

ANÁLISIS GENÉTICO FORENSE:

Técnicas de Vanguardia y

Nuevas Aplicaciones

Autor: Xavier Martí i Pérez

Tutora: Ximena Terra Barbadora

Curso Académico 2019/20

Master en Genética, Física y Química Forense

Universitat Rovira i Virgili

RESUMEN

La genética forense, del mismo modo que el resto de ciencias forenses, presenta una serie de condiciones especiales a la hora de actualizar las técnicas de análisis. Es cada vez más probable que los métodos usados por defecto actualmente, PCR junto con electroforesis capilar, y secuenciación Sanger, se vean reemplazados por tecnologías de secuenciación más modernas. En este trabajo se explican las ventajas que presentaría la implementación de la *Next-Generation Sequencing* como el nuevo *Gold Standard* en el análisis forense, así como las limitaciones por las que no se ha dado la transición a día de hoy. Se detallan también las nuevas aplicaciones que esta tecnología permitiría incorporar o mejorar: identificación de tejidos, autopsia molecular, predicción de características físicas y avances en estudios genealógicos. Se discute, finalmente, los progresos y la potencial aplicación de los secuenciadores de última generación en la elaboración de perfiles genéticos, así como otras tecnologías punteras paralelas o complementarias a los métodos de análisis usados.

ABSTRACT

Forensic genetics, like any other forensic science, requires a specific set of conditions when updating the analysis instrumental. It is becoming increasingly likely that the current methods, a combination of PCR with capillary electrophoresis, and Sanger sequencing, will be replaced with more modern sequencing technologies. In this project, the advantages that Next-Generation Sequencing would provide, as well as the current limitations because of which it has not been implemented yet, are explained. The main novel applications that the incorporation of this technology would either allow or upgrade are also shown: tissue identification, molecular autopsy, prediction of physical characteristics and improvements in genealogic studies. Finally, the advances and potential use of cutting-edge sequencing in genetic profiling, and other new technologies, parallel or complementary to the used analysis methods, are discussed.

TABLA DE CONTENIDO

1. INTRODUCCIÓN	1
1.1. Breve historia de la genética forense	1
1.2. Estado actual de la genética forense	2
2. OBJETIVOS DE LA REVISIÓN	3
3. MATERIALES Y MÉTODOS	4
4. CARACTERÍSTICAS PRINCIPALES DE LOS DISTINTOS TIPOS DE SECUENCIACIÓN	5
4.1. Secuenciación Sanger	5
4.2. <i>Next-Generation Sequencing</i>	7
4.3. <i>Single Molecule Sequencing</i>	7
5. IDENTIFICACIÓN DE TEJIDOS CON ÁCIDOS NUCLEICOS	8
5.1. Introducción	8
5.2. Identificación de tejidos basada en RNA	9
5.3. Identificación de tejidos basada en DNA	11
5.4. Previsión	12
6. AUTOPSIA MOLECULAR	13
6.1. Introducción	13
6.2. La genética y la genómica en muertes súbitas	13
6.3. Estudio toxicogenético post-mortem	14
6.4. Microgenómica	15
6.5. Consideraciones adicionales	16
7. PREDICCIÓN DE CARACTERÍSTICAS FÍSICAS	16
7.1. Introducción	16
7.2. Pigmentación y altura	17
7.3. Elaboración de <i>predictors</i> y validación	18
7.4. Direcciones futuras	19

8. GENEALOGÍA EN LA ERA GENÓMICA.....	20
8.1. Introducción.....	20
8.2. Testeo de DNA haploide	21
8.2.1. Cromosoma Y.....	21
8.2.2. mtDNA	22
8.3. Testeo de DNA autosómico	23
8.4. Discusión	24
9. SECUENCIACIÓN DE ÚLTIMA GENERACIÓN Y LA GENÉTICA FORENSE	24
9.1. Introducción a la secuenciación Nanopore	24
9.2. Elaboración de perfiles de STR.....	25
8.3. Genotipaje de SNPs trialélicos	27
9.4. Discusión	28
10. OTRAS TECNOLOGÍAS DE ANÁLISIS GENÉTICO FORENSE.....	29
10.1. Análisis rápido de DNA.....	29
10.2. Espectrometría de masas.....	30
10.3. Automatización	30
10.4. Consideraciones finales y resumen	31
11. DISCUSIÓN GENERAL	32
12. CONCLUSIONES.....	33
13. REFERENCIAS	34

1. INTRODUCCIÓN

La genética forense puede definirse como la aplicación de la genética a la resolución de conflictos legales. La evolución de esta se ha asociado siempre al estudio de la variación genética humana, así como al progreso tecnológico¹.

1.1. Breve historia de la genética forense

El origen de esta ciencia data de hace más de un siglo, con el descubrimiento de los grupos sanguíneos ABO por Karl Landstainer, el cual se percató de que estas variantes podían ser usadas para resolver casos de paternidad y crímenes. Tal estudio de la variación genética hallada en proteínas, llamada serología forense, devino más compleja y fue usada regularmente hasta la llegada de la tipificación del DNA o *DNA typing*².

El *DNA typing* ofrecía ventajas significativas respecto al estudio de proteínas, puesto que el DNA era mucho más informativo y presentaba una mayor resistencia relativa a la degradación. Además, la información podía ser extraída de cualquier tejido, en contraposición con la necesidad de la presencia específica en los análisis de las células que expresaban las proteínas usadas en serología. El primer caso en el cual se usó este método fue en 1985, mientras que el primer caso de naturaleza criminal fue en 1986. Ambos análisis genéticos se basaron en el estudio de minisatélites mediante *Single Locus Probes* (SLPs) y, más precisamente, *Restriction Fragment Length Polymorphism* (RFLP)^{2,3}.

Con la implementación de una todavía en fase de desarrollo *Polymerase Chain Reaction* (PCR) en 1987 y el descubrimiento de los *Short Tandem Repeats* (STR) o microsatélites en 1989, la genética forense dio otro paso de gigante en su evolución. La rápida amplificación específica de fragmentos de DNA sustancialmente más cortos que los minisatélites -y, por lo tanto, más fácilmente analizables y menos exigentes en la cuestión del estado de la muestra-, combinada con su fácil discriminación por migración mediante electroforesis, dieron como resultado un mejor procedimiento a la hora de analizar el material genético. Esta combinación de técnicas (sometidas evidentemente a constantes mejoras desde su aparición) se convertiría en el *Gold Standard* hasta día de hoy para la principal aplicación de la genética forense, esto es, la identificación y asociación de personas vía contraste de los perfiles genéticos entre muestras².

Posteriormente se descubrieron más potenciales marcadores y se desarrollaron nuevas

técnicas de análisis como: las regiones hipervariables (HVR) del ADN mitocondrial (mtDNA), los marcadores del cromosoma Y y los *Single Nucleotide Polymorphisms* (SNPs); y la secuenciación capilar o Sanger, respectivamente. Estos nuevos marcadores y la secuenciación, sin embargo, no sustituyeron a la amplificación de STRs y su análisis por electroforesis capilar (EC) y fueron destinados a funciones complementarias como el establecimiento de linajes, entre otros^{1,2}.

1.2. Estado actual de la genética forense

Los laboratorios de genética forense siguen, por tanto, adoptando para sus análisis más habituales una combinación de técnicas clásicas como la PCR-EC en vez de usar técnicas punteras de secuenciación masiva.

Con el fin de entender la situación relativamente estanca en la que se halla la genética forense, especialmente si se compara con el resto de subramas de la genética, es necesario comprender las condiciones y requisitos especiales del análisis genético forense.

Para empezar, conviene considerar los requisitos de los marcadores de DNA empleados en genética forense¹:

- ✓ Altamente polimórficos.
- ✓ Caracterización fácil y barata.
- ✓ Perfiles resultantes fáciles de interpretar y comparar entre laboratorios.
- ✓ Nula presión selectiva y en equilibrio de Hardy-Weinberg.
- ✓ Tasa de mutación baja.

Además, debido a la determinante influencia que pueden llegar a tener las conclusiones de los análisis en la resolución en las vidas de los involucrados en los casos donde han sido solicitados, es imprescindible que los resultados sean extremadamente fiables, hecho para el cual se necesita que las técnicas usadas para los análisis tengan una contrastada efectividad y grado de perfeccionamiento^{4,5}.

Vistas entonces las características necesarias, el análisis de STRs mediante PCR-EC resulta el método que más se acerca a todos los requisitos simultáneamente. A pesar de obtener objetivamente menos información si los STRs se analizan por EC en lugar de por Sanger, el coste y tiempo significativamente menores y la contrastada efectividad de

unas técnicas optimizadas constantemente a lo largo de muchos años han prevalecido sobre una ligera mejora informativa, considerada no crucial. Por otro lado, el análisis de otros marcadores, como los SNPs, habría resultado totalmente impracticable en la época de su descubrimiento por limitaciones económicas, temporales y, probablemente, computacionales⁶.

No obstante, cabe la notable posibilidad de que, en un futuro no demasiado lejano, se produzca una progresiva sustitución en los laboratorios de la PCR-EC por técnicas más avanzadas y cada vez más consolidadas que permitan no solo un incremento significativo en la efectividad, versatilidad y calidad de los análisis, sino también la ejecución de análisis a un coste inferior al de las técnicas actuales⁶.

Es por esto que en este trabajo se detallan cuáles son las potenciales ventajas que pueden presentar estos nuevos métodos de análisis en campo de la genética forense, tanto en mejora como en innovación. A pesar de que las primeras máquinas de secuenciación con capacidad *High Throughput* daten de hace 15 años⁷, la llamada *Next-Generation Sequencing* (NGS), caracterizada por la secuenciación masiva y paralela de los fragmentos de DNA, va a ser tratada aquí como una técnica de vanguardia dada su poca o nula implantación en los laboratorios forenses hasta la fecha, por las razones explicadas previamente. Sin embargo, dentro de la propia secuenciación existe un elevado repertorio de técnicas distintas, las más modernas de las cuales sí pueden considerarse de vanguardia en cualquier campo de la genética⁸.

La NGS es la herramienta principal por la cual las aplicaciones noveles detalladas en este trabajo se ven más beneficiadas, pero también se han incluido otras técnicas de base científica radicalmente distinta. La razón por la que se mencionan tangencialmente es su esperada poca progresión o relevancia relativa en comparación con la secuenciación⁵.

2. OBJETIVOS DE LA REVISIÓN

Con este trabajo se pretende sintetizar en un único documento los potenciales avances tecnológicos más relevantes que pueda experimentar la genética forense durante la

próxima década, sirviendo de esta forma como un rápido ejemplar de consulta en caso requerido.

Asimismo, se pretende dar respuesta a la cuestión de si los laboratorios forenses, al servicio del estado y con un presupuesto destinado a la investigación criminal limitado, deberían reforzarse con dichas nuevas tecnologías, argumentando su potencial grado de impacto en el incremento de conflictos legales resueltos contra el esfuerzo económico requerido.

