

Genetska
dijagnostika

ExomeXtra®: Detaljan pregled dijagnostičkih performansi



Stručni pregled



U ovom stručnom pregledu prikazujemo

- ✗ kako ExomeXtra® objedinjuje prednosti sekvenciranja egzoma, sekvenciranja genoma i array CGH analize
- ✗ da ExomeXtra® rješava do 25% više slučajeva od standardne analize egzoma
- ✗ gdje sekvenciranje cijelog genoma (WGS) doseže tehničke granice analitičke osjetljivosti
- ✗ da je ExomeXtra® osjetljiviji od WGS-a u klinički relevantnim regijama
- ✗ da objavljene tvrdnje o nadmoći WGS-a nad sekvenciranjem egzoma mogu dovesti u zabludu
- ✗ da WGS jednog pacijenta ima isti dijagnostički prinos kao Single ExomeXtra®
- ✗ i šta ovi nalazi znače za izbor dijagnostičke strategije prvog izbora

Exome



Dijagnostički izazov

Rijetke genetske bolesti nastaju zbog različitih genetskih varijanti, uključujući varijante sekvence i varijante broja kopija. Otkrivanje varijante koja uzrokuje bolest omogućava molekularnu dijagnozu, često završava dug dijagnostički put, podržava genetsko savjetovanje te, zavisno od bolesti, pruža informacije važne za prognozu i odluke o liječenju.

Cilj genetske dijagnostike je jasan: utvrditi uzročnu varijantu kod što većeg broja pacijenata. Savremene visokoprotodne metode molekularne genetike omogućavaju otkrivanje varijanti broja kopija i varijanti sekvence. Sekvenciranje cijelog egzoma (WES) usmjereno je na regije koje kodiraju proteine, dok sekvenciranje cijelog genoma proširuje analizu na čitav genom. Ipak, više podataka sekvenciranja ne znači automatski i više dijagnoza.

Dijagnostička uspješnost ne zavisi samo od obuhvata genoma. Klinički relevantne regije zahtijevaju dovoljnu i ravnomjernu dubinu sekvenciranja, različite vrste genetskih promjena moraju se pouzdano otkriti, a pronađene varijante protumačiti u odgovarajućem kliničkom i porodičnom kontekstu. Izazov je dobiti prave podatke na pravim mjestima i analizirati ih u pravom kontekstu. Podatke sekvenciranja treba usmjeriti na klinički relevantne regije, a podaci o genetskim srođnicima mogu znatno povećati dijagnostički prinos.

ExomeXtra® je razvijen na ovom principu. Njegov dizajn omogućava osjetljivo otkrivanje širokog spektra dijagnostički relevantnih vrsta varijanti.

Dio 1: ExomeXtra® - Više od egzoma, pametnije od genoma

CeGaT ExomeXtra® objedinjuje prednosti sekvenciranja cijelog egzoma, sekvenciranja cijelog genoma i array CGH analize u jednom dijagnostičkom testu. Proširuje standardnu dijagnostiku egzoma izvan kodirajućih regija i uključuje analizu CNV-a na nivou cijelog genoma. Time uklanja potrebu za dodatnom array CGH analizom koja se kod određenih fenotipova obično naručuje prije dijagnostike egzoma.

Istovremeno, ciljano obogaćivanje omogućava veliku dubinu sekvenciranja klinički relevantnih regija i osjetljivije otkrivanje varijanti nego WGS standardne dubine. Dodatne analize proširuju dijagnostički obuhvat, uključujući otkrivanje odabranih infekcija čiji simptomi mogu podsjećati na genetske poremećaje.

ExomeXtra® obuhvata kodirajuće sekvence iz više baza transkripata (MANE, Gencode, Ensembl, RefSeq Curated). Obuhvat je dodatno proširen na:

(1) sve ručno kurirane kodirajuće i nekodirajuće regije iz više od 20 CeGaT dijagnostičkih panela koji obuhvataju stotine nasljednih bolesti,

(2) sve patogene i vjerovatno patogene nekodirajuće varijante opisane u HGMD-u i ClinVaru ($\geq 53,500$),

(3) 5'UTR mjesta prekranja svih MANE Select transkripata,

(4) cijeli mitohondrijski genom uz visoku pokrivenost,

(5) RNA gene relevantne za bolesti,

(6) regije s farmakogenetički relevantnim varijantama u odabranim genima,

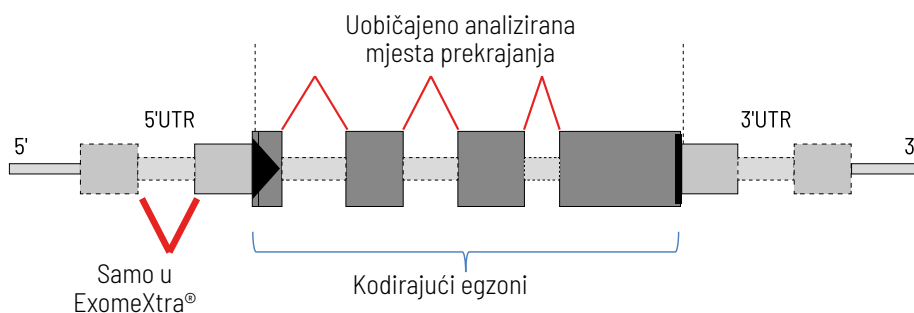
(7) probe za otkrivanje relevantnih prenatalnih infekcija i virusa koji potiču nastanak tumora, te

(8) mrežu ciljnih regija na svim hromosomima za otkrivanje CNV-a na nivou cijelog genoma.



Tabela 1: Poređenje performansi standardnog komercijalnog egzoma i testa CeGaT ExomeXtra®. Procenti označavaju udio dijagnostički relevantnih regija pokrivenih dubinom od najmanje 30x.

