rs183361548	73852726	GCTGCAGCAG C GAATTGCAGC
	E1b	rs191198327	73852649	AATGTTCTAG A AAGAACCCCA
		rs142826759	73852511	CAAATACTTT C TTTAAAAAAA
	E1c	rs187477685	73852499	CTCAAAATAT T CCAAATACTT
		rs183563829	73852331	TATCCACGGC C CCGTTGGGCA
	E1c/d	rs191320229	73852250	CCCCGATGGG C TAAGGAAAAA
		rs186435780	73852113	AATAAAAAAA A AAAAAAGCAGG
		rs36033063	73852112	AAATAAAAAA A AAAAAAGCAG
		rs146029769	73852112	AAATAAAAAA A AAAAAAGCAG
		rs72462994	73852041	GTATCCGCGG C CCGATGGGC
	E1d	rs180675156	73851978	AAACCAACAT T TTTTCATCCA
		rs112864750	73851809	AGTTATGCGG C AAGTCTAAAA
rs138241042		73851795	CTAGCCCTAA G CCGAGTTATG	
rs150561547		73851725	TGCACATGAC T TCCTCTGCCT	

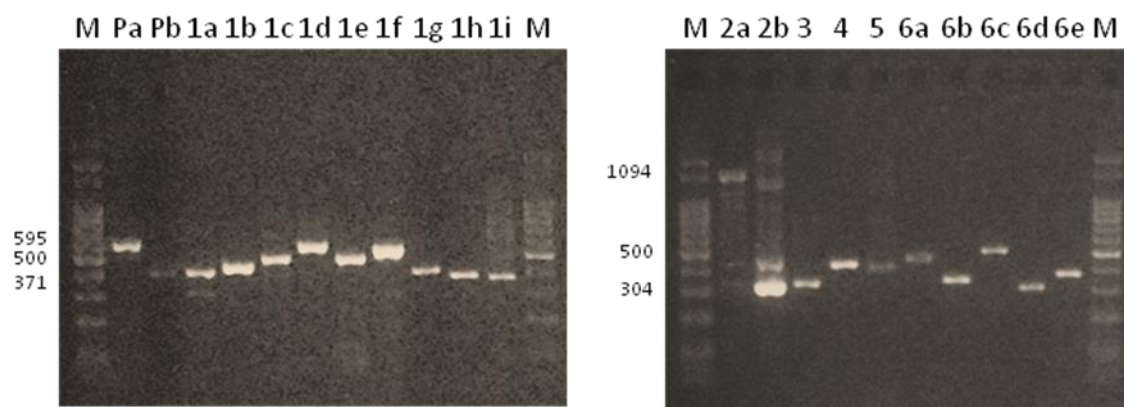
Región	Amplímero	Nombre del SNP	NC_000023.11 g.	Secuencia	
		rs36050632	73851306	GCGGCAAAC C CAACACGAAA	
	E1e/f	rs41305409	73851035	GCTAATTCAC C CAGGTCTTCG	
	E1f	rs189693729	73848879	GGTAGTTCAC A CTATCTAGGA	
	E1g		rs191943671	73848825	CTAATGAGTG G GACTATGGGC
			rs186759648	73848738	TATATGCAAT T ATGCATATCA
			rs6528	73848691	AAGGTGGGAT G GGGTGAGAAT
			rs73486576	73848596	TCTAATGGAC A GGA CTCTGGA
			rs183920959	73847259	AAGGAAGCGG G ATTCTACTCT
	E1h		rs6527	73847056	GTATATAACA C CTAAGATTAT
			rs180923892	73846987	TTAGCACTCT C TGCTTTGATA
		rs41310613	73846987	TTAGCACTCT C TGCTTTGATA	
Exon 3	E3	rs188905149	73833421	GCACCGGTAG A ATACCAAAG	
		rs150466687	73833399	CAGGGTTCAG T AGGCTACCAA	
Exon 4	E4	rs17302743	73831326	TGGGAAAGTC T GTTCTAATGG	
		rs187441106	73831181	CTTTTCCCCT G GAGGATCCTC	
		rs183746781	73831001	CTAAGTGCAA T GTACTGGA	
Exon 5	E5	rs189420347	73829289	TACTGTGCTA G AGTTACATAA	
		rs184904556	73829245	TAAAAAACA T GAGATAAACCC	
		rs138899629	73829014	CAAAAAACAG C TGTTATAACA	
Exon 6	E6a	rs186725726	73828172	ACAATAATAA C AACCACCAAT	
		rs150263261	73827968	AGAAAGAAAG A GACAAAGAAA	
		rs41305399	73827887	AGGCACACAC G AAAGAAATTA	
	E6b	rs34986527	73826794	GGGCCTTCCA T CCTTGGGTTG	
		rs188556573	73826762	TCCTCTATCC A TCTAGGTAGG	
		rs183234357	73826749	TTTTTCGCTT G GGTCTCTAT	
		rs190245821	73826743	GATACCTTTT T CGCTTGGGTC	
		rs16992443	73826722	GATTCCGGCC G TTAGTCTTGA	
		rs151282405	73826721	AGATTCCGGC C GTTAGTCTTG	
	E6c	rs185000623	73825990	CAGTCCACCA A ATTATTTGGC	
		rs16992442	73825974	AGGAGTAGCG T TGGCACAGTC	
		rs180926896	73825972	CCAGGAGTAG C GTTGGCACAG	
		rs12009304	73825966	TTAAACCCAG G AGTAGCGTTG	
		rs41303699	73825910	TATTAACAT G CTCAACAGTC	
		rs1794213	73825905	GAAAGTATTA A ACATGCTCAA	
	E6d	rs1620574	73823982	CTGTAGGCTT T GTAAATCTAC	
		rs141214761	73823963	CATTCCTATC T GTATAGAACT	
		rs148757364	73823958	TAGTTCATTC C TATCTGTATA	
		rs183652815	73823910	CATGTAGTTC C GAGCCCCACA	
		rs144341862	73823899	ATATCCTAGG G CATGTAGTTC	
rs2779095		73823813	CTTGTTAAAT G ACTTTTGGTC		
rs73624269		73823811	AACTTGTTAA A TGACTTTTGG		
rs12835102		73823752	CTCAGCTTAG T AACTGCCTT		
		rs77959471	73823720	ATGTATCTTC T GGACAGGACT	



Región	Amplímero	Nombre del SNP	NC_000023.11 g.	Secuencia
E6e		rs12833040	73823714	GGAAGCATGT A TCTTCTGGAC
		rs138730366	73822494	GCACCCGAAT A TATTTGTAGA
		rs16992436	73822481	CAAGTACCTA T TGGCACCCGA
		rs185674201	73822417	ATTAGGCATT C TCTGGGAAGG
		rs112909669	73822382	ATGCCCTAA C AAGAAAACAA

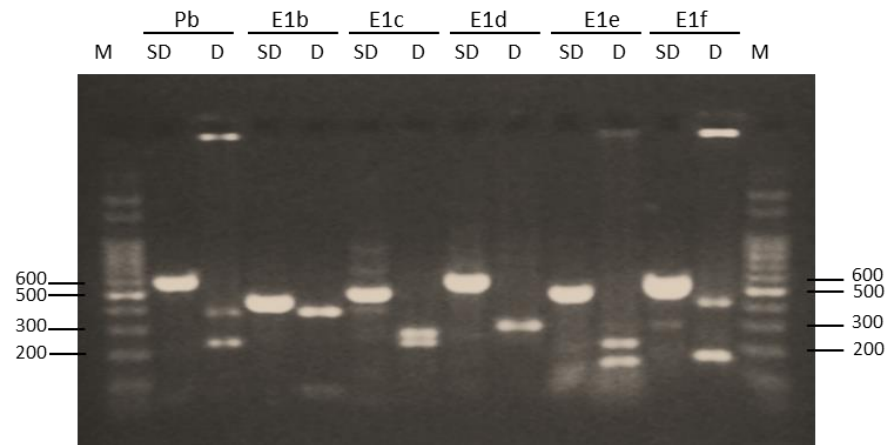
\*Los polimorfismos resaltados en amarillo son los que encontramos en nuestra población de estudio.

La identidad y el tamaño de cada amplímero (Tabla 1) fue analizada por electroforesis en gel de agarosa (Figura 3). En todos los casos y controles el análisis primario mediante electroforesis en gel de agarosa no indico ninguna anomalía aparente, por lo que se realizó el análisis secundario de mutaciones pequeñas por análisis de heterodúplex mediante CSGE-PAGE. Aquellos amplímeros cuyos tamaños exceden los límites recomendados para alcanzar máxima resolución por CSGE ( $\approx 500$  pb) fueron sometidos a digestión con enzimas de restricción para poder ser analizados. En la figura 4 se muestra el esquema digestión de amplímeros.



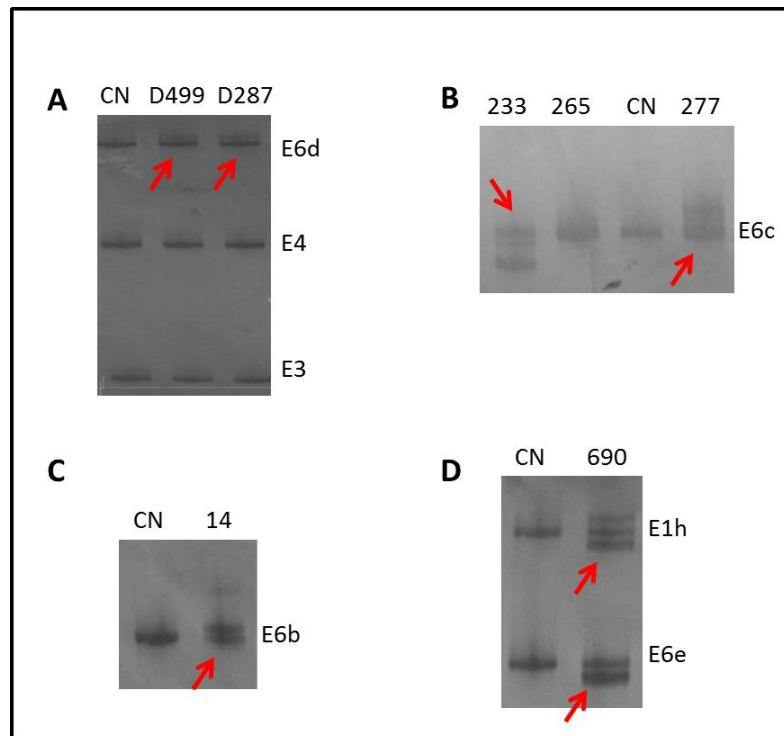
**Figura 3. Amplificación de las regiones clave del gen *XIST*.** Electroforesis en gel de agarosa (1,5%). El panel izquierdo muestra los resultados de los amplímeros correspondientes a la región promotora (Pa-Pb) y al primer exón (1a-1i); en el panel derecho, los amplímeros correspondientes a los exones 2 (2a-2b), 3, 4, 5 y 6 (6a-6e). Se detallan los tamaños de los amplímeros de mayor y menor peso molecular y la señal de mayor intensidad del marcador de peso molecular (500 pb) M indica marcador de peso molecular, escalera de 100 pares de bases.

Amplímero Pb 595 pb ↓ Taql 226/ 369 pb	Amplímero E1b 447 pb ↓ Hinfl 67/ 380 pb	Amplímero E1c 498 pb ↓ Hinfl 231/ 267 pb	Amplímero E1d 590 pb ↓ Bcll 290/ 300 pb	Amplímero E1e 502 pb ↓ Trul 47/38/70/137/ 210 pb	Amplímero E1f 561 pb ↓ Trul 156/ 343 pb
---	--	---	--	--	--



**Figura 4. Esquema de restricción de amplímeros.** En el panel superior se muestra el esquema de digestiones para los amplímeros indicados, se detalla el tamaño del producto antes y después de la digestión con la enzima de restricción correspondiente. En el panel inferior se muestra un gel de agarosa al 1,5% donde se investiga la identidad de cada amplímero antes y después de la digestión, corroborándose los resultados esperados. M: Marcador de peso molecular escalera de 100 pb.

El análisis por CSGE mostró patrones con movilidad diferencial respecto al heteroduplex control de corridas paralelas en los amplímeros: E1c, E1d, E1f, E6b, E6c, E6d y E6e, algunos ejemplos son mostrados en la figura 5.



