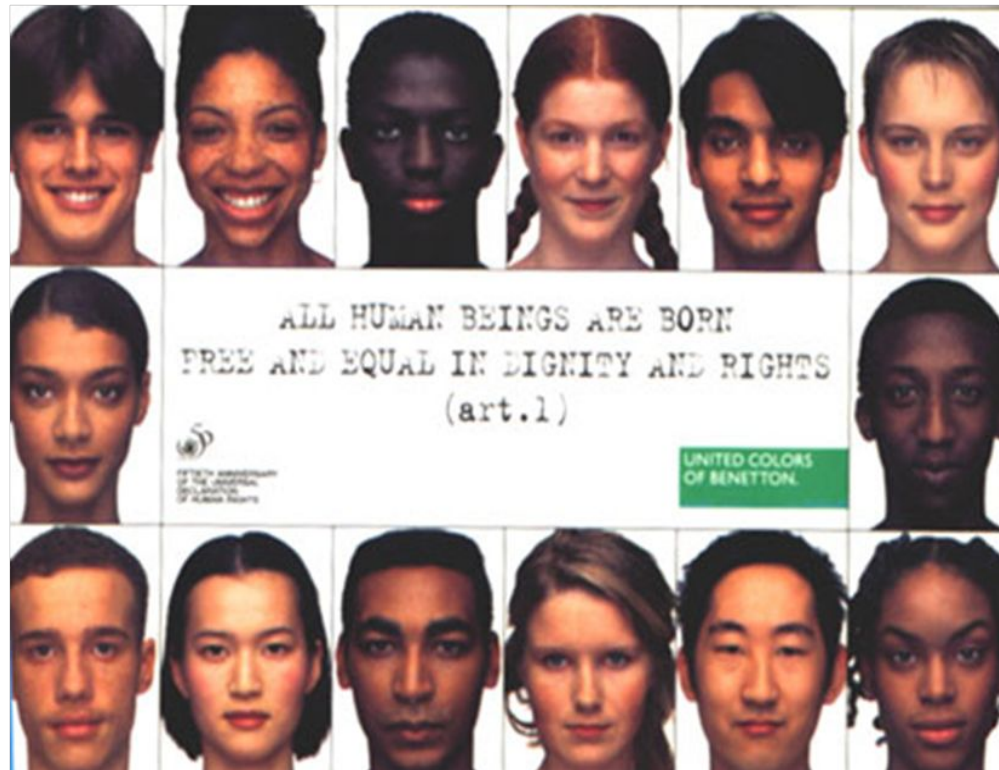


# Tools for human molecular diagnostics/Human Genetic diversity

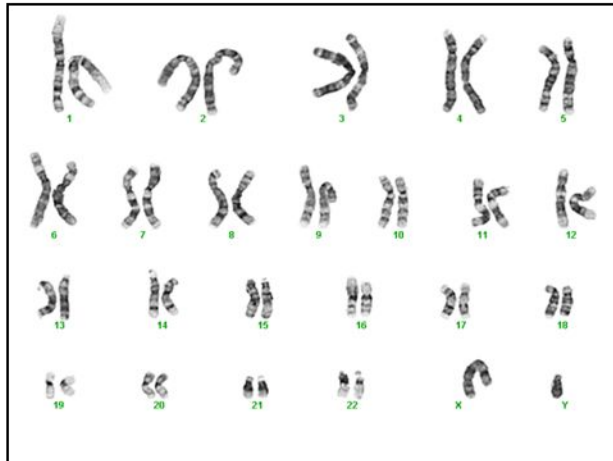
Joris Vermeesch  
BeSHG 2023

# Why are we different?



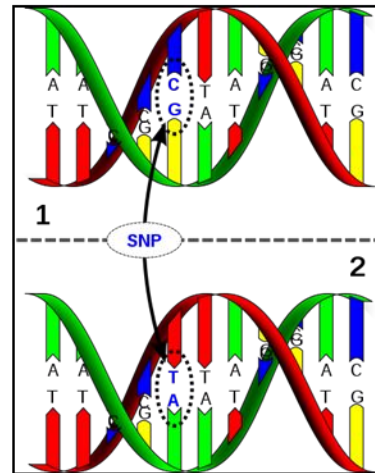
# Causes of genetic variation

Chromosomes  
(1960)



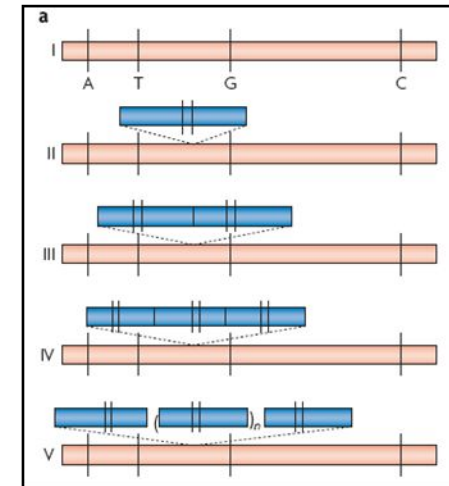
Variant are rare

Single nucleotide  
polymorphisms (SNPs)  
(1980)



Frequent:  
- 1 SNP every 1000 bp  
- 0.1% difference between 2 human genomes  
- 3 Mb difference

Copy number variations  
(CNVs)  
(2004)



Very frequent:  
- 1000 CNVs/2 individuals  
- 0.7% of genome is copy variable between 2 individuals  
- 21 Mb difference!

# Human Genetic Variation

1. Nature of variation
2. Types of mutations and their consequences
3. Variation in individual genomes
4. Origin and frequency of different types of mutation
5. Consequences for molecular diagnostics of WES/WGS

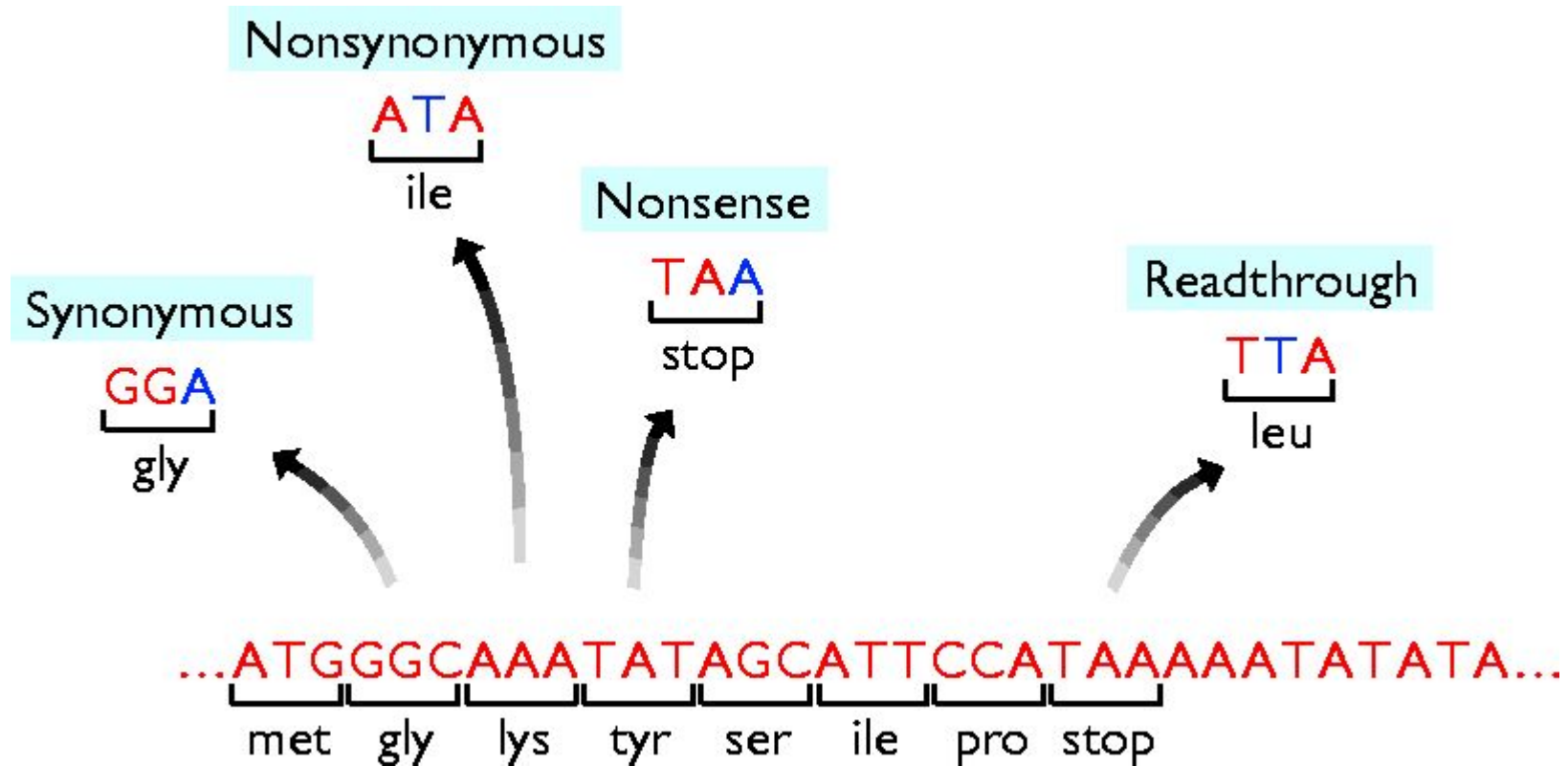
# Types of variation and their consequences

- Chromosomes & Copy number variation => see lesson on chromosomes

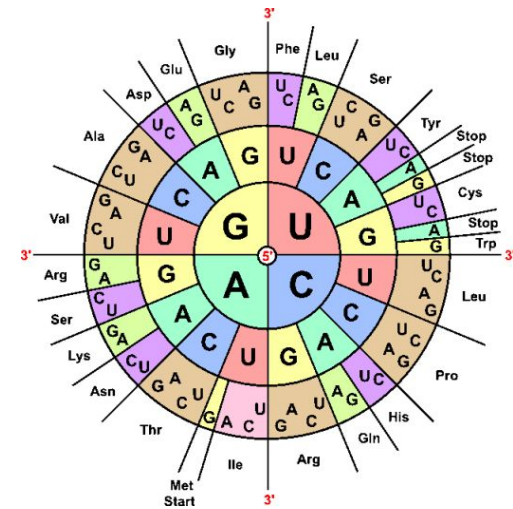
# SNPs : Common variation in the genome

	5	10	15	20																					
<b>Reference sequence</b>	...	G	A	T	T	C	T	A	G	G	T	A	A	C	T	C	A	G	T	C	G	A	...		
<b>SNP</b>	<i>Allele 1</i>	...	G	A	T	T	C	T	A	G	G	T	A	A	C	T	C	A	G	T	C	G	A	...	
	<i>Allele 2</i>	...	G	A	T	T	C	<b>C</b>	A	G	G	T	A	A	C	T	C	A	G	T	C	G	A	...	
<b>Indel A</b>	<i>Allele 1</i>	...	G	A	T	T	C	T	A	G	G	T	A	A	C	T	C	A	G	T	C	G	A	...	
	<i>Allele 2</i>	...	G	A	T	T	C	T	A	G	G	<b>G</b>	T	A	A	C	T	C	A	G	T	C	G	A	...
<b>Indel B</b>	<i>Allele 1</i>	...	G	A	T	T	C	T	A	G	G	T	A	A	C	T	C	A	G	T	C	G	A	...	
	<i>Allele 2</i>	...	G	A	T	<b>-</b>	<b>-</b>	C	T	A	G	G	T	A	A	C	T	C	A	G	T	C	G	A	...

# Point mutations (SNV)

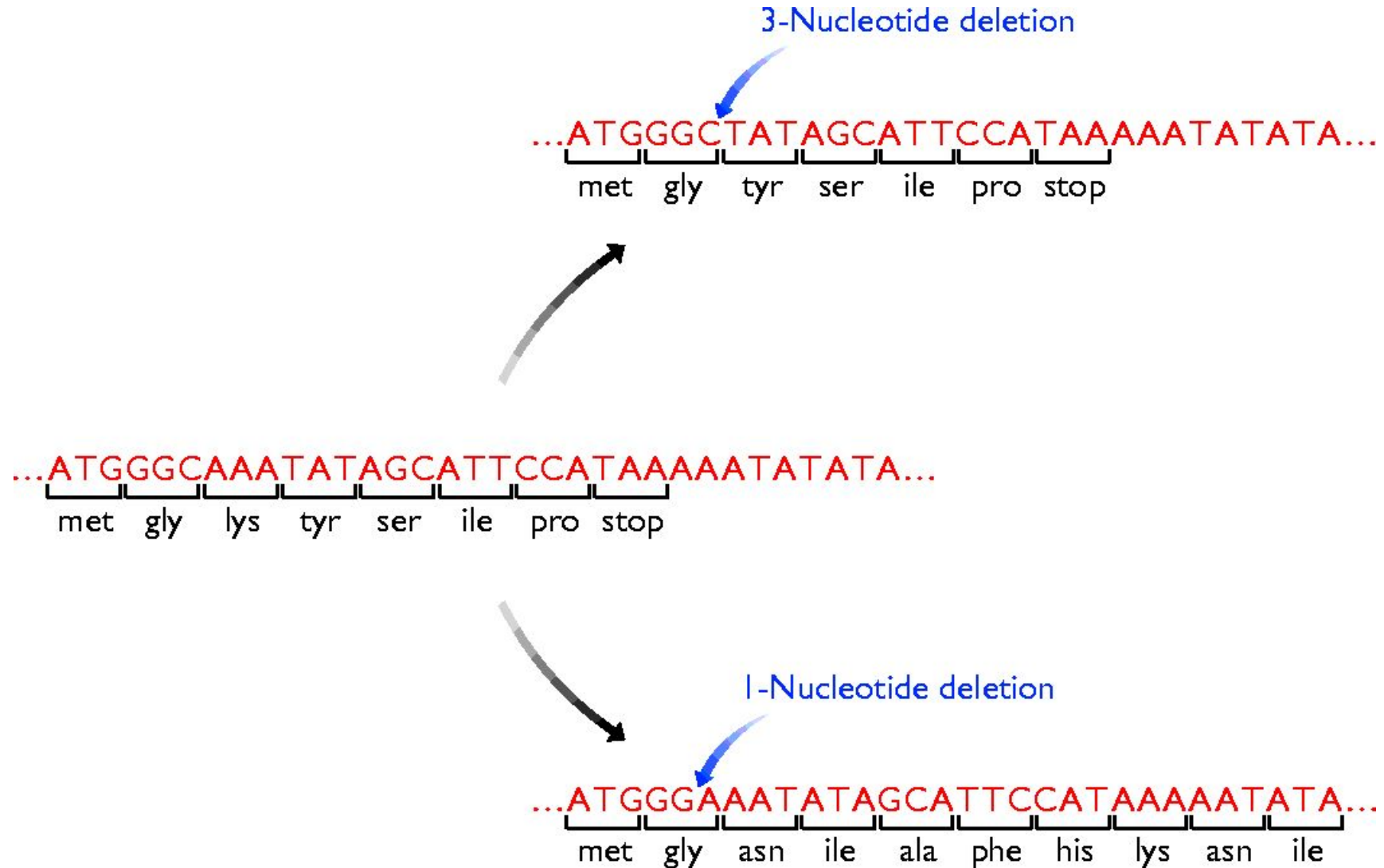


		Second letter				
		U	C	A	G	
First letter	U	UUU } Phe UUC } UUA } Leu UUG }	UCU } UCC } Ser UCA } UCG }	UAU } Tyr UAC } UAA Stop UAG Stop	UGU } Cys UGC } UGA Stop UGG Trp	U C A G
	C	CUU } CUC } Leu CUA } CUG }	CCU } CCC } Pro CCA } CCG }	CAU } His CAC } CAA } Gln CAG }	CGU } CGC } Arg CGA } CGG }	U C A G
	A	AUU } AUC } Ile AUA } AUG Met	ACU } ACC } Thr ACA } ACG }	AAU } Asn AAC } AAA } Lys AAG }	AGU } Ser AGC } AGA } Arg AGG }	U C A G
	G	GUU } GUC } Val GUA } GUG }	GCU } GCC } Ala GCA } GCG }	GAU } Asp GAC } GAA } Glu GAG }	GGU } GGC } Gly GGA } GGG }	U C A G





# Deletions and insertions (indels)



# Point mutations origin

- During replication (1 mutation/cell division)
- DNA damage
  - Estimated to be 10000- 1M nucleotides are damaged/human/day
  - Spontaneous chemical processes: e.g. Depurination, Demethylation, Deamination
  - Chemical mutagens (natural or otherwise)
  - Ionizing and UV radiation
- DNA damage is repaired, but some remain.

# Mutational signatures

**COSMIC**  
Catalogue Of Somatic Mutations In Cancer

Projects ▾ Data ▾ Tools ▾ News ▾ Help ▾ About ▾ Genome Ver

Terms and Conditions have been updated and include

## COSMIC v94, released 28-MAY-21

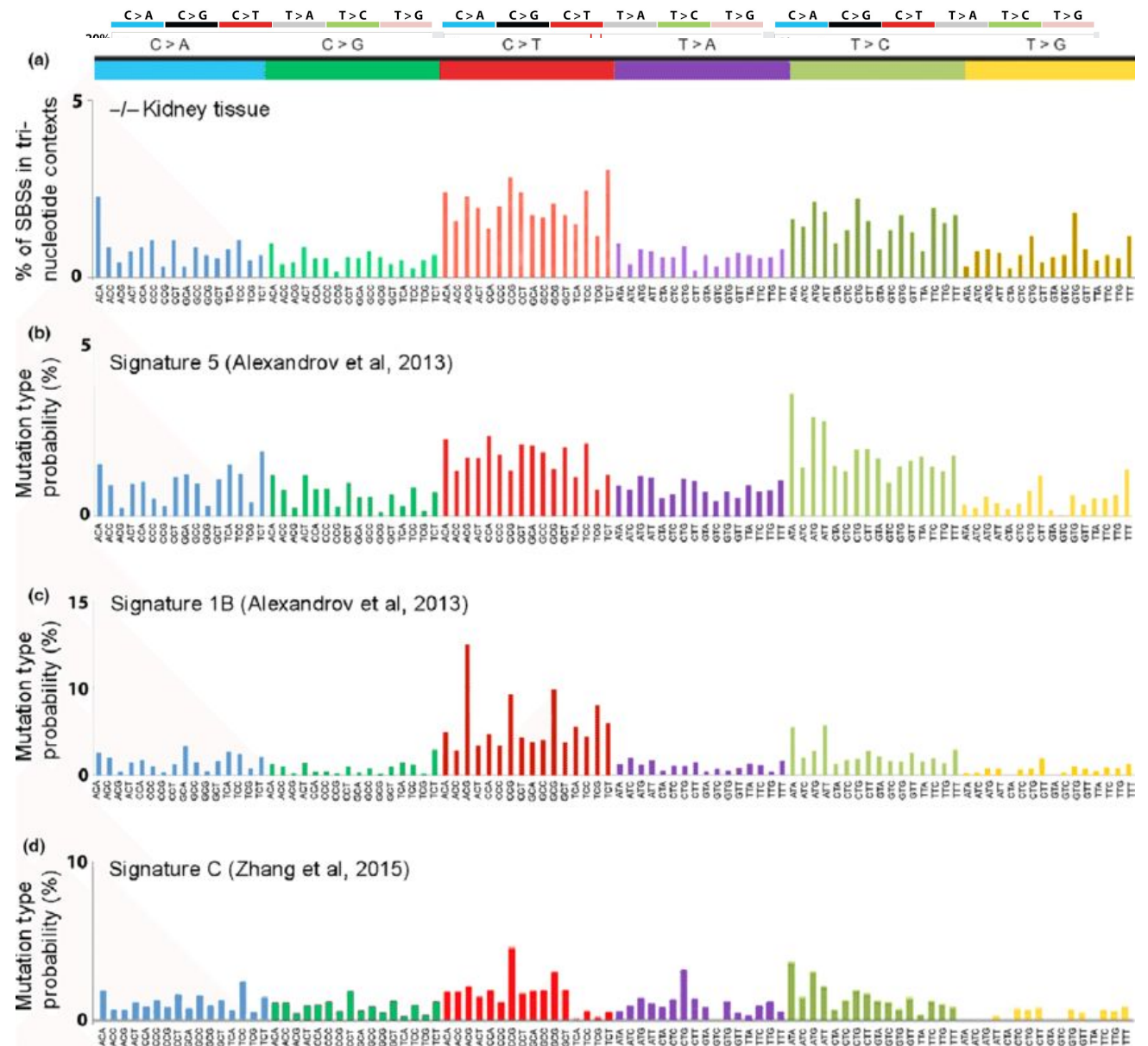
COSMIC, the Catalogue Of Somatic Mutations In Cancer, is the world's largest and most comprehensive resource for exploring the impact of somatic mutations in human cancer.

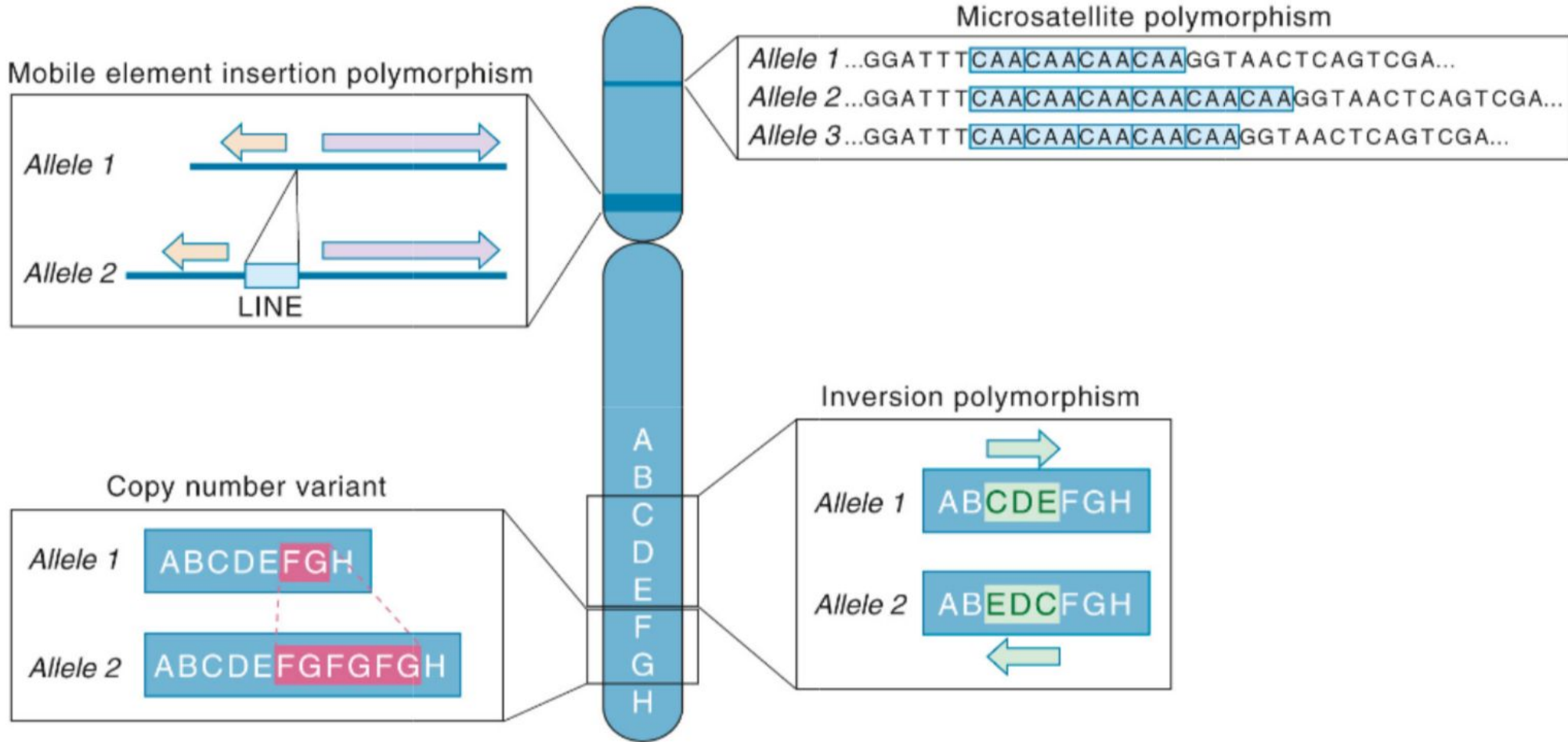
Start using COSMIC by searching for a gene, cancer type, mutation, etc. below.

eg *Braf*, *COLO-829*, *Carcinoma*, *V600E*, *BRCA-UK*, *Campbell* **SEARCH**

### Projects

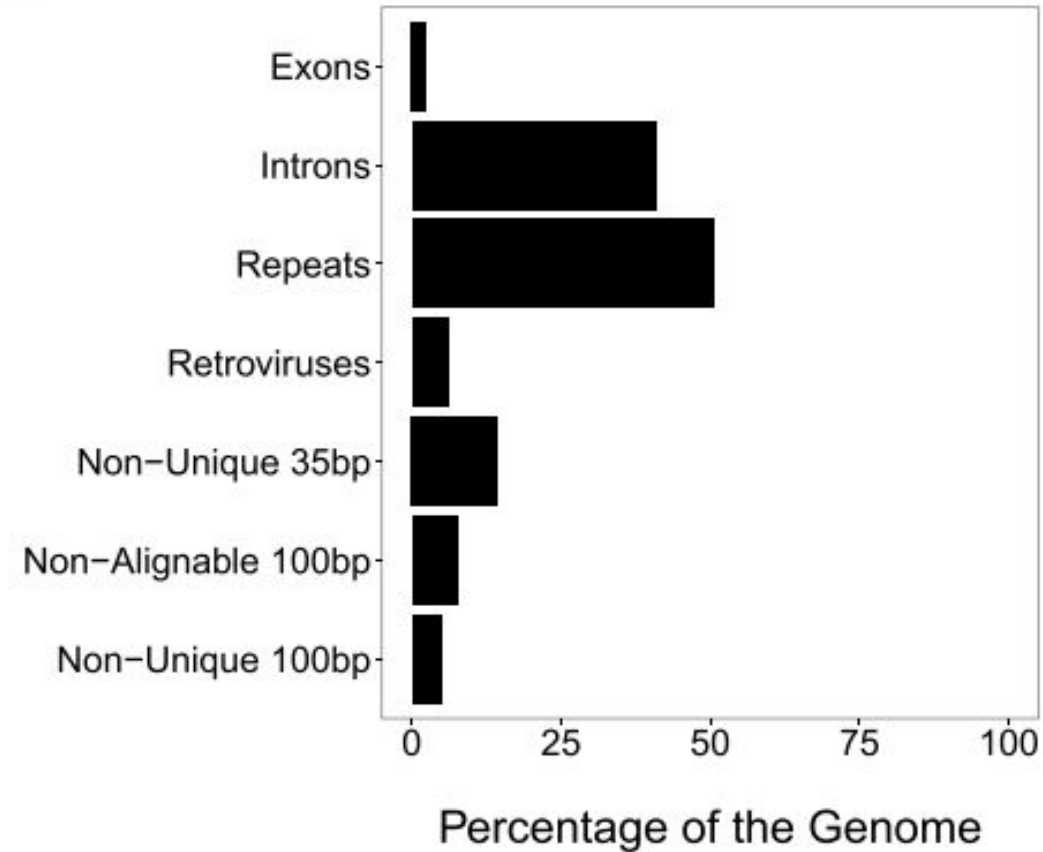
COSMIC is divided into several distinct projects, each presenting a separate dataset or view of our data:



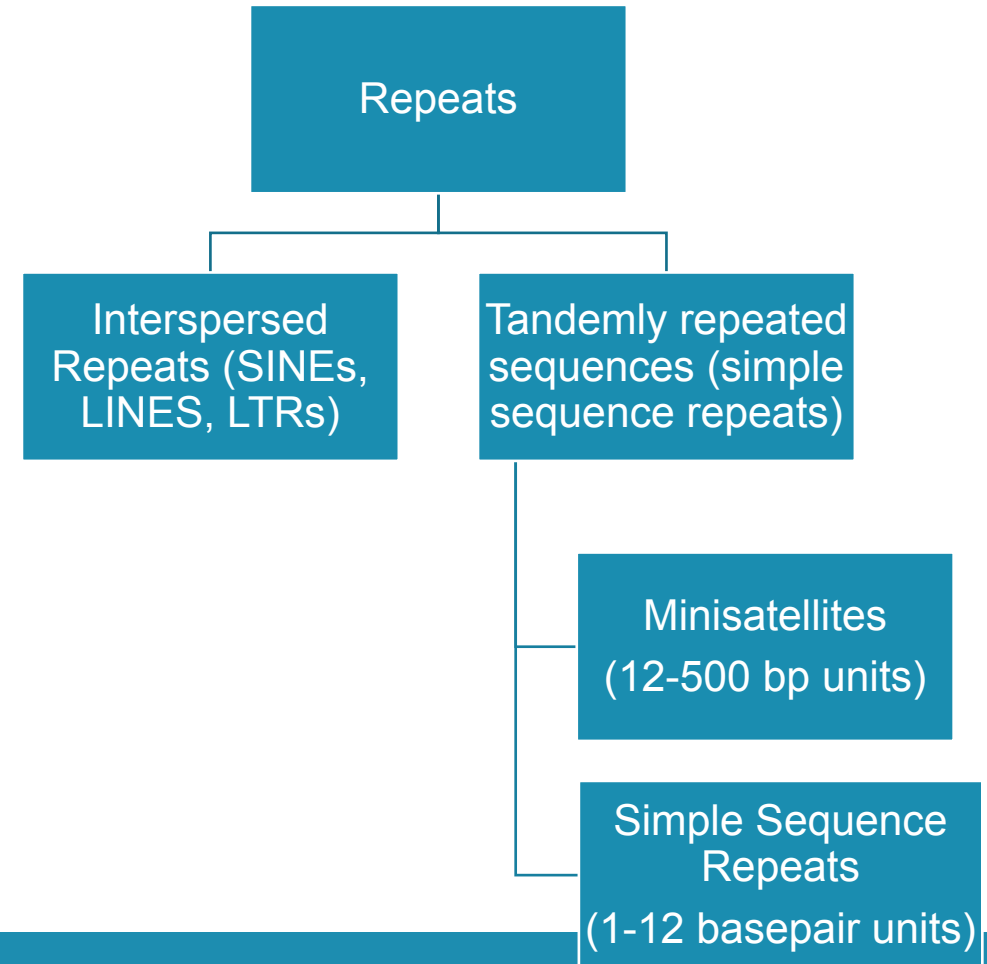
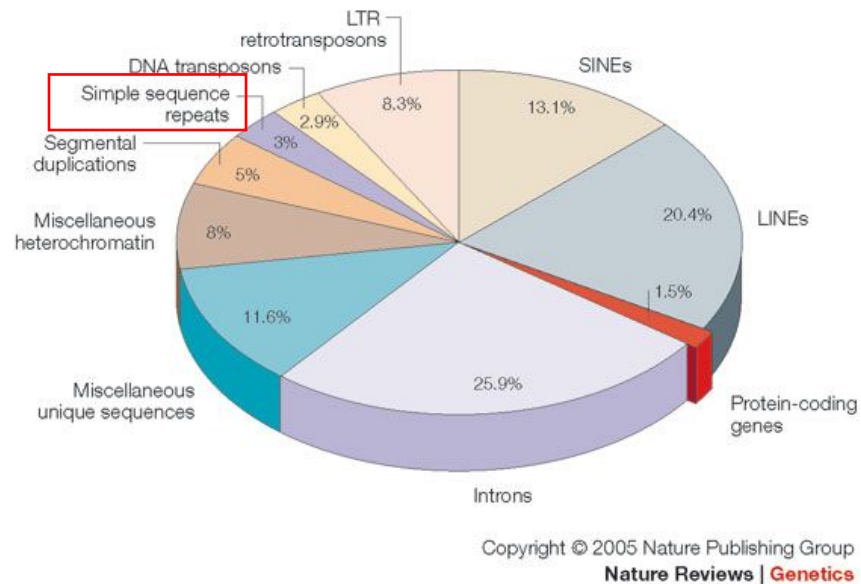


# Het humane genome contains 50% repeats

**A**



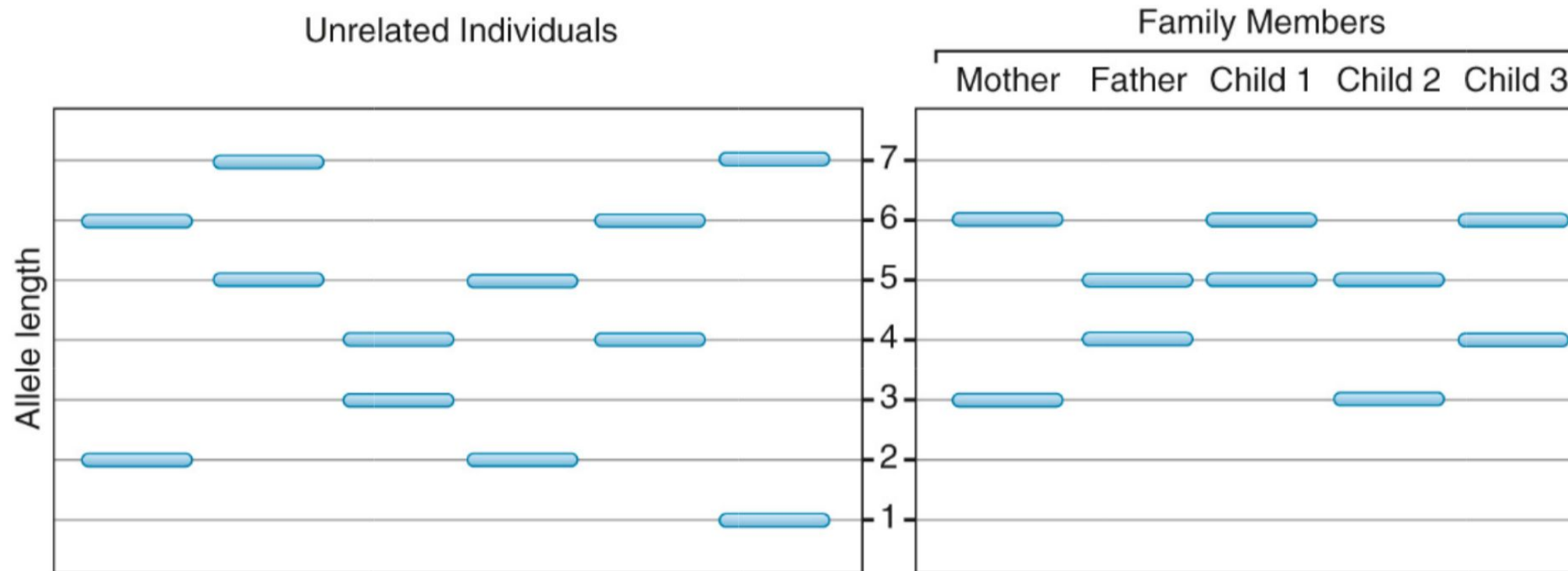
# Het humane genoom bevat 50% repeats



# Short tandem repeat (STR)

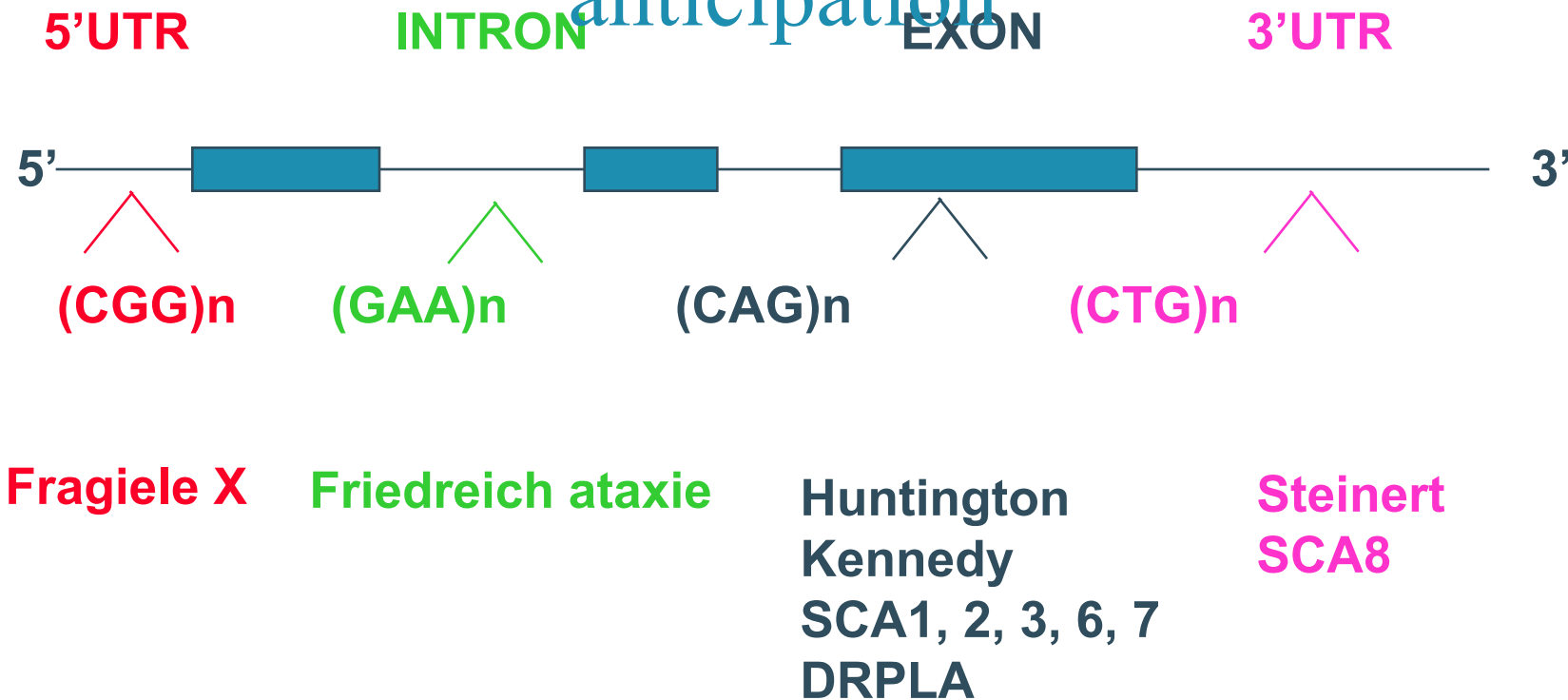
- Short tandem repeat = Microsatellites = Variable number of tandem repeats = simple sequence repeats
- They have specific unit: e.g. **CGG**
- That is repeated:  
**CGGCGGCGGCGGCGGCGG**
- **± 1 miljoen STRs** in the human genome
- Tandem repeats can have a big impact on phenotype

# A schematic of a hypothetical microsatellite marker in human DNA.





are associated with disorders characterized by anticipation

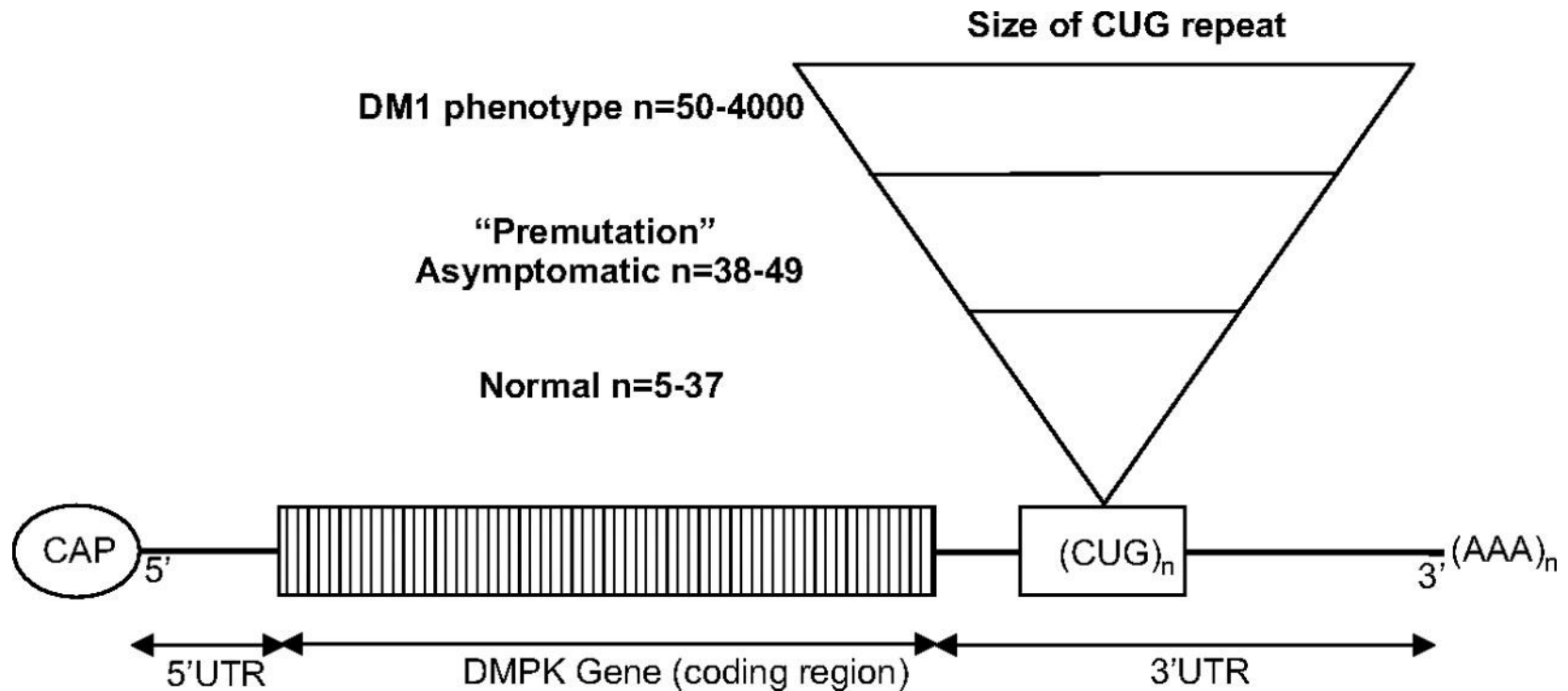


# Myotonic dystrophy (Steinert disease) (as an example)

- *Autosomal dominant*
- *Trinucleotide repeat expansion*



# DMPK pre-mRNA with relationship between CUG repeat size and phenotype.

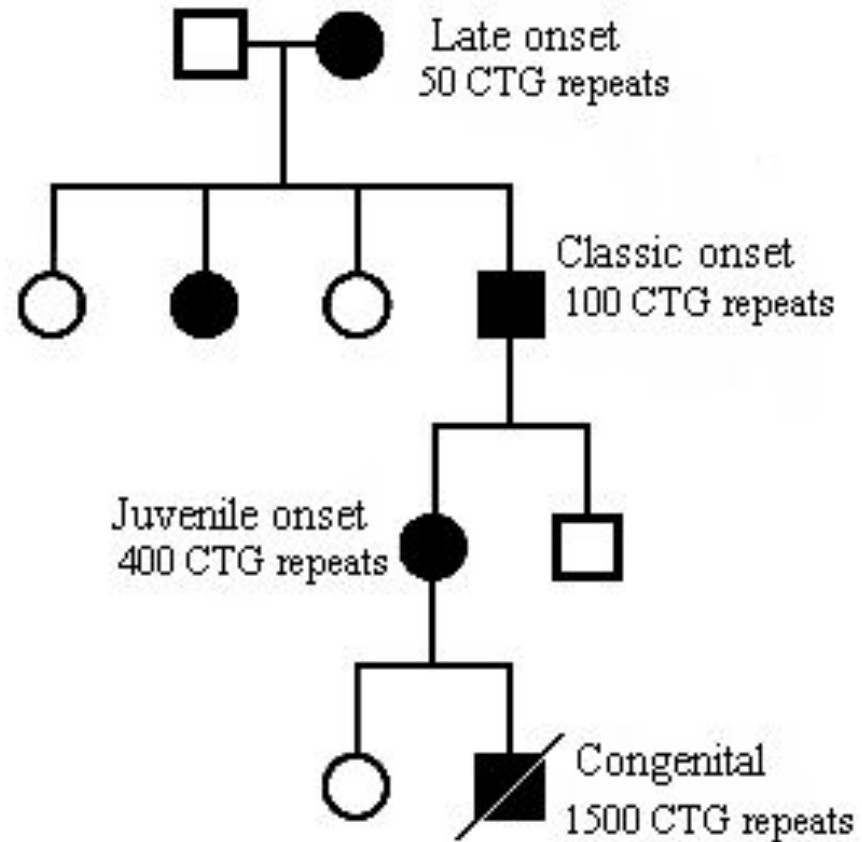
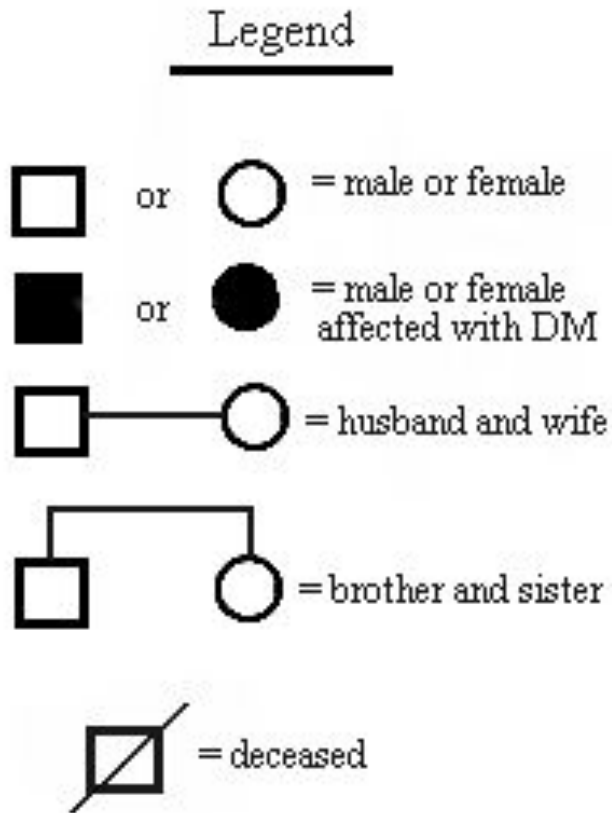


# Myotonic dystrophy (Steinert disease)

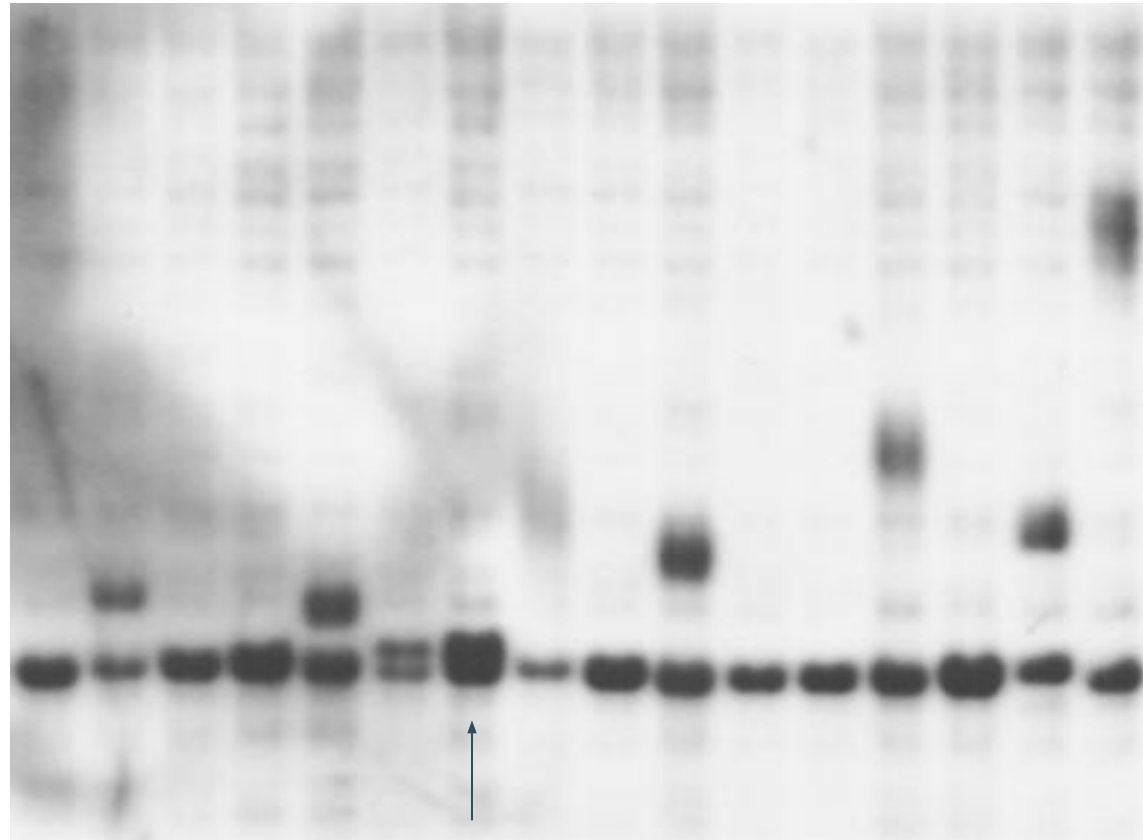
## Hereditary aspects

- Anticipation
  - Increasing severity and successive generations
- Maternal transmission for large expansions
- Often paternal transmission in case of smaller expansions.

# DM: anticipation



# Myotonic dystrophy (Steinert disease): detection



◀ n = 1500

◀ n = 500

◀ n = 350

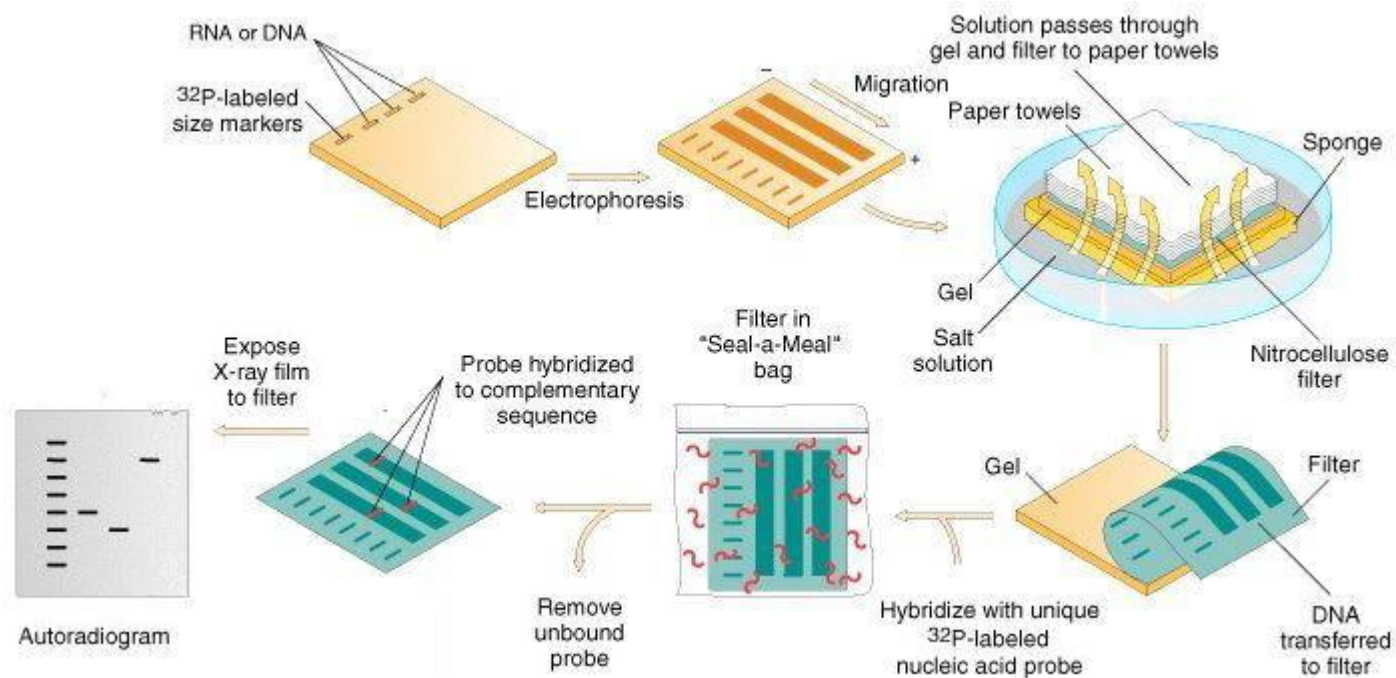
◀ n = 100

◀ n = 4-12

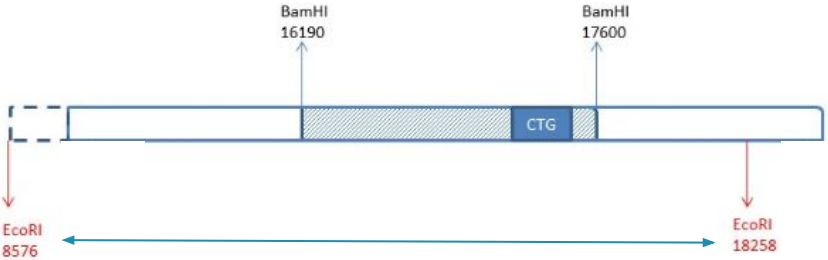
n = 35

# Southern blot om lange expansies te bepalen:

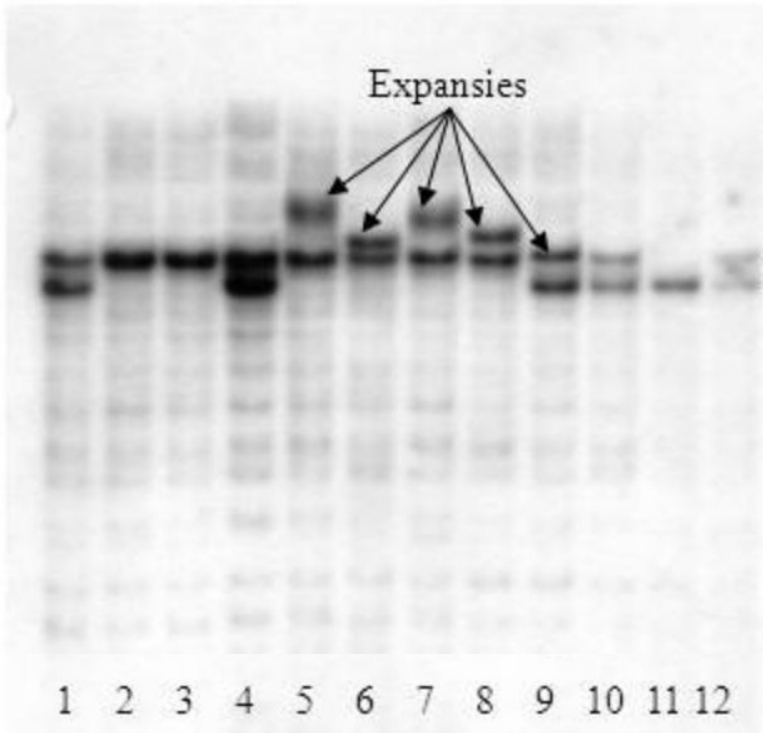
## 1. Knippen van humaan genoom met restrictie enzymen



# Southern blot to detect large expansions



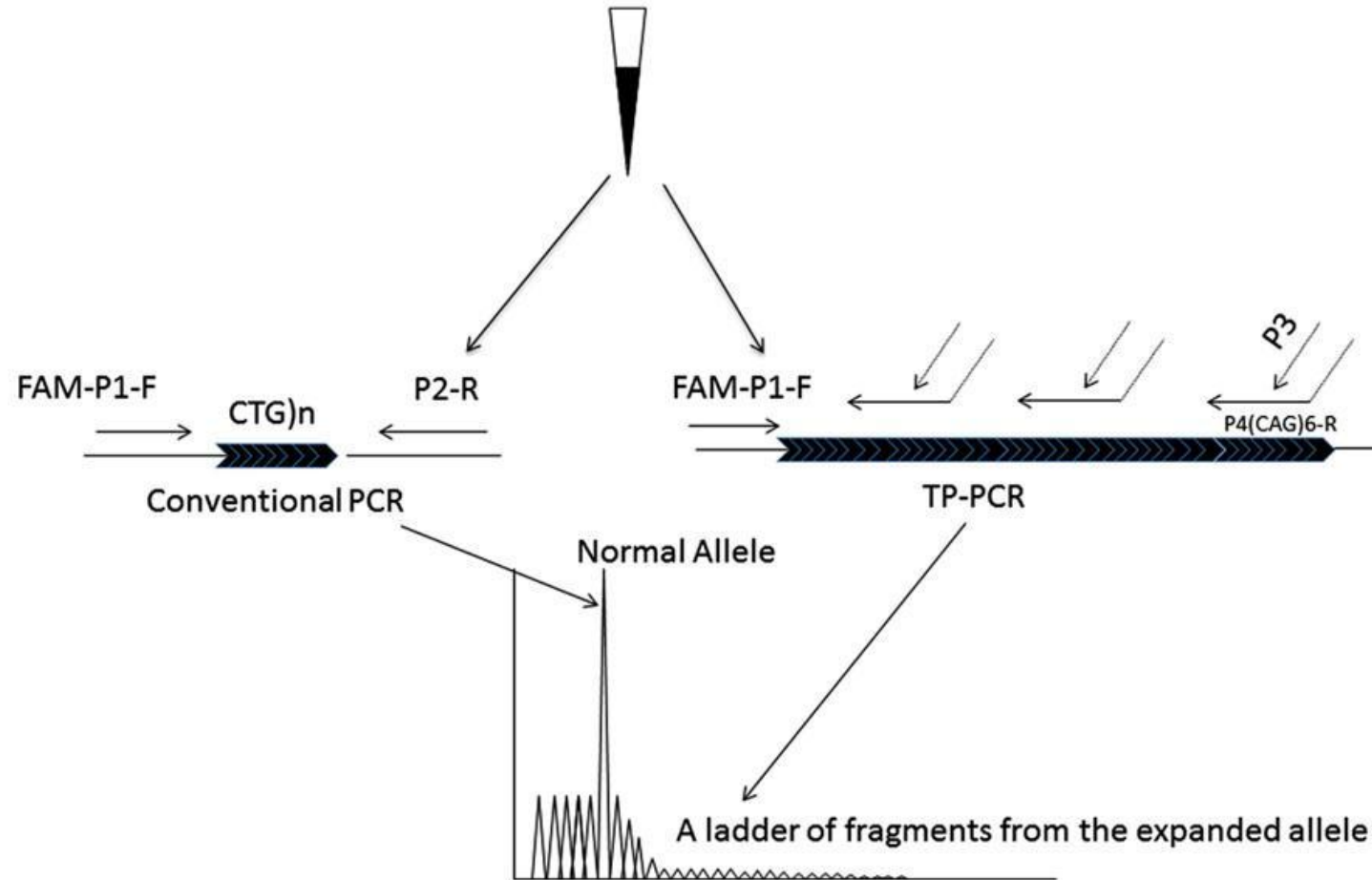
9682 bp



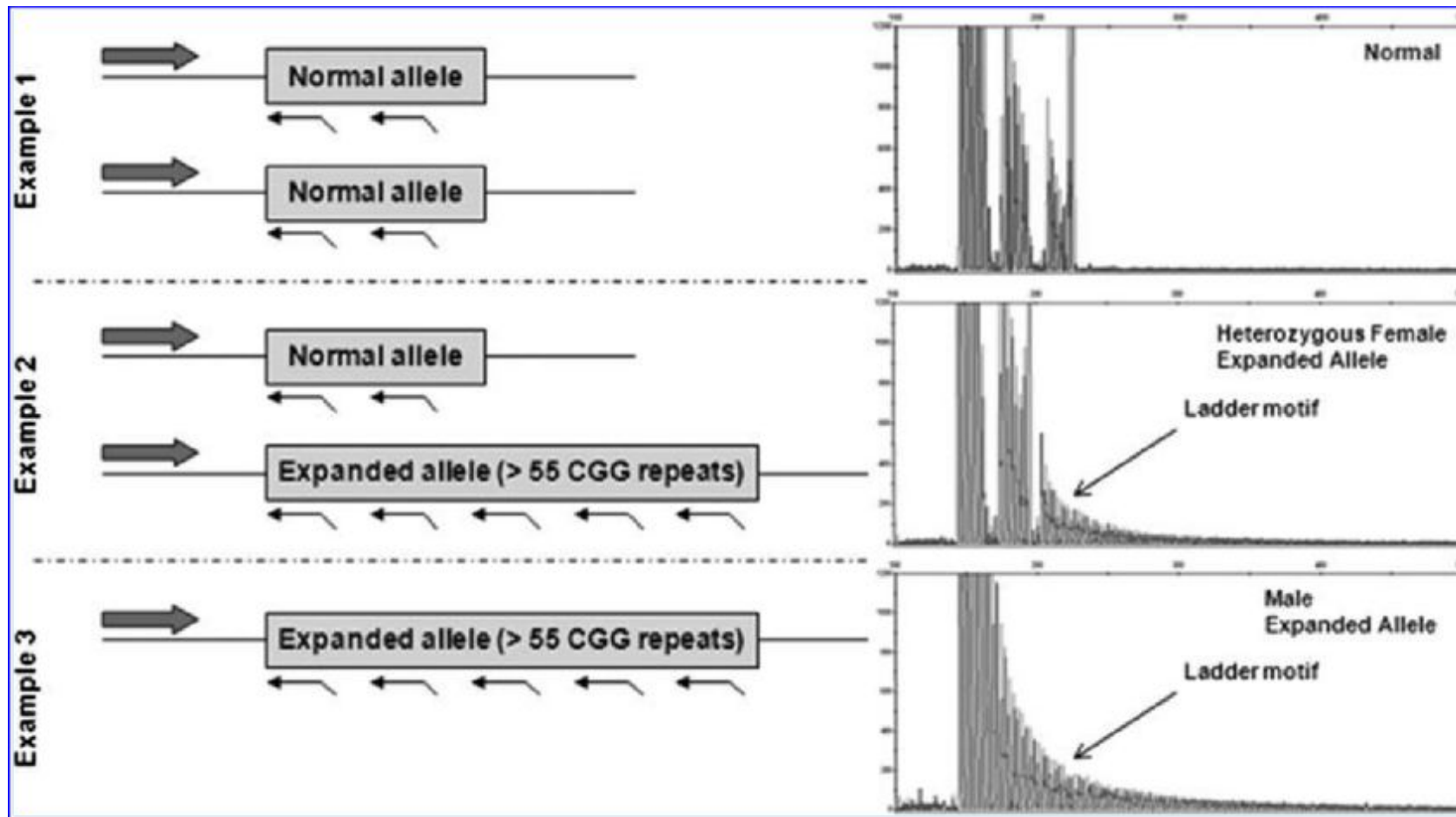
- > Fragment can be recognized via probe
- > Fragment will be larger with larger expansions



