

»BIOINFORMATIKA IN UMETNA INTELIGENCA NA PODROČJU ODKRIVANJA MOLEKULARNIH BIOOZNAČEVALCEV ZA NAPOVED POTEKA IN ZDRAVLJENJA KRONIČNIH IMUNSKIH BOLEZNI«

v okviru » **»Problemsko učenje študentov v delovno okolje: gospodarstvo, negospodarstvo in
neprofitni sektor v lokalnem/regionalnem okolju 2024-2027 (PUŠ v delovno okolje 2024-2027) - 2.
odpiranje«**

Operacijo sofinancirata Republika Slovenija in Evropska unija iz Evropskega socialnega sklada plus

Prenos znanja in izkušenj pedagoškega mentorja partnerju iz delovnega okolja se je izvajal tekom celotnega obdobja izvajanja projektne naloge.

MAREC 2026

V mesecu marcu je potekala intenzivna predaja znanj s področja sekvenciranja DNA in RNA ter bioinformatične analize pridobljenih podatkov. Aktivnosti so bile usmerjene v razumevanje celotnega poteka priprave knjižnic za sekvenciranje naslednje generacije, delovanja sekvenatorja Illumina NextSeq550 ter osnovnih postopkov kontrole kakovosti sekvenčnih podatkov. Poseben poudarek je bil namenjen razumevanju strukture datotek FASTQ, poravnavi sekvenčnih branj na referenčni humani genom ter osnovnim bioinformatičnim pristopom pri analizi podatkov.

V nadaljevanju je bila predstavljena analiza genomske variante, pri kateri so bili obravnavani postopki obdelave sekvenčnih podatkov z uporabo priporočil GATK, funkcionalna anotacija variant z orodjem ANNOVAR ter filtriranje in interpretacija rezultatov. Posebna pozornost je bila namenjena klasifikaciji genetskih različic po smernicah ACMG in podatkovni zbirki ClinVar, kar je omogočilo boljše razumevanje ocenjevanja patogenosti genetskih sprememb.

Pomemben del usposabljanja je predstavljala tudi analiza RNA-seq podatkov v okolju R. Predstavljena je bila uporaba paketov Rsubread, edgeR in Voom za procesiranje, normalizacijo in analizo diferencialnega izražanja genov. Obravnavane so bile tudi metode funkcionalne interpretacije rezultatov s pomočjo analize genske ontologije. Skozi izvedene aktivnosti je bilo pridobljeno razumevanje priprave knjižnic za sekvenciranje, osnov delovanja sekvenatorja, bioinformatične obdelave genomske in transkriptomskih podatkov ter interpretacije rezultatov analiz.

APRIL 2026

Aprila je predaja znanj obsegala področje analize somatskega sekvenciranja ter klicanja različic na RNA-seq podatkih. Na primeru raka dojke je bila prikazana celotna priprava sekvenčnih podatkov za nadaljnjo analizo, vključno z odstranjevanjem adapterskih sekvenc, uporabo ukazne vrstice Bash, pripravo referenčnega genoma GRCh38 ter poravnavo sekvenčnih branj. Predstavljeni so bili različni

pristopi k somatskim analizam, vključno z uporabo prostocelične DNA ter primerjavo tumorskega in zdravega tkiva.

V nadaljevanju so bile obravnavane naprednejše faze obdelave podatkov, kot so pretvorba datotek SAM v BAM, označevanje parnih branj in PCR duplikatov ter rekaliibracija kakovosti baznih odčitkov. Delo je potekalo na realnih podatkih raka dojke, kar je omogočilo vpogled v praktično uporabo bioinformatičnih metod pri analizi tumorskih genomov.

Poseben poudarek je bil namenjen uporabi algoritma Mutect2 za identifikacijo somatskih različic. Predstavljena je bila priprava podpornih referenčnih virov, vključno s panelom normalnih vzorcev in podatki iz baze dbSNP, ter obravnavani koncepti modeliranja napak, »pile-up« analiz in zaznavanja kontaminacije z zarodnimi različicami. Na koncu je bila izvedena funkcionalna anotacija zaznanih različic z orodjem Funcotator.

Poleg tega je bilo izvedeno tudi klicanje različic na RNA-seq podatkih z uporabo algoritma HaplotypeCaller. Obravnavane so bile posebnosti analize RNA-seq podatkov, normalizacija in leva poravnava različic ter interpretacija rezultatov na primeru dveh komercialno dostopnih celičnih linij. Aktivnosti so omogočile poglobljeno razumevanje postopkov somatskega sekvenciranja, samostojnejše delo v ukazni vrstici ter uporabo sodobnih bioinformatičnih pristopov pri analizi genomskih podatkov.

MAJ 2026

V mesecu maju je predaja znanj obsegala področje genotipizacijskih mikromrež in njihove uporabe v genetski diagnostiki. Predstavljeni so bili principi delovanja mikromrež ter njihova uporaba pri odkrivanju genetskih različic v diagnostičnem in raziskovalnem okolju. Udeleženec se je seznanil z delovanjem sistema Illumina iScan, laboratorijskimi postopki priprave vzorcev ter osnovnimi načeli bioinformatične obdelave pridobljenih podatkov.

Poseben poudarek je bil namenjen uporabi programskih orodij GenomeStudio in PLINK za analizo genotipizacijskih podatkov. Obravnavana je bila struktura standardnih datotečnih formatov (.map, .ped, .fam, .bim in .bed) ter njihov pomen pri nadaljnjih genetskih analizah. Na primeru prostodostopnih podatkov je bila prikazana praktična uporaba genotipizacijskih mikromrež za identifikacijo genetskih različic.

Izvedene aktivnosti so omogočile pridobitev znanj o principih delovanja genotipizacijskih mikromrež, razumevanje laboratorijskih postopkov genotipizacije na sistemu Illumina iScan ter osnovno usposobljenost za bioinformatično obdelavo in interpretacijo genotipizacijskih podatkov z uporabo programskih orodij GenomeStudio in PLINK.