**Figura 5. Análisis secundario de mutaciones pequeñas incluyendo SNPs por CSGE.** Se muestran las señales específicas observadas en algunos de los casos presentados. Se indica con flecha roja, las señales con movilidad anómala en el heterodúplex. También se muestra en cada gel el patrón del homodúplex control (CN). Cualquier cambio en el patrón respecto del control indica potencialmente la presencia de un cambio en la secuencia de la muestra analizada. (A) Multiplex 3 (Tabla 2) (E6d, E4 y E3) patrones de señal correspondientes a los heterodúplex de dos casos (D499 y D287). (B) Patrones de señal correspondientes al amplímero E6c donde se analizan dos muestras con patrón anómalo (# 233, 277); (C) amplímero E6b, se ejemplifica una muestra con patrón diferencial (# 14); y (D) Multiplex 1 (Tabla 2) (E1h y E6e) la muestra presenta un patrón diferencial en ambos amplímeros (caso # 690).

El análisis de los 440 amplímeros no mostró mutaciones nuevas, sino sólo las 7 variantes alélicas de los 80 SNPs ya anotados en las regiones estudiadas del *XIST* que están resaltados en la tabla 4. El análisis de las variantes encontradas se detalla en la Tabla 5.

Tabla 5. SNP encontrados

SNP	Alelo Ancestral	Frecuencias alélicas	MAF*	CITA
rs1794213*A>C	A	A=0.9052 C=0.0948	C=0.0948/358	Kucera <i>et al</i> , 2011. Bolduc <i>et al</i> , 2008. Chagnon <i>et al</i> , 2005.
rs1620574*T>C	T	T=0.3536 C=0.6464	T=0.3536/1335	Babbs <i>et al</i> , 2011. Bolduc <i>et al</i> , 2008. Chagnon <i>et al</i> , 2005.
rs41305409*C>G	C	C=0.9078 G=0.0922	G=0.0922/348	-
rs6527*C>A	A	C=0.9078 A=0.0922	A=0.0922/348	Chagnon <i>et al</i> , 2005.
rs16992443*G>T	T	G=0.9081 T=0.0919	T=0.0919/347	-
rs16992436*T>C	T	T=0.9081 C=0.0919	C=0.0919/347	-
rs16992442*T>C	T	T=0.9566 C=0.0434	C=0.0434/164	Ananiev <i>et al</i> , 2011.

\*MAF indica la frecuencia del alelo minoritario (*Minor Allele Frequency*) y el número total de alelos estudiados. Las frecuencias alélicas se refieren a las frecuencias observadas en el Proyecto 1000 Genomas (fase 3, URL: <https://www.ncbi.nlm.nih.gov/variation/tools/1000genomes/>).

Las variantes del *XIST* identificadas en nuestra población, fueron evaluadas en su contexto filogenético, para investigar su posible integración en regiones conservadas en distintas especies de homínidos, familia de primates *hominoideos* que incluyen orangutanes, gorilas, chimpancés y bonobos.

La Figura 6 muestra los alineamientos de las regiones relevantes de los genes *XIST* ortólogos. Cinco SNPs rs1794213, rs41305409, rs16992436, rs16992442 y rs1620574 mostraron estar en secuencias altamente conservadas en todas las especies descritas y, en contraste, dos SNPs, identificados como rs6527 y rs16992443, mostraron la prevalencia del alelo que en humanos corresponde a la variante y en las seis especies descritas está altamente conservada (Figura 6).

<p><b>rs1794213*A&gt;C</b></p> <p>Human GAAAGTATTTA<u>A</u>ACATGCTCAA            Chimpanzee GAAAGTATTTA<u>A</u>ACATGCTCAA            Gorilla GGAAGTATTTA<u>A</u>ACGTGCTCAA            Orangutan GAAAATATTTA<u>A</u>ATGTGCTCAA            Vervet-AGM GAAAGTATTTA<u>A</u>ATGTGCTCAA            Macaque GAAAGTATTTA<u>A</u>ATGTGCTCAA            Olive baboon GAAAGTATTTA<u>A</u>ATGTGCTCAA</p>	<p><b>rs41305409*C&gt;G</b></p> <p>Human GCGGCAAAAC<u>C</u>CAACACGAAA            Chimpanzee GCGGCAAAAC<u>C</u>CAACACGAAA            Gorilla GCGGCAAAAC<u>C</u>CAACACGAAA            Orangutan GCGGCAAAAC<u>C</u>CAACACGAAA            Vervet-AGM GCGGCAAAAC<u>C</u>CAACACGAAA            Macaque GCGGCAAAAC<u>C</u>CAACACGAAA            Olive baboon GCGGCAAAAC<u>C</u>CAACACGAAA</p>
<p><b>rs16992436*T&gt;C</b></p> <p>Human CAAGTACCTA<u>T</u>TGGCACCCGA            Chimpanzee CAAGTACCTA<u>T</u>TGGCACCCGA            Gorilla CAAGTACCTA<u>T</u>TGGCACCCAA            Orangutan CAAGTACCTA<u>T</u>TGGCACCCAA            Vervet-AGM CAAGTACCTA<u>T</u>TGGCACCCAA            Macaque CAAGTACCTA<u>T</u>TGGCACCCAA            Olive baboon CAAGTACCTA<u>T</u>TGGCACCCAA</p>	<p><b>rs16992442*T&gt;C</b></p> <p>Human AGGAGTAGGG<u>T</u>TGGCACAGTC            Chimpanzee AGGAGTAGGG<u>T</u>TGGCACAGTC            Gorilla AGGAGTAGGG<u>T</u>TGGCACAGTC            Orangutan AGGAGTAGGG<u>T</u>TGGCACAGTC            Vervet-AGM AGGAGTAGGG<u>T</u>TGGCACAGTC            Macaque AGGAGTAGGG<u>T</u>TGGCACAGTC            Olive baboon AGGACTAGGG<u>T</u>TGGCATAGTC</p>
<p><b>rs1620574*T&gt;C</b></p> <p>Human CTGTAG--//-----GCTT<u>T</u>GTAATCTAC            Chimpanzee CTGTAG--//-----GCTT<u>T</u>GTAATCTAC            Gorilla CTGTAG--//-----GCTT<u>T</u>GTAATCTAC            Orangutan CTGTAG--//-----GCT-<u>T</u>GTAATCTAC            Vervet-AGM CTGTAGGCTTGTAATCAG//CATTCTATTTGTATAGAACTGTAGGCTT<u>T</u>GTAATCTAC            Macaque CTGTAGGCTTGTAATCAA//CATTCTATTTGTATAGAACTGTAGGCTT<u>T</u>GTAATCTAC            Olive baboon CTGTAGGCTTGTAATCAG//CATTCTATTTGTATAGAACTGTAGGCTT<u>T</u>GTAATCTAC</p>	
<p><b>rs6527*C&gt;A</b></p> <p>Human GTATATAACA<u>C</u>CTAAGATTAT            Chimpanzee GTATATAACA<u>A</u>CTAAGATTAT            Gorilla GTATATAACA<u>A</u>CTAAGATTAT            Orangutan GTATATAACA<u>A</u>CTAAGATTAT            Vervet-AGM ATATCTAACA<u>A</u>CTAAGATTAT            Macaque GTATATAACA<u>A</u>CTAAGATTAT            Olive baboon GTATATAACA<u>A</u>CTAAGATTAT</p>	<p><b>rs16992443*G&gt;T</b></p> <p>Human GATTCCGGCC<u>G</u>TTAGTCTTGA            Chimpanzee GATTCCAGCC<u>T</u>TTAGTCTTGA            Gorilla GATTCCGGCC<u>T</u>TTAGTCTTGA            Orangutan GATTCCGGCC<u>T</u>TTAGTCTTGA            Vervet-AGM GATTCCAGCC<u>T</u>TTAGTCTTGA            Macaque GATTCCGGCC<u>T</u>TTAGTCTTGA            Olive baboon GATTCCGGCC<u>T</u>TTAGTCTTGA</p>

**Figura 6. Alineamiento de secuencias ortólogas del *XIST* sobre los 7 SNP detectados en nuestra población bajo estudio.** Alineamientos obtenido en el servidor de análisis de polimorfismos Ensembl (disponible en la URL [www.ensembl.org](http://www.ensembl.org)). En rojo se destaca el nucleótido analizado en cada caso.

Para estudiar el impacto de los SNPs sobre el *splicing* del *XIST* se aplicaron los programas para análisis teórico de secuencias consensos disponibles en internet. Los siete polimorfismos encontrados en el *XIST* se ubican en regiones exónicas, por lo que se aplicaron los programas *ESE Finder* que estima si la variante estudiada aumenta o disminuye el *splicing* que permite la inclusión/exclusión de ese exón particular (*ESE: Exonic Splicing Enhancer*) según su probabilidad de ser parte o de destruir sitios de unión a proteínas SR (*serine arginine-rich protein*), un subgrupo de proteínas caracterizadas por un dominio de unión al RNA rico en serinas y argininas. Asimismo, se estudió la presencia de las variantes sobre secuencias hexanuclóticas con actividad ESE (estimuladoras del

*splicing*) características de exones con *splicing* constitutivo, pero cercanos sitios 5' o 3' débiles, típicos de exones de *splicing* alternativo (Cáceres y Hurst, 2013). Para esta investigación se utilizó la base de datos del programa *RESCUE-ESE* (Tabla 6).

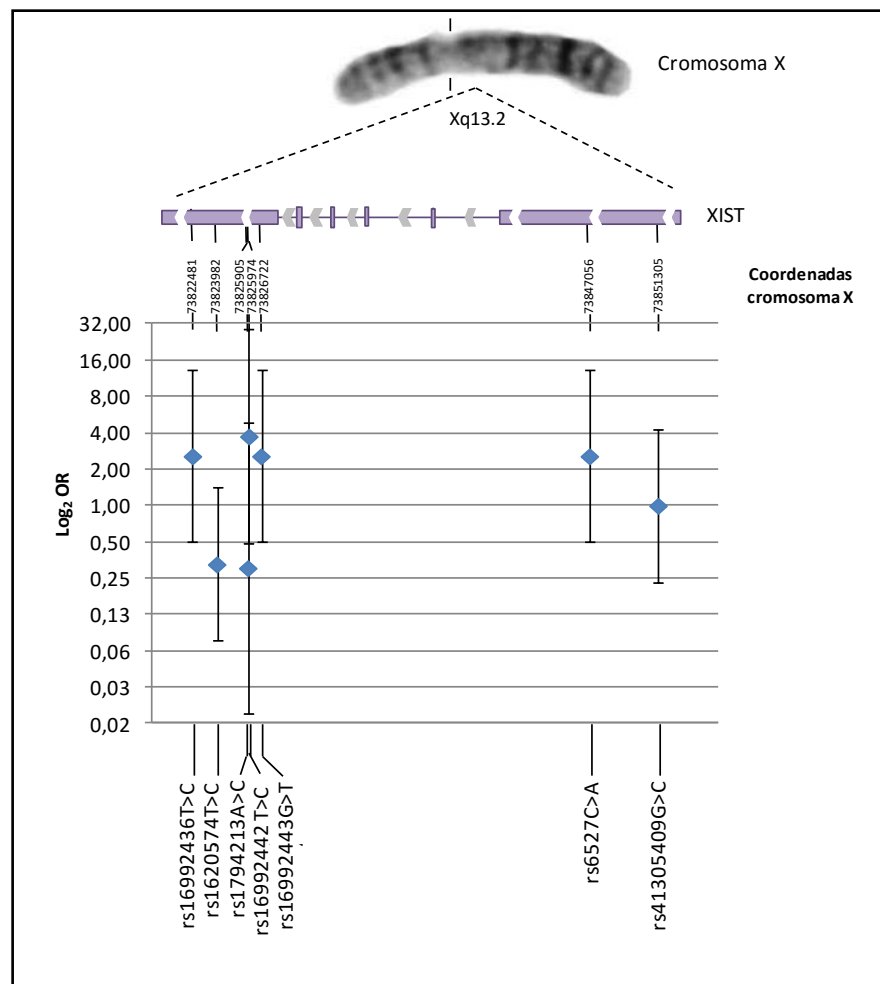
**Tabla 6.** Análisis bioinformático de los SNPs respecto al *splicing* del *XIST*.

