# A Computational Platform for Gene Expression Analysis

Diogo Teixeira[1]

Supervisors: Rui Camacho[2], Nuno Fonseca[3]

[1]Check affiliation

[2]Check affiliation

[3]Check affiliation

July 2014

# Outline


1. Introduction
   - Domain Problem
   - Motivation and Objectives
2. Developed Solution
   - Overview
   - RNA-Seq Analysis Pipeline
   - RBP Analysis Pipeline (PBS Finder)
   - Integration
3. Case Studies
   - RNA-Seq Analysis Pipeline
   - RBP Analysis Pipeline (PBS Finder)
4. Conclusions
   - Objective Fulfilment
   - Future Work

# Domain Problem I

## Introduction

- Molecular biology is a young field of study, with a lot of unknowns and partial knowledge.
- Studying gene expression is crucial to understand the mechanisms that control living organisms.
- We focused on two different areas:
  - differential expression analysis;
  - RNA-binding protein (RBP) discovery and analysis.

# Domain Problem II

## Introduction

Three distinct problems:

- Read alignment against a reference genome and differential expression analysis on the aligned data.
- RBP discovery, analysis and information enrichment.
- Further result analysis using data mining techniques.

**Tools are complex**

Tools for biological data analysis often require a very technical set of skills.

**Tools are complex**
Tools for biological data analysis often require a very technical set of skills.

**Tasks are repetitive**
Analysing high quantities of data can be repetitive, especially if executed manually.

# Motivation and Objectives

## Introduction

### Tools are complex

Tools for biological data analysis often require a very technical set of skills.

### Tasks are repetitive

Analysing high quantities of data can be repetitive, especially if executed manually.

### Information is scattered

Information is easy to acquire, but is often scattered through multiple platforms, services and institutions.

# Motivation and Objectives

## Introduction

### Tools are complex

Tools for biological data analysis often require a very technical set of skills.

### Tasks are repetitive

Analysing high quantities of data can be repetitive, especially if executed manually.

### Information is scattered

Information is easy to acquire, but is often scattered through multiple platforms, services and institutions.

### Create simpler tools

Any user should be able to use the tools, with little to no training.

# Motivation and Objectives

## Introduction

### Tools are complex

Tools for biological data analysis often require a very technical set of skills.

### Create simpler tools

Any user should be able to use the tools, with little to no training.

### Tasks are repetitive

Analysing high quantities of data can be repetitive, especially if executed manually.

### Automate tasks

Automated systems should perform repetitive tasks, so that users can focus on their work.

### Information is scattered

Information is easy to acquire, but is often scattered through multiple platforms, services and institutions.

