

ARTICLE OPEN



Translocating proteins compartment-specifically alter the fate of epithelial-mesenchymal transition in a compartmentalized Boolean network model

Péter Mendik^{1,3}, Márk Kerestély^{1,3}, Sebestyén Kamp², Dávid Deritei¹, Nina Kunšič¹, Zsolt Vassy¹, Péter Csermely¹ and Daniel V. Veres^{1,2}✉

Regulation of translocating proteins is crucial in defining cellular behaviour. Epithelial-mesenchymal transition (EMT) is important in cellular processes, such as cancer progression. Several orchestrators of EMT, such as key transcription factors, are known to translocate. We show that translocating proteins become enriched in EMT-signalling. To simulate the compartment-specific functions of translocating proteins we created a compartmentalized Boolean network model. This model successfully reproduced known biological traits of EMT and as a novel feature it also captured organelle-specific functions of proteins. Our results predicted that glycogen synthase kinase-3 beta (GSK3B) compartment-specifically alters the fate of EMT, amongst others the activation of nuclear GSK3B halts transforming growth factor beta-1 (TGFB) induced EMT. Moreover, our results recapitulated that the nuclear activation of glioma associated oncogene transcription factors (GLI) is needed to achieve a complete EMT. Compartmentalized network models will be useful to uncover novel control mechanisms of biological processes. Our algorithmic procedures can be automatically rerun on the <https://translocaboole.linkgroup.hu> website, which provides a framework for similar future studies.

npj Systems Biology and Applications (2022)8:19; <https://doi.org/10.1038/s41540-022-00228-7>

INTRODUCTION

The spatial and temporal organization of intracellular proteins is well regulated and it often defines the cellular fate¹. Proteins must function specifically at given times with certain functions, thus the eukaryotic cell is organized into different compartments, called organelles². These subcellular organelles must provide a specific chemical environment and a specific set of interaction partners, so that proteins can function physiologically³. The physiological distribution of proteins is essential for the life of a cell and it plays an inevitable role in biological processes such as cellular learning⁴.

Protein translocation refers to the regulated change of a protein's subcellular localization between subcellular compartments⁵. This results in altered interactors and functions of the translocated protein. Disturbances in the regulation of protein translocation lead to altered cellular localization of proteins which may contribute to cellular dysfunction and different pathologies^{6,7}. Pathological alterations in the localization of proteins also function as biomarkers for disease progression⁸. Moreover, pharmaceutical targeting of protein translocation is a promising therapeutic strategy⁹.

We have recently assembled the Translocatome database, which is a comprehensive dataset of human translocating proteins⁵. It contains 13,066 human proteins, each characterized with a translocation probability. 34% of the proteins in the Translocatome database were predicted as translocating proteins. This ratio is supported by the experimental results of Thul et al., which shows that approximately 50% of the human proteome is multi-localized¹⁰. These large datasets provide available data that can be utilized to create in silico models of cellular processes designed in a way that they can represent subcellular dynamics.

Epithelial-mesenchymal transition (EMT) is a biological process which physiologically accompanies early embryonic development

but may also be activated pathologically during cancer progression or tissue fibrosis. The reversion of EMT is called mesenchymal-epithelial transition (MET). Thus there is a dynamic transition between the epithelial and mesenchymal phenotypes, and cells may reside in a spectrum of intermediary phases, so-called hybrid states¹¹. EMT is triggered through signals that cells receive from their environment, like the TGFB signal which is a potent driver of EMT¹². As EMT is a diverse process, research that focuses on EMT must take this into consideration and must understand EMT on a more diverse palette. Monitoring EMT-induced changes on different levels, such as those of gene expression, phenotypical (microscopically observable) changes and functional disturbances, has become increasingly important¹³.

EMT is a spatial process, where several key factors change their localization. The most obvious examples of these EMT-induced localization changes include EMT-related transcription factors, which translocate from the cytoplasm to the nucleus to exert their transcriptional activity¹⁴. One specific regulatory protein of EMT is beta-catenin (CTNNB), which translocates between the plasma membrane and the nucleus¹⁵. Although the role of protein translocations during EMT is evident, yet there is no work that evaluates how the compartmentalization of a network model may create better model predictions and how to systematically create models that incorporate data on subcellular dynamics.

Representing intracellular signalling pathways as signalling networks has become a standard for the evaluation of cellular signalling processes in the last decades^{16–18}. These networks can be easily translated to Boolean models and these models offer a quick and effective option to examine dynamic biological processes. Boolean models define a node's state as active (TRUE/ON/one) or inactive (FALSE/OFF/zero) and convert biological rules into Boolean rules using the logical operators AND, OR

¹Department of Molecular Biology, Institute of Biochemistry and Molecular Biology, Semmelweis University, Budapest, Hungary. ²Turbine Ltd, Budapest, Hungary. ³These authors contributed equally: Péter Mendik, Márk Kerestély. ✉email: veres.daniel1@med.semmelweis-univ.hu

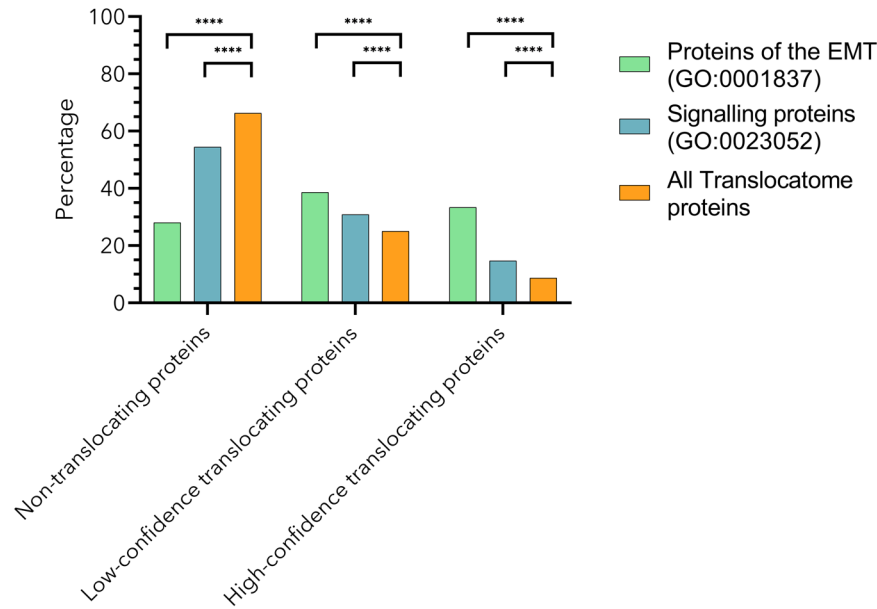


Fig. 1 Translocating proteins are enriched between signalling and EMT proteins. On this figure, we can see the distribution of proteins according to their translocation probability in different datasets (Translocatome proteins, Signalling proteins and EMT proteins, see Methods for details). There were significantly more high-confidence translocating proteins between the Signalling (15%) and EMT (33%) proteins than between the Translocatome proteins (9%). We can observe the same significant difference in the case of low-confidence translocating proteins. The percentage of low-confidence translocating proteins is higher between Signalling (31%) and EMT proteins (39%) than between Translocatome proteins (25%). Oppositely in the case of non-translocating proteins there is a lower percentage of non-translocating proteins between signalling (54%) and EMT (28%) proteins than between Translocatome proteins (66%). For further details about the calculation of protein enrichment refer to the Methods section, and to understand the details of the Translocatome data please refer to its original publication⁵. **** $p < 0.0001$, Chi-square test.

and NOT. These models are simplified but are still very useful to qualitatively evaluate dynamic biological processes¹⁹.

EMT was previously characterized by several computational models²⁰, but these often focus on very specific aspects of the transition²¹ or just on some core regulatory proteins²². An extensive attempt that incorporated a large number of proteins and characterized EMT on a system level was published by ref. ²³. That model revealed the joint Sonic Hedgehog and Wnt pathway activation during TGF β -mediated EMT. This model consisted of 70 nodes and 135 edges, but using an attractor-preserving network simplification method it was reduced to a 19 node core network. In the current work, we present how the addition of compartment-specific information enriches the usefulness of a Boolean network model. For this purpose, we used the 19-node EMT model published in the paper of Steinway et al. as a benchmark and we compared our compartment-extended model to it. We refer to this 19-node EMT model of ref. ²³ as the “original EMT model”.