SNP #	ESE FINDER		RESCUE-ESE	
	WT	SNP	WT	SNP
rs41305409	ATTCAC <b>C</b> CAGGTC	ATTCAC <b>G</b> CAGGTC	ATTCAC <b>C</b> CAGGTC	ATTCAC <b>G</b> CAGGTC
	*	SRSF1A <i>Score</i> 2.13	Un hexámero ESE	Se corre una base 3' el ESE.
rs6527	ATAACA <b>C</b> CTAAGA	ATAACA <b>A</b> CTAAGA	ATAACA <b>C</b> CTAAGA	ATAACA <b>A</b> CTAAGA
	SRSF5 <i>Score</i> 2.91	*	Un hexámero ESE	Dos nuevos hexámeros ESE
rs16992442	GTAGCG <b>T</b> TGGCAC	GTAGCG <b>C</b> TGGCAC	GTAGCG <b>T</b> TGGCAC	GTAGCG <b>C</b> TGGCAC
	*	SRSF1A <i>Score</i> 2.23	*	*
rs1620574	AGGCTT <b>T</b> GTAAAT	AGGCTT <b>C</b> GTAAAT	AGGCTT <b>T</b> GTAAAT	AGGCTT <b>C</b> GTAAAT
	*	SRSF6 <i>Score</i> 3.47	*	*
rs16992436	TACCTA <b>T</b> TGGCAC	TACCTA <b>C</b> TGGCAC	TACCTA <b>T</b> TGGCAC	TACCTA <b>C</b> TGGCAC
	SRSF5 <i>Score</i> 3.03	SRSF5 <i>Score</i> 5.41	*	*
rs16992443	CCGGCC <b>G</b> TTAGTC	CCGGCC <b>T</b> TTAGTC	CCGGCC <b>G</b> TTAGTC	CCGGCC <b>T</b> TTAGTC
	SRSF1 <i>Score</i> 4.55	SRSF1 <i>Score</i> 2.43	*	*
	SRSF1A <i>Score</i> 4.68	SRSF1A <i>Score</i> 3.14		
SRSF2 <i>Score</i> 3.38	SRSF2 <i>Score</i> 3.51			
rs1794213	GTATTA <b>A</b> ACATGC	GTATTA <b>C</b> ACATGC	GTATTA <b>A</b> ACATGC	GTATTA <b>A</b> ACATGC
	*	SRSF1A <i>Score</i> 2.16 SRSF5 2x <i>Scores</i> 3.25 y 3.25.	Un hexámero ESE.	*

\* Sin sitios de unión a las proteínas SRs. ESE: *exonic splicing enhancer*. Cada proteína SR tiene asociado un umbral de unión definido por un *score* mínimo. SRSF1A (SRSF1 IgM-BRCA1) umbral:1.867; SRSF1, 1.956; SRSF2, 2.383; SRSF5, 2.67 y SRSF6, 2.676.

La función estimuladora sobre el *splicing* que ejercen los sitios ESE está condicionada por la distancia a los extremos exónicos 5' y 3' observándose una mayor efectividad cuando se ubican hasta 100 pb del extremo exón-intrón (Cáceres y Hurst, 2013). En este caso todos los SNP observados en los exones del *XIST* están ubicados a distancias mayores de los extremos (1263-5698 pb) por lo cual es poco probable que los cambios observados en la

Tabla 6, con el análisis por ESE FINDER y RESCUE ESE, tengan algún impacto efectivo modificando la normal estructura de *splicing* del *XIST*.



**Figura 7. Análisis estadístico de SNPs del *XIST* y su asociación al sesgo en la inactivación del cromosoma X.** Bajo el esquema genómico del *XIST* localizado en la banda cromosómica Xq13.2 y orientación hacia el centrómero X, se representan los 6 exones (cajas violetas), y cada uno de los 7 SNPs mostrando su identificación (número rs) y sus coordenadas (ubicación) en el archivo NC\_000023.10 correspondiente al cromosoma X (desde Xp-telómero hasta Xq-telómero).

En el gráfico de probabilidades (en escala logarítmica en base 2) se muestran cada uno de los OR con su IC<sub>90%</sub>

El análisis la distribución de las variantes genéticas del *XIST* encontradas en nuestra población de 'casos' (muestras con XIP sesgado extremo) respecto a los 'controles' (muestras con XIP sin sesgo) no mostró diferencias significativas que permitan concluir una asociación estadística en ninguno de los siete SNPs.

A pesar de ello, la inspección de las tendencias de cada uno de los ORs nos permitió destacar dos SNPs de perfil protector del sesgo en el exón 6, rs1794213 con OR(IC<sub>90%</sub>) de 0.304(0.019-4.890) y rs1620574, con 0.327(0.076-1.404); un SNP neutro en el exón 1, rs41305409, con un OR igual a la unidad virtualmente sobre la hipótesis nula (H<sub>0</sub>) 1.00(0.233-4.299); y cuatro SNPs de perfil asociado a sesgo, uno en el exón 1, rs6527, con 2.57(0.495-13.37) y tres en el exón 6, rs16992443, también con 2.571(0.495-13.37); rs16992436, 2.571(0.495-13.37); y rs16992442, con 3.75(0.48-29.24) (Figura 7).

### 4.3. Estudio exhaustivo del gen *FTX*

Se realizó un monitoreo genético de *FTX* por CSGE y secuenciación de Sanger, sobre 17 amplímeros. El promotor junto al exón 1 se amplificó en dos amplímeros (472 y 418 pb, respectivamente); el exón 1 alternativo, los sitios tres sitios MIR, el exón 2, 5 y 6 en un único amplímero cada uno (con tamaños de 473, 381, 439, 439, 415, 329, 439 pb, respectivamente); el exón 3 y 4 (con tamaños de 155 y 101 pb, respectivamente) fueron amplificados en tres reacciones PCR que se solapan entre si y el exón 7 fue amplificado en cinco amplímeros (de 443, 275, 444, 497 y 409 pb).

El análisis de las bases de datos de SNPs en las regiones amplificadas del *FTX* indicó un total 47 SNPs anotados, cuyas secuencias y posiciones genómicas están detalladas en la Tabla 7.

**Tabla 7.** Detalle de los SNPs de *FTX* anotados en base de datos públicas que fueron estudiados.

Región	Amplímero	Nombre del SNP	NC_000023.11 g.	Secuencia
Exon 1	E1a	rs174136	74293707	GGACGCTATC C CACTCTGCTT
		rs174137	74293640	GCGCCGCTCT C CAGCCTCTAG
		rs190422282	74293632	TCCCCACGC G CCGCTCTCCA
		rs73624295	74293607	CTTAACCCCC C CTCGCGCCCC
		rs192433154	74293505	CACAGGACCA G CTTTGCCTCG
	E1b	rs113272485	74293248	TATTCATTT T TTCCTCCCGT
		rs12847490	74293207	CAAATGCTCA A CATCCTTTTG
		rs174138	74293172	AGGAAAGGAA C CCGCATTCT
		rs142244323	74293146	TGCTCAGGTG C CTTGACTCTA
		rs189393699	74293064	TAAGGCCTCG C CCTTTAAATA

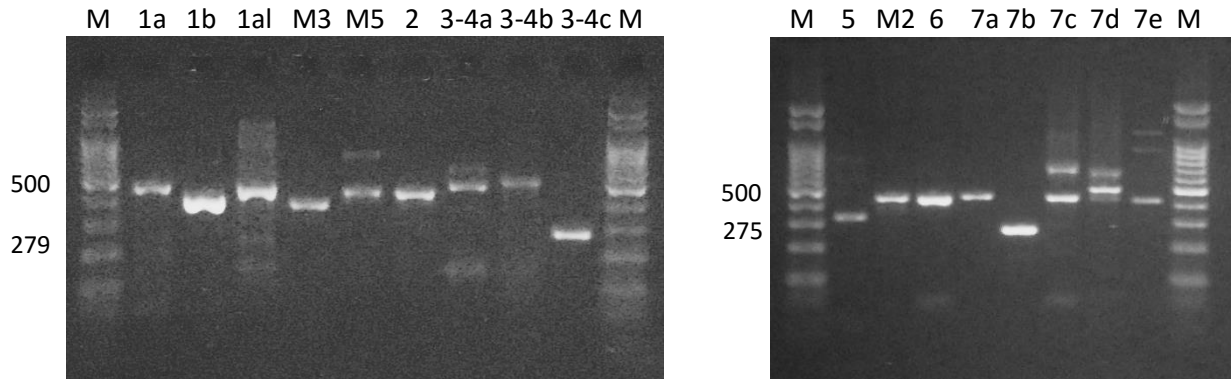


Región	Amplímero	Nombre del SNP	NC_000023.11 g.	Secuencia
Exon 1 Alt	Alt1	rs144911006	74292732	CCTGCGCACC A CGATAGTCCT
MIR 374	MIR374a	rs35911224	74287586	ATCCCCTCCT A TAAAATATGA
		rs188586699	74287455	AAATGCTTTT C GGTTATTTCA
	MIR374b	rs185707904	74287064	ATCAAATTTT C GACCACATGC
Exon 2	E2	rs192819778	74281946	GTCAAGTTAT C CAGCCAGCCA
Exon 3	E3-4a	rs6647468	74281344	CATTTTGTGTC T CACATGGCTT
		rs174152	74281339	AAAAACATTT T TGTCTCACAT
		rs6647467	74281295	ATTTAAAAAA A AGAAAAAGAA
		rs68124822	74281266	GGCAACATAG - TAAGACCCCC
Exon 4	E3-4b	rs191420540	74280865	GGCCAGATCA C GACCAGTTTC
		rs143115067	74280656	AGTTTAACCT A ACCATTTTTA
		rs149089115	74280648	TGTTGCTTAG T TTAACCTAAC
Exon 5	E5	rs187477758	74274420	CCTGAAAGCC A AATTTAGGGT
MIR421	MIR2	rs147961677	74218493	ACAAAGATGA G TAAGGCACTG
		rs191162534	74218340	AGAGCTTGGA C GTTGTGGCA
Exon 6	E6	rs146632102	74031862	AATAACATGA A TAAATTTTGA
		rs144468909	74031746	GTGAACACCC G TGGGAGTTCA
		rs183613416	74031714	GGAAGTTCTG C GCATGCGTGG
		rs181056641	74031568	CAGCCTCTCA C GGAGGCATCT
Exon 7	E7a	rs59818676	74029798	TGATAAAGAA A TTCAGTAAAA
		rs187332478	74029738	AAAAAAAAAA A AAAGAAAGAA
		rs6647427	74029737	AAAAAAAAAA A AAAAGAAAGA
		rs182532788	74029620	AGCTACTCAG G AGGCTGAGGC
		rs5981252	74029472	AATCCAGCA C TTTGGGAGGC
		rs191146271	74029455	CGGTGGTTCA T GCCTGTAATC
	E7b	rs147284851	74029293	CTCTCAGCAC C TCATTCAACC
	E7c	rs188673545	74029150	TAAAAGCTGT C AACCAATTAG
		rs184119881	74029141	TTGATATAAT A AAAGCTGTCA
		rs112479906	74029140	TTTGATATAA - AAAAGCTGTC
		rs113888903	74028959	TAGTGCTTCA C TGTAATTTT
	E7d	rs138861528	74028821	AAAGAGCAGC G TAGAATAAAT
	E7e	rs190351700	74028396	GCAAACGAC G TATACAGAAT
		rs147198639	74028386	GCCAAATAAA G CAACTGACG
		rs12838779	74028331	TGGGCAGATC A TCTGGATCTG
rs185880872		74028127	GCCAGAAACC A TAAAATTTGG	
rs57706478		74028094	AGCTAAACTT - ATTCCATCAA	

\*Los polimorfismos resaltados en amarillo son los que encontramos en nuestra población de estudio.

