

Máster en Bioinformática

Secuenciación Genómica y Análisis De Variantes Para Medicina Personalizada y De Precisión

Curso académico 2024-25

Edición Abril



Universidad
Internacional
de Valencia

Dra. Laura Gutiérrez Macías
laura.gutierrez.m@professor.universidadviu.com

Capítulo 3. ¿Cómo analizamos un genoma eucariota? Paneles de captura de genes vs genoma completo

3.1.

- Introducción al análisis de genomas eucariotas

3.2.

- Estrategias generales de análisis
 - 3.2.1.- Paneles de genes o regiones de interés
 - 3.2.2.- Secuenciación de exoma (WES)
 - 3.2.3.- Secuenciación de genoma completo (WGS)

3.3.

- Protocolo general de análisis
 - 3.3.1.- Extracción de ADN
 - 3.3.2.- Preparación de genotecas o librerías
 - 3.3.3.- Secuenciación de la genoteca/librería

3.4.

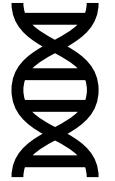
- Test Prenatal No-Invasivo (NIPT)
 - ¿Qué es?
 - ¿Cómo funciona?
 - ¿Cuándo está indicado?
 - Ventajas y limitaciones



3.1

Introducción al análisis de genomas eucariotas

Proyecto Genoma Humano (PGH)



Programa de investigación colaborativo e internacional cuya meta era la del mapeo (cartografía) y entendimiento completo de todos los genes de los seres humanos (genoma).

Los investigadores del PGH han descifrado el genoma humano de tres maneras principales:

- determinado el orden, o "**secuencia**", de todas las bases en el ADN de nuestro genoma;
- **elaborado mapas** que muestran las **ubicaciones de los genes** para las principales secciones de todos nuestros cromosomas;
- y producido lo que se conoce como **mapas de ligamiento**, a través de los cuales puede darse seguimiento a características heredadas (tales como aquellas para enfermedades genéticas) a lo largo de varias generaciones.



El **Consorcio** Internacional de Secuenciación del Genoma Humano (International Human Genome Sequencing Consortium) fue fundado en **1990** por el Departamento de Energía y los departamentos nacionales de salud de los EEUU, liderado por Francis Collins, aunando un grupo de investigación público de múltiples países diferentes. El plazo inicial de realización fue de 15 años.

Se publicó la primera versión preliminar del genoma humano en la revista Nature en **febrero de 2001**, con el **90 por ciento de la secuencia** de los tres mil millones de pares de bases del genoma completo. Un dato sorprendente de esta primera versión preliminar fue que el número de genes humanos parecía ser significativamente menor que las estimaciones anteriores, que variaron desde 50,000 genes hasta tantos como 140,000. **La secuencia completa fue terminada y publicada en abril de 2003.**

Paralelamente, la **Corporación Celera**, fundada en **1998** por **Craig Venter**, trabajó en la secuenciación del genoma humano, quienes publicaron en **2021** en **Science** el esbozo de 5 genomas de diferentes etnias.

En **mayo de 2006** se publicó la secuencia del último cromosoma (cromosoma 1) humano en la revista Nature.

Como extensión del PGH actualmente se encuentra en marcha el **proyecto microbioma humano**.

Junio 2021 se secuenció por primera vez el genoma completo de un ser humano, apoyándose en tecnología de long reads: 3.055 pares de bases que incluían un 8% del mismo que hasta el momento era desconocido.

Ciencia / Materia

ASTROFÍSICA · MEDIO AMBIENTE · INVESTIGACIÓN MÉDICA · MATEMÁTICAS · PALEONTOLOGÍA · ÚLTIM.

GENÉTICA >

Un consorcio internacional secuencia por primera vez el genoma completo de un ser humano

El libro de instrucciones de una persona tiene 3.055 millones de letras, según la nueva lectura, que incluye un 8% del ADN que permanecía oculto por falta de tecnología

Earth's heart of iron begins
to yield its secrets p. 18

Microglia in chronic pain recovery
and relapse pp. 33 & 86

Particle acceleration
in a nova explosion p. 77

Science

\$15
1 APRIL 2022
SPECIAL ISSUE
science.org



FILLING THE GAPS

Closing in on a complete
human genome p. 42



SPECIAL SECTION

COMPLETING THE HUMAN GENOME

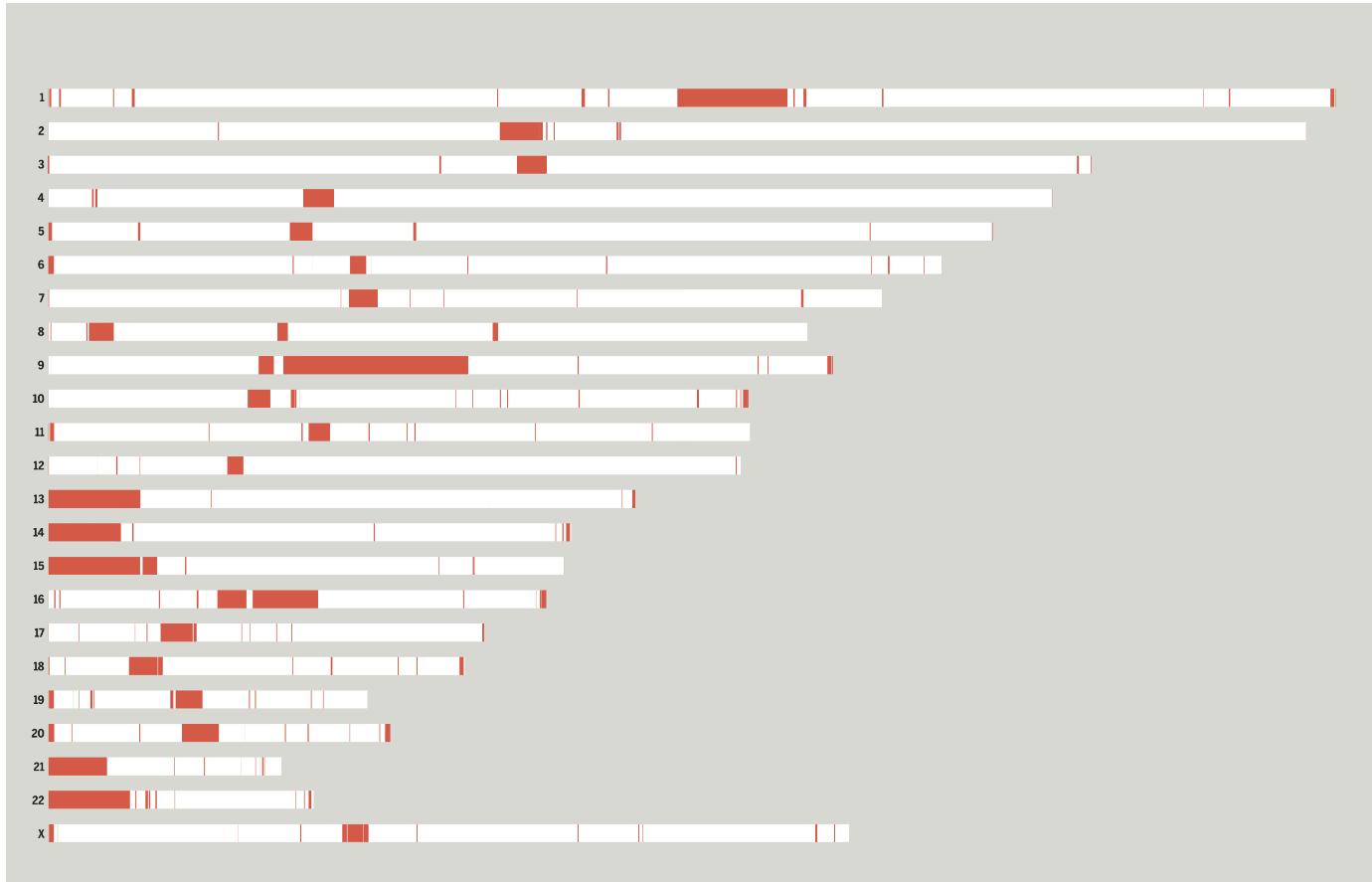
RESEARCH ARTICLE

HUMAN GENOMICS

The complete sequence of a human genome

Sergey Nurk^{1†}, Sergey Koren^{1†}, Arang Rhie^{1†}, Mikko Rautiainen^{1†}, Andrey V. Bzikadze², Alla Mikheenko³, Mitchell R. Vollger⁴, Nicolas Altemose⁵, Lev Uralsky^{6,7}, Ariel Gershman⁸, Sergey Aganezov^{9‡}, Savannah J. Hoyt¹⁰, Mark Diekhans¹¹, Glennis A. Logsdon⁴, Michael Alonge⁹, Stylianos E. Antonarakis¹², Matthew Borchers¹³, Gerard G. Bouffard¹⁴, Shelise Y. Brooks¹⁴, Gina V. Caldas¹⁵, Nae-Chyun Chen⁹, Haoyu Cheng^{16,17}, Chen-Shan Chin¹⁸, William Chow¹⁹, Leonardo G. de Lima¹³, Philip C. Dishuck⁴, Richard Durbin^{19,20}, Tatiana Dvorkina³, Ian T. Fiddes²¹, Giulio Formenti^{22,23}, Robert S. Fulton²⁴, Arkarachai Fungtammasan¹⁸, Erik Garrison^{11,25}, Patrick G. S. Grady¹⁰, Tina A. Graves-Lindsay²⁶, Ira M. Hall²⁷, Nancy F. Hansen²⁸, Gabrielle A. Hartley¹⁰, Marina Haukness¹¹, Kerstin Howe¹⁹, Michael W. Hunkapiller²⁹, Chirag Jain^{1,30}, Miten Jain¹¹, Erich D. Jarvis^{22,23}, Peter Kerpeljiev³¹, Melanie Kirsche⁹, Mikhail Kolmogorov³², Jonas Korlach²⁹, Milinn Kremitzki²⁶, Heng Li^{16,17}, Valerie V. Maduro³³, Tobias Marschall³⁴, Ann M. McCartney¹, Jennifer McDaniel³⁵, Danny E. Miller^{4,36}, James C. Mullikin^{14,28}, Eugene W. Myers³⁷, Nathan D. Olson³⁵, Benedict Paten¹¹, Paul Peluso²⁹, Pavel A. Pevzner³², David Porubsky⁴, Tamara Potapova¹³, Evgeny I. Rogaev^{6,7,38,39}, Jeffrey A. Rosenfeld⁴⁰, Steven L. Salzberg^{9,41}, Valerie A. Schneider⁴², Fritz J. Sedlazeck⁴³, Kishwar Shafin¹¹, Colin J. Shew⁴⁴, Alaina Shumate⁴¹, Ying Sims¹⁹, Arian F. A. Smit⁴⁵, Daniela C. Soto⁴⁴, Ivan Sović^{29,46}, Jessica M. Storer⁴⁵, Aaron Streets^{5,47}, Beth A. Sullivan⁴⁸, Françoise Thibaud-Nissen⁴², James Torrance¹⁹, Justin Wagner³⁵, Brian P. Walenz¹, Aaron Wenger²⁹, Jonathan M. D. Wood¹⁹, Chunlin Xiao⁴², Stephanie M. Yan⁴⁹, Alice C. Young¹⁴, Samantha Zarate⁹, Urvashi Surti⁵⁰, Rajiv C. McCoy⁴⁹, Megan Y. Dennis⁴⁴, Ivan A. Alexandrov^{3,7,51}, Jennifer L. Gerton^{13,52}, Rachel J. O'Neill¹⁰, Winston Timp^{8,41}, Justin M. Zook³⁵, Michael C. Schatz^{9,49}, Evan E. Eichler^{4,53*}, Karen H. Miga^{11,54*}, Adam M. Phillippy^{1*}

Since its initial release in 2000, the human reference genome has covered only the euchromatic fraction of the genome, leaving important heterochromatic regions unfinished. Addressing the remaining 8% of the genome, the Telomere-to-Telomere (T2T) Consortium presents a complete 3.055 billion-base pair sequence of a human genome, T2T-CHM13, that includes gapless assemblies for all chromosomes except Y, corrects errors in the prior references, and introduces nearly 200 million base pairs of sequence containing 1956 gene predictions, 99 of which are predicted to be protein coding. The completed regions include all centromeric satellite arrays, recent segmental duplications, and the short arms of all five acrocentric chromosomes, unlocking these complex regions of the genome to variational and functional studies.



Each bar is a linear visualization of a chromosome, with the chromosome number shown at left. Red segments denote previously missing sequences that the T2T Consortium resolved.

GRAPHIC: V. ALTOUMANI/SCIENCE; DATA: T2T CONSORTIUM

Through the **resolution of previously unsequenceable and unalignable regions, mostly composed of highly repetitive sequences, this reference genome allows for a detailed characterization of the centromeric satellite repeats, transposable elements, and segmental duplications**. Mapping of genomic sequences, including those from previously published studies, resolves aspects of human genetic diversity, including evolutionary comparisons with our primate relatives. Furthermore, **it allows for identification of how changes in methylation density differ within and among centromeres and how epigenetics can affect the transcription of repeat sequences**.

STATISTICS

	GRCH38	T2T-CHM13	DIFFERENCE (%)
Summary			
Assembled bases (Gbp)	2.92	3.05	+4.5
Unplaced bases (Mbp)	11.42	0	-100.0
Gap bases (Mbp)	120.31	0	-100.0
Number of contigs	949	24	-97.5
Contig NG50 (Mbp)	56.41	154.26	+173.5
Number of issues	230	46	-80.0
Issues (Mbp)	230.43	8.18	-96.5
Gene annotation			
Number of genes	60,090	63,494	+5.7
Protein coding	19,890	19,969	+0.4
Number of exclusive genes	263	3,604	
Protein coding	63	140	
Number of transcripts	228,597	233,615	+2.2
Protein coding	84,277	86,245	+2.3
Number of exclusive transcripts	1,708	6,693	
Protein coding	829	2,780	
Segmental duplications			
Percentage of segmental duplications (%)	5.00	6.61	
Segmental duplication bases (Mbp)	151.71	201.93	+33.1
Number of segmental duplications	24097	41528	+72.3
RepeatMasker			
Percentage of repeats (%)	51.89	53.94	
Repeat bases (Mbp)	1,516.37	1,647.81	+8.7
Long interspersed nuclear elements	626.33	631.64	+0.8
Short interspersed nuclear elements	386.48	390.27	+1.0
Long terminal repeats	267.52	269.91	+0.9
Satellite	76.51	150.42	+96.6
DNA	108.53	109.35	+0.8
Simple repeat	36.5	77.69	+112.9
Low complexity	6.16	6.44	+4.6
Retroposon	4.51	4.65	+3.3
rRNA	0.21	1.71	+730.4

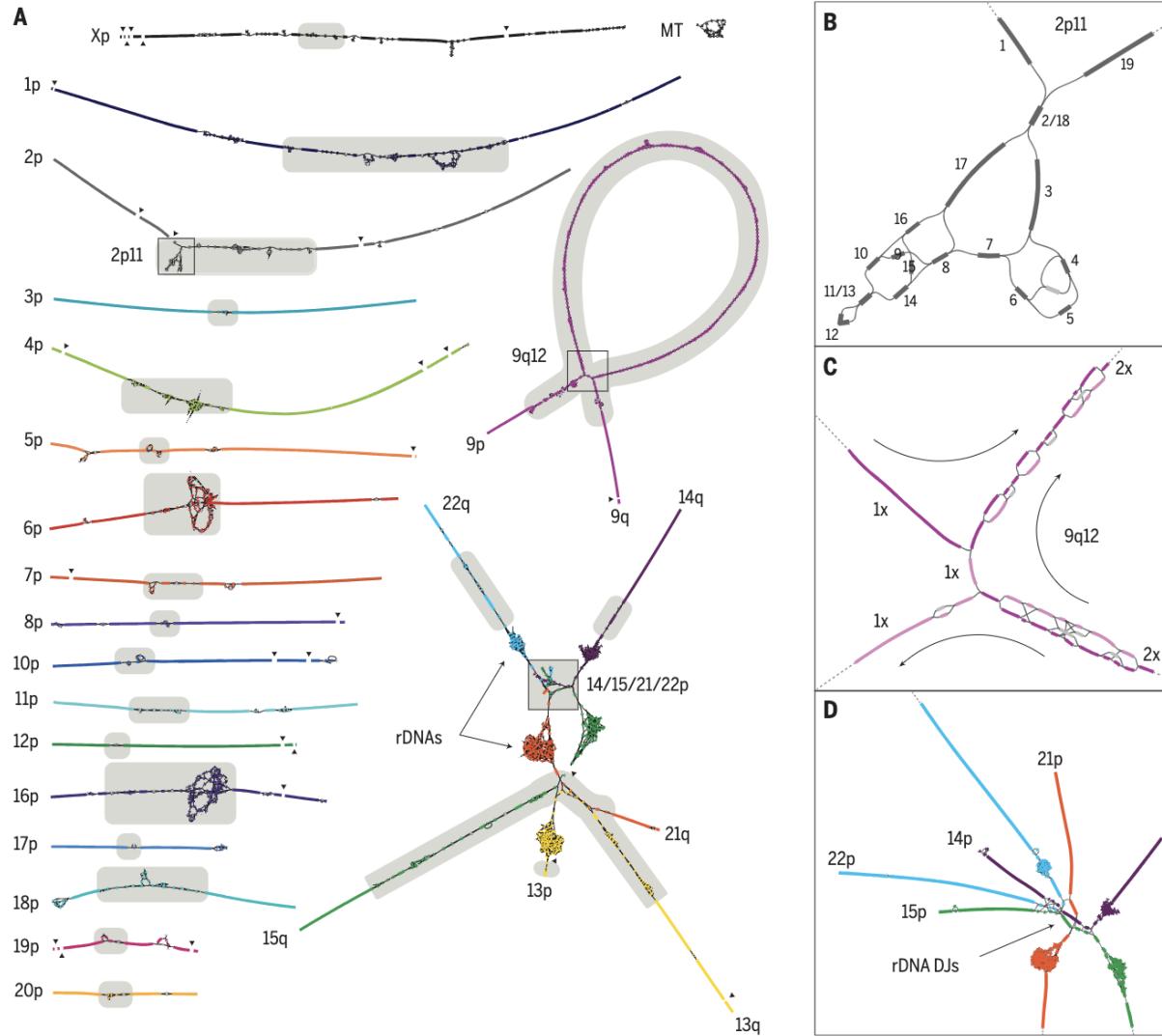


Fig. 2. High-resolution assembly string graph of the CHM13 genome.

(A) Bandage (60) visualization, where nodes represent unambiguously assembled sequences scaled by length and edges correspond to the overlaps between node sequences. Each chromosome is both colored and numbered on the short (p) arm. Long (q) arms are labeled where unclear. The five acrocentric chromosomes (bottom right) are connected owing to similarity between their short arms, and the rDNA arrays form five dense tangles because of their high copy number. The graph is partially fragmented because of HiFi coverage dropout surrounding GA-rich sequence (black triangles). Centromeric satellites (30) are the source of most ambiguity in the graph (gray highlights). MT, mitochondria.

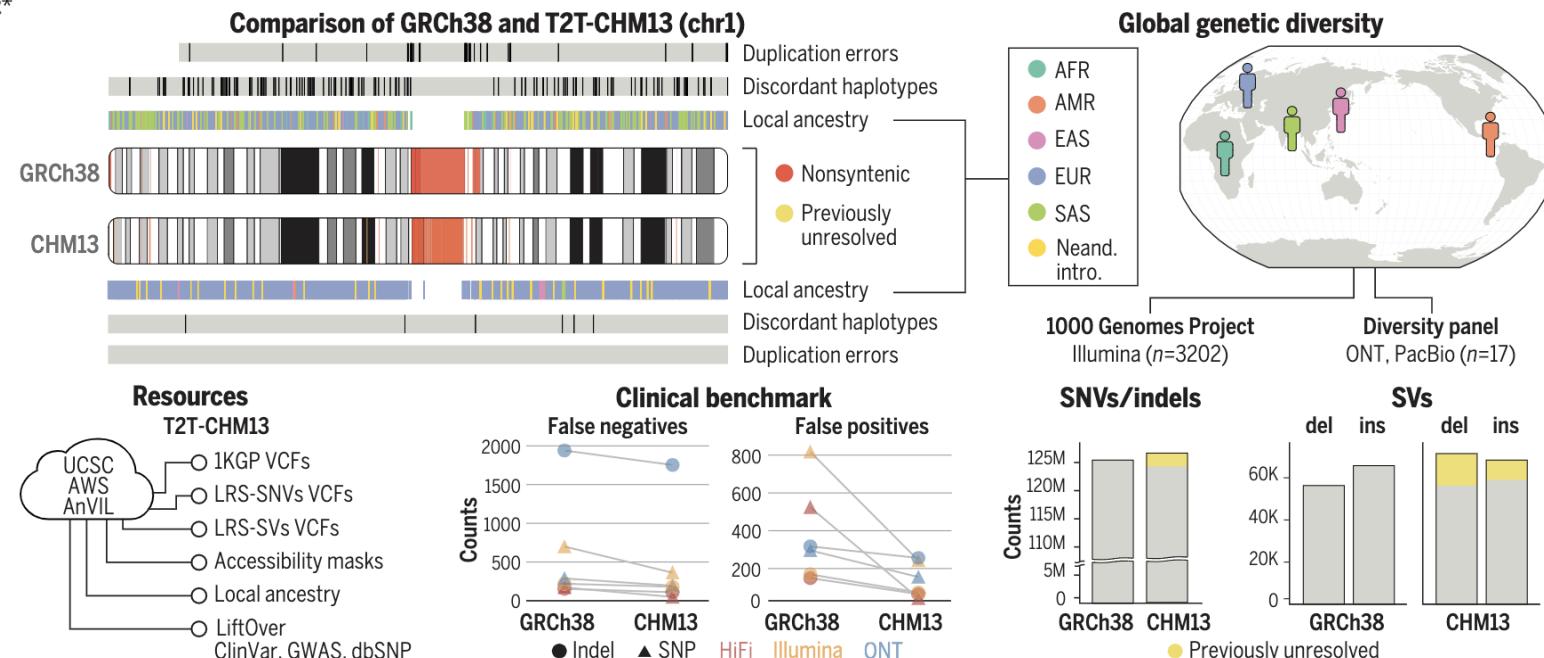
(B) The ONT-assisted graph traversal for the 2p11 locus is given by numerical order. Based on low depth of coverage, the unlabeled light gray node represents an artifact or heterozygous variant and was not used. (C) The multimegapbase tandem HSat3 duplication (9q+) at 9q12 requires two traversals of the large loop structure. (The size of the loop is exaggerated because graph edges are of constant size.) Nodes used by the first traversal are in dark purple, and nodes used by the second traversal are in light purple. Nodes used by both traversals typically have twice the sequencing coverage. (D) Enlargement of the distal short arms of the acrocentrics, showing the colored graph walks and edges between highly similar sequences in the distal junctions (DJs) adjacent to the rDNA arrays.

RESEARCH ARTICLE SUMMARY

HUMAN GENOMICS

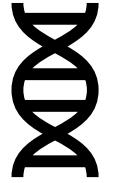
A complete reference genome improves analysis of human genetic variation

Sergey Aganezov†, Stephanie M. Yan†, Daniela C. Soto†, Melanie Kirsch†, Samantha Zarate†, Pavel Avdeyev, Dylan J. Taylor, Kishwar Shafin, Alaina Shumate, Chunlin Xiao, Justin Wagner, Jennifer McDaniel, Nathan D. Olson, Michael E. G. Sauria, Mitchell R. Vollger, Arang Rhee, Melissa Meredith, Skylar Martin, Joyce Lee, Sergey Koren, Jeffrey A. Rosenfeld, Benedict Paten, Ryan Layer, Chen-Shan Chin, Fritz J. Sedlazeck, Nancy F. Hansen, Danny E. Miller, Adam M. Phillippy, Karen H. Miga, Rajiv C. McCoy*, Megan Y. Dennis*, Justin M. Zook*, Michael C. Schatz*



Genomic features and resources available for T2T-CHM13. Comparisons to GRCh38 reveal broad improvements in SNVs, indels, and SVs discovered across diverse human populations by means of short-read (1KGP) and long-read sequencing (LRS). These improvements are due to resolution of complex genomic loci (nonsyntenic and previously unresolved), duplication errors, and discordant haplotypes, including those in medically relevant genes.

Proyecto Genoma Humano (PGH)



<https://www.genome.gov/breve-historia-del-proyecto-del-genoma-humano>

<https://www.genome.gov/11510905/preguntas-maacutees-frecuentes>

<https://www.genome.gov/panorama-general-del-proyecto-del-genoma-humano>

<https://www.nih.gov/news-events/nih-research-matters/first-complete-sequence-human-genome#:~:text=The%20Human%20Genome%20Project%2C%20completed,important%20for%20fundamental%20biological%20processes.>

GRCh38.p14

https://www.ensembl.org/Homo_sapiens/Info/Index?db=core

 Human (GRCh38.p14) ▾

Search Human (Homo sapiens)

Search all categories ▾ Search... Go

e.g. [BRCA2](#) or [17:63992802-64038237](#) or [rs699](#) or [osteoarthritis](#)

Genome assembly: GRCh38.p14 (GCA_000001405.29)

-  [More information and statistics](#)
-  [Download DNA sequence \(FASTA\)](#)
-  [Convert your data to GRCh38 coordinates](#)
-  [Display your data in Ensembl](#)

Other assemblies

GRCh37 Full Feb 2014 archive with BLAST, VEP and BioMart ▾ Go

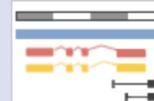
Comparative genomics

What can I find? Homologues, gene trees, and whole genome alignments across multiple species.

