



Презентация научно-исследовательской работы:

«Полногеномный анализ вариантов нуклеотидной последовательности человека способных влиять на прохождение сплайсинга»

Гатупов Михаил
группа 6113



Цель работы

Провести полногеномный поиск вариантов нуклеотидной последовательности, способных по-разному влиять на прохождение сплайсинга в зависимости от геномного окружения.

.



Актуальность работы

- Сплайсинг мРНК играет важную роль в процессе создания белка в клетке. Нарушение сплайсинга зачастую приводит к созданию белка, неспособного правильно выполнять свою функцию, что в свою очередь, приводит к различным заболеваниям.
- Основным инструментом исследования влияния мутации в гене на сплайсинг в настоящее время является биоинформатика
- В этом году на основе алгоритмов машинного обучения был создан инструмент SpliceAI для анализа влияния мутаций на прохождение сплайсинга.

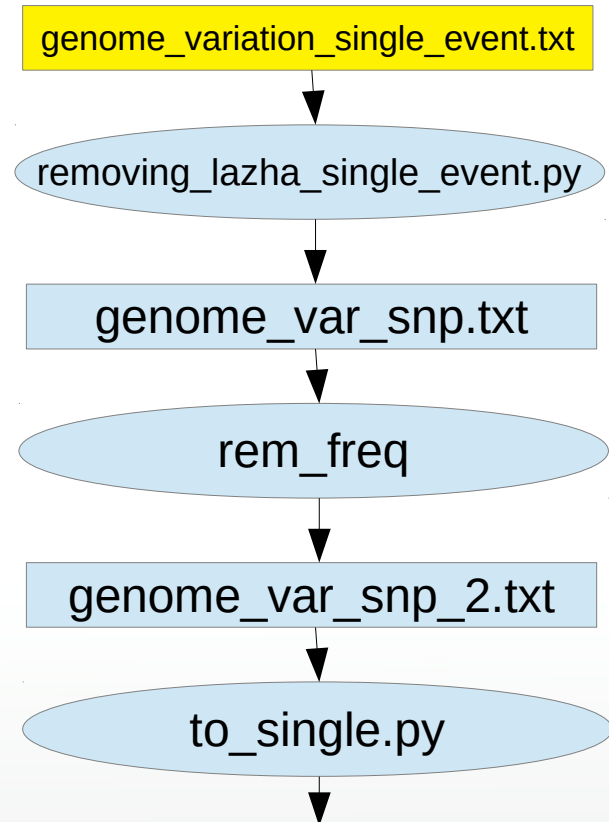


Поставленные задачи

- Создание альтернативной версии генома человека на основе частых однонуклеотидных полиморфизмов.
- Проведение глубокого мутагенеза последовательностей генов человека *in silico*.
- Оценка предсказаний прохождения сплайсинга в различном геномном окружении.

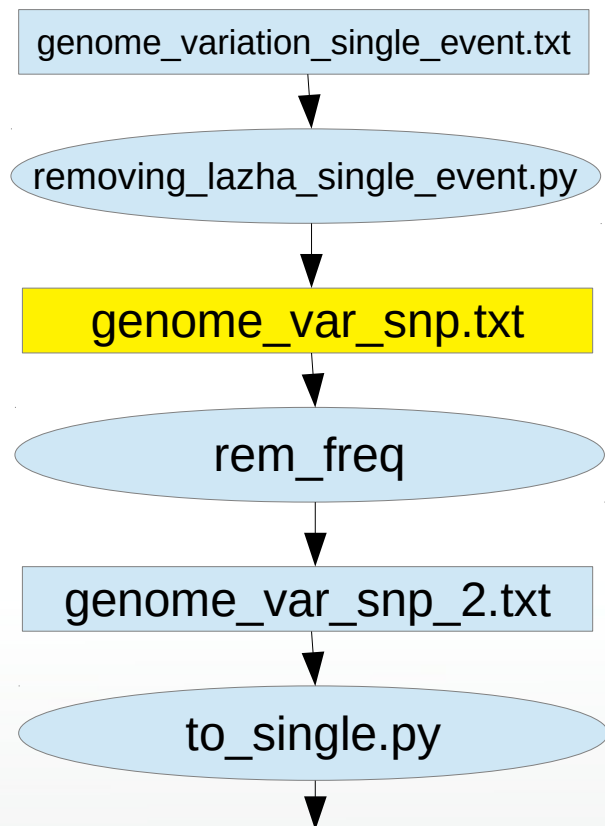
Задача 1: Создание альтернативной версии генома

