



OPEN

A quantitative model of cellular decision making in direct neuronal reprogramming

Adriaan Merlevede¹, Emilie M. Legault³, Viktor Drugge¹, Roger A. Barker², Janelle Drouin-Ouellet³✉ & Victor Olariu¹✉

The direct reprogramming of adult skin fibroblasts to neurons is thought to be controlled by a small set of interacting gene regulators. Here, we investigate how the interaction dynamics between these regulating factors coordinate cellular decision making in direct neuronal reprogramming. We put forward a quantitative model of the governing gene regulatory system, supported by measurements of mRNA expression. We found that nPTB needs to feed back into the direct neural conversion network most likely via PTB in order to accurately capture quantitative gene interaction dynamics and correctly predict the outcome of various overexpression and knockdown experiments. This was experimentally validated by nPTB knockdown leading to successful neural conversion. We also proposed a novel analytical technique to dissect system behaviour and reveal the influence of individual factors on resulting gene expression. Overall, we demonstrate that computational analysis is a powerful tool for understanding the mechanisms of direct (neuronal) reprogramming, paving the way for future models that can help improve cell conversion strategies.

Cell differentiation, the process that establishes cellular identity, traditionally follows a well-established hierarchy from totipotent to nilpotent cell types. However, the past decade has seen great advances in reprogramming procedures, allowing researchers to convert terminally differentiated cells to other cell types in vitro. This technology has promising applications for disease modelling, where an individual's skin fibroblasts can be converted to neurons and be used to determine their own unique neuronal pathology in vitro, as well as regenerative medicine, where replacement tissues can be sourced from a patient's own body. Direct reprogramming, in particular, allows conversion between terminally differentiated cell types without passing through an intermediate pluripotent state. This gives the opportunity to obtain mature neurons quickly (within a month) and allowing for relatively easy handling of dozens of lines by one experimenter. Moreover, they are not clonal, thus avoiding the risk of clonal bias. Importantly, multiple reports have recently demonstrated that directly reprogrammed neurons retain many important aspects of the age signatures of the donor, including age-related changes in the epigenetic clock, the transcriptome and microRNAs, the reactive oxygen species (ROS) levels, DNA damage and telomere lengths, as well as the metabolic profile and mitochondrial defects^{1–4}. As these cellular changes are suspected to play crucial roles in the development of age-associated disorders, this makes direct neuronal reprogramming an ideal approach to model neurodegenerative diseases for which age is the most important risk factor. However, because the source cells do not pass through a proliferating intermediate stage, a high conversion efficiency is required for large scale clinical application.

The working principle of contemporary reprogramming methods is to introduce a select combination of transcription factors and other biomolecules that manipulate the expression of key genes, initiating a cascade of regulatory mechanisms that control all known aspects of cell identity. Finding a combination of appropriate factors that induce the desired conversion with sufficient efficiency is the central difficulty for current reprogramming research, and is typically done by trial and error, using a set of candidate genes picked from databases and published resources based on the characteristic expression profiles of the source and target cell types. Some effort has gone into streamlining this process of selecting candidate genes⁵, but the efficiency of these tools is bounded by our knowledge of the underlying gene interactions.

One of the most notable cases is conversion of somatic dermal cells into neurons, which shows great promise in the study and treatment of neurodegenerative disorders⁶. Dermal fibroblasts are highly suitable source cells

¹Computational Biology and Biological Physics, Department of Astronomy and Theoretical Physics, Lund University, 223 62 Lund, Sweden. ²Cambridge Centre for Brain Repair, University of Cambridge, Forvie Site, Robinson Way, Cambridge CB2 2PY, UK. ³Faculté de Pharmacie, Université de Montréal, Montreal, QC H3T 1J4, Canada. ✉email: janelle.drouin-ouellet@umontreal.ca; victor.olariu@thep.lu.se

because they are abundant and easy to harvest. Many different cocktails of transcription factors have been successfully applied. Originally, Brn2, Ascl1 and Myt1l were shown to be sufficient in adult mouse fibroblasts⁷. In humans, homologous factors could induce neural conversion in fibroblasts at a very low efficiency⁸, which could be improved by the addition of other factors such as Ngn2, Sox2, NeuroD1/2, miR-9/9*, and/or miR-124^{6,9,10}. The efficiency of some of these methods has been increased by concomitant knockdown (KD) of the REST complex^{6,10,11}. Another approach involves knockdown of PTB^{12,13}. Other factors have been used to convert to specific neural subtypes⁹.

The experimental successes in the field of neural reprogramming indicate that, although the specialized identity of the cell constitutes many different aspects of cellular form and function, it is fundamentally controlled by only a few key regulators. Much interest has gone to identifying these key regulators and their roles, but the attempts to integrate this knowledge into a holistic understanding of the gene interaction network are still in the early stages. A quantitative model that captures the essential properties of the neural conversion mechanism represents an invaluable tool for suggesting new experiments and more efficient reprogramming methods. Researchers have engaged in such modelling in closely related fields such as the study of pluripotent cell commitment and reprogramming somatic cells to pluripotency. The resulting models successfully compress many experimental results into a single framework that can be interpreted by researchers, provides further experimental predictions, and can be used to suggest new experiments^{12–14}. With the goal of understanding direct neural reprogramming, an initial network hypothesis has been published¹³, but its ability to reproduce and explain the behaviour of the cell has not yet been verified in a data-driven analysis.

In this study, we built a quantitative model of the gene regulatory system governing direct reprogramming to neurons, based on known gene interactions that have been described in literature. We then measured the expression levels of key transcription factors at different time points during a neuronal reprogramming experiment, and used the resulting experimental data to evaluate the literature-based model. We found that this literature-based model was not able to explain the experimentally observed dynamic behaviour, and that the data suggested a missing feedback loop in the gene regulatory network. Accordingly, specific adjustments to the network topology could drastically increase the match between model simulations and experimental observations. In particular, all the adjustments that were shown to substantially improve the model were characterized by an activation from nPTB to PTB, not previously described in literature. Further, we showed that the gene regulatory network equipped with this additional interaction can successfully predict the reactions of the cells to various overexpression and knockdown perturbation experiments, including several different neuronal conversion strategies. In contrast, models based on network hypotheses without this crucial interaction could not reproduce the correct system behaviour. Together, these results suggest that nPTB feeds back into the neural conversion gene network playing an activating role in the expression of PTB, for example by blocking the negative self-regulation of PTB. Thus nPTB induces genes that are opposing neural conversion. In support to this, we experimentally show that a knockdown of nPTB results in successful neuronal conversion. In addition, we present a novel approach to dissect the model to reveal the influence of individual interactions during the conversion process, potentially illuminating internal mechanisms of the network that are hard to observe in the lab. By incorporating the mechanisms of different conversion methods into one explanatory framework, this work aims towards an integrated and predictive understanding of cellular decision making in direct neuronal reprogramming.

Results

Measured transcription levels during neuronal reprogramming. We have previously shown that cells that are converted by knocking down REST adopt a transcriptome that more closely resembles that of a neuron, compared to the reprogramming approach involving the overexpression of miR-9/9* and miR-124⁶. However, the combination of both the REST knockdown and miR-9/9* and miR-124 overexpression promotes neuronal maturation¹⁰. By building on this work, we sought to have a better understanding of how the different components of the gene regulatory network (GRN) controlling direct conversion to neurons interact. We thus generated induced neurons (iNs) through the knockdown of REST, followed by the forced expression of the transcription factors Ascl1 and Brn2 (Fig. 1a). This reprogramming protocol, which leads to a striking change towards a neuronal morphology during the first 25 days of conversion³¹, generates iNs that express synaptic markers such as synapsin, synaptophysin and that are electrophysiologically functional⁶, and genes of specific neurotransmitter phenotypes, including somatostatin (SSTR1), GABAergic (GABRA1), glutamatergic (GRIA2), acetylcholinergic (CHRNA43), and dopaminergic (DRD1)¹⁰. The neuronal identity of the reprogrammed cells was also confirmed in this study by the expression of the mature neuronal markers TAU and MAP2 at day 25 (Fig. 1b). This reprogramming strategy generates iNs at an efficiency of $43.9 \pm 8.9\%$, as calculated by the percentage of the total number of TAU⁺ cells over the number of starting fibroblasts, as well as a purity of $44.1 \pm 1.1\%$, as calculated by the percentage of the number of TAU⁺ cells over the total number of cells in the dish at day 25 (Fig. 1c). We measured mRNA expression of known key transcription factors in the fibroblast before conversion, as well as three days following the REST knockdown, and 8 h, 1, 2, 3, 5, 7, 14, and 21 days following the induction of viral expression vectors (Fig. 1d). Endogenous Ascl1 and Brn2 were measured separately from their viral counterparts. No endogenous Brn2 transcription was detected.

Our measurements showed intense fluctuations in the expression levels during the first 5 days, followed by stagnation or monotonous increase during the next 2 weeks. This is consistent with previous observations that commitment to neuronal development occurs quickly in the reprogramming process in human cells, followed by a longer period in which downstream genetic and epigenetic control mechanisms are rewired^{15–17}. The observed dynamics are typical of a system responding to an initial input shock before settling into a new stable regime. In particular, the dampened fluctuations that are seen in the expression of these genes over the first few days is characteristic of gene regulatory networks containing a negative feedback loop.