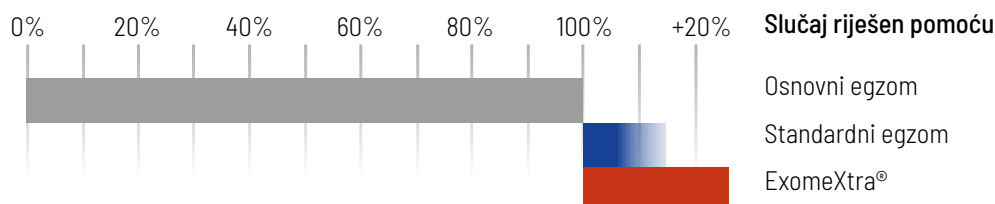
	Komercijalni egzom	CeGaT ExomeXtra®
Kodirajuće sekvence		
✗ Gencode 49	95.7%	97.3%
✗ MANE Select 1.5	96.9%	98.0%
✗ RefSeq Curated 250813	97.0%	98.2%
✗ Ensembl 115	96.3%	96.9%
5'UTR mjesta prekranja	nije obuhvaćeno	obuhvaćeno za otkrivanje varijanti koje mijenjaju prekranja
Nekodirajuće varijante		
✗ ClinVar 260703	97.7%	99.6%
✗ HGMD 2026 Q2	92.1%	99.3%
✗ Obje baze zajedno	93.5%	99.3%
✗ Duboko intronske / udaljene	54.3%	97.3%
Nekodirajući RNA geni	nije obuhvaćeno	obuhvaćeno 37 relevantnih gena
Mitohondrijski genom	relativno niska pokrivenost	visoka pokrivenost omogućava osjetljivu detekciju heteroplazmijskih varijanti
Mreža za CNV analizu cijelog genoma	nije uključeno	uključena uz rezoluciju od približno 50 kbp i dodatne ciljane cis-regulatorne regije odabranih gena
Probir na infekcije	nije uključeno	Humani papilomavirus (HPV), <i>Helicobacter pylori</i> , Epstein-Barr virus (EBV), citomegalovirus (CMV), virus hepatitisa B (HBV), Herpes simplex virus (HSV) 1 i 2, <i>Toxoplasma gondii</i> (toksoplazmoza), Varicella Zoster virus (VZV), parvovirus B19 (peta bolest), <i>Treponema pallidum</i> (sifilis), <i>Plasmodium</i> spp. (malaria), <i>Borrelia</i> spp. (lajmska bolest), <i>Brucella melitensis</i> (bruceloza), <i>Neisseria meningitidis</i> (meningitis), <i>Listeria monocytogenes</i> (listerioza)



Slika 1: Standardni dizajni egzoma ne obuhvataju 5'UTR mjesta prekranja. Varijante na tim mjestima ipak mogu biti patogene, npr. zbog zadržavanja introna ili preskakanja egzona koji sadrži početni kodon.

Dio 2: ExomeXtra® rješava 25% više slučajeva od osnovne analize egzoma

ExomeXtra® može riješiti 9-16% više slučajeva od standardnih visokokvalitetnih egzoma i do 26% više od osnovne analize egzoma.



Slika 2: Na osnovu stvarnih dijagnostičkih slučajeva procijenjeno je koji se mogu riješiti osnovnom analizom egzoma, visokokvalitetnim standardnim egzomom (definicije u tekstu) ili samo testom ExomeXtra®.

Prikupljanje podataka

Procijenjena su 2,592 dijagnostička slučaja riješena testom ExomeXtra® (verzija 6). Obuhvatali su 3,174 varijante klasificirane kao patogene (P) ili vjerovatno patogene (LP), odnosno varijante koje objašnjavaju slučaj. U 2,230 slučajeva uzrok je bila jedna varijanta, a u 362 više njih, uglavnom složena heterozigotnost ili kombinacije SNV-a i CNV-a.

Varijante su razvrstane u 24 klase prema vrsti, metodi detekcije ili učestalosti. Klase su svrstane u četiri glavne grupe: varijante sekvence, varijante broja kopija, varijante u dodatnim nekodirajućim regijama testa ExomeXtra® i varijante koje zahtijevaju naprednu bioinformatiku.

Za svaku klasu varijanti procijenjeno je može li se otkriti „osnovnom“ analizom egzoma, „standardnom visokokvalitetnom“ analizom ili samo testom ExomeXtra®.

Sve analize egzoma mogu otkriti varijante sekvence u kodirajućim regijama. Detekcija CNV-a iz egzomskih podataka nije jednostavna; u nalazima drugih laboratorija redovno vidimo slučajeve bez dijagnoze jer CNV nije otkriven. Posebno su zahtjevni CNV-i jednog egzoma (19% CNV-a u našoj kohorti). Stoga za standardni egzom, u odnosu na ExomeXtra®, pretpostavljamo osjetljivost detekcije CNV-a od 50% do 81%. Nekodirajuće regije ciljane testom ExomeXtra® nisu dostupne drugim analizama, a druge laboratorije rijetko koriste i neke složenije bioinformatičke metode.

Scenariji analize

Definirali smo tri vrste analize:

- ✗ „osnovna“ analiza egzoma: standardni komplet i osnovna bioinformatika za otkrivanje varijanti sekvence u kodirajućim egzomima,
- ✗ „standardna visokokvalitetna“ analiza: standardno obogaćivanje egzoma uz dodatnu detekciju CNV-a, te
- ✗ ExomeXtra®.

Analiza na nivou slučaja

Za objašnjenje genetskog slučaja ponekad je potrebno više od jedne varijante. Takav slučaj smatramo riješenim samo ako je otkrivena i varijanta koju je najteže otkriti. Naprimjer, bolest uzrokovana missense mutacijom i drugom varijantom duboko u intronu može objasniti samo metoda koja analizira duboke intronske regije, poput testa ExomeXtra®.

Dodatnih 20% varijanti otkrivenih testom ExomeXtra®

2,612	Varijante sekvence u egzonima koji kodiraju proteine (CDS) Sve missense, nonsense, frameshift, stop-loss i start-loss varijante, sinonimne varijante i varijante ključnih mjesta prekrajanja, u homozigotnom, heterozigotnom i mozaičnom stanju s visokom učestalošću (>20%).
349 +13.4%	Varijante broja kopija Sve varijante broja kopija koje zahvataju pojedinačne egzone, cijele gene i veće genomske regije koje standardni egzomi ne obuhvataju. ExomeXtra® je visoko osjetljiv i za CNV-e jednog egzona. Samo ExomeXtra® uključuje mrežu za CNV analizu cijelog genoma s rezolucijom sličnom array analizi. Dodatni array CGH nije potreban.
153 +5.9%	Nekodirajući geni i regije Varijante izvan kodirajućih regija (intronske i duboko intronske), u nekodirajućim RNA genima, nekodirajućim egzonima gena za proteine (UTR), mitohondrijskoj DNA i ekspanzijama ponavljanja koje standardni egzomi ne obuhvataju. Samo ExomeXtra® obuhvata sve poznate varijante koje uzrokuju bolesti iz baza ClinVar i HGMD, uključujući duboko intronske i one uzvodno od gena. Uključujemo i ✗ relevantne RNA gene (npr. RNU4-2), ✗ uzvodne netranslatirane regije (UTR) svih MANE Select transkripata kodirajućih gena, ✗ regije s ponavljanjima koje nedostaju u standardnim egzomima, ✗ cis-regulatorne regije koje stupaju u interakciju s promotorima određenih gena.
60 +2.3%	Potrebna je napredna bioinformatika Ovdje spadaju mozaične varijante sekvence i mozaični CNV-i niske učestalosti, uniparentalne disomije, složeni aleli i VNTR-i (varijabilni broj tandemskih ponavljanja). ExomeXtra® uključuje dobro osmišljen komplet za obogaćivanje, ali i naprednu bioinformatiku i pažljivo razvijene analitičke metode. Mozaicizam može biti vrlo relevantan, a mnogi bioinformatički postupci zanemaruju takve varijante. Posebno, analize cijelog genoma ne omogućavaju osjetljivo otkrivanje mozaicizma niske učestalosti. Uniparentalne disomije (UPD) mogu dovesti do efekata genomskog utiskivanja relevantnih za bolest, koji se ne vide direktno iz rezultata detekcije varijanti.
nije kvantificirano	Izvan granica ljudskog genoma ExomeXtra® uključuje i probir na relevantne infekcije, što omogućava dodatne dijagnoze, posebno u prenatalnim slučajevima. To nije bilo dio ove kohortne analize, pa kvantifikacija trenutno nije moguća.

