

Secuenciación masiva: paneles, Exomas, Genomas

Dra. Teresa Aravena

Clínica INDISA

Hospital Dr. Exequiel González Cortés

Introducción

Secuenciación masiva (NGS): estudio de genomas completos y exomas, dirigidos hasta grupo de genes o un solo gen.

NGS permite detectar variaciones de un solo nucleotido, variaciones en números de copias, inserciones, deleciones y translocaciones.

Podemos obtener perfiles de expresión génica (RNA de distintos tipos celulares), variantes que afecten a sitios de *splicing*, secuenciación de ARN no codificante (ncARNnc) entre otros.

Next Generation Sequencing (NGS)

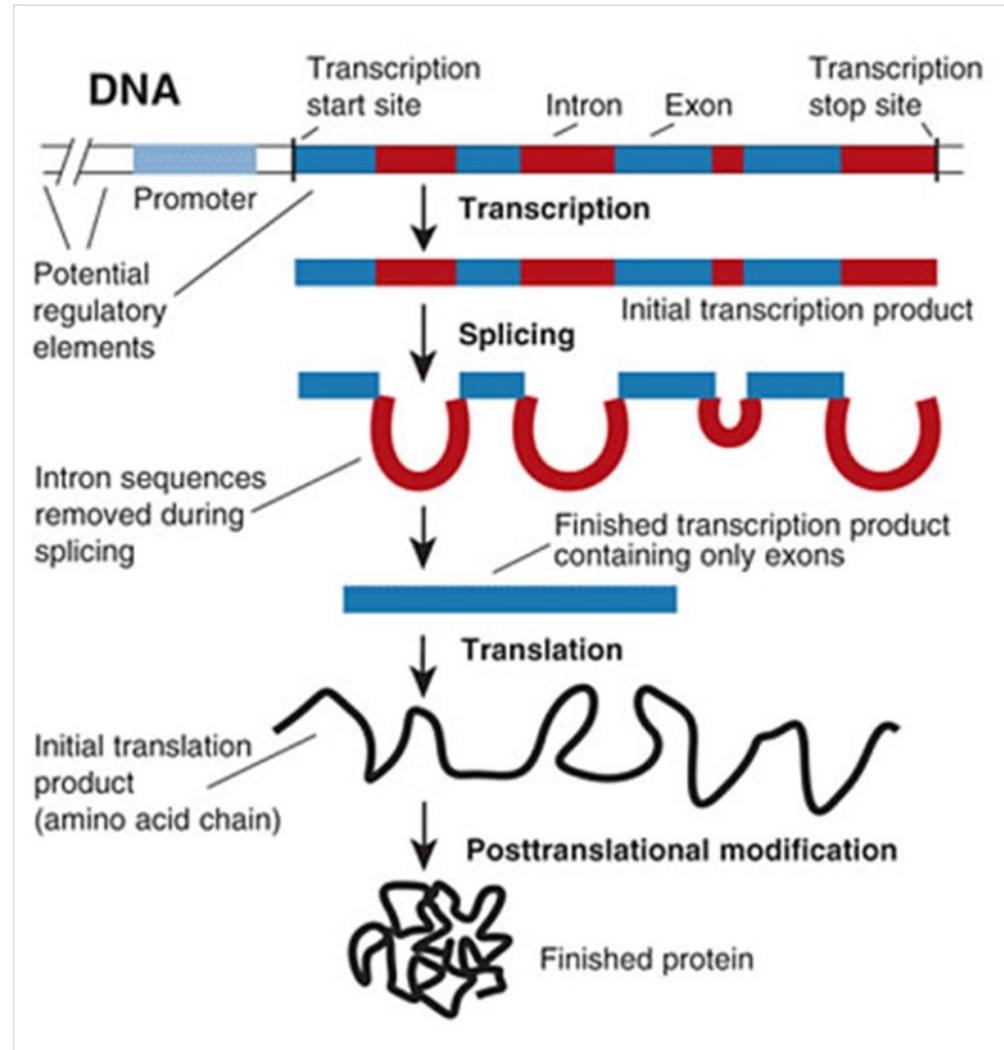
Sanger

NGS

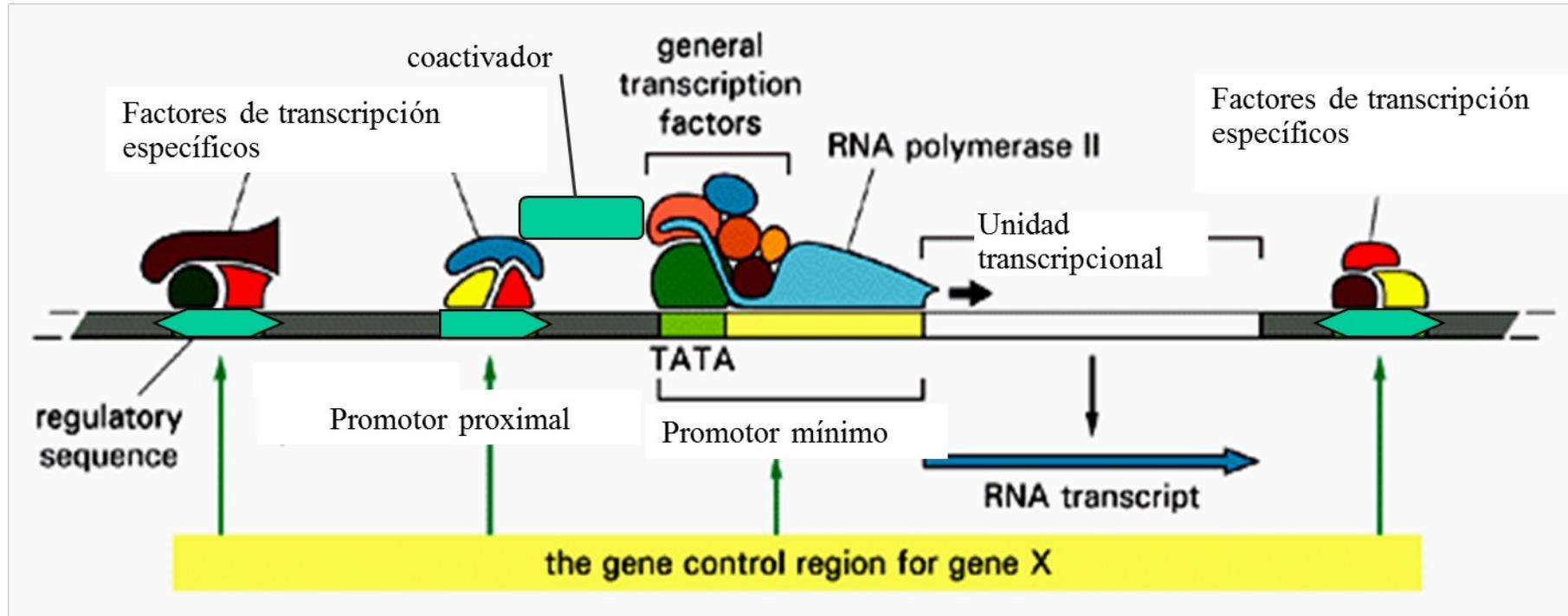
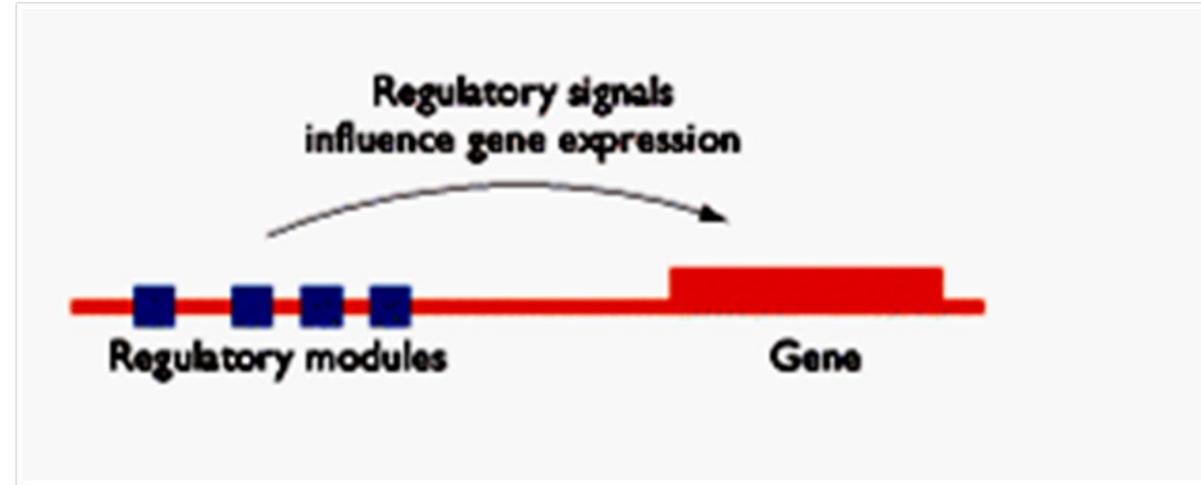
secuenciación automática de un máximo de 96 secuencias de 800 nucleótidos con secuenciadores de primera generación