- Исходный файл
- db_SNP (version 151)
с полиморфизмами
генома человека



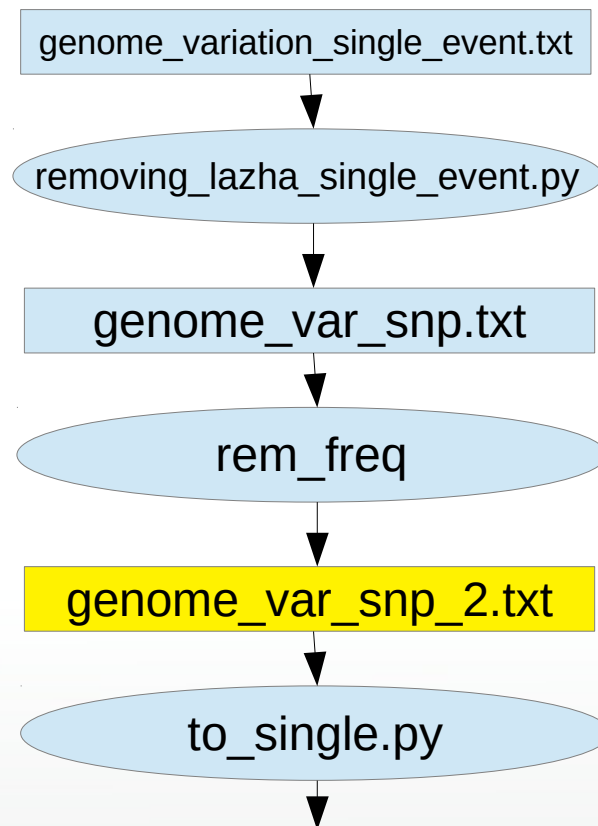
Создание альтернативной версии генома

- Убираем ненужные столбцы
- Делаем из столбца alleles два столбца
(A/G → A G)



Создание альтернативной версии генома

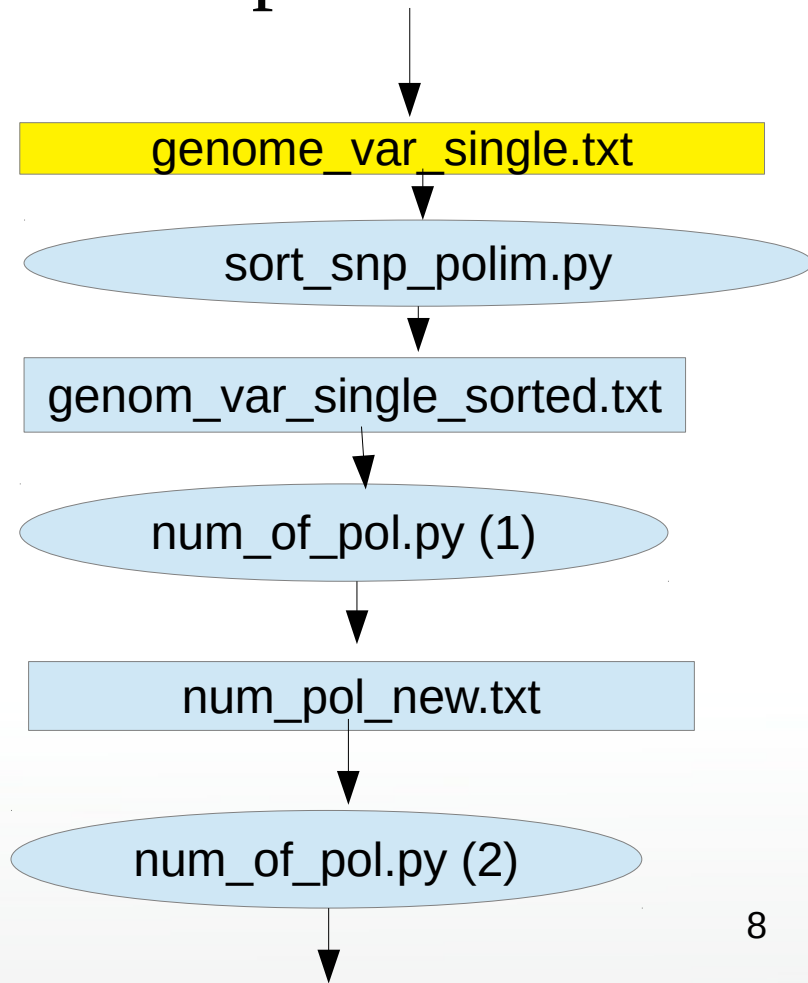
- Убираем слишком редкие полиморфизмы. Было отобрано 11,8 миллионов полиморфизмов.
(alleleFreqs > 0,01)



