

APLICACIONES NO CONVENCIONALES DE MARCADORES BIALÉLICOS (SNPs) EN GENÉTICA FORENSE. *UNCONVENTIONAL APPLICATIONS OF BIALLELIC MARKERS (SNPs) IN FORENSIC GENETICS.*

M.FONDEVILA¹, C. PHILLIPS¹, MV LAREU¹.

RESUMEN:

Revisaremos en este artículo la aplicación de los Single Nucleotide Polymorphisms o SNPs, polimorfismos simples de secuencia, en genética forense dirigida a dos de los principales escenarios de uso en este campo. La predicción del origen biogeográfico de una muestra desconocida y la predicción de rasgos físicos a través del análisis del genoma de muestras biológicas. Se incluirá la base biológica de cada aplicación, la adecuación de los SNPs a la tarea y un breve repaso de la situación actual en el campo forense.

PALABRAS CLAVE: Toma de muestras, intervención corporal, muestras de referencia, vestigios biológicos, consentimiento informado, cadena de custodia.

ABSTRACT:

SNPs have been used as forensic genetics markers even since the very beginning of the field's history. However it has not been until recently, that SNPs have reappeared on the scope. Tools, apparently more powerful, set aside SNP typing into secondary tasks. Currently, the situation is quite different. New applications of SNPs on which these markers appear as the optimal tool, have been developed and are being applied to real criminal casework. We will review on this article two of such applications that perfectly illustrate the current use of SNPs in forensic genetics: Bio-geographical origin prediction of unknown samples and phenotypical trait prediction."

KEY WORDS: DNA, forensic genetics, SNPs, biogeographic origin, physical features, EVCs, phenotype, genetic of populations.

CONTACTO: Manuel Fondevila Álvarez, e-mail: mafondevila@hotmail.com, Instituto de Ciencias Forenses, Facultade de Medicina, Rúa San Francisco s/n 15782, Santiago de Compostela, A Coruña.

1. INTRODUCCIÓN.

El trabajo en genética forense en la actualidad está dominado por el genotipado de polimorfismos STR. El elevado poder estadístico que aporta su análisis, acompañado de la sencillez de sus protocolos de genotipado, estandarización mundial y el desarrollo de kits comerciales de amplificación se bastan para explicar este hecho. Los SNPs, por otro lado, parecen una herramienta más reciente (los primeros ensayos de genotipado forense de SNPs para identificación humana fueron publicados en fechas tan próximas como enero de 2006 [1]) y sin embargo han estado presentes en la genética forense desde su primer momento.

Sir Alec Jeffreys aplicó por primera vez el análisis de polimorfismos de ADN a la resolución de un caso de investigación criminal, mediante

la comparación de patrones de fragmentos de restricción en genomas completos, el "DNA fingerprinting" o huella genética. La variación en la longitud de esos fragmentos viene dada por la presencia de secuencias diana específicas para una serie de enzimas de restricción, la presencia o ausencia de una de estas dianas en una región concreta del genoma viene dada por la presencia de un SNP en ese lugar, cuya variabilidad determina que uno de estos enzimas reconozca o no esa secuencia como diana.

El "DNA fingerprinting" de Sir Alec Jeffreys, está en la actualidad ampliamente superado. Los SNPs, como polimorfismos de secuencia, requieren de protocolos de genotipado más complejos [2] que la combinación directa PCR-Electroforesis de los STRs. Por otro lado, los SNPs como polimorfismos mayoritariamente bialélicos son capaces únicamente, marcador a

¹ Instituto de Ciencias Forenses de la Universidad de Santiago de Compostela.

marcador, de suministrar solo una parte de la información que se podría obtener del genotipado de un solo locus STR [3]. Cálculos realizados por P.Gill, arrojan la estima de que sería necesarios del orden de 30-40 SNPs con máxima heterocigosidad (frecuencias alélicas de 0,5 en ambas variantes), para igualar la potencia estadística del tipado de 15 STRs[3].

Con un panorama como el descrito, parecería que la aplicación de SNPs al trabajo forense no sería interesante en absoluto. Pese a todo, los SNPs si ofrecen una serie de propiedades específicas que los hacen muy interesantes para un conjunto de aplicaciones forenses en las que el genotipado de STRs no garantiza el éxito. Por todo ello, podemos considerar el tipado de SNPs como una herramienta complementaria a los STRs o de aplicación única, cuando dichos STRs fallan en su análisis.

A) PROPIEDADES DE LOS SNPs.

Los SNPs, presentan una serie de propiedades, compartidas en cierto grado por otros marcadores bialélicos como los InDels, que resultan de gran interés en el campo de la genética forense:

1. Son el tipo de variación más común del genoma, se ha estimado que por cada 100 o 1000 bases del genoma humano, existiría una polimórfica [4]. En teoría, se dispondría de un amplio número de marcadores para su estudio, informativos para diversas aplicaciones. Este gran número de polimorfismos a la disposición del investigador forense, permite la adecuada selección de marcadores con una distribución de frecuencias alélicas apropiada para tareas específicas (a revisar en el presente texto), y secuencia flanqueante óptima para la amplificación, sin el problema que presentan factores como el desequilibrio de ligamiento entre marcadores.

2. Los SNPs son marcadores extremadamente estables que tienden a permanecer inalterados de generación en generación. Presentan tasas de mutación tan bajas como

2×10^{-8} por generación y meiosis [5] como valor medio. Un valor extremadamente bajo cuando se compara con la tasa media de mutación de 10^{-3} de los loci STR. Este hecho, presenta un doble interés en genética forense, por un lado resulta de extrema utilidad en casos de paternidad cuestionada (y entre estos casos de especial relevancia en casos de pedigrí complejo) y por otro lado, favorece la aparición de una variabilidad interpoblacional en las frecuencias alélicas.

