Some CIPF tools

CellBase, Variant, GenomeMaps and Renato

Megasequencing Data Analysis

Valencia, 12th December 2012

Marta Bleda Latorre

mbleda@cipf.es

Department of Computational Medicine
Centro Investigación Príncipe Felipe

PRINCIPE FELIPE
CENTRO DE INVESTIGACION

ciberer BIER
PLATAFORMA BIOINFORMÁTICA DE ENFERMEDADES RARAS

er ciberer
CENTRO DE INVESTIGACIÓN BIOMÉDICA EN RED
DE ENFERMEDADES RARAS

Overview

1. CellBase
2. Variant
3. GenomeMaps
4. Renato
5. Summary

CellBase

CellBase

Motivation

- Exponential growth in the number and size of biological databases and repositories. Data size can reach hundreds of gigabytes and involves serious problems of data access through Internet and local disks.
- Biological information is spread out in different databases and repositories (~1380), using different identifiers → Problem when analyzing genome-wide experiments.
- Researchers have to deal with complex scripts (in Perl, generally) or parse horribles XML files.

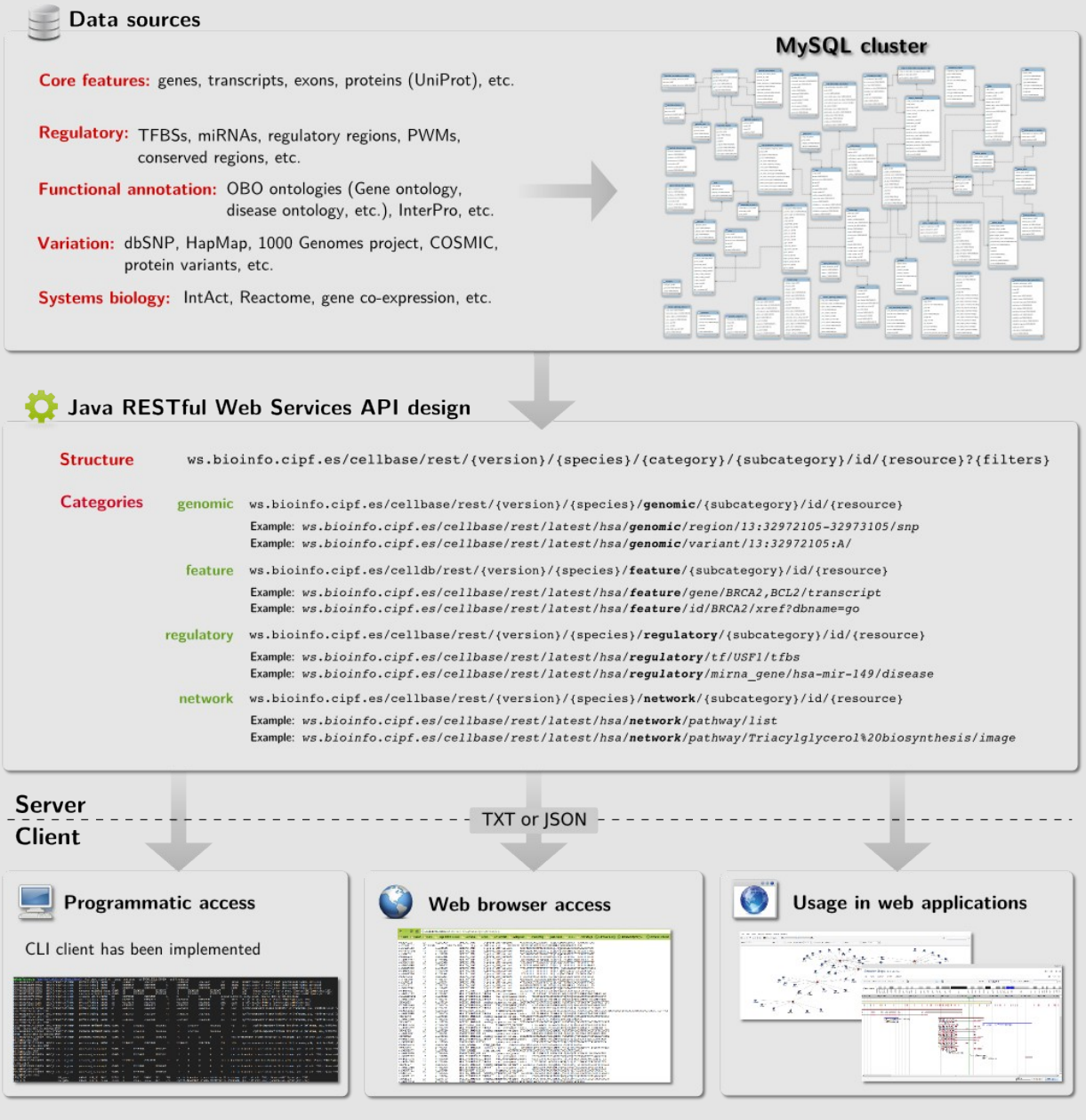
Goals

- Join the most relevant and high quality biological information in a single database.
- Facilitate accessibility to users.

