

NGS wprowadzenie; klasyfikacja wariantów, zapis i wytyczne

Monika Gos

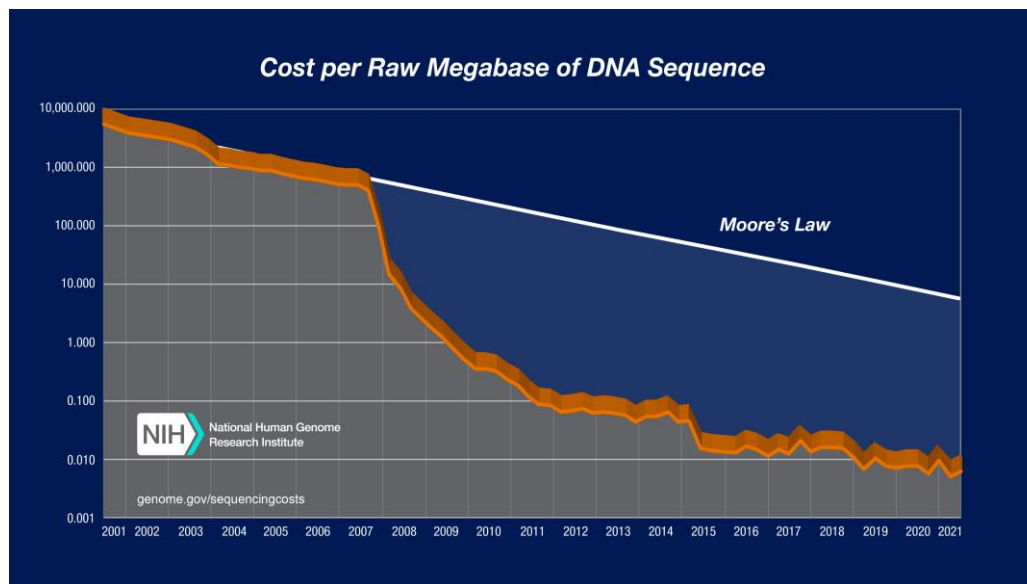
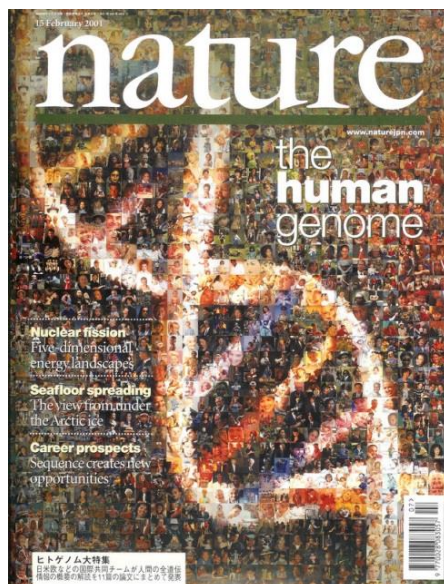
Zakład Genetyki Medycznej

Instytut Matki i Dziecka

Kurs: Warsztaty z analizy wyników NGS - eksomowego i panelowego,
19-20.05.2022

Projekt sekwencjonowania genomu ludzkiego

Genom	Metoda	Koszt
Human Genome Project (13 lat)	Sekwenator kapilarny	2,7 mld \$
Venter Genome (9 miesięcy)	Sekwenator kapilarny (>340tys. reakcji)	70 mln \$
James Watson	Roche, 454 (234 reakcje)	1 mln \$
James Lupski	Life/APG (3 reakcje)	75 tys. \$



Fundusze Europejskie
Wiedza Edukacja Rozwój



Rzeczpospolita
Polska



Instytut
Matki i Dziecka

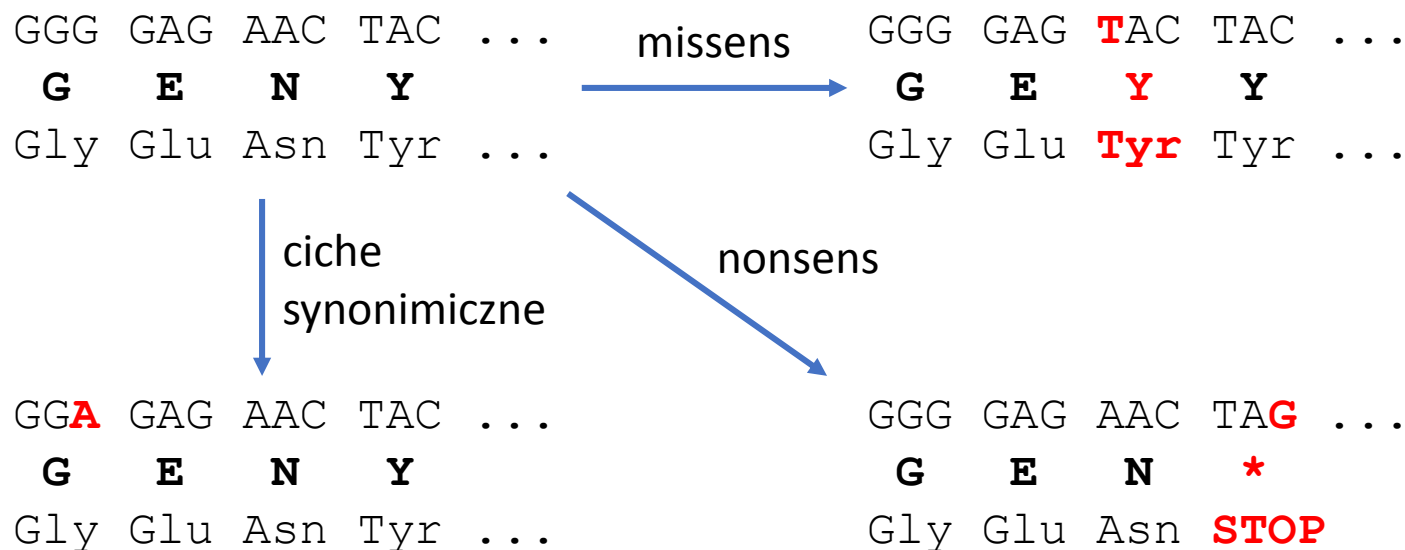
Unia Europejska
Europejski Fundusz Społeczny



edu-metgen.imid.med.pl

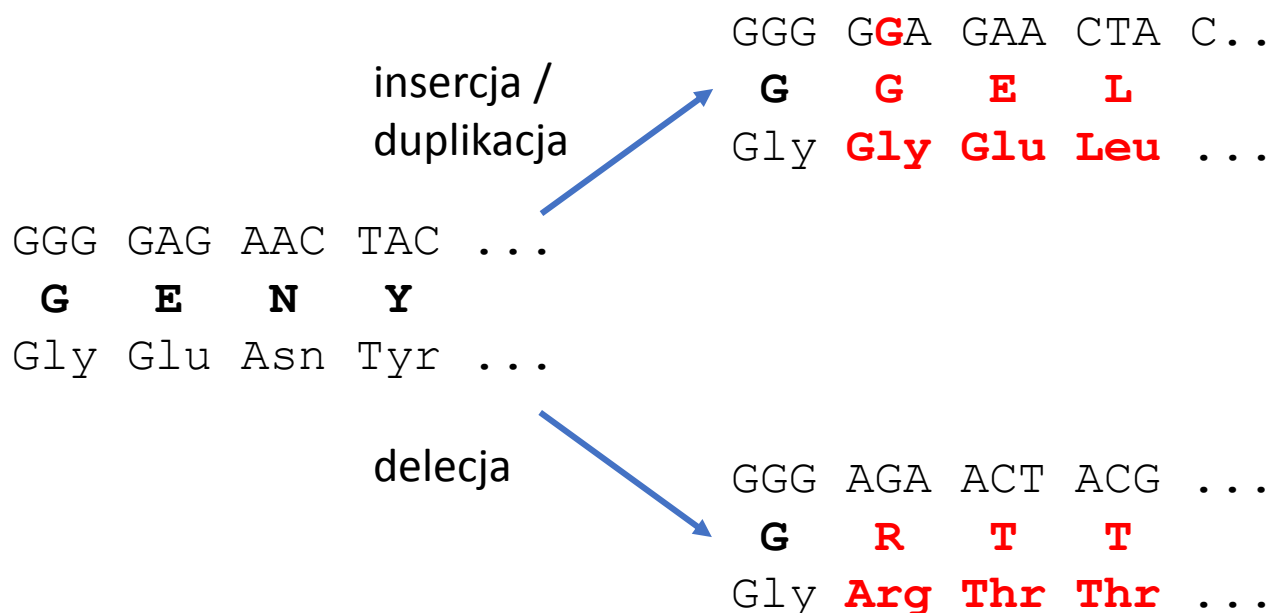
Jakie mutacje identyfikujemy za pomocą sekwencjonowania?

Mutacje o charakterze substytucji



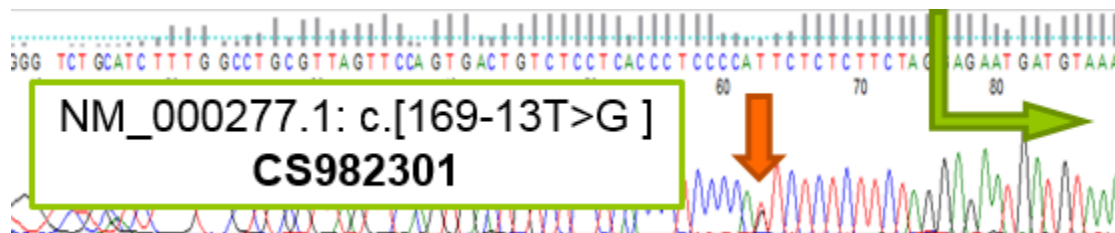
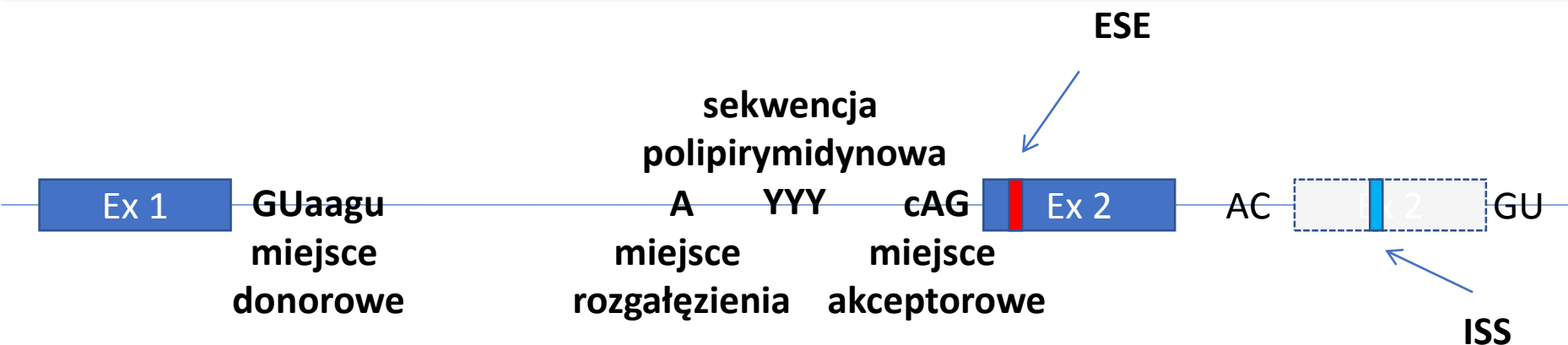
Jakie mutacje identyfikujemy za pomocą sekwencjonowania?

Delecje / insercje – mutacje typu frameshift



Jakie mutacje identyfikujemy za pomocą sekwencjonowania?

