

The Use of Compartmentalization Databases in the Logical Modelling of the Epithelial-Mesenchymal Transition

Márk Kerestély¹, Péter Mendik¹, Sebestyén Kamp², Dávid Deritei^{1,3}, Nina Kunšič^{1,4}, István Narozsny¹, Gábor Máthé¹, Péter Csermely¹, Dániel V. Veres^{1,2}

1: Department of Molecular Biology, Institute of Biochemistry and Molecular Biology, Semmelweis University, Budapest, Hungary

2: Turbine Ltd, Budapest, Hungary.

3: Channing Division of Network Medicine, Brigham and Women's Hospital, Harvard Medical School, Boston, United States of America

4: NIJZ: National Institute of Public Health, Ljubljana, Slovenia

Introduction: Intracellular signalling can be computationally modelled on a compartment-level by using compartmentalization databases such as the ComPPI (COMpartmentalized Protein-Protein Interaction) and Translocatome databases. Translocatome contains manually curated and machine learning predicted data on protein-translocation. One relevant use for these databases is the modelling of the epithelial-mesenchymal transition (EMT), which has a central role in the cancer metastasis process and is substantially regulated by protein-translocations.

Aims: Our aim was to utilize ComPPI and Translocatome in the logical modelling of EMT. We also aimed to identify areas of improvement for their use in logical modelling.

Method: We used a Gene Ontology based enrichment analysis to ascertain the relevance of protein-translocation in the signalling of EMT. We conducted a literature review of translocating EMT proteins based on the predictions of Translocatome. Using information on compartmentalized interactions from ComPPI and the literature, we incorporated these protein-translocations into a previously established Boolean model of TGF- β -mediated EMT. We built our model in the BooleanNet simulation framework.

Results: We found that, in Translocatome, predicted translocating proteins were enriched among EMT proteins ($p < 0.0001$), while predicted non-translocating proteins were de-enriched among EMT proteins ($p < 0.0001$). We manually curated 21 translocating proteins relevant to the previously established EMT model, which had 19 nodes and 70 edges. After incorporating the protein-translocations, the updated model reached 31 nodes and 101 edges. 3 out of 5 instances of differing behaviour in the dynamic simulations between the original and the updated model involved an incorporated protein-translocation.

Conclusion: Our work serves as a proof-of-concept for compartmentalized logical modelling of complex intracellular signalling processes utilizing compartmentalization databases. In the future, the adoption of a data format suitable for logical modelling in the databases could enable the algorithmic inference of compartmentalized logical models.

Funding: Supported by the ÚNKP-23-2-I-SE-59 New National Excellence Program of the Ministry for Culture and Innovation from the source of the National Research, Development and Innovation Fund.