

A MAGYAR
TUDOMÁNY
ÜNNEPE

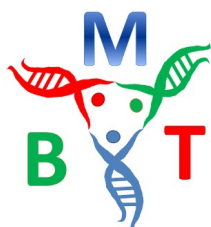


Magyar Tudományos
Akadémia



Bioinformatika 2024

A Magyar Tudományos Akadémia Bioinformatikai Osztályközi Tudományos Bizottsága és a Magyar Bioinformatikai Társaság tudományos konferenciája



2024. november 8.

Természettudományi Kutatóközpont
1117 Budapest Magyar Tudósok krt 2.
Földszint nagyterem

Program

1. szekció - Bioinformatikai algoritmusok 1

Elnök: Dr. Ari Eszter

10:00 - 10:05: **Köszöntő (Prof. Pongor Sándor)**

10:05 - 10:20: **Gerber Dániel**

Bioinformatikai alkalmazások és kihívások az archeogenomikában
MTA-BTK Lendület Biorégészet Kutatócsoport; HUN-REN BTK Archeogenomikai Intézet

10:20 - 10:35: **Bálint Balázs**

Annotált genomok szennyeződéseinek érzékeny kimutatása és eltávolítása a ContScout programmal
HUN-REN SZBK Biokémiai Intézet, Gomba Genomika és Evolúció Csoport

10:35 - 10:50: **Dr Ari Eszter**

mulea - egy többféle ontológiát és empirikus FDR-t alkalmazó dúsulást vizsgáló R csomag
HUN-REN Szegedi Biológiai Kutatóközpont, Biokémiai Intézet; Eötvös Loránd Tudományegyetem, Genetikai Tanszék

10:50 - 11:05: **Hegedűs Dániel és Juhász Márk Hunor**

Filogenetikai fák metszése
ELTE TTK, Matematikai Intézet; ELTE TTK, Biológiai Intézet, Genetika Tanszék

11:05 - 11:20: **Juhász Judit**

Nukleotid szekvenciák kontextusának vizsgálata pángenom és assembly gráfokban valószínűség alapú módszerekkel
Pázmány Péter Katolikus Egyetem, Információs Technológiai és Bionikai Kar

11:20 - 11:35: **Szünet**

2. szekció - Bioinformatikai algoritmusok 2

Elnök: Dr. Hegedűs Tamás

11:35 - 11:50: **Krizsán Dániel**

Génexpresszió előrejelzése Escherichia coli baktériumban genomikai nyelvmodellek segítségével
Pázmány Péter Katolikus Egyetem, Információs Technológiai és Bionikai Kar

11:50 - 12:05: **Nagy Ádám**

Agyhullámok és Mentális Zavarok: A Legkorábbi Jelek Feltárása
Szegedi Tudományegyetem, Szoftverfejlesztési Tanszék

12:05 - 12:20: **Pogány Domonkos**

Metabolit-betegség asszociációk becslése hiperbolikus terekben
BME Mesterséges Intelligencia és Rendszertervezés Tanszéken, Mesterséges Intelligencia Kutatócsoport

12:20 - 12:35: **Stipping Marcel**

Jellemzők terének szűrése utranagy dimenziójú többosztályos adatokon
HUN-REN Wigner Fizikai Kutatóközpont, Részecske- és Magfizikai Intézet, Komputációs Tudományok Osztálya

12:35 - 12:50: **Zsichla Levente**

Nyílt forráskódú vírusgenom-összeszerelő pipeline-ok összehasonlító elemzése
Biológiai Intézet, Eötvös Loránd Tudományegyetem; Egészségbiztonság Nemzeti Laboratórium, Eötvös Loránd Tudományegyetem

12:50 - 13:30: **Szünet**

3. szekció – Rendszerbiológia, Szerkezeti Bioinformatika, Orvosi Bioinformatika

Elnök: Dr. Kerepesi Csaba

13:30 - 13:45: **Csendes Gerold**

Poszt-perturbációs génexpressziót prediktáló gépi tanulási modellek kiértékelése
Turbine Kft

13:45 - 14:00: **Weber Áron**

Hogyan fedez fel egy sejtszimuláció új fehérjekomplexeket?
Pázmány Péter Katolikus Egyetem Információs Technológiai és Bionikai Kar

14:00 - 14:15: **Bruncsis Bence**

Új szimulációs keretrendszer multivalens fehérjék és membránfehérjék közötti kötődés vizsgálatára
BME Mesterséges Intelligencia és Rendszertervezés Tanszéken, Mesterséges Intelligencia Kutatócsoport

14:15 - 14:30: **Kálmán Zsófia**

Kölcsönhatások elemzése a PSD-ben: a GKAP-LC8 példája
Pázmány Péter Katolikus Egyetem Információs Technológiai és Bionikai Kar

14:30 – 14:45: **Tóth Adrienn Gréta**

Klinikai megfontolások az antimikrobiális rezisztenciapotenciálhoz
Bioinformatikai központ, Állatorvostudományi Egyetem

14:45 – 15:00: **Nagy Sára**

Metagenomikai bepillantás egy gabonamúmiába.
Komplex rendszerek fizikája tanszék, Eötvös Loránd Tudományegyetem

15:00 - 15:15: **Szünet**

4. szekció – Orvosi bioinformatika

Elnök: Prof. Gyórfy Balázs

15:15 - 15:30: **Stirling Tamás**

Genomi monitoring és fágterápia összekapcsolása az antibiotikum-rezisztens kórokozók elleni küzdelemben
HUN-REN Szegedi Biológiai Kutatóközpont; Szegedi Tudományegyetem, Biológia Doktori Iskola

15:30 - 15:45: **Dr. Buza Krisztián**

Továbbfejlesztett ROCKET algoritmus tónusos-klónusos epilepsziás rohamok felismerésére
Sapientia Erdélyi Magyar Tudományegyetem, Matematika-Informatika Tanszék, Biointelligencia Kutatócsoport; Budapesti Gazdasági Egyetem, Gazdaságinformatika Tanszék

15:45 - 16:00: **Szatmári Zita Eszter**

A sport hatása a biológiai életkor változására
HUN-REN Számítástechnikai és Automatizálási Kutatóintézet

16:00 - 16:15: **Kerepesi Csaba**

A legújabb öregedés órák és érdekes alkalmazásaik
HUN-REN Számítástechnikai és Automatizálási Kutatóintézet

16:15 – 16:30: **Müller Dalma**

Epigenetikai altípusok és génszintű metilációs változások vizsgálata vastagbél adenokarcinómában

Semmelweis Egyetem, Bioinformatika Tanszék

16:30 - 16:45: **Nagy Gergely**

A makrofág promóterek összetételének vizsgálata

Debreceni Egyetem, Általános Orvostudományi Kar, Biokémiai és Molekuláris Biológiai Intézet

16:45 - 16:50: **Zárszó (Prof. Győrffy Balázs)**

Előadások kivonatai

1. szekció - Bioinformatikai algoritmusok 1

Bioinformatikai alkalmazások és kihívások az archeogenomikában

Gerber Dániel¹, Jakab Kristóf¹, Mező György² és Szécsényi-Nagy Anna¹

¹HUN-REN BTK Archeogenomikai Intézet

²HUN-REN CSFK, HUN-REN Wigner Fizikai Kutatóközpont

Az archeogenomika az utóbbi évtizedben elért a teljes ősi genomok vizsgálatának szintjére. A szakterületre specifikus, az ősi DNS-t ért károsodások miatt fennálló kihívások leküzdése mellett, sikerrel adaptál és alkalmaz számos modern populációkra kidolgozott bioinformatikai és genetikai eljárást. Az előadás a munkacsoport aktuális kutatási eredményeit példának hozva bemutatja a HUN-REN BTK Archeogenomikai Intézetében jelenleg folyó genomikai elemzések kihívásait és sikereit, ezek között kiemelt figyelmet szentelve a leszármazási alapon megegyező kromoszóma szegmensek vizsgálati mélységeinek. Az informatikai fejlesztések a HUN-REN Cloud, Wigner DC és az MTA-BTK Lendület Biorégészeti Program támogatásával valósulnak meg.

