

# Učinkovitost bioinformatičkih alata u predviđanju patogenosti somatskih mutacija dokazanih tehnologijom sekvenciranja sljedeće generacije kod bolesnika s dijagnozom akutne mijeloične leukemije

---

Božac, Toni

Master's thesis / Diplomski rad

2024

Degree Grantor / Ustanova koja je dodijelila akademski / stručni stupanj: **University of Zagreb, Faculty of Pharmacy and Biochemistry / Sveučilište u Zagrebu, Farmaceutsko-biokemijski fakultet**

Permanent link / Trajna poveznica: <https://um.nsk.hr/um:nbn:hr:163:207469>

Rights / Prava: [In copyright](#) / [Zaštićeno autorskim pravom.](#)

Download date / Datum preuzimanja: **2025-01-25**



Repository / Repozitorij:

[Repository of Faculty of Pharmacy and Biochemistry University of Zagreb](#)



**Toni Božac**

**Učinkovitost bioinformatičkih alata u predviđanju  
patogenosti somatskih mutacija dokazanih  
tehnologijom sekvenciranja sljedeće generacije kod  
bolesnika s dijagnozom akutne mijeloične  
leukemije**

**DIPLOMSKI RAD**

Predan Sveučilištu u Zagrebu Farmaceutsko-biokemijskom fakultetu

Zagreb, 2024.

Ovaj diplomski rad je prijavljen na Sveučilištu u Zagrebu Farmaceutsko-biokemijskog fakulteta i izrađen u Kliničkom bolničkom centru Zagreb pod stručnim vodstvom prof. dr. sc. Dunje Rogić te suvoditeljstvom Margarete Radić Antolic, mag.med.biochem i doc.dr.sc. Ivane Lapić.

*Zahvaljujem se prof.dr.sc. Dunji Rogić na povjerenju, podršci i omogućenoj izvedbi ovog diplomskog rada. Posebno se zahvaljujem doc.dr.sc. Ivani Lapić, Margareti Radić Antolic, spec.med.biokem. i lab.medicine i Moniki Kolundžić, mag.med.biochem. na brojnim savjetima i velikoj pomoći u svakom koraku u stvaranju ovog rada.*

*Posebna zahvala ide prijateljima s fakulteta koji su bili uz mene i uljepšali mi ovih 5 godina studiranja. Veliko hvala mojim roditeljima koji su me podržavali tijekom cijelog studija. Hvala Ines na podršci i motivaciji tijekom pisanja diplomskog i studiranja.*

## SADRŽAJ

1. UVOD .....	1
1.1 AKUTNA MIJELOIČNA LEUKEMIJA - DEFINICIJA .....	1
1.2. EPIDEMIOLOGIJA I ETIOLOGIJA .....	2
1.3. PATOFIZIOLOGIJA .....	2
1.4. PODJELA .....	2
1.5. STRATIFIKACIJA RIZIKA .....	6
1.6. DIJAGNOSTIKA AKUTNE MIJELOIČNE LEUKEMIJE .....	8
1.6.1. Citomorfološka dijagnostika .....	8
1.6.2. Citogenetska dijagnostika .....	8
1.6.3. Imunofenotipizacija – protočna citometrija .....	9
1.6.4. Molekularna dijagnostika .....	9
1.7. SEKVENCIRANJE SLJEDEĆE GENERACIJE .....	11
1.7.1. Ciljano sekvenciranje .....	12
1.8. BIOINFORMATIČKA OBRADA PODATAKA .....	12
1.9. INTERPRETACIJA DETEKTIRANIH VARIJANTI .....	14
2. OBRAZLOŽENJE TEME .....	16
3. MATERIJALI I METODE .....	17
3.1. UZORCI .....	17
3.2. IZDVAJANJE NUKLEINSKIH KISELINA .....	17
3.2.1. Određivanje koncentracije izolirane DNA .....	17
3.3. SEKVENCIRANJE SLJEDEĆE GENERACIJE .....	17
3.3.1. Priprema knjižnice .....	18
3.3.2. Ciljano obogaćivanje knjižnice .....	19
3.3.3. Sekvenciranje sintezom .....	19
3.3.4. Sekvenciranje uparenih krajeva .....	21
3.4. BIOINFORMATIČKA OBRADA PODATAKA .....	22
3.4.1. SureCall .....	23
3.4.2. VarSome Clinical .....	24
3.5. OBRADA PODATAKA .....	24
4. REZULTATI I RASPRAVA .....	25
4.1. Vrsta, raspodjela i učestalost patogenih mutacija povezanih s akutnom mijeloičnom leukemijom .....	25
4.2. Usporedba sekundarne analize podataka .....	27
4.3. Usporedba tercijarne analize podataka .....	27
4.4. Rasprava .....	28
5. ZAKLJUČCI .....	35
6. POPIS KRATICA, OZNAKA I SIMBOLA .....	36

7. LITERATURA .....	38
8. SAŽETAK/SUMMARY .....	43
9. TEMELJNA DOKUMENTACIJSKA KARTICA/ BASIC DOCUMENTATION CARD	45

## 1. UVOD

### 1.1 AKUTNA MIJELOIČNA LEUKEMIJA - DEFINICIJA

Akutna mijeloična leukemija (AML) klonalna je bolest krvotvornih matičnih stanica i najčešći tip leukemije u odrasloj populaciji (Vakiti i sur., 2023.). Nastaje uslijed nakupljanja genskih promjena u mijeloidnim prekursorskim stanicama (blastima). Genske promjene uzrokuju prednost preživljavanja i smanjenu sposobnost diferencijacije leukemijskih blasta što dovodi do njihove ekspanzije i nakupljanja u koštanoj srži. Pritom dolazi do potiskivanja normalne granulocitopoeze, eritrocitopoeze te trombocitopoeze. Zbog ekspanzije blasta i potiskivanja normalne funkcije koštane srži dolazi do kliničkih simptoma koji uključuju umor, dispneju i bolove u kostima. Uslijed ekspanzije leukemijskog klona može doći do infiltracije organa leukemijskim stanicama i metaboličkih komplikacija (Labar, 2017.).

Sumnja na AML postavlja se kod bolesnika sa znakovima krvarenja, učestalim infekcijama, hepatomegalijom ili splenomegalijom. Laboratorijski nalazi očituju se leukocitozom, prisustvom nezrelih stanica u razmazu periferne krvi, anemijom i trombocitopenijom.

Za postavljanje dijagnoze neophodan je multidisciplinarnan pristup koji uključuje citološku, citogenetsku i molekularnu analizu te analizu imunofenotipa protočnom citometrijom (Döhner i sur., 2022.). Na temelju sveobuhvatne analize bolesnike se svrstava u rizične skupine s pomoću kojih se može predvidjeti vjerojatnost postizanja remisije i rizik za relaps bolesti.

Unatoč znatnom razvoju terapijskih i dijagnostičkih mogućnosti petogodišnje preživljenje ostaje vrlo nisko. Predviđeno petogodišnje preživljenje kod bolesnika s AML iznosi 31 %. (<https://seer.cancer.gov/>). U mlađih bolesnika petogodišnje preživljenje iznosi 50 %, dok u starijih ono iznosi samo 10 % (Sasaki K i sur., 2021.).

## 1.2. EPIDEMIOLOGIJA I ETIOLOGIJA

AML obuhvaća otprilike 1 % svih tumora odrasle dobi. Incidencija AML-a u svijetu iznosi 3 do 5 oboljelih na 100,000 stanovnika. U odraslih medijan godina pri dijagnozi iznosi 68 godina (Juliusson i sur., 2009.), češće oboljevaju muškarci nego žene i to u omjeru 5:3. Glavni rizični čimbenik za razvoj AML-a je prethodna kronična mijeloproliferativna bolest koja najčešće nije prepoznata (mijelodisplastični sindrom, aplastična anemija, hipoplazija koštane srži i druge) (Shallis i sur., 2019.). Ostali rizični čimbenici uključuju nasljedne bolesti poput Downovog sindroma i Fanconijeve anemije kao i okolišni čimbenici poput izloženosti zračenju, benzenu, dimu cigareta i prijašnje liječenje (Shallis i sur., 2019.).

## 1.3. PATOFIZIOLOGIJA

Akutna mijeloična leukemija heterogena je skupina bolesti karakterizirana mutacijama gena koji su uključeni u hematopoezu. U 97 % bolesnika kojima je dijagnosticirana AML prisutna je barem jedna genska promjena (Papaemmanuil i sur., 2016.). Brojna istraživanja predlažu kako se u patofiziologiji bolesti razlikuju dva važna događaja od kojih je svaki karakteriziran posebnim tipom mutacija (Chen i sur., 2022.). Mutacije prvog tipa leukemijskom klonu omogućuju proliferaciju i prednost preživljavanja pred ostalim stanicama. Mutacije prvog tipa javljaju se u genima koji kodiraju za tirozin kinaze: FMS-slična-tirozin-kinaza (engl. *FMS-like tyrosine kinase-3, FLT3*), tirozin-protein kinaza (c-Kit) i *BCR::ABL1* (engl. *break point cluster region-abelson*) (Zhao i sur., 2020.). Mutacije drugog tipa dovode do nemogućnosti diferencijacije, sazrijevanja i problema s programiranom staničnom smrću. Mutacije drugog tipa najčešće se očituju u genima za nukleofosmin (*NPML1*) i *CEBPA* (engl. *CCAAT/enhancer-binding protein alpha*). Kombinacija navedenih genskih promjena naposljetku rezultira leukemogenezom i razvojem kliničkih simptoma.

## 1.4. PODJELA

Donedavno, podjela mijeloidnih neoplazmi temeljila se na četvrtom izdanju klasifikacije mijeloidnih neoplazmi Svjetske zdravstvene organizacije (SZO) (engl. World Health Organization, WHO) iz 2017. godine (Arber i sur., 2016.). Znatni napretci u području genetike doveli su do revizije smjernica i uvođenja novih podskupina bolesti koje se

primarno temelje na genskim promjenama (Huber i sur., 2023.). Slijedom novih otkrića, 2022. godine izlaze dvije različite klasifikacije mijeloidnih neoplazmi, 5. izdanje SZO klasifikacije hematolimfoidnih tumora (Allaggio i sur., 2022.) te klasifikacija radne grupe Međunarodne konsenzusne klasifikacije (engl. International consensus classification, ICC) (Arber i sur., 2022.).

Prema klasifikaciji SZO-a iz 2022. godine, AML se dijeli u dvije velike skupine: AML s definiranim genskim promjenama i AML koja je definirana diferencijacijom. Fokus nove klasifikacije preusmjerava se s morfoloških značajki na genske i citogenetske promjene. Prva skupina AML-a s definiranim genskim promjenama obuhvaća 11 zasebnih podtipova AML-a temeljenih na kromosomskim rearanžmanima i mutacijama u genima. Tablica 1 prikazuje podjelu AML-a s definiranim genskim promjenama.

Za postavljanje dijagnoze AML-a s jednom od definiranih genskih promjena nije nužno dokazati postojanje  $\geq 20\%$  blasta u perifernoj krvi ili koštanoj srži, osim u slučaju AML-a s fuzijskim prijepisom *BCR::ABL1* i AML s mutacijom u *CEBPA* genu. Navedena promjena u kriteriju za postavljanje dijagnoze temeljila se na istraživanjima koja prikazuju kako bolesnici s nižim brojem blasta i prisutnim genskim ili citogenetskim promjenama imaju slične simptome kao i bolesnici s većim brojem blasta (Alaggio i sur., 2022.). Nadalje, redefinirana je podskupina AML-a povezana s mijelodisplazijom (engl. *AML myelodysplasia-related*, AML-MR). AML-MR definirana je postojanjem  $\geq 20\%$  blasta u perifernoj krvi ili koštanoj srži koji posjeduju specifične citogenetske ili molekularne promjene povezane s mijelodisplastičnim sindromom (MDS) ili mijelodisplastičnim sindromom/mijeloproliferativnim neoplazmama (MDS/MPN). Naposljetku, u novu klasifikaciju uvedena je kategorija AML s drugim definiranim genskim promjenama u koju ulaze slučajevi s rijetko detektiranim genskim promjenama

Tablica 1. Podjela AML s definiranim genskim promjenama (Preuzeto, prevedeno i prilagođeno prema: Alaggio i sur., 2022.)

