

Recomendaciones de la Comisión Técnica Permanente (CTP) de la Comisión Nacional para el Uso Forense del ADN (CNUFADN) sobre valoración de coincidencias mediante el empleo de STR de cromosoma Y (aprobado en la reunión del Pleno de la CNUFADN de 11 de diciembre de 2023)

Justificación

El uso de STR de cromosoma Y en el campo de la genética forense es una herramienta ampliamente consolidada y especialmente ha demostrado su extraordinaria eficiencia en la resolución de aquellos casos en los que existen mezclas de material genético entre individuos de distinto sexo y en los que interesa la identificación del componente masculino, resultando particularmente útil cuando la desproporción a favor del componente femenino es muy acusada. (1-3).

Sin embargo, la ausencia casi completa de recombinación del cromosoma Y le confiere un patrón de herencia exclusivamente paterno, lo cual compromete la discriminación de esta herramienta genética y por tanto su utilidad identificativa. Para ganar en el poder discriminador del análisis de cromosoma Y los laboratorios forenses recurren al empleo de bases de datos poblacionales que recopilan haplotipos aportados por laboratorios de todo el mundo. Sin embargo, cuando debe estimarse estadísticamente el peso de una coincidencia entre una muestra indubitada y otra dubitada hay distintas aproximaciones matemáticas que emplean distintos parámetros estadísticos y que en ocasiones se traduce en valores finales significativamente diferentes.

Recientemente, la International Society of Forensic Genetics (ISFG) ha emitido unas recomendaciones (4) que pretenden abordar una serie de aspectos tales como la estima de frecuencias, el uso de base de datos poblacionales, tratamiento estadístico de coincidencias, parámetros estadísticos...que son necesarios a la hora de interpretar los análisis de marcadores de cromosoma Y en el contexto de la investigación criminal y humanitaria. La estandarización en esta materia, como en otras tantas de la disciplina de la genética forense, resultan absolutamente necesarias para poder transmitir a los Tribunales de justicia una imagen homogénea de la prueba genética y su interpretación. La ISFG anima a los organismos de estandarización nacionales en materia del uso de ADN con fines de investigación criminal/humanitario a pronunciarse en este sentido mediante la emisión de guías o recomendaciones de aplicación en el ámbito local, algunos grupos de estandarización internacionales ya lo han hecho (5-8).

El alcance de estas recomendaciones es de aplicación sobre los laboratorios que realizan su actividad pericial en el campo de la genética forense con fines de investigación criminal a nivel estatal y el objetivo de este documento pretende tender a la máxima estandarización respecto a la valoración estadística en el análisis de marcadores STR de cromosoma Y.

1. Cuando como resultado del análisis genético con marcadores STR de cromosoma Y resulte una coincidencia entre muestra dubitada (MD) y la muestra de referencia (MR) es necesario transmitir al Tribunal el peso probabilístico de dicha coincidencia. Este tipo de valoración en términos numéricos permite al Tribunal adquirir un valor ponderable para evaluar la prueba genética. Distintas guías y recomendaciones internacionales recogen la necesidad de incluir una valoración estadística de la prueba en los informes de carácter evaluativo (9-11).

Recomendación 1

Se recomienda realizar una valoración de dicha coincidencia en términos cuantitativos. Esta recomendación será de aplicación tanto en el campo criminal como en el humanitario.

2. El tamaño y la composición de las bases de datos poblacionales que se emplean para realizar las estimas de frecuencias de marcadores haplotípicos (cromosoma Y y ADN mitocondrial) resultan de especial importancia para garantizar la calidad del dato estadístico obtenido así como para conseguir un mayor grado de discriminación del valor estadístico obtenido. De manera general, y debido al patrón monoparental de herencia de dichos marcadores, cuanto mayor sea el tamaño de las bases de datos poblacionales también será mayor el poder de discriminación que nos ofrece el empleo de marcadores de cromosoma Y con fines de identificación.