Создание альтернативной версии генома

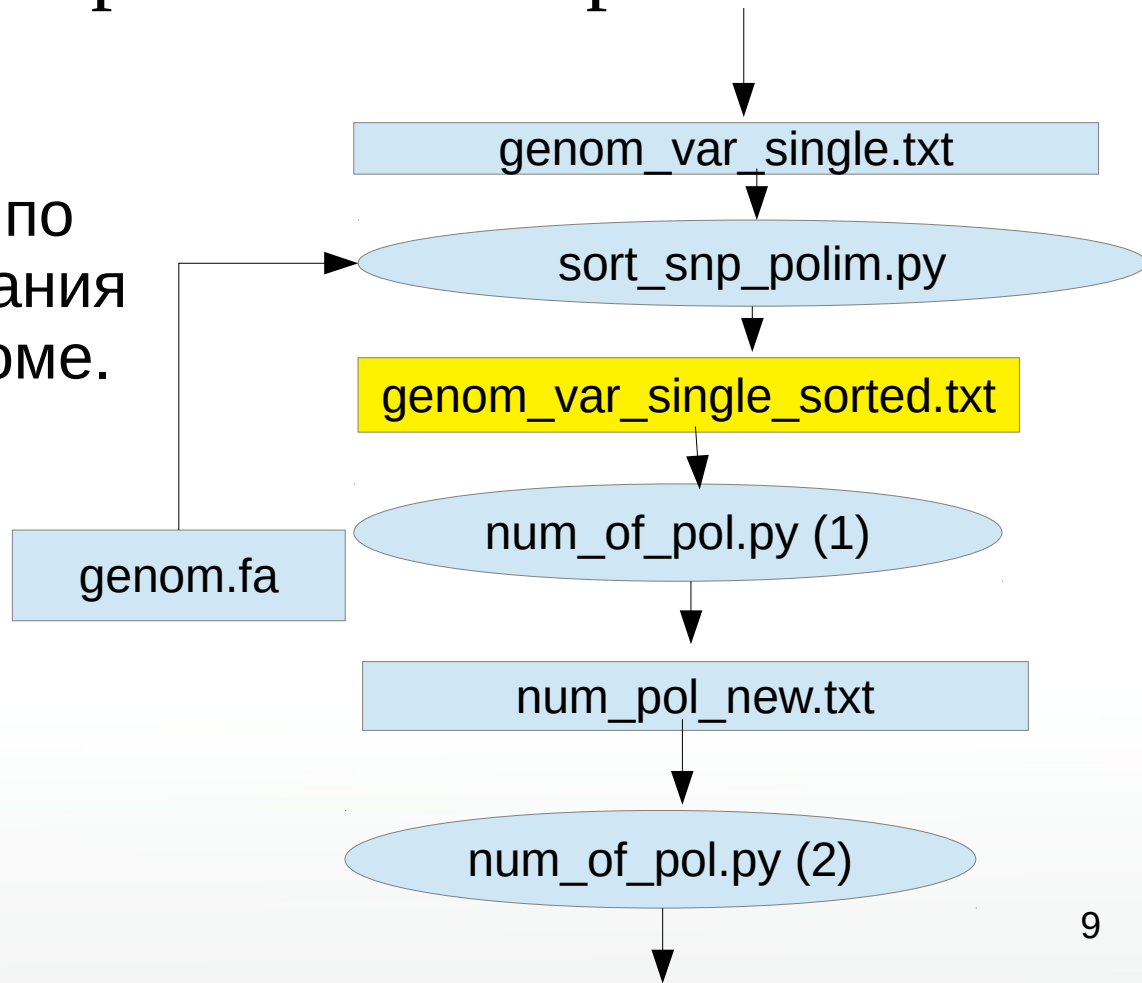
- Убираем все инсерции и делеции
(Используем столбцы ref и alt. В них должно быть по одному смысловому символу)

Было отобрано 10.5 миллионов полиморфизмов



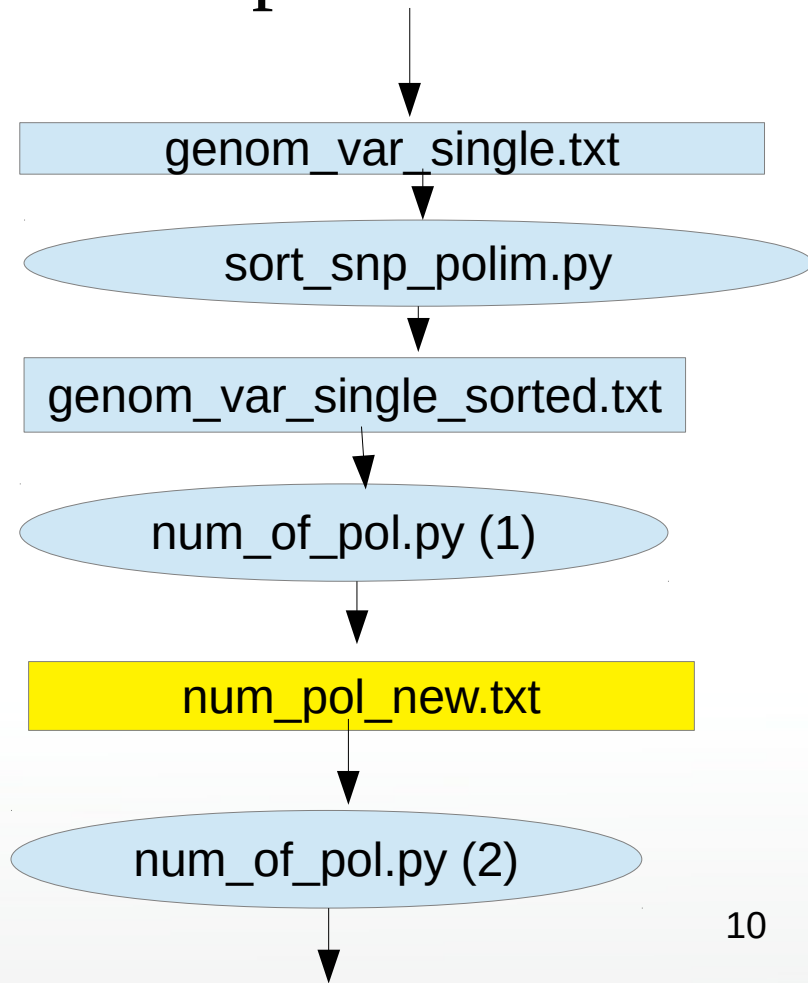
Создание альтернативной версии генома

- Сортируем полиморфизмы по порядку следования хромосом в геноме.



