

WPROWADZENIE DO METOD SEKWENCJONOWANIA NASTĘPNEJ GENERACJI

Monika Gos

Zakład Genetyki Medycznej

Instytut Matki i Dziecka

Warsztaty: Sekwencjonowanie następnej generacji – warsztaty
dla pracowników laboratoriów, 23-24.03.03.2022



Fundusze
Europejskie
Wiedza Edukacja Rozwój



Rzeczpospolita
Polska



Instytut
Matki i Dziecka

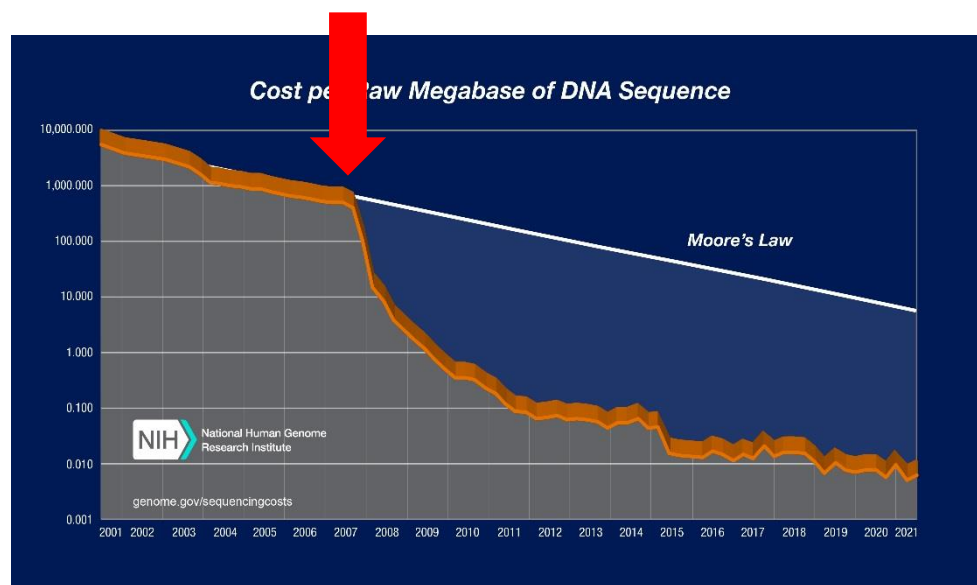
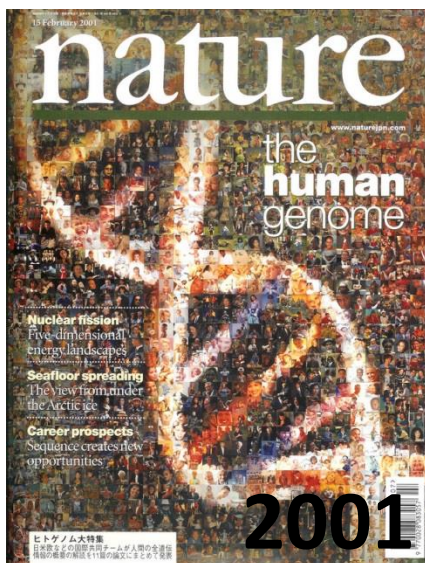
Unia Europejska
Europejski Fundusz Społeczny



edu-metgen.imid.med.pl

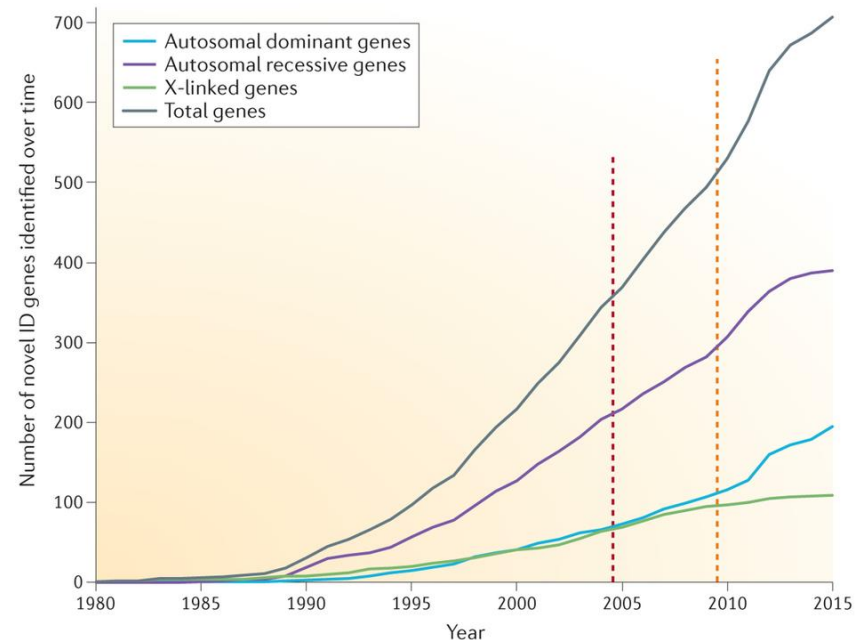
Projekt sekwencjonowania genomu ludzkiego

