

# Определение копийности контигов с использованием информации о сборке

Мелешко Дмитрий  
СПбАУ РАН

Научный руководитель: Son Pham, PhD, UCSD  
Рецензент: Антипов Д.Ю., СПбГУ

03 июня 2015

# Изменение копинойсти (ИК)



Хромосома



Референсный геном



Дупликация



Трипликация



Сложное ИК



Инверсия



Делеция

**В список заболеваний входят как редкие:**

Синдром Вильямса

Аутизм

Шизофрения

**Так и довольно распространены:**

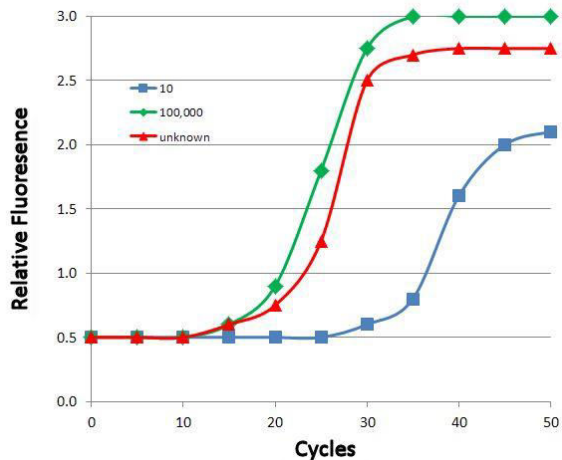
Ожирение

Болезнь Крона

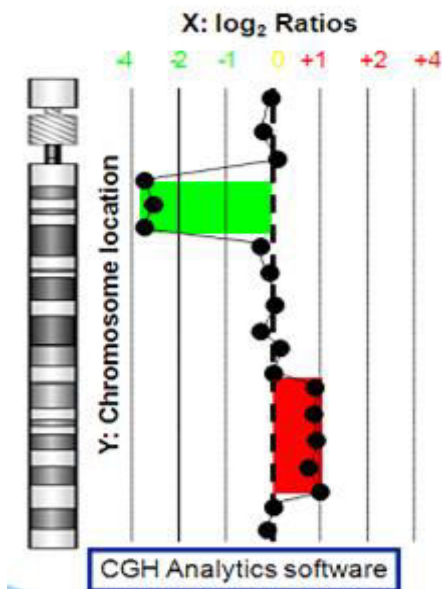
Псориаз

# Методы определения ИК

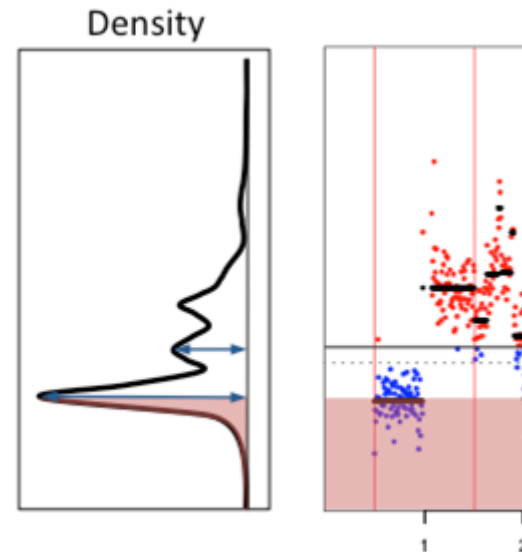
## Биологические



## Чипы

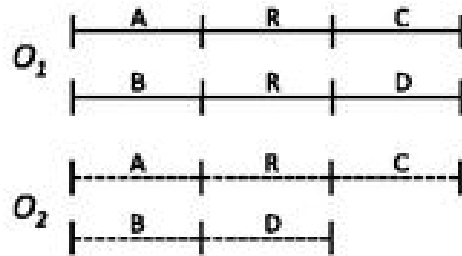
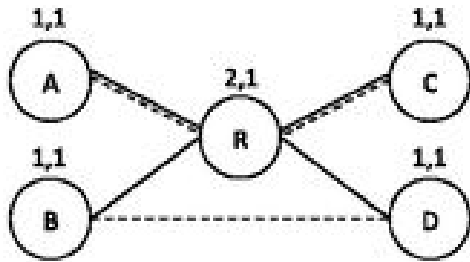


## Секвенирование нового поколения

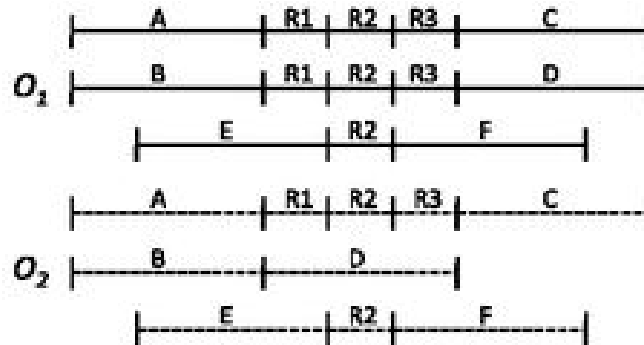
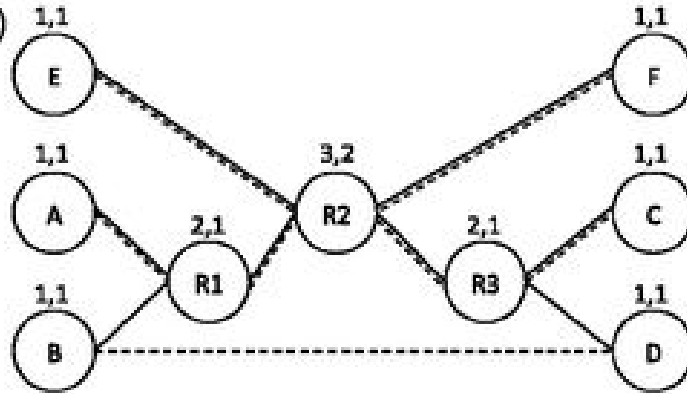


