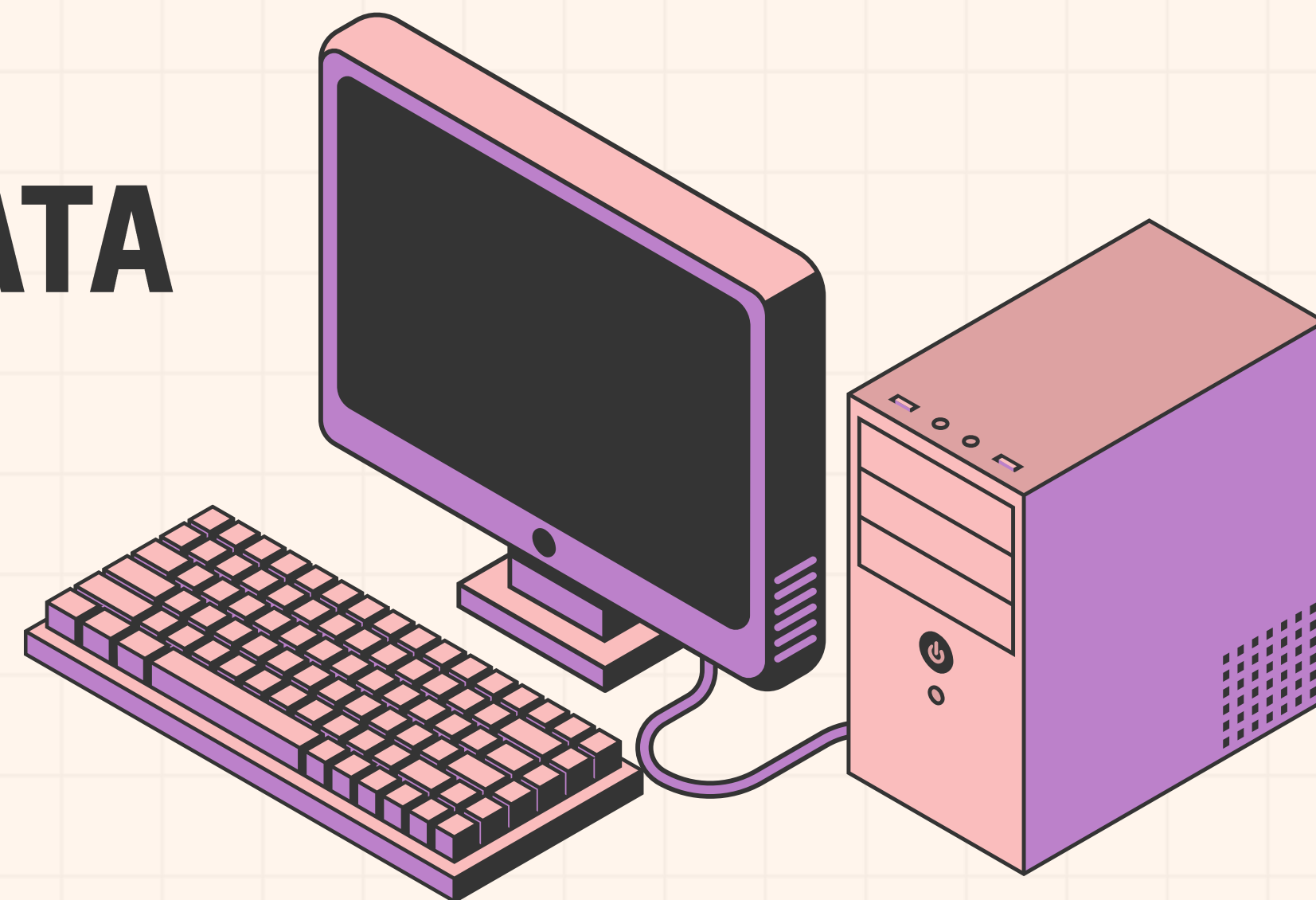


PRIMJENA BIOINFORMATIČKIH ALATA U DIJAGNOSTICI AKUTNIH LEUKEMIJA



Monika Kolundžić

Odjel za laboratorijsku hematologiju i koagulaciju, Laboratorij za molekularnu hematologiju

Klinički zavod za laboratorijsku dijagnostiku

KBC Zagreb

PLAN PREDAVANJA

- UVOD U BIOINFORMATIKU
- SEKVENCIRANJE DRUGE GENERACIJE
- NGS I AKUTNE LEUKEMIJE
- ANALITIČKE FAZE NGS-A
- BIOINFORMATIČKA OBRADA PODATAKA
- PRIMARNA ANALIZA
- SEKUNDARNA ANALIZA
- TERCIJARNA ANALIZA



UVOD U BIOINFORMATIKU

grč. bios: život, engl. informatics: informatika

- disciplina koja koristi tehnike i znanja informatike, uključujući primijenjenu matematiku, računarstvo i statistiku, u svrhu analize, organizacije i prikaza bioloških podataka
- 1960.-te
- visokoprotodne tehnologije – omiksonske studije – sustavno proučavanje bioloških molekula
- sekvenciranje sljedeće generacije – NGS
- generiranje velike količine podataka
- **BIOINFORMATIČKI ALATI** – pretvaranje masovnih sirovih podataka u razumljive i interpretabilne nalaze



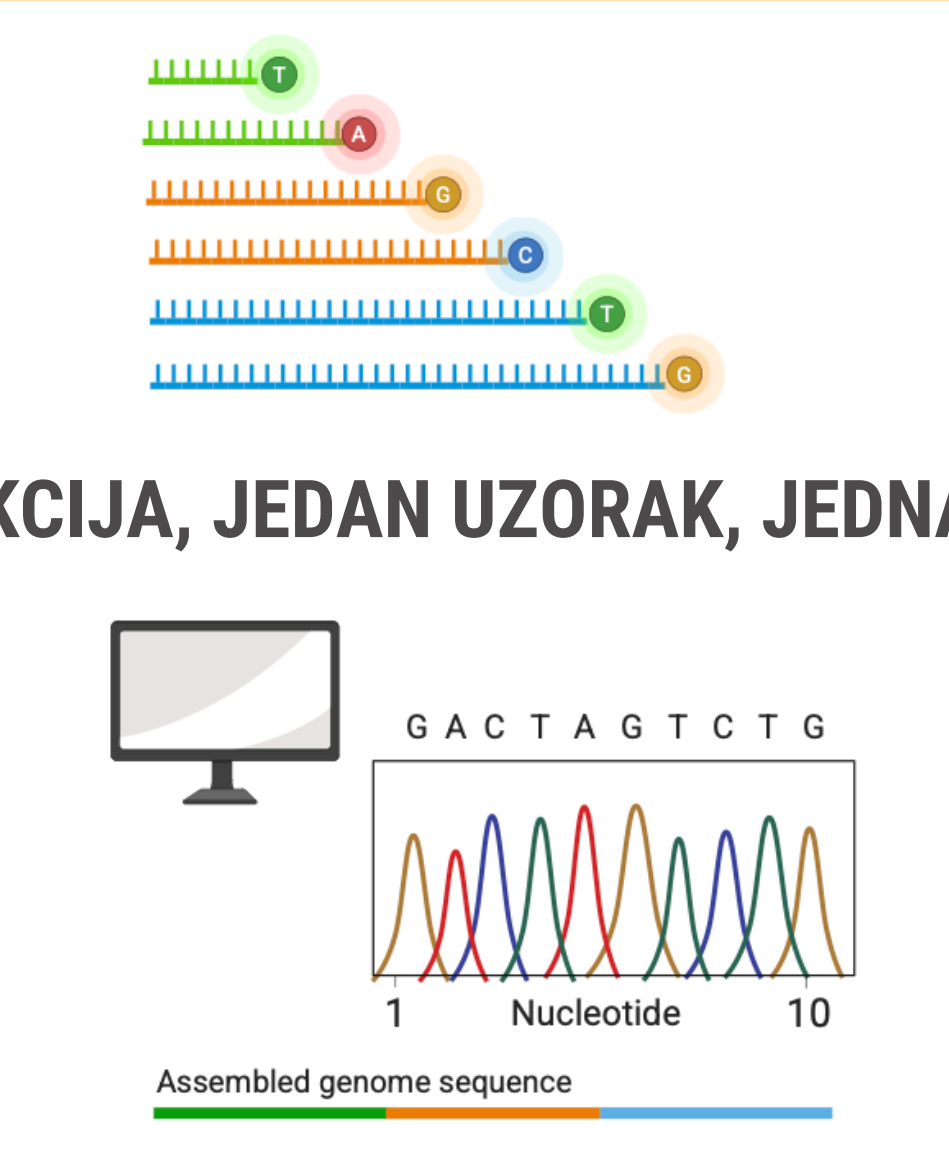
SANGER

RAZLIKA?



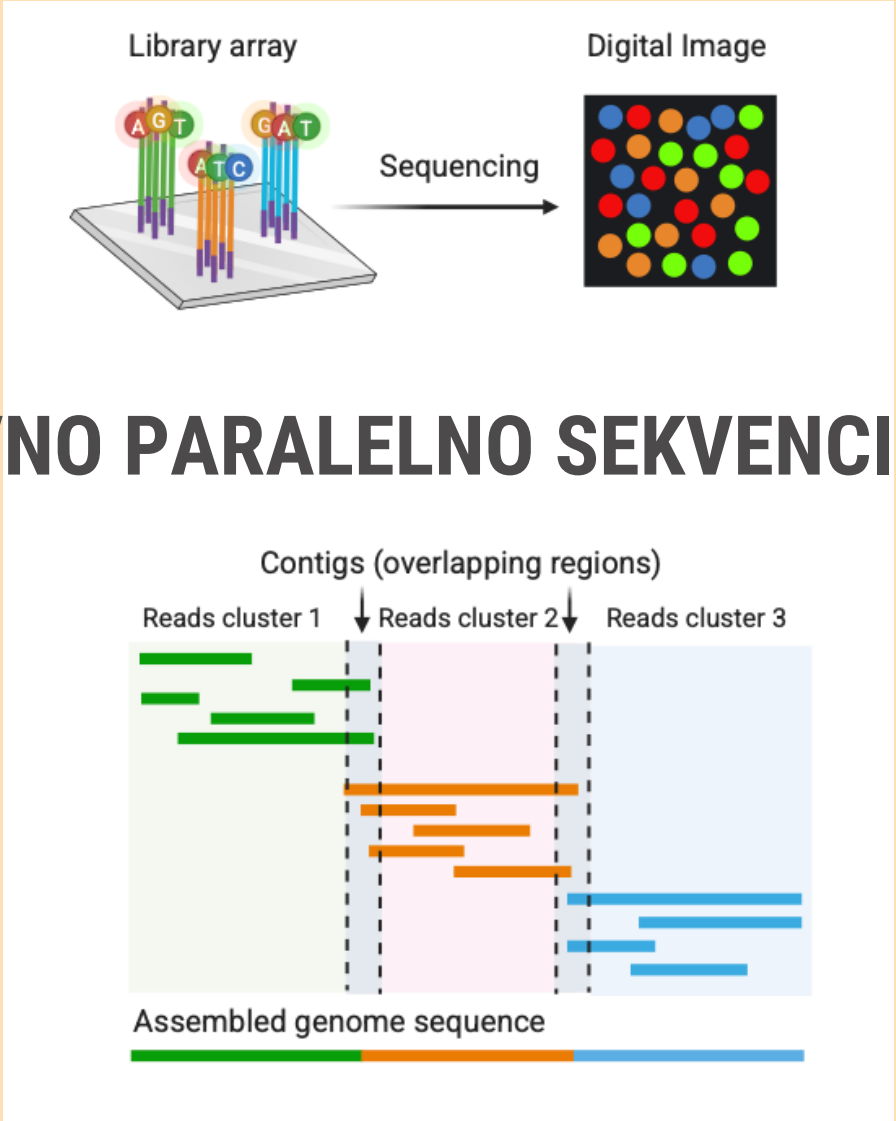
NGS

jedan slijed nukleotida veličine do 1000 pb



JEDNA REAKCIJA, JEDAN UZORAK, JEDNA SEKVENCA

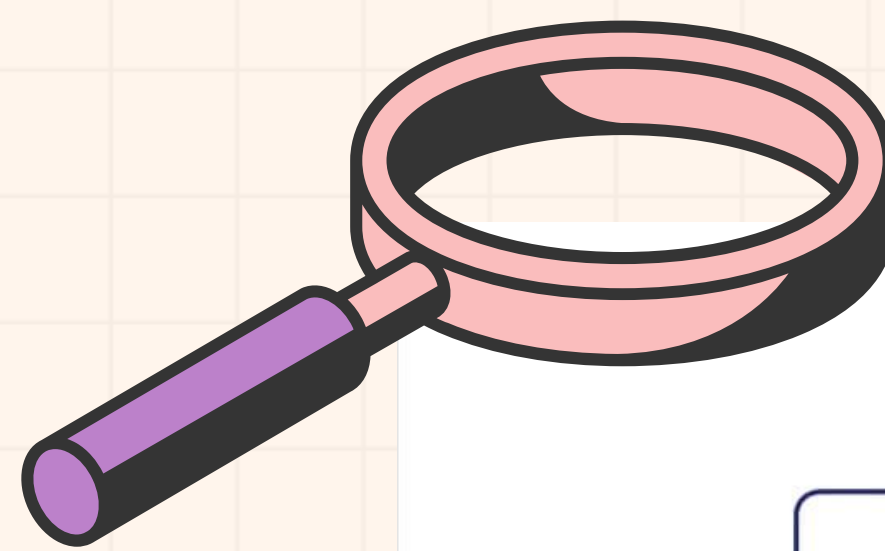
mogućnost sekvenciranja više gena i više različitih uzoraka
ISTOVREMENO