Las bases de datos poblacionales a emplear deben cumplir una serie de características que se resumen en:

- Garantizar una alta calidad de los haplotipos que están incluidos en la base de datos que ofrezcan máxima fiabilidad de los mismos.
- Bases de datos amplias que permita representar los *clusters* dominantes en la población de referencia que pueda ser utilizada.
- Muestreo aleatorio de los individuos seleccionados que permita la mayor representatividad de los parentescos existentes en la población.
- Muestras anonimizadas para evitar la asociación entre un registro haplotípico y la identificación concreta del donante de la muestra.
- Los haplotipos deben contar con una información suficiente (datos geográficos de la toma, grupo étnico, nacionalidad, lenguaje del donante...) que permita encuadrar dicho haplotipo en una determinada metapoblación y/o subpoblación.

La base de datos YHRD (www.yhrd.org) cumple con los requisitos descritos anteriormente para convertirla en una base de datos poblacional de calidad (12). Esta base de datos es la utilizada de manera mayoritaria a nivel internacional por los laboratorios de genética forense para llevar a cabo la evaluación estadística de coincidencias obtenidas con análisis de cromosoma Y. A la fecha de realización de este documento (Release 69, mayo 2023) la base de datos YHRD cuenta con 37 metapoblaciones y 141 bases de datos nacionales.

Recomendación 2

Se recomienda para realizar estimas de frecuencias de haplotipos de cromosoma Y el empleo de la base de datos poblacionales YHRD (www.yhrd.org) y los algoritmos estadísticos que incluye la citada base de datos.

3. La estima de una determinada frecuencia haplotípica puede realizarse mediante el empleo de bases de datos locales (nacionales) o mediante el empleo de metapoblaciones. La metapoblación es un concepto que engloba poblaciones ubicadas en distintos entornos geográficos pero que comparten ancestros comunes. Realizar las estimas de frecuencias sobre una determinada metapoblación, normalmente, permite realizar estimas más discriminativas. De manera general, una metapoblación incluye distintas poblacionales locales y por tanto, a priori, esas búsquedas ofrecen un mayor poder de discriminación. Por otro lado, cuando deba valorarse una coincidencia entre MD y MR debe emplearse como población de referencia la del lugar donde suceden los hechos (salvo indicaciones en contra para contextos muy concretos). La metapoblación adecuada será aquella en la que está incluido la población (área geográfica) en la que suceden los hechos. Por ejemplo, si el individuo sospechoso de una agresión sexual es de origen norte africano y el delito se comete en Alemania la metapoblación a emplear debe ser la del entorno, es decir la de *Western European* (en ella se incluye la población de Alemania).

Recomendación 3

Se recomienda a los laboratorios el empleo de **metapoblación** para realizar estima de frecuencias haplotípicas y en concreto la metapoblación *Western European* (YHRD database), salvo especificaciones concretas por los antecedentes del caso que aconsejen el uso de otra metapoblación alternativa.

4. **La estima de la frecuencia** de un determinado haplotipo puede realizarse siguiendo distintas aproximaciones probabilísticas: mediante *contaje* o utilizando *estimaciones probabilísticas de aparición de un determinado haplotipo*.
- 4.1 Método de contaje: Consiste en realizar el conteo del número de veces (n) que aparece un determinado haplotipo en una base de datos poblacional de tamaño (N).

$$p = \frac{n}{N}$$

A menudo se emplea el *método de contaje aumentado*, esta aproximación permite corregir los errores de muestreo que pueden producirse en bases de datos de tamaño reducido. Consiste en considerar el haplotipo observado tanto en las observaciones como en la propia base de datos.

$$p = \frac{n + 1}{N + 1}$$

En ocasiones la estima de frecuencia mediante este método de contaje aumentado se acompaña de intervalos de confianza (Clopper–Pearson interval (Biometrika 26: 64-69 1934). (ver anexo I)