CellBase

Overview

- A comprehensive integrative database and RESTful Web Services API.
- More than 120GB of data and ~100 SQL tables containing the most relevant biological information.
- Available for 11 species: human, mouse, rat, zebrafish, fruitfly, worm, yeast, dog, pig, mosquito and plasmodium.
- Accessible via CLI, Web Browser and Web applications.



CellBase

WS Structure

General Structure

ws.bioinfo.cipf.es/cellbase/rest/{version}/{species}/{category}/{subcategory}/id/{resource}?{filters}

http://ws.bioinfo.cipf.es/cellbase/rest/latest/hsa/feature/gene/BRCA2/mutation

Categories

- Genomic
  Subcategories: *region*, *variant* and *position*
  http://ws.bioinfo.cipf.es/cellbase/rest/latest/hsa/genomic/region/1:3972105-12973105/gene
- Feature
  Subcategories: *gene*, *transcript*, *exon*, *protein*, *snp* and *karyotype*
  http://ws.bioinfo.cipf.es/cellbase/rest/latest/hsa/feature/snp/rs3934834/phenotype
- Regulatory
  Subcategories: *mirna_gene*, *mirna_mature* and *tf*
  http://ws.bioinfo.cipf.es/cellbase/rest/latest/hsa/regulatory/mirna_gene/hsa-mir-95/disease
- Network
  Subcategories: *pathway*
  http://ws.bioinfo.cipf.es/cellbase/rest/latest/hsa/network/pathway/Triacylglycerol%20biosynthesis/info

CellBase

Documentation

Documentation site

http://docs.bioinfo.cipf.es/projects/cellbase/wiki

Article

Published online 12 June 2012 *Nucleic Acids Research, 2012, Vol. 40, Web Server issue* ***W609–W614***
doi:10.1093/nar/gks575

CellBase, a comprehensive collection of RESTful web services for retrieving relevant biological information from heterogeneous sources

Marta Bleda[1,2], Joaquin Tarraga[1,3], Alejandro de Maria[1], Francisco Salavert[1,2], Luz Garcia-Alonso[1], Matilde Celma[4], Ainoha Martin[4], Joaquin Dopazo[1,2,3,*] and Ignacio Medina[1,3,*]

Variant

Variant

Motivation

- Exome and genome sequencing are a promising instrument for finding novel mutations in human disorders. However, massive sequencing experiments reveal an enormous amount of genomic variation.
- Finding the causal mutation can represent a big challenge.
- Existing tools do not analyze all functional features, require the user to download biological annotations, do not use standard file formats or even need data to be processed several times to extract the necessary information.

Goals

- Create a variant annotation tool which reports rich functional information.
- Avoid installations or data downloads.
- Do not sacrifice speed.