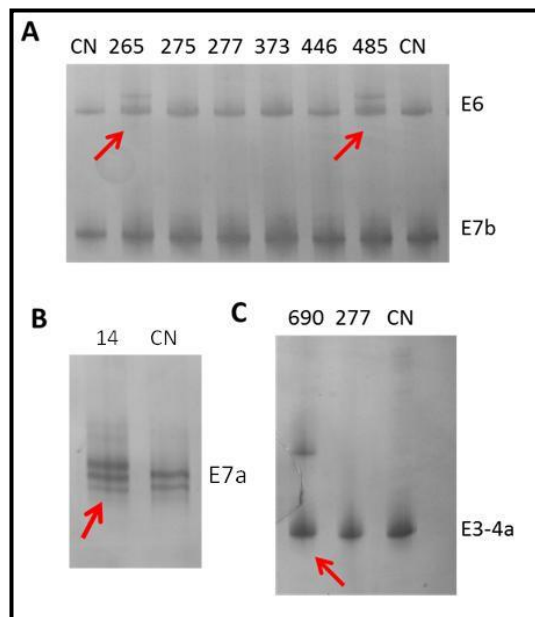
La identidad y el tamaño de cada amplímero (Tabla 1) fue verificada en un primer análisis por electroforesis en gel de agarosa (Figura 8). Mediante este primer análisis todos los casos y controles mostraron amplificación positiva y el tamaño esperado para cada uno,

sin diferencias aparentes, por lo que se realizó el análisis secundario de mutaciones pequeñas por análisis de heterodúplex mediante CSGE-PAGE.



**Figura 8. Amplificación de las regiones clave de *FTX*.** Electroforesis en gel de agarosa (1,5%). El panel izquierdo muestra los resultados de los amplímeros correspondientes a la región del primer exón (1a-1b), exón alternativo 1 (1al), microRNA: MIR374 (M3), MIR545 (M5), exón 2, exones 3 y 4 amplificados juntos en tres amplímeros (3-4a, 3-4b, 3-4c); en el panel derecho, los amplímeros correspondientes a los exones 5, 6, 7 (7a-7e) y microRNA: MIR2 (M2). Se indican los tamaños en pb de los amplímeros de menores pesos moleculares y la señal de mayor intensidad del marcador de peso molecular (500 pb). M indica marcador de peso molecular, escalera de 100 pb.

El análisis por CSGE mostró patrones con movilidad diferencial respecto al heterodúplex control de corridas paralelas en los amplímeros: E1b, E6, E7a, E7c, E3-4a; algunos ejemplos son mostrados en la figura 9.



**Figura 9: Análisis por CSGE de mutaciones pequeñas en *FTX* incluyendo SNPs.** Se muestran señales específicas en algunos de los casos presentados. Se indican con flecha roja las señales con movilidad CSGE anómala en el heterodúplex. También se muestra el patrón homodúplex control (CN). (A) Multiplex 3 (Tabla 2) (E6 y E7b), patrones de señal correspondientes a los heterodúplex de seis casos (#265, #275, #277, #373, #446 y #485). (B) Patrones de señal específicos del amplímero 7a, se muestra el resultado del heterodúplex #14 (C) Patrones de señal correspondientes al amplímero E3-4a donde se analizan dos muestras una de ellas con patrón anómalo (#690); amplímero E7a.

El análisis de los 374 amplímeros no mostró mutaciones nuevas, sino 5 variantes alélicas de las 46 anotadas en las regiones estudiadas de *FTX*: resaltadas en la Tabla 7. El análisis de las variantes encontradas se detalla en la Tabla 8.

**Tabla 8.** SNPs encontrados en *FTX*.

	Alelo Ancestral	Frecuencias alélicas	MAF*	CITA
rs174138*C>T	T	C=0.1168 T=0.8832	C=0.1168/441	-
rs68124822*/-/TT	-	TT=0.1171 - =0.8829	T=0.1171/442	-
rs146632102*A/T	-	A=0.9844 T=0.0156	T=0.0156/59	-
rs187332478*A>G	A	A=0.9097 G=0.0903	G=0.0903/341	-
rs188673545*G>C	C	C=0.9603 G=0.0397	G=0.0397/150	-

\*MAF indica la frecuencia del alelo minoritario (*Minor Allele Frequency*) y el número total de alelos estudiados. Las frecuencias alélicas se refieren a las frecuencias observadas en el Proyecto 1000 Genomas (fase 3, URL <https://www.ncbi.nlm.nih.gov/variation/tools/1000genomes/>).

Las variantes del *FTX* identificadas en nuestra población humana, fueron evaluadas al igual que el *XIST* según su contexto filogenético.

La Figura 10 muestra los alineamientos de las regiones relevantes de los genes ortólogos de *FTX*: el SNP rs188673545 se encuentra en secuencias altamente conservadas en todas las especies descritas y en contraste, los SNPs rs174138 y rs18332478, mostraron alta prevalencia del alelo que en humanos corresponde a la variante minoritaria y en las especies descritas está altamente conservada; en cambio, rs68124822, mostró únicamente el alelo con la variante, que prevalece sólo en humanos ya que en otras especies relacionadas prevalece el alelo ancestral. Para el SNP rs146632102 no se encontraron alineamientos disponibles.

<p><b>rs174138*C&gt;T</b></p> <p>Human AGGAAAGGAA<b>C</b>CCGCATTCT  Chimpanzee AGGAAAGGAA<b>T</b>CCGCATTCT  Gorilla AGGAAAGGAA<b>T</b>CCGCATTCT  Orangutan AGGAAAGGAA<b>T</b>CCGCATTCT  Vervet-AGM AGGTAAGGAG<b>T</b>CCGCATTCT  Macaque AGGTAAGGAG<b>T</b>CCGCATTCT  Olive baboon AGGTAAGGAG<b>T</b>CCGCATTCT</p>	<p><b>rs68124822*-/TT</b></p> <p>Human GGCAACATAGT<b>T</b>TAAAGACCC  Chimpanzee GGCAACATAGT--AAGACCC  Gorilla GGCAACATAGT--AAGACCC  Orangutan GGCAACATAGT--AAGACCC  Vervet-AGM GGCAACATAGT--GAGACCC  Macaque GGCAACATAGT--GAGACCC  Olive baboon GGCAACATAGT--GAGACCC  Marmoset GGCAACACAGT--AAGACCC</p>
<p><b>rs146632102*A/T</b></p> <p>No hay alineamiento disponible</p>	<p><b>rs187332478*A&gt;G</b></p> <p>Human AAAAAAAAAA<b>A</b>AAAGAAAGA  Chimpanzee AAAAAAAAAAGAAAGAAAGA  Orangutan AAAAAAAAAA-AAAAAAGAA</p>
<p><b>rs188673545*G&gt;C</b></p> <p>Human TAAAAGCTGT<b>C</b>AACCAATTAG  Chimpanzee TAAAAGCTGT<b>C</b>AACCAATTAG  Gorilla TAAAAGCTGT<b>C</b>AACCAATTAG  Orangutan TAAAAGCTGT<b>C</b>AACCAATTAG  Olive baboon GTATATAACA<b>A</b>CTAAGATTAT  Marmoset GCATATAACA<b>A</b>CTAAGATTAT</p>	

**Figura 10. Alineamiento de secuencias ortólogas de FTX sobre los 5 SNP detectados en nuestra población bajo estudio.** Alineamiento obtenido en el servidor de análisis de polimorfismos ofrecido por Ensembl (URL: <http://www.ensembl.org>). En rojo se destaca el nucleótido analizado en cada caso.

Para analizar el potencial impacto de cada uno de los SNPs encontrados sobre el *splicing* del *FTX* se utilizaron los mismos programas y condiciones descritas para *XIST*. De los cinco SNPs observados, sólo el SNP rs188673545 afecta una región exónica, a 402 y 1015 pb de los extremos 5' y 3' del exón 7, respectivamente. Según lo discutido previamente para el *XIST*, es poco probable que los cambios observados, mediante el *ESE FINDER* y *RESCUE ESE* (Tabla 9) tengan algún impacto efectivo modificando la normal estructura de *splicing* del *FTX*.

**Tabla 9. Análisis bioinformático teórico del *splicing* del único SNP encontrado en *FTX*.**

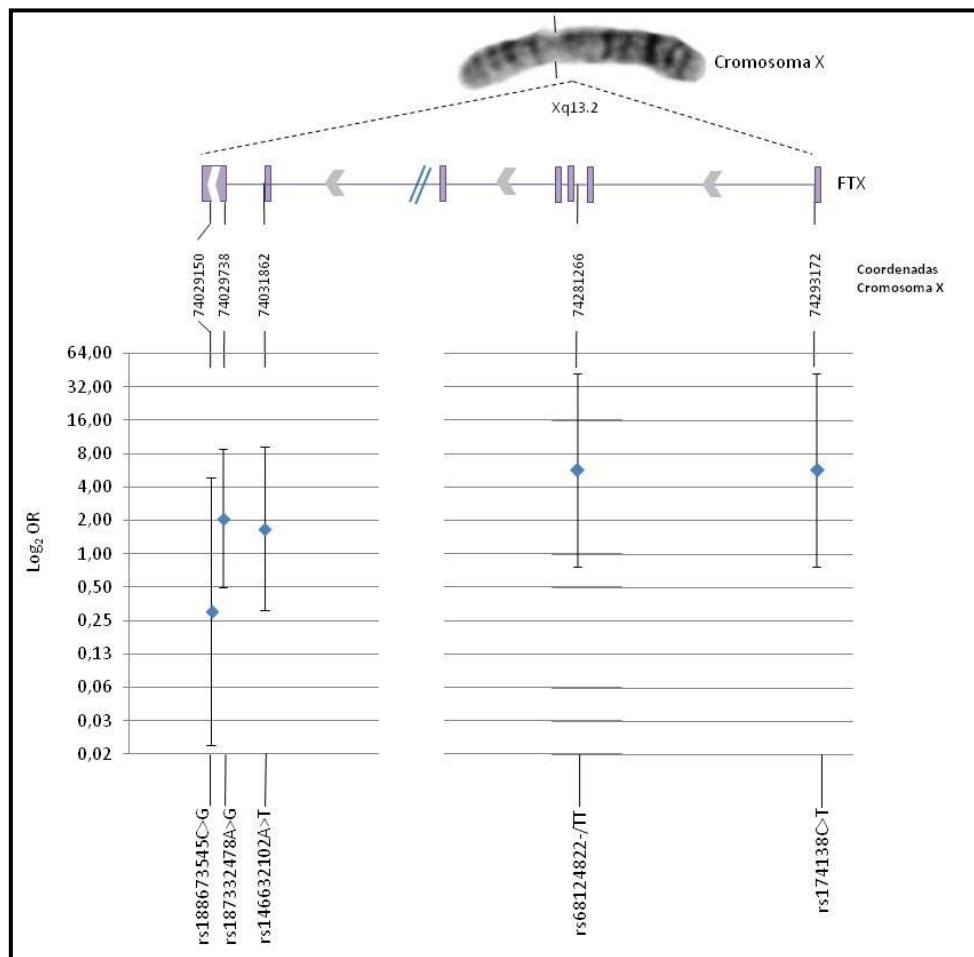
SNP#	ESE FINDER		RESCUE – ESE	
	AGCTGT <b>C</b> AACCAA	AGCTGT <b>G</b> AACCAA	AGCTGT <b>C</b> AACCAA	AGCTGT <b>G</b> AACCAA
rs188673545	SRSF5 Score 3.04	*	*	Un hexámero ESE

\* Sin sitios de unión a las proteínas SRs. ESE: *exonic splicing enhancer*. Cada proteína SR tiene asociada un umbral de unión definido por un score mínimo. SRSF5 umbral: 2.67.

Los cuatro SNPs encontrados en *FTX* ubicados en regiones intrónicas y sobre sitios consenso de *splicing* (rs174138, rs68124822, rs146632102 y rs187332478) fueron

analizados mediante los programas de predicción de sitios de *splicing*, *NNpredict* y *NetGene2*. Los resultados muestran que la presencia de estos polimorfismos no predicen ninguna modificación en el reconocimiento de los sitios normales de *splicing*, ni generan tampoco la activación de sitios crípticos o nuevos sitios de *splicing*.

La distribución de las variantes genéticas del *FTX* encontradas en nuestra población de 'casos' respecto a los 'controles' no mostró diferencias significativas que permitan concluir una asociación entre la inactivación sesgada extrema del cromosoma X y ninguno de los cinco SNPs encontrados en *FTX* (Figura 11).



**Figura 11. Análisis genético del *FTX*, sus variantes y su potencial asociación al sesgo en la inactivación del cromosoma X.** Esquema genómico del *FTX* (en Xq13.2), orientado hacia el centrómero X. Se representan los 7 exones más el exón 1 alternativo (cajas violeta), cada uno de los 5 SNPs identificados (número rs) y sus coordenadas en el cromosoma X (NC\_000023.10). El riesgo relativo se grafica mostrando los OR con su IC<sub>90%</sub> en escala logarítmica base 2.

