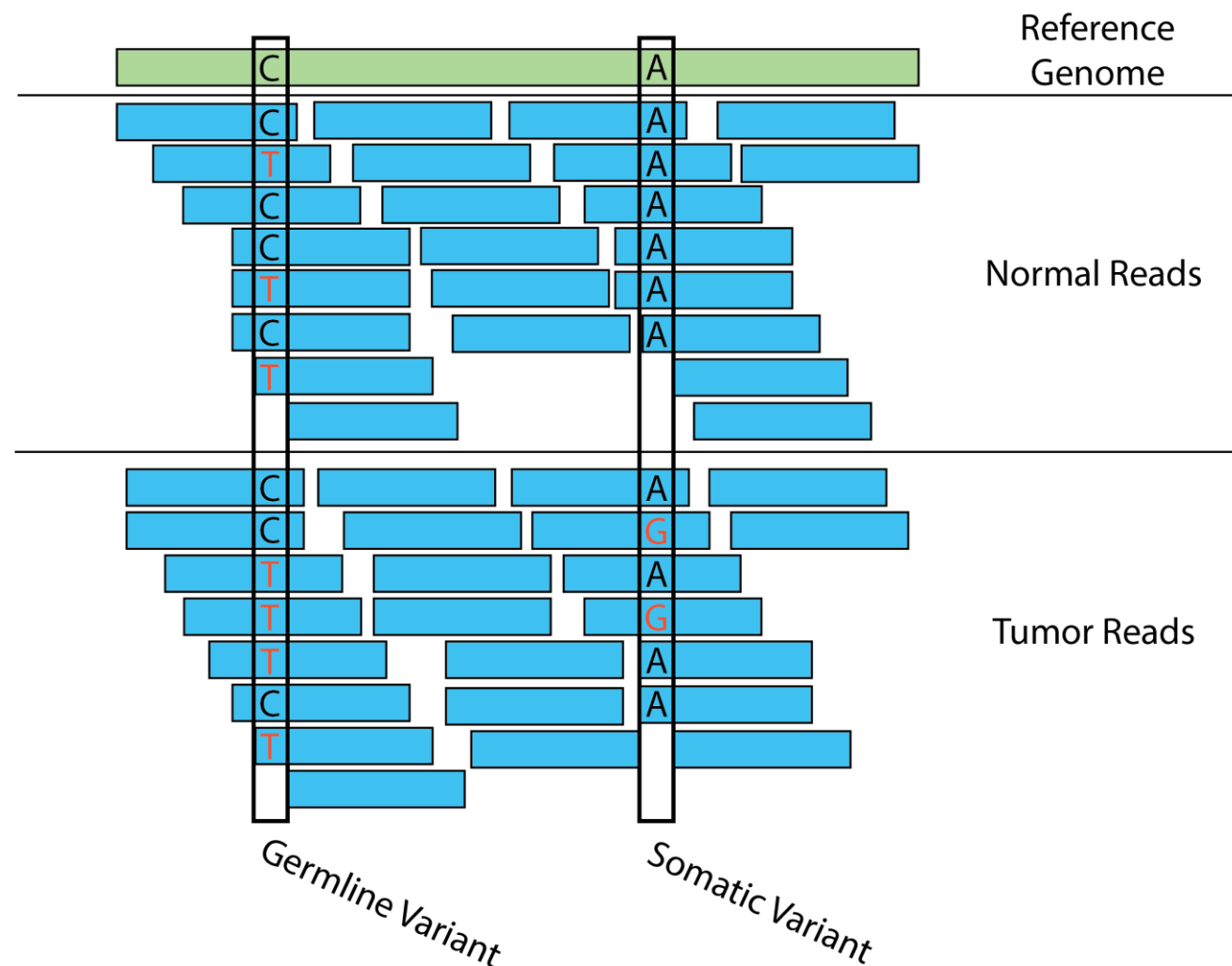


Wykrywanie wariantów (Variant Calling)

ABWG 9 | 3 grudnia 2024

Co to jest Variant Calling (VC)?