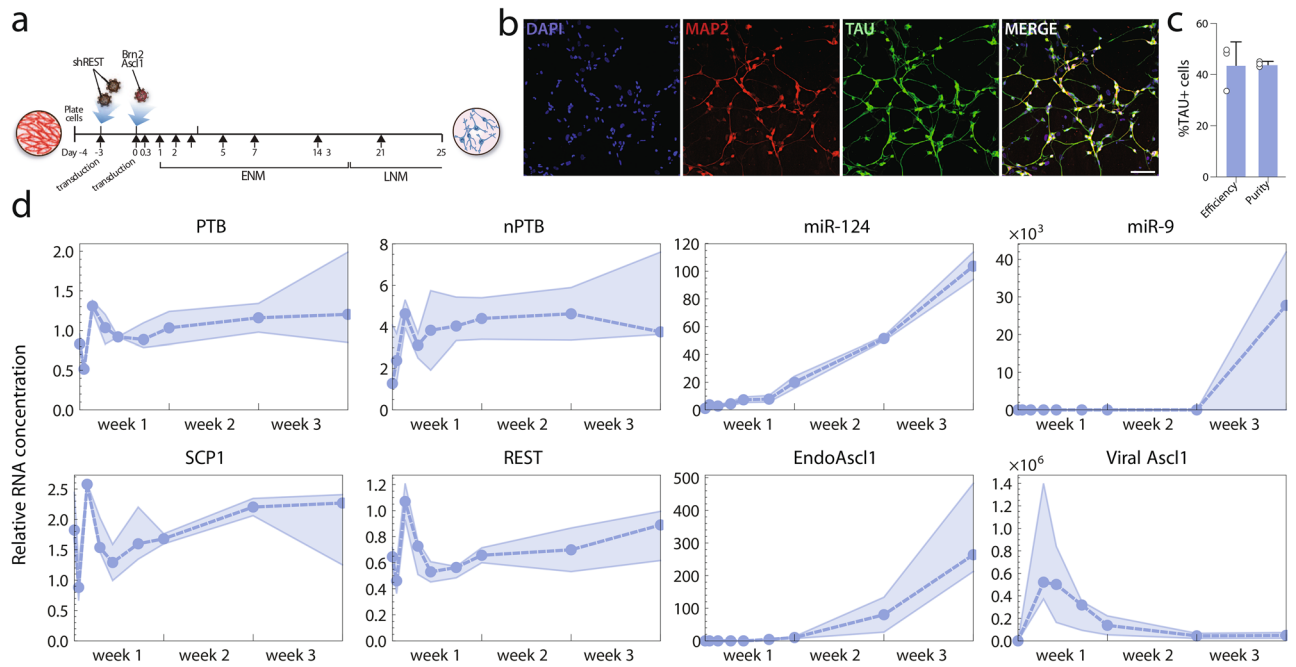


Figure 1. Measured transcription levels during neural reprogramming. **(a)** A schematic representation of the reprogramming process. **(b)** Immunofluorescence staining of MAP2 (in red) and TAU (in green) showing reprogrammed iNs at day 25 post transduction. Cells are counterstained with DAPI (in blue). Scale bar = 100 μ m. **(c)** Quantification of TAU + cells at day 25 post transduction, error bars are defined as S.D. **(d)** Measured transcription levels. The dashed line represents the median of three replicates, surrounded by a shaded region between the minimum and maximum values at each time point. The measurements are normalized so that a value of 1 is equal to the fibroblast level, or to the detection limit in the case of Ascl1, which was not detected in fibroblasts. ENM early neuronal medium, LNM late neuronal medium. The graphs were generated using Graphpad Prism 8, the fluorescent images were put together using Adobe Photoshop 2020. <https://www.graphpad.com/scientific-software/prism/>, <https://www.adobe.com/se/products/illustrator.html>.

We noted that the viral expression was not strongly correlated with the other expression curves, and even at lower observed levels was several orders of magnitude above endogenous expression in fibroblasts and throughout the conversion experiment. In the rest of this study, we assumed that the regulatory mechanisms depending on Ascl1 were saturated during the time frame where it was virally expressed, and considered the absence or presence of Ascl1 as an on-off switch in our models.

Testing the literature-based network. In the last decade, there has been considerable interest in the interaction between several key factors that play an important role in establishing and maintaining neural identity. Based on these studies, we integrated the state of the art for understanding relevant gene interactions (Table 1) into a gene regulatory network model with dynamics suitable for comparison with our dataset (Fig. 2a). This GRN is based especially on previous work by Xue et al.^{12,13,19}. We focused here only on neuronal conversion, not maturation.

Two nodes in our network represent the combined expression of multiple physical regulatory elements: the miRs node represents miR-124 and miR-9/9* while the REST complex node is an aggregate of its multiple components, including the REST gene and SCP1 cofactor. This approach was motivated by high similarity between the relevant transcription levels during our experimental time frame (correlation coefficient of median concentration levels $r = 0.89$ (miR-124–miR-9) and $r = 0.89$ (REST–SCP1)), especially during the first five days ($r = 0.97$ (miR-124–miR-9) and $r = 0.93$ (REST–SCP1)). Thus, we observed almost no quantitative distinction between these regulatory elements during the conversion process. Experimental data gathered from other conversion methods could potentially distinguish the action of these factors in the future.

We developed a quantitative computational model for the GRN in Fig. 2a, using the Shea-Ackers formalism²⁰, yielding a system of ordinary differential equations that predicted the dynamics of RNA concentrations over time for each node in the network. The interaction strengths and reaction rates were unknown parameters in these equations; we used an evolutionary algorithm to find the parameter values which yielded the closest match between predicted and observed transcription dynamics. The model was executed in three modes corresponding to the three stages of our experimental conversion process: the initial fibroblast cell stage, the REST knockdown stage, and the conversion stage with Ascl1 and Brn2 overexpression. The switch between the three modes was modelled by manipulating two parameters representing the action of external factors i.e. REST knockdown and Ascl1 overexpression, respectively. System parameters were optimized to maintain a steady state in the fibroblast stage, to match the observed expression after 3 days of REST inhibition, and to fit as closely as possible the

<p>REST complex RE1 silencing transcription factor and its cofactors repress many neural genes in non-neural cells^{32,33}. Functions as a barrier to neural differentiation, especially in human adults⁶. Knockdown enhances reprogramming efficiency^{6,11,28}, but does not cause reprogramming by itself in vitro³⁴, while forced constitutive expression prevents successful differentiation^{26,35}. Represses the expression of miR-124^{6,36,37}. Represses many neuronal genes including Ascl1³⁷.</p>
<p>miR-124 and miR-9/9* (miRs) Micro-RNAs found in neurons, responsible for repressing a wide array of non-neuronal genes³⁶. Overexpression leads to a neuron-like expression profile^{17,26}, and conversely, forced downregulation leads to a more non-neuronal profile³⁶. Downregulate the activity of the REST complex^{28,37,38}. Represses PTB, and to a lesser extent nPTB, during neuronal differentiation²⁷.</p>
<p>PTB and nPTB Polypyrimidine tract binding protein and its neural paralogue control alternative splicing (often followed by nonsense degradation²³) of many neural genes²¹. nPTB replaces PTB during neural differentiation, and its expression decreases as part of neuronal maturation³⁹. Knockdown of PTB leads to neural reprogramming¹⁶ (but not in all cases²²), while forced expression blocks neuronal differentiation²⁶. PTB competes with the repression of SCP by miRs¹⁶, thus activating REST. PTB represses nPTB as well as itself²⁷; nPTB has the same functionality²².</p>
<p>Ascl1 Most prominent member of a family of proneural basic-helix-loop-helix proteins that control neural differentiation but also proliferation of neural progenitors^{40,41}. Overexpression leads to neural conversion⁴² (but not in all cases²⁶), especially in combination with other factors^{7,29}. Knockout blocks neuronal differentiation in parts of the mouse brain⁴³. miR-124 is overexpressed in cells transfected with Ascl1, indicating an activating interaction or pathway from Ascl1 to miRs²⁹.</p>

Table 1. Genes in the core regulatory network. These factors have been indicated as key to establishing and maintaining neural identity.

expression levels dynamics during the conversion stage after viral activation. The optimisation and modelling procedures are explained in more detail in the "Methods" section.

When fitting our simulation outcomes to the experimental time series data, we found that our initial model based on interactions described in literature was not able to explain the dynamic behaviour of the system (Fig. 2b). The following observations stood out as possible explanations for the inability of the model to fit the data.

While the behaviour of the system during the first five days of observation appeared to be dominated by the shock response to the sudden addition of the viral vector-delivered Ascl1 to the system, there was no interaction in the network that could explain these heavy fluctuations in expression levels. In particular, the strongest fluctuations occurred during the first day, when PTB, REST, and miRs drastically switched from increasing to decreasing transcription levels, or vice versa. These fluctuations are evidence of a negative feedback loop in the network: a gene that activates its own repressor, or represses its own activator, causes a delayed self-repression that can result in several rounds of increase and decrease of transcription after sudden changes. No negative feedback has been suggested before in literature on direct neural reprogramming dynamics.