secuenciación de millones de fragmentos de ADN con equipos de segunda generación

Del ADN a la proteína

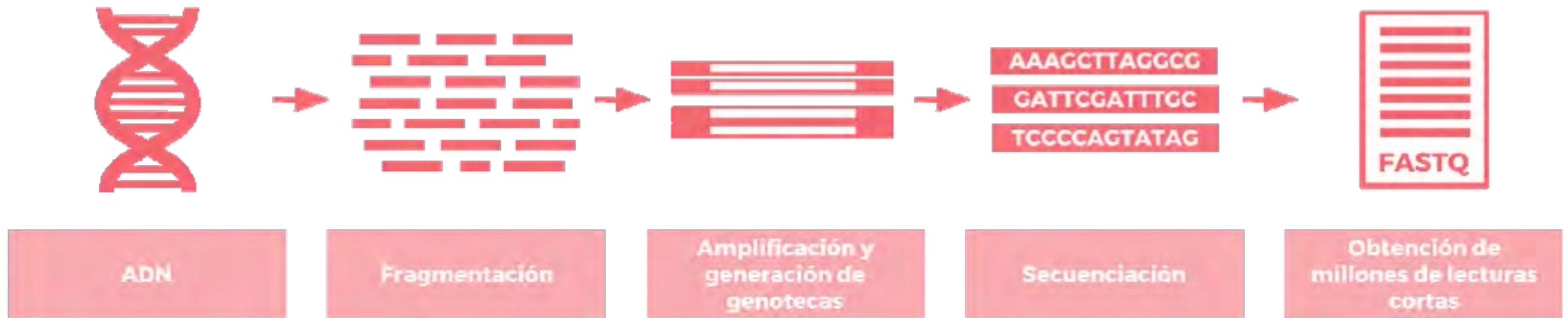


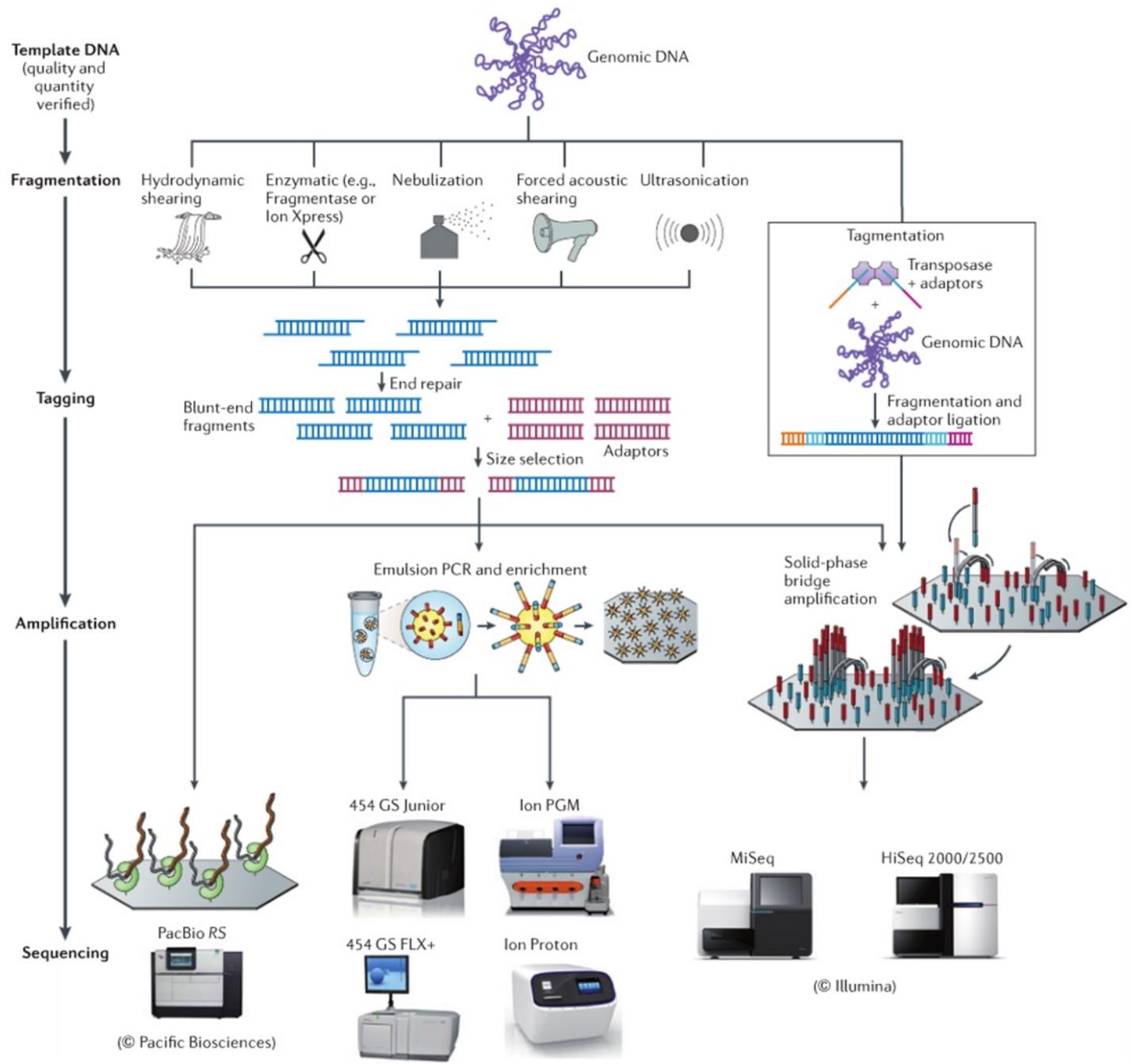
Estructura del gen: promotores, potenciadores, silenciadores.



