

Сборка *de novo* и анализ транскриптомов  
*Platynereis dumerilii* (*Nereididae*, *Annelida*)  
на разных этапах регенерации

Зенкова Наталья, Абасов Руслан

Санкт-Петербургский государственный университет

Руководитель: Нестеренко Максим Алексеевич



Санкт-Петербург  
2018г.

*Регенерация* — восстановление или отращивание клеток, тканей или целых органов — широко, но неравномерно представлена среди различных групп животных.

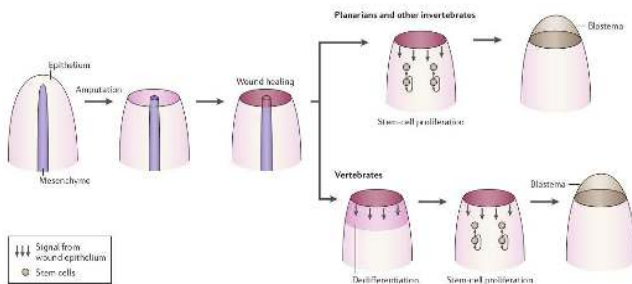


Рис.: Этапы регенерации [Alvarado A.S. and Tsonis P.A., 2006]



- Модельный организм.
- Регенерирует только хвост.
- Живое ископаемое.

## Задача:

- Исследовать различия в молекулярных основах регенерации в «головном» и «хвостовом» сайтах полихеты *P.dumerilii*.

## Вопрос:

- Почему у этой полихеты невозможно восстановление головных структур?

- 1 Проверка качества библиотек прочтений.
- 2 Подготовка прочтений для последующего анализа.
- 3 Сборка транскриптома *de novo* с помощью Trinity.
- 4 Оценка качества и полноты собранного транскриптома.
- 5 Аннотация собранного транскриптома.
- 6 Определение уровней экспрессии генов.
- 7 Сравнение характеров активностей генов между разными временными точками и сайтами регенерации.

- 1 Размер библиотек:

|      |          |
|------|----------|
| min  | 13476893 |
| max  | 17431535 |
| mean | 15151034 |

- 2 Библиотеки характеризуются относительно неплохим качеством.
- 3 После удаления низкокачественных и адаптерных последовательностей «выжило» не менее 95.7% пар прочтений.

```
ILLUMINA_CLIP:$ADAPTERS:2:30:10:2:TRUE  
SLIDINGWINDOW:4:20 MAXINFO:50:0.8 MINLEN:25
```

- 4 После картирования на геном человека (GRCH38) контаминация тканями исследователя в каждой из проанализированных библиотек менее 4%.

## 1 Оценка качества сборки:

- |                         |        |
|-------------------------|--------|
| TransRate score         | 0.2441 |
| TransRate optimal score | 0.3572 |

- 90% собранных контигов являются «хорошими».
- 

BUSCO (vs Metazoa-odb9) = 99.5%.

## 2 TransDecoder:

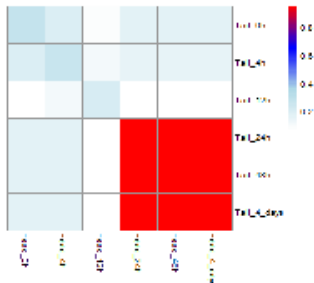
|                                 |        |
|---------------------------------|--------|
| Количество предсказанных белков | 160381 |
|---------------------------------|--------|

## 3 Аннотация белков с помощью Diamond:

|                                |            |
|--------------------------------|------------|
| Количество последовательностей | 97842      |
| E-value                        | 1e-10      |
| База данных                    | Swiss-Prot |

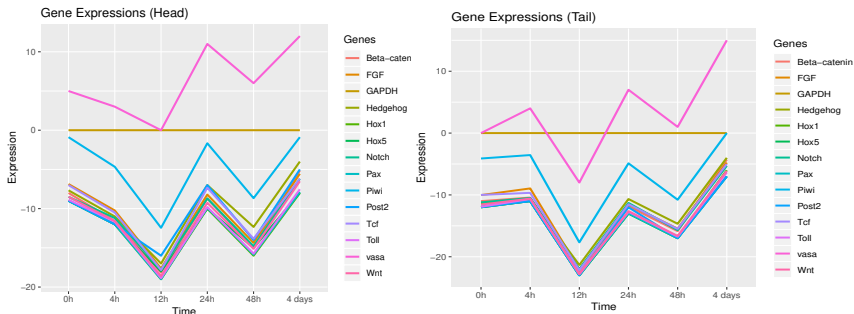
## 4 Картирование с помощью Salmon.

- Ассоциированные с образцом последовательности — это последовательности, которые экспрессируются не во всех образцах и имеют ненулевую экспрессию в анализируемом образце.
- Сравнение множеств последовательностей произведено с помощью критерия Жаккара.
- Наибольшие значения критерия принимает при сравнении данных по последним 3 временным точкам.





Графики значений экспрессий генов (для головы и хвоста), отцентрированных относительно референса и представленных в TPM-единицах:



- Ось x — временные точки (0, 4, 12, 24, 48 часов и 4 дня);
- Ось y — уровень экспрессии генов.

## Результаты:

- 1 Собранный транскриптом характеризуется высокими показателями качества и полноты сборки.
- 2 Выделенные множества «ассоциированных» последовательностей у разных временных точек перекрываются не полностью:
  - это может указывать на активность различных процессов;
  - значимые совпадения между вр.точками — 24, 48 часов и 4 дня — могут быть обусловлены преобразованием тканей после повреждений.
- 3 Для рассматриваемой группы генов можем отметить повышение их уровней активности на таких временных точках, как 12, 28, 48 часов и 4 дня.

Данные результаты могут быть использованы для дальнейшей экспериментальной работы.

Спасибо за внимание!