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. \$



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*

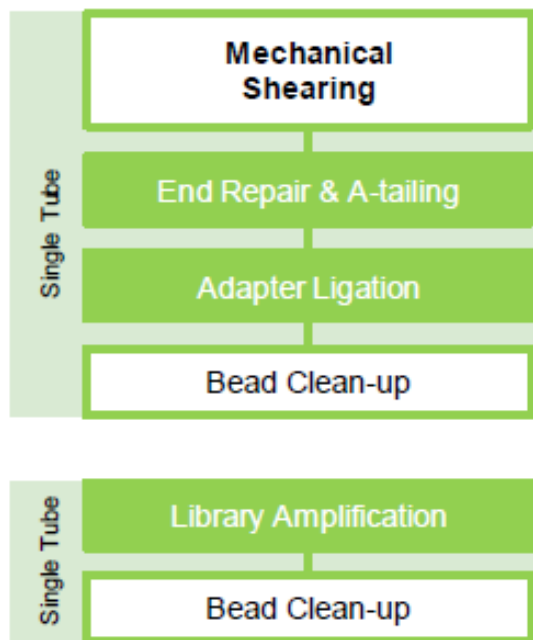
nieswoiste zapalenie jelit

przypominające zespół Crohna - *XIAP*

Protokół

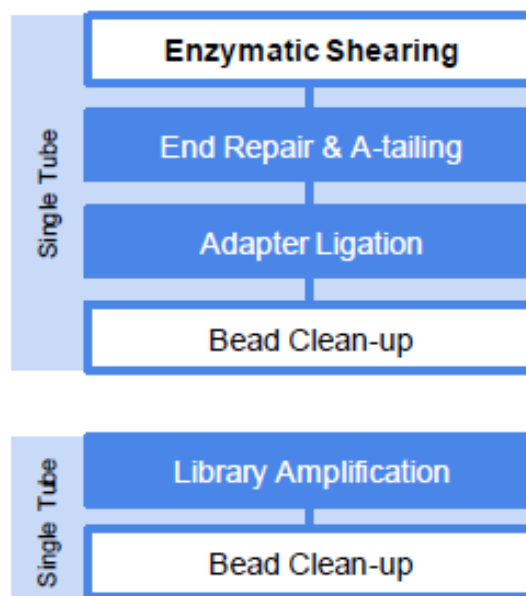
KAPA HyperPrep Kit

Total time ~3.0 hours



KAPA HyperPlus Kit

Total time ~2.5 hours



DNA: 100ng, HQ, ds
 Adaptery: UDI
 Czas fragmentacji: 25 minut
 Kulki: Roche, AMPure Beads
 Praca na lodzie
 Możliwość automatyzacji
 Fragmentacja enzymatyczna: zmniejszona efektywność w przypadku domieszki EDTA, EtOH

PRÓBKA

**Izolacja kwasów
nukleinowych**

Ocena jakości DNA/RNA

**Przygotowanie próbek do
sekwencjonowania**

SEKWENCJONOWANIE



**Fundusze
Europejskie**
Wiedza Edukacja Rozwój



**Rzeczpospolita
Polska**



**Instytut
Matki i Dziecka**

Unia Europejska
Europejski Fundusz Społeczny



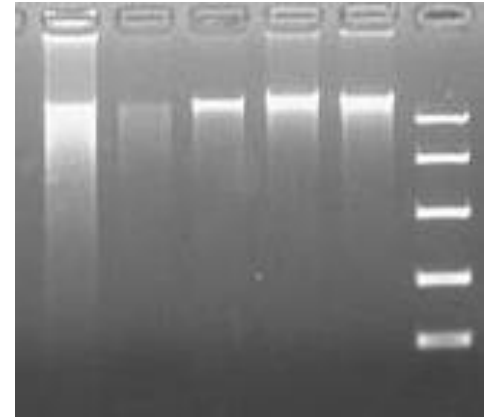
edu-metgen.imid.med.pl

Analiza NGS na poziomie genomowego DNA

- izolacja DNA
- technika w zależności od aplikacji
- krótkie odczyty – może być pofragmentowane; aktualnie izolacja automatyczna (kulki magnetyczne)
- długie odczyty – wysokocząsteczkowy DNA – specjalnie techniki izolacji DNA (kolumny, wytrącanie izopropanolem)

Jakość DNA

- DNA dwuniciowe, nie zdegradowane (żel)
- odpowiednio przechowywane (np. nierozmrażane wielokrotnie, odpowiednie pH i temperatura)
- $OD_{260}/OD_{280} - 1,8-2,0$
- $OD_{260}/OD_{230} \sim 2,0$
- brak wytrąconych kryształów lub resztek kulek magnetycznych
- bez domieszki RNA
- bez domieszki odczynników wykorzystywanych do izolacji DNA (sole guanidynowe, fenol, SDS)
- brak zanieczyszczeń biologicznych (np. hemu, bakterii – wymazówki – 30% DNA z bakterii)
- pomiar ilości i jakości DNA – Nanodrop vs. Qubit/Quantus (!!!)
- QC-PCR – AMEL-XY/GAPDH