A pesar de no obtenerse diferencias que significativas, la inspección de las tendencias de cada uno de los ORs nos permitió destacar un SNP potencialmente protector del sesgo en el exón 7 (rs188673545) con un OR de 0.30 (tres veces menos chances de sesgar cuando la variante está presente que cuando no está presente) y cuatro SNPs de riesgo, en el intrón 1 (rs174138), en el intrón 2 (rs68124822), que mostraron ORs de 5.71 (casi seis veces más chances de sesgar) y dos en el intrón 6 (rs146632102 y rs187332478), con ORs de 1.69 y 2.1 que estiman un leve aumento del riesgo a sesgar asociado a la variante (Figura 11).

#### 4.4. Estudio exhaustivo del gen JPX

Se realizó un *screening* genético completo del *JPX* por CSGE y secuenciación de *Sanger*, en siete amplímeros. El promotor junto al primer exón se amplificó en un único amplímero (360 pb), al igual que los exones 2, 3 y 4, (350, 376 y 397 pb, respectivamente) el exón 5 se amplificó en tres segmentos, el primero por PCR-anidada con un producto de 476 pb sobre un producto primario de 2.114 pb, y los otros dos de 430 y 611 pb (Figura 12).

El esquema de análisis del *JPX* incluye 35 SNPs según los anotados en las bases de datos correspondientes: un SNP en el exón 1 y 4, seis en el exón 2, ocho en el exón 3, y 19 en el exón 5. En la Tabla 11 se detallan las características de cada amplímero y los SNPs esperados en *JPX*, así como la posición en el genoma y su secuencia adyacente.

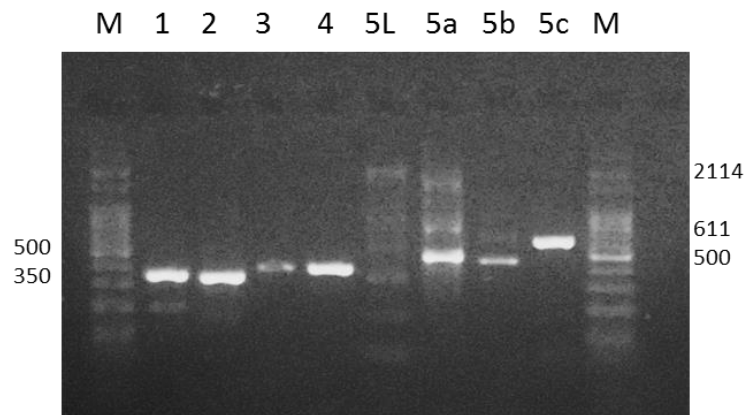
**Tabla 10.** Detalle de los SNPs anotados en base de datos públicas estudiados en *JPX*.

Región	Amplímero	Nombre del SNP	Localización	Secuencia
Exon 1	E1	rs187748140	73944376	GCACGGAGTG C AATCACTGCG
Exon 2	E2	rs113865234	73944515	TAGCGAGGCT A CACACTTTGG
		rs191507498	73944624	TCTCGCTGAT C GCCTAACTTC
		rs149942356	73944641	CTTCTACTGA C GAAGCTTGA
		rs68178357	73944648	TGACGAAGCT T GGAAGTTGCA
		rs148535176	73944653	AAGCTTGAA G TTGCAGAAGG
		rs142856779	73944776	GCTTTCCTAG G AACAGCCTAA
Exon 3	E3	rs55767614	73994778	CACAAGACCT C CCTTCTCCCT
		rs55824328	73994789	CCTTCTCCCT T ATTTGTTTAG
		rs112052329	73994794	TCCCTTATT G TTTAGTACCA
		rs148620997	73994825	ATATTCCTAC C CTTTTTTTTT
		rs191412282	73994857	TCGCTCTTGT T GCCCAGGCTG

		rs111832022	73994863	TTGTTGCCCA G GCTGGAATGC
		rs184109904	73994968	ACAGGTATGC G CTACCACGCC
		rs113529996	73994996	TTTTGTATT T TAGTAGAGAC
Exon 4	E4	rs192213903	74062091	CCTCCTGACT G ACACTGGTGC
Exon 5	E5a	rs189723728	74069173	TTTCAGATTG T TCACTGTTGG
		rs112051669	74069207	TACTGATGCC T CAGAGATGAC
		rs111339997	74069287	TCTCCTGTGA T CCCACGGAGC
		rs73624283	74069335	AATAAGTGCT G AGCAGTTGTC
		rs182767628	74069398	GACAGCCTGA T GACGGTGGTA
		rs187094096	74069409	GACGGTGGTA C AATCAATGAC
		rs192311279	74069411	CGGTGGTACA A TCAATGACAT
	E5b	rs73488787	74069641	GTGTGATGCT A GAAAGCCCCT
		rs183413827	74069680	TAACCGCCAC T GAAAAGAGTG
		rs113449402	74069758	ATCCTGAGGC G GAGCAGGTGA
		rs67649459	74069762	TGAGGCGGAG C AGGTGAGAGG
	E5c	rs188471434	74069899	AGATTAGGCA C CGAGAAGAGG
		rs151057146	74070057	AGCTCCAGAA T GTAGCCAGGC
		rs193106142	74070109	TGGGAGGCCG A GCGGGCAGA
		rs2182287	74070327	GCGACAGAGC G AGACTCTGTC
		rs112560894	74070338	AGACTCTGTC T CTAATAAAT
rs59904201		74070384	AATAAATAAA A TAAAATAAAA	
rs72278733		74070410	ATAAAATAAA TAAA AAGAAAAA	
rs7060923	74070417	TAAAAAA G AAAAAAGTTCC		

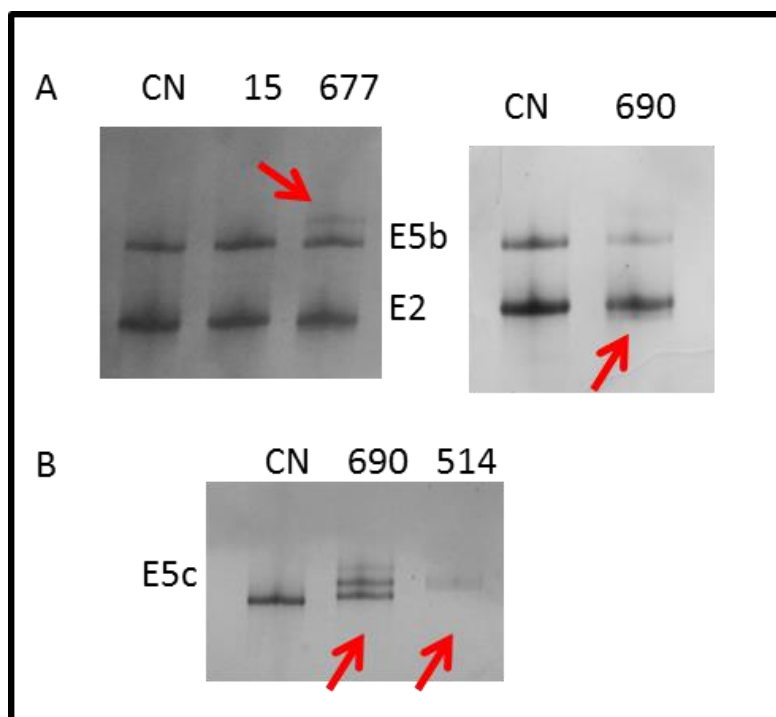
\*Los polimorfismos resaltados en amarillo son los que encontramos en nuestra población de estudio.

Los resultados de la amplificación primaria del *JPX* fue analizada por electroforesis en gel de agarosa (ejemplo mostrado en la Figura 12). En este análisis primario, tanto casos como controles mostraron amplificación positiva sobre los tamaños específicos esperados por lo que se realizó el análisis secundario o monitoreo por CSGE de mutaciones pequeñas (cambios puntuales y pequeñas *indels*).



**Figura 12. Esquema de amplificación de las regiones clave del gen *JPX*.** Electroforesis en gel de agarosa (1,5%). Se muestran los resultados de los amplímeros correspondientes a los exones 1 a 4 y a la amplificación del exón 5. 5L corresponde a un amplímero de 2114pb que fue usado como sustrato para amplificar el amplímero anidado 5a. También se muestran los amplímeros 5b, 5c del exón 5. Se detallan los tamaños de los amplímeros de mayor y menor peso molecular y la señal de mayor intensidad del marcador de peso molecular (500 pb). M indica marcador de peso molecular, escalera de 100 pb.

El análisis por CSGE mostró patrones con movilidad diferencial respecto al heterodúplex control de corridas en paralelo en los amplímeros E2, E3, E5b y E5c (algunos ejemplos son mostrados en la Figura 13).



**Figura 13. Análisis secundario de mutaciones pequeñas y SNPs del *JPX* por CSGE.** Se muestran las señales específicas observadas en algunos de los casos presentados. Se indica con flecha roja, las señales con movilidad anómala en el heterodúplex. También se muestra en cada gel el patrón del homodúplex control (CN). (A) Multiplex 2 (Tabla 2) (E2 y E5b) patrones de señal correspondientes a los heterodúplex de tres casos (15, 677, 690). (B) Patrones de señal correspondientes al amplímero E5c en dos muestras (690, 514) con patrones de señal distintos.



El análisis de los 176 amplímeros no mostró mutaciones nuevas, sino las variantes alélicas reportadas de 6 de los 35 SNPs anotados en las regiones estudiadas del *JPX* (indicados en la Tabla 10). El análisis de las variantes encontradas en *JPX* se detalla en la Tabla 11. Así como en los genes anteriores se estudiaron las variantes del *JPX* en el contexto filogenético para investigar su posible integración en regiones conservadas en distintas especies de homínidos (Figura 14). Según este análisis el alelo ancestral de los SNPs rs68178357, rs2182287 y rs72278733 está en una región conservada en las distintas especies estudiadas, sólo en humanos esta posición se presenta como un polimorfismo. El polimorfismo rs55824328 se localiza en una región ampliamente conservada y el alelo ancestral está presente en todas las especies relacionadas. El SNP rs67649459 no ofrece suficiente información para aplicar este análisis.

**Tabla 11.** Características de los SNPs encontrados en el gen *JPX*.

SNP	Alelo Ancestral	Frecuencias alélicas	MAF	CITA
rs68178357*A>T	A	T=0.9078 A=0.0922	A=0.0922/348	-
rs55824328*C>T	T	T=0.9078 C=0.0922	C=0.0922/348	-
rs67649459*C>G	C	C=0.9091 G=0.0909	G=0.0909/343	-
rs2182287*A>G	G	G=0.6387 A=0.3613	A=0.3613/1364	-
rs72278733*/-TAAAA	ND	ND	ND	-
rs59904201*/-/A	ND	ND	ND	-

\*MAF indica la frecuencia del alelo minoritario (*Minor Allele Frequency*) y el número total de alelos estudiados. Las frecuencias alélicas se refieren a las frecuencias observadas en el Proyecto 1000 Genomas (fase 3, CITA y aquí URL).