Annotált genomok szennyeződéseinek érzékeny kimutatása és eltávolítása a ContScout programmal

Bálint Balázs¹, Merényi Zsolt¹, Hegedűs Botond¹, Igor V. Grigoriev^{2,3}, Zhihao Hou^{1,4}, Földi Csenge^{1,4} és Nagy G. László¹

¹HUN-REN SZBK Biokémiai Intézet, Gomba Genomika és Evolúció Csoport

²US Department of Energy Joint Genome Institute, Lawrence Berkeley National Laboratory

³Department of Plant and Microbial Biology, University of California Berkeley

⁴SZTE TTIK, Biológia Doktori Iskola

A genomi adatok idegen szekvenciákkal szennyezettsége egyre inkább felismert probléma, amely jelentősen befolyásolhatja a későbbi adatelemzési és adatértelmezési lépéseket. Előadásunkban a ContScout programot mutatjuk be, ami nagy pontossággal távolítja el a szennyező szekvenciákat az annotált genomokból. Szintetikus tesztek révén bemutatjuk, hogy az eszköz kiváló érzékenységgel és specifitással működik még akkor is, ha a szennyezés forrása egy közeli rokon faj. Ember által annotált referencia adatokon demonstráljuk, hogy a ContScout pontosságban jelentősen felülmúlja a legtöbb versenytárs eszközt, valamint bizonyítjuk, hogy képes megkülönböztetni a horizontális génátvitelt a szennyeződéstől. Nyilvános adatbázisokból letöltött 844 eukarióta genom ContScout vizsgálata során jelentős mértékű adatbázis szennyeződést azonosítunk, melyek eredete túlnyomórészt bakteriális. Bemutatjuk, hogy a szennyezett adatokon végzett ősi genom rekonstrukciók irreálisan korai gén-eredeteket feltételeznek, amivel a ténylegesnél komplexebb ősi genomokat eredményeznek, a leszármazottakban pedig túlbecslik a génvesztések számát.

A ContScout-tal egy pontos és kényelmes eszközt biztosítunk a genomi szennyeződések azonosítására és eltávolítására az annotált genomokból. A forráskód elérhető a GitHubon ([h836472/ContScout](https://github.com/h836472/ContScout)), míg a Docker-konténer a `docker://h836472/contscout` címről tölthető le. További részletek a közleményben olvashatók: Nat. Commun., DOI:10.1038/s41467-024-45024-5.

mulea - egy többféle ontológiát és empirikus FDR-t alkalmazó dúsulást vizsgáló R csomag

Ari Eszter^{1,2}, Cezary Turek¹, Ölbei Márton¹, Stirling Tamás¹, Fekete Gergely¹, Tasnádi Ervin¹, Gul Leila¹, Bohár Balázs¹, Papp Balázs¹ és Jurkowski Wiktor¹

¹HUN-REN Szegedi Biológiai Kutatóközpont, Biokémiai Intézet

²Eötvös Loránd Tudományegyetem, Genetikai Tanszék

A hagyományos génszett dúsítási elemzések többnyire néhány ontológiára korlátozódnak, és nem veszik figyelembe a génszettek vagy ontológia elemek közötti összefüggéseket, így túlkorrigálják a p-értékeket. Ezen két problémát kívánjuk megoldani a mulea nevű R csomaggal, amely átfogó funkcionális dúsulási elemzéseket kínál. A mulea egy empirikus false discovery rate (eFDR) algoritmust alkalmaz, amelyet kifejezetten az egymástól nem független biológiai adatokra terveztünk. A hagyományos p-érték korrekciós eljárásokkal szemben, az eFDR-t alkalmazva pontosabban azonosíthatjuk a felülreprezentált ontológia kategóriákat. Az eFDR alkalmazása mellett, a mulea-ba 27 modellorganizmusból számos ontológia típust integráltunk, beleértve a Gene Ontology-t, útvonalakat, szabályozó elemeket, genomi régiókat és fehérjedoméneket. Ezáltal lehetővé válik a kutatók számára, hogy az elemzést saját tudományos kérdéseikhez igazítsák. Az ontológiák elérhetőek standardizált GMT formátumban, valamint a muleaData ExperimentData Bioconductor csomag segítségével is. A mulea külső forrásokból származó GMT fájlok integrálását is lehetővé teszi a felhasználók számára, ezzel bővítve annak alkalmazhatóságát különböző kutatási területeken. A mulea elérhető CRAN R csomagként: <https://cran.r-project.org/web/packages/mulea/> és a GitHub-on is: <https://github.com/ELTEbioinformatics/mulea> is.

Filogenetikai fák metszése

Hegedűs Dániel¹, Juhász Márk Hunor¹ és Ari Eszter²

¹ELTE TTK, Matematikai Intézet

²ELTE TTK, Biológiai Intézet, Genetika Tanszék

A filogenetikai fák készítéséhez használt DNS- vagy fehérje-szekvenciaadatok sok hibát tartalmazhatnak, ezeknek okai lehetnek például DNS szennyeződések, rossz minőségű genom összerakás vagy a taxon téves besorolása. Bár ezek sokszor már DNS szinten azonosíthatók, a fel nem ismert hibák gyakran túl hosszú ágakat eredményeznek a filogenetikai fán. Az ilyen hibás ágak felismerése és lemetszése kulcsfontosságú a kapott fa evolúciós értelmezése szempontjából. Bár a hibás szekvenciák száma a növekvő szekvenált genomok és egyéb DNS-adatok számával együtt növekszik, mostanáig csupán a TreeShrink nevű algoritmus és szoftver létezett publikáltan a probléma megoldására. Azonban a TreeShrink néhány esetben nem képes a szemmel láthatóan túl hosszú ágakat levágni a fáról. Az előadás során két új metszési algoritmust mutatunk be: a "Primitive Straight-Forward Approach"-nak (PSFA) elnevezett egyszerű algoritmust, amely csak a leghosszabb ágakat távolítja el, és a "Circular Pruning Algorithm" (CPA) névre hallgató komplexebb algoritmust, amely a fákat nagyjából kör alakúra metszi. Ezeket a módszereket kombináljuk is, és eredményeiket összehasonlítjuk a TreeShrink szoftver eredményeivel. A tesztelést 25 különböző fán végeztük úgy, hogy dátum alapú gyökereztetés után kiszámoltuk az egyes fák R értékeit. A tesztek jelentős részében nagy javulás tapasztalható ezen értékben az eredeti fához képest.