<b>AML s definiranim genskim promjenama</b>
Akutna promijelocitna leukemija s fuzijskim prijepisom <i>PML::RARA</i>
Akutna mijeloična leukemija s fuzijskim prijepisom <i>RUNX1::RUNX1T1</i>
Akutna mijeloična leukemija s fuzijskim prijepisom <i>CBFB::MYH11</i>



Akutna mijeloična leukemija s fuzijskim prijepisom <i>DEK::NUP214</i>
Akutna mijeloična leukemija s fuzijskim prijepisom <i>RBM15::MRTFA</i>
Akutna mijeloična leukemija s fuzijskim prijepisom <i>BCR::ABL1</i>
Akutna mijeloična leukemija s preuredbom <i>KMT2A</i>
Akutna mijeloična leukemija s preuredbom <i>MECOM</i>
Akutna mijeloična leukemija s preuredbom <i>NUP98</i>
Akutna mijeloična leukemija s mutacijom <i>NPM1</i>
Akutna mijeloična leukemija s mutacijom <i>CEBPA</i>
Akutna mijeloična leukemija, povezana s mijelodislazijom
Akutna mijeloična leukemija s drugim definiranim genetskim promjenama

Bolesnici kod kojih nije dokazano postojanje genskih promjena svrstavaju se u kategoriju AML koja je definirana diferencijacijom. Navedena podjela temelji se na specifičnim morfološkim karakteristikama sazrijevanja leukemijskih stanica. Podjela AML-a na temelju morfoloških karakteristika tradicionalno se koristila te posjeduje važnost u procjeni prognoze bolesti. Tablica 2 prikazuje podjelu AML na temelju morfoloških karakteristika sazrijevanja.

Tablica 2. Podjela AML definiranih diferencijacijom (Preuzeto, prevedeno i prilagođeno prema: Alaggio i sur., 2022.)

<b>AML koja je definirana diferencijacijom</b>
Akutna mijeloična leukemija s minimalnom diferencijacijom
Akutna mijeloična leukemija sa sazrijevanjem
Akutna mijeloična leukemija bez sazrijevanja
Akutna bazofilna leukemija
Akutna mijelomonocitna leukemija
Akutna monocitna leukemija
Akutna eritroidna leukemija
Akutna megakarioblastična leukemija

Podjela ICC-a temelji se na hijerarhijskom pristupu koji omogućuje podjelu AML na temelju genskih i citogenetskih promjena. Hijerarhijski pristup nastoji prioritizirati pojedine genske i citogenetske promjene koje su najvažnije za postavljanje dijagnoze i određivanje prognoze

bolesti. Nasuprot SZO-u, ICC zahtijeva postojanje  $\geq 10\%$  blasta u perifernoj krvi ili koštanoj srži kako bi se zadovoljili kriteriji za dijagnozu AML. Izdvaja se posebna podskupina AML-a s mutacijom u protoonkogenom tumorskom proteinu 53 (*TP53*). AML s mutacijom u *TP53* povezana je s vrlo lošom prognozom. Nadalje, uvedena je nova kategorija AML-a povezana s genskim promjenama vezanim uz mijelodisplaziju, a uključuje mutacije u sljedećim genima: *ASXL1*, *BCOR*, *EZH2*, *RUNX1*, *SF3B1*, *SRSF2*, *STAG2*, *U2AF1*, i /ili *ZRSR2* te strukturne i brojčane citogenetske promjene. U Tablici 3. prikazane su molekularne i citogenetske promjene povezane s pojedinom podskupinom AML. U hijerarhijskom pristupu klasifikacije, kategorija AML s ponavljajućim genskim promjenama ima prednost pred ostalim kategorijama. Ako je kod bolesnika dokazana genska ili citogenetska promjena koja je definirana u AML s ponavljajućim genskim promjenama ta promjena ima prioritet nad značajkama povezanim s mijelodisplazijom. Naposljetku, dio bolesnika imat će molekularne ili citogenetske promjene koje se ne mogu svrstati niti u jednu podskupinu te se navedeni bolesnici svrstavaju u AML koju nije moguće drugačije specificirati (engl. *AML- not otherwise specified*, AML-NOS) (Duffield i sur., 2023.).

Tablica 3. Podjela akutnih mijeloičnih leukemija (Preuzeto, prevedeno i prilagođeno prema: Döhner i sur., 2022. )

AML s ponavljajućim genskim promjenama
Kriterij $>10\%$ blasta u perifernoj krvi ili koštanoj srži
APL s t(15;17)(q24.1;q21.2)/ <i>PML::RARA</i>
AML s t(8;21)(q22;q22.1)/ <i>RUNX1::RUNX1T1</i>
AML s inv(16)(p13.1q22) ili t(16;16)(p13.1;q22)/ <i>CBFB::MYH11</i>
AML s t(9;11)(p21.3;q23.3)/ <i>MLLT3::KMT2A</i>
AML s t(6;9)(p22.3;q34.1)/ <i>DEK::NUP214</i>
AML s inv(3)(q21.3q26.2) ili t(3;3)(q21.3;q26.2)/ <i>GATA2</i> , <i>MECOM(EVII)</i>
AML s drugim rijetkim ponavljajućim translokacijama
AML s t(9;22)(q34.1;q11.2)/ <i>BCR::ABL1</i>
Kategorije označene kao AML (ako je $\geq 20\%$ blasta u perifernoj krvi ili koštanoj srži) ili MDS/AML (ako je 10-19% blasta u perifernoj krvi ili koštanoj srži)
AML s mutacijom u <i>TP53</i>

AML s mutacijama gena povezanih s mijelodisplastičnim sindromom - <i>ASXL1</i> , <i>BCOR</i> , <i>EZH2</i> , <i>RUNX1</i> , <i>SF3B1</i> , <i>SRSF2</i> , <i>STAG2</i> , <i>U2AF1</i> , i/ili <i>ZRSR2</i>
--

AML s citogenetskim promjenama povezanim s mijelodisplazijom
--

U klasifikaciji ICC-a uvode se dijagnostički kvalifikatori. U dijagnostičke kvalifikatore spadaju: AML povezana s terapijom, AML koja je napredovala iz prethodnog mijelodisplastičnog sindroma (MDS) te AML povezana s nasljednom predispozicijom. Navedena promjena uvedena je kako bi se smanjilo preklapanje pojedinih podskupina AML-a (Arber i sur., 2022.).

### 1.5. STRATIFIKACIJA RIZIKA

Iako se zadnjih godina uočava napredak u razvoju antileukemijskih lijekova, stope petogodišnjeg preživljavanja ostaju vrlo niske. AML je heterogena bolest i ishod bolesti ovisi o brojnim čimbenicima; dobi, drugim komorbiditetima, općem stanju bolesnika, parametrima krvne slike te genskim promjenama koje karakteriziraju bolest (Boscaro i sur., 2023.). Zbog saznanja koja se temelje na novim molekularnim otkrićima i rezultatima kliničkih ispitivanja, Europska leukemijska mreža (engl. *European Leukemia Net*, ELN) donijela je novi hodogram stratifikacije rizika. Naglasak se stavlja na genske i citogenetske promjene prema kojima se bolesnici svrstavaju u tri kategorije rizika: povoljnu, intermedijarnu i nepovoljnu (Tablica 4). Na temelju tih podataka određuje se pristup liječenju.

Tablica 4. Stratifikacija rizika prema Europskoj leukemijskoj mreži na temelju utvrđenih genskih i citogenetskih abnormalnosti pri postavljanju dijagnoze (Preuzeto, prevedeno i prilagođeno prema: Döhner i sur., 2022. )

Kategorija rizika	Genske promjene
Povoljna	<ul style="list-style-type: none"> <li>- t(8;21)(q22;q22.1)/<i>RUNX1::RUNX1T1</i></li> <li>- inv(16)(p13.1q22) ili t(16;16)(p13.1;q22)/<i>CBFB::MYH11</i></li> <li>- Mutacija <i>NPM1</i> bez <i>FLT3</i>-ITD</li> <li>- bZIP in-frame mutacija <i>CEBPA</i></li> </ul>

Intermedijarnu	<ul style="list-style-type: none"> <li>- Mutacija u <i>NPM1</i> u kombinaciji s <b>FLT3-ITD</b></li> <li>- Divlji tip <i>NPM1</i> zajedno s <i>FLT3-ITD</i> (bez nepovoljnih genskih abnormalnosti)</li> <li>- <i>t(9;11)(p21.3;q23.3)/MLLT3::<i>KMT2A</i></i></li> <li>- Citogenske i/ili molekularne abnormalnosti koje nisu klasificirane kao povoljne ili nepovoljne</li> </ul>
Nepovoljna	<ul style="list-style-type: none"> <li>- <i>t(6;9)(p23.3;q34.1)/DEK::<i>NUP214</i></i></li> <li>- <i>t(v;11q23.3)/KMT2A</i>-rearanžman</li> <li>- <i>t(9;22)(q34.1;q11.2)/BCR::<i>ABL1</i></i></li> <li>- <i>t(8;16)(p11.2;p13.3)/KAT6A::<i>CREBBP</i></i></li> <li>- <i>inv(3)(q21.3q26.2)</i> ili <i>t(3;3)(q21.3;q26.2)/GATA2, MECOM(<i>EVII</i>)</i></li> <li>- <i>t(3q26.2;v)/MECOM(<i>EVII</i>)</i>-rearanžman</li> <li>- 25 ili <i>del(5q); 27; 217/abn(17p)</i></li> <li>- Kompleksni kariotip, monosomni kariotip</li> <li>- Mutacija <i>ASXL1, BCOR, EZH2, RUNX1, SF3B1, SRSF2, STAG2, U2AF1</i>, i/ili <i>ZRSR2</i></li> <li>- Mutacija <i>TP53</i></li> </ul>

U povoljnu kategoriju rizika spadaju bolesnici s CBF (engl. *Core Binding Factor*) leukemijama te bolesnici kojima su dokazane citogenske promjene *t(8;21)(q22;q22)* i *inv(16)(p13.1q22)* ili *t(16;16)(p13.1;q22)*. Navedene promjene češće se javljaju kod mlađih bolesnika i povezane su s boljim odgovorom na kemoterapiju (Boscaro i sur., 2023.). U kategoriji intermedijarnog rizika su bolesnici s *t(9;11)(p21.3;q23.3)/MLLT3::*KMT2A**, koji dobro reagiraju na liječenje kemoterapijom (Meyer i sur., 2018.). Bolesnici s više od tri citogenetske promjene na različitim kromosomima imaju takozvani kompleksni kariotip i spadaju u kategoriju nepovoljnog rizika. Istraživanja pokazuju kako je petogodišnje preživljavanje ovih bolesnika znatno niže nego u ostalih bolesnika (Ezekwudo, 2017.).

Osim kromosomskih promjena ističe se važnost i različitih genskih mutacija. U kategoriju povoljnog rizika ulaze bolesnici s dokazanom mutacijom u genu za *NPM1* a bez mutacije u genu za *FLT3* s internom tandemskom duplikacijom (*FLT3-ITD*). Dodatno u skupinu povoljnog rizika svrstavaju se bolesnici s mutacijama u *CEBPA* genu, neovisno o tome pojavljuju li se mutacije u navedenom genu kao bialelne ili monoalelne. Bolesnici s *FLT3-ITD* svrstavaju se u kategoriju srednjeg rizika neovisno o udjelu alela ili postojećim mutacijama u genu za *NPM1*. Naposljetku, bolesnici kojima je dokazana prisutnost mutacija povezanih s mijelodisplazijom svrstavaju se u kategoriju visokog rizika (Döhner i sur., 2022.).

## **1.6. DIJAGNOSTIKA AKUTNE MIJELOIČNE LEUKEMIJE**

Kako bi se postavila točna dijagnoza primjenjuje se multidisciplinarni pristup koji uključuje citomorfološku analizu, citogenetiku, imunofenotipizaciju i molekularnu dijagnostiku.

### **1.6.1. Citomorfološka dijagnostika**

Citomorfološka analiza provodi se iz uzorka aspirata ili bioptata koštane srži te uzorka periferne krvi. Prvi korak citomorfološke analize uključuje mikroskopsko ispitivanje celularnosti koštane srži. Nakon procjene celularnosti slijedi morfološka analiza stanica prisutnih u koštanoj srži. Ako se morfološkom analizom utvrdi da leukemijski blasti čine  $\geq 20\%$  stanica s jezgrom, može se postaviti sumnja na akutnu leukemiju. Na temelju citomorfološke analize nije moguće odrediti porijeklo blasta već je nužno provesti daljnje korake koje uključuju citokemijsko bojanje i imunofenotipizaciju. Usporedno s analizom uzorka koštane srži provodi se i morfološka analiza stanica prisutnih u perifernoj krvi (McKenzie i sur., 2020.).