3. MATERIALES Y MÉTODOS

Con el fin de realizar esta revisión se han consultado libros y artículos científicos como base informativa. A su vez, la mayor parte de figuras y tablas han sido extraídas de los mismos libros y artículos sobre los cuales se ha basado el redactado, únicamente recurriendo a fuentes externas cuando se estimó oportuno complementar las descripciones textuales y no había material gráfico disponible o suficientemente explicativo en el mismo recurso. Tanto la redacción como las consultas fueron realizadas íntegramente en formato digital, contribuyendo de esta forma a la preservación del medio ambiente.

Por lo que respecta a los criterios de búsqueda, se ha estimado oportuna la consulta de libros y artículos científicos cuya fecha de publicación no excediera los 10 años y 5 años, respectivamente, para procurar que la consulta de las nuevas tecnologías consideradas para realizar análisis forenses y las aplicaciones más novedosas del mismo fuera la más actual posible, a la par que hubiera suficiente material de estudio. Otras fuentes consultadas para la introducción, como la historia general de la genética forense o las características de los marcadores forenses, no han requerido esa criba temporal.

La selección de las aplicaciones expuestas en este trabajo ha sido filtrada, a criterio personal, en base a su utilidad y atractivo científico.

4. CARACTERÍSTICAS PRINCIPALES DE LOS DISTINTOS TIPOS DE SECUENCIACIÓN

Secuenciar el genoma de un organismo consiste en conocer la secuencia de pares de bases (A,T,C,G) que lo componen, obteniéndose así información capaz de ser leída y analizada. La secuenciación puede ser total o parcial y puede aplicarse tanto al DNA como al RNA, dadas las preparaciones necesarias. Conocer la información genética de un organismo supone un sinfín de posibles aplicaciones relacionadas con los campos de medicina, biología del desarrollo, evolución y, en este caso, las ciencias forenses⁹.

4.1. Secuenciación Sanger

La secuenciación enzimática, de la cual derivan todas las técnicas de secuenciación actuales, data de hasta 1975, cuando fue inventada por Frederick Sanger. La base de su secuenciación consistía en imitar el sistema natural de replicación del DNA en las células, esto es, el añadimiento progresivo de nucleótidos complementarios a una cadena de DNA por parte de un enzima llamado DNA polimerasa. En el caso de la secuenciación, sin embargo, en lugar de añadir siempre deoxinucleótidos se añadían ocasionalmente dideoxinucleótidos (ddNTPs) marcados que impedían seguir con la incorporación de nucleótidos en esa cadena. Esto provocaba, terminada la reacción, múltiples cadenas de distinta longitud marcadas al final⁹.

En su estadio más primitivo, cada tubo de reacción presentaba un único tipo de ddNTP marcado, de forma que había 4 tubos en total (uno para cada base nitrogenada distinta). Las cadenas de dichos tubos se separaban posteriormente por electroforesis y se obtenía una pauta ordenada en un eje del gel de marcadores, que se correspondían al nucleótido terminador en función de la migración de las cadenas en el gel, inversamente proporcional a su peso y, en consecuencia, a su longitud. Se procedía entonces al registro manual de la cadena resultante de la unión de los 4 carriles de electroforesis⁹.

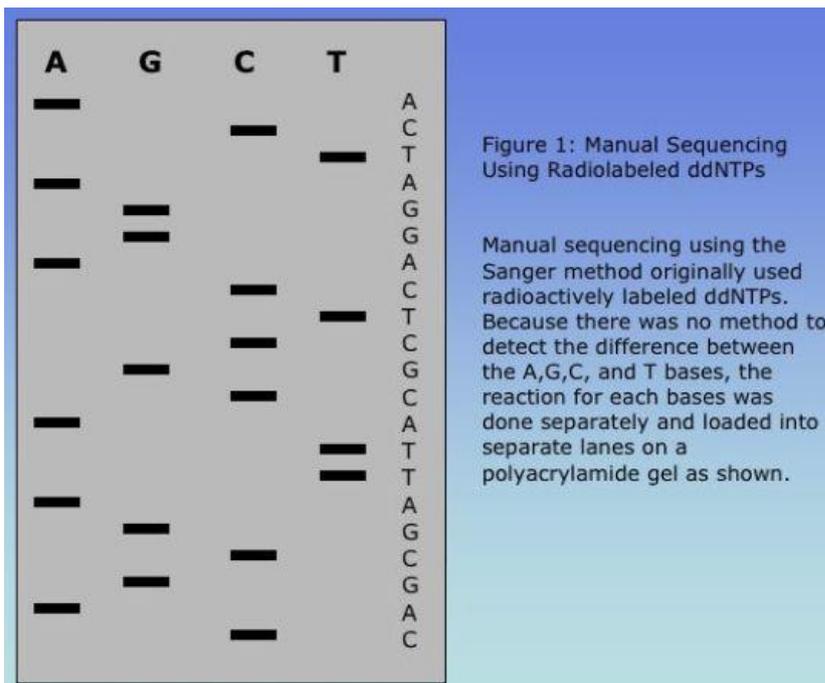


Figura 1. Ilustración de una cadena deducida por la secuenciación Sanger manual (el orden de lectura de la cadena es de abajo a arriba)¹⁰.

La secuenciación Sanger manual era tediosa, lenta y solo podía secuenciar del orden de 100 pares de bases (bp) por ensayo o *run*. En 1998 apareció el primer secuenciador Sanger que automatizaba y simplificaba los pasos anteriores⁹:

- ✚ Amplificación del fragmento a secuenciar por PCR (este paso se implementó tan pronto como surgió la técnica).
- ✚ Un único tubo de ensayo con 4 ddNTPs marcados diferencialmente.
- ✚ La electroforesis en gel se sustituyó por la electroforesis capilar con detector láser.
- ✚ El registro de señal de fluorescencia estaba conectado a un ordenador, que convertía los códigos de color en una secuencia nucleotídica.

Esta secuenciación, llamada secuenciación capilar o de primera generación, fue la que permitió acortar los resultados del Proyecto Genoma Humano de 10 años a, aproximadamente, la mitad. Pudo ser así debido a su notable mayor rapidez y su ampliación a 1000pb por *run*⁹.

Además, con su >99.999% de fiabilidad (por bp), sigue siendo hoy en día la técnica de referencia en la investigación, especialmente donde se requiere secuenciación *de novo* (sin referencia previa del genoma en cuestión), y la usada por la inmensa mayoría de

laboratorios de genética forense en caso de necesitar conocer la secuencia de algún fragmento de DNA, a pesar de su elevado coste y su bajo *throughput*⁹.

4.2. Next-Generation Sequencing

Se habla por primera vez de NGS, o secuenciación de segunda generación, con la invención del secuenciador 454 en 2005 por Jonathan Rothberg. En general, este tipo de secuenciación se diferencia de la secuenciación capilar en los siguientes aspectos⁹:

- Fragmentos de DNA a secuenciar más cortos, de 100-200bp.
- Amplificación y secuenciación de cada fragmento de DNA de forma masiva (de $\times 100$ a $\times 10^6$). Esto permite secuenciar muchísima más información en un solo *run*.
- Fiabilidad relativamente baja, entre el 99-99.9%.
- Coste significativamente menor.
- Mayor rapidez de análisis.

El hecho de que la secuenciación sea masiva y los fragmentos sean más cortos implica, a su vez, que se requiera una capacidad de procesamiento de datos notablemente mayor en el ensamblaje de los *reads* para reconstruir el fragmento de DNA original⁹.

Tras muchos años de constantes mejoras, la tecnología NGS se encuentra en la actualidad en un punto muy próximo al de conseguir el balance rendimiento-coste necesario para que merezca la pena implementar este sistema como el análisis de rutina, tanto en la investigación como en los laboratorios forenses, desbancando de esta forma a la secuenciación Sanger¹¹. Su precio menor y su elevada versatilidad hacen que la NGS sea una herramienta a tener en cuenta para las nuevas aplicaciones surgidas en la genética forense, las más importantes de las cuales se detallan a lo largo de este trabajo.

4.3. Single Molecule Sequencing

La *Single Molecule Sequencing* (SMS), o secuenciación de tercera generación, es el escalón más moderno en lo que respecta a las bases para la secuenciación. Como su propio nombre indica, este tipo de secuenciación se centra en la lectura de la molécula original, sin amplificación por PCR y a tiempo real. Este enfoque supone unos *reads*

resultantes mucho más largos que los vistos hasta ahora, del orden de diversos miles de pares de bases, así como una reducción de material necesario para la secuenciación y un menor esfuerzo computacional⁹.

Este enfoque se encuentra todavía en una fase de desarrollo muy temprana, por lo cual presenta un precio muy elevado y una fiabilidad muy baja (90%). Sin embargo, una vez sea perfeccionada se espera que se pueda efectuar lecturas fiables y muy largas de material genético a un coste muy reducido y de forma muy versátil⁹.

5. IDENTIFICACIÓN DE TEJIDOS CON ÁCIDOS NUCLEICOS

5.1. Introducción

Las muestras biológicas provenientes de fluidos corporales humanos, tejidos líquidos y piel epitelial recogidas de una escena del crimen suelen conllevar DNA útil para la identificación del donante de dicha muestra. Después de dicha identificación, establecer el origen celular de los restos biológicos usados para el perfil de DNA supone una tarea fundamental en la genética forense, puesto que permite reconstruir la secuencia de eventos que ocurrieron durante el crimen, así como evaluar la importancia de las pruebas recogidas¹².

Incluso si el tipo celular que ha originado el rastro biológico parece obvio en base a una inspección ocular, es altamente recomendable que sea validado por un método fiable y objetivo. Es por esto que los investigadores forenses necesitan herramientas para que los rastros sean identificados de forma inequívoca y puedan ser defendidos en un juicio, juntamente con la identidad del donante¹².

Los métodos convencionales para la identificación del tipo celular se basan en generalmente en ensayos químicos, enzimáticos e inmunológicos, los cuales son de naturaleza presuntiva y de tendencia a dar falsos positivos y falsos negativos. Muy pocos tests clásicos y no moleculares, como la identificación por microscopio de espermatozoides, pueden considerarse concluyentes. Otras desventajas de estos ensayos son su habitual destrucción de la muestra y su especificidad, esto es, que suelen

informar únicamente de si la muestra se corresponde a un tipo de tejido en particular o no¹².

Durante la última década se han desarrollado para las ciencias forenses nuevas técnicas basadas en métodos físicos, biológicos y bioquímicos destinadas a confirmar el tipo celular que no presentan los problemas mencionados en el párrafo anterior. Es este apartado se exponen únicamente los basados en ácidos nucleicos¹².

5.2. Identificación de tejidos basada en RNA

El uso del RNA para identificar tejidos y tipos celulares se basa en que, a diferencia del DNA, el RNA -y, concretamente, el RNA mensajero o mRNA- se expresa de forma específica en función del tejido que corresponda. Este fenómeno se conoce como expresión génica diferencial. Los métodos destinados a la cuantificación de moléculas de RNA son muy sensibles y están bien establecidos, a pesar de que inicialmente no fueran pensados para el ámbito forense¹².

