



CAPER2.0 An interactive, configurable and extensible workflow-based platform to analyze datasets from the Chromosome-centric Human Proteome Project

Dan Wang^{1,2,3}#, Zhongyang Liu^{1,2,3}#, Feifei Guo^{1,2,3,4}, Lihong Diao^{1,2,3}, Yang Li^{1,2,3}, Xinlei Zhang⁵, Zechi Huang⁵, **Dong Li^{1,2,3}* and Fuchu He^{1,2,3,4}***

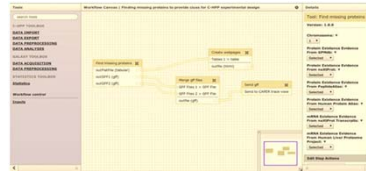
What is the update for CAPER 2.0?

The Chromosome-centric Human Proteome Project (C-HPP) aims to map and annotate the entire human proteome by the “chromosome-by-chromosome” strategy. As the C-HPP proceeds, the increasing volume of proteomic datasets presents a challenge for customized and reproducible bioinformatics data analyses for mining biological knowledge. To address this challenge, we updated the previous static proteome browser CAPER [J Proteome Res. 2013 Jan;12: 179-86] into a higher version, CAPER2.0 - an interactive, configurable and extensible workflow-based platform for C-HPP data analyses. In addition to the previous visualization functions of track-view and heatmap-view, CAPER2.0 presents a powerful toolbox for C-HPP data analyses, including the tools for finding missing proteins, mapping peptides to chromosomes, integrating peptides with ENCODE datasets and protein functional annotation etc.. CAPER2.0 also integrates a configurable workflow system, supporting the view, construction, edit, run and share of workflows. These features allow users to easily conduct their own C-HPP proteomic data analyses and visualization by CAPER2.0. The updated CAPER is available at

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



Index page of CAPER2 including Toolbox



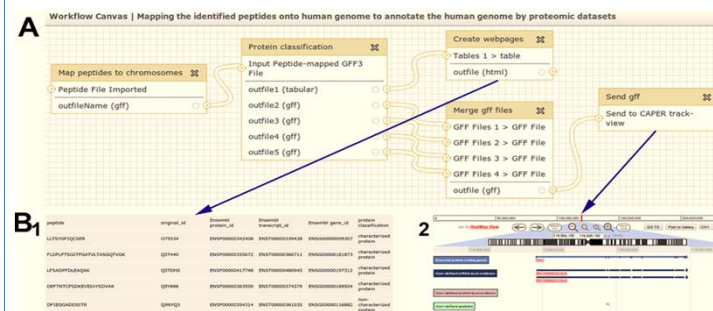
Graphical interface for workflows



Published workflows and their details

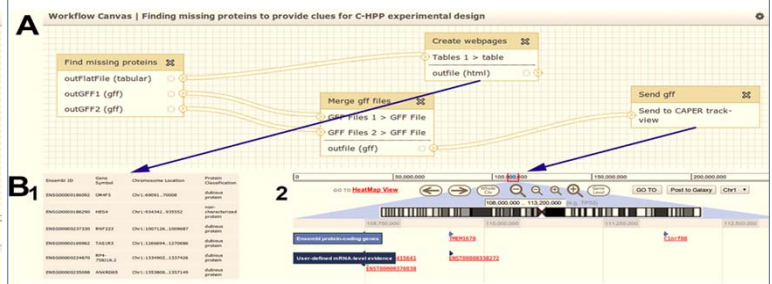
What can CAPER2.0 do?

Mapping the identified peptides onto human genome to annotate the human genome by proteomic datasets



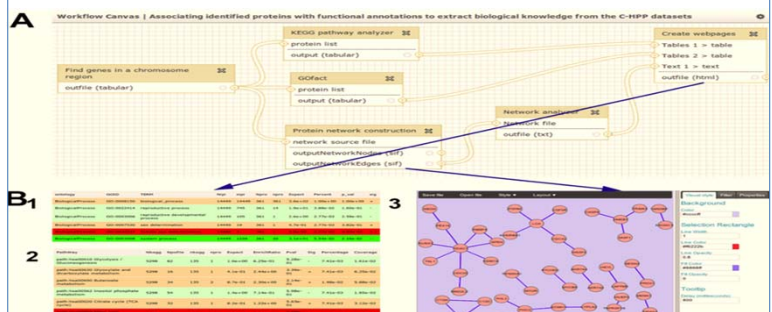
A: The workflow used to map the identified peptides to chromosomes, mainly composed of “Map peptides to chromosomes”, “Protein classification based on existence evidence at different levels” and “Send gff to CAPER track-view” tools
B₁: The webpage displaying the output results of “Protein classification” tool
B₂: Example of visualizing peptides in track-view achieved by the “Send gff to CAPER track-view” tool

Finding missing proteins to provide clues for C-HPP experimental design



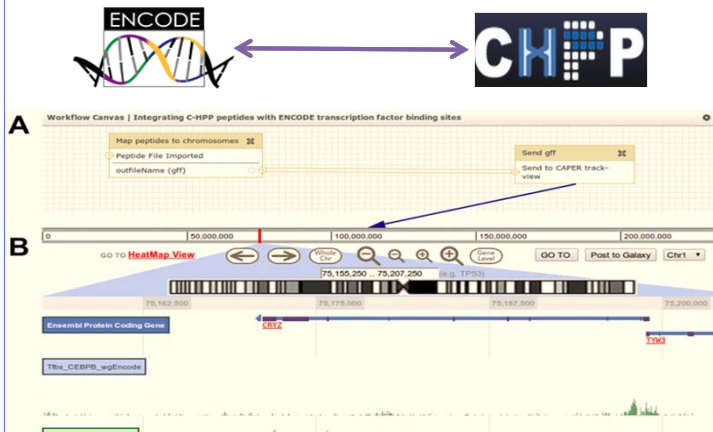
A: The workflow of finding missing proteins, tools “Find missing proteins” and “Send gff to CAPER track-view” included
B₁: The webpage displaying the output results of “Find missing proteins” tool
B₂: Example of exhibiting missing proteins in track-view achieved by the “Send gff to CAPER track-view” tool

Associating identified proteins with functional annotations to extract biological knowledge from the C-HPP datasets



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

Integrating genomic annotation and identified peptides to bridge ENCODE and C-HPP



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 (CRY2) of the identified peptides

These authors contribute equally to this work.
* To whom correspondence should be addressed. E-mail: hefc@nic.bmi.ac.cn.
¹ State Key Laboratory of Proteomics, Beijing Proteome Research Center, Beijing Institute of Radiation Medicine, Beijing 102850, China
² National Center for Protein Sciences Beijing, Beijing 102206, China
³ National Engineering Research Center for Protein Drugs, Beijing 100850, China
⁴ Institute of Basic Medical Sciences and School of Basic Medicine, Peking Union Medical College and Chinese Academy of Medical Sciences, Beijing 100005, China
⁵ Beijing Genestone Technology Ltd., Beijing 100085, China