# Motivation and Objectives

## Introduction

### Tools are complex
Tools for biological data analysis often require a very technical set of skills.

### Create simpler tools
Any user should be able to use the tools, with little to no training.

### Tasks are repetitive
Analysing high quantities of data can be repetitive, especially if executed manually.

### Automate tasks
Automated systems should perform repetitive tasks, so that users can focus on their work.

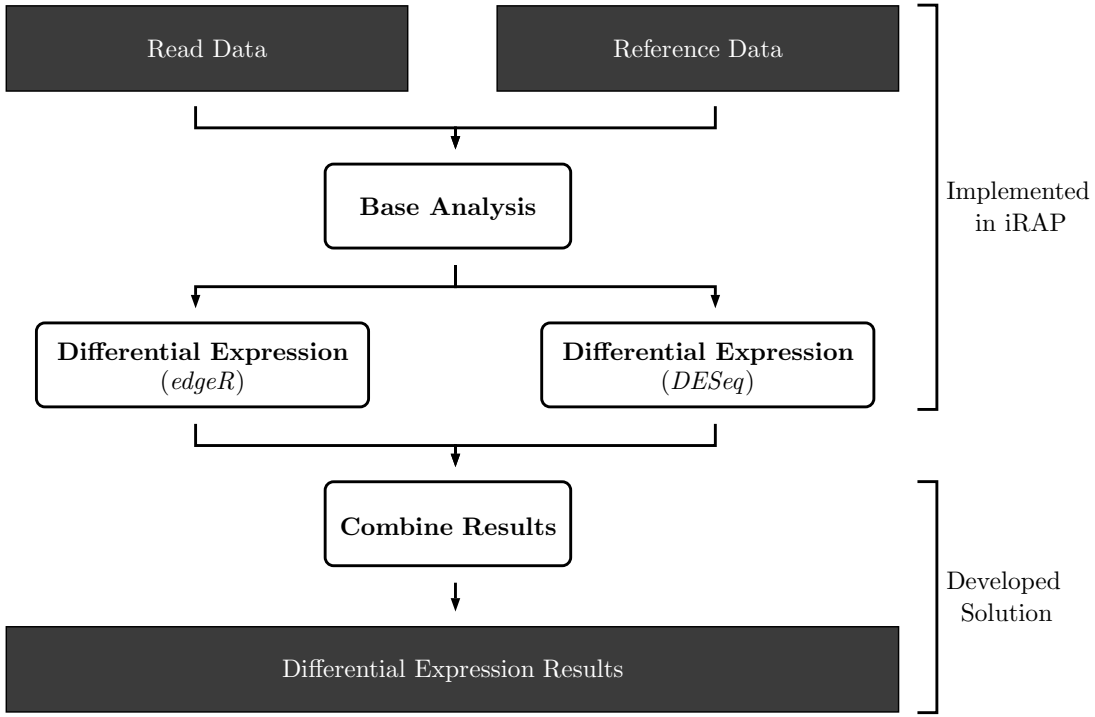
### Information is scattered
Information is easy to acquire, but is often scattered through multiple platforms, services and institutions.

### Gather information
Information should be contextually aggregated, allowing for quick access of relevant information.

- Two distinct problems warrant two different solutions.
- The developed system should be available anywhere, through the internet.
- The system should be as modular as possible, to allow future extensions.

# RNA-Seq Analysis Pipeline I

## Developed Solution

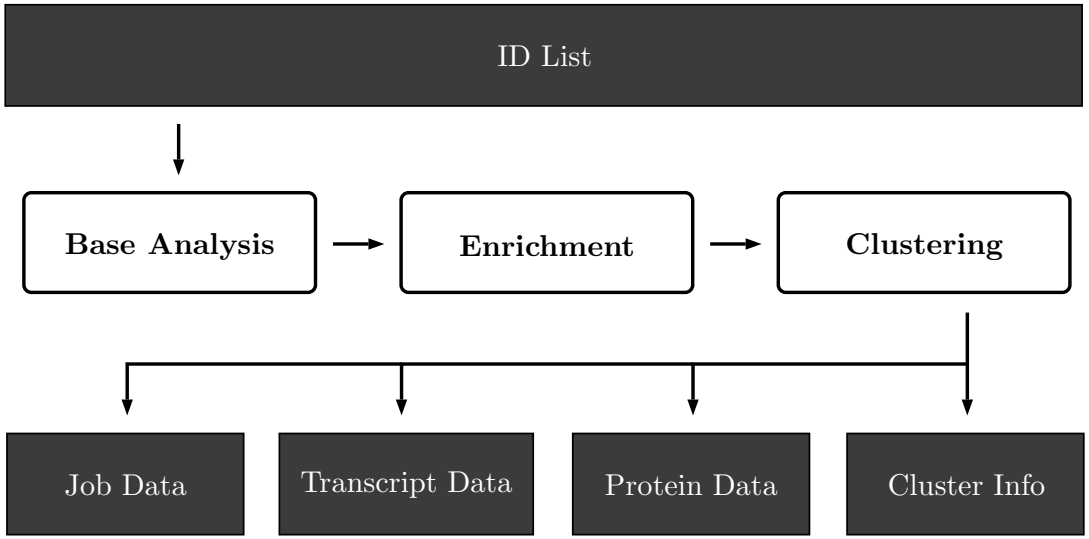



NOTES:

- Show scheme, refer iRAP, the script and the web interface.
- Refer multiple differential expression tools.
- Refer user input and tool configuration.

