

**ПРАВИТЕЛЬСТВО РОССИЙСКОЙ ФЕДЕРАЦИИ
НАЦИОНАЛЬНЫЙ ИССЛЕДОВАТЕЛЬСКИЙ УНИВЕРСИТЕТ
«ВЫСШАЯ ШКОЛА ЭКОНОМИКИ»
Факультет компьютерных наук
Департамент программной инженерии**

СОГЛАСОВАНО
Научный руководитель,
Доцент департамента больших данных и
информационного поиска, заведующая
лабораторией «Научно-учебная
лаборатория биоинформатики» факультета
компьютерных наук

_____ М.С. Попцова
«___» _____ 2020 г.

УТВЕРЖДАЮ
Академический руководитель
образовательной программы
«Программная инженерия», канд. техн.
наук, профессор департамента
программной инженерии факультета
компьютерных наук

_____ В. В. Шилов
«___» _____ 2020 г.

**ПРОГРАММА ДЛЯ ИССЛЕДОВАНИЯ ПРИМЕНИМОСТИ МЕТОДОВ
КЛАСТЕРИЗАЦИИ И СНИЖЕНИЯ РАЗМЕРНОСТИ ДЛЯ АВТОМАТИЧЕСКОГО
СРАВНЕНИЯ ЭКСПЕРИМЕНТОВ ОДНОКЛЕТОЧНОГО СЕКВЕНИРОВАНИЯ**

Техническое задание

ЛИСТ УТВЕРЖДЕНИЯ

RU.17701729.04.13-01 ТЗ 01-1-ЛУ

Исполнитель
студент группы БПИ182
_____ / И. С. Егоров /
«___» _____ 2020 г.

Москва 2020

Подп. и дата	
Инв. № дубл.	
Взам. инв. №	
Подп. и дата	
Инв. № подл	

УТВЕРЖДЕН
RU.17701729.04.13-01 ТЗ 01 1-ЛУ

**ПРОГРАММА ДЛЯ ИССЛЕДОВАНИЯ ПРИМЕНИМОСТИ МЕТОДОВ
КЛАСТЕРИЗАЦИИ И СНИЖЕНИЯ РАЗМЕРНОСТИ ДЛЯ АВТОМАТИЧЕСКОГО
СРАВНЕНИЯ ЭКСПЕРИМЕНТОВ ОДНОКЛЕТОЧНОГО СЕКВЕНИРОВАНИЯ**

Техническое задание

RU.17701729.04.13-01 ТЗ 01-1

Листов 20

Име. № подл	Подп. и дата	Взам. инв. №	Име. № дубл.	Подп. и дата

Москва 2020

АННОТАЦИЯ

Настоящий документ оговаривает набор требований и порядок создания программного продукта. В соответствии с техническим заданием производится разработка программы, ее тестирование и приемка.

Настоящее Техническое задание на разработку «Программы для исследования применимости методов кластеризации и снижения размерности для автоматического сравнения экспериментов одноклеточного секвенирования» содержит следующие разделы: «Введение», «Основания для разработки», «Назначение разработки», «Требования к программе», «Требования к программным документам», «Технико-экономические показатели», «Стадии и этапы разработки», «Порядок контроля и приемки» и приложения [7].

В разделе «Введение» указано наименование и краткая характеристика области применения «Программы для исследования применимости методов кластеризации и снижения размерности для автоматического сравнения экспериментов одноклеточного секвенирования».

В разделе «Основания для разработки» указан документ, на основании которого ведется разработка и наименование темы разработки.

В разделе «Назначение разработки» указано функциональное и эксплуатационное назначение программного продукта.

Раздел «Требования к программе» содержит основные требования к функциональным характеристикам, к надежности, к условиям эксплуатации, к составу и параметрам технических средств, к информационной и программной совместимости, к маркировке и упаковке, к транспортировке и хранению, а также специальные требования.

Раздел «Требования к программным документам» содержит предварительный состав программной документации и специальные требования к ней.

Раздел «Технико-экономические показатели» содержит ориентировочную экономическую эффективность, экономические преимущества разработки «Программы для исследования применимости методов кластеризации и снижения размерности для автоматического сравнения экспериментов одноклеточного секвенирования».

