

EFECTO FUNDADOR EN EL CONGLOMERADO GENÉTICO GEOGRÁFICO DE SÍNDROME X FRÁGIL DE RICAURTE – VALLE DEL CAUCA

JULIÁN ANDRÉS RAMÍREZ CHEYNE. Investigador. Estudiante de Doctorado en Ciencias Biomédicas con énfasis en Genética Médica. Universidad del Valle.

WILMAR SALDARRIAGA GIL. Director. Profesor Universidad del Valle.

MAURICIO ARCOS. Asesor. Profesor Universidad de Antioquia.

CESAR PAYÁN. Asesor. Profesor Universidad del Rosario.



UNIVERSIDAD DEL VALLE
FACULTAD DE SALUD
ESCUELA DE CIENCIAS BÁSICAS
DEPARTAMENTO DE MORFOLOGÍA
Santiago de Cali, Septiembre de 2019

Contenido

Índice de tablas, figuras y anexos	4
Título	7
Resumen y palabras clave	7
Abstract and key words	7
Planteamiento del problema y justificación	8
Objetivos	12
Objetivo general	12
Objetivos específicos	12
Marco teórico	12
Síndrome X frágil.....	12
Epidemiología del síndrome X frágil y trastornos relacionados	13
Diagnóstico molecular de síndrome X frágil y trastornos relacionados.....	14
Variación genética de las poblaciones	15
Emparejamiento no aleatorio	15
Frecuencias alélicas inconstantes	16
Análisis de haplotipos en síndrome X frágil	17
Tecnologías de secuenciación	26
Secuenciación exómica y sus aplicaciones en genética poblacional.....	28
Metodología	30
Muestra.....	31
Análisis bioinformático.....	35
Flujo bioinformático para la obtención del archivo de SNP/Indels	35
Obtención de la matriz de SNPs e Indels	37
Búsqueda de un haplotipo común en los pacientes con mutación completa o premutación en FMR1 de la población de Ricaurte – Valle	44
Obtención de haplotipos.....	44
Construcción de redes de haplotipos.....	45
Haplotipo común.....	48
Determinación de las relaciones filogenéticas entre los individuos de Ricaurte con expansión en FMR1 (premutación o mutación completa), la estructura genética de la población, y distancia genética con respecto a poblaciones del proyecto 1000 genomas.....	51
Equilibrio Hardy-Weinberg.....	51

Heterocigosidad y endogamia	52
Análisis de Identidad por Estado (IBS)	52
Estructura genética de la muestra estudiada	55
Análisis de componentes principales	58
Determinación de distancia genética con poblaciones del proyecto 1000 genomas	63
Búsqueda de SNPs significativamente más frecuentes en los individuos con expansión en FMR1 de Ricaurte que presentan características fenotípicas específicas.....	65
Comparaciones realizadas	65
SNPs con frecuencia significativamente mayor en individuos de Ricaurte con expansión (mutación o premutación) que convulsionan	66
Gen PTCHD3	69
GEN ADAP1	70
SNPs con frecuencia significativamente mayor en mujeres de Ricaurte con expansión (premutación) que presentan falla ovárica prematura	71
Gen TRPM3	75
Gen CHD5.....	76
Gen ADGRF5.....	77
Gen PLCB4	78
SNPs con frecuencia significativamente mayor en individuos de Ricaurte con expansión (premutación) que presentan temblor-ataxia	79
Gen STXBP3.....	84
Gen PEX1.....	85
Gen ZNF430.....	86
Gen ROR1	87
Gen SMCHD1	88
Gen NHEJ1	89
Conclusiones y recomendaciones.....	90
Consideraciones Éticas.....	93
Referencias	93

Índice de tablas, figuras y anexos

Tabla 1: Genotipo para FMR1 y presencia o ausencia de convulsiones, síndrome de temblor-ataxia y síndrome de falla ovárica prematura en cada individuo de la muestra.....	31
Tabla 2: Datos crudos del exoma en secuencias de 150pb para los 55 individuos.....	35
Figura 1: Flujo bioinformático empleado para la obtención de la matriz de SNP e Indels.	36
Tabla 3: Cantidad de SNPs e Indels encontrados por cromosoma (1-22, X, Y, mitocondrial) para los 55 individuos evaluados.	38
Tabla 4: Datos perdidos de SNPs en los individuos.....	38
Figura 2: Proporción de SNPs perdidos por individuo en los 55 individuos estudiados.	40
Figura 3: Frecuencia del menor alelo por SNP para los 55 individuos estudiados.	41
Tabla 5: SNPs conservados para el análisis después de realizar el control de calidad (exclusión de los no secuenciados en al menos 95% de los individuos y exclusión de los monomórficos).	42
Tabla 6: Cantidad y posición de SNPs encontrados por cromosoma (1-22, X, Y) para los 55 individuos evaluados, después de realizar el control de calidad (exclusión de los no secuenciados en al menos 95% de los individuos y exclusión de los monomórficos).	43
Tabla 7: Haplotipos obtenidos del análisis estadístico de parsimonia (TCS), que trata los huecos como un estado de 5º carácter, para un conjunto de SNPs del cromosoma X para 55 individuos (24 hombres y 31 mujeres).	44
Figura 4: Red de haplotipos de los hombres de las familias construido a partir de 236 SNPs. Codificación de las familias: 1 (azul), 2 (amarillo), 4 (morado), 999 (gris) y el control (6:verde).	46
Figura 5: Redes de haplotipos de 55 individuos pertenecientes a las familias.....	47
Tabla 8: Frecuencias alélicas en los 37 haplotipos de 58 SNPs del cromosoma X.	48
Figura 6: Haplotipo conservado en los hombres de Ricaurte con expansión.	50
Tabla 9: Cantidad de SNPs que se encuentran en desequilibrio Hardy-Weinberg por cada cromosoma.	51
Tabla 10: Heterocigosidad observada (H_o), Heterocigosidad esperada (H_e), y FIS por familia de 12381 SNPs evaluados.	52
Tabla 11: Matriz de disimilitud basado en IBS para 12381 SNPs en 55 individuos pertenecientes a seis familias.	53
Figura 7: Gráficos de Identidad por Estado para 12381 SNPs en 55 individuos.	53
Tabla 12: Estadísticas del STRUCTURE resultante del análisis de los 55 individuos utilizando 12381 SNPs. El análisis de Delta K seleccionó $K = 2$ grupos óptimos (en rojo).	56
Figura 8: Proporción de miembros de cada población predefinida en cada uno de los 2 grupos de la estructura genética definida por STRUCTURE.....	56
Tabla 13: Probabilidades de pertenencia de los 55 individuos a los 2 grupos genéticos óptimos inferidos a partir de STRUCTURE.....	57
Figura 9: Resultados de los análisis de STRUCTURE de 55 individuos utilizando 12381 SNPs. Óptimo $K = 2$. Las barras verticales representan la probabilidad de pertenencia de un individuo a un grupo genético con colores que designan los diferentes grupos.	58
Tabla 14: Las probabilidades de pertenencia de la familia a los 2 grupos genéticos óptimos inferidos a partir de STRUCTURE.....	58
Figura 10: Análisis de Componentes principales en los 55 individuos.....	58
Tabla 15: Resultados del Análisis de Varianza Molecular (AMOVA) basado sobre 12381 SNPs para cuatro familias.....	59

Tabla 16: Valores de FST entre seis parejas de familias (debajo de la diagonal). Valores de P-valor para los Fst calculados (encima de la diagonal, en rojo los significativos $P < 0.05$).....	60
Figura 11: Análisis de Componentes principales para 236 SNPs-Indels del cromosoma X de 55 individuos.	61
Figura 12: Análisis de Identidad por Estado para 236 SNPs-Indels del cromosoma X en 55 individuos.	62
Figura 13: Árbol filogenético Neighbor-Joining (NJ) basado en 236 SNPs-Indels del cromosoma X para los 55 individuos.	62
Tabla 17: Identificación de las poblaciones del proyecto 1000 genomas utilizadas en el análisis. ..	63
Figura 14: Análisis de Componentes principales para 95635 SNPs del cromosoma X sin filtrar de 55 individuos más 2504 individuos de 26 orígenes geográficos.	64
Figura 15: Árbol filogenético Neighbor-Joining (NJ) basado en 236 SNPs-Indels para 55 individuos más 2504 individuos de 26 orígenes geográficos.	65
Tabla 18: SNPs significativamente más frecuentes en pacientes con expansión (premutación o mutación completa) que presentaban convulsiones.	66
Tabla 19: Características de los genes que albergan las variantes encontradas con mayor frecuencia en los individuos expandidos que presentan convulsiones.	67
Figura 16: Red de interacciones moleculares de 25 componentes para el gen PTCHD3. Interacciones encontradas: 98. Interacciones esperadas: 1.....	69
Figura 17: Red de interacciones moleculares de 25 componentes para el gen ADAP1. Interacciones encontradas: 33. Interacciones esperadas: 7.....	70
Tabla 20: SNPs significativamente más frecuentes en mujeres con expansión (premutación) que presentaban falla ovárica prematura.....	71
Tabla 21: Características de los genes que albergan las variantes encontradas con mayor frecuencia en mujeres expandidas (premutación) con falla ovárica prematura.	72
Figura 18: Red de interacciones moleculares de 25 componentes para el gen TRPM3. Interacciones encontradas: 47. Interacciones esperadas: 9.....	75
Figura 19: Red de interacciones moleculares de 25 componentes para el gen CHD5. Interacciones encontradas: 260. Interacciones esperadas: 9.....	76
Figura 20: Red de interacciones moleculares de 25 componentes para el gen ADGRF5. Interacciones encontradas: 22. Interacciones esperadas: 1.....	77
Figura 21: Red de interacciones moleculares de 25 componentes para el gen PLCB4. Interacciones encontradas: 281. Interacciones esperadas: 24.	78
Tabla 22: SNPs significativamente más frecuentes en pacientes con expansión (premutación) que presentaban temblor-ataxia.	79
Tabla 23: Características de los genes que albergan las variantes encontradas con mayor frecuencia en los individuos expandidos (premutación) con temblor-ataxia.....	80
Figura 22: Red de interacciones moleculares de 25 componentes para el gen STXBP3. Interacciones encontradas: 129. Interacciones esperadas: 4.	85
Figura 23: Red de interacciones moleculares de 25 componentes para el gen PEX1. Interacciones encontradas: 117. Interacciones esperadas: 7.	86
Figura 24: Red de interacciones moleculares de 25 componentes para el gen ZNF430. Interacciones encontradas: 28. Interacciones esperadas: 4.....	87

Figura 25: Red de interacciones moleculares de 25 componentes para el gen ROR1. Interacciones encontradas: 95. Interacciones esperadas: 11.	88
Figura 26: Red de interacciones moleculares de 25 componentes para el gen SMCHD1. Interacciones encontradas: 22. Interacciones esperadas: 4.	89
Figura 27: Red de interacciones moleculares de 25 componentes para el gen NHEJ1. Interacciones encontradas: 291. Interacciones esperadas: 9.	90
Anexo 1: Acta de aprobación del Comité de Ética	101

Título

Efecto fundador en el conglomerado genético geográfico de síndrome X frágil de Ricaurte – Valle del Cauca.

Resumen y palabras clave

Ricaurte es un corregimiento del municipio de Bolívar ubicado al norte del Departamento del Valle del Cauca donde se presenta un conglomerado geográfico genético de síndrome X frágil. Los casos se concentran en tres núcleos familiares en los cuales se identificaron tres parejas fundadoras de la población, por lo que se ha propuesto un efecto fundador como causa de la alta prevalencia de la enfermedad en la población. Por tanto, se planteó este estudio para, a través del análisis de SNPs obtenidos mediante secuenciación exómica de 54 pacientes del clúster y un individuo normal, confirmar o descartar la presencia del efecto fundador.

Se encontró un haplotipo común de 33 SNPs comprendido en 83.567.899 pb circundantes a FMR1 en todos los hombres de Ricaurte portadores de la expansión, con lo cual se comprobó que el conglomerado genético de síndrome X frágil en Ricaurte se debe a un efecto fundador.

Además, se hicieron algunos análisis exploratorios iniciales, sobre aspectos que constituirán objetivos de investigaciones posteriores más profundas, encontrando que los individuos analizados presentan SNPs en desequilibrio de H-W en todos los autosomas, lo cual es esperado por tratarse de una población pequeña y aislada socialmente por el conglomerado X frágil. Sin embargo, el coeficiente de endogamia es negativo. Al comparar los individuos de este estudio con los de 1000 genomas, los individuos de Ricaurte y zonas aledañas portadores de la expansión, se ubicaron con los colombianos de 1000 genomas, más cercanos a Europa, América y Asia que a África. Se encontraron siete variantes tipo SNP, en seis genes, significativamente más frecuentes en individuos expandidos (mutación completa o premutación) con convulsiones. Solamente las ubicadas en el gen PTCHD3 pasaron Bonferroni. Se encontraron 12 variantes tipo SNP, en 9 genes, significativamente más frecuentes en mujeres expandidas (premutadas) con falla ovárica prematura. Solamente la ubicada en el gen TRPM3 pasó Bonferroni. Se encontraron 15 variantes tipo SNP, en 11 genes, significativamente más frecuentes en individuos expandidos (premutación) con temblor-ataxia. Ninguna pasó Bonferroni.

Palabras clave: exoma, análisis por conglomerados, síndrome del cromosoma X frágil, haplotipos, efecto fundador.

Abstract and key words

Ricaurte is a village in the municipality of Bolívar located in the north of the Department of Valle del Cauca where a genetic geographic conglomerate of fragile X syndrome is presented. The cases are concentrated in three family nuclei in which three founding couples

of the population were identified, for which reason a founder effect has been proposed as a cause of the high prevalence of the disease in the population. Therefore, this study was proposed to, through the analysis of SNPs obtained by means of exome sequencing of 54 patients from the cluster and a normal individual, confirm or rule out the presence of the founder effect.

A common haplotype of 33 SNPs extended in 83,567,899 bp surrounding FMR1 was found in all the Ricaurte men carrying the expansion, which proved that the genetic cluster of fragile X syndrome in Ricaurte is due to a founder effect.

In addition, some initial exploratory analyzes were carried out, on aspects that will constitute objectives of deeper further investigations, finding that the individuals analyzed present SNPs in HW imbalance in all autosomes, which is expected because this is a small and socially isolated population. However, the inbreeding coefficient is negative. When comparing the individuals in this study with those of 1000 genomes, the individuals of Ricaurte and surrounding areas carrying the expansion, were located with the Colombians of 1000 genomes, closer to Europe, America and Asia than to Africa. Seven SNP-type variants were found, in six genes, significantly more frequent in expanded individuals (complete mutation or premutation) with seizures. Only those located in the PTCHD3 gene passed Bonferroni. 12 SNP-type variants were found, in 9 genes, significantly more frequent in expanded (premutated) women with premature ovarian failure. Only the one located in the TRPM3 gene passed Bonferroni. Fifteen SNP-type variants were found, in 11 genes, significantly more frequent in expanded individuals (premutation) with tremor-ataxia. None passed Bonferroni.

Key words: exome; cluster analysis; fragile X syndrome; haplotypes; founder effect.

Planteamiento del problema y justificación

“Seguramente ninguno de los otros 38 bobos de Ricaurte habría hecho lo que César Augusto hizo. Pero él, en sí, también era una excepción en origen, comportamiento y agresividad. Primero, no pertenecía a las familias que tradicionalmente le han aportado, por siglos y siglos, los bobos al pueblo. Segundo, inicialmente, nadie creyó que era bobo y solo cuando Calixto y Maximiliano, los dirigentes de la tribu de intonsos, le fueron enseñando los trucos y los gestos, las carantoñas y las pataletas y se le comenzó a descolgar la cara y empezaron a chorreársele las babas, Ricaurte aceptó que el César Augusto hacía parte de su batallón de anormales. Era también el único bobo que hablaba sin gaguear y sin sonidos guturales intermitentes en sus frases. Eso sí, desde muy chico fue lentísimo para decir las palabras y solo cuando se enfadaba o enfatizaba sus odios, aceleraba un poco su gangosería para aparentar la velocidad de un disco 33 revoluciones tocado en 45. Generalmente ha ido muy bien vestido. Sus hermanos, que trabajaban todos en fincas del otro lado del Cauca y su madre, que ha hecho canastas de alambre para vender en Tuluá y en Pereira, guardan de su

apariencia con celosa rigurosidad. No atinan a explicar por qué resultó bobo, pero custodian sus acciones como si fueran enteramente culpables de su tara.”

“En el momento en el que Rafael Urriago silbó las mulas en medio del aguacero pantanoso que se abría paso por entre la muralla verde de las selvas chocoanas y ellas se arrimaron a reposar debajo de mismo palo y al pie de la misma peña en donde cada viaje descargaban su fatiga y hundían sus hocicos babosos para atragantar la sed de tantos días de viaje, ni él, ni Domingo Victoria, su socio de comercio en Istmina, alcanzaron a darse cuenta del pedazo de tabla que salía por entre medio del torrente que bajaba de la peña.”

“Pero como el aguacero no calmó en toda la tarde y la siguiente jornada se hacía muy larga en la curtida experiencia de Urriago, solo a la mañana siguiente, cuando cansado de oír llover metió en su cabeza la fuerza del chorro de la peña y se topó con la tabla incrustada entre dos piedras y, sin pensar siquiera que al sacarla podría quebrar el equilibrio del peñasco o abrir el boquete del torrente, con las dos manos entumecidas tomó el trozo de madera pulido, perfectamente cuadrado, en donde el barro y el agua habían ido dejando una costra borrosa que no se iba a poner a examinar con las tenues luces de la mañana que alcanzaban a dejarse sentir por entre medio del follaje. Simplemente la sacudió, le quitó un poco de barro y lama con la misma fuerza del chorro y la metió en la jíquera vacía de su mula carrizosa. De pronto le servía para picar las puntas de los arreos y enredar los picos de las amarras.

Veinte días después, cuando la mula campanera pisó los primeros carrizales de la montaña blanca y el ventarrón seco se metió vega abajo buscando el valle para perderse, Rafael Urriago alcanzó a ver allá, en lo profundo, los techos pajizos de los ranchos de su pueblo, el ciruelo braziabierto de su casona y los nueve ranchos iguales que la rodeaban como círculo protector.

Era un paisaje demasiado conocido para quien durante tantos años había ido y vuelto a irse por la misma ruta que los indios de la encomienda de su papá le enseñaron. Un paisaje idéntico, seco, absurdamente diferente del que tenía que recorrer con solo traspasar el filo de la montaña o bajar una legua hacia el Cauca. Por allí no corría el agua dando saltos montaña abajo, sino abriéndose paso ayudada por el viento porque la sombra de los árboles no existía y el tenue verdor de las tunas y las zarzas apenas si marcaba los límites.

Pero como estaba cansado de trajinar y se sentía doblado por los años, Rafael Urriago descargó sus mulas, entregó a sus hijos la mercancía, se quedó para sí y para su vieja con las cáscaras de las resinas y, por supuesto, con la tabla que encontró en la peña. Y cuando descansó sus bríos y arrulló sus caricias en la molicie del ventarrón, totuma en mano, se puso a limpiar la tabla y a encontrar en ella los rasgos diferentes, indefinibles, de una pintura que no era pintura, de un dibujo que no podía ser dibujo y sin advertir qué clase de fuerza interior se iba apoderando de su mano y de su ansiedad, prefirió dejar la tabla sobre el primer estante de su rancho, confiado en que si no era un cuadro podría ser alguna imagen santa que el tiempo hizo inviolable bajo el agua de la peña.

Afortunadamente por su casa seguían pasando las otras mulas que iban y venían de Toro a Popayán y, al primer cura que se detuvo en ella a defenderse del viento huracanado que no dejaba avanzar, le sacó su concepto y con él se satisficieron sus hijos y sus nietos y los hijos de sus biznietos asegurando que desde ese día, cuando el cura se percató que estaba frente a un cuadro del Ecce Homo, Rafael Urriago nunca más volvió a quebrar montaña detrás de sus mulas o a comerciar con los pantanos del Chocó y primero en su propio aposento, después al pie del ciruelo en el rancho que le construyeron, cuidó, hasta el último de sus días, de la imagen que milagrosamente había topado.

A su muerte, el cuadro quedó en manos de María Sixta Urriago, la mujer de Rubén Cruz, su hermana de amor, y quien fuera la primera testigo de las curaciones milagrosas del santo aparecido, que poco a poco, con los años y los rumores, se volvió propiedad de los Cruz y fuente inagotable de leyendas y favores hasta el punto que cuando Simón Pablo Cruz no pudo volver a vender la leche de cabras ni a engordar los marranos de su herencia, se inventó la fórmula para que el santo rindiera y el hambre no le consumiera.

María Cruz de Triviño, su nieta, cargó con la herencia y el manejo, las prebendas y la fama y en un pequeño santuario de paja, al pie del mismo ciruelo donde Rafael Urriago colocara aquel último domingo de agosto la tablita milagrosa, se hizo a su santo particular, a su Divino Ecce Homo y cada año, el mismo domingo de agosto que se dijo que Rafael Urriago bajó con él de la montaña, armó la fiesta, la novena y la procesión, el peregrinar y la lista de milagros (que registraba religiosamente en un libraco mamotrético que había conseguido en Roldanillo) hasta hacer girar el pequeño caserío, que ya tenía diez veces más casas que el paisaje contemplado por Rafael Urriago, alrededor de su devoción.”

Fragmentos de la novela El Divino (1986). Gustavo Álvarez Gardeazabal

Ricaurte es un humilde corregimiento del municipio de Bolívar ubicado al norte del Departamento del Valle del Cauca, en el margen occidental del río Cauca, sobre el piedemonte de la cordillera occidental, conocido por la imagen del Santo Ecce Homo que reposa en su iglesia y en honor a la cual se hacen celebraciones en el mes de agosto de cada año, y también conocido por la alta prevalencia de discapacidad intelectual observada entre sus habitantes.

“Genética es el estudio de la herencia, genética médica es el estudio de la herencia de enfermedades y caracteres anormales, genética médica poblacional es cuando el enfermo es una población. Genética médica poblacional es el estudio de los factores genéticos de enfermedades que son frecuentes en algunas poblaciones.”

Eduardo Castilla.

Genetista del Instituto Nacional de Genética Médica Poblacional de Brasil

Genetistas y epidemiólogos del Instituto Nacional de Genética Médica Poblacional (INAGEMP), creado en 2008 por el gobierno brasileño para el estudio de genética médica a nivel poblacional. Dentro del marco de este instituto, inició en el 2009 el Censo Nacional de Aislamientos (CENISO), un programa específico para el estudio de poblaciones

genéticamente aisladas que funciona como un sistema de vigilancia de sub poblaciones con enfermedades crónicas endémicas, en su mayoría mendelianas, enfocado en su estudio, evaluación de necesidades, prevención y cuidado. (Castilla & Schuler-faccini, 2014)

Los expertos del INAGEMP han estudiado por años este tipo de poblaciones y han descrito una metodología para abordar el estudio de conglomerados genéticos, dicha metodología consiste en los siguientes pasos: (Castilla & Schuler-faccini, 2014)

1. Rumor (siempre se parte de un rumor).
2. Validación del rumor.
3. Generación de una hipótesis.
4. Diseño de protocolos.
5. Recolección de datos.
6. Análisis de datos.
7. Abordaje del problema de salud.

Durante muchos años, la elevada prevalencia de discapacidad intelectual en Ricaurte popularmente se atribuyó a causas ambientales e incluso maldiciones, hasta que a finales de los años noventa un grupo de investigadores de la Universidad del Valle realizó un estudio encontrando mediante cariotipo con déficit de folato que la causa de la discapacidad intelectual era el síndrome X frágil. (Saldarriaga et al., 2014)

Recientemente, el grupo MACOS (Malformaciones Congénitas Perinatales y Dismorfología) de la Universidad del Valle en colaboración con el instituto MIND (Medical Investigation of Neurodevelopmental Disorders) de la Universidad UC Davis - California, mediante tamizaje con PCR (con dos y tres cebadores) aplicado a toda la población y SouthernBlot como prueba confirmatoria, demostró que en Ricaurte existe un conglomerado geográfico genético de síndrome X frágil, con las prevalencias de las variantes alélicas del gen FMR1 (zona gris, mutación completa y premutación) más altas reportadas en la literatura. Además, se construyeron heredogramas sistematizados de las familias en las que se encontraron casos de mutación completa y premutación, encontrando que los casos se concentraban en 3 núcleos familiares en los cuales se identificaron 3 fundadores de la población, por lo que se propuso un efecto fundador. Por tanto, se planteó este estudio para, a través del análisis de SNPs obtenidos mediante secuenciación exómica de pacientes del clúster, responder a las preguntas:

- ¿Los individuos con expansión en FMR1 (premutación o mutación completa) de Ricaurte heredaron la alteración genética de un ancestro común?
- ¿Cuál es la estructura genética de la población estudiada?
- ¿Entre los individuos de Ricaurte con expansión en FMR1 (premutación o mutación completa), existen SNPs significativamente más frecuentes en quienes presentan convulsiones, temblor ataxia o falla ovárica prematura?

Objetivos

Objetivo general

Determinar si existe un efecto fundador en el conglomerado genético geográfico de síndrome X frágil de Ricaurte – Valle del Cauca.

Objetivos específicos

Determinar si existe un haplotipo común en el cromosoma X de los individuos de Ricaurte con expansión en FMR1 (premutación o mutación completa).

Determinar las relaciones filogenéticas entre los individuos de Ricaurte con expansión en FMR1 (premutación o mutación completa), la estructura genética de la población estudiada y las distancias genéticas con otros grupos poblacionales incluidos en la base de datos del proyecto 1000 genomas.

Identificar si en individuos de Ricaurte con expansión en FMR1 (premutación o mutación completa), existen SNPs significativamente más frecuentes entre los que presentan convulsiones, temblor ataxia o falla ovárica prematura.

Marco teórico

Síndrome X frágil

El gen *FMR1* (Fragile X Mental Retardation 1) está localizado en Xq27.3 y codifica para la proteína de unión a RNA denominada FMRP (Fragile X Mental Retardation Protein), cuya función está relacionada con la regulación postranscripcional de la expresión de genes que codifican para proteínas involucradas en plasticidad neuronal y el desarrollo del cerebro y las sinapsis. (Darnell et al., 2011)(V. Brown et al., 2001)(Chen, Yun, Seto, Liu, & Toth, 2003) Este gen alberga una repetición del trinucleótido CGG en su región 5' no traducida. La mutación dinámica con expansión de estas repeticiones, que puede ocurrir en la gametogénesis femenina, produce fenotipos patológicos.(Bagni, Tassone, Neri, & Hagerman, 2012)

Se han establecido 4 rangos de repeticiones de CGG con consecuencias fenotípicas diferentes, los cuales constituyen los 4 tipos de alelo definidos para el gen FMR1: (Saldarriaga et al., 2014)

- Alelos normales: hasta 44 repeticiones.
- Alelos intermedios o de la zona gris: 45 a 54 repeticiones. Considerados precursores de alelos de la premutación.
- Alelos de la premutación: 55 a 200 repeticiones.
- Alelos de la mutación completa: más de 200 repeticiones.

Cuando se tiene un alelo con 55 a 200 repeticiones (premutación), se produce acumulación de mRNA de FMR1 debido a un desbalance entre los procesos de transcripción y traducción, de modo que los niveles de mRNA de FMR1 se elevan 2 a 8 veces por encima de lo normal. Estas moléculas de mRNA acumuladas tienen, en la región correspondiente a la expansión, una estructura de horquilla que les confiere la propiedad de secuestrar proteínas importantes para la función neuronal. Esta toxicidad por mRNA puede explicar el hecho de que los portadores de la premutación tengan riesgo aumentado de presentar ciertas condiciones médicas, psiquiátricas y neurológicas, como síndrome de temblor/ataxia asociado a X frágil, insuficiencia ovárica primaria asociada a X frágil, depresión, ansiedad, migraña, hipertensión, apnea del sueño y enfermedades mediadas por mecanismos inmunes como hipotiroidismo y fibromialgia. (Hagerman & Hagerman, 2013)

Cuando se tiene un alelo con más de 200 repeticiones (mutación completa), se producen cambios epigenéticos complejos que implican una metilación de las repeticiones, la cual se extiende corriente arriba hasta de la isla CpG de la región promotora de FMR1, llevando al silenciamiento del gen, a la no producción de la proteína FMRP por parte de ese cromosoma (Verkerk AJ, Pieretti M, Sutcliffe JS, Fu YH, Kuhl DP, Pizzuti A, Reiner O, Richards S, Victoria MF, 1991) (Chonchaiya, Schneider, & Hagerman, 2009) (Levenga J, de Vrij FM, Oostra BA, 2010) (Fu et al., 1991) y consecuentemente a aumento de los niveles de un amplio rango de proteínas, dismorfogénesis de las espinas dendríticas y desbalance excitación/inhibición (glutamato/GABA). (Bear, Huber, & Warren, 2004) (Gatto & Broadie, 2010) Entre la región de las repeticiones y el promotor, normalmente existe una secuencia límite que al metilarse induce cambios en la disposición de la cromatina e impide que la metilación se extienda hasta la región promotora. Se ha propuesto que cuando este mecanismo protector no funciona, es que se produce el silenciamiento. (Sutcliffe et al., 1992) Si se trata de una mujer portadora de la mutación completa en heterocigosidad, la mayoría de las veces esta tendrá, por contar con un segundo alelo no patológico, un fenotipo normal. Pero si el portador del alelo con la mutación completa es un hombre, la mayoría de las veces este tendrá, por ser hemicigoto para el cromosoma X, ausencia total de FMRP y un fenotipo típico de síndrome X frágil el cual consiste en retraso mental, cara alargada, orejas grandes y macroorquidia. (McKusick-Nathans Institute of Genetic Medicine, 2018)

A pesar de que existen diferentes opciones farmacológicas para el manejo sintomático, (Saldarriaga et al., 2014) no existen terapias curativas con impactos significativos en el fenotipo comportamental, lo cual evidencia la importancia de los trastornos asociados a FMR1 como objeto de investigación.

Epidemiología del síndrome X frágil y trastornos relacionados

La prevalencia de síndrome X frágil es de 1 en 5.000 hombres (Coffee et al., 2009) y 1 en 4.000-6.000 mujeres. (Coffee et al., 2009) (Hirst et al., 1993) Es considerada la causa hereditaria más común de discapacidad intelectual, (Verkerk AJ, Pieretti M, Sutcliffe JS, Fu YH, Kuhl DP, Pizzuti A, Reiner O, Richards S, Victoria MF, 1991)

Para Colombia no existen datos de la prevalencia de síndrome X frágil en población general, sin embargo, es reconocido el caso de Ricaurte, un corregimiento del municipio de Bolívar

ubicado al norte del Departamento del Valle del Cauca, que tiene una prevalencia elevada de discapacidad intelectual. A finales de los años 90, el grupo de investigación MACOS de la Universidad del Valle realizó en Ricaurte un estudio encontrando que la causa de la discapacidad intelectual era el síndrome X frágil. En ese estudio se determinó que, entre los 1124 habitantes, 35 tenían este síndrome, 19 diagnosticados a través de cariotipo con bandas G con déficit de folatos y 16 con diagnóstico clínico. Para el año 1999 el grupo MACOS determinó una prevalencia de síndrome X frágil en Ricaurte de 1 en 38 hombres y 1 en 100 mujeres, lo que excedió 100 veces lo reportado en la literatura. (Saldarriaga et al., 2014)

La prevalencia de la premutación en la población general es 1 en 130-200 mujeres y 1 en 250-450 hombres.(Fernandez-Carvajal et al., 2009)(Tassone et al., 2012) El síndrome de temblor/ataxia asociado a X frágil ocurre en aproximadamente 40% de los hombres con la premutación y 16% de mujeres con la premutación, mientras que la insuficiencia ovárica primaria asociada a X frágil ocurre en aproximadamente 16% de mujeres con la premutación. (Hagerman RJ & Hagerman PJ Editors, 2002)(Hagerman & Hagerman, 2013)

Para Colombia no existen datos de la prevalencia de la premutación en población general. En el año 2016 en un trabajo colaborativo entre el grupo MACOS de la Universidad del Valle y el instituto MIND de la Universidad UC Davis - California se encontró en Ricaurte una prevalencia de la premutación en FMR1 de 1 en 23 mujeres y 1 en 120 hombres. (Saldarriaga et al., 2016)

Diagnóstico molecular de síndrome X frágil y trastornos relacionados

Entre las metodologías de análisis de DNA, el gold-standard para el diagnóstico de síndrome X frágil, es el uso de una combinación de PCR y SouthernBlot. La PCR es particularmente útil para la determinación del número de repeticiones de CGG en el rango de la premutación, mientras que el SouthernBlot, aunque es un procedimiento costoso y demorado que requiere mayores cantidades de DNA genómico de alto peso molecular, es ideal para la determinación de repeticiones en alelos más grandes y provee información sobre el estado de metilación. Individuos con deleciones o mutaciones puntuales, que representan una pequeña minoría, no serán detectados con las pruebas mencionadas.(Wells, 2009)(W. T. Brown et al., 1996)(Snow et al., 1993)

Sin embargo, el número muestras que pueden analizarse simultáneamente mediante SouthernBlot es limitado, por lo cual se han propuesto varias estrategias diagnósticas basadas en PCR para la identificación de repeticiones CGG en el gen FMR1. Las metodologías basadas en PCR convencional no alcanzan una amplificación confiable de los fragmentos grandes de repeticiones CGG debido a que el alto contenido de CG conlleva una tendencia de estos fragmentos a formar estructuras secundarias no desnaturalizables.(Tassone, 2015a)

La mayoría de las aproximaciones para superar este problema, implican la combinación de polimerasas y co-solventes como DMSO y betaína; y el uso de 7-deaza-2'GTP, que son guaninas con un reemplazo del nitrógeno 7 por un grupo NH que evita la formación de 3 puentes de hidrogeno entre las bases Citosina y Guanina, pero que requiere métodos de detección diferentes al bromuro de etidio, como tinciones de plata o hibridación con sondas

radiactivas o quimioluminiscentes. La técnica TR-PCR (TripletRepeat-primed PCR) es capaz de identificar todos los alelos expandidos desde los normales hasta los del rango de la mutación completa, de distinguir entre mujeres normales y mujeres con la mutación completa y proveer información sobre el tamaño de la repetición. (Tassone, 2015b)

En la técnica TR-PCR se utilizan 4 cebadores diferentes. Un primer cebador (FMR1f) tiene una secuencia complementaria a una región flanqueante de las repeticiones CGG en la 5'UTR del gen FMR1. Un segundo cebador, reverso llamado FMR1R, tiene una secuencia exacta a la secuencia flanqueante corriente abajo de la repetición de tripletas junto con una porción de repeticiones CGG. Así, después del segundo ciclo de PCR se obtienen amplicones con el número completo de tripletas (cuando el FMR1R se unió a la secuencia flanqueante) y con un número menor de repeticiones (si el FMR1R se unió únicamente con la secuencia de repeticiones CGG). Dos cebadores adicionales FMR1CCGR y M13 se utilizan para amplificar la señal y el patrón denominado “stutter” o de interrupciones en el electroferograma. El análisis del producto de la reacción se hace utilizando electroforesis capilar que arroja un electroferograma donde el eje Y indica número de señales recibidas y el eje X el tamaño de la hebra en pares de bases estimado utilizando la velocidad de desplazamiento. (Tassone, 2015b)

Variación genética de las poblaciones

La disparidad interpoblacional en frecuencias alélicas llevó al surgimiento de la genética poblacional para el estudio de las frecuencias alélicas en la población, su comportamiento y los factores que las afectan, que pueden ser genéticos como las mutaciones y la reproducción, ambientales como la selección y sociales como la migración. La genética poblacional permite indagar sobre la historia, estructura genética, susceptibilidad genética a enfermedades y frecuencias de alelos patológicos en las poblaciones que permiten el cálculo de riesgos para la asesoría genética. Debido a que existen relaciones directas entre las frecuencias alélicas y las frecuencias genotípicas, se puede aplicar la ley de Hardy – Weinberg a una población, para estimar las frecuencias genotípicas a partir de las frecuencias alélicas, siempre y cuando se cumpla que la población es grande, los emparejamientos ocurren al azar y las frecuencias alélicas son constantes. Estos supuestos no siempre se cumplen. (Nussbaum, McInnes, & Willard, 2016)

Emparejamiento no aleatorio

Una situación de emparejamiento no aleatorio puede ocurrir debido a: (Nussbaum et al., 2016)

- Estratificación: emparejamientos entre miembros de subgrupos genéticamente aislados.
- Emparejamiento dirigido: elección de parejas por rasgos particulares como talla baja, ceguera o sordera.
- Consanguinidad: emparejamiento entre familiares.
- Endogamia: emparejamientos entre miembros de poblaciones pequeñas con ancestros comunes.

