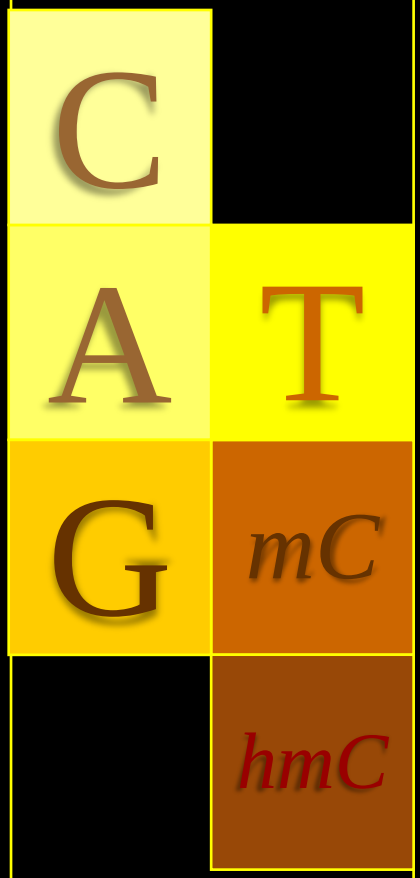
CENTRO PARA LA INVESTIGACIÓN Y
REHABILITACIÓN DE LAS ATAXIAS HEREDITARIAS
"CARLOS J. FINLAY"
DEPARTAMENTO DE NEUROBIOLOGÍA
MOLECULAR
FACULTAD DE BIOLOGÍA
UNIVERSIDAD DE LA HABANA



GEN ATAXIN-2 EN LA POBLACIÓN CUBANA: MECANISMO MUTAGÉNICO DE LA EXPANSIÓN TRINUCLEOTÍDICA Y METILACIÓN EPIGÉNETICA MODIFICADORA DEL FENOTIPO

Tesis presentada en opción al grado científico de Doctor en Ciencias Biológicas

JOSE MIGUEL LAFFITA MESA

HOLGUÍN

2014

CENTRO PARA LA INVESTIGACIÓN Y REHABILITACIÓN DE LAS
ATAXIAS HEREDITARIAS "CARLOS J. FINLAY"
DEPARTAMENTO DE NEUROBIOLOGÍA MOLECULAR
FACULTAD DE BIOLOGÍA
UNIVERSIDAD DE LA HABANA

GEN *ATAXIN-2* EN LA POBLACIÓN CUBANA: MECANISMO MUTAGÉNICO DE LA EXPANSIÓN TRINUCLEOTÍDICA Y METILACIÓN EPIGÉNETICA MODIFICADORA DEL FENOTIPO

Tesis presentada en opción al grado científico de Doctor en Ciencias Biológicas

JOSÉ MIGUEL LAFFITA MESA

HOLGUÍN
2014

CENTRO PARA LA INVESTIGACIÓN Y REHABILITACIÓN DE LAS ATAXIAS HEREDITARIAS "CARLOS J. FINLAY"

DEPARTAMENTO DE NEUROBIOLOGÍA MOLECULAR

FACULTAD DE BIOLOGÍA

UNIVERSIDAD DE LA HABANA

GEN *ATAXIN-2* EN LA POBLACIÓN CUBANA: MECANISMO MUTAGÉNICO DE LA EXPANSIÓN TRINUCLEOTÍDICA Y METILACIÓN EPIGÉNETICA MODIFICADORA DEL FENOTIPO

Tesis presentada en opción al grado científico de Doctor en Ciencias Biológicas

Autor: Lic. José Miguel Laffita Mesa

Tutores: Prof. Tit., Dr. Luís Velázquez Pérez, Dr. Cs.

Prof. Tit., Dra. Vivian Kourí Cardellá, Dra. C.

HOLGUÍN

2014

AGRADECIMIENTOS:

Al pensar que pudiese olvidar algún nombre, extiendo mi gratitud a los que colaboraron de una u otra forma en la culminación del presente trabajo, pero en especial a:

- A mi Dios (Jesús Cristo) y Padre verdadero, el único que he tenido. Por creer en mí, por soportarme. Por darme ideas que no me dejan dormir, por enseñarme que junto a Él todo es posible. A él sea todo mi honor y gloria.
- A mi unigénito angelito Elías Miguel, por estar siempre conmigo y ser la razón de mi inspiración. Hijo disculpa todo el tiempo perdido, el futuro te lo dedico.
- A mi esposa Gretchen Batule Duran que me inunda de felicidad. Gracias porque has sabido soportar este peso que la vida me ha puesto encima. Mi amor puedes contar conmigo. Uds., son mi fiel equipo que al llegar al hogar me llenan de alegría y dicha.
- A mi Madre, por no pedir mucho de mí, ni tampoco por esperar mucho de mí.
- A mi hermano Lázaro por sustentarme y darme cobijo cuando más lo necesité. Tengo guardado dentro de mi todo lo que me diste sin esperar nada a cambio. Hoy se cumple parte de mi locura: Escena 1, Acto 1. A Bernabé, por ser amigo.
- A todos los pacientes de ataxias hereditarias de Cuba y a sus familiares en riesgo, en especial a los de la ataxia espinocerebelosa tipo 2 y a los de ELA, que accedieron a participar en este estudio.
- A las Dras. Rosa Rademaker y Mariely DeJesus-Hernandez de Mayo Clinic por brindarnos controles positivos para la mutación *C9ORF72*.
- A mi amigo y compadre Manuel Cortiña por estimularme en este camino, y por darme apoyo en la impresión.
- A los que se unieron en el camino como el Departamento de Investigaciones de la Universidad de Ciencias Médicas de Holguín, y en especial a su Rectora Dra. Katia Ochoa y la Dra Magalis, Directora de Ciencia e Innovación Tecnológica, por hacer suyo nuestros problemas y cotidianidad y adoptarnos como hijos.

En fin, a todos los que recuerdo y los que se me han olvidado, a los que durante estos 38 años de juventud me han acompañado y han compartido este instante de vida.

Ah, y a los que pensaron que este momento nunca llegaría…

DEDICATORIA:

- A mi Dios.
- A mi Hijo querido, quien en su inocencia me ha hecho sentir muy grande, con el sólo acto de amarme.
- A mi amada esposa Gretchen Batule, porque aún no entendiendo y sacrificando mucho ha sabido esperar por mí. Sé muy bien cuanto te amo y te quiero, nada de esto sería posible si tú, mi amor, no estuvieras.
- A mi Madre querida quien hasta hoy me ha guiado por el sendero estrecho de la verdad.
- A los ignorados, los rebeldes, los revolucionarios, a los rechazados, a los soñadores, a los sinceros y transparentes, a los firmes en su carácter, a los temerarios, audaces y atrevidos, a los cuatro ojos, a los perseverantes, a los que no se cansan, a los inconformes, a los que van contra la lógica y dicen sí contra muchos no, a los que no respetan el *statu quo*, los que rompen las reglas, a esos no los podemos ignorar esos son los que hacen este mundo diferente, los que nos hacen avanzar...

Dedico esta tesis a la excelencia mayor, que es la sabiduría, al conocimiento que es virtud y sólo así se puede ser libre….

LISTA ABREVIATURAS

A2BP1: Proteína unidora de ataxina-2, *en inglés Ataxin-2 Binding Protein*

ADN: Ácido desoxiribonucleico

ALs: Alelos largos

ARN: Ácido ribonucleico

ATG: Codón de ***A***denina ***T***imina ***G***uanina reconocido como sitio consenso para el inicio de la transcripción en eucariontes

ATXN2: Gen Ataxin 2

C9ORF72: Marco de lectura abierto No. 72 en el locus 9p21

CAA: Trinucleótido ***C***itosina ***A***denina ***A***denina

CAG: Trinucleótido ***C***itosina **A**denina ***G***uanina

Ct: Ciclo umbral, en inglés *Cycle threshold*

DNMTs : Enzimas DNA Metil Tranferasas

ELA: Esclerosis Lateral Amiotrófica

EP: Enfermedad de Parkinson

EP: Enfermedad de Parkinson

GE: Gránulos de estrés

HUGO: Organización del genoma, en inglés *Human Genome Organization*

IBRO: Organización de investigaciones del cerebro, en inglés *International Brian Research Organization*

LSm: Motivo parecido a Sm, en inglés *like spliceosomal small nuclear ri-bonucleoproteins motifs*

LSm-AD: Motivo asociado al dominio Lsm**,** en inglés *like spliceosomal small nuclear ri-bonucleoproteins motifs associated domains*

MDS: Sociedad de los trastornos del movimiento, *en inglés Movement Disorder Society*

MethyLight: Tecnicismo para la PCR cuantitativa para metilación

mi-ARNs: Micro ácido ribonucleico

MSP: Técnica de PCR específica para Metilación, en inglés *Methyl Specific PCR*

PSP: Parálisis Supranuclear Progresiva

RER: Retículo Endoplasmático Rugoso

SCA2: Ataxia Espinocerebelosa tipo 2, en inglés *Spinocerebellar Ataxia type 2*

SCAs: Sigla genérica para las Ataxias Espinocerebelosas, en inglés *Spinocerebellar Ataxias*

SFN: Sociedad Norteamericana de Neurociencias, en inglés *Society for Neurosciences*

Sm: Motivos de empalme de las ribo-nucleoproteínas nucleares pequeñas, en inglés *spliceosomal small nuclear ri-bonucleoproteins motifs*

SNP: Polimorfismos de un solo nucleótido, en inglés *Single Nucleotide Polymorphisms*

STR: Polimorfismo de microsatélites, en inglés *Short Tandem Repeat*

TDP-43: TAR DNA-binding protein-43

ZBRK1: en inglés *Zinc-finger BRCA1-interacting protein with a KRAB domain 1*

Tabla de Contenido ***Pág.***

SÍNTESIS

Las expansiones de CAG en el gen *ATXN2* causan SCA2, y se asocian con Esclerosis Lateral Amiotrófica, Parálisis Supra nuclear Progresiva y Enfermedad de Parkinson. No se conocen mecanismos moleculares explicativos del origen de estas alteraciones genéticas. El efecto fenotípico de esta mutación es pleiotrópico existiendo un conocimiento limitado de los fenómenos moleculares asociados a esta variabilidad. Se presenta la más amplia caracterización de los polimorfismos genéticos asociados con la expansión de CAG de *ATXN2*, no afectada por SCA2. Se determina la relación existente entre la frecuencia de alelos largos normales y la de SCA2 ($r2=0.841$, $p<0.000$). Los alelos largos mostraron inestabilidad i) somática y ii) intergeneracional ($\chi^2=159.80$, $p<0.0000$). Se definió el rango premutacional desde expansiones de 25-31CAG. Se identifican factores genéticos que determinan el origen de mutaciones nuevas desde expansiones no patogénicas. Se comprobó que a partir de estas expansiones se originaron expansiones patogénicas causales de ELA y SCA2. Se identificó metilación la cual modificó el debut de la SCA2 ($\chi^2=11.59$, $p<0.0007$). Se aplica esta información al asesoramiento genético. Se demuestra el origen de las mutaciones y la metilación de *ATXN2*, siendo importante para atender y tratar las enfermedades con estas mutaciones en *ATXN2*.

INTRODUCCIÓN

I

1. INTRODUCCIÓN

Las enfermedades neurodegenerativas aumentan su prevalencia con el incremento de la expectativa de vida, lo que representa un claro y serio problema para la salud pública mundial (Trojanowski., 2008).

Una de las mutaciones genéticas causales de un gran número de estas enfermedades neurodegenerativas son las expansiones de nucleótidos repetidos, y dentro de este grupo están las mutaciones dinámicas.

Las mutaciones dinámicas son alteraciones genéticas que al trasmitirse de padres a hijos tienden a disminuir o aumentar en cierto número de unidades repetitivas. Una de las mutaciones dinámicas más comunes son las expansiones del trinucleótido *C*itosina-*A*denina-*G*uanina -CAG-. Esta mutación situada en el primer exón del gen *ATXN2* se traduce en un segmento repetitivo de glutaminas (Gln/Q), y es la causa de la ataxia espinocerebelosa tipo 2 -SCA2-.

Este gen se extiende por una región de ~130Kb en el ADN genómico, abarcando 25 exones, y codificando una proteína de expresión ubicua (ataxina-2). A esta proteína no se le conoce función fisiológica, pero los datos más recientes la vinculan al control de la traducción y la señalización celular. Estos datos son apoyados por la presencia de motivos de unión al ARN, la interacción física con proteínas vinculadas a la traducción, su localización en el Retículo endoplasmático rugoso (RER), su unión a polisomas, así como su influencia sobre receptores tirosina quinasas de membrana, e interacciones con endofilinas (Satterfield & Pallanck, 2006; Nonis *et al*., 2008).

De manera global, se reconocen dos estados de ataxina-2 en relación con la expansión de tripletes de CAG o el segmento de poliglutaminas (polyQ): ataxina-2 salvaje y la mutante. Estos a su vez definen dos poblaciones, la saludable y la enferma lo cual está

determinado por umbrales de acuerdo a la longitud de la expansión reiterativa de CAG o del motivo polyQ (aquí usaremos CAG), si bien existe superposición genética.

El rango de variación de la expansión de CAG en *ATXN2* en la población saludable es de 13-31 unidades repetidas, con 22CAG como el más común. Actualmente, se considera que las expansiones de 27-39CAG incrementan el riesgo para Esclerosis Lateral Amiotrófica (ELA) (Elden *et al*., 2010), siendo 30 unidades la más común. El rango de 34-79 tripletes de CAG corresponde a expansiones que son causa monogénica de la SCA2, con 37CAG como el más frecuente (Pulst *et al*., 1996, Imbert *et al*., 1996, Sanpei *et al*., 1996). Las grandes expansiones ≥200-750CAG causan encefalopatía infantil, y actualmente se considera como una extensión del fenotipo de la SCA2 (Lastres-Becker *et al*., 2008).

Existe un rango de incertidumbre diagnóstica que incluye 32-34 repeticiones de CAG, (Sequeiros *et al*., 2010). En Cuba se encontró un paciente con 32CAG mostrando un cuadro leve de SCA2 (Santos *et al*., 1999). En el espectro correspondiente al rango mutante se han reportado otras enfermedades como la Enfermedad de Parkinson (EP), Parálisis Supranuclear Progresiva (PSP) y atrofia Multisistema (Lastres Becker *et al*., 2008). Todo esto sugiere que esta proteína con este segmento repetitivo anormal, vulnera varios grupos neuronales, desconociéndose el mecanismo patogénico.

La SCA2 alcanza las mayores tasas de prevalencia e incidencia a nivel mundial en la provincia de Holguín, donde existen más de 500 enfermos vivos, y 7000 descendientes. Estos últimos dispersos por toda la isla lo cual representa un serio problema de salud nacional. Anualmente nacen 22 niños portadores de la mutación, enferman 35 nuevos casos y fallecen 15 enfermos (Velázquez-Pérez *et al*., 2003).

En la SCA2 es conocida la relación entre la longitud de la secuencia de CAG y la edad de debut de la ataxia de la marcha. Este determinismo tiene su validez en un intervalo limitado de expansiones de CAG (≥35CAG), dado que existe penetrancia reducida en el intervalo de expansiones ≥33CAG<35CAG, además existen pares de hermanos con mutaciones similares y con edades de inicio muy diferentes. Este comportamiento pudiera considerarse como un primer nivel de variabilidad.

Otro nivel es aquel donde la mutación causa SCA2 combinada con signos de Parkinsonismo o enfermedad de la Motoneurona, aunque también se presenta como estas enfermedades sin combinación con la SCA2 (Infante *et al*., 2004, Charles *et al*., 2007, Elden *et al.,* 2010).

Con el propósito de explicar esta variabilidad fenotípica en SCA2 se ha propuesto la existencia de genes modificadores y otros factores modificadores como polimorfismos genéticos en *ATXN2* u otros genes, así como marcas epigenéticas (patrones de metilación diferenciales, código de histonas). Estos factores serían el punto de partida para la búsqueda de dianas terapéuticas con el objetivo de retardar el comienzo de la SCA2 (Chattopadhyay *et al*., 2003; Bauer *et al*., 2004; Pulst *et al*., 2005).

Sin embargo, sólo se han descrito tres factores genéticos modificadores, los cuales son variantes genéticas en otros genes y genomas que contribuyen, discretamente, a la variabilidad del debut de la SCA2. Estos son los genes *RAI1, CACNA1A* y la variante común A10398G del complejo mitocondrial I (Hayes *et al.,* 2000, Pulst *et al.,* 2005, Simon *et al.,* 2005). Se sospecha, además, que la interrupción del triplete de CAA ejerce un efecto modificador sobre el fenotipo SCA2 derivándolo a EP (Charles et al., 2007), pero no ha sido demostrado científicamente, es sólo una observación.

Las marcas epigenéticas, en específico la metilación del ADN, han sido sugeridas como posible modificadores. La metilación regula la expresión génica, reprimiendo la transcripción, en tanto que su ausencia activa la expresión (Oakeley., 1999). Aguiar *et al*., 1999 sugirieron un origen no metilado para el promotor *ATXN2*, lo cual explicaba la fortaleza de esta secuencia. Esto fue además apoyado por la observación de que el promotor *ATXN2* estaba dentro de un islote de CpG. Por tanto, se infiere la posibilidad de un mecanismo epigenético controlando la expresión de ataxina-2. Empíricamente se sugirió un estado libre de metilación para este gen en estadios embrionarios, *pero esto no se demostró ni tampoco se identificó la existencia de estas marcas epigenéticas*.

Bauer *et al.,* 2003 hipotetizaron que la presencia de metilación *de novo* en el promotor *ATXN2*, podía silenciar la expresión del alelo mutante. Esto, definiría un debut tardío o temprano (en su ausencia) de la SCA2. *Sin embargo, sólo encontraron evidencias indirectas apoyando su postulado*. Por tanto, no hay un estudio que así demuestre la existencia de metilación sobre el promotor *ATXN2* ejerciendo un efecto modificador que explique la variabilidad fenotípica en la SCA2. Más allá de esto, no existe idea clara del mecanismo que controla la expresión de este gen.

Por otra parte, poco se conoce de los factores desencadenantes de la inestabilidad de la reiteración de CAG de este gen. La secuencia repetida de CAG en *ATXN2*, normalmente está interrumpida por tripletes de CAA. El alelo más común de 22CAG tiene la lectura (CAG)8-CAA-(CAG)4-CAA-(CAG)8 u 8+4+8, donde '+' significa CAA. En analogía con otras enfermedades se ha propuesto que las interrupciones de CAA son un anclaje para la estabilidad genética del segmento de CAG, y que su ausencia en alelos intermedios o diferentes de 22CAG predispone a la inestabilidad, a lo cual también contribuye el tamaño del triplete repetido de CAG, el contexto genético o haplotipo

cercano a la mutación y factores fisiológicos como el sexo y edad del progenitor, entre otros. En analogía con otras enfermedades de similar naturaleza se supone que aquellos alelos ≥ 23CAG están predispuestos o "premutados", y se expanden en generaciones sucesivas (Takano *et al*., 1998), lo que no se ha comprobado.

De todo lo anterior se puede resumir que existen *seis temas recurrentes* los cuales son centrales en las investigaciones biomédicas conectadas con la expansión del triplete de CAG en *ATXN2*. Uno de ellos hace referencia a los mecanismos mutagénicos de la expansión de CAG en el locus *ATXN2*, o sea, (1) ¿cómo se origina la mutación SCA2?, el otro es (2) ¿cómo se regula la expresión del gen *ATXN2*?, (3) ¿cuál es su función fisiológica? El otro hace referencia al papel del producto génico de *ATXN2* en la etiopatogenia de SCA2 y otras enfermedades o sintéticamente (4) ¿cuál es el mecanismo patogénico de la proteína ataxina-2? El conocimiento de este mecanismo es la antesala para una posible terapia. Otra pregunta está enfocada en (5) ¿el por qué de la variabilidad fenotípica y el pleiotropismo de la mutación? Y finalmente todas estas se traducirían en (6) el cambio del estado mórbido al saludable.

Esto supone varios **problemas científicos** resumidos en: "*la no identificación de determinantes y contribuyentes moleculares de la mutagénesis del segmento de CAG en ATXN2, y de factores genéticos que una vez expandida la secuencia de CAG, regulan la expresión del alelo mutante contribuyendo a la variabilidad fenotípica observada"*.

Esto, a su vez, puede estar modulado por una visión reduccionista (*tripletológica*) de asumir a la expansión del triplete de CAG y su relación con el fenotipo como *único* factor determinante de SCA2 o enfermedades conectadas con este gen. Por tanto, toda investigación básica o traslacional se limita a la relación: "longitud del triplete CAG→fenotipo", desviando la atención del desarrollo de técnicas o investigaciones

que permitan identificar otros factores moleculares, biomarcadores, u eventos que permitirían comprender y caracterizar mejor el cambio del estado *saludable* al *mórbido*. Sobre la base de estos antecedentes, la presente investigación contribuye a la solución de estos problemas mediante la exploración directa de las preguntas: (1), (2) y (5).

Esto nos lleva a proponer la siguiente ***hipótesis***: *"más allá de la longitud del CAG, pudieran existir factores intrínsecos de este gen, que determinan la mutagénesis y la variabilidad fenotípica asociada a esta mutación en ATXN2".*

Estos factores pudieran estar dados por variaciones genéticas asociadas al crecimiento de la expansión de CAG, lo que predispone a la inestabilidad genética. La existencia de haplotipos en riesgo en nuestra población pudiera ser la causa del estado epidemiológico actual, lo que sugiere la existencia de estados pre-mutacionales contribuyentes con nuevos casos.

Además, dentro de los factores intrínsecos que determinan el fenotipo y su variabilidad, la localización del promotor en un islote de CpG pudiera inducir distintos estados de activación del gen en la población. Estos pudieran estar dados por patrones diferenciales de metilación epigenética del ADN que controlarían tanto la expresión génica y fenotípica así como la inestabilidad del gen.

Para poder aceptar o refutar esta hipótesis nos propusimos los siguientes **objetivos**:

General: Caracterizar los factores genéticos y epigenéticos que determinan el fenotipo y el origen de la SCA2 en Cuba.

Específicos.

(1) Determinar el efecto de factores genéticos tales como el patrón de interrupción de CAA y haplotipo alrededor de la expansión de CAG en la predisposición a la inestabilidad genética.

(2) Demostrar la validez de la relación entre frecuencia relativa de alelos largos y la frecuencia de SCA2.

(3) Identificar posibles mecanismos mutagénicos de la expansión de CAG en las transmisiones intergeneracionales.

(4) Desarrollar técnicas moleculares para analizar la metilación del promotor *ATXN2*.

(5) Determinar la influencia de la metilación del promotor *ATXN2* sobre el fenotipo.

Fundamento Metodológico: La investigación clasifica como analítica y observacional. Para abordarla partimos de un estudio demográfico (2009-2013). Posteriormente, se analizaron relaciones causales mediante el estudio de variables moleculares como el CAG, la metilación y el contexto genético sobre la inestabilidad genética, la edad de inicio y el fenotipo clínico en controles y cohortes SCA2.

Se obtuvo un gran número de muestras de ADN (~3000) de pacientes SCA2, SCA3 y controles las que fueron analizadas mediante técnicas moleculares y análisis estadísticos. Esto incluyó secuenciación, análisis de fragmentos para determinar haplotipos y longitud de la expansión de CAG. Se realizaron análisis de desequilibrio de ligamiento, meta-análisis, análisis *in silico* y otros métodos estadísticos. Además, desarrollamos dos técnicas moleculares: *Methyl Specific PCR –MSP-* y *MethyLight* para detectar la metilación del promotor *ATXN2*.

La **novedad científica**: Previamente se postuló el modelo teórico: *"a alta frecuencia de alelos largos alta frecuencia de ataxias"* lo cual generó una serie de datos contradictorios. En esta investigación se confirmó este modelo al menos para la SCA2 ya que se comprobó, por primera vez en el mundo, que la población con mayor número de alelos largos en *ATXN2* tiene la mayor prevalencia de SCA2. También, se demostró

que las inestabilidades genéticas de CAG en *ATXN2* están afectadas por el haplotipo circundante y la secuencia interna de la expansión de CAG.

Como resultado de este trabajo, se propone un umbral de susceptibilidad genética a inestabilidades de la secuencia repetitiva de CAG del gen *ATXN2*. Este umbral se confirmó experimentalmente, dado que los alelos largos mostraron inestabilidad somática e intergeneracional, proclives a expandirse, alcanzar el rango patogénico, y asociarse a neurodegeneraciones como la ELA.

Se demostró por primera vez en SCA2, que factores genéticos como las interrupciones de CAA son claves para estabilizar la secuencia de CAG. Así se comprobó que la pérdida de estas interrupciones es primordial para que ocurran expansiones nuevas en la secuencia de CAG del gen *ATXN2*.

Esta investigación es pionera al introducir la epigenética en la SCA2. Adicionalmente se desarrollaron, por primera vez, dos técnicas moleculares para estudiar el patrón de metilación del promotor del gen *ATXN2*. De igual modo, se detectaron, por primera vez a nivel mundial, patrones diferenciales de metilación asociados a la SCA2 y se comprobó que afectaban la severidad de la SCA2 y de la enfermedad de Machado Joseph o SCA3. Todo esto tiene implicaciones terapéuticas y preventivas para la SCA2 y otras enfermedades conectadas con alteraciones genéticas en *ATXN2*.

El **aporte teórico**: En este estudio se hace pública la frecuencia génica de las expansiones de CAG en *ATXN2* de parte de la población saludable cubana, lo que posibilita su utilización para el análisis comparativo entre poblaciones. Estas frecuencias génicas sirvieron como modelo predictivo para inferir la frecuencia de nuevas mutaciones causantes de neurodegeneraciones. Este modelo teórico, es una explicación de las elevadas tasas de prevalencias de la SCA2 en Cuba.

A la vez, a partir del estudio molecular de las trasmisiones de la secuencia de CAG y el análisis de su estructura, permitió proponer mecanismos moleculares del surgimiento de enfermedades neurodegenerativas distintas a la SCA2, a partir de variantes de riesgo en la población saludable. Este conocimiento contribuye a la conceptualización de entender el modo de herencia de enfermedades que aún cuando no muestran un patrón de herencia definido, sí están influenciadas o determinadas por factores genéticos.

El descubrimiento de la metilación de ADN del promotor *ATXN2* permite proponer nuevas estrategias para intervenir en la cascada degenerativa de SCA2 y es extrapolable a otras enfermedades en las que *ATXN2,* o sus productos génicos, están involucrados. Estos hallazgos tienen implicaciones fisiológicas para este gen, con función desconocida, porque en dependencia de su estatus de metilación en distintos tejidos puede inferirse la función de *ATXN2* y la delineación de cierto fenotipo o carácter fenotípico.

El **aporte práctico:** El conocimiento derivado de esta investigación mejora el asesoramiento genético en la SCA2. También sirve para orientar a la población saludable que puede ser proclive a inestabilidades del triplete reiterado de CAG de *ATXN2*. Por tanto, el asesoramiento se puede extender a otros trastornos neurodegenerativos que en nuestro país no se contemplaban como enfermedades genéticas y relacionadas a *ATXN2* (Ej. ELA y EP). Al demostrar experimentalmente y mediante meta-análisis que los alelos largos son una variante de riesgo de inestabilidad, y un factor de riesgo para neurodegeneraciones, se ilumina el camino para entender la base de las enfermedades comunes. La demostración de factores modificadores de la severidad de la SCA2, tales como la metilación del promotor *ATXN2*, sirve para el correcto diseño de ensayos clínicos. Por. ej.: la estratificación de las distintas cohortes

de pacientes SCA2 de acuerdo al factor modificador, puede mejorar el efecto de una terapia. En la investigación de la metilación, reportamos por primera vez dos técnicas moleculares de alta sensibilidad y exactitud, las cuales sirven como herramientas analíticas para el estudio de la SCA2, SCA3 y otras enfermedades. Estas técnicas, en la medida que se conozcan y se profundice en el rol de la metilación de *ATXN2,* serán de utilidad para la atención y manejo de la SCA2 y otras enfermedades.

La tesis está dividida en: Lista de Abreviaturas, Síntesis, Tabla de Contenidos, Introducción, Marco Teórico, Resultados y Discusión (Selección de artículos del aspirante sobre el tema de tesis), Discusión general, Consideraciones Finales, Conclusiones y Recomendaciones, Referencias bibliográficas y Anexos.

Estos resultados fueron subvencionados en parte por la Academia Mundial de Ciencias para el desarrollo de la ciencia en los países del sur –TWAS-, y un premio de QIAGEN-BIOLABS. Los resultados contenidos en esta tesis han sido sometidos a la discusión científica en más de 20 eventos científicos internacionales y obtuvieron los siguientes **premios científicos:** Rafaél Estrada, IBRO, IBRO-SFN, QIAGEN, ENS-2009, MDS-2010, MDS-2011,2012, 2014, ACC-2012 (1) nivel provincial, 2 ACC-2013 a nivel provincial y 2 nacional, 3 Premios Anual de Salud 2013 a nivel provincial y 2 nacional.

MARCO TEÓRICO

II

2. MARCO TEÓRICO

2.1 Gen Ataxin-2 y su producto génico

Desde su descubrimiento, que fue el de la mutación, a este gen se le denomina *SCA2* aunque de acuerdo a la nomenclatura HUGO debe ser *ATXN2 o ATX2*.

El gen Ataxin-2 se localiza en el *locus* 12q (Gispert *et al*., 1993) y existen datos no publicados que indican que la expansión puede interactuar genéticamente con otras variantes en otros *loci* y afecta la expresión de genes vecinos en el mismo *locus* (Rub *et al*., 2013).

Inicialmente se sugirió que el promotor de ataxin-2 estaba demetilado, explicando su origen así como la fortaleza de esta secuencia (Aguiar *et al*., 1999).

2.1.1 Gen ataxin-2 (*ATXN2/SCA2*) y su expresión fenotípica

Las características de la SCA2 son la atrofia y pérdida de las células de Purkinje del cerebelo. Clínicamente manifestándose como déficit de la coordinación motora que afecta la mirada, la prosodia, la marcha y la estabilidad postural (Lastres-Becker *et al*., 2008).

La SCA2 es causada por la expansión del dominio de poliglutaminas (polyQ) en la proteína ataxina-2 por encima de 34 unidades de tripletes de CAG (Imbert *et al*., 1996; Pulst *et al*., 1996; Sanpei *et al*., 1996). Este dominio tiene una longitud de 22 unidades en más del 90% de los individuos debido a mecanismos de selección positiva, y usualmente codificado por la estructura (CAG)8CAA(CAG)4CAA(CAG)8 (Yu *et al*., 2005). La inestabilidad de esta estructura parece comenzar con la pérdida de la interrupción de CAA (Choudhry *et al*., 2001).

Las expansiones de ≥27CAG con una interrupción se asocian con riesgo a desarrollar Esclerosis Lateral Amiotrófica (ELA) así como Parálisis Supranuclear Progresiva (PSP) (Elden *et al*., 2010; Gispert *et al*., 2012; Lee *et al*., 2011; Ross *et al*., 2011).

Por encima de las 32 unidades, las expansiones de CAG raramente preservan la interrupción de CAA y se manifiesta como Parkinsonismo respondedor a la L-dopa o como ELA (Charles *et al*., 2007; Daoud *et al*., 2011; Van Damme *et al*., 2011). Usualmente, esta interrupción no está en expansiones mayores, resultando en una secuencia pura de tripletes de CAG y manifestándose como la neurodegeneración multisistémica SCA2 (Auburger., 2012).

2.1.2 Producto génico (Ataxina-2)

Ataxina-2 existe en varias isoformas resultante de empalmes alternativos, y además existen proteínas relacionadas a ataxina-2 -ATXN2L- desconociéndose sus funciones (Affaitati *et al*., 2001; Figueroa & Pulst, 2003).

Tradicionalmente, se ha pensado que las expansiones polyQ actúan mediante mecanismos de ganancia de una función tóxica, tales como potenciación de la agregación de la proteína mutada lo cual lleva a la formación de cuerpos de inclusión y al secuestro de proteínas. Sin embargo, para ataxina-2 hay evidencias que indican mecanismos adicionales (Auburger., 2012; Huynh *et al*., 2000). La ataxina-2 mutada se une al canal liberador de calcio intracelular InsP(3)R1y altera la excitabilidad neuronal (Liu *et al*., 2009).

Curiosamente, existe evidencia de que ataxina-2 tiene un papel importante en otras enfermedades neurodegenerativas. La ausencia de ataxina-2 tiene un potente efecto mitigador de la apoptosis en modelos animales y celulares que sobre-expresan TDP-43, proteína clave en la ruta patogénica de ELA (Elden *et al*., 2010).

La deficiencia de ataxina-2 modula el curso de la SCA1 y la SCA3 en modelos animales de *Drosophila,* útiles para entender la patogénesis de esas enfermedades (Al-Ramahi *et al*., 2007; Lessing & Bonini, 2008). Es muy probable que las interacciones directas de ataxina-2 con ataxina-1 o ataxina-3 medien estos efectos, pero se ha demostrado que las interacciones de ataxina-2 con el ARN es clave (Elden *et al*., 2010; Lim *et al*., 2006).

La expresión de ataxina-2 es muy ubicua, siendo fuerte en poblaciones neuronales con grandes citoplasmas, en aquellas deprivadas de nutrientes, en las que se induce estrés y en las senescentes (Huynh *et al*., 1999).

La localización sub-celular de ataxina-2 es eminentemente en el retículo endoplasmático rugoso (RER) y con los poli ribosomas, con baja presencia en membrana citoplasmática y núcleo de algunas células. Bajo estrés esta se relocaliza a estructuras granulosas del citosol, llamadas gránulos de estrés (GE), donde los ARNm se mantienen en un estado inactivo traduccionalmente (Hallen *et al*., 2011, Nonhoff *et al*., 2007, Satterfield & Pallanck, 2006, van de Loo *et al*., 2009).

En la membrana plasmática, ataxina-2, parece modular la internalización y señalización del receptor de tirosina cinasas mediante interacciones de su propios dominios ricos en prolina, con los dominios SH3 de endofilina/Src y otras proteínas (Nonis *et al*., 2008; Ralser *et al*., 2005b). En este contexto ataxina-2 puede interactuar, además, con Parkina, una ubiquitina ligasa neuroprotectora, la cual se conoce modula la endocitosis del receptor del factor de crecimiento epidérmico, que además tiene riesgo para la Enfermedad de Parkinson (Corti *et al*., 2011; Fallon *et al*., 2006; Huynh *et al*., 2006). En el núcleo la ataxina-2 podría actuar como proteína acompañante y modificadora del factor de transcripción ZBRK1 que contiene un dominio KRAB, y este factor está implicado en la respuesta al daño del ADN. En el RER, poli ribosomas y los GE, la

ataxina-2 interactúa con la proteína unidora de poliadenilatos PABPC1, a través de la asociación con su dominio PAM2 y del MLLE, en competencia con otros moduladores que contienen dominios PAM2. De manera que modularía finamente la traducción y degradación de ARNms (Albrecht & Lengauer, 2004b; Kozlov *et al*., 2010; Ralser *et al*., 2005a; Satterfield & Pallanck, 2006; Siddiqui *et al*., 2007).

Otros motivos relevantes en ataxina-2 relacionados con el procesamiento de ARN son el LSm y el asociado LSm-AD, de los cuales se conoce que aparecen en proteínas cruciales para la degradación del ARN o el empalme, y son importantes para la citotoxicidad adquirida de ataxina-2 (Albrecht *et al*., 2004b; Albrecht & Lengauer, 2004a; Ng *et al*., 2007; Satterfield & Pallanck., 2006; Tharun., 2009). Es interesante notar que los ortólogos de ataxina-2 inferiores a levaduras, conservan la presencia del motivo PAM2 así como la interacción con otras proteínas LSm (Elden *et al*., 2010; Gavin *et al*., 2002; Satterfield & Pallanck, 2006; Swisher & Parker, 2010).

La relevancia de ataxina-2 en el procesamiento de ARN es además resaltada por su interacción con A2BP1, el cual es un factor de empalme importante que regula la excitabilidad neuronal y está implicado en autismo (Gehman *et al*., 2011; Shibata *et al*., 2000; Voineagu *et al*., 2011). En *Drosophila*, el ortólogo de ataxina-2 muestra una fuerte interacción con los genes me31B y ago1, implicándola en la ruta de los microARNs importante para la habituación a largo plazo y asociada a la plasticidad sináptica (McCann *et al*., 2011). En *C. Elegans*, el ortólogo de ataxina-2 previene la traducción de manera inapropiada, en complejo con la proteína unidora de poly(A) y a través de la acción del dominio KH de la proteína MEX-3 (Ciosk *et al*., 2004). En levadura, su ortólogo Pbp1p promueve la formación de GE e inhibe el crecimiento celular, posiblemente a través del reclutamiento de la nucleasa poly(A) a los complejos de ribonucleoproteínas (Mangus *et al*., 2004a,b; Swisher & Parker, 2010). Por tanto, la

alteración de ambos procesos: traducción del ARN y la señalización trófica probablemente cooperen de manera decisiva en la obesidad, fertilidad, mala señalización de insulina y desajuste de la excitabilidad observada en modelos *knock out* o deficientes de ataxina-2 (Huynh *et al*., 2009; Lastres- Becker *et al*., 2008a).

2.2 Epigenética (metilación del ácido desoxirribonucleico)

2.2.1 Metilación del ácido desoxirribonucleico: dinámica del control transcripcional

La metilación de la citosina representa uno de los mecanismos epigenéticos más importantes para la regulación génica, y consiste en la adición del grupo metilo en la posición 5' del anillo pirimidínico de la Citosina, lo cual es mediado por las enzimas metiltransferasas (DNMTs) usando a la S-adenosilmetionina (SAM) como compuesto donante del grupo metilo (Oakeley., 1999). La metilación del ADN en la posición 5' de la citosina (5 metilcitosina) típicamente ocurre en el contexto del dinucleótido CpG, y la metilación de las secuencias de CpG pueden inducir modificaciones conformacionales inhibiendo el acceso de la maquinaria transcripcional a las regiones promotoras, alterando los niveles de expresión génica. La hipermetilación de promotores se asocia con el silenciamiento génico y la demetilación con expresión génica (Oakeley., 1999).

Para explicar el papel de la metilación en la disminución de la actividad transcripcional se han propuesto varios mecanismos. En un primer modelo se propuso que los residuos de citosina podrán interferir directamente en la interacción de los factores de transcripción a sus sitios de unión en el ADN, una vez que el grupo metilo se proyecta hacia el surco mayor de la doble hélice del ADN. Un segundo modelo propone que la metilación de los CpG tiene una consecuencia directa en el posicionamiento nucleosomal precediendo a la formación de esta estructura. De esta forma, una

estructura nucleosómica compacta silencia mucho más eficientemente la transcripción que una estructura cromosómica convencional (Esteller., 2005).

La información almacenada en los islotes de CpG hipermetilados es en parte, interpretada por proteínas unidoras de estos residuos metilados (*en inglés Methyl Binding Domain Protein* -MBDs-). Las MBDs son importantes traductoras entre la metilación del ADN y los genes modificadores de histonas que establecen un ambiente cromatínico transcripcionalmente inactivo (Esteller., 2007). Las MBDs pueden reclutar complejos proteicos que contienen co-represores y desacetilasas de histonas (HDACs) y con base en esta interacción se propuso que la unión de estos complejos al ADN conllevaría a un cambio en la estructura de la cromatina impidiendo la transcripción (Yang *et al*., 2001).

2.2.2 Modificaciones de histonas

El estado de la cromatina representa otro importante modulador de perfiles de la expresión génica. La cromatina existe en una forma no condensada, transcripcionalmente activa (eucromatina) o en su forma condensada inactiva (heterocromatina). Los cambios conformacionales en las histonas o las modificaciones en la que el ADN se enrolla en los octámeros de histonas de los nucleosomas pueden bloquear o facilitar el acceso de la maquinaria transcripcional a las regiones promotoras de los genes, llevando al silenciamiento o activación respectivamente (Luger *et al*., 1997).

La partícula nuclear o *core* del nucleosoma consiste en segmentos de ~147pb de ADN enrollados alrededor de los octámeros de histonas, los cuales consisten de dos copias de histonas núcleos H2A, H2B, H3 y H4. Las partículas *core* están conectadas por segmentos de "ADN *linker*" o unidor e "histonas-*linkers*" o unidoras tales como la H1, y son responsables de la compactación de la cromatina (Berger *et al*., 2007). Las

modificaciones de las colas de histonas incluyen acetilación, metilación, fosforilación, ubiquitilación, sumoilación y otras modificaciones post-transcripcionales.

La acetilación de las colas de histonas es la modificación más estudiada se asocia con la relajación de la cromatina y activación transcripcional, mientras que la desacetilación se relaciona con una forma más condensada de la cromatina y represión transcripcional (Luger *et al*., 1997). La acetilación ocurre en los residuos de lisina de la región amino-terminal de las colas de las histonas, neutralizando la carga positiva de las colas de histonas y disminuyendo su afinidad por el ADN. Como consecuencia, la acetilación de histonas altera la conformación nucleosomal, lo que incrementa la accesibilidad de las proteínas reguladoras de la actividad transcripcional al ADN en las cromatinas (Berger *et al*., 2007).

Las enzimas Histonas Acetil Transferasas (HATs) catalizan la acetilación de los residuos de lisina en las colas de histonas, mientras que la desacetilación de histonas es mediada por las Histonas desacetilasas (HDACs). Otra modificación muy estudiada es la metilación en las lisinas (Lys/K) y argininas (Arg/H). Esta modificación puede asociarse con la condensación o relajación de la estructura de la cromatina, dado que las colas tienen varios sitios para metilación en cada cola de histona existen varias combinaciones (Chouliaras *et al*., 2010). El efecto exacto de todas las combinaciones posibles de modificaciones post-traduccionales de las colas de histonas ("conocido como "código de histonas") sobre la expresión génica es complicado y no se entiende del todo (Martin *et al*., 2005). Algunos ejemplos son la metilación en las lisinas (Lys/K) y argininas (Arg/H) H3K4, H3K36 y H3K79 y generalmente se vincula con genes transcripcionalmente activos, otra es la H3K9 di o trimetilada, y la H3K27, H4K20 metiladas, las cuales se consideran marcas represoras (Chouliaras *et al*., 2010).

Las proteínas lisina metiltransferasas (PKMTs) y las arginina metiltransferasas (PRMTs) son dos enzimas responsables de añadir marcas de metilo a las histonas. Ambas PKMTs y PRMTs requieren S-adenosil-Metionina (SAM) como donador del grupo metilo para las reacciones de metilación de las histonas (Yost *et al*., 2011). Además, la posición relativa de los nucleosomas respecto al sitio de inicio de la transcripción es otro regulador fundamental de la transcripción, en este sentido el posicionamiento del nucleosoma está fuertemente influenciado tanto por la metilación del ADN y las modificaciones de las colas de las histonas. En la mayoría de los casos las regiones cadena arriba cercanas a los ATG de inicio libres de nucleosomas, representan genes activos permitiendo el acceso a la maquinaria transcripcional. Sin embargo, los nucleosomas pueden actuar como barreras a la transcripción para algunos genes y deben desplazarse para que ocurra la activación del gen (Li *et al*., 2007).

2.2.3 Efectos de posición en los cromosomas y control remoto de la transcripción

Los genes, elementos reguladores y el ADN repetitivo están interpuestos, formando un mosaico de regiones condensadas y abiertas en el cromosoma. Estos elementos pueden ser regulados remotamente así como en dependencia de su posición cromosómica. Así, la proximidad de los distintos tipos de cromatina puede influenciar la expresión de genes positiva (proximidad a '*enhancer'* o intensificador) o negativamente (proximidad a '*silencer'* o silenciador) (Schluth-Bolard *et al*., 2011). Por otra parte, los cambios en la estructura cromosómica pueden regular en *trans* a los elementos regulatorios del ADN deslocalizando factores cromatínicos específicos. Después de un reordenamiento, un gen relocalizado en la vecindad de una región heterocromática puede entrar en silenciamiento, trayendo consigo un patrón variegado de expresión como consecuencia de un efecto de posición (conocido como PEV en inglés *position effect variegation*). Los distintos tipos de efecto de posición caen dentro de los globalmente conocidos CPE

del inglés *Chromosomal Position Effect* o efecto de posicionamiento de los cromosomas EPC. Dentro de estos la proximidad a los telómeros recuerda el clásico EPC generando silenciamiento de genes próximos, conociéndose como TPE (del inglés *Telomeric Position Effect* o efecto de posicionamiento de los telómeros) (Schluth-Bolard *et al*., 2011).

En alguna medida la identidad de los dominios de la cromatina se mantienen por factores como los reguladores en *cis*, mientras que los límites borrosos o "*fuzzy boundaries*" y los *insulators* o aisladores limitan la influencia de una región sobre la otra. Los reguladores en *cis* pueden ser secuencias cortas, en tanto que las "*fuzzy boundaries*" o límites borrosos, definen el límite entre la eu y la heterocromatina, lo cual no está precisamente definido, como su nombre lo indica, y puede cambiar en el tiempo (Schluth-Bolard *et al*., 2011).

Los intensificadores o '*enhancers*' tienen el potencial de activar varios genes en regiones cromosómicas, por lo que su acción debe restringirse previniendo la activación inespecífica de genes. Esta actividad limitadora se realiza por secuencias de ADN especializadas, definidas como bloqueadores de *'enhancer'* o *'insulators'* o aisladores los cuales interfieren con la capacidad de unirse a un promotor específico, cuando el intensificador está situado entre dos secuencias promotoras. El potencial de transcripción de un gen puede además ser susceptible al silenciamiento heterocromático originado en su cromatina.

Los aisladores pueden actuar como barreras para la heterocromatinización (Schluth-Bolard *et al*., 2011). Además unen proteínas específicas con actividad bloqueadora de la actividad de los intensificadores.

La proteína CTCF muestra esta actividad en vertebrados y se estima que existen 39000 sitios de unión a CTCF en el genoma en una relación casi de 1:1 con los genes y los

factores de transcripción (Schluth-Bolard *et al*., 2011). Las regiones en el genoma sin estos sitios regularmente corresponden a grupos de genes co-expresados, mientras que aquellas enriquecidas muestran múltiples regiones promotoras alternativas. Los sitios de unión a CTCF además flanquean las expansiones de CTG/CAG de varios loci asociados con enfermedades humanas (Schluth-Bolard *et al*., 2011).

2.2.4 Mecanismos epigenéticos mediados por ácido ribonucleico

Existe evidencia acumulada que apoya la función de los ARN no codificantes en la epigenética así como de la trasmisión de información a través de estos ácidos nucleicos. Los micro-ARNs o micro ácido ribonucleicos (mi-ARNs) son un grupo de ARNs no codificantes de ~22 nucleótidos que se unen a las regiones 3'no traducidas (3'-UTR) de los ARNm y median la regulación post-transcripcional causando la degradación o la inhibición traduccional, dependiendo del grado de complementariedad de la secuencia. Los mecanismos mediados por mi-ARNs son considerados comúnmente como reguladores epigenéticos (Sato *et al*., 2011). Los mi-ARNs tienen como diana al 60% de todos los genes (Sato *et al*., 2011), existiendo una compleja red de interacciones entre los mi-ARNs y otros mecanismos epigenéticos, tales como la metilación del ADN y la modificación de las histonas, con el propósito de organizar todos los perfiles de expresión génica (Moazed *et al*., 2009). Los ARNs cortos interferentes (si-ARNs) y los (piARNs) que interactúan con PIWI (RNasa con dominio de tipo H) son otras clases de ARN cortos (Sato *et al*., 2011). Los si-ARNs son ARNs cortos de 20-25nt que juegan un papel en el silenciamiento pos-transcripcional (degradación del ARN o arresto traduccional) y en las rutas de silenciamiento dependientes de la cromatina, involucrando el ensamblaje de complejos de ARN cortos en transcriptos nacientes (Moazed *et al*., 2009). Los pi-ARNs tienen como longitud promedio 24-31nt y están vinculados con el silenciamiento génico post-transcripcional y en menor grado con el

silenciamiento génico dependiente de las cromatinas (Moazed *et al*., 2009). Además, los ARNs largos no codificantes lnc-ARNs (del inglés *long non coding RNA*) son una clase de transcriptos largos de ARN de longitud mínima de 200nt que no codifican para proteínas, y participan en el silenciamiento específico de genes a través del remodelamiento de la cromatina, la organización nuclear, la formación del dominio de silenciamiento y el control preciso sobre los genes que entran a compartimentos silentes (Moazed *et al*., 2009).

2.2.5 Avances tecnológicos en los estudios epigenéticos

En los últimos años, el progreso en las técnicas bioquímicas usadas para estudiar las modificaciones epigenéticas ha sido impresionante. Esto es particularmente así para el estudio de la metilación del ADN. Esta modificación fue originalmente estudiada a nivel de genoma analizando el contenido total de 5-metilcitosina del ADN por hidrólisis seguida de cromatografía (Saluz *et al*., 1993, Oakeley., 1999). Mucho más reciente, el desarrollo de la electroforesis capilar hizo este ensayo más rápido, barato y mucho más sensible, sin embargo, aunque permite una cuantificación precisa, tiene limitaciones en cuanto a que se necesitan grandes cantidades de ADN y equipos muy caros. Por otra parte, sólo permite determinar el nivel total de metilación, sin obtener información relacionada con secuencias específicas.

Hoy en día, se usan tres estrategias diferentes las cuales se combinan, generándose un gran número de ensayos orientados a estudiar los diferentes aspectos de la metilación del ADN y permitiendo la determinación de patrones de metilación secuencia-específicos: digestión con endonucleasas, con enzimas sensibles a la metilación, modificación química con bisulfito y purificación de fracciones genómicas metiladas con anticuerpos específicos para metilación.

Aunque en el presente las técnicas basadas en la modificación con bisulfito representan al estado actual del conocimiento en el estudio de la metilación del ADN secuencia específico, las enzimas sensibles a metilación han encontrado nuevas aplicaciones en los análisis de metilación a nivel del genoma completo.

El bisulfito sódico reacciona selectivamente con las citosinas no metiladas convirtiéndolas a uracilo, dejando aquellas metiladas como citosinas no modificadas. Esta reacción es altamente dependiente del estado simple cadena del ADN y no ocurre en ADN de doble cadena, por lo que se requiere de una desnaturalización inicial del ADN. Este es un paso crucial del método, dado que la desnaturalización parcial puede causar transformación incompleta de ciertas citosinas no metiladas, consecuentemente esto crea artefactos tales como falsos positivos. La reacción es el fundamento base para diferenciar el ADN metilado del no metilado, pero se necesita combinarlo con otros métodos para valorar el estado de metilación de una secuencia dada.

De manera general las estrategias asociadas con el bisulfito requieren amplificación por PCR del ADN transformado (el cual incorpora T por U) y el diseño de oligos específicos para el estado de metilación (Fuso *et al*., 2006). Sin embargo, el método de análisis de los productos amplificados por PCR puede variar dependiendo del grado de especificidad y detalle de la metilación requerida, lo cual ha dado lugar una gran cantidad de técnicas.

El análisis de restricción combinado con modificación por bisulfito (COBRA) fue considerado en los inicios como referente (ya hoy no es así) y combina el corte enzimático y la modificación con Bisulfito (Xiong *et al*., 1997). También se ha desarrollado una estrategia que usa la espectrometría de masa (Schatz *et al*., 2004).

En las técnicas de PCR se usa también la *High Resolution Melting* (HRM) o análisis de alta resolución de las curvas de disociación del ADN convertido con bisulfito (Wojdacz

et al., 2007). El ADN modificado con Bisulfito puede analizarse directamente mediante PCR específica para metilación (MSP en inglés *Methylation Specific PCR*) (Herman *et al*., 1996). Una versión cuantitativa de la MSP es la *MethyLight*, la cual usa oligos fluorescentes y equipos de PCR en tiempo Real (Eads *et al*., 2000). Se han desarrollado varias tecnologías para el estudio de la metilación a nivel de genoma pero sólo nos enfocaremos a las usadas para análisis de metilación a nivel de genes, y de estas a las dos más usadas actualmente.

2.2.5.1 Técnica de PCR específica para metilación

La *MSP* es sencilla, sensible (0.1% del ADN metilado en una secuencia rica en CpG), rápida y requiere pocas cantidades de ADN, permitiendo el estudio de muestras embebidas en parafina, o tejidos micro-diseccionados. Se basa en el diseño de oligos que contienen dinucleótidos CpG compatibles con las dos versiones del ADN (metilado o no metilado) el cual ha sido previamente tratado con bisulfito, siendo este último el paso complejo de este procedimiento. Es importante considerar que el ADN convertido no es auto complementario, por lo que los oligos diseñados para amplificar la cadena con polaridad positiva de una secuencia son diferentes de los de la cadena con polaridad negativa (Herman *et al*., 1996).

2.2.5.2 *MethyLight*

Este método específicamente usa tecnología TaqMan®, la cual se basa en tres oligonucleótidos: los dos F y R, y una sonda fluorogénica de hibridación al ADN, lo cual brinda la oportunidad para la detección de la metilación a varios niveles, tales como la amplificación y/o la etapa de unión de la sonda al ADN molde. La detección de la fluorescencia causa un aumento de la sensibilidad dado que detecta un único alelo metilado en 10^5 no metilados (Eads *et al*., 2000).

RESULTADOS Y DISCUSIÓN

III

3. RESULTADOS Y DISCUSIÓN

Los resultados que conforman este capítulo de la tesis están recogidos en una serie de artículos científicos que aparecen a continuación.

Cada uno esta precedido por un resumen de los objetivos, razonamientos, hipótesis y premisas que guiaron la investigación, así como una breve descripción del contenido fundamental del mismo que facilitará seguir el orden de pensamiento. Con ese mismo propósito los artículos aparecen en orden lógico y cronológico en la medida que se ha desarrollado esta línea de investigación. Muchos de estos materiales son contemporáneos en su publicación.

Seis artículos están contenidos en esta tesis (Tabla 1). Para mejor comprensión de este documento hemos decidido organizarlos de acuerdo al esquema presentado en la Fig. 1. El primer artículo sirve de marco para generar las dos hipótesis específicas para responder a la hipótesis principal propuesta en la introducción, y a cada serie de artículos le sigue una discusión particular, la cual en el tema de origen de la SCA2 la hemos conformado por dos artículos que son consolidaciones y discusiones prácticas de los resultados presentados en los artículos previos (I, II, III). En el caso de la metilación hemos desarrollado una discusión específica no en forma de publicación.

Tabla 1.

