

Plan upravljanja istraživačkim podacima - MAPinAD

Vugrek, Oliver

Data management plan / Plan upravljanja istraživačkim podacima

Publication year / Godina izdavanja: **2024**

Permanent link / Trajna poveznica: <https://urn.nsk.hr/urn:nbn:hr:241:826349>

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

Download date / Datum preuzimanja: **2025-02-16**



Repository / Repozitorij:

[Fulir DATA - Ruđer Bošković Institute Research Data Repository](#)

Plan upravljanja istraživačkim podacima

| Opće informacije | | |
|------------------|---------------------------------------------------------------------------------------------------------------------------------------------------------|----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| | Ime i prezime predlagatelja | Oliver Vugrek |
| | Matična organizacija | Institut Ruđer Bošković |
| | Naziv projekta | Molekularni aspekti patogenih procesa u AHCY deficijenciji - MAPinAD |
| | Upravitelj podacima | Oliver Vugrek; ovugrek@irb.hr |
| 1. | Prikupljanje podataka i dokumentacija | |
| | Koje ćete podatke prikupljati, obrađivati, stvarati ili se ponovno njima koristiti? (navedite format, vrstu i opseg podataka) | Podaci naprednog sekvenciranja DNA i RNA, FastQ format datoteke; opseg >500 Gb |
| | Kako će se podaci prikupljati, obrađivati ili stvarati? (ukratko navedite metodologiju i procese osiguranja kvalitete te načine organiziranja podataka) | <p>Analiza se razvija na sljedeći način krenuvši od sirovih podataka sekvenciranja (FastQ):</p> <p>A. <u>Generiranje podataka</u></p> <ul style="list-style-type: none"> • Podaci o sekvenciranju – FastQ: tekstualni format za pohranjivanje biološke sekvence (obično nukleotidne sekvence) i njezine odgovarajuće kvalitete. NGS instrument pohrani podatke u obliku FastQ, slično kao što Word pohrani tekst u dokumentu tipa .doc, docx, itd. • Usklađivanje s referentnim genomom (alignment) • Kreiranje datoteke 'Binary alignment map' (BAM) (sadrži pokrivenost po bazi i informacije o kvaliteti) • Kreiranje datoteke 'Browser Extensible Data' (BED): BED sadrži koordinate kromosomskih regija od interesa. • Generiranje datoteke 'Variant Call Format' (VCF) : VCF datoteka sadrži varijacije pronađene u određenom genomu. Sadrži informacije o skupu podataka i relevantnim referentnim izvorima (npr. organizam, inačica izrade genoma itd.), te popis svih otkrivenih varijanti s brojnim dodatnim poljima, kao što su kvaliteta baze, dubina pokrivenosti, itd. <p><u>Potrebni alati za analizu podataka su:</u></p> |

a) Burrows-Wheeler Aligner: BWA

- BWA je softverski paket za mapiranje nisko divergentnih sekvenci u odnosu na veliki referentni genom, kao što je ljudski genom.
- Sastoji se od tri algoritma: BWA-backtrack, BWA-SW i BWA-MEM.
- Prvi algoritam je dizajniran za Illumina sekvence do 100bp, dok su ostala dva za duže sekvence u rasponu od 70bp do 1Mbp.
- BWA-MEM je najnoviji, općenito se preporučuje za visokokvalitetne analize jer je brži i točniji.
- BWA-MEM također ima bolje performanse od BWA-backtrack.

b) GATK: Genome Analysis Toolkit (<https://gatk.broadinstitute.org/hc/en-us>)

- GATK je industrijski standard za identifikaciju SNP-ova i indela u DNA i RNAseq podacima.
- Ovi su alati prvenstveno dizajnirani za obradu WES i WGS podataka generiranih tehnologijom sekvenciranja Illumina
- Premda je izvorno razvijen za ljudsku genetiku, GATK je evoluirao kako može obraditi podatke o genomu bilo kojeg organizma.