# RBP Analysis Pipeline (PBS Finder) I

## Developed Solution

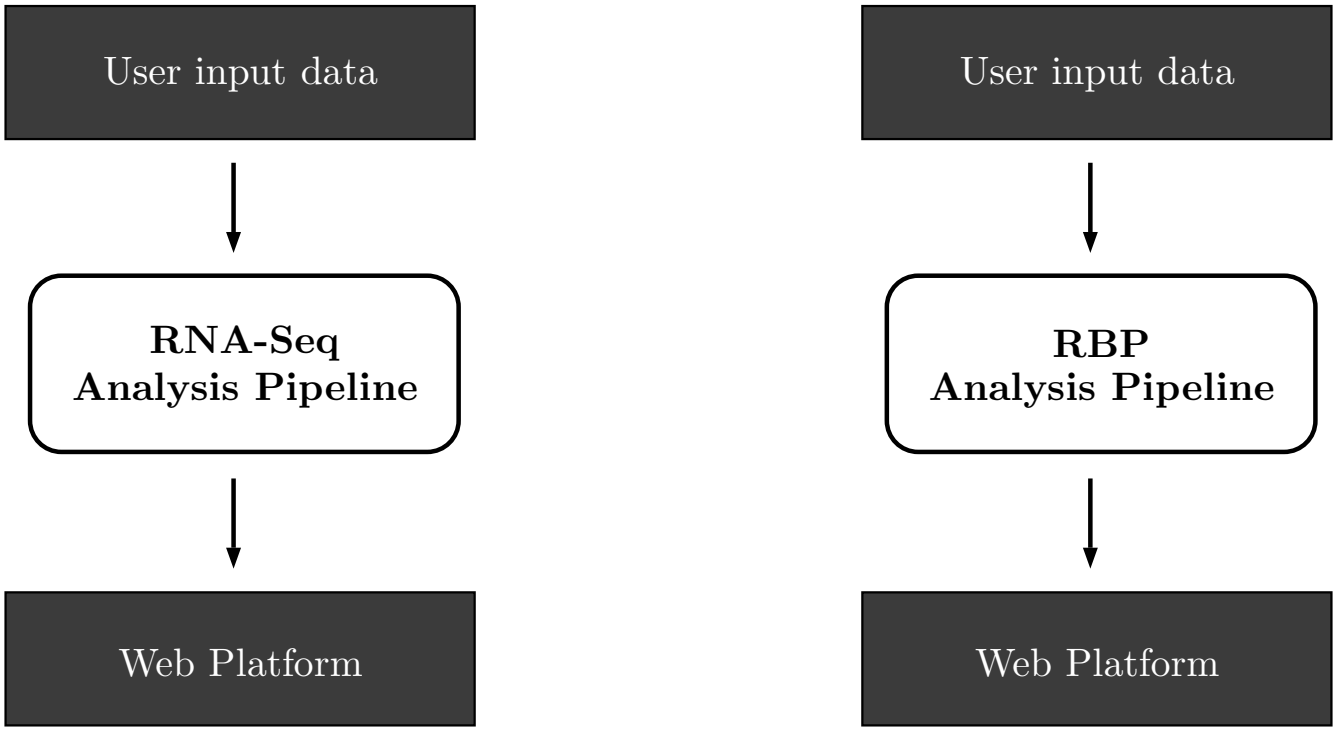



NOTES:
- Show scheme, RBP finding, gene enrichment, clustering.
- Refer web interface. - Refer that the tools is in production for several months, being extensively tested by experts.
- Refer user input and tool configuration.