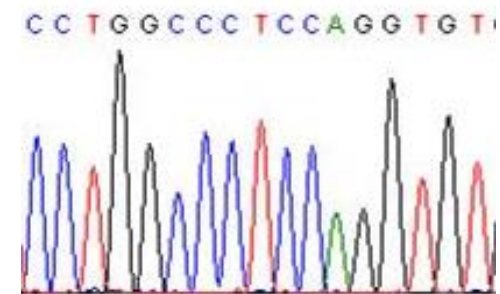
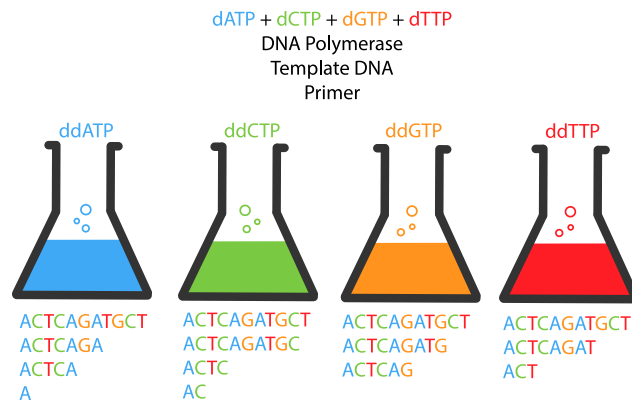
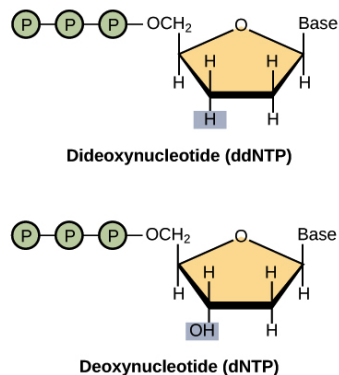
Mutacje splicingowe (najczęściej substytucje)



NGS: SpliceAI, ADA, RF
Sanger: Alamut, Human Splicing Finder

Sekwencjonowanie

- odczyt sekwencji nukleotydowej wybranego pojedynczego fragmentu DNA
- sekwencjonowanie I generacji
Fred Sanger – laureat nagrody Nobla – 1977
Matryca do sekwencjonowania: produkt reakcji PCR



https://www.ensembl.org/Homo_sapiens/

Baza danych



Fundusze
Europejskie
Wiedza Edukacja Rozwój



Rzeczpospolita
Polska

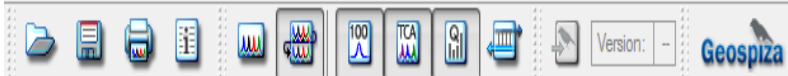


Instytut
Matki i Dziecka

Unia Europejska
Europejski Fundusz Społeczny



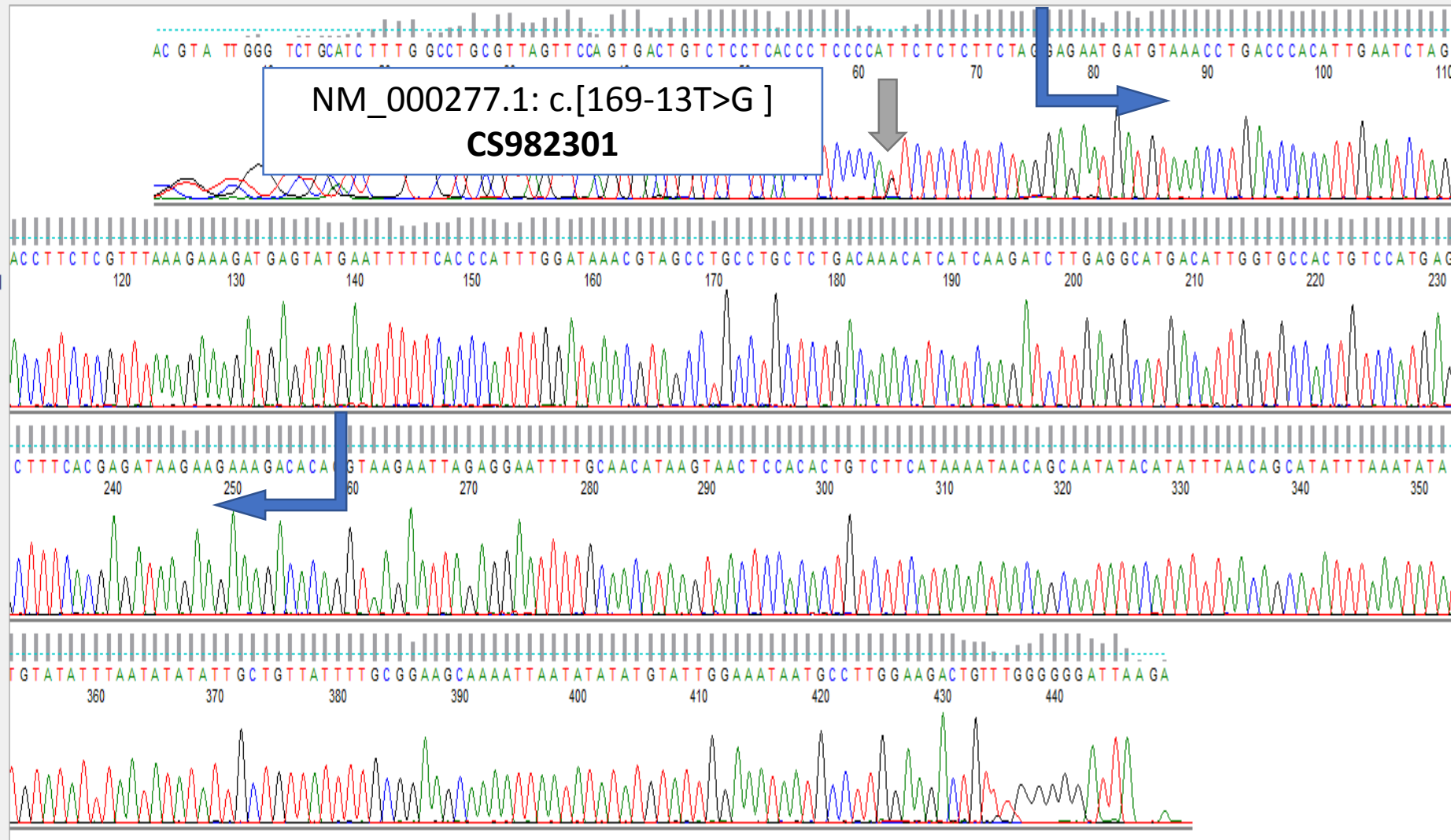
edu-metgen.imid.med.pl

Go to Base No. Find Sequence

A C G T

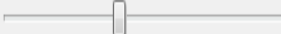
NM_000277.1: c.[169-13T>G]

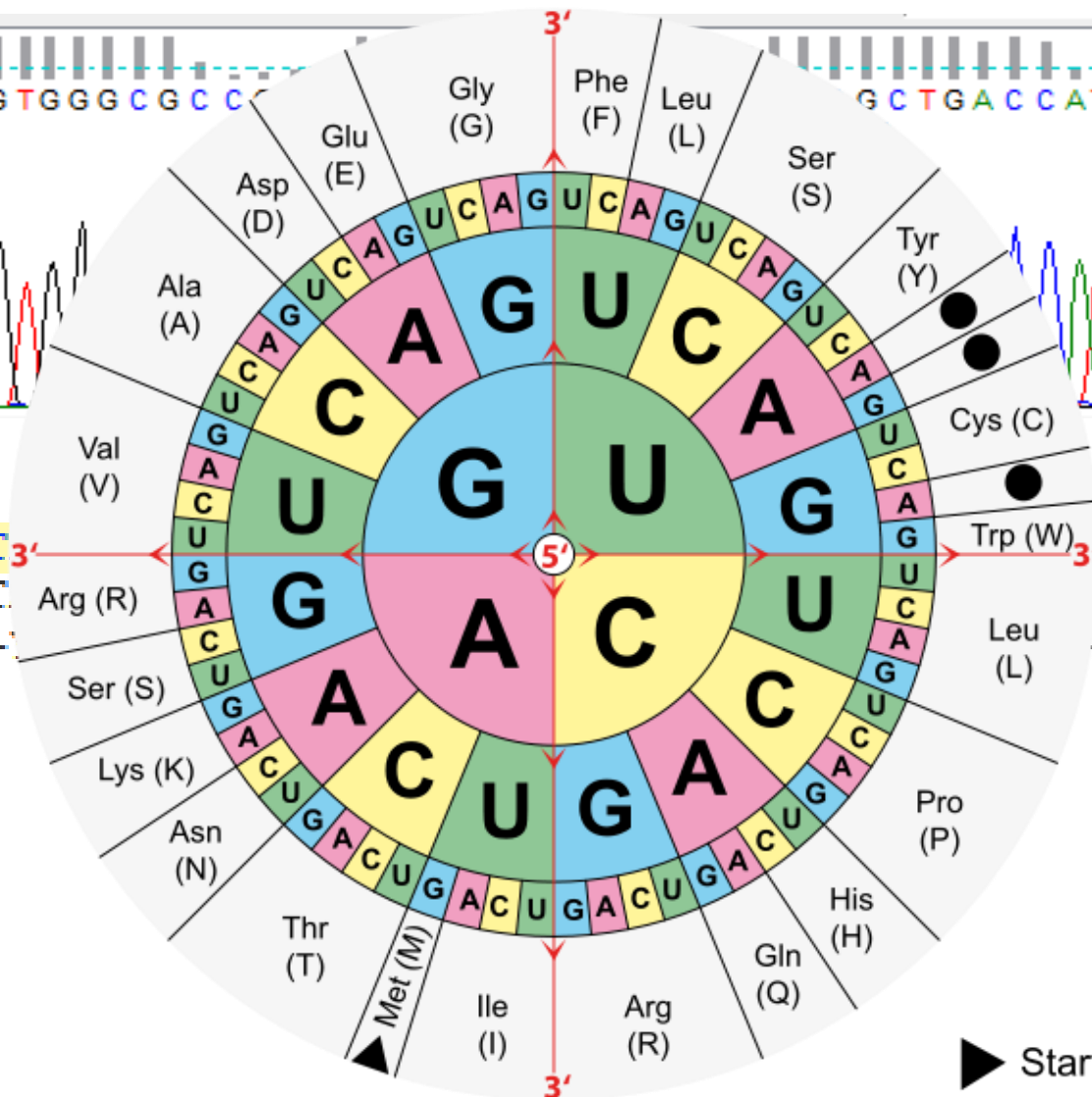
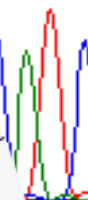
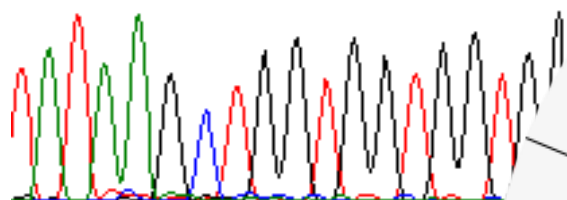
CS982301

Vertical
Scale

Reset Scales

Horizontal Scale





240
52
17

► Start
● Stop

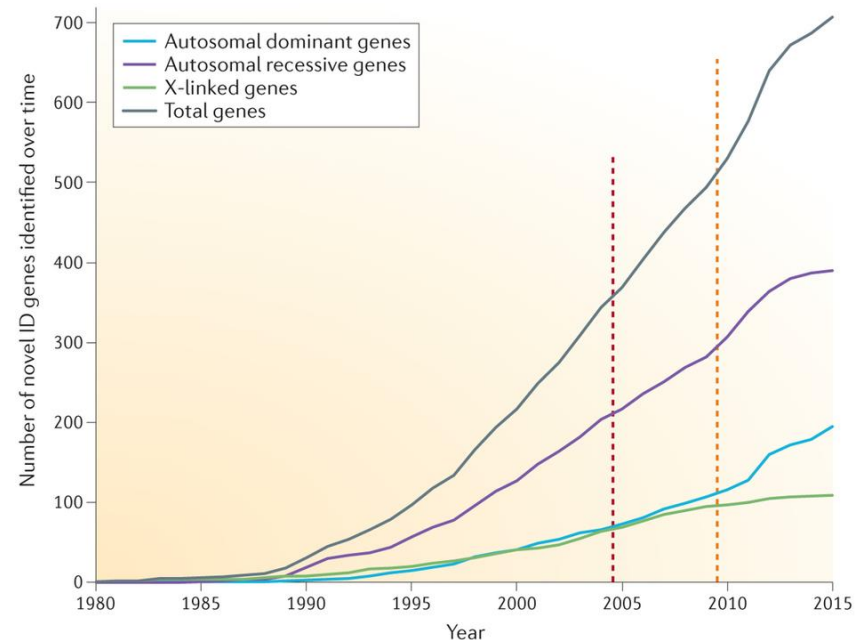
GGC>AGC

p.Gly12Ser

Mutation taster
<https://www.mutationtaster.org/>
Program predykcyjny

Sekwencjonowanie

- sekwencjonowanie następnej generacji (NGS) - wysokoprzepustowe sekwencjonowanie równoległe, pierwotnie opracowane do sekwencjonowania genomów *de novo*; aktualnie resekwencjonowanie
- **sekwencjonowanie II generacji** wymaga amplifikacji materiału przed właściwą reakcją sekwencjonowania
- **sekwencjonowanie III generacji** nie wymaga amplifikacji materiału; sekwencjonowanie bezpośrednie