-  [More about comparative analysis](#)
-  [Download alignments \(EMF\)](#)

Regulation

 View karyotype

 Example region

Gene annotation

What can I find? Protein-coding and non-coding genes, splice variants, cDNA and protein sequences, non-coding RNAs.

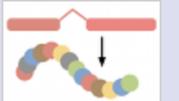
-  [More about this genebuild](#)
-  [Download FASTA files for genes, cDNAs, ncRNA, proteins](#)
-  [Download GTF or GFF3 files for genes, cDNAs, ncRNA, proteins](#)
-  [Update your old Ensembl IDs](#)

Variation

What can I find? Short sequence variants and longer structural variants; disease and other phenotypes

-  [More about variation in Ensembl](#)
-  [Download all variants \(GVF\)](#)
-  [Variant Effect Predictor](#)

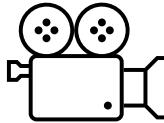
 Example gene

 Example transcript

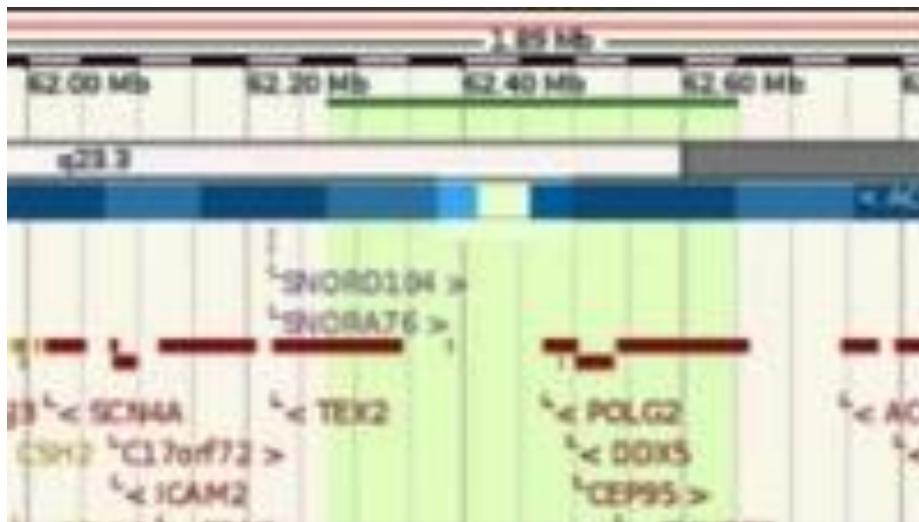
 Example variant

 /2024

GRCh38.p14



Patches & haplotypes in the human genome:
https://www.youtube.com/watch?v=sPE9j_Hw9HU



Statistics	
Summary	
Assembly	GRCh38.p13 (Genome Reference Consortium Human Build 38), INSDC Assembly GCA_000001405.28 , Dec 2013
Base Pairs	3,096,649,726
Golden Path Length	3,096,649,726
Assembly provider	Genome Reference Consortium
Annotation provider	Ensembl
Annotation method	Full genebuild
Genebuild started	Jan 2014
Genebuild released	Jul 2014
Genebuild last updated/patched	Jul 2022
Database version	108.38
Gencode version	GENCODE 42
Gene counts (Primary assembly)	
Coding genes	19,813 (excl 651 readthrough)
Non coding genes	25,972
Small non coding genes	4,864
Long non coding genes	18,887
Misc non coding genes	2,221
Pseudogenes	15,241
Gene transcripts	252,477
Gene counts (Alternative sequence)	
Coding genes	3,028 (excl 26 readthrough)
Non coding genes	1,682
Small non coding genes	297
Long non coding genes	1,198
Misc non coding genes	187
Pseudogenes	1,796
Gene transcripts	21,630
Other	
Genscan gene predictions	51,756
Short Variants	715,081,156
Structural variants	7,097,115

GRCh38.p14

Index of /pub/release-112/fasta/homo_sapiens

Name	Last modified	Size	Description
 Parent Directory	-	-	
 cdna/	2024-04-23 05:26	-	
 cds/	2024-04-23 05:26	-	
 dna/	2024-04-23 05:28	-	
 dna_index/	2024-04-23 05:29	-	
 ncrna/	2024-04-23 05:29	-	
 pep/	2024-04-23 05:29	-	

https://ftp.ensembl.org/pub/release-112/fasta/homo_sapiens/dna/

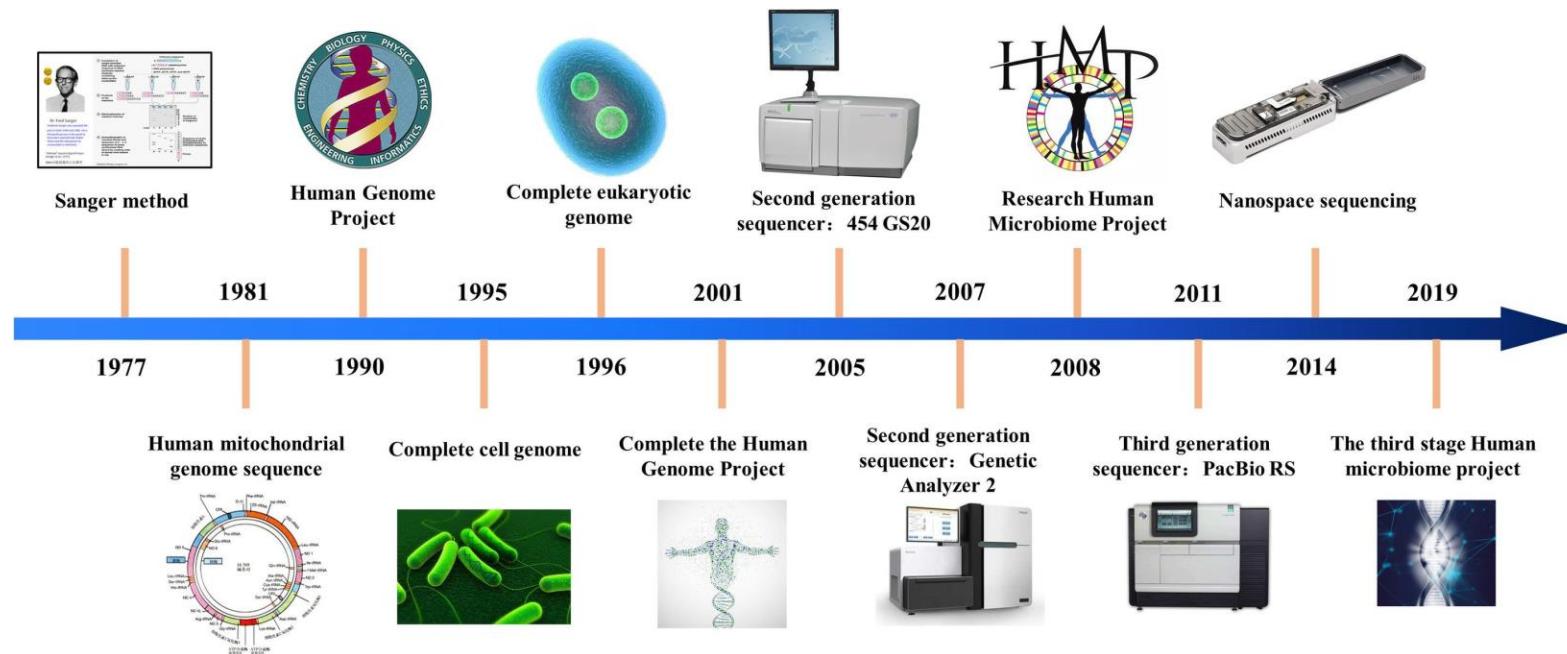
 Homo_sapiens.GRCh38.dna.alt.fa.gz	2024-02-13 18:14	52M
 Homo_sapiens.GRCh38.dna.chromosome.1.fa.gz	2024-02-13 18:14	66M
 Homo_sapiens.GRCh38.dna.chromosome.2.fa.gz	2024-02-13 18:14	69M
 Homo_sapiens.GRCh38.dna.chromosome.3.fa.gz	2024-02-13 18:14	57M
 Homo_sapiens.GRCh38.dna.chromosome.4.fa.gz	2024-02-13 18:14	55M
 Homo_sapiens.GRCh38.dna.chromosome.5.fa.gz	2024-02-13 18:14	52M
 Homo_sapiens.GRCh38.dna.chromosome.6.fa.gz	2024-02-13 18:15	49M
 Homo_sapiens.GRCh38.dna.chromosome.7.fa.gz	2024-02-13 18:15	45M
 Homo_sapiens.GRCh38.dna.chromosome.8.fa.gz	2024-02-13 18:15	42M
 Homo_sapiens.GRCh38.dna.chromosome.9.fa.gz	2024-02-13 18:15	35M
 Homo_sapiens.GRCh38.dna.chromosome.10.fa.gz	2024-02-13 18:14	38M
 Homo_sapiens.GRCh38.dna.chromosome.11.fa.gz	2024-02-13 18:14	38M
 Homo_sapiens.GRCh38.dna.chromosome.12.fa.gz	2024-02-13 18:14	38M
 Homo_sapiens.GRCh38.dna.chromosome.13.fa.gz	2024-02-13 18:14	28M
 Homo_sapiens.GRCh38.dna.chromosome.14.fa.gz	2024-02-13 18:14	26M
 Homo_sapiens.GRCh38.dna.chromosome.15.fa.gz	2024-02-13 18:14	24M
 Homo_sapiens.GRCh38.dna.chromosome.16.fa.gz	2024-02-13 18:14	23M
 Homo_sapiens.GRCh38.dna.chromosome.17.fa.gz	2024-02-13 18:14	23M
 Homo_sapiens.GRCh38.dna.chromosome.18.fa.gz	2024-02-13 18:14	22M
 Homo_sapiens.GRCh38.dna.chromosome.19.fa.gz	2024-02-13 18:14	16M
 Homo_sapiens.GRCh38.dna.chromosome.20.fa.gz	2024-02-13 18:14	18M
 Homo_sapiens.GRCh38.dna.chromosome.21.fa.gz	2024-02-13 18:14	11M
 Homo_sapiens.GRCh38.dna.chromosome.22.fa.gz	2024-02-13 18:14	11M
 Homo_sapiens.GRCh38.dna.chromosome.MT.fa.gz	2024-02-13 18:14	5.3K
 Homo_sapiens.GRCh38.dna.chromosome.X.fa.gz	2024-02-13 18:15	44M
 Homo_sapiens.GRCh38.dna.chromosome.Y.fa.gz	2024-02-13 18:15	7.5M

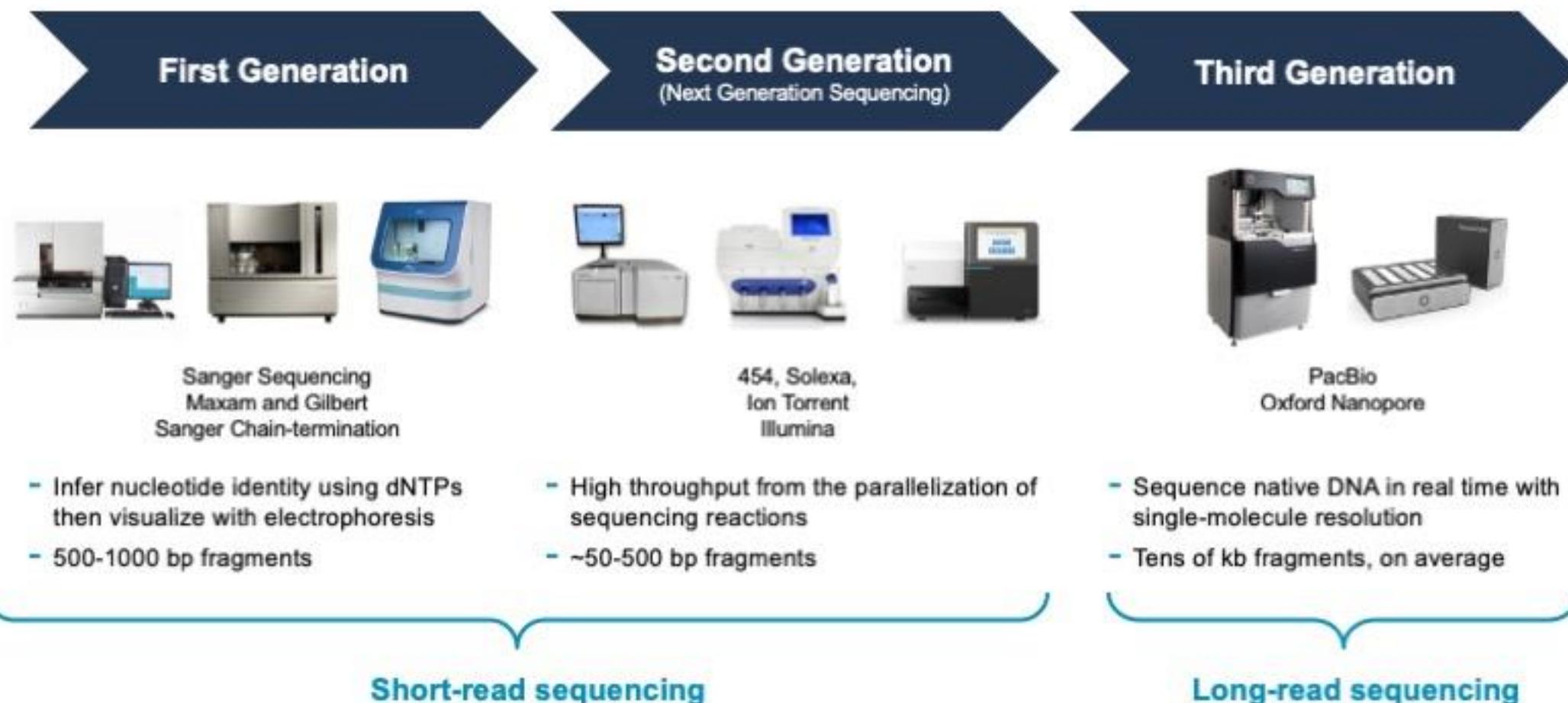
Desde que se completó el Proyecto Genoma Humano (HGP) en 2003, la era de la tecnología de secuenciación masiva (NGS) aplicada a su conocimiento despegó y comenzó a revolucionar la práctica médica y el diagnóstico clínico.



Este tipo de técnicas nos ayudan a solventar problemas de resolución que presentan técnicas moleculares convencionales, como la secuenciación Sanger de genes candidatos, la hibridación comparativa por arrays o el cariotipado.

Nos proporcionan el poder de detectar nuevas variantes e incrementando el diagnóstico de enfermedades raras y complejas de etiología desconocida. Su aplicación no solo se ciñe a enfermedades raras o desconocidas, sino que incluye cáncer, enfermedades complejas o análisis de modificaciones en la expresión génica.





Nota: en el aula virtual tenéis videos explicativos de cada tipo de tecnología y la química que subyace detrás

3.1. Introducción al análisis de genomas eucariotas

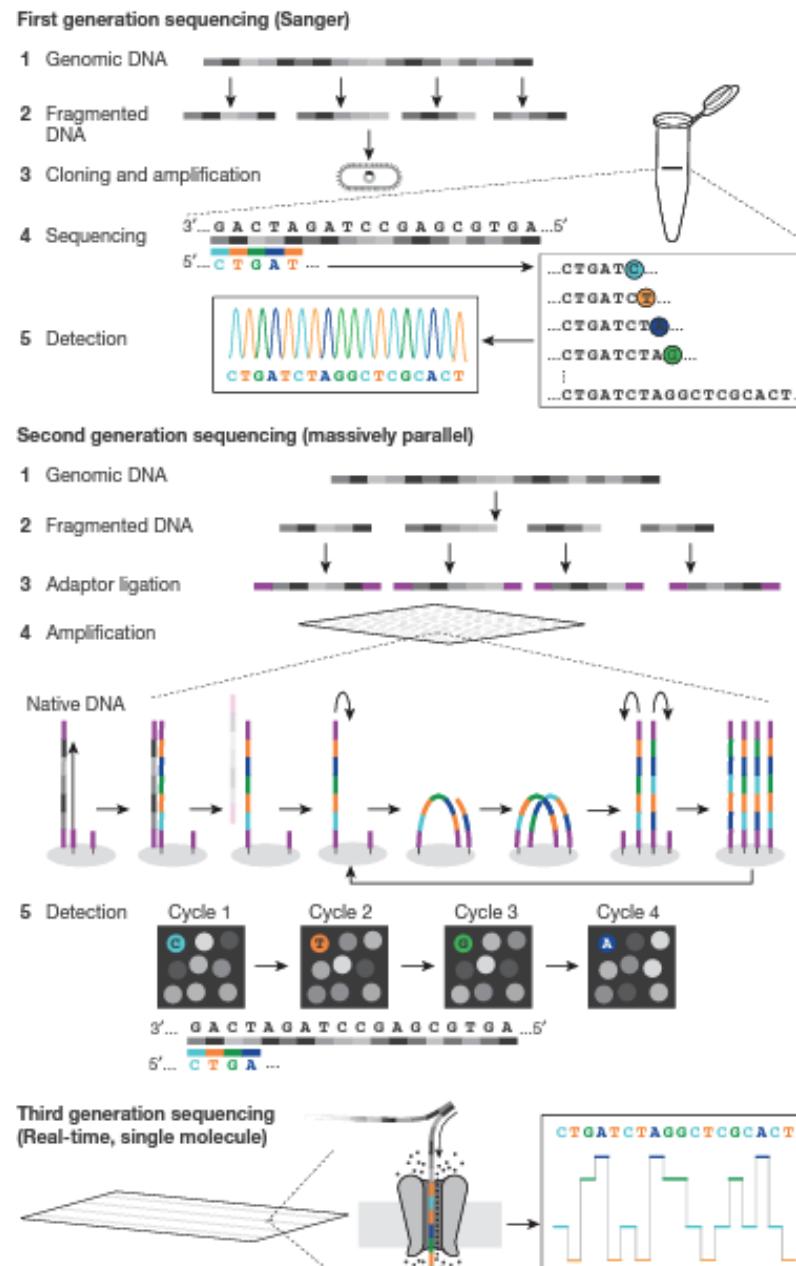
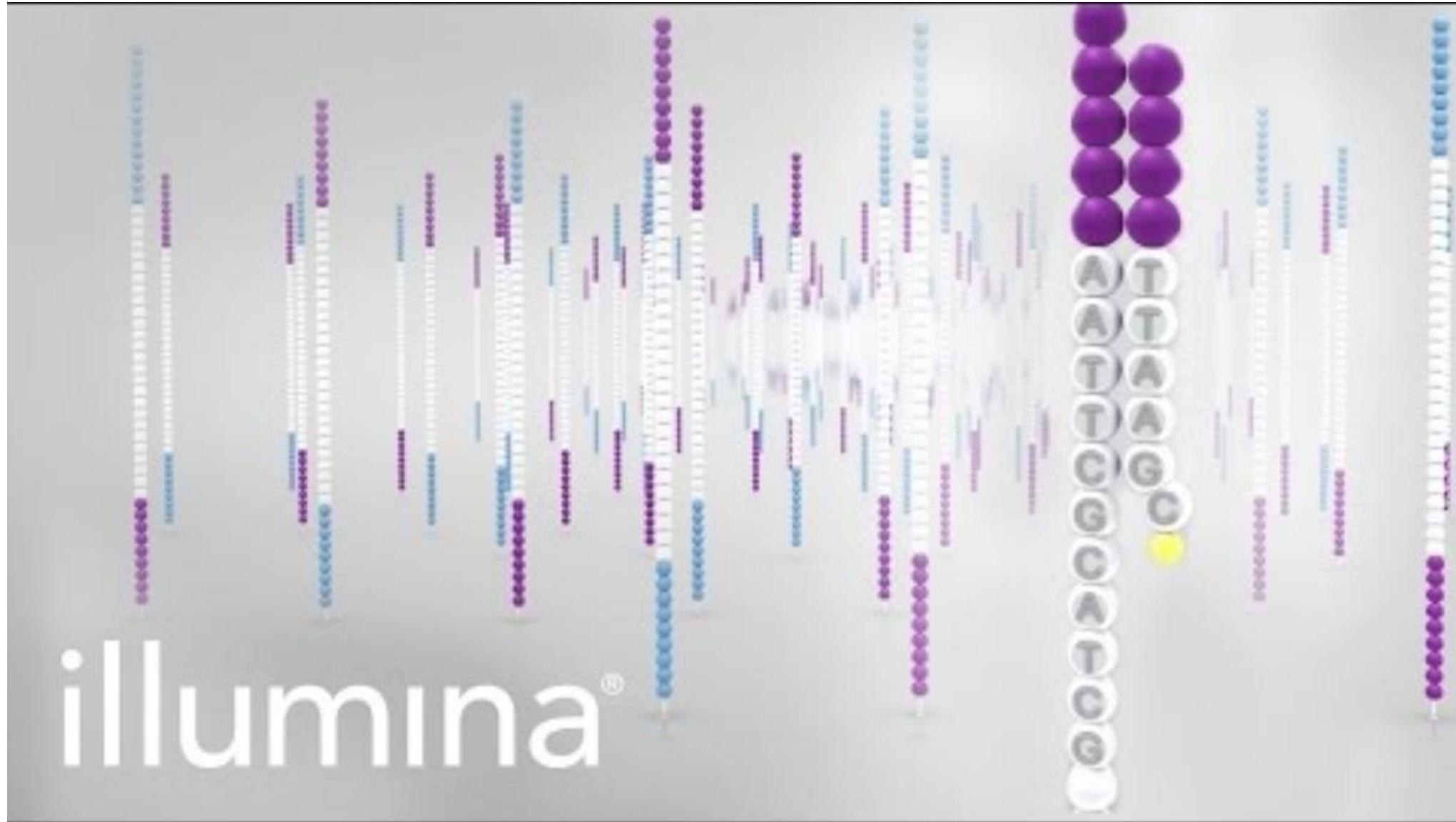
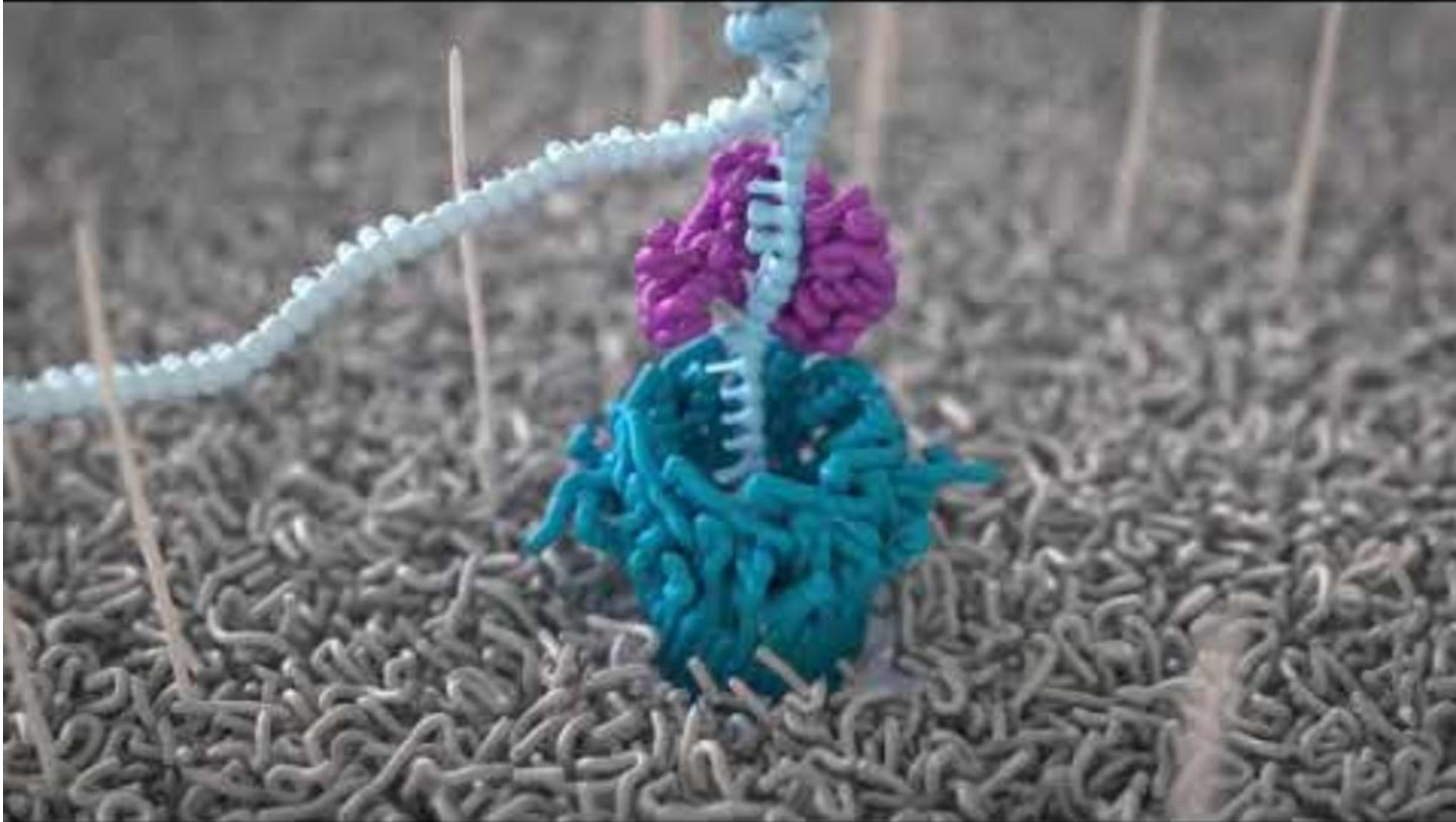


Figure 1 | DNA sequencing technologies. Schematic examples of first, second and third generation sequencing are shown. Second generation sequencing is also referred to as next-generation sequencing (NGS) in the text.