Nukleotid szekvenciák kontextusának vizsgálata pángenom és assembly gráfokban valószínűség alapú módszerekkel

Juhász Judit¹, Réti Márton¹ és Dr. Ligeti Balázs¹

¹Pázmány Péter Katolikus Egyetem, Információs Technológiai és Bionikai Kar

A tradicionális megközelítésben a DNS-t lineáris szöveggént ábrázoljuk. Ez a leírás azonban problémákat okozhat, amikor nagyobb csoportok, telepek, fajok vagy éppen jól ismert patogének, mint pl. Klebsiella genetikai információit elemezzük. Egy lehetséges alternatív ábrázolási mód a pángenom vagy szekvencia gráfok lehetnek, melyek kompakt formában tartalmazzák az egyedek közötti hasonlóságokat és különbségeket is. A kutatás célja egy sketching alapú algoritmus kifejlesztése, amely hasonlósági alapú keresést végez pángenom vagy assembly gráfokon.

Az algoritmus modern, valószínűség alapú módszerekkel dolgozik, melyek kifejezetten alkalmasak a szekvenciák közötti hasonlóság mértékének leírására, valamint az időkomplexitásuk is megfelelő.

A kifejlesztett algoritmust a long-read illesztés megoldására létrehozott eszközökkel hasonlítottuk össze. Az első tesztek alapján az algoritmus futásidő tekintetében teljesít a legjobban, habár jelenleg prototipizálási fázisban van és Pythonban lett implementálva. Továbbá a kontextusfüggetlensége miatt olyan mutációk vagy összeillesztési hibák esetén is megtalálja a keresett szekvenciát, amely helyzetekben a kontextusfüggő megoldások már nem vezetnek eredményre.

A módszert továbbá validáltuk nem csak in-silico hanem kísérleti eredményekkel is. Azt találtuk, hogy az algoritmus sikeresen azonosította a bakteriális kromoszómába beépült fágokat. Ez az eredmény jól bizonyítja a gráfbeli kontextusok kiaknázásának fontosságát.

2. szekció - Bioinformatikai algoritmusok 2

Génexpresszió előrejelzése Escherichia coli baktériumban genomikai nyelvmodellek segítségével

Krizsán Dániel¹, Réti Márton¹ és Dr. Ligeti Balázs¹

¹Pázmány Péter Katolikus Egyetem, Információs Technológiai és Bionikai Kar

A mikroorganizmusok alapvető szerepet játszanak a bioszférában, és számos ipari, biotechnológiai és környezeti alkalmazásuk ismert. A mikroorganizmusok metabolikus folyamatairól többek között génexpressziós mérésekkel és mintázat elemzéssel kaphatunk pontosabb képet, amikhez költséges kísérletekkel kellenek. A genomikai nyelvmodellek azonban ígéretesek lehetnek a probléma megoldásában, mivel képesek figyelembe venni a nemkódoló szabályozási elemeket és a genomi kontextust.

A kiindulópont a prokBERT genomikai nyelvmodell expresszióra specializált változatának fejlesztése. Kutatásunk középpontjában az E. coli, mint modellorganizmus áll. Régóta kutatott és nagy mennyiségben állnak rendelkezésre róla publikus adatok, így alkalmas kiindulópont a modell fejlesztéséhez.

Feldolgoztuk az összes Sequence Read Archive-ban elérhető E. coli transzkriptom adatot (pair-end, Illumina, 4 colour chemistry). Ez 7216 SRA azonosítóhoz tartozó, nagyságrendileg összesen 18 terrabájtnyi E. coli RNA-seq szekvencia letöltését és tisztítását jelentette. A keletkezett adatbázis alkalmas a célul kitűzött genomikai nyelvmodell betanítására. A számításaimat a KIFÜ szuperszámítógépén, a Komondoron végeztem, NextFlow pipeline-t építve a skálázhatóság érdekében.

Lehetővé vált a genomikai nyelvmodellek alkalmazása, ezeken keresztül a tágabb genomi kontextus figyelembe vételére épülő predikciós eszközök fejlesztése. Ezáltal közelebb kerültünk a mikrobiomok pontosabb megértéséhez és tervezésükhöz is.

Agyhullámok és Mentális Zavarok: A Legkorábbi Jelek Feltárása

Nagy Ádám¹, Gubics Flórián², Dombi József¹, Szabó Zoltán¹, Viharos Zsolt János^{3,4}, Anh Tuan Hoang³, Bilicki Vilmos¹ és Szendi István¹

¹Szegedi Tudományegyetem, Szoftverfejlesztési Tanszék

²Szegedi Tudományegyetem, Orvosi Genetikai Intézet, Interdiszciplináris Doktori Képzés

³HUN-REN Európa Kiválósági Központja az információ-technológia és az automatizálás területén (EPIC), MTA

⁴Neumann János Egyetem - Pénzügyi és Gazdasági Kar

A szkizofrénia és bipoláris zavar olyan mentális betegségek, amelyek élettartam-prevalenciája a népesség körében 5-6%. Mindkét betegség súlyos negatív hatást gyakorol egyéni és társadalmi szinten. A tünetek változatossága és átfedése, valamint az elavult diagnosztikai módszerek miatt a diagnózis és a kezelés nehézkes, nem elég hatékony, és jelentős erőforrásokat igényel. Mindkét betegség az idegrendszeri fejlődést érinti, így a korai, még tünetmentes, úgynevezett "prodrómás" állapot felismerése felbecsülhetetlen jelentőséggel bírhat a korai kezelés megkezdése, vagy akár a prevenció szempontjából. Ezáltal a betegek életszínvonala hosszú távon megőrizhető, és állapotuk nem romlik folyamatosan.

Kutatásunk célja olyan biomarkerek azonosítása, amelyek segítségével még a súlyos tünetek megjelenése előtt felismerhető a veszélyeztetett populáció. A kutatás alanyai olyan egyetemisták, akik egészségesek, de szkizotípiá vagy bipolaritás felmérő tesztek alapján a szórásban helyezkednek el, így tünetmentes, prodrómális kockázati csoportba sorolhatók. Ezek mellett a kontrollcsoportot olyan hallgatók alkotják, akik minden teszten negatív eredményt értek el.

A két kockázati és kontrollcsoport körében számos felmérést és adatgyűjtést végeztünk. Eddig aktigráfia és EEG jelek feldolgozásával foglalkoztunk, amelyekből biomarkereket generáltunk, majd innovatív módszertannal vizsgáltuk, melynek alapját tanuló algoritmusok képzik. A legjobb eredményeket EEG biomarkerek felhasználásával értük el: a bipoláris hajlamú csoportot 90%-os, míg a szkizotípiá hajlamú csoportot 80%-os pontossággal sikerült elkülöníteni.

A tanuló algoritmusoknak olyan kihívásokkal kellett szembenézniük, mint a nagy dimenzionalitás és az alacsony mintaszám. Ezt úgy kezeltük, hogy bizonyos metrikák alapján potenciálisan optimális változókat csoportosítottunk, így csökkentve a dimenzionalitást. Az így kialakított csoportokat krosszvalidációval tanítottuk, számos modellt létrehozva. Az alacsony mintaszám torzító hatásainak további minimalizálása érdekében modellmagyarázó metrikákat, úgynevezett Shapley értékeket is számoltunk. Ezek betekintést engednek a modellek belső működésébe, megmutatva, mely biomarkerek milyen szerepet játszanak, akár egymáshoz viszonyítva is. Az eredmények teljesebb értelmezése érdekében aggregáltuk a jól teljesítő algoritmusok Shapley értékeit, csökkentve a véletlenszerű hatásokat és kiemelve az esetleges anomáliákat, ahol a biomarkerek egymásnak ellentmondó szerepeket tölthetnek be.

