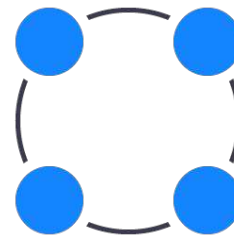


# Поиск генов в LTR-повторах

Студенты: Алешкина Дарья,  
Атаманова Мария

Руководитель: Добрынин Павел,

Центр геномной  
биоинформатики им Ф.Г.  
Добржанского



ИНСТИТУТ  
БИОИНФОРМАТИКИ

Осенний семестр, 2016

# LTR - повторы



- Длинные концевые повторы (Long terminal repeats)
- Расположены на обоих концах ДНК-копии ретровирусов и ретротранспозонов
- Участвуют во внедрении ДНК-копии генома ретровируса в геном клетки-хозяина и регуляции экспрессии генов.

# Роль генов эндогенных ретровирусов

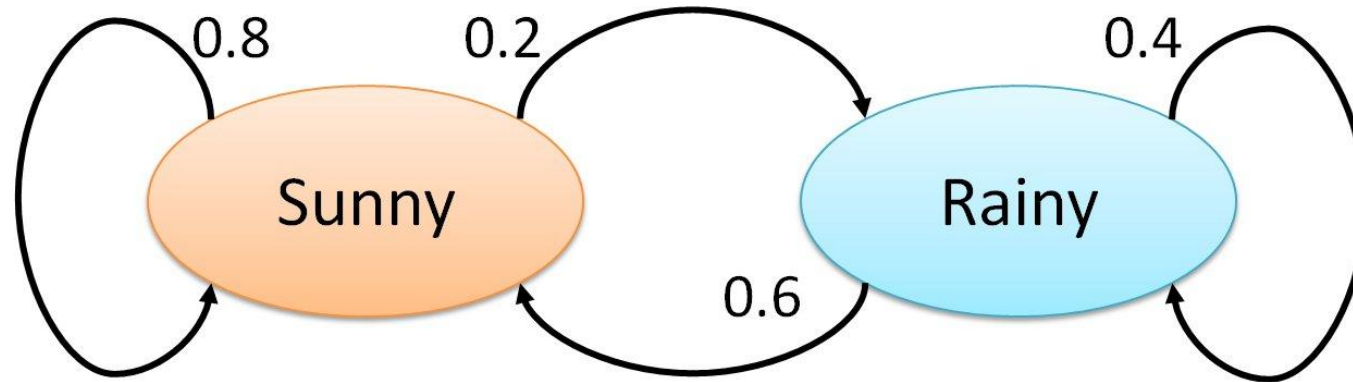
Гены группы env (ENVV1, ENVV2, ERVW-1 и др.) участвуют в формировании плаценты (Dupressoir et al., 2012)

Мутации в гене ERVWE1 (белок syncytin) ассоциированы с рассеянным склерозом и шизофренией (Karlsson et al., 2004; Suntsova et al., 2013; Slokar et al., 2015)

Гены ретровирусов играют роль в иммунных ответах (Patience et al., 1997 и др.)

# Алгоритмы поиска генов

- hidden Markov model (HMM)

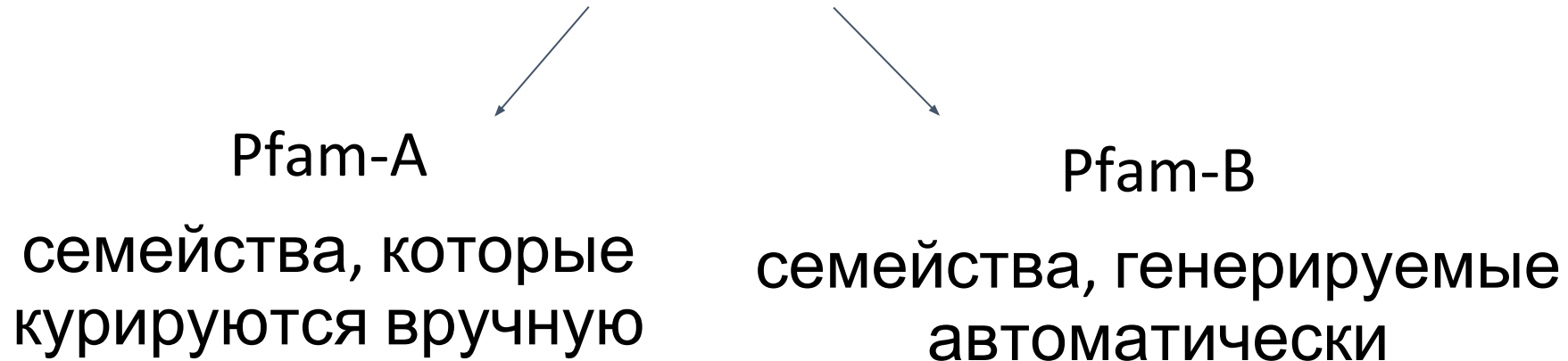


# Программное обеспечение

- GeneMark – S (для поиска генов в последовательностях)
- Interproscan (поиск доменной структуры)
- HMMER с использованием базы данных Pfam-A  
(проверка доменной структуры генов)

# HMMER

- **hmmscan** - ищет последовательности в заданной базе
- **Pfam** - база данных семейств белковых доменов