Our work systematically incorporates protein translocation into a Boolean dynamic signalling model, thus it catalyses the understanding of protein translocation as a general and important aspect of cellular regulation. In the future, similar works will be useful to understand how protein translocations govern different cellular mechanisms and this will increase the understanding of how subcellular dynamics affect cellular behaviours. Extended models that use this compartmentalized approach will be useful to uncover novel biomarkers and to create more complex therapeutic targeting strategies.

In this work, we show the enrichment of translocating proteins in EMT and create a compartmentalized dynamic network model of EMT. Utilizing the potential of our compartmentalized model, we successfully prove that the addition of translocating proteins can result in a more concise decision-making system, by reducing the number of ambiguous attractors. More importantly, our model enables the analysis of compartment-specific protein functions, thus we were able to predict and validate (based on published

experimental data) that both GSK3B and GLI compartment-specifically influence EMT.

RESULTS

Enrichment of translocating proteins between signalling proteins

The Translocatome database contains the translocation probability (stored as Translocation Evidence Score) of approximately 13 thousand human proteins⁵. Based on the literature, translocating proteins are playing a pivotal role in cellular signalling and also in the EMT^{14,24}. To prove this, we assessed the enrichment of translocating proteins between human signalling and EMT proteins. There are 13,066 human proteins in the Translocatome database, out of which 66% are non-translocating proteins, 25% are low-confidence and 9% are high-confidence translocating proteins. We assessed if the same distribution is observed between human signalling and EMT proteins. Signalling and EMT proteins were defined using specific Gene Ontology (GO)²⁵ terms (see Methods for details).

We show that the observed distribution of translocating proteins differs in the case of signalling and EMT proteins, from what we would expect based on the Translocatome database (Fig. 1). Between signalling proteins, we found a higher percentage of translocating proteins (31% are low-confidence and 15% are high-confidence translocating proteins) and fewer non-translocating proteins (54%). Similarly, in the case of the proteins in connection with EMT, we also found a higher percentage of translocating proteins between the EMT proteins (39% are low-confidence and 33% are high-confidence translocating proteins) and the number of non-translocating proteins decreased (28%).

These observed discrepancies are significant ($p < 0.0001$), so we can conclude that translocating proteins become enriched in human signalling processes and in EMT (see details of this enrichment analysis in the Methods section). This observation also underlines

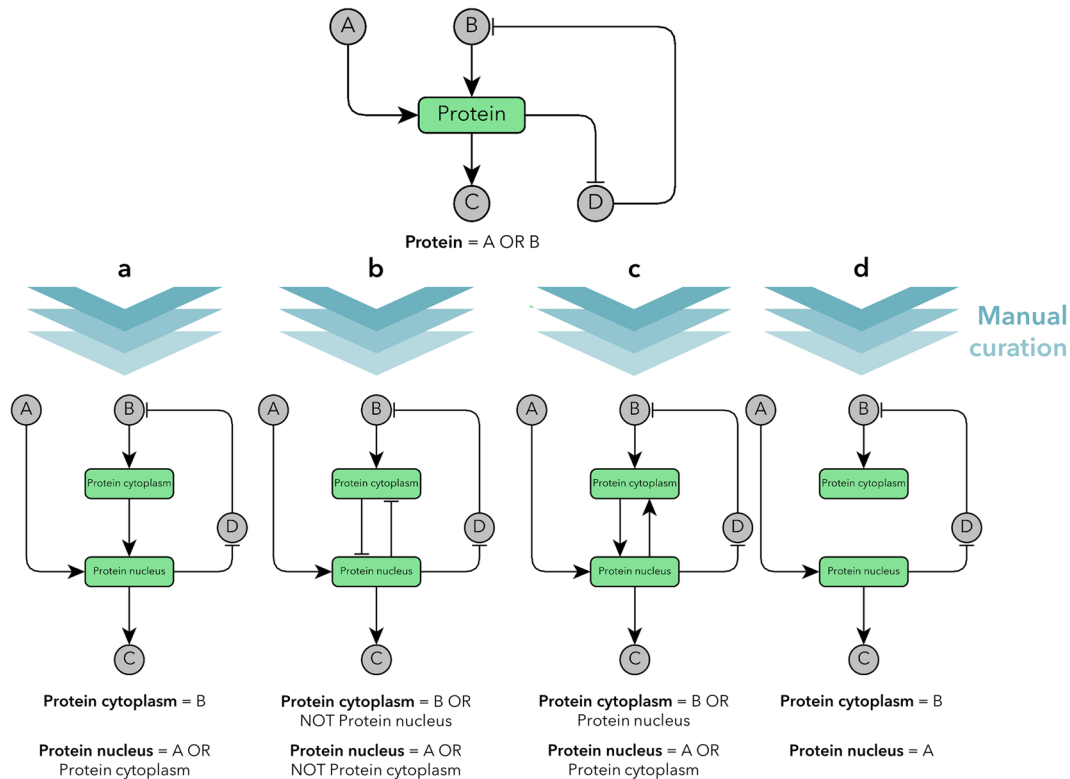


Fig. 2 Creation of compartmentalized Boolean rules. Conventionally in the network representation of signalling processes nodes are individual proteins and edges (either inhibitory or activatory) are interactions between them, but their localization specificity is not considered. We take into consideration the compartment-specific functions of proteins so we can systematically add this information to Boolean models. After the careful revision of available literature data (manual curation) we created a compartmentalized Boolean network. In this hypothetical explanatory example we highlighted a “translocating protein” with green. **a** One possibility is that after e.g., phosphorylation by B the protein translocates to the nucleus and this does not directly affect the cytoplasmic pool of that protein. **b** Another possibility was already contained in the original EMT model (in the case of CTNNB), where the nodes mutually inhibit each other. **c** Some transcription factors can upregulate their own expression which can act as a positive feedback, **d** but it is also possible that both the cytoplasmic and the nuclear pools of a protein have their own regulatory interactions but they don't affect each other directly. There are also other potential combinatorial possibilities (see Supplementary Fig. 7). During compartmentalization we always followed the logic of the available experimental data and tried to interpret into Boolean rules. As the nodes represent the functionality not the mass of proteins the simultaneous activation of both nodes does not negate the law of mass conservation (more on this topic in Supplementary Note 5).

the value of creating a compartmentalized model, enabling the possibility to represent protein translocation in a dynamic model.

Creation of a compartmentalized EMT network

After demonstrating the enrichment of translocating proteins in EMT we became interested in creating a network model of EMT where we can assess protein translocations from a network biology perspective. To model protein translocation, we created a compartmentalized network where we can represent each protein according to its subcellular localization. In this model, if a protein can translocate between 2 subcellular localizations, then it is divided into 2 nodes and each node has only the edges corresponding to that localization (Fig. 2). We call this process the compartmentalization of a network because during this process each protein's regulation will be represented according to their subcellular locations. Since there is no universal solution to split one node into two compartmentalized nodes and there are several possible combinatorial scenarios it is important to capture the functionality of proteins. We always implemented the Boolean rules that best capture the biological behaviour of certain proteins, with a special focus on functionality and less on a universal methodology. This compartmentalization is crucial as many translocating proteins can have very different roles in different subcellular locations, e.g., there are translocating moonlighting proteins^{26,27}.

In this work, our aim was to show the applicability and usefulness of the addition of compartment-specific information to a Boolean network model. To prove this, we used the 19-node EMT model published by ref.²³ as a benchmark and we compared our model to it. As mentioned earlier we refer to this model as the “original EMT model”.

The original EMT model had 19 nodes and 70 edges. The Translocatome database provided sufficient data to examine the translocation probability of these nodes. Out of the 19 nodes 14 nodes were predicted as high-confidence translocating proteins. After this initial signal that there are several important proteins that translocate during EMT, we carried out a thorough manual curation (see details in the Methods section), to validate these translocations, and to see if these translocations could be observed during EMT (summary of the manually curated information in Supplementary Data 1). If we could both validate the translocation itself, and its involvement in EMT, then we could create compartmentalized Boolean logic for that given protein.