In the literature based GRN, Ascl1 was controlled exclusively by inhibition from the REST complex. Therefore, it was activated only by constitutive transcription factors not present in the network. However, REST reaches its lowest expression value during the first day, where little or no Ascl1 is expressed (Fig. 1d). This suggested that another regulatory process might have been necessary to accurately explain why Ascl1 is expressed after five days but not before. Alternatively, it is possible that the low REST expression is simply not sustained for long enough during the first day to accumulate sufficient Ascl1 protein concentration.

Finally, we note that previous publications have described inhibiting interactions between PTB and nPTB, as both proteins have been found to inhibit the correct splicing of their own and each other's mRNA transcripts^{21–23}. Of these interactions, only the inhibition of nPTB by PTB was included in the literature model. Self-inhibition and regulation of PTB by nPTB have not received attention in literature as causal elements influencing the reprogramming process, and have not been included in previous models that integrate important interactions into a GRN^{12,13,19} (Fig. 2a). However, it has not been ruled out that these interactions play a role in the quantitative expression dynamics of the neural reprogramming system.

Comparison of alternative network topologies. To investigate whether a small modification to the GRN topology could improve the match between predicted and observed transcription levels, we attempted to fit models with alternative sets of gene interactions to our data. When adding a new hypothetical interaction to the network, a quantitative model was constructed and trained in the same systematic way as the original literature-based model. Then fitting errors were compared (Fig. 3a).

We hypothesised that a missing control mechanism for Ascl1 is responsible for the lack of fit between the literature model and measurements, as discussed in the previous section. In terms of model fitting, this hypothesis entails that RESTc is constrained to remain at higher expression levels during the first day in order to avoid activation of Ascl1, and that this constraint potentially affects the regulation of other genes by RESTc. To test whether the ability of the model to fit observed data is indeed constrained in this way, we constructed an alternative model where Ascl1 was activated at the correct time, so that RESTc is no longer constrained by the need to regulate Ascl1. Because miRs reach maximal expression at the same time as Ascl1 in our data, it was convenient to implement this in practice as an activation of Ascl1 by miRs, which allows Ascl1 to be activated at the correct time with an appropriate set of parameters without constraining RESTc or miRs. Figure 3a (blue) shows that introducing an activation of Ascl1 by miRs resulted in a small improvement compared to the unmodified literature model.

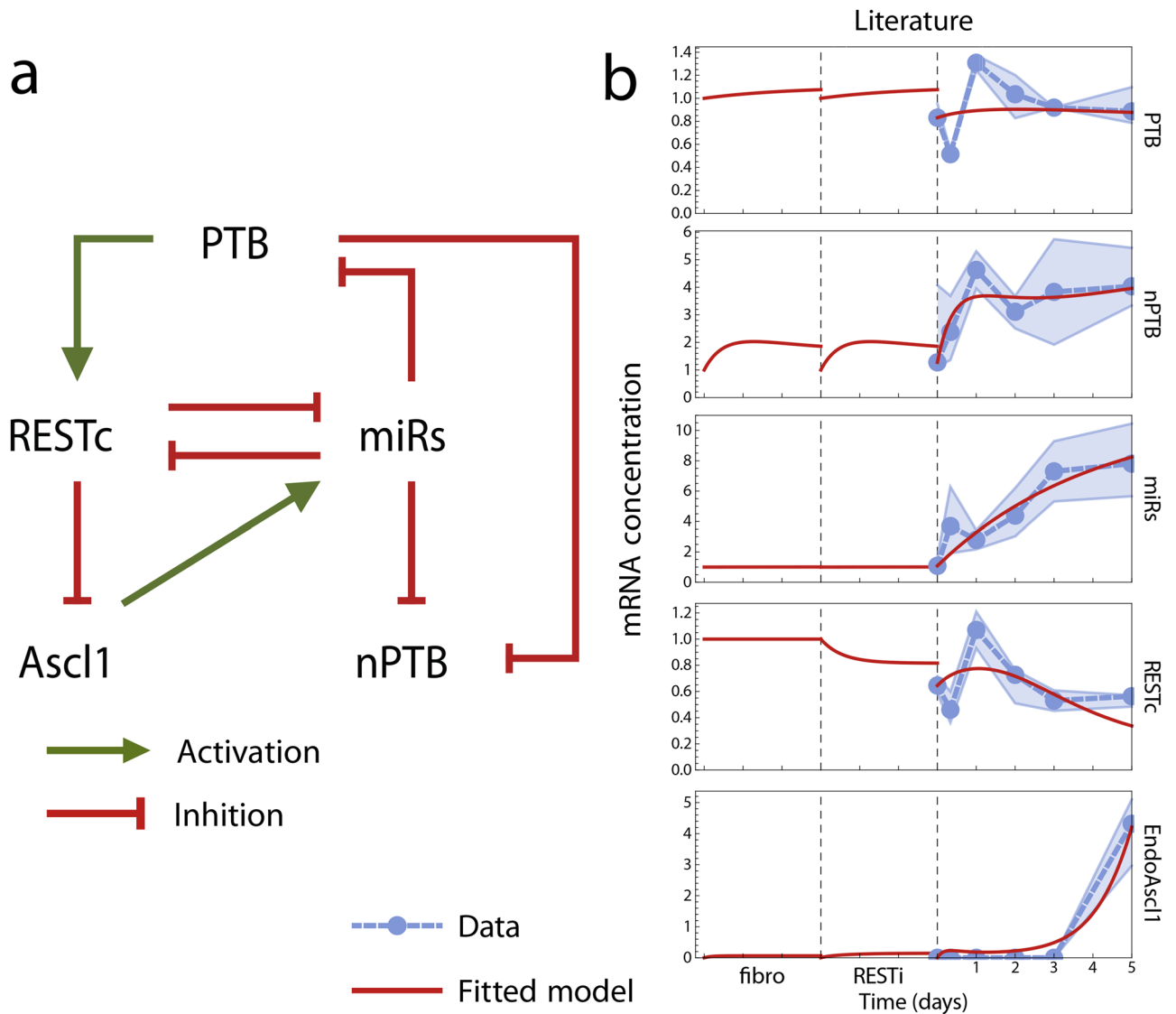


Figure 2. Known interactions are not sufficient to explain observed reprogramming dynamics. **(a)** Literature-based network topology. Green pointed arrows represent activations, red flat arrows represent inhibition interactions. **(b)** Predicted transcription levels according to quantitative model based on gene network in A, fitted to our experimental data. Time scale is divided in three phases, corresponding to fibroblast, REST knockdown, and conversion stages of the experiment. The plots were made with Mathematica 11.3.0 (Linux version). <https://www.wolfram.com/mathematica/>.

We also considered the potential role of self-inhibition of nPTB and PTB, as well as inhibition of PTB by nPTB, which are interactions that have been observed experimentally. The fitting errors of the resulting model are shown in Fig. 3a (yellow). The PTB and nPTB inhibitions did not affect the ability of the model to match observed data.

We further investigated the hypothesis that another, previously unknown interaction is missing in the GRN. To this end, we constructed models for all possible network alternatives where a single interaction was added to the initial literature-based network. The added interactions were between the four active nodes PTB, nPTB, RESTc and miRs, excluding self-interactions (a node activating or repressing itself) and double-interactions (a node acting as an activator and repressor for the same target). The fitting error of these models is shown in Fig. 3a (one additional interaction).

For each of the tested models, there is also an alternative where Ascl1 is activated by miRs, increasing the freedom to fit RESTc (shown in blue); and another alternative where self-inhibition of PTB and nPTB, and inhibition of PTB by nPTB, were included (shown in yellow). We note that this comparison includes many different models, and therefore introduces the potential dangers associated with multiple hypothesis testing. However, comparison between multiple similar models is also a useful indicator of whether any fitting improvements are unique to a particular hypothesis, or simply result from the generic increase in free parameters.

