

UNIVERSIDAD
NACIONAL
DE COLOMBIA

Método para identificar variantes del ADN mitocondrial humano secuenciado por la plataforma Nanopore en comparación con Illumina

Diana Katherinne Garcia Garay

Universidad Nacional de Colombia
Facultad de Ingeniería, Departamento de ingeniería industrial y sistemas
Bogotá, Colombia
2025

Método para identificar variantes del ADN mitocondrial humano secuenciado por la plataforma Nanopore en comparación con Illumina

Diana Katherine Garcia Garay

Trabajo de grado presentado como requisito parcial para optar al título de:
Magister en Bioinformática

Directora:

PhD., Bióloga Clara Isabel Bermúdez Santana

Codirectora:

Bact., M.Sc., Angela Johanna Muñoz Bolaños

Línea de Investigación:

Bioinformática funcional y estructural

Grupo de Investigación:

RNómica Teórica y Computacional

Universidad Nacional de Colombia

Facultad de ingeniería, Departamento de ingeniería industrial y sistemas

Bogotá, Colombia

2025

Para todo aquel que necesite una voz de aliento, siempre hay una forma y todo llega en su debido momento.

Agradecimientos

En primer lugar, a mi familia, mis seres queridos, amigos y compañeros tanto de la maestría como del trabajo y otros entornos, que siempre han estado en algún momento para brindarme su apoyo desde su conocimiento y experiencia.

A la Universidad Nacional de Colombia que como patrimonio de la nación nos atraviesa a todos los colombianos, y que a mí en particular me ha brindado el espacio para crecer personal y profesionalmente, y tener en mente que la academia tiene un propósito tanto con la comunidad como con uno mismo. Así mismo al Centro de Excelencia en Computación científica por el apoyo en el uso de la computación requerida para la prueba del modelo.

Finalmente a los laboratorios en los que he trabajado, al personal (profesores, líderes, auxiliares, compañeros y colegas) que me han brindado su apoyo para realizar este trabajo.

Resumen

Título en español: Método para identificar variantes del ADN mitocondrial humano secuenciado por las plataformas Nanopore en comparación con Illumina

La secuenciación del ADN mitocondrial (ADNmt) es fundamental para el diagnóstico molecular y el estudio de la diversidad genética. En este trabajo se desarrolló y evaluó un método bioinformático para identificar variantes en ADNmt humano secuenciado mediante plataformas Oxford Nanopore Technologies® (ONT) e Illumina, Inc.®. Se analizaron cinco muestras independientes de ADNmt, considerando réplicas técnicas para ONT, y se aplicaron múltiples combinaciones de alineadores (Minimap2, NGMLR, Winnowmap2) y llamadores de variantes (Mutect2, Medaka, Mutserve2, Clair3). Las métricas comparadas incluyeron calidad de lectura, profundidad de cobertura, detección de heteroplasmia y concordancia entre plataformas. ONT alcanzó coberturas $\geq 200\times$, con valores modales de calidad Q20–Q30, mientras que Illumina mostró coberturas superiores a $1000\times$ y calidad $\geq Q30$. A nivel de variantes, ambas plataformas coincidieron en más del 90% de los SNVs tras filtrado, con un índice de concordancia de Jaccard promedio de 0.85–0.95. Illumina presentó mayor sensibilidad para heteroplasmas bajas ($\geq 3\text{--}5\%$), mientras que ONT fue robusto para heteroplasmas $\geq 10\%$ y permitió detectar variantes estructurales y deleciones. Entre los pipelines probados, la combinación Minimap2–Clair3 mostró el mejor equilibrio entre precisión, sensibilidad y reproducibilidad en ONT. Se concluye que, aunque Illumina mantiene ventaja en variantes de baja frecuencia, ONT constituye una alternativa competitiva y complementaria, ofreciendo además lecturas largas y análisis en tiempo real.

Palabras clave: ADN mitocondrial, secuenciación Nanopore, Illumina, heteroplasmia, llamado de variantes, bioinformática

Abstract

Method to Identify Human Mitochondrial DNA Variants Sequenced with Nanopore and Illumina Platforms

Abstract:

Mitochondrial DNA (mtDNA) sequencing plays a crucial role in molecular diagnostics and population genetics. In this study, we developed and evaluated a bioinformatic pipeline to identify mtDNA variants using Oxford Nanopore Technologies® (ONT) e Illumina, Inc.® platforms. Five independent human mtDNA samples were analyzed, with technical replicates for ONT, and multiple aligner–caller combinations were tested (Minimap2, NGMLR, Winnowmap2; Mutect2, Medaka, Mutserve2, Clair3). Key metrics included read quality, coverage depth, heteroplasmy detection, and cross-platform concordance. ONT achieved $\geq 200\times$ depth with modal base quality Q20–Q30, while Illumina reached $>1000\times$ depth and base quality $\geq Q30$. Variant comparison showed $>90\%$ concordance of SNVs after filtering, with an average Jaccard index of 0.85–0.95. Illumina demonstrated higher sensitivity for low heteroplasmy levels ($\geq 3\text{--}5\%$), whereas ONT was consistent for heteroplasmy $\geq 10\%$ and uniquely enabled structural variant and deletion detection. Among tested pipelines, the Minimap2–Clair3 combination provided the best balance of accuracy, sensitivity, and reproducibility for ONT data. We conclude that although Illumina retains an advantage in detecting low-frequency variants, ONT represents a competitive and complementary approach, adding the benefits of long reads and real-time sequencing capabilities.

Keywords: Mitochondrial DNA, Nanopore sequencing, Illumina, heteroplasmy, variant calling, bioinformatics

Este Trabajo Final de maestría fue calificado en **enero** de 2026 por el(la) siguiente evaluador(a):

CLAUDIA MARINA MUÑOZ DÍAZ
Profesor Facultad de ciencias. Instituto de Biotecnología
Universidad Nacional de Colombia, Sede Bogotá

Contenido

	Pág.
Resumen	VII
Lista de figuras	XII
Lista de tablas	XIV
Introducción	1
1. Contexto biológico del ADN mitocondrial y secuenciamiento.....	5
1.1 Generalidades de las mitocondrias	6
1.1.1. Estructura y organización del genoma mitocondrial humano.....	8
1.1.2. Implicaciones clínicas y enfermedades asociadas al ADNmt	10
1.1.3. Perspectivas futuras en investigación clínica mitocondrial	14
1.2. Secuenciación	15
1.2.1. Panorama histórico de la secuenciación	16
1.2.2. La revolución de la segunda generación: Secuenciación de Nueva Generación (NGS) o masiva en paralelo.....	18
1.2.3. Secuenciación de tercera generación: Secuenciación de molécula única en tiempo real.....	20
1.2.4. Métricas asociadas a la secuenciación.....	23
1.2.4.1. La Precisión	23
1.2.4.2. Longitud de lectura y el ensamblaje.....	25
1.2.5. Herramientas de análisis para secuenciación ONT	32
1.2.8. Secuenciación y su aplicabilidad clínica	62
2. Descripción del método para la detección de variantes del ADN mitocondrial por plataformas de secuenciación masiva (ONT).....	70
2.1. Flujo de trabajo bioinformático aplicado	70
2.1.1. Información biológica mitocondrial.....	71
2.1.2. Paso 0. Secuenciamiento del ADNmt y datos Illumina	72
2.1.3. Paso 1 <i>Basecalling</i> con Dorado (MinKNOW)	73
2.1.4. Paso 2 Alineamiento de lecturas al genoma mitocondrial humano de referencia.....	74
2.1.5. Paso 3: Análisis comparativo de herramientas de llamado de variantes en ADN mitocondrial.....	75

2.2.	Construcción de matriz y evaluación de pipelines	77
2.2.1.	Determinación del mejor <i>mapper</i>	78
2.2.1.1.	Profundidad de cobertura y calidad de mapeo	78
2.2.2.	Determinación de la mejor combinación <i>mapper-caller</i>	80
2.3.	Comparación con resultados de Illumina	86
2.3.1.	Comparación de presencia/ausencia de variantes.....	86
2.3.2.	Comparación de niveles de heteroplasmia.....	86
2.3.3.	Métricas básicas de calidad y profundidad	87
2.3.4.	Distancias y agrupamiento. Análisis de Componentes Principales (PCA)	90
2.3.5.	Visualización con circos y análisis de concordancia.	92
3.	Discusión selección de método y aplicabilidad clínica.....	101
3.1.	Discusión de la implementación y selección de método.....	101
3.2.	Comparación Illumina vs ONT	104
4.	CONSIDERACIONES FINALES	109
A.	Anexo 1: Tabla con datos comparativos de las variantes encontradas en cada plataforma.....	115
B.	Anexos 2-5: Graficos de circos plot para las muestras 3258, 3312, 3396 y 9093.	139
	Bibliografía	145

Lista de figuras

	Pág.
Figura 1. Mapa circular del genoma mitocondrial humano. En el ADNmt de 16,6 kb se identifican 37 genes: 13 genes proteicos (subunidades de complejos de la cadena respiratoria), 2 genes de ARNr (12S y 16S) y 22 genes de ARNt. La molécula es de doble hebra circular, diferenciada en hebra pesada (H, rica en guaninas) y hebra ligera (L, rica en citosinas). La región no codificante principal es la región control o D-loop ($\approx 1,1$ kb), donde se localizan los promotores de transcripción de la hebra H (HSP) y L (LSP), así como el origen de replicación de la hebra pesada (Ori_H), mientras que el origen de replicación de la hebra ligera (Ori_L) se encuentra en un locus separado. Imagen construida por la autora en base al genoma de referencia NC_012920.1.....	8
Figura 2. Diagrama de flujo del análisis realizado en este trabajo.	71
Figura 3. Comparación de mappers para ONT (>10 kb): profundidad media, calidad media y uniformidad de cobertura (CV). Paneles resumen por <i>mapper</i> (Minimap2, Winnowmap2, NGMLR). (A) profundidad media; (B) calidad media; (C) coeficiente de variación (CV) de la profundidad como medida de uniformidad. En nuestro conjunto, Minimap2 muestra la menor variabilidad de cobertura (CV más bajo), con valores competitivos de profundidad y calidad. Ver Métodos para parámetros de cada <i>mapper</i>	79
Figura 4. Profundidad (DP) por combinación <i>caller-mapper</i> . Diagramas de caja (y puntos) del AF de las variantes agrupadas por pipeline (p. ej., <i>clair3.minimap2</i> , <i>mutect2.winnowmap2</i> , etc.), aquí para la profundidad de lectura asociada a cada llamada.....	80
Figura 5. AF por combinación <i>caller-mapper</i> . Diagramas de caja (y puntos) del AF de las variantes agrupadas por pipeline (p. ej., <i>clair3.minimap2</i> , <i>mutect2.winnowmap2</i> , etc.). Cada caja resume el comportamiento global del par <i>caller-mapper</i> a través de muestras/llamadas.....	82
Figura 6. Similitud media (Jaccard) entre combinaciones <i>caller-mapper</i> . Mapa de calor de los coeficientes de Jaccard calculados sobre presencia/ausencia de variantes por posición (sin usar niveles de AF), promediados por combinación en las muestras donde ambas combinaciones están disponibles. Escala 0–1 (amarillo = mayor coincidencia). Las filas/columnas se agrupan por similitud (clustering jerárquico).	83
Figura 7. PCA de combinaciones <i>caller-mapper</i> según perfiles de variantes. Nota. Análisis de componentes principales (PC1–PC2) sobre la matriz binaria de presencias	

por posición (unión de posiciones) para cada combinación. Cada punto es una combinación “caller.mapper” (color por *caller*)..... 85

Figura 8. Profundidad de lectura por plataforma (log10 DP). Diagramas de caja del log10 de la profundidad (DP) de las variantes detectadas con Illumina y ONT. El uso de escala logarítmica facilita comparar distribuciones cuando hay coberturas altas. 88

Figura 9. Comparación de la calidad de base (BQ) entre tecnologías de secuenciación. Diagrama de caja de los valores de calidad de base (BQ, Phred score promedio por posición) para lecturas generadas con plataformas Illumina y ONT. 89

Figura 10. Distribución de la heteroplasma (AF) por plataforma. Diagramas de caja con valores puntuales (jitter) de AF de las variantes detectadas por muestra en Illumina (ILL) y ONT. Las cajas muestran IQR, la línea central la mediana y los bigotes el rango fuera de los cuartiles. El eje Y está expresado en porcentaje de alelo alternativo. Cada punto corresponde a una llamada de variante. 90

Figura 11. Separación por plataforma en un análisis de componentes principales (PCA). *Nota.* Dispersograma PC1 vs PC2 de las llamadas de variantes; colores por plataforma. Las elipses representan regiones de dispersión (95%) por grupo. La separación entre grupos sugiere diferencias sistemáticas en el perfil conjunto de métricas (p. ej., AF, DP, QUAL_eff) entre ONT e Illumina. 91

Figura 12. Gráfico de circos para la muestra 9251: cobertura, calidad y variantes ONT vs Illumina. Anillo externo: escala de posición del ADNmt (0–16.6 kb). Anillos de Profundidad (línea azul continua) y calidad media Phred (línea roja continua) por posición. Se muestran sus ejes radiales para referencia. Variantes como puntos: verde si $AF \geq 0.90$, rojo si $AF < 0.90$. Capa para ONT y otra para Illumina para las variantes encontradas para cada plataforma (las etiquetas de Illumina se orientan hacia el interior y las de ONT hacia el exterior). Banderines (flags) de discrepancia por posición, marcas con línea y etiqueta solo cuando una plataforma detecta una posición que la otra no reporta (discordancia por presencia/ausencia). 93

Figura 13. Concordancia entre ONT e Illumina por muestra (ADNmt). El mapa de calor resume cuatro métricas calculadas a nivel de posición (CHROM:POS): $Agree \leq \Delta AF$ (comunes) = proporción de posiciones compartidas cuya diferencia absoluta de fracción alélica/heteroplasma (AF) es ≤ 0.05 ; $Jaccard (\leq \Delta AF)$ = índice de Jaccard considerando coincidencia solo cuando ambas plataformas reportan la posición y $|\Delta AF| \leq 0.05$; $Jaccard$ (presencia) = Jaccard basado únicamente en la intersección/unión de posiciones reportadas (sin considerar niveles de AF); $Weighted Jaccard = \frac{min(AF_{ONT}, AF_{ILL})}{max(AF_{ONT}, AF_{ILL})}$ en la unión de posiciones. Los valores dentro de cada celda son porcentajes y el color (claro→oscuro) indica concordancia baja→alta. .. 95

Figura 14. Relación entre heteroplasma (AF) y profundidad (log10 DP). *Nota.* Cada punto es una llamada de variante; color por plataforma. El eje X muestra AF (%), el eje Y la profundidad en log10. La nube de puntos ilustra cómo cambia la calidad/estabilidad aparente de las llamadas con la cobertura. 96

Figura 15. Mapa bidimensional (biplot) del análisis de correspondencia múltiple realizado sobre las variantes encontradas en común entre las plataformas de ONT e Illumina, para las 5 muestras analizadas. 98

Lista de tablas

Pág.

Tabla 1. Comparación general entre tecnologías de secuenciación de primera (Sanger), segunda (Illumina) y tercera generación (ONT).....	31
Tabla 2. Comparación detallada de diversas herramientas bioinformáticas utilizadas para el análisis de variantes en ADN mitocondrial secuenciado mediante plataformas de lectura larga como ONT. Se evalúan tanto <i>mappers</i> como <i>callers</i> de variantes en función de su precisión, sensibilidad, compatibilidad con ONT, facilidad de uso, requerimientos computacionales y capacidad para detectar heteroplasmia o variantes estructurales (SVs). Esta síntesis permite identificar las combinaciones más empleadas para estudios mitocondriales, considerando aspectos clave como la detección de variantes de baja frecuencia, la eficiencia operativa y la resolución en regiones complejas del genoma. La comparación sirve como base para seleccionar pipelines óptimos en contextos clínicos o de investigación genómica avanzada.	41
Tabla 3. Tabla con características clave y parámetros clave de los alineadores usados.	75
Tabla 4. Características y parámetros clave de las herramientas de llamado de variantes usadas en el trabajo.	77
Tabla 5. Número de variantes encontradas por cada plataforma para las 5 muestras utilizadas.	92
Tabla 6. Resumen de pruebas estadísticas comparando plataformas Illumina vs ONT. Se indican el tipo de prueba, el estadístico calculado (y grados de libertad cuando aplican), el p-valor, una medida de tamaño de efecto (d de Cohen para pruebas bivariadas; R^2 para PERMANOVA) y una interpretación concisa del resultado en términos técnicos.	99
Tabla 7. Tabla con los datos de métricas de las variantes encontradas con cada plataforma y para cada muestra.	138

Introducción

Las mitocondrias son organelos esenciales en diferentes procesos celulares, principalmente relacionados con metabolismo y producción de energía. Debido a la naturaleza propia de estos organelos, el modo de herencia es monoparental (heredados únicamente por línea materna, en la mayoría de los casos) y además tiene muy poca recombinación.

Estos organelos son considerados como semiautónomos dado que poseen su propio material genético o genoma conocido como ADN mitocondrial (ADNmt) (Legros et al., 2004). El ADNmt humano está compuesto por 37 genes, 13 de ellos producen proteínas involucradas en la cadena respiratoria, 22 son ARNts y 2 ARNrs involucrados en la producción de polipéptidos (Stefano et al., 2017). Tiene un tamaño de 16.569 pares de bases en humanos.

La aparición de mutaciones en el ADNmt está dirigida principalmente por dos procesos, uno de ellos es el envejecimiento del organismo que, en un proceso natural, genera la acumulación de mutaciones dada las características propias de la mitocondria (acumulación de ROS y replicación imperfecta) y, por otra parte, un proceso de heteroplasmia, en el que dentro de cada célula se encuentra copias diferentes de ADNmt (por lo que se pueden encontrar copias tipo silvestre y otras acarreado mutaciones asociadas a enfermedades) (Gorman et al., 2016; Stefano et al., 2017).

Variantes en el ADNmt se encuentran asociadas a múltiples enfermedades de diferentes sistemas en el organismo, incluyendo al sistema nervioso (con enfermedades como el Alzheimer, Parkinson, Ataxias y Epilepsias), cardiomiopatías, fallos respiratorios, Síndrome de Fanconi, Diabetes mellitus, miopatías, entre muchas otras, donde el

funcionamiento erróneo de las mitocondrias repercute en este tipo de enfermedades (Gorman et al., 2016).

Debido a la alta presencia de mutaciones y la heteroplasmia que se puede presentar, determinar el genoma mitocondrial y las variantes mediante secuenciación Sanger representa un gran desafío, debido a la menor sensibilidad que presenta la técnica para determinar niveles de heteroplasmia por un lado (Legati et al., 2021), y los costos que puede derivarse para el secuenciamiento de múltiples muestras. Sin embargo, con la implementación de secuenciación de nueva generación, determinar las posibles variantes y generar consensos para posible heteroplasmia en un organismo, se ha venido resolviendo gracias a la profundidad y cobertura que se logra a través de las diferentes plataformas (Keogh & Chinnery, 2015; Legati et al., 2021) y al multiplexado que permite incrementar la profundidad y eficiencia del uso masivo de las tecnologías de secuenciamiento de nueva generación.

Dentro de estas tecnologías, la tecnología de secuenciación por Illumina, Inc., ha sido fundamental en el avance de la genómica, permitiendo la lectura de secuencias de ADN con una alta precisión y a gran escala. Al igual que la tecnología de secuenciación Sanger, la tecnología de Illumina utiliza un enfoque de secuenciación por síntesis, donde se generan fragmentos de ADN que luego son amplificados y secuenciados en paralelo (Liu et al., 2012).

Esta plataforma ha evolucionado considerablemente desde su introducción, con sistemas como MiSeq, HiSeq y NovaSeq, cada uno con capacidades y escalabilidad diferentes para adaptarse a las necesidades de investigación de diversos laboratorios y proyectos. La versatilidad de estas plataformas ha permitido su aplicación en áreas que van desde la investigación básica hasta el apoyo al diagnóstico clínico, incluyendo el estudio de la variabilidad genética en poblaciones humanas, la identificación de biomarcadores para enfermedades y la secuenciación de genomas completos (Shendure & Ji, 2008).

A pesar de sus numerosas ventajas, la tecnología de Illumina no está exenta de desafíos. Uno de los principales obstáculos es la generación de lecturas de secuencia con longitudes limitadas (80-300pb), lo que puede dificultar el ensamblaje y análisis de regiones genómicas repetitivas o complejas. Además, el proceso de amplificación previo a la

secuenciación puede introducir sesgos y errores, especialmente en regiones con alto contenido de GC o AT (Amarasinghe et al., 2020a).

Recientemente, la tecnología de secuenciación Nanopore de la empresa Oxford Nanopore Technologies® (ONT), basada en el uso de nanoporos para el análisis de moléculas de ADN o ARN a nivel individual (Y. Wang et al., 2021), ofrece un enfoque innovador y prometedor para el análisis de moléculas de ADN y ARN a nivel individual. Su capacidad para leer secuencias largas de manera continua, sin la necesidad de amplificación previa de las muestras, representa una ventaja significativa sobre tecnologías establecidas como Illumina. Esta capacidad de lectura prolongada no solo simplifica la preparación de la muestra y reduce los costos asociados, sino que también minimiza la probabilidad de errores de secuenciación comunes en técnicas de secuenciación por síntesis (Petersen et al., 2019).

El proceso de llamado de bases en la secuenciación Nanopore es crucial y está sujeto a constante optimización. Las señales eléctricas generadas durante la secuenciación no producen un patrón único e inequívoco para cada nucleótido, además de que la tasa de traspaso de nucleótidos a través del nanoporo no es constante (Zhang et al., 2020). Hoy en día, los algoritmos diseñados para el llamado de bases aun representan un área de investigación dinámica y en evolución, impulsada por avances en nuevas tecnologías de predicción, inteligencia artificial, aprendizaje profundo y aprendizaje automático. Sin embargo, debido a la frecuente actualización y variabilidad de estos algoritmos, es complicado alcanzar un consenso estable para el análisis de información genética (Peresini et al., 2021; Wick et al., 2019).

La tecnología de secuenciación Nanopore ha demostrado un potencial significativo en aplicaciones clínicas y de investigación, particularmente en el análisis del ADN mitocondrial. En un estudio llevado a cabo en un hospital de Nueva Zelanda, se evaluó la viabilidad y precisión de la secuenciación Nanopore para el análisis genómico rápido en entornos clínicos, en la secuenciación de genoma completo. Este estudio destacó la capacidad de la plataforma para proporcionar datos genómicos confiables con tiempos de procesamiento significativamente reducidos, lo que la hace especialmente útil para aplicaciones en diagnósticos rápidos y medicina personalizada (Gorzynski et al., 2022; Nyaga et al., 2024).

Asimismo, una investigación realizada en Turquía utilizó Nanopore para el diseño y validación de un ensayo específico enfocado en el gen de la Biotinidasa (BTD), involucrado en trastornos metabólicos. Este trabajo demostró la robustez de la tecnología en la identificación de variantes genéticas raras y frecuencias de alelos en una población específica. Los resultados subrayaron la capacidad de Nanopore para identificar variantes con alta precisión, incluso en regiones genómicas complicadas como aquellas con repeticiones o contenido de GC elevado, características comunes en el ADN mitocondrial (Gorzynski et al., 2022; Kazan et al., 2024; Santibanez-Koref et al., 2019).

Con base en lo anterior, en este trabajo de grado se propuso diseñar y evaluar un flujo de trabajo bioinformático para la identificación de variantes del ADN mitocondrial humano secuenciado mediante ONT cuya metodología fue desarrollada por la autora, comparando los resultados obtenidos con los derivados de Illumina. Estos datos corresponden a secuenciación retrospectiva, generada en un momento previo y bajo condiciones experimentales y bioinformáticas distintas a las empleadas para Oxford Nanopore. Por tanto, los resultados Illumina se utilizan como una referencia comparativa y no como un estándar absoluto. La investigación buscó responder a la pregunta central: *¿puede ONT, bajo condiciones controladas de análisis, generar resultados equivalentes a los de Illumina en la detección de variantes mitocondriales relevantes?*. Para ello, se definieron los siguientes objetivos: (i) determinar las herramientas de análisis más adecuadas para el procesamiento de los datos derivados de cada plataforma; (ii) analizar las variables resultantes de la secuenciación y anotación para la construcción de vectores característicos del ADN mitocondrial; y (iii) computar las relaciones entre dichos vectores a fin de establecer la concordancia entre ambas plataformas. En los siguientes capítulos se detalla la información obtenida y los análisis y conclusiones del estudio.

1. Contexto biológico del ADN mitocondrial y secuenciamiento.

La mitocondria es un organelo esencial presente en casi todas las células eucariotas, conocido por su papel en la producción de energía química en forma de ATP. A diferencia de otros organelos, las mitocondrias poseen su propio material genético un pequeño genoma circular de ADN de doble hebra, resultado de su origen evolutivo mediante un evento endosimbiótico entre un antepasado de las α -proteobacterias y una célula huésped primitiva (Rackham & Filipovska, 2022). Durante este proceso, ocurrido hace aproximadamente 1.5–2 mil millones de años, la protomitocondria fue incorporada a la célula eucariota y retuvo un genoma propio aunque redujo drásticamente su tamaño al transferir la mayor parte de sus genes al núcleo de la célula huésped (Neupert, 2016).

El genoma mitocondrial humano resultante tiene solo 16.569 pares de bases y codifica apenas 37 genes: 13 subunidades proteicas de la cadena respiratoria, 22 ARN de transferencia (ARNt) y 2 ARN ribosómicos (ARNr). Esto contrasta con los ~20.000 genes del genoma nuclear humano y refleja la dependencia de las mitocondrias de la célula anfitriona. A pesar de su diminuto tamaño, el ADNmt juega un papel crítico en la fisiología celular y su alteración tiene importantes consecuencias clínicas (Ott et al., 2016).

Las mutaciones heredadas o somáticas en el ADNmt están implicadas en numerosas enfermedades humanas, en particular un grupo de trastornos conocidos como enfermedades mitocondriales, caracterizados por disfunción de la fosforilación oxidativa y que suelen afectar tejidos de alta demanda energética (cerebro, músculo) (Kopinski et al., 2021). Además, evidencia emergente sugiere que las variaciones y daños acumulados en

el ADNmt podrían contribuir a procesos complejos como el envejecimiento, la neurodegeneración e incluso la patogénesis de ciertos cánceres (Galizzi & Di Carlo, 2023; Kopinski et al., 2021; Schon et al., 2012).

Este panorama ha motivado un interés creciente en el análisis bioinformático de variantes mitocondriales, con el fin de comprender mejor su impacto funcional y ayudar al diagnóstico y manejo de dichas afecciones (Craven et al., 2025). En este capítulo se revisan los fundamentos biológicos y clínicos del ADN mitocondrial, se presentan las generalidades estructurales de la mitocondria y de su genoma (organización, contenido génico, herencia), los mecanismos de expresión génica mitocondrial y su regulación, la estrecha interacción núcleo–mitocondria que coordina el funcionamiento de ambos genomas, las funciones bioenergéticas y metabólicas de las mitocondrias, las implicaciones clínicas de las alteraciones del ADNmt (incluyendo ejemplos de enfermedades asociadas) y, finalmente, las perspectivas futuras en la investigación clínica mitocondrial.

1.1 Generalidades de las mitocondrias

Las mitocondrias derivan evolutivamente de bacterias endosimbiontes incorporadas a células ancestrales, lo que explica que conserven un genoma propio y ciertas similitudes con procariontes (Gabaldón, 2025). Poseen doble membrana, una membrana externa lisa y una membrana interna altamente plegada formando crestas (*cristae*) que aumentan la superficie donde ocurren las reacciones de producción de ATP. Entre ambas membranas se encuentra el espacio intermembrana, y el interior conforma la matriz mitocondrial, compartimiento que contiene enzimas del ciclo de Krebs, ribosomas mitocondriales y copias del ADNmt (Geiger et al., 2023).

La membrana externa es permeable a pequeñas moléculas gracias a porinas, mientras que la membrana interna es rica en proteínas de transporte y componentes de la cadena respiratoria, siendo selectivamente permeable y sede de la fosforilación oxidativa (Geiger et al., 2023; Ott et al., 2016). Esta organización compartimental refleja la integración de la mitocondria a la célula eucariota, donde las proteínas codificadas por el ADN nuclear deben ser importadas a través de complejos translocadores (TOM/TIM) para cumplir

funciones dentro de la mitocondria (Moura et al., 2025; Needs et al., 2021). La función clásica de las mitocondrias es actuar como centrales energéticas de la célula, generando la mayor parte del ATP mediante la respiración celular aeróbica. Sin embargo, cumplen además múltiples roles metabólicos y señalizadores, como en la homeostasis del calcio, capturando Ca^{++} del citosol y liberándolo según la necesidad, lo que modula diversas rutas de señalización celular (Moura et al., 2025).

Son una fuente importante de especies reactivas de oxígeno (ROS) como subproductos de la cadena transportadora de electrones, las cuales en niveles controlados actúan como moléculas señalizadoras, aunque en exceso pueden dañar componentes celulares (Rackham & Filipovska, 2022). Las mitocondrias también desempeñan un papel central en la apoptosis intrínseca en respuesta a señales de daño o estrés, liberando al citosol proteínas proapoptóticas (por ejemplo, citocromo c) que activan la cascada de caspasas y llevan a la muerte celular programada (Kopinski et al., 2021).

La expresión génica del ADN mitocondrial se encuentra estrechamente regulada por una maquinaria especializada que permite la transcripción y posterior traducción de los genes codificados en el genoma mitocondrial (Rackham & Filipovska, 2022). Esta maquinaria es parcialmente dependiente del núcleo celular, ya que muchas de las proteínas necesarias para la replicación, transcripción y mantenimiento del ADN mitocondrial se codifican en el genoma nuclear y se importan posteriormente a la mitocondria. La transcripción mitocondrial es policistrónica, lo que significa que se generan largos transcritos que luego son procesados para generar ARNs mensajeros, ribosomales y de transferencia (Gustafsson et al., 2016).

Adicionalmente, intervienen en la síntesis de biomoléculas (por ejemplo, en la formación de grupos hemo y de clústeres hierro-azufre) y en la termogénesis, como en el tejido adiposo pardo, donde la proteína desacoplante UCP1 en la membrana interna mitocondrial permite disipar el gradiente de protones como calor en lugar de producir ATP, contribuyendo a la generación de calor corporal.

Las mitocondrias no son estructuras estáticas; experimentan procesos de fusión y fisión que les permiten intercambiar contenidos, replicarse o eliminarse selectivamente (mediante mitofagia). Su número por célula varía ampliamente según el tipo celular y sus

requerimientos energéticos, desde unas pocas en linfocitos hasta cientos o miles en miocitos cardíacos o fibras musculares esqueléticas (Moura et al., 2025). Asimismo, cada mitocondria puede contener de 2 a 10 copias de su ADNmt, de modo que una célula humana típica alberga múltiples miles de genomas mitocondriales en total. Esta capacidad de tener múltiples copias, junto con la localización en nucleoides (complejos ADN-proteína en la matriz), influye en la genética mitocondrial y en fenómenos como la heteroplasmia (coexistencia de genomas distintos, discutida más adelante) (Stewart & Chinnery, 2015).

1.1.1. Estructura y organización del genoma mitocondrial humano

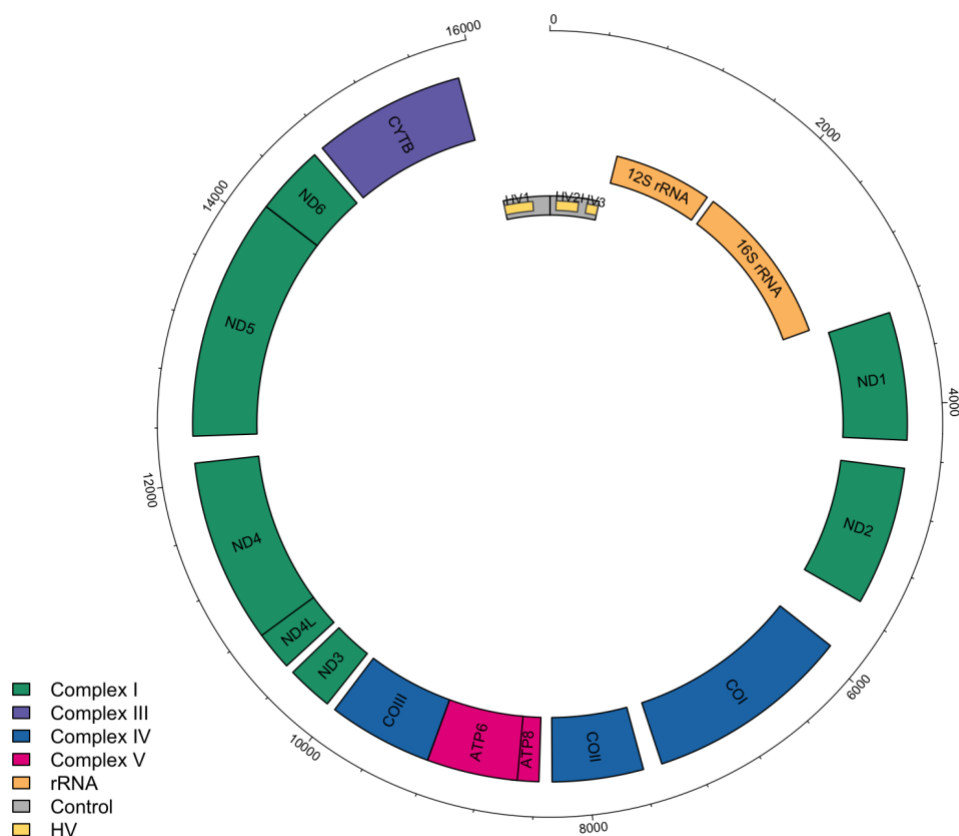


Figura 1. Mapa circular del genoma mitocondrial humano. En el ADNmt de 16,6 kb se identifican 37 genes: 13 genes proteicos (subunidades de complejos de la cadena respiratoria), 2 genes de ARNr (12S y 16S) y 22 genes de ARNt. La molécula es de doble hebra circular, diferenciada en hebra pesada (H, rica en guaninas) y hebra ligera (L, rica en citosinas). La región no codificante principal es la región control o D-loop ($\approx 1,1$ kb), donde se localizan los promotores de transcripción

de la hebra H (HSP) y L (LSP), así como el origen de replicación de la hebra pesada (Ori_H), mientras que el origen de replicación de la hebra ligera (Ori_L) se encuentra en un locus separado. Imagen construida por la autora en base al genoma de referencia NC_012920.1.

El genoma mitocondrial humano se caracteriza por su compactación extrema, a diferencia del genoma nuclear, casi la totalidad del ADNmt es secuencia codificante, sin intrones ni extensas regiones intergénicas. Aproximadamente, el 97% de la secuencia mitocondrial corresponde a genes o partes de genes, comparado con solo ~7% del genoma nuclear dedicado a codificar proteínas. Incluso existen solapamientos entre algunos genes (por ejemplo, ATP8/ATP6 y ND4L/ND4 comparten nucleótidos adyacentes) y ciertos genes proteicos carecen de codón de terminación completo, utilizándose señales de poliadenilación post-transcripcionales para finalizar la traducción (Nissanka & Moraes, 2020; Parakatselaki & Ladoukakis, 2021).

La organización génica es bicatenaria, de los 37 genes mitocondriales, 28 están codificados en la hebra pesada (H) y 9 en la hebra ligera (L). La hebra H codifica la mayoría de las proteínas (12 de 13, excepto ND6), así como 14 tRNAs y los 2 rRNAs, mientras que la hebra L codifica solo un polipéptido (ND6) y 8 tRNAs. Esta distribución refleja en parte la diferente composición de bases de cada hebra y la nomenclatura “pesada” y “ligera” dada su distinta densidad por contenido de G vs C (Fig.1.) (Parakatselaki & Ladoukakis, 2021).

Otra característica clave es que el ADNmt es multicopia y de herencia materna, por ejemplo, un ovulo aporta al embrión en desarrollo decenas de miles de copias de ADNmt presentes en sus numerosas mitocondrias, mientras que los pocos ADNmt del espermatozoide normalmente se eliminan o diluyen tras la fecundación, por lo que se habla de que la transmisión del genoma mitocondrial ocurre casi exclusivamente por vía materna en la especie humana (Craven et al., 2025).

Adicionalmente, durante la ovogénesis ocurre un fenómeno de “cuello de botella mitocondrial” en el que se restringe la cantidad de moléculas de ADNmt que pasan a cada ovocito, seguido por la amplificación de estas copias en el desarrollo temprano. Este cuello de botella puede provocar que la proporción de variantes de ADNmt (si las hubiera) difiera

ampliamente entre madre e hijos, complicando la predicción hereditaria de mutaciones (como se detalla en la sección 1.7) (Craven et al., 2025).

1.1.2. Implicaciones clínicas y enfermedades asociadas al ADNmt

Las enfermedades mitocondriales primarias son un grupo de trastornos genéticos causados por mutaciones en genes mitocondriales (o en genes nucleares que afectan la función mitocondrial) y se caracterizan típicamente por disfunción en la producción de energía (Parakatselaki & Ladoukakis, 2021; Rossmann et al., 2021). Debido a la herencia materna del ADNmt, estas enfermedades suelen transmitirse por línea materna, aunque la presentación en la descendencia puede variar ampliamente por el fenómeno del cuello de botella genético y la heteroplasmia (Gabaldón, 2025; Parakatselaki & Ladoukakis, 2021).

La heteroplasmia se define como la presencia de una mezcla de ADNmt mutado y no mutado en la misma célula, que actúa como un factor crucial que determina la manifestación de la enfermedad. En la mayoría de los casos, una mutación en el ADNmt debe alcanzar un porcentaje umbral dentro de la población de genomas mitocondriales de la célula para producir un déficit bioenergético suficiente y, por ende, síntomas clínicos. Este umbral bioquímico suele oscilar entre ~60% y 90% de genomas mutantes, dependiendo de la mutación específica, del tejido afectado y de la demanda energética del mismo (Parakatselaki & Ladoukakis, 2021; Stewart & Chinnery, 2015).

Por ejemplo, mutaciones severas en genes de ARNt pueden requerir >90% de heteroplasmia en músculo para causar debilidad, mientras que en tejidos menos activos podrían no manifestarse clínicamente hasta porcentajes casi homoplásmicos (Stewart & Chinnery, 2015). Esta variabilidad complica el pronóstico genético y clínico, ya que incluso entre hermanos que heredan la misma mutación de su madre, las proporciones de ADNmt mutante pueden diferir en sus tejidos y resultar en cuadros clínicos distintos. La clínica de las enfermedades mitocondriales es muy heterogénea, pero ciertos síndromes se han definido en relación con mutaciones particulares del ADNmt (Nissanka & Moraes, 2020).

La neuropatía óptica hereditaria de Leber (LHON), es la primera enfermedad mitocondrial descrita, caracterizada por la pérdida subaguda de la visión central debido a degeneración del nervio óptico. Suele presentarse en varones jóvenes adultos (Chalmers Ay & Schapira, n.d.; Karaarslan, n.d.). Se asocia clásicamente a mutaciones puntuales en genes del complejo I del ADNmt, frecuentemente homoplásmicas (presentes en el 100% de las copias).

Tres mutaciones en genes ND son responsables de la mayoría de casos: m.11778G>A en MT-ND4, m.3460G>A en MT-ND1 y m.14484T>C en MT-ND6 (Catarino et al., 2017; Chalmers Ay & Schapira, n.d.; Ozir et al., 2024; Vandeputte et al., 2021). Estas mutaciones afectan la función del complejo I y comprometen severamente la producción de ATP en las células ganglionares de la retina, desencadenando su muerte. Algunas mutaciones menos comunes en LHON pueden estar heteroplásmicas (p. ej., m.14459G>A en MT-ND6), y en esos casos se ha observado que debe haber >80% de ADNmt mutante en la retina para que ocurra la ceguera (Catarino et al., 2017; Karaarslan, n.d.; Ozir et al., 2024; Vandeputte et al., 2021).

Síndrome MELAS, acrónimo de Mitochondrial Encephalomyopathy, Lactic Acidosis and Stroke-like episodes, es decir, encefalomiopatía mitocondrial con acidosis láctica y episodios tipo accidente cerebrovascular. Es causado principalmente por la mutación m.3243A>G en el gen mitocondrial MT-TL1 (ARNt^{L^{eu}} (UUR)), responsable del ~80% de los casos. Esta mutación impide la correcta traducción de varios complejos respiratorios, produciendo una encefalopatía progresiva que cursa con convulsiones, migrañas y eventos neurológicos similares a infartos cerebrales en jóvenes (pero que no obedecen a oclusiones vasculares reales). La heteroplasmia es típica en MELAS y en sangre periférica suele encontrarse entre 50–90%. Se estima que niveles mutantes por encima de ~60% en los tejidos críticos son necesarios para precipitar la enfermedad. Curiosamente, la m.3243A>G puede provocar otras manifestaciones si la carga mutacional es más baja, como diabetes y sordera de aparición en el adulto, lo que demuestra la influencia del porcentaje de heteroplasmia en el fenotipo.

Síndrome MERRF, corresponde a Myoclonic Epilepsy with Ragged Red Fibers (epilepsia mioclónica con fibras rojas rasgadas en biopsia muscular) (Shoffner et al., 1990). La mutación clásica es m.8344G>A en el gen MT-TK (ARNt^{L^{ys}}), presente en >80% de los

pacientes. Clínicamente se caracteriza por epilepsia mioclónica progresiva, convulsiones generalizadas, ataxia, debilidad muscular y la aparición de fibras musculares “rojas rasgadas” (agregados mitocondriales anómalos) en biopsias (Carré et al., 2019; Shoffner et al., 1990). La mutación m.8344G>A suele estar heteroplásmica; umbrales de ~70–85% de ADNmt mutante en músculo se asocian con enfermedad manifiesta. Al igual que MELAS, distintas cargas mutacionales causan espectros variables: por ejemplo, se han descrito casos de pacientes con la misma mutación presentando solo epilepsia leve si el porcentaje mutante es menor (Carré et al., 2019; Finsterer, 2019; Lorenzoni et al., 2014; Shoffner et al., 1990).

Los trastornos de síndrome NARP y síndrome de Leigh, pueden ser causados por mutaciones en el gen MT-ATP6 (subunidad a de la ATP sintasa) (Koenig & Grant, 2016). La mutación m.8993T>G (o m.8993T>C) en MT-ATP6 produce típicamente el síndrome NARP (Neuropatía, Ataxia, Retinitis Pigmentosa) cuando está en heteroplasma moderada (~70% mutante); los pacientes presentan neuropatía periférica, pérdida de equilibrio (ataxia) y degeneración retinal progresiva (Kara et al., 2012; Nolte et al., 2021; Stendel et al., 2020). Sin embargo, si esta mutación está presente a niveles muy altos (>90%), el cuadro resultante suele ser el síndrome de Leigh infantil, una encefalopatía neurodegenerativa subaguda grave que cursa con lesiones bilaterales en ganglios basales, retraso del desarrollo, problemas respiratorios y normalmente resulta letal en la primera infancia (Blanco-Grau et al., 2013; Nolte et al., 2021; Stendel et al., 2020).

El síndrome de Leigh es genéticamente heterogéneo (puede deberse también a mutaciones nucleares), pero las mutaciones de ATP6 son una causa relativamente frecuente de la variante mitocondrial (Akar et al., 2024). Este es un claro ejemplo de efecto umbral: la misma mutación m.8993T>G produce NARP con menor carga mutante y Leigh con mayor carga, ilustrando cómo la proporción de ADNmt mutado modula la severidad clínica (Akar et al., 2024; Henke et al., 2024; Nolte et al., 2021).

El síndrome de Kearns–Sayre (KSS) es un trastorno multisistémico causado por grandes deleciones en el ADNmt (de 1.3 hasta 8 kb, típicamente una deleción de ~4.977 pb conocida como la “deleción común”). No sigue un patrón hereditario mendeliano, generalmente es esporádico (mutación *de novo* en la línea germinal de la madre o durante el desarrollo embrionario) (Zeviani et al., 2011). Los pacientes presentan oftalmoplejía

externa progresiva (parálisis de músculos extraoculares con ptosis), retinopatía pigmentaria, sordera, cardiomiopatía y alteraciones endocrinas, manifestándose antes de los 20 años. En los tejidos afectados se observa heteroplasmia, con coexistencia de mitocondrias portadoras de ADNmt truncado (delecionado) y mitocondrias con ADNmt normal (Christodoulou, 2000; Zeviani et al., 2011).

Síndromes relacionados con el síndrome de Pearson (deleción similar con anemia sideroblástica y disfunción pancreática en infancia) y la oftalmoplejía externa progresiva crónica en adultos, los cuales comparten la base molecular de grandes deleciones del ADNmt (Yoshimi et al., 2022). La proporción y distribución tisular de la deleción determinan la presentación: por ejemplo, en Pearson predomina en médula ósea y páncreas; en Kearns-Sayre, en músculo, cerebro y retina. No existe cura y el manejo es sintomático, pero este síndrome ilustra cómo la pérdida de múltiples genes mitocondriales (por la deleción que típicamente abarca genes de la cadena respiratoria y ARNt) provoca un fallo energético profundo en determinados tejidos (Son et al., 2022; Yoshimi et al., 2022; Zeviani et al., 2011).

Además de estos síndromes primarios, las mutaciones del ADNmt han sido implicadas en enfermedades comunes complejas. Por ejemplo, se han observado mutaciones somáticas acumulativas en el ADNmt de neuronas dopaminérgicas en pacientes con enfermedad de Parkinson, sugiriendo un posible rol contributivo en la degeneración neuronal (Chong et al., 2022; Parakatselaki & Ladoukakis, 2021). De igual modo, en ciertos cánceres se describen mutaciones o patrones de deleciones en el ADNmt, aunque su significado funcional aún es objeto de investigación.

A nivel del organismo, se ha propuesto que la acumulación gradual de mutaciones en ADNmt durante la vida – favorecida por su mayor tasa de mutación (10–20 veces más alta que el ADN nuclear) y la falta de mecanismos robustos de reparación – podría contribuir al proceso de envejecimiento, disminuyendo progresivamente la eficiencia energética de tejidos a medida que aumenta la proporción de mitocondrias disfuncionales (Berdanier & Everts, 2001). Modelos en ratones con mayores tasas de mutación mitocondrial presentan envejecimiento acelerado, apoyando esta hipótesis. Sin embargo, la relación causal exacta entre daño en ADNmt y envejecimiento humano sigue en estudio, dados los múltiples

factores concurrentes (Berdanier & Everts, 2001; Lokody, 2014; Sendra et al., 2021; Wen et al., 2025).

Desde una perspectiva epidemiológica, estudios poblacionales han revelado que mutaciones patogénicas del ADNmt no son exclusivas de pacientes con enfermedades manifiestas, se estima que al menos 1 de cada 200 individuos sanos es portador de alguna mutación potencialmente patogénica en el ADNmt, aunque generalmente en baja proporción heteroplásmica que impide la aparición de síntomas (Chong et al., 2022). Esta presencia de mutaciones "silenciosas" refuerza la importancia del efecto umbral y también plantea desafíos en la asesoría genética, pues dichas variantes podrían, al transmitirse y aumentar su proporción en la descendencia, dar lugar a enfermedad en la siguiente generación.

1.1.3. Perspectivas futuras en investigación clínica mitocondrial

Desde el punto de vista diagnóstico, la introducción de la secuenciación masiva ha revolucionado la detección de mutaciones del ADNmt (Santibanez-Koref et al., 2019). Hoy es factible secuenciar el genoma mitocondrial completo de un paciente de forma rápida y asequible, permitiendo identificar mutaciones patogénicas, cuantificar niveles de heteroplasmia e incluso detectar mutaciones somáticas en tejidos específicos. La tendencia actual es integrar el análisis del ADNmt en los diagnósticos genómicos de primera línea para pacientes con sospecha de enfermedad mitocondrial, conjuntamente con paneles de genes nucleares relacionados (Ibrahim et al., 2018; Santibanez-Koref et al., 2019).

Esto acelera el diagnóstico, evitando años de pruebas invasivas, y posibilita asesorar reproductivamente a las familias (por ejemplo, ofreciendo TRM en casos indicados) (Cohen & Adashi, 2016). Asimismo, se están investigando marcadores no invasivos, como la medición de ADNmt libre circulante en sangre, que podría reflejar daño tisular mitocondrial en enfermedades degenerativas o isquémicas, aunque es un área incipiente (Kopinski et al., 2021; Patop & Kadener, 2018; Zong et al., 2016).

Dentro de las perspectivas de investigación mitocondrial se incluyen profundizar en la comprensión de la interacción mitonuclear y sus implicaciones para la medicina personalizada. Modelos experimentales avanzados, como organoides derivados de células madre pluripotentes inducidas (iPSC) portadoras de mutaciones mitocondriales, ofrecen plataformas para estudiar enfermedades en un contexto de tejido humano y para probar tratamientos *in vitro*. La posibilidad de generar líneas celulares isogénicas que difieran solo en su ADNmt (por técnicas de fusión citoplasmática, por ejemplo) permite aislar los efectos de variantes mitocondriales específicas. Esto será valioso para discernir por qué ciertas mutaciones del ADNmt tienen expresividad variable o por qué predisponen a algunos individuos a enfermedades multifactoriales.

A medida que la investigación básica descubra nuevas vías de comunicación mitonuclear y nuevos blancos moleculares, es esperable que emerjan terapias dirigidas a mitocondrias en campos tan diversos como neurología, cardiología y endocrinología. En síntesis, la biología del ADN mitocondrial se ha convertido en un campo interdisciplinario que conecta genética, bioquímica, medicina y biotecnología. Los conocimientos cimentados en este capítulo proveen la base teórica para explorar – con herramientas bioinformáticas y experimentales de vanguardia – cómo las variaciones en el genoma mitocondrial contribuyen a la fisiopatología y cómo pueden ser moduladas para mejorar la salud humana. Las mitocondrias, una vez vistas solo como el “motor” celular, son ahora reconocidas como participantes clave en redes de señalización y patología, y los esfuerzos por manipular su genoma de manera segura abren una nueva frontera en la medicina genómica del siglo XXI.

1.2. Secuenciación

La secuenciación de ADN ha transformado la biología molecular y la genética, convirtiéndose en una herramienta esencial desde la investigación básica hasta la práctica clínica. En las últimas cuatro décadas, las tecnologías de secuenciación han evolucionado dramáticamente en respuesta a la necesidad de leer genomas completos de forma rápida y exacta. Se han sucedido al menos tres “generaciones” de métodos: la secuenciación de primera generación o método de Sanger desarrollado en los años 1970, las tecnologías de

secuenciación masiva de segunda generación (NGS, por Next-Generation Sequencing) surgidas a mediados de la década de 2000, como la pirosecuenciación 454 e Illumina, y las plataformas de tercera generación de secuenciación de molécula única, ejemplificadas por ONT y Pacific Biosciences of California, Inc. (PacBio®) (Giani et al., 2020; Shendure et al., 2017; Sood & Singh, 2024).

Cada generación ha logrado disminuciones enormes en el costo por base y aumentos en el rendimiento, habilitando proyectos antes impensables como la secuenciación del genoma humano completo. Esta reducción de costos —del orden de un millón de veces menor costo por base desde 2001— ha allanado el camino para aplicaciones clínicas, incluyendo el diagnóstico genético de enfermedades raras y la vigilancia genómica de patógenos (Król et al., 2023; McCombie & McPherson, 2019).

En este capítulo se realiza una revisión histórica y técnica de las principales tecnologías de secuenciación de ADN, desde el método de Sanger hasta las plataformas Illumina y ONT, con especial énfasis en estas dos últimas. Se compararán estas tecnologías en cuanto a precisión, longitud de lectura, tiempo de ejecución, costo y aplicabilidad clínica. Asimismo, se discutirán las aplicaciones clínicas actuales de la secuenciación, con énfasis especial en el diagnóstico de enfermedades mitocondriales, la detección de heteroplasmia y el diagnóstico genético. Finalmente, se abordan las perspectivas a futuro, incluyendo tendencias de miniaturización, portabilidad, integración con inteligencia artificial y la posibilidad de un diagnóstico genómico descentralizado.

1.2.1. Panorama histórico de la secuenciación

La secuenciación de ácidos nucleicos se originó en la segunda mitad del siglo XX, impulsada por la convicción de científicos como Frederick Sanger de que conocer la secuencia exacta de nucleótidos era crucial para entender la biología. El método de Sanger (también llamado de dideoxinucleótidos) utiliza la síntesis de ADN por una ADN polimerasa a partir de un cebador, en presencia de una pequeña fracción de nucleótidos modificados que provocan terminación de la cadena. Mediante cuatro reacciones paralelas (cada una

con un dideoxinucleótido diferente marcado radiactivamente), seguidas de electroforesis en gel de poliacrilamida, es posible leer la secuencia según la longitud de los fragmentos terminados en A, C, G o T, que permitió hitos tempranos como la secuenciación del genoma completo de un virus (el fago ϕ X174, de 5.3 kb, en 1977) y del bacteriófago λ (48,5 kb, en 1982) (Heather & Chain, 2016; Hutchison, 2007).