Az EEG adatok elemzésével mindkét hajlamcsoport esetén sikerült jól elkülöníthető, konzisztens frekvencia- és téreloszlású biomarkereket azonosítani, amelyek összhangban vannak a szakirodalomban leírt kóros agyi elváltozásokkal. Ezen felül a két hajlamcsoport elkülönítése során kiemelkedő, 92%-os pontosságot értünk el, ami még előrehaladott állapotú betegségek esetén is kihívást jelenthet. Korábbi elemzéseinkben az aktigráf adatokkal kevésbé pontos algoritmusokat kaptunk, és néha egymásnak ellentmondó biomarkereket is találtunk. Mind az EEG, mind az aktigráfia további, mélyebb elemzése szükséges, hogy élesebb képet kapjunk a

betegségekről és azok kialakulásáról. Ezenkívül a populációtól származó további adatformákat is elemezzük, mint például a szemmozgás-felvételek, szelfágencia, és biokémiai adatok.

Metabolit-betegség asszociációk becslése hiperbolikus terekben

Pogány Domonkos¹ és Antal Péter¹

¹BME Mesterséges Intelligencia és Rendszertervezés Tanszéken, Mesterséges Intelligencia Kutatócsoport

A gépi tanulás és a nem-euklideszi geometriák ötvözése napjainkban egyre népszerűbb kutatási irányt képvisel. A hiperbolikus geometria ideális fák és heterogén fokszámeloszlású hálózatok modellezésére, mivel az euklideszi térrel ellentétben a negatív görbületű tér exponenciális növekedést mutat.

Bioinformatikai alkalmazásokban gyakran találkozunk ilyen struktúrákkal hierarchiák és skálafüggetlen interakciós hálózatok formájában, így a hiperbolikus beágyazások használata ezen a területen különösen előnyös lehet.

Előadásomban a hiperbolikus geometria főbb tulajdonságait és a bioinformatikában vett alkalmazhatóságát mutatom be. Ez utóbbit egy frissen publikált példán keresztül, egy metabolit-betegség asszociációs hálózat modellezésére alkalmazott módszerrel szemléltetem.

Jellemzők terének szűrése utranagy dimenziójú többosztályos adatokon

Marcell Stippinger¹, Dávid Hanák², Marcell T. Kurucz^{1,3}, Olivér M. Törteli², Ágnes Chripkó², Gergely Hanczár² és Zoltán Somogyvári¹

¹HUN-REN Wigner Fizikai Kutatóközpont, Részecske- és Magfizikai Intézet, Komputációs Tudományok Osztálya

²Cursor Insight Ltd.

³Budapesti Corvinus Egyetem, Adatelemzés és Informatika Intézet

Az időszerelemzés gyakran motívumok és szakértők által meghatározott jellemzők kereséséből áll. Ez a folyamat automatizálható bármely egy- vagy többcsatornás jel kinematikai tulajdonságai alapján több százezer lehetséges jellemző számítógépes generálásával, majd gépi tanulásra bízva az osztálycímkék hozzárendelését. Két kihívással szembesülünk: kevés az ilyen utranagy dimenziójú, több osztályt tartalmazó adathalmaz illetve kevés az ilyen adatokból értelmes jellemzőket kinyerni képes szoftvercsomag.

A jellemzőket szűrni képes szoftverek fontos szerepet játszanak sok tudományterületen azáltal, hogy kiválasztják az eredeti jellemzők leghasznosabb kis részhalmazát. Ezeket használhatjuk az idegtudományban nagy térbeli felbontású EEG- és ECoG-felvételek mintáinak kinyerésére, de többcsatornás biometrikus azonosító adatok, például a kézírásos aláírás jellemzőinek válogatására is. Munkánkban bemutatjuk a BiometricBlender Python-csomagot, amellyel adott paramétereknek megfelelő utranagy dimenziós, több osztályú szintetikus adatok generálhatóak. Ezután bemutatunk egy új módszert, a véletlen erdőkön alapuló többkörös szűrést (RFMS), amelyet kifejezetten az irreleváns jellemzők hatékony kizárására és a bináris és magasabb rendű jellemzők kölcsönhatásainak azonosítására terveztünk. A javasolt algoritmus a jellemzők terét kis részhalmazokra bontja, és részleges modellépítések sorozatát hajtja végre, hogy a jellemzőket körversenyben fontosságuk szerinti sorba rendezze.

Nyílt forráskódú vírusgenom-összeszerelő pipeline-ok összehasonlító elemzése

Levente Zsichla^{1,2}, Marius Zeeb^{3,7}, Dávid Fazekas^{1,4}, Éva Áy^{2,5}, Dalma Müller^{1,2,6}, Karin J. Metzner^{3,7}, Roger 4 Kouyos^{3,7} és Viktor Müller^{1,2}

¹Biológiai Intézet, Eötvös Loránd Tudományegyetem

²Egészségbiztonság Nemzeti Laboratórium, Eötvös Loránd Tudományegyetem

³Fertőző Betegségek és Kórházi Járványügy Osztálya, Zürichi Egyetemi Kórház, Zürichi Egyetem

⁴Earlham Intézet

⁵Retrovírusok Nemzeti Referencia Laboratóriuma, Virologiai Laboratóriumi Osztály, Nemzeti Népegészségügyi és Gyógyszerészeti Központ

⁶Bioinformatikai Tanszék, Semmelweis Egyetem

⁷Orvosi Virologiai Intézet, Zürichi Egyetem

Az újgenerációs szekvenálás térnyerése az orvosi diagnosztikában és a járványtani vizsgálatokban igényt teremtett a gyors, megbízható, automatizált és felhasználóbarát bioinformatikai algoritmusok fejlesztésére. A short-read szekvenálási adatokból teljes vírusgenom összeszerelésre fejlesztett programok közül négy széles körben alkalmazott pipeline (shiver – amihez dshiver néven egy konténerizált változatot is készítettünk, valamint SmaltAlign, viral-ngs, V-Pipe) megbízhatóságát és alkalmazhatóságát vizsgáltuk meg szimulált és empirikusan gyűjtött HIV-1 páros végű szekvenálási eredmények segítségével.

Mind a négy vizsgált pipeline jó minőségű konszenzus genomokat és variánsbecsléseket eredményezett, ha a minta altípusa és az összeszereléshez használt referencia szekvencia kismértékű genetikai különbséget mutatott. Ugyan a shiver és a SmaltAlign megbízható teljesítményt nyújtott akkor is, amikor ez a genetikai távolság megnőtt (eltérő altípusú minta és referencia esetén), a viral-ngs és a V-Pipe teljesítménye jelentősen lecsökkent. Empirikus adatokon végzett elemzések során a SmaltAlign és a viral-ngs lényegesen kisebb futási időt igényelt, mint a shiver és a V-Pipe. Ezen felül figyelembe vettük, hogy egy elemzéshez a legmegfelelőbb pipeline kiválasztását a szoftverek alkalmazhatóságában mutatkozó különbségek is befolyásolhatják. Végezetül, a shiver új, konténerizált változata könnyen alkalmazhatósággal egészíti ki az eredeti algoritmus megbízható teljesítményét.