4. SELECCIÓN DE ARTÍCULOS		**Factor de impacto de la revista**	**Total de citas en Google Académico (autocitas)**	**Total de citas en Scopus (autocitas)**
I.	Velázquez Pérez L, Sánchez Cruz G, Santos Falcón N,… **LAFFITA-MESA JM** et al. Molecular epidemiology of spinocerebellar ataxias in Cuba: Insights into SCA2 founder effect in Holguin. Neuroscience Letters 454 (2009) 157–160.	2.23	46(30)	28(10)
II.	**LAFFITA-MESA JM**, Velázquez-Pérez LC, Cruz-Mariño T et al.Unexpanded and Intermediate CAG Polymorphisms at SCA2 Locus (*ATXN2*) in the Cuban Population: Evidences About the Origin of Expanded SCA2 Alleles. European Journal of Human Genetics (2012) doi:10.1038/ejhg.2011.154.	4.38	16(2)	3(2)
III.	**LAFFITA-MESA JM,** Rodríguez Pupo JM, Moreno Sera R et al. (2013) De Novo Mutations in Ataxin-2 Gene and ALS Risk. PLoSONE 8(8): e70560. Doi:10.1371/journal.pone.0070560	4.09	--	1(0)
IV.	**LAFFITA-MESA JM***, Almaguer-Mederos LE, Vázquez Mojena Y, Kourí Vet al. Large normal alleles and SCA2 prevalence: Lessons from a nationwide study and analysis of the literature. 2013, doi:10.1111/cge.12221.	4.25	--	--
V.	Cruz-Mariño T, **LAFFITA-MESA JM***, Gonzalez-Zaldivar Y et al (2013). Large Normal and Intermediate Alleles in the Context of SCA2 Prenatal Diagnosis. J Genet Counsel. 10.1007/s10897-013-9615-1 ***1era Autoría compartida.**	1.77	--	--
VI.	**LAFFITA-MESA JM**, Bauer Peter O, Kourí Vivian et al. Epigenetics DNA-Methylation in the core ataxin-2 gene promoter: Novel physiological and pathological implications. (2011). Hum Genet Volume 131, (4), 625-638.	5.07	13(1)	9(1)

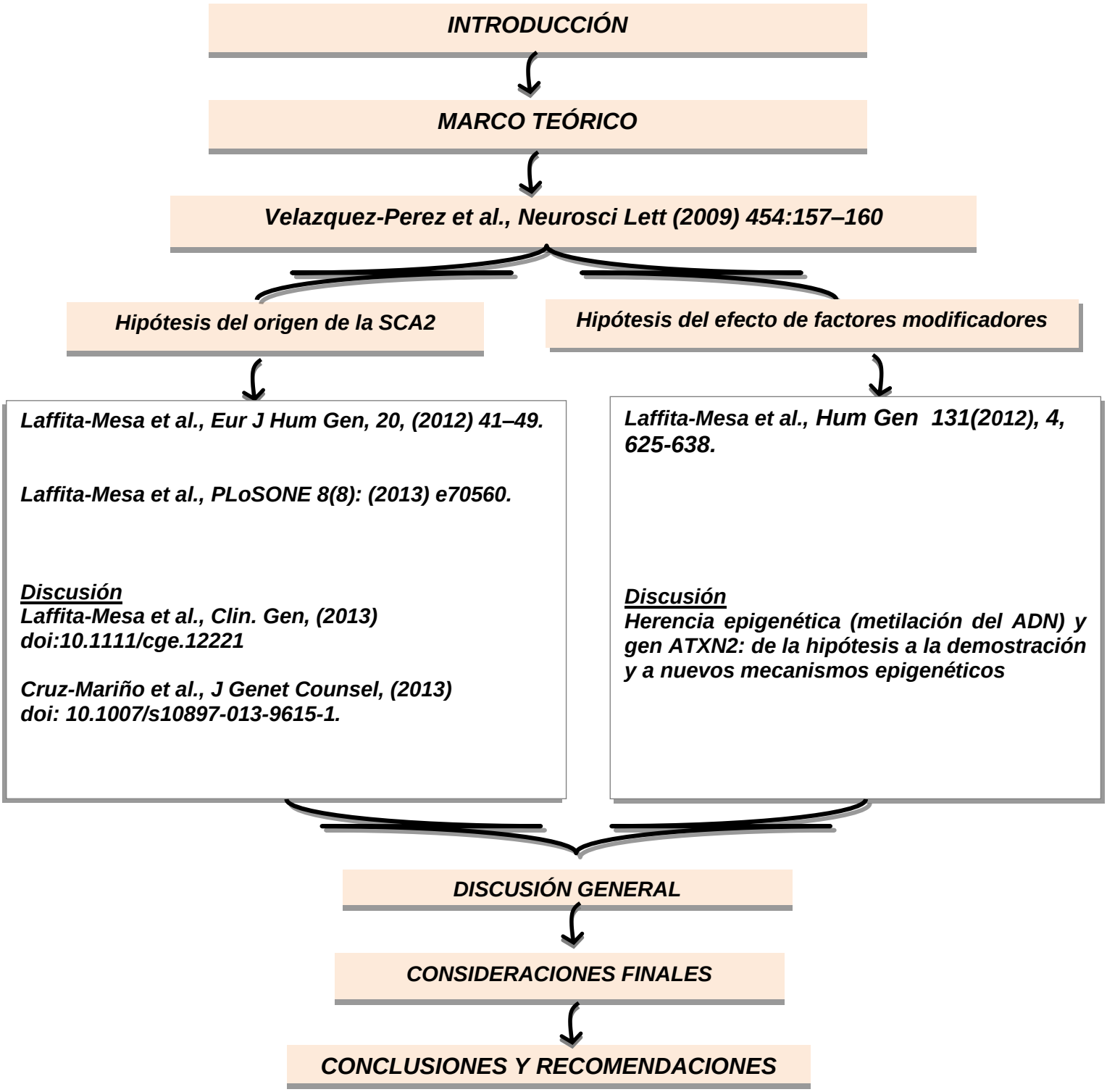


Figura 1. Unidad estructural de los artículos.

3.1: Epidemiología de las Ataxias Espinocerebelosas en Cuba: Postulados sobre el origen de la SCA2 y modificadores del fenotipo

Publicado en: Neurosci. Lett (2009) 454:157-160 (Factor de Impacto: 2.23)

El efecto fundador es la sobre-representación de un alelo específico en uno o varios *loci* en una nueva población, lo cual surge a partir de la separación de un reducido número de individuos los cuales tienen un contenido génico no representativo de la población parental. Este proceso ha sido ampliamente aceptado como origen de la SCA2 en Cuba. Sin embargo, existen tres posibles modelos: 1) expansión *de novo* estocástica, 2) expansión de CAG fundadora y 3) haplotipos predispuestos (Warby et al., 2009). Estos a su vez trascienden al origen y la existencia de la SCA2 en Cuba, y llegan a la pregunta misma de cómo surge la expansión del triplete en las poblaciones humanas.

Esta publicación es un acercamiento a varias cuestiones relacionadas con el origen de la SCA2 así como de la expresión fenotípica de la mutación SCA2 y de la variabilidad observada. Por tanto, este artículo presenta la epidemiología genética de SCA2 en Cuba y algunas características fenotípicas y su objetivo fundamental fue generar nuevas hipótesis, las cuales hoy vemos relacionadas con las seis preguntas fundamentales planteadas en la introducción de esta tesis.

El problema científico que guió a esta investigación fue: *"el desconocimiento de la prevalencia de SCA2 y otras ataxias hereditarias en el país"* esto a su vez partió de la existencia de estudios aislados (a nivel internacional) que brindaban un cuadro epidemiológico inexacto de estas enfermedades. Esto limitaba la correcta caracterización genotípica y fenotípica de la enfermedad, y los investigadores se centraban en la fenomenología de casos aislados, lo cual fertilizó hipótesis incorrectas acerca de estas esferas de la SCA2, y que han podido permear la perspectiva actual de la enfermedad.

Los resultados de la publicación derivan del único estudio epidemiológico nacional de ataxias hereditarias, a nivel internacional. Se identificaron 753 pacientes y 7173 familiares asintomáticos con diferentes ataxias, agrupados en 200 familias. El 86.79 % de los pacientes con ataxias autosómicas dominantes tenían SCA2. En la provincia de Holguín, la prevalencia promedio de esta enfermedad fue de 40.18 por 100 000 habitantes con un valor significativo de 141.66 por 100 000 en el municipio de Báguanos.

El rango de expansiones no patogénicas fue 13-31, siendo la más común 22CAG (76%). El 19% fueron variantes largas que tuvieron entre 23-31CAG y las cortas (de 13-21CAG) fueron el 5%. El rango de las expansiones patogénicas varió continuamente desde 32-79 unidades, siendo 37CAG la mutación más frecuente (Fig.2). No se encontró discontinuidad entre la distribución de alelos no patogénicos y los mutantes.

Las trasmisiones intergeneracionales fueron inestables a favor de las expansiones, y se observó un marcado efecto parental a favor de las trasmisiones paternas (Fig.3).

La expansión anómala del CAG explicó el 80% de la variabilidad fenotípica. Además, se observó anticipación genética con y sin aumento intergeneracional de la expansión del segmento de CAG.

Relacionado al origen o introducción de la SCA2 se manejan dos hipótesis 1) mutación fundadora, y 2) haplotipo predisponente, este último entendemos hoy que es la hipótesis más probable. En esta se postula que: "*las elevadas prevalencias de SCA2 son el resultado de una alta frecuencia de variantes no patogénicas largas (alelos largos) con un haplotipo particular, y estas son el reservorio para alelos mutados, llevando al origen continuo de nuevos casos SCA2"*. Por tanto, la existencia de estas variantes largas es explicativa del sostenimiento de las tasas de prevalencia de SCA2. La hipótesis general fue propuesta por Takano et al., 1998 para todas las Ataxias Autosómicas

Dominantes causadas por expansiones de tripletes. Sin embargo, existen discrepancias en cuanto a su cumplimiento. Esto se trata en la serie de artículos que siguen.

Aquí también se plantea otra hipótesis relacionada con la "*existencia de factores modificadores del dueto: longitud del triplete CAG→fenotipo"* y para comprobar esto se introduce el estudio de la metilación del ADN. Por tanto, la presente investigación asume que la expresión del gen *ATXN2* (como factor intrínseco) determina la variabilidad fenotípica y la pleiotropía de la mutación SCA2, preguntas dos y cinco de la introducción.

Neuroscience Letters 454 (2009) 157–160

Contents lists available at ScienceDirect

Neuroscience Letters

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

Molecular epidemiology of spinocerebellar ataxias in Cuba: Insights into SCA2 founder effect in Holguin

Luis Velázquez Pérez*, Gilberto Sánchez Cruz, Nieves Santos Falcón, Luis Enrique Almaguer Mederos, Karel Escalona Batallan, Roberto Rodríguez Labrada, Milena Paneque Herrera, José Miguel Laffita Mesa, Julio C. Rodríguez Díaz, Raúl Aguilera Rodríguez, Yanetza González Zaldivar, Dany Coello Almarales, Dennis Almaguer Gotay, Humberto Jorge Cedeño

Centre for the Research and Rehabilitation of the Hereditary Ataxias, Holguin, Cuba

ARTICLE INFO

Article history:
Received 7 January 2009
Received in revised form 4 March 2009
Accepted 5 March 2009

Keywords:
SCA2
Cuban ataxia
Epidemiological survey
CAG repeats
Genetic anticipation

ABSTRACT

The objective of this study was to determine the prevalence of hereditary ataxias in Cuba, with a special focus on the clinical and molecular features of SCA2. Clinical assessments were performed by neurological examinations and application of the SARA scale. Molecular analyses of genes SCA1–3, SCA6, SCA17 and DRPLA identified 753 patients with SCA and 7173 asymptomatic relatives, belonging to 200 unrelated families. 86.79% of all SCA patients were affected with SCA2. In the Holguin province, the average population prevalence of SCA2 is 40.18×10^5 inhabitants, with the remarkable figure of 141.66×10^5 in the Baguanos municipality. The high prevalence of the SCA2 mutation in Holguin reflects most likely a founder effect. The stabilization of the prevalence along time suggests the existence of premutated chromosomes with pure CAG, acting as reservoir for further expansions. CAG repeat length correlated inversely with age at onset, accounting for 80% of the variability. Genetic anticipation was observed in the 80% of transmissions. Repeat instability was greater in paternal transmissions whereas CAG expansions without anticipation was observed in 10.97% suggesting the effect of CAA interruptions in the CAG segment, which decrease the toxicity of the abnormal ataxin-2, and/or other protective factors.

The spinocerebellar ataxias (SCAs) include several clinically, pathologically and genetically heterogeneous neurodegenerative disorders, characterized by the loss of balance and motor coordination due to dysfunction of the cerebellum and its afferent and efferent pathways [7]. SCAs are associated with at least 28 loci. The disease gene has been identified in 14 SCA types (SCA1–8, 10–14, 17, 27 and DRPLA) [13].

Most of the few existing epidemiological studies of hereditary ataxias have been performed in isolated geographical regions in families not large enough for linkage analysis. The collective worldwide prevalence is estimated in about 5–7 cases per 100,000 people, although higher figures have been reported in particular populations because of founder effects [24].

Several previous studies have reported a high prevalence of SCA2 in the north-eastern region of Cuba, specifically in the Holguin province [30,14,31]. However, the prevalence of this and other hereditary ataxias in the whole Cuban population has not been so far thoroughly determined.

The underlying mutation of SCA2 is an unstable expansion of a polyglutamine domain within ataxin-2. The size of the polyglutamine expansion has a strong influence on the age at onset as well as the severity of disease [10,17,20].

The aims of the present study were: (a) to determine the prevalence of autosomal dominant SCAs (SCA1–3, SCA6, SCA17, and DRPLA), as well as of recessive and sporadic ataxias and (b) to assess the correlation between the clinical features (age of onset, clinical severity, anticipation) and the length of the expansion and the, intergenerational instability in SCA2 patients.

The study was conducted by the National Center for the Research and Rehabilitation of the Hereditary Ataxias in the city of Holguin, which is the main referral center for these conditions in the country. In order to establish the epidemiological profile of inherited ataxias in Cuba, we conducted a survey to identify all patients with primary ataxic disorder and their asymptomatic relatives.

Clinical examinations and family history enabled the classification of patients as affected with SCA, recessive ataxia, sporadic ataxia or secondary ataxia. The latter group included patients with secondary ataxia due to alcoholism, neoplasias, autoimmune

* Corresponding author at: Carretera Central Km 5 1/2 via Havana, Holguin 80100, Cuba. Tel.: +53 724461693; fax: +53 24424090.
E-mail address: ataxia@cristal.hlg.sld.cu (L. Velázquez Pérez).

0304-3940/$ – see front matter © 2009 Elsevier Ireland Ltd. All rights reserved.
doi:10.1016/j.neulet.2009.03.015

Table 1
Distribution of SCA2 patients and prevalence rates by provinces, Cuba, 2006–2007.

Cuban provinces	SCA2 patients	Prevalence rate
Holguín	413	40.18
Las Tunas	63	11.82
Granma	23	2.75
Santiago de Cuba	24	2.30
Camaguey	14	1.76
Guantánamo	5	0.96
Ciudad Habana	27	1.04
Isla de la Juventud	2	2.48
La Habana	2	0.28
Cienfuegos	2	0.50
Matanzas	3	0.45
Total	578	6.57

or inflammatory diseases, vascular pathology, malformations and other non-genetic causes, and was excluded from further study.

The Ethics Committee of the National Center for the Research and Rehabilitation of the Hereditary Ataxias approved the research protocol and all studied patients signed an informed consent form after being explained the purpose and methods of the research.

Genomic DNA was isolated from peripheral leukocytes from all patients diagnosed with SCA, using standard protocols. PCRs and gel electrophoresis were carried out on SCA1, SCA2, SCA3, SCA6, SCA17 and DRPLA genes according to established techniques.

In the first phase of the study, 753 patients with hereditary ataxias were identified, belonging to 200 unrelated families. Of them, 666 (88.44%) had SCAs, 69 (9.16%) had recessive ataxias and 18 (2.39%) had sporadic ataxias. Out of the 666 patients with SCAs, 578 (86.79%) were found positive for an expansion in SCA2 gene, and 8 patients (1.2%) were positive for an expansion in the SCA3 locus. No mutations were detected in the remaining 80 (12.00%) patients with SCA; in particular, no expanded SCA1, SCA6, SCA7, SCA17 and DRPLA alleles were detected. We identified 7173 asymptomatic at-risk individuals in the SCA2 families, 2060 of which (75.90%) were first-degree relatives.

The second phase of the study was devoted specifically to further define the clinical manifestations and molecular epidemiology of SCA2, which was present in 11 out of the 14 provinces of the country. The distribution of SCA2 patients and prevalence rates by province are shown in Table 1. The highest concentration of SCA2 mutation was observed in the Holguin province, with 413 SCA2 patients and 1384 asymptomatic first-degree relatives. This value represents 71.45% of all Cuban SCA2 patients. The prevalence rate of SCA2 in the Holguin province is 40.18 per 100,000 inhabitants, but there are regions of the province where the prevalence reaches higher values with remarkable figure in Baguanos municipality (141.66 per 100,000 inhabitants) (Fig. 1).

All patients showed a cerebellar syndrome characterized by ataxic gait, cerebellar dysarthria, dysmethria and dysdiadochokinesia. In 558 of these (96.5%), gait ataxia was the first symptom of the disease. Patients exhibited abnormal tandem stance (95%), slow saccadic eye movements (91%), limited voluntary ocular movements (88%), loss of vibration sense (73%), areflexia or hyporeflexia (77%) and abnormal swallowing (76%). Autonomic abnormalities (urinary dysfunction, hypohidrosis, constipation, and sexual dysfunction) were presented in 77.68% of the cases.

The age at onset of SCA2 ranged from 3 to 79 years, with a mean of 32.96 ± 13.10. There was no significant difference in the age at onset between males; (mean 32.1 ± 13.6) and females (mean 32.7 ± 13.4). In order to assess the correlation between the size of the polyglutamine expansion and age of onset of symptoms, simple regression analyses were conducted and best fit were obtained using a $\log(y)$ transformation. A significant negative correlation of age of onset and polyglutamine expansion size ($r = 0.80$; $P = 0.0002$) was revealed. The duration of the disease from first symptoms to death varied from 6 to 39 years; mean 16.49 ± 7.11.

In order to assess the severity of the cerebellar ataxia and its progression from early to late stages, we used the Scale for the Assessment and Rating of Ataxia (SARA) [22] in a subsample 215 SCA2 patients with clinical and molecular diagnosis. The SARA score ranged from 4 to 39, mean 15.8 ± 7.25. There was a significant positive correlation between the SARA score, the size of the CAG repeat ($r = 0.4470$; $P \leq 0.001$) and disease duration ($r = 0.5366$; $P \leq 0.001$). ANOVA followed by post hoc Newman Keuls methods differentiated patients using SARA in three clinical stages defined by Klockgether [12]: stage 1 (slight gait ataxia): 14.09 ± 0.71; stage 2 (loss of independent gait): 19.66 ± 2.34, stage 3 (confinement to wheelchair or bed): $30.33 \pm 2{,}34$; $F = 13.94$, $P \leq 0.0001$. The size range of the SCA2 abnormal allele was 32–79 CAG repeats. The most common size of the abnormal allele in SCA2 patients was $(CAG)_{37}$ (Fig. 2). Unexpanded alleles ranged from 13 to 31 units. Normal alleles with 22 CAG repeats were the most frequents (76%). Nineteen percent of the remaining normal alleles ranged between 23 and 31 CAG units and 5% from 13 to 21.

Genetic anticipation was observed in the 80% of transmissions. We performed a detailed study of intergenerational instability of the repeat in 102 sibships. Expansions occurred in 89.03% and the contractions in 10.97% of the offspring of affected patients. Paternal transmission resulted higher variability in repeat length; ranging from −6 to +38 CAG repeats versus only −5 to +6 CAG repeats via the maternal transmission (Fig. 3). 89.04% of the patients with intergenerational expansion showed anticipation in the age onset compared to the parent and 10.96% showed expansions without anticipation.

Anticipation of age at onset (range 2–29 years) also occurred without intergenerational instability in 75% of a randomized sample of offspring of affected individuals.

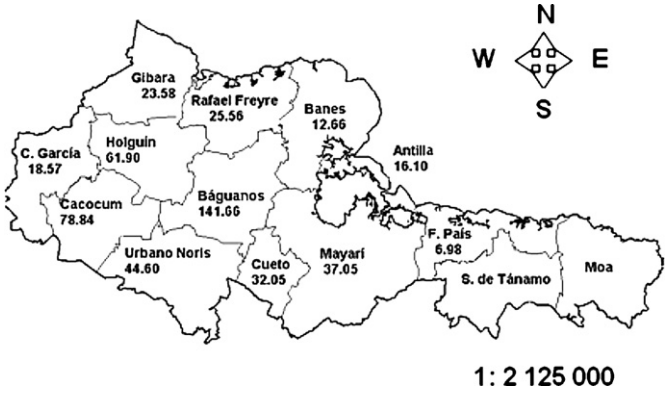


Fig. 1. Prevalence rate of hereditary ataxias in the province of Holguin, Cuba (2006–2007).

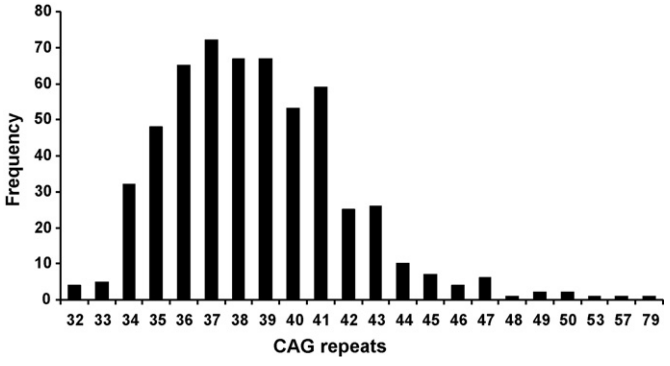


Fig. 2. Frequencies of pathological CAG repeats in 578 Cuban subjects with SCA2 mutation.

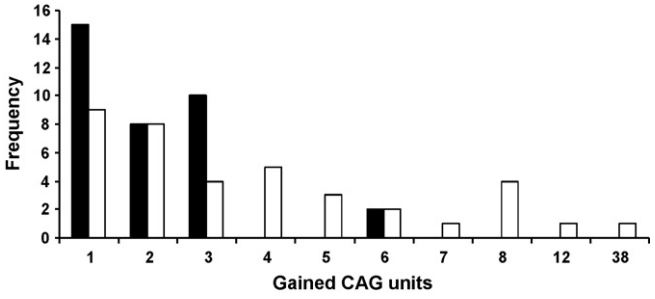


Fig. 3. Intergenerational instability of CAG expansions in SCA2. Dark bars indicate maternal transmissions and white bars indicate paternal transmissions.

The global incidence and prevalence of SCA2 are unknown [12] and there are large regional variations due to founder effects [24]. Large SCA2 families have been found in India [26], Martinique [5], Australia [27], Tunisia [2], Germany [23] and Italy [32].

Probably, the marked SCA2 prevalence in Cuba stems for Hispanics founders arrived to eastern region at XVII. Two main lines of evidences reinforce the above: (1) SCA2 Cuban families have a wide documented story of Hispanic origin and (2) almost all families share the same chromosomal haplotype surrounding in close region of SCA2 mutation [9].

Despite the dissemination of Cuban families through the island, the SCA2 prevalence in Holguin is different from other regions of the country. Cultural and environmental process has long lasting effects on allelic diversity, restricting mutations to selected populations [28]. The endogamous and closed cycle marriages of Holguin people living at XVII altogether with putative restricted environmental unbalancing, limited the SCA2 to Holguin. A similar explanatory mechanism has been proposed in SCA3 founders in Azoreans [15].

The prevalence of SCAs is influenced by the frequency of large normal alleles sharing the same chromosomal haplotype of patients [29]. In Holguin, the stable prevalence of SCA2 along time [30,14,31], despite the existence of genetic anticipation is likely the result of the high frequency of large normal alleles with pure CAG, which represent reservoirs of mutated alleles, driving to the continuous raising of new SCA2 cases. Our Center has an ongoing project to perform haplotype studies in this population to test this hypothesis.

The SARA rating scale is a reliable and valid assessment of ataxia, making it an appropriate test to evaluate its most important neurological manifestations, such as gait, stance and coordination of the upper and lower limbs, as well as to estimate the progression rate and its predictors [22].

The current study is the first in the world to use the SARA rating scale in more than 200 patients with autosomal dominant hereditary ataxia derived most likely from a founder effect. We found a strong correlation between the SARA score and the CAG expansion size and disease duration, indicating that the main clinical manifestations of SCA2 and its progression are a direct consequence of the genetic mutation. Also SARA settled robustly the patients in the SCA2 clinical stages, supporting their use in clinical surveys to assess the progression of cerebellar syndrome.

The strong correlation between CAG size and age of onset of symptoms in our patients is the highest ever reported in SCA2. Several factors can contribute to this finding, such as a large number of patients belonging to the same founder population and a wide range of CAG sizes (32–79 CAG). However, since this correlation has also been found in other studies in different founder populations [11] we cannot rule out that the common haplotype may be the driving force contributing to the correlation. Age of onset–CAG size correlation might be affected by other factors than CAG size: gene modifiers [8,19], ATXN2 polymorphisms or other genes [25], and unknown environmental factors.

The fact that at least 11% of expansions were not associated with anticipation, may reflect the effect of *cis*-acting elements and the presence of CAA interruptions in the CAG segment [3,16,4], decreasing the toxicity of CAG at mRNA or protein levels or other protective factors.

An unexpected result in this study was the presence of genetic anticipation (range 2–29 years) in 75% of a randomized sample of 102 sibships in the absence of intergenerational instability. We are not aware of any similar finding in the literature. A plausible explanation is that the variation in the severity of SCA2 can also be due to gene modifiers combined with environmental factors, although more research is needed to achieve a better understanding of this finding.

The frequent intergenerational expansion of the CAG repeat, in contrast with the infrequency of its contraction, found in the Cuban patients studied, has been described in other populations [1,6]. This could be due to differences in hairpin stability of CAG with respect to CTG, enabling a preferential replication to CAG with respect to CTG [3,16,4,18].

Our finding that a CAG size range between 32 and 79 units is capable to cause SCA2, is similar to what was reported in a previous Cuban study [21].

The prevalence of the SCA2 mutation in the Holguin province of Cuba is the highest reported worldwide, and reflects most likely a founder effect. The stabilization of the prevalence rate along time suggests the existence of premutated chromosomes with pure CAG, acting as a founder reservoir for new expansions. Further epidemiological studies will be focused on the scanning of other SCA mutations in Cuban population, such as SCA7, SCA15, SCA16, SCA18 and other.

The work of our Center has enabled a complete ascertaining of all families affected with SCAs in Cuba and the establishment of national guidelines to improve their clinical management and to provide genetic counselling and presymptomatic testing. The Cuban experience in the molecular epidemiology, as well as in the management and prevention of these conditions, could be expanded in international collaborative studies to design new therapies and better palliative treatments.

Acknowledgments

We are indebted to the Cuban SCA families that participated in this study, to the Cuban Ministry of Public Health, to the physicians, technicians and other specialists that collaborated in this research and to MD Victor B. Penchaszadeh for language corrections.

References

[1] E. Alonso, L. Martinez-Ruano, I. De Biase, C. Mader, A. Ochoa, P. Yescas, R. Gutierrez, M. White, L. Ruano, M. Fragoso-Benitez, T. Ashizawa, S.I. Bidichandani, A. Rasmussen, Distinct distribution of autosomal dominant spinocerebellar ataxia in the Mexican population, Mov. Disord. 22 (2007) 1050–1053.
[2] S. Belal, G. Cancel, G. Stevanin, F. Hentati, C. Khati, C. Ben Hamida, G. Auburger, Y. Agid, M. Ben Hamida, A. Brice, Clinical and genetic analysis of a Tunisian family with autosomal dominant cerebellar ataxia type 1 linked to the SCA2 locus, Neurology 44 (1994) 1423–1426.
[3] J.D. Cleary, C.E. Pearson, Replication fork dynamics and dynamic mutations: the fork-shift model of repeat instability, Trends Genet. 21 (2005) 223–230.
[4] O. Didierjean, G. Cancel, G. Stevanin, A. Dürr, K. Bürk, A. Benomar, A. Lezin, S. Belal, M. Abada-Bendid, T. Klockgether, A. Brice, Linkage disequilibrium at the SCA2 locus, J. Med. Genet. 36 (1999) 415–417.
[5] A. Dürr, D. Smadja, G. Cancel, A. Lezin, G. Stevanin, J. Mikol, R. Bellance, G.G. Buisson, H. Chneiweiss, J. Dellanave, Autosomal dominant cerebellar ataxia type I in Martinique (French West Indies) clinical and neuropathological analysis of 53 patients from three unrelated SCA2 families, Brain 118 (1995) 1573–1581.
[6] P. Giunti, G. Sabbadini, M.G. Sweeney, M.B. Davis, L. Veneziano, E. Mantuano, A. Federico, R. Plasmati, M. Frontali, N.W. Wood, The role of the SCA2 trinucleotide repeat expansion in 89 autosomal dominant cerebellar ataxia families frequency, clinical and genetic correlates, Brain 121 (1998) 459–467.

[7] A.E. Harding, Clinical features and classification of inherited ataxias, in: A.E. Harding, T. Deufel (Eds.), Inherited Ataxias, Raven, New York, 1993, pp. 1–14.
[8] S. Hayes, G. Turecki, K. Brisebois, I. Lopes-Cendes, C. Gaspar, O. Riess, L.P. Ranum, S.M. Pulst, G.A. Rouleau, CAG repeat length in RAI1 is associated with age at onset variability in spinocerebellar ataxia type 2 (SCA2), Hum. Mol. Genet. 9 (2000) 1753–1758.
[9] A. Hernandez, C. Magarino, S. Gispert, N. Santos, A. Lunkes, G. Orozco, L. Heredero, J. Beckmann, G. Auburger, Genetic mapping of the spinocerebellar ataxia 2 (SCA2) locus on chromosome 12q23-q24.1, Genomics 25 (1995) 433–435.
[10] G. Imbert, F. Saudou, G. Yvert, D. Devys, Y. Trottier, J.M. Garnier, C. Weber, J.L. Mandel, G. Cancel, N. Abbas, A. Dürr, O. Didierjean, G. Stevanin, Y. Agid, A. Brice, Cloning of the gene for spinocerebellar ataxia 2 reveals a locus with high sensitivity to expanded CAG/glutamine repeats, Nat. Genet. 14 (1996) 285–291.
[11] T. Klockgether, Ataxias, in: C.G. Goetz (Ed.), Textbook of Clinical Neurology, Saunders, Philadelphia, 2003, pp. 741–757.
[12] T. Klockgether, R. Ludke, B. Kramer, M. Abele, K. Burk, L. Schols, O. Riess, F. Laccone, S. Boesch, I. Lopes-Cendes, A. Brice, R. Inzelberg, N. Zilber, J. Dichgans, The natural history of degenerative ataxia: a retrospective study in 466 patients, Brain 121 (1998) 589–600.
[13] A. Matilla, R. Goold, P. Giunti, Molecular pathogenesis of spinocerebellar ataxias, Brain 129 (2006) 1357–1370.
[14] G. Orozco, A. Nodarse, R. Cordovés, G. Aurburger, R. Estrada, Study of 225 patients with autosomal dominant cerebellar ataxia: presumed founder effect in the province Holguin Cuba, in: J. Plaitakis (Ed.), Hereditary Ataxias, Springer-Verlag, Berlin, 1989, pp. 345–360.
[15] M. Parduey, The pathogenesis of Machado Joseph disease: a high manganese/low magnesium initiated CAG expansion mutation in susceptible genotypes? J. Am. Coll. Nutr. 23 (2004) 715S–729S.
[16] C.E. Pearson, K.N. Edamura, J.D. Cleary, Repeat instability: mechanisms of dynamic mutations, Nat. Rev. Genet. 6 (2005) 729–742.
[17] M.S. Pulst, A. Nechiporuk, T. Nechiporuk, S. Gispert, X.N. Chen, I. Lopes-Cendes, S. Pearlman, S. Starkman, G. Orozco-Diaz, A. Lunkes, P. DeJong, G.A. Rouleau, G. Auburger, J.R. Korenberg, C. Figueroa, S. Sahba, Moderate expansion of a normally biallelic trinucleotide repeat in spinocerebellar ataxia type 2, Nat. Genet. 4 (1996) 269–276.
[18] S.M. Pulst, A. Nechiporuk, S. Starkman, Anticipation in spinocerebellar ataxia type 2, Nat. Genet. 5 (1993) 8–10.
[19] S.M. Pulst, N. Santos, D. Wang, H. Yang, D. Huynh, L. Velazquez, K.P. Figueroa, Spinocerebellar ataxia type 2: polyQ repeat variation in the CACNA1A channel modifies age of onset, Brain 128 (2005) 2297–2303.
[20] K. Sanpei, H. Takano, S. Igarashi, T. Sato, M. Oyake, H. Sasaki, A. Wakisaka, K. Tashiro, Y. Ishida, T. Ikeuchi, R. Koide, M. Saito, A. Sato, T. Tanaka, S. Hanyu, Y. Takiyama, M. Nishizawa, N. Shimizu, Y. Nomura, M. Segawa, K. Iwabuchi, I. Eguchi, H. Tanaka, H. Takahashi, S. Tsuji, Identification of the spinocerebellar ataxia type 2 gene using a direct identification of repeat expansion and cloning technique, DIRECT, Nat. Genet. 14 (1996) 277–284.
[21] N. Santos, J. Aguiar, J. Fernández, M. Vázquez, G. Auburger, S. Gispert, Y. Mendoza, J. Garcia, L. Velazquez, Molecular diagnosis of a sample of the Cuban population with spinocerebellar ataxia type 2, Biotecnología Aplicada 16 (1999) 219–221.
[22] T. Schmitz-Hubsch, S. Tezenas du Montcel, L. Baliko, J. Berciano, S. Boesch, C. Depondt, P. Giunti, C. Globas, J. Infante, J.S. Kang, B. Kremer, C. Mariotti, B. Melegh, M. Pandolfo, M. Rakowicz, P. Ribai, R. Rola, L. Schöls, S. Szymanski, B.P. van de Warrenburg, A. Dürr, T. Klockgether, R. Fancellu, Scale for the assessment and rating of ataxia: development of a new clinical scale, Neurology 66 (2006) 1717–1720.
[23] L. Schöls, Spinocerebellar ataxia type 2. Genotype and Phenotype in German kindreds, Arch. Neurol. 54 (1997) 1073–1080.
[24] L. Schöls, P. Bauer, T. Schmidt, T. Schulte, O. Riess, Autosomal dominant cerebellar ataxias: clinical features, genetics, and pathogenesis, Lancet 3 (2004) 291–304.
[25] D.K. Simon, K. Zheng, L. Velazquez, N. Santos, L. Almaguer, S.M. Pulst, Mitochondrial complex I gene variant associated with early age of onset in SCA2, Arch. Neurol. 64 (2007) 1042–1044.
[26] K.K. Sinha, P.F. Worth, D.K. Jha, S. Sinha, V.J. Stinton, M.B. Davis, N.W. Wood, M.G. Sweeney, K.P. Bhatia, Autosomal dominant cerebellar ataxia: SCA2 is the most frequent mutation in eastern India, J. Neurol. Neurosurg. Psychiatry 75 (2004) 448–452.
[27] E. Storey, D. du Sart, J.H. Shaw, Frequency of spinocerebellar ataxia types 1, 2, 3, 6 and 7 in Australian patients with spinocerebellar ataxia, Am. J. Med. Genet. 95 (2000) 351–357.
[28] S. Sun, C.M.T. Greenwood, I. Thiffault, N. Hamel, G. Chong, W.D. Foulkes, The HNPCC associated MSH2*1906GRC founder mutation probably originated between 1440 CE and 1715 CE in the Ashkenazi Jewish population, J. Med. Genet. 42 (2005) 766–768.
[29] H. Takano, G. Cancel, T. Ikeuchi, D. Lorenzetti, R. Mawad, G. Stevanin, O. Didierjean, A. Durr, M. Oyake, T. Shimohata, R. Sasaki, R. Koide, S. Igarashi, S. Hayashi, Y. Takiyama, M. Nishizawa, H. Tanaka, H. Zoghbi, A. Brice, S. Tsuji, Close associations between prevalences of dominantly inherited spinocerebellar ataxias with CAG-repeat expansions and frequencies of large normal CAG alleles in Japanese and Caucasian populations, Am. J. Hum. Genet. 63 (1998) 1060–1066.
[30] L. Vallés, G.L. Estrada, S.L. Bastecherrea, Algunas formas de heredoataxia en una región de Cuba, Rev. Cub. Neurol. 27 (1978) 163–176.
[31] L. Velázquez, Introduction to hereditary ataxias, in: L. Gonzalez (Ed.), Spinocerebellar Ataxia Type 2: Principal Neurophysiological Aspects in the Diagnosis, Prognosis and Evolution of the Disease, Ediciones Holguin, Holguin, 2008, pp. 17–74.
[32] M. Zortea, M. Armani, E. Pastorello, G.F. Nunez, S. Lombardi, S. Tonello, M.T. Rigoni, L. Zuliani, M.L. Mostacciuolo, C. Gellera, S. Di Donato, C.P. Trevisan, Prevalence of Inherited Ataxias in the Province of Padua, Italy, Neuroepidemiology 23 (2004) 275–280.

FUNDAMENTOS MOLECULARES DE LA MUTAGENESIS DE LAS EXPANSIONES DE CAG EN ATXN2 Y SU APLICACIÓN EN LA GENETICA CLINICA

III.2

3.2 FUNDAMENTOS MOLECULARES DE LA MUTAGENESIS DE LAS EXPANSIONES DE CAG EN ATXN2 Y SU APLICACIÓN EN LA GENETICA CLINICA

3.2.1: Polimorfismos del CAG en alelos de longitud no expandida e intermedia en el locus SCA2 (*ATXN2*) de la población Cubana: Evidencia del origen de las expansiones patológicas causantes de la SCA2.

Publicado en: Eur J Hum Gen, 20, (2012) 41–49. (Factor de Impacto: 4.38)

En el artículo anterior se postula que las variantes genéticas con expansiones de CAG en el rango ≥23-31CAG son proclives a la inestabilidad y por tanto son variantes de riesgo genético. Esto partía de la hipótesis que las grandes diferencias de la frecuencia de SCAs autosómicas pueden explicarse por la frecuencia de alelos largos normales (ALs) (Takano et al., 1998). Sin embargo, varios autores planteaban que la hipótesis no era verificable en la SCA2.

Por otra parte, no se había reportado un estudio en nuestra población dirigido a comprobar esta relación, pero existía la observación de un gran número de alelos largos en Cuba (Velázquez *et al*., 2009). Por tanto, el objetivo de este trabajo fue: *determinar las frecuencias de los alelos normales en la población cubana y su relación con la frecuencia de SCA2.* Para esto se determinó mediante análisis de fragmentos y secuenciación el haplotipo de STR y SNP, así como la variación de la expansión de CAG de ~3000 cromosomas de la población saludable, para darle cumplimiento al objetivo # 2 de la tesis y poder demostrar el postulado de Takano et al. 1998 (objetivo # 1).

Los principales resultados fueron:

1. *Los alelos largos fueron 3.8 veces más frecuentes que los cortos.*
2. *Cuba mostró la mayor frecuencia, a nivel internacional, de alelos largos (Fig.1).*

3. *Cuba mostró la mayor variabilidad en cuanto a la presencia, posición y configuración del triplete de CAA dentro de la secuencia de CAG (Tabla 3).*
4. *Los alelos largos muestran mayor variabilidad en cuanto al polimorfismo del CAA (Tabla 4).*
5. *El triplete de CAA se pierde en la medida que aumenta el la expansión de CAG (Tabla4).*
6. *La expansión de CAG de los alelos largos muestran inestabilidad (contracciones y expansiones) meiótica y mitótica (Fig. 2c, Fig.3).*
7. *El haplotipo reconocido para las familias SCA2 y los alelos largos es 3-4-11.*

Aquí no se probó que existe un relación matemática entre frecuencias de variantes largas y frecuencia de SCA2 pero sí permitió vislumbrarla y contrarrestar, de acuerdo a los datos existentes, las aseveraciones prevalentes de Juvonen et al., 2005 y Sobczak & Krzyzosiak., 2004.

Dado que no habían datos para establecer un modelo explicativo global, en este trabajo si probamos que existe relación entre la frecuencia absoluta de ALs y de SCA2 en Cuba. Por tanto, aquí se generó la poderosa observación de que Cuba tiene *"la mayores frecuencias absolutas de alelos largos y de SCA2 a nivel internacional"*.

La conclusiones fundamentales fueron que esta población a diferencia de otras es altamente polimórfica, lo cual pudiera ser clave para el surgimiento de mutaciones *de novo*. Estas pudieran ser mucho más probables en aquellas variantes largas con pérdida de las interrupciones de CAA, y bajo un haplotipo predispuesto. Por tanto, esto indicaba la posibilidad de que uno de los primeros pasos en la fundación de la SCA2 en Cuba, involucró premutaciones más que la introducción de la expansión patogénica (haplotipo predisponente vs. efecto fundador.

European Journal of Human Genetics (2011), 1–9

www.nature.com/ejhg

ARTICLE

Unexpanded and intermediate CAG polymorphisms at the SCA2 locus (ATXN2) in the Cuban population: evidence about the origin of expanded SCA2 alleles

José Miguel Laffita-Mesa*[,1], Luis C Velázquez-Pérez[1], Nieves Santos Falcón[1], Tania Cruz-Mariño[2], Yanetza González Zaldívar[1], Yaimee Vázquez Mojena[1], Dennis Almaguer-Gotay[1], Luis Enrique Almaguer Mederos[1] and Roberto Rodríguez Labrada[3]

The role of short, large or intermediate normal alleles (ANs) of the *ataxin-2* gene in generating expanded alleles (EAs) causing spinocerebellar ataxia type 2 (SCA2) is poorly understood. It has been postulated that SCA2 prevalence is related to the frequency of large ANs. SCA2 shows the highest worldwide prevalence in Cuban population, which is therefore a unique source for studying the relationship between the frequency of large and intermediate alleles and the frequency of SCA2 mutation. Through genetic polymorphism analyses in a comprehensive sample (~3000 chromosomes), we show that the frequency of large ANs in the *ataxin-2* gene is the highest worldwide, although short ANs are also frequent. This highly polymorphic population displayed also high variability in the CAG sequence, featured by loss of the anchor CAA interruption(s). In addition, large ANs showed germinal and somatic instability. Our study also includes related genotypic, genealogical and haplotypic data and provides substantial evidence with regard to the role of large and intermediate alleles in the generation of pathological EAs.
European Journal of Human Genetics (2011) **0,** 000–000. doi:10.1038/ejhg.2011.154

Keywords: SCA2; genetic polymorphisms; *ataxin-2*; Cuba

INTRODUCTION

Spinocerebellar ataxia type 2 (SCA2) is a neurodegenerative disease showing an autosomal dominant inheritance pattern. This disorder is caused by a CAG expansion crossing a certain threshold in the coding region of the *ataxin-2* gene.[1–3] SCA2 patients exhibit a progressive cerebellar syndrome characterized by ataxic gait, cerebellar dysarthria, dysmetria and dysdiadochokinesia. Normal *SCA2* gene/*ataxin-2* CAG tracts range from 13 to 31 repeats, the most frequent being that of 22 trinucleotides.[4,5] Ataxic phenotype occurs when the repeat is larger than 34 CAG.[6] Triplet repeats between 32–34 fall in the gray zone for penetrance, whereas 37–75 CAG repeats are fully penetrant.[6] Only few patients having 32 and 33 CAG repeats have been reported so far, with very late onset – between 50 and 60 years of age.[4,7,8] Extremely large expansions of 109, 200 and 500 CAG in infants have also been observed,[9–11] but are rarer.

The CAG sequence in the SCA2 locus is cryptic in nature, showing CAA interruptions. Normal SCA2 alleles contain CAA interruptions, whereas the vast majority of expanded alleles are uninterrupted.[1–3] The most common configuration is $(CAG)_8CAA(CAG)_4CAA(CAG)_8$. It is claimed that the CAA interruptions confer genomic stability to the CAG tracts.[12,13] The absence of interruptions, on the other hand, predisposes the CAG tracts to undergo expansion and eventually to reach the pathological threshold in analogy to the absence of interruptions in the CGG and CAG tracts in fragile X syndrome and SCA1, respectively.[12–16] In a first effort to decipher which factors underlie the prevalence of SCAs, it was found that a high frequency of large normal alleles (ANs) was closely linked to a high prevalence of SCA2 in Caucasians, whereas low large ANs results in a very low prevalence of SCA2 in the Japanese population.[17] Later, some authors suggested that it is biased to only use CAG length as a marker for determining the propensity of certain alleles to be predisposed or prone to undergo expansion reaching the pathological range.[18] This argument is also supported by the fact that in East India, where SCA2 prevalence is high,[19–21] large ANs are spared and clustered to limited ethnic groups.[22] Therefore, the role of large ANs as a possible source of SCA2 expansion and the mechanism by which this might happen is poorly understood. Here, we performed a comprehensive study to determine the frequencies of ANs in the Cuban population and their relationship with the high prevalence of SCA2. We analyzed CAG sizes, sequences, as well as genealogical and microsatellite haplotype data to gain insights into the mechanism underlying the prevalence of SCA2 in Cuba. The main questions were as follows: (1) What is the relative frequency of ANs? (2) Are the frequency of large ANs and the prevalence of SCA2 associated? In answering these questions, we compared our data with other studies performed in large populations. We found a strong relationship between the SCA2 prevalence and the relative frequency of large ANs.

[1]Laboratory of Molecular Neurobiology, Centre for the Research and Rehabilitation of Hereditary Ataxias (CIRAH), Holguín, Cuba; [2]Center of Medical Genetics, Holguín, Cuba; [3]Laboratory of Neurophysiology, Center for the Research and Rehabilitation of Hereditary Ataxias (CIRAH), Holguín, Cuba
*Correspondence: Dr JM Laffita-Mesa, Laboratory of Molecular Neurobiology, Department of Molecular Neurobiology, Centre for the Research and Rehabilitation of Hereditary Ataxias, Carretera Central Km 5 ½ Reparto Edecio Pérez, Postal Code 80100, Holguín, Cuba. Tel: +53 24 462296/424090/461564;
E-mails: laffita@ataxia.hlg.sld.cu, laffitamesa@yahoo.es
Received 30 November 2010; revised 11 May 2011; accepted 30 June 2011

METHODS

Design of the study

The study was performed in three phases: (1) a nation-wide screening for SCA, conducted during 2003–2007; (2) a survey for neurological consultations to SCA2 relatives identified in the former phase; and (3) nation-wide recruitment of the general non-SCA2 population (not related healthy subjects), in 2009. The study was conducted by the National Center for the Research and Rehabilitation of the Hereditary Ataxias (CIRAH) in the city of Holguín, which is the main referral center for these conditions and a neurological center of excellence in the country.[22] The Ethics Committee of the National Centre for the Research and Rehabilitation of the Hereditary Ataxias approved the research protocol and all studied subjects signed an informed consent form after being explained the purpose and methods of the research.

CAG length estimation by fragment analysis

Peripheral blood leukocytes were extracted using EDTA as anticoagulant, and genomic DNA was isolated using standard methods. Molecular diagnosis was performed to estimate the number of CAG repeats at the SCA2 locus according to the methods described elsewhere.[1,2] CAG length was determined by using Alfexpress II sequencing system and the PCR fragments run with ReproGel high resolution (GE Healthcare, Buckinghamshire, UK). All repeat sizes were compared with the standards, before which repeat size was confirmed by DNA sequence analysis. Also, internal (100 and 300 bp) and external (50–500-step, 50 bp) Alfexpress ladders were used to extrapolate the fragment size. Traces were analyzed using the software Allelelink according to the manufacturer's specifications. In all, 200 ng of genomic DNA from peripheral blood was used in each PCR aimed at somatic mosaicism determination, which was defined by Matsuura *et al*.[23]

CAG substructure deciphering

In total, 81 normal chromosomes with 13–31 CAG repeats representing the majority of allelic classes were sequenced. Reactions were performed by using either DAN1–DAN2 or SCA2A–SCA2B (Tib-Mol-Biol, Berlin, Germany) and amplified fragments were excised from the gel using GFX band extraction kit (GE Healthcare). Each allele was sequenced directly in both directions, forward and reverse, using the Thermosequenase Cy5 Dye Terminator Sequencing kit (GE Healthcare) and run as explained above. Sequence data were managed by using Alf win Sequence analyzer 2.10.

Data management and statistical analyses

Availability of a large number of SCA2 and non-SCA2 carrier chromosomes (n=2695) and the general population (GP) (n=80) allowed us to determine the distribution of the number of CAG repeats in ANs at the SCA2 locus. Non-SCA2 carrier chromosomes were defined as the new mutation (NM) group given its familial relationship with SCA2 families, because all of them segregated or were related to SCA2 families. CAG distributions were determined by pooling chromosomes. Genotypes only included healthy individuals. Means, variances, ranges and skewness were determined for the distributions of ANs at the SCA2 locus in the individuals. To perform statistical analyses of the differences in the frequencies of large ANs between Cuban and other populations, we defined large ANs as in Takano *et al*.[17] Thus, for the purposes of this study, all alleles smaller than 22 repeats were defined as short alleles and those longer than 22 CAG repeats were considered as large alleles. Differences in the relative frequencies of the large ANs were analyzed by means of the χ^2-test with Yates's correction when necessary. Associations were also assessed using the Fisher's exact test. Data were arranged in excel sheets, managed and analyzed as counts and frequencies. The normality of each distribution was assessed by the Kolmogorov–Smirnov test (K–S). All statistical analyses were performed using the STATISTICA data analysis software system version 6 (StatSoft Inc., 2003, Tulsa, OK, USA). The null hypotheses were rejected at $P \leq 0.05$.

RESULTS

Normal genetic polymorphism according to CAG repeat length (13–31 CAG)

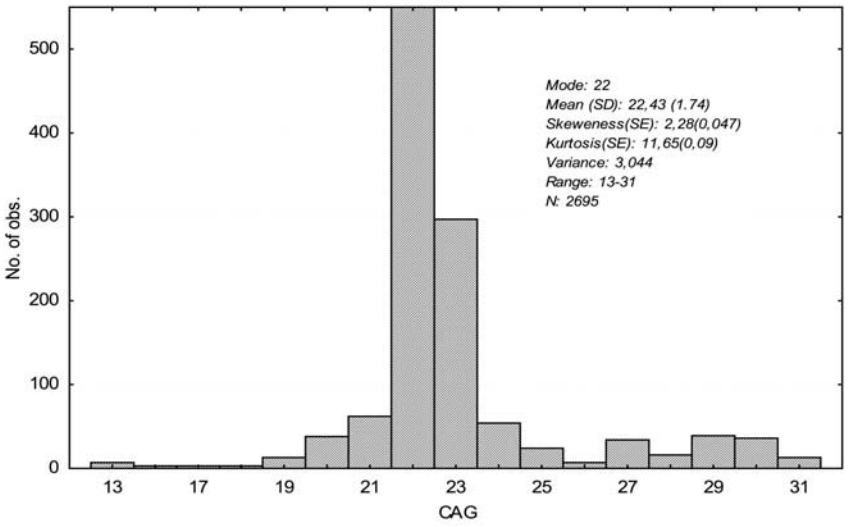
The normal CAG repeat size from 2695 (K–S, $P<0.01$) chromosomes of the Cuban NM population shows a modal distribution (Figure 1a) (mode and median=22 CAG), with 22 CAG being the most frequent allele (n=2046/2695, 75.92%). This frequency is the lowest reported as compared with other populations: East Indian,[24] 91% (χ^2=51.53, d.f.=1, P=0.000); Polish,[18] 91.2% (χ^2=27.85, d.f.=1, P=0.0000); and Finnish,[25] 92.81% (χ^2=71.62, d.f.=1, P=0.0000). CAG mean was 22.43 ± 0.033; while in the populations previously reported, it was as follows: Japanese,[17] 21.86; Caucasians,[17] 22.24; East Indian,[24] 21.95; Finnish,[25] 22.2; Canadian,[26] 22.2; four human populations mixed,[27] 22.15; and Czech, 22.3.[28] The range of the CAG is distributed continuously from 13 to 31 CAG and encompasses almost all the expected allelic classes in this numeric series (17 observed/18 expected) – with a kurtosis of 11.66 and a variance of 3.04 *vs* 1.21 for other populations worldwide.[27] The allele with 13 CAG repeats is exclusively found in the Cuban population and that with 26 CAG in both Cuban and Czech populations.[28] The distribution is non-symmetrical (skewness 2.28), with a clear tendency toward the large CAG range. The mode at 22 CAG splits the distribution into short ANs sized from 13 to 21 CAG and large ANs sized from 23 to 31 CAG, showing 7 and 9 allelic classes, respectively (χ^2=0.03, d.f.=1, P=0.86) (Figure 1a). Short AN frequencies are as follows: 21 (2.30%), 20 (1.41%), 19 (0.48%), while in the large AN zone those sized with 23 (11.02%), 24 (2.00%), 29 (1.44%), 30 (1.33%), 27 (1.26%) and 25 (0.89%) CAG units are represented more (Figure 1a). Alleles falling in the positive tail (ie, large ANs) are highly over-represented compared with those included in the negative tail (ie, short ANs) (χ^2=204.71, d.f.=1, P=0.0000).

The exclusive distribution of Cuban large ANs prompted us to compare the frequencies of such alleles with other populations. It is noteworthy that the frequency of large ANs (>22 CAG) was significantly higher in Cubans than in Japanese[17] (χ^2=73.67, d.f.=1, P=0.0000), Caucasians[17] (χ^2=11.15, d.f.=1, P=0.0000), North Indians[21] (χ^2=15.14, d.f.=1, P=0.0002), East Indians[24] (χ^2=65.15, d.f.=1, P=0.0000) and Polish[18] (χ^2=19.33, d.f.=1, P=0.0000) (Figure 1b). Extending our cutoff values to >23 CAG or >24 CAG disclosed even more significant differences between our and the aforementioned populations. Curiously, there were populations lacking alleles larger than 25 CAG triplets (Japanese, Caucasians, Indians, except Czechs) and others showing gaps (Polish) in the large AN zone, contrasting with Cubans, who showed a continuous distribution of such alleles.