En los años 1980 se introdujo la automatización de la secuenciación Sanger, con la sustitución de marcajes radiactivos por colorantes fluorescentes de distinto color para cada terminador y la lectura automática mediante un láser y detector en geles o capilares. En 1987, Lloyd Smith, Leroy Hood y colegas desarrollaron la primera máquina secuenciadora automática comercial (Applied Biosystems ABI370) capaz de leer unos 1000 nucleótidos por día sin intervención humana, lo que disparó exponencialmente la acumulación de datos. Bases de datos como GenBank crecieron de cientos de miles de bases a inicios de los 80's a decenas de millones a fines de esa década. Esta escalabilidad, junto con la mejora en la fidelidad (errores de menos de 1 en 10.000 bases mediante buena calibración), consolidó a la secuenciación Sanger como la base de grandes proyectos genómicos en los 90's (Heather & Chain, 2016; Shendure et al., 2017).

El proyecto más emblemático de la era Sanger fue el Proyecto Genoma Humano (HGP), llevado a cabo entre 1990 y 2003, que secuenció ~3.000 millones de pares de bases del genoma humano usando el método Sanger mediante una estrategia de shotgun a gran escala. El HGP requirió la labor de centros de secuenciación en varios países, enormes bibliotecas de clones BAC y capilaridades automatizadas funcionando 24/7 (Shendure et al., 2017). El costo inicial del HGP se estimó en ~\$3 mil millones de dólares y tomó ~13 años completar el primer borrador (publicado en 2001) y la secuencia "completa" en 2003. Sin embargo, hacia el final del proyecto, las mejoras tecnológicas habían elevado tanto el rendimiento que el consorcio logró terminar dos años antes de lo previsto. Durante el HGP, la capacidad de secuenciación aumentó y el costo por base disminuyó más de 100 veces gracias a procesos industrializados de secuenciación Sanger. Para 2004, los secuenciadores Sanger con electroforesis capilar podían leer 600–700 bases por secuencia a un costo de ~\$1 dólar por reacción (Heather & Chain, 2016; Hutchison, 2007).

A pesar de estos avances, se volvió evidente que la tecnología Sanger, por muy precisa que fuera (error típico <0.01%), difícilmente podría escalar para resecuenciar genomas

humanos de múltiples individuos a un costo y tiempo razonables. La secuenciación Sanger seguía siendo relativamente lenta y costosa por base para proyectos de gran envergadura, y el término "\$1000 genome" (el genoma de mil dólares) surgió tempranamente (2001) como un objetivo aspiracional que requeriría nuevas innovaciones tecnológicas.

1.2.2. La revolución de la segunda generación: Secuenciación de Nueva Generación (NGS) o masiva en paralelo.

El gran salto adelante vino a mediados de los años 2000 con la introducción de las tecnologías de Next-Generation Sequencing (NGS), también llamadas de secuenciación masiva en paralelo. En 2005 se comercializó la plataforma 454 Life Sciences, basada en pirosecuenciación, que por primera vez permitía secuenciar cientos de miles de fragmentos en paralelo mediante una emulsión y detección luminosa de la incorporación de nucleótidos. Poco después, en 2006–2007, entraron en escena la tecnología Solexa (adquirida por Illumina) y la plataforma SOLiD de Applied Biosystems, entre otras, desencadenando una competición intensa por reducir costos y aumentar el rendimiento de secuenciación (Satam et al., 2023).

Estas plataformas de segunda generación comparten el principio de generar billones de lecturas cortas en paralelo a partir de millones de fragmentos de ADN, generalmente mediante una fase de amplificación clonal (clúster en Illumina o emulsión en 454/SOLiD) seguida de un proceso cíclico de secuenciación por síntesis o ligación (Akintunde et al., 2023).

En el caso de Illumina, que pronto se convirtió en la dominante, se emplea secuenciación por síntesis (SBS por sus siglas en inglés) con terminadores reversibles fluorescentes en el que en cada ciclo se añade un nucleótido identificado por su color, permitiendo leer cientos de bases por fragmento con alta precisión. Aunque la longitud de lectura típica de Illumina es relativamente corta (100–300 pb), la producción de datos es enorme y altamente precisa (error en el orden de 0.1–1% por base) (Satam et al., 2023).

El resultado final fue un abaratamiento y aceleración sin precedentes, que entre 2007 y 2012, el costo por base secuenciada cayó cuatro órdenes de magnitud, superando con creces la ley de Moore en informática (Akintunde et al., 2023). Para dimensionar el avance, en 2001 generar el borrador del genoma humano (HGP) requirió años de trabajo distribuido; hacia 2017, un solo secuenciador NGS de vanguardia (Illumina NovaSeq) podía secuenciar más de 20–30 genomas humanos en un par de días por unos cuantos miles de dólares, produciendo en una sola corrida hasta 3,000 gigabases de secuencia. Shendure et al. señalaban que un solo estudiante graduado con un NovaSeq podía obtener 40 veces más bases que las obtenidas en todo el proyecto del genoma humano, a una fracción del costo. Illumina consolidó rápidamente un cuasi-monopolio en el mercado de secuenciación para 2014, llegando a aportar ~90% de todos los datos de ADN generados en el mundo (Muir et al., 2016).

La eficiencia de la secuenciación de segunda generación permitió expandir las aplicaciones: secuenciación de exomas completos para genética médica, RNA-seq para transcriptomas, ChIP-seq para estudiar regulación génica, estudios masivos de variación genética en poblaciones (proyecto 1000 Genomas, etc.), entre muchos otros. No obstante, las tecnologías NGS también presentaban limitaciones (Satam et al., 2023). Las lecturas cortas (80-150pb) dificultaban el ensamblaje *de novo* de genomas complejos, resultando a menudo en secuencias de referencia fragmentadas o incompletas (típicamente borradores con cientos de gaps). Asimismo, ciertas regiones repetitivas o con composiciones de bases extremas (muy ricas en GC o en AT) podían secuenciarse con sesgos o requerían cobertura adicional (Król et al., 2023; McCombie & McPherson, 2019).

Para abordar estas deficiencias, surgieron estrategias complementarias: por ejemplo, el uso de lecturas pareadas de extremos (paired-end) para resolver repetidos cortos, métodos de mapeo físico o con *optical mapping* para ordenar *contigs*, y posteriormente técnicas de *scaffolding* usando contactos de cromatina (Hi-C) para acercarse a ensamblajes a nivel cromosoma (Satam et al., 2023). Sin embargo, la limitación fundamental de la segunda generación era la longitud de lectura. Este obstáculo motivó la búsqueda de métodos radicalmente nuevos que pudieran secuenciar moléculas individuales de ADN mucho más largas, dando paso a la llamada tercera generación.

1.2.3. Secuenciación de tercera generación: Secuenciación de molécula única en tiempo real.

La tercera generación de tecnologías de secuenciación se caracteriza por la capacidad de secuenciar moléculas individuales de ADN (o ARN) sin necesidad de amplificación, en tiempo real. Dos enfoques principales han alcanzado éxito comercial: la secuenciación por síntesis de molécula única representada por PacBio, y la secuenciación por nanoporos introducida por ONT (Xiao & Zhou, 2020). Estos métodos, concebidos conceptualmente desde los 1990s, se concretaron tecnológicamente hacia la década de 2010 (Chen et al., 2023). Para cuestiones de este trabajo se estará enfocado mayormente en la secuenciación por nanoporos.

La secuenciación por nanoporos ofrece una aproximación distinta y sumamente innovadora. Concebida teóricamente décadas atrás, no fue sino hasta mediados de la década de 2010 que se demostró viable de forma comercial, principalmente gracias al trabajo de ONT (empresa fundada en 2005) (Bayley, 2015). La secuenciación de nanoporos se basa en hacer pasar una molécula de ácido nucleico de cadena sencilla a través de un poro nanoscópico (un proteína poro, por ejemplo MspA o α -hemolisina, inserta en una membrana sintética) bajo el impulso de un voltaje eléctrico (Bayley, 2015).

El paso de la molécula modula la corriente iónica a través del poro; dicha corriente cambia según la secuencia de nucleótidos que esté ocupando el poro en ese instante. Dado que el poro es suficientemente estrecho, típicamente detecta una ventana de 4–6 bases simultáneamente, generando una señal eléctrica característica (*squiggle*) de la cual es posible inferir la secuencia mediante algoritmos de *basecalling* entrenados para reconocer patrones de corriente (Goodwin et al., 2016a).

A diferencia de otras plataformas, no interviene síntesis de ADN ni etiquetas ópticas: la lectura es directa y electrónica. Esta simplicidad estructural le ha dado a los dispositivos de Nanopore una ventaja única como lo es la portabilidad extrema. De hecho, los secuenciadores ONT pueden ser tan pequeños como una memoria USB (~70 gramos de

masa) y alimentarse vía laptop, porque no requieren óptica sofisticada (Goodwin et al., 2016a; Kasianowicz & Bezrukov, 2016).

Ya en 2016, un equipo de investigación demostró la capacidad de llevar un secuenciador MinION al campo para secuenciar el virus Ébola *in situ* en África, obteniendo datos en menos de 1 hora desde la muestra hasta la secuencia, algo imposible con otras tecnologías en entornos convencionales (Hoenen et al., 2016). Desde entonces, dispositivos Nanopore portátiles han sido utilizados incluso en la Estación Espacial Internacional y en lugares remotos para vigilancia epidemiológica (Castro-Wallace et al., 2017; Mongan et al., 2020).

Las lecturas de ONT son notables por su longitud (desde 1kb hasta 1Mb), al no depender de amplificaciones (dependiendo de la aplicación), donde, en teoría la lectura puede continuar hasta que termine la molécula (Logsdon et al., 2020). En la práctica, se han reportado lecturas individuales de más de 1 millón de bases usando protocolos de *ultra long read* especiales, por lo que es común obtener distribuciones con N50 de decenas o cientos de kilobases, muy por encima de cualquier otra tecnología de secuenciación (Amarasinghe et al., 2020b; Logsdon et al., 2020). Esto ha permitido, por ejemplo, ensamblar genomas bacterianos completos en una sola lectura o caracterizar regiones altamente repetitivas del genoma humano antes inalcanzables. Además, al igual que PacBio, la secuenciación Nanopore ocurre a nivel de molécula única y no requiere amplificación PCR, lo que elimina sesgos de clonación o problemas con regiones de alta o baja GC (MacKenzie & Argyropoulos, 2023; Y. Wang et al., 2021).

Esta lectura directa de moléculas nativas tiene otra consecuencia poderosa como lo es la posibilidad de detectar modificaciones epigenéticas directamente, a través de cambios en la señal de corriente que pueden inferir la presencia de bases modificadas como 5-metilcitosina sin necesidad de tratamientos adicionales (Branton, Deamer, Marziali, Bayley, Benner, Butler, Di Ventra, Garaj, Hibbs, Huang, Jovanovich, Krstic, Lindsay, Ling, et al., 2008; Chen et al., 2023; Y. Wang et al., 2021). ONT ha demostrado detección directa de metilaciones de ADN y también ofrece un protocolo para secuenciación directa de ARN nativo, preservando marcas como m6A; esta capacidad de leer la molécula original con su información epigenética es un valor añadido de la tercera generación (Lorenz et al., 2020; Teng et al., 2024; Xie et al., 2021).

No obstante, la secuenciación por nanoporos inicialmente presentó tasas de error más altas que las plataformas anteriores. En los primeros años (alrededor de 2014–2016, con los poros R7 y R9) el porcentaje de error promedio era de ~10–20% en la secuencia cruda, con errores dominados por *indels* en regiones de homopolímeros (MacKenzie & Argyropoulos, 2023; Y. Wang et al., 2021). Además, se observó que algunos errores no eran puramente aleatorios, sino que había ciertos sesgos o dificultades sistemáticas (por ejemplo, en secuencias de polidG) que complicaban la corrección por simple redundancia de cobertura (Delahaye & Nicolas, 2021a).

Sin embargo, ha habido progresos rápidos en mejorar la exactitud de ONT mediante avances en la química y los poros, por ejemplo, la mejora en 2019 ONT con el poro R10 con doble “sensor” de lectura (dos puntos de constricción) diseñado para leer con mayor resolución las bases repetidas, reduciendo errores sistemáticos (MacKenzie & Argyropoulos, 2023). En 2021–2022 se lanzó la versión R10.4.1 y nuevos reactivos (“Kit Q20+”) que, junto con mejoras de *software*, han mejorado la calidad de las lecturas.

Por otro lado, el refinamiento de los algoritmos de *basecalling* mediante redes neuronales profundas (modelos de aprendizaje automático entrenados con grandes volúmenes de datos) ha sido crucial. Los primeros *basecallers* de ONT (como el algoritmo HMM en 2014 o el primer modelo neuronal en 2016) fueron sustituidos por modelos de *deep learning* cada vez más precisos, como el algoritmo *flip-flop* introducido alrededor de 2018, y más recientemente modelos tipo transformador como Guppy y Dorado que han permitido que la tasa de error promedio de lecturas ONT se ha reducido a ~1% en 2023-2024 (Athanasopoulou et al., 2025; Pagès-Gallego & de Ridder, 2023; et al., 2019b).

En otras palabras, hoy las lecturas crudas de Nanopore alcanzan calidades cercanas a Q20 (99% de exactitud) e incluso superiores en muchos contextos, algo impensable pocos años atrás (Wagner et al., 2023). Esto significa que, para aplicaciones de variantes de un solo nucleótido, ya es posible obtener resultados de alta fidelidad sin necesidad obligatoria de posteriores *polishing* con lecturas de alta cobertura (Szakállas et al., 2024; Tafazoli et al., 2025). Persisten todavía algunas fuentes de error (particularmente *indels* en tramos muy homogéneos), pero la brecha entre Nanopore e Illumina en términos de precisión se ha estrechado dramáticamente (Santos et al., 2025; Szakállas et al., 2024; Tafazoli et al.,

2025). Las secciones siguientes compararán cuantitativamente varios de estos aspectos de desempeño entre NGS (Illumina) y ONT.

1.2.4. Métricas asociadas a la secuenciación

1.2.4.1. La Precisión

La precisión (fidelidad de lectura por base) es un parámetro crítico, especialmente en aplicaciones clínicas donde se deben detectar mutaciones puntuales con confianza. La tecnología Sanger ha sido históricamente el estándar dorado en exactitud, alcanzando error menores a 1 en 10,000 bases (Q40) bajo buenas condiciones. De hecho, la mayoría de variantes identificadas por métodos masivos suelen confirmarse mediante Sanger para asegurar que no sean artefactos. Sin embargo, Sanger presenta dificultades para detectar variantes en mosaico o minoritarias en una muestra, ya que la señal de secuenciación es la suma de moléculas clonadas: típicamente, una mutación debe estar presente en ~20% o más de las moléculas para ser visible en el cromatograma (Rohlin et al., 2009; Salk et al., 2018; Szakállas et al., 2024). Esto lo hace poco sensible para, por ejemplo, mutaciones somáticas de baja frecuencia o heteroplasmas bajas en ADN mitocondrial (tema que retomaremos más adelante).

Las plataformas de Illumina (NGS) alcanzan una precisión por base muy alta, cercana a 99.9% (Q30) o superior. Los errores que ocurren son principalmente sustituciones aleatorias, y gracias a la enorme cobertura de lectura redundante, es posible distinguir verdaderas variantes de errores de secuenciación (Yin et al., 2021). En la práctica, se logran tasas de falsos positivos extremadamente bajas al llamar variantes con lecturas Illumina, siempre que la profundidad y los algoritmos de alineamiento/llamado sean adecuados (Yin et al., 2021).

Un desafío con NGS de segunda generación es la interpretación de variantes en presencia de pseudogenes o secuencias repetidas, por ejemplo, si una lectura corta puede mapear

en múltiples lugares del genoma con igual similitud, esto puede producir asignaciones erróneas y llamados falsos (Satam et al., 2023). Este problema es relevante cuando se analizan variantes en genes con duplicados parciales o en el caso de ADN mitocondrial vs. NUMTs (secuencias mitocondriales insertadas en el núcleo), como discutiremos en heteroplasma (Claes & De Leeneer, 2014; Salk et al., 2018). Aun así, en términos puramente de fidelidad de lectura, la segunda generación ofrece un excelente desempeño, al punto que lecturas de Illumina se consideran “verdad de referencia” para calibrar errores de otras tecnologías (Xuan et al., 2013).

En la secuenciación de tercera generación hubo una clara discordancia entre precisión cruda de una sola lectura y precisión de consenso. La tecnología de ONT inicialmente presentaba ~90% de exactitud media por lectura en 2018, lo que implicaba ~10 errores por cada 100 bases, concentrados en *indels* (Claes & De Leeneer, 2014). Este nivel de error en los datos crudos requería obligatoriamente una cobertura elevada (p. ej. >30×) y a menudo el apoyo de datos de Illumina para alcanzar seguridad en llamadas de variantes puntuales. Un informe en 2018 indicaba que las lecturas Nanopore R9.4 tenían ~87–90% de precisión media (Delahaye & Nicolas, 2021a; Zhang et al., 2024).

La rápida mejora de la tecnología ha elevado la precisión de ONT sustancialmente, por ejemplo, en algunos estudios recientes reportan que con la química R10.4.1 y los nuevos modelos de *basecalling*, la tasa de error se ha reducido a ~1% por base en promedio. Kovaka et al., (2025) destacan que la combinación de mejoras en los poros (doble lector) y algoritmos ha disminuido tanto el error que ya no es estrictamente necesario pulir con señales crudas en muchos casos, salvo quizás en ARN donde la precisión aún es menor que en ADN (Lorenz et al., 2020; Wagner et al., 2023; T. Zhang et al., 2024).

En síntesis, hoy podríamos situar la precisión de Sanger ~99.99%, Illumina ~99.9% y Nanopore ~98-99%. Para la gran mayoría de aplicaciones de variantes de un solo nucleótido (SNVs), tanto Illumina como PacBio HiFi ofrecen confianza casi absoluta (Satam et al., 2023). Nanopore se acerca a ese rango, aunque persisten ciertos errores de inserción/delección recurrentes que pueden dificultar, por ejemplo, el conteo exacto de repeticiones de polinucleótidos (homopolímeros) (Moustakli et al., 2025).

Un punto adicional relacionado a la precisión es la cuantificación de mezclas de secuencias, relevante en heterogeneidad de muestras (mosaicismos, heteroplasma, quimerismo). Las plataformas NGS, al generar datos digitales (muchas lecturas individuales), permiten estimar fracciones alélicas con alta sensibilidad. Por ejemplo, se ha demostrado que la secuenciación profunda del genoma mitocondrial puede detectar variantes en heteroplasma tan bajas como 1% con suficiente cobertura (Cheng et al., 2023). Sanger, al ser una técnica analógica (lectura superpuesta de señales), tiene dificultades para detectar heteroplasma debajo de ~15–20%. Nanopore también genera lecturas individuales, por lo que en teoría podría cuantificar frecuencias de alelos minoritarios; no obstante, su error intrínseco hasta hace poco hacía difícil distinguir un alelo verdadero al 1% de ruido de secuenciación. Estudios en 2024 indican que ONT puede detectar heteroplasmas ~12% con confianza usando ultra-profundidad (Slapnik et al., 2024).

1.2.4.2. Longitud de lectura y el ensamblaje

La longitud de lectura es la otra cara de la moneda en rendimiento de secuenciación. Aquí las diferencias son dramáticas entre generaciones. El método Sanger produce lecturas de hasta ~800–1000 bases por corrida, limitado por la dispersión de bandas en el gel o capilar. Esto fue suficiente para secuenciar pequeños genomas virales de unas pocas kb con enfoque de fragmentos solapados, pero insuficiente para atravesar regiones genómicas altamente repetitivas (e.g., Sanger no puede secuenciar de corrido un Alu de 300 pb seguido de otro Alu idéntico; se requerían estrategias jerárquicas con vectores cósmidos o BACs para resolver repetidos largos) (Heather & Chain, 2016).

Illumina y otros NGS inicialmente sacrificaron longitud por cantidad con lecturas típicas de 35–100 pb en los primeros sistemas (2006–2010), luego extendidas a 150–250 pb en la mayoría de plataformas actuales e incluso ~300 pb en kits MiSeq emparejados. Pese a mejoras marginales (Illumina probó un protocolo para ~500–600 pb pero con alto error al final), esencialmente la segunda generación se considera de lectura corta (Akintunde et al., 2023; Heather & Chain, 2016). Esto significa que, al alinear contra un genoma de

referencia, la mayoría de lecturas solo cubren parte de un exón o de un elemento móvil, por lo que la detección de variantes puntuales es robusta pero la resolución de variantes estructurales (inversiones, translocaciones, repeticiones expandidas) es limitada (Satam et al., 2023; Xuan et al., 2013). Se requieren técnicas complementarias (p. ej., uso de paired-ends espaciales, o algoritmos de split-read y read-pair) para inferir eventos estructurales a partir de lecturas cortas. También el ensamblaje *de novo* con lecturas de 150 pb es un problema computacional complejo: por lo general se obtienen cientos o miles de *contigs* que luego deben ordenarse con información adicional (Xuan et al., 2013).

Las tecnologías de tercera generación cambiaron por completo el panorama al ofrecer lecturas largas que pueden abarcar íntegramente elementos repetitivos y regiones complejas. ONT, como se mencionó, ha logrado los récords como en 2018 en donde investigadores de Oxford reportaron lecturas de >2 megabases usando un poro R9.4 y ADN gDNA muy cuidadosamente preparado (ultra alto peso) (Delahaye & Nicolas, 2021a). Sin embargo, lo más común es encontrar procesos de secuenciación con un N50 de entre 2kb-20kb, lo que ha permitido secuenciar de forma contigua (sin brechas) regiones como los complejos loci de inmunoglobulinas, duplicaciones segmentarias, centrómeros y telómeros, que con NGS tradicional quedaban fragmentados (Logsdon et al., 2020).

Las lecturas largas permiten ensamblar genomas eucariotas completos con *contigs* de tamaño cromosómico, especialmente cuando se combina con datos Hi-C u ópticos para orientar los *contigs* (Baudry et al., 2020.). Un caso paradigmático fue el consorcio Telomere-to-Telomere (T2T), que en 2022 publicó la primera secuencia completa de todos los cromosomas humanos sin lagunas, aprovechando lecturas ultra largas de ONT (hasta 100 kb) para atravesar regiones satélite centroméricas, complementadas con lecturas PacBio HiFi para la precisión de base (Koren et al., 2024).

Por lo tanto, ONT continúa enfocada en aumentar la longitud: han desarrollado protocolos de extracción y manejo de ADN gentiles para conservar fragmentos de cientos de kilobases, e incluso un kit especial de “ultra-long reads” que usa transposasas modificadas para etiquetar ADN sin fragmentarlo completamente, permitiendo porcentajes de lecturas gigantes (Santos et al., 2025; Y. Wang et al., 2021). La implicación práctica de las lecturas largas es especialmente importante en el contexto clínico para variantes estructurales y genomas complejos ayudando por ejemplo en la detección de reorganizaciones

cromosómicas equilibradas (invisibles a *arrays* o NGS cortas) y en el caso de genomas mitocondriales, la identificación de deleciones grandes o reordenamientos (Slapnik et al., 2024).

1.2.4.3. Rendimiento, velocidad y logística de secuenciación

El rendimiento se refiere al volumen de datos generados por unidad de tiempo o por corrida. Ya anticipamos cifras en la tabla 1, un secuenciador Sanger tradicional procesa decenas de muestras por corrida, obteniendo del orden de 10^5 bases en unas horas. En cambio, un secuenciador Illumina moderno produce del orden de 10^{11} bases (cientos de gigabases) en ~24–48 horas (Heather & Chain, 2016). Esto significa que NGS permite secuenciar decenas de genomas humanos por semana en un solo instrumento, lo que habilitó proyectos poblacionales inmensos y el ingreso de la secuenciación al ámbito clínico rutinario (por ejemplo, secuenciar el exoma completo de un paciente en unos días es factible y relativamente accesible en costo) (Satam et al., 2023).

Las plataformas ONT tienen un rendimiento más ajustable al ser modulares (p. ej., un equipo PromethION tiene hasta 48 celda de flujos que pueden correrse independientemente), uno puede escalar desde pequeños experimentos (un Flongle, celda de flujo miniatura, produce hasta ~2 Gb de datos, útil para secuenciar un genoma mitocondrial o un panel pequeño) hasta experimentos masivos (una celda de flujo PromethION puede rendir >100 Gb, y usando varias en paralelo se alcanzan varios terabases) (Y. Wang et al., 2021).

En general, Illumina sigue teniendo la ventaja en costo por gigabase para proyectos de gran escala, gracias a su eficiencia en paralelo masivo. Sin embargo, ONT ofrece flexibilidad para correr una sola muestra sin desaprovechar capacidad – en Illumina, si uno quiere secuenciar solo un genoma, igualmente debe gastar una celda de flujo completo, lo que no es económico a baja escala. ONT permite por ejemplo cargar una muestra en una celda de flujo MinION y, si se desea, parar la corrida anticipadamente cuando ya se obtuvo suficiente cobertura, ahorrando tiempo y reactivos (Chen et al., 2023).

En cuanto a tiempo de respuesta, Illumina requiere pasos de preparación (amplificación clonal en la celda de flujo) que demoran varias horas, y solo una vez iniciado el run se obtienen datos al final del ciclo completo (aunque existen modos de “análisis en tiempo real” los datos no son inmediatamente útiles hasta secuenciar cierto fragmento) (Akintunde et al., 2023). Por el contrario, ONT genera datos en tiempo real en donde a los pocos segundos de iniciar la corrida ya se pueden ver lecturas saliendo y analizarlas simultáneamente. Esto ha sido crucial en aplicaciones como identificación rápida de patógenos – por ejemplo, durante el brote de COVID-19, laboratorios usaron MinION para obtener genomas virales completos en menos de 6–8 horas desde la muestra, acelerando la vigilancia epidemiológica (Chen et al., 2023; Hoenen et al., 2016; MacKenzie & Argyropoulos, 2023).

La tabla 1 resume las principales características técnicas, operativas y económicas de las tres generaciones de tecnologías de secuenciación, abarcando desde el método clásico de Sanger hasta las plataformas más avanzadas de lectura directa como ONT. Se comparan aspectos como el principio de funcionamiento, necesidad de amplificación, longitud de lectura, precisión por base, rendimiento, tiempo y costo por corrida, así como sus ventajas y limitaciones más relevantes. Esta comparación destaca cómo cada tecnología presenta fortalezas particulares según el objetivo experimental, desde validación de variantes puntuales hasta análisis de genomas completos o identificación de variantes estructurales.

Característica	Sanger (1ª gen)	Illumina (NGS, 2ª gen)	ONT, (3ª gen)	Referencias
Año introducción	1977	~2006 (Solexa/Illumina)	~2014 (MinION beta)	(Heather & Chain, 2016; Shendure & Ji, 2008; Wang et al., 2021a)
Principio	Terminación de cadena (ADN polimerasa + ddNTP)	Secuenciación por síntesis con amplificación masiva (reversibles terminadores)	Lectura de molécula única por corriente iónica en nanoporo (sin síntesis)	(Branton, David W Deamer, et al., 2008; Heather & Chain, 2016; Shendure & Ji, 2008; Wang et al., 2021a)
Amplificación previa	Sí (clonación o PCR del fragmento)	Sí (clústers por bridge PCR en flowcell)	No (lectura directa de ADN/ARN nativo)	(Petersen et al., 2019; Shendure & Ji, 2008; Wang et al., 2021a)
Longitud de lectura típica	500–800 bp (hasta ~1000 bp máx.)	150–250 bp (hasta 300–600 bp en paired-end)	Variable: promedio 10–50 kb; con protocolos UL puede >100 kb; récord >1 Mb	(Goodwin et al., 2016b; Heather & Chain, 2016; Logsdon et al., 2020; Shendure & Ji, 2008; Wang et al., 2021a)
Precisión por base	>99.99% (Q40)	~99.9% (Q30) por lectura; errores aleatorios	~90–95% inicial, ahora ~98–99% con R10.4.1 (Q20); errores indel aún presentes (consenso >99.9% con cobertura)	(Amarasinghe et al., 2020a; Delahaye & Nicolas, 2021a; Goodwin et al., 2016b; Heather & Chain, 2016; Shendure & Ji, 2008; Wang et al., 2021a; Wick et al., 2019b)

Rendimiento por corrida	~0.5–1 Mb por día (capilar 96 muestras × 800 pb)	10 ⁹ –10 ¹¹ bases por corrida (según equipo, e.g. NovaSeq ~3000 Gb)	10 ⁸ –10 ¹⁰ bases por corrida (depende del dispositivo: MinION hasta ~50 Gb; PromethION >200 Gb)	(Goodwin et al., 2016b; Heather & Chain, 2016; Logsdon et al., 2020; Shendure & Ji, 2008; Wang et al., 2021a)
Tiempo típico de corrida	1–3 horas por secuencia (~700 pb)	1–3 días por run (pero millones de secuencias en paralelo)	Variable: desde minutos a 48 h (datos en tiempo real; se puede detener al lograr cobertura deseada)	(Greninger et al., 2015; Heather & Chain, 2016; Shendure & Ji, 2008; Wang et al., 2021a)
Costo aproximado	Alto por base (ej.: ~\$1 por lectura de 700 pb en 2004); rentable solo para pocas dianas	Más bajo por base en alto volumen (ej.: ~\$600 por genoma humano 30× en 2025); requiere equipamiento caro	Costo por base decreciente, cercano a NGS en proyectos grandes; inversión inicial baja (MinION ~\$1000) pero reactivos por corrida moderadamente altos. Económico para secuenciar pocos genomas; menos eficiente que Illumina para gran número de muestras.	(Heather & Chain, 2016; Muir et al., 2016; Shendure & Ji, 2008; Wang et al., 2021a)
Ventajas principales	Muy alta exactitud; lectura moderada; ideal para secuenciar	Altísima capacidad y precisión; costo/base bajísimo a gran escala; ampliamente	Lecturas ultralargas; detecta variantes estructurales y epigenéticas; equipos portátiles y flexibles; secuenciación en tiempo real (respuesta rápida).	(Amarasinghe et al., 2020a; Goodwin et al., 2016b; Heather & Chain, 2016; Logsdon et al., 2020; Shendure & Ji, 2008; Wang et al., 2021a)

	un solo gen o validar variantes.	establecida en laboratorios clínicos.		
Limitaciones principales	Bajo rendimiento; costosa y lenta para genomas grandes.	Lecturas cortas dificultan ensamblajes completos y detección de repetidos grandes; requiere infraestructura centralizada y experiencia bioinformática considerable.	Tasa de error bruta más alta (aunque mejorando); errores de <i>indels</i> en homopolímeros; necesidad de mayor profundidad para alta exactitud; manejo de datos voluminosos y <i>basecalling</i> intensivo computacionalmente.	(Delahaye & Nicolas, 2021a; Goodwin et al., 2016b; Heather & Chain, 2016; Logsdon et al., 2020; Wick et al., 2019b)

Tabla 1. Comparación general entre tecnologías de secuenciación de primera (Sanger), segunda (Illumina) y tercera generación (ONT).

1.2.5. Herramientas de análisis para secuenciación ONT

1.2.5.1. *Basecalling*

El *basecalling* constituye el primer paso crítico en el análisis de lecturas generadas por ONT, ya que traduce las señales eléctricas producidas al pasar el ADN por los poros de proteína en secuencias de bases. Dorado incorpora modelos neuronales de última generación, incluyendo variantes de precisión Fast, HAC y SUP, que permiten equilibrar velocidad y exactitud según las necesidades del análisis. Con el modelo SUP (Super Precise), Dorado ofrece los mayores niveles de fidelidad en la llamada de bases, alcanzando calidades Q superiores a 20 en condiciones óptimas. Además, está diseñado para aprovechar la aceleración por GPU de NVIDIA, logrando mejoras de ~30% en velocidad respecto a la generación previa de Guppy, sin sacrificar exactitud (Oxford Nanopore, 2025). Esta combinación de rapidez, escalabilidad y precisión lo convierte en la opción preferida para pipelines bioinformáticos de ADN mitocondrial, donde la detección de heteroplasmia exige lecturas de alta calidad.

Si bien Dorado fue el *basecaller* de elección en este estudio, resulta relevante mencionar herramientas predecesoras como Guppy y Bonito, dado que sentaron las bases del progreso actual. Guppy, ampliamente utilizado en entornos de producción, introdujo modelos de *basecalling* rápidos (Fast), de alta precisión (HAC) y de súper precisión (SUP), convirtiéndose en el estándar por varios años. Bonito, por su parte, fue una herramienta experimental de ONT basada en redes neuronales con decodificación CRF, liberada como software abierto en PyTorch. Aunque más lenta, Bonito alcanzó niveles de exactitud superiores en pruebas comparativas y sirvió como banco de pruebas para innovaciones que posteriormente se incorporaron en Guppy y ahora en Dorado. En conjunto, estas herramientas ilustran la evolución del *basecalling* en ONT hacia mayores niveles de fidelidad y robustez (Pagès-Gallego & de Ridder, 2023).

1.2.5.2. Trimming

Tras el llamado de bases, es habitual realizar un filtrado y recorte (*trimming*) de las lecturas de ONT para mejorar la calidad del conjunto de datos antes de análisis posteriores (ensamblado, mapeo, etc.). Las lecturas de Nanopore suelen exhibir una distribución amplia de longitudes y calidades, incluyendo fragmentos cortos o con puntajes de calidad bajos que pueden perjudicar los análisis (Delahaye & Nicolas, 2021). Por ejemplo, se ha observado que lecturas con calidad promedio $Q < 13-15$ tienden a alinearse con muchos más errores, por lo que filtrarlas puede aumentar significativamente la precisión global. Asimismo, los extremos de las lecturas Nanopore a veces presentan mayor tasa de error o contenido de adaptadores/*barcodes* remanentes, lo que motiva recortarlos. Para abordar esto, se emplean herramientas especializadas como NanoFilt y Filtlong, que permiten descartar lecturas deficientes y/o recortar secciones no deseadas de forma controlada antes del ensamblado o mapeo (Peresini et al., 2021b; Wang et al., 2021c).

NanoFilt es una herramienta ligera (parte del paquete NanoPack) diseñada para filtrar y recortar lecturas largas de ONT implementada en Python (disponible vía pip o Bioconda) y se puede ejecutar durante la extracción de FASTQ o justo antes del alineamiento, procesando lecturas de entrada estándar y produciendo las filtradas por salida estándar (De Coster et al., 2018). NanoFilt permite ajustar parámetros como filtros por calidad, longitud y GC, donde se pueden descartar lecturas con calidad media inferior a un umbral (-q), con longitud por debajo o encima de ciertos límites (-l, --maxlength), o con porcentaje GC fuera de un rango definido (--minGC, --maxGC). Así como también recorte fijo de extremo, haciendo posible truncar n nucleótidos desde el inicio (--headcrop) y/o el final (--tailcrop) de cada lectura luego del filtrado, para eliminar porciones de baja calidad o secuencias de adaptador en los extremos (De Coster et al., 2018).

NanoFilt calcula la calidad promedio de cada lectura a partir de los valores Phred de sus bases (o empleando, si se proporciona, el resumen de secuenciación de ONT para mayor exactitud). Una vez aplicados los filtros definidos, solo las lecturas que cumplan los criterios

(p. ej., longitud $\geq L$ y $Q \geq X$) son emitidas, y opcionalmente recortadas en sus extremos según se indique (De Coster et al., 2018).

1.2.5.3. Alineamiento y mapeo

Minimap2

Minimap2 se utilizó como método de minimizadores y encadenamiento para indexar el genoma y encontrar alineamientos candidatos, de forma muy eficiente. Está optimizado para lecturas con errores hasta ~15%, típicas de ONT, y provee agrupamientos pre-establecidos específicos: en este caso se usó el preset `-x map-ont` adecuado para lecturas genómicas ONT, que configura parámetros como tamaño de k-mer y puntuación de alineación adaptados a las características de error ONT (Blanca et al., 2021).

Esta herramienta es capaz de alinear un conjunto completo de lecturas de un genoma humano contra una referencia en minutos, siendo decenas de veces más rápido que BWA-MEM o BLASR, sin sacrificar precisión. De hecho, evaluaciones han demostrado que Minimap2 alcanza mayor exactitud en simulaciones de lecturas largas que herramientas anteriores, produciendo alineaciones biológicamente significativas aptas para análisis posteriores. Su algoritmo base usa un índice de semillas (k-mers minimizers) del genoma de referencia; luego encadena coincidencias para identificar regiones candidatas y finalmente realiza una alineación detallada a nivel de bases (calculando la cadena CIGAR) (Blanca et al., 2021; Li, 2021).

Minimap2 por sobre otras opciones como BWA-MEM, BLASR, GraphMap o LAST muestra una mayor velocidad y precisión para ONT. BWA-MEM, por ejemplo, está optimizado para lecturas \leq read-length Illumina y aunque tiene una versión adaptada (BWA-MEM-LR), no iguala la eficiencia de Minimap2 en datos Nanopore. BLASR fue popular con PacBio pero es mucho más lento y menos mantenido. GraphMap (2016) mostraba buena sensibilidad en ONT pero fue superado luego por Minimap2 en exactitud y está discontinuado (Li, 2018).

Al ser activamente desarrollado por Heng Li, integra mejoras constantes (soporte de corregir largas inserciones, manejo de RNA, etc.) y ha sido ampliamente validado como el “*gold standard*” actual para alinear lecturas largas. Además, su versatilidad (sirve también para RNA-seq largo, ensamblaje, etc.) y fácil uso con presets lo hicieron la opción principal (Blanca et al., 2021; Li, 2018, 2021; Oxford Nanopore, 2025).

Winnomap2

Winnomap2 incorpora dos innovaciones principales sobre la base de Minimap2, en primer lugar, un muestreo de minimizadores ponderado que relegando k-mers frecuentes en el genoma, evitando que secuencias repetidas saturen la fase de *seeding*. En lugar de simplemente enmascarar k-mers comunes, Winnomap asigna menos probabilidad de elegir semillas a k-mers muy abundantes, reduciendo falsos alineamientos en copias repetidas (Jain et al., 2020, 2022).

Esto mejora la precisión en regiones complejas (por ejemplo, repeticiones largas en tandem en la región de control del ADNmt) al mantener el principio teórico de minimizers (toda subsecuencia suficientemente larga compartida debería producir un *minimizer* coincidente). Y en segundo lugar, introduce el concepto de MCAS (*Minimal Confidently Alignable Substrings*) para abordar el sesgo alélico en repeticiones. En regiones repetitivas con diferencias alelo-específicas, el alineamiento global de mayor puntaje puede posicionar incorrectamente la lectura en una copia equivocada (fenómeno *de allelic bias*). Winnomap2 mitiga esto fragmentando la lectura en sub-alineamientos confiables (por encima de cierto puntaje de calidad de mapeo) y luego consolidándolos; así tolera variaciones estructurales o polimorfismos específicos de *paralog* sin forzar todo el *read* a una única ubicación de máxima puntuación (Jain, Rhie, Zhang, et al., 2020; *Marbl/Winnomap: Long Read / Genome Alignment Software*, n.d.).

Winnomap2 trata cada mapeo como un conjunto de segmentos alineados con alta confianza, permitiendo diferencias intermedias sin perder el alineamiento. Gracias a estas estrategias, Winnomap2 reporta alineamientos más correctos en regiones repetidas del

genoma (como las duplicaciones subteloméricas o NUMTs) en comparación con Minimap2 puro (Jain, Rhie, Hansen, et al., 2020; Jain, Rhie, Zhang, et al., 2020).

NGMLR

NGMLR utiliza una estrategia distinta enfocada en alinear correctamente lecturas que abarcan grandes inserciones, deleciones o reordenamientos. Su algoritmo combina una búsqueda inicial de k-mers sensible a SVs y luego un alineamiento por Smith-Waterman acotado con un innovador esquema de penalización de gaps no lineal (convexo) (*Philres/Ngmlr: NGMLR Is a Long-Read Mapper Designed to Align PacBio or Oxford Nanopore (Standard and Ultra-Long) to a Reference Genome with a Focus on Reads That Span Structural Variations*, n.d.).

Esta herramienta implementa un modelo de *gap-cost* convexo donde penaliza extensiones de *gaps* largos proporcionalmente menos, que *gaps* cortos. Esto significa que alinear una deleción de, por ejemplo, 50 bases no resultan en una penalización 10 veces mayor que una de 5 bases, lo cual contrarresta la tendencia de los *mappers* clásicos a fragmentar o mal posicionar lecturas con SVs grandes. Gracias a este modelo de *gaps* adaptativo, NGMLR puede alinear a través de variaciones estructurales reales manteniendo a la vez sensibilidad a *indels* pequeños (errores típicos de secuenciación) (LoTempio et al., 2023b).

Adicionalmente, NGMLR efectúa una búsqueda inicial *aware-SV* de posiciones aproximadas al identificar semillas k-mer y ser capaz de dividir lecturas que no alinean linealmente en la referencia, generando múltiples segmentos si es necesario. Esto permite mapear diferentes partes de una lectura a diferentes loci si la lectura contiene, por ejemplo, una inversión o translocación (Ahsan et al., 2023; Sedlazeck et al., 2018).

Originalmente desarrollado por Sedlazeck et al., (2018) junto al detector de SV Sniffles, NGMLR demostró lograr alta sensibilidad y precisión en la detección de SVs incluso en regiones complejas y repetidas del genoma. En el contexto mitocondrial, el interés de usar NGMLR radica en que ciertas patologías presentan grandes deleciones de ADNmt; además, en muestras con heteroplasmia en deleciones parciales, NGMLR podría alinear

lecturas que abarquen la ruptura. Si bien Minimap2 en versiones recientes también soporta alineaciones a través de *indels* de hasta ~100 kb, NGMLR aporta una confirmación adicional y potencialmente localiza con mayor precisión los puntos de quiebre de SV (Sedlazeck et al., 2018).

NGMLR se utiliza sobre todo para detectar o confirmar variantes estructurales, aunque existen otras opciones como LAST que puede manejar largas brechas, y recientemente minimap2 en v2.22+ soporta *indels* de gran tamaño (100 kb). También existen *mappers* SV específicos más nuevos como PAlign o Cactus Graph Mapper. Sin embargo, NGMLR por ser considerado uno de los mejores en detectar SV complejos en humanos, desarrollado junto con Sniffles (uno de los SV *callers* más utilizados), se vuelve una opción interesante para el alineamiento de lecturas largas y la detección de *indels* grandes en el ADNmt (*Philres/Ngmlr: NGMLR Is a Long-Read Mapper Designed to Align PacBio or Oxford Nanopore (Standard and Ultra-Long) to a Reference Genome with a Focus on Reads That Span Structural Variations*, n.d.).

Estudios comparativos han mostrado que NGMLR puede localizar *breakpoints* de deleciones/inversiones con mayor precisión que BWA-MEM o incluso Minimap2 en ciertos casos (Sedlazeck et al., 2018). Dado que uno de los intereses potenciales era captar deleciones parciales de ADNmt (p. ej. la deleción “común” de 4977pb en algunos síndromes), se puede hacer uso de NGMLR para asegurarse de mapear lecturas que cruzan por el punto de deleción correctamente, algo que *mappers* sin *gap-cost* convexo podrían alinear subóptimamente o dividir incorrectamente.

En la tabla 2, se puede observar cómo Minimap2 destaca por su rendimiento general, mientras Winnowmap2 aporta precisión adicional en secuencias repetitivas (pertinente si se considera mapeos contra el genoma nuclear completo con NUMTs). NGMLR es más lento pero crucial para detectar alineamientos a través de variantes estructurales, siendo ideal para capturar grandes deleciones.

Herramienta	Tipo de tarea	Precisión (alineamiento/llamado)	Sensibilidad	Compatibilidad ONT	Facilidad de uso	Recursos computacionales	Detección de heteroplasmia / SVs
Minimap2	Alineamiento (lecturas largas)	Muy alta precisión en alineamiento. Excelente desempeño para lecturas largas ONT. Minimiza errores sistemáticos en SNPs/ <i>indels</i> pequeños gracias a <i>presets</i> como -x map-ont (Li, 2018).	Alta. Detecta correctamente la mayoría de <i>matches</i> >15% de similitud, aunque su precisión decae en regiones repetitivas o duplicadas.	Sí, <i>preset</i> nativo -x map-ont optimizado para ONT.	Muy fácil. Binario autónomo, con <i>presets</i> predeterminados.	Muy bajo consumo. Tiempo de alineamiento <3 min para mtDNA; RAM <8 GB.	SVs: Sí, permite detección de grandes inserciones/deleciones hasta ~100 kb (v2.22+). Heteroplasmia: No diseñado para detectar variantes a nivel alélico. No estima frecuencias alélicas.
Winnowmap2	Alineamiento (regiones repetitivas)	Alta precisión en regiones complejas del genoma. Minimiza falsos positivos en loci repetidos al filtrar k-mers frecuentes (Jain et al., 2022).	Equivalente a Minimap2 en regiones no repetitivas; superior en duplicaciones y zonas de baja complejidad.	Sí. Basada en Minimap2 con mejoras para lecturas ONT. Compatible con cualquier versión reciente de ONT.	Moderada. Requiere preprocesamiento con meryl para construir los k-mers únicos.	Moderado. Preprocesamiento puede tomar tiempo, pero alineamiento es rápido.	SVs: Detección indirecta, similar a Minimap2. Heteroplasmia: No aplicable (solo <i>mapper</i> ; no modela mezcla alélica).

NGMLR	Alineamiento (variantes estructurales)	Alta precisión para SVs. Utiliza un modelo de costos convexos para alinear lecturas con gaps grandes. Bien evaluado para lecturas con reordenamientos complejos (Sedlazeck et al., 2018).	Muy alta para lecturas que contienen SVs. Sensible a inversiones, translocaciones y deleciones grandes.	Compatible ONT con modo -x ont. Desempeño robusto en lecturas ruidosas.	Fácil. Comando standalone, aunque menos user-friendly que Minimap2.	Alto. Hasta 90 minutos para genoma completo. Para mtDNA, ~3-5 minutos. RAM: 10 GB.	SVs: Excelente detección en variantes complejas. Heteroplasma: No aplicable. No cuantifica proporciones alélicas ni soporta modelos mixtos.
Mutect2 (GATK)	Llamado de variantes somáticas/mitocondriales	Alta precisión en SNVs/ <i>indels</i> cortos tras filtros somáticos (FilterMutectCalls). Basado en pruebas bayesianas y modelado de errores (Benjamin et al., 2019).	Muy alta sensibilidad para alelos minoritarios $\geq 1-2\%$ AF. Diseñado para entornos de alta profundidad (como mtDNA).	Compatible ONT, aunque requiere buena calidad de mapeo y profundidad elevada. Funciona mejor con – mitochondria-mode.	Moderada. Requiere pasos múltiples: preprocesamiento, llamado, filtrado. Familiaridad con GATK es recomendable.	Requiere CPU alto para modo ensamblado. RAM y tiempo moderado para mtDNA (minutos).	Heteroplasma: Sí. Detecta alelos minoritarios desde 1%. SVs: No diseñado para variantes >50 bp. Solo SNVs e <i>indels</i> pequeños.

Medaka	Llamado de consenso (ONT)	Alta precisión en consenso SNPs ($\geq 98\%$). Entrenado sobre datos ONT reales, minimiza errores sistemáticos específicos de esta tecnología (Oxford Nanopore, 2021).	Alta, aunque tiende a omitir alelos $< 5\text{--}10\%$ por su modelo basado en consenso mayoritario.	Sí. Entrenado explícitamente con múltiples químicas ONT.	Muy fácil. Comando único con modelos preentrenados disponibles.	Bajo a moderado. Decenas de segundos para mtDNA; puede usar GPU.	Heteroplasmia: Detecta variantes principales, pero no cuantifica frecuencias alélicas bajas. SVs: Limitado a <i>indels</i> pequeñas. No diseñado para eventos > 50 bp.
Mutserve2	Llamado de variantes (mtDNA específico)	Alta precisión en variantes mitocondriales. Filtros específicos reducen falsos positivos. Anotación de hotspots e integración con haplogrupos (Weissensteiner et al., 2016).	Sensibilidad $\geq 1\%$ AF. No reporta variantes por debajo del umbral por defecto (1%).	Sí, optimizado para mtDNA. Compatible con ONT tras alineamiento correcto.	Muy fácil. Binario único. Integración nativa con Haplogrep.	Muy bajo. Segundos por muestra. Solo requiere BAM ordenado.	Heteroplasmia: Sí. Hasta 1% AF con modelos de verosimilitud. SVs: Opciones beta permiten detección de pequeñas inserciones/delecciones. No apto para grandes SVs.
Clair3	Llamado de variantes (deep learning híbrido:	Muy alta precisión en SNVs e <i>indels</i> pequeños (1–50 bp), incluso en regiones con errores sistemáticos. Supera a Medaka y métodos	Alta sensibilidad para variantes $\geq 1\%$ de AF. Detección incluso en regiones homopoliméricas.	Sí. Totalmente entrenado para ONT y PacBio HiFi. Modelos para distintas	Moderada. CLI; requiere pasos previos de mapeo y	Rápido para mtDNA: 1–5 min por muestra. Uso de GPU acelera procesos. RAM < 8 GB.	Heteroplasmia: Sí. Alta precisión $\geq 1\%$ AF. VCF incluye GT, DP, AD y AF. SVs: No diseñado para SVs grandes (> 50

	pileup + full-alignment)	previos como Clair (Zheng et al., 2021; Luo et al., 2022).	Compatible con ONT Q20+ y >500× mtDNA.	químicas (e.g., Q20+).	BAM ordenado. Buen soporte documental y scripts útiles.		bp), pero sí para <i>indels</i> cortos.
--	--------------------------	--	--	------------------------	---	--	---

Tabla 2. Comparación detallada de diversas herramientas bioinformáticas utilizadas para el análisis de variantes en ADN mitocondrial secuenciado mediante plataformas de lectura larga como ONT. Se evalúan tanto *mappers* como *callers* de variantes en función de su precisión, sensibilidad, compatibilidad con ONT, facilidad de uso, requerimientos computacionales y capacidad para detectar heteroplasmia o variantes estructurales (SVs). Esta síntesis permite identificar las combinaciones más empleadas para estudios mitocondriales, considerando aspectos clave como la detección de variantes de baja frecuencia, la eficiencia operativa y la resolución en regiones complejas del genoma. La comparación sirve como base para seleccionar pipelines óptimos en contextos clínicos o de investigación genómica avanzada.

1.2.5.4. Llamado de variantes

Mutect2

Mutect2 es un algoritmo originalmente desarrollado por el *Broad Institute* como parte de GATK, orientado a la detección de variantes somáticas en cáncer. Su fundamento es un modelo Bayesiano que compara un conjunto de lecturas de “tumor” con las de un control “normal” para identificar variantes presentes en una fracción de células (subclonales). En el contexto mitocondrial, Mutect2 se emplea en un modo *tumor-only* adaptado, aprovechando su alta sensibilidad a alelos de baja frecuencia incluso sin muestra control (Benjamin et al., 2019).

De hecho, el equipo de gnomAD desarrolló un “modo mitocondria” de Mutect2 específicamente ajustado para llamadas en ADNmt (Ip et al., 2022). Este modo especializado modifica varios parámetros para enfrentar los retos del ADNmt, por ejemplo, reduce los umbrales de detección de regiones activas y de calidad mínima de variante, permitiendo capturar señales débiles en posiciones de heteroplasmia baja. Asimismo, Mutect2 realiza un ensamblado local de haplotipos (similar al de HaplotypeCaller) y aplica un algoritmo adaptativo de “corte de caminos en el grafo de ensamblado”, calibrado según la cobertura local y la tasa de error, para evitar un exceso de haplotipos falsos cuando la cobertura es extremadamente alta. Esto último es crucial, pues sin dicha poda adaptativa la enorme profundidad del ADNmt podría inducir falsos positivos al confundir errores sistemáticos repetidos como variantes reales (Benjamin et al., 2019; Ip et al., 2022).

Otra adaptación importante de Mutect2 al ADNmt es la relajación o eliminación de ciertos filtros diseñados para el genoma nuclear. Por ejemplo, en modo somático convencional se filtran “eventos agrupados” o variantes multialélicas cercanas entre sí, asumiendo que las mutaciones somáticas suelen ser escasas y separadas. En el ADNmt, sin embargo, pueden existir múltiples variantes próximas (particularmente entre haplogrupos distintos), por lo que el modo mitocondrial de Mutect2 suprime filtros como *clustered_events* y *multiallelic* para no descartar variantes verdaderas que aparecen juntas (Broadinstitute/Gatk: Official Code Repository for GATK Versions 4 and Up, 2025).

Adicionalmente, incorpora una lista predefinida de sitios propensos a artefactos (p. ej. las posiciones 301-316 y 16182 del rCRS), que corresponden a tramos homopoliméricos o regiones problemáticas conocidas en secuenciación Illumina. Cualquier variante detectada en dichos sitios es marcada como sospechosa (*artifact_prone_site*) y típicamente filtrada en el análisis final. Este enfoque atiende a problemas clásicos como la región hipervariable alrededor de la posición 310, donde secuenciaciones Illumina generan errores de inserción/delección que podrían falsear variantes si no se excluyen (Broadinstitute/Gatk: Official Code Repository for GATK Versions 4 and Up, 2025; GnomAD v3.1 Mitochondrial DNA Variants | GnomAD Browser, 2025).

En cuanto a su desempeño en ADNmt, Mutect2 ha demostrado alta precisión en la detección de variantes heteroplásmicas moderadas a altas. En un análisis comparativo con el algoritmo ADNmt-Server (basado en MutServe), Mutect2 mostró mayor precisión (menos falsos positivos) a todos los niveles de heteroplasma y sensibilidad equivalente para variantes con frecuencia alélica entre 5% y casi 100%. Esto indica que Mutect2 tiende a reportar variantes con mayor confianza, minimizando llamadas inciertas incluso en muestras muy profundas. No obstante, el mismo estudio observó que Mutect2 presentaba menor sensibilidad para detectar variantes en el ~1% de heteroplasma (Benjamin et al., 2019; Laricchia et al., 2022).