Раздел «Стадии и этапы разработки» содержит стадии разработки, этапы и содержание работ.

В разделе «Порядок контроля и приемки» указаны общие требования к приемке работы.

Перед прочтением данного документа рекомендуется ознакомиться с терминологией, приведенной в Приложении 1.

Техническое задание разработано в соответствии со следующими требованиями:

- 1) ГОСТ 19.101-77 Виды программ и программных документов [1];
- 2) ГОСТ 19.102-77 Стадии разработки [2];
- 3) ГОСТ 19.103-77 Обозначения программ и программных документов [3];
- 4) ГОСТ 19.104-78 Основные надписи [4];
- 5) ГОСТ 19.105-78 Общие требования к программным документам [5];
- 6) ГОСТ 19.106-78 Требования к программным документам, выполненным печатным способом [6];
- 7) ГОСТ 19.201-78 Техническое задание. Требования к содержанию и оформлению [7].

Изменения к данному Техническому заданию оформляются согласно ГОСТ 19.603-78 [8], ГОСТ 19.604-78 [9].

Изм.	Лист	№ докум.	Подп.	Дата
RU.17701729.04.13-01 ТЗ 01-1				
Инв. № подл.	Подп. и дата	Взам. инв. №	Инв. № дубл.	Подп. и дата

Оглавление

1. ВВЕДЕНИЕ	5
1.1. Наименование программы.....	5
1.2. Краткая характеристика области применения	5
2. ОСНОВАНИЯ ДЛЯ РАЗРАБОТКИ	6
2.1. Документы, на основании которых ведется разработка.....	6
2.2. Наименование темы разработки	6
3. НАЗНАЧЕНИЕ РАЗРАБОТКИ.....	7
3.1. Функциональное назначение.....	7
3.2. Эксплуатационное назначение	7
4. ТРЕБОВАНИЯ К ПРОГРАММЕ.....	8
4.1. Требования к функциональным характеристикам.....	8
4.1.1. Требования к составу выполняемых функций.....	8
4.1.2. Требования к организации входных данных	8
4.1.3. Требования к организации выходных данных	8
4.2. Требования к интерфейсу	8
4.3. Требования к надежности	8
4.3.1. Требования к обеспечению надежного (устойчивого) функционирования программы	8
4.3.2. Время восстановления после отказа	9
4.3.3. Отказы из-за некорректных действий оператора	9
4.4. Условия эксплуатации	9
4.4.1. Требования к видам обслуживания	9
4.4.2. Требования к численности и квалификации персонала	9
4.5. Требования к составу и параметрам технических средств	9
4.6. Требования к информационной и программной совместимости	9
4.6.1. Требования к информационным структурам и методам решения	9
4.6.2. Требования к программным средствам, используемым программой	9
4.6.3. Требования к исходным кодам и языкам программирования	9
4.6.4. Требования к защите информации и программы	10
4.7. Требования к маркировке и упаковке	10
4.8. Требования к транспортированию и хранению.....	10
4.9. Специальные требования	10
5. ТРЕБОВАНИЯ К ПРОГРАММНОЙ ДОКУМЕНТАЦИИ.....	11
5.1. Предварительный состав программной документации.....	11

Изм.	Лист	№ докум.	Подп.	Дата
RU.17701729.04.13-01 ТЗ 01-1				
Инв. № подл.	Подп. и дата	Взам. инв. №	Инв. № дубл.	Подп. и дата

5.2. Специальные требования к программной документации.....	11
6. ТЕХНИКО-ЭКОНОМИЧЕСКИЕ ПОКАЗАТЕЛИ.....	12
6.1. Ориентировочная экономическая эффективность.....	12
6.2. Предполагаемая потребность	12
6.3. Экономические преимущества разработки по сравнению с отечественными и зарубежными образцами или аналогами	12
7. СТАДИИ И ЭТАПЫ РАЗРАБОТКИ	13
8. ПОРЯДОК КОНТРОЛЯ И ПРИЕМКИ.....	14
8.1. Виды испытаний	14
8.2. Общие требования к приемке работы.....	14
СПИСОК ИСПОЛЬЗУЕМОЙ ЛИТЕРАТУРЫ	15
ПРИЛОЖЕНИЕ 1	16
ПРИЛОЖЕНИЕ 2	17
ПРИЛОЖЕНИЕ 3	19
ЛИСТ РЕГИСТРАЦИИ ИЗМЕНЕНИЙ.....	20

Изм.	Лист	№ докум.	Подп.	Дата
RU.17701729.04.13-01 ТЗ 01-1				
Инв. № подл.	Подп. и дата	Взам. инв. №	Инв. № дубл.	Подп. и дата

1. ВВЕДЕНИЕ

1.1. Наименование программы

Наименование программы – «Программа для исследования применимости методов кластеризации и снижения размерности для автоматического сравнения экспериментов одноклеточного секвенирования».