- Proces identyfikacji i charakterystyki wariantów genetycznych w sekwencjach DNA na podstawie danych z sekwencjonowania.
- Ma na celu wykrywanie różnic genetycznych pomiędzy badaną sekwencją a genomem referencyjnym.
- Kluczowy etap w analizie NGS.



Rodzaje wariantów genetycznych

Polimorfizmy pojedynczych nukleotydów (*Single Nucleotide Polymorphisms, SNPs*)

- Zmiana jednego nukleotydu w określonej pozycji genomu.
- Najczęstszy typ wariantu genetycznego.

Insercje i delecje (indels)

- Wstawienia lub usunięcia jednego lub więcej nukleotydów.
- Mogą prowadzić do przesunięcia ramki odczytu.

Warianty strukturalne

- CNVs (Copy Number Variations): zmiany liczby kopii fragmentów DNA.
- Translokacje: przeniesienie fragmentu DNA między chromosomami.
- Inwersje: odwrócenie orientacji fragmentu DNA w chromosomie.

	SNP1	SNP2	SNP3	SNP4	SNP5	SNP6									
Haplotype 1	C	A	G	A	T	C	G	C	A	T	C	T	G	T	[35%]
Haplotype 2	C	A	G	A	T	C	G	C	T	G	A	A	T	G	[30%]
Haplotype 3	C	G	G	A	T	T	G	C	T	G	C	A	T	G	[15%]
Haplotype 4	C	G	G	A	T	T	G	C	T	G	C	A	T	G	[10%]
Several other haplotypes														[10%]	

Indel examples

wild-type sequence

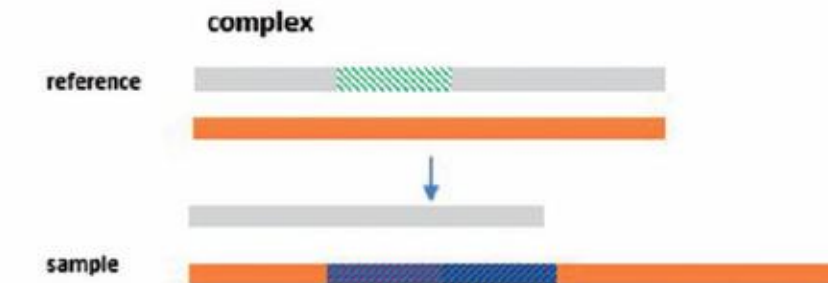
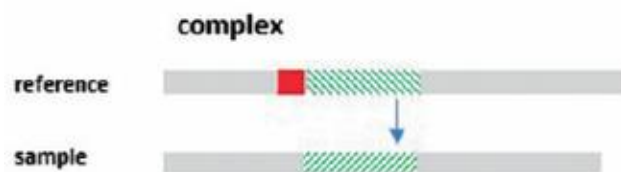
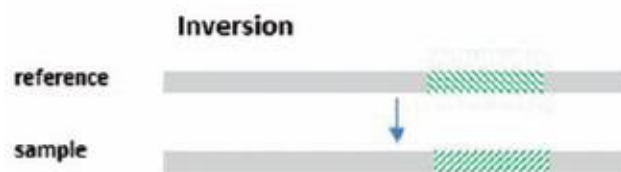
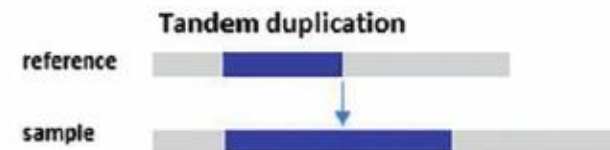
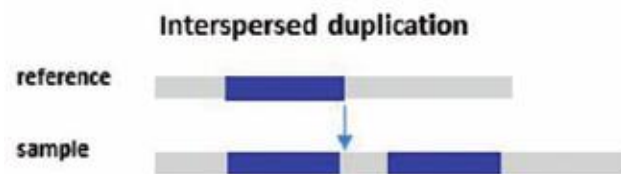
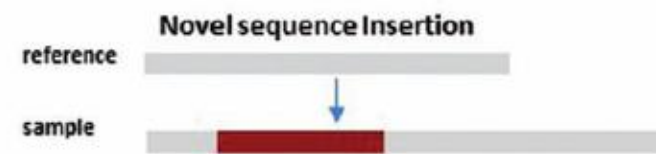
ATCTTCAGCCATAAAAGATGAAGTT

