

Whole-genome *Drosophila* sequence analysis - 2

Косолапова Анастасия
Лебеда Юрий

Руководитель: Захаров Геннадий, EPAM Systems

- *D.melanogaster* - модельный объект
- Результаты изучения генома дрозофилы важны для понимания работы схожих генов у млекопитающих, в том числе, человека
- Линия agnostic потенциально может использоваться для изучения синдрома Вильямса у человека



Исходные данные

- Три линии дикого типа:
 - Canton-S (**CS**)
 - Berlin (**Ber**)
 - Oregon-R (**Or**)
- Два мутанта с нарушениями в строении и функционировании нервной системы:
 - **agnostic^{ts3}**(ts3) - дефекты обучения и памяти, восстанавливаются до дикого типа после теплового шока(37°C, 30 мин)
 - **agnostic^{x1}**(X1) - дефекты обучения и памяти, локомоторного поведения. Гистологические дефекты структуры мозга. Длительное время впадения в эфирный наркоз.

Цели и задачи работы






Цель: Обнаружить мутации, ответственных за формирование мутантных фенотипов

Задачи (план работы):

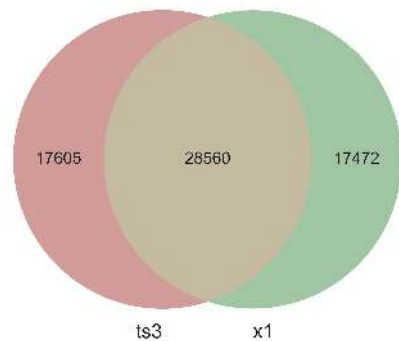
- провести базовый контроль качества ридов
- найти вариации для всех имеющихся линий
- отфильтровать полученные вариации
- установить различия между геномами диких и мутантных линий
- определить мутации, ответственные за формирование мутантных фенотипов

Все стадии включаются в общий пайплайн в виде файла Snakefile

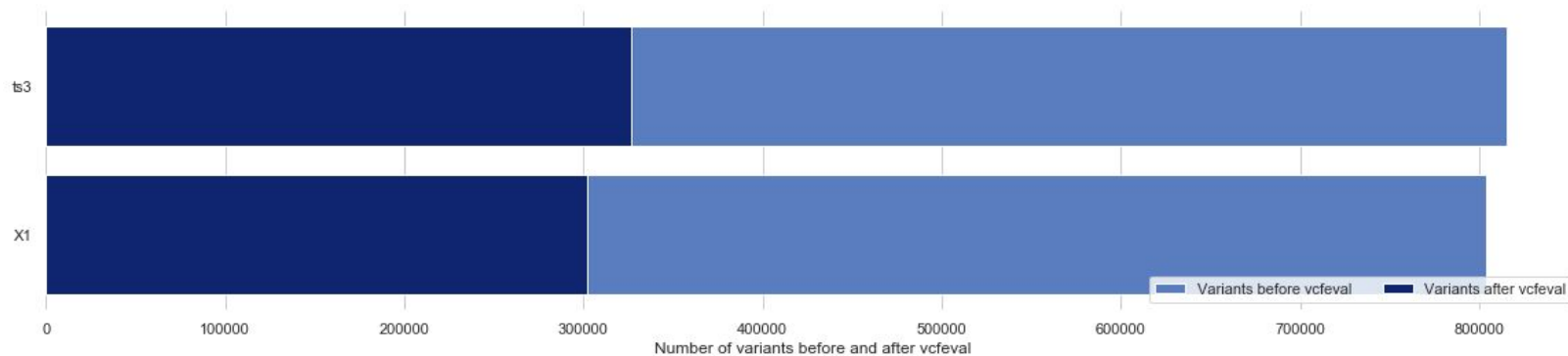
Структура пайплайна

	FastQC, Trimmomatic, BWA, GATK	Выравнивание, нахождение вариаций, их фильтрация
	picard-tools	Получение метрик выравнивания
	rtg-tools vcfEval	Установление различий между линиями
	snpEff	Аннотирование вариаций
	snpSift	Фильтрация незначущих вариаций

Вариации в ts^3 и x^1



	ts^3	x^1
До Vcfeval	815608	803690
После Vcfeval	326780	302019
Из них в chrX	46165	46032



Вариации: ген LIMK1 (без фильтрации)

X¹

Интроны: 6 вариаций, из них

- 2 indel
- 4 SNP

Экзоны:

- 2 SNP (синонимичные)

ts³

Интроны: 7 вариаций, из них

- 2 indel
- 5 SNP

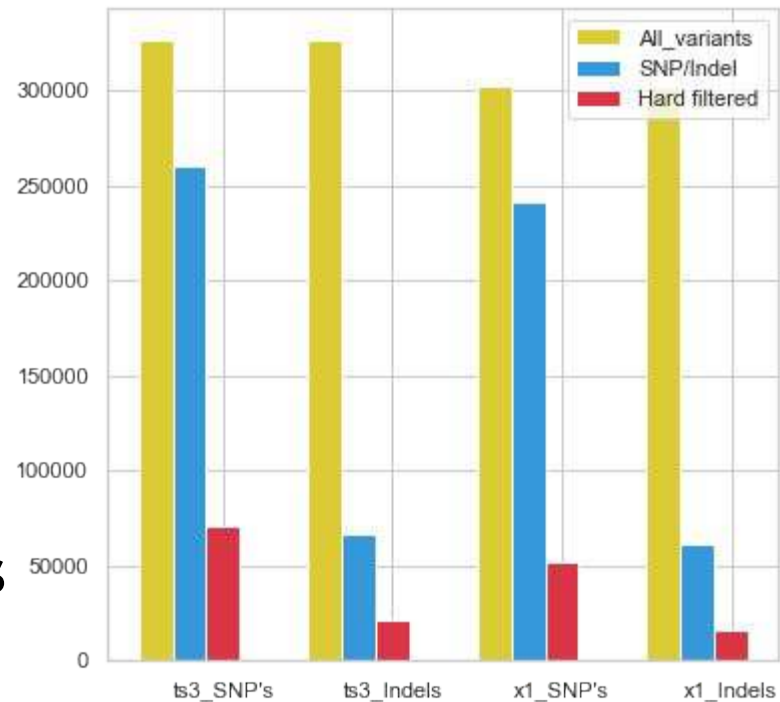
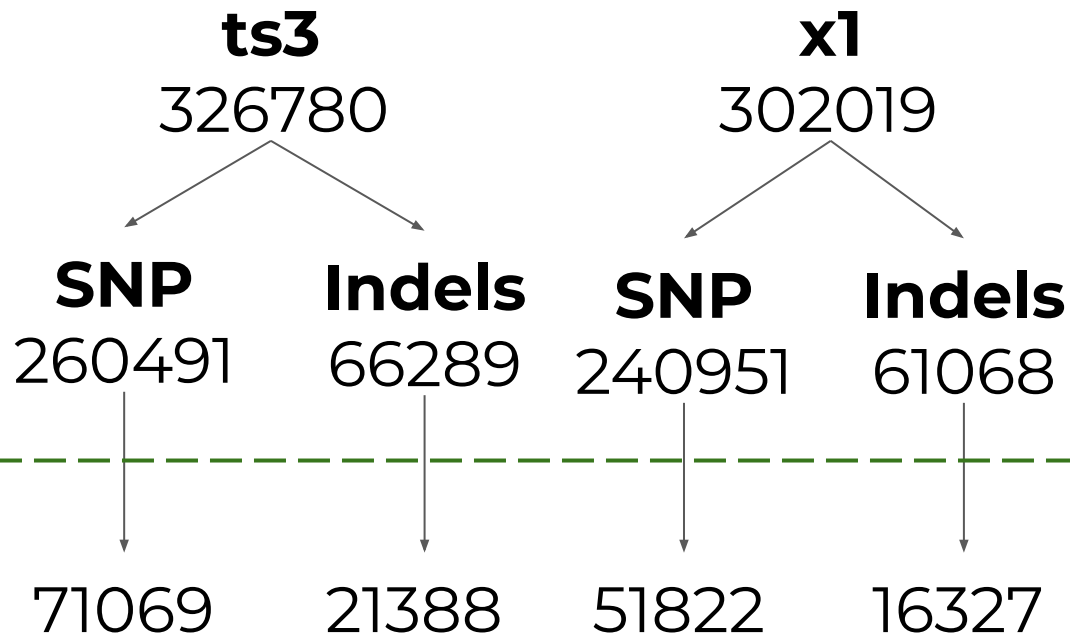
Экзоны:

- 2 SNP (синонимичные)

3-UTR: 1 indel

Вариации в X¹ полностью перекрываются с вариациями в ts³

Вариации: Hard filtering



Вариации: Гомозиготные варианты

Impact				
ts3	High	Moderate	Low	Modifier
Indels	5	8	12	2586
SNPs	4	182	526	4349

Impact				
x1	High	Moderate	Low	Modifier
Indels	5	14	8	1931
SNPs	3	121	434	3363

#Результаты

1. Для каждой из мутантных линий (X_1 , ts_3) получены списки вариаций
2. Данные вариации проаннотированы и профильтрованы
3. Обнаружены мутации в гене LIMK1, полностью перекрывающиеся у двух мутантных линий (до фильтрации)
4. Профильтрованные вариации разделены на группы по значимости для ручного анализа
5. Фильтрацию результатов при помощи GATK на основе машинного обучения сделать не удалось (для *D.melanogaster* отсутствуют качественные наборы вариаций)
6. Весь пайплайн выполнен в Snakemake



Ссылка на проект: https://github.com/IuriyLeb/drosophila_project