Pripadajuće reference za navedene alate su (svi su besplatni):

1. Racz, C. et al. (2013). Isaac: Ultra-fast whole genome secondary analysis on Illumina sequencing platforms. *Bioinformatics* (Oxford, England), 4-5. Doi:10.1093/bioinformatics.btt314 <http://bioinformatics.oxfordjournals.org/content/early/2013/07/01/bioinformatics.btt314>
2. http://www.illumina.com/documents/products/whitepapers/whitepaper_isaac_workflow.pdf Title of Paper: Fast, accurate and easy alignment and variant calling with Isaac Genome Alignment Software and Isaac Variant Caller
3. <https://github.com/Illumina/isaac2>
4. Kim, S. et al. (2017) Strelka2: Fast and accurate variant calling for clinical sequencing applications. *bioRxiv* doi: 10.1101/192872
5. <https://github.com/Illumina/strelka>
6. Chen, X. et al. (2016) Manta: rapid detection of structural variants and indels for germline and cancer sequencing applications. *Bioinformatics*, 32, 1220-1222. doi:10.1093/bioinformatics/btv710
7. <https://github.com/Illumina/manta>
8. Roller, E. et al. (2016). *Bioinformatics* (Oxford England). Doi: 10.1093/bioinformatics/btw163 <http://bioinformatics.oxfordjournals.org/content/early/2016/03/24/bioinformatics.btw163>
9. <https://github.com/Illumina/canvas>

| | | |
|----|----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|--------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| | | <p>10. BWA Genome Alignment Software : Li H. and Durbin R. (2009) Fast and accurate short read alignment with Burrows-Wheeler Transform. <i>Bioinformatics</i>, 25:1754-60. [PMID: 19451168]</p> <p>11. GATK Variant Caller - https://gatk.broadinstitute.org/hc/en-us</p> |
| | Koju ćete dokumentaciju i metapodatke ustupiti osim podataka? (navedite koje su informacije potrebne korisnicima kako bi mogli čitati i interpretirati podatke u budućnosti te koji će se standardi koristiti pri tumačenju podataka) | Ukoliko se ne radi o podacima pacijenata, podaci se mogu ustupiti u obliku FastQ datoteke |
| 2. | Pravna i sigurnosna pitanja | |
| | Jeste li ograničeni sporazumom o povjerljivosti? Imate li potrebna dopuštenja za prikupljanje, obradu, čuvanje i dijeljenje podataka? Jesu li osobe čiji se podaci pohranjuju informirani o tome i jesu li dali privolu? Kojim ćete se metodama koristiti u svrhu zaštite osjetljivih podataka (GDPR - posebne kategorije osobnih podataka)? | <p>Ovaj projektni prijedlog uključuje uporabu životinja (vrsta miš; <i>Mus Musculus</i>). Istraživanjima na životinjama prethodit će višestruki biokemijski eksperimenti <i>in vitro</i> kao i na kulturi stanica. Miševi će se koristiti samo u naprednim fazama istraživanja, kada je moguće odgovoriti na nekoliko specifičnih i visoko relevantnih pitanja na koje se može odgovoriti s ograničenim brojem eksperimenata.</p> <p>Glavni istraživač i istraživački tim radit će u skladu sa svim primjenjivim pravilima, smjernicama i načelima kao što su <i>Direktiva 2010/63/EU</i> o zaštiti životinja korištenim u istraživačke svrhe itd. [navesti sve potrebne pravilnike] Pojediniosti o uporabi životinja. U izvođenju eksperimenata nastojimo se strogo pridržavati <i>3R</i> načela (<i>Replacment, Refinement</i> i <i>Reduction</i> tj. zamjena, usavršavanje i smanjene).</p> <ul style="list-style-type: none"> • Smanjenje (<i>Reduction</i>): Svaki eksperiment dizajnirat će se tako da se smanji broj miševa potrebnih da bi se postigla statistička značajnost. Za predložene farmakokinetičke eksperimente neophodna su 24 miša. • Poboljšanje uvjeta (<i>Refinement</i>): Životinje će boraviti u štali institucije koja se pridržava svih internacionalnih normi. Zdravlje životinja pratit će ovlašteni veterinar. Za smanjenje stresa i nelagodnosti životinja, svi postupci radit će se nakon anesteziiranja životinja. Nakon eksperimenta životinje će se usmrtniti. Također, čim životinje pokažu bilo koji oblik patnje i/ili tjeskobe tumorom tijekom eksperimenta bit će usmrćene dislokacijom vrata nakon što su anestezirane. • Zamjena (<i>Replacment</i>): Tijekom istraživanja alternativa eksperimentiranja na miševima uzet će se u obzir tijekom svih faza projekta. Tamo gdje je moguće, koristit će se druga opcija umjesto miševa. <p>Tečaj. Svi znanstvenici i tehničko osoblje koji rade sa životinjama proći će edukaciju rada sa životinjama.</p> |
| | Kako će se regulirati pristup podacima i njihova | Podaci će se obrađivati i njima upravljati u zaštićenom nemrežnom okruženju koristeći se virtualnom desktop tehnologijom. |