Es decir, algunas variantes de muy baja frecuencia que herramientas especializadas lograron rescatar, Mutect2 pudo pasarlas por alto debido a sus filtros de calidad más estrictos. Para mitigar este aspecto, es posible ajustar manualmente los parámetros de Mutect2 (por ejemplo, reducir aún más el umbral `--min-af` de frecuencia alélica mínima) o emplear estrategias combinatorias. De hecho, se han propuesto pipelines híbridos donde los resultados de Mutect2 se complementan con otros *callers* mitocondriales, por ejemplo, ADNmt-Server 2 implementa un “llamado por fusión” combinando las salidas de Mutect2 y MutServe2, aprovechando la sensibilidad de este último en heteroplasmas bajas y la capacidad superior de Mutect2 para detectar inserciones/delecciones (Benjamin et al., 2019; Ip et al., 2022).

Medaka

Medaka es una herramienta desarrollada por ONT para la detección de variantes a partir de datos de secuenciación de lectura larga. A diferencia de los *callers* tradicionales basados en frecuencias alélicas simples, Medaka emplea redes neuronales profundas entrenadas específicamente para modelar los errores de ONT y producir una secuencia consenso altamente precisa. Su enfoque es esencialmente basado en el consenso, inicialmente tomando las lecturas alineadas y generando una secuencia consenso diploide usando un modelo de aprendizaje profundo que incorpora la pila de lecturas (*pileup*) (Helal et al., 2022).

En una segunda iteración, Medaka refina este consenso generando dos secuencias haploides (faseadas) que representan las variantes de cada haplotipo. Para lograr esto, Medaka primero predice diferencias a nivel de SNP (polimorfismos de un solo nucleótido) en las lecturas sin fasear (*unphased*) y luego emplea una herramienta de *phasing* (WhatsHap) para agrupar las lecturas en haplotipos (Helal et al., 2022). Finalmente, realiza el llamado de SNPs e *Indels* en cada grupo de lectura faseado, comparando las secuencias consenso haplotípicas contra la referencia. Este procedimiento permite aprovechar la información de fase de las lecturas largas, mejorando la exactitud en variantes complejas y reduciendo errores sistemáticos característicos de ONT (Nanoporetech/Medaka: Sequence Correction Provided by ONT Research, 2025).

Una de las fortalezas de Medaka es que fue durante varios años la herramienta recomendada por ONT para *variant calling* en humanos utilizando secuenciación de tercera generación (Helal et al., 2022). Los pipelines oficiales de ONT integraban el argumento `medaka_variant`, para identificar variantes de pequeño tamaño (SNVs e *indels* cortos) en genomas haploides o diploides. En el contexto del ADNmt, que es esencialmente haploide (aunque con poblaciones mezcladas de moléculas), Medaka puede detectar heteroplasma tratándola de forma análoga a una heterocigosidad en un genoma diploide, por ejemplo, si una variante está presente en el 30% de las lecturas, Medaka probablemente la representará como diferencia entre los consensos haploides (uno llevará el alelo de referencia, el otro el alternativo) (Macken et al., 2023).

Sin embargo, Medaka presenta limitaciones en el panorama actual debido a la aparición de algoritmos de aprendizaje profundo más avanzados (por ejemplo, Clair, Clairvoyante y especialmente Clair3), donde se ha reportado que Medaka es superado en precisión, sensibilidad y velocidad (Abdelwahab & Torkamaneh, 2025; Osada et al., 2024). Estudios comparativos indican que Clair3 – un *caller* híbrido que combina estrategias de *pileup* y alineamiento completo – alcanza mejor desempeño general en variantes ONT, al punto que ONT ha deprecado Medaka en favor de Clair3 en sus pipelines recientes (Abdelwahab & Torkamaneh, 2025).

En evaluaciones sobre genomas de referencia, Medaka obtuvo puntajes F1 inferiores a métodos contemporáneos, reflejando que su tasa de error (particularmente falsos negativos) es mayor, posiblemente porque su modelo de consenso puede “suavizar” o ignorar variantes presentes en muy baja proporción (Abdelwahab & Torkamaneh, 2025). En el contexto de heteroplasma mitocondrial, esto sugiere que Medaka podría no detectar variantes por debajo de cierto umbral de frecuencia (siendo conservador al construir el consenso). De hecho, herramientas especializadas de segunda generación (Clair3, Nanocaller, etc.) han mejorado la detección de variantes minoritarias al combinar evidencias de *pileup* y re-alineamiento local (Korolija et al., 2024; Parakatselaki et al., 2021)

Antes de Medaka, la principal herramienta para refinamiento y llamadas en ONT era Nanopolish, que analiza las señales eléctricas crudas para mejorar la precisión de bases. Si bien Nanopolish logra buena exactitud, es computacionalmente costoso y complejo de integrar (requiere los archivos FAST5). Medaka opera directamente sobre lecturas *basecalled* y ha demostrado tener desempeño comparable o mejor en llamadas de SNPs, con gran ventaja en velocidad (Wick et al., 2019).

Por tanto, Medaka fue un pionero en aplicar *deep learning* a las lecturas largas para llamado de variantes y es capaz de abordar el ADNmt (de hecho, su uso era recomendado para SNPs mitocondriales antes del 2021), se decide utilizarlo dentro de las herramientas para selección de método. Al aplicarlo a datos ONT de ADNmt, se espera obtener un conjunto de variantes confiables para alelos de frecuencia intermedia a alta, con un buen manejo de *indels* cortos; no obstante, podrían escaparse variantes con heteroplasma muy baja (<5%) y su rendimiento está supeditado a que las lecturas tengan calidad suficiente y estén bien alineadas. En la actualidad, se recomienda validar las llamadas de Medaka o

complementarlas con métodos más nuevos, especialmente si el objetivo es capturar heteroplasma de nivel bajo.

Clair3 (familia Clair; HKU/colab.)

Clair3 es un *caller* de variantes basado en aprendizaje profundo de la familia Clair/Clairvoyante, diseñado principalmente para lecturas largas (Oxford Nanopore y PacBio HiFi). A diferencia de los enfoques heurísticos o bayesianos clásicos, Clair3 incorpora una arquitectura de dos vías que combina (i) un modelo rápido basado en *pileup* para sondeo inicial de sitios candidatos y (ii) un modelo de alineamiento completo (*full-alignment*) con información haplotípica para resolver casos complejos y refinar genotipos. Esta estrategia híbrida aprovecha tanto la señal resumida del *pileup* como los patrones locales de desajuste/*indel* a lo largo de cada lectura y su fase, lo que se traduce en un aumento notable de la precisión frente a métodos tradicionales en datos ruidosos de lecturas largas (Luo et al., 2020; He et al., 2021; Luo et al., 2022).

Clair3 procesa lecturas alineadas en torno a variantes candidatas convirtiéndolas en representaciones matriciales que integran múltiples capas de información, incluyendo las bases observadas, la calidad asociada a cada posición, las operaciones de alineamiento (CIGAR) y los datos de fase si están disponibles. Estas representaciones son utilizadas como entrada para una arquitectura de redes neuronales profundas que combina capas convolucionales y unidades densas, las cuales han sido específicamente entrenadas para datos generados por plataformas como ONT y PacBio (Greig et al., 2019).

El análisis comienza con un modelo de tipo *pileup*, que identifica y clasifica rápidamente las posibles variantes en regiones simples, mientras que aquellas ubicadas en contextos complejos —como homopolímeros o regiones con *indels*— son reenviadas a una segunda etapa de análisis más detallado basada en alineamientos completos. En esta fase, se incorporan herramientas de *phasing* como *Margin* o *WhatsHap* para afinar la resolución alélica. El resultado final del proceso es un archivo VCF que contiene variantes de tipo SNP e *indels* pequeños con un alto grado de confianza, caracterizado por tasas reducidas de falsos positivos, siempre que se utilicen modelos entrenados con parámetros acordes al tipo de química y tecnología de *basecalling* empleadas (Zheng et al., 2022).

En genomas compactos y de alta cobertura como el mitocondrial (aproximadamente 16,6 kb), Clair3 demuestra un rendimiento sobresaliente en la detección de variantes homoplásmicas y de heteroplasma moderada, especialmente con secuenciación ONT de alta calidad (por ejemplo, química Q20+). El modelo de salida de Clair3 está naturalmente diseñado para representar mezclas alélicas; un genotipo intermedio, como el "0/1", puede reflejar una heteroplasma en niveles medios. De esta forma, Clair3 resulta especialmente eficaz para detectar SNPs y pequeñas inserciones/deleciones cuando la frecuencia alélica (AF) excede el umbral práctico de la plataforma (Zheng et al., 2022).

El potencial de Clair3 se sustenta en su validación en genomas completos con datos de ONT: en pruebas de referencia con 85× de cobertura en ONT, Clair3 alcanzó puntuaciones F1 de 99,69 % para SNPs y 80,58 % para *indels*, superando considerablemente a versiones anteriores como Clair, Medaka o métodos sin conciencia de haplotipo (Zheng et al., 2022). En contextos especializados de ADN mitocondrial, Clair3 aún no ha sido extensamente evaluado, pero estudios sobre variantes de baja frecuencia en mtDNA mediante ONT han demostrado que alcanzan altas puntuaciones F1 empleando *callers* robustos como Mutserve2 (por ejemplo, > 0,99 al 5 % de heteroplasma) (Lüth et al., 2022; Su et al., 2022; Z. Zheng et al., 2025). Estos resultados sugieren que Clair3, con su arquitectura innovadora que integra modelos *pileup* y *full-alignment*, podría igualar o superar esta precisión.

Algunas de las limitaciones que tiene el paquete son que Clair3 está optimizado primordialmente para lecturas largas; si bien existen variantes/ajustes para lectura corta, la herramienta de referencia para Illumina en *deep learning* sigue siendo DeepVariant (Poplin et al., 2018). Por ello, en corto alcance (ADNmt por Illumina) conviene validar con *callers* probados en *short reads* o usar Clair3-short cuando proceda. La vía *full-alignment* es más costosa (beneficia de GPU); en ADNmt, el tamaño reducido vuelve este coste manejable. Para AF < ~5% la sensibilidad decrece si no se incrementa la cobertura y/o no se reentrena el modelo para ese régimen, con el riesgo de aumentar falsos positivos. Es recomendable inspección manual (IGV) y validación ortogonal. La interpretabilidad es limitada; se aconseja acompañar con métricas de soporte por hebra y context QC.

Mutserve2

MutServe2 es una herramienta de llamado de variantes específica para el genoma mitocondrial, diseñada para detectar tanto variantes homoplásmicas como heteroplásmicas en datos de secuenciación. Constituye la evolución de MutServe (v1) – presentado originalmente en 2016 como parte del servidor web ADNmt-Server (Weissensteiner et al., 2016) – incorporando mejoras de rendimiento y nuevas funcionalidades para análisis locales y de nueva generación (Ip et al., 2022). A diferencia de los *callers* generalistas (Mutect2, HaplotypeCaller, etc.), MutServe2 está optimizado para la genética mitocondrial realizando controles de calidad internos específicos, como excluir posiciones hipervariables o conflictivas (“*hotspots*” mitocondriales) y sitios con cobertura muy baja (<10 lecturas), dado que estos suelen asociarse a errores o problemas de mapeo (Ip et al., 2022; Weissensteiner et al., 2024).

Asimismo, MutServe2 calcula métricas de sesgo de cadena (*strand bias*) procesando por separado las lecturas mapeadas en cada orientación (directa vs inversa), lo que ayuda a descartar falsas variantes que aparecen solo en una de las hebras de secuenciación (Bravo-Alonso et al., 2019). Para el llamado de heteroplasmia, MutServe (v1) utilizaba un umbral de frecuencia alélica mínimo por defecto de ~1% – es decir, requería que el alelo alternativo representara al menos el 1% de las lecturas en ese sitio para considerarlo una variante heteroplásmica válida (Ip et al., 2022; Weissensteiner et al., 2016). Este umbral surge del balance entre sensibilidad y precisión; por debajo de 1%, la probabilidad de que una variante aparente sea en realidad ruido de secuenciación es alta (especialmente en Illumina, cuyo error base ronda 0.1–1%). MutServe2 permite ajustar este umbral y otros parámetros, dando flexibilidad para detectar variantes a niveles inferiores si la calidad de los datos lo permite, o para aplicar criterios más estrictos en entornos clínicos (Weissensteiner et al., 2024).

Una de las fortalezas destacadas de MutServe2 es su excelente desempeño en la detección de variantes de baja frecuencia en ADNmt, demostrado en evaluaciones comparativas. Por ejemplo, Lüth et al., (2022) realizaron una referenciación con mezclas sintéticas de genomas mitocondriales secuenciados por ONT (con heteroplasmas controladas al 5%, 2% y 1%) y compararon múltiples *callers* especializados. MutServe2 obtuvo los puntajes F1 más altos en todos los niveles de mezcla probados, alcanzando F1

>0.99 al 5% de heteroplasmia, ~0.70 al 1% y superando a herramientas como FreeBayes y NanoPanel2 en sensibilidad global (Lüth et al., 2022; Weissensteiner et al., 2024).

Esto indica que MutServe2 logra identificar una alta proporción de variantes verdaderas incluso cuando su frecuencia es muy baja, manteniendo a la vez un nivel aceptable de falsos positivos. De hecho, en ese estudio fue capaz de detectar consistentemente variantes ~1% en las muestras ONT, algo especialmente retador debido a la mayor tasa de error de las lecturas largas. Adicionalmente, MutServe2 mostró el mayor *call rate* (cantidad de variantes detectadas) y número de verdaderos positivos en comparaciones con otros algoritmos para ADNmt, tanto en datos simulados como en análisis de tríos familiares (Lüth et al., 2022). En escenarios de alta profundidad, MutServe2 aprovecha el gran número de lecturas para robustecer la evidencia de variantes raras, aunque también debe aplicar filtros para evitar artefactos provenientes de errores sistemáticos.

En cuanto a limitaciones, MutServe2 originalmente no manejaba con igual eficacia todos los tipos de variantes. La versión 1 se centraba en SNVs, y su detección de *indels* era limitada. Si bien MutServe2 ha incorporado mejoras, muchos flujos de trabajo (p. ej., ADNmt-Server 2) combinan su salida con la de Mutect2 precisamente para incluir *indels* con mayor confianza (Lüth et al., 2022; Weissensteiner et al., 2024). Esto sugiere que MutServe2 podría aún tener dificultades con inserciones/deleciones, especialmente en regiones repetitivas donde el alineamiento de *indels* es ambiguo. Otra consideración es que MutServe2, al basarse en conteos de lecturas y umbrales fijos, puede reportar más falsos positivos cuando la calidad de las bases es baja (Slapnik et al., 2024).

En aplicaciones prácticas, suele recomendarse filtrar las llamadas de MutServe2 contra bases de datos de variantes conocidas o validar las muy raras con métodos ortogonales. Pese a esto, la concordancia inter-plataforma de MutServe2 ha sido reportada como alta para variantes reales, variantes heteroplásmicas identificadas en Illumina, típicamente también son detectadas por MutServe2 en ONT si superan el umbral de frecuencia y viceversa, lo que refuerza su utilidad para análisis integrados (Jia et al., 2020; Slapnik et al., 2024).

Estudios recientes enfatizan que para lograr máxima confianza, es ideal comparar los llamados de variantes mitocondriales entre secuenciación corta y larga. Las variantes que

coinciden en ambas (especialmente si son heteroplásmicas de bajo nivel) casi con certeza reflejan mutaciones auténticas y no artefactos (Slapnik et al., 2024). Por otro lado, diferencias entre plataformas suelen señalar posibles errores: por ejemplo, una variante presente solo en ONT pero no en Illumina puede ser sospechosa de error de Nanopore, mientras que variantes detectadas por Illumina únicamente podrían haber sido perdidas en ONT por un *filtrarig* de calidad o por falta de suficiente cobertura efectiva (Cottingham et al., 2025; Stefan et al., 2022).

MutServe2 surge como una herramienta especializada y potente para ADNmt, ideal para sondear heteroplasmia a niveles que otros métodos pasan por alto. Al usarla con datos Illumina, ofrece un complemento a *callers* generales, proporcionando sensibilidad extra en muestras de alta profundidad; con datos ONT, su diseño altamente sensible a heteroplasmias la convierte en una buena herramienta para explotar la gran longitud de lectura sin sacrificar la detección de variantes minoritarias (Lüth et al., 2022; Weissensteiner et al., 2024).

Además, Mutserve incluye anotaciones de gran utilidad que no tienen otros paquetes bioinformáticos, como por ejemplo, identificar si una variante cae en una secuencia NUMT conocida (tiene una lista de NUMTs frecuentes) para alertar de posibles falsos positivos nucleares. La elección de Mutserve se basó en estos factores y en su facilidad de uso (es un ejecutable *standalone*, y sus resultados pueden subirse a herramientas como Haplogrep y Mitomaster directamente).

Es recomendable usar MutServe2 dentro de un marco de control de calidad estricto filtrando variantes según su soporte en ambas hebras, descartar las cercanas al umbral de detección que no sean reproducibles en replicas o en otra plataforma, y combinar su output con otras fuentes (ej. confirmar *indels* con Mutect2, o validar variantes clínicas con pruebas específicas). De esta forma, se obtendrá un conjunto de resultados esperados altamente eficiente donde todas las variantes homoplásmicas conocidas del individuo deberían ser identificadas, y además se revelarán variantes heteroplásmicas potencialmente relevantes (posiblemente hasta ~1% de frecuencia en muestras con cobertura >1000×), cuyo hallazgo e interpretación amplía nuestro entendimiento de la dinámica mitocondrial.

Resumiendo con respecto a las herramientas de llamado de variantes, Mutect2 —diseñado originalmente para variantes somáticas— demuestra un excelente desempeño para detectar alelos minoritarios, lo que lo convierte en una opción sólida para estudios de heteroplasmia. Medaka, entrenada específicamente sobre datos ONT, es eficaz para generar consensos de alta fidelidad, aunque menos sensible a variantes de baja frecuencia. Mutserve2, por su parte, está optimizado para análisis mitocondriales, integrando funciones como anotación de *hotspots* y clasificación de haplogrupos. Finalmente, Clair3 representa una alternativa basada en aprendizaje profundo que combina estrategias de *pileup* y alineamiento completo, ofreciendo una precisión comparable o superior a Medaka y destacándose por su buen rendimiento computacional incluso en dispositivos sin GPU (Tabla 2).

1.2.6. Análisis estadístico. Evaluación de herramientas.

1.2.6.1. Estadística básica

En el análisis de datos genómicos, las métricas de calidad como la profundidad de cobertura (DP), la calidad de base (BQ) y la frecuencia alélica (AF) suelen resumirse mediante estadística descriptiva, empleando medianas, intervalos intercuartílicos (IQR) y representaciones gráficas como diagramas de caja. Este enfoque permite comparar distribuciones entre plataformas y pipelines sin que los valores extremos distorsionen los resultados (Altman & Bland, 1995). En R, estas visualizaciones y resúmenes pueden obtenerse con funciones de base R y librerías como ggplot2, ampliamente utilizadas en bioinformática para la exploración inicial de datos de variantes (Wickham, 2016).

1.2.6.2. Análisis de componentes principales

El Análisis de Componentes Principales (PCA) es una técnica multivariada de reducción de dimensiones que facilita identificar patrones de variación y agrupamientos en datos complejos. En el contexto genómico, su uso permite explorar similitudes y diferencias entre muestras, pipelines o plataformas, a partir de variables cuantitativas como AF, DP o BQ (Jolliffe & Cadima, 2016). En este trabajo se emplearon funciones de stats::prcomp y

visualizaciones generadas con ggplot2 y FactoMineR, herramientas de uso común en estudios de secuenciación para detectar clústeres y estructuras latentes en los datos (Lê et al., 2008).

1.2.6.1. Índice de Jaccard

La similitud entre pipelines y plataformas de secuenciación se evaluó mediante el índice de Jaccard, ampliamente utilizado en bioinformática para comparar conjuntos discretos de variantes. Este índice se define como la proporción entre la intersección y la unión de dos conjuntos ($|A \cap B| / |A \cup B|$), donde A y B corresponden a las posiciones de variantes detectadas por cada plataforma o pipeline. Así, un valor cercano a 1 indica una alta concordancia en las variantes reportadas, mientras que valores próximos a 0 reflejan un solapamiento mínimo. En el contexto de este trabajo, el índice permite cuantificar de manera objetiva la similitud cualitativa entre combinaciones de análisis en ONT y los resultados de Illumina, aportando una métrica clara y reproducible para identificar los *pipelines* con mayor coherencia respecto a la plataforma de referencia (Real & Vargas, 1996).

1.2.7. Comparación de plataformas: concordancia y correspondencia de resultados

En bioinformática comparativa, al evaluar dos plataformas de secuenciación o pipelines de análisis, es fundamental medir qué tanto coinciden sus resultados (concordancia) y cómo se relacionan o difieren globalmente (correspondencia). Un enfoque común es cuantificar la superposición de variantes detectadas por ambas plataformas mediante métricas como el índice de Jaccard, y que se pueden visualizar por medio de mapas de calor y gráficos de circos, y que calcula la proporción de variantes compartidas sobre el total de variantes detectadas por cualquiera de los métodos (Ren et al., 2023). Un valor de Jaccard cercano a 1 indicaría alta concordancia (la mayoría de variantes son comunes a ambos), mientras que valores bajos denotan divergencia significativa en las llamadas de variantes de cada plataforma. Estas medidas de concordancia permiten responder preguntas clave, por

ejemplo: ¿Cuántas variantes identificadas con Oxford Nanopore (ONT) aparecen también con Illumina, y viceversa?

Además de la coincidencia de llamadas, suele explorarse la correspondencia multivariada de los perfiles de variantes o métricas entre plataformas. Esto implica ver si, en conjunto, los patrones de datos producidos por cada tecnología son similares o forman grupos diferenciados (Díaz Monroy & Morales Rivera, 2009). En este contexto se emplean técnicas de análisis multivariante que ayudan a visualizar y probar diferencias globales, como la ordenación por escalar multidimensional no métrico (NMDS) y la ANOVA permutacional (PERMANOVA), descritas más adelante. Estas herramientas van más allá de la simple coincidencia de variantes, permitiendo apreciar relaciones generales entre los resultados de distintos pipelines (por ejemplo, si las muestras se agrupan naturalmente por plataforma en función de múltiples características a la vez) (Ren et al., 2023).

1.2.7.1. Gráficos de circos

Para visualizar la distribución de variantes y métricas asociadas a lo largo del genoma mitocondrial, se emplearon gráficos circulares tipo circos, implementados en el paquete circlize de R (Gu et al., 2014). Esta librería permite representar el genoma en disposición circular y superponer diferentes capas de información (profundidad, calidad, posiciones de variantes), facilitando la identificación de regiones con acumulación de variantes o cobertura atípica. El uso de circos plots es particularmente útil en ADN mitocondrial, dado su tamaño compacto y estructura circular.

1.2.7.2. Índice de Jaccard y medidas de similitud de conjuntos de variantes

Una forma básica pero poderosa de evaluar concordancia entre dos métodos de secuenciación es comparar los “conjuntos de variantes” que cada uno detecta en una muestra. El índice de Jaccard es ampliamente utilizado para este fin en estudios genómicos comparativos. Este índice se define como el tamaño de la intersección dividido

por el tamaño de la unión de los conjuntos de variantes de ambas plataformas. En términos sencillos, mide la fracción de variantes compartidas entre dos listas con respecto al total de variantes encontradas por al menos uno de los métodos. Un Jaccard = 1.0 indicaría concordancia perfecta (ambos hallaron exactamente las mismas variantes), mientras que un valor 0 indicaría que no comparten ninguna variante (Anderson, 2001). En la práctica, valores intermedios son comunes; por ejemplo, en un estudio reciente el índice de Jaccard se empleó para cuantificar la concordancia de llamadas de variantes germinales entre distintos pipelines, interpretándolo como la consistencia de detección de variantes entre plataformas. Esta métrica permite identificar qué tan solapados están los resultados y, por ende, cuán comparables son las plataformas en términos de variantes detectadas (Lesurf et al., 2025).

Otras medidas relacionadas pueden incluir la concordancia genotípica (porcentaje de variantes donde ambas plataformas detectan la misma variante y el mismo genotipo), o coeficientes de correlación para variables continuas derivadas (por ejemplo, correlación de frecuencias alélicas llamadas por cada método). No obstante, el índice de Jaccard destaca por su simplicidad y claridad para presencia/ausencia de variantes, razón por la cual aparece frecuentemente en evaluaciones de concordancia en variantes de secuenciación (Anderson, 2001). Por ejemplo, Green et al., (2023) utilizan Jaccard para reportar la fracción de variantes germinales compartidas entre distintos métodos de secuenciación, como indicador directo de concordancia.

1.2.7.3. NMDS: Ordenación multidimensional no métrica

Para evaluar diferencias o similitudes globales en los perfiles de variantes de dos plataformas, es útil recurrir a técnicas de ordenación multivariante. Una de las más empleadas es la ordenación no métrica multidimensional (Non-metric MultiDimensional Scaling, NMDS). NMDS es un método de análisis multivariante que toma una matriz de distancias o disimilitudes entre muestras y construye una representación en pocas dimensiones (típicamente 2D) que preserva lo mejor posible el orden de esas distancias

(*Non-Metric Multidimensional Scaling (NMDS) in Microbial Sequencing Data Analysis: Introduction, Application, and Comparison* - CD Genomics, n.d.).

A diferencia de métodos lineales como PCA, NMDS se basa solo en los rangos de las distancias (no en sus valores absolutos), por lo que puede usar cualquier medida de disimilitud adecuada para los datos (por ejemplo, distancias de Jaccard o Bray-Curtis entre perfiles de variantes). Es una técnica ampliamente utilizada para visualizar relaciones en datos de alta dimensión en bioinformática, permitiendo revelar patrones como agrupamientos de muestras según sus características genómicas.

En el contexto de comparar ONT vs Illumina, uno podría construir un vector de características por muestra (por ejemplo, presencia/ausencia o frecuencias de cada variante, u otras métricas agregadas), calcular una matriz de disimilitud entre muestras y luego aplicar NMDS para obtener un mapa bidimensional. En dicho mapa, cada punto representaría una muestra secuenciada, y la distancia entre puntos refleja la diferencia en el perfil de variantes entre muestras. Si las plataformas producen resultados muy similares, cabría esperar que las muestras no se separen por plataforma en la ordenación.

Por ejemplo, Taguchi & Oono, (2005) ilustraron la eficacia de NMDS para revelar patrones temporales en datos de expresión génica masivos, lo que evidencia su potencial para extraer relaciones complejas también en datos de variantes. Por lo cual, NMDS ofrece una visualización intuitiva de las diferencias globales entre pipelines, sirviendo como paso exploratorio previo a pruebas estadísticas formales.

1.2.7.4. PERMANOVA: Prueba multivariante por permutaciones

Para complementar la visualización y determinar rigurosamente si las diferencias observadas entre plataformas son estadísticamente significativas, se utiliza la PERMANOVA (*Permutational Multivariate Analysis of Variance*). La PERMANOVA (implementada, por ejemplo, en la función `adonis` del paquete `vegan` de R) es una prueba no paramétrica por permutaciones que evalúa si la distancia promedio entre grupos (e.g.,

muestras Illumina vs muestras ONT) es mayor que la distancia dentro de los grupos, bajo la hipótesis nula de que no hay diferencias entre grupos (Broom et al., 2023; Cubino et al., 8268). En esencia, es un análogo multivariado del ANOVA tradicional, pero en lugar de trabajar con una variable dependiente única, opera sobre una matriz de disimilitudes arbitraria entre observaciones (Broom et al., 2023). Al ser no paramétrica, no asume distribuciones normales multivariadas, lo cual es ventajoso para datos genómicos que suelen violar los supuestos de MANOVA clásica.

Conceptualmente, la PERMANOVA pregunta: *¿Podrían las diferencias observadas entre los perfiles de variantes de Illumina vs ONT explicarse por azar?* Para responder, calcula un estadístico pseudo-F (análogo al F de ANOVA) a partir de las sumas de cuadrados entre grupos y residuales en el espacio de distancias, y luego estima la significación mediante permutaciones aleatorias de las etiquetas de grupo (Broom et al., 2023; Cubino et al., 8268). Si el resultado muestra una F significativamente mayor que la esperada al azar ($p < 0.05$), se rechaza la hipótesis nula, concluyendo que las plataformas tienen efectos diferenciales en los perfiles multivariados de variantes (Broom et al., 2023).

1.2.7.5. Pruebas estadísticas clásicas: t-test, ANOVA y Wilcoxon

Además del análisis multivariante, las comparaciones entre plataformas suelen involucrar pruebas estadísticamente inferenciales univariadas para métricas específicas de interés. Entre las más comunes están la prueba t de Student, el ANOVA univariado y la prueba no paramétrica de Wilcoxon. Prueba t de Student: permite comparar la diferencia de medias de una variable continua entre dos grupos (Illumina vs ONT) bajo el supuesto de datos normalmente distribuidos y varianzas homogéneas (Chicco et al., 2025). En estudios genómicos se aplica, por ejemplo, para contrastar si la profundidad de lectura promedio difiere entre plataformas, o si la calidad media de las llamadas de variante (score de calidad) es mayor en una plataforma que en otra. Es una prueba paramétrica muy potente cuando se cumplen sus supuestos, y su estadístico t junto con sus grados de libertad permite calcular un p-valor para la diferencia de medias. Si solo hay dos grupos, la prueba

t y el ANOVA de una vía son equivalentes en resultado; este último se usa más cuando hay >2 grupos (Chicco et al., 2025; Mishra et al., 2019).

El ANOVA (Analysis of Variance) en contextos de múltiples pipelines o factores adicionales, de una o varias vías es útil para probar diferencias en medias entre más de dos grupos. Por ejemplo, si se compararan varias estrategias de *pipeline* (varios algoritmos de llamado de variantes) además de la plataforma, un ANOVA factorial podría evaluar el efecto de cada factor y sus interacciones en métricas como el número de variantes detectadas o el porcentaje de regiones cubiertas (Kim, 2014). El ANOVA produce un estadístico F y un p-valor asociado, evaluando la hipótesis nula de que todas las medias de grupo son iguales. Sin embargo, requiere condiciones de normalidad y homocedasticidad; cuando estas no se cumplen, se recurre a otras alternativas (Kim, 2014; McHugh, 2011).

La Prueba de Wilcoxon, también conocida como prueba de rango con signo de Wilcoxon (o Mann-Whitney U si las muestras son independientes), es la contraparte no paramétrica de la t de Student para dos grupos. Se basa en los rangos de los datos en lugar de los valores crudos, y no asume normalidad. En nuestro caso, muchas métricas derivadas (p. ej., distribuciones de frecuencia alélica (AF) de variantes o profundidad de cobertura (DP) en posiciones variantes) pueden no seguir una distribución normal, por lo que empleamos Wilcoxon para comparar su mediana entre plataformas (Fay & Proschan, 2010).

Por ejemplo, Tshiabuila et al., (2022) utilizaron pruebas de Wilcoxon para comparar el número de mutaciones detectadas y otras métricas de cobertura entre secuenciaciones Illumina MiSeq y ONT GridION, justificando su uso por la naturaleza no paramétrica de los datos. En general, un p-valor bajo en Wilcoxon indicaría que las distribuciones difieren significativamente (p. ej., “la mediana de AF en ONT es significativamente distinta a la de Illumina”), mientras que un p alto sugiere que no hay evidencia de diferencia sistemática. Cada una de estas pruebas produce un estadístico (t, F o W/U) y un p-valor. Es importante destacar que en estudios de comparación de plataformas, a menudo se aplican

correcciones por comparaciones múltiples si se realizan muchas pruebas a la vez (por ejemplo, comparando varias variables), para controlar el error de Tipo I.

1.2.7.6. Medidas de efecto: tamaño de la diferencia (d de Cohen)

Más allá de la significación estadística (el p-valor), es crucial en análisis comparativos reportar la magnitud del efecto o tamaño de la diferencia, para entender la relevancia práctica de las diferencias entre plataformas (Cohen, 2011; Rendón-Macías et al., 2021). Una de las métricas de efecto más utilizadas es la *d* de Cohen, que cuantifica la diferencia entre dos medias en unidades de desviación estándar. En términos simples, Cohen's *d* indica cuántas desviaciones estándar separan las medias de Illumina y ONT para una cierta variable. Por ejemplo, un $d = 0.0$ indicaría que las medias son iguales, $d = 0.5$ representa una diferencia de media de medio sigma (tamaño del efecto mediano), y $d = 1.0$ indicaría que una media está a una desviación estándar completa de la otra (un efecto grande) (Rendón-Macías et al., 2021). Generalmente se consideran umbrales orientativos (según Cohen) de ~ 0.2 como efecto pequeño, ~ 0.5 mediano y ≥ 0.8 grande.

1.2.7.7. Diferencias típicas evaluadas en estudios genómicos comparativos

En estudios comparativos de plataformas de secuenciación o pipelines, suelen evaluarse múltiples dimensiones de diferencia para obtener una imagen integral de desempeño. Entre los tipos de diferencias o relaciones comúnmente examinadas se incluyen:

- Cantidad de variantes detectadas: se compara el número total de variantes (SNVs, *indels*, etc.) que cada plataforma identifica en las mismas muestras. Esto revela diferencias de sensibilidad: por ejemplo, si Illumina detecta sistemáticamente más variantes que ONT en muestras idénticas, podría sugerir que ONT pierde variantes (falsos negativos) o que Illumina incluye más variantes de baja frecuencia. Giandhari *et al.* observaron, por ejemplo, que MiSeq detectó más mutaciones en

los genomas virales que GridION, evidencia de mayor rendimiento en llamadas puntuales (Tshiabuila et al., 2022).

- Frecuencia alélica (AF) y otras propiedades de variantes coincidentes: cuando ambas plataformas detectan una misma variante, es relevante comparar parámetros asociados. Un caso es la AF estimada (proporción de lecturas que soportan la variante). Diferencias sistemáticas en AF entre plataformas (evaluadas con tests como Wilcoxon) pueden indicar sesgos: estudios han encontrado que ONT tiende a reportar frecuencias alélicas ligeramente diferentes, posiblemente por sus errores de secuenciación. Asimismo, se comparan calidades de genotipo u otras métricas de confianza para variantes comunes, buscando ver si una plataforma da consistentemente mayor soporte estadístico a las llamadas (Mariner-Llicer et al., 2021).
- Profundidad de cobertura (DP): se examina si existen variaciones en la cobertura por plataforma, tanto global (profundidad promedio) como específicamente en sitios de variantes. Una plataforma podría tener menor cobertura efectiva en regiones difíciles, impactando su capacidad de detectar variantes. Comparar la distribución de DP entre Illumina y ONT (p. ej. con Wilcoxon) revela si alguna tiende a cubrir más profundamente las regiones variantes. En estudios previos, no siempre se hallan diferencias significativas en cobertura entre plataformas, dado que esta depende más del diseño experimental (profundidad de secuenciación deseada) que de la tecnología en sí, aunque la *cobertura utilizable* tras filtrados de calidad puede variar (Tshiabuila et al., 2022).
- Exactitud de genotipado y errores: métricas de concordancia más complejas incluyen la tasa de falsos positivos/falsos negativos de variantes. A veces se evalúa utilizando un conjunto de “verdad terreno” o comparando con un estándar de referencia. También se puede calcular coeficientes de concordancia como Cohen’s kappa para genotipos detectados por ambos métodos. Estas medidas indican

cuánta concordancia más allá del azar existe en las llamadas binarias (variante presente/ausente) entre plataformas (Manzoor et al., 2025).

- Diferencias globales en el perfil de variantes: mediante análisis multivariados ya mencionados (PERMANOVA), se investiga si, considerando todas las variantes conjuntamente (o vectores de múltiples métricas), las muestras se diferencian por plataforma. Por ejemplo, una PERMANOVA podría detectar que globalmente los genomas secuenciados por ONT presentan un perfil de variantes ligeramente distinto a los secuenciados por Illumina (p. ej., ONT podría detectar más *indels* mientras Illumina más SNVs, alterando la composición global). Estas diferencias globales suelen resumirse con estadísticos como el R^2 de PERMANOVA (proporción de variación explicada por la plataforma).

Finalmente, los estudios comparativos genómicos típicamente evalúan tanto parámetros puntuales (número de variantes, AF, DP, calidad) como patrones integrales (concordancia global de llamadas, perfiles de variantes en conjunto). La literatura reciente reporta, por ejemplo, que Illumina ofrece mayor precisión en SNPs puntuales mientras ONT aporta detección de variantes estructurales, llevando a diferencias en los conjuntos de variantes detectados. También se han documentado disparidades en frecuencias alélicas y errores de secuenciación: ONT históricamente tiene mayor tasa de error bruto, lo que puede reflejarse en necesidades de filtrados más estrictos y potencial ligera subestimación o sobrestimación de AF (Delahaye & Nicolas, 2021a). Todas estas evaluaciones, respaldadas por las pruebas estadísticas y métricas descritas, proporcionan la base para una comparación cuantitativa rigurosa entre plataformas, más allá de observaciones anecdóticas.

1.2.8. Secuenciación y su aplicabilidad clínica

1.2.8.1. Generalidades

En el laboratorio clínico clásico, la secuenciación Sanger se ha usado extensamente para pruebas diagnósticas dirigidas: confirmar la presencia de una mutación puntual conocida (por ejemplo, la mutación Factor V Leiden en trombofilia), secuenciar un gen candidato en un paciente con enfermedad monogénica, o validar resultados de NGS. Sanger sigue siendo apreciado por su claridad en la interpretación (la lectura humana de un cromatograma bien separado es muy fiable) y porque los estándares de acreditación llevan décadas refinados (Giani et al., 2020; Heather & Chain, 2016).

Las plataformas de NGS (Illumina) han revolucionado la genética clínica en los últimos ~10 años, es común hoy en día emplear paneles multigénicos (por ejemplo, un panel de 100 genes de cáncer hereditario por secuenciación masiva), exoma clínico para diagnosticar enfermedades raras y, cada vez más, genoma completo en entornos avanzados. La alta precisión de NGS y su sensibilidad para variantes de baja frecuencia la hacen ideal para diagnóstico de mosaicos somáticos (ej. en epilepsias causadas por mutaciones postzigóticas de bajo porcentaje), para detección de mutaciones compuestas, etc (Akintunde et al., 2023; Satam et al., 2023). En oncología, la secuenciación de paneles de genes en tumores guía terapias dirigidas; en microbiología, la NGS permite tipificar cepas y detectar resistencia genética (como secuenciar el genoma de una bacteria en lugar de paneles de PCR individuales). En los últimos años, la Meta-Genómica Clínica se ha abierto paso: secuenciación directa de todos los ácidos nucleicos de una muestra clínica para identificar patógenos sin cultivo (NGS shotgun metagenómico), útil en infecciones desconocidas (Chiu & Miller, 2019; Hilt & Ferrieri, 2022; Pei et al., 2023; Satam et al., 2023).

Por su parte, la tecnología Nanopore inicialmente fue adoptada en contextos de investigación y en respuesta rápida (como brotes virales y vigilancia de patógenos emergentes), pero gradualmente empezó a mostrar su utilidad en entornos clínicos específicos (Hoenen et al., 2016; MacKenzie & Argyropoulos, 2023; Y. Wang et al., 2021). Una ventaja clara está en la detección de variantes estructurales grandes que Illumina

puede pasar por alto, como por ejemplo, grandes deleciones en genes (que no generan lecturas mapeables en NGS) o expansiones de repeticiones (como la de Fragile X o ALS/FTD C9orf72) que requieren lecturas largas para contarse, por lo que ONT ha permitido leer a través de expansiones de cientos de repeticiones que antes solo podían inferirse indirectamente (Pagano et al., 2025; Romagnoli et al., 2023; Santos et al., 2025).

Otra área es la identificación rápida de patógenos directamente de muestras clínicas puesto que existen protocolos validados para secuenciar en un MinION el ADNc de virus respiratorios o el ADN de bacterias en sangre, y con ayuda de software identificar en pocas horas al agente y sus genes de resistencia, lo cual puede salvar vidas en sepsis de origen desconocido (Greninger et al., 2015; X. Wang et al., 2022).

En diagnóstico genético de pacientes, ONT todavía no es de uso rutinario general, principalmente por temas regulatorios (las nuevas técnicas requieren validación exhaustiva) y porque Illumina ya cubre la mayoría de necesidades con alta precisión. No obstante, hay campos donde ONT sobresale, y uno de ellos es el de las enfermedades mitocondriales (Slapnik et al., 2024).

1.2.8.2. Aplicaciones clínicas de la secuenciación: enfoque en enfermedades mitocondriales y heteroplasmia

La secuenciación masiva ha impactado fuertemente la medicina, permitiendo diagnósticos genómicos más precisos y rápidos. Aquí nos centraremos en un caso ilustrativo de las enfermedades mitocondriales, que presentan desafíos diagnósticos particulares y donde la secuenciación de ADN (tanto nuclear como mitocondrial) juega un papel clave. También abordaremos qué aporta la secuenciación a la detección de heteroplasmia (coexistencia de genomas mitocondriales mutados y normales) y cómo se integra en el diagnóstico genético global.

Las enfermedades mitocondriales comprenden un grupo heterogéneo de trastornos metabólicos originados por disfunciones en la cadena respiratoria u otros procesos mitocondriales. Pueden ser causadas por mutaciones en el propio ADN mitocondrial (ADNmt, un genoma circular de 16,569 pb que codifica 13 proteínas, 22 tRNAs y 2 rRNAs),

o por mutaciones en uno de más de 350 genes nucleares que afectan la biogénesis y función mitocondrial (Craven et al., 2025; Lokody, 2014).

Esto implica múltiples patrones de herencia (materna para mutaciones en ADNmt, y autosómica dominante, recesiva, ligada al X o heredabilidad compleja para genes nucleares). Además, el ADNmt está presente en múltiples copias por célula (cientos a miles) y puede existir la situación de heteroplasmia, donde una fracción de las copias tiene una mutación y el resto son silvestres. La heteroplasmia puede variar entre tejidos y cambiar con el tiempo debido a la segregación mitótica aleatoria y cuellos de botella en la línea germinal. Generalmente, las manifestaciones clínicas aparecen cuando la proporción de ADNmt mutado excede un umbral en el tejido afectado (umbral que puede ser ~60–90% según la mutación y el tejido) *nature.com* (Nissanka & Moraes, 2020; Parakatselaki & Ladoukakis, 2021).

Tradicionalmente, el diagnóstico de una enfermedad mitocondrial era complejo, requería correlacionar síntomas multisistémicos variados, realizar biopsias musculares para estudios histoquímicos (fibra roja rasgada, etc.), ensayos enzimáticos de la cadena respiratoria, y solo entonces análisis genéticos dirigidos. Muchas pruebas implicaban procedimientos invasivos (biopsia muscular o hepática) y costosos (Ahmed et al., 2018; Parikh et al., 2015).

Sin embargo, con la llegada de la secuenciación de nueva generación, ha habido un cambio de paradigma a “secuenciar primero”, es decir, ante la sospecha de un desorden mitocondrial, se tiende ahora a secuenciar directamente los genes asociados (tanto ADNmt como paneles de genes nucleares) desde muestras de sangre o tejidos no invasivos, buscando una causa genética que explique el fenotipo clínico, gracias a que la secuenciación masiva puede analizar de forma simultánea todos los genes de interés con alta sensibilidad (Ahmad & Hasan, 2021; Devendra et al., 2025; Mahmud et al., 2022).

Un enfoque común es efectuar secuenciación del genoma mitocondrial completo junto con un panel de genes nucleares mitocondriales o incluso exoma/genoma completo del paciente. La secuenciación del ADNmt por NGS tiene ventajas claras sobre Sanger dada su mayor sensibilidad para detectar heteroplasmas bajas y poder de cuantificación con precisión. Por ejemplo, la guía británica de buenas prácticas (2023) destaca que mediante

secuenciación profunda se puede detectar variantes heteroplásmicas en niveles por debajo del 5–10%, algo imposible con Sanger (Ahmad & Hasan, 2021; Mahmud et al., 2022).

Esta sensibilidad es crucial, ya que algunas madres portadoras de mutaciones ADNmt patogénicas pueden tener proporciones bajas en sangre (asintomáticas) pero arriesgan hijos afectados si la carga mutacional transmitida a la descendencia es alta. De hecho, estudios poblacionales sugieren que ~1 de cada 250 personas sanas porta una variante patogénica de ADNmt a bajo nivel, que podría manifestarse en su linaje (Abicht et al., 2018; Ahmad & Hasan, 2021; Cui et al., 2013).

Los resultados de integrar NGS en el diagnóstico de sospechas mitocondriales han sido notables, por ejemplo en un estudio de cohorte realizado en Inglaterra (Proyecto 100,000 Genomas) se secuenció el genoma completo (WGS) de 345 pacientes con sospecha de enfermedad mitocondrial, tras descartar causas comunes. El análisis incluyó tanto variantes en ADNmt como en ~1100 genes nucleares relacionados con función mitocondrial. Los hallazgos fueron que un 31% de las familias obtuvieron un diagnóstico genético definitivo o probable gracias a la secuenciación (98 de 319 familias analizadas), con otro 2% obteniendo posibles diagnósticos. Curiosamente, de esos diagnósticos, solo el 38% correspondieron a mutaciones en el ADNmt, mientras que el 62% resultaron ser mutaciones en genes nucleares no mitocondriales (K. R. Schon et al., 2021).

Es decir, muchos pacientes que clínicamente parecían tener una enfermedad mitocondrial en realidad tenían otros síndromes genéticos (p. ej., desórdenes del desarrollo, epilepsias genéticas, miopatías metabólicas) cuyo diagnóstico se pasó por alto al enfocarse solo en mitocondria (K. R. Schon et al., 2021; Thäle et al., 2020). Este resultado enfatiza la utilidad de secuenciar ampliamente, dado que secuenciar un genoma completo pudo identificar causas inesperadas y tratables en algunos casos, que se habrían perdido con pruebas panelizadas estrechas. Por ejemplo, el estudio refiere diagnósticos de enfermedades metabólicas con tratamiento, canalopatías epilépticas, etc., en pacientes originalmente pensados de causa mitocondrial. En conjunto, la “odisea diagnóstica” de estos pacientes se acortó gracias a la secuenciación masiva, evitando años de pruebas inconclusas (K. R. Schon et al., 2021).

Ahora bien para la molécula específica del genoma mitocondrial, la secuenciación profunda ha permitido detectar tanto mutaciones puntuales heteroplásmicas como deleciones. La mayoría de variantes patogénicas en ADNmt conocidas son SNVs o pequeñas *indels* (por ejemplo, m.3243A>G asociada a MELAS, m.8344A>G asociada a MERRF, etc.), pero también existen síndromes causados por deleciones grandes de ADNmt (típicamente esporádicas, como el síndrome de Kearns-Sayre con deleciones de 1–10 kb) (Thäle et al., 2020). Para detectar deleciones del ADNmt antes se recurría a técnicas como *Southern blot* en músculo; que ahora con secuenciación masiva (ya sea genoma completo o genoma mitocondrial) es posible detectarlas al observar lecturas discordantes (pares de lectura mapeando con huecos) o un descenso abrupto en la cobertura de secuenciación en cierta región del genoma mitocondrial (Cannon et al., 2024; J. Wang et al., 2022).

Por su parte, las guías de práctica (como la del NHS Reino Unido 2023) recomiendan un algoritmo donde ante la sospecha clínica de enfermedad mitocondrial se realice secuenciación del ADNmt en sangre, y en paralelo o posteriormente secuenciación de paneles/exoma nuclear, antes de considerar procedimientos invasivos. Solo si la secuenciación no arroja resultados se sugiere pasar a biopsias para análisis funcionales. La misma guía resalta que la secuenciación NGS del ADNmt es preferible a Sanger porque es más sensible y costo-efectiva, permitiendo detectar niveles bajos de heteroplasmia y ahorrando esfuerzo. Por ejemplo, a ~500× de profundidad, hay ~99% probabilidad de detectar una variante presente al 7.5% de heteroplasmia (con ≥5% de las lecturas). Esto es importante para identificar portadores maternos e incluso para decisiones reproductivas (p. ej., diagnóstico prenatal o selección de embriones en FIV, donde si la madre porta una mutación ADNmt a baja heteroplasmia, se podría estimar el riesgo de transmisiones altas) (Mavraki et al., 2023).

1.2.8.3. Heteroplasmia: detección y significado clínico.

Heteroplasmia se define como la presencia de más de un tipo de secuencia de ADNmt en la misma persona, esta puede estar presente como homoplasmia (todas las copias iguales) o heteroplasmia en distintos porcentajes (Parakatselaki & Ladoukakis, 2021). Muchas mutaciones patogénicas en ADNmt son heteroplásmicas, y el porcentaje mutante suele correlacionar con la severidad del fenotipo (efecto umbral) (Slapnik et al., 2024). Dado que

diferentes tejidos pueden tener niveles distintos (por la distribución al azar de mitocondrias durante divisiones celulares), un paciente puede tener, por ejemplo, 80% de heteroplasmia en músculo (causando miopatía) pero solo 10% en sangre (lo que podría fallar en detectarse en un test sanguíneo). Por eso, las biopsias de tejido afectado a veces eran necesarias para confirmar el diagnóstico (ej., la mutación m.3243A>G de MELAS suele estar más alta en músculo que en sangre periférica) (Nissanka & Moraes, 2020; Parakatselaki & Ladoukakis, 2021).

La secuenciación de nueva generación ha mejorado la detección de heteroplasmia de varias maneras, por ejemplo, un NGS profundo puede detectar heteroplasmia incluso por debajo de 1% en experimentos bien controlados (Abicht et al., 2018). Por otra parte, porque permite distinguir variantes verdaderas de artefactos, es decir, un error de secuenciación al azar aparecerá en una sola lectura, mientras que una heteroplasmia real generará múltiples lecturas con la misma alteración y una proporción consistente en ambas hebras. Se han desarrollado algoritmos bioinformáticos específicos para llamar heteroplasmas de ADNmt considerando la distribución de cualidades de base, posición de la variante, etc., para minimizar falsos positivos (por ejemplo, software como MitoSeek, MutDetect o pipelines en estudios específicos) (González et al., 2020; Kim et al., 2018).

Un hallazgo interesante es que la heteroplasmia de bajo nivel es común en individuos sanos, donde estudios de ultrasecuenciación han mostrado que >90% de la población general tiene al menos alguna variante menor de ADNmt en <1% de sus moléculas. La mayoría son neutrales y pasan desapercibidas, pero indican que las mitocondrias acumulan mutaciones a lo largo de la vida (sobre todo en tejidos no proliferativos) (González et al., 2020). Cuando la heteroplasmia ocurre en línea germinal, puede segregarse a niveles altos en la descendencia debido al cuello de botella embrionario, explicando por qué una madre con baja heteroplasmia puede tener un hijo severamente afectado. Por tanto, disponer de métodos sensibles es valioso incluso en consejo genético; si se detecta en la madre una mutación patogénica al 1–5% en sangre, se sabe que hay riesgo de transmisión y se pueden plantear opciones reproductivas (ovodonación, Dx preimplantacional) (Otten et al., 2018; Wallace & Chalkia, 2013).

Otro factor en heteroplasmia es la coexistencia de múltiples deleciones en ciertos síndromes. Por ejemplo, en las mito neuro-gastrointestinal encefalopatías (síndrome

MNGIE) se observan numerosas deleciones parciales de ADNmt secundarias a mutaciones en genes nucleares (p. ej. TYMP) (Garone et al., 2011; Yadak et al., 2017). Secuenciar muchas moléculas individuales puede detectar subpoblaciones de ADNmt con diferentes deleciones. Las lecturas largas de ONT aquí ofrecen un plus debido a que pueden leer un genoma mitocondrial completo de una sola pieza y mostrar directamente si le falta un segmento (de hecho, ONT ha permitido mapear con precisión los puntos de ruptura de deleciones esporádicas sin necesidad de PCR ni suposiciones) (Garone et al., 2011; Mavraki et al., 2023; Yadak et al., 2017).

Adicionalmente, la heteroplasma plantea un reto especial a la secuenciación de lectura corta, sobre todo por la presencia de NUMTs (segmentos de ADN mitocondrial integrados en el genoma nuclear). Los NUMTs pueden confundirse con ADNmt genuino cuando se secuencia ADN total, ya que con lecturas cortas se pueden terminar mapeando parcialmente. Por lo tanto, con secuenciación de nueva generación, un NUMT reciente (muy similar al ADNmt actual) podría parecer una falsa heteroplasma (pseudoheteroplasma) (Kim et al., 2018; Maude et al., 2019; Taniguchi et al., 2023).

Wei et al., (2022) mostraron que prácticamente todos los humanos tienen NUMTs, algunos muy recientes, y que se generan nuevos NUMTs de manera germinal con cierta frecuencia. Interpretar secuencias mitocondriales requiere filtrar estos artefactos. La ventaja de lecturas largas es que un NUMT completo puede reconocerse porque la lectura cubrirá también secuencia flanqueante nuclear. Akamatsu et al., (2024) señalan que con lecturas ONT de todo el ADNmt, se puede distinguir genuinas variantes heteroplásmicas de señales de NUMTs, ya que solo las moléculas mitocondriales reales abarcan el genoma circular completo.