# Integration

## Developed Solution

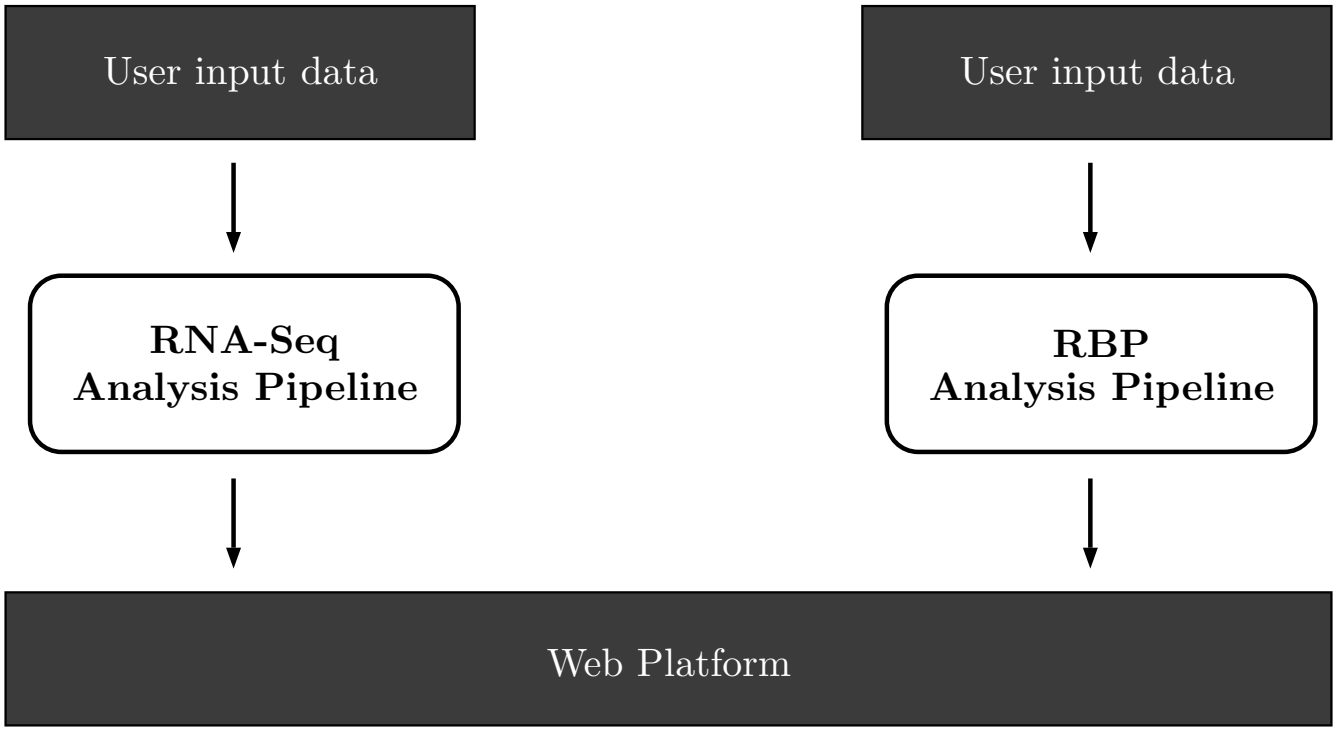
While focusing on aggregation and quick access to information, does it make sense to separate the results into two different platforms?



