

Translocatome

a novel tool for drug target discovery, based on the systematic analysis of protein translocation

Levente Dobronyi^{(1)*}, Péter Mendik^{(1)*}, Péter Csérmely⁽¹⁾, Dániel V. Veres^{(1)(2)*}

⁽¹⁾ Semmelweis University, Department of Medical Chemistry, Budapest, Hungary

⁽²⁾ Turbine Ltd., Budapest, Hungary

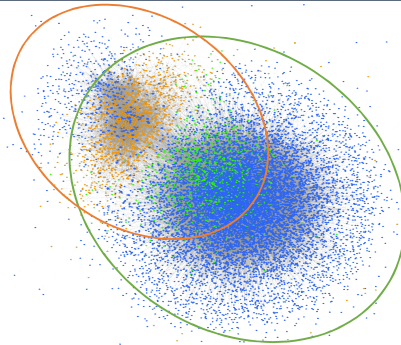
*veres.daniel1@med.semmelweis-univ.hu



What is Translocatome?

Translocatome is a **comprehensive database of translocating proteins and their interactions**, enabling the systematic analysis of their role in cellular signaling. **Predicted** translocating proteins filtered from the ComPPI database (compbi.linkgroup.hu), with an extensive **manually curated** part makes the Translocatome reliable for modelling. After performing network analysis with our in-house developed **AI guided network dynamical simulation tool**, called Turbine (turbine.ai), we are able to get new insight of the pathological role of protein translocations resulting in the prediction of new drug targets with a potential to help patients fighting cancer.

Compartmentalized interactome of the cytoplasm and the mitochondrion



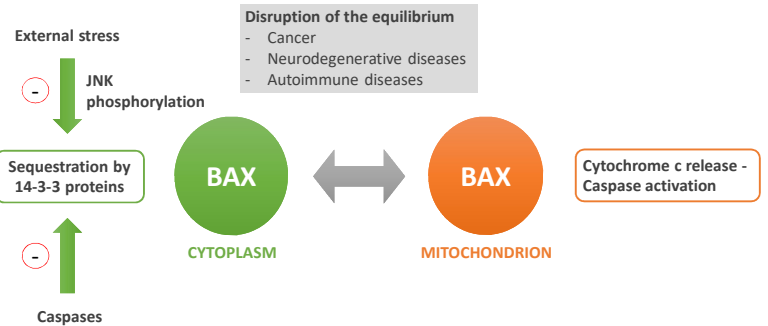
Proteins having both cytoplasmic and mitochondrial localizations act as „bridges” between the two compartments (**orange** and **green** nodes). The two separated clusters are representing the two different compartments, having their own protein-protein interaction network.

The **orange** nodes are representing the **mitochondrial** localization of these potentially translocating proteins, whereas the **green** nodes are their **cytoplasmic** ones. **Blue** nodes have only one, either mitochondrial or cytoplasmic localization.

Pathological role of translocation aberrations

As shown on our example, translocation can **mediate** several **biological processes**. These effects can be explained by looking into the **organelle specific interactome**, showing that translocating proteins are making the **communication between the organelles** possible. Thus, the alteration of these proteins (overexpression, mutation resulting in altered localization) can **disrupt this communication**, potentially causing **cellular dysfunction**, eg.: *neoplastic transformation*.

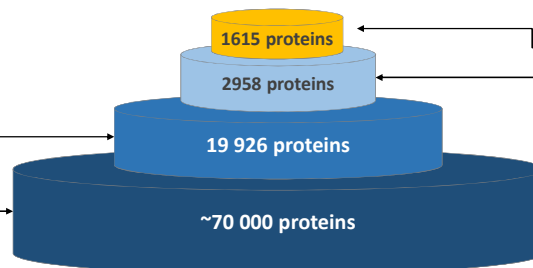
One of the most important translocating proteins is **Bax**, a member of the Bcl-2 family of proteins, and has a key role in *stress-induced apoptosis* by its **translocation to the mitochondria** causing Cytochrome c release⁽¹⁾.



Method of predicting translocating proteins

To avoid the use of unspecific data we filtered the data for human **SwissProt** entries. Number of more reliable proteins is **19 926** now.

The protein set is based on **ComPPI Database**⁽²⁾, which contains localization data of approximately **70 000** human proteins.

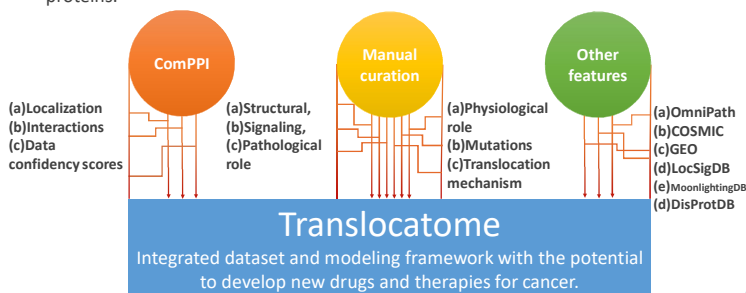


To exclude the noise from the dataset we focused on the „**nucleus-cytoplasm-extracellular space**” axis which contains 9 routes (see later) for translocation. We also excluded the proteins which are found in almost all compartments, to exclude the housekeeping proteins.

There are **2958** proteins with at least **two different experimentally validated** localization. This set also contains the secretory and housekeeping proteins, that need to be filtered.

Structure of the database

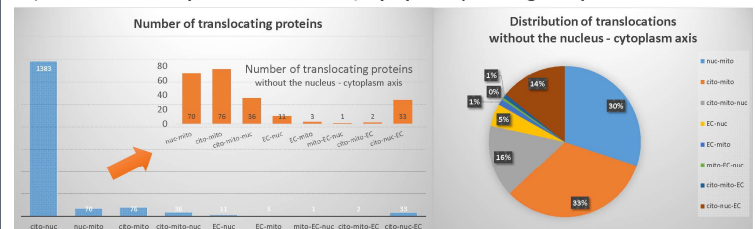
- (I) Set of **predictably translocating proteins** from ComPPI. Translocatome will be able to communicate with ComPPI, and the new versions of the ComPPI database will update the number and information of predictably translocating proteins.
- (II) **Heavily manually curated data** with data of the translocation localizations, mechanism, interactions in each compartment, biological role, signaling, structural data of the protein, pathological role, and druggability. There are also **lightly manually curated proteins** without detailed information, altogether **52** proteins. Our aim is to manually curate at least **10%** of the database.
- (III) Integration with **metadata**. Signaling (*OmniPath*), genomic (*COSMIC*) and transcriptomic (*GEO*) data to enable performing dynamic simulations on cancer cell models. Localization (*LocSigDB*), functional (*MoonlightingProteinsDB*), internal disorganization (*DisProtDB*) data to widen the understanding of translocating proteins.



Routes of translocation

There are 9 potential translocation routes, from which the following are the main four:

- 1.) **Cytoplasm – Nucleus (86%)** takes the biggest part, these proteins include transcription factors
- 2.) **Cytoplasm – Mitochondrion (5%)**
- 3.) **Nucleus – Mitochondrion (4%)**
- 4.) **Extracellular compartment – Nucleus / Cytoplasm (3% altogether)**



Future plans and applications

- **Database improvement:** continue manual curation and apply node attributes for better data handling:
 - (I) **data complexity score:** complexity of manually curated data and metadata
 - (II) **evidence score:** level of data reliability - predicted or experimentally validated
- **Modelling:** developing cancer model framework based on translocations and AI guided dynamic modelling of pathological states – prediction of potential drug targets
- **Online interface:** open access and user-friendly web application
- **Experimental validation**