

**IMPACTO DE LA TECNOLOGÍA DE SECUENCIA DE NUEVA GENERACIÓN (NGS)  
SOBRE EL ANÁLISIS GENÉTICO DE CASOS COMPLEJOS EN IDENTIFICACIÓN  
DE PERSONAS DESAPARECIDAS EN EL CONFLICTO ARMADO COLOMBIANO**

**CONSTANZA FORERO TORRES**

**UNIVERSIDAD DEL ROSARIO  
ESCUELA DE MEDICINA Y CIENCIAS DE LA SALUD  
MAESTRÍA EN GENÉTICA HUMANA  
BOGOTÁ D.C, JUNIO 2019**

**IMPACTO DE LA TECNOLOGÍA DE SECUENCIA DE NUEVA GENERACIÓN (NGS)  
SOBRE EL ANÁLISIS GENÉTICO DE CASOS COMPLEJOS EN IDENTIFICACIÓN  
DE PERSONAS DESAPARECIDAS EN EL CONFLICTO ARMADO COLOMBIANO**

**CONSTANZA FORERO TORRES**

**Trabajo de grado para optar por el título de  
Magister en Genética Humana**

**Director: MANUEL PAREDES LOPEZ, MD, PhD.  
Instituto Nacional de Medicina Legal y Ciencias Forenses**

**Coodirector: CARLOS RESTREPO, MD  
Escuela de Medicina y Ciencias de la Salud  
Universidad Del Rosario**

**UNIVERSIDAD DEL ROSARIO  
ESCUELA DE MEDICINA Y CIENCIAS DE LA SALUD  
MAESTRÍA EN GENÉTICA HUMANA  
BOGOTÁ D.C, JUNIO 2019**

**IMPACTO DE LA TECNOLOGÍA DE SECUENCIA DE NUEVA GENERACIÓN (NGS)  
SOBRE EL ANÁLISIS GENÉTICO DE CASOS COMPLEJOS EN IDENTIFICACIÓN  
DE PERSONAS DESAPARECIDAS EN EL CONFLICTO ARMADO COLOMBIANO**

**CONSTANZA FORERO TORRES**

**APROBADO**

---

**MANUEL PAREDES L.**  
**Director**

---

**CARLOS RESTREPO**  
**Coodirector**

---

**HERNANDO del CASTILLO**  
**Jurado**

---

**DORA FONSECA**  
**Jurado**

## **NOTA DE ADVERTENCIA**

*“La Universidad no se hace responsable por los conceptos emitidos por sus alumnos en sus tesis de grado”. Artículo 23 Resolución No. 13 de Julio de 1996*

## **DEDICATORIA**

*A mis hijos y esposo amados*

*A la memoria de mi madre,*

*Que permanece siempre en mi corazón*

## **AGRADECIMIENTOS**

Al Instituto Nacional de Medicina Legal y Ciencias Forenses, por el apoyo económico, el soporte técnico y logístico para desarrollar este proyecto. Especialmente, a la Subdirección de Ciencias Forenses por aprobarlo y al Grupo Nacional de Ciencias Forenses.

Al doctor Manuel Paredes, por dirigir, acompañar y asesorar este proyecto, así como por compartir sus conocimientos y por su dedicación y constante apoyo.

A la doctora, Andrea Pinzón por su apoyo y su valiosa colaboración y comentarios.

A mi compañero Cristian Méndez, por su gran apoyo en la organización de los datos.

A la doctora Patricia Heredia, coordinadora del Grupo Nacional de Ciencias Forenses, por su apoyo y cooperación.

A mis compañeras y amigas del Instituto Nacional de Medicina Legal y de la Unidad de Búsqueda de Personas dadas por Desaparecidas, especialmente a Natalia Figueroa, María Victoria Estupiñan, Claudia Vega y Edna Buitrago por sus palabras de apoyo, por escucharme, darme ánimo y soportar mis momentos de estrés.

A mi familia, por su constante apoyo, paciencia, amor y por soportar mis largas ausencias y los momentos importantes que deje de compartir con ellos.

A mi mami que siempre me enseñó que las metas no se logran sin constancia, esfuerzo y sacrificio y porque me dejó su ejemplo y enorme amor, que sin ellos no hubiera logrado culminar este proyecto.

A todos aquellos que, aunque no estén mencionados, contribuyeron de alguna forma en la ejecución de este proyecto.

## TABLA DE CONTENIDO

	pág.
1. RESUMEN.....	15
2. INTRODUCCIÓN .....	17
3. OBJETIVOS.....	21
3.1    Objetivo general .....	21
3.2    Objetivos específicos.....	21
4. MARCO TEÓRICO.....	22
4.1    La genética en el contexto de la investigación de grandes violaciones a los DDHH y al DIH en el conflicto armado colombiano .....	22
4.2    Muestras forenses antiguas, ADN degradado y familias incompletas .....	25
4.2.1    Herramientas actuales de análisis molecular y nuevas alternativas .....	25
4.2.1.1    STRs autosómicos. ....	25
4.2.1.2    STRs de cromosomas sexuales.....	26
4.2.1.3    ADN mitocondrial .....	27
4.2.1.4    SNPs (polimorfismos puntuales de secuencia).....	27
4.2.1.5    Tecnologías dirigidas a analizar ADN degradado.....	29
4.2.1.6    Tecnologías de análisis de SNPs actuales: Secuencia de Nueva Generación (NGS) o Secuenciación masiva en paralelo (MPS) .....	31
4.3    Estudios de poblaciones de referencia forense para marcadores SNPs.....	36
5. METODOLOGÍA .....	42
5.1    Selección de muestras de ADN humano .....	42
5.1.1    Muestras para los experimentos de validación.....	42
5.1.1.1    Muestras para los ensayos de Precisión y Reproducibilidad.....	42

5.1.1.2	Muestras para los ensayos de exactitud, reproducibilidad, sensibilidad, Low Copy Number (LCN) y variación del volumen de librerías para secuencia:.....	43
5.1.2	Muestras para el análisis comparativo entre tecnología de nueva generación y electroforesis capilar en casos complejos.....	43
5.1.2.1	Muestras para el ensayo de complejidad asociada a la calidad de la muestra.	43
5.1.2.2	Muestras para el ensayo de complejidad asociada a pedigrís poco informativos, bajo variables de número y tipo de familiares para cotejo .....	44
5.1.3	Muestras para los estudios poblacionales: .....	45
5.2	Transporte y almacenamiento de muestras.....	47
5.3	Extracción de ADN .....	47
5.4	Preparación de librerías y secuenciación.....	48
5.5	Criterios para calificación de resultados.....	49
5.6	Diseño experimental.....	49
5.6.1	Validación del kit “TG MiSeq FGx Reagent kit” para el equipo de NGS MiSeq® de Illumina <sup>49</sup>	
5.6.1.1	Experimento de exactitud y precisión .....	49
5.6.1.2	Experimento de reproducibilidad .....	51
5.6.1.3	Experimentos de sensibilidad y definición de umbrales .....	51
5.6.1.3.1	Sensibilidad.....	52
5.6.1.3.2	Umbrales analíticos.....	52
5.6.1.3.3	Balance de Heterocigotos: Diferencia máxima del número de reads entre alelos hermanos del Heterocigoto .....	53
5.6.1.4	Experimentos de LCN (Low Copy Number) .....	54
5.6.1.5	Experimentos de variación del volumen de librerías para secuencia .....	55
5.6.2	Análisis comparativo entre tecnología de nueva generación y electroforesis capilar en casos complejos .....	56

5.6.2.1	Complejidad asociada a la calidad de la muestra (muestras óseas):.....	56
5.6.2.2	Complejidad asociada a pedigrís poco informativos, bajo variables de número y tipo de familiares para cotejo:.....	56
5.6.3	Estudios poblacionales de 3 regiones colombianas para 89 SNPs de identificación, del kit forenseq illumina.....	57
6.	RESULTADOS Y DISCUSIÓN.....	59
6.1	Experimentos de validación.....	59
6.1.1	Experimento de exactitud y precisión.....	59
6.1.2	Experimento de reproducibilidad.....	66
6.1.3	Experimentos de sensibilidad y definición de umbrales.....	68
6.1.3.1	Sensibilidad.....	68
6.1.3.1.1	Efecto del tipo de muestra en montajes de 40 librerías: ADN purificado línea celular 2800M.....	69
6.1.3.1.2	Efecto del tipo de muestra en montajes de 40 librerías: ADN extraído de muestras calcificadas.....	73
6.1.3.1.3	Efecto del tipo de muestra en montajes de 88 librerías: ADN purificado línea celular 2800M y Restos óseos.....	73
6.1.3.2	Umbrales analíticos.....	77
6.1.3.3	Balance de Heterocigotos: Diferencia máxima del número de reads entre alelos hermanos del Heterocigotos.....	82
6.1.4	Experimentos de LCN (Low Copy Number).....	85
6.1.5	Experimentos de variación de librerías para secuencia.....	88
6.1.6	Efecto de la combinación de los experimentos de incremento de ciclajes y aumento del pool de librerías.....	91
6.2	Análisis comparativo entre tecnología de nueva generación y electroforesis capilar en casos complejos.....	94
6.2.1	Complejidad asociada a la calidad de la muestra.....	94

6.2.2	Complejidad asociada a pedigrís poco informativos, bajo variables de número y tipo de familiares para cotejo.....	95
6.3	Estudios poblacionales de 3 regiones colombianas para 89 SNPs de identificación, del kit forenseq illumina.....	98
7.	CONCLUSIONES .....	104
8.	CONSIDERACIONES BIOÉTICAS, DISPOSICIONES LEGALES VIGENTES Y PROPIEDAD INTELECTUAL.....	106
9.	BIBLIOGRAFIA.....	107
10.	ANEXOS.....	113

## LISTA DE FIGURAS

Fig 1. PCR en emulsión (Mardis.2008) .....	32
Fig 2. Pirosecuenciación (Secuenciador 454 ROCHE). .....	33
Fig 3. Tecnología semiconductora (Ion Torrent).....	34
Fig 4. Amplificación en puente Illumina .....	34
Fig 5. Secuenciación por síntesis Illumina .....	35
Fig 6. Loci del kit ForenSeq DNA Signature Prep.....	48
Fig 7. Experimento de Exactitud, calculada en profundidad de cobertura (DoC) para (A) STRs y (B) SNPs.....	61
Fig 8. Porcentaje de genotipos asignados, precisión y exactitud, para marcadores STRs y SNPs en ensayos con diferente número de muestras, amplificadas con el kit Signature Prep ForenSeq™.....	63
Fig 9. Promedio del N° de reads para ensayos de 32 y 88 librerías en muestras de sangre y células epiteliales bucales sobre tarjetas FTA, amplificadas con el kit Signature Prep ForenSeq™.....	65
Fig 10. Promedio del N° de reads para ensayo de 32 librerías en muestras de sangre y células epiteliales bucales en FTA y Restos óseos, amplificadas con el kit Signature Prep ForenSeq™.....	65
Fig 11. Pantalla de representación del N° total de Reads por muestra, en el ensayo de sensibilidad.....	70
Fig 12. Ensayo de Sensibilidad, calculada en porcentajes de genotipos asignados, exactitud y precisión para 40 librerías con la línea 2800M.....	73
Fig 13. Ensayo de Sensibilidad, calculada en porcentaje de exactitud para 40 y 88 librerías con la línea celular 2800M.....	75
Fig 14. Promedio de IT en reads para loci STRs heterocigotos en la dilución inmediatamente anterior a la cual se observó <i>drop out</i> .....	79
Fig 15. Promedio IT en reads para loci SNPs heterocigotos en la dilución inmediatamente anterior a la cual se observó <i>drop out</i> .....	81
Fig 16. Porcentaje del Balance de Heterocigotos para loci STRs, a partir del ensayo de sensibilidad.....	83

Fig 17. Porcentaje del Balance de Heterocigotos para loci SNPs, obtenidos a partir del ensayo de sensibilidad.....	84
Fig 18. Ensayo de Low Copy Number (LCN).....	88
Fig 19. Experimentos de variación del volumen de librerías para secuencia. ....	91
Fig 20. Efecto de la combinación de los experimentos de incremento de ciclajes y aumento del pool de librerías. ....	92
Fig 21. Ensayos con los mejores resultados variando número de ciclos (LCN), volumen de librerías para secuencia, o combinación de los dos. ....	94

## LISTA DE TABLAS

Tabla 1. Distribución de los 94 loci SNPs del kit Signature Prep ForenSeq en los 22 cromosomas.....	38
Tabla 2. Experimento de Exactitud para la línea celular 2800M a 1ng/uL .....	60
Tabla 3. Métricas de calidad obtenidas para ensayos con diferente número de librerías amplificadas con el kit Signature Prep ForenSeq™.....	62
Tabla 4. Experimento de Reproducibilidad calculada en porcentajes de precisión y número de genotipos asignados.....	67
Tabla 5. Experimento de Reproducibilidad, calculada en porcentajes de exactitud.....	68
Tabla 6. Experimento de Sensibilidad, calculada en porcentajes de genotipos asignados y de exactitud para marcadores A-STRs, Y-STRs, X-STRs y SNPs en montajes de 40 y 88 librerías.....	74
Tabla 7. Umbrales de Interpretación para STRs obtenidos por locus a partir del ensayo de sensibilidad.....	78
Tabla 8. Umbrales de Interpretación para SNPs obtenidos por locus a partir del ensayo de sensibilidad.....	80
Tabla 9. Métricas de calidad obtenidas para ensayos de Low Copy Number (LCN).....	85
Tabla 10. Ensayo de Low Copy Number (LCN), calculado en porcentajes de exactitud y de genotipos recuperados entre experimentos en condiciones del fabricante (ciclaje normal) vs Tratamiento 2 y Tratamiento 3, obtenidos para marcadores A-STRs, Y-STRs, X-STRs y SNPs .....	86
Tabla 11. Experimentos de variación de librerías para secuencia, calculada en porcentajes de genotipos asignados y de exactitud para marcadores A-STRs, Y-STRs, X-STRs y SNPs en condiciones del fabricante (pool final de 7.5ul) vs Ensayo 1 y Ensayo 2. ....	89
Tabla 12. Cuantificación obtenida para muestras anónimas de ADN, extraídas de restos óseos (RO) de casos de identificación, con resultados negativos por CE, según las variables seleccionadas: Muestras óseas antiguas (30-40 años) no incineradas, parcialmente incineradas y totalmente incineradas.....	95

Tabla 13. Resultado del cálculo del valor de LR, realizado con el software Familias, en casos de parentescos complejos, obtenido para 25 marcadores A-STRs y 92 marcadores SNPs incluidos en el kit Signature Prep ForenSeq™.....	95
Tabla 14. Resultado de la probabilidad de parentesco, en casos complejos, combinando los valores de LR obtenidos para 25 marcadores A-STRs y 92 marcadores SNPs incluidos en el kit Signature Prep ForenSeq™. ....	97
Tabla 15. Frecuencias alélicas de 89 SNPs incluidos en el kit Signature Prep ForenSeq™ para las regiones Andina Central, Orinoquía y Pacífico de Colombia. NA datos eliminados. ....	100

## 1. RESUMEN

En Colombia, como consecuencia directa de un conflicto armado de más de 60 años, se estima que hay más de 100.000 desaparecidos, de los cuales 83.000 pueden asociarse a desapariciones forzadas. Dentro del proceso de reparación a las víctimas y en consecuencia con los acuerdos de paz con las FARC-EP alcanzados en La Habana en 2016, sus cuerpos, una vez recuperados deben identificarse y entregarse dignamente a sus familiares.

Para este fin el Instituto Nacional de Medicina Legal y Ciencias Forenses (INMLCF), como miembro del Sistema Nacional de Atención y Reparación a las víctimas <sup>2</sup>, ha construido un modelo de trabajo forense interdisciplinario para la identificación de cuerpos en condición de no identificados (CNI) en el cual, los análisis genéticos hacen parte fundamental de este modelo.

Por lo anterior, y teniendo en cuenta la antigüedad del conflicto, los laboratorios de genética forense deben desarrollar estrategias para enfrentar dos problemas evidentes durante el análisis de estos casos, el primero obtener perfiles genéticos satisfactorios a partir de muestras óseas altamente degradadas y de concentraciones mínimas de ADN y el segundo, realizar cotejos genéticos con familiares de referencia poco informativos, toda vez que, en un lapso tan prolongado, los padres y relativos cercanos han fallecido en muchos de los casos.

Ante este escenario, hemos puesto a consideración como una opción para los casos de identificación en Colombia, la tecnología de secuencia masiva del genoma NGS/MPS, mediante El kit ForensSeq de la firma Illumina, y el equipo MiSeq FGx, , dado que este tipo de plataformas permiten el análisis de un gran número de loci genómicos (más de 200 loci), compensando el problema de familiares deficitarios y adicionalmente incluye polimorfismos de nucleótido simple o SNP (single nucleotide polymorphisms) lo cual, aumenta la probabilidad de éxito en las muestras degradadas y/o escasas, por su pequeño tamaño de amplicón (60-200 pb).

En la presente investigación hemos realizado la validación interna del ensayo NGS, ajustada a los laboratorios del INMLCF, bajo parámetros internacionales de calidad (SWGDM, 2019) para definir bajo qué condiciones y hasta que límites, pueden obtenerse perfiles genéticos

reproducibles e informativos, a partir de muestras óseas complejas. Adicionalmente, se presentan los análisis genético-poblacionales, para 89 marcadores tipo SNPs, (kit Forenseq de Illumina), en 3 regiones etno-geográficas colombianas (Andina-Central, Pacífica y Orinoquía) de gran impacto en el conflicto armado, por la gran cantidad de víctimas que se generaron allí.

## 2. INTRODUCCIÓN

Según el Registro Único de Víctimas (RUV) del Sistema Nacional de Información (SIN), creado por la Ley 1408 de 2011, también llamada Ley de Víctimas, en Colombia, como consecuencia directa del conflicto armado, a la fecha de elaboración del presente proyecto existen 8.581.339 víctimas acumuladas en más de 60 años de confrontación. Entre ellas, se incluyen cerca de 80 mil personas dadas por desaparecidas y la mayoría quizás, por desaparición forzada.

En el marco de la justicia transicional, la Ley de Víctimas y Restitución de tierras, constituye para el país un marco legal que adopta medidas efectivas a favor de las personas que han sufrido a consecuencia de infracciones al Derecho Internacional Humanitario o de violaciones graves y manifiestas a las normas internacionales de Derechos Humanos, ocurridas dentro del conflicto armado colombiano. En este contexto el Instituto Nacional de Medicina Legal y Ciencias Forenses (INMLCF), como miembro del Sistema Nacional de Atención y Reparación a las víctimas, tiene el deber legal de contribuir a los derechos de los familiares en la búsqueda, exhumación e identificación de víctimas en casos de desaparición forzada. Adicionalmente, en virtud de los acuerdos de paz con las FARC-EP alcanzados en La Habana y en particular en el numeral 5, se define la creación de la Unidad de Búsqueda de Personas dadas por Desaparecidas (UBPD), quien deberá dirigir, coordinar y contribuir a la recuperación, identificación y entrega digna de cuerpos esqueletizados, en conjunto con el INMLCF, quien se constituye como el soporte científico forense para las labores de identificación (Presidencia de la República & FARC-EP, 2016).

En cumplimiento de esta demanda social, Medicina Legal está respondiendo con un modelo de trabajo forense interdisciplinario para la identificación de cuerpos en condición de no identificados (CNI), donde la genética forense juega un papel trascendental, toda vez que, en la mayoría de los casos, no se cuenta con cotejos lofoscópicos ya que se trata de cuerpos esqueletizados, ni tampoco con registros odontológicos ante mortem. La conclusión de identidad por lo tanto, debe soportarse en gran medida en los hallazgos genéticos, para lo cual existen dos situaciones complejas asociadas a varias décadas de duración del conflicto, que deben superarse:

Primero, no es fácil obtener ADN de buena calidad y concentración, a partir de tejidos que pueden estar altamente comprometidos debido a su prolongada permanencia bajo condiciones adversas como las que ofrecen los ambientes de inhumación de los suelos tropicales colombianos; y segundo, que muchas veces se cuenta solo con familiares de grados de consanguinidad más distantes y menos informativos que los padres o hijos del desaparecido.

No obstante, la tecnología genética actual permite hacer accesibles al análisis, evidencias biológicas complejas y extrae de ellas información muy útil para demostrar el parentesco con sus relativos y la identidad de un cadáver. Esto es posible gracias al uso de plataformas tecnológicas altamente sensibles basadas en la Electroforesis capilar (CE) y al análisis de marcadores de ADN muy informativos, como los Microsatélites o STRs (Short tandem repeats). Adicionalmente, se estudia el ADN mitocondrial como herramienta complementaria o de rescate en muestras degradadas. Estas estrategias analíticas permiten dar respuesta a más del 85% de las solicitudes.

Lo anterior supone que, en un porcentaje importante de casos, cerca del 8 al 10%, los resultados son negativos. Según estadísticas del INML y CF, los laboratorios de Genética entre los años 2012-2016 han radicado más de 7500 casos de identificación provenientes del conflicto armado, para un promedio de 1920 casos por año aproximadamente (Datos obtenidos de los informes de gestión del Grupo Nacional de Genética del INML y CF). De estos datos, se puede inferir la dimensión del problema de los casos negativos.

A su vez los casos negativos o no concluyentes, requieren un mayor número de reprocesos, buscando obtener un perfil genético de utilidad en el análisis, o un repertorio muy amplio de marcadores moleculares, principalmente para resolver pedigrís deficitarios. En cualquier caso, se prolongan los tiempos de respuesta, se incrementan los costos en los laboratorios y el represamiento de solicitudes.

Ante este escenario, la comunidad forense está evaluando la tecnología de secuencia masiva del genoma NGS como alternativa a las técnicas tradicionales de CE, puesto que permite el análisis de un gran número de loci genómicos, incluyendo polimorfismos de nucleótido simple o SNP (single nucleotide polymorphisms) que por su pequeño tamaño de amplicón (<150pb) aumenta la probabilidad de análisis en las muestras degradadas y/o escasas. El kit ForensSeq de la firma

Illumina, y el equipo MiSeq que hemos elegido para esta validación, permiten el análisis simultáneo en una sola reacción de más de 240 loci de ADN nuclear, 94 de ellos tipo SNPs, sobre plataformas susceptibles de automatización para el análisis masivo de muestras. Esto último es también importante si se tiene en cuenta que los universos de víctimas de desaparición, como se mencionó antes, pueden superar los 80 mil casos y que los familiares reclamantes que deben ingresar al Banco de Perfiles Genéticos de Desparecidos (BPGD) en Colombia se estiman en más de 250 mil personas, considerando al menos 3 relativos por desaparecido. Por tal razón, propusimos evaluar el impacto que la tecnología NGS/MPS pueda tener, como una opción para los casos de identificación en Colombia, ante la expectativa de constituirse en una estrategia de menor costo que la actual, que cumpla también con los criterios de calidad de los laboratorios, que permita el análisis masivo de muestras, que mejore significativamente los resultados que se obtienen en estos escenarios complejos y que reduzca los tiempos de respuesta de los laboratorios.

Por lo anterior, y no obstante las bondades que han demostrado estas plataformas en otros escenarios (Phillips 2009), es necesario realizar una validación interna del ensayo NGS, que permita cumplir con las exigencias mundiales de los laboratorios acreditados por normas internacionales (ISO IEC 17025, SWGDAM, etc.), como los laboratorios forenses colombianos, que incluya además un diseño de validación que contemple los escenarios de nuestra casuística colombiana, en experimentos controlados, que permitan definir las mejores condiciones de uso de esta nueva técnica en los laboratorios del instituto y los alcances reales de la misma en muestras biológicas complejas.

Adicionalmente, los estudios de validación deben incluir los análisis de poblaciones de referencia forense para los marcadores tipo SNPs ya que, aunque están publicados estudios de SNPs asociados a evaluar la ancestralidad en poblaciones colombianas (Rishishwar et al., 2015), no existen en Colombia, estudios genético-poblacionales que evalúen las frecuencias alélicas de marcadores tipo SNPs de identificación, de todas las regiones colombianas y menos aún de todos los SNPs ID del kit Forenseq de Illumina. Al respecto, hemos considerado, muestrear al menos 3 regiones etno-geográficas que, a nuestro juicio, representan los escenarios más comunes del conflicto colombiano.

Adicionalmente, el estudio se justifica porque muchos de los casos complejos, con altos niveles de degradación de las muestras óseas, son de muy alto impacto social y se espera que las nuevas tecnologías permitan avanzar en su análisis.

En resumen, el alto número de casos negativos producto a su vez del alto nivel de degradación y la escasa cantidad del ADN recuperado de los restos óseos por una parte, y de la ausencia en muchos casos, de familiares de desaparecidos en primer grado de consanguinidad, sumado a la ausencia de estudios en Colombia, de los SNPs de identificación (kit ForenSeq de Illumina) de todas las poblaciones forenses de referencia, justifican el presente estudio de validación y constituyen problemas de investigación que pueden en nuestro concepto ser abordados con una alternativa de análisis como la secuencia de nueva generación.

Esperamos que la aplicación de esta nueva tecnología en el campo forense sea una herramienta en la resolución de los problemas mencionados, que contribuya a disminuir costos y tiempo en el reprocesamiento de muestras y casos complejos, como los que se presentan a diario en nuestro país en relación con el conflicto armado.

### **3. OBJETIVOS**

#### **3.1 Objetivo general**

Determinar el impacto de la tecnología de NGS/MPS en el análisis de casos forenses complejos de identificación de personas desaparecidas en el conflicto armado colombiano.

#### **3.2 Objetivos específicos**

**3.2.1.** Establecer las condiciones experimentales bajo las cuales, se asegura la calidad de los ensayos, cuando se aplica la tecnología NGS (el kit Foren Seq y el equipo MiSeq NGS de la firma Illumina), al análisis de casos forenses de identificación, utilizando experimentos convencionales de validación de métodos analíticos.

**3.2.2.** Determinar si NGS permite rescatar casos de Identificación, analizados por CE en los cuales se reportaron resultados negativos, debido al alto nivel de degradación de las muestras óseas y/o la escasa representación de familiares de referencia.

**3.2.3.** Determinar la distribución de frecuencias alélicas de 89 SNPs de identificación del kit Foren Seq-Illumina, en tres poblaciones colombianas representativas de la casuística colombiana de identificación (Andina central, Orinoquía y Región Pacífica) y estimar los parámetros genéticos poblacionales de interés para su aplicación forense.

## **4. MARCO TEÓRICO**

### **4.1 La genética en el contexto de la investigación de grandes violaciones a los DDHH y al DIH en el conflicto armado colombiano**

La Genética forense se ha convertido en un soporte revolucionario para la administración de justicia en varios escenarios, uno de ellos, ha tenido en las últimas décadas un impacto social definitivo: la genética como apoyo a la investigación de grandes violaciones de los DDHH y especialmente, del DIH en los conflictos armados, eventos que dejan cifras muy altas de víctimas fatales y de personas desaparecidas. Al respecto, el Comité de DDHH de Naciones Unidas instó a los países en general, a desarrollar la genética forense e implementar laboratorios para identificación de personas desaparecidas, (Consejo de Derechos Humanos de Naciones Unidas, Resolución 10/26 de 2009 y 15/5 de 2010) para apoyar a la restauración de los derechos de las víctimas. Argentina, Chile, Guatemala, México, El Salvador, entre otros países de la región, han compartido experiencias similares a la colombiana, y han tenido desarrollos muy importantes en genética forense. No obstante, el caso del conflicto armado colombiano es quizás, por su larga duración y su gran complejidad uno de los escenarios donde más se pone a prueba el alcance de la ciencia forense y donde se hace necesario buscar alternativas científicas para hacer accesible al conocimiento, muchas evidencias forenses que puedan aportar a la verdad de lo sucedido, como espera la sociedad colombiana tras más de 60 años de confrontación.

Varios historiadores (Arias, 2014, García-Peña, 2017, Rojas, 2016) coinciden en que el conflicto armado colombiano inicia a mediados de los años cuarenta y particularmente en 1946, desde la presidencia de Mariano Ospina, cuando se generó una ola de violencia contra el movimiento popular liderado por Jorge Eliecer Gaitán y se incrementó en 1947, tras convertirse en jefe único del Partido Liberal, lo que seguramente garantizaría el ascenso al poder del primer gobierno popular en la historia de Colombia. Si bien, estos años dejaron miles de muertos, más de 3500 personas murieron en los combates que se extendieron solo en una semana (García, 2014) del mes abril de 1948, tras el asesinato del caudillo liberal, y se dio inicio al periodo bélico de la historia nacional conocido como "La Violencia" (Castaño, 2014).

Quizás, de no haber ocurrido el magnicidio y el pueblo hubiese logrado el poder por la vía democrática, Colombia no habría tenido un conflicto armado, no se habrían creado las resistencias campesinas de donde se generaron, varios años después, la guerrilla de las FARC-EP y luego el ELN y el EPL.

Desde el año 48 se institucionalizaron los crímenes de estado en Colombia, las policías políticas inicialmente conservadoras, fueron responsables de miles de asesinatos, luego, la institucionalidad tendría el apoyo incondicional de las fuerzas armadas regulares. La Violencia, tuvo un respiro en el famoso pacto de paz que significó el llamado “Frente Nacional” firmado entre las élites liberal y conservadora. Este pacto de silencio generó el modelo de impunidad, perdón y olvido más perverso de la historia y repartió el poder en las dos élites, dejando fuera del ejercicio democrático a cualquier pensamiento político diferente.

Luego a mediados de los años 70’s, el M-19 (Movimiento 19 de abril) apareció en escena, como respuesta al fraude electoral del 19 de abril de 1970. Su incursión en el Palacio de Justicia en Bogotá, en noviembre de 1985, junto con el asesinato de Gaitán, han sido quizás, los hechos de mayor significancia en la historia del conflicto colombiano, la identificación de las víctimas del holocausto del Palacio, es aún una tarea que aún adelantan los forenses colombianos.

También desde los años 70, la guerra en Colombia tomó dimensiones insospechadas con el ingreso de los grandes capitales del narcotráfico al conflicto y finalmente, con la incursión de paramilitares antiguerrillas, como las Autodefensas Unidas de Colombia (AUC), en la década de los 90 (García, 2014). Fue en esta década cuando se generó la mayor parte de las víctimas conocidas en el conflicto armado colombiano. Las AUC se desmovilizaron en el 2006 en el contexto de la Ley de Justicia y Paz (975 de 2005), y aunque confesaron cientos de homicidios como parte del proceso de justicia transicional que se les ofreció en aras de que las víctimas conocieran la verdad de los hechos, solo una mínima proporción del total de sus víctimas se recuperaron.

Entre tanto, Colombia ha avanzado lentamente en el camino de reivindicar los DDHH y el DIH; solo en el año 2000, la desaparición forzada fue reconocida en los códigos colombianos como un delito (ley 589 de 2000). Luego, también gracias a las asociaciones de víctimas y familiares de desaparecidos, se promulgó el año 2010, la Ley 1408, que creó el Banco de perfiles genéticos de

desaparecidos (BPGD), la cual fue reglamentada 5 años después (Decreto 03030 de 2015). Posteriormente, se promulgó la ley 1448 o ley de víctimas, que tiene por objeto [...] establecer un conjunto de medidas judiciales, administrativas, sociales, económicas, individuales y colectivas en beneficio de las víctimas, dentro de un marco de justicia transicional, que posibiliten hacer efectivo el goce de sus derechos a la verdad, la justicia y la reparación con garantía de no repetición, de modo que se reconozca la condición de víctimas y se dignifique a través de la materialización de sus derechos constitucionales (Ley 1448, 2011: artículo 1o). (Barón, 2015).

Finalmente, después de superar un plebiscito en contra, y en medio de la mayor polarización política de los colombianos en su historia, en el 2016 el gobierno de Juan Manuel Santos, logra en la Habana -Cuba, el histórico Acuerdo de Paz para Colombia, con las FARC-EP, iniciándose una nueva etapa denominada postconflicto.

Como se indicó anteriormente, esta historia ha dejado más de 8 millones de víctimas y una deuda inmensa del estado colombiano para con ellas; Los acuerdos establecidos para lograr una paz estable y duradera obligan a aportar al conocimiento de la verdad, a la correcta administración de la justicia sobre los responsables de crímenes atroces, a la reparación del daño social causado y a garantizar la no repetición de los hechos.

La investigación forense de los crímenes cometidos en el conflicto armado y en particular la identificación de restos humanos de personas desaparecidas, apoyada por la genética, aporta a la búsqueda de la verdad de lo sucedido. En ese escenario, la guerra en Colombia ha dejado en todas sus épocas, entre 85.000 y 100.000 desaparecidos, la mayoría generados en las últimas tres décadas, lo cual supera las cifras de todos los conflictos modernos juntos. Toda esta historia que encierra el conflicto y post-conflicto de nuestro país, nos permite entender el panorama al cual nos enfrentamos las instituciones en el proceso de búsqueda e identificación de víctimas del conflicto armado y explica también la complejidad de los casos que debemos abordar; la antigüedad del conflicto por ejemplo, define las condiciones de las muestras óseas que deben analizarse en los laboratorios de genética: muestras degradadas por más de 30 a 40 años

inhumadas en ambientes y suelos tropicales en la mayoría de los casos y por otra parte, un escaso número de familiares directos disponibles para los cotejos genéticos en el BPGD.

## **4.2 Muestras forenses antiguas, ADN degradado y familias incompletas**

### **4.2.1 Herramientas actuales de análisis molecular y nuevas alternativas**

#### **4.2.1.1 STRs autosómicos.**

La tecnología actual aplicada al análisis del ADN en casos forenses, y en particular los kits comerciales de PCR multiplex para el análisis de polimorfismos del tipo short tandem repeats (STRs) que se usan en el mundo entero, requieren de concentraciones entre 0.2 a 1 ng de ADN dúplex, con una integridad de al menos 100 - 500 pares de bases de longitud para poder ser amplificado. Como se dijo atrás, las muestras óseas obtenidas de casos del conflicto armado en Colombia, además de ser muy antiguas, han estado expuestas a ambientes que degradan el ADN, fragmentándolo de tal forma que puede ser demasiado pequeño para ser amplificado eficientemente por PCR o bien afectan la reacción de amplificación por la presencia de inhibidores de las polimerasas de ADN.

El análisis forense de los STRs se inició al finalizar la década de los 80 (Weber & May, 1989), y se han mantenido como la plataforma de elección en todos los escenarios de la investigación judicial donde es posible el estudio de material biológico. Son muchas las ventajas de este tipo de marcador:

- Generan amplicones de pequeño tamaño entre 100 y 500 pb
- Tienen alto poder de discriminación
- Permiten ser amplificados de forma simultánea con PCR multiplex
- Tienen altas tasas de mutación ( $\sim 10^{-3}$  a  $10^{-4}$ ) y por lo tanto alta diversidad
- Herencia codominante
- Alta reproducibilidad entre laboratorios
- Se trabajan loci comunes en el mundo entero posibilitando la comparación de resultados
- Los perfiles STRs son fácilmente sistematizables en bases de datos

Para los laboratorios del IML los STRs resuelven más del 85% de los casos de identificación y criminalística y todos los casos de paternidad. Se analizan por Electroforesis capilar en plataformas policromáticas automatizadas, con software especializados en el análisis de fragmentos. Actualmente se coamplifican hasta 24 STR simultáneos según el kit.

