

Statistical analysis of annotated genomes

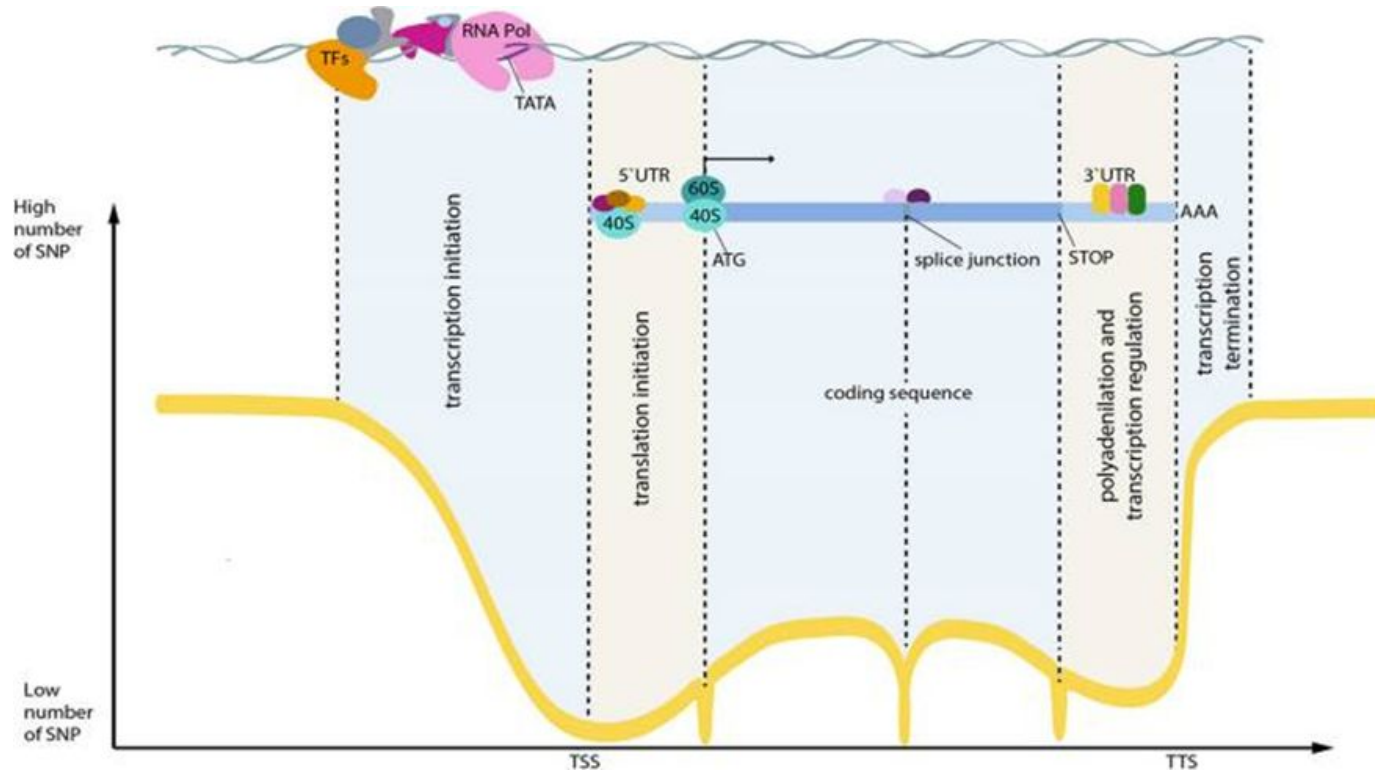


Poleshchuk Olga