Nature Reviews | Genetics

Pierwsze publikacje - 2012

zespół Kabuki – *KMT2D*

zespół Freemana-Sheldona - *MYH3*

zespół Millera – *DHODH*

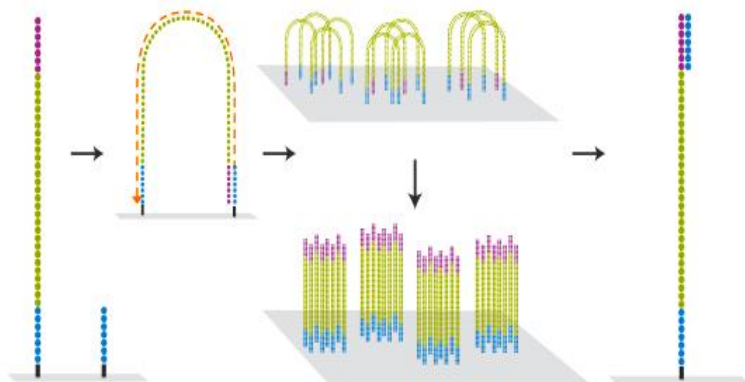
zespół Schinzla-Giediona - *SETBP1*

nieswoiste zapalenie jelit

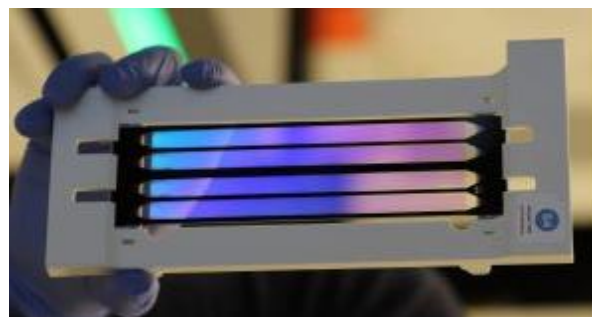
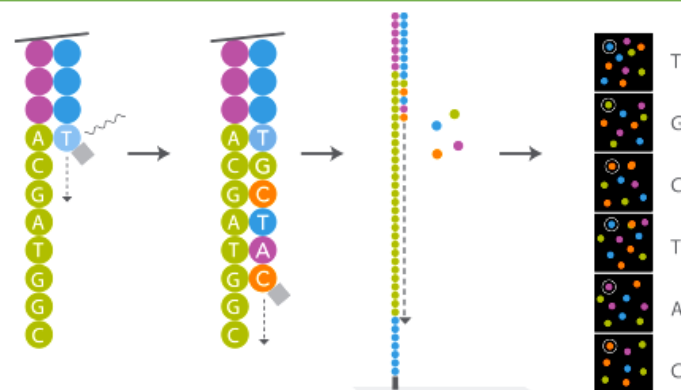
przypominające zespół Crohna - *XIAP*

Sekwencjonowanie II generacji - Illumina

Amplifikacja klastrowa (mostkowa)



Sekwencjonowanie



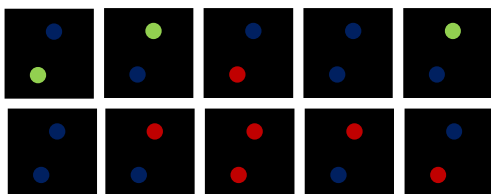
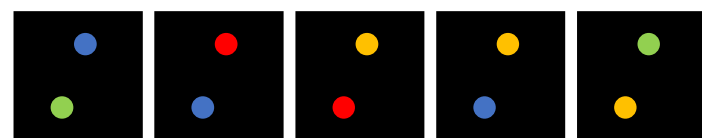
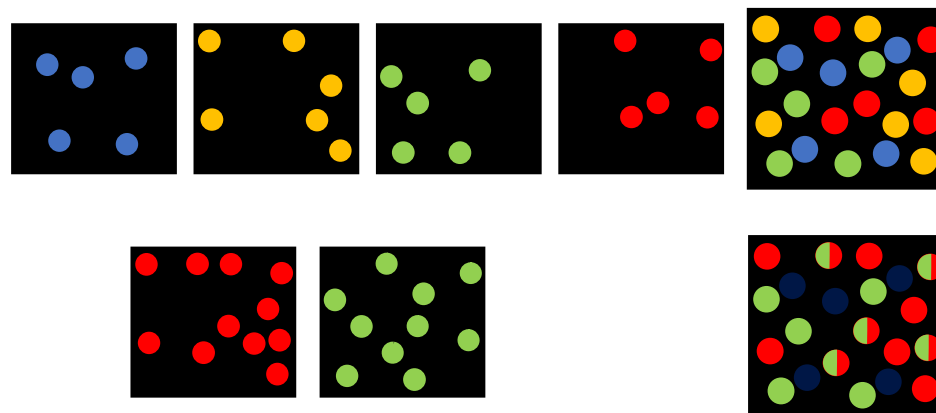
Sekwencjonowanie II generacji - Illumina

4-Channel Chemistry				
	A	G	T	C
Image 1				
Image 2				
Image 3				
Image 4				
Result	A	G	T	C

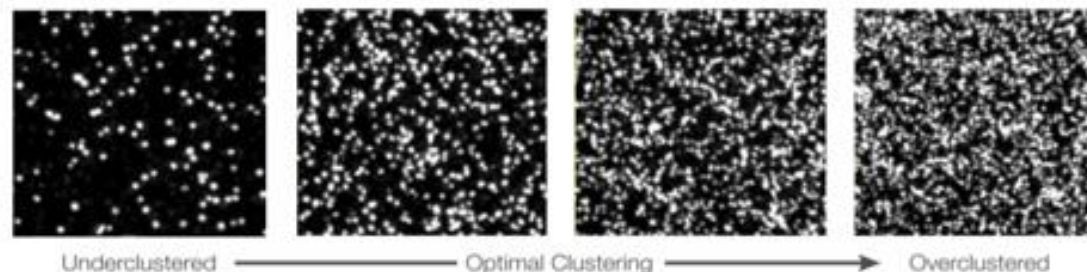
MiSeq, HiSeq

2-Channel Chemistry				
	A	G	T	C
Image 1				
Image 2				
Result	A	G	T	C

NextSeq, NovaSeq



GACCT...
TGAGC...



Underclustered

Optimal Clustering

Overclustered

JPG, BCL → fastq

Plik fastq

The diagram illustrates a single record in the FASTQ format, which is a standard for storing high-throughput sequencing data. The record is represented as a four-line text block:

```
@FORJUSP02AJWD1
CCGTCAATTCATTTAAGTTTAACTTGCGGCCGTA
CTCCCAGGCGGT
+
AAAAAAAAAAAA::99@:::??@@::FFAAAAACCAA:::BB@@?A?
```

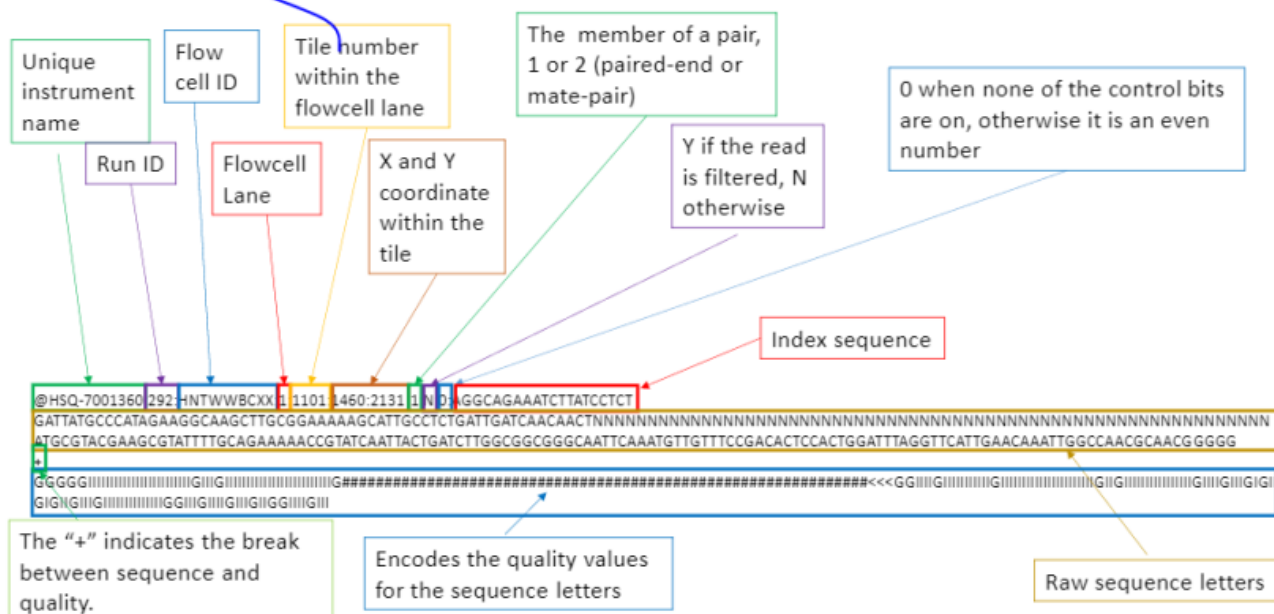
Annotations with arrows point to specific parts of the record:

- Label:** Points to the first line, which contains the sequence identifier.
- Sequence:** Points to the second line, which contains the raw sequencing data (nucleotide bases).
- Q scores (as ASCII chars):** Points to the third line, which contains quality scores represented as ASCII characters.
- Base=T, Q=':'=25:** Points to the first character of the quality score line ('A').