Наименование программы на английском языке – «The software to study the applicability of clustering and dimensionality reduction methods for automatically comparing unicellular sequencing experiments».

1.2. Краткая характеристика области применения

Данная программа предназначена для широкого круга пользователей, работающих в сфере изучения транскриптомов единичных клеток в различных тканях.

Изм.	Лист	№ докум.	Подп.	Дата
RU.17701729.04.13-01 ТЗ 01-1				
Инв. № подл.	Подп. и дата	Взам. инв. №	Инв. № дубл.	Подп. и дата

2. ОСНОВАНИЯ ДЛЯ РАЗРАБОТКИ

2.1. Документы, на основании которых ведется разработка

Разработка выполняется в рамках темы курсовой работы в соответствии с учебным планом подготовки бакалавров по направлению 09.03.04 «Программная инженерия» Национального исследовательского университета «Высшая школа экономики», факультет компьютерных наук, департамент программной инженерии.

Приказ декана факультета компьютерных наук И. В. Аржанцева "Об утверждении тем, руководителей курсовых работ студентов образовательной программы «Программная инженерия» факультета компьютерных наук" № 2.3-02/1112-04 от 11.12.2019.

2.2. Наименование темы разработки

Наименование темы разработки – «Программа для исследования применимости методов кластеризации и снижения размерности для автоматического сравнения экспериментов одноклеточного секвенирования».

Изм.	Лист	№ докум.	Подп.	Дата
RU.17701729.04.13-01 ТЗ 01-1				
Инв. № подл.	Подп. и дата	Взам. инв. №	Инв. № дубл.	Подп. и дата

3. НАЗНАЧЕНИЕ РАЗРАБОТКИ**3.1. Функциональное назначение**

Приложение представляет собой веб-интерфейс для определения сходства входных данных с имеющимися в коллекции.

3.2. Эксплуатационное назначение

Приложение предназначено для использования исследователями, ведущими научную работу в области одноклеточного секвенирования и занимающиеся идентификацией клеточных подтипов в когорте клеток.

Изм.	Лист	№ докум.	Подп.	Дата
RU.17701729.04.13-01 ТЗ 01-1				
Инв. № подл.	Подп. и дата	Взам. инв. №	Инв. № дубл.	Подп. и дата

4. ТРЕБОВАНИЯ К ПРОГРАММЕ

4.1. Требования к функциональным характеристикам

4.1.1. Требования к составу выполняемых функций

Клиент приложения должен реализовывать следующий функционал:

1. Получение необходимых данных от пользователя;
2. Обработка входных данных, содержащих информацию о транскрипционном уровне генов;
3. Представление пользователю графической информации о разбиении отсеивенных единичных клеток на кластеры сходства с имеющимися в коллекции;
4. Предоставление пользователю возможности разбиения на кластеры отсеивенных единичных клеток и получения на основе распределения генов по кластерам предсказания имени и типа клетки.

4.1.2. Требования к организации входных данных

Пользователь должен загрузить базу единичных отсеивенных клеток в формате 10x_mlx (см. Приложение 2), упакованную в архив .zip.

4.1.3. Требования к организации выходных данных

Выходными данными должны быть интуитивно понятные графики, визуализирующие полученные кластеры.

В качестве выходных данных по запросу пользователя должен быть доступен для скачивания .txt файл, в котором по кластерам будут перечислены топ-5 (см. Приложение 3) клеток, наиболее подходящих для данного разбиения.

4.2. Требования к интерфейсу

Данный программный продукт должен иметь Web-интерфейс (см. Приложение 2).