# От ИК к копиям КОНТИГОВ

(a)



(b)



2012 - *De novo*  
detection of copy  
number variation  
by co-assembly,  
Jurgen F. Nijkamp.

# Цель

Разработать метод, который бы использовал всю имеющуюся информацию на этапе сборки и таким образом улучшить действие алгоритма Magnolia

# Задачи

- Выбор ассемблера, с которым предстоит работать

# Задачи

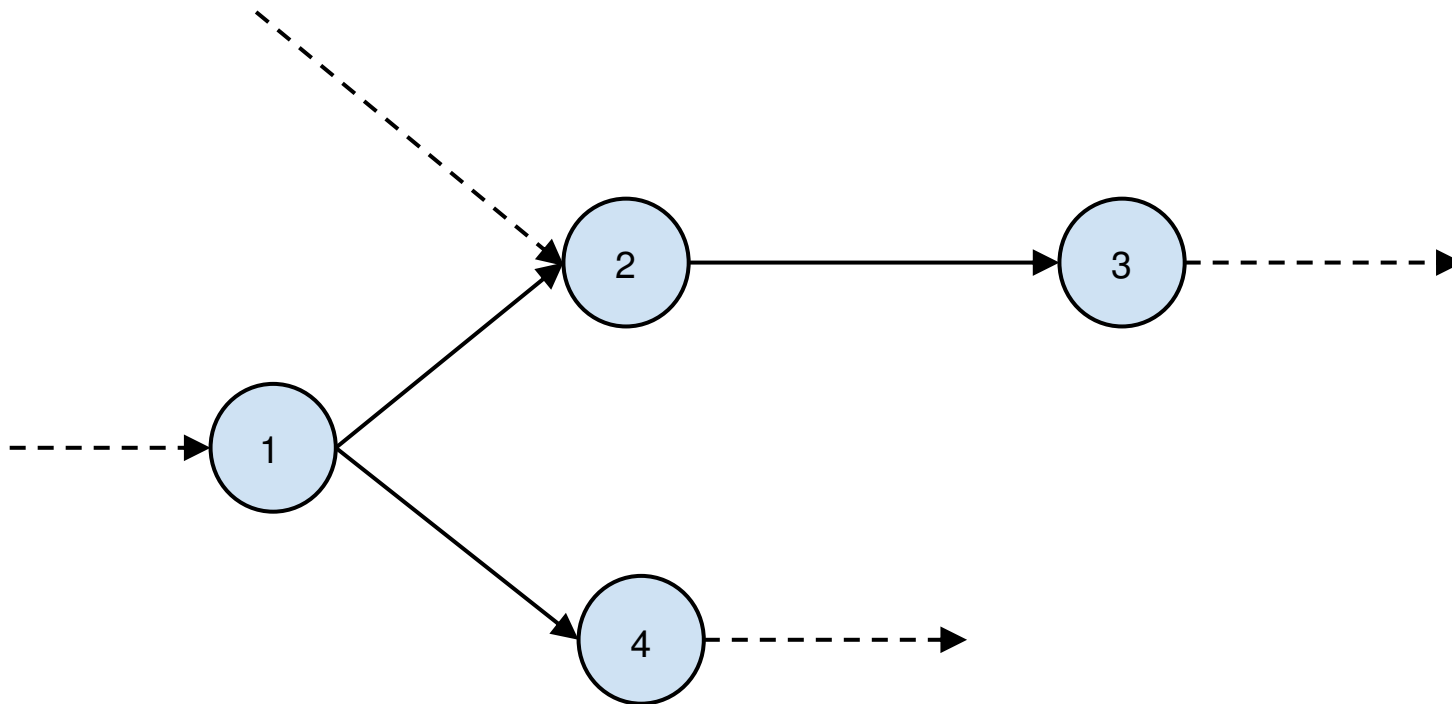
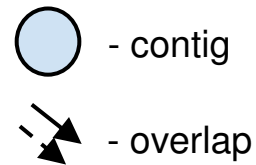
- Выбор ассемблера, с которым предстоит работать
- Сведение задачи о нахождении копийности о нахождении количества путей