The '+' symbol on the fourth line indicates that the quality scores are in Sanger (Phred) format.

https://www.drive5.com/usearch/manual/fastq_files.html

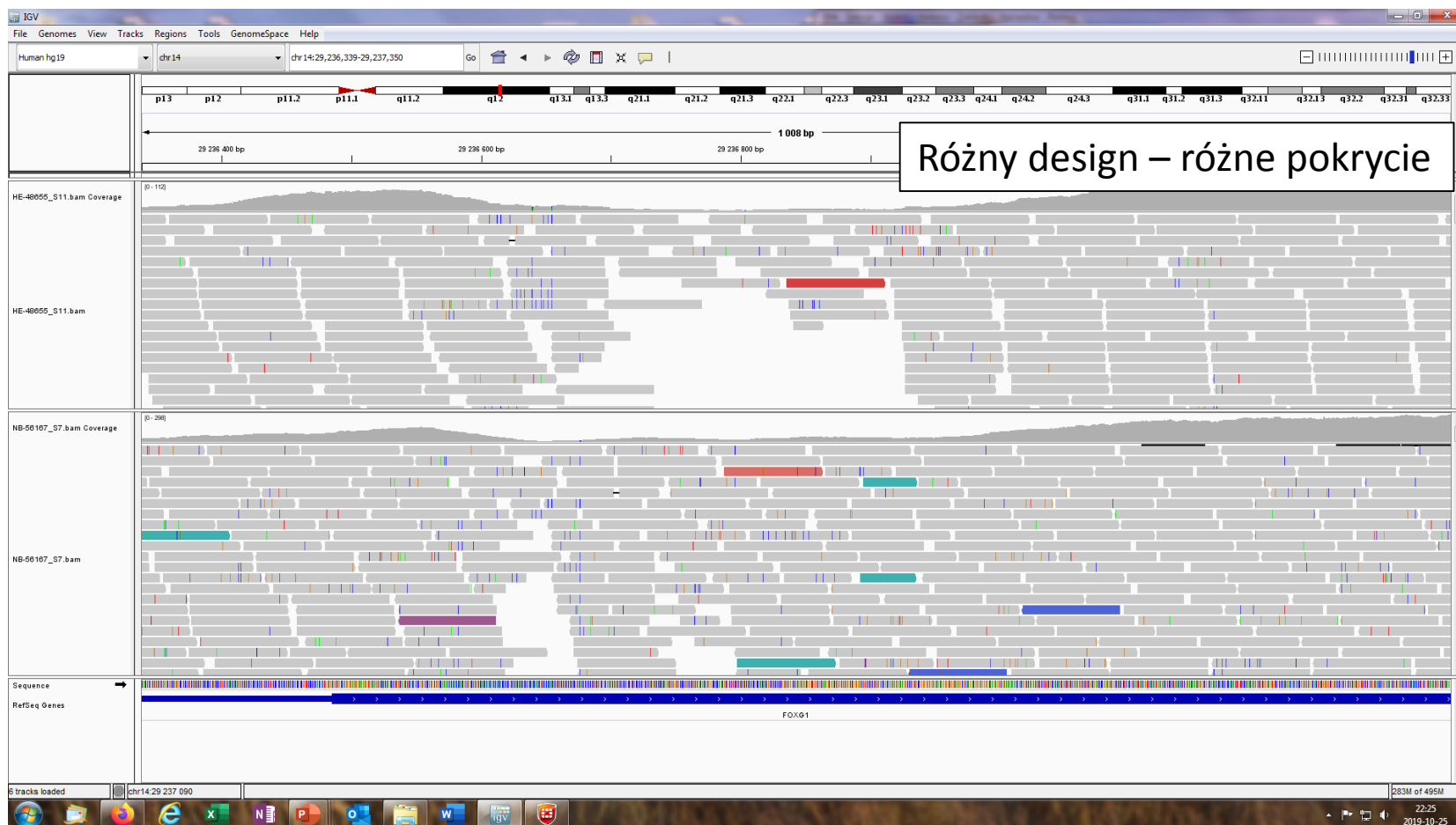
FASTQ File Format Analysis

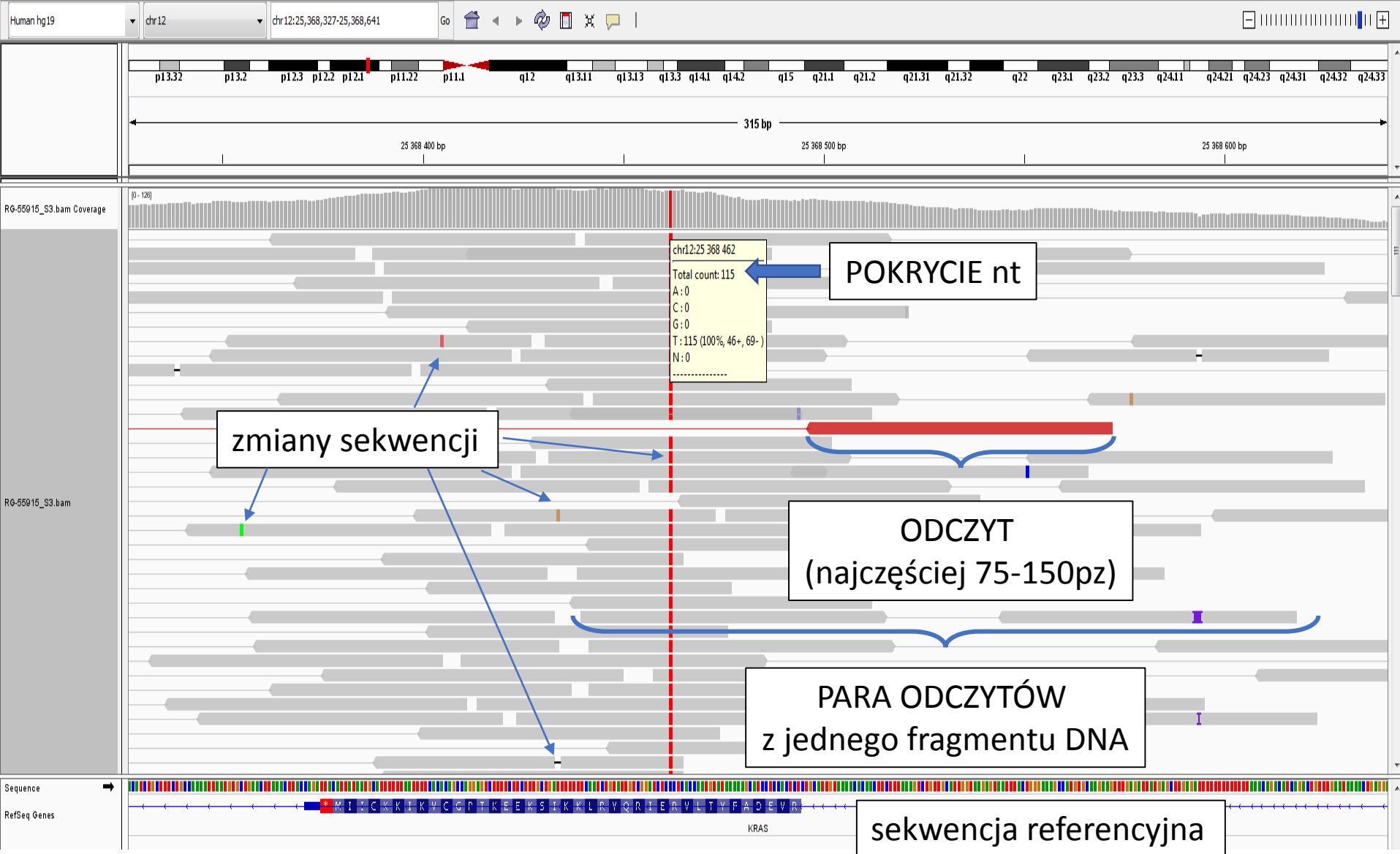


Przyrównanie do sekwencji referencyjnej lub *de novo* assembly
fastq → bam

<https://kscbioinformatics.wordpress.com/2017/02/03/raw-illumina-sequence-data-files-for-dummies-part-1/>

Integrative Genomic Viewer





Minusy „krótkich odczytów”

- Problemy z sekwencjonowaniem regionów bogatych w pary GC
- Słaba rozdzielczość w regionach „kompleksowych” np. MHC
- Regiony powtórzone / paralogi – problemy z uliniowaniem sekwencji
- Brak możliwości identyfikacji rearanżacji chromosomowych (rozległe delecje/duplikacje; warianty strukturalne)
- Brak możliwości „fazowania” wariantów

Parametr	MiniSeq	MiSeq	NextSeq	NovaSeq
Czas reakcji	4–24 godz.	4–55 godz.	12–30 godz.	16–44 godz.
Max ilość danych	7.5 Gb (5Gb)	15 Gb (5Gb)	120 Gb (80Gb)	6000 Gb
Max liczba odczytów	25 mln	25 mln*	400 mln	20 mld
Max długość odczytu	2 × 150 bp	2 × 300 bp	2 × 150 bp	2 × 150 bp
Genom ludzki 30x	21 reakcji		2 reakcje	1 reakcja (3-19)



Można analizować **genom**, ale ...
 To 3 miliardy par zasad, z czego >95% to DNA niekodujący,
 o którym jeszcze mało wiemy
 1.5% genomu – część kodująca (25 tys. genów o wielkości 16,3kpz) tzw. **eksom**

Parametr	Genom	Eksom	Eksom kliniczny	Celowane NGS
koszt	++	++	++	+
pokrycie	+	++	+++	++++
Ilość danych	++++	+++	++	+
CNV	+ (strukturalne)	+/-	+/-	+/-
nowe geny ?	+	+	-	-

Zalecenia EuroGenTest (12.2014)

http://www.eurogentest.org/fileadmin/templates/eugt/pdf/NGS_Guidelines/EuroGentest_NGS_guidelines_2014_-_final_draft_02-12-2014_v2.pdf

STATEMENT 2.03: For diagnostic purpose, only genes with a known (i.e. published and confirmed) relationship between the aberrant genotype and the pathology, should be included in the analysis.

STATEMENT 2.04: For the sake of comparison, to avoid irresponsible testing, for the benefit of the patients, 'core disease gene lists' should be established by the clinical and laboratory experts.

STATEMENT 3.02: The analysis pipeline of diagnostic laboratories should focus on the gene panel under investigation in order to avoid the chance of secondary findings, and be validated accordingly.

STATEMENT 3.04: If a clinical centre or a laboratory decides to offer patients the possibility to get carrier status for unrelated diseases and secondary findings, it should implement an opt-in, opt-out protocol and all the logistics need to be covered.

STATEMENT 5.01: The report of an NGS assay should summarize the patient's identification and diagnosis, a brief description of the test, a summary of results, and the major findings on one page.

Panel App

(<https://panelapp.genomicsengland.co.uk/>)

Panel ▼	Evaluated genes	Reviewers	Actions
growth 6 panels			
Beckwith-Wiedemann syndrome (BWS) and other congenital overgrowth disorders Level 3: Beckwith-Wiedemann syndrome (BWS) and other congenital overgrowth disorders Level 2: Growth disorders Relevant disorders: Atypical Beckwith-Wiedemann syndrome, Classical Beckwith-Wiedemann syndrome, Simpson-Golabi-Behmel syndrome, Sotos syndrome, Weaver syndrome Version 1.103 Panel Types: Rare Disease 100K,	29 of 30 97% Regions: 1	8 reviewers	Download
Growth failure in early childhood Relevant disorders: R147 Version 1.58 Panel Types: GMS Rare Disease Virtual, GMS signed-off, Signed Off v.1.4 on 3 Mar 2020	129 of 148 87% Regions: 5	10 reviewers	Download
IUGR and IGF abnormalities Level 3: Growth hormone disorders Level 2: Endocrine disorders Relevant disorders: Version 1.35 Panel Types: Rare Disease 100K,	108 of 115 94% Regions: 5	13 reviewers	Download

Human Phenotype Ontology (HPO)

The screenshot shows the HPO website with a dark blue header containing navigation links: Tools, Downloads, and Documentation. The main content area features a large search bar with the HPO logo and a dropdown menu set to 'All'. Below the search bar, there are examples of search results: 'e.g. brachnodactyly | Marfan syndrome | FBN1'. The page is divided into two columns. The left column, titled 'The Human Phenotype Ontology', contains a paragraph describing the HPO as a standardized vocabulary of phenotypic abnormalities, followed by a 'Learn More About HPO' button. The right column, titled 'News & Updates', lists three recent releases: 'August 2021 HPO release' (August 2, 2021), 'June 2021 HPO release' (June 8, 2021), and 'April 2021 HPO release' (April 14, 2021), with a 'View All News' button. Below these columns are four tool cards: 'Exomiser' (Evaluate variants based on the predicted pathogenicity), 'Genomiser' (Analyze genome sequence data for non-coding variants), 'Phenomizer' (Rank disease differential diagnosis by clinical features), and 'Profile Search' (Discover diseases with a phenotype profile). At the bottom, the Monarch Initiative logo is displayed alongside the text 'Cross-species disease discovery and diagnosis.' and an 'Explore Monarch Initiative' button.