# Long expansions can be detected by Triplet primed-PCR (TP-PCR)


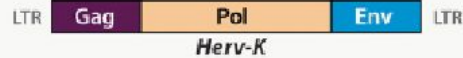
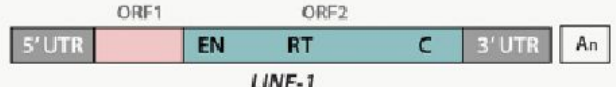


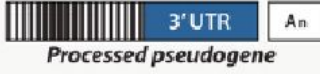


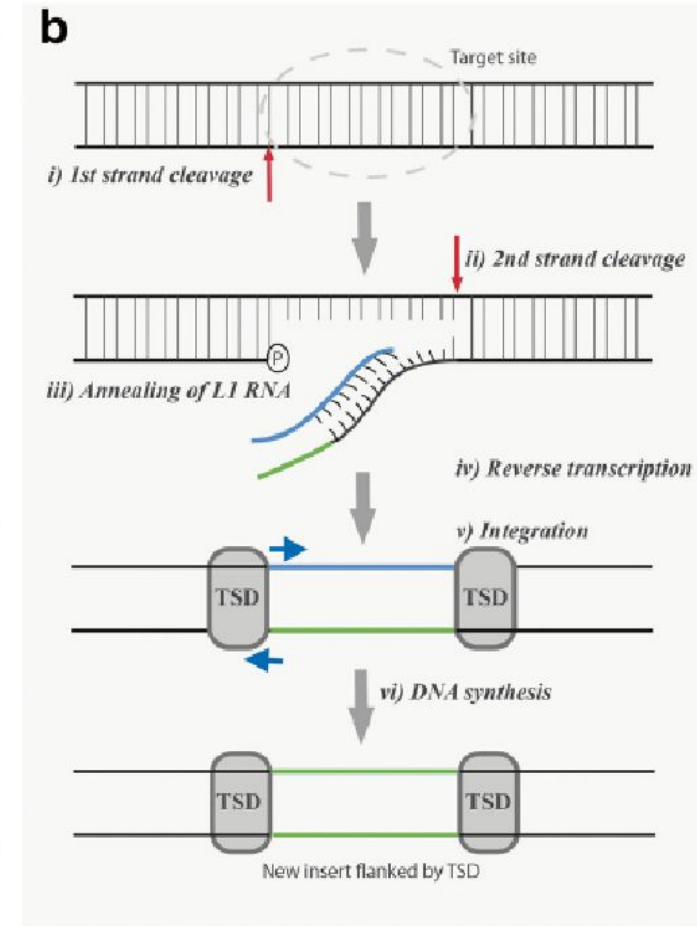
# Triplet primed PCR



# Active mobile elements in the human genome

**a**

Mobile element structure	HGR	Length	Remarks	
<b>DNA transposons:</b>  IR Transposase IR <i>Mariner</i>	2-3%	1.4 kb	Transposes from less than 100 kb to distant sites from original site	
<b>Retrotransposons (autonomous):</b>  LTR Gag Pol Env LTR <i>HerV-K</i>  5' UTR ORF1 EN RT C 3' UTR An <i>LINE-1</i>	7-9%	±1.4 kb	Retrovirus like structure with defective envelope gene, reinserts in the same genome from which they come	
<b>Non-autonomous:</b>  Left arm A B An Right arm Ins An <i>Alu</i>  TSD (CCCTCT) <sub>n</sub> Alu VNTR SINE-R An TSD <i>SVA</i>  3' UTR An <i>Processed pseudogene</i>	17-19%	6 kb	Only autonomously active mobile elements in primates and humans	
		0.3 kb	Alu insertions accounts for over 20 cases of human genetic diseases	
		11-13%	±1.5 kb	SVA insertions occurs at high frequency, so far 3 cases of human disease reported
		Variable	Arises by reverse transcription of cellular mRNA & integration of cDNA in the genome	



# SNPs : Common variation in the genome

Type of Variation	Size Range (approx.)	Basis for the Polymorphism	Number of Alleles
Single nucleotide polymorphisms	1 bp	Substitution of one or another base pair at a particular location in the genome	Usually 2
Insertion/deletions (indels)	1 bp to > 100 bp	<i>Simple:</i> Presence or absence of a short segment of DNA 100-1000 bp in length <i>Microsatellites:</i> Generally, a 2-, 3-, or 4-nucleotide unit repeated in tandem 5-25 times	<i>Simple:</i> 2 <i>Microsatellites:</i> typically 5 or more
Copy number variants	10 kb to > 1 Mb	Typically the presence or absence of 200-bp to 1.5-Mb segments of DNA, although tandem duplication of 2, 3, 4, or more copies can also occur	2 or more
Inversions	Few bp to > 1 Mb	A DNA segment present in either of two orientations with respect to the surrounding DNA	2

# Short read genome sequencing

## Benchtop Sequencers

iSeq™ 100



MiniSeq™



7.5 Gb | 25M | 2x150

MiSeq™



15Gb | 25M | 2x300

## Production-Scale Sequencers

NextSeq™ 550



120Gb | 400M | 2x150

HiSeq™ 4000



1500Gb | 5B | 2x150

HiSeq X™ (5 or 10)



1800Gb | 6B | 2x150

NovaSeq™ 6000



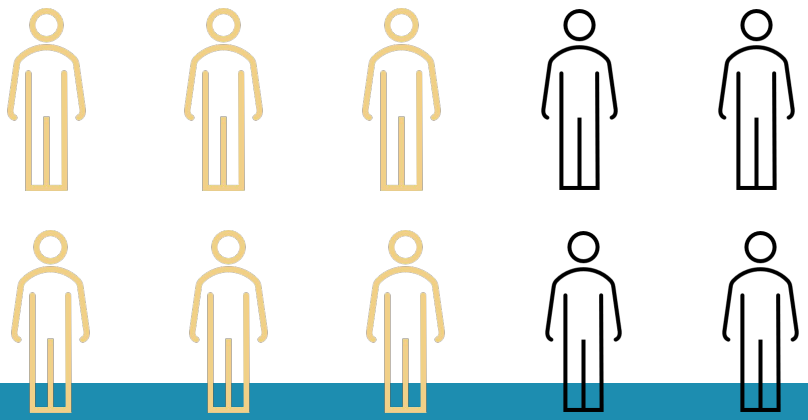
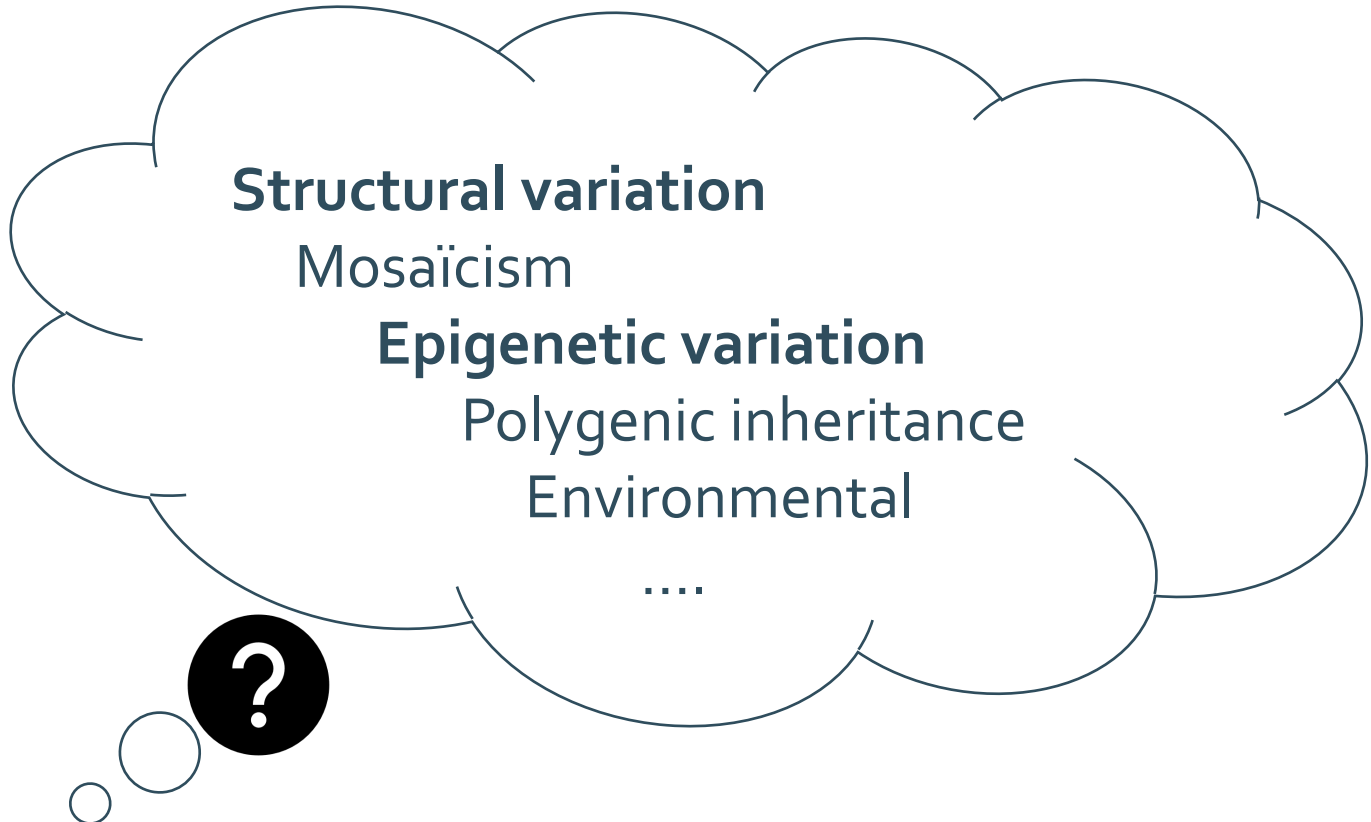
6000Gb | 20B | 2x150

NovaSeq  
Xplus

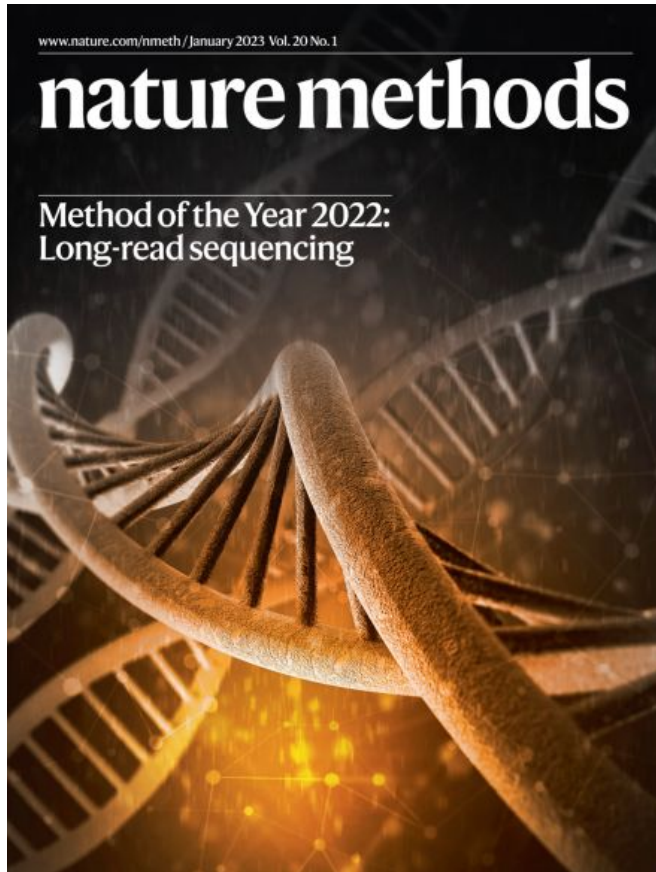


16Tb/  
128 human  
genomes/run

# GAP of short read sequencing

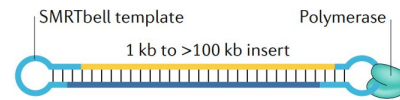


# third generation sequencing bridging the gap

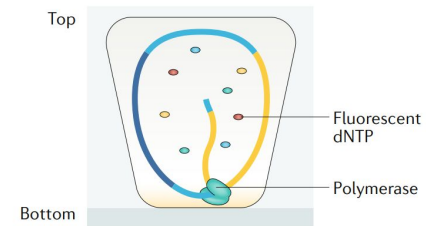


## a PacBio SMRT sequencing

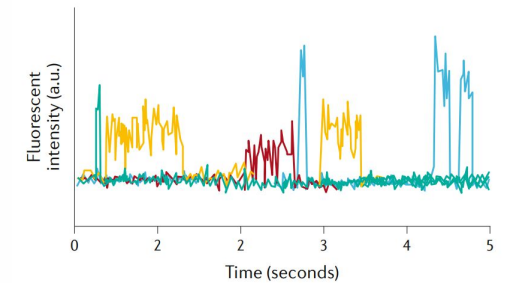
### Template topology



### Single ZMW (cross section)

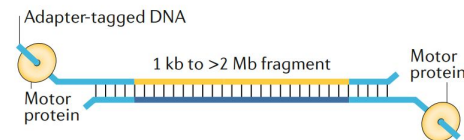


### Readout

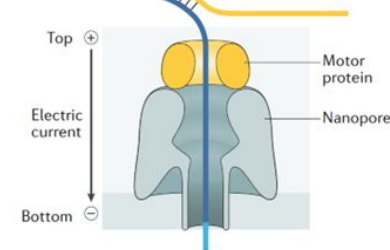


## b ONT sequencing

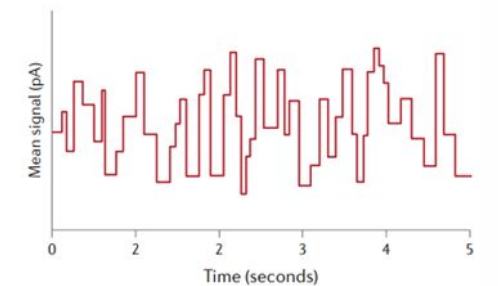
### Template topology

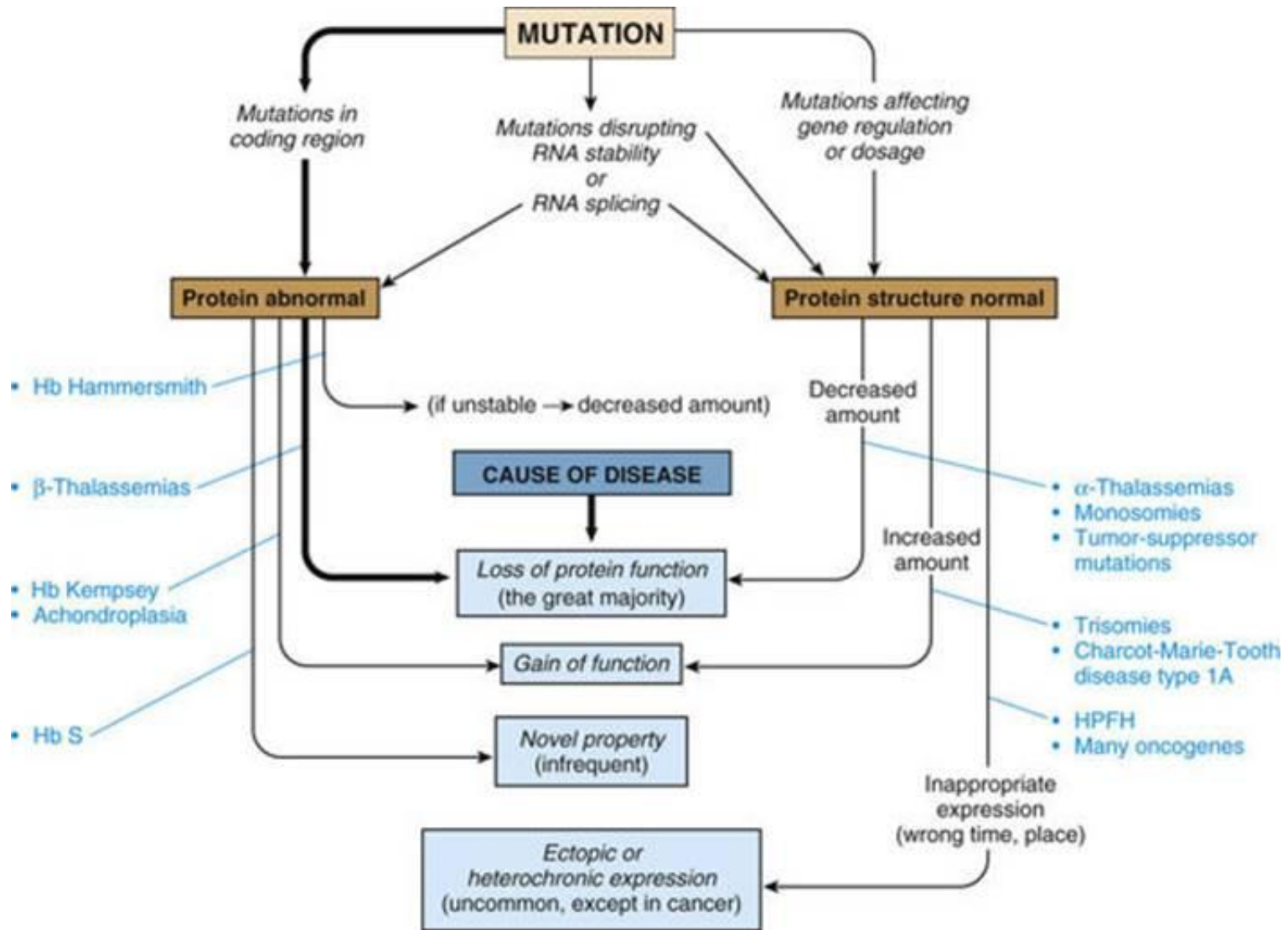


### Single nanopore (cross section)



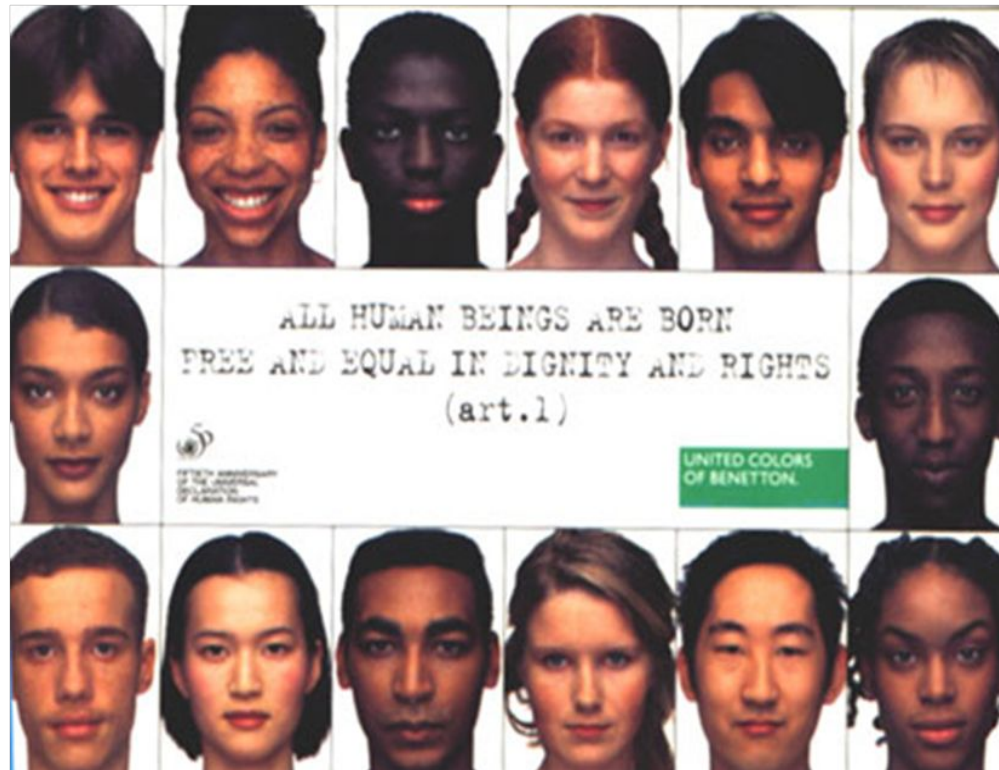
### Readout





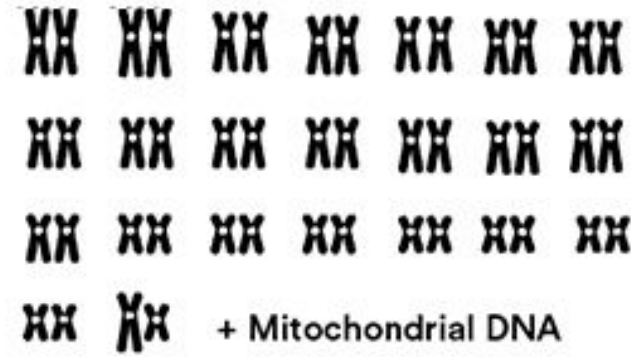
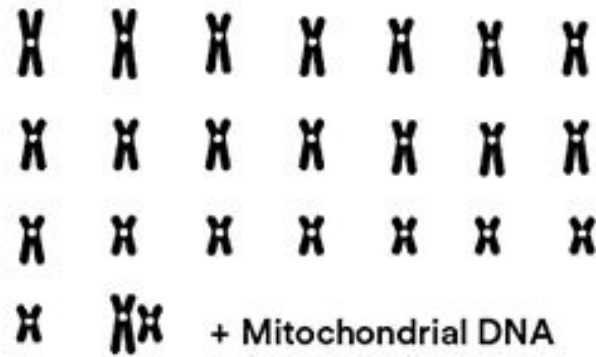


# Variation in individual genomes



## Reference Genome

## A Person's Genome



What is it?

How many chromosomes?

**24**  
(22 + X + Y)

**46**  
(23 PAIRS)

How many letters?

**~ 3.2 bn**

**~ 6.4 bn**

How to think about it?

- The Human Genome Project and its goal of a first draft of "the human genome"
- Serves as a standard for comparison
- A "consensus" genome sequence

- The genome of a person
- The genome within a person's cells
- The whole genome sequence of an individual

Press Releases: 14th April 2003

## The Finished Human Genome - Wellcome To The Genomic Age

Navigation

le.  
est  
d  
e

The  
anno

### 1000 Ge

A Deep Catalog of Human Genetic Variation



OPEN ACCESS Freely available online

PLOS BIOLOGY

## The Diploid Genome Sequence of an Individual Human

Samuel Levy<sup>1\*</sup>, Granger Sutton<sup>1</sup>, Pauline C. Ng<sup>1</sup>, Lars Feuk<sup>2</sup>, Aaron L. Halpern<sup>1</sup>, Brian P. Walenz<sup>1</sup>, Nelson Axelrod<sup>1</sup>, Jiaqi Huang<sup>1</sup>, Ewen F. Kirkness<sup>1</sup>, Gennady Denisov<sup>1</sup>, Yuan Lin<sup>1</sup>, Jeffrey R. MacDonald<sup>2</sup>, Andy Wing Chun Pang<sup>2</sup>, Tsiamouri<sup>1</sup>, Vineet Bafna<sup>3</sup>, Vikas Bansal<sup>3</sup>, Saul A. Kravitz<sup>1</sup>, Dana A. Busam<sup>1</sup>, .. Remington<sup>1</sup>, Josep F. Abril<sup>4</sup>, John Gill<sup>1</sup>, Jon Borman<sup>1</sup>, Yu-Hui Rogers<sup>1</sup>, ert L. Strausberg<sup>1</sup>, J. Craig Venter<sup>1</sup>



Vol 452 | 17 April 2008 | doi:10.1038/nature06884

# nature

International weekly journal of science

Home | News & Comment | Research | Careers & Jobs | Current Issue | Archive | Audio & Video | For Authors

Archive > Volume 532 > Issue 7600 > News > Article

NATURE | NEWS

## AstraZeneca launches project to sequence 2 million genomes

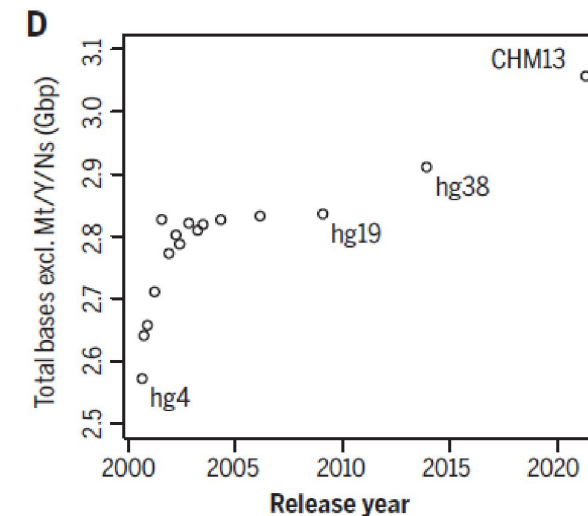
### Drug company aims to pool genomic and medical data in hunt for rare genetic sequences

the DNA's double helix structure, followed in 2007, and later the DNA of gene hunter Craig Venter. Recently the completion of the sequences of two Yoruba Africans was

# The Human Reference genome

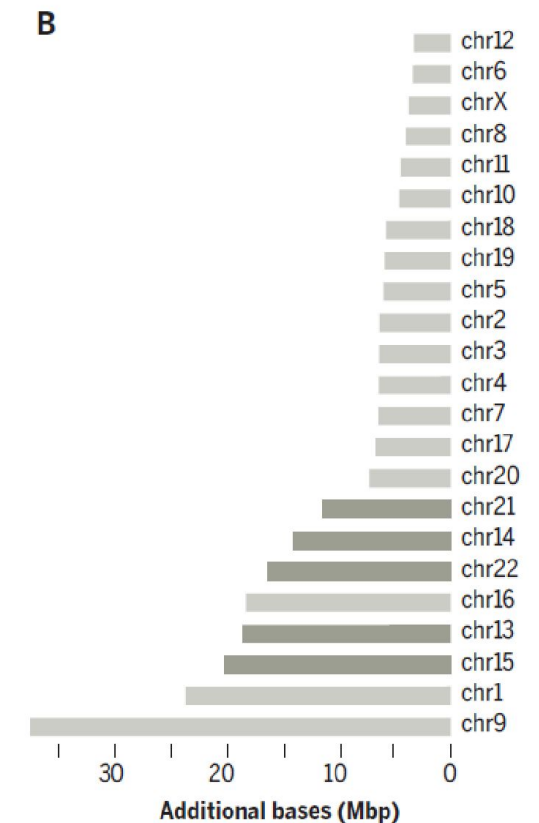
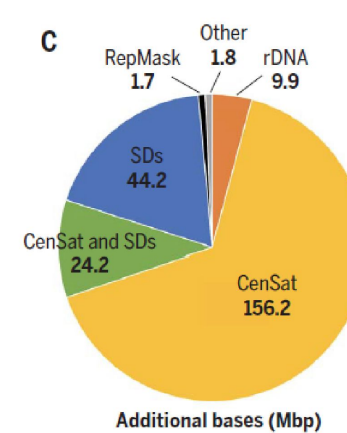
- Genome reference consortium
- Made up of 13 anonymous volunteers. 80% from 8 individuals.
- First version (2001) had 150000 gaps, the latest release, GRCh38 was released December 2013 and contained around 250 gaps.
- Issue: 'The' reference does not exist!