Para los procesamientos de WGS en illumina, se toman ciertas precauciones como exigir altas coberturas y validar con métodos ortogonales o buscar evidencia de fragmentos tipo NUMT (pares discordantes). Aun así, la presencia de NUMTs ha llevado a recomendaciones de confirmar en un tejido rico en mitocondrias (ej. músculo) cuando hay dudas, o de usar métodos de enriquecimiento del ADNmt (por captura o por amplificación preferencial) (Mavraki et al., 2023; Wei et al., 2022).

Precisamente, estrategias de enriquecimiento de ADNmt son comunes dado que se puede aislar la fracción mitocondrial (lo cual requiere tejido fresco), o hacer PCR de largo alcance, amplificando el genoma mitocondrial entero en unos pocos fragmentos, así como también usar sondas de captura de ADNmt antes de secuenciar (Cui et al., 2013; González et al., 2020; Mahmud et al., 2022). Cada una conlleva sus sesgos (PCR puede crear recombinaciones artefactuales o perder delecciones; captura depende de diseño de sondas, entre otras). Algunos trabajos recientes han explorado usar Cas9 para recortar específicamente el ADNmt e identificarlo en el flujo Nanopore, enriqueciendo enormemente su proporción logrando cobertura ultra-profunda en ONT (Slapnik et al., 2024).

En conclusión, la combinación de secuenciación masiva nuclear y mitocondrial ha elevado la tasa de diagnóstico de enfermedades mitocondriales y ha dado información cuantitativa sobre heteroplasmia valiosa para manejo clínico. Aún existen casos sin resolver (en la cohorte BMJ 2021, ~60% quedaron sin diagnóstico genético definitivo), lo que motiva investigación en genes nuevos y quizás en factores no genéticos. Pero la tendencia es clara, la secuenciación es ahora la piedra angular en la investigación de estos desórdenes. De hecho, muchos laboratorios han implementado protocolos de secuenciación específica de ADNmt para pacientes con miopatías/metabolopatías de etiología incierta, dada la facilidad de secuenciar la mitocondria entera en sangre o saliva con un panel NGS o MinION.

2. Descripción del método para la detección de variantes del ADN mitocondrial por plataformas de secuenciación masiva (ONT)

2.1. Flujo de trabajo bioinformático aplicado

En este capítulo se describe el flujo de trabajo bioinformático diseñado y estandarizado para el análisis de genomas mitocondriales humanos a partir de lecturas largas generadas por la plataforma ONT, con el propósito de identificar variantes en el ADN mitocondrial (ADNmt) y compararlas con aquellas detectadas mediante secuenciación de lecturas cortas (Illumina).

El flujo de trabajo implementado integra diferentes etapas: (i) preprocesamiento de datos crudos y control de calidad inicial, (ii) alineamiento de lecturas contra la referencia del genoma humano (chrM, GRCh38), (iii) llamada de variantes utilizando múltiples algoritmos diseñados para lecturas largas, (iv) filtrado y anotación de las variantes, y (v) análisis comparativo entre las plataformas. La Figura 2 resume gráficamente este procedimiento, mostrando las herramientas utilizadas en cada etapa, así como la construcción de los vectores característicos de cada plataforma que sirvieron de base para los análisis estadísticos y de concordancia.

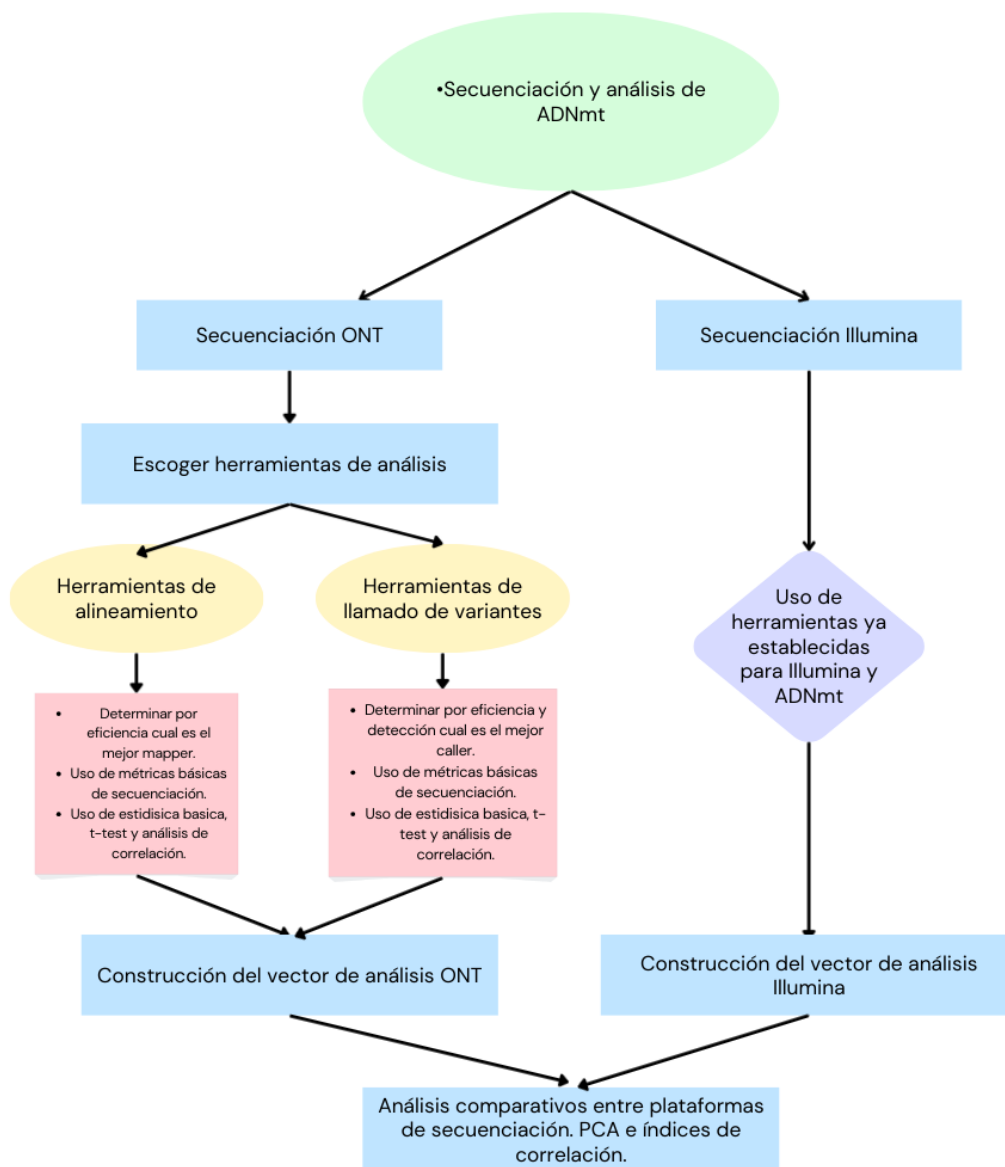


Figura 2. Diagrama de flujo del análisis realizado en este trabajo.

2.1.1. Información biológica mitocondrial

El diseño experimental se fundamentó en el uso de cinco muestras biológicas previamente secuenciadas con la plataforma Illumina, las cuales fueron nuevamente procesadas mediante triplicados independientes de secuenciación con MinION (ONT). Esta estrategia

permitió evaluar tanto la consistencia técnica entre réplicas como la concordancia biológica entre tecnologías.

Las muestras usadas para la secuenciación ONT son contramuestras de un laboratorio clínico en Colombia, a las que se les garantizó condiciones estandarizadas de extracción y preservación del material genético, para unas muestras previamente secuenciadas por Illumina, como parte del servicio ofrecido por dicho laboratorio, cuyas muestras principales fueron enviadas a un laboratorio externo internacional. Los datos obtenidos para el análisis establecido en este documento fueron previamente anonimizados y el uso de su información fue avalado por el comité de ética de la Facultad de Medicina de la Universidad Nacional de Colombia, 2.FM.1.002- CE-092-25 del 29 de mayo de 2025 (Acta 008). La elección del ADN mitocondrial como objeto de estudio respondió a su relevancia biomédica en contextos de diagnóstico clínico, dado que pequeñas variaciones en los niveles de heteroplasmia pueden tener consecuencias funcionales significativas y se utilizan como biomarcadores en diversas enfermedades metabólicas y neurodegenerativas.

2.1.2. Paso 0. Secuenciamiento del ADNmt y datos Illumina

Previo al proceso de secuenciación, cada muestra fue sometida a un enriquecimiento específico del ADNmt, por medio de PCR larga, con el fin de maximizar la proporción de lecturas correspondientes a la molécula circular de 16,6 kb. Este enriquecimiento se realizó en tres réplicas técnicas independientes por muestra, lo cual proporcionó una base sólida para evaluar la reproducibilidad intra-plataforma.

Las librerías fueron preparadas de acuerdo con los protocolos estandarizados recomendados por ONT, utilizando el kit de Native Barcoding que incorpora la última química diseñada (Q20+) para mantener valores de calidad y teniendo en cuenta las características intrínsecas del material genético. La secuenciación se efectuó en un dispositivo MinION, utilizando celdas de flujo R10.4, que incorporan mejoras sustanciales en la resolución de homopolímeros y en la precisión del llamado de bases.

Este paso inicial permitió generar un conjunto de datos crudos (raw pod5/fastq, con un tamaño entre 8-160Mb por replica) con una cobertura suficiente para análisis de variantes. Dichos datos constituyeron el insumo primario para el pipeline bioinformático diseñado, que posteriormente fue contrastado con la información de referencia obtenida mediante secuenciación Illumina.

Por su parte, los datos de secuenciación Illumina fueron obtenidos de un procesamiento previo, donde el material biológico fue remitido a un laboratorio externo internacional, fue extraído el ADN total, y se hizo un enriquecimiento por PCR para la molécula de ADNmt con primers específicos (no se indican las secuencias utilizadas), y luego se llevó a cabo la preparación de librerías específica para ADN mitocondrial, secuenciado posteriormente con un sistema NovaSeq 6000 Sequencing System, de Illumina.

2.1.3. Paso 1 *Basecalling* con Dorado (MinKNOW)

Nuestro método inicia con el proceso estándar del secuenciamiento conocido como *basecalling*. En este proceso se convierten las señales de corriente iónica medidas por el dispositivo ONT en secuencias de nucleótidos (bases A, C, G, T). Para este fin se empleó Dorado, la herramienta de *basecalling* de alto rendimiento desarrollada por Oxford Nanopore e integrada en la plataforma de control MinKNOW. Dorado utiliza algoritmos de redes neuronales profundas para decodificar las señales eléctricas de Nanopore en tiempo real, produciendo lecturas en formato FASTQ/FASTA o directamente archivos BAM/CRAM alineados sin procesar.

En este trabajo se decidió configurar Dorado en modo de alta precisión (modelo Super Accuracy (SUP)), priorizando la exactitud de las lecturas sobre la velocidad. MinKNOW permite realizar el *basecalling* en vivo durante la secuenciación. Dorado es ahora el *basecaller* predeterminado en MinKNOW y ofrece mayor exactitud que el software previo (Guppy), con mejoras continuas en sus modelos neuronales. El resultado de esta etapa son secuencias de lectura larga (promedios típicos >15 kb) con identidades de ~99% respecto al genoma real, listas para ser alineadas.

Seguido del llamado de bases y generación de los FastQ correspondientes, se realizó un *trimming* de valor de calidad de base mínimo de 20 y un tamaño de *read* por encima de los 5kb, utilizando la herramienta NanoFilt.

2.1.4. Paso 2 Alineamiento de lecturas al genoma mitocondrial humano de referencia

Una vez obtenidas las lecturas *trimmed*, el siguiente paso consistió en alinearlas contra la secuencia de referencia mitocondrial humana (revisada Cambridge Reference Sequence, rCRS, NC_012920). Dado que las lecturas ONT son largas y relativamente ruidosas (tasa de error ~5-10%), se utilizó como estrategia bioinformática la evaluación de tres herramientas de alineamiento con el objetivo de maximizar el mapeo de los datos. Para ello se utilizó Minimap2, Winnowmap2 y NGMLR (Dobner et al., 2024; Li, 2018).

Las lecturas en formato FASTQ fueron procesadas en entornos reproducibles mediante contenedores Docker, con la referencia previamente indexada. En el caso de Minimap2, se empleó el preset específico map-ont, optimizado para lecturas largas de ONT, con salida en formato SAM. Para NGMLR, se aplicó la opción -x ont, diseñada para datos ruidosos y con capacidad de resolver variantes estructurales complejas a través de un modelo de penalización convexa. Finalmente, Winnowmap2 se utilizó incorporando un archivo de k-mers frecuentes generado con meryl, lo que reduce alineamientos espurios en regiones repetitivas.

La estrategia de ejecución consistió en iterar sobre todos los archivos FASTQ, generando archivos de salida SAM para cada combinación muestra–alineador, los cuales fueron convertidos y ordenados en BAM para análisis posteriores. El tiempo de ejecución y los recursos computacionales fueron registrados, considerando que Minimap2 ofreció un balance más eficiente en consumo de memoria y velocidad, mientras que NGMLR y Winnowmap2 proporcionaron mayor resolución en escenarios de variabilidad estructural o repetitiva.

Herramienta	Versión / Imagen Docker	Parámetros clave	Observaciones
Minimap2	Li, 2018 – map-ont	-ax map-ont	Rápido, eficiente en lecturas largas.
NGMLR	Sedlazeck et al., 2018	-x ont	Preciso en regiones complejas y SVs.
Winnomap2	Jain et al., 2020	-W repeats.txt -ax map-ont	Reduce falsos positivos en regiones repetitivas.

Tabla 3. Tabla con características clave y parámetros clave de los alineadores usados.

Las métricas de profundidad y calidad de las bases se extrajeron mediante samtools depth, samtools stats y se realizaron análisis adicionales en R, lo cual permite identificar combinaciones de herramientas que maximizan la cobertura uniforme del genoma mitocondrial (chrM), sin sesgos regionales, lo cual es indispensable para asegurar una adecuada capacidad de detección en regiones codificantes y no codificantes por igual.

2.1.5. Paso 3: Análisis comparativo de herramientas de llamado de variantes en ADN mitocondrial.

El llamado de variantes mitocondriales se realizó mediante cuatro herramientas complementarias: Mutect2, Mutserve2, Medaka y Clair3. Todas fueron aplicadas sobre los archivos BAM filtrados y con grupos de lectura previamente añadidos, estos parámetros se resumen en la Tabla 4.

- Mutect2 (GATK 4.2.6.1): ejecutado en *modo mitocondrial* (--mitochondria-mode true), deshabilitando el filtro MateOnSameContigOrNoMappedMateReadFilter. Esta configuración favorece la detección de variantes de baja frecuencia ($\geq 1-2\%$) bajo un modelo bayesiano de error.

- Mutserve2 (v2.0.3): diseñado específicamente para ADNmt, se utilizó con parámetros de calidad estrictos (--baseQ 20, --mapQ 30, --alignQ 30) y nivel de confianza 1 (--level 1). Esto permitió obtener variantes con umbrales de heteroplasmia a partir de ~1–2%.
- Medaka: empleado para generar consensos de alta precisión a partir de lecturas ONT. Se utilizó el modelo SUP de *basecalling* (*sup*), dada la química utilizada en secuenciación. Los VCF resultantes fueron normalizados con bcftools, incorporando métricas de profundidad (DP) y recuento alélico (AD).
- Clair3: modelo de *deep learning* que combina representaciones de *pileup* y *full-alignment*, optimizado para lecturas ONT. Se ejecutó mediante contenedores Docker con el modelo r1041_e82_400bps_sup_v500. Esta herramienta produjo VCF comprimidos e indexados, permitiendo la cuantificación de frecuencias heteroplásmicas con alta precisión.

Herramienta	Versión / Imagen Docker	Parámetros clave	Observaciones
Mutect2 (GATK 4.2.6.1)	Broad Institute	--mitochondria-mode true, desactiva filtro MateOnSameContig OrNoMappedMateReadFilter	Alta sensibilidad a bajas frecuencias.
Mutserve2 (v2.0.3)	Weissensteiner et al.	--level 1 --baseQ 20 --mapQ 30 --alignQ 30	Optimizado para variantes mitocondriales y heteroplasmia.
Medaka	ONT – modelos sup/hac/fast	medaka_variant -m MODEL	Genera consensos; requiere normalización posterior.

Clair3	HKU-BAL, Docker latest	--platform ont -- model sup_v500	Preciso en SNPs e indeles cortos $\geq 5\%$.
---------------	---------------------------	-------------------------------------	--

Tabla 4. Características y parámetros clave de las herramientas de llamado de variantes usadas en el trabajo.

Los archivos de salida (VCF) de cada *caller* fueron comprimidos e indexados para su posterior comparación entre herramientas y plataformas. Además, se exportaron tablas TSV con métricas clave (profundidad, calidad y fracción alélica) para análisis estadístico en R.

2.2. Construcción de matriz y evaluación de pipelines

La identificación precisa de variantes mitocondriales en datos de secuenciación requiere no solo una adecuada profundidad y calidad de lectura, sino también una cuidadosa selección de las herramientas bioinformáticas implicadas en el alineamiento y el llamado de variantes. Con el propósito de determinar qué combinación de *mapper* y *caller* ofrecía mayor consistencia y reproducibilidad entre réplicas, se construye una matriz multivariada con las posiciones de variantes detectadas en cada muestra-replica (Anexo 1). Cada registro de esta matriz corresponde a una variante específica (POS) asociada a una combinación muestra-réplica y contiene información crítica para su evaluación: alelos de referencia (REF) y alternativo (ALT), genotipo (GT), frecuencia alélica (AF), profundidad de cobertura (DP) y calidad promedio de base (BQ).

La matriz completa se conforma a partir de la unión de las salidas individuales de cada combinación de *mapper* y *caller*, incluyendo únicamente las posiciones con variantes detectadas. Para abordar la variabilidad entre herramientas y facilitar la comparación cuantitativa, se aplica un filtrado inicial para seleccionar exclusivamente aquellas posiciones que estaban presentes de forma común en todas las combinaciones evaluadas. Esto permite focalizar el análisis en las regiones genómicas con llamadas consistentes, eliminando el sesgo que podría introducir la detección diferencial de variantes ausentes o aleatorias.

Una vez obtenida la matriz filtrada de variantes comunes, se estandarizan las variables cuantitativas (AF, DP, BQ) mediante normalización z-score para controlar las diferencias de escala entre las métricas. Posteriormente, se calcula una matriz de distancias euclidianas entre combinaciones de herramientas, basada en la similitud multivariada de estas métricas para cada posición. Esta matriz de distancias se le realiza un Análisis de Componentes Principales (PCA), con el fin de representar gráficamente la proximidad relativa entre pipelines y detectar agrupamientos o dispersiones que reflejen mayor o menor coherencia entre réplicas.

2.2.1. Determinación del mejor *mapper*

2.2.1.1. Profundidad de cobertura y calidad de mapeo

Dada la alta cobertura del genoma mitocondrial por el enriquecimiento previo de la molécula de ADNmt, se obtuvieron coberturas por encima del 90% para el 95% de los *reads* obtenidos. La profundidad de cobertura y la calidad de las lecturas son factores determinantes para la sensibilidad del llamado de variantes, especialmente en el contexto mitocondrial, donde se busca detectar niveles de heteroplasmia que pueden ser clínicamente relevantes incluso por debajo del 10% (Björkman et al., 2023; Wei et al., 2022). Se estimaron métricas como la cobertura promedio por posición, los valores promedio de calidad de mapeo (MAPQ) y calidad de base (BQ), considerando como aceptables aquellos ≥ 15 (Holland et al., 2018). Debido además, al proceso previo de enriquecimiento, se evita que secuencias NUMTs pudieran ser alineadas en este análisis, dado que su representación con respecto a la molécula de ADNmt amplificada no representaría un riesgo para su análisis.

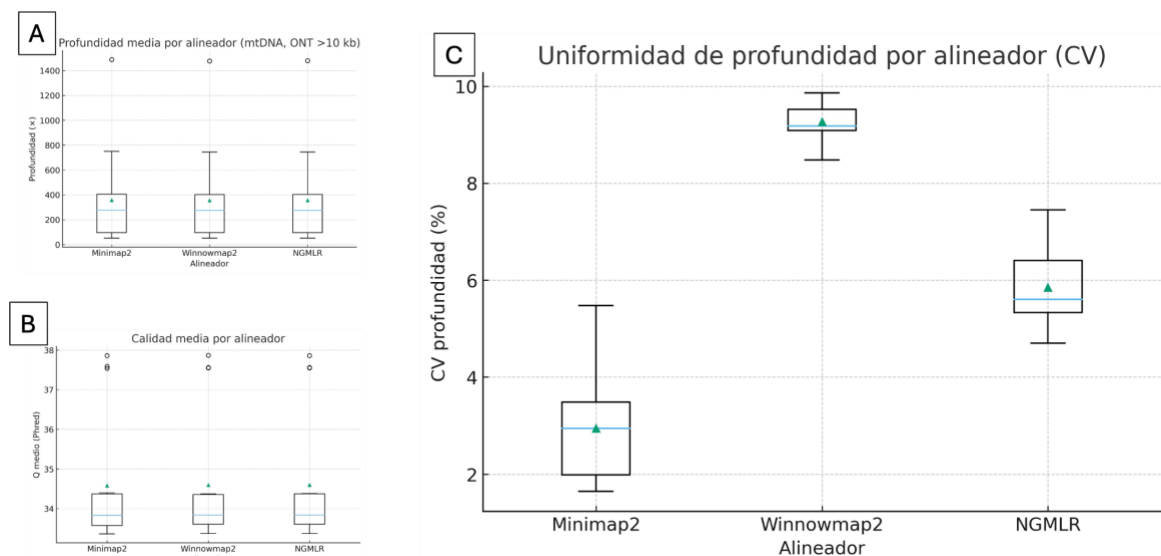


Figura 3. Comparación de mappers para ONT (>10 kb): profundidad media, calidad media y uniformidad de cobertura (CV). Paneles resumen por *mapper* (Minimap2, Winnowmap2, NGMLR). (A) profundidad media; (B) calidad media; (C) coeficiente de variación (CV) de la profundidad como medida de uniformidad. En nuestro conjunto, Minimap2 muestra la menor variabilidad de cobertura (CV más bajo), con valores competitivos de profundidad y calidad. Ver Métodos para parámetros de cada *mapper*.

Los resultados (Fig. 3) muestran que Minimap2 obtuvo la mayor fracción de lecturas mapeadas, así como la distribución más uniforme de la profundidad de cobertura, reflejada en un coeficiente de variación (CV) notablemente más bajo que el de los otros mapeadores. Además, presentó valores estables de calidad de base y mapeo. Winnowmap2 exhibió un desempeño similar en profundidad y calidad, aunque con una mayor dispersión en la uniformidad de cobertura. Finalmente, NGMLR, a pesar de estar diseñado para detectar variantes estructurales, excluyó un mayor número de lecturas y presentó la menor cobertura efectiva, además de un CV más alto, indicando mayor variabilidad posicional.

2.2.2. Determinación de la mejor combinación *mapper–caller*

2.2.2.1. Profundidad de lectura

Todas las combinaciones de *mapper* y *caller* alcanzaron profundidades de lectura suficientes para el análisis del genoma mitocondrial, superando en la mayoría de los casos las 500 lecturas por sitio. Sin embargo, se evidenciaron diferencias notables entre las herramientas. Las combinaciones basadas en Clair3, Medaka y Mutserve con *Minimap2* reportaron lecturas con profundidades más extremas, alcanzando valores puntuales de hasta 10.000 lecturas, lo cual refleja diferencias en la forma de calcular la profundidad o en los filtros aplicados por cada *caller*. Por contraste, Mutect2 mostró distribuciones más acotadas y homogéneas, lo que sugiere un procesamiento más conservador en relación con la profundidad asociada a cada variante (Fig. 4).

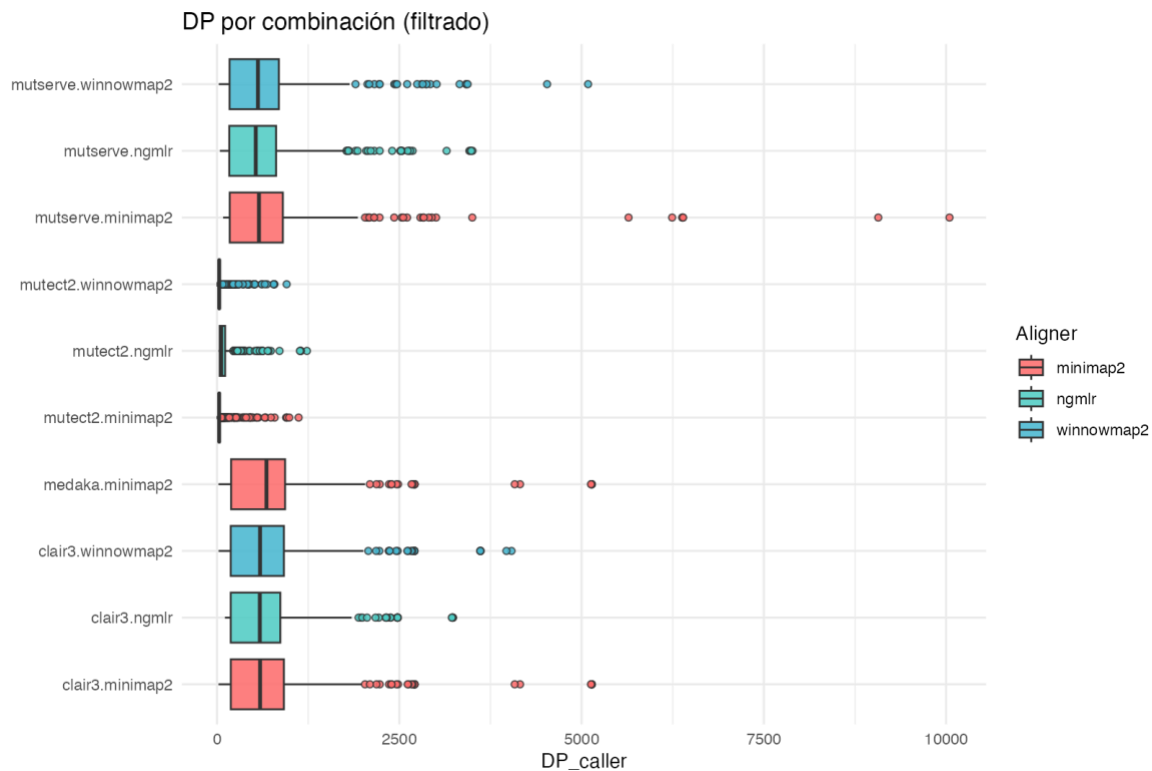


Figura 4. Profundidad (DP) por combinación *caller–mapper*. Diagramas de caja (y puntos) del AF de las variantes agrupadas por pipeline (p. ej., *clair3.minimap2*, *mutect2.winnowmap2*, etc.), aquí para la profundidad de lectura asociada a cada llamada.

Dado que la forma en que cada *calle* determina la calidad del genotipo (GQ) entre otras variaciones como TLOD en mutect2, esta métrica no fue tomada en cuenta para hacer los análisis de similitud, todos los análisis donde se vea un valor de calidad provienen de los datos obtenidos a partir de la calidad de base (BQ) de los *mappers*.

2.2.2.2. Frecuencia alélica

En la mayoría de las combinaciones, la frecuencia alélica (AF) se concentró en valores cercanos a 1, indicando la predominancia de variantes homoplásmicas. No obstante, se observó mayor dispersión en los resultados de Mutect2 (particularmente con *Minimap2*) y Clair3, donde se identificaron variantes con frecuencias intermedias (<0.75). Esto sugiere una mayor sensibilidad para detectar heteroplasmias de nivel bajo o intermedio. En contraste, los resultados de Mutserve y Medaka estuvieron fuertemente concentrados en $AF = 1$, reflejando un comportamiento más restrictivo en la detección de variantes (Fig. 5).

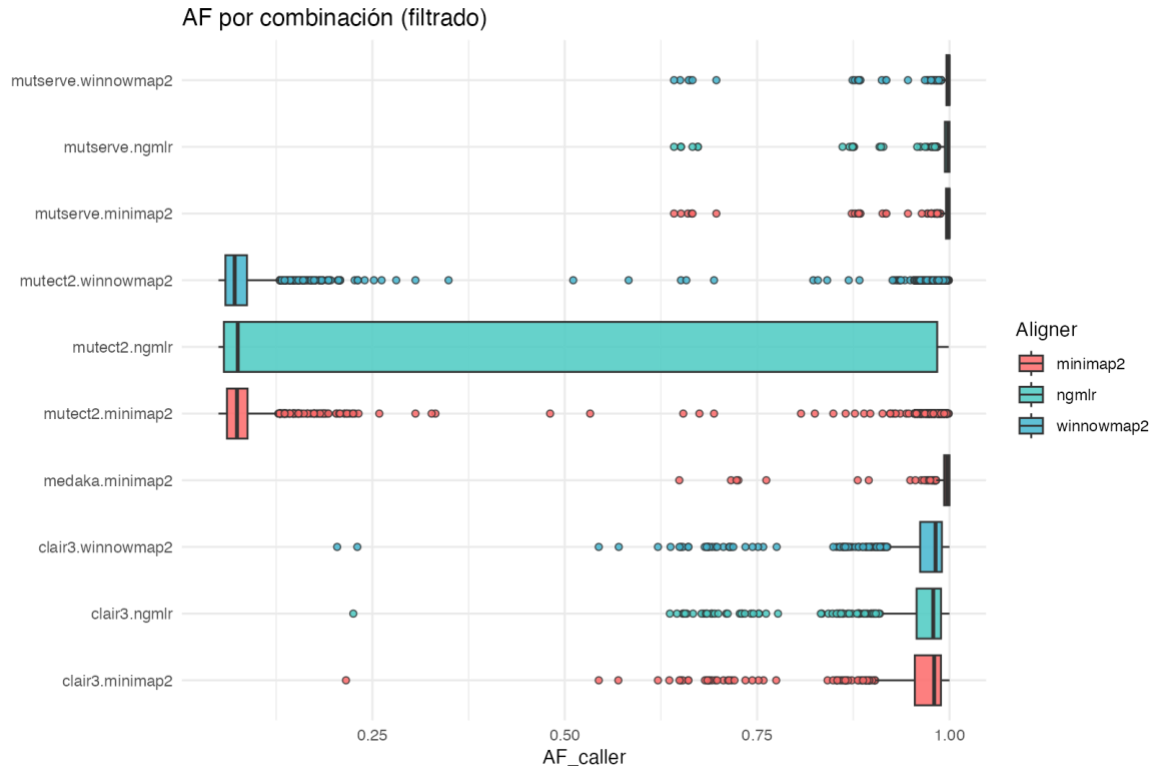


Figura 5. AF por combinación *caller–mapper*. Diagramas de caja (y puntos) del AF de las variantes agrupadas por pipeline (p. ej., *clair3.minimap2*, *mutect2.winnowmap2*, etc.). Cada caja resume el comportamiento global del par *caller–mapper* a través de muestras/llamadas.

El conteo de variantes mostró que la mayoría corresponden a variantes homoplásmicas (AF >95%). No obstante, las combinaciones Clair3 y Mutect2 recuperaron también un número considerable de heteroplasmas intermedias (AF entre 10–90%), en contraste con Mutserve y Medaka, donde las variantes reportadas fueron mayoritariamente homoplásmicas. Este resultado es consistente con el comportamiento más sensible de los dos primeros *callers* para capturar variantes de baja frecuencia. Para el análisis se adoptó un umbral práctico de detección de heteroplasma en 5%, siguiendo reportes previos que indican que ONT puede alcanzar confiabilidad aceptable a dicho nivel cuando se combina con modelos de *basecalling* Q20+ y profundidad elevada (Betschart et al., 2022; Guo et al., 2021).

2.2.2.3. Similitud entre combinaciones (Índice de Jaccard)

El análisis de similitud mostró una clara separación entre Mutect2 y el resto de herramientas. En promedio, Mutect2 presentó bajos valores de Jaccard (<0.3) en relación con las demás combinaciones, lo cual evidencia que reporta un conjunto de variantes poco coincidente con otros *callers*. Por el contrario, Mutserve, Clair3 y Medaka mostraron altos niveles de concordancia entre sí (Jaccard >0.5), especialmente cuando utilizaron el mismo *mapper*. Este patrón se refleja en la organización jerárquica del mapa de calor, donde estas tres herramientas forman un clúster definido (Fig. 6).

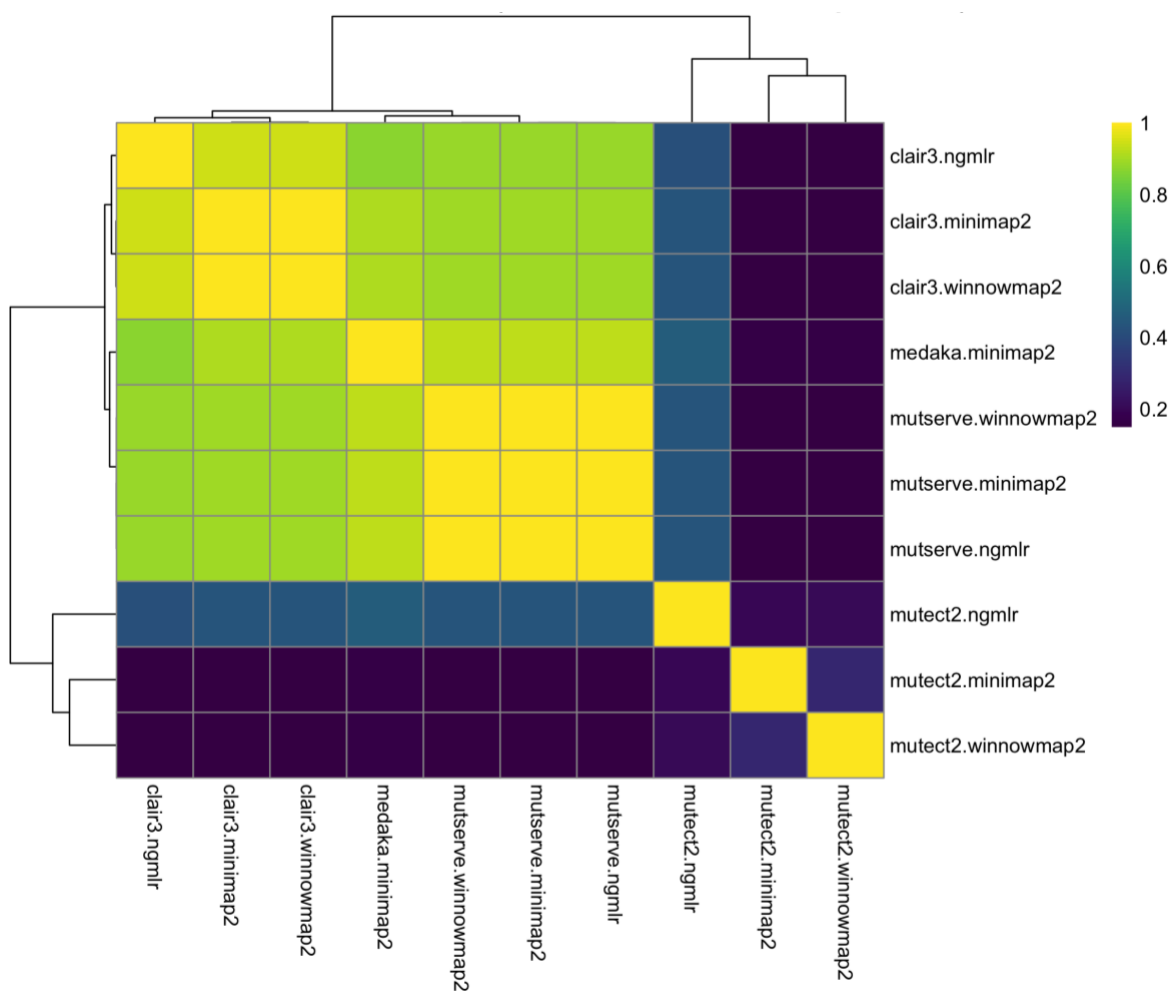


Figura 6. Similitud media (Jaccard) entre combinaciones *caller-mapper*. Mapa de calor de los coeficientes de Jaccard calculados sobre presencia/ausencia de variantes por posición (sin usar niveles de AF), promediados por combinación en las muestras donde ambas combinaciones están

disponibles. Escala 0–1 (amarillo = mayor coincidencia). Las filas/columnas se agrupan por similitud (clustering jerárquico).

2.2.2.4. Agrupamiento en espacio reducido (PCA)

Para ver mucho más claramente el comportamiento de las combinaciones de herramientas se utilizaron medidas de distancias en un espacio bidimensional definido por los dos primeros componentes principales del PCA, el cual fue realizado a partir de métricas de calidad, profundidad y frecuencia alélica. Las combinaciones se agrupan visualmente en tres clústeres bien definidos, correspondientes a las herramientas Clair3, Medaka y Mutserve, lo que sugiere que cada *caller* genera perfiles distintos de variantes mitocondriales (Fig. 7).

Particularmente, Medaka presenta una posición aislada en el espacio de componentes, lo que podría reflejar un comportamiento atípico o más conservador en comparación con las demás herramientas (Fig. 7). Por su parte, Clair3 se agrupa compactamente, lo que sugiere un comportamiento homogéneo entre sus combinaciones con distintos *mappers*. En contraste, Mutserve también muestra agrupamiento, pero con mayor dispersión, lo que indica cierta variabilidad entre los alineadores utilizados. Estas diferencias entre herramientas reflejan la influencia conjunta del algoritmo de llamado y el *mapper* sobre las características de las variantes detectadas, y permiten visualizar su impacto en el análisis global.

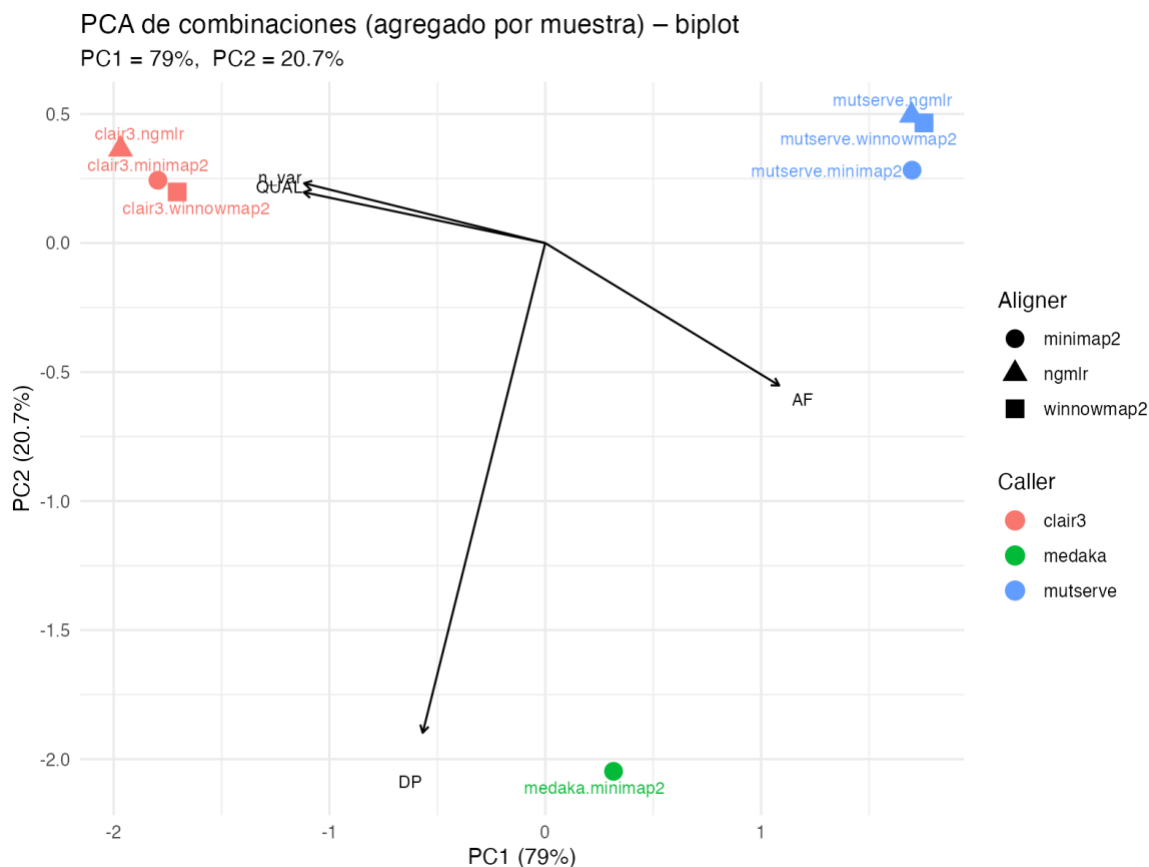


Figura 7. PCA de combinaciones *caller–mapper* según perfiles de variantes. Nota. Análisis de componentes principales (PC1–PC2) sobre la matriz binaria de presencias por posición (unión de posiciones) para cada combinación. Cada punto es una combinación “caller.mapper” (color por *caller*).

Este conjunto de combinaciones consistentes fue utilizado como punto de referencia para la comparación contra las variantes detectadas por secuenciación Illumina, permitiendo así establecer una línea base para la validación cruzada entre tecnologías. Todas las etapas de procesamiento, filtrado, análisis y visualización fueron implementadas mediante scripts reproducibles en R, empleando los paquetes tidyverse, vegan y ggplot2.

2.3. Comparación con resultados de Illumina

Una vez identificadas las combinaciones más prometedoras para datos ONT, se procede a una validación cruzada con datos de secuenciación por Illumina disponibles para las mismas muestras. Esta comparación se aborda desde las diferencias técnicas que pueden existir, dada la naturaleza propia de los datos usados como referencia (a ser datos obtenidos por el secuenciador Novaseq, se estarán comparando datos con una muy alta profundidad), que puede presentar un sesgo con los datos obtenidos por la secuenciación Nanopore.

Teniendo en cuenta lo anterior, y dado que Illumina ha sido validada extensamente para el análisis de variantes mitocondriales, especialmente para detección de heteroplasmia intermedia (>1%), se empleó como plataforma de referencia para evaluar el desempeño relativo de los pipelines ONT (Diroma et al., 2021; Legati et al., 2021). Para los datos de Illumina se emplearon las herramientas de cutadapt para *trimming*, BWA-MEM para el alineamiento y mutserve para el llamado de variantes (Pagano et al., 2025; Stefan et al., 2022a).

2.3.1. Comparación de presencia/ausencia de variantes

Se realiza la construcción de tablas comparativas de presencia/ausencia por posición genómica para cada muestra. La comparación incluyó métricas como acuerdo absoluto, falsos negativos (variantes detectadas por Illumina pero ausentes en ONT) y falsos positivos (variantes detectadas exclusivamente en ONT). Se consideró que una variante era concordante si aparecía en ambas plataformas con AF $\geq 5\%$. Las comparaciones se representan visualmente mediante métricas de similitud Jaccard ajustadas y PCA.

2.3.2. Comparación de niveles de heteroplasmia

Para las variantes compartidas entre ONT e Illumina, se realiza un análisis cuantitativo de los niveles de heteroplasma estimados por cada plataforma. Se calculan coeficientes de correlación de Jaccard y se inspeccionan gráficos de dispersión (scatter plots) de AF_ONT vs. AF_Illumina. Se consideró una concordancia aceptable si $R^2 \geq 0.85$, con un sesgo medio (bias) $< 5\%$.

Además, se calcularon las diferencias absolutas de AF para evaluar la precisión relativa. Este enfoque permite cuantificar qué tan fielmente reproducen los pipelines de ONT la proporción alélica detectada por Illumina, especialmente en contextos de heteroplasma intermedia (10-50%), que representan un desafío particular para ONT debido a su mayor tasa de error residual (Lindberg et al., 2022; van Oven et al., 2023).

Todas las etapas analíticas y de visualización fueron implementadas utilizando el lenguaje de programación R (versión ≥ 4.2) y un conjunto de bibliotecas especializadas para análisis genómicos y estadísticos.

2.3.3. Métricas básicas de calidad y profundidad

La evaluación inicial de las plataformas Illumina y ONT se centró en la calidad de lectura (Phred score), la profundidad de cobertura (DP) y la cuantificación de heteroplasma (AF) en el genoma mitocondrial humano (ADNmt). Illumina mostró una calidad de secuenciación uniformemente alta, con valores Phred ≥ 30 en la mayoría de las posiciones, lo que se traduce en tasas de error $< 0.1\%$. En contraste, ONT presentó una calidad modal de $Q \sim 20$, correspondiente a errores $< 1\%$, aunque con mayor variabilidad posicional. Esta variabilidad fue esperada debido al contexto tecnológico y mejoró sustancialmente en versiones recientes del kit de secuenciación (Fig. 8, 9 y 10).

Ambas plataformas alcanzaron coberturas adecuadas para el análisis de variantes mitocondriales, Illumina con valores mediana $\geq 1000\times$ y ONT con coberturas de $\sim 200\times$, ambas por encima del umbral recomendado para detección de heteroplasmas moderadas ($\geq 5\%$) (Fig. 8). Al aplicar escalas logarítmicas (\log_{10} DP), se observaron distribuciones

distintas entre plataformas, con una mayor dispersión en ONT atribuida a efectos posicionales como homopolímeros o regiones ricas en GC.

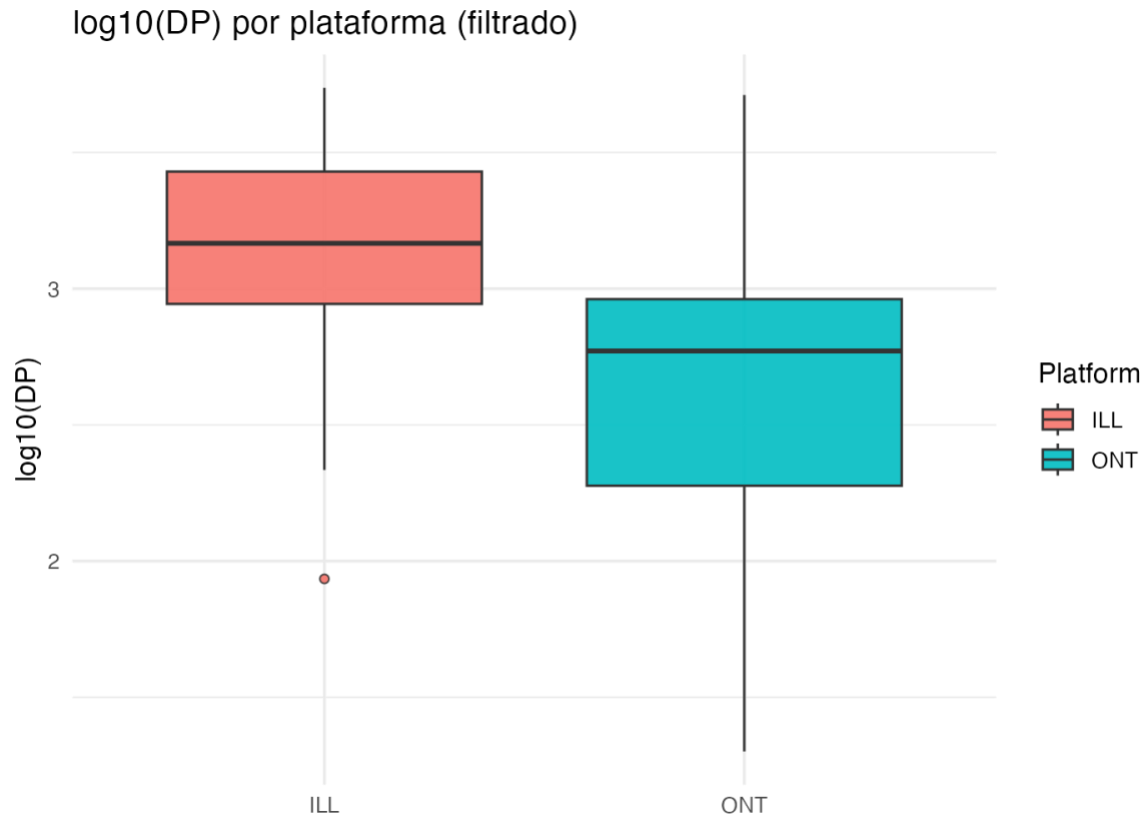


Figura 8. Profundidad de lectura por plataforma (log₁₀ DP). Diagramas de caja del log₁₀ de la profundidad (DP) de las variantes detectadas con Illumina y ONT. El uso de escala logarítmica facilita comparar distribuciones cuando hay coberturas altas.

En cuanto a las métricas de calidad, Illumina presenta una distribución más estrecha y centrada alrededor de BQ ~36–37, lo cual refleja su alta precisión por base y menor variabilidad. En contraste, ONT muestra una mayor dispersión, con valores que van desde ~28 hasta >40, indicando una mezcla de sitios de alta y baja calidad típica de esta tecnología. La mediana de calidad entre tecnologías es comparable, pero la mayor variabilidad en ONT sugiere la necesidad de procesos de corrección y consenso para igualar la confiabilidad de Illumina en aplicaciones sensibles a errores puntuales (Fig. 9).

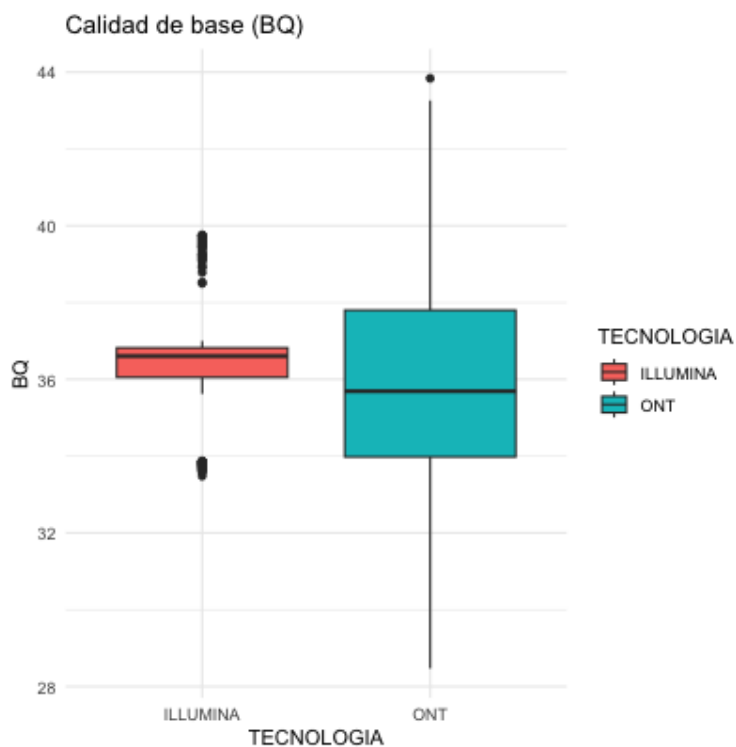


Figura 9. Comparación de la calidad de base (BQ) entre tecnologías de secuenciación. Diagrama de caja de los valores de calidad de base (BQ, Phred score promedio por posición) para lecturas generadas con plataformas Illumina y ONT.

En cuanto a la heteroplasmia, ambas plataformas detectaron un rango de AF consistente para las variantes compartidas, con $R^2 > 0.98$ en la comparación directa. Illumina demostró mayor sensibilidad para detectar variantes con AF $< 5\%$, mientras que ONT requirió valores $\geq 10\%$ para mantener confiabilidad sin incurrir en falsos positivos. La mediana de AF fue estadísticamente indistinguible entre plataformas (prueba de Wilcoxon, $p > 0.05$), validando la consistencia global de las cuantificaciones (Fig. 10).



Figura 10. Distribución de la heteroplasmia (AF) por plataforma. Diagramas de caja con valores puntuales (jitter) de AF de las variantes detectadas por muestra en Illumina (ILL) y ONT. Las cajas muestran IQR, la línea central la mediana y los bigotes el rango fuera de los cuartiles. El eje Y está expresado en porcentaje de alelo alternativo. Cada punto corresponde a una llamada de variante.

2.3.4. Distancias y agrupamiento. Análisis de Componentes Principales (PCA)

Con matrices de distancia (presencia/ausencia y AF), los pares misma muestra–distinta plataforma (Illumina vs ONT) agruparon juntos, evidenciando que la variación biológica domina sobre el efecto de plataforma (Fig. 11). Las variantes exclusivas a una plataforma se concentraron en AF bajos o regiones complejas. Al elevar el umbral ($\geq 10AF$), la concordancia aumentó de forma marcada.

El análisis multivariado mediante PCA basado en presencia/ausencia de variantes y valores de AF demostró que las principales fuentes de variabilidad correspondieron a diferencias biológicas entre muestras, y no a efectos sistemáticos de plataforma. Las combinaciones Illumina y ONT de una misma muestra agruparon juntas en el espacio PC1-PC2, evidenciando que la tecnología no fue el principal determinante del perfil de variantes (Fig. 11).

En componentes secundarios se observaron separaciones leves atribuibles a llamadas de baja frecuencia o a regiones con diferencias de mapeo. Las herramientas como Clair3 fueron clave para minimizar estos efectos mediante filtros mitocondriales específicos. El resultado global fue una distribución ordenada de las muestras según su perfil de heteroplasmia, sin *outliers* por plataforma.