3. Son la forma más simple de polimorfismo de ADN y por tanto sus métodos de detección y genotipado, aún dentro de su relativa complejidad, son susceptible de automatización y permiten el diseño de reacciones multiplex con un alto número de marcadores. Ofrecerían así la posibilidad de optimizar la relación entre el gasto de muestra y información obtenida en el análisis. Dada esta propiedad, no parece ya una seria limitación el hecho de la baja informatividad individual de los SNPs, ya que pueden ser incorporados en número suficientemente alto en una única reacción de genotipado como para superar la potencia estadística de 15 STRs de identificación individual [1].

4. El polimorfismo es una única base. Esto permite diseñar reacciones de amplificación PCR con productos con longitud reducida al mínimo, de forma que en condiciones de ADN degradado, en las que el genoma puede estar en avanzado estado de fragmentación, todavía puedan ser analizados con garantía de éxito [6-9].

B) LA APLICACIÓN DE LOS SNPs EN GENÉTICA FORENSE.

En base la combinación de todas o varias de estas propiedades, existen una serie de escenarios de aplicación de los SNPs en genética forense, en los que se perfilan como la herramienta más adecuada a nuestra disposición. Entre estos escenarios encontramos tareas tales como: los casos de ADN extremadamente degradado, la predicción

del origen biogeográfico o predicción de características fenotípicas.

En un principio, el interés en el genotipado de SNPs en casuística forense se centró en su aplicación a casos de ADN degradado [1]. Tras la muerte o la deposición de una muestra biológica, comienzan a actuar sobre la molécula de ADN una serie de procesos tanto endógenos como exógenos que conducen a la degradación progresiva y acumulativa de las biomoléculas [10-12]. Es común encontrar tales muestras en la casuística forense. En este tipo de muestras es frecuente observar que la eficacia de amplificación se hace menor cuanto más pesado es el producto de PCR, es por ello que la tendencia en el trabajo orientado a muestras degradadas es la selección de marcadores con producto de PCR lo más corto posible. Así, se ha tratado de reducir a la mínima expresión el fragmento amplificado en reacciones multiplex de STRs, los denominados MiniSTRs. Pese a todo, los STRs al tratarse de polimorfismos de longitud multialélicos relativamente grandes presentan un límite en cuanto al número de marcadores que pueden ser introducidos en una reacción multiplex con un tamaño de amplicon reducido [13]. Aún contando con que podemos marcar varios STRs con diferentes fluorocromos de forma que la señal electroforética de cada uno no interfiera con la de los demás, incluso Minifiler solo cuenta con 4 marcadores STR por debajo de 150pbs [6-9].

Es en este caso cuando los SNPs realmente ofrecen su ventaja:

- Al ser polimorfismos de una única base, el tamaño de amplicon puede ser reducido a fragmentos menores de 90pb sin demasiada dificultad, dado que podemos acercar los primers de PCR uno al otro (en teoría) hasta que solo acoten la base polimórfica[1,14].
- Dado que la reacción de genotipado en SNPs es independiente de la PCR de enriquecimiento, no tenemos la limitación presente en STRs en cuanto al número de marcadores con el mismo tamaño de fragmento amplificado. Todos los marcadores de una reacción de PCR de

SNPs pueden ser amplificados con fragmentos menores de 90pbs sin riesgo de interferencia.

- La sencillez del genotipado permite realizar el análisis paralelo en una sola reacción de un elevado número de SNPs con el consiguiente ahorro de material limitado.

En las siguientes páginas se revisarán las aplicaciones que surgieron a medida que el interés del campo forense en los SNPs se iba desarrollando, la predicción de origen biogeográfico y de características fenotípicas.

2. PREDICCIÓN DEL ORIGEN BIOGEOGRÁFICO.

Esta aplicación se basa en la existencia en el genoma de todas las especies de una serie de polimorfismos cuya variabilidad se encuentra entre grupos de individuos o poblaciones (variantes alotípicas), y no entre los individuos de una misma población (variantes idiotípicas). De esta forma si conocemos la frecuencia de aparición de una serie de estas variantes en poblaciones de la especie humana. El estudio de las variantes presentes para un conjunto de estos polimorfismos en el genoma de un individuo concreto, podría entonces ayudarnos a inferir la población con mayor probabilidad de haber sido el marco genético en el que se desarrollara dicha combinación de variantes en particular.

La aplicación de estas competencias a la genética forense se centrarían en aquellos casos en los que aunque se ha hallado una muestra biológica con ADN útil, pero no se cuenta con ninguna muestra de referencia con la que hacer una identificación con marcadores de identificación humana (ya sean sospechosos de haber cometido un crimen o en casos de identificación de víctimas). Las aplicaciones en casos de desapariciones en masa y grandes catástrofes son comunes, a la hora de hacer un screening de posible nacionalidad de las víctimas. Pongamos como ejemplo la identificación de cadáveres de pasados conflictos militares, como las dos guerras

mundiales, en las que bajas de diversas nacionalidades se encuentran en la misma área geográfica.

A) BASE BIOLÓGICA.

Es relativamente frecuente encontrar polimorfismos de esta clase en el genoma de nuestra especie, debido a nuestra historia evolutiva. Pertenece a una especie con una distribución geográfica muy amplia, ya desde los orígenes del Homo Sapiens anatómicamente moderno [15]. Esta amplia distribución desde tiempos remotos, en poblaciones reducidas con escaso contacto entre ellas es la base de la aparición de la variación que nos interesa.