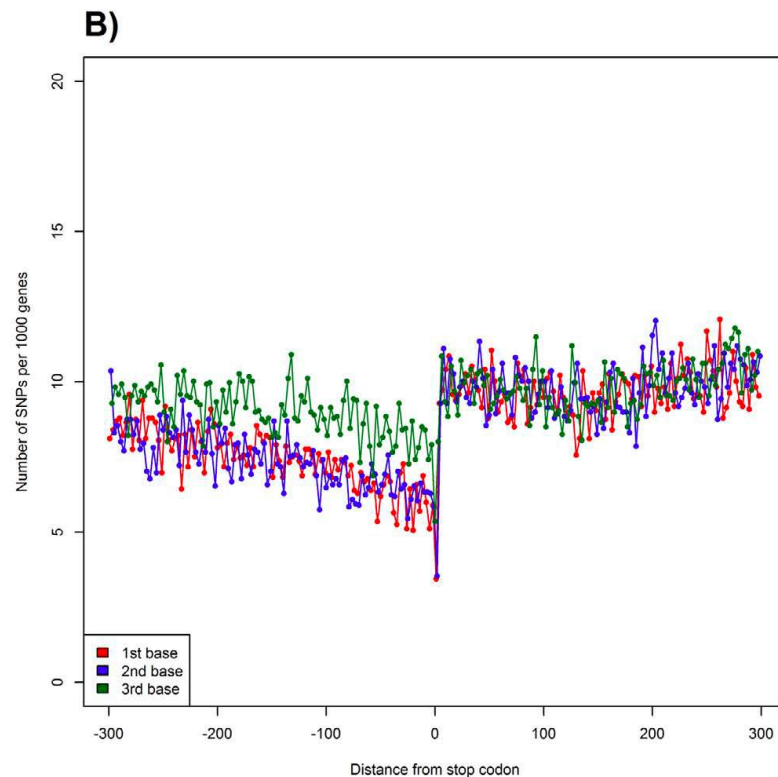
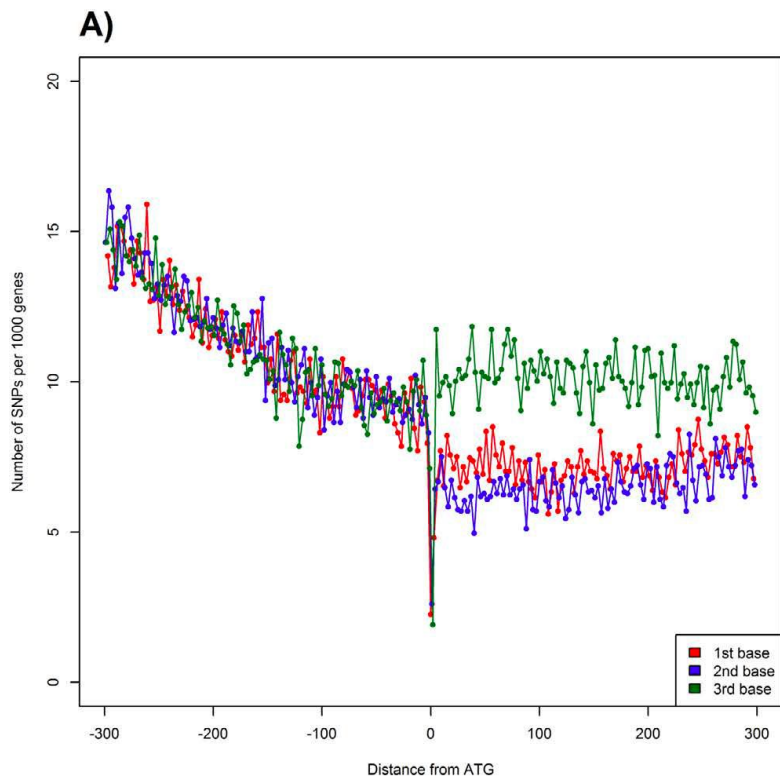
Danchurova Anastasia

