

A Computational Platform for Gene Expression Analysis

Diogo Teixeira¹

Supervisors: Rui Camacho², Nuno Fonseca³

¹Check affiliation

²Check affiliation

³Check affiliation

July 2014

- 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
- 5 References

Domain Problem

Introduction

NOTES:

- Molecular biology is a new subject.
- Many unknowns.
- Gene expression is important...
- Two areas of interest of IBMC experts: differential expression and RBPs.

Motivation and Objectives

Introduction

Tools are complex

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

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.

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.

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.

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.

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.

Overview

Developed Solution

- Two distinct problems warrant two different solutions.
- The developed system should be available anywhere, through the internet.
- The system's footprint should be small enough to allow deployment in almost any available hardware.

RNA-Seq Analysis Pipeline

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)

Developed Solution

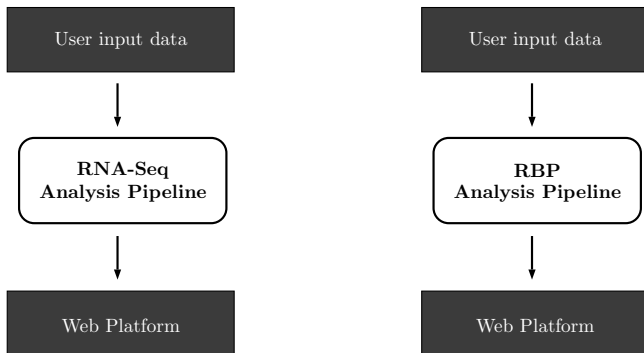
NOTES:

- 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

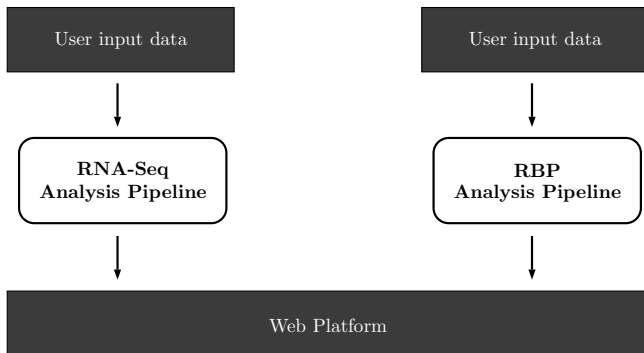
While focusing on aggregation and quick access to information, does it make sense to separate the results into two different platforms?



Integration

Developed Solution

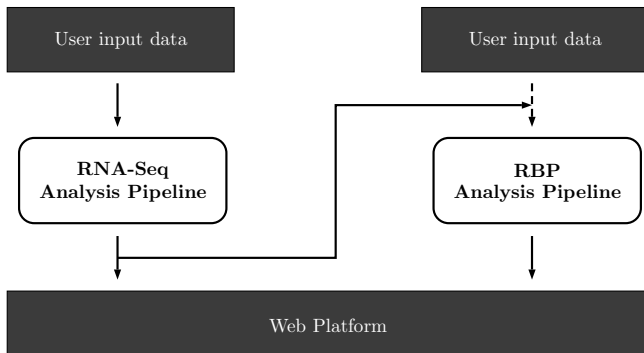
A list of differentially expressed genes is not very useful without further information about those genes. Does it make sense for a user to launch a new gene enrichment task by hand?



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.



RNA-Seq Analysis Pipeline

Case Studies

NOTES:

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

RBP Analysis Pipeline (PBS Finder)

Case Studies

NOTES:

- Refer objectives, data and experimental method.
- Refer results.
- Maybe show screenshot.

Objective Fulfilment

Conclusions

NOTES:

- All PBS Finder functionality implemented.
- Differential expression analysis pipeline deployed, combination script created.
- It was not possible to integrate with web platform, due to time constraints.

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.

- 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
- 5 References

Bibliography

A Computational Platform for Gene Expression Analysis

Diogo Teixeira¹

Supervisors: Rui Camacho², Nuno Fonseca³

¹Check affiliation

²Check affiliation

³Check affiliation

July 2014