chromatin data interpretation

Eleni Mina

***Abstract*—¡p¿This research object, was created in order to further analyse and interpret the results from the research object http://sandbox.wf4ever-project.org/rodl/ROs/HD**$_chromatin_analysis/nbsp;(HDchromatinanalysis).Theworkflowsinthisresearchobjectareusingtheanniwebservices$
$/p>$

I. INTRODUCTION

This document provides a paper-style view of the Research Object (RO) "*chromatin data interpretation*"[1] generated . The RO has been created, managed and preserved via ROHub platform [1]. Please refer to [2] for a general introduction to the RO concept, to [3] for a detailed description of the RO model, and to [4] for more information about ROHub platform.

The RO is of type *"Basic", which represents a general aggregation of related resources.* [2].

An overview of this RO is depicted in Figure 1. In summary, the hypothesis of this work is described in *hypothesis.txt*[3]; Additionally, this RO has been enriched automatically with the following annotations:

- concepts (most frequently mentioned in the RO): *research, HD*
- domains (fields of knowledge in which the main concepts are commonly used): *genetics, memory*
- frequent expressions (most frequently mentioned noun phrases): *deregulate in HD, participate in epigenetic processes, have an epigenetic role*

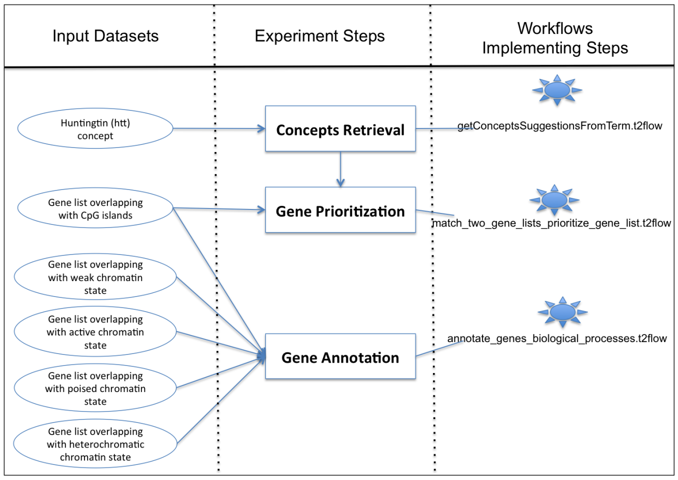

Fig. 1. Research Object Sketch

II. RESOURCES

The resources encapsulated by the RO are summarized in table I.

TABLE I
RESEARCH OBJECT RESOURCES

name	size	type
conclusions.txt	67.0 B	Conclusions
hypothesis.txt	63.0 B	Hypothesis
workflow_sketch_hd_chromatin_interpretation.png	38.6 KB	Sketch
workflow_sketch_hd_chromatin_interpretation.png	298.4 KB	Sketch
getConceptSuggestionsFromTerm.t2flow	40.7 KB	Workflow
listPredefinedConceptSets.t2flow	30.1 KB	Workflow
listPredefinedConceptSets.t2flow	39.9 KB	Workflow
match_two_gene_lists_prioritize_gene_list.t2flow	183.6 KB	Workflow
match_two_gene_lists_prioritize_gene_list.t2flow	446.6 KB	Workflow
explainScoresStringInput2.t2flow	40.6 KB	Workflow
explainScoresStringInput2.t2flow	1.8 MB	Workflow
annotate_genes_biological_processes_xpath_cpids.t2flow	209.6 KB	Workflow
annotate_genes_biological_processes_xpath_cpids.t2flow	216.2 KB	Workflow
Annotate_gene_list_w.wfbundle	50.5 KB	
match_two_gene_lists_prioritize_gene_list-sample.bundle.zip	49.2 MB	

A. Key Resource details

- Name: *workflow_sketch_hd_chromatin_interpretation.png*
  Description: *Sketch of the workflows and their explanation, for data interpretation, plus the connection to the output from the RO for chromatin analysis*
- Name: *workflow_sketch_hd_chromatin_interpretation.png*
  Description: *Sketch of the workflows and their explanation, for data interpretation, plus the connection to the output from the RO for chromatin analysis*
- Name: *getConceptSuggestionsFromTerm.t2flow*
  Description: *This workflow suggests concept ids that match the query term. The user can run this workflow with any term of interest as for example "human", "htt", "Transcription" etc, and will get suggestions for concept ids together with descriptions. Then can choose the concept id that matches the best to her/his needs and use it to the rest of the CPA workflows*
- Name: *listPredefinedConceptSets.t2flow*
  Description: *This workflow lists all IDs and descriptions of the predefined concept set*
- Name: *listPredefinedConceptSets.t2flow*
  Description: *This workflow lists all IDs and descriptions of the predefined concept set*
- Name: *match_two_gene_lists_prioritize_gene_list.t2flow*
  Description: *This workflow can prioritize genes that are related to a specific concept, e.g. HTT. In order to*