PRÓBKA

Fragmentacja

**Przygotowanie biblioteki
(ER, AT, AL, SS)**

Amplifikacja biblioteki

SEKWENCJONOWANIE



Fundusze Europejskie
Wiedza Edukacja Rozwój



**Rzeczpospolita
Polska**



**Instytut
Matki i Dziecka**

Unia Europejska
Europejski Fundusz Społeczny



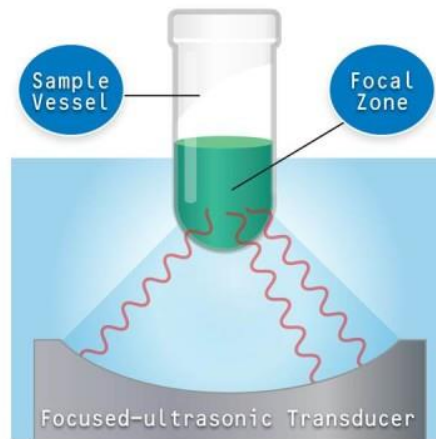
edu-metgen.imid.med.pl

Fragmentacja DNA

Fragmentacja mechaniczna

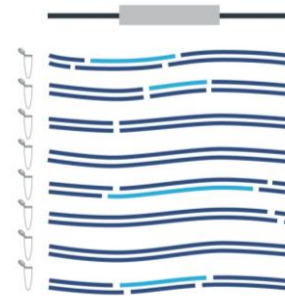


Covaris® M220



Sonikacja / Nebulizacja

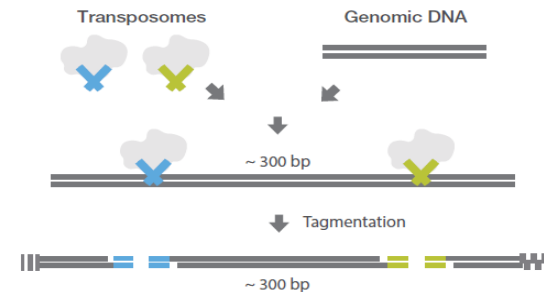
Fragmentacja enzymatyczna



Enzymy
restrykcyjne

Nukleazy
(DNAza I)

Transpozazy



Mieszaniny enzymów (komercja)



Fundusze
Europejskie
Wiedza Edukacja Rozwój



Rzeczpospolita
Polska



Instytut
Matki i Dziecka

Unia Europejska
Europejski Fundusz Społeczny



edu-metgen.imid.med.pl

Fragmentacja DNA

Fragmentacja mechaniczna

- „złoty standard”
 - łatwa kontrola długości fragmentów (amplituda drgań)
 - taki sam protokół niezależnie od ilości DNA i zawartości par GC
-
- znaczny koszt aparatu i materiałów zużywalnych
 - możliwość degradacji DNA
 - stała objętość reakcji
 - czasochłonny (?)

Fragmentacja enzymatyczna

- możliwość fragmentacji dużej liczby próbek
 - zmienna objętość reakcji
 - łatwość wykonania
 - zminimalizowane ryzyko degradacji DNA
-
- zależny od sekwencji DNA
 - wrażliwy na obecność kontaminacji w próbkach
 - może być konieczna optymalizacja metody
 - transpozazy – specyficzne wymagania