Release name	Date of release	Equivalent UCSC version
GRCh38	Dec 2013	hg38
GRCh37	Feb 2009	hg19
NCBI Build 36.1	Mar 2006	hg18
NCBI Build 35	May 2004	hg17
NCBI Build 34	Jul 2003	hg16

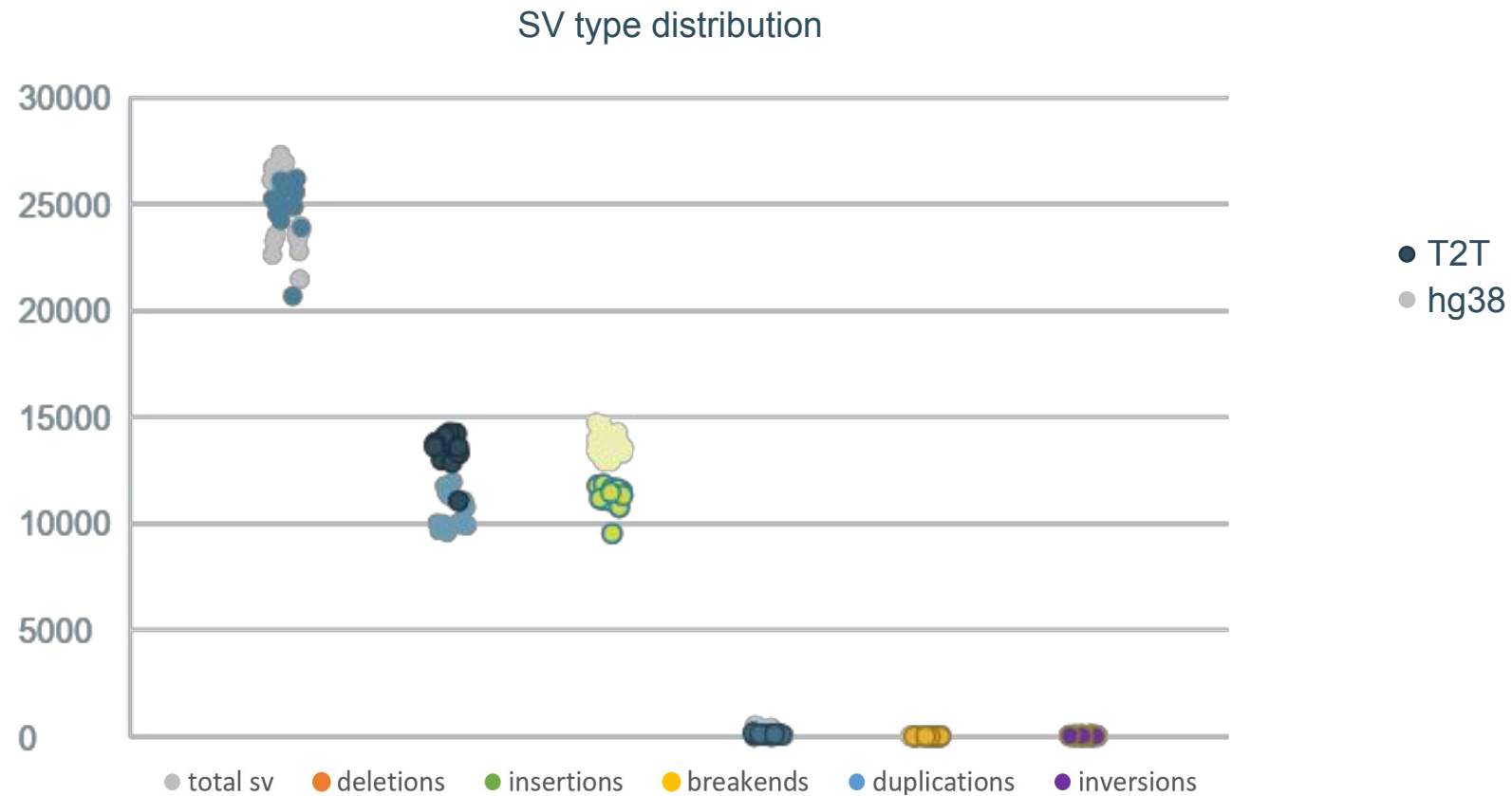


# The T2T reference genome: The complete sequence of a human genome (2022)

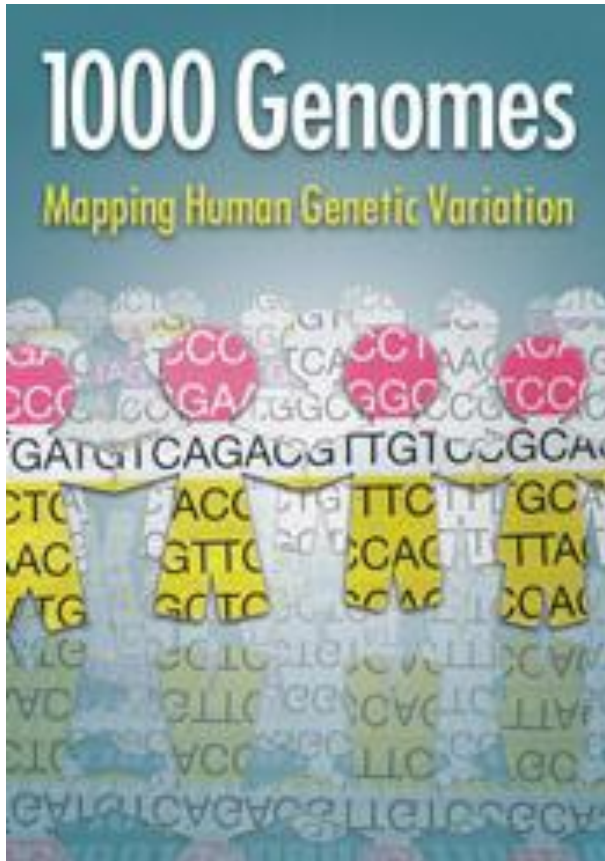
STATISTICS	GRCH38	T2T-CHM13	DIFFERENCE (±%)
<b>Summary</b>			
Assembled bases (Gbp)	2.92	3.05	+4.5
Unplaced bases (Mbp)	11.42	0	-100.0
Gap bases (Mbp)	120.31	0	-100.0
Number of contigs	949	24	-97.5
Contig NG50 (Mbp)	56.41	154.26	+173.5
Number of issues	230	46	-80.0
Issues (Mbp)	230.43	8.18	-96.5
<b>Gene annotation</b>			
Number of genes	60,090	63,494	+5.7
Protein coding	19,890	19,969	+0.4



# the new reference genome adds 200 Mb and improves mapping

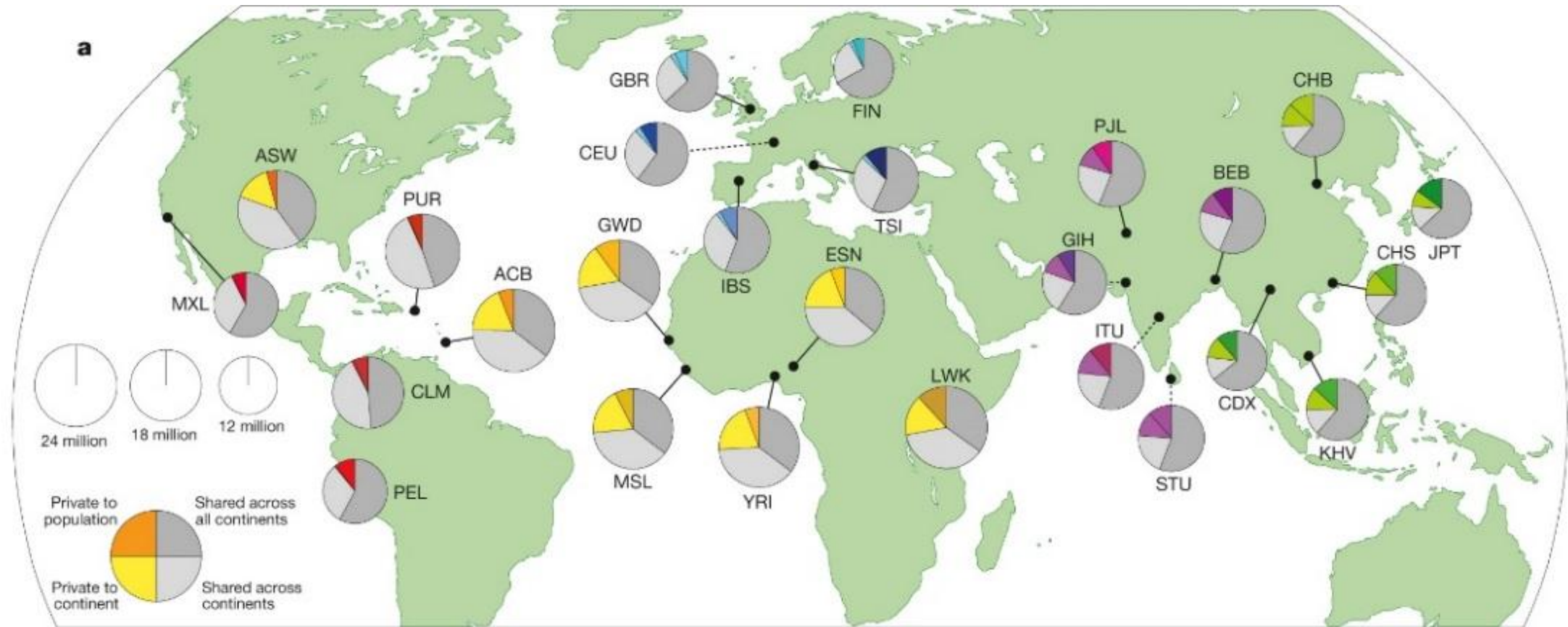


# 1000 genome project/resource



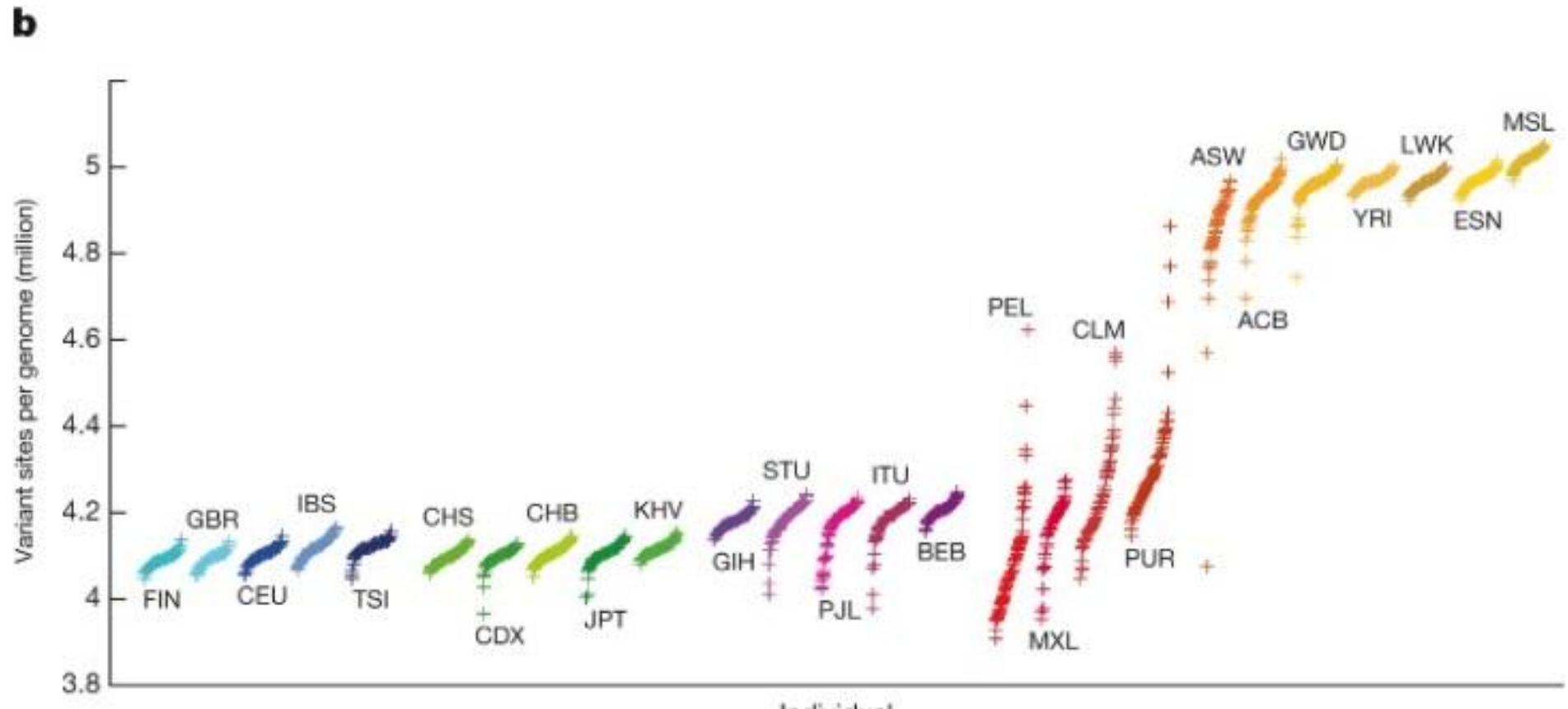
- Comprehensive description of common human genetic variation
- Latest report: genomes of 2,504 individuals from 26 populations using a combination of low-coverage whole-genome sequencing, deep exome sequencing, and dense microarray genotyping.
- Results:
  - over **88 million variants** (84.7 million single nucleotide polymorphisms (SNPs))
  - **3.6 million short insertions/deletions** (indels), and 60,000 structural variants), all phased onto high-quality haplotypes.
  - This resource includes >99% of SNP variants with a frequency of >1% for a variety of ancestries.

# SNP variation/population





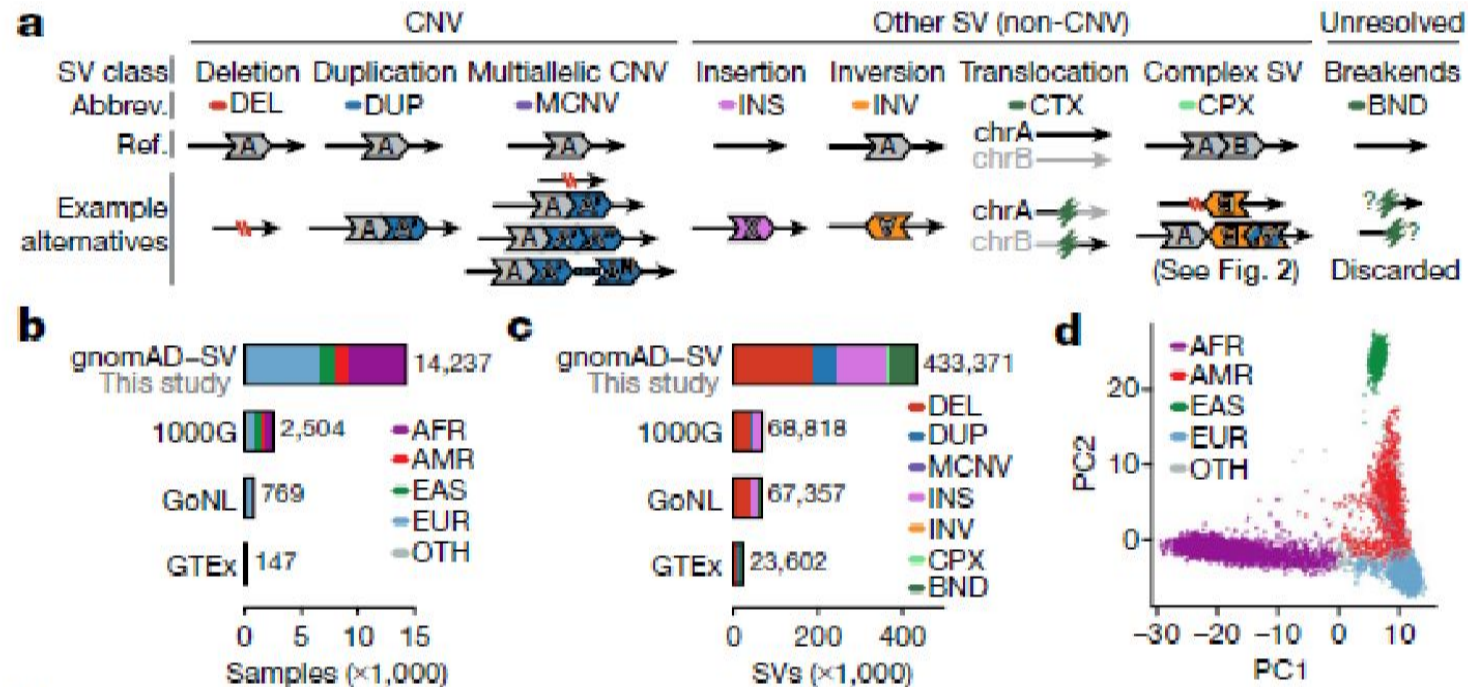
# SNP variation/population



# Structural variation

genome aggregation database or GnomAD

Based on short read sequencing in 14290 genomes



## Article

# A structural variation reference for medical and population genetics

<https://doi.org/10.1038/s41586-020-2287-8>

Received: 2 March 2019

Accepted: 31 March 2020

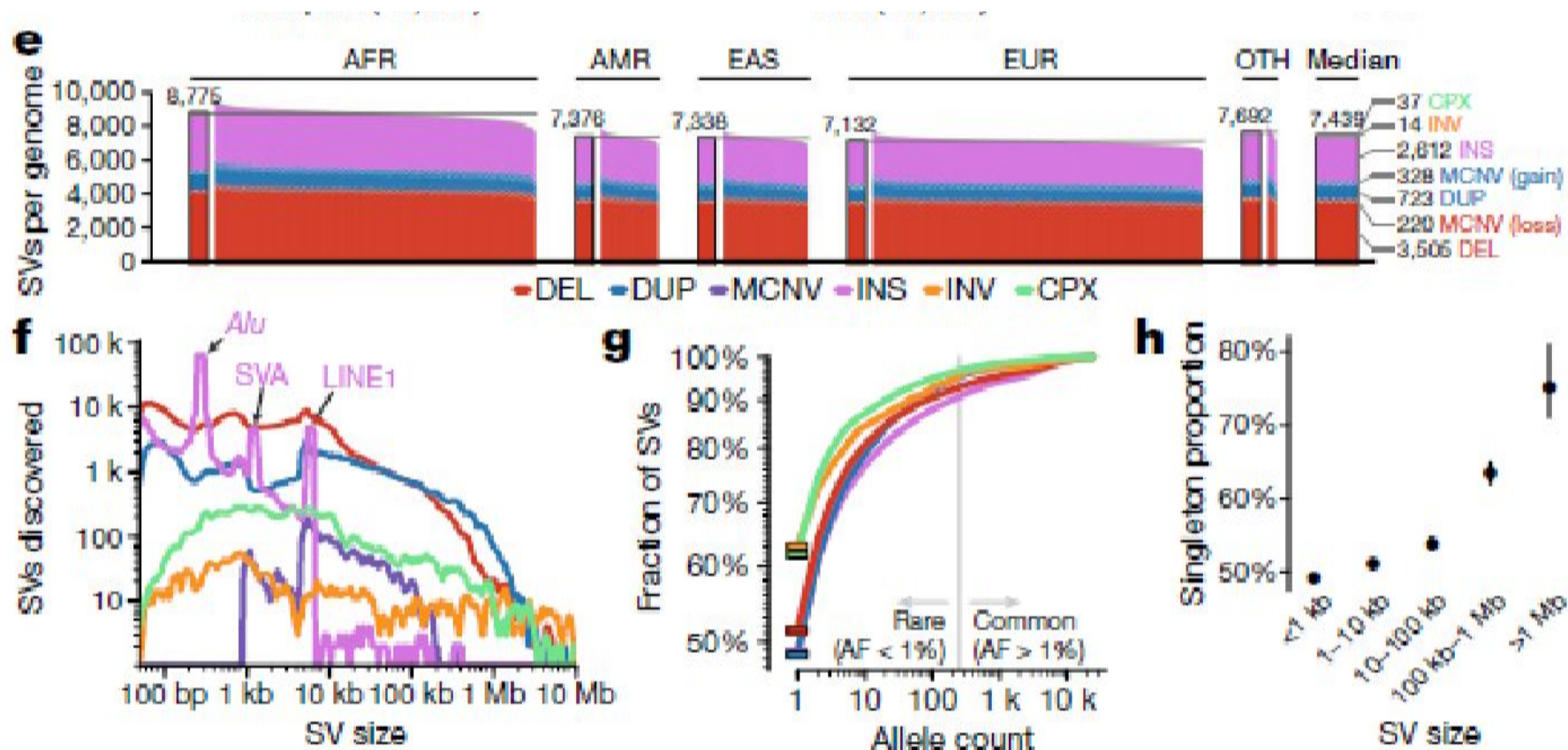
Published online: 27 May 2020

Open access

Check for updates

Ryan L. Collins<sup>1,2,3,10†</sup>, Harrison Brand<sup>1,2,4,10†</sup>, Konrad J. Karczewski<sup>1,5</sup>, Xuefang Zhao<sup>1,2,4</sup>, Jessica Alfoldi<sup>1,5</sup>, Laurent C. Franciolini<sup>1,5,6</sup>, Amit V. Khera<sup>1,2</sup>, Chelsea Lowther<sup>1,2,4</sup>, Laura D. Gauthier<sup>1,2</sup>, Harold Wang<sup>1,2</sup>, Nicholas A. Watts<sup>1,5</sup>, Matthew Solomonson<sup>1,5</sup>, Anne O'Donnell-Luria<sup>1,5</sup>, Alexander Baumann<sup>1,2</sup>, Ruchi Munshi<sup>1,2</sup>, Mark Walker<sup>1,2</sup>, Christopher W. Whelan<sup>1</sup>, Yongqing Huang<sup>1</sup>, Ted Brookings<sup>1</sup>, Ted Sharpe<sup>1</sup>, Matthew R. Stone<sup>1,2</sup>, Elise Valkanas<sup>1,2,3</sup>, Jack Fu<sup>1,2,4</sup>, Grace Tiao<sup>1,5</sup>, Kristen M. Laricchia<sup>1,5</sup>, Valentin Ruano-Rubio<sup>1</sup>, Christine Stevens<sup>1</sup>, Namrata Gupta<sup>1</sup>, Caroline Cusick<sup>1</sup>, Lauren Margolin<sup>1</sup>, Genome Aggregation Database Production Team<sup>1</sup>, Genome Aggregation Database Consortium<sup>1</sup>, Kent D. Taylor<sup>4</sup>, Henry J. Lin<sup>4</sup>, Stephen S. Rich<sup>4</sup>, Wendy S. Post<sup>10</sup>, Yil-Dar Ida Chen<sup>4</sup>, Jerome I. Rotter<sup>4</sup>, Chad Nusbaum<sup>1,5,4</sup>, Anthony Philippakis<sup>1</sup>, Eric Lander<sup>1,10,11</sup>, Stacey Gabriel<sup>1</sup>, Benjamin M. Neale<sup>1,2,3,10</sup>, Sekar Kathiresan<sup>1,2,6,14</sup>, Mark J. Daly<sup>1,2,3,10</sup>, Eric Banks<sup>1</sup>, Daniel G. MacArthur<sup>1,2,5,6,10,15,16</sup> & Michael E. Talkowski<sup>1,2,4,10</sup>

# Structural variation characteristics



# Mapping full spectrum of structural variation

- Human Structural variation consortium
- Comprehensive structural variation analysis with a multitude of techniques.
- 3 parent-child trios (Han, Puerto Rican, Yoruban)

**Table 1 Summary of sequencing statistics**

	Avg. seq. coverage	Avg. frag. length	Physical coverage
Pacific Biosciences	39.6 (child)	8165 (child)	39.6
	20.03 (parent)	9619 (parent)	
Oxford Nanopore	18.9 (HG00733)	11,993	18.9
Illumina short insert	74.5	694	171
Illumina IWGS	3	3475	159
Illumina 7 kb JMP	1.1	6973.2	39.2
10X Chromium	82.4	90,098	53.9
Bionano Genomics	N/A	2.81E + 05	116.7
Trio-Seq SLR	3.47	4900	3.47
Strand-seq	N/A	N/A	5.87
Hi-C	19.49	1.03E + 07	N/A
Total	2235.6		607.08

Physical coverage is given for Illumina short insert, IWGS, 7 kb JMP. 10X Chromium physical coverage is estimated read cloud coverage  
 For Hi-C, fragment length is the distance between two read ends for intra-chromosome read pairs



ARTICLE

<https://doi.org/10.1038/s41467-018-08148-z> OPEN

Multi-platform discovery of haplotype-resolved structural variation in human genomes

Mark J.P. Chaisson et al.<sup>‡</sup>

# Per genome variation

(3-7x more than known from short read sequencing)

818000 indels (<50bp)

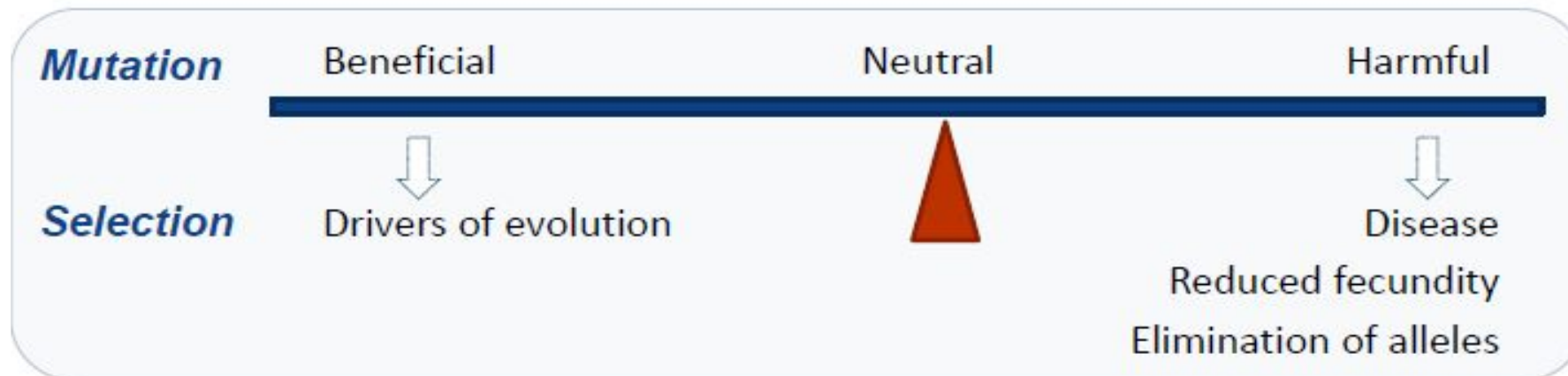
31599 structural variants (>50bp)

156 inversions (>50bp)

# Origin and frequency of de novo variation

# Selection-mutation balance

*“balance between genetic copying errors that turn normal alleles into harmful mutations, and selection eliminating these mutations”*



# Frequency of de novo mutations

## Estimation per generation mutation rate

$7.6 \times 10^{-9}$  to  $2.2 \times 10^{-8}$  = **50-100 *de novo* mutations per genome**

These mutations are under limited selective pressure!