MASIVNO PARALELNO SEKVENCIRANJE

“SEQUENCING BY SYNTHESIS (SBS)”

NGS TEHNOLOGIJA



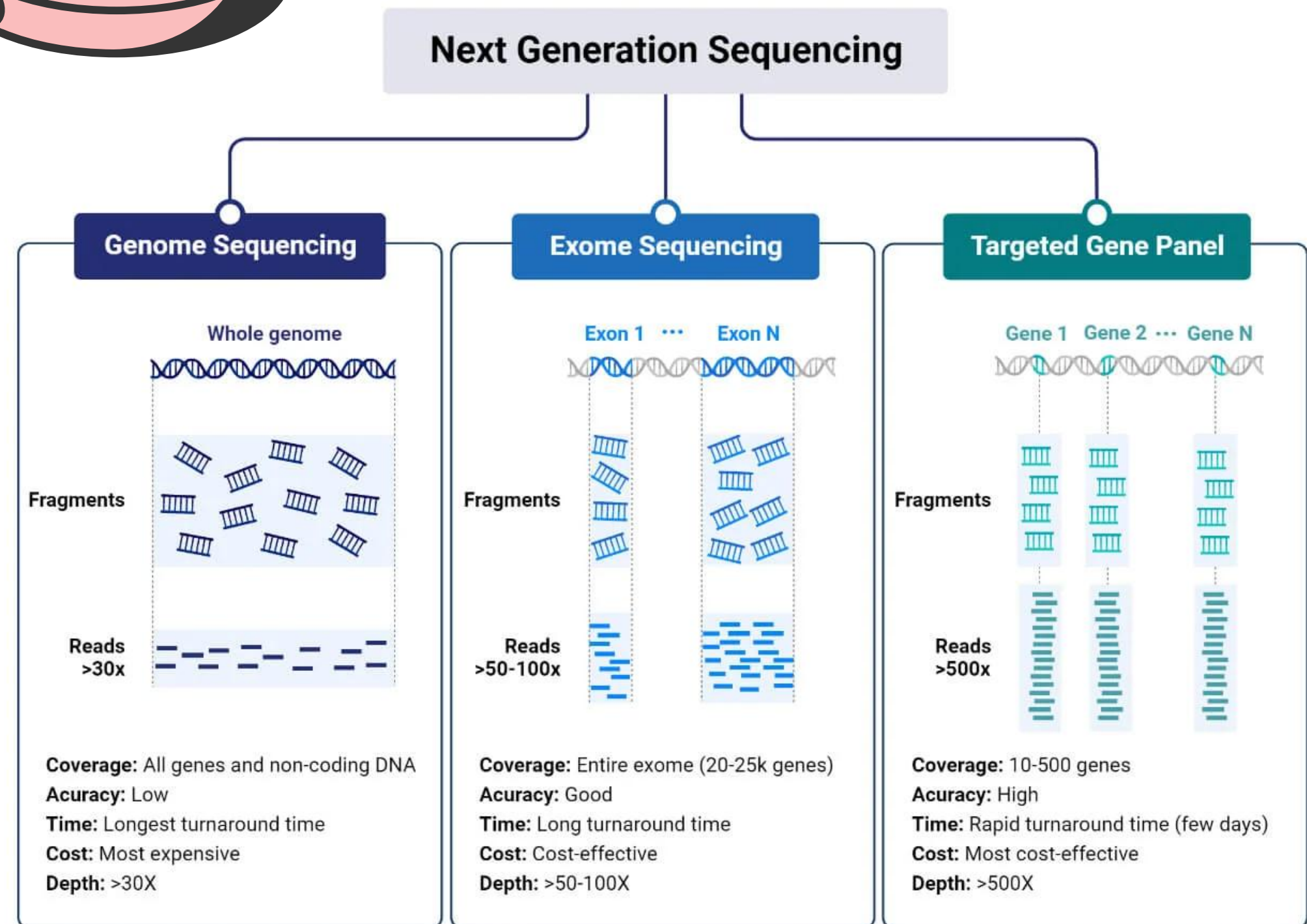
Next Generation Sequencing

- odabir pristupa sekvenciranja ovisi o kliničkom/istraživačkom pitanju

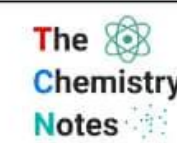
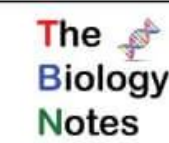
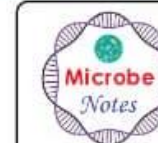
- tri su varijable:

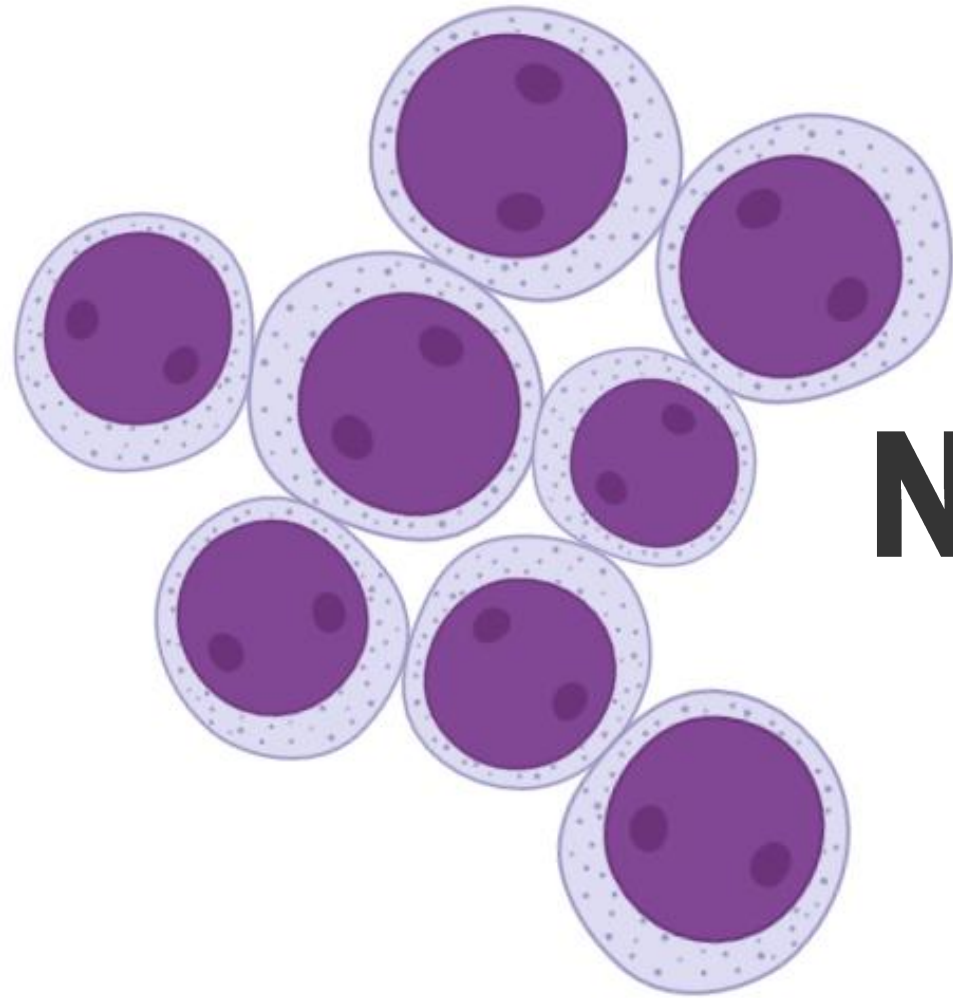
1. osjetljivost metode
2. balansiranje troškova
3. složenost bioinformatičke obrade rezultata

- Osjetljivost = dubina čitanja pojedine baze
- dubina čitanja = broj unikatnih isječaka gena (engl. *read*) u kojima se nalazi točno ta baza.
- veći broj *read*-ova koji se preklapaju i pokrivaju određenu genomsku poziciju = veća osjetljivost i specifičnost određivanja varijante

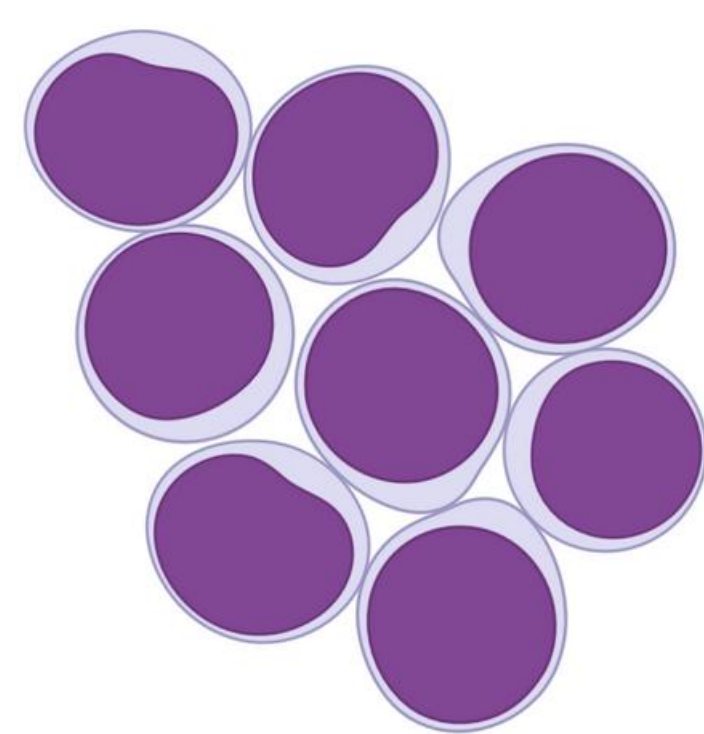
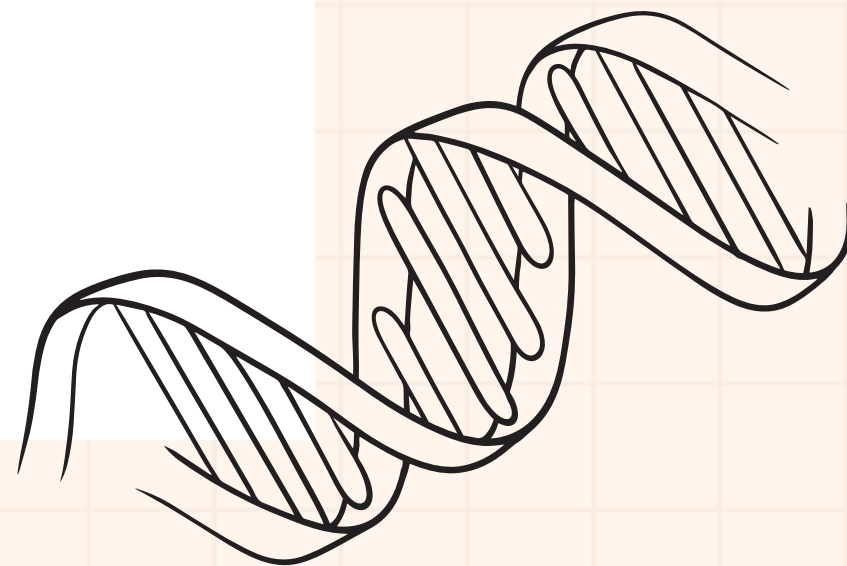


Template adapted from: Dr. Roshini Abraham
Clinical Immunologist at Nationwide Children's Hospital