Frecuencias alélicas inconstantes

Una situación de inconstancia de las frecuencias alélicas puede ocurrir debido a: (Nussbaum et al., 2016)

- Selección y eficacia biológica de los alelos.
- Migración y flujo génico: la migración entendida en el sentido amplio de cruzar una barrera reproductiva, que puede ser racial, étnica o cultural, y no necesariamente geográfica, puede cambiar las frecuencias alélicas de una población. Los genes de una población migrante con sus frecuencias alélicas propias son gradualmente mezclados con el conjunto de genes de la población a la han migrado. Este proceso de mezcla genética que involucra una población grande, una difusión lenta de genes a través de una barrera y un cambio gradual en las frecuencias génicas, se denomina flujo génico.
- Deriva genética: los eventos aleatorios pueden tener un efecto mucho mayor sobre las frecuencias alélicas en poblaciones pequeñas que en poblaciones grandes. Por ejemplo, cuando una mutación nueva ocurre en una población pequeña, su frecuencia está representada por solo una copia entre todas las copias que hay de ese gen en la población. Efectos aleatorios del ambiente u otros eventos aleatorios independientes del genotipo pueden producir cambios significativos en la frecuencia del alelo patológico cuando la población es pequeña. Estos eventos aleatorios causan que la frecuencia alélica cambie de una generación a la siguiente. Este fenómeno conocido como deriva genética, puede explicar cómo las frecuencias alélicas pueden cambiar como resultado del azar. Durante algunas pocas generaciones siguientes, aunque el tamaño de la población del nuevo grupo se mantiene pequeño, puede haber fluctuación considerable en la frecuencia génica hasta que las frecuencias alélicas lleguen a un nuevo equilibrio a medida que la población aumenta su tamaño. En contraste con el flujo génico, un proceso lento en el cual las frecuencias alélicas cambian a causa de mezcla de poblaciones previamente distintas, el mecanismo de la deriva genética es simplemente el azar actuando rápidamente sobre las frecuencias alélicas de una población pequeña.
- Efecto fundador: Cuando una subpoblación pequeña se separa de una más grande, las frecuencias génicas en la población pequeña pueden ser diferentes a las de las frecuencias en la población original, porque el nuevo grupo contiene una muestra pequeña y aleatoria de la grande, y por azar puede no tener las mismas frecuencias génicas. Si uno de los fundadores originales del nuevo grupo es portador de un alelo relativamente raro, este alelo tendrá una frecuencia mucho más alta con respecto a la que tenía en el grupo más grande del cual se derivó. Cuando esto sucede se dice que ha ocurrido una forma especial de deriva genética denominada efecto fundador.

Análisis de haplotipos en síndrome X frágil

El análisis de haplotipos de cromosomas normales y mutantes entre diversos grupos poblacionales ha indicado que una proporción sustancial de todos los cromosomas X frágil muestran desequilibrio de ligamiento con un subconjunto pequeño de haplotipos DXS548-FRAXAC2-FRAXAC1.(Eichler et al., 1996) Incluso, en poblaciones más genéticamente aisladas como Finlandia, el desequilibrio de ligamiento se ha encontrado en el 75% de todos los cromosomas X frágil ocurriendo en un único haplotipo DXS548-FRAXAC2. (Eichler et al., 1996)

Estudios poblacionales a gran escala en caucásicos y población afroamericana respecto al origen del síndrome X frágil han identificado diversos factores potencialmente involucrados en la inestabilidad de la repetición CGG, es así como el análisis de los patrones de repetición y los haplotipos flanqueantes (DXS548-FRAXAC1-FRAXAC2; 3 loci polimórficos ubicados dentro de 150 kb de la repetición CGG FMR1) en 214 individuos caucásicos normales y 16 individuos con premutación, ha demostrado dos vías mutacionales distintas que predisponen hacia la expansión de repeticiones CGG.(Eichler et al., 1996; Huang et al., 2015)

Una de estas vías involucra un haplotipo DSX548-FRAXAC1-FRAXAC2: 2-1-3, que responde al 14% de todos los cromosomas X frágil, mantiene dos interrupciones AGG y progresa lentamente hacia la inestabilidad de forma escalonada. La segunda vía involucra dos haplotipos DSX548-FRAXAC1-FRAXAC2: 6-4-5 y 6-4-4, que responde a aproximadamente el 30% de todos los casos X frágil, implica una pérdida frecuente y recurrente de interrupciones AGG y en los cuales no se presenta enriquecimiento de los alelos de longitud normal.(Eichler et al., 1996; Huang et al., 2015)

Varios grupos han investigado la posibilidad de que la subestructura de AGG de la repetición CGG en FMR1 pueda jugar un papel crítico, al determinar la estabilidad y susceptibilidad a la enfermedad, los resultados de estos análisis revelaron que el tracto más largo de repeticiones CGG ininterrumpidas (puras) determina la inestabilidad en el locus FRAXA, observándose transmisiones intergeneracionales inestables en la longitud umbral de 34-37 repeticiones puras.(Eichler et al., 1996)Así mismo, Crawford y colegas, al analizar los patrones de repetición CGG de 213 afroamericanos no afectados, identificaron una vía mutacional independiente que implica que la ausencia de una interrupción AGG proximal puede incrementar la inestabilidad de la repetición CGG.(D. C. Crawford, Zhang, Wilson, Warren, & Sherman, 2000)

Teniendo en cuenta estos postulados, diversos autores han planteado estudios poblacionales para aclarar la etiología del síndrome X frágil. Ejemplo de esto es un estudio realizado en población finlandesa hacia el año 1996 para explorar la posibilidad de un efecto fundador en el síndrome X frágil mediante el análisis de un conjunto de polimorfismos de microsatélites relacionados (DXS548, FRAXAC1, FRAXE, y Ah-L). En el estudio se analizaron un total de 93 cromosomas X (56 normales y 37 X frágil) de Finlandia del este, amplificándose mediante PCR FRAXAC1, DXS548, FRAXE y las repeticiones CGG FMR1, incluyendo un

polimorfismo de conformación de cadena simple (SSCP) de Alu-L. Adicionalmente se llevó a cabo un análisis de las interrupciones AGG entre las repeticiones CGG. (Zhong et al., 1996)

Se encontró una marcada diferencia en la distribución de los alelos DXS548 entre controles (5 alelos) y cromosomas X frágil (2 alelos), siendo el 95% de los X frágil el alelo de 21 repeticiones CA (196 bp) comparado con el 17% de los controles ($P < 0,001$). Al comparar las frecuencias alélicas de DXS548 de los controles en Finlandia con tres poblaciones caucásicas de control (Nueva York, Francia y Bélgica), se encontró que el 69% de los alelos control de Finlandia tenían 20 repeticiones y que esto era similar a lo encontrado en las otras poblaciones, mientras que ninguno de los alelos X frágil en la población de Finlandia tenía este alelo. Para el caso de las 21 repeticiones, el ratio 5,6:1 (95%:17%) fue mucho mayor al encontrado en las otras poblaciones (21-30% en X frágil: 10-14% controles). Lo anteriormente expuesto provee evidencia para considerar una mutación fundadora en la población Finlandesa. (Zhong et al., 1996)

Respecto a las frecuencias alélicas de FRAXAC1 pese a estar presentes 19 repeticiones CA en el 84% de los cromosomas X frágil, no se encontró una diferencia significativa respecto a los controles (61%) y solo se presentó en el 34-38% de los casos X frágil en las otras comunidades caucásicas. (Zhong et al., 1996)

En cuanto a FRAXE, se encontraron 13 tamaños alélicos diferentes, que iban de 7 a 25 repeticiones GCC entre 56 controles y 36 cromosomas X frágil en la población finlandesa. Se encontró una heterocigosidad del 82% en los controles, pero de 42% en los casos. El alelo de 16 repeticiones FRAXE se presentó en el 75% de los cromosomas X frágil comparado con el 30% de los controles ($P < 0,001$), sin encontrarse estas diferencias entre los cromosomas caucásicos de Nueva York (25% X frágil versus 27% controles). (Zhong et al., 1996)

Al analizar los tres microsatélites (DXS548, FRAXAC1 y FRAXE) juntos se encontró un total de 29 haplotipos diferentes (26 en controles y 6 en cromosomas X frágil). Hubo tres haplotipos (21-19-16, 21-19-15 y 21-19-20) encontrados solamente en cromosomas X frágil y que diferían únicamente en el tipo FRAXE, que explicaban el 75%, 2,8% y 2,8% de los casos, respectivamente. Los individuos con el alelo más común provenían de un área que rodea Kuopio. Un control y un cromosoma X frágil compartían el mismo haplotipo 21-19-25. El haplotipo 21-18-17 se encontró en 4 cromosomas X frágil (11,1%) pero solo en un control (1,9%), los individuos no estaban relacionados, pero provenían de una región estrechamente agrupada en el sureste. Otro haplotipo (25-21-15) se encontró en dos cromosomas X frágil (5,6%) pero solo en un control (1,9%) y puede representar otra mutación reciente. (Zhong et al., 1996)

Los análisis respecto al polimorfismo de Alu-L fueron poco informativos a pesar de encontrarse tipo A y D entre la población finlandesa; en cuanto al análisis de interrupciones AGG se encontró que tanto en las muestras de Finlandia como en las de Nueva York, hubo una asociación con los alelos más largos y los alelos con repeticiones CGG largas y puras (≥ 15) tanto de FRAXAC1 como de DXS548. (Zhong et al., 1996)

Para el año 2001, investigadores reportaron la secuencia interna de FMR1 (CGG)_n y la distribución de los haplotipos microsatélites DXS548-FRAXAC1 en tres poblaciones habitantes de las regiones circumpolares de Europa. Las muestras de sangre se colectaron de población normal del Norte de Noruega y de pacientes con síndrome X frágil de toda Noruega, también se obtuvieron de la población Saami y Nenets habitantes de la Península de Kola y la costa marítima de Kara y Barents. Ningún individuo Saami o Nenets había sido diagnosticado con el síndrome antes. (Larsen et al., 2001)

El número de CGG se determinó en 264 cromosomas X Noruegos normales, 82 cromosomas X Nenets normales y 66 cromosomas X Saami normales. La distribución de las repeticiones (CGG)_n noruegas fue similar a la reportada en poblaciones caucásicas mientras que los Nenets tuvieron una mayor proporción ($P < 0.05$) de alelos 29 CGG comparado con la población noruega y Saami y la distribución alélica de los Saami mostró un pico más amplio alrededor de 30 CGG. (Larsen et al., 2001)

El alelo (CGG)_n predominante en 20/59 de noruegos masculinos tenía una secuencia interna de (CGG)10AGG(CGG)9AGG(CGG)9, es decir, 10A9A9. En 21 Nenets de sexo masculino, los alelos (CGG)_n más prevalentes tenían la secuencia interna de 10A9A9 (9/21) y 9A9A9 (7/21); mientras que en 12 Saami de sexo masculino, se encontraron ocho alelos (CGG)_n diferentes. (Larsen et al., 2001)

Se determinaron los haplotipos de los microsatélites FRAXAC1 y DXS548 en 119 noruegos, 25 Nenets y 17 Saami individuos normales, así como en 24 individuos noruegos de sexo masculino con síndrome X frágil. En la población noruega normal se encontraron 14 haplotipos diferentes, 4 en Nenets normales y 5 en Saami normales. El haplotipo predominante tanto en población noruega como en Saami fue el 7-3 con una frecuencia de 68% y 71%, respectivamente. Para el caso de la población Nenets, tanto los haplotipos 7-3 como 7-4 tuvieron una frecuencia de 44%. (Larsen et al., 2001)

En la población noruega se encontró un desequilibrio de ligamiento entre la mutación del síndrome X frágil y el haplotipo 6-4 ($\chi^2 = 17.4$, $P < 0.01$), y entre los alelos normales y el haplotipo 7-3 ($\chi^2 = 12.9$, $P < 0.01$), en concordancia con lo encontrado previamente en poblaciones caucásicas. (Larsen et al., 2001)

En el año 2006, investigadores deciden realizar un estudio similar en población asiática para analizar las repeticiones CGG en FMR1 y haplotipos flanqueantes en los tres grupos poblacionales principales de Singapur, evaluando a su vez su relación con la predisposición a la inestabilidad de las repeticiones CGG. (Zhou et al., 2006)

Para representar las poblaciones del Sur de China, indígenas del Archipiélago de Malasia y del sur India se tomaron muestras de ADN de sangre de cordón umbilical recogidas en el Hospital Universitario Nacional (Singapur) con información de etnicidad y sexo de 177, 178 y 99 personas, respectivamente. Adicionalmente se analizó un pequeño número de hombres X frágil no relacionados (9 de China y 3 de Malasia) del KK Women's and Children's Hospital cuyas expansiones de mutación completa se determinaron por análisis Southern blot. (Zhou et al., 2006)

Este estudio permitió determinar que entre la población China se encontraron 16 longitudes diferentes de repeticiones CGG y 30 patrones diferentes de repeticiones CGG en 29 haplotipos flanqueantes de FMR1 diferentes. En la población de Malasia se encontraron 18 longitudes diferentes de repeticiones CGG y 25 patrones diferentes en 25 haplotipos flanqueantes diferentes. A su vez, en la población de India se encontraron 15 longitudes diferentes de repeticiones CGG y 22 patrones diferentes en 30 haplotipos flanqueantes diferentes. La mayor heterocigosidad esperada se observó en los cromosomas de población india. Si se analizan en combinación, los 454 cromosomas del sureste de Asia contenían 25 longitudes de repetición CGG diferentes y 60 patrones de repetición CGG diferentes en 54 haplotipos flanqueantes, lo cual indica una heterocigosidad esperada y diversidad mayor a lo reportado previamente pero menor al compararse con poblaciones caucásicas o afroamericanas. (Zhou et al., 2006)

Los patrones de repetición CGG más comúnmente observados en estas tres poblaciones asiáticas fueron 9 + 9 + 9, 10 + 9 + 9 y 9 + 9 + 6 + 9; en la población india se encontró adicionalmente una alta frecuencia de patrón 13+9 ($\geq 5\%$). Respecto a la distribución de los haplotipos flanqueantes se encontró que los haplotipos principales (frecuencias mayores a 5%) en la población China fueron MH1 (194-4-G-146-T), MH2 (194-4-G-147-T), MH4 (193-3-A-151-C), MH7 (193-4-G-146-T), MH8 (196-4-G-146-T) y MH9 (193-4-G-147-T) correspondiendo juntos al 76,3% de los cromosomas X en esta población. En la población de Malasia los haplotipos principales fueron MH1, MH2, MH3 (193-3-A-150-C), MH4 y MH5 (194-3-A-151-C), lo cual correspondía al 71,9% de los cromosomas X en esta población. Por su parte en la población de India, los haplotipos principales fueron MH1, MH3, MH6 (194-3-A-150-C), MH10 (193-3-G-149-C), y MH11 (203-1-G-151-C), los cuales correspondían al 68,7% de los cromosomas X de esta población. Con estas observaciones es importante resaltar la ausencia de MH9 en población de India, ausencia de MH10 en población de China y ausencia de MH11 tanto en población China como en población de Malasia. (Zhou et al., 2006)

Del análisis de 12 cromosomas X frágil no relacionados se obtuvo que 4 de los 9 de población China eran portadores del MH1, 2 cromosomas portaban el MH2, 2 el MH3 y uno portaba el haplotipo menor 193-3-A-154-C, siendo este último observado una única vez en población normal de Malasia. De los 3 cromosomas X frágil de población de Malasia, cada uno portaba MH1, MH3 y el haplotipo menor 194-4-G-148-T. (Zhou et al., 2006)

Todas las poblaciones mostraron patrones similares de desequilibrio de ligamiento (LD), encontrándose LD fuerte entre la repetición CGG FMR1 con los marcadores FRAXAC1, FRAXAC2, ATL1 e IVS10, pero no con DXS548. No obstante, se encontraron diferencias específicas entre poblaciones. (Zhou et al., 2006)

En la población China los valores $|D'|$ entre los patrones de repetición CGG principales (9 + 9 + 9, 10 + 9 + 9, y 9 + 9 + 6 + 9) y los alelos 3 y 4 de FRAXAC1; alelos 146, 147 y 151 de FRAXAC2; ATL1 e IVS10, oscilaron entre 0,72 y 1 ($p < 0,005$). Además, el patrón 10 + 9 + 9 se asoció significativamente con los alelos 3 ($r^2 = 0,65$) y 4 ($r^2 = 0,61$) de FRAXAC1 y con ATL1 ($r^2 = 0,70$) e IVS10 ($r^2 = 0,71$) ($p < 0,005$). (Zhou et al., 2006)

En la población de Malasia, los alelos 146 y 150 de FRAXAC2 mostraron fuerte LD con la repetición CGG. Los valores $|D'|$ entre los 3 patrones de repetición CGG principales y los alelos 3 y 4 de FRAXAC1, alelos 146, 147, 150 y 151 de FRAXAC2, ATL1 e IVS10 oscilaron entre 0.56 y 1 ($p < 0,005$); encontrándose también un fuerte LD entre 10 + 9 + 9 y un alelo 194 único de DXS548 en esta población ($|D'| = 0,82$, $p < 0,005$). Al igual que para el caso de población China, únicamente el patrón 10 + 9 + 9 estuvo fuertemente asociado con los alelos 3 ($r^2 = 0,56$) y 4 ($r^2 = 0,56$) de FRAXAC1 y con ATL1 ($r^2 = 0,64$) e IVS10 ($r^2 = 0,59$) ($p < 0,005$). (Zhou et al., 2006)

En la población de la India un alelo 1 de FRAXAC1 estuvo en fuerte LD con la repetición CGG FMR1. Los valores $|D'|$ entre los 3 patrones de repetición CGG principales y los alelos 1, 3 y 4 de FRAXAC1, alelos 146, 149 y 150 de FRAXAC2, ATL1 e IVS10 oscilaron entre 0.73 y 1 ($p < 0,005$). En contraste a lo encontrado para población China y de Malasia, el patrón 10 + 9 + 9 no mostró ninguna asociación fuerte con ninguno de los marcadores flanqueantes en población de India. (Zhou et al., 2006)

Respecto a la relación entre los patrones de repetición CGG y los haplotipos flanqueantes, los autores encontraron que en población China el patrón 9 + 9 + 9 estuvo asociado positivamente ($p < 0,05$) con MH2 (194-4-G-147-T) y MH9 (193-4-G-147-T) mientras que el patrón 9 + 9 + 6 + 9 estuvo positivamente asociado con MH1 (194-4-G-146-T) ($p < 0,05$). En población de Malasia el patrón 9 + 9 + 9 estuvo positivamente asociado con MH1 y MH2, mientras que el patrón 10 + 9 + 9 estuvo positivamente asociado con MH4 (193-3-A-151-C) y MH3 (193-3-A-150-C), ($p < 0,05$). En población de India el patrón 9 + 9 + 9 estuvo positivamente asociado con MH10 (193-3-G-149-C), mientras que el patrón 10 + 9 + 9 estuvo positivamente asociado con MH3 ($p < 0,05$). (Zhou et al., 2006)

Adicionalmente se encontró que la pérdida de interrupciones AGG se encuentra asociada con mayor inestabilidad de la repetición CGG. (Zhou et al., 2006)

Para el año 2010, investigadores decidieron analizar la estructura de las repeticiones CGG, de los haplotipos flanqueantes y de las interrupciones AGG en hombres de una población africana sub-sahariana. (Peprah, Allen, Williams, Woodard, & Sherman, 2010)

Como población masculina no afectada se analizaron muestras de 350 individuos seleccionados aleatoriamente de un estudio poblacional previo realizado en Sunyani, Ghana. (Peprah et al., 2010; Williams et al., 2000) Como grupo de comparación se tomaron los datos reportados previamente por los autores en afroamericanos y caucásicos. (D. C. Crawford, Schwartz, et al., 2000; D. Crawford et al., 2002; D. Crawford, Zhang, Wilson, Warren, & Sherman, 2000; Peprah et al., 2010)

El estudio reveló que en cuanto a los marcadores flanqueantes (DXS548, FRAXAC1 y FRAXAC2) la población de Ghana tenía un menor o comparable número de alelos que lo encontrado en las otras poblaciones, con un nivel similar de heterocigosidad esperada al compararse con afroamericanos pero mayor al compararse con caucásicos. A excepción de FRAXAC2, la clase modal fue igual en todas las muestras: alelo 7 para DXS548, alelo 0 para DXS548-P y alelo 3 para FRAXAC1. A excepción de FRAXAC2 y DXS548-P, las

distribuciones alélicas de los marcadores flanqueantes no fueron estadísticamente diferentes entre las muestras de Ghana y afroamericanas (DXS548: $\chi^2 = 2.16$, $P = 0.539$, $df = 3$; FRAXAC1: $\chi^2 = 2.14$, $P = 0.543$, $df = 3$), mientras que sí diferían entre las muestras de Ghana y las caucásicas (DXS548: $\chi^2 = 96.7$, $P < 0.0001$, $df = 3$; FRAXAC1: $\chi^2 = 56.1$, $P < 0.0001$, $df = 3$). (Peprah et al., 2010)

Se identificaron en total 72 haplotipos diferentes (DXS548, FRAXAC1 y FRAXAC2) en población de Ghana comparado con 71 y 202 haplotipos encontrados en caucásicos y afroamericanos respectivamente. (D. C. Crawford, Schwartz, et al., 2000; D. Crawford et al., 2002, 2000; Peprah et al., 2010) Los haplotipos más frecuentes en población de Ghana (7-3-5 (0.11), 7-3-3+ (0.09) y 7-4-4 (0.06)) no mostraron un comportamiento similar en las otras poblaciones, encontrándose que en éstas el haplotipo más frecuente era 7-3-4+ (0.13 en afroamericanos y 0.39 en caucásicos), este último correspondiendo a menos del 2% de la población de Ghana. (Peprah et al., 2010)

Adicionalmente se evaluó el polimorfismo de repetición CGG FMR1 (FRAXA) en las muestras de población de Ghana con respecto a la longitud de la repetición y el patrón de interrupciones AGG, comparándolo a su vez con las muestras afroamericanas y de caucásicos. La heterocigosidad esperada para la repetición FRAXA fue similar en las tres poblaciones, aunque el número observado de alelos difirió: 23 para población de Ghana, 37 para afroamericanos y 42 para caucásicos. La distribución alélica en este caso no difirió significativamente entre las muestras de Ghana y las afroamericanas ($\chi^2 = 10.3$, $P = 0.172$, $df = 7$) pero ambas difirieron de lo encontrado para caucásicos (Ghaneses vs. Caucásicos: $\chi^2 = 56.5$, $P < 0.0001$, $df = 7$; Afroamericanos vs. Caucásicos $\chi^2 = 320$, $P < 0.0001$). (Peprah et al., 2010)

Los alelos más comunes en las muestras de Ghana fueron los tamaños de repetición de 29 y 30, similar a lo encontrado en las otras poblaciones. (Peprah et al., 2010)

Los patrones de interrupción AGG para la población de Ghana revelaron 41 estructuras de interrupción, entre estas la estructura 9+9+9 fue el patrón encontrado más frecuentemente seguido del patrón 10+9+9. Evaluado de otra manera, el estudio reveló que la mayoría de las estructuras contenían dos interrupciones, similar a lo encontrado en las otras poblaciones. (Peprah et al., 2010)

También se examinaron la estructura de las repeticiones y las asociaciones con los haplotipos entre los alelos de tamaño intermedio (35-49 repeticiones) encontrándose que la mayoría de estos alelos carecían de interrupciones AGG (44%). (Peprah et al., 2010)

Al momento de comparar los haplotipos encontrados en la muestra de población de Ghana con aquellos hallados en cromosomas con o sin el síndrome X frágil de población caucásica y afroamericana, los autores descubrieron que las muestras de población de Ghana compartían siete haplotipos que habían sido asociados con cromosomas X frágil en caucásicos y cinco que habían sido asociados con cromosomas X frágil en afroamericanos. (Peprah et al., 2010)

Tres haplotipos de cromosomas X frágil (6-4-5, 7-3-4+ y 7-4-5) se compartían entre las tres poblaciones, estableciéndose una frecuencia combinada de 26% en cromosomas X frágil caucásicos y de 22% en cromosomas X frágil afroamericanos. Entre los cromosomas sin la mutación, las frecuencias combinadas de los haplotipos fueron 44% en el caso de caucásicos y 7% en el caso de afroamericanos, lo anterior comparado con una frecuencia de solo 2% encontrada en la población de Ghana. (Peprah et al., 2010)

Por último, se determinó que los haplotipos más comunes en las muestras de población de Ghana entre este grupo compartido con los cromosomas X frágil fueron 7-3-3+ y 6-4-4. Estos últimos se encontraron en baja frecuencia en la población afroamericana no afectada (<1%) y en las muestras de caucásicos no afectados ($\leq 3\%$) y en una ligera mayor frecuencia en población de Ghana no afectada (9% y 4% respectivamente). (Peprah et al., 2010)

En el año 2011 se planteó un estudio similar cuyo objetivo era identificar los haplotipos específicos entre once pacientes con síndrome X frágil procedentes de Letonia y 122 pacientes control con retardo mental y un número normal de repeticiones CGG. (Daneberga, Pronina, Lace, & Lugovska, 2011)

Los alelos de repetición CGG fueron analizados en el grupo control respecto a los alelos ATL1 (alelo A y alelo G), este grupo a su vez se dividió en dos subgrupos de acuerdo al número de alelos de repetición CGG: alelos de tamaño normal y alelos de zona gris; en cuanto a las frecuencias en cada grupo, se encontró una diferencia significativa entre la presencia del alelo G en los alelos de tamaño normal y los alelos de zona gris ($p = 0.0271$). (Daneberga et al., 2011)

Se encontró una asociación estadísticamente significativa entre los alelos CGG individuales y el SNP ATL1, para el alelo de 29 repeticiones CGG y el alelo G ($p = 0.001$); para el alelo de 30 repeticiones CGG y el alelo A ($p < 0.0001$) y para el alelo de 31 repeticiones y el alelo A ($p = 0.0013$). (Daneberga et al., 2011)

Todos los cromosomas del grupo X frágil se asociaron exclusivamente con el alelo G de ATL1, siendo esta asociación estadísticamente significativa ($p = 0.0008$). (Daneberga et al., 2011)

Entre los 26 alelos de zona gris se identificó que 12 cromosomas exhibían 3 interrupciones de AGG entre las repeticiones CGG, 12 cromosomas tenían 2 AGG, un cromosoma tenía una AGG, y un cromosoma era un tracto puro de CGG. En todos los casos, la pérdida de AGG se detectó en el extremo 3' de la secuencia. (Daneberga et al., 2011)

Respecto a los marcadores microsatélites se encontraron 7 diferentes alelos en el locus DXS548, 4 alelos en el locus FRAXAC1 y 9 alelos en el locus FRAXAC2; estos marcadores se combinaron para formar haplotipos encontrándose un total de 27 haplotipos diferentes: 26 en el grupo control y 3 en el grupo X frágil, de los cuales 1 fue único. El haplotipo más común entre los cromosomas del grupo control fue 7-4-A-5+ (RF = 0.327; $p = 0.0336$) mientras que entre los cromosomas del grupo X frágil, el haplotipo prevalente fue 2-2-G-4 (RF = 0.818; $p < 0.0001$); mediante análisis estadístico se determinó que la variación molecular entre ambos

grupos fue de 27,04% y al interior de cada grupo de 72,96%. El índice de fijación (F_{st}) calculado con base en las frecuencias de los haplotipos entre el grupo control y el grupo X frágil fue de 0,27042. (Daneberga et al., 2011)

Al realizar un análisis de haplotipos basados en el patrón del tracto CGG en los alelos de zona gris se encontró que los haplotipos más comunes eran 2-2-G-4 ($RF = 0.308$) y 7-4-A-5+ ($RF = 0.231$), todos los alelos del primer haplotipo tenían una secuencia ininterrumpida larga (≥ 18 repeticiones CGG) en el extremo 3' ($p = 0.0022$), con 6 alelos de 18 con otros haplotipos detectados contando con la misma característica. (Daneberga et al., 2011)

Como conclusión el estudio revela una prevalencia estadísticamente significativa del alelo G de ATL1 entre los alelos de zona gris y los alelos con mutación completa como indicador de inestabilidad, adicionalmente en esta población se encontró que el haplotipo 2-2-G-4 era un marcador de inestabilidad del tracto CGG y que los alelos de zona gris con un tracto CGG largo ininterrumpido en el extremo 3' asociados con el haplotipo 2-2-G-4 tenían una posibilidad mayor de incrementar el número de repeticiones CGG conllevando a premutación y mutación en las generaciones siguientes. (Daneberga et al., 2011)

Debido a la heterogeneidad exhibida en la población de Letonia y la falta de estudios en esta región geográfica, los investigadores concluyen que un efecto fundador no puede ser una explicación a los hallazgos presentados y por tanto proponen que es necesario realizar un estudio más grande de los haplotipos ligados a FMR1 en las regiones de norte y este de Europa para establecer una vía mutacional específica que explique la inestabilidad de los alelos CGG. (Daneberga et al., 2011)

En el año 2015, se realizó un estudio de características similares en 1113 individuos de China etnia Han no afectados y los patrones de repetición CGG en los individuos de sexo masculino se determinaron con secuenciación directa (se excluyeron 32 hombres debido a falta de muestra de ADN). Adicionalmente, se analizaron los haplotipos flanqueantes (DXS548-FRAXAC1-FRAXAC2) en 566 individuos de sexo masculino no afectados y 28 individuos de sexo masculino con síndrome X frágil. (Huang et al., 2015)

En este estudio se identificaron 6 alelos diferentes tanto para DXS548 como para FRAXAC1 y 8 para FRAXAC2; los alelos más frecuentes en esta población fueron el alelo 7 para DXS548 (igual a lo reportado en poblaciones caucásicas y afroamericanas) y el alelo 4 para FRAXAC1, mientras que los alelos más abundantes para FRAXAC2 fueron los alelos 7+, 4 y 6. Adicionalmente se encontraron los alelos 2 para DXS548 y los alelos 1, 2, 5 y 6 para FRAXAC1, los cuales no habían sido reportados previamente en población China o Taiwanesa. Los autores concluyen que en general la población China tiene un menor número de alelos y una menor heterocigosidad esperada que los individuos caucásicos o afroamericanos, con una distribución alélica significativamente diferente entre las tres poblaciones. (Huang et al., 2015)

El análisis realizado identificó 33 haplotipos distintos (DXS548-FRAXAC1-FRAXAC2), siendo los más frecuentes los haplotipos 7-4-7+ (0.440), 7-3-4 (0.173), y 7-4-6 (0.141). (Huang et al., 2015)

Respecto a las longitudes y patrones de las repeticiones CGG en población China no afectada se encontraron 33 longitudes de repetición CGG diferentes en 1692 cromosomas, encontrándose valores similares en población de Indonesia y Japón. El alelo más frecuente fue (CGG)₂₉ seguido del alelo (CGG)₃₀, con un menor número de longitudes de repeticiones CGG y menor heterocigosidad esperada que lo reportado para caucásicos y afroamericanos, lo anterior en concordancia con distribuciones de las longitudes de repetición CGG estadísticamente diferentes entre la población China y las otras dos poblaciones. (Chinos vs. Caucásicos: $\chi^2 = 230.2$, $P < 0.0001$, grados de libertad [df] = 4; Chinos vs. Afroamericanos: $\chi^2 = 144.6$, $P < 0.0001$, df = 4). (Huang et al., 2015)

El alelo (CGG)₃₀ mostró una frecuencia casi idéntica o ligeramente superior que el alelo (CGG)₂₉ en afroamericanos y caucásicos, respectivamente. En la población China no afectada solo se identificó una portadora femenina de la premutación con 89 repeticiones CGG. (Huang et al., 2015)

Se encontraron 62 patrones de repetición CGG diferentes en 32 haplotipos distintos, los más comunes en población China fueron 9 + 9 + 9 (0.489), 10 + 9 + 9 (0.246), y 9 + 9 + 6 + 9 (0.074) (Huang et al., 2015), mientras que en población caucásica el más frecuente es 10 + 9 + 9 (0.275) y la frecuencia del patrón 10 + 9 + 9 (0.211) y del patrón 9 + 9 + 9 (0.211) son casi iguales en afroamericanos; con el patrón 9 + 9 + 6 + 9 ausente tanto en población caucásica como en afroamericanos (D. C. Crawford, Zhang, et al., 2000; Eichler et al., 1996)

La mayoría de los patrones de repetición CGG en población China contaban con dos interrupciones AGG (0.788), lo cual está en concordancia con lo reportado en caucásicos, afroamericanos y ghaneses; mientras que la frecuencia de alelos de repetición puros (0.015) en población China es la más baja comparada con las otras poblaciones. (D. C. Crawford, Zhang, et al., 2000; Eichler et al., 1996; Peprah et al., 2010)-Todos los alelos intermedios (45-54 repeticiones) en esta población contenían dos o más interrupciones AGG. (Huang et al., 2015)

Adicionalmente este estudio muestra que los alelos FRAXAC1 tienen una alta asociación con las longitudes de repetición CGG. En población China, casi el 73,6% del alelo 4 de FRAXAC1 estuvo asociado con (CGG)₂₉, mientras que el 77,3% del alelo 3 de FRAXAC1 estuvo asociado con (CGG)₃₀. Se confirmó además la presencia de LD significativo entre los alelos FRAXAC1 y los patrones de repetición CGG en esta población. El alelo 4 de FRAXAC1 se asocia fuertemente con el patrón 9 + n ($\chi^2 = 367.0$, $P < 0.0001$, df = 1), mientras que el alelo 3 de FRAXAC1 se asocia fuertemente con el patrón 10 + n ($\chi^2 = 163.0$, $P < 0.0001$, df = 1). (Huang et al., 2015)

Así mismo se determinó la relación entre los patrones de repetición CGG y los haplotipos flanqueantes, encontrándose que los patrones 9 + 9 + 9 y 9 + 9 + 6 + 9 se asocian significativamente con el haplotipo 7-4-7 + y que el patrón 10 + 9 + 9 se asociaba significativamente con el haplotipo 7-3-4. Los patrones 9 + 9 + 16 y 9 + 16 + 9 con longitud igual al patrón 9 + 9 + 6 + 9, se encontraron únicamente en el haplotipo 7-4-7 +. (Huang et al., 2015)

En la población con síndrome X frágil estudiada (28 individuos), para FRAXAC1 y FRAXAC2 se encontró que tenían menos alelos, pero mayor heterocigosidad esperada al compararse con la población no afectada. Para el caso de DXS548 la población con síndrome X frágil mostró menos alelos y menor heterocigosidad esperada que la población no afectada. Los alelos más abundantes tanto para DXS548 (alelo 7) como para FRAXAC2 (alelo 7+) fueron idénticos entre las poblaciones X frágil y sanas. Con respecto a FRAXAC1 el alelo más frecuente en población X frágil fue el alelo 3 y en población no afectada el alelo 4.(Huang et al., 2015)

Adicionalmente el estudio identificó siete haplotipos diferentes en la población X frágil, siendo los más frecuentes (7-4-7+ y 7-3-4) iguales que en población no afectada. Cabe resaltar que pese a que el haplotipo 7-3-5+ no tuvo representatividad significativa en el grupo no afectado, corresponde al 17,9% de todos los cromosomas X frágil, difiriendo estadísticamente en su distribución entre ambos grupos ($\chi^2 = 24.0$, $P < 0.0001$, $df = 1$); lo cual puede indicar una asociación significativa entre el haplotipo y la presencia de mutación completa.(Huang et al., 2015)

Como se ve, la mayoría de estudios de haplotipos para exploración de posibles efectos fundadores de síndrome X frágil, se han realizado con conjuntos de loci correspondientes a microsatélites, sin embargo, recientemente la secuenciación de nueva generación, particularmente secuenciación exómica completa, ha permitido la identificación mutaciones fundadoras de enfermedades raras (Rousseau-Nepton et al., 2015) y el análisis de la estructura genética de poblaciones aisladas. (Jeroncic et al., 2016) con la ventaja de que al tratarse de un análisis paralelo masivo a un costo muy reducido (Mardis, 2011), permite secuenciar millones a billones de lecturas en paralelo en unos pocos días u horas, de modo que arroja una gran cantidad de información, que puede ser comparada contra referencias mundiales de variación genética humana como la del proyecto 1000 genomas, para realizar análisis de la estructura genética y evolución molecular de una población. (Zheng-Bradley & Flicek, 2017)

Tecnologías de secuenciación

El descubrimiento de la estructura de doble hélice compuesta de las cuatro bases del ADN por Watson y colaboradores en 1953 (WATSON & CRICK, 1953) ha llevado a la decodificación de las secuencias genómicas y a conocer la composición del ADN de los organismos. La secuenciación del ADN es el descubrimiento que usa la composición del ADN para entender y describir el código de toda la vida biológica sobre la tierra así como también a entender y tratar las enfermedades genéticas (Frio, 2015).

