



Пайплайн для анализа результатов таргетного секвенирования

Руководитель:

Мозгов Сергей Сергеевич, к.т.н.,
инженер ИТ-инфраструктуры, Parseq
Lab

Студент:

Слепченков А.В.

Задачи

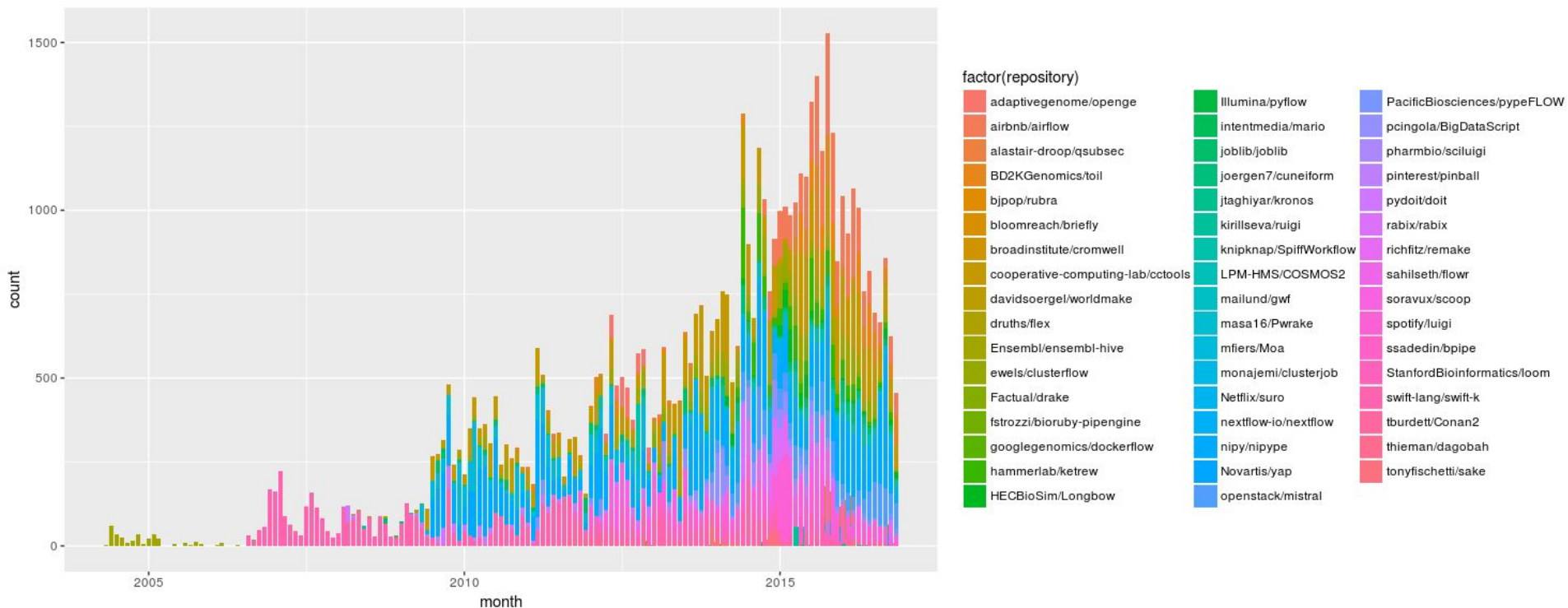
- Обзор наиболее перспективных средств описания и запуска пайплайнов
- Реализация пайплайна для анализа результатов таргетного секвенирования
- Сравнение созданного пайплайна со стандартным пайплайном Ion Torrent

Средства описания пайплайнов



Средства описания пайплайнов

- ~ 100 различных фреймворков и библиотек
- <https://github.com/pditommaso/awesome-pipeline>