The Human Phenotype Ontology

The Human Phenotype Ontology (HPO) provides a standardized vocabulary of phenotypic abnormalities encountered in human disease. Each term in the HPO describes a phenotypic abnormality, such as [Atrial septal defect](#). The HPO is currently being developed using the medical literature, Orphanet, DECIPHER, and OMIM. HPO currently contains over 13,000 terms and over 156,000 annotations to hereditary diseases. The HPO project and others have developed software for phenotype-driven differential diagnostics, genomic diagnostics, and translational research. The HPO is a flagship product of the [Monarch Initiative](#), an NIH-supported international consortium dedicated to semantic integration of biomedical and model organism data with the ultimate goal of improving biomedical research. The HPO, as a part of the Monarch Initiative, is a central component of one of the [13 driver projects](#) in the [Global Alliance for Genomics and Health \(GA4GH\) strategic roadmap](#).

[Learn More About HPO](#)

News & Updates

[August 2021 HPO release](#) August 2, 2021

[June 2021 HPO release](#) June 8, 2021

[April 2021 HPO release](#) April 14, 2021

[View All News](#)

Exomiser
Evaluate variants based on the predicted pathogenicity.

Genomiser
Analyze genome sequence data for non-coding variants.

Phenomizer
Rank disease differential diagnosis by clinical features.

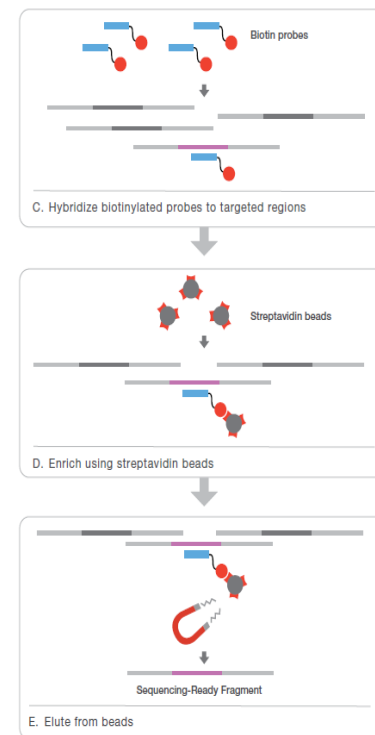
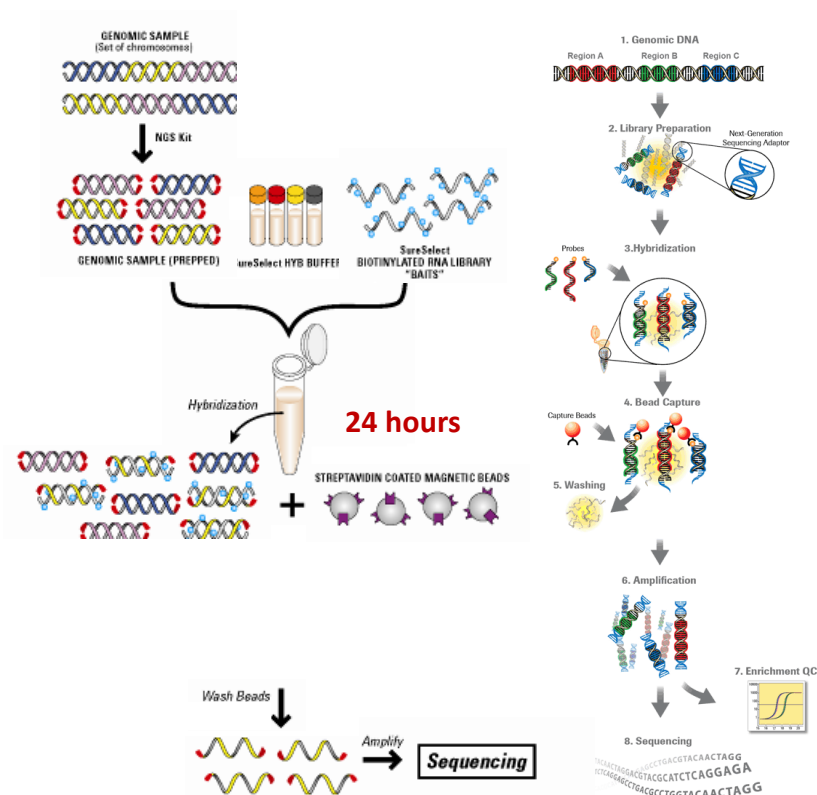
Profile Search
Discover diseases with a phenotype profile.

monarch INITIATIVE

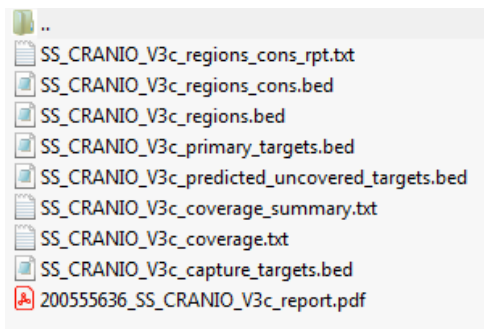
Cross-species disease discovery and diagnosis. [Explore Monarch Initiative](#)

Sekwencjonowanie celowane

dodatkowy
etap
wzbogacania
fragmentów
DNA pod
względem
analizowanej
sekwencji
tzw. TARGET

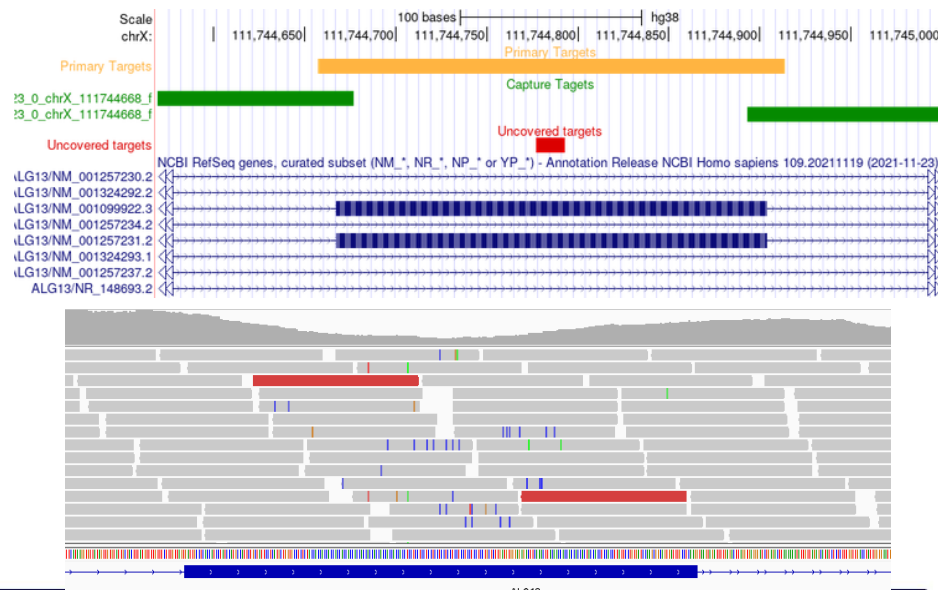
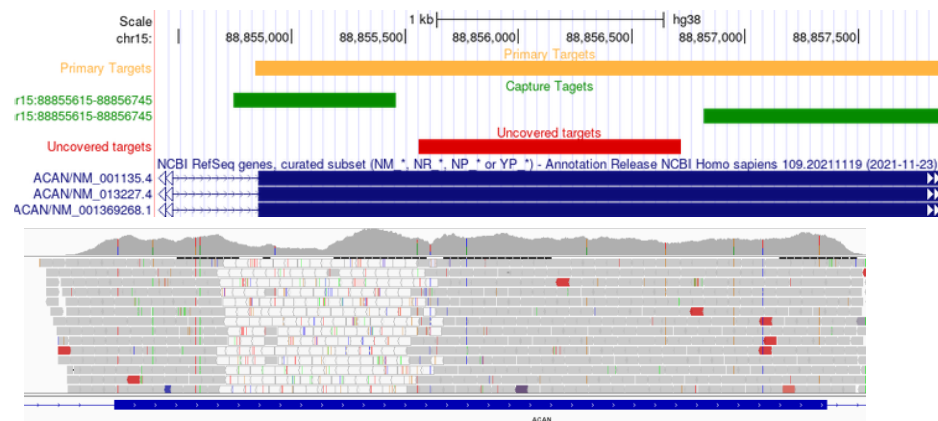
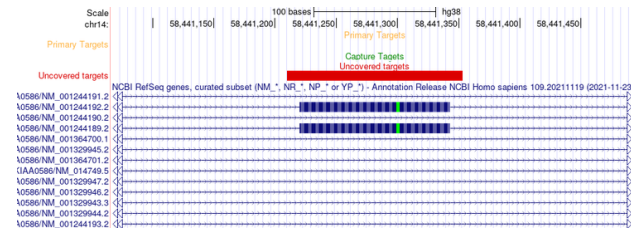


Sekwencjonowanie celowane



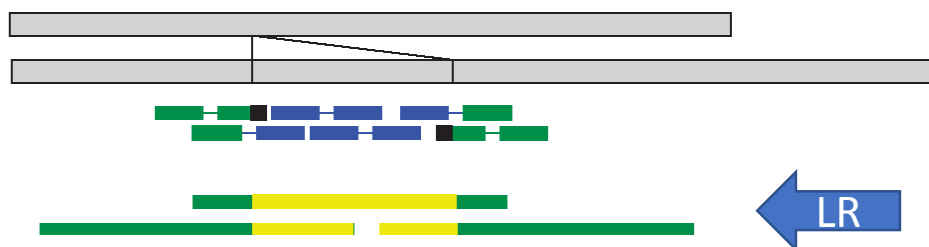
chr1	243612950	243613070
chr12	65914699	65914925
chr12	110167860	110167988
chr14	58441210	58441353
chr15	48729407	48729417
chr15	88855561	88856715
chr16	53956672	53956799
chr2	33323486	33323511
chr2	112566747	112566905
chr20	32372153	32372179
chr7	100116503	100116611
chr7	100118385	100118484
chr7	158905164	158905293
chrX	65847	65867
chrX	53648362	53648629

A, B czy C?

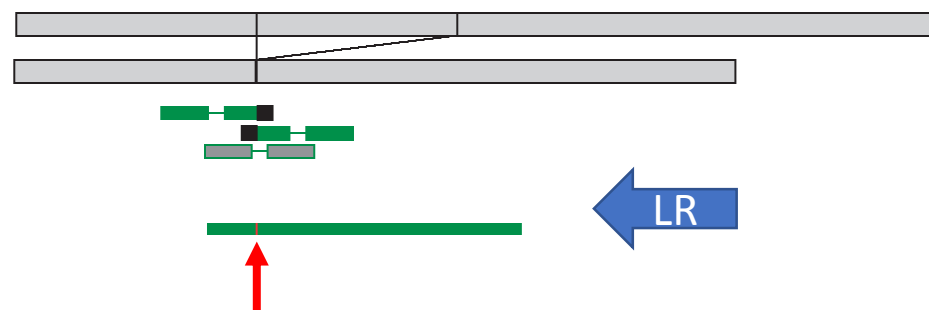


Rozwiązanie – długie odczyty

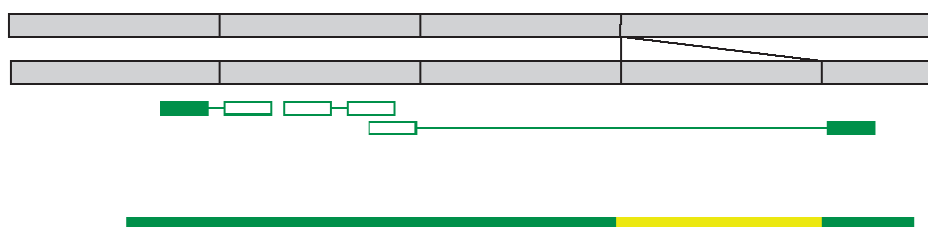
Insercja



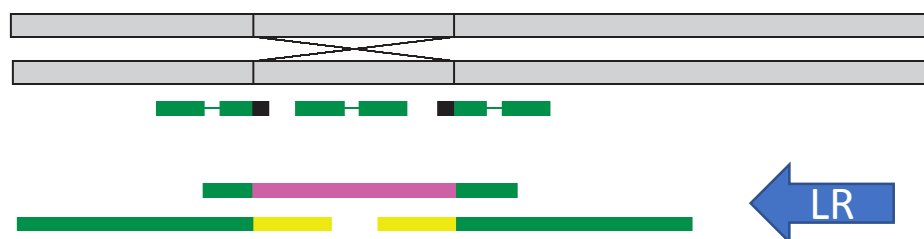
Delecja



Duplikacja (CNV)