3. szekció - Rendszerbiológia, Szerkezeti bioinformatika, Orvosi bioinformatika

Poszt-perturbációs génexpressziót prediktáló gépi tanulási modellek kiértékelése

Csendes Gerold¹, Szalai Kristóf¹ és Szalai Bence¹

¹Turbine Kft

Az egészséges és beteg sejtek működésének megértéséhez elengedhetetlen a sejtválaszok pontos előrejelzése perturbációs hatásokra. Míg a perturbációs sejt adatok ideálisak az ilyen prediktív modellek tanítására, az ilyen adatforrások sokkal szűkösebben állnak rendelkezésre, mint a natív (nem-perturbált) adatok. E limitáció enyhítésére az elmúlt időszakban több, nagy-nyelvi modellek (Large Language Model, LLM) által inspirált módszert is fejlesztettek. Ezeket a modelleket nagy, címkézetlen (natív) egy sejt RNS szekvenálási adaton tanítják, majd specifikus feladatokra, mint például poszt-perturbációs génexpresszió profilokra finomhangolják (fine-tuning). Habár ezek a modellek érik el a legjobb kiértékelési metrikákat (SOTA), az eredmények kontextusba helyezése és benchmarkolása jelenleg is egy megoldatlan probléma.

Munkánk során egy ilyen modellt, az scGPT benchmarkolásával foglalkoztunk és több egyszerű modellel is összemértük. Meglepetésünkre azt tapasztaltuk, hogy az egyszerű modelljeink jobban teljesítettek, mint az scGPT. Továbbá, megállapítottuk, hogy a benchmark adatszetek alacsony perturbáció-specifikus varianciát hordoznak. Emiatt csak limitáltan alkalmasak valódi modell

generalizáció mérésére. Az eredményeink felhívják a figyelmet a jelenlegi benchmarkok limitációira és ötleteket ad arra, hogyan érdemes a poszt-perturbációs génexpressziós feladatokat kiértékelni.

Hogyan fedez fel egy sejtsszimuláció új fehérjekomplexeket?

Weber Áron^{1,2}, Kovács Szabolcs Cselgő^{1,2}, Ecker András¹, Noto Santiago¹, Fichó Erzsébet^{1,2}, Paccanaro Alberto³ és Csikász-Nagy Attila^{1,2}

¹Cytocast Hungary Kft.

²Pázmány Péter Katolikus Egyetem Információs Technológiai és Bionikai Kar

³Fundação Getúlio Vargas

A fehérjék jelentős része nem önmagában látja el funkcióját, hanem más fehérjékkel kölcsönhatva, makromolekuláris komplexeket alkotva. Emiatt a sejt komplexomjának, azaz fehérjekomplexeinek összességének vizsgálatával potenciálisan még több információt szerezhetünk a sejt működéséről, mint ha csak a proteomot, illetve az interaktomot vizsgálnánk.

A rendelkezésre álló irodalmi adatokat felhasználva több, fehérjekomplexeket leíró adatbázis is kifejlesztésre került. Mivel azonban ezen adatbázisok adatai in vitro mérésekből származnak, ezek bővítése rendkívül költséges és időigényes. Ezen hiányosságok okán több, komplexek predikcióját célzó számítási módszer is létrejött, mint például a fehérjeinterakciókat felhasználó ClusterOne és hu.MAP. Munkánk során kidolgoztunk egy eljárást, mely humán sejtek adatainak használatával, sejtsszimulációk eredményeiből képes lehetséges komplexek kvalitatív előrejelzésére. Ehhez a Cytocast Digital Twin Cellt, egy adatvezérelt, ágensalapú sejtsszimulátort használtunk, mely a fehérjék kötésén, és ezek felbomlásán keresztül jelzi előre a sejt lehetséges komplexeit, és azok mennyiségét. Saját módszerünk kimeneteit később összehasonlítottuk a ClusterONE és a hu.MAP predikcióival, megvizsgálva az átfedésüket. Az egymáshoz hasonló jóslott komplexek struktúráját ezután az AlphaFold3 segítségével validáltuk.

Az általunk kidolgozott módszer robusztus módon képes eddig ismeretlen fehérjekomplexeket felismerni, ezzel elősegítve a sejtek működésének egyik alapvető építőkövének minél teljesebb leírását.

Új szimulációs keretrendszer multivalens fehérjék és membránfehérjék közötti kötődés vizsgálatára

Brunccsis Bence¹ és Antal Péter¹

¹BME Mesterséges Intelligencia és Rendszertervezés Tanszéken, Mesterséges Intelligencia Kutatócsoport

Bemutatunk egy új szimulációs keretrendszert, amely a fehérjealapú gyógyszerek, például a bispecifikus (két eltérő antigénköti helyet tartalmazó) antitestek és membránhoz kötött fehérjék közötti kötődési interakciókat vizsgálja. A modell integrálja a 3D térre jellemző diffúziót és kinetikát a membránfelületen történő laterális diffúzióval, valamint a multivalens fehérjekötődési kinetikával. Ezáltal pontosabb leírást ad olyan biológiai rendszerekről, ahol a korábban alkalmazott, hagyományos kinetikai modellek már nem megfelelőek. A keresztkötések, a membrán laterális diffúziója és a lokális fehérjekoncentráció figyelembevételével a modell hatékony eszközt nyújt a multispecifikus membránköti terápiás molekulák kötési affinitásának és kinetikájának előrejelzésére, valamint szerkezeti optimalizálására. Különösen fontos a bispecifikus antitestek esetében, amelyek szerepe a modern gyógyszerfejlesztésben egyre növekszik, de kötődési tulajdonságaik még nem teljesen feltártak és tervezésük a gyakorlatban intuíción alapszik.

Ez a keretrendszer nemcsak a kötések kinetikai leírására alkalmas, hanem lehetőséget biztosít az új, hatékonyabb ligandstruktúrák tervezésére is. A Minnesotai Egyetem és a Budapesti Műszaki és

Gazdaságtudományi Egyetem közötti együttműködés révén a modell hozzájárulhat a következő generációs terápiák tervezéséhez, amelyek előmozdítják a kutatásokat egy dinamikusan fejlődő területen.

Kölcsönhatások elemzése a PSD-ben: a GKAP-LC8 példája

Nagy-Kanta Eszter¹, Kálmán E. Zsófia¹, Helena Tossavainen², Juhász Tünde³, Farkas Fanni¹, Hegedűs József¹, Keresztes Melinda¹, Beke-Somfai Tamás³, Gáspári Zoltán¹, Perttu Permi^{2,4,5} és Péterfia Bálint¹

¹Pázmány Péter Katolikus Egyetem Információs Technológiai és Bionikai Kar

²Department of Biological and Environmental Science, University of Jyväskylä

³HUN-REN Természettudományi Kutatóközpont, Anyag és Környezetkémiai Intézet

⁴Department of Chemistry, University of Jyväskylä

⁵Institute of Biotechnology, Helsinki Institute of Life Science, University of Helsinki

A posztszinaptikus denzitás (PSD) a szinapszis alapvető funkcionális alegysége, amelyet körülbelül 2 000 egyedi fehérje alkot. A PSD kritikus szerepet tölt be a memória kialakításában, a tanulási folyamatokban, valamint az agy kognitív funkcióinak megvalósításában. Az itt expresszálandó fehérjék mutációit számos neuropszichiátriai és neurológiai rendellenességgel/állapottal társították, többek között az autizmus spektrum zavarokkal, a skizofréniával és a Parkinson-kórral. A PSD strukturális felépítésében több fehérje játszik szerepet, köztük különös jelentőséggel bírnak az úgynevezett állvány fehérjék.

