

## **Integrating ontological representation and reasoning into a disease map: application to Alzheimer's disease**

Vincent Henry<sup>1, 2</sup>, Ivan Moszer<sup>2</sup>, Olivier Dameron<sup>3</sup>, Marie-Claude Potier<sup>2</sup>, Martin Hofmann-Apitius<sup>4</sup>, Olivier Colliot<sup>1</sup>,

*Inria, Aramis project-team, F-75013, Paris, France<sup>1</sup>, Institut du Cerveau et de la Moelle épinière, ICM, Inserm U 1127, CNRS UMR 7225, Sorbonne Université, F-75013, Paris, France<sup>2</sup>, Inria, Dyliss project-team, Rennes, France<sup>3</sup>, Fraunhofer Institute for Algorithms and Scientific Computing, SCAI, Sankt Augustin, Germany<sup>4</sup>, AP-HP, Pitié-Salpêtrière Hospital, Departments of Neurology and Neuroradiology, F-75013, Paris, France<sup>5</sup>*

**Background and Objectives:** Systems medicine disease maps are ongoing projects that provide fine curated knowledge on pathophysiology at the molecular and phenotypic level (<http://disease-maps.org/projects>). They are based on a common framework that includes the use of standards and guidelines for defining biological processes and entities. Despite this standardisation effort, there remains some formal and semantics ambiguities: for instance, processes with unexpected participants (e.g. translation of genes instead of transcripts) or phenotypes named with process names (e.g. apoptosis). Ontologies provide a consistent framework to deal with these issues: axiomatic-based definition and logical reasoning properties allow one to underline formal inconsistencies and refine semantic descriptions.

In the present work, we propose to use an ontology in order to increase the consistency of the AlzPathway map [1].

**Approach:** We first built an ontological model based on the definition and axiomatization of high-level classes offered by the CellDesigner diagram [2] including:

- the disjunction of gene, gene product (RNA, protein and complex) and phenotype.
- the definition of processes according to their actions (e.g. transcription is a process that has as input at least one gene and as output at least one RNA).
- the definition of phenotype naming as a state of the system and not as an action.

Then, the knowledge contained in AlzPathway was integrated as subclasses of the previously defined high-level classes. This resulted in 2,429 subclasses.

We then performed automatic reasoning. It allowed us to identify 285 redundancies, 53 inconsistent processes and 14 inconsistent participants. Moreover, the fine exploration of the ontology led to the identification of 55 phenotypes that refer to biological aggregated processes. We manually fixed these inconsistencies and validated the correction with automatic reasoning. Finally, we validated the ability of this modified map to manage biomedical data using multiomics data from a clinical study.

**Conclusions:** In conclusion, we integrated ontological properties into AlzPathway. Our model clears out ambiguities in the gene, gene product, metabolite, phenotype and biological process specifications, and thus, facilitates the integration of multiomics data. Furthermore, our work points out the lack of consensual definition of phenotypes and the need to manage process granularity. Interestingly, the integration of ontological standard into AlzPathway opens perspectives to link AlzPathway with the Alzheimer Disease Ontology [3].

### References

1. Mizuno et al. BMC Syst Biol, 2012, 30; 6:52
2. Kitano et al. Nat biotechnol, 2005, 23; 961-6
3. Malhotra et al. Alzheimers Dement, 2014, 10; 238-46