Вихідний номер: 20240315-1 15 березня 2024 р., Київ

Відгук на план освітньо-професійної програми “Біоінформатика” Прикарпатського національного університету ім. Василя Стефаника

Ми, керівництво та представники громадської організації “Геноміка ЮА”, надсилаємо структурований відгук на план магістерської програми з біоінформатики Прикарпатського національного університету (ПНУ) ім. Василя Стефаника.

Останні п’ять років наша організація займається неформальною освітою, кар’єрним консультуванням та взаємодією з академічними та приватними установами у сфері біоінформатики та наук про життя. Керівництво громадської організації представлено вченими та фахівцями з досвідом роботи в академії та індустрії ЄС та Північної Америки. Нами була детально проаналізована вищезгадана програма відносно відповідності світовим кращим практикам та сучасному ринку праці. В цілому, ми вітаємо появу нової спеціальності з біоінформатики у ПНУ, а також вдячні за звернення до спільноти Геноміка ЮА. Нижче ми приводимо ключові пункти які, на нашу думку, могли б зробити цю освітньо-професійну програму кращою:

1. Кадри

Для навчання біоінформатиків, фахівець повинен мати підтверджений досвід в цій сфері. Наприклад, регулярні публікації в профільних журналах рівня BMC Bioinformatics та вище (фаховий перелік журналів МОН не є релевантним в цьому випадку), або мати регулярний досвід в дослідницьких або виробничих установах, або так званих “core facility”, “bioinformatics core” тощо як біоінформатик або консультант.

Не зважаючи на те, що наукові здобуття окремих представників робочої групи є солідними не лише за українськими, а і за світовими мірками, серед них нема фахівців з біоінформатики. Відповідно, закладу вищої освіти потрібно буде знайти та найняти таких інструкторів та лекторів для цієї програми. На нашу думку, підхід, щоб на магістерській програмі викладали переважно представники суміжних напрямів не є життєздатним, бо не дозволить підготувати студентів до вимог ринку праці.

З гарних новин: ця проблема не є унікальною для України, і у світі майже нема країн, які завдяки “внутрішнім” кадрам можуть закрити усі необхідні позиції для таких складних програм. Для цього є можливість залучити а) міжнародний рекрутинг, б) так звані “adjunct”-позиції, та в) організовувати запрошені лекції та практичні курси від визнаних фахівців для студентів.

а) Складно прогнозувати, скільки людей з релевантним досвідом оберуть зараз релокуватися до України для повної зайнятості в освіті в подібній програмі. До того невідомо, чи здатний Прикарпатський університет конкурувати за компенсацією зі, скажімо, аналогічними позиціями в інститутах Чехії або Польщі. Проте, є гіпотетична ймовірність залучити українців на початковому рівні кар’єри (постдоки+), які хотіли б продовжити працювати вдома.

б) Щодо adjunct-позицій, це може бути не менш гарним рішенням: на такі позиції запрошується фахівець, який/яка продовжуватиме працювати за основним місцем роботи, проте паралельно викладає або проводить дослідження на магістерській програмі. Чи то в

дистанційному форматі, чи приїжджаючи на конкретні курси. Скоріше за все, на таку позицію можливо залучити навіть враховуючи звичайну компенсацію у Прикарпатському університеті, тобто фактично pro bono, через те, що як фахівці-українці, так і іноземні колеги, що співчувають нам, були б не проти допомогти.

в) Гарною можливістю було б включити до програми серії лекцій-семінарів від різних провідних вчених у галузі (наприклад: Wolfgang Huber (EMBL), Katherine Pollard (UCSF), Jesse Bloom (UW), вчених українського походження з підтвердженим треком впливових публікацій та досвідом в індустрії тощо). Це потребує менше часу та забов’язань від лекторів, та це відчиняє можливість запросити більш видатних вчених на одноразові лекції.