Es especies con estructura poblacional como la que exhibía el Homo Sapiens durante gran parte de su recorrido, la combinación de la acción de una serie de fuerzas evolutivas aleatorias sobre el ADN no codificante lleva a la aparición de diferencias alotípicas en las frecuencias alélicas como las que nos interesan. Las fuerzas evolutivas que actúan es estas zonas son a grandes rasgos: 1. La migración, 2. la deriva genética y 3. la mutación.

1. La migración actuaría como un factor homogenizador de las frecuencias, introduciendo cambios anteriormente inexistentes en el pool génico de la población receptora, o aumentando su frecuencia. La deriva y la mutación (en menor medida), al actuar al azar, pueden ocasionar cambios en las frecuencias génicas en cualquiera de los sentidos, bien aumentando la diferencia entre poblaciones o bien reduciéndolas.

2. La deriva génica es un factor que se debe al emparejamiento al azar de los individuos de la población (el emparejamiento es estocástico al menos en cuanto a estos marcadores, ya que no tienen influencia alguna en el suceso), es decir, la muestra de la población que se toma para formar las variantes de cada polimorfismo en la siguiente generación incluye un número aleatorio de moléculas con cada una de las

variantes, con lo que bien se puede aumentar la frecuencia de un determinado alelo o bien eliminar una variante del pool génico de la población. En general la deriva génica actuará aumentando la varianza entre grupos y disminuyendo la varianza intragrupal.

3. La mutación ocurre con una determinada probabilidad en cada posición nucleotídica, aunque dicha probabilidad suele ser muy baja, de forma que es muy probable que por mutación aparezca un cambio en una población y no en otras aumentando la variación interpoblacional, y con menos probabilidad que el cambio introducido sí esté presente en otras poblaciones, reduciendo la diferencia. En general, la mutación y la migración introducen cambios en el pool génico de una población (la primera aumentando heterogeneidad y la segunda la homogeneidad interpoblacional) y la deriva modifica las frecuencias aleatoriamente, fijando o eliminando los cambios introducidos.

B) LOS SNPs COMO MARCADORES DE ORIGEN BIOGEOGRÁFICO.

Estas fuerzas evolutivas afectan, por supuesto, a todo el genoma, por mucho que este sea más aparente en las regiones no selectivas. Así, las variantes haplotípicas de la región hipervariable del ADN mitocondrial y el cromosoma Y, siguen un patrón de marcada variabilidad interpoblacional bien conocido por la comunidad científica y pueden ser aplicados a la predicción de origen biogeográfico con cierta garantía de éxito [16-17]. Los loci STR, son parte fundamental del ADN no selectivo y sobre sus variantes también operan estas fuerzas evolutivas. Rosemberg y colaboradores en múltiples publicaciones demostraron la capacidad de los STRs [18] para predecir el posible origen biogeográfico, así como otra serie de trabajos publicados por nuestro grupo en los últimos años [19-20].

Ahora bien, el uso de SNPs autosómicos se adapta con mucha mayor facilidad a esta tarea [14,21-22]. Sus propiedades los hacen una opción idónea que aúna las ventajas tanto del

ADN mitocondrial como las de los STRs.

Tanto la molécula de ADN mitocondrial como el cromosoma Y, presentan una amplísima variabilidad interpoblacional, bien documentada que puede ser explotada con éxito para predecir el origen biogeográfico de perfiles de ADN. Por otro lado, el ADNmt es un marcador muy resistente a la degradación portmortem (si no el más resistente), lo que constituye una importante ventaja en el campo forense. Sin embargo estos marcadores debido a la ausencia de recombinación, presentan la limitación de que no permite asignar un perfil a un individuo si no a un linaje extenso. La escala de tiempo que cubren los cambios en los polimorfismos de Cromosoma Y y ADNmt es mucho mayor que la de los marcadores autosómicos, lo que acompañado de su transmisión en bloques haplotípicos a través de familias extensas nos lleva a discutir sobre si la aplicación de estos datos refleja o no el estado actual de los individuo, si mas bien deberíamos hablar del origen biogeográfico de su linaje, de la ancestralidad del linaje, no del individuo actual. En este sentido, quizás los términos ancestralidad y origen biogeográfico no sean intercambiables, sino derivados de evidencias diferentes.

Por su lado, la variabilidad de STRs refleja el estado actual del individuo, sin embargo, presentan una dificultad derivada de la misma ventaja que los hace adecuados para la identificación individual, multiallelismo e hipervariabilidad. Pese a que existe un sesgo apreciable en las frecuencias interpoblacionales para algunos de los alelos de ciertos STRs [19-20,23] que puede ser explotado para la predicción del origen biogeográfico, casi todos los alelos frecuentes están presentes en ambas poblaciones, si no en todas, lo que reduce su informatividad en cuanto a origen biogeográfico. La elevada tasa de mutación de estos marcadores, colabora en la estabilización de las frecuencias alélicas entre poblaciones.

Los SNPs, por su lado, están presentes en grandes números en los cromosomas autosómicos, por lo que la recombinación meiótica en cada generación, garantiza que una determinada combinación de alelos en un perfil reflejará fielmente el estado actual del individuo

que las porta, en cuanto a origen poblacional. Su extremadamente baja tasa de mutación impide que la mutación recurrente de forma independiente en diferentes poblaciones establezca la variabilidad entre poblaciones. En este caso, su limitado número de variantes posibles por locus juega a nuestro favor, de forma que si existe una variante representativa de la población, es bastante más probable que la porte un gran porcentaje de los individuos de ese grupo, en contra de lo esperado en sistemas multialélicos. Además, los SNPs, como ya fue discutido en las páginas anteriores, se adaptan muy bien al trabajo con muestras degradadas, siendo el marcador genético más resistente, solo por debajo del ADNmt.