# Задачи

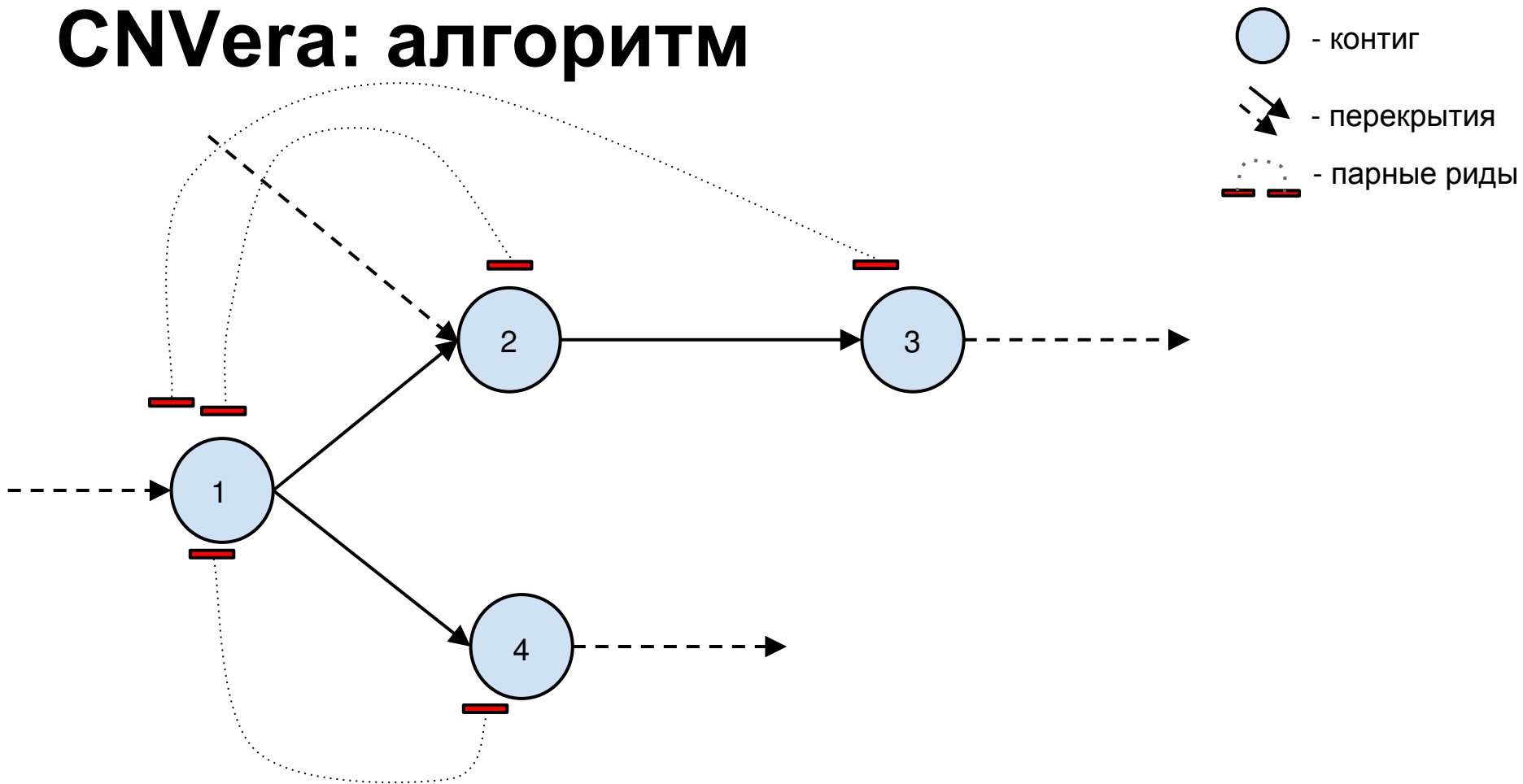
- Выбор ассемблера, с которым предстоит работать
- Сведение задачи о нахождении копийности о нахождении количества путей
- Сравнение результатов