Rezultati prema broju riješenih slučajeva

2,061	Riješeno osnovnom analizom egzoma Bolest objašnjavaju samo pojedinačne varijante sekvence ili njihove kombinacije
2,233 - 2,269 +8-15%	Riješeno standardnim visokokvalitetnim egzomom Bolest objašnjavaju CNV-i, kombinacije SNV-a i CNV-a te neke proksimalne intronske varijante
2,592 +26%	Riješeno samo testom ExomeXtra® Rješenje zahtijeva detekciju nekodirajućih varijanti, mozaicizma niskog nivoa ili specijaliziranu bioinformatiku

Dio 3: I duboko sekvenciranje cijelog genoma ima ograničenu osjetljivost

Jedan od argumenata u korist sekvenciranja cijelog genoma jeste da obuhvata „potpuni” genom.

Pokrivenost cijelog genoma znači da očitavanja sekvenciranja obuhvataju svaku genomsku lokaciju i da je njihov broj dovoljan za osjetljivo otkrivanje varijanti na svakoj od tih lokacija.

Međutim, nisu sve regije klinički relevantne. Smatramo da je bolje stvarati podatke tamo gdje su važni: usmjerenim sekvenciranjem duboko pokriti klinički relevantne regije, umjesto rijetko rasporediti očitavanja svuda.

U odnosu na WES, WGS mijenja dubinu pokrivenosti (osjetljivost) za širinu obuhvata (udio zastupljenog genoma). U sljedećem dijelu razmatramo posljedice te smanjene osjetljivosti.

Najprije utvrđujemo gornju granicu pokrivenosti sekvenciranjem kratkih očitavanja. Analiziramo duboko sekvenciran WGS skup podataka (350 Gbp sirovih podataka, prosječna pokrivenost 100×) kako bismo procijenili obuhvat svih klinički relevantnih regija.

Te regije definiramo kao sve kodirajuće sekvence iz baze MANE Select i sve duboko intronske varijante koje uzrokuju bolesti iz HGMD-a i ClinVara. Za osjetljivu detekciju zahtijevamo pokrivenost od najmanje 30×.

Teoretski, WGS skup s pokrivenošću 100× trebao bi omogućiti više od 30× u cijelom genomu. Međutim, regije s ekstremnim sadržajem GC-a ili visoko repetitivnim sekvencama (npr. telomere, centromere i duge regije ponavljanja) gube se pri pripremi biblioteke. Ni sva preostala očitavanja ne mogu se mapirati na

referentni genom, što stvara praznine u pokrivenosti.

Naprimjer, očitavanja mogu biti prekratka za mapiranje na repetitivne regije ili referentna sekvenca može predstavljati varijabilne regije na način nespojiv s podacima dobijenim iz određenog uzorka.

Klinička dijagnostika zahtijeva i strogo filtriranje kako bi mapirana očitavanja tačno predstavljala genom pacijenta. To uključuje uklanjanje duplikata, očitavanja niske kvalitete mapiranja i onih koja se ne poravnavaju jedinstveno na jedan genomski lokus, odnosno imaju više poravnanja s jednakim rezultatom. Takva očitavanja mogu iskriviti alelne učestalosti, dati lažno pozitivne nalaze ili dovesti do propuštanja varijanti.

Nakon ovog filtriranja, WGS skup s pokrivenošću 100× adekvatno pokriva 97.5% CDS-a i 98.8% duboko intronskih varijanti. To nije „potpuna” pokrivenost, ali predstavlja gornju granicu ostvarivog i cilj kojem treba težiti svaki kvalitetan dijagnostički test zasnovan na NGS-u.

Zaključno, čak ni stvaranje ogromne količine podataka ne osigurava 100% pokrivenost cijelog genoma. Ne osigurava ni potpuni obuhvat medicinski relevantnih regija.

Kvalitetan dijagnostički test treba se što više približiti optimumu u regijama u kojima je to važno.

Dio 4: ExomeXtra® je osjetljiviji od standardnog WGS-a

Iako je WGS odličan istraživački alat, dijagnostička primjena NGS-a zahtijeva visoku osjetljivost, a ograničenje predstavlja mogućnost tumačenja otkrivenih varijanti. Naša analiza pokazuje da su medicinski relevantne regije bolje zastupljene u dobro osmišljenom egzomu.

Niža pokrivenost medicinski relevantnih regija povećava mogućnost propuštanja uzročne varijante, posebno kod mozaicizma niskog nivoa (<20%), a naročito kod malih mozaičnih CNV-a niskog nivoa.