- 1) В программном продукте должна быть предусмотрена возможность ввода данных пользователем.
- 2) В программном продукте должна быть предусмотрена возможность вывода пользователю графиков с визуализированными кластерами.
- 3) В программном продукте должна быть предусмотрена возможность вывода на скачивания .txt файла с предсказанными именами и типами клеток.

4.3. Требования к надежности

4.3.1. Требования к обеспечению надежного (устойчивого) функционирования программы

Приложение должно корректно обрабатывать все обращения пользователя и отображать его результаты. Должны быть предусмотрены возможные некорректные действия пользователей, в случае которых не должно возникать ошибок.

Изм.	Лист	№ докум.	Подп.	Дата
RU.17701729.04.13-01 ТЗ 01-1				
Инв. № подл.	Подп. и дата	Взам. инв. №	Инв. № дубл.	Подп. и дата

4.3.2. Время восстановления после отказа

Если отказ был вызван какими-либо внешними факторами, например DDOS-атакой, и при этом не произошел непоправимый сбой web-приложения, то время восстановления не должно превышать времени, требующегося на перезагрузку сервера.

Если отказ был вызван неисправностью технических средств или непоправимым сбоем сервера, то время восстановления не должно превышать времени, необходимого для устранения неисправностей технических и программных средств.

4.3.3. Отказы из-за некорректных действий оператора

Для предотвращения сбоев программы из-за некорректных действий пользователь должен ознакомиться с руководством оператора перед использованием данного приложения.

4.4. Условия эксплуатации

4.4.1. Требования к видам обслуживания

Сервера, на которых располагается приложения должны обслуживаться с качеством, достаточным для бесперебойной работы web-приложения.

4.4.2. Требования к численности и квалификации персонала

Пользователем программы может являться исследователь в области биологии, имеющий навыки работы с web-интерфейсом и иметь входной файл нужного формата, который будет описан в “Руководстве оператора”.

4.5. Требования к составу и параметрам технических средств

Для использования приложения пользователю необходим бесперебойный доступ в Интернет.

4.6. Требования к информационной и программной совместимости

4.6.1. Требования к информационным структурам и методам решения

Программа должна обрабатывать входные данные пользователя при помощи нейросетей и корректно кластеризовать клетки.

4.6.2. Требования к программным средствам, используемым программой

Для работы клиента программы необходима любая операционная система с web-браузером.

4.6.3. Требования к исходным кодам и языкам программирования

Изм.	Лист	№ докум.	Подп.	Дата
RU.17701729.04.13-01 ТЗ 01-1				
Инв. № подл.	Подп. и дата	Взам. инв. №	Инв. № дубл.	Подп. и дата

Интерфейс программы должен быть написан на языке программирования Java Script. Программный код, выполняющий обработку входных данных и получающий некоторый результат для вывода на экран пользователю, должен быть реализован на языке Python.

4.6.4. Требования к защите информации и программы

Программа не должна допускать утечек личной информации пользователей.

4.7. Требования к маркировке и упаковке

Код для развёртывания приложения на сервере должен быть выложен на github.

На github или ftp-сервере должны быть размещены все необходимые файлы с обученными классификаторами.

4.8. Требования к транспортированию и хранению

Никаких особых требований к транспортировке и хранению не представляется.

4.9. Специальные требования

Специальные требования к данной программе не предъявляются.

Изм.	Лист	№ докум.	Подп.	Дата
RU.17701729.04.13-01 ТЗ 01-1				
Инв. № подл.	Подп. и дата	Взам. инв. №	Инв. № дубл.	Подп. и дата

5. ТРЕБОВАНИЯ К ПРОГРАММНОЙ ДОКУМЕНТАЦИИ

5.1.Предварительный состав программной документации

- 1) «Программа для исследования применимости методов кластеризации и снижения размерности для автоматического сравнения экспериментов одноклеточного секвенирования». Техническое задание (ГОСТ 19.201-78);
- 2) «Программа для исследования применимости методов кластеризации и снижения размерности для автоматического сравнения экспериментов одноклеточного секвенирования». Программа и методика испытаний (ГОСТ 19.301-78);
- 3) «Программа для исследования применимости методов кластеризации и снижения размерности для автоматического сравнения экспериментов одноклеточного секвенирования». Текст программы (ГОСТ 19.401-78);
- 4) «Программа для исследования применимости методов кластеризации и снижения размерности для автоматического сравнения экспериментов одноклеточного секвенирования». Пояснительная записка (ГОСТ 19.404-79);
- 5) «Программа для исследования применимости методов кластеризации и снижения размерности для автоматического сравнения экспериментов одноклеточного секвенирования». Руководство оператора (ГОСТ 19.505-79).

