**PLAN UPRAVLJANJA ISTRAŽIVAČKIM PODACIMA (PUP)**

|  |  |  |
| --- | --- | --- |
| Opće informacije | | |
|  | Ime i prezime predlagatelja | Maja Sabol |
|  | Matična organizacija | Institut Ruđer Bošković |
|  | Naziv projekta | Uloga signalnog puta Hedgehog-GLI u komunikaciji tumora i strome posredovanoj faktorima rasta u 2D i 3D modelima in vitro |
|  | Upravitelj podacima | Maja Sabol |
| 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) | Svi eksperimentalni postupci i ispisi rezultata zapisivat će se na dnevnoj bazi u fizički Laboratorijski dnevnik (bilježnica). U njima se opisuju svi eksperimentalni postupci, uzorci, odstupanja od standardnih protokola, nazivi i opisi kemikalija kao što su protutijela i početnice, te komentari na dobivene rezultate. Rezultati će se u dnevnik bilježiti u obrađenom obliku (grafovi, zbirne tablice, ispisi slika isl.).  Elektronički podaci će se prikupljati na sljedeći način:  Numeričke vrijednosti u Excel dokumentima  Tekstualni podaci u Word dokumentima  Slike u .tiff i .jpg formatima  Dokumenti proizašli sa specijaliziranih instrumenata u pripadajućim formatima (.pcrd dokumenti sa qPCR instrumenta, .fcs dokumenti s MUSE Cell Analyzes i Cell Sorter instrumenata, .dna dokumenti iz SnapGene Viewer programa, rezultati sekvenciranja u obliku .abi dokumenata)  Očekuje se prikupiti oko 200 GB podataka, od čega će većinu zauzeti slike i rezultati sekvenciranja |
|  | Kako će se podaci 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) | Podaci će se generirati na specijaliziranim instrumentima za pojedinu analizu.  Rezultati qPCR analize (.pcrd) će se generirati i spremati na qPCR instrumentu (neograničeno vrijeme), te kopirati na osobno računalo istraživača, gdje će se analizirati i procesirati u numeričke i/ili tekstualne podatke.  Slike će se pohranjivati (neograničeno vrijeme) na specijaliziranim instrumentima (npr. UVITEC uređaj za snimanje Western blotova, računalo vezano uz mikroskop i s njim povezanu kameru). Podaci će se kopirati na osobno računalo istraživača, gdje će se analizirati i procesirati u numeričke i/ili tekstualne podatke, uz zadržavanje originalnih neobrađenih slika.  Podati sa MUSE Cell Analyzer i Cell Sorter instrumenata (.fcs) pohranjivat će se na samim instrumentima (neograničeno vrijeme), te kopirati na osobno računalo istraživača, gdje će se analizirati i procesirati u numeričke i/ili tekstualne podatke.  Rezultati RNA-seq analize će se generirati u kompaniji koja pruža navedene usluge, i elektronički podijeliti putem on-line računa sa voditeljem istraživanja. Originalni podaci će se kopirati na računalo glavnog istraživača i podijeliti s bioinformatičarem koji će te podatke obraditi u numeričke podatke. Dobiveni rezultati pohranit će se na računalima istraživača i dalje obrađivati.  Svi numerički/tekstualni podaci pohranjivat će se na osobnim računalima istraživača, te jednom mjesečno kopirati na eksterni čvrsti disk za dugoročnu pohranu. |
|  | 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.) | Podaci će se organizirati u datoteke podijeljene prema projektnim aktivnostima, i pod-datoteke podijeljene po metodama. Ovakav način organizacije podataka koristili smo i u prethodnom projektu, tako da ga na jednostavan način može koristiti više istraživača.  Tablice i liste sa specifikacijama i popisima kemikalija i potrošnog materijala (npr. protutijela, početnice, skevence vektora i slični podaci) podijelit će se putem institucijskog Oblaka sa svima članovima laboratorija.  Za praćenje provođenja i dovršenosti eksperimenata generirat će se Excel tablice sa specifičnim grupama uzoraka i eksperimentalnim postupcima. Ovi podaci također će biti dostupni svim članovima laboratorija putem institucijskog Oblaka. |