Intermediate alleles (32 and 33 CAG)

We extended our analysis to include those alleles sized 32 and 33 CAG (intermediate alleles (IAs)), thus enlarging our sample to 2722 chromosomes. Cuban chromosomes carrying 32 and 33 CAG repeats were as frequent in our population (0.55% and 0.44%, respectively) as those with 28 (0.58%) and 31 (0.47%) CAG repeats. The paired comparison between IAs and the frequency of short ANs disclosed highly significant differences in the frequencies of IAs *vs* those short ANs with sizes ranging from 13 to 18 CAG (IAs: 0.99% *vs* short ANs: 0.58%, χ^2=85.80, d.f.=1, P=0.0000, Fisher's exact test P=0.0000). No differences were observed when the rest of the short ANs with 18, 19 and 21 CAG repeats were added (data not shown). In all, 25 genotypes with IAs were found. All these alleles were found in SCA2 families, that is, the NM population. The genotypic frequencies were 22/32 (n=11/25, 44%); 22/33 (n=7/25, 28%); 23/32 (n=3/25, 12%); 30/32 (n=1/25, 4%); 23/33 (n=1/25, 4%); 33/41 (n=1/25, 4%); and 33/43 (n=1/25, 4%). Only two IAs were associated with age at disease onset at 34 and 45 years, with genotypes of 22/32 CAG and 23/32 CAG, respectively (penetrance 8%) (Table 1).

a



b

COMPARISON OF FREQUENCIES, OF CUBAN LARGE ANs AT SCA2 LOCUS VS OTHER POPULATIONS

Cutoff values when CAG for Large AN =	>22	>23	>24	>25	>26	>27	>28	>29	>30	
Frequencies of SCA2 Large AN in different Populations (N)										Skewness
Cuba (2695)	0.19	0.08	0.06	0.05	0.05	0.04	0.03	0.02	0.01	2.28
Japanese[1](359)	0.01 (χ^2=73.67, p=0.0000)	0.01 (χ^2=23.61, p=0.0000)	0.00 (χ^2=22.99, p=0.0000)	—	—	—	—	—	—	-0.97
Caucasians[1](355)	0.12 (χ^2=11.15, p=0.0000)	0.03 (χ^2=11.86, p=0.0006)	0.00 (χ^2=5.11, p=0.0238)	—	—	—	—	—	—	-4.81
North India[2] (135)	0.06 (χ^2=15.14, p=0.0002)	0.02 (χ^2=6.41, p=0.0114)	0.02 (χ^2=4.90, p=0.0268)	—	—	—	—	—	—	nd
Eastern India[3](448)	0.038 (χ^2=65.15, p=0.0000)	—	—	—	—	—	—	—	—	–0.13
Polish[4](234)	0.08 (χ^2=19.33, p=0.0000)	0.02 (χ^2=11.30 p=0.0000)	0.02 (χ^2=8.06 p=0.0045)	¶	¶	0.004 (χ^2=7.34 p=0.0068)	¶	—	—	nd
Finnish[5] (501)	0.07 (χ^2=44.65, p=0.0000)	0.04 (χ^2=11.03, p=0.0009)	0.03 (χ^2=8.36, p=0.0038)	—	—	—	—	—	—	4.47

1: Takano et al. 1998, 2: Salem et al. 2000, 3: Basu et al. 2000, 4: Sobczack & Krzyzosiak., 2004, 5, Juvonen et al., 2005. nd: *not documented*, ¶: This population don´t shows alleles sized with 25, 26 and 28CAG.

Figure 1 Unexpanded genotype frequency distribution at the *ataxin-2* gene in 2695 NM Cuban chromosomes and frequency of large ANs in Cuba *vs* different populations. (**a**) CAG distribution at the SCA2 locus of the ANs in the Cuban population related to SCA2 families (NM). The distribution is skewed toward large ANs. The shortest alleles found are those sized 13 and 14 CAG and the largest are 30 and 31 CAG. Alleles with >23 CAG repeats are over-represented over all large and short ANs. (**b**) Comparison of the CAG size frequency of large ANs in Cuba *vs* other populations. Allele frequencies in Cuba were grouped by CAG size; frequencies of large normal allele (>22 CAG) and other alleles (≤22 CAG) were compared with the frequency of large ANs in other populations by χ^2 or Fisher's exact test. Frequencies were tabulated in a 2×2 contingency table with 2 d.f. for comparison. Because the frequency of alleles sized 22 CAG may be higher with respect to both groups, short (<22 CAG) and large alleles (>22 CAG), we also applied a component analysis by χ^2 and Fisher's exact test, excluding such alleles, and included in the table only alleles either >22 CAG or <22 CAG. In the table each line shows the frequency of large ANs in each population and the resulting comparison of the frequency of each allele when grouped according to CAG cutoff (ie, >22 CAG, >23 CAG, >24 CAG, and so on). Frequencies of alleles in each region were taken from the literature as shown in the table and the Cuban frequency was determined in the current work.

Normal genotypic distribution

The genotypic distribution of 836 healthy subjects from the NM group revealed that the most frequent combination was 22/22 CAG (60.88%), followed by 22/23 (13.75%), 23/23 (2.99%), 22/29 (2.15%), 22/30 (2.15%), 22/27(1.91%), 21/21(1.55%) and 20/20 (0.47%). The distribution was skewed toward the positive tail of the distribution, with 28 *vs* 12 genotypic classes in the negative tail (comparison of number of allelic classes χ^2=0.36, d.f.=1, P=0.55) (Supplementary Table S1 and Figure S1). Excluding the central genotype 22/22, we divided the distribution into short (with at least one short AN) and large genotypes (with at least one large AN). Both of these categories were significantly different with regard to the presence of the second allele (ie, short ANs in large genotypes and *vice versa*) in each group (χ^2=288.46, d.f.=1, P=0.0000, Fisher's exact test P=0.0000). Moreover, large genotypes were more enriched, with 22 CAG alleles as compared with the short genotypes (200/572 (35%) *vs* 20/114 (18%), respectively, χ^2=13.24, d.f.=1, P=0.0003, Fisher's exact test P=0.0000) (Table 2). Therefore, the large chromosomes associate preferentially with 22 CAG and with large than with short ANs. We focused on the comparison excluding the 22/23 genotype

Table 1 Genotypic frequency of IAs

Genotype	*Freq (%)*	*Cases sick (age at onset years)*
22/32	14	1 (34 years)
22/33	28	—
23/32	12	1 (45 years)
30/32	4	—
23/33	4	—
33/41	4	—
33/43	4	—

Table 2 Comparisons of the frequencies of short ANs, large ANs and 22 CAG normal alleles in the short, normal or large genotypes

	Counts (frequency of alleles)			
Genotypes (n)	*Short ANs*	*Large ANs*	*22 CAG*	*Total (2n)*
Short heterozygotes (57)	78 (0.68)[a]	16 (0.14)	20 (0.17)	114
Large heterozygotes(286)	16 (0.02)	356 (0.62)[b]	200 (0.34)[c]	572

[a]χ^2=288.46, d.f.=1, P=0.0000 (short ANs in short heterozygote group *vs* short ANs in large heterozygote group).
[b]χ^2=288.46, d.f.=1, P=0.0000 (large ANs in large heterozygote group *vs* large ANs in short heterozygote group).
[c]χ^2=13.24, d.f.=1, P=0.0003 (22 CAG ANs in each large or short heterozygote groups).

that accounts for the 42.59% of the large genotypes, but the association yet remained highly significant (155 *vs* 57, χ^2=7.28, d.f.=1, P=0.007, Fisher's exact test P=0.0078).

Large ANs in the GP

By active screening, we randomly selected 80 chromosomes in the GP (K–S, $P<0.01$) (Supplementary Figure S2). The analysis revealed the same combination, with at least 25% (n=20/80) of large AN_{GP}'s. Short AN_{GP}'s were only sized with 16, 20 and 21 repeats (n=30/80, 28.75%), but were over-represented compared with short AN_{NM}'s (n=129/2695 (4.78%), χ^2=153.93, d.f.=1, P=0.0000). The paired comparison between the frequencies of short AN_{GP} and large AN_{GP} revealed only marginal differences (n=30/80 (28.75%) *vs* n=20/80 (25%), χ^2=2.91, d.f.=1, P=0.08, Fisher's exact test P=0.06). Variance of the CAG repeats belonging to the GP distribution was 5.31, with a positive kurtosis of 4.44 ± 0.532. Interestingly, we found an expanded allele with 35 CAG repeats without direct evidence of any other pathogenic CAG expansion in the family; this allele was not previously registered during the time of our national screening.[5] The overall distribution in GP samples was skewed, although less (1.66 ± 0.26 *vs* 2.29 ± 0.047) than in the NM samples. Large AN_{GP}'s appeared much more frequently than large AN_{NM}'s, but did not yield significant differences (n=20/80, 25.00% *vs* n=536/2695, 19.29%; χ^2=1.27, d.f.=1, P=0.2604).

CAG>CAA substructure of Cuban ATXN2 alleles with 13–33 CAG repeats

To characterize the allelic diversity of the CAG tract substructure, 81 normal chromosomes were analyzed, encompassing the whole spectrum of Cuban normal CAG alleles (range 13–31 CAG). Of the 36 allelic worldwide classes, 24 were found in the Cuban population (Table 3). Other populations and ethnic groups are much less diverse, for example, Indian[29] (14), Yoruba in Ibadan, Nigeria (YRI)[30] (11), Polish[18](11) and French[2] (7) (Supplementary Table S2).

Nine new alleles with regard to the CAG/CAA configuration were exclusively linked to the Cuban population and were sized 13, 16, 22, 24, 28, 30 and 31 CAG, respectively: $(CAG)_{13}$; $(CAG)_{16}$; $(CAG)_8CAA(CAG)_8CAA(CAG)_4$; $(CAG)_8CAA(CAG)_6CAA(CAG)_8$; $(CAG)_8CAA(CAG)_{19}$; CAG_{30}; $(CAG)_8CAA(CAG)_7CAA(CAG)_{13}$; $(CAG)_{31}$; and $(CAG)_8CAA(CAG)_8CAA(CAG)_4(CAG)_8$ (Table 3 and Supplementary Table S2).

Among the short ANs, those sized with 15, 17 and 19 trinucleotides had one CAA interruption, whereas alleles with 13 and 16 units did not contain such interruptions. Four CAA interruption patterns, $(CAG)_8CAA(CAG)_4CAA(CAG)_8$ or 8+4+8, $(CAG)_8CAA(CAG)_8$-$CAA(CAG)_4$ or 8+8+4, $(CAG)_{13}CAA(CAG)_8$ or 13+8 and inverted 8+13, were found in alleles with 22 repeats (Table 4). The large AN zone was more enriched in allelic variants with regard to the configuration of the internal CAA interruption 10/20 classes (50%). Alleles with 27, 29 and 31 CAG repeats had three CAA interruptions, alleles with 24 and 30 CAG repeats contained two CAA interruptions, those with 23, 25 and 28 CAG repeats had one CAA interruption, and those with 30 and 31 CAG repeats were uninterrupted (Table 4).

According to the previous criteria stated by Choudhry *et al*,[29] the polar variation of the internal CAG tracts was analyzed. Five alleles showed a continuous pure CAG configuration (alleles: 13, 16, 30, 31 and 33, range 13–33 CAG). In the remaining 16 alleles, the 5′ and 3′ tracts were much more variable (range 6–16 and 8–19, respectively) than the middle tract (range 4–8). Large ANs accounted for the vast majority of the length variability at each end (5′ tract: range 8–16 CAG; 3′ tract: range 8–19 CAG; and middle tract: range 4–8 CAG). On the contrary, short ANs displayed variability only at the 3′ tract (range 6–10 CAG). IAs (32–33 CAG alleles) showed CAG/CAA polymorphism as well. Three of these alleles had uninterrupted CAG tracts. CAA interruptions were frequent in 33 CAG alleles (14 chromosomes), all having the $(CAG)_{24}CAA(CAG)_8$ structure (Table 4).

CAG>CAA, short tandem repeat and CAG length polymorphism analysis in large ANs in SCA2 families

A combined analysis of four genetic polymorphisms, short tandem repeat (STR) haplotype, CAG length, the CAG⇒CAA change (SNP database, SNP ID No. rs4098854) and sequence architecture was carried out in two SCA2 Cuban families with expanded and large ANs (Figures 2a and b). Marked association between the disease locus and the haplotypic run 3-G-4-11 at D12S1332–(A/G)–D12S1672–D12S1333 loci (allele 3, 202 bp; allele 4, 287 bp; allele 11, 235 bp) was observed. In the case of family 86, the allele 3 at D12S1332 in II-1 expanded and became allele 2 (204 bp). It was then passed with the rest of the truncated haplotype 4-11 to the successive generations IV and V. Homozygotes for the whole haplotype block was a feature in the family SCA2-44 (Go) (individuals II-2 and II-3) (Figure 2a). In the first branch of this pedigree, discrete vertical, sib–ships and horizontal, sib–sib CAG instabilities (± 1 CAG) in individuals III-1, III-3, III-5, III-6 and IV-1 were observed (Figures 2a and c). These CAG instabilities reached the peak normality, 31 CAG, in the son (III-5) and grandson (IV-1) of II-2. Furthermore, reverse mutations or contractions to 30 repeats resulted in non-ataxic phenotype in individuals II-2, III-8, III-12 and III-13 of the SCA2-44 (Go) family. Under a similar back mutation, however, the ataxic phenotype was retained with onset at 45 years in the female III-6, belonging to family SCA2-86 (Os) (Figure 2b). All alleles were sequenced and no interruptions were detected in the CAG tracts.

Large ANs are somatically unstable

To provide insights into the behavior of large ANs, we analyzed somatic mosaicism in 245 alleles, CAG range: 20–31. Large ANs

Table 3 CAG length polymorphism and its relationship with CAA interruption in unexpanded and IAs

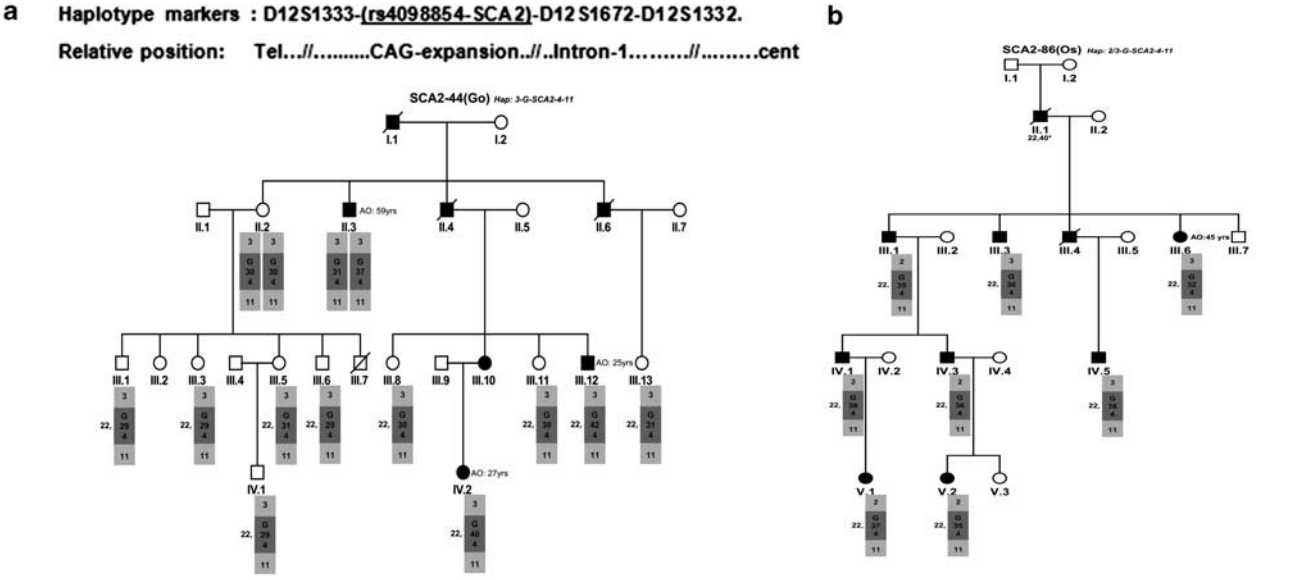
CAG repeat Length	*Allele type*	*CAA interruption Pattern within CAG repeat*	*CAG length polymorphism (Normals, Short, large ANs or Intermediate Alleles)*
13	13	○○○○○○○○○○○○○	Short ANs
15	6+8	○○○○○○●○○○○○○○○	Short ANs
16	16	○○○○○○○○○○○○○○○○	Short ANs
17	8+8	○○○○○○○○●○○○○○○○○	Short ANs
19	10+8	○○○○○○○○○○●○○○○○○○○	Short ANs
22	13+8	○○○○○○○○○○○○○●○○○○○○○○	ANs
22	8+13	○○○○○○○○●○○○○○○○○○○○○	ANs
22	8+4+8	○○○○○○○○●○○○○●○○○○○○○	ANs
22	8+8+4	○○○○○○○○●○○○○○○○●○○○○	ANs
23	13+9	○○○○○○○○○○○○○●○○○○○○○○○	Large ANs
23	14+8	○○○○○○○○○○○○○○●○○○○○○○○	Large ANs
24	8+6+8	○○○○○○○○●○○○○○○●○○○○○○○○	Large ANs
25	16+8	○○○○○○○○○○○○○○○○●○○○○○○○○	Large ANs
27	8+4+4+8	○○○○○○○○●○○○○●○○○○●○○○○○○○○	Large ANs
28	8+19	○○○○○○○○●○○○○○○○○○○○○○○○○○○	Large ANs
29	8+4+4+10	○○○○○○○○●○○○○●○○○○●○○○○○○○○○○	Large ANs
30	30	○○○○○○○○○○○○○○○○○○○○○○○○○○○○○○	Large ANs
30	13+7+8	○○○○○○○○○○○○○●○○○○○○○●○○○○○○○○	Large ANs
30	8+7+13	○○○○○○○○●○○○○○○○●○○○○○○○○○○○○	Large ANs
31	31	○○○○○○○○○○○○○○○○○○○○○○○○○○○○○○○	Large ANs
31	8+8+4+8	○○○○○○○○●○○○○○○○●○○○○●○○○○○○○○	Large ANs
32	32	○○○○○○○○○○○○○○○○○○○○○○○○○○○○○○○○	IAs
33	33	○○○○○○○○○○○○○○○○○○○○○○○○○○○○○○○○○	IAs
33	24+8	○○○○○○○○○○○○○○○○○○○○○○○●○○○○○○○○	IAs

CAG repeats are represented by empty circles and CAA interruptions by filled circles.

Table 4 Relationship between CAG size, sequence structure, allelic class and the purity of *ataxin-2* CAG repeat tracts

Alleles	*Allelic classes according CAA interruption triplets*	*CAG length*	*Number of CAA*	*CAG>CAA structure*	*Site of the polar variation of CAG ends or CAG purity (5′/middle/3′)*
Short	***2***	13	0	○○○○○○○○○○○○○	**pure**
		16		○○○○○○○○○○○○○○○○	**pure**
		15	1	○○○○○○●○○○○○○○○	**3′**
		17		○○○○○○○○●○○○○○○○○	**5′/3′**
		19		○○○○○○○○○○●○○○○○○○○	**5′**
		--	2	--	--
		--	3	--	--
Normal	***4***	--	0	--	--
		22	1	○○○○○○○○●○○○○○○○○○○○○ ○○○○○○○○○○○○○○○●○○○○○○○○	**3′** **5′**
		22	2	○○○○○○○○●○○○○●○○○○○○○○ ○○○○○○○○●○○○○○○○○●○○○○	--- **middle**
		--	3	--	--
Large	***10***	30	0	○○○○○○○○○○○○○○○○○○○○○○○○○○○○○○	**pure**
		31		○○○○○○○○○○○○○○○○○○○○○○○○○○○○○○○	**pure**
				○○○○○○○○○○○○○○○●○○○○○○○○	**5′**
		23	1	○○○○○○○○○○○○○○○●○○○○○○○○	**5′**
		25		○○○○○○○○○○○○○○○○○●○○○○○○○○	**5′**
		28		○○○○○○○○●○○○○○○○○○○○○○○○○○○	**3′**
		24	2	○○○○○○○○●○○○○○○●○○○○○○○○	--
		30		○○○○○○○○●○○○○○○○●○○○○○○○○○○○	**3′**
		27		○○○○○○○○●○○○○●○○○○●○○○○○○○○	--
		29	3	○○○○○○○○●○○○○●○○○○●○○○○○○○○○	**3′**
		31		○○○○○○○○●○○○○○○○○●○○○○●○○○○○○○○	**middle**
Intermediates	***2***	32	0	○○○○○○○○○○○○○○○○○○○○○○○○○○○○○○○○	**pure**
		33		○○○○○○○○○○○○○○○○○○○○○○○○○○○○○○○○○	**pure**
		33	1	○○○○○○○○○○○○○○○○○○○○○○○○○●○○○○○○○	**5′**
		--	2	--	--
		--	3	--	--

displayed different morphologies as compared with 22 CAG alleles (Supplementary Figure S3c). Through associative analysis using CAG cutoff ranges and stability criteria according to the fragment morphology, we clearly discriminated two CAG ranges (20–26 CAG and 27–31 CAG) with different stabilities (χ^2=159.80, Fisher's exact test P=0.0000, OR: infinitum and ROC curve sensitivity: 100%,



c

Familiy	Elder sib (repeat size)	Younger sib (repeat size)	Difference in repeat size (Younger sib-Elder sib)	Transmission	Status of transmitting Parent
SCA2-44(Go)	II-2 (30)[a]	II-3 (37)	-7	Paternal	affected
	II-2 (30)	II-3 (31)	1	Paternal	affected
	III-8 (30)[a]	III-11 (42)	-12	Paternal	affected
	III-8 (30)	III-12 (30)	0	Paternal	affected
SCA2-86(Os)	III-1 (35)	III-3 (36)	1	Paternal	affected
	III-1 (35)	III-6 (32)[b]	-3	Paternal	affected
	III-3 (36)	III-6 (32)	-4	Paternal	affected
	IV-1 (36)	IV-3 (36)	0	Paternal	affected

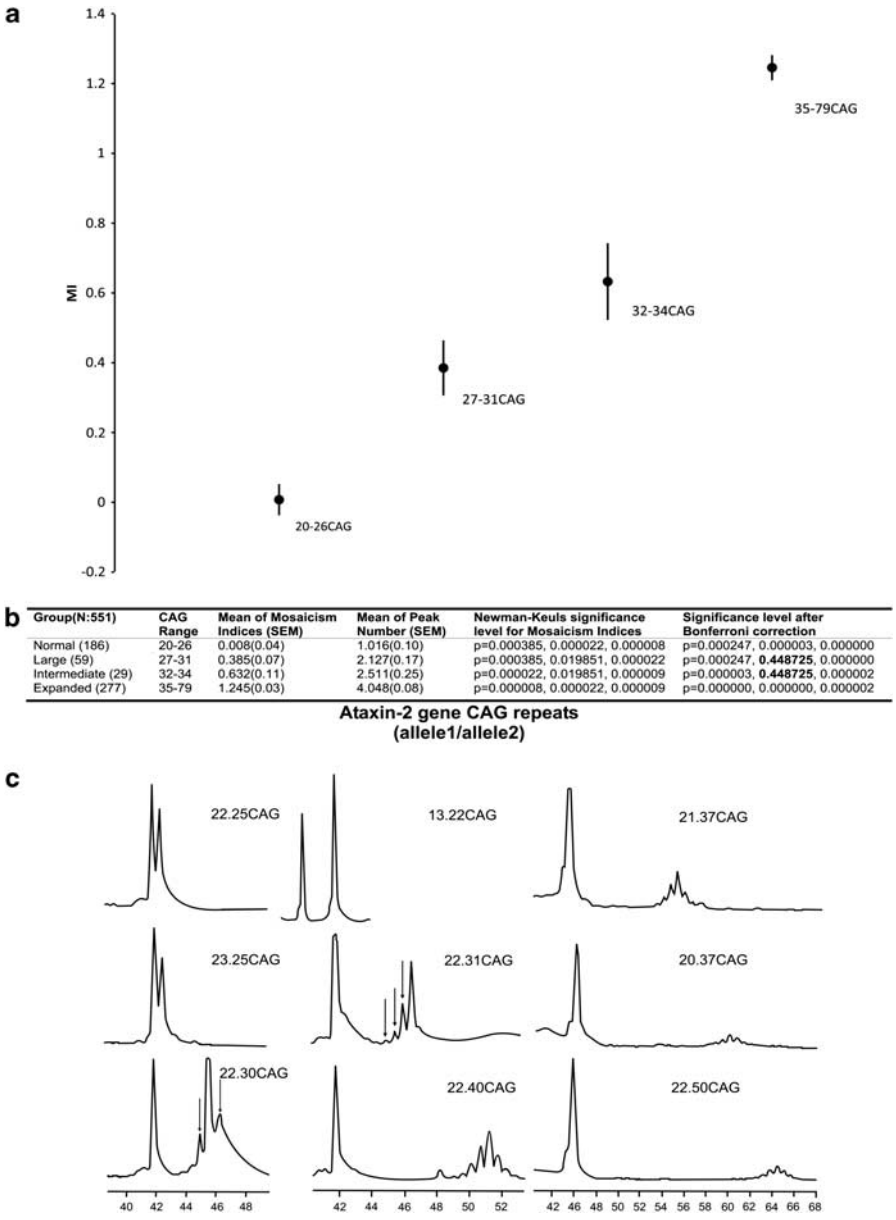
a: Rescued individual by reversal mutation, b: Affected individual after a reversal mutation.

Figure 2 Haplotype of STR microsatellites around the SCA2 CAG in two pedigrees and analysis of sib–sib instability. STR haplotype segregating into two proven SCA2 families with large ANs. Un/expanded alleles and the STR microsatellite haplotypes are indicated in the boxes (gray). Genotype of the individual II-1 was taken from Santos *et al*.[4] and is marked with asterisk. Age at onset in individuals II-3, III-12, IV-2 and III-6, of pedigrees 44 and 86, respectively, are denoted. (**a** and **b**) Families SCA2-44 (Go) and SCA2-86 (Os) of the Cuban kindred. Note that in the first branch of family 44, there are slight instabilities in descendants of the homozygous female II-2 with the haplotype 3-4-11. (**c**) Analysis of CAG instabilities in both pedigrees in sib–sib pairs. CAG contractions are prominent, leading to back or reversal mutations that rescued 2/5 individuals of the SCA2 phenotype. The contraction in the female III-6 was able to retain the SCA2 phenotype with a mild ataxic course. Note: The normal alleles of the STR have been omitted, and only disease alleles are shown.

specificity: 91%) (Supplementary Figures S3a and b). Furthermore, the comparison disclosed highly significant differences and strong effect of the CAG length in the somatic mosaicism, assessed either as mosaicism index (MI) or as peak numbers ($F(1,243)=131.57$, $P=0.00$, ANOVA followed by Bonferroni *post hoc* test, 20–26 *vs* 27–31 CAG range, MI ± SEM: 0.00 ± 0.016, 0.385 ± 0.028, respectively, $P=0.000$) (Supplementary Figures S3a and b). According to our current results with regard to the differences between somatic mosaicism and the known phenotypic range of SCA2, we extended the previous comparison to a cohort of 551 alleles with the following ranges: normal (20–26), large (27–31), intermediate (32–34) and expanded (35–79). In addition to the differences in normal CAG repeat ranges, the levels of somatic mosaicism of 32–34 CAG alleles were lower than those observed in the full penetrant expanded alleles (35–79 CAG) (MI ± SEM: 0.632 ± 0.11 and 2.51 ± 0.25, respectively, $P=0.000$) (Figures 3a–c). These results remained significant after adjustment for multiple comparisons. However, the levels of somatic mosaicism for the 27–31 CAG and 32–34 CAG repeat ranges were not different (MI ± SEM: 0.385 ± 0.07 and 0.632 ± 0.11, respectively, $P=0.448725$) (Figure 3b).

DISCUSSION

We found that large alleles and IAs are very frequent in the Cuban population. This may be relevant for the understanding of SCA2 origin in Cuba, because though the prevalence of SCA2 in Cuba is the highest worldwide,[5] the basis for this[31] is poorly understood yet. The frequency of large ANs in Cuba is the highest compared with other populations having a high prevalence of SCA2[20,21] and even other populations with a high frequency of such alleles.[18] This highly significant frequency of large ANs and other alleles distinct from 22 CAG suggests that the SCA2 locus in Cuba is highly polymorphic. The similar distribution in the sample of non-related chromosomes in the general population, collected for our study, further supports the relative abundance of large ANs in Cuba. Taken together, the genetic variance in the CAG > CAA sequence was also the highest, showing novel alleles situated at the peak of normality (29, 30, 31 CAG; Table 3), completely lacking stability-mediating CAA interruptions (Figure 3). Collectively, the data may indicate a causal relationship between the overall abundance of large ANs with CAG purity and the frequency of expanded SCA2 alleles. It is well known that Cuba has the highest frequency of SCA2 mutation worldwide,[5] a fact that supports



Group(N:551)	CAG Range	Mean of Mosaicism Indices (SEM)	Mean of Peak Number (SEM)	Newman-Keuls significance level for Mosaicism Indices	Significance level after Bonferroni correction
Normal (186)	20-26	0.008(0.04)	1.016(0.10)	p=0.000385, 0.000022, 0.000008	p=0.000247, 0.000003, 0.000000
Large (59)	27-31	0.385(0.07)	2.127(0.17)	p=0.000385, 0.019851, 0.000022	p=0.000247, **0.448725**, 0.000000
Intermediate (29)	32-34	0.632(0.11)	2.511(0.25)	p=0.000022, 0.019851, 0.000009	p=0.000003, **0.448725**, 0.000002
Expanded (277)	35-79	1.245(0.03)	4.048(0.08)	p=0.000008, 0.000022, 0.000009	p=0.000000, 0.000000, 0.000002

Figure 3 Multiple comparisons of the somatic mosaicism indices harboring 20–79 CAG repeats in the *ataxin-2* gene. (**a**) One-way ANOVA comparison for somatic mosaicism in a cohort of 551 un/expanded alleles. Each range showed differences either in the somatic mosaicism indices or in peak number. (**b**) *Post hoc* Newman–Keuls multiple comparisons of the mosaicism indices in the studied sample. Only alleles sized between 27–31 and 32–34 CAG were not different after corrections. Data are presented as mean and standard error of mean (SEM). (**c**) Somatic mosaicism of unexpanded and expanded pathological (causing SCA2) ATXN2 alleles; arrowheads point to the somatic mosaicism in large ANs.

our hypothesis. It is conceivable that alleles with high normal CAG provide the source of *de novo* mutations that further contributes to the high SCA2 prevalence in the Cuban population.

Our study also revealed that large ANs were more frequent than short ANs. Previously, it has been stated that longer alleles in the normal range present a higher probability of pathogenic expansion than do shorter alleles.[32,33] Therefore, this makes more likely the contribution of the former group to the onset of *de novo* mutations. In this state, IAs with 32–35 CAG repeats that may have evolved from large ANs with 23–31 CAG repeats are intermediates, which in successive generations would give rise to full penetrant alleles (37–79 CAG) associated with SCA2, ALS,[34] FTDP-U,[34] or pure parkinsonism.

Alleles with 27–31 CAG were somatically unstable, similar to expanded alleles (Figures 3a and b), providing significant evidence for the increased instability of large ANs, and pointing to 27 CAG as the threshold for IAs. The high odds ratio and the specificity seen in the 27–31 CAG alleles suggest that instabilities would be associated to 27–31 CAG alleles and are more likely to occur in these rather than in the shorter ones (20–26 CAG alleles). The proposed 27 CAG as the lower threshold for intermediate or indeterminate penetrance alleles is in agreement with the situation in SCA7, for which *de novo* mutations have been proven, and the lower threshold for mitotic and meiotic instabilities has been set at 27 CAG.[35] In addition, polyglutamine stretches between 27 and 33 CAG repeats in *ataxin-2* are associated with an increased risk for ALS,[34] suggesting a different molecular

nature and behavior of ≥ 27 CAG alleles, which is similar to that of 32–34 CAG alleles (Figure 3b). The association of IAs with other phenotypes alongside SCA2, for example, ALS, FTDP-U and PD in Cuba, is currently being studied because it cannot be neglected in our population (Laffita-Mesa *et al*, in preparation).

Given the fact that somatic mosaicism could be responsible for the juvenile onset and different SCA2 course and phenotypes (unpublished data), and that mosaicism has been found in SCA2 brain and cerebellum,[23] our finding of large ANs being somatically unstable is important for considering these alleles as risk alleles for intergenerational instabilities. Moreover, they are likely to contribute to other neurodegenerative pathologies (ie, ALS and FTDP-U).[34] The lack of differences between this group and the 32–33 CAG places these alleles in the category of pathological CAG expansions, which is supported by recent findings.[34] Somatic instability in blood might be used as a diagnostic and prognostic measurement in these pathologies given that *ataxin-2* is ubiquitously expressed, although it eminently affects CNS.

The distorted assortment shown in Table 2 may reflect some selective advantage. Significant distortion either of homozygous or of heterozygous Ataxin-2-deficient mice with regard to wild type might be reminiscent of the current distortion.[36] Short ANs might be variants with partial loss of function and large ANs variants with partial gain of function. The former group may be negatively selected. As it has been shown previously, levels of Ataxin-2 protein have a very important relationship with its biological function.[37,38] The partial gain of function may confer better functionality (translatability and/or protein stability), leading to an increase in Ataxin-2 levels and a positive selection in the population, a notion that is in agreement with the current selection of pre-expansions in the ATXN2 locus.[30]

In conclusion, we have shown that the highest worldwide concentration of large ANs underlies the highest worldwide prevalence and incidence rates of SCA2. We have found that Cuban population is highly polymorphic at the SCA2 locus. Precisely, these polymorphisms are central for the generation of NMs. These polymorphisms deviate from the focal length (22 CAG) and harbor short and large CAG repeats with polar loss of the CAA interruptions, which are essential for genomic stability. The genomic instability of larger-than-27 CAG alleles supports the proposal that expansions arise from those alleles showing augmented CAG, with either 5′ or 3′ CAA loss, under a predisposed haplotype. The weight of the evidence seems to involve a premutation with a predisposed haplotype as an early step in the foundational event of SCA2 in Cuba; it is more likely than the introduction of the mutation properly. It seems that the enlargement of the CAG is linked to the conservation of certain adaptive functions, with a further selection of the more advantageous variants.

CONFLICT OF INTEREST

The authors declare no conflict of interest.

ACKNOWLEDGEMENTS

We are indebted to our Cuban SCA2 patients and their relatives. We are grateful to Dr Peter O Bauer for his critical review and correction of the English. This work was supported in part by Third World Academy of Sciences -TWAS- in the framework of the Research Grant Project 06-329 RG/BIO/LA UNESCO FR: 3240157855 all to JMLM and a grant of the Cuban Ministry of Health and the Science and Technology Ministry.

AUTHOR CONTRIBUTIONS

Study concept and design: Laffita-Mesa. Acquisition of data: Laffita-Mesa, Santos-Falcón, González-Zaldívar, Almaguer-Mederos, Almaguer-Gotay, Vázquez-Mojena, Rodríguez-Labrada. Analysis and interpretation of data: Laffita-Mesa, Velázquez-Pérez, Rodríguez-Labrada, Cruz-Mariño. Drafting of the manuscript: Laffita-Mesa. Critical revision of the manuscript for important intellectual content: Vázquez-Mojena, Rodríguez-Labrada, Almaguer-Gotay. Statistical analysis: Laffita-Mesa. Obtained funding: Laffita-Mesa and Velázquez-Pérez. Study supervision: Laffita-Mesa.

1 Pulst SM, Nechiporuk A, Nechiporuk T *et al*: Moderate expansion of a normally biallelic trinucleotide repeat in spinocerebellar ataxia type 2. *Nat Genet* 1996; **14**: 269–276.
2 Imbert G, Saudou F, Yvert G *et al*: Cloning of the gene for spinocerebellar ataxia 2 reveals a locus with high sensitivity to expanded CAG/glutamine repeats. *Nat Genet* 1996; **14**: 285–291.
3 Sanpei K, Takano H, Igarashi S *et al*: Identification of the spinocerebellar ataxia type 2 gene using a direct identification of repeat expansion and cloning technique, DIRECT. *Nat Genet* 1996; **14**: 277–284.
4 Santos N, Aguiar J, Fernandez J *et al*: Molecular diagnosis of a sample of the Cuban population with spinocerebellar ataxia type 2. *Biotecnol Aplic* 1999; **16**: 219–221.
5 Velázquez Pérez L, Sánchez Cruz G, Santos Falcón N *et al*: Molecular epidemiology of spinocerebellar ataxias in Cuba: Insights into SCA2 founder effect in Holguin. *Neurosci Lett* 2009; **454**: 157–160.
6 Cancel G, Durr A, Didierjean O *et al*: Molecular and clinical correlations in spinocerebellar ataxia 2: a study of 32 families. *Hum Mol Genet* 1997; **6**: 709–715.
7 Futamura N, Matsumura R, Fujimoto Y *et al*: CAG repeat expansions in patients with sporadic cerebellar ataxia. *Acta Neurol Scand* 1998; **98**: 55–59.
8 Fernandez M, McClain ME, Martinez RA *et al*: Late-onset SCA2: 33 CAG repeats are sufficient to cause disease. *Neurology* 2000; **55**: 569–572.
9 Spears MD, Melton S, Mao Q *et al*: Ataxia and progressive encephalopathy in a 4-year-old girl. *Lab Med* 2010; **41**: 5–9.
10 Mao R, Aylsworth AS, Potter N *et al*: Childhood-onset ataxia: testing for large CAG-repeats in SCA2 and SCA7. *Am J Med Genet* 2002; **110**: 338–345.
11 Babovic-Vuksanovic D, Snow K, Patterson MC *et al*: Spinocerebellar ataxia type 2 (SCA 2) in an infant with extreme CAG repeat expansion. *Am J Med Genet* 1998; **79**: 383–387.
12 Pearson CE, Eichler EE, Lorenzetti D *et al*: Interruptions in the triplet repeats of SCA1 and FRAXA reduce the propensity and complexity of slipped strand DNA (S-DNA) formation. *Biochemistry* 1998; **37**: 2701–2708.
13 Chung MY, Ranum LP, Duvick LA *et al*: Evidence for a mechanism predisposing to intergenerational CAG repeat instability in spinocerebellar ataxia type I. *Nat Genet* 1993; **5**: 254–258.
14 Kunst CB, Leeflang EP, Iber JC *et al*: The effect of FMR1 CGG repeat interruptions on mutation frequency as measured by sperm typing. *J Med Genet* 1997; **34**: 627–631.
15 Eichler EE, Holden JJ, Popovich BW *et al*: Length of uninterrupted CGG repeats determines instability in the FMR1 gene. *Nat Genet* 1994; **8**: 88–94.
16 Zhong N, Ju W, Pietrofesa J *et al*: Fragile X 'gray zone' alleles: AGG patterns expansion risks and associated haplotypes. *Am J Med Genet* 1996; **64**: 261–265.
17 Takano H, Cancel G, Ikeuchi T *et al*: Close associations between prevalences of dominantly inherited spinocerebellar ataxias with CAG-repeat expansions and frequencies of large normal CAG alleles in Japanese and Caucasian populations. *Am J Hum Genet* 1998; **63**: 1060–1066.
18 Sobczak K, Krzyzosiak WJ: Patterns of CAG repeat interruptions in SCA1 and SCA2 genes in relation to repeat instability. *Hum Mutat* 2004; **24**: 236–247.
19 Wadia NH, Swami RK: A new form of heredo-familial spinocerebellar degeneration with slow eye movements (nine families). *Brain* 1971; **94**: 359–374.
20 Sinha KK, Worth PF, Jha DK *et al*: Autosomal dominant cerebellar ataxia: SCA2 is the most frequent mutation in eastern India. *J Neurol Neurosurg Psychiatry* 2004; **75**: 448–452.
21 Saleem Q, Choudhry S, Mukerji M *et al*: Molecular analysis of autosomal dominant hereditary ataxias in the Indian population: high frequency of SCA2 and evidence for a common founder mutation. *Hum Genet* 2000; **106**: 179–187.
22 Becker M: For neurologists in Cuba, hope is not embargoed. *Lancet Neurol* 2009; **8**: 1088–1089.
23 Matsuura T, Sasaki H, Yabe I *et al*: Mosaicism of unstable CAG repeats in the brain of spinocerebellar ataxia type 2. *J Neurol* 1999; **246**: 835–839.
24 Basu P, Chattopadhyay B, Gangopadhaya PK *et al*: Analysis of CAG repeats in SCA1, SCA2, SCA3, SCA6, SCA7 and DRPLA loci in spinocerebellar ataxia patients and distribution of CAG repeats at the SCA1, SCA2 and SCA6 loci in nine ethnic populations of eastern India. *Hum Genet* 2000; **106**: 597–604.
25 Juvonen V, Hietala M, Kairisto V *et al*: The occurrence of dominant spinocerebellar ataxias among 251 Finnish ataxia patients and the role of predisposing large normal alleles in a genetically isolated population. *Acta Neurol Scand* 2005; **111**: 154–162.
26 Butland SL, Devon RS, Huang Y *et al*: CAG-encoded polyglutamine length polymorphism in the human genome. *BMC Genom* 2007; **8**: 126.
27 Andres AM, Lao O, Soldevila M, Calafell F, Bertranpetit J: Dynamics of CAG repeat loci revealed by the analysis of their variability. *Hum Mutat* 2003; **21**: 61–70.
28 Bauer PO, Kotliarova SE, Matoska V: Fluorescent multiplex PCR – fast method for autosomal dominant spinocerebellar ataxias screening. *Russ J Genet* 2005; **41**: 675–682.
29 Choudhry S, Mukerji M, Srivastava AK *et al*: CAG repeat instability at SCA2 locus: anchoring CAA interruptions and linked single nucleotide polymorphisms. *Hum Mol Genet* 2001; **10**: 2437–2446.

30 Yu F, Sabeti PC, Hardenbol P *et al*: Positive selection of a pre-expansion CAG repeat of the human SCA2 gene. *PLoS Genet* 2005; **1**: e41.
31 Hernandez A, Magarino C, Gispert S *et al*: Genetic mapping of the spinocerebellar ataxia 2 (SCA2) locus on chromosome 12q23–q24.1. *Genomics* 1995; **25**: 433–435.
32 Rubinsztein DC, Amos W, Leggo J *et al*: Mutational bias provides a model for the evolution of Huntington's disease and predicts a general increase in disease prevalence. *Nat Genet* 1994; **7**: 525–530.
33 Cleary JD, Pearson CE: The contribution of *cis*-elements to disease-associated repeat instability: clinical and experimental evidence. *Cytogenet Genome Res* 2003; **100**: 25–55.
34 Elden AC, Kim HJ, Hart MP *et al*: Ataxin-2 intermediate-length polyglutamine expansions are associated with increased risk for ALS. *Nat Genet* 2010; **466**: 1069–1075.
35 Stevanin G, Giunti P, Belal GD *et al*: *De novo* expansion of intermediate alleles in spinocerebellar ataxia 7. *Hum Mol Genet* 1998; **7**: 1809–1813.
36 Kiehl TR, Nechiporuk A, Figueroa KP *et al*: Generation and characterization of Sca2 (ataxin-2) knockout mice. *Biochem Biophys Res Commun* 2006; **339**: 17–24.
37 Satterfield TF, Pallanck LJ: Ataxin-2 and its *Drosophila* homolog, ATX2, physically assemble with polyribosomes. *Hum Mol Genet* 2006; **15**: 2523–2532.
38 Nonhoff U, Ralser M, Welzel F *et al*: Ataxin-2 interacts with the DEAD/H-box RNA helicase DDX6 and interferes with P-bodies and stress granules. *Mol Biol Cell* 2007; **18**: 1385–1396.

Supplementary Information accompanies the paper on European Journal of Human Genetics website (http://www.nature.com/ejhg)

3.2.2: Mutaciones *de novo* en el gen Ataxin-2 y riesgo a ELA

Publicado en: PLoSONE 8(8): (2013) e70560. (Factor de Impacto: 4.09)

En el artículo anterior, se sugiere que el mismo mecanismo mutagénico subyacente a la SCA2, podía explicar la aparición de otras enfermedades como ELA, EP y demencia con expansiones anómalas de CAG en el gen *ATXN2*.

Esto fue, en parte, fertilizado por el hallazgo de expansiones intermedias asociadas con ELA (Elden et al., 2010) y la evidencia en nuestra población de casos ELA que eran el resultado de expansiones *de novo*. Sin embargo, la idea prevaleciente es que la expansión de polyQ en ataxina-2, más allá de ser una causa genética, es un factor de riesgo o gen modificador para esta enfermedad neurodegenerativa, a pesar de que en muchos casos alcanza expansiones penetrantes para la SCA2.

En este trabajo mediante el haplotipaje y la secuenciación de la expansión de CAG del gen *ATXN2* se reporta el hallazgo de tres casos ELA, los cuales eran el resultado de expansiones *de novo* que provenían de variantes paternas largas no patogénicas.

El mecanismo mutagénico en la trasmisión *padres-hijos* involucró la pérdida del triplete de CAA del extremo 5' (Fig. 4a).

La existencia de una expansión *de novo* de 35CAG, la cual es patogénica y completamente penetrante para SCA2, sugiere un solo mecanismo mutagénico involucrando la perdida de CAA como paso clave. Además, se encuentra que todo esto ocurría bajo el haplotipo CC-4, el cual es el mismo encontrado en familias SCA2 fundadoras, siendo una evidencia a favor del modelo de *haplotipo predisponente* (Fig. 4b). Estos hallazgos demuestran el efecto desestabilizador del haplotipo y de la interrupción de CAA, haciendo al segmento de CAG inestable genéticamente (objetivos # 1 y 3).

¿Pudieron estos casos ser producto de una mala clasificación de SCA2 como ELA? Aquí se prueba que no, porque de acuerdo al ESCORIAL (Brooks et al., 2000), los mismos eran ELA definida, tenían una sobrevida corta y no pertenecían a familias conocidas SCA2. Las expansiones del triplete de CAG en algunos de ellos, eran no penetrantes para SCA2, y se excluyó la contribución de mutaciones causales para ELA, p. ej. la expansión hexanucleotídica de GGGGCC en el *locus C90RF72* (Fig.2d) que es principal contribuyente a ELA.

Se realizó un meta-análisis con los datos de otras poblaciones y se excluyeron a los alelos largos (en el rango 24-27CAG) como riesgo para ELA. Sin embargo, el riesgo fue especifico y significativamente mayor (OR = 2.16, p≤ 0.000) para expansiones ≥30CAG (Figure 5b). La eliminación de las fuentes de heterogeneidad arrojó un riesgo relativo homogéneo, lo cual imputa un riesgo positivo de 1.77; 95% CI 1.55–2.03 a los alelos intermedios ≥32CAG. Estos datos cuestionan la idea de que las expansiones de CAG en *ATXN2*, son un factor riesgo o gen modificador de ELA, y abren la hipótesis de que estas mutaciones pueden causar esta enfermedad.

Lo hallazgos presentados son de absoluta novedad, hasta el momento son la única evidencia de esta naturaleza en la literatura porque se demuestra la segregación de la nueva mutación con el fenotipo de la enfermedad, lo cual es un criterio dorado para considerar una mutación como causa monogénica.

OPEN ACCESS Freely available online

PLOS ONE

De Novo Mutations in Ataxin-2 Gene and ALS Risk

José Miguel Laffita-Mesa[1,2]*, Jorge Michel Rodríguez Pupo[3], Raciel Moreno Sera[7], Yaimee Vázquez Mojena[1], Vivian Kourí[4], Leonides Laguna-Salvia[3], Michael Martínez-Godales[4], José A. Valdevila Figueira[3], Peter O. Bauer[6], Roberto Rodríguez-Labrada[1,2], Yanetza González Zaldívar[1], Martin Paucar[5], Per Svenningsson[5], Luís Velázquez Pérez[1,2]*

1 Center for the Research and Rehabilitation of Hereditary Ataxias, Holguín, Cuba, **2** Branch of Biomedicine of the Cuban Academy of Sciences -ACC-, Havana, Cuba, **3** Clinical and Surgical Hospital Lucía Íñiquez Landín, Holguín, Cuba, **4** Laboratory of Virology of Department of ITS, Institute of Tropical Medicine Pedro Kourí, Havana, Cuba, **5** Department of Clinical Neuroscience, Center for Molecular Medicine, Karolinska Institute, Sweden, **6** Department of Neuroscience, Mayo Clinic, Jacksonville, Florida, United States of America, **7** Policlinic and Medical Faculty, Tacajó, Báguanos, Holguín, Cuba

Abstract

Pathogenic CAG repeat expansion in the ataxin-2 gene (*ATXN2*) is the genetic cause of spinocerebellar ataxia type 2 (SCA2). Recently, it has been associated with Parkinsonism and increased genetic risk for amyotrophic lateral sclerosis (ALS). Here we report the association of *de novo* mutations in *ATXN2* with autosomal dominant ALS. These findings support our previous conjectures based on population studies on the role of large normal *ATXN2* alleles as the source for new mutations being involved in neurodegenerative pathologies associated with CAG expansions. The *de novo* mutations expanded from ALS/SCA2 non-risk alleles as proven by meta-analysis method. The ALS risk was associated with SCA2 alleles as well as with intermediate CAG lengths in the *ATXN2*. Higher risk for ALS was associated with pathogenic CAG repeat as revealed by meta-analysis.

Citation: Laffita-Mesa JM, Rodríguez Pupo JM, Moreno Sera R, Vázquez Mojena Y, Kourí V, et al. (2013) *De Novo* Mutations in Ataxin-2 Gene and ALS Risk. PLoS ONE 8(8): e70560. doi:10.1371/journal.pone.0070560

Editor: Bart Dermaut, Pasteur Institute of Lille, France

Received November 2, 2012; **Accepted** June 26, 2013; **Published** August 6, 2013

Funding: This work was supported in part by Third World Academy of Sciences (TWAS) in the framework of the Research Grant Project 06-329 RG/BIO/LA UNESCO FR: 3240157855 and BIOLABS Research Award all to JMLM and a grant of the Cuban Ministry of Health and the Science and Technology Ministry. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

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

* E-mail: laffita@ataxia.hlg.sld.cu (JMLM); cirahsca2@cristal.hlg.sld.cu (LVP)

Introduction

Amyotrophic lateral sclerosis (ALS), also known as Lou Gehrig's disease, is a devastating adult-onset neurodegenerative disease with no cure and is fatal within 2 to 5 years after the disease onset. Phenotypically, it is characterized by progressive motor neuron loss resulting in muscle weakness, wasting, fasciculations, spasticity, and hyperreflexia. The vast majority of cases (~90%) have no family history, while about 10% of patients have a genetic locus as causal entity for ALS. So far, mutations in *SOD1*, encoding for Cu/Zn superoxide dismutase, have been identified in 20% of familial ALS cases. Other frequent disease-causing genes include *C9ORF72*, TAR DNA-binding protein 43 (*TARDBP*), and fused in sarcoma/translocated in liposarcoma protein (*FUS/TLS*) [1]. Recently, it has been shown that large intermediate CAG repeat expansions in the ataxin-2 gene (*ATXN2*) contribute to almost 5% of the sporadic or familial ALS cases [2]. Intermediate CAG repeats are now recognized as ALS13 associated locus (MIM183090). Based on several studies, it appears that CAG expansion in *ATXN2* might account for more familial ALS cases than *SOD1* mutations with an overall incidence of 2% [2–5]. SCA2 is another neurodegenerative disorder with no available cure exhibiting progressive cerebellar syndrome characterized by ataxic gait, cerebellar dysarthria, dysmetria, and dysdiadochokinesia.

Normal CAG length in *ATXN2* gene ranges from 13–31 repeats, with 22 CAG being most common. Alleles with ≥26 CAG are considered as large alleles [6]. While inconsistencies exist in the literature regarding the CAG range and nomenclature of the intermediate or indeterminate alleles, those with ≥27–33 CAG are associated with increased and specific risk for ALS [2,3,7].

Only few cases with 32–34 CAG expansions have been reported so far in SCA2 [8–11]. Of note, none of the individuals with intermediate allele displayed typical clinical SCA2 picture. Therefore, starting point for SCA2 mutant range has been widely considered ≥34 CAG, while 37–75 CAG repeats are full penetrant [12] and massive expansions has been observed with severe disease starting in infants [13,14].

First report relating ALS with intermediate alleles considered 27 repeats as the threshold [2] while subsequently it has been refined to >30 CAG [3,7]. However, in SCA2, these alleles are considered as large normal, which might be prone for intergenerational instability leading to full penetrant disease-causing alleles, but this instability has not been reported so far. Given the rarity of large and intermediate alleles (23–34 CAG), little is known about the origin of ALS-related alleles. Only one *de novo* mutation has been reported, but the data about the unstable behavior of the involved alleles were not revealed [5]. Investigation in these patients would be important to understand events leading to genetic instability and for the accurate presymptomatic and prenatal diagnostics as well as genetic counseling.

Recently, we have shown that the highest worldwide concentration of large normal alleles (≥23 CAG) underlies the highest worldwide prevalence and incidence rates of SCA2 in Cuba.

Based on this observation we postulated that *ATXN2* intermediate alleles with 32–35 CAG may evolve from large ANs with 23–31 CAG, explaining familial or sporadic cases associated either with SCA2, ALS, and Parkinsonism [5]. In support of our postulates, we uncovered familial ALS cases resulting from large unstable alleles reaching ALS intermediate length and SCA2 pathogenic expansions at the *ATXN2* locus. In this study, we provide data on the global risk of intermediate *ATXN2* alleles for ALS.

Results

Clinical Phenotype of ALS with *de novo* CAG Expansions in Ataxin-2 Gene

We identified a pedigree with three ALS (two fully developed ALS: III-13 and III-16; and one in early stages of ALS: III-6) cases segregating with intermediate *ATXN2* CAG repeat which originated from unexpanded allele (Fig. 1). Average age at onset in this family was 54.25±1.26 years and the survival was 25±21.28 months. Ataxia, slow saccades or other symptoms which would indicate co-morbidity of ALS and SCA2 were not found in these patients. Clinical records of two cases showed neither ataxic signs nor cerebellar anomalies excluding SCA2 presenting initially as motor neuron disease. Additional clinical information of these cases is presented in Table 1 and described below.