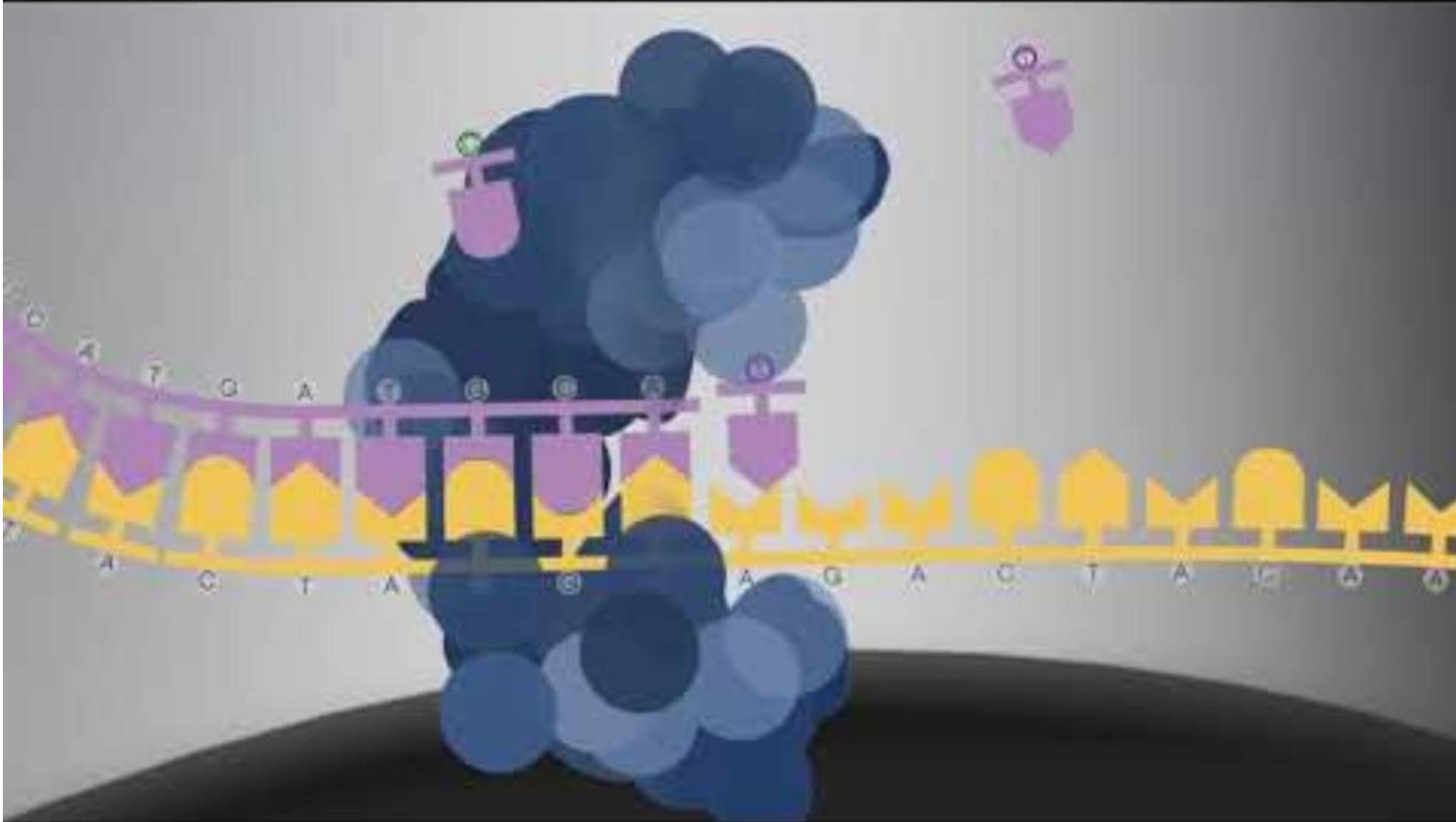
<https://www.youtube.com/watch?v=fCd6B5HRaZ8>

09/07/2024

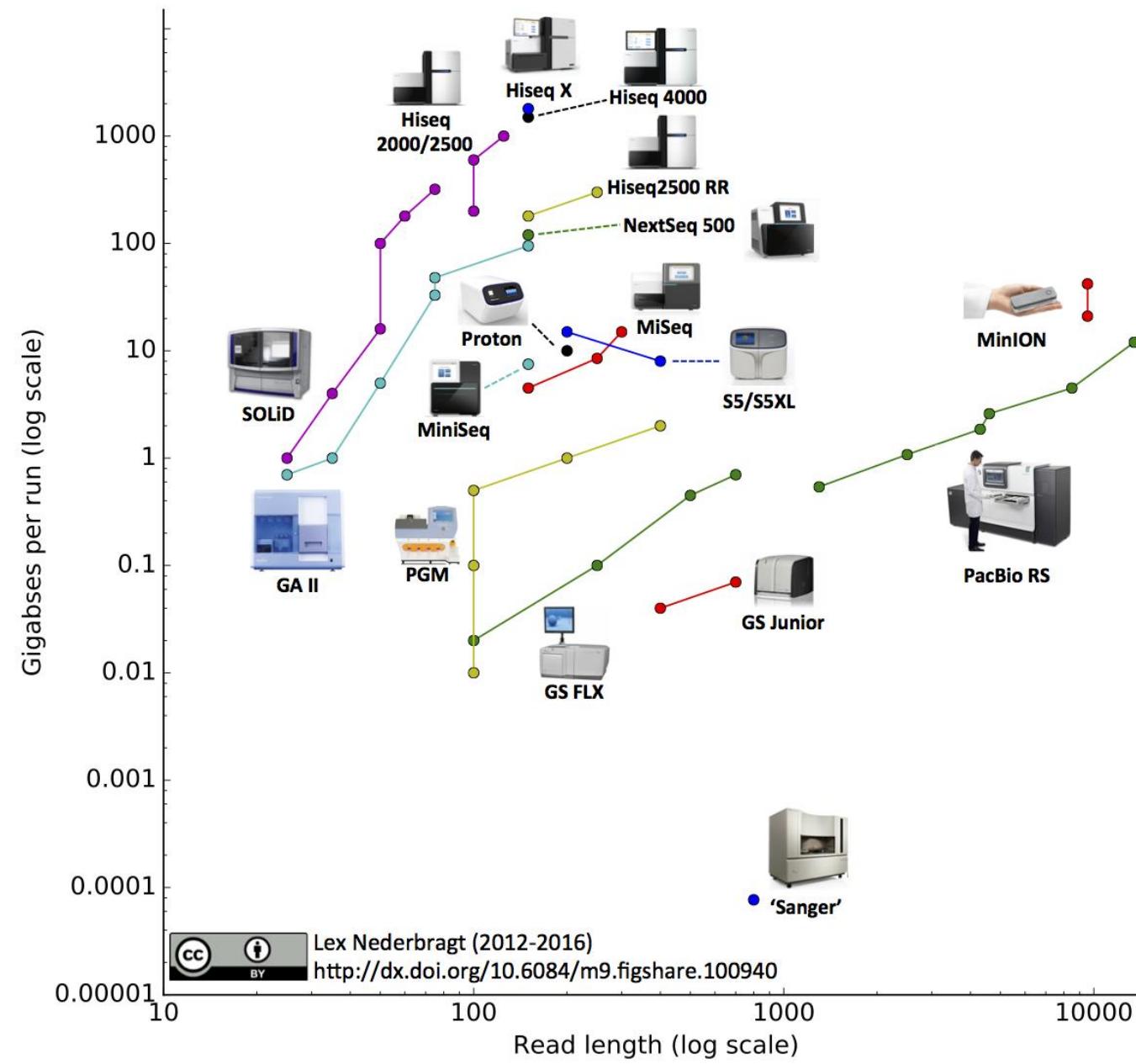


<https://www.youtube.com/watch?v=RcP85JHLmnl>

09/07/2024



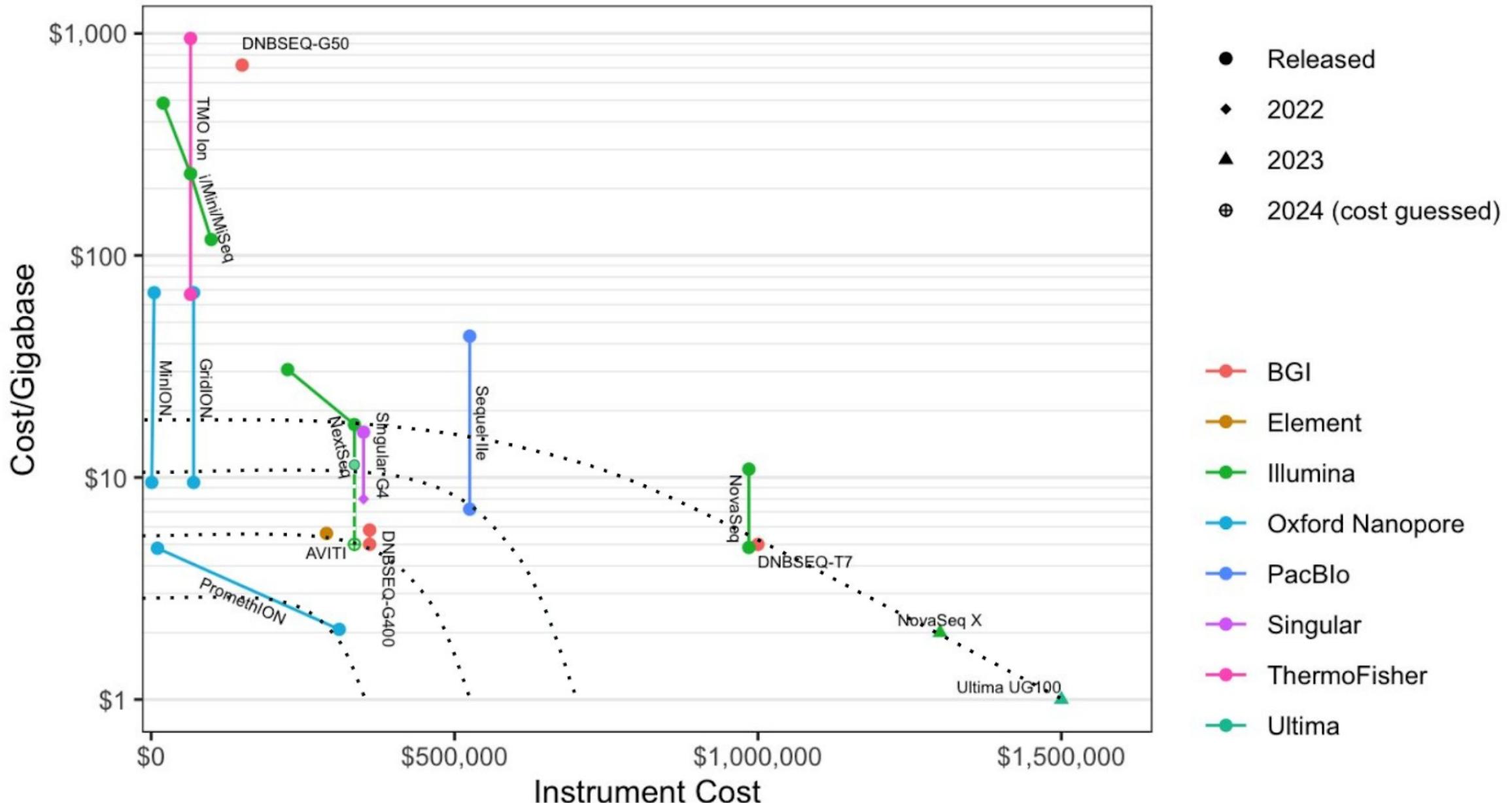
https://www.youtube.com/watch?v=_ID8JyAbwEo



Sequencing Platform	SBS Kit Version	Maximum Read Length
iSeq 100	v1	2 x 151bp
	v2	2 x 151bp
MiniSeq	MO*	2 x 151bp
	HO*	2 x 151bp
MiSeq	v2	2 x 251bp
	v3	2 x 301bp
NextSeq 500/550	MO*	2 x 151bp
	HO*	2 x 151bp
NextSeq 1000/2000	P1, P2	2 x 301bp
	P3	2 x 151bp
HiSeq 1000/1500/2000/2500	HO* v3	2 x 101bp
	HO* v4	2 x 126bp
	RR** v4	2 x 251bp
HiSeq 3000/4000	N/A	2 x 151bp
HiSeq X	N/A	2 x 151bp
NovaSeq 6000	SP	2 x 251bp
	S1, S2, S4	2 x 151bp

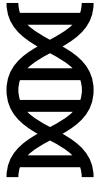
Whole Genome Sequencing Costs

Data from @AlbertVilella (<http://bit.ly/ngsspecs>)
Visualization by @brianlandry23



	Sanger	NGS
Beneficios	<p>Rápido</p> <p>Coste efectivo para un número bajo de genes a analizar (1-20 genes)</p> <p>Análisis de datos sencillo y asequible</p> <p>Lecturas largas (hasta 1500 - 2000 pb)</p> <p>10 ng ADN = 1 Kb de datos</p> <p>96 muestras en paralelo</p>	<p>Profundidad de secuenciación muy alta, por lo que permite una sensibilidad mayor.</p> <p>Alto poder de descubrimiento de nuevas variantes.</p> <p>Resolución de mutaciones mayor (no solo variantes de nucleótido, sino grandes reorganizaciones cromosómicas)</p> <p>10 ng ADN = 300 Kb de datos</p> <p>384 muestras en paralelo</p>
Desventajas	<p>Baja sensibilidad</p> <p>Bajo poder de descubrimiento de novo</p> <p>Para gran número de genes encarecimiento</p> <p>Baja escalabilidad</p>	<p>Poco coste-efectivo para bajo número de genes</p> <p>Lecturas cortas (máximo 2x300 pb)</p>

	<i>Second-generation sequencing</i>	<i>Third-generation sequencing</i>
Ventajas	Alta precisión de secuencia Bajo coste económico Posibilidad de secuenciar ADN fragmentado	Posibilidad de empezar con fragmentos grandes de DNA Fácil preparación de librería (portable) * Epigenética: patrones de metilación y modificaciones de histona Genera secuencias largas, muy largas
Desventajas	Solo produce secuencias cortas (hasta 300bp PE)* No resuelve variantes estructurales No distingue regiones homólogas del genoma Problemas en las regiones altamente repetitivas No detecta epigenética	Baja precisión de secuencia Elevado coste económico



El análisis del genoma humano se realiza por **resecuenciación**, es decir, secuenciamos el genoma o parte de él, y lo comparamos frente a un genoma que consideramos modelo.

El genoma completo mide aproximadamente 3 billones de pares de bases, pero solo un 1-2 % codifica para proteínas. Es lo que denominamos **exoma**. La mayor parte de las variaciones genómicas conocidas alteran la secuencia de la proteína.

Conocemos aún muy poco sobre la función del DNA no-codificante, aunque el proyecto ENCODE nos acerca poco a poco a este conocimiento.

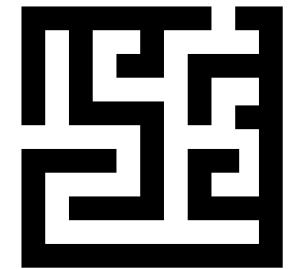
Dado que se estima que el **85% de las mutaciones que causan patología residen en el exoma**, esta técnica suele ser de elección preferente frente a la secuenciación del genoma completo.

Es la opción intermedia entre un coste asequible, cobertura del genoma, rendimiento diagnóstico y utilidad para la interpretación (Seaby et al., 2016).

3.2

Estrategias generales de análisis

De manera general existen tres estrategias para abordar el diagnóstico molecular empleando técnicas de secuenciación masiva. A lo largo de este capítulo desgranaremos sus ventajas e inconvenientes.



3.2.1. Paneles de genes o regiones de interés

3.2.2. Secuenciación de exoma clínico o completo (WES)

3.2.3. Secuenciación de genoma completo (WGS)



Paneles de genes
30 – 200 genes

Exoma clínico

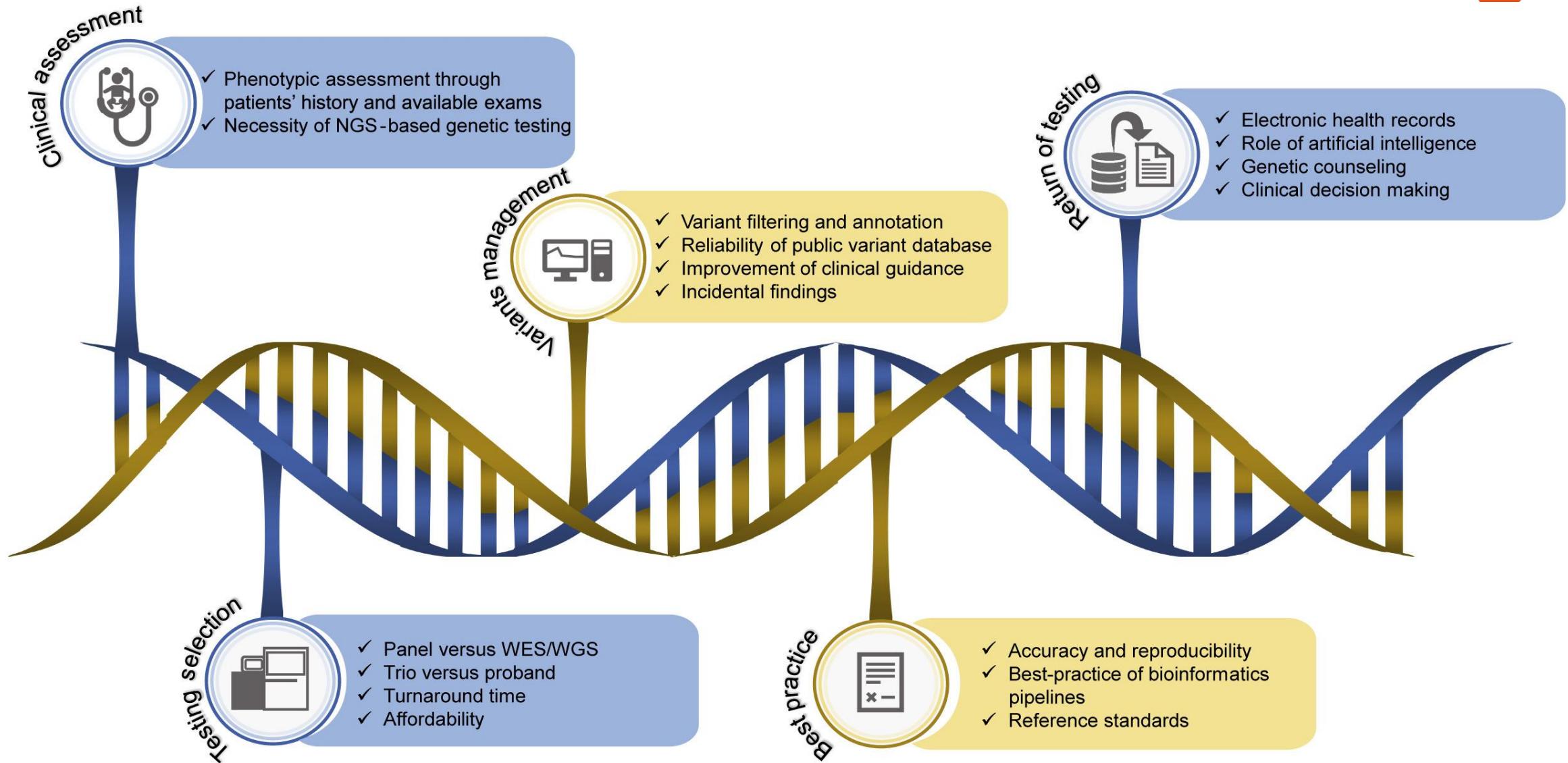
5.000 genes

Exoma WES
(Whole Exome Sequencing)

20.000 genes

Genoma WGS
(Whole Genome Sequencing)
Genes + ADN no codificante





Trends in Genetics

Figure 1.- The Workflow of NGS-Based Genetic Testing in Rare Disease Diagnosis. Liu Z, Trends in Genetics, Nov 2022, Vol. 35, No. 11

09/07/2024

3.2.1. Paneles de genes o regiones de interés

Los paneles de genes o regiones de interés analizan un número limitado de genes que están asociados a un grupo de enfermedades determinadas o fenotipos.

Están indicados en caso de **enfermedades mendelianas** bien descritas genéticamente, para los que se conoce certamente los genes y mutaciones implicadas.

Ventajas:

- Menor coste económico, ya que son regiones bien delimitadas.
- Rapidez.
- Gran cobertura de secuenciación, lo que permite la detección de variantes de baja frecuencia.
- Seleccionar a priori regiones de interés facilita el análisis posterior de datos.

Limitaciones:

- No es posible detectar mutaciones en genes y/o regiones que no han sido amplificadas o capturadas por el diseño inicial.

3.2.2. Secuenciación de exoma clínico o completo (WES)

La **secuenciación del exoma completo**, también llamado WES (Whole Exome Sequencing) es un panel donde se incluyen las regiones codificantes de todos los genes, aproximadamente un 2% del genoma completo.

Existen variaciones sobre este panel para reducirlo a zonas de mayor interés clínico, más focalizado, siendo un intermedio entre el panel de genes y el exoma completo.

Esta es la metodología elegida cuando un panel génico no puede identificar la causa de la enfermedad, ya que permite detectar mutaciones en nuevos genes implicados en la misma. También está indicado en casos cuyo fenotipo implica varios genes o no se conoce una asociación clara con alguna enfermedad previamente descrita.

Ventajas: Es más rápido y barato, tanto en preparación como en análisis, que el estudio del genoma completo.

Limitaciones:

La cobertura no es uniforme

No incluye regiones intrónicas ni reguladoras

En muchos casos no detecta variantes estructurales

3.2.3. Secuenciación de genoma completo (WGS)

La secuenciación del genoma completo, llamada WGS (Whole Genome Sequencing) cubre teóricamente la información completa codificada en el ADN del individuo, **incluyendo regiones intrónicas y reguladoras.**

Ventajas:

- Permite detectar variaciones estructurales complejas, como cambios en el número de copias y traslocaciones cromosómicas.
- Permite detectar mutaciones en genes previamente no asociados a enfermedad, e incluso en regiones intrónicas o reguladoras.

Limitaciones:

- Coste de secuenciación. Pese a que el coste de secuenciar un genoma completo humano ha descendido notablemente en los últimos años, actualmente se tasa en aproximadamente menos de 1000 euros por genoma.
- Complejidad en el análisis de los datos, tanto a nivel computacional (necesidad de gran número de recursos) como a nivel de interpretación de los mismos.