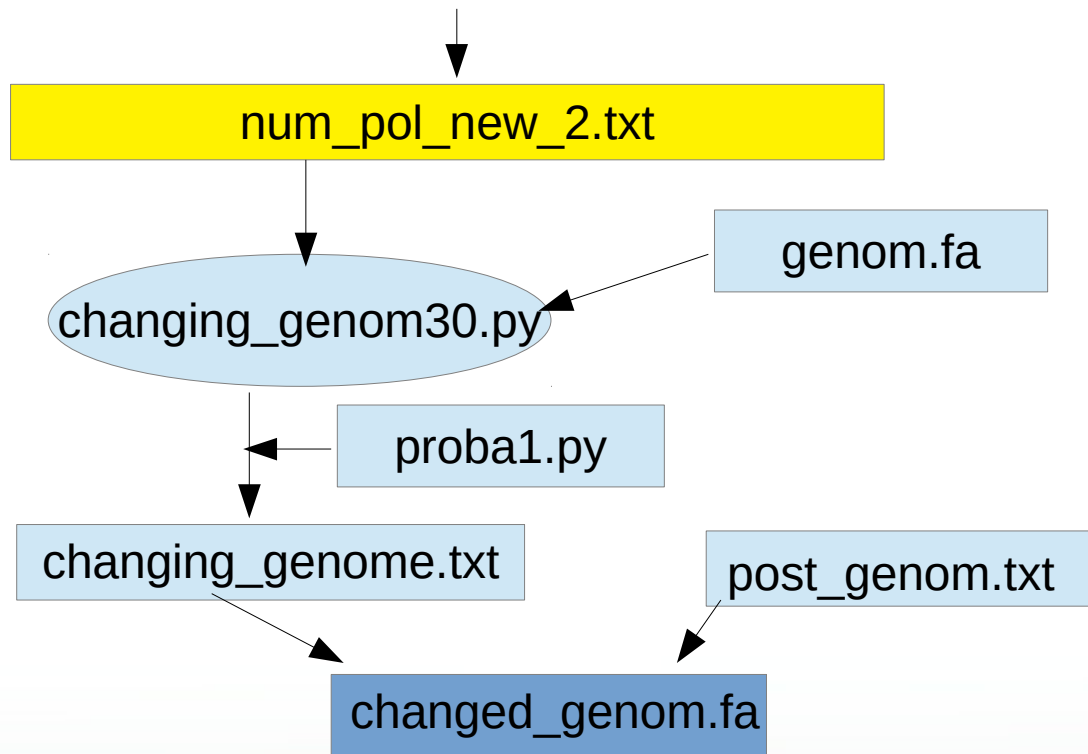
Создание альтернативной версии генома

- Убираем всю лишнюю информацию — подготавливаем файл для замены оснований на месте полиморфизмов в геноме. (1 этап)



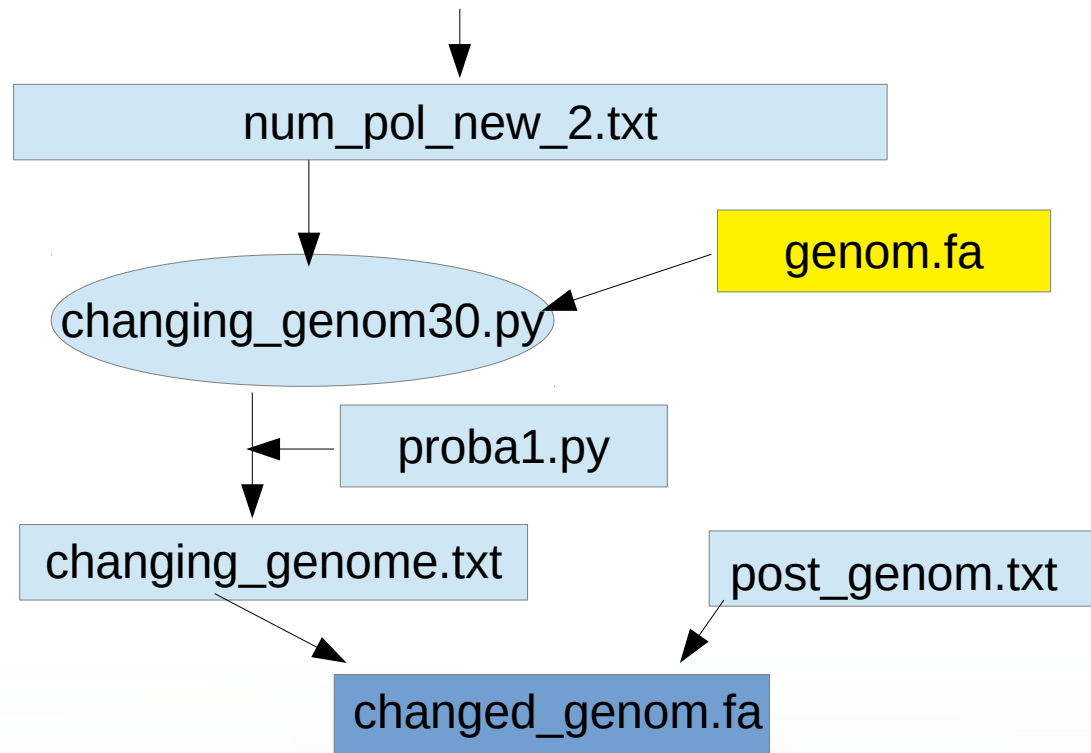
Создание альтернативной версии генома

- Убираем всю лишнюю информацию — подготавливаем файл для замены оснований на месте полиморфизмов в геноме. (2 этап)



