

Biolinux

Выполнила: Старостина Екатерина

Руководитель: Суворов Владимир, ЕМС

Задача

- Сравнение существующих сборок `bioinformatics` и отдельных инструментов для типовых задач препроцессинга генома для клинических и фармацевтических исследований

Препроцессинг

- Short read mapping
- Variant calling

Тестируемые инструменты

- Инструменты
 - Mosaik
 - Bwa
 - Bowtie
 - Bowtie2

Тестируемые сборки

	samtools	bcftools	bwa	bowtie	bowtie2	mosaik
BioLinux	0.1.7	-	0.5.5	0.12.7	-	-
BioKnoppix	-	-	-	-	-	-
BioSlax	-	-	-	-	-	-
OpenDiscovery2	-	-	-	-	-	-

- Рассматриваемые сборки, кроме BioLinux не предназначены для препроцессинга генома

Galaxy

Открытая веб-ориентированная платформа для биомедицинских исследований. Предоставляет удобный интерфейс для работы со многими из рассматриваемых инструментов. Может быть использована для задач сборки генома по референсу.

Результаты сравнения

1. E coli K12 MG1655

Объем, Гб	Технология	Количество ридов	#	Tool	Покрытие, %	Время работы			
						Построение индекса мин.сек	Выравнивание	Получение отсортированного bam	Получение vcf
1,9	Illumina	5729470	1	BWA	95.21	0.03	2:34	1.49	2.11
			2	Bowtie	95.27	0.04	1:00	4.00	12.52
			3	Bowtie2	99,32	0.04	2.16	3.03	20.39
			4	Mosaik	98.67	0.30	8.19	2.13	46.30
1,9 + 1,9	Illumina, paired-end	12197454	1	BWA	98.49	0.03	10. 20	8.06	31.23
			2	Bowtie	87.91	0.04	5.11	8.34	25.37
			3	Bowtie2	99,21	0.04	9.05	8.46	32.50
			4	Mosaik	97.65	1.09	8.15	2.17	59.43
4,1	IonTorrent	6668556	1	Bwa	97.74	0.02	34.00	3.54	26.21
			2	Mosaik	96.48	0.51	90.31	4.28	37.23

2. PhiX

Объем, Гб	Техноло гия	Количество ридов	#	Tool	Покрытие, %	Время работы			
						Построен ие индекса, мин.сек	Сборка	Получени е отсортир ованного bam	Получени е vcf
10,14	Illumina	18306599	1	BWA	92.77	0	9.26	8.30	52.23
			2	Bowtie	91.95	0.01	4.30	6.00	43.00
			3	Bowtie2	95.53	0	9.30	16.08	59.45
			4	Mosaik	94.95	2.02	52.04	8.42	56.54

Результаты

- Написаны скрипты для тестирования рассматриваемых инструментов, их можно найти по адресу https://github.com/mydatascience/bio_bundle
- Проведено тестирование на бактериях E.coli и PhiX
- На тестовых данных наиболее точным и быстрым оказался bowtie2 для Illumina, bwa для IonTorrent