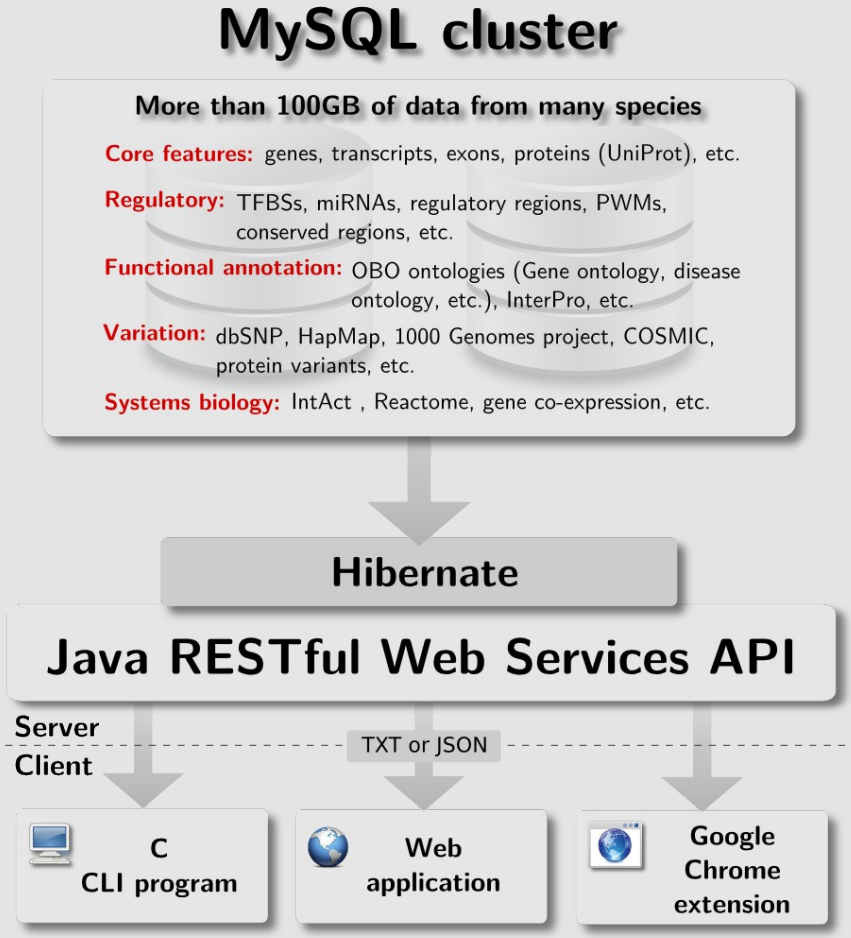
Variant

Overview

- A tool for predicting the effect of a genomic variant.

  VARIANT = VARIant ANalysis Tool

- Is a *cloud* variant annotator. Requires no installation or updates.
- Report
  - Conventional consequence type
  - Regulatory information
  - Conserved region
  - Annotated SNPs and variants
  - Phenotypic information
- Accessibility
  - Web application and web services
  - C Client
  - Chrome Extension



Variant

Web Service

http://ws.bioinfo.cipf.es/cellbase/rest/latest/hsa/genomic/variant/13:32332472:T:G/consequence_type

http://ws.bioinfo.cipf.es/cellbase/rest/latest/hsa/genomic/variant/10:52575931:G/mutation_phenotype

Web application

http://variant.bioinfo.cipf.es/

Variant

Documentation

Documentation site

http://docs.bioinfo.cipf.es/projects/variant/wiki

Article

W54–W58 Nucleic Acids Research, 2012, Vol. 40, Web Server issue Published online 11 June 2012
doi:10.1093/nar/gks572

VARIANT: Command Line, Web service and Web interface for fast and accurate functional characterization of variants found by Next-Generation Sequencing

Ignacio Medina[1], Alejandro De Maria[1], Marta Bleda[1,2], Francisco Salavert[1,2], Roberto Alonso[1], Cristina Y. Gonzalez[1] and Joaquin Dopazo[1,2,3,*]

GenomeMaps

GenomeMaps

Introduction

Motivation

- Genome browsers are extremely useful to represent, compare genomic information. The most popular: Ensembl and UCSC.
- However, with the increase of genomic data, these browsers are becoming slower, less efficient and difficult to manipulate.
- Cannot visualize you own data.

