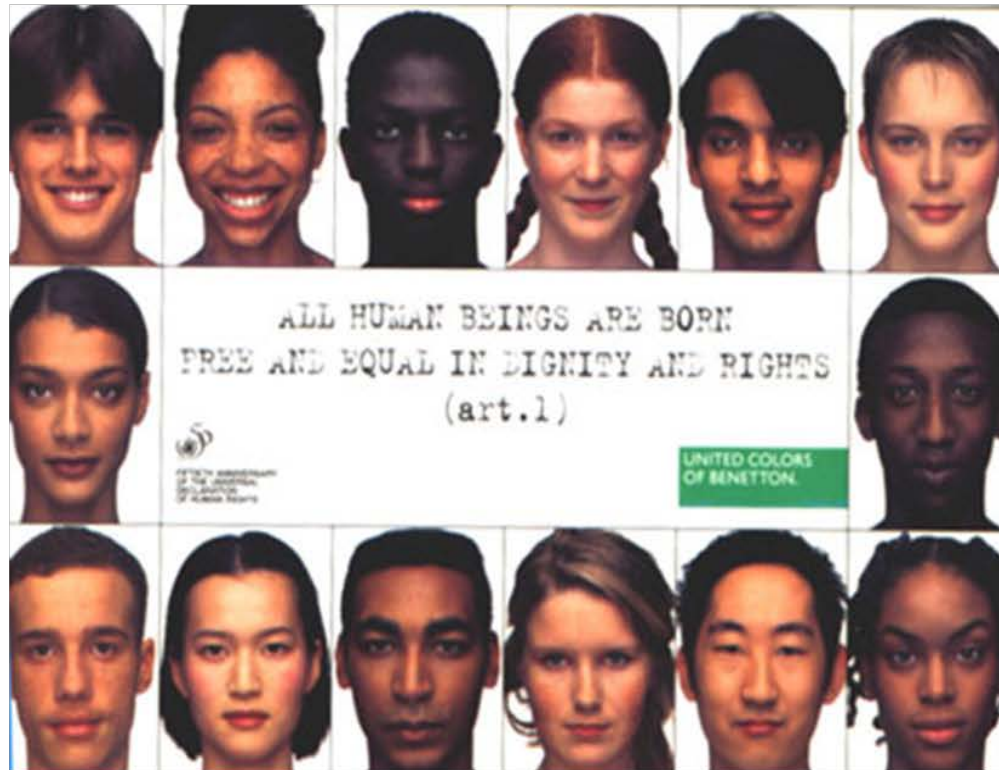


Tools for human molecular diagnostics/Human Genetic diversity

Joris Vermeesch
BeSHG 2021

Why are we different?

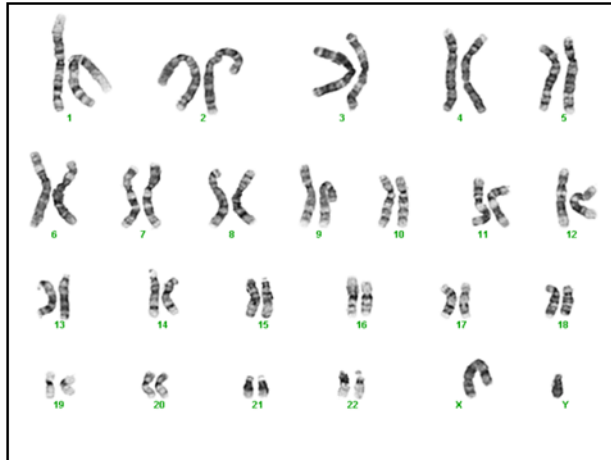


Human Genetic Variation

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

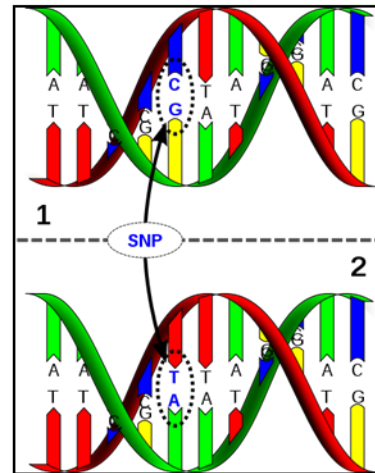
Causes of genetic variation

Chromosomes
(1960)



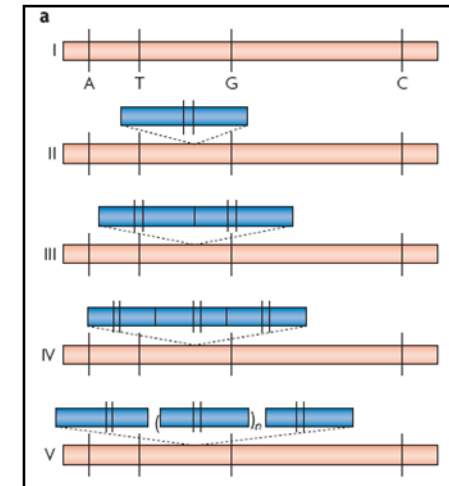
Variant are rare

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!

Types of variation and their consequences

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

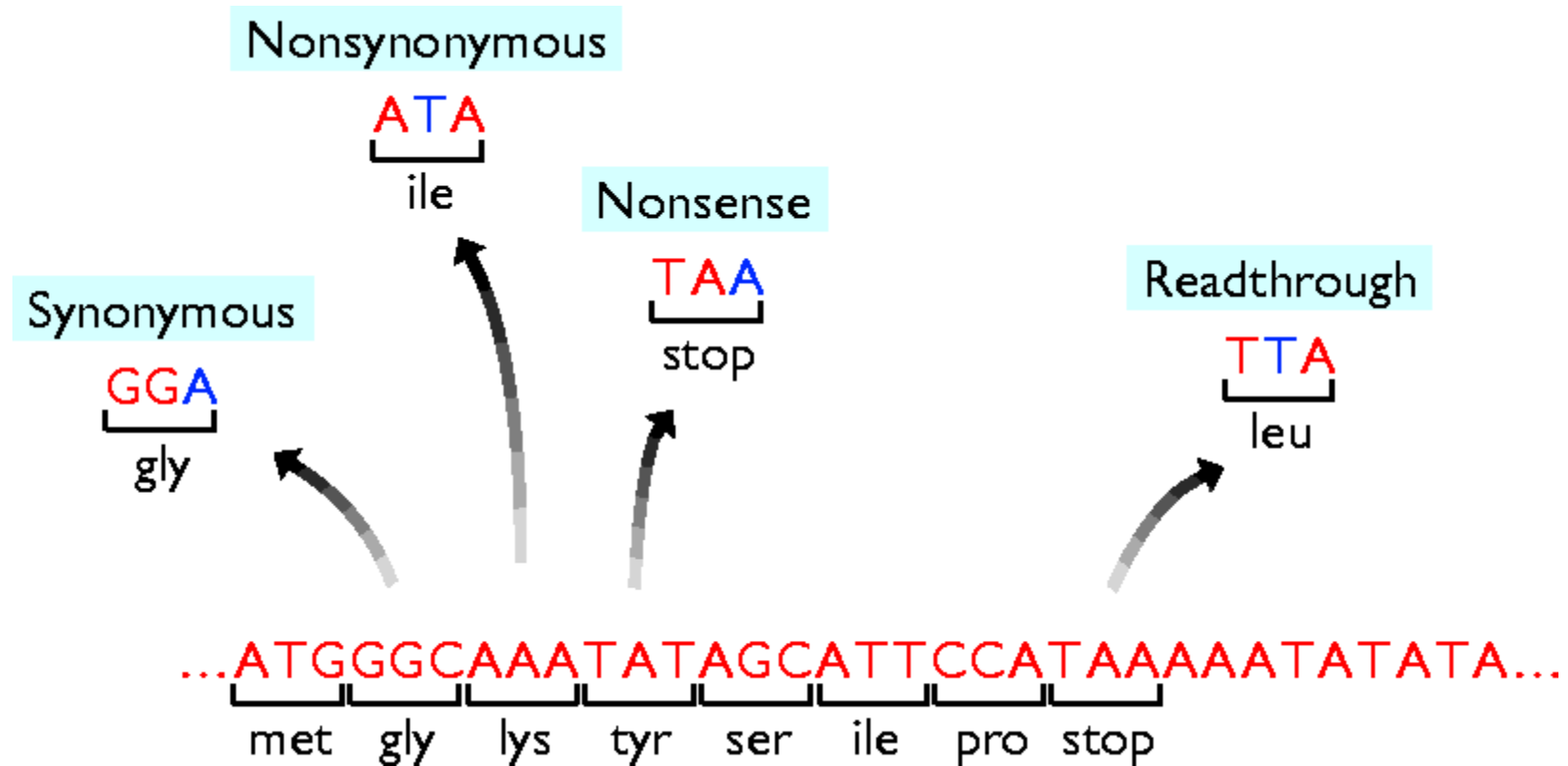
SNPs : Common variation in the genome

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

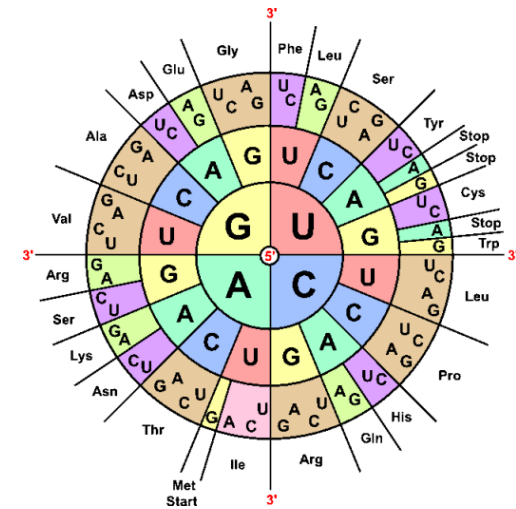
SNPs : Common variation in the genome

	5	10	15	20
Reference sequence	...GGATT	TTCTAGG	TAACTC	AGTCGA...
SNP	<i>Allele 1</i>	...GGATT	TTCTAGG	TAACTCAGTCGA...
	<i>Allele 2</i>	...GGATT	TTCT C AGG	TAACTCAGTCGA...
Indel A	<i>Allele 1</i>	...GGATT	TTCTAGG	TAACTCAGTCGA...
	<i>Allele 2</i>	...GGATT	TTCTAGG G	TAACTCAGTCGA...
Indel B	<i>Allele 1</i>	...GGATT	TTCTAGG	TAACTCAGTCGA...
	<i>Allele 2</i>	...GGAT	- - CTAGG	TAACTCAGTCGA...

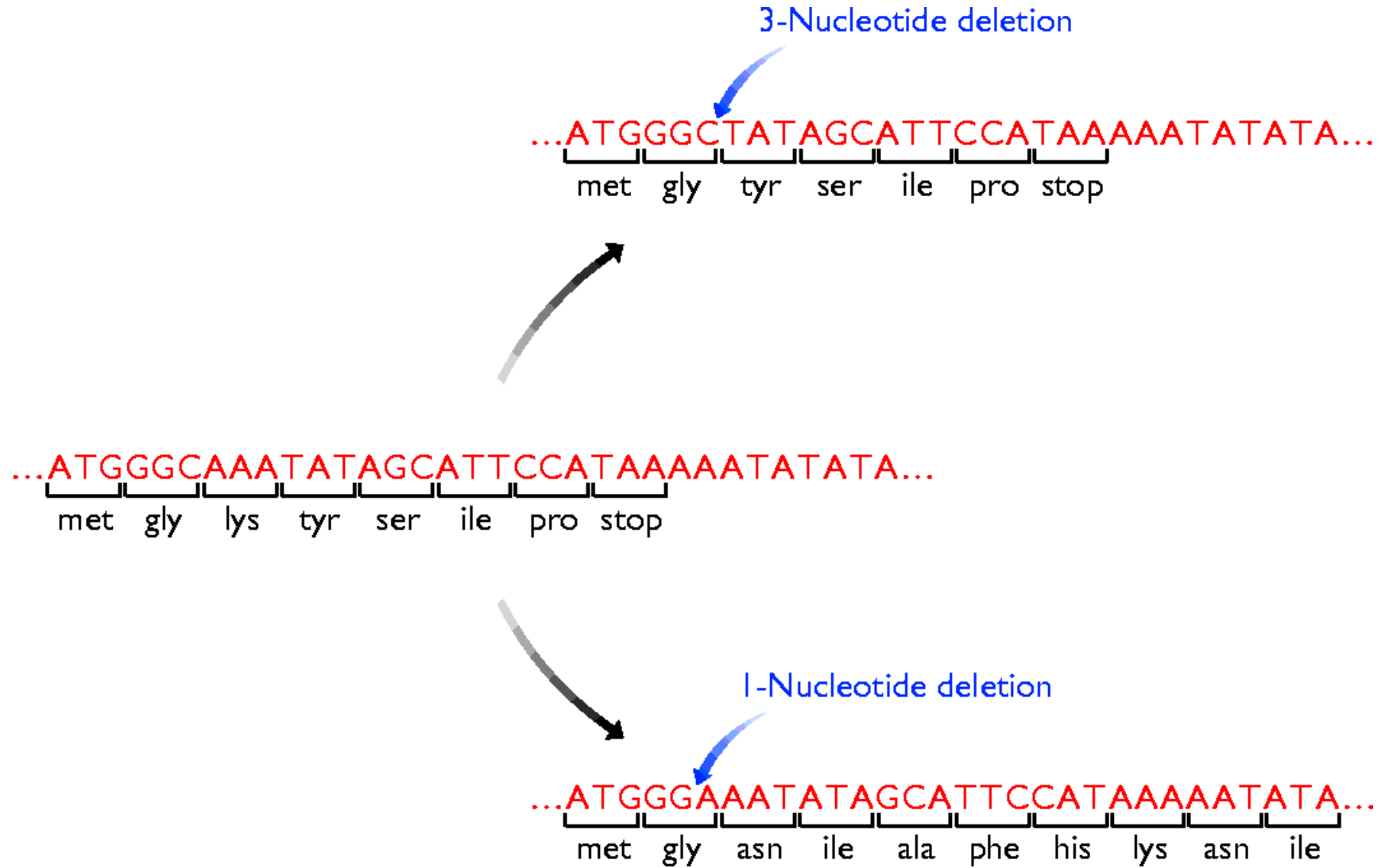
Point mutations (SNV)



		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 } Ile AUC } 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 } Val GUC } GUA } GUG }	GCU } Ala GCC } GCA } GCG }	GAU } Asp GAC } GAA } Glu GAG }	GGU } Gly GGC } 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

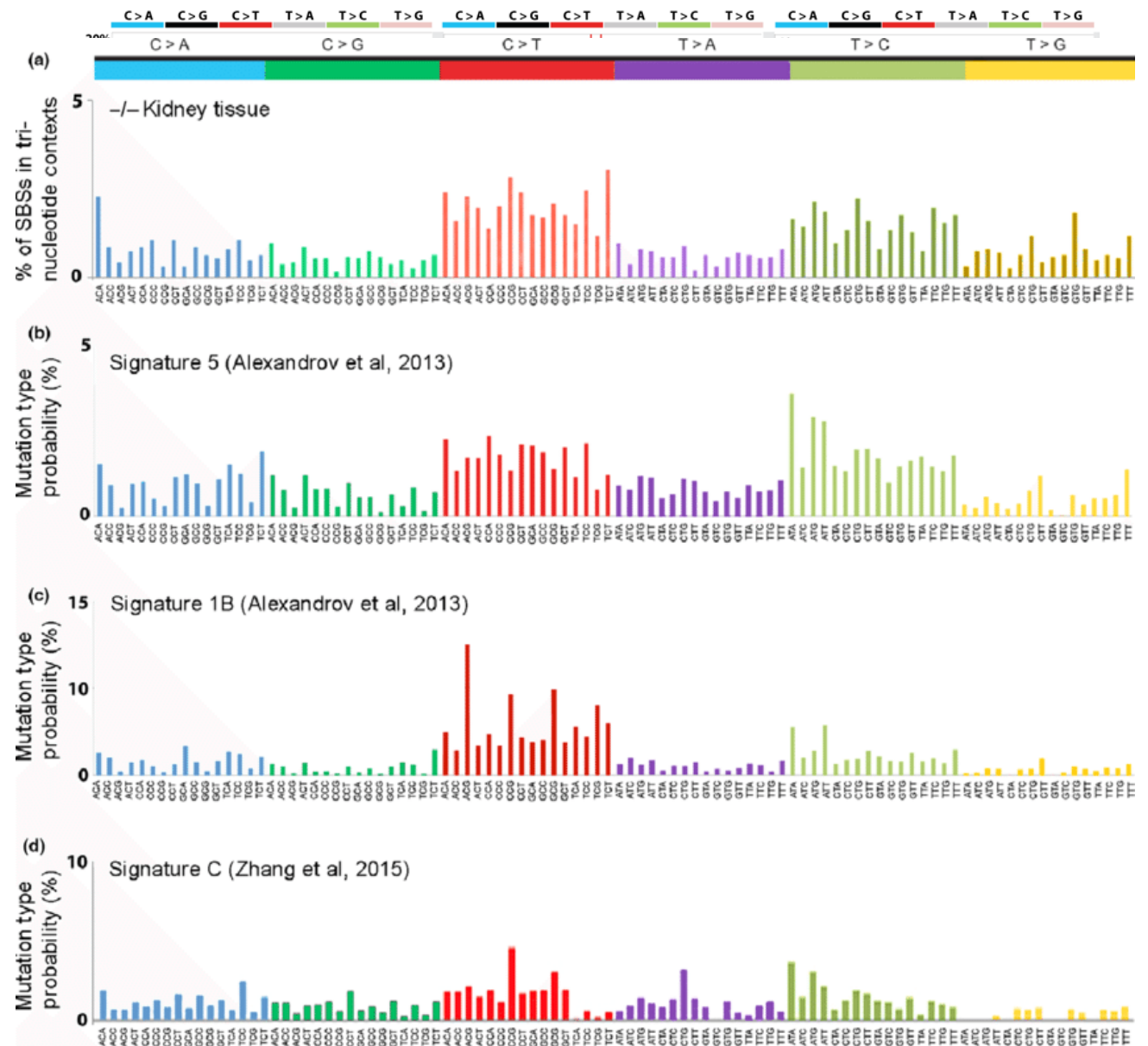
COSMIC, the Catalogue Of Somatic Mutations In Cancer, is the world's largest and most comprehensive resource for exploring the impact of somatic mutations in human cancer.