| 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 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)? | Jedan dio istraživanja uključuje rad na materijalu iz pacijenata. Materijal uključuje primarne stanice tumora pohranjene zamrznute u tekućem dušiku te izoliranu RNA pohranjenu zamrznutu na -80°C. Ovaj materijal prikupljen je u skladu sa svim etičkim principima, uključujući odluke Etičkih povjerenstava i informirane pristanke pacijenata. Uzorci su de-identificirani, te istraživači koji rade s tim uzorcima ne mogu povezati uzorke s pacijentima na temelju oznaka. Ovaj materijal se smatra arhivskim materijalom laboratorija. Dokument kojim se oznake uzoraka povezuju s imenima pacijenata spremljen je lokalno na računalima liječnika koji je prikupljao uzorke te glavnog istraživača. |
|  | Kako će se regulirati pristup podacima i njihova sigurnost? Koji su potencijalni rizici koje treba uzeti u obzir? Kako ćete osigurati sigurnost pohrane osjetljivih podataka? | Pristup podacima o pacijentima ima samo liječnik koji je prikupljao uzorke i glavni istraživač na projektu (kako je i definirano u odobrenju Etičkog povjerenstva), i podaci su spremljeni lokalno na dva računala. Razlog za spremanje na dva mjesta je sigurnosni, kako ne bi zbog kvara računala došlo do gubitka jedine kopije dokumenta. Na taj način druga kopija služi kao sigurnosna kopija. Ovi podaci nisu pohranjeni u nikakvom repozitoriju niti na Oblaku, jer se smatraju povjerljivima. Razmjena ovih podataka dopuštena je između liječnika koji je prikupio uzorke i glavnog istraživača. |
|  | 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? | Ne očekujemo probleme s pravima intelektualnog vlasništva.  Intelektualno vlasništvo (copyright) je definirano časopisima u kojima će se objaviti rezultati istraživanja, a cilj je objaviti rezultate u časopise s otvorenim pristupom, jer u tom slučaju autori zadržavaju pravo intelektualnog vlasništva nad objavljenim podacima. |
| 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)? | Svi prikupljeni podaci označavat će se datumom i opisom, s pripadajućim unosima u Laboratorijskom dnevniku na pripadajuće datume. Na taj način svi podaci se mogu povezati s fizičkom kopijom dnevnika.  Podaci će se sortirati u datoteke i pod-datoteke. Projektne aktivnosti će biti glavne datoteke, s pod-datotekama koje uključuju podjelu po metodologiji. Svi novi podaci pohranjivat će se pod odgovarajaućim datumom i opisom u odgovarajuću datoteku.  Za podatke koji će se kontinuirano nadopunjavati (npr. radne verzije znanstvenih radova tijekom pisanja), svaka nova verzija pohranjivat će se s novim datumom, uz zadržavanje prethodnih verzija. Kada podaci u starijim verzijama više ne budu relevantni, obrisat će se (npr. starije verzije znanstvenih radova u pripremi).  Procjenjujemo da će se prikupiti oko 200 GB podataka. Većinu toga zauzet će originalni rezultati RNA-seq analize (otprilike 1-2 GB po uzorku, procijenjeno 40-50 GB ukupno) te slike (ovisno o veličini i kvaliteti slika, otprilike 100 GB). Preostalih 50 GB zauzet će numerički, tekstualni i sirovi podaci kao što je gore opisano. |
