Bioinformatics Unit

Centre de Regulació Genòmica

Centres de recerca de Catalunya

OVERVIEW

Starting in September 2009, this unit has contributed to different research projects by providing support and expertise in programming and advanced data analysis, focusing primarily on high-throughput genomics technologies including microarrays, genotyping and next-generation sequencing. The unit is open to all researchers at the PRBB and also to external users anywhere in the world. For further information please contact Guglielmo Roma (email: guglielmo.roma@crg.es; phone: +933160181)

MEMBERS

Alibes. Andreu
Cozzuto, Luca
Hermoso, Antonio
Lowy, Ernesto
Mancuso, Francesco
Roma, Guglielmo

NGS Technologies

Our group is heavily involved in the analysis and interpretation of deep sequencing data. During 2010 we have implemented several pipelines capable of analyzing different types of data such as RNA-seq, resequencing data for polymorphism detection, *De novo* assembly , ChIP-Seq and functional annotation in new genome projects. These pipelines can be customized depending on the customer needs.

Samples

•Chromatin immunoprecipitation
• Immunoprecipitation of an RNA-binding protein
• Whole transcriptome
• Whole genome
• Targeted & enriched sequences
• Methylated DNA

Sequencers

De novo assembly (genome or transcriptome)

• Definition of large stretch of DNA or whole chromosomes by analyzing both 454 and HISEQ data.
• Setup of custom Genome Browser for showing results.

Functional annotation

•Annotation of coding and non coding genes
•Functional domain prediction
•Molecular Function assignment an specification of Gene Ontology (GO) terms
•Protein-protein interactions
•Mapping of proteins into metabolic pathway and reactions

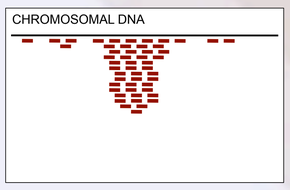
Short reads

AATGCGTACATGCACCANTTCAGGTC
TGTCANNTGCATTACATGCATTGACC
TGTCANNTGCATTACATGCATTGACC
AATGCGTACATGCACCANTTCAGGTC
TGTCTTTTGCATNACATGCAAAAACC
TGTCTTTTGCATNACATGCAGGGACC

Detection of alternative splicing, differential expression (RNA-seq)

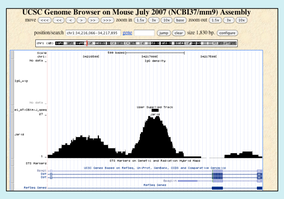
• Analysis of transcripts abundance
• Detection of novel splicing sites
• Measure of differential gene expression

Mapping to reference genome

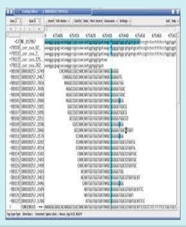


Detecting enriched regions (Chip-seq, RIP-seq)

• Statistical evaluation of enriched regions
• Data displayed in Genome Browser
• Detection of enriched motifs in binding sites
• Analysis of gene enrichment



SNP, indels, structural variation

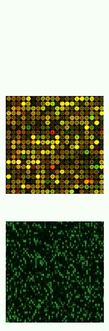


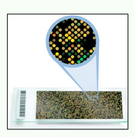
Whole Genome resequencing

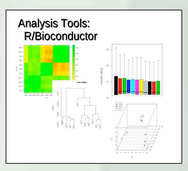
Targeted resequencing

Qualilty control with TEQC (http://www.bioconductor.org)

Gene expression arrays







preprocessing

Normalization

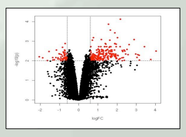
Raw Data

Data inspection & Quality Control

Statistical testing

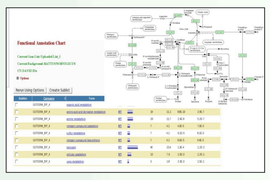
Filtering for differentially expressed genes