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:



Splice consensus signals

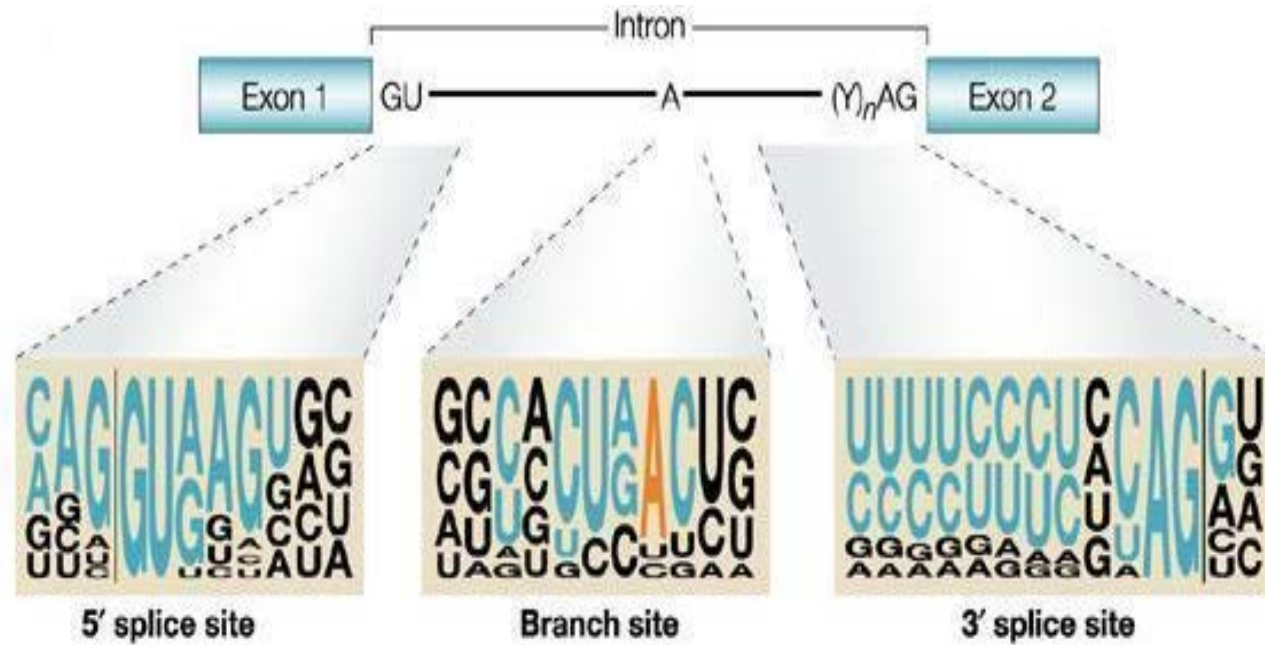
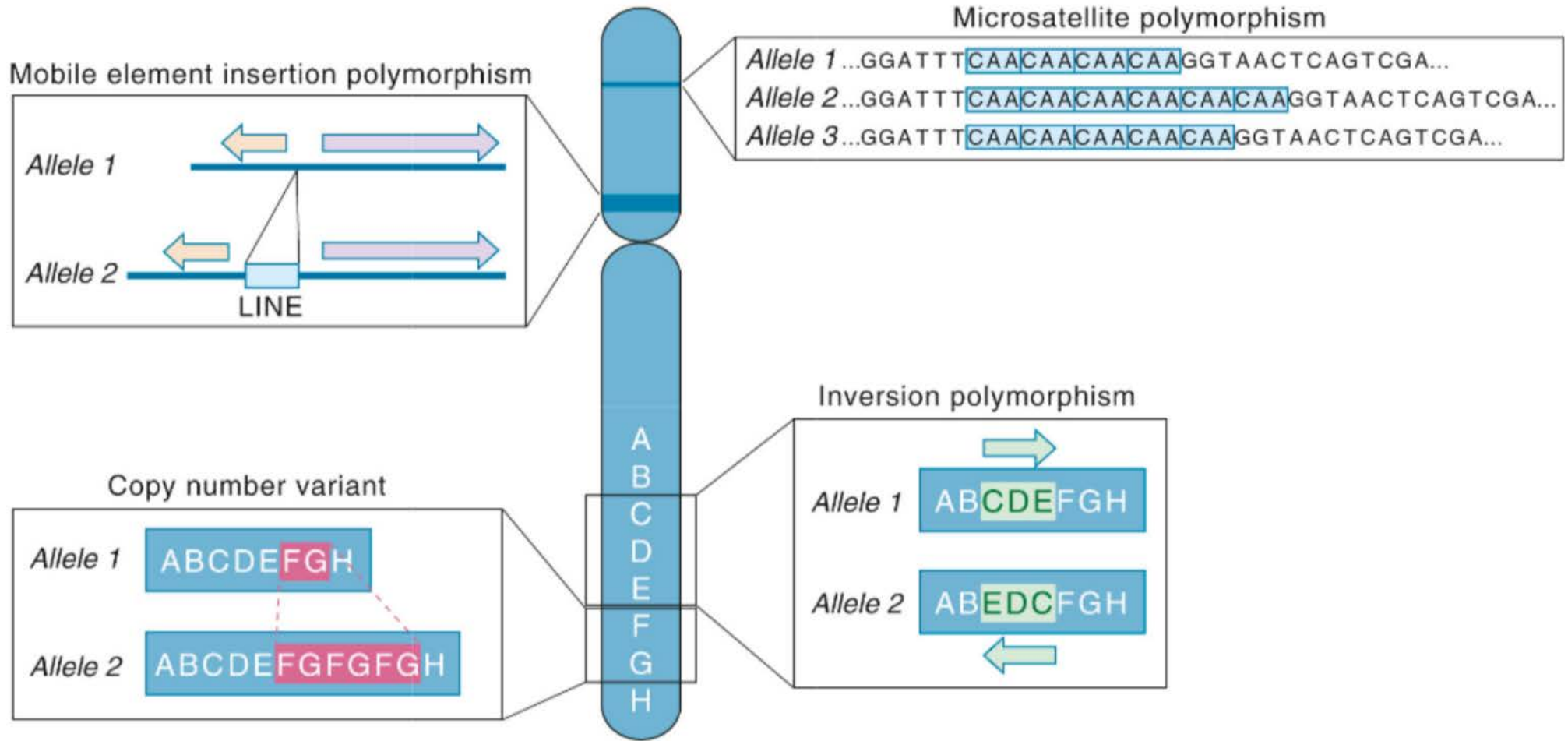


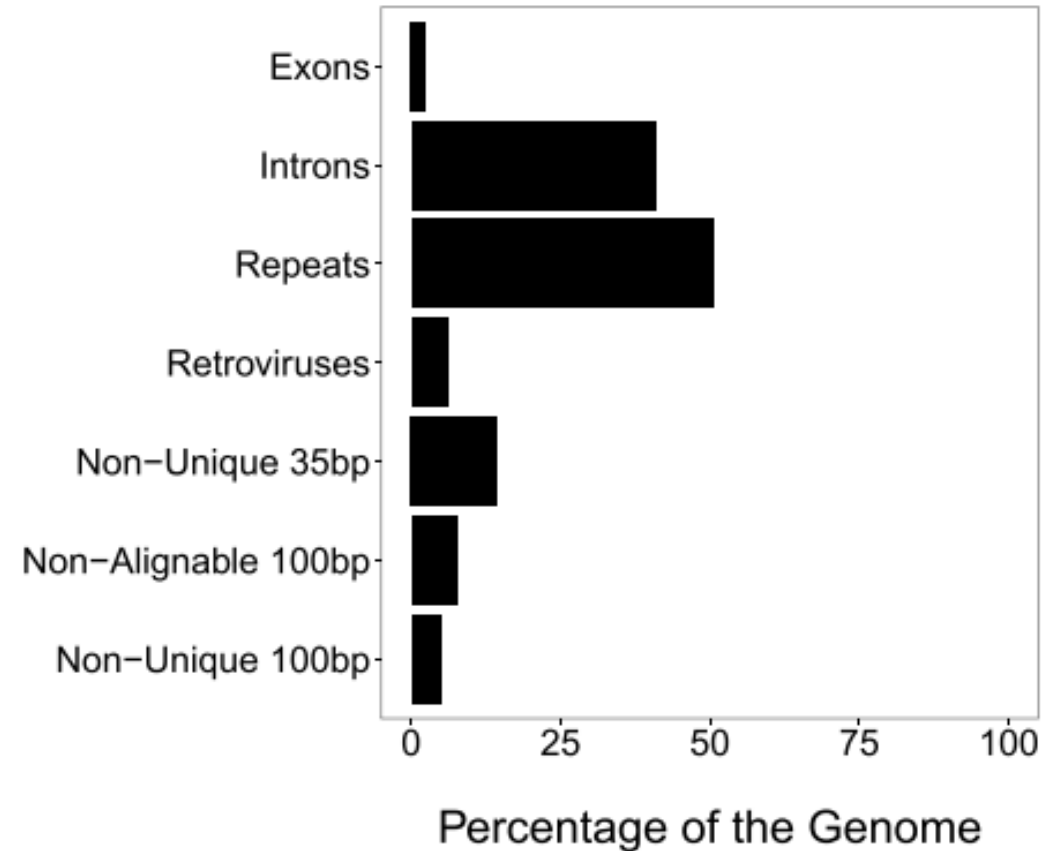
Figure 1

Nature Reviews | Genetics

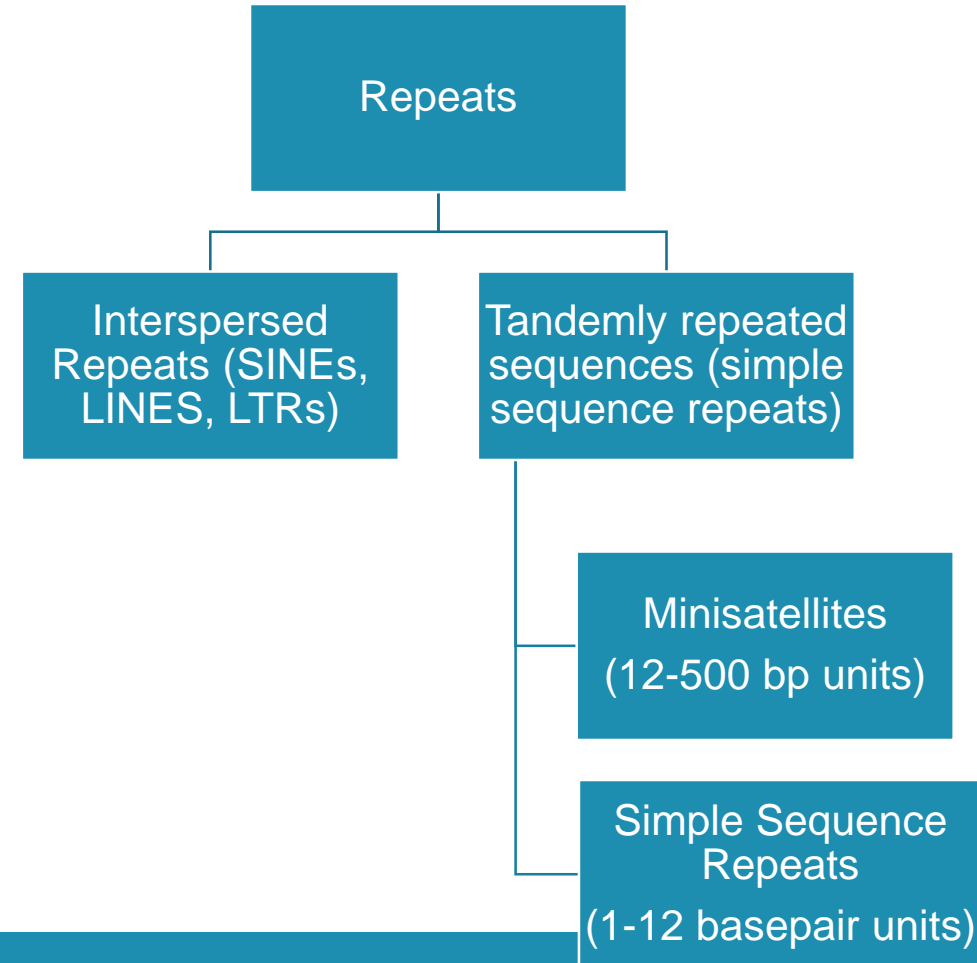
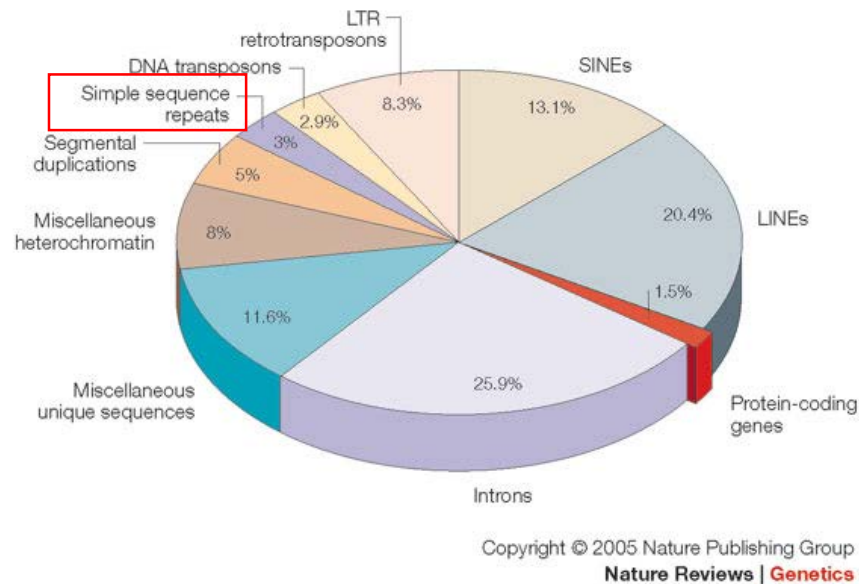


Het humane genoom bevat 50% repeats

A



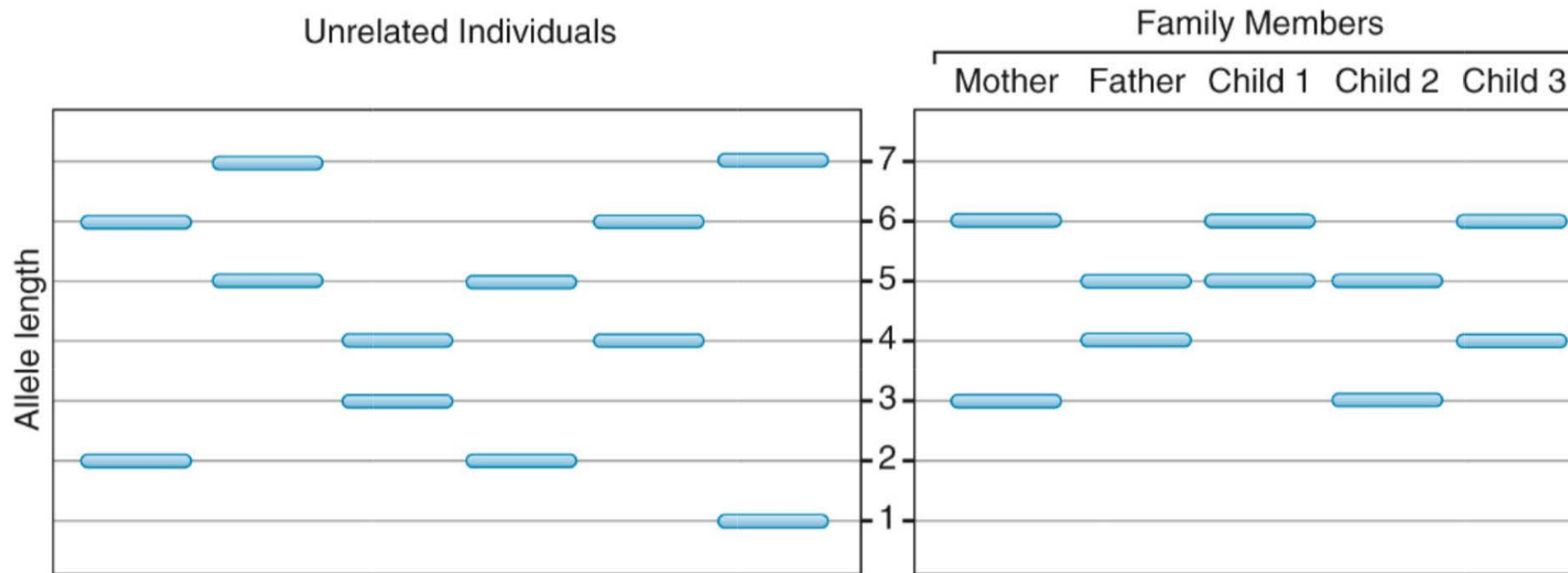
Het humane genoom bevat 50% repeats



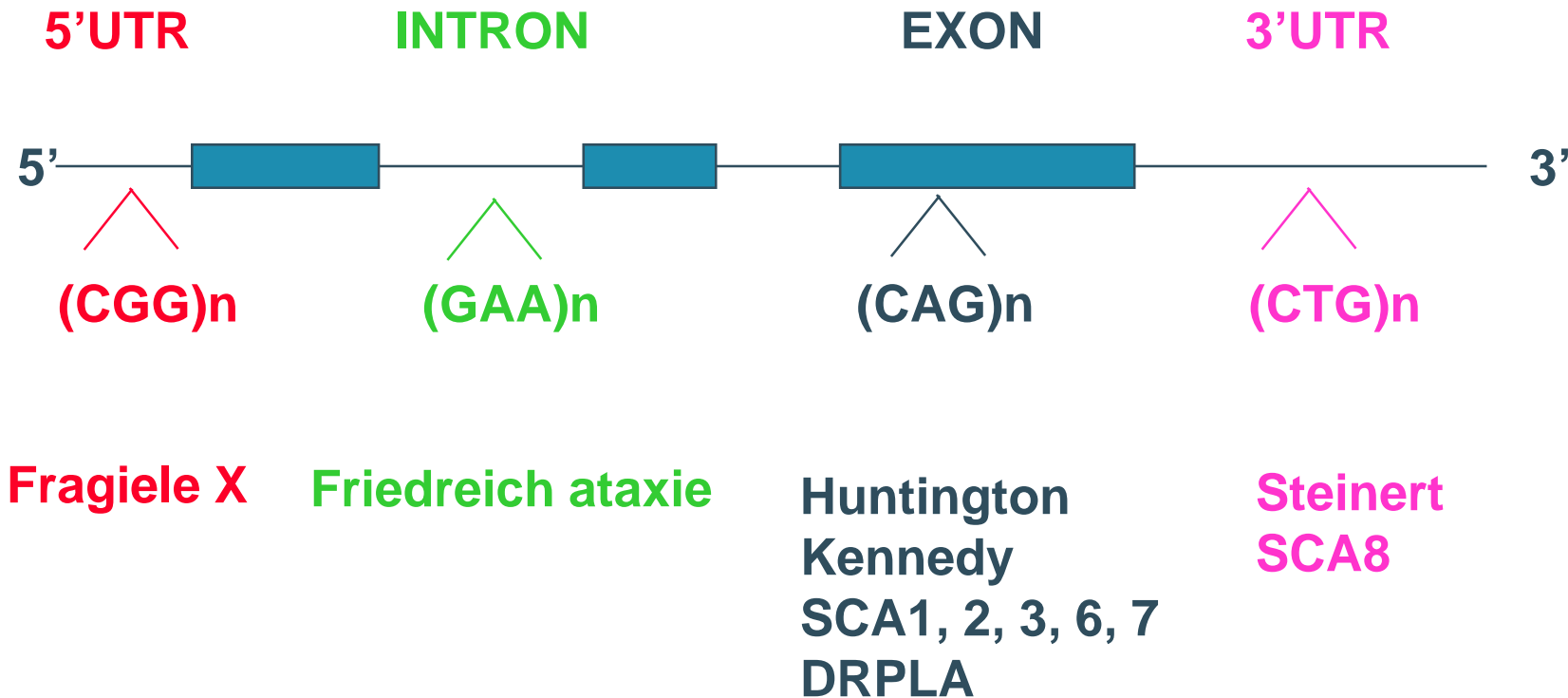
Short tandem repeat (STR)

- 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.



Tandem repeats in genes are associated with disorders characterized by anticipation

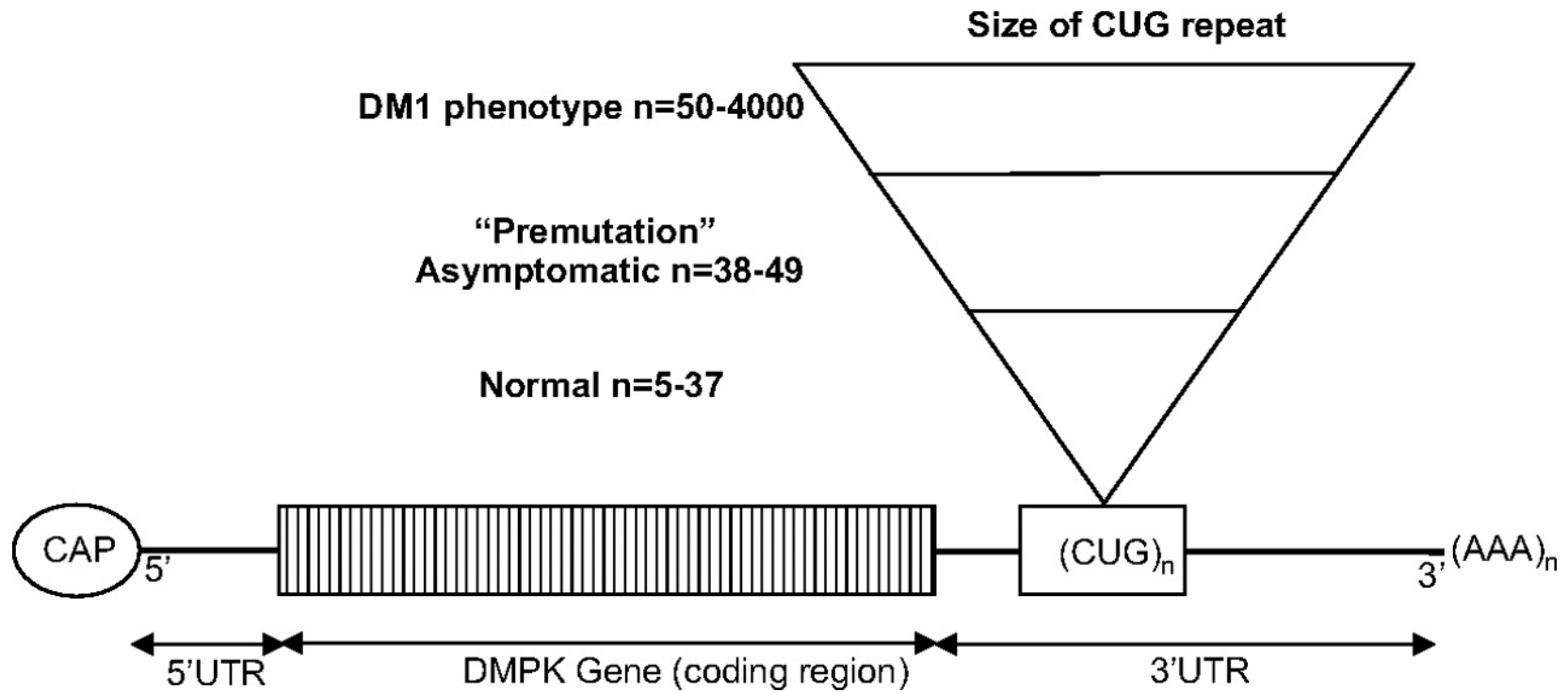


Myotonic dystrophy (Steinert disease) (as an example)