### **1.6.2. Citogenetska dijagnostika**

Citogenetska analiza ima ključnu ulogu u klasifikaciji podskupina AML-a. Citogenetske promjene detektirane su u 50 do 60% slučajeva AML (Kumar, 2011.). U obje klasifikacije, SZO i ICC postoje kategorije AML s ponavljajućim genetskim abnormalnostima. Uzorak za citogenetsku analizu dobiva se aspiracijom ili biopsijom koštane srži. Nakon prikupljanja

uzorka potrebno je uzgojiti kratkotrajnu staničnu kulturu iz koje se zatim provode daljnje analize. Citogenetske analize koje se provode u polju molekularne hematologije uključuju metode konvencionalne citogenetike i fluorescentnu *in situ* hibridizaciju (FISH). U metode klasične citogenetike ubraja se GTG-pruganje (Giemsa-Tripsin-Giemsa). Uz pomoć metoda klasične citogenetike prikupljaju se informacije o morfoloiji i kompleksnosti promjena u kariotipu (Keinanen M i sur. 1986). FISH je tehnologija koja se temelji na vezanju fluorescentno obilježenih proba na komplementarna mjesta u genomu. S pomoću FISH-a se mogu odrediti poznate translokacije, delecije, insercije ili duplikacije pojedinih dijelova kromosoma.

### **1.6.3. Imunofenotipizacija – protočna citometrija**

Imunofenotipizacija je metoda temeljena na principu protočne citometrije. U sklopu imunofenotipizacije izvodi se analiza uzorka periferne krvi i aspirata koštane srži. Postupak imunofenotipizacije omogućuje analizu karakteristika leukemijskih stanica na principu vezanja fluorescentno obilježenih antitijela na specifične unutarstanične i izvanstanične antigene. U sklopu dijagnostike AML-a provodi se analiza sljedećih antigena koji su povezani s mijeloidnim stanicama: CD2 (engl. *Cluster of differentiation* -CD), CD3, CD5, CD7, CD10, CD13, CD14, CD15, CD19, CD22, CD33, MPO (mijeloperoksidaza), HLA-DR (humani leukocitni antigen), CD34, CD56, CD45 i CD117 (Pereira, 2006.). Jedna od glavnih uloga imunofenotipizacije u dijagnostici AML-a jest razlikovanje akutne mijeloidne leukemije od akutne limfoblastične leukemije (ALL). Nadalje, imunofenotipizacija se koristi u praćenju minimalne ostatne bolesti (engl. *Minimal residual disease* - MRD). Praćenjem minimalne ostatne bolesti uz pomoć imunofenotipizacije utvrđuje se prisutnost početnog leukemijskog klonova i nastanak novih klonova što je korisno u praćenju kinetike bolesti i odgovora na terapiju (Tembhare, 2023.).

### **1.6.4. Molekularna dijagnostika**

Molekularno-dijagnostički testovi omogućuju detekciju promjena u deoksiribonukleinskoj kiselini (engl. *DeoxyriboNucleid Acid*, DNA) i ribonukleinskoj kiselini (engl. *RiboNucleid Acid*, RNA). Na temelju detektiranih promjena u DNA i RNA donose se odluke o: klasifikaciji bolesti, postavljanju dijagnoze, stratifikaciji rizika i odabiru terapije (Dahui,

2022.). U okviru dijagnostike AML-a najčešće se koriste metode molekularne dijagnostike koje se zasnivaju na principu umnožavanja specifičnih sljedova DNA uz pomoć enzima DNA polimeraze - lančana reakcija polimerazom (engl. *Polymerase chain reaction*, PCR) te detekciji umnoženog fragmenta DNA različitim metodama. Uz klasični PCR koristi se još Sangerovo sekvenciranje i sekvenciranje sljedeće generacije (engl. *Next generation sequencing*, NGS). Kao uzorak za molekularno dijagnostičke testove koristi se aspirat koštane srži i uzorak periferne krvi. U aspektu dijagnostike AML-a, PCR se koristi za postavljanje dijagnoze i praćenje kinetike bolesti. Prilikom postavljanja dijagnoze PCR omogućuje brzu detekciju mutacija u genima koji su često mutirani u bolesnika oboljelih od AML-a. Neki od primjera uključuju dokazivanje mutacija u: *FLT3*, *NPM1* i *CEBPA*. Lančana reakcija polimeraze u stvarnom vremenu (engl. *Real-time quantitative polymerase chain reaction*, RT-qPCR) koristi se u praćenju minimalne ostatne bolesti (engl. *Minimal residual disease*, MRD) te pruža dodatne informacije o uspješnosti liječenja i doprinosi kliničkoj odluci. Sangerovo sekvenciranje omogućuje detekciju točkastih mutacija, delecija i insercija. Primjena Sangerovog sekvenciranja u dijagnostici AML-a odnosi se na potvrdu mutacija dokazanih drugim metodama poput PCR-a ili NGS-a.

Sekvenciranje sljedeće generacije (engl. *Next generation sequencing* - NGS) tehnologija je koja omogućuje istovremeno sekvenciranje velikog broja gena. Tehnologija sekvenciranja sljedeće generacije temelji se na višestrukum umnožavanju pojedinih DNA fragmenata te njihovom istovremenom sekvenciranju. Postoji više vrsta NGS-a koje uključuju: sekvenciranje cijelog genoma, sekvenciranje cijelog egzoma i sekvenciranje pojedinih regija u genomu (ciljano sekvenciranje). U okviru dijagnostike AML-a koriste se ciljani paneli koji obuhvaćaju najčešće mutirane gene. Ciljanim panelima moguće je obuhvatiti sve gene koji su povezani s klasifikacijom bolesti, stratifikacijom rizika i procjenom odgovora na terapiju (Haferlach, 2020.). Trenutne smjernice Europske leukemijske mreže preporučuju gensku analizu svih novodijagnosticiranih bolesnika s AML-om (Döhner i sur., 2022. ). Tablica 5 prikazuje gene na koje je potrebno izvršiti probir prilikom dijagnosticiranja novodijagnosticiranih bolesnika s AML-om.

Tablica 5. Geni povezani s postavljanjem dijagnoze i praćenjem bolesti (Preuzeto, prevedeno i prilagođeno prema: Döhner i sur., 2022. )

<i>FLT3</i>	<i>IDH1</i>	<i>IDH2</i>	<i>NPM1</i>	<i>CEBPA</i>
<i>DDX41</i>	<i>TP53</i>	<i>ASXL1</i>	<i>BCOR</i>	<i>EZH2</i>
<i>RUNX1</i>	<i>SF3B1</i>	<i>SRSF2</i>	<i>STAG2</i>	<i>U2AF1</i>
<i>ZRSR2</i>	<i>ANKRD26</i>	<i>BCORL1</i>	<i>BRAF</i>	<i>CBL</i>
<i>CSF3R</i>	<i>DNMT3A</i>	<i>ETV6</i>	<i>GATA2</i>	<i>JAK2</i>
<i>KIT</i>	<i>KRAS</i>	<i>NRAS</i>	<i>NF1</i>	<i>PHF6</i>
<i>PPM1D</i>	<i>PTPN11</i>	<i>RAD21</i>	<i>SETBP1</i>	<i>TET2</i>
<i>WT1</i>				

## 1.7. SEKVENCIRANJE SLJEDEĆE GENERACIJE

Sekvenciranje sljedeće generacije (engl. *Next Generation Sequencing*, NGS) je skup tehnologija temeljenih na mogućnosti masivnog paralelnog sekvenciranja DNA. Metode NGS-a omogućuju istovremeno sekvenciranje stotina gena iz više uzoraka. S pomoću NGS-a moguće je detektirati različite genske promjene poput: mutacija pogrešnog smisla (engl. *missense mutations*), besmislenih mutacije (engl. *nonsense mutations*) te malih delecija i insercija, a dodatno NGSom se mogu detektirati strukturalne promjene, promjene u broju kopija i promjene u ekspresiji gena. Mogućnost masivnog paralelnog sekvenciranja dovela je do smanjenja cijene i vremena potrebnog za analizu.

Postupak sekvenciranja sastoji se od tri koraka: pripreme knjižnice, klonalne amplifikacije te sekvenciranja. Tijekom procesa pripreme knjižnice dolazi do fragmentacije DNA molekule te dodavanja adapterskih sekvenci. U procesu klonalne amplifikacije fragmenti DNA se klonalno umnožavaju. Naposljetku dolazi do čitanja sekvence umnoženih fragmenata DNA. Postoji više tehnologija sekvenciranja, a neke od najčešće korištenih uključuju sekvenciranje na temelju sinteze, pirosekvenciranje, sekvenciranje pomoću reverzibilne terminatorne kemije te sekvenciranje nanoporama.



### **1.7.1. Ciljano sekvenciranje**

U slučajevima kada se istražuje samo specifični dio genoma pogodno je koristiti ciljano sekvenciranje. Ciljano sekvenciranje osigurava veću dubinu čitanja te smanjuje cijenu i vrijeme potrebno za sekvenciranje (Xuan i sur., 2012.). Dubina čitanja daje informaciju o ukupnom broju očitavanja jednog nukleotida u reakciji sekvenciranja. Velika dubina čitanja osigurava pouzdanost prilikom detekcije varijante te omogućava detekciju varijanti s niskom frekvencijom. Sekvenciranje uz veliku dubinu čitanja je posebno važno prilikom analize somatskih mutacija u tumorima. Zaključno, ciljano sekvenciranje omogućuje brzo postavljanje dijagnoze brojnih genetskih bolesti (Grada i Weinbrecht, 2013.).

Postoje dva tipa ciljanog sekvenciranja: ciljano obogaćivanje i sekvenciranje na temelju kloniranih segmenata. Sekvenciranje na temelju kloniranih segmenata vrsta je ciljanog sekvenciranja koja omogućuje detekciju rijetkih genskih promjena u specifičnim genskim regijama. Ovakva vrsta sekvenciranja omogućuje visoku dubinu čitanja pojedinih regija genoma te je korisna u detekciji somatskih varijanti koje imaju nisku frekvenciju u općoj populaciji. Temelji se na umnožavanju specifičnih regija u genomu. To se postiže dizajnom specifičnih početnica koje obuhvaćaju samo ispitivane gene. Obuhvaćene regije se zatim umnožavaju višestrukim PCR-om te naposljetku sekvenciraju uz pomoć NGS tehnologije (<https://www.illumina.com/>).

Ciljano obogaćivanje je vrsta ciljanog sekvenciranja koja se temelji se na izdvajanju i umnožavanju specifičnih regija genoma. Princip ciljanog obogaćivanja temelji se na hibridizaciji DNA na kuglice koje su obložene RNA i biotinom. Tijekom procesa ciljanog obogaćivanja DNA se hibridizira na kuglice koje se zatim izoliraju iz smjese te se vezana DNA eluira s njih i umnaža PCR reakcijom. Proces ciljanog obogaćivanja omogućuje izdvajanje većeg broja gena u odnosu na sekvenciranje kloniranih segmenata (<https://www.illumina.com/>).

## **1.8. BIOINFORMATIČKA OBRADA PODATAKA**

Procesi sekvenciranja sljedeće generacije rezultiraju nastankom velike količine sirovih podataka. Analiza NGS podataka kompleksan je i bioinformatički zahtjevan proces. U analizi podataka razlikuju se tri osnovna koraka: primarna, sekundarna i tercijarna analiza (<https://www.thermofisher.com/>).

Primarna analiza najčešće se odvija na samom uređaju za sekvenciranje i obuhvaća konverziju sirovih signala u nukleotidne baze s povezanom ocjenom kvalitete (Oliver, 2015.). Primarna analiza osigurava kvalitetu podataka koji ulaze u daljnje korake obrade. Prilikom procesa sekvenciranja dolazi do istovremenog sekvenciranja uzoraka na istom instrumentu. Prije početka daljnje analize podataka nužno je razdvojiti podatke o pojedinim uzorcima u zasebne datoteke. Proces razdvajanja podataka u zasebne datoteke naziva se demultipleksiranje (engl. *demultiplexing*).

U sekundarnoj analizi dolazi do poravnanja detektiranih sekvenci s referentnim genomom. Poravnanje s referentnim genomom se provodi pomoću BWA-MEM algoritma (Li i Durbin, 2009.). Tijekom postupka poravnanja pojedinom očitavanju dodjeljuje se lokacija u genomu. Proces poravnanja s referentnim genomom rezultira nastankom BAM (engl. *Binary Alignment Map*, BAM) datoteke koja se zatim koristi u daljnjim procesima obrade podataka. Posljednji korak sekundarne analize podataka je pozivanje varijanti (engl. *variant calling*). Pozivanje varijanti je proces tijekom kojeg dolazi do identifikacije razlika u nukleotidima između referentnog i ispitivanog genoma. Pomoću procesa pozivanja varijanti dolazi do detekcije polimorfizama jednog nukleotida (engl. *single nucleotide polymorphisms*, SNP), malih delecija, insercija i strukturnih varijanti. Bioinformatički alati koji provode proces pozivanja varijanti koriste statističke modele na temelju kojih procjenjuju vjerojatnost postojanja varijante na određenom mjestu u genomu. Prilikom pozivanja varijanti uzimaju se u obzir kvaliteta očitavanja baze, frekvencija alela u uzorku i broj očitavanja s promjenom.