PRÓBKA

Fragmentacja

**Przygotowanie biblioteki
(ER, AT, AL, SS)**

Amplifikacja biblioteki

SEKWENCJONOWANIE



Fundusze Europejskie
Wiedza Edukacja Rozwój



**Rzeczpospolita
Polska**



**Instytut
Matki i Dziecka**

Unia Europejska
Europejski Fundusz Społeczny



edu-metgen.imid.med.pl

Przygotowanie biblioteki

DNA



Naprawa końców



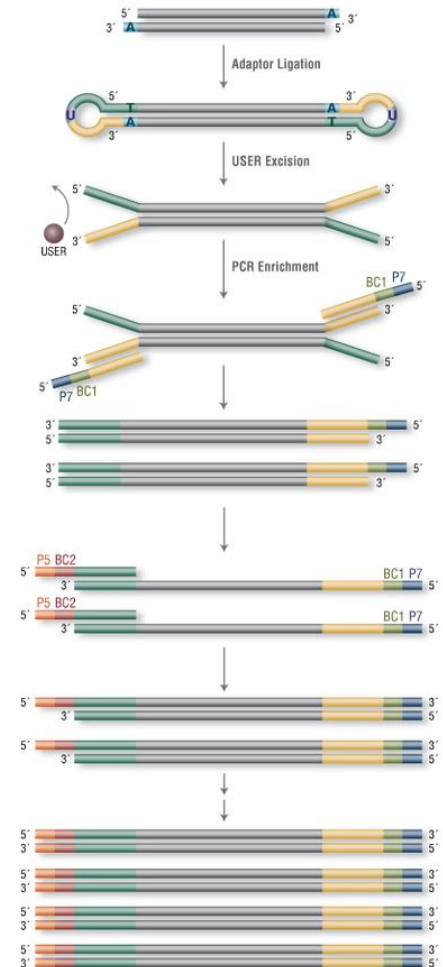
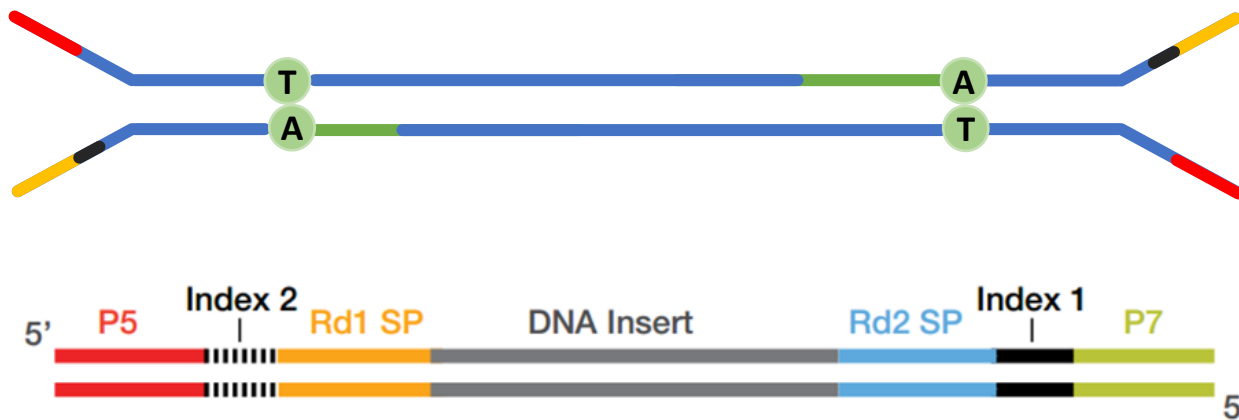
A-tailing



POLIMERAZA, KINAZA

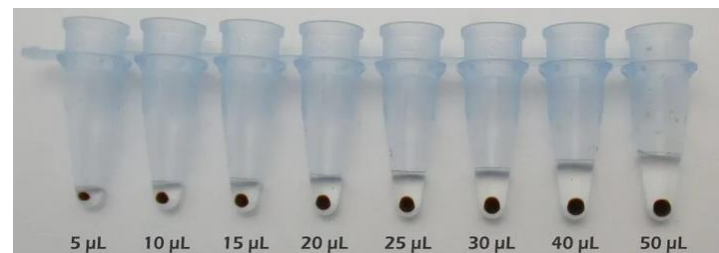
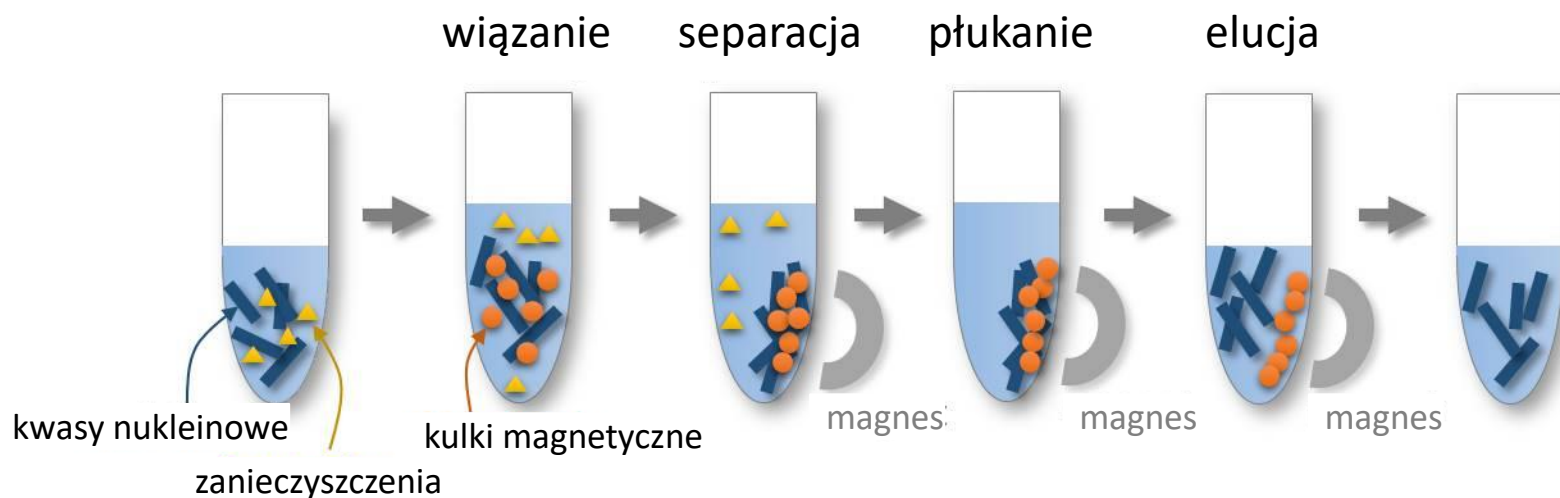
Przygotowanie biblioteki

