

Structured review on Huntington's disease iron hypothesis

Karolis Cremers^{1,*}, Marco Roos¹, Katy Wolstencroft², Eleni Mina¹ and N ria Queralt-Rosinach^{1,*}

¹Leiden University Medical Center, Albinusdreef 2, 2333 ZA Leiden, The Netherlands

²Leiden Institute of Advanced Computer Science, Niels Bohrweg 1, 2333 CA Leiden, The Netherlands

Abstract

Here we present a Structured Review (SR) on the relationships of iron with Huntington's Disease. Including relationship predictions made by different edge prediction models, the results of the inclusion of the Gene Ontology structure on relationship predictions and experimental data representation within the SR.

Keywords

Structured Review, Knowledge graph, Ontologies, FAIR, Network Analysis, Linked Data

Motivation

Structured Reviews (SRs) organize and semantically represent the current knowledge around a research hypothesis in a structured manner, enabling semantic querying and data mining [1].

In this work we present the application of a SR to explore the relationship of iron with Huntington's Disease (HD). HD (OMIM:143100) is a heritable rare neurodegenerative disease caused by an elongated CAG repeat within the huntingtin (HTT, HGNC:4851) gene. The exact mechanisms that lead to disease pathogenesis remain unclear, however one of the current hypotheses implicates the accumulation of iron in HD brain. Abnormal accumulation of iron in the brain has been associated with several other neurodegenerative diseases. Therefore, current therapies often include iron chelators to combat iron build up.

Our SR is a knowledge graph that includes information surrounding the iron hypothesis in HD. We constructed a HD knowledge graph that integrates genes, anatomy, genotypes, variants, physiology and disorders as concepts and their relationships such as "role of" (*RO_0000081*) and "in homology relationship with" (*RO_HOM0000001*). Every instance of concepts and relationships within the SR is annotated with references, similar to a normal review article. These concepts and relationships integrated in the SR are extracted from two curated sources. First, the Monarch Initiative knowledge base is queried for the retrieval of relevant gene and

SWAT4HCLS 2023: The 14th International Conference on Semantic Web Applications and Tools for Health Care and Life Sciences

*Corresponding author.

✉ k.m.p.cremers@lumc.nl (K. Cremers); n.queralt_rosinach@lumc.nl (N. Queralt-Rosinach)

ORCID 0000-0003-0169-8159 (K. Cremers); 0000-0002-8691-772X (M. Roos); 0000-0002-1279-5133 (K. Wolstencroft); 0000-0002-8972-9206 (E. Mina); 0000-0002-1756-3905 (N. Queralt-Rosinach)



© 2023 Copyright for this paper by its authors. Use permitted under Creative Commons License Attribution 4.0 International (CC BY 4.0).

CEUR Workshop Proceedings (CEUR-WS.org)

phenotype information. Second, the TFtargets library is used to obtain transcription regulatory information. In addition to the curated databases, the SR integrates gene expression information from a HD RNA-seq experiment (GSE64810) to include genes that are specifically altered in HD. Finally, we include a set of concepts and relationships of interest pre-selected by domain experts.

We want to place emphasis on the fact that a SR has utility at multiple points of the research cycle: First it can be constructed at the start of the research cycle to describe and explore a specific hypothesis. Second, it can be used during research as a reference for interdisciplinary collaboration. Finally it can be used as a tool to contextualize experimental results at the end of the cycle. In order to encourage the use of the SR throughout the cycle it is hosted on a Wikibase server, where Wikipedia style pages represent node information in the KG. This allows users to review, edit and update the graph with knowledge involving the hypothesis.

To aid the use of computational analysis, including hypothesis generation, and knowledge exploration we load the SR into a NEO4J instance. We use the Graph Data Science Library (GDSDL) to apply relationship prediction algorithms that provide insight on missing information in the SR and potential research hypotheses. In addition to the relationship predictions, we improve the semantic richness of the SR by using NEOsemantics to integrate OWL Ontologies.

Ontologies are used both as independent nodes and as descriptors of concepts and relationships within the SR. These ontologies provide consensus definitions for users, such as data scientists, unfamiliar with the underlying biological concepts. These definitions, combined with the concept-relationship structure of the SR, allow for sharing of contextual information surrounding a hypothesis across disciplinary fields. This is especially important in highly interdisciplinary activities such as disease research.

Here, we will present our ongoing results on the HD iron SR. This will include three parts: First, relationship predictions between disease related genes and iron related concepts made by topology based edge prediction models. Second, the Gene Ontology based improvement of the SR. Third, the graph based representation of the RNA-seq experimental results and their relationship to Iron.

Acknowledgments

The work leading to this poster is supported by grants from European Joint Project on Rare Diseases (EJP RD, COFUND-EJP N° 825575), the collaboration project Trusted World of Corona (TWOC) co-funded by the PPP Allowance made available by Health Holland, Top Sector Life Sciences & Health; to stimulate public-private partnerships and the Leiden Center for Computational Oncology (LCCO): A strategic initiative of the LUMC Oncology Center-Building Individual Digital Tumor-Host Twins For Precision Medicine.

References

- [1] N. Queralt-Rosinach, G. S. Stupp, et al., Structured reviews for data and knowledge-driven research, Database 2020 (2020). doi:10.1093/database/baaa015.