Tercijarna analiza posljednji je korak interpretacije podataka sekvenciranja. Tercijarna analiza započinje utvrđivanjem i dodjeljivanjem (engl. *annotation*) varijanti. Prilikom procesa anotacije varijanti prikupljaju se funkcionalne i klinički važne informacije o varijantama. Prvi korak anotacije odnosi se na dodjeljivanje dodatnih informacija o varijanti koje uključuju: lokaciju varijante unutar genoma, vrstu varijante i HGVS nazivlje (engl. *Human Genome Variation Society*).

Zatim slijedi funkcionalna anotacija tijekom koje se dodjeljuju informacije o utjecaju varijante na trodimenzionalnu strukturu proteina, promjenu aminokiseline i izračunava se rezultat evolucijske očuvanosti (Oliver i ostali, 2015.). Pomoću funkcionalnih anotacija nastoji se predvidjeti utjecaj promjene nukleotida na funkciju proteina. Informacije o utjecaju promjene nukleotida na funkciju proteina prikupljaju se uz pomoć predikcijskih programa

kao što su: PolyPhen-2, SIFT (engl. *Sorting Intolerant From Tolerant*), CADD (engl. *Combined Annotation Dependent Depletion*), VEP (engl. *Variant Effect Predictor*).

Informacije o učestalosti varijanta u populaciji prikupljaju se iz populacijskih baza podataka kao što su 1000Genome ili gnomAD. Pregledom populacijskih baza podataka utvrđuje se frekvencija varijante u općoj populaciji, a ukoliko je varijanta u općoj populaciji zastupljena >1% smatra se polimorfizmom.

Podatci o kliničkom značaju pojedine varijante prikupljaju se pretraživanjem literature i kliničkih baza podataka. Prilikom prikupljanja klinički važnih informacija pretražuju se baze podataka poput ClinVar, OMIM (engl. *Online Mendelian Inheritance in Man*), dbSNP (engl. *Single Nucleotide Polymorphism Database*). U procesu analize tumorskih varijanti dodatno se mogu koristiti baze podataka poput COSMIC-a (engl. *Catalogue Of Somatic Mutations In Cancer*) koji sadrži informacije o prisutnosti somatskih varijanta u pojedinim tumorima.

## 1.9. INTERPRETACIJA DETEKTIRANIH VARIJANTI

Interpretacija detektiranih varijanti uključuje proces kojim se nastoji utvrditi ima li pojedina varijanta utjecaj na bolest s kojom se bolesnik očituje. Kako bi se ujednačio postupak interpretacije varijanti Američko društvo za medicinsku genetiku i genomiku (engl. *American College of Medical Genetics and Genomics - ACMG*) je 2015. godine razvilo preporuke za klasifikaciju nasljednih varijanti (Richards i sur., 2015.). Naglasak se stavlja na razmatranje i uključivanje višestrukih linija dokaza kako bi se točno utvrdila značajnost pojedine varijante. Prilikom interpretacija uzimaju se u obzir: zastupljenost varijante u populaciji, utjecaj varijante na funkciju proteina, obiteljska anamneza i brojni ostali čimbenici.

Kako bi se ujednačilo klasificiranje varijanti smjernice preporučuju podjelu u pet kategorija: benigne (engl. *Benign - B*), vjerojatno benigne (engl. *Likely benign - LB*), varijante nepoznatog značaja (*Variants of uncertain significance - VUS*), vjerojatno patogene (engl. *Likely pathogenic - LP*) i patogene (engl. *Pathogenic - P*). Podjela varijanti prema ACMG prikazana je u Tablici 6.

Tablica 6. Podjela varijanti prema ACMG 2015:

Benigne varijante (engl. <i>Benign - B</i> )
Vjerojatno benigne (engl. <i>Likely benign - LB</i> )
Varijante nepoznatog značaja(engl. <i>Variants of uncertain significance - VUS</i> )
Vjerojatno patogene (engl. <i>Likely pathogenic - LP</i> )
Patogene (engl. <i>Pathogenic - P</i> )

Prilikom određivanja patogenosti varijante svakoj varijanti se dodjeljuju pravila s određenom snagom dokaza. Na temelju dodijeljenih pravila varijanta se svrstava u jednu od pet kategorija patogenosti (Richards i sur., 2015.).

Kako bi se olakšao proces interpretacije somatskih varijanti, 2017. godine Udruga za molekularnu patologiju, (engl. *Association for Molecular Pathology, AMP*) donosi smjernice za interpretaciju somatskih varijanti (Li i sur., 2017.). Prema navedenim smjernicama somatske varijante se dijele u četiri razreda. Prvi razred uključuje varijante od snažne kliničke značajnosti, drugi razred uključuje varijante od potencijalne kliničke značajnosti, treći razred uključuje varijante od nepoznate kliničke značajnosti i četvrti razred uključuje benigne varijante (Li i sur., 2017.).

## 2. OBRAZLOŽENJE TEME

Akutna mijeloična leukemija je genski heterogena bolest. Posljednje smjernice za postavljanje dijagnoze i stratifikaciju rizika bolesnika oboljelih od AML naglašavaju važnost detekcije genskih i citogenetskih promjena (Döhner i sur., 2022.). Bolesnici se na temelju detektiranih genskih i citogenetskih promjena svrstavaju u rizične skupine na temelju kojih se procjenjuje vjerojatnost postizanja remisije i preživljenja. Jedan od procesa određivanja mutacija u genima vezanim uz AML je sekvenciranje sljedeće generacije. Sekvenciranjem sljedeće generacije moguće je istovremeno detektirati velik broj genskih promjena u ispitivanim regijama genoma.

Tijekom procesa sekvenciranja dolazi do detekcije velikog broja promjena u ljudskom genomu od kojih većina nije patogena ni klinički značajna. Kako bi se proces interpretacije detektiranih varijanti standardizirao i ubrzao, razvijeni su bioinformatički alati koji imaju sposobnost automatske klasifikacije detektiranih varijanti. Bioinformatički alati omogućuju medicinskim djelatnicima bržu klasifikaciju i interpretaciju detektiranih varijanti.

Cilj ovog rada je evaluacija *in house* postupka analize podataka dobivenih sekvenciranjem sljedeće generacije za mijeloidni panel, utvrđivanje stupnja povezanosti dva bioinformatička alata za analizu podataka dobivenih tehnologijom NGS-a te utvrditi postoje li odstupanja između njih.

### **3. MATERIJALI I METODE**

#### **3.1. UZORCI**

U istraživanju je korišten 41 uzorak aspirata koštane srži bolesnika oboljelih od akutne mijeloične leukemije. Uzorci su obrađeni od 2020. do 2023. godine u Kliničkom zavodu za laboratorijsku dijagnostiku, Kliničkog bolničkog centra Zagreb.

#### **3.2. IZDVAJANJE NUKLEINSKIH KISELINA**

Izolacija genomske DNA iz leukemijskih stanica provedena je iz uzorka aspirata koštane srži. Genomska DNA izolirana je automatiziranom metodom temeljenom na korištenju silikatnih stupaca pomoću uređaja QIAcube Kit (QIAGEN, Aarhus, Danska). Za izolaciju DNA korišten je reagens QIAamp DNA Blood Mini Kit (QIAGEN, Aarhus, Danska). Proces izolacije uključuje uporabu lizirajućeg pufera i proteinaze K uz pomoć kojih dolazi do liziranja stanica i razgradnje proteina. Oslobođena DNA veže se na QIAamp silika membranu tijekom procesa centrifugiranja. Elucija DNA sa silika membrane izvodi se uz pomoć elucijskog pufera pri sobnoj temperaturi.

##### **3.2.1. Određivanje koncentracije izolirane DNA**

Nakon izolacije određena je koncentracija izolirane DNA. Za određivanje koncentracije DNA korišten je fluorimetar Qubit 3.0 (Thermo Fisher Scientific, Waltham, Massachusetts, SAD) s reagensom Qubit dsDNA Broad Range Assay kit (Thermo Fisher Scientific, Waltham, Massachusetts, SAD). Qubit određuje fluorescenciju koja nastane nakon što fotodioda obasja fluorescentnu boju koja se veže na dvolančanu DNA. Uz pomoć kalibracijske krivulje određuje se koncentracije DNA u ng/mL.

#### **3.3. SEKVENCIRANJE SLJEDEĆE GENERACIJE**

Sekvenciranje je provedeno na uređaju MiSeq (Illumina, San Diego, Kalifornija, SAD) pomoću reagensa Illumina paired-end cluster generation kit (Illumina, San Diego, Kalifornija, SAD) s 150 ciklusa i duljine čitanja 76pb na uparenom kraju. Korišten je prilagođeno razvijen panel sonda proizvođača Agilent Technologies (Agilent Technologies,

Santa Clara, Kalifornija, SAD). Sonde su obuhvatile gene povezane s akutnom mijeloičnom leukemijom. Geni obuhvaćeni ovim panelom sonda prikazani su u Tablici 7.

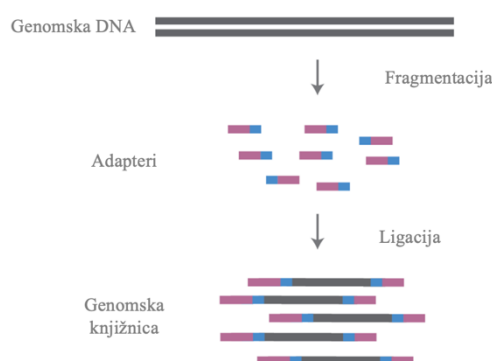
Tablica 7. Geni obuhvaćeni panelom sonda za potrebe sekvenciranja genoma bolesnika oboljelih od akutne mijeloične leukemije.

<i>ABL1</i>	<i>CEBPA</i>	<i>HRAS</i>	<i>MYD88</i>	<i>SF3B1</i>
<i>ASXL1</i>	<i>CSF3R</i>	<i>IDH1</i>	<i>NOTCH1</i>	<i>SMC1A</i>
<i>ATRX</i>	<i>CUX1</i>	<i>IDH2</i>	<i>NPM1</i>	<i>SMC3</i>
<i>BCOR</i>	<i>DNMT3A</i>	<i>IKZF1</i>	<i>NRAS</i>	<i>SRSF2</i>
<i>BCORL1</i>	<i>ETV6/TEL</i>	<i>JAK2</i>	<i>PDGFRA</i>	<i>STAG2</i>
<i>BRAF</i>	<i>EZH2</i>	<i>JAK3</i>	<i>PHF6</i>	<i>TET2</i>
<i>CALR</i>	<i>FBXW7</i>	<i>KDM6A</i>	<i>PTEN</i>	<i>TP53</i>
<i>CBL</i>	<i>FLT3</i>	<i>KIT</i>	<i>PTPN11</i>	<i>U2AF1</i>
<i>CBLB</i>	<i>GATA1</i>	<i>KRAS</i>	<i>RAD21</i>	<i>WT1</i>
<i>CBLC</i>	<i>GATA2</i>	<i>MLL</i>	<i>RUNX1</i>	<i>ZRSR2</i>
<i>CDKN2A</i>	<i>GNAS</i>	<i>MPL</i>	<i>SETBP1</i>	

### 3.3.1. Priprema knjižnice

Priprema knjižnice obuhvaća pripremu DNA za sekvenciranje te ciljano obogaćivanje specifičnih regija u genomu. Tijekom procesa pripreme knjižnice izvodi se fragmentacija DNA, dodavanje adapterskih sekvenci i umnožavanje fragmentiranih oligonukleotida. Proces enzimatskog cijepanja uz dodavanje adapterskih sekvenci naziva se tagmentacija DNA. Adapterske sekvence građene su od tri dijela: indeksa, veznog mjesta za početnice i komplementarnog slijeda oligonukleotida. Indeksi su kratki sljedovi DNA koji omogućuju označavanje pojedinih DNA fragmenata. Indeksi omogućuju identifikaciju DNA fragmenata unutar knjižnice koja sadrži više različitih uzoraka. Vezna mjesta za početnice omogućuju vezanje početnica tijekom procesa sekvenciranja. Adapterske sekvence omogućuju vezanje fragmentirane DNA na protočnu ćeliju (engl. *flow cell*). Naposljetku dolazi do PCR amplifikacije te pročišćavanja dobivene knjižnice uz pomoć magnetnih kuglica. Proces pripreme knjižnice objašnjen je na Slici 1.

