Some CIPF tools

CellBase, Variant and GenomeMaps

Megasequencing Data Analysis
Valencia, 12th July 2012

Marta Bleda Latorre
mbleda@cipf.es
Department of Computational Genomics
Centro Investigación Príncipe Felipe

PRINCIPE FELIPE
CENTRO DE INVESTIGACION

ciberer
CENTRO DE INVESTIGACIÓN BIOMÉDICA EN RED
DE ENFERMEDADES RARAS

Overview

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

CellBase

CellBase

Introduction

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.

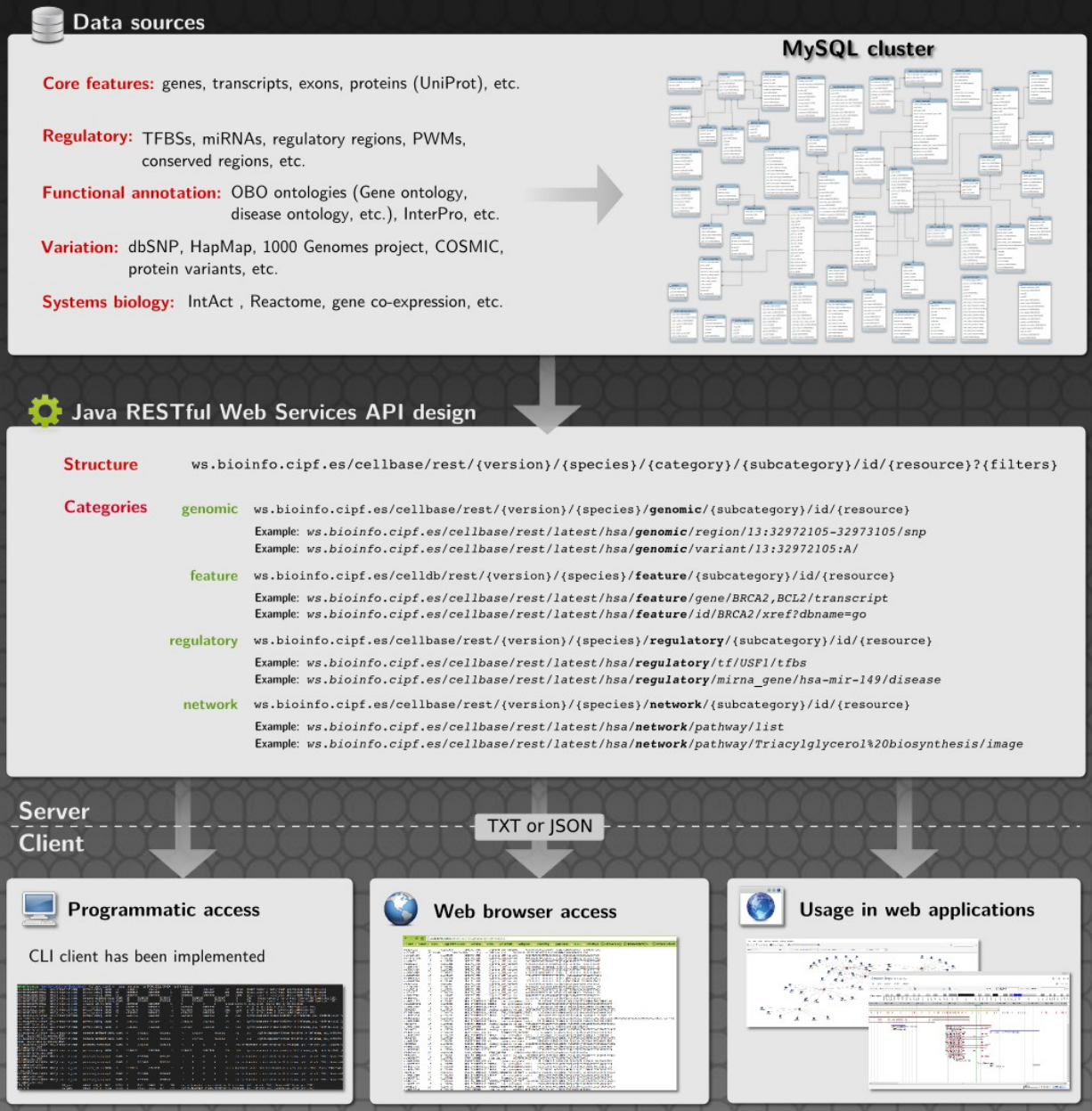
Goals

- 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

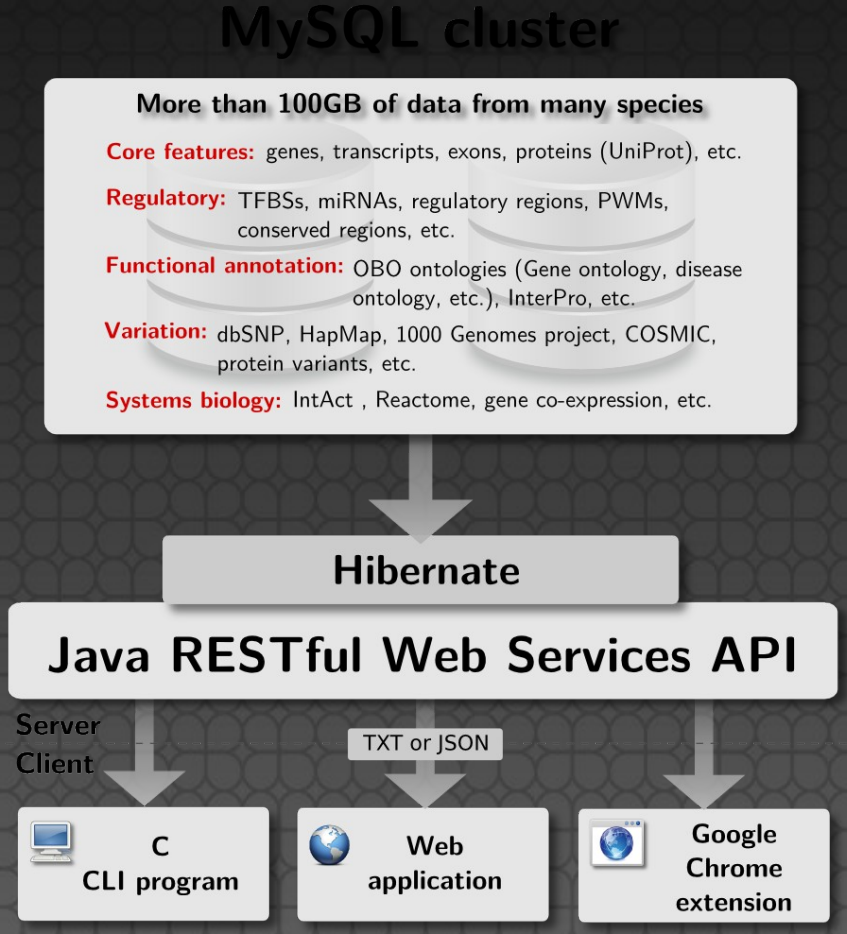
