

SVEUČILIŠTE U RIJECI

TEHNIČKI FAKULTET

Diplomski sveučilišni studij računarstva

Evolucijsko računarstvo

TRANSFORMER MODEL ZA PREDVIĐANJE HLA PEPTIDA

1. Kontrolna točka

Rijeka, 05.2024.

Luka Buršić

Patrik Paunović

Matteo Plišić

1. UVOD

Naš se projekt bavi replikacijom studije o povezivanju peptida s HLA molekulama, radom s drugim datasetovima kao što su antimikrobni i antiviralni te proučavanje njihovog povezivanja s peptidima. Prvi je korak replikacija modela za predviđanje povezivanja peptida sa sustavom leukocitnog antigena, nakon čega ćemo isti model pokušati primijeniti na slične skupove podataka. Dodatno, radit ćemo sa peptidima duljine 8-14 aminokiselina koji imaju do 4 mutacije.

2. METODE RJEŠAVANJA ZADATKA

U radu je korišteno 112 HLA molekula, i svaka od tih molekula ima drugačiji raspon duljina peptida s kojim se mogu povezati, dakle ukupno ima 366 mogućih kombinacija. U radu je za svaku kombinaciju uzeto 10 primjera, odnosno ukupno 3660. Preko Api-a od IEDB probali smo za svaku kombinaciju dobiti 10 parova koji su eksperimentalno pokazani da su negativni. No, za neke kombinacije nije bilo 10 primjera pa zbog toga imamo 1696 primjera. Ti se primjeri klasificiraju pomoću TransPHLA modela, čime se dobiju *true negative* i *false positive* uzorci. Za dobivene *true negative* uzorke se mutacije vrše pomoću AOMP modela te se uspješne mutacije provjeravaju pomoću IEDB api-a.

3. REZULTATI

Korišteni su podaci peptida koji su već dostupni u samoj studiji[1], uzeti su uzorci koji su samo usitinu pozitivni (*true positive*). Također, u daljnjim koracima biti će korišteni antiviralni i antimikrobni skupovi podataka.

Svih 1696 je obrađeno pomoću TransPHLA modela. Od njih 1696, model je njih 965 uspješno klasificirao kao negativne, što nam daje uspješnost od 56.9%. Generirani su i *heatmapovi* koji prikazuju koje aminokiseline na kojim pozicijama imaju najveći utjecaj na to hoće li par biti klasificiran kao negativan ili pozitivan

Program je odradio 248 *true negative* parova gdje je za svaki peptid odrađen određen broj mutacija (od 1 do 4 mutacija) te je za svaki mutirani peptid dobivena y_{prob} vrijednost koja označava vjerojatnost povezivanja tog mutiranog peptida s HLA. Ukoliko je ona veća od 0.5, y_{pred} poprima vrijednost 1 što znači uspješnost povezivanja. Kako smo nakon izvršavanja

prethodnog koraka dobili 248 csv datoteka, iz svake smo uzeli po jedan par sa što manjim brojem mutacija, gdje je y_{pred} rezultat barem 0.5, te spremili u novu csv datoteku.

Za svaki od tih 248 parova smo odredili koliki je najmanji broj mutacija potreban da se dobije par kojeg TransPHLA klasificira kao pozitivnog. Od ukupno 248, njih 197 je trebalo 1 mutaciju, 38 parova je trebalo 2 mutacije, a 13 parova je trebalo 3 mutacije, čime se pokazuje da je homolognost mutiranih peptida vrlo visoka.

Konačno, dobivene uspješne mutacije smo usporedili s onima iz IEDB te smo time uspjeli potvrditi 234 mutirana para, što znači da je točnost AOMP modela bila 94.35%.

4. ZAKLJUČAK

Uspješnost TransPHLA modela u našem projektu je značajno manja od uspješnosti dobivene u znanstvenom radu. To se može opravdati time što smo mi koristili uzorke koji su eksperimentalno pokazani kao negativni, dok su se u znanstvenom radu koristili uzorci koji su pomoću NetMHCpan_ba metode klasificirani kao negativni.

Zbog ograničenosti naših laptopa, uspjeli smo obraditi sa 248 od 965 *true negative* uzoraka pomoću AOMP modela. Većina uzoraka ima samo 1 mutaciju, dok je 2, 3 i 4 mutacije značajno manje. Za daljnji rad projekta uzeti su uzorci peptida s najvećim „Score-om“, a koji je minimalno bio 0.5.

5. LITERATURA

[1] A transformer-based model to predict peptide_HLA class I binding and optimize mutated_peptides for vaccine design.