3 bp deletion

ATCTTCAGCCAAAGATGAAGTT

4 bp insertion (orange)

ATCTTCAGCCATATGTGAAAGATGAAGTT



Wpływ wariantów na funkcje genów

Mutacje w regionach kodujących

- Synonimiczne (ciche): nie zmieniają aminokwasu w białku
- Niesynonimiczne
 - Missense: zmiana aminokwasu na inny
 - Nonsense: powstanie kodonu STOP (skrócenie białka)

Mutacje w regionach niekodujących

- wpływ na regulację ekspresji genów (promotory, enhancery)
- zmiany w miejscach splicingowych wpływające na składanie mRNA

Zastosowanie w badaniach podstawowych

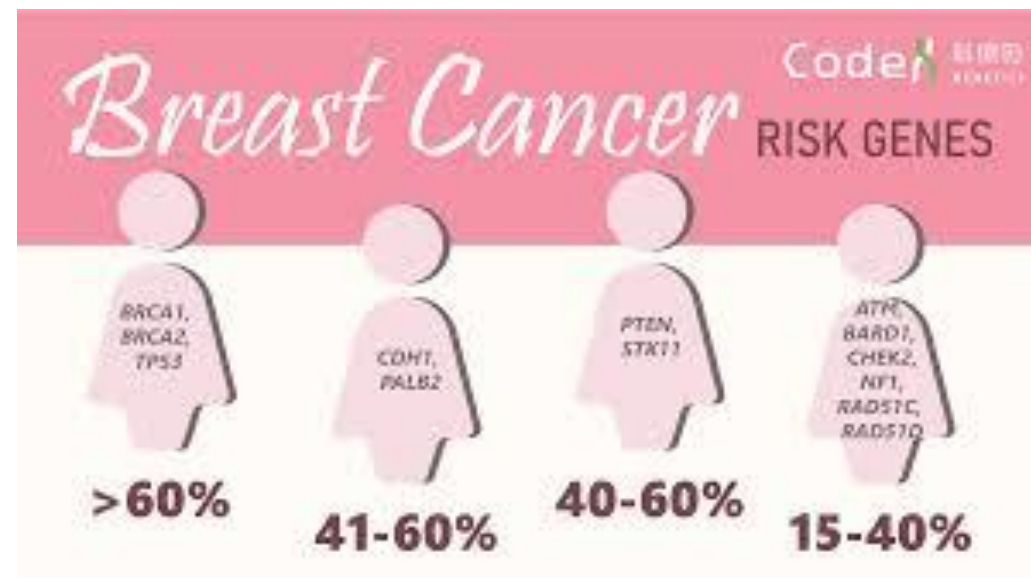
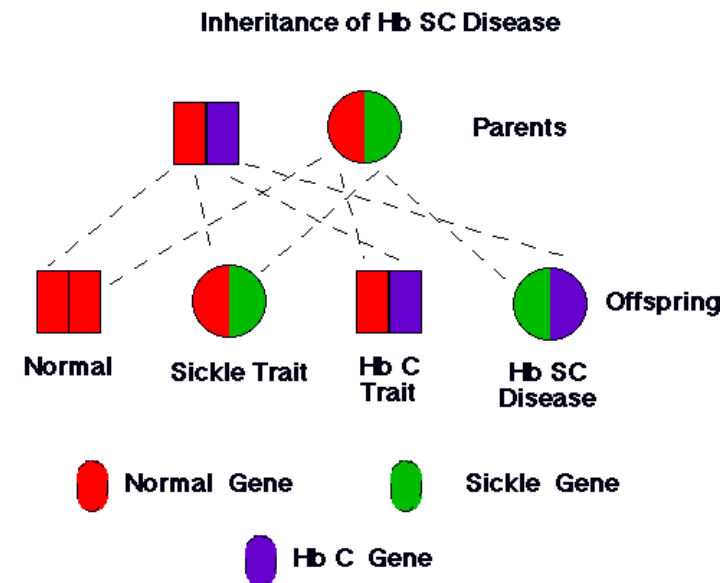
Zróżnicowanie genetyczne
populacji