In the compartmentalization phase, we created Boolean rules based on the literature, which was collected by manual curation. During this process, we reviewed 64 publications and created 10 compartmentalized nodes (see the list of these nodes and their Boolean rules in Supplementary Data 2) and thus by node-duplications extended the 19-node network into a 30 node-network (the NOTCH node was divided into 3 subcellular nodes, since it has validated localization and activity in the plasma

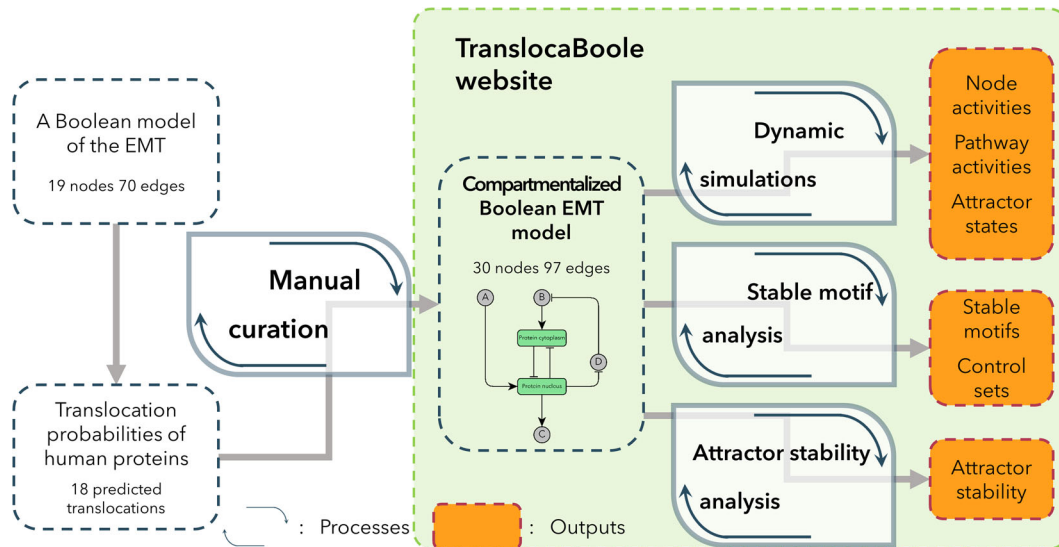


Fig. 3 Workflow for the creation and evaluation of a compartmentalized Boolean model. Starting from previously published Boolean models of the EMT with the help of the Translocatome database⁵ it is possible to uncover translocating nodes that need to be compartmentalized. The careful revision of available literature i.e., manual curation, through the creation of new Boolean functions (based on the collected experimental data) enables the creation of a compartmentalized model. This is followed by dynamic simulation which could be analysed as different outputs on different levels, such as node or pathway activities or the identification of stable attractors of a system. Other analyses can identify the stable motifs and control sets of a model or calculate attractor stability measures. The website (<https://translocaboole.linkgroup.hu/>) enables users to modify our model and to automatically run simulations with our standard settings (25 iterations and 20,000 steps, every node is perturbed) and download the results.

membrane, in the cytosol and in the nucleus as well). After this, we had a model which was ready for further dynamic network analysis to understand the changes caused by the compartmentalization and to validate the model with the comparison to experimental results published earlier or after our predictions. The final model is defined in Supplementary Data 3 and we will refer to the node names defined in this Supplementary Data 3.

This work can be generalized to a workflow shown in Fig. 3, where first the predicted translocations are identified in a previously published dynamic signalling network model, then these are manually validated and a compartmentalized network model is created according to the considerations shown in Fig. 2. Finally, computational simulations can be implemented on the compartmentalized network model and different outputs can be evaluated. The <https://translocaboole.linkgroup.hu> website provides a surface to edit compartmentalized Boolean rules and to rerun computational analysis (the simulation runs with the default settings of 25 iterations and 20,000 steps, every node is perturbed). Starting from the Boolean rules of our model users are able to change them. Once this is done, the server supporting the website automatically generates the required inputs needed for the dynamic simulations in the background, and users can download the final output files of these analyses (Fig. 3).

Users who need more features and modifications can also opt to download the input files and with those files, they can use the relevant codes available in our GitHub repository (https://github.com/deriteidavid/compartmentalized_EMT_Boolean_model_Mendik_et_al_2021) to rerun the analyses (for details see Supplementary Fig. 1). The website and GitHub tools make it possible to also use our workflow as a framework for similar future studies.

Characteristics and dynamical repertoire of the compartmentalized EMT model

In the previous steps of our work, we created a compartmentalized signalling network that could easily be translated to a

dynamic model based on the work of ref. ²³. The resulting model is a discrete dynamic Boolean model that characterizes every node by a state variable (ON or OFF) corresponding to the node's activity and a Boolean function that represents the node's regulation. During the dynamic simulations, an asynchronous updating algorithm randomly selects a node and updates its state according to its Boolean function (more details in Methods). The arguments of the Boolean functions represent the transcriptional, post-transcriptional, translational, post-translational or localizational regulation of a certain node. The main predictive power of Boolean models of biological systems lies in their long-term convergent behaviour into stable states, which qualitatively correspond to real biological phenotypes^{28,29}. The attractors of a Boolean model are such terminal stable states. The predictive value of a Boolean model of a biological system can be assessed by evaluating its attractors from a biological viewpoint. Moreover, it is also crucial to understand how the attractor states emerge from the dynamical landscape of the model. This understanding can enhance further validation of the model and give insight into control and finding potential therapeutic targets^{30,31}.

Using a state-of-the-art method³² we mapped out the dynamical repertoire of our new model in the form of a succession of stable motif stabilizations. Stable motifs are generalized positive feedback loops (also known as maximal trap spaces) which represent sets of self-sustaining node-states, which once activated permanently lock into those node-states³³. Different successive combinations of stable motifs determine the steady state attractors of the system. The new compartmentalized EMT model has seven attractors summarized in Supplementary Table 1. The two attractors most relevant in our case are the ones corresponding to the epithelial (E) and mesenchymal (M) states (Fig. 4a). The E and M attractors are polar opposite states, differing in every node state (with the exception of SOS_GRB2 which is 0 in both attractors). The remaining 5 attractors are intermediary states, which we call hybrid attractors. Hybrid states can emerge during incomplete or partial EMT processes²⁴. The emergence of the hybrid attractors of our model can be explained with three specific