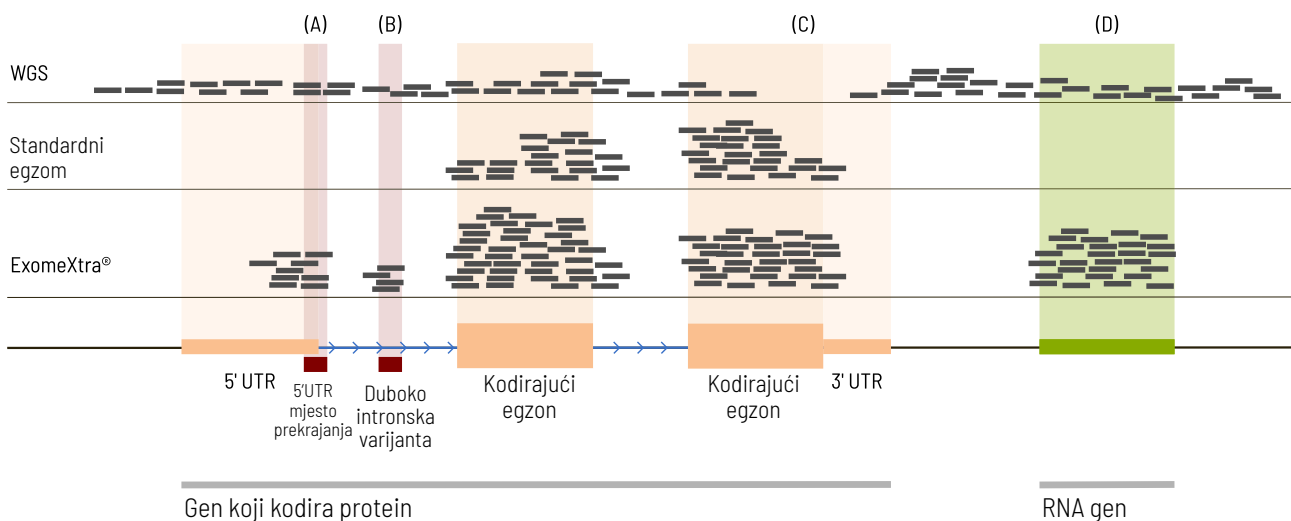
Dijagnostički WGS obično se izvodi uz prosječnu pokrivenost 30×. U prethodnom dijelu prikazali smo gornju granicu pokrivenosti na mnogo dublje sekvenciranom uzorku. Sada poredimo realan WGS skup s podacima standardnog egzoma i testa ExomeXtra®, oba sekvencirana uz prosječnu pokrivenost oko 130×.

Kao u prethodnom dijelu, koristimo klinički relevantne kodirajuće i nekodirajuće regije iz javnih baza podataka

i za osjetljivu detekciju zahtijevamo pokrivenost 30×. WGS s prosjekom 30× očigledno ne može osigurati 30× za sve ciljane regije jer pokrivenost nije potpuno ravnomjerna. Zato za poređenje primjenjujemo i blaži prag od 20×.

Ipak, zbog stohastičkog uzorkovanja molekula DNA tokom pripreme biblioteke i sekvenciranja, niža pokrivenost povećava vjerovatnoću da varijanta potpuno promakne. Smanjenje sa 30× na 20× povećava ovaj rizik sa 0.003% na 0.6%. Drugim riječima, bit će propuštena 1 od 170 heterozigotnih varijanti. Za mozaične varijante ovaj efekat je još izraženiji.

Većina poznatih varijanti koje uzrokuju bolesti nalazi se u kodirajućim regijama (85%) ili blizu granica CDS-a (10%), ali se opisuje sve više relevantnih intergenskih i duboko intronskih varijanti. Istovremeno, nove varijante na takvim položajima rijetko se mogu klinički procijeniti. Prilagođena obogaćivanja egzoma poput testa ExomeXtra® mogu uključiti duboko intronske, UTR i intergenske regije, uz fokus na poznate mutacije koje uzrokuju bolesti.



Slika 3: Shema raspodjele očitavanja duž genoma za WGS (gore), standardno obogaćivanje egzoma i ExomeXtra® (dolje), u odnosu na klinički relevantna obilježja transkriptata. WGS pokriva većinu genoma uz relativno nisku dubinu, ali u nekim regijama može doći do izostanka pokrivenosti (C). Standardni egzomi imaju veću prosječnu pokrivenost, ali propuštaju medicinski relevantne varijante u nekodirajućim regijama i RNA genima (A, B, D). CeGaT ExomeXtra® objedinjuje visoku pokrivenost egzoma s relevantnim nekodirajućim regijama i povećava pokrivenost tamo gdje standardni egzomi i WGS mogu biti nedovoljni.

Podaci pokazuju da ExomeXtra® nadmašuje sekvenciranje cijelog genoma u pokrivenosti klinički relevantnih regija. Praznine u pokrivenosti ne proizlaze iz razlike između egzomskog i genomskog pristupa (odnosno pristranosti obogaćivanja), nego odgovaraju granici mogućeg koju određuju priprema biblioteke, mogućnost mapiranja očitavanja i potrebno filtriranje, kao što je prikazano ranije.

Prvo, podaci standardnog egzoma već pokazuju bolju pokrivenost kodirajućih sekvenci od WGS-a.

Drugo, kod nekodirajućih varijanti koje uzrokuju bolesti iznenađuje da kvalitetan komercijalni egzomski komplet, koji pokriva oko polovine tih regija, nadmašuje WGS. ExomeXtra® je osmišljen za njihovo ciljano obogaćivanje i nadmašuje i WGS i WES pri pragovima 20× i 30×. Za ove varijante ExomeXtra® doseže 98.2% pri pragu 20×, dok je gornja granica duboko sekvenciranog WGS skupa bila 98.8% (vidjeti prethodni dio).

Kao što je navedeno, priprema biblioteke i sekvenciranje su stohastički procesi uzorkovanja, a niža pokrivenost povećava rizik propuštanja varijante. Homozigotne varijante lako se otkrivaju, ali rizik neotkrivanja raste sa smanjenjem alelne učestalosti, od heterozigotnih do mozaičnih varijanti niske učestalosti.

Efekat smanjene osjetljivosti može se simulirati na 2,592 stvarna slučaja prikazana u dijelu 1. Konačni nalazi sadrže 2,775 varijanti sekvence (SNV-i i indeli, bez mitohondrijskog genoma).

Provjerili smo pokrivenost položaja varijanti u WGS skupu 30×. Zajedno s alelnim učestalostima zabilježenim u egzomskim podacima i binomnom raspodjelom, ove vrijednosti omogućavaju procjenu vjerovatnoće lažno negativnih nalaza zbog stohastičkog uzorkovanja.

Utvrđeno je da bi vjerovatno promaklo 1% varijanti: WGS vjerovatno ne bi otkrio 27 varijanti koje su u ovoj kohorti omogućile dijagnozu testom ExomeXtra®. To pokazuje vrijednost usmjeravanja analize i stvaranja više podataka tamo gdje su važni. Povećana osjetljivost glavna je prednost sekvenciranja egzoma. Kod testa ExomeXtra® ona ne zahtijeva kompromis u otkrivanju varijanti relevantnih za bolest u nekodirajućim regijama, jer su obje grupe obuhvaćene velikom dubinom.