NGS I AKUTNE LEUKEMIJE



DIJAGNOSTIČKA OBRADA

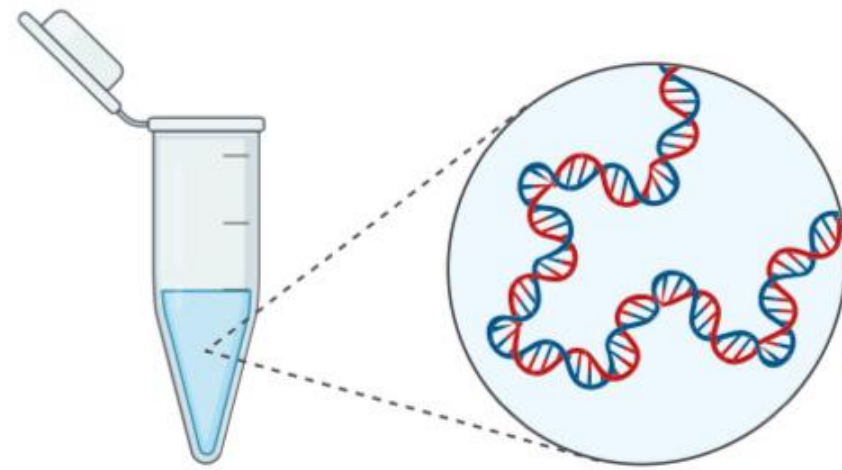
- ciljani panel gena
- optimalni omjer osjetljivosti, informativnosti, troškova i složenosti bioinformatičke obrade podataka
- u svrhu postavljanja precizne dijagnoze, stratifikacije rizika bolesti, određivanje prognoze, odabira ciljane terapije
- somatske mutacije

PRAĆENJE BOLESTI

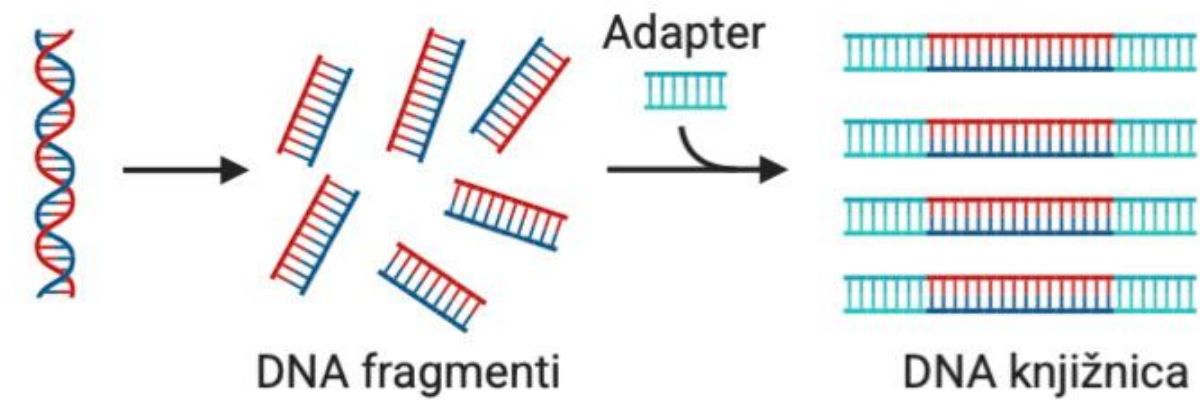
- praćenje minimalne ostatne bolesti (MRD)
- vrlo osjetljiva metoda praćenja bolesti molekularnim biljezima
- rano otkrivanje klonalne evolucije bolesti
- komplicirana standardizacija metode