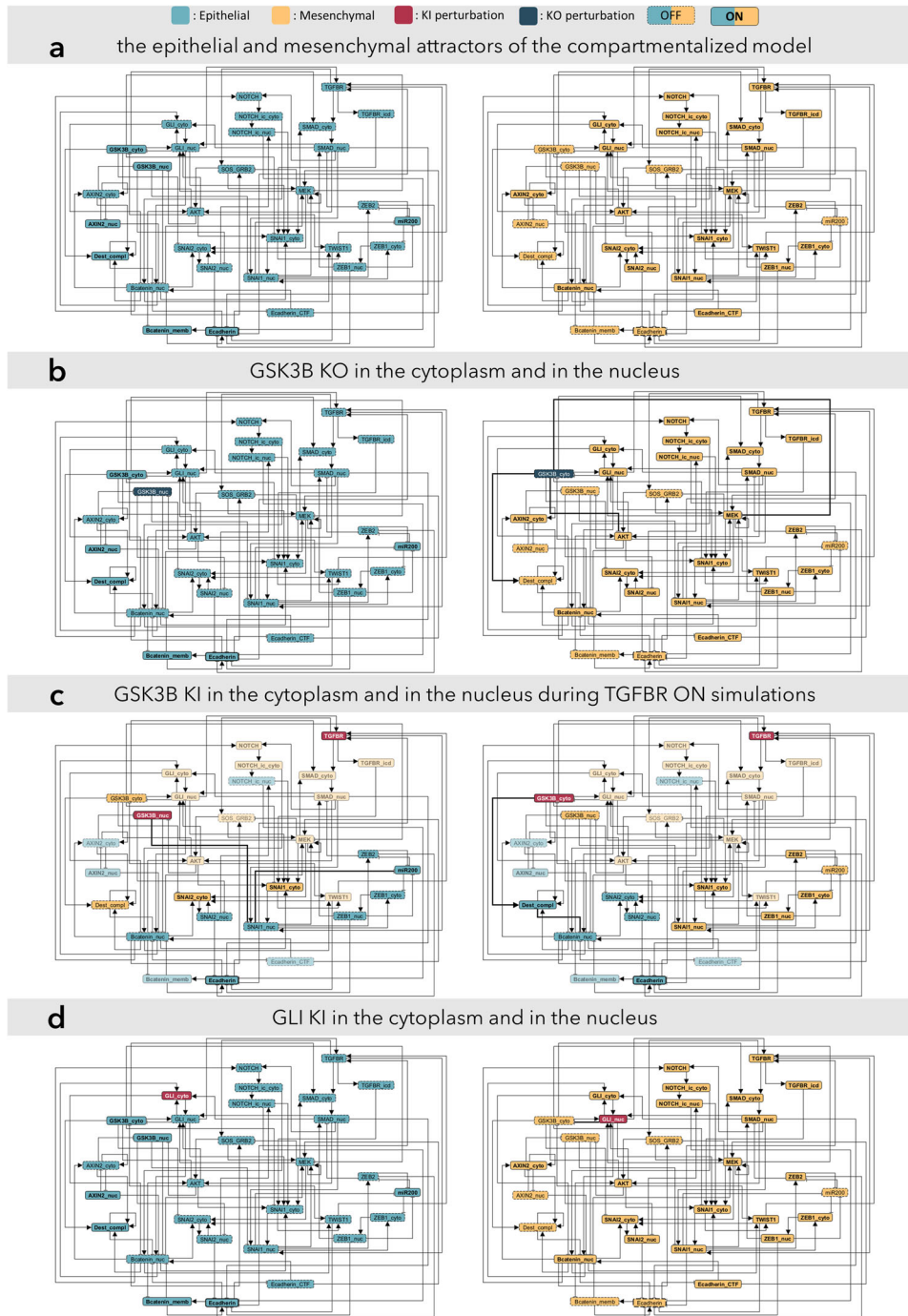


Fig. 4 Attractors of the compartmentalized model showing localization-specific functions of GSK3B and GLI. **a** The attractors corresponding to the epithelial and mesenchymal state can be seen on the left and right respectively. During the dynamic simulations, perturbations were introduced to the initial epithelial state. **b** In case of the TGFBR OFF simulations the cytoplasmic KO of GSK3B resulted in EMT, while the nuclear perturbation had no effect. The signal from GSK3B_{cyto} KO propagated through the loss of the Dest_{compl} and the activation of AKT and MEK. **c** During TGFBR ON simulations both the KI of the cytoplasmic and nuclear node of GSK3B inhibited the TGFBR mediated EMT. The nuclear perturbation had a greater inhibitory effect by preventing the loss of miR200 and consequently the activation of ZEB1 and ZEB2. The nuclear perturbation also led to the inhibition of SNAI2_{nuc} and Bcatenin_{nuc}, despite the loss of the Dest_{compl} and the activation of SNAI2_{cyto} compared to the cytoplasmic perturbation. GSK3B_{cyto} functions both to stabilize the epithelial state and to inhibit TGFBR mediated EMT, whereas GSK3B_{nuc} functions specifically to inhibit TGFBR mediated EMT. **d** TGFBR OFF simulations show that the nuclear perturbation of GLI led to EMT, while the cytoplasmic perturbation alone was insufficient to destabilize the epithelial state. As part of a transcription factor family, the GLI proteins exert their main function in the nucleus and if sequestered in the cytoplasm, GLI2 gets truncated into a repressor form that further decreases transcriptional activity. The compartmentalized model reproduced these localization-specific functions.

stable motifs, made up of only 5 nodes in total, as discussed in Supplementary Note 1 (see also Supplementary Fig. 2 and Supplementary Table 2).

Furthermore, based on the method of refs.^{32,34} we compared the control sets of our model with the original EMT model. A control set of an attractor is a given group of nodes, which, when forced to the corresponding state, drive the whole system into a specific attractor³⁵. As the data in Supplementary Data 4 show the simplest control set of the epithelial state is a smaller set of nodes (5 nodes, see Supplementary Data 4, Epithelial control sets of the Compartmentalized model No. 1) compared to the original EMT model (6 nodes, see Supplementary Data 4, Epithelial control sets of the Original model No. 1), thus our model reveals that forcing the network into the epithelial state is more feasible if we use compartment-specific perturbations.

We also found that in the presence of noise (introduced to our model as update errors) the mesenchymal attractor is by far the most stable and thus has the highest stationary probability from all states (for details see Supplementary Note 2). This results from the fact that this original EMT model focuses on the transition from epithelial to mesenchymal state and thus nodes and interactions driving towards the mesenchymal states are overrepresented.

It is also worth mentioning that although the overall stability profile of attractors still overwhelmingly favours the mesenchymal state, the balance is somewhat improved towards epithelial state as compared to the original EMT model of ref.²³ (see Supplementary Table 3).

Comparison of the original and compartmentalized model

As mentioned above our simulations showed that the compartmentalized model has both a stable epithelial and mesenchymal state. Our compartmentalized model was validated by showing that it could reproduce key experimental outcomes (detailed in Supplementary Table 4) that had been used for the validation of the original EMT model.

For additional comparison and analysis of the two models, we used two simulation setups: one we call single node TGFBR node OFF (simply referred to as TGFBR ON or OFF), the other TGFBR ON simulations. In both setups the network started from the epithelial state and was perturbed by permanently setting a node's state to ON or OFF, thus overriding its Boolean rule. These perturbations were similar to wet-lab knock-out (KO) or knock-in (KI) experiments, but importantly, our perturbations were always compartment-specific. This resulted in scenarios when a protein's functionality was maintained in one subcellular compartment and inhibited in another. During the TGFBR ON simulations the epithelial initial state was altered by setting the TGFBR node's state to ON and an additional KI/KO perturbation was also introduced to the network. This represented the biological scenario of an active TGFBR signal which is a potent driver of EMT. In the latter case, we observed if a node's perturbation can block TGFBR driven EMT.

Comparison of TGFBR OFF simulations showed differences between the original and our model in two cases: NOTCH or MEK KI led to EMT in the original but not in the compartmentalized model. In the case of NOTCH activation, our model better simulated the experimental results showing that Neurogenic locus notch homolog protein 1 (NOTCH) activation alone without the induction of TGFBR was not sufficient to induce EMT^{36,37} (Supplementary Figs. 3 and 4). This result stemmed from the fact that in our model the activation of Zinc finger protein SNAIL1 (SNAIL1) is captured more complexly, and NOTCH alone could not activate this key transcription factor of EMT.

In the case of MEK activation, there was a more complex situation. The activation of the MEK node did not lead to a mesenchymal phenotype in the compartmentalized model,

whereas it did in the original EMT network (Supplementary Figs. 3 and 4). There is available experimental validation for both outcomes, as experiments have shown that MEK/ERK pathway activation alone did not, while a combined TGFBR and MEK/ERK pathway activation did lead to a perfect EMT phenotype in a lung adenocarcinoma cell line³⁸. However, in intestinal epithelial cells overexpression of a constitutively active form of MEK1 activation was sufficient to induce EMT³⁹. Moreover, in the case of TGFBR ON simulations only the compartmentalized model showed correctly that MEK inhibition could prevent TGFBR induced EMT⁴⁰ (Supplementary Fig. 5).

Compartmentalized functions of GSK3B and GLI

The division of translocating nodes by compartments allows for the exploration of compartment-specific functions, providing the major advantage of our method. In TGFBR OFF simulations GSK3B KO in the original EMT model led to a mesenchymal phenotype, which matched the outcome of GSK3B_cyto KO in the compartmentalized model. Experimental evidence underlines this outcome as the inhibition of cytoplasmic GSK3B by LiCl led to EMT in ovarian adenocarcinoma⁴¹. The compartmentalized model uncovered that GSK3B_nuc KO did not lead to EMT (Fig. 4b). Because GSK3B localizes mainly to the cytoplasm^{41,42} and translocates to the nucleus during EMT⁴³, it is expected that in the epithelial state GSK3B_nuc KO does not have an effect.

Furthermore, in the case of TGFBR ON simulations, we could predict other compartment-specific functions of the GSK3B kinase. Our results showed that the KI of GSK3B both in the cytoplasm (GSK3B_cyto) and in the nucleus (GSK3B_nuc) acted as a repressor of EMT, but this inhibitory effect was stronger in the nucleus (Fig. 4c). In accordance with that, experimental results proved that the inhibition of the PI3K/AKT pathway suppressed EMT through the induction of GSK3B in hepatocellular carcinoma (HCC)⁴³. The more robust inhibition in the nucleus is a consequence of the fact, that nuclear translocation of GSK3B prevents EMT through the downregulation of SNAIL transcription factor⁴³ and that nuclear GSK3B is highly active relative to its cytoplasmic counterpart⁴⁴.

Another node that showed compartment-specific functions is the GLI node. GLI KI resulted in EMT in the original model just as GLI_nuc KI led to EMT in the compartmentalized model. Importantly, it has been reported in the literature that independent GLI activation can induce EMT⁴⁵. Additionally, the compartmentalized model shows that GLI_cyto KI did not result in EMT (Fig. 4d). This is due to the fact that in the absence of an upstream signal cytoplasmic GLI2 gets truncated into a transcriptional repressor form which inhibits GLI-induced gene transcription⁴⁶ and there is a simultaneous cytoplasmic sequestration of GLI1⁴⁷. Overall, we have shown that our compartmentalized model compared to the original could provide justifiable results on the subcellular regulation and compartmentalized functions of different kinds of proteins like the GSK3B kinase and the transcription factor family GLI.

