

Практикум 6. Базы данных KEGG, GO и другие

Отчёт по этому заданию необходимо оформить в виде документа pdf объемом не менее трех(!!!) и не более восьми(!!!) страниц и прикрепить ссылку на него в очередь с соответствующим номером практикума.

Назовите файл согласно шаблону: **IvanovI_y24_pr6_bioinf.pdf**

Отчет оформите в виде СВЯЗНОГО ТЕКСТА с иллюстрациями. Все таблицы и иллюстрации должны быть подписаны и пронумерованы, к каждому рисунку и таблице в тексте отчета должно быть приведено краткое текстовое описание и обсуждение с ссылкой на соответствующий номер рисунка или таблицы.

Не стесняйтесь проявлять свою фантазию и испытывать чувство юмора проверяющих, главное, чтобы это не мешало содержательной части отчета. В отчете должна быть представлена вся необходимая информация для понимания сути и цели исследования.

Мягкий дедлайн: **12 апреля в 23:59**. Жесткий дедлайн: **20 апреля в 23:59**.

Каждому из Вас дан список ID генов человека (каждый файл назван как **ListN.txt**, смотрите свой N в ведомости)).

На занятии мы проходили несколько баз данных/сервисов, которые можно разделить на 2 группы:

- для **группового** анализа полного списка ID: GO (любой сервис), KEGG, Reactome, STRING, DAVID, Panther
- для **индивидуального** анализа одного любого ID из списка: Human Protein Atlas, KEGG, STRING, Genome browser, Gene Cards, Uniprot

Вам необходимо из **каждой** группы выбрать **одну** базу или сервис и исследовать возможности этой базы или сервиса, используя в качестве входных данных выданные вам ID.

Обратите внимание, что некоторые ресурсы могут принимать на вход и один ID и список ID. Учтите это при формулировке соответствующих задач, заранее решите, рассматриваете ли вы конкретный сервис для группового или для индивидуального анализа.

В отчете **обязательно** должно быть следующее:

1. Кратко опишите входные данные (что они из себя представляют, сколько генов в списке, ссылка на список)
2. Групповая база/сервис.
 - a. Какие задачи вы можете решать с помощью выбранной базы/сервиса? На примере трех задач. Сформулируйте в виде утверждения.
 - b. Какова общая «тема» списка ваших ID по результатам обогащения?
 - Укажите параметры запуска анализа: метод, поправка на множественное тестирование.
 - Приведите ссылку на таблицу с результатами, отсортированную по p-value с поправкой.
 - Сколько терминов (GO terms, путей в KEGG или др.) оказалось в выдаче?
 - В тексте опишите лучшие находки.
 - Проиллюстрируйте результаты, если возможности выбранного ресурса это позволяют. **Подписи к иллюстрациям должны быть понятными без обращения к тексту Вашего отчета. В тексте должны быть ссылки на все рисунки.**
 - Все полученные результаты должны быть обсуждены - ЭТОТ пункт САМЫЙ главный, нам нужны ВАШИ мысли по поводу сделанного.
3. Индивидуальная база/сервис.
 - a. Какие задачи вы можете решать с помощью выбранной базы/сервиса? На примере трех задач. Сформулируйте в виде утверждения.
 - b. На примере одного ID из вашего списка проиллюстрируйте решение одной из заявленных выше задач.
 - c. Соотнесите характеристику вашего ID с результатами, полученными в пункте 2. Например, поясните, какую конкретно функцию выполняет выбранный ID в общем процессе, полученном по результатам анализа обогащения в предыдущем пункте.
 - d. Проиллюстрируйте результаты, если возможности выбранного ресурса это позволяют. **Подписи к иллюстрациям должны быть понятными без обращения к тексту Вашего отчета. В тексте должны быть ссылки на все рисунки.**
 - e. Все полученные результаты должны быть обсуждены - ЭТОТ пункт САМЫЙ главный, нам нужны ВАШИ мысли по поводу сделанного.

Сохраняйте(!!!!) предложенную выше структуру отчета с формулировками заданий и пунктами!

Оцениваться будет качество вашего отчета - связность и осмысленность обсуждения результатов, оформление.

При желании можете исследовать **больше баз данных** и сервисов за **дополнительные** баллы. При этом можно добавить в отчет не более двух страниц на каждую дополнительную базу данных или сервис.

Постарайтесь отнестись к заданию, как к возможности ознакомиться с разными ресурсами, проявить научную самостоятельность и узнать что-то новое и интересное.

Если у вас не получилось найти значимого обогащения терминами, напишите преподавателям. Хотя такое крайне маловероятно, потому что Вам даны списки генов, участвующие в определенных процессах (это подсказка).

Внимательно прочитайте правила оформления задания. При несоответствии объема, формата и прочих требований работа не будет проверена.

При возникновении трудностей не стесняйтесь нам писать.