PLAN UPRAVLJANJA ISTRAŽIVAČKIM PODATCIMA (PUP)

|  |  |  |
| --- | --- | --- |
| Opće informacije | | |
|  | Ime i prezime predlagatelja | Prof. dr. sc. Siniša Volarević |
|  | Matična organizacija | Medicinski Fakultet u Rijeci, Sveučilište u Rijeci |
|  | Naziv projekta | Međudjelovanje heterozigotne mutacije gena za ribosomski protein L24 i p53 u patogenezi bolesti |
|  | Upravitelj podatcima | Prof. dr. sc. Siniša Volarević, email: [sinisa.volarevic@uniri.hr](mailto:sinisa.volarevic@uniri.hr) |
| 1. | Prikupljanje podataka i dokumentacija | |
|  | Koje ćete podatke prikupljati, obrađivati, stvarati ili se ponovno njima koristiti? (navedite formate, vrste i opseg svih podataka s kojima ćete raditi, a ne samo krajnji skup podataka koji će biti rezultat istraživanja) | Tijekom projekta proizvest ćemo nekoliko vrsta neobrađenih podataka:  1. Digitalne slike Western blota (detekcija i kvantifikacija pojedinačnih proteina u lizatima stanica)  2. Digitalne slike elektroforeze na akrilamidnom ili agaroznom gelu (detekcija i kvantifikacija razina tRNA i rRNA u stanicama)  3. Digitalne slike obojenih mišjih embrionalnih fibroblasta (MEF) i presjeka mišjih novookoćenih tkiva dobivene konfokalnom fluorescentnom mikroskopijom.  4. Digitalne histološke slike novookoćenih tkiva miševa dobivene svjetlosnom mikroskopijom.  5. Podatci protočne citometrije  6. Skupovi podataka prikupljeni mjerenjem enzimske aktivnosti (antioksidacijski enzimi i reaktivni kisikovi spojevi)  7. Podatci o preživljenju novookoćenih miševa (u satima)  8. Podatci dobiveni sekvenciranjem RNA  9. Podatci dobiveni pokusima kvantitativne lančane reakcije polimerazom (PCR) u stvarnom vremenu  Sve digitalne slike bit će pohranjene u izvornim formatima koje je proizveo uređaj. TIF format za slike Western blota; JPG za slike elektroforeze u akrilamidnom i agaroznom gelu; CZI za slike dobivene konfokalnom fluorescentnom mikroskopijom; JPEG za slike dobivene svjetlosnom mikroskopijom; FCS za podatke protočne citometrije; CSV za skupove podataka prikupljene mjerenjem enzimske aktivnosti, podatke o preživljavanju novookoćenih miševa, podatke dobivene pokusima PCR-a u stvarnom vremenu; fastaq za podatke dobivene sekvenciranjem RNA. Rezultati mjerenja i kvantifikacije digitalnih slika pohraniti će se u xmls formatu, a za dugotrajnu arhivu pretvoriti će se u CSV format. Za pohranjivanje slika prikupljenih tijekom projekta bit će potrebno između 1TB i 2TB. Između 2TB i 3TB bit će potrebno za pohranjivanje podataka dobivenih sekvenciranjem RNA. Za podatke dobivene ostalim mjerenjima i kvantifikacijama neće biti potrebno više od 500MB. |
|  | Kako će se podatci prikupljati, obrađivati ili stvarati? (ukratko navedite metodologiju i procese osiguranja kvalitete, načine organiziranja podataka te alate i instrumente kojima ćete se koristiti za prikupljanje i obradu) | Digitalne slike Western blota bit će prikupljene pomoću uređaja ImageQuant™ LAS 4000. Slike će se najprije snimiti u automatskom načinu rada. U slučaju preintenzivnog signala, bit će snimljen veći broj slika s kraćim vremenima ekspozicije (do deset fotografija). Uređaj za snimanje slika dobivene će slike izravno spremiti kao TIF dokument. U ovom koraku slike će se imenovati i organizirati unutar odgovarajućih mapa prema epitopu i datumu kada je snimanje obavljeno. Pojedinačne slike bit će imenovane prema epitopu, a ime svake fotografije sadržavat će vrijeme ekspozicije u sekundama. Digitalne slike elektroforeze na akrilamidnom i agaroznom gelu bit će snimljene pomoću transiluminatora (Cleaver Scientific) s digitalnim fotoaparatom Canon Power shot G10 podešenim u automatskom načinu rada. Slike će se zatim razvrstati u zasebne mape na računalu i imenovati prema sljedećem principu: Naziv Pokusa\_Broj\_Datum. Signali Western blota te akrilamidnih