Signalling pathway analysis uncovers pathway crosstalk

To understand how different signalling pathways activate during the EMT we created a method to assess not only node level but signalling pathway activities as well (see the details of this in the methods section). We monitored the activity of the most significant signalling pathways that lead EMT¹⁴. These pathways are:

- TGF- β SMAD-dependent signalling
- TGF- β SMAD-independent signalling
- receptor tyrosine kinase (RTK) signalling
- Wnt signalling
- Notch signalling

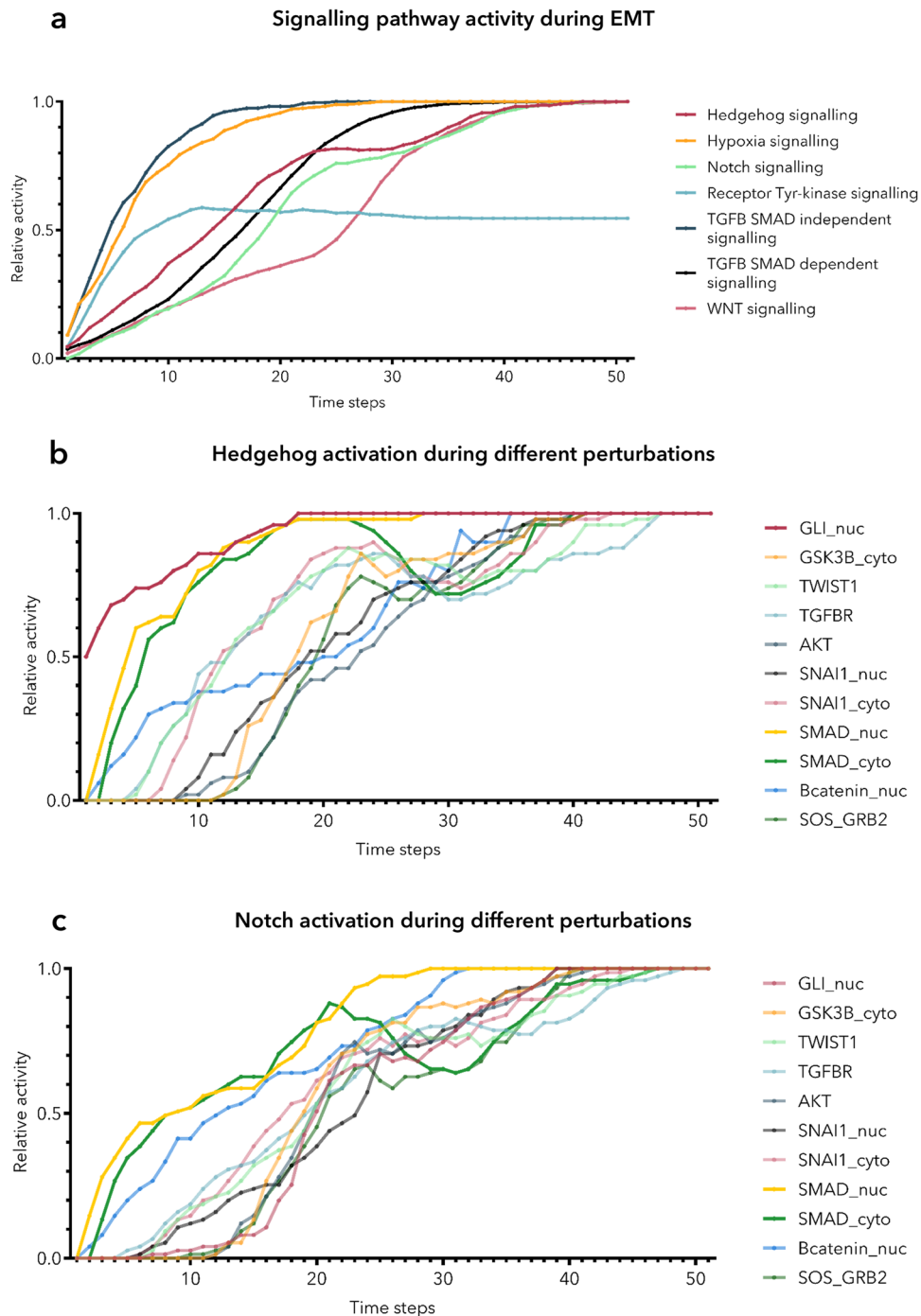


Fig. 5 Signalling pathway activities during EMT. **a** We plotted the activity of the main pathways during EMT on this panel. There are 11 single node perturbations that led to EMT, here we show the average of the pathway activities that can be observed during these 11 perturbations. We can see that there was a difference in the kinetics of the activation of different pathways. Importantly, we could still observe that TGFB activation is accompanied by the activation of the Hedgehog and Wnt pathways as captured by the original EMT model²³. **b** On this panel we show the activation of the Hedgehog signalling pathway for different perturbations. Each line symbolizes one specific perturbation which led to EMT and thus the activation of the Hedgehog pathway. We note here that the perturbation of the TGFB pathway member SMAD proteins (yellow and green) resulted in a quicker and more robust activation of the Hedgehog pathway. The most robust activation could be observed when we activated the nuclear pool of GLI proteins, with the perturbation of the GLI_nuc node (brown). **c** Here, we represent the activation of the Notch signalling pathway for different perturbations similarly as we showed it in the case of Hedgehog signalling on panel b. Here the activation of SMAD proteins resulted in a quick and robust activation of the Notch pathway which underlines the crosstalk between these pathways.

- Hedgehog signalling
- Hypoxia signalling.

This signalling pathway-level analysis confirmed that WNT and Hedgehog pathways are jointly activated during EMT as it was

highlighted in the case of the original publication^{23,48} (Fig. 5a), but also captured the activation of other important pathways (RTK, Notch and Hypoxia signalling). Moreover, the activatory perturbation of SMAD proteins resulted in a quicker activation of the

Hedgehog signalling pathway highlighting validated cross-talks^{49,50} between the TGFB and Hedgehog signalling pathways (Fig. 5b). Further analysis also captured the synergic functions between TGFB and Notch signalling^{51,52}, as the activation of TGFB pathway-member SMAD proteins led to quicker activation of the Notch pathway (Fig. 5c).

DISCUSSION

In this work, we presented the importance of translocating proteins in the signalling of the EMT and created a framework of how to create a compartmentalized Boolean model using a previously published reference work²³. Using this model, we simulated the EMT and presented key outcomes which can be considered as emerging properties of this new model. These new outcomes described experimental results better than the original EMT model due to our model's capability of better capturing compartment-specific functions of proteins.

It is known that a large proportion of proteins may localize to multiple compartments inside a cell¹⁰. This subcellular organization and the underlying dynamics are essential for physiological functions of human cells⁵³. We have numerically proven that translocating proteins are overrepresented between human proteins in the subset of human signalling and human EMT associated proteins (Fig. 1). We then created a framework which enables the incorporation of translocating proteins into cellular signalling models. We believe that the addition of translocating proteins to other models will greatly enhance *in silico* models and future studies.

First, we have presented that our model is able to recapitulate the results of the original EMT model. Our model maintains a stable epithelial and mesenchymal steady state (Fig. 4a), and certain perturbations (e.g., TGFB signal) can lead to a mesenchymal phenotype. It is important to note that this model simulates the transition from the epithelial state to the mesenchymal state so the mesenchymal state is the most robust attractor (Supplementary Table 1), and the inverse process of MET cannot be studied in this model (due to the fact that the regulations of this latter process are not involved in our model). With the above in mind, we can still conclude that our model produces biologically valid stable attractors with a slight enhancement in the stability of the epithelial attractor (which was very vulnerable to perturbations in the original model. See Supplementary Note 2 for details).

Based on the control sets of our model (Supplementary Data 4) we can also see, that the compartmentalization of a model reveals new control possibilities as expected. This is recapitulated by the fact, that our model reveals nontrivial compartmentalized node interventions. Hypothetically based on our model the control of the epithelial state can be reached by knocking out nuclear and cytoplasmic pools of GLI and AXIN2 proteins, respectively. In case of these compartment-specific perturbations there is no need to additionally knock out SNAI2 as in the case of the original EMT model (see Supplementary Data 4 Epithelial control sets No. 1 of the Compartmentalized and the Original model). Perturbations based on control methods may signal only *in vitro* intervention points, thus the control sets identified in compartmentalized models should be later validated through wet-lab experiments.

