



ФЕДЕРАЛЬНОЕ ГОСУДАРСТВЕННОЕ БЮДЖЕТНОЕ УЧРЕЖДЕНИЕ
ВЫСШЕГО ОБРАЗОВАНИЯ И НАУКИ

**САНКТ-ПЕТЕРБУРГСКИЙ НАЦИОНАЛЬНЫЙ ИССЛЕДОВАТЕЛЬСКИЙ
АКАДЕМИЧЕСКИЙ УНИВЕРСИТЕТ ИМЕНИ Ж.И. АЛФЕРОВА
РОССИЙСКОЙ АКАДЕМИИ НАУК**

ПРОГРАММА

**вступительного экзамена в аспирантуру по Математической биологии и
биоинформатике**

**научная специальность 1.5.8 Математическая биология и
биоинформатика**

Санкт-Петербург

2023г.

Введение

В основу настоящей программы положены сведения из стандартного курса по биоинформатике.

Основы молекулярной биологии и генетики

Основные процессы передачи информации в клетке. Репликация, транскрипция, сплайсинг, трансляция. Различие и сходство основных процессов в эукариотах и в прокариотах. Посттранскрипционные модификации РНК. Посттрансляционные модификации белков. Структура ДНК в прокариотах и в эукариотах. Хроматин, основные уровни организации, гистоны, модификации хроматина. Роль хроматина в регуляции экспрессии генов. Регуляция экспрессии генов. Основные уровни регуляции экспрессии. Регуляция транскрипции, транскрипционные факторы. Сходство и различие регуляции транскрипции в прокариотах и в эукариотах. Понятие о регуляторных модулях. Регуляция трансляции. Регуляция инициации трансляции в прокариотах и в эукариотах. МикроРНК и механизм регуляции с помощью микроРНК. Стабильность РНК. РНК интерференция. Полиморфизмы. Признаки стабилизирующего и движущего отбора. Дрейф генов. Видообразование. Медицинская генетика. Моногенные и полигенные заболевания, анализ семей, ассоциации и молекулярные причины заболеваний. Понятие об основных экспериментальных методах молекулярной биологии. ПЦР, секвенирование, секвенирование нового поколения, микрочипы, белок-белковые взаимодействия, иммунопреципитация хроматина, ChIP-seq, масс-спектрометрия.

Математика и информатика

Случайные величины, распределения, математическое ожидание и дисперсия, основные распределения. Основы методов анализа данных. Выборка, нулевая гипотеза. Критерии χ^2 -квадрат, Фишера, Стьюдента, Колмогорова. Коэффициент корреляции и регрессия. Непараметрические критерии. Множественное тестирование. Дисперсионный анализ. Байесовский подход. Регрессионный анализ. Дискриминантный анализ. Методы кластеризации. Факторный анализ. Понятия Data mining и Text mining. Теоретическая информатика. Основные структуры данных: списки, стек, очередь, бинарное дерево поиска. Понятие алгоритма. Вычислительная сложность алгоритмов. Методы сравнения алгоритмов. Алгоритмы на графах, Эйлеров цикл, поиск оптимального пути. Алгоритмы для строк. Конечные автоматы, суффиксное дерево и суффиксный массив, регулярные выражения. Понятие об NP-полных задачах. Примеры NP-полных задач. Стохастические алгоритмы. Реляционные базы данных, язык SQL. Языки программирования в биоинформатике. Сравнительный анализ средств программирования.

Основы биоинформатики

Типы и качество данных. Биологические базы данных. Первичные (архивные), курируемые и производные базы данных. Выравнивание. Методы выравнивания: парное и множественное, локальное и глобальное. Алгоритм глобального выравнивания Нидльмана-Вунша (Needleman-Wunsh). Алгоритм локального выравнивания Смита-Уотермана (Smith-Waterman). Биологический смысл выравнивания. Понятие о «золотом стандарте». Алгоритмы динамического программирования. Статистическая значимость выравнивания. Линейное и логарифмическое поведение веса выравнивания. Методы быстрого поиска сходства BLAST, FASTA. Скрытые Марковские модели. Определение параметров моделей. Скрытые Марковские модели для выравнивания. Алгоритм Витерби. Субоптимальные выравнивания. Множественное выравнивание последовательностей. Динамическое

программирование для множественного выравнивания. Прогрессивное выравнивание. Улучшение выравнивания. Реконструкция эволюции по последовательностям. Укоренённые и неукоренённые филогенетические деревья. Основные методы реконструкции филогении.