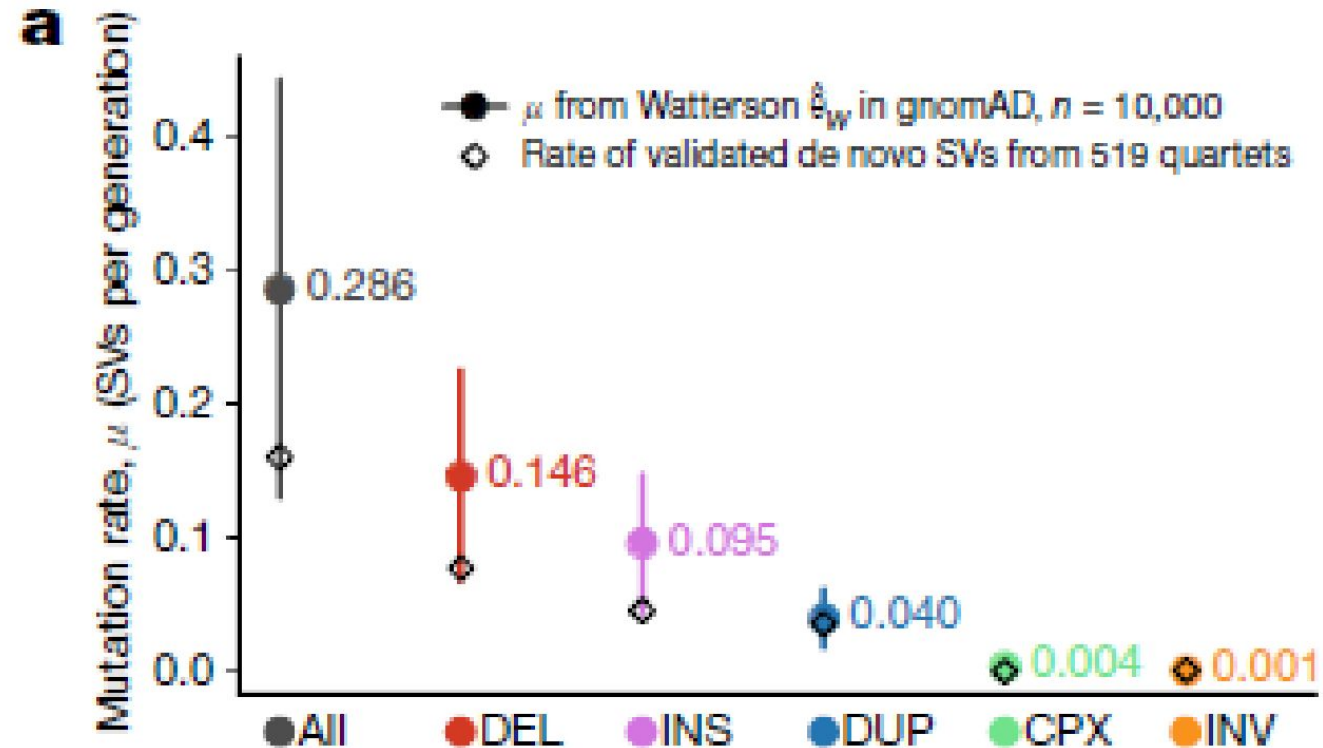
Estimated **de novo mutations per exome:**  
**1.4 exonic mutations/ individual**



# Frequency of de novo structural variation

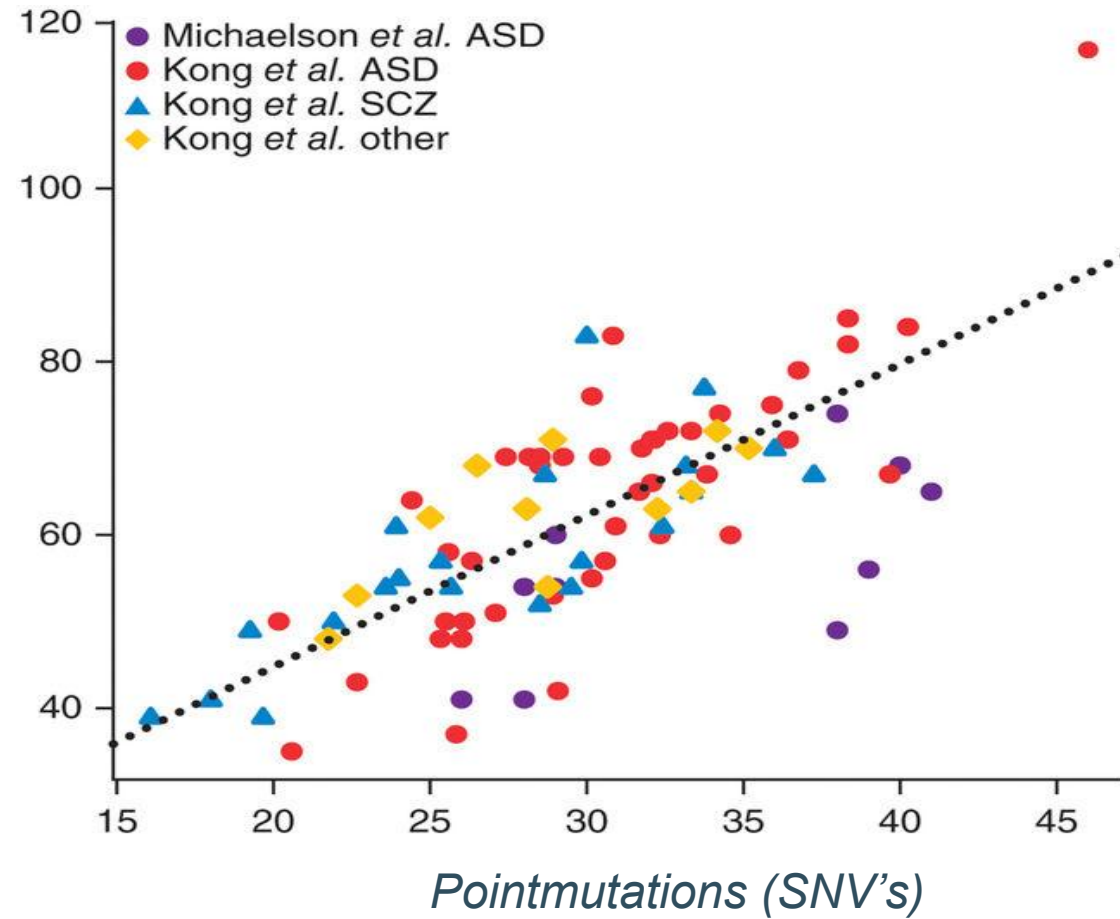
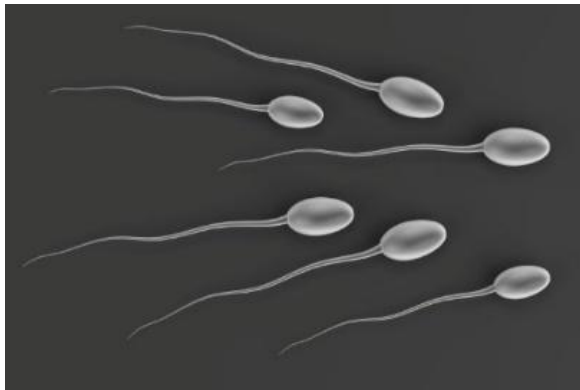
0.29 de novo SVs per generation in regions of the genome accessible to short-read WGS or 1 per 2-8 live births

# Frequency varies along types of SVs

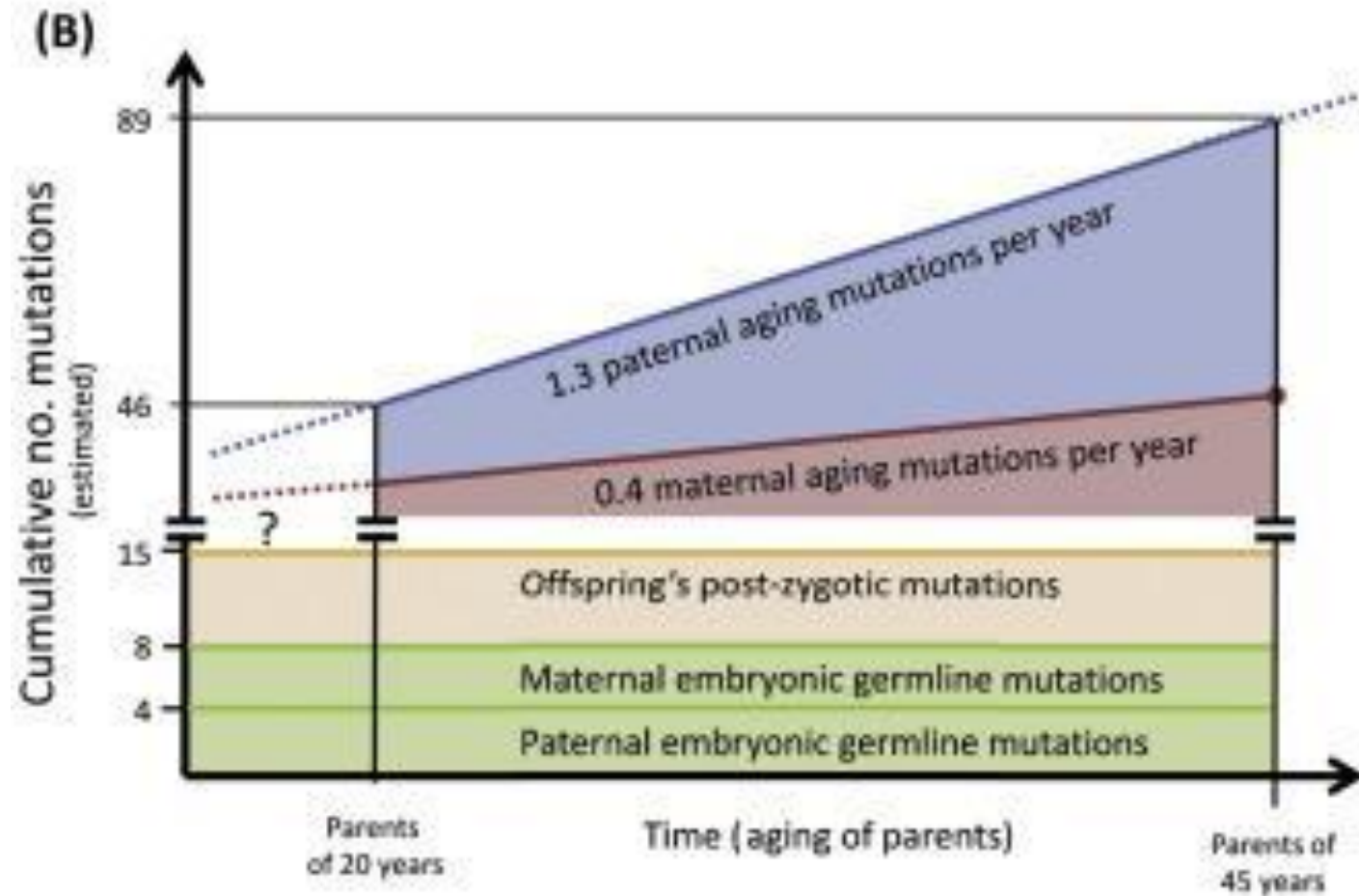


# SNV frequency increases with paternal age

in zaadcel



The de novo mutation rate is the som of mutations in sperm, oocyte and first 100 cell divisions



# Overview

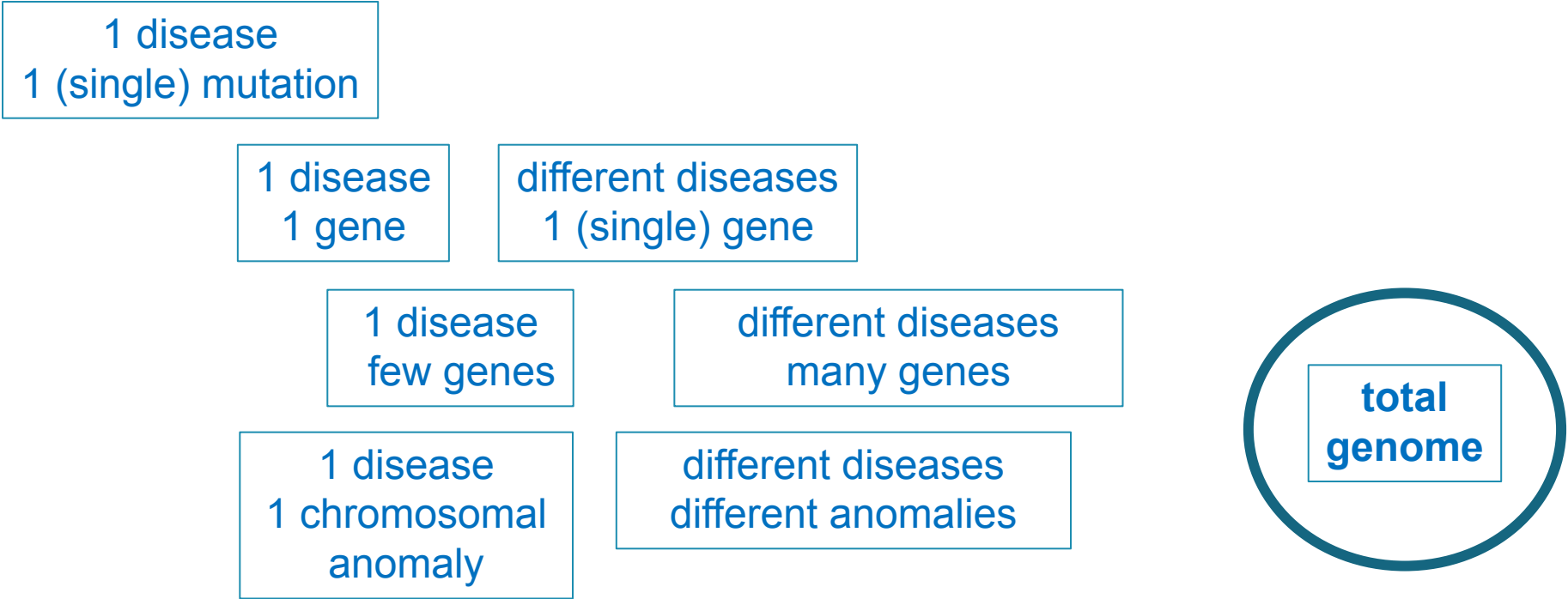
## Variation Detected in a Typical Human Genome

Individuals vary greatly in a wide range of biological functions, determined in part by variation among their genomes. Any individual genome will contain the following:

- ≈5-10 million SNPs (varies by population)
- 25,000-50,000 rare variants (private mutations or seen previously in < 0.5% of individuals tested)
- ≈75 new base pair mutations not detected in parental genomes
- 3-7 new CNVs involving ≈500 kb of DNA
- 200,000-500,000 indels (1-50 bp) (varies by population)
- 500-1000 deletions 1-45 kb, overlapping ≈200 genes
- ≈150 in-frame indels
- ≈200-250 shifts in reading frame
- 10,000-12,000 synonymous SNPs
- 8,000-11,000 nonsynonymous SNPs in 4,000-5,000 genes
- 175-500 rare nonsynonymous variants
- 1 new nonsynonymous mutation
- ≈100 premature stop codons
- 40-50 splice site-disrupting variants
- 250-300 genes with likely loss-of-function variants
- ≈25 genes predicted to be completely inactivated

# Genetic testing

## Molecular testing

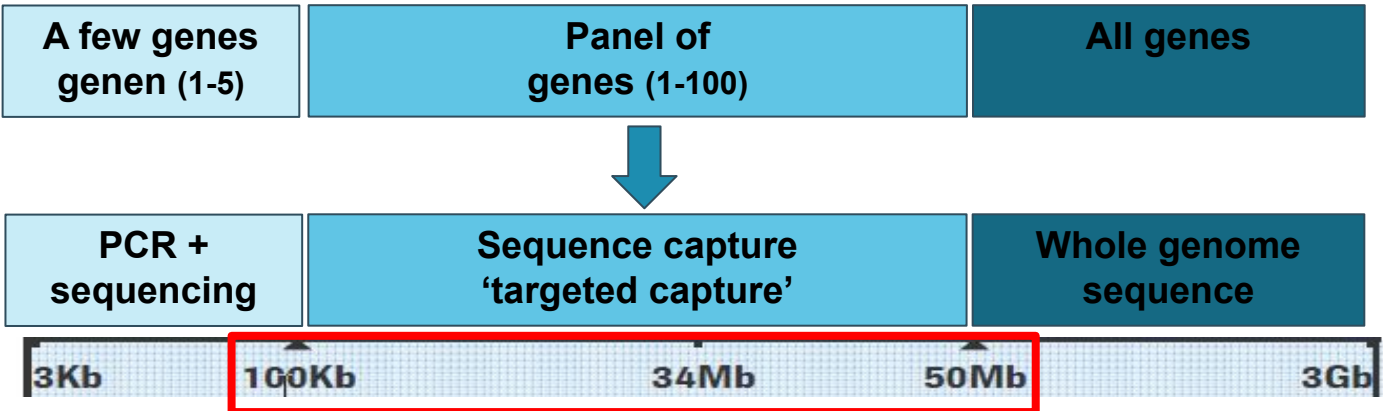
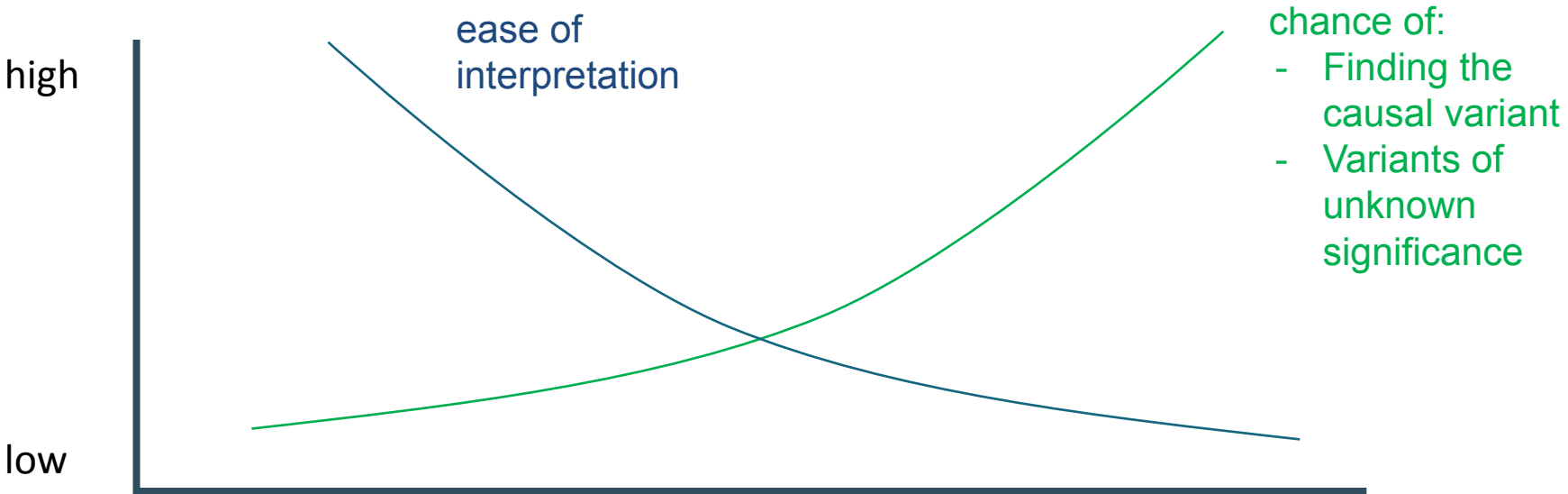


(un)known defect  
karyotyping

## Molecular cytogenetic testing

## Cytogenetics

# Molecular diagnostics WES/WGS



Θ Ζ ΕΥΧΕΡΙΑΝ ΟΣ ΤΡΑΣΧΙΑ ΦΑΒΡΙΚΑ ΨΑΙΝ Π Δ Ι C Τ Α Ε C C Ι Α Ε Ρ Ι G I  
S T A C I M E T O R D I N A M V O L E S A C I D E G E R N E E S D I C T A F A B R I C A E S S E  
D E P A T A A T O D S T I C T A A B O I A I A M S A S E V M A S S A Z O B V S A I I S T V R I B P R E  
T E E C C L E Z A D I L P R E G I M Z A D M I S T R A T I O N E E I G I D E B E S I G V I S A N I S P C A  
N I C O S E I D E E C C I E D V O S E X E I S Z T O T I I D C V E S R O M A O S B O N E A C L A  
B I L V E C O V S A T I O I S Z F A M E O S O T T E T A P A I I C V I H O S P I T A L N O E X  
A T O B P N O S Z S V C C E S S O E S N R O S R O M A O S P O N T F A D D I E T A D I N G I  
A T V N E P S I R O M A C V R I A R E S I D S Q O E S O P E R A R I I V O C E T C V A O  
A B R I C E D E F E S O R E P E C T O R E Z C O S E V A T O R E C O S T I T V I M A T O D E  
T A M A N T I O V I O R E C A D I N A L E I E A D E C V R I A P T P E E X I S T E N T E C V I P D I C  
S O P A R I O S I T E G R A R A T I O N E A D M I S T R A T I O N E P I D E F A B R I C E S I G V I S A  
R E D D V O I V M Z E O S A D I D P I P S V C A D N A I E C O G I P O S S E E T D E B E Q V A  
O P E V E R I T Q B O I B P E N A Z O I M O D A M S V P P M I S S I S A V C T O R I T A E A  
T O I I C A C O C E D M F A C V L T A E Z V T X P I F I D L E S A D T A N T V P I V E T L A M D  
P F E R V E T I O E S R E D D A T Q V O M N I O R A E X I N D E N O V E R I T A I A R S V A R C O  
D A A D P I S C I D O I P O E T I S D E I M I S E R I C O D A A C B E A T O R P E T R I Z P A V L A  
T O L O R E I V S A V C T O R C O F I S I C I B Z S I G V I X P I F I D L B P D C T I S V E R E P  
B Z C O F E S S I S T A P R E S E T T B O V A F V T V R Q V I E I D M F A B R I C E O V I O V A  
O R E Q S M O N E T E R O M A N V E I R O R V A I O R E A D M I N D O N A V T R I

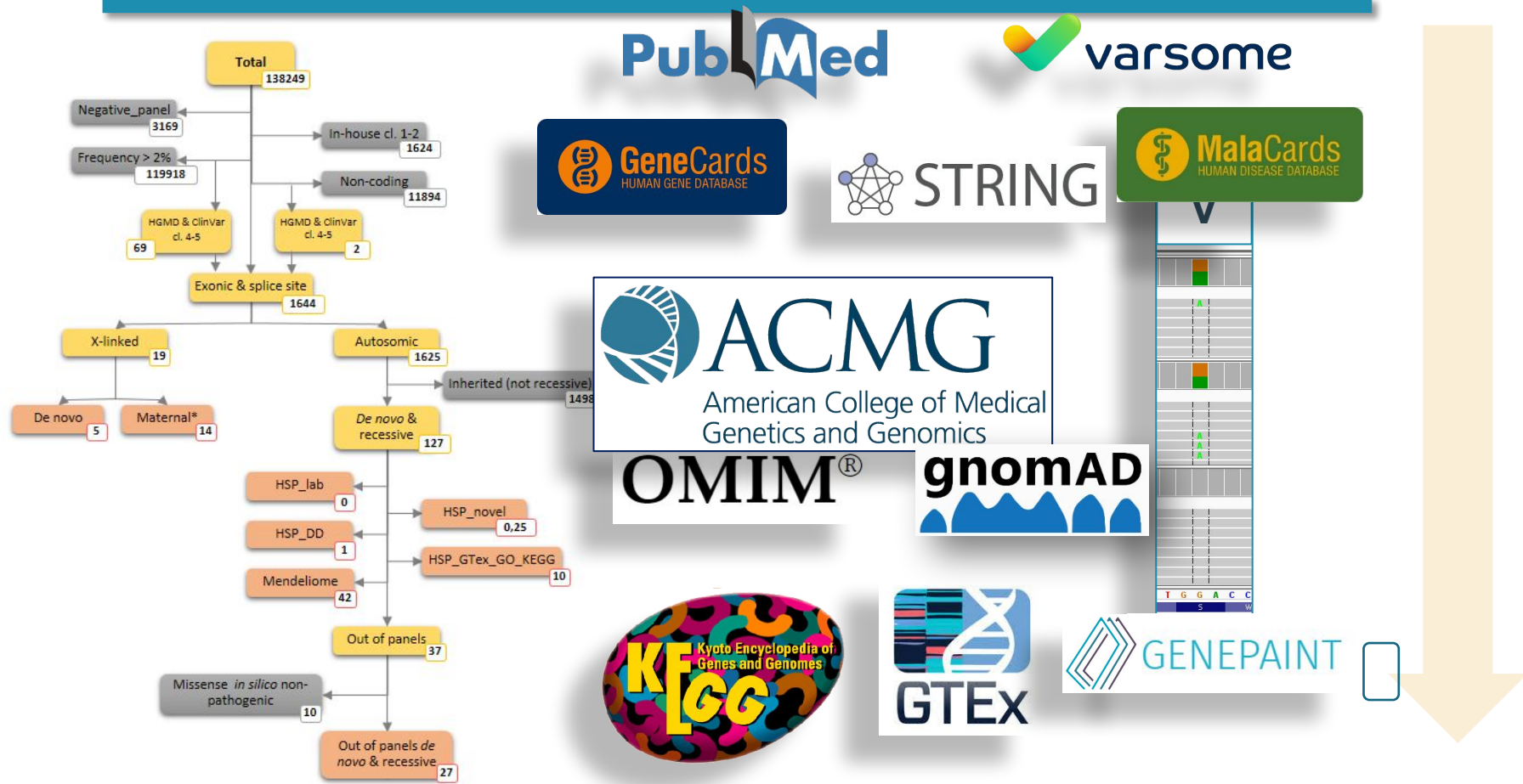


# Interpretation of variants

- Databases
  - SNP Databases > population frequenties
  - (internationale en lokale) mutation databases

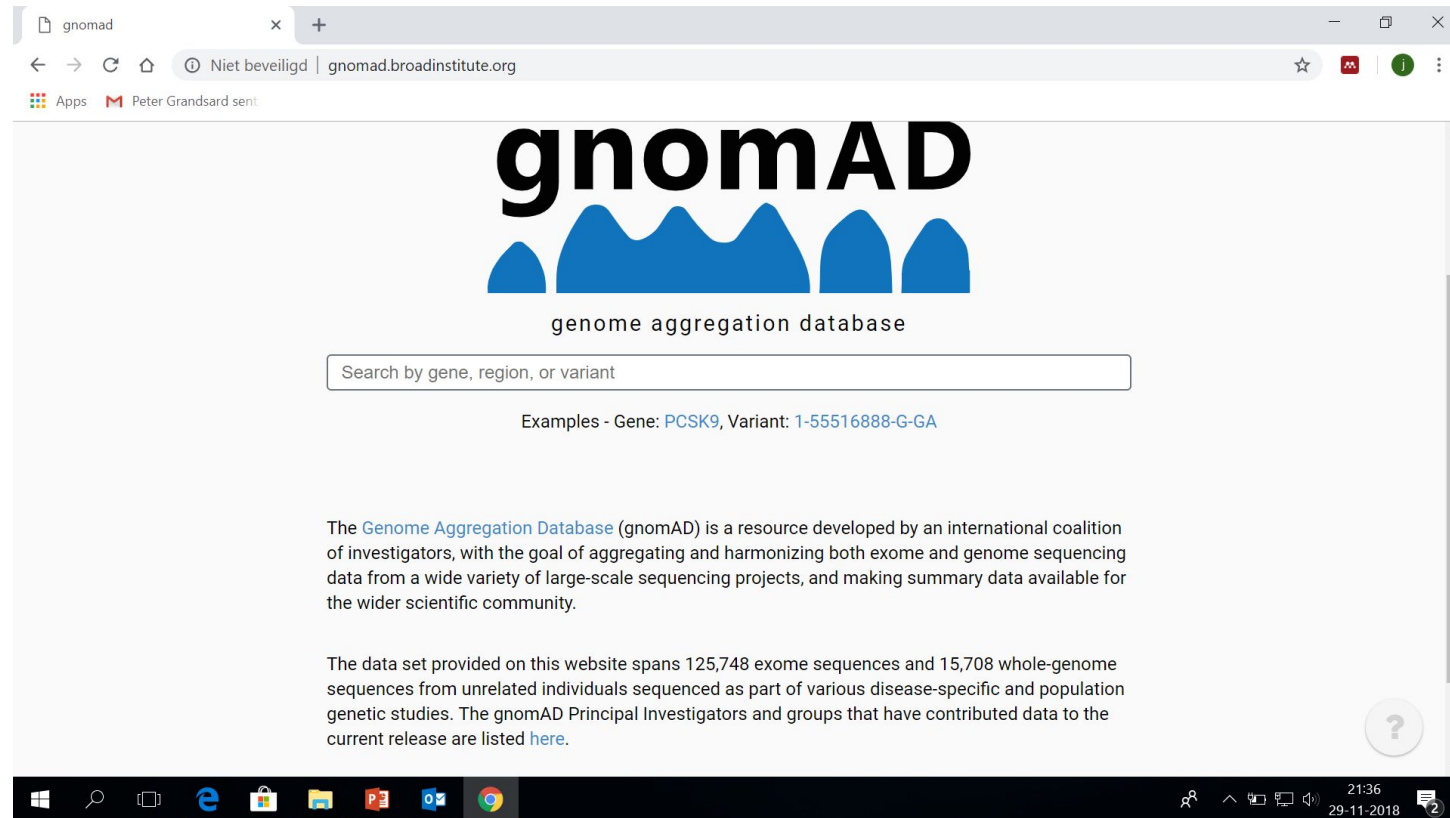


# WES/WGS analysis: Databases



# gnomAD (past Exac) database

Gene identification from genome wide population sequencing data based on 140000 exomes



The screenshot shows the gnomAD website homepage in a web browser. The browser's address bar displays "gnomad.broadinstitute.org". The page features the gnomAD logo, which consists of the text "gnomAD" in a large, bold, black font, with a blue silhouette of a mountain range below it. Underneath the logo is the text "genome aggregation database". A search bar is present with the placeholder text "Search by gene, region, or variant". Below the search bar, there are examples: "Examples - Gene: [PCSK9](#), Variant: [1-55516888-G-GA](#)". The main content area contains two paragraphs of text. The first paragraph states: "The [Genome Aggregation Database](#) (gnomAD) is a resource developed by an international coalition of investigators, with the goal of aggregating and harmonizing both exome and genome sequencing data from a wide variety of large-scale sequencing projects, and making summary data available for the wider scientific community." The second paragraph states: "The data set provided on this website spans 125,748 exome sequences and 15,708 whole-genome sequences from unrelated individuals sequenced as part of various disease-specific and population genetic studies. The gnomAD Principal Investigators and groups that have contributed data to the current release are listed [here](#)." A small circular icon with a question mark is visible in the bottom right corner of the page content. The Windows taskbar is visible at the bottom of the screenshot, showing the time as 21:36 on 29-11-2018.