Supervisor: Tatarinova Tatiana

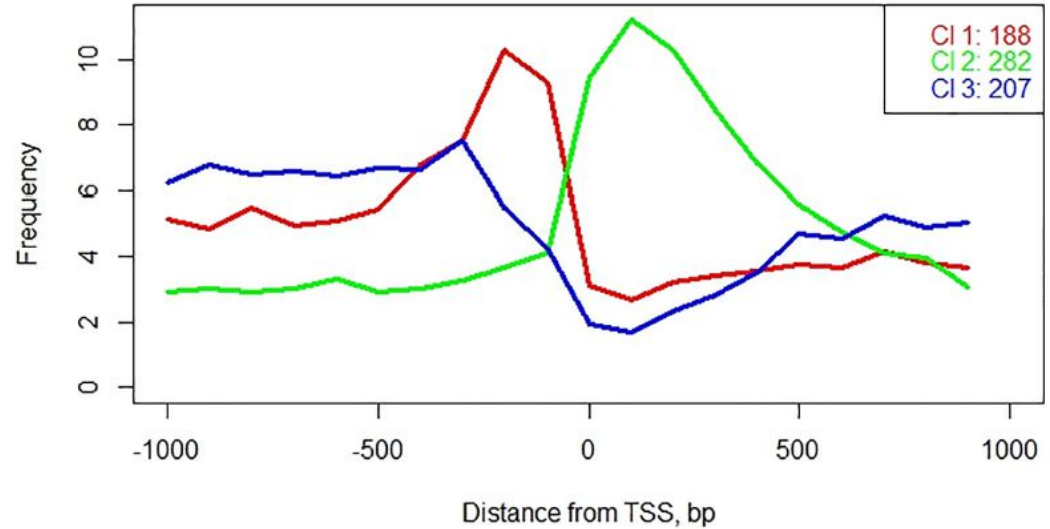
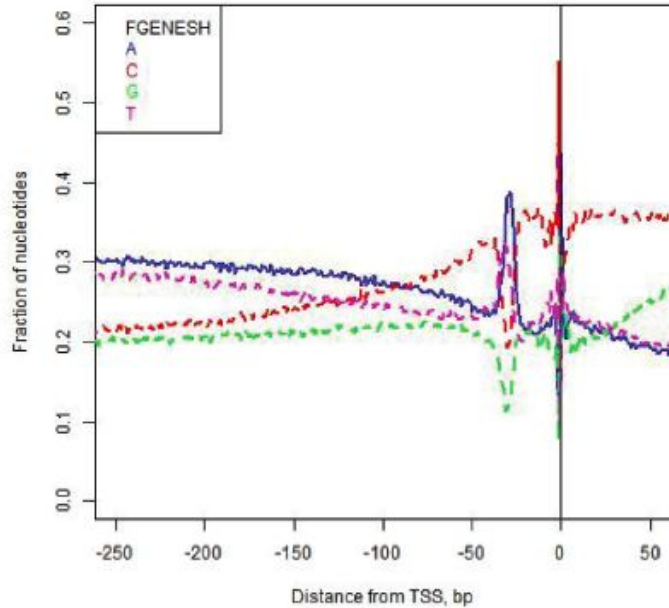
SNP неравномерно распределены по геному



SNP неравномерно распределены по геному



Распределение нуклеотидов и TFBS



Цель и задачи

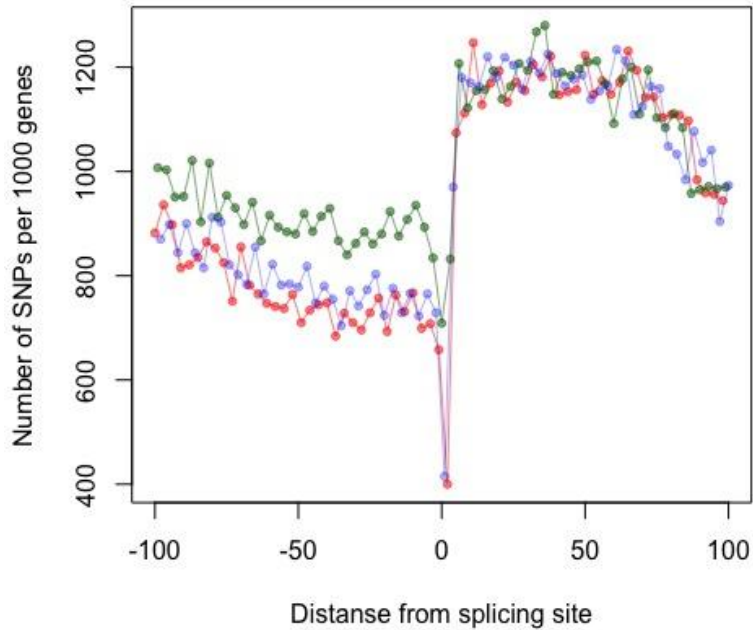
Найти корреляцию между свойствами генома и его функциональными регионами у разных организмов

- 1) Построить распределение таких свойств генома, как SNP и метилирование
- 2) Наложить это распределение на области промоторов/терминаторов
- 3) Проверить воспроизводимость этого распределения для нескольких организмов
- 4) Сделать выводы о возможном качестве аннотаций и использовании подобных метрик для аннотирования геномов