Secuenciación masiva.

- La secuenciación masiva consiste en fragmentar el ADN en pequeños trozos, amplificar- los mediante PCR y procesarlos todos a la vez

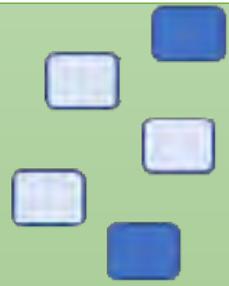




(© Pacific Biosciences)

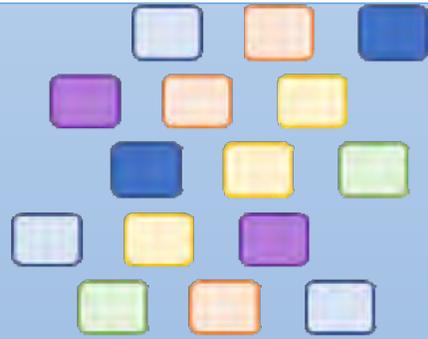
(© Illumina)

- Aunque las plataformas de NGS difieren en la tecnología utilizada, todas ellas comparten la capacidad de secuenciar moldes de ADN amplificados clonalmente.
- La amplificación se lleva a cabo sobre ADN inmovilizado en una superficie sólida (nanopartícula esférica o *bead*), o deposita en el fondo de los nanopocillos de una placa (chip o *flowcell*) en los que se llevará a cabo la reacción de secuenciación.
- Esto permite la lectura en paralelo de millones de secuencias y la reducción drástica del tiempo y del costo de secuenciación.
- Por otro lado, la gran cantidad de datos generados en este proceso ha supuesto un gran reto para ingenieros y bioinformáticos, que han tenido que desarrollar programas específicos de análisis de más fácil manejo.



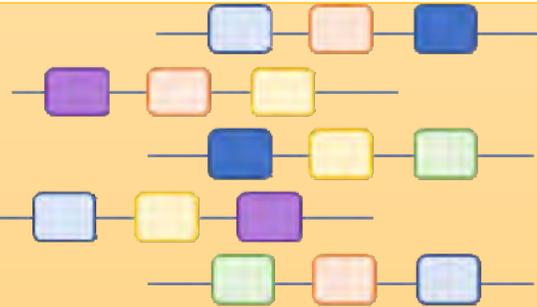
PANEL DE GENES

Secuenciación de un determinado grupo de genes de interés.



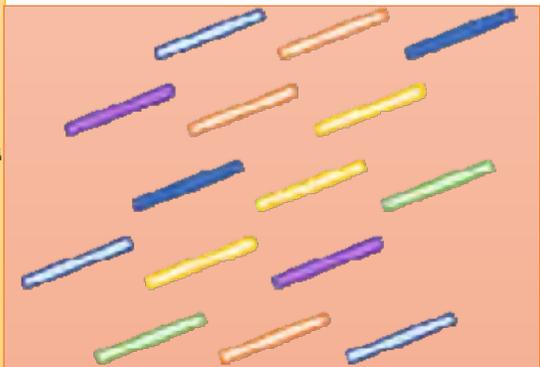
SECUENCIACIÓN DEL EXOMA (WES)

Secuenciación del exoma, la parte del genoma correspondiente a las regiones codificantes (exones).



SECUENCIACIÓN DEL GENOMA (WGS)

Secuenciación del genoma completo de un individuo, incluyendo ADN cromosómico y mitocondrial.

