

# Инструментальная среда на базе технологий облачных вычислений для геномного секвенирования CloudNA

НИЦ «Курчатовский институт»

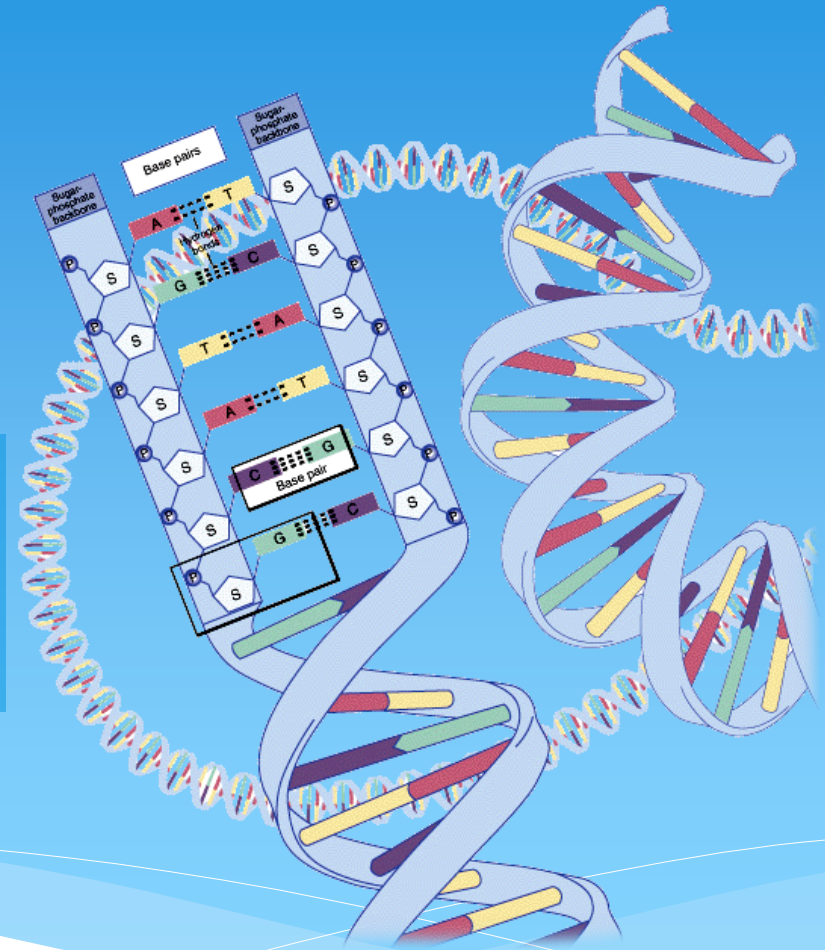
А. И. Кураченко

В. А. Ильин

# Определения

**Секвенирование** - определение первичной аминокислотной или нуклеотидной последовательности белков и нуклеиновых кислот

**МикроРНК** – некодирующие молекулы РНК, длина 21-25 нуклеотидов отвечающий за экспрессию генов



# Секвенирование нового поколения



СНП – единовременное  
«прочтение» нескольких  
участков генома

## Illumina HiSeq1500

- Используются в лаборатории геномики НБИК-Цentra НИЦ «Курчатовский институт»
- 150 ГБ/день данных

# Актуальные проблемы анализа данных секвенирования

- \* Увеличение скорости генерации данных существенно повысило требования к производительности обработки;
- \* Многоэтапность обработки – цепочка атомарных вычислительных задач;
- \* Разнородность требований к ресурсам атомарных задач;

# Актуальные проблемы анализа данных секвенирования

## Распределенные вычисления

- \* Увеличение скорости генерации данных существенно повысило требования к производительности обработки;

## Использование композитных приложений

- \* Многоэтапность обработки – цепочка атомарных вычислительных задач;

## Балансировка нагрузки

- \* Разнородность требований к ресурсам атомарных задач;