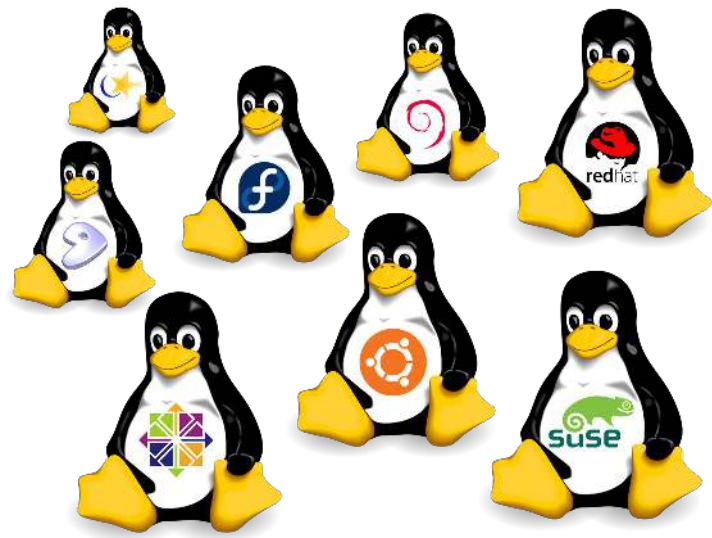


Common Workflow Language

Переносимость



amazon
web services



Google Cloud Platform

Common Workflow Language

Простота работы с docker
контейнерами