# Probability of being LOF intolerant

- Haploinsufficiency to estimate the total number of autosomal recessive human protein-coding genes based on mutation tolerance
- Haploinsufficient genes do not tolerate loss-of-function (LOF) variants in one of the two alleles.
- Their probability of being LOF intolerant (pLI) is thus close to 1.
- In Gnomad pLI is measure by analysis of 140k exomes

# Database of genomic variants

(curated structural variation)

## Database of Genomic Variants

*A curated catalogue of human genomic structural variation*

[About the Project](#) [Downloads](#) [Links](#) [Statistics](#) [FAQ](#)  
[Genome Browser](#) [Query Tool](#) [Submissions](#) [Contact Us](#) [Training Resources](#)

Keyword, Landmark or Region Search:  Search GRCh37/hg19 ▾

**Examples:** RP11-34P13; CFTR, 7q11.21; chr7:71890181-72690180

### Find DGV Variants

[by Study](#) [by Sample](#)  
[by Method](#) [by Variant](#)  
[by Platform](#) [by Chromosome](#)

### Summary Statistics

Stat	Merged-level	Sample-level
CNVs:	983845	7021692
Inversions:	4083	32044

**Number of Studies:** 75

[News: February 2020 Update and Newsletter has been issued](#)

# Human Gene mutation database

The Human Gene Mutation Database (HGMD®) represents an attempt to collate all known (published) gene lesions responsible for human inherited disease and is maintained in Cardiff by D.N. Cooper, E.V. Ball, P.D. Stenson, A.D. Phillips, K. Evans, S. Heywood, M.J. Hayden, M.M. Chapman, M.E. Mort, L. Azevedo and M. Mort

**Get HGMD Professional** \*Please note that this less up-to-date public version of our database is freely available only to [registered](#) users from academic institutions/non-profit organisations. All commercial users are required to purchase a license from QIAGEN®, our commercial partner. A license to [HGMD Professional](#) is available to both commercial and academic/non-profit users wishing to access the most up-to-date version of the database (visit QIAGEN® to request a [free trial](#) of HGMD Professional). Read more about how HGMD is [funded](#). You may not copy, store or re-distribute HGMD data without express written permission (i) from the curators or (ii) via your license agreement. Copyright © Cardiff University 2017. All rights reserved.

[Register for Public Version](#)

Table:	Description:	Public entries: <small>This site. Academic/non-profit users only</small>	Total entries: <small>HGMD Professional 2018.3</small>
<b>Mutation totals (as of 2018-11-29)</b>			
		<b>157114</b>	<b>240269</b>
Gene symbol	The gene description, gene symbol (as recommended by the HUGO Nomenclature Committee) and chromosomal location is recorded for each gene. In cases where a gene symbol has not yet been made official, a provisional symbol has been adopted which is denoted by lower-case letters.	6531	9976
cDNA sequence	cDNA reference sequences are provided, numbered by codon.	6531	10339
Genomic coordinates	Genomic (chromosomal) coordinates have been calculated for missense/nonsense, splicing, regulatory, small deletions, small insertions and small indels.	0	214308
HGVS nomenclature	Standard HGVS nomenclature has been obtained for missense/nonsense, splicing, regulatory, small deletions, small insertions and small indels.	0	214691
Missense/nonsense	Single base-pair substitutions in coding regions are presented in terms of a triplet change with an additional flanking base included if the mutated base lies in either the first or third position in the triplet.	87397	137354
Splicing	Mutations with consequences for mRNA splicing are presented in brief with information specifying the relative position of the lesion with respect to a numbered intron donor or acceptor splice site. Positions given as positive integers refer to a 3' (downstream) location, negative integers refer to a 5' (upstream) location.	14317	21222
Regulatory	Substitutions causing regulatory abnormalities are logged in with thirty nucleotides flanking the site of the mutation on both sides. The location of the mutation relative to the transcriptional initiation site, initiation codon, polyadenylation site or termination codon is given.	3046	4189

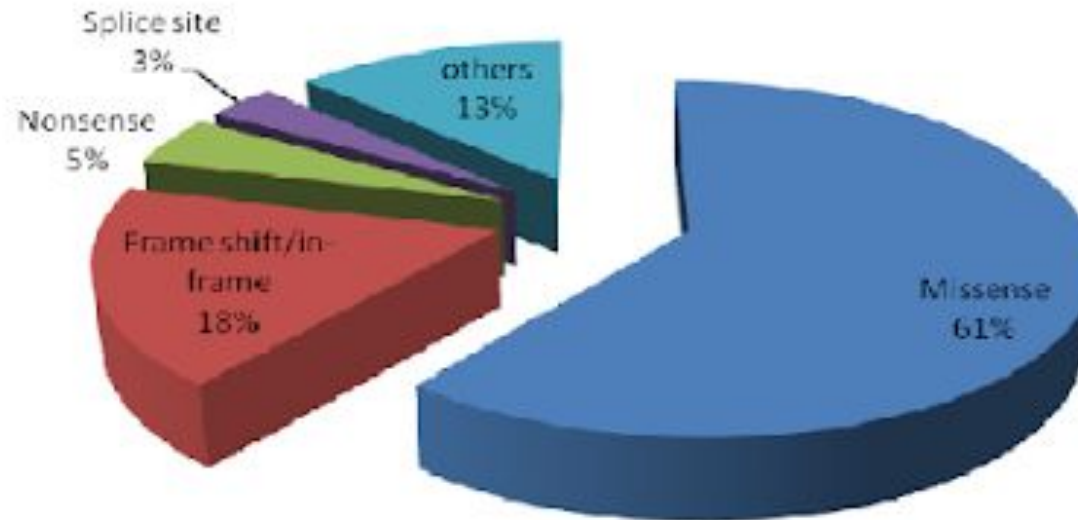
# Clinvar

The screenshot shows the ClinVar website in a browser window. The address bar displays <https://www.ncbi.nlm.nih.gov/clinvar/>. The page header includes the NCBI logo, navigation links for Resources and How To, and a Sign in to NCBI button. A search bar is prominently displayed with the text "Search ClinVar for gene symbols, HGVS expressions, conditions, and more" and a "Search" button. Below the search bar is a navigation menu with links for Home, About, Access, Help, Submit, Statistics, and FTP. The main content area features a large DNA sequence: ACTGATGGTATGGGGCCAAGAGATATATCT CAGGTACGGCTGTCATCACTTAGACCTCAC CAGGGCTGGGCATAAAAAGTCAGGGCAGAGC CCATGGTGCATCTGACTCCTGAGGAGAAGT GCAGGTTGGTATCAAGGTTACAAGACAGGT GGCCTGACTCTCTGCCTATTGGTCTAT. To the right of the sequence is a dark blue box with the ClinVar logo and the text "ClinVar aggregates information about genomic variation and its relationship to human health." Below this are three columns of links: "Using ClinVar" (About ClinVar, Data Dictionary, Downloads/FTP site, FAQ, Contact Us, RSS feed/What's new?, Factsheet), "Tools" (ACMG Recommendations for Reporting of Incidental Findings, ClinVar Submission Portal, Submissions, Variation Viewer, Clinical Remapping - Between assemblies and RefSeqGenes, RefSeqGene/LRG), and "Related Sites" (ClinGen, GeneReviews®, GTR®, MedGen, OMIM®, Variation). At the bottom, there is a "Submitter highlights" section. The Windows taskbar at the bottom shows the time as 21:40 on 29-11-2018.

## LQTS Gene LOVD Database



Tao Zhang<sup>1,2\*</sup>, Arthur Moss<sup>3,\*</sup>, Peikuan Cong<sup>2,\*</sup>, Min Pan<sup>2,\*</sup>, Bingxi Chang<sup>4</sup>, Liangrong Zheng<sup>5</sup>, Quan Fang<sup>4</sup>, Wojciech Zareba<sup>3</sup>, Jennifer Robinson<sup>3</sup>, Changsong Lin<sup>2</sup>, Zhongxiang Li<sup>6</sup>, Junfang Wei<sup>7</sup>, Qiang Zeng<sup>8</sup>, Long QT International Registry Investigators, HVP-China Investigators, and Ming Qi<sup>1,2,9\*\*</sup>





## **Standards and guidelines for the interpretation of sequence variants: a joint consensus recommendation of the American College of Medical Genetics and Genomics and the Association for Molecular Pathology**

Sue Richards, PhD<sup>1</sup>, Nazneen Aziz, PhD<sup>2,16</sup>, Sherri Bale, PhD<sup>3</sup>, David Bick, MD<sup>4</sup>, Soma Das, PhD<sup>5</sup>, Julie Gastier-Foster, PhD<sup>6,7,8</sup>, Wayne W. Grody, MD, PhD<sup>9,10,11</sup>, Madhuri Hegde, PhD<sup>12</sup>, Elaine Lyon, PhD<sup>13</sup>, Elaine Spector, PhD<sup>14</sup>, Karl Voelkerding, MD<sup>13</sup> and Heidi L. Rehm, PhD<sup>15</sup>; on behalf of the ACMG Laboratory Quality Assurance Committee

<b>Class of risk</b>	<b>Clinical significance</b>
<b>1</b>	<b>not pathogenic</b>
<b>2</b>	<b>likely not pathogenic</b>
<b>3</b>	<b>uncertain</b>
<b>4</b>	<b>likely pathogenic</b>
<b>5</b>	<b>definitely pathogenic</b>

Plon SE, Eccles DM, Easton D, et al. Sequence variant classification and reporting: recommendations for improving the interpretation of cancer susceptibility genetic test results. Hum Mutat. 2008; 29:1282–1291. [PubMed: 18951446]

**Table 3** Criteria for classifying pathogenic variants

Evidence of pathogenicity	Category
Very strong	<p>PVS1 null variant (nonsense, frameshift, canonical <math>\pm 1</math> or 2 splice sites, initiation codon, single or multiexon deletion) in a gene where LOF is a known mechanism of disease</p> <p>Caveats:</p> <ul style="list-style-type: none"> <li>• Beware of genes where LOF is not a known disease mechanism (e.g., <i>GFAP</i>, <i>MYH7</i>)</li> <li>• Use caution interpreting LOF variants at the extreme 3' end of a gene</li> <li>• Use caution with splice variants that are predicted to lead to exon skipping but leave the remainder of the protein intact</li> <li>• Use caution in the presence of multiple transcripts</li> </ul>
Strong	<p>PS1 Same amino acid change as a previously established pathogenic variant regardless of nucleotide change</p> <p>Example: Val→Leu caused by either G&gt;C or G&gt;T in the same codon</p> <p>Caveat: Beware of changes that impact splicing rather than at the amino acid/protein level</p> <p>PS2 De novo (<u>both</u> maternity and paternity confirmed) in a patient with the disease and no family history</p> <p>Note: Confirmation of paternity only is insufficient. Egg donation, surrogate motherhood, errors in embryo transfer, and so on, can contribute to nonmaternity.</p> <p>PS3 Well-established in vitro or in vivo functional studies supportive of a damaging effect on the gene or gene product</p> <p>Note: Functional studies that have been validated and shown to be reproducible and robust in a clinical diagnostic laboratory setting are considered the most well established.</p> <p>PS4 The prevalence of the variant in affected individuals is significantly increased compared with the prevalence in controls</p> <p>Note 1: Relative risk or OR, as obtained from case-control studies, is &gt;5.0, and the confidence interval around the estimate of relative risk or OR does not include 1.0. See the article for detailed guidance.</p> <p>Note 2: In instances of very rare variants where case-control studies may not reach statistical significance, the prior observation of the variant in multiple unrelated patients with the same phenotype, and its absence in controls, may be used as moderate level of evidence.</p>
Moderate	<p>PM1 Located in a mutational hot spot and/or critical and well-established functional domain (e.g., active site of an enzyme) without benign variation</p>

	Benign		Pathogenic			
	Strong	Supporting	Supporting	Moderate	Strong	Very Strong
<b>Population Data</b>	MAF is too high for disorder <i>BA1/BS1</i> OR observation in controls inconsistent with disease penetrance <i>BS2</i>			Absent in population databases <i>PM2</i>	Prevalence in affecteds statistically increased over controls <i>PS4</i>	
<b>Computational And Predictive Data</b>		Multiple lines of computational evidence suggest no impact on gene /gene product <i>BP4</i> Missense in gene where only truncating cause disease <i>BP1</i> Silent variant with non predicted splice impact <i>BP7</i>	Multiple lines of computational evidence support a deleterious effect on the gene /gene product <i>PP3</i>	Novel missense change at an amino acid residue where a different pathogenic missense change has been seen before <i>PM5</i> Protein length changing variant <i>PM4</i>	Same amino acid change as an established pathogenic variant <i>PS1</i>	Predicted null variant in a gene where LOF is a known mechanism of disease <i>PVS1</i>
<b>Functional Data</b>	Well-established functional studies show no deleterious effect <i>BS3</i>		Missense in gene with low rate of benign missense variants and path. missenses common <i>PP2</i>	Mutational hot spot or well-studied functional domain without benign variation <i>PM1</i>	Well-established functional studies show a deleterious effect <i>PS3</i>	
<b>Segregation Data</b>	Non-segregation with disease <i>BS4</i>		Co-segregation with disease in multiple affected family members <i>PP1</i>	Increased segregation data →		
<b>De novo Data</b>				<i>De novo</i> (without paternity & maternity confirmed) <i>PM6</i>	<i>De novo</i> (paternity & maternity confirmed) <i>PS2</i>	
<b>Allelic Data</b>		Observed in <i>trans</i> with a dominant variant <i>BP2</i> Observed in <i>cis</i> with a pathogenic variant <i>BP2</i>		For recessive disorders, detected in <i>trans</i> with a pathogenic variant <i>PM3</i>		
<b>Other Database</b>		Reputable source w/out shared data = benign <i>BP6</i>	Reputable source = pathogenic <i>PP5</i>			
<b>Other Data</b>		Found in case with an alternate cause <i>BP5</i>	Patient's phenotype or FH highly specific for gene <i>PP4</i>			

	Benign		Pathogenic			
	Strong	Supporting	Supporting	Moderate	Strong	Very Strong
<b>Population Data</b>	MAF is too high for disorder <i>BA1/BS1</i> OR observation in controls inconsistent with disease penetrance <i>BS2</i>			Absent in population databases <i>PM2</i>	Prevalence in affecteds statistically increased over controls <i>PS4</i>	
<b>Computational And Predictive Data</b>		Multiple lines of computational evidence suggest no impact on gene /gene product <i>BP4</i> Missense in gene where only truncating cause disease <i>BP1</i> Silent variant with non predicted splice impact <i>BP7</i>	Multiple lines of computational evidence support a deleterious effect on the gene /gene product <i>PP3</i>	Novel missense change at an amino acid residue where a different pathogenic missense change has been seen before <i>PM5</i> Protein length changing variant <i>PM4</i>	Same amino acid change as an established pathogenic variant <i>PS1</i>	Predicted null variant in a gene where LOF is a known mechanism of disease <i>PVS1</i>
<b>Functional Data</b>	Well-established functional studies show no deleterious effect <i>BS3</i>		Missense in gene with low rate of benign missense variants and path. missenses common <i>PP2</i>	Mutational hot spot or well-studied functional domain without benign variation <i>PM1</i>	Well-established functional studies show a deleterious effect <i>PS3</i>	
<b>Segregation Data</b>	Non-segregation with disease <i>BS4</i>		Co-segregation with disease in multiple affected family members <i>PP1</i>	Increased segregation data →		
<b>De novo Data</b>				<i>De novo</i> (without paternity & maternity confirmed) <i>PM6</i>	<i>De novo</i> (paternity & maternity confirmed) <i>PS2</i>	
<b>Allelic Data</b>		Observed in <i>trans</i> with a dominant variant <i>BP2</i> Observed in <i>cis</i> with a pathogenic variant <i>BP2</i>		For recessive disorders, detected in <i>trans</i> with a pathogenic variant <i>PM3</i>		
<b>Other Database</b>		Reputable source w/out shared data = benign <i>BP6</i>	Reputable source = pathogenic <i>PP5</i>			
<b>Other Data</b>		Found in case with an alternate cause <i>BP5</i>	Patient's phenotype or FH highly specific for gene <i>PP4</i>			

	Benign		Pathogenic			
	Strong	Supporting	Supporting	Moderate	Strong	Very Strong
<b>Population Data</b>	MAF is too high for disorder <i>BA1/BS1</i> OR observation in controls inconsistent with disease penetrance <i>BS2</i>			Absent in population databases <i>PM2</i>	Prevalence in affecteds statistically increased over controls <i>PS4</i>	
<b>Computational And Predictive Data</b>		Multiple lines of computational evidence suggest no impact on gene /gene product <i>BP4</i> Missense in gene where only truncating cause disease <i>BP1</i> Silent variant with non predicted splice impact <i>BP7</i>	Multiple lines of computational evidence support a deleterious effect on the gene /gene product <i>PP3</i>	Novel missense change at an amino acid residue where a different pathogenic missense change has been seen before <i>PM5</i> Protein length changing variant <i>PM4</i>	Same amino acid change as an established pathogenic variant <i>PS1</i>	Predicted null variant in a gene where LOF is a known mechanism of disease <i>PVS1</i>
<b>Functional Data</b>	Well-established functional studies show no deleterious effect <i>BS3</i>		Missense in gene with low rate of benign missense variants and path. missenses common <i>PP2</i>	Mutational hot spot or well-studied functional domain without benign variation <i>PM1</i>	Well-established functional studies show a deleterious effect <i>PS3</i>	
<b>Segregation Data</b>	Non-segregation with disease <i>BS4</i>		Co-segregation with disease in multiple affected family members <i>PP1</i>	Increased segregation data →		
<b>De novo Data</b>				<i>De novo</i> (without paternity & maternity confirmed) <i>PM6</i>	<i>De novo</i> (paternity & maternity confirmed) <i>PS2</i>	
<b>Allelic Data</b>		Observed in <i>trans</i> with a dominant variant <i>BP2</i> Observed in <i>cis</i> with a pathogenic variant <i>BP2</i>		For recessive disorders, detected in <i>trans</i> with a pathogenic variant <i>PM3</i>		
<b>Other Database</b>		Reputable source w/out shared data = benign <i>BP6</i>	Reputable source = pathogenic <i>PP5</i>			
<b>Other Data</b>		Found in case with an alternate cause <i>BP5</i>	Patient's phenotype or FH highly specific for gene <i>PP4</i>			

	Benign		Pathogenic			
	Strong	Supporting	Supporting	Moderate	Strong	Very Strong
<b>Population Data</b>	MAF is too high for disorder <i>BA1/BS1</i> OR observation in controls inconsistent with disease penetrance <i>BS2</i>			Absent in population databases <i>PM2</i>	Prevalence in affecteds statistically increased over controls <i>PS4</i>	
<b>Computational And Predictive Data</b>		Multiple lines of computational evidence suggest no impact on gene /gene product <i>BP4</i> Missense in gene where only truncating cause disease <i>BP1</i> Silent variant with non predicted splice impact <i>BP7</i>	Multiple lines of computational evidence support a deleterious effect on the gene /gene product <i>PP3</i>	Novel missense change at an amino acid residue where a different pathogenic missense change has been seen before <i>PM5</i> Protein length changing variant <i>PM4</i>	Same amino acid change as an established pathogenic variant <i>PS1</i>	Predicted null variant in a gene where LOF is a known mechanism of disease <i>PVS1</i>
<b>Functional Data</b>	Well-established functional studies show no deleterious effect <i>BS3</i>		Missense in gene with low rate of benign missense variants and path. missenses common <i>PP2</i>	Mutational hot spot or well-studied functional domain without benign variation <i>PM1</i>	Well-established functional studies show a deleterious effect <i>PS3</i>	
<b>Segregation Data</b>	Non-segregation with disease <i>BS4</i>		Co-segregation with disease in multiple affected family members <i>PP1</i>	Increased segregation data →		
<b>De novo Data</b>				<i>De novo</i> (without paternity & maternity confirmed) <i>PM6</i>	<i>De novo</i> (paternity & maternity confirmed) <i>PS2</i>	
<b>Allelic Data</b>		Observed in <i>trans</i> with a dominant variant <i>BP2</i> Observed in <i>cis</i> with a pathogenic variant <i>BP2</i>		For recessive disorders, detected in <i>trans</i> with a pathogenic variant <i>PM3</i>		
<b>Other Database</b>		Reputable source w/out shared data = benign <i>BP6</i>	Reputable source = pathogenic <i>PP5</i>			
<b>Other Data</b>		Found in case with an alternate cause <i>BP5</i>	Patient's phenotype or FH highly specific for gene <i>PP4</i>			

	Benign		Pathogenic			
	Strong	Supporting	Supporting	Moderate	Strong	Very Strong
<b>Population Data</b>	MAF is too high for disorder <i>BA1/BS1</i> OR observation in controls inconsistent with disease penetrance <i>BS2</i>			Absent in population databases <i>PM2</i>	Prevalence in affecteds statistically increased over controls <i>PS4</i>	
<b>Computational And Predictive Data</b>		Multiple lines of computational evidence suggest no impact on gene /gene product <i>BP4</i> Missense in gene where only truncating cause disease <i>BP1</i> Silent variant with non predicted splice impact <i>BP7</i>	Multiple lines of computational evidence support a deleterious effect on the gene /gene product <i>PP3</i>	Novel missense change at an amino acid residue where a different pathogenic missense change has been seen before <i>PM5</i> Protein length changing variant <i>PM4</i>	Same amino acid change as an established pathogenic variant <i>PS1</i>	Predicted null variant in a gene where LOF is a known mechanism of disease <i>PVS1</i>
<b>Functional Data</b>	Well-established functional studies show no deleterious effect <i>BS3</i>		Missense in gene with low rate of benign missense variants and path. missenses common <i>PP2</i>	Mutational hot spot or well-studied functional domain without benign variation <i>PM1</i>	Well-established functional studies show a deleterious effect <i>PS3</i>	
<b>Segregation Data</b>	Non-segregation with disease <i>BS4</i>		Co-segregation with disease in multiple affected family members <i>PP1</i>	Increased segregation data →		
<b>De novo Data</b>				<i>De novo</i> (without paternity & maternity confirmed) <i>PM6</i>	<i>De novo</i> (paternity & maternity confirmed) <i>PS2</i>	
<b>Allelic Data</b>		Observed in <i>trans</i> with a dominant variant <i>BP2</i> Observed in <i>cis</i> with a pathogenic variant <i>BP2</i>		For recessive disorders, detected in <i>trans</i> with a pathogenic variant <i>PM3</i>		
<b>Other Database</b>		Reputable source w/out shared data = benign <i>BP6</i>	Reputable source = pathogenic <i>PP5</i>			
<b>Other Data</b>		Found in case with an alternate cause <i>BP5</i>	Patient's phenotype or FH highly specific for gene <i>PP4</i>			



**Table 5** Rules for combining criteria to classify sequence variants

Pathogenic	<ul style="list-style-type: none"> <li>(i) 1 Very strong (PVS1) <i>AND</i></li> <li style="padding-left: 20px;">(a) <math>\geq 1</math> Strong (PS1–PS4) <i>OR</i></li> <li style="padding-left: 20px;">(b) <math>\geq 2</math> Moderate (PM1–PM6) <i>OR</i></li> <li style="padding-left: 20px;">(c) 1 Moderate (PM1–PM6) and 1 supporting (PP1–PP5) <i>OR</i></li> <li style="padding-left: 20px;">(d) <math>\geq 2</math> Supporting (PP1–PP5)</li> <li>(ii) <math>\geq 2</math> Strong (PS1–PS4) <i>OR</i></li> <li>(iii) 1 Strong (PS1–PS4) <i>AND</i></li> <li style="padding-left: 20px;">(a) <math>\geq 3</math> Moderate (PM1–PM6) <i>OR</i></li> <li style="padding-left: 20px;">(b) 2 Moderate (PM1–PM6) <i>AND</i> <math>\geq 2</math> Supporting (PP1–PP5) <i>OR</i></li> <li style="padding-left: 20px;">(c) 1 Moderate (PM1–PM6) <i>AND</i> <math>\geq 4</math> supporting (PP1–PP5)</li> </ul>
Likely pathogenic	<ul style="list-style-type: none"> <li>(i) 1 Very strong (PVS1) <i>AND</i> 1 moderate (PM1–PM6) <i>OR</i></li> <li>(ii) 1 Strong (PS1–PS4) <i>AND</i> 1–2 moderate (PM1–PM6) <i>OR</i></li> <li>(iii) 1 Strong (PS1–PS4) <i>AND</i> <math>\geq 2</math> supporting (PP1–PP5) <i>OR</i></li> <li>(iv) <math>\geq 3</math> Moderate (PM1–PM6) <i>OR</i></li> <li>(v) 2 Moderate (PM1–PM6) <i>AND</i> <math>\geq 2</math> supporting (PP1–PP5) <i>OR</i></li> <li>(vi) 1 Moderate (PM1–PM6) <i>AND</i> <math>\geq 4</math> supporting (PP1–PP5)</li> </ul>
Benign	<ul style="list-style-type: none"> <li>(i) 1 Stand-alone (BA1) <i>OR</i></li> <li>(ii) <math>\geq 2</math> Strong (BS1–BS4)</li> </ul>
Likely benign	<ul style="list-style-type: none"> <li>(i) 1 Strong (BS1–BS4) and 1 supporting (BP1–BP7) <i>OR</i></li> <li>(ii) <math>\geq 2</math> Supporting (BP1–BP7)</li> </ul>
Uncertain significance	<ul style="list-style-type: none"> <li>(i) Other criteria shown above are not met <i>OR</i></li> <li>(ii) the criteria for benign and pathogenic are contradictory</li> </ul>

**ClinGen**  
Clinical Genome Resource


**PATHOGENICITY CALCULATOR**

Users of the calculator can contribute their interpretation, evidence codes, evidence, and assertion in the **Pathogenicity Calculator Evidence Repo (PCER)** by clicking 'Export to PCER'. The shared data is instantly available through [ClinGen Allele Registry](#) and [PCER](#).

ClinGen Pathogenicity Calculator team is thankful to our [distinguished users](#) who donated their interpretations in ClinVar.

LOG IN

## WHAT IS THE CLINGEN PATHOGENICITY CALCULATOR?



The shift from genetic testing of individual genes to exome and genome sequencing has been accompanied by new challenges in genome interpretation. The American College of Medical Genetics and Genomics and the Association for Molecular Pathology (ACMG/AMP) have published [Standards and Guidelines for the Interpretation of Sequence Variants](#). To enable wide application of the ACMG/AMP and similar guidelines and the development of collective knowledge by the community, ClinGen has developed the ClinGen Pathogenicity Calculator. By automating the formal reasoning, the Calculator eliminates errors in rule application and makes it possible to automatically calculate provisional conclusions based on latest evidence. Moreover, the Calculator makes reasoning explicit by documenting applicable rules, evidence codes, and links to supporting data. By explicitly communicating the reasoning behind a conclusion about pathogenicity of any specific variant, the Calculator enables critical evaluation of the reasoning and facilitates resolution of conflicting conclusions.

Report generated dynamically by BCM's ClinGen Pathogenicity Calculator. Powered by [Genobase](#).