- A tool for predicting the effect of a genomic variant.

  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

- Accessibility
  - Web application and web services
  - C Client
  - Chrome Extension



Variant

Usage

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.

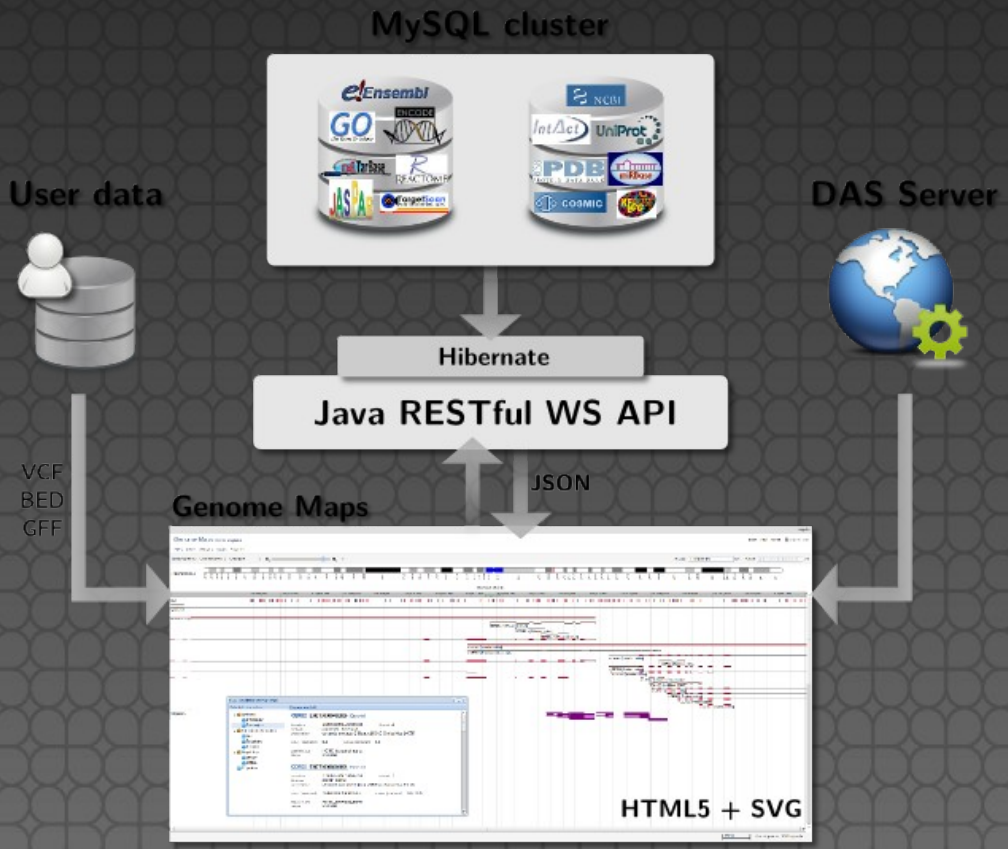
Goals

- 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/



Summary

- CellBase

  Extract heterogeneous biological information easily and fast

- Variant

  Annotate genomic variants

- GenomeMaps

  Explore the genome and visualize variants in the genomic context

Thank you!

Questions?