Filogeneza i badania ewolucyjne

Identyfikacja genów
odpowiedzialnych za cechy
fenotypowe

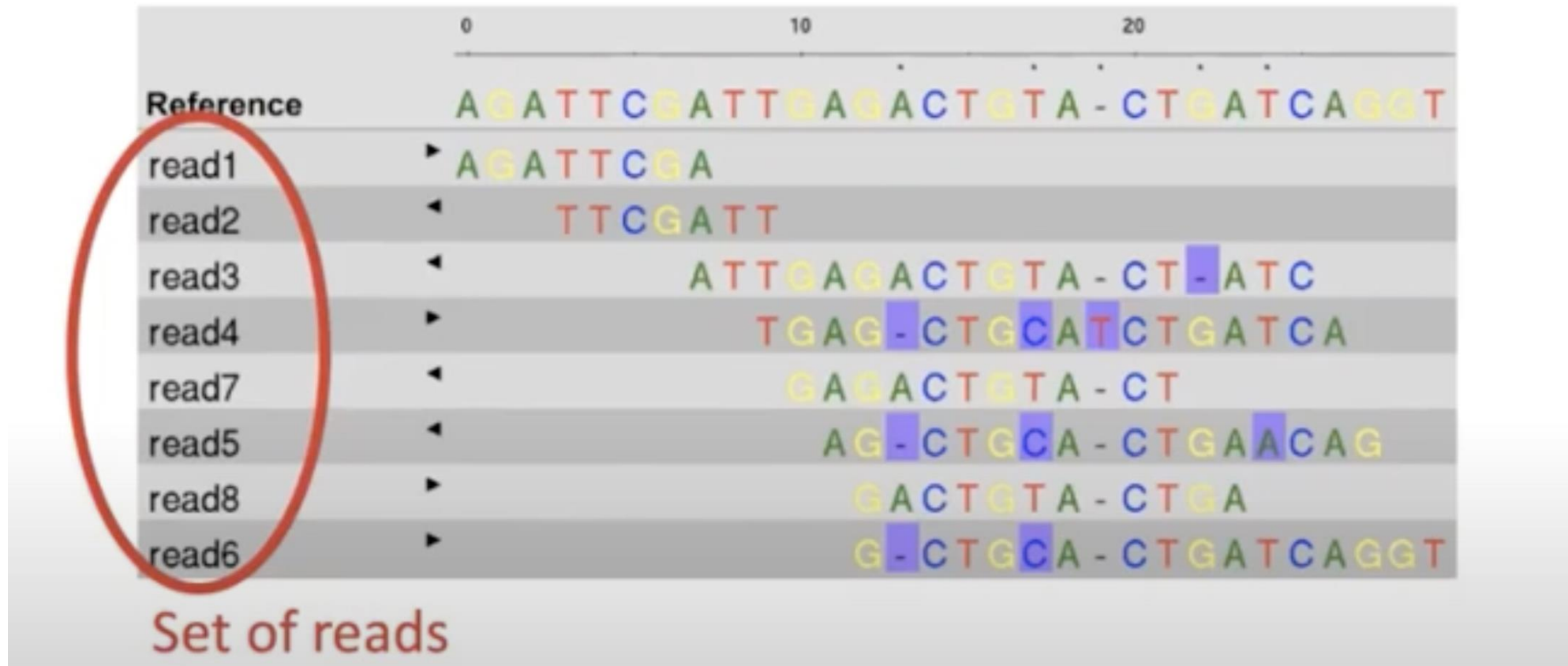
Przykłady z badań klinicznych

- Choroby monogenowe (np. mukowiscydoza, anemia sierpowata)
- Onkogenetyka (np. *HER2* wraku piersi)
- Medycyna spersonalizowana i farmakogenomika