- *Autosomal dominant*
- *Trinucleotide repeat expansion*



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

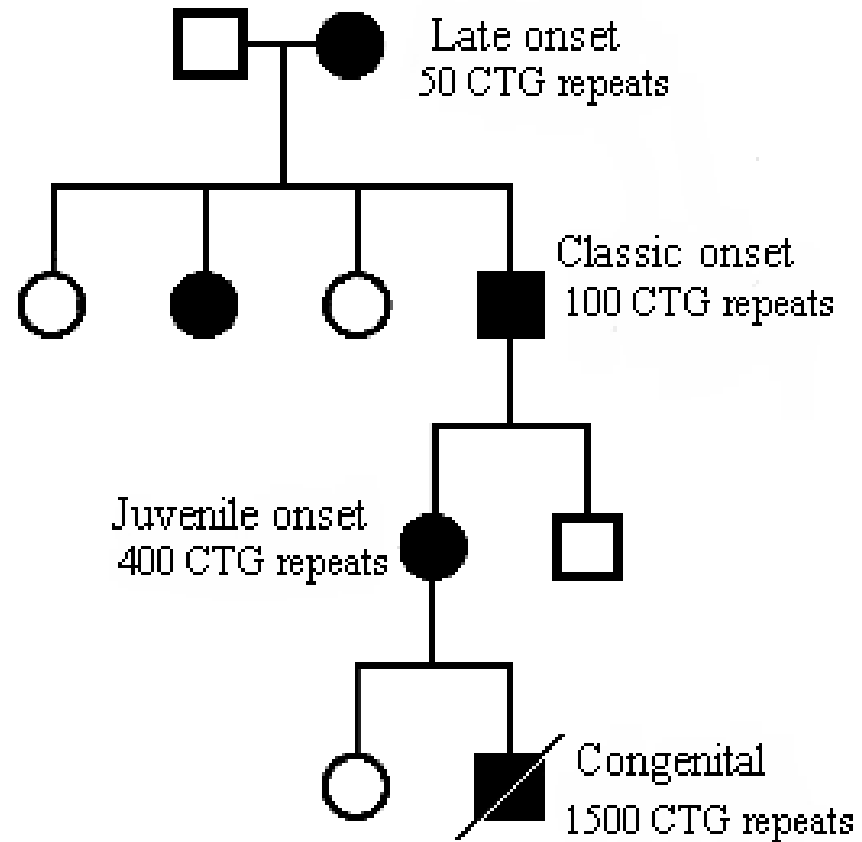
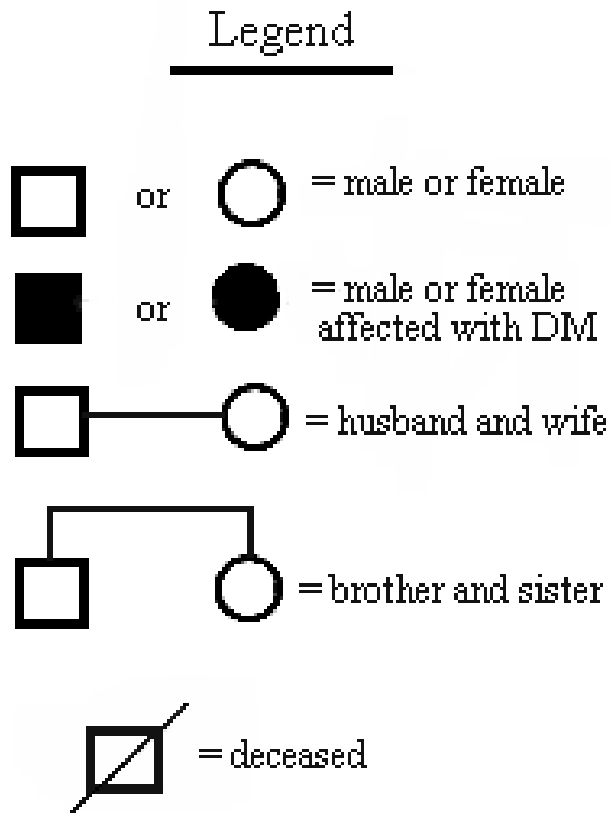


Myotonic dystrophy (Steinert disease)

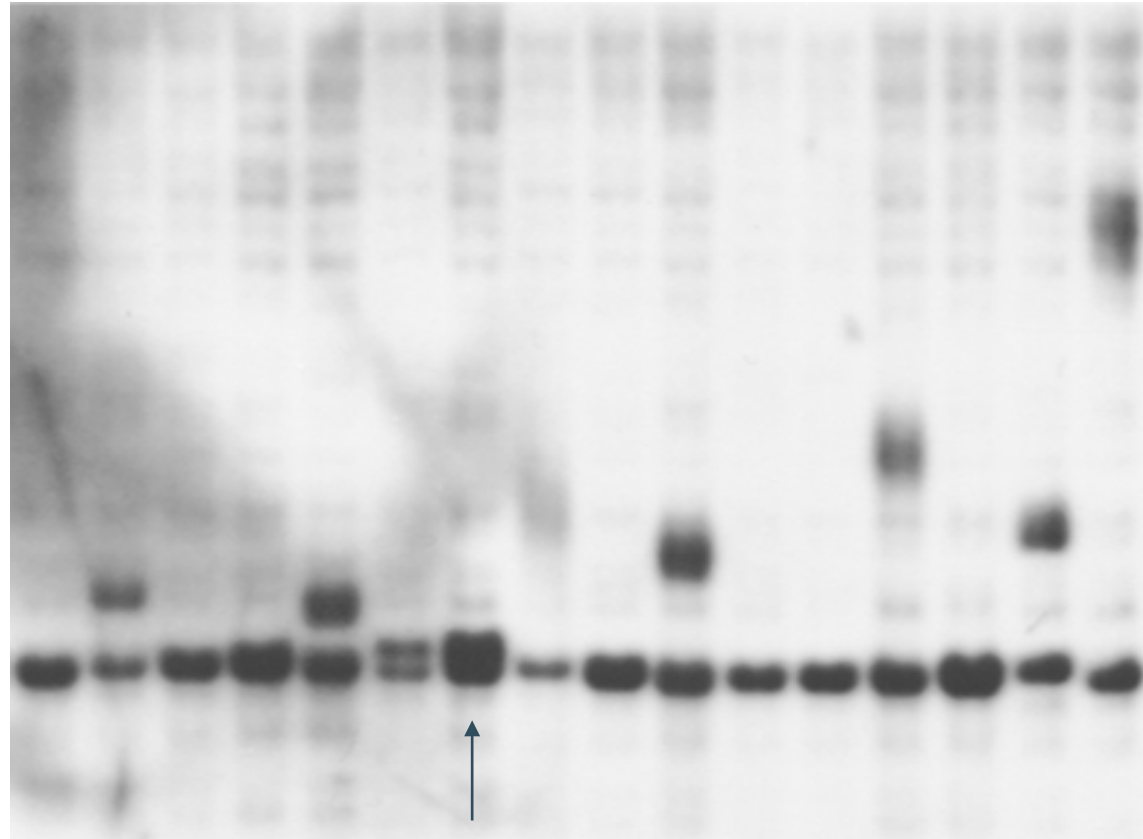
Hereditary aspects

- 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

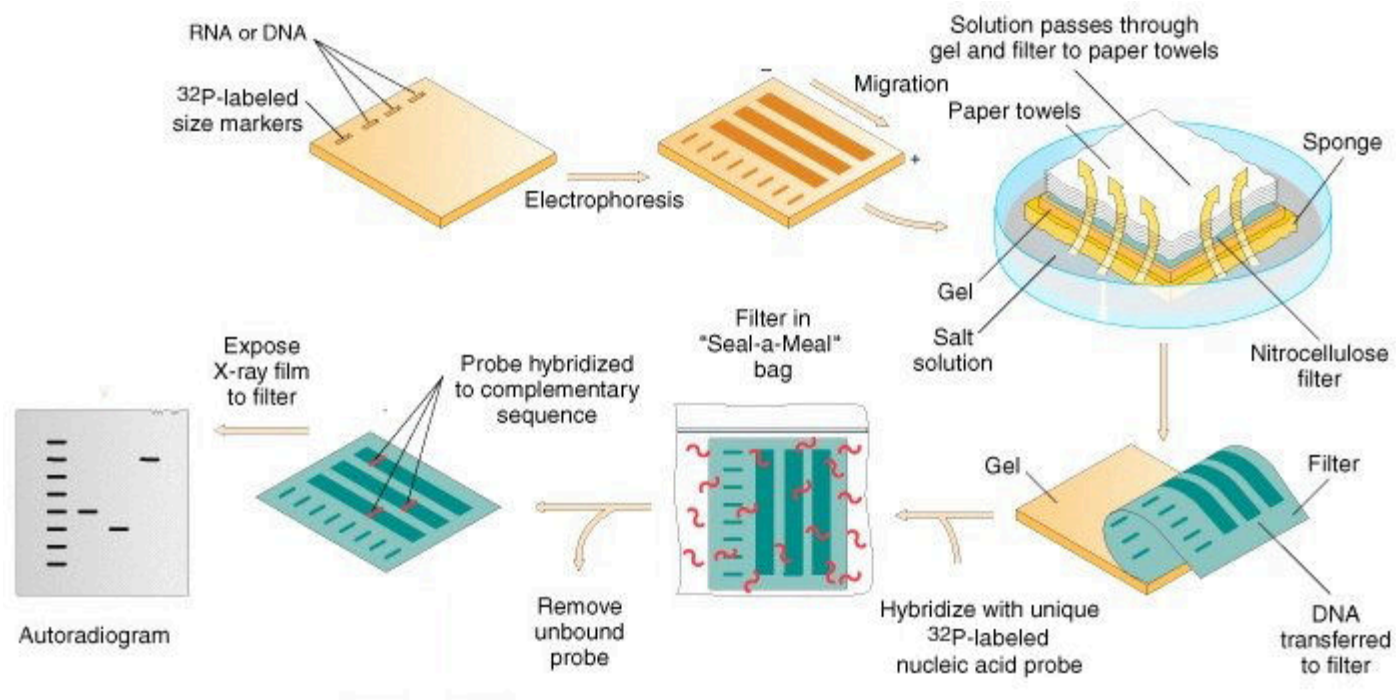
← n = 100

← n = 4-12

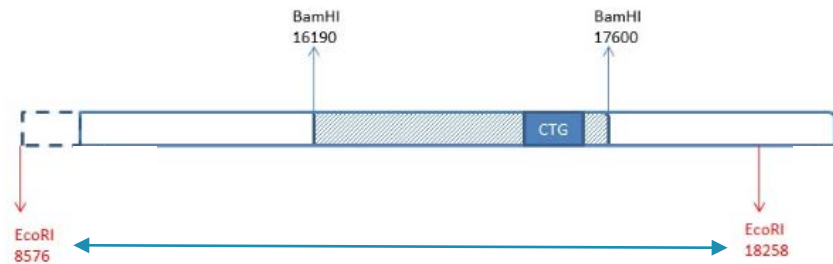
n = 35

Southern blot om lange expansies te bepalen:

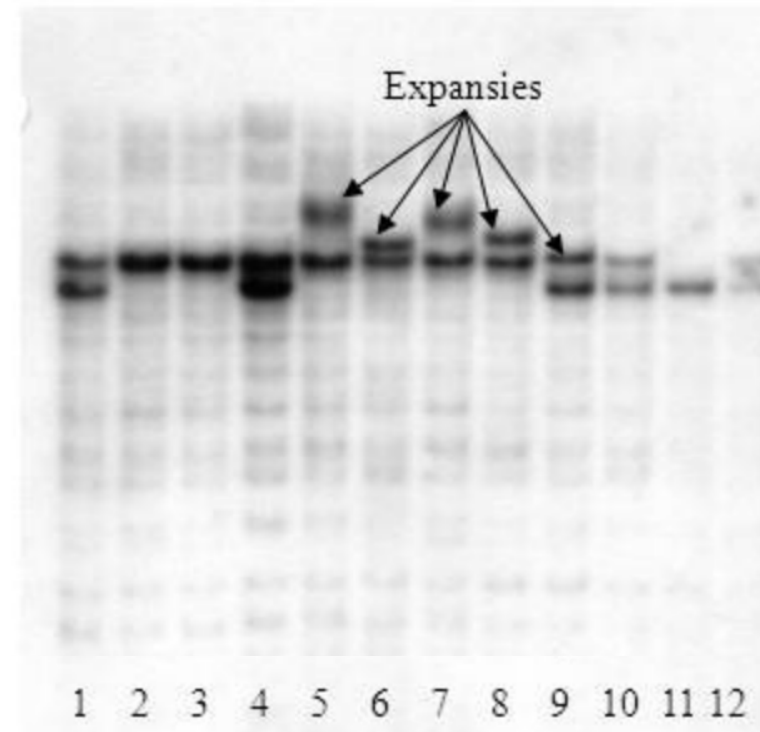
1. Knippen van humaan genoom met restrictie enzymen



Southern blot to detect large expansions

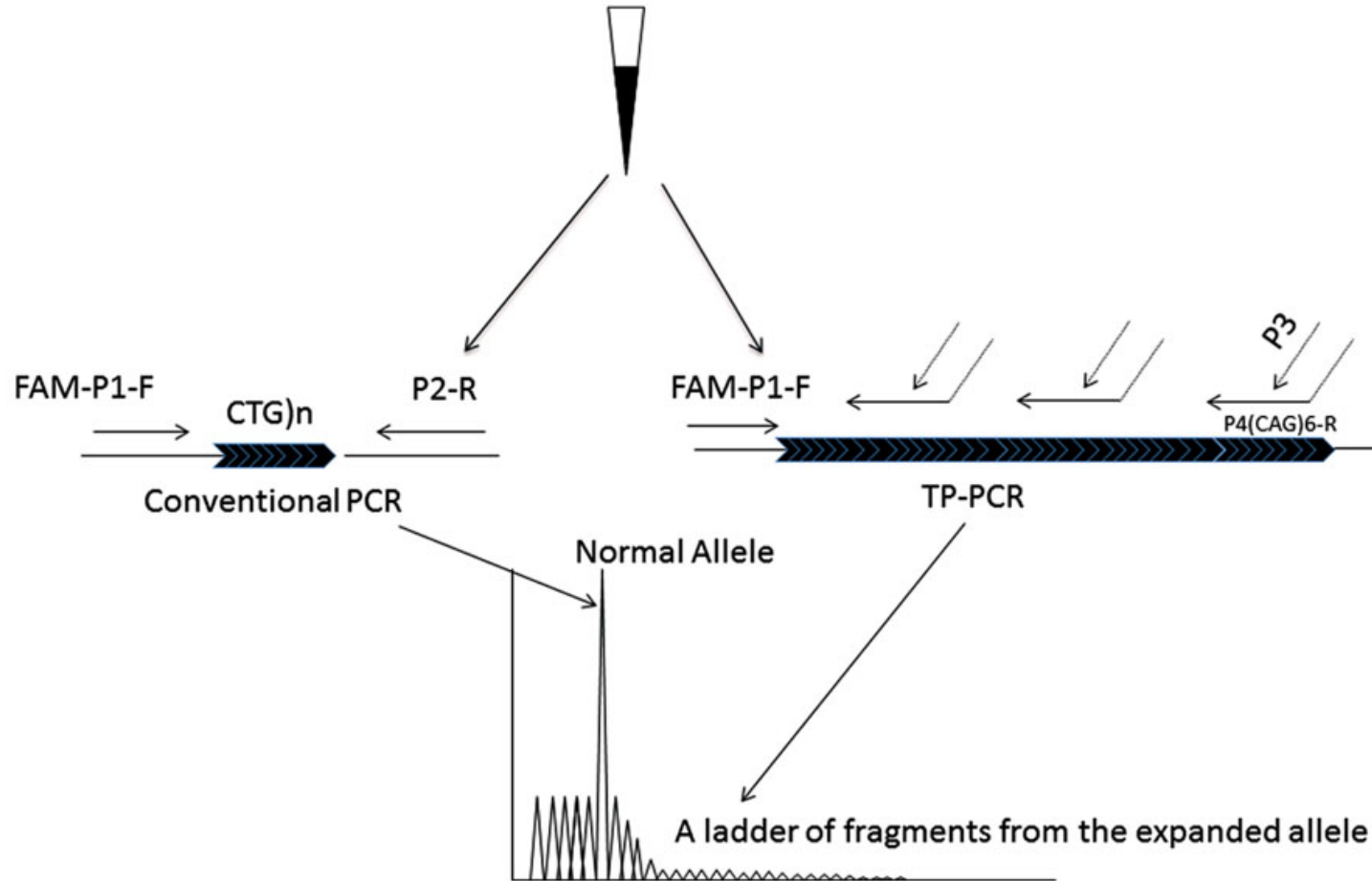


9682 bp

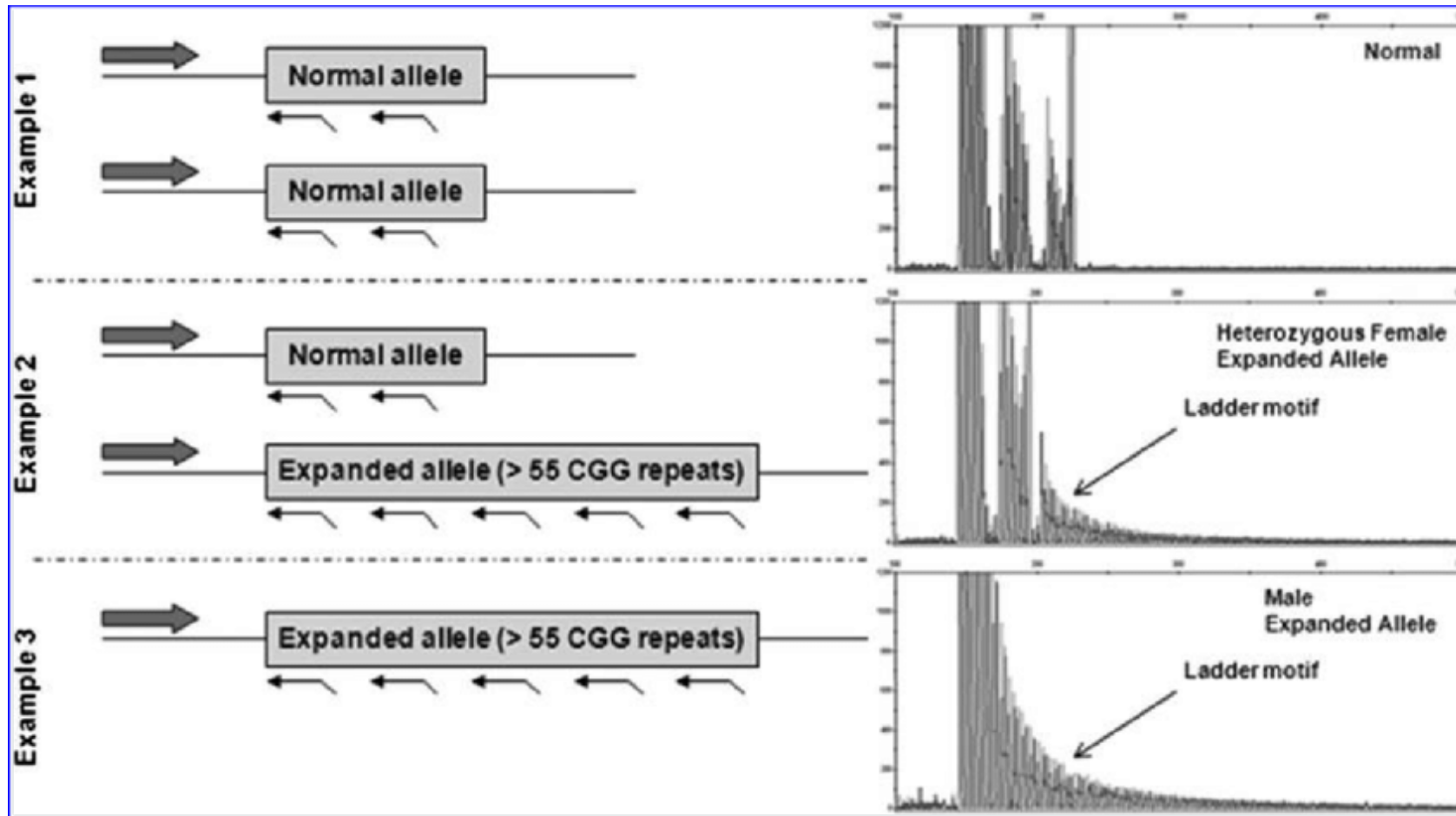


- > Fragment can be recognized via probe
- > Fragment will be larger with larger expansions

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


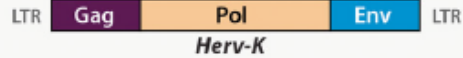
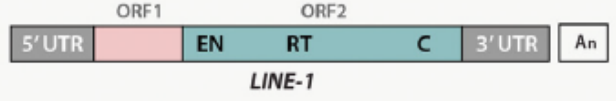