WGS

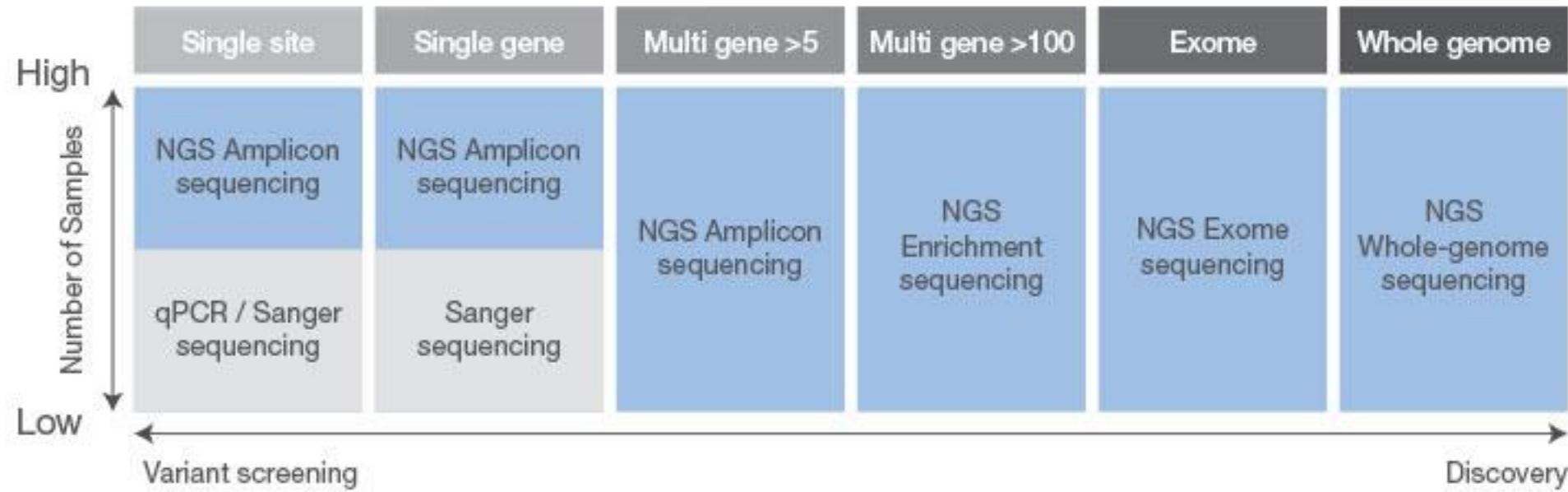
- Variantes a lo largo de todo el genoma (codificante o no)
- Variaciones estructurales (limitaciones short reads)
- Variaciones en número de copias

WES

- Variantes en regiones codificantes y **sitios de splicing**
- **Variaciones en número de copias**

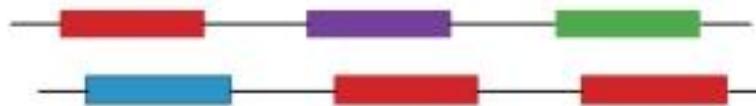
Panel

- Variantes en genes pre-seleccionados
- Deleciones



Targeted panel sequencing

- Categorical genetic disorders
- Up to thousands of genes
- High coverage and depth
- Lowest cost



capture

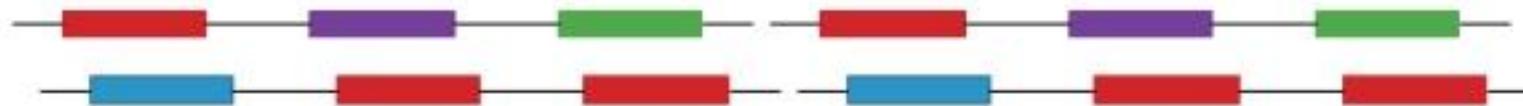


High accuracy



Whole-Exome Sequencing

- Whole exome
- Intermediate coverage and depth



capture

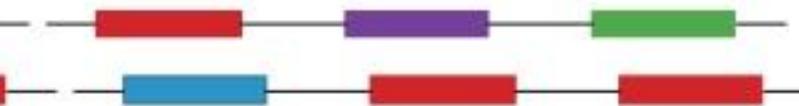


good accuracy

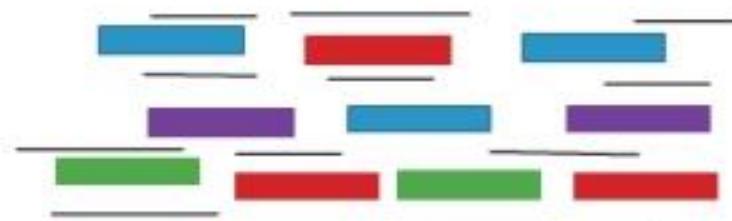


Whole-Genome Sequencing

- All genes and non-coding DNA
- Lower coverage
- Highest cost



No capture



Lower accuracy



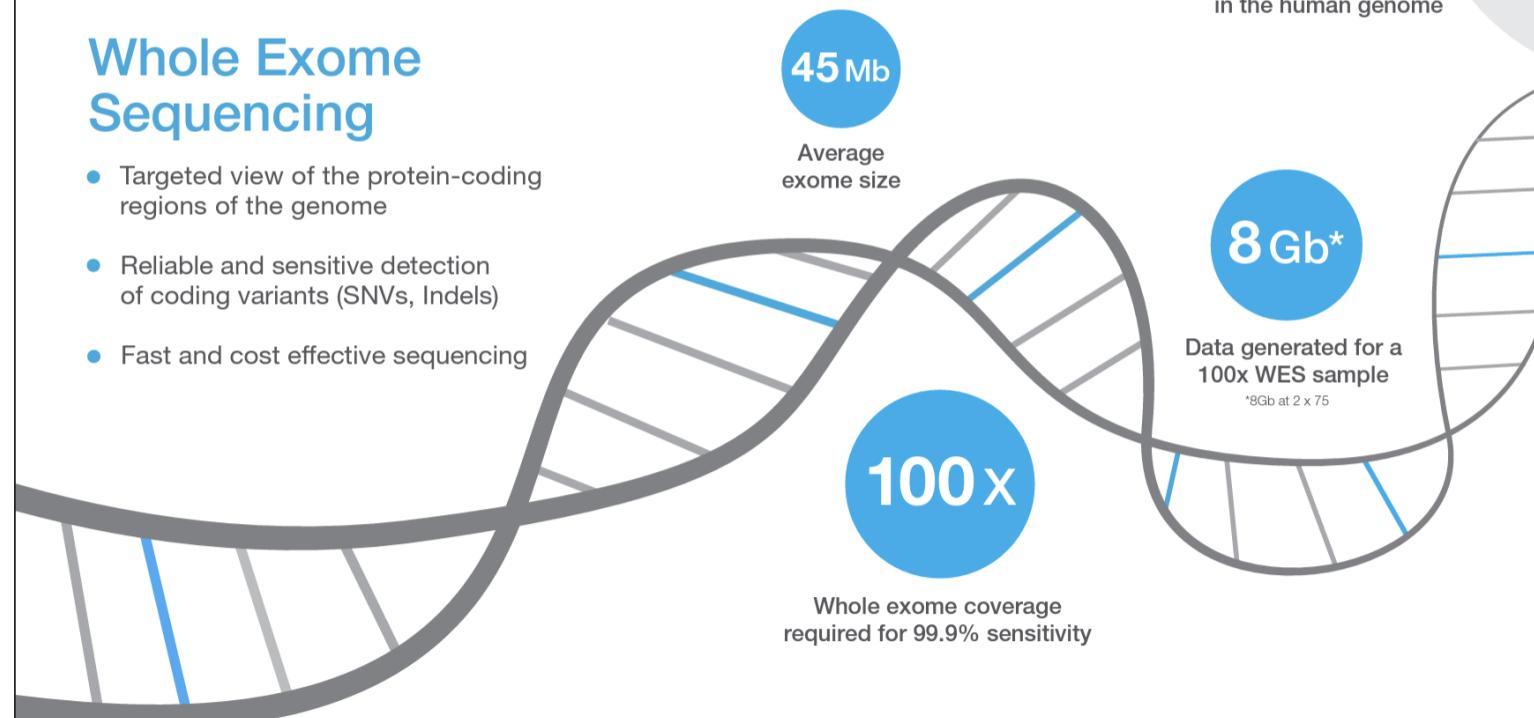
Find the method that's right for your research

Next-generation sequencing (NGS) allows researchers to explore genetic variations like never before – down to single nucleotide resolution. Whether you need to identify specific variants from a focused set of regions, or look more broadly for potential causative variations, whole exome sequencing (WES) or whole genome sequencing (WGS) both offer effective solutions.

Explore the benefits of these approaches to understand which method is best for your research.

Whole Exome Sequencing

- Targeted view of the protein-coding regions of the genome
- Reliable and sensitive detection of coding variants (SNVs, Indels)
- Fast and cost effective sequencing

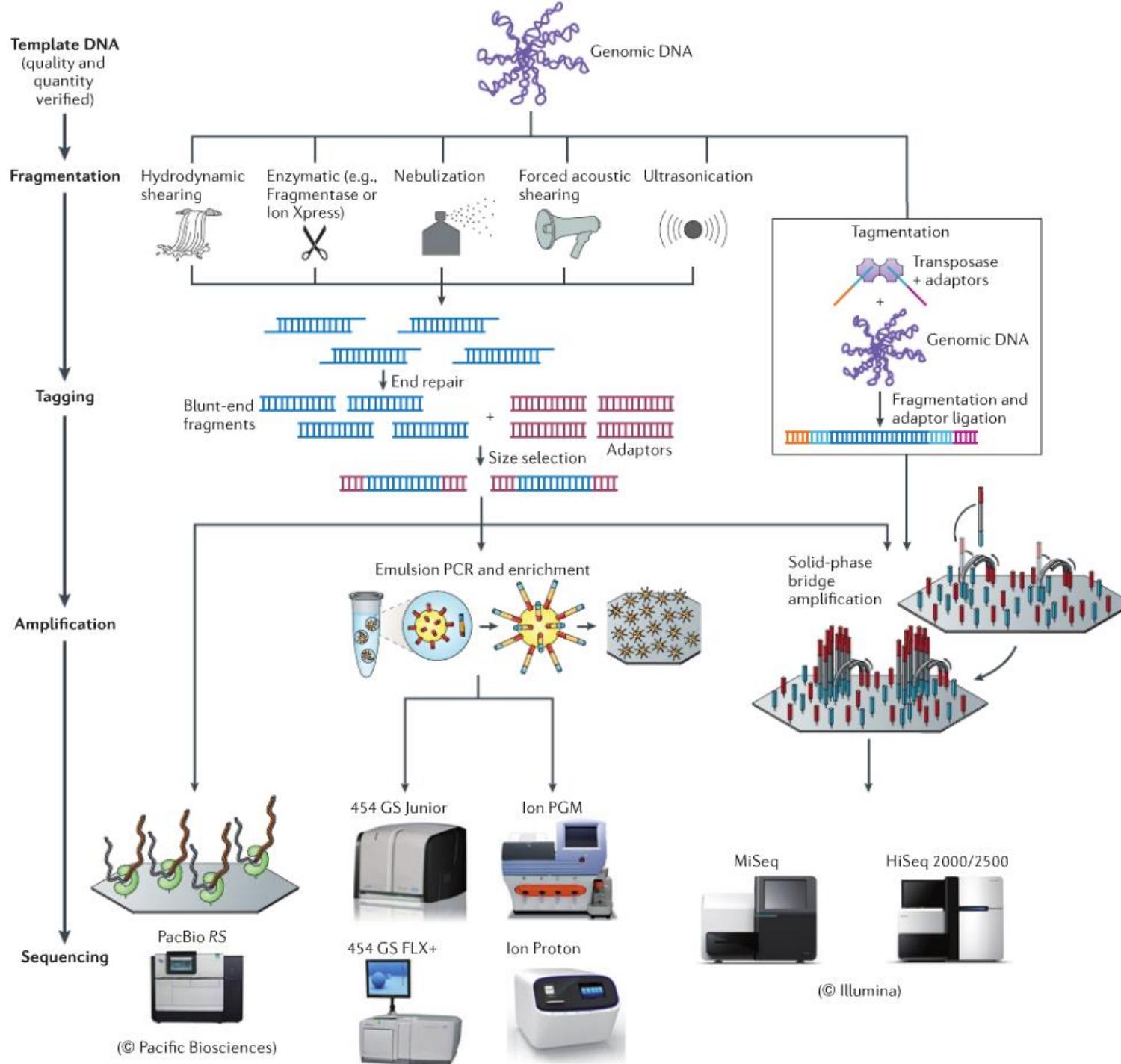


Whole Genome Sequencing

- Comprehensive view of the genome (coding, non-coding and mtDNA)
- Reliable and sensitive detection of all variant types (SNVs, Indels, SVs, CNVs)
- Low cost, fast library preparation

3.3

Protocolos de análisis



3.3.1. Extracción del ADN

El primer paso en cualquiera de los tres protocolos consiste en la **obtención de un ADN de alta calidad** a partir de muestras biológicas.

Habitualmente se obtiene de leucocitos de sangre periférica, aunque también son comunes muestras como saliva o incluso tejidos parafinados (Formamin-Fixed Paraffin-Embedded, FFPE).

La saliva presenta problemas de posible contaminación con ADN de la microbiota oral, y las muestras de FFPE suelen obtenerse de tejidos cancerosos, dando un ADN de peor calidad debido a alta degradación (Seaby et al., 2016).

Este ADN **se cuantifica utilizando fluorimetría** y se **cualifica** utilizando **electroforesis capilar** (Fragment Analyzer, Bioanalyzer).

3.3.2. Preparación de la genoteca/librería

En el caso de los paneles de genes y WES la preparación de la genoteca sigue metodologías similares: amplificación mediante PCR de las regiones de interés (sobre las que previamente hemos realizado un diseño con cebadores específicos) o bien, captura mediante sondas de estas regiones.

En realidad, podemos pensar en que un análisis WES no es más que un panel de genes muy ampliado, con un mayor número de regiones a capturar y secuenciar. Sin embargo, en el proceso de preparación de un genoma completo, no existe una captura de regiones, sino que simplemente se secuencia todo el ADN que tenemos en la muestra.

Targeted panel sequencing

- Categorical genetic disorders
- Up to thousands of genes
- High coverage and depth
- Lowest cost



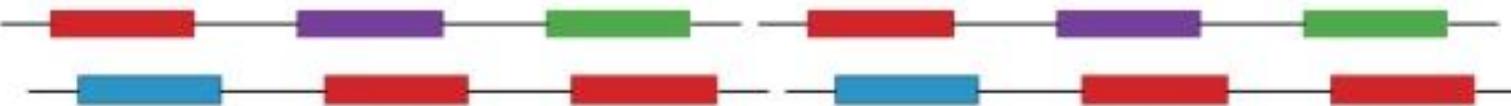
capture



capture

Whole-Exome Sequencing

- Whole exome
- Intermediate coverage and depth



No capture

High accuracy



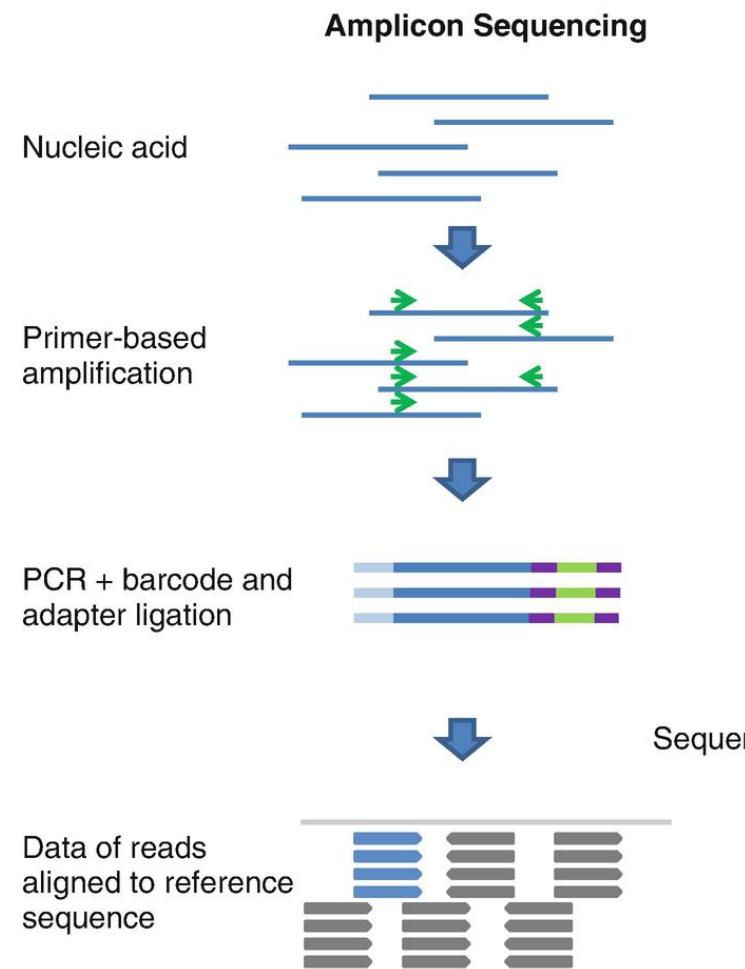
good accuracy



Lower accuracy



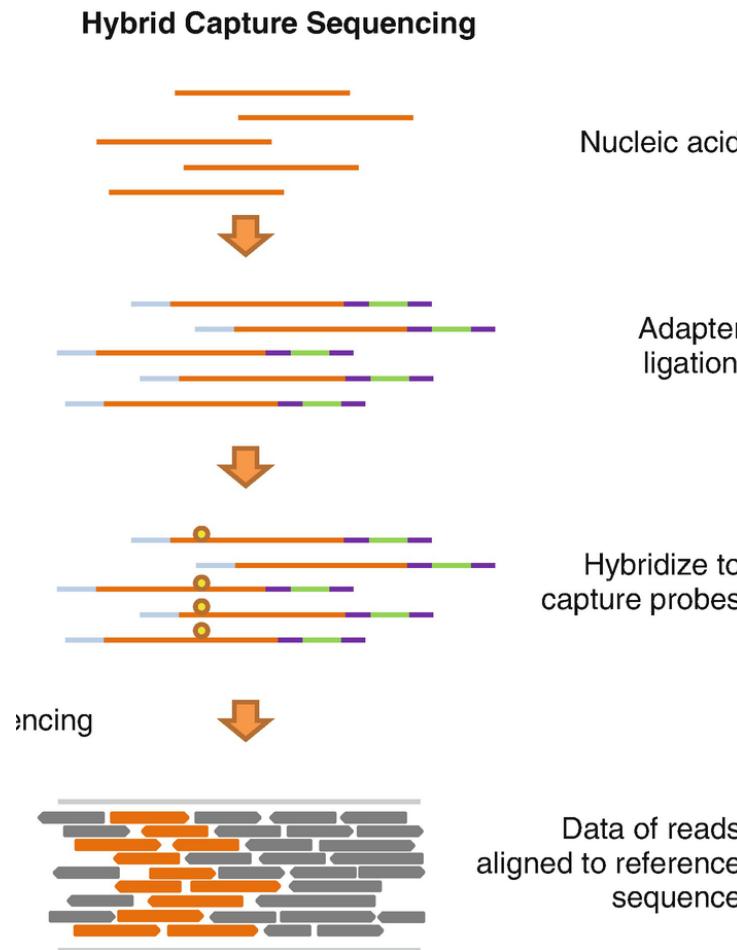
3.3.2. Preparación de la genoteca/librería: WES y paneles



Secuenciación por amplicones.

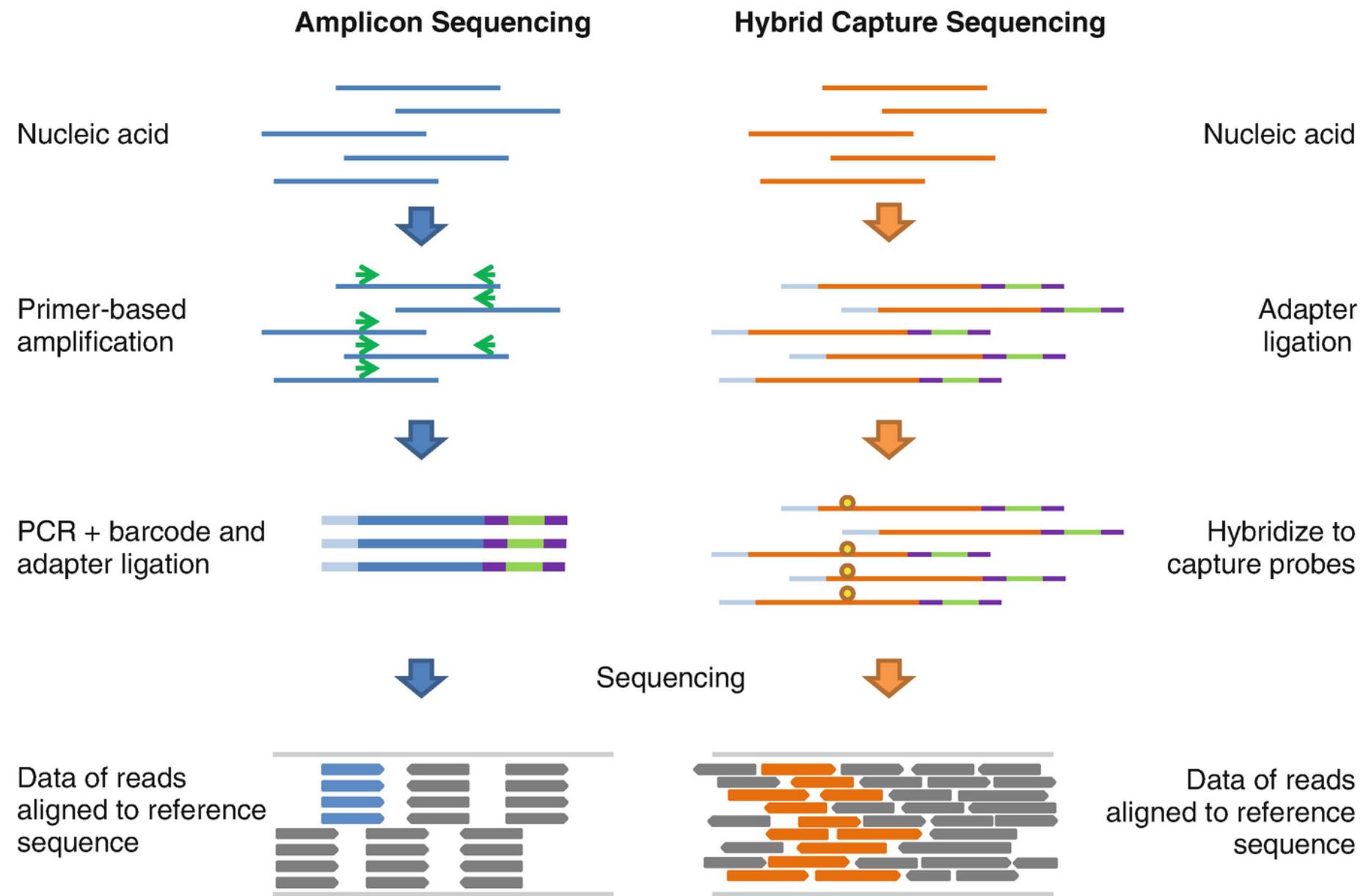
- Se realiza un diseño de cebadores específicos sobre los genes de interés y se amplifican mediante PCR.
- Al pool amplificado se adicionan los adaptadores e índices, previamente a la secuenciación.
- Es un proceso rápido y sencillo, **permite partir de menor cantidad de ADN**, además de tener un menor coste por muestra que otros procedimientos; sin embargo, está limitado a un máximo de unos 10.000 amplicones e inevitablemente el proceso de amplificación por PCR puede producir errores y sesgos.
- Está indicado para el genotipado mediante secuenciación, detección de variantes ya conocidas a enfermedad y detección de SNPs e indels en células germinales.

3.3.2. Preparación de la genoteca/librería: WES y paneles



Captura por hibridación.

- Este tipo de captura se realiza mediante sondas de RNA que hibridan con el ADN de interés, previamente fragmentado bien enzimáticamente o por métodos de sonicación focalizada (método mecánico).
- Tras la captura, se liberan los fragmentos capturados, se adicionan los adaptadores e índices y se secuencia la genoteca.
- En este procedimiento, dependiendo de la concentración del ADN de partida, la amplificación por PCR para enriquecer los fragmentos capturados es opcional.
- Aunque es un proceso más largo y laborioso, así como el coste puede ser más elevado que con amplicones, **su sensibilidad es mayor que la técnica anterior.**
- Está indicado para genotipado, secuenciación de exoma, análisis de mutaciones oncológica, descubrimiento de variantes raras, descubrimiento de genes, detección de variaciones somáticas de baja frecuencia y análisis de número de copias de un gen (Copy Number Variants, CNVs).



3.3.2. Preparación de la genoteca/librería: WES y paneles

Aunque existen muchas soluciones comerciales en el mercado, en la tabla siguiente, adaptada del artículo de (Seaby et al., 2016), se detallan algunas características de los kits de captura mediante sondas de exoma disponibles.

Las principales diferencias entre los kits disponibles radican en dos cuestiones principales.

- La primera es el **tamaño de las regiones capturadas**. Los diseños en **kits de captura de exoma** **recogen los genes más relevantes en patologías clínicas de interés**, pero adicionalmente pueden extenderse a otras regiones, de ahí sus principales diferencias. También, **la captura de lugares promotores o UTR, sitios de splicing o variantes intrónicas**.
- En cuanto a la preparación de la genoteca, el **método de fragmentación** determina la cantidad de ADN inicial del que partir. Los **métodos mecánicos** producen **genotecas con una dispersión de tamaños menor, pero necesitan unas cantidades de ADN inicial mayores** y por supuesto, el equipamiento de sonicación mecánica adecuado, equipamiento de alto coste no disponible en todos los laboratorios. Por otra parte, los **métodos enzimáticos** son más rápidos y menos laboriosos, así como más baratos y **parten de cantidades de ADN menores, pero el tamaño medio de la genoteca resultante es más disperso**.

3.3.2. Preparación de la genoteca/librería: WES y paneles

Tabla 5

Kits de captura de exoma. Comparativa.