### A. Priprema knjižnice



Slika 1. Proces pripreme genomske knjižnice (Preuzeto, prevedeno i prilagođeno s <https://www.illumina.com/>)

### 3.3.2. Ciljano obogaćivanje knjižnice

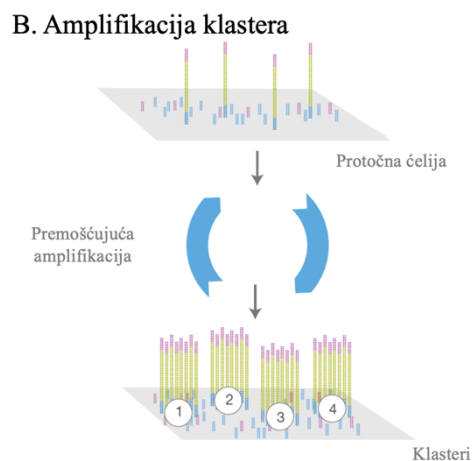
Proces ciljanog obogaćivanja omogućuje amplifikaciju i izolaciju specifičnih regija DNA uz pomoć biotiniziranih kuglica. Ciljano obogaćivanje izvršeno je pomoću reagensa SureSelect<sup>XT</sup> Target Enrichment System for the Illumina Platform (Agilent Technologies, Santa Clara, Kalifornija, SAD). Za izdvajanje željenih regija genoma korištene su biotinizirane sonde. Ciljane DNA regije hibridiziraju se sa specifičnim biotiniziranim sondama. Nakon hibridizacije, kompleksi DNA i biotiniziranih sonda vežu se za magnetne kuglice obložene streptavidinom. Magnetne kuglice na kojima je vezana DNA se zatim izoliraju iz smjese, slijedi elucija DNA s magnetnih kuglica uz pomoć elucijskog pufera, a aposljjetku dolazi do amplifikacije eluirane DNA uz pomoć PCR-a.

### 3.3.3. Sekvenciranje sintezom

Reakcija sekvenciranja odvija se na protočnoj ćeliji. Protočna ćelija je staklena podloga koja sadrži oligonukleotide komplementarne adapterskim sekvencama. Fragmentirana DNA nanosi se na protočnu ćeliju te dolazi do hibridizacije adapterskih sekvenci i oligonukleotida na protočnoj ćeliji. Nakon vezanja fragmentirane DNA dolazi do stvaranja komplementarnog lanca uz pomoć DNA polimeraze. Novonastala dvolančana DNA se zatim denaturira i dolazi



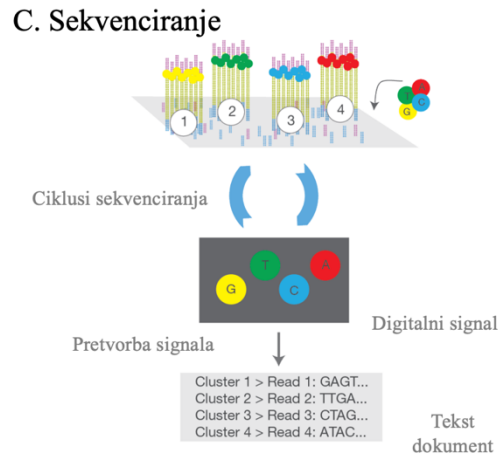
do ispiranja oligonukleotida koji su bili vezani na protočnu ćeliju. Oligonukleotidni lanci koji su ostali vezani za protočnu ćeliju se zatim umnožavaju procesom premošćujućeg PCR-a (engl. *bridge polymerase chain reaction; bridge amplification*). Tijekom procesa premošćujućeg PCR-a dolazi do stvaranja tisuća kopija jednog DNA fragmenta što rezultira nastankom lokalnih klastera. Lokalni klasteri (engl. *local clusters*) predstavljaju područja na protočnoj ćeliji gdje se nalaze tisuće kopija određenog DNA fragmenta. Nastankom lokalnih klastera povećava se jačina signala koja je potrebna za sekvenciranje (<https://www.illumina.com/>). Proces premošćujućeg PCR-a i nastanka lokalnih klastera prikazan je na Slici 2.



Slika 2. Proces premošćujućeg PCR-a i nastanka lokalnih klastera (Preuzeto, prevedeno i prilagođeno s <https://www.illumina.com/>)

Sekvenciranje sintezom temelji se na sposobnosti DNA polimeraze da ukomponira fluorescentno obilježene deoksiribonukleotid trifosfate (engl. *deoxyribose nucleotide triphosphate, dNTP*) u rastući lanac DNA. 3' kraj dNTP-ova zaštićen je protektivnom grupom koja osigurava da se u rastući lanac DNA dodaje samo jedan dNTP. Nakon ugradnje dNTP-a dolazi do ispiranja DNA polimeraze i slobodnih dNTP-ova. Zatim se izvodi proces ekscitacije fluorofora uz pomoć lasera te detekcije signala optičkim čitačem. Uz pomoć kompjuterskog programa optički signal se prevodi u binarni signal. Nakon detekcije signala dodaje se regens koji uklanja protektivnu skupinu s 3' kraja dNTP-a i

započinje sljedeći ciklus (<https://www.illumina.com/>). Proces sekvenciranja sintezom prikazan je na slici 3.



Slika 3. Proces sekvenciranja sintezom (Preuzeto, prevedeno i prilagođeno s <https://www.illumina.com/>)

### 3.3.4. Sekvenciranje uparenih krajeva

Sekvenciranje uparenih krajeva je tehnologija sekvenciranja sljedeće generacije koja omogućava dobivanje više informacija o ispitivanoj regiji genoma. Sekvenciranje uparenih krajeva temelji se na sekvenciranju oba kraja DNA fragmenta. Za provedbu sekvenciranja uparenih krajeva potrebna je specifična priprema knjižnice tijekom koje se na fragmente DNA dodaju adapteri s obje strane DNA fragmenta. Dvostruki adapteri omogućuju da se sekvenciranje DNA vrši s obje strane DNA fragmenta. Sekvenciranje započinje s čitanjem vodiljnog DNA lanca (engl. *forward strand*) prilikom kojeg se dodaje početnica koja omogućuje početak sekvenciranja. Tijekom sekvenciranja vodiljnog DNA lanca ciklično ugrađuju fluorescentno obilježeni nukleotidi i detektira se signal u pojedinom klasteru na protočnoj ćeliji. Nakon završetka sekvenciranja vodiljnog lanca DNA, dolazi do odstranjivanja lanca proizvedenog reakcijom sekvenciranja. Vodiljni lanac DNA ostaje vezan na protočnu ćeliju te se premošćuje na susjedni oligonukleotid koji se nalazi na protočnoj ćeliji. U sljedećem koraku DNA polimeraza sintetizira komplementarni DNA lanac prilikom čega se stvara dvolančana DNA molekula. Novonastala dvolančana DNA molekula se zatim denaturira, nakon čega dolazi do ispiranja vodiljnog lanca DNA. Nakon ispiranja vodiljnog

lanca na protočnoj ćeliji ostaje vezan samo obrnuti lanac DNA (engl. *reverse strand*). Naposljetku započinje sekvenciranje obrnutog lanca DNA prilikom čega dolazi do ugradnje fluorescentno obilježenih oligonukleotida i detekcije signala u pojedinom klasteru.

### 3.4. BIOINFORMATIČKA OBRADA PODATAKA

Podaci dobiveni sekvenciranjem obrađeni su pomoću bioinformatičkih alata. Za obradu podataka korišteni su programi: SureCall (verzija 4.2.1.) (Agilent Technologies, Santa Clara, Kalifornija, SAD) i VarSome Clinical (verzija 11.9, pristupljeno siječanj 2023.).

Prvi korak obrade podataka je demultipleksiranje (engl. *demultiplexing*) i odvija se na samom uređaju. Prilikom procesa demultipleksiranja dolazi do razdvajanja podataka o pojedinim uzorcima u zasebne datoteke.

Prije daljnje obrade izvršeno je uklanjanje adapterskih sekvenci koje su dodane prilikom pripreme knjižnice. Proces detekcije i uklanjanja adapterskih sekvenci naziva se podrezivanje (engl. *trimming*).

Proces poravnanja s referentnim genomom proveden je s pomoću BWA-MEM algoritama (Li i Durbin, 2009.) Za potrebe ovog istraživanja provedeno je poravnanje s referentnim ljudskim genomom GRCh37 (hg19, veljača 2009.). Kako bi se osiguralo pouzdano određivanje frekvencije alela u uzorku izvršeno je uklanjanje duplikata i isključivanje preklapanja.

Proces pozivanja varijanti (engl. *variant calling*) proveden je uz pomoć *in-house* algoritma za detekciju polimorfizama jednog nukleotida te malih delecija i insercija. Algoritam omogućuje detekciju statistički značajnih varijanti na temelju kvalitete očitavanja baze, minimalne frekvencije alela u uzorku i minimalnog broja očitavanja s promjenom. U Tablici 8. navedeni su pragovi korišteni tijekom procesa pozivanja varijanti.

Tablica 8. Pragovi korišteni tijekom procesa pozivanja varijanti

Donja granica kvalitete očitavanja baze	30
Minimalna frekvencija alela u uzorku	5%
Minimalni broj očitavanja s promjenom	10

Prilikom procesa kategorizacije varijanti izabrane su samo varijante sa sljedećim učincima:

- Pomak okvira čitanja (engl. *frameshift*)
- Terminacija translacije u RNA (engl. *stop-gained*)
- Nesinonimne promjene baza koje rezultiraju promjenom slijeda aminokiselina
- Mutacije u dijelovima gena koji utječu na ekspresiju

Filtracija varijanti prema zastupljenosti u općoj populaciji izvodi se uz pomoć populacijske baze podataka 1000Genomes. Uklanjaju se sve varijante čija je učestalost u populaciji veća od 1%, time se smatraju polimorfizmima. Klasifikacija varijanti provodi se uz pomoć javno dostupnih baza podataka, znanstvenih publikacija i predikcijskih alata. Korišteni alati uključuju VarSome, COSMIC, gnomAD i ClinVar. Prilikom klasifikacije varijanti uzima se obzir: utjecaj varijante na funkciju gena, rezultira li promjena gena određenom bolesti i relevantnost fenotipa na određeno stanje prisutno kod bolesnika. U klinički nalaz ulaze samo varijante klasificirane kao patogene i vjerojatno patogene za koje je dokazano da su povezane s hematološkim bolestima.

### 3.4.1. SureCall

Za obradu podataka korišten je bioinformatički program SureCall (verzija 4.2.1.) (Agilent Technologies, Santa Clara, Kalifornija, SAD). SureCall je programsko rješenje koje omogućuje cjelokupnu obradu podataka dobivenih tehnologijom sekvenciranja sljedeće generacije. SureCall integrira više koraka analize sirovih podataka, omogućuje uklanjanje adaptera, poravnanje s referentnim genomom, pozivanje varijanti, filtraciju varijanti te ocjenu i kategorizaciju varijanti. Za poravnanje detektiranih sekvenci s referentnim genomom SureCall koristi BWA-MEM algoritam (Li i sur., 2009.). Nakon poravnanja s referentnim genomom SureCall omogućuje uklanjanje duplikata i preklapanja. Pozivanje varijanti izvodi se uz pomoć Agilentova SNP-PET SNP algoritma (<https://www.agilent.com/>). Za svaku detektiranu varijantu provjerava se njena lokalizacija, promjena u aminokiselinskom slijedu te utjecaj na funkciju proteina. SureCall provodi kategorizaciju varijanti uz pomoć sljedećih baza podataka: dbSNP, 1000 Genomes project, ClinVar, OMIM. Naposljetku, SureCall omogućuje kreiranje kliničkog izvještaja koji uključuje samo varijante klasificirane kao patogene i vjerojatno patogene.

### 3.4.2. VarSome Clinical

VarSome je mrežno dostupan pretraživač humanih genskih promjena. VarSome sakuplja podatke o traženoj varijanti i omogućuje korisnicima jednostavan pregled svih informacija o traženoj varijanti. Podatci se sakupljaju iz 30-ak baza podataka i uključuju informacije o učestalosti varijante u populaciji, utjecaju varijante na funkciju proteina i rezultate kliničkih istraživanja povezanih s navedenom varijantom (Kopanoss i sur., 2019.).