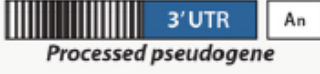


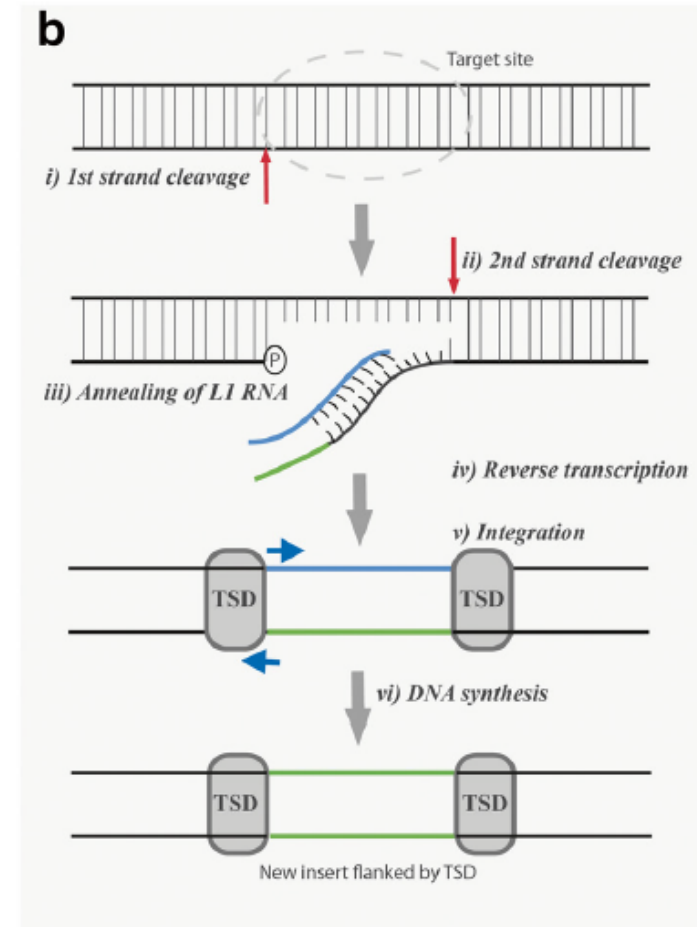
Triplet primed PCR

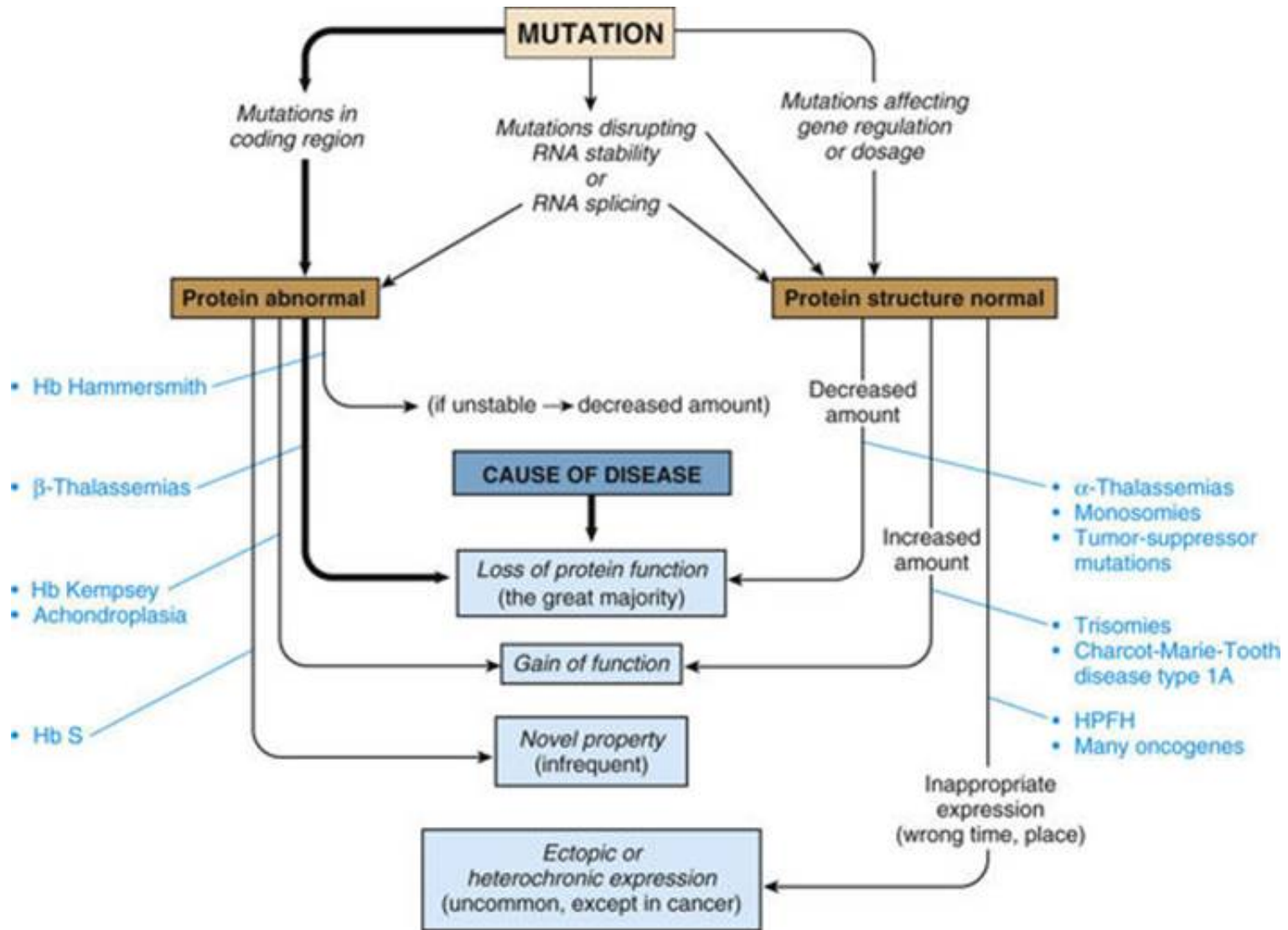


Active mobile elements in the human genome

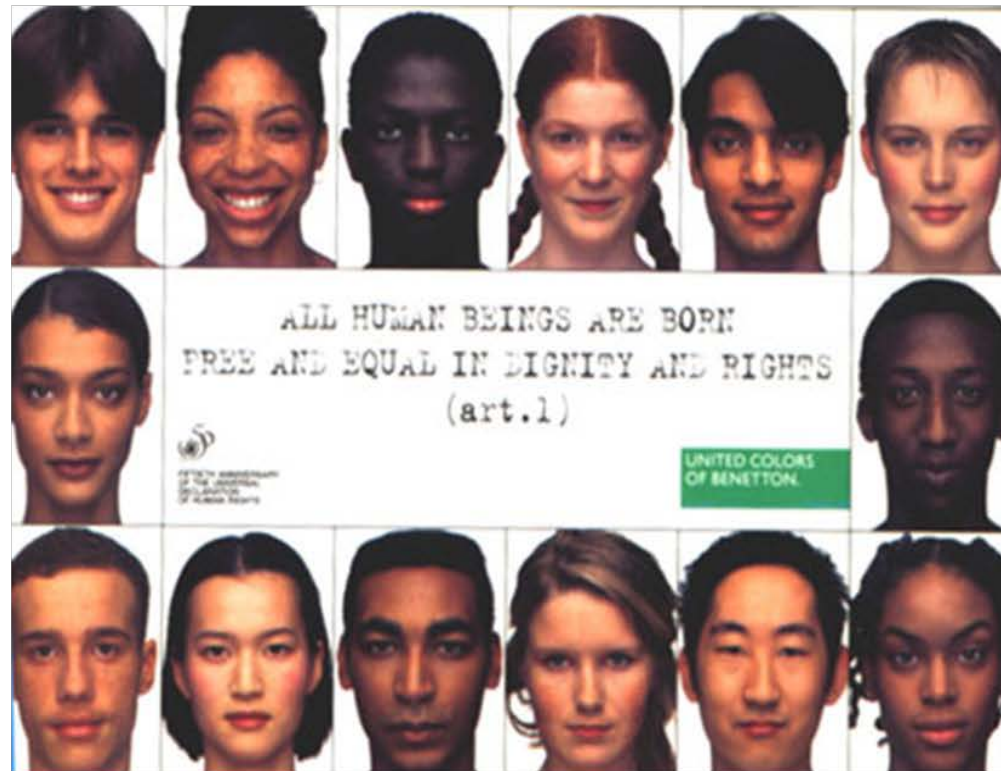
a

Mobile element structure	HGR	Length	Remarks
DNA transposons:  IR Transposase IR <i>Mariner</i>	2-3%	1.4 kb	Transposes from less than 100 kb to distant sites from original site
Retrotransposons (autonomous):  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
Non-autonomous:  Left arm A B An Right arm Ins An <i>Alu</i>  TSD (CCCTCT) _n 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
		Variable	SVA insertions occurs at high frequency, so far 3 cases of human disease reported Arises by reverse transcription of cellular mRNA & integration of cDNA in the genome

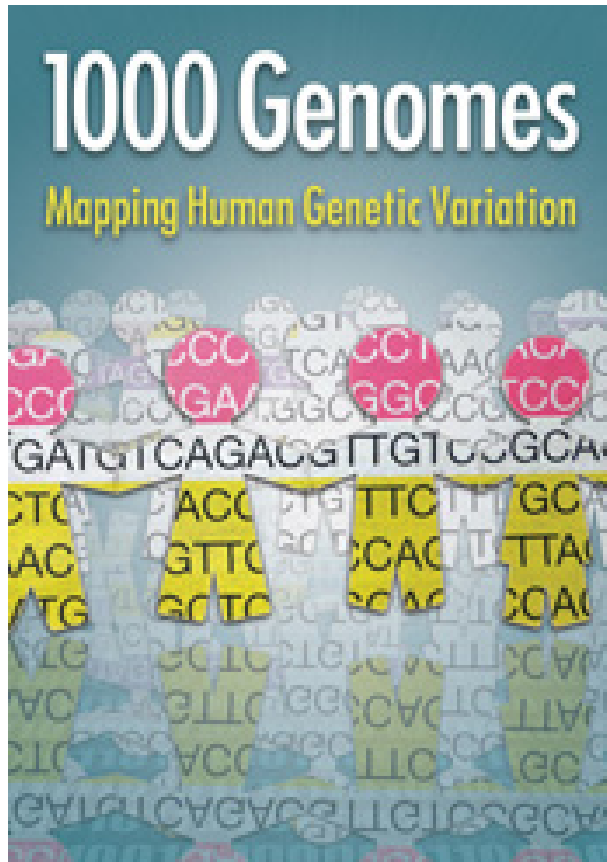




Variation in individual genomes

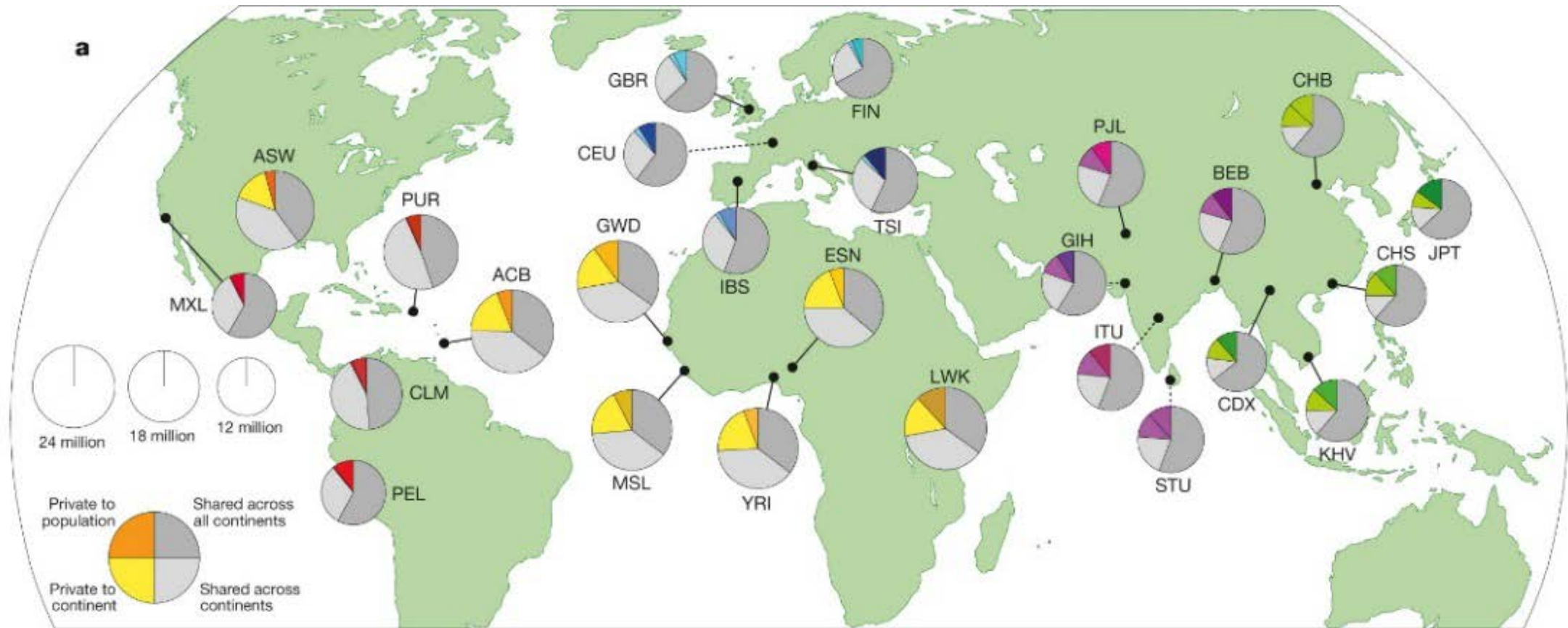


1000 genome project/resource

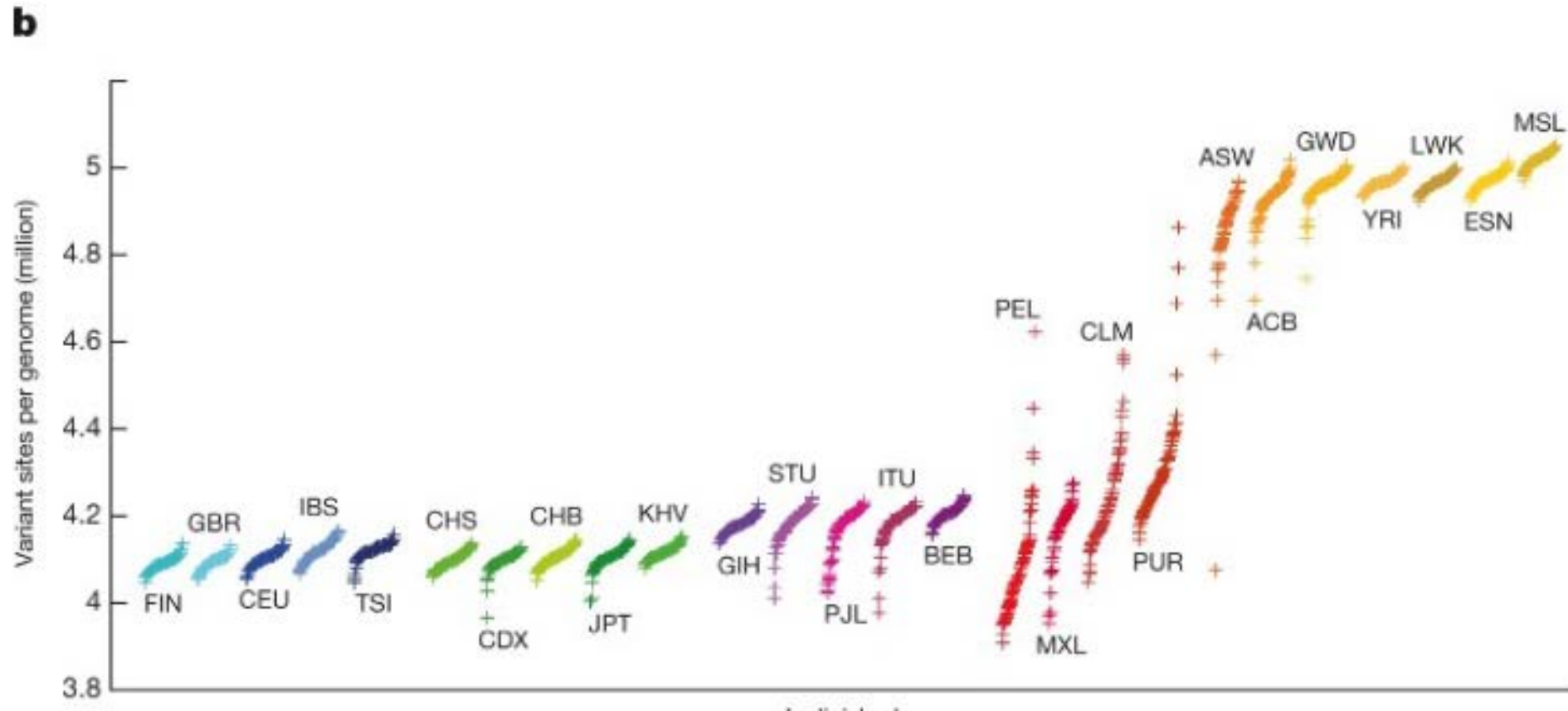


- Comprehensive description of common human genetic variation
- Latest report: genomes of 2,504 individuals from 26 populations using a combination of low-coverage whole-genome sequencing, deep exome sequencing, and dense microarray genotyping.
- Results:
 - over **88 million variants** (84.7 million 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