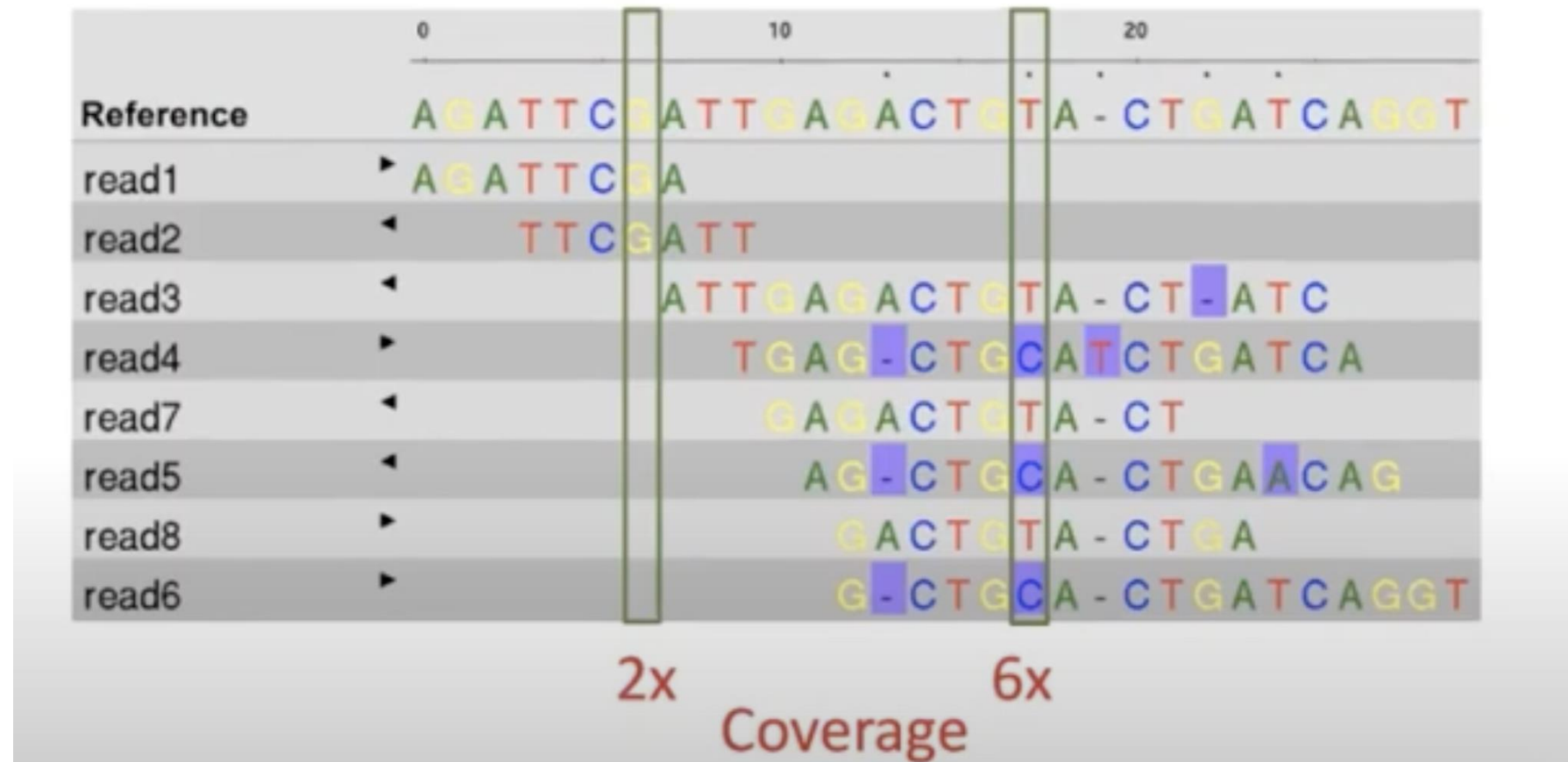


Variant Calling w praktyce

Zmapowane odczyty



Pokrycie genomu



Wykrywanie wariantów

	0	10	20																						
Reference	A G A T T C G A T T G A G A C T G T A - C T G A T C A G G T																								
read1	▶	A G A T T C G A																							
read2	◀	T T C G A T T																							
read3	◀											A T T G A G A C T G T A - C T - A T C													
read4	▶											T G A G - C T G C A T C T G A T C A													
read7	◀											G A G A C T G T A - C T													
read5	◀											A G - C T G C A - C T G A A C A G													
read8	▶											G A C T G T A - C T G A													
read6	▶											G - C T G C A - C T G A T C A G G T													

Variations: Deletion & Single-nucleotide variant

Błędy sekwencjonowania

	0	10	20																					
Reference	A G A T T C G A T T G A G A C T G T A - C T G A T C A G G T																							
read1	▶	A G A T T C G A																						
read2	◀	T T C G A T T																						
read3	◀	A T T G A G A C T G T A - C T												-	A T C									
read4	▶	T G A G - C T G C A T C						G A T C A																
read7	◀	G A G A C T G T A - C T																						
read5	◀	A G - C T G C A - C T G A		A C A G																				
