

Министерство образования и науки Российской Федерации
Федеральное государственное автономное образовательное учреждение
высшего образования
«Санкт-Петербургский политехнический университет Петра Великого»

УДК

№

Инв. №

У Т В Е Р Ж Д А Ю

Зав. НИЛ «Математическая биология
и биоинформатика», ИПММ
ФГАОУ ВО «СПбПУ»,
д.б.н.

_____ М. Г. Самсонова
«____» _____ 2016 г.

ОТЧЕТ
О НАУЧНО-ИССЛЕДОВАТЕЛЬСКОЙ РАБОТЕ

по теме:
«Стохастическое моделирование экспрессии генов»

Выполнил студент гр. №53601/4

_____ Д. В. Яковлев
«____» _____ 2016 г.

Руководитель НИР, к.ф.-м.н.

_____ В.В. Гурский
«____» _____ 2016 г.

Санкт-Петербург 2016

РЕФЕРАТ

Отчёт 21 стр., 1 часть, 8 рис., 1 таблица, 5 источников

СТОХАСТИЧЕСКОЕ МОДЕЛИРОВАНИЕ ЭКСПРЕССИИ ГЕНОВ

СОДЕРЖАНИЕ

Введение	4
1 Модель	5
2 Инструменты для моделирования	8
2.1 StochPy, GillesPy	8
2.2 fern	8
2.3 StochKit	8
2.4 DEEP	8
3 Результаты	9
3.1 Описание реакций	9
3.2 Разработка программного обеспечения	10
3.3 Анализ результатов	10
СПИСОК ИСПОЛЬЗОВАННЫХ ИСТОЧНИКОВ	14

Введение

Экспрессия генов - это процесс преобразования последовательности нуклеотидов ДНК в функциональный продукт РНК и белок. Существует два основных направления моделирования экспрессии генов - аналитическое и стохастическое.

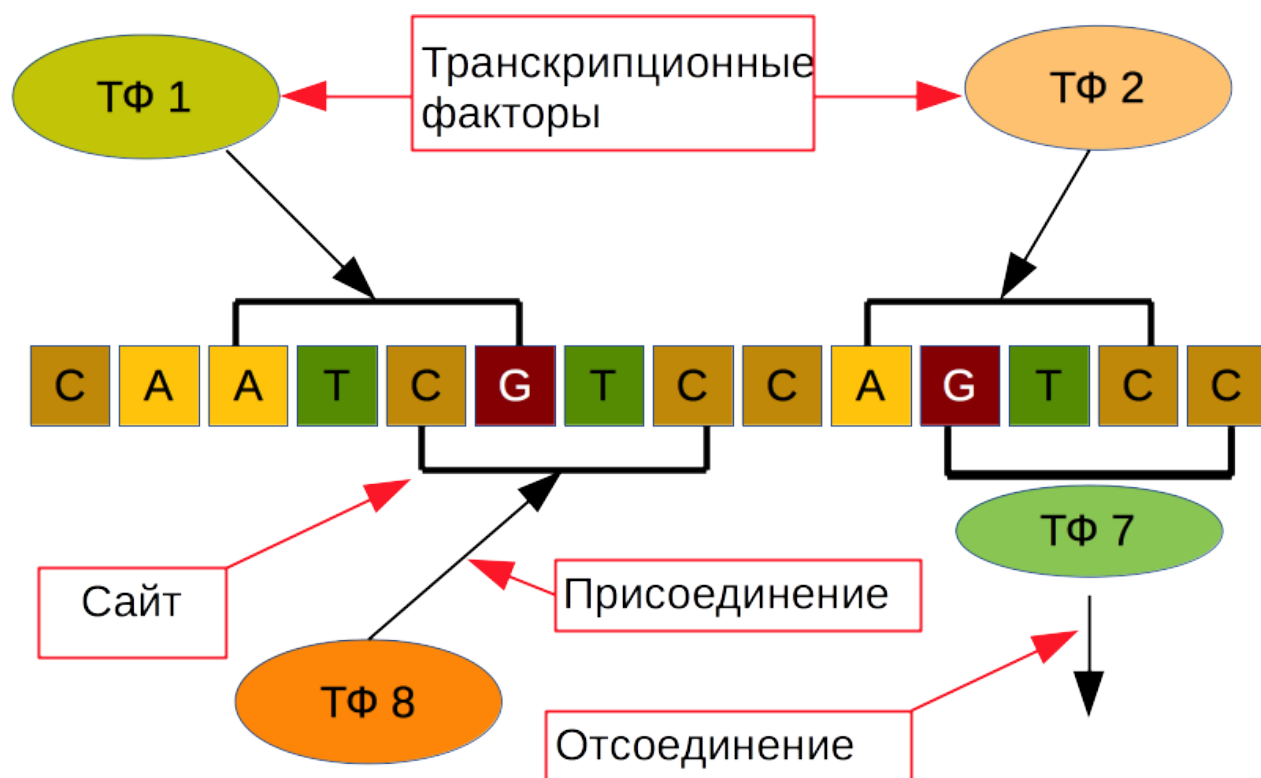
Цель научно-исследовательской работы:

- Описать реакции экспрессии генов
- Разработка программного обеспечения для стохастического моделирования экспрессии генов.
- Анализ экспериментальных расчётов
- Выводы

1 Модель

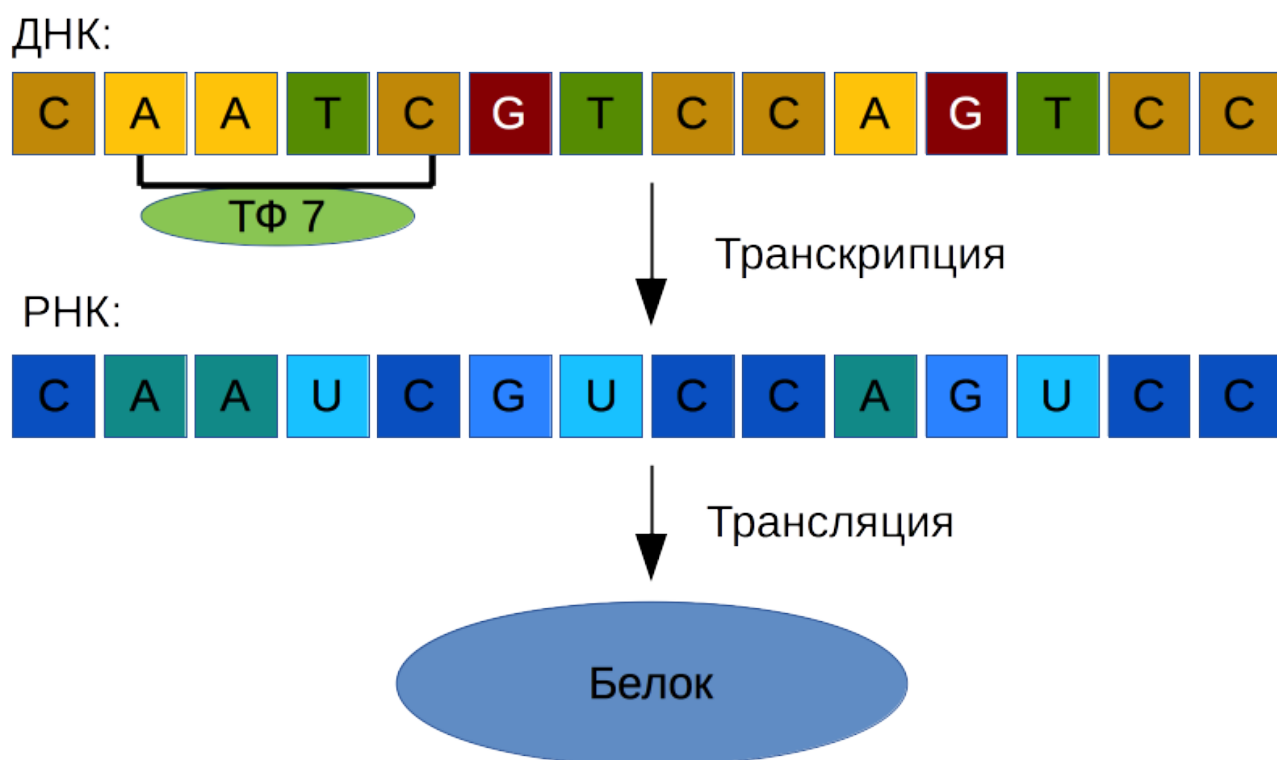
В данной работе будут рассмотрены 5 основных видов реакций в процессе экспрессии генов:

- Присоединение и отсоединение транскрипционных факторов (Рис. 1)
- Начало транскрипции (Рис. 2)
- Трансляция (Рис. 2)
- Деградация белка и тф (Рис. 3)
- Диффузия между ядрами (Рис. 3)



Для каждого сайта определено какой ТФ может присоединиться. Сайтов – 611 и ТФ – 8.