Ligacja adapterów



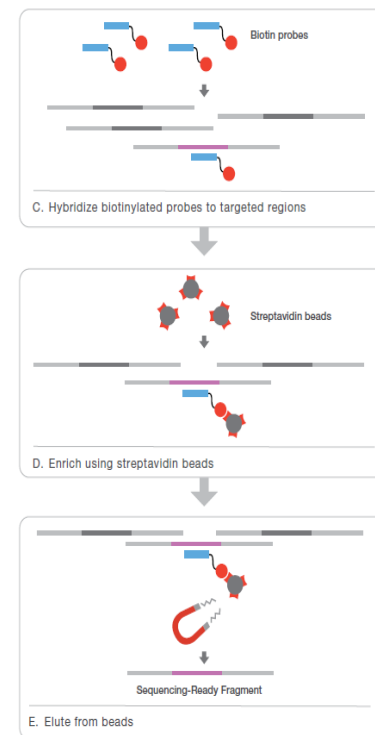
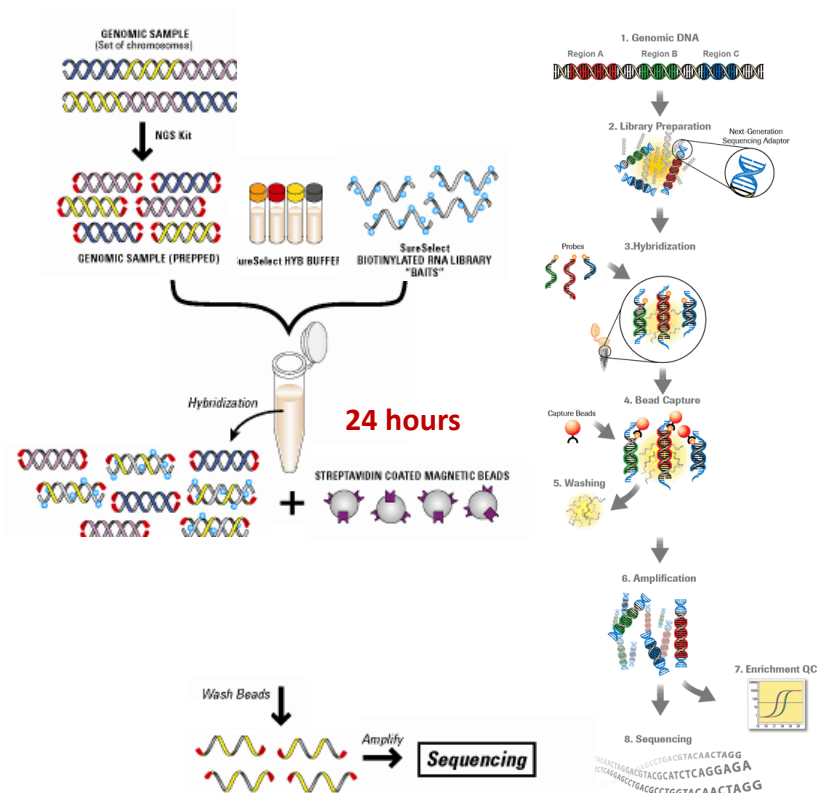
Przygotowanie biblioteki