Common Workflow Language

Разрабатывается при поддержке
крупных организаций



Таргетное секвенирование

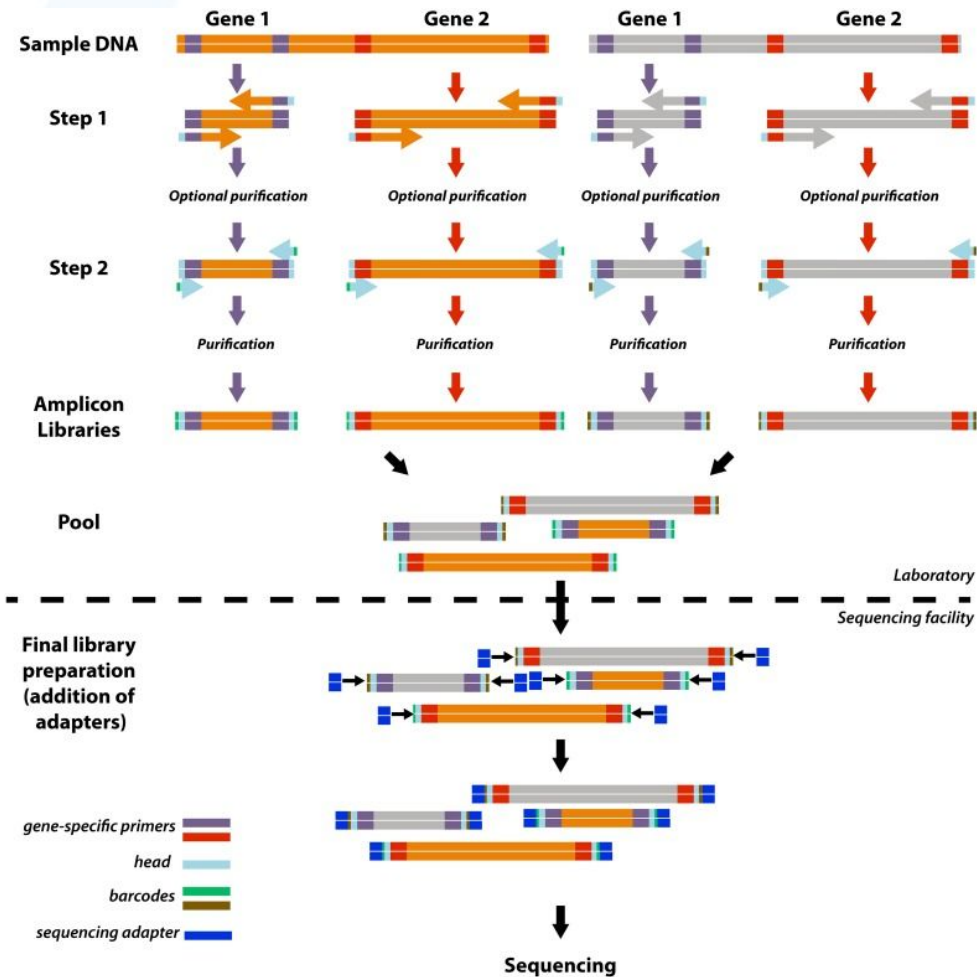
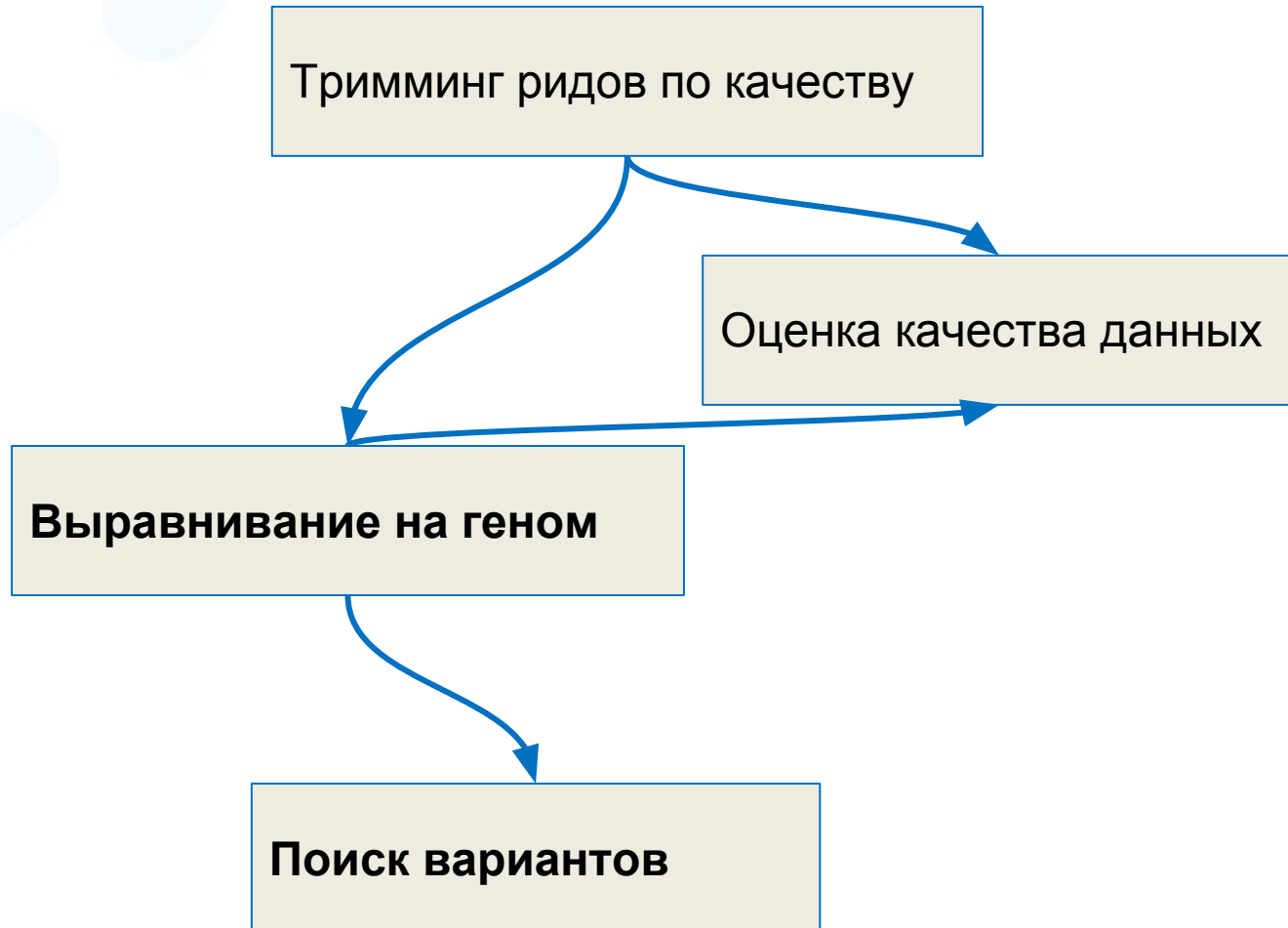