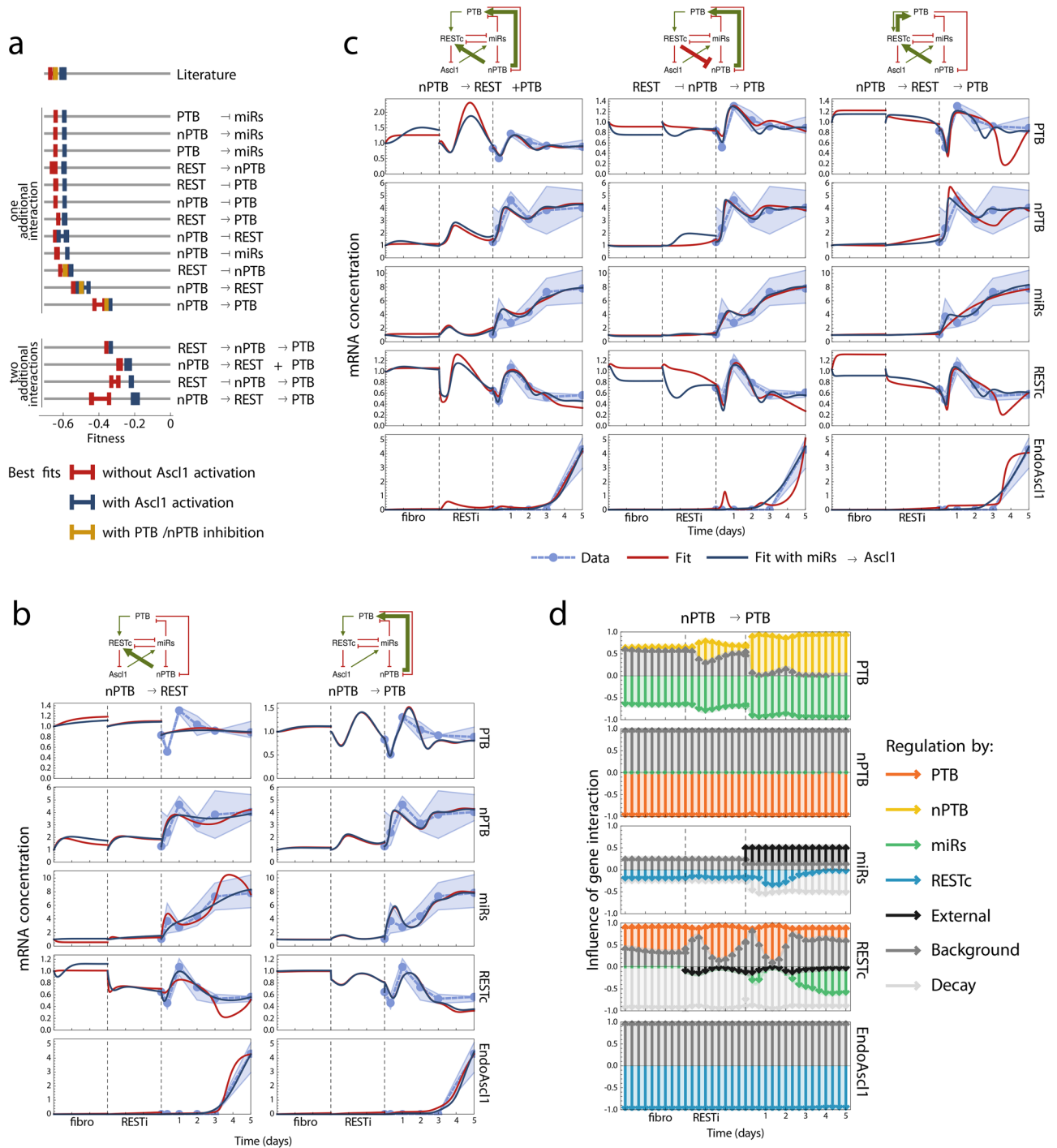


Figure 3. Model performance increases drastically when nPTB activates PTB. **(a)** Model performance as measured by the objective (fitness) function of the optimization procedure, that is, the negative fitting error. Performance of the three best fits is shown. Models are grouped by the hypothetical interaction they have in addition to the literature-based network, where \rightarrow denotes activation, and \dashv denotes inhibition. **(b)** The time evolution of the RNA concentrations predicted by the nPTB \rightarrow REST and nPTB \rightarrow PTB models. Time scale is divided in the three experimental phases as in Fig. 2b. **(c)** The time evolution of the DNA concentrations in the best models with two additional interactions. **(d)** Dissection of the simulated time evolution in the nPTB \rightarrow PTB model (see **(b)**), decomposing the transcription rates of each factor into the separate contributions of its activators and inhibitors. Activating interactions are represented by arrows pointing upwards from the zero line, inhibiting interactions point downwards. Arrow length indicates relative influence of the regulator. See Methods for an extended description. nPTB and Ascl1 are controlled by changes in inhibition regulation that are too small to see on the figure. The plots were made with Mathematica 11.3.0 (Linux version). <https://www.wolfram.com/mathematica>.

Adding new interactions (and thus more free parameters) did not affect the fitting error of these models in most cases. Only two models showed a marked improvement, and in particular the match between simulated and observed gene expression was enhanced when nPTB activated PTB (nPTB \rightarrow PTB). To a lesser extent, fitting also improved when nPTB activated REST (nPTB \rightarrow REST). The expression levels predicted by these models over the course of the conversion experiment are shown in Fig. 3b. Notably, the nPTB \rightarrow PTB model introduces direct negative feedback between nPTB and PTB, a network motif that was strongly suggested by visual inspection of the data. The two models (nPTB \rightarrow PTB and nPTB \rightarrow REST) are similar in the sense that each implies the other through an intermediate reaction chain (nPTB \rightarrow [PTB \rightarrow] REST and nPTB \rightarrow [REST \rightarrow miRs \rightarrow] PTB).

To further investigate these interactions, we questioned whether an even better fit to our data could be obtained by combining two different interactions (not including the additional activation of Ascl1 by miRs). We constructed several models with two interactions between PTB, nPTB and RESTc. These combinations of multiple interactions resulted in a small decrease of the fitting error in some cases, more so when RESTc was not constrained to the regulation of Ascl1, as shown in Fig. 3a (two additional interactions). However, on visual inspection, there is no substantial change in the fit between the simulated and observed expression dynamics, despite the added free parameters, as shown in Fig. 3c. These experiments indicate that the improvement obtained by adding the nPTB \rightarrow RESTc interaction are subsumed by the nPTB \rightarrow PTB model, since combining both interactions (labelled nPTB \rightarrow PTB + RESTc) did not result in a marked improvement compared to the simpler nPTB \rightarrow PTB model. Thus, we can consider the nPTB \rightarrow RESTc model a weaker version of nPTB \rightarrow PTB.

These results single out nPTB \rightarrow PTB as the most valuable hypothetical interaction with respect to improving the fitting error of our models. Two other interactions could decrease fitting error (nPTB \rightarrow RESTc and miRs \rightarrow Ascl1), but when combined with nPTB \rightarrow PTB these interactions are both shown to be obsolete. Indeed, the nPTB \rightarrow PTB model could not be improved by the added freedom in RESTc conferred by the miRs \rightarrow Ascl1 interaction (Fig. 3b), unlike all other models (Fig. 3a). Together, these results show that an activating interaction between nPTB and PTB was highly compatible with our data.

We finish this section with a more detailed view of the nPTB \rightarrow PTB model. The equations governing the quantitative models presented here were built up from the described gene interactions using a systematic procedure. Different terms in these equations correspond to specific gene interactions. By comparing the magnitude of partial expressions that represent contributions of individual gene interaction, we can measure and visualize the relative influence of each activator or inhibitor on the total transcription rate of each gene at different time points during a simulation. This dissection of the model allows us to distinguish the roles of different gene interactions by directly observing the (model) system's internal mechanisms in full detail. This is a unique benefit of computational modelling, since in the lab it is only possible to observe the behaviour of the network as a whole, for example in response to overexpression or knockdown of genes, leaving the contributions of individual interactions to inexact inference. Figure 3d shows such a dissection analysis for a simulation of the nPTB \rightarrow PTB model (the corresponding timed expression profile is shown in Fig. 3b). See Methods for the analytical details of this analysis. Figure 3d shows that the inhibition from miRs has a negligible impact on nPTB expression in the model after training on the observed data. This interaction may still contribute to the behaviour of the real cell, but apparently its function is redundant (i.e. is functionally equivalent to the inhibition by PTB) in the conditions of our conversion strategy.

Model performance in overexpression and knockdown scenarios. Having compared several network hypotheses that are highly compatible with gene expression data during one conversion process, we set out to investigate whether the fitted models would also reproduce the behaviour of the cell in response to other stimuli. To this end, each model was subjected to simulated overexpression (OX) or knockdown (KD) of factors that are known to cause or block neural conversion (see Table 1). The results could then be compared to experimental observations. Since these qualitative data are independent of our training data, they are suitable for model testing and validating comparison.

In order to simulate analogues to these experiments in our models, we explored the effect of modifying relevant parameters on the system behaviour. Viral induction of Ascl1 and REST knockdown were explicitly represented in the system equations and thus could be controlled by manipulating the dedicated system parameters. The same method was used when we simulated the switch between the three stages of the experimental procedure. All other cases of overexpression and knockdown were simulated by increasing (5 \times) or decreasing (1/5 \times) the transcription rate coefficient of the factor under consideration. The resulting system of equations were then integrated to yield the evolution of transcription levels over time, starting from the fibroblast mRNA expression level as an initial state. Because many direct reprogramming methods operate at low efficiency, the deterministic integration representing bulk average expression levels of a population was supplemented by a stochastic approach, to capture noise in individual cells. We used the Gillespie algorithm to produce stochastic simulations over a simulated time span of 14 days²⁵ which introduces gene regulatory network intrinsic noise in the system. Endogenous expression of Ascl1 was taken as a convenient indication that a simulated cell had committed to a neuronal fate; a maximum endogenous Ascl1 value of 1 is taken as the threshold value to indicate conversion. Results are shown in Table 2 and Supplementary Fig. 1 for the models with the lowest fitting error (alternative models with miRs \rightarrow Ascl1 or PTB/nPTB inhibition are not shown).