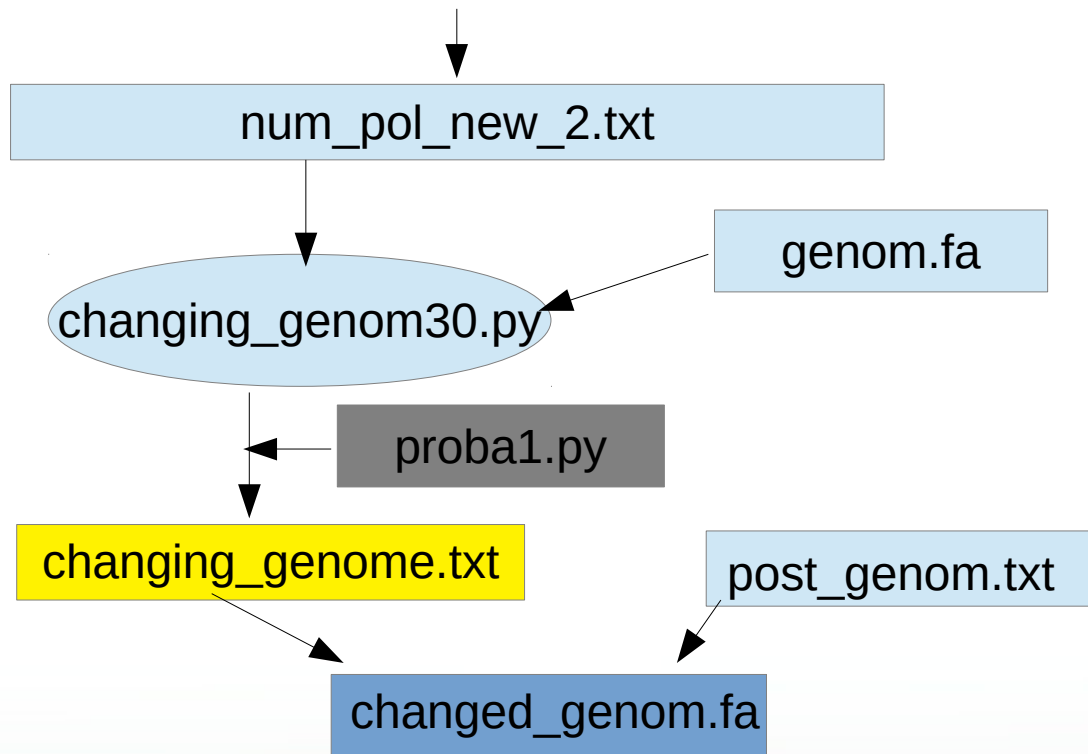
Создание альтернативной версии генома

Последовательность генома человека hg19. В ней будут заменены основания, стоящие в позициях полиморфизмов



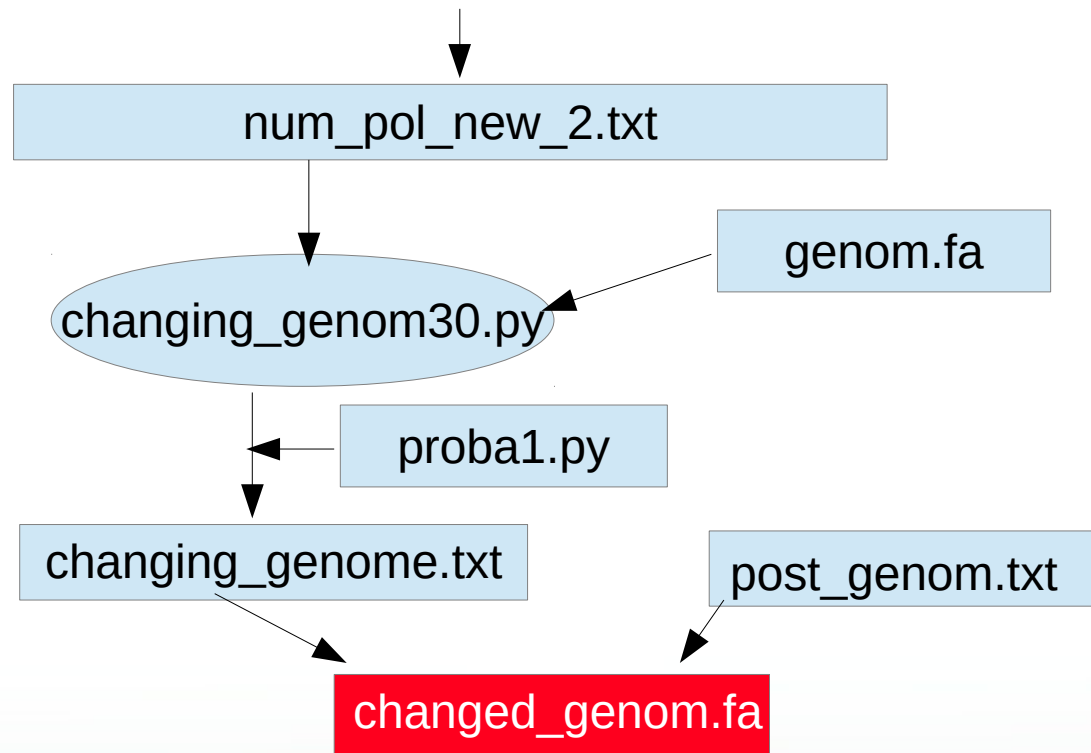
Создание альтернативной версии генома

- Получим геном человека с внесенными изменениями.
Проверим правильность его составления.