# Integration

## Developed Solution

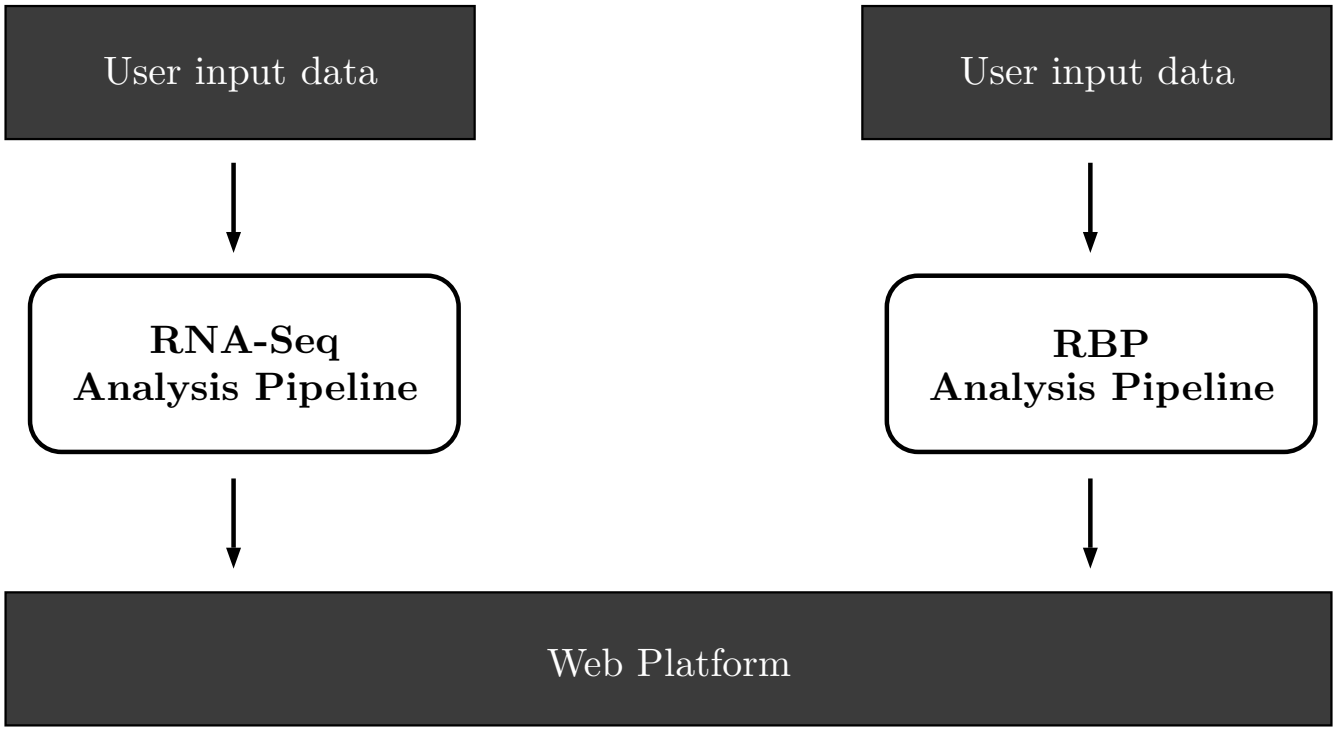
While focusing on aggregation and quick access to information, does it make sense to separate the results into two different platforms?