Out of all the models presented here, nPTB \rightarrow PTB (Fig. 4a) provided the most accurate predictions of system behaviour in different OX/KD scenarios, improving substantially upon the purely literature-based model. The stochastic and deterministic simulation results corresponding to OX/KD scenarios with the nPTB \rightarrow PTB model are shown in Fig. 4b,c.

Stochasticity of these simulations is necessary to capture heterogeneity of cellular populations. In particular, overexpression of Ascl1 only caused neuronal conversion (endogenous Ascl1 expression) in some simulations,

		Fibroblast	OX Ascl1	OX miRs	KD REST	OX Ascl1 + KD REST	OX Ascl1 + OX REST	OX Ascl1 + OX PTB	KD PTB
Simulation outcome	Lab outcome	No	Yes	Yes	No	Yes+	No	No	Yes
	Literature model	No	Yes	Yes	No	Yes	Yes	Yes	No
	nPTB → REST	No	Yes	Yes	Yes	Yes	No	Yes	No
	nPTB → PTB	No	Yes	Yes	No	Yes+	No	Yes	No
	nPTB → REST + PTB	Yes	Yes	Yes	Yes	Yes	Yes	Yes	No
	nPTB → REST → PTB	No	No	No	Yes	Yes+	No	No	No
	REST → nPTB → PTB	Yes	Yes	Yes	Yes	Yes	No	No	Yes

Table 2. The nPTB → PTB model correctly predicts conversion outcome of most overexpression and knockdown scenarios. Conversion outcome is shown for several OX and KD experiments described in literature (discussed in Table 1), and compared to outcome predicted by quantitative models. Simulated outcome was considered to predict conversion (“Yes”) if Ascl1 expression exceeded 1 at any time point during at least 1 out of 50 simulations, and no conversion (“No”) otherwise. In the case of OX Ascl1 + KD REST, all models predict conversion because it is an explicit feature of the training data. For this reason, a higher conversion efficiency was required (indicated by a plus sign) in this experiment as compared to OX Ascl1. Bold cells indicate outcomes that correspond to lab experiments. Full simulation results are shown in Supplementary Fig. 1.

and the conversion efficiency (fraction of simulations with high endogenous Ascl1 levels) could be greatly increased by addition of REST knockdown, in the nPTB → PTB model. This is in accordance with experimental results⁶. However, it bears mentioning that simulated experiments with Ascl1 overexpression correspond most closely to the conditions of the training data for these models. In other conditions, simulated heterogeneity is less realistic, for example in the case of miRs overexpression. In general, single-cell data would be required to make quantitative predictions of conversion efficiencies.

Experimentally validating the nPTB → PTB model. In order to test the existence of feed-back from nPTB to the core neural conversion network by inducing PTB and subsequently opposing the neural conversion, we conducted a nPTB knockdown experiment. Our results showed that a nPTB knockdown generates iNs at a similar efficiency to the one obtained through the repression of REST (Fig. 5a,b). Furthermore, the complexity of the neuronal morphology, as assessed by measurements of the total surface area of neurites and the number of branchpoints, was also similar, suggesting that both strategies are equipotent (Fig. 5c,d). Thus, our results showed that repression of PTB by itself and nPTB does not improve neuronal conversion while a knockdown of nPTB, which is a PTB inducer, leads to iNs conversion. We also observed that simultaneous depletion of REST and nPTB during the initial phase of conversion results in a lower conversion efficiency and neuronal complexity (Fig. 5c,d). We speculate here that this double depletion leads to expression of REST which is under the levels necessary for successful conversion to iNs.

To test whether the nPTB activation is crucial at the initial conversion stage but not required for the neuronal maturation, we have tested experimentally whether a nPTB knockdown done after the initiation stage would improve maturation. We have previously identified that the major changes in gene expression occur within the first 5 days after the initiation of conversion^{6,10}. We thus initiated the nPTB knockdown on day 7 at the beginning of the maturation phase, which did not improve on the number of neurons, and significantly gave rise to less complex neuronal morphology (Fig. 5c,d). This suggests that nPTB is important for the maturation of the neurons when the conversion is initiated through a depletion of REST.

Finally, to compare between the REST KD- and the nPTB KD-based reprogramming, we have performed gene expression profiling of the main actors in our model at 8hrs intervals during the first 48 h post transduction with Ascl1 and Brn2. Please note that through this experiment we also performed a zoom in exercise by monitoring the gene expression more often than in our training data. Our results show that PTB, REST and endogenous Ascl1 follow the same transcriptional fluctuations during the first 48 h during both the REST KD and the nPTB KD experiments. The resulting experimental time series data showed PTB, REST and endogenous Ascl2 expression downregulation early on followed by an increase around the second day of reprogramming (Fig. 5e), suggesting the presence of a negative feedback loop in the network in both KD cases. Furthermore, the knockdown of nPTB results in a slight downregulation of PTB expression confirming our model prediction of positive interactions between nPTB and PTB. We also observed that nPTB KD lead to diminished expression of REST which is in line with the double positive input on REST from nPTB via PTB. However, the REST knockdown did not impact on nPTB expression levels, suggesting that the double negative interaction between REST and nPTB via miRs is not a crucial mechanism in the network. These results suggest that the successful neuronal conversion seen in (Fig. 5a,b) could be the result of the REST downregulation in these cells either directly or through depletion of nPTB which positively induces REST through PTB.

Dissecting the model response to PTB knockdown. Generally, there was a remarkable agreement between the system behaviour predicted by the nPTB → PTB model and experimental observations, even though

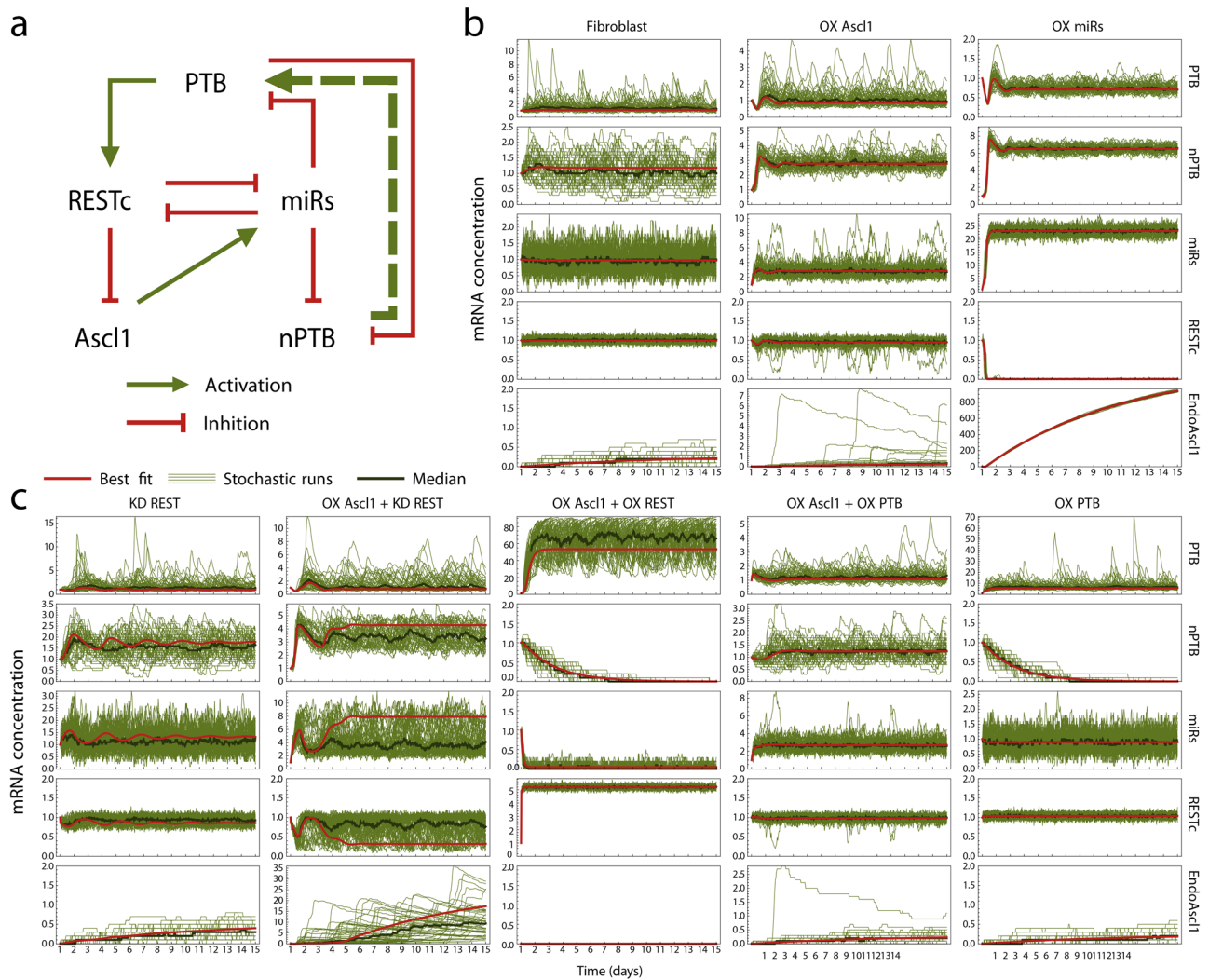


Figure 4. The nPTB → PTB model correctly predicts conversion outcome of most overexpression and knockdown experiments. **(a)** Network topology of the nPTB → PTB model. **(b,c)** The deterministic (red) and stochastic (green) simulation outcomes are shown for OX/KD scenarios in the nPTB → PTB model. Similar simulations for all models and experiments are summarized in Table 2 and are shown in Supplementary Fig. 1. The plots were made with Mathematica 11.3.0 (Linux version). <https://www.wolfram.com/mathematica/>.