Oczyszczanie biblioteki - separacja magnetyczna



Sekwencjonowanie celowane - wzbogacenie

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

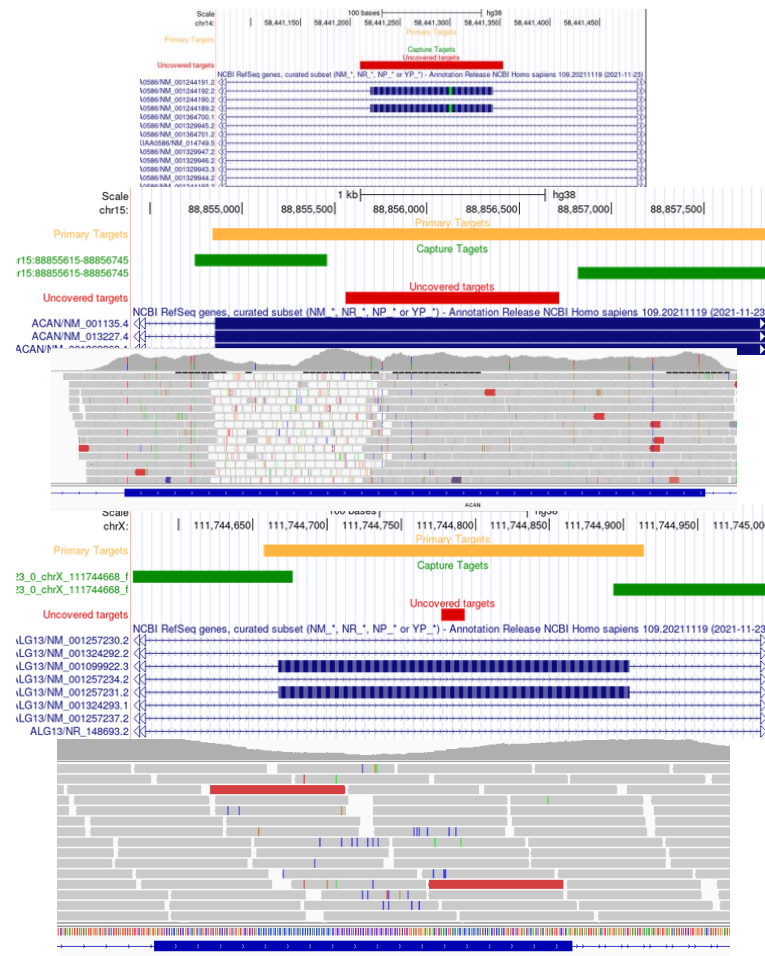


Sekwencjonowanie celowane - wzbogacenie



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	658747	658867
chrX	53648562	53648629

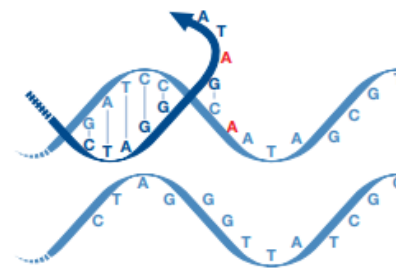
A, B czy C?



Sekwencjonowanie celowane - wzbogacenie



Wykorzystanie
amplikonów –
konieczność
zaprojektowania
odpowiednich
starterów;
allele drop-out
ALE: do reakcji brana
jest mniejsza ilość DNA



PRÓBKA

Fragmentacja

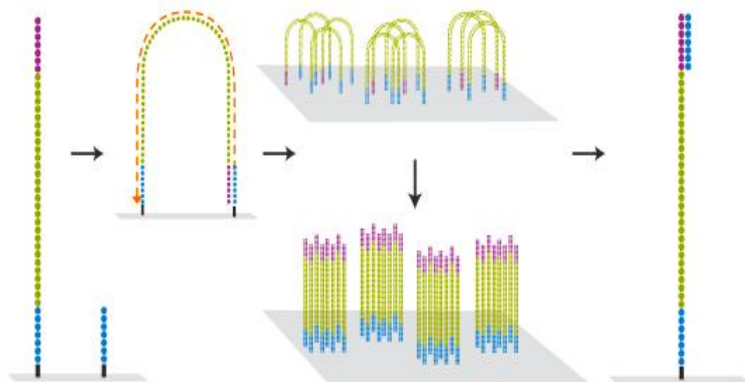
**Przygotowanie biblioteki
(ER, AT, AL, SS)**

Amplifikacja biblioteki

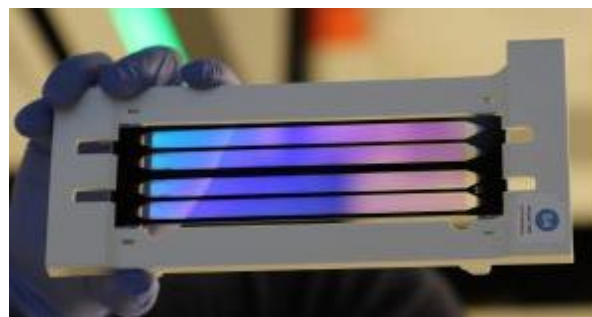
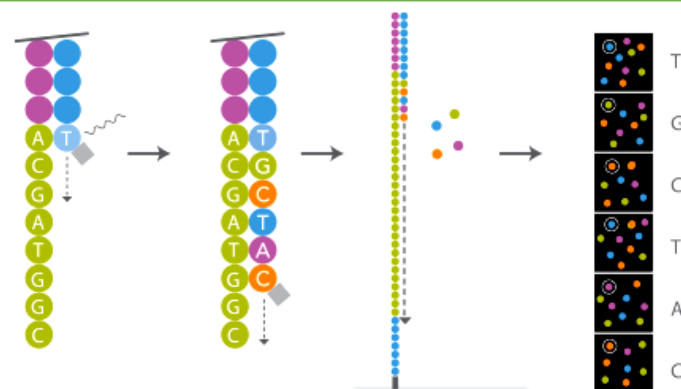
SEKWENCJONOWANIE

Sekwencjonowanie II generacji - Illumina

Amplifikacja klastrowa (mostkowa)