# Integration

## Developed Solution

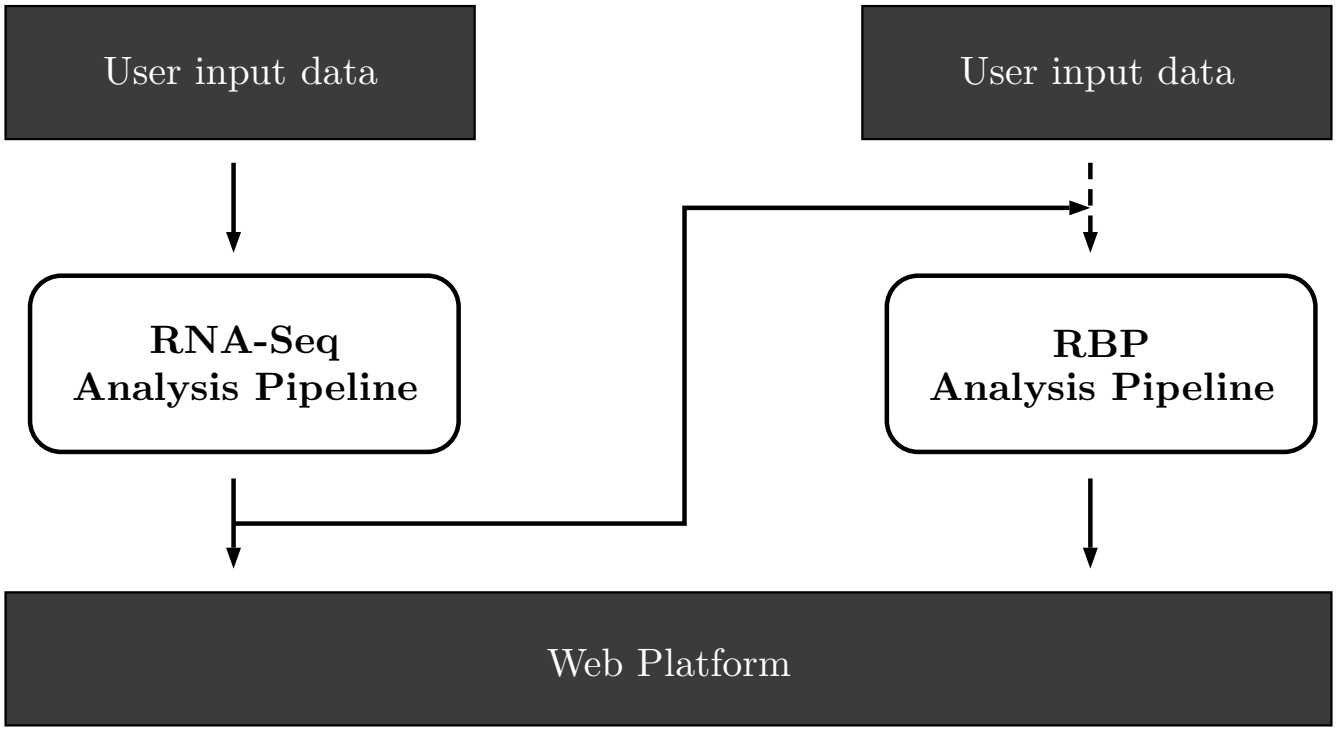
A list of differentially expressed genes is not very useful without further information about those genes. Does it make sense for a user to launch a new gene enrichment task by hand?