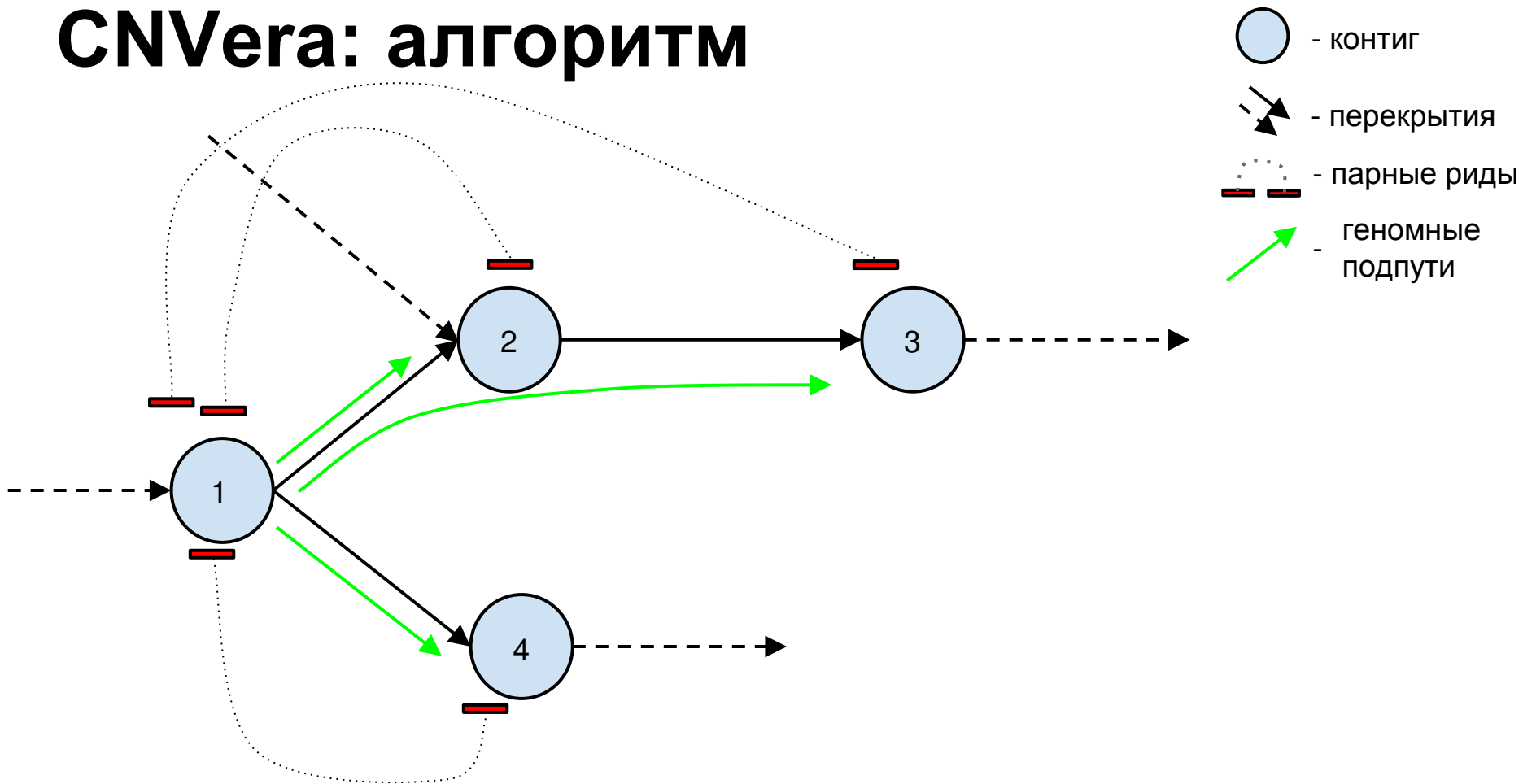
# CNVera: алгоритм



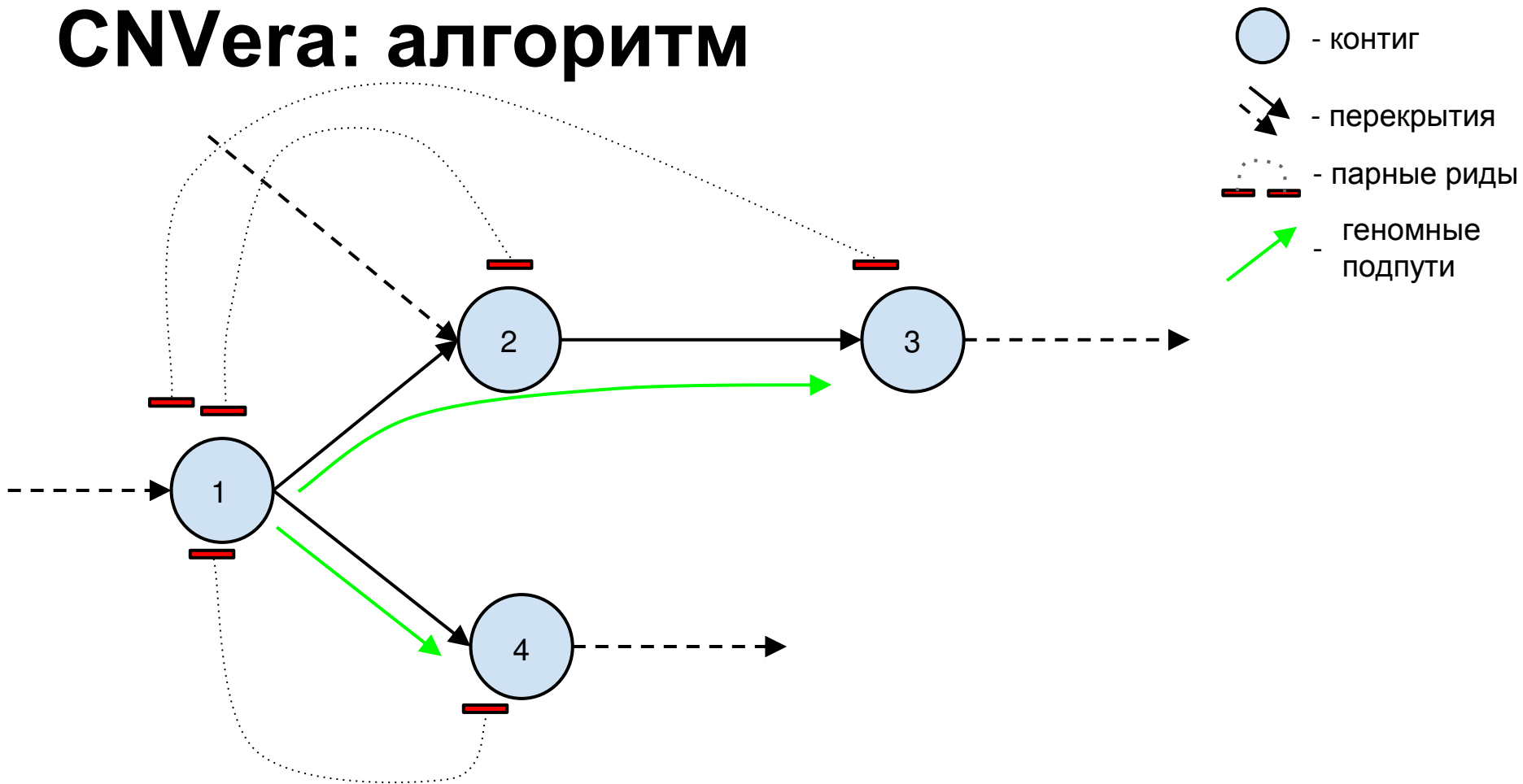
# CNVera: алгоритм