Inwersja

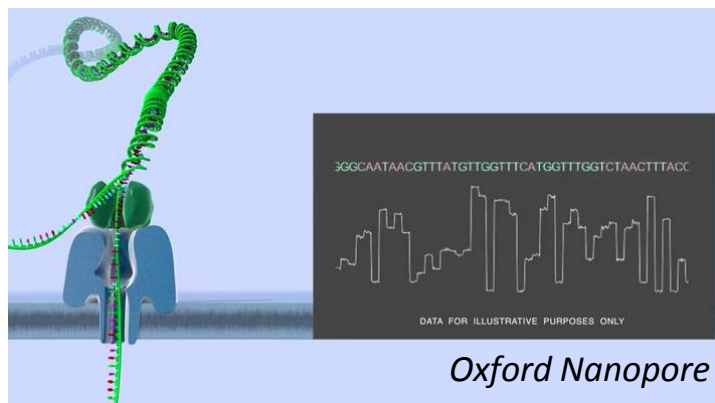


Krótkie odczyty: błędne mapowanie odczytów / brak mapowania

Długie odczyty: sekwencjonowanie punktów złamań

<https://academic.oup.com/hmg/article/27/R2/R234/4996216>

Sekwencjonowanie III generacji



- + długie odczyty (do 15kb) co umożliwia identyfikację rozległych rearanżacji genomu i sekwencjonowanie sekwencji powtórzonych
- duża ilość generowanych błędów (do 15%)

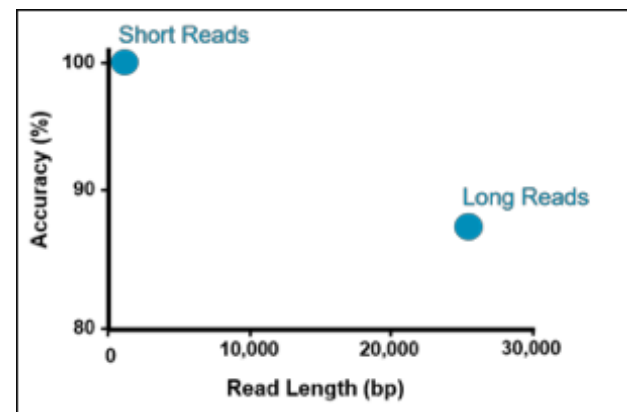
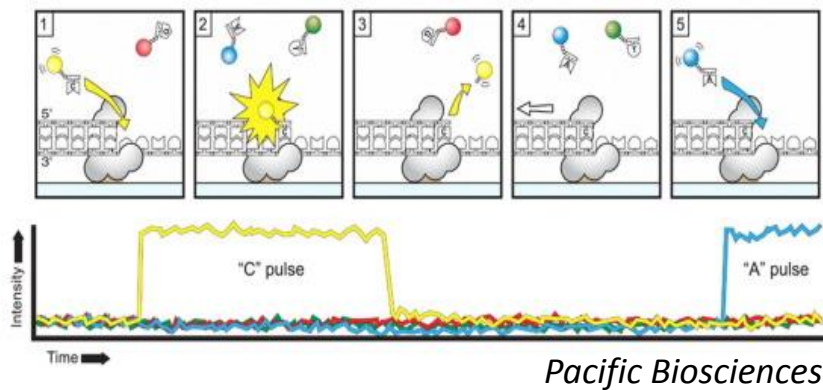
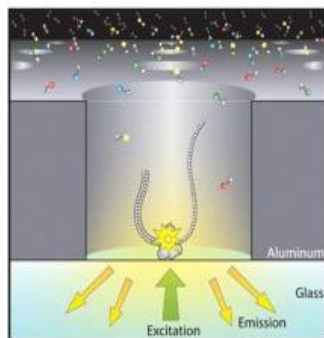
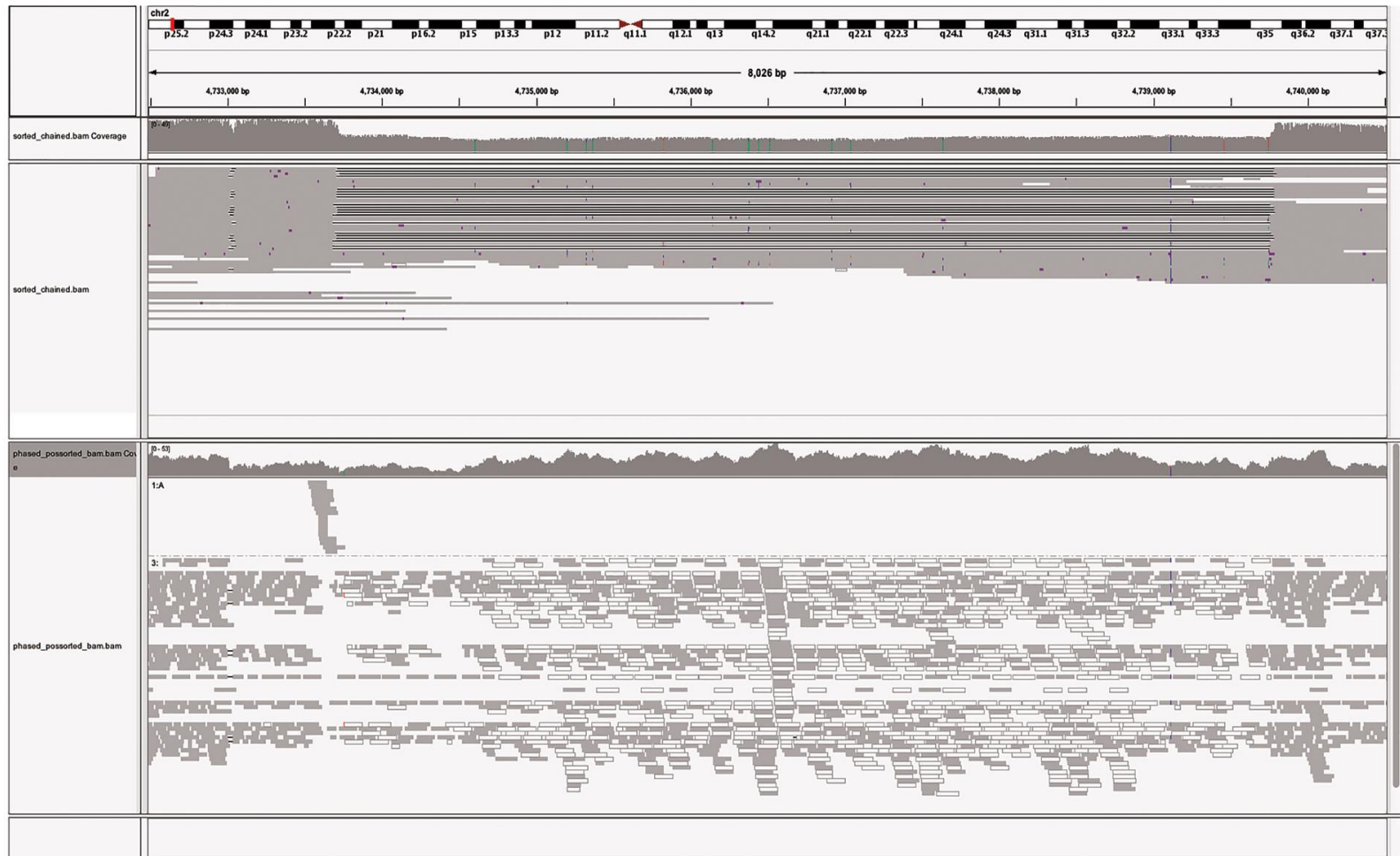
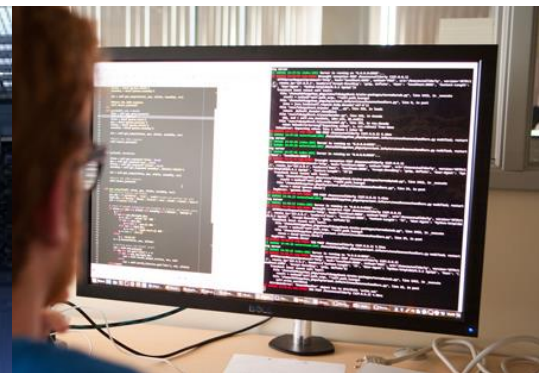


Figure 3. Long reads span and call variations that short reads cannot. IGV ...





DANE, które należy zinterpretować



pliki *.bcl → plik *.fastq (CASAVA, bcl2fastq, basespace)

dopasowanie wygenerowanych przez sekwenator
sekwencji do genomu referencyjnego (hg19,hg38)



plik *.bam (BWA)

analiza sekwencji pod kątem obecności zmian – podstawień pojedynczych
nukleotydów lub insercji / delecji + anotacja



plik *.vcf (m.in. GATK, SAMTools)

analiza funkcjonalna – odniesienie do populacyjnych baz danych
(gnomAD, ClinVar), analiza predykcyjna

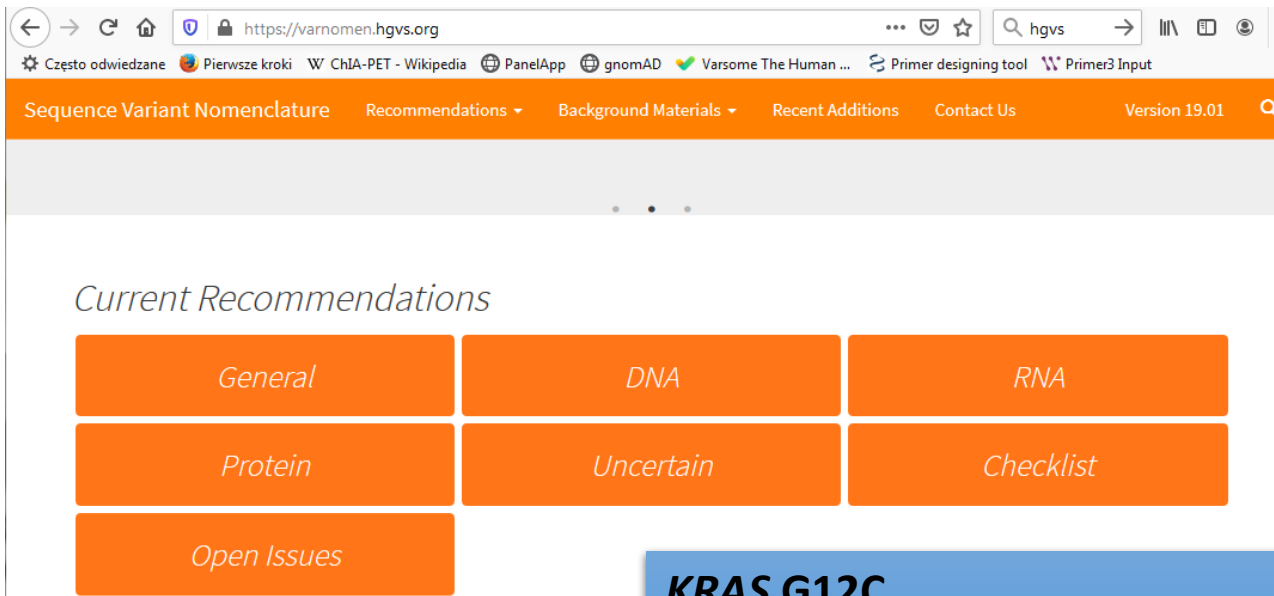


plik Excel (*.tsv, *.txt, *.csv) lub oprogramowanie komercyjne

wybór wariantu związanego z patogenezą choroby

HGVS Recommendations for the Description of Sequence Variants: 2016 Update

Johan T. den Dunnen,^{1*} Raymond Dalglish,² Donna R. Maglott,³ Reece K. Hart,⁴ Marc S. Greenblatt,⁵ Jean McGowan-Jordan,⁶ Anne-Francoise Roux,⁷ Timothy Smith,⁸ Stylianos E. Antonarakis,⁹ and Peter E.M. Taschner¹⁰ on behalf of the Human Genome Variation Society (HGVS), the Human Variome Project (HVP), and the Human Genome Organisation (HUGO)