The proband (III-16), appeared as a sporadic ALS case when assisted by consultants. This patient from pedigree A1 first experienced weakness and pain in the right heel at age of 53 years, progressing to paralysis of the leg, the left leg was also compromised within another month. MRI was indicated and misdiagnosis of lumbar radiculopathy was set. Treatment with non-steroidal anti-inflammatory drugs and Vitamin-B12 did not lead to an improvement. Four months later, upper right limb was started to be affected with wasting and weakness and within a month, it progressed to the left side. In all four extremities, hyperreflexia (+4) and moderate spasticity was present. Marked distal hypotrophy at first dorsal interosseum in both upper and lower limbs was observed with severe deterioration of first interosseum as well as in thenar and hypothenar eminences. Right hand adopted simian-like shape, requiring right-to-left hand change. About 6 months later, atrophy of arms, forearms, deltoid, and pectoralis muscles, accentuated pyramidalism featured by hyperreflexia, bilateral Hoffman and Troemmer signs predominating on the right side and Babinski reflex were evident. Sensation and coordination were normal. Both heels showed non-exhausting and exhausting clonus. Bulbar involvement was confirmed with remarked atrophy of the tongue edges, and tongue and palate fasciculations. Somatosensory evoked potentials of upper and lower limbs were normal. Electromyography (EMG) in muscles in all extremities (m. biceps brachii, gastrocnemius,

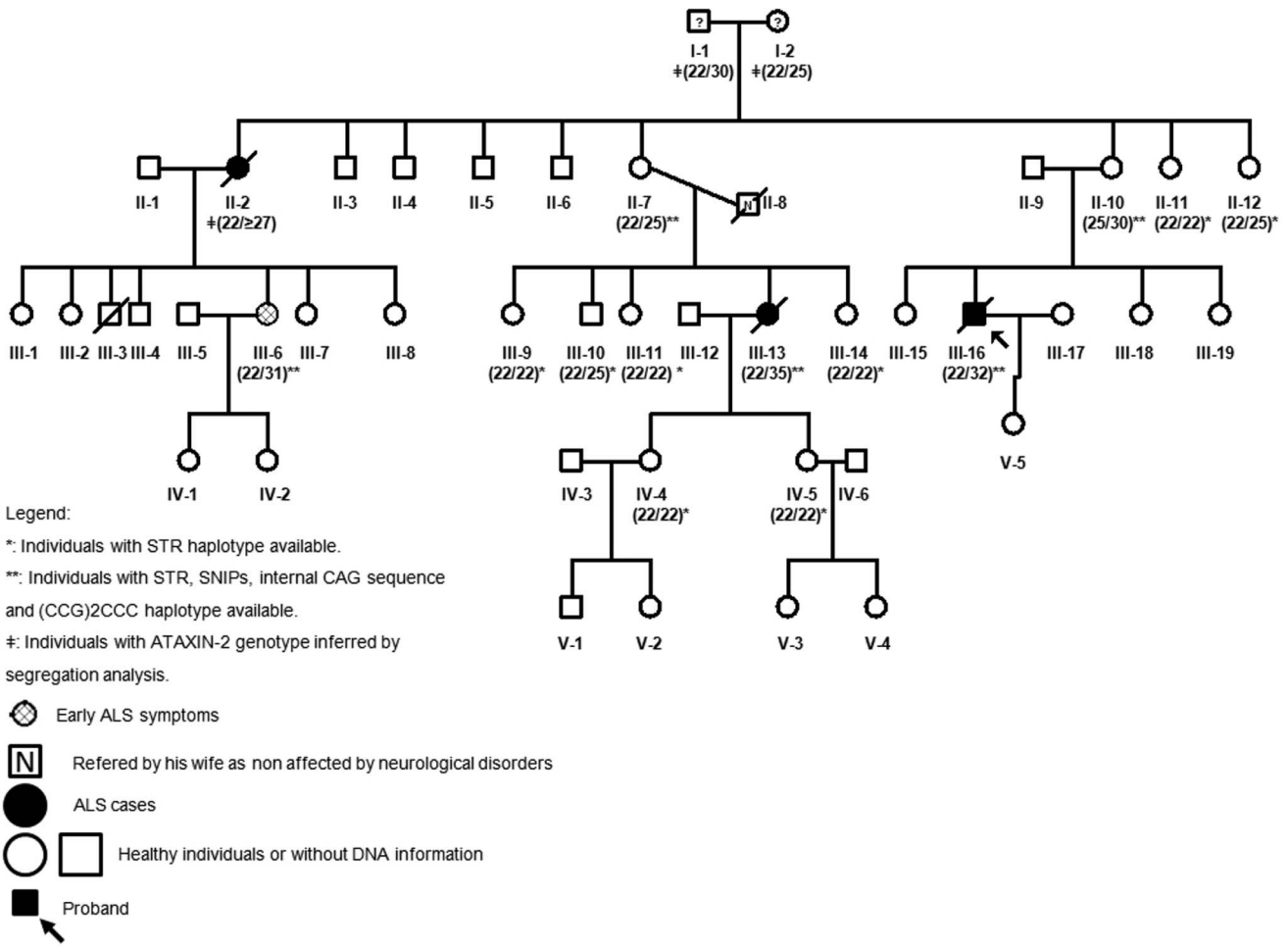


Figure 1. Pedigree of family in which novel sequence variants and three *de novo* mutations were identified.
doi:10.1371/journal.pone.0070560.g001

tibialis, rectus femoris, pectoralis, spinalis) and the tongue showed diffused signs, widespread fasciculations, fibrillations and sharp-waves. At rest, there were spontaneous discharges of motor unit action potentials in interosseal muscles. On contrast, dropout of the number of motor units with increased firing rate was evident. In summary, EMG report was indicative of neurogenic changes with signs of active denervation and renervation, spontaneous fasciculations (Fig. 2B) and fibrillations in all four limbs and bulbar segments, reflecting acute and progressive denervation of the second motor neuron of the brain stem and the anterior horn of spinal cord and a definite diagnosis of a classical ALS was concluded. Magnetic resonance imaging showed generalized cortical atrophy (not shown). No cerebellar damage was evident on a recovered sagittal MRI image.

Weakness in all the extremities progressed, and the patient subsequently developed dysarthria and dysphagia. Approximately 7 months later, the patient started to experience gait and later stance abnormalities and became quadriplegic, anarthric and aphagic. Neurological examination showed that patient was alert and oriented until his death. At this clinical stage he showed vasomotor disturbance congruent with Sudeck's atrophy. The entire clinical picture progressed several months until death caused by respiratory failure at age of 55 years. *ATXN2* CAG repeats length expansion was 22 and 32 CAG repeats (Fig. 2A, lane 14, Table 1) with the longer allele belonging in the intermediate range associated with ALS.

The sister of proband's mother (II-2) lived between years 1933–1991 and started to experience similar symptoms at age of 54 years. The disease progressed to death within 4 years (at age of 58 years). According to the clinical records her first symptom was hoarseness progressing throughout~15 years with pain (at age of 54 years) in the right heel and weakness in arms and legs. As she deteriorated, the course of her clinical picture occurred without dementia, cognitive or social dysfunction. A diagnosis of lumbar radiculopathy was proposed based on tomography performed in 1989. She subsequently developed dysarthria and dysphagia, was confined to wheelchair, and died due to respiratory failure. Although her genetic material was not available, a mutation near or beyond 31 CAG in *ATXN2* could be suspected, as her daughter's (III-6) genetic status of *ATXN2* is 22/31 CAG. Importantly, III-6 referred that, similarly to her mother, she has been suffering from progressive weakness and pain in lower limbs in last 7 months. Additionally, she complained about frequent cramps in the right arm and leg. At neurological examination, no signs of ataxia signs were found. She had motor neuron signs in all extremities and provoked fasciculations were evident in arms. Muscular strength in limbs was: superior limb was 4/5 distal, 4/5 proximal bilateral while in right inferior limb 4/5, 4/5 proximal bilateral and in left 4/5 distal. This clinical information is in agreement with early symptoms involving bombardment of motor neurons progressing toward typical ALS. In addition, her status highlights the possibility that the same genetic variants in *ATXN2* causing ALS in her mother might be the responsible for her clinical signs indicative for ALS. EMG was not performed in this patient due to her disapproval of this examination. While the definite ALS diagnosis in II-2 cannot be rule out, and was confounded in her time with lumbar radiculopathy, it seem clear that she presented a progressive neurodegenerative disorder involving motor neurons, and for her interviewed relatives this presentation resembled the picture of her first order nephews, our proband and III-13 (see below), both with a motor-neuron disease and where a hereditary pattern can be applied.

The examination of the patient III-13 (cousin of III-16) revealed weakness in the right leg and dysarthria lasting approximately 4 months. Similarly to the previous cases, the symptoms occurred at age of 54 years and led to death within 6 months. For about 3 months, the patient experienced muscle twitching in upper limbs and in m. gastrocnemius, as well as dysarthria, dysphagia and mild tongue atrophy. Neurological examination revealed spontaneous and provoked fasciculations in upper limbs, m. gastrocnemius and spinal muscles. Muscular strength in right upper limbs was 3/5 distal, 4/5 proximal bilateral while in right lower limb 3/5, 4/5 proximal bilateral and in left 4/5 distal. Both spatia interossea displayed hypotrophy with more profound picture on the right side and with marked deterioration in the first interosseum. Hyperreflexia (+3) and exhausting clonus in both heels was also observed. Babinski signs were discrete on the right side. Cutaneous sensibility was normal. On neuropsychological grounds cognitive and social function were preserved. Somatosensorial evoked potentials of upper and lower limbs were normal and MRI only showed generalized cortical atrophy and similarly to III-16 EMG of both upper and lower limb muscles (m. tibialis anterior, vastus medialis, gastrocnemius, first dorsal interosseous and biceps brachii) suggested denervation of second motor neuron compatible with ALS. On muscle contraction, decrease of recruitment pattern was evident, and in addition to motor neuron anomalies shown in III-16, a remarkably clear neurogenic EMG record was obtained in tongue all compatible with definite ALS. Genotype analysis revealed 22/35 CAG repeat size in *ATXN2* (Table 1), which is in the pathogenic range for SCA2. Since the first consultation, the clinical picture worsened progressively until quadriplegia in the last 3 months and the patient's death at age of 55 years due to respiratory failure.

All these cases showed intermediate *ATXN2*≥31 CAG, which are scattered in 80 chromosomes from the general population (frequency 0.025, and see Supplementary Figure S2 in reference 6) but overrepresented only in the SCA2/New Mutation deriving sample [6]. In that study an isolated allele with 35 CAG was found with no connections to SCA2 founder families [6], and as far as we know, the carrier has not developed neurological signs. This patient is now 37 years old, 18 years below the average onset age of ALS individuals presented in this study.

Segregation Analysis of Intermediate-length *De Novo* Mutated Alleles

DNA for genetic analysis was available from sib-ship pairs and revealed that at least three *de novo* mutations occurred. They have been proven in two patients (III-13 and III-16) and in one it was inferred (II-2) (Fig. 1). Female II-2 was the first person in the family with ALS symptoms and she was the first daughter of I-1 and I-2 deceased at age of 87 and 79 years, respectively, with no neurological disorders referred by their children (Fig. 1). They had eight other children (Fig. 1) with following *ATXN2* CAG repeat sizes: 22/25 (II-7), 25/30 (II-10), 22/22 (II-11) and 22/25 (II-12). The genetic material from II-2, II-3, II-4 and II-5 was not available. Neurological symptoms were observed only in one sibling (II-11) while the others appeared healthy despite advanced ages at examination (II-7, 63 years; II-10, 72 years; II-11, 67 years; II-12, 76 years). Siblings II-3, II-4 and II-5 were not examined as they live abroad. According to the family members, they are all healthy. Individual II-11 showed resting and kinetic tremor unrelated with ataxin-2 (normal genotype of 22/22 CAG) (Fig. 1) and different haplotypes in each allele (Table 2). According to the segregation analysis of this pedigree, the most probable genotype in the parents was 22/25 and 22/30. As an evidence for parental genotype serves the fact that the alleles with 25 and 30 CAG had different STR haplotype in the healthy 72-years old daughter (II-10) who is also the mother of the proband III-16 (Table 2).

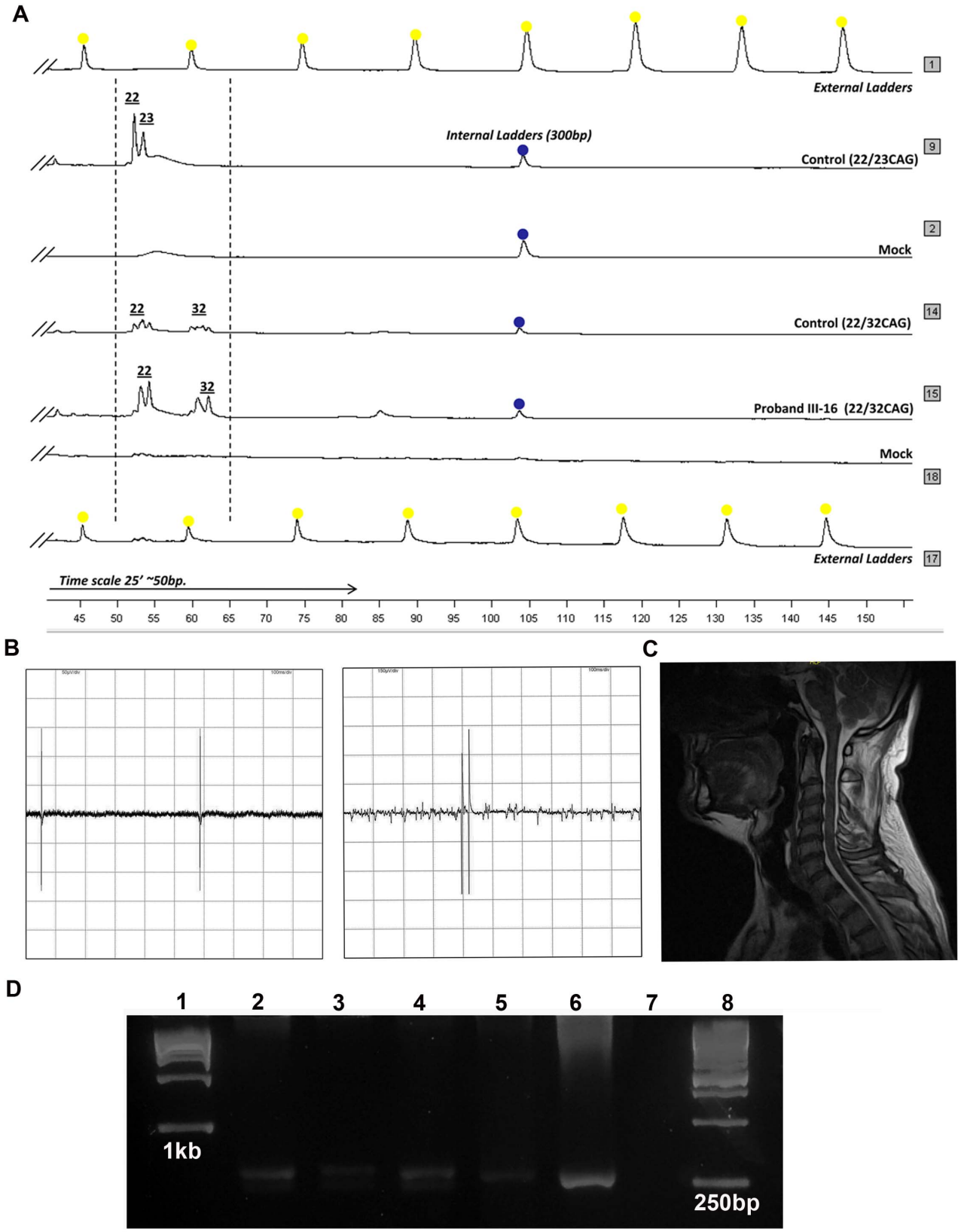


Figure 2. Genetics, EMG and MRI analysis. A) Electrophoresis of fluorescent fragment analysis of *ATXN2* CAG repeat. In each lane, the content is specified. B) Fasciculation patterns in proband's tongue and in biceps brachii recorded by EMG. C) Midsagittal MRI image of proband, no cerebellar atrophy is evident. D) Representative 3% agarose electrophoresis of *C9ORF72* analysis in index case and parents (lanes 2, 3, 4). Lanes 1and 8: MW markers (Ready Load™ 1–12.216 Kb ladder and 250–3500 bp ladder in multiples of 250 bp (Invitrogen), respectively; Lane 7: mock; lanes 2, 3, 4: case III-16, and both parents II-9 and II-10, respectively. Note that each DNA showed two defined bands despite PCR products with 7-deaza-2-deoxy GTP

stain poorly with ethidium bromide. These 3 samples were heterozygous with bands higher than 250 bp but bellow 350 bp (hex repeat ~11units) using Renton et al. primers anchoring 280 bp from hex-repeat. Lanes 5, 6 are unrelated ALS cases from the Cuban population.
doi:10.1371/journal.pone.0070560.g002

Moreover, 22/22 CAG and 22/25 CAG combinations were found in the other children (e.g. 22/25 in II-7, 22/22 in II-11, and 22/25 in II-12), suggesting the presence of a fourth allele in parents with 22 CAG repeats (for haplotypes see also Table 2 and Fig. 3A).

One of the longer alleles (25 or 30 CAG) would segregate in II-2 as a *de novo* ≥27 CAG ALS mutation which was further transmitted to III-6 (22/31 CAG) and might be responsible for the early signs of ALS (transmissions T1 and T2 in Fig. 4). The allele with 30 CAG could be suggested as the possible source for *de novo* ALS cases with *ATXN2* CAG expansions. However, segregation analysis for the transmission T4 (Fig. 4A) conformed by II-7 (22/25 CAG) and III-13 (22/35 CAG) revealed that instead of the 30 CAG allele, the one with 25 CAG in II-7 (age 76 years) underwent a pathogenic expansion by 10 CAG and contributed to ALS onset at age of 54 years. This *de novo* mutated allele in III-13 crossed the threshold for ALS and reached the minimal repeat length for SCA2 (34 CAG) (Fig. 4A). The origin of these expansions from the 25 CAG allele with one CAA interruption was also confirmed in the unstable transmission T3, sib-ship pair of II-10 (25/30 CAG)/III-16 (22/32 CAG) (Fig. 3B, Fig. 4A,B) expanded by 8 repeats and was found associated with classical ALS phenotype described above. The contribution of proband's father to *ATXN2* mutagenesis is not likely as he was a carrier of 22/22 CAG alleles. One of these alleles was identified in proband with 8+4+8 CAG/CAA pattern. Despite the absence of genetic data in mother of III-6, we could identify the increase in CAG length by 6 repeats as compared with CAG in the aunts (II-7 and II-10) (Fig. 4B). Normal CAG expansions were found in two of her sisters (III-2 and III-7) with22/22 CAG genotype. In summary, the average for intergenerational instability in all

Table 1. Phenotype and *ATXN2* genotype for ALS case series.

Case	Age	Clinical status	Gender	Ataxin-2 Genotype	D12S1333	D12S1672	D12S1332
I.1/I.2	–	Healthy	Female/Male	[22/25]	[251]	[289]	[190]
					NA	NA	NA
II.2	54	ALS*	Female	[22/≥27]	[239]	[289]	[198]
					NA	NA	NA
II.7	76	Healthy	Female	22/25	245	125	170
					251	289	190
II.10	72	Healthy	Female	22/25	239	125	170
					257	289	190
II.11	67	Tremor	Female	22/22	245	125	170
					245	289	198
II.12	63	Healthy	Female	22/25	231	125	170
					251	289	190
III.6	57	ALS	Female	22/31	231	129	170
					239	289	198
III.9	59	Healthy	Female	22/22	227	287	202
					245	125	170
III.10	57	Healthy	Male	22/25	245	125	170
					257	289	190
III.11	50	Healthy	Female	22/22	245	125	170
					257	283	196
III.13	54	ALS	Female	22/35	239	133	172
					239	289	202
III.14	38	Healthy	Female	22/22	245	125	170
					231	289	202
III.16	54	ALS	Male	22/32	231	129	170
					239	289	198
IV.4	38	Healthy	Female	22/22	239	133	172
					245	287	196
IV.5	29	Healthy	Female	22/22	239	133	172
					245	185	202

‡Affected by history and genotype inferred from daugther.
*De novo mutation causing disease. **EERC**: El Escorial Revised Classification.
doi:10.1371/journal.pone.0070560.t001

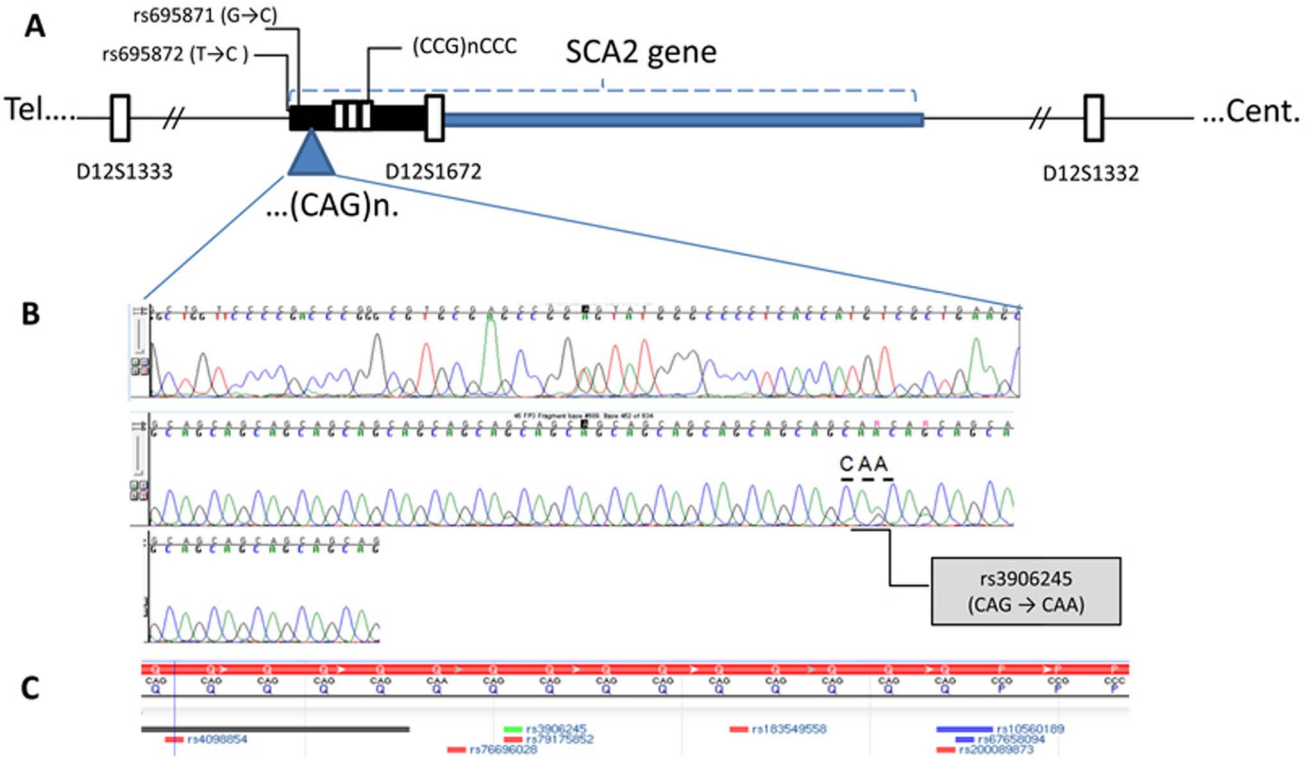


Figure 3. Genetic markers for *ATXN2* haplotyping and gene sequencing. A) *ATXN2* gene schematic maps and microsatellite, D12S1333 (telomeric at 200 kb from *ATXN2*), D12S1672 (intragenic at exon 1) and D12S1332 (centromeric at 350 kb from *ATXN2*) and SNIPs markers, rs695871 (at 177 bp upstream CAG expansion), rs695872 (at 106 bp upstream CAG expansion) and rs390624 (within the expanded CAG) used for haplotyping cases involved in *de novo* mutations, the polymorphic (CCG)nCCC/poly-proline adjacent to the CAG expansion is also indicated. B) Sequencing for case II-10, mother of the proband III-16 (25 CAG with only one CAA interruption). C) Relative position for other SNPs situated either within or near the expanded CAG.
doi:10.1371/journal.pone.0070560.g003

described transmissions was 6.25±2.87 gained CAG units in alleles sized 25.5±1 CAG, resulting in *de novo* mutated alleles with ~32CAG (transmission A in Fig. 4A) which is a greater CAG repeat than the threshold for ALS risk ≥30 CAG [3,7].

Haplotype of Ataxin-2 Intermediate-length Polyglutamine Expansions

To clarify the mechanism of *de novo* CAG mutations, we performed SNIPs and STR haplotyping. The analysed markers span ~550 kb and are lined up on chromosome 12q24.1 as follows: centromere-D12S1332-D12S1672-(CCG)nCCC-CAG-rs695871-rs695872-D12S1332-telomere (Fig. 3A). The mothers of patients III-13 and III-16 (II-7 and II-10) shared the truncated haplotype 6-4 (289 bp and 257 bp) for markers D12S1672 and D12S1332 embedding the 350 kb region including the SCA2 gene (Fig. 3A,B). For the D12S1333 telomeric marker at 20 kb, variation of 3 CA/GT dinucleotides was found (allele 4, 257 bp and allele 6, 251 bp), yielding closely related haplotype variants H1 (6-4-4) for II-7 and H′1 (6-4-6) for II-10.

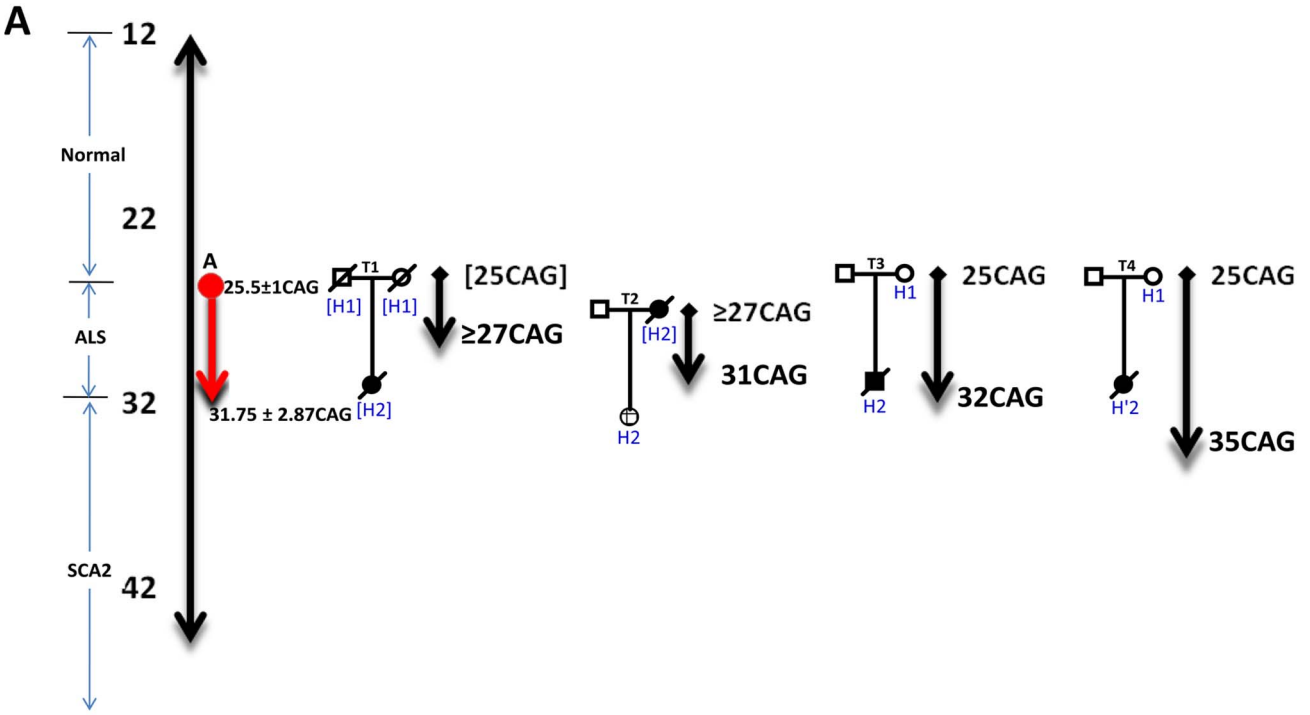
As for the intergenerational transmissions, STR haplotype analysis revealed that all CAG instabilities involving parental haplotypes H1 or H′1 lead to significant changes in offsprings with the resulting haplotype H2: 4-4-10 in III-6, III-16 and H′2: 3-4-10 in III-13 (Fig. 4A,B). Further DNA haplotyping in III-6 showed that H2 variant was shared with III-16 and II-2.

The haplotypes in generation III are distinct from the original parental haplotypes regarding the microsatellites D12S1333 and D12S1332 but not the intragenic marker D12S1672. The more

Table. 2. Microsatellite haplotypes.

No.	Genotype	Gender (Birth)	Clinical status	Age at onset/death	Survival (mo)	Site of onset, EERC
II-2	22/≥30*	Female (1933)	Affected ǂ	54/58	48 (deceased)	Spinal right (lower limb), ALS possible
III-6	22/31	Female (1955)	Early symptoms	56, Alive	–	–
III-16	22/32*	Male (1956)	Affected	53/55	21 (deceased)	Spinal right (lower limb), ALS definite
III-13	22/35*	Female (1957)	Affected	54/55	6 (deceased)	Spinal right (lower limb), ALS definite

NA: Not available, Inferred haplotype are bracketed,
*sick by history.
doi:10.1371/journal.pone.0070560.t002



B

CAG length and haplotype relationship

No.	Kinship	D12S1332	rs695871	rs695872	CAG internal structure	(CCG)nCCC	D12S1672	D12S1333	Difference in repeat size sib-ship/
II-2	Mother	[4]	[C]	[C]	ND	[(CCG)2CCC]	[4]	[10]	ND
II-7	Mother	6	C	C	16+8	(CCG)2CCC	4	4	
II-10	Mother	6	C	C	16+8	(CCG)2CCC	4	6	
III-6	Daugther	4	C	C	8+8+4+8	(CCG)2CCC	4	10	6CAG (III-6 vs II-10)
III-16	Son	4	C	C	23+8	(CCG)2CCC	4	10	8CAG (III-16 vs II-10)
III-13	Daugther	3	C	C	26+8	(CCG)2CCC	4	10	10 CAG(III-13 vs II-7)

Figure 4. CAG repeat instabilities and *de novo ATXN2* mutations associated with ALS in relationship with the haplotype. A) Schematic presentation of the CAG instabilities in four transmissions (T1–T4) of SCA2 (drawn to scale). Haplotypes are indicated as H1, H2, H′1and H′2 (see explanation in the text). Red (A), represents the average for CAG instabilities in the four transmissions. All transmissions reach either ALS (27–33 CAG) or SCA2 (≥34 CAG) thresholds. B) Scheme representing the haplotype and CAG(CCG)2CCC sequence for the individuals involved in the unstable CAG transmissions. Data between brackets is inferred from the segregation analysis. D12S1332, 3: 202 bp, 4: 198 bp, 6: 190 bp, D12S1672, 4: 289 bp, D12S1333, 4: 257 bp, 6: 251 bp, 10: 239 bp.
doi:10.1371/journal.pone.0070560.g004

significant changes of STR were evident in the vicinity of SCA2 locus, for example D12S1333 decreased 9 and 6 CA/GT and D12S1332 increased 6 and 4 CA/GT dinucleotides in III-13, III-6 and III-16, respectively, while no changes were seen in D12S1672 marker (Figure 4A,B).

As for the intragenic SNIPs and the polymorphic CCG/CCC tract, we found C-C and $(CCG)_2CCC$ variants for SNIPs rs695871 and rs695872, respectively, as the haplotype co-segregating with the intermediate and the normal (25 CAG) alleles (Fig. 4B, Table 2).

These results showed that the intragenic markers (SNIPs and D12S1672) are conserved but the flanking markers (D12S1332, D12S1333) are variable in the transmitted alleles, i.e.: 6-C-C-4-Intermediate CAG-(CCG_2CCC)-6/4 and 3/4-C-C-4-Intermediate CAG-(CCG_2CCC)-10, respectively.

Mechanism of De Novo Mutation: the Role of CAA Interruption Pattern

Next, we analyzed CAA interruptions of the CAG tract by DNA sequencing. The interruption patterns were as follows: 16+8 (II-7 and II-10), 23+8 (III-16), 26+8 (III-13), and 8+8+4+8 (III-6) (Fig. 4B). This analysis suggested sequential lengthening of the 5′ repeat tract in the alleles with the concurrent absence of the most proximal CAA interruption. Lack of this interruption enhances the enlargement of CAG tract, but no instability was found in the $(CCG)_2CCC$ triplets flanking the CAG repeat (Fig. 4B).

Ataxin-2 Intermediate Alleles and ALS Risk

Genetic overlap between SCA2 and ALS has been reported [5] without co-segregation of mutations in the most common ALS-related genes (*C9ORF72*, *TDP43*, *FUS*,*UBQLN2*,*ANG*,*OPTN*, *SPG11*, *PLEKHG5*, *VAPB*) in ALS cohorts with *ATXN2* intermediate-length (Table S1). We also found no segregation of *C9ORF72* hexanucleotide repeat expansion in our cases. Moreover, *de novo* mutations in our pedigree originated from the 25 CAG allele, which we consider as "large normal" [6] while other authors as intermediate ALS-risk alleles [2].Therefore we investigated whether this allele, which is overrepresented in our population [6], could be considered as a risk factor for ALS. Meanwhile, we aimed to address the refinement of currently accepted threshold of intermediate *ATXN2* alleles as risk factors for ALS. In the first step of the meta-analysis we used 24 or 27 CAG as threshold for ALS risk (Table 3). The combined risk was heterogeneous among 12 studies with $\chi^2 = 29.97$, df = 11, P = 0.0016. At this starting point we detected a higher prevalence of intermediate alleles in the ALS cases as compared with controls (OR = 1.23 95% CI 1.09–1.39). This modest value of association is also supported by the appearance of healthy controls with intermediate alleles and low

risk for this range in some populations [15]. In the population study in Belgium and Netherlands, the contribution of the 30 CAG alleles to the ALS risk has been shown, although the threshold at 27 CAG was confirmed [5] (Table 3). To clarify this, we performed meta-analysis with CAG range between 24 and 27 repeats, and the heterogeneity of the risk alleles dropped ($\chi^2 = 16.95$, df = 11, P = 0.11) (data not shown) with no significant risk of the alleles in this range (OR = 1.02; 95% CI 0.91–1.16). These results point to >27 triplet repeats length as one of the sources for heterogeneity seen in the first meta-analysis with threshold between 24 and 27 CAG. Furthermore, the data confirmed that the *de novo* mutations reported in this study originated from alleles with low or no risk for ALS with a propensity to cross the pathogenic range in further generations (Fig. 5A, Table 3 and Fig. 6).

When the meta-analysis was performed using repeat length threshold of ≥30 CAG, the ALS risk was estimated more accurately (Fig. 5B). We detected specific risks estimates in all populations (Table 3) and a higher global risk (OR = 2.16; 95% CI 1.76–2.65, $\chi^2 = 187.88$, P = 0.000) representing a two-fold excess of intermediate alleles enriching ALS cases versus healthy control population. Excluding the most precise studies [2,3,15] (Figure 5B) and those with the most extreme odds ratios [16,17] (Table 3) resulted in homogenous relative risk >1 (P = 0.12), which confidently input positive combined risk of 1.77; 95% CI 1.55–2.03 for intermediate allele in ALS patients versus controls. Although the risk for ALS was associated with the ≥30C AG alleles, it was higher in carriers of ≥32 CAG repeats (Fig. 5C,F) in all populations except Chinese [18] with no alleles with ≥32 CAG. The combined risk was "significantly heterogeneous" (2.62; 95% CI 2.23–3.09, $\chi^2 = 633.97$, df = 10, P = 0.0000) and eight cases-control series showed odds ratio greater than 2 (Table 3). In three populations, including USA [3]; French Canadian [19] and European [7] controls with >32 CAG were found. After independent exclusions of more precise risk estimates (Fig. 5C), such as Van Damme et al., 2011 (OR 2.70; 95% CI 2.20–3.29) [5]; Lee et al., 2011(OR 2.73; 95% CI 2.33–3.20) [3]; Elden et al., 2010 (OR 2.68; 95% CI 2.22–3.25) [2]; and Gellera et al., 2012 (OR 2.71; 95% CI 2.26–3.27) [20] the direction and magnitude of risk effect was no longer modified. Similar approach with exclusion of the most influential case-control series (the above data and Ross et al., 2011 [17], Van Langenhove et al., 2012 [16]) resulted in significantly heterogeneous risk estimates (combined OR = 2.34; 95% CI 1.94–2.83, $\chi^2 = 40.04$, df = 4, P = 0.0000) for CAG ≥32 repeats associated with ALS risk across populations. Overall our meta-analysis confirmed that the more accurate estimates in all populations are achieved when alleles with ≥30 CAG are used as cut-off for intermediates and the effect is more robust when CAG tract expansion is close to the upper limit of the *ATXN2*intermediate length. Moreover, it supported our findings on the pathological nature of *de novo* mutations found in the family reported here.

Discussion

In SCA2, similarly to other triplet repeat disorders, the *de novo* mutations are thought to result from enlargement of the CAG repeat tract during father-to-child transmissions of pre-expanded alleles (23–30 CAG) to a pathogenic range of ≥34 CAG. Here we report three ALS cases coinciding with ≥30 CAG *ATXN2* alleles. For the first time, expansion of the normal repeat tract of ≤25 CAG, were shown to expand to a CAG repeat range causing or contributing to disease. *De novo* mutations in polyglutamine (polyQ)-related genes causing diseases have been reported in few SCA7 [21–23], one SCA6 [24] and a small number of Huntington's Disease cases [25]. Regarding the *ATXN2* gene,

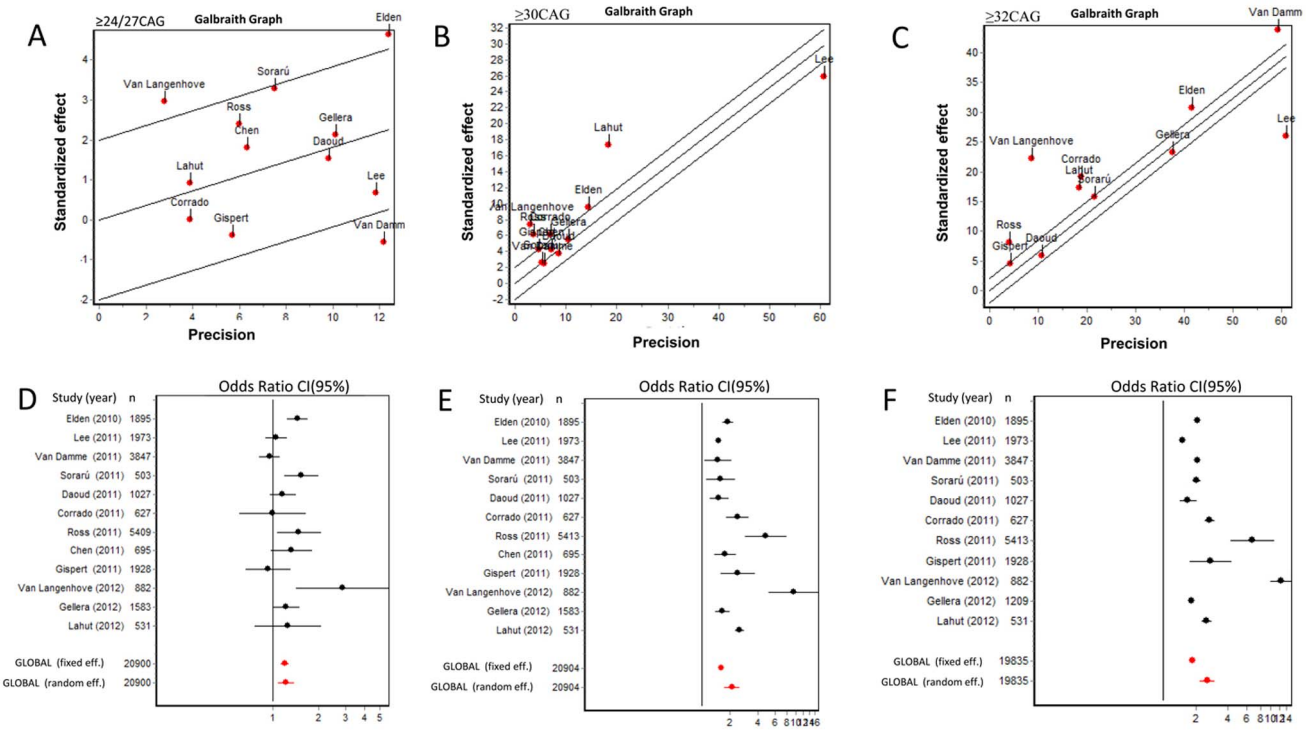


Figure 5. Meta-analyses for large and intermediate CAG lengths in *ATXN2*. Galbraith (A–C) and forest plots (D–F) for different *ATXN2* CAG length across populations. A) and D) ≥24/27 CAG. B) and E) ≥30 CAG. C) and F) ≥32 CAG.
doi:10.1371/journal.pone.0070560.g005

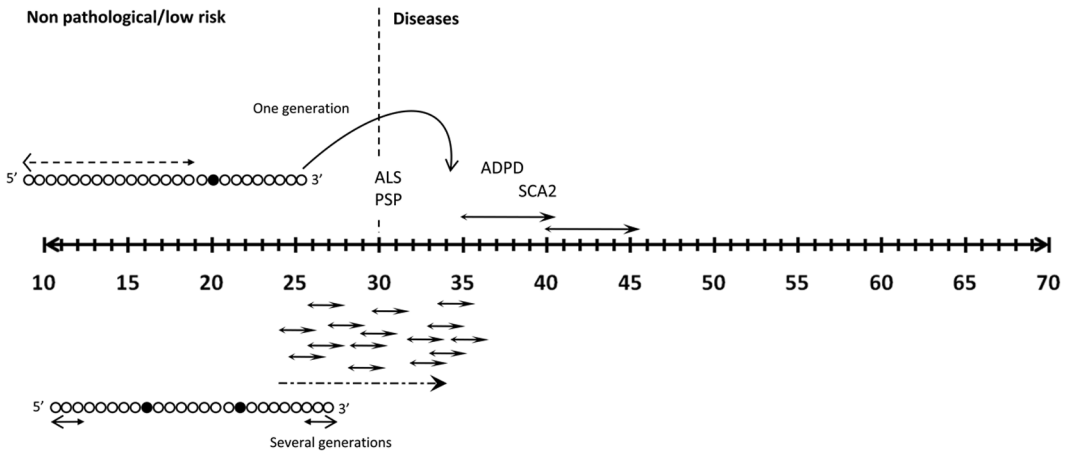


Figure 6. General mechanisms for *ATXN2* gene *de novo* mutagenesis in the population. Two models can be proposed for explanation of *de novo* CAG expansions in *ATXN2*. Both involve loss of the CAA interruption in large alleles resulting in a minimal length of pure repeat within the CAG expansion. CAA interruptions break the CAG tract in discrete repeat arrays protecting it from instability [17]. According to this study, the minimal length of the internal pure repeat leading to *de novo* mutations is 8 CAG.
doi:10.1371/journal.pone.0070560.g006

there is no consistent evidence of novel mutations arising from non-pathogenic alleles (13–31 CAG). Schöls et al., [26] interpreted as *de novo* mutation a father-to-daugther transmission with 34 CAG (currently recognized as pathogenic) expansion to 41 CAG in daugther presenting as a sporadic SCA2 case. The father died from cancer at age of 65 years without exhibiting ataxia. Alonso et al., [27] reported on a Mexican sib-ship pair where the parents (both with 22/22 CAG) appeared to transmit an intermediate-length allele with 33 CAG to a child with early SCA2 onset. Unfortunately, in the previous studies, neither STR nor SNPs haplotypes were shown preventing a clear conclusion on *de novo* mutations. Van Damme et al. [5] suggested in a cases-control study a possible *de novo* origin for expansions of ≥32 CAG linked with sporadic ALS in one family, but no clear conclusion was made as DNA samples from siblings and parents were not available. In this study, we present three confirmed *de novo* mutations associated with ALS (transmissions T1, T3 and T4) in a single family. ALS in this family appeared first as sporadic but presented as familial following Mendelian pattern of inheritance. The familial form of ALS connected with intermediate *ATXN2* alleles has further been supported in a French-Canadian cohort showing a stronger association of alleles with more than 29 CAG in *ATXN2* with familial rather than sporadic ALS [19]. In our family, autosomal dominant inheritance pattern may be applied for the occurrence of ALS with *de novo ATXN2* expansions. The lacking of genetic heterogeneity with other prominent ALS genes (Table S1), the concomitant segregation of the intermediate and full CAG repeat expansion in our proband (III-13) and III-16 respectively (both individuals affected), and its absence in 80 chromosomes of the Cuban population satisfied the gold standard

Table 3. Meta-analysis for case-controls series for ATXIN-2 polyQ intermediate length and ALS.

Study, Population	ALS cases/ Controls (ratio)	ATXN2 polyQ cutoff, P-value*	PolyQ>24/27 OR (95%CI)	24≥PolyQ≤27 OR (95%CI)	PolyQ≥30 OR (95%CI)	PolyQ≥32 OR (95%CI)
Elden et al., 2010, North America [2].	915/980 (0.93)	≥27, 3.6×10^{-5}	1.45 (1.24–1.70)	1.24 (0.99–1.55)	1.91 (1.67–2.19)	2.08 (1.99–2.18)
Lee et al., 2011, Northern Europe [3].	1294/679 (1.91)	>30, 6.2×10^{-3}	1.06 (0.90–1.25)	0.94 (0.75–1.16)	1.53 (1.48–1.58)	1.53 (1.48–1.58)
Van Damme et al., 2011, Belgium and the Netherlands [5].	1845/2002 (0.92)	≥32, 3.6×10^{-2}	0.96 (0.81–1.12)	0.91 (0.76–1.08)	1.49 (1.07–2.08)	2.09 (2.02–2.16)
Sorarú et al., 2011, Italy [28].	247/256 (0.96)	≥24, 2.6×10^{-2}	1.54 (1.19–2.00)	1.47 (1.05–2.08)	1.60 (1.12–2.30)	2.06 (1.88–2.25)
Daoud et al., 2011, French and French Canadian [19].	556/471 (1.18)	≥29, 2.4×10^{-4}	1.17 (0.96–1.42)	1.02 (0.79–1.33)	1.53 (1.22–1.93)	1.71 (1.43–2.04)
Corrado et al., 2011, Italy [4].	232/395 (0.59)	>30, 8.9×10^{-4}	1.00 (0.61–1.65)	0.42 (0.15–1.19)	2.41 (1.82–3.19)	2.73 (2.46–3.03)
Ross et al., 2011, North America [17].	536/4877 (0.11)	>30, 1×10^{-3}	1.49 (1.07–2.06)	1.19 (0.82–1.74)	4.81 (2.89–8.01)	6.8 (4.25–10.87)
Chen et al., 2011, Chinese [18].	345/350 (0.99)	≥32, 4×10^{-2}	1.33 (0.98–1.81)	1.08 (0.67–1.74)	1.78 (1.35–2.34)	–
Gispert et al., 2012, Europe [7].	559/1369 (0.41)	≥30, 4×10^{-3}	0.93 (0.66–1.32)	0.75 (0.50–1.13)	2.43 (1.61–3.67)	2.77 (1.78–4.32)
Van Langenhove et al., 2012, Flanders-Belgian [16].	72/810 (0.09)	≥30, 1.2×10^{-2}	2.86 (1.43–5.73)	1.79 (0.70–4.57)	9.54 (5.19–17.55)	12.41 (9.93–15.51)
Gellera et al., 2012, Italy [20].	658/551 (1.19)	>30, 1.4×10^{-3}	1.23 (1.02–1.50)	0.92 (0.68–1.24)	1.66 (1.38–2.00)	1.85 (1.75–1.95)
Lahut et al., 2012, Turkish [15].	212/319 (0.66)	>30, 2.6×10^{-2}	1.26 (0.76–2.08)	0.83 (0.37–1.86)	2.53 (2.28–2.82)	2.53 (2.27–2.81)

*As reported by authors in the original data.
doi:10.1371/journal.pone.0070560.t003

criteria pointing to a causative role. In two pedigrees, in which *SOD1*, *TARDBP*, *FUS*, and *ANG* mutations had been excluded, a co-segregation of the ALS phenotype with expanded *ATNX2* alleles has been reported [5]. The CAG expansions were either in the SCA2-parkinsonism or classic cerebellar SCA2, which, together with the reported genetic overlap [5] and without clinical heterogeneity or ALS phenotype modification [4,5,15,19,28] support our hypothesis of CAG expansion in *ATXN2* being monogenic ALS cause (at least for longer intermediate CAG expansions). In addition, *de novo ATXN2* CAG expansions may also contribute to apparent sporadic cases as reported by Elden et al [2]. This repeat behavior is supported by the meiotic and mitotic instability associated with large alleles [6]. All these observations may serve as a starting point for intergenerational risk estimations for CAG repeat instabilities in the large non-SCA2 expansions in *ATXN2* with ≥25 CAG.

Intermediate-length CAG expansions in *ATXN2* can lead to atypical SCA2 phenotype [8–10], however their association with ALS phenotype featured by shorter survival and onset at ~54years of age helps to avoid misdiagnosis of ALS as SCA2 with motor neuron disease [29]. In our SCA2 population, ataxia combined with motor neuron phenotype has been noted [30], but such cases segregated within SCA2 pedigrees with cerebellar symptoms and a long disease duration excluding ALS. The intermediate *ATXN2* alleles associated either with familial or sporadic ALS arose as *de novo* CAG expansions unrelated to SCA2 families in our population.

Interestingly, one of the *de novo* mutations contributing to ALS crossed the pathogenic SCA2 threshold (35 CAG; T4 transmission) suggesting a similar mutagenic mechanism for both diseases (Fig. 6). The 35 CAG repeat expanded from a large normal allele which are overrepresented in Cuba [6] confirming their role as a source for new expansions.

The allele with 25 CAG repeat with only one CAA interruption in II-7 was transmitted as the 35 CAG expansion to III-13 (Fig. 4A,B). Similar allele would be the disease-causing mutation in II-2 and it is plausible to assume that in the preceding generation (grandparents and parents healthy), sequential loss of CAA in the 25 CAG allele resulted in CAG enlargements leading to ALS phenotype in successive generations (bottom model in Fig. 6).

Retrospective studies made in mortality archives between years 2001 and 2006 in Cuba showed low frequency of ALS, but Holguín province was third in the rating of death by ALS [33]. An introduction of a predisposed founder chromosome in this region would be proposed to contribute considerably to the ALS incidence in this region.

While the CAG sequence in SCA2 is pure, the ALS-related expansions are interrupted by CAA [4,34] suggesting different origins. Two internal CAG repeat structure either with single or three CAA interruptions in sporadic ALS case series was interpreted as multiple mechanisms for *ATXN2* repeat expansions [17]. We observed a trend of pure CAG tracts gaining and removal of CAA interruptions in successive generations as shown in case II-10 (CAG 23+8) versus III-6 (CAG 8+8+4+8), respectively, indicating that both expansions share the same ancestor as previously proposed [35], and that mutational mechanism driving the removal or gaining CAA interruptions might occur concurrently.

Using microsatellites haplotype analysis, we found that the *de novo* mutations originate from the same ancestors although the extragenic microsatellites underwent some changes. On the other hand, we found SCA2 families in which the haplotype remained unchanged across generations [6,36,37]. This suggests that there may be regions in the vicinity of *ATXN2* which can interact genetically and epigenetically [38] and delineate and/or contribute further to ALS phenotype. This hypothesis is in accordance with the study Lahut et al. [15].

Two intragenic SNPs were used for haplotype analysis: rs695871 (G/C) and rs695872 (T/C) in the first exon of *ATXN2* gene situated 177 and 106 bp, respectively downstream from the CAG tract. The G→C substitution results in Val to Leu change while the T to C polymorphism is silent (Arg residue).The haplotype variants are CC or GT [17]. The CC haplotype was present in the 25 CAG allele, which expanded to intermediate CAG (Fig. 3B).This haplotype was confirmed also by the identification of the $(CCG)_2CCC$ sequence which is useful for haplotype discrimination [39] and evolutionarily conserved in primates [34]. The CC haplotype was connected with uninterrupted CAG tract and associated with intergenerational instability and ALS disease development in our three patients. This haplotype was reported in Indian population as enriched inSCA2 founders and in a subset of predisposed normal repeats, while the GT variant was found in the healthy population reducing the repeat sequence instabilities due to the presence of several CAA interruptions [34].Taken together, these data suggest that the presence of the CC haplotype, loss of CAA interruptions and the repeat size are prerequisite for CAG tract instability while the GT variant appears to stabilize the repeat even in large normal and intermediate CAG expansions and to prevent the SCA2-related mutagenesis. Surprisingly, however, the ALS cases carrying GT variant with three CAA interruptions developed the disease earlier than those with CC variant and less interruptions. Moreover, the ALS patients with three interruptions had shorter CAG than patients with fewer CAA interruptions [40].These observations suggest that besides the CAG repeat size, the discrete changes in *ATXN2* gene sequence probably affecting ataxin-2's physiological function may be a risk factors for ALS development and 'genetic hits' facilitating the disease.

The effect of $(CCG)_2CCC$ (encoding poly-proline tail varying from 1–4) on CAG instability and possibly ALS phenotype should be also considered as it may contribute to both to the CAG (in)stability increasing the number of continuous triplets folding as hairpin, and to the content of prolines residues which lowers the complexity of unstructured sequence regions in the N-terminal tail of ataxin-2 [41].

Our meta-analysis further confirms the observation that the increased risk for ALS is specifically associated with long intermediate *ATXN2* repeats of ≥30 CAG and is greater with larger expansions. The heterogeneity observed in the combined risk is explained by longer intermediate alleles as the main factor. Pending other factors such as sample selection, gender ratio, SNPs haplotype or CAG calling approaches must be warranted and homogenized in further studies. Regarding this last technical point, the heterogeneity may reside in the range where authors don't use sequencing methods, but for ≥27 CAG, none heterogeneity is expected, and great part is driven for the risk associated with CAG size variation only.

High number of intermediate alleles with 27–31 CAG not associated with any phenotype [6] complicates the interpretation of the "ALS threshold". Only two out of 25 cases carrying intermediate CAG (32 and 33 CAG) had mild SCA2 phenotype and all intermediate alleles segregated within SCA2 pedigrees [6]. However, in a single family here, we showed three *de novo* mutated alleles penetrating with ALS disease. The question is what factors differentiate intermediate-length alleles segregating in SCA2 families from those associated with ALS? Our study point to a possible cis-acting factor, as STR haplotype changes were identified in all the *de novo* mutations.

In several back-to-back reports, *ATXN2* intermediate alleles associated with ALS have been defined differently across the populations, e.g. in the USA ≥27 CAG [2], in Europe ≥30 CAG [3], in Belgium and Netherland ≥32 CAG [5], suggesting a range of at-risk alleles with 27–30 CAG and alleles causing the disease with ≥31 CAG. This view is corroborated by recent studies showing escalating effect for alleles sized with 27Q, 29Q, 31Q and 32Q in the accumulation of phosphorylated and truncated TDP-43 in response to cell stress [42,43]. Both post-translational modifications of TDP-43 are hallmark features of ALS pathology [44].