Создание альтернативной версии генома

В альтернативный геном были добавлены последовательности, принадлежность которых к определенной хромосоме не установлена. Они оставлены неизменными.

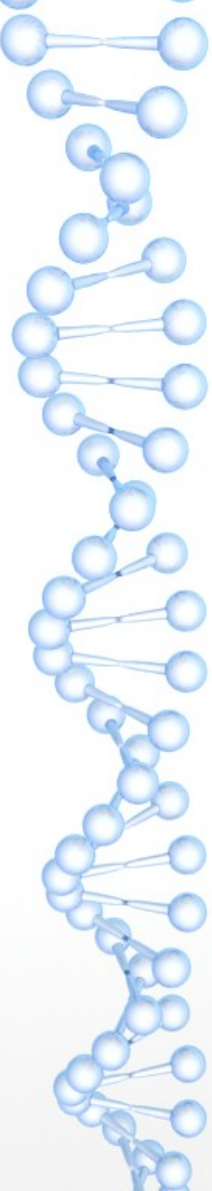




Итог выполнения 1 задачи

- Из базы данных db_SNP (version 151) Было получено 11,8 млн SNP полиморфизмов с частотой более 1 %. Среди этих полиморфных вариантов были отобраны только однонуклеотидные замены. (их кол-во 10,5 млн).
- На основании полученных полиморфизмов и генома hg19 был получен альтернативный геном.

Задача 2: Глубокий мутагенез in silico

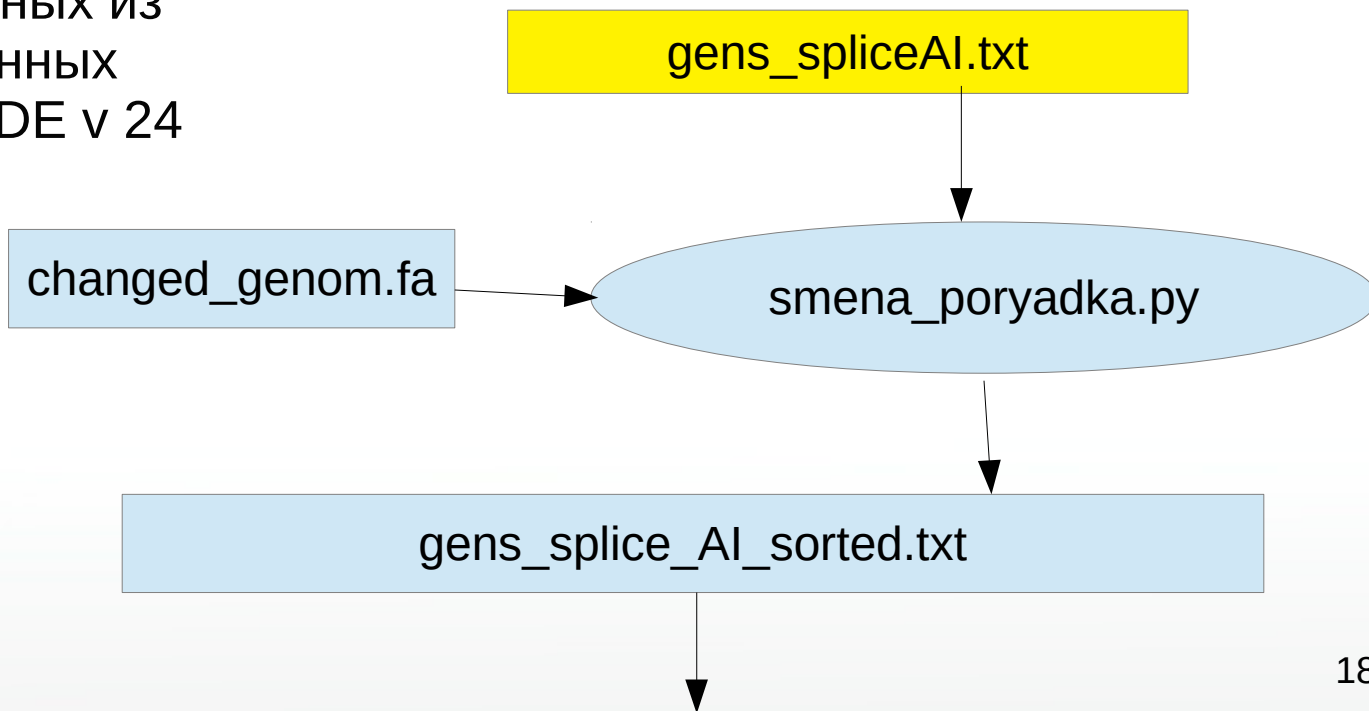


Терминал ▾

Файл	Правка	Вид	Поиск	Терминал	Справка
#CHOM	POS	ID	REF	ALT	
chr1	69090	.	T	A	
chr1	69090	.	T	G	
chr1	69090	.	T	C	
chr1	69091	.	A	T	
chr1	69091	.	A	G	
chr1	69091	.	A	C	
chr1	69092	.	T	A	
chr1	69092	.	T	G	
chr1	69092	.	T	C	
chr1	69093	.	G	A	
chr1	69093	.	G	T	
chr1	69093	.	G	C	
chr1	69094	.	G	A	
chr1	69094	.	G	T	
chr1	69094	.	G	C	
chr1	69095	.	T	A	
chr1	69095	.	T	G	
chr1	69095	.	T	C	
chr1	69096	.	G	A	
chr1	69096	.	G	T	
chr1	69096	.	G	C	