Zaključno, ExomeXtra® pokriva više relevantnih regija većom dubinom i zato pruža višu osjetljivost od WGS-a 30×. To povećava udio riješenih slučajeva i omogućava otkrivanje mozaicizma koji se sistematski propušta u većini WGS analiza. ExomeXtra® postiže potrebnu pokrivenost u 98.0% kodirajuće sekvence i 97.3% duboko intronskih nekodirajućih varijanti, dok WGS 30× doseže samo 54.9%, odnosno 52.6%.

Zanimljivo je da smanjena osjetljivost WGS-a u odnosu na sekvenciranje egzoma naizgled proturječi literaturi. Više publikacija tvrdi da sekvenciranje genoma umjesto egzoma znatno povećava dijagnostički prinos. U sljedećem dijelu pokazujemo zašto to može dovesti u zabludu.

Tabela 2: Prosječna pokrivenost i udio (%) ciljnih regija pokrivenih najmanje 20×/30× nakon filtriranja poravnanja za WGS, standardni WES i optimizirani CeGaT egzom. Kodirajuće sekvence: MANE Select 1.5; varijante koje uzrokuju bolesti: HGMD i ClinVar. Dijagnostička pokrivenost izračunata na CDS-u iz MANE Select 1.5.

	WES	WGS	ExomeXtra®
Ukupno sekvenciranih GB	12	100	17.5
Prosječna dijagnostička pokrivenost	~135×	~30×	~125×
Pokrivenost >20×			
Kodirajuće sekvence	97.0	91.4	98.2
Duboko intronske varijante koje uzrokuju bolesti	59.4	89.0	98.4
Pokrivenost >30×			
Kodirajuće sekvence	96.9	54.9	98.0
Duboko intronske varijante koje uzrokuju bolesti	54.3	52.6	97.3

Dio 5: Dijagnostički prinos sekvenciranja genoma niži je nego što tvrde neke publikacije

S padom cijena sekvenciranja, stručna zajednica počela se okretati WGS-u u nadi da će ovaj „najsveobuhvatniji“ test konačno popuniti preostale praznine u dijagnostičkom prinosu. Uprkos većim i bržim uređajima, WGS i dalje zahtijeva znatno više resursa za sekvenciranje, pohranu, obradu podataka i tumačenje. Očekivalo se da taj dodatni napor znatno poveća udio riješenih slučajeva. Istovremeno, dijagnostičke ustanove poput naše stalno su optimizirale WES kako bi obuhvatio sve više varijanti koje uzrokuju bolesti.

Očekivane prednosti WGS-a i prestiž najnovije tehnologije stvaraju snažan pritisak za prelazak s WES-a na WGS u dijagnostici. Međutim, javni zdravstveni sistemi strogo uređuju pokriće troškova, a troškovna efikasnost ostaje vrlo važna. Dodatni resursi za WGS moraju biti opravdani dodatnom koristi. Zato je više grupa pokušalo dokazati korist WGS-a kroz viši dijagnostički prinos i kraći put pacijenta do dijagnoze. Takve studije analiziraju kohorte testirane objema metodama kako bi utvrdile koji slučajevi nisu riješeni WES-om, ali jesu WGS-om.

Uprkos tim nastojanjima, velike dijagnostičke prednosti WGS-a dosad nisu potvrđene: **publikacije prikazuju znatan dodatni prinos, ali se pri pažljivijem uvidu on ne može pripisati novoj tehnologiji. Tri su glavna moguća razloga. Prvo, ako je WES urađen mnogo prije WGS-a, nova naučna saznanja omogućila bi iste dijagnoze i ponovnom analizom WES podataka uz aktuelne baze. Drugo, ponekad se nekvalitetni WES podaci porede s najsavremenijim WGS-om. Treće, WGS varijante smatraju se uzročnim iako su klasificirane kao VUS (varijante neizvjesnog značaja) i ne omogućavaju konačno objašnjenje slučaja.** Slučaj smatramo riješenim samo kada se može izdati značajan nalaz, odnosno kada je otkrivena vjerovatno patogena ili patogena varijanta (ACMG klasa 4 ili 5).

Razmatramo tri publikacije koje navode viši dijagnostički prinos WGS-a u odnosu na WES. U prvoj analiziranoj publikaciji autori navode porast od 14.5 procentnih

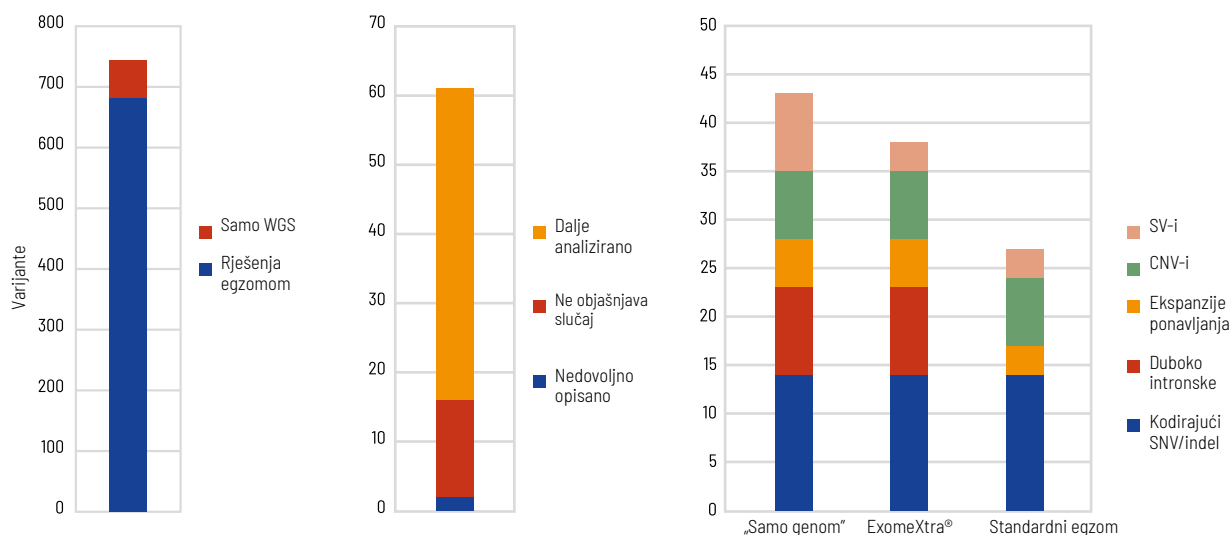
poena (p. p.) [4]. Naša ponovna analiza pokazuje porast od najviše 1.4 p. p. [5]. Međutim, taj dodatni prinos prati gubitak prinosa u drugim područjima, kako je objašnjeno u nastavku.