Рисунок 1 – Процесс присоединения и отсоединения ТФ



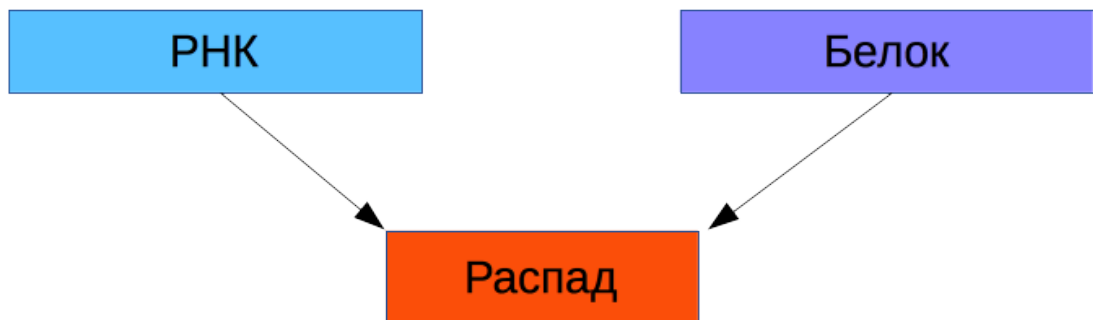
Вероятность транскрипции зависит от количества присоединённых ТФ, трансляции – от количества РНК.

Рисунок 2 – Процесс транскрипции и трансляции

Диффузия:



Деградация:



Вероятности деградации и диффузии линейно зависят от концентраций.

Рисунок 3 – Процесс диффузии и деградации

2 Инструменты для моделирования

Были рассмотрены следующие пакеты для стохастического моделирования - fern[1], StochKit[2], StochPy[3], GillesPy[2]. Каждый из них рассчитан на языки программирования - java, c++, python соответственно. Далее будет каждый пакет рассмотрен более подробно, но наш выбор остановился на StochKit, как на самом быстром.

2.1 *StochPy, GillesPy*

Является обвёрткой над пакетом StochKit с некоторыми модификациями. В комплекте с моделированием есть возможность строить графики. Содержит две модификации алгоритма гиллеспи. На вход не принимает файлы в формате sbml (нужно конвертировать в xml).

2.2 *fern*

Является самым медленным из предложенных пакетов стохастического моделирования. Содержит более семи модификаций алгоритма гиллеспи. На вход принимает файлы в формате sbml, что является большим плюсом. Данный пакет подходит для моделирования небольшого количества реакций.

2.3 *StochKit*

Является самым быстрым, так как написан на C и все остальные пакеты либо опираются на него, либо являются обвёрткой. В комплекте идёт программа, написанная на Matlab, для построения графиков, но к сожалению данные графики неинтересны. В пакете реализовано несколько модификаций алгоритма гиллеспи. Выходом программы StochKit является текстовый файл с концентрациями веществ в любой момент времени (частота задаётся параметром).

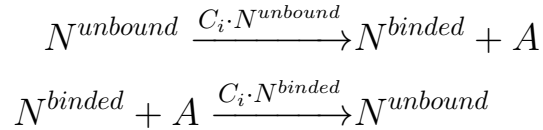
2.4 *DEEP*

Наша модель будет содержать параметры и для моделирования мы воспользуемся генетическими алгоритмами, реализованными в DEEP [4].

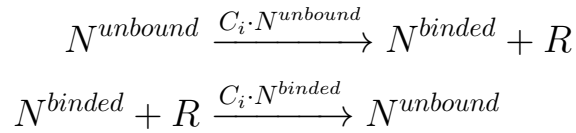
3 Результаты

3.1 Описание реакций

Опишем реакции для одного ядра (для остальных ядер аналогично). Для каждого сайта присоединения реакции присоединения и отсоединения будут иметь одинаковый вид. Основным отличием будет какой именно ТФ присоединяется/отсоединяется. Для активаторов (в нашей работе это **cad** или **bcd**) реакции примут следующий вид:

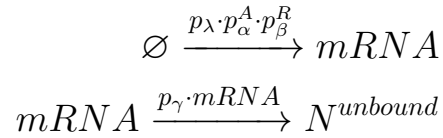


Для репрессоров реакции будут выглядеть так:



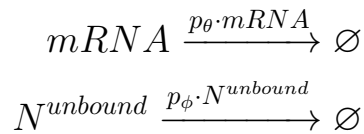
где C_i - коэффициент присоединения к i -ому сайту, R - количество репрессоров и A - количество активаторов.

Теперь запишем реакции транскрипции и трансляции:



где p_λ , p_α , p_β , p_γ - коэффициенты транскрипции, активации, регрессии и трансляции соответственно.