Estos métodos, a diferencia de los ensayos convencionales, no tienden a falsos positivos derivados de contaminación por agentes ambientales o material genético de animales debido a que la gran mayoría de tests basados en RNA son específicos para el ser humano. Además, la extracción del RNA puede realizarse de forma paralela a la del DNA, de forma que el principal material para el perfil de identificación no se desaprovecha. La principal limitación del RNA es su relativamente alta inestabilidad, en comparación con el DNA, y su consecuente tendencia a la degradación en condiciones *in vitro*. Pese a tal hecho, diversos estudios demostraron la persistencia de marcadores de RNA en diversos escenarios post-mortem y en muestras de diversos tejidos preservadas *in vitro*¹².

Tabla 1. Marcadores de mRNA validados por el European DNA Profiling Group para la identificación de sangre, saliva y semen¹².

Targeted Tissue/Body Fluid	mRNA Markers	Suitability of Markers
Peripheral blood	HBB (Hemoglobin beta)	Good candidates: high sensitivity
	HBA (Hemoglobin alpha)	
	ALAS2 (Aminolevulinic acid synthase)	Good candidates: medium sensitivity
	CD3G (T-cell surface glycoprotein CD3, gamma chain)	
	ANK1 (Ankyrin 1, erythrocytic)	
	PBGD (Porphobilinogen deaminase)	
	SPTB (β -Spectrin)	
AQP9 (Aquaporin 9)	Not specific	
Saliva	HTN3 (Histatin 3)	Good candidates
	STATH (Statherin)	
	MUC7 (Mucin 7)	
	PRB1-3 (Proline-rich protein BstNI subfamily 1)	Not sensitive
	PRB4 (Proline-rich protein BstNI subfamily 4)	
	SPRR2A (Small proline rich protein 2A)	Not specific
KRT13 (Keratin 13)		
Semen	PRM1 (Protamin 1)	Good candidates
	PRM2 (Protamin 2)	
	TGM4 (Transglutaminase 4)	
	KLK3 (Kalikrein 3)	
	SEMG1 (Semenogelin 1)	
	SPANXB [SPANX family member B]	Inconsistent results during singleplex testing
	HSFY [Heat shock transcription factor, Y-linked]	
	SZPBP [Zona pellucida binding protein]	
	ODF1 [Outer dense fiber of sperm tails 1]	
	BPY2 [Basic charged, Y-linked 2]	

La principal ventaja del uso del mRNA para la identificación de tejidos es la posibilidad de análisis simultáneo de distintos marcadores. Para ello se emplea la *Reverse Transcriptase* PCR (RT-PCR) de tipo multiplex, una PCR donde, inicialmente, se convierten los diversos fragmentos de mRNA a cDNA en un único tubo de ensayo para su posterior amplificación. Los productos de la amplificación son posteriormente analizados mediante CE, conociendo previamente la longitud constante, específica y diferencial de cada mRNA marcador de los fluidos correspondientes¹².

Además de los mRNA, se está investigando sobre la viabilidad de otro tipo de RNA, llamado miRNA, para la identificación de fluidos corporales. Estos RNA se diferencian fundamentalmente de los mRNA en que, en lugar de codificar para proteínas, son

reguladores de tipo negativo de la expresión génica y, por tanto, con potencial de representación de tejido específica, ya sea a nivel cualitativo o cuantitativo. Tales RNA sostienen grandes expectativas por su corto tamaño, lo cual los hace más resistentes a la degradación, así como su relativa abundancia en las células¹².

Tabla 2. miRNA propuestos para la identificación forense de distintos tejidos¹².

Targeted Tissue/ Body Fluid	Hanson et al. (2009)	Zubakov et al. (2010)		Courts and Madea (2011)		Wang et al. (2013)	
	SYBR Green qPCR	Microarray with LNA Capture Probes	TaqMan qPCR	Geniom Biochips	SYBR Green qPCR	TaqMan Array Human MicroRNA Cards	TaqMan qPCR
Peripheral blood	miR451, miR16	miR20a, miR106a, miR185	miR20a, miR106a, miR185, miR144	miR126, miR150, miR451	miR126, miR150, miR451	miR486, miR16	miR486, miR16
Menstrual blood	miR451, miR412	miR185*, miR144				miR214	miR214
Semen	miR135b, miR10b	miR943, miR135a, miR10a, miR507	miR943, miR135a, iR10a, miR507, miR891a			miR888, miR891a	miR888, miR891a
Saliva	miR658, miR205	miR583, miR518c*		miR200c, miR203,	miR200c, miR203,	miR138-2	
Vaginal secretion	miR124a, miR372	miR617, miR891a				miR124a	

5.3. Identificación de tejidos basada en DNA

Otro método actualmente investigado para la aplicación forense de identificación de tejidos es el estudio epigenético del DNA y, más concretamente, la metilación diferencial de este en función del origen. Dicho método requiere otra PCR especial, denominada *Methylation-Sensitive Restriction Endonuclease* PCR (MSRE-PCR), seguida de EC. Los principales puntos fuertes de este enfoque son la notable baja cantidad de muestra inicial necesaria (1ng aprox.) y la posibilidad de integrar el análisis de tejido y el perfil genético en una sola amplificación de multiplex PCR¹².

Finalmente, la última posibilidad explorada en este apartado será el análisis de DNA de las especies de bacterias residentes en el cuerpo humano y, en particular, de las específicas de cada zona del cuerpo. Este tipo de estudio generalmente emplea PCR, ya sea de DNA o RNA, de los genes 16S y 23S de rRNA, con gran variabilidad interespecie¹².

Tabla 3. Especies bacterianas propuestas para identificación forense de tejidos¹².

Targeted Tissue/ Body Fluid	Genus Name	Species Name	DNA/ RNA	Marker (Gene Name)	Reference	
Saliva	<i>Streptococcus</i>	<i>S. salivarius</i>	DNA	Glucosyltransferase (<i>GTF</i>)	Nakanishi et al. (2009)	
	<i>Streptococcus</i>	<i>S. mutans</i>	DNA			
Mouth- expired blood	<i>Streptococcus</i>	<i>S. mutans</i>	DNA	Glucosyltransferase (<i>GTF</i>)	Donaldson et al. (2010)	
	<i>Streptococcus</i>	<i>S. sanguinis</i>	DNA			
	<i>Streptococcus</i>	<i>S. gordonii</i>	DNA			
	<i>Streptococcus</i>	<i>S. salivarius</i>	DNA			
	<i>Streptococcus</i>	<i>S. sp.</i>	DNA	16S rRNA	Power et al. (2010)	
Vaginal secretion	<i>Lactobacillus</i>	<i>L. crispatus</i>	RNA	16S–23S rRNA intergenic spacer	Fleming and Harbison (2010), Jakubowska et al. (2013)	
	<i>Lactobacillus</i>	<i>L. gasseri</i> / <i>L. johnsonii</i>	RNA			
	<i>Lactobacillus</i>	<i>L. crispatus</i>	DNA	16S rRNA	Giampaoli et al. (2012)	
	<i>Lactobacillus</i>	<i>L. gasseri</i>	DNA			
	<i>Lactobacillus</i>	<i>L. iners</i>	DNA	16S rRNA	Akutsu et al. (2012)	
	<i>Lactobacillus</i>	<i>L. crispatus</i>	DNA			
	<i>Lactobacillus</i>	<i>L. jensenii</i>	DNA			
	<i>Lactobacillus</i>	<i>L. gasseri</i>	DNA			
	<i>Gardnerella</i>	<i>G. vaginalis</i>	DNA			
	<i>Atopobium</i>	<i>A. vaginae</i>	DNA			
	<i>Lactobacillus</i>	<i>L. sp.</i>	DNA	Whole genome sequencing	Benschop et al. (2012)	
	<i>Gardnerella</i>	<i>G. vaginalis</i>	DNA			
	Feces	<i>Bacteroides</i>	<i>B. uniformis</i>	DNA	RNA polymerase β -subunit (<i>rpoB</i>)	Nakanishi et al. (2013)
		<i>Bacteroides</i>	<i>B. vulgatus</i>	DNA		
<i>Bacteroides</i>		<i>B. thetaiotaomicron</i>	DNA	α -1-6 Mannanase (<i>MANBA</i>)		

5.4. Previsión

La identificación de tejidos y fluidos corporales vía DNA es un campo que ha progresado significativamente durante esta década. Algunos países como Países Bajos y Nueva Zelanda ya han incorporado los sistemas de análisis mencionados, y se espera un seguimiento progresivo del resto a medida que más kits comerciales aparezcan en el mercado¹².

Asimismo, se espera que los progresos realizados en el campo correspondan a minimizar la principal limitación de este tipo de estudio, esto es, la dificultad para distinguir tejidos biológicamente muy similares, como la saliva y la secreción vaginal. Es factible que esta limitación se vea solventada con la entrada de nuevas tecnologías como la NGS vía un

screening de los elementos diferenciales mucho más amplio que los realizados hasta ahora, basado en el estudio entero del transcriptoma y el epigenoma¹².

6. AUTOPSIA MOLECULAR

6.1. Introducción

La autopsia se puede definir como la examinación de un cuerpo después de su muerte para determinar la causa de la muerte (CM) o bien la naturaleza y extensión de cambios producidos por una enfermedad. Una autopsia forense pretende examinar un cuerpo post-mortem con la intención de especificar la causa de la muerte, así como su forma legal (natural, accidental, suicidio, homicidio o indeterminada)^{13,14}.

Las autopsias son realizadas por profesionales cualificados, ya sea un médico forense o un *coroner*, y juegan un rol capital en el procesamiento de crímenes con cadáveres, dado que el establecimiento de la CM y su forma legal constituyen el eje de la investigación. Además, también tienen importancia en aspectos médicos (para asegurar que la víctima fue tratada correctamente), materia relacionada con la salud pública, identificación de restos, y como fuente de datos para estimar la salud pública y el estilo de vida de una población^{13,14}.

Habiendo esclarecido esta serie de conceptos elementales, la autopsia molecular puede ser definida como una herramienta nueva, emergente en la patología molecular, consistente en el estudio de los mecanismos celulares bioquímicos y biofísicos como los factores más básicos de las enfermedades. Suele ser descrita como un conjunto de ensayos genéticos realizados en los tejidos extraídos durante la autopsia para confirmar la presencia de mutaciones genéticas asociadas a la muerte súbita e inesperada de ciertos cadáveres que no presentan signos morfológicos. Por este hecho, a este tipo de autopsia también se la conoce bajo los nombres “autopsia genética”, “autopsia silente” y también “autopsia líquida”^{13,14}.

6.2. La genética y la genómica en muertes súbitas

Los examinadores médicos no solamente investigan difuntos relacionados con violencia, suicidios, etc., sino que también se les pueden ser asignadas ciertas muertes por causas aparentemente o verdaderamente naturales. De entre estas últimas, la gran mayoría

corresponden a muertes súbitas e inesperadas que, a su vez, suelen ser causadas por trastornos cardiovasculares^{13,14}.

La capacidad de identificación de CM originadas por alteraciones génicas detrimenales, además de contribuir al esclarecimiento de la propia CM -especialmente en las “autopsias negativas”, donde no se han hallado marcadores fisiológicos-, puede ayudar a prevenir tragedias relacionadas en la familia del difunto, dada la naturaleza hereditaria de la enfermedad. Para ello, los estudios de esta índole están transicionando del estudio de unos pocos genes candidatos al de un número significativamente mayor o incluso todo el genoma (estudio genómico), donde el papel e interacción de los genes y de las vías metabólicas puede ser interpretada en conjunto. Para ello, la emergencia de la NGS como herramienta de preferencia es fundamental^{13,14}.