In conclusion, we identified de novo CAG expansions in *ATXN2* causing ALS. The mutational mechanism involved the loss of CAA anchors in large normal alleles on a predisposed genetic background, leading to a subsequent CAG instability. Intermediate ATXN2-ALS alleles segregated as *de novo* mutations in families in the general population, highlighting the necessity for providing *ATXN2* genetic testing in ALS patients.

Methods

Patient's Resources and Clinical Characterization

First diagnosis of all patients was performed in the neuromuscular consultation at Clinical and Surgical Hospital Lucía Íñiquez Landín, Holguín, Cuba. Genetic status of the patients was not known at the time of first consultation. These individuals were enrolled in the research since they either met the revised El Escorial revisited criteria [45] for definite ALS or were related to ALS patients.

This study included a cohort of 17 individuals, four with familial ALS and 13 relatives. All studied subjects signed an informed consent form after being explained the purpose and methods of the study. All studies were approved by the review boards and ethics bureau of Center for Research and Rehabilitation of Hereditary in Holguin. Specifically, genealogical data were obtained for each family member and genetic DNA information was generated in regard to CAG expansions in the *ATXN2* gene aimed at determination of the founder allele.

The DNA testing of *ATXN2* was performed at the Department of Molecular Neurobiology of the National Center for the Research and Rehabilitation of the Hereditary Ataxias (CIRAH), Holguín. Since 1993, specialized database is available in this center enabling to confirm that there were no relationships between ALS cases and the 124 SCA2 families in the Cuban population. The genealogy database includes information from 124 multigenerational SCA2 families with more than 10.000 members across 15 generations, with names, demographic, genetic and clinical information.

Determination of *ATXN2* CAG Repeat Size and SNIPs Haplotype

Peripheral blood leukocytes were extracted using EDTA as anticoagulant and genomic DNA was isolated using standard methods after informed consent was signed by patients or their guardians. In addition to SCA2, a complete set of polyQ diseases-related genes (SCA1, 3, 6, 7, 8, 17, HD) were analysed, with no remarkable findings. CAG repeat length in *ATXN2* was determined using Alfexpress II sequencing system and CAG fragments were separated with ReproGel^{TM} high resolution system (GE Healthcare, Buckinghamshire, UK). PCR was performed as previously reported [6]. For detecting single nucleotide polymorphisms (SNPs), DNA was amplified with SCA2-FP3 and SCA2-RP3 primers [17] and using PCR conditions as previously reported in Ramos et al., 2010 [35], using 1 unit of TopTaq polymerase per reaction (Qiagen, Mainz Germany). For Short Tandem Repeat (STR) sequencing, oligonucleotides used for fragment analysis were also used for sequence reactions. Gel extraction was performed using GFX^{TM} pCR DNA and Gel Band Purification kit (GE Healthcare, Buckinghamshire, UK) and the sequencing reactions were developed on Beckman Coulter CEQ8000 automated sequencing equipment with GenomeLab 10.2 software (Beckman Coulter, Inc. Belgium) at IPK. For sequencing, samples were blinded for the IPK investigators. Reactions were also done in both sense using FP3-RP3, and SCA2A primers, resulting contigs were assembled and compared with the Reference Sequence NG_011572.1. ALS patients were not genotyped for SOD1 and FUS mutations, but we exclude the contribution of the pathogenic GGGGCC expansions of C9ORF72 by using both repeat-primed PCR assays [31,32].

Short Tandem Repeat Haplotyping

Three chromosome 12 microsatellite CA repeat markers were used to establish the haplotypes in this study. Markers D12S1333 and D12S1332, with approximate 200 kb telomeric and 350 kb centromeric location, respectively from the gene, and the third marker, Dl2S1672, is located in the first intron of the *ATXN2* gene. PCR reaction was followed by Gene Scan analysis for STR allele identification. Reactions were performed in 13 µl containing ~ 10 mM Tris-HCl (pH 8), 1.5 mM $MgCl_2$, 50 mM KCl, 10% DMSO, 250 mM dNTPs, 100 ng of each of the primers, 1.5 U Taq-polymerase, and ~100 ng genomic DNA. The following primer pairs and annealing temperatures were used: Dl2S1332a-b Cy5-GCC AGG TAC AGT GGC TC/CTG GGA CCA CAG GTG TAG at 60°C, D12S l333a-b Cy5-TTC AGG TGG TAC AGC CGT/CAT CAG AAG GCT TCA TAG GAA T at 50°C and D12Sl672a-b Cy5-CAG AGG GAG ATT CCA TCC AA/ CGG TTT GAC AAG TTTCGA GA at 50°C. STR lengths were determined by Alfexpress II sequencing system and the PCR fragments were separated with ReproGel high resolution system. Internal (100 and 300 bp) and external (50–500-step, 50 bp) Alfexpress ladders were used to determine the fragment size. Traces were analyzed using the software Allelelink according to the manufacturer's specifications.

Meta-analysis

Meta-analysis was performed using odds ratio as measure of overall ALS risk, and the data were retrieved from the literature. Data were tabulated and processed in the program EPIDAT 3.1 of the Pan-American Health Organization (PAHO). We searched the literature investigating the *ATXN2* intermediate alleles as risk factors for ALS and performed a meta-analysis when applicable of case–control association (Table 3). Potentially useful studies were identified in review articles and through PubMed and Scopus, using the key words: ALS, ataxin-2 intermediate alleles and polyQ. The content of studies was also assessed manually to avoid redundant or unrelated data. We considered all studies published between 2010 until 2012. We avoided double counting using only the largest available published dataset from any study described in more than one published article. Given the heterogeneity in the nomenclature and thresholds of intermediate alleles we first considered threshold for intermediate alleles as originally authors reported in each work (i.e. 24, 27 and ≥30) (see also Table 3).

Thresholds for 24, 27, 30, 32 CAG repeats used here as starting point in our meta-analysis, were in accordance to the current literature reporting, where these thresholds were previously tested either as statistical significant or functional in each study (for details about rationale for thresholds see Text S1). For discriminating specific thresholds we applied component analysis, first

analyzing the risks beginning with ≥24/27 CAG, ≥30 CAG, then we eliminated the contribution of the last alleles and the last cut-off was ≥32 CAG. The Dersimonian and Laird's test was used to estimate the heterogeneity between relative risks in each metanalysis. When significant heterogeneity was absent, odds ratio and 95% confidence intervals (CIs) were calculated using a fixed-effects model while when significant heterogeneity was present, a random-effects model was used. Publication bias was investigated using funnels plots and both Begg's and Egger's tests. The significance level was set to $P = 0.05$.

We included 7, 471 ALS cases and 13,059 controls (ratio ~1.75 control/cases) harboring 12 studies and including samples from North America, Europe, French-Canadian population, China, Italy, Belgium (Flanders), Germany and Turkey, published between 2010 and 2012 (Table 3). The studies were identified by the search strategy described above, met the inclusion criteria, and contributed to the meta-analysis (Table 3). Reports related with cases-controls studies in intermediate-length *ATXN2* CAG expansions and ALS in other languages than English were not found. There was no overlap in reporting ALS cases from the same country as authors did not specify whether samples were the same, which is not possible regarding that are different referral centers representing different regions. Lattante et al., 2012 [46] reported cases of sporadic ALS with *ATXN2* CAG expansions, but this study had no cases-controls design, therefore it was not included in our meta-analysis. Other two North-American studies [40,47] were not included because were considered overlapping with Elden et al., 2010 [2]. During the revision of the article two additional studies has been published recently, two original from Italy and China, but were not included and did not modify our metanalysis [48,49]. In the Chinese study, the repeat in *ATXN2* ranged between 30–35 CAG [48].

Supporting Information

Table S1 Major Mutations in ATXN2-ALS case series.
(XLSX)

Text S1 Rational for polyQ thresholds used in Meta-Analysis.
(DOC)

Text S2 Abstract in Japanese.
(DOCX)

Acknowledgments

We thank the reviewers for their thorough reading of the manuscript and their useful comments. We are indebted to the family A1 involved in this study for all their cooperation. Support of Dr. Ocilia Rodriguez is acknowledged. We would like to thank Dr. Rosa Rademaker, Patricia Brown and Mariely DeJesus-Hernandez from Mayo Clinic for providing the DNA positive control for *C9ORF72* expansion and primers for C9FTD/ALS hex-repeat expansion detection and for the technical assistance Acknowledgement belongs also to our colleagues from Neurobiology Department at CIRAH.

Author Contributions

Conceived and designed the experiments: JMLM. Performed the experiments: JMLM JMRP RMS YVM VK LLS JAVF MMG RRL. Analyzed the data: JMLM JMRP RMS YVM VK LLS JAVF MMG YGZ RRL. Contributed reagents/materials/analysis tools: JMLM PS LVP. Wrote the paper: JMLM POB MP PS LVP.

References

1. Andersen PM, Al-Chalabi A (2011) Clinical genetics of amyotrophic lateral sclerosis: what do we really know? Nat. Rev. Neurol 7: 603–615.
2. Elden AC, Kim HJ, Hart MP, Chen-Plotkin AS, Johnson BS, et al. (2010) Ataxin-2 intermediate-length polyglutamine expansions are associated with increased risk for ALS. Nature 466: 1069–1075.
3. Lee T, Li YR, Ingre C, Weber M, Grehl T, et al. (2011) Ataxin-2 intermediate-length polyglutamine expansions in European ALS patients. Hum Mol Genet 20: 1697–1700.
4. Corrado L, Mazzini L, Oggioni GD, Luciano B, Godi M, et al. (2011) ATXN-2 CAG repeat expansions are interrupted in ALS patients. Hum Genet. doi10.1007/s00439-011-1000-2.
5. Van Damme P, Veldink JH, van Blitterswijk M, Corveleyn A, van Vught P, et al. (2011) Expanded ATXN2 CAG repeat size in ALS identifies genetic overlap between ALS and SCA2. Neurology 76: 2066–2072.
6. Laffita-Mesa JM, Velázquez-Pérez LC, Santos Falcón N, Cruz-Mariño T, González Zaldívar Y, et al. (2012) Unexpanded and intermediate CAG polymorphisms at the SCA2 locus (ATXN2) in the Cuban population: evidence about the origin of expanded SCA2 alleles. Eur J Hum Genet. 2011. doi:10.1038/ejhg.2011.154.
7. Gispert S, Kurz A, Waibel S, Bauer P, Liepelt I, et al. (2011) The modulation of Amyotrophic Lateral Sclerosis risk by Ataxin-2 intermediate polyglutamine expansions is a specific effect. Neurobiol. Dis: doi:10.1016/j.nbd.2011.08.021.
8. Futamura N, Matsumura R, Fujimoto Y, Horikawa H, Suzumura A, et al. (1998) CAG repeat expansions in patients with sporadic cerebellar ataxia. Acta Neurol Scand 98: (1), 55–9.
9. Fernandez M, McClain ME, Martinez RA, Snow K, Lipe H, et al. (2000) Late-onset SCA2: 33 CAG repeats are sufficient to cause disease. Neurology 22, 55(4), 569–72.
10. Costanzi-Porrini S, Tessarolo D, Abbruzzese C, Liguori M, Ashizawa T, et al. (2000) An interrupted 34-CAG repeat SCA-2 allele in patients with sporadic spinocerebellar ataxia. Neurology 54: 491–493.
11. Santos N, Aguiar J, Fernández J, Vázquez M, Auburger, etal. (1999) Molecular diagnosis of a sample of the Cuban population with spinocerebellar ataxia type 2. BiotecnologiaAplicada, 16: 219–221.
12. Pulst SM, Nechiporuk A, Nechiporuk T, Gispert S, Chen XN, et al. (1996) Moderate expansion of a normally biallelic trinucleotide repeat in spinocerebellar ataxia type 2. Nat Genet 14: 269–276.
13. Mao R, Aylsworth AS, Potter N, Wilson WG, Breningstall G, et al. (2002) Childhood-onset ataxia: testing for large CAG-repeats in SCA2 and SCA7. Am J Med Genet 15: 110(4), 338–45.
14. Paciorkowski A, Shafrir Y, Hrivnak J, Patterson MC, Tennison MB, et al. (2011) Massive expansion of SCA2 with autonomic dysfunction, retinitis pigmentosa, and infantile spasms Neurology 77, 1055–1060.
15. Lahut S, Omur O, Uyan O, Agım ZS, Ozoguz A, et al. (2012) ATXN2 and Its Neighbouring Gene SH2B3 Are Associated with Increased ALS Risk in the Turkish Population. PLoS ONE 7(8): e42956. doi:10.1371/journal.pone.0042956.
16. Van Langenhove T, van der Zee J, Engelborghs S, Vandenberghe R, Santens P, et al. (2011) Ataxin-2 polyQ expansions in FTLD-ALS spectrum disorders in Flanders-Belgian cohorts. Neurobiol Aging. doi:10.1016/j.neurobiolaging.2011.09.025.
17. Ross OA, Rutherford NJ, Baker M, Soto-Ortolaza AI, Carrasquillo MM, et al. (2011) Ataxin-2 repeat-length variation and neurodegeneration. Hum MolGenet 20: 3207–3212.
18. Chen Y, Huang R, Yang Y, Chen K, Song W, et al. (2011) Ataxin-2 intermediate-length polyglutamine: a possible risk factor for Chinese patients with amyotrophic lateral sclerosis. Neurobiol Aging 32: 1925 e1921–1925.
19. Daoud H, Belzil V, Martins S, Sabbagh M, Provencher P, et al. (2011) Association of long ATXN2 CAG repeat sizes with increased risk of amyotrophic lateral sclerosis. Arch Neurol 68: 739–742.
20. Gellera C, Ticozzi N, Pensato V, Nanetti L, Castucci A et al. (2012) ATAXIN2 CAG-repeat length in Italian patients with amyotrophic lateral sclerosis: risk factor or variant phenotype? Implication for genetic testing and counseling. Neurobiol Aging. doi:10.1016/j.neurobiolaging.2012.02.004.
21. Stevanin G, Giunti P, Belal GD, Belal S, Durr A et al. (1998) De novo expansion of intermediate alleles in spinocerebellar ataxia 7. Hum Mol Genet 7: (11), 1809–13.
22. Bauer P, Kraus J, Matoska V, Brouckova M, ZumrovaAet al. (2004). Large de novo expansion of CAG repeats in patient with sporadic spino-cerebellar ataxia type 7. J Neurol 251, 1023–1024.
23. Mittal U, Sanghamitra R, Satish J, Srivastava AK, Mukerji M et al. (2005). Post-zygotic de novo trinucleotide repeat expansion at spinocerebellar ataxia type 7 locus: evidence from an Indian family J Hum Genet 50: 155–157.
24. Shizuka M, Watanabe M, Ikeda Y, Mizushima K, Okamoto K, Shoji M (1998) Molecular analysis of a de novo mutation for spinocerebellar ataxia type 6 and (CAG)n repeat units in normal elder controls. J Neurol Sci 161: 85–87.
25. Almqvist EW, Elterman DS, MacLeod PM, Hayden MR (2001) High incidence rate and absent family histories in one quarter of patients newly diagnosed with Huntington disease in British Columbia. Clin Genet 60: 198–205.

26. Schöls L, Szymanski S, Peters S, Przuntek H, Epplen JT et al. (2000) Genetic background of apparently idiopathic sporadic cerebellar ataxia. Hum Genet, 107: 132–137.
27. Alonso E, Martínez-Ruano L, De Biase I, Mader C, Ochoa A, et al. (2007) Distinct distribution of autosomal dominant Spinocerebellar Ataxia in the Mexican population. Movement Disorders 22: 1050–1053.
28. Soraru G, Clementi M, Forzan M, Orsetti V, D'Ascenzo C, et al. (2011) ALS risk but not phenotype is affected by ataxin-2 intermediate length polyglutamine expansion. Neurology 76: 2030–2031.
29. Fischbeck KH, Pulst SM (2011) Amyotrophic lateral sclerosis and spinocerebellar ataxia type 2. Neurology 76: 2050–2051.
30. Velázquez-Pérez L, Rodríguez-Labrada R, García-Rodríguez JC, Almaguer-Mederos LE, Cruz-Mariño T, et al. (2011) A Comprehensive Review of Spinocerebellar Ataxia Type 2 in Cuba. Cerebellum10: 184–198.
31. DeJesus-Hernandez M, Mackenzie IR, Boeve BF, Boxer AL, Baker M, et al. (2011) Expanded GGGGCC hexanucleotide repeat in noncoding region of C9ORF72 causes chromosome 9p-linked FTD and ALS. Neuron 72: 245–256.
32. Renton AE, Majounie E, Waite A, Simón-Sánchez J, Rollinson S, et al. (2011) A hexanucleotide repeat expansion in C9ORF72 is the cause of chromosome 9p21-linked ALS-FTD. Neuron 72: 257–268.
33. Zaldivar T, Gutiérrez J, Lara G, Carbonara M, Logroscino G, et al. (2009) Reduced frequency of ALS in an ethnically mixed population A population-based mortality study. Neurology 72: 1640–45.
34. Choudhry S, Mukerji M, Srivastava AK, Jain S, Brahmachari. (2001) CAG repeat instability at SCA2 locus: anchoring CAA interruptions and linked single nucleotide polymorphisms. Hum Mol Genet 10(21): 2437–46.
35. Ramos EM, Martins S, Alonso I, Emmel VE, Saraiva-Pereira ML, et al. (2010) Common origin of pure and interrupted repeat expansions in spinocerebellar ataxia type 2 (SCA2). Am J Med Genet B Neuropsychiatr Genet153B: 524–531.
36. Gispert S, Twells R, Orozco G, Brice A, Weber J, et al. (1993) Chromosomal assignment of the second locus for autosomal dominant cerebellar ataxia (SCA2) to chromosome 12q23–24.1. Nat Genet 4(3), 295–9.
37. Hernandez A, Magarino C, Gispert S, Santos N, Lunkes A, et al. (1995) Genetic mapping of the spinocerebellar ataxia 2 (SCA2) locus on chromosome 12q23-q24.1. Genomics 20: 25(2): 433–5.
38. Laffita-Mesa JM, Bauer PO, Kourí V, Peña Serrano L, Roskams J, et al. (2012) Epigenetics DNA-Methylation in the core ataxin-2 gene promoter: Novel physiological and pathological implications. Hum Genet. doi:10.1007/s00439-011-1101-y.
39. Mizushima K, Watanabe M, Kondo I, Okamoto K, Shizuka M, et al. (1999) Analysis of spinocerebellar ataxia type 2 gene and haplotype analysis: (CCG) 1–2 polymorphism and contribution to founder effect. J Med Genet 36: 112–114.
40. Yu Z, Zhu Y, Chen-Plotkin AS, Clay-Falcone D, McCluskey L, et al. (2011) PolyQ repeat expansions in ATXN2 associated with ALS are CAA interrupted repeats. PLoS One 6: e17951.
41. Albrecht M, Golatta M, Wullner U, Lengauer T (2004) Structural and functional analysis of ataxin-2 and ataxin-3. Eur J Biochem 271: 3155–70.
42. Hart M, Gitler A (2012) ALS-Associated Ataxin 2 PolyQ Expansions Enhance Stress-Induced Caspase 3 Activation and Increase TDP-43 Pathological Modifications. doi:10.1523/JNEUROSCI.0996-12.2012.
43. Hart MP, Brettschneider J, Lee VM, Trojanowski JQ, Gitler AD (2012) Distinct TDP-43 pathology in ALS patients with ataxin 2 intermediate-length polyQ expansions. Acta Neuropathol. doi:10.1007/s00401-012-0985-5.
44. Neumann M, Kwong LK, Lee EB, Kremmer E, Flatley A et al. (2009) Phosphorylation of S409/410 of TDP-43 is a consistent feature in all sporadic and familial forms of TDP-43 proteinopathies. Acta Neuropathol 117: 137–149.
45. Brooks BR, Miller RG, Swash M, Munsat TL, for the World Federation of Neurology Research Group on Motor Neuron Diseases (2000) El Escorial revisited: revised criteria for the diagnosis of amyotrophic lateral sclerosis. Amyotroph Lateral Scler Other Motor Neuron Disord 1: 293–299.
46. Lattante S, Conte A, Zollino M, Luigetti M, Del Grande A, et al. (2012) Contribution of major amyotrophic lateral sclerosis genes to the etiology of sporadic disease. Neurology 79: 66–72.
47. Lee T, Li YR, Chesi A, Hart MP, Ramos D, et al. (2011) Evaluating the prevalence of polyglutamine repeat expansions in amyotrophic lateral sclerosis. Neurology 76: 2062–2065.
48. Liu X, Lu M, Tang L, Zhang N, Chui D, et al. (2012) ATXN2 CAG repeat expansions increase the risk for Chinese patients with amyotrophic lateral sclerosis. Neurobiology of Aging.dx.doi.org/10.1016/j.neurobiolaging.2013.04.009.
49. Conforti FL, Spataro R, Sproviero W, Cavalcanti F, Condino F, et al. (2012) Ataxin-1 and ataxin-2 intermediate-length PolyQ expansions in amyotrophic lateral Sclerosis. Neurology. doi:10.1212/WNL.0b013e318278b618.

DISCUSIÓN

III.2.3

3.2.3.1: Alelos largos y prevalencia de la SCA2: lecciones del estudio nacional y análisis de la literatura

Publicado en: Clin Gen, (2013) doi10.1111/cge.12221. (Factor de Impacto: 4.25)

La poderosa observación de la cual se habla en el resumen del acápite 3.2.1permitía sostener en parte el postulado de Takano *et al*, pero al no existir una relación matemática comprobada nuestras conclusiones no eran del todo rigurosas.

Por otra parte, el esquema que se propone al final del artículo 3 (Fig.6 Articulo 3 del acápite 3.2.2), sirve también para interpretar las expansiones *de novo* para SCA2.

Este artículo es una revisión sistemática, no meta-análisis, que retoma los hallazgos del artículo 2 y 3 para demostrar la relación entre alelos largos y frecuencia de SCA2 así como la influencia del haplotipo dando cumplimiento a los objetivos 1 y 2 de la tesis. Aquí se demuestra que la frecuencia relativa de SCA2 es explicada en un 84% por la frecuencia de alelos largos ($p<0.000$) (Fig.1c). Además, se reporta una familia con inestabilidad de la expansión de CAG desde 31-33CAG, aconteciendo bajo el haplotipo CC-4 (Fig.1d), el cual es universal para SCA2. Este evento mutagénico sigue un modelo *paso a paso*, a diferencia de lo observado para los casos ELA que es saltatorio.

Se discuten una serie de errores que no se han tenido en cuenta por otros autores y lo cual pudo distorsionar la percepción acerca de la relación entre alelos largos y la frecuencia de SCA2.

Por lo que esta "carta al editor" es la apología, consolidación, complementación y generalización del postulado, así como la extensión del modelo mutagénico de pérdidas de interrupciones CAA para alcanzar umbrales anómalos de la expansión de CAG en SCA2.

Este conocimiento es aplicable a la práctica de la genética clínica, dado que no está claro si estas variantes se asocian con riesgo para la descendencia, esto se discute en el artículo posterior a este.

CLINICAL GENETICS An International Journal of Genetics, Molecular and Personalized Medicine

Clin Genet 2013
Printed in Singapore. All rights reserved

CLINICAL GENETICS
doi: 10.1111/cge.12221

Letter to the Editor

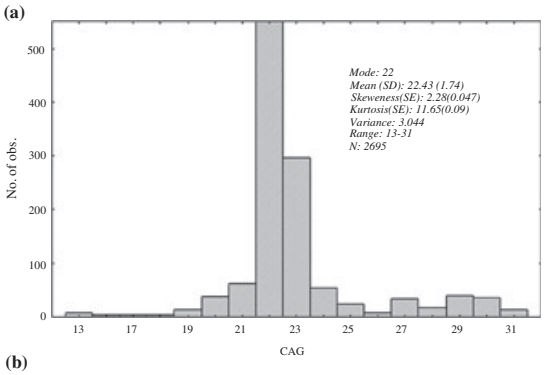
Large normal alleles and SCA2 prevalence: lessons from a nationwide study and analysis of the literature

To the Editor:

The relationship between high frequency of large normal alleles (LNA) and high prevalence of autosomal dominant cerebellar ataxias (ADCAs) is not completely understood. Inconsistent reports regarding this relationship have been published, some of them postulating strong correlation between LNA and expanded alleles (EA) of the dominant spinocerebellar ataxias (SCAs) (1). Here, we provide epidemiological and genetic data supporting the hypothesis of large normals alleles (LNA) being a reservoir of EA in *SCA2(ATXN2)*. We also discuss the reasons of conflicting data in the literature.

We developed nationwide genetic ataxia screening (2), and was found that the highest prevalence of SCA2 worldwide relates to the highest frequency of 23-31CAG repeats in Cuban population (Fig. 1a,b). Pre-mutated alleles in SCA2 families displayed meiotic and mitotic instability (3). Analysis of the epidemiological data revealed a linear correlation ($r^2 = 0.8408$, $p < 0.000$) between LNA-SCA2 frequencies (Fig. 1c). We also present a family with a progressive CAG elongation (from *31-to-33CAG*), linked to the ancestral C-C-SNP-haplotype (Fig. 1d), supporting the original postulate (1).

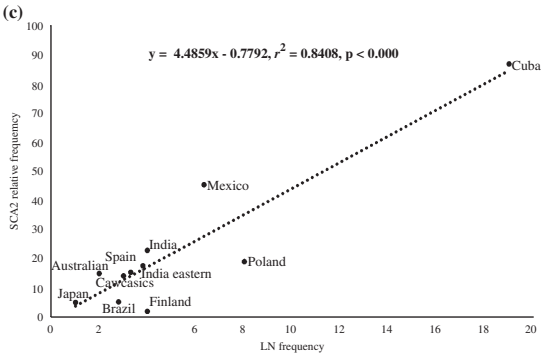
The hypothesis outlined by Takano et al. (1) was assessed in 11 populations. In summary, the LNA–EA relationship has been confirmed in the majority of populations and only one study involved a nationwide screening (Table 1). Therefore, inconsistencies between



(b)

Cut off values when CAG for Large AN =	>22	>23	>24	>25	>26	>27	>28	>29	>30	
Frequencies of SCA2 Large AN in different Populations (N)										Skewness
Cuba (2695)	0.19	0.08	0.06	0.05	0.05	0.04	0.03	0.02	0.01	2.28
Japanese[1](359)	0.01 (X^2=73.67, P=0.0000)	0.01 (X^2=23.61, P=0.0000)	0.00 (X^2=22.99, P=0.0000)	–	–	–	–	–	–	-0.97
Caucasians[1](355)	0.12 (X^2=11.15, P=0.0000)	0.03 (X^2=11.86, P=0.0006)	0.00 (X^2=5.11, P=0.0238)	–	–	–	–	–	–	-4.81
North India[2](135)	0.06 (X^2=15.14, P=0.0002)	0.02 (X^2=6.41, P=0.0114)	0.02 (X^2=4.90, P=0.0268)	–	–	–	–	–	–	nd
Eastern India[3](448)	0.038 (X^2=65.15, P=0.0000)	–	–	–	–	–	–	–	–	-0.13
Polish[4](234)	0.08 (X^2=19.33, P=0.0000)	0.02 (X^2=11.30, P=0.0000)	0.02 (X^2=8.06, P=0.0045)	¶	¶	0.004 (X^2=7.34, P=0.0068)	¶	–	–	nd
Finnish[5](501)	0.07 (X^2=44.65, P=0.0000)	0.04 (X^2=11.03, P=0.0009)	0.03 (X^2=8.36, P=0.0038)	–	–	–	–	–	–	4.47

1: Takano et al. 1998, 2: Salem et al. 2000, 3: Basu et al. 200, 4: Sobczack & Krzyzosiak., 2004, 5, Juvonen et al., 2005. nd: *not documented*. ¶: This population don't shows alleles sized with 25, 26 and 28CAG.



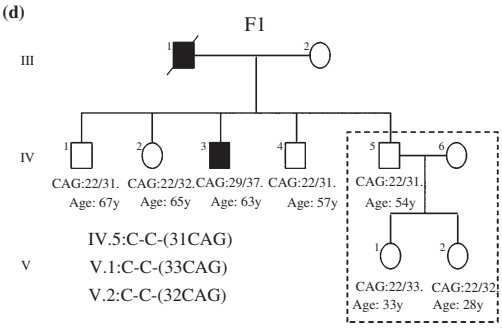


Fig. 1. (**a**) CAG repeats size distribution at the SCA2 locus in the Cuban population. The distribution is shifted toward ANs according to our previous data (3). (**b**) Comparison of large ANs frequency of in Cuban *vs* other populations. (**c**) Correlation between large normal (LN) frequency and relative SCA2 frequencies in different populations. LN and SCA2 frequencies were collected from reports in Table 1. (**d**) Genealogy and genotype of the SCA2 family showing expansion from a LN to abnormal SCA2 repeat expansion. SNP haplotypes and CAG repeat sizes were revealed by sequencing using FP3-RP3 primers. Intragenic SNPs rs695871 (G/C) and rs695872 (T/C) are situated at positions 106 and 177 nucleotides upstream of the CAG repeat. CAG for other cases were obtained by fragment analysis as reported in (3).

Table 1. Positions respecting Takano et al. postulate in SCA2 locus

Population	LNA-SCA2	Nationwide study	References
Caucasian	S	N	Takano et al. (1)
Japanese	S	N	Takano et al. (1)
Spanish	S	N	Pujana et al. (4)
Australian	U	N	Storey et al.(5)
Indian	S	N	Basu et al. (6) and Saleem et al. (7)
Brazilian	S	N	Silveira et al.(8)
Portuguese	S	N	Silveira et al.(8)
Polish	D/S	N	Sobczak and Krzyzosiak (9)
Finnish	D	N	Juvonen et al. (10)
Cuban	S	Y	Laffita-Mesa et al. (3)
Mexican	S	N	Magaña et al. (11)

D, discordant; LNA, large normal alleles; N, not; S, supportive; U, unknown; Y, yes.

studies could be the consequence of restricted representation of normal/pathological CAG-repeat spectrum. Collection of control samples is simpler than obtaining samples from ataxics and differential distribution may result from selection bias related to unequal access to specialized medical services.

Polish and Finnish studies provided discrepant results. Poland had high prevalence of LNA while SCA2 being recently found as the second most common SCA subtype (9, 12), supporting the correlation of LNA-SCA2 prevalence.

Finnish population have different occurrence of genetic diseases (10), explaining the lack of controls and *ATXN2*-pathological alleles relationship. No excess of LNA in Indian controls compared with Finns was observed, but would not be conclusive. Note, that nine Indian populations were not enriched with LNA as only in Brahmin and Mahishya ethnicities, >27 CAG has been found. India is a *multiethnic* (~200) country of one billion peopleí, and genetic diseases are clustered in ethnicities. Phenotypically similar families to SCA2 have been reported but without genetic localization, thus it is not clear whether SCA2 is the most frequent SCA in India, unless nationwide screening is undertaken (6, 13).

The clinical spectrum of SCA2 is the broadest among all SCAs, and may present as cerebellar, parkinsonism, motoneuron, psychiatric phenotypes or infantile encephalopathy with genetic overlap. Moreover, differences in age-of-onset and non-Mendelian inheritance represent serious diagnostic challenges for clinicians. These cases are often gathered not for the CAG expansion, modifying the real picture regarding mutation incidence. Penetrance is uncertain between 32-35CAG-repeats. The carriers of such expansions usually have a very late disease onset resulting in limited accessibility, correct clinical assessment and hence classification within the SCA subtypes. There are inconsistencies in the consensus thresholds for LNA/intermediate and pathological CAG expansions. In fragment analysis, the (CCG/CCG)*n* polymorphic sequence repeats adjacent to CAG is counted as part of the repeat expansion. All the above-mentioned issues contribute to discrepancies between predisposed alleles and the disease incidence, in SCA2 as well as in other SCAs.

What should be therefore kept in mind? In the populations with highest SCA2 incidence, the highest LNA frequency is found. *De novo* pathogenic CAG expansions are more likely to occur from LNA. Their presence represents an increased intergenerational risk (>0%) for disease development which should be communicated with the carriers. Wide clinical expressivity, penetrance, pleiotropism of *ATXN2* mutation, and technical errors contribute to '*discordant results*' concerning the hypothesis on LNA being reservoirs for EA in SCA2 and conditions connected with *ATXN2*-CAG expansions.

Acknowledgements

We would like to thank our SCA2 patients and families who have been contributing to the genetic and clinical studies for more than 30 years. We deeply acknowledge to Dr Christopher Gomez for language correction. This work was supported by Third World Academy of Sciences – TWAS – grant from the Project 06–329 RG/BIO/LA UNESCO FR: 3240157855 and a grant from the Cuban Ministry of Health and the Ministry of Science and Technology all to JMLM.

JM Laffita-Mesa[a,b,*]
LE Almaguer-Mederos[a,b]
V Kourí[c]
PO Bauer[d]
YV Mojena[a]
TC Mariño[a]
LV Pérez[a,b]

[a]Center for Research and Rehabilitation of Hereditary Ataxias - CIRAH, Holguín, Cuba
[b]Branch of Biomedicine of the Cuban Academy of Sciences - ACC, Havana, Cuba
[c]Virology Department, Institute of Tropical Medicine "Pedro Kourí", Havana, Cuba, and
[d]Department of Neuroscience, Mayo Clinic, Jacksonville, FL, USA

References

1. Takano H, Cancel G, Ikeuchi T et al. Close associations between prevalences of dominantly inherited spinocerebellar ataxias with CAG-repeat expansions and frequencies of large normal CAG alleles in Japanese and Caucasian populations. Am J Hum Genet 1998: 63 (4): 1060–1066.
2. Velázquez L, Sánchez G, Santos N et al. Molecular epidemiology of spinocerebellar ataxias in Cuba: insights into SCA2 founder effect in Holguin. Neurosci Lett 2009: 454: 157–160.
3. Laffita-Mesa JM, Velázquez LC, Santos N et al. Unexpanded and intermediate CAG polymorphisms at the SCA2 locus (ATXN2) in the Cuban population: evidence about the origin of expanded SCA2 alleles. Eur J Hum Genet 2012: 20: 41–49.
4. Pujana MA, Corral J, Gratacos M et al. Spinocerebellar ataxias in Spanish patients: genetic analysis of familial and sporadic cases. Hum Genet 1999: 104: 516–522.
5. Storey E, du Sart D, Shaw JH et al. Frequency of spinocerebellar ataxia types 1, 2, 3, 6, and 7 in Australian patients with spinocerebellar ataxia. Am J Med Genet 2000: 95: 351–357.
6. Basu P, Chattopadhyay B, Gangopadhaya PK et al. Analysis of CAG repeats in SCA1,SCA2,SCA3,SCA6,SCA7 and DRPLA loci in spinocerebellar ataxia patients and distribution of CAG repeats at the SCA1,SCA2 and SCA6 loci in nine ethnic populations of eastern India. Hum Genet 2000: 106 (6): 597–604.
7. Saleem Q, Choudhry S, Mukerji M et al. Molecular analysis of autosomal dominant hereditary ataxias in the Indian population: high frequency of SCA2 and evidence for a common founder mutation. Hum Genet 2000: 106 (2): 179–187.
8. Silveira I, Miranda C, Guimaraes L et al. Trinucleotide repeats in 202 families with ataxia: a small expanded (CAG)n allele at the SCA17 locus. Arch Neurol 2002: 59: 623–629.
9. Sobczak K, Krzyzosiak WJ. Patterns of CAG repeat interruptions in SCA1 and SCA2 genes in relation to repeat instability. Hum Mutat 2004: 24 (3): 236–247.
10. Juvonen V, Hietala M, Kairisto V et al. The occurrence of dominant spinocerebellar ataxias among 251 Finnish ataxia patients and the role of predisposing large normal alleles in a genetically isolated population. Acta Neurol Scand 2005: 111 (3): 154–162.
11. Magaña JJ, Tapia-Guerrero YS, Velázquez L et al. Analysis of CAG repeats in five SCA loci in Mexican population: epidemiological Evidence of a SCA7 founder effect. Clin Genet 2013. DOI: 10.1111/cge.12114.
12. Sułek-Piatkowska A, Zdzienicka E, Raczyńska-Rakowicz M et al. The occurrence of spinocerebellar ataxias caused by dynamic mutations in Polish patients. J Neurol Neurochir Pol 2010: 44 (3): 238–245.
13. Chakravarty A, Mukherjee SC. Autosomal dominant cerebellar ataxias in ethnic Bengalees in West Bengal an-Eastern Indian state. Acta Neurol Scand 2002: 105: 202–208.

Correspondence:
José Miguel Laffita-Mesa, PhD
Calixto García 152b
Antilla, Holguín
Cuba
Tel.: +53 24 461564
Fax: +53 24 463579
e-mail: laffita@ataxia.hlg.sld.cu

3.2.3.2: Alelos largos normales e intermedios en el contexto del diagnóstico prenatal de SCA2

Publicado en: J Genet Counsel, (2013) doi10.1007/s10897-013-9615-1. (Factor de Impacto: 1.77)

El Centro para la Investigación y Rehabilitación de las Ataxias Hereditarias, ubicado en la Provincia de Holguín desarrolla desde el año 2001 el programa de diagnóstico predictivo de la SCA2. Esta es una experiencia única a nivel global, siendo una de las aplicaciones directas en la genética clínica, la identificación de la mutación SCA2. Sin embargo, nunca se consideraron a los alelos largos, y las variantes intermedias en este programa, en parte por la creencia de que estas variantes no tenían riesgo asociado (Pulst *et al*., 1996, Sequeiros *et al*., 2010).

En base a los hallazgos obtenidos en estudios anteriores, es que se decide realizar este trabajo descriptivo. El objetivo fue: *valorar las implicaciones de los alelos largos (23-31CAG) y los intermedios (32-34CAG) en el contexto del programa cubano de diagnóstico prenatal de la SCA2*.

Los antecedentes fueron la presencia de un elevado número de alelos largos en Cuba, encontrados en la población sana que no presentaba SCA2, muchos de ellos eran portadores de un genotipo con dos alelos largos (Art.2).

Por tanto, este trabajo es la aplicación en la genética clínica del conocimiento relacionado con el mecanismo mutagénico de la expansión de CAG *ATXN2 y* el haplotipo asociado, y más allá es el impacto positivo de conocer cómo se origina la mutación SCA2.

Aquí se presentan tres escenarios reales, los cuales involucraron: 1) expansiones, 2) contracciones, 3) estabilidad en la trasmisión de la expansión de CAG de una generación a la otra.

De este trabajo se derivan las siguientes conclusiones:

1. *El riesgo mendeliano asociado a estas variantes es >0%, habiendo parejas con genotipo (AL/AL), en las que este riesgo es >50 %.*

2. *El hecho de ser portador de una variante larga o intermedia no significa necesariamente que habrá expansiones. Pero existe un riesgo, que debe ser conocido y discutido con los portadores.*

3. *Aunque podemos predecir cuál cromosoma pudiera estar predispuesto, cuándo ocurrirá la expansión de CAG es mucho más estocástico.*

La complejidad de estas situaciones sugiere que el diagnóstico predictivo, no solo debe estar basado en la determinación de la longitud del CAG, sino que a la luz de estos hallazgos la información del haplotipo es útil para predecir la proclividad o riesgo a expandirse en sucesivas generaciones de una variante larga o intermedia. También es útil para asesorar a pacientes con ELA, con expansiones anómalas de CAG en este gen, acerca del pronóstico de la enfermedad.

J Genet Counsel
DOI 10.1007/s10897-013-9615-1

ORIGINAL RESEARCH

Large Normal and Intermediate Alleles in the Context of SCA2 Prenatal Diagnosis

Tania Cruz-Mariño · Jose Miguel Laffita-Mesa · Yanetza Gonzalez-Zaldivar · Miguel Velazquez-Santos · Raul Aguilera-Rodriguez · Annelie Estupinan-Rodriguez · Yaime Vazquez-Mojena · Patrick Macleod · Milena Paneque · Luis Velazquez-Perez

Received: 6 October 2012 / Accepted: 22 May 2013

Abstract In 2001 a program for predictive testing of Spinocerebellar Ataxia type 2 was developed in Cuba, based on the detection of an abnormal CAG trinucleotide repeat expansion in the *ATXN2* gene. A descriptive study was designed to assess the implications of *ATXN2* large normal and intermediate alleles in the context of the SCA2 Prenatal Diagnosis Program. Four clinical scenarios were selected based upon the behaviour of large normal and intermediate alleles when passing from one generation to the next, showing expansions, contractions, or stability in the CAG repeat size. In some populations, traditional Mendelian risk figures of 0 % or 50 % may not be applicable due to the high frequency of unstable large normal alleles. Couples with no family history of SCA2 may have a >0 % risk of having an affected offspring. Similarly, couples in which there is both an expanded and a large normal allele may have a recurrence risk >50 %. It is imperative that these issues be addressed with these couples during genetic counseling. These recurrence risks have to be carefully estimated in the presence of such alleles (particularly alleles ≥27 CAG repeats), carriers need to be aware of the potential risk for their descendants, and programs for prenatal diagnosis must be available for them.

Keywords Hereditary ataxias · Intermediate alleles · Large normal alleles · Predictive testing · Prenatal diagnosis · Spinocerebellar ataxia type 2

Tania Cruz-Mariño and Jose Miguel Laffita-Mesa has equal contribution.

T. Cruz-Mariño · M. Velazquez-Santos
Predictive Genetics Department, Center for the Research and Rehabilitation of Hereditary Ataxias, Reparto Edecio Pérez, Carretera Vía Habana, Holguín, Cuba 80100

J. M. Laffita-Mesa · Y. Gonzalez-Zaldivar · Y. Vazquez-Mojena
Neurobiology Department, Center for the Research and Rehabilitation of Hereditary Ataxias, Reparto Edecio Pérez, Carretera Vía Habana, Holguín, Cuba 80100

R. Aguilera-Rodriguez · A. Estupinan-Rodriguez
Clinical Department, Center for the Research and Rehabilitation of Hereditary Ataxias, Reparto Edecio Pérez, Carretera Vía Habana, Holguín, Cuba 80100

P. Macleod
Centre for Biomedical Research, University of Victoria, Victoria, BC, Canada

M. Paneque
Centre for Predictive and Preventive Genetics (CGPP), IBMC, University of Porto, Porto, Portugal

L. Velazquez-Perez
Neurophysiology Department, Center for the Research and Rehabilitation of Hereditary Ataxias, Reparto Edecio Pérez, Carretera Vía Habana, Holguín, Cuba 80100

T. Cruz-Mariño (✉)
Calle 4 de Abril. Número 6. Entre Final y Carretera de Gibara, Holguín, Cuba 80100
e-mail: taniacruz@hotmail.ca

Published online: 28 June 2013

Introduction

Spinocerebellar ataxias (SCAs) are a group of more than 30 autosomal dominant neurodegenerative disorders primarily affecting gait and movement coordination.

The degenerative process affects the cerebellum and its afferent and efferent connections, and may additionally involve the ponto-medullar systems, pyramidal tracts, basal ganglia, cerebral cortex and peripheral nerves. The molecular mechanism for a group of these diseases consists of a pathological CAG trinucleotide repeat expansion in the coding region of the mutated genes; such is the case of SCA1, SCA2, SCA3/MJD, SCA6, SCA7, SCA17 and Dentatorubral-pallidoluysian atrophy (DRPLA) (Matilla et al. 2006; Fam et al. 2012; Velázquez-Pérez et al. 2012).

The Holguín province in Cuba has the world's highest prevalence of SCA2 with 753 affected individuals belonging to 200 unrelated families. In addition, there are 7 000 at risk individuals living in this area (Velázquez-Pérez et al. 2009; Velázquez-Pérez et al. 2011).

With an autosomal dominant mode of inheritance, SCA2 is characterized by ataxic gait, cerebellar dysarthria, dysmethria, dysdiadochokinesia, and an age at onset that in Cuba ranges from 3 to 79 years. Variation in the CAG repeat number is the reason for this wide spectrum.

Anticipation, or an increase in CAG repeat number upon transmission, is estimated to occur in 80 % of transmissions of abnormal *ATXN2* CAG repeats (Velázquez-Pérez et al. 2009). There is a significant negative correlation between CAG repeat size of the *ATXN2* gene and age at onset, the former accounting for 80 % of the variability; and frequent intergenerational expansions contrast with infrequent contractions of the CAG repeat (Velázquez-Pérez et al. 2009). Individuals from the affected families should be therefore recipients of systematic clinical and psychological support, as well as genetic counseling.

Slowing of horizontal saccades at 60° of target displacement is present as the earliest subclinical sign approximately 15 years before the onset of ataxia, followed by the reduction of REM sleep percentage with decreased rapid eye movements' density, decrease of sensory amplitudes, increased P40 latency, motor performance deficits demonstrated through the prism adaptation task and reduced capabilities to identify odors in a smell identification test (Velázquez-Pérez et al. 2012).

Cuba started a national program for predictive testing of SCA2 in 2001 (both presymptomatic testing and prenatal diagnosis) (Paneque et al. 2007) which has been developed at the Centre for the Research and Rehabilitation of Hereditary Ataxias (CIRAH). DNA-testing is based on the detection of an abnormal CAG trinucleotide repeat expansion in the *ATXN2* gene through fragment analysis (Laffita-Mesa et al. 2012).

Alleles up to 31 CAG repeats are normal (NA); those with ≥23 to 31 repeats are considered large normal (LNA). Alleles with 32 repeats or larger are expanded (EA); those with 32 to 34 CAG are considered intermediate alleles (IAs: they fall in the gray zone for penetrance and may have evolved from large normal alleles representing intermediates that in successive generations would give rise to fully penetrant alleles), while alleles from 35 to 500 CAG are fully penetrant (Sanpei et al. 1996; Imbert et al. 1996; Pulst et al. 1996; Choudhry et al. 2001; Sequeiros et al. 2010; Laffita-Mesa et al. 2012) (Table 1).

The observation that the genotype of some SCA2 Cuban patients consisted of an expanded allele together with a large normal allele (EA/LNA) led to research documenting the frequency of large normal alleles and the relationship between the large normal alleles and the high frequency of SCA2 in this population (Laffita-Mesa et al. 2012). At the same time, some individuals who consented to be analysed as healthy controls, had a genotype with both a normal allele and a large normal allele (NA/LNA).

In addition to having the highest worldwide prevalence of SCA2, large normal alleles and intermediate alleles are also very frequent in Cuba (Laffita-Mesa et al. 2012). Alleles with 23 and 24 repeats have frequencies of 0.08 and 0.06 respectively, which are significatively higher than those described in other populations like Japanese (0.01-0.00) (Takano et al. 1998), Caucasians (0.03-0.00) (Takano et al. 1998), North Indians (0.02-0.02) (Saleem et al. 2000), Polish (0.02-0.02) (Sobczak and Krzyzosiak 2004), and Finnish (0.04-0.03) (Juvonen et al. 2005). Besides, Japanese, Caucasians, Indians, and Finnish are populations lacking alleles larger than 25 CAG triplets contrasting with Cubans, who show a continuous distribution of alleles from 25 to 30 repeats (Takano et al. 1998; Saleem et al. 2000; Juvonen et al. 2005; Laffita-Mesa et al. 2012).

Though the CAG repeat meiotic instability of expanded alleles of the *ATXN2* gene has been well documented (Saleem et al. 2000), the large normal alleles and intermediate alleles have received less attention, bringing about complex scenarios like the ones presented here.

Table 1 Reference ranges for the CAG repeats at the *ATXN2* gene

CAG repeats	Allele type
≤31	Normal (NA)
23 to 31	Large normal (LNA)
≥32	Expanded (EA)
32 to 34	Intermediate (IA)
35 to 500	Fully penetrant

Methods

A descriptive study was designed to assess the implications of *ATXN2* large normal alleles (23 to 31 CAG repeats) and intermediate alleles (32 to 34 CAG repeats) in the context of the Cuban program for the SCA2 Prenatal Diagnosis (PND). The protocol for the Cuban predictive testing program has been published elsewhere (Paneque et al. 2007).

In brief, access to PND was given to couples with one member carrying the *ATXN2* expanded allele, having expressed their intention to terminate the pregnancy if the fetus was diagnosed as an expanded allele carrier (which is sustained in the aim to protect the autonomy of the unborn child).

All participants were informed about the predictive testing procedures and protocol, as well as the possibility of using information from the clinical records in clinical research and they gave their separate written consent for both. The study was approved by the Institutional Ethics Committee.

In the years 2001–2011, 58 couples requested their inclusion in the SCA2 PND. Their medical records were retrospectively analyzed. Twenty eight fetuses were carriers of an expanded allele and three of them had an intermediate allele genotype (22/32, 22/32 and 22/34). Among the non-carrier fetuses, eleven large normal alleles were identified with genotypes 22/23 (four cases), 22/27 (two cases), 22/30 (three cases) and 22/31 (two cases).

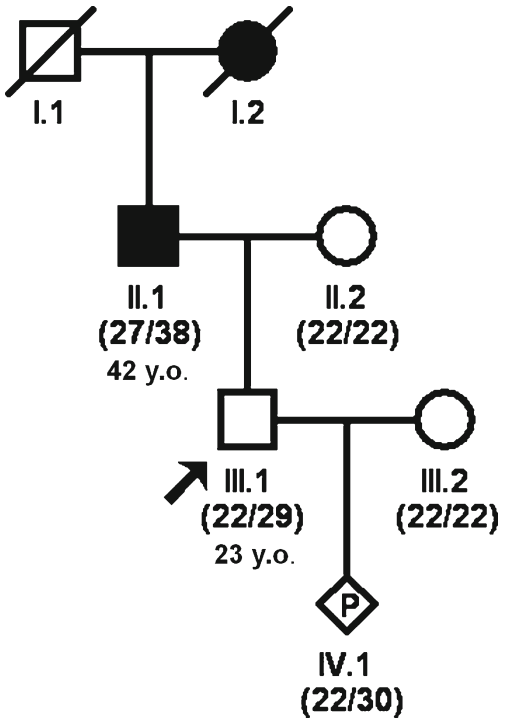
The selection of cases presented here were based upon the behaviour of the *ATXN2* gene large normal alleles and intermediate alleles when passed from one generation to the next, showing 1) expansions, 2) contractions, or 3) stability in the CAG repeat size.

Results and Discussion

Through his inclusion in the presymptomatic testing for SCA2, individual III.1 (Case 1) was identified as a carrier of both a normal allele and a large normal allele (genotype 22/29). Because his father is affected, a CAG repeat contraction seemed to be a good rationale for the presence of the large normal allele. Nonetheless, tracing his family history, it was revealed that his father (II.1) presented both a large normal allele and a fully penetrant allele (genotype 27/38).

Once III.2 became pregnant, the dilemma of a probable meiotic instability was discussed with the couple, providing them with reproductive options, one of them being PND. The couple agreed to a prenatal test and a normal fetus (IV.1) was diagnosed. The fetus carried a large normal allele with an additional CAG repeat (genotype 22/30). The expansions have taken place in the paternal transmissions.

In SCA2 Cuban families paternal expansions are more frequent than contractions (Velázquez-Pérez et al. 2009). Therefore, the participant (III.1) probably did not inherit the large normal allele via a contraction of the paternal expanded allele, but from the paternal large normal allele instead, gaining two CAG repeats −27 to 29 CAG.



Case 1 Family history of case 1. Squares = men; circles = women; solid circle and solid square = affected; slash marks = deceased; rhomb = sex unknown; P = pregnancy; arrow = consultand; y.o = years old

Although DNA samples from I.1 are not available, it is possible to infer that he was a carrier of a large normal allele (27 CAG repeats or fewer), based on the fact that he was healthy and married to I.2, a clinically affected woman who represents the most probable source of the fully penetrant allele inherited by II.1.

As in India (Saleem et al. 2000), there exists a Cuban pool of "ancestral" or "at risk" haplotypes. Not only are families with clinically diagnosed SCA2 at risk, but also families harboring the large normal allele at-risk haplotypes. It is a well-known fact that the occurrence of expanded alleles in Huntington disease, can come from unaffected relatives who carry large normal alleles predisposing haplotypes (Warby et al. 2009).

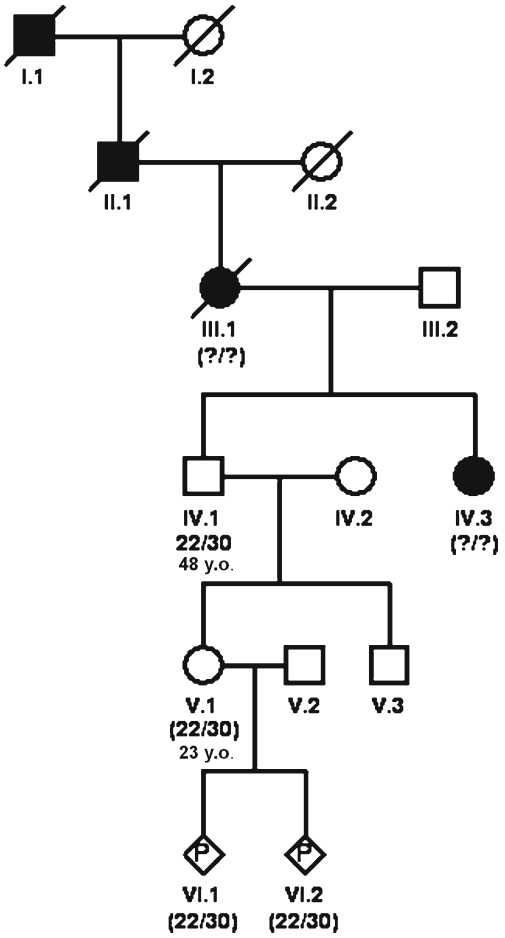
This distribution of large normal alleles mirrors the relative frequency of SCA2 in populations. Caucasians show an abundance, in contrast to the Japanese who show a paucity of alleles between 23 and 31 CAG (Takano et al. 1998). The lack of gaps between the normal (13–31) and the pathological threshold (32–79) in SCA2 suggest that large normal alleles are a reservoir for new expanded alleles in our population.

As with expanded alleles, large normal alleles of 27–31 CAG repeats are somatically unstable; they have a different molecular nature and behaviour resembling intermediate

alleles (32–34 CAG). Meiotic instability is more likely to occur in these 27–31 CAG repeats than in shorter large normal alleles (23–26 CAG), pointing to 27 CAG repeats as the threshold for a higher instability risk (Laffita-Mesa et al. 2012). This is why our group considers that PND must be available not only for carriers of expanded alleles but for those persons carrying large normal alleles ≥27 CAG repeats.

In addition large normal alleles ≥27 CAG repeats and intermediate alleles may also be a common genetic risk factor for Amyotrophic Lateral Sclerosis (ALS) (Elden et al. 2010). This is another issue adding complexity to the genetic counseling situation presented here.

Individual V.1 (Case 2) had participated in the presymptomatic testing for SCA2. Her test results revealed the presence of a large normal allele (genotype 22/30) and she requested PND. The molecular study showed that the normal fetus inherited the large normal allele with the same repeat size as her mother (genotype 22/30). Five years later she became pregnant again, requesting the PND one more time and receiving the same result (genotype 22/30).