### Allele Information

**Allele Registry ID**  
<http://reg.genome.network/allele/CA021883>

**HGVS**  
NC\_000023.11:g.101399747C>T, CM000685.2:g.101399747C>T, NC\_000023.10:g.100654735C>T, CM000685.1:g.100654735C>T, NC\_000023.9:g.100541391C>T, NG\_007119.1:g.13217G>A, LRG\_672:g.13217G>A, NM\_001169.2:c.640-801G>A, LRG\_672t1:c.640-801G>A, NM\_001199973.1:c.408+4290C>T, NM\_001199974.1:c.285+7925C>T, XR\_938397.1:n.721G>A, ENST00000218516.3:c.640-801G>A, ENST00000409170.3:c.300+4290C>T, ENST00000409338.5:c.177+7925C>T, ENST00000468823.1:n.189-801G>A, ENST00000480513.5:n.478-801G>A, ENST00000486121.5:n.685-801G>A, ENST00000493905.6:c.\*24G>A

**Gene**  
GLA

**Phenotype**  
Fabry disease

**Mode of Inheritance**  
X-linked Recessive

### Evidence

**PP1**  
**Category** : Pathogenic » Supporting » Segregation Data  
**ACMG Text** : cosegregation with disease in multiple affected family members in a gene definitively known to cause the disease  
**User Summary** : Allele cosegregated with disease  
**Supporting Links** :  
◦ Taiwan population [\[Link\]](#)

**PS4**  
**Category** : Pathogenic » Strong » Population Data  
**ACMG Text** : Prevalence in affecteds statistically increased over controls  
**User Summary** : Higher prevalence over control  
**Supporting Links** :  
◦ Paper reporting unexpected high prevalence of the cardiac variant IVS4+919G>A among both newborns and patients with idiopathic hypertrophic cardiomyopathy in the Taiwan Chinese population [\[Link\]](#)

**PS3**  
**Category** : Pathogenic » Strong » Functional Data  
**ACMG Text** : Well-established functional studies show a deleterious effect  
**User Summary** : Functional studies support this tag.  
**Supporting Links** :  
◦ Plasma  $\beta$ -galactosidase A activity assay was 10.47±11.2% of normal in the men and 48.67±19.5% of normal in the women [\[Link\]](#)

**PVS1-Strong**  
**Category** : Pathogenic » Strong » Computational And Predictive Data  
**ACMG Text** : PVS1 downgraded in strength to Strong  
**User Summary** : Null variant but incomplete alternate splicing

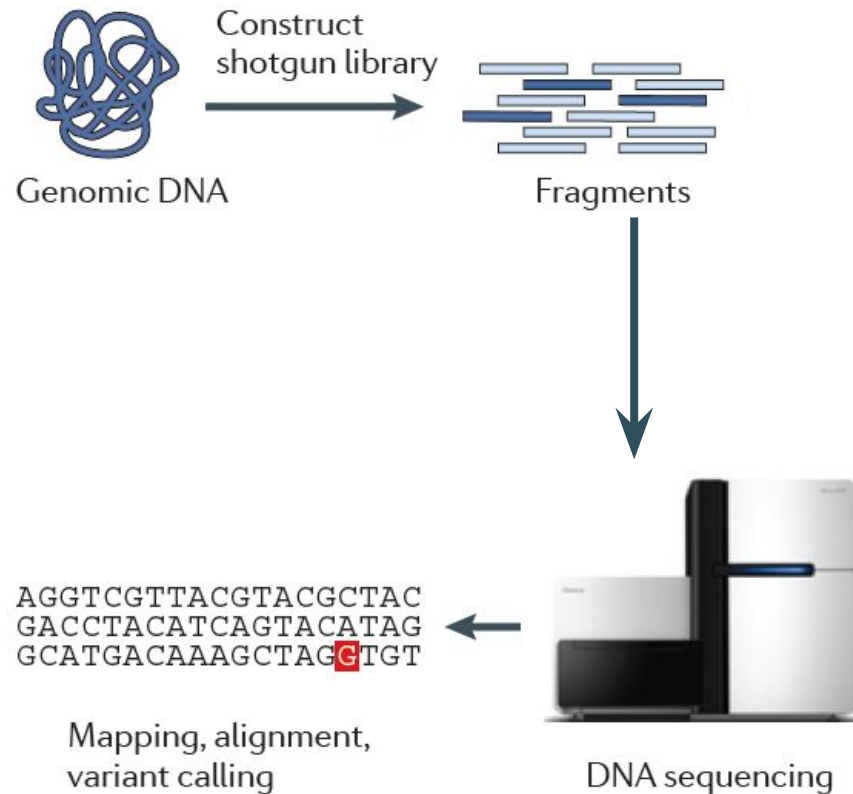
### Assertions and Reasoning

**Final Call** : Pathogenic

**Rules Passed** :  
◦ Pathogenic.Strong >=2

Fig. 3 A sample summary report generated by Pathogenicity Calculator. The report itself is printable as PDF and downloadable by the user

# Whole-Genome SEQUENCING



## MCA/ID: trio-based WGS analysis of 3 patients with unknown clinical diagnosis

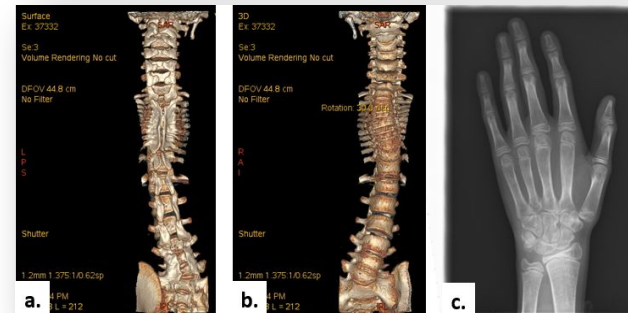
### Anamnesis:

- Only child of an healthy unrelated couple.
- At birth facial dysmorphism with polymalformative syndrome
- Large anterior fontanella
- Microcephaly
- Right-turning
- Abdominal wa
- Cleft palate
- Hypoplasia of
- Cryptorchidia
- Major hearing
- Hypertension
- Feeding problems
- Short stature
- Congenital thoracic vertebral fusion  severe torsional scoliosis

#### Molecular analysis:

- Karyotyping
- FISH for 22q11.2 and 9p-
- Array-CGH
  - 8q12.1(56899737-57048789)x3mat
  - 16p13.3(4379999-4443009)x3mat
  - likely benign

### Patient 1



## MCA/ID: trio-based WGS analysis of 3 patients with unknown clinical diagnosis

### Patient 2

- Hypotonia
- Failure to thrive
- Progressive macrocephaly (H&W at p3, OFC at p97)
- Periventricular leukomalacia on imaging

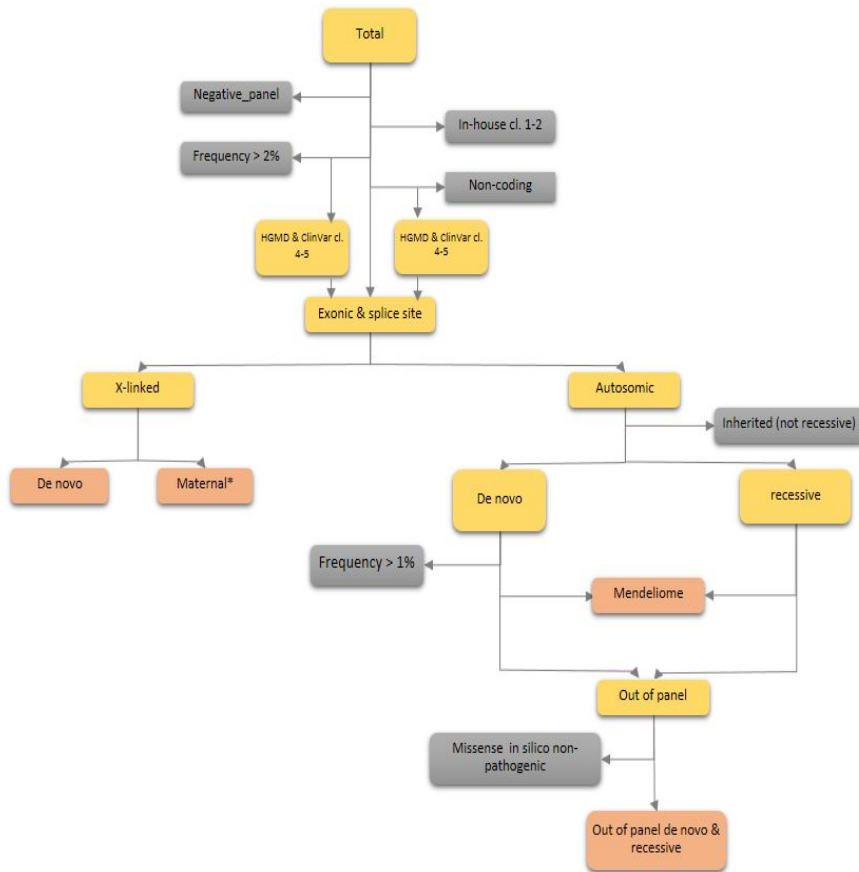
- Epilepsy
- Frontal bossing
- Deep-set eyes
- Downslanting palpebral fissures
- Mild hypotonia
- Mild intellectual disability
- Clear picture of a child with a developmental delay
- Very small and fragile teeth
- Mildly delayed speech
- Small and fragile teeth

#### Molecular analysis:

- Array CGH
- Fragile X
- PTEN, MID1 and NEMO genes
- gene panel for Rasopathies (PTPN11, SOS1, RAF1, RIT1, KRAS, BRAF, MEK1, MEK2 and HRAS)
- Mendeliome in 2015



## MCA/ID: trio-based WGS analysis of 3 patients with unknown clinical diagnosis



### SNVs Analysis:

Filtering criteria	Number of variants	
	Patient 1	Patient 2
Total	108071	105933
Not in negative panel	104375	102349
Not in-house class 1-2	102901	100890
Frequency < 2%	8337	8083
Exonic and splice site variants	1450	1351
x-linked + recessive + de novo	95	75
AD filtering	60	58



↓  
**Negative**

# MCA/ID: trio-based WGS analysis of 3 patients with unknown clinical diagnosis



## Analysis:



	Number of variants	
Filtering criteria	Patient 1	Patient 2
<i>de novo</i> Total	108071	105933
Not in negative controls <b>c.430dupG</b>	104375	102349
Not in-house class 1-2 <b>p.Ala144Glyfs*52</b>	102901	100890
Frequency < 2%	8337	8083
Exonic and splice sites <b>HIST1H1E</b>	1450	1351
x-linked + recessive + de novo	95	75
		58

**HISTONE GENE CLUSTER 1, H1 HISTONE FAMILY, MEMBER E; HIST1H1E**

OMIM®

HGNC Approved Gene Symbol: [H1-4](#)  
 Cytogenetic location: [6p22.2](#)  
 Genomic coordinates (GRCh38): [6:26,156,330-26,157,114](#) (from NCBI)

Location	Phenotype	Phenotype MIM number	Inheritance
<a href="#">6p22.2</a>	<b>Rahman syndrome</b>	<a href="#">617537</a>	AD

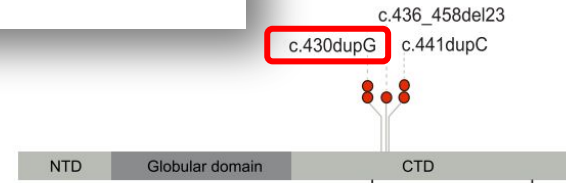
HGMD accession	Reported disease/phenotype	Variant class	Gene symbol	...
CI176502	Intellectual disability with overgrowth		<a href="#">HIST1H1E</a>	GCGACG*GG
Literature citation		Citation type	Support	
1. Tatton-Brown (2017) <i>Am J Hum Genet</i> <b>100</b> : 725. PubMed: <a href="#">28475857</a> Mutations in Epigenetic Regulation Genes Are a Major Cause of Overgrowth with Intellectual Disability.		Primary literature report		See Table S1.
Extra information				
Coding strand genomic sequence (GRCh38)		CGGCGAAGAAGCCCAAGAAGGCGACGGGGG(-g)CGGCC		
<a href="#">Genomic coordinate</a> (GRCh38)		chr6:26156820-26156821		
Genome viewers		<a href="#">UCSC</a> ; <a href="#">NCBI MapViewer</a> ; <a href="#">NCBI SeqViewer</a>		
HGVS nomenclature		<a href="#">NM_005321.2</a> : c.430dupG; <a href="#">NP_005312.1</a> : p.(Ala144Glyfs*52)		

# MCA/ID: trio-based WGS analysis of 3 patients with unknown clinical diagnosis

## ARTICLE

### Mutations in Epigenetic Regulation Genes Are a Major Cause of Overgrowth with Intellectual Disability

Katrina Tatton-Brown,<sup>1,2</sup> Chey Loveday,<sup>1</sup> Shawn Yost,<sup>1</sup> Matthew Clarke,<sup>1</sup> Emma Ramsay,<sup>1</sup> Anna Zachariou,<sup>1</sup> Anna Elliott,<sup>1</sup> Harriet Wylie,<sup>1</sup> Anna Ardissonne,<sup>3</sup> Olaf Rittinger,<sup>4</sup> Fiona Stewart,<sup>5</sup> I. Karen Temple,<sup>6,7</sup> Trevor Cole,<sup>8</sup> Childhood Overgrowth Collaboration, Shazia Mahamdallie,<sup>1</sup> Sheila Seal,<sup>1</sup> Elise Ruark,<sup>1</sup> and Nazneen Rahman<sup>1,9,10,\*</sup>



c.430dupG p.Ala144Glyfs\*52



Frame	Mutation	Sequence	Length	Charge
1	Wildtype	...KKATGAATPKKSAKKTPKKAKKPAAGAKKAKSPKKAKAAKPKKAPKSPAKAKAVKPK...	219	44
2	c.430dupG	...KKATGGGHPQEERQEDPKEGEEAGCSCWSQKSEKPEKGESSQAKKGAQEPSEGQSS	194	7
2	c.441dupC	...KKATGAATPQEERQEDPKEGEEAGCSCWSQKSEKPEKGESSQAKKGAQEPSEGQSS	194	7
2	c.436_458del23	...KKATGAA-----DPKRGEEAGCSCWSQKSEKPEKGESSQAKKGAQEPSEGQSS	186	9
3		...KKATGRPPPRRPRRQRRRRRLQLLEPKKRRKARRKRRKQPSQKRRPRAQRRPKQLNPR...	227	48



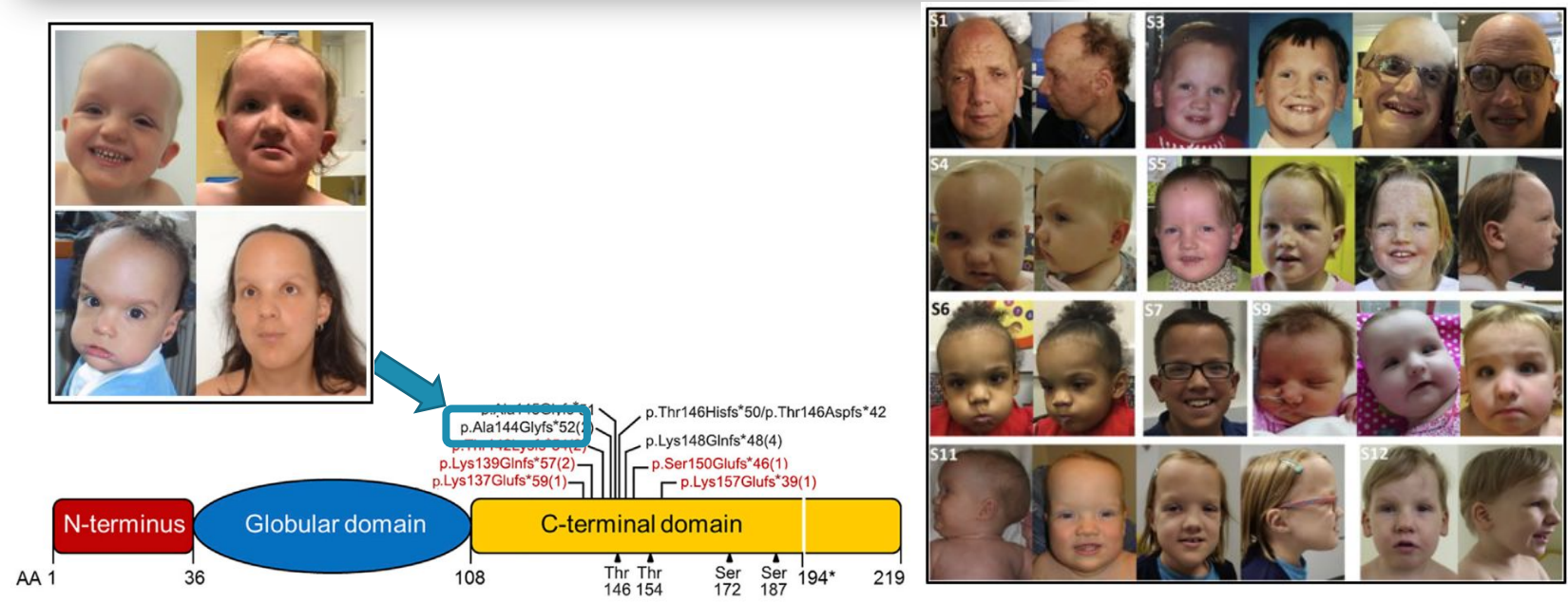


# MCA/ID: trio-based WGS analysis of 3 patients with unknown clinical diagnosis

Am J Hum Genet. 2019 Sep 5;105(3):493-508. doi: 10.1016/j.ajhg.2019.07.007. Epub 2019 Aug 22.

## Aberrant Function of the C-Terminal Tail of HIST1H1E Accelerates Cellular Senescence and Causes Premature Aging.

Flex E<sup>1</sup>, Martinelli S<sup>2</sup>, Van Dijk A<sup>3</sup>, Ciolfi A<sup>4</sup>, Cecchetti S<sup>5</sup>, Coluzzi E<sup>6</sup>, Pannone L<sup>7</sup>, Andreoli C<sup>8</sup>, Radio FC<sup>4</sup>, Pizzi S<sup>4</sup>, Carpentieri G<sup>7</sup>, Bruselles A<sup>2</sup>, Catanzaro G<sup>9</sup>, Pedace L<sup>10</sup>, Miele E<sup>10</sup>, Carcarino E<sup>11</sup>, Ge X<sup>12</sup>, Chijiwa C<sup>13</sup>, Lewis MES<sup>13</sup>, Meuwissen M<sup>14</sup>, Kenis S<sup>15</sup>, Van der Aa N<sup>14</sup>, Larson A<sup>16</sup>, Brown K<sup>16</sup>, Wasserstein MP<sup>17</sup>, Skotko BG<sup>18</sup>, Begtrup A<sup>19</sup>, Person R<sup>19</sup>, Karayiorgou M<sup>20</sup>, Roos JL<sup>21</sup>, Van Gassen KL<sup>22</sup>, Koopmans M<sup>22</sup>, Bijlsma EK<sup>23</sup>, Santen GWE<sup>23</sup>, Barge-Schaapveld DQCM<sup>23</sup>, Ruivenkamp CAL<sup>23</sup>, Hoffer MJV<sup>23</sup>, Lalani SR<sup>24</sup>, Streff H<sup>24</sup>, Craigen WJ<sup>24</sup>, Graham BH<sup>25</sup>, van den Elzen APM<sup>26</sup>, Kamphuis D<sup>27</sup>, Ünnap K<sup>28</sup>, Reinson K<sup>28</sup>, Pajusalu S<sup>29</sup>, Wojcik MH<sup>30</sup>, Viberti C<sup>31</sup>, Di Gaetano C<sup>31</sup>, Bertini E<sup>4</sup>, Petrucci S<sup>32</sup>, De Luca A<sup>33</sup>, Rota R<sup>10</sup>, Ferretti E<sup>34</sup>, Matullo G<sup>31</sup>, Dallapiccola B<sup>4</sup>, Sgura A<sup>6</sup>, Walkiewicz M<sup>35</sup>, Kooy RF<sup>36</sup>, Tartaglia M<sup>37</sup>.



seqnames	start	end	width	change_ty	QUAL_lun	GC069450	GC069451	GC069452	Combined_geno	ncbiRefSeq	geneHanc	Num_exo	DGV	freq	GNOMAD	lumpy_en	lumpy_ch	OMIM_ye	OMIM_gr
chr1	25405501	25410000	4500	DEL	NOT_LUM	1/1	0/1	0/1	1/1;0/1;0/1	RHCE	.	11	DGV	NA	NA				RHCE
chr5	70508001	70512000	4000	DEL	NOT_LUM	1/1	0/1	0/1	1/1;0/1;0/1	SMA5	.	1	NA	NA	NA				
chr7	38348501	38358000	9500	DEL	NOT_LUM	1/1	0/1	0/1	1/1;0/1;0/1	TRG-AS1	TRGV4	0	NA	NA	NA				
chr5	1178001	1181000	3000	DEL	LowQualit	1/1	0/1	0/1	1/1;0/1;0/1	CTD-3080f	CTD-3080f	0	NA	NA	NA			1180617	chr5

MCA/ID: trio-based WGS analysis of 3 patients with unknown clinical diagnosis

CNVs Analysis: Negative

SVs Analysis



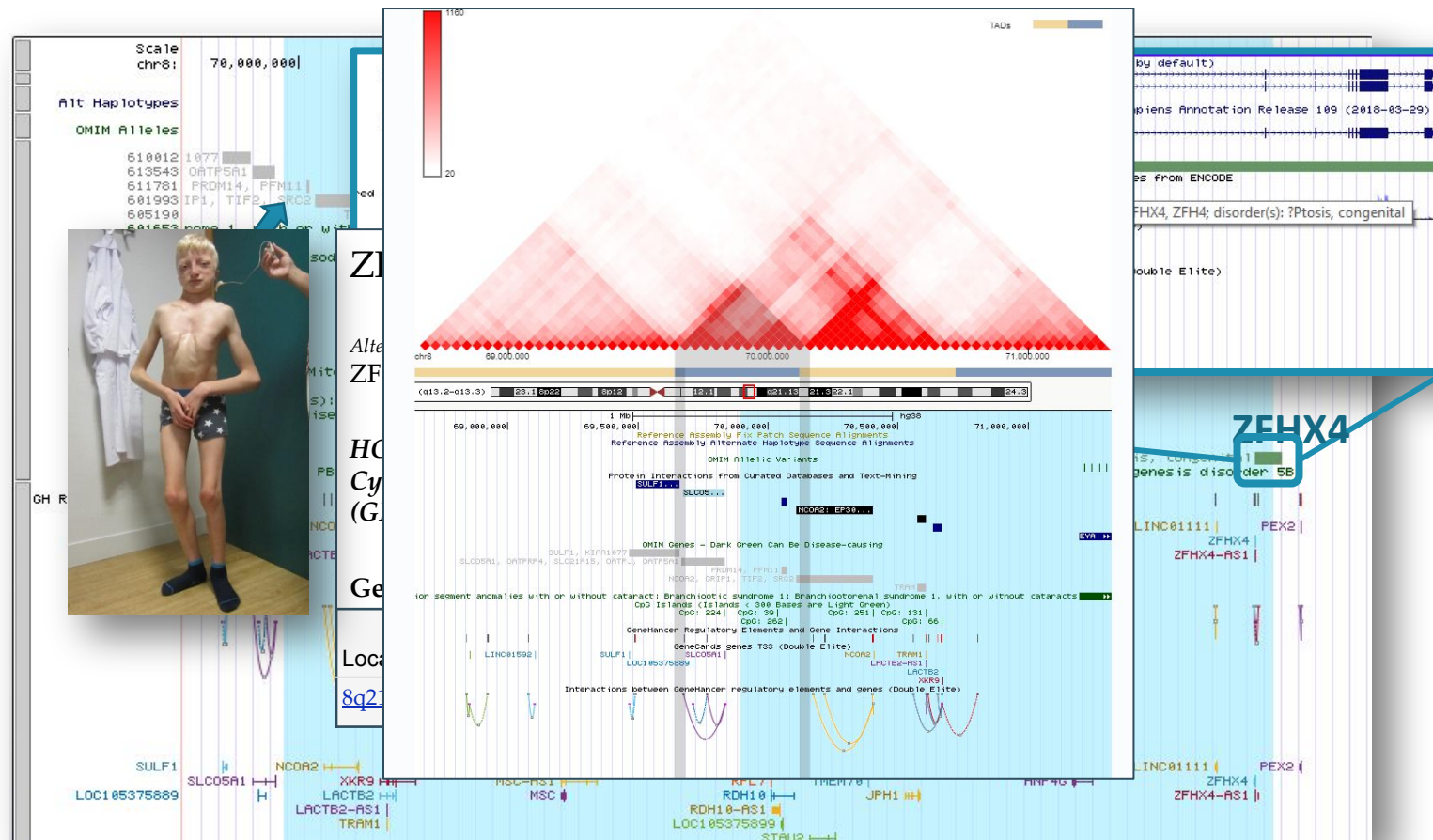
		Number of variants		
Total	20414			
<b>Filtering criteria</b> Good quality and rare > 10 bp De novo and recessive containing gene and/or regulatory regions IGV manual filtering	<b>DEL/DUP/INV</b>	<b>INS</b>	<b>BND</b>	
	737	3795	2827	
	*	165	*	
	44	20	40	
IGV manual filtering	3	4	0	

seqnames	start	end	width	change_ty	QUAL_lumpy	50_conser	51_conser	52_conser	Combined_genotypes	ncbiRefSeq	Curat	geneHanc	num_exon	DGV	freq	GNOMAD	lumpy_end	lumpy_chro	MIM_yello	MIM_gree
chr6	32637001	32643000	6000	DEL	NOT_LUM	0/1	0/0	0/0	0/1;0/0;0/0	HLA-DQA1	.	HLA-DQA1	7	NA	NA	NA				HLA-DQA1
chr6	32643001	32649000	6000	DEL	NOT_LUM	0/1	0/0	0/0	0/1;0/0;0/0	HLA-DRB5	.	HLA-DRB5	10	NA	NA	NA				HLA-DRB5
chr8	10111001	10117000	6000	DEL	NOT_LUM	0/1	0/0	0/0	0/1;0/0;0/0	OR4F21	.	OR4F21	7	NA	NA	NA				
chr15	10111001	10117000	6000	DEL	NOT_LUM	0/1	0/0	0/0	0/1;0/0;0/0	OR4F4	.	OR4F4	1	NA	NA	NA				
chr14	71111001	71117000	6000	DEL	NOT_LUM	0/1	0/0	0/0	0/1;0/0;0/0	ACOT1,HEATR4	.	ACOT1,HEATR4	12	NA	NA	NA				ACOT1
chr8	61111001	61117000	6000	DEL	NOT_LUM	0/1	0/0	0/0	0/1;0/0;0/0	FA1	.	FA1	1	NA	NA	NA				FA1
chr19	51111001	51117000	6000	DEL	NOT_LUM	0/1	0/0	0/0	0/1;0/0;0/0	2DL3	.	2DL3	1	NA	NA	NA				2DL3
chr22	21111001	21117000	6000	DEL	NOT_LUM	0/1	0/0	0/0	0/1;0/0;0/0	GGT2	.	GGT2	1	NA	NA	NA				GGT2
chr19	54829001	54845000	16000	DEL	NOT_LUM	0/1	0/0	0/0	0/1;0/0;0/0	2DS4	.	2DS4	1	NA	NA	NA				2DS4
chr17	36175001	36181000	6000	DEL	NOT_LUM	0/1	0/0	0/0	0/1;0/0;0/0	KIR3DL1	.	KIR3DL1	1	NA	NA	NA				KIR3DL1
chr8	72960001	72966000	6000	DEL	NOT_LUM	0/1	0/0	0/0	0/1;0/0;0/0	TBC1L1	.	TBC1L1	1	NA	NA	NA				TBC1L1
chr11	19450001	19456000	6000	DEL	NOT_LUM	0/1	0/0	0/0	0/1;0/0;0/0	PL23	.	PL23	1	NA	NA	NA				PL23
chr3	130082001	130088000	6000	DEL	NOT_LUM	0/1	0/0	0/0	0/1;0/0;0/0	1D3F	.	1D3F	1	NA	NA	NA				1D3F
chr17	36390001	36396000	6000	DEL	NOT_LUM	0/1	0/0	0/0	0/1;0/0;0/0	DRB5	.	DRB5	1	NA	NA	NA				DRB5
chr7	77009001	77015000	6000	DEL	NOT_LUM	0/1	0/0	0/0	0/1;0/0;0/0	GOLC	.	GOLC	1	NA	NA	NA				GOLC
chr6	32524001	32530000	6000	DEL	NOT_LUM	0/1	0/0	0/0	0/1;0/0;0/0	TED	.	TED	1	NA	NA	NA				TED
chr15	34387001	34393000	6000	DEL	NOT_LUM	0/1	0/0	0/0	0/1;0/0;0/0	SRGA	.	SRGA	1	NA	NA	NA				SRGA
chr21	13598001	13604000	6000	DEL	NOT_LUM	0/1	0/0	0/0	0/1;0/0;0/0		.		1	NA	NA	NA				
chr1	143541001	143547000	6000	DEL	NOT_LUM	0/1	0/0	0/0	0/1;0/0;0/0		.		1	NA	NA	NA				
chr2	87709001	87715000	6000	DEL	NOT_LUM	0/1	0/0	0/0	0/1;0/0;0/0		.		1	NA	NA	NA				
chr3	46754001	46760000	6000	DEL	NOT_LUM	0/1	0/0	0/0	0/1;0/0;0/0		.		1	NA	NA	NA				
chr11	19645001	19785000	14000	DEL	NOT_LUM	0/1	0/0	0/0	0/1;0/0;0/0		.		1	NA	NA	NA				
chr15	84362001	84383500	21500	DEL	NOT_LUM	0/1	0/0	0/0	0/1;0/0;0/0		.		1	NA	NA	NA				
chr19	27399501	27633500	234000	DEL	NOT_LUM	0/1	0/0	0/0	0/1;0/0;0/0		.		1	NA	NA	NA				
chr22	21459501	21511500	52000	DEL	NOT_LUM	0/1	0/0	0/0	0/1;0/0;0/0	PI4KAP2,TMEM19	.	PI4KAP2,TMEM19	10	NA	NA	NA				
chr7	74885501	74894000	8500	DEL	NOT_LUM	0/1	0/0	0/0	0/1;0/0;0/0	PMS2P5,STAG3L2	.	PMS2P5,STAG3L2	2	NA	NA	NA				
chr22	24251001	24252500	1500	DEL	NOT_LUM	0/1	0/0	0/0	0/1;0/0;0/0	POM121L9P	POU121L5	0	NA	NA	NA					
chr7	38348501	38358000	9500	DEL	NOT_LUM	1/1	0/1	0/1	1/1;0/1;0/1	TRG-AS1	TRGV4	0	NA	NA	NA					
chr5	70508001	70512000	4000	DEL	NOT_LUM	1/1	0/1	0/1	1/1;0/1;0/1	SMA5	.	SMA5	1	NA	NA	NA				

Inversion  
chr8:69893659-76806725  
6.9 Mb



# MCA/ID: trio-based WGS analysis of 3 patients with unknown clinical diagnosis



## MCA/ID: trio-based WGS analysis of 3 patients with unknown clinical diagnosis

Am J Hum Genet. 2011 Aug 12;89(2):295-301. doi: 10.1016/j.ajhg.2011.06.012. Epub 2011 Jul 28.

### Characterization of a 8q21.11 microdeletion syndrome associated with a recognizable phenotype.

Palomares M<sup>1</sup>, Delicado A, Mansilla E, de Torres ML, Vallespín E, Fernandez L, Martinez-Glez V, Garcia-Miñaur S, Lynch SA, Sharkey FH, Thuresson AC, Annerén G, Belligni EF, Martínez-Fernández ML, Bermejo E, Nowakowska B, Obersztyn E, Martínez-Frías ML, Hennekam RC, Lapunzina P.