La aparición de las tecnologías de secuenciación ha jugado un papel importante en el análisis de las secuencias genómicas de los organismos. Un secuenciador de ADN produce archivos que contienen secuencias de ADN (Shendure & Ji, 2008). Estas secuencias son cadenas llamadas lecturas sobre un alfabeto formado por 5 letras (A, T, C, N). El símbolo N es

representado para representar una ambigüedad. Las primeras tecnologías de secuenciación fueron desarrolladas en 1977 por Sanger y colaboradores (Sanger, Nicklen, & Coulson, 1977) de la universidad de Cambridge lo que les mereció el premio Nobel de química en 1980, y Maxam y colaboradores (Maxam & Gilbert, 1977) de la universidad de Harvard. Sus descubrimientos abrieron la puerta al estudio del código genético de los seres vivos y trajo inspiración a los investigadores para el desarrollo de tecnologías de secuenciación eficientes y más rápidas. La secuenciación de Sanger fue comercializada y automatizada como Tecnología de Secuenciación Sanger y debido a su alta eficiencia y baja radiactividad fue por un buen tiempo la técnica de secuenciación más aplicada (Pareek, Smoczynski, & Tretyn, 2011)

Las tecnologías de secuenciación de Sanger y Maxam fueron las tecnologías de secuenciación más comúnmente usadas por los biólogos hasta la aparición de la nueva era de las tecnologías de secuenciación que abrió nuevas perspectivas para la exploración y análisis de genomas. La primera de estas tecnologías de la nueva era fue la 454 de Roche en 2005 (Qiang-long, Shi, Peng, & Fei-shi, 2014) y fueron comercializadas como tecnologías capaces de producir secuencias con muy alto rendimiento y a un costo mucho más bajo que las primeras tecnologías de secuenciación. Estas nuevas tecnologías de secuenciación son generalmente conocidas como Tecnologías de Secuenciación de Nueva Generación (NGS) o Tecnologías de Secuenciación de Alto Rendimiento. Las tecnologías NGS producen un análisis paralelo masivo con alto rendimiento de múltiples muestras a un costo muy reducido (Mardis, 2011). Las tecnologías NGS pueden secuenciar millones a billones de lecturas en paralelo en una sola corrida y el tiempo requerido para generar lecturas de tamaños de gigabases es solamente unos pocos días u horas, haciendo que sea mejor que la secuenciación de primera generación. El genoma humano, por ejemplo, consiste en 3 billones de pares de bases y está hecho de macromoléculas de ADN de longitudes que varían desde 33 hasta aproximadamente 247 millones de pares de bases, distribuidos en los 23 cromosomas localizados en los núcleos celulares, la secuenciación del genoma humano usando Sanger tomó aproximadamente 15 años, requirió la cooperación de muchos laboratorios alrededor del mundo y costó aproximadamente 100 millones de dólares, mientras que la secuenciación por NGS usando un secuenciador genómico 454 FLX tomó dos meses y costó aproximadamente una centésima parte (Wheeler et al., 2008). Desafortunadamente la NGS es incapaz de leer la secuencia completa del ADN, ella está limitada a secuenciar fragmentos pequeños de ADN y generar millones de lecturas. Este límite representa un punto negativo especialmente para proyectos de ensamblaje de genomas porque requiere grandes recursos informáticos. La NGS continúa mejorando y el número de secuenciadores aumentó en los últimos años. Sin embargo, la literatura divide la NGS en dos tipos (Shendure & Ji, 2008)(Thudi, Li, Jackson, May, & Varshney, 2012). Se pueden distinguir tecnologías de Secuenciación de Segunda Generación las cuales se refieren a las tecnologías más nuevas, desarrolladas después de la primera generación (Metzker, 2010)(Schatz, Delcher, & Salzberg, 2010), estas se caracterizan por la necesidad de preparar bancos de secuenciación amplificados antes de empezar la secuenciación de clones de ADN (Kulski, 2016) y hay

Tecnologías de secuenciación de Tercera Generación, que son las tecnologías de secuenciación que aparecieron recientemente (Pareek et al., 2011), en contraste con las de segunda generación, estas tecnologías se clasifican como Tecnologías de Secuenciación de Molécula Única (Kchouk, Gibrat, & Elloumi, 2017) porque ellas pueden secuenciar una molécula única sin necesidad de crear librerías de amplificación y son capaces de generar lecturas más largas a costos muchos más bajos y tiempos más cortos.

Secuenciación exómica y sus aplicaciones en genética poblacional

Sin lugar a dudas, la secuenciación exómica (SE) es ahora la herramienta más comúnmente usada para el descubrimiento de genes causantes de enfermedades mendelianas.(Gilissen, Hoischen, Brunner, & Veltman, 2011) (Gilissen, Hoischen, Brunner, & Veltman, 2012) Cuando se considera la SE para diagnóstico clínico, se encuentra que el rendimiento diagnóstico de esta prueba es 25 a 28%. De hecho, una tasa mucho más alta de aproximadamente 50% se observó en un estudio piloto en el cual se aplicaron criterios más rigurosos para seleccionar los pacientes adecuados. (Yang et al., 2014) Desde una perspectiva económica, se ha mostrado que la SE tiene el potencial de proveer una buena relación costo-beneficio para pacientes que permanecen sin diagnóstico después de las aproximaciones tradicionales.(Shashi et al., 2014) La SE ha mostrado éxito más frecuentemente en los casos en los cuales no se tiene conocimiento previo de la condición genética. En un reporte, la tasa más alta de diagnóstico positivo estuvo en el grupo de pacientes con trastornos neurológicos inespecíficos, indicando que la SE es el último recurso para diagnóstico cuando no haya signos específicos apunten a herramientas diagnósticas específicas.(Xue, Ankala, Wilcox, & Hegde, 2015)

Sin importar qué impulsa a un clínico a solicitar una SE, es crítico que cualquier información y hallazgos clínicos de otras evaluaciones estén disponibles para que el genetista molecular clínico pueda interpretar el vasto número de variantes de secuencia generadas a través de SE. Aunque la SE puede eludir la aproximación basada en genes candidatos, la interpretación de variantes depende en gran medida de tener los fenotipos bien caracterizados. El componente más crucial de cualquier algoritmo diagnóstico es la evaluación clínica. La SE no es un sustituto para la realización de una historia clínica completa y una fenotipificación adecuada. Algunas evaluaciones adicionales pueden ser requeridas después de la SE para hacer un diagnóstico final. Las habilidades diagnósticas de los médicos especialistas serán cambiadas de un diagnóstico diferencial pre-SE a uno post-SE. Esto significa que la SE puede ser usada como una herramienta de tamizaje y ofrecer un abanico de posibilidades, permitiendo a los clínicos realizar evaluaciones más enfocadas y eficientes. Un esfuerzo colaborativo entre médicos y laboratorios diagnósticos será cada vez más importante. (Xue et al., 2015)

A pesar de que los kits de SE disponibles comercialmente en el mercado fueron diseñados para cubrir el exoma entero, estos no ofrecen cubrimiento completo de todos los exones de los genes, incluyendo el subgrupo de genes asociados a enfermedades conocidas. Para que la transición de la SE desde una prueba de tamizaje a una prueba independiente de primer nivel, esta debe primero ofrecer un cubrimiento de todos los exones de al menos los 4.600 genes asociados a enfermedades conocidas y mantener el cubrimiento de los restantes 18.000 genes,

y tener la habilidad de detectar variantes a través del espectro completo de mutaciones de cada uno de los genes. (Xue et al., 2015) En el momento esta técnica es predominantemente usada en investigación, siendo uno de sus usos la detección de variantes de nucleótido único (SNPs) que pueden estar asociadas a fenotipos específicos, (Gilissen et al., 2011) pudiendo ser causantes o estar en desequilibrio de ligamiento con las realmente causantes.

Sin embargo, además de los usos mencionados, las tecnologías de secuenciación de nueva generación han abierto la puerta a una nueva dimensión de estudios poblacionales, permitiendo miradas claras al pasado de las poblaciones y la corroboración de evidencia proveniente de la paleontología, la arqueología y la historia con una precisión sin precedentes, lo cual resulta de interés tanto pensado a gran escala, (Abecasis et al., 2012) como a una escala local, particularmente en el caso de poblaciones aisladas. (Jeroncic et al., 2016) Como ya se mencionó previamente, utilizando secuenciación de nueva generación, particularmente secuenciación exómica completa, se han identificado mutaciones fundadoras de enfermedades raras (Rousseau-Nepton et al., 2015) y se han realizado análisis de la estructura genética de poblaciones aisladas. (Jeroncic et al., 2016)

Las poblaciones aisladas retienen su aislamiento genético y singularidades debido a varios factores posibles como barreras geográficas, étnicas y lingüísticas, y se ha estimado que, en conjunto, solo en Europa, este tipo de poblaciones comprenden aproximadamente 11,5 millones de individuos. (Jeroncic et al., 2016)

Cada individuo inmigrante lleva consigo una contribución genética particular. Diferentes elementos como el tamaño de las poblaciones fundadoras y su aislamiento por geografía u otras razones, pueden llevar a una alta prevalencia de condiciones genéticas específicas en diversos grupos étnicos. Aunque una ese tipo de variaciones en la frecuencia de enfermedades ya se había observado en el mundo, y los principios de la genética poblacional han estado allí desde hace tiempo, ha sido la aplicación de la tecnología de secuenciación de nueva generación moderna la que ha permitido la exploración precisa de la estructura genética de las poblaciones y la que tiene el potencial de conducir a mejoras de los pronósticos de los pacientes, sus familias y sus comunidades. (Evans, 2015)

Existen múltiples ejemplos de poblaciones aisladas con conglomerados genéticos de enfermedades genéticas raras en el mundo:

- En los inuits son comunes dos enfermedades autosómicas recesivas: deficiencia congénita de isomaltasa-sucrasa y glucogenosis tipo IIIa, en ambas enfermedades hay expresividad variable y el diagnóstico temprano con intervención dietaria oportuna ayuda a disminuir las consecuencias clínicas. Macardier y cols., (Macardier et al., 2015) documentaron el hallazgo de una mutación común en el gen SI responsable de deficiencia congénita de isomaltasa-sucrasa, y Rousseau-Nepton y colaboradores (Rousseau-Nepton et al., 2015) describieron el uso de secuenciación exómica completa en niños con glucogenosis tipo IIIa para identificar una mutación fundadora relativamente reciente en el gen AGL, específica de inuits habitantes de Nunavik. Los análisis de haplotipos pueden proveer información sobre la edad relativa de mutaciones fundadoras porque las regiones de homocigosidad alrededor del gen de

interés se van acortando con el tiempo como resultado de la recombinación meiótica. (Evans, 2015)

- En población chippewayan en Manitoba y Cree en el norte de Quebec existe evidencia molecular e histórica de que una mutación fundadora de Leucoencefalopatía infantil recesiva encontrada, puede haber venido de comerciantes de pieles británicos de la empresa Hudson Bay Company en los 1770s. (Fogli et al., 2002)

Globalmente, los efectos fundadores han llevado a patrones similares de alta prevalencia de enfermedades raras en poblaciones específicas, sin embargo, los alelos fundadores pueden también contribuir al riesgo para enfermedades comunes:

- Oji-Cree en el noroeste de Ontario es una población con altas tasas de diabetes mellitus 2 en la que se ha encontrado que cada copia de una mutación única en HNF1A (hepatocyte nuclear factor 1 homeobox A) disminuye 7 años la edad media de inicio de esta enfermedad. (Triggs-Raine et al., 2002)
- En familias islandesas con cáncer de mama es común el alelo 999del5 de BRCA2.(Thorlacius et al., 1997)
- En Newfoundland y Labrador se encontró una mutación fundadora en MHS2 (MutS protein homolog 2) (Stuckless et al., 2007)

La identificación de la variación en la distribución de enfermedades ha permitido el planteamiento de estrategias de salud preventivas como el uso del background étnico para identificar personas a las que se debe ofrecer tamizaje para condiciones como drepanocitosis, talasemias, enfermedad de Tay Sachs, (Wattendorf & Hadley, 2005) (Evans, 2015) deficiencia de carnitina palmitoiltransferasa I A,(Prasad et al., 2001) entre otras.

La atención a la ancestría ayudará a evaluar la probabilidad de enfermedad genética y se podrá reducir la carga de enfermedad y brindar a los pacientes y sus familias un cuidado personalizado. (Evans, 2015)

Metodología

En un estudio previo realizado por nuestro grupo, denominado Genética médica poblacional en un corregimiento colombiano con elevada prevalencia de síndrome de frágil X, se realizó un censo de la población del corregimiento de Ricaurte, ubicado en Bolívar-Valle, encontrando que había 1186 habitantes, 543 hombres y 634 mujeres. Se recolectaron muestras de sangre en papel filtro de 926 personas (424 hombres y 502 mujeres). En ese estudio se realizaron pruebas moleculares a todas las muestras recolectadas para determinar el genotipo *FMRI*, encontrando que 64 de los 926 individuos tenían expansión en FMR1, más específicamente, entre los 424 hombres 22 tenían la mutación completa, 3 tenían la premutación y el resto tenían un alelo normal, y entre las 502 mujeres, 11 tenían la mutación completa, 28 tenían la premutación y el resto tenían alelos normales. Con esta información se pudo demostrar que Ricaurte constituye un conglomerado genético de síndrome X frágil, con las prevalencias más altas de mutación completa y premutación reportadas en la literatura. Adicionalmente se logró construir, con ayuda del software Progeny®, un heredrograma de 485 individuos.

En dicho estudio, además, dos genetistas clínicos (Dr Julián Ramírez-Cheyne y Dr Wilmar Saldarriaga) y un neurólogo (Dra Randi Hagermann) realizaron a todos los pacientes una evaluación fenotípica mediante anamnesis y examen físico, en la que se indagó, entre otras cosas, por la presencia de convulsiones, síndrome de temblor-ataxia y síndrome de falla ovárica prematura. En el momento de recolección de muestras y datos de ese estudio se solicitó autorización para su uso en estudios posteriores.

Muestra

Se incluyeron todas las muestras de ADN de todos los pacientes expandidos (con mutación completa o premutación) de las cuales se disponía, y se incluyó la muestra de una mujer no expandida ni perteneciente a las familias afectadas con X frágil del conglomerado. De todas las muestras se tenía la autorización para uso. Se excluyeron las muestras que no pasaron el control de calidad para la realización del exoma. Finalmente, la muestra analizada estuvo conformada por ADN de 55 individuos del conglomerado: 49 mutados o premutados de las familias afectadas oriundas de Ricaurte (familia 1, familia 2, familia 999), 4 mutados o premutados de una familia afectada (familia 4) oriunda de Robledo (población ubicada a 8 km de Ricaurte), 1 mujer premutada (familia 5) que habitaba Ricaurte pero que afirmaba ser oriunda de La Herradura (población ubicada a 4,8 km de Ricaurte) y 1 mujer normal (familia 6-control) oriunda de Ricaurte que no pertenecía a ninguna de las familias afectadas. El genotipo para FMR1 y la presencia o ausencia de convulsiones, síndrome de temblor-ataxia y síndrome de falla ovárica prematura en cada individuo de la muestra se observa en la tabla 1.

Tabla 1: Genotipo para FMR1 y presencia o ausencia de convulsiones, síndrome de temblor-ataxia y síndrome de falla ovárica prematura en cada individuo de la muestra.

Código del individuo	Código de la familia	Sexo	Edad	Genotipo FMR1 por Southern Blot	Convulsiones	FXTAS	FXPOI
166	1	M	33	Mutación completa	SI	NO	NA
167	1	F	72	Premutación	NO	NO	NO
169	1	F	28	Mutación completa	NO	NO	NA
171	1	F	68	Premutación	NO	NO	NO
172	1	M	49	Mutación completa	SI	NO	NA
174	1	F	45	Premutación	NO	NO	SI

175	1	M	24	Mutación completa mosaico	NO	NO	NA
176	1	M	26	Mutación completa	NO	NO	NA
177	1	F	36	Mutación completa	NO	NO	NA
178	1	M	46	Mutación completa mosaico	NO	NO	NA
179	1	M	54	Mutación completa	NO	NO	NA
180	1	F	80	Premutación	NO	NA	NO
181	1	M	57	Mutación completa mosaico	NO	NO	NA
182	1	F	33	Mutación completa	NO	NO	NA
187	1	F	53	Premutación	SI	SI	SI
195	1	F	58	Premutación	NO	SI	NO
196	1	M	29	Mutación completa	SI	NO	NA
197	1	F	60	Premutación	SI	SI	NO
199	1	M	69	Mutación completa	NO	NO	NA
200	1	M	64	Mutación completa	NO	NO	NA
204	1	F	51	Premutación	NO	NO	NO
205	1	M	31	Mutación completa	SI	NO	NA
206	1	F	57	Premutación	NO	NO	NO
250	1	F	27	Mutación completa	NO	NO	NA

251	1	M	6	Mutación completa	NO	NO	NA
252	1	M	8	Mutación completa	NO	NO	NA
253	1	F	35	Mutación completa	NO	NO	NA
259	1	F	26	Premutación	NO	NA	NA
260	1	M	70	Premutación	NO	NO	NA
282	1	F		Premutación	NO	NO	NO
299	1	F	62	Premutación	NO	NO	NO
301	1	F	38	Premutación	NO	NO	NA
165	1	M	85	Premutación	NO	SI	NA
168	1	M	30	Mutación completa mosaico	SI	NO	NA
198	1	M	61	Mutación completa	NO	NO	NA
207	1	F	50	Mutación completa	NO	NO	NO
255	1	F	43	Premutación	NO	NO	SI
256	1	M	14	Premutación	NO	NO	NA
258	1	F	27	Premutación	NO	NO	NA
158	2	F	69	Mutación completa	NO	NO	NO
159	2	M	45	Mutación completa mosaico	NO	NO	NA

160	2	F	67	Mutación completa	NO	NO	NO
161	2	M	7	Mutación completa mosaico	NO	NO	NA
162	2	F	37	Mutación completa	NO	NO	NA
163	2	F	66	Mutación completa	NO	NO	NO
186	2	M	71	Mutación completa mosaico	NO	NO	NA
188	4	M	26	Mutación completa mosaico	NO	NO	NA
189	4	F	44	Mutación completa	NO	NO	NO
190	4	F	55	Mutación completa	NO	NO	NO
191	4	F	84	Mutación completa	NO	NA	NO
173	5	F	71	Premutación	NO	NO	SI
203	6	F	59	Normal	NA	NA	NA
170	999	F	66	Mutación completa	NO	NO	NO
193	999	M	69	Mutación completa	NO	NO	NA
192	999	M	63	Mutación completa	NO	NO	NA

Se realizó secuenciación exómica (preparación de librerías, PCR y secuenciación del exoma) utilizando Agilent SureSelect Human All Exon V6 (58 M) (<https://en.novogene.com/next-generation-sequencing-services/human-genome/whole-exome-sequencing-service/>), obteniendo un archivo sin procesar fastq (Tabla 2) con secuencias de 150 pb del exoma para cada uno de los 55 individuos. Este procedimiento fue llevado a cabo en Novogene Corporation®.

Tabla 2: Datos crudos del exoma en secuencias de 150pb para los 55 individuos.

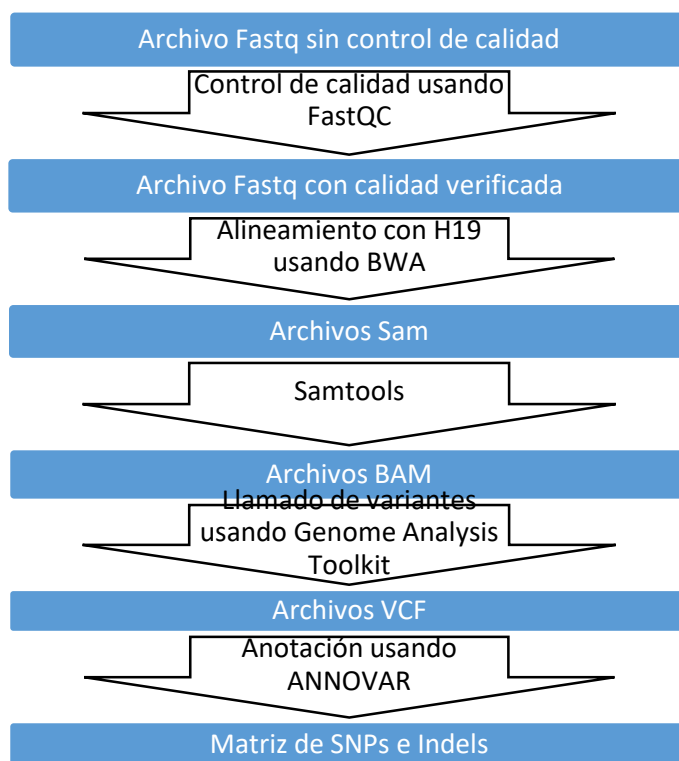
Identificación del individuo	Número de lecturas	Identificación del individuo	Número de lecturas
158	20979219	190	23857413
159	22100894	191	30815010
160	21669788	192	25157074
161	21488574	193	21990629
162	19531617	195	21451641
163	20977892	196	21205375
165	18819391	197	23128430
166	27486437	198	22555313
167	20532885	199	23917104
168	23119313	200	21870937
169	21883474	203	23954927
170	22332102	204	19906421
171	18903165	205	20732879
172	22884754	206	20427516
173	21274445	207	22422854
174	17729686	250	21006922
175	19773979	251	20219209
176	23445135	252	20626522
177	22195880	253	20079671
178	19022157	255	24494530
179	22422312	256	19848118
180	22311514	258	21143374
181	22268680	259	20348031
182	24197713	260	21805756
186	17543118	282	21468239
187	20590337	299	22565165
188	18389899	301	18980734
189	22802284		

Análisis bioinformático.

Flujo bioinformático para la obtención del archivo de SNP/Indels

El esquema de la figura 1 muestra un resumen del flujo bioinformático empleado para la obtención del archivo de SNPs e Indels (matriz de SNP e Indels) a partir del archivo Fastq (datos crudos de secuenciación).

Figura 1: Flujo bioinformático empleado para la obtención de la matriz de SNP e Indels.



Los pasos de este flujo bioinformático se describen con más detalle a continuación:

Control de calidad. Se empleó el programa FastQC (Andrews, 2010), con el fin de verificar la calidad de los datos sin procesar.

Indexación. El genoma humano 37 (GRCh37) del Consorcio de referencia del genoma (ucsc.hg19.fasta; <https://github.com/creggian/ucsc-hg19-fasta>) se indexó para su uso mediante el programa BWA (v0.7.5a) (Li & Durbin, 2009)

Alineación. Usando BWA (v0.7.5a), que se utiliza para la alineación de lecturas cortas (Li & Durbin, 2009), las secuencias crudas en formato fastq se mapearon con el genoma de referencia previamente indexado, obteniendo resultados en formato SAM (alineación de secuencias) (Li, 2009).

Conversión de a archivos SAM a archivos BAM. Empleando el programa Samtools (Li, 2009), el cual es una colección de herramientas para manipular la alineación resultante en formato SAM / BAM, se convirtió el archivo SAM a BAM (formato mapeado binario) . Posteriormente, se clasificaron los archivos BAM, ordenándolos para el procesamiento de datos optimizado con el fin de evitar una alineación adicional al cargar en la memoria. Subsecuentemente, el archivo ordenado BAM se indexó en un archivo BAM indexado. Finalmente, se empleó el comando *mpileup* para analizar múltiples muestras, brindando cobertura a todas las lecturas asignables.

Llamado de variantes. Se empleó el programa Genome Analysis Toolkit (DePristo et al., 2011)(McKenna et al., 2010) para llamar las variantes tipo SNP e indels pequeñas de los datos de secuenciación de lectura corta alineados con el genoma de referencia.

Anotación de variantes. Después del llamado de variantes, se utilizó ANNOVAR (Wang, Li, & Hakonarson, 2010), tomando los archivos de entrada basados VCF para SNPs e Indels, para anotar funcionalmente las variantes genéticas detectadas en el genoma humano hg19.

Posteriormente, la matriz de SNPs e Indels obtenida se analizó mediante múltiples softwares.

Obtención de la matriz de SNPs e Indels

Se generaron archivos PLINK (PED y MAP) a partir de los VCF obtenidos. Un archivo PED es un archivo delimitado por tabulación donde cada línea representa a un individuo y las primeras seis columnas son obligatorias: identificación de la familia, identificación del individuo, identificación del padre, identificación de la madre, sexo (1 = masculino, 2 = femenino, 0 = perdido) y fenotipo (1 = no afectado, 2 = afectado, 0 = perdido). Las columnas siguientes denotan genotipos que pueden ser de cualquier carácter (por ejemplo 1, 2, 3, 4 o A, C, G, T). 0 denota un genotipo perdido. Cada SNP debe tener dos alelos (es decir, ambos alelos están presentes o ausentes). El orden de los SNP en el archivo PED se da en el archivo MAP, donde cada línea denota un solo marcador y las cuatro columnas separadas por espacios en blanco son: Cromosomas (1-22, X, Y o 0 para no colocados), Nombre del marcador (por lo general, un número rs), Distancia genética en Morgans (esto se puede fijar a 0) y Posición en pares de bases (unidades bp).

Se llevó a cabo un control de calidad de las matrices, empleando una estadística básica en PLINK v1.90b4 (Purcell et al., 2007), que incluyó la estimación de las frecuencias de los SNPs (--freq), de los datos perdidos de SNPs por individuos (--missing; imissing), de datos perdidos de individuos por SNPs (--missing; lmissing) y finalmente, la estimación del Equilibrio Hardy-Weinberg (--hardy); adicionalmente como los datos contienen secuencias no colocadas fue necesario agregar la función --allow-extra-chr.

Se obtuvieron cuatro archivos denominados: *.frq (parámetro a revisar el MAF), *.hwe (Ho vs He), *.imiss y *.lmiss, los cuales se visualizaron usando el programa Haploview (Barrett, Fry, Maller, & Daly, 2005). El MAF, es la frecuencia del alelo menor y se define como la frecuencia con la que aparece el alelo secundario, a, en un determinado locus, dentro de una población. A partir de ese valor se definieron los SNPs, considerando únicamente aquellos cambios de base cuya frecuencia fue mayor del 1%, ya que la potencia estadística es extremadamente baja para variantes raras.

En primera instancia se filtraron los SNPs que no estuvieron presentes en el 90% de todos los individuos empleando la función --geno 0.1, y se calcularon las estadísticas para frecuencia del alelo menor (MAF) y desequilibrio de Hardy-Weinberg ($p < 0.05$). Finalmente, se filtraron los SNP por $MAF = 0$, y con esa matriz se llevaron a cabo los análisis posteriores.

Se obtuvieron dos matrices, una de 3356141 SNPs y otra de 688668 Indels (inserciones o deleciones). Con un promedio de 133460.24 SNPs y 27454.56 Indels por cromosoma (Tabla 3).

Tabla 3: Cantidad de SNPs e Indels encontrados por cromosoma (1-22, X, Y, mitocondrial) para los 55 individuos evaluados.

Cromosoma	Cantidad de SNPs	Cantidad de Indels
1	282993	57958
2	244138	51540
3	194142	41304
4	180608	37717
5	167478	35786
6	195292	39316
7	188710	39410
8	149155	29925
9	142789	27858
10	164333	33293
11	158547	31589
12	161579	34311
13	95863	20522
14	107297	22136
15	109830	22223
16	131340	24621
17	143265	28703
18	78268	16776
19	146026	29303
20	84350	16999
21	48665	10163
22	71267	13331
MT	136	10
X	87 010	20 963
Y	3425	607

La estimación de los datos perdidos de SNPs por individuo, permitió determinar que las muestras solo están genotipadas en promedio en un 0.0851212 (Tabla 4).

Tabla 4: Datos perdidos de SNPs en los individuos.

Familia a la que pertenece	Identificación del individuo	Número de SNPs perdidos	Número de SNPs genotipados	Índice F
-----------------------------------	-------------------------------------	--------------------------------	-----------------------------------	-----------------

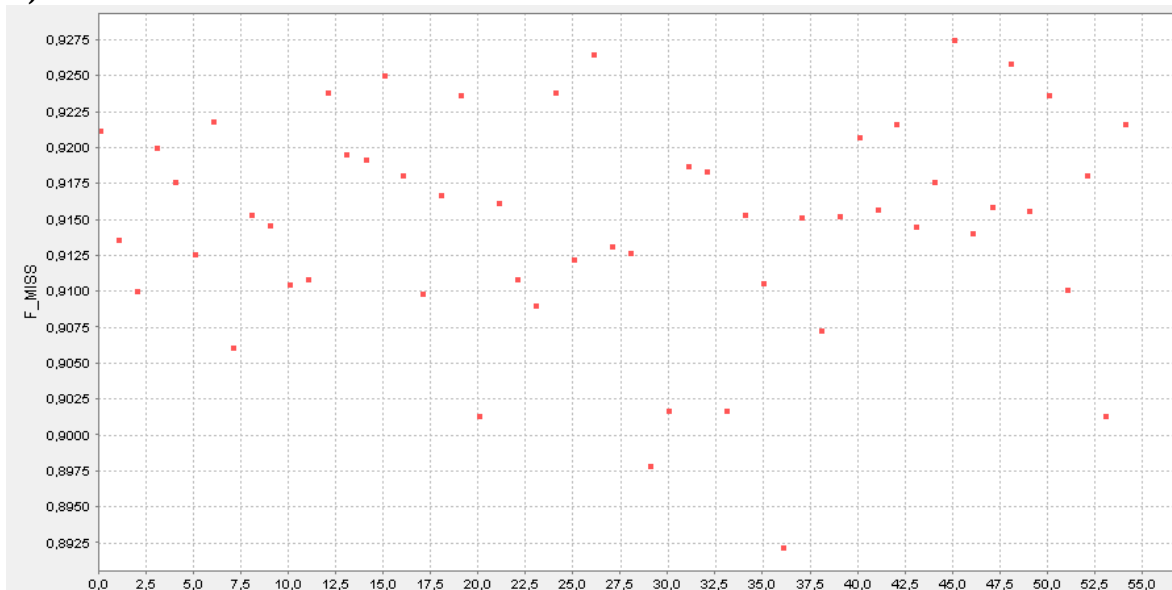
el individuo				
2	200	3088833	3352716	0.9213
2	159	3066522	3356141	0.9137
2	160	3051348	3352716	0.9101
2	161	3088072	3356141	0.9201
2	162	3076950	3352716	0.9177
2	163	3060051	3352716	0.9127
1	165	3094002	3356141	0.9219
1	166	3041331	3356141	0.9062
1	167	3069150	3352716	0.9154
1	168	3069954	3356141	0.9147
1	169	3053029	3352716	0.9106
999	170	3054133	3352716	0.9109
1	171	3097554	3352716	0.9239
1	172	3086469	3356141	0.9196
5	173	3082272	3352716	0.9193
1	174	3101500	3352716	0.9251
1	175	3081749	3356141	0.9182
1	176	3053620	3356141	0.9099
1	177	3073716	3352716	0.9168
1	178	3099999	3356141	0.9237
1	179	3025312	3356141	0.9014
1	180	3071677	3352716	0.9162
1	181	3057133	3356141	0.9109
1	182	3047838	3352716	0.9091
999	186	3100802	3356141	0.9239
1	187	3058651	3352716	0.9123
4	188	3109795	3356141	0.9266
4	189	3061754	3352716	0.9132
4	190	3060263	3352716	0.9128
4	191	3010486	3352716	0.8979
999	192	3026707	3356141	0.9018
999	193	3083727	3356141	0.9188
1	195	3079016	3352716	0.9184
1	196	3026507	3356141	0.9018
1	197	3069219	3352716	0.9154
1	198	3056428	3356141	0.9107
1	199	2994793	3356141	0.8923
1	200	3071412	3356141	0.9152
6	203	3042197	3352716	0.9074
1	204	3068630	3352716	0.9153

1	205	3090336	3356141	0.9208
1	206	3070444	3352716	0.9158
1	207	3090179	3352716	0.9217
1	250	3066480	3352716	0.9146
1	251	3080024	3356141	0.9177
1	252	3113138	3356141	0.9276
1	253	3064847	3352716	0.9141
1	255	3071062	3352716	0.916
1	256	3107358	3356141	0.9259
1	258	3070207	3352716	0.9157
1	259	3096754	3352716	0.9237
1	260	3054704	3356141	0.9102
1	282	3078476	3352716	0.9182
1	299	3022209	3352716	0.9014
1	301	3090054	3352716	0.9217

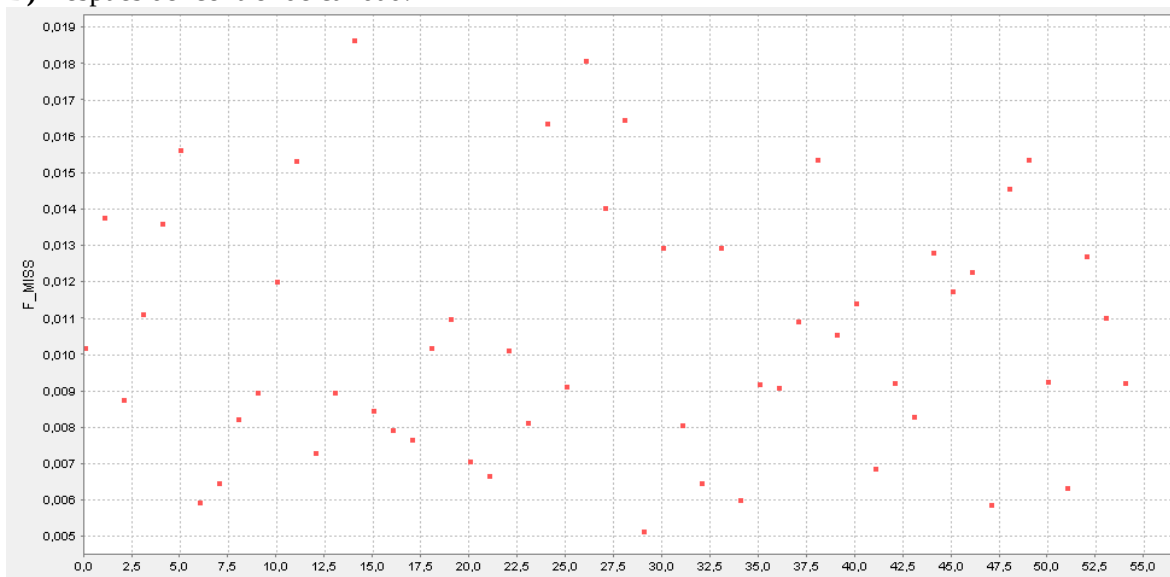
Debido a que todos los individuos se encontraron con 89 a 92% de SNPs perdidos, fue necesario filtrar los SNPs que no estuvieron presentes en el 95% de los individuos, de manera que se pasó de tener 3356141 a 23357 SNPs, mejorando la tasa de genotipado de 0.0851212 a 0.989427 (Figura 2).