Nota. Adaptado de (Seaby et al., 2016). Datos proporcionados por las casas comerciales.

	xGen Exome Research Panel v2 (IDT)	SureSelect Focused Exome (Agilent)	SureSelect Human All Exon V8 (Agilent) ¹	SureSelect Clinical Research Exome V2 (Agilent)	Illumina DNA Prep with Enrichment Exome ⁴ (Illumina)	Illumina DNA Prep with Enrichment Exome ⁵ (Illumina)	TruSeq DNA Exome (Illumina) ³
Tamaño regiones (target size)	34 Mb	12 Mb	35.1 Mb	67.3 Mb	12 Mb	16,5 Mb	45 Mb
Número de sondas	415,115	231,855	436,193	843,912	125,395	183,809	429,826
Número de genes/regiones capturados	19,433	4,800	>20,000 ²	1,009 genes >800 promotores 75K sitios splicing no- codificantes >12,000 variantes intrónicas	4,813	6,704	19,396
Lecturas sobre la diana (reads on target)	>95%	>83%	>60%	>97%	>85%	>85%	>85%
% de pb cubiertas ≥20x	>90%	>97%	>96%	>95%	>93%	>93%	>90%
Método fragmentación	Enzimático	Enzimático/mecánico	Enzimático/mecánico	Enzimático/mecánico	Enzimático	Enzimático	Mecánico
ADN inicial	100 ng	50 ng / 200 ng – 3 ug o 100 ng – 1 ug (QXT, XT, XT2) ³	50 ng / 200 ng – 3 ug o 100 ng – 1 ug (QXT, XT, XT2) ³	50 ng / 200 ng – 3 ug o 100 ng – 1 ug (QXT, XT, XT2) ³	50 ng	50 ng	100 ng

Nota: todos los diseños se han comparado en base al genoma *Homo sapiens* Hg19

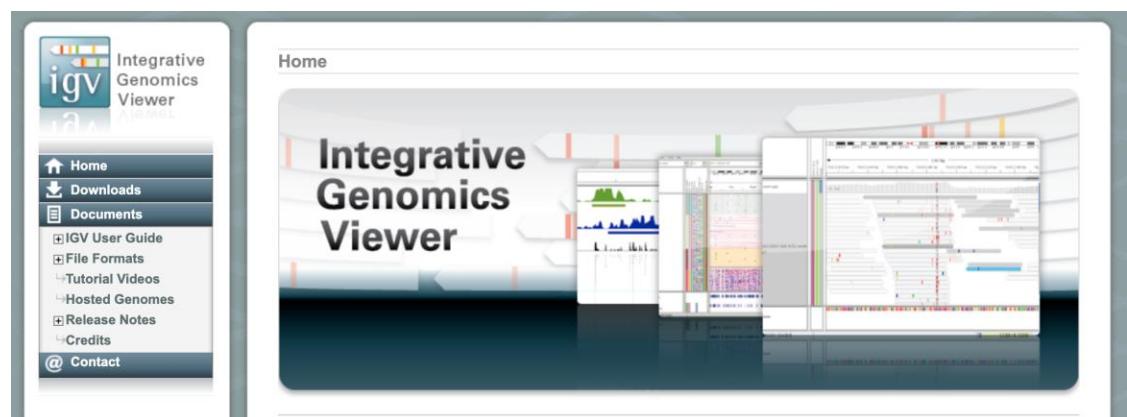
¹Existen versiones del kit "extendidas", donde se cubren también regiones UTR.

Ejemplo: visualizar las sondas de captura

Vamos a visualizar las sondas de captura de exoma sobre el genoma de referencia *Homo sapiens* Hg19 (GRCh37) en el programa IGV. Las sondas están en formato BED.

Vamos a hacer uso de la máquina virtual AWS, para lo que tenemos instalado el programa de visualización IGV en un entorno Conda al llamado "04MBIF_humano".

Nota: IGV es un programa que puede instalarse fácilmente en cualquier ordenador de cualquier tipo.



Install IGV 2.16.1

See the [Release Notes](#) for what's new in each IGV release.

Users of the new M1 Mac: Apple's Rosetta software is required to run the IGV MacOS App that includes Java. If you run IGV with your own Java installation, Rosetta may not be required if your version of Java runs natively on M1.

Linux users: The 'IGV for Linux' download includes AdoptOpenJDK (now Eclipse Temurin) version 11 for x64 Linux. See [their list of supported platforms](#). If this does not work on your version of Linux, download the 'Command line IGV for all platforms' and use it with your own Java installation.

About log4j: IGV versions 2.4.1 - 2.11.6 used log4j2 code that is subject to the log4jShell vulnerability. We recommend using version 2.11.9 (or later), which removed all dependencies on log4j.



Ejemplo: visualizar las sondas de captura

Descarga de archivo

En *Campus virtual > Aula de la asignatura > Recursos y materiales > Materiales_docentes > Ejemplo_Tema3*, podrás encontrar los archivos de extensión BED correspondientes a las sondas de estos kits de captura de exoma que vamos a exemplificar.



Tema 3 - Ejemplo 1

- Archivos adjuntos:
-  [AllExonV8_S33266436_Covered.bed](#) (44,105 MB)
 -  [FocusedExome_S07084713_AllTracks.bed](#) (56,08 MB)
 -  [IDT_probes_Hg19.bed](#) (15,293 MB)
 -  [SureSelect_ClinicalResearchExomeV2_S30409818_AllTracks.bed](#) (59,623 MB)
 -  [TruSeq_Exome_TargetedRegions_v1.2.bed](#) (10,779 MB)
 -  [TruSight_One_TargetedRegions_v1.1.bed](#) (3,257 MB)

Os adjunto los archivos BED con las sondas de captura de distintos paneles .

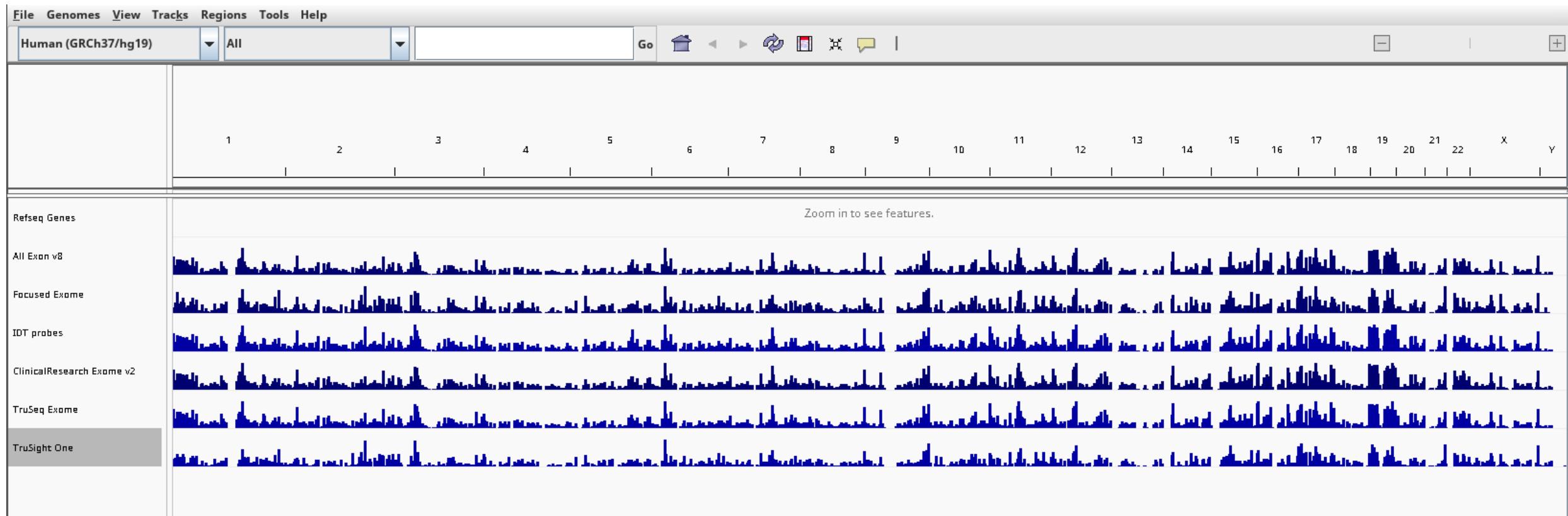
Importante! Genoma de referencia de estos archivos Hg19

En caso de error tenéis estos archivos en: [Tema3_Ejemplo1_kitsExoma](#)

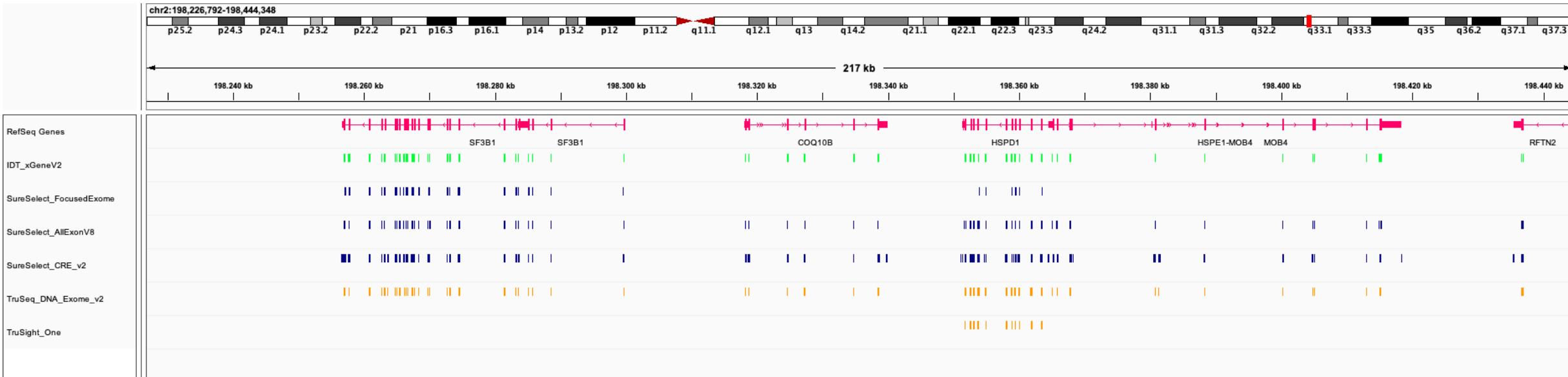
Abrimos IGV Desde la terminal y cargamos el genoma de referencia Hg19 (por defecto).

Cargamos los archivos de sondas en formato BED. Recomendación: cargar una a una y nombrar el “track” adecuadamente.

3.3. Protocolo de análisis

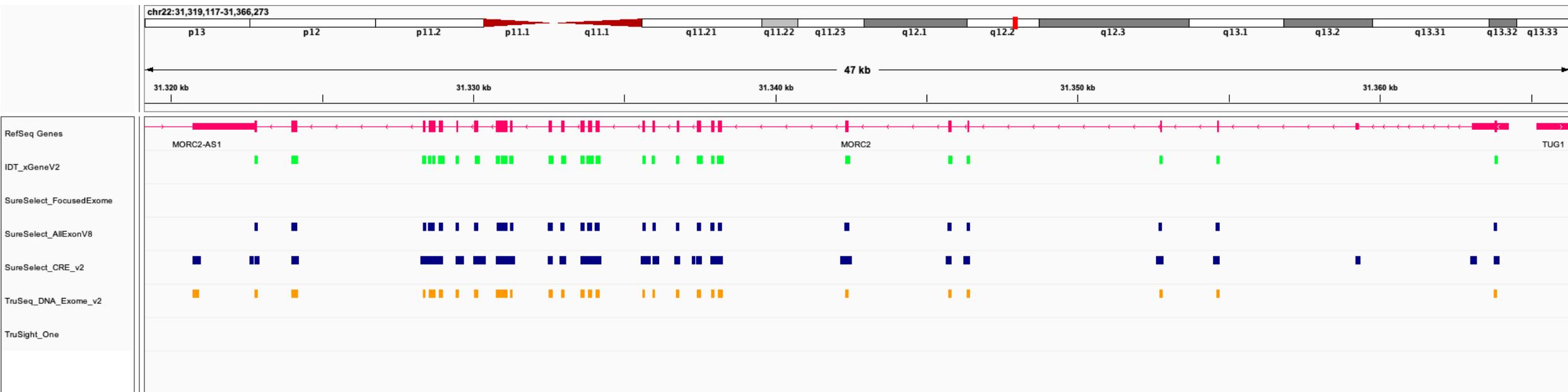


(A) La región del genoma Hg19 representada es chr2:198,226,792-198,444,348



Se ha tomado una región del cromosoma 2 (posiciones 198 226 792 a 198 444 348), donde ya podemos ver que no todos los diseños de panel cubren de la misma manera todos los genes de esta zona. El gen *COQ10B*, codificador de la proteína Coenzima Q10B, no está cubierto por los paneles Sure Select Focused Exome (Agilent) ni por TruSight One (Illumina); mientras que el panel SureSelect Clinical Research Exome (CRE) v2 cubre adicionalmente una región más extendida de su zona 3'-UTR. Una situación similar ocurre con el gen *HSPD1*, presente también en este cromosoma. Aunque todos los paneles presentan sondas frente a él, no todos lo cubren de igual manera. Este gen, codificador de una chaperonina, está involucrado en leucodistrofias y paraplejias (<https://www.omim.org/entry/118190?search=hspd1&highlight=hspd1>).

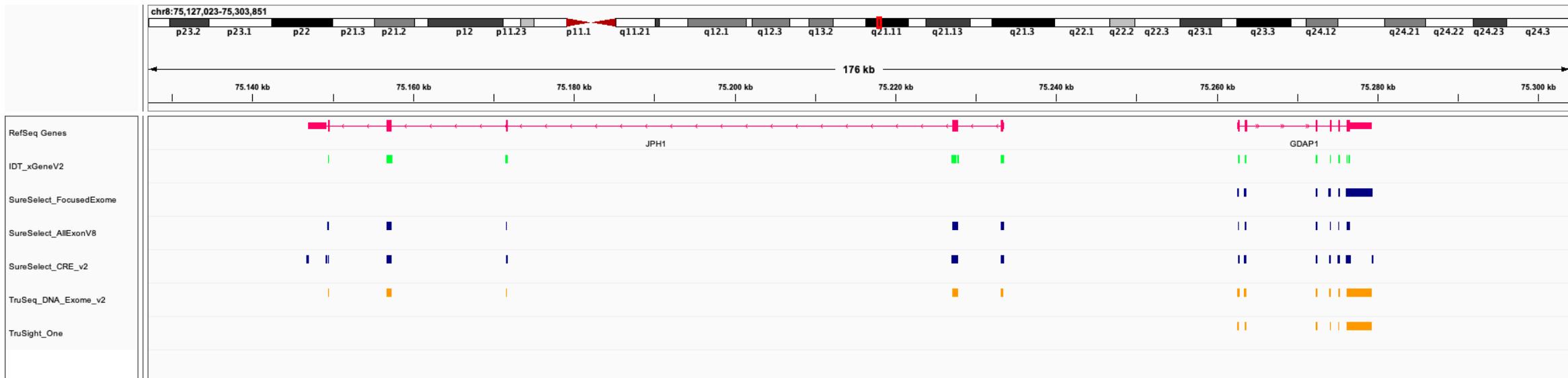
Por lo tanto, si es uno de los genes que está dentro del enfoque de nuestro estudio debemos ser cautos en analizar qué sondas debemos adquirir para cubrirlo correctamente



Aquí tenemos la zona de sondas del gen *MORC2* (chr22:31,319,117-31,366,760), involucrado en un subtipo de la enfermedad de Charcot-Marie-Tooth, denominada axonal (<https://www.omim.org/entry/616661?search=morc2&highlight=morc2>).

El panel de sondas “Focused Exome” ya veis que no cubre ninguna región de este gen. Sin embargo, el resto de paneles, en mayor o menor medida si que lo hacen.

Comparativa de sondas que cubren la región chr8:75,127,023-75,303,851, involucrado en la enfermedad de Charcot-Marie-Tooth.



Los genes *JPH1* y *GDAP1*, localizados en la región aproximada chr8:75,127,023-75,303,851 también se correlacionan con esta neuropatía, y deben ser analizados conjuntamente con el gen previamente mencionado.

Sin embargo, como vemos en la Figura, solo *GDAP1* está cubierto en todos los kits mostrados.

¿Creéis que todos estos kits os permitirían el diagnóstico certero de esta neuropatía de igual manera? Si sospecháis de ella, ¿utilizaríais cualquiera de ellos?

Ejemplo: visualizar las sondas de captura

Adicionalmente, en este link, donde se puede visualizar el genoma humano en sus distintas versiones, os permite analizar las sondas de los kits comerciales más frecuentemente utilizados.

https://genome-euro.ucsc.edu/cgi-bin/hgTrackUi?hgSID=277312691_qnaKZAVtoze5TjXzI2ZxP3t2eqrw&db=hg38&c=chr1&g=exomeProbesets

Ejemplo: visualizar las sondas de captura

Exome Probesets Track Settings

Exome Capture Probesets and Targeted Region ([^All Mapping and Sequencing tracks](#))

Display mode: [Reset to defaults](#)

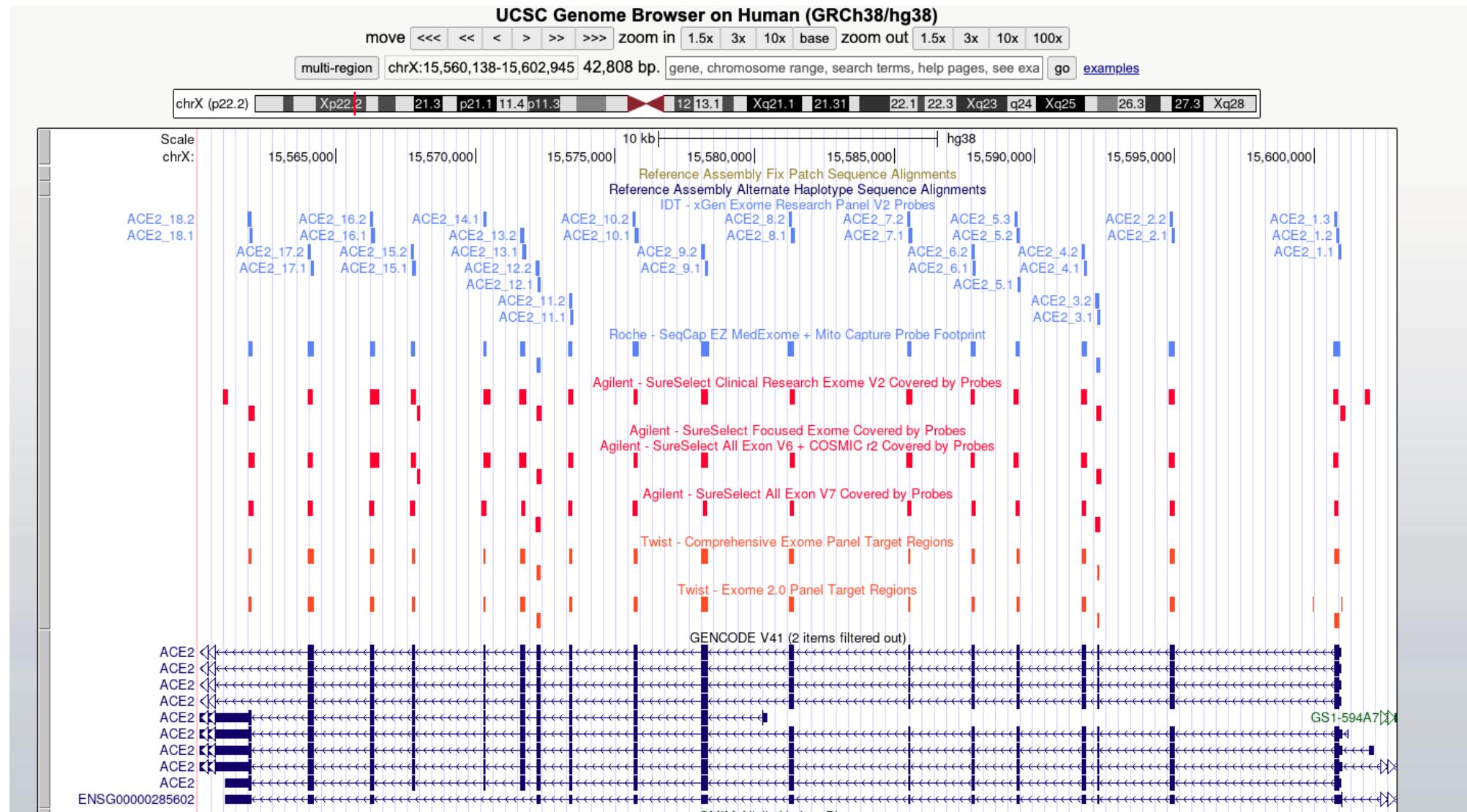
Select all subtracks

List subtracks: only selected/visible all (10 of 28 selected)

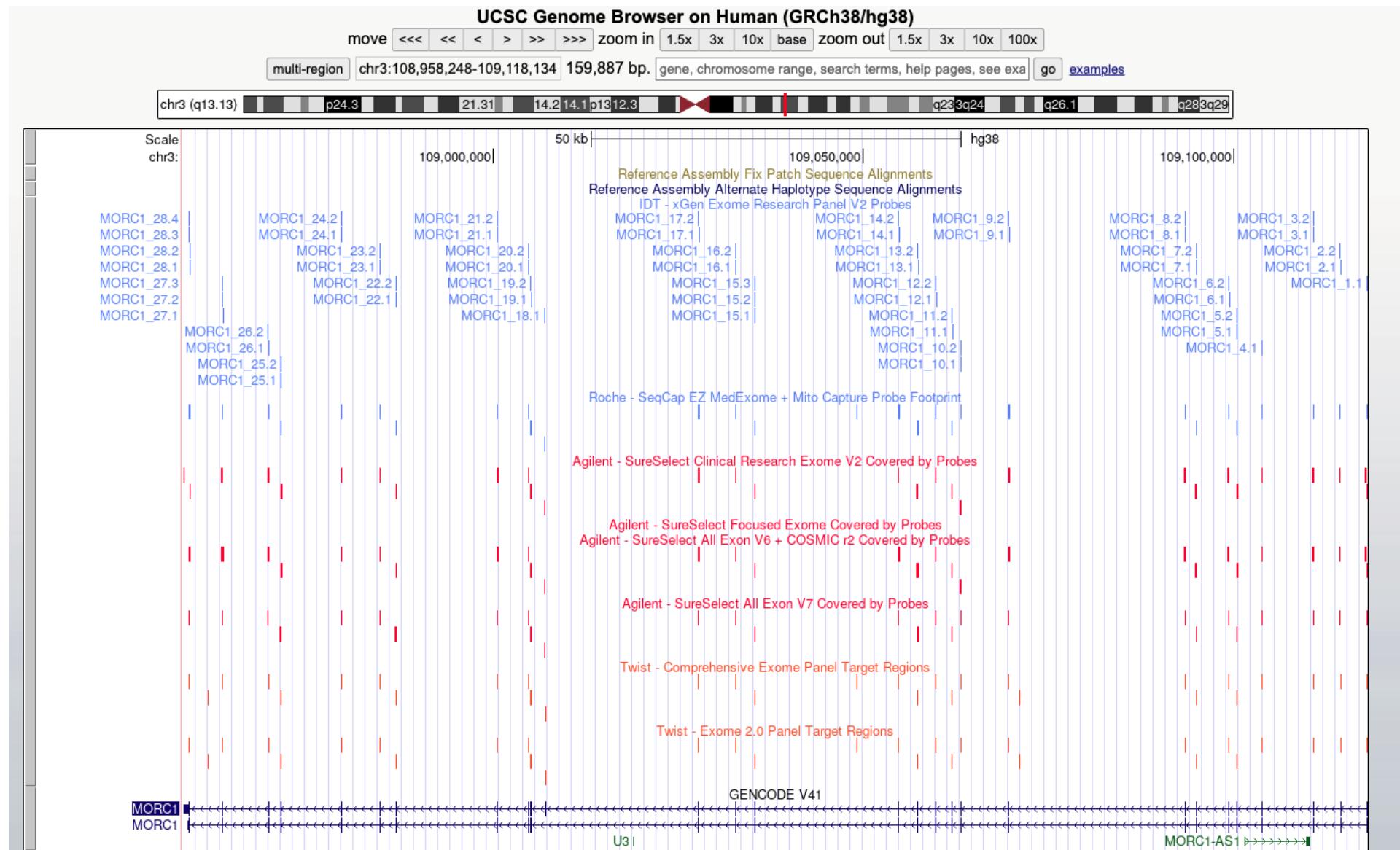
<input checked="" type="checkbox"/> hide	IDT xGen V2 P	IDT - xGen Exome Research Panel V2 Probes	Schema
<input checked="" type="checkbox"/> hide	IDT xGen V2 T	IDT - xGen Exome Research Panel V2 Target Regions	Schema
<input type="checkbox"/> hide	KAPA Hyper P	Roche - KAPA HyperExome Capture Probe Footprint	Schema
<input type="checkbox"/> hide	KAPA Hyper T	Roche - KAPA HyperExome Primary Target Regions	Schema
<input type="checkbox"/> hide	SeqCap EZ Med P	Roche - SeqCap EZ MedExome Capture Probe Footprint	Schema
<input type="checkbox"/> hide	SeqCap EZ Med T	Roche - SeqCap EZ MedExome Empirical Target Regions	Schema
<input checked="" type="checkbox"/> hide	SeqCap EZ Med+Mito P	Roche - SeqCap EZ MedExome + Mito Capture Probe Footprint	Schema
<input checked="" type="checkbox"/> hide	SeqCap EZ Med+Mito T	Roche - SeqCap EZ MedExome + Mito Empirical Target Regions	Schema
<input checked="" type="checkbox"/> hide	SureSel. Clinical V2 P	Agilent - SureSelect Clinical Research Exome V2 Covered by Probes	Schema
<input checked="" type="checkbox"/> hide	SureSel. Clinical V2 T	Agilent - SureSelect Clinical Research Exome V2 Target Regions	Schema
<input type="checkbox"/> hide	SureSel. Focused P	Agilent - SureSelect Focused Exome Covered by Probes	Schema
<input type="checkbox"/> hide	SureSel. Focused T	Agilent - SureSelect Focused Exome Target Regions	Schema
<input type="checkbox"/> hide	SureSel. V4+UTR P	Agilent - SureSelect All Exon V4 + UTRs Covered by Probes	Schema
<input type="checkbox"/> hide	SureSel. V4+UTR T	Agilent - SureSelect All Exon V4 + UTRs Target Regions	Schema
<input type="checkbox"/> hide	SureSel. V5+UTR P	Agilent - SureSelect All Exon V5 + UTRs Covered by Probes	Schema
<input type="checkbox"/> hide	SureSel. V5+UTR T	Agilent - SureSelect All Exon V5 + UTRs Target Regions	Schema
<input type="checkbox"/> hide	SureSel. V6 +UTR T	Agilent - SureSelect All Exon V6 + UTR r2 Target Regions	Schema
<input type="checkbox"/> hide	SureSel. V6 P	Agilent - SureSelect All Exon V6 r2 Covered by Probes	Schema
<input type="checkbox"/> hide	SureSel. V6 T	Agilent - SureSelect All Exon V6 r2 Target Regions	Schema
<input type="checkbox"/> hide	SureSel. V6+COSMIC P	Agilent - SureSelect All Exon V6 + COSMIC r2 Covered by Probes	Schema
<input type="checkbox"/> hide	SureSel. V6+COSMIC T	Agilent - SureSelect All Exon V6 + COSMIC r2 Target Regions	Schema
<input type="checkbox"/> hide	SureSel. V6+UTR P	Agilent - SureSelect All Exon V6 + UTR r2 Covered by Probes	Schema
<input checked="" type="checkbox"/> hide	SureSel. V7 P	Agilent - SureSelect All Exon V7 Covered by Probes	Schema
<input checked="" type="checkbox"/> hide	SureSel. V7 T	Agilent - SureSelect All Exon V7 Target Regions	Schema
<input checked="" type="checkbox"/> hide	Twist Compr. T	Twist - Comprehensive Exome Panel Target Regions	Schema
<input type="checkbox"/> hide	Twist Core T	Twist - Bioscience - Core Exome Panel Target Regions	Schema
<input checked="" type="checkbox"/> hide	Twist Exome 2.0	Twist - Exome 2.0 Panel Target Regions	Schema
<input type="checkbox"/> hide	Twist RefSeq T	Twist - RefSeq Exome Panel Target Regions	Schema