Por ello, el genotipado de SNPs autosómicos se perfila como la herramienta ideal para la predicción de origen biogeográfico.

C) APLICACIÓN DE LOS SNPs.

Tomando una serie de posibles orígenes biogeográficos, por ejemplo Europeoide, Extremo oriente y África subsahariana como los utilizados en nuestro trabajo original en este tema de 2008 [14], se hará una búsqueda de marcadores en bases de datos. Se tratará de seleccionar una serie de marcadores en los que uno de los alelos tenga una frecuencia de aparición diferente para una de las poblaciones con respecto a las otras dos. Esta diferencia haría de este marcador concreto, un predictor de origen poblacional asociado al grupo humano para el que muestra una frecuencia alélica diferencial. Buscaríamos entonces una serie de marcadores predictivos para cada una de las poblaciones incluidas en el cálculo planeado, que se incluirían en una reacción de genotipado multiplex. Estos SNPs podrían ser:

- SNPs específicos de población: Aquellas posiciones del genoma en las que sólo existe variabilidad en una población concreta, siendo las restantes, monomórficas.
- AIMS propiamente dichos: Aquellas posiciones que siendo variables en toda población, presentan una baja heterocigosidad, siendo el alelo mayoritario

diferente entre los grupos en estudio.

- SNPs no binarios: Existen SNPs en el genoma con un tercer o incluso cuarto alelo. Estos pueden ser un caso especial de las otras clases comentadas, en las que es posible que cada uno de los múltiples alelos de la posición polimórfica sean indicativos de uno de los orígenes en la hipótesis.
- Variantes fijadas: Incluimos aquí posiciones invariables dentro de cada población, pero en las que por deriva, se ha fijado una variante diferente en cada grupo Fig.1. Ejemplos de SNPs con variación alotípica

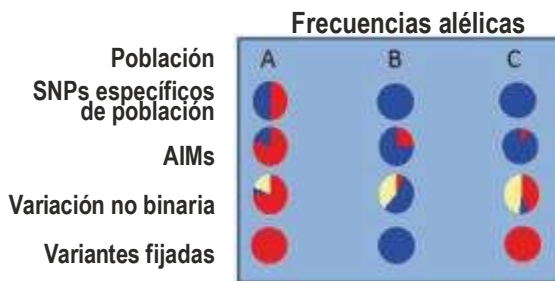


Fig.1. Ejemplos de SNPs con variación alotípica

Así, tendremos una serie de SNPs informativos de origen poblacional dirigidos a una predicción concreta, cuyo análisis combinado nos ayudaría a calcular de forma estadística la población de origen más probable de entre los grupos de referencia incorporados a nuestra hipótesis inicial [14,21]. Evidentemente esta predicción debe realizarse mediante un cálculo matemático con el que podamos obtener una LR entre la probabilidad de pertenencia del perfil a cada una de las poblaciones contempladas en la hipótesis inicial (toda vez que existe una variabilidad intrapoblacional paralela a la interpoblacional para muchos de estos marcadores que hará que diferentes individuos de la misma población clasifiquen con una LR diferente según combinación alélica individual).

A la hora de llevar a cabo esta tarea, y a fin de conseguir una selección óptima de marcadores que nos permita hacer el mejor uso de los SNPs

como herramienta de predicción de origen biogeográfico, se debe buscar un equilibrio de información para cada uno de los orígenes incluidos en la hipótesis, de forma que no se incurra en ningún sesgo hacia ninguno de ellos, a riesgo de sobreestimar la LR asociada a esos grupos [14,24].

D) DIFICULTADES DE LA PREDICCIÓN DE ORIGEN BIOGEOGRÁFICO.

Hemos revisado brevemente el efecto de la variabilidad intrapoblacional sobre los cálculos. Esta variabilidad, no resulta dificultosa en los casos en que las poblaciones incluidas en la hipótesis inicial estén filogenéticamente alejadas. En este caso, independientemente de la variabilidad individual en cada población, el rango de LR's posibles partirá de valores muy elevados en ambos extremos.

Sin embargo, no podemos contemplar las poblaciones humanas como elementos alejados o independientes unos de otros. Lejos de ser unidades discretas, las poblaciones humanas tienen una distribución continua, cuyos límites se diluyen.

Es más, en la actualidad gracias a movimientos masivos de población, voluntarios o no, facilitados por la mejora en los medios de transporte, muchos grupos poblacionales conviven en el mismo espacio geográfico. Esto conduce a que estos grupos se entremezclen diluyendo la divergencia entre ellos. Se denomina a la situación resultante de un flujo génico entre dos poblaciones, admixture o mezcla poblacional. Esta situación puede llegar a introducir una cantidad significativa de ruido en la clasificación en caso de que las poblaciones de referencia en la hipótesis inicial presenten un cierto grado de admixture [24-26].