ZA TOČNU ANALIZU I INTERPRETACIJU REZULTATA NUŽNA JE KVALITETNA IZVEDBA I POZNAVANJE SVIH FAZA RADA

1. korak izolacija DNA



2. korak priprema knjižnice

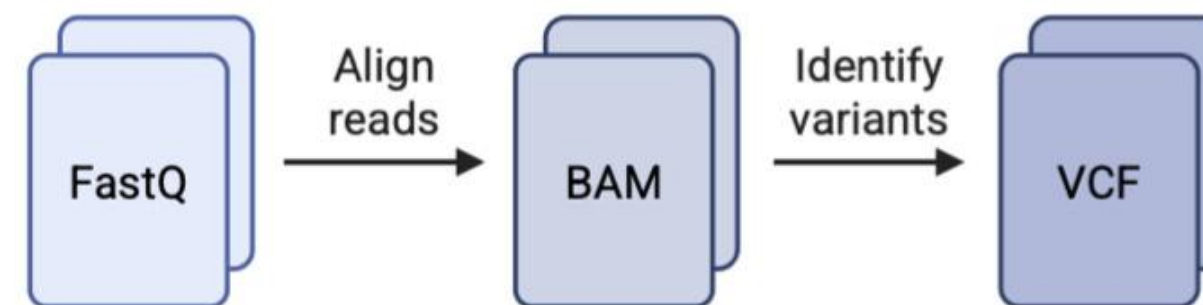


NGS tijekom rada

3. korak sekvenciranje

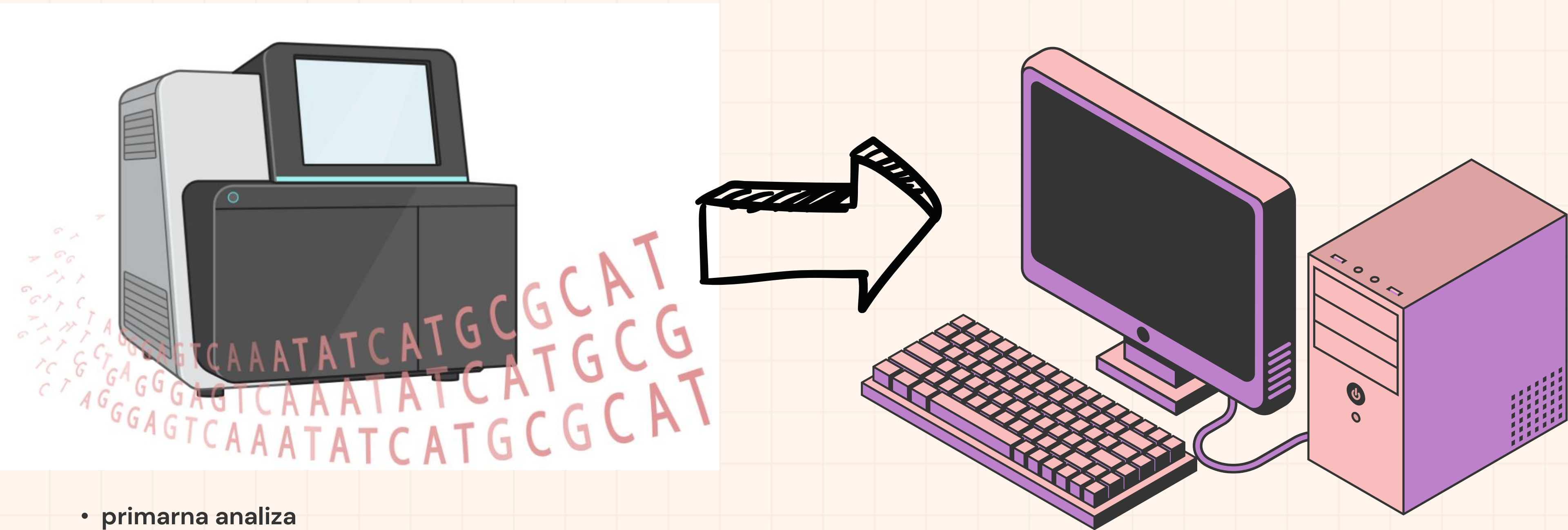


4. korak bioinformatička analiza rezultata



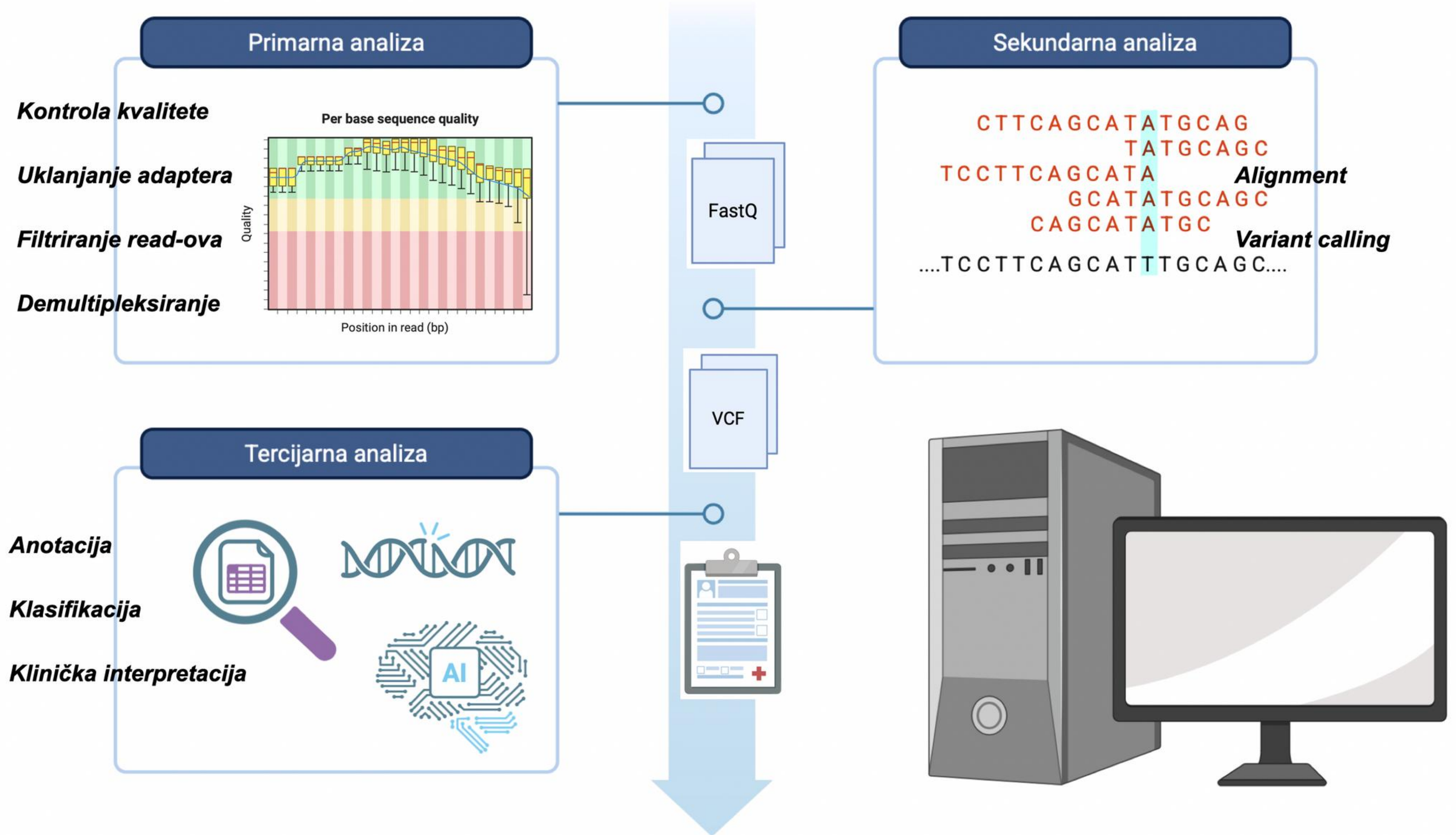
POSLIJEANALITIČKA FAZA

primjena bioinformatičkih alata



- primarna analiza
- sekundarna analiza
- tercijarna analiza

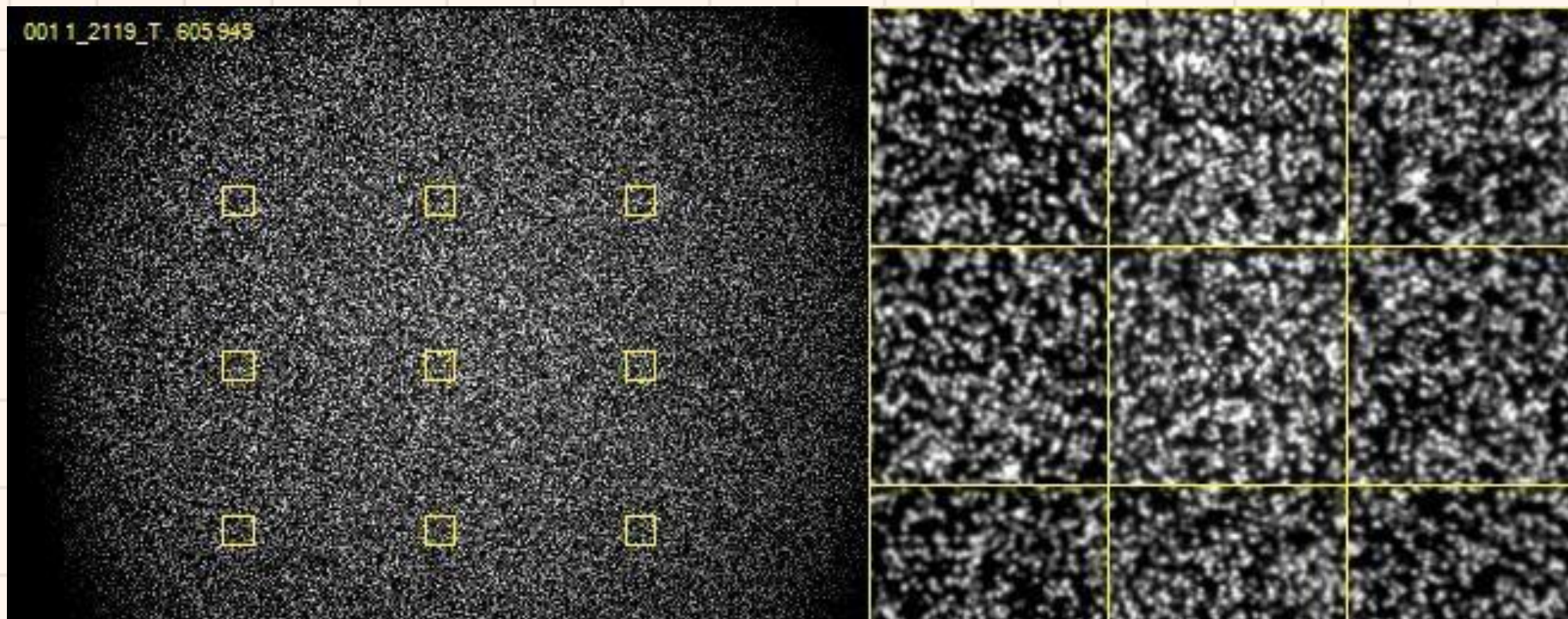
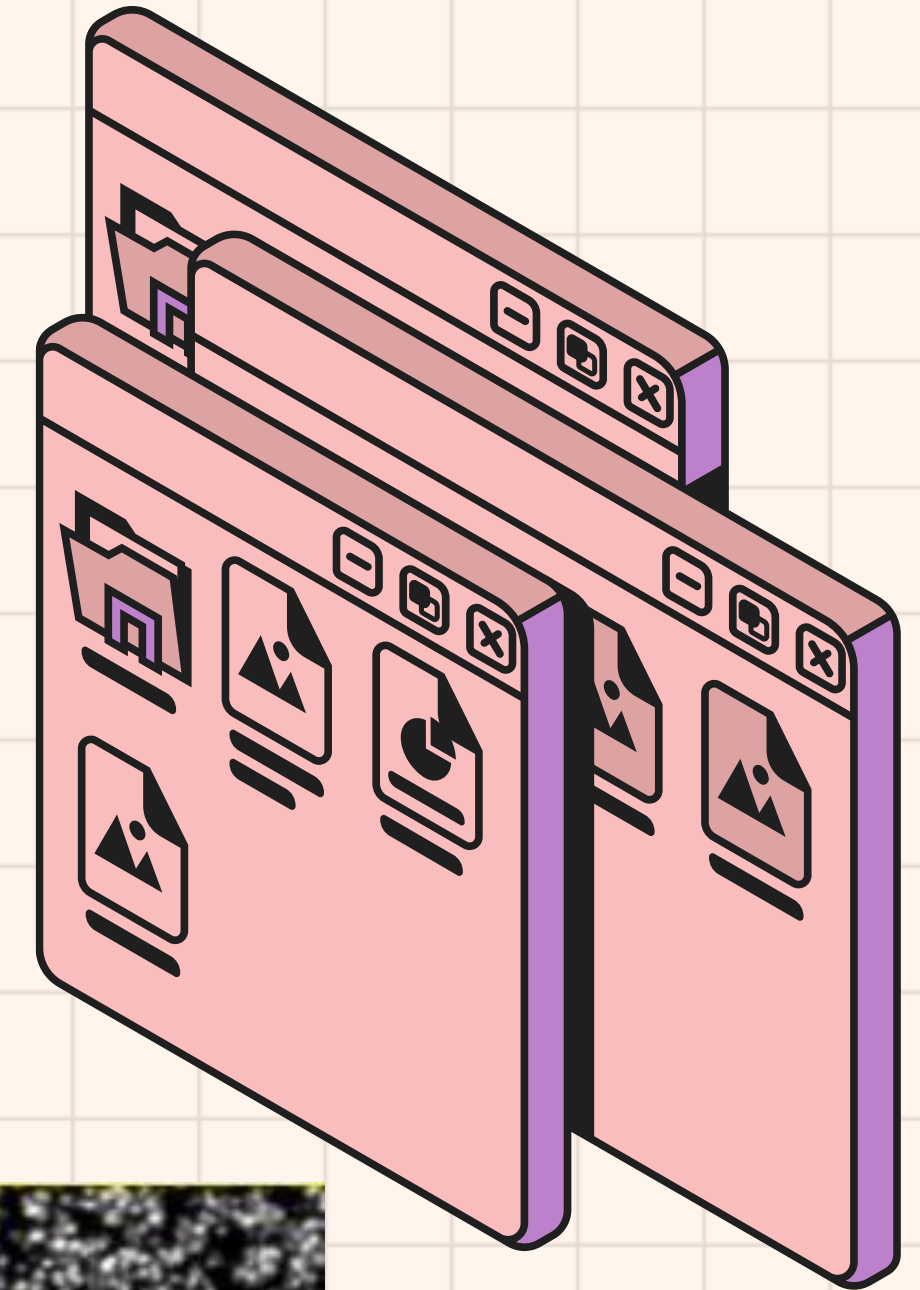
BIOINFORMATIČKI PIPELINE



PRIMARNA ANALIZA

sirovi podaci sekvenciranja

- **rezultat sekvenciranja** – slikovne datoteke očitavanja baza iz tzv. fluorescentnih klastera na protočnici analizatora
- **ocjenjivanja kvalitete** pojedinog baznog očitavanja,
- **uklanjanja artefakata**
- **demultipleksiranje** – serije slikovnih očitavanja (engl. base calls) pretvaraju se u digitalni zapis slijeda baza koje čine nukleotidnu sekvencu (*read*)
 - *read*-ovi svakog pojedinog uzorka (knjižnice) pohranjuju se u obliku zasebnih **FastQ datoteka** i čine sirove podatke sekvenciranja





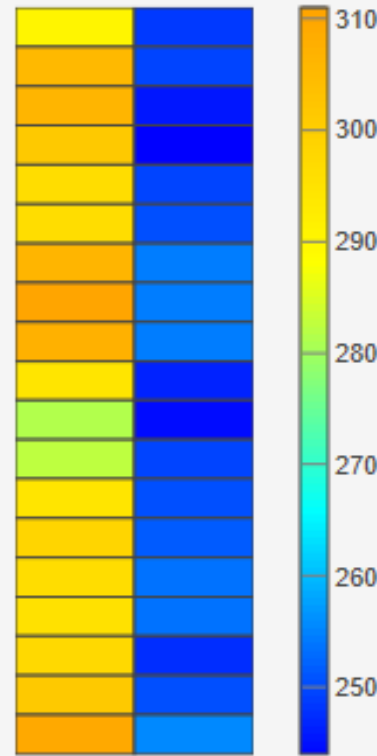
Sequencing Complete

#MS3390617-600V3 "20251201LMH"

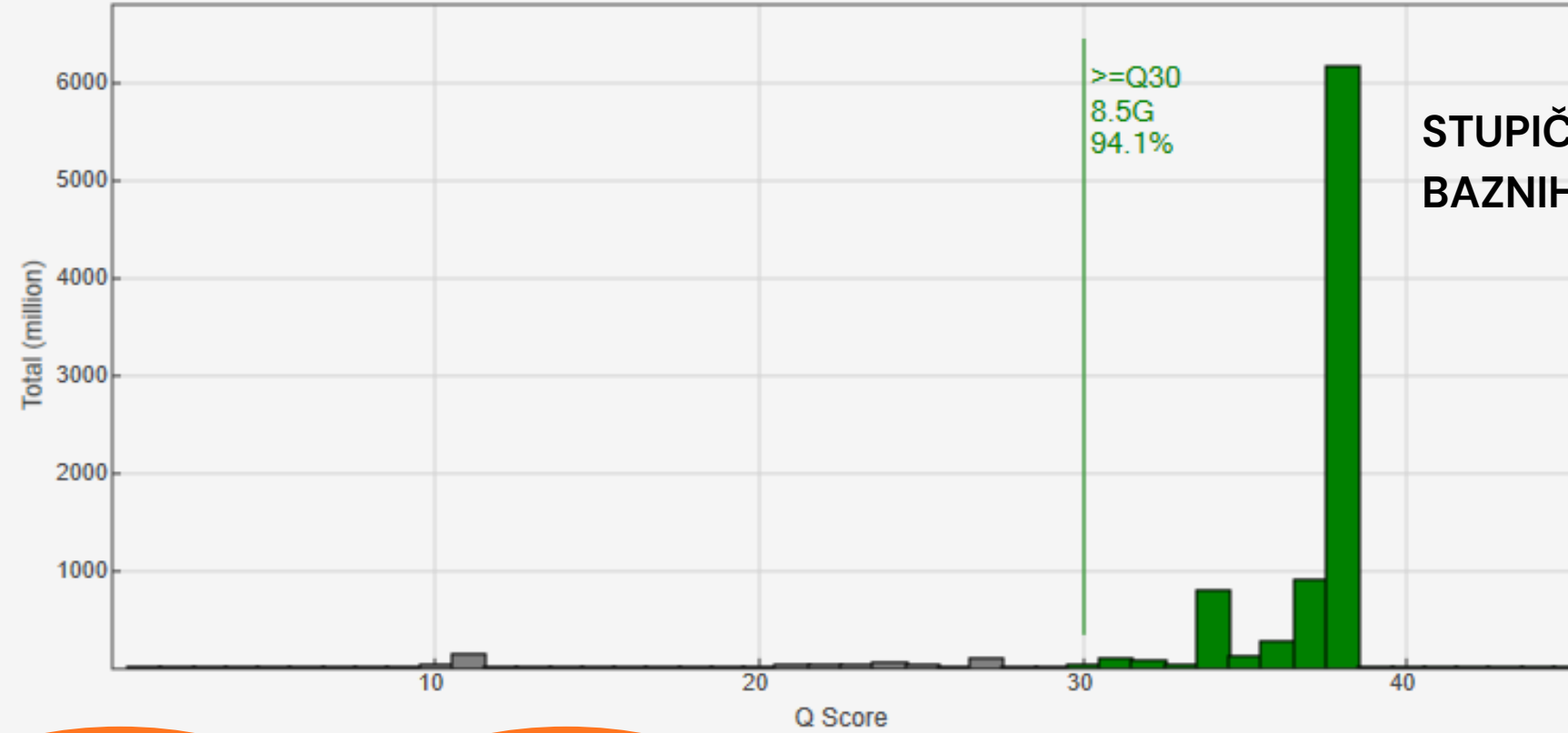
Module: GenerateFASTQ

Elapsed Time: 28:23 Cycles: 151 Cycles: 8 Cycles: 8 Cycles: 151

Intensity



Q Score All Cycles



STUPIČASTI DIJAGRAM KVALITETE
BAZNIH OČITANJA

Parametri kvalitete
pokazuju da je
sekvenciranje bilo vrlo
uspješno i tehnički
kvalitetno izvedeno.