6.3. Estudio toxicogenético post-mortem

Otro aspecto del estudio forense para el cual el estudio masivo del material genético promete ser de notable utilidad es la toxicogenética. Dicha subrama está estrechamente relacionada con la farmacogenética (FGt) y la farmacogenómica (FGm), que no son sino el estudio de determinadas secuencias de DNA o de su totalidad, respectivamente, en relación con la respuesta particular del individuo a sustancias (generalmente drogas o fármacos). Esencialmente, tanto la FGt como la FGm se focalizan en comprender cómo la variación en los mecanismos moleculares relacionados con la metabolización afecta a la respuesta a drogas o fármacos administrados. Si se aplica este enfoque al estudio sobre por qué un cierto individuo ha sufrido una reacción adversa o incluso letal a determinadas sustancias, se obtiene la toxicogenética o, si el estudio es global, toxicogenómica^{13,14}.

En el ámbito forense la toxicogenética puede resultar, consecuentemente, de gran utilidad en casos de “autopsia negativa” con análisis complementarios toxicológicos donde, *a priori*, las hipotéticas sustancias presentes en el cadáver no debieran haber resultado letales. De nuevo, esta información, además de aportar información de sumo valor en el caso, puede ser usada para evitar tragedias recurrentes en familiares^{13,14}.

6.4. Microgenómica

La microgenómica puede definirse como el campo de análisis molecular cuantitativo de macromoléculas obtenidas de una célula o conjunto muy pequeño de células aisladas. Estas células se seccionan, extraen y examinan mediante técnicas de micromanipulación muy sensibles, la más habitual de las cuales son las Técnicas de Microdissección Láser (LMTs). Este conjunto de técnicas tiene como fundamento el corte literal de las células de interés del tejido o conglomerado de células de corresponda, para su posterior captura y manipulación, todo asistido vía microscopio. A su vez, la preparación de la extracción debería efectuarse con material que no interfiera en el posterior tratamiento de las células extraídas, en este caso, extracción del DNA, cuantificación y amplificación masiva^{13,14}.

La tecnología de extracción por LMT combinada con un análisis genético o genómico presenta aplicaciones diversas en el campo forense, incluyendo ensayos de paternidad en casos criminales, hallazgo de material genético en escombros, telógenos capilares individuales, así como la separación en mezclas de tipos celulares (principalmente tejido epitelial vs. espermatozoides) provenientes de agresiones sexuales^{13,14}.

Dicha tecnología también puede ser usada para tratar potenciales problemas en laboratorios de patología clínica, como contaminación de tejidos y mezclas o inserción de “tejido maligno” en un trasplante de órganos^{13,14}.

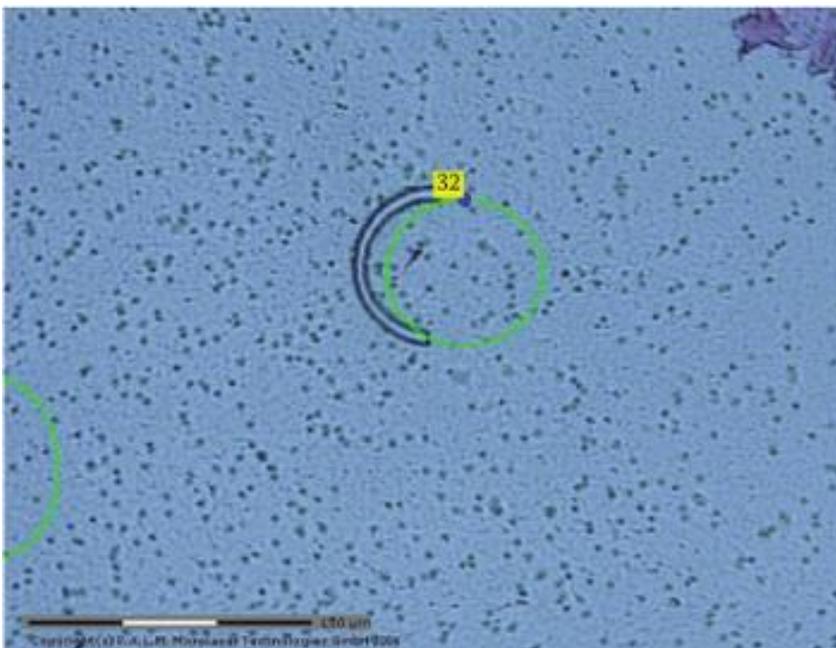


Figura 2.
Espermatozoide
montado en una
membrana de
naftalato de polietileno
listo para ser aislado
mediante un láser de
UV. Tinción NFR, 20
aumentos, PALM-
Zeiss¹⁴.

6.5. Consideraciones adicionales

Actualmente los ensayos moleculares no suelen realizarse durante las autopsias. No obstante, cabe esperar que la autopsia molecular devenga una herramienta complementaria para investigaciones forenses en un futuro próximo. Un enfoque grupal multidisciplinario por parte de examinadores médicos, investigadores de escenas del crimen, genetistas, epidemiólogos e investigadores ha permitido comprender mejor las causas genéticas de las muertes súbitas inesperadas, y se tiene que seguir investigando en esta dinámica si se pretenden realizar más progresos^{13,14}.

No cabe olvidar aspectos importantes a tener en cuenta a la hora de realizar ensayos genéticos como son el consentimiento familiar, la privacidad, el estrés emocional, la incertidumbre a nivel de resultados de ciertas variantes genéticas todavía no contrastadas, y la posibilidad de desvelar falsas paternidades. Asimismo, se espera que el consejo genético adopte un rol importante en análisis moleculares post-mortem y asesoramiento de familiares^{13,14}.

7. PREDICCIÓN DE CARACTERÍSTICAS FÍSICAS

7.1. Introducción

La variación en distintos loci del genoma tiene múltiples aplicaciones en el campo de la genética. De estas, la que más interesa en este apartado es la capacidad de predicción de características físicas. Esta (de momento) pequeña guía visual adquiere especial importancia cuando el lanzamiento de perfiles genéticos obtenidos contra la base de datos da resultados negativos, o bien como herramienta auxiliar para identificar restos de personas desaparecidas^{15,16}.

La base de variación molecular, por tanto, no se centra en este caso en los STRs usados en los perfiles, sino en la variación a nivel de lectura de la secuencia de DNA, de la cual presentan especial interés los *Single Nucleotide Polymorphisms* (SNPs). Los SNPs pueden tener consecuencias funcionales ocasionadas por una gran variedad de causas e, independientemente de cuál sea, la identificación de un SNP asociado a un rasgo

fenotípico pasa por la comparación del genotipo con el fenotipo, juntamente con un análisis estadístico para determinar la significatividad de la asociación^{15,16}.

Se puede pensar que la predicción de las características fenotípicas más básicas requiere de pocos tests y ofrecen información precisa y detallada, pero la realidad dista mucho de estos ideales, puesto que la inmensa mayoría de variables se consideran complejas debido a la interacción entre múltiples genes, así como con el ambiente. Por este hecho es razonable pensar que únicamente técnicas modernas con capacidad de análisis y procesamiento de mucha información pueden satisfacer las necesidades de esta aplicación auxiliar en el campo forense, la cual presenta un enorme potencial y margen de mejora. La investigación y progreso de este campo se centra casi exclusivamente en los llamados *Genome-Wide Association Studies* (GWAS), en los que se tiene en cuenta la variación de muchísimos marcadores (SNPs) distribuidos por todo el genoma^{15,16}.

7.2. Pigmentación y altura

La pigmentación de iris, cabello y piel ha sido, de momento, el único factor del cual se ha hecho investigación profunda y, pese a esto, la precisión de las predicciones todavía no alcanza el 100%. Tal hecho puede justificarse alegando que hasta 100 genes intervienen en la síntesis de melanina y que, además, dicha síntesis puede verse regulada en función del tejido corporal y es influida por distintos factores ambientales comunes, como enfermedades, la edad, radiación UV y drogas^{15,16}.

Por otro lado, la altura es probablemente la característica corporal cuantitativa que más puede explicarse por el componente genético, alrededor del 80-90%. Los factores ambientales que más afectan en este caso son la dieta y la morbosidad. Diversos GWAS han demostrado que existen diversos centenares de loci relacionados con este factor, a mayor parte de los cuales pertenecen a vías metabólicas relativas al crecimiento esquelético^{15,16}.

Tabla 4. Tipos de variación para los SNPs seleccionados en la pigmentación¹⁵.

Protein Function	Gene	SNP	Variation Type/Location
Possible involved in protein ubiquitination	HERC2	rs12913832	Located in predicted transcription factor binding site of OCA2
		rs1129038	3' UTR
		rs11636232	Synonymous
Integral membrane protein involved in transport of small molecules	OCA2	rs1800407	Missense
		rs1545397	Intron
Transporter protein that mediates melanin synthesis	SLC45A2	rs16891982	Missense
Potassium-dependent sodium/calcium exchanger protein	SLC24A5	rs28777	Intron
		rs1426654	Missense
Potassium-dependent sodium/calcium exchanger protein	SLC24A4	rs12896399	5' region
		rs2402130	Intron
Melanosomal enzyme involved in melanin biosynthetic pathway	TYRP1	rs1408799	5' region
		rs683	3' UTR
Involved in conversion of tyrosine to melanin	TYR	rs1393350	Intron
Regulation of interferons	IRF4	rs1042602	Missense
		rs12203592	Intron
Exocyst complex	EXOC2	rs4959270	3' region (between IRF4 and EXOC2)
Receptor protein that controls melanogenesis	MC1R	rs201326893	Nonsense
		N29insA	INDEL
		rs1805006	Missense
		rs11547464	Missense
		rs1805007	Missense
		rs1805008	Missense
		rs1805009	Missense
		rs1805005	Missense
		rs2228479	Missense
		rs1110400	Missense
Signaling protein in pigmentation	ASIP	rs885479	Missense
		rs6119471	5' region, predicted transcription factor binding site
Ligand of tyrosine kinase receptor	KITLG	rs2424984	Intron
		rs12821256	5' region
Membrane protein with GPI anchor activity	PIGU	rs2378249	Intron

7.3. Elaboración de *predictors* y validación

Como se ha mencionado anteriormente, las características físicas son notablemente complicadas de predecir con detalle. No obstante, para algunas de estas se ha descubierto que un número particular de SNPs no demasiado elevado puede llegar a estimar el fenotipo resultante con una significación estadística aceptable. De estos conjuntos de SNPs provienen los tests usados en los laboratorios forenses llamados *predictors*, de los cuales los más perfeccionados hasta la fecha corresponden al color del ojo y al color de piel^{15,16}.

Tabla 5. Eye-Color-Predictor-2¹⁵.