# Integration

## Developed Solution

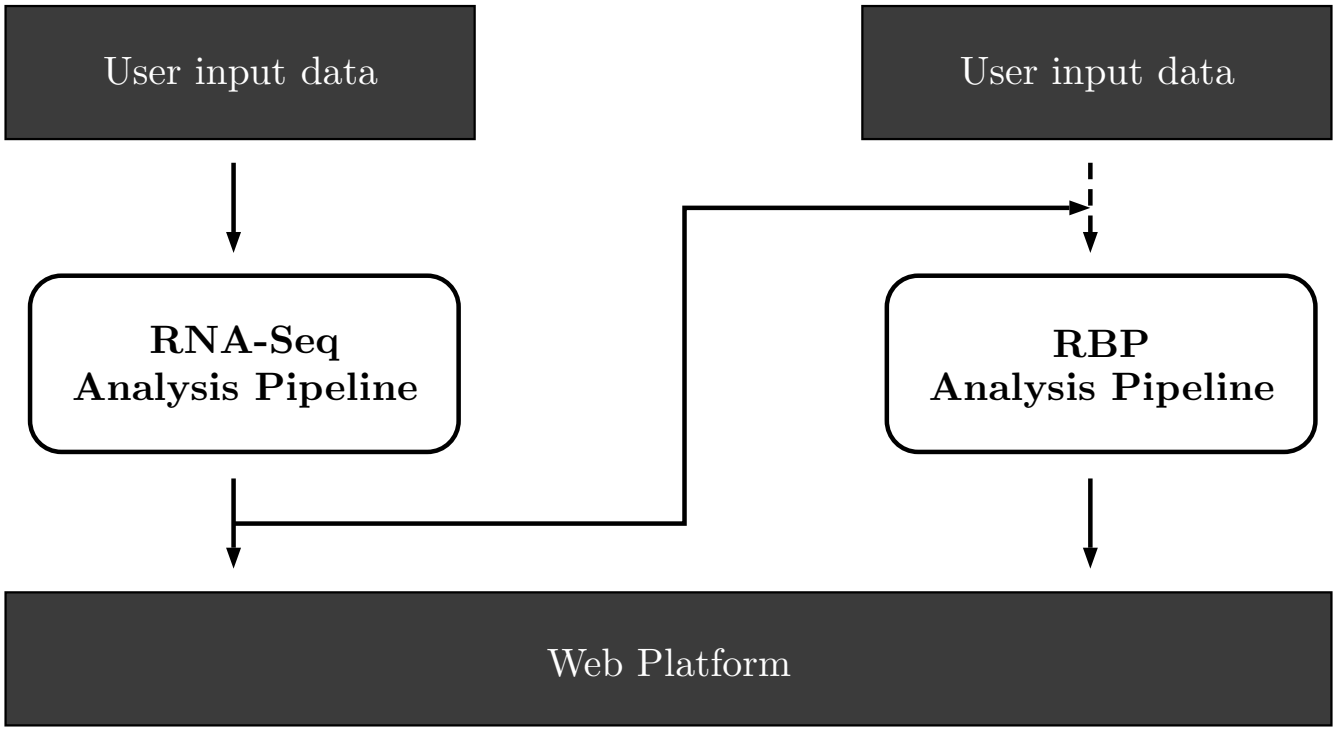
A list of differentially expressed genes is not very useful without further information about those genes. Does it make sense for a user to launch a new gene enrichment task by hand?



# Integration

## Developed Solution

A fully integrated solution: the analysis pipelines can be used separately or automatically executed in sequence; result visualization for both pipelines is isolated.