Topic	Author documents	Topic Field-Weighted Citation Impact
Single-Cell Analysis; Small Cytoplasmic RNA; Single Cell RNA Seq	4	2.63
Interleukin 6; Rheumatology; COVID-19	3	3.47
Ibrutinib; B-Cell Chronic Lymphocytic Leukemia; Venetoclax	2	1.11
Extracellular Traps; Peptidylarginine Deiminase Type IV; Inflammation	2	1.76
Cancer; B-Lymphocyte Subset; Interleukin-10	2	1.09
Brahma; Chromatin Assembly and Disassembly; Complex	2	1.28
T-Cells; Graft Vs Host Disease; Interleukin-2	1	1.15
Nivolumab; Pembrolizumab; Immunotherapy	1	2.08
Exosomes; Extracellular Vesicles; Microrna	1	2.11
COVID-19; Coronavirus; Severe Acute Respiratory Syndrome	1	3.84
Delta Protein; Signal Transduction; Amyloid Precursor Protein Secretases	1	1.08

Приклад: профіль біоінформатика з досвідом, потрібним для викладання на магістерській програмі. Джерело: Scopus, де topic field-weighted citation impact > 1 показує, що автора цитують більше, ніж в середньому у сфері.

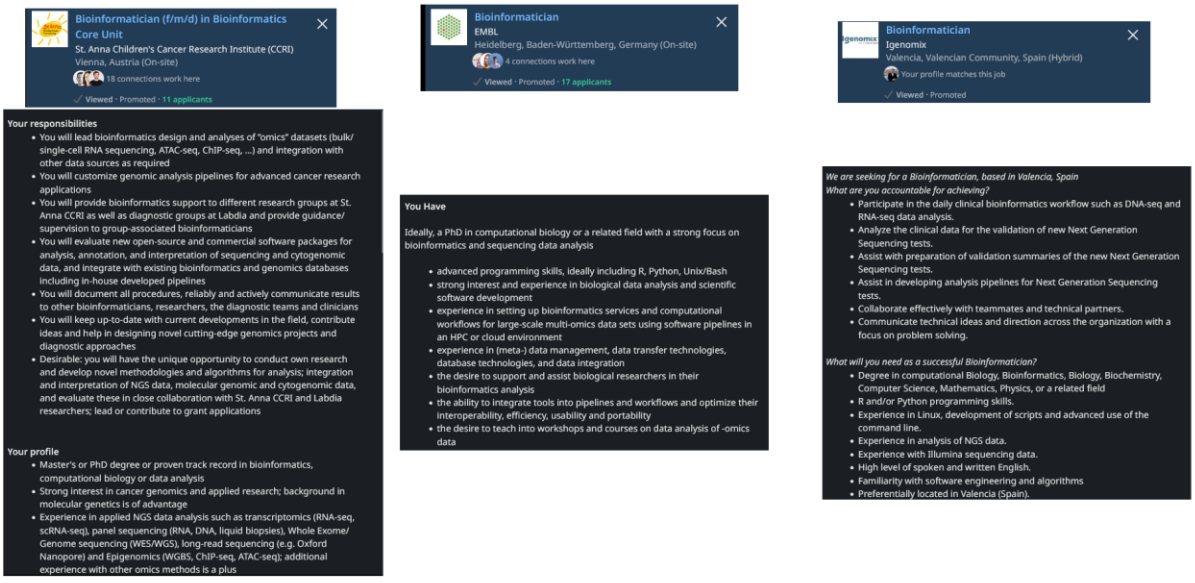
За умов дефіциту ресурсів (час кваліфікованих викладачів), ми би порадили наступні заходи, щоб забезпечити високий рівень програми з біоінформатики: а) кооперація з іншими установами, які вже мають розроблені курси; б) імплементація вже готових курсів, які існують в англомовному просторі (замість розробки курсів з початку); в) орієнтація на англомовні підручники з біоінформатики.