Gene	SNP ID	Genotype	Eye Color (Predicted)
HERC2	rs12913832	G/G	Not brown
		G/A	Not blue
		A/A	
HERC2	rs12913832	A/A	Not blue
IRF4	rs12203592	T/T	
HERC2	rs12913832	G/A	Green
IRF4	rs12203592	T/T	
HERC2	rs12913832	G/G	Green
SLC45A2	rs16891982	C/C	
HERC2	rs12913832	A/A or G/A	Brown
SLC45A2	rs16891982	C/C	
HERC2	rs12913832	A/A or G/A	Brown
OCA2	rs1545397	T/T	
HERC2	rs12913832	A/A or G/A	Brown
MC1R	rs885479	A/A	
HERC2	rs12913832	A/A or G/A	Brown
ASIP	rs6119471	G/G	

A su vez, la validación de estos tests predictivos permite definir las tasas de error y de resultados inconcluyentes, incrementando de esta forma el nivel de confianza. La validación se basa simplemente en la comparación de la predicción con el fenotipo real correspondiente^{15,16}.

Tabla 6. Validación del Skin-Color-Predictor en 805 muestras de diversas poblaciones¹⁵.

Population	Skin Color (Self-Reported)	Skin Color Predicted	Errors in Prediction	Inconclusive
AA (33 ^b + 10 ^b)	Medium (4 + 0)	Not white (1 + 0)	0	3 + 0
	Dark (29 + 10)	Not white (11 + 4)	0	18 + 6
SA (25 + 2)	Medium (25 + 0)	Not dark (2 + 0)	0	23 + 0
	Dark (0 + 2)	Not dark (0 + 1)	0 + 1	0 + 1
EA (22 + 13)	Light (20 + 11)	Not dark (7 + 2)	0	13 + 9
	Medium (2 + 2)	Not dark (0 + 1)	0	2 + 1
E (379 + 178)	Light (379 + 178)	Not dark (345 + 169)	0	34 + 9
Mixed (95 + 48)	Light (44 + 29)	Not dark (18 + 21)	0	26 + 8
	Medium (35 + 13)	Not dark (6 + 1)/not white (3 + 1)	0	26 + 11
	Dark (16 + 6)	Not white (3 + 2)/not dark (2 + 1)	2 + 1	11 + 3
Total	554 + 251		4	156 + 48
%			1% ^b	Range 28% to 19%

7.4. Direcciones futuras

Se espera que bastante pronto los tests descriptivos de DNA sean incorporados como rutina en los laboratorios de genética forense. También se espera que en un futuro próximo los *predictors* se vean mejorados y sean capaces de describir características físicas con más detalle^{15,16}.

Entre las mejoras en este campo cabe mencionar, sobre todo, la inclusión de cada vez más características, preferentemente representadas por pocos marcadores. Este sería el caso de la recientemente descubierta variación en la cera de las orejas, húmeda o seca, explicada íntegramente por un solo SNP. Otros elementos fenotípicos prometedores son el grosor y la ondulación del pelo, la calvicie y la pecosidad^{15,16}.

La última tecnología en este campo pertenece a la relectura constante de la secuencia de diversas partes del genoma para obtener los rasgos faciales del sujeto. El perfeccionamiento e implantación de esta aplicación supondrá la herramienta auxiliar definitiva a la hora de predecir características físicas de interés forense¹⁷.

8. GENEALOGÍA EN LA ERA GENÓMICA

8.1. Introducción

La genealogía es el estudio de la historia familiar, el cual incluye el rastreo del linaje del individuo atrás en el tiempo a través de la documentación vital de cada ancestro como fechas y lugares de nacimiento, de defunción y matrimonios. La investigación genealógica depende a menudo de registros históricos escritos para determinar las relaciones familiares, lo cual suele conllevar toparse con lo que se conoce en inglés como *brick wall*, literalmente “muro de ladrillos”. Las causas más comunes son la destrucción de los documentos o directamente la ausencia de estos. Para solventar estas limitaciones se introdujo en los 90 el enfoque molecular¹⁸.

Inicialmente se usaron STR del cromosoma Y con el fin de contrastar los individuos supuestamente emparentados. Posteriormente se añadieron otros marcadores como las regiones hipervariables (HVR) del cromosoma mitocondrial (mtDNA) y el propio genoma mitocondrial completo, hasta pasar a marcadores autosómicos, adoptando inicialmente un estudio basado en *microarrays* o *microchips*¹⁸.

Los perfiles genéticos obtenidos con estos fines, los cuales son fundamentalmente creados por empresas a petición del cliente, son depositados en bases de datos cada vez más completas, creando de esta forma repositorios de variación humana considerables

que pueden ser usados en diversos campos de estudio, entre los cuales se encuentra la genética forense¹⁸.

8.2. Testeo de DNA haploide

El DNA haploide, esto es, mtDNA y cromosoma Y, presenta un interés especial en el estudio genealógico debido a dos características que no posee el DNA autosómico. La primera es su patrón de herencia estricto, holándrica para el cromosoma Y y matrilineal para el mtDNA; mientras que la segunda es la ausencia de recombinación. Tal combinación de fenómenos permite a estos tipos de DNA ser usado como un reloj molecular que permite rastrear linajes no solo en la historia reciente, sino desde un pasado muy lejano¹⁸.

8.2.1. Cromosoma Y

Además de ser el primer tipo de DNA usado en la genealogía molecular, el cromosoma Y sigue siendo el favorito de los especialistas para analizar linajes paternos. La herencia de este cromosoma se corresponde con la del apellido en la mayoría de sociedades occidentales, lo cual supone una gran ayuda para los genealogistas. Estos estudios de relación Crom. Y – Apellido están bastante extendidos, permitiendo a parientes más o menos lejanos descubrir los orígenes de su apellido y rastrear los linajes familiares comunes cada vez más hacia el pasado, así como descartar pistas falsas¹⁸.

Otra prueba de la popularidad del cromosoma Y es el tamaño de sus bases de datos, las cuales son las más grandes y sobre las que se indaga con mayor frecuencia¹⁸.

Los ensayos más comunes relativos al cromosoma Y corresponden al análisis de 12-111 STRs localizados en la región no recombinante del cromosoma para establecer relaciones de ascendencia próxima. No obstante, muchos tests también incluyen un componente para el análisis del haplogrupo para ofrecer una perspectiva ancestral muy lejana¹⁸.

El futuro en el estudio de este cromosoma pasa por la elaboración de perfiles informativos mucho más completos, de escala del cromosoma íntegro, basados en el análisis de SNPs¹⁸.

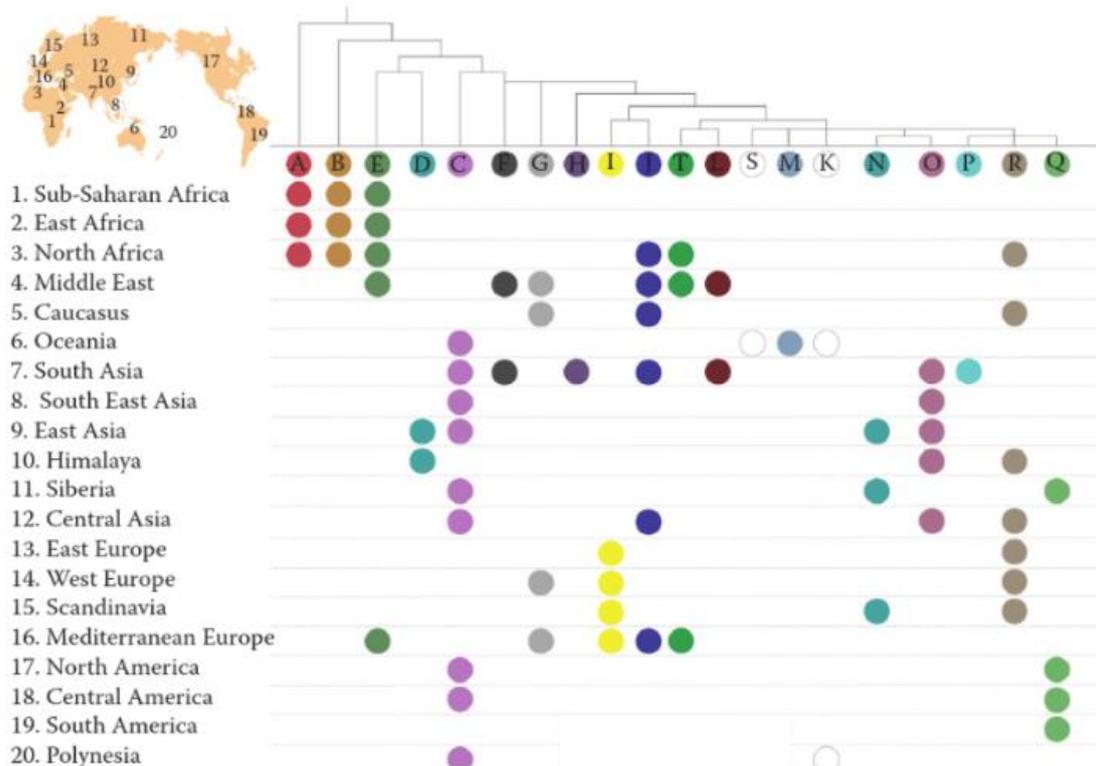


Figura 2. Relaciones filogenéticas entre los 20 principales haplogrupos del cromosoma Y y sus asignaciones geográficas típicas¹⁸.

8.2.2. mtDNA

El patrón de herencia del mtDNA también es estricto, siendo pasado de la madre a todos sus descendientes, aunque únicamente las hijas de esta serán las que continúen transmitiendo ese mtDNA. Los análisis de mtDNA generalmente se usan para confirmar o desmentir una supuesta relación materna en lugar de identificar el progenitor reciente más próxima en una ascendencia no lejana en el tiempo. Cuando se da el segundo caso, el análisis más usado es el de las HVR del mtDNA, puesto que su tasa de mutación más alta que el resto del cromosoma mitocondrial permite maximizar la variabilidad y reducir el coste al máximo, logrando así que los ensayos sean viables¹⁸.

No obstante, también existe la opción de secuenciar todo el mtDNA, más cara, pero que ofrece una mayor resolución tanto en la comparación de perfiles como en la elaboración de haplogrupos. De forma similar al cromosoma Y, los haplotipos de mtDNA son almacenados en bases de datos y frecuentemente se encuentran asociados a información biogeográfica y de ascendencias particulares¹⁸.

Asimismo, el futuro del análisis del mtDNA y, en general, del DNA haploide, consiste en el incremento de marcadores con el fin de poder elaborar perfiles y haplotipos mucho más refinados¹⁸.

8.3. Testeo de DNA autosómico

Los estudios del cromosoma Y y del mtDNA resultan muy útiles para los estudios de ascendencia y son muy fáciles de interpretar, pero se ven limitados a herencias muy limitadas que constituyen únicamente una pequeña parte del árbol genealógico completo. Además, la historia asociada únicamente al linaje paterno o materno puede que esté condicionada¹⁸.

Por este conjunto de factores, juntamente con el progreso tecnológico, se está incorporando progresivamente el análisis del DNA autosómico. Esto supone una perspectiva significativamente más amplia, puesto que el mtDNA y el cromosoma Y solamente constituyen aproximadamente el 2% del genoma¹⁸.