10 of 28 selected

Ejemplo: visualizar las sondas de captura

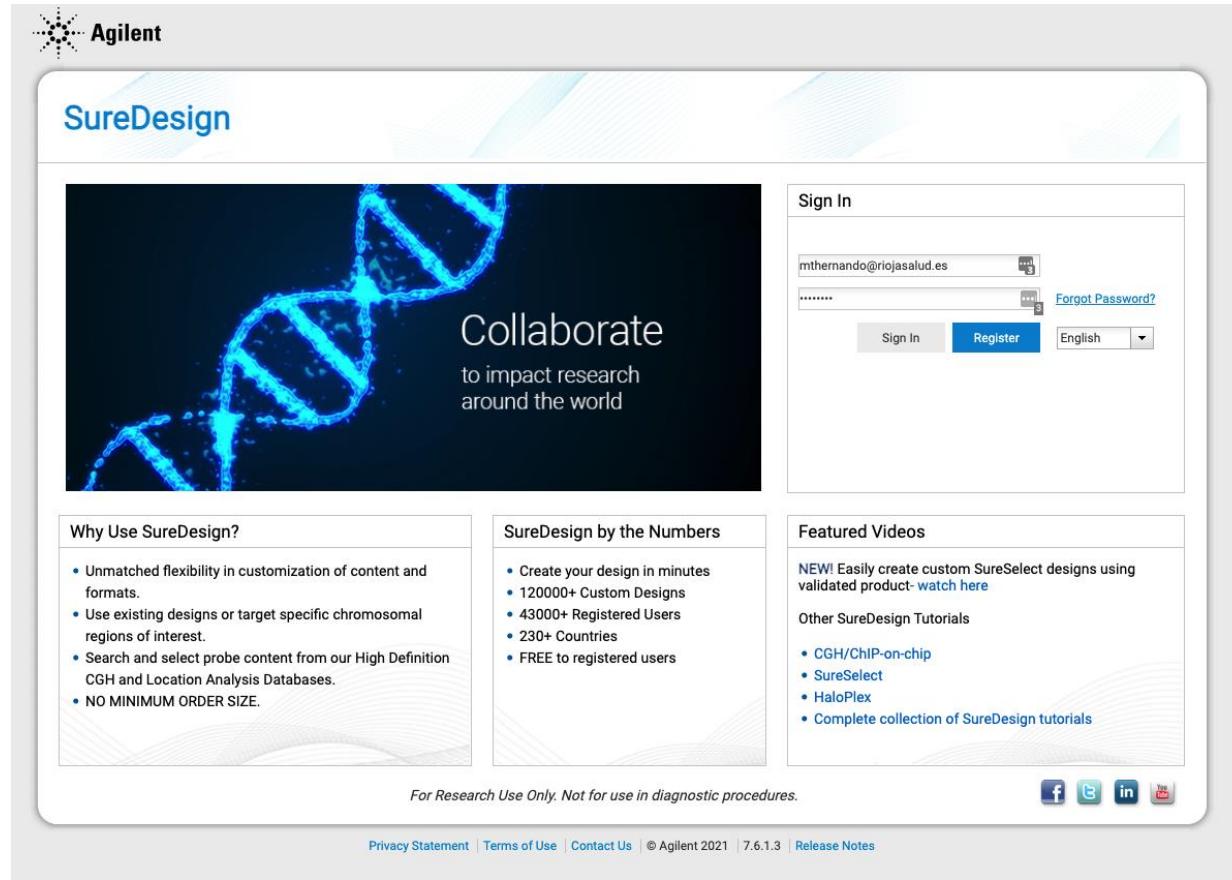


Ejemplo: visualizar las sondas de captura



Ejemplo: Diseño de un panel de captura

De manera análoga a estos paneles de captura de exoma, los paneles de captura de genes aislados siguen la misma metodología. Las casas comerciales principales permiten en su web diseñar sondas para los genes seleccionados, simplemente incluyendo el identificador de los genes de interés.



The screenshot shows the Agilent SureDesign website. At the top left is the Agilent logo. The main header reads "SureDesign". Below the header is a large image of a DNA double helix with the text "Collaborate to impact research around the world". To the right of the image is a "Sign In" form with fields for email and password, and buttons for "Sign In", "Register", and language selection ("English"). Below the sign-in form is a "Featured Videos" section with a link to "watch here". On the left, there's a "Why Use SureDesign?" section listing benefits like unmatched flexibility, existing designs, probe content from databases, and no minimum order size. In the center, there's a "SureDesign by the Numbers" section with statistics such as over 120,000 custom designs, 43,000 registered users, 230+ countries, and it's free for registered users. At the bottom, there's a note "For Research Use Only. Not for use in diagnostic procedures." and social media links for Facebook, Twitter, LinkedIn, and YouTube. The footer includes links to Privacy Statement, Terms of Use, Contact Us, and Release Notes, along with a copyright notice for Agilent 2021.

SureDesign



Home

Find
DesignsCreate
Designs

Create Designs

SureSelect DNA

 Show Advanced Options | Maria de Toro (CIBIR-Genomics Platform) Cart (0) | Logout | Settings

1. Choose Wizard

 Standard

The standard wizard takes you through the steps of creating a design. You simply define the target regions, set the probe selection parameters, and then submit the probe selection job to SureDesign.

 Advanced

The advanced wizard allows you to create a design composed of multiple probegroups. You can select new probes for defined target regions; you can upload probes that you have designed yourself; or you can include probegroups that already exist in your account.

2. Choose Design or Probegroup

 Create Design

A design is a set of one or more probegroups that is manufactured as a single library for use in a SureSelect DNA or HaloPlex protocol.

 Create Probegroup

A probegroup is a set of probes that were selected as part of the same probe selection job or that were uploaded into SureDesign as part of the same text file. Probegroups cannot be manufactured without including them in a design.

 Create Combined Design

A combined design is the set of base and spike-in design which can be used to order as a single design in SureSelect DNA.

 Create OneSeq Design

A OneSeq design is comprised of a set of CNV backbone plus a custom or catalog spike-in design.

 Create All-In-One Design

An All-In-One design is comprised of probes designed for specific types of mutations including SNV, gene level CNV, and translocation.



SureSelect DNA (Advanced)

Define Design

▶ Define Design

Add/Review Content

Select Optimized Probes

Tile Genes or Regions

Upload Probes

Select Existing Probes

Finalize

SureSelect DNA Design

Name: NA
Species: NA
Category: SureSelect DNA
Hybridization: NA

Probes

Probes: NA
Size: NA
Price Tier: [\(i\)](#) NA

UCSC View

Download

* Design Name: [\(i\)](#)

* Species: [Select](#) [\(i\)](#)

Build: [\(i\)](#)

* Create In: [Select](#)

Hybridization: [\(i\)](#) SureSelect XT HS2 / XT HS / XT LI / QXT

Description:

Keywords:

Cancel

Next

SureDesign

Help - Add/Review Content

SureSelect DNA (Advanced) Add/Review Content

Define Design 

► Add/Review Content

- Select Optimized Probes
- Design new probes by tiling genes or regions
- Upload probes
- Select probes from an existing Design or Probegroup

Select probes for specific genes or regions

SureSelect DNA Design

Name: prueba_clase

Species: H. sapiens (hg19)

Category: SureSelect DNA

Hybridization: 90 Minutes

Probes

Probes: NA

Size: NA

Price Tier:  NA

UCSC View Download

Cancel Back Next

SureDesign

SureSelect DNA (Advanced) > Add/Review Content > Tile Genes or Regions

Define Design ✓
 Add/Review Content ✓
 Select Optimized Probes
 ▶ Tile Genes or Regions
 Upload Probes
 Select Existing Probes
 Finalize

* Targets:

- BRCA2
- CDKN2A
- EPCAM
- MLH1
- MSH2
- MSH6
- PALB2
- STK11
- TP53

Upload

#Prostate cancer

- ATM
- BRCA1
- BRCA2
- CHEK2
- HOXB13
- PALB2
- EPCAM
- MLH1
- MSH2
- MSH6
- PMS2

Example

Clear

SureSelect DNA Design

Name: prueba_clase
 Species: H. sapiens (hg19)
 Category: SureSelect DNA
 Hybridization: 90 Minutes

Probes

Probes: NA
 Size: NA
 Price Tier:  NA

Databases

- RefSeq
- Ensembl
- CCDS
- Gencode
- VEGA
- SNP
- CytoBand

Regions of interest

Coding Exons
 Coding Exons + UTRs 5' UTR 3' UTR
 Entire Transcribed Region

Include Flanking Bases:
 3': bp 5': bp

Allow Synonyms

UCSC View Download

Cancel Back Next

#Breast cancer in woman	#Endometrial cancer	#Melanoma
ATM	BRCA1	BAP1
BARD1	EPCAM	BRCA2
BRCA1	MLH1	CDK4
BRCA2	MSH2	CDKN2A
BRIP1	MSH6	PTEN
CHEK2	PMS2	TP53
CDH1	PTEN	
NF1	STK11	#Pancreatic cancer
PALB2		ATM
PTEN	#Ovarian cancer	BRCA1
RAD51C	ATM	BRCA2
RAD51D	BRCA1	CDKN2A
STK11	BRCA2	EPCAM
TP53	BRIP1	MLH1
	EPCAM	MSH2
#Breast cancer in men	MLH1	MSH6
BRCA1	MSH2	PALB2
BRCA2	MSH6	STK11
CHEK2	NBN	TP53
PALB2	PALB2	
	RAD51C	#Prostate cancer
#Colorectal cancer	RAD51D	ATM
APC	STK11	BRCA1
EPCAM		BRCA2
MLH1	#Gastric cancer	CHEK2
MSH2	APC	HOXB13
MSH6	CDH1	PALB2
PMS2	STK11	EPCAM
CHEK2	EPCAM	MLH1
PTEN	MLH1	MSH2
STK11	MSH2	MSH6
TP53	MSH6	PMS2
MUTYH	PMS2	

SureDesign

SureSelect DNA (Advanced) Add/Review Content Tile Genes or Regions Help - Tile Genes or Regions

Define Design ✓
 Add/Review Content ✓
 Select Optimized Probes
► Tile Genes or Regions
 Upload Probes
 Select Existing Probes
 Finalize

Target Summary

- 26 Target IDs resolved to 31 targets comprising 579 regions.
- 0 Target IDs were not found.

Target Details

[View targets in UCSC](#)

Target ID	# Regions	Base Pairs	Position
APC	20	13475	chr5:112043170-112181961
APC	7	2852	chr2:128175971-128186847
ATM	63	25504	chr11:108093186-108239854
BAP1	52	20473	chr3:52434999-66024618
BAP1	8	4473	chr1:185014471-185071765
BARD1	13	6712	chr2:215590345-215674453
BRCA1	26	9578	chr17:41196287-41277525
BRCA2	29	13715	chr13:32889586-32974428
BRIP1	24	11325	chr17:59756475-59940914
BRIP1	3	1565	chr5:1798474-1801505
CDH1	20	7177	chr16:68771103-68869476
CDH1	15	7265	chr19:3506246-3538357
CDK4	9	3653	chr12:58141485-58149821
CDKN2A	9	4908	chr9:21967726-21995348
CHEK2	22	5445	chr22:29083706-29138435
EPCAM	10	3019	chr2:47572272-47614765
HOXB13	2	3567	chr17:46802100-46806565
MLH1	24	4948	chr3:37034798-37107405
MSH2	19	10347	chr2:47630083-47794320
MSH6	18	11890	chr2:47922644-48037265
MUTYH	12	4540	chr1:45794810-45806167

[UCSC View](#) [Download](#)

[Cancel](#) [Back](#) [Next](#)

SureSelect DNA (Advanced)

Add/Review Content

Tile Genes or Regions

Define Design



Add/Review Content



Select Optimized Probes

► Tile Genes or Regions

Upload Probes

Select Existing Probes

Finalize

* Probegroup Name: prueba clase_1

Selection Parameters

Density: 3x

Masking: Moderately Stringent

Boosting: Optimized Performance XT HS/XT HS2/XT LI/QXT (Recommended)

* Extension into Masked: 20

Reset

SureSelect DNA Design

Name: prueba clase
Species: H. sapiens (hg19)
Category: SureSelect DNA
Hybridization: 90 Minutes

Probes

Probes: NA
Size: NA
Price Tier:  NA

UCSC View

Download

Cancel

Back

Begin Probe Selection

SureDesign

SureSelect DNA (Advanced) Add/Review C

Define Design ✓

Add/Review Content

Select Optimized Probes

Tile Genes or Regions

Upload Probes

Select Existing Probes

Finalize

Probes: - 0

SureSelect DNA Design

Name: prueba clase

Species: H. sapiens (hg19)

Category: SureSelect DNA

Hybridization: 90 Minutes

Probes

Probes: 20659

Size: 277.395 kbp

Price Tier: i Tier 1

UCSC View Download

This screenshot shows the SureDesign software interface. On the left, a sidebar lists steps: Define Design (marked with a green checkmark), Add/Review Content, Select Optimized Probes, Tile Genes or Regions, Upload Probes, Select Existing Probes, and Finalize. Below this is a summary of the 'SureSelect DNA Design' with details: Name: 'prueba clase', Species: 'H. sapiens (hg19)', Category: 'SureSelect DNA', Hybridization: '90 Minutes'. It also shows probe statistics: # Probes: 20659, Size: 277.395 kbp, and Price Tier: Tier 1. At the bottom are 'UCSC View' and 'Download' buttons. The main area is titled 'Add/Review C'.

Download Help - Download ×

Name: prueba clase

Species: H. sapiens (hg19)

Please select files to download.

prueba clase

prueba clase_1

prueba clase_1_AllTracks.bed

prueba clase_1_Covered.bed

prueba clase_1_Regions.bed

prueba clase_1_Report.txt

prueba clase_1_Targets.txt

This screenshot shows a 'Download' dialog box. It displays the name and species of the design ('prueba clase', 'H. sapiens (hg19)'). Below, it says 'Please select files to download.' A tree view shows several files under two main categories: 'prueba clase' and 'prueba clase_1'. All files under both categories have checkboxes checked, indicating they are selected for download. The files listed are: 'prueba clase_AllTracks.bed', 'prueba clase_Covered.bed', 'prueba clase_Regions.bed', 'prueba clase_Report.txt', and 'prueba clase_Targets.txt'.

Files for custom SureSelect DNA designs

SureSelect DNA designs created in SureDesign with the standard wizard have PDF, BED, and text design files available for download.

PDF Report file

The PDF report file has the file name **[design ID]_Report.pdf**.

This file contains summary information on the submitted targets, probe selection parameters, [coverage](#) statistics, recommended [minimum sequencing](#), and overall success of the design in covering the targets. This same information, and additional information, is provided in spreadsheet format in the Report text file.

BED files

The BED-format track files that SureDesign creates for custom SureSelect DNA designs are described below. You can import these files into a compatible genome browser to graphically view the locations of the tracks in the genome. For detailed information on the tracks and how they can help you analyze your design, see [Design analysis using tracks](#).

[design ID]_Regions.bed - This BED file contains a single track of the target regions of interest that SureDesign used to select the probes. You can use this track to see the exact regions that the program was attempting to cover when selecting the probes.

[design ID]_Covered.bed - This BED file contains a single track of the genomic regions that are covered by one or more probes in the design. The fourth column of the file contains annotation information. You can use this file for assessing coverage metrics.

[design ID]_AllTracks.bed - This multitrack BED file includes the following tracks:

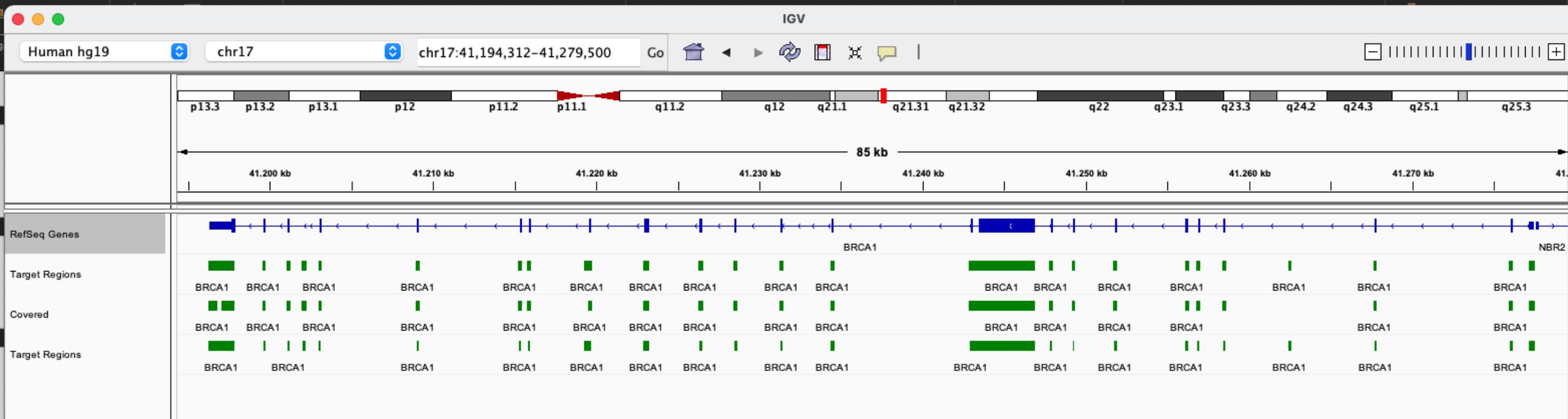
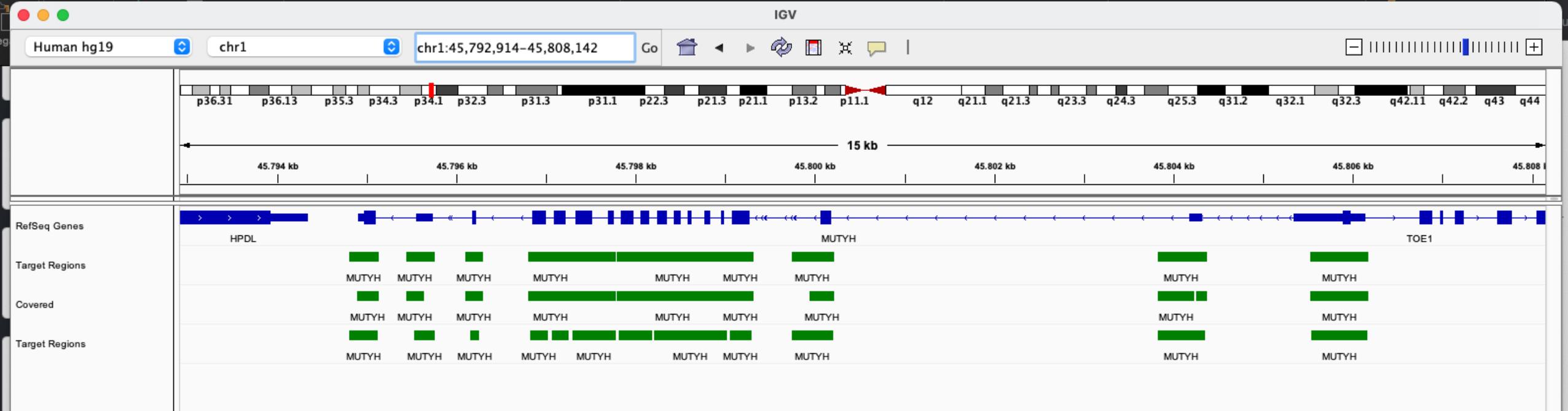
- The *Target Regions* track is identical to the track in the Regions BED file.
- The *Covered probes* track is identical to the track in the Covered BED file.
- The *Missed Regions* track contains any regions from the *Target Regions* track that are not included in the *Covered probes* track.

Text files

The three text files for a custom SureSelect DNA design are described below. You can view these files in any text editor program (e.g., NotePad) or spreadsheet program (e.g., Excel). Any tables embedded in the text files are tab-delimited and contain column headers. Lines of text that start with a # character are comment lines.

[design ID]_Targets.txt - This file contains a list of the target identifiers that you entered when creating the design.

[design ID]_Report.txt - This file contains summary information on the design, the probes, the targets, and the parameters used to create the design. </>



BaseSpace

DESIGNSTUDIO
SEQUENCING

START DESIGN

VIEW DESIGNS

VIEW REPORTS



Maria

Design a custom solution using our interactive tool

Start A New Design

What's new in DesignStudio?

Introducing AmpliSeq for Illumina

DesignStudio has been improved to include new AmpliSeq for Illumina designers to walk you through the steps of personalizing your custom amplicon panels. Designers for the following application types are supported:

- AmpliSeq Gene Design
- AmpliSeq Hotspot Design
- AmpliSeq On-Demand Design
- AmpliSeq RNA Expression Design
- AmpliSeq RNA Fusion Design

[Start a new design](#) [To learn more, read the release notes.](#)

Your Recent Designs

ID	NAME	ASSAY TECHNOLOGY	STATUS
193510	Huntington	AmpliSeq DNA Gene	Ready to Order
122573	CMS project	TruSeq Targeted RNA	Designed

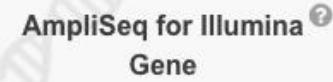
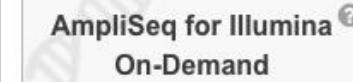
[Learn about the available assays](#) [See all](#)

07/2024

Assay Type



Assay Technology

[Compare Technologies](#) AmpliSeq for Illumina
Gene AmpliSeq for Illumina
Hotspot AmpliSeq for Illumina
On-Demand

Species



Design Name

prueba_clase_illumina

Description

enter a description



START DESIGN

VIEW DESIGNS

VIEW REPORTS



? ▾



Logout

1 Select Assay

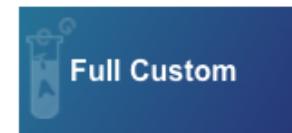
2 Configure Design

3 Manage Targets

4 Review Design

 Design Summary

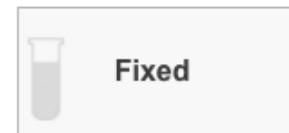
Configure Design

Design type 

Full Custom



Add-on



Fixed

 Tip : Full Custom - Create a project using targets of your choosing.

Create Project

Need help? [Contact us](#): Monday - Friday 7:00am - 5:00pm (PT) | ☎ 800.809.4566 (N. America) | ☎ 858.202.4566 (Outside) | 📧 Tech Support | 📧 Customer Service

© 2022 Illumina, Inc. All rights reserved.