- Round face with full cheeks
- High forehead
- Ptosis
- Corneal opacities
- Wide nasal bridge
- Underdeveloped alae
- Short philtrum
- Cupid's bow of the upper lip
- Downturned corners of the mouth
- Micrognathia
- Low-set and prominent ears
- Short neck
  
- Camptodactyly
- Syndactyly
- Broadening of the first rays
- Hypotonia
- Impaired balance
- Sensorineural hearing loss
- Underdeveloped corpus callosum
- Unusual behavior

## MCA/ID: trio-based WGS analysis of 3 patients with unknown clinical diagnosis

### 22 years old

- Non-consanguineous, healthy parents
- Ventricular septum defect
- Coarctation of the aorta
- Horseshoe kidney
- Bilateral choanal atresia
- Clinodactyly of the third and fourth finger
- Bilateral sandal gap
- Short stature
- Hyperextension of the knees and slumped shoulders
- Hypogenesis of the abdominal mesentery
- Mild intellectual disability
- Facial dysmorphism
  - Midfacial hypoplasia
  - Short palpebral fissures
  - High-arched palate
  - Undersized maxilla resulting in a nasal speech
  - Ptosis of the upper eyelids
  - Smallmouth and ears
  - Horner's syndrome



#### Molecular analysis:

- CHD7 negative
- Array-CGH negative



## MCA/ID: trio-based WGS analysis of 3 patients with unknown clinical diagnosis

### SNVs Analysis:

<b><i>Filtering criteria</i></b>	<b><i>Number of variants</i></b>
<i>Total</i>	6,912,472
<i>De novo variants</i>	102,190
<i>Rare variants (MAF&lt;1%)</i>	69,071
<i>Exonic and splice-site variants</i>	223
<i>CADD &gt; 20</i>	142
<i>Excluding synonymous variants</i>	91



**Negative**

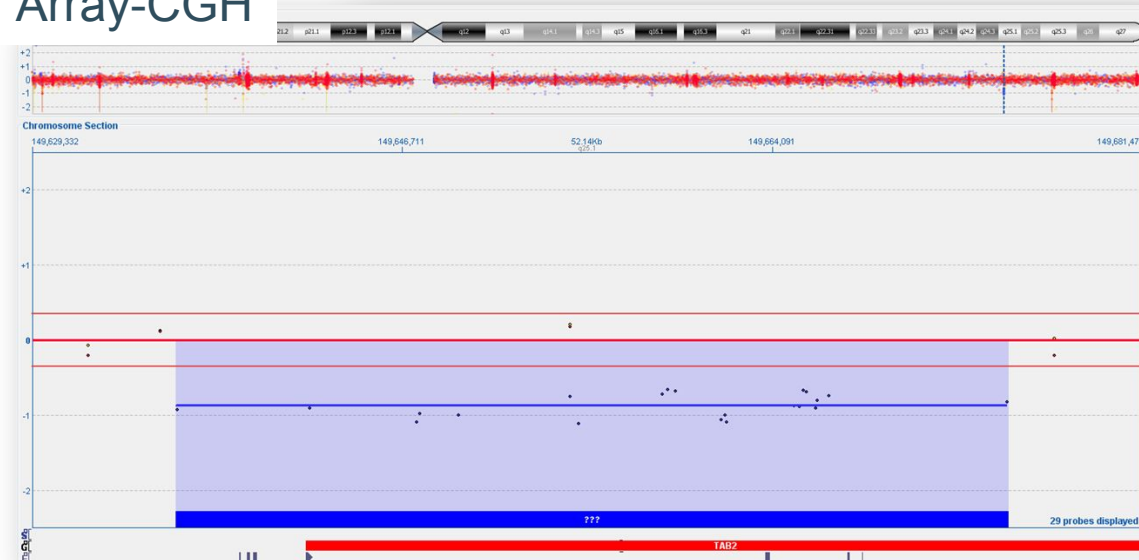
## MCA/ID: trio-based WGS analysis of 3 patients with unknown clinical diagnosis

### CNVs Analysis:

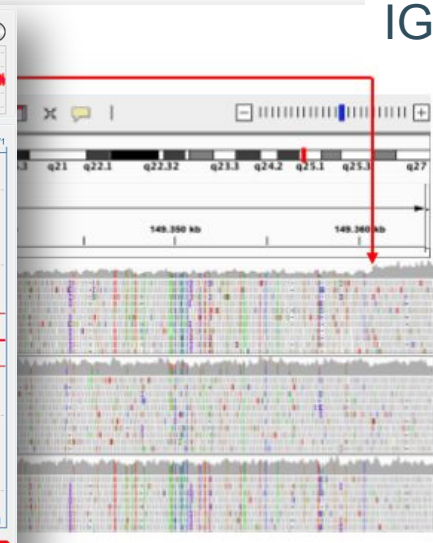
- fold change under 0.7 and above 1.3
- good mappability
- *de novo*

→ deletion  
**chr6: 149,308,196 - 149,360,335**  
in *TAB2* gene

### Array-CGH

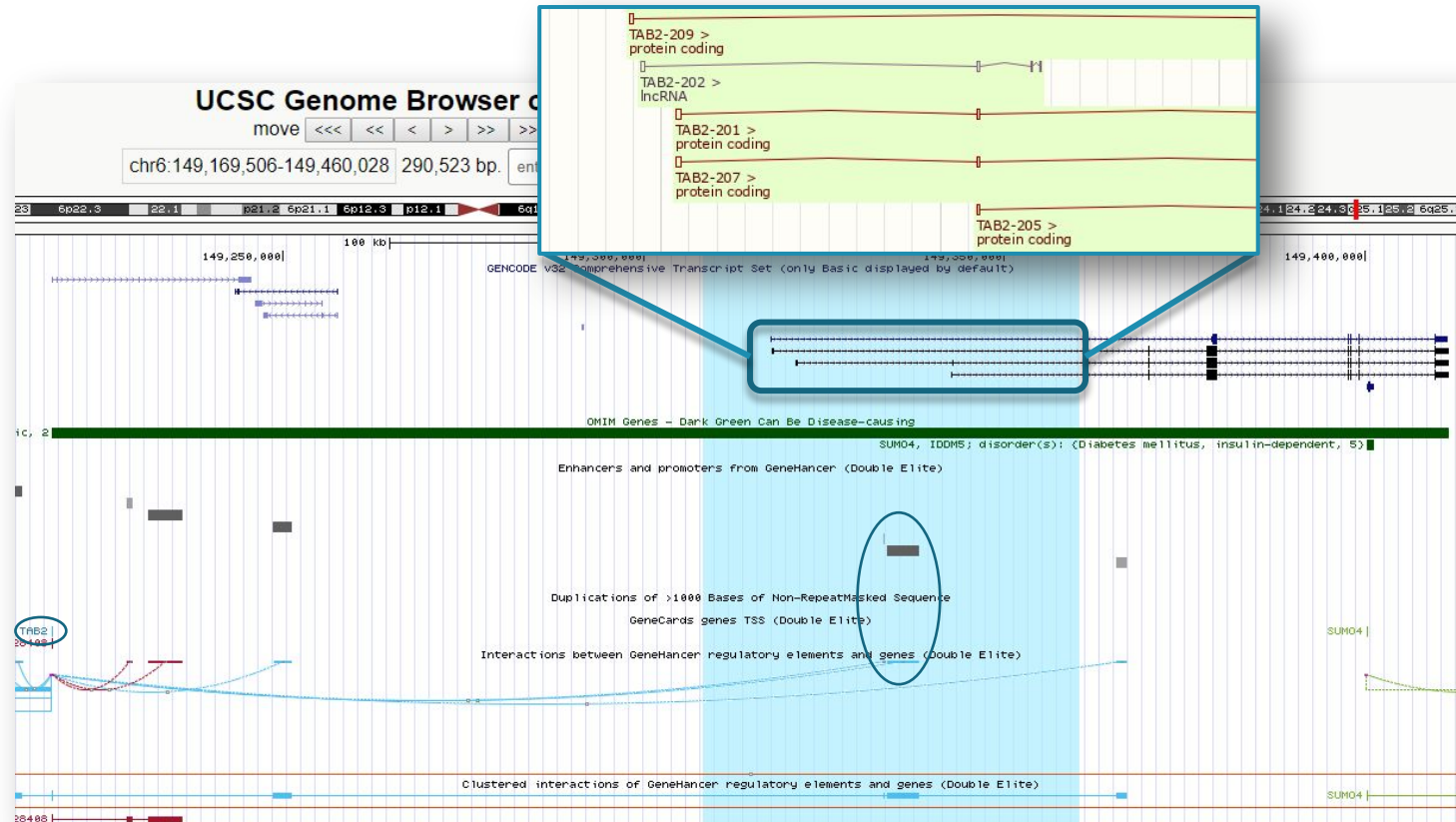


### IGV





## MCA/ID: trio-based WGS analysis of 3 patients with unknown clinical diagnosis

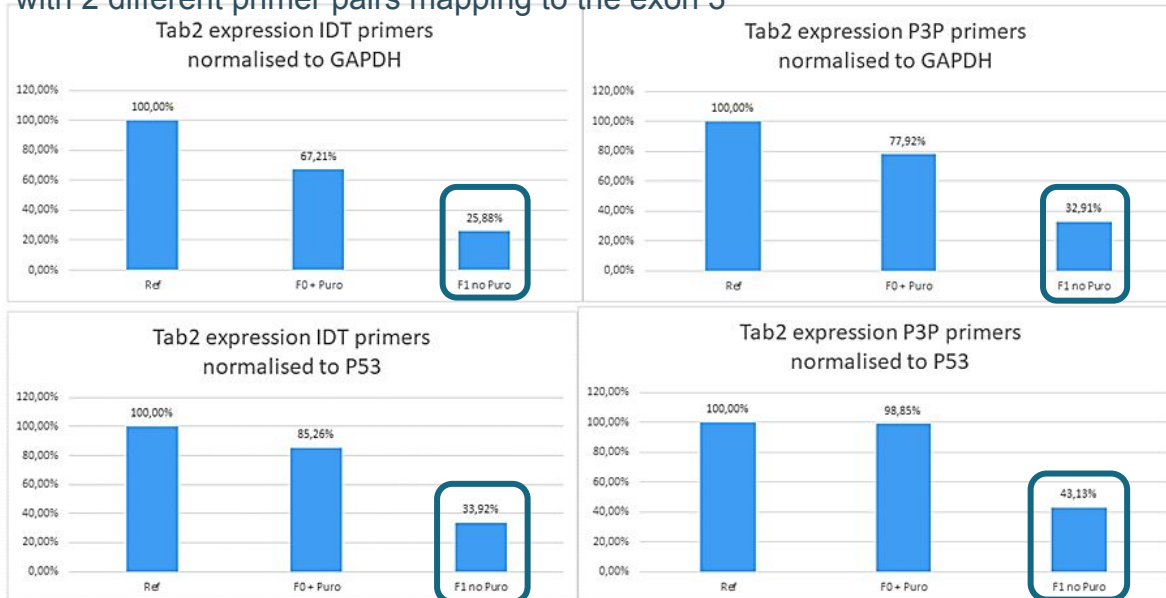


## MCA/ID: trio-based WGS analysis of 3 patients with unknown clinical diagnosis

### Gene expression

#### qRT-PCR

with 2 different primer pairs mapping to the exon 3



## MCA/ID: trio-based WGS analysis of 3 patients with unknown clinical diagnosis

### Whole Exome Sequencing, Familial Genomic Triangulation, and Systems Biology Converge to Identify a Novel Nonsense Mutation in *TAB2*-encoded TGF-beta Activated Kinase 1 in a Child with Polyvalvular Syndrome

Jaeger P. Ackerman, BA,\* John A. Smestad, BS,† David J. Tester, BS,\* Muhammad Y. Qureshi, MBBS,\* Beau A. Crabb, MS, CGC,‡ Nancy J. Mendelsohn, MD,‡ and Michael J. Ackerman, MD, PhD\*

### A recognizable systemic connective tissue disorder with polyvalvular heart dystrophy and dysmorphism associated with *TAB2* mutations

M. Ritelli<sup>1</sup> | S. Morlino<sup>2</sup> | E. Giacomuzzi<sup>1</sup> | L. Bernardini<sup>3</sup> | B. Torres<sup>3</sup> | G. Santoro<sup>1</sup> | V. Ravasio<sup>1</sup> | N. Chiarelli<sup>1</sup> | D. D'Angelantonio<sup>2</sup> | A. Novelli<sup>4</sup> | P. Grammatico<sup>2</sup> | M. Colombi<sup>1</sup> | M. Castori<sup>5</sup>

### A 2.6 Mb deletion of 6q24.3–25.1 in a patient with growth failure, cardiac septal defect, thin upper lip and asymmetric dysmorphic ears

R. Caselli<sup>a</sup>, M.A. Mencarelli<sup>a</sup>, F.T. Papa<sup>a</sup>, V. Uliana<sup>a</sup>, S. Schiavone<sup>b</sup>, M. Strambi<sup>b</sup>, C. Pescucci<sup>a</sup>, F. Ariani<sup>a</sup>, V. Rossi<sup>c</sup>, I. Longo<sup>a</sup>, I. Meloni<sup>a</sup>, A. Renieri<sup>a,\*</sup>, F. Mari<sup>a</sup>

### Familial *TAB2* Microdeletion and Congenital Heart Defects Including Unusual Valve Dysplasia and Tetralogy of Fallot

Karin Weiss,<sup>1</sup> Carolyn Applegate,<sup>2</sup> Tao Wang,<sup>2,3</sup> and Denise A. S. Batista<sup>2,4,5\*</sup>

### A De Novo 0.63 Mb 6q25.1 Deletion Associated with Growth Failure, Congenital Heart Defect, Underdeveloped Cerebellar Vermis, Abnormal Cutaneous Elasticity and Joint Laxity

Vincenzo Salpietro,<sup>1,2\*</sup> Martino Ruggieri,<sup>3</sup> Kshitij Mankad,<sup>4</sup> Gabriella Di Rosa,<sup>5</sup> Francesca Granata,<sup>6</sup> Italia Loddo,<sup>7</sup> Emanuela Moschella,<sup>8</sup> Maria Pia Calabro,<sup>7</sup> Anna Capalbo,<sup>8</sup> Laura Bernardini,<sup>8</sup> Antonio Novelli,<sup>9</sup> Agata Polizzi,<sup>10,11</sup> Daniela G. Seidler,<sup>12</sup> Teresa Arrigo,<sup>2</sup> and Silvana Briuglia<sup>2</sup>

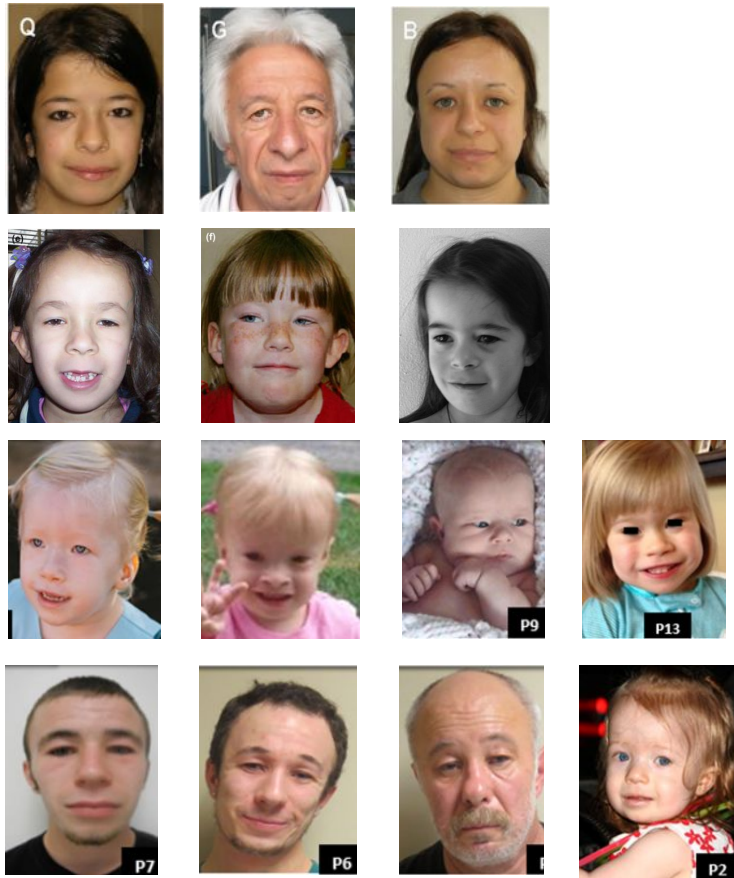
### Haploinsufficiency of *TAB2* Causes Congenital Heart Defects in Humans

Bernard Thienpont,<sup>1,14</sup> Litu Zhang,<sup>2,15</sup> Alex V. Postma,<sup>3</sup> Jeroen Breckpot,<sup>1</sup> Léon-Charles Tranchevent,<sup>4</sup> Peter Van Loo,<sup>5,6</sup> Kjeld Møllgård,<sup>7</sup> Niels Tommerup,<sup>2</sup> Iben Bache,<sup>2</sup> Zeynep Tümer,<sup>2,8</sup> Klaartje van Engelen,<sup>9</sup> Björn Menten,<sup>10</sup> Geert Mortier,<sup>10,11</sup> Darrel Waggoner,<sup>12</sup> Marc Gewillig,<sup>13</sup> Yves Moreau,<sup>4</sup> Koen Devriendt,<sup>1</sup> and Lars Allan Larsen<sup>2,\*</sup>

### 6q25.1 (*TAB2*) microdeletion syndrome: Congenital heart defects and cardiomyopathy

Andrew Cheng<sup>1</sup> | Mary Beth P. Dinulos<sup>2</sup> | Whitney Neufeld-Kaiser<sup>3</sup> | Jill Rosenfeld<sup>4</sup> | McKenna Kyriss<sup>5</sup> | Suneeta Madan-Khetarpal<sup>6</sup> | Hiba Risheg<sup>7</sup> | Peter H. Byers<sup>3</sup> | Yajuan J. Liu<sup>3</sup>

## MCA/ID: trio-based WGS analysis of 3 patients with unknown clinical diagnosis

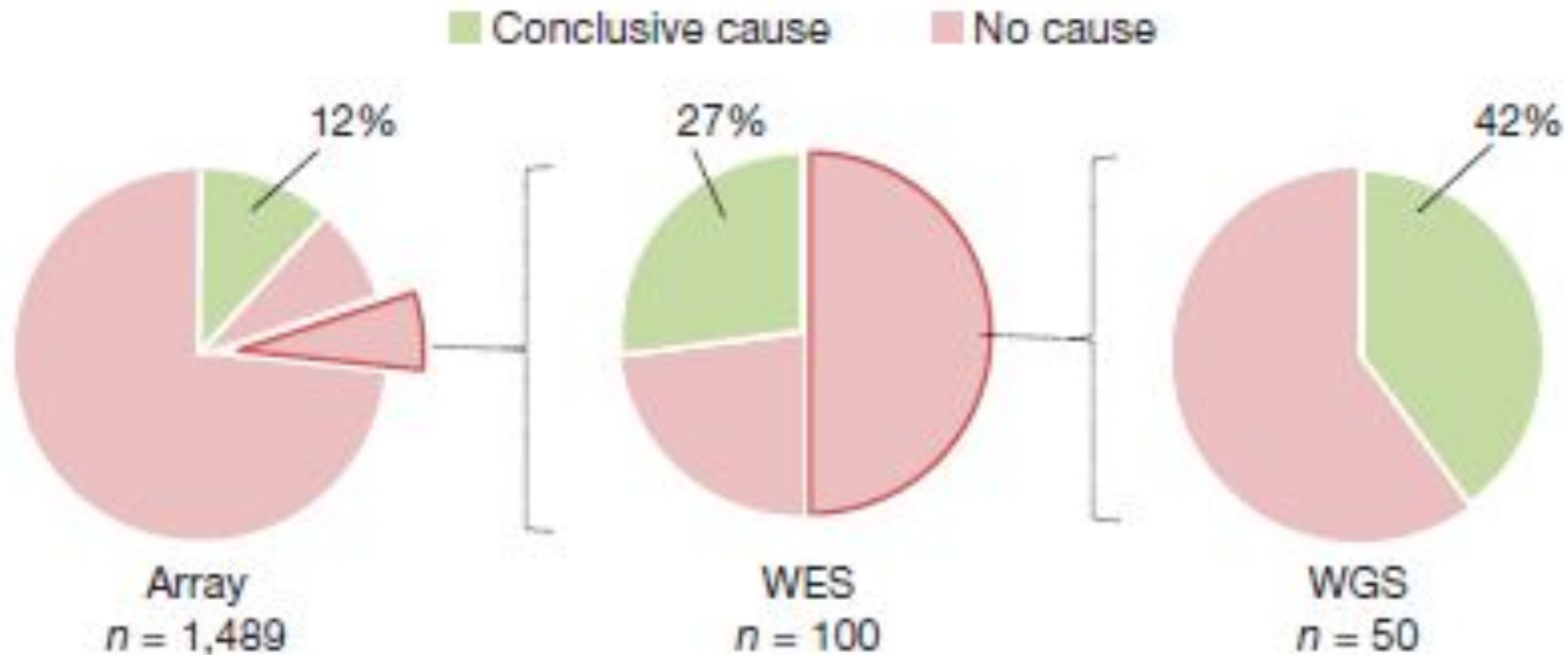


- CHDs
- Facial dysmorphism
- Growth failure
- Joint laxity
- Hypotonia
- Connective tissue abnormalities
- Developmental or intellectual disability

- Horseshoe kidney
- Bilateral choanal atresia
- Maldevelopment of the abdominal cavity



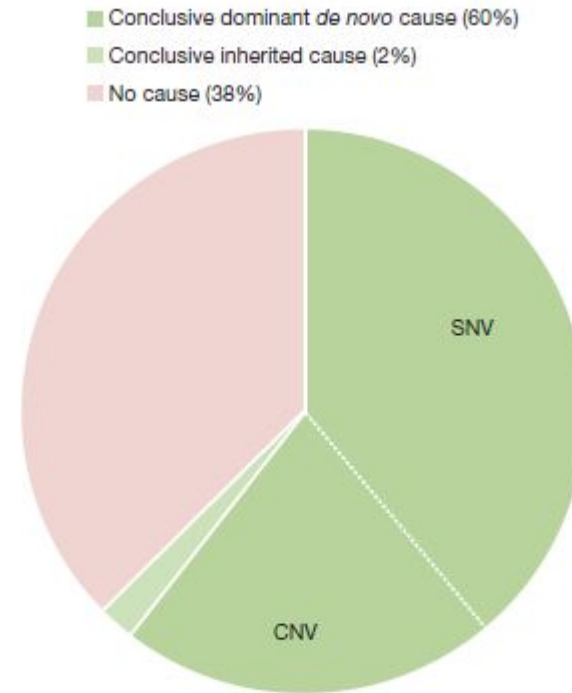
# Analysis of the complete genome (SNPs + CNVs)



Genome sequencing identifies major causes of severe intellectual disability

Christian Gilissen<sup>1\*</sup>, Jayne Y. Hehir-Kwa<sup>2\*</sup>, Dje Tjwan Thung<sup>1</sup>, Maartje van de Vorst<sup>1</sup>, Bregje W.M. van Bon<sup>1</sup>, Marjolijn H. Willenssen<sup>1</sup>, Michael Kwint<sup>1</sup>, Irene M. Janssen<sup>1</sup>, Alexander Hoischen<sup>1</sup>, Annette Schenck<sup>1</sup>, Richard Liang<sup>3</sup>, Robert Klein<sup>4</sup>, Rick Teasdale<sup>5</sup>, Tan Bo<sup>6,7</sup>, Ralph Pfundt<sup>1</sup>, Helger G. Yntema<sup>1</sup>, Bert B. A. de Vries<sup>1</sup>, Tjibbe de Graaf<sup>1</sup>, Hanneke van den Veyver<sup>1,8</sup>, Lisinka E. L. M. Vissers<sup>1\*</sup> & Joris A. Veltman<sup>1\*</sup>

# Analysis of the complete genome



Diagnostic yield: 62%

Genome sequencing identifies major causes of severe intellectual disability

Christian Gilissen<sup>1\*</sup>, Jayne Y. Hehir-Kwa<sup>1\*</sup>, Dje Tjwan Thung<sup>1</sup>, Maartje van de Vorst<sup>1</sup>, Bregje W.M. van Bon<sup>1</sup>, Marjolijn H. Willenssen<sup>2</sup>, Michael Kwint<sup>1</sup>, Irene M. Janssen<sup>1</sup>, Alexander Hoischen<sup>1</sup>, Annette Schenck<sup>3,4</sup>, Richard Liang<sup>5</sup>, Robert Klein<sup>2</sup>, Rick Teare<sup>6</sup>, Tan Bo<sup>7,8</sup>, Ralph Pfundt<sup>1</sup>, Helger G. Yntema<sup>1</sup>, Bert B. A. de Vries<sup>1</sup>, Tjibbe de Graaf<sup>1</sup>, Hanneke van der Vliet<sup>1,4\*</sup>, Lisienka E. L. M. Vissers<sup>2\*</sup> & Joris A. Veltman<sup>1,4\*</sup>

University

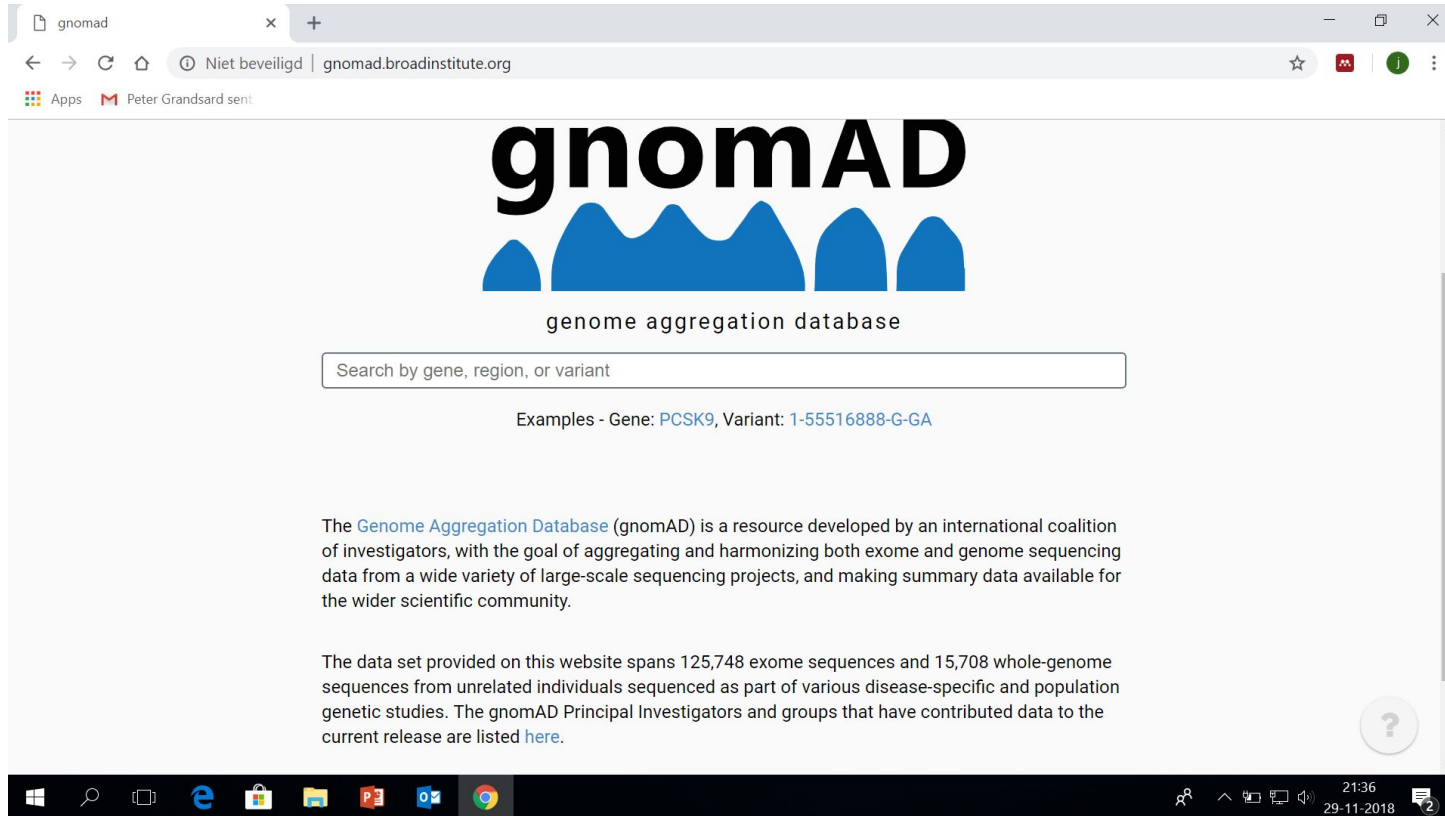
KU LEUVEN

# The burden of autosomal recessive diseases in rare developmental disorders

- DDD study
  - 3.6% autosomal recessive
  - 40% de novo coding mutations
- Pakistani study:
  - 30.9% autosomal recessive
  - 30% de novo dominant

# gnomAD (past Exac) database

## exercise



The screenshot shows a web browser window with the URL `gnomad.broadinstitute.org`. The page features the gnomAD logo, which consists of the text "gnomAD" in a large, bold, black font, with a blue silhouette of a mountain range underneath. Below the logo is the text "genome aggregation database". A search bar is present with the placeholder text "Search by gene, region, or variant". Below the search bar, there is an example search: "Examples - Gene: [PCSK9](#), Variant: [1-55516888-G-GA](#)". The main content area contains two paragraphs of text. The first paragraph states: "The [Genome Aggregation Database](#) (gnomAD) is a resource developed by an international coalition of investigators, with the goal of aggregating and harmonizing both exome and genome sequencing data from a wide variety of large-scale sequencing projects, and making summary data available for the wider scientific community." The second paragraph states: "The data set provided on this website spans 125,748 exome sequences and 15,708 whole-genome sequences from unrelated individuals sequenced as part of various disease-specific and population genetic studies. The gnomAD Principal Investigators and groups that have contributed data to the current release are listed [here](#)." The browser's taskbar at the bottom shows various application icons and the system tray with the time 21:36 and date 29-11-2018.



# We are all mutants!