En los últimos años, estos kits han sido mejorados incluyendo dos estrategias que aumentaron notablemente el número de resultados exitosos en muestras moderadamente degradadas: la inclusión de buffers que disminuían el efecto de sustancias inhibidoras de la PCR y el diseño de primers que permitían generar amplicones aún más cortos, generando fragmentos amplificados de cercas de 100 a 200 pb, llamados **mini-STRs**.

Aun así, los restos óseos obtenidos de exhumaciones en casos de desapariciones forzadas o incluso en desastres masivos, actos terroristas, etc, pueden contener ADN aún más degradado, de modo que el estudio con STRs mejorados puede ser aún insuficiente para obtener información de utilidad en la identificación de un cuerpo.

#### ***4.2.1.2 STRs de cromosomas sexuales***

Por otra parte, ya sea como herramienta complementaria o a manera de tamiz en el análisis de grandes números de casos, los STR de cromosomas sexuales X o Y (Mo et al., 2016), han resultado de gran utilidad para el análisis forense, dadas sus características: Los Y-STRs por ejemplo:

- Muestran moderado nivel de polimorfismo
- Tienen un tipo de herencia exclusivamente paterna que se mantiene de generación en generación sin cambios, excepto los producidos por eventos mutacionales
- Permiten obtener haplotipos altamente discriminativos de grupos familiares y útiles para trazar la evolución de los linajes paternos.

Sin embargo, por si solos, estos marcadores no excluyen como padre a ningún otro varón de esa línea paterna o no podrán excluirse los individuos que presenten el mismo perfil haplotípico

como autores de un delito, por lo que el estudio debe complementarse con marcadores autosómicos en el caso de que sea necesario eliminar estas posibilidades (Forense, Gen, Gen, & Landsteiner, 1960).

#### **4.2.1.3 ADN mitocondrial**

Finalmente, el estudio del ADN mitocondrial (ADNmt) ha sido también una herramienta tradicional en la batería de marcadores con que cuentan los laboratorios de genética forense en el mundo y de hecho en los laboratorios forenses colombianos. La importancia del ADNmt en los estudios forenses radica básicamente en:

- Tipo de herencia materna, que permite que todos los individuos de un mismo linaje materno presenten una misma secuencia de ADNmt
- Elevada tasa de mutación
- Existencia de miles de moléculas por célula, lo cual permite el estudio de muestras en las que el material biológico a analizar se encuentra degradado o en concentraciones insuficientes para ensayar cualquier otro marcador nuclear.

Finalmente, los marcadores de amplicón corto, cuya variación está concentrada en un espacio muy pequeño de un fragmento de ADN amplificable, (InDels y SNPs), constituyen ahora una de las mejores alternativas para el abordaje de muestras altamente degradadas.

#### **4.2.1.4 SNPs (*polimorfismos puntuales de secuencia*)**

Estos marcadores, están siendo muy utilizados en la resolución de casos complejos de parentesco o identificación, donde se tienen muestras muy degradadas o en casos donde se necesita aumentar la cantidad de información genética obtenida de muestras forenses impugnadas.

Los SNP son sustituciones de bases, inserciones o eliminaciones que se producen en una posición determinada del genoma de cualquier organismo. La mayoría de éstos son bialélicos,

por lo que son menos informativos que los STR y necesitan por lo tanto ser estudiados en un mayor número (50-100) para obtener el mismo poder de discriminación que existe en STR.

Debido a su baja tasa de mutación (aproximadamente  $10^{-8}$ ) que los convierte en marcadores genéticos estables y, a que aproximadamente el 85% de la variación humana se deriva de los SNP, se vuelven idóneos para análisis forenses: SNPs de identificación humana (HID SNPs) (análisis basados en linaje, casos de parentesco, de personas desaparecidas y de situaciones, donde no puede haber una muestra de referencia directa disponible).

Adicionalmente, otra de las grandes utilidades de los SNPs radica en la determinación del origen etnogeográfico de una muestra biológica (SNPs de ancestría) o de características fenotípicas como el color del pelo o el color de los ojos (SNPs de fenotipo) (Budowle & Van Daal, 2008) (A. D. Ambers et al., 2016); (Mo et al., 2016) refiere la incorporación de dos sistemas principales de SNP de identificación humana (HID SNPs): el ensayo múltiple SNPforID que incluye 52 loci SNP (Sanchez et al., 2006), y el panel de SNP para identificación individual (IISNP) que incluye 92 loci SNP desarrollados por Kidd et al (Pakstis et al., 2010); (Dixon et al., 2005), en investigaciones corrientes y aplicaciones de pruebas de paternidad y parentesco.

Para casos forenses complejos, se ha visto la necesidad de ampliar el número de marcadores o usar diferentes tipos de ellos (STR autosomales, de cromosoma X, cromosoma Y y SNPs) si se quiere obtener valores de probabilidad adecuados para resolver los casos o por el estado de degradación de las muestras.

Actualmente, para una misma muestra biológica, se suelen analizar hasta 23 STRs (estandarizados y validados) usando PCR multiplex y secuenciadores automáticos mediante electroforesis Capilar (CE), esta tecnología requiere del uso de varias reacciones de PCR y diferentes flujos de trabajo para obtener los resultados requeridos. Por esta razón, recientemente, se han desarrollado grandes plataformas como secuenciación de nueva generación (NGS), que comparada con la electroforesis capilar convencional, permite la tipificación simultánea y proporciona genotipado sensible y preciso aplicable al análisis forense (Mardis, 2008).

#### **4.2.1.5 Tecnologías dirigidas a analizar ADN degradado**

Se conoce muy poco sobre los mecanismos de degradación del ADN e inhibición de los procesos metodológicos comúnmente asociados a las muestras forenses (Mccord, Opel, Funes, & Jantz, 2011), si bien existen tratamientos que mejoran la condición de las muestras degradadas sin afectar el perfil genético del ADN, actualmente no existen métodos para evitar este problema (Battista, 2012), más aún los kits de tipificación están limitados a la condición y concentración de la muestra.

Se han propuesto protocolos para restaurar el ADN degradado, que permiten la tipificación por STRs, una tarea que requiere la capacidad de aumentar la cantidad de fragmentos de ADN no dañados disponibles en el rango necesario de 100-500 pb antes de la amplificación a través del proceso de PCR.

Una de las alternativas para mejorar el éxito de la tipificación de STRs a partir de ADN degradado y / o de baja copia, es la amplificación del genoma completo (WGA), la cual fue descrita por primera vez a principios de la década de 1990, y tiene esencialmente dos categorías: amplificación de desplazamiento múltiple (MDA) y métodos que implican variaciones de la PCR.

La primera requiere de ADN de alto peso molecular (generalmente > 2 kb) para tener éxito, lo que la hace inviable en la mayoría de muestras forenses; por el contrario la segunda (Ej DOP-PCR modificado) se afecta menos por la cantidad y calidad del ADN, pudiendo ser usado para muestras forenses complejas (A. Ambers et al., 2016).

Dentro de esta última categoría se han reportado estudios con diferentes metodologías para aumentar la sensibilidad de detección, logrando la tipificación de muestras de bajo número de copias (Low copy number o LCN), que se define como el análisis de cualquier muestra de ADN donde los resultados están por debajo del umbral estocástico para una interpretación confiable. Estas metodologías comprenden modificaciones durante la reacción en cadena de la polimerasa (PCR) y / o manipulaciones posteriores a la PCR.

Las dos técnicas pueden generar resultados indeseables como desbalance de alelos en loci heterocigotos, *dropout* alélico, incremento en el número de *stutters* o aumento en el riesgo de contaminación con ADN foráneo. La fiabilidad en el uso de estas metodologías, obliga a la reproducibilidad de los análisis, validaciones confiables y al exhaustivo análisis en la interpretación e incertidumbre de los resultados (Budowle, Eisenberg, & van Daal, 2009).

Budowle describe varias maneras de llevar a cabo la tipificación LCN para aumentar la sensibilidad del ensayo, estos incluyen:

- Aumento del número de ciclos de PCR
- PCR anidada
- Reducción del volumen de la PCR
- Amplificación del genoma completo antes de la PCR
- Señal de colorante fluorescente mejorada
- Uso de formamida de mayor pureza en la preparación de muestras para electroforesis capilar (CE)
- Limpieza post-PCR para eliminar iones que compiten con ADN durante la inyección electrocinética
- Aumento de los tiempos de inyección en la CE

Aunque, las técnicas de LCN, pueden ser ventajosas en ciertos estudios, Budowle hace un llamado a la precaución y sugiere que la tipificación de LCN debe usarse solo para identificación de personas desaparecidas (incluidas las víctimas de desastres masivos) sugiere estudiar SNPs como un marcador de primera elección para el análisis de ADN degradado y de baja copia, dado que sus amplicones pueden ser más cortos (hasta 50 pb) incluso que los de los mini-STRs convencionales y por lo tanto, la amplificación podría ser más robusta y los efectos estocásticos pueden ser menores que para los amplicones STRs grandes.

Se han realizado intentos con muestras degradadas en análisis de genotipificación de SNPs, inicialmente demostrando que los SNP basados en la bioquímica de la extensión de primer como GenomeLab™ SNPStream1 (Beckman Coulter) y sistemas multiplex SNaPshot™ (Applied Biosystems™) son capaces de genotipificar miles de SNP en un solo análisis, pero requieren un mayor volumen de ADN inicial o de producto de PCR y tienen la desventaja de ser

procedimientos de múltiples etapas. Adicionalmente el kit Foren-SNPs™, permitió la amplificación de 21 loci en una sola reacción y necesita amplicones mucho más pequeños, no obstante, por la naturaleza bialélica de los SNP se hace difícil la interpretación de mezclas (Dixon et al., 2006);

Posteriormente, se ha intentado optimizar aún más los sistemas multiplex SNP, en el 2008 (Babol-Pokora & Berent, 2008) demostraron con el método de SNP minisequencing que son mucho más efectivos para análisis de LCN-DNA en comparación con STR multiplex comerciales. Luego a partir del 2010, se han publicado artículos, tanto en el ámbito forense como clínico caracterizando SNP adecuados para la identificación humana (Pakstis et al., 2010).

A comienzos del 2014 se reportaron ensayos con kits comerciales de SNP multiplex para aplicaciones de identidad humana (HID), con el panel de identidad HID-Ion AmpliSeq en el Ion Torrent (Ion PGM) (Thermo Fisher Scientific, Waltham, MA, EE. UU.), mostrando perfiles de SNP casi completos partiendo desde 0.1 ng de ADN inicial (Seo et al., 2013); (Gettings, Kiesler, & Vallone, 2015) y en el 2015 una evaluación con el kit ForenSeq DNA Signature Prep usado en el secuenciador MiSeq, de la firma Illumina, mostró perfiles completos con cantidades de DNA de 1 ng y en un (94,4%) se observaron datos de secuencia con cantidades de ADN inicial de hasta 100 pg indicando su potencial para alcanzar niveles de sensibilidad comparables con los actuales tecnologías de tipado de ADN (Jennifer D. Churchill,\*, Sarah E. Schmedes, Jonathan L. King, 2016; Xavier & Parson, 2017)

#### ***4.2.1.6 Tecnologías de análisis de SNPs actuales: Secuencia de Nueva Generación (NGS) o Secuenciación masiva en paralelo (MPS)***

Recientemente, la Secuencia de Nueva Generación está siendo utilizada en el campo de la genética forense y específicamente para casos de identificación y análisis de rasgos fenotípicos. Si se comparan las tecnologías tradicionales con las ventajas que ofrece NGS, se encuentra un amplio número de marcadores tanto de STRs como de SNPs, que ofrecen ventajas al ser obtenidos en una sola reacción, con un alto rendimiento y a un precio competitivo para los laboratorios (Børsting & Morling, 2015).

El método general consiste en 4 pasos: Fragmentación del ADN (o ARN), ligación de oligonucleótidos (adaptadores) en cada uno de los extremos, amplificación y secuenciación. Estos dos últimos, varían según las diferentes empresas que han desarrollado equipos con la plataforma NGS (Mardis, 2008)

La primera de ellas fue publicada en el 2005 y desarrollada por la empresa 454 (luego adquirida por Roche Co.), ésta se basa en amplificación clonal, mediante PCR en emulsión y en secuenciación por síntesis usando un protocolo de pirosecuenciación (Baracchinis & Bremer, 1991).

En síntesis, la PCR en emulsión permite realizar en un único tubo miles reacciones de PCR independientes. Cada fragmento obtenido con su adaptador se une a esferas que contienen secuencias complementarias en su superficie, luego se realiza una emulsión (agua-aceite) de tal forma que en cada gota de aceite se encuentra una única esfera, junto con los reactivos para la PCR y finalmente se lleva a cabo una PCR convencional y simultánea, para obtener una amplificación clonal de cada fragmento unido a una esfera.

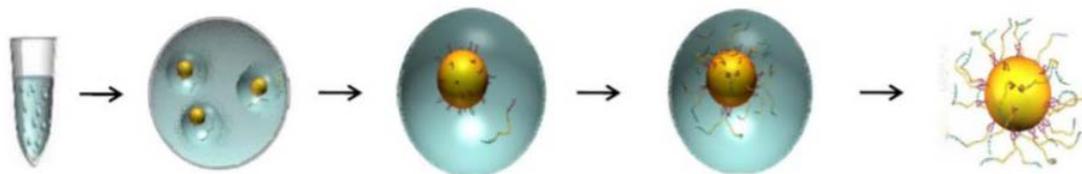


Fig 1. PCR en emulsión (Mardis.2008)

La pirosecuenciación, se basa en la adición de un (dNTP) a perlas que contienen fragmentos de DNA amplificados.

Brevemente, el DNA se fragmenta en tamaños aleatorios, se agregan adaptadores universales a los extremos y se adhieren a cada perla que contiene miles de adaptadores, quedando atrapados en una emulsión (solo un fragmento de DNA por perla). El DNA se amplifica en esta emulsión generando cientos de copias. A continuación, éstas perlas son transferidas a pozos sobre un chip de fibra óptica, las cuales se incuban con las enzimas ADN polimerasa, ATP sulfurilasa, luciferasa y apirasa, y con los sustratos luciferina y adenosin-5-fosfosulfato (APS). Allí, la ADN polimerasa incorpora el dNTP a la cadena naciente y se liberará pirofosfato que proviene de la

formación del enlace fosfodiéster y es convertido a (ATP), en presencia de la APS. El ATP producido reaccionará con la luciferasa para generar oxiluciferina y producirse luz en proporciones equivalentes a las cantidades de ATP producidas. La emisión de luz será detectada por una cámara CCD. Finalmente, la apirasa removerá el ATP y dNTPs no incorporados. Una vez realizado esto, se repite el ciclo con un nuevo dNTP.

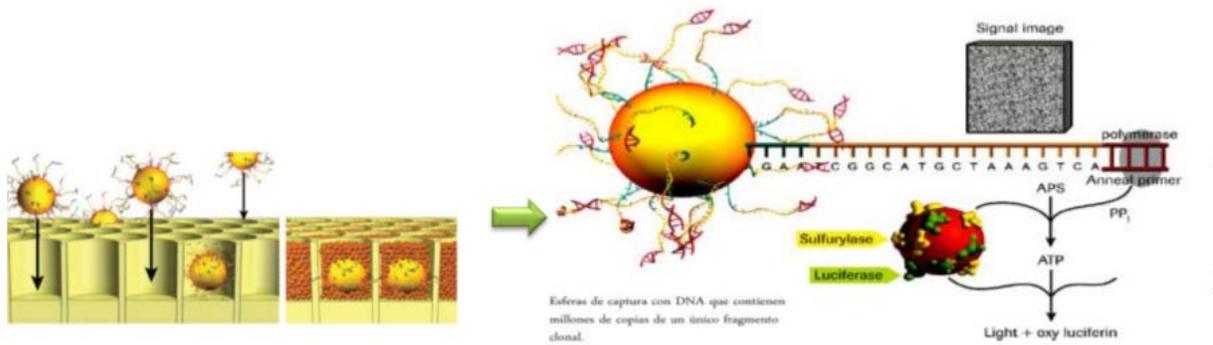


Fig 2. Pirosecuenciación (Secuenciador 454 ROCHE).  
(Tomado de (Mardis 2008) y <http://symposium.com/2013/05/roche-454-pyrosequencing/>)

La segunda tecnología de secuenciación, ha sido desarrollada por Life Technologies con el Ion Torrent, (PGM) en el año 2010, que se basa en el registro de los cambios de pH mediante un ion semiconductor. Brevemente, durante el proceso de síntesis, cada vez que se incorpora un nuevo nucleótido a la cadena, se libera un protón ( $H^+$ ) modificando el pH. Este ciclo se repite varias veces con la adición de un único nucleótido, que es detectado con los cambios de pH (LifeTechnologies, 2014).

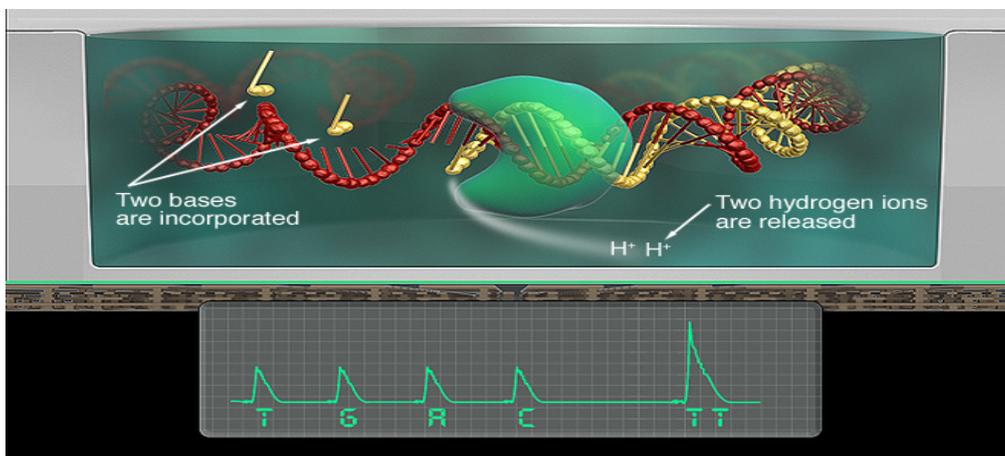


Fig 3. Tecnología semiconductor (Ion Torrent).

(Tomado de <https://www.thermofisher.com/co/en/home/life-science/sequencing/next-generation-sequencing/ion-torrent-next-generation-sequencing-technology.html>)

Finalmente, la tecnología de secuenciación masiva más utilizada que salió al mercado se dio a conocer en el 2006 con la empresa Solexa (luego adquirida por Illumina), está utiliza un mecanismo de amplificación en puente para la generación de clusters, que permite la amplificación de bibliotecas ciclo a ciclo, no dependiente de la procesividad de la polimerasa.

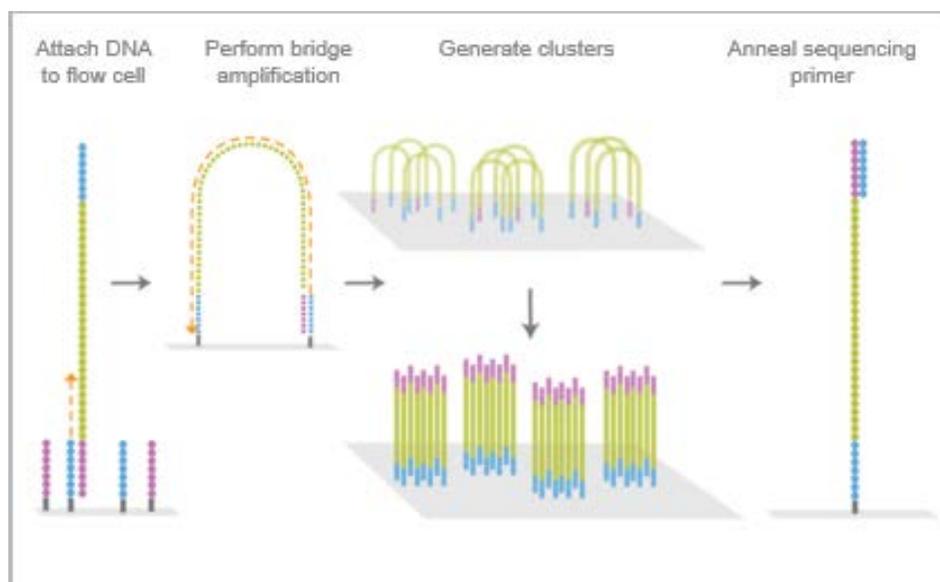


Fig 4. Amplificación en puente Illumina  
(Tomado de [www.illumina.com/applications/forensics.ilmn](http://www.illumina.com/applications/forensics.ilmn))

El método de secuenciación usado por Illumina es la secuenciación por síntesis (SBS), que se basa en la incorporación base a base, uso de nucleótidos terminadores y detección del fluoróforo reversible mediante la toma de imágenes, a medida que se incorpora cada dNTP. Luego los fluoróforos son eliminados (revertidos) para permitir la incorporación de la siguiente base sin acumulación de señales. Cada detección de cada ciclo de incorporación es independiente de la anterior, con lo cual se lleva a cabo una competencia entre los cuatro dNTPs terminadores presentes durante cada ciclo de secuenciación, minimizando el riesgo de incorporaciones erradas

por excesos de algún nucleótido y eliminando los errores específicos del contexto de secuencia, incluso dentro de regiones de secuencias repetitivas u homopoliméricas.

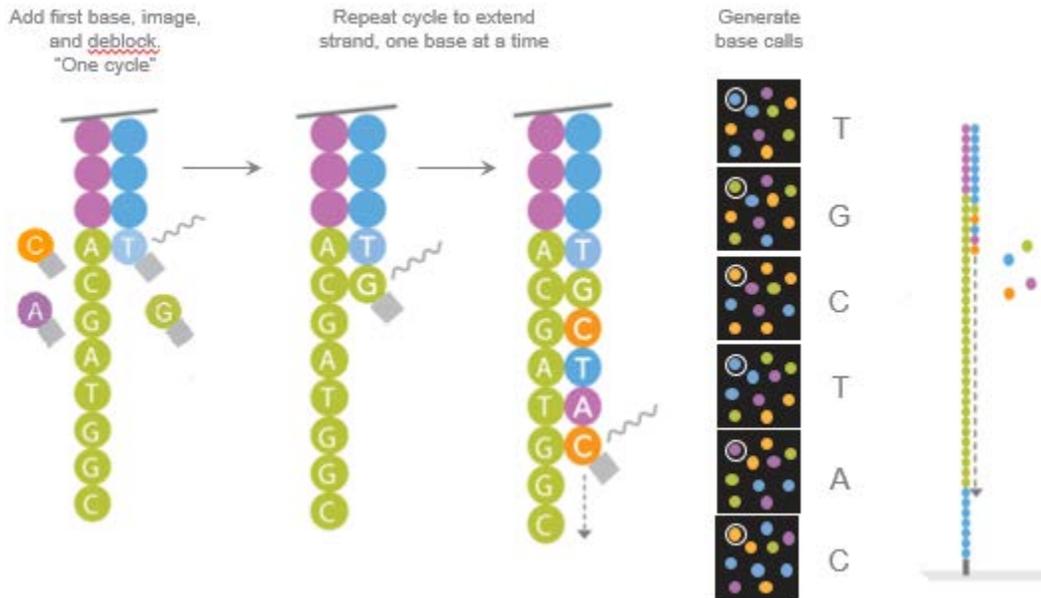


Fig 5. Secuenciación por síntesis Illumina  
(Tomado de [www.illumina.com/applications/forensics.ilmn](http://www.illumina.com/applications/forensics.ilmn))

Adicionalmente, la marca Illumina ha diseñado un kit para el uso en genómica forense con el equipo de secuenciación masiva MiSeq FGx™, que incluye además de marcadores de identificación, marcadores de ancestría biogeográfica y fenotípica (Illumina, n.d.).

Esta empresa, ha realizado estudios de validación de desarrollo basados en las directrices de validación forense establecidas por la SWGDAM (A. Swgdam, 2012) y (A. Swgdam, 2016) y centró sus ensayos en los parámetros de sensibilidad, análisis de mezcla, estabilidad (inhibidores y degradación), exactitud y precisión.

Los resultados de su validación mostraron valores mayores al 99% de exactitud y precisión en la tipificación de alelos de la línea celular control 2800M, tanto para STR como para SNPs. La tasa de éxito, asociada con la obtención de perfiles completos usando 96 muestras con el set de primers (DPMA para 153 loci), fue mayor al 88% y en 32 muestras, con el set de primers (DPMB para 231 loci), fue del 68%.

Adicionalmente, reportaron tres loci SNP's (rs10776839, rs7041158, rs6955448) con un mayor nivel de desbalance intralocus que la mayoría de los loci presentes. Por otra parte, en los ensayos de sensibilidad se observaron genotipos completos, precisos y reproducibles en concentraciones de 62.5 pg a 1 ng de ADN de entrada.

Finalmente, en los ensayos de mezclas se detectaron alelos compartidos y no compartidos del contribuyente minoritario al menos en el 5% de las señales del contribuyente mayoritario. Los ensayos de degradación mostraron información genética en más del 50% de los 230 amplicones blanco menores a 205 nucleótidos de longitud y se observaron perfiles parciales del templado en cantidades de 7 pg. Todos los resultados muestran un sistema robusto, preciso y semi-automatizado (Jager et al., 2017).

### **4.3 Estudios de poblaciones de referencia forense para marcadores SNPs**

Son varios los aspectos que deben tenerse en cuenta para diseñar un estudio poblacional con marcadores tipo SNP, que garanticen tanto la representatividad de la muestra en la población de referencia que se quiere estudiar, como la eficiencia a priori del conjunto de loci analizados, para ser usados con fines de identificación forense.

En primer lugar, la cantidad de loci que deben ser analizados para alcanzar un Poder de Discriminación (PD) comparable con marcadores multialélicos. En segundo lugar, el tamaño muestral que se requiere, para alcanzar frecuencias alélicas mínimas, definidas a priori, con un límite de confianza alto (Chakraborty, 1992). En tercer lugar, la variación que pueda presentarse entre las poblaciones objeto de estudio y su impacto sobre el PD y la Probabilidad de coincidencia al azar (RPM), el Equilibrio genético (HWE) y el Desequilibrio por ligamiento (LD). En cuarto lugar, la posibilidad de que exista estructura al interior de las poblaciones definidas en el muestreo ( $F_{st}$ ), para lo que es necesario una documentación suficiente de la composición étnica de la población, su historia de poblamiento, migraciones masivas y estudios previos realizados sobre la ancestría de los pobladores. En quinto lugar, la distribución demográfica de las poblaciones en estudio, según los datos censales disponibles y proyectados a la actualidad. Un aspecto adicional refiere a la necesidad de demostrar que los individuos

muestreados no tengan vinculo de parentesco y Finalmente, es necesario considerar aspectos moleculares o genómicos que puedan sesgar los resultados de un locus SNP muestreado.

En principio, la información que aportan los SNPs para identificar la población a la que pertenece una persona, o para su individualización, está restringida al polimorfismo que presentan con solo 3 posibles genotipos por marcador, por lo cual el poder de identificación depende en gran medida de la cantidad de SNP usados. De aquí que se necesite analizar alrededor de 50 - 70 marcadores, para alcanzar un PD equivalente a 13 - 27 STRs. Lo anterior, puede representarse como el mínimo número de loci SNPs que se requieren para diferenciar dos individuos tomados al azar de la población en estudio. Al respecto, Vallone y Cols, clasificaron 70 loci SNPs, según los niveles observados de heterocigocidad, encontrando que para diferenciar los 189 individuos de su estudio, necesitaron solo de 12 SNPs siempre que su Heterocigocidad fuese superior a 0.45 (Vallone, Decker, & Butler, 2005), lo mismo, se consigue solo con 3 STRs de alta Heterocigocidad. En relación con la evaluación de la hipótesis de Equilibrio Hardy Weinberg, encontraron 10 de 210 pruebas que indicaban una desviación de los resultados esperados del modelo, lo cual es razonablemente esperado solo por azar; Se observaron 6 resultados significativos en caucásicos, y 2 en afroamericanos e hispanos.

El Poder de Discriminación (Jones, 1972) o la Probabilidad de Identidad que se alcanza en diferentes muestras poblacionales, con un kit multiplex de última generación, que puede co-amplificar 24 loci STRs (*Global filer™* o *Fusion™*) en plataformas CE, está entre  $1 - (2 \text{ a } 7 \times 10^{-26})$  [www.promega.es/products/pm/genetic-identity/population.../power-of-discrimination/](http://www.promega.es/products/pm/genetic-identity/population.../power-of-discrimination/). Por su parte, un kit como *ForenSeq* de *Illumina* que analiza 95 SNPs de identidad, en plataformas NGS/MPS, supera al parecer notablemente el poder de identidad. Casals y Cols, encontraron valores de  $1 - (3.1 \times 10^{-38})$  en poblaciones de Cataluña, que se incrementan hasta,  $1 - (5.9 \times 10^{-73})$ , si se tienen en cuenta además los 56 STRs que incluye el kit (Casals et al., 2017). La expectativa entonces sobre la potencialidad de los múltiples SNPs /NGS, para desempeñarse en aplicaciones forenses es muy alta y está justificando la transición de las tecnologías actuales de identificación humana a las plataformas NGS/MPS.

Al tiempo que se incrementa el poder de identificación al incorporar más loci SNP al estudio, también se incrementa la posibilidad de que estos no sean independientes en su segregación

meiótica y tengan índices de recombinación muy bajos, pudiendo heredarse ligados, con lo cual no podría aplicarse la ley del producto para las frecuencias genotípicas entre loci. Al respecto Kidd et al, analizaron 19 SNPs distribuidos en 9 cromosomas, 4 de ellos conteniendo más de un SNP. En 14 marcadores no observaron desviación significativa de la hipótesis de equilibrio, pero los valores observados en 5 loci no tienen explicación clara sobre su origen, ni pudo establecerse una relación coherente con una posible asociación no al azar entre loci (Kidd et al., 2006). Por su parte Dixon et al, analizaron 20 SNPs encontrando valores no significativos de desequilibrio o asociación entre ninguno de los análisis pareados realizados (Dixon, 2005).

Para el kit ForenSeq de Illumina, objeto de esta validación, los 94 loci SNP de identidad que lo conforman están distribuidos en todos los 23 cromosomas humanos así:

Cromosoma	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22
Nº de SNPs	6	5	5	4	5	4	4	4	5	5	4	5	4	4	3	4	4	4	2	4	5	4

Tabla 1. Distribución de los 94 loci SNPs del kit Signature Prep ForenSeq en los 22 cromosomas.

Es necesario por lo tanto determinar si existe desequilibrio de ligamiento o asociación no al azar entre ellos para definir su estado de independencia entre loci.

Con respecto al tamaño muestral requerido para estudios poblacionales con marcadores polimórficos de ADN, hemos tomado los criterios publicados por Chakraborty y Cols en 1992; En resumen, se puede inferir, a partir de estimas conservadoras, la probabilidad de obtener frecuencias alélicas representativas del polimorfismo de cada locus en niveles de confianza muy altos. Por ejemplo, para un STR que tiene en promedio 15 alelos, una muestra de 300 individuos de una población puede ser suficiente para detectar todos los alelos del locus en frecuencias superiores al mínimo definido, así no aparezcan en la muestra todos los genotipos posibles. Para el caso de los SNPs, por su condición de marcadores bialélicos, es decir, que solo se generan tres genotipos posibles, las muestras pueden ser mucho más pequeñas, siempre que se establezca un umbral conservador para las frecuencias alélicas mínimas (Science, 1996). En el estudio de Vallone et al, 10 alelos de 630 estuvieron por debajo del umbral  $5/2N$  en tres poblaciones estudiadas (caucásicos, africanos e hispanos). Así, una estima conservadora de la frecuencia

alélica mínima sería 0,034, 0,035 y 0,057 respectivamente para aquellos alelos raramente observados (Vallone et al., 2005).

En relación con el análisis entre poblaciones, debe tenerse en cuenta que los STRs tienen muchos alelos con bajas frecuencias en la mayoría de las poblaciones, por lo que la probabilidad de coincidencia (RMP), no es muy diferente entre ellas, en su lugar, los SNPs pueden mostrar frecuencias muy diferentes entre poblaciones, pudiendo variar de 0 a 1, lo que genera que la RMP dependa mucho de las frecuencias alélicas usadas para el cálculo (Kidd et al, 2006). El SNP ideal para propósitos de identificación, será entonces aquel que tenga alta heterocigocidad y frecuencias alélicas idénticas en todas las poblaciones independientemente de la población estudiada. Alta heterocigocidad maximiza la información en cada SNP y un bajo  $F_{st}$  minimiza el efecto del azar entre poblaciones. Así, la combinación de altas heterocigocidades y bajos valores de  $F_{st}$  incrementarían la eficiencia de un panel forense de SNPs. Afortunadamente, no todos los SNPs son variantes en sus frecuencias alélicas entre poblaciones; Estos principios genético poblacionales han sido tenidos en cuenta para la selección de marcadores SNP de más impacto en la identificación individual por su homogeneidad de frecuencias, así como para determinar cuáles son más útiles como marcadores de ancestría (AIMs).

Por otra parte, la ancestría conocida de las poblaciones a estudio, debe dirigir también el diseño del muestreo poblacional; En América Latina, se observa una amplia variación de ancestría entre países, así como entre individuos dentro de los países, debido a la gran mezcla entre nativos americanos e inmigrantes de Europa y África (Ruiz-Linares et al., 2014). La población colombiana también ha demostrado gran variabilidad genómica, distribuida en diferentes regiones etnogeográficas.

Para (Ruiz-Linares et al., 2014), el análisis por regiones proporciona un entorno ventajoso, dado que permite analizar factores genéticos, físicos y sociales en relación con la etnia y la autopercepción racial. A partir del análisis de 1.659 individuos basados en 30 marcadores tipo SNPs, informativos de ancestría, encontró una variación geográfica significativa en ancestría a través de las regiones, la cual fue consistente con la historia demográfica e información censal y reportó que en Colombia se evidencia mayor ascendencia africana en las regiones costeras (particularmente en el Pacífico), ancestría europea más alta en las áreas andinas centrales y mayor ascendencia nativa en el sudoeste y este del país (Amazonia), aunque el estudio no contó

con muchos datos para éstas áreas. Los valores de ascendencia porcentual hallados, según la contribución de las poblaciones ancestrales fueron: 60% europeos, 29% nativos americanos y 11% africanos.

Adicionalmente, (Rishishwar et al., 2015), realizaron un estudio sobre la ascendencia y patrones de mezcla para secuencias del genoma completo de 60 colombianos no emparentados, de Medellín, que fueron secuenciados como parte del Proyecto de los 1000 genomas, los resultados de sus análisis mostraron ascendencia predominantemente europea; la secuencia del genoma promedio mostró un 74,6% de europeos, un 18,1% Asiático (Nativo Americano) y 7.3% de ascendencia Africana.