---

[1] *http://sandbox.rohub.org/rodl/ROs/data_interpretation-2/*
[2] See RO types definitions at http://w3id.org/ro/earth-science#
[3] http://sandbox.rohub.org/rodl/ROs/data_interpretation-2/hypothesis.txt

obtain the concept id of the term that is going to be matched against the gene list, the workflow Get concept suggestions from term, needs to run first. matchConceptProfileList: the gene list we want to match (order) against a particular concept queryConceptProfileList: the concept (or gene list) we want to match the query against

- Name: *match_two_gene_lists_prioritize_gene_list.t2flow*

  Description: *This workflow can prioritize genes that are related to a specific concept, e.g. HTT. In order to obtain the concept id of the term that is going to be matched against the gene list, the workflow Get concept suggestions from term, needs to run first. matchConceptProfileList: the gene list we want to match (order) against a particular concept queryConceptProfileList: the concept (or gene list) we want to match the query against*
- Name: *explainScoresStringInput2.t2flow*

  Description: *This workflow takes two concept ids as input and returns the top ranking ”B” concepts according to Swanson’s ABC model of discovery, where the relationships AB and BC are known and reported in the literature, and the implicit relationship AC is a putative new discovery. It might also be the case that AC is already known. In that case AC does not represent a new discovery but will still be returned (see workflow example values). The B concepts are returned sorted on the percentage of the contributions of the individual concepts to the coherence score (the average of the inner product scores of all possible concept pairs within the group). This workflow can be used together with other workflows in this pack: http://www.myexperiment.org/packs/282 for functional gene and SNP annotation and knowledge discovery.*

- Name: *explainScoresStringInput2.t2flow*

  Description: *This workflow takes two concept ids as input and returns the top ranking ”B” concepts according to Swanson’s ABC model of discovery, where the relationships AB and BC are known and reported in the literature, and the implicit relationship AC is a putative new discovery. It might also be the case that AC is already known. In that case AC does not represent a new discovery but will still be returned (see workflow example values). The B concepts are returned sorted on the percentage of the contributions of the individual concepts to the coherence score (the average of the inner product scores of all possible concept pairs within the group). This workflow can be used together with other workflows in this pack: http://www.myexperiment.org/packs/282 for functional gene and SNP annotation and knowledge discovery.*

- Name: *annotate_genes_biological_processes_xpath_cpids.t2flow*

  Description: *This workflow annotates a comma separated gene list with a predefined concept set as for example Biological processes or Disease/syndrome. To obtain the particular id for each concept set (e.g. ”5” for Biological processes), the workflow listPredefinedConceptSets needs to run first. The workflow is using the anni web services*

- Name: *annotate_genes_biological_processes_xpath_cpids.t2flow*

  Description: *This workflow annotates a comma separated gene list with a predefined concept set as for example Biological processes or Disease/syndrome. To obtain the particular id for each concept set (e.g. ”5” for Biological processes), the workflow listPredefinedConceptSets needs to run first. The workflow is using the anni web services*

ACKNOWLEDGMENT

The Research Object was uploaded to ROHub by *Eleni Mina*. ROHub portal development was supported by EVEREST EU project (HORIZON 2020 grant 674907).

REFERENCES

[1] The Research Object Management Platform - ROHub http://www.rohub.org/.

[2] K. Belhajjame, O. Corcho, D. Garijo, J. Zhao, P. Missier, D. Newman, R. Palma, S. Bechhofer, E. García Cuesta, J. M. Gómez-Pérez, S. Soiland-Reyes, L. Verdes-Montenegro, D. De Roure, and C. Goble “Workflow-Centric Research Objects: First Class Citizens in Scholarly Discourse”, Proceedings of Workshop on the Semantic Publishing, SePublica Crete, Greece 28 May 2012.

[3] Belhajjame K., Zhao J., Garijo D., Gamble M., Hettne K., Palma R., Mina E., Corcho O., Gómez-Pérez J. M., Bechhofer S., Klyne G., Goble C. “Using a suite of ontologies for preserving workflow-centric research objects”, Journal of Web Semantics: Science, Services and Agents on the World Wide Web Available online 11 February 2015 ISSN 1570-8268.

[4] Palma R., Corcho O., Gómez-Pérez J. M., Mazurek, C. “ROHub - A Digital Library of Research Objects Supporting Scientists Towards Reproducible Science”. In Semantic Publishing Challenge of Proc. Extended Semantic Web Conference (ESWC) Crete, Greece 25-29 May 2014