KRAS G12C

- LRG_344t1:c.12G>C (NM_004985.5:c.12G>C)
- LRG_344t1:c.34G>T (p.Gly12Cys or G12C)
- LRG_344t1:c.37G>T (p.Gly13Cys)

Human Gene Variation Society (HGVS)

<http://varnomen.hgvs.org/> Version 20.05

1. należy używać nazw genów zgodnych z **HGNC** (np. nie *MLL2*, tylko *KMT2D*)
2. wszystkie opisywane zmiany muszą być opisane na poziomie DNA
3. zawsze w opisie mutacji należy podać literę opisującą sekwencję referencyjną:
 - > kodujący DNA – **c.35A>G**
 - > sekwencja genomowa – **g.375C>T**
 - > sekwencja mitochondrialna – **m.8993T>C**
 - > sekwencja RNA – **r.67c>g**
 - > sekwencja aminokwasowa – **p.Arg408Trp**
 - > niekodujący RNA – **n....**
4. w opisie zmiany powinien być podany numer sekwencji referencyjnej.
5. sekwencja referencyjna nukleotydowa – główny i najdłuższy transkrypt
6. sekwencja referencyjna białkowa – pierwotny produkt translacji
7. DNA – duże litery, RNA – małe litery, białko – kod trójliterowy, pierwsza litera duża



Human Gene Variation Society (HGVS)

<http://varnomen.hgvs.org/>

Rekomendowane sekwencje referencyjne:

RefSeq sequences with the prefixes NC_, NT_, NW_, NG_, NM_, NR_ or NP_

chromosome - NC_000023.11

genomic contigs or scaffolds - NT_010718.17, NW_003315950.2

gene/genomic region - NG_012232.1

coding transcript - NM_004006.2

non-coding transcript - NR_004430.2

protein - NP_003997.1

LRG sequences with the prefixes LRG_#, LRG_#t#, LRG_#p# (see examples below)

gene/genomic region - LRG_199

coding transcript (or non-coding transcript) - LRG_199t1

protein - LRG_199p1

Ensembl transcript (ENST) and protein (ENSP) which are not identified by Ensembl as being incomplete

gene/genomic region - ENSG00000198947.15

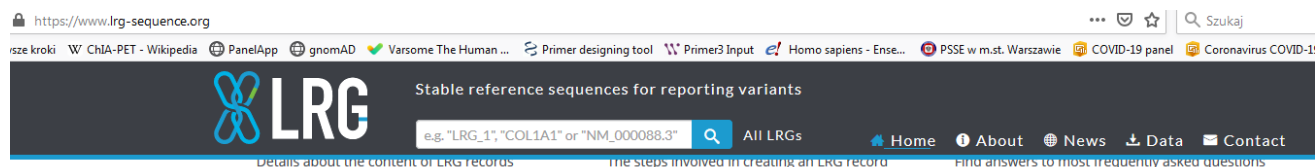
coding transcript - ENST00000357033.8

non-coding transcript - ENST00000383925.1

protein - ENSP00000354923.3

Sekwencje referencyjne – baza LRG

<https://www.lrg-sequence.org/>



What is unique about LRG reference sequence records?

Created specifically for clinical reporting by manual curation

- Include a minimum set of transcripts for reporting at a locus (ideally one)
- Manually curated by expert scientists

Stable

- Include transcript sequences that are stable and independent of changes to transcript models (RefSeq and GENCODE)
- Establish a coordinate system that is independent from upgrades to the reference genome assembly, and provides mappings to present and past assemblies
- Use a unique and stable identifier that is not versioned

Flexible and collaborative

- Allow inclusion of legacy or community-requested reference sequences
- Individual records created at the request of the clinical community and in collaboration with gene-specific experts

Connect the past, present and future of clinical variant reporting

- Part of the effort to rationalise differences in NCBI (RefSeq) and EMBL-EBI (Ensembl/GENCODE) gene sets
- Aim to achieve faster convergence between NCBI (RefSeq) and EMBL-EBI (Ensembl/GENCODE) on key high value annotations to provide a common minimal set of transcripts per gene
- Facilitate unambiguous multi-directional data exchange between NCBI (RefSeq), EMBL-EBI (Ensembl/GENCODE) and the reference genome assemblies (GRCh37, GRCh38)
- Define the relationship between legacy and community-requested sequences and the reference assembly
- Ensure compatibility with genome reference assembly-based NGS variant reporting systems

Well-supported

- Integrated into the Ensembl, NCBI and UCSC genome browsers to allow visualization in genomic context, with all other existing annotations
- Compatible with the Human Genome Variation Society (HGVS) nomenclature and supported by nomenclature checker systems (e.g. Mutalyzer, VariantValidator)
- Generated and maintained by the NCBI and EMBL-EBI

Sekwencje referencyjne – baza LRG

<https://www.lrg-sequence.org/>

← → ↺ ↻ ftp.ebi.ac.uk/pub/databases/lrgex/LRG_614.xml

Często odwiedzane Pierwsze kroki W ChIA-PET - Wikipedia PanelApp gnomAD Varsome The Human ... Primer designing tool Primer3 Input Homo sapiens - Ense... PSSM

 **LRG_614**
Gene: PTPN11

 Fixed Annotation
 Updatable Annotation
 Additional source
 Requester info

PUBLIC
This LRG has been made public (finalised) i.e. the fixed reference sequences will not change

HGNC Gene Symbol (Identifier): PTPN11 ([HGNC:9644](#)) **Made Public:** 10 April 2014 **Last Update:** 23 April 2020

 **Fixed reference sequences in this record**

Number of sequences  **Genomic:** 1 / **Transcript:** 1 / **Protein:** 1

Genomic			Transcript				Protein			
Name	Length	Source	Name	Length	Source	MANE type	Name	Length	Source	CCDS
LRG_614	98,182 nt	NG_007459.1	t1	6,283 nt	NM_002834.3	-	p1	593 aa	NP_002825.3 ENSP00000340944.2	CCDS9163.1

 Download LRG_614 data: [XML](#) - [FASTA](#)

 **Fixed Sequence Annotation**

LRG_614 genomic sequence

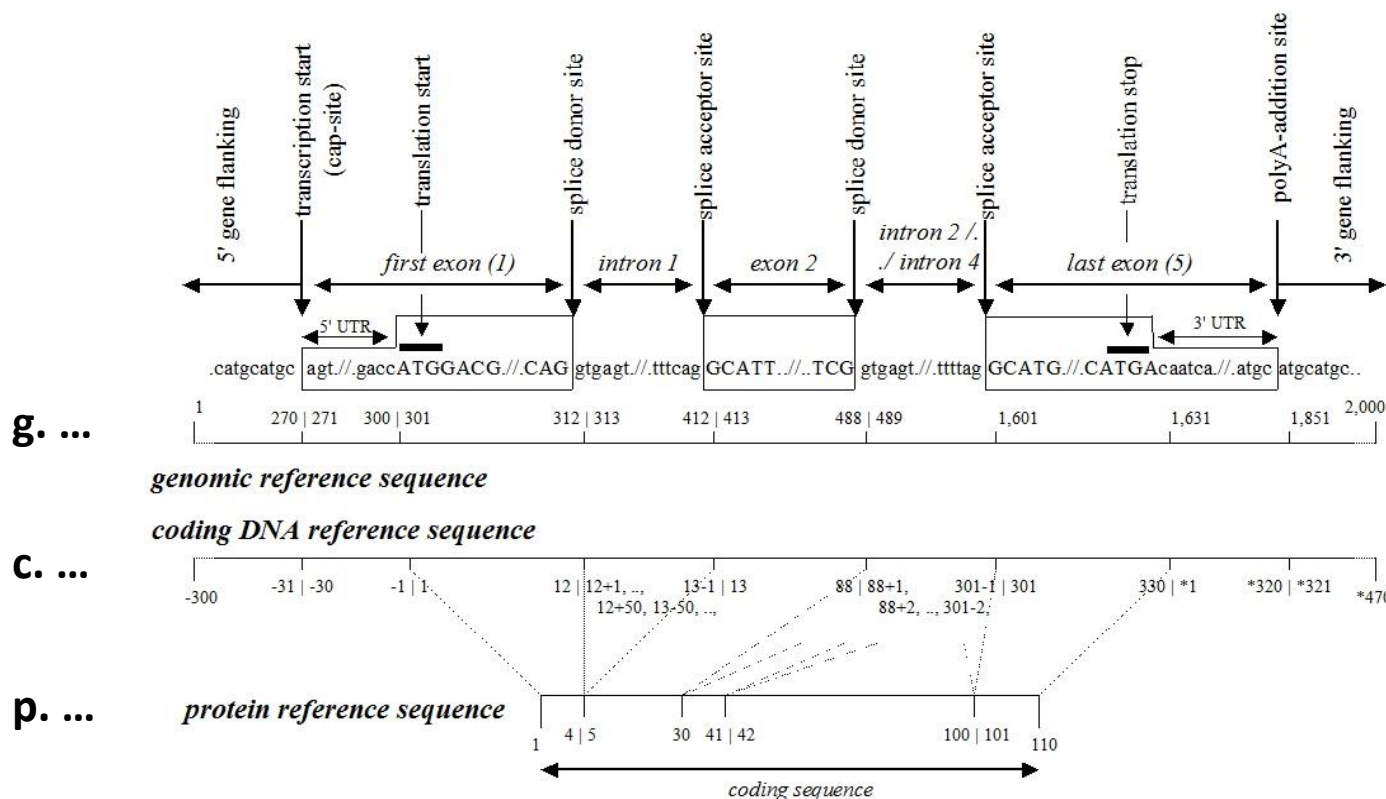
Genomic sequence length: 98,182 nt 

Sequence source: [NG_007459.1](#)

Comment(s):
 The LRG genomic sequence is identical to [NG_007459.1](#) and to the Primary genome assembly (GRCh38).

Human Gene Variation Society (HGVS)

<http://varnomen.hgvs.org/>



Human Gene Variation Society (HGVS)

<http://varnomen.hgvs.org/>

Opis zmian

- substytucja c.34A>T
- zakres zmiany c.34_38del (c.34_38delATGCA)
- delecja c.34del (c.34delA)
- duplikacja c.34dup
- insercje c.34_35insG
- ALE: c.35dup, a nie c.35_36insT
- inwersja c.34_76inv
- konwersja c.123_678conNM_004006.1:c.123_678
- allel []

w przypadku sekwencji powtórzonych zakładamy, że zmianie uległ nukleotyd/-y znajdujące się na końcu 3' sekwencji.

Human Gene Variation Society (HGVS)

<http://varnomen.hgvs.org/>

Nazewnictwo na poziomie białka

- tylko jako dodatek do opisu na poziomie DNA
- jeśli zmiana aminokwasowa nie została zbadana funkcjonalnie (RNA lub białko) zapisuje się ją w nawiasie np. p.(Arg45Trp)
- format p.Arg45Trp, NIGDY: p.45Arg>Trp
ZALECA SIĘ UŻYWANIE KODU TRZYLITEROWEGO
- zmiany ciche: p.= (ew. p.(=), NIGDY: p.Arg45Arg
- zmiany prawdopodobne, nie badane: p.?
- mutacje typu frameshift: p.Arg97Glyfs*26, p.Arg97GlyfsTer26, p.Arg97fs
- mutacje typu extension: p.(Ter110GlnextTer17), p.(*110Glnext*17), p.Met1ext-5 (5'UTR), p.Met1_Leu2insArgSerThrVal (ATG)

Human Gene Variation Society (HGVS)

<http://varnomen.hgvs.org/>

Zapis mutacji zgodny z HGVS

- NM_000277.1:c.1222C>T
- NC_000012.12(NM_000277.1):c.1315+1G>A
- NG_008690.2(NM_000277.1):c.1315+1G>A
 - NP_000268.1:p.Arg408Trp

Zapis genotypu zgodny z HGVS

- NM_000277.1:c.[=];[=]
- NM_000277.1:c.[1222C>T];[=]
- NM_000277.1:c.[1222C>T];[?]
- NM_000277.1:c.1222C>T(;)1315+1G>A
- NM_000277.1:c.1222C>T(;)(1222C>T)
- NM_000277.1:c.[1222C>T];[1222C>T]
- NM_000277.1:c.[1222C>T;1155C>G];[1315+1G>A]

Analiza konkretnego wariantu – zapis genotypu

- NM_000277.1:c.[1222C>T];[1222=]
- NP_000268.1:p.[Arg408Trp];[Arg408=]

Mutalyzer - <https://mutalyzer.nl/>

Welcome to the Mutalyzer website

The aim of this program suite is to support checks of [Sequence Variant Nomenclature](#) according to the guidelines of the [Human Genome Variation Society](#).