Figura 2: Proporción de SNPs perdidos por individuo en los 55 individuos estudiados.

A) Antes del control de calidad.



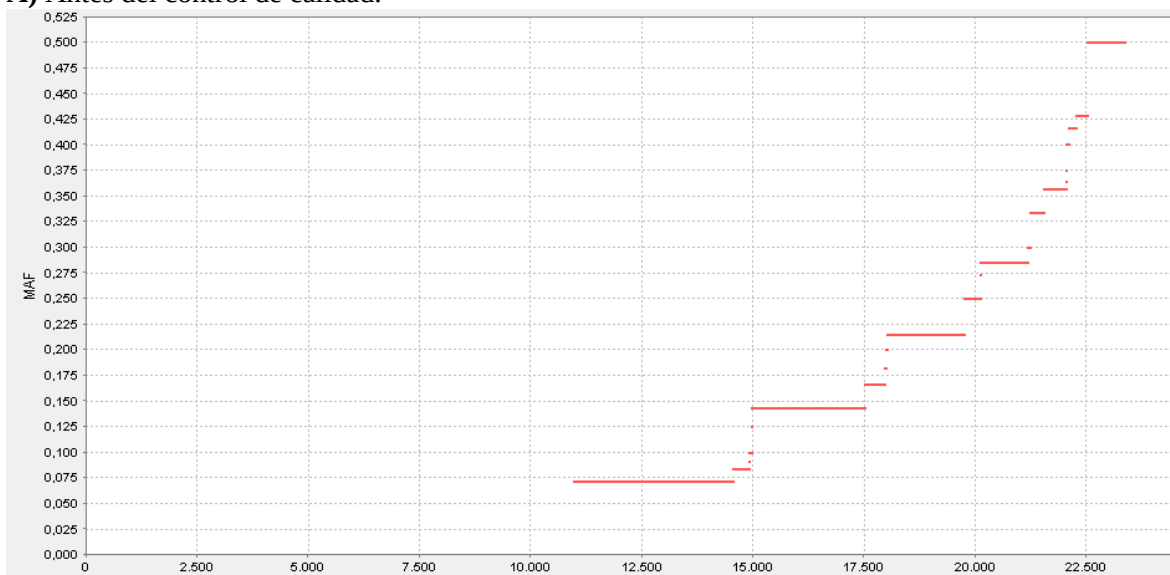
B) Después del control de calidad.



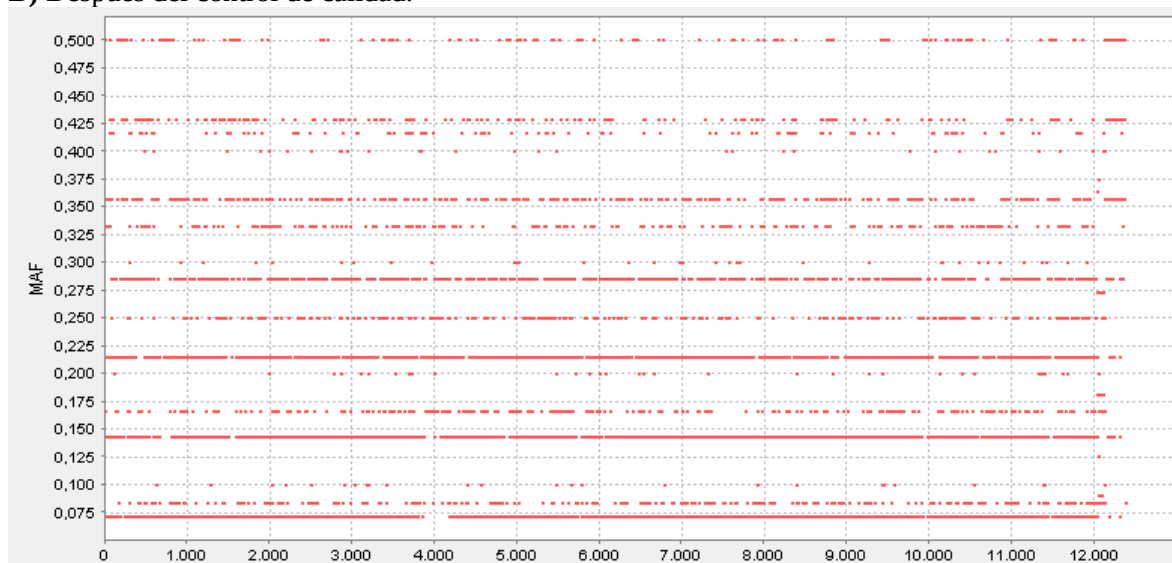
Posteriormente, cuando se calculó la frecuencia del alelo menos frecuente, se encontró que para 10976 variantes era igual a 0, lo que significa que tales SNPs son monomórficos en la muestra. Por lo tanto, se filtraron nuevamente los datos para eliminar los SNPs monomórficos, pasando de una matriz con 23357 a 12381 SNPs (Figura 3).

Figura 3: Frecuencia del menor alelo por SNP para los 55 individuos estudiados.

A) Antes del control de calidad.



B) Después del control de calidad.



La tabla 5 muestra el porcentaje de SNPs conservados para el análisis después de realizar el control de calidad descrito.

Tabla 5: SNPs conservados para el análisis después de realizar el control de calidad (exclusión de los no secuenciados en al menos 95% de los individuos y exclusión de los monomórficos).

Cromosoma	Número de SNPs antes del control de calidad	Número de SNPs después del control de calidad	Proporción de SNPs conservados después del control de calidad	Porcentaje de SNPs conservados después del control de calidad
1	282993	1294	0.00457255	0.457255126
2	244138	768	0.00314576	0.314576182
3	194142	585	0.00301326	0.301325834
4	180608	538	0.00297883	0.297882707
5	167478	507	0.00302726	0.302726328
6	195292	902	0.00461872	0.461872478
7	188710	561	0.00297282	0.297281543
8	149155	467	0.00313097	0.313097114
9	142789	647	0.00453116	0.453116136

10	164333	507	0.0030852	0.308519896
11	158547	705	0.00444663	0.444663097
12	161579	665	0.00411563	0.411563384
13	95863	228	0.00237839	0.237839417
14	107297	404	0.00376525	0.376524973
15	109830	513	0.00467085	0.467085496
16	131340	461	0.00350997	0.350997411
17	143265	594	0.00414616	0.414616271
18	78268	184	0.0023509	0.235089692
19	146026	750	0.00513607	0.513607166
20	84350	187	0.00221695	0.221695317
21	48665	281	0.00577417	0.577417035
22	71267	298	0.00418146	0.418145846
MT	136	0	0	0
X	87010	92	0.00105735	0.105734973
Y	3425	3	0.00087591	0.087591241

Finalmente, se obtuvo una matriz de 12381 SNPs con representación de SNPs en todos los cromosomas, excepto el mitocondrial, y una tasa de genotipado de 0.984681 a 0.989427 (Tabla 6).

Tabla 6: Cantidad y posición de SNPs encontrados por cromosoma (1-22, X, Y) para los 55 individuos evaluados, después de realizar el control de calidad (exclusión de los no secuenciados en al menos 95% de los individuos y exclusión de los monomórficos).

Cromosoma	No. SNPs	Posición
1	1294	0 (14907) - 1293 (248845499)
2	768	1294 (676457) - 2061 (242818097)
3	585	2062 (391100) - 2646 (196750016)
4	538	2647 (49792) - 3184 (189659608)
5	507	3185 (230980) - 3691 (180582604)
6	902	3692 (348051) - 4593 (169637911)
7	561	4594 (550791) - 5154 (159025180)
8	467	5155 (494804) - 5621 (145678020)
9	647	5622 (325817) - 6268 (141070509)
10	507	6269 (93945) - 6775 (135381927)
11	705	6776 (193194) - 7480 (134234406)
12	665	7481 (92252) - 8145 (133797649)
13	228	8146 (19240707) - 8373 (114792842)

14	404	8374 (19123663) - 8777 (107170398)
15	513	8778 (20192951) - 9290 (102251993)
16	461	9291 (109671) - 9751 (90133137)
17	594	9752 (6115) - 10345 (80895933)
18	184	10346 (12976) - 10529 (77420703)
19	750	10530 (372551) - 11279 (59025685)
20	187	11280 (286047) - 11466 (62374219)
21	281	11467 (9907345) - 11747 (48022375)
22	298	11748 (16350323) - 12045 (51064818)
X	92	12046 (200860) - 12137 (155251512)
Y	3	12138 (21154466) - 12140 (59016450)

Búsqueda de un haplotipo común en los pacientes con mutación completa o premutación en FMR1 de la población de Ricaurte – Valle

Obtención de haplotipos

Para la obtención de los haplotipos se consideraron las variantes del cromosoma X, obteniendo una matriz sin filtrar de 87010 SNPs. Posteriormente, debido a que esta matriz contenía un gran número de datos perdidos se procedió a filtrar aquellos que estuvieron presentes en el 100% de los 55 individuos, obteniendo una matriz de 236 SNPs.

Se hizo un análisis de redes de haplotipos en el cual el número de haplotipos fue calculado a partir de la matriz de 236 SNPs del cromosoma X utilizando un análisis de parsimonia en el programa TCS 1.21 imponiendo un límite de conexión del 95% con 11 pasos. Se obtuvieron 37 haplotipos repartidos entre los 55 individuos (Tabla 7).

Tabla 7: Haplotipos obtenidos del análisis estadístico de parsimonia (TCS), que trata los huecos como un estado de 5ª carácter, para un conjunto de SNPs del cromosoma X para 55 individuos (24 hombres y 31 mujeres).

Haplotipos	Individuos (ID)
h1	172, 193
h2	179, 181, 186, 188, 192
h3	168, 199, 205
h4	165, 175, 251, 260
h5	178
h6	161, 176, 256
h7	196
h8	159, 252
h9	200
h10	198

h11	166
h12	203
h13	170
h14	169
h15	160
h16	158
h17	177
h18	258
h19	206
h20	190
h21	187, 250, 299
h22	167
h23	204
h24	171, 174, 255
h25	173
h26	253
h27	259
h28	195
h29	207
h30	282, 301
h31	189
h32	162
h33	182
h34	163
h35	197
h36	191
h37	180)

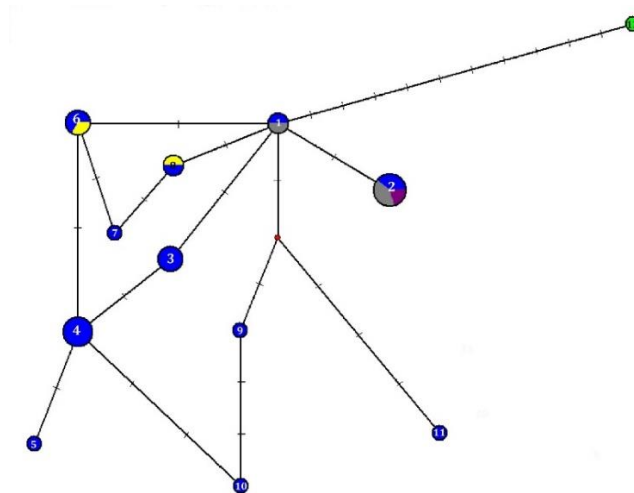
Los haplotipos h1 a h11 son los de los hombres, todos portadores hemicigotos de premutación o mutación completa y se obtuvieron directamente de la secuenciación. El resto de haplotipos, pertenecientes a las mujeres, una de ellas homocigota para el FMR1 normal, y el resto heterocigotas para el FMR1 patológico, son producto de la unificación de los 2 haplotipos obtenidos directamente de la secuenciación. Para esta unificación, cada vez que se encontró heterocigosidad en alguno de los SNPs incluidos en el haplotipo, se eligió siempre el más raro. Los haplotipos h13 a h37, son los de todas las portadoras heterocigotas de premutación o mutación completa, pero dado que esta es una muestra donde predominan los cromosomas expandidos (86 cromosomas X en total. 54 expandidos, 31 no expandidos), al elegir el alelo más raro para los haplotipos femeninos, se buscó obtener los haplotipo de los cromosomas no expandidos, de manera que, junto con el haplotipo h12, de la mujer homocigota para el alelo normal, constituyeran controles de los haplotipos de los cromosomas expandido de los hombres.

Construcción de redes de haplotipos

Entre los individuos con premutación o mutación completa, se observa mayor número de haplotipos en las mujeres que en los hombres (11 haplotipos para 24 hombres y 26 haplotipos para 30 mujeres), desde luego, también se observan muchos más haplotipos exclusivos en las

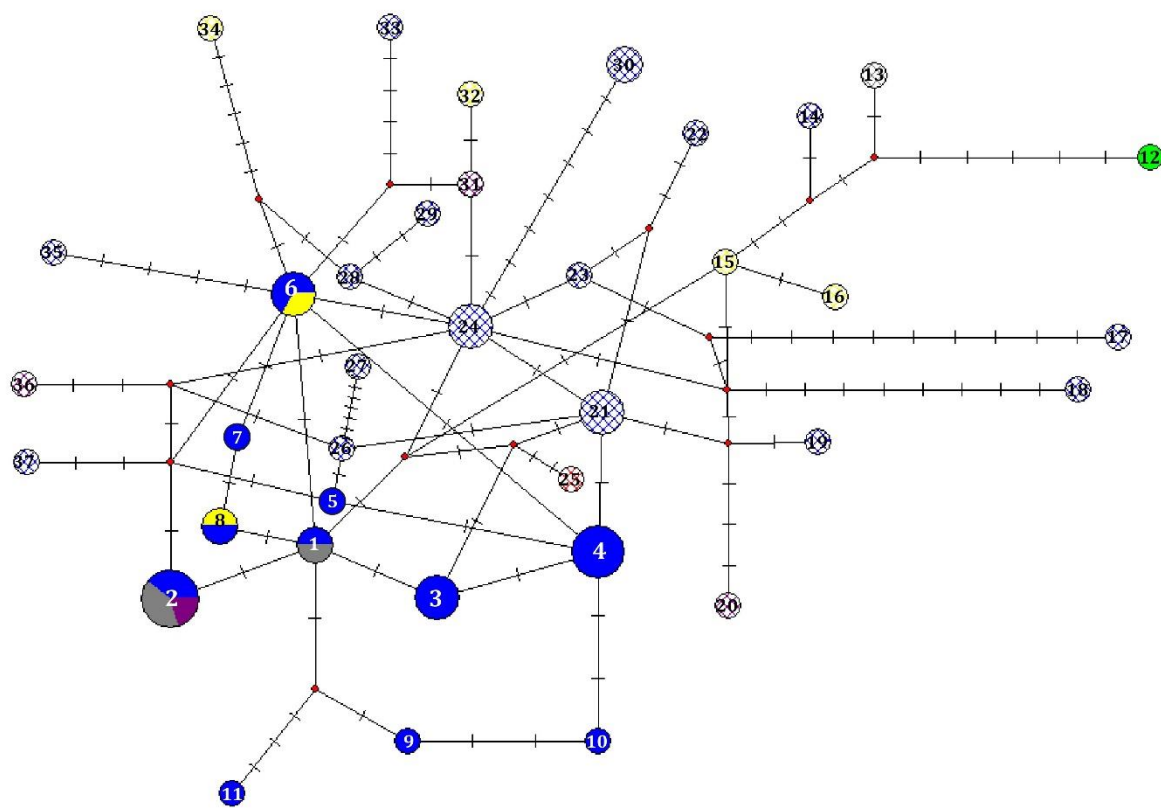
mujeres que en los hombres (5 hombres con haplotipos exclusivos y 22 mujeres con haplotipos exclusivos). Dado que los haplotipos de las mujeres, que son heterocigotas para el FMR1 patológico, resultan de la unificación de 2 haplotipos (el presente en cada uno de los cromosomas X), y solo uno de los 2 es el realmente asociado al FMR1 patológico, no hay seguridad de que el haplotipo obtenido para las mujeres sea exactamente el asociado al FMR1 patológico, de hecho, al seleccionar al alelo menos frecuente para la construcción del haplotipo, precisamente se pretendió obtener el haplotipo no asociado al FMR1 patológico. Es esperada la presencia de un alto número de haplotipos exclusivos. Para los hombres en cambio, por ser hemicigotos, y haberse obtenido los haplotipos directamente de la secuenciación, los haplotipos son los realmente asociados con el FMR1 patológico. Por lo tanto, se construyeron 2 redes de haplotipos, una incluyendo solo a los 24 hombres y la mujer control (Figura 4), y otro con la totalidad de los individuos (Figura 5). Esperando que la primera red fuera ilustrativa del probable efecto fundador.

Figura 4: Red de haplotipos de los hombres de las familias construido a partir de 236 SNPs. Codificación de las familias: 1 (azul), 2 (amarillo), 4 (morado), 999 (gris) y el control (6: verde).



Cada haplotipo es representado por un círculo con tamaño proporcional al número de individuos que comparten el haplotipo. Los números identifican el haplotipo. Los puntos rojos indican haplotipos faltantes y las líneas divisorias indican las mutaciones.

Figura 5: Redes de haplotipos de 55 individuos pertenecientes a las familias.



Codificación de las familias: 1 (azul), 2 (amarillo), 4 (morado), 5 (rojo), 999 (gris) y el control (6: verde), construido a partir de 236 SNPs. Cada haplotipo es representado por un círculo con tamaño proporcional al número de individuos que comparten el haplotipo. Los números identifican los haplotipos. Los puntos rojos indican haplotipos faltantes, las líneas divisorias indican las mutaciones, los círculos de colores sólidos corresponden a los hombres y los círculos con líneas perpendiculares corresponden a las mujeres.

Al analizar la red de haplotipos construida a partir de los 24 hombres y la mujer control, que incluyó 12 haplotipos, se encontró que, entre los haplotipos de los hombres, el más ancestral fue el h1, presente en individuos de las familias 1 y 999, los haplotipos h6 y h8 se presentaron en individuos de las familias 1 y 2, el haplotipo más diverso fue el h2, presente en individuos de las familias 1, 4 y 999, y 7 haplotipos (h3- h5, h7, h9-h11) fueron exclusivos de la familia 1. También se encontró que, como era de esperarse en el contexto de un probable efecto fundador, los 11 haplotipos de los hombres fueron bastante similares entre sí, y difirieron entre ellos únicamente por 1, 2 o 3 SNPs (promedio 1.2), además, el haplotipo de la mujer control, es decir, la homocigota para el FMR1 normal, se separó del resto de haplotipos por 12 mutaciones.

Al analizar la red de los 37 haplotipos de los 55 individuos (24 hombres y 31 mujeres), se evidenció de nuevo el hallazgo ya descrito para los hombres, pero además se observó que los haplotipos de las mujeres son diferentes a los de los hombres y diferentes entre ellos. Se

observaron haplotipos únicos para la familia 2 (h15, h16, h32 y h34), para la familia 4 (h20, h31 y h36), para la familia 999 (h13) y para la familia 5 (h25). En total, para la familia 1 se encontraron 16 haplotipos (h14, h17, h18, h19, h21, h22, h23, h24, h26, h27, h28, h29, h30, h33, h35, h37). La mujer control (homocigota para el FMR1 normal) se agrupó en el lado de las mujeres, separándose por 7 mutaciones del haplotipo femenino más cercano h13, y muy separada de los haplotipos masculinos.

En contraste con lo observado en el análisis de endogamia descrito previamente, en el cual se consideraron todos los cromosomas y se encontró una heterocigosidad mayor a la esperada, este análisis de los cromosomas X muestra una población más homogénea y evidencia que el cromosoma X que alberga la expansión en FMR1 se encuentra relativamente conservado.

Haplotipo común

Considerando los 236 SNPs secuenciados en el 100% de los pacientes, se encontraron 37 haplotipos en la población estudiada. Si de los 37 haplotipos se excluyen los SNPs monomórficos en toda la población, se obtienen 37 haplotipos de 58 SNPs. La frecuencia de los alelos de esos 58 SNPs se muestra en la tabla 8.

Tabla 8: Frecuencias alélicas en los 37 haplotipos de 58 SNPs del cromosoma X.

Posición	Frecuencia del alelo 1	Frecuencia del alelo 2
229137	G:0.990909	A:0.00909091
970046	T:0.954545	C:0.0454545
970843	G:0.972727	A:0.0272727
1314764	G:0.972727	T:0.0272727
1404936	C:0.836364	T:0.163636
1536641	A:0.990909	G:0.00909091
1560998	A:0.981818	G:0.0181818
2609749	C:0.818182	T:0.181818
2724760	C:0.990909	T:0.00909091
2726396	G:0.990909	A:0.00909091
11313027	C:0.972727	T:0.0272727
15817963	G:0.981818	A:0.0181818
15827524	G:0.972727	T:0.0272727
16762002	G:0.990909	A:0.00909091
17092152	T:0.990909	G:0.00909091
19041592	T:0.909091	C:0.0909091
40456961	C:0.990909	G:0.00909091
45011340	C:0.990909	T:0.00909091
47372250	T:0.990909	C:0.00909091
48053428	A:0.990909	C:0.00909091
48053521	C:0.990909	T:0.00909091

48054740	G:0.990909	C:0.00909091
48432791	C:0.990909	T:0.00909091
48435928	T:0.981818	C:0.0181818
69262423	C:0.981818	T:0.0181818
69652762	G:0.981818	A:0.0181818
71357093	A:0.990909	G:0.00909091
71357160	T:0.990909	G:0.00909091
71363415	T:0.990909	C:0.00909091
71364479	C:0.990909	A:0.00909091
91456621	T:0.990909	G:0.00909091
100911066	T:0.990909	C:0.00909091
114426171	T:0.990909	C:0.00909091
117817266	T:0.990909	C:0.00909091
125953932	G:0.963636	A:0.0363636
128957640	C:0.990909	T:0.00909091
130416958	T:0.990909	C:0.00909091
138774209	T:0.990909	C:0.00909091
138811003	A:0.990909	G:0.00909091
148038321	G:0.981818	A:0.0181818
148049264	G:0.763636	T:0.236364
148055144	G:0.981818	A:0.0181818
148622730	A:0.954545	G:0.0454545
148622745	A:0.954545	T:0.0454545
148890964	G:0.981818	T:0.0181818
149680221	C:0.972727	G:0.0272727
149680554	T:0.972727	C:0.0272727
149826503	A:0.972727	G:0.0272727
150336814	A:0.990909	G:0.00909091
150869267	G:0.981818	A:0.0181818
151283336	C:0.981818	T:0.0181818
152018832	G:0.990909	T:0.00909091
152807665	G:0.981818	A:0.0181818
152825414	T:0.990909	G:0.00909091
152830322	G:0.972727	A:0.0272727
153129556	A:0.963636	G:0.0363636
153668172	G:0.972727	A:0.0272727
155004437	T:0.909091	G:0.0909091

La observación detallada de los 11 haplotipos masculinos de 58 SNPs (Figura 6) evidencia que todos son idénticos para 33 SNPs (color verde comprendido en una región de 83.567.899 pb) circundantes a FMR1, e incluso 10 de los 11 haplotipos masculinos son idénticos para

Figura 6: Haplotipo conservado en los hombres de Ricaurte con expansión.



La presencia de un mismo haplotipo de 33 SNPs, comprendido en una región de 83.567.899 pb y circundantes a FMR1, todos los hombres (con mutación completa o premutación) de la muestra, prueba que el conglomerado X frágil de Ricaurte es producto de un efecto fundador. Llama la atención que incluso el individuo masculino de la familia 4, oriunda de Robledo, también comparte ese haplotipo. Si bien, la familia 4 se separó claramente del resto de familias en los análisis realizados a partir de la matriz de SNPs de todos los cromosomas, al realizar análisis del cromosoma X, esta familia se separa claramente de las demás, lo que sugiere que el cromosoma que alberga la expansión en esta familia, tiene el mismo origen que cromosoma con la expansión de Ricaurte.

Determinación de las relaciones filogenéticas entre los individuos de Ricaurte con expansión en FMR1 (premutación o mutación completa), la estructura genética de la población, y distancia genética con respecto a poblaciones del proyecto 1000 genomas

Los datos se analizaron para equilibrio Hardy-Weinberg (H-W), heterocigosidad y endogamia, y relaciones familiares. Además, tanto para el conjunto de todos los SNPs obtenidos del llamado de variantes como para el conjunto de SNPs filtrados, se llevaron a cabo análisis de componentes principales (ACP) y análisis de agrupamiento jerárquico.

Equilibrio Hardy-Weinberg

La desviación de los SNPs con respecto al equilibrio Hardy-Weinberg (H-W) para cada locus, para cada familia de muestreo y para la población en general, se calculó mediante el programa Arlequín v3.01 (Excoffier, Laval, & Schneider, 2005) usando el método de Cadenas de Markov para 100000 permutaciones (Guo & Thompson, 1992). Se encontraron 749 SNPs en desequilibrio de H-W con un $P < 0.05$, repartidos en todos los autosomas (Tabla 9).

Tabla 9: Cantidad de SNPs que se encuentran en desequilibrio Hardy-Weinberg por cada cromosoma.

Cromosoma	No. SNP
1	199
2	58
3	7
4	7
5	4
6	20
7	43
8	15
9	43
10	39
11	2
12	21
13	1
14	8
15	17
16	26
17	45
18	9
19	3
20	9
21	41
22	4

La presencia de SNPs en desequilibrio de H-W en todos los autosomas es esperada para una población como Ricaurte. En primer lugar, es pequeña de menos de 1200 habitantes, y en segundo lugar, el hecho de albergar el conglomerado de síndrome X frágil ha podido llevar durante años a un aislamiento social, en el que los miembros de las poblaciones aledañas evitan emparejarse con personas de Ricaurte. Hay que considerar que estos hallazgos son resultado del análisis de una matriz filtrada de apenas 12381 SNPs, pero podría esperarse que muchos más SNPs estuvieran en desequilibrio de H-W, por ejemplo los 10976 que fueron eliminados en el control de calidad por ser monomórficos en la población.

Heterocigosidad y endogamia

La diversidad genética fue estimada con una serie de indicadores: heterocigosidad alélica observada (H_o), heterocigosidad esperada (H_e) (Levene, 1949) y heterocigosidad esperada insesgada de Nei (Nei, 1978) para familia usando como estimación la media con su respectiva desviación estándar empleando el programa Genepop (ROUSSET, 2008). El coeficiente de endogamia (F_{is}) para las familias muestreadas, el cual mide la proporción en que disminuye la frecuencia de heterocigotos de una población en H-W se halló por medio del programa Genepop y resultó negativo (F_{is} ; -0.33).

El coeficiente de endogamia como medida indirecta de consanguinidad, mide la reducción en la heterocigosidad individual debida a las desviaciones de los cruzamientos al azar. Interesantemente, el resultado negativo de este coeficiente sugiere que hay un exceso de heterocigotos en el total de la población evaluada. Incluso, en cada familia también se observan valores negativos, indicando exceso de heterocigotos. La familia 4 es la que presenta un H_o más alto con 0.41 mientras los individuos de la familia 1 son los que presentan el menor valor de H_o de 0.29 (Tabla 10).

Tabla 10: Heterocigosidad observada (H_o), Heterocigosidad esperada (H_e), y F_{is} por familia de 12381 SNPs evaluados.

Familias	H_o	H_e	F_{is}
Familia 2	0,32	0,23	-0,38
Familia 1	0,29	0,22	-0,31
Familia 6 (control)	-	-	-
Familia 4	0,41	0,29	-0,39
Familia 5	-	-	-
Familia 999	0,33	0,25	-0,34
Total			-0,33

Una posible explicación a este hecho es que, al ser conocedores de la posibilidad de portar una anomalía genética, por la presencia de casos de afectados en sus familias, los pobladores de Ricaurte intencionalmente buscan parejas no oriundas de esta población.

Análisis de Identidad por Estado (IBS)

Se realizó un análisis de relaciones familiares mediante la estimación de la disimilitud usando la identidad por estado (IBS) (Tabla 11).

Tabla 11: Matriz de disimilitud basado en IBS para 12381 SNPs en 55 individuos pertenecientes a seis familias.

	Familia 1	Familia 2	Familia 4	Familia 5	Familia 6 (control)	Familia 999
Familia 1	0,153	0,168	0,193	0,181	0,183	0,167
Familia 2	0,168	0,142	0,198	0,183	0,189	0,174
Familia 4	0,193	0,198	0,181	0,216	0,223	0,196
Familia 5	0,181	0,183	0,216	NaN	0,206	0,194
Familia 6 (control)	0,183	0,189	0,223	0,206	NaN	0,196
Familia 999	0,167	0,174	0,196	0,194	0,196	0,162

Este análisis muestra que los coeficientes de disimilitud más altos (mayores a 0,2) se presentan entre la familia 4 (oriundos de Robledo con casos X frágil) y el individuo 5 (mujer con premutación oriunda de la Herradura); la familia 4 (oriundos de Robledo con casos X frágil) y el individuo 6 (mujer normal oriunda de Ricaurte pero no perteneciente a las familias con X frágil); y el individuo 5 (mujer con premutación oriunda de la Herradura) y el individuo 6 (mujer normal oriunda de Ricaurte pero no perteneciente a las familias con X frágil).

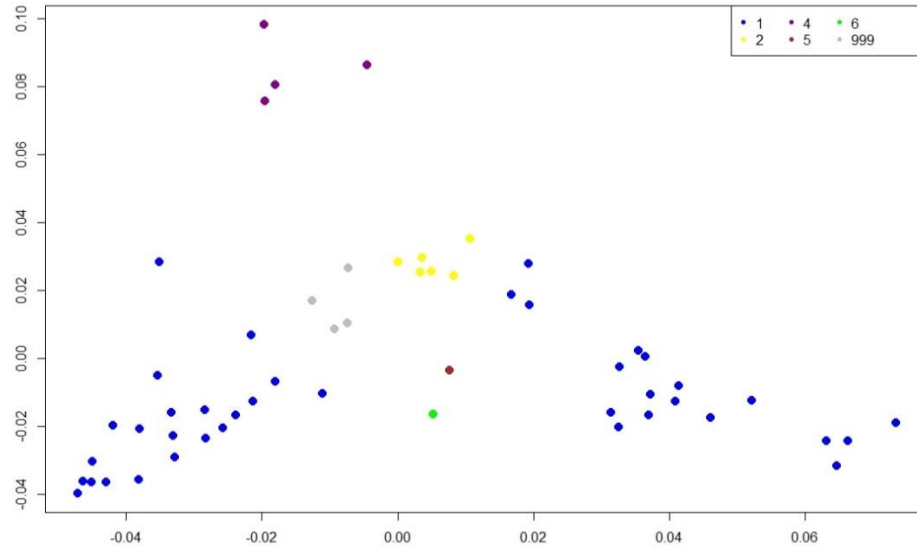
La familia con coeficientes más altos de disimilitud es la 4 (oriundos de Robledo con casos X frágil), que con un coeficiente de 0.19 se separó de las familias oriundas de Ricaurte con casos X frágil (familias 1, 2 y 999).

Los coeficientes más bajos de disimilitud se observaron entre las familias oriundas de Ricaurte con casos X frágil: 0.168 entre las familias 1 y 2, 0.174 entre las familias 2 y 999, y 0.167 entre las familias 1 y 999.

Considerando los SNPs de la matriz filtrada se realizó un análisis de identidad por estado (IBS) y se obtuvo un dendograma a partir de un análisis de disimilaridad usando IBS, evidenciando una separación de los individuos de la familia 4 (oriundos de Robledo con casos X frágil) de las demás familias, y mostrando una subestructuración por núcleos familiares (Figura 7).

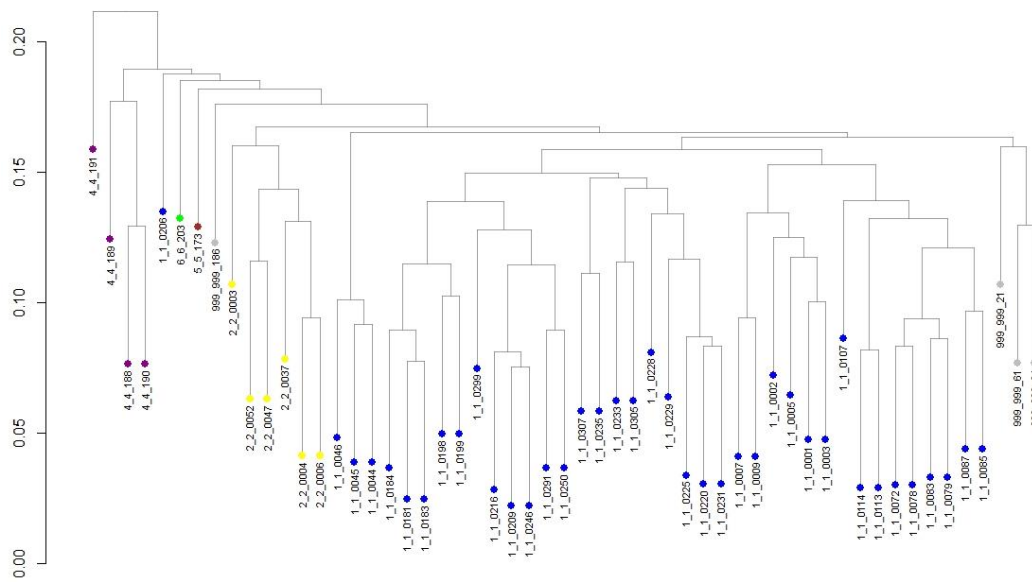
Figura 7: Gráficos de Identidad por Estado para 12381 SNPs en 55 individuos.