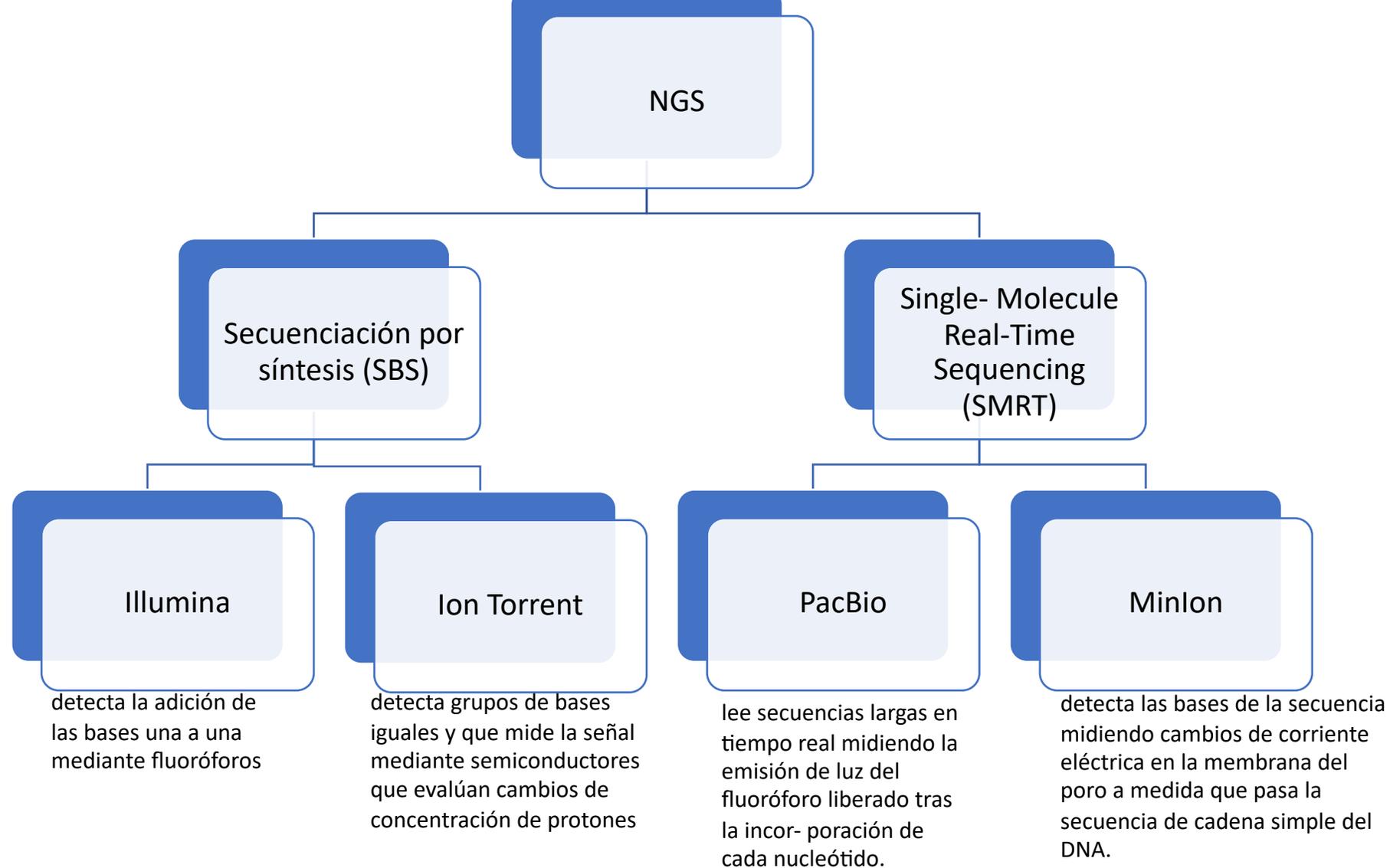


SECUENCIACIÓN DEL TRANSCRPTOMA (WTS)

Secuenciación del ARN transcrito que se traducirá finalmente a proteínas.

Preparación de las muestras.

- La extracción del ADN es un proceso sencillo: extraer los leucocitos de la sangre (o células de tejidos como la mucosa bucal, biopsias o heces), romper su pared celular y la del núcleo, y precipitar el ADN mediante sales y alcohol.
- Inhibir las proteínas que degradan el ADN cuando se lisan las células; obtener solo ADN o ARN (o los dos juntos según se necesite) o únicamente el ADN mitocondrial.
- el ADN obtenido debe ser lo más puro posible y no contenga trazas de los productos usados para la lisis y extracción del ADN (detergentes, fenol, cloroformo, etc.), que interferirían en la PCR y procesos de secuenciación



- A día de hoy hay dos modelos principales de técnicas de secuenciación NGS: secuenciación por síntesis (SBS) que implica fragmentos cortos y *Single- Molecule Real-Time Sequencing (SMRT)* que permite fragmentos de varias kilobases pero con mayor tasa de error.

NGS vs SANGER

- La secuenciación masiva de lecturas cortas como la de Illumina, tiene la misma o más calidad que la secuenciación Sanger (Gold Standard actual) si se realiza a una cobertura de 200x.
- Si bien la secuenciación masiva tiene una mayor tasa de error por lectura, esto no representa un problema ya que su potencia viene de leer la misma secuencia entre 30 y 200 veces.
- La secuenciación Sanger sigue siendo útil en regiones del genoma donde hagan falta secuencias largas para poder alinear/amplificar de forma unívoca una región y asegurarse de que las variantes que se van a estudiar son realmente las de esa zona y no las de un pseudogen o región homóloga.



Dada la importancia de la NGS en el área clínica, el procesado de los datos se ofrece ya muchas veces como un software cerrado al que se proporcionan unos datos de entrada y una configuración y se obtienen "mágicamente" unos resultados. Pero cuidado, esos resultados hay que interpretarlos teniendo en cuenta las medidas de calidad de cada parámetro en la interpretación y visualizar los alineamientos para descartar artefactos.