the training data cover only a narrow set of circumstances compared to the many perturbation experiments used as validation. The fact that a computational model could learn the outcome of one experiment from training data based exclusively on another experiment suggests that they operate in the same region of the system phase space, and that these different conversion strategies invoke a comparable response from the GRN. Notably, the exception to this rule is the system response to PTB knockdown (which can cause neuronal differentiation¹⁶) or overexpression (which is known to block neuronal differentiation with Ascl1 overexpression²⁶). This suggests that neuronal conversion induced by PTB knockdown is affected by a different cascade of gene interactions compared to the other conversion strategies. Based on the gene interaction network, one way in which the PTB knockdown strategy stands out is that it affects REST (and through it the rest of the network) by removing a crucial activator (PTB), whereas other conversion strategies directly or indirectly induce miR-124-9/9*, a repressor of the REST complex.

We sought to understand this in more detail by applying the model dissection shown in Fig. 3d to relevant overexpression and knockdown scenarios for the nPTB → PTB model. Simulated PTB knockdown affected the model by an increase of nPTB, but with no noticeable change in the expression of other factors (Fig. 6a, left). While different from the expectation that PTB knockdown results in neuronal conversion (Table 1), this outcome has also been found experimentally²². In our model, this diminished system response is caused by the fact that RESTc expression remained activated constitutively, maintaining nominal concentrations, even when RESTc was no longer activated by PTB (Fig. 6b, left). Thus, PTB knockdown does not affect RESTc concentrations and therefore its effects do not cascade through the rest of the GRN. Interestingly, when the constitutive activation of REST was reduced, the system became permissible to neuronal conversion by PTB knockdown (Fig. 6a,b, middle). In the absence of PTB knockdown, this lower constitutive RESTc expression did display different system

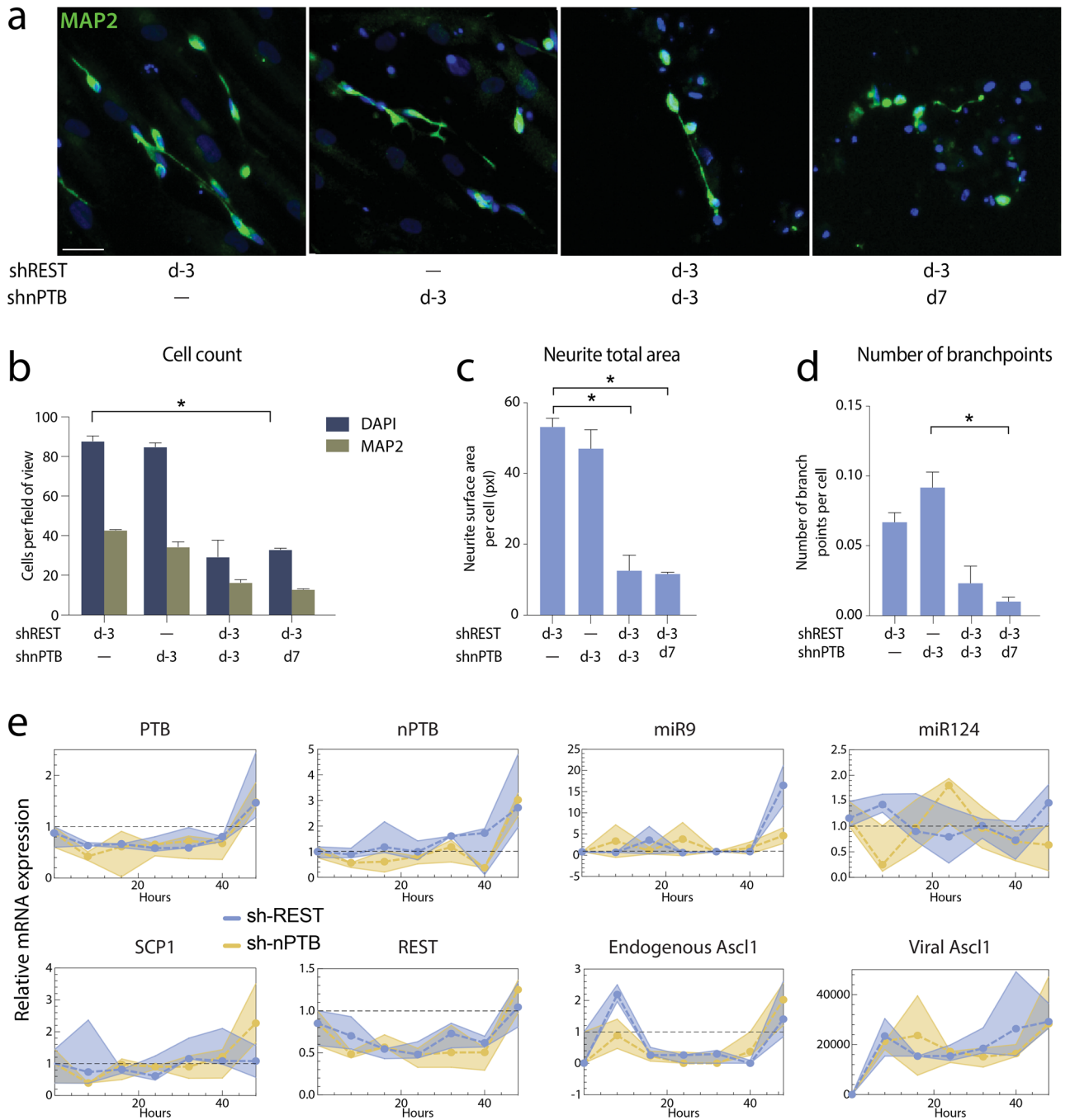


Figure 5. Experimental validation of the nPTB → PTB model. **(a)** Immunofluorescence staining of MAP2 (in green) showing reprogrammed iNs at day 25 post transduction. Cells are counterstained with DAPI (in blue). Scale bar = 150 μm. **(b)** Quantification of DAPI+ and MAP2+ cells at day 25 post transduction, error bars are defined as S.E.M. **(c,d)** Quantification of neurite total area **(c)** and number of branch points, error bars are defined as S.E.M. **(d)** in iNs at day 25 post transduction. **(e)** Measured transcription levels. The dashed line represents the median of three replicates, surrounded by a shaded region between the minimum and maximum values at each time point. The measurements are normalized so that a value of 1 is equal to the fibroblast level, or to the detection limit in the case of Ascl1, which was not detected in fibroblasts. The graphs were generated using Graphpad Prism 8, the fluorescent images were put together using Adobe Photoshop 2020, the figures were ensemble using Adobe Illustrator 2020. <https://www.graphpad.com/scientific-software/prism/>, <https://www.adobe.com/se/products/illustrator.html>.

behaviour compared to the unmodified parameter set (Fig. 6a,b, right). Moreover, our experimental data in Fig. 5 showed that nPTB KD results in neural conversion therefore an nPTB KD could lead to depletion of the constitutive REST expression. Thus, the constitutive RESTc activation is largely a free parameter when fitting to

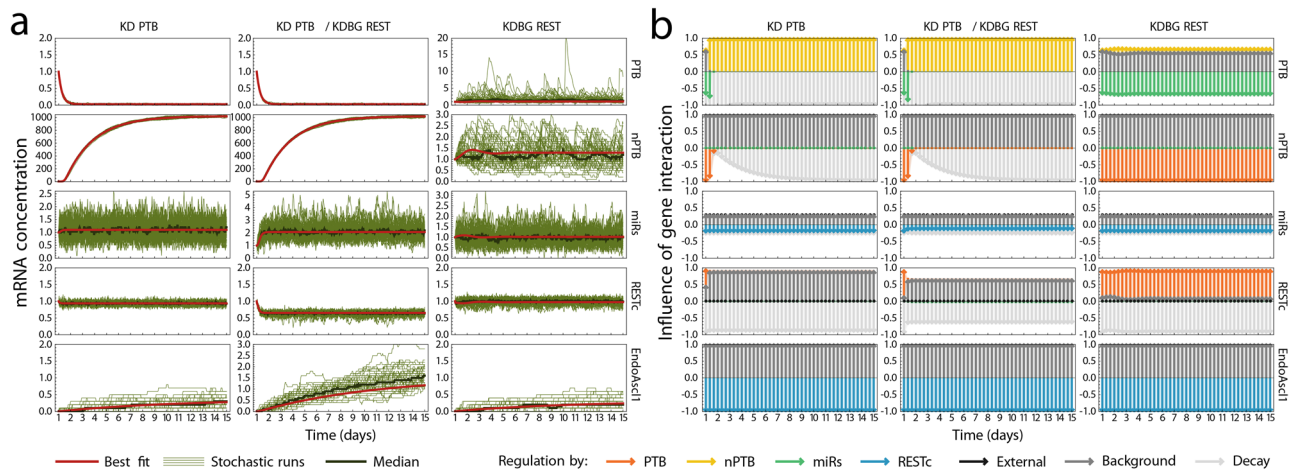


Figure 6. Strong constitutive activation of REST expression blocks neuronal reprogramming through PTB knockdown in models. **(a)** Stochastic (green) and deterministic (red) time evolution of the system. Left: simulated knockdown of PTB. The system remains in non-neuronal state, as seen by the lack of Ascl1 expression. Middle: combining PTB knockdown with a diminished constitutive REST activation induces Ascl1 expression. Right: with regular PTB expression, reducing constitutive REST activation results in a stable non-neuronal profile, with a steady state that strongly resembles baseline fibroblast levels. **(b)** The influence of different regulatory interactions on the expression of each factor during the deterministic experiments shown in panel (a) (see also Fig. 3d). See Methods for an extended description. The plots were made with Mathematica 11.3.0 (Linux version). <https://www.wolfram.com/mathematica/>.