Esta situación se puede compensar de dos maneras: de forma estadística, limitando los cálculos únicamente a las poblaciones fuente del admixture [26] o bien aumentando el poder estadístico del test con el genotipado de un mayor número de marcadores. Esta estrategia ha sido estudiada por nuestro grupo de forma sistemática [21,24-26], lo que nos ha permitido

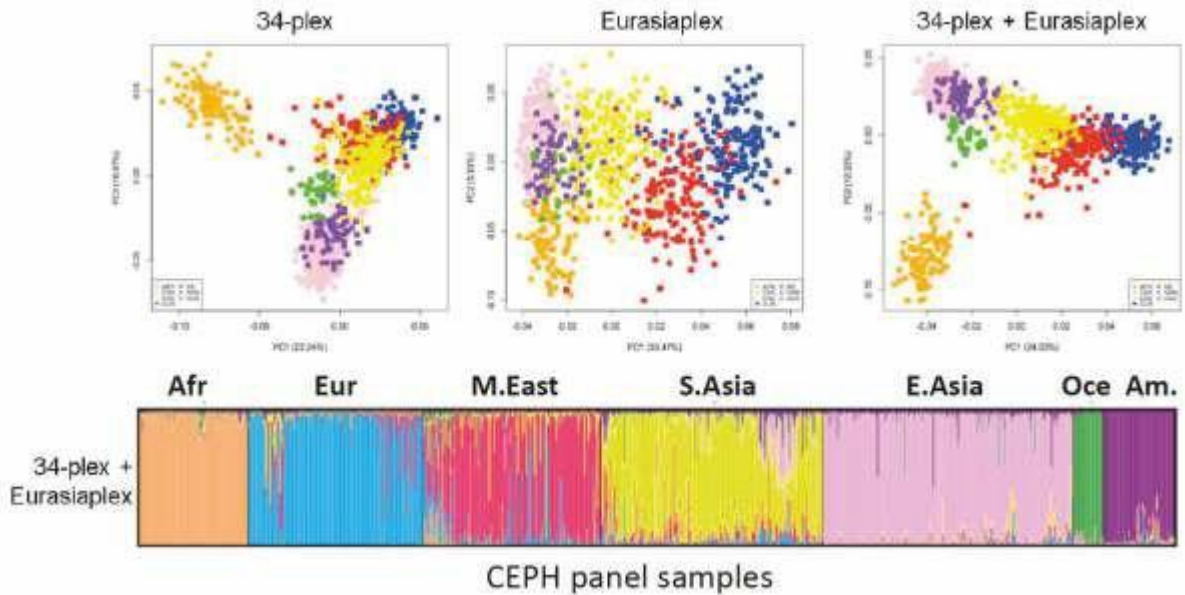


Fig .2. Resultados del análisis con el software Structure y de PCAs de cerca de 1000 muestras de distribución mundial (CEPH panel) con 34 SNPs de predicción de origen biogeográfico generalista, un nuevo set de SNPs de variabilidad alotípica entre grupos contenidos en Eurasia, y ambos combinados. La agrupación de estos dos sets, nos permite no solo detectar la divergencia entre grupos muy alejados geográficamente, sino también entre aquellas poblaciones eurasiáticas genéticamente próximas pese al elevado grado de admixture entre ellas. "Phillips C. FSI Genet, 2013 ;7(3):359-66".

no solo aumentar el número inicial de orígenes geográficos incluidos en nuestros cálculos, si no también aumentar la resolución de estos para llegar a un cierto grado de predicción de origen dentro 2 de los grandes grupos iniciales [21-22,25].

En los últimos años, hemos tenido la oportunidad de aplicar nuestros multiplexes de SNPs a la predicción de origen biogeográfico en casos reales de investigación policial. Consiguiendo resultados exitosos en todas las ocasiones independientemente de las dificultades encontradas, incluso cuando se requirió la predicción de origen entre poblaciones muy próximas [25]. Con este bagaje, debería quedar despejada toda duda sobre la adecuación de los SNPs a la predicción forense de origen poblacional en muestras desconocidas.

3. PREDICCIÓN DE CARACTERÍSTICAS FENOTÍPICAS.

La predicción de rasgos fenotípicos o EVCs (External Visible Characteristics) es la aplicación más novedosa de los SNPs al campo forense. Se encuentra en la actualidad en estudio y solo un pequeño número de rasgos de interés forense pueden ser predichos con seguridad en la actualidad, pese que se está llevando a cabo una intensa labor investigadora en esta área en numerosos grupos [26-28].

Esta aplicación trata, de forma paralela a la predicción de origen biogeográfico ya tratada, de hallar información relativa a un individuo completamente desconocido, en base a un resto biológico que nos de acceso a la variabilidad de su genoma. En este caso se estudiarán aquellas posiciones polimórficas cuya variabilidad esté asociada estadísticamente al espectro de un rasgo físico cuantificable.

A) BASE BIOLÓGICA.

Es un hecho ampliamente conocido como parte de la cultura general, es incluso una afirmación intuitiva. Los rasgos físicos (fenotípicos) están regulados por las variantes presentes en las secuencias codificantes del ADN. De hecho, las características físicas, entre las que están aquellas de interés forense, como la altura, peso, pigmentación e incluso la estructura ósea y rasgos faciales están regulados por una serie de polimorfismos génicos heredables y un efecto ambiental variable.

La mayor parte de estas variantes, los alelos de esos genes (que se traducen en variantes proteicas) estarán definidas por una particular combinación de bases posibles de una serie de SNPs contenidos tanto en secuencias codificantes como reguladoras de la expresión génica.

La aplicación del tipado de SNPs a esta tarea pasaría a través del análisis de una serie de posiciones polimórficas en el genoma, cuyas variantes estarían asociadas (no se requiere que la variante a tipar sea codificantes, solo estar en desequilibrio de ligamiento intenso con las posiciones definitorias de un alelo) en cierto grado con una variante en particular de un rasgo fenotípico.