DesignStudio Sequencing Assay Designer 8.1.0.00001

[+ Design Summary](#)
- Add Targets

Selected Targets: 86 / 86 Cumulative Target (bp): 1,720,410 Estimated Probes: 23,092 Coverage: 99%

Gene Free Text Coordinate File

Enter targets as free text separated by comma or new line

Enter a list of targets below separated by comma or new line (ex. brca1, Chr1:9000-10000). Coordinates are treated as Full Regions.

Add Target

Submit less than 500 targets at a time

⌚ What is FULLREGION ? :

← Specified Coordinates →

Targeting Option ⓘ:

CDS Only Exon Only Full Region

Probe Spacing ⓘ Standard ▾

Label:

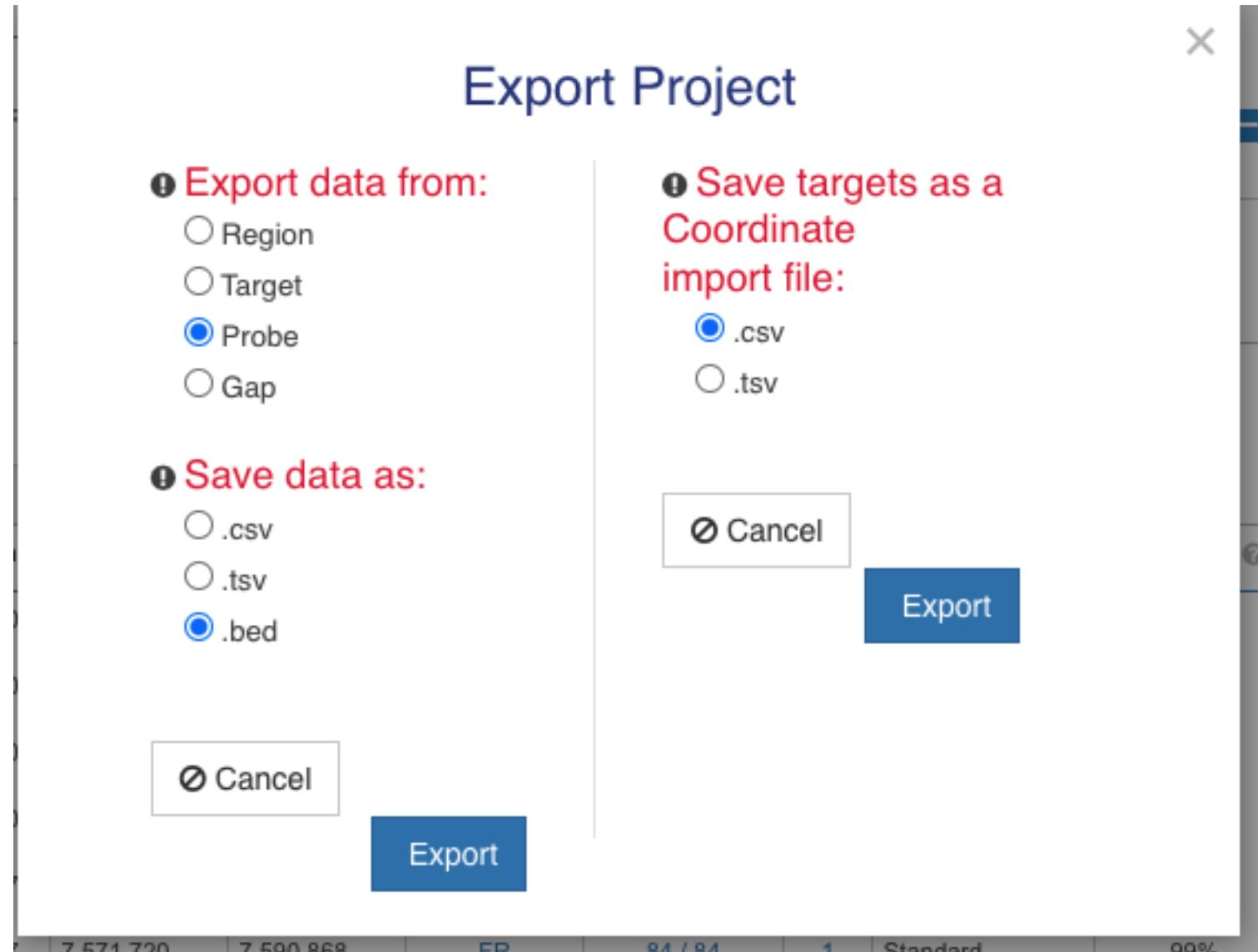
Para que sea comparable a Agilent, seleccionar CDS + UTRs

✖ Delete cloud Export arrow-right Review Design

	Regions ⓘ	Targets ⓘ	Probes	Gaps	+ Add ⓘ	- Remove ⓘ	Filter ▾	Filtered (0) ▾				
	Target Region	Chr	Start	Stop	Targets ⓘ	Probes ⓘ	Gaps	Probe Spacing...	Coverage ⓘ	Design ...	Labels	Added
<input type="checkbox"/>	PTEN ucsc ⓘ	10	89,623,195	89,728,532	FR	458 / 458	2	Standard	99%	Poor GC ...		01/02/2022
<input type="checkbox"/>	PTEN ucsc ⓘ	10	89,623,195	89,728,532	FR	458 / 458	2	Standard	99%	Poor GC ...		01/02/2022
<input type="checkbox"/>	PTEN ucsc ⓘ	10	89,623,195	89,728,532	FR	458 / 458	2	Standard	99%	Poor GC ...		01/02/2022
<input type="checkbox"/>	PTEN ucsc ⓘ	10	89,623,195	89,728,532	FR	458 / 458	2	Standard	99%	Poor GC ...		01/02/2022
<input type="checkbox"/>	TP53 ucsc ⓘ	17	7,571,720	7,590,868	FR	84 / 84	1	Standard	99%	Low Spe...		01/02/2022
<input type="checkbox"/>	TP53 ucsc ⓘ	17	7,571,720	7,590,868	FR	84 / 84	1	Standard	99%	Low Spe...		01/02/2022
<input type="checkbox"/>	TP53 ucsc ⓘ	17	7,571,720	7,590,868	FR	84 / 84	1	Standard	99%	Low Spe...		01/02/2022
<input type="checkbox"/>	TP53 ucsc ⓘ	17	7,571,720	7,590,868	FR	84 / 84	1	Standard	99%	Low Spe...		01/02/2022
<input type="checkbox"/>	STK11 ucsc ⓘ	19	1,205,798	1,228,434	FR	99 / 99	3	Standard	99%	Poor GC ...		01/02/2022
<input type="checkbox"/>	STK11 ucsc ⓘ	19	1,205,798	1,228,434	FR	99 / 99	3	Standard	99%	Poor GC ...		01/02/2022

Total Items: 86 Selected Item(s):

Page Size: 100 ▾ 1 / 1



prueba_clase_illumina

Design ID: 196260 – Ready to Order

[Edit](#)[Copy](#)[Transfer](#)[Delete](#)[+ Design Summary](#)

Review Design

Selected Targets: 86 / 86
Cumulative Target Length (bp): 1,720,410
Number of Probes: 23,092

Overlap: 74 %
Number of Gaps: 63
Total Gap Length (bp): 1,550

Duplicate Targets: 79
Average Region Size (bp): 51,466

Coverage:

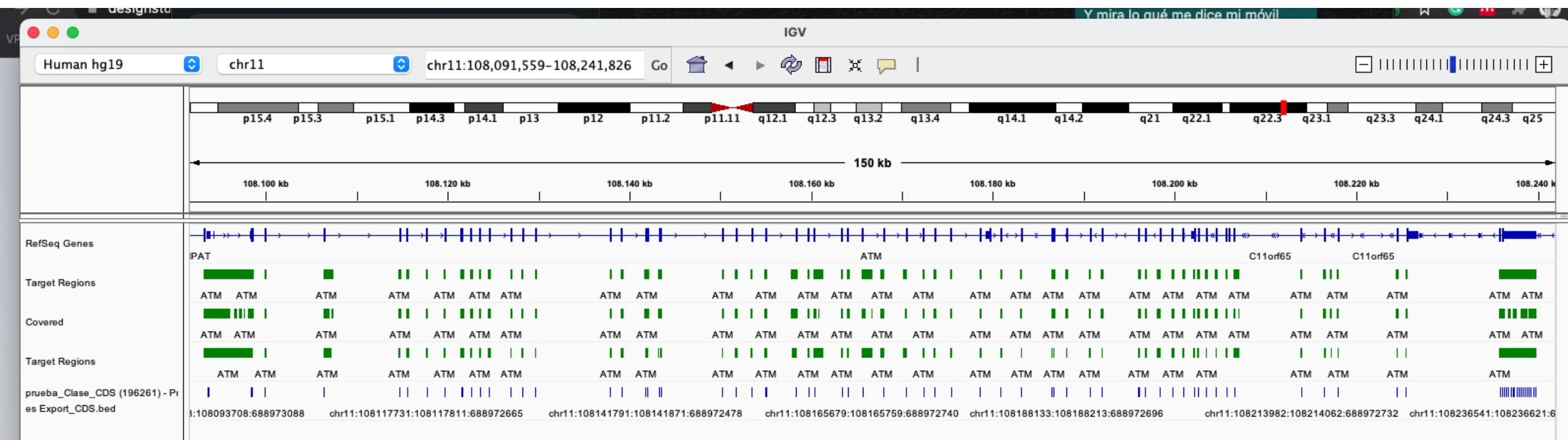
99%



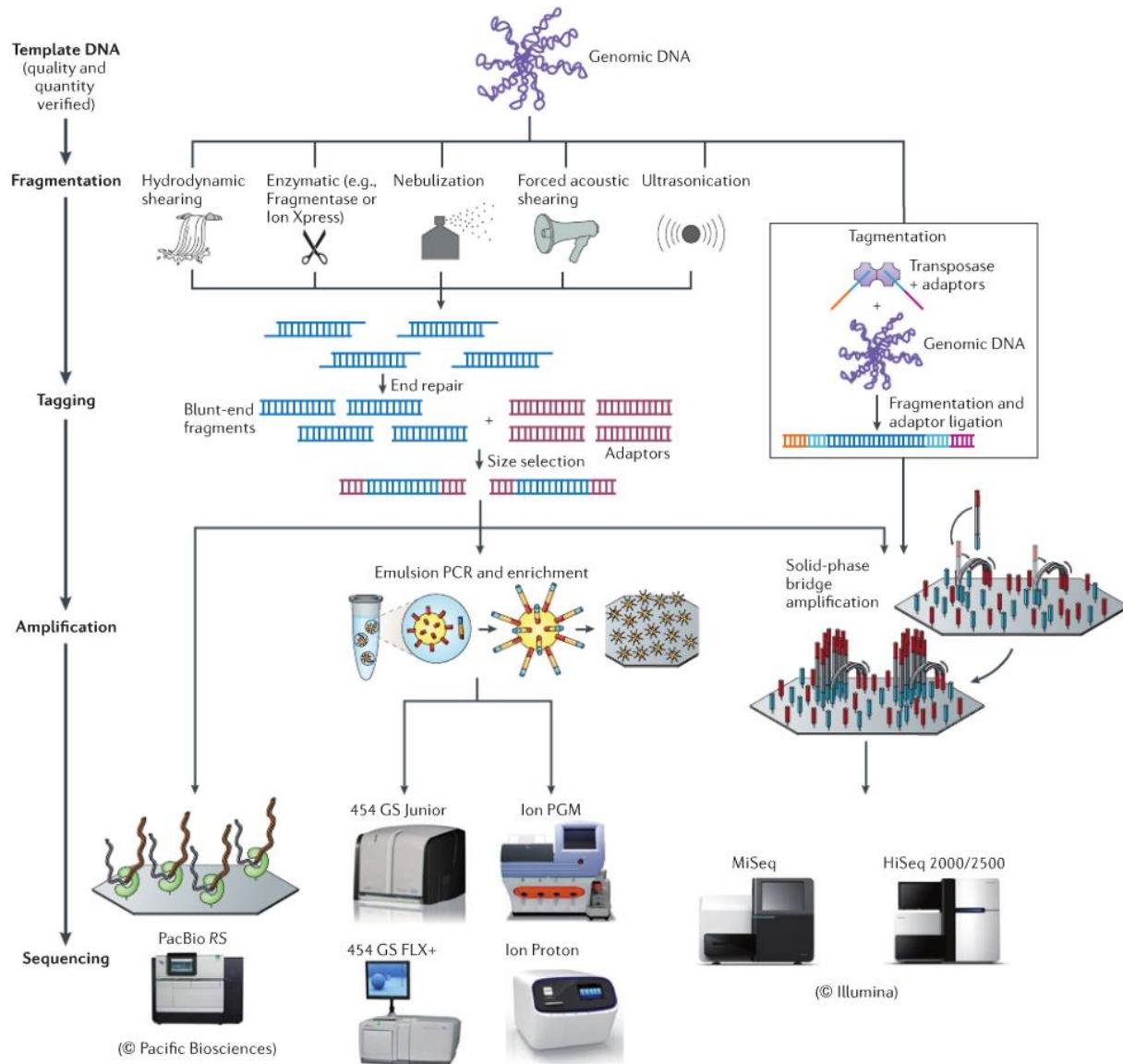
⚠ Warnings: There are duplicated targets in your design: 79

[Export](#) [Edit Design](#)

Regions	Targets	Probes	Gaps	Filter	Filtered (0)							
<input type="checkbox"/> Target Region		Chr	Start	Stop	Targets	Probes	Gaps	Probe Spacing...	Coverage	Design ...	Labels	Added
<input type="checkbox"/> TP53 ucsc		17	7,571,720	7,590,868	FR	84 / 84	1	Standard	99%	Duplicate...		01/02/2022
<input type="checkbox"/> TP53 ucsc		17	7,571,720	7,590,868	FR	84 / 84	1	Standard	99%	Duplicate...		01/02/2022
<input type="checkbox"/> STK11 ucsc		19	1,205,798	1,228,434	FR	99 / 99	3	Standard	99%	Duplicate...		01/02/2022
<input type="checkbox"/> STK11 ucsc		19	1,205,798	1,228,434	FR	99 / 99	3	Standard	99%	Duplicate...		01/02/2022
<input type="checkbox"/> STK11 ucsc		19	1,205,798	1,228,434	FR	99 / 99	3	Standard	99%	Duplicate...		01/02/2022
<input type="checkbox"/> STK11 ucsc		19	1,205,798	1,228,434	FR	99 / 99	3	Standard	99%	Duplicate...		01/02/2022
<input type="checkbox"/> STK11 ucsc		19	1,205,798	1,228,434	FR	99 / 99	3	Standard	99%	Duplicate...		01/02/2022
<input type="checkbox"/> STK11 ucsc		19	1,205,798	1,228,434	FR	99 / 99	3	Standard	99%	Duplicate...		01/02/2022
<input type="checkbox"/> STK11 ucsc		19	1,205,798	1,228,434	FR	99 / 99	3	Standard	99%	Duplicate...		01/02/2022
<input type="checkbox"/> CDKN2A ucsc		9	21,967,751	21,994,490	FR	117 / 117	1	Standard	99%	Duplicate...		01/02/2022

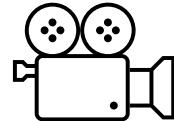


3.3.3. Secuenciación de la genoteca/librería



Una vez obtenida la genoteca, se procede a la secuenciación de ésta. La elección del equipo de secuenciación depende fundamentalmente de la cobertura que queremos conseguir para el genoma analizado. Los millones de lecturas, o gigabases secuenciados,

3.3.3. Secuenciación de la genoteca/librería



Secuenciación por síntesis (Illumina):

<https://www.youtube.com/watch?v=fCd6B5HRaZ8>



3.3.3. Secuenciación de la genoteca/librería

La **cobertura** de secuenciación (*coverage*, en inglés), como ya indicamos en temas anteriores, es el número medio de veces que las lecturas alinean para cubrir las bases conocidas de referencia. Este valor de cobertura determina si el descubrimiento de una variante genómica puede realizarse con un cierto grado de confianza.

La cobertura requerida varía en función de la aplicación, teniendo en cuenta que, a mayor cobertura, mayor confianza en la base nucleotídica secuenciada, lo que nos permitirá realizar la determinación de mutaciones con mayor precisión.

Sin embargo, debemos tener en cuenta que las lecturas no se distribuyen uniformemente sobre el genoma, sino que lo hacen de manera aleatoria e independiente. Por lo tanto, muchas bases estarán cubiertas por menos lecturas que la cobertura promedio, mientras que otras bases estarán cubiertas por más lecturas que el promedio.

3.3.3. Secuenciación de la genoteca/librería

No existen una regla única y estricta para la cobertura requerida, ya que depende del tipo de estudio, tamaño del genoma de referencia, y en el caso de experimentos de expresión génica, el nivel de esta expresión.

Sin embargo, actualmente, para experimentos de genómica humana se recomienda:

- para **genoma humano completo (WGS)** una **cobertura recomendada entre 30 a 50x**;
- mientras que para **exoma completo (WES)** o **paneles de genes** se recomienda una **cobertura 100x**.

Estos estándares suelen fijarse por las sociedades de genética, en base a los resultados publicados en las diversas publicaciones científicas.

3.3.3. Secuenciación de la genoteca/librería

La ecuación que calcula la cobertura media estimada es la ecuación de Lander/Waterman, expresada como:

$$C = \frac{L \times N}{G}$$

Donde C es la cobertura, G es el tamaño del genoma haploide en Megabases (Mb), L es la longitud de la lectura (si es pareada, se tiene en cuenta la suma de las dos lecturas) y N es el número de lecturas en millones.

$$\frac{\# \text{ of clusters} \cdot (\text{paired-end})}{\text{read length}} = \frac{40 \cdot 10^9}{3.2 \cdot 10^9} = 12.5$$

coverage

of clusters → 200M • (2 • 100)

paired-end →

read length →

human genome size → 3.2GB

3.3.3. Secuenciación de la genoteca/librería

Veamos qué cobertura obtendríamos para la secuenciación en los siguientes casos:

- Genoma completo (WGS) de humano (3 000 Mb) en un formato 2x100 pb, con un total de 60 M de lecturas para cada muestra:
 - $C = LN/G = (2 \times 100) * (60 \times 10^6) / (3 \times 10^9) = 4x$
- Genoma completo (WGS) de humano (3 000 Mb) en un formato 2x150 pb con un total de 60 M de lecturas por muestra:
 - $C = LN/G = (2 \times 150) * (60 \times 10^6) / (3 \times 10^9) = 6x$
- Genoma completo (WGS) de *Caenorhabditis elegans* (100 Mb) en formato 2x100 pb con un total de 180 M de lecturas por muestra:
 - $C = LN/G = (2 \times 100) * (180 \times 10^6) / (100 \times 10^6) = 360x$
- ¿Cuántos millones de lecturas necesitaríamos para llegar a una cobertura 50x en un genoma completo humano? (nota: lecturas pareadas 150 pb)
 - $C = LN/G \rightarrow N = CG/L = 50 * (3 \times 10^9) / (2 \times 150) = 500 \times 10^6$ (500 M de lecturas)

3.3.3. Secuenciación de la genoteca/librería

En el uso de equipos Illumina, en el siguiente enlace se encuentran unas directrices generales, así como una calculadora de coberturas para distintas aplicaciones.

https://emea.support.illumina.com/downloads/sequencing_coverage_calculator.html

Sequencing Coverage Calculator

Support Center:
Sequencing Coverage Calculator

4. Select the instruments you want to perform the estimation for.
5. Click Submit.

The estimator writes a table containing the number of lanes or flow cells you need to use for the desired product and parameters. You can also download the results in a comma-separated values file, so you can share data or use the tables in Excel.

DNA Applications RNA Applications DNA and RNA Applications

Additional Considerations

- Output is an estimation based on recommended cluster density and read length; real output will vary.
- For more information about estimating coverage estimates, see the technical note [Estimating Sequencing Coverage](#).
- If you plan a targeted resequencing or enrichment experiment, make sure to read the technical note [Optimizing Coverage for Targeted Resequencing](#).
- The estimator uses an estimate of clusters passing filter commonly found for balanced genomes (such as PhiX or the human genome). If you plan to sequence an unbalanced genome, you may have a lower number of clusters passing filter and consequently a lower output per lane

Support Center:
Sequencing Coverage Calculator

Application or product: TruSeq DNA Exome

Coverage: 20 %

Duplicates: 20 %

Genome or region size (in million bases): 45 Mb

Total read length (e.g. 200 for 2x100): 150 cycles

Benchtop Sequencers

- iSeq
- MiSeq
- MiSeq / MiSeq Dx in RUO mode
- NextSeq 500/550

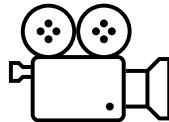
Production-Scale Sequencers

- NextSeq 500/550
- NovaSeq 6000
- HiSeq 3000/4000
- HiSeq 1500/2500 Rapid Run
- HiSeq 1500/2500 High Output
- NextSeq 1000 Sequencing System
- NextSeq 2000 Sequencing System

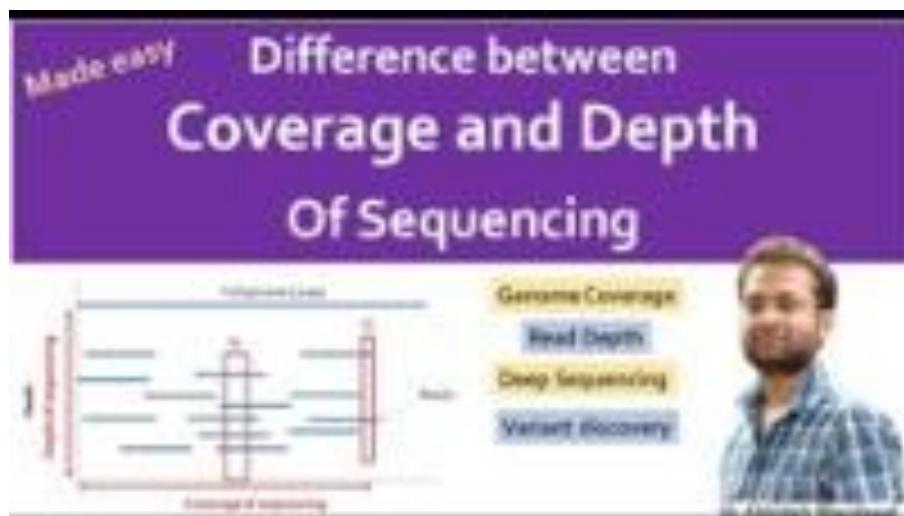
[Submit](#)

3.3.3. Secuenciación de la genoteca/librería

Nota: **(no confundir cobertura con profundidad!)**



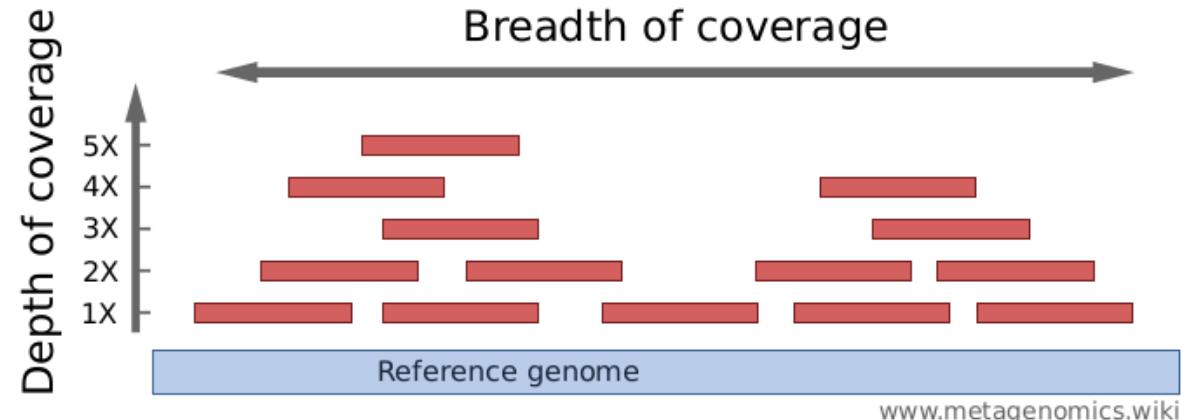
<https://www.youtube.com/watch?v=n1QbW1w1ThU>



Coverage and read depth are metrics that look at how well your library has been sequenced. Visually you can picture coverage looking at it horizontally, while read depth looks at it vertically.

Coverage / Breadth of coverage: “how much of the sample is covered by sequencing?”