Case 2 Family history of case 2. Squares = men; circles = women; solid circle and solid square = affected; slash marks = deceased; rhomb = sex unknown; P = pregnancy; y.o = years old

The large normal allele with 30 CAG repeats has remained stable in generations IV, V, and VI, independently of the sex of the transmitting parent.

The information concerning individuals III.1 and IV.3 genotype is not available. The clinical record of individual IV.3 shows she had an early age at onset (15 years), suggesting that she inherited a fully penetrant expanded allele from her affected mother (III.1).

It is interesting that there are another three Cuban families where alleles with 30 or 31 repeats remain stable from one generation to the next (unpublished observation). This fact supports the notion that carrying a large normal allele is not a sufficient reason for the transmission of instability of large normal alleles, leading to a particular at risk haplotype.

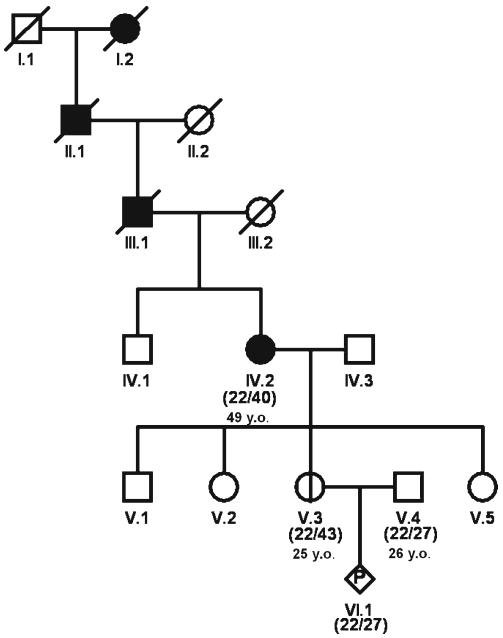
The clinical manifestations of SCA2 in the mother of individual IV.1, together with the early age at onset on his sister (IV.3), raise some questions about the origin of the large normal allele present in IV.1: Did it arise through a contraction of the maternal expanded allele? Was it inherited from the father? Was the father then a carrier of a large normal allele coming from the general population?

Despite the origin of the large normal allele, it has remained stable in four members of the family along three generations. This behaviour, nevertheless, should not be interpreted as a guarantee of stability in subsequent generations until all mechanisms responsible for dynamic mutations can be completely clarified.

Cleary and Pearson (2003) reported that gender-, tissue-, developmental-, and locus-specific *cis*-elements in conjunction with *trans*-factors may facilitate instability through the processes of DNA replication, repair and/or recombination. In diverse tissues at different developmental times and at specific loci, repetitive elements display variable levels of instability, suggesting vastly different mechanisms may be responsible for repeat instability amongst the disease loci and between various tissues. Some examples of *cis*-elements are the sequence of the repeat units, the length and purity of the repeat tracts, the sequences flanking the repeat, as well as the surrounding epigenetic environment, including DNA methylation and chromatin structure.

Individual V.3 (Case 3) was enrolled in the presymptomatic testing and was diagnosed as a presymptomatic carrier (genotype 22/43). The expansion occurred in the maternal transmission (40 to 43).

Subsequently, the PND revealed a 22/27 fetal genotype (VI.1). Initially, it was not clear if the 27 repeats allele was paternally inherited, or if it arose through a contraction of the mother's 43-repeats allele. This was an important issue to resolve as these results theoretically could also represent maternal contamination of a fetal sample (that has the 27 repeats allele from the father) or an extremely large maternally inherited expansion that exceeds the detection limits of the assay.



Case 3 Family history of case 3. Squares = men; circles = women; solid circle and solid square = affected; vertical bar = presymptomatic carrier; slash marks = deceased; rhomb = sex unknown; P = pregnancy; y.o = years old

Both a normal individual from the general population and the fetus were found to be large normal alleles carriers. This case highlighted the need of analyzing both parental alleles especially in cases related to PND, not only due to the possibility of maternal contamination of the fetal sample as has been reported by Sequeiros et al. (2010), but also because it enhances the correct interpretation of fetal results.

Carriers of large normal alleles in the general population are often not aware of their genetic status because they have no family history of SCA2. In a population where large normal alleles are abundant, such as in this region of Cuba, it is actually possible for a couple to have an SCA2 risk exceeding 50 % if one member of the couple is affected and the other unknowingly harbors a large normal allele.

Explaining these facts may represent new ethical dilemmas, notably in those cases when an individual carries large normal alleles and decides to have children disregarding the prenatal test while trusting his/her "not at-risk status".

Semaka et al. (2006) refer to a small pilot study (n=6) on Huntington disease which suggested that new mutation intermediate alleles (NM:IA) and general population intermediate alleles (GP:IA) carriers segregate into two groups based on their understanding and predictive testing experience. GP:IA carriers did not have the initial expectation that an intermediate allele result was a possibility and thus experienced great shock when they received the result. They also experienced the most uncertainty about the implications of an intermediate allele result and had a higher level of incorrect knowledge regarding these implications. NM:IA carriers were aware that an intermediate allele result was a possibility and therefore they did not experience surprise when they received this result.

The fact that diagnosing a large normal allele coming from the general population has implications for the relatives of the large normal allele carrier too, also must be taken into account; genetic counseling and presymptomatic testing may be expanded to reach them.

Cases 4 and 5 (individuals V.2 and V.3 respectively) belong to one family where intermediate alleles have been identified (Cases 4 and 5).

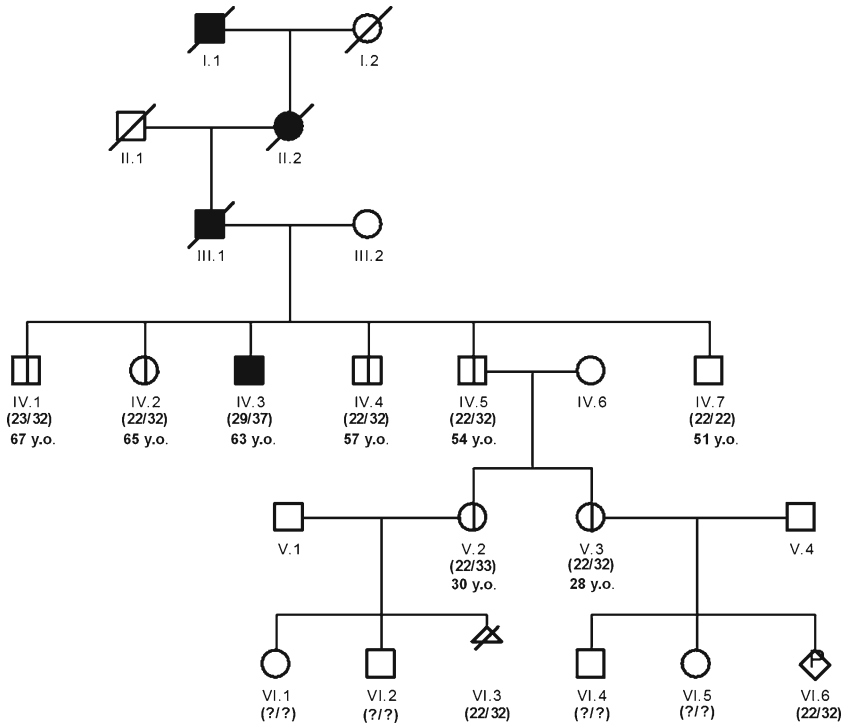
Individual III.1 died affected with SCA2 which began in his 60's (genotype 22/32). The four individuals in generation IV who inherited a $(CAG)_{32}$ repeats intermediate allele are not clinically affected, while the other individual (IV.3) is symptomatic and presents an expanded allele with $(CAG)_{37}$ repeats; again, the expansion took place in the paternal transmission, together with a large normal allele of $(CAG)_{29}$ repeats. This is another case where a large normal allele came from the mother, a member of the general population.

When individual V.2 requested to participate in the presymptomatic testing, she already had given birth to 2 children without PND, because it was not yet available. Once she was diagnosed as an expanded allele carrier (genotype 22/33), she requested the PND during her third pregnancy. The fetus inherited a $(CAG)_{32}$ repeats allele (genotype 22/32). Genetic counseling was offered in order to support the couple's decision making. Non-directiveness and respect for the autonomy of both members were observed during the process. The couple decided to terminate the pregnancy.

Two years later, her sister (V.3) was in a similar situation, having had two pregnancies without PND. Presymptomatic testing showed that she carried a 32 repeats allele and she decided to undergo PND for her third pregnancy. The fetus was shown to be also a carrier of a 32 repeats allele. This couple nevertheless decided to carry the pregnancy to term. They expressed that their decision was based mainly on the late age of onset in the affected grandfather.

So far, 25 individuals with intermediate alleles genotypes have been identified in the Cuban population (Laffita-Mesa et al. 2012). Only two intermediate alleles have been associated with age at disease onset at 34 and 45 years with genotypes of 22/32 CAG and 23/32 CAG, respectively (Laffita-Mesa et al. 2012; Santos et al. 1999). A very late onset (after 50's or 60's) had been previously reported for a few patients with intermediate alleles (Fernandez et al. 2000; Futamura et al. 1998).

It seems likely that the $(CAG)_{32}$ intermediate allele expanded to $(CAG)_{37}$ when passing from III.1 to IV.3, because in Cuban families intergenerational instability is produced by expansions in 89.03 % and by contractions in only 10.97 % of the offspring of affected patients (Velázquez-Pérez et al. 2009).



Cases 4 and 5 Family history of cases 4 and 5. Squares = men; circles = women; solid circle and solid square = affected; vertical bars = presymptomatic carriers; slash marks = deceased; rhomb = sex unknown; P = pregnancy; slash mark plus triangle = therapeutic abortion; y.o = years old

Alonso and colleagues (Alonso et al. 2007) reported three interesting SCA2 pedigrees. In the first family they documented three individuals carrying $(CAG)_{33}$ repeats alleles (intermediate alleles); the $(CAG)_{33}$ allele remains stable while passing from the affected mother to her yet asymptomatic son, but an unknown allele in her brother gives rise to a 37 repeats allele; subsequently a 40 repeats allele passes to the next generation. Up to this point, the allele has been expanding, but in generation V the allele contracts to 33 repeats. In the second family a $(CAG)_{43}$ allele (fully penetrant allele) contracts to $(CAG)_{33}$ (intermediate allele). In the third family it seems that the $(CAG)_{33}$ allele (intermediate allele) arose via a *de novo* mutation from a $(CAG)_{22}$ allele (normal allele).

Cases 4 and 5 represent two out of the three intermediate alleles already identified through SCA2 PND. We consider that the 32 repeats allele represents the expression of an unstable allele. Based on our current knowledge this allele should be considered as prone to instability, keeping in mind the possibility of contracting, remaining stable or even expanding in next generations.

Santos et al. (1999) reported a Cuban 45 year-old female (genotype 23/32) who inherited the $(CAG)_{32}$ allele via a contraction of a $(CAG)_{40}$ allele from her affected father, referring to the $(CAG)_{32}$ allele as the shortest CAG repeat reported for SCA2 that causes neurodegeneration. Interruptions of the CAG sequence by CAA triplets are considered related to the phenotypical expression of such alleles, because the 32-CAG repeat without CAA interruptions can also produce the disease phenotype (Santos et al. 1999).

The CAA interruptions may play an important role in conferring stability to *ATXN2* repeat and their absence predisposes alleles towards instability and pathological expansion. A minimal length of pure repeats is required to initiate instability, and the presence of interruptions break the repeat into smaller repeat tracts and thus protects it from instability by reducing the length of continuous uninterrupted repeats (Choudhry et al. 2001).

Individual V.2 is the only one in the family with a 33 repeats allele (genotype 22/33). Nevertheless the allele with 32 repeats has had a high stability in the last three generations, independent of the sex of the transmitting parent. This is why for this case, apart from a paternal expansion event, we may also consider one CAG as an acceptable margin of error, since it is the limit of accuracy of this testing.

Sequeiros et al. (2010) point out that size is critical for interpretation of intermediate alleles, and more importantly, of alleles at or near the boundaries of normal and expanded ranges.

Intrafamilial and interfamilial variability in age at onset is frequent in SCA2. That is why predictions relating to the age at onset based on the allele size and the family history are not a simple matter but an issue to be carefully discussed during genetic counseling sessions.

In the case of Huntington disease, Tibben (2007) explains that the demonstrated inverse association between age at onset and repeat length is not sufficient to be used clinically to predict age at onset in individual cases; individuals at risk look for something to hold on to, no matter how (un)realistic this is. An argument in favor of providing the repeat length is that given their need for information and the lack of treatment

options in the near future, there is nothing else to offer to individuals at risk other than greater control over their future (Tibben 2007).

The fact that large normal alleles and intermediate alleles are present in many SCA2 Cuban pedigrees, mainly coming from healthy individuals without a family history of SCA2, is relevant for clinical practice. It is related to the high prevalence of large normal alleles and intermediate alleles in the general population, which correlates with the particular high prevalence of SCA2 in Cuba (Laffita-Mesa et al. 2012).

The factors associated with an increased risk of repeat instability have to be considered when assessing the risk of expansion into the disease-associated range, and they must be discussed with carriers during genetic counseling. There are potential complications of developing risk estimates that account for all factors known to influence instability; haplotype analysis may offer additional clinical information when assessing the likelihood that an allele will undergo expansion into the disease range during transmission (Semaka et al. 2006).

Practice Implications of the Present Work

In some populations, traditional Mendelian risk figures of 0 % or 50 % may not be applicable due to the high frequency of unstable large normal alleles. Couples with no family history of SCA2 may, in fact, have a >0 % risk of having affected offspring. Similarly, couples in which there are both an expanded and a large normal allele may have a recurrence risk >50 %.

So far, the identification of large normal alleles carriers coming from the general population has been random; the absence of expectation about their genetic status predisposes them to shock, uncertainty and incorrect knowledge. The lack of awareness related to the risk for descendants hinders genetic counseling and PND as a reproductive option. At the same time, it is necessary to reach other family members, in order to provide them with genetic counseling and to offer them predictive testing options. Affected individuals among these families could have been misdiagnosed as sporadic or recessive cases.

The high prevalence of large normal alleles in the Cuban general population is the reason for the identification of EAs/LNAs genotypes in affected individuals, as well as NAs/LNAs genotypes in "healthy" persons. For large normal alleles, particularly those ≥27 CAG repeats which are near the boundaries of expanded alleles, the interpretation of their significance as a source for new expansions, and the difficulties generated by possible instability regarding recurrence risk estimation are issues to be discussed extensively during genetic counseling sessions. The analysis of both paternal and maternal DNA samples is mandatory for PND purposes.

The termination of pregnancy is always a complex decision for the couple to make, especially when a fetus is diagnosed as an intermediate allele carrier. The $(CAG)_{32}$ repeats allele is generally related to a very late age at onset; nevertheless, two cases with an earlier onset have been described. Precise predictions of age of onset are not accurate.

Our genetic counseling practice is challenged by the complexity of the clinical situations and the information that needs to be discussed with the couples. Provision of support during the decision making process is an important component of genetic counseling, particularly in this specific context. Once again, advanced counseling competences and ethical awareness are mandatory.

Limitations of the Study

The CAG repeat meiotic instability of large normal alleles and intermediate alleles of the *ATXN2* gene is a new chapter where many pages remain yet to be written. Attempts to identify an at-risk haplotype have been made, but further research is yet necessary to document the haplotype correlating with instability, and to understand the mechanisms underlying this event.

Despite the fact that large normal alleles with 27–31 CAG molecularly resemble intermediate alleles with 32–34 CAG, a closer look is needed to verify if these alleles are also prone to expansions instead of contractions, as it happens with fully penetrant alleles.

Both, the identification of the at-risk haplotype and the accurate estimation of the probability of instability are important aspects in the provision of genetic counseling.

We selected four clinical scenarios related to couples who underwent PND for SCA2 in Cuba; this is a very small sample if we consider the high frequency of large normal alleles and intermediate alleles in our population. Given the fact that these alleles may be risk factors for ALS, it is important to extend their study to a larger sample including the Cuban ALS population.

As we developed our study through a retrospective review of the medical records, information such as ages, age of death and CAG repeats sizes of some family members is missing.

The CAG trinucleotide size estimation has been performed through fragment analysis, with a margin of error of one CAG repeat rather than sequencing, a more accurate estimation of the repeat size.

Future Research Recommendations

A research project to identify the particular haplotype conferring susceptibility for CAG meiotic instability in SCA2 Cuban families is already being conducted. The population of Holguín province is at high risk for SCA2, and the identification of large normal alleles and intermediate alleles carriers through a voluntary population screening program could have a favourable impact, increasing the capability of early disease diagnosis and prevention.

Conclusions

ATXN2 gene large normal alleles and intermediate alleles are able to expand into the full penetrance range under certain conditions; therefore some cases previously diagnosed as "sporadic or recessive ataxias" mainly based on family history must be reviewed. Recurrence risks have to be carefully estimated in the presence of large normal alleles and intermediate alleles. Individuals carrying such alleles (particularly alleles ≥27 CAG repeats) need to be aware of the potential risk for their descendants and programs for PND must be available for them. The analysis of both paternal and maternal DNA samples before performing a PND, particularly in populations with a high frequency of large normal alleles and intermediate alleles, is both useful and necessary. These alleles represent a challenge for genetic counselors, especially in the context of PND.

Acknowledgments The authors are deeply indebted to patients for their valuable collaboration and to the Cuban Health Care Ministry. We are thankful to José Luis Guisao Martínez for the language corrections.

References

Alonso, E., Martínez-Ruano, L., De Biase, I., Mader, C., Ochoa, A., Yescas, P., et al. (2007). Distinct distribution of autosomal dominant Spinocerebellar Ataxia in the Mexican population. *Movement Disorders, 22*, 1050–1053.

Choudhry, S., Mukerji, M., Srivastava, A. K., Jain, S., & Brahmachari, S. K. (2001). CAG repeat instability at SCA2 locus: anchoring CAA interruptions and linked single nucleotide polymorphisms. *Human Molecular Genetics, 10*, 2437–2446.

Cleary, J. D., & Pearson, C. E. (2003). The contribution of *cis*-elements to disease-associated repeat instability: clinical and experimental evidence. *Cytogenetic and Genome Research, 100*, 25–55.

Elden, A. C., Kim, H. J., Hart, M. P., Chen-Plotkin, A. S., Johnson, B. S., Fang, X., et al. (2010). Ataxin-2 intermediate-length polyglutamine expansions are associated with increased risk for ALS. *Nature, 466*(7310), 1069–1075.

Fam, H.K., Chowdhury, M.K., Boerkoel, C.F. Spinocerebellar Ataxia with Axonal Neuropathy (SCAN1): A disorder of nuclear and mitochondrial DNA Repair. In J Gazulla (Ed.), Spinocerebellar Ataxia 2012. (pp 41–58), InTech.

Fernandez, M., McClain, M. E., Martinez, R. A., Snow, K., Lipe, H., Ravits, J., et al. (2000). Late-onset SCA2: 33 CAG repeats are sufficient to cause disease. *Neurology, 55*, 569–572.

Futamura, N., Matsumura, R., Fujimoto, Y., Horikawa, H., Suzumura, A., & Takayanagi, T. (1998). CAG repeat expansions in patients with sporadic cerebellar ataxia. *Acta Neurologica Scandinavica, 98*, 55–59.

Imbert, G., Saudou, F., Yvert, G., Devys, D., Trottier, Y., Garnier, J. M., et al. (1996). Cloning of the gene for spinocerebellar ataxia 2 reveals a locus with high sensitivity to expanded CAG/glutamine repeats. *Nature Genet, 14*, 285–291.

Juvonen, V., Hietala, M., Kairisto, V., & Savontaus, M. L. (2005). The occurrence of dominant spinocerebellar ataxias among 251 Finnish ataxia patients and the role of predisposing large normal alleles in a genetically isolated population. *Acta Neurologica Scandinavica, 111*, 154–162

Laffita-Mesa, J. M., Velázquez-Pérez, L. C., Santos-Falcon, N., Cruz-Mariño, T., González-Zaldívar, Y., Vázquez-Mojena, Y., et al. (2012). Unexpanded and intermediate CAG polymorphisms at the SCA2 locus (*ATXN2*) in the Cuban population: evidence about the origin of expanded SCA2 alleles. *European Journal of Human Genetics, 20*, 41–49.

Matilla, A., Goold, R., & Giunti, P. (2006). Molecular pathogenesis of Spinocerebellar ataxias. *Brain, 129*, 1357–1370.

Paneque, H. M., Prieto, A. L., Reynaldo, R. R., Cruz, M. T., Santos, F. N., Almaguer, M. L., et al. (2007). Psychological aspects of presymptomatic diagnosis of Spinocerebellar Ataxia Type 2 in Cuba. *Community Genetics, 10*, 132–139.

Pulst, S. M., Nechiporuk, A., Nechiporuk, T., Gispert, S., Chen, X. N., Lopes-Cendes, I., et al. (1996). Moderate expansion of a normally biallelic trinucleotide repeat in spinocerebellar ataxia type 2. *Nature Genet, 14*, 269–276.

Saleem, Q., Choudhry, S., Mukerji, M., Bashyam, L., Padma, M. V., Chakravarthy, A., et al. (2000). Molecular analysis of autosomal dominant hereditary ataxias in the Indian population: high frequency of SCA2 and evidence for a common founder mutation. *Human Genetics, 106*, 179–187.

Sanpei, K., Takano, H., Igarashi, S., Sato, T., Oyake, M., Sasaki, H., et al. (1996). Identification of the Spinocerebellar ataxia type 2 gene using a direct identification of repeat expansion and cloning technique, DIRECT. *Nature Genet, 14*, 277–284.

Santos, N., Aguiar, J., Fernández, J., Vázquez, M., Auburger, G., Gispert, S., et al. (1999). Molecular diagnosis of a sample of the Cuban population with spinocerebellar ataxia type 2. *Biotecnologia Aplicada, 16*, 219–221.

Semaka, A., Creighton, S., Warby, S., & Hayden, M. R. (2006). Predictive testing for Huntington disease: interpretation and significance of intermediate alleles. *Clinical Genetics, 70*, 283–294.

Sequeiros, J., Seneca, S., & Martindale, J. (2010). Consensus and controversias in best practices for molecular genetic testing of spinocerebellar ataxias. *European Journal of Human Genetics, 18*, 1188–1195.

Sobczak, K., & Krzyzosiak, W. J. (2004). Patterns of CAG repeat interruptions in SCA1 and SCA2 genes in relation to repeat instability. *Human Mutation, 24*, 236–247.

Takano, H., Cancel, G., Ikeuchi, T., Lorenzetti, D., Mawad, R., Stevanin, G., et al. (1998). Close associations between prevalences of dominantly inherited spinocerebellar ataxia with CAG-repeat expansions and frequencies of large normal CAG alleles in Japanese and Caucasian populations. *American Journal of Human Genetics, 63*, 1060–1066.

Tibben, A. (2007). Predictive testing for Huntington's disease. *Brain Research Bulletin, 72*, 165–171.

Velázquez-Pérez, L., Cruz, G. S., Santos Falcon, N., Enrique Almaguer Mederos, L., Escalona Batallan, K., Rodríguez Labrada, R., et al. (2009). Molecular epidemiology of spinocerebellar ataxias in Cuba: insights into SCA2 founder effect in Holguín. *Neuroscience Letters, 454*, 157–160.

Velázquez-Pérez, L., Rodriguez-Labrada, R., García-Rodríguez, J. C., Almaguer-Mederos, L. E., Cruz-Mariño, T., & Laffita-Mesa, J. M. (2011). *Cerebellum, 10*(2), 184–198.

Velázquez-Pérez, L., Rodriguez-Labrada, R., Freund, H.J., Auburger, G. Spinocerebellar Ataxia type 2. In J Gazulla (Ed.), Spinocerebellar Ataxia 2012. (pp 77–102), InTech.

Warby, S. C., Montpetit, A., Hayden, A. R., Carroll, J. B., Butland, S. L., Vissher, H., et al. (2009). CAG expansion in the Huntington disease gene is associated with a specific and targetable predisposing haplogroup. *American Journal of Human Genetics, 84*, 351–66.

METILACIÓN EPIGENETICA DEL ADN MODIFICADORA DEL FENOTIPO SCA2

III.3

3.3 METILACIÓN EPIGENETICA DEL ADN MODIFICADORA DEL FENOTIPO SCA2

3.3.1: Metilación epigenética del ADN en el promotor del gen ataxin-2: Nuevas implicaciones fisiológicas y patológicas.

Publicado en: Hum Gen, 131, (4), 625-638. (Factor de Impacto: 5.07)

En el primer artículo de esta tesis hipotetizamos de la "*existencia de factores modificadores del dueto: longitud del triplete CAG→fenotipo"*, para esto se han propuesto modificadores endógenos como exógenos, siendo la epigenética uno de estos, la que aún no ha sido evaluada. En este sentido, Bauer et al., 2003 propusieron que mediante la metilación *de novo* la célula silenciaba el alelo mutante. Estos autores realizaron experimentos, *in vitro*, con la droga demetilante 5-aza-desoxicitidina, en células derivadas de casos con genotipo similar y edades de inicio discordantes y encontraron diferencias en el nivel de expresión de *ATXN2*. Sin embargo, estos autores no detectaron metilación en el gen ATXN2 que validara sus observaciones.

En el presente artículo, dándole cumplimiento al objetivo # 5, se demuestra que existe metilación en la región promotora del gen *ATXN2*, por lo que el efecto observado por Bauer et al, pudo estar determinado precisamente por un silenciamiento directo sobre la variante mutada. Para poder demostrar la existencia de este fenómeno se desarrollaron dos técnicas moleculares de PCR (objetivo # 4), que permitían detectar metilación en dinucleótidos CpG de cualquier región del genoma. En el presente, en la región promotora del gen *ATXN2*, esto fue a su vez facilitado por la premisa de que este elemento está dentro de un gran islote de CpG.

La primera de ellas fue la *Methyl Specific PCR (MSP)*, que es un método cualitativo y fue utilizada por su sencillez para detectar el estado de metilación (generaba variables categóricas dicotómica) (Herman *et al*., 1996, Trinh *et al*. 2001). Precisamente se detectó que había metilación, y esto condujo a desarrollar la otra técnica *MethyLight* la

cual es una PCR en tiempo real cuantitativa, mucho más sensible y específica que la *MSP*. Por tanto, este método permitió conocer si la metilación observada era variable, lo cual fue comprobado.

En SCA2 solo existe una técnica molecular para el estudio de este gen, y es la PCR para el análisis de la expansión de CAG. Por tanto, el mérito de haber desarrollado estas técnicas está en que son las únicas que permiten estudiar otra faceta de la biología molecular del gen *ATXN2*, y se contraponen a la actual visión *tripletológica*, demostrando que más allá de la expansión del triplete de CAG, existen factores intrínsecos del gen que determinan el fenotipo o su variabilidad.

La casuística para estudiar este fenómeno fue amplia. Se analizaron 50 muestras de ADN de sujetos en diferentes diseños de estudio y aproximaciones: pares padres-hijos, familias SCA2, controles, casos SCA2 (heterocigóticos y homocigóticos), SCA3.

Las conclusiones fundamentales de este trabajo son:

1. *Se demostró la existencia de metilación en el promotor ATXN2, tanto en pacientes SCA2, SCA3 como en controles. Si esto es así, entonces es posible que la misma gobierne la expresión génica y por tanto la fenotípica.*

2. *La metilación justificó la anticipación genética sin inestabilidad intergeneracional. Si esto es así, la expresión asemeja el efecto de la expansión patogénica, siguiendo el modelo de ganancia de función.*

3. *La metilación se asocia preferentemente con el estado de portador de la mutación SCA2. Por tanto, el silenciamiento ocurre sobre la variante mutante limitando su efecto patogénico (silenciamiento alelo específica).*

4. *La metilación retarda o evita el debut de la SCA2. Por tanto, es posible que sea un mecanismo de defensa celular para contrarrestar el efecto tóxico de la ataxina-2 mutante.*

5. *La metilación es variable. Si esto es así, existe una gradación de estados de metilación en la población explicando la amplia variabilidad fenotípica y el pleiotropismo de la expansión de CAG o del gen ATXN2.*

Estas conclusiones permiten responder las preguntas de investigación enunciadas:

(2) ¿cómo se regula la expresión del gen *ATXN2*?, R/ controlando la activación de su estructura de este gen. (5) ¿el por qué de la variabilidad fenotípica y el pleiotropismo de la mutación? R/ por la existencia de metilación variable que modifica la expresión del gen, la cual pudiera ocurrir en diferentes estadios de la vida y ser diferente en cada tipo celular.

Hum Genet
DOI 10.1007/s00439-011-1101-y

ORIGINAL INVESTIGATION

Epigenetics DNA methylation in the core ataxin-2 gene promoter: novel physiological and pathological implications

José Miguel Laffita-Mesa · Peter O. Bauer · Vivian Kourí · Leodani Peña Serrano · Jane Roskams · Dennis Almaguer Gotay · Julio Cesar Montes Brown · Pedro Ariel Martínez Rodríguez · Yanetza González-Zaldívar · Luís Almaguer Mederos · Dany Cuello-Almarales · Jorge Aguiar Santiago

Received: 15 July 2011 / Accepted: 29 September 2011

Abstract Pathogenic CAG (cytosine-adenine-guanine) expansions beyond certain thresholds in the ataxin-2 (ATXN2) gene cause spinocerebellar ataxia type 2 (SCA2) and were shown to contribute to Parkinson disease, amyotrophic lateral sclerosis and frontotemporal lobar degeneration. Regulation of ATXN2 gene expression and the function of the protein product are not known. SCA2 exhibits an inverse correlation between the size of the CAG repeat and the age at disease onset. However, a wide range of age at onset are typically observed, with CAG repeat number alone explaining only partly this variability. In this study, we explored the hypothesis that ATXN2 levels could be controlled by DNA methylation and that the derangement of this control may lead to escalation of disease severity and influencing the age at onset. We found that CpG methylation in human ATXN2 gene promoter is associated with pathogenic CAG expansions in SCA2 patients. Different levels of methylation in a SCA2 pedigree without an intergenerational CAG repeat instability caused the disease anticipation in a SCA2 family. DNA methylation also influenced the disease onset in SCA2 homozygotes and SCA3 patients. In conclusion, our study points to a novel regulatory mechanism of ATXN2 expression involving an epigenetic event resulting in differential disease course in SCA2 patients.

Electronic supplementary material The online version of this article (doi:10.1007/s00439-011-1101-y) contains supplementary material, which is available to authorized users.

J. M. Laffita-Mesa (✉) · L. Peña Serrano · D. Almaguer Gotay · Y. González-Zaldívar · L. Almaguer Mederos · D. Cuello-Almarales
Center for Research and Rehabilitation of Hereditary Ataxia, CIRAH, Carlos Juan Finlay, Carretera Central Vía Habana km 51/2, Reparto Edecio Pérez, CP 80100 Holguín, Cuba
e-mail: laffita@ataxia.hlg.sld.cu; laffitamesa@yahoo.es

P. O. Bauer
Laboratory for Structural Neuropathology, RIKEN Brain Science Institute, 2-1 Hirosawa, Wako, Saitama 351-0198, Japan

V. Kourí · P. A. Martínez Rodríguez
Virology Department, Institute of Tropical Medicine "Pedro Kourí", Autopista Novia del Mediodía, km 6½, La Lisa, PO Box 601, Marianao, Havana, 13, Cuba

J. Roskams
Department of Zoology, University of British Columbia, Vancouver, BC V6T 1Z4, Canada

J. C. Montes Brown
Clinical and Surgical Hospital Ernesto Guevara, Las Tunas, Cuba

J. Aguiar Santiago
Division of Pharmaceutics, Department of Gene Therapy, Center for Genetic Engineering and Biotechnology (CIGB), Ave 31 Between 158 and 190, CP 10 600 Havana, Havana, Cuba

Introduction

Spinocerebellar ataxia type 2 is caused by CAG expansions in the coding region of the ataxin-2 gene. Pathologic feature in SCA2 is the atrophy and loss of Purkinje neurons from the cerebellar cortex. This leads clinically to deficits in motor coordination that affect gaze, speech, gait and balance (Lastres-Becker et al. 2008).

The ataxin-2 gene (*atxn2*) is located on the chromosome 12q24.1 and it harbors 25 exons encoding the ubiquitously expressed 140 kDa protein named ataxin-2 (ATXN2) (Pulst et al. 1996). Previously, it has been shown that the *atxn2* promoter is embedded in a typical CpG island (CGI) devoid of TATA box and appeared to be nonmethylated. The promoter activity is localized in the exon 1 of *atxn2*

Published online: 30 October 2011

and it is stronger than CMV promoter (Aguiar et al. 1999). Recently, an interaction between the KRAB-containing zinc-finger transcriptional regulator (ZBRK1) and ATXN2 was reported. The ZBRK1/ATXN2 complex activates the *atxn2* transcription, where ATXN2 acts as co-activator of ZBRK1 (Hallen et al. 2011). Other modulating mechanism regulating the atxn2 expression was not found highlighting that regulation of ATXN2 expression is poorly understood. Gaining insights in this mechanism is crucial, because pathogenic CAG expansions in the *atxn2* gene are directly and/or indirectly linked to the development of devastating neurodegenerative disorders such as SCA2, Parkinson disease, ALS and frontotemporal lobar degeneration (Neumann et al. 2006; Socal et al. 2008; Elden et al. 2010). In addition, the expression levels of ATXN2 have been found to be important in the development of some malignancies (Wiedemeyer et al. 2003; Hallen et al. 2011), suggesting a tumor suppressing activity for this gene. Cellular concentrations of ATXN2 were reported to be important in mRNA processing and translation under stress conditions through interaction with DEAD/H-box RNA helicase DDX6 and poly (A)-binding protein and consequent interference with the assembly of cellular P-body structures and stress granules (Nonhoff et al. 2007).

DNA methylation in mammalian cells is almost exclusively limited to the cytosine residues in the CpG dinucleotides and is a key epigenetic process involved in the control of gene expression (Miranda and Jones 2007; Lande-Diner et al. 2007). DNA methylation is mediated by DNA methyltransferases, which transfer the methyl group to the C5 position of the cytosine residues and as a result, 5-methylcytosine is created (reviewed by Jurkowska et al. 2011).

Here, we address the question whether ATXN2 gene promoter is methylated and if it is related to the variability seen in the age at onset of SCA2. We explored the hypothesis that the levels of mutant ATXN2 may be influenced by changes in gene expression and it is translated to phenotypic differences. Specifically, we believe that a key mechanism involving gene silencing is DNA methylation in the CpG of the promoter region of *atxn2* gene. Previously, it has been proposed that the allele-specific de novo methylation may result in a milder phenotype of SCA2. Different abilities of the de novo DNA methylation in SCA2 patients might then account for the broad range of the age at disease manifestation (Bauer et al. 2004) that cannot be explained simply by the length of CAG stretch.

We found that DNA methylation was associated with pathogenic CAG expansions in *atxn2* in SCA2 and SCA3 patients suggesting that the gene dosage of the mutant ATXN2 might be controlled by this epigenetic mechanism. This observation provides an alternative mechanistic platform for current insight into the pathogenesis of SCA2 and other disorders involving ATXN2 CAG expansions. It is also important for the diagnostic, prognostic and potential future therapeutic strategies in SCA2.

Methods

Design of the study

Our study was performed in DNA samples from SCA2 ($n = 23$) patients and control individuals ($n = 25$), considering atxn2 CAG repeats expansion (see also information in clinical description and table in supplemental material information and table 1). The methylation status of *atxn2* promoter was analyzed using two approaches, including qualitative methylation-specific PCR (MSP) (Herman et al. 1996) and semi-quantitative MethyLight assays (Eads et al. 2000). Primers for the MSP amplified the region from the nucleotides 92–242, 94–240 unmethylated and methylated reactions, respectively, while MethyLight oligos anchor precisely in the first ATG site leading atxn2 transcription (Sahba et al. 1998; Aguiar et al. 1999) and its close proximity as can be shown in Fig. 2a. In association studies, individuals were divided into groups based on either CAG repeat expansions in *atxn2* gene or their neurological status (normal controls ($n = 13$), SCA2 ($n = 23$) or SCA3 ($n = 12$) cases; see breakdown and clinical data in supplementary material. MethyLight data was processed as dichotomous (methylated vs. unmethylated) in the study involving the controls and SCA2 patients and as quantitative in the other study related to homozygotes and SCA3 patients, which also used absolute measures for ATXN2 methylation (Trinh et al. 2001).

Bioinformatics

Ataxin-2 gene sequence was retrieved from GenBank (accession number NG_011572.1). For CpG island prediction and analysis, the Windows-based program CpG analyzer was used (Xu et al. 2005), and the putative methylation CpG sites were determined by the Support Vector Machine software available at http://bio.dfci.harvard.edu/Methylator/ref.html (Bhasin et al. 2005). We utilized Promoter Scan software (http://www-bimas.cit.nih.gov/molbio/proscan/) for identification of putative Transcription Factor (TF) binding sites and putative promoter.

DNA bisulfite conversion

Genomic DNA was modified with Epitect® 96 Bisulfite Kit (Qiagen, Mainz Germany). This process exchanges non-methylated cytosine residues for uracil, whereas methylated cytosines remain intact if were methylated originally.

The efficiency of the conversion was confirmed by PCR at SCA2 locus (including the CAG stretch) in DNA samples treated with the bisulfite kit, where no PCR products were generated.

Methyl-specific PCR (MSP)

Firstly, we developed a MSP protocol using customized buffers and thermal conditions according to Herman et al. (1996), with 65°C as annealing temperature and 1.5 mM $MgCl_2$. The efficiency of the method was validated by Epitect MSP kit (Qiagen, Mainz Germany) on bisulfite-converted DNA. Equivalent DNA amounts were used for MSP and each assay was replicated in three independent experiments. Each converted DNA sample was quantified by Spectronic BioMate 3 Spectrophotometer (Merck, Germany) at 260 nm. Two sets of oligos were designed (http://www.itsa.ucsf.edu/urolab/methprimer) (Li and Dahiya 2002) and used for methylated- and unmethylated-specific MSP for determining the methylation status of five CpG at positions between sites 92 and 242 nt of the SCA2 gene. Primer sequences: methylated—At2-2m-F (forward) 5′-CGATTTTCGGTAAAGAGTTTTTATTC-3′, At2-2m-R (reverse) 5′-GAACCACCTAACTACGACGAA-3′ and unmethylated—At2-2u-F (forward) 5′-TTTGATTTTTGGTAAAGAGTTTTTATTT-3′, At2-2u-R (reverse) 5′-CCAAACCACCTAACTACAACAAA-3′ (CpG sites underlined).

Validation of the MSP

The performance and specificity of this MSP was evaluated using commercial bisulfate converted DNA (Qiagen, Mainz, Germany) in two different frameworks (1) *Heminested PCR* where a third oligonucleotide (TM-meth/TM-Unmeth) that is compatible with the 196–218 nt region within of the first amplicon amplified with primers used in the qualitative MSP (i.e. At2-2m-F and At2-2m-R), and most render a band with ~120 bp, which was corroborated as shown in Fig. 2b, lane 4. (2) *HotStartTaq d-Tect polymerase*, this reaction was performed by EpiTect MSP Kit and was able to discriminate between single-base mismatches at the 3′-end of primers during the annealing and extension steps. This kit was also used according to the manufacturer's suggestions such as different annealing temperatures and confirmed the specificity of our MSP.

Quantitative assay for methylation in ATXN2 promoter

Two sets of primers for amplifying unmethylated and methylated CpG islands by quantitative MSP by TaqMan® (Q-MSP) were custom designed either by us or by Tib-Mol-Biol (Germany). The sequences of the forward and reverse primers able to detect fully states were Meth-S: 5′-GTTTTGGTGTGTTTGTTTTTTGATGT-3′, MethR: 5′-CCACCCAAACCACCTAACTACA-3′; Unmeth-S: 5′-GCGTTCGTTTTTCGATGC-3′, A-unmeth: 5′-CACCCGAACCACCTAACTACG-3′. TaqMan® probes were as follows: TM-Meth: 6FAM5′-CgACgAAACTCgATAACCACCgC-3′BBQ, TM-Unmeth: 6FAM5′-CAACAAAACTCAATAACCACCACAAAACT-3′BBQ. These primers and probes are able to detect the methylation in 13 CpG dinucleotides, while each forward primer anneals at the ATG starting site (162 nt) and the six CpG dinucleotides (Fig. 2a). Q-MSP discriminates methylation only under successful annealing of primers and probes to the target DNA (reviewed in Eads et al. 2000). The methylation was measured and expressed by fractional cycle number referred to as the crossing point (C_t). Standard curve necessary for calculations of the percent of methylation in tested DNA samples was generated using fully methylated and fully unmethylated control DNA samples. The concentrations of methylated and unmethylated DNA were extrapolated from these curves converting Cp to ng of DNA, and the percentage of methylation was calculated as $\%M = [\text{ng methylated } atxn2\ (M)/(\text{ng methylated } atxn2\ (M) + \text{ng unmethylated } atxn2\ (U))]$, where total target DNA was expressed as the sum of $U + M$ (Eads et al. 2000; Trinh et al. 2001). Reactions were performed in Light Cycler 1.5 Instrument using LightCycler® TaqMan® Master mix (Roche, Mannheim, Germany). EpiTect® PCR Control DNA Set was used for positive real-time reactions (Qiagen, Mainz, Germany) and reactions without DNA template were included as negative control.

Statistics

The correlation of *atxn2* methylation status with CAG expansions or phenotypes was determined via Chi-square test of association and fisher exact probability test. Odds ratios were determined at http://faculty.vassar.edu/lowry/odds2x2.html. All statistical analyses were performed using STATISTICA data analysis software system version 6 (StatSoft Inc., 2003, Tulsa, OK, USA).The null hypotheses were rejected at $P \leq 0.05$.

Results

In silico predictions related to *atxn2* gene promoter methylation status and regulation

As a first step in determining the methylation status of *atxn2* promoter, we used Support Vector Machine to predict possible methylatable CpG sites (Bhasin et al. 2005). We found that 33 out of 100 CpG (33%) dinucleotides are methylatable in the core promoter between genomic positions 230 and

Table 1 CpG dinucleotides in the core ataxin-2 gene promoter

(Ordinal) and genomic position in the gene sequence							
(1)	6	(2)	22	(3)	24	(4)	27
(5)	31	(6)	40	(7)	51	(8)	57
(9)	70	(10)	81	(*11*)	*94*	(*12*)	*101*
(*13*)	*119*	(14)	127	(15)	137	(16)	140
(17)	145	(18)	148	(19)	150	(20)	154
(**21**)	**161**	(22)	166	(23)	173	(24)	177
(25)	187	(*26*)	*195*	(*27*)	*197*	(*28*)	*207*
(*29*)	*214*	(*30*)	*217*	(*31*)	*222*	(*32*)	*225*
(33)	240	(*34*)	*247*	(35)	251	(36)	259
(*37*)	*263*	(*38*)	*266*	(*39*)	*268*	(*40*)	*271*
(*41*)	*276*	(*42*)	*280*	(*43*)	*285*	(*44*)	*288*
(*45*)	*294*	(*46*)	*296*	(47)	302	(48)	308
(49)	318	(50)	321	(51)	332	(52)	336
(53)	342	(*54*)	*358*	(55)	368	(*56*)	*378*
(57)	402	(*58*)	*408*	(*59*)	*411*	(*60*)	*414*
(61)	417	(*62*)	*419*	(63)	424	(*64*)	*429*
(*65*)	*436*	(*66*)	*439*	(67)	453	(68)	456
(69)	466	(70)	474	(71)	477	(*72*)	*480*
(73)	495	(74)	503	(75)	507	(76)	515
(77)	519	(78)	521	(79)	529	(80)	531
(81)	539	(82)	543	(*83*)	*546*	(84)	550
(85)	554	(86)	560	(87)	563	(88)	565
(89)	570	(90)	573	(91)	584	(92)	586
(*93*)	*590*	(94)	597	(95)	607	(96)	611
(97)	614	(98)	618	(99)	623	(**100**)	**647**

In bold are the CpG close to the putative TSS in the SCA2 gene. Italic entries are the potentially methylated CpG

471 bp (Table 1). The core promoter is close to the two ATG start codons at 162 and 642 nt (Fig. 1a), harboring the *Sac*I restriction site previously used in cloning and identification of *atxn2* promoter (Aguiar et al. 1999). Abundant Sp1 transcription factor binding sites are present throughout whole promoter with strongest affinity in the core region (Fig. 1b). A promoter activity was also predicted at the 3′UTR of *atxn2* (Fig. 1d). Using the same approaches we also predicted the more likely promoter sequences in genes of the ataxins network previously reported in Lim et al. (2006) and sequence homology was searched in those genes. We found high homology (80%) between RBPMS gene promoter and the ATXN2, the former showing 25% methylatable CpG sites (not shown).

Development and validation of MSP conditions for mapping CpG methylation in *atxn2* promoter

Due to the lack of a specific method for studying methylation in the *atxn2* promoter, we developed and validated MSP conditions for this analysis. Our method was able to detect methylation status of five CpG sites covered by the oligonucleotides shown in Fig. 2a, b. Our MSP was specific for bisulfite-converted DNA and resulted in a single band at expected molecular weight (in each version of the MSP, belonging either to the methylated or to the unmethylated reaction) (Fig. 2b). We validated the MSP specificity described in "Methods". The result of this analysis yielded a single band with the expected molecular weight as can be shown in the right panel of Fig. 2b (lanes 1 and 2), representing products of the Heminested after the use of the amplicon generated in the first MSP (MSP product in lane 4). This reaction was also made on genomic DNA, and no products were found, lane 3 in the same gel. The described experiments and results represented by specific bands in several MSP reactions using control and tested DNA samples confirmed that our MSP protocol is suitable for the methylation detection in the specific CpG sites covered by the designed primers.

DNA methylation and pathogenic CAG expansions causing SCA2

Next, we used the MSP assay in conditions related to abnormal behavior of the *atxn2* gene to investigate the association of

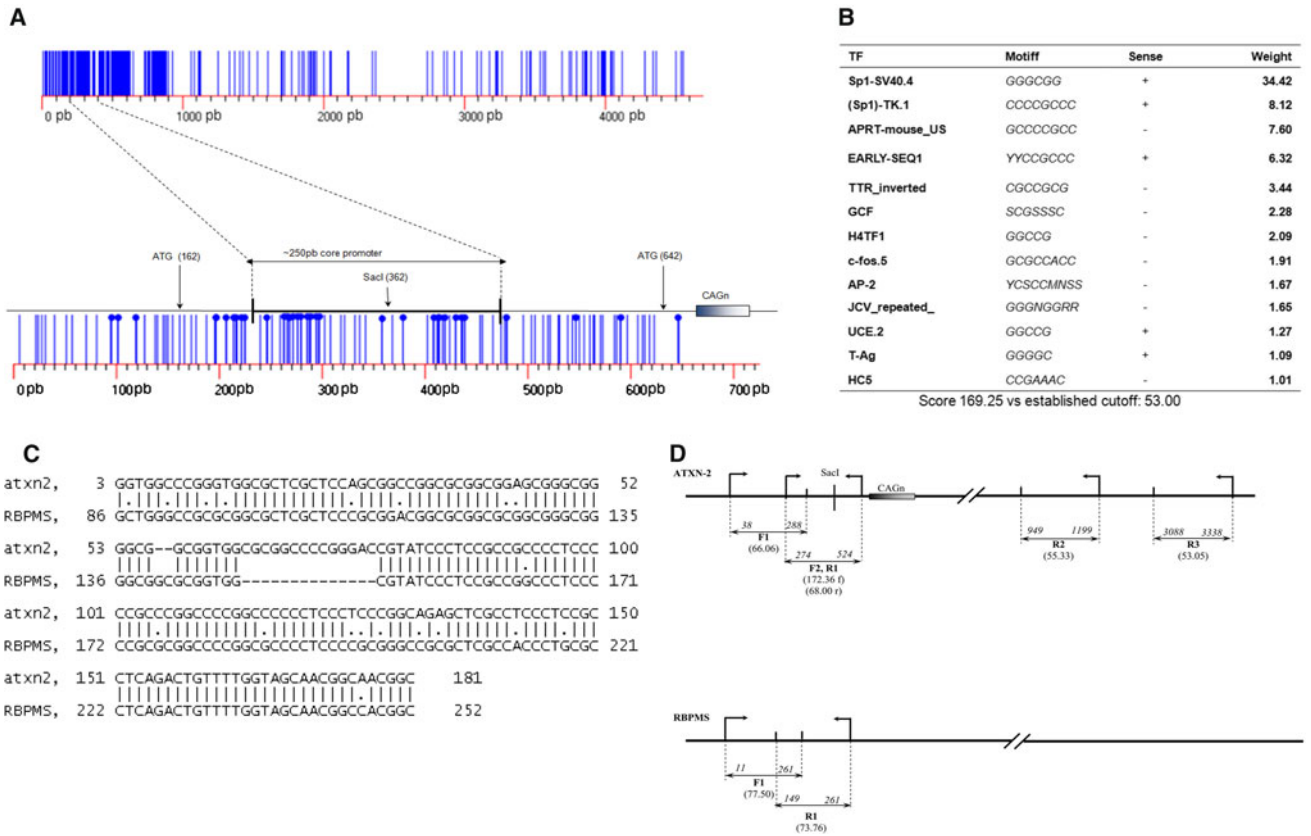


Fig. 1 In silico predictions of DNA methylation in *atxn2*. **a** *Top panel* Scheme of CpG islands distribution in ***atxn2*** gene sequence. *Horizontal lines* depict CpG dinucleotides which are overrepresented in exon 1. *Bottom panel* Predicted gene promoter and its core region with potentially methylated CpG (*lines* with *circle* at the *top*). Two putative transcription starting sites at positions 162 and 642 bp, and a genomic reference site (*Sac*I restriction) are pointed by *arrows*. CAG tract is shown as *rectangle*. **b** Putative binding sites for transcription factors (TF) in the core promoter region, sequence and weights of the prediction. The sites are sorted by the prediction weight. Below the table is the score of the predicted core region respecting the established cut off value of 53, which is lower than the predicted for *atxn2* gene. **c** Sequence alignment between *atxn2* and the RBMPS gene promoters. **d** Promoter predictions in the entire sequence of the *atxn2* (*top*) and the RBMPS (*bottom*) genes. Several promoter sequences were predicted both in sense and antisense orientation. The relative positions and corresponding scores of each promoter are shown. In the sequence with F/R, two sense promoters are predicted. *Vertical lines* with *arrows* indicate the orientation of the predicted promoter

methylation status at the *atxn2* promoter. For that we collected DNA samples from a cohort of SCA2 patients ($n = 14$) and healthy control individuals with normal CAG length in *atxn2* ($n = 13$) (see clinical information and breakdown table in supplementary material). MSP analysis resulted only in marginal association of the methylated/unmethylated (M/U) status with SCA2 (χ^2 3.63, $P = 0.056$, Fisher's exact test $P = 0.09$) (Fig. 2d). Visual inspection of the PCR products in agarose gels, however, revealed high variability in the intensities of "methylated" and "unmethylated" bands, suggesting that the separate quantification of methylation status for alleles with expanded and normal CAG repeat may be more appropriate to detect association between levels of methylation and CAG repeat length (Fig. 2b, left panel and lanes 2–9). In addition, MSP not only shown to be highly specific for the sequence of the atxn2 gene promoter but also shown to be very sensitive, as it detected as little as 1 ng of bisulfite-converted methylated or unmethylated DNA (Fig. 2e, right panel).

Variable methylation in the *atxn2* promoter and anticipation without CAG expansion change

It has been shown that the CAG repeat instability in *atxn2* leads to genetic anticipation (earlier age at onset, more severe disease course) in successive generations (Lorenzetti et al. 1997). In a nationwide epidemiological screening in Cuba, we found that a considerable portion of anticipation in SCA2 families was not justified by increase of the CAG repeat between generations. Moreover, a wide variance was observed in the age at disease onset (2–29 years), suggesting that other factors than the CAG repeat length could significantly contribute to this effect (Velázquez-Pérez et al. 2009). To analyze the effect of methylation pattern in *atxn2* promoter on SCA2 disease onset, we performed the MSP analysis in a SCA2 family (for additional clinical data see supplementary material) with evident anticipation without a CAG intergenerational

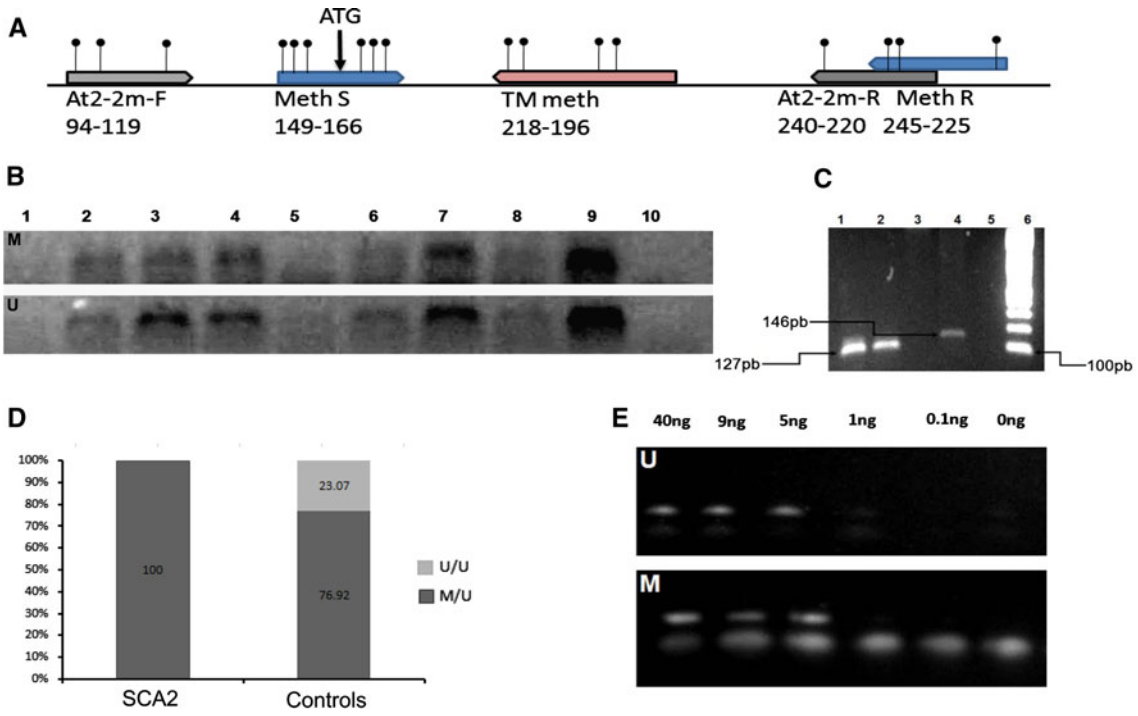


Fig. 2 Methyl Specific PCR for *atxn2* gene promoter. **a** Relative genomic positions of the annealing sites for methylated and unmethylated primers the *atxn2* promoter (only methylated versions displayed). Primers At2-2m-F/R are for qualitative MSP. Other oligos Meth-S/R and TMmeth are for the quantitative assay. **b** *Left panel* Representative PCR signal for methylation in the *atxn2* promoter in SCA2 patients and controls. *Lanes 1* and *10* mock, *lanes 2–9* MSP products for SCA2 patients and controls (*lanes 4* and *6*). **c** *Right panel* Specificity of the MSP PCR assay showing hemi-nested MSP products (*lanes 1* and *2*) using products from an initial MSP (*lane 4*). *Lanes 3* and *5* Genomic DNA and mock, respectively. **d** *Left panel* Association analysis using MSP data from controls and SCA2 patients. Marginal association of single allele methylation (M/U condition, one allele methylated and another unmethylated) with SCA2 was found. **e** *Right panel* Mixing experiment as suggested by manufacturers to determine the sensitivity of MSP assay in the *atxn2* gene. (The lowest detection was 1 ng, representing ~500 genomic copies). In the methylation reaction are observed primer dimers due to high content of CpG dinucleotides

instability. As shown in Fig. 3a, b, MSP ($n = 4$) in the analyzed SCA2 family (So) revealed differences in the methylation levels (visual inspection) between family members, suggesting variable expression of *atxn2* in each subject. Specifically, the mother (I.1; AO: 43 years; CAG: 22/39) (co-twin of I.2) transmitted the CAG expansion to her son (II.1; AO: 37 years; CAG: 22/39) who developed the disease 6 years earlier. Regarding the hypomethylation status in mother (I.1), it was similar to her healthy daughter (II.2; CAG: 22/22), but markedly higher than that seen in her affected son (II.1), who also shares the STR haplotype with his mother (Fig. 3a, b). Similar level of methylation was observed in the daughter of II.1, III.1 (11 years old, no clinical signs of SCA2), who inherited an expanded allele with 47 CAG repeats. The difference in *atxn2* promoter methylation level (I.1 vs. II.1) may provide an alternative explanation for the anticipation without CAG repeat instability observed in this SCA2 family. No fully methylated signals were identified suggesting that different variants of partial methylation exist in the *atxn2* promoter. A low detection ability of the MSP assays could be another reason; however, this possibility seems unlikely as our method detected as little as 1 ng (~500 genome copies) of methylated or unmethylated bisulfite-converted DNA (Fig. 2e). In conclusion, we found that hypomethylation in *atxn2* promoter in II.1 patient as compared to I.1 individual is likely a contributory factor to the observed intergenerational anticipation without CAG instability (age at onset shifted by 6 years). This anticipation might be the consequence of the relative overexpression of either normal or pathological ATXN2 alleles in II.1 patient. Partial methylations observed in the SCA2 family members suggest that a quantitative analysis of the methylation levels and/or testing in others CpG loci would be even more informative.