A GKAP állványfehérje és a dinenin könnyűlánc (DYNLL2, LC8) kölcsönhatásának esszenciális szerepe van a PSD-95 komplex megfelelő lokalizációjában. A komplexet többféle kísérletes módszerrel, például NMR-spektroszkópiával vizsgáltuk. A rendszer jellemzéséhez ezeket a kísérletes vizsgálatokat az AlphaFold szerkezetbecslésével és molekuladinamikai szimulációkkal egészítettük ki. Ennek eredményeként leírtuk a rendszer dinamikus szerkezeti változásait, emellett javaslatot tettünk a komplex sztöchiometriájára. Jellemeztük a GKAP hosszú rendezetlen régiójában található két nem-kanonikus motívum LC8-hoz történő kötését. Emellett további bioinformatikai elemzéseket végeztünk a LC8:GKAP kölcsönhatás kapcsán. Érdekes, hogy amíg az LC8 fehérje kötősebe konzervált, a GKAP fehérje kölcsönhatást kialakító részei változatosságot mutatnak különböző eukarióta organizmusok esetében. Hasonló elvet követve, a szekvenciális és szerkezeti elemzést kiterjesztettük több PSD fehérje kölcsönhatást kialakító moduljaira is. A jelenleg elfogadott nézet szerint az egyes fajokban mutató eltérő kognitív képességek nem a PSD-ben található fehérje készlet összetételéből adódnak, hanem az egyes fehérjék lehetséges kölcsönhatásaival állnak összefüggésben. Az itt bemutatott bioinformatikai módszerek alkalmazásával nagyskálán, a teljes PSD proteomra is vizsgálni szeretnénk a különböző fajok interakciós repertoárjának változatosságát.

Klinikai megfontolások az antimikrobiális rezisztenciapotenciálhoz

Tóth Adrienn Gréta^{1,2}, Solymosi Norbert^{1,2}, Nagy Sára Ágnes², Csabai István², Feczkó Csongor¹, Reibling Tamás¹ és Németh Tibor³

¹Bioinformatikai Központ, Állatorvostudományi Egyetem

²Komplex Rendszerek Fizikája Tanszék, Eötvös Loránd Tudományegyetem

³Sebészeti Tanszék, Állatorvostudományi Egyetem

Az antimikrobiális rezisztencia (AMR) a jelen és a közeljövő egyik legnagyobb közegészségügyi kihívása. Az AMR megjelenési gyakoriságában és terjedésében enyhülést hozhat az antibiotikumok célzott alkalmazása. Ehhez a kórképeket kialakító kórokozó baktériumok előzetes antibiotikum érzékenységi vizsgálatára van szükség. A mai rutin diagnosztikai eljárások során az antibiotikum érzékenységi vizsgálatokat csak a kórképekben szerepet játszó baktériumok

hagyományos módszerekkel tenyészthető tagjain végzik el. A második és harmadik generációs szekvenálásnak köszönhetően azonban a teljes bakteriális életközösség antimikrobiális rezisztenciagén (ARG) készlete feltérképezhető. A komplex, sok fajból álló bakteriális életközösségek ezen tulajdonságát antimikrobiális rezisztenciapotenciálnak (AMRP) is nevezhetjük. Az AMRP ráadásul anélkül is praktikus klinikai, közegészségügyi jelentőséggel bírhat, hogy az egyes ARG-ket baktériumtaxonokhoz kötnénk. A fenti, genomikai alapú AMR diagnosztikai megközelítési lehetőség 574 *Escherichia coli* törzs, valamint egy kutya külső hallójaratgyulladásból származó összetett mikrobiológiai minta példáján keresztül kerül bemutatásra. Az AMRP használata hozzásegíthet a hatékonyan alkalmazható antibiotikumok gyors megválasztásához, így akár az AMR térnyerésének hátráltatásában is szerepet játszhat.

Metagenomikai bepillantás egy gabonamúmiába.

Nagy Sára Ágnes¹, Solymosi Norbert^{1,2}, Pap Bernadett³, Tóth Adrienn Gréta^{1,2}, Kevély Flóra Judit⁴, Maróti Gergely³, Csabai István¹, Kóthay Katalin⁴ és Magyar Donát⁵

¹*Komplex rendszerek fizikája tanszék, Eötvös Loránd Tudományegyetem*

²*Bioinformatikai központ, Állatorvostudományi Egyetem*

³*Növénybiológiai Intézet, HUN-REN Szegedi Biológiai Kutatóközpont*

⁴*Szépművészeti Múzeum, Egyiptomi gyűjtemény*

⁵*Nemzeti Népegészségügyi és Gyógyszerészeti Központ*

Számos tanulmány kimutatta, hogy a metagenomika új dimenziót nyitott a régészeti maradványok tartalmának mint időkapcsoláknak az értelmezésében. A gabonamúmiák az ókori Egyiptomban készített víz, föld és gabonaszemek keverékét tartalmazó embert ábrázoló rituális tárgyak.

Egy nagyjából a Kr. e. 3. század második feléből származó gabonamúmiából vett minták metagenomikai elemzését végeztük el. Számos egyértelműen modern szennyeződés mellett azonosítottunk olyan organizmusokat is, amelyek esetében nem zárható ki, hogy történelmi eredetűek. A *Bacillus*, *Coprococcus*, *Metabacillus*, *Niallia* és *Roseburia* nemzetségekbe tartozó baktériumszekvenciák jelentős mennyisége mellett növények, állatok és emberek nyomait is megtaláltuk.

A *Triticum* nemzetséghez rendelt szekvenciák a legnagyobb hasonlóságot az Egyiptomból és a Levante déli részéről származó ősi *T. turgidum* ssp. *dicoccum* példányokkal mutatták. A *Populus* nemzetségbe sorolt szekvenciák leginkább a *P. nigra* genomjával mutattak hasonlóságot. A *Lepidoptera* eredetűként azonosított fragmentumok a legnagyobb hasonlóságot a *Sphingidae* genomjaival mutatták. A *H. sapiens*-szekvenciák haplotípus-elemzése a mitogenom és az Y-kromoszóma alapján az L3 és az E haplocsoportokat jelöli, amelyek a mai szubszaharai Afrikában két gyakori genetikai vonal.