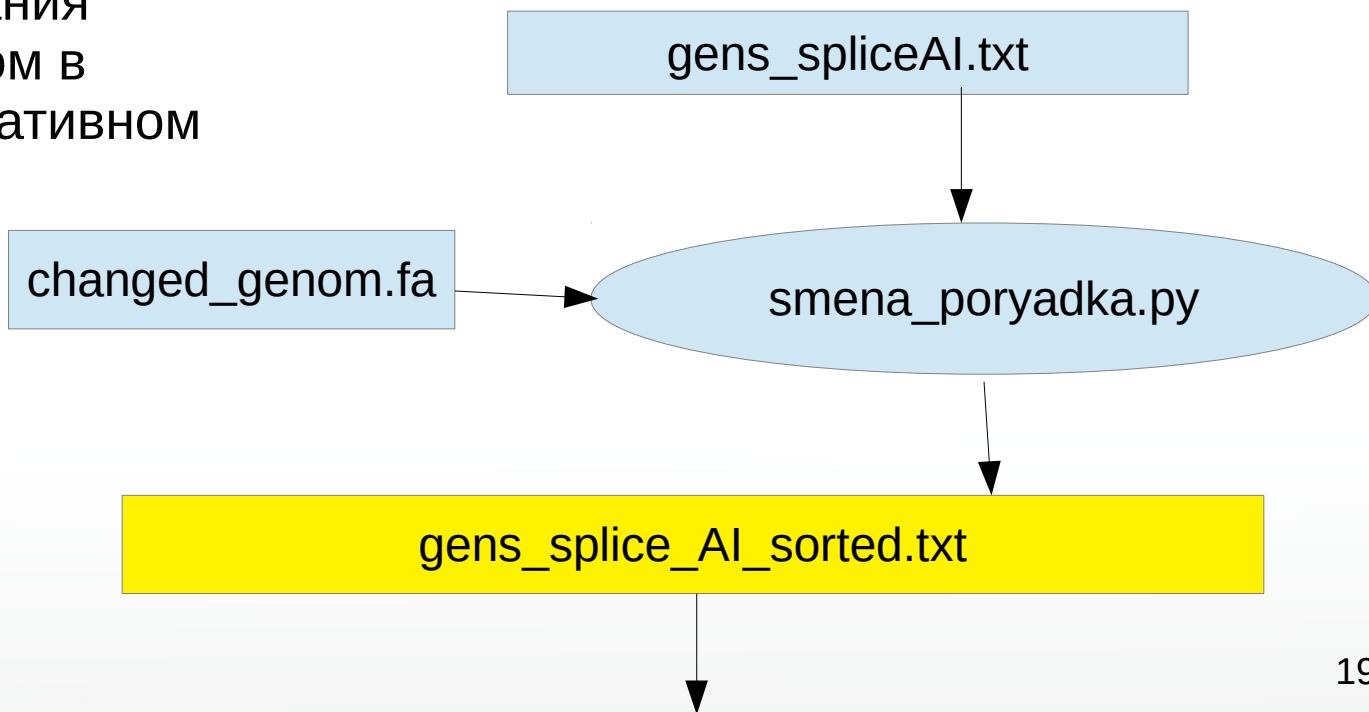
Глубокий мутагенез in silico

- Список 20274 генов человека, полученных из базы данных GENCODE v 24



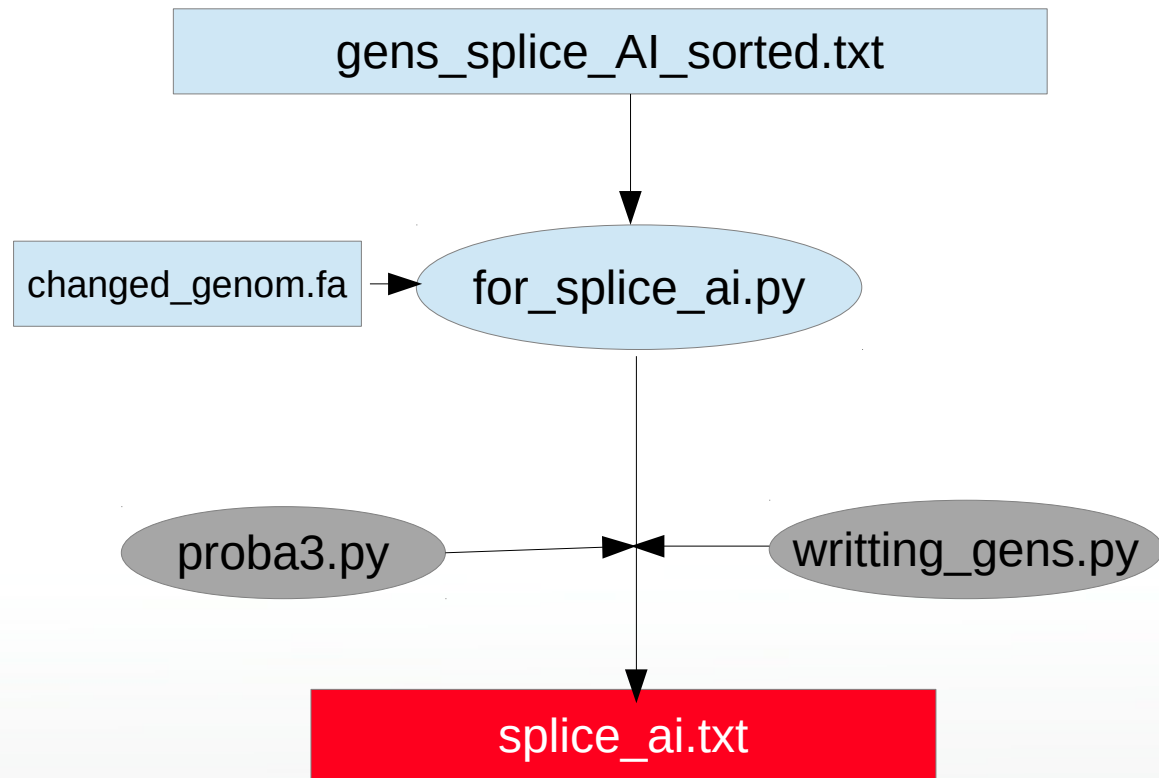
Глубокий мутагенез in silico

- Сортируем гены по порядку следования хромосом в альтернативном геноме.

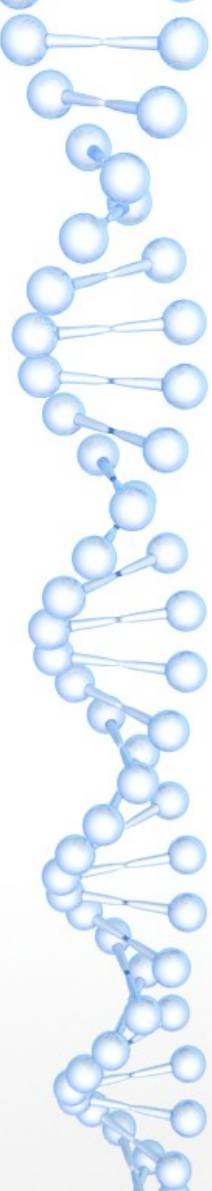


Глубокий мутагенез in silico

- Получаем итоговый файл для SpliceAI. Проверяем правильность его составления.



Полученный файл



Терминал gatupov@gatupov

Файл	Правка	Вид	Поиск	Терминал	Справка				
#CHOM	POS	ID	REF	ALT	QUAL	FILTER	INFO	FORMAT	SAMPLE
chr1	69090	.	T	A	100	PASS	MQ=50	AF:GT	0.5:0/1
chr1	69090	.	T	G	100	PASS	MQ=50	AF:GT	0.5:0/1
chr1	69090	.	T	C	100	PASS	MQ=50	AF:GT	0.5:0/1
chr1	69091	.	A	T	100	PASS	MQ=50	AF:GT	0.5:0/1
chr1	69091	.	A	G	100	PASS	MQ=50	AF:GT	0.5:0/1
chr1	69091	.	A	C	100	PASS	MQ=50	AF:GT	0.5:0/1
chr1	69092	.	T	A	100	PASS	MQ=50	AF:GT	0.5:0/1
chr1	69092	.	T	G	100	PASS	MQ=50	AF:GT	0.5:0/1
chr1	69092	.	T	C	100	PASS	MQ=50	AF:GT	0.5:0/1
chr1	69093	.	G	A	100	PASS	MQ=50	AF:GT	0.5:0/1
chr1	69093	.	G	T	100	PASS	MQ=50	AF:GT	0.5:0/1
chr1	69093	.	G	C	100	PASS	MQ=50	AF:GT	0.5:0/1