Methylation analysis in the core *atxn2* promoter by MethyLight technology

To overcome the MSP limitation and test this proof of concept of quantitative methylation, we next determined the methylation levels in the *atxn2* promoter using TaqMan® real-time PCR. In real-time PCR, the fluorescence increment is monitored during the amplification and cycle threshold values are obtained which inversely

A

Identities	Age	Age at Examination	Age at Onset	DD	Gender	ATXN2 CAG1/CAG2
I.1	66	58	43	23	F	22/39
I.2	66	58	50	16	F	22/39
II.1	45	39	37	8	M	22/39
II.2	--	--	--	--	F	22/22
II.3	--	--	--	--	M	22/22
III.1	11	9	--	--	F	22/47

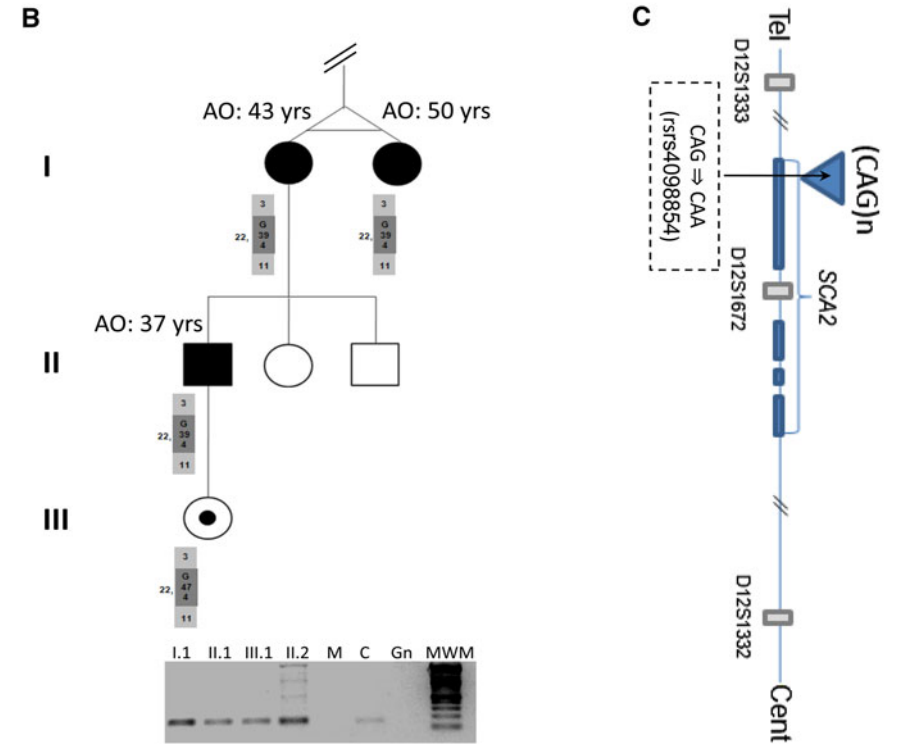


Fig. 3 Methylation and genetic anticipation in SCA2. **a** Phenotypes of the SCA2 family members. AO is lower in II.1 as compared to his mother with same CAG number. This anticipation is consistent with lower methylation of the *atxn2* promoter in II.1. **b** Pedigree showing the genealogy, haplotypic and methylation data of the familiy "So". In I.1 and her healthy daughter (II.2), the methylation levels appeared similar and were higher than in II.1 (gel at the *bottom* shows a representative analysis). *M* mock, *C* positive control-unmethylated bisulfite converted DNA, *Gn* genomic DNA, *MWM* molecular weight marker. **c** Analyzed STR and SNIPs polymorphisms close to atxn2 locus. Schematic map of the genetic polymorphisms used to study segregation of expanded CAG tract in family "So". The relative position at the chromosome 12 of each polymorphism is indicated. Ataxin-2 pathological CAG expansion in family "So" segregated with the haplotype was 3-[SCA2exp(G)]4-11, allele 3, 202 bp; allele 4, 287 bp; allele 11, 235 bp. STR markers were D12S1333 (~200kb from ATXN2), the intragenic D12S1672 (~20kb) and D12S1332 (~350kb) (Pulst et al. 1996, Sahba et al. 1998). The SNIP No. rs4098854, repesenting the change CAG to CAA is in the 5' CAG tract of ATXN2 gene

correlate with the amount of DNA templates. In the methylation experiments, a sample with more copies of hypermethylated or hypomethylated DNA display smaller cycle threshold value (crossing point, C_t). Oligos used in this assay (colored arrows oligos Meth S, TM meth and Meth R and their unmethylated version in Fig. 2a; "Methods") cover 13 CpG sites in the *atxn2* core promoter. More specifically, the forward primer's binding site covers two CpG dinucleotides in dyad configuration (CG**ATG** CG), within the recognized canonical TSS of ataxin-2 at 162 bp (Sahba et al. 1998; Aguiar et al. 1999 and present study in Figs. 1a, 2a). After serially diluting a fully methylated control DNA in solutions containing variable amounts of unmethylated DNA, the detection limit of our real-time PCR assay was determined to be around 0.03 ng of DNA corresponding to approximately four genomic copies (correlation coefficient: 1, error: 0.051; Fig. 4a) This assay is then 500-fold more sensitive than the qualitative method described above. We found that both full hypermethylation and hypomethylation states coexist in the *atxn2* promoter, with hypermethylation being more variable as compared to hypomethylation, which was more homogenous (Fig. 4b, c). Mean C_t for unmethylated signal was 22.97 ± 0.03, while methylation showed robust detectable signals, with mean $C_t = 34.73 \pm 0.8$, indicating that *atxn2* promoter is usually at least partly methylated.

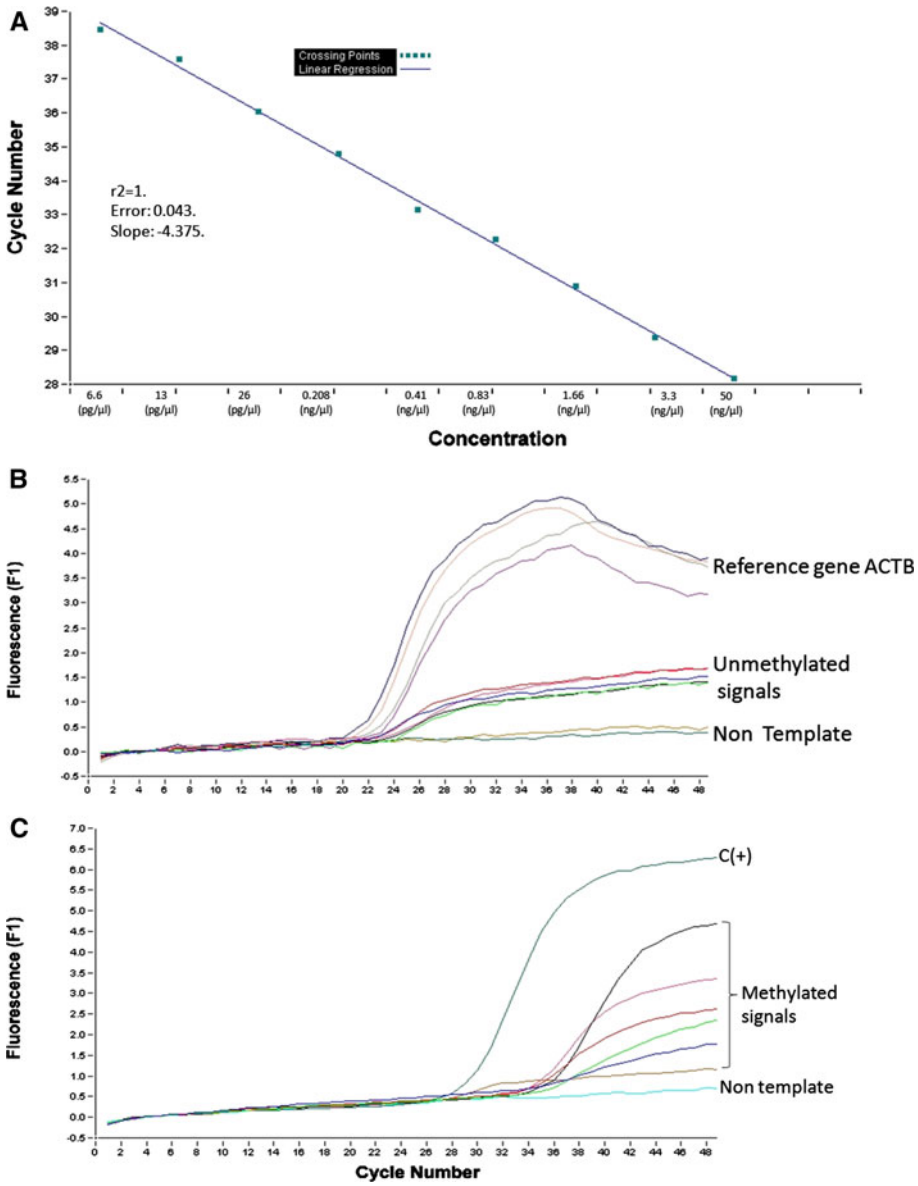
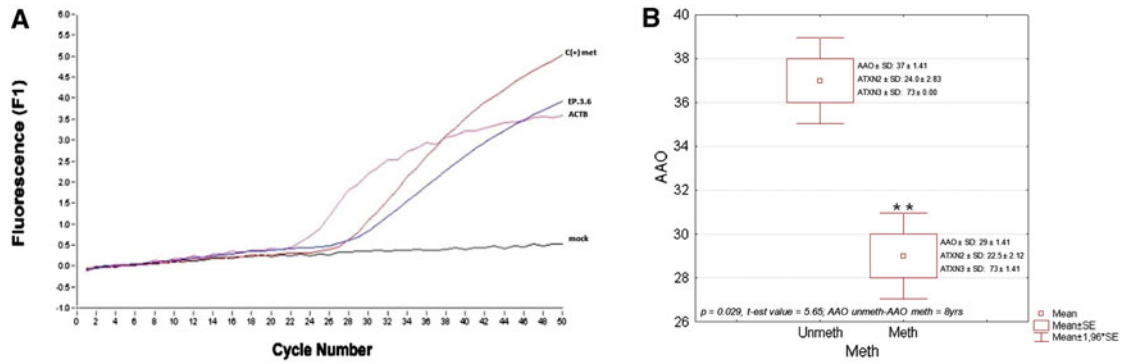


Fig. 4 Sensitivity and performance of QMSP for *atxn2* promoter. **a** Standard curve plots (C_t vs. DNA quantity) of serially diluted methylated DNA. The slope of −4.375 reflects high efficiency for amplification. The correlation coefficient (R^2) of 1 shows high degree of linearity over the entire range. **b** Amplification curves for unmethylated bisulfite converted DNA in a typical QMSP reaction for *atxn2* promoter. Actin (ACTB) reaction was used as internal reference for DNA input to illustrate that there are differences in Ct for "unmethylated" (**b**) and "methylated" (**c**) amplifications. **c** Reactions for "methylated" of the *atxn2* promoter. "Unmethylated" reactions are more homogenously grouped than "methylated"

Atxn2 hypermethylation accounts for early onset of MJD/SCA3 subtype 2

In a proposed network, ATXN2 and ATXN3 are closely connected by hub RBMPS gene that showed similarities with the *atxn2* promoter (present work). In addition, other hubs like A2BP1 connect ATXN2 to ATXN3, suggesting that the proposed functional linking might be biologically relevant for the involved phenotypes (Lim et al. 2006). Interestingly, methylation in *atxn3* promoter had no significant effect on AO; however, it was not be excluded as a larger sample or differential methylation patterns in MJD patients' brains should be investigated to make a conclusion on this epigenetic control (Emmel et al. 2011). Given the observed connections between ataxin-2 and ataxin-3 genes, we asked whether methylation in *atxn2* promoter could be a modifying factor of SCA3 phenotype. We determined methylation of the *atxn2* core promoter in the only 12 Cuban SCA3 patients from six families. In this analysis, two patients, Ep.3.3 and Ep.3.6, having normal ATXN2 alleles (21/21 and 22/24 CAG) and pathological ATXN3 alleles of 73 and 74 CAG, showed hypermethylation in the *atxn2* gene promoter (Fig. 5a, c). In other two patients, Ep.3.10 and Ep.3.5 (ATXN2: 26/26 and 22/22 CAG; ATXN3: 23/73 and 27/73 CAG), only *atxn2* hypomethylation was observed (Fig. 5c). All displayed patients fall in MJD clinical subtype 2 (MJD-2) (Durr et al. 1996). Despite the broad AO range reported for in MJD-2 (20–50 years), the AO in our MJD patients with *atxn2* hypermethylation was significantly lower (by 8 years) as compared to the patients with hypomethylated *atxn2* (t test $= 5.56$; $P < 0.003$) (Fig. 5a–c). In the patient



C

Cases Identities	ATXN2 CAG1, CAG2	ATXN3 CAG1, CAG2	AAO	Cp Meth ± SEM	Cp Unmeth ± SEM	Absolute PMR
Ep.3.10	26, 26	23, 73	36	0 ± 0.0	26.83 ±0.53	0
Ep.3.5	22, 22	27, 73	38	0 ± 0.0	23.04 ±0.81	0
Ep.3.3	21, 21	15, 72	28	33.45 ± 0.52	22.04 ±0.46	77.88
Ep.3.6	22, 24	37, 74	30	34.81 ± 0.54	0 ± 0.0	100

Cp: Crossing points, PMR: Percent of methylation relative.

Fig. 5 Effect of *atxn2* promoter methylation on SCA3/MJD onset. **a** Representative real-time PCR for MJD subtype 2 patients (EP.3.6) with different AO carrying normal *atxn2* alleles. **b**, **c** Onset of MJD symptoms was delayed in patients with hypermethylated *atxn2* promoter. In **c**, phenotypes and genotypic information is shown

Ep.3.6, a fully methylated (100%) signal was seen, while in the patient Ep.3.3, the degree of hypermethylation was 78%. This observation indicates a modifying effect *atxn2* methylation on SCA3/MJD disease severity, as no significant differences in CAG repeat lengths were seen between the patients (Fig. 5b). Our data show that the hypermethylation in *atxn2* promoter leading to a partial or complete epigenetic silencing of the *atxn2* gene expression influences the age at onset of SCA3 symptoms, and suggests that the development of this disease involves physiological functions of atxn2.

Hypermethylation in the *atxn2* core promoter is associated with CAG expansions in the SCA2 locus

Next we asked whether the methylation in the *atxn2* promoter accounting for phenotypic variability in ATXN2-related pathologies is linked to certain CAG lengths or whether this is a random event. We investigated the methylation status in the CpG loci detected as methylated in Cuban SCA3 patients ($n = 12$) and compared the frequency of hypermethylation to a cohort of SCA2 patients ($n = 9$; see Meth Data, clinical and genotypic information breakdown in supplementary material) (Fig. 6a). Using the real-time technique, we found that hypermethylation of *atxn2* was significantly more frequent in SCA2 ($n = 8/9$, 88%) compared to SCA3 patients ($n = 2/12$, 15%; Chi-square = 11.59, $P = 0.0007$, Fisher's exact test: $P = 0.001$). Extremely high odds ratio ([OR] = 44, 95% CI from 3.37 to 573.43) for this association suggested that alleles with pathogenic CAG expansions are preferentially hypermethylated (compared to normal alleles), probably as a part of the cellular self-defense mechanism to reduce the burden of cytotoxic mutant ATXN2 (Bauer et al. 2004). All SCA3 patients carried normal *atxn2* alleles of 20–26 CAG (mean 22.04 ± 1.84 CAG) but they cannot be considered as *bona fide* controls, because atxn2 and atxn3 share functional pathways involved in the ataxic phenotype (Lim et al. 2006; Lessing and Bonini 2008), and the observed association might be biased. To clarify this, we joined the data generated in the MSP and those of real-time PCR, as CpG analyzed are close tightly in the promoter sequence analyzed. In addition, we divided our cohort according to two classifiers: firstly, according to the CAG repeat length in *atxn2*, to expanded (≥27 CAG; $n = 25$) and unexpanded (≤27 CAG; $n = 23$) cases; and secondly according to the neurological status to patients with SCA2 ($n = 23$) or SCA3 ($n = 12$), and healthy control individuals ($n = 13$). We used this cut-off CAG length of 27, given novel results showing that alleles higher than 27 are pathological (reviewed in Elden et al. 2010), and independently in our population we found that alleles sized above 27CAG behave as those intermediates from 32 to 34 according somatic instability, and differently to those lower than 27 repeats (Laffita-Mesa et al. 2011).

In all comparisons, hypermethylation was associated with longer expansions in *atxn2* in either single-allele or the sum-of-both-alleles CAG number analysis (Fig. 6a, b).

A

No. of Subjects (n = 48)	Categories	Frequencies (%)	p-value	Association
Expanded (n = 25)	M/U	22/25 (88)	0.0007	positive
	M/M	2/25 (8)		
	U/U	1/25 (4)		
Unexpanded (n = 23)	M/U	9/23 (39)		
	M/M	1/23 (4)		
	U/U	13/23 (56)	0.00007	positive
SCA2 (n = 23)	M/U	20/23 (86)	0.000009	positive
	M/M	2/23 (8)		
	U/U	1/23 (4)		
SCA3 (n = 12)	M/U	1/12 (8)		
	M/M	1/12 (8)		
	U/U	10/12 (83)	<0.001a,b	positive
Controls (n = 13)	M/U	10/13 (76)	0.004	positive
	M/M	0/13 (0)		
	U/U	3/13 (23)		
No. of chromosomes (n = 96)				
Expanded (n = 50)	M	26/50 (52)	0.006c,d	positive
	U	24/50 (48)		
Unexpanded (n = 46)	M	11/46 (24)		
	U	35/46 (76)	0.006	positive
SCA2 (n = 46)	M	24/46 (52)	≤0.005e,f	positive
	U	22/46 (47)		
SCA3 (n = 24)	M	3/24 (13)		
	U	21/24 (88)	0.05	positive
Controls (n = 26)	M	10/26 (38)		
	U	16/26 (62)		

B

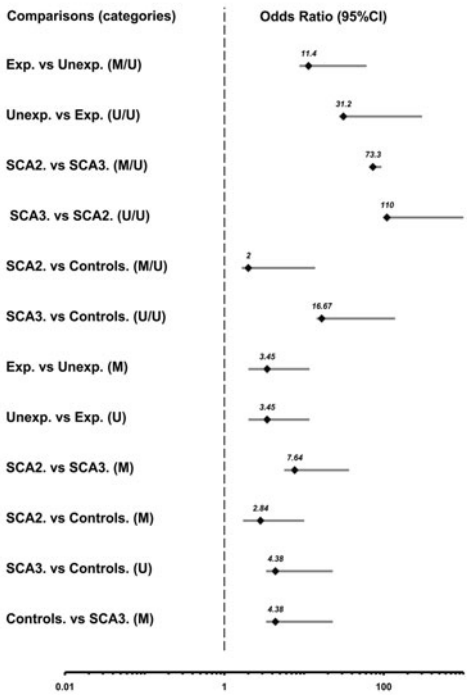


Fig. 6 Joint association analysis for methylation using Real-Time PCR and MSP. Epigenetics conditions were analyzed in the patients joining results of the MSP and real-time PCR to determine the level of association of certain condition (either methylated or unmethylated or the combinations) with CAG expansion or with the clinical status of individuals. **a** *Top table* is the association analysis of methylation taken the genotype (single allele combinations) and *bottom table* as chromosomes (the sum-of-both-alleles). **a** Both alleles unmethylated (U/U) was more frequently found in SCA3 compared to SCA2 (*a*) and controls (*b*). Methylated chromosomes were associated with expanded alleles than with unexpanded (*c*) and unmethylated with unexpanded (*d*). Similar association was shown between methylation and the clinical status SCA2 vs SCA3 (*e*) and controls (*f*). **b** *Right panel* indicates panel odds ratios with 95% CI for each comparison. In the *left part* are indicated each comparison and in *parentheses* the positive association for which odds ratio are significant

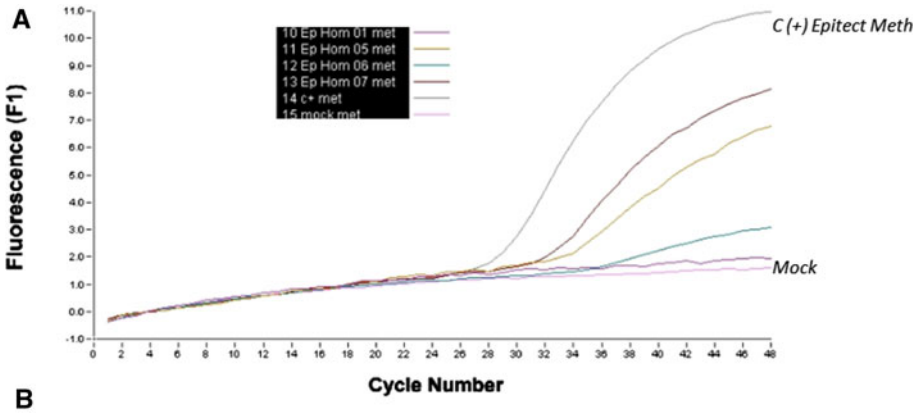
The M/U or M/M conditions were more frequent in the group of the expanded (≥27 CAG), compared to the unexpanded alleles with the most predominant, unmethylated (U/U) combination, representing a "double dose" of expressed ATXN2.

Genetic doses in SCA2 is under epigenetic control by DNA methylation in the *atxn2* promoter

Homozygosis in SCA2 is rare with only two cases reported so far (Spadafora et al. 2007; Ragothaman et al. 2008). Authors of these studies suggested that the presence of two pathogenic alleles increases the severity of the disease. It is plausible to hypothesize that the higher expression rate results in enhanced ATXN2 accumulation but has not been proven yet. We quantitatively assessed the methylation degree in the *atxn2* promoter in four homozygotes from our SCA2 cohort having both alleles with a pathogenic CAG expansion. For each individual in this study, haplotype and sequence data were available. From the two homozygotes with 33/34 CAG, hypermethylation was present in Hom07 patient, who started to experience imbalances at the age of 61 years while in the other proband (Hom01), who is still asymptomatic, full methylation was not detected (Fig. 7a, b). This unusual case may escape from the scope of our assay because we only asses a limited number of CpG; therefore, hypermethylation in other CpG in the vicinity may not be excluded. In the other two homozygous patients, Hom05 and 06 with CAG expansions of 43 and 39 CAG, respectively, either a marked delay in AO or still asymptomatic phase was associated with hypermethylation. Hypermethylation resulted in the absence of ataxia symptoms in the 47-year-old female patient Hom06 (Fig. 7b), where the expected AO is delayed at least by 11 years compared to AO in heterozygous SCA2 patients with 39 CAG repeats. In the patient Hom05 (43 CAG), a delay of 10 years according to expected AO heterozygous SCA2 patients with the same pathogenic expansion was observed (Fig. 7a).

Discussion

Here, we show for the first time that the *atxn2* promoter can be hypermethylated. Previously, it has been shown that the expression of mutant *atxn2* mRNA is regulated by allele-



B

ID (Pedigree)	CAG length , Reference Onset	Age (yrs)	AAO(yrs)	Meth (%)	ATXN2 CAG1/CAG2	CAG sequence	STR Haplotype
Hom07 (63)	--	83	61	20	33/34	24+8/25+8	3-4-11/3-4-11
Hom01 (114)	--	>65	0	0	33/34	24+8/25+8	3-4-11/3-4-11
Hom05 (89)	CAG = 43, 19(18-20)	37	30	0.15	33/43	24+8/43	3-4-11/3-4-11
Hom06 (2)	CAG = 39, 30(28-32)	47	0	0.41	34/39	25+8/39	3-4-11/3-4-11

Fig. 7 Hypermethylation in SCA2 homozygotes. **a** Representative real-time PCR traces of Q-MSP for the *atxn2* promoter in Cuban SCA2 homozygotic patients carrying intermediate and fully penetrant alleles. **b** Phenotypes and haplotypes of the patients, epigenetic information in regard to the clinical status is shown. Reference AO (AAO) for each CAG repeat number is also listed (Almaguer-Mederos et al. 2010)

specific DNA methylation and that the higher level of *atxn2* gene methylation lead to a delayed age at disease onset and to a milder SCA2 phenotype (Bauer et al. 2004). The *atxn2* promoter can be methylated in general population but occurs more frequently in alleles with pathogenic CAG repeat expansions. Such modification at regulatory regions (particularly gene promoters) correlates with the transcriptional state of the respective genes: hypermethylation represses while hypomethylation enhances transcription levels (Gimelbrant et al. 2007). The CpG sites situated close to the transcription start are predominantly targeted by methylation suggesting that these regions might be important in epigenetic control of ATXN2 expression. Recent findings on transcriptional activation of *atxn2* gene by ATXN2 protein (through an interaction with ZBRK1) (Hallen et al. 2011) and our data showing hypermethylation in this gene suggest that the self-activating function of ATXN2 may be modified by the methylation state of its own promoter and may influence the feedback loop for ATXN2 regulation during cellular stress (Hallen et al. 2011). In this light, the reduced or enhanced ATXN2 expression by hyper- or hypomethylation leading to reduced ATXN2 levels or increased cellular burden of expanded polyQ protein, respectively, may lead to an maladaptive stress response.

The proximity of the *atxn2* CAG tract to both TSS makes credible that noncanonical CAG-associated transcription overriding the main ATG start codon takes place, and this mechanism has been found recently leading to the homopolymeric toxicity at RNA and protein level (Zu et al. 2011), enforcing our hypothesis. MSP assays detected methylation signals upstream from the ATG start codon, suggesting that methylation may harbor wider region than we explored, targeting also other regions important for *atxn2* regulation. Candidate elements could include the CTCF binding site situated upstream of the CAG sequence (Filippova et al. 2001). Our in silico predictions showing possible several sense and antisense in *atxn2* promoter regions with methylated CpG dinucleotides is striking as convergent regulation of the *atxn7* trough usage an alternative promoter transcribing antisense noncoding RNA repressing transcription, together with the necessary participation of the CTCF regulator (Sopher et al. 2011) has been found recently.

Moreover, methylation status of *atxn2* may affect CAG repeat stability in somatic and germinal tissues, thus influencing both the pathology and intergenerational changes in the CAG repeat length. SCA7 shares several genetic features with SCA2 in terms of pathogenic CAG repeat length threshold, somatic stability and allele length (Laffita-Mesa et al. 2011). A study showing that DNA methylation at the CTCF binding element close to the CAG tract in ATXN7 promotes repeat instability (Libby et al. 2008), supports our postulate for SCA2. CAA interruptions of CAG tracts are considered a cis-acting factor influencing CAG repeat stability in ATXN2 (Choudhry et al. 2001; Laffita-Mesa et al. 2011). CAA interruptions in CAG repeat in ATXN2 have been associated with risk haplotypes for CAG instability in Cuban pedigrees (Laffita-Mesa et al. 2011) and in Parkinsonism (Charles et al. 2007). In addition, imperfections in CAG tract may synergize with CpG methylation in expanded triplet repeat and alter nucleosome assembly as reported in SCA1 and Fragile X syndrome (Mulvihill et al. 2005).

The differential methylation of *atxn2* might also have modifying effects in other pathological conditions connected

with ≥27 CAG repeat expansions in *atxn2*, such as ALS (Elden et al. 2010; Lee et al. 2011), neuropsychiatry disorders (Rottnek et al. 2008; Goyal et al. 2010), and cancer (Wiedemeyer et al. 2003; Hallen et al. 2011) or as the causative mutant gene in autosomal dominant parkinsonism (Charles et al. 2007; Payami et al. 2003). An intuitive relationship between the methylation pattern in *atxn2* promoter and phenotypes resulting from CAG expansions in the *atxn2* locus would be based on the balance of the methylation pattern and consecutive expression levels of ATXN2 at given condition and cell types. For example, ATXN2 level has been shown to be a prognostic marker in human neuroblastomas (Wiedemeyer et al. 2003). In ALS, ATXN2 is a modifier of the pathology as reduced levels of ATXN2 suppress TDP-43 toxicity (Elden et al. 2010). In addition, dysregulation of ATXN2 levels enhance or suppresses disease severity in Drosophila models of SCA1 and SCA3 (Al-Ramahi et al. 2007; Lessing and Bonini 2008). Derangement of this control linked to a pathogenic CAG expansion in the population may further define some human disorders and/or delineate phenotypes. The unexplained variance in the correlation between AO and CAG repeat length observed in SCA2 could be explained by methylation status of *atxn2* as shown in this study in SCA2 homozygous patients. It is not well understood whether indeed methylation and/or other epigenetic modification as has been shown in Purkinje cells and brain (Kriaucionis and Heintz 2009; Tahiliani et al. 2009) controls the ATXN2 expression. Our immediate objective therefore would be to address the potential connection of methylation and its related modifications (Kriaucionis and Heintz 2009; Tahiliani et al. 2009) on mRNA or even *atxn2* protein levels and the influence of methylation in the expression of ataxin-2 locus in different human conditions and cell type. Although we analyzed DNA methylation in lymphocytes, our observations are still important because of ubiquitous expression of ATXN2 and most of the CAG repeat-related disorders are multisystemic.

We also noticed that other genes causing or modifying different types of SCA (Lim et al. 2006) might be target for promoter methylation, as seen in *A2BP1* in Rett Syndrome models (Chahrour et al. 2008). Another target for methylation might be RBPMS, a gene with high content of methylatable CpG, which acts as a critical hub in protein networks involving ataxins. Hypermethylation of *atxn2* was pronounced in SCA3 patients than in SCA2, and this was related with earlier onset of SCA3 symptoms; this result most be extended to larger SCA3 populations. However, despite the reduced number of our sample, the observation that overexpression of *atxn2* synergistically enhances degeneration in drosophila model of SCA3 conversely reducing *atxn2* level mitigated ataxin3 toxicity (Lessing and Bonini 2008) makes our findings relevant for the SCA3 pathogenesis. In cases with early onset as we presented here a cellular strategy to palliate a potential synergistic damage by polyQ accumulation may be the epigenetic silencing of *atxn2* by DNA methylation, with a surrogate participation of *atxn2* motifs provided by homologues like ataxin-2-related protein (A2RP) (Figueroa and Pulst 2003).

Methods for methylation status analysis introduced in our study (e.g. Q-MSP) could also be used in clinical or diagnostic centers and provide information important for gaining insight and for improving management of these neurodegenerative diseases. As no cure is available for diseases involving CAG expansions in *atxn2* gene, the DNA methylation and the underlying mechanism could provide a rational target.

In conclusion, we found that atxn2 gene promoter is methylated, and this epigenetic mark was found in the context of pathological CAG expansions, associated with different clinical expressivity in SCA2 and SCA3. This findings accent the functional cooperation between pathways for ataxin proteins which may lead or limit the pathogenesis. The epigenetic phenomenon in atxn2 gene might explain other phenotypes linked to atxn2 levels. While this study was under review, Dick et al. 2011 found by Genome Wide Methylation Analysis that atxn2 was significantly methylated in a case series of Coronary Artery Disease, extending the relevance of atxn2 methylation to other pathologies and providing an anchor point that independently confirms our results on SCAs. Collectively, all the data support that the control of atxn2 concentration is pivotal both for its physiological function and the pathogenesis, this control may serve as targetable mechanism and rational for therapies.

Acknowledgments We are indebted to our Cuban patients. We would like to thank Dr. Guadalupe Guzman for providing real-time PCR and other facilities at IPK. Support of Dr Ocilia Rodriguez is acknowledged. We also thank all members belonging to the Dr. Vivian Kourí laboratory for their help and expertise in real-time PCR. Acknowledgement also belongs to our colleagues from Neurobiology Department and Dr Velázquez-Pérez for his support. This work was supported in part by Third World Academy of Sciences (TWAS) in the framework of the Research Grant Project 06-329 RG/BIO/LA UNESCO FR: 3240157855, a BIOLABS Research Award and a grant of the Cuban Ministry of Health and the Science and Technology Ministry all to JMLM.

Conflict of interest Authors have nothing to disclose.

References

Aguiar J, Santurlidis S, Nowok J et al (1999) Identification of the physiological promoter for spinocerebellar ataxia 2 gene reveals a CpG island for promoter activity situated into the exon 1 of this

gene and provides data about the origin of the nonmethylated state of these types of islands. Biochem Biophys Res Commun 254(2):315–318

Almaguer-Mederos LE, Falcon NS, Almira YR et al (2010) Estimation of the age at onset in spinocerebellar ataxia type 2 Cuban patients by survival analysis. Clin Genet 78:169–174

Al-Ramahi I, Perez AM, Lim J, Zhang M, Sorensen R et al (2007) dAtaxin-2 mediates expanded Ataxin-1-induced neurodegeneration in a Drosophila model of SCA1. PLoS Genet 3(12):e234. doi:10.1371/journal.pgen.0030234

Bauer PO, Zumrova A, Matoska V, Mitsui K, Goetz P (2004) Can ataxin-2 be down-regulated by allele-specific de novo DNA methylation in SCA2 patients? Med Hypotheses 63(6):1018–1023

Bhasin M, Zhang H, Reinherz E, Reche P (2005) Prediction of methylated CpGs in DNA sequences using a support vector machina. FEBS Lett 579(20):4302–4308

Chahrour M, Jung SY, Shaw C et al (2008) MeCP2, a key contributor to neurological disease, activates and represses transcription. Science 320:1224–1229

Charles P, Camuzat A, Benammar N, Sellal F, Destée A, Bonnet AM, Lesage S, Le Ber I, Stevanin G, Dürr A, Brice A (2007) French Parkinson's Disease Genetic Study Group. Are interrupted SCA2 CAG repeat expansions responsible for parkinsonism? Neurology 69:1970–1975

Choudhry S, Mukerji M, Srivastava AK et al (2001) CAG repeat instability at SCA2 locus: anchoring CAA interruptions and linked single nucleotide polymorphisms. Hum Mol Genet 10(21):2437–2446

Dick KJ, Nelson CP, Braund PS et al (2011) Genome wide methylation analysis in coronary artery disease. Heart 97:A42

Durr A, Stevanin G, Cancel G et al (1996) Spinocerebellar ataxia 3 and Machado-Joseph disease: clinical, molecular, and neuropathological features. Ann Neurol Apr 39(4):490–499

Eads CA, Danenberg KD, Kawakami K et al (2000) MethyLight: a high-throughput assay to measure DNA methylation. Nucleic Acids Res 28:E32

Elden AC, Kim HJ, Hart MP et al (2010) Ataxin-2 intermediate-length polyglutamine expansions are associated with increased risk for ALS. Nat Genet 466:1069–1075

Emmel VE, Alonso I, Jardim LB et al (2011) Does DNA methylation in the promoter region of the ATXN3 gene modify age at onset in MJD (SCA3) patients? Clin Genet 79:100–102

Figueroa KP, Pulst SM (2003) Identification and expression of the gene for human ataxin-2 related protein on chromosome 16. Exp Neurol 184:669–678

Filippova GN, Thienes CP, Penn BH et al (2001) CTCF-binding sites flank CTG/CAG repeats and form a methylation-sensitive insulator at the DM1 locus. Nat Genet 28:335–343

Gimelbrant A, Hutchinson JN, Thompson BR (2007) Widespread monoallelic expression on human autosomes. Science 318:1136–1140

Goyal N, Sur S, Sinha VK (2010) Childhood mood disorder in spinocerebellar ataxia type 2: a case report. J Neuropsychiatry Clin Neurosci 22:21–22

Hallen L, Klein H, Stoschek C (2011) The KRAB-containing zinc-finger transcriptional regulator ZBRK1 activates SCA2 gene transcription through direct interaction with its gene product, ataxin-2. Hum Mol Genet 20(1):104–114

Herman J, Graff JR, Myohanen S, Nelkin BD, Baylin BS (1996) Methylation-specific PCR: a novel assay for methylation status of CpG islands. Proc Natl Acad Sci 93:9821–9826

Jurkowska RZ, Jurkowski TP, Jeltsch A (2011) Structure and function of mammalian DNA methyltransferases. Chembiochem 12(2):206–222

Kriaucionis S, Heintz N (2009) The nuclear DNA base 5-hydroxymethylcytosine is present in Purkinje neurons and the brain. Science 324:929–930

Laffita-Mesa JM, Velázquez-Pérez LC, Santos Falcón N et al (2011) Unexpanded and intermediate CAG polymorphisms at SCA2 locus (ATXN2) in the Cuban population: evidences about the origin of expanded SCA2 alleles. Eur J Hum Gen. doi:10.1038/ejhg.2011.154

Lande-Diner L, Zhang J, Ben-Porath I, Amariglio N, Keshet I, Hecht M, Azuara V, Fisher AG, Rechavi G, Cedar H (2007) Role of DNA methylation in stable gene repression. J Biol Chem 20;282(16):12194–12200

Lastres-Becker I et al (2008) Spinocerebellar ataxia 2 (SCA2). Cerebellum 7:115–124

Lee T, Li YR, Ingre C et al (2011) Ataxin-2 intermediate-length polyglutamine expansions in European ALS patients. Hum Mol Genet 20:1697–1700

Lessing D, Bonini NM (2008) Polyglutamine genes interact to modulate the severity and progression of neurodegeneration in Drosophila. PLoS Biol 6:e29

Li LC, Dahiya R (2002) MethPrimer: designing primers for methylation PCRs. Bioinformatics 18:1427–1431

Libby RT, Hagerman KA, Pineda VV et al (2008) CTCF cis-regulates trinucleotide repeat instability in an epigenetic manner: a novel basis for mutational hot spot determination. PLoS Genet 4:e1000257

Lim J, Hao T, Shaw C et al (2006) A Protein–protein interaction network for human inherited ataxias and disorders of purkinje cell degeneration. Cell 125:801–814

Lorenzetti D, Bohlega S, Zoghbi HY (1997) The expansion of the CAG repeat in ataxin-2 is a frequent cause of autosomal dominant spinocerebellar ataxia. Neurology 49(4):1009–1013

Miranda TB, Jones PA (2007) DNA methylation: the nuts, bolts of repression. J Cell Physiol 213(2):384–390

Mulvihill DJ, Edamura NK, Hagerman KA et al (2005) Effect of CAT or AGG interruptions and CpG methylation on nucleosome assembly upon trinucleotide repeats on spinocerebellar ataxia, type 1 and fragile X syndrome. J Biol Chem 280:4498–4503

Neumann M, Sampathu DM, Kwong LK et al (2006) Ubiquitinated TDP-43 in frontotemporal lobar degeneration and amyotrophic lateral sclerosis. Science 314:130–133

Nonhoff U, Ralser M, Welzel F et al (2007) Ataxin-2 interacts with the DEAD/H-box RNA helicase DDX6 and interferes with P-bodies and stress granules. Mol Biol Cell 18(4):1385–1396

Payami H, Nutt J, Gancher S et al (2003) SCA2 may present as levodopa-responsive parkinsonism. Mov Disord 18:425–429

Pulst SM, Nechiporuk A, Nechiporuk T et al (1996) Moderate expansion of a normally biallelic trinucleotide repeat in spinocerebellar ataxia type 2. Nat Genet 14:269–276

Ragothaman M, Muthane U (2008) Homozygous SCA2 mutations changes phenotype and hastens progression. Mov Disord 23(5):770–701

Rottnek M, Riggio S, Byne W et al (2008) Schizophrenia in a patient with Spinocerebellar Ataxia 2: coincidence of two disorders or a neurodegenerative disease presenting with psychosis? Am J Psychiatry 165(8):964–967

Sahba S, Nechiporuk A, Figueroa KP, Nechiporuk T, Pulst SM (1998) Genomic structure of the human gene for spinocerebellar ataxia type 2 (SCA2) on chromosome 12q24.1. Genomics 47:359–364

Socal MP, Emmel VE, Rieder CR et al (2008) Intrafamilial variability of Parkinson phenotype in SCAs: novel cases due to SCA2 and SCA3 expansions. Parkinsonism Relat Disord 15(5):374–378

Sopher BL, Ladd PD, Pineda VV et al (2011) CTCF regulates ataxin-7 expression through promotion of a convergently transcribed, antisense noncoding RNA. Neuron 70:1071–1084

Spadafora P, Annesi G, Liguori M, Tarantino P, Cutuli N, Carrideo S et al (2007) Gene dosage influences the age at onset of SCA2 in a family from southern Italy. Clin Genet 72:381–383

Tahiliani M, Koh KP, Shen Y et al (2009) Conversion of 5-methylcytosine to 5-hydroxymethylcytosine in mammalian DNA by MLL partner TET1. Science 324:930–934

Trinh BN, Long TI, Laird PW (2001) DNA methylation analysis by MethyLight technology. Methods 25:456–462

Velázquez-Pérez LP, Sánchez GC, Santos NF et al (2009) Molecular epidemiology of spinocerebellar ataxias in Cuba: insights into SCA2 founder effect in Holguin. Neurosci Lett 454(2):157–160

Wiedemeyer R, Westermann F, Wittke I et al (2003) Ataxin-2 promotes apoptosis of human neuroblastoma cell. Oncogene 22:401–411

Xu YH, Manoharan HT, Pitot HC (2005) CpG Analyzer, a Windows-based utility program for investigation of DNA methylation. BioTechniques 39:656–662

Zu T, Gibbens B, Doty NS et al. (2011) Non-ATG-initiated translation directed by microsatellite expansions. Proc. Natl Acad. Sci. USA 108:260–265

DISCUSIÓN

III.3.2

3.3.2.1: Herencia epigenética (metilación del ADN) y gen *ATXN2:* de la hipótesis a la demostración y a nuevos mecanismos epigenéticos

Bauer *et al*., 2003, basados en observaciones en dos pacientes SCA2, sugirieron que la expresión de ataxina-2 podía controlarse mediante silenciamiento alelo específico. Estos autores postularon que este control podía ser mediado por metilación *de novo* del alelo mutante. Esto se comprobó, *in vitro*, en células derivadas de dos mujeres con SCA2 y genotipo 22/37CAG y edades de inicio discordantes (67 *vs.* 20 años). Estas células se trataron con la droga demetilante 5-aza-desoxicitidina (5-Aza-CdR), y se determinó ***cualitativamente*** los niveles del ARNm de *ATXN2*. Dado que se observaron cambios en la expresión del ARNm del alelo mutante en las células tratadas, se concluyó que existía metilación alelo específica. El paciente de mayor edad biológica y edad de debut para SCA2, mostró mayor patrón de metilación en el alelo expandido que el paciente con menor edad de debut de la ataxia. Todo esto era congruente con el efecto silenciador de la expresión del ARNm patológico, lo cual se interpretó como evidencia directa de metilación del promotor SCA2. Estos hallazgos sirvieron de marco conceptual, pero no se probó la existencia de metilación sobre el gen SCA2.

Se reconocen 5 puntos de control de la expresión génica: activación de la estructura génica, inicio de la transcripción, procesamiento del transcripto, transporte al citosol, traducción (Lewin., 2004). La metilación se localiza jerárquicamente en el primero de estos puntos. Por tanto, aunque estos hallazgos sirvieron de marco conceptual para el presente trabajo, no probaron rigurosamente la existencia de metilación sobre el gen SCA2 porque el efecto pudo deberse al control en cualquiera de estos puntos. De hecho, a pesar de observarse un efecto dosis-respuesta, esto pudo estar dado por una *trans*-activación regulando/modulando la expresión de este gen. También se conoce que esta droga ejerce el mismo efecto sobre las histonas de la cromatina, pudiendo ser una explicación alternativa a lo observado.

Sobre la base de la hipótesis: *existencia de factores modificadores del dueto: longitud del triplete CAG→fenotipo,* del postulado de Bauer et al., 2003 y las observaciones de Aguiar et al., 1999 en la presente investigación se identifica metilación de los sitios CpG de la región promotora de este gen. Esto además se relacionó con un control sobre el debut de los síntomas de SCA2. La metilación mostró además ser variable en los casos SCA2, lo cual es congruente con la variabilidad fenotípica de esta enfermedad.

Por tanto, estas son evidencias categóricas de un componente epigenético directamente vinculado al gen *SCA2/ATXN2*. No existe reporte previo de este tipo de hallazgos, y representa una nueva ruta en la investigación de la genética humana relacionada con la SCA2. Después de la identificación de la mutación SCA2, este es el segundo estudio en orden de importancia relacionado con la genética de este segmento de ADN y su influencia sobre el fenotipo.

Existen otros hallazgos que vinculan a *ATXN2* y a su producto génico con procesos epigenéticos. La función de ataxina-2 es desconocida, sin embargo recientemente se situó en una ruta por antonomasia epigenética. Ataxina-2 es requerida para la óptima represión mediada por micro-ARNs de varios transcriptos en los glomérulos de receptores del olor (McCann *et al.*, 2011). Además, se ha observado que especies largas de ARN no codificantes -en inglés *long-non-coding* RNA- que son ejecutores epigenéticos median la expresión del alelo mutante (Scoles *et al*., 2013).

Nuevos reportes sugieren una interacción entre el regulador transcripcional ZBRK1 y ataxina-2. El complejo ZBRK1/ATXN2 activa la expresión de ataxina-2 (Hallen *et al*., 2011). Esto sugiere además una acción en *cis* de ataxina-2 mediando su propia expresión, donde la metilación de su promotor sería clave en un ciclo de regulación de la expresión de este gen mediante *long-non-coding* RNA específicos.

Basado en las marcas epigenéticas en el promotor de *ATXN2* y en la disminución de ataxina-2 en células pluripotentes se plantea que las células en estadios embrionarios, son susceptibles a la hipermetilación y que tanto la sobreexpresión como la pérdida por silenciamiento contribuyen a la neurodegeneración. Hallazgos que apoyan esto son la alta expresión de ataxina-2 en neuronas diferenciadas (Xia *et al*., 2013) y retardo en el desarrollo de portadores de expansiones masivas (Babovic-Vuksanovic *et al*., 1998). Por otra parte, la deficiencia de esta proteína es detrimental durante el desarrollo embrionario (Lastres-Becker *et al*., 2008, Kiehl *et al*., 2006). Las expansiones masivas ≥200CAG se asocian con fenotipos neonatales, infantiles y con un cuadro distinto en cuanto a debut, y progresión comparado al de las expansiones patogénicas más comunes. Mucho más reciente se demostró, en modelos animales, que contrario a lo que se pensaba, la transcripción de este gen comienza por el segundo sitio de inicio de la transcripción y no por el primero, ambos muy próximos en la secuencia separados sólo por ~300pb (Scoles *et al*., 2013). Aguiar *et al*., 1999 encontraron que la región promotora se extendía desde -436pb a 362 relativo al primer ATG (768pb). A esta secuencia se une el factor ETS1, (en inglés *E-twenty six*) necesario para la expresión de *ATXN2*, lo cual además involucra remodelamiento cromatínico de la región (Scoles *et al*., 2013). Estos hallazgos son consistentes con un efecto regulatorio de la metilación sobre el primer ATG, posiblemente silenciando, actuando como control epigenético de la expresión de variantes ataxina-2 potencialmente más patogénicas.

El conocimiento a profundidad de estos procesos será crucial no sólo para entender la patogénesis de esta enfermedad, sino también su inicio, gravedad y pronóstico y nos permitirá avanzar en el tratamiento mediante la terapia epigenética.

DISCUSIÓN GENERAL

VI

4. DISCUSIÓN GENERAL

Alelos largos en Cuba y mecanismo mutagénico de la expansión trinucleotídica de CAG

La observación inicial del elevado número de variantes largas llevó a revisar la validez de la linealidad entre frecuencia de alelos largos y de SCA2. En Cuba se demostró que la frecuencia de estas variantes se relaciona con la frecuencia absoluta de SCA2. En Méjico también se cumple esto, donde además esta SCA es la de mayor frecuencia, y es el segundo país de más frecuencia de SCA2 (Magaña *et al*., 2013, Alonso *et al*., 2007). El complemento a estas observaciones es la demostración de una relación lineal, donde las variantes largas explican la frecuencia relativa de la SCA2 a nivel global. De acuerdo con esto, un incremento en la frecuencia de alelos largos se amplifica en múltiplos de cinco en la frecuencia relativa de SCA2. Esta relación es además congruente con el cuadro epidemiológico de la SCA2 a nivel internacional: Cuba, Méjico e India como los de mayor frecuencia de SCA2 y Japón el de menor frecuencia de SCA2.

A pesar de la no universalidad para otros *loci* con expansiones trinucleotídicas causantes de SCAs, esta ley refuta lo planteado por Juvonen *et al*., 2005 en la población finlandesa y por Sobczak & Krzyzosiak., 2004 en Polonia, argumentando en contra de la existencia de esta relación. Por tanto, se puede concluir que en regiones donde exista elevada frecuencia de alelos largos, hay una alta frecuencia de SCA2 y viceversa.

Al analizar 109 trasmisiones de mutaciones patogénicas encontramos que existían expansiones y contracciones (artículo 1), con tendencia a las expansiones. Además, se observa que las mutaciones eran capaces de revertir hasta los alelos intermedios de 29, 30 y 31 repeticiones, los que podían expandirse nuevamente hasta alcanzar la categoría

de expansiones patogénicas, lo que sostendría la prevalencia de la SCA2 en Cuba. Esto sugiere un balance entre la pérdida (por reversión o selección negativa de grandes expansiones) y la ganancia de mutaciones.

Según esta concepción, lejos de existir haplotipos predisponentes, los alelos expandidos generan, a través de las reversiones, alelos largos los que nuevamente se expanden y sostienen la prevalencia de la enfermedad.

De manera que el prerrequisito para sostener la prevalencia sería un tamaño de CAG aumentado y no la existencia de factores en *cis* en desequilibrio de ligamiento con haplotipo(s) particular(es). Por tanto, todas las variantes con CAG largos de igual tamaño tendrían la misma probabilidad de expandirse, minimizando la influencia de la secuencia local del ADN sobre la inestabilidad. Esto a su vez sugiere igual diversidad haplotípica entre pacientes y la población general, lo cual no se cumple en SCA2. En contraposición, se ha visto que el haplotipo CC brinda proclividad a la inestabilidad de la expansión, mientras que GT protege y ambos haplotipos tienen frecuencias distintas (Choudhry et al., 2001).

En nuestro caso las observaciones de expansiones patogénicas esporádicas así como en familias de la población en general no relacionadas con las familias SCA2 (artículos 2 y 3), es indicativo de múltiples haplotipos coexistiendo como punto de partida, los que contribuyen decisivamente a aumentar y sostener la prevalencia de mutaciones patogénicas. Es válido añadir que en la familia ELA, el haplotipo extragénico de microsatélites en padres fue muy distinto al de los hijos, aunque se conserva el bloque haplotípico intragénico.

Las siguientes observaciones argumentan a favor de la introducción de la SCA2 en Cuba, mediante variantes predispuestas a inestabilidad. 1) Las variantes largas (con CAG >23-31) que son no patogénicas pierden interrupciones de CAA, en la medida

que la expansión de CAG aumenta. 2) Estas variantes comparten el haplotipo con familias SCA2, pero a la vez este haplotipo es frecuente en la población en general. 3) A partir de las expansiones con CAG>26 se observa inestabilidad somática e intergeneracional, relacionándolas con expansiones CAG patogénicas y definiendo un umbral premutacional.

Por tanto en vez de un simple balance, el origen de las expansiones patogénicas parte de un *crosstalk* o diafonía reforzada donde existen reversiones y expansiones, pero dada la existencia de un haplotipo predisponente, contribuyen a la frecuencia de mutaciones *de novo* tanto los revertantes como los alelos normales con haplotipo predispuesto (Fig. 2A-C, Anexo 1).

La expansión de CAG en alelos largos es el resultado de un proceso que, según la hipótesis del haplotipo predisponente, es independiente de la secuencia del CAG. Esta última, una vez comenzada la mutagénesis pudiera potenciar su crecimiento por la estabilidad de la estructura no convencional formada por la secuencia repetitiva de CAG (Cleary *et al*., 2003) pero lo determinante inicialmente serían los factores en *cis*.