4. szekció - Orvosi bioinformatika

Genomi monitoring és fágterápia összekapcsolása az antibiotikum-rezisztens kórokozók elleni küzdelemben

Mihály Koncz^{1,2,30}, Tamás Stirling^{1,2,3,30}, Hiba Hadj Mehdi^{1,2,3,30}, Orsolya Méhi^{1,30}, Bálint Eszenyi¹, András Asbóth^{1,2,4}, Gábor Apjok¹, Ákos Tóth⁵, László Orosz⁶, Bálint Márk Vásárhelyi¹, Eszter Ari^{1,4,7}, Lejla Daruka¹, Tamás Ferenc Polgár^{8,9}, György Schneider¹⁰, Sif Aldin Zalokh¹, Mónika Számel¹, Gergely Fekete^{1,7}, Balázs Bohár^{1,11}, Karolina Nagy Varga¹, Ádám Visnyovszki^{12,13}, Edit

Székely^{14,15}, Monica-Sorina Licker^{16,17}, Oana Izemendi, Carmen Costache¹⁹, Ina Gajic²⁰, Bojana Lukovic²¹, Szabolcs Molnár²², Uzonka Orsolya Szocs-Gazdi²³, Csilla Bozai²⁴, Marina Indreas²⁵, Katalin Kristóf²⁶, Charles Van der Henst^{27,28}, Anke Breine^{27,28}, Csaba Pál¹, Balázs Papp^{1,7,29} és Bálint Kintses^{1,2,31}

¹HUN-REN Szegedi Biológiai Kutatóközpont, Biokémiai Intézet, Biotechnológia Nemzeti Laboratórium, Szintetikus és Rendszer-biológiai Egység

²HCEMM-SZBK Transzlációs Mikrobiológia Laboratórium, Szeged

³Szegedi Tudományegyetem, Biológia Doktori Iskola

⁴Eötvös Loránd Tudományegyetem, Genetika Tanszék, Budapest

⁵Nemzeti Népegészségügyi és Gyógyszerészeti Központ, Budapest

⁶Szegedi Tudományegyetem, Szent-Györgyi Albert Orvostudományi Kar, Orvosi Mikrobiológiai Intézet

⁷HCEMM-SZBK Metabolikus Rendszerbiológia Kutatócsoport, Szeged

⁸HUN-REN Szegedi Biológiai Kutatóközpont, Biofizikai Intézet

⁹Szegedi Tudományegyetem, Elméleti Orvostudományok Doktori Iskola

¹⁰Pécsi Tudományegyetem, Általános Orvostudományi Kar, Orvosi Mikrobiológiai és Immunitástani Intézet

¹¹Faculty of Medicine, Department of Metabolism, Digestion and Reproduction, Imperial College London, London, UK

¹²Dél-Pesti Centrumkórház, Országos Hematológiai és Infektológiai Intézet

¹³Szegedi Tudományegyetem, Interdiszciplináris Orvostudományok Doktori Iskola

¹⁴George Emil Palade University of Medicine, Pharmacy, Science and Technology of Targu Mures, Targu Mures, Romania

¹⁵County Emergency Clinical Hospital of Targu Mures, Targu Mures, Romania

¹⁶Microbiology Department, Multidisciplinary Research Center on Antimicrobial Resistance, Victor Babes University of Medicine and Pharmacy, Timisoara, Romania

¹⁷Microbiology Laboratory, Pius Branzu Emergency Clinical County Hospital, Timisoara, Romania

¹⁸Doctoral School, Victor Babes University of Medicine and Pharmacy, Timisoara, Romania

¹⁹Department of Microbiology, University of Medicine and Pharmacy Iuliu Hatieganu Cluj-Napoca, Cluj-Napoca, Romania

²⁰Institute of Microbiology and Immunology, Faculty of Medicine, University of Belgrade, Belgrade, Serbia

²¹Academy of Applied Studies Belgrade, College of Health Sciences, Belgrade, Serbia

²²Emergency County Hospital Miercurea-Ciuc, Miercurea Ciuc, Romania

²³County Hospital Odorheiu-Secuiesc, Odorheiu-Secuiesc, Romania

²⁴County Emergency Hospital Satu Mare, Satu Mare, Romania

²⁵Bacau County Emergency Hospital, Bacau, Romania

²⁶Semmelweis Egyetem, Laboratóriumi Medicina Intézet, Budapest

²⁷Microbial Resistance and Drug Discovery, VIB-VUB Center for Structural Biology, VIB, Flanders Institute for Biotechnology, Brussels, Belgium

²⁸Structural Biology Brussels, Vrije Universiteit Brussel (VUB), Brussels, Belgium

²⁹HUN-REN Szegedi Biológiai Kutatóközpont, Egészségbiztonság Nemzeti Laboratórium

Az antibiotikum-rezisztens kórházi fertőzések elleni küzdelemben egyre nagyobb az érdeklődés a fágterápia iránt. A bakteriofágok szűk gazdaspecifitása ugyanakkor hátráltatja a széles hatásspektrumú fág készítmények fejlesztését és precíziós megoldásokat tesz szükségessé, amelyek nehezen skálázhatók. Jelen munkában ötvözzük a baktériumok nagy volumenű filogeográfiai elemzését és a fágérzékenység nagy áteresztőképességű vizsgálatát, hogy célzott fágoktatókat fejlesszünk egy kritikus kórokozó, a karbapenem rezisztens *Acinetobacter baumannii* ellen. Elemzésünk rávilágít arra, hogy egy-egy világrégióban a fertőzéseket mindössze néhány törzstípus dominálja, és ezek földrajzi eloszlása 6 éven keresztül stabil. Kelet-Európa példáján bemutatjuk, hogy a baktérium populáció tér- és időbeli változásai alapján előre láthatóvá válik, hogy milyen típusú fágokra van szükség a leggyakrabban előforduló helyi fertőzések kezelésére. A gyakori törzstípusok ellen fágoktatókat fejlesztettünk, majd ezek hatékonyságát in vitro és állati fertőzéses modelleken igazoltuk. Munkánk megmutatja, hogy a genomi monitoring segítségével világszerte azonosíthatók azok a betegek, akik ugyannal a fágkészítménnyel kezelhetők, így a precíziós fágterápia hatékonysága jelentősen növelhető.

Továbbfejlesztett ROCKET algoritmus tónusos-klónusos epilepsziás rohamok felismerésére

Buza Krisztián^{1,2}, Alexandros Nanopoulos¹ és Varga Noémi Ágnes¹

¹Sapientia Erdélyi Magyar Tudományegyetem, Matematika-Informatika Tanszék, Biointelligencia Kutatócsoport

²Budapesti Gazdasági Egyetem, Gazdaságinformatika Tanszék

A tónusos-klónusos epilepsziás rohamok mindennapi életben történő automatizált felismerése viselhető eszközökből származó gyorsulásmérős adatok alapján elősegítené az epilepsziás betegek diagnosztikáját és nyomonkövetését. Az előadásban bemutatunk egy algoritmust, amely hozzájárulhat a tónusos-klónusos epilepsziás rohamok felismeréséhez gyorsulásmérős adatok alapján, amelyeket mobil- és viselhető eszközökből lehet begyűjteni. Ezt a feladatot többváltozós idősor-osztályozási problémaként kezeljük. Az ilyen problémák megoldásának élvonalbeli megközelítései gépi tanulási technikákon alapulnak, mint például a Random Convolutional Kernel Transform (ROCKET). A ROCKET-et továbbfejlesztjük azzal, hogy a hagyományos konvolúciót dinamikus konvolúcióval helyettesítjük. A dinamikus konvolúció eredetileg egyváltozós idősorokhoz lett definiálva, ezért kiterjesztjük többváltozós idősorokra. Kísérleteket végzünk nyilvánosan elérhető, tónusos-klónusos rohamokhoz kapcsolódó valós gyorsulásmérős adatokon. A kísérleti eredmények azt mutatják, hogy a ROCKET algoritmus javasolt fejlesztései az átlagos osztályozási hibát közel egy nagyságrenddel csökkentik, konkrétan 1,74%-ról akár 0,18%-ra. Ráadásul a megoldásunk felülmúl más idősor-osztályozókat, beleértve számos, az idősorok osztályozása területén gyakran használt mély neurális hálót. Implementációnk nyilvánosan elérhető a következő címen: <https://github.com/kr7/ROCKET-DTW>.