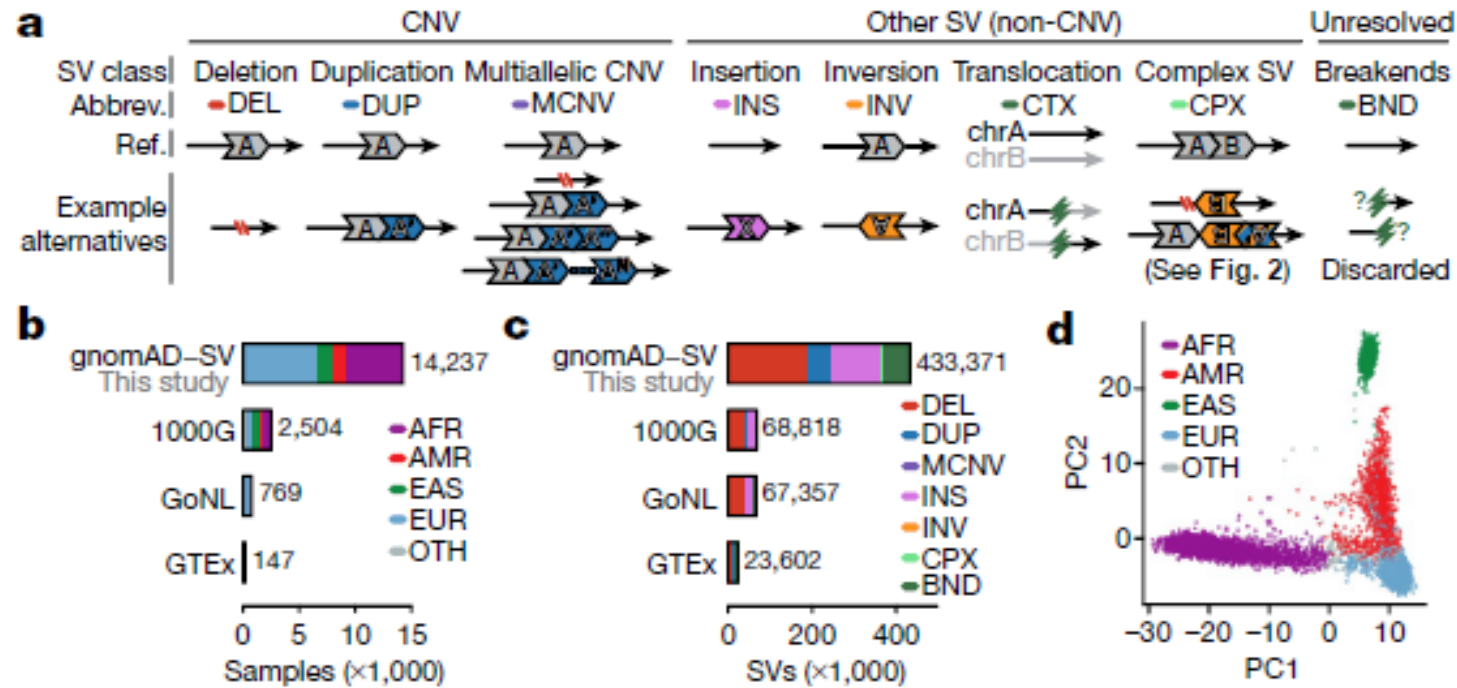
SNP variation/population



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

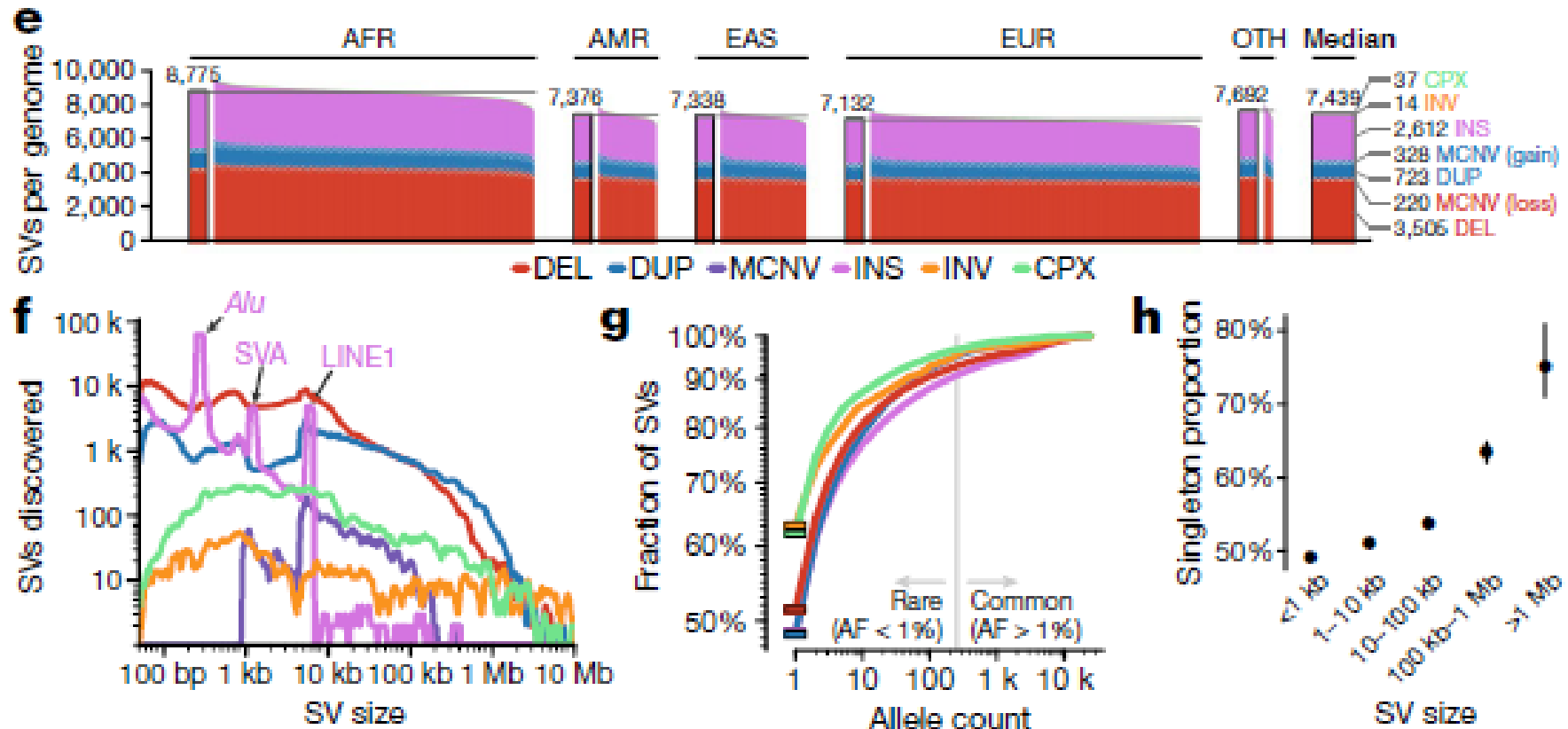
Published online: 27 May 2020

Open access

Check for updates

Ryan L. Collins^{1,2,3,10†}, Harrison Brand^{1,2,4,10†}, Konrad J. Karczewski^{1,5}, Xuefang Zhao^{1,2,4}, Jessica Alfoldi^{1,5}, Laurent C. Franciolli^{1,6,4}, Amit V. Khera^{1,5}, Chelsea Lowther^{1,2,4}, Laura D. Gauthier^{1,7}, Harold Wang^{1,7}, Nicholas A. Watts^{1,8}, Matthew Solomonson^{1,5}, Anne O'Donnell-Luria^{1,5}, Alexander Baumann¹, Ruchi Munshi¹, Mark Walker^{1,7}, Christopher W. Whelan¹, Yongqing Huang¹, Ted Brookings¹, Ted Sharpe¹, Matthew R. Stone^{1,9}, Elise Valkanas^{1,2,3}, Jack Fu^{1,2,4}, Grace Tiao^{1,5}, Kristen M. Laricchia^{1,5}, Valentin Ruano-Rubio¹, Christine Stevens¹, Namrata Gupta¹, Caroline Cusick¹, Lauren Margolin¹, Genome Aggregation Database Production Team⁴, Genome Aggregation Database Consortium⁴, Kent D. Taylor⁴, Henry J. Lin⁴, Stephen S. Rich⁴, Wendy S. Post¹⁰, Yil-Der Ida Chen⁴, Jerome I. Rotter⁴, Chad Nusbaum^{1,5,4}, Anthony Philippakis¹, Eric Lander^{1,10,11}, Stacey Gabriel¹, Benjamin M. Neale^{1,2,3,11}, Sekar Kathiresan^{1,2,4,11}, Mark J. Daly^{1,2,3,11}, Eric Banks¹, Daniel G. MacArthur^{1,2,3,4,5,6,9,10} & Michael E. Talkowski^{1,2,4,10} □

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 MWGS	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
Tau-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, MWGS, 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.[†]

Per genome variation

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

818000 indels (<50bp)

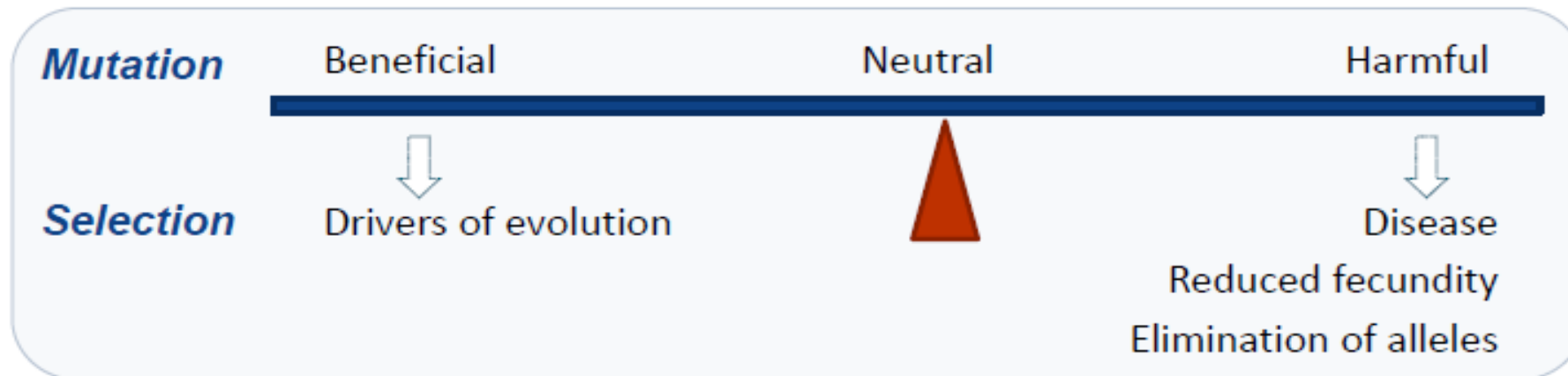
31599 structural variants (>50bp)

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

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

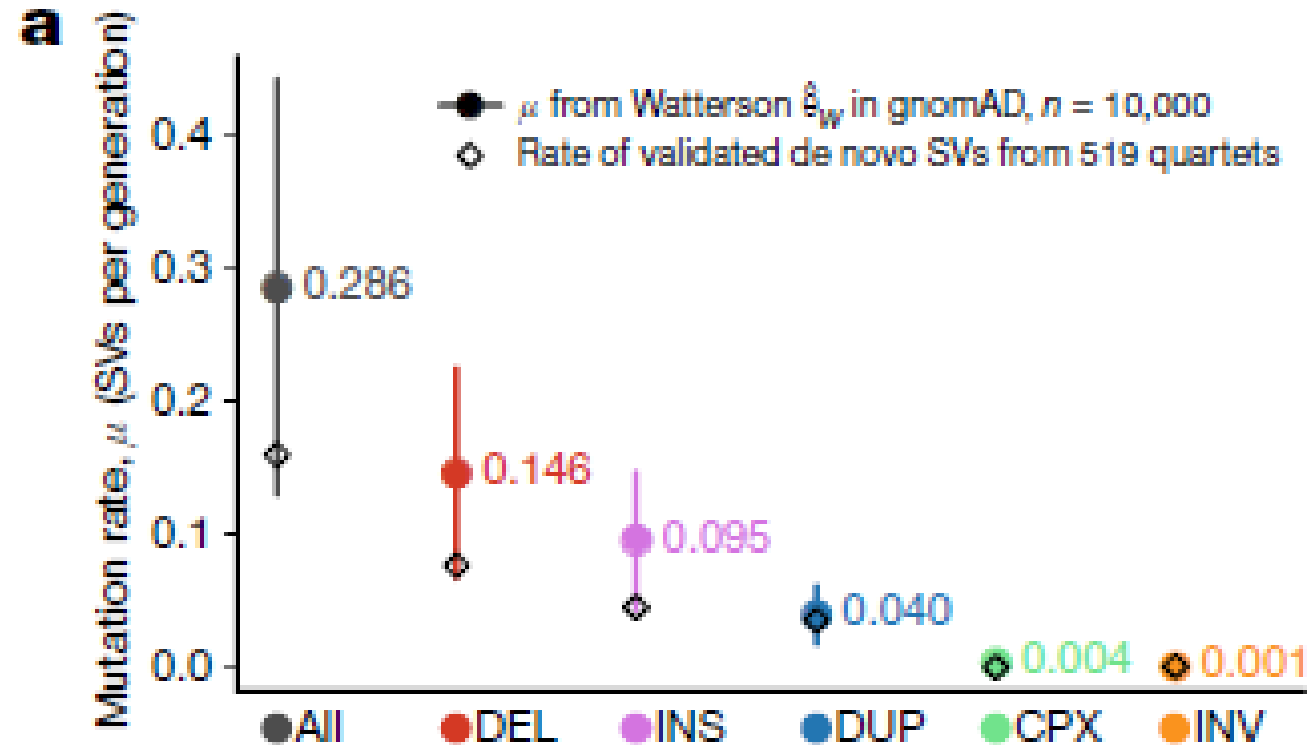
These mutations are under limited selective pressure!

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



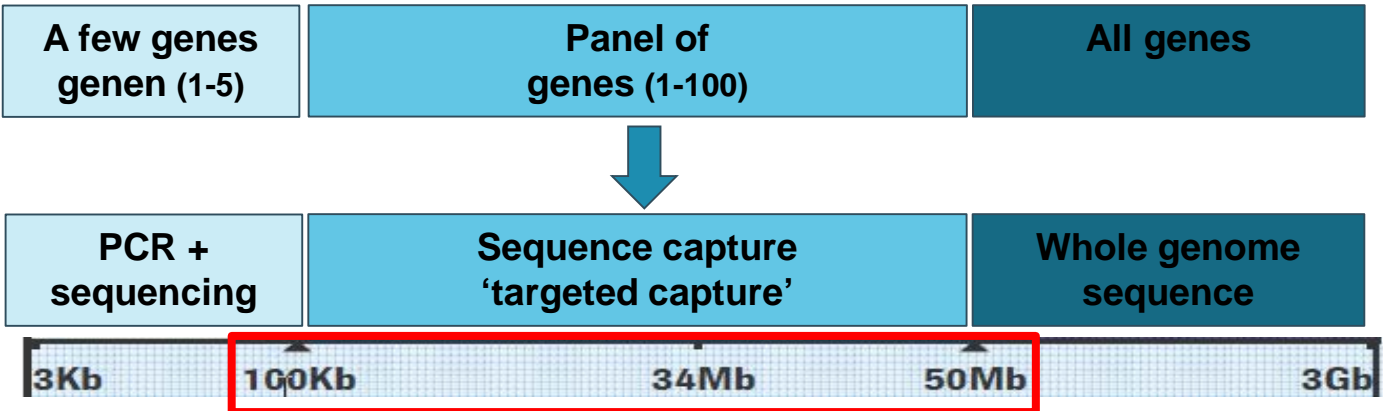
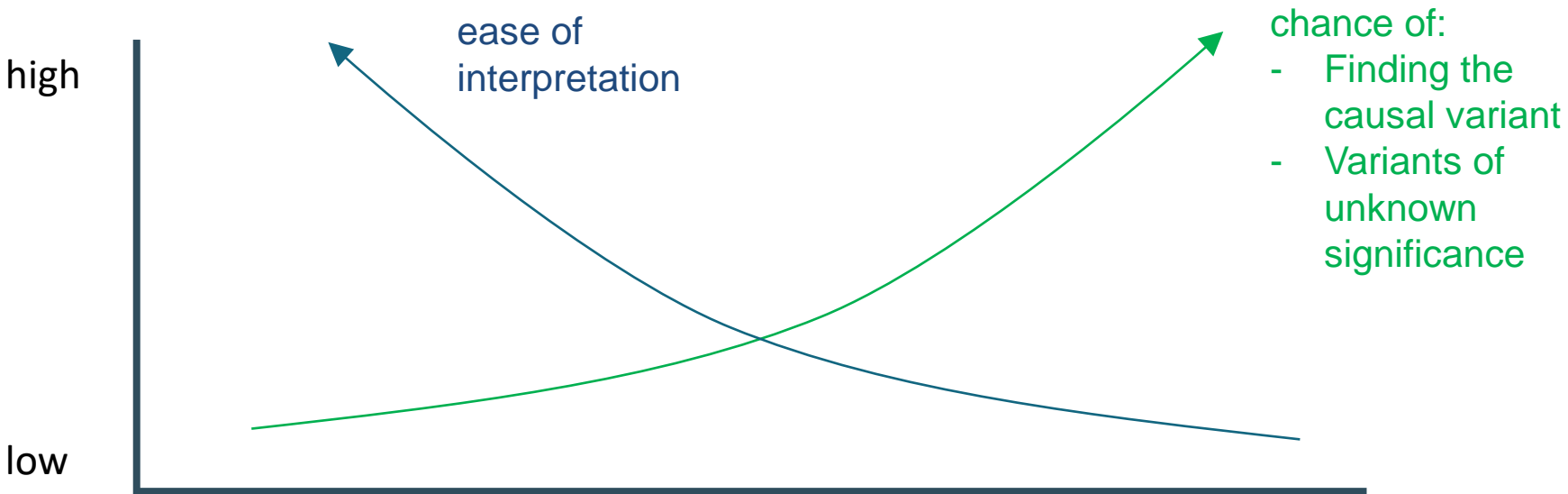
Overview

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

Molecular diagnostics WES/WGS



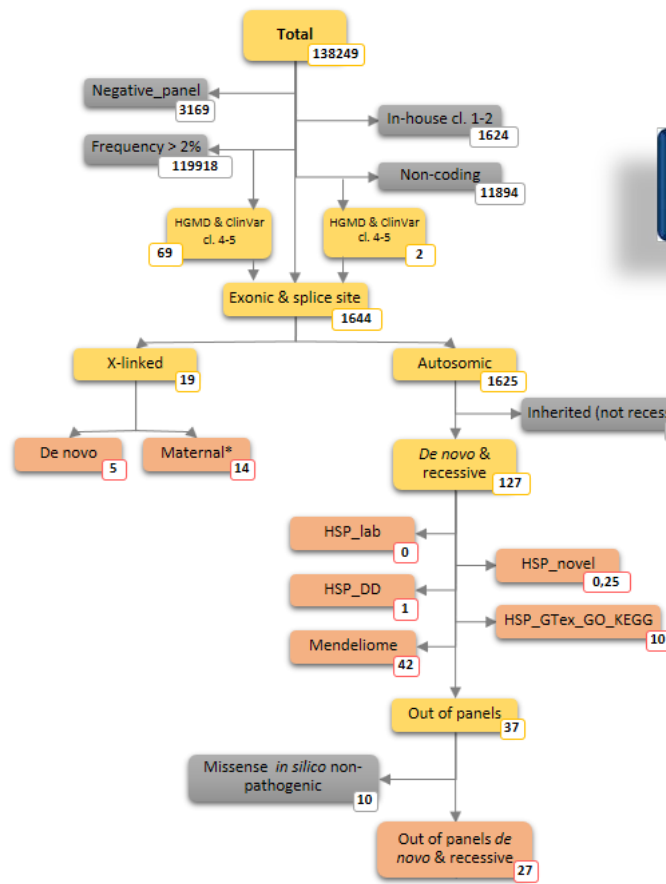
Θ Ζ Ε Χ Ε R T A N Q S T R A S C I A F A B R I C A I P S A I N P D I C T A E C C L I A E R I G I
S T A C I M E T O R D I N A M V O L E S A C D E G E R N E E S D I C T A F A B R I C A E S S E
D S E P A T A A T Q D S T I C T A A B O I A I A M S A S E V M A S S A Z O B V S A I I S T V R I B P R E
T E E C 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
M 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 D C V E S R O M A O S B O N E A C L A
B I L V I 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 Q B P N O S Z S V 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 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 Q 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 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 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 C A C O 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 V D I
P F E R V E T T 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 O I P O T 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 I L A
T O I 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 O S M O N E T E R O M A N V E I F O R V A I O R E A D M I N D O N A V T R I N

Interpretation of variants

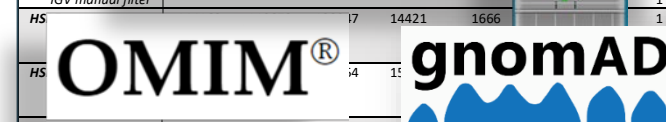
- Databases
 - SNP Databases > population frequencies
 - (internationale en lokale) mutation databases



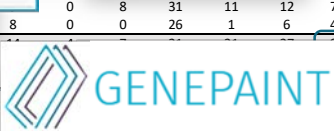
HSP: trio-based WES analysis of a cohort of 8 patients



Patient	AD filter	IG	recessiv	de novo	recessiv	de novo	Out of F	Total of
HSP1	7	12	22	20	71			
HSP2	3	11	5	11	32			
HSP3	8	22	28	19	87			
HSP4	2	20	2	3	34			
HSP7	8	33	19	33	120			
HSP8	1	33	11	11	83			
	0	31	2	3	52			
	0	23	22	28	94			



HSP7	138251	135430	133714	14
AD filter				6
IGV manual filter				6
HSP8	141432	138077	136365	13863
AD filter				0
IGV manual filter				8



Aims

Material and methods

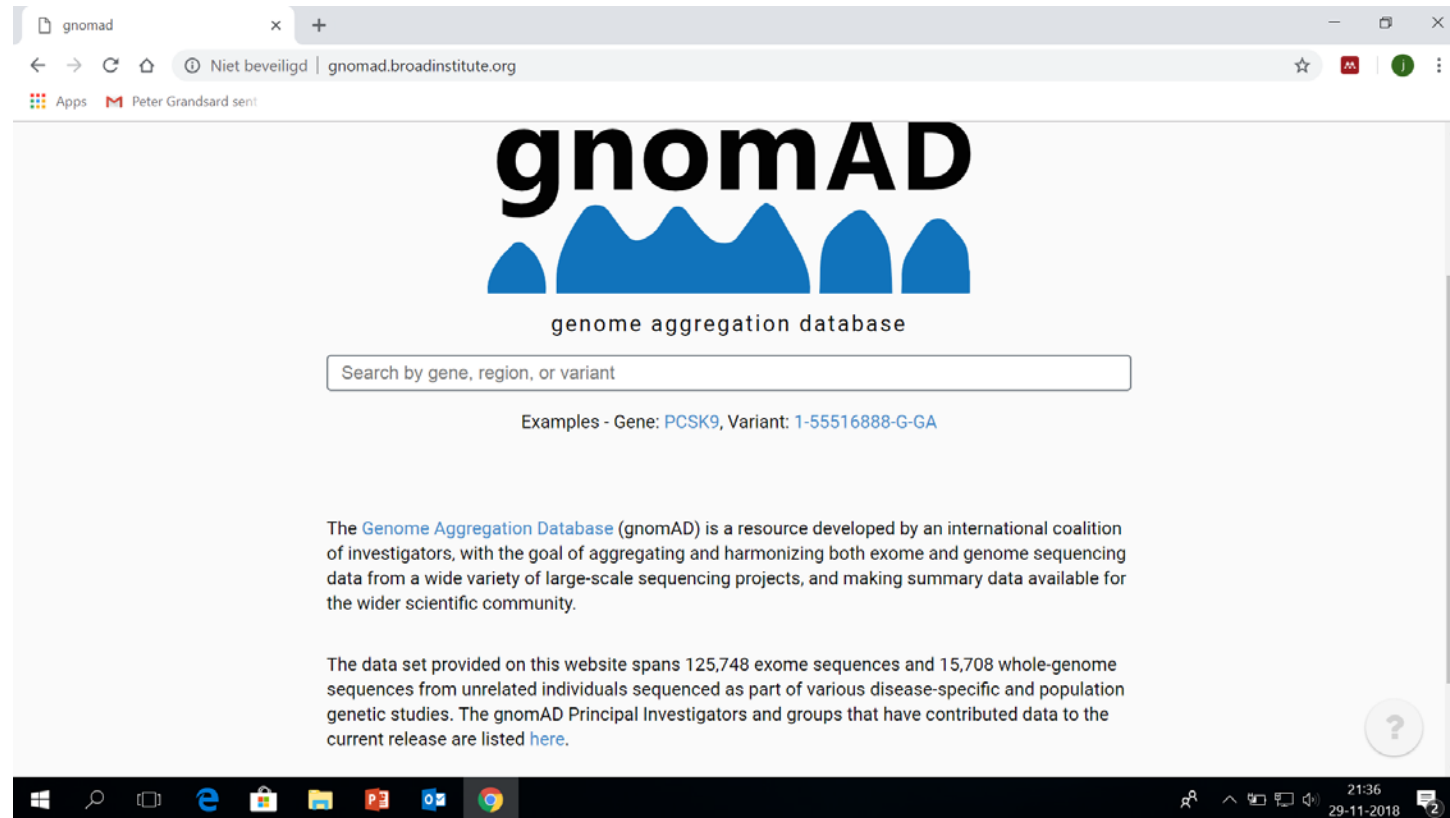
Added value of NGS in heterogeneous disorder

Added value of NGS in very rare disorders

Conclusions

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 describes the database as a resource developed by an international coalition of investigators, aiming to aggregate and harmonize exome and genome sequencing data. The second paragraph states that the data set spans 125,748 exome sequences and 15,708 whole-genome sequences from unrelated individuals. A help icon (a question mark in a circle) is visible in the bottom right corner of the page content.

gnomAD
genome aggregation database

Search by gene, region, or variant

Examples - Gene: [PCSK9](#), Variant: [1-55516888-G-GA](#)

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 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](#).