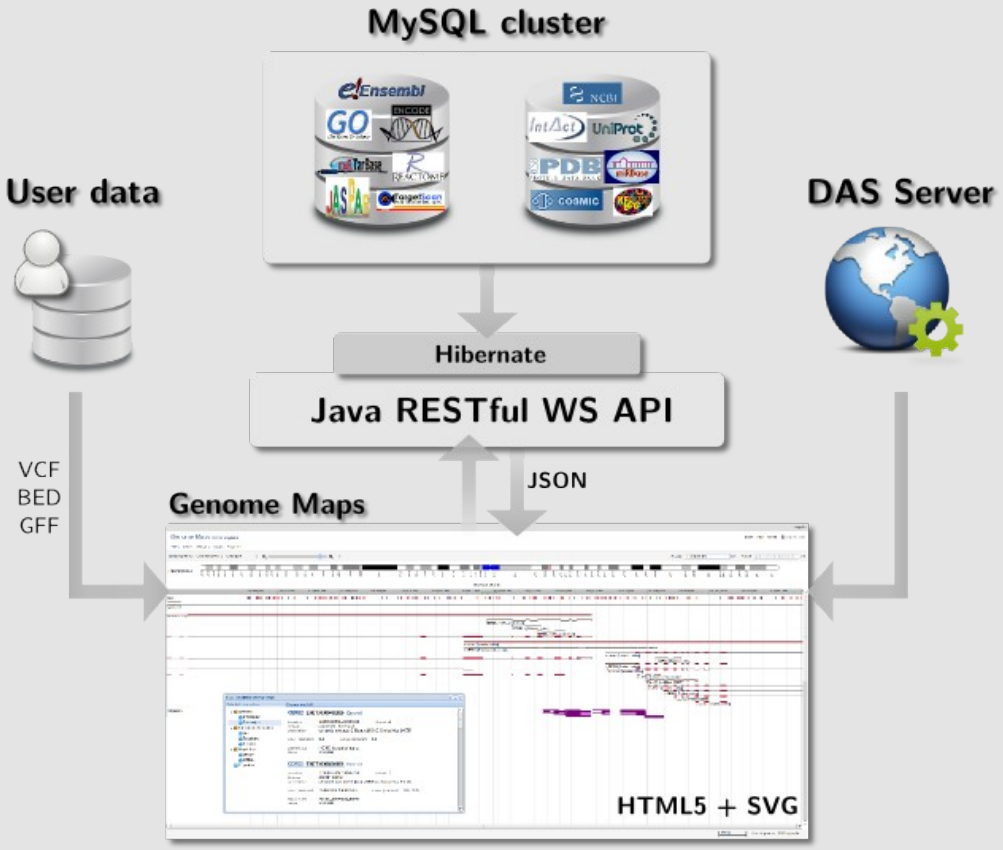
Goals

- Create a useful and user-friendly genome browser rich in biological information.
- Fast and efficient.
- Permit users visualize their own data.

GenomeMaps

Overview

http://www.genomemaps.org/

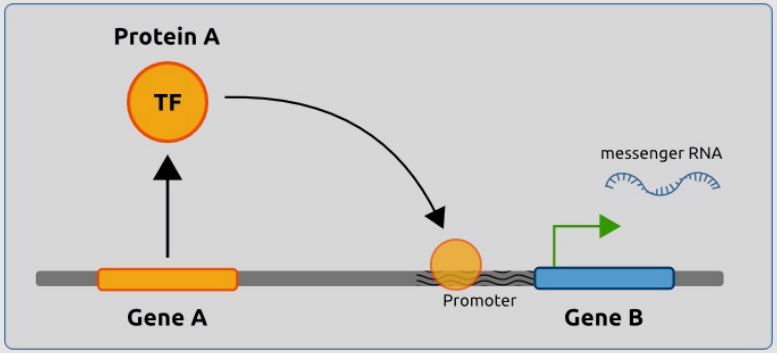


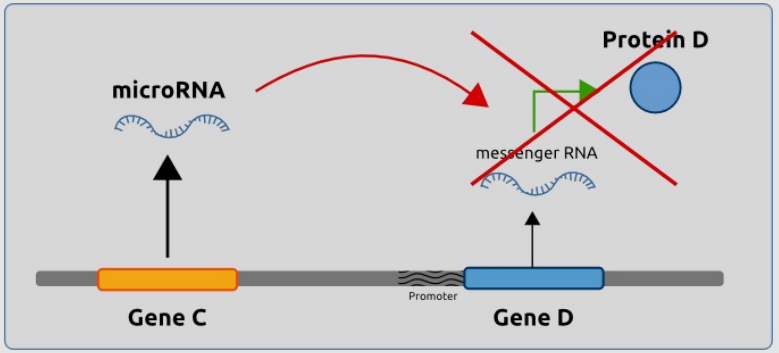
Renato

Renato

Motivation: Importance of studying Gene Regulatory Networks

- Genes are not independent. They regulate each other and act collectively → Action of Transcription Factors (Tfs) and microRNAs (miRNAs).
- Control of gene expression is crucial for normal development and maintenance of healthy cells.
- Potential discovery of triggering mechanism and treatments for disease.





Renato

Hypothesis

There are not hundreds of errors (mutations, deregulation, etc.) but one or few common causes leading to the abnormal/different expression of these genes. Regulatory elements, which generally interact and regulate several genes, are features of interest because of its potential to cause this deregulated profile.