<p><b>rs68178357*A&gt;T</b></p> <p>Human TGACGAAGCT<u>T</u>GGAAGTTGCA            Chimpanzee TGACGAAGCT<u>A</u>GGAAGTTGCA            Gorilla TGACGAGGCT<u>A</u>GGAAGTTGCA            Orangutan TGACGAGGCT<u>A</u>GGAAGTTGCA            Vervet-AGM TGACGAGGCT<u>A</u>GGAAGTTGCA            Macaque TGACGAGGCT<u>A</u>GGAAGTTGCA            Olive baboon TGACGAGGCT<u>A</u>GGAAGTTGCA</p>	<p><b>rs55824328*C&gt;T</b></p> <p>Human CCTTCTCCCT<u>T</u>ATTTGTTTATAG            Chimpanzee CCTTCTCCCT<u>T</u>ATTTGTTTATAG            Gorilla CCTTCTCCCT<u>T</u>ATTTGTTTATAG            Orangutan CCTTCTCCCT<u>T</u>ATTTGTTTATAG            Vervet-AGM CCTTCTCCCT<u>T</u>ATTTGTTTATAG            Macaque CCTTCTCCCT<u>T</u>ATTTGTTTATAG            Olive baboon CCTTCTCCCT<u>T</u>ATTTGTTTATAG</p>
<p><b>rs67649459*C&gt;G</b></p> <p>No hay alineamiento disponible</p>	<p><b>rs2182287*A&gt;G</b></p> <p>Human GCGACAGAGC<u>G</u>GAGACTCTGTGTC            Chimpanzee GCGACAGAGC<u>G</u>GAGACTCTGTGTC            Orangutan GCGACAGAGC<u>A</u>GAGACTCTGTGTC            Olive baboon -----ACTCTGTGTC</p>
<p><b>rs72278733*-/TAAAA</b></p> <p>Human TAAAA<u>TAAAA</u>TAAAAAAGA            Chimpanzee TAAAA<u>TAAAA</u>TAAAAAAGA            Orangutan TAAAA<u>TAAAA</u>TAAAAAAGA            Olive baboon TAAAT----TTAAAAAAGA</p>	<p><b>rs59904201*-/A</b></p> <p>Human AATAAATAAAA<u>A</u>TAAAAATAAAA            Chimpanzee AATAAATA-----AATAAAA            Orangutan -----AA            Olive baboon AGTTAGTT-----AATTAAT</p>

**Figura 14. Alineamiento de secuencias ortólogas de JPX sobre los 6 SNP detectados en nuestra serie.** Alineamientos obtenidos en el servidor Ensembl (URL [www.ensembl.org](http://www.ensembl.org)).

Al igual que en los genes *XIST* y *FTX*, para analizar el posible impacto asociado a las variantes de los SNPs sobre el *splicing* del *JPX* se utilizaron los mismos programas y condiciones descriptos previamente. Cuatro de los seis SNPs observados se encontraron en regiones exónicas, y sólo dos de ellos se localizan dentro de región de 100 pb del límite exón-intrón requeridas para que un ESE pueda ejercer su función estimuladora del *splicing*. A pesar de su potencialidad, el análisis de estos SNPs, rs68178354 (ubicado a 15 pb del extremo 3' y a 65 pb del extremo 5' del exón 2 del *FPX*) y rs2182287 (a 58 pb del 3' y 159 pb del 5' del exón 5), no mostró cambios en los mapas de ESEs (Tabla 12). En los SNPs ubicados fuera de los límites de riesgo para impactar el *splicing*, uno no mostró cambios (rs59904201) y el otro, rs67649459 (a 626 pb del extremo 3' y 591 pb del 5' del exón 5 del *JPX*), aunque muestra cambios en la presencia de la variante no se espera impacto en el *splicing* a causa de su posición (Tabla 12).

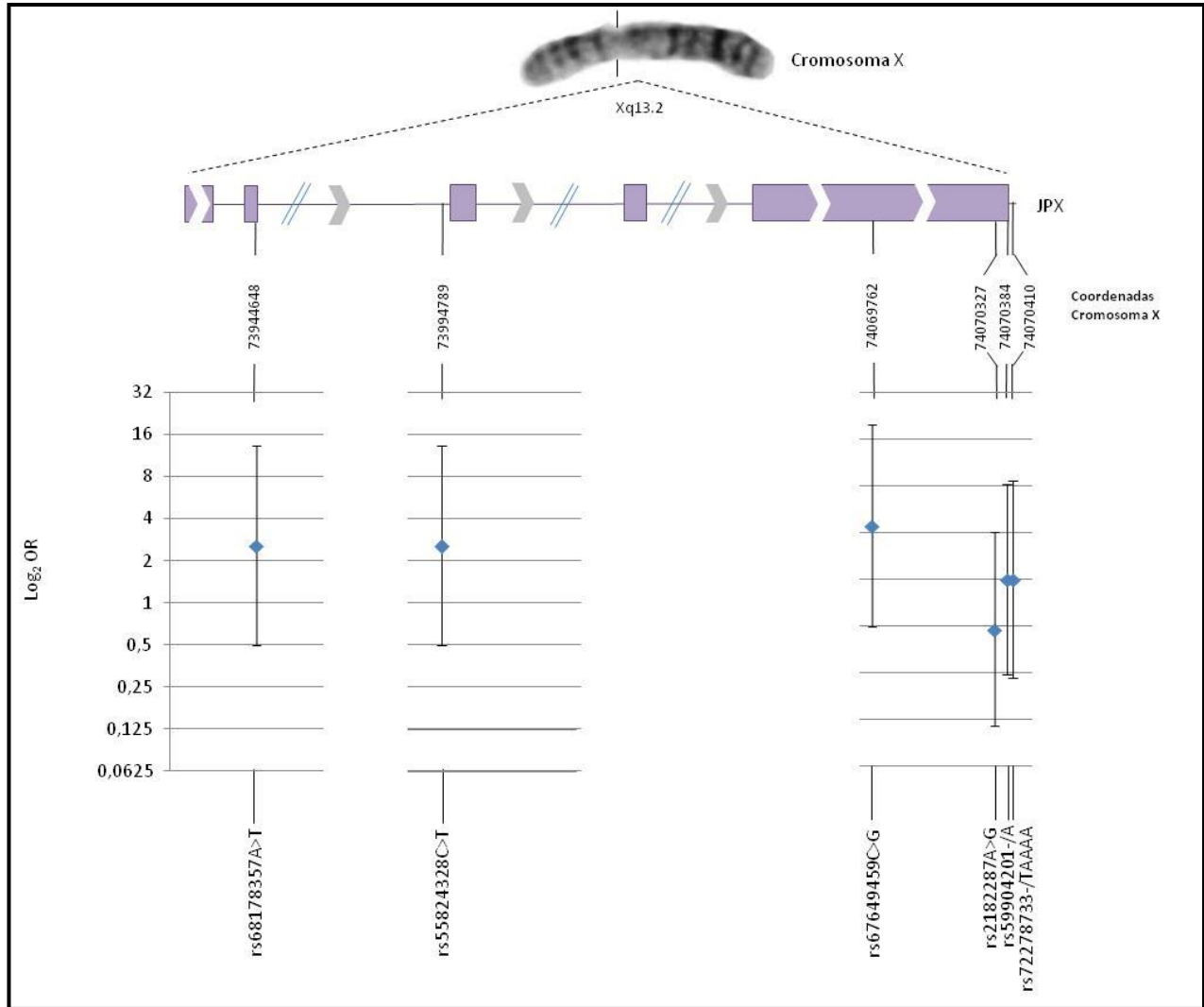
**Tabla 12.** Análisis bioinformático de los SNPs respecto al *splicing* del *JPX*.

SNP#	ESE FINDER		RESCUE-ESE	
rs68178354	GAAGCT <b>T</b> GGAAGT	GAAGCT <b>A</b> GGAAGT	GAAGCT <b>T</b> GGAAGT	GAAGCT <b>A</b> GGAAGT
	*	*	Un hexámero ESE	Un hexámero ESE
rs67649459	GCGGAG <b>C</b> AGGTGA	GCGGAG <b>C</b> AGGTGA	GCGGAG <b>C</b> AGGTGA	GCGGAG <b>C</b> AGGTGA
	SRSF1 2x <i>Score</i> 2.71 y 2.67 SRSF1A 2x <i>Score</i> 3.46 y 2.12	SRSF1 2x <i>Score</i> 5.29 y 2.43 SRSF1A <i>Score</i> 5.11	*	Dos hexámeros ESE.
rs2182287	CAGAGC <b>G</b> AGACTC	CAGAGC <b>G</b> AGACTC	CAGAGC <b>G</b> AGACTC	CAGAGC <b>G</b> AGACTC
	*	*	*	Dos hexámeros ESE
rs59904201	AATAAA <b>A</b> TAAAAT	AATAAA <b>A</b> TAAAAT	AATAAA <b>A</b> TAAAAT	AATAAA <b>A</b> TAAAAT
	*	*	*	*

(\*) Sin sitios de unión a las proteínas SRs. ESE: *exonic splicing enhancer*. Cada proteína SR tiene asociado un *score* mínimo. SFRSF1A (SRSF1 IgM-BRCA1) umbral:1.867; SRSF1, 1.956.

Los dos polimorfismos que se encuentran en regiones intrónicas (SNPs rs55824328 y rs72278733), también analizados respecto a su potencial impacto sobre el *splicing*, tampoco indicaron cambios potencialmente funcionales asociados a sus variantes.

El análisis de asociación entre cada uno de los seis SNPs y sus variantes genéticas del *JPX* encontradas en nuestra población en *casos* y *controles* no mostró diferencias significativas que permitan concluir una asociación (causal o por ligamiento físico/genético) con el sesgo en la inactivación del X. Sin embargo, la inspección de las tendencias de cada uno de los ORs permite destacar, un SNPs potencialmente protector en el exón 5, rs2182287 con OR(IC<sub>90%</sub>) de 0.4762 (0.11 a 1.99), dos SNPs netamente neutros en el exón 5 rs72278733 y rs59904201 con OR(IC<sub>90%</sub>) de 1 (0,23 a 4,29 y de 0,24 a 4,09, respectivamente); y tres SNPs potencialmente asociados a sesgo, uno en el exón 2, (rs68178357), uno en el intrón 2 (rs55824328) mostrando ORs (IC<sub>90%</sub>) iguales: 2,57 (0,49 a 13,37) y otro SNP en el exón 5 (rs67649459) con un OR(IC<sub>90%</sub>) de 2,22 (0,49 a 9,9) (Figura 15).



**Figura 15. Análisis estadístico de SNPs del gen *JPX* y su potencial asociación al sesgo en la inactivación del X.** Bajo el esquema genómico del *JPX* localizado en la banda cromosómica Xq13.2 y orientación hacia el telómero Xq, se representan los 5 exones (cajas violetas), y cada uno de los 6 SNPs mostrando su identificación (# rs) y coordenadas en el cromosoma X del genoma humano archivo NC\_000023.10. Se indican cada uno de los OR con su IC<sub>90%</sub> asociados a cada uno de los seis SNPs en escala logarítmica.

## 5. DISCUSIÓN

En las enfermedades recesivas ligadas al cromosoma X, los varones resultan afectados mientras que las mujeres, debido a su condición de portadoras, resultan generalmente asintomáticas. La presencia de mujeres portadoras sintomáticas de enfermedades ligadas al X como HA y DMD con determinación molecular de su estatus de portadora heterocigota promovió el desarrollo de distintos estudios anteriores al inicio de este trabajo de Tesis en el Laboratorio de Genética Molecular de la Hemofilia (GMH) incluyendo el establecimiento de la estrategia de análisis del patrón de inactivación del cromosoma X (XIP) mediante el sistema HUMARA. El patrón de inactivación sesgado observado junto al diagnóstico molecular de portadora heterocigota, permitió identificar la causa de la expresión de la enfermedad en las mujeres sintomáticas de estas patologías (HA, Radic *et al*, 2015; DMD, Giliberto *et al*, 2014). Sin embargo, si hacemos foco en las causas de la inactivación sesgada del cromosoma X, no sólo se sabe poco, sino el poco conocimiento que se tiene proviene del terreno teórico con escasa o nula evidencia experimental.

La inactivación incluye un proceso inicial de reconocimiento ligado a la selección del cromosoma X a ser inactivado y otros eventos tempranos que están regulados por una región de control, región única del cromosoma X localizada en Xq13, el centro de inactivación del cromosoma X (XIC, *X-inactivation centre*) (Lee y Jaenisch, 1997).

En los últimos años, se han descrito genes clave la regulación de XCI, e interesantemente, muchos de estos producen ARN no codificantes (RNAnC), una clase de genes no traducidos a proteína con funciones no específicas aunque en general regulatorias de la expresión.

El candidato principal necesario, y suficiente, para iniciar y mantener la inactivación del cromosoma X, ubicado dentro del XIC, fue descrito en 1992 (Brown *et al*, 1992), el gen *XIST* que abarca 17 kb de ADN genómico y genera un RNAnC de gran tamaño. La regulación positiva del RNAnC *XIST* desencadena el revestimiento del cromosoma X en *cis*, silenciando todo el cromosoma sobre el cual se expresa (Penny *et al*, 1996; Wutz *et al*, 2002). Estudios posteriores (Kolesnikov y Elisaphenkoc, 2010), indicaron al gen *XIST/Xist* con un rol protagónico clave en el proceso de inactivación del cromosoma X en

humanos/ratones, respectivamente, pero además agregaron que en el proceso de XCI son importantes otros RNAnc como *JPX/jpx* y *FTX/Ftx*. A partir de este trabajo distintos grupos estudiaron éstos últimos dos RNAnc sobre modelos de ratón, el gen *Ftx* (RNAnc que escapa parcialmente a la inactivación del cromosoma X) descrito como regulador positivo del *Xist* (Chureau et al, 2011) y el gen *Jpx* que actuaría sobre el promotor del *Xist* como un activador transcripcional desalojando a las proteínas CTCF (inhibidores transcripcionales) de su secuencias de unión específica (Sun et al, 2013).