read8	▶	G A C T G T A - C T G A																						
read6	▶	G - C T G C A - C T G A		T C A G G T																				

Sequencing errors: Insertions, deletions & basecalling errors

Transformacja danych

Alignment

	0	10	20
Reference	AATTCTGACTGTA-CTGATCAGGT		
read1	AATTCTGA		
read2	TTCGATT		
read3	ATTGACTGTA-CT-ATC		
read4	TGAG-CTGCACTGATCA		
read7	GACTGTA-CT		
read5	AG-CTGCA-CTGAACAG		
read8	GACTGTA-CTGA		
read6	G-CTGCA-CTGATCAGGT		



List of Variants

CHR	POS	ID	REF	ALT	GT
chr1	12	.	GA	G	0/1
chr1	17	rs123	T	C	0/1

Genotype (GT):
0/0: hom. reference
0/1: heterozygous
1/1: hom. alternative

format VCF/BCF

Formaty danych

	0	10	20																			
Reference	AGATTCTGACTGTA-CTGATCAGGT																					
read1	▶ AGATTCTGA																					
read2	◀ TTCGATT																					
read3	◀ ATTGAGACTGTA-CT-ATC																					
read4	▶ TGAG-CTGCACTCTGATCA																					
read7	◀ GAGACTGTA-CT																					
read5	◀ AG-CTGCA-CTGAACAG																					
read8	▶ GACTGTA-CTGA																					
read6	▶ G-CTGCA-CTGATCAGGT																					

SAM/BAM file

QNAME	FLAG	RNAME	POS	MAPQ	CIGAR
read2	17	chr1	3	60	7M
read4	1	chr1	9	30	4M1D5M1I7M



VCF/BCF file

CHR	POS	ID	REF	ALT	GT
chr1	12	.	GA	G	0/1
chr1	17	rs123	T	C	0/1