# Analyzing gene deregulation in Huntington's disease with respect to epigenetic information

*Eleni Mina*

***Abstract*—This RO is comprised by all workflows used for the integration and the analysis of Huntington's Disease (HD) gene expression data and epigenetic datasets in order to establish links between HD and epigenetic regulation in disease. In addition we include all related to our experiment documents, papers and datasets**

## I. INTRODUCTION

This document provides a paper-style view of the Research Object (RO) "*Analyzing gene deregulation in Huntington's disease with respect to epigenetic information*"[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_data_analysis.txt*[3]; Additionally, this RO has been enriched automatically with the following annotations:

- concepts (most frequently mentioned in the RO): *motif, islands, BA, changes, clusters, activity, HD, cell, cell type, enhancers, state, genes, enrichment, disease, kB, chromatin, mRNA*
- domains (fields of knowledge in which the main concepts are commonly used): *anatomy, genetics*
- frequent expressions (most frequently mentioned noun phrases): *unfolded protein response protein, HD brain, frontal cortex, a number of mRNA, channel subunit, cell change, mRNA change, gene deregulation, s disease brain, caudate nucleus, aberrantprotein protein interaction, mRNAs encod ing proton channel subunit, Huntington's disease gene deregulation*
- named entities (most frequently mentioned):
  - Places: *New Hampshire, United States of America*

## II. RESOURCES

The resources encapsulated by the RO are summarized in table I

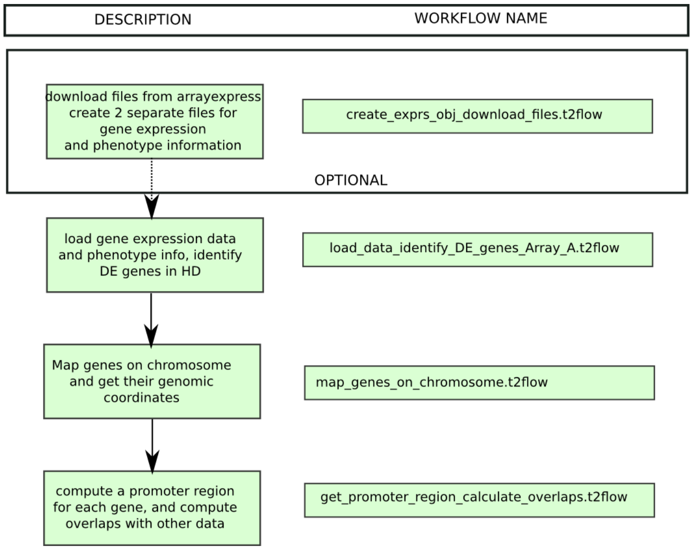

Fig. 1. Research Object Sketch

TABLE I
RESEARCH OBJECT RESOURCES

| name | size | type |
|---|---|---|
| create_exprs_obj_download_files.t2flow | 41.3 KB | Workflow |
| load_data_identify_DE_genes_Array_B.t2flow | 182.0 KB | Workflow |
| load_data_identify_DE_genes_Array_B.t2flow | 182.0 KB | Workflow |
| get_promoter_region_calculate_overlaps.t2flow | 207.3 KB | Workflow |
| load_data_identify_DE_genes_Array_A.t2flow | 213.3 KB | Workflow |
| map_genes_on_chromosome.t2flow | 120.1 KB | Workflow |
| workflow_sketch_hd_chromatin_analysis.png | 123.6 KB | Sketch |
| chr_state_dynamics_in_nine_human_cell_types_nature.pdf | 688.3 KB | Paper |
| CpG_islandsINVertebrateGenomes.pdf | 2.3 MB | Paper |
| Hodges06_human_brain_Affy.pdf | 351.4 KB | Paper |
| cpg_islands_info.txt | 1.7 MB | Dataset |
| cpg_islands_info.txt | 1.7 MB | Dataset |
| broad_hmm_1_Active_Promoter.txt | 1.1 MB | Dataset |
| broad_hmm_1_Active_Promoter.txt | 1.1 MB | Dataset |
| broad_hmm_2_Weak_Promoter.txt | 2.1 MB | Dataset |
| broad_hmm_2_Weak_Promoter.txt | 2.1 MB | Dataset |
| broad_hmm_3_Poised_Promoter.txt | 338.8 KB | Dataset |
| broad_hmm_3_Poised_Promoter.txt | 338.8 KB | Dataset |
| broad_hmm_13_Heterochrom_lo.txt | 7.2 MB | Dataset |
| broad_hmm_13_Heterochrom_lo.txt | 7.2 MB | Dataset |
| hypothesis_data_analysis.txt | 78.0 B | Hypothesis |
| conclusions.txt | 91.0 B | Conclusions |
| conclusions.txt | 91.0 B | Conclusions |
| dummy_results.txt | 558.0 B | Results |
| brain_data_assession_number.txt | 29.0 B | ExampleInput |