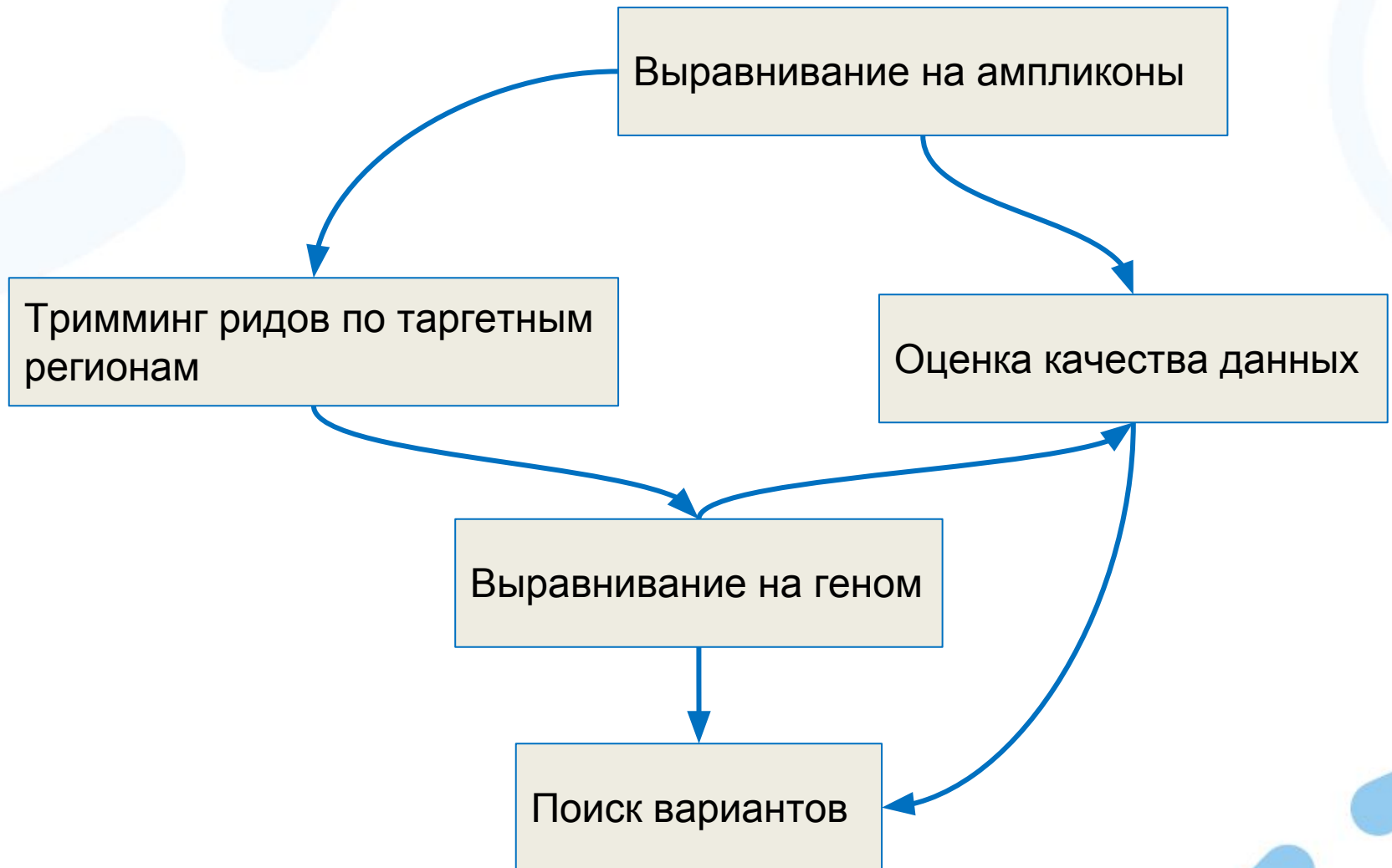
Схема пайплайна



Проблемы с перекрывающимися регионами

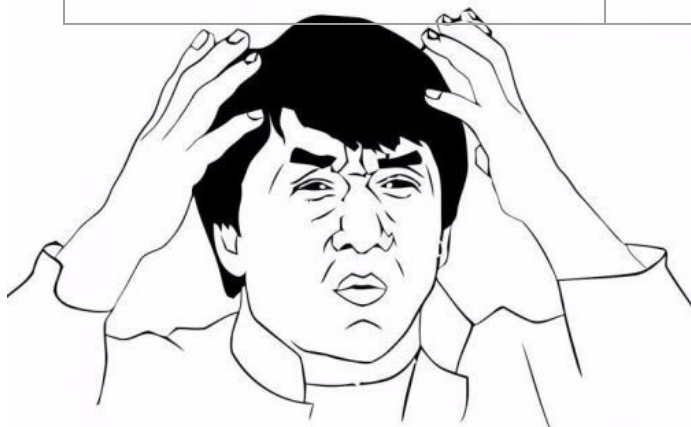


Схема пайплайна



Последствия тримминга

Тримминг	Ложно отрицательные варианты	Ложно положительные варианты
+	40	6
-	4	2



Теги Ion Torrent'a



Теги Ion Torrent'a

- На чип подаются нуклеотиды в определенном порядке
- Включение нуклеотида вызывает изменение pH
- Последовательность сигналов об изменении pH включается в данные в виде тега -
ZM:B:s,252,0,262,-4,2,224,-6,240,-2,194,466,214,258,256,28,472,248,-6,
- Torrent Suite Variant Caller использует эти данные для оценки значимости варианта

Последствия тримминга

Тримминг	Ложно отрицательные варианты	Ложно положительные варианты
+	40	6
+ с изменением ZM тега	4	2
-	4	2