Por lo anterior, dada la diversidad étnica de nuestro país, las diferencias tan extremas en la densidad de población entre regiones y la autopercepción de ascendencia relativa, la cual está sesgada a la influencia de la apariencia física y no a ascendencia estimada genéticamente (Ruiz-Linares et al., 2014), es necesario integrar, tanto los criterios de representatividad forense en una muestra poblacional, como los criterios de representatividad genético poblacional. Los primeros, se refieren a poblaciones de referencia requeridas para valorar estadísticamente hallazgos genéticos en los cotejos de rutina que hacen los laboratorios. Esto es, generalmente, poblaciones residentes de ciudades o regiones donde suceden los hechos delictivos, sin importar su origen o su ancestría. Los segundos están definidos por la dinámica poblacional, su historia de poblamiento, los sesgos de apareamiento, migraciones masivas dadas por razones históricas, etc.

Para Colombia, y más en el contexto de esta investigación, es necesario considerar desde la visión forense, cuáles son las regiones más afectadas por el conflicto armado y en particular por el fenómeno de desaparición forzada, y además, la posible estructura genético poblacional de estas zonas de conflicto.

Por ejemplo, las regiones donde se han observado altos índices de desaparición son, entre otras, los llanos orientales y la región pacífica. Las primeras se han conformado principalmente de poblaciones de colonos procedentes de las capitales andinas orientales, que migraron en la primera violencia de los años 40 y 50. Aunque, originalmente, en estas vastas regiones colombianas predominaban poblados indígenas, su población siempre ha sido minoritaria y la población inmigrante fue mayor y creciente y ha desplazado a las ancestrales. Por esta razón, no

esperamos mayor diferenciación de las poblaciones llaneras, en relación con lo observado en las poblaciones de la cordillera oriental.

Por el contrario, las poblaciones que conforman los departamentos del Pacífico colombiano, deben diferenciarse étnicamente ya que tienen dos claros componentes: Mestizo, Indígena o incluso caucásico en las capitales de departamento, como Pasto y Popayán, ubicadas sobre la cordillera occidental, y al tiempo, un componente afrodescendiente predominante en pueblos del litoral y las riveras de los ríos, que caracteriza también el litoral pacífico de los departamentos del Valle y principalmente del departamento del Chocó, donde en la mayoría de poblaciones, más del 85% de sus habitantes son negros.

Así, vemos conveniente documentar, a través de marcadores bialélicos, estas diferencias, de modo que contemos con referentes poblacionales más fundamentados y realísticos para ser usados como referencias forenses en los casos de identificación.

Adicionalmente, los tamaños muestrales deben guardar proporción con la distribución demográfica documentada en datos censales; al respecto, el censo de 2005 proyectado a 2015-2020 permite hacer estimas más aproximadas a la realidad demográfica nacional actual y es una ayuda para definir el tamaño de las muestras que representan mejor las poblaciones de interés forense, no obstante, se hace necesario fijar un tamaño mínimo de muestra para aquellas regiones de muy baja densidad poblacional.

Un último factor a considerar, es que los SNPs de uso forense, deberían estar localizados exclusivamente en secuencias de los genomas humanos y representados en una sola copia. Recientemente, varios investigadores han descubierto una abundancia de variación submicroscópica en el número de copias de segmentos de ADN en rangos de kilobases a megabases de tamaño. Deleciones, inserciones, duplicaciones y complejas variantes multi-sitios, colectivamente denominadas CNVRs o variaciones en número de copias, se han encontrado en humanos y otros mamíferos. Un total de 1447 copias de CNVRs que pueden abarcar solapamientos, ganancias o pérdidas adyacentes, cubriendo el 12% del genoma, fueron recientemente identificadas. El hallazgo del SNP rs999842, por ejemplo, localizado dentro de un CNVR, no debería ser considerado un buen marcador para propósitos forenses (Giardina et al., 2007).

## 5. METODOLOGÍA

### 5.1 Selección de muestras de ADN humano

#### 5.1.1 Muestras para los experimentos de validación

##### 5.1.1.1 *Muestras para los ensayos de Precisión y Reproducibilidad*

Muestras de 5 individuos masculinos del *staff* del Laboratorio de Genética del INML y CF, quienes aceptaron participar y firmaron un consentimiento informado autorizando el uso de su muestra en este estudio. Se seleccionaron individuos de sexo masculino, para evaluar los marcadores asociados a cromosoma Y, los cuales se encuentran incluidos en el kit TG ForenSeq DNA Signature Prep usado para este proyecto.

Adicionalmente, consideramos importante incluir muestras del personal del *staff* del laboratorio en los experimentos, ya que sus perfiles genéticos analizados con el kit objeto del estudio, pueden servir de control de contaminación, una vez esta tecnología sea implementada en sus laboratorios para el análisis de casos forenses.

De cada uno de los cinco (5) individuos se obtuvieron muestras de sangre y saliva, colectadas sobre tarjetas tipo FTA<sup>®</sup> y se analizaron por triplicado.

El número de individuos a participar fue escogido teniendo en cuenta recomendaciones de las sociedades científicas internacionales, tanto en el ámbito metrológico como forense quienes afirman que, para realizar validaciones de métodos analíticos, con un número total de 50 muestras analizadas (usando de 5-10 muestras por experimento), se observan resultados con una significancia estadística robusta, muy similares a los obtenidos con un número elevado de muestras (A. Swgdam, 2016), (Butler, 2006).

Adicionalmente, para estos ensayos, se usaron muestras de ADN extraídas de restos óseos de casos procesados en el laboratorio de Genética de la Seccional Meta del INML y CF, con cuantificaciones previas superiores a 1 ng y que se encontraban previamente tipificadas para marcadores STR. Una vez seleccionadas las muestras, se cuantificaron nuevamente con el kit Quantifiler™ Trio DNA Quantification de Life Technologies, Carlsbad, CA, en el sistema de PCR en tiempo real de Applied Biosystems®, 7500, de acuerdo a las instrucciones del fabricante, para normalizarlas a una concentración de 1 ng/uL y luego amplificarlas con el kit ForenSeq™.

***5.1.1.2 Muestras para los ensayos de exactitud, reproducibilidad, sensibilidad, Low Copy Number (LCN) y variación del volumen de librerías para secuencia:***

Estos ensayos se realizaron con ADN genómico humano masculino, de la Línea celular 2800M (Promega®), de la cual se conoce el perfil genético para todos los loci incluidos en el Kit ForenSeq – Illumina. Por esta razón, también se usó como control positivo de amplificación y secuenciación. De cada una de las muestras, se realizaron tres réplicas, en cada ensayo.

**5.1.2 Muestras para el análisis comparativo entre tecnología de nueva generación y electroforesis capilar en casos complejos**

***5.1.2.1 Muestras para el ensayo de complejidad asociada a la calidad de la muestra***

Para este ensayo, se utilizaron muestras óseas de casos de identificación, asociados al conflicto armado colombiano, que no presentaron resultados positivos con kits STR multiplex convencionales, analizados por tecnología de electroforesis multi capilar policromática.

Interesó evaluar piezas óseas que presentaron bajas concentraciones de ADN o signos de degradación de la molécula (EFG con efecto *slope*); que estuvieron sometidas a ambientes de inhumación extremos, particularmente, exposición a altas temperaturas y que se pudo

documentar la antigüedad de los restos. Según lo anterior, los criterios de selección fueron los siguientes:

- Cuantificación de ADN: menor a 62 pg.
- Resultado Negativo: 0 sistemas A-STRs detectados, luego de 2 o más reprocesos de extracción de ADN.
- Antigüedad: de 30-40 años, medida desde la fecha de desaparición.
- Ambiente peri mortem: restos óseos con niveles o no de exposición a altas temperaturas, se eligieron las muestras, según la apariencia del fragmento óseo (No incinerado, parcial y totalmente Incinerado)

#### ***5.1.2.2 Muestras para el ensayo de complejidad asociada a pedigrís poco informativos, bajo variables de número y tipo de familiares para cotejo***

Para los casos de muestras con escasa representación de familiares de referencia, se seleccionaron 9 casos pertenecientes a la base de datos del Instituto Nacional de Medicina Legal y Ciencias Forenses -Convenio ICBF, con parentescos previamente confirmados, a partir de los cuales se pudieron realizar simulaciones de parentescos complejos comunes en casos de identificación Ej. Dúos padre-hijo, Hermanos, medio hermanos, abuelos.

Se tuvieron en cuenta los siguientes criterios de selección:

- Número y tipo de familiares analizados (Casos de filiación con presunto padre, madre y 2 o más hijos, abuelos)
- Resultado del cotejo (Resultado no excluyente)
- Número de marcadores STRs analizados (con mínimo 15 marcadores biparentales analizados)
- Población de referencia asociada al caso: Lugar de radicación de la demanda de filiación (Radicadas en algún departamento de la región Andina Central)
- Consentimiento informado para uso de estudios poblacionales firmado

### 5.1.3 Muestras para los estudios poblacionales:

Las muestras para el estudio poblacional de los polimorfismos SNPs, fueron obtenidas del biobanco de tarjetas FTA de casos de filiación del Instituto Nacional de Medicina Legal y Ciencias Forenses, Convenio ICBF-INML.

Se adelantó el estudio poblacional de tres regiones etnogeográficas del país, de relevancia forense, dado que allí se concentra un gran número de víctimas del conflicto armado, y porque constituyen casos de mayor complejidad, ya sea por el estado de degradación de las muestras óseas disponibles para análisis o por la ausencia de familiares directos de los desaparecidos, escenarios característicos del conflicto. Son ellas: Región Andina central, Región de la Orinoquia o Llanos orientales y Región Pacífica.

A partir de un estudio demográfico, con datos de poblaciones colombianas, obtenidos del censo de 2005 proyectado a 2018 según el DANE, y otras fuentes de información de los gobiernos departamentales, se elaboró un listado representativo de los municipios del país que componen cada una de las 3 regiones de interés y se definió el porcentaje de población en cada uno de ellos con respecto al total nacional y al total por región analizada.

Posteriormente, se realizó una preselección de casos de paternidad de los últimos 5 años (2012-2016) atendidos y radicados en la base de datos SIFMELCO del INML y CF, esperando que este periodo representara además la estructura actual de las poblaciones colombianas que han sido objeto de desplazamiento forzado interno.

Los casos seleccionados cumplieron los siguientes requisitos:

- Estar registradas en la base de datos del Instituto Nacional de Medicina Legal y Ciencias Forenses -Convenio ICBF.
- Las muestras provenían de las de las regiones Andina central, Orinoquia y Pacífica de Colombia, aclarando que esta región corresponde al municipio donde fueron atendidas, que generalmente es el mismo sitio de residencia o de origen de los muestradantes.

- Los apellidos de padre y madre no se compartían, si era así se eliminaba uno de los dos individuos, como medida de seguridad para evitar que se seleccionaran personas emparentadas.
- Todas las muestras tenían el respectivo consentimiento informado de aceptación para estudios poblacionales firmado.
- Todas las muestras habían sido previamente tipificadas para marcadores STRs

Para lo anterior, se realizó una matriz en Excel que permitió hacer la revisión de los casos preseleccionados a partir del archivo del Convenio ICBF-INML, capturando la siguiente información:

- Año, N° caso, región, departamento y municipio
- Consentimiento Informado Si-No-Nd (No Diligenciado)
- Año, Mes, Lugar Nacimiento (Departamento y municipio) y apellidos del padre
- Año, Mes, Lugar Nacimiento (Departamento y municipio) y apellidos de la madre

Teniendo en cuenta los criterios anteriores, y lo reportado por Chakraborty, 1992, con el fin de observar todos los genotipos esperados en cada locus SNP, se estableció un número de 100 muestras (sexo 1:1) para poblaciones con alta densidad poblacional (Andes central) y de 50 personas, para poblaciones con densidades menores (Orinoquía y Pacífico), con el fin de obtener frecuencias alélicas superiores a 0,01 con un intervalo de confianza de 95%, en el rango de edades reproductivas, lo cual se supone, ya que la muestra se obtendrá a partir de individuos parentales de investigaciones de paternidad.

Una vez tuvimos la información de los casos en la matriz descrita anteriormente, se hizo un muestreo aleatorio en Excel sobre los números de casos definidos para cada región seleccionada, más un 20% de casos adicionales, con el fin de elegir únicamente los casos que cumplieran con los criterios de selección.

Todas las muestras, fueron anonimizadas, para guardar la confidencialidad de la información.

Finalmente se solicitaron las muestras en tarjeta FTA de las personas seleccionadas y se procedió a su análisis genético por NGS como lo indica el fabricante.

## **5.2 Transporte y almacenamiento de muestras**

Las muestras de sangre o saliva colectadas sobre tarjetas tipo FTA<sup>®</sup>, se encuentran almacenadas a temperatura ambiente en el Instituto Nacional de Medicina Legal y Ciencias Forenses.

Las muestras de ADN extraídas de restos óseos, de casos procesados en el laboratorio de Genética de la Seccional Meta del INML y CF, se transportaron vía valija con su respectiva cadena de custodia (Resolución 000606 del 25 de septiembre de 2006, por la cual se adopta el instructivo para la utilización de los rótulos de Cadena de Custodia de elementos materiales probatorios o evidencias físicas en el INML y CF y Resolución No. 0-23969 del 11 de Julio de 2016, por medio de la cual se adopta el Manual de Procedimientos para Cadena de Custodia).

- Todos los extractos de ADN fueron embalados en tubos eppendorf, sellados y rotulados con la siguiente información:
- Código de la muestra
- Tipo de muestra
- Fecha de extracción

Y se transportaron en cadena de frío desde la ciudad de Villavicencio hasta Bogotá en donde fueron analizadas por NGS.

## **5.3 Extracción de ADN**

Las muestras colectadas sobre tarjetas tipo FTA<sup>®</sup>, se recortaron en círculos de 1.2 mm, luego se lavaron con 100 µl de TBE 1X, 2 min de agitación y se centrifugaron a 1800 rpm eliminando el buffer residual de cada pozo; el recorte con el ADN adherido, se sometió directamente a amplificación con el kit ForenSeq<sup>™</sup>.

## 5.4 Preparación de librerías y secuenciación

La amplificación y preparación de librerías se realizó en los termocicladores Biorad fast C1000 y/o Veriti del laboratorio de Bogotá del INML, siguiendo las recomendaciones del fabricante, igualmente que la purificación y normalización de las librerías (Illumina, 2015b).

La secuenciación se realizó en un equipo secuenciador de próxima generación NGS tipo MiSeq FGx® de Illumina, usando los kits ForenSeq prep ® DNA, con la mezcla A, siguiendo las especificaciones de la casa comercial (Illumina, 2015d, 2015c, 2015b). Finalmente el análisis de datos se realizó utilizando el software UAS (ForenSeq™ Universal Analysis Software) (Illumina, 2015a).

El Kit ForenSeq prep ® incluye 2 tipos de marcadores: SNPs y STRs; los primeros diseñados tanto para identificación individual (SNP-ID), Investigación de ancestría (AIM-SNPs), como para determinación del fenotipo (Phe-SNPs). Por su parte los STRs son autosomales (A-STRs), de cromosoma Y (Y-STRs) y de cromosoma X (X-STRs)

Feature	Number of Markers <sup>a</sup>	Amplicon Size Range (bp)	Included in DNA Primer Mix A	Included in DNA Primer Mix B <sup>b</sup>
Global Autosomal STRs	27	61–467	Yes	Yes
Y-STRs	24	119–390	Yes	Yes
X-STRs	7	157–462	Yes	Yes
Identity SNPs	94	63–231	Yes	Yes
Phenotypic SNPs	22	73–227	No	Yes
Biogeographical Ancestry SNPs	56	67–200	No	Yes

Fig 6. Loci del kit ForenSeq DNA Signature Prep (Tomado de [www.illumina.com/applications/forensics.ilmn](http://www.illumina.com/applications/forensics.ilmn))

## 5.5 Criterios para calificación de resultados

Para el análisis de los genotipos en las muestras se tuvieron en cuenta los siguientes resultados:

- Genotipo concordante: cuando los alelos del genotipo resultan idénticos a los esperados, siempre que superen tanto el umbral analítico, como el umbral estocástico o de interpretación establecidos.
- Genotipo ambiguo: Alelos cuyo número de reads se encuentren entre el umbral analítico y el de interpretación.
- Alelo discordante: Alelo supera el umbral de análisis, pero difiere del reportado por CE
- Alelo no asignado: Alelos con conteo de reads por debajo del umbral de análisis

## 5.6 Diseño experimental

### 5.6.1 Validación del kit “TG MiSeq FGx Reagent kit” para el equipo de NGS MiSeq® de Illumina

#### 5.6.1.1 Experimento de exactitud y precisión

El experimento de **exactitud** se realizó con el fin de determinar mediante análisis repetitivos, bajo las condiciones propias del laboratorio y siguiendo las recomendaciones de la casa comercial, en qué medida la asignación de alelos por locus (% de concordancia en la asignación alélica) de una misma muestra, coincide con los valores reportados. Para este ensayo se analizaron 6 réplicas de la Línea celular 2800M a una concentración de 1 ng, tipificada y publicada como control positivo previamente en la literatura, en un montaje de 32 muestras.

Parámetros analizados:

- EXACTITUD (STRs o SNPs) =  $100 * \frac{\text{Total de alelos concordantes}}{\text{Total del número de alelos}}$

*Ecuación 1*

**La precisión** en los corridos se evaluó mediante análisis repetitivos bajo las mismas condiciones del experimento de exactitud y recomendaciones de casa comercial, para caracterizar el grado de acuerdo entre la genotipificación de una serie de muestras comparadas entre ellas mismas. Como se indicó inicialmente, para este ensayo, se analizó por triplicado, 5 muestras de sangre y 5 muestras de células epiteliales bucales sobre tarjetas FTA de las mismas 5 personas del Staff de personal masculino del Laboratorio de Genética de la Seccional Meta del INML y 5 muestras anónimas de ADN extraídas de restos óseos (RO) que fueron obtenidas de casos procesados en el mismo laboratorio, previa cuantificación y normalización a una concentración de 1 ng.

Adicionalmente, para evaluar si existían diferencias en los resultados al realizar montajes con diferente número de librerías secuenciadas simultáneamente en un mismo kit, se efectuaron montajes con 32, 40, 88 y 96 muestras por ensayo.

En cada ensayo, se analizó el número de loci observados, la precisión como concordancia entre replicas y la exactitud en el genotipo con respecto a un patrón. El perfil genético usado como patrón para las muestras de sangre, saliva y restos óseos, fue el obtenido previamente por CE convencional e incluye, A-STRs y Y-STRs. Para la línea celular 2800M, se utilizó el perfil reportado en el Kit Signature Prep ForenSeq™, para 153 marcadores (A-STRs, Y-STRs, X-STRs e ID-SNPs).

Finalmente, se calculó la desviación estándar y el coeficiente de variación del número de *reads* obtenidos por locus, entre las tres réplicas de cada muestra (teniendo en cuenta que los alelos superaran los umbrales de análisis e interpretación pre-definidos).

Parámetros analizados:

- PRECISION:  $100 * \frac{\# \text{ de observaciones del genotipo/haplotipo llamado más frecuente}}{\text{Número total de todos los genotipos}}$

Número total de todos los genotipos

***Ecuación 2***

- EXACTITUD DE GENOTIPO (STRs) =  $100 * \frac{\text{Total de genotipos concordantes}}{\text{Total del número de genotipos}}$

***Ecuación 3***

- Profundidad de cobertura (DoC): Desviación estándar del promedio del número de *reads* por locus.

**5.6.1.2 Experimento de reproducibilidad**

Se realizó para determinar si existen diferencias en cualquiera de los parámetros cualitativos o cuantitativos, comparando los resultados obtenidos entre dos peritos del Instituto Nacional de Medicina Legal, con el mismo proceso de capacitación. Cada uno procesó en dos ensayos diferentes de 96 muestras cada uno y por triplicado: la línea Celular 2800M y 3 muestras de sangre del Staff del laboratorio, tipificadas previamente por electroforesis capilar. Los ensayos se realizaron usando los mismos procedimientos y bajo las mismas condiciones de los experimentos de exactitud y precisión.

Los resultados obtenidos entre los dos analistas se compararon, aplicando la prueba *t-student* para determinar si existen o no diferencias estadísticamente significativas (< 0.05).

Parámetros de comparación:

- Profundidad de cobertura (DoC): Desviación estándar del promedio del número de reads por locus.
- Concordancia de genotipo: Para cada locus, la asignación genotípica obtenida por los dos analistas es la esperada. Se calcularía con la *Ecuación 3*.

**5.6.1.3 Experimentos de sensibilidad y definición de umbrales**

#### 5.6.1.3.1 Sensibilidad

Se determinaron los niveles de sensibilidad de la prueba y los efectos estocásticos, resultantes de bajas concentraciones de ADN en las muestras analizadas.

Para este ensayo se realizaron diluciones seriadas por triplicado, entre 0.0078 ng y 1 ng (1000, 500, 250, 125, 62.5, 31.2, 15.6 y 7.8 pg), de la línea celular 2800M y de tres muestras de RO previamente tipificadas por CE y secuenciadas en otros experimentos de esta misma validación, y se amplificaron bajo las mismas condiciones y equipos de los experimentos anteriores.

El límite de detección del ensayo se evaluó teniendo en cuenta el efecto del número de librerías por montaje, por lo cual se diseñaron 2 ejercicios de sensibilidad en montajes de 40 y 88 librerías simultáneas.

Parámetros de comparación:

- Profundidad de cobertura (DoC): Desviación estándar del promedio del número de *reads* por locus.
- Concordancia de genotipo: Calculada en términos de exactitud y precisión, aceptando los alelos que superaron los umbrales por defecto programados en el equipo (AT: 5% y IT 4,5%). Se calculó con la *Ecuación 3*.

#### 5.6.1.3.2 Umbrales analíticos

Para la validación de desarrollo, la firma Illumina basados en estudios empíricos, usó 1,5% como Umbral analítico (AT) y 4,5% como Umbral estocástico o de interpretación (IT); los dos umbrales fueron determinados para un locus, multiplicando el valor en porcentaje del parámetro analítico, por la suma del conteo de lecturas (*reads*) en ese locus y en caso de baja cobertura, se asumió un mínimo de 600 *reads* para la determinación de los valores umbrales. En la validación interna se propuso determinar los umbrales analíticos de la siguiente manera:

- *Umbral de análisis (AT)*

Para determinar el umbral de análisis en la asignación alélica tanto de loci STRs como de SNPs, se amplificaron por triplicado 8 controles negativos y se tomaron 3 desviaciones estándar por encima del promedio del número de reads de las secuencias que aparezcan por locus.

**Método 1.**  $UA=2(Y_{max}-Y_{min})$ .

**Método 2.**  $UA=Promedio + (10 \times Desviación\ Estándar)$

**Método 3.**  $UA=3 \times Desviación\ Estándar$

- *Umbral Estocástico o de Interpretación (IT)*

Se determinó un umbral en número de reads con el cual puede asumirse que por encima de este valor en un locus determinado no ha habido pérdida alélica de un heterocigoto (alelo hermano) y al detectar un solo alelo, éste podrá considerarse homocigoto.

Este umbral se estableció, mediante el seguimiento de los sistemas heterocigotos observados en una muestra, luego de realizar diluciones seriadas por triplicado en el experimento de sensibilidad. Se determinó el número de reads del alelo sobreviviente observado en cada locus (Gill et al., 2008).

#### *5.6.1.3.3 Balance de Heterocigotos: Diferencia máxima del número de reads entre alelos hermanos del Heterocigoto*

Este análisis permitió establecer un valor de referencia por encima del cual, el investigador puede sospechar la presencia de más de un individuo en la muestra analizada.

Se estableció el porcentaje de diferencia entre alelos de cada heterocigoto de las muestras analizadas en el experimento de sensibilidad, así:

$$\text{Balance Heterocigotos} = 100 * \frac{\text{Número de reads del alelo de menor altura}}{\text{Número de reads del alelo de mayor altura}}$$

***Ecuación 4***

Se calculó el promedio de los porcentajes de diferencia obtenidos y se multiplicó por 3 veces la desviación estándar (SD) de los datos.

Concordancia de genotipo: Para cada locus, la asignación alélica obtenida es la esperada. Se calculó con la *Ecuación 3*

#### **5.6.1.4 Experimentos de LCN (Low Copy Number)**

Estudio de comparación entre los protocolos de termociclaje del fabricante del Kit ForenSeq y 2 protocolos con ciclos aumentados en los ciclajes de la PCR1.

Se tomaron las diluciones realizadas en el ensayo de sensibilidad con la línea celular 2800M por triplicado. Incluyendo control positivo de 1ng/ul y control negativo.

Las muestras se amplificaron durante la PCR1, teniendo en cuenta dos modificaciones LCN a los termociclajes sugeridos por el fabricante (8 y 10 ciclos en los dos ciclajes de la PCR targeting), así:

- Tratamiento 2: Dos (2) ciclos adicionales en cada termociclaje  
98 °C durante 3 minutos  
**10** ciclos de:  
96 °C durante 45 segundos  
80 °C durante 30 segundos  
54 °C durante 2 minutos, con el modo de rampa especificado  
68 °C durante 2 minutos, con el modo de rampa especificado  
**12** ciclos de:  
96 °C durante 30 segundos  
68 °C durante 3 minutos, con el modo de rampa especificado  
68 °C durante 10 minutos. Temperatura a 10 °C.
- Tratamiento 3: Un (1) ciclo adicional en el primer ciclaje y manteniendo el segundo igual  
98 °C durante 3 minutos  
**9** ciclos de:  
96 °C durante 45 segundos

80 °C durante 30 segundos  
54 °C durante 2 minutos, con el modo de rampa especificado  
68 °C durante 2 minutos, con el modo de rampa especificado  
10 ciclos de:  
96 °C durante 30 segundos  
68 °C durante 3 minutos, con el modo de rampa especificado  
68 °C durante 10 minutos. Temperatura a 10 °C.

Cada uno de los tratamientos se analizó por separado en montajes de 40 muestras. Los procesos de indexación, purificación, normalización y secuenciación de las librerías se realizaron siguiendo las instrucciones del fabricante.

Posteriormente, se seleccionó el tratamiento con el mejor resultado en concordancia y número de genotipos obtenidos y se aplicó a muestras de ADN de concentraciones inferiores a 62 pg obtenidas de restos óseos, en un volumen total de reacción de 15ul.

#### ***5.6.1.5 Experimentos de variación del volumen de librerías para secuencia***

Se realizaron 4 ensayos de 40 muestras cada uno, variando el volumen de cada librería o del pool de librerías que van a secuenciarse en un mismo montaje, con el objetivo de determinar con que volumen se obtiene el mayor número de marcadores con los genotipos esperados.

Para los tratamientos 1 y 2, se tomaron todas las diluciones realizadas en el ensayo de sensibilidad con la línea celular 2800M por triplicado. Para los tratamientos 3 y 4 se tomaron solo las diluciones entre 62.5 pg y 7.8 pg, incluyendo control positivo de 1ng/ul y control negativo para los 4 ensayos.

Las muestras se amplificaron en las mismas condiciones sugeridas por el fabricante y al momento de realizar la secuenciación se realizaron los siguientes ensayos para crear el pool de librerías y adicionarlo al cartucho de reactivos:

- Ensayo 1: 5 ul de cada librería para un volumen de pool final de 8.5 ul
- Ensayo 2: 5 ul de cada librería para un volumen de pool final de 10 ul
- Ensayo 3: 8.5 ul de cada librería para un volumen de pool final de 7.5 ul
- Ensayo 4: 10 ul de cada librería para un volumen de pool final de 7.5 ul

Finalmente, se seleccionaron los ensayos que mostraron los mejores resultados y se combinaron 2 o más estrategias a concentraciones de ADN más bajas, con el fin de observar si el resultado en el número de genotipos asignados concordantes podría mejorar.

## **5.6.2 Análisis comparativo entre tecnología de nueva generación y electroforesis capilar en casos complejos**

### **5.6.2.1 Complejidad asociada a la calidad de la muestra (muestras óseas):**

Teniendo en cuenta, los criterios para la selección de las muestras descritos anteriormente, se analizaron 3 muestras anónimas de ADN extraídas de restos óseos (RO), para cada variable:

- Muestras óseas antiguas (30-40 años) y no incineradas
- Muestras óseas antiguas (30-40 años) y parcialmente incineradas
- Muestras óseas antiguas (30-40 años) y totalmente incineradas

Todas las muestras seleccionadas fueron cuantificadas por PCR en tiempo real, se tuvo en cuenta la concentración máxima de ADN alcanzada debido al deterioro de la pieza y se analizaron por NGS, teniendo en cuenta las modificaciones realizadas al protocolo del fabricante que mejores resultados mostraron en los experimentos de LCN y variación del volumen de librerías para secuencia.

### **5.6.2.2 Complejidad asociada a pedigrís poco informativos, bajo variables de número y tipo de familiares para cotejo:**

Se escogieron 5 escenarios diferentes de parentescos complejos en 9 casos de filiación de las bases de datos del convenio ICBF-INML, como se describió anteriormente, con resultados no excluyentes y un valor de probabilidad de paternidad mayor de 99.9999%, con el mayor número de marcadores STRs analizados por CE previamente.

Los 9 casos seleccionados se analizaron nuevamente con el kit Signature Prep ForenSeq™ en el equipo MiSeq FGx™, y se calculó el valor de LR con 92 SNPs y 25 A-STRs (Se eliminaron 2 SNPs y 2 A-STRs que no amplificaron en algunas de las muestras) en el software Familias. Se usaron las frecuencias poblacionales de la región Andina Central, obtenidas en este proyecto, para determinar la variación del valor del LR obtenido entre los dos tipos de marcadores (A-STRs y SNPs).

Por cada escenario de parentesco, se realizaron 4 ejemplos de simulaciones diferentes, estos fueron:

- 2 hermanos completos: tomando 1 de referencia y el otro dubitado
- 2 medio hermanos: tomando 1 de referencia y el otro dubitado
- Padre e hijo con madre ausente
- Un abuelo paterno y nieto con madre ausente
- Un abuelo paterno, nieto y madre

### **5.6.3 Estudios poblacionales de 3 regiones colombianas para 89 SNPs de identificación, del kit forenseq illumina**

Las frecuencias alélicas resultantes del estudio poblacional, se evaluaron por conteo directo utilizando una plantilla en Excel de diseño propio del laboratorio y por medio del software de análisis genéticos; GDA-Power Marker v3.2; Se asumió como frecuencia alélica mínima (f.min), el valor obtenido a partir de la ecuación de Budowle, B.1996

$$p_{\min} = 1 - [1 - (1 - \alpha)^{\frac{1}{c}}]^{2n}$$

Los parámetros de genética poblacional y forense como: frecuencias alélicas y genotípicas, frecuencia alélica mínima (f.min.), heterocigocidad observada y esperada (***Ho***, ***He***), Equilibrio Hardy-Weinberg (***HWE***), Desequilibrio por ligamiento (***LD***), Subestructura poblacional (***Fst***) fueron estimados utilizando el programa GDA-Power Marker v3.2; además el Poder de discriminación (***PD***), información del contenido polimórfico (***PIC***), el poder de exclusión (***CE***), probabilidad de coincidencia (***RMP***) se estimaron con el programa Power stat – Promega.

## 6. RESULTADOS Y DISCUSIÓN

Los datos obtenidos para todos los ensayos fueron analizados usando el software UAS (ForenSeq™ Universal Analysis Software) (Illumina, 2015a); se revisó que cumplieran con las métricas de calidad (QC) establecidas por el software:

- Densidad de Cluster entre 400 a 1650 k/mm<sup>2</sup>
- Porcentaje de Cluster que pasaron el filtro  $\geq 80\%$
- Phasing  $\leq 0.25\%$
- Prephasing  $\leq 0.15\%$

Adicionalmente, se tuvo en cuenta el mínimo de profundidad de cobertura de 650 reads para la determinación de los umbrales por locus. El umbral analítico fue estimado en 10 reads por alelo (1.5% por 650 reads) y el umbral de interpretación en 30 reads por alelo (4.5% por 650 reads) pre-definidos en el software.

### 6.1 Experimentos de validación

#### 6.1.1 Experimento de exactitud y precisión

**Exactitud:** Se analizaron 6 réplicas de la línea celular 2800M a una concentración de 1 ng/μl, en un montaje de 32 muestras, utilizando el kit ForenSeq™, con los primers (DPMA para 153 loci). Todas las muestras superaron ampliamente los 85.000 reads, en un rango de lecturas entre 250.000 a 350.000.

Las métricas de calidad obtenidas en el MiSeq FGx™, para este ensayo fueron las siguientes:

- Densidad de cluster 1119 k/mm<sup>2</sup>
- Porcentaje de Cluster que pasaron el filtro 93.91%
- Phasing 0.203%
- Pre-Phasing 0.085%.

Como se indicó en la metodología, los alelos asignados en el ensayo de exactitud, fueron considerados como concordantes si eran idénticos a los reportados para los mismos loci STR por CE tradicional o para los mismos SNPs, en el kit Signature Prep ForenSeq™. Además, debían superar los umbrales analíticos y de interpretación pre-establecidos.

Para el análisis se tuvo en cuenta, cada alelo del genotipo de un locus de forma independiente, contando como 2 si ambos eran iguales al esperado, o 1 si solo uno de ellos era concordante.

El porcentaje de exactitud se calculó como lo indica la ecuación 1:

- $$\text{EXACTITUD (STRs o SNPs)} = 100 * \frac{\text{Total de alelos concordantes}}{\text{Total del número de alelos}}$$

***Ecuación 1***

Todos los 153 genotipos esperados fueron detectados con un 100% de exactitud, en todos los tipos de marcadores (STRs y SNPs) analizados y en las 6 réplicas de la Línea celular 2800M incluidas, como lo muestra la Tabla 2 y Anexo 1.

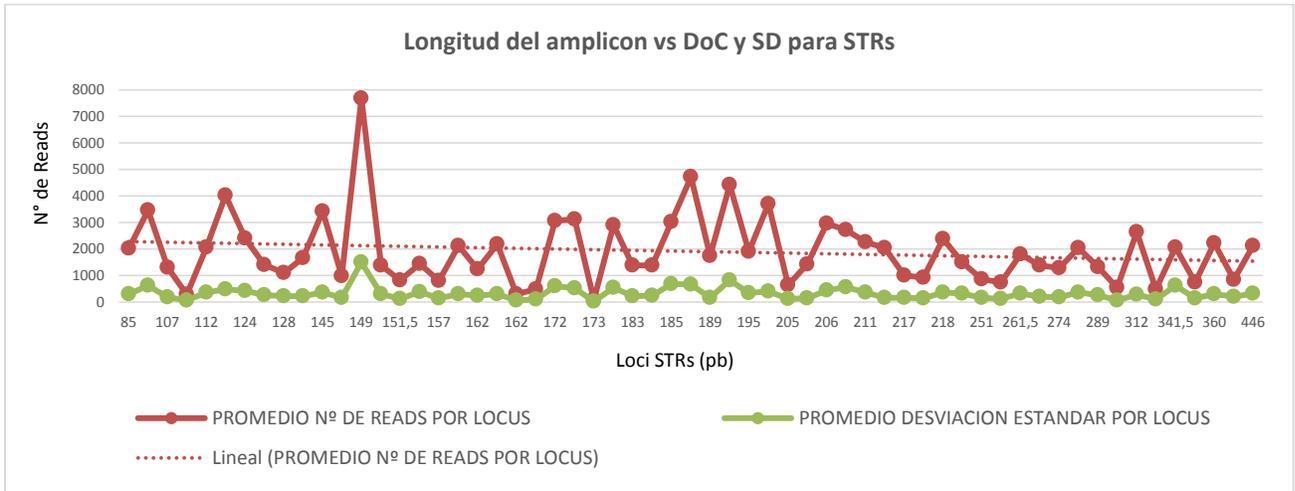
	<b>A-STR</b>	<b>Y-STR</b>	<b>X-STR</b>	<b>SNPs</b>	<b>TOTAL STR</b>	<b>TOTAL STR Y SNPs</b>
Nº Genotipos	28	24	7	94	59	153
% Exactitud	100	100	100	100	100	100

Tabla 2. Experimento de Exactitud para la línea celular 2800M a 1ng/uL

El promedio de *reads* observados para A-STRs, está ente 1591 ±141, para Y-STRs entre 2077±1589 y para X-STRs entre 2307 ± 1150 (Promedio ± Desviación estándar). Para el locus DXS10103, la profundidad obtenida fue solo de 73 ± 26 *reads*, lo cual es consistente con lo reportado en la literatura (Vallone et al., 2005). Este locus también se ha asociado con una mayor probabilidad de *dropout*. Por su parte, los marcadores SNPs mostraron una menor profundidad de cobertura (DoC), comparada con los *loci* microsatélites (417 ± 307).