2. Структура програми

Запропонована програма - дуже потужний і близький погляд біологів на те, як може виглядати освітня програма з біоінформатики. Разом з запланованою сертифікацією програми як біологічної, це збігається із розумінням біоінформатики як вивчення біології на комп'ютері, а не біоінформатики як аналізу даних в будь-якій галузі, наприклад, в біології. Проте, якщо

ультимативна мета - щоб випускники отримували кар’єрні можливості десь поза закладами НАНУ, на нашу думку, її треба модифікувати.

Перед тим, як перейти до деталей, ми пропонуємо виконати просту вправу. Необхідно відкрити LinkedIn і ввести там bioinfmatician в пошуку - окремо для Європи і окремо для Північної Америки. Перше, що ви побачите - що більшість вакансій вимагає MSc + досвід АБО PhD. Тобто, очікується, що вищезгадана програма готуватиме фахівців, які можуть конкурувати на ринку з випускниками ETH Zurich, UCL та MIT - а їм доведеться, оскільки на одні й ті самі вакансії подаються десятки та сотні людей з усього світу. Друге, що можна побачити - вимоги та необхідні компетенції. Пропонуємо розглядати програму саме через призму них. Наприклад, автори програми вказують, що випускники працюватимуть в лабораторіях, які займаються аналізом геному та динаміки поширення вірусів. На жаль, це дуже нішеві напрями, які обмежують можливості для роботи. Повертаючись до LinkedIn, умовно, усі вакансії для біоінформатиків можна розділити на структурні/CADD (мало і, можливо, не профільно для ЗВО?), software-орієнтовані (мало і не профіль ЗВО) та омікснi (більшість). В останніх найбільший запит на таких, які займаються дослідженнями в сфері мультиомік, CRISPR/drug screens, imaging analysis, spatial biology.



Приклад: перші результати пошуку за запитами Bioinfomatician / European Union”.

Стратегічна проблема в тому, що, якщо фахівець у 2024 році вміє лише працювати з геномними та протеомними даними, як зазначено у проєкті програми, то він/вона фактично не здатні конкурувати на ринку. Бо що геномні дані, що протеомні сьогодні аналізують за допомогою pipelines; щобільше, якщо лабораторія сфокусована на геномних даних, то, скоріше за все, в них є доступ до комерційних систем від виробників обладнання (див. Illumina BaseSpace, Illumina DRAGEN), які роблять більшість рутинного аналізу таким, з яким здатні працювати technical assistant (TA) з бакалавратом. Те ж саме стосується і протеоміки: саме спектральні дані (LC/MS proteomics, metabolomics), як правило, є найменш складною проблемою для аналізу. Щобільше, зазвичай в core facilities первинну обробку даних та навіть певний downstream здатні з робити ті ж

самі ТА, бо для цього використовують готові рішення на кшталт ThermoFisher Protome Discoverer; якщо ж в людей є бажання працювати самостійно з матрицями спектральних прочитань і робити власний аналіз, як раджу робити студентам я, то для цього не треба півтора року навчатися на магістратурі - це здатен зробити й бакалавр з computer science. До того ж ми не побачили у проєкті програмі жодної згадки того, що вже дуже багато років є “королем” омік - bulk RNA-seq і більш сучасних scRNA-seq/snRNA-seq у різних модифікаціях. На відміну від протеоміки, про це як раз можна розмовляти півтора року (а краще разом з протеомікою в контексті інтеграції), бо це, по-перше, є найбільш затребуваним в академії, по-друге, відкриває шлях до spatial biology, яка як раз буде на пику коли ваші перші випускники отримують ступені, по-третє, використовується майже в кожній біофарм-компанії.