Conociendo la frecuencia con la que una variante de un SNP se encuentra presente cuando un rasgo es observado, permitiría hacer una predicción a través de un cálculo matemático objetivo en base la conjunción de una serie de variantes observadas.

A primera vista, parece una tarea fácil, pero no tardan en surgir dificultades. La mayor parte de los ECVs presentan una variación continua, no un número pequeño de variantes discretas fácilmente identificables y relacionables con una variante genotípica concreta. Además, cada rasgo individual está regulado por la actividad conjunta de un gran número de secuencias génicas, cada una de ellas con un pequeño efecto en el fenotipo final, y cada uno

de ellos conteniendo un número significativo de alelos (aún mas, cada variante funcional del gen puede ser definida por un cierto número de combinaciones de SNPs diferentes). Para complicar la situación aún más, una misma secuencia puede afectar múltiples caracteres al mismo tiempo y ver su contribución al fenotipo modulada o anulada por otras secuencias.

La investigación en este tema se centra en desenmarañar el comportamiento de aquellas características de interés forense cuya variabilidad descomponerse en una serie de variantes más o menos discretas reguladas por un número relativamente pequeño de secuencias. Un ejemplo de esto son: la pigmentación del iris principalmente, y cercanamente relacionados la pigmentación del cabello y la piel [26-28].

B) APLICACIÓN DE LOS SNPs.

La aplicación de marcadores genéticos como los SNPs a la predicción del fenotipo en base al análisis de la variabilidad genotípica, es conceptualmente semejante a la ya revisada en el apartado de origen poblacional.

Se debe, para cada rasgo establecer una población de individuos que comparten una misma variante del EVC, del mismo modo que lo hacíamos con los posibles orígenes biogeográficos. La correcta definición de cada variante del ECV en estudio es clave para el éxito de la predicción.

Se debe seleccionar un panel de SNPs análogo a los explicados en el apartado anterior, cuya variabilidad presente una componente alotípica entre los grupos de individuos de referencia para cada variante del EVC.

Como en el caso anterior, se precisa de una utilidad matemática en la que se calcularían las LR de pertenencia a cada grupo de referencia en contra de los restantes.

C) LA PREDICCIÓN DE VARIANTES DE PIGMENTACIÓN DEL IRIS.

Como se anticipaba, el rasgo fenotípico cuya predicción se encuentra más avanzada es la coloración del iris. Se trata de un rasgo de variabilidad continua y compleja, pero que permite agrupar las variantes por colores bien diferenciados (ignorando la microvariabilidad responsable del tono concreto dentro de cada coloración). El metabolismo de la melanina en el iris y la implicación de variantes estructurales del tejido de la coroides es bien conocido, lo que permite realizar una selección de SNPs asociados al EVC, dirigida a las secuencias implicadas en la regulación del carácter.

Varios grupos de investigación han trabajado recientemente para conseguir la predicción más exacta de la coloración del iris [26-28].

El trabajo realizado con muestras de diferentes grupos de referencia con el ensayo desarrollado por nuestro grupo de investigación, nos llevó a definir tres grupos: Azules, Marrones e Intermedio (que incluye verdes, avellana y otras coloraciones que no pueden ser clasificadas como azules o marrones). Tanto en nuestros trabajos, como en los de grupos independientes, se seleccionaron los mismos SNPs de efecto mayor sobre el carácter de forma que la predicción presenta resultados similares hasta cierto punto. Adicionalmente, se añadieron SNPs de efecto menor, marcadores con menor asociación con el carácter, que modulan de forma sutil el efecto final, pero que resultaron cruciales para la predicción de variantes intermedias [26].

Como se puede ver en la imagen que acompaña al texto, en la que se muestra la progresión de la línea AUC que proporciona la adición de cada nuevo SNP de menor informatividad para cada uno de los posibles colores, la predicción de cada una de las variantes fenotípicas, varía. Al igual que el efecto sobre el valor final de cada nuevo SNP. Tenemos, por ejemplo una muy buena predicción de las variantes extremas, como el azul y el marrón, mientras que el grupo

intermedio es mucho más reducida. Esto se debe al hecho de que las variantes extremas están relacionadas con polimorfismos de secuencia que afectan al proceso de forma intensa en los primeros eslabones de la cadena metabólica mientras que las variantes intermedias deben su característica a efectos menores concatenados de una multiplicidad de SNPs sin efecto mayor. El caso concreto del SNP mayor rs12913832 es muy gráfico. Este marcador en solitario consigue predicciones de coloración azul del iris de más del 0.9 [26-28]. Esto se debe a que el alelo asociado a este único marcador irrumpe o limita la efectividad de los Fig.3- curvas AUC obtenidas con 18 SNPs relacionados con la pigmentación del iris, obtenida de Y. Ruiz (FSI:Genet 2013 Jan;7(1):28-40) procesos restantes casi al principio de la cadena metabólica de la melanina, dando lugar a pigmentaciones claras.