La expansión, por tanto, es el resultado final del fenómeno más que la causa en sí, aunque a los efectos de la genética clínica y la poblacional este tamaño de la expansión indique riesgo genético o frecuencia de SCA2. Aquí, se brindan evidencias apoyando esta premisa que además propone al haplotipo como evidencia de la existencia factores intrínsecos del *locus* que vulneran la estabilidad de la secuencia del CAG.

Estos factores pueden ser elementos en *cis* que introducen inestabilidad o que su afinidad por factores estabilizadores de la repetición está comprometida. La sustitución sinónima de A→G en el triplete de CAA (llevando al cambio de CAA→CAG) es uno de estos y lleva a aumentar la pureza del segmento de CAG lo que contribuye a la

formación de estructuras secundarias desestabilizadoras proclives a resbalones (Cleary *et al*., 2003).

Además del efecto de este triplete, los haplotipos GT o CC pudieran representar pérdida o ganancia de cajas de GC del islote de CpG respectivamente, lo cual influya tanto en el control de la expresión como en la susceptibilidad a inestabilidad. De acuerdo con esta hipótesis aquellas variantes con ganancia de cajas GC, tienen más potencial para metilarse y este estado desestabilizaría el segmento de tripletes de CAG haciéndolo más inestables (Fig.3A, B, Anexo 2). Se conoce que tanto la mutación como la metilación de sitios reguladores adyacentes a la secuencia de CAG de *ATXN7* promueven la inestabilidad de la repetición en un modelo murino para SCA7 (Libby *et al*., 2008). Esto permite vislumbrar un vínculo entre los estados epi y genéticos de *ATXN2* en el origen de las mutaciones.

Relacionado con el mecanismo mutagénico que lleva desde la expansión normal a la patogénica bajo el haplotipo CC. La pérdida de CAA (o mejor la sustitución sinónima A→G) es uno de los disparadores de la mutagénesis. Este evento ocurre por el extremo 5' y puede llevar a la pureza de la expansión de CAG, como ocurre mayoritariamente en los cromosomas con expansiones penetrantes SCA2, lo que supone pérdida ulterior del CAA del extremo 3' (Fig.4, Anexo 3). Estas pérdidas desestabilizarían el segmento de tripletes y llevarían al crecimiento de la expansión de CAG. En caso de las expansiones intermedias asociadas a ELA o EP, la pérdida del CAA 3'no es necesaria para la patogénesis.

Este mecanismo es exclusivo para interrupciones de CAA, pero no para los pocos casos Parkinsonianos con expansiones en *ATXN2* con interrupciones como CGG (Arg), o CCG (Pro) en vez de CAA (Furtado *et al*., 2004). Aunque es el mecanismo más común, no es universal, porque no explica la aparición de casos ELA, de demencia fronto

temporal, bajo el haplotipo GT con 32 unidades de CAG y patrón 8+4+9+8 u 8+9+4+8 encontrado en EUA (Yu *et al*., 2011, Ross *et al*., 2011).

Termodinámicamente los segmentos menos estructurados tienen un ΔG mayor que confórmeros más ramificados (Sobczak & Krzyzosiak., 2005), lo que supone diferencias en la estabilidad de la estructura no convencional del ADN y sugiere un mecanismo distinto al de resbalón, para el origen de alelos *ATXN2* con 3 interrupciones de CAA. Presumiblemente, conversión génica, intercambio desigual entre cromátidas hermanas por uno de los segmentos internos de CAG 5' ó 3' o recombinación entre alelos de 22 y 23 repeticiones de CAG con estructura 8+4+8 (22CAG) o 9+4+8 (23CAG).

En esta investigación se encuentra, por primera vez a nivel internacional, que la ELA segrega asociada a mutaciones (expansiones de CAG) *de novo* en este gen, mostrando un fenotipo distinto al de SCA2 y contribuyendo a la prevalencia de la expansión anómala de CAG en *ATXN2*. Sobre la base de diferencias genotípicas, fenotípicas y neuropatológicas entre estas dos enfermedades (Yu *et al*., 2011, Hart & Gilter., 2012) que segregan bajo mutaciones de similar naturaleza, pero donde los umbrales son distintos (como mostramos en el meta-análisis), nos indica la existencia de heterogeneidad genética de tipo alélica, en este continuo genético de expansiones trinucleotídicas.

Metilación epigenética modificadora del fenotipo

En esta investigación se identificó metilación de 18 dinucleótidos CpG en la región promotora de este gen. Este patrón se asoció con la variante mutante y con ausencia o retardo del debut de la SCA2 en una muestra de 50 sujetos. Así como justificó anticipación genética sin inestabilidad genética. La relevancia de estos hallazgos es que

junto a la expansión SCA2, y más allá de la inestabilidad del triplete de CAG el patrón de metilación dicta el comienzo de la enfermedad.

Se ha identificado metilación en otros sitios del gen y en enfermedades distintas a SCA2 siguiendo la regla: *silenciamiento asociado a un mejor pronóstico de la enfermedad* (Buckley *et al*., 2010, Ohm *et al*., 2010, Dick *et al*., 2012, Cicek *et al*., 2013). Estas observaciones sugieren que las desviaciones no fisiológicas en los niveles de expresión de ataxina-2, perturban de manera global varias rutas celulares. Estas desviaciones pudieran alcanzarse bien por mayor estabilidad del polipéptido (proteostasis), o por sobreexpresión. Por tanto, para contrarrestar esto *ATXN2* está sometido a un estricto control de su expresión, no sólo para garantizar una correcta economía celular sino por el carácter tóxico *per se* de los transcriptos y/o la proteína polyQ. Por lo que la metilación es una primera línea de defensa ante 'golpes degenerativos'.

Es importante notar que este patrón de metilación en periferia, sangre, se relacionó con un cambio 'crucial' en el fenotipo SCA2. Este último es consecuencia de una serie de eventos neurodegenerativos en neuronas, las cuales son muy diferentes a las células hematopoyéticas. Esto resalta el valor potencial como biomarcador de la metilación en la SCA2, pudiendo ser fácilmente detectado.

En relación con las técnicas introducidas para detectar metilación, es válido agregar algunos comentarios. Tanto la "*MSP*" como la "*MethyLight*" son las más actuales y resolutivas para el análisis específico de metilación a nivel génico. Para el gen SCA2, no existían técnicas para estudiar este fenómeno y los análisis contemporáneos son a nivel de genoma, y si bien tienen una gran utilidad pueden no detectar la señal en los 18 CpG en cuestión. Estos estudios pueden detectar variaciones comunes, pero son poco sensibles para detectar variantes raras (<5%) aún en muestras relativamente grandes.

La tecnología "*MethyLight*" es cuantitativa, y nos permitió detectar metilación en 13 sitios CpG contiguos, e incluyó CpG que flanqueaban uno de los dos ATG de la región promotora. Nuestra variante de *Methylight* sólo detectaba dos permutaciones de las 13^2=169 posibles. Estas representan los estados extremos completamente metilado o demetilado y aun así fueron identificados en nuestra muestra. Por tanto, esto sugiere que la hipermetilación de este gen ocurre naturalmente y puede ser detectada corrientemente. Además, dada la variabilidad en los valores de los *Ct* nos sugiere que los CpG identificados son sitios de metilación variable en cada sujeto (Fig. 5A, Anexo 4). Esto último podría estar relacionado con los niveles de expresión de ataxina-2, y a su vez con la intensidad clínica del fenotipo.

Por su parte la técnica *MSP* es una prueba de *'todo o nada'*, o sea detecta estados metilados o no metilados (Herman *et al*., 1996). Con esta se identificaron 5 CpG adicionales, para un total de 18 dinucleótidos contando los 13 de la *MethyLight*, sugiriendo que existen 324 posibles patrones de metilación de este gen considerando solo estos CpG. Por tanto, esto sugiere que puede existir un amplio espectro de variantes/estados metilados de este gen que son aún mucho más efectivas y 'sutiles' para una regulación fina de este gen, o la inactivación del alelo mutante (Fig.5B, Anexo 5). Estas además pueden variar de acuerdo al tejido y durante la vida en cuestión justificando la pleiotropía de este gen.

La región que se identificó como metilada en el *core* del promotor, fue identificada posteriormente por Scoles *et al*., 2013 como necesaria y suficiente para sostener la expresión de ataxina-2. Nosotros nos anticipamos a estos hallazgos dado que usando la técnica *MSP* identificamos metilación en la secuencia (-)GCTGAAGGCC, que es compatible con el oligo At2-2m-F. Scoles *et al*., 2013 identificaron este elemento secuencia como sitio de unión del factor de transcripción ETS1 (del inglés *E twenty Six*)

el cual controla la expresión de *ATXN2*. Mientras que aquí se observó metilación variable de este sitio, justificando la anticipación genética de SCA2. En este sentido, ETS1 utiliza el trinucleótido GGA como *core* para su conjugación con el ADN. Esto es mediado por tres motivos unión, que incluyen 3 α-hélice y la cadena 4, siendo la tercera hélice la fundamental en la unión a la región *core* GGA. En esta hélice, hay dos residuos de Arg muy conservados en toda la familia de estos factores. El primer residuo se une con gran afinidad al dinucleótido GA, mientras que la segunda Arg, se une a la G del CpG 9/10 de la secuencia entera (GCTGAAG**GC**C). La cadena 4 en ETS1 contiene una Tirosina (Tyr) en la posición 410 que contacta con el esqueleto del ADN, y se une al primer CpG de la secuencia consenso (Wei *et al*., 2010) (Fig. 6A, Anexo 5).

Se sabe además, que el estado de metilación de la secuencia *core* interfiere en la unión de ETS1, de los genes *housekeeping* con promotores similares a *ATXN2* (Gaston & Fired., 1995). En estos la secuencia actúa como *'portero molecular'* el cual en dependencia de su estado de metilación, permite o previene la unión de un amplio número de factores sensibles a la metilación incluyendo a CREB/ATF y NF-κB. El mecanismo propuesto incluye la exclusión de estos factores limitando el acceso al sitio diana en los promotores, mediante la inducción de cambios conformacionales (Fig.6b, Anexo 5). Todos estos hallazgos sugieren que la metilación del promotor *ATXN2*, es un mecanismo fisiológico intrínseco de este gen el cual vincula activación de la estructura génica controlando unión de factores transcripcionales. La capacidad de mantener un equilibrio entre estados epigenéticamente (in)activos -metilostasis- de *ATXN2,* pudiera además explicar la patogénesis de SCA2 (Fig. 7, Anexo 6).

Resumiendo esta parte, la metilación se asoció con el estado de portador lo que sugiere que el silenciamiento se ejerce sobre el alelo mutante, y esto fue compatible con la vida indicando que el alelo no silenciado es capaz de suplir la falta del silenciado

(haplosuficiencia del alelo salvaje). Este último, el alelo mutante, se asocia con el haplotipo universal CC. Por lo que estas observaciones sirven de marco conceptual para exógenamente apagar/silenciar terapéuticamente la expresión del alelo mutante, usando interferencia por ARN ligada al reconocimiento del haplotipo (silenciamiento alelo especifico).

Se pueden avizorar impactos potenciales de estos descubrimientos epigenéticos. Por primera vez, poder contar con marcadores moleculares, como la metilación del promotor, íntimamente vinculados con el gen *ATXN2* y con la mutación causal de SCA2, permitirá tener un biomarcador específico para poder conocer el estado de activación de este gen. Esto permitiría homogenizar a los casos de acuerdo al nivel de activación del gen *SCA2* para la aplicación de futuras terapias. También permitiría conocer los momentos claves durante el desarrollo, cuando ataxina-2 tiene activa su expresión lo cual unido a tratamientos paliativos pudiera potenciar su efecto beneficioso. También, puede encontrar aplicación en el asesoramiento genético. Hoy sólo se usa la expansión de CAG como dato molecular para sugerir una posible edad de debut, lo cual se realiza con gran nivel de incertidumbre.

Estos hallazgos tienen incluso un potencial impacto en otras áreas. Una ausencia de esta proteína por silenciamiento mediado por metilación sugeriría una pérdida de función, y por lo tanto daría luz acerca de la causa molecular. En enfermedades con similitudes fenotípicas en las que ataxina-2 contribuye, o en las que subfenotipos de SCA2 están presentes, este marco conceptual indicaría el fenómeno de metilación como alteración genética candidata.

Un modelo de ratón deficiente (*knock-out*) para ataxina-2 no mostró fenotipo neurológico, sino obesidad y resistencia a insulina vinculada a fallos en la regulación de

la traducción varios genes vinculados al metabolismo lipídico. Los animales nulicigóticos fueron pocos fértiles, lo cual extiende los efectos deletéreos del descontrol de la expresión de ataxina-2 a otros fenotipos (Kiehl *et al*., 2000, Lastres-Becker *et al*., 2008). Esto sugiere que en el humano, la ausencia crónica de esta proteína por metilación aberrante sería un defecto epigenético candidato en enfermedades diferentes a la SCA2. Por tanto, no sólo las alteraciones genéticas en *ATXN2* se vincularían con las enfermedades, sino que las marcas epigenéticas pudieran ser 'golpes' que dictan la severidad de SCA2 y ser determinantes genéticos para la emergencia de otras enfermedades.

Dado que la mutación SCA2 está vinculada con ELA y EP, gran parte de la epigenética de este gen puede influenciar a estas enfermedades.

Es evidente que todos estos hallazgos tienen un fuerte carácter traslacional en SCA2, e impone un nuevo nivel de análisis en el cual no sólo el tamaño de la expansión de tripletes (CAG(A)/polyQ) es importante, sino su estado de activación y contexto genético. Queda a la ciencia futura poder entender y aplicar este nivel e interrelacionarlo con otros biomarcadores de la SCA2, y otras enfermedades sensibles a CAG expandidos en este gen así como a otros tipos de alteraciones genéticas.

CONSIDERACIONES FINALES

V

5. CONSIDERACIONES FINALES

Esta tesis comienza introduciendo preguntas centrales que guían la investigación relacionada con el gen *ATXN2* y la SCA2. Hasta aquí se ha respondido a tres de ellas relativas a la mutagénesis y los factores que la desencadenan, así como la metilación del ADN como mecanismo que regula la expresión de este gen y por tanto la expresión fenotípica. Para no perder en la objetividad se plasmarán aquí de acuerdo a los hallazgos presentados en los artículos.

(1) ¿cómo se origina la mutación SCA2? Esta surge bajo un haplotipo predispuesto, lo cual está determinado por factores propios del gen que desestabilizan el segmento de repeticiones de CAG, y esto potencia el crecimiento durante las trasmisiones de padres-hijos. Uno de los factores es la interrupción de CAA y su existencia parece responder a procesos evolutivos para conservar la función de este polipéptido, lo cual esta intrínsecamente ligado a la secuencia global del gen más que a la extensión del segmento repetido de CAG.

(2) ¿cómo se regula la expresión del gen *ATXN2*? La metilación de varios sitios CpG en la secuencia promotora es uno de los reguladores de la expresión de este gen, y presumiblemente actúa como *'portero molecular'* en dependencia de su estado epigenético evitando o permitiendo la unión de factores de trascripción. Este es un mecanismo fisiológico, el cual a su vez es una defensa primaria para controlar la sobreexpresión así como la *proteostasis* de ataxina-2 mutante o la salvaje.

(3) ¿el porqué de la variabilidad fenotípica y el pleiotropismo de la mutación? de acuerdo con los hallazgos de metilación, la presencia de patrones de metilación diferenciales en la población es central para la variabilidad fenotípica de SCA2. A la

vez que esto mismo, pero a nivel tisular, y el fallo de retener este mecanismo en las distintitas poblaciones neuronales y durante la vida explicaría el pleiotropismo de este gen. De manera que entre esta pregunta y la anterior existe un fuerte vínculo, así como con la primera si tenemos en cuenta que la metilación también influye en la estabilidad de segmentos de CAG en otros genes similares a *ATXN2*.

Los datos presentados en los artículos así como lo discutido previamente, son congruentes con la existencia factores intrínsecos del gen que pueden gobernar, influenciar, modificar tanto el origen de la mutación, como su expresión fenotípica independientemente de la existencia o no de la expansión de CAG (salvaje o mutante). Más allá, esta depende de estos factores y siendo específico la presencia del haplotipo CC o GT es exclusiva en *homo sapiens*, que pudo responder la mejora de la función de ataxina-2 o al control de su expresión, y en presencia o no de las expansiones mutantes se expresa como una u otra enfermedad. En tanto, que la presencia del islote de CpG es independiente de la existencia de CAG porque está presente en *Mus musculus* el cual tiene un solo CAG/Q (Nechiporuk *et al*., 1998). Este islote le imparte un potencial predeterminado para poder regular la expresión de este gen, que a la vez pudiera influenciar la estabilidad del segmento repetitivo de CAG.

CONCLUSIONES Y RECOMENDACIONES

VI

8. CONCLUSIONES Y RECOMENDACIONES

(1) Las interrupciones de tripletes de CAA y el haplotipo asociados a la expansión repetitiva de CAG, determinan la inestabilidad de la mutación en *ATXN2*. Por tanto, son factores de riesgo, predisponen a la inestabilidad intergeneracional del segmento repetitivo de CAG de este gen.

(2) Existe predisposición genética a la inestabilidad de la repetición de CAG en *ATXN2*, dada por variantes normales con haplotipos particulares las cuales contribuyen a la existencia de un estado premutacional, y este a la generación de nuevas mutaciones en *ATXN2*.

(3) La linealidad entre la frecuencia de variantes largas y la de SCA2, predice la abundancia de la enfermedad a nivel global. Esto está respaldado por la mayor frecuencia de variantes largas, explicando la mayor frecuencia de la enfermedad a nivel mundial en Cuba.

(4) Las expansiones patogénicas en *ATXN2* pueden originarse *de novo* o *paso a paso*, lo cual subyace a la linealidad entre variantes largas y frecuencia de enfermedades causadas por este tipo mutación en este gen.

(5) Las mutaciones *de novo* con expansiones intermedias y las completamente penetrantes pueden segregar con fenotipos distintos al de SCA2, p. ej. ELA, esto sugiere que más allá de ser un factor modificador, esta mutación pudiera causar otras enfermedades por mecanismos patogénicos diferentes al de SCA2.

(6) El principal mecanismo mutagénico de las expansiones repetitivas de CAG en *ATXN2* es: pérdida de la interrupción CAA 5' en un contexto genético predispuesto, conllevando al crecimiento intergeneracional de esta mutación.

(7) Las técnicas desarrolladas permitieron detectar metilación variable en el promotor *ATXN2* asociada a expansiones patogénicas. Esto permite esclarecer el papel de la epigenética, en la expresión y variabilidad fenotípica de la SCA2.

(8) La metilación del promotor *ATXN2* modifica el fenotipo SCA2, retardando o evitando el debut de la ataxia. Por tanto, pudiera ser usado como biomarcador pronóstico específico de las enfermedades relacionadas con este gen.

(9) La influencia de la metilación sobre el fenotipo, modula el efecto de la mutación y el haplotipo, la dosis génica y la estructura interna de la mutación SCA2. Esto sirve de marco conceptual para modular, de manera exógena, la expresión de la variante mutada mitigando la neurodegeneración y retardando el comienzo de la enfermedad.

RECOMENDACIONES

(1) Integrar al diagnóstico de SCA2, el haplotipo para poder determinar la propensión a inestabilidad y pronóstico a ELA o SCA2.

(2) Evaluar el uso del patrón de metilación como biomarcador pronóstico de la severidad de la SCA2 y otras enfermedades relacionadas con este gen.

(3) Realizar pesquisas de la expansión de CAG de *ATXN2* en la población general, para poder determinar la frecuencia de mutaciones *de novo*, así como el riesgo intergeneracional asociado. Estos valores de riesgos serán útiles para el correcto asesoramiento genético de las enfermedades vinculadas a expansiones de CAG en este gen.

REFERENCIAS BIBLIOGRAFICAS

VI

7. REFERENCIAS BIBLIOGRAFICAS (107)

- Affaitati A, de Cristofaro T, Feliciello A et al. *Gene* 267 (1), 89 (2001).
- Aguiar J, Santurlidis S, Nowok J et al. *Biochem Biophys Res Commun* 254 (2), 315 (1999).
- Albrecht M, Golatta M, Wullner U et al. *Eur J Biochem* 271 (15), 3155 (2004a).
- Albrecht M, Lengauer T. *Biochem Biophys Res Commun* 316 (1), 129 (2004b).
- Albrecht M, Lengauer T. *FEBS Lett* 569 (18-26) (2004a).
- Alonso E, Martínez L, De Biase I et al. *Mov Disord* 22, 1050 (2007).
- Al-Ramahi I, Perez AM, Lim J et al. *PLoS Genet* 3 (12), e234 (2007).
- Auburger, G.W, 103, pp. 423, Amsterdam (2012).
- Babovic-Vuksanovic D, Snow K, Patterson MC et al. *Am J Med Genet*, 79:383–387 (1998).
- Bauer PO, Zumrova A, Matoska V, et al. *Med Hypotheses* 63 (6), 1018 (2004).
- Berger, SL. *Nature* 447 (7143), 407 (2007).
- Buckley PG, Alcock L, Kenneth B et al. *Clin. Cancer Res* 16, 2971 (2010).
- Charles P, Camuzat A, Benammar N et al. *Neurology* 69, 1970 (2007).
- Chattopadhyay B, Ghosh S, Gangopadhyay PK et al. *Neurosci Lett.* 17 (345), 93 (2003).
- Choudhry S, Mukerji M, Srivastava AK et al. *Hum Mol Genet.* 10 (21), 2437 (2001).
- Chouliaras L, Rutten BP, Kenis G et al. *Prog Neurobiol* 90 (4), 498 (2010).
- Cicek MS, Koestler DC, Fridley BL et al. *Hum. Mol. Genet.* 22, 3038 (2013).
- Ciosk, RM. DePalma, J.R. Priess. *Development* 131 4831–4841(2004).
- Cleary J & Pearson C. *Cytog and Gen Res*, 100, 25–55 (2003).
- Corti O, Lesage S & Brice A. *Physiol Rev* 91, 1161 (2011).
- Daoud H, Belzil V, Martins S et al. *Arch Neurol* 68, 739 (2011).
- Dick KJ, Nelson CP, Braund PS et al. *Heart* 97, A42(Abstract) (2011).
- Eads CA, Danenberg KD, Kawakami K et al. *Nucleic Acids Res* 28, E32 (2000).

- Elden AC, Kim HJ, Hart MP et al. *Nature.* 466, 1069 (2010).
- Esteller M. *Annu Rev Pharmacol Toxicol* 45, 629 (2005).
- Esteller M. *Hum. Mol. Genet* 16, R50 (2007).
- Fallon L, Belanger CM, Corera AT et al. *Nature Cell Biology* 8 (834-842.) (2006).
- Figueroa KP, Pulst SM. *Exp. Neurol.* 184 669 (2003).
- Furtado S, Payami H, Lockhart PJ et al. *Mov Disord*. 19 (6), 622 (2004).
- Fuso A, Scarpa S, Grandoni F et al. *Anal Biochem* 350, 24 (2006).
- Gaston K, Fired M. Gene 157 (1-2), 257 (1995).
- Gavin AC, Bosche M, Krause R et al. *Nature* 415, 141 (2002).
- Gehman LT, Stoilov P, Maguire J et al. *Nat Genetics* 43, 706 (2011).
- Gispert S, Kurz A, Waibel S et al. *Neurobiol. Dis* 10.1016/j.nbd.2011.08.021 (2011).
- Gispert S, Twells R, Orozco G, et al. *Nat Genet* 4 (3), 295 (1993).
- Hallen L, Klein H, Stoschek C et al. *Hum Mol Genet* 20, 104 (2011).
- Hart & Gilter., J NEUROSCI 10.1523/JNEUROSCI.0996-12.2012.(2012).
- Hayes S, Turecki G, Brisebois K et al. Hum Mol Genet; 9 (12): 1753-58 (2000).
- Herman J, Graff JR, Myohanen S et al. *Proc. Natl. Acad. Sci.* 93, 9821 (1996).
- Huynh DP, Del Bigio MR, Ho DH et al. *Ann Neurol* 45, 232 (1999).
- Huynh DP, Figueroa K, Hoang N et al. *Nat Genet* 26, 44 (2000).
- Huynh DP, Maalouf M, Silva AJ et al. *PLoS One* 4, e6235 (2009).
- Huynh DP, Nguyen DT, Pulst JB et al. *Exp Neurol* 203 (531-541) (2006).
- Imbert G, Saudou F, Yvert G et al. *Nature Genet* 14 (285-291.) (1996).
- Infante J, Berciano J, Volpini V et al. *Mov Disord* 19 (7), 848 (2004).
- Juvonen V, Hietala M, Kairisto V et al. *Acta Neurol Scand.* 111 (3), 154 (2005).
- Kiehl TR, Figueroa KP et al, *Biochem Biophys Res Commun*. 339, 17 (2006).
- Kozlov G, Safaee N, Rosenauer A et al. *J. Biol. Chem.* 285, 13599 (2010).
- Lastres-Becker I, Brodesser S, Lütjohann D et al. *Hum Mol Genet* 17 (10), 1465 (2008).
- Lastres-Becker, I et al. *Cerebellum.* 7, 115 (2008).
- Lee T, Li YR, Ingre C et al. *Hum Mol Genet.* 20, 1697 (2011).
- Lessing D, Bonini NM. *Plos Biol* 6 (2), 266 (2008).

- Lewin, B. Genes VIII., 22 pp63. NJ, (2004).
- Li B, Carey M, Workman JL. *Cell* 128 (4), 707 (2007).
- Libby R, Hagerman K, Pineda V et al. PLoS Genet 4(11): e1000257 (2008).
- Lim J, Hao T, Shaw C et al. *Cell* 125, 801 (2006).
- Liu J, Tang TS, Tu H et al. *J. Neurosci* 29, 9148 (2009).
- Luger K, Mäder AW, Richmond RK et al. *Nature* 389 (6648), 251 (1997).
- Magaña JJ, Tapia YS, Velázquez L et al. *Clin Genet* (2013).
- Mangus DA, Evans MC, Agrin NS et al. *Mol Cell Biol* 24 5521 (2004a).
- Mangus DA, Smith MM, McSweeney JM et al. *Mol Cell Biol* 24, 4196 (2004b).
- Martin C, Zhang Y. *Nat Rev Mol Cell Biol* 6 (11), 838 (2005).
- McCann C, Holohan EE, Das S et al. *Proc. Natl. Acad. Sci.* 108, 655 (2011).
- Moazed, D. *Nature* 457 (7228), 413 (2009).
- Nechiporuk T, Huynh DP, Figueroa K, et al. *Hum Mol Genet*. 7: 1301-9 (1998).
- Ng H, Pulst SM and Huynh DP. *Exp. Neurol.* 208, 207 (2007).
- Nonhoff U, Ralser M, Welzel F et al. *Mol Biol Cell.* 18 (4), 1385 (2007).
- Nonis D, Schmidt MH, van de Loo S et al. *Cell Signal* 20, 1725 (2008).
- Oakeley, EJ. *Pharmacology & Therapeutics* 84 (3), 389 (1999).
- Ohm JE, Mali P, Van Neste L et al. *Cancer Res.* 70, 7662 (2010).
- Pulst SM, Nechiporuk A, Nechiporuk T et al. *Nat Genet.* 14 (3), 269 (1996).
- Pulst SM, Santos N, Wang D et al. *Brain.* 128, 2297 (2005).
- Ralser M, Albrecht M, Nonhoff U et al. *J Mol Biol* . 346, 203 (2005a).
- Ralser M, Nonhoff U, Albrecht M et al. *Hum Mol Genet* 14, 2893 (2005b).
- Ross OA, Rutherford NJ, Baker M et al. *Hum Mol Genet* 20, 3207 (2011).
- Rub U, Schols L, Henry P et al. *Prog. Neurobiol.* (2013).
- Saluz H, Jost JP. *EXS* 64, 11 (1993).
- Sanpei K, Takano H, Igarashi S et al. *Nat Genet* 14, 277 (1996).
- Santos N, Aguiar J, Fernández J et al. *Biot Apl* 16, 219 (1999).
- Sato F, Tsuchiya S, Meltzer SJ, Shimizu K. *FEBS J* 278 (10), 1598 (2011).
- Satterfield TF, Pallanck LJ. *Hum Mol Genet.* 15 (16), 2523 (2006).
- Schatz P, Dietrich D, Schuster M. *Nucleic Acids Res* 32, e167 (2004).
- Schluth-Bolard C, Ottaviani A, Tollefsbol T (ed), pp. 77 Amsterdam, (2011)
- Scoles DR, Pflieger LT, Khanh K et al. *Hum Mol Gen* (2013).

- Sequeiros J, Seneca S, Martindale J. *Eur J Hum Gen* 18 1188 (2010).
- Shibata H, Huynh DP, Pulst SM. *Hum Mol Genet* 9 (9), 1303 (2000).
- Siddiqui N, Mangus DA, Chang TC et al. *J Biol Chem* 282, 25067 (2007).
- Simon DK, Zheng K, Velazquez L et al. *Arch Neurol* 64, 1042 (2007).
- Sobczak K, Krzyzosiak WJ. *Hum Mutat.* 24 (3), 236 (2004).
- Swisher & Parker, *PLoS ONE* 5, e10006. (2010).
- Takano H, Cancel G, Ikeuchi T et al. *Am J Hum Genet*. 63 (4), 1060 (1998).
- Tharun, S. *Int Rev Cell Mol Biol* 272 (149-189) (2009).
- Trojanowski, JQ. *Neurosignals* 16 (5-10) (2008).
- Van Damme P, Veldink JH, van Blitterswijk M et al. *Neurology* 76, 2066 (2011).
- Van de Loo S, Eich F, Nonis D et al. *Exp Neurol* 215, 110 (2009).
- Velázquez L, Sánchez G et al. *Rest Neur* Neursc 21 (6), 273, (Abstract) (2003).
- Velázquez Pérez L, Sánchez G, Santos N et al. *Neu Lett.* 454, 157 (2009).
- Voineagu I WX, Johnston P et al. *Nature* 474, 380 (2011).
- Wei GH, Badis G, Berger MF et al. *The EMBO Journal* 29, 2147 (2010).
- Wiedemeyer R, Westermann F, Wittke I et al. *Oncogene* 22, 401 (2003).
- Wojdacz TK, Dobrovic A. *Nucleic Acids Res* 35, e41 (2007).
- Xia G, Santostefano K, Hamazaki T et al. *J Mol Neurosci*, DOI 10.1007/s12031-012-9930-2 (2013).
- Xiong Z, Laird PW. *Nucleic Acids Res* 25, 2532 (1997).
- Yang X, Yan L, Davidson NE. *Endoc Relat Cancer* 8 115 (2001).
- Yost JM, Korboukh I, Liu F et al. *Curr Chem Genomics* 5 (1), 72 (2011).
- Yu F, Sabeti PC, Hardenbol P et al. *PLoS Genet*. 1 (3), e41 (2005).
- Yu Z, Zhu Y, Chen-Plotkin AS, Clay-Falcone D et al. *PLoS One* 6, e17951. (2011).

ANEXOS

VIII

8. ANEXOS

Anexo 1

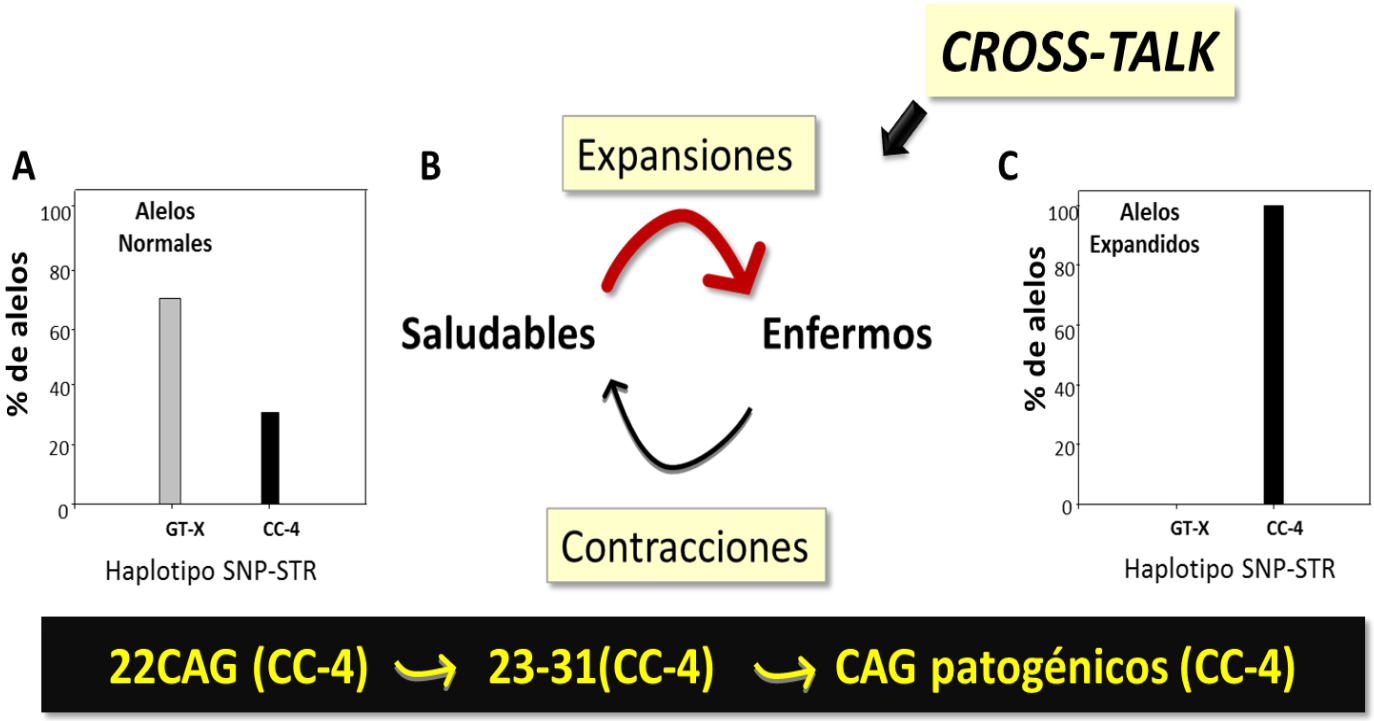


Figura 2. Efecto del haplotipo sobre el origen y mantenimiento de las expansiones patogénicas en *ATXN2*. A) Distribución del haplotipo de SNP en una población ideal (p. ej. la cubana) de portadores de alelos normales no patogénicos. B) *Crosstalk* o diafonía que explica como contribuyen los alelos normales con un haplotipo predisponente, al origen de variantes patogénicas causantes de la SCA2 u otras enfermedades. C) Los alelos patogénicos mediante contracciones, generan variantes de penetrancia reducida, y estos eventos pueden revertir hasta convertirse en variantes pre-mutadas. Estas últimas también pueden originarse por la expansión del CAG a partir de variantes predispuestas con 22CAG y el haplotipo CC, el cual está predispuesto a perder interrupciones de CAA. La expansión progresiva del CAG pudiera representar una mejora en la función de la proteína ataxina-2. El copyright© de esta figura es de José Miguel Laffita Mesa 2014.

Anexo 2

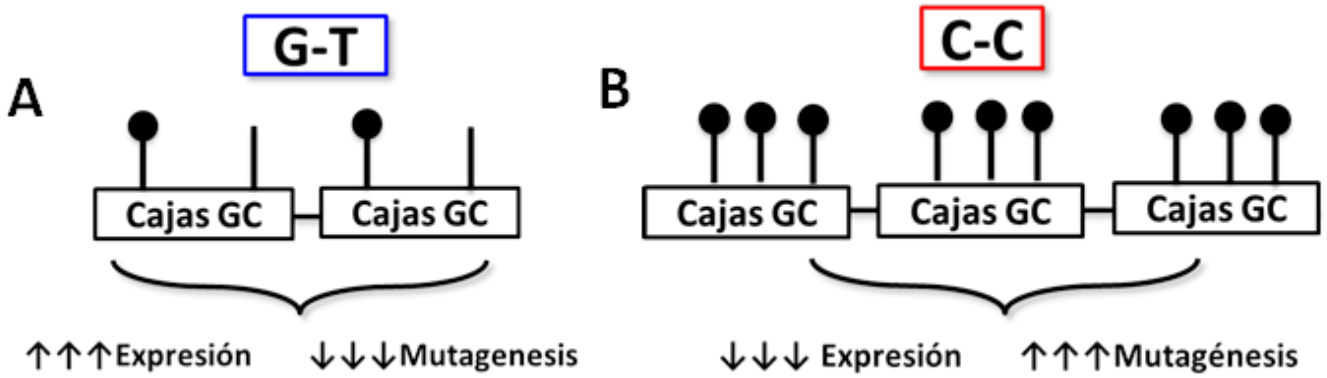


Figura 3. Relación entre el haplotipo y el potencial de metilación del gen *ATXN2*. A) A nivel de la secuencia promotora del gen *ATXN2*, el haplotipo G-T equivale a la pérdida de cajas GC, y por tanto tiene menos potencial de metilación. La consecuencia sería menos control sobre la activación del gen y por tanto menos expresión. A la vez este estado de menos metilación favorece la estabilidad de la expansión de CAG que esta adyacente. B) En cambio, el haplotipo C-C equivale a ganancia de cajas GC las que pueden metilarse, y representa un mayor potencial de metilación. Consecuentemente, esto inactivaría al gen y se expresaría menos ataxina-2. Sin embargo, este estado hipermetilado desestabiliza la secuencia repetitiva del CAG que está a 200pb y se tornaría inestable. El copyright© de esta figura es de José Miguel Laffita Mesa 2014.

Anexo 3

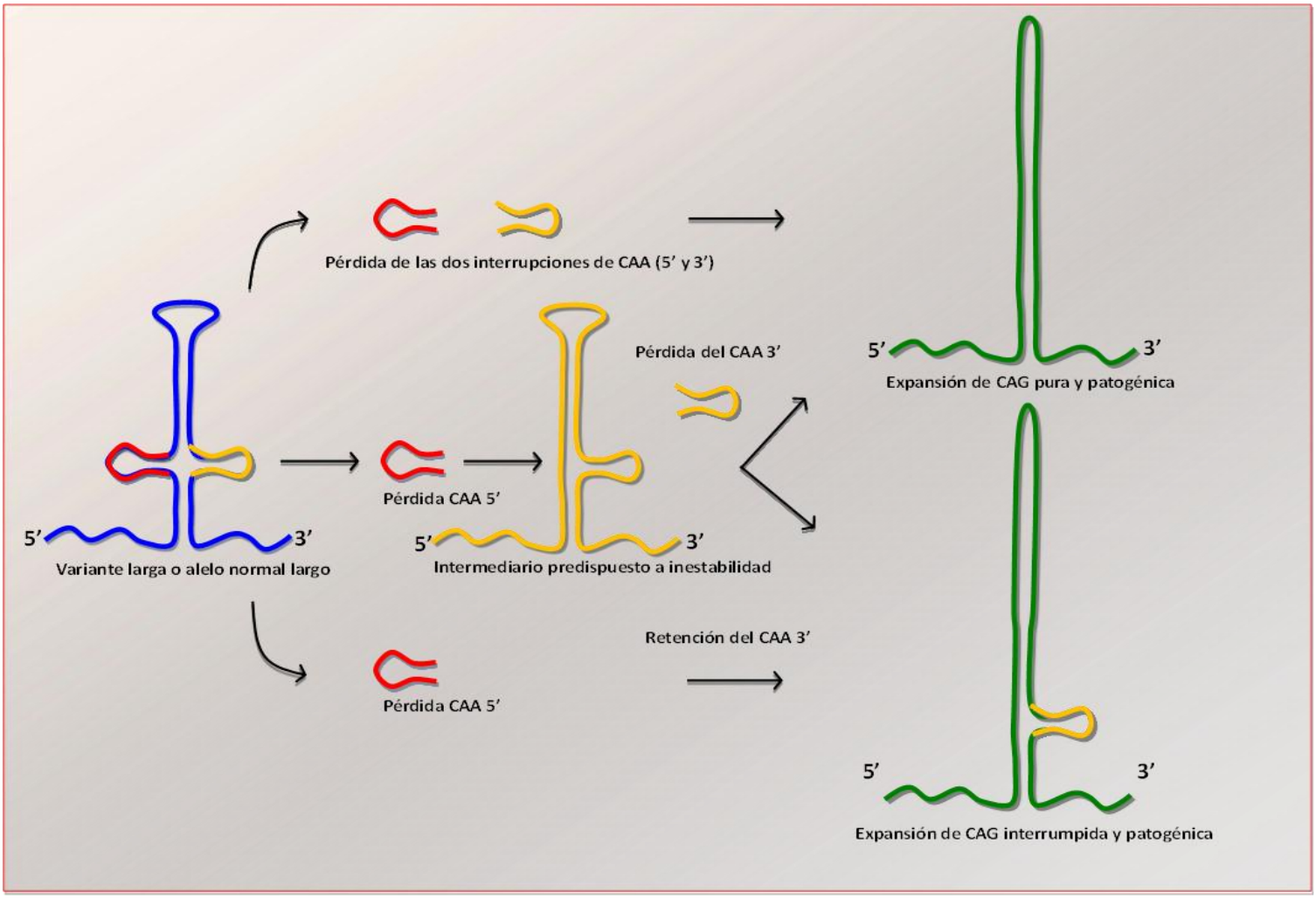


Figura. 4. Mecanismos mutagénicos del segmento de CAG en el gen *ATXN2*. Cada una de las rutas mutagénicas involucra la formación de estructuras no convencionales del ADN, con pérdida de tripletes de CAA del extremo 5' (lazo rojo) como generadores del crecimiento de CAG. Para esto proponemos la existencia de un intermediario predispuesto a inestabilidad genética que puede llevar a la pérdida total de todos los CAA o la retención de del triplete del extremo 3' (lazo amarillo). El copyright© de esta figura es de José Miguel Laffita Mesa 2014.

Anexo 4

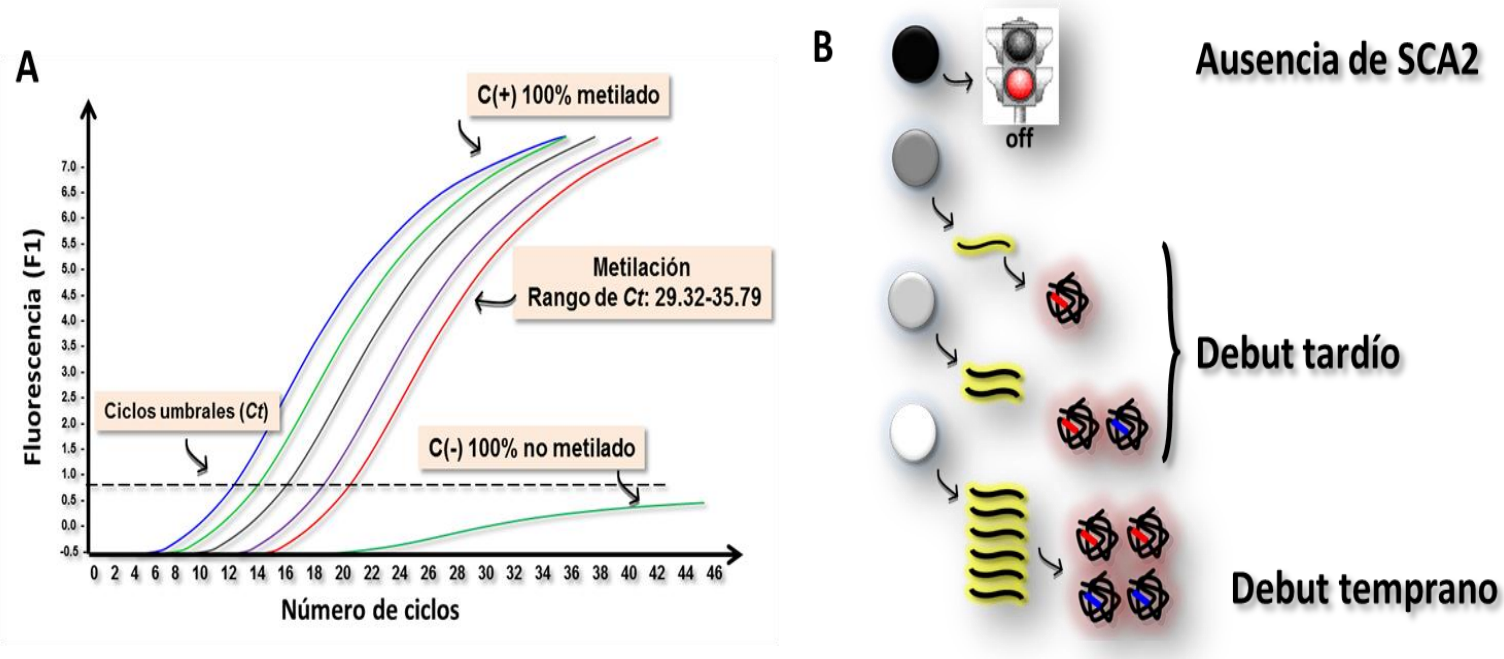


Figura 5. Metilación variable y modulación del debut y fenotipo de la SCA2. A) Registro representativo de una corrida típica de PCR en tiempo real correspondiente a la reacción de metilación para el promotor *ATXN2*. Se señalan con rótulos las reacciones de los controles positivos y negativos. También se indican los valores de ciclo umbral (*Ct*) de cada una de las muestras y su rango de variación en varias muestras de ADN de pacientes cubanos con SCA2. B) Influencia de la metilación variable sobre los niveles de expresión del gen *ATXN2* y sobre la expresión fenotípica, específicamente sobre el debut de la SCA2, a mayores niveles de metilación no aparece la enfermedad, ralentizándose o apareciendo más temprano a medida que disminuye la marca epigenética. El copyright© de esta figura es de José Miguel Laffita Mesa 2014.

Anexo 5

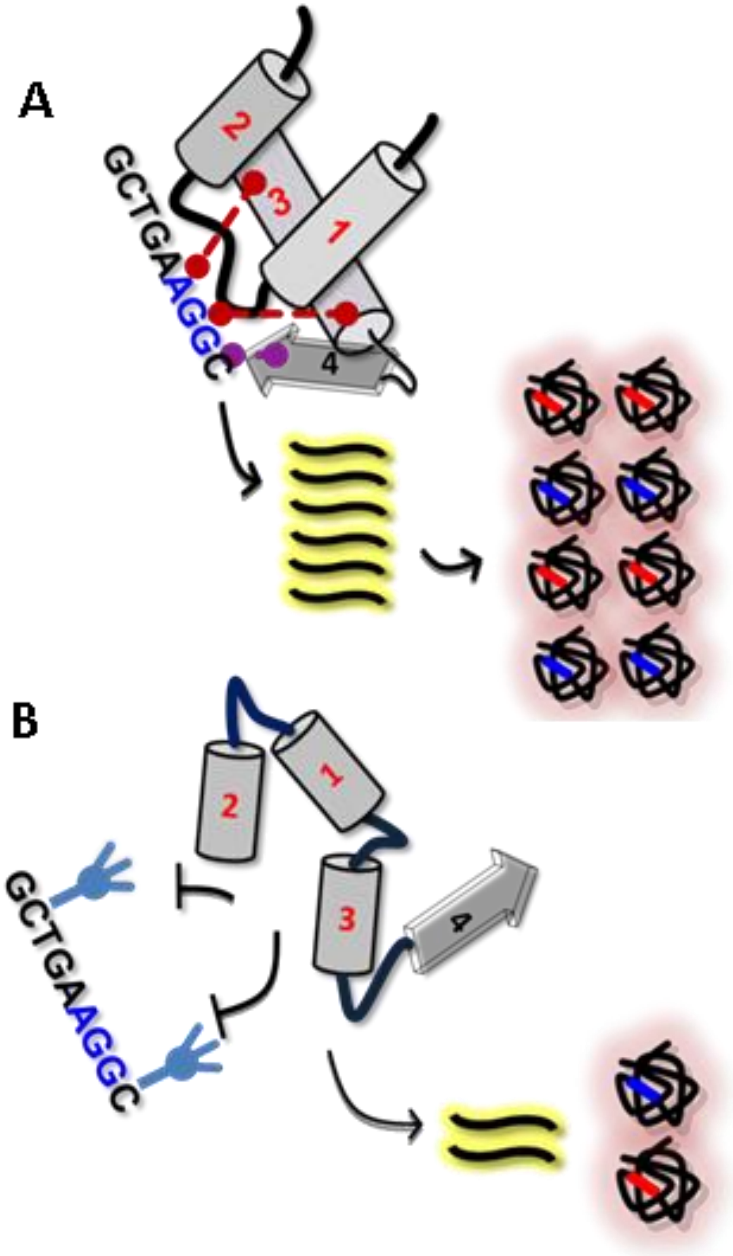


Figura 6. Conjugación de ETS1 a su elemento secuencia en el promotor *ATXN2* e influencia de la metilación sobre la expresión. A) Tres aminoácidos determinan la conjugación de ETS1 a su secuencia en el ADN y están localizados en la hélice 3. Estos residuos son dos Argininas y una Tirosina las cuales se unen cercanas o directamente a sitios CpG. B) Efecto de la metilación sobre la unión de ETS1. Al estar metilado alguno de los sitios CpG, se excluye a ETS1 de esta secuencia, la cual actúa como portero molecular. Cuando esta metilado no hay conjugación de ETS1, sufriendo un cambio conformacional. Mientras que cuando esta demetilado, se une y activa la expresión del gen cadena abajo (Wei et al., 2010, Gaston & Fired., 1995). El copyright© de esta figura es de José Miguel Laffita Mesa 2014.

Anexo 6

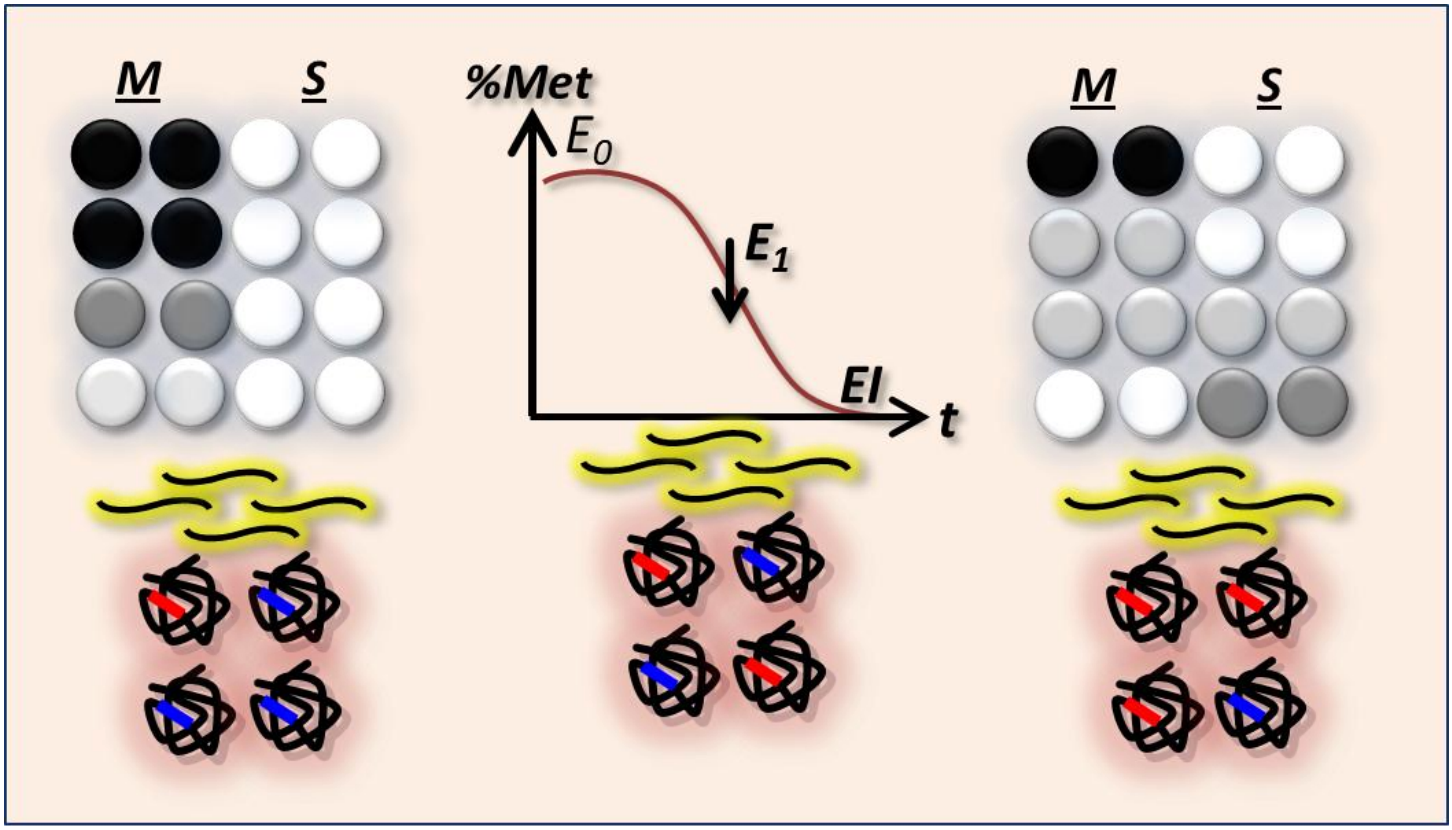


Figura 7. Patogénesis y metilación de la variante mutante de *ATXN2*. La patogénesis es el resultado de la pérdida de la capacidad de mantener el equilibrio (metilostasis) entre alelos mutantes metilados y por tanto inactivos (*M*) y los salvajes activos (*S*). Estos estados a través de la vida pueden variar. En los estadios iniciales (E_0), los mecanismos de inactivación mediante hipermetilación del alelo mutante son más eficaces, y la cantidad de ataxina-2 mutante (roja) soluble no alcanza tasas críticas para la patogénesis. En etapas posteriores de la vida (E_1) se compensan la expresión de ambas variantes de ataxina-2 estableciéndose un estado de metilostasis, que protege de la patogenicidad de la ataxina mutante, y a la vez se suple la función de ataxina-2 salvaje (azul). Este estado se pierde durante la adultez, la vejez, por estrés celular o fallos en la metilación de mantenimiento o la *de novo*, sobrepasando umbrales críticos y llevando al debut de los síntomas de SCA2 (*EI*). Los círculos representan los estados de metilación de los alelos mutantes y salvajes del gen en poblaciones neuronales sensibles a la ataxina-2 mutada (negros: alelos mutantes metilados, y blancos: no metilados). La intensidad del color indica distintos niveles de metilación (negro 100% hipermetilado, blanco 100% no-metilado, grises metilación intermedia). El copyright© de esta figura es de José Miguel Laffita Mesa 2014.