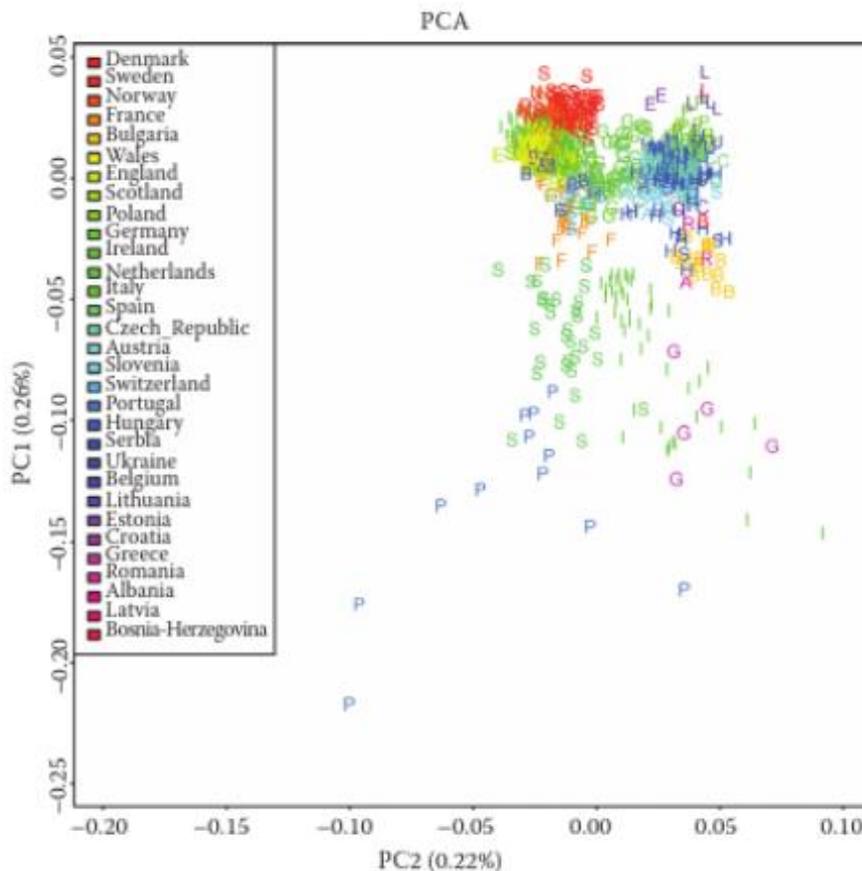


Figura 3. Inferencia del origen ancestral en base a datos autosómicos, 135.000 SNPs de 1000 individuos de origen europeo¹⁸. Los dos componentes principales muestran cómo individuos de la misma zona geográfica tienden a ser genéticamente similares.

Las principales aplicaciones que ofrece el estudio de DNA autosómico son 2: la determinación de las poblaciones históricas que han contribuido significativamente en la aportación de material genético a un individuo, y la identificación de familiares. Ambos usos pueden lograrse también con el análisis del DNA haploide, pero los resultados son más precisos y fiables con el DNA autosómico¹⁸.

8.4. Discusión

En este apartado se han visto los dos aspectos principales de estudio en la genealogía, inferencia del origen ancestral o bien identificación de familiares cercanos. Visto desde una perspectiva más amplia, puede considerarse que ambos temas son, en realidad, el mismo, pero con escalas temporales distintas¹⁸.

El avance de las tecnologías relativas a los dos propósitos continúa cerrando el espacio que los considera distintos, esperando como objetivo final el rastreo de fragmentos particular de DNA a lo largo de espacio y tiempo mientras se captura toda la historia de su herencia, desde una gran población continental primigenia hasta una familia de hace pocas generaciones¹⁸.

9. SECUENCIACIÓN DE ÚLTIMA GENERACIÓN Y LA GENÉTICA FORENSE

9.1. Introducción a la secuenciación Nanopore

De entre las nuevas técnicas de secuenciación, correspondientes ya a las de tercera generación, la secuenciación Nanopore, desarrollada por Oxford, es la que presenta más opciones de devenir la técnica de preferencia cuando se realicen análisis genéticos forenses. Su secuenciador, denominado MinION, cabe en un bolsillo, puede alimentarse con un cable USB y su adquisición ronda actualmente los 1000\$. No obstante, el precio del secuenciador no es aquí el factor económico limitante, como en los secuenciadores de primera y segunda generación; sino sus *flow cells*, de naturaleza desechable. Se espera que este componente de la secuenciación disminuya considerablemente su precio a lo largo de los siguientes años. Adicionalmente, el procesador de muestras VolTRAX, también portátil, permite convertir de forma rápida y programable las

muestras biológicas en una librería compatible con el análisis Nanopore, permitiendo que todo el *workflow* sea ejecutado sin necesidad de intervención humana^{8,19}.

El *software* propio de Oxford Nanopore Technologies permite, a su vez, realizar análisis sobre la marcha, con un poder computacional de un teléfono inteligente (como mínimo) y conexión a Internet como únicos requisitos. Este hecho hace posible la secuenciación de muestras biológicas virtualmente en cualquier lugar, y también en sitios donde el acceso a laboratorios está limitado, como en países en estado de desarrollo precario o zonas que han sufrido alguna catástrofe. La posibilidad de trabajar a instancias independientes de un laboratorio también desbloquea la identificación de personas en fronteras, controles policiales, estaciones de policía, etc., sin la necesidad de personal altamente cualificado^{8,19}.

Si se consideran los requisitos moleculares para este tipo de secuenciación, es necesaria la incorporación de un fragmento *leader* en un extremo del segmento del DNA y otro de tipo *hairpin* en el otro extremo para poder preparar la librería. A pesar de que teóricamente no se requiera una PCR para este tipo de secuenciación, en su adaptación al campo forense sí se realiza una PCR para amplificar los loci forense hasta que se obtiene la cantidad de DNA adecuada para la preparación de la librería (1µg). Además, la longitud mínima analizable actual por este método son 100bp, requisito que no cumplen algunos marcadores. Para solucionar esta limitación se suelen concatenar los amplicones de PCR en un paso de ligación previo al de la preparación de la librería^{8,19}.

En este capítulo se contemplan dos aplicaciones de la secuenciación Nanopore: la elaboración de perfiles clásicos analizando STRs, y la elaboración de otro tipos de perfil basado en SNPs específicamente trialélicos^{8,19}.

9.2. Elaboración de perfiles de STR

Para evaluar la capacidad analítica de Nanopore para STRs se compararon los resultados del mismo kit multiplex de PCR de 14 loci, aplicado al DNA de un mismo donante. Los datos fueron contrastados con las bases de datos para los loci y se prestó atención especial al marcador sexual, en este caso, el gen de la amelogenina¹⁹.

Tras el pertinente análisis de calidad y selección de los *reads*, se observó en los 13 loci autosómicos la siguiente distribución, mostrada en la figura posterior, en la que se

aprecia cómo la mayoría de *subreads* se corresponden con alelos verdaderos (p.e. D3S1358), pero un nombre relativamente grande de estos se alinea también con otras secuencias que no son los alelos verdaderos (p.e. SE33) creando, consiguientemente, ruido. También se observa cómo los loci heterocigotos muestran buen balance en algunos loci (p.e. D3S1358), mientras que en otros no se logra (p.e. TPOX).

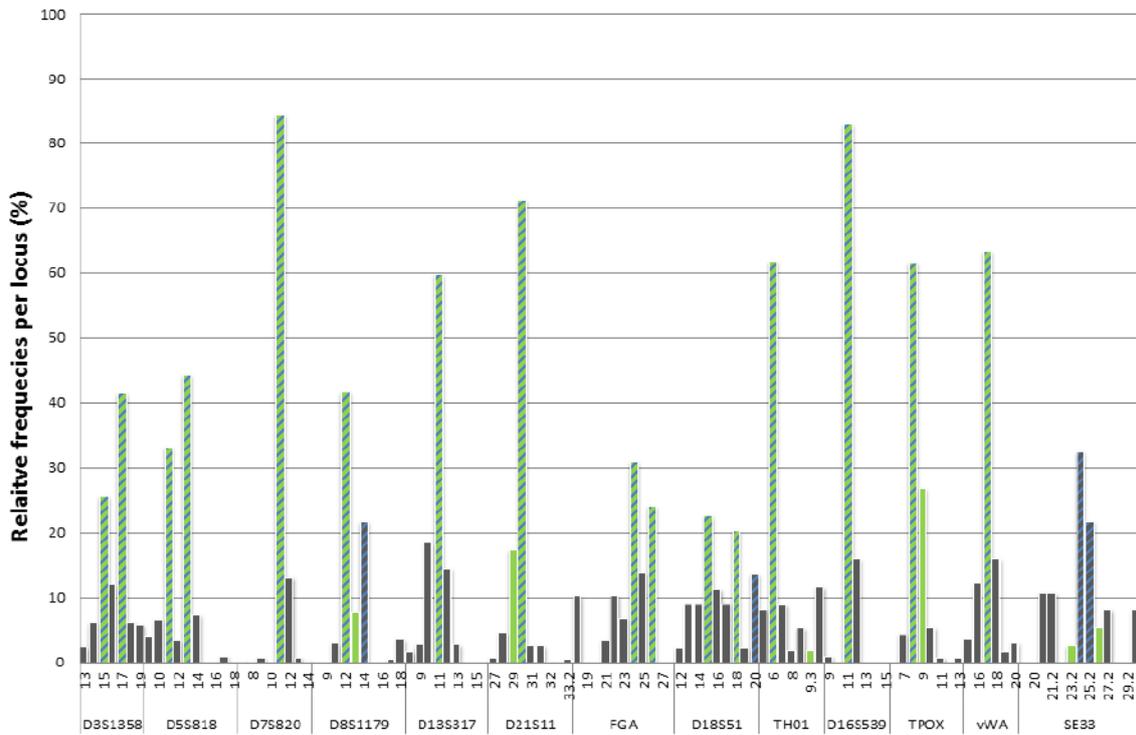


Figura 4. Frecuencia relativa de los subreads mapeados de forma única, en función del locus. El verde indica los alelos verdaderos; las bandas azules, cuando el alelo se considera el observado. Verde con rallas azules corresponde a alelos correctamente genotipados. Verde sin rallas azules implica un error de identificación del alelo correcto (*drop-out*), mientras que gris con rallas azules indica que se observó un alelo incorrecto (*drop-in*)¹⁹.

A su vez, el análisis del gen de la amelogenina permitió observar la siguiente fracción de resultados, donde se observan numerosos *mismatches* en regiones idénticas entre los dos tipos de amelogenina, además de algunas *indels*; no obstante, la más notoria de las cuales, la delección de 6bp entre los 40 y los 60, es conocida y esperada.