| | | |
|----|------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|-------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| | sigurnost? Koji su potencijalni rizici koje treba uzeti u obzir? Kako ćete osigurati sigurnost pohrane osjetljivih podataka? | |
| | Kako ćete upravljati zaštitom autorskih prava i intelektualnog vlasništva? Tko će biti vlasnik podataka? Koje će se licencije primjenjivati na podatke? Koja će se ograničenja primjenjivati na ponovnu uporabu osobnih podataka? | Ne očekuje se da će rezultat istraživanja dovesti do patenta. Ostali problemi intelektualnog vlasništva će se u rješavati prema preporukama institucije [Institut Ruđer Bošković]. Budući da podaci nisu podvrgnuti ugovoru, te se neće patentirati, objavit će se kao otvoreni podaci pod licencijom <i>Creative Commons CC0</i> . |
| 3. | Pohrana i čuvanje podataka | |
| | Kako će podaci biti pohranjeni i kako će biti napravljena sigurnosna kopija podataka (<i>backup</i>) tijekom istraživanja? Koji su kapaciteti čuvanja podataka kojim raspolazete? Kojim se procedurama koristite za sigurnosnu kopiju (<i>backup</i>)? | Podaci su pohranjeni lokalno (računalo, NAS sustav, eksterni diskovi), te u Cloud sustavu 'Basespace' proizvođača instrumenta za napredno sekvenciranje DNA i RNA Illumina. Kapaciteti na Basespace su vezani uz pretplatu. |
| | Koji je vaš plan čuvanja podataka? U kojim će se formatima čuvati? | |

| 4. Dijeljenje i ponovna uporaba podataka | | |
|------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|--|
| Kako i gdje će se podaci dijeliti? Na kojem repozitoriju planirate dijeliti podatke? Kako će potencijalni korisnici doznati za podatke? | Konačnu verziju skupa podataka voditelj projekta podijelit će putem institucijskog repozitorija IRBa uspostavljenog u nacionalnom sustavu Dabar gdje će biti pohranjene i publikacije i ostala projektna dokumentacija. Podaci će biti objavljeni pod CC0 licencom. Institucijski repozitorij u sustavu Dabar odabrali smo jer podržava FAIR principe: skupovima dodjeljuje trajni identifikator URN:NBN, osigurava vidljivost podataka putem OpenAIRE portala i Google Scholar a te tražilice dabar.srce.hr, a ujedno doprinosi vidljivosti i transparentnosti rada IRBa. | |
| Ako postoje podaci koji se ne smiju dijeliti (prijavitelji vezani zakonskim, etičkim, autorskim pravila, povjerljivošću i sl.), pojasnite razloge ograničenja. | Podaci neophodni za bilo koju publikaciju bit će dostupni u trenutku objavljivanja. Podaci pacijenata ne će se ustupiti zbog etičkih pravila. | |
| Potvrdite da ćete se koristiti digitalnim repozitorijem koji je u skladu s načelima FAIR-a. | Koristit će se digitalni repozitorij koji je u skladu s načelima FAIR-a. | |
| Potvrdite da ćete se koristiti digitalnim repozitorijem koji održava neprofitna organizacija (ako ne, objasnite zašto ne možete dijeliti podatke na digitalnom repozitoriju koji nije komercijalan). | Potvrđujem kako će se koristiti digitalni repozitorij koji održava neprofitna organizacija. | |

Ref:

[1] Celjak, D., Dorotić Malič, I., Matijević, M., Poljak, Lj., Posavec K. i Turk, I.: „Istraživački podaci - što s njima?“ [Istraživački podaci - što s njima?: priručnik o upravljanju istraživačkim podacima | Digitalni repozitorij Srca \(unizg.hr\)](#)