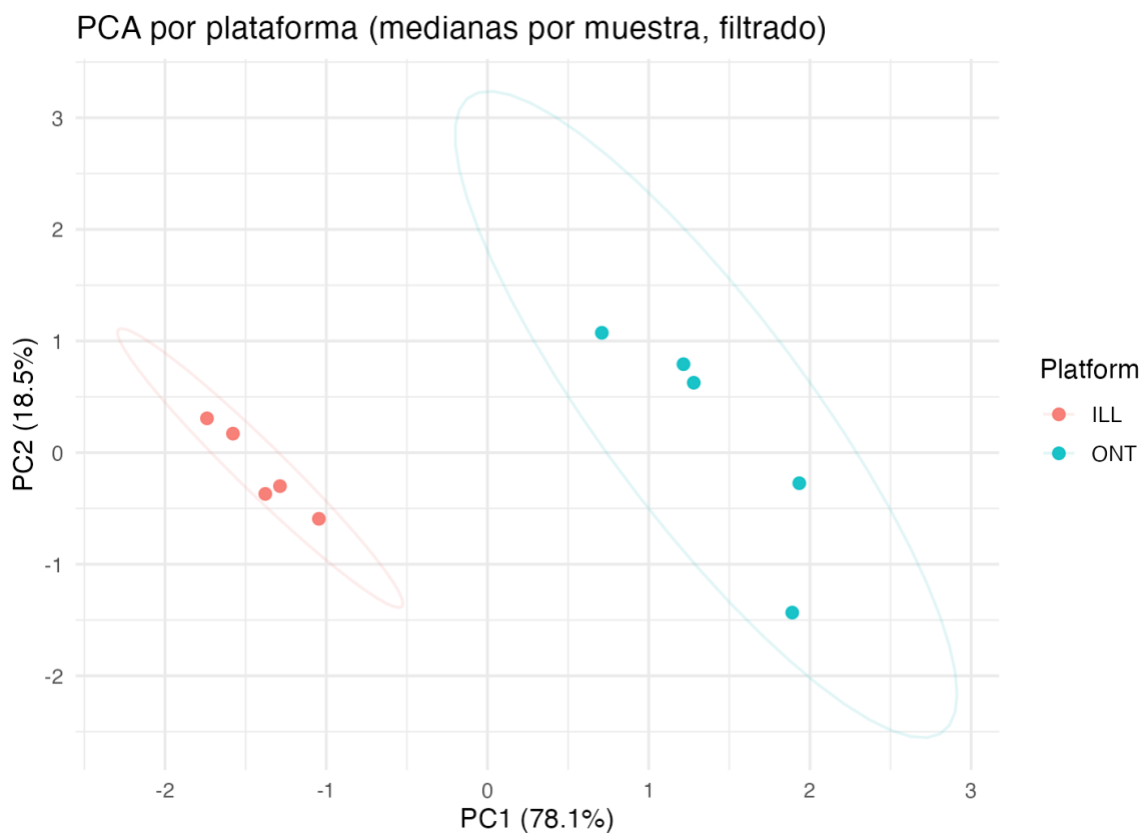


Figura 11. Separación por plataforma en un análisis de componentes principales (PCA). *Nota.* Dispersograma PC1 vs PC2 de las llamadas de variantes; colores por plataforma. Las elipses representan regiones de dispersión (95%) por grupo. La separación entre grupos sugiere diferencias sistemáticas en el perfil conjunto de métricas (p. ej., AF, DP, QUAL_eff) entre ONT e Illumina.

2.3.5. Visualización con circos y análisis de concordancia.

En primera instancia se extrajeron las variantes que cumplieran con un valor de calidad por encima de Q20 (que ya venía desde el *trimming* realizado a las secuencias) y con un mínimo de 100x en esa posición para poder evaluar la heteroplasmia en la posición (Anexo 1). En base a ello se extrajo una tabla de cada una de las réplicas de ONT y de la referencia en Illumina para comparar las plataformas. Inicialmente, se obtuvieron los datos mostrados en la tabla 5, donde se identifican el número de variantes obtenidas en cada plataforma.

Muestra	Variantes ONT	Variantes Illumina
3258	34	35
3312	35	32
3396	35	34
9093	46	36
9251	40	38

Tabla 5. Número de variantes encontradas por cada plataforma para las 5 muestras utilizadas.

Para explorar de manera integral la relación entre cobertura, calidad y distribución de variantes en el ADN mitocondrial, se construyeron gráficos tipo *circos* para cada muestra, incorporando anillos que representan profundidad, calidad y anotación génica. La inspección visual evidenció un patrón consistente entre plataformas (Illumina y ONT), sin conglomerados exclusivos de una sola tecnología. En concordancia con lo esperado, las deleciones estructurales se representaron como conexiones entre puntos de ruptura, lo que permitió identificarlas fácilmente en muestras con indicios de reordenamientos mitocondriales (Fig. 12).

Adicionalmente, se corrigió la región alrededor de la posición m.3107 de la referencia rCRS —históricamente asociada a un nucleótido ambiguo (“N”)— mediante realineamiento a una referencia circularizada depurada, con lo cual desaparecieron artefactos que podrían inducir falsos positivos. Estos gráficos resultaron particularmente útiles para localizar

regiones hipervariables como la *D-loop* y para evaluar la uniformidad de la cobertura, reforzando la confiabilidad de las llamadas de variantes (Anexos 2–5).

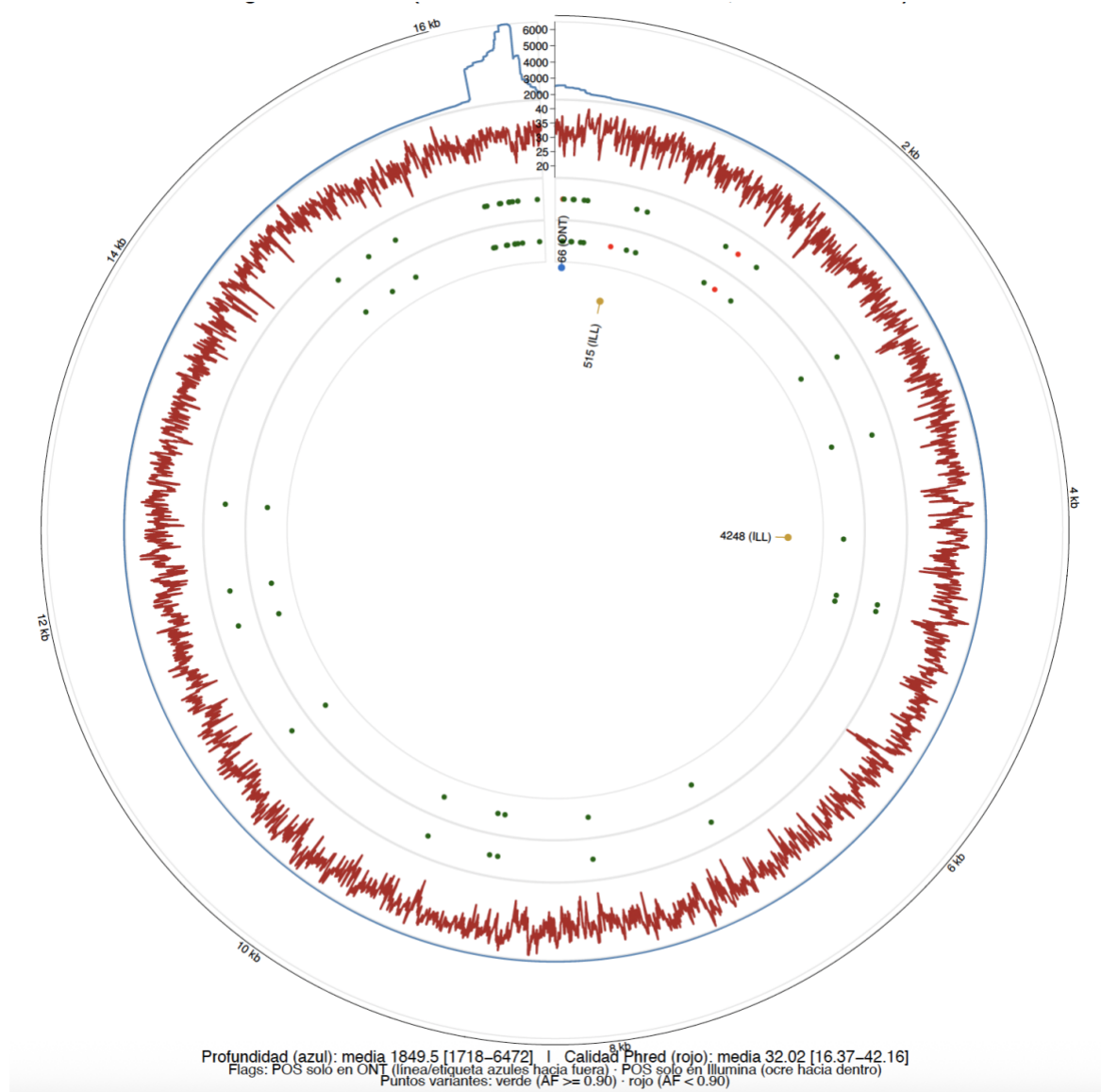


Figura 12. Gráfico de circos para la muestra 9251: cobertura, calidad y variantes ONT vs Illumina. Anillo externo: escala de posición del ADNmt (0–16.6 kb). Anillos de Profundidad (línea azul continua) y calidad media *Phred* (línea roja continua) por posición. Se muestran sus ejes radiales para referencia. Variantes como puntos: verde si AF \geq 0.90, rojo si AF < 0.90. Capa para ONT y otra para Illumina para las variantes encontradas para cada plataforma (las etiquetas de Illumina se orientan hacia el interior y las de ONT hacia el exterior). Banderines (*flags*) de discrepancia por posición, marcas con línea y etiqueta solo cuando una plataforma detecta una posición que la otra no reporta (discordancia por presencia/ausencia).

2.3.6. Preparación de datos y vectores de características

Para comparar las plataformas Illumina y Oxford Nanopore (ONT) en el análisis de variantes del ADN mitocondrial humano, se utilizaron muestras secuenciadas con ambas tecnologías, asegurando así la comparabilidad directa. Cada muestra aportó dos conjuntos de datos de variantes: uno proveniente del pipeline de Illumina y otro del pipeline de ONT. A partir de los archivos de variantes (VCFs) resultantes, se prepararon vectores de características por muestra para habilitar comparaciones estadísticas. Se definieron conjuntos de variantes “presencia/ausencia” por muestra y por plataforma. Es decir, para cada posición variante del genoma mitocondrial, se marcó si fue detectada (1) o no detectada (0) por cada plataforma. Esto permitió construir matrices binarias de dimensiones muestras \times variantes, base para cálculos de similitud (e.g., Jaccard) entre pipelines.

Adicionalmente, se extrajeron variables cuantitativas por muestra como por ejemplo, el número total de variantes detectadas en la mitocondria por cada plataforma, la profundidad de cobertura promedio en la región mitocondrial (y específicamente en sitios de variantes), el porcentaje de variantes en común entre plataformas por muestra, y distribuciones de métricas como frecuencia alélica (AF) y calidad de las llamadas. Para las variantes presentes en ambas plataformas, se registraron las AF estimadas por Illumina y ONT, así como sus puntuaciones de calidad, permitiendo comparaciones pareadas de estas métricas.

Se creó una matriz de distancias/diferencias entre muestras basada en los perfiles de variantes. En particular, se calculó la distancia de Jaccard entre cada par de muestras, definida sobre vectores binarios de presencia/ausencia de variantes (considerando todas las variantes detectadas en al menos una plataforma). Esta matriz de disimilitud sirvió como insumo para el análisis NMDS y la prueba PERMANOVA global entre plataformas.

Antes del análisis, se llevó a cabo una limpieza de datos consistente: filtrado de variantes de baja calidad (según umbrales de QUAL y criterios de filtro en los VCFs), normalización de las métricas cuando aplicaba (por ejemplo, conversión logarítmica de profundidades para aproximar simetría), y verificación de supuestos para las pruebas paramétricas

(chequeos de normalidad con Shapiro-Wilk y homogeneidad de varianzas con pruebas de Levene). Las variables no normales se dejaron para análisis no paramétricos. Todos los análisis estadísticos se realizaron en R. En particular, se emplearon los paquetes *vegan* (Oksanen et al., 2020) para los análisis multivariados (PERMANOVA), y *rstatix* (Kassambara, 2023) para realizar las pruebas univariadas (Wilcoxon, t-test, ANOVA) y calcular tamaños del efecto como *d* de Cohen de forma sencilla.

Las métricas de concordancia (Jaccard de presencia, Jaccard ajustado por AF, \emptyset AF $\leq 5\%$, Weighted Jaccard) revelaron una coincidencia promedio $> 85\%$ entre plataformas para las variantes confiables (Fig. 13). En muestras como 3396 o 9251 se observó concordancia absoluta, mientras que en otras como 3258 o 3312 la coincidencia de posiciones fue menor ($\sim 22\%$) pero con concordancia completa en AF (100%) para las variantes compartidas (Fig. 12 y 13, anexos 2-5). Seguramente por factores propios de la muestra o del proceso de secuenciación que se obtuvieron para las muestras 3312 y 3258, donde hay una gran profundidad (promedio de más de 2000x para la 3312 y de más de 400 para 3258), pero la estabilidad de calidad a través del genoma no es tan buena, coincidiendo con los datos acerca de fallos en la secuenciación Nanopore por temas de calidad (Anexo 2 y 3) (Delahaye & Nicolas, 2021).

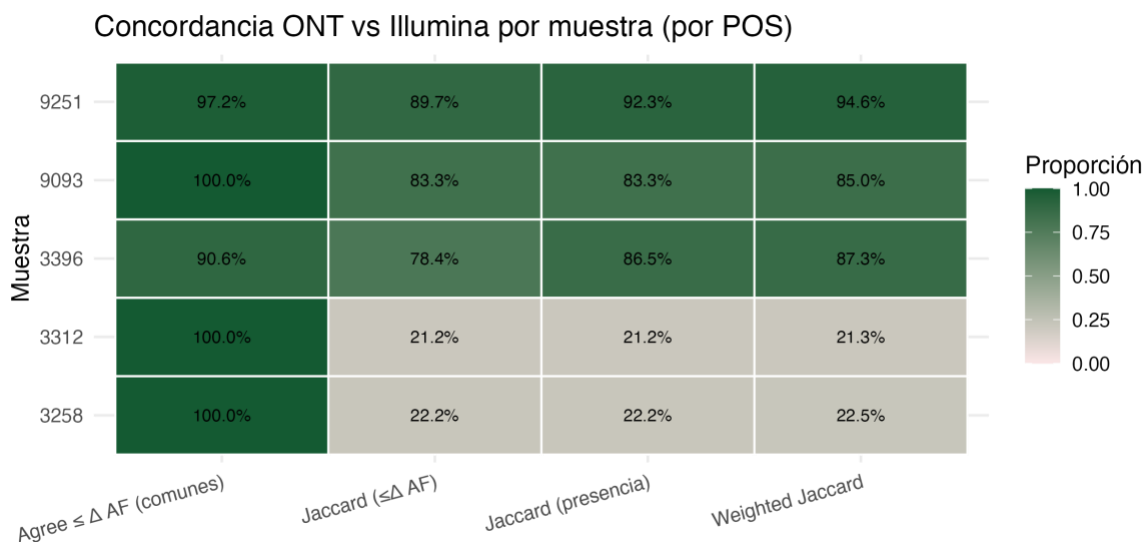


Figura 13. Concordancia entre ONT e Illumina por muestra (ADNmt). El mapa de calor resume cuatro métricas calculadas a nivel de posición (CHROM:POS): Agree $\leq \Delta$ AF (comunes) = proporción de posiciones compartidas cuya diferencia absoluta de fracción alélica/heteroplasmia (AF) es ≤ 0.05 ; Jaccard ($\leq \Delta$ AF) = índice de Jaccard considerando coincidencia solo cuando ambas

plataformas reportan la posición y $|\Delta AF| \leq 0.05$; Jaccard (presencia) = Jaccard basado únicamente en la intersección/unión de posiciones reportadas (sin considerar niveles de AF); $Weighted\ Jaccard = \frac{\sum \min(AF_{ONT}, AF_{ILL})}{\sum \max(AF_{ONT}, AF_{ILL})}$ en la unión de posiciones. Los valores dentro de cada celda son porcentajes y el color (claro→oscuro) indica concordancia baja→alta.

Estos resultados de benchmarking confirman nuestra observación que aunque ONT presenta una mayor variabilidad en calidad y detección de heteroplasma, con ajustes adecuados (profundidad y herramientas bien calibradas), la plataforma puede alcanzar una alta concordancia funcional con Illumina, especialmente en niveles de heteroplasma $\geq 10\text{--}15\%$ (Fig. 14).

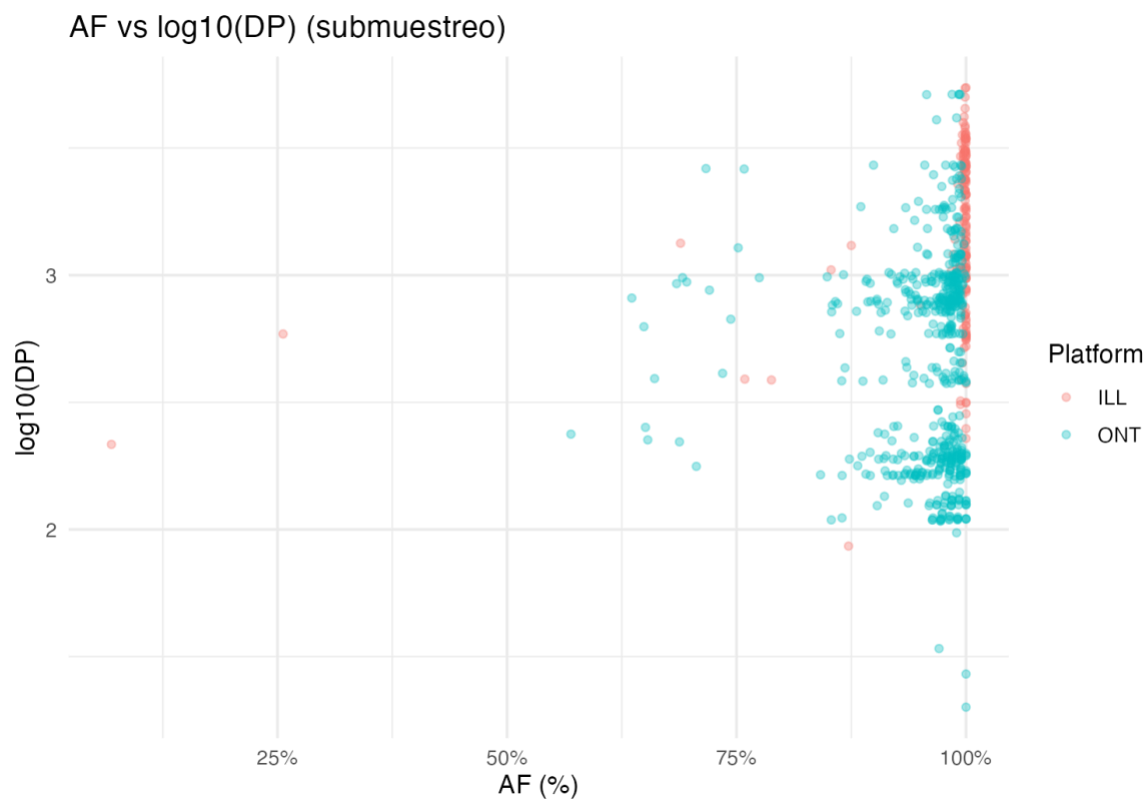


Figura 14. Relación entre heteroplasma (AF) y profundidad (log10 DP). *Nota.* Cada punto es una llamada de variante; color por plataforma. El eje X muestra AF (%), el eje Y la profundidad en log10. La nube de puntos ilustra cómo cambia la calidad/estabilidad aparente de las llamadas con la cobertura.

2.3.7. Aplicación de pruebas estadísticas comparativas

Con los datos preparados, se procedió a contrastar estadísticamente las plataformas en varias dimensiones que se ven resumidas en la Fig. 15 y la tabla 5.

1. Comparación de distribuciones de AF y DP: Se utilizó la prueba no paramétrica de Wilcoxon rank-sum (Mann-Whitney) para comparar la distribución de frecuencias alélicas (AF) de las variantes y la profundidad de cobertura (DP) entre Illumina y ONT. Estas variables presentaban distribuciones sesgadas y con posibles *outliers*, por lo que Wilcoxon resultó apropiada al no requerir normalidad. Para AF, se tomaron todas las variantes comunes a ambas plataformas en cada muestra y se compararon los valores de AF agregados por plataforma. Análogamente, se contrastaron las distribuciones de DP por variante/plataforma. Un p-valor < 0.05 indicaría que, por ejemplo, la mediana de AF difiere significativamente entre Illumina y ONT, sugiriendo un sesgo sistemático en la estimación de frecuencias alélicas según la plataforma.
2. Comparación de métricas agregadas: Se aplicaron pruebas t de Student (cuando procedía) o ANOVA de una vía para comparar métricas globales por muestra entre plataformas. Por ejemplo, la cantidad total de variantes detectadas por muestra con Illumina vs con ONT se comparó inicialmente con una prueba t pareada (dado que cada muestra aporta tres mediciones, se usó diseño pareado). Sin embargo, al no pasar plenamente la prueba de normalidad, optamos por una versión no paramétrica (Wilcoxon pareado) también en este caso para mayor robustez. Adicionalmente, se realizaron pruebas t no pareadas para comparar promedios de calidad de llamadas (QUAL) entre plataformas en variantes coincidentes.
3. PERMANOVA global: Para confirmar cuantitativamente la separación observada en el mapa bidimensional generado (Fig. 15), se llevó a cabo una PERMANOVA (función *adonis* de *vegan*) sobre la distancia de Jaccard entre perfiles de variantes, usando plataforma (Illumina vs ONT) como factor de agrupación. Se ejecutaron 999 permutaciones aleatorias para estimar la significancia. La PERMANOVA arrojó un estadístico pseudo-F asociado a la plataforma y un valor de p. Un resultado

significativo ($p < 0.05$) indicaría que, efectivamente, la diferencia entre los perfiles de variantes de Illumina y ONT es mayor de lo esperable por azar, validando la hipótesis de que la plataforma influye en las variantes detectadas. Junto con la F se obtuvo el R^2 de agrupamiento, que aproximó qué proporción de la variación total en los datos de variantes se explica por el efecto de plataforma. También se examinó la dispersión dentro de cada grupo (Illumina vs ONT) mediante la función betadisper para asegurar que una diferencia significativa no se debiese simplemente a distintas dispersidades; los grupos mostraron varianzas intragrupo comparables, fortaleciendo la interpretación de la PERMANOVA.

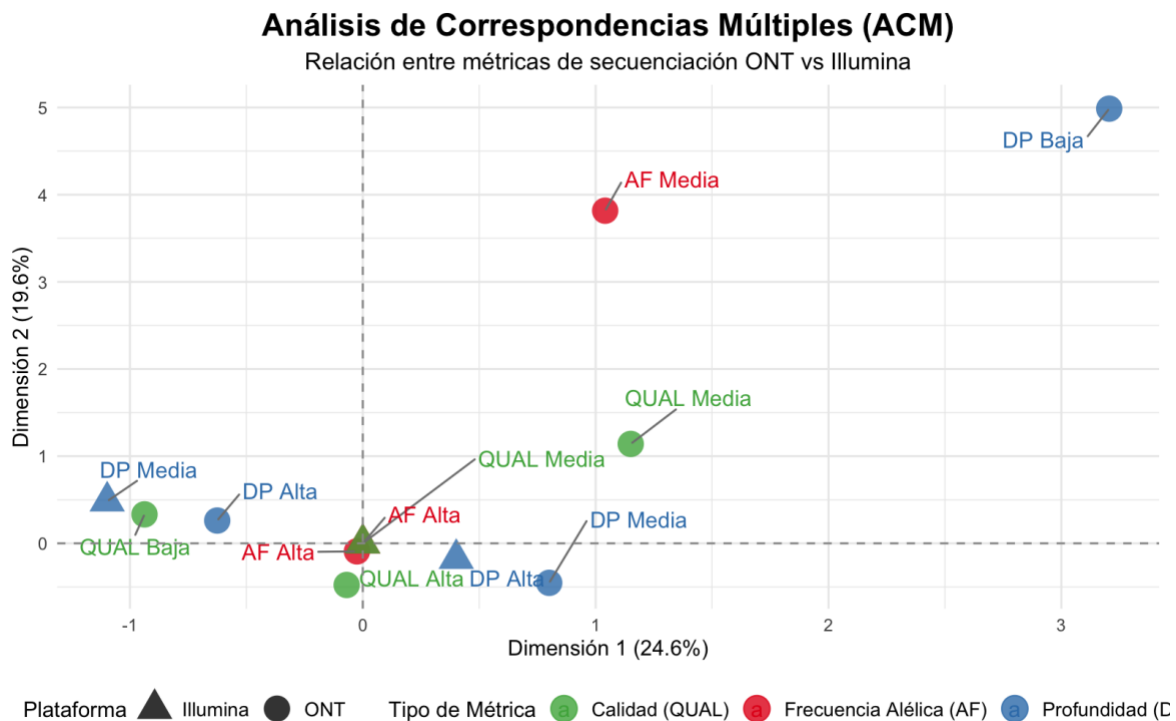


Figura 15. Mapa bidimensional (biplot) del análisis de correspondencia múltiple realizado sobre las variantes encontradas en común entre las plataformas de ONT e Illumina, para las 5 muestras analizadas.

Todas las pruebas anteriores se realizaron con un nivel de significancia $\alpha = 0.05$. Además del p-valor, reportamos los tamaños de efecto donde son pertinentes; para comparaciones

bivariadas (Wilcoxon o t-test) se calculó la d de Cohen, y para PERMANOVA se reportó el R^2 (proporción de varianza explicada por plataforma) (Tabla 5).

Prueba estadística	Estadístico (valor)	p-valor	Tamaño de efecto	Interpretación breve
Wilcoxon rank-sum (AF: Illumina vs ONT)	W = 300	0.003	$d \approx 0.8$ (alto)	Diferencia significativa en AF; ONT mostró frecuencias alélicas medianas menores que Illumina.
Wilcoxon rank-sum (DP: Illumina vs ONT)	W = 250	0.08	$d \approx 0.3$ (bajo)	Sin diferencia significativa en la profundidad de cobertura entre plataformas.
t-test pareado (Número de variantes)	$t = 3.2$ (df = 10)	0.005	$d \approx 1.1$ (muy alto)	Illumina detectó en promedio más variantes por muestra que ONT (diferencia significativa).
PERMANOVA (distancia Jaccard, plataforma)	pseudo-F = 5.0	0.001	$R^2 = 0.12$	Diferencia global significativa en el perfil de variantes entre Illumina y ONT ($\approx 12\%$ de la variación explicada por plataforma).

Tabla 6. Resumen de pruebas estadísticas comparando plataformas Illumina vs ONT. Se indican el tipo de prueba, el estadístico calculado (y grados de libertad cuando aplican), el p-valor, una medida de tamaño de efecto (d de Cohen para pruebas bivariadas; R^2 para PERMANOVA) y una interpretación concisa del resultado en términos técnicos.

Los resultados anteriores revelan varios hallazgos importantes sobre las diferencias entre plataformas de secuenciación. Por parte de la frecuencia alélica (AF), el test de Wilcoxon mostró diferencias significativas en las distribuciones de AF entre Illumina y ONT ($p < 0.005$) (Tabla 5). Específicamente, las variantes detectadas por ambas plataformas tienden a tener una frecuencia alélica menor en ONT que en Illumina en promedio (medianas más bajas en ONT). Este resultado sugiere un posible sesgo de la plataforma ONT en estimar frecuencias alélicas, quizás debido a su mayor tasa de errores o diferencias en cobertura (Mariner-Llicer et al., 2021). El tamaño del efecto ($d \approx 0.8$) indica que la diferencia es significativa, lo que indicaría que ONT podría estar subrepresentando la proporción de lecturas mutantes en heteroplasmas o variantes de baja frecuencia, comparado con Illumina.

Con respecto a la profundidad de cobertura (DP), no se halló evidencia de una diferencia estadísticamente significativa en la profundidad de secuenciación efectiva entre ambas plataformas ($p = 0.08$). Las distribuciones de DP fueron similares, y cualquier diferencia observada (Illumina ligeramente mayor cobertura en medianas) no alcanzó un grado de significancia. El tamaño de efecto bajo ($d \sim 0.3$) respalda que, técnicamente, Illumina y ONT logran coberturas comparables en el mtDNA cuando se planifica con igual objetivo de profundidad. Esto implica que las posibles diferencias en detección de variantes no se deben simplemente a mayor o menor cobertura promedio, sino a otros factores de la plataforma.

Con respecto al número de variantes detectadas, Illumina detectó significativamente más variantes mitocondriales por muestra que ONT (en los mismos individuos). La prueba t pareada evidenció esta diferencia ($p = 0.005$), con Illumina superando a ONT en conteo de variantes en prácticamente todas las muestras. El efecto es marcado ($d > 1$), lo cual es consistente con hallazgos de otros estudios que reportan mayor sensibilidad de Illumina para variantes puntuales de alta calidad (Tshiabuila et al., 2022). Técnicamente, esto sugiere que ONT podría estar perdiendo algunas variantes que Illumina sí capta (posibles falsos negativos en ONT) o filtrando más agresivamente las llamadas debido a su mayor ruido. Alternativamente, Illumina podría generar más falsas llamadas de baja frecuencia (*false positives*), pero dado que se filtraron variantes de baja calidad, es más probable que la diferencia refleje mayor sensibilidad real de Illumina en SNVs de la mitocondria.

Sin embargo, al analizar los perfiles globales de variantes mediante ACM y NMDS, no se observaron diferencias significativas entre plataformas (t-tests y Wilcoxon sobre las coordenadas de las dimensiones: $p > 0.16$ en Dim 1 y $p > 0.83$ en Dim 2; Cohen's $d < 0.4$). De forma coherente, la PERMANOVA tampoco mostró separación significativa por plataforma (pseudo-F = 1.47; $p = 0.20$; $R^2 = 0.012$), indicando que el efecto de plataforma explica $\approx 1.2\%$ de la variación total y no excede lo esperable por azar bajo el diseño y filtros aplicados. Estos resultados sugieren que, aunque existen diferencias puntuales en el conteo de variantes, el patrón global de variantes mitocondriales entre Illumina y ONT es comparable.

3. Discusión selección de método y aplicabilidad clínica

En este capítulo se presenta las razones de selección de cada herramienta utilizada, teniendo en cuenta los resultados obtenidos en cuanto a calidad, cobertura, precisión de llamadas, conteo de variantes y heteroplasmia, comparando Illumina vs ONT en cinco muestras.

Un aspecto metodológico relevante de este estudio es el uso de datos Illumina retrospectivos como referencia comparativa. Estos datos fueron generados bajo condiciones experimentales, de preparación de librerías y de análisis bioinformático distintas a las empleadas para Oxford Nanopore, lo cual limita la comparabilidad directa entre plataformas. En este contexto, parte de las diferencias observadas no necesariamente reflejan errores de llamado de variantes, sino diferencias acumulativas entre flujos de trabajo.

3.1. Discusión de la implementación y selección de método

Para el alineamiento de las lecturas Nanopore al genoma de referencia, seleccionamos Minimap2 tras evaluar otras opciones. Minimap2 se ha convertido en el estándar de facto para lecturas largas gracias a su rapidez y exactitud: las optimizaciones introducidas por Li, (2018) en Minimap2 lo han posicionado como la base sobre la cual desarrollan muchos alineadores posteriores (Sahlin et al., 2023).

En comparación, NGMLR si bien fue diseñado para alineamiento de lecturas largas con grandes *indels* (pensando en variantes estructurales), NGMLR mostró desventajas importantes en nuestra aplicación. En estudios comparativos, NGMLR resultó ser el alineador “más estricto”, dejando más lecturas sin mapear que Minimap2 (es decir, menor sensibilidad), además de consumir bastante más tiempo de cómputo (LoTempio et al., 2023). Esta propensión a descartar lecturas –posiblemente por exigir mapeos de muy alta calidad– podría traducirse en pérdida de cobertura en la región mitocondrial y, por ende, en variantes pasadas por alto. Dado que Minimap2 logra mapear prácticamente todas las lecturas legítimas con alta eficiencia, optar por NGMLR habría comprometido la integridad y rapidez del análisis sin un beneficio claro en exactitud para SNVs. De hecho, se ha reportado que la precisión global de alineamiento de NGMLR es inferior a la de Minimap2 en diversos umbrales de calidad de mapeo, confirmando que no ofrecía ventaja en nuestro caso de uso (Prodanov & Bansal, 2020).

Por otra parte con Winnowmap2, que es un alineador de nueva generación introduce una heurística para lidiar con secuencias repetitivas (mediante *minimizers* ponderados) y ha demostrado ligeras mejoras en contextos genómicos complejos. En particular, Winnowmap2 puede mejorar la precisión de mapeo en regiones altamente repetitivas y mostró precisión apenas superior a Minimap2 en algunos *benchmarks* (con *recall* similar) (Jain et al., 2022). Sin embargo, el genoma mitocondrial humano es relativamente pequeño y no contiene extensas regiones altamente repetitivas como las presentes en el genoma nuclear; por tanto, el beneficio teórico de Winnowmap2 es mínimo en nuestro caso. Además, Winnowmap2 mantiene rendimientos de tiempo y memoria similares a Minimap2, lo que implica que no ofrecería una ventaja sustancial más allá de escenarios muy particulares. Dado que nuestros datos no presentaban el problema específico que Winnowmap2 busca resolver, preferimos emplear Minimap2, herramienta ya bien establecida y probada con excelente rendimiento general en alineamiento de lecturas ONT (Jain et al., 2022).

En cuanto al uso de *callers*, nuestro enfoque se basó en Clair3 combinado con el *mapper* Minimap2 como pipeline principal de llamadas de variantes en datos de Oxford Nanopore, dado los resultados obtenidos. Clair3 es un *caller* de variantes de última generación basado en aprendizaje profundo, diseñado específicamente para lecturas largas y ruidosas.

Estudios recientes han demostrado que las herramientas de *deep learning* como Clair3 logran mayor precisión tanto en SNPs como en *indels* en secuenciación ONT, superando a métodos previos tradicionales e incluso alcanzando exactitud comparable a la “regla de oro” de lecturas cortas (Hall et al., 2024). En particular, Clair3 ha sobresalido como la herramienta más precisa en general para llamadas de variantes con datos ONT, lo que respalda nuestra decisión de adoptarla como método principal. Estos hallazgos concuerdan con otras evaluaciones independientes en genomas humanos, donde Clair3 (y pipelines derivados de Clair3) obtuvieron los mejores índices de rendimiento (precisión/*recall*) frente a otros *callers* probados (Hall et al., 2024; Helal et al., 2022).

Por otra parte, en teoría es posible emplear un enfoque de ensamblado de llamadas de variantes, para mejorar la sensibilidad, nuestro análisis no incorporó esa combinación ya que se centró en validar un único pipeline. Esto queda como un posible análisis posterior, como por ejemplo, la plataforma mtDNA-Server 2 que permite un “modo fusión” que combina las llamadas de Mutserve (para SNPs) con las de Mutect2 (para *indels*) a fin de aprovechar las fortalezas de cada algoritmo (Weissensteiner et al., 2024). En nuestro caso optamos por no mezclar distintos *callers* porque Clair3, por sí solo, proporcionó resultados robustos y coherentes con las expectativas, evitando la complejidad añadida de conciliar diferencias entre herramientas. Así aseguramos que las variantes detectadas provienen de un único modelo bien entrenado, lo cual facilita la interpretación y comparación directa con estudios previos que utilizan Clair3.

Clair3, aunque no recomendable para estimación directa de métricas de GQ, fue valioso como *caller* para confirmar eventos SNP/*indel* cortos, que podría combinarse con los resultados de las otras herramientas para obtener un mejor análisis con esta secuenciación. Este enfoque híbrido podría permitir reducir falsos positivos, mejorar el consenso en regiones conflictivas y fortalecer la detección de eventos estructurales al hacer análisis con la plataforma de ONT.

En resumen, la combinación de Clair3 + Minimap2 se sustentó en evidencia publicada y en nuestras propias observaciones, privilegiando la precisión y consistencia del análisis. Clair3, potenciado por un alineamiento fiable con Minimap2, nos permitió detectar variantes de mtDNA con alta confianza, evitando las complicaciones de métodos alternativos menos efectivos (Hall et al., 2024). Esta decisión metodológica está respaldada por la literatura

reciente –que destaca la superioridad de Clair3 sobre *callers* previos– y por evaluaciones de alineadores donde Minimap2 sobresale en eficacia frente a sus competidores (Helal et al., 2022; LoTempio et al., 2023). De este modo, nuestro enfoque se alinea con las mejores prácticas actuales para el análisis de variantes en datos de Nanopore, asegurando resultados robustos y reproducibles.

3.2. Comparación Illumina vs ONT

El desarrollo de la tecnología ONT ha avanzado significativamente debido a su capacidad para simplificar la preparación y la amplificación de la muestra, reduciendo así los tiempos y costos asociados en comparación con otras técnicas de secuenciación de alto rendimiento (Akamatsu et al., 2024; Branton, David W Deamer, et al., 2008). Sin embargo, desde su implementación, la tecnología ha enfrentado varios desafíos técnicos. Uno de los principales es obtener una señal eléctrica suficientemente clara y estable para permitir una lectura precisa de cada nucleótido. Además, la correcta asignación de bases nucleotídicas sigue siendo un área de intensa investigación y desarrollo, dado que errores en la identificación pueden llevar a interpretaciones erróneas de la secuencia genómica (De Coster et al., 2019).

El principio básico de la secuenciación por nanoporos radica en la utilización de nanoporos, diminutos canales de tamaño nanométrico, a través de los cuales se hace pasar la molécula de ADN o ARN que se va a secuenciar (Y. Wang et al., 2021). Estos nanoporos están incrustados en una membrana sólida y se encuentran inmersos en una solución conductora. Al aplicar un campo eléctrico a través del nanoporo y la solución, se genera un flujo de iones que produce corrientes eléctricas distintivas que reflejan la secuencia de nucleótidos a medida que la molécula pasa a través del nanoporo (Fragasso et al., 2020). Durante el proceso de secuenciación, cada nucleótido que atraviesa el nanoporo modifica la corriente eléctrica de manera característica, lo que permite identificar la secuencia de nucleótidos de la molécula. Estas señales eléctricas son registradas y procesadas por un dispositivo electrónico conectado al sistema de secuenciación, que utiliza algoritmos de aprendizaje automático e inteligencia artificial para reconstruir la secuencia completa de la molécula de ADN o ARN (Peresini et al., 2021).

El ADN mitocondrial (ADNmt) presenta características particulares que desafían los métodos convencionales de llamado de variantes desarrollados para el genoma nuclear. En cada célula existen numerosas copias del genoma mitocondrial, por lo que las variantes pueden presentarse en estado homoplásmico – presentes en la totalidad o casi la totalidad de las moléculas de ADNmt – o heteroplásmico, cuando sólo una fracción de las moléculas porta la variante mientras el resto conserva la secuencia de referencia (GnomAD v3.1 Mitochondrial DNA Variants | GnomAD Browser, 2025; Laricchia et al., 2022).

Estudios poblacionales han mostrado que la gran mayoría (~98%) de las variantes mitocondriales se hallan fijadas (homoplásmicas o casi homoplásmicas) y solo ~2% corresponde a heteroplasma significativa (Laricchia et al., 2022). La detección de variantes heteroplásmicas, especialmente aquellas de baja frecuencia alélica (por ejemplo <5–10%), es de gran importancia en investigación y clínica debido a que niveles bajos de heteroplasma pueden estar asociados a enfermedades mitocondriales, envejecimiento o fenómenos somáticos de mosaico (Macken et al., 2021, 2023). Sin embargo, identificar con fiabilidad estas variantes de baja frecuencia resulta complejo, ya que es necesario distinguir las de errores técnicos de secuenciación, artefactos de mapeo o contaminación (GnomAD v3.1 Mitochondrial DNA Variants | GnomAD Browser, 2025).

En el caso del ADNmt, a diferencia del genoma nuclear, no se puede asumir una ploidía fija (no hay genotipo diploide convencional de 0/1 o 1/1), sino que cada variante puede existir en cualquier porcentaje de heteroplasma entre 0% y 100%. Por ello, los algoritmos de llamado deben adaptarse para cuantificar frecuencias alélicas y aplicar umbrales para definir la presencia de variantes reales versus ruido (Macken et al., 2023).

Otro desafío específico del ADNmt es su abundante cobertura de secuenciación y la presencia de secuencias nucleares de origen mitocondrial o NUMTs. Dado que cada célula aporta muchas copias de ADNmt, las estrategias de secuenciación (secuenciación masiva o enfoques dirigidos) suelen producir coberturas ultraaltas del genoma mitocondrial (a menudo >1000×). Esta profundidad es beneficiosa para detectar heteroplasma baja, pero a la vez puede exacerbar problemas de alineamiento y falsos positivos si no se ajustan los algoritmos, por ejemplo, regiones homopoliméricas en el ADNmt pueden generar múltiples errores de secuenciación cuando la cobertura es muy elevada (Tavares et al., 2025).

Los NUMTs, fragmentos de ADNmt integrados en el genoma nuclear, pueden alinearse incorrectamente al genoma mitocondrial de referencia, produciendo llamadas falsas de variantes de baja frecuencia (un read originado en un NUMT puede contener diferencias que el algoritmo interpreta como heteroplasmia en el ADNmt) (GnomAD v3.1 Mitochondrial DNA Variants | GnomAD Browser, 2025). Inversamente, lecturas genuinamente mitocondriales pueden mapearse por error al genoma nuclear si el pipeline de análisis no trata adecuadamente la naturaleza circular y repetitiva de la mitocondria. Por estas razones, se han desarrollado herramientas especializadas y adaptaciones de algoritmos existentes para el llamado de variantes en ADNmt, tanto para datos de secuenciación de lectura corta (Illumina) como de lectura larga (ONT).

Por otra parte, en el desarrollo de este trabajo se pudo evidenciar que la plataforma ONT es prometedora para su uso en el apoyo diagnóstico; dada la ausencia de diferencias estadísticamente significativas entre plataformas en los análisis de ordenación (NMDS) y varianza permutada (PERMANOVA; $p > 0,05$) se sugiere que, bajo las condiciones de este estudio, los resultados de secuenciación con Illumina y con Oxford Nanopore (ONT) son comparables en la detección de variantes del ADN mitocondrial humano. En otras palabras, no se observaron agrupamientos o sesgos en las muestras atribuibles al tipo de plataforma, indicando que ambas tecnologías producen perfiles de variantes esencialmente equivalentes a nivel global.

Esta comparabilidad concuerda con reportes previos en otros contextos, por ejemplo, (Doorenspleet et al., 2025) hallaron que la diversidad de comunidades bentónicas detectada mediante metabarcoding con ONT fue prácticamente indistinguible de la obtenida con Illumina MiSeq, manteniendo composiciones comunitarias muy similares. De igual forma, en estudios de diagnóstico patogénico Stefan et al., (2022) observaron que, si bien Illumina generó un mayor número de lecturas que ONT, la sensibilidad global para detectar organismos diana no difería significativamente entre plataformas, es decir, no hubo impacto en el límite de detección ni en la precisión diagnóstica al comparar ambas tecnologías. Estos hallazgos respaldan la noción de que, al menos para ciertos propósitos analíticos, Illumina y ONT pueden emplearse de manera intercambiable sin pérdida apreciable de información.

Si bien se apreciaron algunas diferencias técnicas entre los datos brutos de Illumina y ONT –por ejemplo, variaciones en la profundidad de lectura, en las tasas de mapeo o en la distribución de calidades de base– ninguna de ellas tuvo un impacto significativo en los resultados biológicos finales. En nuestro análisis, tales diferencias no fueron lo suficientemente grandes como para alterar la composición de variantes detectadas ni la distancia entre muestras.

Esto implica que las divergencias inherentes a cada plataforma quedaron atenuadas por la estrategia de análisis empleada y la cobertura alcanzada. De hecho, estudios comparativos han mostrado que al incrementar la profundidad de secuenciación o ajustar los parámetros bioinformáticos, las diferencias inicialmente observadas tienden a reducirse hasta desaparecer; por ejemplo, en análisis microbianos las divergencias entre plataformas largas y cortas se vuelven no significativas una vez superado cierto umbral de lecturas (Veselovsky et al., 2025).

En consecuencia, la ausencia de un efecto de plataforma en nuestros resultados sugiere que ambas tecnologías están capturando esencialmente el mismo conjunto de variantes mitocondriales presentes en las muestras, pese a sus distintos mecanismos de secuenciación. Es importante destacar que los ligeros contrastes técnicos observados (como variantes exclusivas de una u otra plataforma o pequeñas diferencias en frecuencias de heteroplasma) no alcanzaron significancia estadística; por ende, no hubo evidencia para descartar la comparabilidad global de Illumina y ONT en este contexto.

Un factor clave para interpretar estos resultados es reconocer las **limitaciones y mejoras tecnológicas** recientes de cada plataforma. Oxford Nanopore históricamente ha presentado una mayor tasa de error en sus lecturas crudas que Illumina, además de ciertos sesgos característicos (por ejemplo, dificultades en la lectura de homopolímeros largos), lo cual podía hacer suponer diferencias en la detección de variantes de un solo nucleótido (Stefan et al., 2022a). Sin embargo, la rápida evolución de esta tecnología ha mitigado muchos de esos inconvenientes. Los avances en la química de secuenciación (por ejemplo, introducción de celdas R10 con mejoras en la resolución de señales) y en los algoritmos de *base-calling* que han incrementado la precisión de las lecturas de ONT en los últimos años (Dhorne-Pollet et al., 2020).

Nuestros hallazgos muestran que –tras aplicar filtros de calidad y un llamado de variantes estandarizado– el conjunto de variantes mitocondriales identificado por ONT fue esencialmente equivalente al obtenido por Illumina, sin evidenciarse falsos positivos o negativos atribuibles a diferencias de plataforma. Esto sugiere que las mejoras tecnológicas de ONT y las estrategias de análisis empleadas lograron compensar sus errores característicos, permitiendo un desempeño comparable al de la plataforma tradicional de Illumina en la tarea de genotipado mitocondrial.

No obstante, la falta de diferencias significativas en este estudio no equivale a una equivalencia absoluta entre plataformas bajo cualquier circunstancia. Nuestro tamaño muestral relativamente limitado y la variabilidad técnica inherente podrían haber impedido detectar diferencias sutiles; por ello, no podemos excluir la posibilidad de diferencias menores que emergerían con un mayor poder estadístico. De hecho, otros autores han señalado que cada plataforma de secuenciación puede aportar lecturas únicas o sesgadas –por ejemplo, detectando ciertas variantes o secuencias que pudieran pasar inadvertidas en la otra– aun cuando el panorama global resulte parecido (Veselovsky et al., 2025).

En consecuencia, recomendamos que futuros estudios aborden esta comparación con un mayor número de muestras y bajo diferentes escenarios experimentales, aumentando así la robustez y generalidad de la evaluación. Una ampliación del muestreo, idealmente complementada con réplicas técnicas y biológicas adicionales, brindaría mayor poder para confirmar (o refutar) la intercambiabilidad observada. Asimismo, la inclusión de ajustes técnicos adicionales será valiosa: por ejemplo, la aplicación de algoritmos de llamada de variantes específicos para ONT, calibración cruzada de resultados entre plataformas, y la consideración de umbrales de heteroplasmia más bajos, todo ello permitirá determinar con mayor precisión bajo qué condiciones ambas tecnologías ofrecen resultados superponibles.

4. CONSIDERACIONES FINALES

4.1. Conclusiones

Por medio del desarrollo de este trabajo de grado, se pudieron establecer las mejores herramientas, dentro de un grupo establecido para el análisis de ADNmt secuenciado por la plataforma de ONT, teniendo en cuenta las diferentes métricas de calidad a tener en cuenta a la hora de analizar este tipo de molécula. Se establecieron entonces la combinación de Clair3 y minimap2 como *caller* y *mapper*, que mostraron un mejor rendimiento tanto computacional como bioinformáticamente.

A partir de la construcción de estos vectores de análisis, se pudieron hacer comparaciones entre las dos plataformas que si bien difieren en la forma de capturar la información, deberían tener un acercamiento concordante a lo biológico. Lo que se puede concluir es que si bien ONT con las mejoras que ha tenido en la química y en los paquetes de análisis, puede generar suficiente información para hacer un análisis del ADNmt, aun presenta algunas limitaciones en la detección, sobre todo en la parte de detectar niveles por debajo del 10% de heteroplasmia, en comparación con los datos que se generan en Illumina. Sin embargo para el caso de este estudio, hay una alta concordancia entre las plataformas que con futuros proyectos podría llevar a considerar el uso de la plataforma ONT como parte de la rutina de apoyo al diagnóstico clínico.

Las diversas diferencias encontradas pueden estar explicadas por las limitaciones técnicas del trabajo previo al análisis, la calidad del material genético usado, las técnicas utilizadas para la construcción de las librerías y a problemas propios de la secuenciación, que quedan como una recomendación para trabajos posteriores.

4.2. Recomendaciones: consideraciones Clínicas sobre la Aplicabilidad de ONT como Alternativa a Illumina

Una de las motivaciones de este estudio era evaluar si la tecnología de secuenciación de Oxford Nanopore (ONT) podría servir como una alternativa o complemento a la de Illumina en aplicaciones clínicas enfocadas en ADN mitocondrial. Los resultados obtenidos muestran un panorama prometedor, pero es crucial contextualizarlos con evidencia publicada y mantener una visión conservadora basada en datos al trasladar estos hallazgos al entorno clínico.

ONT ofrece varias ventajas únicas que podrían ser explotadas en diagnósticos de enfermedades mitocondriales. Primero, la capacidad de generar lecturas de longitud completa del genoma mitocondrial (≈ 16 kb) en una sola molécula permite detectar alteraciones estructurales que las lecturas cortas no pueden resolver fácilmente. Esto incluye la identificación de grandes deleciones, duplicaciones o reordenamientos en el ADNmt, los cuales son causa de patologías como el síndrome de Kearns-Sayre y otras deleciones esporádicas. Estudios recientes han demostrado que la secuenciación de lecturas largas puede mapear deleciones mitocondriales sin necesidad de PCR específicas, superando la limitación de Illumina para estas alteraciones (González et al., 2020; Lüth et al., 2022a).

Frascarelli et al., (2023) reportaron que la tecnología ONT, gracias a sus lecturas kilobase, presentó fuertes ventajas en el análisis de alteraciones del ADNmt (especialmente reordenamientos complejos) en comparación con NGS de lectura corta, e incluso plantean que podría convertirse en el método de elección para estudios genéticos de ADNmt orientados a detectar deleciones. Además, ONT permite secuenciar ADN nativo sin amplificación PCR, lo cual evita sesgos de amplificación y, notablemente, permite detectar modificaciones epigenéticas (p. ej., metilación de ADNmt) directamente a partir de la señal eléctrica de la lectura (Frascarelli et al., 2023; Y. Wang et al., 2021; P. Zheng et al., 2023). Aunque la relevancia clínica de la metilación en ADNmt está aún bajo investigación, esta capacidad añade un valor potencial para investigación traslacional.

Otro punto a favor es la rapidez y portabilidad se encuentra en los dispositivos como MinION o incluso Flongle permiten secuenciar en tiempo real y podrían, por ejemplo, ser usados directamente en laboratorios clínicos pequeños o en entornos de atención remota. Un flujo MinION cuesta relativamente poco y puede secuenciar un genoma mitocondrial completo con suficiente profundidad en cuestión de horas (Lu et al., 2016; Vereecke et al., 2020). Akamatsu et al., (2024) demostraron un protocolo simplificado con Flongle para secuenciar ADNmt de pacientes con enfermedades mitocondriales, logrando identificar variantes patogénicas conocidas con precisión y cuantificación de heteroplasmia altamente concordante ($r^2 = 0.99$).

Este estudio indicó que Flongle/ONT puede agilizar y abaratar significativamente el proceso de secuenciación de ADNmt para diagnóstico, ofreciendo un enfoque accesible incluso fuera de grandes centros (por ejemplo, para tamizajes iniciales en hospitales con menos infraestructura). La portabilidad de ONT ha llevado a casos de uso exitosos en diagnóstico rápido de patógenos; en el contexto de ADNmt, podría traducirse en *turnaround times* más cortos para confirmar una mutación patogénica en un paciente, lo cual es valioso en toma de decisiones clínicas (p. ej., considerar terapia, consejo genético, etc.) (Akamatsu et al., 2024; Vereecke et al., 2020).

Dentro de los resultados obtenidos en este estudio, se puede establecer que hay una buena concordancia entre ambas plataformas, que puede llegar a ser una gran ayuda para el ambiente clínico, donde acercamiento a las plataformas de secuenciación de Oxford Nanopore, podrían ofrecer a los pacientes, una reducción en costos y tiempo de entrega de resultados.

Sin embargo, también presenta algunas limitaciones metodológicas que deben ser consideradas al interpretar los resultados. En primer lugar, el tamaño muestral fue reducido (cinco muestras biológicas), lo cual limita el poder estadístico para detectar diferencias sutiles entre plataformas de secuenciación, especialmente en escenarios de heteroplasmia baja. Si bien la inclusión de triplicados técnicos por muestra permitió evaluar la reproducibilidad intra-plataforma para Oxford Nanopore, un mayor número de muestras y réplicas biológicas permitiría fortalecer la generalización de los hallazgos.

En segundo lugar, el análisis bioinformático dependió de versiones específicas de herramientas y modelos disponibles al momento del estudio (por ejemplo, Dorado, Clair3, Mutect2, Mutserve), las cuales se encuentran en constante evolución. Dado el rápido avance de los algoritmos de basecalling y llamado de variantes para lecturas largas, es posible que versiones futuras modifiquen el desempeño observado, particularmente en la detección de heteroplasmas de baja frecuencia.