Las profundidades observadas muestran una leve tendencia a disminuir a medida que se incrementa la longitud del amplicón (Figura 7 A y B), sin embargo, también se observan algunos loci de amplicones de menor tamaño, con bajas profundidades, que no podrían explicarse con este argumento (Anexo 3).

(A)



(B)

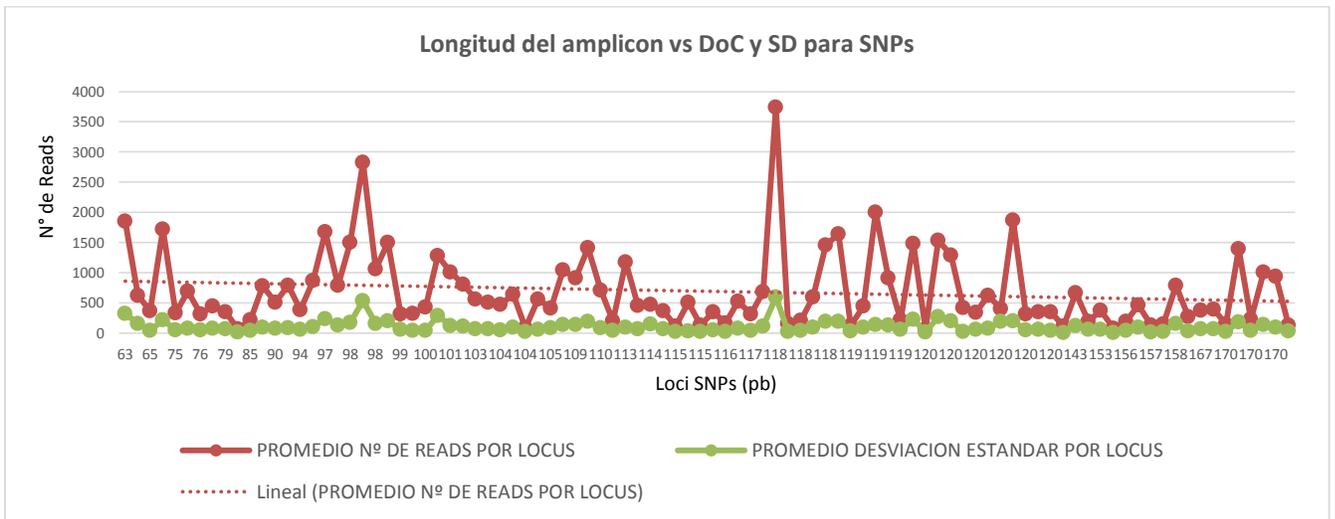


Fig 7. Experimento de Exactitud, calculada en profundidad de cobertura (DoC) para (A) STRs y (B) SNPs. Promedio desviación estándar-longitud del amplicón (verde), Promedio Número de reads por locus (rojo) para réplicas de la línea celular 2800M, amplificadas con el kit Signature Prep ForenSeq™. Eje horizontal en pares de bases.

Si bien existe una relación entre longitud del amplicón y su profundidad, no parece ser de alto impacto; (Guo, Yu, Zhang, & Li, 2017), valoraron la relación dependiente del tamaño alélico, solo en el 7,58% de los datos comparables, entre marcadores tipo STRs y 14.27% para los SNPs, mediante un estudio similar al actual (Vallone et al., 2005). Todo parece indicar que las causas de la variación en la profundidad de las secuencias son múltiples: la longitud del fragmento amplificado es solo una de ellas, la cantidad e integridad del ADN templado, la eficiencia y fidelidad de las amplificaciones que se realizan durante el procedimiento de NGS: PCR inicial, PCR enriquecedora, los ciclos de extensión de fragmentos durante la formación de *clusters* y finalmente la secuenciación.

**Precisión:** Como se explicó en la metodología, las muestras usadas para este experimento, fueron secuenciadas en cuatro ensayos con diferente número de muestras: 32, 40, 88 y 96. Se determinó el número de loci observados, la precisión entre réplicas y la exactitud en el genotipo con respecto a un patrón.

En la tabla 3 se muestran las métricas de calidad obtenidas en los 4 montajes:

Numero de muestras por ensayo	32	40	88	96
Densidad de clúster (k/mm2)	1119	976	875	1012
% de clusters que pasaron el filtro	93.91	94.94	96.49	94.05
Phasing %	0.203	0.172	0.165	0.151
Pre-Phasing %	0.085	0.080	0.057	0.063

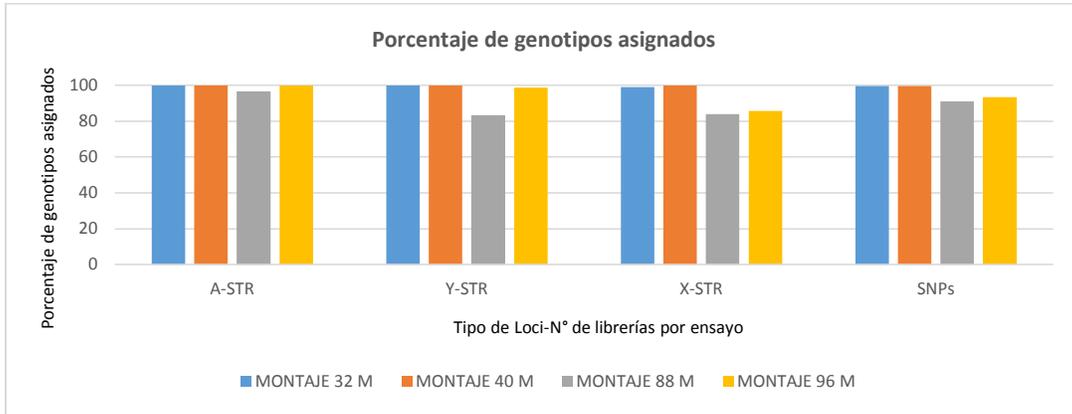
Tabla 3. Métricas de calidad obtenidas para ensayos con diferente número de librerías amplificadas con el kit Signature Prep ForenSeq™

No obstante haber superado las métricas de calidad, el porcentaje de genotipos asignados en los montajes con 88 y 96 librerías, disminuye hasta el 83% para Y-STRs, 84% para X-STRs y 91% para SNPs. Para los A-STRs solo se observa disminución en el porcentaje de genotipos, para el montaje de 88 librerías (Figura 8 (A) y Anexo 2).

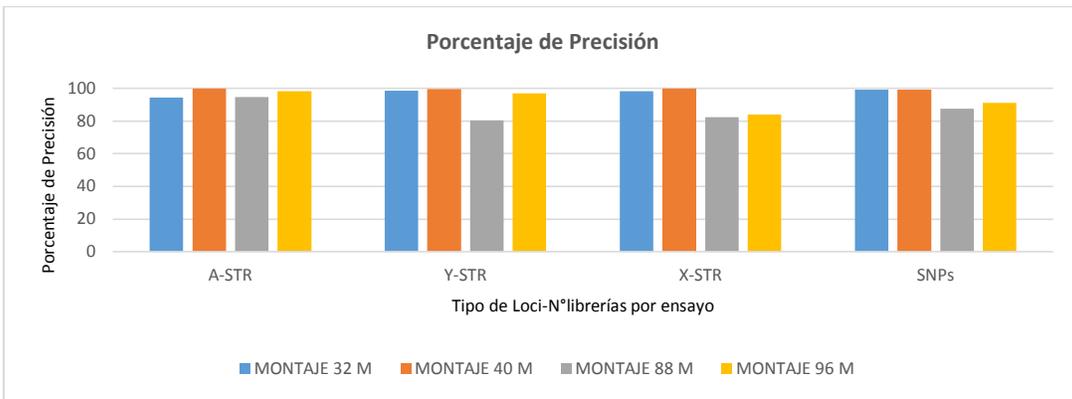
Por otro lado, el porcentaje de exactitud es menor para marcadores Y-STRs en el ensayo de 88 muestras (Figura 8 (C) y Anexo 2), mientras que la precisión disminuye hasta el 80% y 82%,

para marcadores Y-STRs y X-STRs respectivamente y hasta el 87% para SNPs en los ensayos de mayor número de muestras (Figura 8 (B) y Anexo 2).

(A)



(B)



(C)

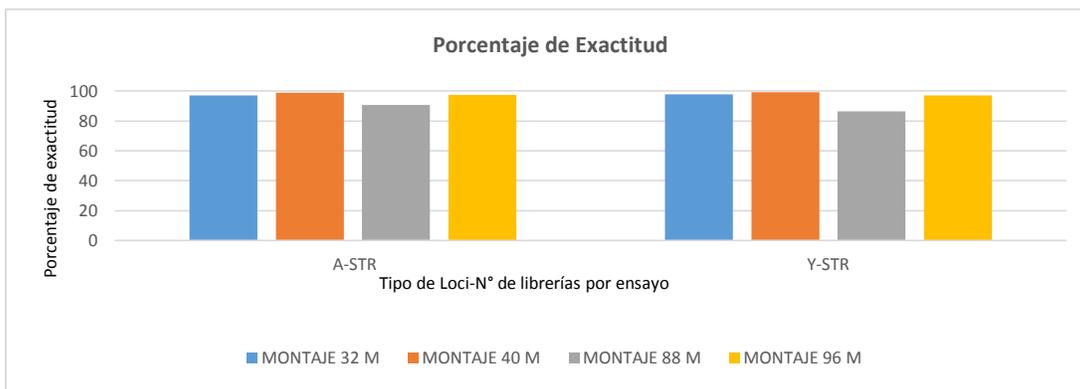


Fig 8. Porcentaje de genotipos asignados, precisión y exactitud, para marcadores STRs y SNPs en ensayos con diferente número de muestras, amplificadas con el kit Signature Prep ForenSeq™.

(A) Porcentaje de genotipos asignados (B) Porcentaje de Precisión (C) Porcentaje de exactitud para marcadores A-STR, Y-STR y SNPs en ensayos con 32 (azul), 40 (rojo), 88 (verde) y 96 (morado) muestras.

Los resultados anteriores muestran una menor eficiencia en los ensayos de mayor número de muestras con respecto al porcentaje de genotipos asignados y a la precisión entre muestras, excepto para los A-STR donde solo en el ensayo de 88 muestras disminuye el porcentaje de genotipos asignados. Del mismo modo, se observa un menor porcentaje de exactitud para los marcadores A-STR y Y-STRs en este ensayo.

Lo anterior, podría explicarse por la menor densidad de *cluster* ( $875 \text{ k/mm}^2$ ) obtenida, que indica un bajo número de librerías amplificadas.

Algunos marcadores parecen ser especialmente sensibles al montaje de un número alto de muestras; Así, los sistemas DYS392, rs1294331, rs12997453, rs1357617, rs13182883, rs13218440, rs10092491, rs2920816, rs1031825 no fueron detectados en los ensayos de 88 y 96 muestras, mientras que los sistemas DXS10103, rs1355366, rs7041158, rs1736442, rs719366 no se detectaron en ninguno de los montajes realizados en la mayoría de muestras (Anexo 2).

Es posible que el tipo de muestra influya de alguna manera en la eficiencia del ensayo, ya que como se indicó en el experimento anterior de exactitud con la línea celular a 32 muestras, se obtuvo el 100% de los genotipos concordantes, a diferencia de este ensayo con muestras de sangre y saliva en tarjetas FTA y restos óseos, en el cual, a la misma cantidad de muestras ensayadas, el porcentaje de genotipos obtenidos para marcadores X-STRs y SNPs fue del 99% y la exactitud fue del 97%.

Como se ha reportado en la validación de desarrollo (Jager et al., 2017) y en otras publicaciones (Almalki et al., 2017; Vallone et al., 2005; Xavier & Parson, 2017), el sistema DXS10103 es uno de los sistemas más sensibles a *dropout* y 6 de los SNPs no detectados en este estudio (rs1294331, rs1357617, rs10092491, rs1031825, rs7041158 y rs1736442) también fueron reportados anteriormente con *locus dropout*. Compartimos la sugerencia de otros autores para los fabricantes, de realizar algunas mejoras en el desarrollo de los primers de estos loci que presentan menor rendimiento y que están incluidos en el Kit Signature Prep ForenSeq.

Por otro lado, observamos que el número de librerías secuenciadas en un mismo montaje, también afecta el número de *reads* por muestra y define el número de perfiles completos. De esta manera, a mayor número de librerías, es menor el número de *reads* y por lo tanto se disminuye también el porcentaje de perfiles completos obtenidos, lo cual apoya el hallazgo descrito anteriormente en cuanto a exactitud y precisión dependientes del número de librerías secuenciadas simultáneamente (Figura 9).

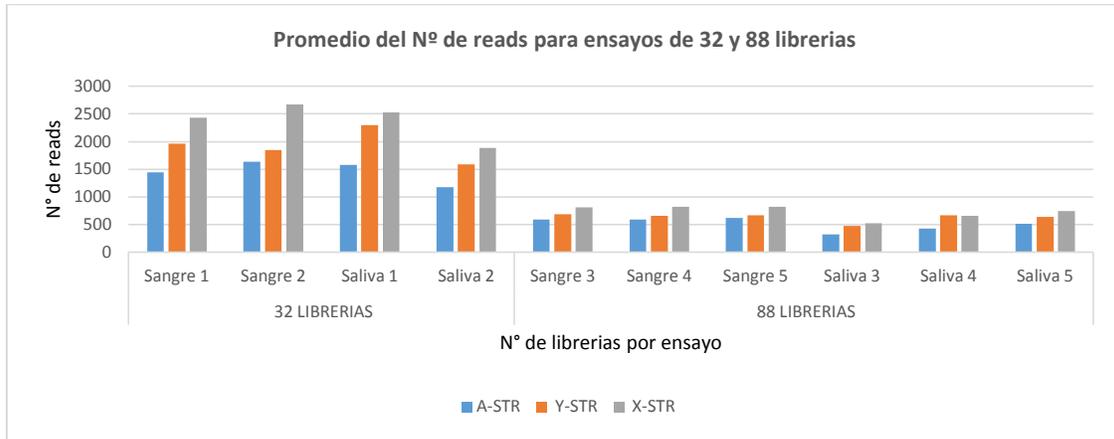


Fig 9. Promedio del N° de reads para ensayos de 32 y 88 librerías en muestras de sangre y células epiteliales bucales sobre tarjetas FTA, amplificadas con el kit Signature Prep ForenSeq™ (Azul) A-STR, (rojo) Y-STR y (Verde) X-STR

Al comparar los tipos de muestras analizadas, se observa un mejor desempeño del método en las muestras sobre tarjetas FTA (Figura 10). Los restos óseos, aun a una concentración comúnmente usada en los laboratorios forenses de 1ng/uL muestra resultados inferiores en número de *reads*. Los resultados no difieren mucho de los obtenidos comúnmente en CE.

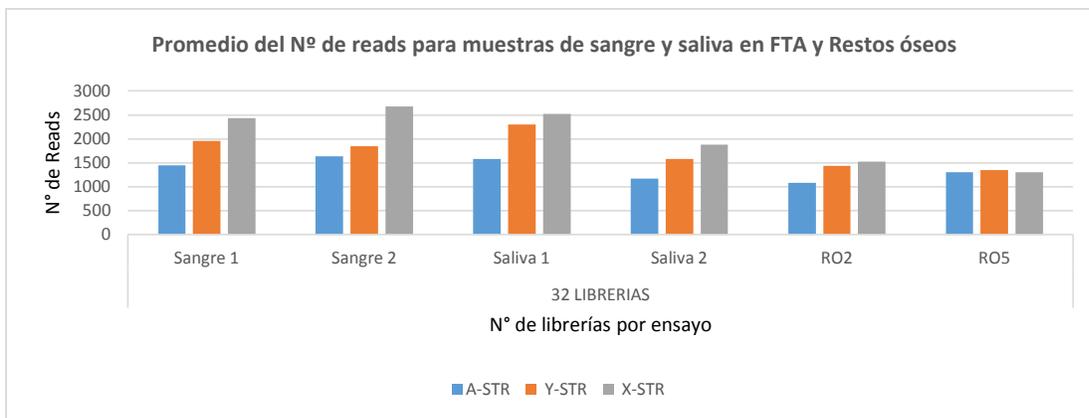


Fig 10. Promedio del N° de reads para ensayo de 32 librerías en muestras de sangre y células epiteliales bucales en FTA y Restos óseos, amplificadas con el kit Signature Prep ForenSeq™ (Azul) A-STR, (rojo) Y-STR y (Verde) X-STR

Ante condiciones similares de concentración de ADN de entrada y de *clusters* obtenidos, así como en el número de librerías secuenciadas simultáneamente, se observa que el número de reads entre los alelos de un mismo locus se mantiene sin mayor variación, sin importar la muestra analizada y varían en magnitud, dependiendo del locus analizado; por ejemplo, los alelos detectados en el locus DXS10103 se mantienen entre 30 y 100 *reads*, mientras que para DYS576 entre 2000 y 6700 *reads*. Esta observación permite inferir que la variación en profundidades podría depender de la especificidad o de la concentración de los *primers* usados en la técnica, en la PCR target. (Anexo 3).

Por otra parte, la gran diferencia entre valores de profundidad de cobertura que se observan entre diferentes loci no tendría impacto si estos siempre superaran los umbrales de análisis establecidos. No obstante, para muestras críticas, como pueden ser la mayoría de las evidencias forenses, esta variabilidad puede ser un problema a la hora de definir aportantes a una mezcla, o de reconocer efectos de degradación de una muestra, etc, y habría que tener en cuenta los loci que presenten los rangos menores de cobertura como los más propensos al error.

En consecuencia, la variación observada en el método para los valores de profundidad de cobertura puede estar determinada en función de varios factores: el número de librerías secuenciadas simultáneamente, el tipo de muestra, el tipo de marcador, el nivel de optimización de las reacciones de amplificación y específicamente el diseño de *primers* para la PCR inicial. Quizás, una mejor optimización en el diseño de *primers* o las condiciones de PCR podrían mejorar u homogenizar las profundidades entre alelos, loci e incluso entre muestras.

El estudio y el diseño experimental realizado en el presente trabajo son buenos porque permiten hacer evidentes estas diferencias. Todos estos factores deben tenerse en cuenta al momento de evaluar muestras forenses.

### **6.1.2 Experimento de reproducibilidad**

Como se mencionó en la metodología, cada perito procesó, dos ensayos diferentes de 96 muestras con la línea Celular 2800M y 3 muestras de sangre del *staff* del laboratorio tipificadas

usando los mismos procedimientos y bajo las mismas condiciones de los experimentos de Exactitud y Precisión descritos previamente.

Los resultados obtenidos entre los dos analistas se compararon teniendo en cuenta el porcentaje de genotipos asignados y la concordancia entre ellos, medida en términos de Precisión y Exactitud de genotipo, esta última analizada únicamente para la línea celular 2800M, que permite la comparación entre todos los 156 loci del kit ForenSeq, calculadas con las Ecuaciones 2 y 3 respectivamente. (Tabla 4 y 5)

No se observaron diferencias entre analistas. Se obtuvo entre el 99% y el 100% de genotipos asignados para los marcadores Y-STRs y A-STRs respectivamente, con una precisión mayor al 97% en los dos ensayos y para los dos peritos analizados. (Tabla 4).

El porcentaje de X-STRs fue mayor al 87%, con porcentajes de precisión entre el 82% y el 87% y por su parte, para SNPs, se obtuvo más del 90% en genotipos con precisiones entre el 88% y el 93% para los dos peritos analizados entre los dos ensayos considerados. (Tabla 4)

		ENSAYO/ PERITO					PROMEDIO STR Y SNPs (%)
		A-STR (%)	Y-STR (%)	X-STR (%)	PROMEDIO STR (%)	SNPs (%)	
Precisión	<b>E1/P2</b>	97	98	86	96	92	94
Genotipos asignados		100	99	88	98	93	95
Precisión	<b>E1/P1</b>	97	99	82	96	93	94
Genotipos asignados		100	99	87	98	94	96
Precisión	<b>E2/P2</b>	99	98	85	97	88	92
Genotipos asignados		100	99	91	99	90	94
Precisión	<b>E2/P1</b>	99	98	87	97	89	92
Genotipos asignados		100	99	91	99	91	94

Tabla 4. Experimento de Reproducibilidad calculada en porcentajes de precisión y número de genotipos asignados

La exactitud para el experimento de reproducibilidad, fue medida únicamente para la línea celular 2800M, comparando con los genotipos reportados para CE tradicional (Para STRs) o para el kit Signature Prep ForenSeq™ (para SNPs). No se observan diferencias significativas entre los 2 peritos y los porcentajes de exactitud son cercanos al 100% para Y-STRs, entre el 81%-86% para X-STRs y entre 76% y el 87% para SNPs entre los dos ensayos. (Tabla 5)

	ENSAYO/PERITO	Y-STR (%)	X-STR (%)	SNPs (%)
EXACTITUD	E1/P2	100	81	87
	E1/P1	100	81	86
	E2/P2	99	86	76
	E2/P1	100	86	82

Tabla 5. Experimento de Reproducibilidad, calculada en porcentajes de exactitud

Adicionalmente se observa que para ambos ensayos de 96 muestras los marcadores en los que se observan *dropouts* son los mismos reportados en el experimento de Precisión, sumando el SNP rs10488710.

Los resultados obtenidos para el promedio del número de *reads* por locus entre los dos analistas se compararon, aplicando la prueba *t-student* para determinar si existen o no diferencias estadísticamente significativas ( $p < 0.05$ ). La prueba indica que existen diferencias significativas entre los dos peritos analizados, variación que podría generarse durante el proceso de preparación de las librerías, teniendo en cuenta que esta etapa no se encuentra automatizada y es de suma importancia obtener una librería adecuada para entrar al proceso de secuenciación, no obstante todas las muestras superaron el número de *reads* establecido como umbrales analítico y de interpretación y los porcentajes de genotipos asignados, exactitud y precisión de los datos entre los 2 analistas son mayores de 87%, 76% y 82%, respectivamente. (Anexo 4).

### 6.1.3 Experimentos de sensibilidad y definición de umbrales

#### 6.1.3.1 Sensibilidad

Se realizaron diluciones seriadas por triplicado, entre 1 y 0.0078 ng (1000, 500, 250, 125, 62.5, 31.2, 15.6 y 7.8 pg), de las muestras de ADN objeto del estudio.

Se decidió analizar el efecto de la disminución de concentración de ADN *in put*, en relación con dos variables que, según la experiencia analítica, afectan la calidad de los resultados:

- El tipo de muestra: Optamos por determinar si los resultados de un experimento de sensibilidad son extrapolables a las muestras calcificadas que son el interés principal de este trabajo, teniendo en cuenta la alta variabilidad demostrada en análisis forenses tradicionales, entre muestras de restos óseos.
- El número de librerías por ensayo: Aunque la validación de desarrollo señala la relación existente entre profundidad de cobertura y número de librerías por ensayo, los ensayos previos realizados en el presente estudio demuestran que este factor es determinante sobre la variación en la profundidad y amerita por lo tanto definir límites de detección específicos para cada tipo de montaje.

#### *6.1.3.1.1 Efecto del tipo de muestra en montajes de 40 librerías: ADN purificado línea celular 2800M*

Las métricas de calidad obtenidas para este ensayo se describieron en la tabla 3; todas superaron los valores recomendados por el fabricante. En la Figura N°11, se observa el Número total de *reads* por muestra, que se obtiene sumando todas las profundidades de cada alelo y de cada locus analizado.

Este indicador permite evaluar de forma general la eficiencia del proceso analítico en los ensayos de sensibilidad, comparando visualmente la altura en *reads* alcanzada por cada muestra del montaje. La validación de desarrollo fijó en 85.000 *reads* el umbral por encima del cual se obtienen perfiles completos y concordantes en todos los marcadores del kit Forenseq. Por debajo de este valor, se pueden presentar *drop out*, *drop in*, *imbalances extremos*, *stutters elevados*, etc por lo que la casa Illumina recomienda revisar cuidadosamente los resultados de las muestras que no alcancen los 85.000 *reads*, más aún si se trata de muestras complejas.

En nuestro ensayo se observó que a concentraciones menores de 125 pg, la muestra no alcanza los 85.000 reads pre-establecidos en el software UAS (Figura 11). Este hallazgo es consistente con lo reportado por los fabricantes, ya que se observan *drop out* en algunos marcadores, desde los 62.5 pg.

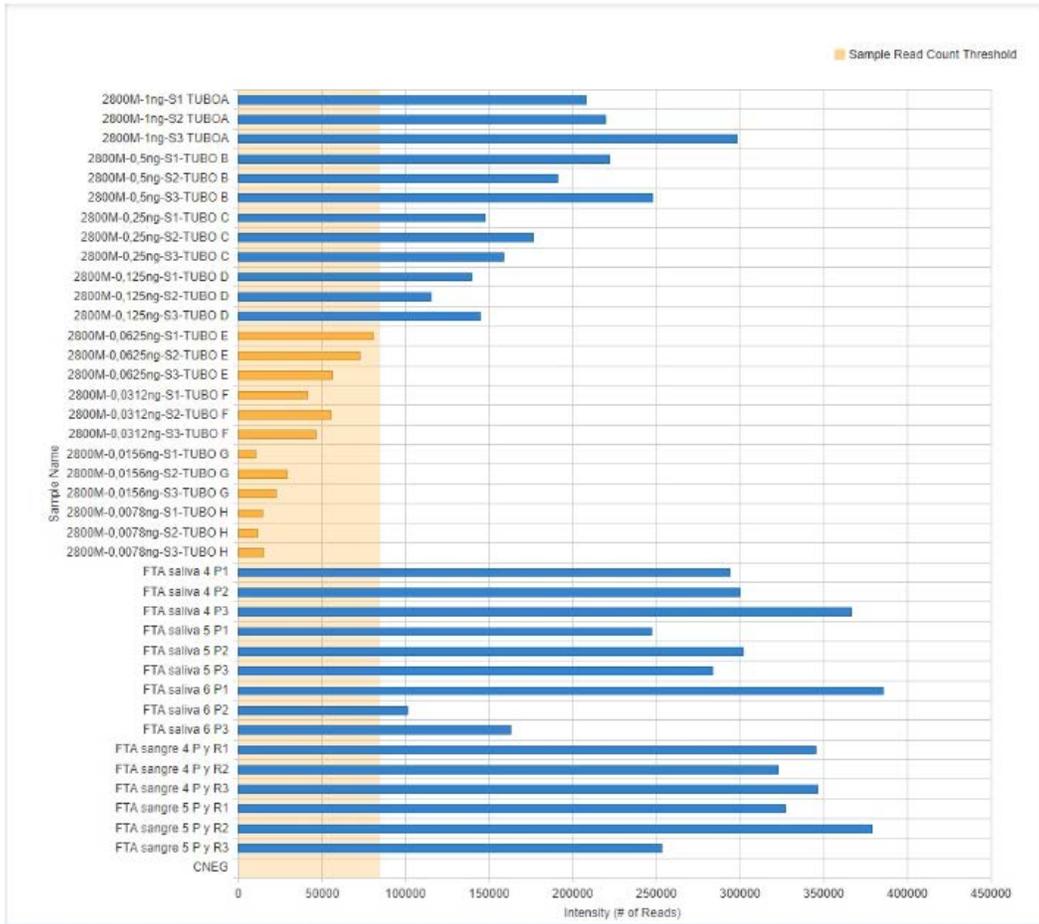


Fig 11. Pantalla de representación del N° total de Reads por muestra, en el ensayo de sensibilidad. Diluciones seriadas entre 1 y 0.0078 ng (1000, 500, 250, 125, 62.5, 31.2, 15.6 y 7.8 pg) de la línea celular 2800M con el kit Signature Prep ForenSeq™

Para la línea celular 2800M en el ensayo de 40 muestras, los resultados mostraron que a concentraciones de ADN entre 1 ng y 62.5 pg, se obtiene el 100% de loci A-STRs y Y-STRs, con valores de concordancias superiores al 93%. Este resultado es coincidente con el obtenido en la validación de desarrollo (Jager et al., 2017), adicionalmente, demuestra el alto nivel de reproducibilidad obtenido como laboratorio usuario de la técnica.

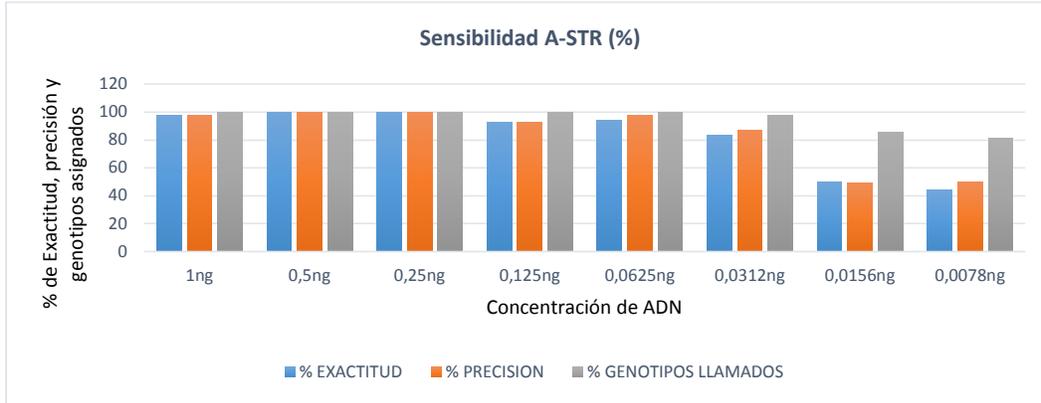
Para concentraciones de hasta 7 pg se recupera el 81 % de A-STRs y el 61% de Y-STRs con valores del 44 y 56% de concordancia respectivamente. Llama la atención que, aunque la asignación genotípica se mantiene en valores relativamente altos para las 2 últimas concentraciones probadas en A-STRs, no se acompaña de valores igualmente altos en la exactitud y precisión de los datos. Solo el 58% de los genotipos asignados son los esperados. Por su parte para los marcadores de Y, la relación entre asignación y concordancia es más estable en todas las concentraciones; incluso en las 2 concentraciones más críticas la proporción de genotipos asignados correctamente es superior al 91%. (Figura 12 (A) (B) y Anexo 5).

En lo que respecta a marcadores X-STRs, se obtuvo el 100% de loci solamente en concentraciones de 0.5 y 1 ng de ADN, con el 100% de exactitud y precisión, mientras que a 7 pg solo el 62% de los loci asignados, fueron exactos. Al igual que los Y-STRs, el comportamiento de los X-STR, es más estable que los A-STRs, pero tanto la asignación genotípica como la exactitud de los datos, comienza a disminuir lentamente desde una concentración de ADN de 0,25 ng. (Figura 12 (C) y Anexo 5).

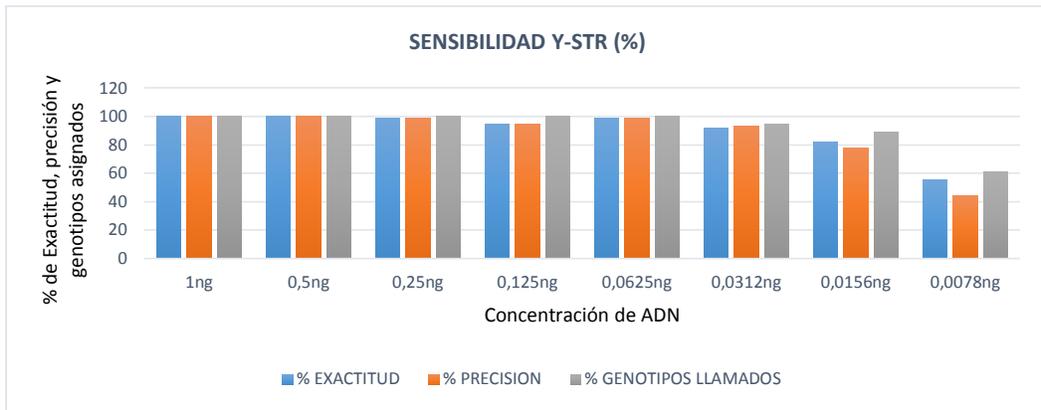
Por el contrario, no se alcanzó el 100% en asignación de genotipos, en ninguna de las diluciones probadas para loci tipo SNPs. Los valores máximos alcanzados fueron del 95% al 99% para concentraciones de ADN entre 125 pg y 1 ng y concordancias entre el 92% y el 98% respectivamente. En la dilución de ADN más baja analizada (7pg), se obtuvo solo el 39% de SNPs asignados y su concordancia fue por mucho del 26 al 30%. (Figura 12 (D) y Anexo 5).

En comparación con los marcadores microsatélites, a partir de 62,5 pg, los marcadores SNPs estudiados demuestran una caída continua en todas las mediciones de exactitud y precisión realizadas. (Figura 12 (D)). El efecto es particularmente notable para 6 loci: rs1294331, rs1357617, rs13182883, rs13218440, rs1736442, rs1031825, en los cuales no se detectó ningún genotipo por debajo de los 125 pg. Una situación más extrema mostro el locus rs1355366, que no fue detectado para ninguna concentración de ADN *in put*.