Реакции деградации:



где p_θ , p_ϕ - коэффициенты деградации мРНК и белка соответственно.

Так же нужно учесть диффузию - перетекание белка в соседние ядра:

$$N_i \xrightarrow{p_\xi \cdot \max(N_i - N_{i+1}, 0)} N_{i+1}$$

$$N_i \xrightarrow{p_\xi \cdot \max(N_i - N_{i-1}, 0)} N_{i-1}$$

где p_ξ - параметр диффузии, i - номер ядра. В результате наша модель имеет 7 параметров - p_λ , p_α , p_β , p_γ , p_θ , p_ϕ , p_ξ .

3.2 Разработка программного обеспечения

Пакет был разделён на четыре компоненты:

- Генерация системы реакций для пакета StochKit с помощью языка программирования Python
- Стохастическое моделирование с помощью пакета StochKit
- Оптимизация параметров модели с помощью DEEP
- Анализ результатов с помощью языка программирования R

Основной трудностью было наладить работу DEEP совместно с StochKit. Для этого нужно было изменить исходники StochKit, чтобы после каждого запуска он выводил ошибку моделирования при заданных параметрах.

3.3 Анализ результатов

В силу ограниченности возможности работы на кластере, было произведено несколько стартов. Было построены графики зависимости концентрации белка **kni** от ядра к концу моделирования (в нашем случае это 60 минута).

На графиках (4, 5) отчётливо видно, что поведение концентраций у белка **kni** и мРНК схожи с экспериментальными расчётами.

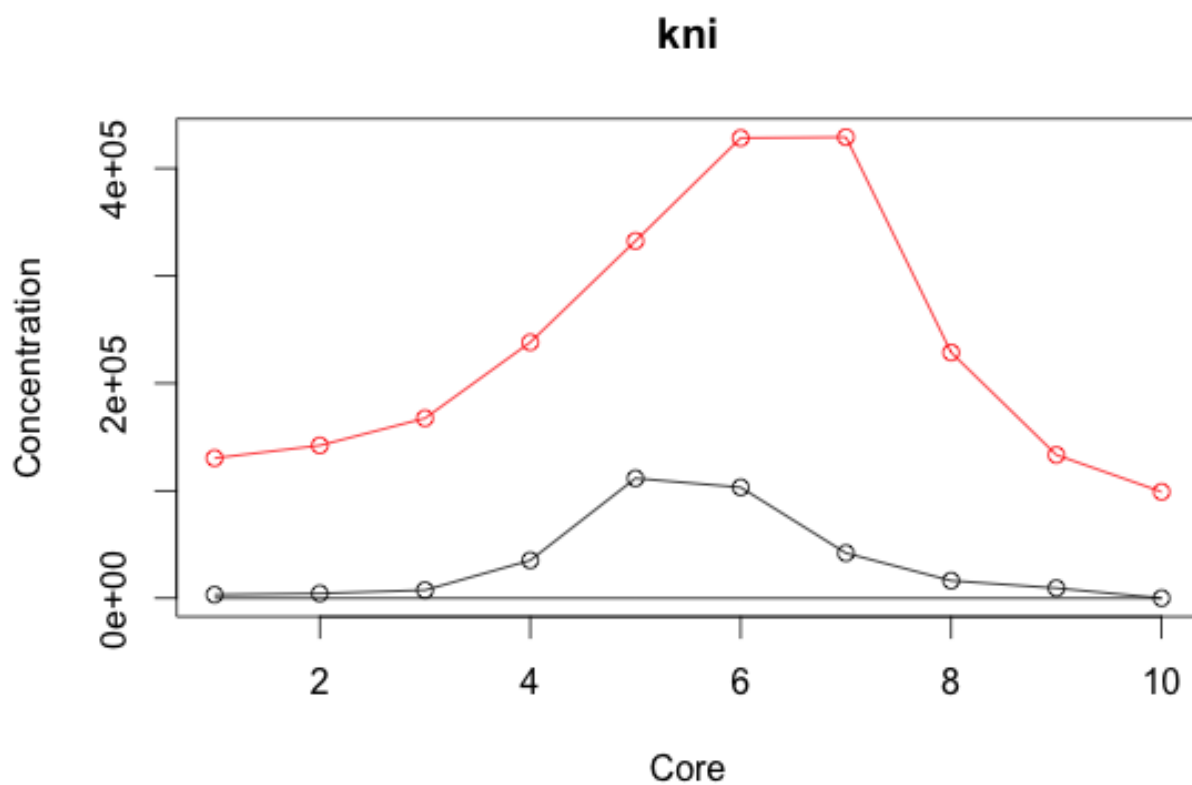


Рисунок 4 – Зависимость концентрации белка kni от ядра. Чёрным цветом - результат моделирования, красный цвет - экспериментальный расчёты