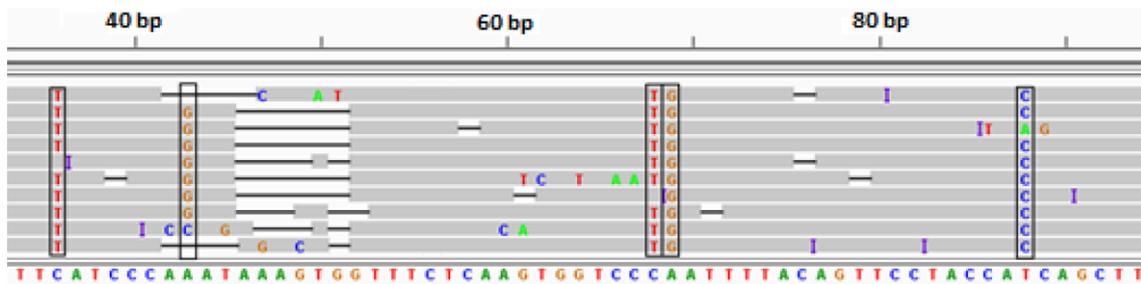
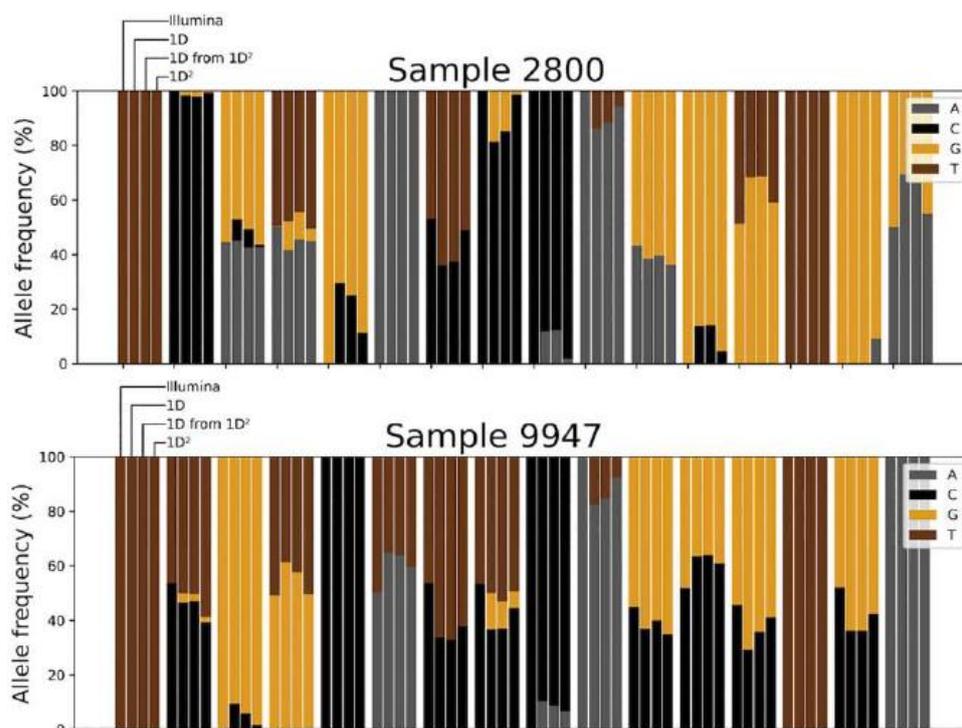


Figura 5. Subreads de la Amelogenina X contrastados contra la secuencia de Amelogenina Y de referencia. Los rectángulos negros indican las variaciones conocidas. Los otros nucleótidos sin coincidencia se muestran separadamente con su respectiva letra¹⁹.

8.3. Genotipaje de SNPs trialélicos

En este caso se genotiparon 3 controles positivos y 2 muestras de la *German DNA Profiling (GEDNAP)*, que fueron usados como casos problema. Todos los perfiles obtenidos con la secuenciación Nanopore fueron comparados con los de referencia, obtenidos mediante secuenciación Illumina. Se aplicaron nuevamente la filtración y selección de *reads* de calidad más alta⁸.

Los resultados se ven sintetizados en la siguiente figura. En esta, se aprecian cómo las secuenciaciones realizadas con Nanopore son distintas de las de referencia, convirtiendo algunos loci homocigóticos en heterocigóticos y heterocigóticos en loci con 3 SNPs -lo cual es imposible-, ambos cambios generados por errores en la secuenciación.



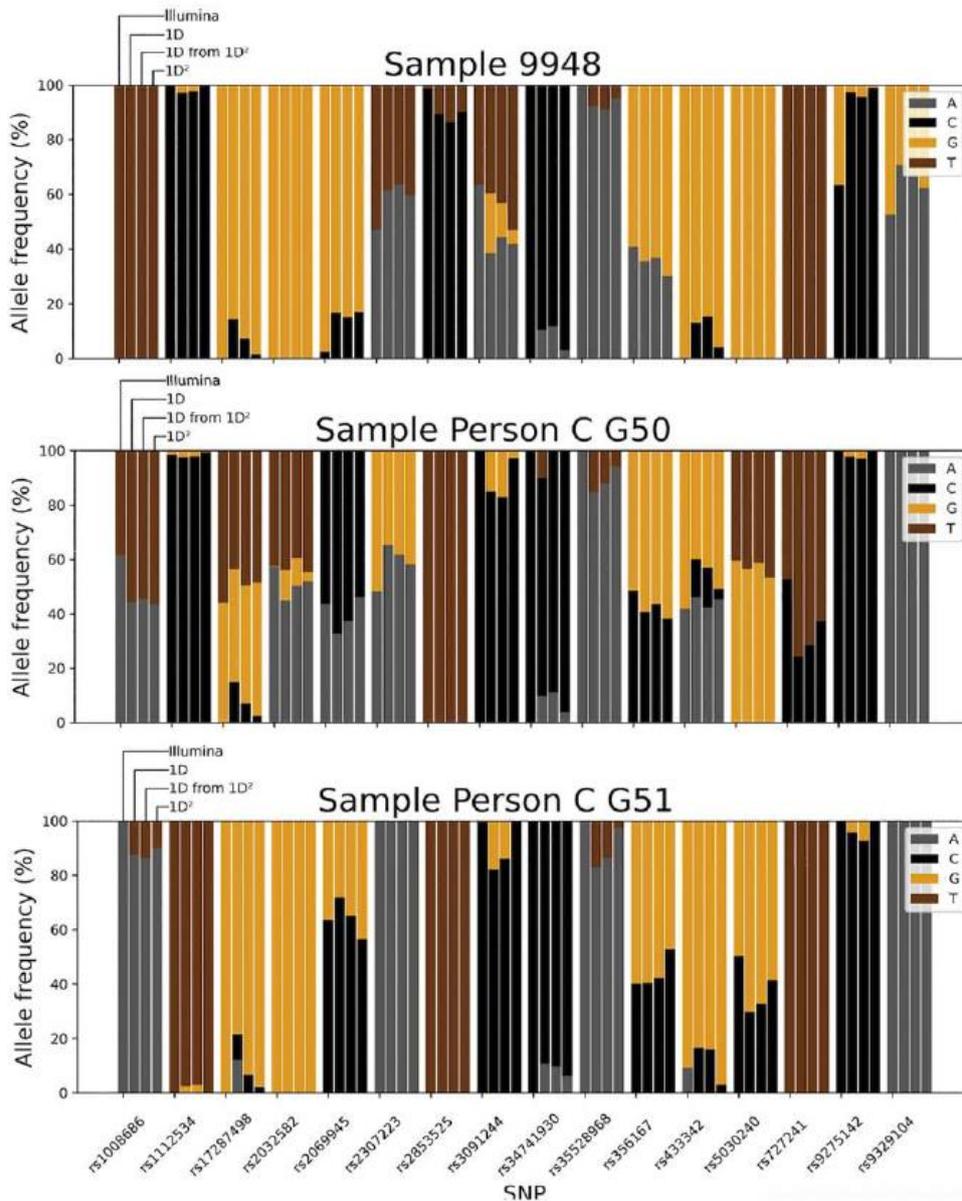


Figura 6. Frecuencia relativa de cada nucleótido para cada SNP y para cada perfil. Las 4 barras para cada SNP corresponden a los distintos tipos de secuenciación realizada⁸.

9.4. Discusión

La elaboración de perfiles basados en secuencias requiere un alineamiento único de los *reads* respecto a una secuencia de referencia. Un porcentaje muy elevado de los *subreads* se mapean contra más de un alelo de la base de datos, pero, por suerte, existen métodos para filtrar estos *reads* ambiguos. La mayoría restante se corresponde con los alelos verdaderos, aunque también se han visto casos donde los *subreads* siguen asociándose a alelos erróneos. Estos alineamientos incorrectos pueden ser debidos a la gran cantidad de errores en la secuenciación producidos por el secuenciador MinION, y

también a la gran similitud de los alelos de un mismo locus presentes en la base de datos^{8,19}.

Por lo que respecta a los errores en la secuenciación, se observaron, a parte de los errores aleatorios en la incorporación de pares de bases, un error sistemático consistente en deleciones parciales en regiones homopoliméricas. También se apreció una distribución amplia en la longitud de los fragmentos, puesto que MinION es propenso a generar errores de tipo inserción-delección o *indel*^{8,19}.

Con la finalidad de solventar los problemas de secuenciación que presenta MinION y seguir generando perfiles genéticos forenses viables, los autores de los perfiles basados en STRs proponen pasar a usar como marcadores los SNPs¹⁹.

10. OTRAS TECNOLOGÍAS DE ANÁLISIS GENÉTICO FORENSE

A parte de los progresos realizados con la NGS, también se ha estudiado la viabilidad de otras tecnologías con el fin de mejorar la velocidad y resolución de los análisis genéticos enfocados al ámbito forense. No obstante, tanto la NGS como el enfoque tradicional PCR-EC son susceptibles a verse beneficiados por estos progresos tecnológicos⁵.

10.1. Análisis rápido de DNA

La progresiva mejora de las PCR vía polimerasas especializadas y cambios de temperatura más rápidos ha permitido llevar amplificaciones multiplex a incluso 30 minutos. Asimismo, de forma similar al secuenciador MinION, se intenta, de forma general, minimizar el tamaño y maximizar la portabilidad del instrumental. La reducción en tamaño también suele conllevar el beneficio extra de la reducción del tiempo de análisis. Este sería el caso de las EC en miniatura⁵.

No obstante, cabe mencionar que todavía no se ha logrado establecer rutinas de “análisis rápido” de DNA con la resolución suficiente para ser considerado en casos forenses, a pesar de que muchos grupos de investigación estén trabajando extensamente en esta área⁵.

10.2. Espectrometría de masas

La espectrometría de masas (MS) es una técnica muy versátil que se centra en la detección de iones y en su relación masa-carga (M/Z). Debido a que los iones son separados en condiciones de vacío el tiempo de análisis puede ser extremadamente rápido, de cuestión de segundos. Si se combina con una preparación de muestras robótica, la MS ofrece la posibilidad de numerosas muestras de DNA de forma rápida y automatizada⁵.

Con esta técnica se está caracterizando la masa real de la molécula de DNA en cuestión, lo cual supone un análisis más preciso que una comparación relativa tamaño como es el caso de la electroforesis. Para inducir el estado gaseoso en las moléculas de DNA se han registrado usos tanto de MALDI como de ionización por *electrospray* (ESI). Actualmente, MS basados en *ESI-Time Of Flight* tienen un rango de análisis de secuencias de hasta 250bp, el cual engloba la inmensa parte de STR que se analizan rutinariamente⁵.