REZULTAT PRIMARNE ANALIZE
FASTQ

❄ 4.10 °C 📖 23.20 °C

Cluster Density
1330K/mm2

Clusters Passing Filter
89.9%

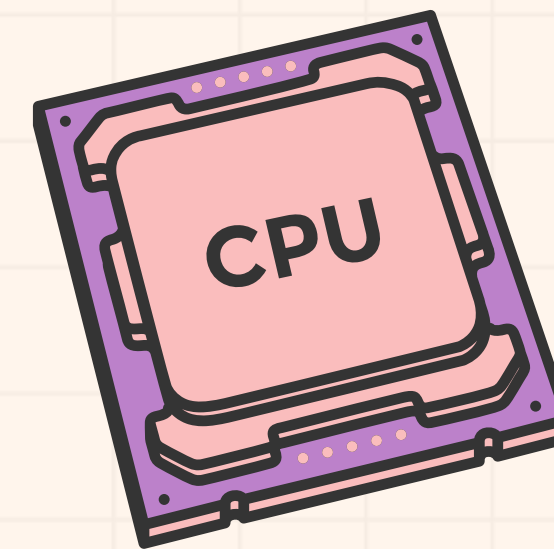
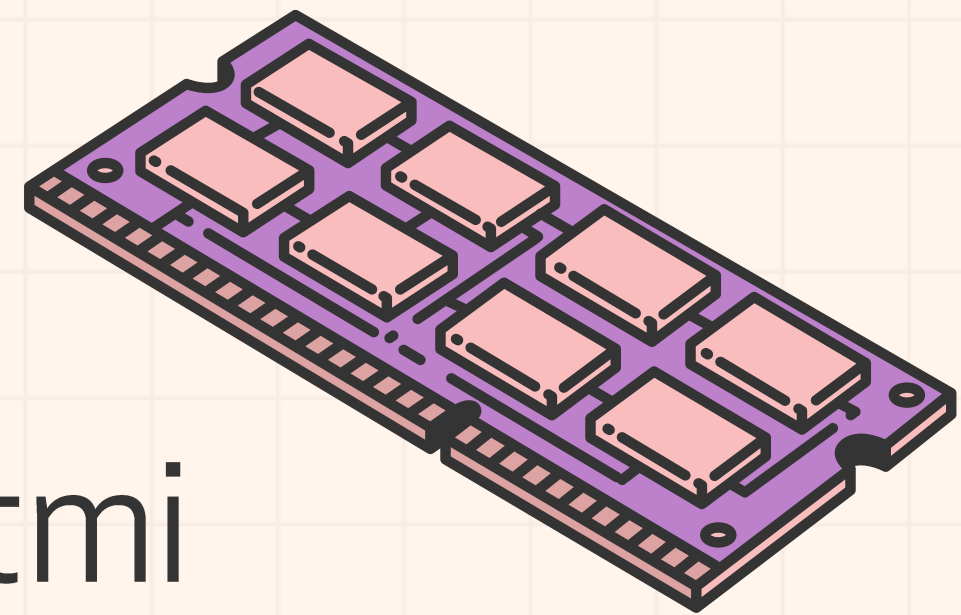
Estimated Yield
9020.1MB

Next >

```
@M04305:150:000000000-M7LVH:1:1101:11189:1559 1:N:0:5
TTGTCTTATTTTGTAACTTAAACCCTATCAACTGTTTAAAAGATTATAATCAAAGGCTAAAATTTATATTCTCTTTTGTATTGTGAGGCCAAAGTCCGATTGAATATAAGTCTGCTTTATTTTATATTCTTATTCTAAT
+
BC<CCGGAFFGGGAFFAFAA,,,CEF999966@E,CC,,,C,,,<,,,,,;C,C,,,C@,,,,6,,,,<CCF,C,<C<CCDEE,CC,,,<,<C,,,,8,,,,9CEC,+BC,,,<C,,,<E,6BEE,<@EE,C,,,99E,A,,,,,9
@M04305:150:000000000-M7LVH:1:1101:11124:1561 1:N:0:5
CCGTGTGTAATTGTAAACTTTCTAAGTGTCTTTCTCTCTCCACAGGTGGATTTTATTCTTTTATGTGTGTTAGTTTTTGTCTTATTTATCTTTACTTCTGTTCTATAACACCCAGGTTGTTTTTCTTTCTCTGTTCTGTAGTT
+
-8@B9C-E<-EF<FA<99CFGGFCE,,,=,,,;CEE@C@D<@<6,,,,,;C,,,6,,,,<,<,<,,,,,++9<C,,,<9EF,,6CCCF9EE@CF,E,,6,:86,,,,,9:B,49AABE@E,C,95,,,,,;
@M04305:150:000000000-M7LVH:1:1101:13466:1567 1:N:0:5
CTTCTCTGGGATCCTGCAGGAACCCCTCACTTGCCTCTTCTTCTGTCCCCTCACCTTTTATCATTTGAACCTGGGCCATTTTCTTCTCCTTGTTTTACCGGGGCTGTGCGCATGAGGTGGTGGATGTTTTTCCCGAAAGACATAGTT
+
<CC@CFG9--<EFGGFF8<87,CC@BBEE8@E,,6;C@CE@E,,C,C@C@@C,,6,,,C<,,6,<,,,,6:C,,,;,,,<9:CC<,66,,,,;C@9C6,6++@C7EE7C=,8,,,,;++8+++;,,,,++955+++++,;,,,,;
@M04305:150:000000000-M7LVH:1:1101:11196:1587 1:N:0:5
CAGTCTCCCCTTGCAGAGCGGCTGGTGTACGTGTTTCTGCAGAATTGCCGGAACATTGATTGTAATTTTTTCTTGGAACCTTGCACCTCCTAGTGTGGCATGGGGAGAGCAGATCGGAATAGCACACGTCTGAACTCCAGTCACTCAT
```

SEKUNDARNA ANALIZA

pametni algoritmi



```
CTTCAGCATATGCAG
      TATGCAGC
TCCTTCAGCATA
      GCATATGCAGC
      CAGCATATGC
....TCCTTCAGCATTGCAGC....
```

PORAVNANJE (ENGL. *ALIGNMENT*)

Poravnanje očitavanja slijedova baza na referentni genom tako da traži najprimjereniji izvorni slijed očitnog read-a u referentnom genomu uz algoritamsko prihvaćanje mogućeg odstupanja od referentnog genoma što u tom slučaju predstavlja **mutaciju, polimorfizam ili grešku sekvenciranja**

DETEKCIJA VARIJANTI (ENGL. *VARIANT CALLING*)

Identifikacija varijanti s pomoću **niza pravila, tj. postupnika** koji će osigurati razlikovanje stvarnih varijanti od slučajnih grešaka sekvenciranja. Da bi se neka promjena, odnosno odstupanje od referentnog genoma, pouzdano prozvala varijantom, treba biti zabilježena u točno određenom broju različitih očitavanja slijedova baza

REFERENTNA
SEKVENCA





bioinformatički hodogram

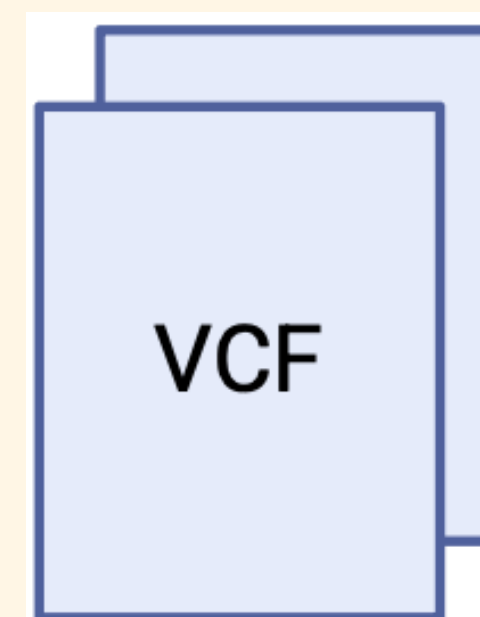
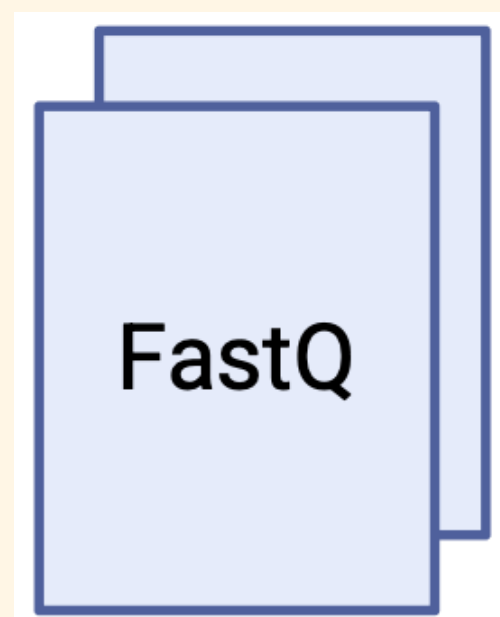
SUČELJE NAREDBENOG RETKA

- engl. command line interface
- rad u terminalu u kojem se analize pokreću s pomoću tekstnih naredbi i skripti
- "programiranje"
- potrebno više bioinformatičkog znanja
- vrlo precizna kontrola nad parametrima i prilagođavanje složenih pravila

PROGRAMSKI PAKETI S GRAFIČKIM SUČELJEM

- vizualno, klikom vođeno okruženje u kojem se analize sastavljaju i pokreću kroz izbornike i okvire
- intuitivniji za korisnike bez programerskog iskustva
- manja fleksibilnost i manja mogućnost prilagođavanja parametara analize

U rutinskoj kliničkoj praksi najčešće se poseže za **gotovim komercijalnim programskim paketima** koji su prilagođeni specifičnom NGS pristupu, ciljanoj bolesti, vrsti tkiva i skupini varijanti koje želimo detektirati.