Kung y colaboradores (2013), en su revisión “*Long Noncoding RNAs: Past, Present, and Future*” refiere algunas controversias respecto a que casi la totalidad del genoma de mamíferos se transcribe en transcritos no codificantes funcionales, documenta el gran interés que pone en foco a los RNAnc porque los involucra tanto en la evolución biológica como en la función del genoma, con roles identificados en el desarrollo embrionario y en la expresión de ciertas enfermedades. Sin embargo, también se indica que el conocimiento de los mecanismos por los que actúan estos ARNnc sigue siendo incompleto o limitado.

Teniendo en cuenta la bibliografía donde se han indicado los genes involucrados en el proceso de inactivación del X, este trabajo de Tesis presenta un estudio exploratorio de las variantes genéticas en los tres genes no codificantes del XIC y su eventual asociación estadística al proceso de inactivación mediante la búsqueda de diferencias entre un grupo de pacientes con sesgo (extremo) en la inactivación del cromosoma X, “casos”, y un grupo de pacientes sin sesgo referidos como “controles”.

La variantes genéticas constitucionales estudiadas fueron identificadas desde ADN genómico extraído de leucocitos de sangre periférica por amplificación PCR de las regiones exónicas y secuencias de *splicing* adyacentes a los exones de los genes *XIST*, *JPX* y *FTX*.

La metodología aplicada no permitió el monitoreo y la detección de grandes deleciones (aquellas que remuevan la presencia de un amplicón entero y no aquellas que sólo modifican el tamaño del amplicón como las pequeñas *indel* (deleciones y/o inserciones de 1-10 pb)) en heterocigosis porque en estudios de genes ligados al X la presencia del alelo normal enmascara el eventual alelo con la deleción no permitiendo su identificación directa por este abordaje. La secuenciación de los amplicómeros con movilidad diferencial en CSGE

no reveló deleciones pequeñas, mutaciones puntuales nuevas, ni cambios en las secuencias consenso de *splicing* que permitan explicar, sobre base racional, un defecto funcional en la expresión del gen correspondiente y, finalmente, el sesgo en la XCI observado en alguna o varias de las pacientes bajo estudio. En cambio, se determinó la presencia de 18 SNPs entre los tres genes estudiados (siete SNPs en *XIST*, cinco en *FTX* y seis en *JPX*) todos ellos reportados con anterioridad en distintas poblaciones humanas. Sin embargo, a pesar de la pequeña población explorada en este trabajo hasta el presente (n=22), la magnitud de los OR observados entre casos y controles nos permite estimar que 11 de las 18 variantes observadas podrían estar asociadas al proceso de XCI en humanos (cuatro variantes en *XIST*, cuatro en *FTX* y tres en *JPX*) aunque esta asociación debiera ser confirmada en series más grandes en el futuro.

Como era de esperar, el mayor número de variantes genéticas fueron encontradas en los exones de mayor tamaño que a su vez listaban un mayor número de variantes anotadas en las bases de datos de SNPs correspondientes. Por ejemplo, de las 80 variantes anotadas en el *XIST*, todos los siete SNPs encontrados estaban localizados en los dos mayores exones, dos en el exón 1 y cinco en el exón 6. Entre éstos siete, sólo cuatro SNPs estarían potencialmente asociados a sesgo de la XCI (exón 1: rs6527C/A, OR: 2.57; exón 6: rs16992443G/T, OR: 2.57, rs16992442T/C, OR: 3.75 y rs16992436T/C, OR: 2.57).

En *FTX* se encontraron cinco SNPs informativos de las 46 variantes anotadas cuatro de los cuales muestran valores OR que sugieren una potencial predisposición a sesgar la XCI, todos ellos ubicados en regiones intrónicas del *FTX*: rs74138C/T: OR: 5.71 ; rs68124822-/TT, OR: 5.71; rs146632102A/T OR: 1.69 y rs187332478A/G OR: 2.1.

En *JPX* se detectaron seis SNPs informativos de las 35 variantes anotadas, y sólo tres de ellas muestran OR sugerentes de una potencial asociación a sesgo de XCI: rs68178357A/T OR: 2.57; rs55824328C/T OR: 2.57 y rs67649459C/G OR: 2.22.

La falta de un conocimiento bien establecido sobre los mecanismos y las estructuras funcionales de los ARNnc no nos permite afinar la búsqueda de variantes sobre aquellas regiones con roles bien asignados e importantes para el funcionamiento de los transcritos de *XIST*, *FTX* y *JPX*. Sin embargo, aún con este panorama poco definido, es razonable suponer que tanto variantes que afecten a las regiones regulatorias de estos genes con

potencial para alterar los niveles de expresión como las mutaciones que modifiquen las estructuras exónicas de estos genes (e.g., defectos de *splicing* o grandes deleciones de uno o más exones), todas ellas tienen suficiente potencial deletéreo para abolir o reducir el funcionamiento del alelo afectado en su capacidad para iniciar o desarrollar la XCI. Aún en este escenario complejo, las variantes encontradas como potencialmente involucradas podrían ser: o bien causales *per se* del estado de inactivación sesgada del X o, alternativamente, estar físicamente cerca de la variante causal desconocida y heredarse en conjunto como un haplotipo ligado genéticamente. Bajo ambas situaciones se verificará una asociación estadística entre la variante en estudio y el sesgo extremo de la XCI observada en "casos" vs "controles". Aunque los resultados obtenidos mostraron una mera tendencia a la asociación variante alélica-sesgo XCI, con valores OR de riesgo mayores a la hipótesis nula ( $H_0: 1,00$ ), en ningún caso esta tendencia se tradujo en diferencias estadísticamente significativas ( $p < 0,05$ ). Esta evidencia podría estar asociada al reducido tamaño de las series de casos y controles estudiadas aunque en cualquier caso se necesitará de la profundización de este trabajo para confirmar o descartar nuestra hipótesis que aún queda abierta sin una respuesta definitiva.

Asimismo, los estudios bio-informáticos sobre los SNPs encontrados enfocados a elucidar una posible asociación causal de la variante *per se* con el sesgo en la XCI (i.e., investigación *in silico* de los sitios exónicos de *enhancer* de *splicing* (ESE) y de sitios consenso de *splicing*) tampoco mostraron resultados netos o significativos, evidencia que nos permitió descartarlos como la causa directa de un eventual cambio de estructura exónica de estos ARNnc por *splicing* defectuoso.

En la literatura médica, el único SNP reportado hasta el momento como asociado a sesgo en la inactivación del cromosoma X fue descrito en dos trabajos y corresponde a la variante rs187705242A/G (Plenge *et al*, 1997; Tomkins *et al*, 2002), este polimorfismo no fue encontrado como informativo en las mujeres de nuestra población, en coincidencia con otro estudio que incluye más de 50 mujeres con sesgo (Pereira y Zatz, 1999).

Cabe destacar que en nuestras series de casos y controles hay individuos que mostraron ausencia de variantes heterocigotas en todos los amplímeros estudiados: dos casos y tres



---

controles en *XIST* (D514, 444, 275, 15 y D60), un caso y un control en *FTX* (482 y 373) y un caso y tres controles en *JPX* (444, 15, D60 y 277). En este grupo de pacientes esta ausencia de variantes en heterocigosis deja abierta la posibilidad de presentar grandes deleciones, que deberían investigarse por ejemplo por algún sistema de medición de dosis génica, y de encontrarse estas deleciones heterocigotas permitirían explicar la causa directa del sesgo de la XCI observado.

En el año 2004, el *International Human Genome Sequencing Consortium*, en la publicación "*Finishing the euchromatic sequence of the human genome*", reveló la existencia de aproximadamente 10 millones de SNPs que pueden estar localizados en regiones intrónicas, exónicas, en regiones reguladoras y en regiones intergénicas. Los SNPs, a diferencia de otro tipo de marcadores como los microsatélites, caracterizados por la repetición de un motivo o secuencia pequeña (típicamente de 2-6 pb), presentan una tasa menor de mutación por lo que son de gran ayuda en estudios de genética de poblaciones para investigar fenómenos biológicos como por ejemplo la evolución filogenética de la raza humana. En nuestro estudio se realizaron los alineamientos de secuencias ortólogas en las regiones que contenían las variantes informativas en casos y controles de los tres genes del XIC, para ser evaluadas en su contexto filogenético a fin de determinar si estas variantes estaban o no conservadas en las distintas especies lo que podría indicar un rol definido e importante en la función de estos ARNnc. La mayoría de los SNPs informativos identificados en nuestras series humanas resultaron en regiones altamente conservadas en las distintas especies de homínidos incluidas en el estudio, encontrándose tanto el alelo ancestral como el correspondiente a la variante polimórfica. Nueve de los 18 SNPs informativos encontrados en nuestras series mapeaban sobre regiones conservadas en todas las especies, mientras que seis 6 SNPs mostraron una variante distinta en humanos respecto a los otros grupos incluidos, y tres SNPs no presentan datos. Otra vez, estos datos no permiten aportar evidencia concluyente acerca de las regiones funcionalmente importantes de los ARNnc de *XIST*, *FTX* y *JPX* en el proceso de XCI, pero si nos estimulan a refocalizar los futuros estudios sobre aquellas regiones altamente conservadas por sobre las más variables en los ortólogos.

---

Si bien nuestro estudio no permitió elucidar la/s causa/s del sesgo en la XCI, este análisis exploratorio de SNPs permitió recopilar los primeros datos sobre variantes genéticas en los genes *XIST*, *FTX* y *JPX* en nuestra población de mujeres con XCI sesgada extrema y XCI al azar (no sesgada) en muestras de leucocitos de sangre periférica.

Como primera perspectiva, nos proponemos aumentar en lo posible el número de "casos" con sesgo extremo de la XCI y sobretodo de los "controles" a, por lo menos, más de 150 mujeres de la población general sin problemas de salud explícitos, a fin de evaluar la variabilidad genética natural de cada una de las regiones de estos tres genes asociados a ARNnc y, eventualmente, encontrar diferencias significativas en el análisis estadístico. Asimismo, la ausencia de SNPs en heterocigosis en algunos casos y controles de nuestra serie nos pone como objetivo inmediato la búsqueda de grandes deleciones en la región.

La comprensión del complejo proceso de inactivación del cromosoma X y de las causas asociadas al sesgo en la XCI en el interjuego entre los dos alelos XIC nos permitirá, por ejemplo, predecir cuál de los dos cromosomas X se inactivará preferencialmente y cuál no (i.e., el ligado a la mutación o el normal) en una mujer portadora heterocigota determinando el nivel de expresión del fenotipo específico en enfermedades hereditarias ligadas al cromosoma X recesivas tales como la HA, la HB y la DMD entre muchas otras. Finalmente, la futura elucidación de los mecanismos involucrados y los defectos genéticos asociados al proceso de XCI permitirán diseñar algún abordaje terapéutico específico para intervenir morigerando o evadiendo, si fuera posible, el sesgo extremo en la XCI en la portadora sintomática, sólo sobre el *locus* causal específico del fenotipo de la enfermedad ligada al X correspondiente.

## 6. CONCLUSIÓN.

Se determinó el patrón de inactivación del cromosoma X mediante el sistema HUMARA en muestras de ADN genómico de leucocitos de sangre periférica en mujeres portadoras y no portadoras de Hemofilia A y Distrofia Muscular de Duchenne y controles de la población Argentina general. Se clasificaron las muestras en "Casos" (n=11), con XCI sesgada extrema ( $90\% < XIP \leq 100\%$ ) y "Controles" (n=11), con XCI al azar (XCI no-sesgada) ( $50\% \leq XIP \leq 55\%$ ) para abordar un estudio tipo Caso/Control exploratorio de variantes genéticas en tres genes del XIC.