VarSome Clinical je sveobuhvatan alat koja omogućuje detekciju, kategorizaciju i interpretaciju varijanti detektiranih tehnologijom sekvenciranja sljedeće generacije (<https://varsome.com/>). VarSome Clinical sadrži *in silico* alate predviđanja patogenosti nasljednih i somatskih genskih varijanti. Predviđanje patogenosti provodi se uz pomoć računalnih algoritama koji prikupljaju podatke iz 140 baza podataka i znanstvenih publikacija. VarSome Clinical implementira pravila za klasifikaciju varijanti i predviđanje patogenosti iz ACMG (Richards i sur., 2015.) i AMP (Li i sur., 2017.) preporuka za klasifikaciju varijanti.

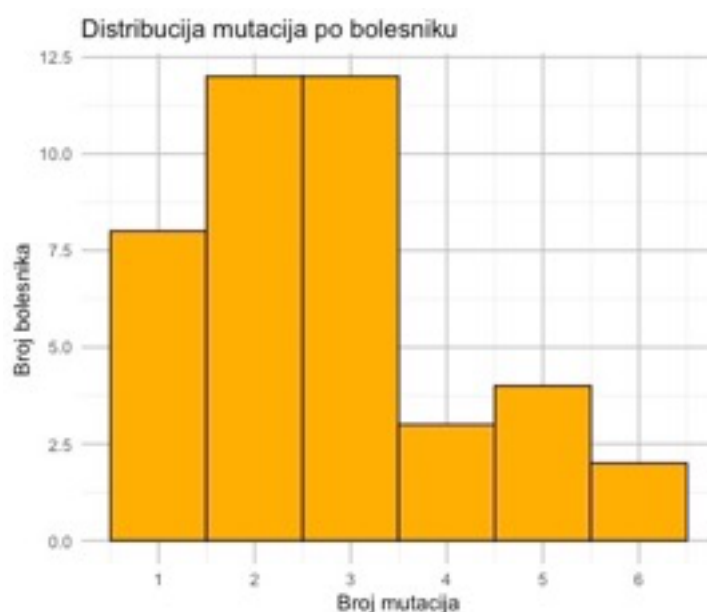
### 3.5. OBRADA PODATAKA

Prilikom obrade podataka korišten je program Microsoft Excel za MacOS verzija 16.76 te statistički programi R (R Core Team, 2021.) i RStudio (Rstudio Team, 2020.).

## 4. REZULTATI I RASPRAVA

### 4.1. Vrsta, raspodjela i učestalost patogenih mutacija povezanih s akutnom mijeloičnom leukemijom

Ciljanim sekvenciranjem genoma tehnologijom NGS-a utvrđeno je postojanje patogenih mutacija kod ukupno 41 ispitanika. Medijan broja pronađenih patogenih mutacija u bolesnicima je bio 3 (IQR 2-3). Raspodjela broja patogenih mutacija po bolesniku prikazana je na Slici 4. Ukupno je utvrđeno postojanje mutacija u 33 različita gena. Učestalost mutacija u pojedinim genima prikazana je u Tablici 9.



Slika 4. Raspodjela broja mutacija po bolesniku

Tablica 9. Učestalost mutacija u pojedinim genima povezanim s AML-om

Gen	Broj mutacija (N=118)
ASXL1	12
DNMT3A	11
NPM1	9
TET2	7
NRAS	7
TP53	6
SRSF2	5
JAK2	5
RUNX1	5
WT1	5
PTPN11	4
IDH2	4
FLT3	4
CBL	3
SF3B1	3
SETBP1	3
NOTCH1	3
IDH1	3
STAG2	2
BCOR	2
PDGFRA	2
PHF6	2
KRAS	2
ZRSR2	1
U2AF1	1
KIT	1
KDM6A	1
CEBPA	1
SMC3	1
ABL1	1
IKZF1	1
EZH2	1

Utjecaj varijante na funkciju proteina određen je uz pomoć VEP-a (engl. *Variant Effect Predictor*) (McLaren W i sur, 2016.). Od ukupno utvrđenih 118 varijanti, 79 varijanti su varijante pogrešnog smisla (engl. *Missense mutations*), 23 su varijante pomaka okvira čitanja (engl. *Frameshift mutations*), 3 su insercije unutar okvira čitanja (engl. *Inframe insertion mutations*) i 13 varijanti su varijante koje uzrokuju dobiveni stop kodon (engl. *Stop gained mutations*). Tablica 10. prikazuje vrste detektiranih varijanti.

Tablica 10. Vrste detektiranih varijanti

Vrsta detektiranih varijanti	N=118
Varijante pogrešnog smisla	79
Varijante pomaka okvira čitanja	23
Dobiveni stop kodon	13
Insercije unutar okvira čitanja	3

#### 4.2. Usporedba sekundarne analize podataka

U procesima sekundarne analize dobivenih podataka SureCall-om dokazano je postojanje 118 mutacija, dok je sekundarnom analizom podataka uz pomoć VarSome Clinical-a dokazano postojanje 115 mutacija. Usporedba detektiranih mutacija je prikazana u Tablici 11.

Tablica 11. Usporedba detekcije varijanti između SureCall-a i VarSome Clinical-a

	SureCall	VarSome Clinical
Detektirana	118	115
Nedetektirana	0	3

Podudaranje u detektiranim varijantama između između SureCall-a i VarSome Clinical-a je 97,5%.

#### 4.3. Usporedba tercijarne analize podataka

Nadalje *in house* tercijarnom analizom detektirano je 114 patogenih varijanti, dok je automatskom klasifikacijom uz pomoć VarSome Clinical-a detektirano 96 patogenih mutacija. Raspodjela klasifikacije mutacija prikazana je u Tablici 12.



Tablica 12. Usporedba *in house* tercijarne analize i klasifikacije VarSome Clinical-a

	<i>in house</i> tercijarna analiza	VarSome Clinical
Patogena	114	96
Nepatogena	0	18

Podudaranje u predviđanju patogenosti između *in house* tercijarne analize i VarSome Clinical-a je 84,2%.

#### 4.4. Rasprava

Sekvenciranje sljedeće generacije u okviru dijagnostike akutne mijeloične leukemije postaje sve važniji alat za postavljanje dijagnoze. Bioinformatički alati olakšavaju proces filtriranja varijanti i omogućuju prepoznavanje patogenosti varijanti. Nužno je da navedeni alati na ispravan način detektiraju i klasificiraju varijante kako bi se postavila ispravna dijagnoza. U ovom istraživanju provedena je usporedba bioinformatičke analize u dvije razine; u prvom koraku uspoređena je sekundarna analiza podataka SureCall-a i VarSome Clinical-a. Nadalje uspoređena je *in house* tercijarna analiza i automatizirana klasifikacija varijanti korištenjem VarSome Clinical algoritma.

Usporedbom sposobnosti programa u detekciji mutacija dokazano je podudaranje SureCall-a i VarSome Clinical-a koje iznosi iznosi 97,5%. SureCall je detektirao veći broj mutacija (118) u odnosu na VarSome Clinical (115). Neslaganja u detekciji mutacija između SureCall-a i VarSome Clinical-a odnose se na 3 mutacije u FLT3 genu. U dva slučaja radi se o NM\_004119.3:c.2503G>T, p.Asp835Tyr mutaciji, dok se u jednom slučaju radi o NM\_004119.3:c.2503G>C, p.Asp835His. Prilikom sekundarne analize sirovih NGS podataka SureCall i VarSome Clinical koriste iste algoritme za poravnanje s referentnim genomom, a razlikuju se u algoritmima za pozivanje varijanti. U procesu pozivanja varijanti SureCall koristi SNPPET SNP, dok VarSome Clinical koristi VarDict algoritam. Iako su korišteni različiti algoritmi za pozivanje varijanti, nemogućnost detekcije NM\_004119.3:c.2503G>T, p.Asp835Tyr i NM\_004119.3:c.2503G>C, p.Asp835His varijanti proizlazi iz drugačijih pragova detekcije postavljenih u VarSome Clinical u odnosu na SureCall. Ovo otkriće

ukazuje na potrebu za standardizacijom pragova detekcije u korištenim alatima i boljim usklađivanjem algoritama koji se koriste u sekundarnoj analizi podataka.

Usporedbom klasifikacije varijanti provedene *in house* tercijarnom analizom i VarSome Clinical-om utvrđeno je slaganje od 84,2%. Prilikom postupka *in house* tercijarne analize varijanti koriste se VarSome, COSMIC, gnomAD i ClinVar. Na temelju podataka iz navedenih alata uz kombinaciju s kliničkim dokazima proizašlih iz istraživanja donosi se krajnja odluka o klasifikaciji pojedine varijante.

S druge strane klasifikacija somatskih varijanti od strane VarSome Clinical-a temelji se na preporukama AMP-a iz 2017. (Li i sur., 2017.). VarSome-ov algoritam automatski svrstava somatske varijante u jedan od četiri razreda kliničke značajnosti. AMP preporuke (Li i sur., 2017.) nemaju čvrsto i jasno definirana pravila koja moraju biti zadovoljena kako bi se varijanta klasificirala u određeni razred kliničke značajnosti. Stoga VarSome koristi vlastiti algoritam prilikom klasifikacije varijanti VarSome Clinical. Neslaganje između *in house* provedene tercijarne analize i automatizirane klasifikacije VarSome Clinical-a primarno proizlazi iz različitog pristupa obradi informacija. Iako VarSome Clinical ima pristup istim bazama podataka i alatima koje su korištene u *in house* tercijarnoj analizi, uočen je različit pristup u korištenju tih baza podataka u predviđanju patogenosti varijanti. U Tablici 13. prikazane su nesukladnosti u klasifikaciji varijanti između *in house* tercijarne analize i VarSome Clinical-a. Jedna od glavnih uočenih razlika proizlazi iz podataka koji su dostupni u COSMIC-u. Na temelju *in house* tercijarne analize sve mutacije prisutne u COSMIC bazi podataka klasificiraju se kao patogene. Iako VarSome Clinical ima sposobnost pretraživanja COSMIC-a, podatci s COSMIC-a nisu integrirani u VarSome-ov algoritam predviđanja patogenosti varijanti.

Posljednja istraživanja naglašavaju važnost korištenja podataka COSMIC-a u predviđanju patogenosti somatskih varijanti. Froyen i sur. (2019) donijeli su preporuke za klasifikaciju somatskih varijanti u tumorima, gdje zaključuju kako je prisutnost varijante u COSMIC-u od glavnih kriterija koji ukazuje na patogeni učinak varijante. Česta pojava određena varijante u COSMIC-u ukazuje na njezinu vjerojatnu povezanost s razvojem raka. Važno je naglasiti kako VarSome Clinical prilikom klasifikacije koristi druge baze podataka koje sadrže varijante prisutne u hematološkim neoplazmama. Ukoliko je varijanta prisutna s visokom frekvencijom u jednoj od korištenih baza podataka klasificira se kao klinički značajna. Na temelju ovih otkrića postavlja se zaključak o važnosti usklađivanja i validiranja

baza podataka korištenih u predviđanju patogenosti.

Iako brojni laboratoriji provode klasifikaciju i predviđanje patogenosti somatskih varijanti, trenutno ne postoje međunarodno usvojene smjernice koje na jasan i strukturiran način objašnjavaju proces predviđanja patogenosti somatskih varijanti detektiranih u tumorima. Programi za automatiziranu klasifikaciju poput VarSome Clinical-a značajno su napredovali u predikciji patogenosti nasljednih varijanti. Međutim i dalje postoji potreba za napretkom u području predviđanja patogenosti somatskih varijanti. Jedan od razloga za značajan napredak u klasifikaciji nasljednih varijanti je pretvorba ACMG preporuka (Richards i sur., 2015.) u Bayseov okvir. Korištenjem Bayseovog okvira računalni algoritmi mogu bolje integrirati i analizirati složene podatke, što naposljetku dovodi do boljih predikcija patogenosti (Tavtigian i sur., 2018.). Ovo istraživanje dokazuje potrebu za uspostavljanjem harmoniziranih smjernica za klasifikaciju klinički relevantnih somatskih varijanti detektiranih hematoonkološkim neoplazmama.