In comparison with other models, we must emphasize that there are not many system level models of the EMT²⁰. Kinetic and ordinary differential equation (ODE) models were shown to be useful in deciphering certain aspects of EMT, but they focus on a very limited number of nodes. To highlight some of such available models, one focused on a general minimal dynamic model and numerically described EMT based on among others the analysis of the phase diagram²². Another specific model addressed the shuttling of CTNNB and its antagonists²¹, and a third one focused on the bistable switches underlying EMT⁵⁴. Although these models are useful in the quantification of dynamics but have a

high computational demand. Thus, these models cannot be scaled up to model processes on a system level and they cannot be used to understand system level complexity and to plan system level interventions. The model of ref.⁵⁵ also builds on the Boolean model of ref.²³ but uses a slightly modified modelling approach and incorporates the LIF/KLF4 pathway to the model. Their model showed a more symmetric and more diverse representation of the attractor states involving a multitude of metastable hybrid phenotypic states. To test the robustness of our results we included the LIF/KLF4 pathway to our model. In this modified model we still observed the same compartment-specific functions (Supplementary Note 3), and no notable difference in the attractor landscape. This gives an important additional proof for the robustness of our method. Throughout this paper, we extensively compared our model with the Boolean model of ref.²³, which models EMT systematically, and our model succeeded the original EMT model in predicting compartment-specific functions.

The most important aspects of this study are those which highlight the compartment-specific functions of proteins. One of our most significant results is that GSK3B compartment-specifically inhibited EMT. We've predicted that nuclear activation of GSK3B can interfere with EMT and it prevents the activation of key factors of the mesenchymal state (Fig. 4c). In their paper, Lee et al.⁴³ created an experimental constellation in which exactly this phenomenon was examined. It is important to note that their paper was published after our initial results, so our results could be considered as computational predictions that are validated by this very recent experimental study. Lee et al.⁴³ claimed that enhancing the nuclear translocation of GSK3B i.e., the activation of GSK3B in the nucleus suppresses EMT in HCC cells⁴³.

Another result showed that GLI transcription family members compartment-specifically induced EMT (Fig. 4d). Our findings coincide with other studies showing that GLI1 aberrantly activates in DNA damage and carcinogenesis⁵⁶ and it activates EMT due to its transcriptional activity mainly through SNAI1⁵⁶. There are promising therapeutic targets among the SHH pathway members. Both GLI antagonists (GANTs) and Arsenic trioxide (ATO) (which also inhibits GLI functions) have been shown effective⁵⁷. Our simulations correctly identified GLI as an important factor during EMT, which is highlighted by the above examples. Further simulations that incorporate the actors of the SHH pathway complexly, could shed light on more impactful interventions that target GLI proteins.

Lately EMT is described as a multifaceted and often reversible process. To define EMT in its enormous complexity and plasticity in diverse developmental and pathological settings, studies must focus on complex molecular markers and changes in cellular properties simultaneously¹³. With this mindset we focused on multiple aspects of EMT during this study. We analysed attractors in their entirety and did not rely solely on some specific markers (e.g., loss of E-cadherin (CDH1) alone, or a dedicated EMT node). We also analysed signalling pathway activities (Fig. 5) which represented EMT more on a spectrum rather than just a binary outcome. Furthermore, we aimed to create a functional readout based on GO²⁵ terms, but this has limited usability as GO terms currently cannot capture the compartment-specific traits of proteins (see Supplementary Note 4 and Supplementary Fig. 6 for more on this matter).

In summary, we presented the enrichment of translocating proteins in EMT and to model their role we created a compartmentalized Boolean dynamic model of the EMT. This model was used to prove that translocating proteins, namely GSK3B and GLI proteins, can alter the fate of EMT in a compartment-specific manner. Our study also serves as a framework for future studies thanks to the algorithmic procedures available on the <https://translocaboole.linkgroup.hu> website and in the GitHub repository of this project. The results of our model underline that proteins should be modelled in their specific

location with compartment-specific functionality in order to maintain their physiological cellular behaviours. Also, the understanding of subcellular dynamics and the utilization of similar frameworks as ours in more complex networks will help uncover new biomarkers and therapeutic targeting strategies.

METHODS

Enrichment analysis of translocating proteins

To validate the enrichment of translocating proteins in human signalling and specifically in EMT we assessed the enrichment of translocating proteins as detailed below. The only available proteome level database of human translocating proteins is the Translocatome database, containing 13,066 human proteins. The database characterizes each protein with a Translocation Evidence Score (reflecting the translocation probability of that given protein). We found that in the Translocatome database 66.28% of the proteins are non-translocating proteins, 25.04% are low-confidence translocating proteins and 8.68% are high-confidence translocating proteins.

We used the Gene Ontology (GO)²⁵ database to identify the proteins, which take part in human signalling processes. To achieve this, we downloaded the list of proteins annotated with the GO term “signalling (GO:0023052)”. There were 6867 proteins annotated with this GO term and the translocation probability data (from the Translocatome database) was available in the case of 4991 proteins.

Similarly, we used the GO database to identify the proteins in connection with EMT. To achieve this, we downloaded the list of proteins annotated with the GO term “epithelial to mesenchymal transition (GO:0001837)”. There were 156 proteins annotated with this GO term and the translocation probability data (from the Translocatome database) was available in the case of 133 proteins.

In order to validate that the observed distribution of translocating proteins—in the case of signalling and EMT proteins—indeed significantly differs from the expected, we conducted Chi-square tests where we validated the observed datasets against the Translocatome database. In both cases, we found that the observed distributions significantly differ from the Translocatome database with a *P* value of smaller than 0.0001.

Manual curation

We conducted manual curation of translocating EMT proteins guided by translocation probability predictions of the Translocatome database⁵ that classified 13,066 human proteins into three groups: high- and low-confidence translocating and non-translocating proteins. We utilized the PubMed and Google Scholar search engines to find the relevant scientific publications for protein-translocation in EMT. It consisted of two phases: the selection of translocating proteins of the 19-node network and the exploration of the regulatory relations of these proteins.

In the selection-phase, we searched for the evidence of proteins translocating during the signalling of EMT to assess how extensive the literature was on the translocation of these proteins. In the exploration-phase, we gathered the necessary information for the system-level incorporation and Boolean rule creation of validated translocations into the model. Similarly to the methodology that was used in the manual curation of the Translocatome database⁵ the following preferences/conditions were set for the refined keyword searches:

- selection-phase:
 - Word combinations in quotation marks and “title” or “title/abstract” field-restrictions were used to refine keyword searches.
 - The publication had to be peer-reviewed and had to connect the translocation of a human protein to EMT.
 - The described translocation had to meet the requirements of the definition established in ref.⁵.
 - Recent publications (last 5 years) were preferred.
- exploration-phase:
 - Experimental evidence from hepatocellular carcinoma (HCC) cell lines were preferred otherwise other human carcinoma cell lines were preferred.
 - Publications exploring TGFβ mediated EMT were preferred
 - Some publications were tracked down from the references or the “cited by” page of review publications or relevant experimental publications not quite congruous with our conditions.

- Conditions/preferences of the selection-phase were also applied here.

In the selection-phase, we focused on finding evidence in the literature for the fact of a protein of the 19-node network translocating during EMT. Based on 64 publications we collected data about 21 proteins (see Supplementary Data 1) in the following topics:

1. name set, gene name and UniProt⁵⁸ accession number and link,
2. PubMed ID(s) and link(s) to peer-reviewed article(s) describing the experimental evidence of translocation,
3. initial and target localizations of the translocating protein,
4. interacting partners and biological functions (both in the initial and target compartments),
5. translocation mechanism,
6. the used detection method,
7. protein structural information on translocation mechanism,
8. disease group, exact disease involved and pathological role,
9. signalling pathways affected.

In the exploration-phase, we first searched for a review article of the molecular signalling mechanisms of EMT¹⁴ that could show us the well-established translocation events in the signalling of EMT. We then gathered information on the mechanism of the translocations (e.g., its regulation or dynamics) and the compartmentalized protein-protein interactions of the validated translocating proteins. We prioritized the review of publications that described interactions of a translocating protein with other proteins that were in the original network. For a detailed transcript of the manually curated information please refer to Supplementary Data 1.

Boolean network modelling and dynamic simulations

For the modelling of the compartmentalized EMT network inferred from the manually curated information, we utilized the Boolean framework described by ref.²³. In this every node is characterised by one of two qualitative states: ON (also referred to as TRUE or 1) or OFF (FALSE or 0) meaning above or below threshold activity/abundance, respectively. The state of a node is determined by its Boolean rule, which is a Boolean function that expresses the upstream regulatory relationships of the node. The Boolean rules of the compartmentalized model can be seen in Supplementary Data 2.