Sekwencjonowanie



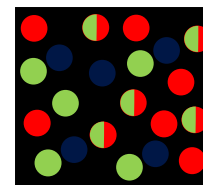
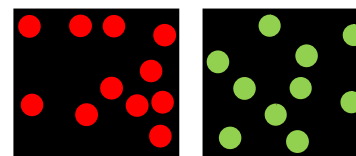
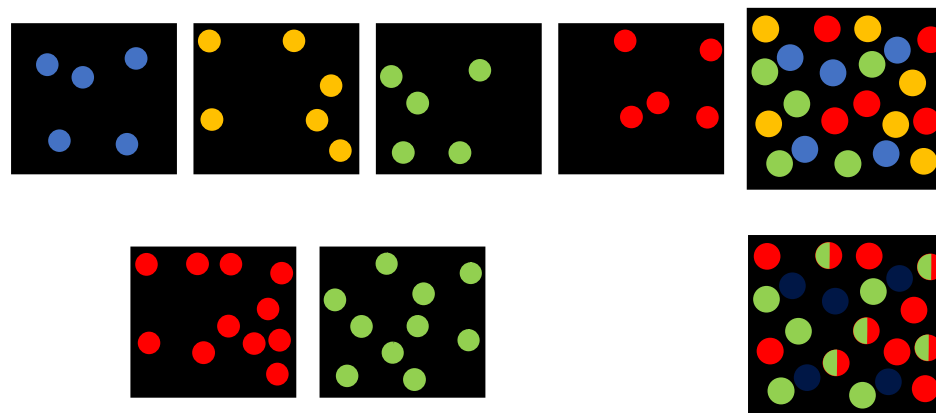
Sekwencjonowanie II generacji - Illumina

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

MiSeq, HiSeq

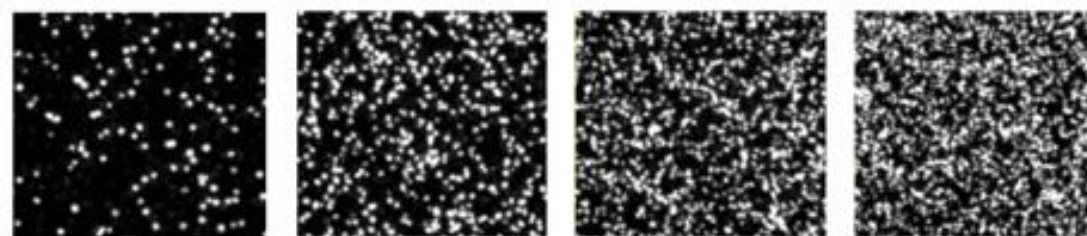
2-Channel Chemistry				
	A	G	T	C
Image 1				
Image 2				
Result	A	G	T	C

NextSeq, NovaSeq



GACCT...

TGAGC...



Underclustered

Optimal Clustering

Overclustered

JPG, BCL → fastq

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)

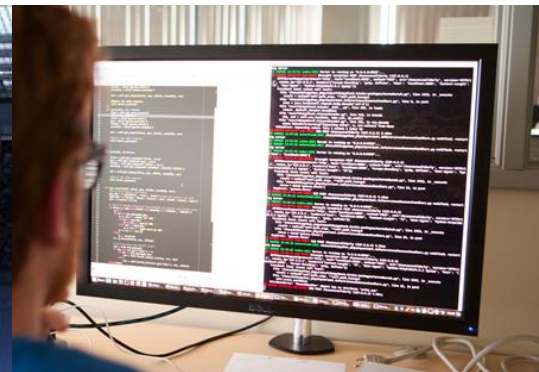


Można analizować **genom**, ale ...
 To 3 miliardy par zasad, z czego >95% to DNA niekodujący,
 o którym jeszcze mało wiemy
 1.5% genomu – część kodująca (25 tys. genów o wielkości 16,3kbp) tzw. **eksom**