our data. Note that constitutive here refers to any regulation by factors that are not explicitly present in our GRN. It does not imply that this regulation is identical for every cell, or independent of the cell's state and environment outside the range of our experiments. These results show that our data and model predict that some cells are more susceptible to neuronal conversion through PTB knockdown than others, due to variation in constitutive expression of members of the REST complex.

Discussion

In this study, we put forward a quantitative model of the core GRN governing direct reprogramming of human adult fibroblasts to neurons. Our analysis showed that a straightforward interpretation of the known interactions between key transcription factors was unable to explain the full dynamic behaviour that we observed. In particular, transcription data gathered during a conversion experiment showed evidence of a negative feedback loop in the GRN, which has not previously been described.

Adding a hypothetical interaction to the network where nPTB activates PTB introduced such a negative feedback loop and drastically increased the match between measured mRNA expression levels and those predicted by the model. In addition, in contrast to various alternative hypotheses, this model correctly predicted the response of the cell in a range of different overexpression and knockdown scenarios. Taken together, these findings suggest that such an interaction between these two genes may play a role in neuronal conversion. In support to this, we show that modulating this interaction through a nPTB repression results into successful neuronal reprogramming. However, the possibilities for how this interaction is manifested biochemically should be interpreted broadly. The model is agnostic of specific regulatory mechanisms, so the cause of the predicted interaction is not necessarily restricted to the transcriptional level and may include mediation by one or more intermediate regulators.

Interestingly, there is a possible mechanism by which nPTB may function as an activator of PTB, in line with known features of PTB and nPTB regulation. Both PTB and nPTB are known to regulate the splicing of PTB mRNA, effectively inhibiting its expression by inducing nonsense-mediated decay^{21–23,27}. Furthermore, it has been shown that PTB and nPTB have similar RNA-binding properties, however, nPTB is a weaker splicing regulator than PTB due to a difference in recruitment of cofactors^{21,22,27}. Therefore, whether the net effect of nPTB on the correct splicing of PTB is positive or negative depends on the present concentration of PTB. If the concentration of PTB is low, nPTB simply acts as a splicing regulator and decreases PTB expression. On the other hand, if the concentration of PTB is high, binding of nPTB to the PTB mRNA is likely to displace PTB as a splicing regulator, and cause a net increase in correctly spliced RNA due to its relatively weaker activity. It is possible that this mechanism explains our findings, wherein nPTB is best modelled and understood as an activator (that is, increasing the expression) of PTB in the relevant cellular conditions. Interpreted this way, our computational analysis suggests that known interactions between key regulators are enough to explain much of the cellular decision-making process during direct neuronal reprogramming, although competition between PTB and nPTB for the regulation of PTB plays a more important role than expected. Note that our quantitative approach does not distinguish between weaker or stronger regulators with the same binding strength, so that explicitly including repression of PTB by itself and nPTB did not improve our models.

When investigating the response of the different models to simulated overexpression and knockdown experiments, we found a remarkable ability to predict experimental outcomes. The fact that information gained from only one experiment generalises to other conversion strategies suggests that the different methods operate via the same interaction mechanisms and guide the cell to similar trajectories in the phase space of the system. One consequence of this insight is that we may expect greater conversion efficiencies when compatible strategies are combined. Indeed, several publications have combined REST knockdown, *Ascl1* overexpression and miR-124-9/9* overexpression, each of which can induce conversion on its own, with the effect of increasing conversion efficiency and neuronal maturation^{6,10,11,28,29}. On the other hand, lower conversion efficiencies can be observed when incompatible strategies are combined, as we show in this study with the double repression of REST and nPTB.

Our analysis did not predict that the conversion strategy based on PTB knockdown would result in cells with a neuron-like identity, but instead showed an increase in nPTB expression without affecting other genes. Interestingly, although PTB knockdown has been used successfully in a number of conversion experiments^{16,30}, this negative experimental outcome has also been observed in the lab²². We provide a speculative but plausible explanation for this variation in results by observing that, in our best model, the conversion was blocked by constitutive activation of REST. According to the model, cells with a lower constitutive REST activation would have similar REST expression levels, but would be more susceptible to conversion after PTB depletion. We also showed experimentally that nPTB KD leads to successful production of iNs which confirms that nPTB feeds back in the core gene regulatory network controlling direct neural conversion. Moreover, nPTB KD reduced the amount of REST in the cells suggesting that diminished nPTB might be able to block the constitutive REST activation.

Overall, this work illustrates that the current understanding of the genes and interactions governing neural reprogramming can be unified into a computational framework that explains much of the observed behaviour of the cell. Our analysis brings a new perspective to some aspects of the cellular decision process that establishes neuronal fate, and demonstrates a novel analytical tool to dissect internal model dynamics. In the future, more detailed quantitative models informed by data from varied experiments can increase the power of this approach, and single-cell data would improve the ability to make predictions regarding heterogeneity and conversion efficiency. This and future models will help to provide a better understanding of the existing conversion strategies, and suggest improvements that could result in higher yield or faster conversion to neurons.

Methods

Experimental data. *Cell culture and cell lines.* Adult dermal fibroblasts of a 61-, 67- and 75-year-old healthy female donors were obtained with written, informed consent from the Parkinson's Disease Research clinic at the John van Geest Centre for Brain Repair (Cambridge, UK). All experimental protocols were approved by the East of Cambridgeshire Research Ethics Committee (REC 09/H0311/88) as well as the University of Montreal *Comité d'éthique de la recherche en sciences et en santé* (CERSES-18-004-D). All experiments were performed in accordance with the Declaration of Helsinki.

For details on the skin biopsy sampling method, please refer to⁶. Primary fibroblasts were expanded and cultured at 37 °C in 5% CO₂ in fibroblast medium (Dulbecco's Modified Eagle Medium (DMEM) + Glutamax (Gibco) with 100 mg/mL penicillin/streptomycin (Sigma), and 10% FBS (Biosera)). The cells were then dissociated with 0.05% trypsin, spun, and frozen in 50/50 DMEM/FBS with 10% DMSO (Sigma).

Viral vectors and virus transduction. DNA plasmids expressing mouse open reading frames (ORFs) for *Ascl1* and *Brn2* on the same construct (pB.pA, see Ref.⁶) in a third-generation lentiviral vector containing a non-regulated ubiquitous phosphoglycerate kinase (PGK) promoter were used (Fig. 1a). The knockdown of REST and nPTB were done using short hairpin RNAs^{6,31}. The shRNAs plasmid DNA for nPTB was purchased from Sigma-Aldrich. All the constructs have been verified by sequencing. Lentiviral vectors were produced and titrated as previously described⁶. Transduction was performed at a MOI of 10 for the pB.pA vector and at an MOI of 5 for the shRNA vectors against REST. 0.5 µl of the sh-nPTB lentiviral vector was added to the cells, which provided a 50% downregulation on day 3 (Supplementary Fig. 2).