primjeri programskih alata



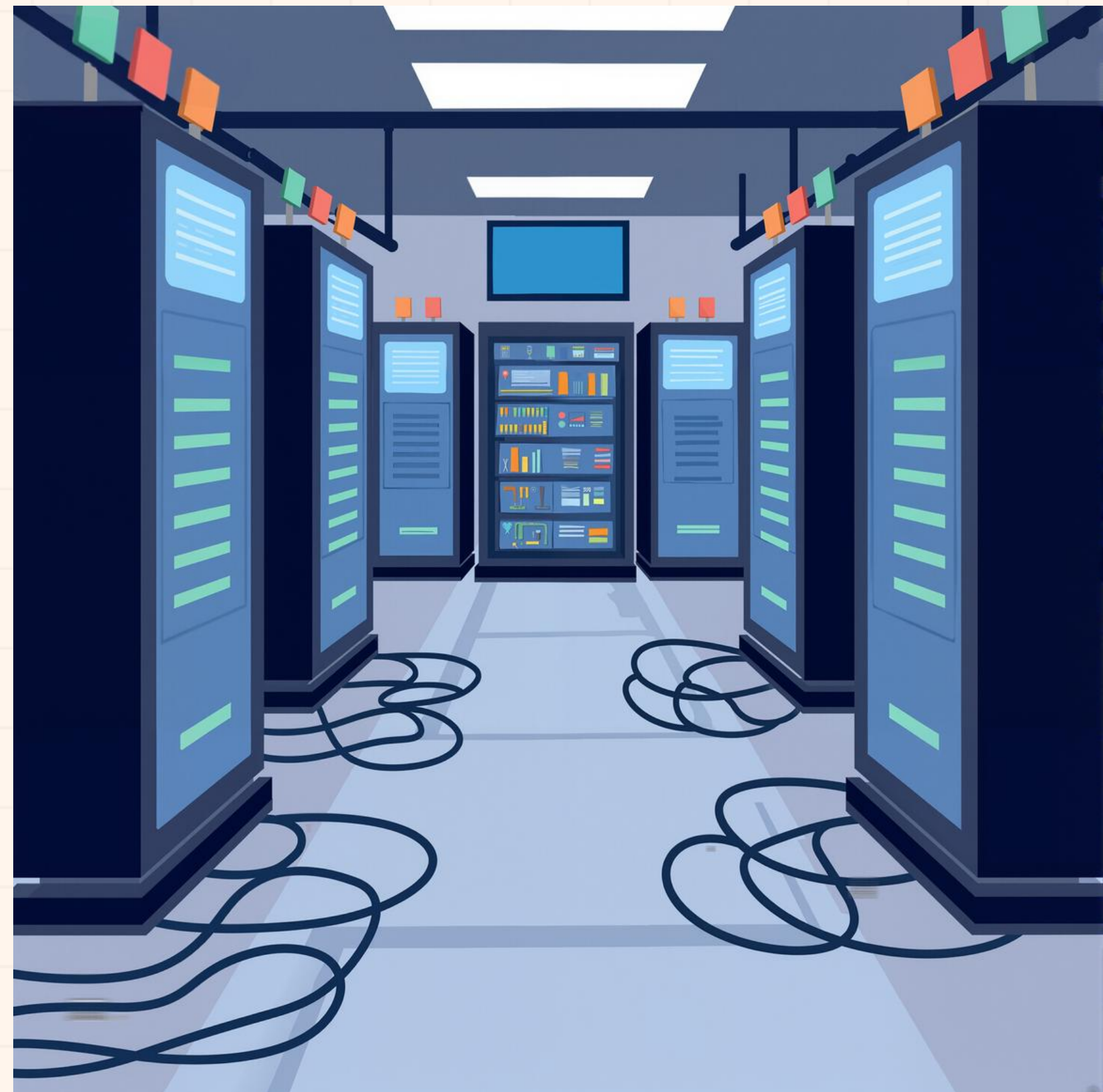
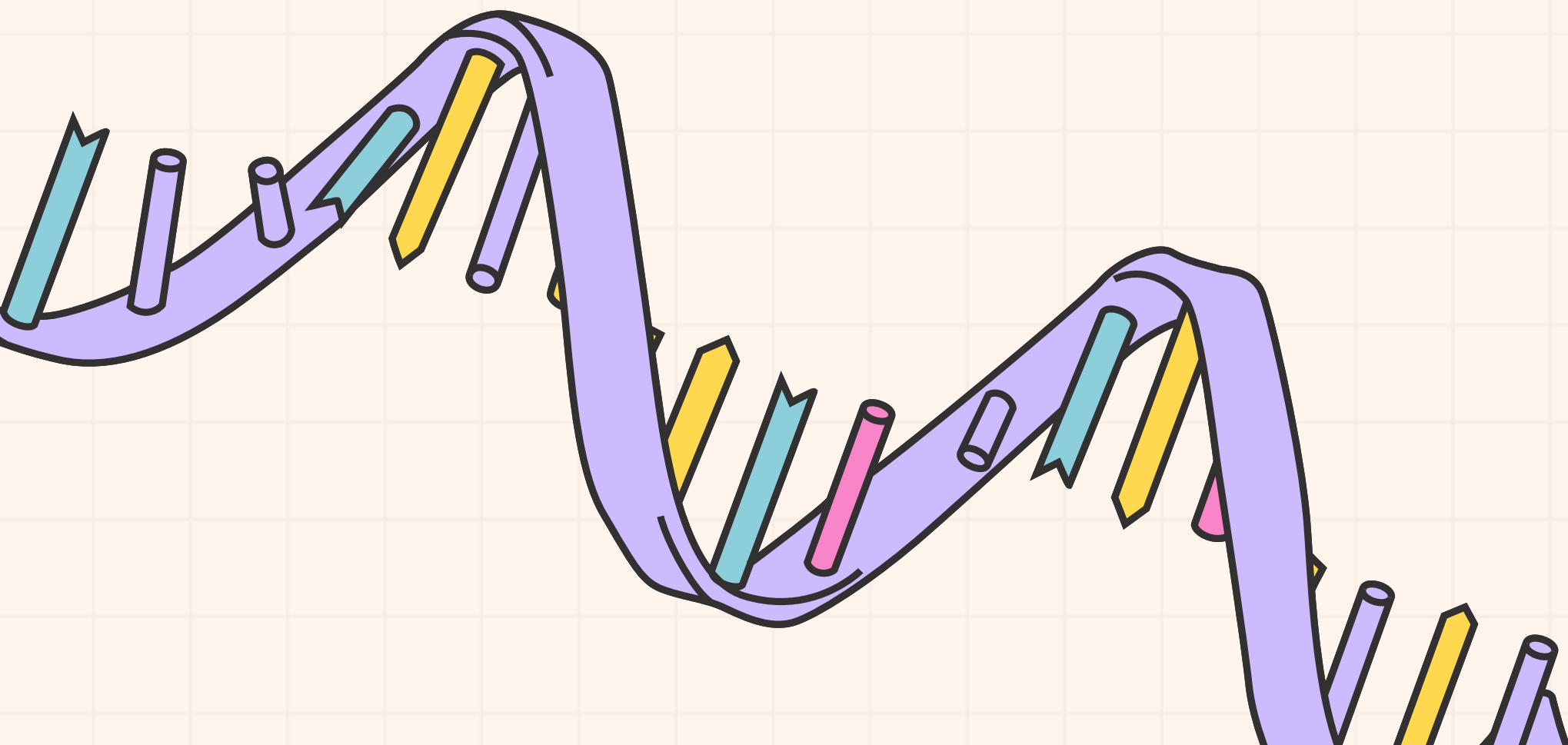
REZULTAT SEKUNDARNE ANALIZE

Search for gene, chr, etc.													
Variant	Gene Symbol	Variant Type		Somatic Tier	Sample Metrics	Allelic Balance	Coverage	HGVS	HGVS Protein	HGVS Coding	OncoKB Mutation Effect	Somatic	
☒ chr20:31022277 C⇒T	ASXL1	SNV		Tier I	★ 📄 👤 🌐 ○ 🔍	0.43	2 397 (JB/IGV)	NM_015338.6:c.1762C>T p.(Gln588Ter)	Q588* p.(Gln588T...	c.1762C>T	Likely Loss-of-function		
☐ chr17:74732936 delG...	SRSF2	Deletion (24)		Tier I	★ 📄 👤 🌐 ○ 🔍	0.45	974 (JB/IGV)	NM_001195427.2:c.284_307del p.(Pro95_Arg102del)	P95_R102del p.(P...	c.284_307del	Likely Loss-of-function		
☐ chr15:90631934 C⇒T	IDH2	SNV		Tier II	★ 📄 👤 🌐 ○ 🔍	0.49	1 500 (JB/IGV)	NM_002168.4:c.419G>A p.(Arg140Gln)	R140Q p.(Arg140...	c.419G>A	Switch-of-function		
☐ chr9:139390945 delGTG	NOTCH1	Deletion (3)		Tier II	★ 📄 👤 🌐 ○ 🔍	0.008	1 513 (JB/IGV)	NM_017617.5:c.7244_7246del p.(Pro2415del)	P2415del p.(Pro2...	c.7244_7246del	Likely Gain-of-function		
☐ chr3:128200138 delC	GATA2	Deletion (1)		Tier II	★ 📄 👤 🌐 ○ 🔍	0.34	1 642 (JB/IGV)	NM_032638.5:c.1167del p.(Lys390ArgfsTer87)	K390Rfs*87 p.(Lys...	c.1167del	Likely Loss-of-function		
☐ chr3:128200140 delTC	GATA2	Deletion (2)		Tier II	★ 📄 👤 🌐 ○ 🔍	0.34	1 643 (JB/IGV)	NM_032638.5:c.1164_1165del p.(Met388IlefsTer147)	M388Ilefs*147 p.(...	c.1164_1165del	Likely Loss-of-function		
☐ chr20:57478807 C⇒T	GNAS	SNV		Tier II	★ 📄 👤 🌐 ○ 🔍	0.5	2 863 (JB/IGV)	NM_000516.7:c.393C>T p.(Ile131=)	I131= p.(Ile131=)	c.393C>T	-		
☐ chr9:21975017 C⇒T	CDKN2A	SNV		Tier II	★ 📄 👤 🌐 ○ 🔍	1	105 (JB/IGV)	ENST00000304494.5:c.-191G>A p.?	p.?	c.-191G>A	-		
☐ chr11:119149356 del...	CBL	Deletion (3)		Tier III	★ 📄 👤 🌐 ○ 🔍	0.004	1 846 (JB/IGV)	NM_005188.4:c.1380_1382del p.(Asp460del)	D460del p.(Asp4...	c.1380_1382del	-		
☐ chr9:133759490 delA...	ABL1	Deletion (3)		Tier III	★ 📄 👤 🌐 ○ 🔍	0.017	2 101 (JB/IGV)	NM_005157.6:c.1826_1828del p.(Lys609del)	K609del p.(Lys60...	c.1826_1828del	-		
☐ chrX:76763680 delA	ATRX	Deletion (1)		Tier III	★ 📄 👤 🌐 ○ 🔍	0.077	117 (JB/IGV)	NM_000489.6:c.*149del p.?	p.?	c.*149del	-		
☐ chr7:101892133 delC...	CUX1	Deletion (3)		Tier III	★ 📄 👤 🌐 ○ 🔍	0.015	472 (JB/IGV)	NM_181552.4:c.4342_4344del p.(Ser1448del)	S1448del p.(Ser1...	c.4342_4344del	-		
☐ chr4:55131002 delAAA	PDGFRA	Deletion (3)		Tier III	★ 📄 👤 🌐 ○ 🔍	0.047	193 (JB/IGV)	NM_006206.6:c.629-70_629-68del p.?	p.?	c.629-70_629-68...	-		
☐ chr9:133759492_3 ins...	ABL1	Insertion (3)		Tier III	★ 📄 👤 🌐 ○ 🔍	0.002	2 101 (JB/IGV)	NM_005157.6:c.1826_1828dup p.(Lys609dup)	K609dup p.(Lys60...	c.1826_1828dup	-		
☐ chrX:76763679_0 insA	ATRX	Insertion (1)		Tier III	★ 📄 👤 🌐 ○ 🔍	0.068	117 (JB/IGV)	NM_000489.6:c.*149dup p.?	p.?	c.*149dup	-		
☐ chr11:119155962 del...	CBL	Deletion (3)		Tier III	★ 📄 👤 🌐 ○ 🔍	0.002	2 328 (JB/IGV)	NM_005188.4:c.1635_1637del p.(Pro548del)	P548del p.(Pro54...	c.1635_1637del	-		

TERCIJARNA ANALIZA ANOTACIJA VARIJANTI

funkcionalna i klinička interpretacija varijanti

- integracija znanja iz genomike i molekularne patogeneze bolesti
- integracija podataka iz javnih genomskih repozitorija