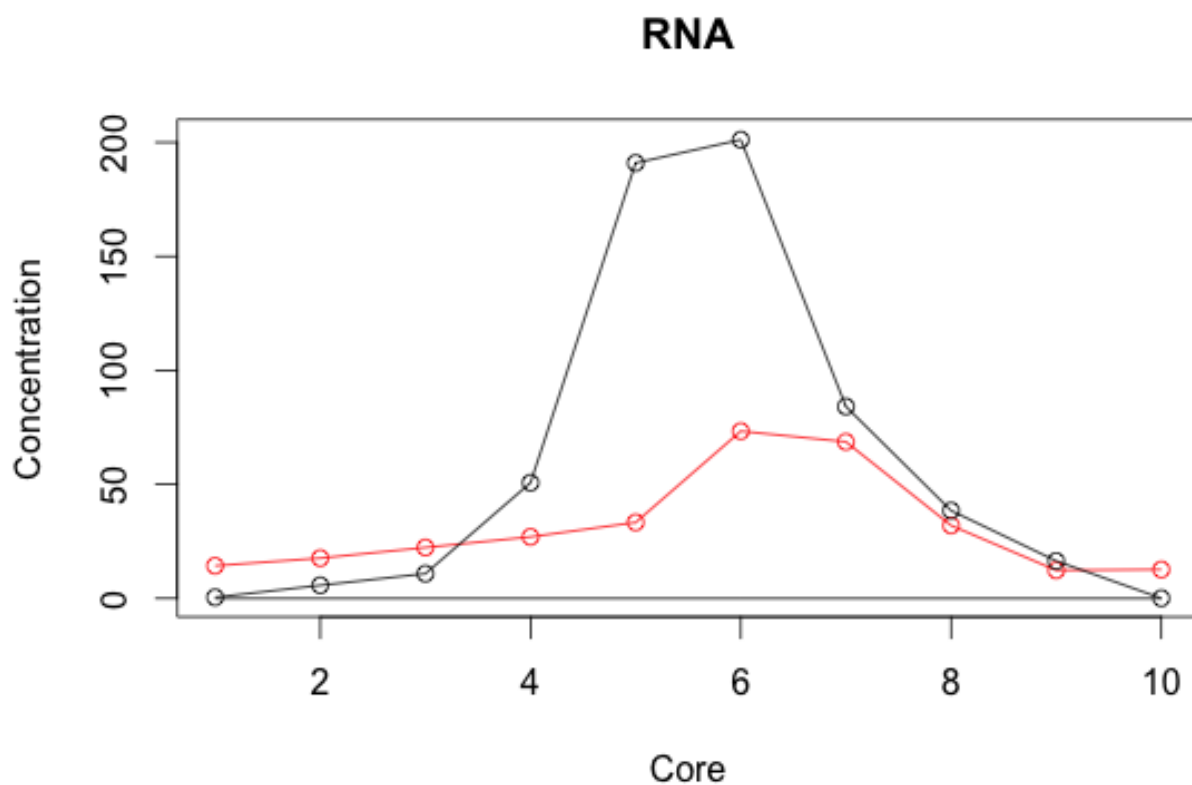


Рисунок 5 – Зависимость концентрации мРНК от ядра. Чёрным цветом - результат моделирования, красный цвет - экспериментальные расчёты

ВЫВОДЫ

Основной сложностью моделирования является вычислительная сложность, так как в сумме получается нужно рассчитывать вероятности для 12404 реакций. Но даже при малом времени работы DEEP удалось подобрать параметры модели, чтобы концентрации белка **kni** и мРНК имели схожее поведение с экспериментальными расчётами.

СПИСОК ИСПОЛЬЗОВАННЫХ ИСТОЧНИКОВ

- 1 FERN – a Java framework for stochastic simulation and evaluation of reaction networks. — <https://bmcbioinformatics.biomedcentral.com/articles/10.1186/1471-2105-9-356>. — Accessed: 2016-30-04.
- 2 StochKit. — <https://sourceforge.net/projects/stochkit/>. — Accessed: 2016-30-04.
- 3 StochPy. — <http://stochpy.sourceforge.net>. — Accessed: 2016-30-04.
- 4 *Kozlov K., Samsonov A.* DEEP—differential evolution entirely parallel method for gene regulatory networks. — <http://link.springer.com/article/10.1007/s11227-010-0390-6>. — Accessed: 2016-30-04.