A) Análisis de Identidad por Estado.



Los colores asignados a cada familia se especifican en la esquina superior derecha de la gráfica.

B) Dendrograma obtenido a partir de un análisis de disimilaridad usando IBS



La codificación de colores es la misma de la figura 7A.

En general, el análisis de identidad por estado basado en la matriz de SNPs filtrada de 12381 SNPs muestra a la familia 4 (oriunda de Robledo con casos X frágil) alejada del resto de familias con casos y del control normal, lo cual concuerda con lo afirmado por ellos mismos sobre la ausencia de nexos familiares con Ricaurte.

Estructura genética de la muestra estudiada

La estimación de la estructura genética se realizó por medio de dos métodos, el método Bayesiano empleando el programa STRUCTURE (Pritchard, Stephens, & Donnelly, 2000) y realizando un Análisis de Varianza Molecular (AMOVA) (Excoffier, Smouse, & Quattro, 1992) mediante el programa Arlequín versión 3.1 (Excoffier et al., 2005). El AMOVA fue aplicado a la variación observada en los SNPs, con el fin de determinar la proporción de esa variación, analizada bajo dos niveles jerárquicos: entre familias de muestreo y dentro de las familias de muestreo.

Para la comparación entre y dentro de las familias de muestreo se calcularon los F-estadísticos de Wright (Wright, 1965): F_{IS} (coeficiente de consanguinidad), F_{ST} (índice de fijación) y F_{IT} (coeficiente de consanguinidad total).

F_{IS} mide el déficit de heterocigotos en cada población y se representa por

$$F_{IS} = \frac{\bar{H}_S - H_I}{\bar{H}_S} = 1 - \frac{H_I}{\bar{H}_S},$$

donde $\bar{H}_S$ es el promedio de las heterocigosidades esperadas y H_I es el promedio de las heterocigosidades observadas.

F_{ST} mide el déficit de heterocigotos entre poblaciones y se representa por

$$F_{ST} = \frac{H_T - \bar{H}_S}{H_T} = 1 - \frac{\bar{H}_S}{H_T}$$

F_{IT} calcula el déficit global de heterocigotos y se representa por

$$F_{IT} = \frac{H_T - H_I}{H_T} = 1 - \frac{H_I}{H_T},$$

donde H_T es la heterocigosidad esperada considerando las poblaciones analizadas como un todo virtual.

Se valoraron las desviaciones significativamente diferentes de cero usando una prueba con permutaciones, se obtuvieron valores promedios mediante repeticiones sobre todas las localidades de muestreo y los intervalos de confianza (95%) mediante un submuestreo con reemplazo “bootstrap”, usando el método de Weir y Cockerham (Weir & Cockerham, 1984) el cual incorpora el tamaño de las muestras y el número de localidades de muestreo, por medio de un AMOVA (Excoffier et al., 1992) y empleando el programa Arlequín. Los niveles de significancia de los componentes de la varianza y los índices de fijación se evaluaron utilizando 9999 permutaciones.

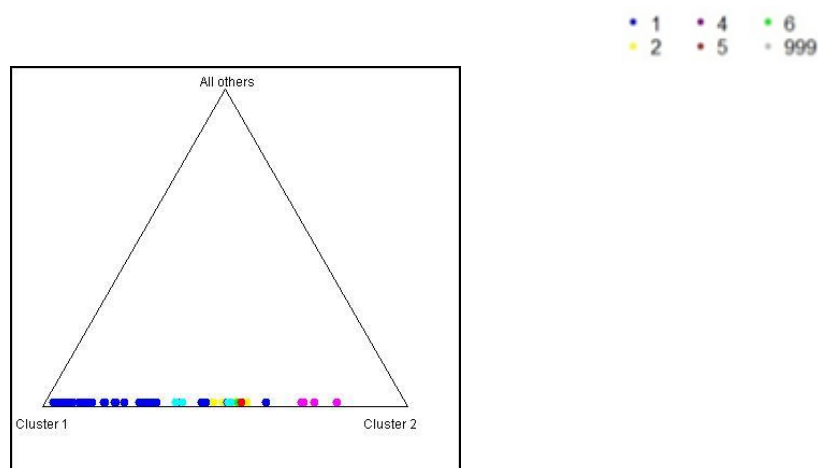
Para la estimación de la estructura genética a partir del programa STRUCTURE (Pritchard et al., 2000), basado en los valores de $\ln P$ y el enfoque δK de Evanno (Evanno, Regnaut, & Goudet, 2005), se encontró que un modelo con $K = 2$ grupos genéticos se adaptaba mejor a los datos (Tabla 12).

Tabla 12: Estadísticas del STRUCTURE resultante del análisis de los 55 individuos utilizando 12381 SNPs. El análisis de Delta K seleccionó K = 2 grupos óptimos (en rojo).

K	Reps	Mean LnP(K)	Stdev LnP(K)	Ln'(K)	Ln''(K)	Delta K
1	10	-523785.75000	144.64634	NA	NA	NA
2	10	-825523.60000	403569.39751	-301737.85000	1709256.16111	4.23535
3	9	-2836517.61111	1613072.81103	-2010994.01111	2743213.96778	1.70061
4	10	-7590725.59000	3511542.01788	-4754207.97889	2902938.13111	0.82668
5	10	-15247871.70000	3645249.39006	-7657146.11000	5409123.68000	1.48388
6	10	-17495894.13000	3074142.76852	-2248022.43000	5441851.48000	1.77020
7	10	-14302065.08000	5680049.13622	3193829.05000	NA	NA

La figura 8 evidencia que los dos grupos definidos por STRUCTURE correspondieron, respectivamente, a los individuos de la Familia 4 (grupo 2) vs todas las demás familias (grupo 1), indicando la divergencia de las frecuencias alélicas entre los grupos.

Figura 8: Proporción de miembros de cada población predefinida en cada uno de los 2 grupos de la estructura genética definida por STRUCTURE



Los colores asignados a cada familia se especifican en la esquina superior derecha de la gráfica.

Las probabilidades de pertenencia de individuos a sus grupos respectivos fueron altas, especialmente dentro del grupo 1 definido por STRUCTURE, que incluye a los individuos de la familia 1, que en su mayoría mostraron porcentajes de asignación mayores a 63%. El grupo 2, conformado por los individuos pertenecientes a la familia 4, presentó porcentajes de

asignación mayores a 70%. El resto de familias e individuos mostraron mezcla genética de sus genotipos con ambos grupos (Tablas 13 y 14. Figura 9).

Tabla 13: Probabilidades de pertenencia de los 55 individuos a los 2 grupos genéticos óptimos inferidos a partir de STRUCTURE.

Individuo	Id Familia	Clúster 1	Clúster 2	Individuo	Id_Familia	Clúster 1	Clúster 2
158	1	0,539	0,461	200	2	0,571	0,429
159	1	0,491	0,509	204	2	0,98	0,02
160	1	0,459	0,541	205	2	0,822	0,178
161	1	0,48	0,52	206	2	0,998	0,002
162	1	0,511	0,489	207	2	0,908	0,092
163	1	0,443	0,557	250	2	0,886	0,114
165	2	1	0	251	2	0,703	0,297
166	2	0,713	0,287	252	2	0,628	0,372
167	2	0,999	0,001	253	2	0,819	0,181
168	2	0,849	0,151	255	2	0,942	0,058
169	2	0,92	0,08	256	2	0,796	0,204
171	2	0,969	0,031	258	2	0,387	0,613
172	2	0,916	0,084	259	2	0,914	0,086
174	2	0,995	0,005	260	2	1	0
175	2	0,904	0,096	282	2	0,966	0,034
176	2	0,925	0,075	299	2	0,999	0,001
177	2	0,75	0,25	301	2	0,948	0,052
178	2	0,738	0,262	203	3	0,467	0,533
179	2	0,731	0,269	188	4	0,279	0,721
180	2	0,854	0,146	189	4	0,287	0,713
181	2	0,643	0,357	190	4	0,248	0,752
182	2	0,706	0,294	191	4	0,185	0,815
187	2	0,975	0,025	173	5	0,458	0,542
195	2	0,968	0,032	186	6	0,488	0,512
196	2	0,824	0,176	192	6	0,629	0,371
197	2	0,978	0,022	193	6	0,647	0,353
198	2	0,559	0,441	170	6	0,494	0,506
199	2	0,498	0,502				

Figura 9: Resultados de los análisis de STRUCTURE de 55 individuos utilizando 12381 SNPs. Óptimo K = 2. Las barras verticales representan la probabilidad de pertenencia de un individuo a un grupo genético con colores que designan los diferentes grupos.

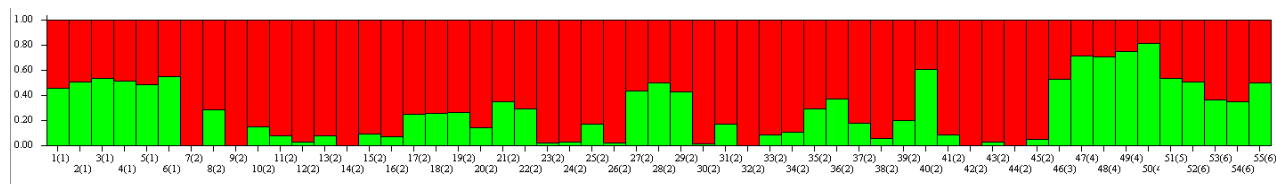


Tabla 14: Las probabilidades de pertenencia de la familia a los 2 grupos genéticos óptimos inferidos a partir de STRUCTURE.

Familia	Clúster 1	Clúster 2	No. Individuos
2	0.487	0.513	5
1	0.838	0.162	39
6 (control)	0.467	0.533	1
4	0.250	0.750	4
5	0.458	0.542	1
999	0.564	0.436	4

De nuevo se muestra a la familia 4 (oriunda de Robledo con casos X frágil) como un grupo diferente del resto de familias con casos y del control normal.

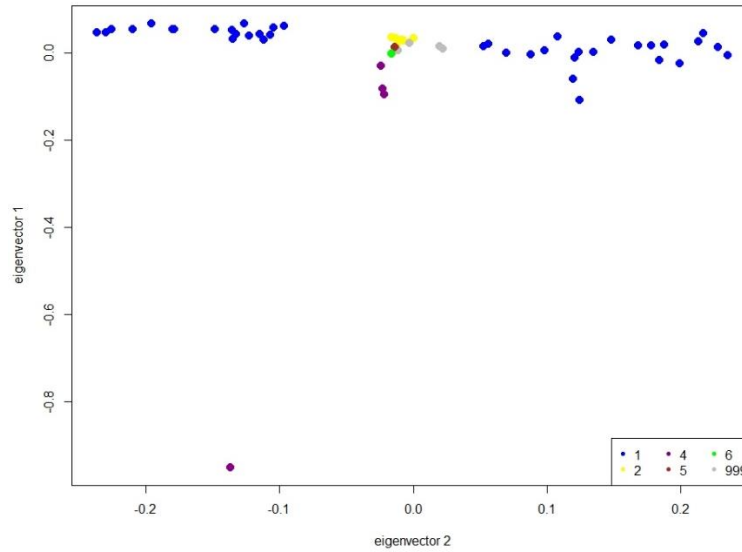
Análisis de componentes principales

Para hacer observables las correlaciones genéticas entre los individuos se llevaron a cabo varios análisis de componentes principales (ACP) independientes usando la matriz de SNPs sin filtrar, es decir con 3356141 SNPs (Figura 10A), y la matriz filtrada, es decir con 12381 SNPs (Figura 10B), observándose una topología similar.

El ACP para la matriz sin filtrar, permitió explicar el 3.2% de la varianza en el componente 1 y el 3.1% de la varianza en el componente 2, observándose una separación de los individuos de la familia 4. Por su parte, el ACP de la matriz filtrada permitió explicar el 5.6% de la varianza en el primer componente y el 4.8% de la varianza en el segundo componente. El primer componente permitió evidenciar con mayor claridad la separación de los individuos de la familia 4 de las demás familias, y el segundo componente permitió evidenciar algunos subgrupos de la familia 1.

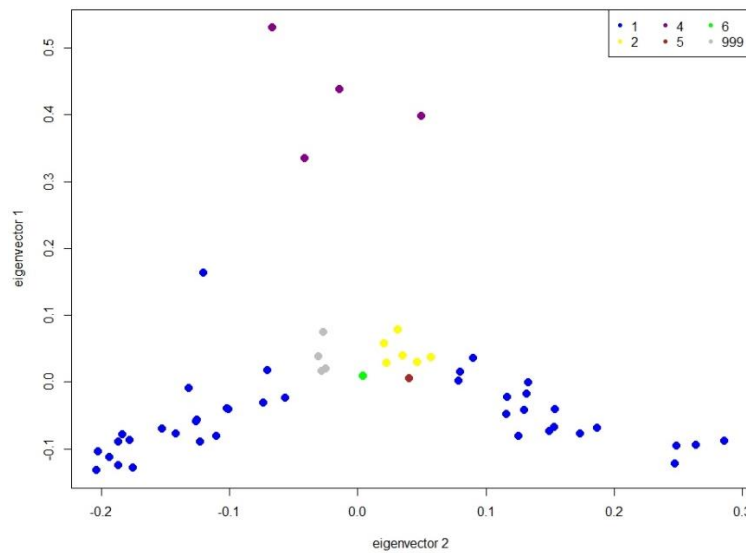
Figura 10: Análisis de Componentes principales en los 55 individuos.

A) Para la matriz sin filtrar (3356141 SNPs).



Los colores asignados a cada familia se especifican en la esquina superior derecha de la gráfica.

B) Para la matriz filtrada (12381 SNPs).



Los colores asignados a cada familia se especifican en la esquina superior derecha de la gráfica.

Se calculó el índice de diferenciación genética (F_{st}), para las seis familias estudiadas empleando el método de Weir & Cockerham (Weir & Cockerham, 1984), obteniendo un valor de 0.017, que es muy bajo. Por consiguiente, se realizó un análisis de varianza molecular (AMOVA) para las cuatro familias con más de un individuo (1, 2, 4 y 999), obteniendo un valor de 0.0241 significativo con un $p < 0.001$ (Tabla 15).

Tabla 15: Resultados del Análisis de Varianza Molecular (AMOVA) basado sobre 12381 SNPs para cuatro familias.

Fuente de Variación	d.f.	Suma de Cuadrados	Varianza de los componentes	Porcentaje de Variación
Entre Familias	3	5754,81	34,34	2,41*
Dentro de las Familias	102	141916,87	1391,34	97,59

Total	105	147671,67	1425,68	
-------	-----	-----------	---------	--

Índices de Fijación: $*(p < 0.001)$

Adicionalmente, valores pareados de F_{st} fueron calculados entre las familias, encontrando valores muy bajos, indicativos de alta cercanía, entre las parejas de familias 999 y 1 ($F_{st} = 0.002$), y 999 y 2 ($F_{st} = 0,008$). Mientras que se encontraron valores más altos, indicativos de mayor diferenciación, entre las parejas de familias 2 y 1 ($F_{st} = 0.029$), 4 y 2 ($F_{st} = 0.034$) y 4 y 1 ($F_{st} = 0.043$) (Tabla 16)

Tabla 16: Valores de F_{ST} entre seis parejas de familias (debajo de la diagonal). Valores de P-valor para los F_{st} calculados (encima de la diagonal, en rojo los significativos $P < 0.05$).

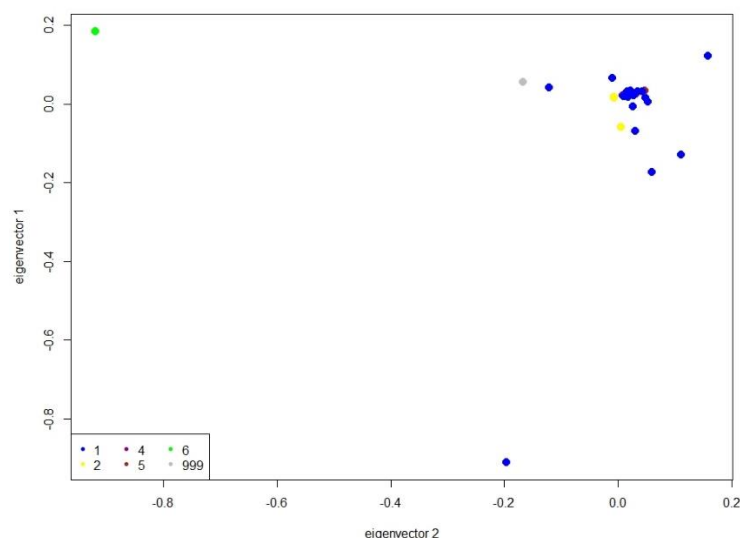
	Familia 2	Familia 1	Familia 6 (control)	Familia 4	Familia 5	Familia 999
Familia 2	0	0	0,99	0	0,99	0
Familia 1	0,029	0	0,99	0	0,99	0
Familia 6 (control)	-0,044	-0,031	0	0,13	0,99	0,14
Familia 4	0,034	0,043	-0,083	0	0,99	0,02
Familia 5	-0,061	-0,037	-0,329	-0,095	0	0,15
Familia 999	0,008	0,002	-0,083	-0,005	-0,089	0

Este ACP muestra una vez más que la familia 4 (oriunda de Robledo con casos X frágil) es un grupo diferente del resto de familias con casos y del control normal.

Teniendo en cuenta únicamente el contenido del cromosoma X, se llevaron a cabo varios ACP independientes usando una matriz sin filtrar con de 95635 SNP-Indels y otra matriz filtrada con 236 SNPs-Indels, observándose una topología similar.

En el ACP de la matriz filtrada para las seis familias estudiadas, el primer componente permitió explicar el 13% de la varianza total y separó el individuo 258 (mujer con premutación de la familia 1) de los demás, y el segundo componente explicó el 12% de la varianza, y permitió separar el individuo 203 (mujer normal oriunda de Ricaurte no perteneciente a las familias con casos X frágil) de los demás (Figura 11).

Figura 11: Análisis de Componentes principales para 236 SNPs-Indels del cromosoma X de 55 individuos.



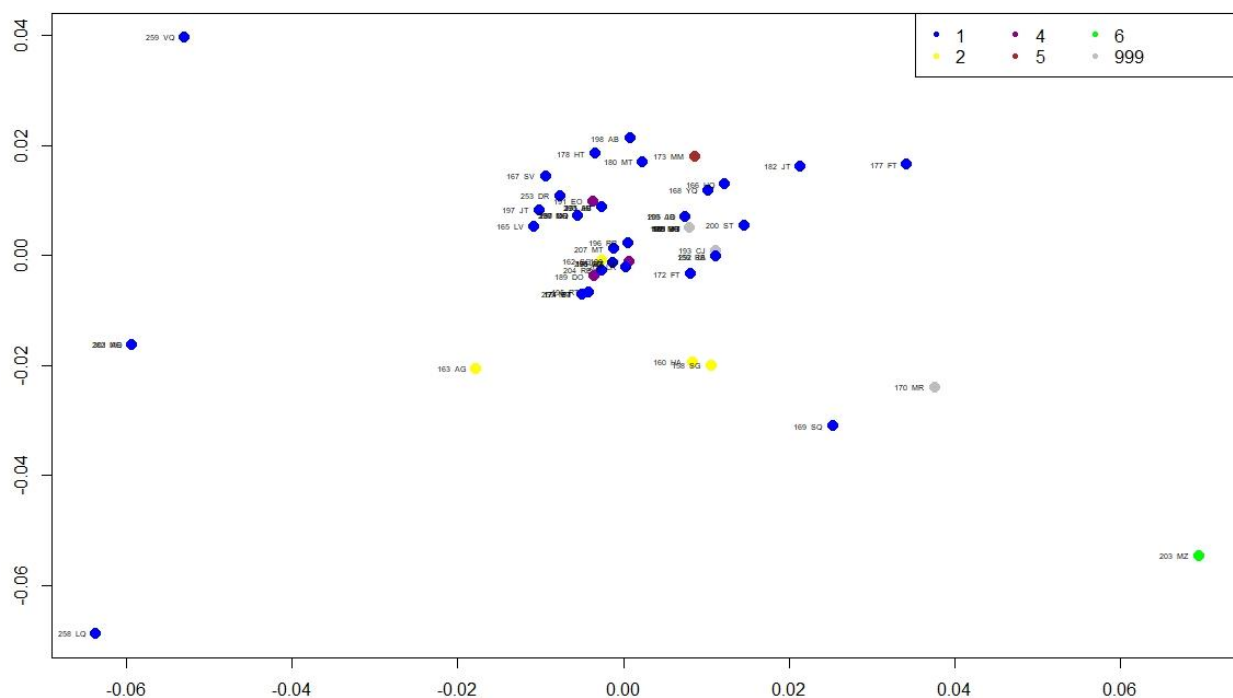
Los colores asignados a cada familia se especifican en la esquina inferior izquierda de la gráfica.

La separación del individuo 203 (mujer normal oriunda de Ricaurte no perteneciente a las familias con casos X frágil) de los demás, es esperada dado que porta dos cromosomas X disímiles a los que portan la expansión en FMR1 en la población. La separación del individuo 258 (mujer con premutación de la familia 1) de los demás, es llamativa, pues si bien, uno de sus cromosomas X es disímil al que porta la expansión, el otro sí la porta.

Se realizó un análisis de IBS (Figura 12) que muestra un comportamiento similar al ACP (Figura 11), separando a los individuos 203 y 258 del resto, pero además separa también a los individuos 259, 282 y 301 de los demás (Valores en x, superiores a -0.05)

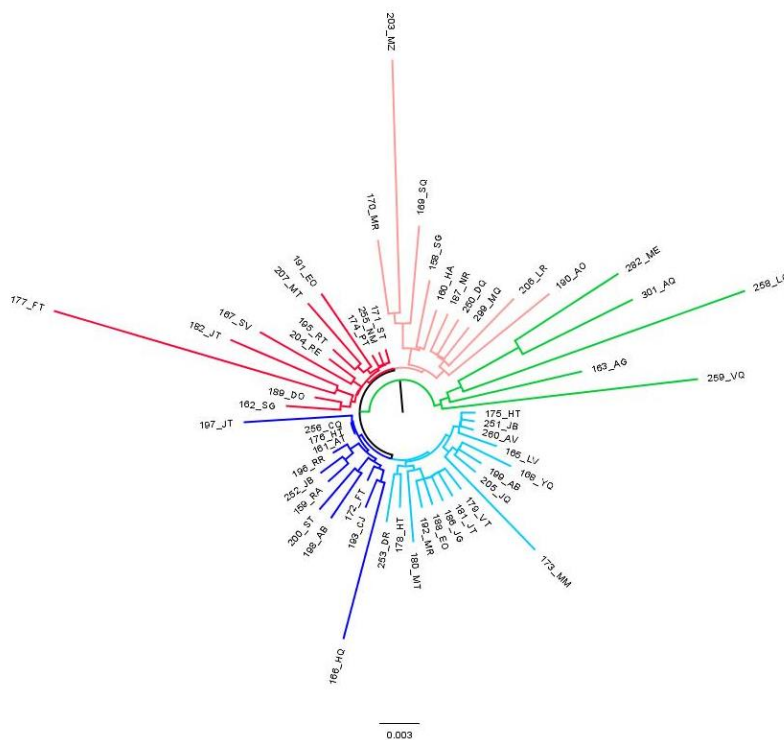
Adicionalmente, usando Neighbor joining, el método de agrupación de unión de vecinos fue calculado, obteniéndose un árbol filogenético que muestra a los individuos distribuidos en tres grupos (azul, rojo y verde) (Figura 13).

Figura 12: Análisis de Identidad por Estado para 236 SNPs-Indels del cromosoma X en 55 individuos.



Los colores asignados a cada familia se especifican en la esquina superior derecha de la gráfica.

Figura 13: Árbol filogenético Neighbor-Joining (NJ) basado en 236 SNPs-Indels del cromosoma X para los 55 individuos.



Determinación de distancia genética con poblaciones del proyecto 1000 genomas

Posteriormente, las matrices fueron concatenadas con los datos disponibles de 1000 genomas (Sudmant et al., 2010) con 2504 individuos de diferentes orígenes geográficos (26 orígenes). Se empleó el conjunto de datos de 1000 genomas (Gibbs et al., 2015) y 2504 individuos, cuyos datos resumen la población mundial (Descarga en línea, http://ftp.1000genomes.ebi.ac.uk/vol1/ftp/release/20130502/ALL.chrX.phase3_shapeit2_mvncall_integrated_v1b.20130502.genotypes.vcf.gz) (Tabla 17).

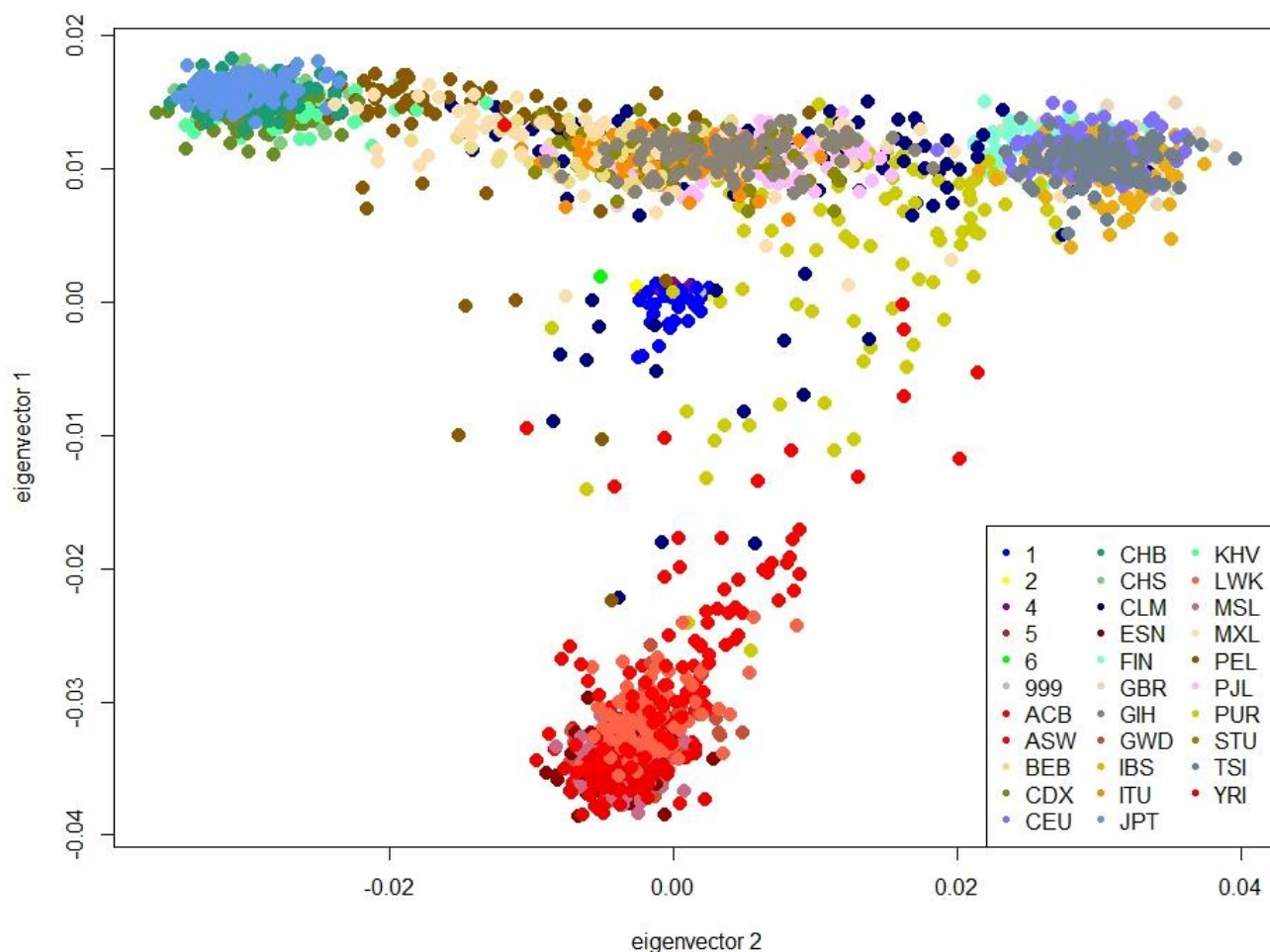
Tabla 17: Identificación de las poblaciones del proyecto 1000 genomas utilizadas en el análisis.

ID	No. Individuos	Población
ACB	96	African Caribbean in Barbados
ASW	61	African Ancestry in Southwest US
ESN	99	Esan in Nigeria
GWD	113	Gambian in Western Division, The Gambia
LWK	99	Luhya in Webuye, Kenya
MSL	85	Mende in Sierra Leone
YRI	108	Yoruba in Ibadan, Nigeria
CLM	94	Colombian in Medellin, Colombia
MXL	64	Mexican Ancestry in Los Angeles, California
PEL	85	Peruvian in Lima, Peru
PUR	104	Puerto Rican in Puerto Rico
CDX	93	Chinese Dai in Xishuangbanna, China
CHB	103	Han Chinese in Beijing, China
CHS	105	Southern Han Chinese, China
JPT	104	Japanese in Tokyo, Japan
KHV	99	Kinh in Ho Chi Minh City, Vietnam
CEU	99	Utah residents with Northern and Western European ancestry
FIN	99	Finnish in Finland
GBR	91	British in England and Scotland
IBS	107	Iberian populations in Spain
TSI	107	Toscani in Italy
BEB	86	Bengali in Bangladesh
GIH	103	Gujarati Indian in Houston, TX
ITU	102	Indian Telugu in the UK
PJL	96	Punjabi in Lahore, Pakistan
STU	102	Sri Lankan Tamil in the UK

Se realizó un ACP de los 95635 SNP-Indels para las seis familias estudiadas, que en el primer componente permitió explicar el 8.6% de la varianza total y separó los individuos de diferentes poblaciones de África (ACB, ASW, ESN, GWD, LWK, MSL, YRI) de los demás orígenes (Europa, América y Asia), y en el segundo componente explicó el 3.61% de la varianza. En general, los individuos de este estudio se ubicaron con los colombianos de 1000

genomas, más cercanos a Europa, América y Asia que a África (Figura 14). Esto se correlaciona con el hecho de que, tanto en el departamento de Antioquia (de donde son originarios los colombianos de 1000 genomas) como en Ricaurte y zonas aledañas, la población afrodescendiente es escasa.

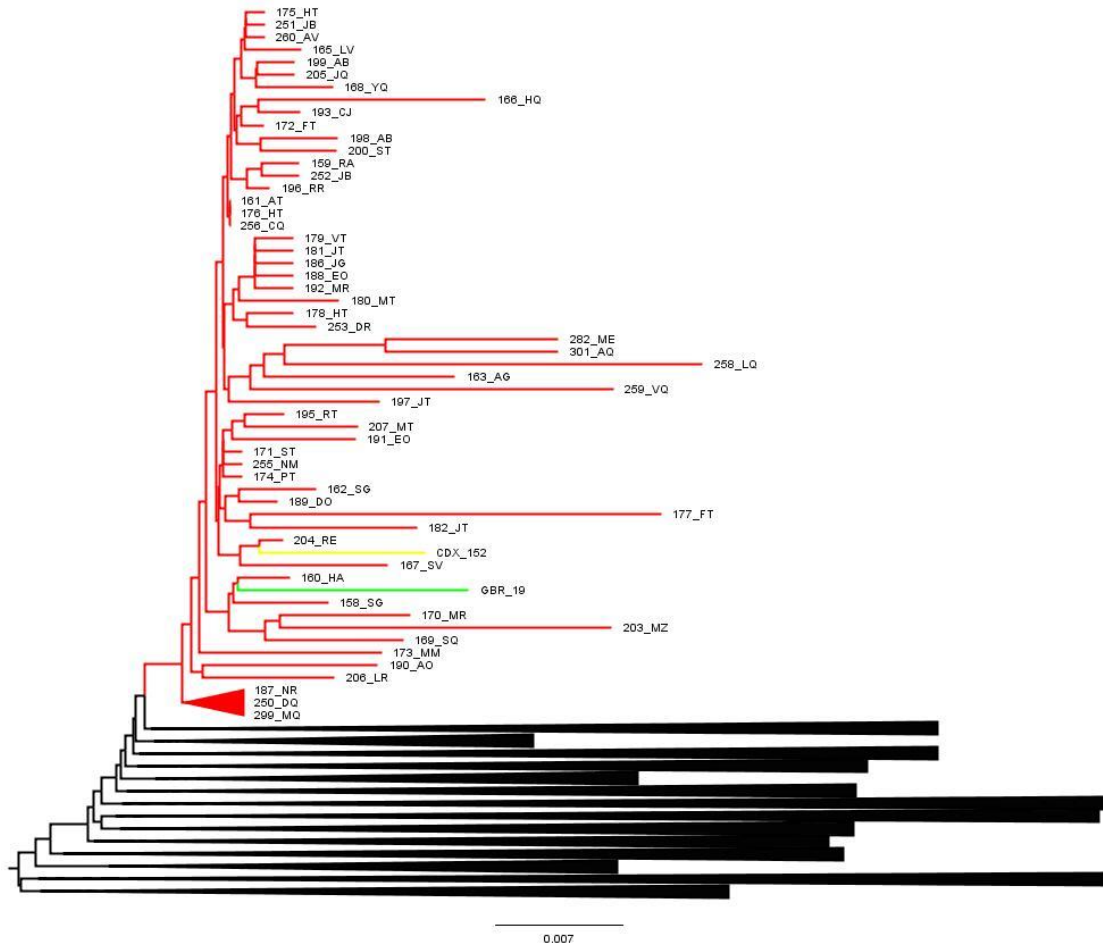
Figura 14: Análisis de Componentes principales para 95635 SNPs del cromosoma X sin filtrar de 55 individuos más 2504 individuos de 26 orígenes geográficos.



Los colores asignados a cada familia de Ricaurte y grupo poblacional de 1000 genomas se especifican en la esquina inferior derecha de la gráfica.

Se obtuvo un árbol filogenético usando Neighbor joining para la matriz concatenada con 2504 individuos y 236 SNPs-Indels de cromosoma X, en el que interesantemente se pudo observar que hay dos individuos de 1000 genomas, el CDX_152 (de China) y el GBR_19 (de Reino Unido) que se agrupan con los datos de este estudio, lo cual podría estar indicando cierto grado de ancestría en regiones específicas de Asia y Europa (Figura 15).

Figura 15: Árbol filogenético Neighbor-Joining (NJ) basado en 236 SNPs-Indels para 55 individuos más 2504 individuos de 26 orígenes geográficos.



Búsqueda de SNPs significativamente más frecuentes en los individuos con expansión en FMR1 de Ricaurte que presentan características fenotípicas específicas.

Se compararon las frecuencias de cada SNP en casos y controles.