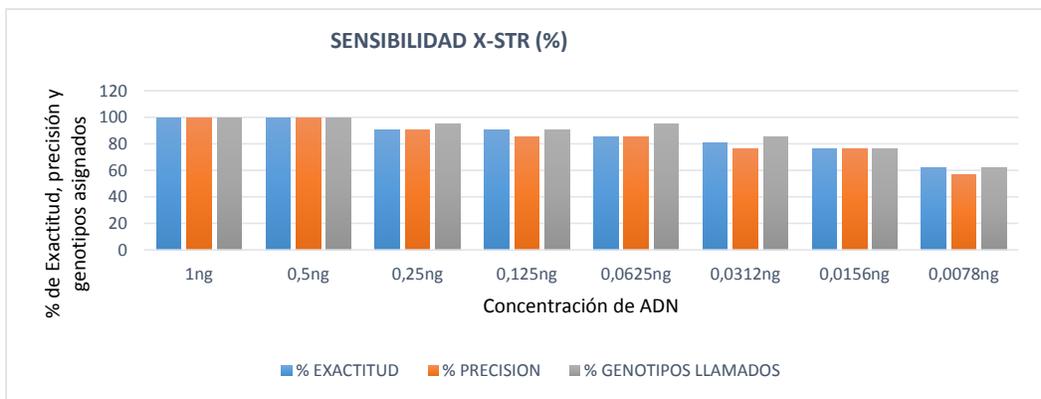
(A)



(B)



(C)



(D)

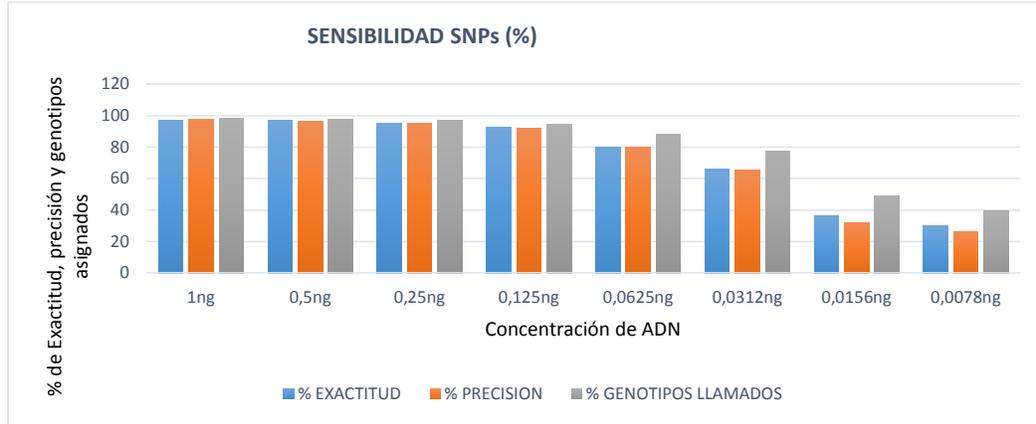


Fig 12. Ensayo de Sensibilidad, calculada en porcentajes de genotipos asignados, exactitud y precisión para 40 librerías con la línea 2800M. Porcentaje de genotipos asignados (verde), porcentajes de concordancia: Exactitud (azul) y Precisión (rojo) obtenidos para A-STRs (A), Y-STRs (B), X-STRs (C) y SNPs (D), al amplificar por triplicado diluciones seriadas entre 1 y 0.0078 ng (1000, 500, 250, 125, 62.5, 31.2, 15.6 y 7.8 pg) de la línea celular 2800M con el kit Signature Prep ForenSeq™

#### 6.1.3.1.2 Efecto del tipo de muestra en montajes de 40 librerías: ADN extraído de muestras calcificadas

Para los restos óseos en el ensayo de 40 muestras, los porcentajes de asignación de genotipos y de concordancia, disminuyen con respecto a la línea celular 2800M.

Solo para loci A-STRs, se obtiene el 100% de genotipos asignados a una concentración de 1 ng y con una concordancia del 92%. Mientras que para loci SNPs y X-STRs se alcanza el 91% y el 93% de genotipos llamados respectivamente, solo hasta concentraciones de 0.125 ng.

Por su parte los Y-STRs, mostraron el mayor nivel de disminución en asignación de genotipos, a la misma concentración de 0.125 ng, la asignación fue del 87%.

#### 6.1.3.1.3 Efecto del tipo de muestra en montajes de 88 librerías: ADN purificado línea celular 2800M y Restos óseos

La sensibilidad observada para los 2 tipos de muestras que se analizaron, disminuye, al aumentar de 40 a 88 librerías por montaje.

Mientras que en un montaje de 40 librerías se detectan todos los 27 loci A-STR de la línea celular 2800M, aun a concentraciones de ADN tan bajas como 62,5pg, en el montaje de 88 librerías, este nivel de éxito solo se logra a concentraciones de ADN superiores a 250pg, manteniendo los mismos niveles de concordancia que el ensayo de 40 librerías (>93%). Por su parte, en el montaje de 88 librerías, los Y-STRs, X-STRs y SNPs nunca se detectan en su totalidad, ni aun a la concentración de 1ng.

Se observa que, tanto la asignación de genotipos como la exactitud disminuyen para todos los marcadores, los resultados se muestran a continuación (Tabla 6):

NUMERO DE LIBRERIAS		40	88	40	88	40	88	40	88
LINEA CELULAR 2800M		A-STR (%)		Y-STR (%)		X-STR (%)		SNPs (%)	
1ng	Exactitud	98	98	100	94	100	86	97	90
	Genotipos asignados	100	100	100	94	100	86	99	91
0,5ng	Exactitud	100	98	100	94	100	90	97	84
	Genotipos asignados	100	100	100	94	100	90	98	88
0,25ng	Exactitud	100	93	99	90	90	86	95	79
	Genotipos asignados	100	100	100	92	95	86	97	83
0,125ng	Exactitud	93	76	94	75	90	86	93	60
	Genotipos asignados	100	92	100	79	90	86	95	71
0,0625ng	Exactitud	94	60	99	63	86	62	80	55
	Genotipos asignados	100	86	100	78	95	81	88	66
0,0312ng	Exactitud	83	42	92	50	81	48	66	24
	Genotipos asignados	98	56	94	53	86	57	77	29
0,0156ng	Exactitud	50	37	82	43	76	71	37	20
	Genotipos asignados	86	69	89	57	76	76	49	24
0,0078ng	Exactitud	44		56		62		30	
	Genotipos asignados	81		61		62		39	

Tabla 6. Experimento de Sensibilidad, calculada en porcentajes de genotipos asignados y de exactitud para marcadores A-STRs, Y-STRs, X-STRs y SNPs en montajes de 40 y 88 librerías.

De la tabla anterior resaltamos el comportamiento de los valores de exactitud en los dos tipos de montajes: 40 y 88 librerías; A mayor número de librerías, la exactitud disminuye notablemente hasta un 40% en STRs, y es más evidente para los marcadores tipo SNPs donde solo se alcanza un 20% de concordancia con los perfiles esperados en las concentraciones más críticas probadas. (Figura 13).

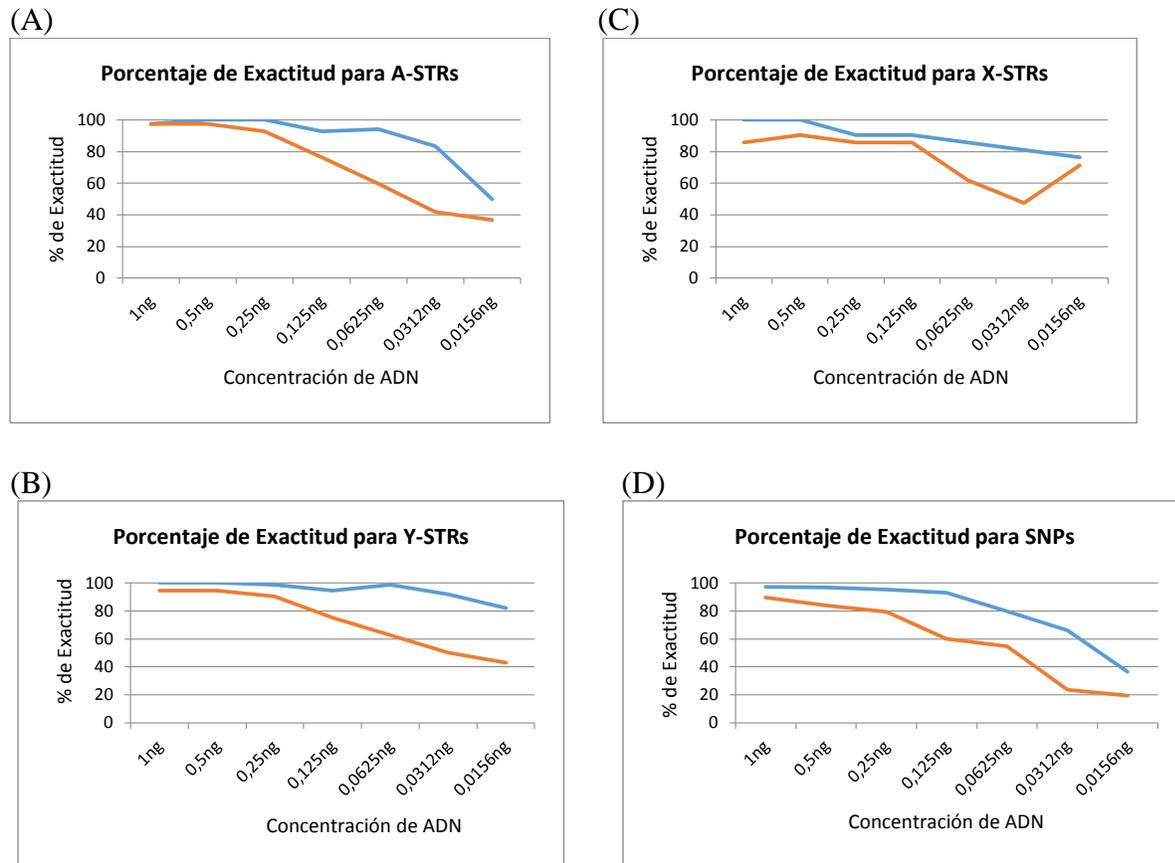


Fig 13. Ensayo de Sensibilidad, calculada en porcentaje de exactitud para 40 y 88 librerías con la línea celular 2800M Marcadores A-STRs (A), Y-STRs (B), X-STRs (C) y SNPs (D), al amplificar por triplicado diluciones seriadas entre 1 y 0.0078 ng (1000, 500, 250, 125, 62.5, 31.2, 15.6 y 7.8 pg) de la línea celular 2800M con el kit Signature Prep ForenSeq™ (Azul) para 40 librerías y (rojo) para 88 librerías.

Para los restos óseos el comportamiento en el ensayo de 88 muestras es similar al observado para la línea celular 2800M, aunque los valores tanto de asignación de genotipos como de concordancia alcanzan una diferencia hasta de 14% a favor de la línea celular para loci A-STRs y Y-STRs.

Los resultados anteriores, demuestran que la sensibilidad del kit Signature Prep ForenSeq™ en el equipo MiSeq FGx™, varía según el tipo de marcadores STRs o SNPs, el número de librerías trabajadas simultáneamente y el tipo de muestras analizadas.

Si bien es cierto que el porcentaje de genotipos obtenidos es importante, no menos lo es la exactitud que alcanzan las secuencias. En otras palabras, aunque se obtengan valores altos de llamados genotípicos, será necesario considerar el nivel de exactitud de los loci detectados. Esto puede ser un factor limitante en el análisis de muestras forenses escasas y/o degradadas, donde el nivel de exactitud puede ser bajo; no obstante, una de las ventajas de la tecnología es la posibilidad de analizar cientos de loci en un mismo proceso, lo cual permite rescatar un número considerable de loci en muestras complejas. Podrá evaluarse en estos casos el Poder de identificación de los loci sobrevivientes. De todos modos, aunque el porcentaje de genotipos asignados no sea del 100%, con este kit se obtienen muchos más marcadores útiles, que en técnicas tradicionales como la electroforesis capilar.

Adicionalmente, es posible modificar los parámetros de análisis, disminuyendo el umbral de interpretación hasta tanto no se observen artefactos como *drop in* o *stutter*, que interfieran con el análisis de los verdaderos alelos. Esto permitirá ampliar el porcentaje de llamados alélicos.

Teniendo en cuenta los resultados obtenidos en este experimento, se establecen los siguientes límites de detección:

Ensayos con máximo 40 librerías simultáneas:

- 62.5 pg para loci A-STRs y Y-STRs
- 500 pg para loci X-STRs
- 125 pg para loci SNPs

Ensayos con 64-96 librerías simultáneas, (obteniendo un porcentaje mayor al 85% de genotipos):

- 250 pg para todos los loci A-STRs, Y-STRs, X-STRs y SNPs, verificando la exactitud de los genotipos, mediante el análisis de *stutter* y *drop out* observados.

### 6.1.3.2 Umbrales analíticos

Para la validación de desarrollo, la firma Illumina basándose en estudios empíricos, usó 1,5% como Umbral analítico (AT) y 4,5% como Umbral estocástico o de interpretación (IT); los dos umbrales fueron determinados para un locus, multiplicando el valor en porcentaje del parámetro analítico, por la suma del conteo de lecturas (reads) en ese locus y en caso de baja cobertura, se asumió un mínimo de 650 reads para la determinación de los valores umbrales como se mencionó antes.

- *Umbral Analítico (AT).*

Los resultados obtenidos para los 8 controles negativos no mostraron ninguna señal de secuenciación por encima de los 10 reads; umbral establecido por defecto en la validación de desarrollo. Por lo anterior asumimos en este trabajo el mismo valor, como umbral analítico.

- *Umbral Estocástico (IT).*

El seguimiento de los sistemas heterocigotos observados en la línea 2800M a través de concentraciones decrecientes de ADN, permitió evidenciar profundidades de cobertura mínimas por debajo de las cuales, no puede asumirse una muestra como homocigota; en otras palabras, por debajo de estos valores, la probabilidad de *drop out* es alta y no debería interpretarse una señal alélica como homocigota (Gill et al., 2008). Se asumió como valor IT o umbral estocástico para STRs y SNPs, el promedio de DoC calculado de los 6 alelos correspondientes a las 3 réplicas del genotipo heterocigoto evaluado, en la dilución inmediatamente anterior a la cual se observó *drop out*.

Adicionalmente, observamos que los umbrales establecidos, varían notablemente de locus a locus en los dos tipos de marcadores; De nuevo, como se mencionó antes, la optimización de los procesos de amplificación en el desarrollo del kit presenta diferencias de locus a locus. Este hallazgo también ha sido recientemente reportado en las guías de validación NGS de la SWGDAM (S. W. G. Swgdam & Methods, 2019)

A continuación, se presenta en la Tabla 7 los valores de IT obtenidos en 23 sistemas con genotipos heterocigotos, de los 27 A-STR analizados:

MUESTRA	LOCUS	IT PROMEDIO LOCUS
2800M-0,125ng	Amelogenin	130
2800M-0,0625ng	D13S317	374
2800M-0,0625ng	D19S433	119
2800M-0,0625ng	D7S820	202
2800M-0,0625ng	PentaD	398
2800M-0,0312ng	D10S1248	237
2800M-0,0312ng	D16S539	282
2800M-0,0312ng	D17S1301	283
2800M-0,0312ng	D18S51	342
2800M-0,0312ng	D1S1656	74
2800M-0,0312ng	D21S11	225
2800M-0,0312ng	D2S1338	225
2800M-0,0312ng	D3S1358	377
2800M-0,0312ng	D6S1043	303
2800M-0,0312ng	D9S1122	99
2800M-0,0312ng	FGA	559
2800M-0,0312ng	PentaE	160
2800M-0,0312ng	TH01	820
2800M-0,0312ng	vWA	87
2800M-0,0156ng	D12S391	164
2800M-0,0156ng	D20S482	227
2800M-0,0156ng	D2S441	161
2800M-0,0156ng	D8S1179	239

Tabla 7. Umbrales de Interpretación para STRs obtenidos por locus a partir del ensayo de sensibilidad

En general, la dispersión de valores IT es muy alta entre loci (Figura 14), no obstante, proponemos definir 3 grupos de loci con valores similares de IT y asignar el mayor IT de cada grupo, aplicable a todos los marcadores asociados. De otra manera, si se usa el promedio general de 265 reads, se aumentaría el riesgo de aceptar falsos homocigotos como reales; del mismo

modo, si la decisión fuera usar el valor de Doc más alto obtenido entre todos los loci, se estaría castigando aquellos loci donde el número de reads promedio es bajo, porque excluiría muchos homocigotos reales.

El promedio de IT definido para los 3 grupos de marcadores STRs se muestra a continuación:

- 164 Reads para los loci D1S1656, vWA, D9S1122, D19S433, Amelogenina, PentaE, D2S441 y D12S391 con IT entre 74 y 164 reads.
- 398 Reads para los loci D7S820, D21S11, D2S1338, D20S482, D10S1248, D8S1179, D16S539, D17S1301, D6S1043, D18S51, D13S317, D3S1358 y PentaD con IT entre 202 y 398.
- 820 Reads para los loci FGA y TH01 con IT entre 559 y 820 reads

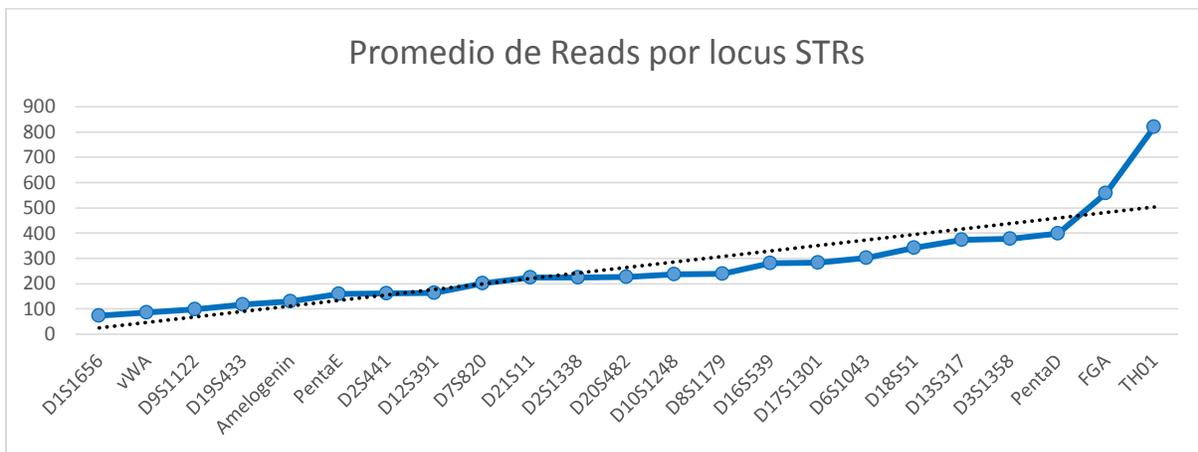


Fig 14. Promedio de IT en reads para loci STRs heterocigotos en la dilución inmediatamente anterior a la cual se observó *drop out*. Obtenidos al amplificar por triplicado diluciones seriadas entre 1 y 0.0078 ng (1000, 500, 250, 125, 62.5, 31.2, 15.6 y 7.8 pg) de la línea celular 2800M con el kit Signature Prep ForenSeq™ para 40 librerías.

Por otro lado, teniendo en cuenta los resultados obtenidos en el ensayo de sensibilidad con muestras de restos óseos, montadas con el mismo número de librerías (40) que, para la línea celular 2800M, se encontró que las concentraciones a las cuales se comienzan a observar *drop out* alélicos, son más altas que las observadas en la línea celular y por ende los umbrales de interpretación deben ser analizados con más detalle. Adicionalmente, se sugiere que se trabajen replicas para muestras complejas, con el fin de obtener reproducibilidad de los genotipos asignados.

Para el caso de los marcadores SNPs, los valores de IT obtenidos en 35 sistemas heterocigotos de los 94 SNPs analizados, se muestran en la tabla 8.

<b>MUESTRA</b>	<b>LOCUS</b>	<b>IT PROMEDIO LOCUS</b>
2800M-0,5ng	rs13182883	59
2800M-0,25ng	rs10488710	94
2800M-0,25ng	rs445251	111
2800M-0,125ng	rs2342747	54
2800M-0,125ng	rs907100	71
2800M-0,125ng	rs1498553	74
2800M-0,125ng	rs10092491	76
2800M-0,125ng	rs10773760	78
2800M-0,125ng	rs914165	91
2800M-0,125ng	rs4606077	91
2800M-0,125ng	rs763869	106
2800M-0,125ng	rs740598	121
2800M-0,125ng	rs733164	157
2800M-0,125ng	rs964681	165
2800M-0,125ng	rs576261	205
2800M-0,125ng	rs279844	227
2800M-0,125ng	rs1886510	262
2800M-0,0625ng	rs6955448	54
2800M-0,0625ng	rs891700	70
2800M-0,0625ng	rs2076848	75
2800M-0,0625ng	rs1463729	101
2800M-0,0625ng	rs873196	103
2800M-0,0625ng	rs560681	124
2800M-0,0625ng	rs987640	129
2800M-0,0625ng	rs8078417	209
2800M-0,0312ng	rs722098	105
2800M-0,0312ng	rs2399332	108
2800M-0,0312ng	rs1454361	111
2800M-0,0312ng	rs1028528	136
2800M-0,0312ng	rs1382387	140
2800M-0,0312ng	rs1058083	162
2800M-0,0312ng	rs2107612	168

Tabla 8. Umbrales de Interpretación para SNPs obtenidos por locus a partir del ensayo de sensibilidad.

La dispersión de valores IT en SNPs es muy alta entre loci (Figura 15), al igual que en los STRs, por lo cual nuevamente proponemos definir 3 grupos de loci con valores similares de IT y asignar el mayor IT de cada grupo, aplicable a todos los marcadores asociados, estos grupos son:

- 94 Reads para los loci rs2342747, rs6955448, rs13182883, rs891700, rs907100, rs1498553, rs2076848, rs10092491, rs10773760, rs914165, rs4606077, rs10488710 con IT entre 54 y 94 reads.
- 168 Reads para los loci, rs1463729, rs873196, rs722098, rs763869, rs2399332, rs1454361, rs445251, rs740598, rs560681, rs987640, rs1028528, rs1382387, rs733164, rs1058083, rs964681, rs2107612 con IT entre 101 y 168.
- 262 Reads para los loci rs576261, rs8078417, rs279844, rs1886510 con IT entre 205 y 262 reads

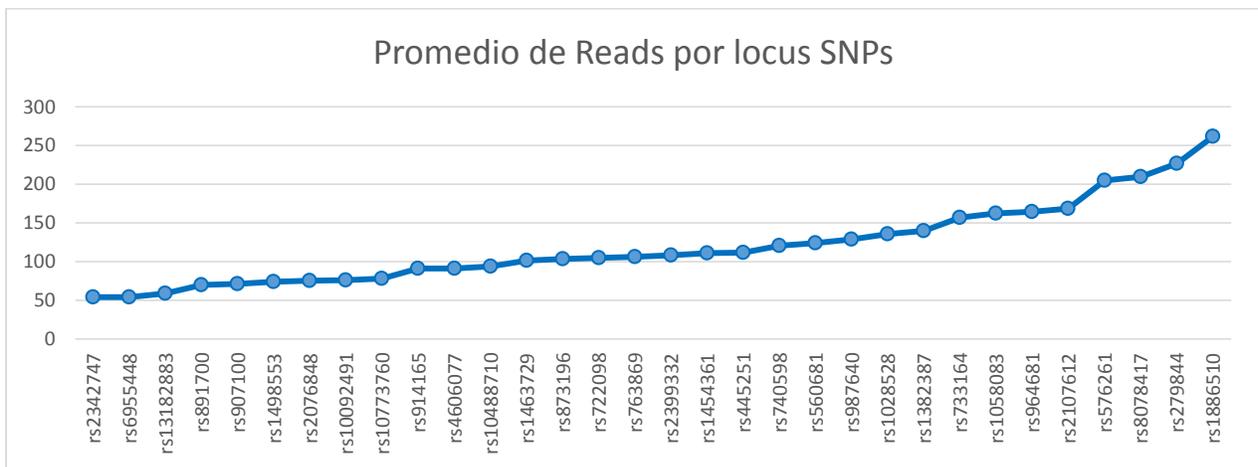


Fig 15. Promedio IT en reads para loci SNPs heterocigotos en la dilución inmediatamente anterior a la cual se observó *drop out* Obtenidos al amplificar por triplicado diluciones seriadas entre 1 y 0.0078 ng (1000, 500, 250, 125, 62.5, 31.2, 15.6 y 7.8 pg) de la línea celular 2800M con el kit Signature Prep ForenSeq™ para 40 librerías.

Los umbrales definidos anteriormente, nos indican que por debajo de estos valores existe una alta probabilidad de que un homocigoto sea en realidad un heterocigoto afectado por *drop out*. Corrobora lo anterior, el hecho de encontrar falsos homocigotos a 51 reads en la concentración de 1 ng, a 64 reads en 0.5 ng, a 72 reads en 0.25 ng y a 83 reads en 0.125ng y en 0.625ng. Lo cual indica que los umbrales establecidos (94, 168 y 262) proporcionan mayor seguridad para el análisis.

Mientras que para muestras heterocigotas, observamos que aún a concentraciones de 31 pg de ADN, podríamos encontrar heterocigotos reales incluso con 11 reads en proporciones de hasta 85:15, que no son leídos con las métricas del UAS de IT de 30 reads. Así que un buen límite de seguridad en estos casos, sería asumir alelos con reads que superen el umbral analítico de 10 reads, en la proporción indicada.

### ***6.1.3.3 Balance de Heterocigotos: Diferencia máxima del número de reads entre alelos hermanos del Heterocigotos***

Se calculó el balance entre alelos de todos los heterocigotos de la línea celular 2800M como se indicó en la metodología en las diferentes diluciones de ADN *in put* diseñadas en el experimento de sensibilidad.

Los resultados obtenidos muestran que a medida que se disminuye la concentración del ADN de entrada, también disminuye la proporción entre los alelos del heterocigoto; los valores fueron tomados del promedio de *reads* de cada alelo en las 3 réplicas realizadas en el experimento. La Figura N°16, muestra la tendencia observada entre concentraciones de ADN *in put* y el Balance de heterocigotos calculado. Esta tendencia ha sido reportada en validaciones de marcadores STRs anteriormente (Hansson, Egeland, & Gill, 2017).

Como se mencionó antes, hasta la concentración de 0.0625ng el nivel de concordancia con los genotipos esperados fue del 100% (ningún *drop out*, *drop in* o discordancias totales del genotipo esperado fueron observados); por lo tanto, se conservaron todos los heterocigotos esperados y fue posible evaluar el balance entre sus alelos. Los valores obtenidos desde 1 ng hasta 15pg en el balance, son todos superiores al 60% establecido en la validación de desarrollo del kit Forenseq – Illumina.

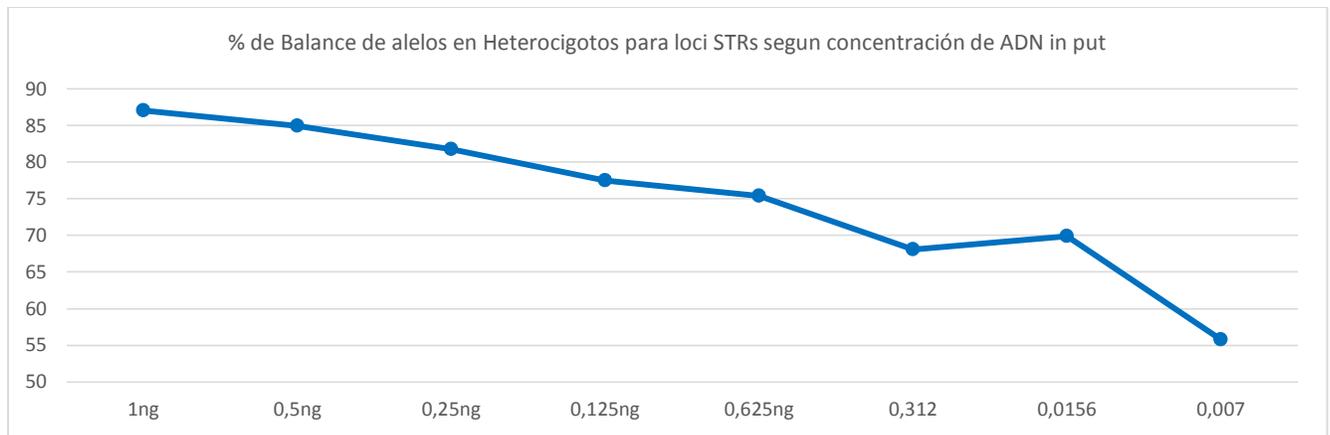


Fig 16. Porcentaje del Balance de Heterocigotos para loci STRs, a partir del ensayo de sensibilidad Al amplificar por triplicado diluciones seriadas entre 1 y 0.0078 ng (1000, 500, 250, 125, 62.5, 31.2, 15.6 y 7.8 pg) de ADN in put de la línea celular 2800M, con el kit Signature Prep ForenSeq™ para 40 librerías.

Hansson y Cols han observado este mismo comportamiento tanto en ejercicios de simulación como en ensayos de diluciones seriadas y sobre muestras forenses. En general, hay acuerdo en que la varianza en el balance de heterocigotos incrementa a medida que la cantidad de ADN disminuye lo cual puede estar relacionado con la eficiencia y la cinética de la PCR, quizás porque en concentraciones límite, se favorece la amplificación de un alelo sobre otro, como un efecto estocástico. En STR-CE es común ver este comportamiento entre alelos muy diferentes en el número de repeticiones ej. 7,15, donde el alelo de menor tamaño se favorece, si la muestra es escasa y más aún si esta degradada. En NGS puede esperarse un efecto parecido si se tiene en cuenta que hay al menos 4 amplificaciones durante todo el proceso de preparación de librerías y secuenciación. La cantidad de ADN templado y la eficiencia de la extracción del ADN, pueden también influir, de hecho la PCR directa parece disminuir notablemente el imbalance (Hansson et al., 2017).

Por otro lado, los SNPs presentan valores muy similares a los STRs, en la proporción de los alelos de heterocigotos, hasta la dilución de 0.625 ng y de igual forma, muestran una tendencia a la disminución de estos porcentajes en función de la concentración de ADN (Figura 17).

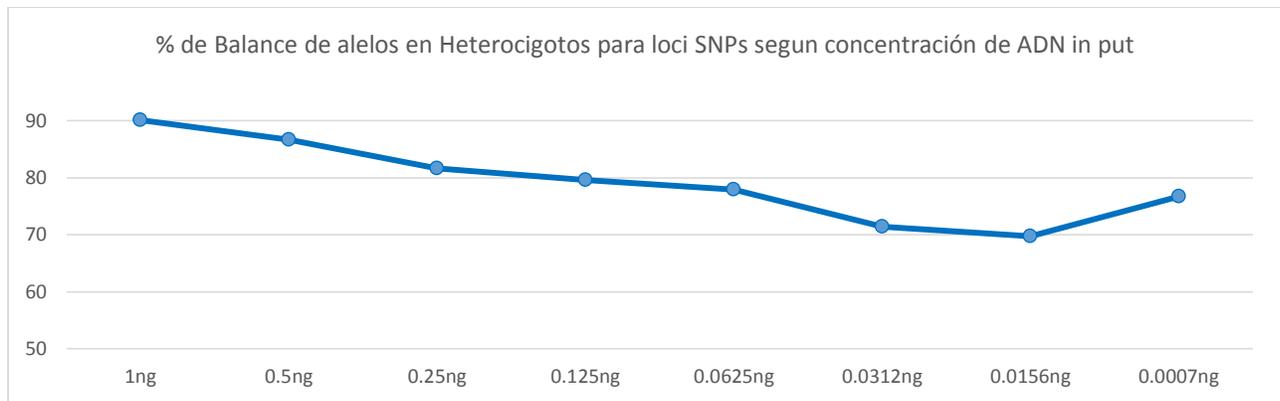


Fig 17. Porcentaje del Balance de Heterocigotos para loci SNPs, a partir del ensayo de sensibilidad Al amplificar por triplicado diluciones seriadas entre 1 y 0.0078 ng (1000, 500, 250, 125, 62.5, 31.2, 15.6 y 7.8 pg) de ADN in put de la línea celular 2800M, con el kit Signature Prep ForenSeq™ para 40 librerías.

Sin embargo, se observa que el imbalance compromete más a los loci STR que a los SNPs, principalmente en las concentraciones de ADN templado más bajas, donde prácticamente, el porcentaje de balance para SNPs no baja del 69% y si llega hasta el 55% en STRs (Figuras 16 y 17).

El resultado anterior, permite tener un límite de seguridad adicional para el análisis de marcadores SNPs en muestras heterocigotas, en donde podría interpretarse un genotipo heterocigoto como verdadero si sus dos alelos superan cada uno al menos los 15 reads (podría ser solo superior al AT, como se dijo antes). Así, en análisis de mezclas se puede generar mayor confianza para el analista, de modo que desbalances hasta del 70% entre los dos alelos, pueden considerarse aún como proveniente de un solo aportante a la muestra, valores menores podrían relacionarse con alelos de un segundo aportante.

En conclusión, los resultados de esta validación permiten recomendar que el balance de heterocigotos se evalúe sobre muestras forenses, en función de la concentración del ADN input, teniendo en cuenta que por debajo de 0.625ng, empiezan a presentarse fenómenos de *drop out* que hace más difícil la valoración de imbalances de heterocigotos en los 2 tipos de marcadores, como lo indicó el experimento de sensibilidad.

Finalmente, por debajo de los valores calculados de balance de heterocigotos, en una muestra de ADN de concentración conocida, puede sospecharse la presencia de más de un aportante al perfil genético STR/SNPs.

#### 6.1.4 Experimentos de LCN (Low Copy Number)

Con el ánimo de cumplir uno de nuestros objetivos, y poder determinar si la tecnología utilizada en este estudio, puede ser usada como una herramienta para rescatar muestras complejas, con altos niveles de degradación y escasa cantidad del ADN recuperado, se probó una de las técnicas de LCN, en la modificación de ciclos de PCR, para observar si es posible mejorar los resultados obtenidos en la validados de nuestro ensayo de sensibilidad cuando se usa la técnica como lo sugiere el fabricante.

Se tomaron las diluciones realizadas en el ensayo de sensibilidad con la línea celular 2800M por triplicado y se analizaron en montajes de 40 muestras, aumentando el número de ciclos en la PCR1 targeting, como se describió en la metodología:

- Tratamiento 2: Dos (2) ciclos adicionales en cada termociclaje
- Tratamiento 3: Un (1) ciclo adicional en el primer ciclaje y manteniendo el segundo igual

Las métricas de calidad, alcanzadas en cada ensayo (2 tratamientos diferentes), se muestran a continuación:

Tratamiento utilizado en la PCR 1	2	3
Densidad de clúster (k/mm <sup>2</sup> )	654	703
% de clusters que pasaron el filtro	97.60	96.74
Phasing %	0.202	0.261
Pre-Phasing %	0.072	0.062

Tabla 9. Métricas de calidad obtenidas para ensayos de Low Copy Number (LCN)

Debido a las bajas concentraciones de ADN que se analizaron, la densidad de cluster obtenida en los dos casos, no superó los 1000 k/mm<sup>2</sup>, aun así, el resultado cae dentro del rango exigido por el fabricante, lo cual nos permitió analizar los resultados adecuadamente. Por otro lado, el porcentaje de *Phasing* observado para el tratamiento 3, superó en 0.011 el límite exigido. Un aumento en estos valores podría permitir la incorporación de errores en la secuencia del ADN y

por lo tanto afectar la concordancia en el genotipo, pero en nuestro caso el valor superado es mínimo, lo que no ocasionó ningún efecto.

