# A bioinformatika gyakorlati beadandó feladat leírása

*3 fős csoport által benyújtott 3-6 oldalas magyar nyelven írt dolgozat*

*Beadás:* 2019. január 7, hétfőig **pdf** mellékletként elküldve az ELTEbioinfo@gmail.com e-mail címre. (Aki előbb küldi el, annak előbb javítjuk.)

*Téma*: A csoporttagok által a félév elején választott különböző fehérjék bizonyos adatainak és tulajdonságainak összehasonlítása és összegzése.

*Leírás:* A feladatok megoldásának minden részletét írjátok le, a rész- és végeredményeket is illesszétek bele a dokumentumba. Kommentáljátok, magyarázzátok, hogy mit miért csináltatok és mi a véleményetek az eredményekről. A beadott dokumentumnak mindenfajta egyéb forrás nélkül is képesnek kell lennie bemutatni az eredményeket és megmagyarázni, hogy mit miért csináltatok.

*Csoport:* A csoportokat három ember alkotja, függetlenül attól, hogy melyik gyakorlati csoportba tartoznak. A csoportot olyan emberek alkossák, akik különböző fehérjéket választottak. A benyújtott dolgozatba írjátok bele a csoporttagok nevét és csoportját.

***Feladatok***:

1. A választott fehérjék bemutatása az Uniprot adatbázisban található legfontosabb adatok szerint. Milyen közös jellemzői vannak a csoporttagok választott fehérjéinek az Uniprot adatbázis és az ugyanitt található GeneOntology leírás alapján? (A GeneOntology adatokat talán könnyebb az Enbembl BioMart segítségével leszedni és összehasonlítani.)
2. Készítsetek többségi szabály (*majority rule* és nem *extended majority rule*) konszenzus fát[[1]](#footnote-1) a fehérjék illesztett mRNS szekvenciái alapján kapott maximum-likelihood fákból (fajok: ember, csimpánz, kutya, szarvasmarha, egér, oposszum). (A fákat számolhatjátok az egész mRNS szekvenciák alapján, vagy a rövidebbre vágott illesztésekből. Elvileg nem kell a fákat újból kiszámolni, ezeket készítettétek el a 6. gyakorlaton.) Milyen következtetést tudtok levonni a konszenzus fa alapján? Mi a különbség az eredeti fák és a konszenzus fa között? Illesszétek be a fák képeit. Ügyeljetek az olvasható betűméretre és arra, hogy a konszenzus fán látszódjon az elágazások alátámasztottsága.
3. Tervezettek primereket mindhárom génetekre UGENE primer3 program segítségével. A következő beállításokat alkalmazzátok: a termék legyen 50-200 bázispár nagyságú, a program adjon vissza 5 terméket, a maximum komplementaritás legyen 4.00, a maximum 3’ komplementaritás legyen 3.00, és maximum 4 hosszú szakasz lehessen azonos bázisból. A target a 40. bázitól kezdődjön és 78 bp-t öleljen fel. Írjátok le mennyi primert találtatok az egyes génekre, és ezek mennyire voltak specifikusak más online elérhető adatbázisok (NCBI) alapján? Készítsetek screen shot-ot a primerekről a UGENE-ban úgy, hogy a legfontosabb információk látszódjanak. A képeket és a válaszokat illesszétek be a jegyzőkönyvetekbe. Előfordulhat, hogy valamelyik génhez nem lehet primert tervezni, ezt az esetet is írjátok le.
4. Cytoscape programmal töltsétek le mindhárom fehérje első-szomszédainak hálózatát két publikus adatbázisból. Egyesítsétek a hálózatokat. Távolítsátok el a nem emberi fehérjéket és minden szubgráfot ami nem kapcsolódik a legnagyobb komponenshez. Készítsetek a hálózatról olyan képet, amelyen jól elkülönül a három fehérje (eltér a többi ponttól) és jól olvashatóak a neveik (csak ennek a háromnak). Jelöljétek valamilyen vizuális tulajdonsággal az adott fehérje első-szomszédait valamit a közös első-szomszédokat. Az így elkészült ábrát illesszétek be a beadandó dokumentumba és készítsetek hozzá ábra magyarázatot.
5. Keressétek meg a GEO2R adatbázisban (https://www.ncbi.nlm.nih.gov/geo/geo2r/) a GSE21510 GEO azonosítójú adatszettet, majd készítsetek két csoportot (Stage1\_2) a Stage oszlop alapján. A csoportokba csak azok a minták kerüljenek be, amelyek rákos szövetekből származtak. Mentsétek le a táblázatot, majd ismételjétek meg a műveletet a másik két Stage csoporttal (Stage3\_4). Importáljátok be ezeket a táblázatokat egy ensembl Biomart-ról letöltött Affy HG U133-PLUS-2 probeset-et tartalmazó táblázattal együtt a Cytoscape-be, a három fehérjétek első szomszédait tartalmazó session-höz lásd 4. feladat. Jelenítsétek meg az expressziós adatok (logFC) változásait a Stage-ek (Stage1\_2 és Stage3\_4) között, (használjatok a (–) értékekre piros színezést, a (+) értékekre pedig zöldet) egy rövid, 10 másodperc hosszú animáció segítségével, amelyet a CyAnimator-ral készítsetek el. Az animációban legyen benne legalább 3 kulcs képkocka. Az animációt videó formába exportáljátok és töltsétek fel YouTube-ra. A videó linkjét illesszétek be a dokumentumba.

\*2. feladat kiegészítés:

Mivel a UGENE programmal ez nem megvalósítható, ezért mentsétek ki a fákat newick formátumban, másoljátok a 3 fát egy fileba, majd a <http://bioinfo.nhri.org.tw/cgi-bin/emboss/fconsense> oldalon olvassátok be a filet. Készítsétek el a konszenzust, mentsétek le a outtreefile-t, majd Figtreevel annak a képét.

2018.12.03.

1. [↑](#footnote-ref-1)