i agaroznih gelova mjerit će se pomoću slobodno dostupnog programa ImageJ, a dobiveni podatci će se uvesti u Excel i pretvoriti u Excel tablicu. Za procjenu pozadinskog intenziteta na slikama fluorescentne mikroskopije, uzorci će biti obojeni fluorescentno obilježenim sekundarnim protutijelima. Nakon toga, pozadinski signal oduzet će se od eksperimentalnih signala. Kontrolna, a potom i eksperimentalna skupina će se snimati kako bi se došlo do valjanog zaključka o nastaloj promjeni intenziteta i lokalizacije promatranog signala. Slike će se organizirati u zasebne mape ovisno o vrsti uzorka, analiziranim epitopima i datumu snimanja. Naziv svake slike sadržavat će tip uzorka, genotip, naziv epitopa i fluorofora, ukupno povećanje objektiva i redni broj slike. U slučaju snimanja niza slika u osi z, nazivi slika zadržat će dodatak z\_stack prije serijskog broja. Digitalne slike obojenih MEF-ova i presjeka tkiva novookoćenih miševa dobivene konfokalnom fluorescentnom mikroskopijom bit će pohranjene u izvornom CZI formatu, ali će se za potrebe analize pomoću programa ImageJ pretvoriti u TIF format, a sirovi podatci dobiveni mjerenjem kopirat će se iz ImageJ programa u Excel tablice koje će se zatim arhivirati u CSV formatu. Digitalne histološke slike novookoćenih tkiva miševa dobivene svjetlosnom mikroskopijom obradit će se i pohraniti u izvornom JPG formatu. Mjerenja će se provesti u programu ImageJ, a sirovi podatci dobiveni mjerenjem kopirat će se iz programa ImageJ u Excel tablice koje će se zatim arhivirati u CSV formatu. Podatci protočne citometrije obraditi će se programom FlowJo™ i arhivirati u FCS formatu. Svi podatci protočne citometrije biti će organizirani u mape prema datumu snimanja, nazivu genotipa i vrsti pokusa kojem pripadaju. Konačni obrađeni rezultati bit će pohranjeni u CSV formatu. Skupovi podataka prikupljeni mjerenjem enzimske aktivnosti i podatci o preživljenju novookoćenih miševa u satima izravno će se unijeti i organizirati u Excel tablicama te pohraniti u CSV formatu. Podatci dobiveni sekvenciranjem RNA bit će pohranjeni u izvornom strojno generiranom formatu fastaq s naknadnom kompresijom u format fastq.gz. Svi podatci dobiveni RNA sekvenciranjem bit će organizirani u mape nazvane prema genotipu MEF-ova. Podatci dobiveni pokusima kvantitativnog PCR-a u stvarnom vremenu obradit će se u Excelu, a dobiveni rezultati će se izvesti i arhivirati u CSV formatu. Statistička obrada dobivenih podataka, osim podataka dobivenih sekvenciranjem RNA, provest će se korištenjem programskog paketa Statistica 13. Obradu podataka RNA sekvenciranja provest će naši suradnici Jiri Bartek i Mikael Lindstrom (Karolinska) uz korištenje odgovarajućih programa za statističku analizu tih podataka. |
|  | Koju ćete dokumentaciju i metapodatke izraditi osim podataka? (dokumentacija mora sadržavati informacije i standarde potrebne korisnicima kako bi mogli samostalno čitati i interpretirati podatke u budućnosti, primjerice, kodne knjige, ReadMe datoteke i sl.) | Svi podatci će biti popraćeni dokumentacijom koja ih detaljnije opisuje. Jedan dokument sadržavat će popis kratica koje se koriste za imenovanje podataka i njihovo značenje. Za sve prikupljene podatke napisat će se protokol pokusa s popisom korištenih kemikalija i načinom prikupljanja podataka. Za svaku grupu slika Western blota, elektroforeze na akrilamidnom ili agaroznom gelu proizvest će se txt dokument s rednim brojevima uzoraka. Digitalne slike dobivene konfokalnom fluorescentnom mikroskopijom ili svjetlosnom mikroskopijom, podatci protočne citometrije, skupovi podataka prikupljeni mjerenjem enzimske aktivnosti, podatci o preživljenju novookoćenih miševa i podatci dobiveni pokusima kvantitativnog PCR-a u stvarnom vremenu sadrže vlastite kvalitativne metapodatke i nisu potrebni nikakvi dodatni dokumenti osim protokola provedbi pokusa. |