Tablica 13. Neslaganja u klasifikaciji varijanti između *in house* tercijarne analize i VarSome Clinical

Gen	HGVS nazivlje	Promjena u proteinu	<i>in house</i> tercijarna analiza	VarSome Clinical	Objašnjenje
TET2	NM_001127208.3:c.1088C>T	p.Pro363Leu	Patogena	Benigna	Varijanta je prisutna u COSMIC bazi podataka te je dovedena u vezu s hematološkim bolestima
TET2	NM_001127208.3:c.100C>T	p.Leu34Phe	Patogena	Benigna	Varijanta je prisutna u COSMIC bazi podataka te je dovedena u vezu s hematološkim bolestima
PDGFRA	NM_006206.6:c.236G>A	p.Gly79Asp	Patogena	Benigna	Varijanta je prisutna u COSMIC bazi podataka te je dovedena u vezu s hematološkim bolestima
PDGFRA	NM_006206.6:c.236G>A	p.Gly79Asp	Patogena	Benigna	Varijanta je prisutna u

					COSMIC bazi podataka te je dovedena u vezu s hematoonkološkim bolestima
IDH1	NM_005896.4:c.548A>G	p.Tyr183Cys	Patogena	Benigna	Varijanta je prisutna u COSMIC bazi podataka
ASXL1	NM_015338.6:c.548A>G	p.Gly1397Ser	Patogena	Benigna	Varijanta je prisutna u COSMIC bazi podataka te je dovedena u vezu s hematoonkološkim bolestima
JAK2	NM_004972.4:c.3188G>A	p.Arg1063His	Patogena	Benigna	Varijanta je prisutna u COSMIC bazi podataka te je dovedena u vezu s hematoonkološkim bolestima
ZRSR2	NM_005089.4:c.1338_1343dup	p.Ser447_Arg448dup	Patogena	Benigna	Varijanta je prisutna u COSMIC bazi podataka te je dovedena u vezu s hematoonkološkim bolestima
RUNX1	NM_001754.5:c.167T>C	p.Leu56Ser	Patogena	Benigna	Varijanta je prisutna u

					COSMIC bazi podataka te je dovedena u vezu s hematoonkološkim bolestima
ASXL1	NM_015338.6:c.3306G>T	p.Glu1102Asp	Patogena	Benigna	Varijanta je prisutna u COSMIC bazi podataka te je dovedena u vezu s hematoonkološkim bolestima
NOTCH1	NM_017617.5:c.4168C>A	p.Pro1390Thr	Patogena	Benigna	Varijanta je prisutna u COSMIC bazi podataka
TET2	NM_001127208.3:c.2242C>T	p.Gly355Asp	Patogena	Benigna	Varijanta je prisutna u COSMIC bazi podataka te je dovedena u vezu s hematoonkološkim bolestima
JAK2	NM_004972.4:c.3188G>A	p.Arg1063His	Patogena	Benigna	Varijanta je prisutna u COSMIC bazi podataka te je dovedena u vezu s hematoonkološkim bolestima
RUNX1	NM_001754.5:c.167T>C	p.Leu56Ser	Patogena	Benigna	Varijanta je prisutna u

					COSMIC bazi podataka te je dovedena u vezu s hematoonkološkim bolestima
SETBP1	NM_015559.3:c.161G>A	p.Arg54His	Patogena	Benigna	Varijanta je prisutna u COSMIC bazi podataka te je dovedena u vezu s hematoonkološkim bolestima

## 5. ZAKLJUČCI

I. Podudaranje SureCall-a i VarSome Clinical-a u postupcima sekundarne analize iznosi 97,5%.

II. Podudaranje *in house* tercijarne analize i VarSome Clinical-a iznosi 84,2%.

III. Nužno je donijeti harmonizirane smjernice za klasifikaciju varijanti kako bi se uskladilo izvještavanje somatskih varijanti u hematološkim neoplazmama.



## 6. POPIS KRATICA, OZNAKA I SIMBOLA

ACMG – Američko društvo za medicinsku genetiku i genomiku (engl. American College of Medical Genetics and Genomics)

ALL - akutna limfoblastična leukemija

AML - akutna mijeloična leukemija

AML-MR – akutna mijeloična leukemija povezana s mijelodisplazijom

AML-NOS – akutna mijeloična leukemija koja nije drugačije klasificirana (engl. *AML- not otherwise specified*)

AMP - Udruga za molekularnu patologiju (engl. *Association for Molecular Pathology*)

BAM – (engl. *Binary Alignment Map*)

BCL – (engl. *binary base call*)

BCR-ABL – engl. break point cluster region-abelson

c-KIT - tirozin-protein kinaza Kit

CADD - (engl. Combined Annotation Dependent Depletion),

CBF – srednji vezujući faktor (engl. *core binding factor*)

CD – klaster diferencijacije (engl. *Cluster of differentiation*)

CEBPA – (engl. *CCAAT/enhancer-binding protein alpha*)

COSMIC - (engl. *Catalogue Of Somatic Mutations In Cancer*)

dbSNP - (engl. *Single Nucleotide Polymorphism Database*)

DNA – deoksiribonukleinska kiselina

EDTA - Etilendiamintetraoctena kiselina

ELN - Europska leukemijska mreža

FISH - fluorescentnu in situ hibridizacija

FLT3 - FMS-slična-tirozin- kinaza (engl. *FMS-like tyrosine kinase-3*)

HLA - humani leukocitni antigen

ICC- međunarodna konsenzusna klasifikacija (engl. *International consensus classification, ICC*)

ITD – unutarnja tandemska duplikacija (engl. *internal tandem duplication*)

MDS – mijelodisplastični sindrom

MDS/MPN - mijelodisplastični sindrom/mijeloproliferativna neoplazma

MPO - mijeloperoksidaza

MRD - minimalna ostatna bolest

NGS - sekvenciranje sljedeće generacije (engl. *Next generation sequencing*)

NPM1 - nukleofizmin 1

OMIM - (engl. *Online Mendelian Inheritance in Man*)

PCR – reakcija umnažanja polimerazom (engl. *Polymerase chain reaction*)

SIFT - (engl. *Sorting Intolerant From Tolerant*)

SNP - Single Nucleotide Polymorphism

SZO - svjetska zdravstvena organizacija (engl. *World health organization, WHO*)

TP53 - tumorski protein 53

VAF - udio varijantnog alela (engl. *variant allele frequency*)

VEP - (engl. *Variant Effect Predictor*).

RNA - ribonukleinska kiselina

## 7. LITERATURA

Alaggio, R., Amador, C., Anagnostopoulos, I., Attygalle, A.D., Araujo, I.B. de O., Berti, E., Bhagat, G., Borges, A.M., Boyer, D., Calaminici, M., Chadburn, A., Chan, J.K.C., Cheuk, W., Chng, W.-J., Choi, J.K., Chuang, S.-S., Coupland, S.E., Czader, M., Dave, S.S. and de Jong, D. The 5th edition of the World Health Organization Classification of Haematolymphoid Tumours: Lymphoid Neoplasms. *Leukemia*, 2022, 36, 1720–1748.

An introduction to Next-generation Sequencing Technology. Preuzeto s: [https://www.illumina.com/content/dam/illumina-marketing/documents/products/illumina\\_sequencing\\_introduction.pdf](https://www.illumina.com/content/dam/illumina-marketing/documents/products/illumina_sequencing_introduction.pdf). (pristupljeno: 6.5.2024.)

Arber, D.A., Orazi, A., Hasserjian, R., Thiele, J., Borowitz, M.J., Le Beau, M.M., Bloomfield, C.D., Cazzola, M. and Vardiman, J.W. The 2016 Revision to the World Health Organization Classification of Myeloid Neoplasms and Acute Leukemia. *Blood*, 2016, 127, 2391–405.

Boscaro, E., Urbino, I., Catania, F. M., Arrigo, G., Secreto, C., Olivi, M., D'Ardia, S., Frairia, C., Giai, V., Freilone, R., Ferrero, D., Audisio, E., & Cerrano, M. Modern Risk Stratification of Acute Myeloid Leukemia in 2023: Integrating Established and Emerging Prognostic Factors. *Cancers*, 2023, 15, 3512.

Chen Y, Li J, Xu L, Găman MA, Zou Z. The genesis and evolution of acute myeloid leukemia stem cells in the microenvironment: From biology to therapeutic targeting. *Cell Death Discov*, 2022, 8, 397.

Döhner H, Wei AH, Appelbaum FR, et al. Diagnosis and management of AML in adults: 2022 recommendations from an international expert panel on behalf of the ELN. *Blood*. 2022, 140, 1345-77.

Duffield, A.S., Mullighan, C.G. and Borowitz, M.J. International Consensus Classification of acute lymphoblastic leukemia/lymphoma. *Virchows Archiv*. 2022, 428, 11-26.

Canaani J, Labopin M, Itälä-Remes M, et al. Prognostic significance of recurring chromosomal abnormalities in transplanted patients with acute myeloid leukemia, *Leukemia*, 2019, 33, 1944-1952.

Froyen, G., Le Mercier, M., Lierman, E., Vandepoele, K., Nollet, F., Boone, E., Van der Meulen, J., Jacobs, K., Lambin, S., Vander Borgh, S., Van Valckenborgh, E., Antoniou, A., Hébrant, A. Standardization of Somatic Variant Classifications in Solid and Haematological Tumours by a Two-Level Approach of Biological and Clinical Classes: An Initiative of the Belgian ComPerMed Expert Panel. *Cancers*, 2019, 11, 2030.

Grada, A. and Weinbrecht, K. Next-generation sequencing: Methodology and application, *Journal of Investigative Dermatology*, 2013, 133, 1–4.

Haferlach T. Advancing leukemia diagnostics: Role of Next Generation Sequencing (NGS) in acute myeloid leukemia, *Hematol Rep*, 2020, 12, 8957.

Huber, S., Baer, C., Hutter, S., Dicker, F., Manja Meggendorfer, Pohlkamp, C., Kern, W., Torsten Haferlach, Haferlach, C. and Hoermann, G. AML classification in the year 2023: How to avoid a Babylonian confusion of languages. *Leukemia*, 2023, 37, 1413–1420.

Juliusson G, Antunovic P, Derolf A, et al. Age and acute myeloid leukemia: real world data on decision to treat and outcomes from the Swedish Acute Leukemia Registry. *Blood*, 2009, 113, 4179-4187.

Keinanen M, Knuutila S, Bloomfield CD, et al. The proportion of mitoses in different cell lineages changes during short-term culture of normal human bone marrow, *Blood*, 1986, 67, 1240–1243.

Kopanos, C., Tsiolkas, V., Kouris, A., Chapple, C.E., Albarca Aguilera, M., Meyer, R. i Massouras, A. VarSome: the human genomic variant search engine. *Bioinformatics*, 2019, 35, 1978-1980.

Kumar, C.C. Genetic Abnormalities and Challenges in the Treatment of Acute Myeloid Leukemia, *Genes & Cancer*, 2011, 2, 95–107.

Labar B, Hauptman E, Aurer I, Batinić D, Boban A, Brkljačić-kerhin V, Davidović-Mrsić, S, Golubić-Ćepulić B, Huić D, Malenica B i ostali. Hematologija, Zagreb: Školska knjiga, 2007, str. 293.

Li H, Durbin R. Fast and accurate short read alignment with Burrows-Wheeler transform, *Bioinformatics*, 2009, 25, 1754-1760.

Li MM, Datto M, Duncavage EJ, et al. Standards and Guidelines for the Interpretation and Reporting of Sequence Variants in Cancer: A Joint Consensus Recommendation of the Association for Molecular Pathology, American Society of Clinical Oncology, and College of American Pathologists, *The Journal of molecular diagnostics*, 2017, 19, 4-23.

Mckenzie, S.B., Landis-Piwowar, K. and Joanne Lynne Williams, Clinical laboratory hematology. Pearson, 2020, str. 558.

McLaren W, Gil L, Hunt SE, et al. The Ensembl Variant Effect Predictor, *Genome Biol*, 2016, 17, 122.

Meyer, C., Burmeister, T., Gröger, D., Tsauro, G., Fehina, L., Renneville, A., Sutton, R., Venn, N.C., Emerenciano, M., Pombo-de-Oliveira, M.S., Barbieri Blunck, C., Almeida Lopes, B., Zuna, J., Trka, J., Ballerini, P., Lapillonne, H., De Braekeleer, M., Cazzaniga, G., Corral Abascal, L. and van der Velden, V.H.J. The MLL recombinome of acute leukemias in 2017. *Leukemia*, 2017, 32, 273–284.

NGS Data Analysis for Illumina Platform—Overview and Workflow, 2018, <https://www.thermofisher.com/>, pristupljeno 6.5.2024.

Nguyen B, Williams AB, Young DJ, et al. FLT3 activating mutations display differential sensitivity to multiple tyrosine kinase inhibitors, *Oncotarget*, 2017, 8, 10931-10944.

Oliver GR, Hart SN, Klee EW. Bioinformatics for clinical next generation sequencing, *Clinical chemistry*, 2015, 61, 124-135.

Papaemmanuil E, Gerstung M, Bullinger L, et al. Genomic classification and prognosis in acute myeloid leukemia. *N Engl J Med*, 2016, 374, 2209-2221.

Park, D.J., Kwon, A., Cho, B.-S., Kim, H.-J., Hwang, K.-A., Kim, M. and Kim, Y. Characteristics of DNMT3A mutations in acute myeloid leukemia. *Blood research*, 2020, 55, 17–26.