Druga publikacija nije sadržavala dovoljno podataka za ponovnu analizu, ali istaknuti porast prinosa od 9 procenata [6] odgovara samo 1.8 procentnih poena [7]. Malo je vjerovatno i da bi WES propustio neke spomenute vrste varijanti (npr. egzonske CNV-e). Zanimljivo, očekivanja izgleda nadjačavaju ono što podaci pokazuju: suočeni s neskladom, autori upućuju na treću publikaciju [8]. Iako prihvataju da njihovi podaci pokazuju porast od 1.8 p. p., a nova publikacija 8.2 p. p., zaključuju da je „naša procjena, čini se, slična onima iz drugih GS studija“, praktično tvrdeći da svi znaju da WGS mora znatno povećati prinos.

Sada detaljnije razmatramo treću publikaciju [9]. U kohorti od 744 slučaja autori nalaze 61 varijantu koja objašnjava slučaj i navodno se može otkriti samo WGS-om, što je porast od 8.2 p. p. u odnosu na WES. Podaci su dostupni pa možemo ispitati svih 61 rješenje.

Ponovnu analizu proveli smo u dva koraka. Prvo smo za svaku varijantu iz dodatnih podataka procijenili odluku autora da ona objašnjava slučaj. Zatim smo provjerili tvrdnju da se varijante koje objašnjavaju slučajeve ne mogu otkriti WES-om. Gdje god je postojala neizvjesnost, zbog nedovoljnih podataka ili mogućih razlika u tumačenju, slijedili smo izvornu procjenu autora.

Za dvije od 61 varijante opis nije dovoljan: LINE insercija bez preciznog položaja, koja može, ali ne mora zahvatiti kodirajući egzon, te tandemski duplikacija/delecija navodno u genu MED13L, ali s koordinatama daleko izvan njega. Obje smo tretirali kao varijante koje objašnjavaju slučaj i mogu se otkriti samo WGS-om. Od preostalih 59, koje se mogu procijeniti, 16 zapravo ne objašnjava slučaj: sami autori klasificirali su ih kao VUS.



Slika 4: Ponovna analiza varijanti za koje se tvrdi da su dostupne samo WGS-u. Lijevo: početno se navodi 61 od 744 varijante. U sredini: dvije od 61 nije moguće procijeniti zbog nedostatka podataka, a 16 su VUS i ne objašnjavaju slučaj. Desno: od 43 varijante pogodne za dalju analizu, 26 može otkriti svaki savremeni egzom, a 36 ExomeXtra®.

Preostale 43 varijante koje objašnjavaju slučajeve pripadaju sljedećim kategorijama:

- ✗ 14 su SNV-i ili indeli u kodirajućoj sekvenci. Tvrdnja da WES ne može otkriti varijante u kodirajućim egzonomima svakako zahtijeva preispitivanje. Sve su, naravno, dostupne WES-u, uključujući varijante mjesta prekrajnja 1-2 para baza izvan kodirajućeg egzona, koje WES također obuhvata.
- ✗ 9 su duboko intronski SNV-i ili indeli, predaleko od kodirajuće regije za standardni WES. Međutim, naš prilagođeni ExomeXtra® sve ih obuhvata ciljnim regijama obogaćivanja.
- ✗ 5 su ekspanzije tandemskih ponavljanja. Ove varijante općenito je teško otkriti iz kratkih očitavanja. Uz dovoljnu pokrivenost, ipak nema razloga očekivati bolju detekciju WGS-om nego WES-om. Tri od pet lokacija obuhvaćene su komercijalnim egzomima, a svih pet našim prilagođenim testom ExomeXtra®.
- ✗ 7 su delecije/duplikacije koje zahvataju najmanje jedan cijeli kodirajući egzon (CNV-i). Sve se mogu otkriti iz WES podataka, pa za njihovu detekciju nije potreban WGS.
- ✗ 8 su veće ili složenije strukturne varijante (SV-i): jedna velika UTR delecija, jedna delecija polovine egzona, pet inverzija i jedna složena varijanta. Tri bi se barem djelimično vidjele u WES podacima, iako oni ne bi pokazali njihovu punu složenost. I djelimičan nalaz bio bi dovoljan razlog za detaljniju analizu.

Zaključno, od 61 dodatne varijante dobijene WGS-om u kohorti od 744 slučaja, samo 45 zaista objašnjava slučaj, a samo 17 zahtijeva WGS za otkrivanje (odnosno samo sedam ako se WES radi testom ExomeXtra®). Dodatni prinos WGS-a iznosi 2.3 p. p. u odnosu na standardni egzom i 0.9 p. p. u odnosu na ExomeXtra®, što vjerovatno poništava gubitak detekcije mozaičnih varijanti ili varijanti u slabo pokrivenim regijama.

Kao WGS, i WES se s vremenom unapređivao. Ne iznenađuje da savremeni WGS nadmašuje WES izveden zastarjelim protokolima ili interpretiran bez aktuelnih naučnih saznanja. Savremeni WES koristi nove dokaze i za tumačenje varijanti i za dizajn ciljnih regija, npr. uključivanjem poznatih uzročnih varijanti u nekodirajućim regijama. Ipak, moguće je da WGS donese dodatni prinos i u odnosu na najsavremeniji WES, posebno kroz detekciju složenih strukturnih varijanti. Vjerovatan porast iznosi oko 1 p. p., kao što smo pokazali u komentarima na tri publikacije.

Istovremeno, WGS propušta mozaične varijante niske učestalosti koje WES otkriva. U našoj kohorti takve varijante doprinose oko 1 p. p. dijagnostičkog prinosa. WGS ne pokriva ni sve relevantne regije jednakom dubinom, pa se varijante gube tamo gdje je pokrivenost niska.

Zaključno: WGS jednog pacijenta i Single ExomeXtra® imaju isti dijagnostički prinos, uz malu razliku u fokusu.

Dio 6: Najbolja dijagnostička strategija počinje sekvenciranjem Trio ExomeXtra®

WGS obećava veći dijagnostički prinos od WES-a. U praksi to nije tako. Dijagnostičko tumačenje obično se usmjerava na klinički relevantne regije. Veliki WGS skupovi zato se praktično filtriraju na podskupove slične egzomu, uz zadržavanje nedostataka niže prosječne pokrivenosti. Ciljana WES obogaćivanja poput testa ExomeXtra® imaju suprotan pristup: koncentriraju dubinu sekvenciranja tamo gdje se donose dijagnostičke odluke.