У програмі, яку підсумували автори, відведено 30% на програмування/інформатику, 30% на загальні навички, і аж 40% на біологію. Тут є декілька проблем. По-перше, навчати людей як працюють функції в Python (чомусь пріоритет відданий йому) - це не завдання магістерських програм, а, скоріше, бакалаврату та/або неформальної (само)освіти. І навпаки, слухачами програми можуть бути люди з вже наявною підготовкою у computer science. Відповідно, має сенс винести усі зайняття з Unix/Bash/SQL/Python/R/Julia/Rust/інше у факультатив, й зробити цей факультатив інтенсивним курсом прямо на старті програми. А ось що треба додати обов’язково - алгоритми (не моделювання, як вказано в програмі). До того ж у програми наразі нема нічого з таких тем як networks analysis, causal inference, machine learning, manifolds, dimensionality reduction, experimental design - це ті знання, які сьогодні є очікуваними від біоінформатиків на ринку праці.

По-друге, на біологію відведено аж 40% часу, з 3 кредитами та екзаменом (!) з ПЛР та секвенування, куди також чомусь віднесена підготовка бібліотек. Ми усвідомлюємо, що, оскільки програма буде сертифікована як біологічна, у МОН можуть бути певні вимоги з цього приводу. Однак, ми наполегливо рекомендуємо стейкхолдерам з МОН переглянути їх, бо в цьому конкретному випадку це скоріше завдає шкоди. Максимальний обсяг знань, який потрібен фахівцям-біоінформатикам про ПЛР та підготовку бібліотек секвенування - це один слайд на вступній лекції програми. Це - те, чим НЕ займаються біоінформатики, і окрім як “добре знати” не є релевантним. Випускник цієї спеціальності або це вже знає, або ніколи не буде працювати в “wet lab” умовах. Також, рекордні 6 кредитів відведені на філогенетику і молекулярну еволюцію - див. пораду з LinkedIn, це не є вирішальною навичкою, яка допоможе у працевлаштуванні студентам.

В цілому, в програмі критично не вистачає сучасних курсів: з аналізу даних ДНК-секвенування, РНК-секвенування, секвенування РНК поодиноких клітин, протеоміки, епігенетика. В технічній частині наразі не включено таких важливих розділів як Git та Github, Docker, code reproducibility.

А ось щодо soft skills/загальних навичок - на нашу думку, це дуже гарний напрям і те, чого зазвичай нема в українських університетах. Ми радимо залучити до таких курсів представників Enamine/Receptor.AI/Explogen, які можуть служитити прикладами вдалих українських стартапів та компаній.

Для складення сучасної конкурентної програми з біоінформатики ми радимо взяти за приклад програми для цієї спеціальності провідних Європейських та Американських університетів. З цього приводу можна звернути увагу на ці програми:

1. Технічний університет Мюнхену https://www.cit.tum.de/cit/studium/studiengaenge/bachelor-bioinformatik/modulkatalog/
2. UCLA (майте на увазі те що окрім курсів за спеціальності американський бакалаврат потребує проходження деяких загальноосвітніх курсів, як то “введення в біологію”, “вища математика” та ін.) https://bioinformatics.ucla.edu/undergraduate-bioinformatics/undergraduate-courses/
3. UCSD https://biology.ucsd.edu/education/undergrad/maj-min/majors/fall20-later/bioinformatics.html#Upper-Division-Requirements
4. Технологічний університет Іллінойсу https://bulletin.iit.edu/undergraduate/colleges/science-letters/biology/bs-bioinformatics/#programrequirementstext

Із конкретних заходів для співпраці ми могли б запропонувати:

- Викладання вже готових курсів “аналіз даних РНК-секвенування” та “аналіз даних РНК-секвенування поодиноких клітин” командою Геноміки.
- Викладання вже готового курсу “вступ у R” командою Геноміки.
- Викладання вже готового курсу “вступ до Git” командою Геноміки.
- Забезпечення студентів ПНУ сучасним посібником з біоінформатики С. Науменко.
- Експертний і незалежний аналіз програм курсів, які будуть подаватися викладачами програми.
- Участь в організації регулярного журнального і книжкового клубу (але він повинен бути відкритим для всіх охочих, не тільки студентів програми ПНУ, для яких він повинен бути обов’язковим).
- Антиплагіат-аналіз студентських робіт.