Name Checker

The Name Checker takes the complete sequence variant description as input and checks whether it is correct.

Examples: `AB026906.1:c.40_42del` , `NG_012337.1(SDHD_v001):c.274G>T` , `LRG_24t1:c.159dup`

[Check variant description](#)

Syntax Checker

Takes the complete sequence variant description as input and checks whether the syntax is correct.

Position Converter

Converts chromosomal positions to transcript orientated positions and vice versa.

SNP Converter

Allows you to convert a dbSNP rsId to HGVS notation.

Name Generator

A user friendly interface that helps to make a valid HGVS variant description.

Mutalyzer - <https://v3.mutalyzer.nl/>



Welcome!

The following tools are still to be tested/extended/etc. Use them with care. Feedback is greatly appreciated.

Name Checker

Takes a variant description as input and checks whether it is correct.

Examples: `NG_012337.1(NM_003002.2):c.274G>T` `LRG_24:g.5525_5532del`

HGVS Description

NORMALIZE

Position Converter

Converts reference positions to selector orientated positions and vice versa.

Description Extractor

Generates the HGVS variant description from an observed and a reference sequence.

Variant Validator (<https://variantvalidator.org/>)

VariantValidator.org

Home Tools Information Login

Validator



Input Variant Description

Enter variant description in the **HGVS** format or in VCF format as shown below:

- NM_000088.3:c.589G>T
- NC_000017.10:g.48275363C>A
- NG_007400.1:g.8638G>T
- LRG_1:g.8638G>T
- LRG_1t1:c.589G>T
- 17-50198002-C-A
- 17:50198002:C:A
- GRCh38:17:50198002:C:A
- chr17:50198002C>A
- chr17:g.50198002C>A
- chr17(GRCh38):g.50198002C>A

Variant Description:

<accession>.<version>:<type>.<variant description>

☐ GRCh38 (hg38, build38) ☐ GRCh37 (hg19, build19)

Submit


The University of Manchester

Copyright © 2016-2022 VariantValidator Contributors

 UNIVERSITY OF
LEICESTER

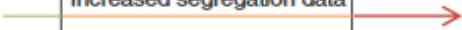
Klasyfikacja wariantów zgodna z ACMG (2015)

<https://www.nature.com/articles/gim201530>

- 1) Patogenne
- 2) Potencjalnie patogenne
- 3) O niepewnym znaczeniu klinicznym
- 4) Potencjalnie łagodne
- 5) Łagodne

Kryteria oparte na:

- częstości występowania wariantu w populacji
- danych obliczeniowych i predykcyjnych
- badaniach funkcjonalnych
- danych dotyczących segregacji wariantu w rodzinie (*de novo*, segregacja z fenotypem)
- obecności innych wariantów u pacjenta (np. *in trans* z patogennym wariantem)

Benign Pathogenic						
	Strong	Supporting	Supporting	Moderate	Strong	Very strong
Population data	MAF is too high for disorder BA1/BS1 OR observation in controls inconsistent with disease penetrance BS2			Absent in population databases PM2	Prevalence in affecteds statistically increased over controls PS4	
Computational and predictive data		Multiple lines of computational evidence suggest no impact on gene /gene product BP4 Missense in gene where only truncating cause disease BP1 Silent variant with non predicted splice impact BP7 In-frame indels in repeat w/out known function BP3	Multiple lines of computational evidence support a deleterious effect on the gene /gene product PP3	Novel missense change at an amino acid residue where a different pathogenic missense change has been seen before PM5 Protein length changing variant PM4	Same amino acid change as an established pathogenic variant PS1	Predicted null variant in a gene where LOF is a known mechanism of disease PVS1
Functional data	Well-established functional studies show no deleterious effect BS3		Missense in gene with low rate of benign missense variants and path. missenses common PP2	Mutational hot spot or well-studied functional domain without benign variation PM1	Well-established functional studies show a deleterious effect PS3	
Segregation data	Nonsegregation with disease BS4		Cosegregation with disease in multiple affected family members PP1	Increased segregation data 		
De novo data				De novo (without paternity & maternity confirmed) PM6	De novo (paternity and maternity confirmed) PS2	
Allelic data		Observed in <i>trans</i> with a dominant variant BP2 Observed in <i>cis</i> with a pathogenic variant BP2		For recessive disorders, detected in trans with a pathogenic variant PM3		
Other database		Reputable source w/out shared data = benign BP6	Reputable source = pathogenic PP5			
Other data		Found in case with an alternate cause BP5	Patient's phenotype or FH highly specific for gene PP4			

	MOCNE	WSKAZUJĄCE
Dane populacyjne	MAF wyższe niż częstość występowania choroby w populacji (BA1/BS1) lub częstość występowania w populacji niezgodna z częstością występowania choroby (BS2)	
Wyniki analiz predykcyjnych	Wariant łagodny: BA1 lub ≥ 2 BS1-BS4 Wariant potencjalnie łagodny: 1 BS1-BS4 i 1 BP1-BP7 lub ≥ 2 BP1-BP7	Predykcje nie wskazują na to, że wariant może wpływać na produkt białkowy/gen (BP4) Mutacje missens w genach, w których mutacje LoF są przyczyną choroby (BP1) Wariant „cichy” nie zaburzający składania mRNA (BP7) Delecje/insercje w ramce odczytu w regionie powtórzeń, nieistotnym dla funkcji białka (BP3)
Analizy funkcjonalne	Nie wykazano znaczącego wpływu na białko/gen (BS3)	
Analiza segregacji	Brak segregacji z chorobą (BS4)	
Analiza <i>de novo</i>		
Dane z alleli		W układzie <i>in trans</i> z wariantem dominującym lub <i>in cis</i> z patogennym (BP2)
Inne bazy danych		Wiarygodne źródło niedostępne publicznie (BP6)
Inne dane		Zidentyfikowany u probanta z inną przyczyną choroby (BP5)

	BARDZO MOCNE	MOCNE	UMIARKOWANE	WSKAZUJĄCE
Dane populacyjne		Częstość wśród pacjentów znacząco wyższa niż w grupie kontrolnej (PS4)	Brak w bazach populacyjnych (PM2)	
Wyniki analiz predykcyjnych	wariant LoF, w genie w którym tego typu mutacje są odpowiedzialne za chorobę (PVS1)	Taka sama zmiana aminokwasowa jak w przypadku znanej mutacji (PS1)	Nowa mutacja missens w kodonie, w którym opisano mutację (PM5) Wariant zmieniający długość białka (PM4)	Wyniki analiz predykcyjnych wskazują na patogenny efekt obecności mutacji (PP3)
Analizy funkcjonalne		Badania funkcjonalne potwierdzające patogenny charakter wariantu (PS3)	Hot-spot lub znana domena funkcjonalna, w której nie występują zmiany łagodne (PM1)	Mutacja missens w genie z małą liczbą łagodnych zmian missens lub częstymi, patogennymi wariantami missens (PP2)
Analiza segregacji		Segregacja z fenotypem w rodzinie (PP1)		
Analiza <i>de novo</i>		<i>De novo</i> (potwierdzone ojcostwo – PS2)	<i>De novo</i> (bez badania ojcostwa – PM6)	
Dane z alleli	Wariant patogenny: 1) PVS1 i • ≥ 1 PS1-PS4 • ≥ 2 PM1-PM6 • 1PM1-PM6 i 1 PP1-PP5 • ≥ 2 PP1-PP5		<i>In trans</i> z wariantem patogennym (PM3)	
Inne bazy danych				Wiarygodne źródło niedostępne publicznie (PP5)
Inne dane		Wariant potencjalnie patogenny: 1) PVS1 i 1 PM1-PM6 2) 1 PS1-PS4 i 1-2 PM1-PM6 3) 1 PS1-PS4 i ≥ 2 PP1-PP5 4) ≥ 3 PM1-PM6 5) 2 PM1-PM6 i ≥ 2 PP1-PP5 6) 1 PM1-PM6 i ≥ 4 PP1-PP5		Feonotyp pacjenta lub historia rodziny odpowiadająca obecności mutacji w danym genie (PP4)

Varsome (https://varsome.com/)

varsome

NM_002834.5(PTPN11):c.922A>G

hg19

Search

Editions

About

Community

Blog

chr12-112915523-A-G (PTPN11:p.N308D)

[Link a publication](#)[Classify](#)[Community contributions](#)[Favorites](#)[UCSC genome browser](#)[Mastermind](#)[TraP Score](#)This variant has been viewed **408** times on VarSome.

Connect with past and future viewers of this variant.

ACMG Classification - Educational use only Version: 8.2.5

[Terms of use](#)[Documentation](#)

Options

CLOSE

Verdict

Pathogenic

Transcript NM_002834.5, canonical, protein length 594, gene PTPN11, missense variant

Rules

Franklin (http://franklin.genoox.com)

Search Page · PTPN11:c.922A>G

PTPN11:c.922A>G

chr12-112477719 A>G | p.Asn308Asp | NM_002834.4 | dbSNP: rs28933386 | UCSC | gnomAD

Improve Interpretation
Add your case details to help Franklin show you the most relevant data 0%

Franklin ACMG Classification Variant Assessment Publications Gene Assessment Associated Conditions

Summary

Region Viewer
Clinical evidence
Predictions
Population Frequencies
Transcripts

Franklin highlights

★ This variant was classified 2 times in Franklin community... | ★ Franklin found 123 variant scope publications | ★ This variant was submitted to Clinvar

Franklin ACMG Classification

Pathogenic

PS4 PM2 PP3 PM1 PP2 PP5

See Details

Conditions Associated with PTPN11

Noonan Syndrome 1 AD
OMIM | Monarch | Decipher

LEOPARD Syndrome 1 AD
OMIM | Monarch | Decipher

Metachondromatosis AD
OMIM | Monarch | Orphanet

Juvenile Myelomonocytic Leukemia N/A
OMIM | Monarch | Orphanet

5 More Conditions

Population Freq

PM2

<0.01%

0% 1% 100%

See all

Relevant Articles

★

123
Variant scope articles
Out of 2528 articles

See all

Prediction

PP3

Revel Deleterious
MetaLR Deleterious
Splice AI N/A

See all Predictions

My Organization Classification

★

No classification

Clinical Evidence

★

Conflict
See Details

Region Viewer

Chr12:112477696-112477742

Select Tracks

Val Ser Asp Tyr Ile Asn Ala Asn Ile Ile Met

T G T T T C A G A T T A C A T C A A T G C A A T A T C A T C A T G G T A A G C T T T G C

Domains

Assessments (Clinvar / Uniprot)

P LP VUS LB B

VARIANT COMMUNITY FEED

Expanding Genomic Clinical Knowledge - Together!

Be the first to add insight for this variant, contribute to the community!

Share Classification Post a Note Link a Publication



Dziękuję za uwagę 😊

dr hab. Monika Gos, prof. IMiD
Zakład Genetyki Medycznej

Kontakt: monika.gos@imid.med.pl, 223277176



Fundusze
Europejskie
Wiedza Edukacja Rozwój



Rzeczpospolita
Polska



Instytut
Matki i Dziecka

Unia Europejska
Europejski Fundusz Społeczny



edu-metgen.imid.med.pl