El estudio de las mutaciones y polimorfismos de los genes *XIST*, *JPX* y *FTX* se abordó mediante la estrategia de monitoreo de variantes pequeñas por análisis de heterodúplex por CSGE incluyendo el análisis de 990 amplímeros donde se identificó la presencia de variantes en 18 SNPs, sobre los 161 SNPs anotados en las bases de datos internacionales sobre las regiones del XIC correspondientes.

El estudio Caso/Control (n=22) exploratorio de variantes genéticas asociadas al establecimiento de sesgo en la XCI indicó ciertas tendencias aunque nunca asociadas a diferencias estadísticamente significativas quizás debidas a los reducidos tamaños muestrales estudiados hasta el presente. En *XIST*, las variantes rs1794213, rs1620574 mostraron ORs sin tendencia de riesgo ( $OR < 1$ ), la variante rs41305409 mostró un OR neutro ( $OR = 1$ ), y las variantes rs6527, rs16992443, rs16992436 y rs16992442 mostraron un OR potencialmente asociado a sesgo ( $OR > 1$ ) (ORs 3-4). En *FTX*, la variante rs188673545 mostró un OR sin tendencia de riesgo y las variantes rs174138, rs68124822, rs146632102 y rs187332478 mostraron OR con tendencia a sesgar (ORs 2-6). En *JPX*, la variante rs2182287 no mostró tendencia a sesgar y dos variantes, rs72278733 y rs59904201, mostraron ORs neutros; mientras que tres variantes, rs68178357, rs55824328 y rs67649459, mostraron cierta tendencia a sesgar (ORs 2-3).

Los resultados de análisis bioinformático para establecer eventual afectación de la estructura exónica por *splicing* espúreo asociado a las variantes y los análisis de regiones

conservadas filogenéticamente de estos tres genes asociados a los ARNnc del XIC tampoco reveló evidencia concluyente que nos permita inferir una relación causal con el sesgo extremo de la XCI en las series de mujeres de nuestra población sobre muestras de ADN genómico de leucocitos de sangre periférica.

---

## 7. BIBLIOGRAFÍA

Allen R, Zoghbi H, Moseley A, Rosenblatt H, Belmont J. 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.* 1992; 51: 1229-39.

Ananiev G, Williams EC, Li H, Chang Q. Isogenic pairs of wild type and mutant induced pluripotent stem cell (iPSC) lines from Rett syndrome patients as in vitro disease model. *PLoS One.* 2011; 6(9):e25255.

Babbs C, Stewart HS, Williams LJ, Connell L, Goriely A, Twigg SR, Smith K, Lester T, Wilkie AO. Duplication of the EFNB1 gene in familial hypertelorism: imbalance in ephrin-B1 expression and abnormal phenotypes in humans and mice. *Hum Mutat.* 2011 Aug; 32(8):930-8.

Biocchi MP, Migeon BR, Pasino M, Lanza T, Bottini F, Boeri E, Molinari AC, Corsolini F, Moreiro C, Acquila M. Familial nonrandom inactivation linked to the inactivation centre in heterozygotes manifesting haemophilia A. *European Journal of Human Genetic* 2005; 13: 635-40.

Bolduc V, Chagnon P, Provost S, Dubé MP, Belisle C, Gingras M, Mollica L, Busque L. No evidence that skewing of X chromosome inactivation patterns is transmitted to offspring in humans. *J Clin Invest.* 2008 Jan; 118(1):333-41.

Brown CJ, Hendrich BD, Rupert JL, Lafrenière RG, Xing Y, Lawrence J, Willard HF. The human XIST gene: analysis of a 17 kb inactive X-specific RNA that contains conserved repeats and is highly localized within the nucleus. *Cell.* 1992 Oct 30; 71(3):527-42.

Brown CJ, Robinson WP. The causes and consequences of random and non-random X chromosome inactivation in humans. *Clin Genet* 2000; 58: 353–63.

Brown CJ, Willard HF. Localization of a gene that escapes inactivation to the X chromosome proximal short arm: implications for X inactivation. *Am J Hum Genet* 1990; 46(2): 273-9.

Cáceres EF, Hurst LD. The evolution, impact and properties of exonic splice enhancers. *Genome Biol.* 2013 Dec 20; 14(12): R143.

- Chagnon P, Provost S, Belisle C, Bolduc V, Gingras M, Busque L. Age-associated skewing of X-inactivation ratios of blood cells in normal females: a candidate- analysis approach. *Exp Hematol*. 2005 Oct; 33(10):1209-14.
- Chow JC, Hall LL, Clemson CM, Lawrence JB, Brown CJ. Characterization of expression at the human XIST locus in somatic, embryonal carcinoma, and transgenic cell lines. *Genomics*. 2003 Sep; 82(3): 309-22.
- Chureau C, Chantalat S, Romito A, Galvani A, Duret L, Avner P, Rougeulle C. Ftx is a non-coding RNA which affects Xist expression and chromatin structure within the X-inactivation center region. *Hum Mol Genet*. 2011 Feb 15; 20(4): 705-18.
- Chureau C, Prissette M, Bourdet A, Barbe V, Cattolico L, Jones L, Eggen A, Avner P, Duret L. Comparative Sequence Analysis of the X-Inactivation Center Region in Mouse, Human, and Bovine. *Genome Research*. 2002 Jun; 12(6) 894.
- Esteller M. Non-coding RNAs in human disease. *Nat Rev Genet*. 2011 Nov 18, 12(12):861-74. doi: 10.1038/nrg3074.
- Giliberto F, Radic CP, Luce L, Ferreiro V, De Brasi C, Szijan I. Symptomatic female carriers of Duchenne muscular dystrophy (DMD): genetic and clinical characterization. *J Neurol Sci*. 2014 Jan 15; 336(1-2):36-4.
- Hinks JL, Winship PR, Makris M, Preston FE, Peake IR, Goodeve AC. A rapid method for haemophilia B mutation detection using conformation sensitive gel electrophoresis. *British Journal of Haematology* 1999, 104: 915-918.
- Hoffman EP, Arahata K, Minetti C, Bonilla E, Rowland LP. Dystrophinopathy in isolated cases of myopathy in females. *Neurology* 1992;42:967-75.
- Hong YK, Ontiveros SD, Strauss WM. A revision of the human XIST gene organization and structural comparison with mouse Xist. *Mamm Genome*. 2000 Mar; 11(3): 220-4.
- International Human Genome Sequencing Consortium. Finishing the euchromatic sequence of the human genome. *Nature*. 2004 Oct 21; 431(7011): 931-45.
- Kling S, Coffey AJ, Ljung R, Sjörin E, Nilsson IM, Holmberg L, Giannelli F. Moderate haemophilia B in a female carrier caused by preferential inactivation of the paternal X chromosome. *Eur J Haematol*. 1991; 47 (4): 257-61.

Kolesnikov NN, Elisafenko EA. Comparative organization and the origin of noncoding regulatory RNA genes from X-chromosome inactivation center of human and mouse. *Genetika*. 2010 Oct; 46(10): 1386-91.

Kucera KS, Reddy TE, Pauli F, Gertz J, Logan JE, Myers RM, Willard HF. Allele specific distribution of RNA polymerase II on female X chromosomes. *Hum Mol Genet*. 2011 Oct 15; 20(20): 3964-73.

Kung JT, Colognori D, Lee JT. Long noncoding RNAs: past, present, and future. *Genetics* 2013 Mar; 193(3): 651-69.

Lafreniere RG, Brown CJ, Rider S, Chelly J, Taillon-Miller P, Chinault AC, Monaco AP, Willard HF. 2.6 Mb YAC contig of the human X inactivation center region in Xq13: physical linkage of the RPS4X, PHKA1, XIST and DXS128E genes. *Hum. Molec. Genet*. 2: 1105-1115, 1993.

Lahiri DK, Nuremberg JI. A rapid non-enzymatic method for the preparation of HMW DNA from blood for RFLP studies. *Nucleic Acids Res*. 1991; 19: 5444.

Lee JT, Jaenisch R. The (epi)-genetic control of mammalian X-chromosome inactivation. *Curr Opin Genet Dev*. 1997 Apr; 7(2): 274-80.

Lee, JT. Disruption of imprinted X inactivation by parent-of-origin effects at Tsix. *Cell*. 2000 Sep 29; 103(1):17-27.

Lyon MF. Gene action in the X-chromosome of the mouse (*Mus Musculus*). *Nature*. 1961; 190: 372-3.

Migeon BR, Chowdhury AK, Dunston JA, McIntosh I. Identification of TSIX, encoding an RNA antisense to human XIST, reveals differences from its murine counterpart: implications for X inactivation. *Am J Hum Genet*. 2001 Nov; 69(5): 951-60.

Mitchell M., Keeney S., Goodeve A. The molecular analysis of haemophilia B: a guideline from the UK haemophilia centre doctors' organization haemophilia genetics laboratory network. *Haemophilia* 2005, 11: 398-404.

Ng K, Pullirsch D, Leeb M, Wutz A. Xist and the order of silencing. *EMBO reports*. 2007; 8(1): 34-9.

Nisen PD, Waber PG. Nonrandom X chromosome DNA methylation patterns in hemophilic females. *Journal Clinical Investigation*. 1989; 83: 1400-03.

Orstavik KH, Orstavick RE. Skewed X chromosome inactivation in a female with haemophilia B and in her non carrier daughter: a genetic influence on X chromosome inactivation? 1999; 36: 865-66.

Penny GD, Kay GF, Sheardown SA, Rastan S, Brockdorff N. Requirement for Xist in X chromosome inactivation. *Nature*. 1996 Jan 11; 379(6561): 131-7.

Pereira LV, Zatz M. Screening of the C43G mutation in the promoter region of the XIST gene in females with highly skewed X-chromosome inactivation. *Am J Med Genet*. 1999 Nov 5; 87(1): 86-7.

Plenge RM, Hendrich BD, Schwartz C, Arena JF, Naumova A, Sapienza C, Winter RM, Willard HF. A promoter mutation in the XIST gene in two unrelated families with skewed X-chromosome inactivation. *Nat Genet*. 1997 Nov; 17(3): 353-6.

Puck J and Willard H. X inactivation in females with X-linked disease. *The New England Journal of Medicine*. 1998; 338(29): 325-8.

Radic CP, Rossetti LC, Abelleyro MM, Tetzlaff T, Candela M, Neme D, Sciuccati G, Bonduel M, Medina-Acosta E, Larripa IB, de Tezanos Pinto M, De Brasi CD. Phenotype-genotype correlations in hemophilia A carriers are consistent with the binary role of the phase between F8 and X-chromosome inactivation. *J Thromb Haemost*. 2015 Apr; 13(4): 530-9.

Sado T, Wang Z, Sasaki H, Li E. Regulation of imprinted X-chromosome inactivation in mice by Tsix. *Development*. 2001 Apr; 128(8): 1275-86.

Sambrook J, Fritsch EF, Maniatis T. *Molecular cloning. A laboratory manual*. Second Edition. Cold Spring Harbor Laboratory Press. New York/USA, 1989.

Starmer J, Magnuson T. A new model for random X chromosome inactivation. *Development*. 2009; 136:1-10

Sun S, Del Rosario BC, Szanto A, Ogawa Y, Jeon Y, Lee JT. Jpx RNA activates Xist by evicting CTCF. *Cell* 2013 Jun 20; 153(7):1537-51.

Tian D, Sun S, Lee JT. The long noncoding RNA, Jpx, is a molecular switch for X chromosome inactivation. *Cell*. 2010 Oct 29;143(3):390-403.



Tomkins DJ1, McDonald HL, Farrell SA, Brown CJ. Lack of expression of XIST from a small ring X chromosome containing the XIST locus in a girl with short stature, facial dysmorphism and developmental delay. *Eur J Hum Genet.* 2002 Jan; 10(1): 44-51.

Williams IJ, Abuzenadah A, Winship PR, Preston FE, Dolan G, Wright J, Peake IR, Goodeve AC. Precise carrier diagnosis in families with haemophilia A: use of conformation sensitive gel electrophoresis for mutation screening and polymorphism analysis. *Thromb Haemost.* 1998 Apr; 79(4): 723-6.

Wutz A, Rasmussen TP, Jaenisch R. Chromosomal silencing and localization are mediated by different domains of Xist RNA. *Nat Genet.* 2002 Feb; 30(2): 167-74.