# Существующие облачные системы для анализа секвенирования

## **Crossbow**

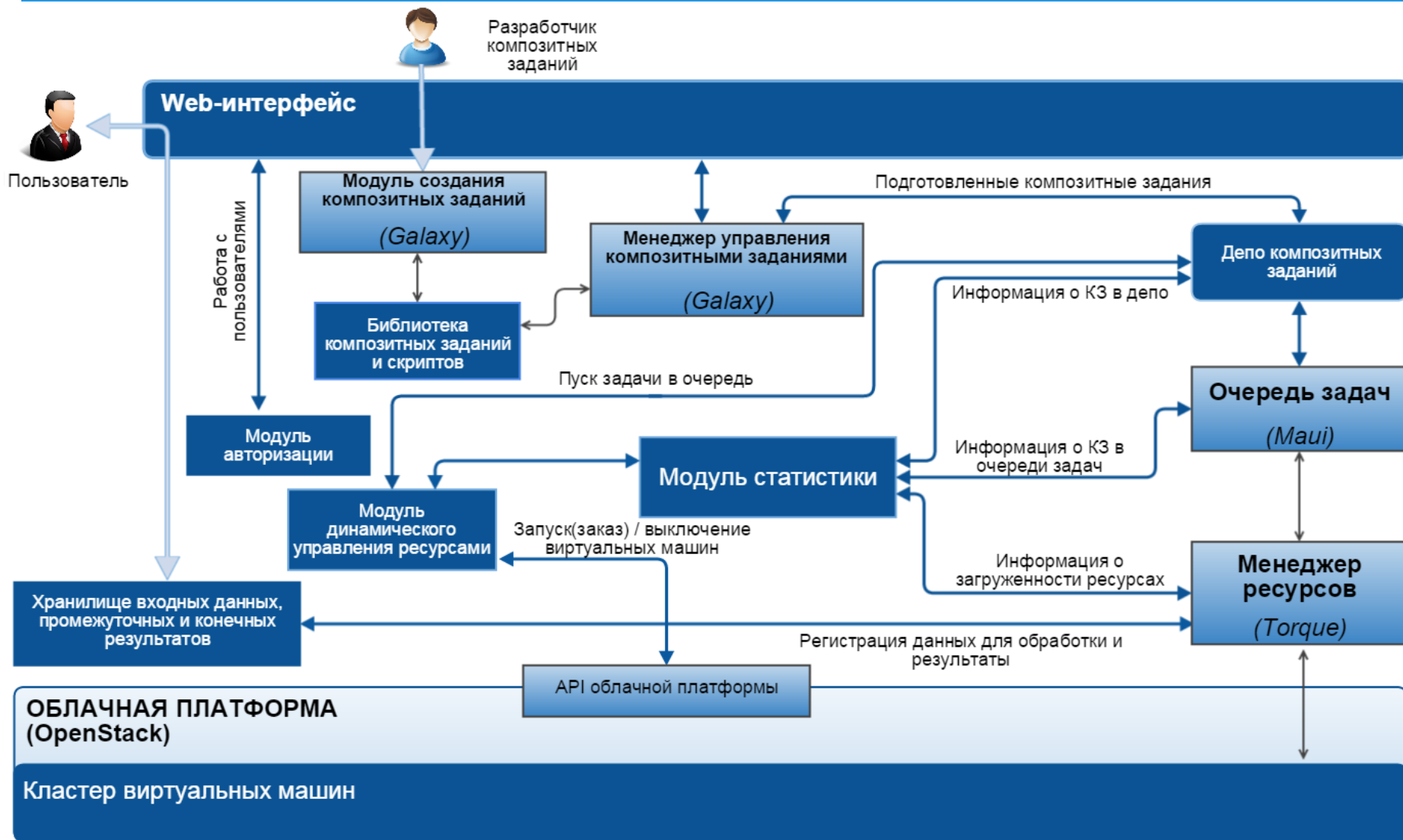
Genotyping from short reads using cloud computing