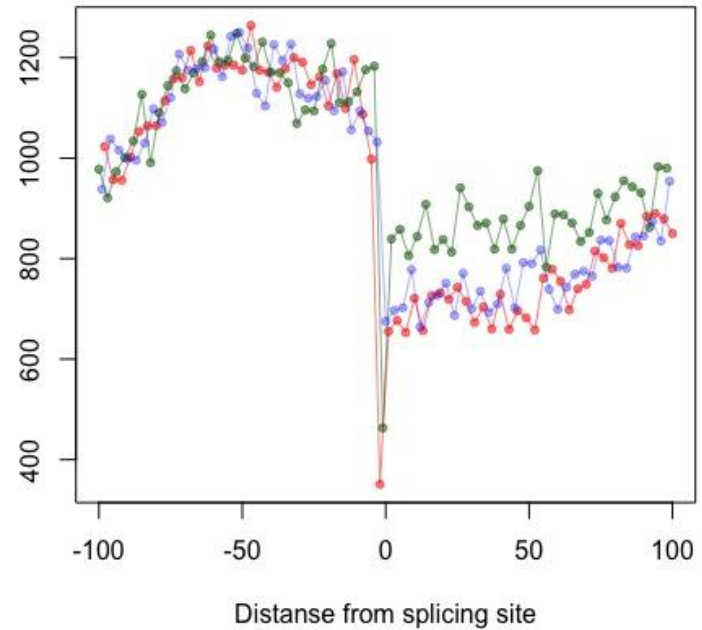
Результаты (*Arabidopsis thaliana*)

Распределение SNP

Number of SNPs near splicing site exon -> intron



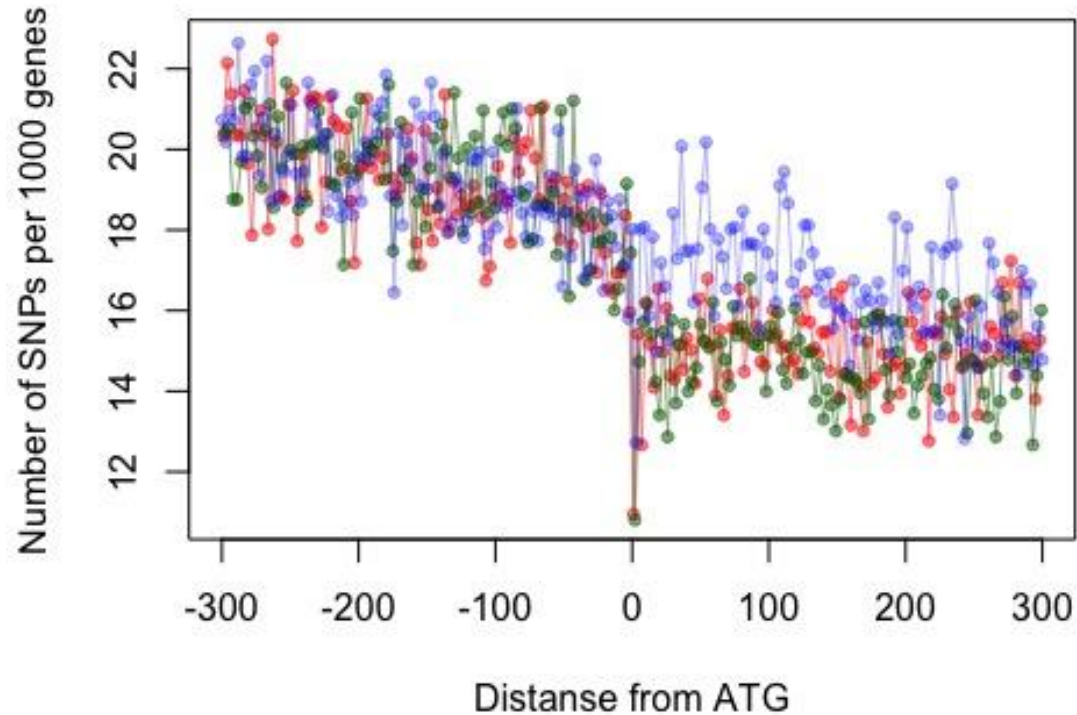
Number of SNPs near splicing site intron -> exon



Результаты (*Arabidopsis thaliana*)