Por otro lado, también es posible corregir el número de observaciones incluyendo un factor kappa (k). Kappa method” (Brenner. FSI Genet 4(5): 281-291 (2010)) (ver anexo I).

$$p = \frac{(n + 1) (1 - k)}{N + 1} \qquad \kappa = \frac{\alpha}{N}$$

$\alpha = n^{\circ}$ de haplotipos únicos en la base de datos

El método de contaje tiene un importante inconveniente ya que subestima los haplotipos que no están representados en la base de datos consultada, y ello lo convierte en una estima excesivamente conservativa. Por el contrario, y como ventaja, su empleo es de fácil explicación en el informe pericial y por tanto facilita la comprensión del dato por parte de los Tribunales.

- 4.2 Estima de frecuencias: mediante el uso del *método discreto de Laplace (DL)*, (anexo I), Este método paramétrico, en base a la propia composición de la base de datos poblacionales, tiene la ventaja de calcular la probabilidad estimada de haplotipos que no están representados en la base de datos, y por tanto no subestima los haplotipos no representados en la base de datos por cuestiones de muestreo insuficiente.

El método de cálculo asume que los distintos haplotipos representados en la base de datos se agrupan en *clusters*, los cuales parten de un haplotipo central que es común a todos los haplotipos del mismo *cluster*.

El método basado en el cálculo de DL no es posible aplicarlo con haplotipos que incluyen alelos intermedios o marcadores con alelos múltiples. Además requiere que el haplotipo incluya los marcadores que componen el haplotipo central en las distintas metapoblaciones incluidas en la base de datos. Por otro lado, no todas las metapoblaciones representadas en la base de datos de cromosoma Y de referencia (YHRD) www.yhrd.org puede aplicarse el método DL.

Recomendación 4

Se recomienda realizar la **estima de frecuencias utilizando el método de discreta de Laplace (DL)**, siempre que las características del haplotipo obtenido y del caso analizado lo permitan. Si no es posible la estima de frecuencias mediante DL se recomienda el empleo del método de contaje aumentado incluyendo el factor kappa de corrección a criterio del laboratorio.

5. Atendiendo a lo expuesto en el recomendación 1 de este documento y según lo recogido en distintas recomendaciones internacionales (9-10) y de esta propia CNUFADN (11), los informes evaluativos deben expresar la valoración de la prueba (p. ej. coincidencia entre los haplotipos de cromosoma Y) utilizando el valor de LR (*likelihood ratio* o razón de verosimilitud) planteando hipótesis alternativas y mutuamente excluyentes en los que se evalúe probabilísticamente el hallazgo científico dadas ambas hipótesis, lo cual permite dotar de la objetividad y equilibrio exigible a los dictámenes periciales de naturaleza evaluativa.

Es importante que el informe incluya el planteamiento correcto de ambas hipótesis alternativas en los términos siguientes:

Hipótesis H1: La persona A es la donante (fuente) del ADN presente en la muestra X.
Hipótesis H2: Otra persona B tomada al azar de la población de referencia es el donante del ADN presente en la muestra X.

La verbalización del valor del LR debe incluirse en el informe evaluativo, por ejemplo, bajo la siguiente expresión:

El perfil de STR de cromosoma Y detectado en la muestra X es VALOR DEL LR número de veces más probable observarlo bajo la hipótesis H1 que hacerlo bajo la hipótesis H2. A pesar de esto, los parientes paternos tienen una alta probabilidad de tener el mismo perfil Y-STR y en ese caso tendrán el mismo valor de LR.

Como continuidad de lo expuesto en los puntos 3 y 4 de este documento el empleo de un método de estima de frecuencias u otro (p. ej. de conteo / Discreta de Laplace) o una metapoblación u otra pueden generar cambios importantes en dicha valoración y por tanto en la valor final de estimación, por ello, el informe pericial debería incluir el método empleado para llevar a cabo la estima de frecuencia así como la metapoblación empleada en la búsqueda y la versión (*release*) de la base de datos YHRD utilizada.