STATION 

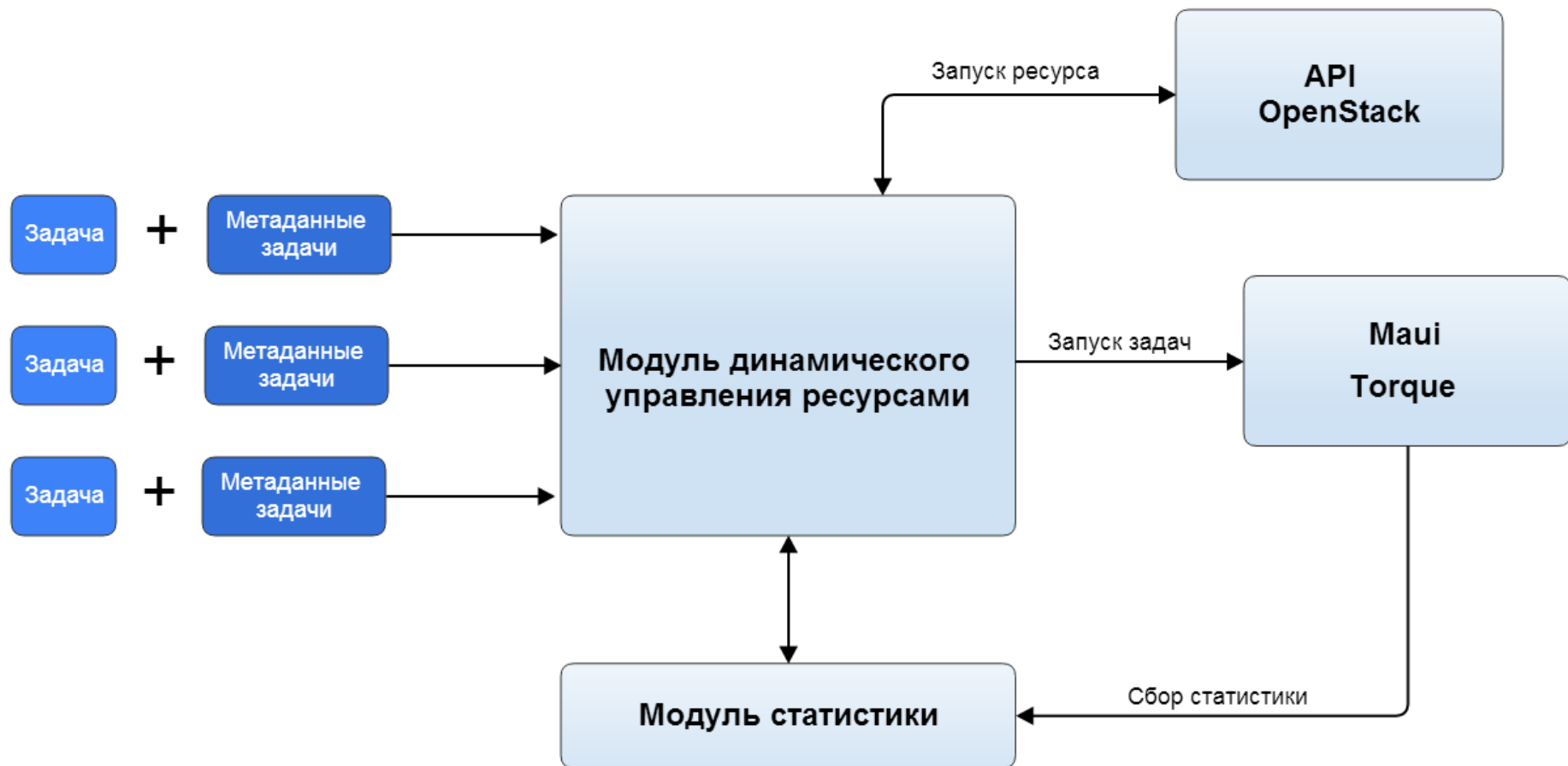


 CloudBLAST

# Архитектура CloudNA



# Архитектура модуля динамической балансировки ресурсов





# Тестовая прикладная задача

Произведен анализ данных секвенирования молекул микроРНК из трех типов тканей хризантемы садовой.

Результаты опубликованы в журнале “Информационное общество” «Использование облачной инфраструктуры для анализа данных секвенирования микроРНК» 2013 г. вып. 1-2, с. 26-38.

<http://infosoc.iis.ru/>

# Заключение

Архитектура позволяет:

- Повысить скорость обработки данных;
- Экономить вычислительные ресурсы за счет балансировки;
- Создавать и использовать композитные задания