- **E-value** - показывает, насколько случайно полученное выравнивание, default = 10 (пороговое значение 0.01)

# Цель и задачи

Цель:

Поиск и анализ генов в LTR-повторах

Задачи:

- Провести анализ литературы
- Провести поиск генов в последовательностях, которые соотносятся с ретротранспозонами, с помощью программы GeneMark-S
- Определить доменную структуру найденных генов
- Определить возможные функции найденных генов
- Обобщить результаты

# Результаты

Исходных последовательностей LTR-повторов:  $\approx 26200$

Количество обнаруженных генов:  $\approx 38700$

Количество генов, обладающих различной доменной структурой:  $\approx 4200$

## Фильтры:

- исключение “шума”
- статистическая значимость результатов программ
- литературные источники

```
>> HERV-K_env_2 Retro-transcribing viruses envelope glycoprotein
#   score  bias  c-Evalue  i-Evalue  hmmfrom  hmm to  alifrom  ali to  envfrom  env to  acc
-----
1 !  56.3   0.1   1.8e-19   2.9e-15    2       92 ..    56      145 ..    55      152 ..  0.91
```

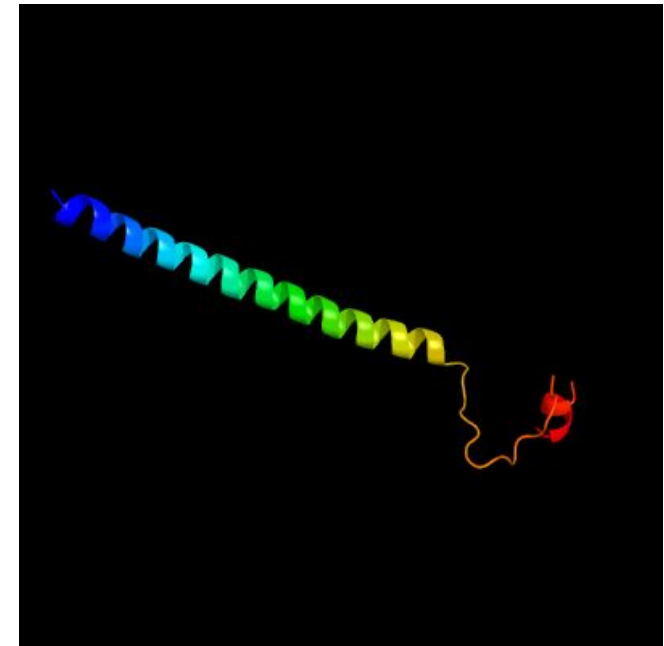
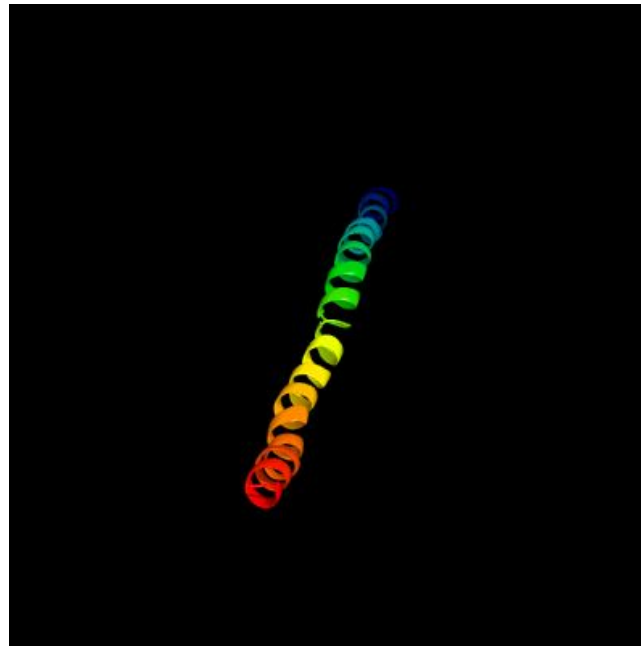


# Результаты

• **Обнаружены гены, связанные с врожденным иммунитетом (Chuong et al., 2016)**

1. STAT\_alpha STAT protein, all-alpha domain
2. STAT6\_C STAT6 C-terminal

**Phyre2 - Protein  
Homology/analogY  
Recognition Engine V 2.0**



# Результаты

- **Гены человеческого эндогенного вируса HERV-K**

1. HERV-K\_env\_2 Retro-transcribing viruses envelope glycoprotein
2. HERV-K\_REC Rec (regulator of expression encoded by corf) of HERV-K-113 (Kassiotis, 2014; Schmitt et al., 2015)

- **Гены, взаимодействующие с приобретенным иммунитетом млекопитающих**

1. MMTV\_SAg Mouse mammary tumour virus superantigen  
Способен заражать не только мышей, но и других млекопитающих (Hsu et al., 2010)

# Результаты

## • Гены млекопитающих связанные с развитием и плацентой

1. PGC7\_Stella PGC7/Stella/Dppa3 domain  
Предотвращает деметилирование импринтированных районов
2. MAGE MAGE family  
Экспрессия генов этого семейства ограничена репродуктивными тканями и опухолями
3. PLAC9 Placenta-specific protein 9 (один домен)

Спасибо за внимание!