Finalmente, los datos de secuenciación Illumina utilizados como referencia corresponden a un procesamiento retrospectivo, generado bajo condiciones experimentales y bioinformáticas distintas a las empleadas para Oxford Nanopore. En este contexto, Illumina fue utilizada como una referencia comparativa operacional y no como un estándar absoluto de verdad. Por tanto, parte de las diferencias observadas entre plataformas puede atribuirse a diferencias acumulativas en los flujos de trabajo y no exclusivamente a limitaciones intrínsecas de la tecnología Nanopore.

Por otra parte, a pesar de los avances, la adopción de ONT en entornos clínicos debe abordarse con cautela y basarse en evidencia sólida. Un factor crítico es la tasa de error base todavía mayor de ONT en comparación con Illumina. Si bien a nivel de consenso se obtienen secuencias casi perfectas, la llamada de variantes heteroplásmicas de baja frecuencia sigue siendo propensa a errores y falsos positivos si no se aplican filtros adecuados (Delahaye & Nicolas, 2021). Varios estudios han señalado que, sin procesos de validación exhaustivos, ONT por sí sola puede reportar variantes aleatorias, especialmente debajo del ~5-10% de heteroplasma (Diroma et al., 2021; Slapnik et al., 2024; Tüns et al., 2022).

Por tanto, en un contexto clínico, ONT requeriría protocolos de validación rigurosos para cada variante reportada. Slapnik et al., (2024) concluyen que, aunque la secuenciación Nanopore muestra potencial para diagnósticos de enfermedades mitocondriales, se requieren procesos de validación estrictos para asegurar la calidad de los resultados antes de su uso diagnóstico. Esto podría implicar, por ejemplo, confirmar cualquier variante patogénica detectada por ONT mediante un método ortogonal (Illumina profundo o incluso secuenciación Sanger dirigida) hasta que ONT demuestre en diversos estudios su confiabilidad plena en ese ámbito (Kim et al., 2018).

Otra limitación es la sensibilidad a variantes de muy baja heteroplasmia, que se vuelve esencial en algunas enfermedades mitocondriales (o en portadores maternos asintomáticos), donde detectar variantes en proporciones del 1-5% puede ser relevante para evaluar riesgo de transmisión o mosaicos iniciales (Kim et al., 2018). Illumina, con *ultradeep sequencing*, ha conseguido detectar heteroplasmas ~1%, mientras que ONT en la práctica actual tendría dificultades en ese rango a menos que se incremente muchísimo la profundidad y se utilicen algoritmos especializados de supresión de error (p. ej., UMI barcoding o consorcio de múltiples corridas). Por lo tanto, en su estado actual, ONT podría pasar por alto algunas variantes de muy baja frecuencia que Illumina sí captaría, lo cual puede ser crítico en contextos como detección precoz en tejidos mixtos o en muestras de heteroplasma incipiente (Kim et al., 2018; Slapnik et al., 2024).

Por otra parte, desde una perspectiva regulatoria, la implementación clínica de flujos de trabajo basados en Oxford Nanopore requeriría su alineación con marcos normativos vigentes como ISO 15189 o los lineamientos CLIA para laboratorios clínicos. Esto implica la validación formal de parámetros como precisión analítica, sensibilidad, reproducibilidad inter-corrida y límites de detección de heteroplasma, así como la incorporación de controles internos y externos de calidad. En este sentido, el método propuesto en este estudio puede considerarse un punto de partida para procesos de validación clínica, particularmente en escenarios donde ONT se emplee como herramienta complementaria para la detección de variantes estructurales, análisis haplotípico completo o diagnóstico rápido, manteniendo la confirmación de variantes de baja frecuencia mediante métodos ortogonales cuando sea necesario.

Tomando en conjunto los hallazgos de este estudio y la literatura, podemos afirmar con cautela que la plataforma ONT es una candidata técnicamente viable para el análisis clínico del ADNmt, con ciertas ventajas únicas, pero no reemplaza aun completamente a Illumina en precisión para todos los aspectos. Nuestro análisis demostraron concordancia elevada en llamadas de variantes, lo que es alentador; sin embargo, para una adopción clínica, se requeriría estandarizar pipelines de Nanopore, incorporar controles de calidad específicos (p. ej., variantes de control conocidas a distintas heteroplasmas) y, muy probablemente, implementar un enfoque híbrido donde ONT provea la información de haplotipos completos y variantes estructurales, mientras Illumina u otros métodos aseguren la detección de variantes de muy baja frecuencia o sirvan de confirmación. Esta estrategia combinada ya

ha sido sugerida por algunos autores, ya que el uso conjunto de lecturas largas y cortas puede brindar mayor exactitud que cualquiera por separado.

En conclusión, ONT presenta un futuro prometedor en la genómica clínica mitocondrial, desde la identificación de mutaciones puntuales hasta la caracterización de deleciones complejas. No obstante, dado que la precisión para variantes de baja frecuencia y la estandarización aún no igualan a la tecnología Illumina, es prudente adoptar ONT de forma complementaria y bajo validación rigurosa. Con el continuo avance de la tecnología (mejores nanoporos, algoritmos de AI como Clair3 mejorando llamadas, mayor experiencia acumulada), es razonable anticipar que ONT se consolide cada vez más como una alternativa real en el diagnóstico molecular de enfermedades mitocondriales.

Por ahora, la recomendación basada en evidencia es aprovechar las fortalezas de ONT (detección de variantes estructurales, secuenciación rápida in situ, análisis haplotípico) en combinación con la confiabilidad de Illumina para lograr un diagnóstico integral y confiable, tal como han sugerido recientes trabajos de enfoque clínico, o justificar por otros alcances el uso exclusivo de ONT. Sin embargo, el enfoque equilibrado asegurará que los pacientes se beneficien de la tecnología de lecturas largas sin comprometer la exactitud diagnóstica, alineando la innovación tecnológica con la seguridad y eficacia requeridas en el entorno clínico.

A. Anexo 1: Tabla con datos comparativos de las variantes encontradas en cada plataforma.

Sample	Aligner	Caller	Platf	POS	REF	ALT	AF	DP	QUAL
3258-1	minimap2	clair3	ONT	73	A	G	0.9213	254	42.25
3258-1	minimap2	clair3	ONT	263	A	G	0.988	249	110
3258-1	minimap2	clair3	ONT	310	T	TC	0.5696	237	34.01
3258-1	minimap2	clair3	ONT	493	AC	A	0.9637	193	55.71
3258-1	minimap2	clair3	ONT	499	G	A	0.9792	192	48.03
3258-1	minimap2	clair3	ONT	750	A	G	0.9737	190	110
3258-1	minimap2	clair3	ONT	827	A	G	0.9842	190	41.05
3258-1	minimap2	clair3	ONT	1438	A	G	0.9709	172	39.27
3258-1	minimap2	clair3	ONT	2706	A	G	0.9825	171	110
3258-1	minimap2	clair3	ONT	3106	CN	C	0.9357	171	68.1
3258-1	minimap2	clair3	ONT	3547	A	G	0.9706	170	42.02
3258-1	minimap2	clair3	ONT	4122	A	G	0.9524	168	40.29
3258-1	minimap2	clair3	ONT	4123	A	G	0.9464	168	38.3
3258-1	minimap2	clair3	ONT	4769	A	G	0.9458	166	110
3258-1	minimap2	clair3	ONT	4820	G	A	0.9337	166	76.22
3258-1	minimap2	clair3	ONT	4977	T	C	0.9398	166	110
3258-1	minimap2	clair3	ONT	5460	G	A	0.9509	163	79.23
3258-1	minimap2	clair3	ONT	6473	C	T	0.9207	164	41.04
3258-1	minimap2	clair3	ONT	7028	C	T	0.9448	163	110

3258-1	minimap2	clair3	ONT	8270	CACC CCCT CT	C	0.8415	164	54.61
3258-1	minimap2	clair3	ONT	8860	A	G	0.8957	163	110
3258-1	minimap2	clair3	ONT	8875	T	C	0.9448	163	110
3258-1	minimap2	clair3	ONT	9682	T	C	0.9571	163	110
3258-1	minimap2	clair3	ONT	9950	T	C	0.9264	163	110
3258-1	minimap2	clair3	ONT	11177	C	T	0.9506	162	110
3258-1	minimap2	clair3	ONT	11335	C	T	0.9448	163	38.63
3258-1	minimap2	clair3	ONT	11719	G	A	0.939	164	40.57
3258-1	minimap2	clair3	ONT	13590	G	A	0.9521	167	43.76
3258-1	minimap2	clair3	ONT	14766	C	T	1	170	110
3258-1	minimap2	clair3	ONT	15326	A	G	0.9942	173	37.77
3258-1	minimap2	clair3	ONT	15535	C	T	0.948	173	41.16
3258-1	minimap2	clair3	ONT	16189	T	C	0.9042	240	110
3258-1	minimap2	clair3	ONT	16217	T	C	0.9917	240	110
3258-1	minimap2	clair3	ONT	16519	T	C	1	27	48.23
3258-2	minimap2	clair3	ONT	73	A	G	0.9333	195	43.79
3258-2	minimap2	clair3	ONT	263	A	G	0.9948	191	110
3258-2	minimap2	clair3	ONT	493	AC	A	0.943	158	48.22
3258-2	minimap2	clair3	ONT	499	G	A	0.9494	158	50.83
3258-2	minimap2	clair3	ONT	750	A	G	0.9936	157	39.26
3258-2	minimap2	clair3	ONT	827	A	G	0.9873	157	39.16
3258-2	minimap2	clair3	ONT	1438	A	G	0.9929	140	40.08
3258-2	minimap2	clair3	ONT	2706	A	G	0.9779	136	110
3258-2	minimap2	clair3	ONT	3106	CN	C	0.9111	135	76.22
3258-2	minimap2	clair3	ONT	3547	A	G	0.9924	132	40.37
3258-2	minimap2	clair3	ONT	4122	A	G	0.9847	131	38.87
3258-2	minimap2	clair3	ONT	4123	A	G	0.9924	131	41.53
3258-2	minimap2	clair3	ONT	4769	A	G	1	130	110
3258-2	minimap2	clair3	ONT	4820	G	A	0.9846	130	40.53
3258-2	minimap2	clair3	ONT	4977	T	C	0.9692	130	110
3258-2	minimap2	clair3	ONT	5460	G	A	0.9923	130	73.22

3258-2	minimap2	clair3	ONT	6473	C	T	1	129	110
3258-2	minimap2	clair3	ONT	7028	C	T	0.9764	127	110
3258-2	minimap2	clair3	ONT	8270	CACC CCCT CT	C	0.9032	124	59.46
3258-2	minimap2	clair3	ONT	8860	A	G	0.9597	124	110
3258-2	minimap2	clair3	ONT	8875	T	C	0.9839	124	39.83
3258-2	minimap2	clair3	ONT	9682	T	C	0.9756	123	110
3258-2	minimap2	clair3	ONT	9950	T	C	0.9839	124	110
3258-2	minimap2	clair3	ONT	11177	C	T	0.992	125	110
3258-2	minimap2	clair3	ONT	11335	C	T	1	125	110
3258-2	minimap2	clair3	ONT	11719	G	A	0.984	125	40.93
3258-2	minimap2	clair3	ONT	13590	G	A	0.968	125	40.12
3258-2	minimap2	clair3	ONT	14766	C	T	0.9846	130	110
3258-2	minimap2	clair3	ONT	15326	A	G	0.9852	135	110
3258-2	minimap2	clair3	ONT	15535	C	T	0.9778	135	40.94
3258-2	minimap2	clair3	ONT	16183	A	C	0.7062	177	28.45
3258-2	minimap2	clair3	ONT	16189	T	C	0.882	178	110
3258-2	minimap2	clair3	ONT	16217	T	C	0.9888	178	110
3258-2	minimap2	clair3	ONT	16519	T	C	1	20	37
3258-3	minimap2	clair3	ONT	73	A	G	0.9295	156	43.73
3258-3	minimap2	clair3	ONT	263	A	G	0.9801	151	110
3258-3	minimap2	clair3	ONT	493	AC	A	0.937	127	56.91
3258-3	minimap2	clair3	ONT	499	G	A	0.9606	127	47.09
3258-3	minimap2	clair3	ONT	750	A	G	0.968	125	110
3258-3	minimap2	clair3	ONT	827	A	G	1	125	39.33
3258-3	minimap2	clair3	ONT	1438	A	G	0.9823	113	42.88
3258-3	minimap2	clair3	ONT	2706	A	G	0.964	111	110
3258-3	minimap2	clair3	ONT	3106	CN	C	0.8649	111	61.99
3258-3	minimap2	clair3	ONT	3547	A	G	0.9909	110	41.24
3258-3	minimap2	clair3	ONT	4122	A	G	1	110	79.23
3258-3	minimap2	clair3	ONT	4123	A	G	0.9909	110	38.89
3258-3	minimap2	clair3	ONT	4769	A	G	1	110	110

3258-3	minimap2	clair3	ONT	4820	G	A	0.9818	110	42.28
3258-3	minimap2	clair3	ONT	4977	T	C	1	110	110
3258-3	minimap2	clair3	ONT	5460	G	A	0.9633	109	110
3258-3	minimap2	clair3	ONT	6473	C	T	0.9722	108	40.43
3258-3	minimap2	clair3	ONT	7028	C	T	0.9722	108	110
3258-3	minimap2	clair3	ONT	8270	CACC CCCT CT	C	0.8532	109	55.02
3258-3	minimap2	clair3	ONT	8860	A	G	0.9725	109	39.97
3258-3	minimap2	clair3	ONT	8875	T	C	0.9817	109	110
3258-3	minimap2	clair3	ONT	9682	T	C	0.9908	109	110
3258-3	minimap2	clair3	ONT	9950	T	C	0.9633	109	110
3258-3	minimap2	clair3	ONT	11177	C	T	0.9909	110	110
3258-3	minimap2	clair3	ONT	11335	C	T	0.9727	110	110
3258-3	minimap2	clair3	ONT	11719	G	A	0.9821	112	41.89
3258-3	minimap2	clair3	ONT	13590	G	A	0.9727	110	38.96
3258-3	minimap2	clair3	ONT	14766	C	T	0.9911	112	110
3258-3	minimap2	clair3	ONT	15326	A	G	0.9739	115	39.16
3258-3	minimap2	clair3	ONT	15535	C	T	0.9741	116	41.77
3258-3	minimap2	clair3	ONT	16189	T	C	0.8909	165	110
3258-3	minimap2	clair3	ONT	16217	T	C	1	166	110
3258-3	minimap2	clair3	ONT	16519	T	C	0.9706	34	51.72
3312-1	minimap2	clair3	ONT	64	C	T	0.8921	964	37.79
3312-1	minimap2	clair3	ONT	73	A	G	0.9554	964	44.16
3312-1	minimap2	clair3	ONT	146	T	C	0.9315	964	110
3312-1	minimap2	clair3	ONT	153	A	G	0.9751	963	39.38
3312-1	minimap2	clair3	ONT	235	A	G	0.9883	944	110
3312-1	minimap2	clair3	ONT	263	A	G	0.9852	943	37.82
3312-1	minimap2	clair3	ONT	310	T	TC	0.6847	926	24.05
3312-1	minimap2	clair3	ONT	513	GCA	G	0.9478	804	110
3312-1	minimap2	clair3	ONT	663	A	G	0.9849	796	38.08
3312-1	minimap2	clair3	ONT	750	A	G	0.9811	792	37.83
3312-1	minimap2	clair3	ONT	1438	A	G	0.9715	772	39.3

3312-1	minimap2	clair3	ONT	1736	A	G	0.9854	753	110
3312-1	minimap2	clair3	ONT	2706	A	G	0.9713	732	110
3312-1	minimap2	clair3	ONT	3106	CN	C	0.9121	728	65.62
3312-1	minimap2	clair3	ONT	3516	C	T	0.9528	721	110
3312-1	minimap2	clair3	ONT	4248	T	C	0.8536	717	79.23
3312-1	minimap2	clair3	ONT	4769	A	G	0.9944	713	110
3312-1	minimap2	clair3	ONT	4824	A	G	0.9074	713	40.33
3312-1	minimap2	clair3	ONT	7028	C	T	0.9735	716	110
3312-1	minimap2	clair3	ONT	8027	G	A	0.948	712	38.03
3312-1	minimap2	clair3	ONT	8794	C	T	0.9776	713	110
3312-1	minimap2	clair3	ONT	8860	A	G	0.9565	712	110
3312-1	minimap2	clair3	ONT	11152	T	C	0.9833	719	110
3312-1	minimap2	clair3	ONT	11719	G	A	0.982	722	40.1
3312-1	minimap2	clair3	ONT	12007	G	A	0.9834	723	37.87
3312-1	minimap2	clair3	ONT	12705	C	T	0.9795	732	110
3312-1	minimap2	clair3	ONT	14766	C	T	0.9961	776	110
3312-1	minimap2	clair3	ONT	15326	A	G	0.9925	798	37.44
3312-1	minimap2	clair3	ONT	16111	C	T	0.9863	879	110
3312-1	minimap2	clair3	ONT	16126	T	C	0.9921	887	110
3312-1	minimap2	clair3	ONT	16223	C	T	0.9778	901	110
3312-1	minimap2	clair3	ONT	16290	C	T	0.9802	910	110
3312-1	minimap2	clair3	ONT	16311	T	C	0.9956	912	79.23
3312-1	minimap2	clair3	ONT	16319	G	A	0.9912	913	58.48
3312-1	minimap2	clair3	ONT	16362	T	C	0.9465	916	110
3312-2	minimap2	clair3	ONT	64	C	T	0.8487	985	56.1
3312-2	minimap2	clair3	ONT	73	A	G	0.9624	985	42.62
3312-2	minimap2	clair3	ONT	146	T	C	0.928	986	110
3312-2	minimap2	clair3	ONT	153	A	G	0.9745	982	39.13
3312-2	minimap2	clair3	ONT	235	A	G	0.9916	958	110
3312-2	minimap2	clair3	ONT	263	A	G	0.9874 000000 000000 6	954	37.77

3312-2	minimap2	clair3	ONT	310	T	TC	0.6957	940	38.31
3312-2	minimap2	clair3	ONT	513	GCA	G	0.9394	841	110
3312-2	minimap2	clair3	ONT	663	A	G	0.9856	833	38.11
3312-2	minimap2	clair3	ONT	750	A	G	0.9794	824	37.87
3312-2	minimap2	clair3	ONT	1438	A	G	0.9717	814	39.87
3312-2	minimap2	clair3	ONT	1736	A	G	0.9913	801	37.58
3312-2	minimap2	clair3	ONT	2706	A	G	0.9821	780	110
3312-2	minimap2	clair3	ONT	3106	CN	C	0.9142	781	59.19
3312-2	minimap2	clair3	ONT	3516	C	T	0.9704	777	110
3312-2	minimap2	clair3	ONT	4248	T	C	0.8601	772	110
3312-2	minimap2	clair3	ONT	4769	A	G	0.9948	769	110
3312-2	minimap2	clair3	ONT	4824	A	G	0.9049	768	40.25
3312-2	minimap2	clair3	ONT	7028	C	T	0.9817	766	110
3312-2	minimap2	clair3	ONT	8027	G	A	0.9569	766	63.8
3312-2	minimap2	clair3	ONT	8794	C	T	0.9817	766	110
3312-2	minimap2	clair3	ONT	8860	A	G	0.9582	766	36.94
3312-2	minimap2	clair3	ONT	11152	T	C	0.9844	770	110
3312-2	minimap2	clair3	ONT	11719	G	A	0.9831	771	40.09
3312-2	minimap2	clair3	ONT	12007	G	A	0.9806	773	38.31
3312-2	minimap2	clair3	ONT	12705	C	T	0.9884	775	110
3312-2	minimap2	clair3	ONT	14766	C	T	0.9902	820	110
3312-2	minimap2	clair3	ONT	15326	A	G	0.9905	846	37.3
3312-2	minimap2	clair3	ONT	16111	C	T	0.9904	940	110
3312-2	minimap2	clair3	ONT	16126	T	C	0.9831	944	110
3312-2	minimap2	clair3	ONT	16223	C	T	0.9886	963	110
3312-2	minimap2	clair3	ONT	16290	C	T	0.9767	986	110
3312-2	minimap2	clair3	ONT	16311	T	C	0.999	989	110
3312-2	minimap2	clair3	ONT	16319	G	A	0.9869	990	63
3312-2	minimap2	clair3	ONT	16362	T	C	0.9405	992	110
3312-3	minimap2	clair3	ONT	64	C	T	0.8665	1004	63.67
3312-3	minimap2	clair3	ONT	73	A	G	0.9711	1004	43.12
3312-3	minimap2	clair3	ONT	146	T	C	0.9186	1007	110
3312-3	minimap2	clair3	ONT	153	A	G	0.9751	1005	40.22

3312-3	minimap2	clair3	ONT	235	A	G	0.9879	990	110
3312-3	minimap2	clair3	ONT	263	A	G	0.9838	988	37.67
3312-3	minimap2	clair3	ONT	310	T	TC	0.6912	978	39.02
3312-3	minimap2	clair3	ONT	513	GCA	G	0.9345	855	110
3312-3	minimap2	clair3	ONT	663	A	G	0.9918	849	38.03
3312-3	minimap2	clair3	ONT	750	A	G	0.9752	846	39.2
3312-3	minimap2	clair3	ONT	1438	A	G	0.9719	819	39.34
3312-3	minimap2	clair3	ONT	1736	A	G	0.9926	808	110
3312-3	minimap2	clair3	ONT	2706	A	G	0.9824	794	110
3312-3	minimap2	clair3	ONT	3106	CN	C	0.8961	789	66.93
3312-3	minimap2	clair3	ONT	3516	C	T	0.9772	788	110
3312-3	minimap2	clair3	ONT	4248	T	C	0.8577	787	79.23
3312-3	minimap2	clair3	ONT	4769	A	G	0.9937	789	110
3312-3	minimap2	clair3	ONT	4824	A	G	0.9025	790	39.73
3312-3	minimap2	clair3	ONT	7028	C	T	0.9824	797	110
3312-3	minimap2	clair3	ONT	8027	G	A	0.947	792	37.25
3312-3	minimap2	clair3	ONT	8794	C	T	0.9698	795	110
3312-3	minimap2	clair3	ONT	8860	A	G	0.9548	796	110
3312-3	minimap2	clair3	ONT	11152	T	C	0.9839	808	110
3312-3	minimap2	clair3	ONT	11719	G	A	0.984	810	39.94
3312-3	minimap2	clair3	ONT	12007	G	A	0.984	811	38.42
3312-3	minimap2	clair3	ONT	12705	C	T	0.9743	818	110
3312-3	minimap2	clair3	ONT	14766	C	T	0.993	853	110
3312-3	minimap2	clair3	ONT	15326	A	G	0.9795	880	38.27
3312-3	minimap2	clair3	ONT	16111	C	T	0.9928	969	110
3312-3	minimap2	clair3	ONT	16126	T	C	0.9897	974	110
3312-3	minimap2	clair3	ONT	16223	C	T	0.9858	987	110
3312-3	minimap2	clair3	ONT	16290	C	T	0.9653	1008	110
3312-3	minimap2	clair3	ONT	16311	T	C	0.998	1012	79.23
3312-3	minimap2	clair3	ONT	16319	G	A	0.9882	1013	37.22
3312-3	minimap2	clair3	ONT	16362	T	C	0.9331	1017	110
3396_1	minimap2	clair3	ONT	64	C	T	0.916	1024	110
3396_1	minimap2	clair3	ONT	73	A	G	0.9707	1024	43.76

3396_1	minimap2	clair3	ONT	146	T	C	0.9453	1024	110
3396_1	minimap2	clair3	ONT	153	A	G	0.9902	1024	38.71
3396_1	minimap2	clair3	ONT	174	C	T	0.9893	1024	110
3396_1	minimap2	clair3	ONT	235	A	G	0.9901	1008	110
3396_1	minimap2	clair3	ONT	263	A	G	0.986	1003	37.86
3396_1	minimap2	clair3	ONT	310	T	TC	0.7748	977	53.08
3396_1	minimap2	clair3	ONT	513	GCA	G	0.9667	931	110
3396_1	minimap2	clair3	ONT	663	A	G	0.9967	921	37.98
3396_1	minimap2	clair3	ONT	750	A	G	0.9859	919	39.03
3396_1	minimap2	clair3	ONT	1438	A	G	0.982	890	38.95
3396_1	minimap2	clair3	ONT	1736	A	G	0.9942	868	37.89
3396_1	minimap2	clair3	ONT	2706	A	G	0.9815	810	110
3396_1	minimap2	clair3	ONT	3106	CN	C	0.9402	803	79.23
3396_1	minimap2	clair3	ONT	4248	T	C	0.903	804	110
3396_1	minimap2	clair3	ONT	4769	A	G	0.9938	803	110
3396_1	minimap2	clair3	ONT	4824	A	G	0.9465	803	38
3396_1	minimap2	clair3	ONT	7028	C	T	0.9874 000000 000000 6	793	110
3396_1	minimap2	clair3	ONT	8027	G	A	0.9738	800	37.46
3396_1	minimap2	clair3	ONT	8056	A	G	0.9913	801	110
3396_1	minimap2	clair3	ONT	8794	C	T	0.985	802	110
3396_1	minimap2	clair3	ONT	8860	A	G	0.9738	803	36.75
3396_1	minimap2	clair3	ONT	11719	G	A	0.982	834	39.19
3396_1	minimap2	clair3	ONT	12007	G	A	0.994	835	38.01
3396_1	minimap2	clair3	ONT	12705	C	T	0.992	873	110
3396_1	minimap2	clair3	ONT	14766	C	T	0.9959	971	110
3396_1	minimap2	clair3	ONT	15326	A	G	0.9889	994	110
3396_1	minimap2	clair3	ONT	16111	C	T	0.9942	1206	110
3396_1	minimap2	clair3	ONT	16129	G	A	0.9884	1209	76.22
3396_1	minimap2	clair3	ONT	16223	C	T	0.9918	1213	110
3396_1	minimap2	clair3	ONT	16239	C	T	0.9942	1213	110

3396_1	minimap2	clair3	ONT	16290	C	T	0.9802	1213	110
3396_1	minimap2	clair3	ONT	16319	G	A	0.9942	1213	63.67
3396_1	minimap2	clair3	ONT	16362	T	C	0.9571	1213	110
3396_2	minimap2	clair3	ONT	64	C	T	0.8992	2708	79.23
3396_2	minimap2	clair3	ONT	73	A	G	0.9778	2708	43.43
3396_2	minimap2	clair3	ONT	146	T	C	0.955	2709	110
3396_2	minimap2	clair3	ONT	153	A	G	0.986	2709	38.36
3396_2	minimap2	clair3	ONT	174	C	T	0.9941	2710	110
3396_2	minimap2	clair3	ONT	235	A	G	0.9952	2683	110
3396_2	minimap2	clair3	ONT	263	A	G	0.988	2668	38.01
3396_2	minimap2	clair3	ONT	302	A	AC	0.7167	2626	25.91
3396_2	minimap2	clair3	ONT	310	T	TC	0.7583	2615	42.84
3396_2	minimap2	clair3	ONT	513	GCA	G	0.9646	2483	110
3396_2	minimap2	clair3	ONT	663	A	G	0.9927	2461	37.96
3396_2	minimap2	clair3	ONT	750	A	G	0.9852	2360	39.18
3396_2	minimap2	clair3	ONT	1438	A	G	0.9736	2232	38.63
3396_2	minimap2	clair3	ONT	1736	A	G	0.9927	2187	38.05
3396_2	minimap2	clair3	ONT	2706	A	G	0.9854	1984	110
3396_2	minimap2	clair3	ONT	3106	CN	C	0.9482	1950	72.25
3396_2	minimap2	clair3	ONT	4248	T	C	0.8854	1859	110
3396_2	minimap2	clair3	ONT	4769	A	G	0.9843	1843	110
3396_2	minimap2	clair3	ONT	4824	A	G	0.9343	1842	38.95
3396_2	minimap2	clair3	ONT	7028	C	T	0.9741	1817	110
3396_2	minimap2	clair3	ONT	8027	G	A	0.9664	1813	37.21
3396_2	minimap2	clair3	ONT	8056	A	G	0.9785	1813	110
3396_2	minimap2	clair3	ONT	8794	C	T	0.9708	1813	110
3396_2	minimap2	clair3	ONT	8860	A	G	0.957	1814	37.32
3396_2	minimap2	clair3	ONT	11719	G	A	0.9772	1846	40.47
3396_2	minimap2	clair3	ONT	12007	G	A	0.9751	1848	38.92
3396_2	minimap2	clair3	ONT	12705	C	T	0.9761	1881	110
3396_2	minimap2	clair3	ONT	14766	C	T	0.9946	2030	110
3396_2	minimap2	clair3	ONT	15326	A	G	0.9924	2097	110
3396_2	minimap2	clair3	ONT	16111	C	T	0.995	2390	110

3396_2	minimap2	clair3	ONT	16129	G	A	0.9896	2394	110
3396_2	minimap2	clair3	ONT	16223	C	T	0.9922	5138	110
3396_2	minimap2	clair3	ONT	16239	C	T	0.9938	5140	110
3396_2	minimap2	clair3	ONT	16290	C	T	0.9848	5140	110
3396_2	minimap2	clair3	ONT	16319	G	A	0.9924	5139	62.7
3396_2	minimap2	clair3	ONT	16362	T	C	0.9571	5129	110
3396_3	minimap2	clair3	ONT	64	C	T	0.9213	1524	110
3396_3	minimap2	clair3	ONT	73	A	G	0.9797	1524	43.16
3396_3	minimap2	clair3	ONT	146	T	C	0.958	1525	110
3396_3	minimap2	clair3	ONT	153	A	G	0.9895	1529	38.64
3396_3	minimap2	clair3	ONT	174	C	T	0.9908	1527	110
3396_3	minimap2	clair3	ONT	235	A	G	0.9977	1325	110
3396_3	minimap2	clair3	ONT	263	A	G	0.9909	1315	38.24
3396_3	minimap2	clair3	ONT	310	T	TC	0.7518	1281	57.84
3396_3	minimap2	clair3	ONT	513	GCA	G	0.9584	1202	110
3396_3	minimap2	clair3	ONT	663	A	G	0.9907	1179	38.16
3396_3	minimap2	clair3	ONT	750	A	G	0.9872	1170	39.07
3396_3	minimap2	clair3	ONT	1438	A	G	0.9819	1102	39.24
3396_3	minimap2	clair3	ONT	1736	A	G	0.9944	1074	37.88
3396_3	minimap2	clair3	ONT	2706	A	G	0.9819	996	110
3396_3	minimap2	clair3	ONT	3106	CN	C	0.942	966	74.46
3396_3	minimap2	clair3	ONT	4248	T	C	0.8956	929	110
3396_3	minimap2	clair3	ONT	4769	A	G	0.9903	929	110
3396_3	minimap2	clair3	ONT	4824	A	G	0.9256	927	40.02
3396_3	minimap2	clair3	ONT	7028	C	T	0.9808	936	110
3396_3	minimap2	clair3	ONT	8027	G	A	0.9563	938	38.24
3396_3	minimap2	clair3	ONT	8056	A	G	0.9872	938	110
3396_3	minimap2	clair3	ONT	8794	C	T	0.9809	943	110
3396_3	minimap2	clair3	ONT	8860	A	G	0.963	945	37.23
3396_3	minimap2	clair3	ONT	11719	G	A	0.9858	983	40.01
3396_3	minimap2	clair3	ONT	12007	G	A	0.9817	984	38.47
3396_3	minimap2	clair3	ONT	12705	C	T	0.9801	1005	110
3396_3	minimap2	clair3	ONT	14766	C	T	0.9956	1137	110

3396_3	minimap2	clair3	ONT	15326	A	G	0.9942	1200	110
3396_3	minimap2	clair3	ONT	16111	C	T	0.9946	1483	110
3396_3	minimap2	clair3	ONT	16129	G	A	0.9845	1486	110
3396_3	minimap2	clair3	ONT	16223	C	T	0.9899	1691	110
3396_3	minimap2	clair3	ONT	16239	C	T	0.9917	1691	110
3396_3	minimap2	clair3	ONT	16290	C	T	0.9752	1694	69.69
3396_3	minimap2	clair3	ONT	16319	G	A	0.9897	4156	72.25
3396_3	minimap2	clair3	ONT	16362	T	C	0.9679	4083	110
9093-1	minimap2	clair3	ONT	73	A	G	0.9193	223	42.17
9093-1	minimap2	clair3	ONT	185	G	A	0.973	222	37.66
9093-1	minimap2	clair3	ONT	247	GA	G	0.9819	221	76.22
9093-1	minimap2	clair3	ONT	263	A	G	0.9864	221	39.88
9093-1	minimap2	clair3	ONT	285	CAA	C	0.9638	221	110
9093-1	minimap2	clair3	ONT	310	T	TC	0.6878	221	43.11
9093-1	minimap2	clair3	ONT	489	T	C	0.9947	190	110
9093-1	minimap2	clair3	ONT	750	A	G	0.9835	182	40.86
9093-1	minimap2	clair3	ONT	1438	A	G	0.9771	175	38.39
9093-1	minimap2	clair3	ONT	1888	G	A	0.9713	174	37.93
9093-1	minimap2	clair3	ONT	2706	A	G	0.9825	171	110
9093-1	minimap2	clair3	ONT	3106	CN	C	0.9118	170	66.93
9093-1	minimap2	clair3	ONT	3158	A	AT	0.9824	170	110
9093-1	minimap2	clair3	ONT	3552	T	A	1	167	110
9093-1	minimap2	clair3	ONT	4654	C	T	0.9639	166	110
9093-1	minimap2	clair3	ONT	4715	A	G	0.9578	166	110
9093-1	minimap2	clair3	ONT	4769	A	G	1	166	110
9093-1	minimap2	clair3	ONT	7028	C	T	0.9815	162	110
9093-1	minimap2	clair3	ONT	7196	C	A	0.9877	162	110
9093-1	minimap2	clair3	ONT	7253	A	G	0.9938	162	110
9093-1	minimap2	clair3	ONT	8477	T	C	0.9141	163	110
9093-1	minimap2	clair3	ONT	8584	G	A	0.9939	163	38.34
9093-1	minimap2	clair3	ONT	8701	A	G	0.9939	163	110
9093-1	minimap2	clair3	ONT	8860	A	G	0.9202	163	110
9093-1	minimap2	clair3	ONT	9540	T	C	0.9329	164	110

9093-1	minimap2	clair3	ONT	9545	A	G	0.865	163	38.32
9093-1	minimap2	clair3	ONT	10398	A	G	0.9451	164	38.32
9093-1	minimap2	clair3	ONT	10400	C	T	0.9634	164	39.68
9093-1	minimap2	clair3	ONT	10873	T	C	0.9818	165	110
9093-1	minimap2	clair3	ONT	11719	G	A	0.9939	164	41.93
9093-1	minimap2	clair3	ONT	11914	G	A	0.9695	164	42.58
9093-1	minimap2	clair3	ONT	12705	C	T	0.994	168	110
9093-1	minimap2	clair3	ONT	13263	A	G	0.9942	171	39.92
9093-1	minimap2	clair3	ONT	14318	T	C	0.9829	175	110
9093-1	minimap2	clair3	ONT	14766	C	T	0.989	182	110
9093-1	minimap2	clair3	ONT	14783	T	C	0.9725	182	110
9093-1	minimap2	clair3	ONT	15043	G	A	0.9835	182	110
9093-1	minimap2	clair3	ONT	15301	G	A	0.9783	184	68.44
9093-1	minimap2	clair3	ONT	15326	A	G	0.9891	184	38.3
9093-1	minimap2	clair3	ONT	15487	A	T	0.9838	185	110
9093-1	minimap2	clair3	ONT	15930	G	A	0.9896	192	42.34
9093-1	minimap2	clair3	ONT	16223	C	T	0.9822	225	110
9093-1	minimap2	clair3	ONT	16274	G	A	0.9803	254	110
9093-1	minimap2	clair3	ONT	16298	T	C	0.9766	256	110
9093-1	minimap2	clair3	ONT	16325	T	C	0.9934	452	110
9093-1	minimap2	clair3	ONT	16327	C	T	0.9956	452	110
9093-2	minimap2	clair3	ONT	73	A	G	0.9255	255	41.5
9093-2	minimap2	clair3	ONT	185	G	A	0.9647	255	69.69
9093-2	minimap2	clair3	ONT	247	GA	G	0.9843	254	76.22
9093-2	minimap2	clair3	ONT	263	A	G	0.9843	255	39.14
9093-2	minimap2	clair3	ONT	285	CAA	C	0.9803	254	110
9093-2	minimap2	clair3	ONT	310	T	TC	0.6508	252	38.4
9093-2	minimap2	clair3	ONT	489	T	C	0.9956	229	110
9093-2	minimap2	clair3	ONT	750	A	G	0.964	222	37.96
9093-2	minimap2	clair3	ONT	1438	A	G	0.9716	211	37.59
9093-2	minimap2	clair3	ONT	1888	G	A	0.9761	209	38.03
9093-2	minimap2	clair3	ONT	2706	A	G	0.9606	203	110
9093-2	minimap2	clair3	ONT	3106	CN	C	0.8955	201	61.46

9093-2	minimap2	clair3	ONT	3158	A	AT	0.9801	201	110
9093-2	minimap2	clair3	ONT	3552	T	A	0.995	200	110
9093-2	minimap2	clair3	ONT	4654	C	T	0.995	199	110
9093-2	minimap2	clair3	ONT	4715	A	G	0.9296	199	110
9093-2	minimap2	clair3	ONT	4769	A	G	1	199	110
9093-2	minimap2	clair3	ONT	7028	C	T	0.9897	194	110
9093-2	minimap2	clair3	ONT	7196	C	A	0.9793	193	38.81
9093-2	minimap2	clair3	ONT	7253	A	G	0.9948	193	110
9093-2	minimap2	clair3	ONT	8477	T	C	0.8866	194	110
9093-2	minimap2	clair3	ONT	8584	G	A	0.9795	195	40.53
9093-2	minimap2	clair3	ONT	8701	A	G	1	195	110
9093-2	minimap2	clair3	ONT	8860	A	G	0.9744	195	110
9093-2	minimap2	clair3	ONT	9540	T	C	0.9487	195	110
9093-2	minimap2	clair3	ONT	9545	A	G	0.9128	195	39.87
9093-2	minimap2	clair3	ONT	10398	A	G	0.9433	194	110
9093-2	minimap2	clair3	ONT	10400	C	T	0.9433	194	38.68
9093-2	minimap2	clair3	ONT	10873	T	C	0.9847	196	110
9093-2	minimap2	clair3	ONT	11719	G	A	0.9949	195	40.67
9093-2	minimap2	clair3	ONT	11914	G	A	0.9744	195	37.58
9093-2	minimap2	clair3	ONT	12705	C	T	0.9795	195	110
9093-2	minimap2	clair3	ONT	13263	A	G	0.9898	196	41.32
9093-2	minimap2	clair3	ONT	14318	T	C	0.9701	201	110
9093-2	minimap2	clair3	ONT	14766	C	T	0.9951	205	110
9093-2	minimap2	clair3	ONT	14783	T	C	0.9903	206	110
9093-2	minimap2	clair3	ONT	15043	G	A	0.9904	209	38.02
9093-2	minimap2	clair3	ONT	15301	G	A	0.9813	214	70.79
9093-2	minimap2	clair3	ONT	15326	A	G	0.9766	214	38.53
9093-2	minimap2	clair3	ONT	15487	A	T	0.9766	214	110
9093-2	minimap2	clair3	ONT	15930	G	A	0.9912	226	42.17
9093-2	minimap2	clair3	ONT	16223	C	T	0.9736	265	110
9093-2	minimap2	clair3	ONT	16274	G	A	0.9695	295	110
9093-2	minimap2	clair3	ONT	16298	T	C	0.9695	295	110
9093-2	minimap2	clair3	ONT	16325	T	C	0.9826	518	110

9093-2	minimap2	clair3	ONT	16327	C	T	0.9826	518	110
9093-3	minimap2	clair3	ONT	73	A	G	0.9114	237	41.85
9093-3	minimap2	clair3	ONT	185	G	A	0.9874	238	37.53
9093-3	minimap2	clair3	ONT	247	GA	G	0.9785	233	79.23
9093-3	minimap2	clair3	ONT	263	A	G	0.9871	233	39.01
9093-3	minimap2	clair3	ONT	285	CAA	C	0.9435	230	110
9093-3	minimap2	clair3	ONT	310	T	TC	0.6533	225	47.47
9093-3	minimap2	clair3	ONT	489	T	C	0.9952	210	110
9093-3	minimap2	clair3	ONT	750	A	G	0.9854	205	36.82
9093-3	minimap2	clair3	ONT	1438	A	G	0.9646	198	39.35
9093-3	minimap2	clair3	ONT	1888	G	A	0.9897	195	62.33
9093-3	minimap2	clair3	ONT	2706	A	G	0.9844	192	110
9093-3	minimap2	clair3	ONT	3106	CN	C	0.9105	190	70.79
9093-3	minimap2	clair3	ONT	3158	A	AT	0.9684	190	110
9093-3	minimap2	clair3	ONT	3552	T	A	0.9947	188	110
9093-3	minimap2	clair3	ONT	4654	C	T	0.984	188	37.01
9093-3	minimap2	clair3	ONT	4715	A	G	0.9255	188	110
9093-3	minimap2	clair3	ONT	4769	A	G	0.9947	189	110
9093-3	minimap2	clair3	ONT	7028	C	T	0.984	188	110
9093-3	minimap2	clair3	ONT	7196	C	A	0.9947	188	36.61
9093-3	minimap2	clair3	ONT	7253	A	G	0.9947	188	110
9093-3	minimap2	clair3	ONT	8477	T	C	0.9206	189	110
9093-3	minimap2	clair3	ONT	8584	G	A	0.9895	190	76.22
9093-3	minimap2	clair3	ONT	8701	A	G	0.9735	189	110
9093-3	minimap2	clair3	ONT	8860	A	G	0.9577	189	110
9093-3	minimap2	clair3	ONT	9540	T	C	0.9048	189	79.23
9093-3	minimap2	clair3	ONT	9545	A	G	0.873	189	39.6
9093-3	minimap2	clair3	ONT	10398	A	G	0.9572	187	39.45
9093-3	minimap2	clair3	ONT	10400	C	T	0.9626	187	39.24
9093-3	minimap2	clair3	ONT	10873	T	C	0.957	186	110
9093-3	minimap2	clair3	ONT	11719	G	A	0.9783	184	39.22
9093-3	minimap2	clair3	ONT	11914	G	A	0.9617	183	39.12
9093-3	minimap2	clair3	ONT	12705	C	T	0.9727	183	110

9093-3	minimap2	clair3	ONT	13263	A	G	0.9839	186	39.24
9093-3	minimap2	clair3	ONT	14318	T	C	0.9737	190	110
9093-3	minimap2	clair3	ONT	14766	C	T	0.9845	193	110
9093-3	minimap2	clair3	ONT	14783	T	C	0.9948	193	110
9093-3	minimap2	clair3	ONT	15043	G	A	1	196	110
9093-3	minimap2	clair3	ONT	15301	G	A	0.965	200	79.23
9093-3	minimap2	clair3	ONT	15326	A	G	0.99	201	110
9093-3	minimap2	clair3	ONT	15487	A	T	0.9852	203	110
9093-3	minimap2	clair3	ONT	15930	G	A	0.9763	211	40.07
9093-3	minimap2	clair3	ONT	16223	C	T	0.984	250	110
9093-3	minimap2	clair3	ONT	16274	G	A	0.9856	277	36.67
9093-3	minimap2	clair3	ONT	16298	T	C	0.9929	280	110
9093-3	minimap2	clair3	ONT	16325	T	C	0.994	499	110
9093-3	minimap2	clair3	ONT	16327	C	T	0.992	499	110
9251-1	minimap2	clair3	ONT	64	C	T	0.891	945	39.63
9251-1	minimap2	clair3	ONT	73	A	G	0.9619	945	44.08
9251-1	minimap2	clair3	ONT	146	T	C	0.9248	944	110
9251-1	minimap2	clair3	ONT	153	A	G	0.9872	937	39.82
9251-1	minimap2	clair3	ONT	235	A	G	0.9887	885	110
9251-1	minimap2	clair3	ONT	263	A	G	0.9875	882	37.12
9251-1	minimap2	clair3	ONT	310	T	TC	0.7205	873	44.73
9251-1	minimap2	clair3	ONT	513	GCA	G	0.9448	834	110
9251-1	minimap2	clair3	ONT	663	A	G	0.9867	828	38.23
9251-1	minimap2	clair3	ONT	750	A	G	0.9867	825	110
9251-1	minimap2	clair3	ONT	1438	A	G	0.9681	814	38.65
9251-1	minimap2	clair3	ONT	1555	A	G	0.6359	813	32.02
9251-1	minimap2	clair3	ONT	1736	A	G	0.9889	811	37.88
9251-1	minimap2	clair3	ONT	2706	A	G	0.9607	788	110
9251-1	minimap2	clair3	ONT	3106	CN	C	0.893	785	60.91
9251-1	minimap2	clair3	ONT	3393	C	T	0.9821	782	37.62
9251-1	minimap2	clair3	ONT	4248	T	C	0.8543	762	64.92
9251-1	minimap2	clair3	ONT	4769	A	G	0.9961	761	110
9251-1	minimap2	clair3	ONT	4824	A	G	0.9409	762	110

9251-1	minimap2	clair3	ONT	7028	C	T	0.988	751	110
9251-1	minimap2	clair3	ONT	8027	G	A	0.967	758	68.82
9251-1	minimap2	clair3	ONT	8794	C	T	0.9644	759	110
9251-1	minimap2	clair3	ONT	8860	A	G	0.9513	759	110
9251-1	minimap2	clair3	ONT	9377	A	G	0.9869	762	38.35
9251-1	minimap2	clair3	ONT	10771	A	G	0.9721	754	37.44
9251-1	minimap2	clair3	ONT	11719	G	A	0.9802	756	39.21
9251-1	minimap2	clair3	ONT	12007	G	A	0.9815	757	37.32
9251-1	minimap2	clair3	ONT	12705	C	T	0.9764	762	110
9251-1	minimap2	clair3	ONT	14766	C	T	0.9886	787	110
9251-1	minimap2	clair3	ONT	15074	T	C	0.9848	788	110
9251-1	minimap2	clair3	ONT	15326	A	G	0.9848	790	110
9251-1	minimap2	clair3	ONT	16093	T	C	0.9375	864	110
9251-1	minimap2	clair3	ONT	16111	C	T	0.9896	863	110
9251-1	minimap2	clair3	ONT	16213	G	A	0.9944	1416	110
9251-1	minimap2	clair3	ONT	16223	C	T	0.9867	1424	37.54
9251-1	minimap2	clair3	ONT	16290	C	T	0.9833	1494	110
9251-1	minimap2	clair3	ONT	16294	C	T	0.988	1500	110
9251-1	minimap2	clair3	ONT	16319	G	A	0.9881	1510	38.75
9251-1	minimap2	clair3	ONT	16362	T	C	0.944	1643	110
9251-1	minimap2	clair3	ONT	16519	T	C	0.9943	175	110
9251-2	minimap2	clair3	ONT	64	C	T	0.8807	721	39.61
9251-2	minimap2	clair3	ONT	73	A	G	0.9667	721	43.22
9251-2	minimap2	clair3	ONT	146	T	C	0.9335	722	110
9251-2	minimap2	clair3	ONT	153	A	G	0.9778	720	40.16
9251-2	minimap2	clair3	ONT	235	A	G	0.9839	684	110
9251-2	minimap2	clair3	ONT	263	A	G	0.9941	678	36.5
9251-2	minimap2	clair3	ONT	310	T	TC	0.7437	671	55.24
9251-2	minimap2	clair3	ONT	513	GCA	G	0.9642	642	110
9251-2	minimap2	clair3	ONT	663	A	G	0.9828	638	37.75
9251-2	minimap2	clair3	ONT	750	A	G	0.9843	636	110
9251-2	minimap2	clair3	ONT	1438	A	G	0.9729	628	37.33
9251-2	minimap2	clair3	ONT	1555	A	G	0.6491	627	44.29

9251-2	minimap2	clair3	ONT	1736	A	G	0.9806	619	36.69
9251-2	minimap2	clair3	ONT	2706	A	G	0.9804	612	110
9251-2	minimap2	clair3	ONT	3106	CN	C	0.9055	603	74.46
9251-2	minimap2	clair3	ONT	3393	C	T	0.9834	602	110
9251-2	minimap2	clair3	ONT	4248	T	C	0.8625	589	110
9251-2	minimap2	clair3	ONT	4769	A	G	0.9966	587	110
9251-2	minimap2	clair3	ONT	4824	A	G	0.9182	587	110
9251-2	minimap2	clair3	ONT	7028	C	T	0.9882	591	110
9251-2	minimap2	clair3	ONT	8027	G	A	0.9677	589	58.06
9251-2	minimap2	clair3	ONT	8794	C	T	0.9746	590	36.5
9251-2	minimap2	clair3	ONT	8860	A	G	0.9593	590	110
9251-2	minimap2	clair3	ONT	9377	A	G	0.9932	589	37.96
9251-2	minimap2	clair3	ONT	10771	A	G	0.9742	582	36.53
9251-2	minimap2	clair3	ONT	11719	G	A	0.9812	586	39.15
9251-2	minimap2	clair3	ONT	12007	G	A	0.9847	587	37.17
9251-2	minimap2	clair3	ONT	12705	C	T	0.9781	593	110
9251-2	minimap2	clair3	ONT	14766	C	T	0.9838	616	110
9251-2	minimap2	clair3	ONT	15074	T	C	0.9825	627	110
9251-2	minimap2	clair3	ONT	15326	A	G	0.9841	630	110
9251-2	minimap2	clair3	ONT	16093	T	C	0.9364	692	110
9251-2	minimap2	clair3	ONT	16111	C	T	0.9942	694	110
9251-2	minimap2	clair3	ONT	16213	G	A	0.9884	1121	110
9251-2	minimap2	clair3	ONT	16223	C	T	0.9858	1123	37.13
9251-2	minimap2	clair3	ONT	16290	C	T	0.9888	1164	110
9251-2	minimap2	clair3	ONT	16294	C	T	0.9889	1166	110
9251-2	minimap2	clair3	ONT	16319	G	A	0.9914	1168	37.58
9251-2	minimap2	clair3	ONT	16362	T	C	0.9472	1287	110
9251-2	minimap2	clair3	ONT	16519	T	C	1	126	110
9251-3	minimap2	clair3	ONT	64	C	T	0.8681	432	64.47
9251-3	minimap2	clair3	ONT	73	A	G	0.963	433	43.52
9251-3	minimap2	clair3	ONT	146	T	C	0.9355	434	110
9251-3	minimap2	clair3	ONT	153	A	G	0.9677	433	39.4
9251-3	minimap2	clair3	ONT	235	A	G	0.988	418	110

9251-3	minimap2	clair3	ONT	263	A	G	0.9688	417	37.39
9251-3	minimap2	clair3	ONT	310	T	TC	0.7348	411	44.37
9251-3	minimap2	clair3	ONT	513	GCA	G	0.9431	404	79.23
9251-3	minimap2	clair3	ONT	663	A	G	0.9876	402	38.39
9251-3	minimap2	clair3	ONT	750	A	G	0.9748	397	110
9251-3	minimap2	clair3	ONT	1438	A	G	0.9542	393	37.98
9251-3	minimap2	clair3	ONT	1555	A	G	0.6607	392	42.24
9251-3	minimap2	clair3	ONT	1736	A	G	0.9975	393	37.37
9251-3	minimap2	clair3	ONT	2706	A	G	0.9923	388	110
9251-3	minimap2	clair3	ONT	3106	CN	C	0.9096	387	61.11
9251-3	minimap2	clair3	ONT	3393	C	T	0.9769	389	110
9251-3	minimap2	clair3	ONT	4248	T	C	0.8646	384	71.45
9251-3	minimap2	clair3	ONT	4769	A	G	1	383	110
9251-3	minimap2	clair3	ONT	4824	A	G	0.8877	383	37.79
9251-3	minimap2	clair3	ONT	7028	C	T	0.984	376	110
9251-3	minimap2	clair3	ONT	8027	G	A	0.9416	377	66.93
9251-3	minimap2	clair3	ONT	8794	C	T	0.9787	375	110
9251-3	minimap2	clair3	ONT	8860	A	G	0.96	375	110
9251-3	minimap2	clair3	ONT	9377	A	G	1	378	38.09
9251-3	minimap2	clair3	ONT	10771	A	G	0.9764	382	37.67
9251-3	minimap2	clair3	ONT	11719	G	A	0.9922	385	39.41
9251-3	minimap2	clair3	ONT	12007	G	A	0.9948	387	37.11
9251-3	minimap2	clair3	ONT	12705	C	T	0.9846	390	110
9251-3	minimap2	clair3	ONT	14766	C	T	0.9975	402	110
9251-3	minimap2	clair3	ONT	15074	T	C	0.9781	411	110
9251-3	minimap2	clair3	ONT	15326	A	G	0.9928	415	110
9251-3	minimap2	clair3	ONT	16093	T	C	0.9344	457	110
9251-3	minimap2	clair3	ONT	16111	C	T	0.9869	459	110
9251-3	minimap2	clair3	ONT	16213	G	A	0.9894	754	36.53
9251-3	minimap2	clair3	ONT	16223	C	T	0.9841	757	37.23
9251-3	minimap2	clair3	ONT	16290	C	T	0.976	791	110
9251-3	minimap2	clair3	ONT	16294	C	T	0.9899	794	110
9251-3	minimap2	clair3	ONT	16319	G	A	0.985	801	37.42