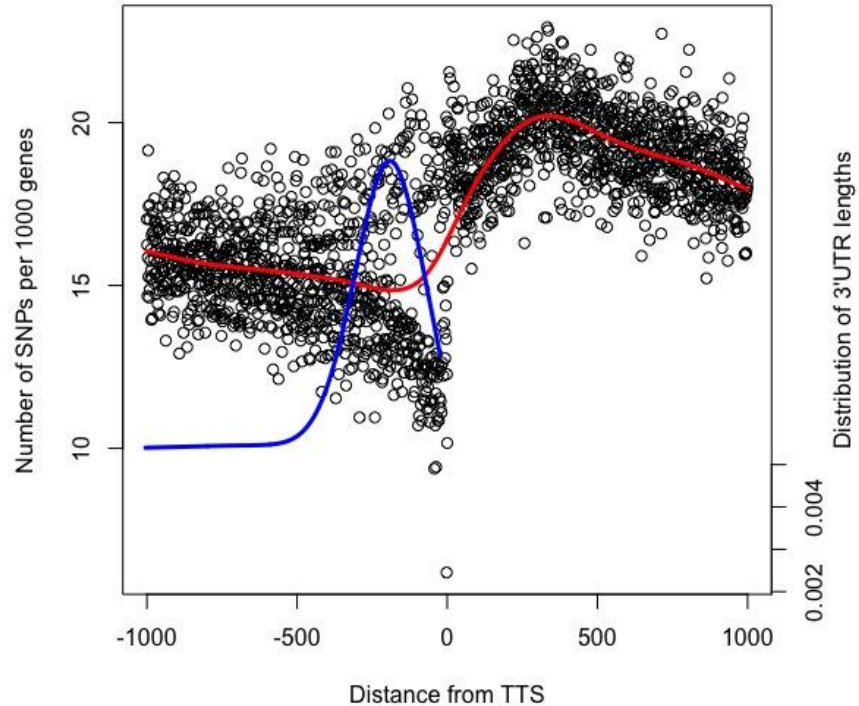
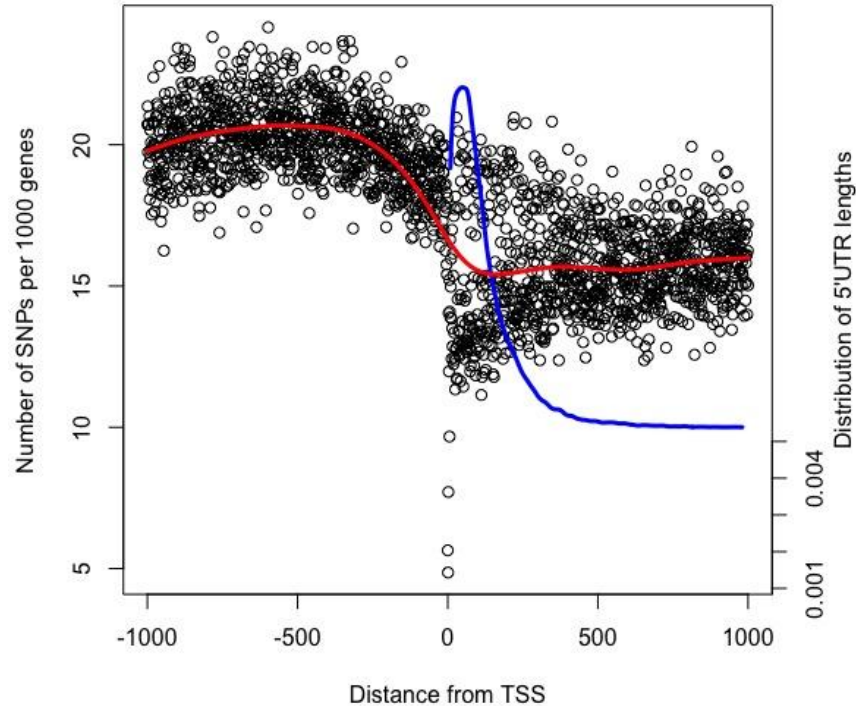
Распределение SNP

Number of SNPs near ATG



Результаты (*Arabidopsis thaliana*)