5.2. Специальные требования к программной документации

- 1) Все документы к программе должны быть выполнены в соответствии с ГОСТ 19.106-78 [6] и ГОСТ к этому виду документа (см. п. 5.1.).
- 2) Пояснительная записка должна быть загружена в систему Антиплагиат через ЛМС НИУ ВШЭ. Лист, подтверждающий загрузку пояснительной записки, сдается в учебный офис вместе со всеми материалами не позже, чем за день до защиты курсовой работы.
- 3) Вся документация и программа также сдается в электронном виде в формате .pdf или .docx. в архиве формата .rar или .zip.
- 4) За один день до защиты комиссии все материалы курсового проекта:
 - техническая документация,
 - программный проект,
 - исполняемый файл,
 - отзыв руководителя
 - лист Антиплагиата

должны быть загружены одним или несколькими архивами в проект дисциплины «Курсовой проект 2019-2020» в личном кабинете в информационной образовательной среде LMS (Learning Management System) НИУ ВШЭ

Изм.	Лист	№ докум.	Подп.	Дата
RU.17701729.04.13-01 ТЗ 01-1				
Инв. № подл.	Подп. и дата	Взам. инв. №	Инв. № дубл.	Подп. и дата

6. ТЕХНИКО-ЭКОНОМИЧЕСКИЕ ПОКАЗАТЕЛИ**6.1. Ориентировочная экономическая эффективность**

В рамках данной работы расчет экономической эффективности не предусмотрен.

6.2. Предполагаемая потребность

Данный продукт может быть интересен биологам-экспериментаторам и биоинформатикам, работающим в области одноклеточного секвенирования и занимающиеся идентификацией клеточных подтипов в когорте клеток.

6.3. Экономические преимущества разработки по сравнению с отечественными и зарубежными образцами или аналогами

В рамках данной работы расчет экономической эффективности не предусмотрен.

Изм.	Лист	№ докум.	Подп.	Дата
RU.17701729.04.13-01 ТЗ 01-1				
Инв. № подл.	Подп. и дата	Взам. инв. №	Инв. № дубл.	Подп. и дата

7. СТАДИИ И ЭТАПЫ РАЗРАБОТКИ

Стадии и этапы разработки были выявлены с учетом ГОСТ 19.102-77 [2]:

Стадии разработки	Этапы работ	Содержание работ
1. Техническое задание	Обоснование необходимости разработки программы	Постановка задачи.
		Сбор исходных материалов.
	Научно-исследовательские работы	Определение структуры входных и выходных данных.
		Определение требований к техническим средствам.
		Обоснование принципиальной возможности решения поставленной задачи.
	Разработка и утверждение технического задания	Определение требований к программе.
		Определение стадий, этапов и сроков разработки программы и документации на неё.
		Согласование и утверждение технического задания.
2. Рабочий проект	Разработка программы	Разработка алгоритма решения задачи.
		Окончательное определение конфигурации технических средств.
		Программирование и отладка программы.
	Разработка программной документации	Разработка программных документов в соответствии с требованиями ГОСТ 19.101-77 [1].
	Испытания программы	Разработка, согласование и утверждение порядка и методики испытаний.
		Проведение испытаний.
3. Внедрение	Подготовка и защита программного продукта.	Корректировка программы и программной документации по результатам испытаний.
		Подготовка программы и программной документации для презентации и защиты.
		Утверждение дня защиты программы.
		Презентация программного продукта.
		Передача программы и программной документации в архив НИУ ВШЭ.

Изм.	Лист	№ докум.	Подп.	Дата
RU.17701729.04.13-01 ТЗ 01-1				
Инв. № подл.	Подп. и дата	Взам. инв. №	Инв. № дубл.	Подп. и дата

8. ПОРЯДОК КОНТРОЛЯ И ПРИЕМКИ

8.1. Виды испытаний

Производится проверка корректного выполнения программой заложенных в нее функций, указанных в пункте 4.1 настоящего технического задания. Также осуществляется визуальная проверка интерфейса программы на соответствие пункту 4.2 настоящего технического задания.