| 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 podatci obrađuju informirani o tome i jesu li dali privolu? Kojim ćete se metodama koristiti u svrhu zaštite osjetljivih podataka (GDPR - posebne  kategorije osobnih podataka, navesti metode anonimizacije podataka)? | Projektni prijedlog uključuje korištenje pokusnih miševa (vrsta miša; Mus musculus). U pripremi za ovaj projekt izolirali smo primarne MEF-ove iz gravidnih ženki miševa specifičnih genotipova na embrionalnom danu 12.5. Gdje god je to moguće, koristit ćemo te MEF-ove u kulturi kako bismo razjasnili molekularne mehanizme koji doprinose patološkom fenotipu miševa. Vivarij na Medicinskom fakultetu Sveučilišta u Rijeci osigurava izvrsne uvjete za uzgoj i križanje pokusnih miševa. Osoblje vivarija stručno je i ima veliko iskustvo u radu s laboratorijskim miševima. Pokusi će se u potpunosti provoditi u skladu s postojećim nacionalnim zakonodavstvom i direktivama EU te ćemo poštivati sva etička načela i standarde koji su relevantni za predloženi projekt. Naš Kodeks ponašanja čvrsto je usmjeren na primjenu načela "3 R": dugoročno smanjenje, zamjena i poboljšanje eksperimenata prilikom upotrebe miševa u pokusnim protokolima. Članovi tima prof. Volarevića posjeduju certifikate o osposobljenosti za rad s laboratorijskim životinjama u istraživačke svrhe. Imaju značajno iskustvo i stručnost u upravljanju kolonijama miševa, smanjenju broja životinja u rasporedima uzgoja, vođenju točne evidencije uzgoja i provedbi in vivo pokusa na miševima. S obzirom prikupljati podatke u pokusima na miševima koji se uzgajaju dugi niz godina uz svu potrebnu dokumentaciju, nikakav sporazum o povjerljivosti nije primjenjiv u ovom slučaju. |
|  | Kako će se regulirati pristup podatcima i njihova sigurnost? Koji su potencijalni rizici koje treba uzeti u obzir? Kako ćete osigurati sigurnost pohrane osjetljivih podataka? | Objavljeni podatci prikupljeni istraživanjima bit će javno dostupni uz poštivanje standarda u području istraživanja putem neke od javnih baza podataka istraživanja (tj. FigShare). Pristup objavljenim i neobjavljenim podatcima prikupljenim istraživanjima nakon objave bit će omogućen članovima znanstvene zajednice na zahtjev. Razlog tome je nemogućnost dijeljenja svih dobivenih podataka preko mrežnih infrastruktura zbog njihove ogromne količine. Veliki skupovi podataka dugotrajno će se pohraniti lokalno na Zavodu u najmanje dvije kopije kako bi se zaštitio integritet podataka u slučaju nepredvidljivih nezgoda. Neće postojati klasificirani podatci ili podatci zaštićeni GDPR-om. Stoga, osiguranje visoke razine sigurnosti je potrebno u ovom slučaju. |
|  | Kako ćete upravljati zaštitom autorskih prava i drugog intelektualnog vlasništva? Tko će biti vlasnik podataka? Koje će se licencije primjenjivati na podatke? Koja će se ograničenja primjenjivati na ponovnu  uporabu osobnih podataka? | Svi podatci prikupljeni u sklopu istraživanja bit će licencirani pod licencom CC BY 4.0 prema preporuci Europske komisije kako bi se omogućilo višekratno i ispravno korištenje podataka od strane članova znanstvene zajednice. Vlasnik intelektualnog vlasništva je Zavod za molekularnu medicinu i biotehnologiju Medicinskog fakulteta u Rijeci na kojemu će se podatci prikupiti. S obzirom da nema osobnih i povjerljivih podataka, osim intelektualnih dostignuća, na podatke se neće primjenjivati nikakva ograničenja. |
| 3. | Pohrana i čuvanje podataka | |