Recomendación 5

Se recomienda el empleo del LR (likelihood ratio) (13) como parámetro para evaluar el ratio de probabilidades para la ocurrencia del hallazgo científico dadas ambas hipótesis alternativas. Las hipótesis deben constar en el informe pericial así como la verbalización del resultado del valor del LR.

Así mismo, se recomienda incluir en el informe pericial emitido el método empleado para realizar la estima de frecuencia haplotípica, la metapoblación que ha sido empleada y la actualización (*release*) de la base de datos

Bibliografía

- [1] L. Roewer, Y chromosome STR typing in crime casework, *Forensic Sci. Med. Pathol.* 5 (2) (2009) 77–84.
- [2] M. Kayser, Forensic use of Y-chromosome DNA: a general overview, *Hum. Genet.* 136 (5) (2017) 621–635.
- [3] L. Roewer, Y-chromosome short tandem repeats in forensics - sexing, profiling, and matching male DNA, *WIREs Forensic Sci.* 1 (4) (2019) e1336.
- [4] L. Roewer, M. Meyer Andersen, J. Ballantyne, J. M. Butler, A. Caliebe, D. Corach, M. E. D'Amato, L. Gusmão, Y. Hou, P. de Knijff, W. Parson, M. Prinz, P. M. Schneider, D. Taylor, M. Vennemann, S. Willuweit, DNA commission of the international society of forensic genetics (IGSF): Recommendations on the interpretation of Y-STR results in forensic analysis. *Forensic Science International. Genetics* (2020) 48: 102308.
- [5] C. Leuker, G. Molsberger, U. Pich, S. Razbin, T. Rothmel, H. Schneider, P. M. Schneider, M. Templin, M. Vennemann, A. Wächter, V. Weirich, H. Zierdt, L. Roewer, Gemeinsame

Empfehlungen der Projektgruppe „Biostatistische DNABerechnungen“ und der Spurenkommission zur biostatistischen Bewertung von Ychromosomalen DNA-Befunden, Rechtsmedizin 28 (2) (2018) 138–142.

- [6] Scientific Working Group on DNA Analysis Methods (SWGDM), Interpretation Guidelines for Y Chromosome STR Typing (published Jan 9), (2014).
http://media.wix.com/ugd/4344b0_da25419ba2dd4363bc4e5e8fe7025882.pdf
- [7] Recommendations of the Polish Speaking Working Group of the International Society for Forensic Genetics on forensic Y chromosome typing. Arch Med Sadowej Kryminol 70(1): 1-18 (2020).
- [8] Guidance for Y-STR analysis delivered into the criminal justice system in England and Wales <https://www.gov.uk/government/publications/y-str-profiling>
- [9] ENFSI Guideline for Evaluative Reporting in Forensic Science. Disponible en:
https://enfsi.eu/wp-content/uploads/2016/09/m1_guideline.pdf
- [10] Forensic Science Regulator Codes of Practice and Conduct Development of Evaluative Opinions FSR-C-118 Issue1
https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/960051/FSR-C-118_ Interpretation Appendix Issue 1_002 .pdf
- [11] Recomendaciones emitidas por la CNUFADN en materia del *informe pericial y la expresión de resultados en materia de análisis genéticos forenses*.
https://www.mjusticia.gob.es/es/ElMinisterio/OrganismosMinisterio/Documents/Recomendaciones_sobre_el_informe_pericial_y_la_expresion_de_resultados_en_materia_de_analisis_genet.PDF
- [12] S. Willuweit and L Roewer. The new Y Chromosome Haplotype Reference Database. Forensic Science International: Genetics 15 (2015) 43–48
- [13] I.W. Evett, B.S. Weir, Interpreting DNA Evidence: Statistical Genetics for Forensic Scientists, Sinauer Associates, Sunderland, 1998.