

«СОГЛАСОВАНО»

Президент Академии наук
Республики Узбекистан
акад. Юлдашев Б.С.

“ ____ ” _____ 201__ год

« УТВЕРЖДАЮ»

Директор Центра геномики и
биоинформатики АН РУз
проф. Абдурахмонов И.Ю.

“ ____ ” _____ 201__ год

**Концепция развития
лаборатории биоинформатики
на 2016-2025 гг.**

Введение

Биоинформатика - область науки, выделяющаяся в единую дисциплину образованной на стыке биологии, информатики и информационных технологий. Эта новая область, возникла из-за стремительного развития биологических наук и появления сопутствующих новых технологий. Открытие структуры ДНК, новых концепций генов, глубокое изучение механизмов биологических процессов и механизмов привело к генерации и накоплению огромных массивов данных в частности из-за развития технологии ДНК секвенирования. Объем этих данных продолжает неуклонно расти благодаря технологиям секвенирования нового поколения (NGS). Анализировать столь большой объем информации вручную представляется невозможным, поэтому для анализа экспериментальных данных в биологических дисциплинах, а также медицине все больше и больше применяются компьютерные вычислительные системы.

Таим образом биоинформатика это новая область, которая имеет дело с применением компьютеров для сбора, организации, анализа, обработки, представления и обмена данными для решения биологических задач на молекулярном уровне, где биоинформатические, математические и статистические методы направлены на решение поставленных задач с использованием нуклеотидных и/или аминокислотных последовательностей и связанной с ними информацией.

В настоящее время биология, наряду с технологическими областями науки, определяет состояние прогресса в мире. От развития биологии напрямую зависят состояние экологии, биомедицины, биотехнологии, биофармацевтики и продовольственное обеспечение страны. Биомедицина в совокупности с биофармацевтикой обеспечивает высокое качество жизни человека, биотехнология решает проблему его продовольственного обеспечения, а экология оценивает качество среды обитания человека. По развитию биологии определяется уровень развития страны.

В последние десятилетия, развитие биологических наук и появление новых технологий, проявляется в создании новых высокопродуктивных сортов сельскохозяйственных культур с улучшенными признаками методами трансгеномики, изучении этиологии рака и других мультифакторных заболеваний, получении нового поколения лекарств с минимальными побочными эффектами, создании ферментов, обладающих высокой активностью и многое другое посредством методов генной инженерии, геномики, протеомики, биотехнологии. Отличительной чертой современной биологии является генерация и накопление массивов данных в огромных масштабах. Обработка такого объема информации вручную представляется невозможным, поэтому для их анализа исследователи стали применять компьютерные вычислительные системы. Использование компьютерных вычислительных систем и программ для решения биологических проблем стало причиной появления нового направления биологии – биоинформатики. Биоинформатика - позволяет анализировать данные полученные при секвенировании биологических последовательностей, организовывать эти данные и управлять ими с помощью баз данных, предсказывать трехмерную структуру белков и белок-белковые и лиганд-белковые взаимодействия и многое другое с помощью достижений в области информационных технологий. Биоинформатика отличается от других биологических направлений тем, что она оперирует не непосредственно с какими-то живыми объектами, а с результатами проведенных экспериментов: отдельный ген, полный геном какого-то организма, белковые взаимодействия, эволюционные исследования и так далее. Ученые-биоинформатики применяют различные инструменты для решения проблем биологии, в то время как исследователи, занимающиеся компьютерной биологией, разрабатывают теории, алгоритмы и другие техники для разработки инструментов, используемых учеными.

Биоинформатические анализы проводятся в различных лабораториях, однако структурированного методологического подхода к биоинформатике не наблюдаются. Создание лаборатории Биоинформатики на базе Центра Геномики и Биоинформатики стало первым шагом к развитию Биоинформатики в Узбекистане не просто как инструмента решения биологических задач и анализа полученных в лабораториях данных, а как

самостоятельного научного направления биологии. За прошедшие годы в Центре создана современная материально-техническая база для проведения исследований в области геномики, протеомики, метаболомики, биотехнологии на уровне мировой науки. Для проведения биоинформатических исследований в Центре накоплено и продолжается генерироваться большое количество информации для анализа и дальнейшей интерпретации.

Миссия лаборатории

Миссия лаборатории заключается в создании в Республике Узбекистан отечественной научно-исследовательской школы биоинформатики, как важнейшего элемента современного развития научных направлений в области биологии, медицины и фармакологии, что будет способствовать укреплению научно-технической базы страны на международном уровне.

Цели и задачи лаборатории Биоинформатики

Целью лаборатории является вывод Центра Геномики и Биоинформатики (ЦГБ) в лидеры научно-исследовательских центров в области биоинформатики и вычислительной биологии регионального уровня с международным признанием.

Исходя из цели, необходимо выполнить следующие задачи:

- **Научное направление**

1. Работа над приоритетными направлениями исходя из анализа научных направлений в области биоинформатики наиболее подходящих для реализации в условиях проектов и научно-исследовательских работ Центра;
2. Аннотация генов. Конверсия анонимных микросателлитных и других типов маркеров в генетических картах в функциональные маркеры путем выявления генов в маркированных регионах;
3. Определение структуры генов, представляющих интерес и их регуляторных элементов. Сравнительный анализ генов для выявления ключевых факторов, с целью дальнейших манипуляций с генами для инновационных решений. Публикация новых идентифицированных в Центре генов в международной генетической базе NCBI;

Полученные данные о генах и их структурах в дальнейшем будут использоваться лабораториями Центра для выведения новых экспериментальных линий растений.