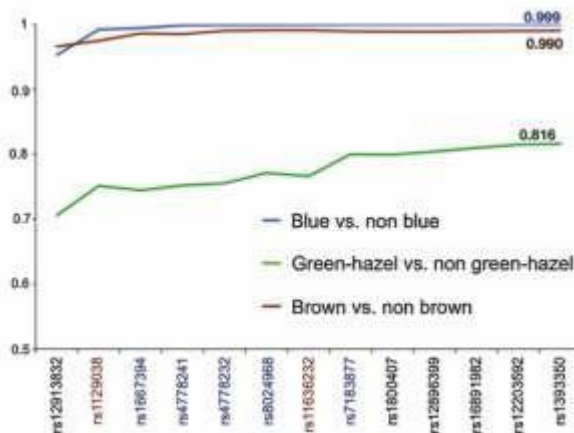


Fig.3. Curvas AUC obtenidas con 18 SNPs relacionados con la pigmentación del iris, obtenida de Y. Ruiz (FSI: Genet 2013 Jan; 7(1): 28-40)

Así pues, los trabajos realizados en pigmentación tanto del iris como de la piel o el cabello ofrecen resultados muy esperanzadores para el campo, y pueden ser aplicados con elevada garantía de éxito, si bien esta aplicación aún requiere estudio para poder aprovechar todas las posibilidades que ofrece.

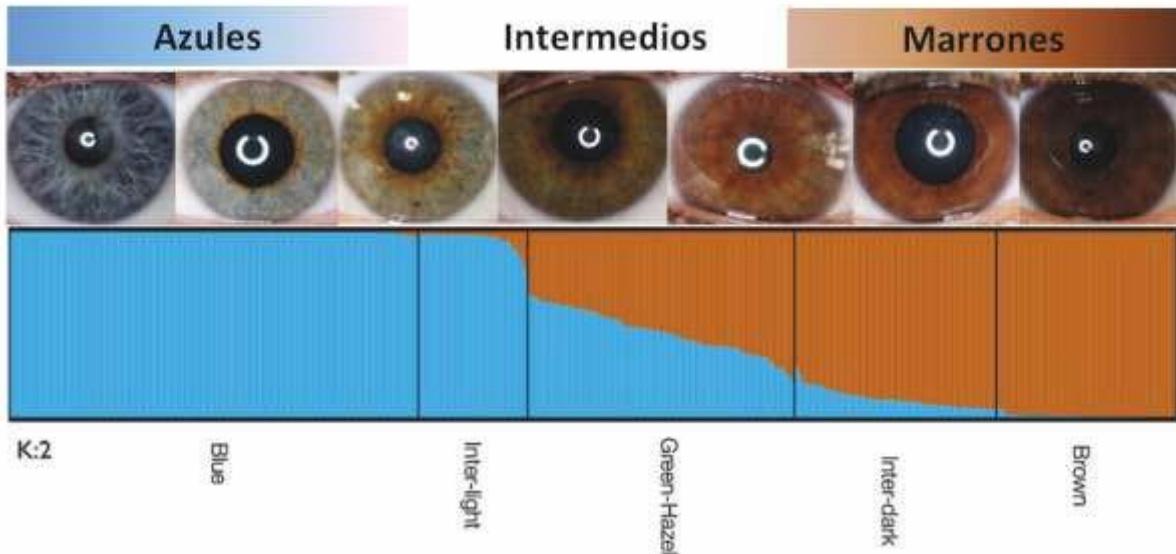


Fig.4. Diagrama de Structure asumiendo K=2 grupos, de las muestras de iris Azul, marrón e intermedio. Se aprecia una clara subestructura entre el grupo "Azul" y los restantes, mientras que el cambio entre "Marrón" e "intermedio" es más sutil, pero perceptible.

4. EN CONCLUSIÓN.

Hemos visto como el genotipado de SNPs autosómicos se adapta a la perfección a la predicción de origen biogeográfico, aunando las ventajas de los marcadores de transmisión uniparental y STRs. Se ha utilizado con éxito esta herramienta en múltiples ocasiones desde su desarrollo en la resolución de casos reales de investigación criminal. Incluso en aquellos en los que la a priori limitación de esta aplicación, el admixture, incrementaba la dificultad significativamente. El continuo desarrollo de un mayor número de marcadores ha llevado en los últimos años a expandir los límites de la predicción de origen biogeográfico y la resolución del cálculo entre poblaciones próximas.

Los estudios genómicos de última generación, cada vez más abordables, han hecho aparente la asociación de entre un grupo de ciertas posiciones variables con una serie de genes relacionados con la pigmentación. La aplicación de estos métodos de análisis genómico masivo podría abrir a estudio EVCs cada vez más complejos, con lo que en el futuro

próximo podría llegar a ser abordable la predicción de características más complejas e interesantes.

En todo caso, la investigación en caracteres relativamente sencillos, como la discutida en las páginas anteriores, constituye la base sobre la que asentarse para acometer el estudio de patrones mucho más complejos y su contribución al campo forense no debe ser en absoluto desdeñada. Y puede ser aplicada con éxito a la resolución de casos reales de investigación criminal.

BIBLIOGRAFÍA:

1. JJ SÁNCHEZ, C PHILLIPS, C BORSTIG, K BALOGH, M BOGUS, M FONDEVILA et al. A multiplex assay with 52 single nucleotide polymorphisms for human identification. *Electrophoresis* 2006, 27, 1713–1724.
2. AC Syanen. Accessing genetic variation: Genotyping single nucleotide polymorphisms. *Nature Reviews* vol. 2 Dec 2001, 930-942.
3. P GILL. An assessment of the utility of single nucleotide polymorphisms (SNPs) for forensic purposes. *Int J Legal Med.* 2001, 114 :204–210.