EXOMA
(Decenas o centenares de miles de variantes)



Filtrado por calidad



Filtrado por tipo de variante



Filtrado por frecuencia de variante



Filtrado por efecto de variante



Filtrado por función del gen
(5 o menos posibles variantes)



Confirmación de la variante

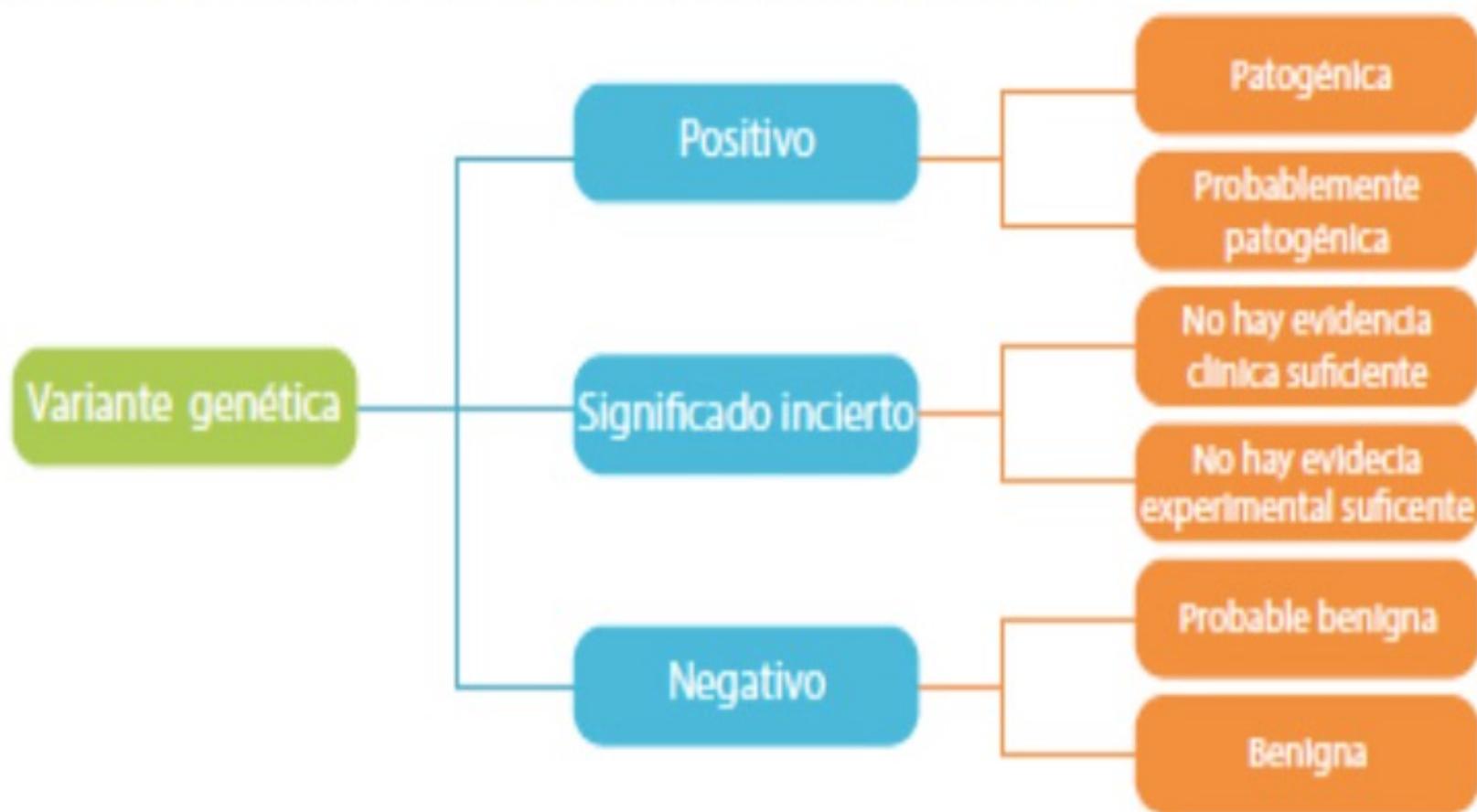
Grandes variaciones estructurales: grandes inserciones, deleciones o translocaciones de muchas kilobases o megabases que engloban uno o varios genes en su totalidad son patogénicas.

En variantes de un solo nucleótido: si hay dudas, se puede hacer mediante secuenciación por Sanger de la zona donde está localizada la mutación.

Una prueba muy contundente a favor de patogenicidad es localizar casos independientes de pacientes con la misma enfermedad y la misma variante.

Otra forma de confirmarlo es realizar un estudio funcional de la variante: mediante mutagénesis dirigida, colocar en un organismo modelo la variante y estudiar los efectos biológicos que puedan tener.

FIGURA 5. CLASIFICACIÓN DE VARIANTES GENÉTICAS.



Análisis bioinformático

- Existen varias herramientas informáticas que proporcionan una estimación de CNVs. Todas ellas emplean diferentes algoritmos que comparan el número de secuencias de un gen (conocido como la cobertura del gen) respecto a las esperadas.
- Por ejemplo, una delección heterocigótica debería tener la mitad de cobertura de la esperada, mientras una duplicación homocigótica debería tener el doble. Una vez más, existen numerosas herramientas, pero las cuatro más utilizadas en la actualidad sonXHMM, CoNIFER, ExomeDepth y CONTRA.