4. Изучение сайтов метилирования регуляторных областей генов является одним из интересных направлений для исследования. В данное время появляются новые биоинформатические технологии, позволяющие с определенной долей вероятности определять возможные сайты метилирования ДНК последовательностей;
5. Анализ транскриптом – перспективное направление исследовательских работ в Центре, с его помощью можно целенаправленно выявлять ключевые гены для последующего их применения в области генной инженерии. В Центре имеется полный пул мРНК (так называемый RNAseq) средневолокнистого хлопчатника *G. hirsutum*, работы над анализом которого инициированы Центром;
6. Создание базы данных агробιοιολογических дескрипторов гермплазмы хлопчатника Центра Геномики и Биоинформатики и коллекции винограда Института Растениеводства;
7. Организация на базе Центра вычислительных ресурсов - запуск высокопродуктивного компьютерного кластера для решения задач в области биоинформатики;

- **Образование и проведение тренингов**

8. Подготовка специалистов по направлению «Биоинформатика»;
9. Подготовка и проведение семинаров и тренингов по разделам статистики, биоинформатики и прикладного программирования для действующих сотрудников центра с целью подготовки специалистов по двум основным направлениям:
 - а) Квалифицированных пользователей современного коммерческого и некоммерческого биоинформатического программного обеспечения (UGENE, Sequecher, FastPCR, GALAXY, GeneFlow, Arlequin, CLUSTAL, MUSCLE и др.), баз данных (BLAST, SWISSProt, Reactome, CMD и др.) и популярных биоинформатических платформ (Linux, BioLinux);
 - б) Специалистов, способных к самостоятельному созданию программ, утилит и скриптов на признанных в биоинформатике языках программирования (Python, Java и др.).

Важно отметить, что современная концепция высшего образования направлена на выпуск кадров узкой специализации, а биоинформатика является междисциплинарной наукой, требующей от специалиста компетенции в таких областях как биология, математика, статистика и информатика. Поэтому первостепенным для Центра является организация

курсов для действующих и тренингов для новых сотрудников лаборатории с целью дать им комплексное представление и систематическое образование в области биоинформатики.

- **Оказание услуг для сторонних организаций и физических лиц по следующим направлениям:**

10. геномные, транскриптомные и протеомные исследования;
11. анализ данных секвенирования, генетическое картирование;
12. аннотация генов;
13. анализ многомерных данных (филогенетический анализ, кластерный анализ, многомерное шкалирование и проч.);
14. корреляционный и регрессионный анализ;
15. анализ больших объемов данных;
16. пространственный анализ (GIS).

Для реализации поставленных задач, требуется привлечение следующих категорий ресурсов:

Финансовые ресурсы

- государственное финансирование: получение инвестиций и грантов от государственных организаций;
- стороннее финансирование: привлечение финансирования на реализацию грантов и стипендий для стажировки сотрудников за границей от локальных (Istedod, UMID, TEMUR) и международных (CASIA, DAAD, Humboldt, TWAS, ICARDA), как государственных, так и неправительственных организаций. Привлечение международных финансовых возможностей (например, USDA-ARS, CRDF, TWAS, ICARDA, FAO, NATO, KOPIA, RDA).
- самостоятельное рефинансирование: проведение по отработанным образовательным программам учебных курсов для слушателей из других учреждений, включая НИИ, ВУЗы, колледжи и лицеи с выдачей сертификатов.

Технические ресурсы

Для успешной реализации поставленных задач необходимо организовать техническую инфраструктуру, состоящую из следующих компонентов:

- Высокопроизводительный суперкомпьютер, состоящих из не менее чем 4-х вычислительных кластеров с использованием сопроцессоров NVidia Tesla.
- 3 сервера для автоматизации служебных функций компьютерной инфраструктуры.

- 3-5 высокопроизводительных рабочих станций с применением сопроцессоров для ведущих научных сотрудников лаборатории.
- Неограниченное высокоскоростное подключение к сети Интернет

Программные ресурсы

Для реализации поставленных задач необходимо приобретение программного обеспечения и лицензий на его использование, в случае коммерческого ПО. Для сокращения финансовых расходов на приобретение лицензий необходимо заключение договоров и партнерских соглашений между ЦГБ с крупными поставщиками ПО, такими как компании Microsoft, JetBrains и др., для получения бесплатных или льготных лицензий на использования их фирменного ПО, а так же использование, по возможности, бесплатного ПО.

Человеческие ресурсы

Развивая программу подготовки кадров на основе отдела тренингов, ЦГБ должен стать независимым от сторонних квалифицированных специалистов, так как штат сотрудников может быть пополнен из числа лучших выпускников собственной программы подготовки. В случае необходимости штат сотрудников может быть пополнен за счет подходящих сторонних специалистов. Организация тренингов для сотрудников Центра как локально, так и за рубежом.

Заключение

В результате реализации предлагаемой концепции развития лаборатории биоинформатики ожидается подготовка высококвалифицированных кадров, обладающих знаниями в области биоинформатики, развитие данного направления в Узбекистане, и как результат – становление Центра Геномики и Биоинформатики региональным лидером ресурсоемких научных (биоинформатических) вычислений в области биологии, медицины и фармакологии.