Решта коментарів відносно програми є меншими, але достатньо важливі, щоб їх включити до відгуку:

- Phython > Python, треба поправити у розділі 2.2.
- Було б гарно, якби були критерії щодо елективних курсів - тобто студенти їх будуть обирати, але викладачі заздалегідь повинні написати, про що цей курс буде, яка буде практика.
- Потрібно продумати та впровадити до початку програми комп’ютерне середовище, в якому студенти зможуть працювати під час практичних семінарві та для виконання магістерської роботи. Власний linux-сервер для обчислень, AWS або GCP - залежить від обставин, але щось з цього повинно бути. В останніх двох може існувати знижка для освітніх програм, особливо в Україні - за необхідністю, можемо допомогти з неогіацією знижки зі своєї сторони.

NGO “Genomics UA”
ГО “Геноміка ЮА”

genomics.org.ua
i@genomics.org.ua

Резюме

Програма, про яку йдеться в обговорені, потрібна для українського ринку освіти. Також ми всіляко вітаємо те, що для учасників програми заплановано багато елективних курсів. Проте, у наявній формі, без залучення фахових біоінформатиків, є високий ризик, що вона опиниться на цвинтарі освітніх програм - там же, де вже знаходиться багато інших українських програм з біоінформатики. Програмі потрібна масштабна ревізія щодо освітньої програми і введення “важких” об’єктивних фільтрів стосовно того, хто може на ній викладати, а хто - ні. Враховуючи реалії українського освітнього ринку, чи не єдиною запорукою успіху є колаборації з тими закладами та людьми, в яких є необхідний вам досвід. Наприклад:

1) Партнерство з представниками українських Data Science програм (наприклад, УКУ/КШЕ) , а також з визнаними фахівцями-біоінформатиками, що підтверджується адекватним треком публікацій та досвіду, для ревізії вашої computer science частини. Залучення таких фахівців до викладання.

2) Партнерство з КАУ щодо успішного досвіду “практичних” магістратур, які найкраще готують студентів до ринку праці (приєднання до їх програм або залучення до програми).

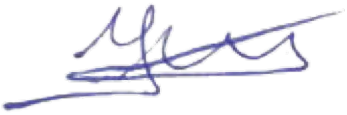
3) Партнерство з вченими, в тому числі українського походження, які мають підтверджений великий досвід в біоінформатиці й могли б викладати на вашому курсі як adjunct-faculty. Наприклад, Аліна Фролова (ІМБіГ), Тарас Олексик (OU), Сергій Науменко (CHEO), Serghei Mangul (UCSD), Олексій Рухленко (UCD), Ростислав Кузяків (ETH). Врешті-решт, з нами як найбільшою організацією з біоінформатики та омік в Україні (ГО “Геноміка ЮА”). Також більш широке коло провідних вчених може бути запрошено на одноразові лекції за темами в межах курсу, це розширило би можливість наукового спілкування для студентів.

З найкращими побажаннями,

Дмитро Казанжи
Директор та співзасновник організації “Геноміка ЮА”,
інженер по роботі з даними в Lyft (Україна)

Олександр Петренко
Менеджер проєктів та співзасновник організації “Геноміка ЮА”,
дослідник у Медичному університеті Відня (Австрія)

Сергій Науменко
Викладач курсу РНК-сік аналізу,
автор українського посібника з біоінформатики (https://v.gd/9gW0lj),
спеціаліст з геноміки (геномної біоінформатики)
у програмі скринінгу новонароджених провінції Онтаріо (Канада)



Юлія Ситнікова
Principal Scientist з біоінформатики,
Principal at Iaso Ventures (США)