Functional analysis



STATISTICAL TESTING
• Performed between compared sample groups (e.g. Treated vs. Controls)
• Provides measures that describe the **size** and **reliability of the change in the** expression levels between conditions
- **Fold-change for the size of the change**
- **P-values and false discovery rates for** the reliability of the change

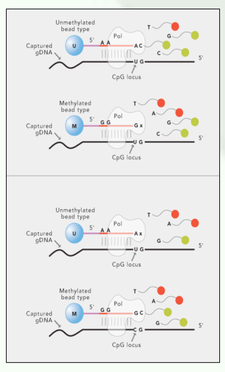
FUNCTIONAL ANALYSIS
• David (GO, KEGG, other ontologies)
• Ingenuity
• GSEA
• Motif analysis (known and *de-novo*)
• Specific request



Summary of the data analysis workflow for gene expression data

Methylation arrays

Infinium HumanMethylation27



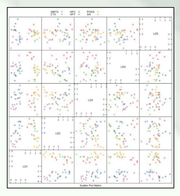
Preprocessing with HumMeth27QCReport (*)

Raw Data from BeadStudio

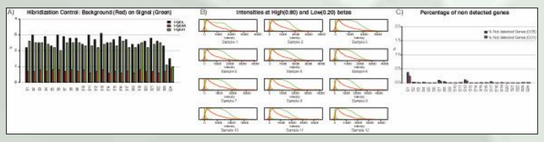
Data inspection & Quality Control

Normalization

Statistical testing



(*) **HumMeth27QCReport** is an R package developed in our group in collaboration with the CRG Genotyping Unit available in the CRAN repository (http://cran.r-project.org/)

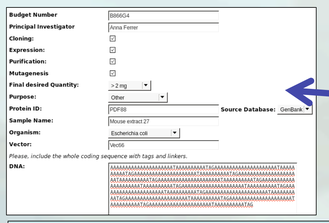


STATISTICAL TESTING
• Linear Discriminant Analysis (lda)
• Principal Component Analysis (PCA)
• Clustering
• Specific requests

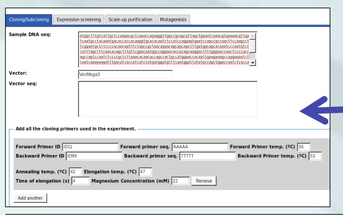
Summary of the data analysis workflow for methylation data

Web & database development

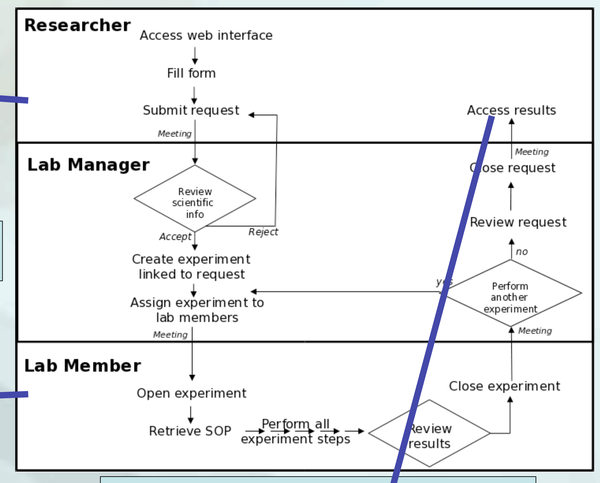
We are developing and deploying easy-to-use lab data management systems based on Wikipedia technology



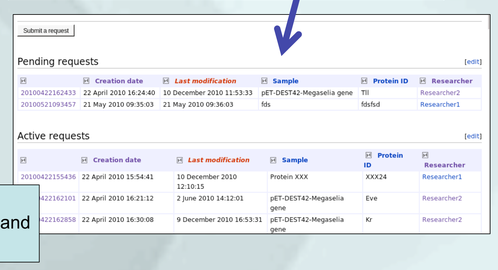
Request page: Researchers can fill themselves the request to be submitted.