Línea celular 2800M	CICLAJE NORMAL				TRATAMIENTO 2				TRATAMIENTO 3			
	A-STRs (%)	Y-STRs (%)	X-STRs (%)	SNPs (%)	A-STRs (%)	Y-STRs (%)	X-STRs (%)	SNPs (%)	A-STRs (%)	Y-STRs (%)	X-STRs (%)	SNPs (%)
Exactitud	98	100	100	97	96	98	100	95	100	92	86	95
Genotipos asignados 1ng	100	100	100	99	100	100	100	97	100	92	86	98
Exactitud	93	94	90	93	89	97	86	91	95	92	86	77
Genotipos asignados 0,125ng	100	100	90	95	100	100	100	93	99	97	90	81
Exactitud	94	99	86	80	80	92	76	85	94	93	86	83
Genotipos asignados 0,0625ng	100	100	95	88	100	100	95	91	100	99	86	89

Tabla 10. Ensayo de Low Copy Number (LCN), calculado en porcentajes de exactitud y de genotipos recuperados entre experimentos en condiciones del fabricante (ciclaje normal) vs Tratamiento 2 y Tratamiento 3, obtenidos para marcadores A-STRs, Y-STRs, X-STRs y SNPs

La tabla 10 resume los resultados obtenidos para los tratamientos 2 y 3, comparados con los obtenidos en el ciclaje sugerido por el fabricante (ciclaje normal). Se observa que a concentraciones entre 1 ng y 62 pg, aunque no se presentan grandes variaciones en el número de genotipos, en comparación con el ciclaje normal, si se empieza a comprometer la concordancia en la mayoría de los tipos de marcadores, con mayor evidencia en el tratamiento 3.

Nos centramos en el análisis de bajas concentraciones (15 y 7 pg), dado que nuestro interés radica en el resultado de muestras con escasa cantidad de ADN (Figura 18 A y B).

Para el tratamiento 2, en una concentración de ADN de 15 pg, se observó un porcentaje de genotipos y de exactitud mayor, que en el ciclaje normal; Sin embargo, al bajar a 7 pg, aunque los porcentajes de recuperación de genotipos son mayores en todos los marcadores, la exactitud decrece notablemente para marcadores A-STRs y Y-STRs y se mantiene sin variación para marcadores X-STRs y SNPs.

Los resultados obtenidos cuando se prueba el tratamiento 3, no superan al anterior; a 15 pg el porcentaje de genotipos obtenidos es mayor que en el ciclaje normal, excepto para Y-STRs y la exactitud solo se incrementa para marcadores A-STRs y SNPs. Pero a 7 pg, el porcentaje de recuperación de genotipos aumenta únicamente para marcadores Y-STRs y SNP y la exactitud solo para SNPs.

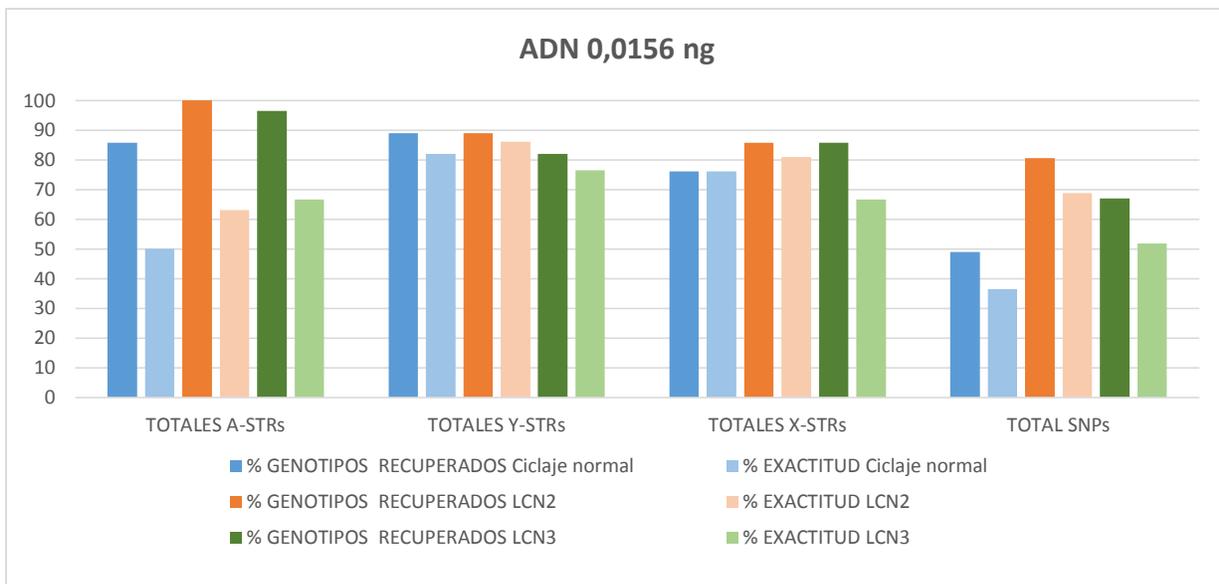
Comparando los resultados entre los 3 ciclajes (ciclaje normal, tratamiento 2 y 3), a bajas concentraciones de ADN, se observa un mejor desempeño de la técnica, con el tratamiento 2. Sin embargo, solo hasta una concentración de 15 pg, es posible incrementar el porcentaje de genotipos recuperados y concordantes.

En el trabajo forense con muestras óseas, en los casos donde se recupere ADN en concentraciones cercanas a 15 pg podría ser útil usar el tratamiento 2, siempre y cuando se tenga especial cuidado en el análisis de *stutter* con marcadores STRs, por lo cual podría ser más fácil el análisis y la interpretación de los marcadores tipo SNPs.

Más aún a 7 pg, con el tratamiento 2 se obtiene un aumento en el porcentaje de genotipos para los marcadores SNPs del 27% con relación al ciclaje normal y del 3 % en la exactitud. Para el tratamiento 3, aunque solo se incrementan en un 10 % los genotipos, la exactitud alcanza un 6% más que en el ciclaje normal. Esto resultados podrían ser prometedores para los casos con ADN degradados y de escasa cantidad.

Por otro lado, no se recomienda usar estas modificaciones de LCN, en estudios en donde se procesen manchas, o se sospeche de mezcla, dado que en la interpretación de los genotipos STRs, se podrían confundir alelos verdaderos con la presencia de *drop in* o *stutter*.

(A)



(B)

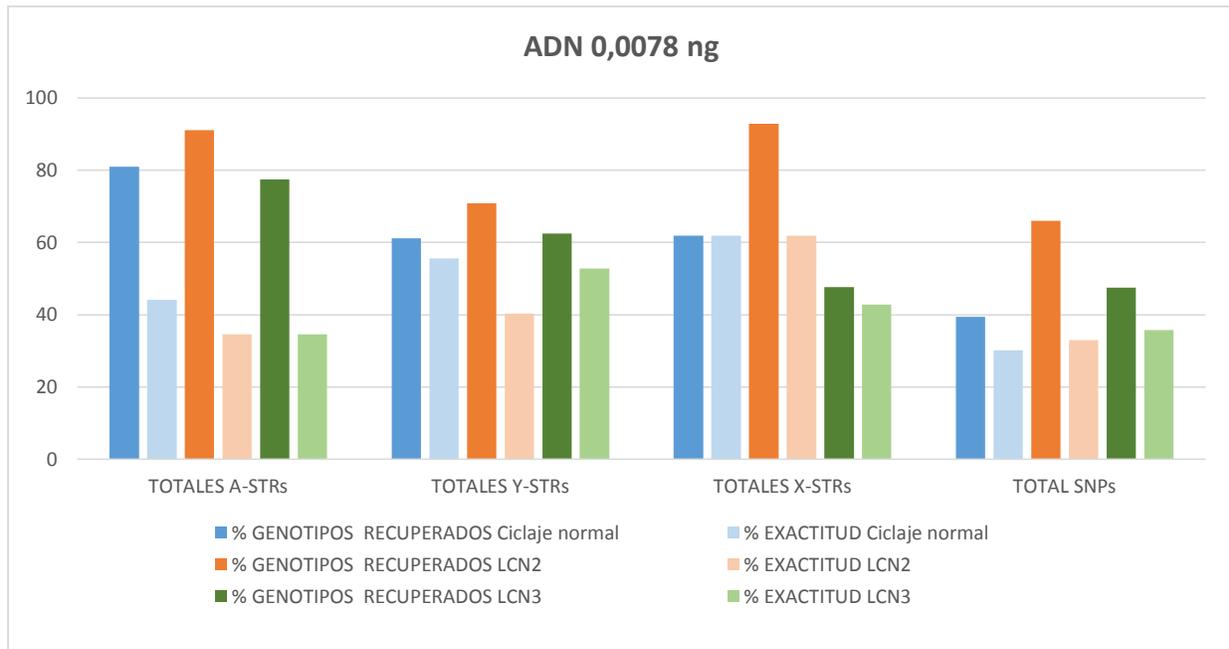


Fig 18. Ensayo de Low Copy Number (LCN)

Se muestra comparativamente el porcentaje de exactitud y de genotipos recuperados entre experimentos en condiciones del fabricante (ciclaje normal) vs Tratamiento 2 y Tratamiento 3, (Colores azul, naranja y verde respectivamente) obtenidos para marcadores A-STRs, Y-STRs, X-STRs y SNPs, al amplificar por triplicado diluciones seriadas de la línea celular 2800M con el kit Signature Prep ForenSeq™ (A) 15,6 pg y (B) 7,8 pg

### 6.1.5 Experimentos de variación de librerías para secuencia

Siguiendo con el objetivo de determinar si NGS puede ser usada como herramienta para rescatar muestras complejas, se realizaron ensayos variando el volumen de cada librería o del *pool* de librerías final que van a secuenciarse en un mismo montaje y se determinó con cuál de los ensayos se obtiene el mayor número de marcadores con los genotipos esperados.

Las muestras se amplificaron en las mismas condiciones sugeridas por el fabricante y al momento de realizar la secuenciación se hicieron los siguientes ensayos para crear el pool de librerías y adicionarlo al cartucho de reactivos:

- Ensayo 1: 5 ul de cada librería para un volumen de pool final de 8.5 ul
- Ensayo 2: 5 ul de cada librería para un volumen de pool final de 10 ul
- Ensayo 3: 8.5 ul de cada librería para un volumen de pool final de 7.5 ul
- Ensayo 4: 10 ul de cada librería para un volumen de pool final de 7.5 ul

Para los ensayos 1 y 2, se tomaron todas las diluciones realizadas en el ensayo de sensibilidad con la línea celular 2800M por triplicado (1ng - 7.8 pg). Para los ensayos 3 y 4 se tomaron solo las diluciones entre 62.5 pg y 7.8 pg, incluyendo control positivo de 1ng/ul y control negativo para los 4 ensayos. Lo anterior para analizar con mayor énfasis el comportamiento de los ensayos en muestras con ADN de concentraciones escasas.

Línea celular 2800M	POOL FINAL DE 7.5 ul (Indicaciones fabricante)				E1 POOL 8.5				E2 POOL 10			
	A-STRs (%)	Y-STRs (%)	X-STRs (%)	SNPs (%)	A-STRs (%)	Y-STRs (%)	X-STRs (%)	SNPs (%)	A-STRs (%)	Y-STRs (%)	X-STRs (%)	SNPs (%)
Exactitud	98	100	100	97	100	100	100	95	99	100	100	98
Genotipos asignados	100	100	100	99	100	100	100	96	100	100	100	99
Exactitud	93	94	90	93	93	99	95	84	57	96	81	89
Genotipos asignados	100	100	90	95	100	100	95	89	100	100	90	91
Exactitud	94	99	86	80	93	93	81	71	51	89	90	70
Genotipos asignados	100	100	95	88	100	96	86	80	100	97	90	77

Tabla 11. Experimentos de variación de librerías para secuencia, calculada en porcentajes de genotipos asignados y de exactitud para marcadores A-STRs, Y-STRs, X-STRs y SNPs en condiciones del fabricante (pool final de 7.5ul) vs Ensayo 1 y Ensayo 2.

Se observa que, al aumentar el volumen del pool final, tanto el porcentaje de genotipos obtenidos, como su exactitud varían a medida que se va disminuyendo la concentración, reduciendo hasta al 51% la exactitud de los A-STRs y hasta al 70% la de SNPs, por lo tanto, no es conveniente modificar las condiciones de volumen del pool, cuando se tienen concentraciones de ADN entre 1 ng y 62.5 pg (Tabla 11)

El análisis de los cuatro ensayos a concentraciones más bajas, nos muestra los siguientes resultados (Figura 19):

El Ensayo 1, con un pool final de 8.5 ul no mejora ninguna de las dos variables de estudio, ni el porcentaje de genotipos recuperados, ni la concordancia. De los 3 ensayos restantes, el que muestra mejores resultados es el Ensayo 2 con un volumen de pool final de 10 ul; para este ensayo el porcentaje de genotipos obtenidos se aumenta en todos los tipos de marcadores en las 2 concentraciones más bajas (15 y 7 pg), excepto para A-STRs a 7 pg. La concordancia se mantiene estable o se aumenta para 15 pg, mientras a 7 pg no se observa un incremento significativo.

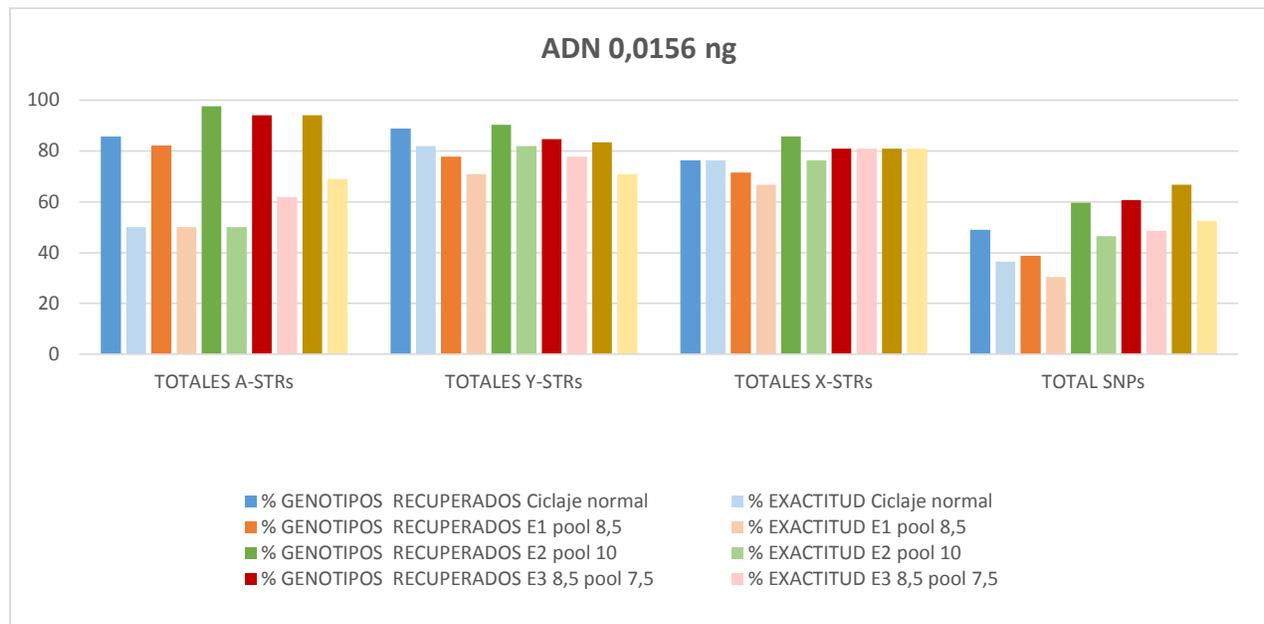
Para los Ensayos (3 y 4) en general, los resultados no son mejores que los anteriores. No obstante, a concentraciones de ADN entre 7 y 15 pg, el Ensayo 4 muestra un incremento, tanto en el porcentaje de genotipos SNPs recuperados, como en su concordancia.

Los marcadores Y-STRs y X-STRs muestran los más altos porcentajes de exactitud en las dos concentraciones probadas (>70% y > 60% respectivamente) y principalmente en el Ensayo 2, si se compara con los A-STRs y los SNPs que resultan más comprometidos en los 4 ensayos. Lo anterior podría sugerir un mejor desarrollo en los *primers* usados para la amplificación de los marcadores Y-STRs y X-STRs.

El efecto observado en estos ensayos es similar a lo obtenido en los experimentos de LCN. A concentraciones cercanas a los 15 pg, se recomienda usar el Ensayo 2 para el trabajo forense, teniendo cuidado en el análisis de *stutter* y de *drop in* con marcadores STRs; de igual forma, si se tiene un especial interés en los marcadores tipo SNPs también es posible usar el ensayo 4.

Sin embargo, en concentraciones de ADN más bajas el análisis de los genotipos puede ser muy dispendioso y poco útil debido a la baja concordancia obtenida en todos los ensayos.

(A)



(B)

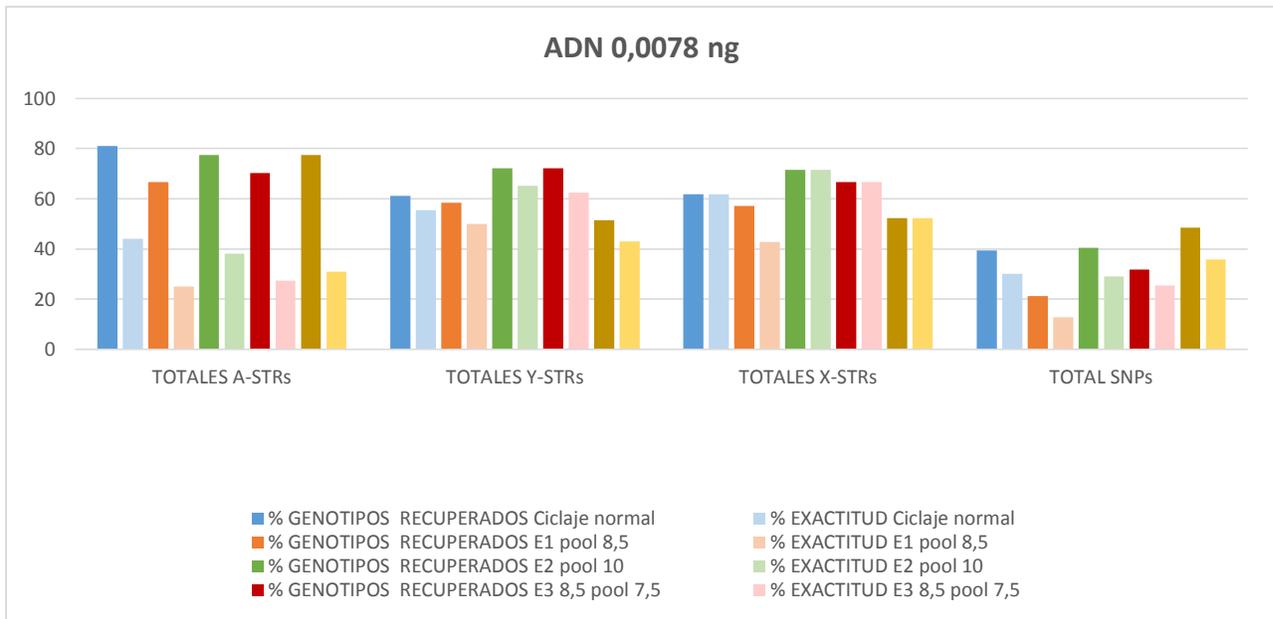


Fig 19. Experimentos de variación del volumen de librerías para secuencia.

Se muestra comparativamente el porcentaje de exactitud y de genotipos recuperados entre experimentos con variación del volumen de librerías (ensayos 2 al 4) y las condiciones del fabricante con pool final de 7.5 ul, obtenidos para A-STRs, Y-STRs, X-STRs y SNPs, al amplificar por triplicado diluciones seriadas de la línea celular 2800M con el kit Signature Prep ForenSeq™ (A) 15,6 pg y (B) 7,8 pg. Condiciones del fabricante pool 7.5 ul (azul), Ensayo 1 (naranja), Ensayo 2 (verde), Ensayo 3 (rojo) y Ensayo 4 (amarillo).

### 6.1.6 Efecto de la combinación de los experimentos de incremento de ciclajes y aumento del pool de librerías

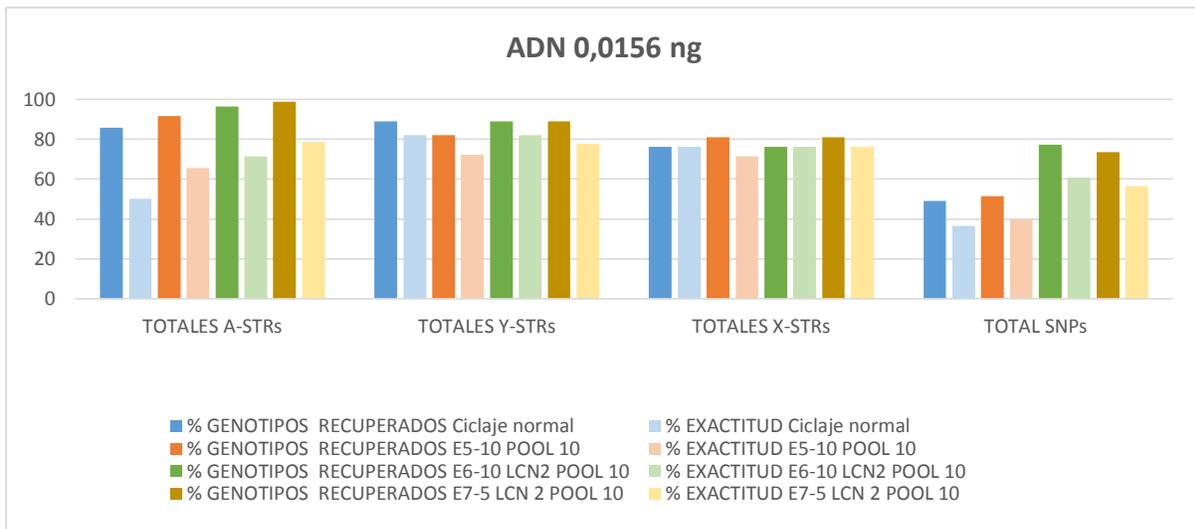
Adicionalmente, se quiso probar si a concentraciones de ADN más bajas, el resultado en número de genotipos asignados concordantes, podría mejorar, combinado 2 o más de las estrategias analizadas en los ensayos anteriores que mejores resultados mostraron. Las combinaciones ensayadas, fueron las siguientes:

- Ensayo 5: 10 ul de cada librería para un volumen de pool final de 10 ul
- Ensayo 6: LCN2 10 ul de cada librería para un volumen de pool final de 10 ul
- Ensayo 7: LCN2 5 ul de cada librería para un volumen de pool final de 10 ul

Analizando la información obtenida, se observa un mayor porcentaje de genotipos STRs recuperados y concordantes en los ensayos 6 y 7, para concentraciones entre 7 y 15 pg,

comparados con el ciclaje y pool recomendados por el fabricante; solo para marcadores X-STRs el porcentaje es mayor en el ensayo 7. Mientras que para marcadores tipo SNPs se muestran mejores resultados con el ensayo 6. Sin embargo, a 7 pg la concordancia de los genotipos en el ensayo 7 para marcadores A-STRs no supera el 44% y para SNPs en el ensayo 6 no supera el 46% (Figura 20).

(A)



(B)

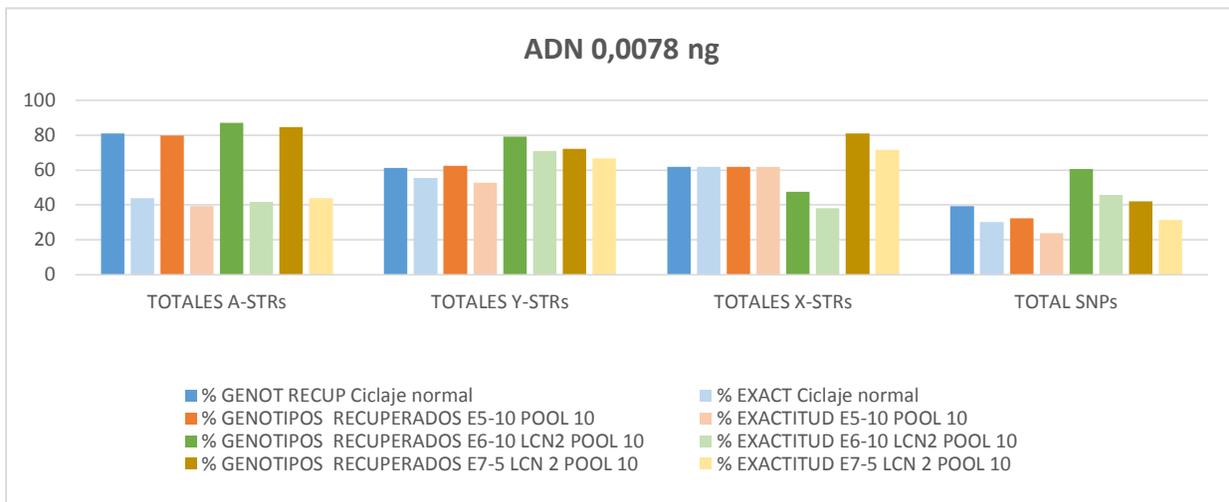


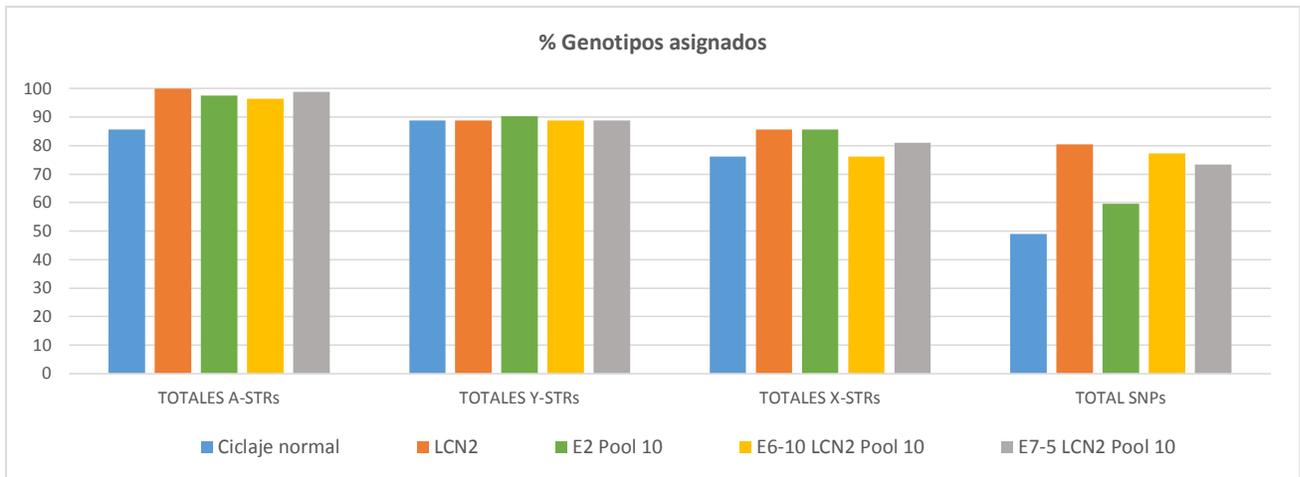
Fig 20. Efecto de la combinación de los experimentos de incremento de ciclajes y aumento del pool de librerías. Se muestra comparativamente el porcentaje de genotipos recuperados y de exactitud, entre experimentos con los mejores resultados obtenidos para variación del número de ciclos en la PCR1 (LCN) y volumen de librerías (ensayos 5 al 7) comparados con los de ciclaje normal, obtenidos para marcadores A-STRs, Y-STRs, X-STRs y SNPs, al amplificar por triplicado diluciones seriadas de la línea celular 2800M con el kit Signature Prep ForenSeq™ (A) 15,6 pg y (B) 7,8 pg. Ciclaje normal (azul), ensayo 5 (naranja), ensayo 6 (verde), ensayo 7 (amarillo).

Más aun, en concentraciones menores a los 7 pg (1.4 pg y 0.7 pg), no fue posible obtener genotipos concordantes, los porcentajes se encuentran en un rango del 2 al 28 % (Anexo 6).

Nuevamente, si comparamos todos los ensayos realizados, ya sea variando el número de ciclos en la PCR1, el número de pool de librerías, o combinando los dos, observamos que los resultados a concentraciones de ADN alrededor de los 15 pg pueden ser favorecidos con los ensayos 6 o 7, pero por debajo de estas concentraciones no se recomienda usar la técnica de NGS, dada la baja concordancia obtenida.

En conclusión, el ensayo con mejores resultados en porcentaje de genotipos obtenidos y concordantes fue el de LCN tratamiento 2, aunque el ensayo 7 (5 LCN2 pool 10) mostró una concordancia mayor para marcadores A-STRs y muestra buenos resultados también en el porcentaje de genotipos obtenidos. Los resultados anteriores nos permiten sugerir estos 2 ensayos, cuando se trabaje con ADN de concentraciones de hasta 15 pg, siempre realizando un análisis exhaustivo de *stutter*, *drop in* y *drop out* y en todo caso logrando siempre la reproducibilidad de los genotipos obtenidos (Figura 21).

(A)



(B)

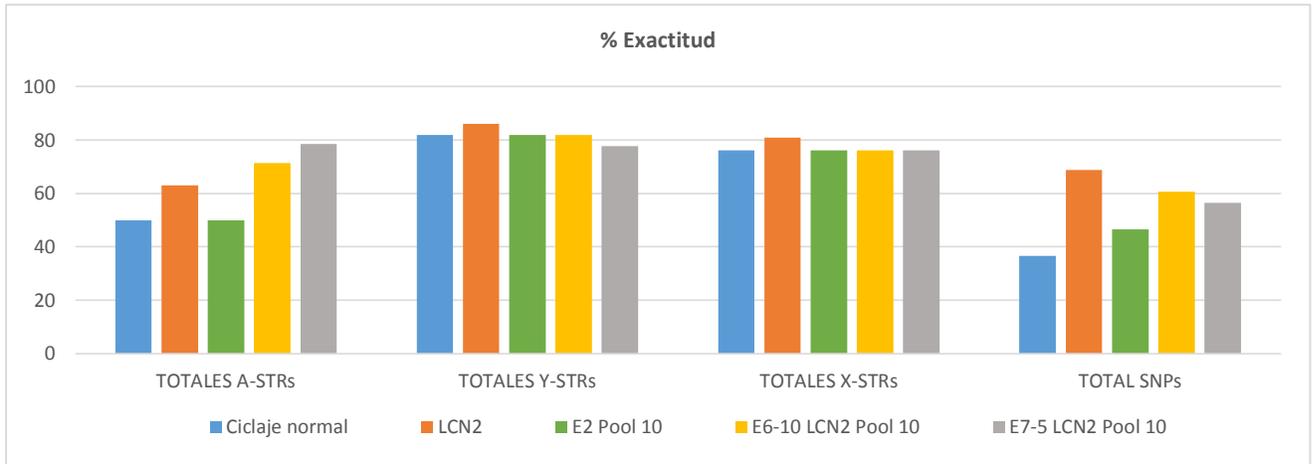


Fig 21. Ensayos con los mejores resultados variando número de ciclos (LCN), volumen de librerías para secuencia, o combinación de los dos. Se muestra comparativamente el porcentaje de genotipos asignados (A) y de exactitud (B), entre los ensayos con los mejores resultados obtenidos, cuando se varía el número de ciclos en la PCR1 (LCN), o el volumen de librerías, o combinando los 2, comparados con el ciclaje normal; obtenidos para A-STRs, Y-STRs, X-STRs y SNPs, al amplificar por triplicado diluciones seriadas de la línea celular 2800M con el kit Signature Prep ForenSeq™

## 6.2 Análisis comparativo entre tecnología de nueva generación y electroforesis capilar en casos complejos

### 6.2.1 Complejidad asociada a la calidad de la muestra

Teniendo en cuenta, el análisis de los resultados obtenidos en los experimentos anteriores, elegimos los mejores ensayos para aplicarlos a las muestras seleccionadas, en montajes de máximo 40 muestras. Así, optamos por los protocolos de LCN tratamiento 2 y el usado en el ensayo 7 (5 LCN2 pool 10), para probar si obteníamos algunos marcadores STRs o SNPs reproducibles.

Como se puede observar en la tabla 12, el valor de cuantificación obtenido para cada una de las muestras, estuvo en el orden de los 0.2 a 15 pg. Aunque estos valores no sobrepasaron la concentración mínima sugerida para análisis (15 pg), según nuestra validación, decidimos aun así observar los resultados alcanzados con los 2 protocolos elegidos.

CUANTIFICACION pg/ul		
Muestras No incineradas	Muestras parcialmente incineradas	Muestras totalmente incineradas
7,1	1,2	0,4
0,2	2	0,48
0,4	0,16	15

Tabla 12. Cuantificación obtenida para muestras anónimas de ADN, extraídas de restos óseos (RO) de casos de identificación, con resultados negativos por CE, según las variables seleccionadas: Muestras óseas antiguas (30-40 años) no incineradas, parcialmente incineradas y totalmente incineradas

Solo para 3 muestras se evidenciaron algunos marcadores; 9 STRs con profundidades entre 32 y 205 reads y 6 SNPs todos homocigotos entre 35 y 81 reads, ninguna de ellas superó los umbrales de interpretación establecidos y en ninguno de los 2 ensayos realizados logramos datos reproducibles, por lo cual no recomendamos esta tecnología para muestras con concentraciones de ADN por debajo de los 15 pg.

### 6.2.2 Complejidad asociada a pedigrís poco informativos, bajo variables de número y tipo de familiares para cotejo

Como se mencionó en la metodología, utilizando el software FAMILIAS v: 3.2.2, se calculó el valor de LR con 94 SNPs incluidos en el kit Signature Prep ForenSeq™, en 5 escenarios simulados de diferentes parentescos y el valor obtenido se comparó con el hallado para 25 STRs.

Los resultados obtenidos se muestran en la tabla 13:

	CASO 1		CASO 2		CASO 3		CASO 4		CASO 5	
	LR – STRs	LR – SNPs	LR – STRs	LR – SNPs	LR – STRs	LR – SNPs	LR – STRs	LR – SNPs	LR – STRs	LR – SNPs
Hermanos completos	$21 \times 10^{18}$	$101 \times 10^6$	$9 \times 10^3$	$61 \times 10^6$	1,36	$82 \times 10^3$	$3 \times 10^9$	$247 \times 10^3$	$25 \times 10^3$	$23 \times 10^3$
Medio hermanos	220,38	6,76	754,60	481,67	$86 \times 10^3$	389,47	59,04	0,50	0,25	76,19
Padre e hijo y madre ausente	$87 \times 10^6$	$18 \times 10^6$	$1 \times 10^9$	$25 \times 10^6$	$26 \times 10^6$	$4 \times 10^6$	$44 \times 10^3$	$30 \times 10^6$	$97 \times 10^6$	$505 \times 10^3$
Abuelo paterno y nieto con madre ausente	16,38	5,93	0,67	1,11	335,19	76,34	757,78	42,39		
Abuelo paterno, nieto y madre	$1 \times 10^3$	120,54	557,46	833,24	$7 \times 10^3$	186,08				

Tabla 13. Resultado del cálculo del valor de LR, realizado con el software Familias, en casos de parentescos complejos, obtenido para 25 marcadores A-STRs y 92 marcadores SNPs incluidos en el kit Signature Prep ForenSeq™.

Como se observa en la tabla anterior, en casos de hermanos completos solo en dos ejemplos el valor del LR para SNPs aumenta substancialmente, similar a lo que ocurre con los casos de padre e hijo y madre ausente, en los que solo en un ejemplo, el valor del LR calculado para SNPs es mayor al de A-STRs.

