CAPER 2.0
An interactive, configurable and extensible workflow-based platform to analyze datasets from C-HPP

Dong Li and Fuchu He
Chinese C-HPP consortium
2013.9.16
CAPER was developed to visualize human proteomic datasets based on chromosomes.

Track view
To show the sequence and site information and the corresponding relationship between proteome, transcriptome and genome.

HeatMap view
To show evidence (traffic light) and quantitative (heat color) functional annotation information for proteins.

Protein annotation page

http://www.bprc.ac.cn/CAPE

Why do we need CAPER 2.0?

**Requirements of Web browser from C-HPP Research**

- **Analyses**
  - Not only static presentation, but also dynamic data analyses
- **Extensibility**
  - Easy addition of new bioinformatics tools and workflows integrating multiple tools
- **Configurability**
  - Simple configuration of research analyses pipelines based on flexible requirements
- **Interactivity**
  - Graphical visualization and excellent interoperability
CAPER 2.0 was developed based on the Galaxy framework

http://wiki.g2.bx.psu.edu/Admin/Get%20Galaxy
Main interface of CAPER 2.0

Tools

Workflow edition

Tool configuration
Published tools and workflows

**Published tools**

**C-HPP tools**
- Map peptides to chromosomes to find their chromosome locations
- Protein classification based on existence evidence at different levels
- Find missing proteins based on public databases
- Find genes in a chromosome region based on the public database

**Galaxy toolbox**
- Statistics tool from R package

**Published workflows**

- To find missing proteins
- To map peptides to chromosomes
- To integrate C-HPP and ENCODE
- To associate C-HPP datasets with functional annotation
Finding missing proteins

A: The workflow of finding missing proteins, tools “Find missing proteins” and “Send gff to CAPER track-view” included

B1: The webpage displaying the output results of “Find missing proteins” tool

B2: Example of exhibiting missing proteins in track-view achieved by the “Send gff to CAPER track-view” tool
Mapping your identified peptides onto human genome

A: The workflow used to map the identified peptides to chromosomes, mainly composed of “Map peptides to chromosomes”, “Protein classification based on existence evidence at different levels” and “Send gff to CAPER track-view” tools

B₁: The webpage displaying the output results of “Protein classification” tool

B₂: Example of visualizing peptides in track-view achieved by the “Send gff to CAPER track-view” tool
Integrating genomic annotation and identified peptides

A: The workflow to integrate C-HPP peptides with ENCODE transcription factor binding sites

B: Example of the output of this workflow exhibiting the binding signal profile of transcription factor CEBPB which is supposed to target the coding gene (CRYZ) of the identified peptides
Associating identified proteins with functional annotations

A: The workflow for the functional annotation of the protein list

B₁ and B₂: The output of the “GOfact” and “KEGG pathway analyzer” tools

B₃: The protein interaction network graph showing the interaction relationship between the submitted proteins
• CAPER has been updated into a higher version
  CAPER 2.0 which is an interactive, configurable and
  extensive workflow-based platform to analyze the
  datasets from C-HPP

• More tools and workflows will be developed
  to meet the requirement from C-HPP......