Comparaciones realizadas

Se realizaron las siguientes comparaciones:

- 7 individuos con expansión (mutación o premutación) que presentan convulsiones VS. 47 individuos con expansión (mutación o premutación) que no presentan convulsiones.
- 4 mujeres con expansión (premutación) que presentan síndrome de falla ovárica prematura VS. 21 mujeres con expansión (premutación) con 40 años o más que no presentan síndrome de falla ovárica prematura. Se excluyeron todos los hombres y las mujeres menores de 40 años con menstruación presente.

- 4 individuos con expansión (premutación) que presentan síndrome de temblor-ataxia VS. 48 individuos con expansión (premutación) que no presentan síndrome de temblor-ataxia. Se excluyeron 3 casos dudosos.

El análisis de asociación se llevó a cabo empleando PLINK, la asociación entre SNP y un resultado binario (valor 1 = no afectado y valor 2 = afectado; 0 y -9 representan faltantes) se probó con la opción --assoc. Esta opción realiza una prueba de asociación X2 que no permite la inclusión de covariables.

En estudios con múltiples comparaciones se debe utilizar correcciones que permitan no incrementar la probabilidad de cometer error tipo I. El procedimiento de Bonferroni consiste en dividir el alfa por el número de comparaciones a estimar: si existen aproximadamente 10^6 variantes en el genoma, el umbral de valor de p corregida para todas las comparaciones sería $p = 0,05/10^6 = 5 \times 10^{-8}$. Se utilizó aplicó Bonferroni a las comparaciones aquí realizadas.

SNPs con frecuencia significativamente mayor en individuos de Ricaurte con expansión (mutación o premutación) que convulsionan

Se encontraron siete variantes, en seis genes, significativamente más frecuentes en los pacientes con expansión (premutación o mutación completa) que presentan convulsiones (Tabla 18), de las cuales solamente una pasó Bonferroni.

Tabla 18: SNPs significativamente más frecuentes en pacientes con expansión (premutación o mutación completa) que presentaban convulsiones.

C H R	SNP	BP	A 1	F A	F U	A 2	CH IS Q	P	O R	BO NF	GEN	Significancia clínica
10	rs1638630	2,77 E+07	T	0,2 14 3	0	C	20,72	5,32 E-06	N a N	0.0 654 7	PTCHD3, patched domain containing 3	Mutación de cambio de sentido no reportada en ClinVar https://www.ncbi.nlm.nih.gov/snp/rs1638630
7	rs1881121	944683	C	0,2 85 7	0,0 212 8	T	16,24	5,58 E-05	18,4	0.6 869	ADAP1, ArfGAP with dual PH domains 1	Variante intrónica no reportada en ClinVar https://www.ncbi.nlm.nih.gov/snp/rs1881121
6	rs55873330	1,61 E+08	G	0,2 85 7	0,0 217 4	A	15,86	6,84 E-05	18	0.8 411	LPAL2, lipoprotein(a) like 2, pseudogene	Variante intrónica no reportada en ClinVar https://www.ncbi.nlm.nih.gov/snp/rs55873330
2	rs6751781	1,71 E+08	T	0,3 57 1	0,0 425 5	G	15,79	7,09 E-05	12,5	0.8 725	MYO3B, myosin IIIB (plus strand)	Variante intrónica no reportada en ClinVar https://www.ncbi.nlm.nih.gov/snp/rs6751781
6	rs1009324	1,61 E+08	A	0,2 85 7	0,0 319 1	G	12,95	3,20 E-04	12,1 3	1.0	LPAL2, lipoprotein(a) like 2, pseudogene	Variante intrónica no reportada en ClinVar https://www.ncbi.nlm.nih.gov/snp/rs1009324
5	rs259096	7,91 E+07	G	0,4 16 7	0,0 744 7	A	12,41	4,26 E-04	8,87 8	1.0	CMYA5, cardiomyopathy associated 5	Variante intrónica no reportada en ClinVar https://www.ncbi.nlm.nih.gov/snp/rs259096
19	rs2937157	2,16 E+07	G	0,5	0,1 277	A	11,65	6,41 E-04	6,83 3	1.0	ZNF738, zinc finger protein 738	Variante intrónica no reportada en ClinVar https://www.ncbi.nlm.nih.gov/snp/rs2937157

La variante en rojo superó Bonferroni.

La tabla 19 resume las características de los seis genes que albergan las siete variantes encontradas con mayor frecuencia en los individuos expandidos que presentan convulsiones. De cada gen se especifican los tejidos de expresión, la función del producto y las enfermedades u otras características fenotípicas asociadas:

Tabla 19: Características de los genes que albergan las variantes encontradas con mayor frecuencia en los individuos expandidos que presentan convulsiones.

Gen (locus)	Tejidos de expresión	Función del producto	Fenotipos y/o enfermedades asociadas
PTCHD3 (10p12.1)	Testículo, omento, páncreas, mama, células germinales testiculares, sistema nervioso, pulmones. https://www.genecards.org/cgi-bin/carddisp.pl?gene=PTCHD3	Puede jugar un papel en el desarrollo o función de espermatozoides https://www.genecards.org/cgi-bin/carddisp.pl?gene=PTCHD3	Escoliosis idiopática de la adolescencia, radio cintura:cadera, asma de inicio en la niñez, respuesta a lamotrigina, erupción maculopapular, epilepsia, índice de masa corporal, nivel de insulina sanguínea en ayuno https://www.genecards.org/cgi-bin/carddisp.pl?gene=PTCHD3
ADAP1 (7p22.3)	Cerebro embrionario (hipocampo, cerebelo, corteza cerebral, ventrículos laterales), testículo embrionario (células de Leydig), sangre, corteza frontal, médula espinal, estómago, intestino, pulmón https://www.genecards.org/cgi-bin/carddisp.pl?gene=ADAP1	Proteína activadora de GTPasas para la familia de factor de ribosilación de ADP (probable). Se une a fosfatidilinositol 3,4,5-trisfosfato (PtdInsP3) e inositol 1,3,4,5-tetrakisfosfato (InsP4). https://www.genecards.org/cgi-bin/carddisp.pl?gene=ADAP1	Insomnio, venas varicosas, valor de presión arterial media, fibrilación atrial, valor de presión sanguínea diastólica https://www.genecards.org/cgi-bin/carddisp.pl?gene=ADAP1
LPAL2 (6q25.3)	Hígado, corazón. https://www.genecards.org/cgi-bin/carddisp.pl?gene=LPA L2	Pseudogen similar al gen Lp(a), pero que produce transcritos truncados. https://www.genecards.org/cgi-bin/carddisp.pl?gene=LPA L2	Niveles de lipoproteína A, niveles de apolipoproteína A1, enfermedad arterial coronaria, carcinoma de próstata. https://www.genecards.org/cgi-bin/carddisp.pl?gene=LPA L2
MYO3B (2q31.1)	Riñón (corteza), testículo, arteria (tibial), mama, retina. https://www.genecards.org/cgi-bin/carddisp.pl?gene=MYO3B	Probable función motora basada en actina. Actividad de protein quinasa. Requerida para el	Grado de unión del lóbulo de la oreja a la cabeza, radio cintura:cadera, habilidad matemática,

	bin/carddisp.pl?gene=MYO3B	desarrollo coclear y la audición. https://www.genecards.org/cgi-bin/carddisp.pl?gene=MYO3B	metilación del ADN, calvicie. https://www.genecards.org/cgi-bin/carddisp.pl?gene=MYO3B
CMYA5 (5q14.1)	Músculo esquelético, corazón (ventrículo izquierdo, atrios), líquido cefalorraquídeo, hígado. https://www.genecards.org/cgi-bin/carddisp.pl?gene=CMYA5	Puede servir como proteína de anclaje que media la compartimentalización subcelular de protein quinasa A vía unión a PRKAR2A. puede funcionar como represor de la actividad transcripcional mediada por calcineurina. Puede atenuar la habilidad de calcineurina para inducir el programa génico de fibras lentas en el músculo y modular negativamente la regeneración del músculo esquelético. Tiene papel en el ensamblaje de clústers de receptor de rianodina en el músculo estriado. https://www.genecards.org/cgi-bin/carddisp.pl?gene=CMYA5	Obesidad, respuesta a diisocianato, asma, seropositividad para antígenos de superficie del virus de hepatitis B, infección crónica por virus de hepatitis B, niveles de selenio. https://www.genecards.org/cgi-bin/carddisp.pl?gene=CMYA5
ZNF738 (19p12)	Suero. https://www.genecards.org/cgi-bin/carddisp.pl?gene=ZNF738	Puede estar involucrado en regulación de la transcripción. https://www.genecards.org/cgi-bin/carddisp.pl?gene=ZNF738	Ninguno.

Entre las variantes encontradas con mayor frecuencia, la única que pasó Bonferroni fue la encontrada en el gen PTCHD3, que interesantemente se expresa en el sistema nervioso central y ha sido previamente asociado a epilepsia. (<https://www.genecards.org/cgi-bin/carddisp.pl?gene=PTCHD3>)

Aunque el resto de variantes no pasaron Bonferroni, llama la atención que una de ellas se encuentra en el gen ADAP1 que se expresa en cerebro. (<https://www.genecards.org/cgi-bin/carddisp.pl?gene=ADAP1>)

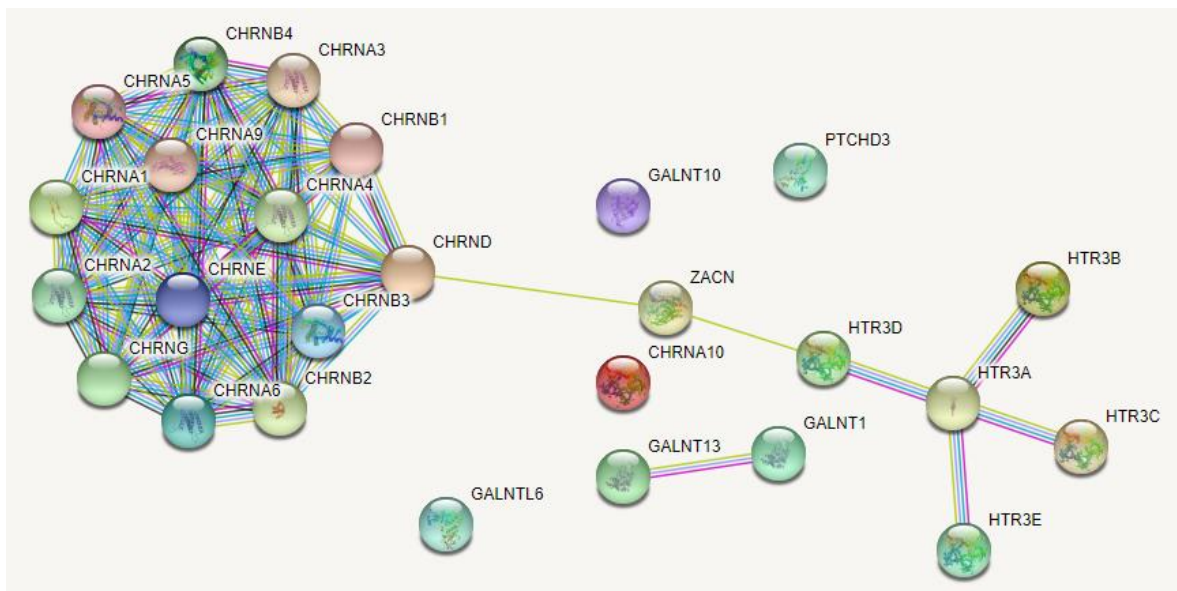
A continuación, se presenta una revisión más detallada de los genes que albergan SNPs significativamente más frecuentes en pacientes con expansión (premutación o mutación completa) que convulsionan, y que, por su expresión, procesos en que se involucra la red a la que pertenecen o fenotipos asociados, podrían asociarse a la presencia de convulsiones.

Gen PTCHD3

Este gen codifica para una proteína que a nivel celular se localiza principalmente en la membrana plasmática, aunque también se encuentra en el espacio extracelular, mitocondria, núcleo y citoplasma. (<https://www.genecards.org/cgi-bin/carddisp.pl?gene=PTCHD3>)

La red de interacciones de 25 componentes arrojada por String para este gen muestra que las proteínas involucradas tienen más interacciones entre ellas que las esperadas para un grupo aleatorio de proteínas, lo cual indica que biológicamente tienen conexión, al menos parcial, como grupo (<https://version11.string-db.org/cgi/network.pl?taskId=jO7sFJ0BS0zF>) (Figura 16).

Figura 16: Red de interacciones moleculares de 25 componentes para el gen PTCHD3. Interacciones encontradas: 98. Interacciones esperadas: 1.



Fuente: <https://version11.string-db.org/cgi/network.pl?taskId=jO7sFJ0BS0zF>

Al revisar las implicaciones funcionales de esta red, se encuentra que se ha relacionado con múltiples funciones y/o vías moleculares, entre las cuales las siguientes podrían tener relación con el fenotipo de convulsiones: (<https://version11.string-db.org/cgi/network.pl?taskId=jO7sFJ0BS0zF>)

- Actividad de canal iónico activado por transmisor.
- Actividad de canal catiónico activado por ligando.
- Actividad de canal selectivo de cationes activado por acetilcolina.

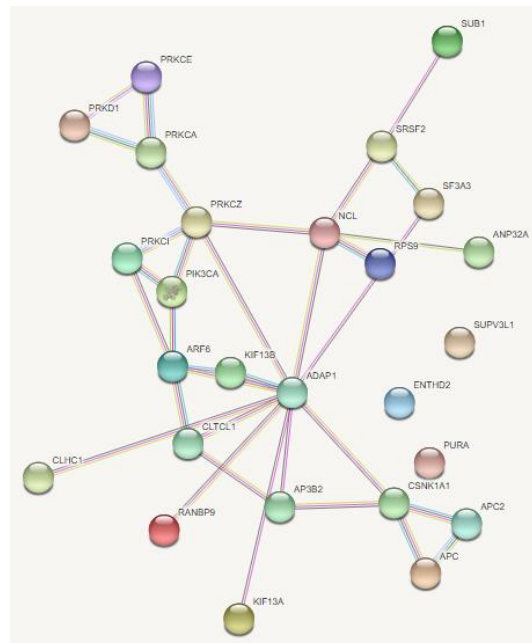
- Actividad de receptor de acetilcolina.
- Actividad de receptor de señalización transmembrana.
- Unión a acetilcolina.
- Actividad de canal selectivo de catión activado por serotonina.
- Unión a neurotransmisores.
- Unión a cationes.
- Actividad de canal de calcio.
- Vía de interacción ligando neuroactivo-receptor.
- Vía de sinapsis serotoninérgica.
- Vía de sinapsis colinérgica.
- Regulación de la morfogénesis de dendritas.

GEN ADAP1

Este gen codifica para una proteína que a nivel celular se localiza principalmente en membrana plasmática, núcleo y citoplasma, aunque también se encuentra en citoesqueleto. <https://www.genecards.org/cgi-bin/carddisp.pl?gene=ADAP1>

La red de interacciones de 25 componentes arrojada por String para este gen muestra que las proteínas involucradas tienen más interacciones entre ellas que las esperadas para un grupo aleatorio de proteínas, lo cual indica que biológicamente tienen conexión, al menos parcial, como grupo (<https://version11.string-db.org/cgi/network.pl?taskId=A4NDcGmdvavz>) (Figura 17).

Figura 17: Red de interacciones moleculares de 25 componentes para el gen ADAP1. Interacciones encontradas: 33. Interacciones esperadas: 7.



Fuente: <https://version11.string-db.org/cgi/network.pl?taskId=A4NDcGmdvavz>

Al revisar las implicaciones funcionales de esta red, se encuentra que se ha relacionado con múltiples funciones y/o vías moleculares, entre las cuales las siguientes podrían tener relación con el fenotipo de convulsiones: <https://version11.string-db.org/cgi/network.pl?taskId=A4NDcGmdvavz>

- Vía de la sinapsis colinérgica.
- Regulación negativa y positiva de procesos biológicos.
- Regulación de transducción de señal.
- Desarrollo del sistema nervioso central.
- Desarrollo de las proyecciones neuronales.
- Vía de señalización de guía de axones.

SNPs con frecuencia significativamente mayor en mujeres de Ricaurte con expansión (premutación) que presentan falla ovárica prematura

Se encontraron 12 variantes, en 9 genes, significativamente más frecuentes en las mujeres con expansión (premutación) que presentan falla ovárica prematura (Tabla 20), de las cuales solamente 3 pasaron Bonferroni

Tabla 20: SNPs significativamente más frecuentes en mujeres con expansión (premutación) que presentaban falla ovárica prematura.

CH R	SN P	BP	A 1	F_A	F_U	A2	CH ISQ	P	OR	BO NF	GEN	Significancia clínica
9	rs7033976	7,32E+07	C	0,5	0	T	22,4	2,21E-06	Na N	0,02659	TRPM3, transient receptor potential cation channel subfamily M member 3	Variante intrónica no reportada en ClinVar https://www.ncbi.nlm.nih.gov/snp/rs7033976
9	rs4744605	7,32E+07	A	0,5	0	G	22,4	2,21E-06	Na N	0,02659		Variante intrónica no reportada en ClinVar https://www.ncbi.nlm.nih.gov/snp/rs4744605
9	rs4744606	7,32E+07	A	0,5	0	C	22,4	2,21E-06	Na N	0,02659		Variante intrónica no reportada en ClinVar https://www.ncbi.nlm.nih.gov/snp/rs4744606
1	rs6696489	6162054	T	0,5	0,02381	C	15,58	7,89E-05	41	0,9477	CHD5, chromo domain helicase DNA binding protein 5	Variante de la región no trducida 3' no reportada en ClinVar https://www.ncbi.nlm.nih.gov/snp/rs6696489
6	rs3799276	4,78E+07	A	0,5	0,025	G	14,83	1,18E-04	39	1	PTCHD4, patched domain containing 4	Variante sinónima no reportada en ClinVar https://www.ncbi.nlm.nih.gov/snp/rs3799276
6	rs678312	4,69E+07	G	0,3333	0	A	14,61	1,32E-04	Na N	1	ADGRF5, G protein-coupled receptor F5	Variante sinónima no reportada en ClinVar https://www.ncbi.nlm.nih.gov/snp/rs678312
8	rs1549849	1,33E+08	T	0,3333	0	C	14,61	1,32E-04	Na N	1	KCNQ3, potassium voltage-gated channel subfamily Q member 3	Variante intrónica no reportada en ClinVar https://www.ncbi.nlm.nih.gov/snp/rs1549849
15	rs2305647	4,06E+07	C	0,3333	0	T	14,61	1,32E-04	Na N	1	PLCB2, phospholipase C beta 2	Variante intrónica no reportada en ClinVar

											https://www.ncbi.nlm.nih.gov/snp/rs2305647
15	rs1869902	4,06E+07	G	0,3333	0	C	14,61	1,32E-04	Na	1	Variante intrónica no reportada en ClinVar https://www.ncbi.nlm.nih.gov/snp/rs1869902
15	rs2854446	4,54E+07	C	0,3333	0	T	14,61	1,32E-04	Na	1	<u>SORD, sorbitol dehydrogenase</u> Variante transcrita no codificada no reportada en ClinVar. https://www.ncbi.nlm.nih.gov/snp/rs2854446
20	rs2206389	9453567	T	0,3333	0	C	14,61	1,32E-04	Na	1	<u>PLCB4, phospholipase C beta 4</u> Variante intrónica no reportada en ClinVar https://www.ncbi.nlm.nih.gov/snp/rs2206389
22	rs2800960	1,91E+07	C	0,5	0,04762	T	11,51	6,91E-04	20	1	<u>DGCR2, DiGeorge syndrome critical region gene 2</u> Variante intrónica no reportada en ClinVar https://www.ncbi.nlm.nih.gov/snp/rs2800960

Las variantes en rojo superaron Bonferroni.

La tabla 21 resume las características de los 9 genes que albergan las 12 variantes encontradas con mayor frecuencia en las mujeres expandidas (premutación) con falla ovárica prematura. De cada gen se especifican los tejidos de expresión, la función del producto y las enfermedades u otras características fenotípicas asociadas:

Tabla 21: Características de los genes que albergan las variantes encontradas con mayor frecuencia en mujeres expandidas (premutación) con falla ovárica prematura.

Gen (locus)	Tejidos de expresión	Función del producto	Fenotipos y/o enfermedades asociadas
TRPM3 (<u>9q21.12-q21.13</u>)	Ojo embrionario (cristalino), corazón embrionario (tubo endocárdico), cerebro (cerebelo, plexos coroideos, locus coeruleus, tálamo posterior, sustancia negra), intestino embrionario, riñón (médula, región periglomerular), médula espinal, testículo, ovario, páncreas. https://www.genecards.org/cgi-bin/carddisp.pl?gene=trpm3	Canal de calcio que media la entrada constitutiva de calcio iónico. Su actividad se incrementa por reducción de la osmolaridad extracelular, por depleción de depósitos y por activación del receptor muscarínico. https://www.genecards.org/cgi-bin/carddisp.pl?gene=TRPM3	Ancho de distribución del componente plaquetario, habilidad matemática, volumen plaquetario medio, edad de menarca, logros educativos autorreportados. https://www.genecards.org/cgi-bin/carddisp.pl?gene=TRPM3
CHD5	Cerebro y tubo neural embrionario (tallo cerebral, médula espinal, corteza frontal, pares craneales X y VII, hipocampo), testículo embrionario, cerebro (corteza, pituitaria,	Proteína remodeladora de cromatina. Juega un papel en el desarrollo del sistema nervioso central activando genes que promueven la diferenciación terminal neuronal. Es supresor de tumor. Regula	Intervalo QT, nivel de zinc en sangre. https://www.genecards.org/cgi-bin/carddisp.pl?gene=CHD5

	cerebelo), líquido sinovial, vesícula biliar, glándula adrenal, testículo, páncreas. https://www.genecards.org/cgi-bin/carddisp.pl?gene=CHD5	negativamente genes de proliferación. Regula positivamente p53. Regula paso crítico de la condensación de la cromatina en la espermatogénesis. https://www.genecards.org/cgi-bin/carddisp.pl?gene=CHD5	
PTCHD4	Tubo neural y cerebro embrionario (tallo cerebral). https://www.genecards.org/cgi-bin/carddisp.pl?gene=PTCHD4	Puede actuar, por antagonización de los efectos SMO, como represor de la señalización canónica Hedgehog. https://www.genecards.org/cgi-bin/carddisp.pl?gene=PTCHD4	Habilidad matemática, nivel de ligando inductor de apoptosis relacionado con TNF, nivel de ácido graso poliinsaturado omega 6, grado de susceptibilidad a otitis en la niñez, percepción sensitiva de olores. https://www.genecards.org/cgi-bin/carddisp.pl?gene=PTCHD4
ADGRF5	Endotelio embrionario, mesodermo extraembrionario, tejido adiposo embrionario, células germinales y gónadas embrionarias, suero y plasma, sistema nervioso, médula ósea, pulmón, riñón. https://www.genecards.org/cgi-bin/carddisp.pl?gene=ADGRF5	Receptor con papel crítico en la homeostasis del surfactante pulmonar. Puede jugar un papel en el control de la función del adipocito. https://www.genecards.org/cgi-bin/carddisp.pl?gene=ADGRF5	Nivel de proteínas sanguíneas, conteo leucocitario, conteo de monocitos, volumen plaquetario medio, escoliosis idiopática del adolescente. https://www.genecards.org/cgi-bin/carddisp.pl?gene=ADGRF5
KCNQ3	Tubo neural embrionario, cerebro (corteza, ganglios basales, amígdala), jugo pancreático, corazón. https://www.genecards.org/cgi-bin/carddisp.pl?gene=KCNQ3	Se asocia con KCNQ2 o KCNQ5 para formar un canal de potasio importante en la regulación de la excitabilidad neuronal. https://www.genecards.org/cgi-bin/carddisp.pl?gene=KCNQ3	Convulsiones neonatales familiares benignas. https://omim.org/entry/121201 Nivel de interleuquina 8, medida del diámetro de la partícula pico de LDL, medida de la ingesta de grasa total, intervalo QT, calcificación de la arteria coronaria. https://www.genecards.org/cgi-bin/carddisp.pl?gene=KCNQ3

			bin/carddisp.pl?gene=KCNQ3
PLCB2	Hueso embrionario, bazo embrionario, sangre, bazo, hígado, sistema nervioso, intestino, pulmón. https://www.genecards.org/cgi-bin/carddisp.pl?gene=PLCB2	Producción de diacilglicerol e inositol trifosfato. Crítica en transducción de señales de quimioatracción. https://www.genecards.org/cgi-bin/carddisp.pl?gene=PLCB2	Deficiencia de fosfolipasa C beta 2 plaquetaria. https://omim.org/entry/604114 Radio FEV/FEC, presión arterial sistólica, capacidad vital, niveles hormonales, diabetes mellitus 2. https://www.genecards.org/cgi-bin/carddisp.pl?gene=PLCB2
SORD	Riñón embrionario, hígado embrionario, próstata embrionaria, tiroides embrionaria, hígado, próstata, tiroides, epitelio nasal, epidídimo, sistema nervioso, sangre, riñón, pulmón, estómago. https://www.genecards.org/cgi-bin/carddisp.pl?gene=sord	Convierte sorbitol a fructosa. Hace parte de la vía del poliol que es importante en la fisiología del espermatozoide. Puede tener papel como fuente de energía para movilidad del espermatozoide. https://www.genecards.org/cgi-bin/carddisp.pl?gene=sord	Catarata congénita. https://omim.org/entry/182500 . Ancho de distribución eritrocitario, escoliosis idiopática del adolescente, enfermedad de Alzheimer. https://www.genecards.org/cgi-bin/carddisp.pl?gene=sord
PLCB4	Cerebro y tubo neural embrionario (hipotálamo, pituitaria, tallo cerebral, corteza cerebral), sangre embrionaria, bazo embrionario, tubo digestivo embrionario, cerebro (cerebelo), glándulas salivares, retina, ojo, pulmón. https://www.genecards.org/cgi-bin/carddisp.pl?gene=PLCB4	Producción de diacilglicerol e inositol trifosfato. Papel en transducción de señales en retina. https://www.genecards.org/cgi-bin/carddisp.pl?gene=PLCB4	Síndrome auriculocondilar 2. https://omim.org/entry/614669 Conteo de neutrófilos, índice de masa corporal, esquizofrenia, aceleración de la búsqueda inicial, trastorno bipolar, trastorno esquizoafectivo, descanso diurno. https://www.genecards.org/cgi-bin/carddisp.pl?gene=PLCB4
DGCR2	Sangre embrionaria (granulocitos), riñón fetal, hueso, cerebro, corazón, hígado, riñón, médula ósea, pulmón, páncreas. https://www.genecards.org/cgi-bin/carddisp.pl?gene=DGCR2	Receptor de adhesión putativo. Puede estar involucrado en interacciones célula-célula o célula-matriz. Requerido para diferenciación y migración celular. https://www.genecards.org/cgi-bin/carddisp.pl?gene=DGCR2	Síndrome de DiGeorge. https://omim.org/entry/600594 Metabolitos en sangre, niveles de proteína, niveles de aminoácidos, conteo de leucocitos. https://www.genecards.org/cgi-bin/carddisp.pl?gene=DGCR2

		bin/carddisp.pl?gene=DGCR2	bin/carddisp.pl?gene=DGCR2
--	--	---	---

Entre las variantes encontradas con mayor frecuencia, las únicas que pasaron Bonferroni fueron las 3 encontradas en el gen TRPM3, que interesantemente se expresa en ovario y ha sido previamente asociado a la edad de la menarca. <https://www.genecards.org/cgi-bin/carddisp.pl?gene=TRPM3>

Aunque el resto de variantes no pasaron Bonferroni, llaman la atención 4 de ellas, una por encontrarse en el gen CHD5 que se expresa en la glándula pituitaria, <https://www.genecards.org/cgi-bin/carddisp.pl?gene=CHD5> otra por encontrarse en el gen ADGRF5 que se expresa en las gónadas embrionarias, <https://www.genecards.org/cgi-bin/carddisp.pl?gene=ADGRF5> y otras 2 por encontrarse en el gen PLCB que se expresa en el hipotálamo y la pituitaria embrionarios. <https://www.genecards.org/cgi-bin/carddisp.pl?gene=PLCB2>

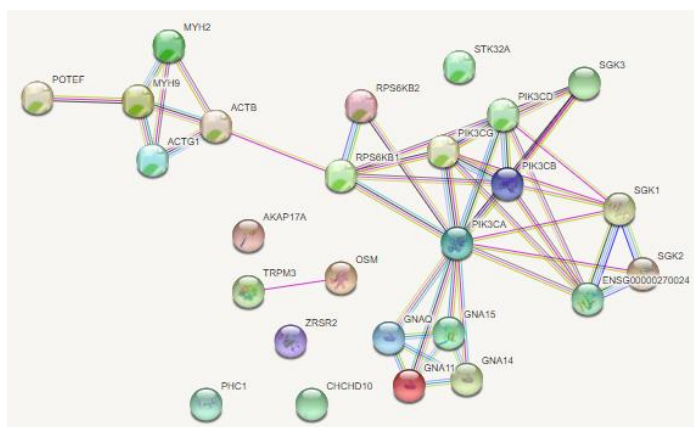
A continuación, se presenta una revisión más detallada de los genes que albergan SNPs significativamente más frecuentes en mujeres con expansión (premutación) que presentan falla ovárica prematura, y que, por su expresión, procesos en que se involucra la red a la que pertenecen o fenotipos asociados, podrían asociarse a la presencia de falla ovárica prematura.

Gen TRPM3

El producto de este gen se localiza a nivel celular principalmente en membrana plasmática, aunque también se encuentra en espacio extracelular, retículo endoplásmico y citoplasma. (<https://www.genecards.org/cgi-bin/carddisp.pl?gene=TRPM3>)

La red de interacciones de 25 componentes arrojada por String para este gen muestra que las proteínas involucradas tienen más interacciones entre ellas que las esperadas para un grupo aleatorio de proteínas, lo cual indica que biológicamente tienen conexión, al menos parcial, como grupo. <https://version11.string-db.org/cgi/network.pl?taskId=BvwJLV68hxbZ> (Figura 18).

Figura 18: Red de interacciones moleculares de 25 componentes para el gen TRPM3. Interacciones encontradas: 47. Interacciones esperadas: 9.



Fuente: <https://version11.string-db.org/cgi/network.pl?taskId=BvwJLV68hxbZ>

Al revisar las implicaciones funcionales de esta red, se encuentra que se ha relacionado con múltiples funciones y/o vías moleculares, entre las cuales las siguientes podrían tener relación con el fenotipo de falla ovárica prematura: <https://version11.string-db.org/cgi/network.pl?taskId=BvwJLV68hxbZ>

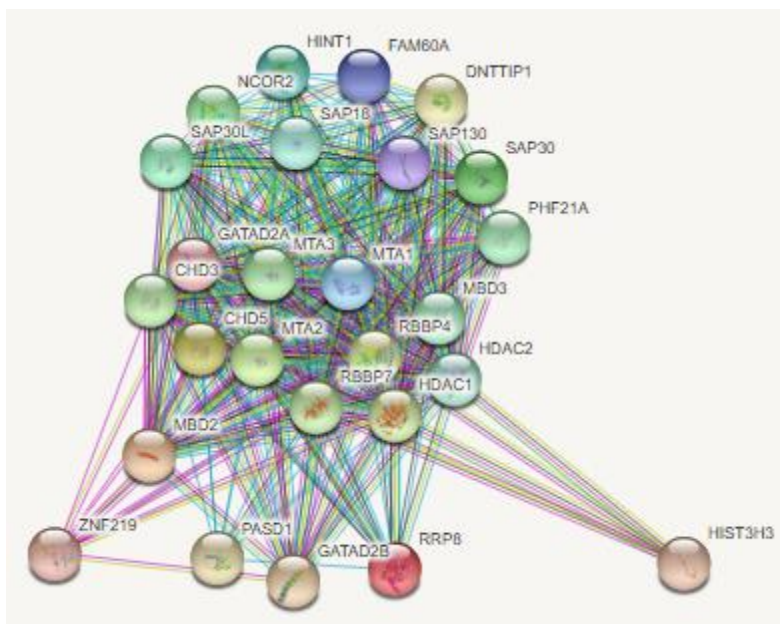
- Respuestas celulares a estímulos hormonales.
- Ritmo circadiano.

Gen CHD5

El producto de este gen se localiza a nivel celular principalmente en el núcleo y el citoplasma. <https://www.genecards.org/cgi-bin/carddisp.pl?gene=CHD5>

La red de interacciones de 25 componentes arrojada por String para este gen muestra que las proteínas involucradas tienen más interacciones entre ellas que las esperadas para un grupo aleatorio de proteínas, lo cual indica que biológicamente tienen conexión, al menos parcial, como grupo. <https://version11.string-db.org/cgi/network.pl?taskId=IuqW7cgl3Qgq> (figura 19)

Figura 19: Red de interacciones moleculares de 25 componentes para el gen CHD5. Interacciones encontradas: 260. Interacciones esperadas: 9



Fuente: <https://version11.string-db.org/cgi/network.pl?taskId=IuqW7cgl3Qgq>

Al revisar las implicaciones funcionales de esta red, se encuentra que se ha relacionado con múltiples funciones y/o vías moleculares, entre las cuales las siguientes podrían tener relación con el fenotipo de falla ovárica prematura: <https://version11.string-db.org/cgi/network.pl?taskId=IuqW7cgl3Qgq>

- 77

- Procesos del desarrollo involucrados en reproducción.
- Diferenciación del ovocito.
- Desarrollo de estructuras reproductivas.
- Desarrollo gonadal.
- Procesos rítmicos.
- Generación de gametos.
- Vía de señalización de la relaxina.

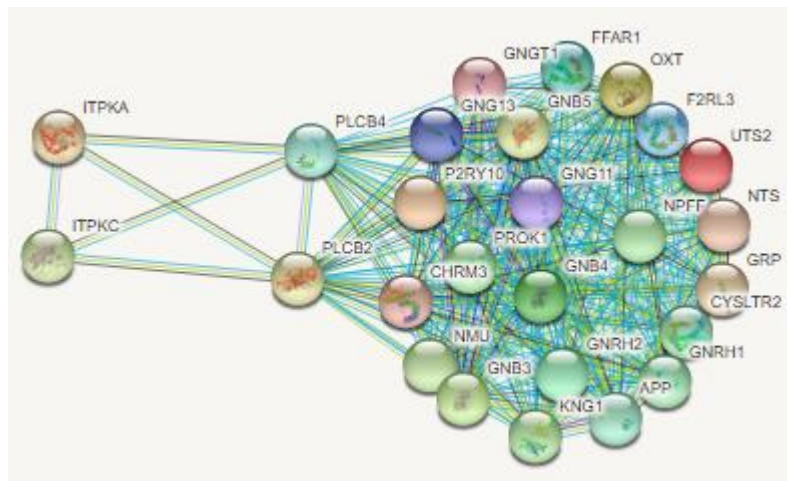
Gen PLCB4

El producto de este gen se localiza a nivel celular principalmente en citosol, aunque también se encuentra en núcleo, retículo endoplásmico y membrana plasmática.