Mediante el análisis de la masa total de un STR o de un fragmento de mtDNA puede saberse su composición de bases e identificar diferencias entre secuencias internas. De esta forma, STR aparentemente homocigotos en electroforesis pueden ser subdivididos en alelos únicos cuando existan polimorfismos de secuencia. No obstante, es imposible saber la posición exacta de la variación en la secuencia⁵.

10.3. Automatización

La automatización en los laboratorios es un tema importante, puesto que la demanda de análisis forenses de DNA está aumentando. La automatización contribuye a la disminución de la tasa de error e incrementa la eficiencia. Reducir en tiempo que los analistas dedican al procesamiento de la muestra les permite centrarse más en la examinación de pruebas y en la interpretación de resultados. Además, se reducen potenciales errores a través de un mejor seguimiento y manipulación⁵.

La automatización en los laboratorios está creciendo en distintas áreas, desde manipulación de líquidos hasta rastreo de muestras y análisis de datos. Un ejemplo de la efectividad que puede conllevar este factor es la capacidad del FBI en para procesar 408.000 muestras de DNA en el margen de un año, entre 2009 y 2010⁵.



Figura 7. Estación de trabajo robótica Tecan Freedom EVO 150, usada en la preparación de muestras para electroforesis capilar⁵.

10.4. Consideraciones finales y resumen

Además de la necesidad de que los métodos analizados en el ámbito forense sean validados en cuanto a su precisión y reproducibilidad, se necesita también una entrada de resultados comparables con los ya existentes. Por estos dos motivos, la adopción de nuevas tecnologías en el campo forense requiere de tiempo y necesita aplicarse de forma gradual⁵.

Asimismo, para la implementación de alguna nueva técnica y sustitución de los *Gold Standard*, se precisa que ésta ofrezca ventajas significativas y resulte más económica, con el fin de que se considere reformar la cantidad ingente de información almacenada en las bases de datos forenses⁵.

Un breve resumen de los principales beneficios y limitaciones de las técnicas expuestas en este capítulo, juntamente con la NGS y otro tipo de secuenciación poco usada, la pirosecuenciación, se muestra en la siguiente tabla:

Tabla 7. Beneficios y limitaciones de diversas tecnologías de análisis de DNA⁵.

Technology	Benefits	Limitations
Rapid DNA analysis	Potentially portable system; faster results that could enable new applications of DNA testing	Some compromises in DNA result quality may be required
Mass spectrometry base composition	Enables measurement of both length and nucleotide variability with compatibility to existing STR databases; less expensive than sequencing for mtDNA	Higher instrument cost and more difficult to maintain than a CE system; unable to determine the location of a sequence change; cannot multiplex to a high-level (would likely require more DNA)
Pyrosequencing	Enables rapid sequence-based analysis of STRs or other PCR products	Cannot multiplex (would likely require more DNA); cannot cope with complex and long repeats
Next-generation DNA sequencing	Enables recovery of sequence information from highly degraded DNA (e.g., ancient DNA samples)	Significant increase in cost compared to current methods; short read lengths limit ability to capture STR repeat variation
Automation	Improved efficiency and speed in processing samples (particularly with expert systems data interpretation)	Initial cost for equipment and software

11. DISCUSIÓN GENERAL

La cuestión sobre la incorporación rutinaria de la NGS a los laboratorios forenses ya no reside en si esta tendrá lugar o no, sino en el cuándo, sobre todo para cada país.

Es indiscutible que, por lo menos al parecer de uno, las ventajas que ofrece la NGS, ya sea como herramienta usada en análisis centrales o auxiliares, provocan que resulte una herramienta extremadamente útil para cualquier laboratorio forense, tanto en el presente como especialmente en el futuro próximo. Los estados, en consecuencia, deberían plantearse hacer la inversión necesaria para que sus expertos forenses puedan maximizar la información derivada de las muestras biológicas y, en consecuencia, se resuelvan los conflictos de manera cada vez más fiable.

No obstante, a pesar de que prácticamente todos los autores consultados defiendan que el precio del instrumental y coste de los *runs* será cada vez más bajo, el aspecto económico va a ser, de todas formas, un factor limitante en muchos casos. El presupuesto destinado a los laboratorios forenses proviene de fondos públicos, la cual cosa limita sustancialmente la posibilidad de modernizar el instrumental, sobre todo si se compara con proyectos de naturaleza privada. Este factor se ve todavía más acentuado si se toman como referencia países con un PIB o PIB per cápita moderado o

bajos, como sería el caso de España; en comparación con los países anglosajones, principalmente EEUU, de donde provienen la mayoría de libros y artículos consultados para este trabajo. Suponiendo que se hayan desarrollado nuevos sistemas más eficaces y más económicos, resulta mucho más complicado que dichos sistemas se apliquen a los países del primer tipo, debido a que la inversión inicial necesaria quede fuera del presupuesto destinado a los Ministerios o Departamentos de Justicia. Es más probable, por tanto, que su implementación se realice de forma mucho más paulatina.

Con respecto a un potencial nuevo *Gold Standard* en la elaboración de perfiles genéticos, cabe esperar dos posibilidades. La primera y más probable, que la secuenciación de 2ª generación acabe por resultar económicamente factible y acabe imponiéndose, pudiendo realizar perfiles tanto de STRs como de SNPs. La segunda, que la secuenciación de 3ª generación, principalmente Nanopore, consiga reducir su precio y optimice su eficacia lo suficiente para que esta devenga el Gold Standard, saltándose de esta forma la de 2ª generación. Esta opción es menos probable debido a su mayor distancia del objetivo en comparación con los secuenciadores de 2ª generación, y por el hecho de que sería necesario cambiar el perfilado a únicamente SNPs, lo cual, a fecha de hoy, es extremadamente poco factible debido a los cambios masivos que requerirían las bases de datos y la pérdida de utilidad de todos los perfiles ya realizados basados en STRs.

12. CONCLUSIONES

En consecuencia, tras el presente estudio se puede extraer que la *Next-Generation Sequencing* será una herramienta imprescindible para los laboratorios de genética forense, y su implementación dependerá exclusivamente de la capacidad económica de los estados correspondientes. Este mismo carácter de necesidad es extensible a todas las aplicaciones descritas a lo largo del trabajo, pues cada una de estas puede aportar información vital para la resolución correcta de los casos.

Asimismo, la secuenciación de última generación se encuentra todavía en un estadio demasiado prematuro para ser considerado en los laboratorios forenses, pero presenta una serie de características muy prometedoras.

Por último, es imperativo que, además de realizar avances en métodos analíticos, se implementen también las mejoras realizadas en otras fases del proceso de análisis, como la automatización del tratado de muestras o herramientas más potentes de procesamiento de datos, con el fin de optimizar la efectividad de proceso y no sufrir cuellos de botella.

13. REFERENCIAS

1. Goodwin W, Linacre A, Hadi S. Introduction to forensic genetics. In: Goodwin W, Linacre A, Hadi S, eds. *An Introduction to Forensic Genetics*. 2nd ed. John Wiley & Sons; 2011:2016.
2. Carracedo A. Forensic Genetics: History. In: *Encyclopedia of Forensic Sciences*. ; 2013:206-210. doi:10.1016/B978-0-12-382165-2.00037-4
3. Jeffreys AJ, Brookfield JFY, Semeonoff R. Positive identification of an immigration test-case using human DNA fingerprints. *Nature*. 1985;317(6040):818-819. doi:10.1038/317818a0
4. Machado H, Granja R. DNA Technologies in Criminal Investigation and Courts. In: *Forensic Genetics in the Governance of Crime*. ; 2020:114.
5. Butler JM. New Technologies and Automation. In: Butler JM, ed. *Advanced Topics in Forensic DNA Typing: Interpretation*. Academic Press; 2014:608.
6. Calloway CD, Erlich H. Evolving Technologies in Forensic DNA Analysis. In: Primorac D, Schanfield M, eds. *Forensic DNA Applications: An Interdisciplinary Perspective*. CRC Press; 2014:647.
7. Liu L, Li Y, Li S, et al. Comparison of next-generation sequencing systems. *J Biomed Biotechnol*. 2012;2012:251364. doi:10.1155/2012/251364
8. Cornelis S, Gansemans Y, Vander Plaetsen AS, et al. Forensic tri-allelic SNP genotyping using nanopore sequencing. *Forensic Sci Int Genet*. 2019;38:204-210. doi:10.1016/j.fsigen.2018.11.012
9. Ping Chiu K. Evolution of DNA Sequencing Technologies. In: *Next-Generation Sequence and Sequence Data Analysis*. Bentham Science Publishers; 2018:162.
10. Sanger Sequencing: Historical Development of Automated DNA Sequencing. Published 2012. <https://agctsequencing.wordpress.com/2012/08/01/sanger-sequencing-historical-development-of-automated-dna-sequencing/>
11. Beck TF, Mullikin JC, Program NCS, Biesecker LG. Systematic Evaluation of Sanger Validation of Next-Generation Sequencing Variants. *Clin Chem*. 2016;62(4):647-654. doi:10.1373/clinchem.2015.249623
12. Zubakov D, Kayser M. Forensic Tissue Identification with Nucleic Acids. In:

- Primorac D, Schanfield M, eds. *Forensic DNA Applications: An Interdisciplinary Perspective*. CRC Press; 2014:647.
13. Sajantila A, Budowle B. Molecular Autopsy. In: Amorim A, Budowle B, eds. *Handbook Of Forensic Genetics: Biodiversity And Heredity In Civil And Criminal Investigation*. World Scientific; 2016:652.
 14. Axler-Diperte G, Bieber FB, Budimlija ZM, Sajantila A, Siegel D, Tang Y. Molecular Autopsy. In: Primorac D, Schanfield M, eds. *Forensic DNA Applications: An Interdisciplinary Perspective*. CRC Press; 2014:647.
 15. Wurmbach E. Prediction of Physical Characteristics, such as Eye, Hair, and Skin Color Based Solely on DNA. In: Primorac D, Schanfield M, eds. *Forensic DNA Applications: An Interdisciplinary Perspective*. CRC Press; 2014:647.
 16. Walsh S, Kayser M. Predicting Human Appearance from DNA for Forensic Investigations. In: Amorim A, Budowle B, eds. *Handbook Of Forensic Genetics: Biodiversity And Heredity In Civil And Criminal Investigation*. World Scientific; 2016:652.
 17. Sero D, Zaidi A, Li J, et al. Facial recognition from DNA using face-to-DNA classifiers. *Nat Commun*. 2019;10(1). doi:10.1038/s41467-019-10617-y
 18. Byrnes JK, Myres NM, Underhill PA. Genetic Genealogy in the Genomic Era. In: Primorac D, Schanfield M, eds. *Forensic DNA Applications: An Interdisciplinary Perspective*. CRC Press; 2014:647.
 19. Cornelis S, Willems S, Neste C Van, et al. Forensic STR profiling using Oxford Nanopore Technologies' MinION sequencer. *bioRxiv*. 2018;(October):433151. doi:10.1101/433151