Por el contrario, los valores obtenidos tanto para marcadores A-STRs como para SNPs, en casos de medios hermanos y abuelo con o sin madre, son bajos y semejantes entre los dos tipos de marcadores.

Los resultados anteriores podrían explicarse por el efecto de la distribución de las frecuencias alélicas de los dos tipos de marcadores: En los STRs, por su naturaleza multialélica y su distribución normal en la población, es común observar frecuencias bajas en varios alelos de un locus, en algunos casos alcanzan la frecuencia alélica mínima FAM (0,01), el rango de valores LR es muy amplio sobre todo en casos de hermanos (1,3 hasta  $1 \times 10^{18}$ ); en su lugar, las frecuencias de marcadores SNPs son muy estables; en el muestreo poblacional que realizamos se observó que la mayoría de los loci SNPs tiene frecuencias cercanas a 0,5, solo un 11% muestran frecuencias extremas ( $> 0,8$  /  $< 0,2$ ), y la FAM fue de 0.112, por lo cual puede concluirse que el efecto sobre el LR de los SNP, depende más del número de loci que se analicen.

Por otra parte, en relación con la independencia entre loci STRs y SNPs y apoyándonos en el artículo de Sánchez 2006 y otros, en donde reportan que 52 de los marcadores SNPs usados en el kit de Illumina se encuentran mapeados en regiones distales de los brazos p y q de cada autosoma, a una distancia mínima de 100 kb entre ellos y que no presentan ninguna asociación probable con los loci STR más comúnmente utilizados en el análisis forense; podemos asumir que los dos tipos de marcadores (SNPs y STRs) se segregan de manera independiente y es válido aplicar entonces la regla del producto para calcular la probabilidad de ocurrencia simultánea de los dos genotipos STR y SNP en un mismo individuo (*match probabilities*).

Adicionalmente, el mismo autor demostró, mediante pruebas de desequilibrio de ligamiento (LD) para pares de SNPs en el mismo cromosoma, que no había desviación significativa, es decir no se encontró desequilibrio de ligamiento entre ninguna de las parejas de SNPs (Pereira et al., 2008; Sanchez et al., 2006). Este resultado se corroboró en este estudio con los 94 SNPs incluidos en el kit de Illumina y se discute más adelante.

Así las cosas, asumiendo un valor *a priori* de 0,5 usado comúnmente en pruebas de filiación, realizamos cálculos combinados con ambos marcadores, obteniendo resultados con probabilidades de parentesco mayores a 99,99% (tabla 14). Se demostró que en 11 de los 16 casos que no alcanzaban la probabilidad de parentesco usando solo marcadores STRs, se superó el valor, logrando un 69% de casos que podrían convertirse en concluyentes (valores en color rojo).

	COMBINADO EJ 1	COMBINADO EJ 2	COMBINADO EJ 3	COMBINADO EJ 4	COMBINADO EJ 5
Caso 1: 2 hermanos completos	>99,99999	>99,99999	99,998	>99,99999	>99,99999
Caso 2: 2 medios hermanos	99,866	99,999	>99,99999	93,655	90,498
Caso 3: Padre e hijo con madre ausente	>99,99999	>99,99999	>99,99999	>99,99999	>99,99999
Caso 4: Un abuelo paterno y nieto con madre ausente	97,983	27,106	99,992	99,994	
Caso 5: Un abuelo paterno, nieto y madre	99,999	>99,99999	>99,99999		

Tabla 14. Resultado de la probabilidad de parentesco, en casos complejos, combinando los valores de LR obtenidos para 25 marcadores A-STRs y 92 marcadores SNPs incluidos en el kit Signature Prep ForenSeq™.

Estos datos nos demuestran la utilidad de combinar los marcadores STRs y SNPs para casos con familiares deficitarios, en pruebas de paternidad y de gran importancia para nuestro país, en casos donde aún existen cuerpos que permanecen sin identificar y que debido al largo tiempo del conflicto armado, sus padres y familiares cercanos ya han fallecido. Por el amplio número de loci que pueden analizarse con la tecnología NGS y la posibilidad de combinar los marcadores STR y SNP en un mismo kit, se propone que todos los casos complejos de parentesco donde los pedigrí sean deficitarios, sean abordados desde un comienzo con esta tecnología, lo cual reduciría tiempos y recursos de forma significativa.

### **6.3 Estudios poblacionales de 3 regiones colombianas para 89 SNPs de identificación, del kit forenseq illumina**

Los datos obtenidos para este estudio cumplieron con las métricas de calidad (QC) establecidas en la validación de desarrollo e incorporadas por defecto en el software UAS y los genotipos asignados fueron verificados, teniendo en cuenta los umbrales establecidos en esta validación como se mencionó anteriormente.

En la tabla 15, se encuentran resumidas las frecuencias alélicas observadas para un total de 200 muestras estudiadas, distribuidas en 3 poblaciones: Andes central (n=101), Pacifico (n=50) y Orinoquia (n=49), analizadas como se describió en la metodología. No se tuvieron en cuenta para este estudio, loci con más del 20% de datos perdidos, por lo que, para la región de la Orinoquía se eliminaron 3 loci (rs1294331, rs1355366, rs13218440) y 5 (rs1294331, rs1355366, rs13218440, rs2920816, rs1736442) para las regiones Pacifica y Andina Central.

#### **Frecuencias alélicas**

El promedio de frecuencias alélicas obtenido para las 3 regiones fue de 0,499 y la frecuencia alélica mínima observada fue de 0,112 para el alelo T en el locus rs1357617 en la región Pacifica. El 64% de los marcadores SNPs mostraron frecuencias alélicas entre 0,4 y 0,69, dada su homogeneidad pueden ser muy informativos como herramientas de identificación forense para las 3 regiones. Sin embargo, las frecuencias alélicas para 10 de los SNPs analizados, fueron superiores a 0,8 en alguna de las regiones, por lo que no aportarían mucho al incremento de los valores de PD, pero afortunadamente resultan ser la minoría y en general, los valores alcanzados de PD y CE son altamente discriminantes (rs93828, rs2056277, rs1360288, rs1028528, rs1357617, rs873196, rs2107612, rs1015250, rs719366 y rs876724)

N°	SNPs	ALELOS	ANDINA CENTRAL		ORINOQUIA		PACIFICO		N°	SNPs	ALELOS	ANDINA CENTRAL		ORINOQUIA		PACIFICO	
			ALELO 1	ALELO 2	ALELO 1	ALELO 2	ALELO 1	ALELO 2				ALELO 1	ALELO 2	ALELO 1	ALELO 2	ALELO 1	ALELO 2
1	159606	A/G	0,312	0,688	0,49	0,51	0,38	0,62	48	1357617	A/T	0,764	0,236	0,888	0,112	0,833	0,167
2	214955	A/G	0,446	0,554	0,52	0,48	0,44	0,56	49	1360288	C/T	0,822	0,178	0,765	0,235	0,78	0,22
3	221956	C/T	0,653	0,347	0,673	0,327	0,62	0,38	50	1382387	G/T	0,332	0,668	0,255	0,745	0,44	0,56
4	251934	C/T	0,302	0,698	0,276	0,724	0,36	0,64	51	1413212	A/G	0,297	0,703	0,245	0,755	0,37	0,63
5	279844	A/T	0,48	0,52	0,551	0,449	0,5	0,5	52	1454361	A/T	0,564	0,436	0,643	0,357	0,64	0,36
6	321198	C/T	0,6	0,4	0,633	0,367	0,6	0,4	53	1463729	A/G	0,495	0,505	0,561	0,439	0,64	0,36
7	338882	C/T	0,44	0,56	0,583	0,417	0,602	0,398	54	1490413	A/G	0,505	0,495	0,561	0,439	0,5	0,5
8	354439	A/T	0,454	0,546	0,316	0,684	0,49	0,51	55	1493232	A/C	0,67	0,33	0,564	0,436	0,628	0,372
9	430046	C/T	0,599	0,401	0,52	0,48	0,54	0,46	56	1498553	C/T	0,46	0,54	0,367	0,633	0,48	0,52
10	445251	C/G	0,411	0,589	0,378	0,622	0,47	0,53	57	1523537	C/T	0,307	0,693	0,296	0,704	0,45	0,55
11	560681	A/G	0,752	0,248	0,704	0,296	0,67	0,33	58	1528460	C/T	0,37	0,63	0,306	0,694	0,592	0,408
12	576261	A/C	0,599	0,401	0,561	0,439	0,53	0,47	59	1736442	A/G	NA	NA	0,346	0,654	NA	NA
13	717302	A/G	0,653	0,347	0,714	0,286	0,61	0,39	60	1821380	C/G	0,47	0,53	0,378	0,622	0,41	0,59
14	719366	C/T	0,244	0,756	0,244	0,756	0,14	0,86	61	1886510	C/T	0,53	0,47	0,469	0,531	0,71	0,29
15	722098	A/G	0,49	0,51	0,408	0,592	0,43	0,57	62	1979255	C/G	0,52	0,48	0,48	0,52	0,61	0,39
16	722290	C/G	0,545	0,455	0,531	0,469	0,53	0,47	63	2040411	A/G	0,649	0,351	0,765	0,235	0,76	0,24
17	727811	A/C	0,495	0,505	0,469	0,531	0,41	0,59	64	2046361	A/T	0,574	0,426	0,653	0,347	0,45	0,55
18	729172	A/C	0,347	0,653	0,408	0,592	0,37	0,63	65	2056277	C/T	0,842	0,158	0,837	0,163	0,88	0,12
19	733164	A/G	0,277	0,728	0,296	0,704	0,24	0,76	66	2076848	A/T	0,277	0,728	0,255	0,745	0,214	0,786
20	735155	A/G	0,475	0,525	0,52	0,48	0,41	0,59	67	2107612	A/G	0,792	0,208	0,816	0,184	0,64	0,36
21	737681	C/T	0,748	0,252	0,796	0,204	0,64	0,36	68	2111980	A/G	0,554	0,446	0,571	0,429	0,61	0,39
22	740598	A/G	0,58	0,42	0,694	0,306	0,73	0,27	69	2269355	C/G	0,485	0,515	0,5	0,5	0,51	0,49
23	740910	A/G	0,634	0,266	0,592	0,408	0,74	0,26	70	2342747	A/G	0,418	0,582	0,408	0,592	0,459	0,541
24	763869	C/T	0,405	0,595	0,347	0,653	0,408	0,592	71	2399332	A/C	0,396	0,604	0,316	0,684	0,4	0,6
25	826472	C/T	0,688	0,312	0,694	0,306	0,79	0,21	72	2830795	A/G	0,755	0,243	0,765	0,235	0,79	0,21
26	873196	C/T	0,282	0,718	0,184	0,816	0,2	0,8	73	2831700	A/G	0,54	0,46	0,571	0,429	0,65	0,35
27	876724	C/T	0,782	0,218	0,796	0,204	0,81	0,19	74	2920816	C/T	NA	NA	0,409	0,591	NA	NA
28	891700	A/G	0,342	0,658	0,439	0,561	0,42	0,58	75	3780962	C/T	0,42	0,58	0,337	0,663	0,44	0,56
29	901398	C/T	0,366	0,634	0,449	0,551	0,34	0,66	76	4364205	G/T	0,663	0,337	0,694	0,306	0,61	0,39
30	907100	C/G	0,424	0,576	0,378	0,622	0,561	0,439	77	4530059	A/G	0,436	0,564	0,408	0,592	0,39	0,61
31	914165	A/G	0,58	0,42	0,388	0,612	0,6	0,4	78	4606077	C/T	0,693	0,307	0,75	0,25	0,694	0,306
32	917118	C/T	0,624	0,376	0,582	0,418	0,47	0,53	79	6444724	C/T	0,505	0,495	0,541	0,459	0,53	0,47
33	938283	C/T	0,158	0,842	0,122	0,878	0,17	0,83	80	6811238	G/T	0,46	0,54	0,51	0,49	0,56	0,44
34	964681	C/T	0,446	0,554	0,408	0,592	0,34	0,66	81	6955448	C/T	0,748	0,252	0,7654	0,235	0,75	0,25
35	987640	A/T	0,446	0,554	0,459	0,541	0,52	0,48	82	7041158	C/T	0,719	0,281	0,625	0,375	0,446	0,554
36	993934	C/T	0,292	0,708	0,245	0,755	0,28	0,72	83	8037429	C/T	0,545	0,455	0,582	0,418	0,56	0,44
37	1005533	A/G	0,46	0,54	0,48	0,52	0,38	0,62	84	8078417	C/T	0,599	0,401	0,592	0,408	0,61	0,39
38	1015250	C/G	0,237	0,763	0,194	0,806	0,4	0,6	85	9905977	A/G	0,228	0,772	0,245	0,755	0,42	0,58
39	1024116	A/G	0,653	0,347	0,592	0,408	0,46	0,54	86	9951171	A/G	0,475	0,525	0,449	0,551	0,42	0,58

40	1028528	A/G	0,757	0,243	0,827	0,173	0,54	0,46	87	10092491	C/T	0,598	0,402	0,633	0,367	0,653	0,347
41	1031825	A/C	0,28	0,72	0,326	0,674	0,393	0,607	88	10488710	C/G	0,578	0,422	0,582	0,418	0,643	0,357
42	1058083	A/G	0,426	0,574	0,459	0,541	0,41	0,59	89	10495407	A/G	0,332	0,668	0,337	0,663	0,22	0,78
43	1109037	A/G	0,485	0,515	0,469	0,531	0,45	0,55	90	10773760	A/G	0,58	0,42	0,551	0,449	0,66	0,34
44	1294331	A/G	NA	NA	NA	NA	NA	NA	91	10776839	G/T	0,52	0,48	0,418	0,582	0,57	0,43
45	1335873	A/T	0,441	0,559	0,255	0,745	0,57	0,43	92	12997453	A/G	0,429	0,571	0,378	0,622	0,388	0,612
46	1336071	A/G	0,545	0,455	0,561	0,439	0,44	0,56	93	13182883	A/G	0,356	0,644	0,378	0,622	0,446	0,554
47	1355366	A/G	NA	NA	NA	NA	NA	NA	94	13218440	A/G	NA	NA	NA	NA	NA	NA

Tabla 15. Frecuencias alélicas de 89 SNPs incluidos en el kit Signature Prep ForenSeq™ para las regiones Andina Central, Orinoquía y Pacífico de Colombia. NA datos eliminados.

En los loci SNP que han sido también analizados en otros estudios que incluyen población colombiana, ej: 1000 genomas (Birney & Soranzo, 2015) (Ibarra et al., 2014), se observan distribuciones similares a las obtenidas en el presente estudio, en poblaciones que pueden compararse con las delimitadas en nuestro diseño de estudio poblacional.

### **Indicadores genético poblacionales y genético forenses**

Los resultados obtenidos por región se encuentran resumidos en los anexos 7, 8 y los datos tomando las tres regiones como una sola en el anexo 9.

### **Equilibrio HWE**

Se observa que 19 loci de los 89 analizados mostraron desviación en los resultados para el equilibrio de Hardy-Weinberg ( $p < 0.05$ ); 5 en región Andina Central (rs1015250, rs2076848, rs873196, rs2342747, rs9951171), 3 en Orinoquía (rs12997453, rs1979255, rs717302) y 11 en Pacífico (rs993934, 12997453, rs717302, rs1015250, rs1335873, rs1058083, rs873196, rs1382387, rs445251, rs1005533 y rs733164), 5 de estos loci en aparente desequilibrio, se repiten en más de una población; sin embargo, estos resultados no tienen significancia si se comparan con la totalidad de los marcadores. Luego de la corrección de Bonferroni, todos los loci se ajustaron a la hipótesis de equilibrio de Hardy-Weinberg.

### **Heterocigocidad**

En relación con las frecuencias genotípicas, el 41% de los datos se encontraron en el rango de 0,4 a 0,69, el promedio para las 3 poblaciones estudiadas en todos los loci SNPs, fue de 0,34,

siendo el genotipo C/C del locus rs938283 el de más baja frecuencia, observándose solo en el 1% de la población colombiana.

El promedio de heterocigocidad esperada para las 3 regiones fue de 0,452 (Andina 0,44; Orinoquia 0,48 y Pacífico 0,43) aunque más del 63% de los loci alcanzaron heterocigocidades mayores a 0,4 demostrando un alto nivel de diversidad de las poblaciones estudiadas.

La alta heterocigocidad esperada, junto con la estabilidad mutacional propia de los SNPs los convierte en valiosas herramientas en la investigación del parentesco. Solo 2 SNPs de la población pacífica (rs719366 y rs1979255) y 1 en Orinoquia (rs1357617) tienen heterocigocidades, muy bajas comparados con las otras poblaciones.

De manera similar, el índice de contenido polimórfico (PIC) estimado a partir de las frecuencias alélicas para los marcadores SNPs estuvo en un rango entre 0,180 y 0,380, con un valor promedio de 0,345, son valores bajos comparados con los PIC que se observan en sistemas multialélicos; lo cual es de esperarse en SNPs aun así, para fines de identificación, el alto número de loci compensa la escasez de polimorfismo.

### **Fst**

Por otro lado, el  $F_{st}$  estimado para los SNPs analizados en las 3 regiones fue de 0,011 con un rango entre -0,021 para rs1360288 y 0,084 para rs1335873, mostrando que no hay diferencia significativa entre las poblaciones y que podría usarse una base de datos de frecuencias alélicas de los 89 SNPs unificada, para las 3 regiones de la población colombiana, ya que no hay estructura (subdivisión) dentro de la población. Sin embargo, para fines forenses y asumiendo el concepto de población de referencia, es decir el sitio donde ocurrieron los hechos, sería más conveniente usar la base de datos por región, dado que una autoridad puede solicitar que los resultados sean comparados con una base de datos local, por el conocimiento que se tiene de los hechos.

### **LD**

Al evaluar el desequilibrio de ligamiento entre loci SNPs, se encontró que existe independencia entre parejas de loci SNPs que se encuentran en un mismo cromosoma, solo 21 parejas de 277 probadas mostraron valores de  $P < 0.05$ , lo cual una vez hecha la corrección de Bonferroni no se

observó ningún valor significativo para LD. Este hallazgo es consecuente con lo reportado en la literatura, por Sánchez y otros, en donde no encuentran asociación probable entre 52 de los marcadores SNPs usados en el kit de Illumina entre ninguna de las parejas de SNPs ubicadas en el mismo cromosoma (Pereira et al., 2008; Sanchez et al., 2006).

Por otra parte, al evaluar la posibilidad de ligamiento o asociación entre SNPs y STRs, se tuvo en cuenta que el tamaño muestral usado en este estudio es muy pequeño para realizar un análisis poblacional con marcadores STRs, por lo tanto, se asumió lo referido en la literatura, en donde se reporta que no existe asociación probable con STR de uso rutinario comunes en el análisis forense (Pereira et al., 2008; Sanchez et al., 2006).

Por lo anterior, asumimos válido el uso de la regla del producto entre LR de marcadores STRs y SNPs, dada la independencia entre ellos, para aumentar los porcentajes de probabilidad (*match probabilities*) en casos complejos de parentescos, lo cual es de suma importancia para los casos de identificación en el marco del conflicto armado colombiano.

### **Indicadores de eficiencia a priori**

Adicionalmente, se evaluaron los indicadores de eficiencia *a priori* genético forenses: Poder de discriminación (PD), Probabilidad de coincidencia (PM) y Probabilidad de exclusión (PE).

Observamos que el poder de discriminación (PD) promedio obtenido para los 89 loci en las tres regiones fue de 0,598, con un rango limitado de 0,414 y 0,652 (rs6444724 y rs2056277), lo cual nos indica que la probabilidad de que dos individuos no relacionados y tomados al azar puedan ser diferenciados genéticamente mediante el análisis de estos marcadores es alta. Ahora bien, considerando la totalidad de los SNPs estudiados, la Probabilidad de coincidencia acumulada (PMcum) alcanzada, fuera de  $3,83 \times 10^{-36}$  para las tres regiones;  $1,07 \times 10^{-35}$  para región Andina Central,  $1,41 \times 10^{-34}$  en región Pacífica y  $2,05 \times 10^{-35}$  en la Orinoquía, por lo tanto, el Poder de discriminación acumulado de los 89 SNPs en las tres poblaciones colombianas estudiadas es notablemente alto  $> 1 - (3 \times 10^{-36})$ .

Por su parte, la probabilidad de exclusión acumulada (PE) de paternidad, alcanzó valores de 0,99999996 para los 89 loci analizados; los rangos por locus variaron entre 0,813 y 0,889 con un promedio de 0,825.

Los altos valores de PD y PE, estimados con las frecuencias alélicas de los 89 SNPs para las 3 regiones estudiadas, nos muestran, por una parte, la utilidad de estos marcadores en el análisis de uniprocedencias, es decir la tarea de asociar partes corporales en casos de cuerpos fragmentados y mezclados; situación muy común en muchas fosas colombianas y por otra parte el alto impacto que se logra en LR's al investigar el parentesco con déficit de familiares informativos.

Adicionalmente pueden ser marcadores de gran valor en el estudio de casos criminalísticos donde interesa evaluar la posibilidad de que una mancha de sangre, semen o saliva provenga de un individuo en particular.

## 7. CONCLUSIONES

### **Validación del kit Foren Seq y el equipo MiSeq NGS de la firma Illumina, al análisis de casos forenses de identificación:**

Bajo las condiciones del laboratorio de Genética de la Regional Bogotá del INMLCF y en muestras óptimas, fue posible detectar todos los 153 genotipos que conforman el Kit ForenSeq de Illumina en el equipo MiSeq, con un 100% de exactitud.

Así mismo, la reproducibilidad de los ensayos, en las mismas condiciones anteriores, no mostró diferencias significativas entre analistas en cuanto a la proporción y la exactitud de loci detectados, aunque si se encontró diferencia para el promedio del número de *reads* por locus, debido a la variación generada por la ausencia de automatización, durante el proceso de preparación de librerías.

Por otro lado, se pudo determinar que con muestras complejas (baja concentración de ADN) y montajes de alto número de librerías, la eficiencia del proceso de secuenciación tanto en genotipos asignados como en concordancia disminuye.

Con respecto al tipo de muestra analizada, las muestras de restos óseos, aun a 1ng/uL, presentan un desempeño menor en número de reads, comparadas con muestras de sangre y saliva en tarjetas FTA y ADN purificado de alto peso molecular (2800M)

Por lo anterior, se sugiere que para el análisis de muestras óseas complejas (ADN hasta 15 pg), se usen modificaciones al protocolo del fabricante (LCN2 y LCN2 pool 10), realizando análisis exhaustivo de *stutter*, *drop in* y *drop out* y verificando la reproducibilidad de los genotipos obtenidos.

Se observó una alta variación en los valores de profundidad de cobertura en función de:

- Número de librerías secuenciadas simultáneamente
- Tipo de muestra
- Tipo de marcador

Adicionalmente, algunos marcadores tanto STRs como SNPs parecen ser especialmente sensibles al montaje de un número alto de muestras, por lo que, se recomienda la optimización de las reacciones de amplificación y el diseño de primers en PCR target.

Finalmente, se establecieron los límites de detección en función del tipo de marcador y el número de librerías simultáneas, así: 62,5 pg a 125 pg para 40 librerías y 250 pg para más de 64 librerías, demostrando que la sensibilidad alcanzada en este kit, supera la reportada en los kits para marcadores STRs más sensibles, los cuales llegan hasta 100 o 125 pg.

Los umbrales de detección establecidos fueron:

- Umbral analítico (AT): 10 reads
- Umbral estocástico o de interpretación (AT): por grupos de loci, entre 164 y 820 reads para STRs y entre 94 y 262 reads para SNPs.

### **Análisis comparativo entre tecnología de nueva generación y electroforesis capilar en casos complejos y estudios poblacionales de 3 regiones colombianas para 94 SNPs de identificación, del kit forenseq illumina.**

Los SNPs analizados (89) mostraron valores de LR hasta de 106; además, debido al amplio número de loci que pueden analizarse, la posibilidad de combinar los dos tipos de marcadores STR y SNP en un mismo kit y la de aplicar la regla del producto dada su independencia, se propone que todos los casos complejos de parentesco, donde no existen familiares cercanos o los familiares disponibles no son suficientes para concluir los casos, sean abordados desde un comienzo con esta tecnología, lo cual reduciría tiempos y recursos de forma significativa.

La distribución alélica de los 89 loci SNPs en poblaciones colombianas, su alto PD y PE los convierte en marcadores ideales para fines de identificación humana:

- Análisis de Uniprocedencias
- Parentesco con déficit de familiares informativos.
- Investigaciones criminalísticas

## **8. CONSIDERACIONES BIOÉTICAS, DISPOSICIONES LEGALES VIGENTES Y PROPIEDAD INTELECTUAL**

El proyecto fue sometido y aprobado por el comité de ética de la Universidad del Rosario, principalmente en relación con las siguientes consideraciones:

Las muestras objeto del estudio, contaron todas con el consentimiento informado respectivo.

En todos los casos, para cualquier publicación de resultados, las muestras serán anonimizadas utilizando una numeración aleatorizada y diferente al código del caso institucional, solamente se conservarán los orígenes geográficos de las muestras ya que esta es una variable a considerar dentro del estudio poblacional; se tomarán los apellidos y municipio de origen de las personas, solo con el fin de tamizar los posibles casos de parentescos y eliminar a uno de los posibles familiares.

Los marcadores utilizados en el estudio solo están asociados a fines de identificación y no se relacionan con rasgos físicos o mentales, normales o patológicos de las personas muestradantes. No se informaron discordancias del sexo observadas en las muestras.

No se realizaron estudios de herencia entre las personas estudiadas; de hecho, solo se ingresaron al estudio personas no emparentadas; no obstante, cuando detectamos relaciones de parentesco inesperadas; estas no fueron informadas.

El estudio no generó riesgo alguno para las personas que aportaron su muestra. Igualmente, el estudio no planteó la toma de muestras, sino que utilizó el biobanco de muestras en tarjetas FTA del INML relacionadas con casos de investigación de la paternidad o la maternidad.

Consideramos que el derecho a la confidencialidad de la información personal y sensible, esta salvaguardado en el diseño de la investigación.

Como lo designa la ley actual, los derechos patrimoniales corresponden al INML ya que fue el Instituto el que cubrió todos los costos de equipos, insumos, reactivos, y personal científico. Los derechos morales corresponden exclusivamente a los autores.

## 9. BIBLIOGRAFIA

- Almalki, N., Chow, H. Y., Sharma, V., Hart, K., Siegel, D., & Wurmbach, E. (2017). Systematic assessment of the performance of illumina's MiSeq FGx™ forensic genomics system. *Electrophoresis*, 38(6), 846–854. <https://doi.org/10.1002/elps.201600511>
- Ambers, A. D., Churchill, J. D., King, J. L., Stoljarova, M., Gill-King, H., Assidi, M., ... Budowle, B. (2016). More comprehensive forensic genetic marker analyses for accurate human remains identification using massively parallel DNA sequencing. *BMC Genomics*, 17(S9), 750. <https://doi.org/10.1186/s12864-016-3087-2>
- Ambers, A., Turnbough, M., Benjamin, R., Gill-King, H., King, J., Sajantila, A., & Budowle, B. (2016). Modified DOP-PCR for improved STR typing of degraded DNA from human skeletal remains and bloodstains. *Legal Medicine*. <https://doi.org/10.1016/j.legalmed.2015.10.013>
- Babol-Pokora, K., & Berent, J. (2008). SNP-minisequencing as an excellent tool for analysing degraded DNA recovered from archival tissues. *Acta Biochimica Polonica*, 55(4), 815–819. <https://doi.org/20081654> [pii]
- Barón, M. D. (2015). Las víctimas del conflicto armado colombiano en la Ley de Víctimas y Restitución de Tierras: Apropiación y resignificación de una categoría jurídica. *Perfiles Latinoamericanos*, 23(46), 121–145. <https://doi.org/10.18504/pl2346-121-2015>
- Battista, J. R. (2012). *Final Technical Report Title: Tools for Improving the Quality of Aged, Degraded, Damaged, or Otherwise Compromised DNA Evidence*.
- Birney, E., & Soranzo, N. (2015). The end of the start for population sequencing. *Nature*, 526(7571), 52–53. <https://doi.org/10.1038/526052a>
- Børsting, C., & Morling, N. (2015). Next generation sequencing and its applications in forensic genetics. *Forensic Science International: Genetics*, 18, 78–89. <https://doi.org/10.1016/j.fsigen.2015.02.002>

- Budowle, B., Eisenberg, A. J., & van Daal, A. (2009). Validity of Low Copy Number Typing and Applications to Forensic Science. *Croatian Medical Journal*, 50(3), 207–217. <https://doi.org/10.3325/cmj.2009.50.207>
- Budowle, B., & Van Daal, A. (2008). Forensically relevant SNP classes. *BioTechniques*, 44(5), 603–610. <https://doi.org/10.2144/000112806>
- Butler, J. (2006). Debunking Some Urban Legends Surrounding Validation Within the Forensic DNA Community. *Profiles in DNA*, 3–6.
- Casals, F., Anglada, R., Bonet, N., Rasal, R., van der Gaag, K. J., Hoogenboom, J., ... Calafell, F. (2017). Length and repeat-sequence variation in 58 STRs and 94 SNPs in two Spanish populations. *Forensic Science International. Genetics*, 30, 66–70. <https://doi.org/10.1016/j.fsigen.2017.06.006>
- Dixon, L. A., Dobbins, A. E., Pulker, H. K., Butler, J. M., Vallone, P. M., Coble, M. D., ... Gill, P. (2006). Analysis of artificially degraded DNA using STRs and SNPs-results of a collaborative European (EDNAP) exercise. *Forensic Science International*, 164(1), 33–44. <https://doi.org/10.1016/j.forsciint.2005.11.011>
- Dixon, L. A., Murray, C. M., Archer, E. J., Dobbins, A. E., Koumi, P., & Gill, P. (2005). Validation of a 21-locus autosomal SNP multiplex for forensic identification purposes. *Forensic Science International*, 154(1), 62–77. <https://doi.org/10.1016/j.forsciint.2004.12.011>
- Forense, G., Gen, A. D. N. La, Gen, L., & Landsteiner, K. (1960). *Genética forense I*. (1900), 357–362.
- Garcia, Y. E. G. (2014). Las políticas de justicia y paz en el reconocimiento a las víctimas del conflicto armado en Colombia. *Revista Pensamiento Americano*, 6(10), 23–29. Retrieved from <http://www.coruniamericana.edu.co/publicaciones/ojs/index.php/pensamientoamericano/article/view/134>
- Gettings, K. B., Kiesler, K. M., & Vallone, P. M. (2015). Performance of a next generation

- sequencing SNP assay on degraded DNA. *Forensic Science International: Genetics*, 19(2015), 1–9. <https://doi.org/10.1016/j.fsigen.2015.04.010>
- Giardina, E., Predazzi, I., Pietrangeli, I., Asili, P., Marsala, P., Gabriele, L., ... Novelli, G. (2007). Frequency assessment of SNPs for forensic identification in different populations. *Forensic Science International: Genetics*, 1(3–4), 2–4. <https://doi.org/10.1016/j.fsigen.2007.05.002>
- Guo, F., Yu, J., Zhang, L., & Li, J. (2017). Massively parallel sequencing of forensic STRs and SNPs using the Illumina® ForenSeq™ DNA Signature Prep Kit on the MiSeq FGx™ Forensic Genomics System. *Forensic Science International: Genetics*, 31(September 2017), 135–148. <https://doi.org/10.1016/j.fsigen.2017.09.003>
- Hansson, O., Egeland, T., & Gill, P. (2017). Characterization of degradation and heterozygote balance by simulation of the forensic DNA analysis process. *International Journal of Legal Medicine*, 131(2), 303–317. <https://doi.org/10.1007/s00414-016-1453-x>
- Ibarra, A., Freire-Aradas, A., Martínez, M., Fondevila, M., Burgos, G., Camacho, M., ... Gusmão, L. (2014). Comparison of the genetic background of different Colombian populations using the SNPforID 52plex identification panel. *International Journal of Legal Medicine*, 128(1), 19–25. <https://doi.org/10.1007/s00414-013-0858-z>
- Illumina. (n.d.). *System Specification Sheet: Forensic Genomics*. Retrieved from <http://www.illumina.com/content/dam/illumina-marketing/documents/products/datasheets/miseq-fgx-system-spec-sheet-1470-2014-004.pdf>
- Illumina. (2015a). *ForenSeq™ Universal Analysis Software User Guide*. (August).
- Illumina. (2015b). *Guía de referencia de ForenSeq™ DNA SignaturePrep*.
- Illumina. (2015c). *MiSeq FGx Instrument Reference Guide*. (February).
- Illumina. (2015d). *MiSeq FGx™ Reagent Kit Reference Guide*. (January 2015).
- Jager, A., Alvarez, M., Davis, C. P., Guzmán, E., Han, Y., Way, L., ... Stephens, K. M. (2017). Developmental Validation of the MiSeq FGx Forensic Genomics System for Targeted Next

- Generation Sequencing in Forensic DNA Casework and Databasing. *Forensic Science International: Genetics*, 28, 52–70. <https://doi.org/10.1016/j.fsigen.2017.01.011>
- Jennifer D. Churchill\*, Sarah E. Schmedes, Jonathan L. King, B. B. (2016). Evaluation of the Illumina?? Beta Version ForenSeq??? DNA Signature Prep Kit for use in genetic profiling. *Forensic Science International: Genetics*, 20, 20–29. <https://doi.org/10.1016/j.fsigen.2015.09.009>
- Jones, D. A. (1972). Blood Samples : Probability of Discrimination. *Journal of the Forensic Science Society*, 12(2), 355–359. [https://doi.org/10.1016/S0015-7368\(72\)70695-7](https://doi.org/10.1016/S0015-7368(72)70695-7)
- Kidd, K. K., Pakstis, A. J., Speed, W. C., Grigorenko, E. L., Kajuna, S. L. B., Karoma, N. J., ... Kidd, J. R. (2006). Developing a SNP panel for forensic identification of individuals. *Forensic Science International*, 164(1), 20–32. <https://doi.org/10.1016/j.forsciint.2005.11.017>
- LifeTechnologies. (2014). *Targeted Sequencing Solutions*. Retrieved from <https://tools.thermofisher.com/content/sfs/brochures/Targeted-Sequencing-Brochure.pdf>
- Mardis, E. R. (2008). Next-Generation DNA Sequencing Methods. *Annual Review of Genomics and Human Genetics*, 9(1), 387–402. <https://doi.org/10.1146/annurev.genom.9.081307.164359>
- Mccord, B. R., Opel, K., Funes, M., & Jantz, L. M. (2011). An Investigation of the Effect of DNA Degradation and Inhibition on PCR Amplification of Single Source and Mixed Forensic Samples. *Project 2006-DN-BX-K006. Final Report to the US Department of Justice*, 1–66.
- Mo, S.-K., Liu, Y.-C., Wang, S., Bo, X.-C., Li, Z., Chen, Y., & Ni, M. (2016). Exploring the efficacy of paternity and kinship testing based on single nucleotide polymorphisms. *Forensic Science International. Genetics*, 22, 161–168. <https://doi.org/10.1016/j.fsigen.2016.02.012>
- Pakstis, A. J., Speed, W. C., Fang, R., Hyland, F. C. L., Furtado, M. R., Kidd, J. R., & Kidd, K. K. (2010). SNPs for a universal individual identification panel. *Human Genetics*, 127(3),

