

A bioinformatika gyakorlati beadandó feladat leírása

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

Beadás: 2018. január 2, keddig 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 Ensembl BioMart segítségével leszedni és összehasonlítani.)
2. Készítsetek többségi szabály (*majority rule* és nem *extended majority rule*) konszenzus fát¹ 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. Ügyeljenek az olvasható betűméretre és arra, hogy a konszenzus fán látszódjon az elágazások alátámasztottsága.
3. Keressetek két HG-U133_Plus_2 Affimetrix platformon (*platform ID* vagy *reporter*: GPL570) végzett génexpressziós vizsgálatot a [GEO](#) adatbázisban. Írjátok le, hogy miről szólnak vizsgálatok, milyen csoportokat hasonlítottak össze. A differenciál expressziós vizsgálatban milyen korrigált *p*-érték és milyen logFC jött ki a három fehérjétek géneire? Szignifikáns a különbség? Tegyétek be a gének expressziós profiljáról készült képet.
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. Készítsetek egy rövid animációt a [CyAnimator](#) segítségével, melyben a 4. feladatban létrehozott hálózaton megjelenítitek a 3. feladat során a GEO2R-ből letöltött differenciál expressziós eredmények logFC értékeit. Előbb az egyik vizsgálat logFC értékeit, majd a

¹ Mivel a UGENE programmal ez nem megvalósítható, ezért mentsetek 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, mentsetek le a outtreefile-t, majd Figtreevel annak a képét.

másikét. Az animáció legyen legalább 10 másodperc hosszú és legyen benne legalább 3 kulcs képkocka. Az animációt videó formába exportáljátok és töltsétek fel YouTube-ra. A videó linkjét illesszétek be a dokumentumba.

2017.II.30.