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



DANE, które należy zinterpretować



Fundusze Europejskie
Wiedza Edukacja Rozwój



Rzeczpospolita
Polska



Instytut
Matki i Dziecka

Unia Europejska
Europejski Fundusz Społeczny



edu-metgen.imid.med.pl

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



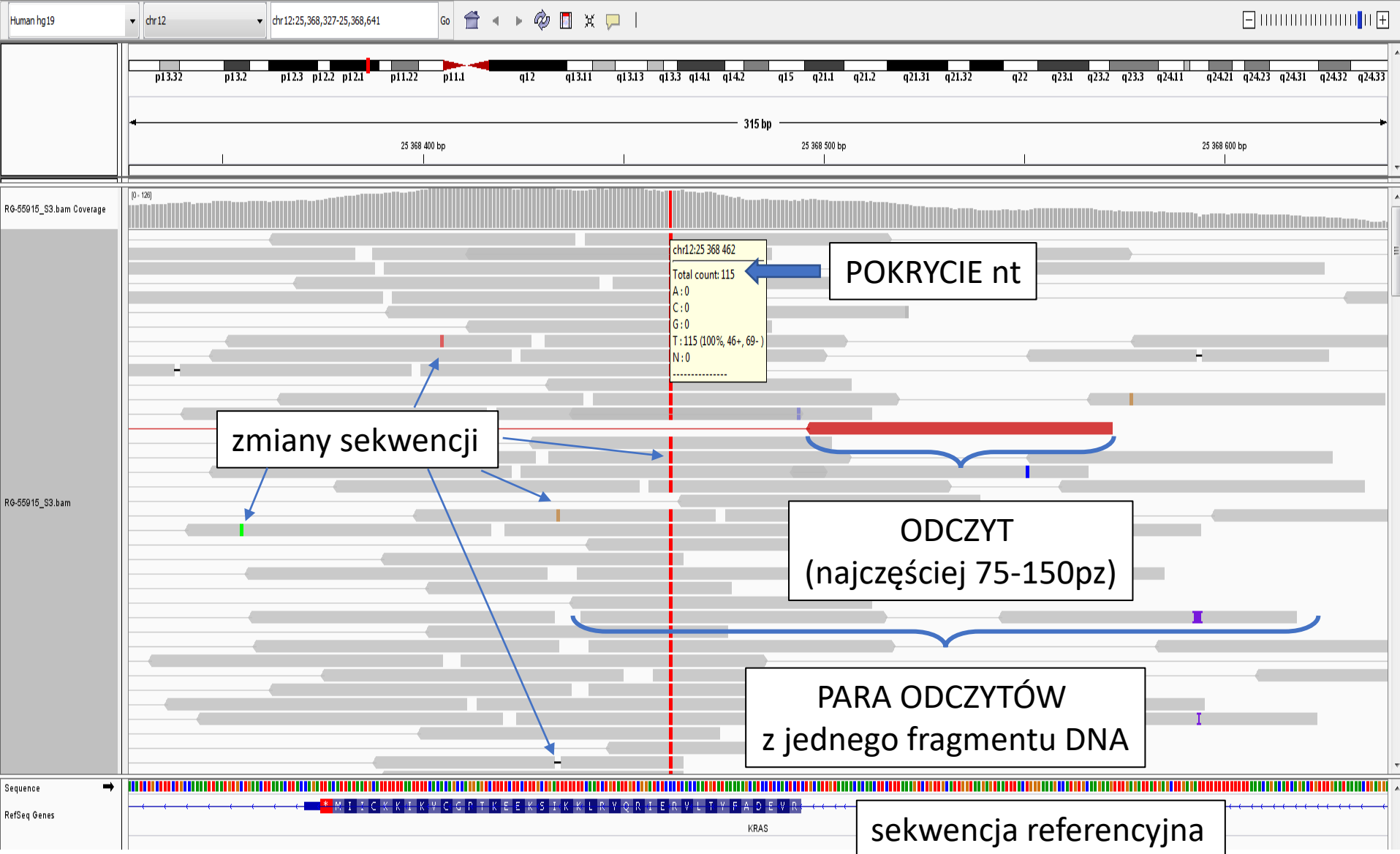
plik *.vcf (m.in. GATK, SAMTools)

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



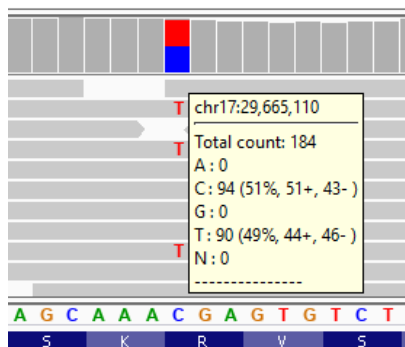
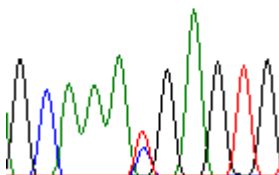
plik Excel (*.tsv, *.txt, *.csv) lub oprogramowanie komercyjne

wybór wariantu związanego z patogenezą choroby



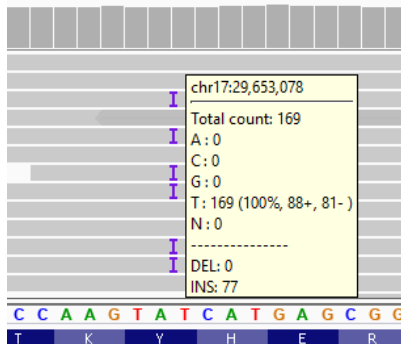
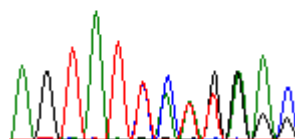
Jakie mutacje identyfikujemy za pomocą sekwencjonowania?

G C A A A T G A G T G
830



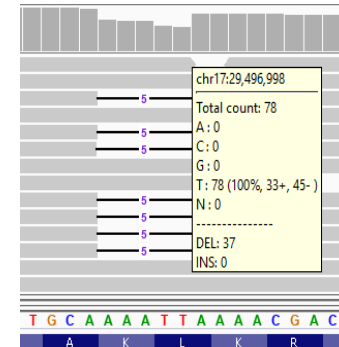
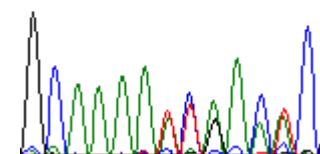
NM_001042492.3:c.6772C>T
(p.Arg2258Ter)

A G T A T C C A G G A C
350



NM_001042492.3:c.5076dup
(p.His1693Serfs*2)

G C A A A A T C A A C T C I
310



NM_001042492.3:c.568_572del
(p.Leu190Thrfs*9)