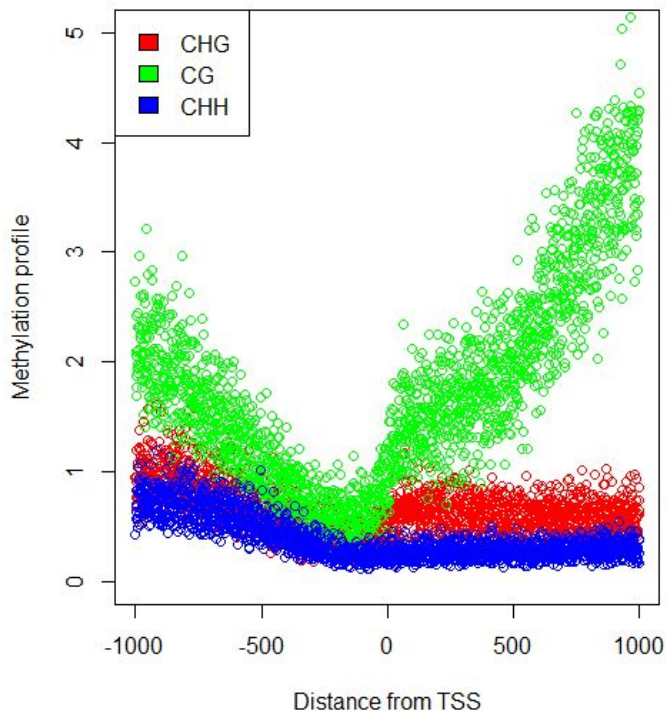
Распределение SNP



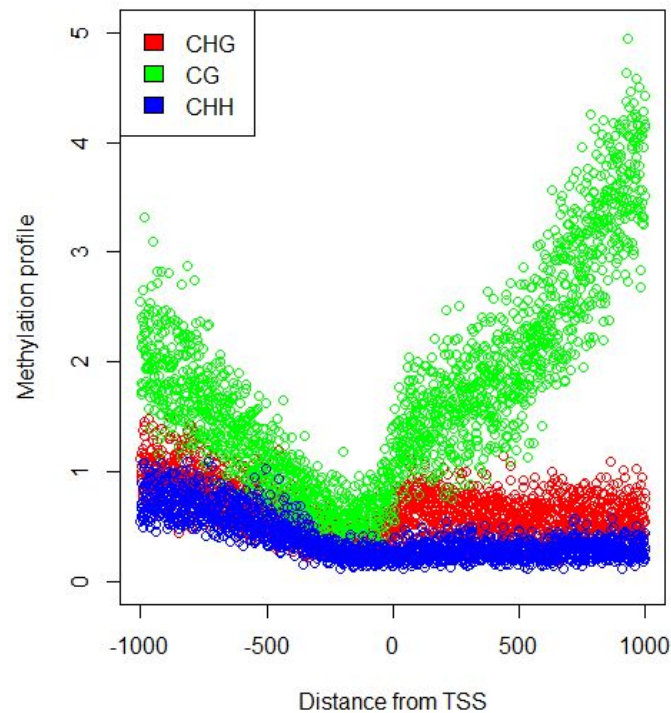
Результаты (*Arabidopsis thaliana*)

Метилирование

Methylation profile for vegetative cell promoters



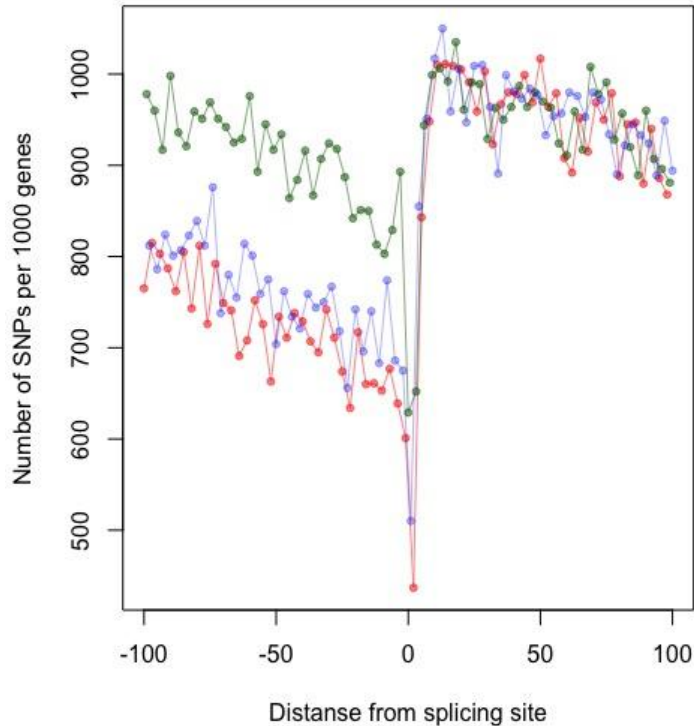
Methylation profile for sperm cell promoters