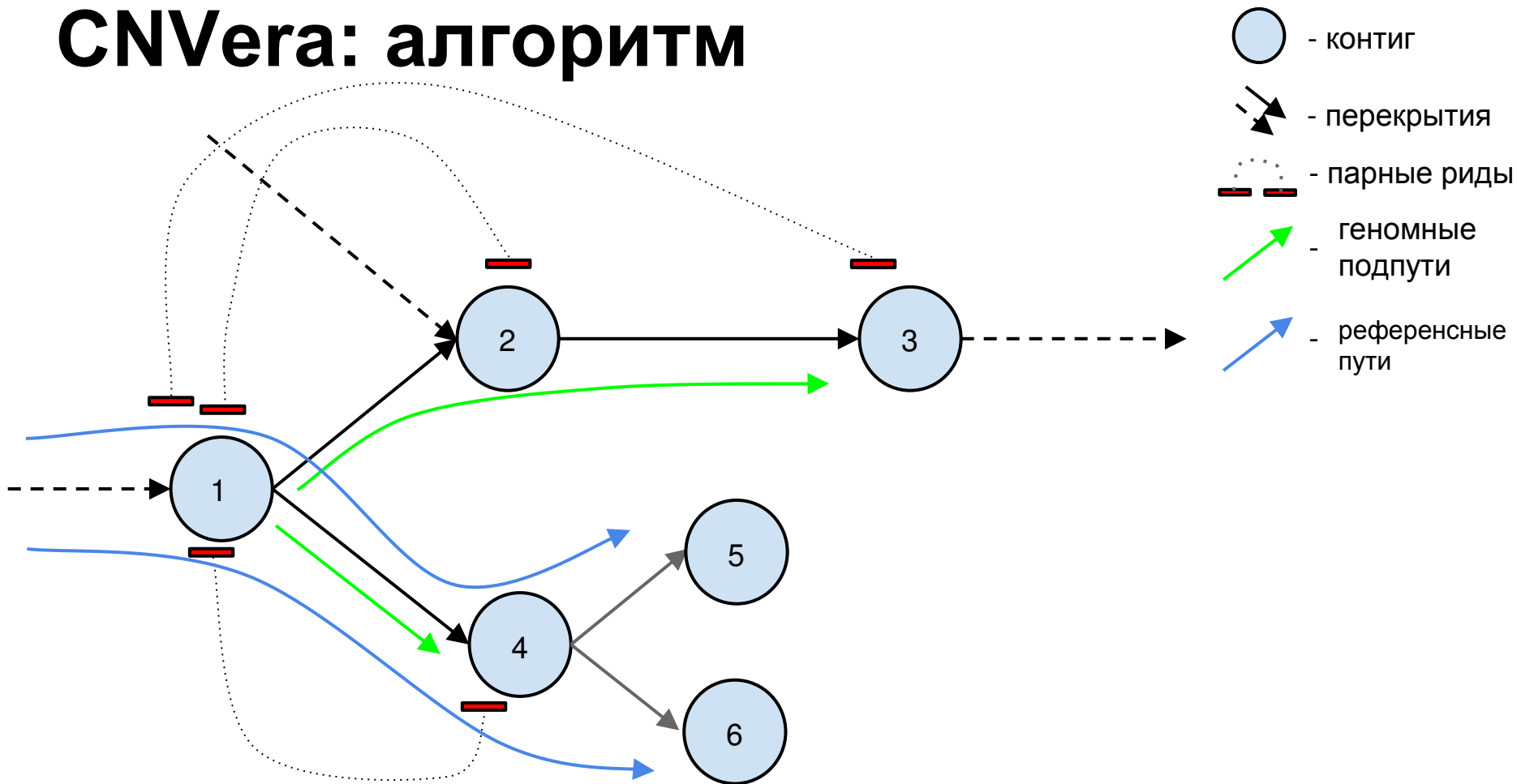
# CNVera: алгоритм



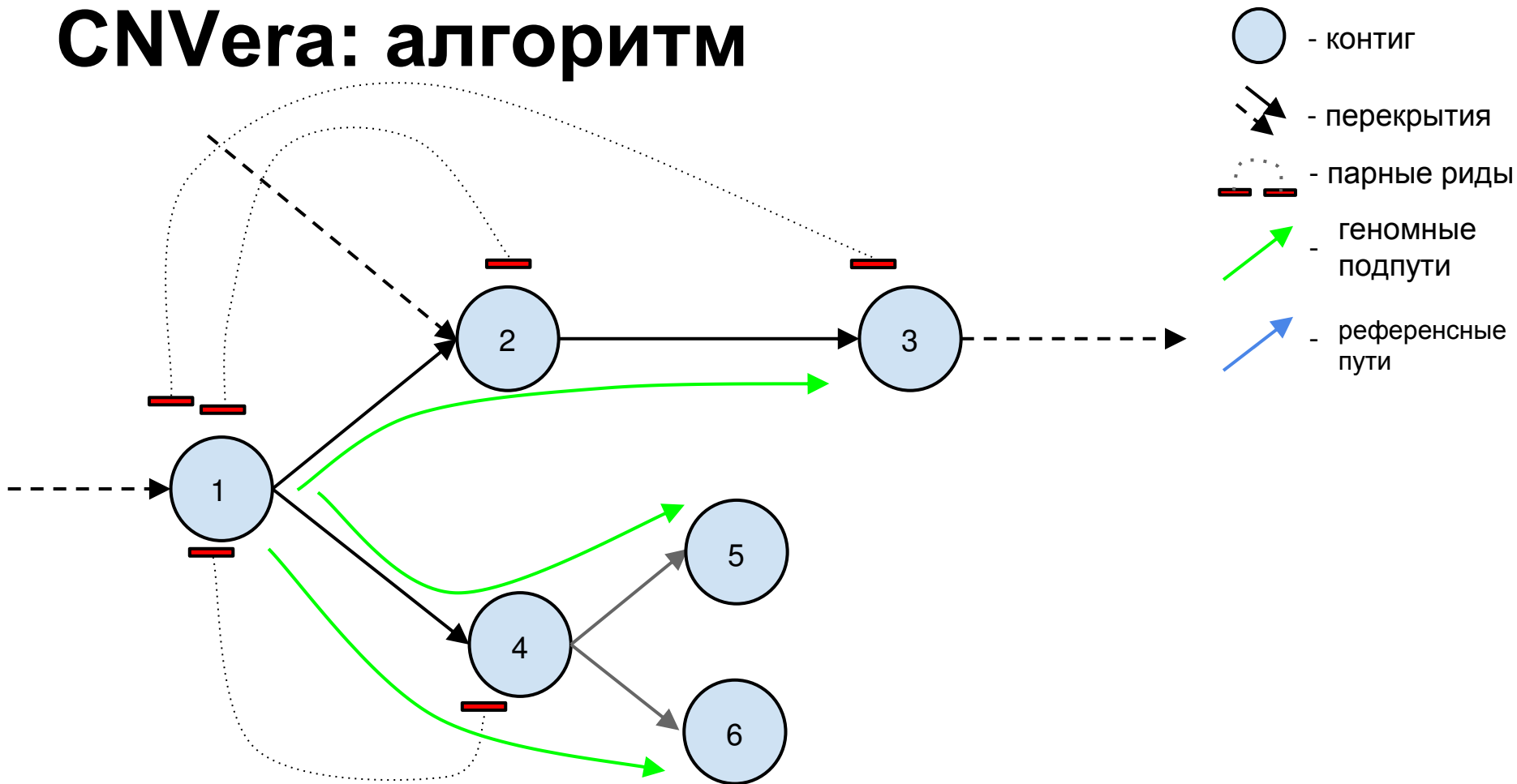
# CNVera: алгоритм