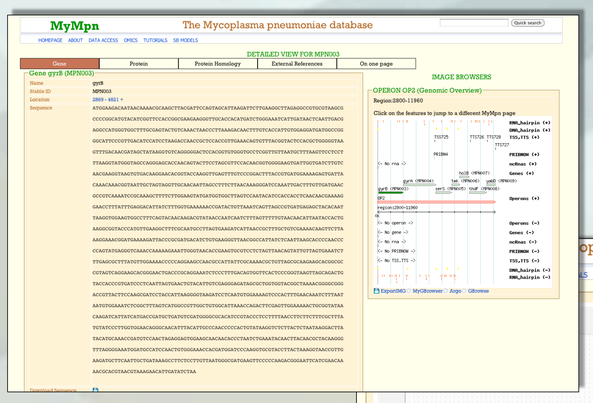
Experiment page: Lab members enter individually or collaboratively all the data related to the ongoing experiments.



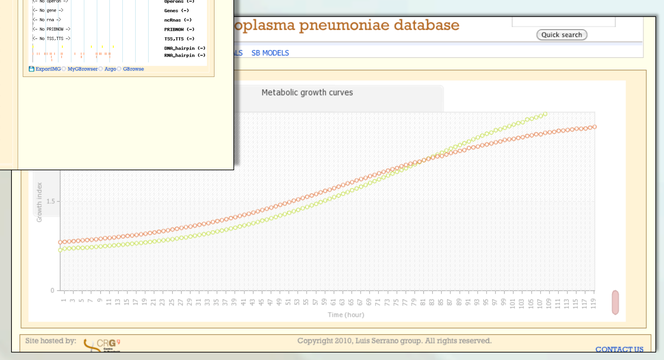
Schematic workflow behind Lab Service wiki



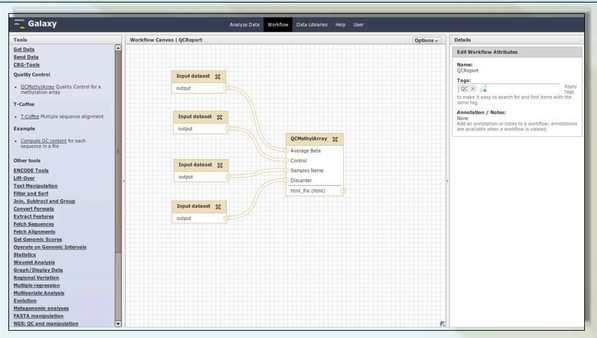
Requests and experiments listing: researchers and lab members can track the state and progress of their requests and their associated experiments.



We are developing the **Mycoplasma pneumoniae** webserver, including a customized Biomart

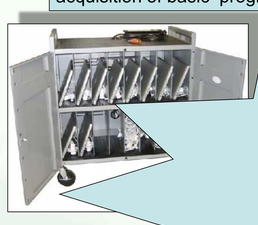


And we plan to have the bioinformatics tools designed at CRG available through a local implementation of **Galaxy**



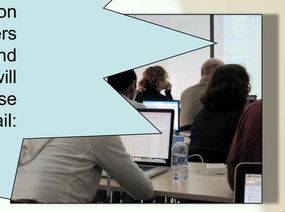
Bioinformatics tutorials & Workshops

The unit is also focused on the education of biologists in the analysis and interpretation of their experimental data through the provision of custom advice and training in bioinformatics as well as in the acquisition of basic programming skills.



Announce!

On February the 7th, Robert Kuhn, Associate Director of the genome browser will give a hands-on interactive workshop designed to help new users become familiar with the basic paradigm and navigation of the site. Experienced users will discover some of the less obvious tools. Please contact Magali Bartomeus to register (email: magali.bartomeus@crg.es; phone : 933160161)



Special Thanks to:
Bioinformaticians from other units : Manuela Hummel, Sarah Bonnin, Debayan Datta

And to:
Giancalrlo Castellano (Gene Regulation),