A sport hatása a biológiai életkor változására

Torma Ferenc¹, Szatmári Eszter Zita², Kerepesi Csaba², Jókai Mátyás¹ és Radák Zsolt¹

¹Magyar Testnevelési és Sporttudományi Egyetem

²HUN-REN Számítástechnikai és Automatizálási Kutatóintézet

Az előadásomban Prof. Radák Zsolt (TF) kutatócsoportja által vezetett és mért fél éven át tartó edzésprogram adatainak bioinformatikai elemzését mutatom be. A kutatásom során rendszeresen sportoló egyének adatait elemeztem, a fizikai aktivitás öregedésre gyakorolt hatására fókuszálva. A kutatásban résztvevő személyeknél vérminták álltak rendelkezésre a program elején és végén, hogy megvizsgálhassuk, milyen hatással van a rendszeres testmozgás a biológiai életkorra, és tapasztalunk-e lassulást az öregedés folyamatában. A biológiai életkort a különböző öregedési órák segítségével határoztuk meg, hogy általános összefüggéseket találhassunk. Ezen felül elemeztem a fizikai teljesítmény két fontos mutatóját is: a szorítóerőt, amely az izomerő mérőszáma, valamint a VO₂max értéket, amely a szív- és érrendszer maximális oxigénfelvevő képességét mutatja. A kutatás célja annak felderítése volt, hogy van-e összefüggés a biológiai életkor változása és a fizikai teljesítmény javulása között. Az eredmények alapján arra kerestem választ, hogy a rendszeres sportolás valóban lassítja-e az öregedést, valamint, hogy a szorítóerő és a VO₂max értékek összefüggnek-e a biológiai életkor változásával.

A legújabb öregedés órák és érdekes alkalmazásai

Kerepesi Csaba¹

¹HUN-REN Számítástechnikai és Automatizálási Kutatóintézet

Az öregedés jelentős hatással van az egészségre, a gazdaságra és a társadalomra, de a közelmúltig nem volt lehetséges az öregedési folyamat olyan pontosságú mérése, hogy az gyakorlati alkalmazásokban is felhasználható legyen. A biológiai életkor eltér a kronológiai életkortól: egy egyénnek lehet gyorsult vagy lassult öregedése, melynek mérésére jelenleg az

életkort előrejelző gépi tanulási modellek ("öregedés órák") a legalkalmasabbnak. Az előadásban bemutatom a kutatócsoportunk legfrissebb eredményeit, amelyekben a biológiai életkor különféle aspektusait közelítettük meg gépi tanulási eszközökkel, például vérképekből, arcfotókból, betegségek kódokból, transzkriptomikai és epigenetikai adatokból kiindulva.

Epigenetikai altípusok és génszintű metilációs változások vizsgálata vastagbél adenokarcinómában

Müller Dalma¹ és Györffy Balázs^{1,2}

¹Semmelweis Egyetem, Bioinformatika Tanszék

²Természettudományi Kutatóközpont, Molekuláris Élettudományi Intézet, Onkológiai Biomarker Kutatócsoport

Bevezetés A metilációs mintázat átalakulása szolid tumorokban évtizedek óta kutatott folyamat. A vastagbélrák heterogén betegség, melynek a CpG-sziget metilátor fenotípusa (CIMP) sajátos klinikopatológiai jellemzőkhöz kapcsolható. A metilációt teljes genom szinten vizsgáló módszerek közül az Illumina Human Methylation array-alapú platformjai a legnépszerűbb, nagy mintaszámmal foglalkozó tanulmányok számára alkalmas platform. Kutatásunk célja a globális metiláció vizsgálata vastagbél adenokarcinómában, illetve egy ezt elősegítő felhasználóbarát webplatform létrehozása volt.

Módszerek Létrehoztunk egy normál vastagbél, adenoma és adenokarcinóma minták Illumina HumanMethylation450K és EPIC microarray-alapú metilációs adatait összegző adatbázist. Az adatokat 20 nyilvánosan elérhető tanulmányból gyűjtöttük össze. A metilációt gén és gén régió szinten tanulmányoztuk Mann-Whitney teszt és ROC elemzés segítségével. A CIMP fenotípust és a mismatch hibajavító rendszer rendellenes működését génszintű metilációs adatok alapján becsültük meg.

Eredmények A létrehozott adatbázis 2646 mintát tartalmaz. Teljes genom szinten, a metiláció csökkenése adenokarcinóma szövetekben és prekancerózus mintákban egyaránt megfigyelhető volt. A hipermetiláció a proximális promóterek közelében elhelyezkedő régiókra volt jellemző. Az összesített adatokon végzett elemzés alapján a C1orf70, a PITX1, IRX5, IGFL1, és az LRRC4 gének metilációja bizonyult a legjobban teljesítő biomarkernek. Az HM450K array adatok alapján elkülönülő CIMP adatszett a minták 14,8%-át (277 minta), az EPIC array adatok között azonosított pedig a minták 5,44%-át (36 minta) alkotta. A különböző CIMP epigenotípusú minták szignifikánsan hipermetilált gén régiói átlagosan 12,3 és 16,4%-ban (HM450K és EPIC) mutattak átfedést. Az átfedés tartalmazta a ROC elemzés által azonosított legjobb 20 biomarker jelöltet.

Következtetés Publikus adatok felhasználásával létrehoztunk egy adatbázist és egy webplatformot, mely kiindulópontja lehet új DNS-metilációs biomarkerek azonosításának vastagbél daganatban. A platform az epigenplot.com webcímen érhető el.

A makrofág promóterek összetételének vizsgálata

Dr. Nagy Gergely¹

¹Debreceni Egyetem, Általános Orvostudományi Kar, Biokémiai és Molekuláris Biológiai Intézet

Az emlős promótereket az őket felépítő elemek kombinációinak és elrendeződésének a sokfélesége teszi egyedivé. Ez az egyediség biztosítja a különböző sejtípusok különböző körülmények között szükséges promótereinek, ill. transzkript változatainak a kiválasztását. A promóterek közvetlen DNS-transzkripció faktor (TF) kölcsönhatásai azonban nagyrészt ismeretlenek. Az egér csontvelői makrofág (BMDM) egy széles körben használt sejtípus a génexpresszió szabályozásának tanulmányozására. Ez a makrofág modell különféle NGS adatszettek, pl. nagyszámú TF cisztróm forrásául is szolgált és szolgál. Munkám során a BMDM-ekből rendelkezésre álló epigenomikai és transzkriptomikai adatok feldolgozásával és

integrálásával jellemeztem a makrofágokban megfigyelhető közvetlen DNS-TF kölcsönhatásokat, különös tekintettel a promóterekre jellemzőekre.

Elemzéseim során azt találtam, hogy míg a makrofágokban aktív promóterek gazdagok metilálható, tehát CG-t tartalmazó elemekben, nagyobb részük nem-metilálható és prototipikusan promóter-disztális elemeket is magában foglal. Érdekes módon nagyjából minden hetedik promóter kizárólag ez utóbbi elemekből tevődik össze, melyek a makrofágokra jellemző TF-ok által biztosítják a sejttípus-specifikus génszabályozást. A CG-ben gazdag promóterekhez hasonlóan ezek is tartalmaznak jelentős részben demetilált CG helyeket, és magas polimeráz aktivitást mutatnak. Feltételezhető, hogy más sejtekben is működnek hasonló jellegű, sejttípus-meghatározó promóter osztályok.