Анализ качества результатов

- На основании Base Quality и покрытия каждому прочитанному основанию выставляется определенное качество
- Участки таргетных регионов или известные варианты, попавшие в плохо отсеквенированные регионы, записываются в метаданные VCF файла

Сравнение со стандартным пайплайном

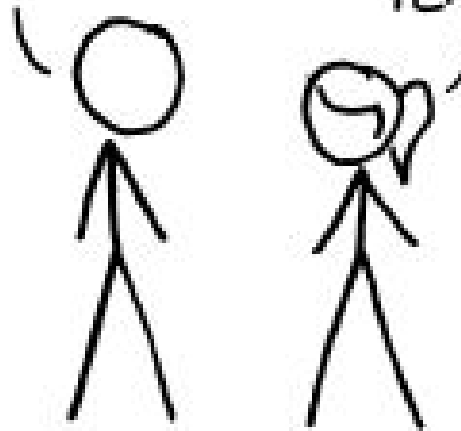
Parameter	Torrent Suite	Custom
Total bases	1,259,274	1,254,424
Common variants	1080	1082
Sensitivity, %	98	98
Specificity, %	99,26	99,54

HOW STANDARDS PROLIFERATE:

(SEE: A/C CHARGERS, CHARACTER ENCODINGS, INSTANT MESSAGING, ETC.)

SITUATION:
THERE ARE
14 COMPETING
STANDARDS.

14?! RIDICULOUS!
WE NEED TO DEVELOP
ONE UNIVERSAL STANDARD
THAT COVERS EVERYONE'S
USE CASES.



SOON:

SITUATION:
THERE ARE
15 COMPETING
STANDARDS.