Read Depth / Sequencing Depth / Depth of coverage: how many reads detected a specific nucleotide. This can be a strong indicator of the reliability of a base call.



www.metagenomics.wiki

3.4

Test Prenatal No-Invasivo (NIPT)

Test Prenatal No-Invasivo (NIPT)



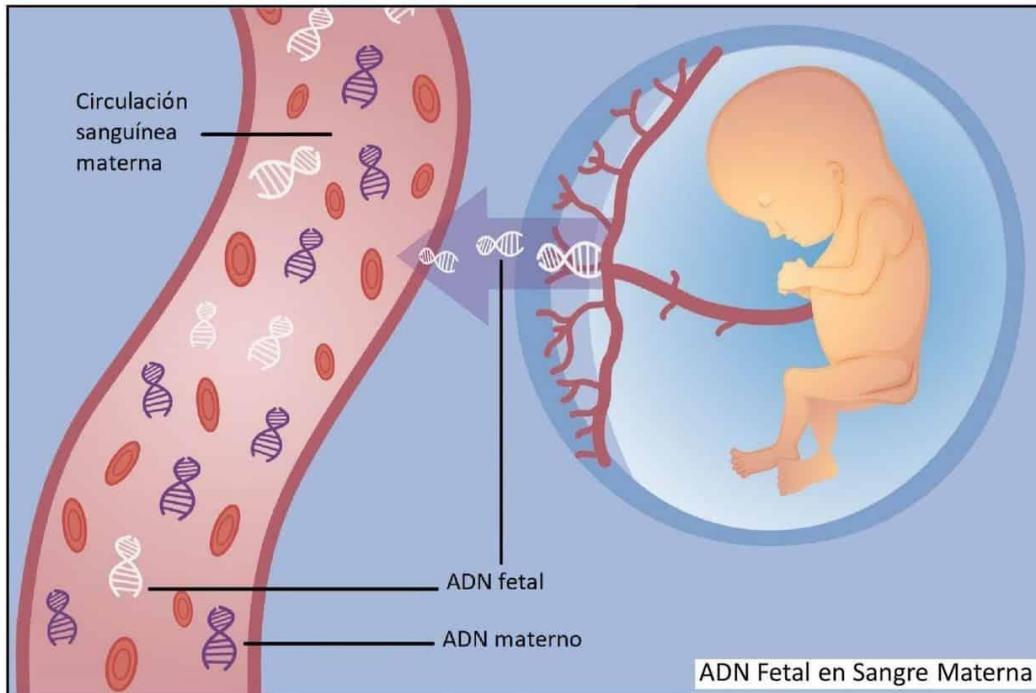
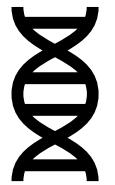
El test de cribado NIPT permite analizar a partir de la sangre materna (10-12 semanas embarazo) si el feto tiene altas probabilidades de tener algunas aneuploidías tales como Síndrome de Down, Edwards o Patau, lo que no significa que el bebé lo tenga definitivamente. --> Cribado de aneuploidías/trisomías en ADN fetal en sangre materna

Hay que destacar que NIPT es un CRIBADO; no una prueba diagnóstica, y si aparece riesgo en este cribado, debe complementarse con un test diagnóstico como una biopsia corial o una amniocentesis. Su utilidad radica en que nos permite disminuir el número de biopsias coriales y amniocentesis que se realizan, y que son más invasivas, con un riesgo de aborto.

Habitualmente, el NIPT está indicado en mujeres con un resultado de riesgo medio/alto en los cribados primarios (translucencia nucal + edad materna + pruebas bioquímicas). No está indicado si la mujer está embarazada de 3 o más bebés.

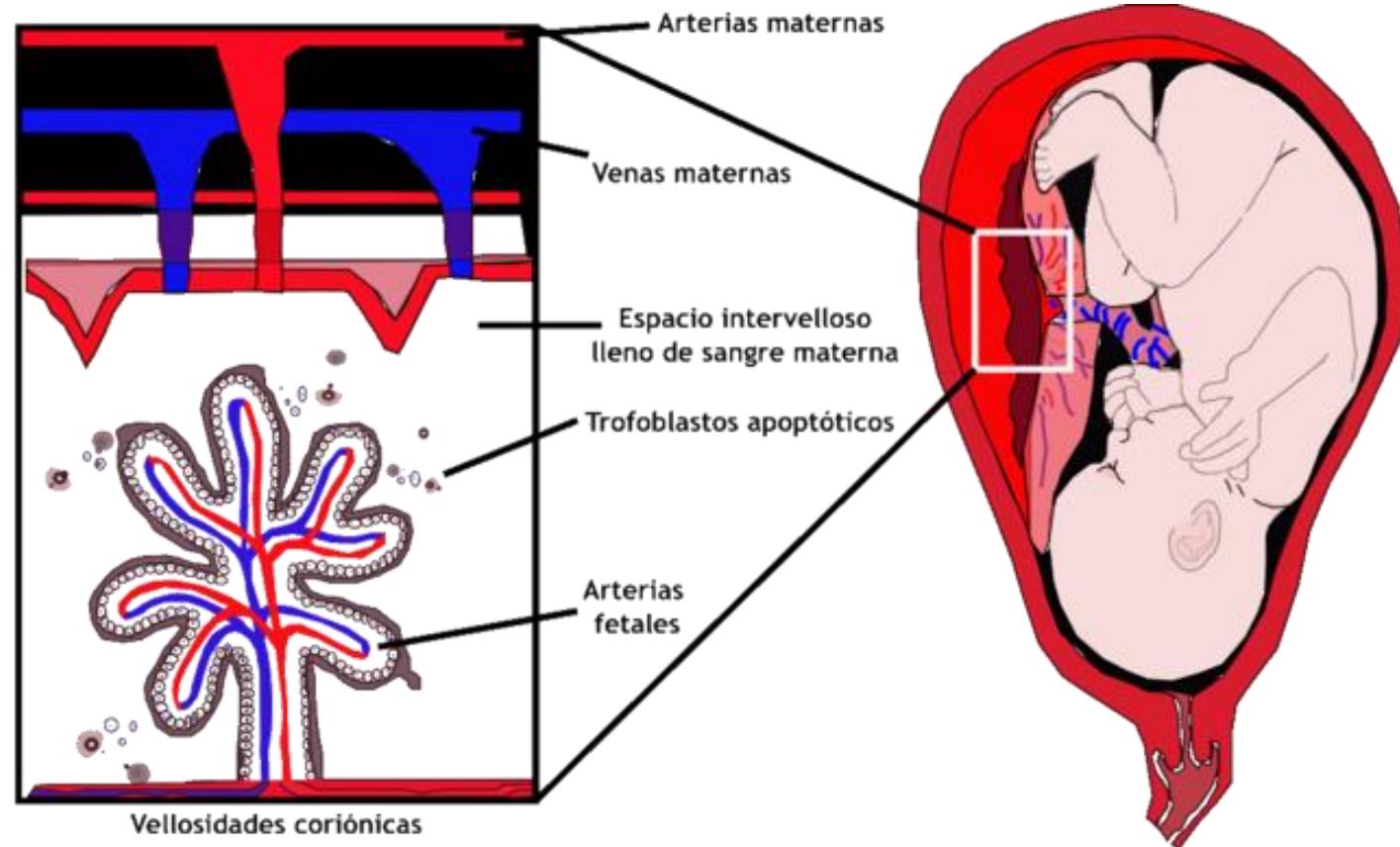
La prueba no selecciona (captura) las regiones a secuenciar/analizar, sino que se hace una **secuenciación de genoma completo**. Según la profundidad de secuenciación se podrán analizar sólo las trisomías, o bien, hasta microdelecciones.



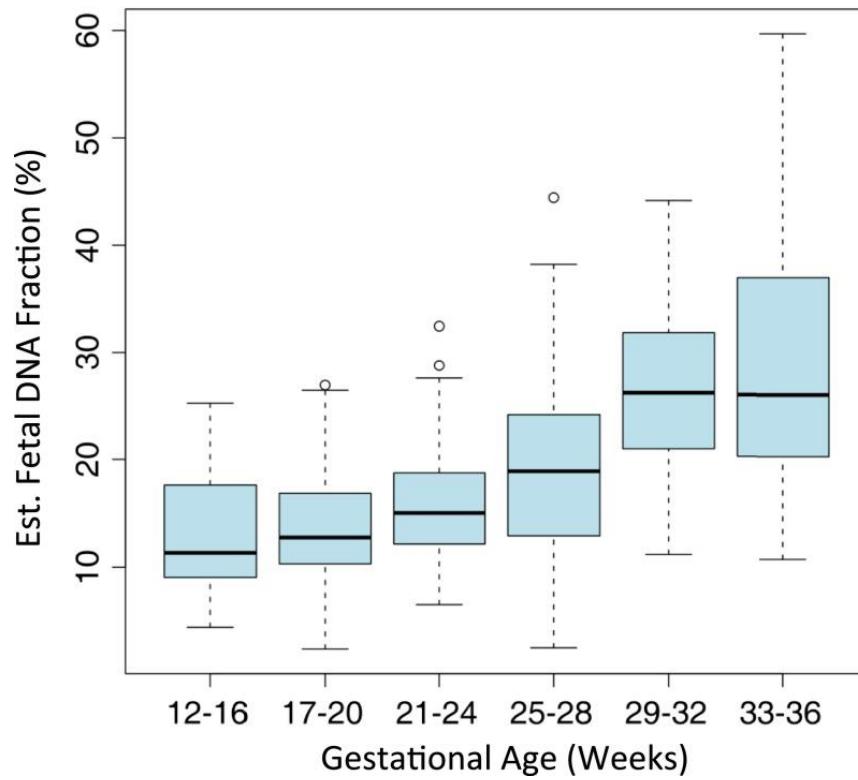


- DNA placentario circulante en sangre materna
- Origen: células del trofoblasto en apoptosis
- DNA fetal es más pequeño (146 nt) que el DNA materno (170 nt).
- Detectable a partir de la semana 9 de gestación, aumenta con la edad gestacional y disminuye con el peso materno (efecto dilucional)
 - Baja fracción fetal (FF) <4%
 - Necesario FF >8% aprox
- Indetectable a las 2 horas del nacimiento
- Vida media: 16 minutos

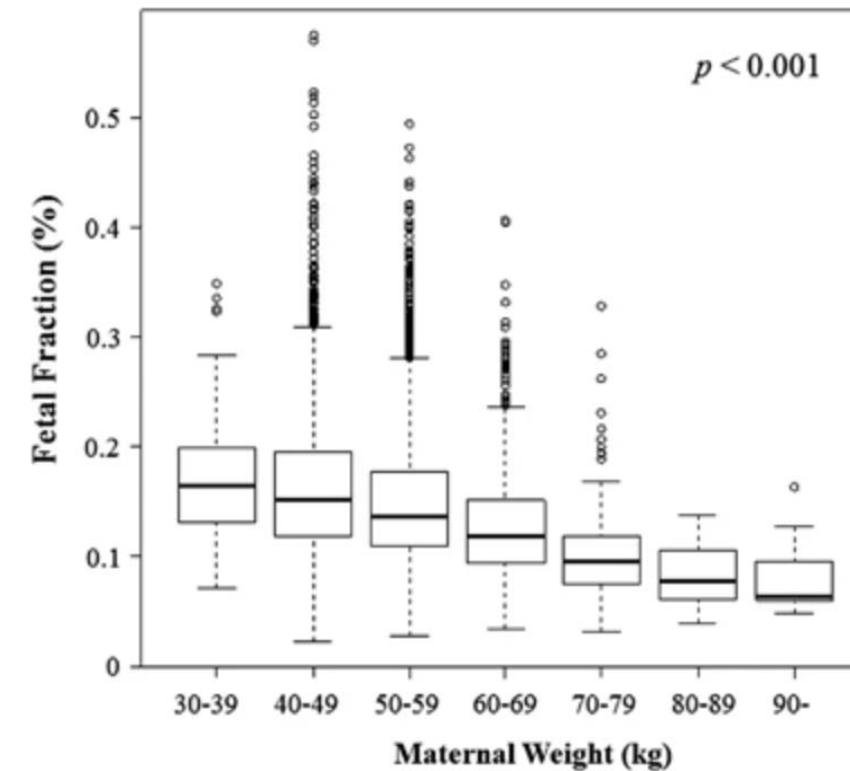
- T21: Síndrome Down
- T18: Síndrome Edwards
- T13: Síndrome Patau
- Otros (a mayor profundidad seq)
 - Cromosoma Y : sexo fetal
 - microdelecciones



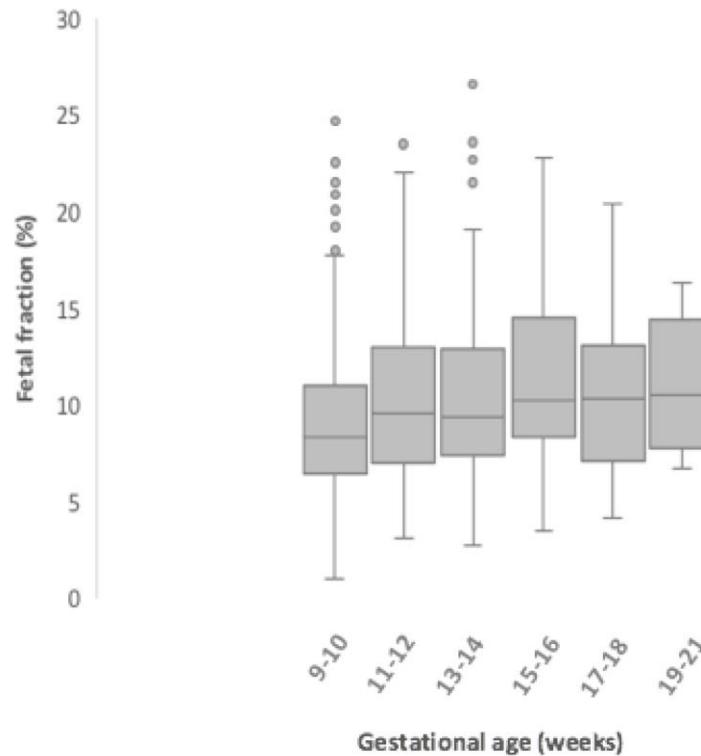
3.4. Test Prenatal No-Invasivo (NIPT)



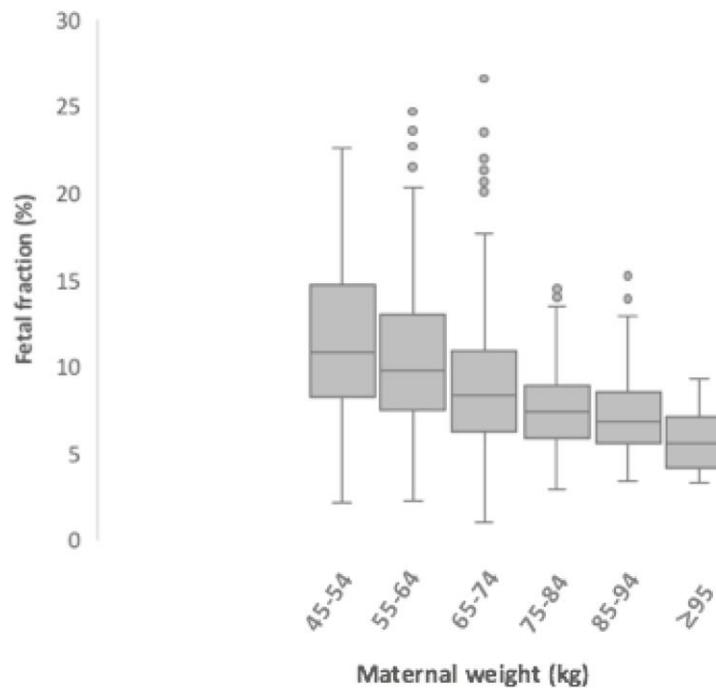
Yin AH, Peng CF, Zhao X, et al. Noninvasive detection of fetal subchromosomal abnormalities by semiconductor sequencing of maternal plasma DNA. *Proceedings of the National Academy of Sciences of the United States of America*. 2015 Nov;112(47):14670-14675. DOI: 10.1073/pnas.1518151112. PMID: 26554006; PMCID: PMC4664371.



Suzumori, N., Ebara, T., Yamada, T. et al. Fetal cell-free DNA fraction in maternal plasma is affected by fetal trisomy. *J Hum Genet* 61, 647–652 (2016).
<https://doi.org/10.1038/jhg.2016.25>



(a)

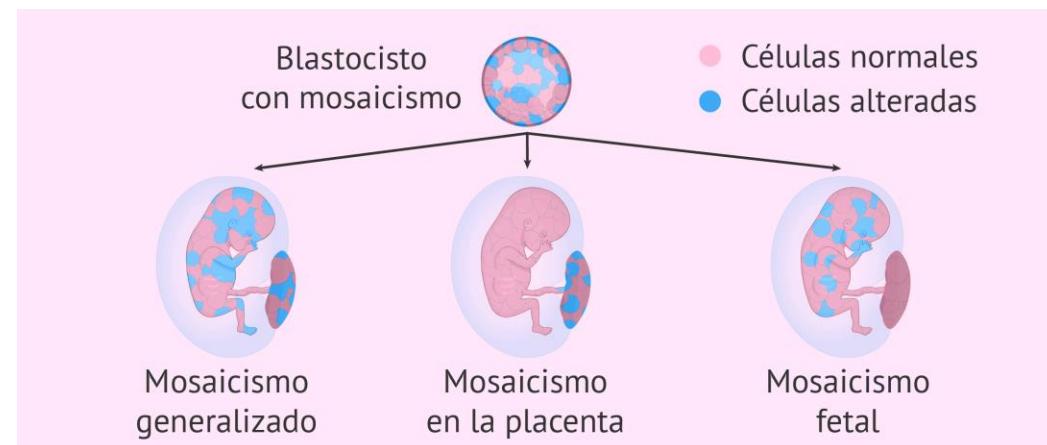


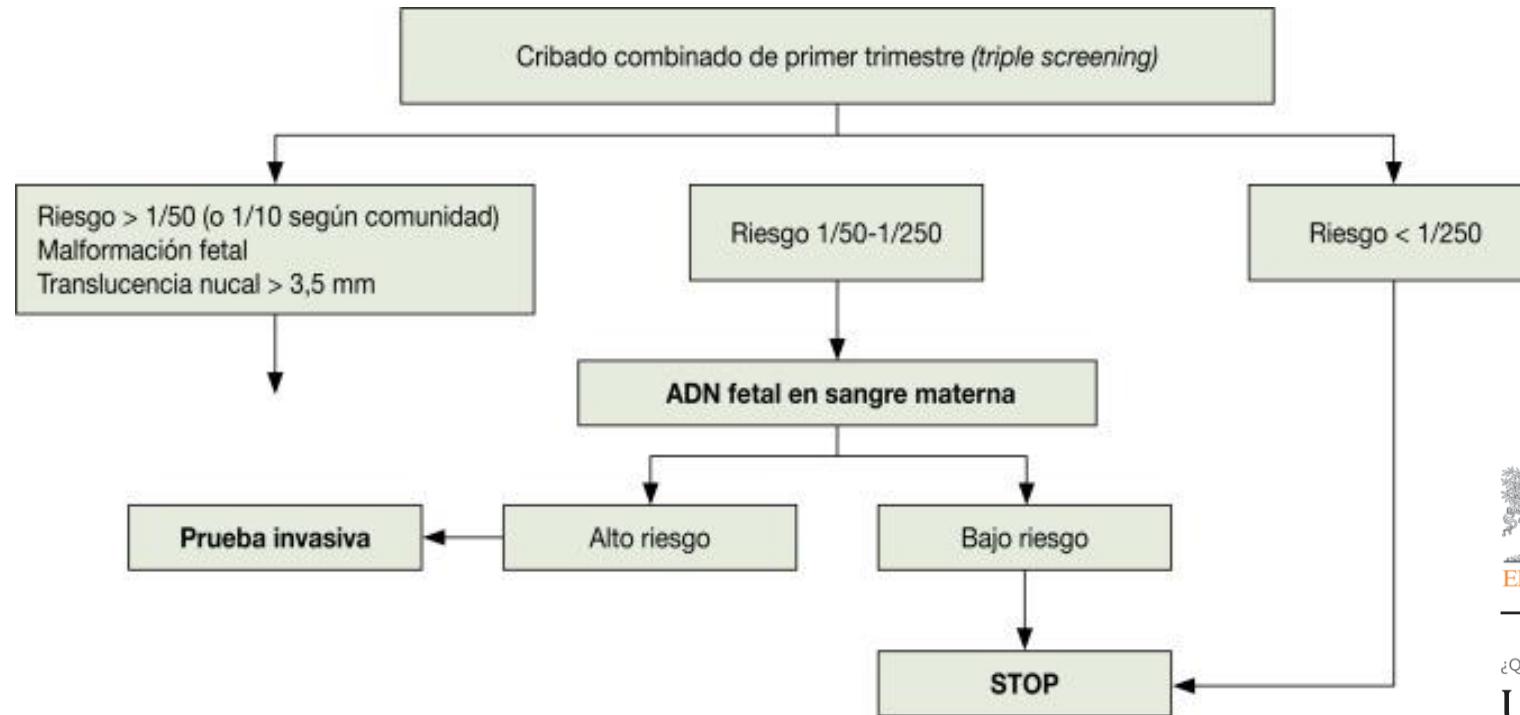
(b)

Serapinas D, Boreikaitė E, Bartkevičiūtė A, Norvilaitė K, Narbekovas A, Bartkevičienė D. The Level of Free Fetal DNA as Precise Noninvasive Marker for Chromosomal Aneuploidies: First Results from BALTIC Region. *Medicina*. 2020; 56(11):579. <https://doi.org/10.3390/medicina56110579>

Figure 2. Relationship between the fetal fraction and the gestational age, (a) the fetal fraction and the maternal weight (b). Boxplot description: The top line of the box is 75th percentile. The horizontal line inside each box is the median. The bottom line of the box is 25th percentile. The vertical lines out of the box represent the minimum and maximum values. The circle outside of the box are outliers.

Ventajas		Limitaciones
No invasivo	<ul style="list-style-type: none"> - 10 mL de sangre periférica - NO hay riesgo de infección intrauterina ni abortos 	<ul style="list-style-type: none"> - Es un test de cribado prenatal, no es una prueba diagnóstica
Alta sensibilidad	<ul style="list-style-type: none"> - Tecnología validada clínicamente - Múltiples publicaciones científicas - Sensibilidad y especificidad >99% 	<ul style="list-style-type: none"> - Un resultado positivo puede indicar alteración materna o en la placenta, y no en el feto
Detección precoz de aneuploidías	<ul style="list-style-type: none"> - Primer trimestre (a partir semana 9-10) - Detección precoz de trisomía - Mejor decisión clínica 	<ul style="list-style-type: none"> - Aneuploidías cromosómicas maternas (ej. Células tumorales)
		<ul style="list-style-type: none"> - Mosaicismo confinado a la placenta





FMC - Formación Médica Continuada en
Atención Primaria

Volume 27, Issue 8, October 2020, Pages 411-414



¿Qué debería saber el médico de familia sobre...?

Utilidad del ADN fetal en sangre materna para el cribado neonatal

[Paula Díaz Rodríguez](#)

Show more ▾

+ Add to Mendeley  Share  Cite

<https://doi.org/10.1016/j.fmc.2019.12.008>

[Get rights and content](#)

Nota: dependiente de hospitales/CA, puede haber variación en los límites de riesgo aplicados

Comparativa pruebas de diagnóstico prenatal

Prueba	¿Invasivo?	¹ SG	Riesgo aborto	Detección (trisomías 21, 18 y 13)
Cribado combinado 1er trimestre	No	11-14	No existe	² Sensibilidad: 60-80% ³ Especificidad: 5-10%
Biopsia vellosidad corial	Sí	10-14	1-2%	Sensibilidad: 100% Especificidad: 100%
Amniocentesis	Sí	16-21	0,5-1%	Sensibilidad: 100% Especificidad: 100%
Prueba ADN fetal	No	10-24	No existe	Sensibilidad: 99,9% ⁴ Especificidad: 99,9% ⁴

1. SG = Semana de gestación. 2. Sensibilidad = Capacidad de detectar una anomalía cuando de verdad existe. Es decir, para un valor del 99,9% de cada 1.000 casos con trisomía sólo uno no será detectado (falso negativo) . 3. Especificidad = Capacidad de no detectar la anomalía cuando no existe. Es decir, para un valor del 99,9% de cada 1.000 casos sin trisomía uno será clasificado como alto riesgo (falso positivo). 4. Valor que ofrecen gran parte de las pruebas que existen actualmente en el mercado.

¿Qué se analiza de modo más extendido?

- **Trisomía del cromosoma 21: Síndrome de Down;**
- **Trisomía del cromosoma 18: Síndrome de Edwards;**
- **Trisomía del cromosoma 13: Síndrome de Patau;**
- Monosomía del cromosoma X: Síndrome de Turner (45,X);
- Trisomía del cromosoma X: 47,XXX
- Síndrome de Klinefelter: 47,XXY;
- Síndrome de Jacobs; 47,XYY
- Microduplicaciones o microdelecciones
- Desequilibrios cromosómicos completos
- Variación del número de copias (CNVs): CNVs a todo el genoma o síndromes de microdeleción dirigidos, como el síndrome de cri-du-chat (5p-), el síndrome de Digeorge (22q11.2), el síndrome de Prader-Willi y Angelman (15q-), el síndrome de Wolf-Hirschhorn (4p-), el síndrome de Jacobsen (11q-), el síndrome de Langer-Giedion (8q-) y 1p36.

¿CUÁN CERTEROS SON LOS RESULTADOS
DEL ADN FETAL LIBRE?



<https://www.youtube.com/watch?v=V2d-Z53TtpE>

¡Gracias!

The logo consists of the lowercase letters "viu" in white, centered within a dark orange, rounded rectangular shape.

viu

Universidad
Internacional
de Valencia

universidadviu.com

De:
 Planeta Formación y Universidades