---

[1] *http://sandbox.rohub.org/rodl/ROs/HD_chromatin_analysis/*

[2] See RO types definitions at http://w3id.org/ro/earth-science#

[3] http://sandbox.rohub.org/rodl/ROs/HD_chromatin_analysis/hypothesis_data_analysis.txt

### A. Key Resource details

- Name: *create_exprs_obj_download_files.t2flow*
  Description: *download file from NCBI and create expression and phenotype file*
- Name: *load_data_identify_DE_genes_Array_B.t2flow*
  Description: *This workflow gets as input the gene expression file, and the corresponding phenotype dataset in order to compute differentially expressed genes. This specific workflow is for hgu133B array*
- Name: *load_data_identify_DE_genes_Array_B.t2flow*
  Description: *Load expression and phenotype file - get DE genes (array B)*
- Name: *get_promoter_region_calculate_overlaps.t2flow*
  Description: *This workflow uses the output from another workflow in this folder http://sandbox.wf4ever-project.org/rodl/ROs/HD_chromatin_analysis/workflows/map_genes_on_chromosome.t2flow.*
- Name: *load_data_identify_DE_genes_Array_A.t2flow*
  Description: *This workflow gets as input the gene expression file, and the corresponding phenotype dataset in order to compute differentially expressed genes. This specific workflow is for hgu133A array*
- Name: *map_genes_on_chromosome.t2flow*
  Description: *This workflow given a set of genes with their entrez gene identifiers, maps the genes to their corresponding chromosomal location. Uses the R library biomart*
- Name: *workflow_sketch_hd_chromatin_analysis.png*
  Description: *a sketch that describes the relation between the different workflows of the experiment, as well as some brief explanation for each one of them*
- Name: *cpg_islands_info.txt*
  Description: *This file was downloaded by the ucsc genome browser, http://genome.ucsc.edu/*
- Name: *cpg_islands_info.txt*
  Description: *This data was obtained from the ucsc genome browser, hg19 assembly.*
- Name: *broad_hmm_1_Active_Promoter.txt*
  Description: *Active promoter chromatin state for mammary epithelial cells (HMEC)*
- Name: *broad_hmm_1_Active_Promoter.txt*
  Description: *Open chromatin region, associated with promoters, transcriptionally active, defined by the most highly observed chromatin marks : H3K4me2,H3K4me3, H3K27ac, H3K9ac*
- Name: *broad_hmm_2_Weak_Promoter.txt*
  Description: *Weak promoter chromatin state for mammary epithelial cells (HMEC)*
- Name: *broad_hmm_2_Weak_Promoter.txt*
  Description: *Open chromatin region, associated with promoters, weak transcription activity, defined by the most highly observed chromatin marks : H3K4me1, H3K4me2,H3K4me3, H3K9ac*
- Name: *broad_hmm_3_Poised_Promoter.txt*
  Description: *Poised promoter chromatin state for mammary epithelial cells (HMEC)*
- Name: *broad_hmm_3_Poised_Promoter.txt*
  Description: *Open chromatin region, associated with promoters, described as a bivalent domain that has strong signals of both active and inactive histone marks. Most highly observed histone marks: H3K27me3, H3K4me2, H3K4me3.*
- Name: *broad_hmm_13_Heterochrom_lo.txt*
  Description: *Heterocrhomatic chromatin state for mammary epithelial cells (HMEC)*
- Name: *broad_hmm_13_Heterochrom_lo.txt*
  Description: *Heterochromatic state is associated with closed chromatin formation, transcriptionally inactive. It is also associated with none histone marks*
- Name: *hypothesis_data_analysis.txt*
  Description: *this file contains the general hypothesis of the experiment*
- Name: *conclusions.txt*
  Description: *conclusion of the experiment*
- Name: *conclusions.txt*
  Description: *this is a dummy conclusion*
- Name: *dummy_results.txt*
  Description: *this file contains dummy results. The results will be updated after the article publication*
- Name: *brain_data_assession_number.txt*
  Description: *This file contains the NCBI Gene Expression Omnibus entry GSE3790, that was used to download the brain gene expression data. the workflow that downloads the data is : http://sandbox.wf4ever-project.org/rodl/ROs/HD_chromatin_analysis/workflows/create_exprs_*

## 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