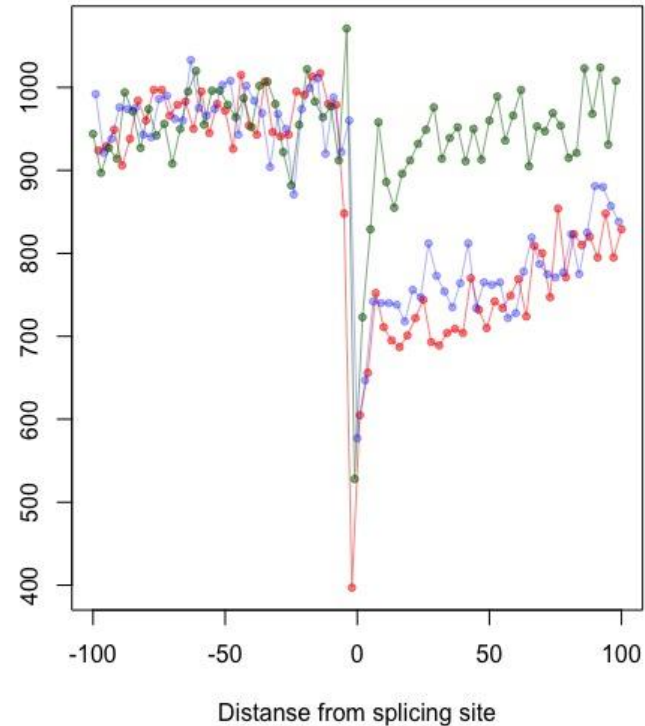
Результаты (*Medicago truncatula*)