315–324. <https://doi.org/10.1007/s00439-009-0771-1>

- Pereira, R., Fondevila, M., Phillips, C., Amorim, A., Carracedo, A., & Gusmão, L. (2008). Genetic characterization of 52 autosomal SNPs in the Portuguese population. *Forensic Science International: Genetics Supplement Series*, 1(1), 358–360. <https://doi.org/10.1016/j.fsigss.2007.10.160>
- Presidencia de la República, & FARC-EP. (2016). Que Como Resultado De Los Diálogos Exploratorios Referidos Se Produjo Un. *Acuerdo Final Para La Terminación Del Conflicto Y La Construcción De Una Paz Estable Y Duradera*, 1–310.
- Rishishwar, L., Conley, A. B., Wigington, C. H., Wang, L., Valderrama-Aguirre, A., & King Jordan, I. (2015). Ancestry, admixture and fitness in Colombian genomes. *Scientific Reports*, 5(1), 12376. <https://doi.org/10.1038/srep12376>
- Ruiz-Linares, A., Adhikari, K., Acuña-Alonzo, V., Quinto-Sanchez, M., Jaramillo, C., Arias, W., ... Gonzalez-José, R. (2014). Admixture in Latin America: Geographic Structure, Phenotypic Diversity and Self-Perception of Ancestry Based on 7,342 Individuals. *PLoS Genetics*, 10(9). <https://doi.org/10.1371/journal.pgen.1004572>
- Sanchez, J. J., Phillips, C., Børsting, C., Balogh, K., Bogus, M., Fondevila, M., ... Morling, N. (2006). A multiplex assay with 52 single nucleotide polymorphisms for human identification. *Electrophoresis*, 27(9), 1713–1724. <https://doi.org/10.1002/elps.200500671>
- Science, N. R. C. (US) C. on D. F. (1996). The Evaluation of Forensic DNA Evidence Library of Congress Cataloging-in-Publication Data The evaluation of forensic DNA evidence / Committee on DNA Forensic Science: an Update, Commission on DNA Forensic. In N. A. P. (US); 1996. (Ed.), *E94* (1996., Vol. 51). Retrieved from <http://www.nap.edu/catalog/5141.html>
- Seo, S. B., King, J. L., Warshauer, D. H., Davis, C. P., Ge, J., & Budowle, B. (2013). Single nucleotide polymorphism typing with massively parallel sequencing for human identification. *International Journal of Legal Medicine*, 127(6), 1079–1086. <https://doi.org/10.1007/s00414-013-0879-7>

- Swgdam, A. (2012). *Scientific Working Group on DNA Analysis Methods Validation Guidelines for DNA Analysis Methods*. (July), 1–13.
- Swgdam, A. (2016). *Scientific Working Group on DNA Analysis Methods Validation Guidelines for DNA Analysis Methods*. (MAY), 1–13.
- Swgdam, S. W. G., & Methods, D. N. A. A. (2019). *SWGDAM Interpretation Guidelines for Autosomal STR Typing* *SWGDAM Interpretation Guidelines for Autosomal STR Typing by Forensic DNA Testing Laboratories Scientific Working Group on DNA Analysis Methods (SWGDAM)*. 1–28. Retrieved from [http://www.forensicdna.com/assets/swgdam\\_2010.pdf](http://www.forensicdna.com/assets/swgdam_2010.pdf)
- Vallone, P. M., Decker, A. E., & Butler, J. M. (2005). Allele frequencies for 70 autosomal SNP loci with U.S. Caucasian, African-American, and Hispanic samples. *Forensic Science International*, 149(2–3), 279–286. <https://doi.org/10.1016/j.forsciint.2004.07.014>
- Weber, J. L., & May, P. E. (1989). Abundant class of human DNA polymorphisms which can be typed using the polymerase chain reaction. *American Journal of Human Genetics*, 44(3), 388–396. Retrieved from <http://www.ncbi.nlm.nih.gov/pubmed/2916582> <http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=PMC1715443>
- Xavier, C., & Parson, W. (2017). Evaluation of the Illumina ForenSeq™ DNA Signature Prep Kit – MPS forensic application for the MiSeq FGx™ benchtop sequencer. *Forensic Science International: Genetics*, 28(2017), 188–194. <https://doi.org/10.1016/j.fsigen.2017.02.018>

## 10. ANEXOS

Los anexos del 1 al 6, se encuentran adjuntos como archivos suplementarios

**Anexo 1.** Ensayo de Exactitud. Genotipos y análisis

**Anexo 2.** Ensayo de Precisión. Genotipos y análisis

**Anexo 3.** N° de Reads para ensayos de Exactitud y Precisión. Análisis

**Anexo 4.** Ensayo de Reproducibilidad. Genotipos, N° de reads y análisis

**Anexo 5.** Ensayo de Sensibilidad. Genotipos y análisis

**Anexo 6.** Ensayo de Variación del volumen de librerías y LCN. Análisis

## Anexo7. Indicadores genético poblacionales por región

SNP	EQUILIBRIO DE HARDY WEINBERG (H-W)			HETEROCIGOSIDAD ESPERADA			ESTRUCTURA POBLACIONAL (Fst)			CONTENIDO DE INFORMACION POLIMORFICA (PIC)		
	ANDINA CENTRAL	ORINOQUIA	PACIFICO	ANDINA CENTRAL	ORINOQUIA	PACIFICO	ANDINA CENTRAL	ORINOQUIA	PACIFICO	ANDINA CENTRAL	ORINOQUIA	PACIFICO
159606	1,0000	0,1580	0,3594	0,426	0,612	0,520	0,107	-0,046	0,014	0,340	0,370	0,360
214955	0,2420	0,3910	0,2468	0,436	0,429	0,560	0,002	-0,014	-0,001	0,370	0,370	0,370
221956	0,2036	0,3040	0,5615	0,396	0,367	0,440	0,010	0,033	-0,036	0,350	0,340	0,360
251934	1,0000	0,4760	0,5390	0,426	0,347	0,440	0,029	0,076	-0,066	0,330	0,320	0,350
279844	0,8417	1,0000	0,5626	0,505	0,490	0,560	-0,003	0,001	-0,010	0,370	0,370	0,380
321198	0,2133	0,4060	0,3616	0,420	0,408	0,440	-0,010	0,017	-0,015	0,360	0,360	0,360
338882	0,1218	0,1410	0,1910	0,571	0,375	0,386	0,013	0,022	0,035	0,370	0,370	0,360
354439	0,3035	0,0950	0,2724	0,454	0,551	0,420	-0,033	0,094	-0,046	0,370	0,340	0,370
430046	0,8324	0,7920	0,5763	0,465	0,469	0,560	-0,002	-0,047	-0,042	0,370	0,370	0,370
445251	0,1041	0,7840	0,0251	0,406	0,510	0,340	0,014	0,038	-0,020	0,370	0,360	0,370
560681	0,1764	0,5100	0,1090	0,317	0,469	0,340	0,083	-0,031	-0,094	0,300	0,330	0,340
576261	0,2159	1,0000	0,0935	0,545	0,510	0,380	0,015	-0,015	-0,027	0,370	0,370	0,370
717302	0,3859	0,0420	0,0156	0,416	0,286	0,340	-0,049	0,049	-0,108	0,350	0,320	0,360
719366	0,2577	0,2410	0,1630	0,330	0,444	0,186	0,011	0,006	0,353	0,300	0,300	0,300
722098	0,5550	1,0000	0,1492	0,525	0,490	0,380	-0,013	0,016	0,002	0,370	0,370	0,370
722290	0,1586	0,3740	0,5944	0,436	0,571	0,460	-0,004	-0,013	-0,013	0,370	0,370	0,370
727811	0,1019	0,7760	0,7698	0,416	0,531	0,500	-0,013	-0,015	0,014	0,370	0,370	0,370
729172	0,2618	0,2600	0,1364	0,515	0,571	0,380	0,028	-0,043	-0,006	0,350	0,370	0,360
733164	0,8046	1,0000	0,0441	0,406	0,429	0,280	0,085	0,033	0,153	0,320	0,330	0,300
735155	0,3217	0,2580	0,7773	0,455	0,592	0,500	-0,006	-0,012	0,019	0,370	0,370	0,370
737681	0,7914	0,6680	0,7636	0,366	0,367	0,480	-0,041	0,099	-0,277	0,310	0,270	0,350
740598	0,6773	0,7560	0,1484	0,460	0,408	0,300	-0,057	0,074	0,141	0,370	0,330	0,320
740910	0,1909	0,7770	0,4693	0,416	0,449	0,440	0,033	-0,012	0,194	0,360	0,370	0,310
763869	0,6810	1,0000	0,3850	0,470	0,449	0,408	-0,018	0,038	-0,026	0,370	0,370	0,370
826472	1,0000	0,3000	0,1128	0,426	0,367	0,260	-0,112	-0,106	0,136	0,340	0,330	0,280
873196	0,0255	0,1960	0,0139	0,307	0,245	0,240	-0,156	0,140	0,082	0,320	0,250	0,270
876724	0,5515	0,0890	0,6666	0,356	0,245	0,340	-0,004	0,037	0,088	0,280	0,270	0,260
891700	0,8244	0,7680	0,1509	0,446	0,469	0,400	0,050	-0,045	-0,034	0,350	0,370	0,370
901398	0,2961	0,3920	0,7600	0,515	0,571	0,440	0,027	-0,043	0,054	0,360	0,370	0,350
907100	0,2079	0,7710	0,5730	0,424	0,510	0,551	0,015	0,048	0,002	0,370	0,360	0,370
914165	0,1572	0,3750	0,5688	0,560	0,408	0,520	0,019	0,039	0,029	0,370	0,360	0,360
917118	0,2699	0,5760	0,3913	0,416	5,511	0,580	0,032	-0,008	-0,032	0,360	0,370	0,370
938283	0,2583	1,0000	0,3136	0,297	0,245	0,340	-0,053	0,147	-0,120	0,230	0,190	0,240
964681	0,7069	0,5820	0,2128	0,515	0,531	0,520	-0,038	-0,020	0,052	0,370	0,370	0,350
987640	0,8467	0,4130	0,4084	0,495	0,429	0,560	0,006	-0,004	-0,009	0,370	0,370	0,370
993934	1,0000	0,4410	0,0053	0,426	0,327	0,240	-0,046	0,060	-0,025	0,330	0,300	0,320
1005533	0,1117	1,0000	0,0169	0,426	0,510	0,320	-0,032	-0,043	0,016	0,370	0,370	0,360
1015250	0,0113	0,6480	0,0032	0,273	0,347	0,280	0,071	0,193	-0,238	0,300	0,260	0,360
1024116	1,0000	0,7630	1,0000	0,455	0,449	0,520	0,069	0,002	-0,027	0,350	0,370	0,370
1028528	0,2862	1,0000	0,1729	0,327	0,306	0,600	0,081	0,279	-0,248	0,300	0,250	0,370
1031825	0,3949	0,4820	1,0000	0,373	0,391	0,500	0,055	-0,034	-0,124	0,320	0,340	0,360
1058083	0,8417	0,1400	0,0431	0,475	0,388	0,620	-0,017	-0,038	-0,011	0,370	0,370	0,370
1109037	0,5598	0,0890	0,1025	0,535	0,367	0,380	-0,010	-0,012	-0,006	0,370	0,370	0,370
1335873	0,0722	0,7030	0,0225	0,584	0,429	0,340	0,002	0,226	0,002	0,370	0,310	0,370

1336071	0,2323	0,3600	0,4122	0,566	0,429	0,440	0,002	0,004	0,003	0,370	0,370	0,370
1357617	0,5279	0,0980	1,0000	0,392	0,143	0,289	-0,217	0,326	0,060	0,300	0,180	0,240
1360288	1,0000	0,4150	0,6812	0,297	0,306	0,320	0,037	-0,188	-0,134	0,250	0,290	0,280
1382387	0,0787	0,4520	0,0461	0,366	0,347	0,360	0,000	0,138	-0,117	0,350	0,310	0,370
1413212	0,6279	0,2670	0,7670	0,436	0,449	0,460	0,009	0,117	-0,112	0,330	0,300	0,360
1454361	0,8382	1,0000	0,7628	0,475	0,469	0,480	-0,016	0,047	0,043	0,370	0,350	0,350
1463729	0,2412	1,0000	0,1395	0,436	0,510	0,560	-0,022	-0,012	0,053	0,370	0,370	0,350
1490413	0,2168	1,0000	0,5699	0,436	0,510	0,560	-0,008	0,002	-0,013	0,370	0,370	0,380
1493232	0,4671	0,0850	1,0000	0,409	0,362	0,489	0,013	-0,104	-0,049	0,340	0,370	0,360
1498553	0,3058	1,0000	0,4147	0,440	0,490	0,560	0,004	0,063	-0,006	0,370	0,360	0,370
1523537	0,8176	0,4710	0,2696	0,416	0,469	0,420	0,069	0,083	-0,089	0,330	0,330	0,370
1528460	0,5343	0,7350	0,3630	0,440	0,408	0,408	0,043	0,124	0,003	0,360	0,330	0,370
1821380	0,5459	0,3650	0,2539	0,525	0,551	0,420	-0,017	0,036	0,007	0,370	0,360	0,370
1886510	0,8551	1,0000	0,7311	0,505	0,490	0,420	-0,009	-0,014	0,162	0,370	0,370	0,330
1979255	0,6940	0,0360	0,2230	0,485	0,347	0,000	-0,003	-0,008	0,039	0,370	0,370	0,360
2040411	0,3804	0,4280	0,2480	0,505	0,306	0,320	-0,068	0,154	0,141	0,350	0,290	0,300
2046361	0,5416	0,3470	0,2777	0,455	0,531	0,420	-0,013	0,056	-0,031	0,370	0,350	0,370
2056277	0,1198	0,3090	0,5224	0,238	0,327	0,240	-0,125	-0,159	0,104	0,230	0,240	0,190
2076848	0,0057	0,2730	1,0000	0,287	0,306	0,347	-0,053	-0,015	0,101	0,320	0,310	0,280
2107612	0,5598	0,6310	0,7636	0,317	0,286	0,480	0,054	0,134	-0,330	0,280	0,250	0,350
2111980	0,6894	1,0000	0,2292	0,515	0,490	0,420	-0,001	0,002	0,031	0,370	0,370	0,360
2269355	1,0000	0,2480	0,5812	0,510	0,592	0,460	-0,004	-0,010	-0,009	0,370	0,380	0,370
2342747	0,0116	0,1120	0,5820	0,361	0,367	0,551	0,007	0,008	-0,019	0,370	0,370	0,370
2399332	0,6787	1,0000	0,5558	0,455	0,429	0,520	-0,027	0,067	-0,036	0,360	0,340	0,360
2830795	0,7947	0,7150	0,6646	0,386	0,347	0,340	0,014	0,031	0,105	0,300	0,290	0,280
2831700	0,5512	0,5800	1,0000	0,525	0,449	0,460	-0,013	-0,004	0,067	0,370	0,370	0,350
3780962	0,2174	0,7540	0,1630	0,420	0,429	0,400	-0,012	0,067	-0,029	0,370	0,350	0,370
4364205	0,2670	0,3370	1,0000	0,495	0,367	0,500	-0,009	0,035	-0,080	0,350	0,330	0,360
4530059	0,6907	1,0000	0,8078	0,515	0,490	0,500	-0,020	-0,008	0,008	0,370	0,370	0,360
4606077	1,0000	1,0000	0,7470	0,427	0,375	0,408	-0,006	0,109	-0,009	0,340	0,300	0,330
6444724	0,2260	0,3840	0,2529	0,436	0,429	0,420	-0,004	-0,003	-0,006	0,370	0,370	0,370
6811238	0,8314	0,2240	0,1625	0,505	0,408	0,600	0,003	-0,008	0,006	0,370	0,370	0,370
6955448	0,4219	0,1030	0,4747	0,347	0,265	0,340	-0,001	0,043	0,001	0,310	0,290	0,300
7041158	0,1582	1,0000	0,3450	0,479	0,458	0,405	0,104	-0,043	-0,103	0,320	0,360	0,370
8037429	0,7014	0,1540	0,1617	0,515	0,388	0,600	0,002	0,016	0,003	0,370	0,370	0,370
8078417	0,6793	0,7570	0,8025	0,505	0,449	0,500	-0,028	-0,040	-0,024	0,370	0,370	0,360
9905977	0,3902	0,2480	0,2370	0,317	0,449	0,560	0,068	0,015	-0,297	0,290	0,300	0,370
9951171	0,0157	0,3810	0,4000	0,614	0,571	0,440	-0,006	-0,003	0,012	0,370	0,370	0,370
10092491	0,8292	0,7410	0,7740	0,495	0,449	0,490	-0,009	0,020	0,044	0,370	0,360	0,350
10488710	0,5226	0,5340	1,0000	0,469	0,551	0,469	-0,060	-0,062	-0,002	0,370	0,370	0,350
10495407	0,0810	0,2230	0,6820	0,366	0,551	0,320	-0,009	-0,022	0,215	0,350	0,350	0,280
10773760	0,5540	0,7430	0,2116	0,520	0,531	0,520	-0,008	-0,029	0,067	0,370	0,370	0,350
10776839	0,1147	0,5550	0,1502	0,426	0,551	0,580	-0,001	0,019	0,012	0,370	0,370	0,370
12997453	0,4022	0,0130	0,0380	0,531	0,306	0,327	-0,025	0,011	0,001	0,370	0,360	0,360
13182883	0,8164	0,5310	0,7430	0,444	0,429	0,457	0,039	0,009	-0,042	0,350	0,360	0,370
PROM	<b>0,473</b>	<b>0,534</b>	<b>0,428</b>	<b>0,441</b>	<b>0,487</b>	<b>0,429</b>	<b>-0,002</b>	<b>0,024</b>	<b>-0,006</b>	<b>0,348</b>	<b>0,340</b>	<b>0,347</b>
PROM 3 regiones				<b>0,452</b>			<b>0,005</b>			<b>0,345</b>		

## Anexo 8. Indicadores genético forenses por región

SNP	PODER DE DISCRIMINACION (PD)			PODER DE EXCLUSION (PE)			PROBABILIDAD DE MATCH (PM)		
	ANDINA CENTRAL	ORINOQUIA	PACIFICO	ANDINA CENTRAL	ORINOQUIA	PACIFICO	ANDINA CENTRAL	ORINOQUIA	PACIFICO
159606	0,583	0,550	0,586	0,130	0,306	0,206	0,417	0,450	0,414
214955	0,645	0,652	0,582	0,137	0,132	0,246	0,355	0,348	0,418
221956	0,614	0,605	0,621	0,112	0,095	0,140	0,386	0,395	0,379
251934	0,575	0,566	0,610	0,130	0,085	0,140	0,425	0,434	0,390
279844	0,622	0,625	0,590	0,192	0,179	0,246	0,378	0,375	0,410
321198	0,635	0,623	0,630	0,127	0,119	0,140	0,365	0,377	0,370
338882	0,575	0,650	0,642	0,258	0,990	0,106	0,425	0,350	0,358
354439	0,641	0,528	0,655	0,150	0,236	0,127	0,359	0,472	0,345
430046	0,620	0,638	0,586	0,159	0,162	0,246	0,379	0,362	0,414
445251	0,643	0,590	0,665	0,118	0,197	0,081	0,357	0,410	0,335
560681	0,539	0,556	0,609	0,071	0,162	0,081	0,461	0,444	0,391
576261	0,580	0,612	0,662	0,230	0,197	0,102	0,420	0,388	0,338
717302	0,609	0,571	0,642	0,124	0,058	0,081	0,391	0,429	0,358
719366	0,536	0,518	0,374	0,076	0,143	0,026	0,464	0,482	0,626
722098	0,612	0,613	0,654	0,210	0,179	0,102	0,388	0,387	0,346
722290	0,647	0,580	0,641	0,137	0,258	0,155	0,353	0,420	0,359
727811	0,656	0,606	0,609	0,124	0,216	0,188	0,344	0,394	0,391
729172	0,570	0,565	0,630	0,201	0,258	0,102	0,430	0,435	0,370
733164	0,555	0,570	0,527	0,118	0,132	0,056	0,445	0,430	0,473
735155	0,643	0,566	0,609	0,151	0,281	0,188	0,357	0,434	0,391
737681	0,542	0,490	0,595	0,095	0,095	0,171	0,458	0,510	0,405
740598	0,630	0,583	0,559	0,155	0,119	0,064	0,370	0,417	0,441
740910	0,621	0,630	0,534	0,124	0,147	0,140	0,379	0,370	0,466
763869	0,621	0,600	0,641	0,163	0,147	0,119	0,379	0,400	0,359
826472	0,583	0,590	0,490	0,130	0,095	0,048	0,417	0,140	0,510
873196	0,571	0,455	0,474	0,066	0,043	0,042	0,429	0,545	0,526
876724	0,507	0,480	0,474	0,090	0,043	0,081	0,493	0,520	0,526
891700	0,598	0,631	0,647	0,144	0,162	0,114	0,402	0,369	0,353
901398	0,582	0,576	0,598	0,201	0,258	0,140	0,418	0,424	0,402
907100	0,643	0,590	0,588	0,129	0,197	0,236	0,357	0,410	0,412
914165	0,577	0,633	0,594	0,246	0,119	0,206	0,423	0,367	0,406
917118	0,626	0,582	0,574	0,124	0,236	0,268	0,374	0,418	0,426
938283	0,431	0,370	0,449	0,062	0,043	0,081	0,569	0,630	0,551
964681	0,611	0,591	0,563	0,201	0,216	0,206	0,389	0,409	0,437
987640	0,622	0,650	0,589	0,183	0,132	0,246	0,378	0,350	0,411
993934	0,567	0,536	0,557	0,130	0,075	0,042	0,433	0,464	0,443
1005533	0,651	0,619	0,638	0,130	0,197	0,072	0,349	0,381	0,362
1015250	0,523	0,479	0,642	0,053	0,085	0,056	0,477	0,521	0,358
1024116	0,597	0,630	0,611	0,151	0,147	0,206	0,403	0,370	0,389
1028528	0,534	0,452	0,557	0,075	0,066	0,291	0,466	0,548	0,443
1031825	0,567	0,601	0,602	0,099	0,109	0,188	0,433	0,990	0,398
1058083	0,625	0,659	0,527	0,167	0,107	0,316	0,375	0,341	0,473
1109037	0,605	0,663	0,658	0,220	0,095	0,102	0,395	0,337	0,342
1335873	0,565	0,533	0,657	0,272	0,132	0,081	0,435	0,467	0,343

1336071	0,582	0,646	0,642	0,252	0,132	0,140	0,418	0,354	0,358
1357617	0,523	0,312	0,441	0,109	0,016	0,059	0,477	0,688	0,559
1360288	0,458	0,525	0,510	0,062	0,066	0,072	0,542	0,475	0,490
1382387	0,608	0,546	0,658	0,095	0,085	0,091	0,392	0,454	0,342
1413212	0,569	0,516	0,609	0,137	0,147	0,155	0,431	0,484	0,391
1454361	0,628	0,598	0,595	0,167	0,162	0,171	0,372	0,402	0,405
1463729	0,651	0,612	0,550	0,137	0,197	0,246	0,349	0,388	0,450
1490413	0,651	0,612	0,590	0,137	0,197	0,246	0,349	0,388	0,410
1493232	0,600	0,657	0,598	0,120	0,092	0,178	0,400	0,343	0,402
1498553	0,646	0,595	0,589	0,140	0,179	0,246	0,354	0,405	0,411
1523537	0,582	0,556	0,650	0,124	0,162	0,127	0,418	0,444	0,350
1528460	0,616	0,583	0,641	0,140	0,119	0,119	0,384	0,417	0,359
1821380	0,610	0,566	0,639	0,210	0,236	0,127	0,390	0,434	0,361
1886510	0,621	0,628	0,567	0,192	0,179	0,127	0,379	0,372	0,433
1979255	0,631	0,666	0,631	0,175	0,085	0,127	0,369	0,334	0,369
2040411	0,578	0,525	0,531	0,192	0,066	0,072	0,422	0,475	0,469
2046361	0,633	0,561	0,650	0,151	0,216	0,127	0,367	0,439	0,350
2056277	0,420	0,440	0,365	0,041	0,075	0,042	0,580	0,560	0,635
2076848	0,560	0,546	0,503	0,058	0,066	0,086	0,440	0,454	0,497
2107612	0,496	0,463	0,595	0,071	0,058	0,017	0,504	0,537	0,405
2111980	0,611	0,620	0,631	0,201	0,179	0,127	0,389	0,380	0,369
2269355	0,619	0,566	0,642	0,196	0,281	0,155	0,381	0,434	0,358
2342747	0,652	0,648	0,592	0,092	0,095	0,236	0,348	0,352	0,408
2399332	0,623	0,586	0,594	0,151	0,132	0,206	0,377	0,414	0,406
2830795	0,530	0,526	0,498	0,106	0,085	0,081	0,470	0,474	0,502
2831700	0,609	0,636	0,598	0,210	0,147	0,155	0,391	0,364	0,402
3780962	0,643	0,600	0,653	0,127	0,132	0,114	0,357	0,400	0,347
4364205	0,574	0,590	0,601	0,183	0,095	0,188	0,426	0,410	0,399
4530059	0,609	0,613	0,601	0,201	0,179	0,188	0,391	0,387	0,399
4606077	0,579	0,539	0,583	0,131	0,099	0,119	0,421	0,461	0,417
6444724	0,651	0,650	0,654	0,137	0,132	0,127	0,349	0,350	0,346
6811238	0,619	0,658	0,553	0,192	0,119	0,291	0,381	0,342	0,447
6955448	0,544	0,519	0,542	0,085	0,050	0,081	0,456	0,481	0,458
7041158	0,539	0,612	0,653	0,170	0,154	0,117	0,461	0,388	0,347
8037429	0,613	0,649	0,553	0,201	0,107	0,291	0,387	0,351	0,447
8078417	0,603	0,630	0,601	0,192	0,147	0,188	0,397	0,370	0,399
9905977	0,518	0,516	0,577	0,071	0,147	0,246	0,480	0,484	0,423
9951171	0,547	0,576	0,637	0,308	0,258	0,140	0,453	0,424	0,363
10092491	0,608	0,611	0,583	0,183	0,147	0,179	0,392	0,389	0,417
10488710	0,627	0,582	0,598	0,162	0,236	0,162	0,373	0,418	0,402
10495407	0,608	0,542	0,510	0,095	0,236	0,072	0,392	0,458	0,490
10773760	0,602	0,603	0,563	0,206	0,216	0,206	0,398	0,397	0,437
10776839	0,623	0,582	0,566	0,130	0,236	0,268	0,347	0,418	0,434
12997453	0,598	0,636	0,641	0,216	0,066	0,075	0,402	0,364	0,359
13182883	0,606	0,623	0,638	0,143	0,132	0,152	0,394	0,377	0,362
PROM	<b>0,592</b>	<b>0,578</b>	<b>0,588</b>	<b>0,148</b>	<b>0,155</b>	<b>0,146</b>	<b>0,407</b>	<b>0,425</b>	<b>0,412</b>
PROM 3 regiones	0,586			0,150			0,415		
ACUMULADO POR REGION				2,10852E-77	6,66144E-79	2,14234E-80	1,07719E-35	1,41015E-34	2,05819E-35

## Anexo 9. Indicadores genético poblacionales y genético forenses de las 3 regiones consolidadas

SNP	EQUILIBRIO DE HARDY WEINBERG (H-W)	HETEROCIGOSIDAD ESPERADA	ESTRUCTURA POBLACIONAL (Fst)	CONTENIDO DE INFORMACION POLIMORFICA (PIC)	PODER DE DISCRIMINACION (PD)	PODER DE EXCLUSION (PE)	PROBABILIDAD DE MATCH (PM)
159606	0,4775	0,495	0,017	0,358	0,595	0,821	0,405
214955	0,3343	0,465	-0,001	0,374	0,638	0,813	0,362
221956	0,0895	0,400	-0,006	0,351	0,615	0,824	0,385
251934	0,5041	0,410	-0,003	0,336	0,586	0,832	0,414
279844	0,6779	0,515	-0,002	0,375	0,617	0,813	0,383
321198	0,1073	0,422	-0,005	0,363	0,631	0,818	0,369
338882	0,3632	0,472	0,029	0,375	0,637	0,813	0,363
354439	0,5716	0,469	0,027	0,370	0,629	0,815	0,371
430046	0,8893	0,490	-0,003	0,371	0,621	0,815	0,379
445251	0,0280	0,415	-0,001	0,368	0,643	0,816	0,357
560681	0,1176	0,360	-0,013	0,322	0,569	0,839	0,431
576261	1,0000	0,495	-0,006	0,370	0,617	0,815	0,383
717302	0,0077	0,365	0,003	0,349	0,616	0,826	0,384
719366	0,3882	0,324	0,030	0,283	0,508	0,858	0,492
722098	0,5730	0,480	0,003	0,373	0,630	0,814	0,370
722290	0,4856	0,475	-0,008	0,374	0,634	0,813	0,366
727811	0,3187	0,465	0,002	0,374	0,639	0,813	0,361
729172	0,3670	0,495	-0,008	0,357	0,592	0,822	0,408
733164	0,4679	0,380	-0,004	0,317	0,558	0,842	0,442
735155	0,8860	0,500	0,004	0,374	0,623	0,813	0,377
737681	0,8482	0,395	0,016	0,315	0,553	0,842	0,447
740598	0,0907	0,407	0,040	0,353	0,616	0,824	0,384
740910	0,4579	0,430	0,020	0,351	0,608	0,824	0,392
763869	0,4698	0,449	-0,001	0,363	0,623	0,819	0,377
826472	0,1687	0,370	0,019	0,325	0,572	0,838	0,428
873196	0,0006	0,275	0,036	0,297	0,524	0,852	0,476
876724	0,8293	0,325	0,001	0,275	0,495	0,863	0,505
891700	0,3055	0,440	-0,004	0,361	0,623	0,819	0,377
901398	0,2282	0,510	0,003	0,360	0,591	0,820	0,409
907100	0,5434	0,477	0,024	0,372	0,630	0,814	0,370
914165	0,6671	0,513	0,042	0,374	0,616	0,813	0,384
917118	0,8858	0,490	0,014	0,369	0,619	0,815	0,381
938283	0,0263	0,295	0,005	0,225	0,423	0,887	0,577
964681	0,3982	0,520	0,011	0,367	0,598	0,817	0,402
987640	0,8942	0,495	-0,002	0,374	0,625	0,813	0,375
993934	0,1070	0,355	0,001	0,321	0,567	0,840	0,433
1005533	0,0309	0,420	0,003	0,372	0,649	0,814	0,351
1015250	0,0002	0,293	0,033	0,315	0,556	0,842	0,444
1024116	0,6675	0,470	0,026	0,367	0,622	0,817	0,378
1028528	0,6099	0,390	0,078	0,322	0,565	0,839	0,435
1031825	0,4798	0,411	-0,004	0,341	0,594	0,829	0,406
1058083	0,8908	0,490	-0,006	0,370	0,620	0,815	0,380
1109037	0,2081	0,455	-0,007	0,374	0,643	0,813	0,357
1335873	0,8900	0,485	0,084	0,370	0,622	0,815	0,378
1336071	1,0000	0,500	0,008	0,374	0,624	0,813	0,376
1357617	0,7932	0,292	0,041	0,253	0,461	0,873	0,539

1360288	0,3896	0,305	-0,021	0,271	0,488	0,865	0,512
1382387	0,0069	0,360	0,024	0,348	0,614	0,826	0,386
1413212	0,4030	0,445	0,008	0,333	0,570	0,834	0,430
1454361	0,8848	0,475	0,009	0,364	0,616	0,818	0,384
1463729	0,7644	0,485	0,021	0,373	0,628	0,814	0,372
1490413	0,5734	0,485	-0,003	0,375	0,632	0,813	0,368
1493232	0,2063	0,418	-0,003	0,357	0,621	0,821	0,379
1498553	0,6859	0,482	0,008	0,372	0,627	0,814	0,373
1523537	0,4443	0,430	0,014	0,348	0,601	0,826	0,399
1528460	0,0762	0,424	0,070	0,367	0,638	0,817	0,362
1821380	0,7335	0,505	0,008	0,370	0,613	0,815	0,387
1886510	0,6631	0,480	0,053	0,371	0,627	0,814	0,373
1979255	0,0917	0,435	0,008	0,374	0,649	0,813	0,351
2040411	0,7357	0,410	0,040	0,329	0,574	0,835	0,426
2046361	0,4915	0,465	0,029	0,371	0,633	0,814	0,367
2056277	0,7779	0,260	0,006	0,222	0,414	0,889	<b>0,586</b>
2076848	0,0079	0,307	0,007	0,307	0,544	0,846	0,456
2107612	0,5532	0,350	0,022	0,298	0,531	0,851	0,469
2111980	0,8825	0,485	-0,002	0,370	0,622	0,815	0,378
2269355	0,5619	0,518	-0,007	0,375	0,616	0,813	0,384
2342747	0,0194	0,410	-0,007	0,369	0,647	0,815	0,353
2399332	0,7942	0,465	0,006	0,360	0,611	0,820	0,389
2830795	0,6977	0,365	0,000	0,293	0,522	0,853	0,478
2831700	0,8812	0,490	0,011	0,369	0,619	0,815	0,381
3780962	0,0607	0,417	0,006	0,366	0,638	0,817	0,362
4364205	0,7142	0,465	-0,002	0,349	0,591	0,826	0,409
4530059	0,5682	0,505	-0,002	0,368	0,609	0,816	0,391
4606077	0,8608	0,409	0,003	0,328	0,572	0,836	0,428
6444724	0,0323	0,430	-0,007	0,375	0,652	0,813	0,348
6811238	0,8848	0,505	0,004	0,375	0,622	0,813	0,378
6955448	0,0854	0,325	-0,005	0,303	0,539	0,848	0,461
7041158	0,6185	0,456	0,054	0,358	0,612	0,821	0,388
8037429	0,6652	0,505	-0,005	0,372	0,616	0,814	0,384
8078417	0,7627	0,490	-0,007	0,365	0,610	0,818	0,390
9905977	1,0000	0,410	0,026	0,322	0,561	0,839	0,439
9951171	0,0602	0,560	-0,002	0,373	0,586	0,814	0,414
10092491	0,7706	0,482	0,001	0,360	0,604	0,820	0,396
10488710	0,7682	0,490	0,001	0,366	0,612	0,817	0,388
10495407	0,4105	0,400	0,022	0,334	0,584	0,833	0,416
10773760	0,2904	0,523	0,007	0,366	0,596	0,817	0,404
10776839	0,7749	0,495	0,014	0,375	0,627	0,813	0,373
12997453	0,0799	0,423	-0,002	0,366	0,637	0,817	0,363
13182883	0,4512	0,443	-0,005	0,361	0,622	0,819	0,378
<b>PROM</b>	<b>0,487</b>	<b>0,436</b>	<b>0,011</b>	<b>0,349</b>	<b>0,598</b>	<b>0,825</b>	<b>0,402</b>
<b>ACUMULADO</b>					<b>1-3,8355E-36</b>	<b>3,80011E-08</b>	<b>3,8355E-36</b>
						<b>0,9999999620</b>	