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.

Table:	Description:	Public entries: <small>This site. Academic non-profit users only</small>	Total entries: <small>HGMD Professional 2018.3</small>
Mutation totals (as of 2018-11-29)			
		157114	240269
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 interface. At the top, there is a search bar 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 browser's address bar shows "https://www.ncbi.nlm.nih.gov/clinvar/" and the Windows taskbar at the bottom indicates the date and time as 21:40 on 29-11-2018.

LQTS Gene LOVD Database



Tao Zhang^{1,2*}, Arthur Moss^{3,*}, Peikuan Cong^{2,*}, Min Pan^{2,*}, Bingxi Chang⁴, Liangrong Zheng⁵, Quan Fang⁴, Wojciech Zareba³, Jennifer Robinson³, Changsong Lin², Zhongxiang Li⁶, Junfang Wei⁷, Qiang Zeng⁸, Long QT International Registry Investigators, HVP-China Investigators, and Ming Qi^{1,2,9**}



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, PhD¹, Nazneen Aziz, PhD^{2,16}, Sherri Bale, PhD³, David Bick, MD⁴, Soma Das, PhD⁵, Julie Gastier-Foster, PhD^{6,7,8}, Wayne W. Grody, MD, PhD^{9,10,11}, Madhuri Hegde, PhD¹², Elaine Lyon, PhD¹³, Elaine Spector, PhD¹⁴, Karl Voelkerding, MD¹³ and Heidi L. Rehm, PhD¹⁵; on behalf of the ACMG Laboratory Quality Assurance Committee

Class of risk	Clinical significance
1	not pathogenic
2	likely not pathogenic
3	uncertain
4	likely pathogenic
5	definitely pathogenic

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 ± 1 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"> • Beware of genes where LOF is not a known disease mechanism (e.g., <i>GFAP</i>, <i>MYH7</i>) • Use caution interpreting LOF variants at the extreme 3' end of a gene • Use caution with splice variants that are predicted to lead to exon skipping but leave the remainder of the protein intact • Use caution in the presence of multiple transcripts
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>C or G>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 >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
Population Data	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>	
Computational And Predictive Data		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>
Functional Data	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>	
Segregation Data	Non-segregation with disease <i>BS4</i>		Co-segregation with disease in multiple affected family members <i>PP1</i>	Increased segregation data →		
De novo Data				<i>De novo</i> (without paternity & maternity confirmed) <i>PM6</i>	<i>De novo</i> (paternity & maternity confirmed) <i>PS2</i>	
Allelic Data		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>		
Other Database		Reputable source w/out shared data = benign <i>BP6</i>	Reputable source = pathogenic <i>PP5</i>			
Other Data		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
Population Data	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>	
Computational And Predictive Data		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>
Functional Data	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>	
Segregation Data	Non-segregation with disease <i>BS4</i>		Co-segregation with disease in multiple affected family members <i>PP1</i>	Increased segregation data →		
De novo Data				<i>De novo</i> (without paternity & maternity confirmed) <i>PM6</i>	<i>De novo</i> (paternity & maternity confirmed) <i>PS2</i>	
Allelic Data		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>		
Other Database		Reputable source w/out shared data = benign <i>BP6</i>	Reputable source = pathogenic <i>PP5</i>			
Other Data		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
Population Data	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>	
Computational And Predictive Data		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>
Functional Data	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>	
Segregation Data	Non-segregation with disease <i>BS4</i>		Co-segregation with disease in multiple affected family members <i>PP1</i>	Increased segregation data →		
De novo Data				<i>De novo</i> (without paternity & maternity confirmed) <i>PM6</i>	<i>De novo</i> (paternity & maternity confirmed) <i>PS2</i>	
Allelic Data		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>		
Other Database		Reputable source w/out shared data = benign <i>BP6</i>	Reputable source = pathogenic <i>PP5</i>			
Other Data		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
Population Data	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>	
Computational And Predictive Data		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>
Functional Data	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>	
Segregation Data	Non-segregation with disease <i>BS4</i>		Co-segregation with disease in multiple affected family members <i>PP1</i>	Increased segregation data →		
De novo Data				<i>De novo</i> (without paternity & maternity confirmed) <i>PM6</i>	<i>De novo</i> (paternity & maternity confirmed) <i>PS2</i>	
Allelic Data		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>		
Other Database		Reputable source w/out shared data = benign <i>BP6</i>	Reputable source = pathogenic <i>PP5</i>			
Other Data		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
Population Data	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>	
Computational And Predictive Data		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>
Functional Data	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>	
Segregation Data	Non-segregation with disease <i>BS4</i>		Co-segregation with disease in multiple affected family members <i>PP1</i>	Increased segregation data →		
De novo Data				<i>De novo</i> (without paternity & maternity confirmed) <i>PM6</i>	<i>De novo</i> (paternity & maternity confirmed) <i>PS2</i>	
Allelic Data		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>		
Other Database		Reputable source w/out shared data = benign <i>BP6</i>	Reputable source = pathogenic <i>PP5</i>			
Other Data		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"> (i) 1 Very strong (PVS1) <i>AND</i> <li style="padding-left: 20px;">(a) ≥ 1 Strong (PS1–PS4) <i>OR</i> <li style="padding-left: 20px;">(b) ≥ 2 Moderate (PM1–PM6) <i>OR</i> <li style="padding-left: 20px;">(c) 1 Moderate (PM1–PM6) and 1 supporting (PP1–PP5) <i>OR</i> <li style="padding-left: 20px;">(d) ≥ 2 Supporting (PP1–PP5) (ii) ≥ 2 Strong (PS1–PS4) <i>OR</i> (iii) 1 Strong (PS1–PS4) <i>AND</i> <li style="padding-left: 20px;">(a) ≥ 3 Moderate (PM1–PM6) <i>OR</i> <li style="padding-left: 20px;">(b) 2 Moderate (PM1–PM6) <i>AND</i> ≥ 2 Supporting (PP1–PP5) <i>OR</i> <li style="padding-left: 20px;">(c) 1 Moderate (PM1–PM6) <i>AND</i> ≥ 4 supporting (PP1–PP5)
Likely pathogenic	<ul style="list-style-type: none"> (i) 1 Very strong (PVS1) <i>AND</i> 1 moderate (PM1–PM6) <i>OR</i> (ii) 1 Strong (PS1–PS4) <i>AND</i> 1–2 moderate (PM1–PM6) <i>OR</i> (iii) 1 Strong (PS1–PS4) <i>AND</i> ≥ 2 supporting (PP1–PP5) <i>OR</i> (iv) ≥ 3 Moderate (PM1–PM6) <i>OR</i> (v) 2 Moderate (PM1–PM6) <i>AND</i> ≥ 2 supporting (PP1–PP5) <i>OR</i> (vi) 1 Moderate (PM1–PM6) <i>AND</i> ≥ 4 supporting (PP1–PP5)
Benign	<ul style="list-style-type: none"> (i) 1 Stand-alone (BA1) <i>OR</i> (ii) ≥ 2 Strong (BS1–BS4)
Likely benign	<ul style="list-style-type: none"> (i) 1 Strong (BS1–BS4) and 1 supporting (BP1–BP7) <i>OR</i> (ii) ≥ 2 Supporting (BP1–BP7)
Uncertain significance	<ul style="list-style-type: none"> (i) Other criteria shown above are not met <i>OR</i> (ii) the criteria for benign and pathogenic are contradictory

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.

LOOK

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.

Allele Information

Allele Registry ID

<http://reg.genome.network/allele/CA021883>

RGV

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 [\[Link\]](#)

PVS3

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 β -galactosidase A activity assay was 10.47±11.2% of normal in the men and 48.6±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 :

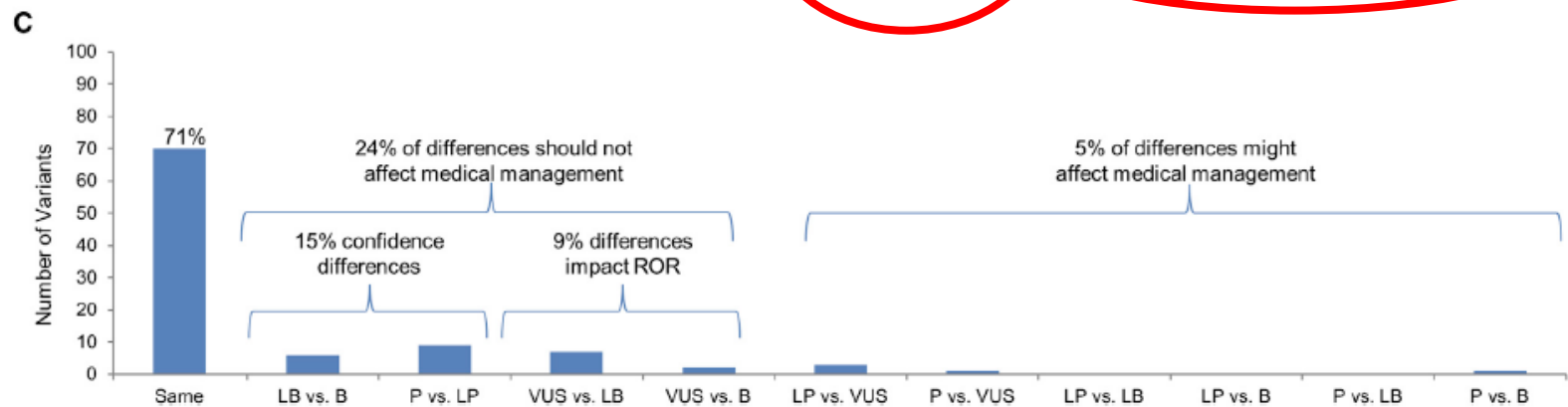
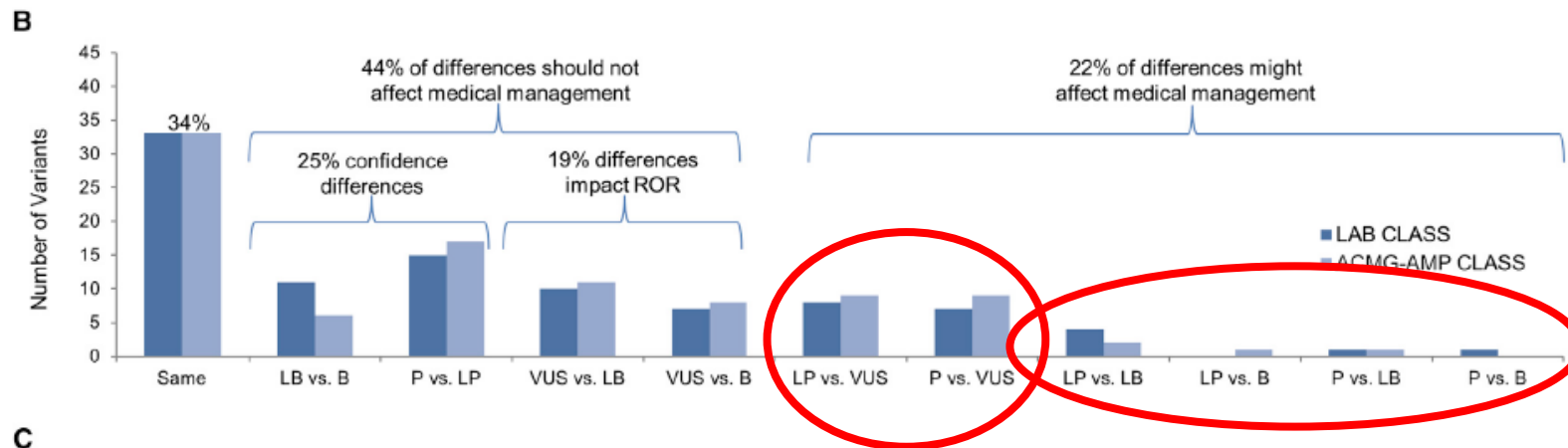
- Pathogenic.Strong >=2

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

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

Performance of ACMG-AMP Variant-Interpretation Guidelines among Nine Laboratories in the Clinical Sequencing Exploratory Research Consortium

Laura M. Amendola,^{1,16} Gail P. Jarvik,^{1,16,*} Michael C. Leo,² Heather M. McLaughlin,³ Yasmine Akkari,⁴ Michelle D. Amaral,⁵ Jonathan S. Berg,⁶ Sawona Biswas,⁷ Kevin M. Bowling,⁵ Laura K. Conlin,⁷ Greg M. Cooper,⁵ Michael O. Dorschner,⁸ Matthew C. Dulik,⁹ Arezou A. Ghazani,¹⁰ Rajarshi Ghosh,¹¹ Robert C. Green,^{3,12,15} Ragan Hart,¹ Carrie Horton,¹³ Jennifer J. Johnston,¹⁴ Matthew S. Lebo,^{3,12} Aleksandar Milosavljevic,¹¹ Jeffrey Ou,¹ Christine M. Pak,⁴ Ronak Y. Patel,¹¹ Sumit Punj,⁴ Carolyn Sue Richards,⁴ Joseph Salama,¹ Natasha T. Strande,⁶ Yaping Yang,¹¹ Sharon E. Plon,¹¹ Leslie G. Biesecker,¹⁴ and Heidi L. Rehm^{3,12,15,*}



(B) Inter-laboratory concordance of 97 variants. This graph compares the same calls, based on either the ACMG-AMP rules or the site's rules, between laboratories.

(C) Inter-laboratory concordance after consensus efforts. This graph shows a final comparison of calls between sites after consensus-building efforts.

Table 1. ACMG-AMP Rule Clarifications and Suggestions for Modification

Rule	Description	Clarifications and/or Suggestions
PVS1	variant predicted null where LOF is a mechanism of disease	do not apply to variants that are near the 3' end of the gene and escape nonsense-mediated decay
PS1	variant with the same amino acid change as a previously established pathogenic variant, regardless of nucleotide change	does not include the same variant being assessed because it is not yet pathogenic, and the rule is intended for variants with a different nucleotide change
PS2	de novo variant with confirmed maternity and paternity	apply this rule as moderate or supporting if the variant is mosaic and its frequency in tissue is consistent with the phenotype
PS3	variant shown to have a deleterious effect by a well-established functional study	reduce the strength for assays that are not as well validated or linked to the phenotype
PM1	variant located in a mutational hotspot and/or critical and well-established functional domain	not meant for truncations; more clarification is needed for applying this rule
PM2, BS1	variant absent in population databases or with an allele frequency too high for the disease	cannot assume longer indels would be detected by next-generation sequencing use a published control dataset if its size is at least 1,000 individuals cannot be applied for low-quality calls or non-covered regions must define the condition and inheritance pattern
PM3	for recessive disorders, variant in <i>trans</i> with a pathogenic variant	invoke this rule as supporting if the phase is not established can upgrade if more than one proband is reported
PM4	protein-length-changing variant	applicable for in-frame deletions, insertions, or stop-loss variants, but not frameshifts, nonsense, and splice variants
PM5	novel missense variant at amino acid with different pathogenic missense change	ensure pathogenicity of previously reported variant suggest changing "novel" to "different" because some variants that are not novel might require assessment with this rule
PP3, BP4	variant with multiple lines of computational evidence	all lines must agree
PP4	the patient's phenotype or family history is highly specific to the genotype	not meant to be used for genetically heterogeneous conditions or conditions with unsolved etiology not typically applied for an analysis of incidental findings, but it could be applied for prior observations
PP5, BP6	variant called pathogenic or benign by a reputable source	only applicable when evidence is not available (e.g., Sharing Clinical Reports Project)
BS2	variant observed in a healthy adult for a disorder with full penetrance at an early age	populations might not have been screened or excluded for the phenotype
BP1	variant in a gene in which truncations primarily cause disease	clarify the meaning of "primary"; suggest >90%
BP2, BP5	variant in <i>trans</i> with a dominant pathogenic variant (BP2) or in an individual with an alternate molecular basis for disease (BP5)	clarify that one should apply BP2 when the pathogenic variant is seen in the same gene as the variant being evaluated and apply BP5 when the pathogenic variant is in a different gene

Clinical laboratories collaborate to resolve differences in variant interpretations submitted to ClinVar

Steven M. Harrison, PhD^{1,2}, Jill S. Dolinsky, MS³, Amy E. Knight Johnson, MS⁴,
Tina Pesaran, MA, MS³, Danielle R. Azzariti, MS¹, Sherri Bale, PhD⁵, Elizabeth C. Chao, MD^{3,6},
Soma Das, PhD⁴, Lisa Vincent, PhD⁵ and Heidi L. Rehm, PhD^{1,2,7,8}

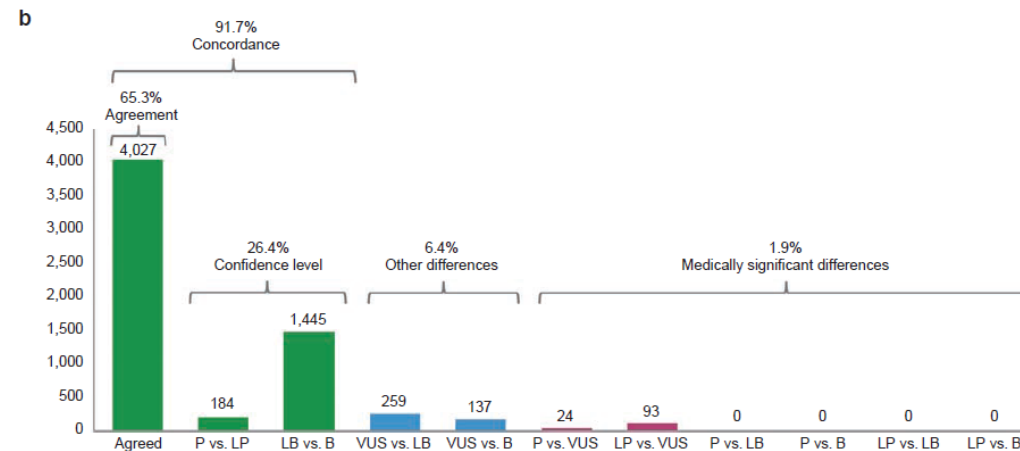
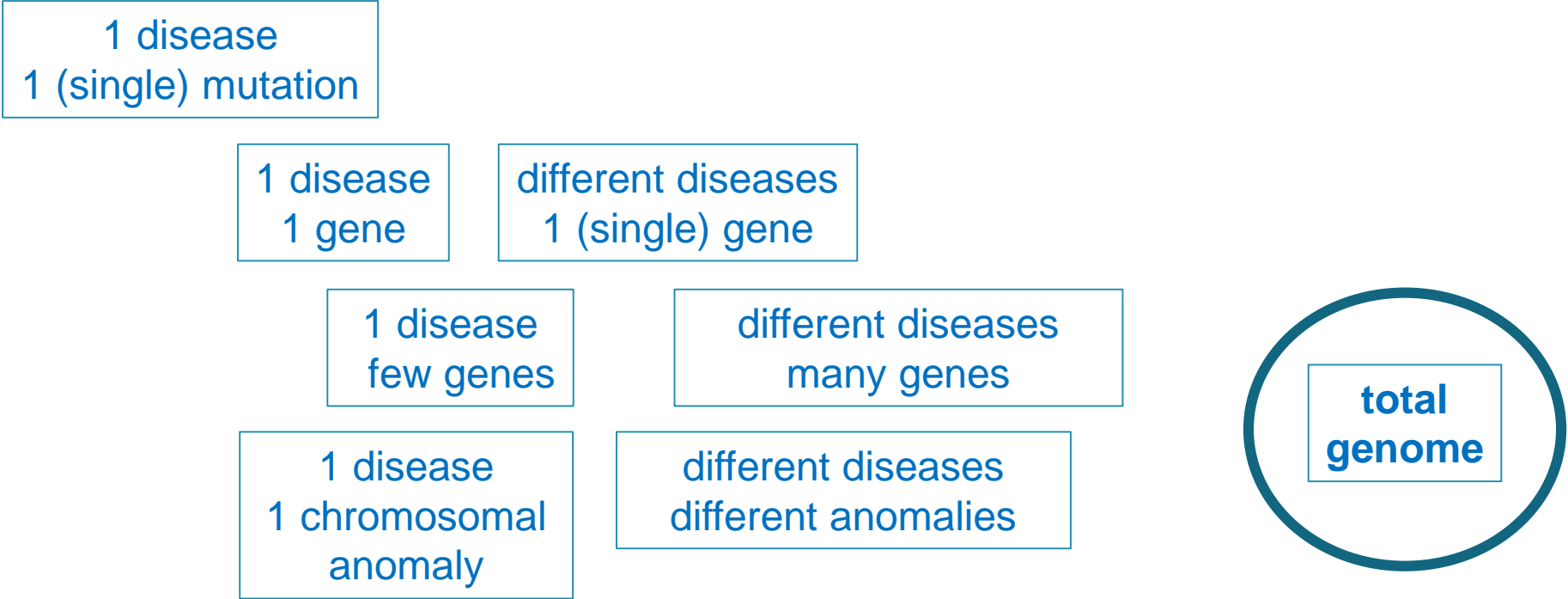


Figure 1 Distribution of variant interpretation differences between four clinical laboratories. (a) Interpretation comparison of data in ClinVar (as of January 1, 2016) before resolution efforts. (b) Interpretation comparison after reassessing 33% (242/724) of shared variants with interpretation differences.