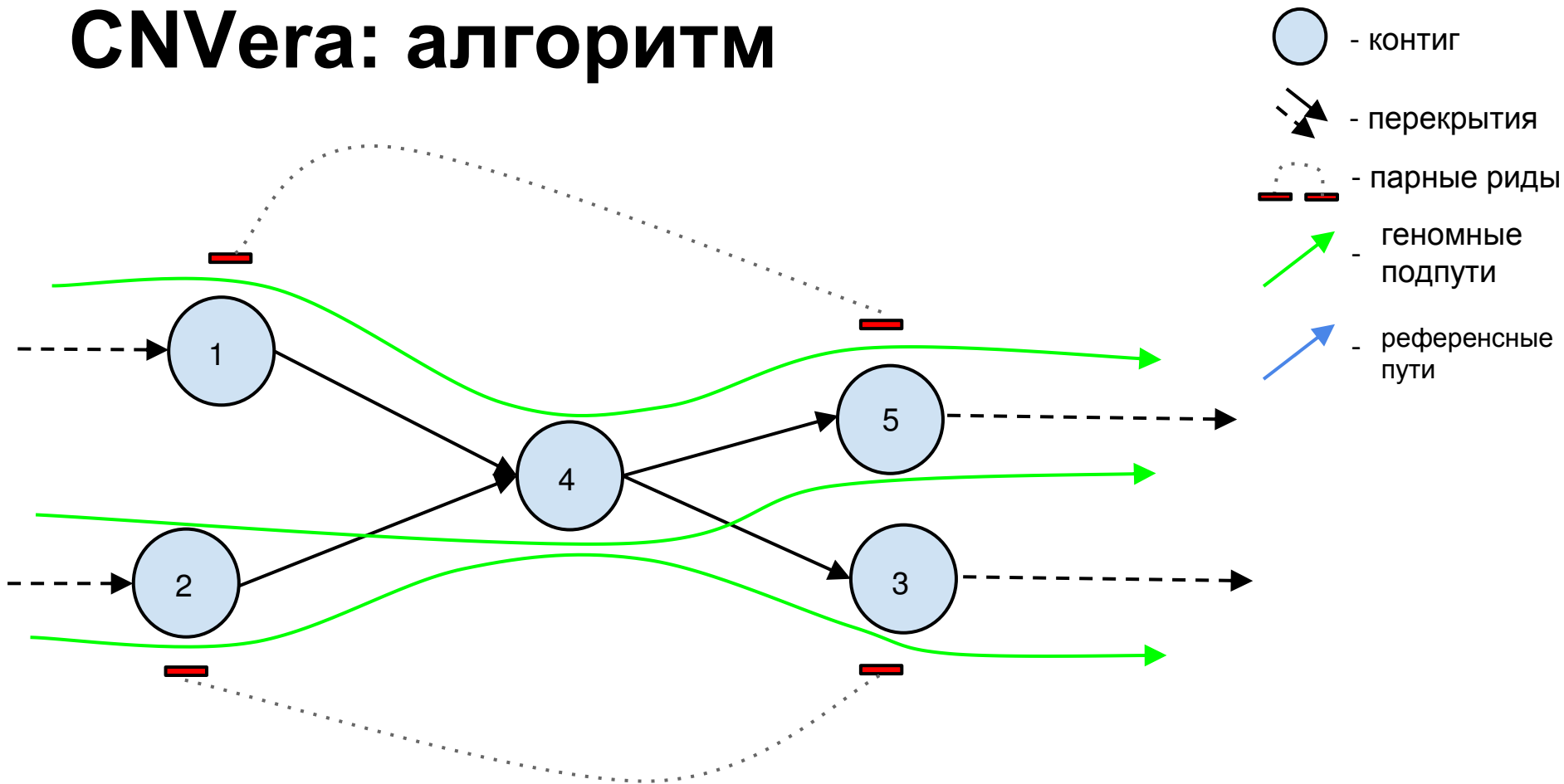
# CNVera: алгоритм



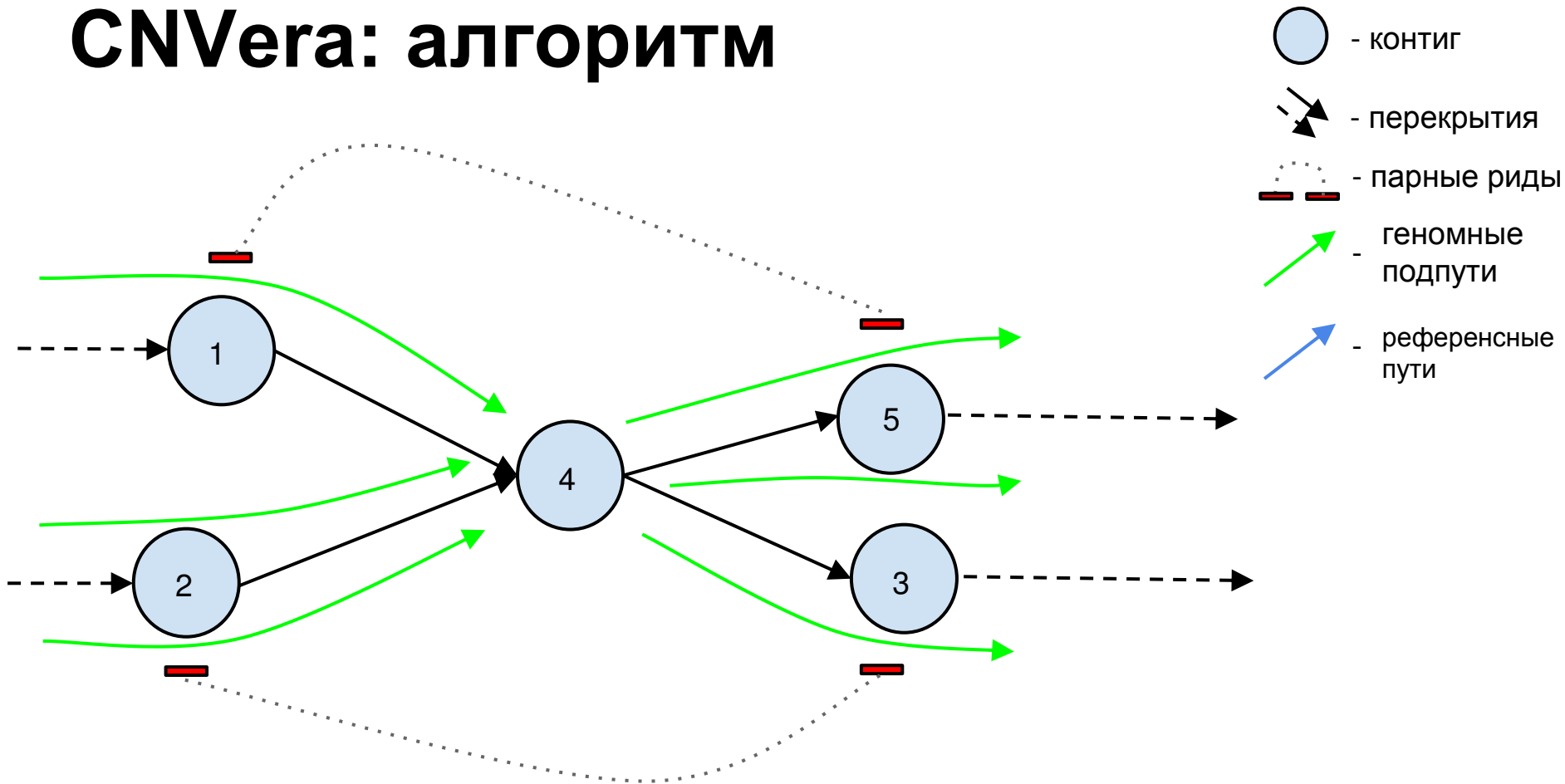
# CNVera: алгоритм



# CNVera: алгоритм



# CNVera: алгоритм





# Описание хода работы

Ассемблирование: SGA (с агрессивными параметрами)