Функциональное тестирование осуществляется в соответствии с документом «Программа для исследования применимости методов кластеризации и снижения размерности для автоматического сравнения экспериментов одноклеточного секвенирования». Программа и методика испытаний (ГОСТ 19.301-78), в котором указывают [17]:

1. перечень функций программы, выделенных в программе для испытаний, и перечень требований, которым должны соответствовать эти функции (со ссылкой на пункт 4.1.1 настоящего технического задания);
2. перечень необходимой документации и требования к ней (со ссылкой на пункт 5 настоящего технического задания);
3. методы испытаний и обработки информации;
4. технические средства и порядок проведения испытаний;

Сроки проведения испытаний обсуждаются дополнительно.

8.2. Общие требования к приемке работы

Прием программного продукта происходит при полной работоспособности программы при различных входных данных, при выполнении указанных в пункте 4.1.1 настоящего документа функций, при выполнении требований, указанных в пунктах 4.2 настоящего документа и при наличии полной документации к программе, указанной в пункте 5.1, выполненной в соответствии со специальными требованиями, указанными в пункте 5.2 настоящего технического задания.

Изм.	Лист	№ докум.	Подп.	Дата
RU.17701729.04.13-01 ТЗ 01-1				
Инв. № подл.	Подп. и дата	Взам. инв. №	Инв. № дубл.	Подп. и дата

СПИСОК ИСПОЛЬЗУЕМОЙ ЛИТЕРАТУРЫ

1. ГОСТ 19.101-77 Виды программ и программных документов. //Единая система программной документации. – М.: ИПК Издательство стандартов, 2001.
2. ГОСТ 19.102-77 Стадии разработки. //Единая система программной документации. – М.: ИПК Издательство стандартов, 2001.
3. ГОСТ 19.103-77 Обозначения программ и программных документов. //Единая система программной документации. – М.: ИПК Издательство стандартов, 2001.
4. ГОСТ 19.104-78 Основные надписи. //Единая система программной документации. – М.: ИПК Издательство стандартов, 2001.
5. ГОСТ 19.105-78 Общие требования к программным документам. //Единая система программной документации. – М.: ИПК Издательство стандартов, 2001.
6. ГОСТ 19.106-78 Требования к программным документам, выполненным печатным способом. //Единая система программной документации. – М.: ИПК Издательство стандартов, 2001.
7. ГОСТ 19.201-78 Техническое задание. Требования к содержанию и оформлению. //Единая система программной документации. – М.: ИПК Издательство стандартов, 2001.
8. ГОСТ 19.603-78 Общие правила внесения изменений. //Единая система программной документации. – М.: ИПК Издательство стандартов, 2001.
9. ГОСТ 19.604-78 Правила внесения изменений в программные документы, выполненные печатным способом. //Единая система программной документации. – М.: ИПК Издательство стандартов, 2001.
10. Устинов В. Надежность оптических дисков: как их правильно хранить и использовать. //Журнал «625» №7. М.: Издательство «625», 2005.
11. ГОСТ Р 7.02-2006 Консервация документов на компакт-дисках. Общие требования. – М.: ИПК Издательство стандартов, 2006.
12. ГОСТ 19.602-78 Правила дублирования, учета и хранения программных документов, выполненных печатным способом. //Единая система программной документации. – М.: ИПК Издательство стандартов, 2001.
13. ГОСТ 19.301-79 Программа и методика испытаний. Требования к содержанию и оформлению. //Единая система программной документации. – М.: ИПК Издательство стандартов, 2001.

Изм.	Лист	№ докум.	Подп.	Дата
RU.17701729.04.13-01 ТЗ 01-1				
Инв. № подл.	Подп. и дата	Взам. инв. №	Инв. № дубл.	Подп. и дата

ПРИЛОЖЕНИЕ 1

Клетка – структурно-функциональная элементарная единица строения и жизнедеятельности всех организмов.

Ген – структурная и функциональная единица наследственности живых организмов.

Кластер сходства – некие группы, на которые разделяются данные по какому-либо признаку.