Genetic testing

Molecular testing

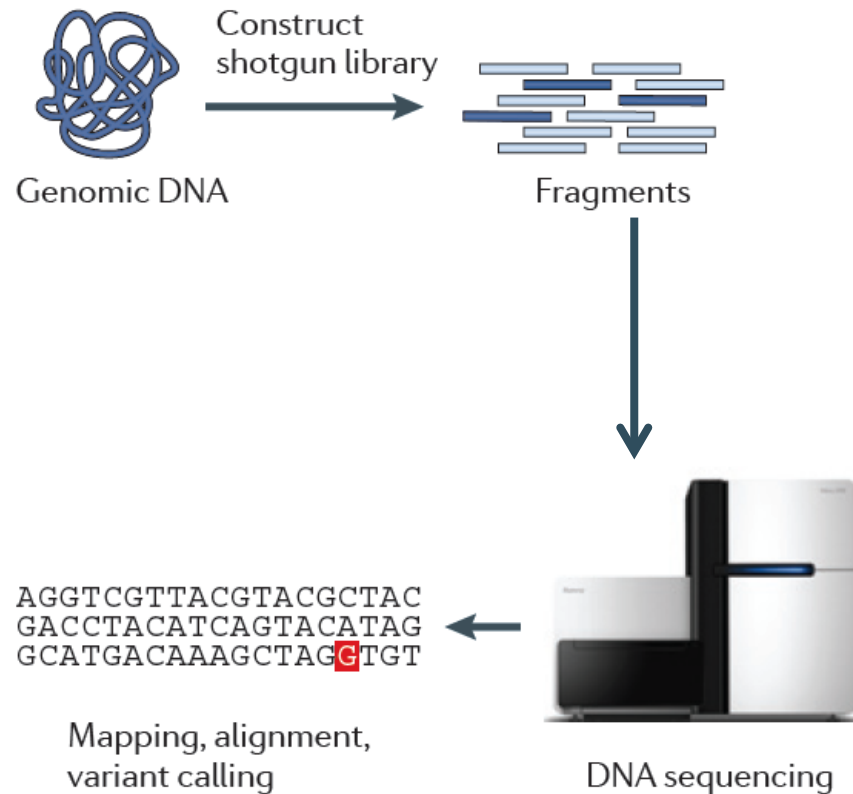


(un)known defect
karyotyping

Molecular cytogenetic testing

Cytogenetics

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 wall
- 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 picket fence

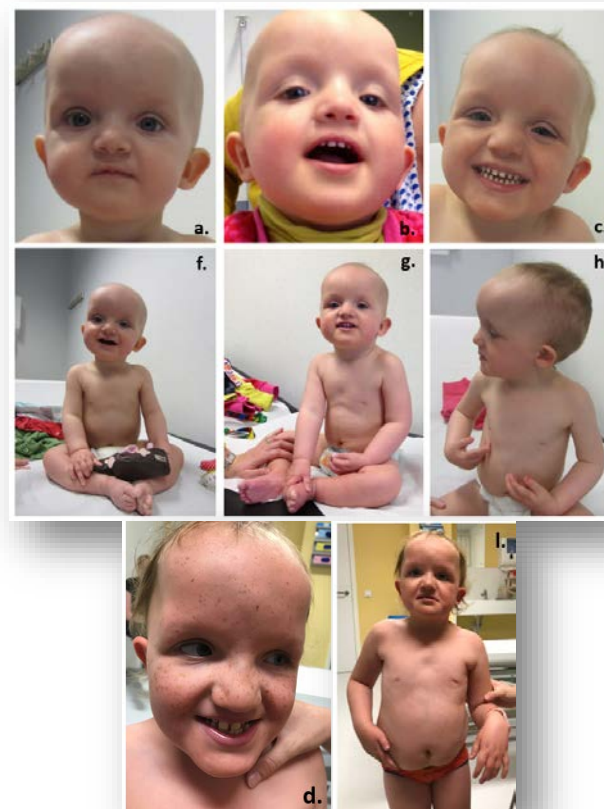
• Very

• Mild

- 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
de novo	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
		58

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



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
6p22.2	Rahman syndrome	617537	AD



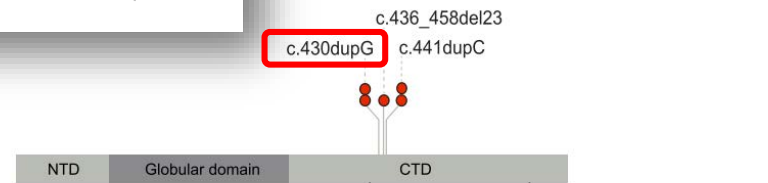
HGMD accession	Reported disease/phenotype	Variant class	Gene symbol	Inheritance
CI176502	Intellectual disability with overgrowth		HIST1H1E	GCGACG*GG
Literature citation		Citation type	Support	
1. Tatton-Brown (2017) <i>Am J Hum Genet</i> 100 : 725. PubMed: 28475857 Mutations in Epigenetic Regulation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

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

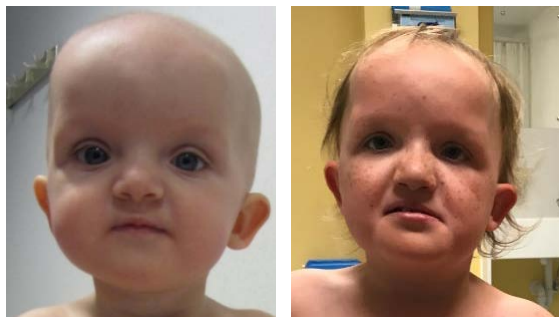
ARTICLE

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

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



c.430dupG p.Ala144Glyfs*52



Frame	Mutation	Sequence	Length	Charge
1	Wildtype	...KKATGAATPKSAKKTPKKAKKPAAGAKKAKSPKKAKAAKPKKAPKSPAKAKAVKPK...	219	44
2	c.430dupG	...KKATGGGHPQEERQEDPKEGEEAGCCSCWSQKSEKPEKGESSQAKKGAQEPSEGQSS	194	7
2	c.441dupC	...KKATGAATPQEERQEDPKEGEEAGCCSCWSQKSEKPEKGESSQAKKGAQEPSEGQSS	194	7
2	c.436_458del23	...KKATGAA-----DPKEGEEAGCCSCWSQKSEKPEKGESSQAKKGAQEPSEGQSS	186	9
3		...KKATGRPPPRRAPRRPQRRRRSRLQLLEPKKRKARKRRKQPSQKRRPRAQRRPKQLNPR...	227	48

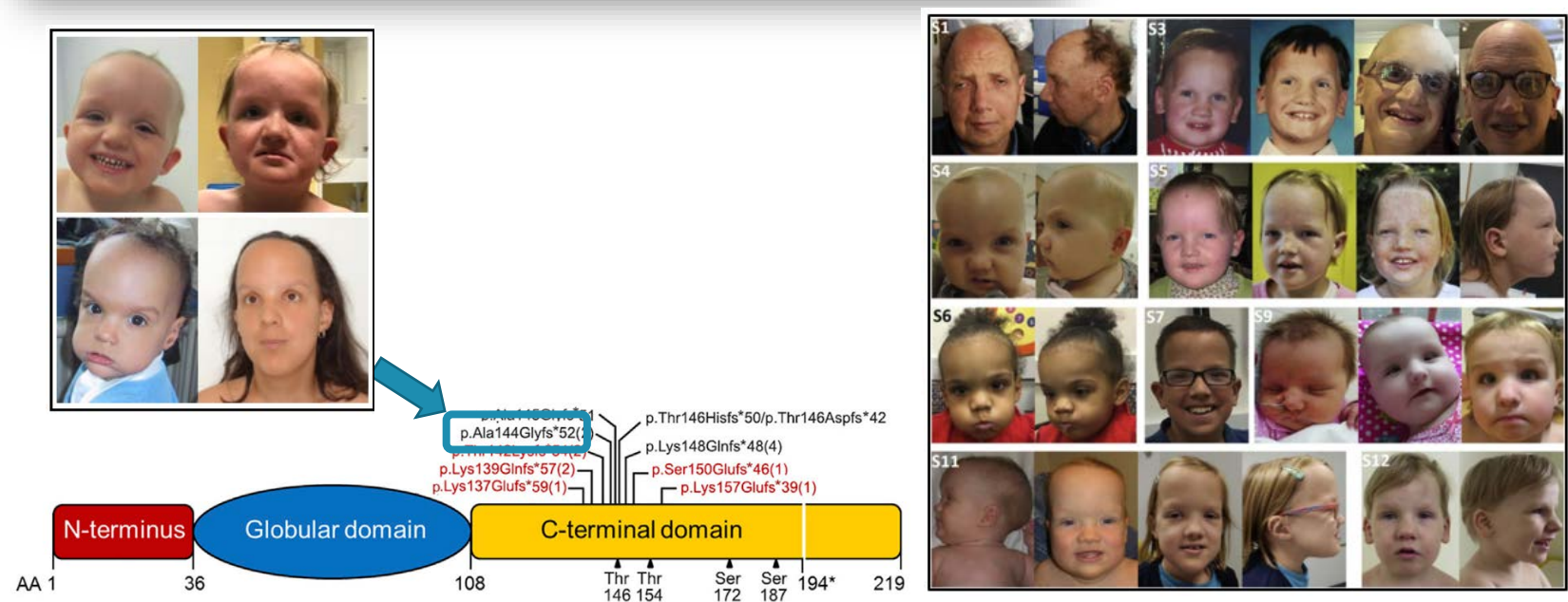


MCA/ID: trio-based WGS analysis of 3 patients with unknown clinical diagnosis

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 E¹, Martinelli S², Van Dijk A³, Ciolfi A⁴, Cecchetti S⁵, Coluzzi E⁶, Pannone L⁷, Andreoli C⁸, Radio FC⁴, Pizzi S⁴, Carpentieri G⁷, Bruselles A², Catanzaro G⁹, Pedace L¹⁰, Miele E¹⁰, Carcarino E¹¹, Ge X¹², Chijiwa C¹³, Lewis MES¹³, Meuwissen M¹⁴, Kenis S¹⁵, Van der Aa N¹⁴, Larson A¹⁶, Brown K¹⁶, Wasserstein MP¹⁷, Skotko BG¹⁸, Begtrup A¹⁹, Person R¹⁹, Karayiorgou M²⁰, Roos JL²¹, Van Gassen KL²², Koopmans M²², Bijlsma EK²³, Santen GWE²³, Barge-Schaapveld DQCM²³, Ruivenkamp CAL²³, Hoffer MJV²³, Lalani SR²⁴, Streff H²⁴, Craigen WJ²⁴, Graham BH²⁵, van den Elzen APM²⁶, Kamphuis D²⁷, Ünnap K²⁸, Reinson K²⁸, Pajusalu S²⁹, Wojcik MH³⁰, Viberti C³¹, Di Gaetano C³¹, Bertini E⁴, Petrucci S³², De Luca A³³, Rota R¹⁰, Ferretti E³⁴, Matullo G³¹, Dallapiccola B⁴, Sgura A⁶, Walkiewicz M³⁵, Kooy RF³⁶, Tartaglia M³⁷.



seqnames	start	end	width	change_ty	QUAL_lun	GC069450	GC069451	GC069452	Combined_genotype	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
chr19																			
chr19																			
chr2																			
chr5	70508001	70512000	4000	DEL	NOT_LUM	1/1	0/1	0/1	1/1;0/1;0/1	SMAS5	.	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

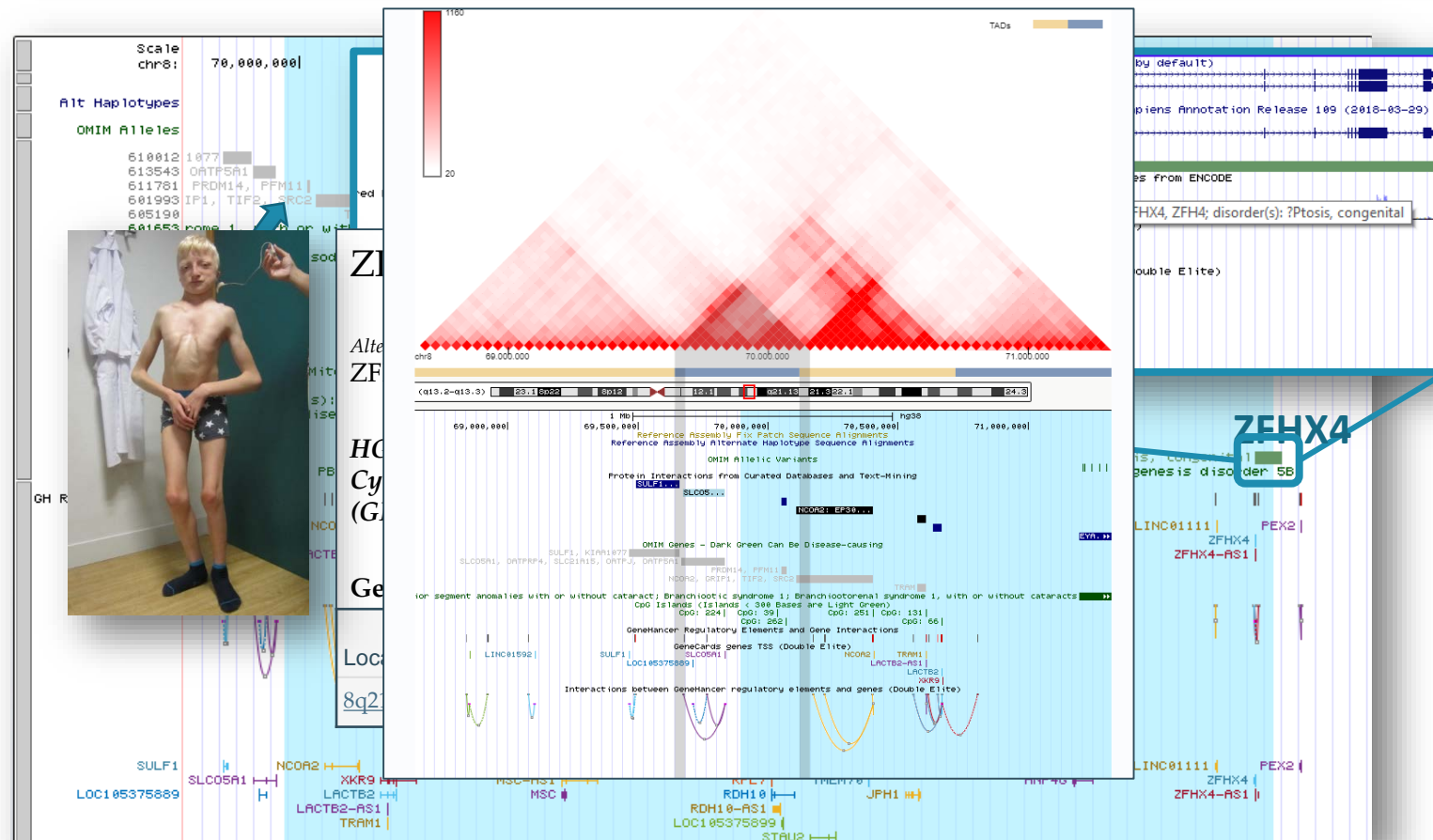


		Number of variants		
Total		20414		
Filtering criteria Good quality and rare > 10 bp De novo and recessive containing gene and/or regulatory regions IGV manual filtering	DEL/DUP/INV	737	3795	2827
	INS	*	165	*
	BND	44	20	40
		3	4	0

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	32637001	32643000	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										OR4F21	.	OR4F21	7	NA	NA	NA					
chr15	10					0/1	0/0	0/0	0/1;0/0;0/0	OR4F4	.	OR4F4	1	NA	NA	NA					
chr14	7					0/1	0/0	0/0	0/1;0/0;0/0	ACOT1,HEATR4	.	ACOT1,HEATR4	12	NA	NA	NA				ACOT1	
chr8	6																			FA1	
chr19	5																			2DL3	
chr22	2																			GGT2	
chr19	54829001	54845000	16000	DEL	NOT_LUM	0/1	0/0	0/0	0/1;0/0;0/0											2DS4	
chr17	36175																				KIR3DL1
chr8	72960																				TBC1L1
chr11	19450																				PL23
chr3	130082																				LD3F
chr17	36390																				
chr7	77009																				
chr6	32524																				DRB5
chr15	34387																				GOLC
chr21	13598																				TED
chr1	143541																				SRGA
chr2	87709																				
chr3	46754																				
chr11	1964501	1978500	14000	DEL	NOT_LUM	0/1	0/0	0/0	0/1;0/0;0/0												
chr15	84362001	84383500	21500	DEL	NOT_LUM	0/1	0/0	0/0	0/1;0/0;0/0	PI4KAP2,TMEM19	.	PI4KAP2,TMEM19	10	NA	NA	NA					
chr19	27399501	27633500	234000	DEL	NOT_LUM	0/1	0/0	0/0	0/1;0/0;0/0	PMS2P5,STAG3L2	.	PMS2P5,STAG3L2	2	NA	NA	NA					
chr22	21459501	21511500	52000	DEL	NOT_LUM	0/1	0/0	0/0	0/1;0/0;0/0	POM121L9P	POU121L5	0	NA	NA	NA						
chr7	74885501	74894000	8500	DEL	NOT_LUM	0/1	0/0	0/0	0/1;0/0;0/0												
chr22	24251001	24252500	1500	DEL	NOT_LUM	0/1	0/0	0/0	0/1;0/0;0/0												
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	SMAS5	.	SMAS5	1	NA	NA	NA					

Inversion
chr18:69893659-76806725
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 M¹, 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:

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



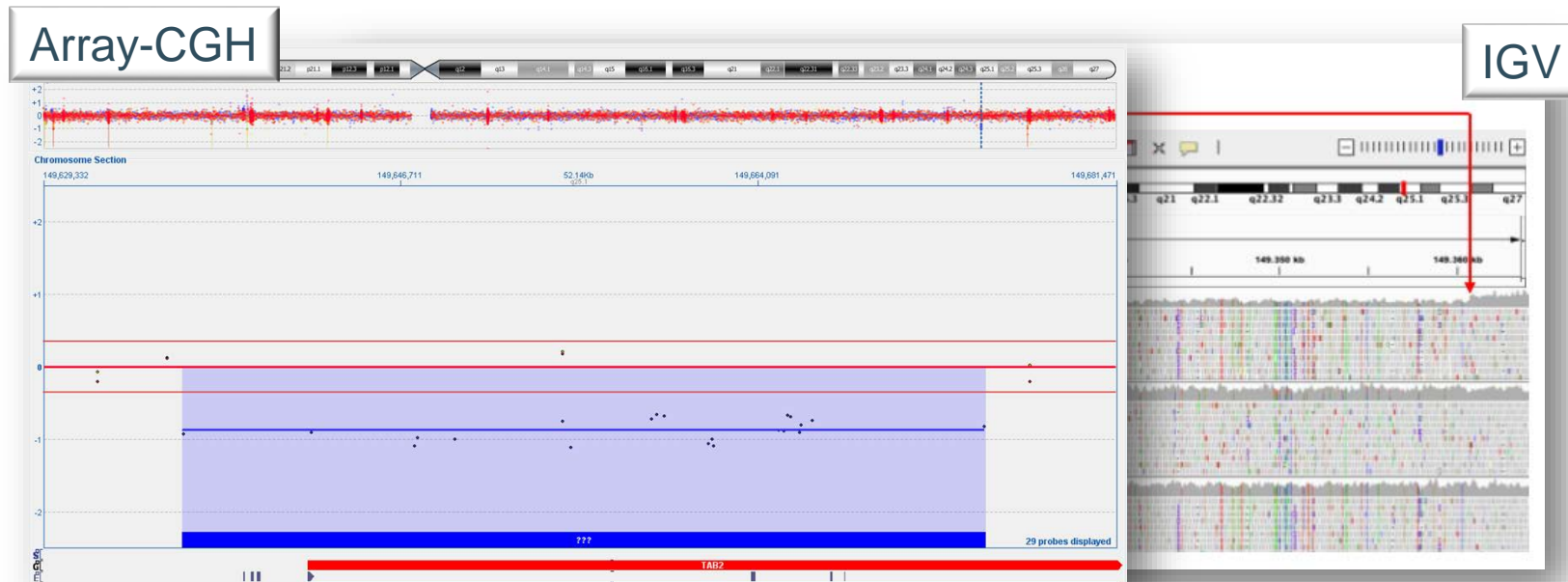
Negative

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

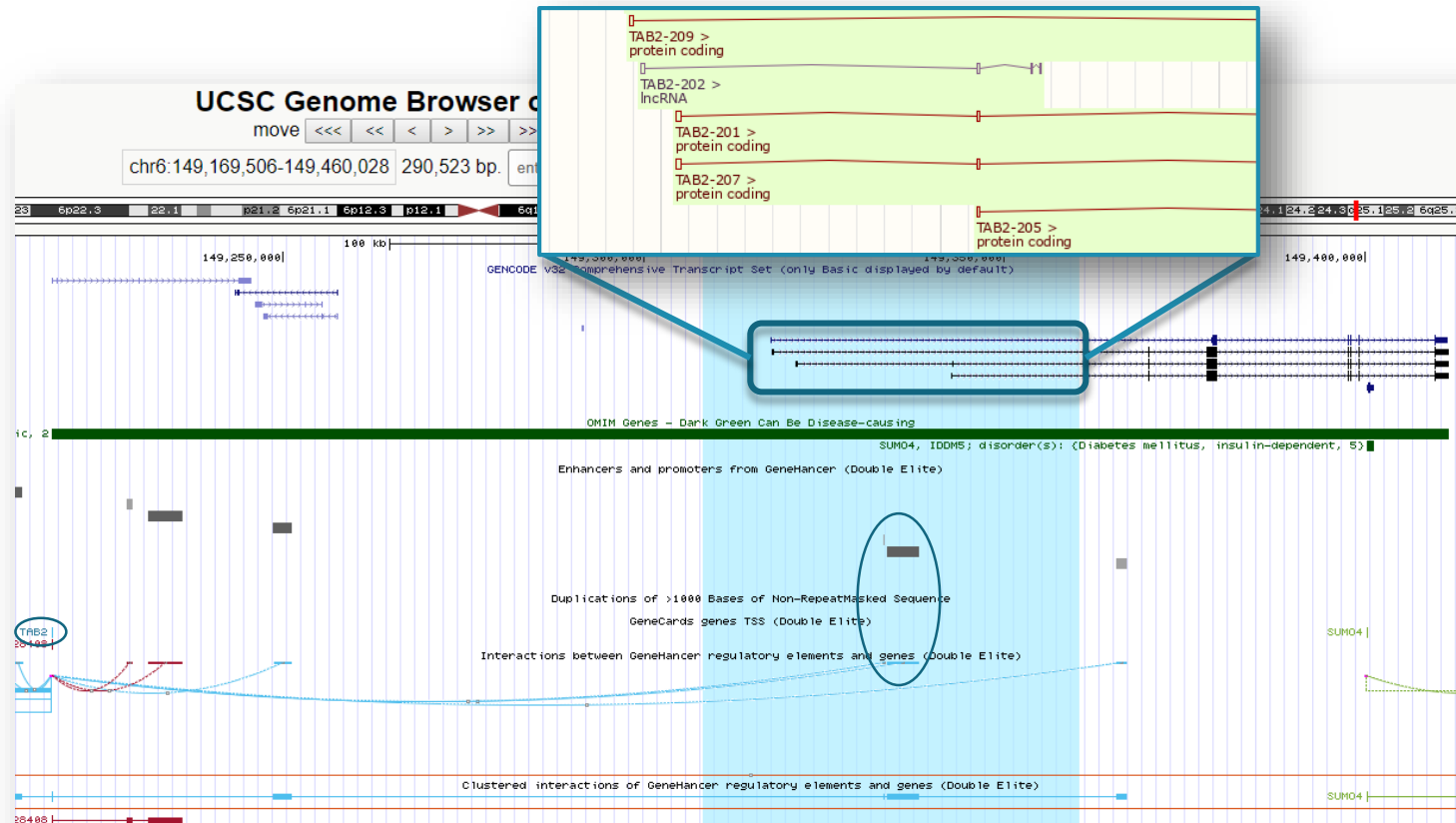
CNVs Analysis:

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

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



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

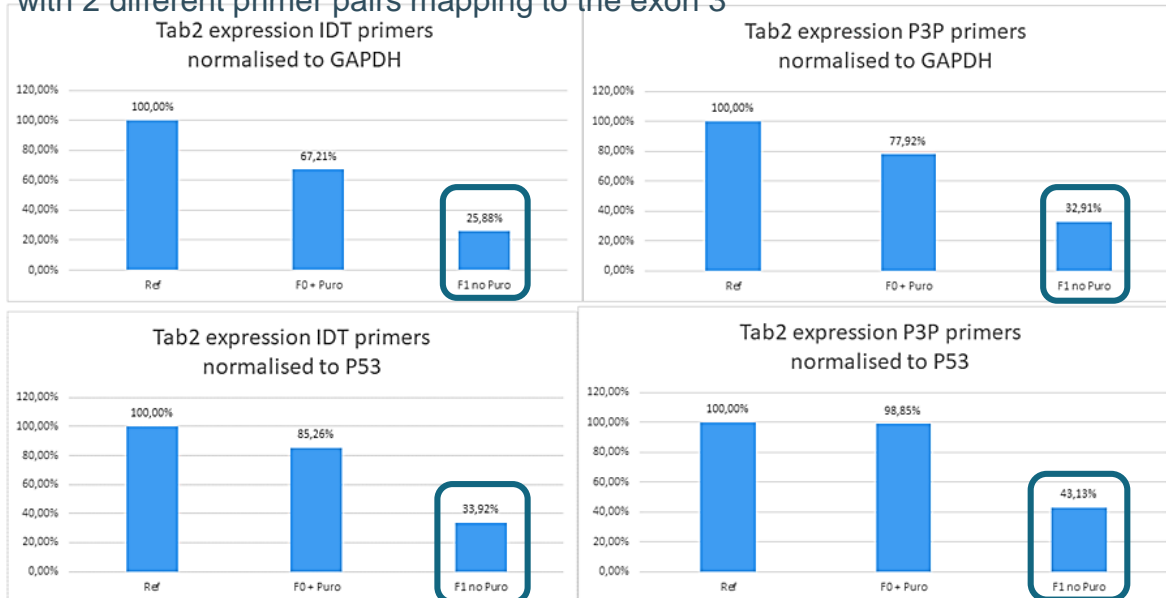


MCA/ID: trio-based WGS analysis of 3 patients with unknown clinical diagnosis

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. Ritelli¹ | S. Morlino² | E. Giacomuzzi¹ | L. Bernardini³ | B. Torres³ | G. Santoro¹ | V. Ravasio¹ | N. Chiarelli¹ | D. D'Angelantonio² | A. Novelli⁴ | P. Grammatico² | M. Colombi¹ | M. Castori⁵

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^a, M.A. Mencarelli^a, F.T. Papa^a, V. Uliana^a, S. Schiavone^b, M. Strambi^b, C. Pescucci^a, F. Ariani^a, V. Rossi^c, I. Longo^a, I. Meloni^a, A. Renieri^{a,*}, F. Mari^a

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

Karin Weiss,¹ Carolyn Applegate,² Tao Wang,^{2,3} and Denise A. S. Batista^{2,4,5*}

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,^{1,2*} Martino Ruggieri,³ Kshitij Mankad,⁴ Gabriella Di Rosa,⁵ Francesca Granata,⁶ Italia Loddo,⁷ Emanuela Moschella,² Maria Pia Calabro,⁷ Anna Capalbo,⁸ Laura Bernardini,⁸ Antonio Novelli,⁹ Agata Polizzi,^{10,11} Daniela G. Seidler,¹² Teresa Arrigo,² and Silvana Briuglia²

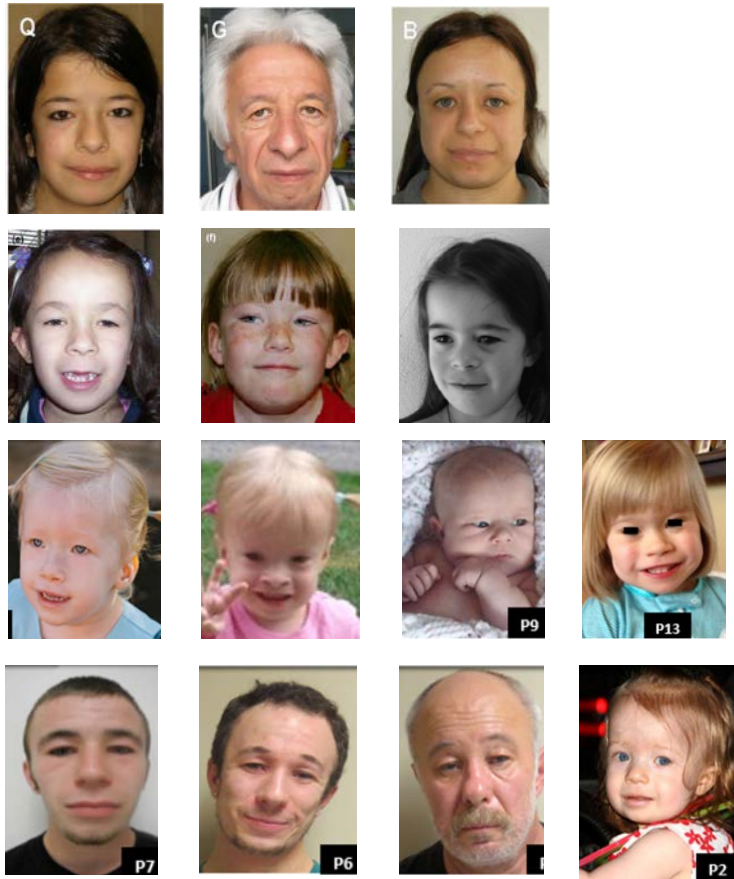
Haploinsufficiency of *TAB2* Causes Congenital Heart Defects in Humans

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

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

Andrew Cheng¹ | Mary Beth P. Dinulos² | Whitney Neufeld-Kaiser³ | Jill Rosenfeld⁴ | McKenna Kyriss⁵ | Suneeta Madan-Khetarpal⁶ | Hiba Risheg⁷ | Peter H. Byers³ | Yajuan J. Liu³

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

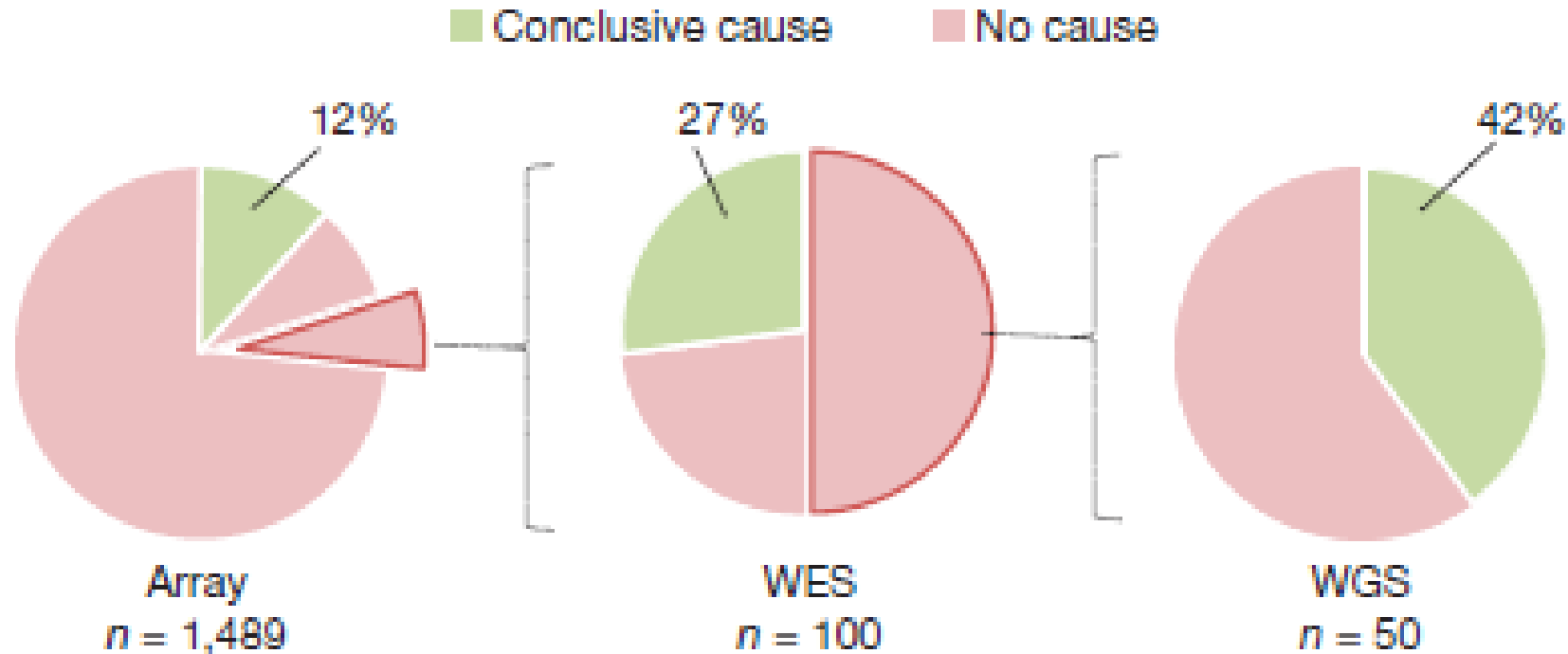


- CHDs
- Facial dysmorphism
- Growth failure
- Joint laxity
- Hypotonia
- Connective tissue abnormalities
- Developmental or intellectual disability

- Horseshoe kidney
- Bilateral choanal atresia
- Hypoplasia of the alveolar process
- Hypertrophy of the alveolar process



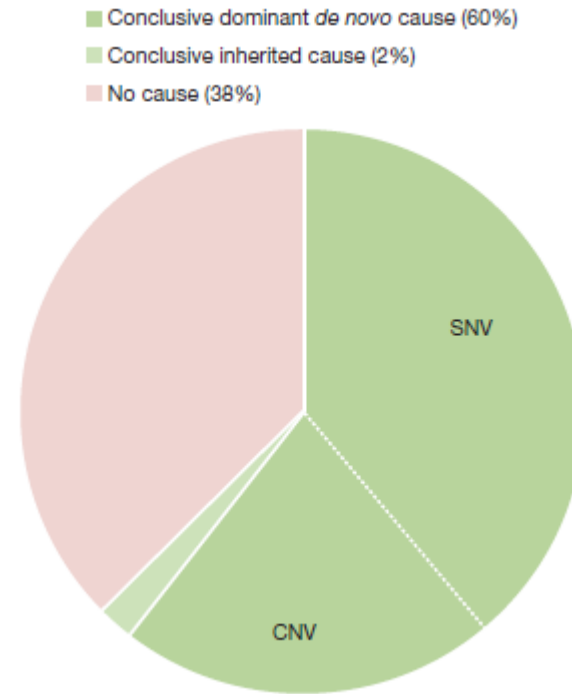
Analysis of the complete genome (SNPs + CNVs)



Genome sequencing identifies major causes of severe intellectual disability

Christian Gillissen^{1*}, Jayne Y. Hehir-Kwa^{2*}, Dje Tjwan Thung³, Maartje van de Vosse⁴, Bregje W.M. van Bon¹, Marjolijn H. Willenssen⁵, Michael Kwint⁶, Irene M. Janssen⁷, Alexander Holschen⁸, Annette Schenck⁹, Richard J. Leahy¹⁰, Robert Klein¹¹, Rick Teasdale¹², Yanbo¹³, Ralph Phundir¹⁴, Helger G. Yntema¹⁵, Bert B. A. de Vries¹⁶, Joris J. A. B. van der Vliet¹⁷, Lisienka E. L. M. Visser^{18*} & Joris A. Vekemans^{19*}

Analysis of the complete genome



Diagnostic yield: 62%

Genome sequencing identifies major causes of severe intellectual disability

Christian Gillissen^{1*}, Jayne Y. Hehir-Kwa^{2*}, Dje Tjwan Thung³, Maartje van de Vosse⁴, Bregje W.M. van Bon¹, Marjolijn H. Willenssen⁵, Michael Kwint⁶, Irene M. Janssen⁷, Alexander Holschen⁸, Annette Schenck⁹, Richard J. Leahy¹⁰, Robert Klein¹¹, Rick Teasdale¹², Tanbo¹³, Ralph Phundir¹⁴, Helger G. Yntema¹⁵, Bert B. A. de Vries¹⁶, Tjibbe de Graaf¹⁷, Peter van der Vliet¹⁸, Lisienka E. L. M. Visser^{19*} & Joris A. Vekemans^{20*}

University

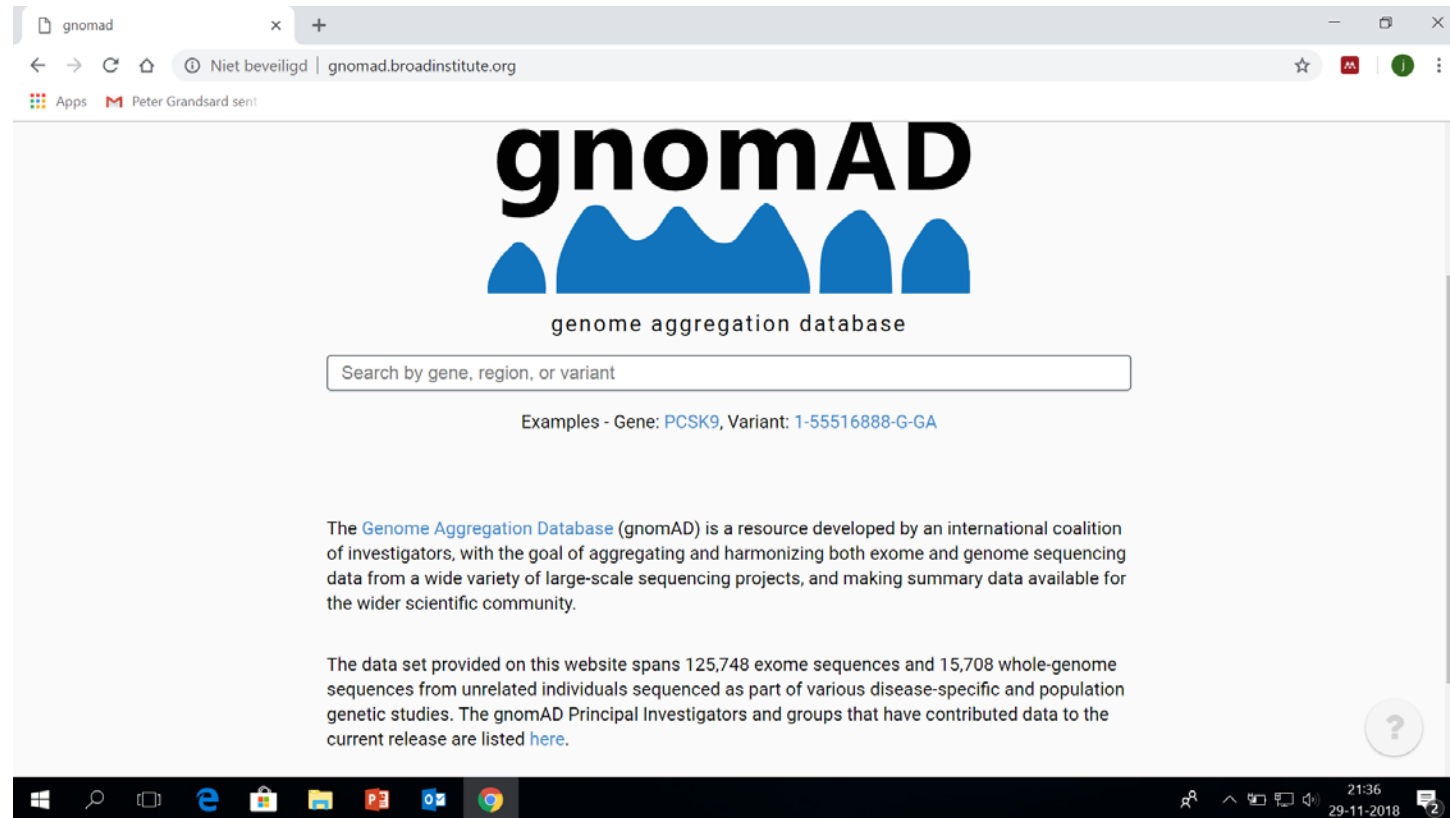
KU LEUVEN

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 are examples: "Examples - Gene: [PCSK9](#), Variant: [1-55516888-G-GA](#)". The main text on the page describes the database as 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. It also states that 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 the Windows logo, search icon, and several application icons, including Edge, File Explorer, PowerPoint, and Chrome. The system tray shows the time as 21:36 and the date as 29-11-2018.

We are all mutants!