4. ST SHERRY, WARD M, SIROTKIN K. dbSNP- database for Single Nucleotide Polymorphisms and Other Classes of Minor Genetic Variation.1999, Genome Research 9: 667-679.
5. FR SANTOS, C TYLER-SMITH. Reading the human Y chromosome: the emerging DNA markers and human genetic history. J Genet 19: 665-670.
6. M FONDEVILA, C PHILLIPS, N NAVERAN, M CEREZO, A RODRIGUEZ, R. CALVO, et al. Challenging DNA: Assessment of a range of genotyping approaches for highly degraded forensic samples. FSI Genet: supp series. 2008, 1-26–28.
7. M FONDEVILA, C PHILLIPS, N NAVERAN, L FERNANDEZ, M CEREZO, A. SALAS et al. Case report: Identification of skeletal remains using short-amplicon marker analysis of severely degraded DNA extracted from a decomposed and charred femur. FSI Genet. 2 (2008) 212–218.
8. C ROMANINI, CATELLI ML, BOROSKYA, PEREIRA R, ROMERO M, SALADO PUERTO M et al. Typing short amplicon binary polymorphisms: supplementary SNP and Indel genetic information in the analysis of highly degraded skeletal remains. FSI Genet. 6 (2012) 469–476.
9. A FREIRE, FONDEVILA M, KRIEGEL AK, PHILLIPS C, GILL P, PRIETO L et al. A new SNP assay for identification of highly degraded human DNA. Forensic Sci Int Genet. 2012 May;6(3):341-9.
10. EM GOLEMBERG, AB AND P WEIHS. Effect of highly fragmented DNA on PCR. Nucleic Acids Research. 1996, vol24, No 24.
11. JA SYKORSKY DA PRIMERANO, W FENGER, J DENVIR. DNA damage reduces Taq DNA polymerase fidelity and PCR amplification efficiency. Biochem and Biophysical Res Communications 355 (2007) 431-437.
12. R ALAEDDINI. Forensic implications of PCR inhibition-A review. FSI Genet. 6 (2012) 297–305.
13. MD COBLE, JM BUTLER. Characterization of new miniSTR loci to aid analysis of degraded DNA. J. Forensic Sci. Jan 2005, Vol. 50 No. 1, 43-53.
14. C PHILLIPS, A. SALAS, JJ SANCHEZ, M FONDEVILA, A GOMEZ-TATO, J ALVAREZ-DIOS et al. Inferring ancestral origin using a single multiplex assay of ancestry-informative marker SNPs. FSI Genet. 1 (2007) 273–280.
15. RL CANN, M STONEKING, AC WILSON. Mitochondrial DNA human evolution. Nature, 325, 31-6.
16. LB JORDE, WS WATKINS, MJ BAMSHAD, ME DIXON, CE RICKER, MT SEIELSTAD, et al. The Distribution of Human Genetic Diversity: A Comparison of Mitochondrial, Autosomal, and Y-Chromosome Data. Am. J. Hum. Genet. 66:979–988, 2000.
17. C PHILLIPS, KIND S, FERNANDEZ-FORMOSO L, GELABERT-BESADA M, CARRACEDO A, LAREU MV. Global population variability in Promega PowerPlex CS7, D6S1043, and Penta B STRs. Int J Legal Med. 2013 Sep;127(5):901-6.
18. C PHILLIPS et al. D9S1120, a simple STR with a common Native American-specific allele: forensic optimization, locus characterization and allele frequency studies. FSI Genet. 3 (2008) 7–13.
19. C. PHILLIPS, RODRIGUEZ A, MOSQUERA-MIGUELA, FONDEVILA M, PORRAS-HURTADO L, RONDON F et al. FSI Genet. 7 (2013) 359–366.
20. O LAO, VAN DUIJN K, KERSBERGEN P, DE KNIJFF P, KAYSER M. Proportioning whole-genome single-nucleotide-polymorphism diversity for the identification of geographic population structure and genetic ancestry. Am J Hum Genet. 2006 Apr;78(4):680-90.
21. KA MILLS, EVEN D, MURRAY JC. Tetranucleotide repeat polymorphism at the human alpha fibrinogen locus (FGA). Hum Mol Genet (1992) 1 (9):779.
22. M FONDEVILA, PHILLIPS C, SANTOS C, FREIRE ARADAS A, VALLONE PM, BUTLER JM et al. Revision of the SNPforID 34-plex forensic ancestry test: Assay enhancements, standard reference sample genotypes and extended population studies. FSI Genet. 7 (2013) 63–74.
23. L PORRAS-HURTADO, RUIZ Y, SANTOS C, PHILLIPS C, CARRACEDO A, LAREU MV. An overview of STRUCTURE: applications, parameter settings, and supporting software. Front Genet. 2013 May 29;4:98.
24. VN SILBIGER, HIRATA MH, LUCHESSI AD, GENVIGIR FD, CERDA A, RODRIGUES AC. Differentiation of African components of ancestry to stratify groups in a case-control study of a Brazilian urban population. Genet Test Mol Biomarkers. 2012 Jun;16(6):524-30.
25. C PHILLIPS, PRIETO L, FONDEVILA M, SALAS A, GÓMEZ-TATO A, ALVAREZ-DIOS J et al. Ancestry analysis in the 11-M Madrid bomb attack investigation. PLoS ONE 4(8):e6583.
26. Y RUIZ, Phillips C, Gomez-Tato A, Alvarez-Dios J, Casares de Cal M, Cruz R. Further development of forensic eye color predictive tests Forensic Sci Int: Genet 7 (2013) 28–40.
27. FLIU, WEN B, KAYSER M. Colorful DNA polymorphisms in humans. Semin Cell Dev Biol. 2013 Jun-Jul;24(6-7):562-75.
28. M KAYSER, LIU F, JANSSENS AC, RIVADENEIRA F, LAO O, VAN DUIJN K et al. Three genome-wide association studies and a linkage analysis identify HERC2 as a human iris color gene. Am J Hum Genet. 2008 Feb;82(2):411-23.