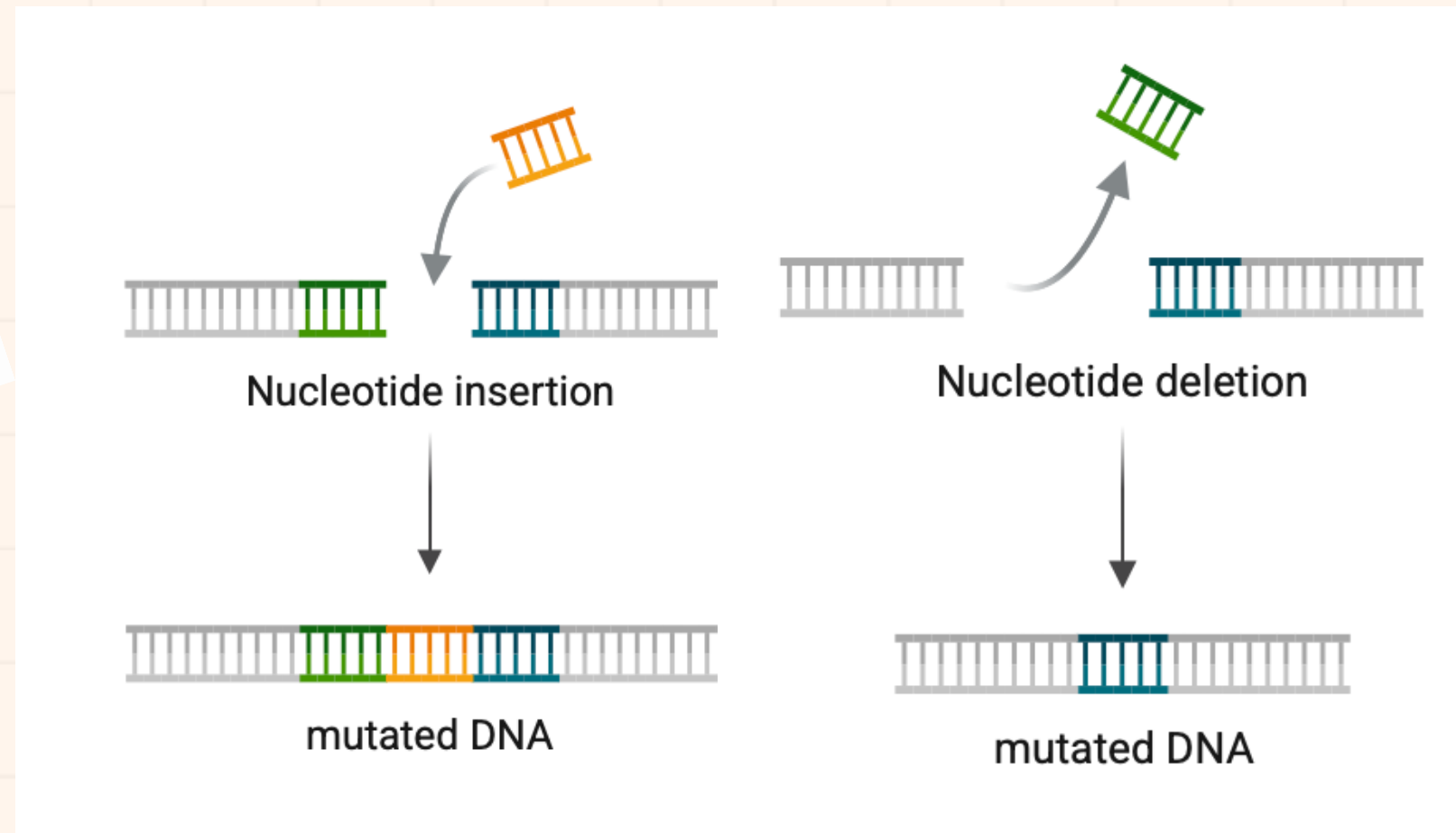
MUTACIJE

SOMATSKE

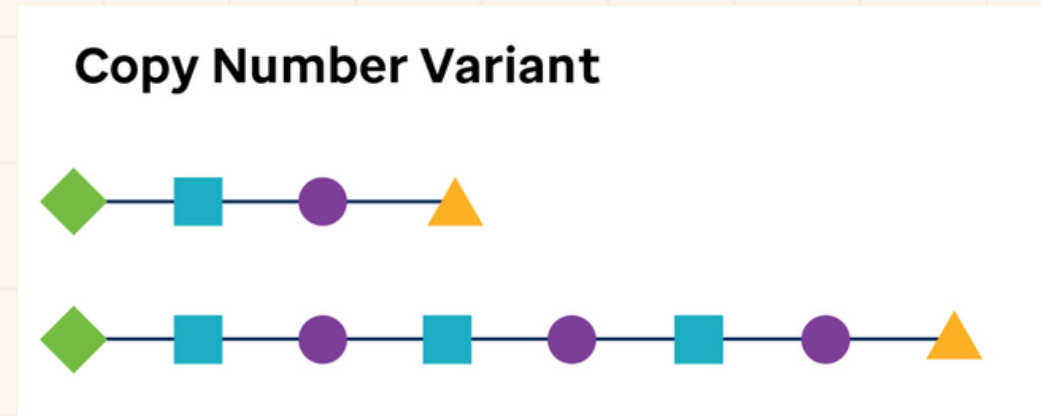
NASLJEDNE



TOČKASTE MUTACIJE
(SNV)



MALE INSERCIJE/ DELECIIJE
(INDELS)



Copy Number Variant

STRUKTURNE PROMJENE

varijacije broja kopija (CNV)
velike delecije, insercije ili duplikacije

- prema učinku na funkciju proteina – **neutralne**, **aktivirajuće** (engl. *gain of function*) ili **inaktivirajuće** (engl. *loss of function*)
- prema genomskom položaju razlikujemo varijante u **kodirajućim regijama** (*missense*, *nonsense*, *frameshift*, *in-frame*), u **splice-site regijama** (narušavaju pravilno spajanje egzona) te u **nekodirajućim/ regulatornim regijama** (UTR, promotorske i intronske varijante) koje mogu utjecati na razinu i regulaciju ekspresije gena

JAVNE BAZE PODATAKA

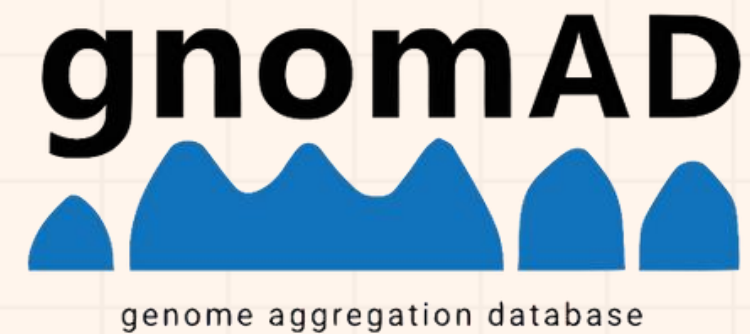
BAZE REFERENTNIH SEKVENCI

- informacija o točnom položaju varijante u genomu
- omogućuju pravilno i standardizirano opisivanje genske varijante HGVS nomenklaturom (engl. Human Genome Variation Society) za nedvosmisleno prikazivanje varijante u DNA, RNA i proteinu



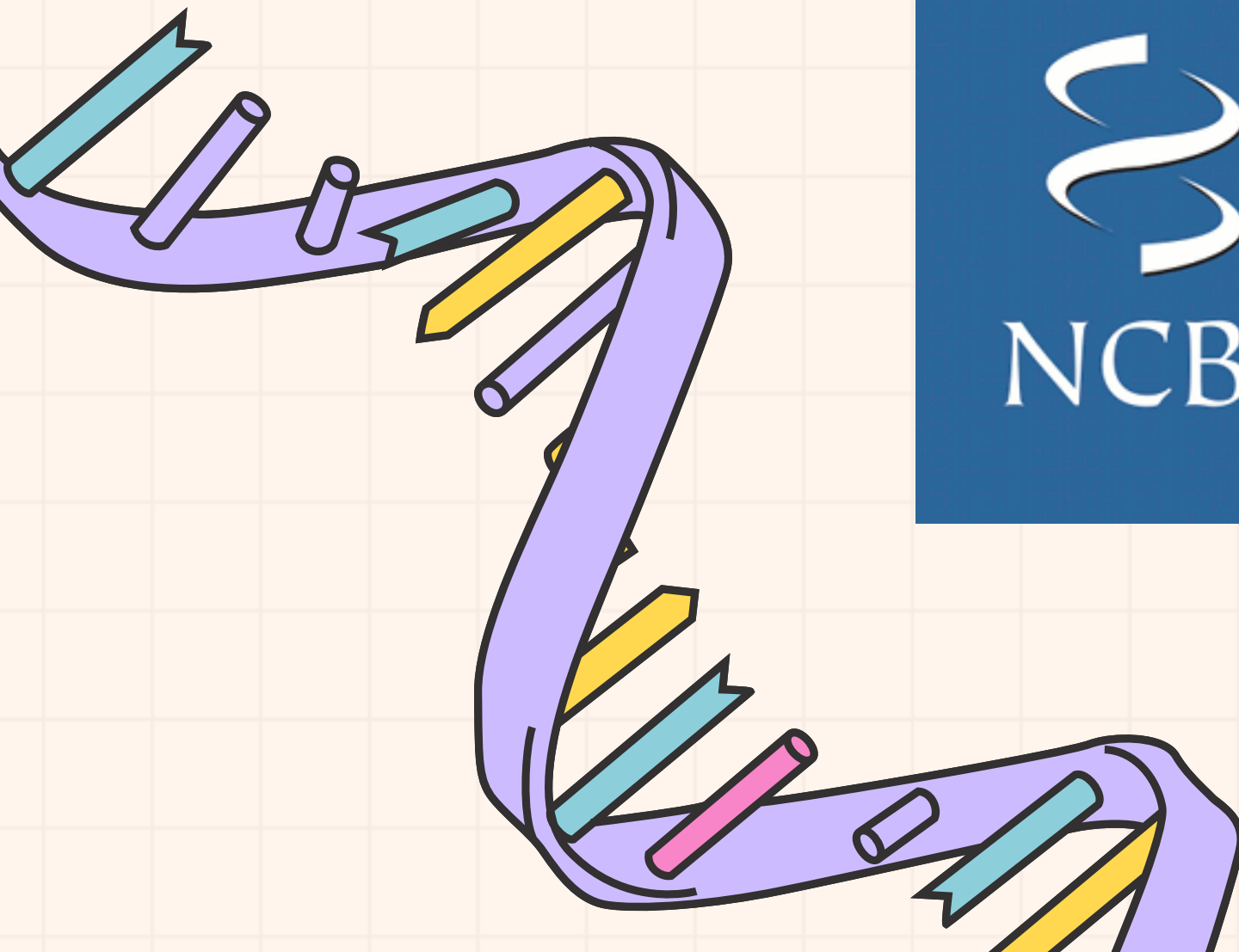
POPULACIJSKE BAZE

- razlikovanje mutacija od polimorfizama



BAZE SOMATSKIH I NASLJEDNIH VARIJANTI

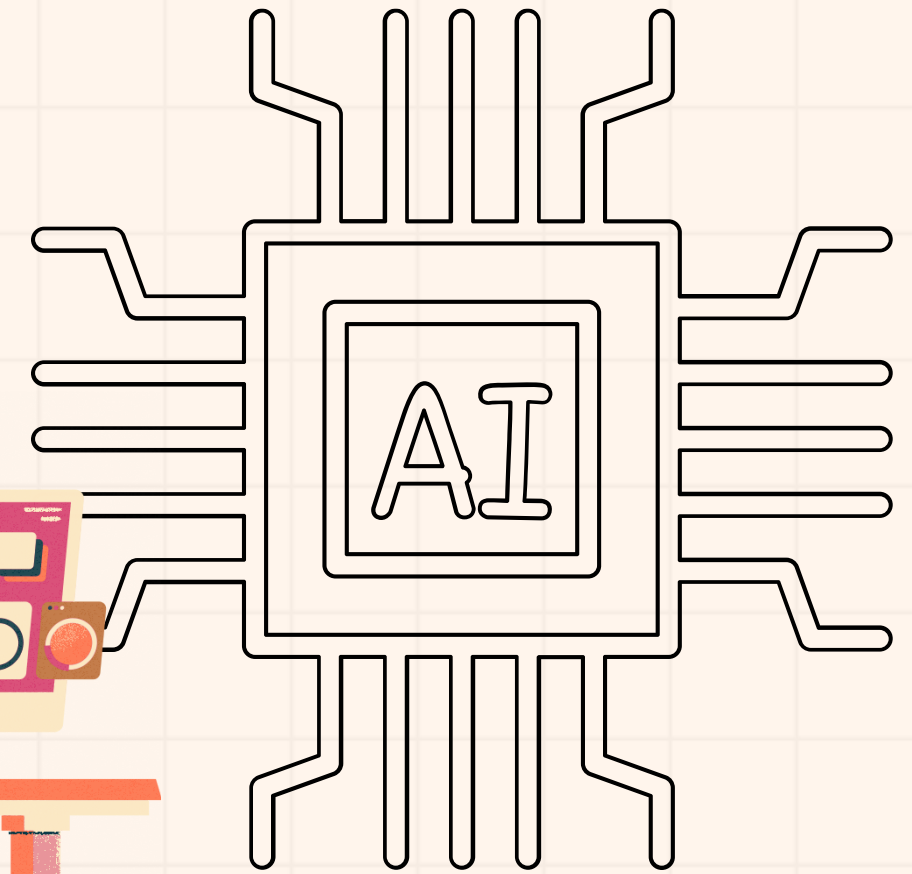
- informacije o varijantama detektiranim u sklopu velikih projekata sekvenciranja
- varijante opisane u sklopu funkcionalnih i kliničkih studija