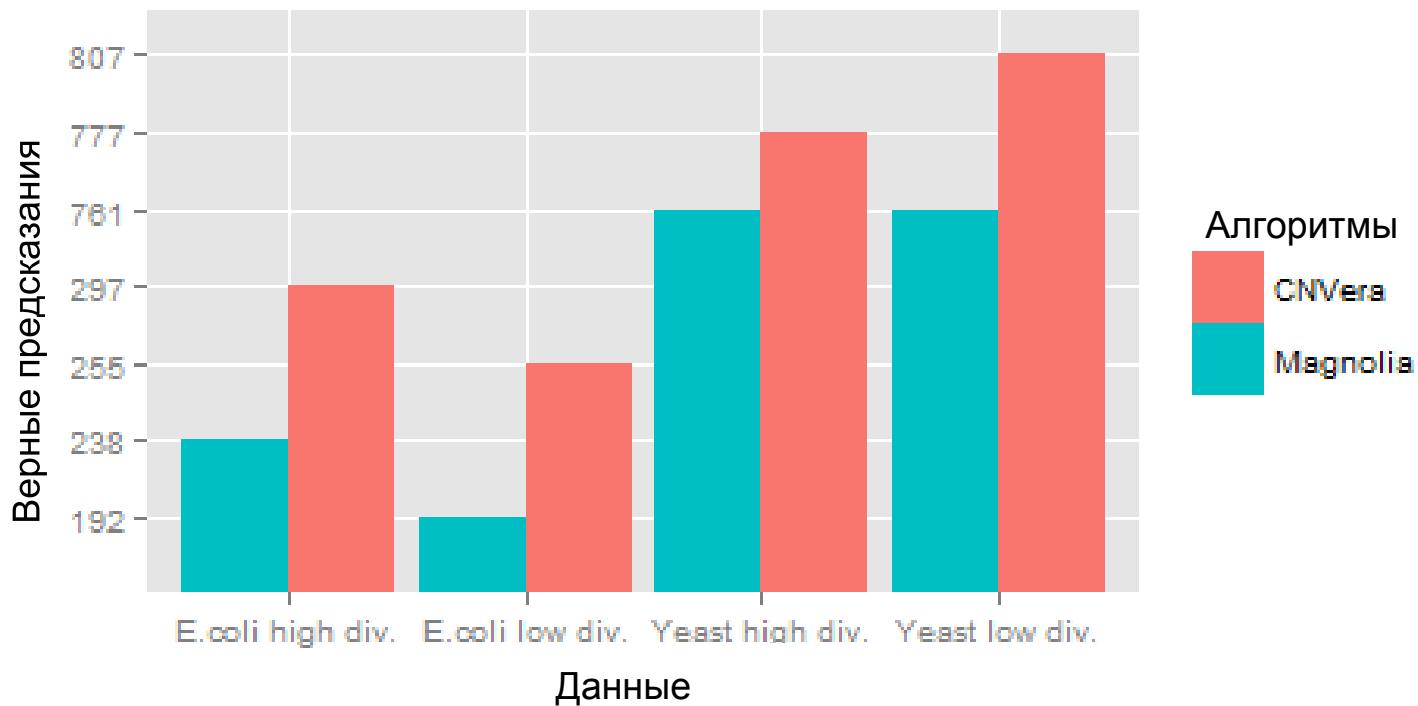
Получение парной информации: Bowtie + DistanceEst

Конкурент: Magnolia

Проверка результатов: BLAST + скрипты на python

Данные: два бактериальных набора данных, два набора данных дрожжей (простейшие эукариоты)

# Сравнение: верные предсказания



# Сравнение: сумма ошибок

Данные: E.Coli, референс с высокой степенью расхождения

Алгоритм	Сумма ошибок	Короткие	Средние	Длинные
CNVera	426	300	113	13
Magnolia	812	751	122	14

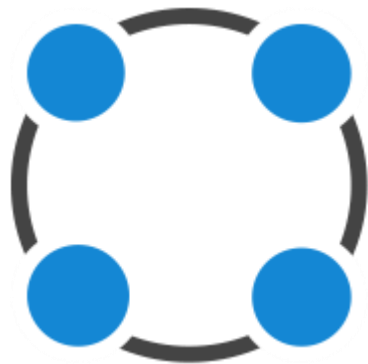
# Выводы

Разработан новый подход поиска ИК, основанный на анализе графа ассемблирования.

Задача о поиске ИК сведена к задаче поиска количества путей, проходящих через вершину графа ассемблирования

Показанные результаты говорят о том, что в настоящее время этот подход мог бы конкурировать с подходами, ориентирующимися на глубину покрытия

По результатам работы принят постер на HiTSeq-2015



**Спасибо за внимание!**