Секвенирование - общее название методов, которые позволяют установить последовательность нуклеотидов в молекуле ДНК.

Одноклеточное секвенирование – подход, позволяющий получить данные о последовательности ДНК отдельной клетки с помощью секвенирования.

Экспрессия генов — процесс, в ходе которого наследственная информация от гена (последовательности нуклеотидов ДНК) преобразуется в функциональный продукт — РНК или белок.

Изм.	Лист	№ докум.	Подп.	Дата
RU.17701729.04.13-01 ТЗ 01-1				
Инв. № подл.	Подп. и дата	Взам. инв. №	Инв. № дубл.	Подп. и дата

ПРИЛОЖЕНИЕ 2

В данном разделе будет описан формат 10x_mx, используемый как формат входных данных проекта.

Формат 10x_mx представляет собой набор из трёх файлов: barcodes.tsv, genes.tsv, matrix.mx.

Файл barcodes.tsv содержит подотрезки ДНК генов (рис. 1).

Файл genes.tsv содержит описание генов (рис. 2).

Файл matrix.mx содержит тройки, задающие пары ген – подотрезок ДНК и значение экспрессии (рис. 3).

```
TTGGACCATCTCGCT
CGAATGTTCCCTAACC
GATTCAGAGCTAGCCC
CGGCTAGCAGCTGTAT
GTATTCTTCAGATAAG
GAGCAGAAAGTGGGCTA
TAAACCGGTAGCGTGA
TGCTACCTCATCACCC
ATTATCCAGACTAGAT
GGTATTGGTACATCCA
ACAGCCGTCTGCCAGG
ATCATCTCAATGGTCT
GGCTGGTGTCCATGAT
TCAGGTAAGAGTCGGT
TCCACACGTAAGTCAAC
AACGTTGGTAGAAGGA
CGAGCCATCTTCTGGC
CTGAAGTGTCTCCACT
GGACAGAGTCCGCTGA
GGTGGTCAAGCCCCAC
GGACAGAAAGGATGGAA
AACTCCCGTAGGCTGA
```

Рис. 1 Шапка файла barcodes.tsv

ENSG00000000005	TNMD	protein_coding	chrX	100584802	100599885	True	Homo sapiens
ENSG000000000419	DPH1	protein_coding	chr20	50934867	50958555	True	Homo sapiens
ENSG000000000457	SCVL3	protein_coding	chr1	169849631	169894267	True	Homo sapiens
ENSG000000000460	C1orf112	protein_coding	chr1	169662007	169854080	True	Homo sapiens
ENSG000000000938	FRG	protein_coding	chr1	27612064	27635277	True	Homo sapiens
ENSG000000000971	CFH	protein_coding	chr1	196651878	196747504	True	Homo sapiens
ENSG000000001036	FUCA2	protein_coding	chr6	143494811	143511690	True	Homo sapiens
ENSG000000001084	GCLC	protein_coding	chr6	53497341	53616970	True	Homo sapiens
ENSG000000001167	NFYA	protein_coding	chr6	41072945	41099976	True	Homo sapiens
ENSG000000001460	STPG1	protein_coding	chr1	24356999	24416934	True	Homo sapiens
ENSG000000001461	NIPAL3	protein_coding	chr1	24415794	24472976	True	Homo sapiens
ENSG000000001497	LAS1L	protein_coding	chrX	65512582	65534775	True	Homo sapiens
ENSG000000001561	ENPP4	protein_coding	chr6	46129993	46146699	True	Homo sapiens
ENSG000000001617	SEMA3F	protein_coding	chr3	50155045	50189075	True	Homo sapiens
ENSG000000001626	CFTR	protein_coding	chr7	117465784	117715971	True	Homo sapiens
ENSG000000001629	ANKIB1	protein_coding	chr7	92246234	92401384	True	Homo sapiens
ENSG000000001630	CYP51A1	protein_coding	chr7	92112151	92142952	True	Homo sapiens
ENSG000000001631	KRIT1	protein_coding	chr7	92198969	92246166	True	Homo sapiens
ENSG000000002016	RAD52	protein_coding	chr12	912077	990053	True	Homo sapiens
ENSG000000002079	MYH16	transcribed_unitary_pseudogene	chr7	99238829	99311130	True	Homo sapiens
ENSG000000002330	RAD	protein_coding	chr11	64269830	64284704	True	Homo sapiens
ENSG000000002549	LAP3	protein_coding	chr4	17577192	17607972	True	Homo sapiens
ENSG000000002586	CD99	protein_coding	chrX	2691179	2741309	True	Homo sapiens
ENSG000000002586	CD99	protein_coding	chrY	2691179	2741309	True	Homo sapiens
ENSG000000002587	HS3ST1	protein_coding	chr4	11393150	11429765	True	Homo sapiens
ENSG000000002726	AOC1	protein_coding	chr7	150824627	150861504	True	Homo sapiens
ENSG000000002745	WNT16	protein_coding	chr7	121325367	121341104	True	Homo sapiens