We updated the BooleanNet software package¹⁹ for the dynamic simulations of the model (see Data/Code availability section for the GitHub repository). The initial state of the simulations was the epithelial state. If perturbed by setting a node's state to a fixed OFF or ON state, the system may have converged into an altered attractor (corresponding to the steady mesenchymal state or a partial epithelial/mesenchymal state). Due to the stochastic nature of the unranked asynchronous update mode the outcome of a perturbation and the order of node-state updates could differ between simulations, so we performed 100 iterations of TGFβ OFF and 25 iterations of TGFβ ON simulations to assess these differences (there was no difference in the outcome of TGFβ simulations with 25 or 100 iterations). For validation, we compared the results of the perturbation analysis to experimental evidence found in the literature.

Signalling pathway analysis

In order to observe the activation of different signalling pathways during the course of our simulation we created a method to observe signalling pathway activity. Based on the work of Gonzalez et al.¹⁴ we mapped each node in our model to one of six major pathways contributing to EMT and identified the node's Boolean state which corresponds to signalling activity. This data is summarized in Supplementary Table 5. With this data, we were able to compute the activity of each of the six tracked signalling pathways in each timestep during a simulation. The activity of a signalling pathway is the average activity of the nodes which belong to that pathway.

Gene Ontology based functional analysis

Similarly, to the signalling pathway analysis, we aimed to create a functional analysis of our model as well. The Gene Ontology (GO) database²⁵ annotates all the human proteins with the biological processes connected to them. We monitored the activity changes of nodes during EMT and with the simultaneous utilization of the GO data we were able to

translate this node activity into the activity of biological processes. These mapped biological processes correspond to the core changes of the EMT.¹³

- cytoskeleton organization (GO:0007010)
- establishment or maintenance of epithelial cell apical/basal polarity (GO:0045197)
- epithelial cell-cell adhesion (GO:0090136)
- cell-matrix adhesion (GO:0007160)
- cell motility (GO:0048870).

Besides these, we also monitored the activity of the GO term “epithelial to mesenchymal transition (GO:0001837)”.

In order to translate node activities to process activities we identified which nodes’ activity represent which process’ activity, then the activity of a process is the average activity of the underlying nodes’ activity.

Attractor stability analysis

We were interested in how the stability of attractors in our compartmentalized EMT model compares to the 19 node Steinway et al.²³ model. To measure this, we conducted a numerical experiment on both models: we introduced an update noise, which returned the wrong state for every update with probability $p_{\text{error}} = 0.001$. This made it possible for the system to leave any attractors. Next, we simulated the long-term time-evolution (106 general asynchronous update steps) of the model initialized from one of the attractors and monitored how much time it spends in each visited state. We did this for every attractor and summarized the visitation probabilities for each attractor.

DATA AVAILABILITY

We made all relevant data available in the Supplementary Information, the Boolean model is defined in Supplementary Data 3 which is the most minimal dataset to reproduce our results. The codes used during the project are available in the GitHub repository: https://github.com/deriteidavid/compartmentalized_EMT_Boolean_model_Mendik_et_al_2021. In case of any further inquiry please contact the corresponding author.

CODE AVAILABILITY

The relevant codes used for data analysis and simulations are available in the GitHub repository: https://github.com/deriteidavid/compartmentalized_EMT_Boolean_model_Mendik_et_al_2021.

Received: 15 June 2021; Accepted: 20 May 2022;

Published online: 09 June 2022

REFERENCES

1. Scott, J. D. & Pawson, T. Cell signaling in space and time: where proteins come together and when they're apart. *Science* **326**, 1220–1224 (2009).
2. Gabaldón, T. & Pittis, A. A. Origin and evolution of metabolic sub-cellular compartmentalization in eukaryotes. *Biochimie* **119**, 262–268 (2015).
3. Andrade, M. A., O'Donoghue, S. I. & Rost, B. Adaptation of protein surfaces to subcellular location 11 Edited by F. E. Cohen. *J. Molecular Biol.* **276**, 517–525, (1998).
4. Csermely, P. et al. Learning of signaling networks: molecular mechanisms. *Trends Biochem Sci.* **45**, 284–294 (2020).
5. Mendik, P. et al. Translocatome: a novel resource for the analysis of protein translocation between cellular organelles. *Nucleic Acids Res.* **47**, D495–D505 (2019).
6. Laurila, K. & Vihinen, M. Prediction of disease-related mutations affecting protein localization. *BMC Genomics* **10**, 122 (2009).
7. Park, S. et al. Protein localization as a principal feature of the etiology and comorbidity of genetic diseases. *Mol. Syst. Biol.* **7**, 494 (2011).
8. Dhillon, P. K. et al. Aberrant cytoplasmic expression of p63 and prostate cancer mortality. *Cancer Epidemiol. Biomark. Prev.* **18**, 595–600 (2009).
9. Wang, J. et al. XPO1 inhibition synergizes with PARP1 inhibition in small cell lung cancer by targeting nuclear transport of FOXO3a. *Cancer Lett.* **503**, 197–212 (2021).
10. Thul, P. J. et al. A subcellular map of the human proteome. *Science* **356**, eaal3321 (2017).
11. Nieto, M. A., Huang, R. Y., Jackson, R. A. & Thiery, J. P. EMT: 2016. *Cell* **166**, 21–45 (2016).
12. Yang, J. & Weinberg, R. A. Epithelial-mesenchymal transition: at the crossroads of development and tumor metastasis. *Dev. Cell* **14**, 818–829 (2008).
13. Yang, J. et al. Guidelines and definitions for research on epithelial–mesenchymal transition. *Nat. Rev. Mol. Cell Biol.* **21**, 341–352 (2020).
14. Gonzalez, D. M. & Medici, D. Signaling mechanisms of the epithelial-mesenchymal transition. *Sci. Signal* **7**, re8 (2014).
15. Chaw, S. Y. et al. Epithelial to mesenchymal transition (EMT) biomarkers – E-cadherin, beta-catenin, APC and Vimentin – in oral squamous cell carcinogenesis and transformation. *Oral. Oncol.* **48**, 997–1006 (2012).
16. Csermely, P., Korcsmáros, T., Kiss, H. J. M., London, G. & Nussinov, R. Structure and dynamics of molecular networks: A novel paradigm of drug discovery: A comprehensive review. *Pharmacol. Therapeutics* **138**, 333–408 (2013).
17. Hastings, J. F., O'Donnell, Y. E. I., Fey, D. & Croucher, D. R. Applications of personalised signalling network models in precision oncology. *Pharmacol. Therapeutics* **212**, 107555 (2020).
18. Hyduke, D. R. & Palsson, B. Ø. Towards genome-scale signalling-network reconstructions. *Nat. Rev. Genet.* **11**, 297–307 (2010).
19. Albert, I., Thakar, J., Li, S., Zhang, R. & Albert, R. Boolean network simulations for life scientists. *Source Code Biol. Med.* **3**, 16 (2008).
20. Burger, G. A., Danen, E. H. J. & Beltman, J. B. Deciphering epithelial-mesenchymal transition regulatory networks in cancer through computational approaches. *Front Oncol.* **7**, 162 (2017).
21. Schmitz, Y., Rateitschak, K. & Wolkenhauer, O. Analysing the impact of nucleocytoplasmic shuttling of β -catenin and its antagonists APC, Axin and GSK3 on Wnt/ β -catenin signalling. *Cell. Signal.* **25**, 2210–2221 (2013).
22. He, P., Qiu, K. & Jia, Y. Modeling of mesenchymal hybrid epithelial state and phenotypic transitions in EMT and MET processes of cancer cells. *Sci. Rep.* **8**, 14323 (2018).
23. Steinway, S. N. et al. Network modeling of TGF β signaling in hepatocellular carcinoma epithelial-to-mesenchymal transition reveals joint sonic hedgehog and Wnt pathway activation. *Cancer Res.* **74**, 5963–5977 (2014).
24. Katsuno, Y. & Derynck, R. Epithelial plasticity, epithelial-mesenchymal transition, and the TGF- β family. *Dev. Cell* **56**, 726–746 (2021).
25. Ashburner, M. et al. Gene ontology: tool for the unification of biology. The Gene Ontology Consortium. *Nat. Genet.* **25**, 25–29 (2000).
26. min, K.-W., Lee, S.-H. & Baek, S. J. Moonlighting proteins in cancer. *Cancer Lett.* **370**, 108–116 (2016).
27. Mani, M. et al. MoonProt: a database for proteins that are known to moonlight. *Nucleic Acids Res.* **43**, D277–D282 (2015).
28. Kauffman, S. A. Metabolic stability and epigenesis in randomly constructed genetic nets. *J. Theor. Biol.* **22**, 437–467 (1969).
29. Wang, R.-S., Saadatpour, A. & Albert, R. Boolean modeling in systems biology: an overview of methodology and applications. *Phys. Biol.* **9**, 055001 (2012).
30. Zaňudo, J. G. T. & Albert, R. Cell fate reprogramming by control of intracellular network dynamics. *PLOS Computational Biol.* **11**, e1004193 (2015).
31. Steinway, S. N. et al. Combinatorial interventions inhibit TGF β -driven epithelial-to-mesenchymal transition and support hybrid cellular phenotypes. *NPJ Syst. Biol. Appl.* **1**, 15014 (2015).
32. Rozum, J. C., Gomez Tejada Zaňudo, J., Gan, X., Deritei, D. & Albert, R. Parity and time reversal elucidate both decision-making in empirical models and attractor scaling in critical Boolean networks. *Sci. Adv.* **7**, eabf8124 (2021).
33. Zaňudo, J. G. T. & Albert, R. An effective network reduction approach to find the dynamical repertoire of discrete dynamic networks. *Chaos: Interdiscip. J. Nonlinear Sci.* **23**, 025111 (2013).
34. Rozum, J. C., Deritei, D., Park, K. H., Zaňudo, J. G. T. & Albert, R. pystablemotifs: Python library for attractor identification and control in Boolean networks. *Bioinformatics* **38**, 1465–1466 (2022).
35. Borriello, E. & Daniels, B. C. The basis of easy controllability in Boolean networks. *Nat. Commun.* **12**, 5227 (2021).
36. Natsuizaka, M. et al. Interplay between Notch1 and Notch3 promotes EMT and tumor initiation in squamous cell carcinoma. *Nat. Commun.* **8**, 1758 (2017).
37. Chanrion, M. et al. Concomitant Notch activation and p53 deletion trigger epithelial-to-mesenchymal transition and metastasis in mouse gut. *Nat. Commun.* **5**, 5005 (2014).
38. Kurimoto, R. et al. Drug resistance originating from a TGF- β /FGF-2-driven epithelial-to-mesenchymal transition and its reversion in human lung adenocarcinoma cell lines harboring an EGFR mutation. *Int. J. Oncol.* **48**, 1825–1836 (2016).
39. Lemieux, E. et al. Constitutively active MEK1 is sufficient to induce epithelial-to-mesenchymal transition in intestinal epithelial cells and to promote tumor invasion and metastasis. *Int. J. Cancer* **125**, 1575–1586 (2009).