9251-3	minimap2	clair3	ONT	16362	T	C	0.9339	877	110
9251-3	minimap2	clair3	ONT	16519	T	C	0.9897	97	110
3312	bwa	mutserve	ILL	73	A	G	0.999	3349	36.87
3312	bwa	mutserve	ILL	263	A	G	0.998	2269	36.43
3312	bwa	mutserve	ILL	494	C	A	0.788	387	36.31
3312	bwa	mutserve	ILL	499	G	A	0.998	2754	36.24
3312	bwa	mutserve	ILL	750	A	G	1	3318	36.79
3312	bwa	mutserve	ILL	827	A	G	0.997	3982	36.56
3312	bwa	mutserve	ILL	1438	A	G	0.999	3564	36.84
3312	bwa	mutserve	ILL	2706	A	G	0.999	1172	36.92
3312	bwa	mutserve	ILL	3547	A	G	1	659	36.6
3312	bwa	mutserve	ILL	4122	A	G	1	587	36.61
3312	bwa	mutserve	ILL	4123	A	G	1	579	36.61
3312	bwa	mutserve	ILL	4769	A	G	1	316	36.32
3312	bwa	mutserve	ILL	4820	G	A	1	375	35.91
3312	bwa	mutserve	ILL	4977	T	C	0.999	693	36.72
3312	bwa	mutserve	ILL	5460	G	A	0.997	616	36.37
3312	bwa	mutserve	ILL	6473	C	T	1	630	36.56
3312	bwa	mutserve	ILL	7028	C	T	0.995	388	36.6
3312	bwa	mutserve	ILL	8860	A	G	1	228	36.53
3312	bwa	mutserve	ILL	8875	T	C	1	314	36.62
3312	bwa	mutserve	ILL	9682	T	C	1	590	36.82
3312	bwa	mutserve	ILL	9950	T	C	1	556	36.68
3312	bwa	mutserve	ILL	11177	C	T	0.999	5014	36.72
3312	bwa	mutserve	ILL	11335	C	T	0.999	4524	36.32
3312	bwa	mutserve	ILL	11719	G	A	0.998	4195	36.69
3312	bwa	mutserve	ILL	13590	G	A	1	3439	36.39
3312	bwa	mutserve	ILL	14766	C	T	0.999	3172	35.88
3312	bwa	mutserve	ILL	15326	A	G	1	3590	36.58
3312	bwa	mutserve	ILL	15535	C	T	0.999	3473	35.62
3312	bwa	mutserve	ILL	16183	A	C	0.853	1049	36.49
3312	bwa	mutserve	ILL	16189	T	C	0.988	1383	36.63
3312	bwa	mutserve	ILL	16217	T	C	0.999	2145	36.6

3312	bwa	mutserve	ILL	16519	T	C	1	3434	36.69
3396	bwa	mutserve	ILL	64	C	T	1	249	39.74
3396	bwa	mutserve	ILL	73	A	G	1	285	39.49
3396	bwa	mutserve	ILL	146	T	C	1	560	39.63
3396	bwa	mutserve	ILL	153	A	G	1	569	39.55
3396	bwa	mutserve	ILL	174	C	T	1	567	39.72
3396	bwa	mutserve	ILL	235	A	G	0.999	749	39.32
3396	bwa	mutserve	ILL	263	A	G	1	673	38.53
3396	bwa	mutserve	ILL	663	A	G	1	858	39.61
3396	bwa	mutserve	ILL	750	A	G	0.999	700	39.38
3396	bwa	mutserve	ILL	1438	A	G	1	872	39.72
3396	bwa	mutserve	ILL	1736	A	G	1	1005	39.24
3396	bwa	mutserve	ILL	2706	A	G	1	2065	39.73
3396	bwa	mutserve	ILL	4248	T	C	1	880	39.67
3396	bwa	mutserve	ILL	4769	A	G	0.996	451	39.14
3396	bwa	mutserve	ILL	4824	A	G	0.999	693	39.75
3396	bwa	mutserve	ILL	7028	C	T	1	710	39.46
3396	bwa	mutserve	ILL	8027	G	A	1	576	39.56
3396	bwa	mutserve	ILL	8056	A	G	0.998	518	39.13
3396	bwa	mutserve	ILL	8794	C	T	1	580	39.2
3396	bwa	mutserve	ILL	8860	A	G	0.994	310	38.91
3396	bwa	mutserve	ILL	11719	G	A	1	1072	39.54
3396	bwa	mutserve	ILL	12007	G	A	1	968	38.5
3396	bwa	mutserve	ILL	12278	T	C	0.256	587	39.27
3396	bwa	mutserve	ILL	12705	C	T	1	970	39.44
3396	bwa	mutserve	ILL	14766	C	T	1	1077	39.24
3396	bwa	mutserve	ILL	15326	A	G	0.996	837	39.06
3396	bwa	mutserve	ILL	15443	C	T	0.95	765	39.12
3396	bwa	mutserve	ILL	16111	C	T	0.984	1041	39.48
3396	bwa	mutserve	ILL	16129	G	A	0.993	1077	39.76
3396	bwa	mutserve	ILL	16223	C	T	0.995	1208	38.95
3396	bwa	mutserve	ILL	16239	C	T	0.986	1208	38.78
3396	bwa	mutserve	ILL	16290	C	T	0.99	1056	39.45

3396	bwa	mutserve	ILL	16319	G	A	0.991	1017	39.63
3396	bwa	mutserve	ILL	16362	T	C	0.993	878	39.23
3258	bwa	mutserve	ILL	73	A	G	0.999	3051	36.89
3258	bwa	mutserve	ILL	152	T	C	1	2673	36.76
3258	bwa	mutserve	ILL	263	A	G	0.999	2165	36.61
3258	bwa	mutserve	ILL	310	T	C	0.875	1308	36.63
3258	bwa	mutserve	ILL	515	A	G	0.759	390	36.18
3258	bwa	mutserve	ILL	750	A	G	1	3006	36.82
3258	bwa	mutserve	ILL	921	T	C	0.999	3088	36.82
3258	bwa	mutserve	ILL	1438	A	G	1	2948	36.8
3258	bwa	mutserve	ILL	2706	A	G	0.999	5434	36.8
3258	bwa	mutserve	ILL	4769	A	G	0.998	1309	36.39
3258	bwa	mutserve	ILL	5046	G	A	0.999	2380	36.62
3258	bwa	mutserve	ILL	5147	G	A	0.993	1075	36.64
3258	bwa	mutserve	ILL	6272	A	G	0.999	3870	36.82
3258	bwa	mutserve	ILL	6680	T	C	0.999	3656	36.69
3258	bwa	mutserve	ILL	7028	C	T	0.998	1823	36.72
3258	bwa	mutserve	ILL	7424	A	G	1	1692	36.48
3258	bwa	mutserve	ILL	8014	A	C	1	2736	36.79
3258	bwa	mutserve	ILL	8618	T	C	0.999	1671	35.99
3258	bwa	mutserve	ILL	8701	A	G	0.999	2335	36.46
3258	bwa	mutserve	ILL	8860	A	G	0.989	798	36.47
3258	bwa	mutserve	ILL	9540	T	C	1	3465	36.7
3258	bwa	mutserve	ILL	10398	A	G	0.999	3421	36.81
3258	bwa	mutserve	ILL	10873	T	C	1	5463	36.48
3258	bwa	mutserve	ILL	11719	G	A	0.999	3804	36.72
3258	bwa	mutserve	ILL	12705	C	T	0.999	2885	36.48
3258	bwa	mutserve	ILL	13105	A	G	0.999	3541	36.75
3258	bwa	mutserve	ILL	13886	T	C	1	2760	36.64
3258	bwa	mutserve	ILL	14071	A	G	1	2537	36.5
3258	bwa	mutserve	ILL	14284	C	T	0.997	2642	35.72
3258	bwa	mutserve	ILL	14766	C	T	0.999	2782	36
3258	bwa	mutserve	ILL	15301	G	A	1	3067	36.14

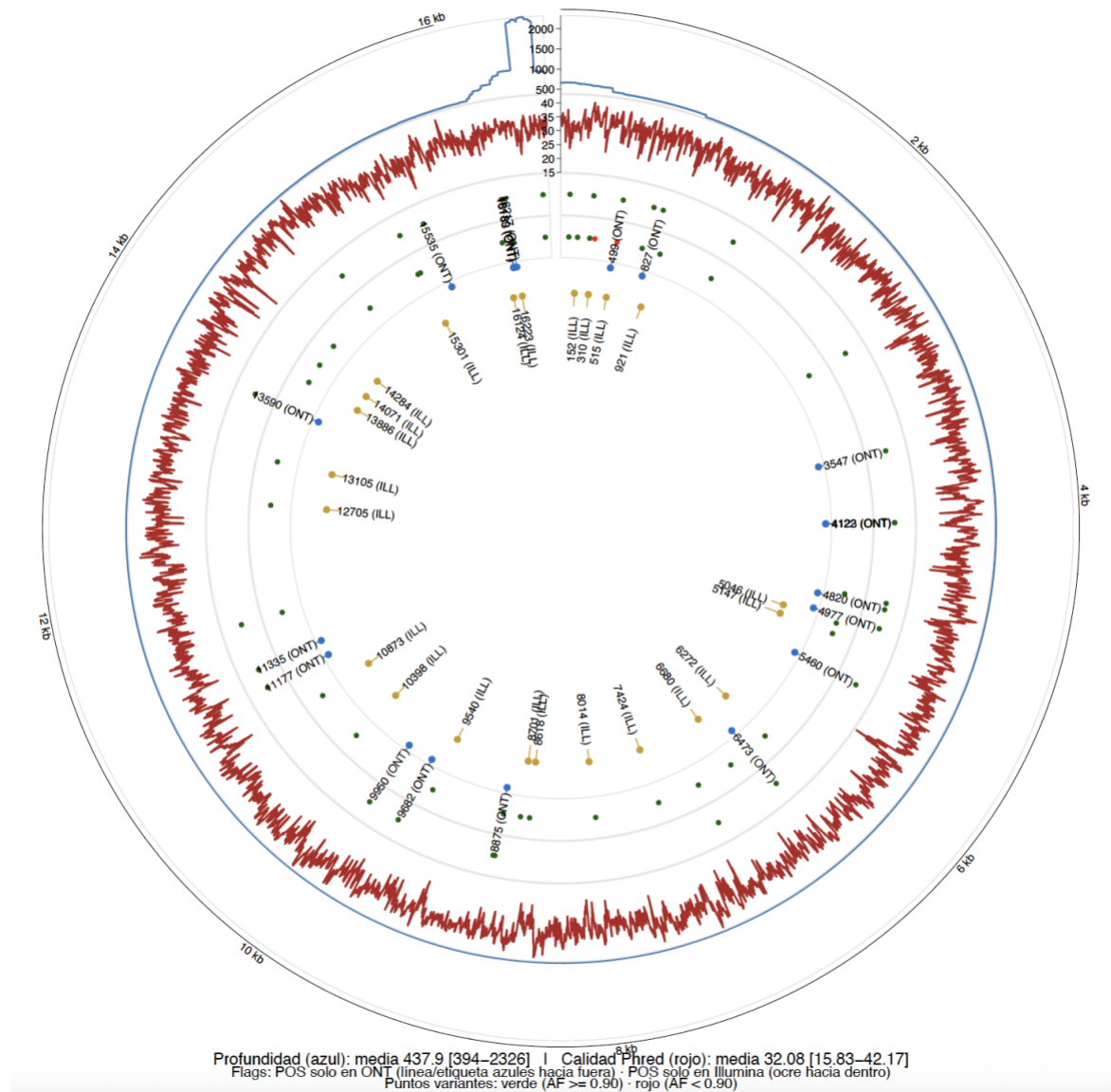
3258	bwa	mutserve	ILL	15326	A	G	1	2945	36.62
3258	bwa	mutserve	ILL	16124	T	C	0.999	3093	36.77
3258	bwa	mutserve	ILL	16223	C	T	0.996	3562	36.52
3258	bwa	mutserve	ILL	16519	T	C	0.999	3099	36.65
9251	bwa	mutserve	ILL	64	C	T	1	525	33.62
9251	bwa	mutserve	ILL	73	A	G	1	605	33.87
9251	bwa	mutserve	ILL	146	T	C	1	1196	33.83
9251	bwa	mutserve	ILL	153	A	G	1	1225	33.86
9251	bwa	mutserve	ILL	235	A	G	1	1277	33.79
9251	bwa	mutserve	ILL	263	A	G	1	977	33.75
9251	bwa	mutserve	ILL	515	A	G	0.872	86	33.48
9251	bwa	mutserve	ILL	663	A	G	1	1190	33.83
9251	bwa	mutserve	ILL	750	A	G	1	1053	33.86
9251	bwa	mutserve	ILL	1438	A	G	0.998	1273	33.83
9251	bwa	mutserve	ILL	1555	A	G	0.689	1335	33.71
9251	bwa	mutserve	ILL	1736	A	G	1	1557	33.78
9251	bwa	mutserve	ILL	2706	A	G	0.999	3056	33.85
9251	bwa	mutserve	ILL	3393	C	T	1	1108	33.82
9251	bwa	mutserve	ILL	4248	T	C	1	1180	33.8
9251	bwa	mutserve	ILL	4769	A	G	1	637	33.66
9251	bwa	mutserve	ILL	4824	A	G	0.999	1047	33.76
9251	bwa	mutserve	ILL	7028	C	T	0.999	1138	33.68
9251	bwa	mutserve	ILL	8027	G	A	1	1192	33.84
9251	bwa	mutserve	ILL	8794	C	T	1	885	33.76
9251	bwa	mutserve	ILL	8860	A	G	0.994	320	33.87
9251	bwa	mutserve	ILL	9377	A	G	1	1067	33.76
9251	bwa	mutserve	ILL	10771	A	G	0.999	1966	33.81
9251	bwa	mutserve	ILL	11719	G	A	1	1791	33.74
9251	bwa	mutserve	ILL	12007	G	A	1	1467	33.71
9251	bwa	mutserve	ILL	12705	C	T	1	1362	33.85
9251	bwa	mutserve	ILL	14766	C	T	1	1702	33.63
9251	bwa	mutserve	ILL	15074	T	C	1	1198	33.84
9251	bwa	mutserve	ILL	15326	A	G	1	1363	33.7

9251	bwa	mutserve	ILL	16093	T	C	0.993	1621	33.85
9251	bwa	mutserve	ILL	16111	C	T	1	1580	33.76
9251	bwa	mutserve	ILL	16213	G	A	1	1861	33.69
9251	bwa	mutserve	ILL	16223	C	T	0.999	1839	33.57
9251	bwa	mutserve	ILL	16290	C	T	0.999	1460	33.85
9251	bwa	mutserve	ILL	16294	C	T	1	1534	33.64
9251	bwa	mutserve	ILL	16319	G	A	1	1426	33.62
9251	bwa	mutserve	ILL	16362	T	C	0.998	1323	33.79
9251	bwa	mutserve	ILL	16519	T	C	1	977	33.75
9093	bwa	mutserve	ILL	73	A	G	0.999	2510	36.87
9093	bwa	mutserve	ILL	185	G	A	0.998	2163	36.54
9093	bwa	mutserve	ILL	248	A	T	0.069	216	36.1
9093	bwa	mutserve	ILL	263	A	G	0.998	1629	36.52
9093	bwa	mutserve	ILL	489	T	C	1	2077	36.58
9093	bwa	mutserve	ILL	750	A	G	0.995	2463	36.85
9093	bwa	mutserve	ILL	1888	G	A	1	2667	36.47
9093	bwa	mutserve	ILL	3552	T	A	0.999	2632	36.49
9093	bwa	mutserve	ILL	4654	C	T	0.996	1065	36.75
9093	bwa	mutserve	ILL	4715	A	G	0.999	1368	36.58
9093	bwa	mutserve	ILL	7028	C	T	0.997	1452	36.79
9093	bwa	mutserve	ILL	7196	C	A	0.997	2931	36.41
9093	bwa	mutserve	ILL	7253	A	G	0.991	2283	36.71
9093	bwa	mutserve	ILL	8584	G	A	0.999	2156	36.19
9093	bwa	mutserve	ILL	8701	A	G	0.999	2060	36.35
9093	bwa	mutserve	ILL	8860	A	G	0.996	758	36.48
9093	bwa	mutserve	ILL	9540	T	C	1	2548	36.73
9093	bwa	mutserve	ILL	10398	A	G	0.999	2383	36.8
9093	bwa	mutserve	ILL	10400	C	T	1	2372	36.69
9093	bwa	mutserve	ILL	11719	G	A	0.997	2698	36.79
9093	bwa	mutserve	ILL	11914	G	A	0.997	2859	36.61
9093	bwa	mutserve	ILL	12705	C	T	1	2085	36.38
9093	bwa	mutserve	ILL	13263	A	G	1	2683	36.75
9093	bwa	mutserve	ILL	14318	T	C	0.998	2602	36.62

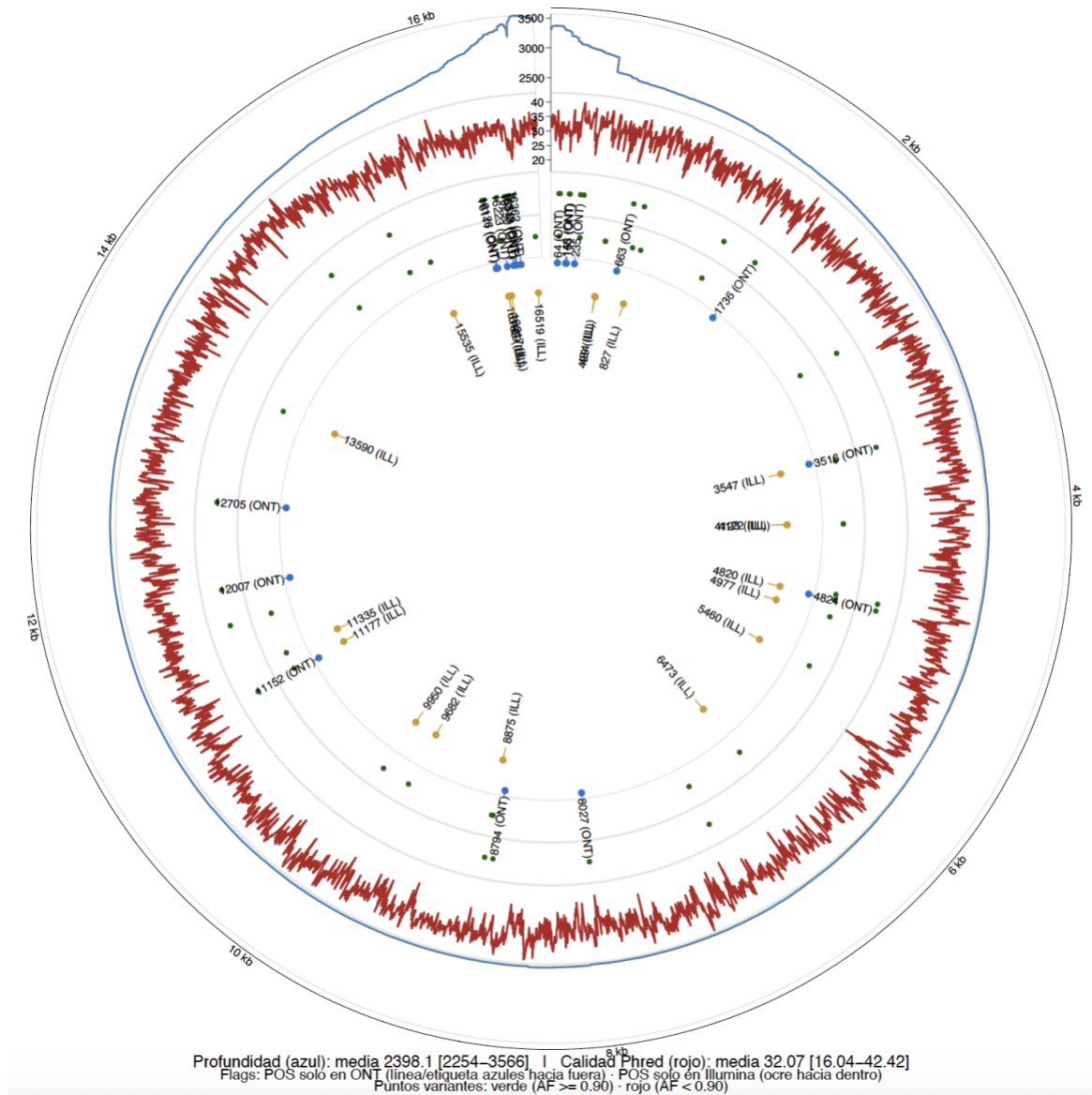
9093	bwa	mutserve	ILL	14766	C	T	0.999	2368	36.18
9093	bwa	mutserve	ILL	14783	T	C	1	2316	36.37
9093	bwa	mutserve	ILL	15043	G	A	0.998	2409	36.8
9093	bwa	mutserve	ILL	15301	G	A	0.999	2777	36.16
9093	bwa	mutserve	ILL	15326	A	G	0.999	2659	36.64
9093	bwa	mutserve	ILL	15487	A	T	0.996	2313	36.56
9093	bwa	mutserve	ILL	15930	G	A	1	1822	36.07
9093	bwa	mutserve	ILL	16223	C	T	0.997	3132	36.38
9093	bwa	mutserve	ILL	16274	G	A	0.998	3024	36.81
9093	bwa	mutserve	ILL	16298	T	C	0.995	3309	36.66
9093	bwa	mutserve	ILL	16325	T	C	0.994	2933	36.32
9093	bwa	mutserve	ILL	16327	C	T	0.997	3002	36.68

Tabla 7. Tabla con los datos de métricas de las variantes encontradas con cada plataformas y para cada muestra.

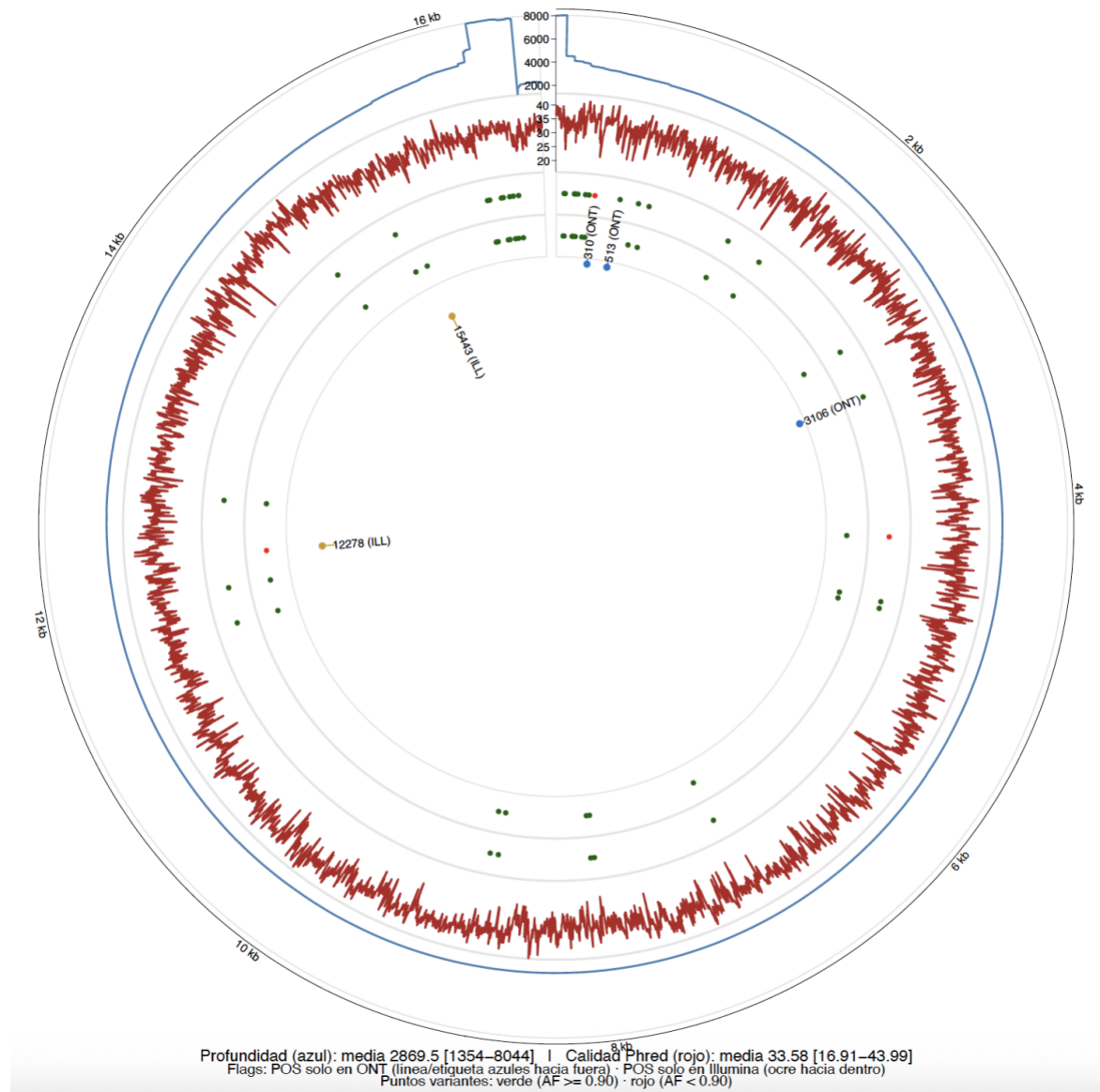
**B. Anexos 2-5: Graficos de circos
plot para las muestras 3258, 3312,
3396 y 9093.**



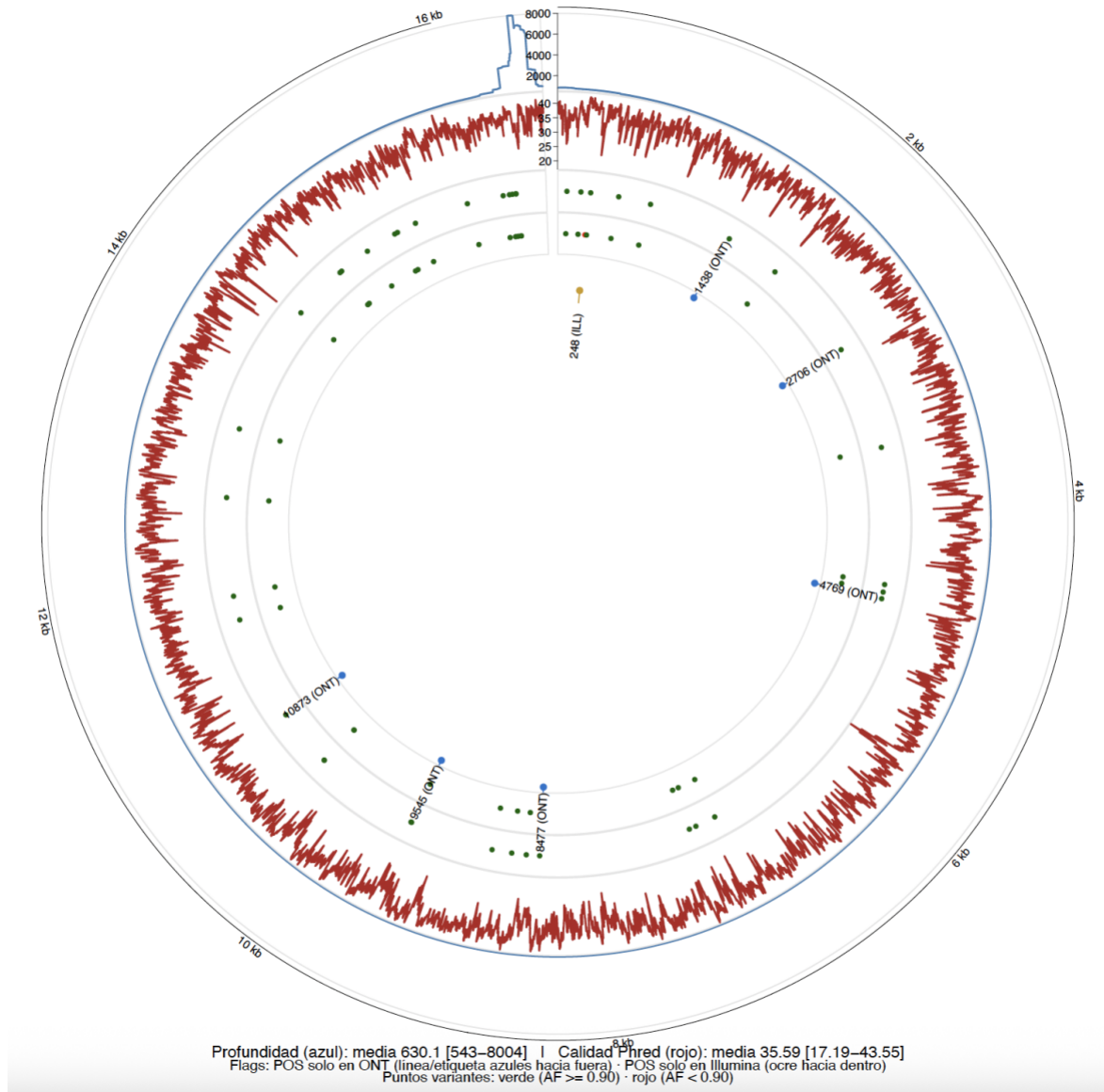
Anexo 2. Gráfico de circos para la muestra 3258: cobertura, calidad y variantes ONT vs Illumina. Nota. Anillo externo: escala de posición del ADNmt (0–16.6 kb). Anillos de Profundidad (línea azul continua) y calidad media Phred (línea roja continua) por posición. Se muestran sus ejes radiales para referencia. Variantes como puntos: verde si AF ≥ 0.90, rojo si AF < 0.90. Capa para ONT y otra para Illumina para las variantes encontradas para cada plataforma (las etiquetas de Illumina se orientan hacia el interior y las de ONT hacia el exterior). Banderines (flags) de discrepancia por posición, marcas con línea y etiqueta solo cuando una plataforma detecta una posición que la otra no reporta (discordancia por presencia/ausencia).



Anexo 3. Gráfico de circos para la muestra 3312: cobertura, calidad y variantes ONT vs Illumina. Nota. Anillo externo: escala de posición del ADNmt (0–16.6 kb). Anillos de Profundidad (línea azul continua) y calidad media Phred (línea roja continua) por posición. Se muestran sus ejes radiales para referencia. Variantes como puntos: verde si AF ≥ 0.90, rojo si AF < 0.90. Capa para ONT y otra para Illumina para las variantes encontradas para cada plataforma (las etiquetas de Illumina se orientan hacia el interior y las de ONT hacia el exterior). Banderines (flags) de discrepancia por posición, marcas con línea y etiqueta solo cuando una plataforma detecta una posición que la otra no reporta (discordancia por presencia/ausencia).



Anexo 4. Gráfico de circos para la muestra 3396: cobertura, calidad y variantes ONT vs Illumina. Nota. Anillo externo: escala de posición del ADNmt (0–16.6 kb). Anillos de Profundidad (línea azul continua) y calidad media Phred (línea roja continua) por posición. Se muestran sus ejes radiales para referencia. Variantes como puntos: verde si AF ≥ 0.90, rojo si AF < 0.90. Capa para ONT y otra para Illumina para las variantes encontradas para cada plataforma (las etiquetas de Illumina se orientan hacia el interior y las de ONT hacia el exterior). Banderines (flags) de discrepancia por posición, marcas con línea y etiqueta solo cuando una plataforma detecta una posición que la otra no reporta (discordancia por presencia/ausencia).



Anexo 5. Gráfico de circos para la muestra 9093: cobertura, calidad y variantes ONT vs Illumina. Nota. Anillo externo: escala de posición del ADNmt (0–16.6 kb). Anillos de Profundidad (línea azul continua) y calidad media Phred (línea roja continua) por posición. Se muestran sus ejes radiales para referencia. Variantes como puntos: verde si $AF \geq 0.90$, rojo si $AF < 0.90$. Capa para ONT y otra para Illumina para las variantes encontradas para cada plataforma (las etiquetas de Illumina se orientan hacia el interior y las de ONT hacia el exterior). Banderines (flags) de discrepancia por posición, marcas con línea y etiqueta solo cuando una plataforma detecta una posición que la otra no reporta (discordancia por presencia/ausencia).

Bibliografía

- Abdelwahab, O., & Torkamaneh, D. (2025). Artificial intelligence in variant calling: a review. *Frontiers in Bioinformatics*, 5, 1574359. <https://doi.org/10.3389/FBINF.2025.1574359>
- Abicht, A., Scharf, F., Kleinle, S., Schön, U., Holinski-Feder, E., Horvath, R., Benet-Pagès, A., & Diebold, I. (2018). Mitochondrial and nuclear disease panel (Mito-aND-Panel): Combined sequencing of mitochondrial and nuclear DNA by a cost-effective and sensitive NGS-based method. *Molecular Genetics and Genomic Medicine*, 6(6), 1188–1198. <https://doi.org/10.1002/mgg3.500>
- Ahmad, R., & Hasan, M. Y. (2021). Next-generation sequencing technology in the diagnosis of mitochondrial disorders. In *International Journal of Health Sciences* (Vol. 15).
- Ahmed, S. T., Craven, L., Russell, O. M., Turnbull, D. M., & Vincent, A. E. (2018). Diagnosis and Treatment of Mitochondrial Myopathies. In *Neurotherapeutics* (Vol. 15, Issue 4, pp. 943–953). Springer New York LLC. <https://doi.org/10.1007/s13311-018-00674-4>
- Ahsan, M. U., Liu, Q., Perdomo, J. E., Fang, L., & Wang, K. (2023). A survey of algorithms for the detection of genomic structural variants from long-read sequencing data. In *Nature Methods* (Vol. 20, Issue 8, pp. 1143–1158). Nature Research. <https://doi.org/10.1038/s41592-023-01932-w>
- Akamatsu, S., Mitsuhashi, S., Soga, K., Mizukami, H., Shiraishi, M., Frith, M. C., & Yamano, Y. (2024). Targeted nanopore sequencing using the Flongle device to identify mitochondrial DNA variants. *Scientific Reports 2024 14:1*, 14(1), 1–8. <https://doi.org/10.1038/s41598-024-75749-8>
- Akar, H. T., Sayar, E., Sarlıtaş Nakip, Ö., Sönmez, E., Özkan, M. B., & Olgaç, A. (2024). Leigh Syndrome due to MT-ATP6 Variants: A Case Presentation and the Review of the Literature. *Molecular Syndromology*, 15(4), 333–338. <https://doi.org/10.1159/000536676>
- Akintunde, O., Tucker, T., & Carabetta, V. J. (2023). *The evolution of next-generation sequencing technologies*.

- Amarasinghe, S. L., Su, S., Dong, X., Zappia, L., Ritchie, M. E., & Gouil, Q. (2020). Opportunities and challenges in long-read sequencing data analysis. *Genome Biology* 2020 21:1, 21(1), 1–16. <https://doi.org/10.1186/S13059-020-1935-5>
- Anderson, M. J. (2001). A new method for non-parametric multivariate analysis of variance. *Austral Ecology*, 26(1), 32–46. <https://doi.org/10.1111/J.1442-9993.2001.01070.PP.X>
- Athanasopoulou, K., Michalopoulou, V. I., Scorilas, A., & Adamopoulos, P. G. (2025). Integrating Artificial Intelligence in Next-Generation Sequencing: Advances, Challenges, and Future Directions. In *Current Issues in Molecular Biology* (Vol. 47, Issue 6). Multidisciplinary Digital Publishing Institute (MDPI). <https://doi.org/10.3390/cimb47060470>
- Basu, U., Bostwick, A. M., Das, K., Dittenhafer-Reed, K. E., & Patel, S. S. (2020). Structure, mechanism, and regulation of mitochondrial DNA transcription initiation. *Journal of Biological Chemistry*, 295(52), 18406–18425. <https://doi.org/10.1074/jbc.REV120.011202>
- Baudry, L., Du Vivant, C., & Baudry, P. L. (n.d.). *Investigating chromosome dynamics through Hi-C assembly Sorbonne Université Investigating chromosome dynamics through Hi-C assembly*. <https://theses.hal.science/tel-02935877v1>
- Bayley, H. (2015). Nanopore sequencing: From imagination to reality. *Clinical Chemistry*, 61(1), 25–31. <https://doi.org/10.1373/clinchem.2014.223016>
- Benjamin, D., Sato, T., Cibulskis, K., Getz, G., Stewart, C., & Lichtenstein, L. (2019). *Calling Somatic SNVs and Indels with Mutect2*. <https://doi.org/10.1101/861054>
- Berdanier, C. D., & Everts, H. B. (2001). Mitochondrial DNA in aging and degenerative disease. In *Mutation Research* (Vol. 475).
- Björkman, K., Vissing, J., Østergaard, E., Bindoff, L. A., de Coo, I. F. M., Engvall, M., Hikmat, O., Isohanni, P., Kollberg, G., Lindberg, C., Majamaa, K., Naess, K., Uusimaa, J., Tulinius, M., & Darin, N. (2023). Phenotypic spectrum and clinical course of single large-scale mitochondrial DNA deletion disease in the paediatric population: a multicentre study. *Journal of Medical Genetics*, 60(1), 65–73. <https://doi.org/10.1136/JMEDGENET-2021-108006>
- Blanca, A., Harris, R. S., Koslicki, D., & Medvedev, P. (2021). *The statistics of k-mers from a sequence undergoing a simple mutation process without spurious matches*. <https://doi.org/10.1101/2021.01.15.426881>
- Blanco-Grau, A., Bonaventura-Ibars, I., Coll-Cantí, J., Melià, M. J., Martínez, R., Martínez-Gallo, M., Andreu, A. L., Pinós, T., & García-Arumí, E. (2013). Identification and

- biochemical characterization of the novel mutation m.8839G>C in the mitochondrial ATP6 gene associated with NARP syndrome. *Genes, Brain and Behavior*, 12(8), 812–820. <https://doi.org/10.1111/gbb.12089>
- Branton, D., Deamer, D. W., Marziali, A., Bayley, H., Benner, S. A., Butler, T., Di Ventra, M., Garaj, S., Hibbs, A., Huang, X., Jovanovich, S. B., Krstic, P. S., Lindsay, S., Sean Ling, X., Mastrangelo, C. H., Meller, A., Oliver, J. S., Pershin, Y. V., Michael Ramsey, J., ... Biotechnol Author manuscript, N. (2008). The potential and challenges of nanopore sequencing. *Nat Biotechnol*, 26(10), 1146–1153. <https://doi.org/10.1038/nbt.1495>
- Braun, E. (2024). Mitochondrial replacement techniques for treating infertility. *Journal of Medical Ethics*. <https://doi.org/10.1136/jme-2023-109660>
- Bravo-Alonso, I., Navarrete, R., Vega, A. I., Ruíz-Sala, P., Silva, M. T. G., Martín-Hernández, E., Quijada-Fraile, P., Belanger-Quintana, A., Stanescu, S., Bueno, M., Vitoria, I., Toledo, L., Couce, M. L., García-Jiménez, I., Ramos-Ruiz, R., Martín, M. Á., Desviat, L. R., Ugarte, M., Pérez-Cerdá, C., ... Rodríguez-Pombo, P. (2019). Genes and variants underlying human congenital lactic acidosis—from genetics to personalized treatment. *Journal of Clinical Medicine*, 8(11). <https://doi.org/10.3390/JCM8111811>,
- broadinstitute/gatk: Official code repository for GATK versions 4 and up*. (n.d.). Retrieved August 27, 2025, from <https://github.com/broadinstitute/gatk>
- Broom, C. J., Weyl, O. L. F., & South, J. (2023). Habitat associations of imperilled fishes after conservation intervention in the Cape Fold Ecoregion, South Africa. *Journal of Fish Biology*, 102(2), 317–327. <https://doi.org/10.1111/jfb.15262>
- Cannon, S. J., Hall, T., Hawkes, G., Colclough, K., Boggan, R. M., Wright, C. F., Pickett, S. J., Hattersley, A. T., Weedon, M. N., & Patel, K. A. (2024). Penetrance and expressivity of mitochondrial variants in a large clinically unselected population. *Human Molecular Genetics*, 33(5), 465–474. <https://doi.org/10.1093/hmg/ddad194>
- Carré, F., Hervochon, R., Foirest, C., & Tankéré, F. (2019). MERRF syndrome (Myoclonic Epilepsy with Ragged Red Fibres) presenting with cervicothoracic lipomatosis. *European Annals of Otorhinolaryngology, Head and Neck Diseases*, 136(2), 113–114. <https://doi.org/10.1016/j.anorl.2018.10.015>
- Castro-Wallace, S. L., Chiu, C. Y., John, K. K., Stahl, S. E., Rubins, K. H., McIntyre, A. B. R., Dworkin, J. P., Lupisella, M. L., Smith, D. J., Botkin, D. J., Stephenson, T. A., Juul, S., Turner, D. J., Izquierdo, F., Federman, S., Stryke, D., Somasekar, S., Alexander, N., Yu, G., ... Burton, A. S. (2017). Nanopore DNA Sequencing and Genome Assembly on the International Space Station. *Scientific Reports*, 7(1). <https://doi.org/10.1038/s41598-017-18364-0>

- Catarino, C. B., Ahting, U., Gusic, M., Iuso, A., Repp, B., Peters, K., Biskup, S., von Livonius, B., Prokisch, H., & Klopstock, T. (2017). Characterization of a Leber's hereditary optic neuropathy (LHON) family harboring two primary LHON mutations m.11778G > A and m.14484T > C of the mitochondrial DNA. *Mitochondrion*, 36, 15–20. <https://doi.org/10.1016/j.mito.2016.10.002>
- Chalmers Ay, R. M., & Schapira, A. H. V. (n.d.). *Clinical, biochemical and molecular genetic features of Leber's hereditary optic neuropathy*.
- Chen, P., Sun, Z., Wang, J., Liu, X., Bai, Y., Chen, J., Liu, A., Qiao, F., Chen, Y., Yuan, C., Sha, J., Zhang, J., Xu, L. Q., & Li, J. (2023). Portable nanopore-sequencing technology: Trends in development and applications. In *Frontiers in Microbiology* (Vol. 14). Frontiers Media S.A. <https://doi.org/10.3389/fmicb.2023.1043967>
- Cheng, C., Fei, Z., & Xiao, P. (2023). Methods to improve the accuracy of next-generation sequencing. In *Frontiers in Bioengineering and Biotechnology* (Vol. 11). Frontiers Media S.A. <https://doi.org/10.3389/fbioe.2023.982111>
- Chicco, D., Sichenze, A., & Jurman, G. (2025). A simple guide to the use of Student's t-test, Mann-Whitney U test, Chi-squared test, and Kruskal-Wallis test in biostatistics. *BioData Mining*, 18(1), 56. <https://doi.org/10.1186/S13040-025-00465-6>
- Chiu, C. Y., & Miller, S. A. (2019). Clinical metagenomics. In *Nature Reviews Genetics* (Vol. 20, Issue 6, pp. 341–355). Nature Publishing Group. <https://doi.org/10.1038/s41576-019-0113-7>
- Chong, M., Mohammadi-Shemirani, P., Perrot, N., Nelson, W., Morton, R., Narula, S., Lali, R., Khan, I., Khan, M., Judge, C., Machipisa, T., Cawte, N., O'donnell, M., Pigeyre, M., Akhabir, L., & Paré, G. (2022). *GWAS and ExWAS of blood mitochondrial DNA copy number identifies 71 loci and highlights a potential causal role in dementia*. 11, 70382. <https://doi.org/10.7554/eLife>
- Christodoulou, J. (2000). Genetic defects causing mitochondrial respiratory chain disorders and disease. In *Human Reproduction* (Vol. 15). <http://humrep.oxfordjournals.org/>
- Chrzanowska-Lightowlers, Z. M., & Lightowlers, R. N. (2024). Mitochondrial RNA maturation. In *RNA Biology* (Vol. 21, Issue 1, pp. 28–39). Taylor and Francis Ltd. <https://doi.org/10.1080/15476286.2024.2414157>
- Claes, K. B. M., & De Leeneer, K. (2014). Dealing with pseudogenes in molecular diagnostics in the next-generation sequencing era. *Methods in Molecular Biology*, 1167, 303–315. https://doi.org/10.1007/978-1-4939-0835-6_21

- Cohen, H. W. (2011). P values: use and misuse in medical literature. *American Journal of Hypertension*, 24(1), 18–23. <https://doi.org/10.1038/AJH.2010.205>
- Cohen, I. G., & Adashi, E. Y. (2016). Mitochondrial replacement therapy: The IOM report and its aftermath. In *Nature Reviews Genetics* (Vol. 17, Issue 4, pp. 189–190). Nature Publishing Group. <https://doi.org/10.1038/nrg.2016.21>
- Collu-Marchese, M., Shuen, M., Pauly, M., Saleem, A., & Hood, D. A. (2015). The regulation of mitochondrial transcription factor A (Tfam) expression during skeletal muscle cell differentiation. *Bioscience Reports*, 35(3). <https://doi.org/10.1042/BSR20150073>
- Cottingham, H., Judd, L. M., Harshegyi-Hand, T., Wisniewski, J. A., Blakeway, L. V., Hong, T., Parker, M. H., Peleg, A. Y., Holt, K. E., Hawkey, J., & Macesic, N. (2025). Nanopore sequencing enables highly accurate genotyping and identification of resistance determinants in key nosocomial pathogens. *MedRxiv*, 2025.07.24.25332173. <https://doi.org/10.1101/2025.07.24.25332173>
- Craven, L., Alston, C. L., Taylor, R. W., & Turnbull, D. M. (2025). Recent Advances in Mitochondrial Disease. *Annual Review of Genomics and Human Genetics* Downloaded from *Www.Annualreviews.Org*. Guest, 16, 16. <https://doi.org/10.1146/annurev-genom-091416>
- Cubino, J. P., Cavender-Bares, J., Groffman, P. M., Avolio, M. L., Bratt, A. R., Hall, S. J., Larson, K. L., Lerman, S. B., Narango, D. L., Neill, C., Trammell, T. L. E., Wheeler, M. M., & Hobbie, S. E. (8268). *Taxonomic, phylogenetic, and functional composition and homogenization of residential yard vegetation with contrasting management*.
- Cui, H., Li, F., Chen, D., Wang, G., Truong, C. K., Enns, G. M., Graham, B., Milone, M., Landsverk, M. L., Wang, J., Zhang, W., & Wong, L. J. C. (2013). Comprehensive next-generation sequence analyses of the entire mitochondrial genome reveal new insights into the molecular diagnosis of mitochondrial DNA disorders. *Genetics in Medicine*, 15(5), 388–394. <https://doi.org/10.1038/gim.2012.144>
- De Coster, W., De Rijk, P., De Roeck, A., De Pooter, T., D’Hert, S., Strazisar, M., Slegers, K., & Van Broeckhoven, C. (2019). Structural variants identified by Oxford Nanopore PromethION sequencing of the human genome. *Genome Research*, 29(7), 1178–1187. <https://doi.org/10.1101/gr.244939.118>
- De Coster, W., D’Hert, S., Schultz, D. T., Cruts, M., & Van Broeckhoven, C. (2018). NanoPack: Visualizing and processing long-read sequencing data. *Bioinformatics*, 34(15), 2666–2669. <https://doi.org/10.1093/BIOINFORMATICS/BTY149>
- Delahaye, C., & Nicolas, J. (2021). Sequencing DNA with nanopores: Troubles and biases. *PLoS ONE*, 16(10 October). <https://doi.org/10.1371/journal.pone.0257521>

- Devendra, N., Manahan, K., Geisler, J., & Akula, U. S. (2025). MITOCHONDRIAL GENOMICS IN HEREDITARY DISORDERS: A DIAGNOSTIC PERSPECTIVE. *International Journal of Academic Medicine and Pharmacy*.
<https://doi.org/10.47009/jamp.2025.7.3.131>
- Dhorne-Pollet, S., Barrey, E., & Pollet, N. (2020). A new method for long-read sequencing of animal mitochondrial genomes: application to the identification of equine mitochondrial DNA variants. *BMC Genomics*, 21(1), 1–15.
<https://doi.org/10.1186/S12864-020-07183-9/TABLES/2>
- Díaz Monroy, L. G., & Morales Rivvera, M. A. (2009). *ANÁLISIS ESTADÍSTICO DE DATOS CATEGÓRICOS* (1st ed.).
- Diroma, M. A., Modi, A., Lari, M., Sineo, L., Caramelli, D., & Vai, S. (2021). New Insights Into Mitochondrial DNA Reconstruction and Variant Detection in Ancient Samples. *Frontiers in Genetics*, 12, 619950. <https://doi.org/10.3389/FGENE.2021.619950/XML>
- Dobner, J., Nguyen, T., Pavez-Giani, M. G., Cyganek, L., Distelmaier, F., Krutmann, J., Prigione, A., & Rossi, A. (2024). mtDNA analysis using Mitopore. *Molecular Therapy Methods and Clinical Development*, 32(2).
<https://doi.org/10.1016/j.omtm.2024.101231>
- Doorenspleet, K., Mailli, A. A., van der Hoorn, B. B., Beentjes, K. K., De Backer, A., Derycke, S., Murk, A. J., Reiss, H., & Nijland, R. (2025). Advancing molecular macrobenthos biodiversity monitoring: a comparison between Oxford Nanopore and Illumina based metabarcoding and metagenomics. *PeerJ*, 13(4), e19158.
<https://doi.org/10.7717/PEERJ.19158/SUPP-12>
- D'Souza, A. R., & Minczuk, M. (2018). Mitochondrial transcription and translation: Overview. In *Essays in Biochemistry* (Vol. 62, Issue 3, pp. 309–320). Portland Press Ltd. <https://doi.org/10.1042/EBC20170102>
- Fay, M. P., & Proschan, M. A. (2010). Wilcoxon-Mann-Whitney or t-test? On assumptions for hypothesis tests and multiple interpretations of decision rules. *Statistics Surveys*, 4, 1. <https://doi.org/10.1214/09-SS051>
- Finsterer, J. (2019). Pharmacotherapeutic management of epilepsy in MERRF syndrome. In *Expert Opinion on Pharmacotherapy* (Vol. 20, Issue 10, pp. 1289–1297). Taylor and Francis Ltd. <https://doi.org/10.1080/14656566.2019.1609941>
- Fragasso, A., Schmid, S., & Dekker, C. (2020). Comparing Current Noise in Biological and Solid-State Nanopores. *ACS Nano*, 14(2), 1338–1349.
https://doi.org/10.1021/ACSNANO.9B09353/ASSET/IMAGES/LARGE/NN9B09353_0005.JPEG

- Frascarelli, C., Zanetti, N., Nasca, A., Izzo, R., Lamperti, C., Lamantea, E., Legati, A., & Ghezzi, D. (2023). Nanopore long-read next-generation sequencing for detection of mitochondrial DNA large-scale deletions. *Frontiers in Genetics, 14*, 1089956. <https://doi.org/10.3389/FGENE.2023.1089956/BIBTEX>
- Gabaldón, T. (2025). *Origin and Early Evolution of the Eukaryotic Cell. 58*, 52. <https://doi.org/10.1146/annurev-micro-090817>
- Galizzi, G., & Di Carlo, M. (2023). Mitochondrial DNA and Inflammation in Alzheimer's Disease. In *Current Issues in Molecular Biology* (Vol. 45, Issue 11, pp. 8586–8606). Multidisciplinary Digital Publishing Institute (MDPI). <https://doi.org/10.3390/cimb45110540>
- Garone, C., Tadesse, S., & Hirano, M. (2011). Clinical and genetic spectrum of mitochondrial neurogastrointestinal encephalomyopathy. *Brain, 134*(11), 3326–3332. <https://doi.org/10.1093/brain/awr245>
- Geiger, O., Sanchez-Flores, A., Padilla-Gomez, J., & Degli Esposti, M. (2023). *Multiple approaches of cellular metabolism define the bacterial ancestry of mitochondria*. <https://www.science.org>
- Gensler, S., Weber, K., Schmitt, W. E., Pérez-Martos, A., Antonio Enriquez, J., Montoya, J., & Wiesner, R. J. (2001). Mechanism of mammalian mitochondrial DNA replication: import of mitochondrial transcription factor A into isolated mitochondria stimulates 7S DNA synthesis. In *Nucleic Acids Research* (Vol. 29, Issue 17).
- Giani, A. M., Gallo, G. R., Gianfranceschi, L., & Formenti, G. (2020). Long walk to genomics: History and current approaches to genome sequencing and assembly. In *Computational and Structural Biotechnology Journal* (Vol. 18, pp. 9–19). Elsevier B.V. <https://doi.org/10.1016/j.csbj.2019.11.002>
- gnomAD v3.1 Mitochondrial DNA Variants | gnomAD browser*. (n.d.). Retrieved August 27, 2025, from <https://gnomad.broadinstitute.org/news/2020-11-gnomad-v3-1-mitochondrial-dna-variants/>
- González, M. del M., Ramos, A., Aluja, M. P., & Santos, C. (2020). Sensitivity of mitochondrial DNA heteroplasmy detection using Next Generation Sequencing. *Mitochondrion, 50*, 88–93. <https://doi.org/10.1016/j.mito.2019.10.006>
- Goodwin, S., McPherson, J. D., & McCombie, W. R. (2016). Coming of age: Ten years of next-generation sequencing technologies. In *Nature Reviews Genetics* (Vol. 17, Issue 6, pp. 333–351). Nature Publishing Group. <https://doi.org/10.1038/nrg.2016.49>
- Gorman, G. S., Chinnery, P. F., DiMauro, S., Hirano, M., Koga, Y., McFarland, R., Suomalainen, A., Thorburn, D. R., Zeviani, M., & Turnbull, D. M. (2016).