Estudios genéticos moleculares

- PANELES
 - Síndromes con heterogeniedad genética
 - Grupo de patologías específicas
- EXOMA
 - Todos los genes con utilidad clínica conocida
 - Incluye exones codificantes
 - \pm 10 pares de bases de intrones adyacentes
 - secuencias no codificantes con variantes patológicas conocidas
 - ADN mitocondrial

Panel

Invitae Osteogenesis Imperfecta and Bone Fragility Panel

- A Pathogenic variant, c.2429del (p.Pro810Leufs*298), was identified in COL1A1.
- The COL1A1 gene is associated with autosomal dominant osteogenesis imperfecta. This result is consistent with a diagnosis of COL1A1-related conditions.
- Osteogenesis imperfecta (OI) is associated with bone fragility resulting in fractures with minimal trauma (PMID: 24715559). Additional features may include skeletal deformities, short stature, tooth erosion referred to as dentinogenesis imperfecta, hearing loss, and blue or gray sclera. COL1A1/2-related OI is divided into four subtypes based on the number, type, and severity of clinical features, however there is considerable overlap between these clinical subtypes.
- The milder end of the spectrum is associated with only a mild predisposition to fractures, while the severe end of the spectrum is perinatal lethal, with multiple rib fractures and long-bone deformities occurring in utero.
- The more severe types of OI are almost exclusively caused by de novo variants, while the milder types may be inherited from an affected parent.

Invitae Aortopathy Comprehensive Panel

- A Variant of Uncertain Significance, c.992C>A (p.Thr331Lys), was identified in MYLK.
- The MYLK gene is associated with autosomal dominant thoracic aortic aneurysms and dissections (TAAD) and autosomal recessive megacystis-microcolon-intestinal hypoperistalsis syndrome (MMIHS).
- Not all variants present in a gene cause disease. The clinical significance of the variant(s) identified in this gene is uncertain.
- Testing of up to two family members for the Variant(s) of Uncertain Significance (VUS) identified in MYLK is available at no additional cost.

Variant details

- MYLK, Exon 10, c.992C>A (p.Thr331Lys), heterozygous, Uncertain Significance.
- This sequence change replaces threonine with lysine at codon 331 of the MYLK protein (p.Thr331Lys). The threonine residue is weakly conserved and there is a moderate physicochemical difference between threonine and lysine.
- This variant is not present in population databases (ExAC no frequency).
- This variant has not been reported in the literature in individuals with MYLK-related disease.
- Algorithms developed to predict the effect of missense changes on protein structure and function output the following: SIFT: "Tolerated"; PolyPhen-2: "Possibly Damaging"; Align-GVGD: "Class C0".
- The lysine amino acid residue is found in multiple mammalian species, suggesting that this missense change does not adversely affect protein function. These predictions have not been confirmed by published functional studies and their clinical significance is uncertain.
- In summary, the available evidence is currently insufficient to determine the role of this variant in disease. Therefore, it has been classified as a Variant of Uncertain Significance.

Exoma: Aplicaciones clínicas

- Diagnóstico de enfermedades con clínica confundente
- Método diagnóstico ante falta de sospecha clínica
- Diagnóstico de enfermedades con heterogeneidad genética
- Paneles: sospecha de diagnóstico clínico, asegura buena cobertura de todos los genes estudiados. Costo?
- Exoma: ventajas del exoma + ventajas del array
- Exoma actual incluye secuenciación de ADN mitocondrial
- Exoma como examen de primera a segunda línea (cariograma o estudio molecular específico)
- No diagnostica enfermedades por triplete ni alteraciones balanceadas (portadores)

Exoma

- INFORMACIÓN CLÍNICA:
- Anomalía de la ceja; Apéndice cutáneo preauricular; Aplasia / hipoplasia de las uñas; Ceja poco pobladas en su porción lateral
- Clinodactilia del quinto dedo de la mano; Composición anormal de heces; Convulsiones febriles; Dermatitis atópica; Distancia intermamilar aumentada; Epicanto; Filtro liso; Frente prominente; Hendidura palpebral ascendente
- Hernia inguinal; Hipotonía axial infantil; Ictericia neonatal prolongada; Irritabilidad; Nevus azul; Pelo ausente; Pelo ralo; Pestañas rizadas; Pliegues palmares anómalos con uniones anormales; Punta nasal prominente
- Reflujo gastroesofágico
- Retardo global del desarrollo; Retraso del lenguaje y la comunicación; Retrusión del tercio medio de la cara; Talla baja
- Ventriculomegalia

(Información clínica reportada atendiendo a la nomenclatura HPO).