Goals

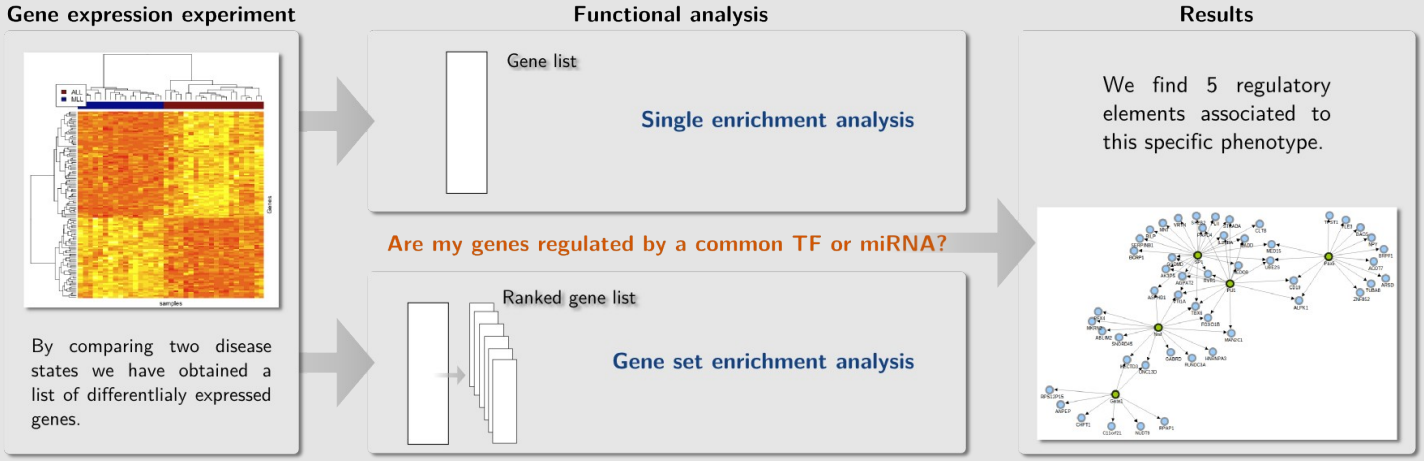
- Integrate regulatory information from different sources in a single regulatory database.
- Implement a Web Service API (Application Programming Interface) to make these resources programmatically accessible.
- Implement a tool designed to identify common regulatory elements in a list of genes.

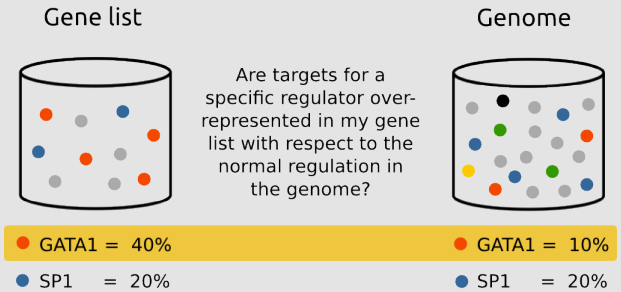
Renato

Overview

- RENATO has been designed to identify common regulatory elements in a list of genes. It maps such genes to the regulatory network, extracts the corresponding regulatory connections and evaluates each regulator for significant over-representation of targets in the list.

  RENATO = REgulatory Network Analysis TOol





The gene list is over-enriched for GATA1 targets.

	Gene list	Genome
Regulated by GATA1	4	2
Not regulated by GATA1	6	18

Renato

Usage

Web Services

- Get all genes regulated by a TF called GATA1:
  http://ws.bioinfo.cipf.es/cellbase/rest/latest/hsa/regulatory/tf/GATA1/target_gene

- Get all miRNAs that regulate a gene called BRCA2:
  http://ws.bioinfo.cipf.es/cellbase/rest/latest/hsa/feature/gene/BRCA2/mirna_target

- Get all diseases related with a miRNA:
  http://ws.bioinfo.cipf.es/cellbase/rest/latest/hsa/regulatory/mirna_gene/hsa-mir-95/disease

Web application

http://renato.bioinfo.cipf.es/

Renato

Documentation

Documentation site

http://bioinfo.cipf.es/docs/renato/

Article

W168–W172 *Nucleic Acids Research, 2012, Vol. 40, Web Server issue* *Published online 11 June 2012*
doi:10.1093/nar/gks573

Inferring the regulatory network behind a gene expression experiment

Marta Bleda[1,2], Ignacio Medina[1], Roberto Alonso[1], Alejandro De Maria[1], Francisco Salavert[1,2] and Joaquin Dopazo[1,2,3,*]

Summary

- CellBase

  Extract heterogeneous biological information easily and fast

- Variant

  Annotate genomic variants

- GenomeMaps

  Explore the genome and visualize variants in the genomic context

- Renato

  Identify common regulatory elements in a list of genes

Thank you!

Questions?