40. Buonato, J. M. & Lazzara, M. J. ERK1/2 blockade prevents epithelial-mesenchymal transition in lung cancer cells and promotes their sensitivity to EGFR inhibition. *Cancer Res* **74**, 309–319 (2014).
41. Mitra, T. & Roy, S. S. Co-Activation of TGF β and Wnt signalling pathways abrogates EMT in ovarian cancer cells. *Cell Physiol. Biochem.* **41**, 1336–1345 (2017).
42. Li, S. et al. The deubiquitinating enzyme USP48 stabilizes TRAF2 and reduces E-cadherin-mediated adherens junctions. *Faseb J.* **32**, 230–242 (2018).
43. Lee, S. et al. Inhibition of PI3K/Akt signaling suppresses epithelial-to-mesenchymal transition in hepatocellular carcinoma through the Snail/GSK-3/ beta-catenin pathway. *Clin. Mol. Hepatol.* **26**, 529–539 (2020).
44. Li, J. et al. Glycogen synthase kinase 3beta induces apoptosis in cancer cells through increase of survivin nuclear localization. *Cancer Lett.* **272**, 91–101 (2008).
45. Zhang, J., Tian, X. J. & Xing, J. Signal transduction pathways of EMT induced by TGF- β , SHH, and WNT and their crosstalks. *J Clin Med.* **5**, 41 (2016).
46. Niewiadomski, P. et al. Gli proteins: regulation in development and cancer. *Cells* **8**, 147 (2019).
47. Szczepny, A. et al. Overlapping binding sites for importin β 1 and suppressor of fused (SuFu) on glioma-associated oncogene homologue 1 (Gli1) regulate its nuclear localization. *Biochem J.* **461**, 469–476 (2014).
48. Behan, F. M. et al. Prioritization of cancer therapeutic targets using CRISPR-Cas9 screens. *Nature* **568**, 511–516 (2019).
49. Javelaud, D. et al. TGF- β /SMAD/GLI2 signaling axis in cancer progression and metastasis. *Cancer Res* **71**, 5606–5610 (2011).
50. Javelaud, D., Pierrat, M. J. & Mauviel, A. Crosstalk between TGF- β and hedgehog signaling in cancer. *FEBS Lett.* **586**, 2016–2025 (2012).
51. Lindsey, S. & Langhans, S. A. Crosstalk of oncogenic signaling pathways during epithelial-mesenchymal transition. *Front Oncol.* **4**, 358 (2014).
52. Klüppel, M. & Wrana, J. L. Turning it up a Notch: cross-talk between TGF beta and Notch signaling. *Bioessays* **27**, 115–118 (2005).
53. Gut, G., Herrmann, M. D. & Pelkmans, L. Multiplexed protein maps link subcellular organization to cellular states. *Science* **361**, eaar7042 (2018).
54. Tian, X.-J., Zhang, H. & Xing, J. Coupled reversible and irreversible bistable switches underlying TGF β -induced epithelial to mesenchymal transition. *Bio-physical J.* **105**, 1079–1089 (2013).
55. Font-Clos, F., Zapperi, S. & La Porta, C. A. M. Topography of epithelial-mesenchymal plasticity. *Proc. Natl Acad. Sci. USA* **115**, 5902–5907 (2018).
56. Palle, K., Mani, C., Tripathi, K. & Athar, M. Aberrant GLI1 activation in DNA damage response, carcinogenesis and chemoresistance. *Cancers (Basel)* **7**, 2330–2351 (2015).
57. Rimkus, T. K., Carpenter, R. L., Qasem, S., Chan, M. & Lo, H. W. Targeting the sonic hedgehog signaling pathway: review of smoothed and GLI inhibitors. *Cancers (Basel)* **8**, 22 (2016).
58. UniProt Consortium. UniProt: the universal protein knowledgebase in 2021. *Nucleic Acids Res.* **49**, D480–D489 (2020).

ACKNOWLEDGEMENTS

The authors appreciate the support of Prof Réka Albert, her ideas and comments in finalizing the manuscript. We are also grateful for the contribution of Orsolya Utasi during the data collection and of Donát Buszlai for creating our webserver. This work was supported by the Hungarian National Research, Development and Innovation Office [K131458], by the Higher Education Institutional Excellence Programme of the Ministry of Human Capacities in Hungary, within the framework of the Molecular Biology thematic programmes of Semmelweis University, by the Thematic Excellence Programme (Tématerületi Kiválósági Program, 2020-4.1.1.-TKP2020, TKP2021-EGA-24) of the Ministry for Innovation and Technology in Hungary, within the framework of

the Molecular Biology thematic programme of the Semmelweis University, and supported by the UNKP-20-III-2-SE-23 New National Excellence Program of the Ministry for Innovation and Technology from the source of the National Research, Development and Innovation fund.

AUTHOR CONTRIBUTIONS

P.M. and D.V.V. conceived the idea of the study. P.M. designed the study. P.M. and M.K. carried out manual curation and Boolean model creation. N.K. updated the code of the BooleanNet software and provided some scripts to the website. P.M. and M.K. ran and analysed the simulations. D.D. performed the stable motif and attractor stability analysis. P.M., M.K., S.K., and D.D. performed the data analysis and visualization and generated the figures and tables. Z.V. created the website and data analysis scripts. P.C.S. and D.V.V. provided support and supervised the study. P.M., M.K., and D.D. wrote the first draft of the manuscript. All authors read and contributed to the final version of the manuscript. P.M. and M.K. wish it to be known, that they should be regarded as Joint First Authors.

FUNDING

Open access funding provided by Semmelweis University.

COMPETING INTERESTS

The authors declare no competing interests.

ADDITIONAL INFORMATION

Supplementary information The online version contains supplementary material available at <https://doi.org/10.1038/s41540-022-00228-7>.

Correspondence and requests for materials should be addressed to Daniel V. Veres.

Reprints and permission information is available at <http://www.nature.com/reprints>

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2022