...IZVORI TERCIJARNE ANALIZE U KARAKTERIZACIJI VARIJANTI

ZNANSTVENA LITERATURA

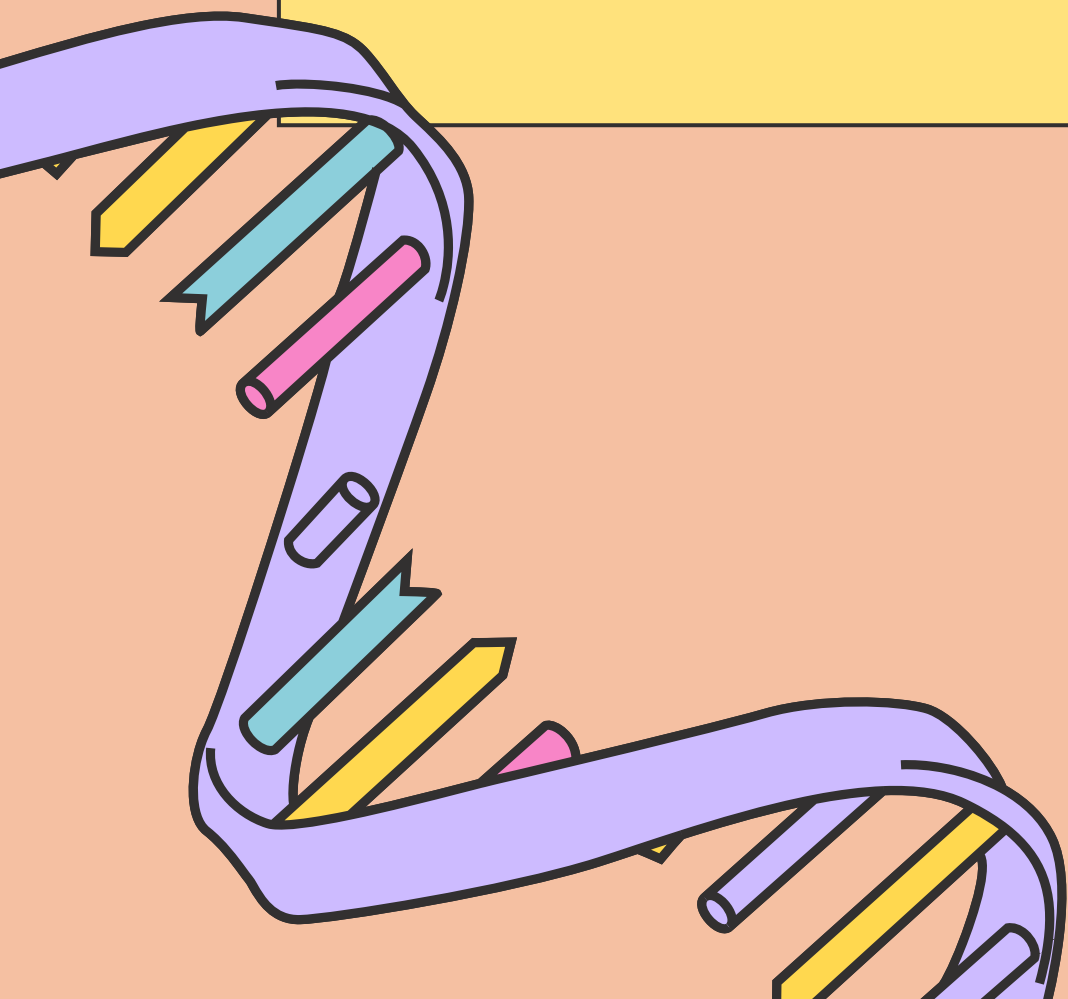
- recenzirane i pouzdane reference u kojima se opisuju varijante od interesa u kontekstu bolesti



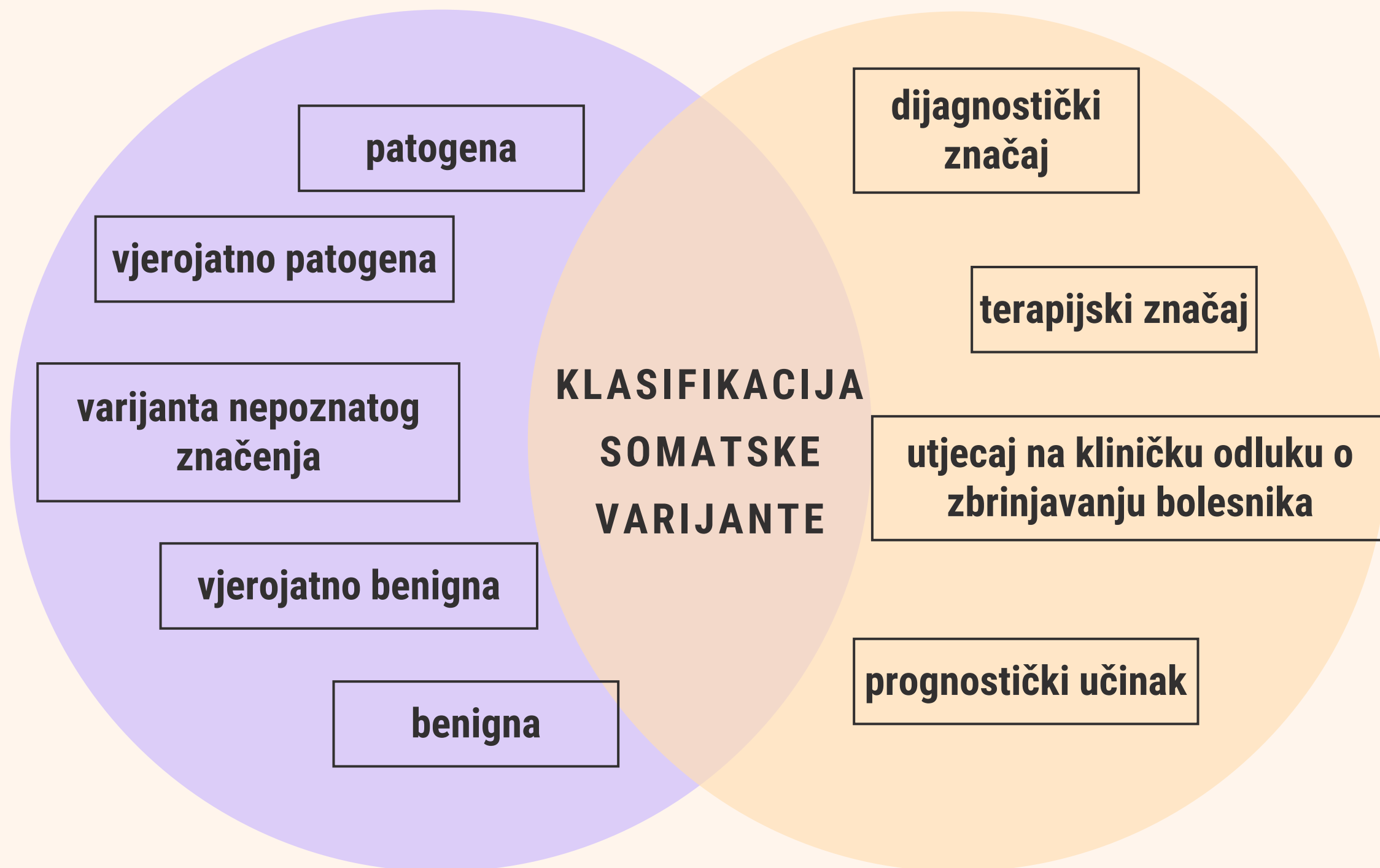
IN SILICO ALATI

- alati predviđanja patogenosti varijante
- *in silico* – izvedenica po uzoru na pojmove *in vivo* (lat. u živom) i *in vitro* (lat. u epruveti)
- koriste računalne simulacije i analize modela u virtualnom okruženju
- algoritam temeljen na AI tehnologiji (engl. *Artificial Intelligence*) i strojnom učenju (engl. *Machine Learning*)
- npr. PolyPhen2 i SIFT

KLJUČ ANOTACIJE



integracija svih navedenih izvora znanja, dokaza i informacija o varijanti za određivanje patogenosti i kliničkog značaja varijante u kontekstu bolesti



Curated Variants Tier I	Tier I	Patient cancer matched: Myelodysplastic Syndromes, diagnostic and prognostic, curated, from OncoKB, citing 24695057 , 24220272 , 22919025 , 21714648 , and 21576631 .
	Tier II	Tier I evidence found for other cancer types: Acute Myeloid Leukemia, Chronic Myelomonocytic Leukemia, Polycythemia Vera, AML with Myelodysplasia-Related Changes, and 4 more, prognostic and diagnostic, guideline and curated, from CKB and OncoKB, citing 29296692, 29176559, 25860933, 25596267, and 9 more. Tier II evidence found for other cancer types: Acute Myeloid Leukemia, 2 therapies (CG-808 and Sunitinib), resistant and sensitive, pre-clinical, from CKB.
	Tier III	Tier III evidence found for other cancer types: Kidney Wilms Tumor and Myelodysplastic Syndrome, prognostic and risk factor, clinical study and guideline, from CKB, citing 24220272 and 21714648. Variant not found in CIVIC.
Drugs & Treatments Tier II	Tier II	Azacitidine + Venetoclax, Sensitive, phase I clinical trial, from AACT. Patient cancer matched Myeloid Cancer. 4 therapies (Azacitidine + Venetoclax, Azacitidine, CG-808, and Sunitinib), from AACT, CKB, and DGI. Citing 25224413. Tier II evidence found for other cancers: Acute Myeloid Leukemia, Chronic Lymphocytic Leukemia/Small Lymphocytic Lymphoma, Chronic Lymphocytic Lymphoma, Leukemia, and 7 more.
	Tier III	No additional relevant drugs or clinical trials found in CIVIC genes or PharmGKB.
Germline Evidence Tier I	Tier I	This nonsense variant is classified Pathogenic by the germline classifier, using rules PVS1, PP5_Moderate, and PM2_Supporting.
Pathways Tier I	Tier I	CKB Genes reports that ASXL1, additional sex combs like 1, transcriptional regulator, is involved in epigenetic regulation of gene expression and forms complexes with several proteins to regulate transcription (23736028). Mutations in ASXL1 are frequent in chronic myelomonocytic leukemia (26771811 , 23736028), and have been identified in a number of different tumor types including acute myeloid leukemia and liver cancer (23736028). CKB Genes classifies ASXL1 as tumor suppressor based on article 30606230 .
	Tier II	GHR reports that the ASXL1 gene provides instructions for making a protein that is involved in a process known as chromatin remodeling. Also, associates ASXL1 with the following cancer: Systemic Mastocytosis. Mondo associates gene ASXL1 with the following 5 cancers: Acute Myeloid Leukemia with Multilineage Dysplasia, Aggressive Systemic Mastocytosis, Chronic Myelomonocytic Leukemia, Myelodysplastic Syndrome, and Systemic Mastocytosis with An Associated Clonal Hematologic Non-Mast Cell Lineage Disease.
Publications Tier III	Tier III	No publications have been linked by VarSome users.
Somatic Evidence Tier I	Tier I	This variant is reported in 22 out of the 4 218 somatic samples for gene ASXL1, which has 2 317 reported somatic variants. It is not found in GnomAD. This is statistically rated Tier I.
	Tier III	CBioPortal reports 20 samples, cancer type = 6 x Leukemia, 3 x Bladder Cancer, 2 x Appendiceal Cancer, and 8 more, sample type = 5 x Primary and 2 x Metastasis, tissue = 9 x Myeloid, 3 x Bladder/Urinary Tract, 3 x Bowel, and 4 more. CancerHotspots reports 1 sample, bio-type = 1 x protein_coding, cancer type = 1 x Bladder Cancer, tissue site = 1 x Bladder/Urinary Tract. GDC reports 1 sample, outcomes = 1 x Disease Free, sex = 1 x Male, tissues = 1 x Bladder/Urinary Tract. Variant not found in ICGC.
	Tier II	Variant not found in gnomAD exomes, good gnomAD exomes coverage = 88.4.
Variant Impact Tier I	Tier I	Null variant (nonsense) in gene ASXL1, not predicted to cause NMD. Loss-of-function is a known mechanism of disease (gene has 248 reported pathogenic LOF variants). The exon affects 1 functional domain: UniProt protein ASXL1_HUMAN region of interest 'Interaction with NCOA1'.
	Tier III	Variant is not predicted splicing: no prediction from MaxEntScan.
Predictions Tier III	Tier III	In-silico predictions are disabled for nonsense variants.

Rezultat tercijarne analize - NALAZ KLASIFICIRANIH VARIJANTI

GEN	Referentna sekvenca	Transkript	Protein	Kliničko značenje	VAF (%)	ID/patogena
ASXL1	NM_015338.6	c.1762C>T	p.(Gln588Ter)	Tier I	43.0	COSV60103144 rs1486082302
SRSF2	NM_001195427.2	c.284_307del	p.(Pro95_Arg102del)	Tier I	45.0	COSV57969801 rs766200080
IDH2	NM_002168.4	c.419G>A	p.(Arg140Gln)	Tier II	49.0	COSV57468751 rs121913502
GATA2	NM_032638.5	c.1167del c.1164_1165del	p.(Lys390Argfs Ter87) p.(Met388Ilefs Ter147)	Tier II	34.0	*

** U ovom slučaju moguće su dvije verzije promjena na razini gena GATA2, a koje je nemoguće sa sigurnošću utvrditi zbog visoko homologne ponavljajuće sekvence. Obje varijante uzrokuju pomak okvira čitanja te imaju terminirajući efekt na protein stoga se smatraju vjerojatno patogenima i imaju kliničku značajnost u kontekstu hematološke bolesti.*

BIOINFORMATIČKI ALATI

INTEGRACIJA, AUTOMATIZACIJA, STANDARDIZACIJA, HARMONIZIRANOST

Rezultati NGS analize u dijagnostici akutnih leukemija imaju veliku diferencijalno dijagnostičku vrijednost te **moгу promijeniti kompletni smjer kliničkog vođenja bolesnika** stoga je od iznimne važnosti odgovarajuća interpretacija i anotacija pronađenih varijanti, a samo NGS izvješće **treba biti u obliku razumljivog, nedvosmislenog i koherentnog nalaza**