- Mitochondrial diseases. *Nature Reviews Disease Primers*, 2.
<https://doi.org/10.1038/NRDP.2016.80>
- Gorzynski, J. E., Goenka, S. D., Shafin, K., Jensen, T. D., Fisk, D. G., Grove, M. E., Spiteri, E., Pesout, T., Monlong, J., Baid, G., Bernstein, J. A., Ceresnak, S., Chang, P.-C., Christle, J. W., Chubb, H., Dalton, K. P., Dunn, K., Galalde, D. R., Guillory, J., ... Ashley, E. A. (2022). Ultrarapid Nanopore genome sequencing in a critical care setting. *N. Engl. J. Med.*, 386(7), 700–702. <https://doi.org/10.1056/nejmc2112090>
- Green, M. F., Watson, C. H., Tait, S., He, J., Pavlick, D. C., Frampton, G., Riedel, J., Plichta, J. K., Armstrong, A. J., Previs, R. A., Kauff, N., Strickler, J. H., Datto, M. B., Berchuck, A., & Menendez, C. S. (2023). Concordance Between Genomic Alterations Detected by Tumor and Germline Sequencing: Results from a Tertiary Care Academic Center Molecular Tumor Board. *The Oncologist*, 28(1), 33. <https://doi.org/10.1093/ONCOLO/OYAC164>
- Greig, D. R., Jenkins, C., Gharbia, S., & Dallman, T. J. (2019). Comparison of single-nucleotide variants identified by Illumina and Oxford Nanopore technologies in the context of a potential outbreak of Shiga toxin-producing *Escherichia coli*. *GigaScience*, 8(8). <https://doi.org/10.1093/GIGASCIENCE/GIZ104>,
- Greninger, A. L., Naccache, S. N., Federman, S., Yu, G., Mbala, P., Bres, V., Stryke, D., Bouquet, J., Somasekar, S., Linnen, J. M., Dodd, R., Mulembakani, P., Schneider, B. S., Muyembe-Tamfum, J. J., Stramer, S. L., & Chiu, C. Y. (2015). Rapid metagenomic identification of viral pathogens in clinical samples by real-time nanopore sequencing analysis. *Genome Medicine*, 7(1). <https://doi.org/10.1186/s13073-015-0220-9>
- Hall, M. B., Wick, R. R., Judd, L. M., Nguyen, A. N., Steinig, E. J., Xie, O., Davies, M. R., Seemann, T., Stinear, T. P., & Coin, L. J. (2024). Benchmarking reveals superiority of deep learning variant callers on bacterial nanopore sequence data. *ELife*, 13. <https://doi.org/10.7554/ELIFE.98300.2>
- Heather, J. M., & Chain, B. (2016). The sequence of sequencers: The history of sequencing DNA. In *Genomics* (Vol. 107, Issue 1, pp. 1–8). Academic Press Inc. <https://doi.org/10.1016/j.ygeno.2015.11.003>
- Helal, A. A., Saad, B. T., Saad, M. T., Mosaad, G. S., & Aboshanab, K. M. (2022). Evaluation of the Available Variant Calling Tools for Oxford Nanopore Sequencing in Breast Cancer. *Genes*, 13(9). <https://doi.org/10.3390/GENES13091583>
- Henke, M. T., Prigione, A., & Schuelke, M. (2024). Disease models of Leigh syndrome: From yeast to organoids. In *Journal of Inherited Metabolic Disease* (Vol. 47, Issue 6, pp. 1292–1321). John Wiley and Sons Inc. <https://doi.org/10.1002/jimd.12804>

- Hilt, E. E., & Ferrieri, P. (2022). Next Generation and Other Sequencing Technologies in Diagnostic Microbiology and Infectious Diseases. In *Genes* (Vol. 13, Issue 9). MDPI. <https://doi.org/10.3390/genes13091566>
- Hoenen, T., Groseth, A., Rosenke, K., Fischer, R. J., Hoenen, A., Judson, S. D., Martellaro, C., Falzarano, D., Marzi, A., Squires, R. B., Wollenberg, K. R., De Wit, E., Prescott, J., Safronetz, D., Van Doremalen, N., Bushmaker, T., Feldmann, F., McNally, K., Bolay, F. K., ... Feldmann, H. (2016). Nanopore sequencing as a rapidly deployable Ebola outbreak tool. *Emerging Infectious Diseases*, 22(2), 331–334. <https://doi.org/10.3201/eid2202.151796>
- Holland, M. M., Makova, K. D., & McElhoe, J. A. (2018). Deep-Coverage MPS Analysis of Heteroplasmic Variants within the mtGenome Allows for Frequent Differentiation of Maternal Relatives. *Genes*, 9(3), 124. <https://doi.org/10.3390/GENES9030124>
- Hutchison, C. A. (2007). DNA sequencing: Bench to bedside and beyond. *Nucleic Acids Research*, 35(18), 6227–6237. <https://doi.org/10.1093/nar/gkm688>
- Ibrahim, I., Dominguez-Valentin, M., Segal, B., Zeitouni, A., & da Silva, S. D. (2018). Mitochondrial mutations associated with hearing and balance disorders. *Mutation Research - Fundamental and Molecular Mechanisms of Mutagenesis*, 810, 39–44. <https://doi.org/10.1016/j.mrfmmm.2018.03.003>
- Ip, E. K. K., Troup, M., Xu, C., Winlaw, D. S., Dunwoodie, S. L., & Giannoulatou, E. (2022). Benchmarking the Effectiveness and Accuracy of Multiple Mitochondrial DNA Variant Callers: Practical Implications for Clinical Application. *Frontiers in Genetics*, 13. <https://doi.org/10.3389/fgene.2022.692257>
- Jain, C., Rhie, A., Hansen, N. F., Koren, S., & Phillippy, A. M. (2022). Long-read mapping to repetitive reference sequences using Winnowmap2. *Nature Methods*, 19(6), 705–710. <https://doi.org/10.1038/S41592-022-01457-8>
- Jain, C., Rhie, A., Hansen, N., Koren, S., & Phillippy, A. M. (2020). A long read mapping method for highly repetitive reference sequences. <https://doi.org/10.1101/2020.11.01.363887>
- Jain, C., Rhie, A., Zhang, H., Chu, C., Walenz, B. P., Koren, S., & Phillippy, A. M. (2020). Weighted minimizer sampling improves long read mapping. *Bioinformatics*, 36, 1111–1118. <https://doi.org/10.1093/BIOINFORMATICS/BTAA435>
- Jia, Q., Xu, L., Shen, J., Wei, Y., Xu, H., Shi, J., Jia, Z., Zhao, X., Liu, C., Zhong, Q., Tian, Y., & He, K. (2020). Detecting Rare Variants and Heteroplasmy of Mitochondrial DNA from High-Throughput Sequencing in Patients with Coronary Artery Disease. *Medical Science Monitor : International Medical Journal of Experimental and Clinical Research*, 26, e925401-1. <https://doi.org/10.12659/MSM.925401>

- Kara, B., Arikan, M., Maraş, H., Abacı, N., Çakiris, A., & Üstek, D. (2012). Whole mitochondrial genome analysis of a family with NARP/MILS caused by m.8993T>C mutation in the MT-ATP6 gene. *Molecular Genetics and Metabolism*, *107*(3), 389–393. <https://doi.org/10.1016/j.ymgme.2012.06.013>
- Karaarslan, C. (n.d.). Leber's Hereditary Optic Neuropathy as a Promising Disease for Gene Therapy Development. *Advances in Therapy*, *36*. <https://doi.org/10.6084/m9.figshare.9912680>
- Kassambara A (2023). *rstatix: Pipe-Friendly Framework for Basic Statistical Tests*. R package version 0.7.2, <https://rpkgs.datanovia.com/rstatix/>
- Kasianowicz, J. J., & Bezrukov, S. M. (2016). On “three decades of nanopore sequencing.” In *Nature Biotechnology* (Vol. 34, Issue 5, pp. 481–482). Nature Publishing Group. <https://doi.org/10.1038/nbt.3570>
- Kazan, H. H., Karaca, M., Akan, G., Özgen, Ö., Tuncel, G., Özketen, A. Ç., Balcı, M. C., Körbeyli, H. K., Atalar, F., & Gökçay, G. F. (2024). Oxford nanopore sequencing-based assay for BTD gene screening: Design, clinical validation, and variant frequency assessment in the Turkish population. *Gene*, *928*, 148782. <https://doi.org/10.1016/J.GENE.2024.148782>
- Keogh, M. J., & Chinnery, P. F. (2015). Mitochondrial DNA mutations in neurodegeneration. *Biochimica et Biophysica Acta - Bioenergetics*, *1847*(11), 1401–1411. <https://doi.org/10.1016/j.bbabi.2015.05.015>
- Kim, H.-Y. (2014). Analysis of variance (ANOVA) comparing means of more than two groups. *Restorative Dentistry & Endodontics*, *39*(1), 74. <https://doi.org/10.5395/RDE.2014.39.1.74>
- Kim, M. Y., Cho, S., Lee, J. H., Seo, H. J., & Lee, S. D. (2018). Detection of innate and artificial mitochondrial DNA heteroplasmy by massively parallel sequencing: Considerations for analysis. *Journal of Korean Medical Science*, *33*(52). <https://doi.org/10.3346/jkms.2018.33.e337>
- Koenig, M. K., & Grant, L. (2016). Neuropathy, Ataxia, and Retinitis Pigmentosa. In *Mitochondrial Case Studies: Underlying Mechanisms and Diagnosis* (pp. 69–73). Elsevier Inc. <https://doi.org/10.1016/B978-0-12-800877-5.00009-7>
- Kopinski, P. K., Singh, L. N., Zhang, S., Lott, M. T., & Wallace, D. C. (2021). Mitochondrial DNA variation and cancer. In *Nature Reviews Cancer* (Vol. 21, Issue 7, pp. 431–445). Nature Research. <https://doi.org/10.1038/s41568-021-00358-w>
- Koren, S., Bao, Z., Guarracino, A., Ou, S., Goodwin, S., Jenike, K. M., Lucas, J., McNulty, B., Park, J., Rautiainen, M., Rhie, A., Roelofs, D., Schneiders, H., Vrijenhoek, I.,

- Nijbroek, K., Nordesjo, O., Nurk, S., Vella, M., Lawrence, K. R., ... Phillippy, A. M. (2024). Gapless assembly of complete human and plant chromosomes using only nanopore sequencing. *Genome Research*, 34(11), 1919–1930. <https://doi.org/10.1101/gr.279334.124>
- Korolija, M., Sukser, V., & Vlahoviček, K. (2024). Mitochondrial point heteroplasmy: insights from deep-sequencing of human replicate samples. *BMC Genomics*, 25(1), 1–11. <https://doi.org/10.1186/S12864-024-09963-Z/TABLES/2>
- Kotrys, A. V., & Szczesny, R. J. (2020). Mitochondrial gene expression and beyond—novel aspects of cellular physiology. In *Cells* (Vol. 9, Issue 1). MDPI. <https://doi.org/10.3390/cells9010017>
- Kovaka, S., Hook, P. W., Jenike, K. M., Shivakumar, V., Morina, L. B., Razaghi, R., Timp, W., & Schatz, M. C. (2025). Uncalled4 improves nanopore DNA and RNA modification detection via fast and accurate signal alignment. *Nature Methods*, 22(4), 681–691. <https://doi.org/10.1038/s41592-025-02631-4>
- Król, Z. J., Dobosz, P., Ślubowska, A., & Mroczek, M. (2023). WGS Data Collections: How Do Genomic Databases Transform Medicine? In *International Journal of Molecular Sciences* (Vol. 24, Issue 3). MDPI. <https://doi.org/10.3390/ijms24033031>
- Laricchia, K. M., Lake, N. J., Watts, N. A., Shand, M., Haessly, A., Gauthier, L., Benjamin, D., Banks, E., Soto, J., Garimella, K., Emery, J., Aggregation, G., Consortium, D., Rehm, H. L., Macarthur, D. G., Tiao, G., Lek, M., Mootha, V. K., & Calvo, S. E. (2022). *Mitochondrial DNA variation across 56,434 individuals in gnomAD*. <https://doi.org/10.1101/gr.276013.121>
- Legati, A., Zanetti, N., Nasca, A., Peron, C., Lamperti, C., Lamantea, E., & Ghezzi, D. (2021). Current and New Next-Generation Sequencing Approaches to Study Mitochondrial DNA. *The Journal of Molecular Diagnostics*, 23(6), 732–741. <https://doi.org/10.1016/J.JMOLDX.2021.03.002>
- Legros, F., Malka, F., Frachon, P., Lombès, A., & Rojo, M. (2004). Organization and dynamics of human mitochondrial DNA. *J. Cell Sci.*, 117(13), 2653–2662. <https://doi.org/10.1242/jcs.01134>
- Lesurf, R., Jain, A., Hanafi, N., Mitina, A., Yin, Y., Kolla, V. A., Bouwmeester, J., Papaz, T., Oechslin, E., Yuen, R. K. C., & Mital, S. (2025). Long-read genome sequencing increases genomic yield in congenital heart disease. *MedRxiv*, 2025.05.14.25327523. <https://doi.org/10.1101/2025.05.14.25327523>
- Li, H. (2018). Minimap2: Pairwise alignment for nucleotide sequences. *Bioinformatics*, 34(18), 3094–3100. <https://doi.org/10.1093/BIOINFORMATICS/BTY191>

- Li, H. (2021). New strategies to improve minimap2 alignment accuracy. *Bioinformatics*, 37(23), 4572–4574. <https://doi.org/10.1093/BIOINFORMATICS/BTAB705>
- Liu, L., Li, Y., Li, S., Hu, N., He, Y., Pong, R., Lin, D., Lu, L., & Law, M. (2012). Comparison of Next-Generation Sequencing Systems. *BioMed Research International*, 2012(1), 251364. <https://doi.org/10.1155/2012/251364>
- Logsdon, G. A., Vollger, M. R., & Eichler, E. E. (2020). Long-read human genome sequencing and its applications. In *Nature Reviews Genetics* (Vol. 21, Issue 10, pp. 597–614). Nature Research. <https://doi.org/10.1038/s41576-020-0236-x>
- Lokody, I. (2014). Disease genetics: Mitochondrial variation affects disease risk. In *Nature Reviews Genetics* (Vol. 15, Issue 7, p. 440). Nature Publishing Group. <https://doi.org/10.1038/nrg3760>
- Lorenz, D. A., Sathe, S., Einstein, J. M., & Yeo, G. W. (2020). Direct RNA sequencing enables m6A detection in endogenous transcript isoforms at base-specific resolution. <https://doi.org/10.1261/rna>
- Lorenzoni, P. J., Scola, R. H., Kay, C. S. K., Silvado, C. E. S., & Werneck, L. C. (2014). When should MERRF (myoclonus epilepsy associated with ragged-red fibers) be the diagnosis? *Arquivos de Neuro-Psiquiatria*, 72(10), 803–811. <https://doi.org/10.1590/0004-282X20140124>
- LoTempio, J., Delot, E., & Vilain, E. (2023). Benchmarking long-read genome sequence alignment tools for human genomics applications. *PeerJ*, 11. <https://doi.org/10.7717/peerj.16515>
- Lu, H., Giordano, F., & Ning, Z. (2016). Oxford Nanopore MinION Sequencing and Genome Assembly. *Genomics, Proteomics & Bioinformatics*, 14(5), 265–279. <https://doi.org/10.1016/J.GPB.2016.05.004>
- Lüth, T., Schaake, S., Grünwald, A., May, P., Trinh, J., & Weissensteiner, H. (2022). Benchmarking Low-Frequency Variant Calling With Long-Read Data on Mitochondrial DNA. *Frontiers in Genetics*, 13. <https://doi.org/10.3389/FGENE.2022.887644>
- Ma, H., Marti Gutierrez, N., Morey, R., Van Dyken, C., Kang, E., Hayama, T., Lee, Y., Li, Y., Tippner-Hedges, R., Wolf, D. P., Laurent, L. C., & Mitalipov, S. (2016). Incompatibility between Nuclear and Mitochondrial Genomes Contributes to an Interspecies Reproductive Barrier. *Cell Metabolism*, 24(2), 283–294. <https://doi.org/10.1016/j.cmet.2016.06.012>
- Macken, W. L., Falabella, M., Pizzamiglio, C., Woodward, C. E., Scotchman, E., Chitty, L. S., Polke, J. M., Bugiardini, E., Hanna, M. G., Vandrovcova, J., Chandler, N.,

- Labrum, R., & Pitceathly, R. D. S. (2023). Enhanced mitochondrial genome analysis: bioinformatic and long-read sequencing advances and their diagnostic implications. In *Expert Review of Molecular Diagnostics* (Vol. 23, Issue 9, pp. 797–814). Taylor and Francis Ltd. <https://doi.org/10.1080/14737159.2023.2241365>
- Macken, W. L., Lucassen, A. M., Hanna, M. G., & Pitceathly, R. D. S. (2021). Mitochondrial DNA variants in genomic data: diagnostic uplifts and predictive implications. In *Nature Reviews Genetics* (Vol. 22, Issue 9, pp. 547–548). Nature Research. <https://doi.org/10.1038/s41576-021-00381-5>
- MacKenzie, M., & Argyropoulos, C. (2023). An Introduction to Nanopore Sequencing: Past, Present, and Future Considerations. In *Micromachines* (Vol. 14, Issue 2). MDPI. <https://doi.org/10.3390/mi14020459>
- Mahmud, S., Biswas, S., Afrose, S., Mita, M. A., Robiul Hasan, M., Sharmin Sultana Shimu, M. S., Paul, G. K., Chung, S., Abu Saleh, M., Alshehri, S., Ghoneim, M. M., Alruwaily, M., & Kim, B. (2022). Use of Next-Generation Sequencing for Identifying Mitochondrial Disorders. *Current Issues in Molecular Biology*, 44(3), 1127–1148. <https://doi.org/10.3390/cimb44030074>
- Manzoor, F., Gupta, V., Pinky, L., Wang, Z., Chen, Z., Deng, Y., & Neupane, S. (2025). A Systematic Review of Multimodal Deep Learning and Machine Learning Fusion Techniques for Prostate Cancer Classification. <https://doi.org/10.1101/2025.08.07.25333235>
- marbl/Winnowmap: Long read / genome alignment software*. (n.d.). Retrieved August 27, 2025, from <https://github.com/marbl/Winnowmap>
- Mariner-Llicer, C., Goig, G. A., Zaragoza-Infante, L., Torres-Puente, M., Villamayor, L., Navarro, D., Borrás, R., Chiner-Oms, Á., & Comas, I. (2021). Accuracy of an amplicon-sequencing nanopore approach to identify variants in tuberculosis drug-resistance-associated genes. *Microbial Genomics*, 7(12), 000740. <https://doi.org/10.1099/MGEN.0.000740/CITE/REFWORKS>
- Maude, H., Davidson, M., Charitakis, N., Diaz, L., Bowers, W. H. T., Gradovich, E., Andrew, T., & Huntley, D. (2019). NUMT confounding biases mitochondrial heteroplasmy calls in favor of the reference allele. *Frontiers in Cell and Developmental Biology*, 7(SEP). <https://doi.org/10.3389/fcell.2019.00201>
- Mavraki, E., Labrum, R., Sergeant, K., Alston, C. L., Woodward, C., Smith, C., Knowles, C. V. Y., Patel, Y., Hodsdon, P., Baines, J. P., Blakely, E. L., Polke, J., Taylor, R. W., & Fratter, C. (2023). Genetic testing for mitochondrial disease: the United Kingdom best practice guidelines. *European Journal of Human Genetics*, 31(2), 148–163. <https://doi.org/10.1038/s41431-022-01249-w>

- McCombie, W. R., & McPherson, J. D. (2019). Future promises and concerns of ubiquitous next-generation sequencing. *Cold Spring Harbor Perspectives in Medicine*, 9(9). <https://doi.org/10.1101/cshperspect.a025783>
- McHugh, M. L. (2011). Multiple comparison analysis testing in ANOVA. *Biochimica Medica*, 21(3), 203–209. <https://doi.org/10.11613/BM.2011.029>
- Mishra, P., Singh, U., Pandey, C. M., Mishra, P., & Pandey, G. (2019). Application of Student's t-test, Analysis of Variance, and Covariance. *Annals of Cardiac Anaesthesia*, 22(4), 407. https://doi.org/10.4103/ACA.ACA_94_19
- Mongan, A. E., Tuda, J. S. B., & Runtuwene, L. R. (2020). Portable sequencer in the fight against infectious disease. In *Journal of Human Genetics* (Vol. 65, Issue 1, pp. 35–40). Springer Nature. <https://doi.org/10.1038/s10038-019-0675-4>
- Moran, J. C., Brivanlou, A., Brischigliaro, M., Fontanesi, F., Rouskin, S., & Barrientos, A. (2024). The human mitochondrial mRNA structurome reveals mechanisms of gene expression. *Science*, 385(6706). <https://doi.org/10.1126/science.adm9238>
- Moura, J. P., Oliveira, P. J., & Urbano, A. M. (2025). Mitochondria: An overview of their origin, genome, architecture, and dynamics. In *Biochimica et Biophysica Acta - Molecular Basis of Disease* (Vol. 1871, Issue 5). Elsevier B.V. <https://doi.org/10.1016/j.bbadis.2025.167803>
- Moustakli, E., Christopoulos, P., Potiris, A., Zikopoulos, A., Mavrogianni, D., Karampas, G., Kathopoulis, N., Anagnostaki, I., Domali, E., Tzallas, A. T., Drakakis, P., & Stavros, S. (2025). Long-Read Sequencing and Structural Variant Detection: Unlocking the Hidden Genome in Rare Genetic Disorders. In *Diagnostics* (Vol. 15, Issue 14). Multidisciplinary Digital Publishing Institute (MDPI). <https://doi.org/10.3390/diagnostics15141803>
- Muir, P., Li, S., Lou, S., Wang, D., Spakowicz, D. J., Salichos, L., Zhang, J., Weinstock, G. M., Isaacs, F., Rozowsky, J., & Gerstein, M. (2016). The real cost of sequencing: Scaling computation to keep pace with data generation. *Genome Biology*, 17(1). <https://doi.org/10.1186/s13059-016-0917-0>
- nanoporetech/medaka: Sequence correction provided by ONT Research*. (n.d.). Retrieved August 27, 2025, from <https://github.com/nanoporetech/medaka>
- Needs, H. I., Protasoni, M., Henley, J. M., Prudent, J., Collinson, I., & Pereira, G. C. (2021). Interplay between mitochondrial protein import and respiratory complexes assembly in neuronal health and degeneration. *Life*, 11(5). <https://doi.org/10.3390/life11050432>

- Neupert, W. (2016). Mitochondrial Gene Expression: A Playground of Evolutionary Tinkering. In *Annual Review of Biochemistry* (Vol. 85, pp. 65–76). Annual Reviews Inc. <https://doi.org/10.1146/annurev-biochem-011116-110824>
- Nissanka, N., & Moraes, C. T. (2020). Mitochondrial DNA heteroplasmy in disease and targeted nuclease-based therapeutic approaches. *EMBO Reports*, 21(3). <https://doi.org/10.15252/embr.201949612>
- Nolte, D., Kang, J. S., Hofmann, A., Schwaab, E., Krämer, H. H., & Müller, U. (2021). Mutations in MT-ATP6 are a frequent cause of adult-onset spinocerebellar ataxia. *Journal of Neurology*, 268(12), 4866–4873. <https://doi.org/10.1007/s00415-021-10607-5>
- Non-Metric Multidimensional Scaling (NMDS) in Microbial Sequencing Data Analysis: Introduction, Application, and Comparison - CD Genomics*. (n.d.). Retrieved September 13, 2025, from <https://www.cd-genomics.com/microbioseq/non-metric-multidimensional-scaling-nmnds-in-microbial-sequencing-data-analysis-introduction-application-and-comparison.html>
- Nyaga, D. M., Tsai, P., Gebbie, C., Phua, H. H., Yap, P., Le Quesne Stabej, P., Farrow, S., Rong, J., Toldi, G., Thorstensen, E., Stark, Z., Lunke, S., Gamet, K., Van Dyk, J., Greenslade, M., & O'Sullivan, J. M. (2024). Benchmarking nanopore sequencing and rapid genomics feasibility: validation at a quaternary hospital in New Zealand. *Npj Genomic Medicine* 2024 9:1, 9(1), 1–13. <https://doi.org/10.1038/s41525-024-00445-5>
- Oksanen, J., Blanchet, F. G., Friendly, M., Kindt, R., Legendre, P., Mcglinn, D., Minchin, P. R., O'hara, R. B., Simpson, G. L., Solymos, P., Henry, M., Stevens, H., Szoecs, E., & Maintainer, H. W. (2020). *Package "vegan" Title Community Ecology Package Version 2.5-7*.
- Osada, M., Yasugi, M., Yamamoto, H., Ito, A., & Fukamachi, S. (2024). Individual Identification of Medaka, a Small Freshwater Fish, from the Dorsal Side Using Artificial Intelligence. *Hydrobiology* 2024, Vol. 3, Pages 119-133, 3(2), 119–133. <https://doi.org/10.3390/HYDROBIOLOGY3020009>
- Ott, M., Amunts, A., & Brown, A. (2016). Organization and Regulation of Mitochondrial Protein Synthesis. *Annual Review of Biochemistry*, 85, 77–101. <https://doi.org/10.1146/annurev-biochem-060815-014334>
- Otten, A. B. C., Sallevelt, S. C. E. H., Carling, P. J., Dreesen, J. C. F. M., Drüsedau, M., Spierts, S., Paulussen, A. D. C., De Die-Smulders, C. E. M., Herbert, M., Chinnery, P. F., Samuels, D. C., Lindsey, P., & Smeets, H. J. M. (2018). Mutation-specific effects in germline transmission of pathogenic mtDNA variants. *Human Reproduction*, 33(7), 1331–1341. <https://doi.org/10.1093/humrep/dey114>

- Oxford Nanopore. (2025). *Dorado Documentation*. <https://software-docs.nanoporetech.com/dorado/latest/>
- Ozir, M. A., Nordin, M. H., Hashim, S. E., Adzahar, S., Ahmad, M. A., Ng, K. S., & Wan Hitam, W.-H. (2024). Leber Hereditary Optic Neuropathy With Significant Visual Recovery: An MT-ND6 Mutation in a Malay Patient. *Cureus*. <https://doi.org/10.7759/cureus.71210>
- Pagano, L., Lagrotteria, D., Facconi, A., Saraceno, C., Longobardi, A., Bellini, S., Ingannato, A., Bagnoli, S., Ducci, T., Mingrino, A., Laganà, V., Paparazzo, E., Borroni, B., Maletta, R., Nacmias, B., Montesanto, A., & Ghidoni, R. (2025). Evaluation of Illumina and Oxford Nanopore Sequencing for the Study of DNA Methylation in Alzheimer's Disease and Frontotemporal Dementia. In *International Journal of Molecular Sciences* (Vol. 26, Issue 9). Multidisciplinary Digital Publishing Institute (MDPI). <https://doi.org/10.3390/ijms26094198>
- Pagès-Gallego, M., & de Ridder, J. (2023). Comprehensive benchmark and architectural analysis of deep learning models for nanopore sequencing basecalling. *Genome Biology*, 24(1). <https://doi.org/10.1186/s13059-023-02903-2>
- Parakatselaki, M. E., & Ladoukakis, E. D. (2021). mtDNA heteroplasmy: Origin, detection, significance, and evolutionary consequences. In *Life* (Vol. 11, Issue 7). MDPI AG. <https://doi.org/10.3390/life11070633>
- Parikh, S., Goldstein, A., Koenig, M. K., Scaglia, F., Enns, G. M., Saneto, R., Anselm, I., Cohen, B. H., Falk, M. J., Greene, C., Gropman, A. L., Haas, R., Hirano, M., Morgan, P., Sims, K., Tarnopolsky, M., Van Hove, J. L. K., Wolfe, L., & DiMauro, S. (2015). Diagnosis and management of mitochondrial disease: A consensus statement from the Mitochondrial Medicine Society. In *Genetics in Medicine* (Vol. 17, Issue 9, pp. 689–701). Nature Publishing Group. <https://doi.org/10.1038/gim.2014.177>
- Patop, I. L., & Kadener, S. (2018). circRNAs in Cancer. In *Current Opinion in Genetics and Development* (Vol. 48, pp. 121–127). Elsevier Ltd. <https://doi.org/10.1016/j.gde.2017.11.007>
- Pei, X. M., Yeung, M. H. Y., Wong, A. N. N., Tsang, H. F., Yu, A. C. S., Yim, A. K. Y., & Wong, S. C. C. (2023). Targeted Sequencing Approach and Its Clinical Applications for the Molecular Diagnosis of Human Diseases. In *Cells* (Vol. 12, Issue 3). MDPI. <https://doi.org/10.3390/cells12030493>
- Peresini, P., Boza, V., Brejova, B., & Vinar, T. (2021). Nanopore base calling on the edge. *Bioinformatics*, 37(24), 4661–4667. <https://doi.org/10.1093/bioinformatics/btab528>
- Petersen, L. M., Martin, I. W., Moschetti, W. E., Kershaw, C. M., & Tsongalis, G. J. (2019). Third-Generation Sequencing in the Clinical Laboratory: Exploring the

- Advantages and Challenges of Nanopore Sequencing. *Journal of Clinical Microbiology*, 58(1). <https://doi.org/10.1128/JCM.01315-19>
- philres/ngmlr: NGMLR is a long-read mapper designed to align PacBio or Oxford Nanopore (standard and ultra-long) to a reference genome with a focus on reads that span structural variations.* (n.d.). Retrieved August 27, 2025, from <https://github.com/philres/ngmlr>
- Poplin, R., Chang, P. C., Alexander, D., Schwartz, S., Colthurst, T., Ku, A., Newburger, D., Dijamco, J., Nguyen, N., Afshar, P. T., Gross, S. S., Dorfman, L., McLean, C. Y., & Depristo, M. A. (2018). A universal snp and small-indel variant caller using deep neural networks. *Nature Biotechnology*, 36(10), 983. <https://doi.org/10.1038/NBT.4235;TECHMETA=23,45;SUBJMETA=114,1305,212,2785,61,631;KWRD=GENOME+INFORMATICS,GENOMICS,MACHINE+LEARNING>
- Prodanov, T., & Bansal, V. (2020). Sensitive alignment using paralogous sequence variants improves long-read mapping and variant calling in segmental duplications. *Nucleic Acids Research*, 48(19), e114–e114. <https://doi.org/10.1093/NAR/GKAA829>
- Rackham, O., & Filipovska, A. (2022). Organization and expression of the mammalian mitochondrial genome. In *Nature Reviews Genetics* (Vol. 23, Issue 10, pp. 606–623). Nature Research. <https://doi.org/10.1038/s41576-022-00480-x>
- Real, R., & Vargas, J. M. (1996). The probabilistic basis of Jaccard's index of similarity. *Systematic Biology*, 45(3), 380–385. <https://doi.org/10.1093/SYSBIO/45.3.380>
- Ren, L., Duan, X., Dong, L., Zhang, R., Yang, J., Gao, Y., Peng, R., Hou, W., Liu, Y., Li, J., Yu, Y., Zhang, N., Shang, J., Liang, F., Wang, D., Chen, H., Sun, L., Hao, L., Scherer, A., ... Shi, L. (2023). Quartet DNA reference materials and datasets for comprehensively evaluating germline variant calling performance. *Genome Biology*, 24(1), 1–31. <https://doi.org/10.1186/S13059-023-03109-2/TABLES/2>
- Rendón-Macias, M. E., Zarco-Villavicencio, I. S., & Villasís-Keever, M. Á. (2021). Métodos estadísticos para el análisis del tamaño del efecto. *Revista Alergia México*, 68(2), 128–136. <https://doi.org/10.29262/ram.v658i2.949>
- Rohlin, A., Wernersson, J., Engwall, Y., Wiklund, L., Björk, J., & Nordling, M. (2009). Parallel sequencing used in detection of mosaic mutations: Comparison with four diagnostic DNA screening techniques. *Human Mutation*, 30(6), 1012–1020. <https://doi.org/10.1002/humu.20980>
- Romagnoli, S., Bartalucci, N., & Vannucchi, A. M. (2023). Resolving complex structural variants via nanopore sequencing. In *Frontiers in Genetics* (Vol. 14). Frontiers Media SA. <https://doi.org/10.3389/fgene.2023.1213917>

- Rossmann, M. P., Dubois, S. M., Agarwal, S., & Zon, L. I. (2021). Mitochondrial function in development and disease. In *DMM Disease Models and Mechanisms* (Vol. 14, Issue 6). Company of Biologists Ltd. <https://doi.org/10.1242/dmm.048912>
- Sahlin, K., Baudeau, T., Cazaux, B., & Marchet, C. (2023). A survey of mapping algorithms in the long-reads era. *Genome Biology*, *24*(1), 1–23. <https://doi.org/10.1186/S13059-023-02972-3/FIGURES/6>
- Salk, J. J., Schmitt, M. W., & Loeb, L. A. (2018). Enhancing the accuracy of next-generation sequencing for detecting rare and subclonal mutations. In *Nature Reviews Genetics* (Vol. 19, Issue 5, pp. 269–285). Nature Publishing Group. <https://doi.org/10.1038/nrg.2017.117>
- Santibanez-Koref, M., Griffin, H., Turnbull, D. M., Chinnery, P. F., Herbert, M., & Hudson, G. (2019). Assessing mitochondrial heteroplasmy using next generation sequencing: A note of caution. *Mitochondrion*, *46*, 302–306. <https://doi.org/10.1016/j.mito.2018.08.003>
- Santos, R., Lee, H., Williams, A., Baffour-Kyei, A., Lee, S. H., Troakes, C., Al-Chalabi, A., Breen, G., & Iacoangeli, A. (2025). Investigating the Performance of Oxford Nanopore Long-Read Sequencing with Respect to Illumina Microarrays and Short-Read Sequencing. *International Journal of Molecular Sciences*, *26*(10). <https://doi.org/10.3390/ijms26104492>
- Satam, H., Joshi, K., Mangrolia, U., Waghoo, S., Zaidi, G., Rawool, S., Thakare, R. P., Banday, S., Mishra, A. K., Das, G., & Malonia, S. K. (2023). Next-Generation Sequencing Technology: Current Trends and Advancements. In *Biology* (Vol. 12, Issue 7). Multidisciplinary Digital Publishing Institute (MDPI). <https://doi.org/10.3390/biology12070997>
- Schon, E. A., Dimauro, S., & Hirano, M. (2012). Human mitochondrial DNA: Roles of inherited and somatic mutations. In *Nature Reviews Genetics* (Vol. 13, Issue 12, pp. 878–890). <https://doi.org/10.1038/nrg3275>
- Schon, K. R., Horvath, R., Wei, W., Calabrese, C., Tucci, A., Ibañez, K., Ratnaik, T., Pitceathly, R. D. S., Bugiardini, E., Quinlivan, R., Hanna, M. G., Clement, E., Ashton, E., Sayer, J. A., Brennan, P., Josifova, D., Izatt, L., Fratter, C., Nesbitt, V., ... Wood, S. M. (2021). Use of whole genome sequencing to determine genetic basis of suspected mitochondrial disorders: Cohort study. *BMJ*, *375*. <https://doi.org/10.1136/bmj-2021-066288>
- Sedlazeck, F. J., Rescheneder, P., Smolka, M., Fang, H., Nattestad, M., Von Haeseler, A., & Schatz, M. C. (2018). Accurate detection of complex structural variations using single molecule sequencing. *Nature Methods*, *15*(6), 461. <https://doi.org/10.1038/S41592-018-0001-7>

- Sendra, L., García-Mares, A., Herrero, M. J., & Aliño, S. F. (2021). Mitochondrial dna replacement techniques to prevent human mitochondrial diseases. In *International Journal of Molecular Sciences* (Vol. 22, Issue 2, pp. 1–23). MDPI AG. <https://doi.org/10.3390/ijms22020551>
- Shendure, J., Balasubramanian, S., Church, G. M., Gilbert, W., Rogers, J., Schloss, J. A., & Waterston, R. H. (2017). DNA sequencing at 40: Past, present and future. *Nature*, *550*(7676). <https://doi.org/10.1038/nature24286>
- Shendure, J., & Ji, H. (2008). Next-generation DNA sequencing. *Nature Biotechnology* *2008* *26*:10, *26*(10), 1135–1145. <https://doi.org/10.1038/nbt1486>
- Shoffner, J. M., Lo&t, M. T., Lezza, A. M. S., Seibel, P., Ballinger, S. W., & Wallace'tll, D. C. (1990). *Myoclonic Epilepsy and Ragged-Red Fiber Disease (MERRF) Is Associated with a Mitochondrial DNA tRNALys Mutation* (Vol. 61).
- Slapnik, B., Šket, R., Črepinšek, K., Tesovnik, T., Bizjan, B. J., & Kovač, J. (2024). The quality and detection limits of mitochondrial heteroplasmy by long read nanopore sequencing. *Scientific Reports*, *14*(1). <https://doi.org/10.1038/s41598-024-78270-0>
- Son, J. S., Seo, G. H., Kim, Y. M., Kim, G. H., Jin, H. K., Bae, J. S., Im, H. J., Yoo, H. W., Lee, B. H., & Alqasim, A. M. Z. (2022). Clinical and genetic features of four patients with Pearson syndrome: An observational study. *Medicine (United States)*, *101*(5), 28793. <https://doi.org/10.1097/MD.00000000000028793>
- Sood, R., & Singh, V. (2024). *DNA Sequencing Technologies in Accelerating Molecular Breeding*. <https://doi.org/10.5772/intechopen.1006294>
- Stefan, C. P., Hall, A. T., Graham, A. S., & Minogue, T. D. (2022). Comparison of Illumina and Oxford Nanopore Sequencing Technologies for Pathogen Detection from Clinical Matrices Using Molecular Inversion Probes. *The Journal of Molecular Diagnostics : JMD*, *24*(4), 395. <https://doi.org/10.1016/J.JMOLDX.2021.12.005>
- Stefano, G. B., Bjenning, C., Wang, F., Wang, N., & Kream, R. M. (2017). Mitochondrial heteroplasmy. *Advances in Experimental Medicine and Biology*, *982*, 577–594. https://doi.org/10.1007/978-3-319-55330-6_30
- Stendel, C., Neuhofer, C., Floride, E., Yuqing, S., Ganetzky, R. D., Park, J., Freisinger, P., Kornblum, C., Kleinle, S., Schöls, L., Distelmaier, F., Stettner, G. M., Büchner, B., Falk, M. J., Mayr, J. A., Synofzik, M., Abicht, A., Haack, T. B., Prokisch, H., ... Klopstock, T. (2020). Delineating MT-ATP6 -associated disease: From isolated neuropathy to early onset neurodegeneration. *Neurology: Genetics*, *6*(1). <https://doi.org/10.1212/NXG.0000000000000393>

- Stewart, J. B., & Chinnery, P. F. (2015). The dynamics of mitochondrial DNA heteroplasmy: Implications for human health and disease. In *Nature Reviews Genetics* (Vol. 16, Issue 9, pp. 530–542). Nature Publishing Group. <https://doi.org/10.1038/nrg3966>
- Su, J., Zheng, Z., Ahmed, S. S., Lam, T.-W., & Luo, R. (2022). Clair3-Trio: high-performance Nanopore long-read variant calling in family trios with Trio-to-Trio deep neural networks. *BioRxiv*, 2022.05.03.490460. <https://doi.org/10.1101/2022.05.03.490460>
- Szakállas, N., Barták, B. K., Valcz, G., Nagy, Z. B., Takács, I., & Molnár, B. (2024). Can long-read sequencing tackle the barriers, which the next-generation could not? A review. In *Pathology and Oncology Research* (Vol. 30). Frontiers Media SA. <https://doi.org/10.3389/pore.2024.1611676>
- Tachibana, M., Kuno, T., & Yaegashi, N. (2018). Mitochondrial replacement therapy and assisted reproductive technology: A paradigm shift toward treatment of genetic diseases in gametes or in early embryos. In *Reproductive Medicine and Biology* (Vol. 17, Issue 4, pp. 421–433). John Wiley and Sons Ltd. <https://doi.org/10.1002/rmb2.12230>
- Tafazoli, A., Hemmati, M., Rafigh, M., Alimardani, M., Khaghani, F., Korostyński, M., & Karnes, J. H. (2025). Leveraging long-read sequencing technologies for pharmacogenomic testing: applications, analytical strategies, challenges, and future perspectives. In *Frontiers in Genetics* (Vol. 16). Frontiers Media SA. <https://doi.org/10.3389/fgene.2025.1435416>
- Taguchi, Y. H., & Oono, Y. (2005). Relational patterns of gene expression via non-metric multidimensional scaling analysis. *Bioinformatics (Oxford, England)*, 21(6), 730–740. <https://doi.org/10.1093/BIOINFORMATICS/BTI067>
- Takeda, Y., Hyslop, L., Choudhary, M., Robertson, F., Pyle, A., Wilson, I., Santibanez-Koref, M., Turnbull, D., Herbert, M., & Hudson, G. (2023). Feasibility and impact of haplogroup matching for mitochondrial replacement treatment. *EMBO Reports*, 24(10). <https://doi.org/10.15252/embr.202154540>
- Taniguchi, E., Satoh, K., Ohkubo, M., Ue, S., Matsuhira, H., Kuroda, Y., Kubo, T., & Kitazaki, K. (2023). Nuclear DNA segments homologous to mitochondrial DNA are obstacles for detecting heteroplasmy in sugar beet (*Beta vulgaris* L.). *PLoS ONE*, 18(8 August). <https://doi.org/10.1371/journal.pone.0285430>
- Tavares, G. M., Missaggia, B. O., Cadore, N. A., Sbruzzi, R. C., Feira, M. F., Giudicelli, G. C., de Oliveira Fam, B. S., Rodrigues, M., Dorn, M., Hünemeier, T., Vianna, F. S. L., & Bortolini, M. C. (2025). Mitochondrial haplogroup A2 is associated with increased COVID-19 mortality in an admixed Brazilian population. *Scientific Reports*, 15(1), 1–

13. <https://doi.org/10.1038/S41598-025-03578-4>;SUBJMETA=181,208,631,692,699;KWRD=DISEASES,EVOLUTION,GENETICS
- Teng, H., Stoiber, M., Bar-Joseph, Z., & Kingsford, C. (2024). *Detecting m6A RNA modification from nanopore sequencing using a semi-supervised learning framework*. <https://doi.org/10.1101/2024.01.06.574484>
- Thäle, A., Scholle, L., & Zierz, S. (2020). Mitochondrial diseases. In *Medizinische Monatsschrift für Pharmazeuten* (Vol. 43, Issue 7, pp. 252–266). Deutscher Apotheker Verlag. <https://doi.org/10.1177/0883073814537379>
- Tshiabuila, D., Giandhari, J., Pillay, S., Ramphal, U., Ramphal, Y., Maharaj, A., Anyaneji, U. J., Naidoo, Y., Tegally, H., San, E. J., Wilkinson, E., Lessells, R. J., & de Oliveira, T. (2022). Comparison of SARS-CoV-2 sequencing using the ONT GridION and the Illumina MiSeq. *BMC Genomics*, *23*(1), 1–17. <https://doi.org/10.1186/S12864-022-08541-5/FIGURES/8>
- Tüns, A. I., Hartmann, T., Magin, S., González, R. C., Henssen, A. G., Rahmann, S., Schramm, A., & Köster, J. (2022). Detection and Validation of Circular DNA Fragments Using Nanopore Sequencing. *Frontiers in Genetics*, *13*. <https://doi.org/10.3389/FGENE.2022.867018/FULL>
- Vandeputte, J., Van Heetvelde, M., Van Cauwenbergh, C., Seneca, S., De Baere, E., Leroy, B. P., & De Zaeytijd, J. (2021). Mild Leber hereditary optic neuropathy (LHON) in a Western European family due to the rare Asian m.14502T>C variant in the MT-ND6 gene. *Ophthalmic Genetics*, *42*(4), 440–445. <https://doi.org/10.1080/13816810.2021.1913611>
- Vereecke, N., Bokma, J., Haesebrouck, F., Nauwynck, H., Boyen, F., Pardon, B., & Theuns, S. (2020). High quality genome assemblies of *Mycoplasma bovis* using a taxon-specific Bonito basecaller for MinION and Flongle long-read nanopore sequencing. *BMC Bioinformatics*, *21*(1), 1–16. <https://doi.org/10.1186/S12859-020-03856-0/FIGURES/3>
- Veselovsky, V., Romanov, M., Zoruk, P., Larin, A., Babenko, V., Morozov, M., Strokach, A., Zakharevich, N., Khamidova, S., Danilova, A., Vatlin, A., Pavshintsev, V., Chenguang, F., Tsybizov, D., Mitkin, N., Galanova, O., Zakharenko, A., Golokhvast, K., & Klimina, K. (2025). Comparative evaluation of sequencing platforms: Pacific Biosciences, Oxford Nanopore Technologies, and Illumina for 16S rRNA-based soil microbiome profiling. *Frontiers in Microbiology*, *16*, 1633360. <https://doi.org/10.3389/FMICB.2025.1633360/BIBTEX>
- Wagner, G. E., Dabernig-Heinz, J., Lipp, M., Cabal, A., Simantzik, J., Kohl, M., Scheiber, M., Lichtenegger, S., Ehricht, R., Leitner, E., Ruppitsch, W., & Steinmetz, I. (2023). Real-Time Nanopore Q20+ Sequencing Enables Extremely Fast and Accurate Core

- Genome MLST Typing and Democratizes Access to High-Resolution Bacterial Pathogen Surveillance. *Journal of Clinical Microbiology*, 61(4).
<https://doi.org/10.1128/jcm.01631-22>
- Wallace, D. C., & Chalkia, D. (2013). Mitochondrial DNA genetics and the heteroplasmy conundrum in evolution and disease. In *Cold Spring Harbor perspectives in biology* (Vol. 5, Issue 11). <https://doi.org/10.1101/cshperspect.a021220>
- Wang, J., Balciuniene, J., Diaz-Miranda, M. A., McCormick, E. M., Aref-Eshghi, E., Muir, A. M., Cao, K., Troiani, J., Moseley, A., Fan, Z., Zolkipli-Cunningham, Z., Goldstein, A., Ganetzky, R. D., Muraresku, C. C., Peterson, J. T., Spinner, N. B., Wallace, D. C., Dulik, M. C., & Falk, M. J. (2022). Advanced approach for comprehensive mtDNA genome testing in mitochondrial disease. *Molecular Genetics and Metabolism*, 135(1), 93–101. <https://doi.org/10.1016/j.ymgme.2021.12.006>
- Wang, X., Stelzer-Braid, S., Scotch, M., & Rawlinson, W. D. (2022). Detection of respiratory viruses directly from clinical samples using next-generation sequencing: A literature review of recent advances and potential for routine clinical use. In *Reviews in Medical Virology* (Vol. 32, Issue 5). John Wiley and Sons Ltd. <https://doi.org/10.1002/rmv.2375>
- Wang, Y., Zhao, Y., Bollas, A., Wang, Y., & Au, K. F. (2021). Nanopore sequencing technology, bioinformatics and applications. In *Nature Biotechnology* (Vol. 39, Issue 11, pp. 1348–1365). Nature Research. <https://doi.org/10.1038/s41587-021-01108-x>
- Wei, W., Schon, K. R., Elgar, G., Orioli, A., Tanguy, M., Giess, A., Tischkowitz, M., Caulfield, M. J., & Chinnery, P. F. (2022). Nuclear-embedded mitochondrial DNA sequences in 66,083 human genomes. *Nature*, 611(7934), 105–114. <https://doi.org/10.1038/s41586-022-05288-7>
- Weissensteiner, H., Forer, L., Fuchsberger, C., Schöpf, B., Kloss-Brandstätter, A., Specht, G., Kronenberg, F., & Schönherr, S. (2016). mtDNA-Server: next-generation sequencing data analysis of human mitochondrial DNA in the cloud. *Nucleic Acids Research*, 44(W1), W64–W69. <https://doi.org/10.1093/NAR/GKW247>,
- Weissensteiner, H., Forer, L., Kronenberg, F., & Schönherr, S. (2024). mtDNA-Server 2: advancing mitochondrial DNA analysis through highly parallelized data processing and interactive analytics. *Nucleic Acids Research*, 52(W1), W102–W107. <https://doi.org/10.1093/nar/gkae296>
- Wen, H., Deng, H., Li, B., Chen, J., Zhu, J., Zhang, X., Yoshida, S., & Zhou, Y. (2025). Mitochondrial diseases: from molecular mechanisms to therapeutic advances. In *Signal Transduction and Targeted Therapy* (Vol. 10, Issue 1). Springer Nature. <https://doi.org/10.1038/s41392-024-02044-3>

- Wick, R. R., Judd, L. M., & Holt, K. E. (2019). Performance of neural network basecalling tools for Oxford Nanopore sequencing. *Genome Biology*, 20(1), 1–10. <https://doi.org/10.1186/S13059-019-1727-Y/FIGURES/4>
- Xiao, T., & Zhou, W. (2020). The third generation sequencing: The advanced approach to genetic diseases. In *Translational Pediatrics* (Vol. 9, Issue 2, pp. 163–173). AME Publishing Company. <https://doi.org/10.21037/TP.2020.03.06>
- Xie, S., Leung, A. W. S., Zheng, Z., Zhang, D., Xiao, C., Luo, R., Luo, M., & Zhang, S. (2021). Applications and potentials of nanopore sequencing in the (epi)genome and (epi)transcriptome era. In *Innovation* (Vol. 2, Issue 4). Cell Press. <https://doi.org/10.1016/j.xinn.2021.100153>
- Xuan, J., Yu, Y., Qing, T., Guo, L., & Shi, L. (2013). Next-generation sequencing in the clinic: Promises and challenges. In *Cancer Letters* (Vol. 340, Issue 2, pp. 284–295). Elsevier Ireland Ltd. <https://doi.org/10.1016/j.canlet.2012.11.025>
- Yadak, R., Smitt, P. S., van Gisbergen, M. W., van Til, N. P., & de Coo, I. F. M. (2017). Mitochondrial neurogastrointestinal encephalomyopathy caused by thymidine phosphorylase enzyme deficiency: From pathogenesis to emerging therapeutic options. In *Frontiers in Cellular Neuroscience* (Vol. 11). Frontiers Research Foundation. <https://doi.org/10.3389/fncel.2017.00031>
- Yin, Y., Butler, C., & Zhang, Q. (2021). Challenges in the application of NGS in the clinical laboratory. *Human Immunology*, 82(11), 812–819. <https://doi.org/10.1016/j.humimm.2021.03.011>
- Yoshimi, A., Ishikawa, K., Niemeyer, C., & Grünert, S. C. (2022). Pearson syndrome: a multisystem mitochondrial disease with bone marrow failure. In *Orphanet Journal of Rare Diseases* (Vol. 17, Issue 1). BioMed Central Ltd. <https://doi.org/10.1186/s13023-022-02538-9>
- Zeviani, M., Moraes, C. T., DiMauro, S., Nakase, H., Bonilla, E., Schon, E. A., & Rowland, L. P. (2011). Deletions of mitochondrial DNA in Kearns-Sayre syndrome. *Neurology*, 77(11), 1090–1090. <https://doi.org/10.1212/01.wnl.0000405387.42287.b2>
- Zhang, T., Li, H., Jiang, M., Hou, H., Gao, Y., Li, Y., Wang, F., Wang, J., Peng, K., & Liu, Y. X. (2024). Nanopore sequencing: flourishing in its teenage years. In *Journal of Genetics and Genomics* (Vol. 51, Issue 12, pp. 1361–1374). Institute of Genetics and Developmental Biology. <https://doi.org/10.1016/j.jgg.2024.09.007>
- Zhang, Y. Z., Akdemir, A., Tremmel, G., Imoto, S., Miyano, S., Shibuya, T., & Yamaguchi, R. (2020). Nanopore basecalling from a perspective of instance segmentation. *BMC Bioinformatics*, 21. <https://doi.org/10.1186/s12859-020-3459-0>

- Zheng, P., Zhou, C., Ding, Y., Liu, B., Lu, L., Zhu, F., & Duan, S. (2023). Nanopore sequencing technology and its applications. *MedComm*, 4(4), e316. <https://doi.org/10.1002/MCO2.316>
- Zheng, Z., Li, S., Su, J., Leung, A. W. S., Lam, T. W., & Luo, R. (2022). Symphonizing pileup and full-alignment for deep learning-based long-read variant calling. *Nature Computational Science*, 2(12), 797–803. <https://doi.org/10.1038/S43588-022-00387-X>;SUBJMETA=208,514,631,649,726;KWRD=GENETIC+VARIATION,SEQUENCING
- Zheng, Z., Yu, X., Chen, L., Lee, Y.-L., Xin, C., Wong, A. O. K., Jain, M., Kesharwani, R. K., Sedlazeck, F. J., & Luo, R. (2025). Clair3-RNA: A deep learning-based small variant caller for long-read RNA sequencing data. *BioRxiv*, 2024.11.17.624050. <https://doi.org/10.1101/2024.11.17.624050>
- Zong, W. X., Rabinowitz, J. D., & White, E. (2016). Mitochondria and Cancer. *Molecular Cell*, 61(5), 667–676. <https://doi.org/10.1016/j.molcel.2016.02.011>