|  | Kako će radne verzije podataka biti pohranjene tijekom projekta?  Kako će se napraviti sigurnosne kopije tih podataka (backup)?  Koja je očekivana količina podataka koja će se prikupiti i čuvati tijekom projekta (izraženo u MB/GB/TB)? | S obzirom na veliku količinu ukupno prikupljenih podataka (4-5TB), učinkovita pohrana na MojOblak ili PUH platformu nije prihvatljivo rješenje. Stoga će se u sklopu projekta kupiti nekoliko (2) vanjskih tvrdih diskova kapaciteta 6TB koji će nam omogućiti učinkovito pohranjivanje i pristup podatcima. Također će se kupiti nekoliko (4) unutarnjih tvrdih diskova kapaciteta 2TB za privremenu pohranu i obradu proizvedenih podataka. Jedan od vanjskih tvrdih diskova koristit će se za izradu sigurnosne kopije. Ovaj kapacitet diskova omogućuje nam pohranu svih podataka u najmanje dvije kopije, a onih značajnijih u najmanje tri kopije. Osim toga, tijekom trajanja ovog istraživanja jedna kopija svih podataka bit će pohranjena na lokalnim računalima kako bi se omogućio brzi pristup i jednostavnija obrada podataka. |
|  | Kako će se završne verzije podataka dugotrajno pohraniti i čuvati (i nakon završetka projekta)?  U kojim će se formatima čuvati podatci? Koja je očekivana količina podataka koja će se trajno pohraniti (izraženo u MB/GB/TB)? | Svi objavljeni i neobjavljeni podatci prikupljenim istraživanjima bit će pohranjeni najmanje tri godine nakon završetka projekta u javno dostupnim nekomercijalnim bazama podataka Hrvatske znanstvene infrastrukture (Dabar, PUH) i na tvrdim diskovima instituta. Većina podataka dobivenih sekvenciranjem RNA bit će pohranjena u formatu fastq bez potrebe za kompresijom, budući da ovaj format već koristi složene biblioteke za optimizaciju kompresije, a veličina je malo smanjena standardnim metodama. Za pohranu fastq datoteka koristit će se standardno adaptivno Lempel-Ziv kodiranje kroz program gunzip kako bi se podatci komprimirali u format fastq.gz i time drastično smanjila njihova veličina (standardni omjer za konverziju fastq u fastq.gz je 6:1) . Ostale vrste prikupljenih podataka neće biti komprimirane kako bi se izbjegle moguće pogreške tijekom procesa kompresije/dekompresije. Tablični podatci će biti spremljeni u CSV formatu, tekstualni podatci u docx, txt ili PDF formatu. |
| 4. | Dijeljenje i ponovna uporaba podataka | |
|  | Kako i gdje će se podatci dijeliti? Koji repozitorij će se koristit za dijeljenje podataka? Kako će potencijalni korisnici doznati za podatke? | Velika količina temeljnih i neobrađenih podataka bit će pohranjena na unutarnjim i vanjskim tvrdim diskovima na Zavodu za molekularnu medicinu. Ti će podatci biti dostavljeni svim znanstvenicima koji zatraže dopuštenje za njihovo korištenje, osobno ili na zahtjev putem predviđenih protokola za dijeljenje podataka (izravni FTP/SFTP prijenos ili putem programa za prijenos kao što su WinSCP ili FileZilla). Podatci dobiveni izračunima (tablice, genomska poravnanja) dijelit će se na javnim repozitorijima koji se koriste u tu svrhu (National Center for Biotechnology Information, FigShare, European Nucleotide Archive), dok će informacija o postojanju tih podataka biti dostupna u objavljenim publikacijama. |
|  | Ako postoje podatci koji se ne smiju dijeliti (prijavitelji vezani zakonskim, etičkim, autorskim pravila, povjerljivošću i sl.), pojasnite razloge ograničenja. | Svi podatci prikupljenim istraživanjima mogu se dijeliti s atribucijom pod licencom CC BY 4.0. |
|  | Potvrdite da ćete se koristiti digitalnim repozitorijem koji je u skladu s načelima FAIR-a. | Digitalni repozitorij slijedit će sva 4 načela FAIR, Findable - dostupan na internetu s poveznicama iz objavljenih publikacija, Accessible - može mu se pristupiti stabilnom internetskom vezom, Interoperable - podatci se preuzimaju u standardnim formatima za pohranu podataka kao što je .zip, Reusable - svi metapodatci bit će objavljeni zajedno s glavnim dijelom podataka. |
|  | 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). | Jedna kopija digitalnih podataka bit će pohranjena u javno dostupnim nekomercijalnim bazama podataka Hrvatske znanstvene infrastrukture (Dabar, PUH) s platformom Figshare. |