Pereira, F.G., Metze, K., Costa, S., Lima, P. and I. Lorand-Metze. Phenotypic quantitative features of patients with acute myeloid leukemia, *PubMed*, 2006, 53, 155–60.

QIAamp DNA Mini and Blood Mini Handbook QIAGEN, Aarhus, Denmark

Qubit 3.0, Thermo Fisher Scientific, Waltham, Massachusetts, SAD

Qin, D. Molecular testing for acute myeloid leukemia, *Cancer Biology and Medicine*, 2021, 19, 4-13.

Release Notes SureCall 4.2.2. Preuzeto s:

<https://www.agilent.com/cs/library/software/public/Release%20Notes%20SureCall%204.2.pdf> (pristupljeno: 15.5.2024.)

Richards S, Aziz N, Bale S, et al. 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, *Genetics in medicine : official journal of the American College of Medical Genetics*, 2015, 17, 405-424.

Sasaki K, Ravandi F, Kadia TM, et al. De novo acute myeloid leukemia: a population-based study of outcome in the United States based on the Surveillance, Epidemiology, and End Results (SEER) database, 1980 to 2017. *Cancer*, 2021, 127, 2049-2061.

SEER data base: Cancer Stat Facts: Leukemia - Acute Myeloid Leukemia (AML),

<https://seer.cancer.gov/statfacts/html/amyl.html>, pristupljeno: 12.4.2024.

Shallis RM, Wang R, Davidoff A, Ma X, Zeidan AM. Epidemiology of acute myeloid leukemia: Recent progress and enduring challenges. *Blood Rev*, 2019, 36, 70-87.

SureSelectXT Target Enrichment System for the Illumina Platform, Agilent Technologies, Santa Clara, Kalifornija, SAD)

Tavtigian, S. V., Greenblatt, M. S., Harrison, S. M., Nussbaum, R. L., Prabhu, S. A., Boucher, K. M., Biesecker, L. G., & ClinGen Sequence Variant Interpretation Working Group (ClinGen SVI). Modeling the ACMG/AMP variant classification guidelines as a Bayesian classification framework. *Genetics in medicine : official journal of the American College of Medical Genetics*, 2018, 20, 1054–1060.

Tembhare, P. Monitoring Measurable/Minimal Residual Disease in Acute Myeloid Leukemia: Multiparametric Flow Cytometry-Based Approach, *Indian Journal of Medical and Paediatric Oncology*, 2023, 44, 554-565.

Vakiti A, Mewawalla P. Acute Myeloid Leukemia. *StatPearls Publishing*, 2024

VarSome Clinical. Preuzeto s: <https://4384097.fs1.hubspotusercontent-na1.net/hubfs/4384097/Brochures/VarSome%20Brochure%20May%202023.pdf> (pristupljeno 15. 5 2024.)

Zhao X, Liu HQ, Wang LN, Yang L, Liu XL. Current and emerging molecular and epigenetic disease entities in acute myeloid leukemia and a critical assessment of their therapeutic modalities, *Semin Cancer Biol*, 2020, 83, 121-135.

## 8. SAŽETAK/SUMMARY

### SAŽETAK

Akutna mijeloična leukemija genski je heterogena bolest kod koje se je u 97% bolesnika dokazana barem jedna genska promjena. Posljednje smjernice za postavljanje dijagnoze i stratifikaciju rizika bolesnika oboljelih od AML naglašavaju važnost detekcije genskih i citogenetskih promjena. Unatoč znatnom napretku dijagnostike i liječenja AML-a, petogodišnje preživljenje ostaje vrlo nisko. Kako bi se poboljšao ishod liječenja bolesnici se na temelju detektiranih genskih i citogenetskih promjena svrstavaju u rizične skupine na temelju kojih se procjenjuje vjerojatnost postizanja remisije i preživljenja. Jedan od procesa određivanja mutacija u genima vezanim uz AML je sekvenciranje sljedeće generacije. Sekvenciranjem sljedeće generacije moguće je istovremeno detektirati velik broj genskih promjena u ispitivanim regijama genoma. Bioinformatički alati omogućuju medicinskim djelatnicima bržu klasifikaciju i interpretaciju detektiranih varijanti. Cilj ovog rada je evaluacija *in house* postupka analize podataka dobivenih sekvenciranjem sljedeće generacije za mijeloidni panel, utvrđivanje stupnja povezanosti dva bioinformatička alata za analizu podataka dobivenih tehnologijom NGS-a te utvrditi postoje li odstupanja između njih. U prvom koraku ispitana je sposobnost detekcije mutacija, dok je u drugom koraku uspoređena sposobnost klasifikacije varijanti VarSome Clinical algoritma u odnosu na *in house* tercijarnu analizu. Podudaranje SureCall-a i VarSome Clinical-a u postupcima sekundarne analize iznosi 97,5%. Nadalje dokazano je da podudaranje *in house* tercijarne analize i VarSome Clinical-a iznosi 84,2%



## SUMMARY

Acute myeloid leukemia (AML) is a genetically heterogeneous disease, with genetic changes being detected in 97% of all patients. The latest guidelines for the diagnosis and risk stratification of patients with AML emphasize the importance of detecting genetic and cytogenetic changes. Despite significant advances in the diagnosis and treatment of AML, five-year survival remains very low. In order to improve the outcome of treatment, patients are classified into risk groups that are based on detected genetic and cytogenetic changes. One of the methods for detecting mutations in AML-related genes is next-generation sequencing (NGS). Using NGS is possible to simultaneously detect a large number of gene changes in the examined regions of the genome. Bioinformatic tools enable medical professionals to classify and interpret detected variants more efficiently. This thesis aimed to evaluate *in-house* procedure for analyzing data obtained by next-generation sequencing for the myeloid panel, determine the degree of connection between two bioinformatics tools for analyzing data obtained by NGS technology, and determine whether there are discrepancies between them. In this research, a comparison of two bioinformatics tools was carried out at two levels. Firstly, the ability to detect mutations was tested, while in the second step, VarSome Clinical algorithm ability to classify mutations was compared in relation to the *in-house* tertiary analysis. The agreement between SureCall and VarSome Clinical in secondary analysis procedures is 97.5%. It was further proven that the agreement between the *in-house* tertiary analysis and VarSome Clinical is 84.2%

**9. TEMELJNA DOKUMENTACIJSKA KARTICA/ BASIC DOCUMENTATION  
CARD**

## Temeljna dokumentacijska kartica

Sveučilište u Zagrebu  
Farmaceutsko-biokemijski fakultet  
Studij: Medicinska biokemija  
Zavod za medicinsku biokemiju i hematologiju  
Domagojeva 2, 10000 Zagreb, Hrvatska

Diplomski rad

### Učinkovitost bioinformatičkih alata u predviđanju patogenosti somatskih mutacija dokazanih tehnologijom sekvenciranja sljedeće generacije kod bolesnika s dijagnozom akutne mijeloične leukemije

Toni Božac

#### SAŽETAK

Akutna mijeloična leukemija genski je heterogena bolest kod koje se je u 97% bolesnika dokazana barem jedna genska promjena. Posljednje smjernice za postavljanje dijagnoze i stratifikaciju rizika bolesnika oboljelih od AML naglašavaju važnost detekcije genskih i citogenetskih promjena. Unatoč znatnom napretku dijagnostike i liječenja AML-a, petogodišnje preživljenje ostaje vrlo nisko. Kako bi se poboljšao ishod liječenja bolesnici se na temelju detektiranih genskih i citogenetskih promjena svrstavaju u rizične skupine na temelju kojih se procjenjuje vjerojatnost postizanja remisije i preživljenja. Jedan od procesa određivanja mutacija u genima vezanim uz AML je sekvenciranje sljedeće generacije. Sekvenciranjem sljedeće generacije moguće je istovremeno detektirati velik broj genskih promjena u ispitivanim regijama genoma. Bioinformatički alati omogućuju medicinskim djelatnicima bržu klasifikaciju i interpretaciju detektiranih varijanti. U ovom diplomskom radu uspoređena je učinkovitost bioinformatičkih alata u predviđanju patogenosti somatskih mutacija dokazanih tehnologijom sekvenciranja sljedeće generacije kod bolesnika s dijagnozom akutne mijeloične leukemije. Cilj ovog rada je evaluacija in house postupka analize podataka dobivenih sekvenciranjem sljedeće generacije za mijeloidni panel, utvrđivanje stupnja povezanosti dva bioinformatička alata za analizu podataka dobivenih tehnologijom NGS-a te utvrditi postoje li odstupanja između njih. U prvom koraku ispitana je sposobnost detekcije mutacija, dok je u drugom koraku uspoređena sposobnost klasifikacije varijanti VarSome Clinical algoritma u odnosu na tercijarnu analizu. Podudaranje SureCall-a i VarSome Clinical-a u postupcima sekundarne analize iznosi 97,5%. Nadalje dokazano je da podudaranje in house tercijarne analize i VarSome Clinical-a iznosi 84,2%

Rad je pohranjen u Središnjoj knjižnici Sveučilišta u Zagrebu Farmaceutsko-biokemijskog fakulteta.

Rad sadrži: 45 stranica, 4 grafičkih prikaza, 13 tablica i 42 literaturnih navoda. Izvornik je na hrvatskom jeziku.

Ključne riječi: Sekvenciranje sljedeće generacije, akutna mijeloična leukemija, predviđanje patogenosti, bioinformatika

Mentor: **Dr. sc. Dunja Rogić**, redoviti profesor Sveučilišta u Zagrebu Farmaceutsko-biokemijskog fakulteta.

Ocjenjivači: **Dr. sc. Dunja Rogić**, redoviti profesor Sveučilišta u Zagrebu Farmaceutsko-biokemijskog fakulteta.  
**Dr. sc. Ksenija Fumić**, redoviti profesor Sveučilišta u Zagrebu Farmaceutsko-biokemijskog fakulteta.  
**Dr. sc. Ivana Lapić**, docent Sveučilišta u Zagrebu Farmaceutsko-biokemijskog fakulteta.

Rad prihvaćen: srpanj 2024

## Basic documentation card

University of Zagreb  
Faculty of Pharmacy and Biochemistry  
Study: Medical Biochemistry  
Department of Medical Biochemistry and Haematology  
Domagojeva 2, 10000 Zagreb, Croatia

Diploma thesis

### The effectiveness of bioinformatics tools in predicting the pathogenicity of somatic mutations proven by next-generation sequencing technology in patients diagnosed with acute myeloid leukemia

Toni Božac

#### SUMMARY

Acute myeloid leukemia (AML) is a genetically heterogeneous disease, with genetic changes being detected in 97% of all patients. The latest guidelines for the diagnosis and risk stratification of patients with AML emphasize the importance of detecting genetic and cytogenetic changes. Despite significant advances in the diagnosis and treatment of AML, five-year survival remains very low. In order to improve the outcome of treatment, patients are classified into risk groups that are based on detected genetic and cytogenetic changes. One of the methods for detecting mutations in AML-related genes is next-generation sequencing (NGS). Using NGS is possible to simultaneously detect a large number of gene changes in the examined regions of the genome. Bioinformatic tools enable medical professionals to classify and interpret detected variants more efficiently. This thesis aimed to evaluate in-house procedure for analyzing data obtained by next-generation sequencing for the myeloid panel, determine the degree of connection between two bioinformatics tools for analyzing data obtained by NGS technology, and determine whether there are discrepancies between them. In this research, a comparison of two bioinformatics tools (SureCall and VarSome Clinical) was carried out at two levels. Firstly, the ability to detect mutations was tested, while in the second step, VarSome Clinical algorithm ability to classify mutations was compared in relation to the in-house tertiary analysis. The agreement between SureCall and VarSome Clinical in secondary analysis procedures is 97.5%. It was further proven that the agreement between the in-house tertiary analysis and VarSome Clinical is 84.2%

The thesis is deposited in the Central Library of the University of Zagreb Faculty of Pharmacy and Biochemistry.

Thesis includes: 45 pages, 4 figures, 13 tables and 42 references. Original is in Croatian language.

Keywords: Next generation sequencing, acute myeloid leukemia, prediction of pathogenicity, bioinformatics, NGS

Mentor: **Dunja Rogić, Ph.D** *Full Professor*, University of Zagreb Faculty of Pharmacy and Biochemistry

Reviewers: **Dunja Rogić, Ph.D** *Full Professor*, University of Zagreb Faculty of Pharmacy and Biochemistry  
**Ksenija Fumić, Ph.D.** *Full Professor*, University of Zagreb Faculty of Pharmacy and Biochemistry  
**Ivana Lapić, Ph.D.** *Assistant Professor*, University of Zagreb Faculty of Pharmacy and Biochemistry

The thesis was accepted: July 2024.