

Biológiai „Big Data” – Transzkripció zaj elemzése nagy mennyiségű génexpressziós adaton

TDK vagy szakdolgozati téma a Semmelweis Egyetem I. sz. Patológiai es Kísérleti Rákkutató Intézetének Bioinformatika csoportjában, Dr. Sebestyén Endre vezetésével.

Az elmúlt 10 évben a különféle „high throughput” módszereket alkalmazó daganat genom és transzkriptom szekvenáló projektek hatalmas mennyiségű adatot generáltak. Emellett kisebb kutatási projektek is rendszeresen több száz DNS vagy RNS mintát szekvenálnak, amelyek nagyrésze publikusan elérhető. Ezen adatokat felhasználva és új statisztikai módszereket alkalmazva rengeteg, általában nem vizsgált molekuláris folyamat részletes elemzésére nyílik lehetőség. Ezek közé tartozik amikor különféle kondíciók, például normál és tumoros minták között nem az átlagos génexpressziós eltérést és azok okait, hanem a génexpresszió változó varianciáját, vagy komplex génszabályozó hálózatok hibáit vizsgáljuk. Másrészt az is elemezhető, hogy a humán genom egyébként nem funkcionális, „evolúciós hulladék” szekvenciái milyen expressziós mintázatokat mutatnak különböző betegségekben. A kutatás során ezeket az aberráns expressziós folyamatokat szeretnénk részletesebben feltárni nagy mennyiségű publikus adat segítségével. Vizsgáljuk az expressziós változások kapcsolatát különféle daganatok, autoimmun és neurológiai betegségek kialakulásával, fejlődésével is.

A téma, valamint informatika és molekuláris biológia iránt egyaránt érdeklődő hallgatókat keresünk a Semmelweis Egyetem I. sz. Patológiai es Kísérleti Rákkutató Intézetének Bioinformatika csoportjába minimum 1 évre. A munka folyamán lehetőség lesz különböző bioinformatikai módszerek és kutatási területek megismerésére. Elsősorban R, alkalmanként Python vagy Perl programozási nyelvet használva megismerhetők a legújabb, genomikában és transzkriptomikában használt statisztikai és gépi tanulási módszerek, programcsomagok. Biológiai szempontból részletes ismeretek szerezhetők a transzkripció és poszt-transzkripció szabályozásról, genomszerkezetről, mutációs mechanizmusokról és különböző betegségek molekuláris hátteréről. Motivációtól függően később pályázatokban, PhD képzésben, projektmenedzselésben való részvételre is van lehetőség. A jelentkezéshez egy rövid (1-2 bekezdés) motivációs levelet és önéletrajtot (maximum 1 oldal) kérünk a következő címre: endre.sebestyen@gmail.com

Szükséges előismeretek: szakirodalom feldolgozásához elegendő angol nyelvtudás, alapvető informatikai és molekuláris biológiai tudás. Előnyt jelent (de nem szükséges) a Linux operációs rendszer, valamilyen script nyelv vagy statisztikai módszer ismerete.

További információ: <https://su-compbio.github.io>

Dr. Sebestyén Endre

Semmelweis Egyetem I. sz. Patológiai es Kísérleti Rákkutató Intézet