Геномика, транскриптомика, протеомика, системная биология

Геномы, размер геномов бактерий и эукариот. Метагеномы. Контиги. Расшифровка геномов и сборка контигов. Структура геномов прокариот. Особенности бактериальных геномов. Особенности геномов эукариот. Геном человека и млекопитающих. Полиморфизмы человека. Аннотация геномов. Предсказание генов. Функциональная аннотация. Использование сходства. Сравнительный анализ геномов. Доменные перестройки. Семейства доменов. Методы идентификации доменов в последовательности. Гомологи, ортологи и паралоги. Методы определения ортологичности. Метаболическая реконструкция. Совместная представленность генов в геномах, колокализация, корегуляция, коэкспрессия. Базы данных метаболических путей. Транскриптом. Методы определения транскриптомов. Методы анализа транскриптомов. Тканевая специфичность транскриптомов. Состав транскриптома, анализ сплайсинга. Приложения к исследованию заболеваний и диагностике. Протеом. Методы определения протеома. Пост-трансляционные модификации белков. Определение посттрансляционных модификаций. Участие модификаций белков в регуляторных каскадах. Эпигеномика. Методы определения эпигенома. Роль эпигенома в регуляции экспрессии генов. Типы регуляторных взаимодействий. Регуляторные каскады. Системная биология. Построение и анализ регуляторных сетей. Роль системной биологии в поиске мишеней для лекарственных средств.

Пространственная организация биомакромолекул

Нуклеиновые кислоты. Биологическая роль. Роль ДНК в биосинтезе белка. Химическое строение и пространственная структура нуклеиновых кислот. Разнообразие формы структур ДНК и РНК. Вода и нуклеиновые кислоты. Взаимодействие нуклеиновых кислот с белками. Белки. Физико-химические свойства и системы классификации. Иерархия уровней пространственной организации белков. Регулярные структуры полипептидной цепи: α -спираль, β -тяжи и β -листы, их распространение и роль в структурах белков. Пространственная структура белков. Отличия в структурной организации глобулярных и фибриллярных белков. Распознавание функциональных сайтов и мотивов в белках. Биологическая роль белков. Ферменты: классификация; соответствие структуры и функции. Регуляторные белки. Мембранные белки, особенности их строения и функции. Основные методы расшифровки пространственных структур биомакромолекул. Особенности моделей, получаемых этими методами. Методы оценки качества пространственной модели белка. Основы анализа пространственной структуры макромолекул. Поверхность макромолекулы, алгоритмы её вычисления. Гидрофобное ядро молекулы белка, алгоритмы его нахождения. Структурные домены белков, алгоритмы их нахождения. Пространственное выравнивание структур белков. Структурные классификации доменов. Физические взаимодействия, определяющие пространственную структуру биомолекул. Конформации и конформационная подвижность биомакромолекул. Понятие эмпирического силового поля. Параметризация валентных и невалентных взаимодействий в биополимерах. Роль растворителя в структурной организации биополимеров. Гидрофобные взаимодействия в биомолекулярных системах. Шкалы гидрофобности. Методы учета влияния растворителя в расчетах энергии биомолекулярных систем. Самоорганизация пространственной структуры биополимеров. Парадокс Левинтала. Динамика конформаций. Проблема сворачивания (фолдинга) биополимеров. Ферментативный катализ химических

реакций. Понятие молекулярного докинга. Докинг в разработке лекарственных средств. Молекулярная динамика биомакромолекул. Подготовка системы к моделированию молекулярной динамики. Типы силовых полей. Моделирование динамики при постоянной энергии и постоянной температуре. Равновесная и направленная (управляемая) молекулярная динамика. Возможности и ограничения моделирования молекулярной динамики.

Литература

1. Альбертс Б.И. Основы молекулярной биологии клетки М.: Лаборатория знаний, 2018.
2. Understanding bioinformatics. Marketa Zvelebil, Jeremy O. Braum, Garland Science, Taylor and Francis Group, LLC 2008.
3. Леск, А. Введение в биоинформатику / А. Леск; пер. с англ. под ред. д.б.н., проф. А.А. Миронова и д.х.н., проф. В.К. Швядаса. Москва: БИНОМ. Лаборатория знаний, 2009.- 318 с.