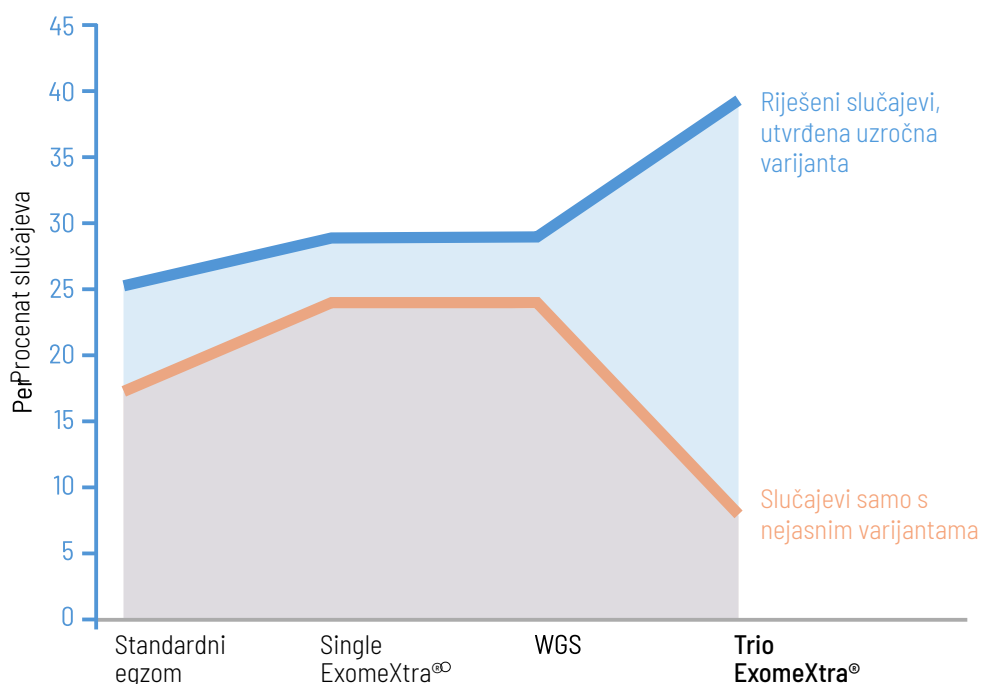
Čak i kada se podaci cijelog genoma analiziraju sveobuhvatno, dodatni prinos WGS-a u odnosu na optimizirani egzom potiče od relativno malog broja složenih strukturnih varijanti. Istovremeno, standardni WGS 30× slabije otkriva mozaične varijante niske učestalosti od duboko sekvenciranog egzoma poput testa ExomeXtra®. WGS također daje velik broj nalaza izvan medicinski relevantnih regija. Oni gotovo uvijek ostaju varijante neizvjesnog značaja i rijetko daju dijagnostičke informacije koje usmjeravaju postupanje. Više podataka samo po sebi rijetko donosi bolji dijagnostički ishod.

prinos za približno 10-15 procentnih poena, mnogo više od dodatne koristi samog prelaska sa sekvenciranja egzoma na genom [11-13].

U praksi ExomeXtra® jednog pacijenta već rješava većinu slučajeva koje bi riješio kvalitetan WGS, a otkriva i neka rješenja koja WGS propušta. Po potrebi se brzo može proširiti na Trio ExomeXtra®. Kod složenih slučajeva opravdano je odmah početi testom Trio ExomeXtra®. WGS je najbolje sačuvati za mali dio neriješenih slučajeva koji zahtijevaju podatke cijelog genoma radi otkrivanja složenih strukturnih varijanti. U odabranim neriješenim slučajevima komplementarne metode, poput sekvenciranja RNA ili analize metilacije, mogu pružiti dodatne funkcionalne dokaze za tumačenje varijanti.

Ključno dijagnostičko pitanje glasi: Koja metoda prvog izbora daje najveću šansu da se objasni moj slučaj? Na osnovu ovdje prikazanih podataka, **Trio ExomeXtra® pruža najveći dijagnostički prinos među procijenjenim strategijama.**

Znatnu razliku donosi dodatna dimenzija podataka: informacije o nasljeđivanju uključene kroz trio analizu. To obično povećava dijagnostički



Slika 5: Izvori podataka za udio riješenih slučajeva, Single/Trio: CeGaT slučajevi s razvojnim kašnjenjem; WGS: prinos procijenjen iz literature; nedostaje do 100%: negativni slučajevi bez prijavljene varijante (Battke, Schulze & Schulte (2024), What Is the Real Diagnostic Benefit of Whole-Genome Sequencing?).

Metode

Dio 1: Pokrivenost je izračunata na tri replike vanjskih referentnih standarda (hg001, hg96, hg97) za svaki egzomski komplet, uz poduzorkovanje sirovih podataka radi direktnog poređenja. Očitavanja 2×100 bp mapirana su pomoću bwa-mem2 (2.1) na referentni genom GRCh38, a pokrivenost je računata vlastitim softverom. Duboko intronske nekodirajuće varijante definirane su kao udaljene više od 30 bp od bilo kojeg poznatog CDS-a; time je dobijeno 7,003 varijante.

Dio 2: Ručno objedinjavanje varijanti koje objašnjavaju 2,592 slučaja prikupljena od uvođenja testa ExomeXtra® verzije 6. Razmotreni su svi klinički slučajevi, uz jedini kriterij odabira: genetski nalaz mora sadržavati genetsko objašnjenje fenotipa („riješeni” slučajevi).

Dijelovi 3 i 4: Sve analize započete su visokokvalitetnom DNA. Egzomi su obogaćeni kompletom Twist Exome 2.0 + Comprehensive Spike-in ili CeGaT ExomeXtra® verzije 7. Sekvenciranje WGS i WES uzoraka izvedeno je

na uređajima Illumina. Dužine očitavanja bile su 2×150 za WGS i 2×100 za WES. Očitavanja su poravnata pomoću bwa-mem2 (2.1, WES uzorci) ili Illumina DRAGEN (4.2.4, WGS uzorci) na referencu GRCh38 s maskiranim pseudoautosomalnim regijama hromosoma Y. Uklanjanje duplikata, filtriranje i izračun pokrivenosti provedeni su vlastitim softverom. Za kodirajuće ciljne regije korištena je baza MANE Select (v1.5): 191,627 regija koje ukupno obuhvataju 33.4 Mbp. Duboko intronske varijante koje uzrokuju bolesti izdvojene su iz HGMD Professional (izdanje 26.2) i ClinVara (preuzeto 2026-07-03). Zadržane su samo nekodirajuće varijante udaljene više od 30 bp od bilo kojeg poznatog CDS-a, ukupno 7,003. Prikazani su prosjeci tri različita uzorka.