Распределение SNP

Number of SNPs near splicing site exon -> intron

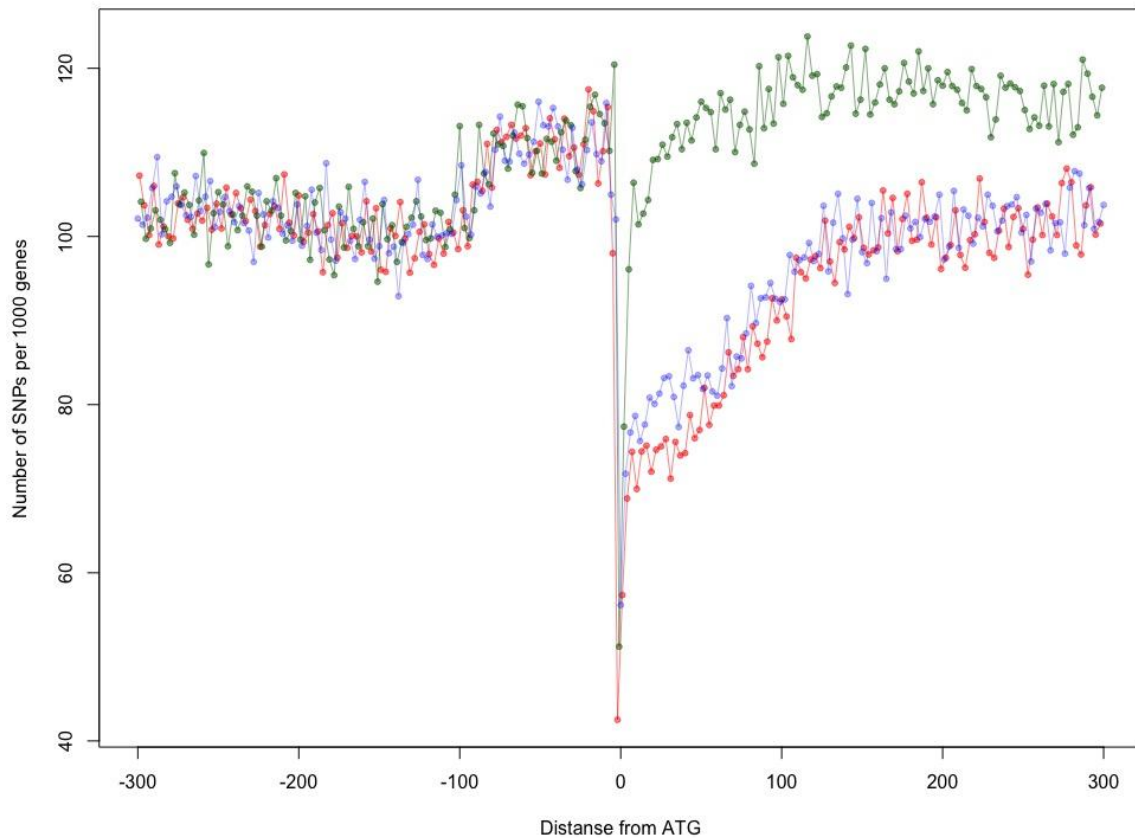


Number of SNPs near splicing site intron -> exon



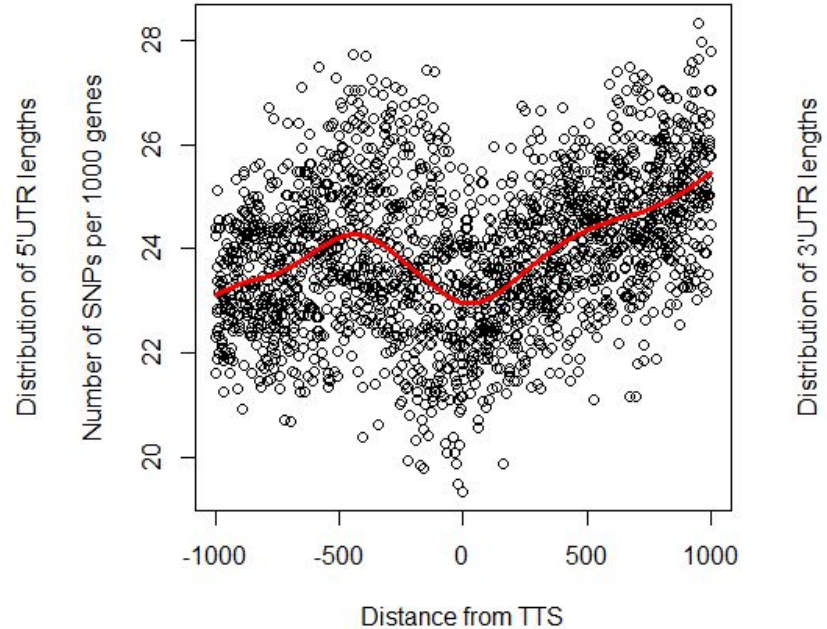
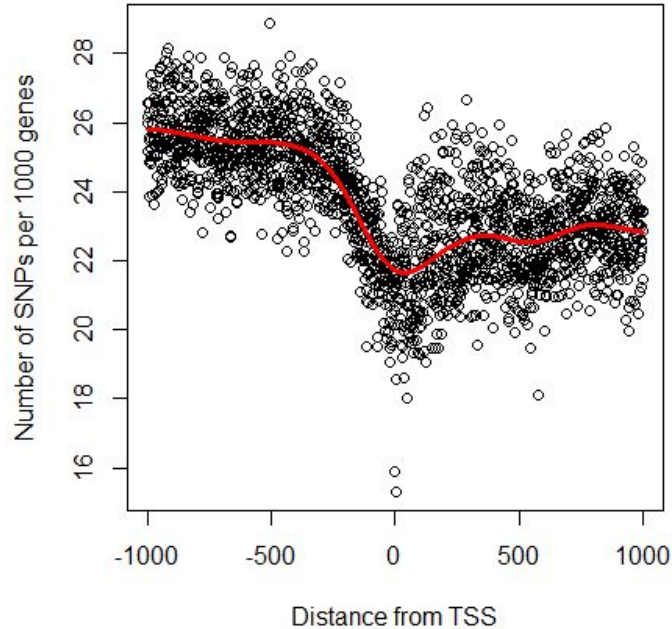
Результаты (*Medicago truncatula*)

Распределение SNP



Результаты (*Medicago truncatula*)

Распределение SNP



Выводы

- 1) Распределение SNP воспроизводится в области сайта начала/конца транскрипции, старт-кодона, стыке экзона/интрона
- 2) Метилирование имеет схожий паттерн в области сайта начала транскрипции у вегетативных и половых клеток у *A.thaliana*
- 3) Такие распределения можно использовать для оценки качества аннотации генома, а также для предсказания функциональных элементов в не аннотированных геномах.

Планы

- 1) Тщательная фильтрация используемых датасетов
- 2) Анализ распределения TFBS

Спасибо за внимание

September



May



https://github.com/danchurova/Genomes_project

Transcription Factor Binding Sites (TRANSFAC database)