|  | Kako će se završne verzije podataka dugotrajno pohraniti i čuvati (i nakon završetka projekta)?  U kojim će se formatima čuvati podaci? Koja je očekivana količina podataka koja će se trajno pohraniti (izraženo u MB/GB/TB)? | Tijekom trajanja projekta podaci će se prikupljati na dnevnoj bazi, te pohranjivati lokalno na osobna računala istraživača. Većina obrade podataka odvijat će se na tim računalima. Obrađeni i skupljeni rezultati podijelit će se s glavnim istraživačem po potrebi, te dodatno pohraniti na računalu glavnog istraživača. Sigurnosna kopija prikupljenih podataka radit će se jednom mjesečno kopiranjem podataka na vanjski čvrsti disk za dugotrajnu pohranu.  Dio podataka dijelit će se putem e-maila, SKYPE-a, OneNote programa ili putem institucijskog Oblaka, kada postoji potreba da više istraživača ia pristup istom setu podataka.  Dokumenti će se pohranjivati u obliku neobrađenih sirovih podataka dobivenih sa specijaliziranih instrumenata, te kao obrađeni i skupljeni rezultati u oblicu Excel i Word dokumenata.  Svi prikupljeni podaci (otprilike 200 GB) pohranit će se na vanjski čvrsti disk na neograničeno vrijeme, dok će se lokalno spremljene kopije podataka na računalima istraživača brisati po potrebi (npr. u slučaju manjka mjesta za pohranu) ali tek nakon objave znanstvenog rada. |
| 4. | Dijeljenje i ponovna uporaba podataka | |
|  | Kako i gdje će se podaci dijeliti? Koji repozitorij će se koristit za dijeljenje podataka? Kako će potencijalni korisnici doznati za podatke? | Većina podataka će se obraditi da bude prikladna za objavu u znanstvenim časopisima. U većini slučajeva to neće uključivati originalne sirove podatke, nego obrađene rezultate u grafičkom obliku (grafovi, slike s više dijelova isl.). Mnogi časopisi zahtijevaju slanje neobrađenih slika western blot analiza tijekom slanja rada, tako da će se i takvi podaci poslati u časopis po potrebi.  Za neke tipove podataka, kao što su podaci RNA-seq analiza, postoji obaveza pohranjivanja u on-line repozitorije (npr. ArrayExpress baza), te je to obavezni dio u tijeku objave znanstvenog rada. Podaci o dostupnosti navedenih podataka moraju se navesti u znanstvenom radu tijekom objave, tako da i drugi istraživači mogu pristupiti navedenim podacima. |
|  | Ako postoje podaci koji se ne smiju dijeliti (prijavitelji vezani zakonskim, etičkim, autorskim pravila, povjerljivošću i sl.), pojasnite razloge ograničenja. | Povjerljivi podaci koji s tiču identiteta pacijenata se neće dijeliti ni na koji način. Svi podaci koji se tiču kliničkih uzoraka prikazivat će se na način da su deidentificirani te zbirno, jer je to stadardni način prikazivanja podataka o pacijentima. Obično se prikazuje zbirna tablica sa karakteristikama uzoraka, te se pojedine točke kod prikaza rezultata ne mogu povezati sa pojedinim pacijentom. |
|  | Potvrdite da ćete se koristiti digitalnim repozitorijem koji je u skladu s načelima *FAIR-a*. | ArrayExpress je baza podataka u koju se podaci velikih eksperimenata funkcionalne genomike pohranjuju i stavljaju na raspolaganje istraživačkoj zajednici. Prema publikaciji Bonaretti i and Willighagen (doi: <https://doi.org/10.1101/739334>), ArrayExpress baza ispunjava potrebne kriterije I indikatore. |
|  | 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). | ArrayExpress je jedan od repozitorija koji se preporučuje akademskoj zajednici za pohranu podataka funkcionalne genomike dobivenih na platformama za mikročipove i masivno paralelno sekvenciranje. Dio je ELIXIR programa (set europskih resursa od fundamentalne važosti za širu znanstvenu zajednicu i za dugoročnu pohranu bioloških podataka). Ovaj repozitorij održava EMBL-EBI, koji je međunarodna istraživačka organizacija. |