Рис. 2 Шапка файла genes.tsv

Изм.	Лист	№ докум.	Подп.	Дата
RU.17701729.04.13-01 ТЗ 01-1				
Инв. № подл.	Подп. и дата	Взам. инв. №	Инв. № дубл.	Подп. и дата

```

%%MatrixMarket matrix coordinate real general
3500 10000 936416
31 1 1.0
98 1 1.0
232 1 2.0
250 1 1.0
474 1 1.0
713 1 1.0
895 1 1.0
955 1 1.0
1310 1 1.0
1318 1 1.0
1737 1 1.0
1807 1 1.0
1834 1 1.0
1837 1 1.0
1838 1 1.0
1865 1 1.0
2114 1 1.0
2263 1 3.0
2344 1 1.0
2704 1 1.0
2937 1 1.0
3049 1 1.0
3063 1 1.0
3091 1 2.0
3151 1 1.0
3222 1 1.0

```

Рис. 3 Шапка файла matrix.mtx

Изм.	Лист	№ докум.	Подп.	Дата
RU.17701729.04.13-01 T3 01-1				
Инв. № подл.	Подп. и дата	Взам. инв. №	Инв. № дубл.	Подп. и дата

ПРИЛОЖЕНИЕ 3

В данном разделе будет описан формат файла .txt, который будет доступен пользователю для скачивания после работы программы.

В файле .txt будут содержаться предсказанные имена и типы клеток для кластеров (рис. 4).

Here are predicted cells by genes significance in each topic.

Topic 0

Cell name: Ureteric bud cell; Cell type: Normal cell

Cell name: Theca interna cell; Cell type: Normal cell

Cell name: Natural regulatory T (Treg) cell; Cell type: Normal cell

Cell name: Lake et al.Science.Ex1; Cell type: Normal cell

Cell name: Pancreatic polypeptide cell; Cell type: Normal cell

Topic 1

Cell name: Colony-forming cell; Cell type: Normal cell

Cell name: Neural stem cell; Cell type: Cancer cell

Cell name: Type I pneumocyte; Cell type: Normal cell

Cell name: Basal progenitor cell; Cell type: Normal cell

Cell name: Luminal progenitor; Cell type: Normal cell

Topic 2

Cell name: Colony-forming cell; Cell type: Normal cell

Cell name: Neural stem cell; Cell type: Cancer cell

Cell name: Plasmablast; Cell type: Normal cell

Cell name: Abnormal plasma cell; Cell type: Cancer cell

Cell name: Common lymphoid progenitor; Cell type: Normal cell

Topic 3

Cell name: Peritubular myoid cell; Cell type: Normal cell

Cell name: Myofibroblast; Cell type: Normal cell

Cell name: Lymphocyte; Cell type: Cancer cell

Cell name: Ito cell (hepatic stellate cell); Cell type: Normal cell

Cell name: Neural cell; Cell type: Normal cell

Рис. 4 Шапка файла с предсказанными клетками

Изм.	Лист	№ докум.	Подп.	Дата
RU.17701729.04.13-01 ТЗ 01-1				
Инв. № подл.	Подп. и дата	Взам. инв. №	Инв. № дубл.	Подп. и дата

[illegible]

Изм.	Лист	№ докум.	Подп.	Дата
RU.17701729.04.13-01 ТЗ 01-1				
Инв. № подл.	Подп. и дата	Взам. инв. №	Инв. № дубл.	Подп. и дата