RESULTADO POSITIVO

Variante probablemente patogénica identificada

- Se identificó una variante probablemente patogénica en heterocigosis en el gen SETD5.
- Este resultado es consistente con un diagnóstico genético de trastorno del desarrollo intelectual autosómico dominante tipo 23.
- SETD5,c.2347-2A>T p.?
- Se predice que la variante detectada SETD5, c.2347-2A>T pueda disrumpir el sitio aceptor de splicing altamente conservado del exón 16/22 (NM_001080517.1). Esta variante fue confirmada mediante secuenciación Sanger.
- Se clasifica como probablemente patogénica de acuerdo a las recomendaciones de CENTOGENE y la ACMG.
- La discapacidad intelectual-23 (MRD23) de herencia autosómica dominante se caracteriza por discapacidad intelectual de diversa gravedad, hipotonía, dificultades de alimentación, rasgos dismórficos, autismo y problemas de comportamiento.
- Raramente se han asociado retraso en el crecimiento, anomalías cardíacas congénitas, defectos gastrointestinales y genitourinarios. Modo de herencia: Autosómica dominante (OMIM®: 615761).

Exoma

- Información clínica:
- Atrofia de nervio óptico, atrofia óptico-quiasmática bilateral, hipoacusia conductiva bilateral, insuficiencia tricuspídea, insuficiencia aórtica, epilepsia, talla baja, pectum carinatum, disfunción vesical, enfermedad pulmonar crónica restrictiva.
- INTERPRETACIÓN: POSITIVO, las variantes detectadas contribuyen al menos parcialmente al fenotipo descrito.

DESCRIPCIÓN DE LOS HALLAZGOS

- La variante c.46C>T, ubicada en una copia del gen SNX10, genera un cambio del aminoácido Arginina en la posición 16 por un codón de término, por lo que se considera una variante truncante de proteína (Criterio PVS1). Esta variante ha sido reportada en una frecuencia muy baja para una patología autosómica recesiva en la base de datos de controles gnomAD de exomas (rs587777490, criterio PM2). También ha sido reportada en la literatura (PMID:23280965; PMID:23123320) y en bases de datos de pacientes (HGMD ID:CM130184; ClinVar ID:139565)(Criterio PP5). Según los criterios reunidos por las guías ACMG 2015 (USA) y ACGS 2020 (UK), esta variante se interpreta como patogénica.
- La variante c.389T>C, ubicada en una copia del gen SNX10, genera un cambio del aminoácido Isoleucina en la posición 130 por una Treonina. Esta variante no ha sido reportada en ninguna base de datos de controles (Criterio PM2) y afectaría un dominio (IPR036871) requerido para la función e interacción de SNX10 con otras proteínas (PMID: 21844891; criterio PM1). Según los criterios reunidos por las guías ACMG 2015 (USA) y ACGS 2020 (UK), esta variante se interpreta como de significado incierto.

Interpretación

- Las variantes bi-alélicas en el gen SNX10 generan la condición Osteopetrosis autosómica recesiva #8 (OMIM #615085), cuyas manifestaciones son variables y se superponen con lo descrito para esta paciente (PMID:34624559; PMID:23280965). Se sugiere correlacionar estos hallazgos con lo descrito en la literatura y las manifestaciones de la paciente. Se sugiere estudiar estas variantes en ambos progenitores para determinar si están en cis o en trans, de manera de contribuir en la interpretación final de la variante c.389T>C. En caso de realizarse este estudio en nuestro laboratorio, por favor señalar número de examen.

DESCRIPCIÓN DE LOS HALLAZGOS

- La variante c.1753A>G, ubicada en una copia del gen HUWE1, genera un cambio del aminoácido Metionina en la posición 585 por una Valina.
- Esta variante no ha sido reportada en ninguna base de datos de controles (Criterio PM2) y afectaría un dominio (IPR010314) requerido para la función de HUWE1 (PMID: 29180823; criterio PM1).
- La variante afecta un gen intolerante a variantes con sentido erróneo, las cuales constituyen un mecanismo frecuente de patogenicidad (PMID: 29180823; gnomAD missense Z-score:8.87; criterio PP2).
- Según los criterios reunidos por las guías ACMG 2015 (USA) y ACGS 2020 (UK), esta variante se interpreta como de significado incierto.

Interpretación

- Las variantes heterocigotas o hemicingotas del gen HUWE1 generan la condición Trastorno del Desarrollo Intelectual ligado a X, tipo Turner (OMIM #309590), de herencia ligada a X (tanto recesiva como dominante) y expresividad altamente variable, y cuyas manifestaciones se superponen con lo descrito para esta paciente (PMID:29180823).
- Se sugiere correlacionar estos hallazgos con lo descrito en la literatura y las manifestaciones de la paciente. Se sugiere estudiar esta variante en ambos progenitores para determinar si fue heredada o es de novo, de manera de contribuir en su interpretación.
- En caso de realizarse este estudio en nuestro laboratorio, por favor señalar número de examen y si algún progenitor presenta manifestaciones similares a la paciente y cuáles.

Genoma

- Secuenciación y análisis molecular del genoma completo: ADN que codifica para proteínas (intrones y exones), elementos regulatorios y estructurales.
- 3.200 millones de pares de bases nucleares y 16.569 mitocondriales.
- Permite estudio de portadores de alteraciones balanceadas y enfermedades por tripletes.
- Falta definir:
 - cómo interactúan los componentes del genoma a nivel genético y proteico.
 - Variación normal en el genoma humano.
- Dificultades en interpretación de resultados.
- Aún estudio de investigación.