<https://www.genecards.org/cgi-bin/carddisp.pl?gene=PLCB4>

La red de interacciones de 25 componentes arrojada por String para este gen muestra que las proteínas involucradas tienen más interacciones entre ellas que las esperadas para un grupo aleatorio de proteínas, lo cual indica que biológicamente tienen conexión, al menos parcial, como grupo. <https://version11.string-db.org/cgi/network.pl?taskId=ANTUw63c8cF5> (Figura 21).

Figura 21: Red de interacciones moleculares de 25 componentes para el gen PLCB4. Interacciones encontradas: 281. Interacciones esperadas: 24.



Fuente: <https://version11.string-db.org/cgi/network.pl?taskId=ANTUw63c8cF5>

Al revisar las implicaciones funcionales de esta red, se encuentra que se ha relacionado con múltiples funciones y/o vías moleculares, entre las cuales las siguientes podrían tener relación con el fenotipo de falla ovárica prematura: <https://version11.string-db.org/cgi/network.pl?taskId=ANTUw63c8cF5>

- Regulación de secreción de hormonas peptídicas.
- Comportamiento de apareamiento.

- Regulación de niveles hormonales.
- Embarazo.
- Actividad hormonal.
- Actividad de hormonas neuropeptídicas.
- Entrenamiento circadiano.
- Vía de señalización de la relaxina.
- Vía de señalización de GnRH.
- Vía de liberación de oxitócica.
- Síntesis y secreción de cortisol.
- Vía de señalización de estrógenos.

SNPs con frecuencia significativamente mayor en individuos de Ricaurte con expansión (premutación) que presentan temblor-ataxia

Se encontraron 15 variantes, en 11 genes, significativamente más frecuentes en los pacientes con expansión (premutación) que presentan síndrome de temblor-ataxia (Tabla 22), de las cuales ninguna pasó Bonferroni.

Tabla 22: SNPs significativamente más frecuentes en pacientes con expansión (premutación) que presentaban temblor-ataxia.

CH R	SNP	BP	A 1	F_ A	F_U	A 2	CHI SQ	P	OR	BO NF	GEN	Link
1	rs1277016	1,09E+08	G	0,375	0,03191	A	15,67	7,52E-05	18,2	0,9256	<u>STXBP3, syntaxin binding protein 3</u>	Variante intrónica no reportada en ClinVar https://www.ncbi.nlm.nih.gov/snp/rs1277016
4	rs4956033	1,09E+08	C	0,25	0,01064	T	14,8	1,20E-04	31	1	<u>HADH, hydroxyacyl-CoA dehydrogenase</u>	Variante intrónica no reportada en ClinVar https://www.ncbi.nlm.nih.gov/snp/rs4956033
4	rs4956034	1,09E+08	A	0,25	0,01064	G	14,8	1,20E-04	31	1	<u>nase</u>	Variante intrónica no reportada en ClinVar https://www.ncbi.nlm.nih.gov/snp/rs4956034
7	rs38809	9,21E+07	C	0,25	0,01064	T	14,8	1,20E-04	31	1	<u>PEX1, peroxisomal biogenesis factor 1</u>	Variante intrónica reportada en CLinVar como benigna. https://www.ncbi.nlm.nih.gov/snp/rs38809
19	rs2081054	2,12E+07	C	0,5	0,08333	T	12,56	3,94E-04	11	1	<u>ZNF430, zinc finger protein 430</u>	Variante en región 5' no traducida no reportada en ClinVar https://www.ncbi.nlm.nih.gov/snp/rs2081054
19	rs7252277	2,12E+07	A	0,5	0,08333	G	12,56	3,94E-04	11	1		Variante intrónica no reportada en ClinVar https://www.ncbi.nlm.nih.gov/snp/rs7252277
1	rs4362024	6,46E+07	G	0,125	0	A	12,12	5,00E-04	NaN	1	<u>ROR1, receptor tyrosine kinase like</u>	Variante intrónica no reportada en ClinVar https://www.ncbi.nlm.nih.gov/snp/rs4362024

											orphan receptor 1	
3	rs9840317	1,19E+08	A	0,125	0	G	12,12	5,00E-04	Na	1	UPK1B, uroplakin 1B	Variante de cambio de sentido no reportada en ClinVar https://www.ncbi.nlm.nih.gov/snp/rs9840317
3	rs3796356	1,19E+08	T	0,125	0	C	12,12	5,00E-04	Na	1		Variante intrónica no reportada en ClinVar https://www.ncbi.nlm.nih.gov/snp/rs3796356
3	rs3796357	1,19E+08	C	0,125	0	T	12,12	5,00E-04	Na	1		Variante intrónica no reportada en ClinVar https://www.ncbi.nlm.nih.gov/snp/rs3796357
11	rs2156719	1,11E+08	G	0,125	0	A	12,12	5,00E-04	Na	1	BTG4, BTG anti-proliferation factor 4	Variante intrónica no reportada en ClinVar https://www.ncbi.nlm.nih.gov/snp/rs2156719
18	rs6507018	3,10E+07	G	0,125	0	A	11,87	5,72E-04	Na	1	CCDC178, coiled-coil domain containing 178	Variante intrónica no reportada en ClinVar https://www.ncbi.nlm.nih.gov/snp/rs6507018
18	rs637651	2688798	G	0,5	0,07447	A	11,35	7,56E-04	12,43	1	SMCHD1, structural maintenance of chromosomes flexible hinge domain containing 1	Variante intrónica no reportada en ClinVar https://www.ncbi.nlm.nih.gov/snp/rs637651
2	rs6436120	2,20E+08	C	0,375	0,05208	T	10,84	9,91E-04	10,92	1	NHEJ1, non-homologous end joining factor 1	Variante intrónica no reportada en ClinVar https://www.ncbi.nlm.nih.gov/snp/rs6436120
10	rs930509	5,45E+07	C	0,375	0,05208	G	10,84	9,91E-04	10,92	1	MBL2, mannose binding lectin 2	Variante intrónica no reportada en ClinVar https://www.ncbi.nlm.nih.gov/snp/rs930509

La tabla 23 resume las características de los 11 genes que albergan las 15 variantes encontradas con mayor frecuencia en los individuos expandidos (premutación) con temblor-ataxia. De cada gen se especifican los tejidos de expresión, la función del producto y las enfermedades u otras características fenotípicas asociadas:

Tabla 23: Características de los genes que albergan las variantes encontradas con mayor frecuencia en los individuos expandidos (premutación) con temblor-ataxia.

Gen asociado en este estudio (locus)	Tejidos de expresión	Función del producto	Fenotipos y/o enfermedades asociadas
STXBP3	Sangre embrionaria, células mononucleares,	Junto con STX4 y VAMP2, tiene papel en el	Infección por VIH-1, distribución de la grasa

	megacariocitos, plaquetas, sistema nervioso, pulmón, hígado, músculo, páncreas. https://www.genecards.org/cgi-bin/carddisp.pl?gene=STXB3	movimiento dependiente de insulina de GLUT 4 y en la fusión de las vesículas que contienen GLUT 4 con la superficie celular de los adipocitos. https://www.genecards.org/cgi-bin/carddisp.pl?gene=STXB3	corporal. https://www.genecards.org/cgi-bin/carddisp.pl?gene=STXB3
HADH	Corazón (ventrículo izquierdo), hígado, riñón, páncreas, músculo esquelético. https://www.genecards.org/cgi-bin/carddisp.pl?gene=HADH	Papel esencial en betaoxidación mitocondrial de ácido grasos de cadena corta, https://www.genecards.org/cgi-bin/carddisp.pl?gene=HADH	Deficiencia de 3 hdroxiacil Co A deshidrogenasa https://omim.org/entry/601609 Hipoglicemia hiperinsulinémica familiar tipo 4 https://omim.org/entry/601609
PEX1	Tiroides y paratiroides embrionarias, intestino embrionario (duodeno), ovario embrionario, tuba uterina embrionaria, esófago, médula ósea (células madre), pulmón, sistema nervioso, corazón. https://www.genecards.org/cgi-bin/carddisp.pl?gene=PEX1	Requerida para estabilidad de PEX5 e importe de proteína en la matriz peroxisomal. Anclada por PEX26 a la membrana peroxisomal. https://www.genecards.org/cgi-bin/carddisp.pl?gene=PEX1	Síndrome de Heimler, Desorden de la biogénesis de peroxisomas 1A (Zellweger), Desorden de la biogénesis de peroxisomas 1B (Refsum infantil). https://omim.org/entry/602136 Medida de la presión de pulso, presión sanguínea sistólica, presión sanguínea diastólica, talla, longevidad. https://www.genecards.org/cgi-bin/carddisp.pl?gene=PEX1
ZNF430	Sangre, sistema nervioso, corazón, intestino, glándulas exocrinas y endocrinas, ovario, útero, testículo, próstata. https://www.genecards.org/cgi-bin/carddisp.pl?gene=znf430	Puede estar involucrado en regulación transcripcional. https://www.genecards.org/cgi-bin/carddisp.pl?gene=znf430	Ninguno. https://www.genecards.org/cgi-bin/carddisp.pl?gene=znf430
ROR1	Sangre embrionaria, dientes embrionarios (esmalte), cordón umbilical (células	Es receptor de WNT5A que activa la vía de señalización NFkB y resulta en inhibición de la	Sordera autosómica recesiva 108. https://omim.org/entry/602336

	mesenquimáticas), arterias (tibial), suero, líquido sinovial, corazón, pulmones, riñón, sistema nervioso. https://www.genecards.org/cgi-bin/carddisp.pl?gene=ROR1	señalización WNT3A. Esencial para innervación en oído interno. https://www.genecards.org/cgi-bin/carddisp.pl?gene=RO R1	Nivel de proteínas sanguíneas, medida de asimetría facial, consumo de alcohol, diabetes mellitus tipo I, respuesta a broncodilatador, radio FEV/FEC. https://www.genecards.org/cgi-bin/carddisp.pl?gene=ROR1
UPK1B	Tejido adiposo embrionario, sangre embrionaria, tubo digestivo embrionario, riñón embrionario, hueso embrionario, vejiga, mama, orina. https://www.genecards.org/cgi-bin/carddisp.pl?gene=UPK1B	Componente de la membrana especializada de las células uroteliales. Puede tener papel en la fisiología del epitelio vesical, posiblemente en regulación de la permeabilidad o estabilidad de la membrana apical. https://www.genecards.org/cgi-bin/carddisp.pl?gene=UP K1B	Neutropenia, respuesta a gemcitabine, carcinoma pancreático. https://www.genecards.org/cgi-bin/carddisp.pl?gene=UPK1B
BTG4	Testículo embrionario, testículo, epitelio olfatorio. https://www.genecards.org/cgi-bin/carddisp.pl?gene=BTG4	Muestra marcada actividad antiproliferativa. Es capaz de inducir detenimiento en G1. https://www.genecards.org/cgi-bin/carddisp.pl?gene=BT G4	Eczema, conteo eosinofílico, linfoma Hodgkins, alergia, asma, logros educativo autorreportados. https://www.genecards.org/cgi-bin/carddisp.pl?gene=BTG4
CCDC178	Testículo, omento, corazón. https://www.genecards.org/cgi-bin/carddisp.pl?gene=CCD C178	Desconocida.	Nivel del ligando inductor de apoptosis relacionado con TNF, respuesta a diisocianato, asma, respuesta a broncodilatador, radio FEV/FEC, carcinoma de próstata, nivel de hidroxileucina. https://www.genecards.org/cgi-bin/carddisp.pl?gene=CCD C178
SMCHD1	Mesodermo intermedio, sangre embrionaria, gónada embrionaria, hueso embrionario, testículo embrionario, sangre, nodo	Promueve heterocromatinización en autosomas y en el cromosoma X. Papel en inactivación del X en	Síndrome de arrinia, atresia de coanas, microftalmia e hipogonadismo hipogonadotrófico,

	linfoide, células mononucleares periféricas, hígado, ojo, sistema nervioso. https://www.genecards.org/cgi-bin/carddisp.pl?gene=SMCHD1	mujeres. Reclutado al cromosoma x inactivo por Xist. Requerido para silenciamiento de clústers de loci autosómicos como el locus de DUX4. Tiene actividad de ATPasa. Localiza sitios de ruptura de doble cadena para promover su reparación. https://www.genecards.org/cgi-bin/carddisp.pl?gene=SMCHD1	distrofia muscular fascioescapohumeral 2. https://omim.org/entry/614982 Ancho de distribución eritrocitario, intervalo PR, seropositividad para tripanosoma cruzi, habilidad matemática, duración del QRS, nivel de síntomas asociados a la retirada de nicotina. https://www.genecards.org/cgi-bin/carddisp.pl?gene=SMCHD1
NHEJ1	Cerebro y tubo neural embrionarios, hueso embrionario, testículo, linfocitos B, células mononucleares de sangre periférica, médula ósea, piel. https://www.genecards.org/cgi-bin/carddisp.pl?gene=NHEJ1	Proteína de reparación del ADN involucrada en unión de extremos no homólogos (NHEJ) requerida para reparación de rupturas de la doble cadena y recombinación V(D)J. Puede servir como puente entre XRCC4 y otros factores NHEJ para permitir el acceso de XRCC4 a los extremos de ADN. Puede actuar con XRCC6/XRCC5 (ku) para estimular la unión de extremos romos y varios tipos de extremos no complementarios mediada por XRCC4. https://www.genecards.org/cgi-bin/carddisp.pl?gene=NHEJ1	Inmunodeficiencia combinada severa con microcefalia, retraso del crecimiento y sensibilidad a la radiación ionizante. https://omim.org/entry/611290 Calvicie, ancho de distribución plaquetario, talla, alopecia androgénica, conteo de reticulocitos. https://www.genecards.org/cgi-bin/carddisp.pl?gene=NHEJ1
MBL2	Testículo embrionario, hígado embrionario, hígado, vesícula biliar, sangre, médula ósea, riñón. https://www.genecards.org/cgi-bin/carddisp.pl?gene=MBL2	Lectina dependiente de calcio involucrada en la respuesta inmune innata. Se une a manosa, fucosa y N-acetilglucosamina de diferentes microorganismos y activa la vía del complemento vía lectina. Se une a células apoptóticas tardías, cuerpos apoptóticos y células necróticas, pero no	Deficiencia de lectina de unión a manosa. https://omim.org/entry/614372 Niveles de proteína sanguínea, densidad mineral ósea del hueso del talón, nivel de proteína C de unión a manosa, fracturas óseas, grado de sinofris. https://www.genecards.org

		a células apoptóticas tempranas, facilitando su captación por macrófagos. Puede unirse a ADN. https://www.genecards.org/cgi-bin/carddisp.pl?gene=MBL2	/cgi-bin/carddisp.pl?gene=MBL2
--	--	--	--

Ninguna de las variantes encontradas con mayor frecuencia en los individuos con temblor-ataxia pasó Bonferroni. Sin embargo, llaman la atención 7 de ellas, una por encontrarse en el gen STXBP3 que se expresa en sistema nervioso, <https://www.genecards.org/cgi-bin/carddisp.pl?gene=STXBP3> otra por encontrarse en el gen PEX1 que se expresa en sistema nervioso y se asocia a desórdenes peroxisomales monogénicos con manifestaciones neurológicas, <https://www.genecards.org/cgi-bin/carddisp.pl?gene=PEX1> otras 2 por encontrarse en el gen ZNF430 que se expresa en sistema nervioso central, <https://www.genecards.org/cgi-bin/carddisp.pl?gene=znf430> otra por encontrarse en el gen ROR1 que se expresa en el sistema nervioso central, <https://www.genecards.org/cgi-bin/carddisp.pl?gene=ROR1> otra por encontrarse en el gen SMCHD1 que se expresa en sistema nervioso central, <https://www.genecards.org/cgi-bin/carddisp.pl?gene=SMCHD1> y otra por encontrarse en el gen NHEJ1 que se expresa en cerebro y tubo neural embrionarios. <https://www.genecards.org/cgi-bin/carddisp.pl?gene=NHEJ1>

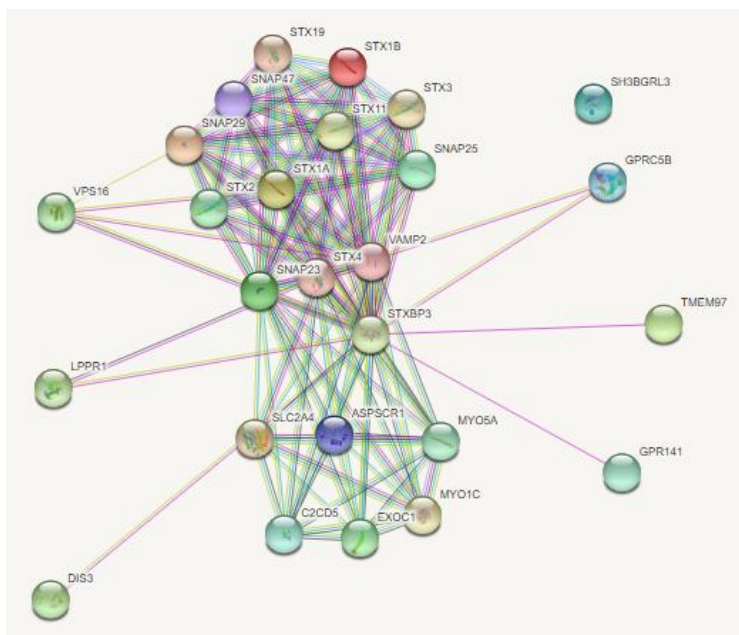
A continuación, se presenta una revisión más detallada de los genes que albergan SNPs significativamente más frecuentes en individuos con expansión (premutación) que presentan temblor-ataxia, y que, por su expresión, procesos en que se involucra la red a la que pertenecen o fenotipos asociados, podrían asociarse a la presencia de temblor-ataxia.

Gen STXBP3

El producto de este gen se localiza a nivel celular principalmente en membrana plasmática y citosol, aunque también se encuentra en núcleo, citoesqueleto, endosoma y lisosoma. <https://www.genecards.org/cgi-bin/carddisp.pl?gene=STXBP3>

La red de interacciones de 25 componentes arrojada por String para este gen muestra que las proteínas involucradas tienen más interacciones entre ellas que las esperadas para un grupo aleatorio de proteínas, lo cual indica que biológicamente tienen conexión, al menos parcial, como grupo. <https://version11.string-db.org/cgi/network.pl?taskId=7L1qdmynb7J2> (Figura 22).

Figura 22: Red de interacciones moleculares de 25 componentes para el gen STXBP3. Interacciones encontradas: 129. Interacciones esperadas: 4.



Fuente: <https://version11.string-db.org/cgi/network.pl?taskId=7L1qdmynb7J2>

Al revisar las implicaciones funcionales de esta red, se encuentra que se ha relacionado con múltiples funciones y/o vías moleculares, entre las cuales las siguientes podrían tener relación con el fenotipo de temblor-ataxia: <https://version11.string-db.org/cgi/network.pl?taskId=7L1qdmynb7J2>

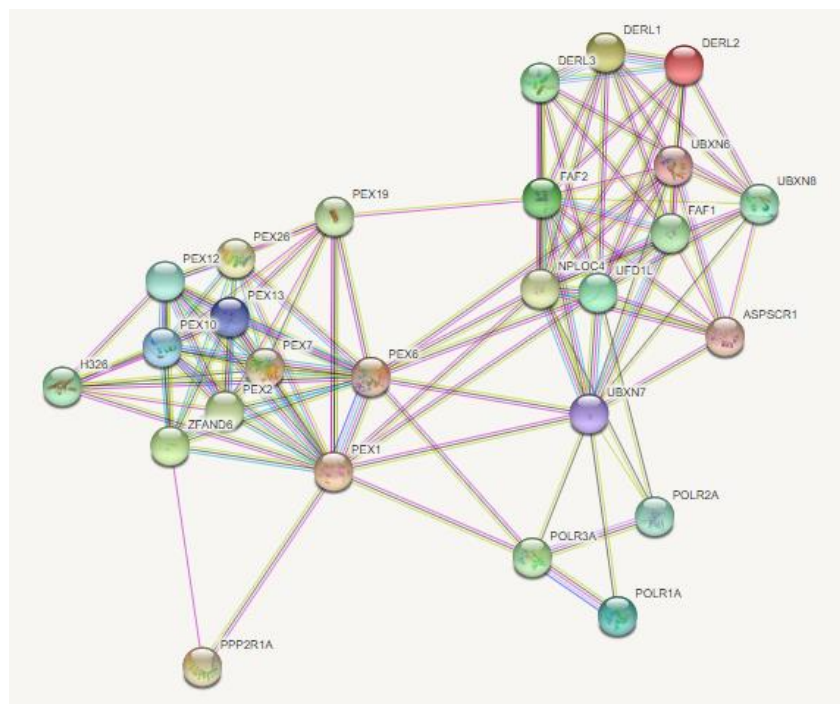
- Regulación positiva de la secreción de neurotransmisores.
- Regulación positiva del potencial postsináptico excitatorio.
- Ciclo de la vesícula sináptica.

Gen PEX1

El producto de este gen se localiza a nivel celular principalmente en peroxisoma y citosol, aunque también se encuentra en núcleo. <https://www.genecards.org/cgi-bin/carddisp.pl?gene=PEX1>

La red de interacciones de 25 componentes arrojada por String para este gen muestra que las proteínas involucradas tienen más interacciones entre ellas que las esperadas para un grupo aleatorio de proteínas, lo cual indica que biológicamente tienen conexión, al menos parcial, como grupo <https://version11.string-db.org/cgi/network.pl?taskId=k6AXjpyl7Gr0> (Figura 23).

Figura 23: Red de interacciones moleculares de 25 componentes para el gen PEX1. Interacciones encontradas: 117. Interacciones esperadas: 7.



Fuente: <https://version11.string-db.org/cgi/network.pl?taskId=k6AXjpql7Gr0>

Al revisar las implicaciones funcionales de esta red, se encuentra que se ha relacionado con múltiples funciones y/o vías moleculares, entre las cuales las siguientes podrían tener relación con el fenotipo de temblor-ataxia: <https://version11.string-db.org/cgi/network.pl?taskId=k6AXjpql7Gr0>

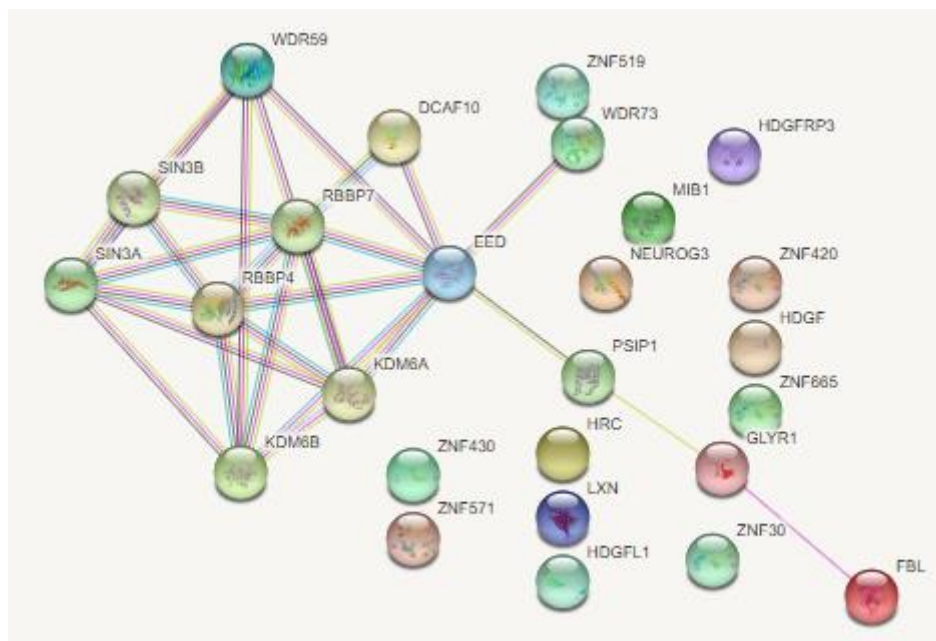
- Vías peroxisomales.

Gen ZNF430

El producto de este gen se localiza a nivel celular principalmente en núcleo. <https://www.genecards.org/cgi-bin/carddisp.pl?gene=ZNF430>

La red de interacciones de 25 componentes arrojada por String para este gen muestra que las proteínas involucradas tienen más interacciones entre ellas que las esperadas para un grupo aleatorio de proteínas, lo cual indica que biológicamente tienen conexión, al menos parcial, como grupo <https://version11.string-db.org/cgi/network.pl?taskId=nai8EpNLrOWX> (Figura 24).

Figura 24: Red de interacciones moleculares de 25 componentes para el gen ZNF430. Interacciones encontradas: 28. Interacciones esperadas: 4.



Fuente: <https://version11.string-db.org/cgi/network.pl?taskId=nai8EpNLrOWX>

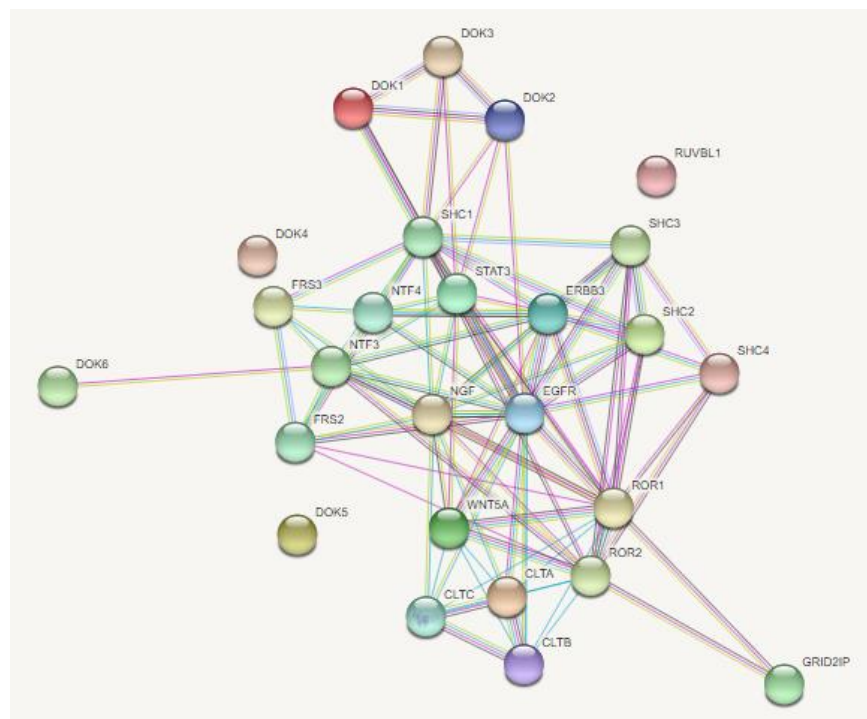
Al revisar las implicaciones funcionales de esta red, se encuentra que se ha relacionado con múltiples funciones y/o vías moleculares de control epigenético. <https://version11.string-db.org/cgi/network.pl?taskId=nai8EpNLrOWX>

Gen ROR1

El producto de este gen se localiza a nivel celular principalmente en membrana plasmática, aunque también se encuentra en citoesqueleto y espacio extracelular. <https://www.genecards.org/cgi-bin/carddisp.pl?gene=ROR1>

La red de interacciones de 25 componentes arrojada por String para este gen muestra que las proteínas involucradas tienen más interacciones entre ellas que las esperadas para un grupo aleatorio de proteínas, lo cual indica que biológicamente tienen conexión, al menos parcial, como grupo <https://version11.string-db.org/cgi/network.pl?taskId=OAwJnyGtcwWo> (Figura 25).

Figura 25: Red de interacciones moleculares de 25 componentes para el gen ROR1. Interacciones encontradas: 95. Interacciones esperadas: 11.



Fuente: <https://version11.string-db.org/cgi/network.pl?taskId=OAwwJnyGtcwWo>

Al revisar las implicaciones funcionales de esta red, se encuentra que se ha relacionado con múltiples funciones y/o vías moleculares, entre las cuales las siguientes podrían tener relación con el fenotipo de temblor-ataxia: <https://version11.string-db.org/cgi/network.pl?taskId=OAwwJnyGtcwWo>

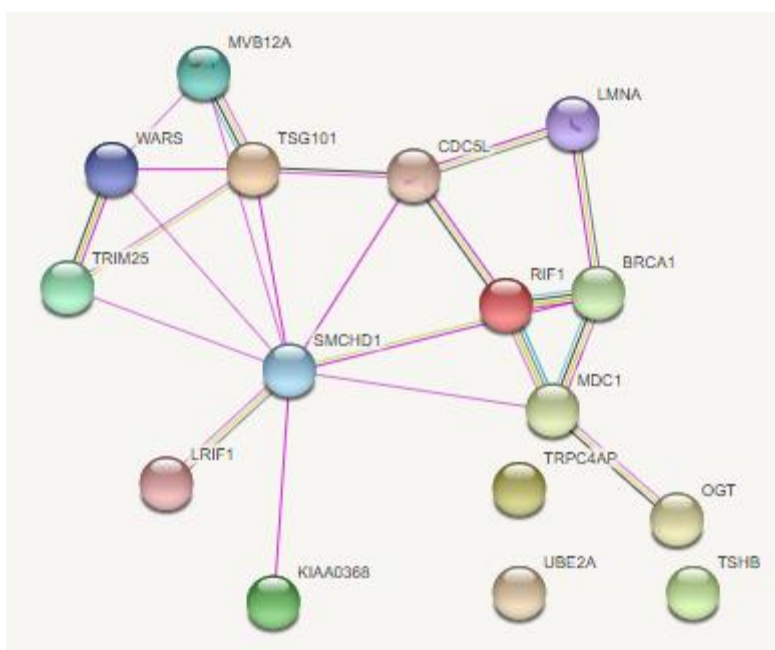
- Morfogénesis de las proyecciones neuronales.
- Neurogénesis.
- Diferenciación de neuronas, astrocitos y glial.
- Guía axonal.
- Regulación de muerte neuronal.
- Desarrollo del sistema nervioso central.
- Regulación positiva de la transmisión sináptica glutamatérgica.
- Desarrollo de nervios.
- Modulación de la transmisión sináptica clínica.
- Vía de señalización de la neurotrofinas.
- Ciclo de la vesícula sináptica.

Gen SMCHD1

El producto de este gen se localiza a nivel celular principalmente en el núcleo. <https://www.genecards.org/cgi-bin/carddisp.pl?gene=SMCHD1>

La red de interacciones de 25 componentes arrojada por String para este gen muestra que las proteínas involucradas tienen más interacciones entre ellas que las esperadas para un grupo aleatorio de proteínas, lo cual indica que biológicamente tienen conexión, al menos parcial, como grupo <https://version11.string-db.org/cgi/network.pl?taskId=zvlfYLPZkIJl> (Figura 26).

Figura 26: Red de interacciones moleculares de 25 componentes para el gen SMCHD1. Interacciones encontradas: 22. Interacciones esperadas: 4.



Fuente: <https://version11.string-db.org/cgi/network.pl?taskId=zvlfYLPZkIJl>

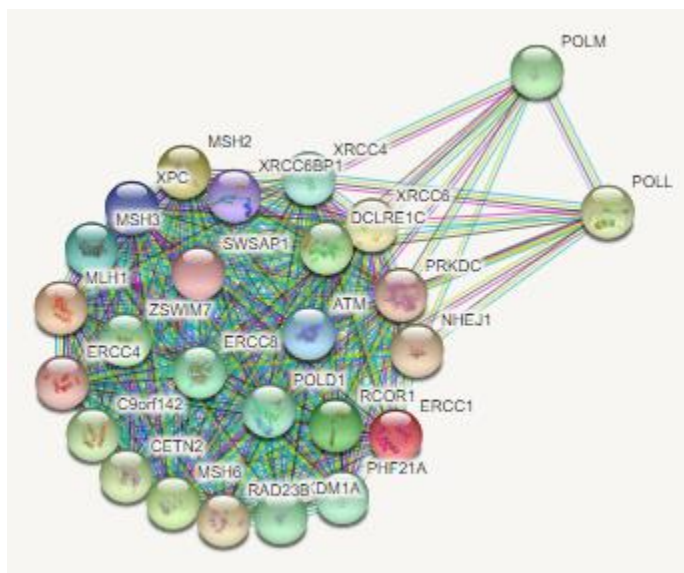
Al revisar las implicaciones funcionales de esta red, se encuentra que se ha relacionado con múltiples funciones y/o vías moleculares de reparación de ADN y control epigenético, entre otras. <https://version11.string-db.org/cgi/network.pl?taskId=zvlfYLPZkIJl>

Gen NHEJ1

El producto de este gen se localiza a nivel celular principalmente en el núcleo, aunque también se encuentra en el citosol. <https://www.genecards.org/cgi-bin/carddisp.pl?gene=NHEJ1>

La red de interacciones de 25 componentes arrojada por String para este gen muestra que las proteínas involucradas tienen más interacciones entre ellas que las esperadas para un grupo aleatorio de proteínas, lo cual indica que biológicamente tienen conexión, al menos parcial, como grupo <https://version11.string-db.org/cgi/network.pl?taskId=n01IDHg1eJdz> (Figura 27).

Figura 27: Red de interacciones moleculares de 25 componentes para el gen NHEJ1. Interacciones encontradas: 291. Interacciones esperadas: 9.



Fuente: <https://version11.string-db.org/cgi/network.pl?taskId=n01IDHg1eJdz>

Al revisar las implicaciones funcionales de esta red, se encuentra que se ha relacionado con múltiples funciones y/o vías moleculares, entre las cuales las siguientes podrían tener relación con el fenotipo de temblor-ataxia: <https://version11.string-db.org/cgi/network.pl?taskId=n01IDHg1eJdz>

- Reparación del ADN.
- Envejecimiento.
- Desarrollo del sistema nervioso central.
- Desarrollo cerebral.

Conclusiones y recomendaciones

- Se encontró un haplotipo común de 33 SNPs comprendido en 83.567.899 pb circundantes a FMR1 en todos los hombres de Ricaurte portadores de la expansión. Con esto se demuestra que el conglomerado genético de síndrome X frágil en Ricaurte se debe a un efecto fundador. Es la primera vez que se demuestra un efecto fundador de síndrome X frágil utilizando la tecnología de secuenciación exómica y haplotipos de SNPs, pues hasta el momento solo se había realizado con haplotipos de microsatélites.
- Los individuos analizados presentan SNPs en desequilibrio de H-W en todos los autosomas, lo cual es esperado por tratarse de una población pequeña y aislada socialmente por el conglomerado X frágil. Sin embargo, el coeficiente de endogamia es negativo, lo podría sugerir que hay un exceso de heterocigotos secundario a una búsqueda

dirigida de parejas de otras poblaciones en un intento de evitar la recurrencia de casos de X frágil, o ser producto del tamaño de muestra.