Dio 5: Ručna ponovna analiza svih varijanti koje objašnjavaju slučajeve iz dodatnih podataka rada Wojcik i sar. [9].

Dio 6: Udio riješenih Single/Trio slučajeva iz CeGaT kohorte s indikacijom razvojnog kašnjenja. Dijagnostički prinos WGS-a procijenjen je pregledom literature.



Reference

- [1] Trujillano D, Bertoli-Avella AM, Kumar Kandaswamy K, *et al.* Clinical exome sequencing: results from 2819 samples reflecting 1000 families. *Eur J Hum Genet.* 2017;25(2):176-182. <https://doi.org/10.1038/ejhg.2016.146>.
- [2] Alotibi RS, Sannan NS, AlEissa M, *et al.* The diagnostic yield of CGH and WES in neurodevelopmental disorders. *Front Pediatr.* 2023;11:1133789. Published 2023 Mar 1. <https://doi.org/10.3389/fped.2023.1133789>.
- [3] Arteche-López, A., Ávila-Fernández, A., Riveiro Álvarez, R. *et al.* Five years' experience of the clinical exome sequencing in a Spanish single center. *Sci Rep* 12, 19209 (2022). <https://doi.org/10.1038/s41598-022-23786-6>.
- [4] Bertoli-Avella AM, Beetz C, Ameziane N, Rocha ME, Guatibonza P, Pereira C, *et al.* Successful application of genome sequencing in a diagnostic setting: 1007 index cases from a clinically heterogeneous cohort. *Eur J Hum Genet.* 2020. <https://doi.org/10.1038/s41431-020-00713-9>.
- [5] Battke, F., Schulte, B., Schulze, M. *et al.* The question of WGS's clinical utility remains unanswered. *Eur J Hum Genet* 29, 722–723 (2021). <https://doi.org/10.1038/s41431-021-00823-y>.
- [6] Park, J.; Sturm, M.; Seibel-Kelemen, O.; Ossowski, S.; Haack, T.B. Lessons Learned from Translating Genome Sequencing to Clinical Routine: Understanding the Accuracy of a Diagnostic Pipeline. *Genes* 2024, 15, 136. <https://doi.org/10.3390/genes15010136>.
- [7] Battke, F.; Schulze, M.; Schulte, B.; Biskup, S. Comment on Park *et al.* Lessons Learned from Translating Genome Sequencing to Clinical Routine: Understanding the Accuracy of a Diagnostic Pipeline. *Genes* 2024, 15, 136. *Genes* 2024, 15, 1322. <https://doi.org/10.3390/genes150101322>.
- [8] Park, J.; Sturm, M.; Haack, T.B. Reply to Battke *et al.* Comment on "Park *et al.* Lessons Learned from Translating Genome Sequencing to Clinical Routine: Understanding the Accuracy of a Diagnostic Pipeline. *Genes* 2024, 15, 136". *Genes* 2024, 15, 1323. <https://doi.org/10.3390/genes150101323>.
- [9] Wojcik, M.H.; Lemire, G.; Berger, E.; Zaki, M.S.; Wissmann, M.; Win, W.; White, S.M.; Weisburd, B.; Wieczorek, D.; Waddell, L.B.; *et al.* Genome Sequencing for Diagnosing Rare Diseases. *N. Engl. J. Med.* 2024, 390, 1985–1997.
- [10] Ewans, L.J., Minoche, A.E., Schofield, D. *et al.* Whole exome and genome sequencing in mendelian disorders: a diagnostic and health economic analysis. *Eur J Hum Genet* 30, 1121–1131 (2022). <https://doi.org/10.1038/s41431-022-01162-2>.
- [11] Malmgren H, Kvarnung M, *et al.* (2025) Diagnostic yield of 1000 trio analyses with exome and genome sequencing in a clinical setting. *Front. Genet.* 16:1580879. doi: 10.3389/fgene.2025.1580879.
- [12] Farwell, Kelly D. *et al.* (2015) Enhanced utility of family-centered diagnostic exome sequencing with inheritance model-based analysis: results from 500 unselected families with undiagnosed genetic conditions. *Genetics in Medicine* (2015), Volume 17, Issue 7, 578 – 586.
- [13] Retterer, Kyle *et al.* (2016) Clinical application of whole-exome sequencing across clinical indications. *Genetics in Medicine*, Volume 18, Issue 7, 696 – 704.



O nama

CeGaT je globalni pružalac genetskih analiza za širok spektar medicinskih, istraživačkih i farmaceutskih primjena.

Osnovana 2009. u Tübingenu, Njemačka, kompanija povezuje najsavremeniju tehnologiju sekvenciranja s medicinskom ekspertizom kako bi otkrila genetske uzroke bolesti i podržala brigu o pacijentima. Istraživačima i farmaceutskim kompanijama CeGaT nudi širok izbor usluga sekvenciranja i analiza tumora. Stvara podatkovnu osnovu za kliničke studije i medicinske inovacije te vlastitim saznanjima doprinosi napretku nauke.

Kompanija kojom upravljaju vlasnici posvećena je nezavisnosti, sveobuhvatnoj ličnoj podršci klijentima i vrhunskoj kvaliteti. CeGaT laboratorija akreditirana je prema CAP/CLIA, DIN EN ISO 15189 i DIN EN ISO/IEC 17025 te ispunjava najviše međunarodne standarde. Za vrhunske rezultate, svi procesi odvijaju se u vlastitoj ustanovi pod naučnim i medicinskim nadzorom. Posvećenost kvaliteti dopunjena je certifikacijom prema DIN EN ISO/IEC 27001, međunarodno priznatom standardu upravljanja informacijskom sigurnošću.

GenoLab d.o.o.

Informacije i naručivanje testiranja

Mis Irbina 24, 71000 Sarajevo
Bosna i Hercegovina
Telefon: +387 67 1300 035
E-mail: office@genolab.ba

www.genolab.ba/konsultacije

Za informacije o testiranju i dogovor
o daljim koracima kontaktirajte
GenoLab ili skenirajte QR kod.



Accredited according to DIN EN ISO 15189



CLIA CERTIFIED ID: 99D2130225

CeGaT GmbH
Paul-Ehrlich-Str. 23
72076 Tübingen
Njemačka

Tel.: +49 7071 56544-55
Fax: +49 7071 56544-56
Email: info@cegat.com