chr1	69094	.	G	A	100	PASS	MQ=50	AF:GT	0.5:0/1
chr1	69094	.	G	T	100	PASS	MQ=50	AF:GT	0.5:0/1
chr1	69094	.	G	C	100	PASS	MQ=50	AF:GT	0.5:0/1
chr1	69095	.	T	A	100	PASS	MQ=50	AF:GT	0.5:0/1
chr1	69095	.	T	G	100	PASS	MQ=50	AF:GT	0.5:0/1
chr1	69095	.	T	C	100	PASS	MQ=50	AF:GT	0.5:0/1
chr1	69096	.	G	A	100	PASS	MQ=50	AF:GT	0.5:0/1
chr1	69096	.	G	T	100	PASS	MQ=50	AF:GT	0.5:0/1
chr1	69096	.	G	C	100	PASS	MQ=50	AF:GT	0.5:0/1



Итог выполнения 2 задачи

- Из базы GENCODE v 24 были получены координаты проаннотированных 20274 генов человека.
- Для каждой нуклеотидной позиции генов человека были сгенерированы все возможные альтернативные варианты нуклеотидной последовательности (1.25 млрд вариантов)



Задачи следующего семестра

- Используя последовательность альтернативного генома и данные о глубоком мутагенезе с помощью SpliceAI получить предсказания для вариантов нуклеотидной последовательности на прохождение сплайсинга.
- Сравнить полученные данные с данными, полученными на основе референсного генома и найти варианты нуклеотидной последовательности, способные по-разному влиять на прохождение сплайсинга в зависимости от геномного окружения.



Спасибо за внимание!