- Todas las familias de Ricaurte y zonas aledañas con casos X frágil son genéticamente similares. La familia 4 oriunda de Robledo se aleja del resto al considerar los SNPs de todo el genoma, pero no es disímil del resto cuando se consideran los SNPs del cromosoma X.
- Al comparar los individuos de este estudio con los de 1000 genomas, los individuos de Ricaurte y zonas aledañas portadores de la expansión, se ubicaron con los colombianos de 1000 genomas, más cercanos a Europa, América y Asia que a África. Se realizarán estudios más profundos sobre análisis de ancestría en esta población.
- Se encontraron siete variantes tipo SNP, en seis genes, significativamente más frecuentes en individuos expandidos (mutación completa o premutación) con convulsiones. Solamente las ubicadas en el gen PTCHD3 pasaron Bonferroni, por lo que este gen podría ser un candidato a estudiar como posible modificador del fenotipo de convulsión en pacientes con expansión (mutación completa o premutación). Sin embargo, por su expresión en sistema nervioso central y las implicaciones funcionales de su red molecular, es importante tener en cuenta al gen ADAP1. Se realizarán estudios más profundos, en busca de genes modificadores de fenotipo, en los cuales se prestará atención especial a los propios genes encontrados aquí y a los que están en desequilibrio de ligamiento con ellos.
- Se encontraron 12 variantes tipo SNP, en 9 genes, significativamente más frecuentes en mujeres expandidas (premutadas) con falla ovárica prematura. Solamente la ubicada en el gen TRPM3 pasó Bonferroni, por lo que este gen podría ser un candidato a estudiar como posible modificador del fenotipo de falla ovárica prematura en mujeres con expansión (premutación). Sin embargo, por su expresión en ovario, hipotálamo o pituitaria, y las implicaciones funcionales de su red molecular, es importante tener en cuenta los genes CHD5, ADGRF5, PLCB4. Se realizarán estudios más profundos, en busca de genes modificadores de fenotipo, en los cuales se prestará atención especial a los propios genes encontrados aquí y a los que están en desequilibrio de ligamiento con ellos.
- Se encontraron 15 variantes tipo SNP, en 11 genes, significativamente más frecuentes en individuos expandidos (premutación) con temblor-ataxia. Ninguna pasó Bonferroni. Sin

embargo, por su expresión en sistema nervioso central, y las implicaciones funcionales de su red molecular, es importante tener en cuenta los genes STXPB3, PEX1, ZNF430, ROR1, SMCHD1, NHEJ1. Se realizarán estudios más profundos, en busca de genes modificadores de fenotipo, en los cuales se prestará atención especial a los propios genes encontrados aquí y a los que están en desequilibrio de ligamiento con ellos.

- Se planteará la conformación de un Centro de Referencia para el manejo de síndrome X frágil y otras fraxopatías en el Hospital Universitario del Valle, con un enfoque clínico para también de salud pública, que incluya una intervención especial en Ricaurte, con vigilancia epidemiológica, seguimiento a las familias, e intensificación de la consejería genética.

“Con los otros bobos no ha sido así. Como pertenecen a familias que por generaciones enteras le han aportado su tanda de androides al cultivo del Divino, y no poseen culpa alguna por haber procreado un hijo en tales condiciones, los otros 38 bobos de Ricaurte viven en condiciones muy diferentes a las de César Augusto. Buena parte de ellos hablan con dificultad y los que pueden modular lo hacen con tal cantidad de interrupciones y tormentas relampagueantes en sus frases, que casi puede decirse que se requiere de un aparato de rayos laser para entenderlos. Lógicamente, tampoco se visten como César Augusto. Generalmente, o al menos unos treinta o treinta y tres, pertenecen a familias que no poseen recursos económicos permanentes o tienen apenas los tres o cuatro bobos como únicas fuentes de trabajo y, por consiguiente, lo que les pagan a ellos por arriar el ganado, por cuidar de las ovejas africanas o por limpiar los parrales, no alcanza sino para evitar que se mueran de hambre.”

Fragmento de la novela El Divino (1986). Gustavo Álvarez Gardeazabal.

El panorama socioeconómico de Ricaurte, descrito hace 33 años por Gardeazabal, aún permanece, con el agravante de que muchos de estos pacientes ya son adultos mayores, cuyos acudientes en muchos casos ya han fallecido, no tiene medios de subsistencia y viven gracias a la solidaridad de los vecinos.

A lo largo de muchos años, en Ricaurte se ha llevado a cabo todo el proceso descrito por los expertos para el abordaje de conglomerados genéticos, (Castilla & Schuler-faccini, 2014) había un rumor sobre alta frecuencia de pacientes con discapacidad intelectual, se validó el rumor, se generó la hipótesis de que podría tratarse de un conglomerado de síndrome X frágil, lo cual se comprobó en los años 90s mediante cariotipo para fragilidad cromosómica, se diseñaron diferentes protocolos de estudio para determinar prevalencias exactas del fenotipo, de los genotipos de premutación y mutación completa, de los fenotipos asociados a la premutación, en este trabajo, específicamente se demostró que la causa del conglomerado es un efecto fundador, y plantearemos muchos más estudios. Sin embargo, con respecto al paso final del proceso de abordaje de conglomerados genéticos, que es el abordaje del problema

de salud, no se ha pasado de la realización de la consejería genética, cuando lo realmente adecuado es un abordaje integral que incluya una atención multidisciplinaria de los problemas de salud y un abordaje de los problemas psicosociales que afectan a esta población. Se ha realizado ya la notificación de los casos al sistema de vigilancia epidemiológica nacional para que las autoridades sanitarias realicen las intervenciones necesarias, y se está trabajando en la conformación de un centro de referencia para el manejo de síndrome X frágil en el Hospital Universitario del Valle, con atención especial para los pacientes de Ricaurte.

Consideraciones Éticas

El presente trabajo cumplió con la normatividad y consideraciones éticas dispuestas en la Resolución No. 008430 de 1993 del Ministerio de Salud, la Ley 84 de 1989 y la Decisión 391 de la Comunidad Andina de Naciones. No se trabajó aquí con organismos genéticamente modificados.

No hubo riesgos físicos, psicológicos, sociales ni legales puesto que las muestras y los datos para la investigación fueron los obtenidos previamente en un estudio realizado por el grupo de investigación MACOS de la Universidad del Valle con el objetivo de determinar la prevalencia de mutación completa y premutación en FMR1 en esa población. Dichas muestras fueron almacenadas para estudios posteriores con la debida autorización de los pacientes o sus representantes legales. Para el mantenimiento de la confidencialidad se mantuvo la codificación asignada a cada paciente en el estudio previo, de modo que los formatos fueron marcados con dichos códigos y nunca con los datos personales de los individuos. Además, solamente los investigadores tuvieron acceso a los archivos documentales. En las publicaciones resultantes no se revelarán datos personales que pudieran dar lugar a la identificación de los individuos.

Para este trabajo se obtuvo aval del Comité Institucional de Revisión de Ética Humana de la Facultad de Salud de la Universidad del Valle (Acta de aprobación N° 002-018).

Referencias

- Abecasis, G., Auton, A., Brooks, L., DePristo, M., Durbin, R., Handsaker, R., ... McVean, G. (2012). An integrated map of genetic variation from 1,092 human genomes. *Nature*, 491, 56–65. <https://doi.org/10.1038/nature11632>
- Andrews, S. (2010). FastQC: a quality control tool for high throughput sequence data. Retrieved from <http://www.bioinformatics.babraham.ac.uk/projects/fastqc>
- Bagni, C., Tassone, F., Neri, G., & Hagerman, R. (2012). Fragile X syndrome: causes, diagnosis, mechanisms, and therapeutics. *The Journal of Clinical Investigation*, 122(12), 4314–4322. <https://doi.org/10.1172/JCI63141>
- Barrett, J. C., Fry, B., Maller, J., & Daly, M. J. (2005). Haploview: analysis and visualization of LD and haplotype maps. *Bioinformatics*, 21(2), 263–265. <https://doi.org/10.1093/bioinformatics/bth457>

- Bear, M., Huber, K., & Warren, S. (2004). The mGluR theory of fragile X mental retardation. *Trends Neurosci*, 27, 370–377.
- Brown, V., Jin, P., Ceman, S., Darnell, J. C., O'Donnell, W. T., Tenenbaum, S. A., ... Warren, S. T. (2001). Microarray identification of FMRP-associated brain mRNAs and altered mRNA translational profiles in fragile X syndrome. *Cell*, 107(4), 477–487.
- Brown, W. T., Nolin, S., Houck, G. J., Ding, X., Glicksman, A., Li, S. Y., ... Jenkins, E. (1996). Prenatal diagnosis and carrier screening for fragile X by PCR. *American Journal of Medical Genetics*, 64(1), 191–195. [https://doi.org/10.1002/\(SICI\)1096-8628\(19960712\)64:1<191::AID-AJMG34>3.0.CO;2-G](https://doi.org/10.1002/(SICI)1096-8628(19960712)64:1<191::AID-AJMG34>3.0.CO;2-G)
- Castilla, E. E., & Schuler-faccini, L. (2014). From rumors to genetic isolates. *Genetics and Molecular Biology*, 37(1), 186–193.
- Chen, L., Yun, S. W., Seto, J., Liu, W., & Toth, M. (2003). The fragile X mental retardation protein binds and regulates a novel class of mRNAs containing U rich target sequences. *Neuroscience*, 120(4), 1005–1017.
- Chonchaiya, W., Schneider, A., & Hagerman, R. (2009). Fragile X: A family of disorders. *Adv Pediatr*, 56, 165–186.
- Coffee, B., Keith, K., Albizua, I., Malone, T., Mowrey, J., Sherman, S. L., & Warren, S. T. (2009). Incidence of fragile X syndrome by newborn screening for methylated FMR1 DNA. *American Journal of Human Genetics*, 85(4), 503–514. <https://doi.org/10.1016/j.ajhg.2009.09.007>
- Crawford, D. C., Schwartz, C. E., Meadows, K. L., Newman, J. L., Taft, L. F., Gunter, C., ... Sherman, S. L. (2000). Survey of the fragile X syndrome CGG repeat and the short-tandem-repeat and single-nucleotide-polymorphism haplotypes in an African American population. *American Journal of Human Genetics*, 66(2), 480–493. <https://doi.org/10.1086/302762>
- Crawford, D. C., Zhang, F., Wilson, B., Warren, S. T., & Sherman, S. L. (2000). Fragile X CGG repeat structures among African-Americans: identification of a novel factor responsible for repeat instability. *Hum. Mol. Genet*, 9, 1759–1769.
- Crawford, D., Meadows, K., Newman, J., Taft, L., Scott, E., Leslie, M., ... Sherman, S. (2002). Prevalence of the fragile X syndrome in African-Americans. *Am J Med Genet*, 110(3), 226–233.
- Crawford, D., Zhang, F., Wilson, B., Warren, S., & Sherman, S. (2000). Fragile X CGG repeat structures among African-Americans: identification of a novel factor responsible for repeat instability. *Hum Mol Genet.*, 9(12), 1759–1769.
- Daneberg, Z., Pronina, N., Lace, B., & Lugovska, R. (2011). FMR1 Linked haplotype analysis in a mentally retarded male population. *Central European Journal of Medicine*, 6(6), 750–757. <https://doi.org/10.2478/s11536-011-0089-y>
- Darnell, J. C., Driesche, S. J. Van, Zhang, C., Ying, K., Hung, S., Mele, A., ... Darnell, R. B. (2011). HHMI Author Manuscript FMRP stalls ribosomal translocation on mRNAs linked to synaptic function and autism, 146(2), 247–261.

- DePristo, M., Banks, E., Poplin, R., Garimella, K., Maguire, J., Hartl, C., ... Dal, M. (2011). A framework for variation discovery and genotyping using next-generation DNA sequencing data. *Nat. Genet.*, 43, 491–498.
- Eichler, E. E., Macpherson, J. N., Murray, A., Jacobs, P. A., Chakravarti, A., & Nelson, D. L. (1996). Haplotype and interspersed analysis of the FMR1 CGG repeat identifies two different mutational pathways for the origin of the fragile X syndrome. *Human Molecular Genetics*, 5(3), 319–330. <https://doi.org/10.1093/hmg/5.3.319>
- Evanno, G., Regnaut, S., & Goudet, J. (2005). Detecting the number of clusters of individuals using the software STRUCTURE: a simulation study. *Molecular Ecology* 14, 2611–2620. *Molecular Ecology* , 14, 2611–2620.
- Evans, J. (2015). in genetic disease. *CMAJ : Canadian Medical Association Journal = Journal de l'Association Médicale Canadienne*, 187(2), 93–94. <https://doi.org/10.1503/cmaj.140657>
- Excoffier, L., Laval, G., & Schneider, S. (2005). Arlequin (version 3.0): An integrated software package for population genetics data analysis. *Evolutionary Bioinformatics*, 1, 47–50.
- Excoffier, L., Smouse, P., & Quattro, J. M. (1992). Analysis of molecular variance inferred from metric distances among DNA haplotypes: application to human mitochondrial DNA restriction data. *Genetics*, 131(2), 479–491.
- Fernandez-Carvajal, I., Walichiewicz, P., Xiaosen, X., Pan, R., Hagerman, P. J., & Tassone, F. (2009). Screening for expanded alleles of the FMR1 gene in blood spots from newborn males in a Spanish population. *The Journal of Molecular Diagnostics : JMD*, 11(4), 324–329. <https://doi.org/10.2353/jmoldx.2009.080173>
- Fogli, A., Wong, K., Eymard-Pierre, E., Wenger, J., Bouffard, J., Goldin, E., ... Boespflug-Tanguy, O. S. R. (2002). Cree leukoencephalopathy and CACH/VWM disease are allelic at the EIF2B5 locus. *Ann Neurol*, 52(4), 506–510.
- Frio, T. R. (2015). High-Throughput Technologies: DNA and RNA Sequencing Strategies and Potential. In *Pan-cancer Integrative Molecular Portrait Towards a New Paradigm in Precision Medicine* (pp. 47–68). Cham: Springer International Publishing. https://doi.org/10.1007/978-3-319-22189-2_5
- Fu, Y.-H., Kuhl, D. P. A., Pizzuti, A., Pieretti, M., Sutcliffe, J. S., Richards, S., ... Caskey, C. T. (1991). Variation of the CGG repeat at the fragile X site results in genetic instability: Resolution of the Sherman paradox. *Cell*, 67(6), 1047–1058. [https://doi.org/10.1016/0092-8674\(91\)90283-5](https://doi.org/10.1016/0092-8674(91)90283-5)
- Gatto, C. L., & Broadie, K. (2010). Genetic controls balancing excitatory and inhibitory synaptogenesis in neurodevelopmental disorder models. *Frontiers in Synaptic Neuroscience*, 2, 4. <https://doi.org/10.3389/fnsyn.2010.00004>
- Gibbs, R. A., Boerwinkle, E., Doddapaneni, H., Han, Y., Korchina, V., Kovar, C., ... Rasheed, A. (2015). A global reference for human genetic variation. *Nature*, 526(7571), 68–74. <https://doi.org/10.1038/nature15393>

- Gilissen, C., Hoischen, A., Brunner, H. G., & Veltman, J. A. (2012). Disease gene identification strategies for exome sequencing. *European Journal of Human Genetics*, 20(5), 490–497. <https://doi.org/10.1038/ejhg.2011.258>
- Gilissen, C., Hoischen, A., Brunner, H. G., & Veltman, J. a. (2011). Unlocking Mendelian disease using exome sequencing. *Genome Biology*, 12(9), 228. <https://doi.org/10.1186/gb-2011-12-9-228>
- Guo, S., & Thompson, E. (1992). Performing the exact test of Hardy-Weinberg proportion for multiple alleles. *Biometrics*, 48, 361–372.
- Hagerman, R., & Hagerman, P. (2013). Advances in clinical and molecular understanding of the FMR1 premutation and fragile X-associated tremor/ataxia syndrome. *The Lancet. Neurology*, 12(8), 786–798. [https://doi.org/10.1016/S1474-4422\(13\)70125-X](https://doi.org/10.1016/S1474-4422(13)70125-X)
- Hagerman RJ, & Hagerman PJ Editors. (2002). Fragile X Syndrome: Diagnosis, treatment, and research. 3rd ed. Baltimore. *The Johns Hopkins University Press*, 540.
- Hirst, M. C., Knight, S. J., Christodoulou, Z., Grewal, P. K., Fryns, J. P., & Davies, K. E. (1993). Origins of the fragile X syndrome mutation. *Journal of Medical Genetics*, 30(8), 647–650.
- Huang, W., Xia, Q., Luo, S., He, H., Zhu, T., Du, Q., & Duan, R. (2015). Distribution of *fragile X mental retardation 1* CGG repeat and flanking haplotypes in a large Chinese population. *Molecular Genetics & Genomic Medicine*, 3(3), 172–181. <https://doi.org/10.1002/mgg3.128>
- Jeroncic, A., Memari, Y., Ritchie, G. R. S., Hendricks, A. E., Kolb-kokocinski, A., Matchan, A., ... Perica, V. B. (2016). Whole-exome sequencing in an isolated population from the Dalmatian island of Vis. *European Journal of Human Genetics*, 24(April), 1479–1487. <https://doi.org/10.1038/ejhg.2016.23>
- Kchouk, M., Gibrat, J. F., & Elloumi, M. (2017). Generations of Sequencing Technologies: From First to Next Generation. *Biology and Medicine*, 09(03). <https://doi.org/10.4172/0974-8369.1000395>
- Kulski, J. K. (2016). Next-Generation Sequencing — An Overview of the History, Tools, and “Omic” Applications. In *Next Generation Sequencing - Advances, Applications and Challenges*. InTech. <https://doi.org/10.5772/61964>
- Larsen, L. a, Vuust, J., Nystad, M., Evseeva, I., Van Ghelue, M., & Tranebjaerg, L. (2001). Analysis of FMR1 (CGG)(n) alleles and DXS548-FRAXAC1 haplotypes in three European circumpolar populations: traces of genetic relationship with Asia. *European Journal of Human Genetics : EJHG*, 9(9), 724–727. <https://doi.org/10.1038/sj.ejhg.5200697>
- Levene, H. (1949). On a Matching Problem Arising in Genetics. *The Annals of Mathematical Statistics*, 20(1), 91–94. <https://doi.org/10.1214/aoms/1177730093>
- Levenga J, de Vrij FM, Oostra BA, W. R. (2010). Potential therapeutic interventions for fragile X syndrome. *Trends Mol Med*, 16, 516–527.

- Li, H. (2009). The sequence alignment/map format and SAMtools. *Bioinformatics*, 25, 2078–2079.
- Li, H., & Durbin, R. (2009). Fast and accurate short read alignment with Burrows Wheeler transform. *Bioinformatics*, 25, 1754–1760.
- Marcadier, J. L., Boland, M., Scott, C. R., Issa, K., Wu, Z., McIntyre, A. D., ... Lines, M. A. (2015). Congenital sucrase–isomaltase deficiency: identification of a common Inuit founder mutation. *CMAJ : Canadian Medical Association Journal = Journal de l'Association Médicale Canadienne*, 187(2), 102–107. <https://doi.org/10.1503/cmaj.140840>
- Mardis, E. R. (2011). A decade's perspective on DNA sequencing technology. *Nature*, 470(7333), 198–203. <https://doi.org/10.1038/nature09796>
- Maxam, A. M., & Gilbert, W. (1977). A new method for sequencing DNA. *Proceedings of the National Academy of Sciences*, 74(2), 560–564. <https://doi.org/10.1073/pnas.74.2.560>
- McKenna, A., Hanna, M., Banks, E., Sivachenko, A., Cibulskis, K., Kernytsky, A., ... DePristo, M. A. (2010). The Genome Analysis Toolkit: A MapReduce framework for analyzing next-generation DNA sequencing data. *Genome Research*, 20(1297–1303).
- McKusick-Nathans Institute of Genetic Medicine, J. H. U. S. of M. (2018). OMIM. Retrieved December 1, 2015, from <http://www.ncbi.nlm.nih.gov/omim>
- Metzker, M. L. (2010). Sequencing technologies — the next generation. *Nature Reviews Genetics*, 11(1), 31–46. <https://doi.org/10.1038/nrg2626>
- Nei, M. (1978). Estimation of average heterozygosity and genetic distance from a small number of individuals. *Genetics*, 89(3), 583–590. <https://doi.org/10.1093/genetics/89.3.583>
- Nussbaum, R., McInnes, R., & Willard, H. (2016). *Thompson y Thompson. Genética en Medicina* (8^o). Barcelona: Elsevier.
- Pareek, C. S., Smoczynski, R., & Tretyn, A. (2011). Sequencing technologies and genome sequencing. *Journal of Applied Genetics*, 52(4), 413–435. <https://doi.org/10.1007/s13353-011-0057-x>
- Peprah, E. K., Allen, E. G., Williams, S. M., Woodard, L. M., & Sherman, S. L. (2010). Genetic diversity of the fragile X syndrome gene (FMR1) in a large Sub-Saharan West African population. *Ann Hum Genet*, 74(4), 316–325. <https://doi.org/10.1111/j.1469-1809.2010.00582.x>
- Prasad, C., Johnson, J., Bonnefont, J., Dilling, L., Innes, A., Haworth, J., ... Greenberg, C. (2001). Hepatic carnitine palmitoyl transferase 1 (CPT1 A) deficiency in North American Hutterites (Canadian and American): evidence for a founder effect and results of a pilot study on a DNA-based newborn screening program. *Mol Genet Metab*, 73(1), 55–63.
- Pritchard, J., Stephens, M., & Donnelly, P. (2000). Inference of population structure using multilocus genotype data. *Genetics*, 155, 945–959.

- Purcell, S., Neale, B., Todd-Brown, K., Thomas, L., Ferreira, M. A. R., Bender, D., ... Sham, P. C. (2007). PLINK: A Tool Set for Whole-Genome Association and Population-Based Linkage Analyses. *The American Journal of Human Genetics*, 81(3), 559–575. <https://doi.org/10.1086/519795>
- Qiang-long, Z., Shi, L., Peng, G., & Fei-shi, L. (2014). High-throughput Sequencing Technology and Its Application. *Journal of Northeast Agricultural University (English Edition)*, 21(3), 84–96. [https://doi.org/10.1016/S1006-8104\(14\)60073-8](https://doi.org/10.1016/S1006-8104(14)60073-8)
- Rousseau-Nepton, I., Okubo, M., Grabs, R., Mitchell, J., Polychronakos, C., & Rodd, C. (2015). A founder AGL mutation causing glycogen storage disease type IIIa in Inuit identified through whole-exome sequencing: a case series. *CMAJ: Canadian Medical Association Journal = Journal de l'Association Médicale Canadienne*, 187(2), E68–E73. <https://doi.org/10.1503/cmaj.140657>
- ROUSSET, F. (2008). genepop'007: a complete re-implementation of the genepop software for Windows and Linux. *Molecular Ecology Resources*, 8(1), 103–106. <https://doi.org/10.1111/j.1471-8286.2007.01931.x>
- Saldarriaga, W., Forero-Forero, J. V., Fandiño-Losada, A., Ramírez-Cheyne, J., Rodríguez-Forero, J. T., & Tassone, F. (2016). Consejería genética y reproductiva en portadores de premutación en un conglomerado geográfico de síndrome X frágil. *Revista Colombiana de Obstetricia y Ginecología*, 67(2), 159–160.
- Saldarriaga, W., Tassone, F., González-Teshima, L. Y., Forero-Forero, J. V., Ayala-Zapata, S., & Hagerman, R. (2014). Fragile X syndrome. *Colombia Médica (Cali, Colombia)*, 45(4), 190–198. <https://doi.org/10.1002/ajmg.c.30062>
- Sanger, F., Nicklen, S., & Coulson, A. R. (1977). DNA sequencing with chain-terminating inhibitors. *Proceedings of the National Academy of Sciences*, 74(12), 5463–5467. <https://doi.org/10.1073/pnas.74.12.5463>
- Schatz, M. C., Delcher, A. L., & Salzberg, S. L. (2010). Assembly of large genomes using second-generation sequencing. *Genome Research*, 20(9), 1165–1173. <https://doi.org/10.1101/gr.101360.109>
- Shashi, V., McConkie-Rosell, A., Rosell, B., Schoch, K., Vellore, K., McDonald, M., ... Goldstein, D. B. (2014). The utility of the traditional medical genetics diagnostic evaluation in the context of next-generation sequencing for undiagnosed genetic disorders. *Genetics in Medicine*, 16(2), 176–182. <https://doi.org/10.1038/gim.2013.99>
- Shendure, J., & Ji, H. (2008). Next-generation DNA sequencing. *Nature Biotechnology*, 26(10), 1135–1145. <https://doi.org/10.1038/nbt1486>
- Snow, K., Doud, L. K., Hagerman, R., Pergolizzi, R. G., Erster, S. H., & Thibodeau, S. N. (1993). Analysis of a CGG sequence at the FMR-1 locus in fragile X families and in the general population. *American Journal of Human Genetics*, 53(6), 1217–1228.
- Stuckless, S., Parfrey, P., Woods, M., Cox, J., Fitzgerald, G., Green, J., & Green, R. (2007). The phenotypic expression of three MSH2 mutations in large Newfoundland families with Lynch syndrome. *Fam Cancer*, 6(1), 1–12. <https://doi.org/10.1007/s10689-006->

- Sudmant, P., Kitzman, J., Antonacci, F., Alkan, C., Malig, M., Tsalenko, A., ... 1000 Genomes Project, Eichler, E. (2010). Diversity of human copy number variation and multicopy genes. *Science*, 330, 641–646.
- Sutcliffe, J. S., Nelson, D. L., Zhang, F., Pieretti, M., Caskey, C. T., Saxe, D., & Warren, S. T. (1992). DNA methylation represses FMR-1 transcription in fragile X syndrome. *Human Molecular Genetics*, 1(6), 397–400.
- Tassone, F. (2015a). Advanced technologies for the molecular diagnosis of fragile X syndrome. *Expert Review of Molecular Diagnostics*, 7159(October), 1–9. <https://doi.org/10.1586/14737159.2015.1101348>
- Tassone, F. (2015b). Advanced technologies for the molecular diagnosis of fragile X syndrome. *Expert Review of Molecular Diagnostics*, 7159(October), 1–9. <https://doi.org/10.1586/14737159.2015.1101348>
- Tassone, F., Iong, K. P., Tong, T.-H., Lo, J., Gane, L. W., Berry-Kravis, E., ... Hagerman, R. J. (2012). FMR1 CGG allele size and prevalence ascertained through newborn screening in the United States. *Genome Medicine*, 4(12), 100. <https://doi.org/10.1186/gm401>
- Thorlacius, S., Sigurdsson, S., Bjarnadottir, H., Olafsdottir, G., Jonasson, J., Tryggvadottir, L., ... Eyfjörð, J. (1997). Study of a single BRCA2 mutation with high carrier frequency in a small population. *Am J Hum Genet*, 60, 1079–1084.
- Thudi, M., Li, Y., Jackson, S. A., May, G. D., & Varshney, R. K. (2012). Current state-of-art of sequencing technologies for plant genomics research. *Briefings in Functional Genomics*, 11(1), 3–11. <https://doi.org/10.1093/bfpg/blr045>
- Triggs-Raine, B., Kirkpatrick, R., Kelly, S., Norquay, L., Cattini, P., Yamagata, K., ... Hegele, R. (2002). HNF-1alpha G319S, a transactivation-deficient mutant, is associated with altered dynamics of diabetes onset in an Oji-Cree community. *Proc Natl Acad Sci U S A*, 99(7), 4614–4619.
- Verkerk AJ, Pieretti M, Sutcliffe JS, Fu YH, Kuhl DP, Pizzuti A, Reiner O, Richards S, Victoria MF, Z. F. (1991). Identification of a gene (FMR-1) containing a CGG repeat coincident with a breakpoint cluster region exhibiting length variation in fragile X syndrome. *Cell*, 65, 905–914.
- Wang, K., Li, M., & Hakonarson, H. (2010). ANNOVAR: Functional Annotation of Genetic Variants from High-Throughput Sequencing Data. *Nucleic Acids Research*, 38, e164.
- WATSON, J. D., & CRICK, F. H. (1953). Molecular structure of nucleic acids; a structure for deoxyribose nucleic acid. *Nature*, 171(4356), 737–738. Retrieved from <http://www.ncbi.nlm.nih.gov/pubmed/13054692>
- Wattendorf, D., & Hadley, D. (2005). Family history: the three-generation pedigree. *Am Fam Physician*, 72(3), 441–448.

- Weir, B., & Cockerham, C. (1984). Estimating F-statistics for the analysis of population structure. *Evolution*, 38, 1358–1370.
- Wells, R. D. (2009). Mutation spectra in fragile X syndrome induced by deletions of CGG*CCG repeats. *The Journal of Biological Chemistry*, 284(12), 7407–7411. <https://doi.org/10.1074/jbc.R800024200>
- Wheeler, D. A., Srinivasan, M., Egholm, M., Shen, Y., Chen, L., McGuire, A., ... Rothberg, J. M. (2008). The complete genome of an individual by massively parallel DNA sequencing. *Nature*, 452(7189), 872–876. <https://doi.org/10.1038/nature06884>
- Williams, S. M., Stocki, S., Jiang, L., Brew, K., Gordon, S., Vaughan, D. E., ... Moore, J. H. (2000). A population-based study in Ghana to investigate inter-individual variation in plasma t-PA and PAI-1. *Ethnicity & Disease*, 17.
- Wright, S. (1965). The interpretation of population structure by F-statistics with special regard to systems of mating. *Evol*, 19, 395–420.
- Xue, Y., Ankala, A., Wilcox, W. R., & Hegde, M. R. (2015). Solving the molecular diagnostic testing conundrum for Mendelian disorders in the era of next-generation sequencing: Single-gene, gene panel, or exome/genome sequencing. *Genetics in Medicine*, 17(6), 444–451. <https://doi.org/10.1038/gim.2014.122>
- Yang, Y., Ph, D., Muzny, D. M., Sc, M., Reid, J. G., Bainbridge, M. N., ... Lupski, J. R. (2014). Clinical Whole-Exome Sequencing for the Diagnosis of Mendelian Disorders. *N Engl J Med*, 369(16), 1502–1511. <https://doi.org/10.1056/NEJMoa1306555>.Clinical
- Zheng-Bradley, X., & Flicek, P. (2017). Applications of the 1000 Genomes Project resources. *Brief Funct Genomics*, 16(3), 163–170. <https://doi.org/10.1093/bfgp/elw027>
- Zhong, N., Kajanoja, E., Smits, B., Pietrofesa, J., Curley, D., Wang, D., ... Brown, W. T. (1996). Fragile X Founder Effects and New Mutations in Finland. *American Journal of Medical Genetics*, 64, 226–233.
- Zhou, Y., Tang, K., Law, H. Y., Ng, I. S. L., Lee, C. G. L., & Chong, S. S. (2006). FMR1 CGG repeat patterns and flanking haplotypes in three Asian populations and their relationship with repeat instability. *Annals of Human Genetics*, 70(6), 784–796. <https://doi.org/10.1111/j.1469-1809.2006.00265.x>



Comité Institucional de Revisión de Ética Humana
Facultad de Salud

ACTA DE APROBACIÓN N° 002 - 018

Proyecto: "ANÁLISIS DE EXOMA EN PACIENTES DE UN CLUSTER DE SÍNDROME DE X FRÁGIL"

Presentado por: **JULIÁN ANDRÉS RAMÍREZ CHEYNE / WILMAR SALDARRIAGA GIL**
OSCAR MAURICIO ARCOS

Código Interno: 008 - 018 **Fecha en que fue sometido:** 29 / 01 / 2018

El Consejo de la Facultad de Salud de la Universidad del Valle, ha establecido el Comité Institucional de Revisión de Ética Humana (CIREH), el cual está regido por la Resolución 008430 del 4 de octubre de 1993 del Ministerio de Salud de Colombia por la cual se establecen las normas científicas, técnicas y administrativas para la investigación en salud; los principios de la Asamblea Médica Mundial expuestos en su Declaración de Helsinki de 1964, última revisión en 2002; y el Código de Regulaciones Federales, título 45, parte 46, para la protección de sujetos humanos, del Departamento de Salud y Servicios Humanos de los Institutos Nacionales de Salud de los Estados Unidos 2009.

Este Comité certifica que:

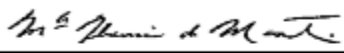
- Sus miembros revisaron los siguientes documentos del presente proyecto:

☒ Resumen del proyecto	☒ Protocolo de investigación
☒ Formulario de consentimiento informado	☒ Instrumento de recolección de datos
☐ Folleto del investigador (si aplica)	☒ Cartas de las instituciones participantes
☐ Resultados de evaluación por otros comités (si aplica)	
- El presente proyecto fue evaluado y aprobado por el Comité.
- Según las categorías de riesgo establecidas en el artículo 11 de la Resolución N° 008430 de 1993 del Ministerio de Salud, el presente estudio tiene la siguiente **Clasificación de Riesgo:**

☐ Sin riesgo	☒ Riesgo Medio	☐ Riesgo Mayor del Medio
-------------------------------------	--	---
- Que las **medidas** que están siendo tomadas para proteger a los sujetos humanos son adecuadas.
- La forma de obtener el **consentimiento** informado de los participantes en el estudio es adecuada.
- Este proyecto será **revisado nuevamente** en la próxima reunión plenaria del Comité, sin embargo, el Comité puede ser convocado a solicitud de algún miembro del Comité o de las directivas institucionales para revisar cualquier asunto relacionado con los derechos y el bienestar de los sujetos involucrados en este estudio.
- Informará** inmediatamente a las directivas institucionales:
 - Todo desacato de los investigadores a las solicitudes del Comité.
 - Cualquier suspensión o terminación de la aprobación por parte del Comité.
- Informará** inmediatamente a las directivas institucionales toda información que reciba acerca de:
 - Lesiones a sujetos humanos.

Calle 4B 36 -50 edificio Decanato Teléfono: 5155677 email: etica@univalle.edu.co

- Problemas imprevistos que involucren riesgos para los sujetos u otras personas.
- b. Cualquier cambio o modificación a este proyecto que no haya sido revisado y aprobado por el Comité.
9. El presente proyecto ha sido **aprobado** por un periodo de **1 año** a partir de la fecha de aprobación. Los proyectos de duración mayor a un año, deberán ser sometidos nuevamente con todos los documentos para revisión actualizados.
10. El **investigador principal** deberá informar al Comité:
- Cualquier cambio que se proponga introducir en este proyecto. Estos cambios no podrán iniciarse sin la revisión y aprobación del Comité excepto cuando sean necesarios para eliminar peligros inminentes para los sujetos.
 - Cualquier problema imprevisto que involucre riesgos para los sujetos u otros.
 - Cualquier evento adverso serio dentro de las primeras 24 horas de ocurrido, al secretario(a) y al presidente (Anexo 1).
 - Cualquier conocimiento nuevo respecto al estudio, que pueda afectar la tasa riesgo/beneficio para los sujetos participantes.
 - cualquier decisión tomada por otros comités de ética.
 - La terminación prematura o suspensión del proyecto explicando la razón para esto.
 - El investigador principal deberá presentar un informe al final del año de aprobación. Los proyectos de duración mayor a un año, deberán ser sometidos nuevamente con todos los documentos para revisión actualizados.

Firma:  Fecha: 04 12 2018
Nombre: **MARIA FLORENCIA VELASCO DE MARTINEZ**
Capacidad representativa: **PRESIDENTA** Teléfono: 5185677

CERTIFICACIÓN DE LA FACULTAD DE SALUD DE LA UNIVERSIDAD DEL VALLE

Por medio de la presente, certifico que la Facultad de Salud de la Universidad del Valle aprueba el proyecto arriba mencionado y respeta los principios, políticas y procedimientos de la Declaración de Helsinki de la Asamblea Médica Mundial, de la Resolución 8430 de 1993 del Ministerio de Salud y de la reglamentación vigente en investigación de la Universidad del Valle.

Firma:  Fecha: 04 12 2018
Nombre: **ADALBERTO SÁNCHEZ G.**
Capacidad representativa: **VICEDECANO DE LA FACULTAD DE SALUD** Teléfono: 5185680