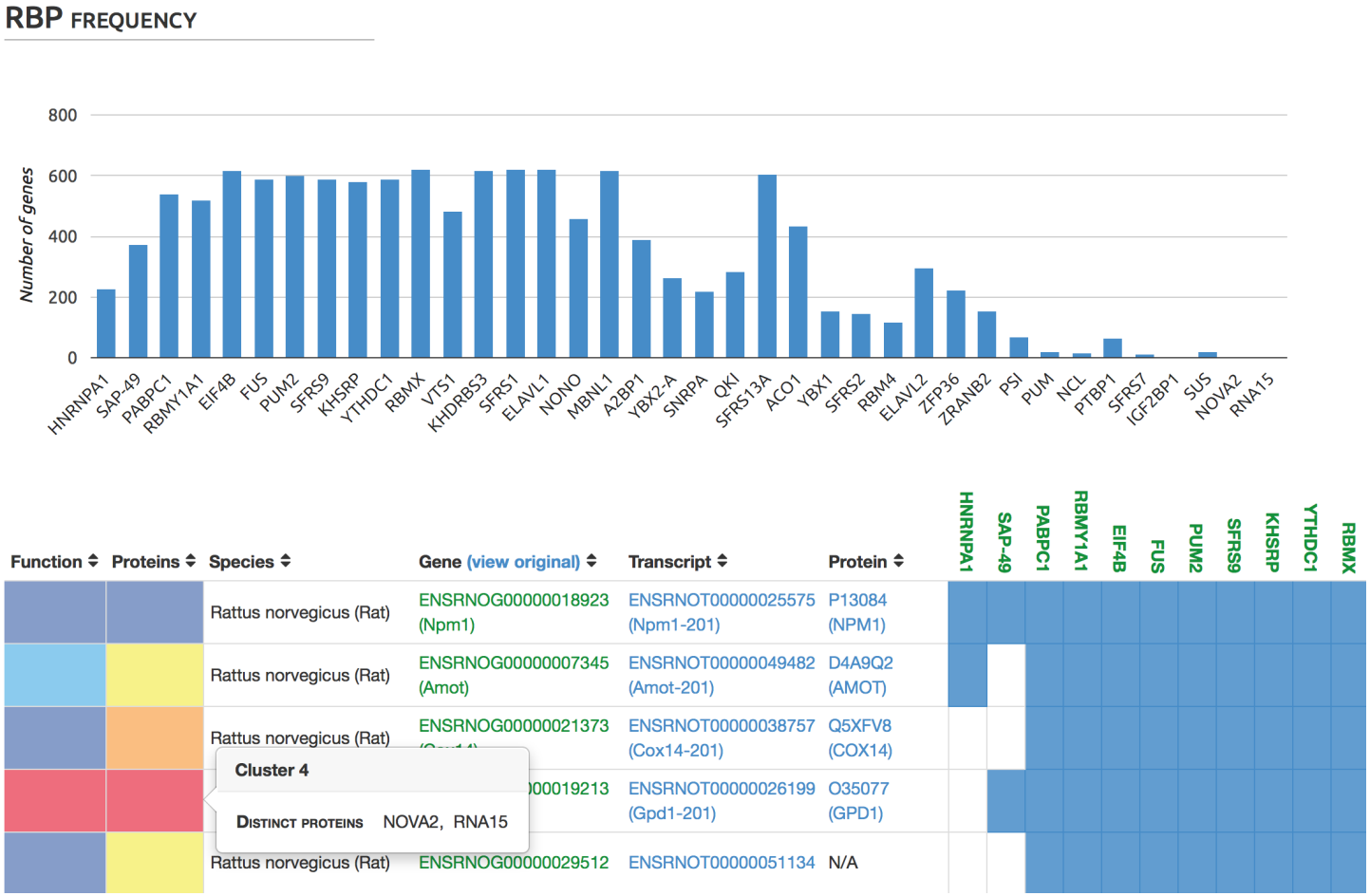


NOTES:
- Refer objectives, data and experimental method.
- Refer results.

# RBP Analysis Pipeline (PBS Finder) I

## Case Studies

Case study results viewed in PBS Finder:



# RBP Analysis Pipeline (PBS Finder) II

## Case Studies

NOTES:
- Refer objectives, data and experimental method.
- Refer results.
- Show screenshot.
- Show table.

# Objective Fulfilment

## Conclusions

- RBP analysis pipeline and web platform (PBS Finder) implemented and tested. PBS Finder has been in production for several months; during this time it was thoroughly tested by IBMC experts.
- RNA-Seq analysis pipeline implemented and tested (iRAP deployed and result consolidation tool implemented).
- Integration of both tools could not be accomplished.

# Future Work

## Conclusions

- Fully integrate the RNA-Seq analysis pipeline with the web platform (automatic job configuration, result visualization, etc.).

- Study the requirements for deploying the platform in large scale, and assess the feasibility of making it available internet-wide.

# Review

1. Introduction
   - Domain Problem
   - Motivation and Objectives
2. Developed Solution
   - Overview
   - RNA-Seq Analysis Pipeline
   - RBP Analysis Pipeline (PBS Finder)
   - Integration
3. Case Studies
   - RNA-Seq Analysis Pipeline
   - RBP Analysis Pipeline (PBS Finder)
4. Conclusions
   - Objective Fulfilment
   - Future Work

# A Computational Platform for Gene Expression Analysis

Diogo Teixeira[1]

Supervisors: Rui Camacho[2], Nuno Fonseca[3]

[1]Check affiliation

[2]Check affiliation

[3]Check affiliation

July 2014