Neuronal reprogramming. For direct neuronal reprogramming, fibroblasts were plated at a density of 26,300 cells per cm² in 24-well plates (Nunc) coated with 0.1% gelatin (Sigma). On the next day, cells were transduced with shRNAs against REST or nPTB and the medium was changed the following day. Three days after this viral transduction, cells were transduced with the pB.pA vector. Fibroblast medium was replaced by neural differentiation medium (NDiff227; Takara-Clontech) supplemented with growth factors at the following concentrations: LM-22A4 (2 µM, R&D Systems), GDNF (2 ng/mL, R&D Systems), NT3 (10 ng/mL, R&D Systems) and the small molecules CHIR99021 (2 µM, Axon), SB-431542 (10 µM, Axon), noggin (0.5 µg/mL, R&D Systems), LDN-193189 (0.5 µM, Axon), as well as valproic acid sodium salt (VPA; 1 mM, Merck Millipore) and db-cAMP (0.5 mM, Sigma), 3 days after the second transduction. Half of the neuronal conversion medium was replaced every 2–3 days. 18 days post-transduction, the small molecules were stopped, and the neuronal medium was supplemented with only the factors LM-22A4, GDNF, NT3 and db-cAMP until the end of the experiment.

qRT-PCR analysis. Total RNA, including miRNA, was extracted from cells from the same line at different stages of conversion (Fig. 1a) using the micro miRNeasy kit (Qiagen) followed by Universal cDNA synthesis kit (Fermentas, for RNA analysis; Exiqon or a miRCURY LNA kit for miRNA expression). Three reference genes were used for each qPCR analysis (ACTB, GAPDH and YWHAZ). Primer sequences can be found in Supplementary Table 1. LNA-PCR primer sets, specific for hsa-miR-9-5p, hsa-miR-124-3p and hsa-miR-103 (the latter used as normalization miRNA), were purchased from Exiqon and used for the miRNA qPCR analysis. All primers were

used together with LightCycler 480 SYBR Green I Master (Roche). Standard procedures of qRT-PCR were used, and data were quantified using the $\Delta\Delta\text{Ct}$ -method. Analyses were performed in triplicates at each time point.

Immunocytochemistry, quantification, and imaging. Cells were fixed on day 25 in 4% paraformaldehyde, permeabilized with 0.1% Triton-X-100 in 0.1 M PBS for 10 min. Thereafter, cells were blocked for 30 min in a solution containing 5% normal serum in 0.1 M PBS. The following primary antibodies were diluted in the blocking solution and applied overnight at 4 °C: mouse anti-TAU clone HT7 (1:500, Thermo Scientific), chicken anti-MAP2 (1:10,000, Abcam). Fluorophore-conjugated secondary antibody (Jackson ImmunoResearch Laboratories) was diluted in blocking solution and applied for 2 h. Cells were counterstained with DAPI for 15 min followed by three washes in PBS. The total number of DAPI⁺, TAU⁺ and MAP2⁺ cells per well were quantified and imaged using the Cellomics Array Scan (Array Scan VTI, Thermo Fischer), which is an automated process ensuring unbiased measurements between groups. Applying the programs “Target Activation” and “Neuronal profiling”, 289 fields (10X magnification) were acquired in a spiral fashion starting from the center.

Building quantitative models from GRNs. The Shea-Ackers formalism. The Shea-Ackers formalism is a mathematical model that describes the activity of a regulatory sequence as a function of the concentrations of activating and inhibiting factors²⁰. Assuming that the regulated gene is active when the binding sequence is occupied by an activator ($\in act$), and inactive when it is unoccupied or in the presence of an inhibitor ($\in inh$), the formalism borrows techniques from statistical mechanics to compute the average activity at the site assuming thermodynamic equilibrium:

$$SA(X) = \frac{\sum_{T \in act} \left(\frac{[T]}{k_{T,X}} \right)^{h_{T,X}}}{1 + \sum_{T \in act} \left(\frac{[T]}{k_{T,X}} \right)^{h_{T,X}} + \sum_{T \in inh} \left(\frac{[T]}{k_{T,X}} \right)^{h_{T,X}}}, \quad (1)$$

where $k_{T,X}$ are dissociation constants describing the interaction strength between factor T and the regulatory sequence of X , and $h_{X,T}$ is a non-linearity parameter representing cooperativity between multiple factors of the same type.

We use this equation to build a system of ordinary differential equations that describes the time evolution of concentrations of the relevant regulators X . Assume that the expression rate of X is proportional to the activity of the regulatory site (defined by Eq. 1) as the average fraction of time the site is occupied by an activator). Assume also that each gene product undergoes exponential decay, that is, its concentration decreases at a rate proportional to itself. Then the change in gene product concentration over time can be described as

$$\frac{d[X]}{dt} = \alpha_X SA(X) - \delta_X [X], \quad (2)$$

where α_X and δ_X are scaling coefficients for the rate of transcription and decay, respectively (decay half-life is $\ln(2)/\delta_X$).

This approach requires that Eq. (2), for each node X in the network, depends only on other variables that are explicitly modelled in the GRN. In practice, most genes are influenced by other genes that are not present in the network because they are not considered relevant in the context of the GRN model. The influence of these interactions on the master equation of the quantitative model (Eq. 2) can be greatly simplified with a change of variables.

Considering Eq. (1), each constitutive inhibitor and activator T will be represented by a term $([T]/k_{T,X})^{h_{T,X}}$ in the rate of expression $\alpha_X SA(X)$. Assuming that the concentration $[T]$ is constant in the relevant conditions for the model, this term can be simplified to the constant value $\beta_{T,X} = ([T]/k_{T,X})^{h_{T,X}}$. Furthermore, the combined influence of all constitutive activators is also constant $\beta_{act,X} = \sum_{T \in act} \beta_{T,X}$, and for the constitutive inhibitors we have that their sum is equal to the constant $\beta_{inh,X} = \sum_{T \in inh} \beta_{T,X}$. This leads to the equation

$$SA(X) = \frac{\beta_{act,X} + \sum_{T \in act'} \left(\frac{[T]}{k_{T,X}} \right)^{h_{T,X}}}{1 + \beta_{act,X} + \beta_{inh,X} + \sum_{T \in act'} \left(\frac{[T]}{k_{T,X}} \right)^{h_{T,X}} + \sum_{T \in inh'} \left(\frac{[T]}{k_{T,X}} \right)^{h_{T,X}}},$$

where act' and inh' are the sets of non-constitutive activators and inhibitors. By dividing the numerator and denominator by $1 + \beta_{inh,X}$ and defining $\beta_X = \beta_{act,X}/(1 + \beta_{inh,X})$, and $k'_{T,X} = k_{T,X}(1 + \beta_{inh,X})^{h_{T,X}}$, the parameter set of the equation can be further reduced:

$$SA(X) = \frac{\beta_X + \sum_{T \in act'} \left(\frac{[T]}{k'_{T,X}} \right)^{h_{T,X}}}{1 + \beta_X + \sum_{T \in act'} \left(\frac{[T]}{k'_{T,X}} \right)^{h_{T,X}} + \sum_{T \in inh'} \left(\frac{[T]}{k'_{T,X}} \right)^{h_{T,X}}}. \quad (3)$$

For nodes which are not activated by non-constitutive factors, the background parameter β_X is also redundant. By dividing the numerator and denominator by $1 + \beta_X$ and defining $\alpha'_X = \alpha_X \beta_X / (1 + \beta_X)$ and $k''_{T,X} = k'_{T,X} / (1 + \beta_X)^{h_{T,X}}$ we obtain the expression

$$\alpha'_X \frac{1}{1 + \sum_{T \in \text{inh}'} \left(\frac{[T]}{k_{T,X}} \right)^{h_{T,X}}} \tag{4}$$

Henceforth we will not write the prime ' on these transformed variables.

The Shea–Ackers formalism is derived with the assumption that all activations and inhibitions are physically realised by binding competition on a gene’s regulatory sequence. The rate equations above assume that there is transcription at a constant rate whenever an activator is bound, thus ignoring differences in the rate at which transcription factors recruit the transcription machinery. Overlooking the complexity of different regulatory mechanisms allows for a mathematical approximation that can describe a wide range of dynamic behaviours that are realistic for different gene interaction mechanisms while using only a small number of parameters.

Viral inhibition and activation. For REST and miRs, formula (3) or (4) must also take into account the externally added factors shREST and viral Ascl1. The direct effect of shREST on the system is inhibition of REST. According to the GRN we use as a basis for our models, the effect of viral Ascl1 is the activation of miRs. Therefore, we could model the influence of these factors as a constant inhibitor acting on REST ($\beta_{\text{REST}i}$) and a constant activator acting on miRs (β_{vAscl1}). These values are parameters of the model and are added to the denominator of SA(REST), or the numerator and denominator of SA(miRs), respectively.

These parameters $\beta_{\text{REST}i}$ and β_{vAscl1} can be added or removed from the model depending on whether the external influence they represent is present. Thus, the same model equations can represent the three stages of the experiment: (a) the fibroblast stage (i.e. $\beta_{\text{REST}i} = \beta_{\text{vAscl1}} = 0$), (b) the three days of REST inhibition ($\beta_{\text{REST}i} \neq 0$) and (c) the conversion stage, where both parameters are non-zero. The value of the parameters when not fixed to zero is trained to the data in the same way as other model parameters.

Note that viral influences are modelled as constant with respect to time, because the measured values are at least 10^5 times greater than the baseline value for fibroblasts during the time frame under consideration, and the system behaviour shown in Fig. 1c does not suggest a fluctuating regulatory influence of Ascl1.

Model equations. Following this procedure to construct one ordinary differential equation (ODE) for each factor in the network yields an ODE system that can be numerically integrated to yield a time evolution of concentration levels $f_{p,c_0}^X(t)$ for each factor X , given initial concentration values $c_0 = f_{p,c_0}^X(0)$ and an appropriate parameter set $p = \{\alpha_X, \delta_X, \beta_X\}_X \cup \{k_{T,X}, h_{T,X}\}_{T,X} \cup \{\beta_{\text{REST}i}, \beta_{\text{vAscl1}}\}$. For the initial literature-based model, the full ODE system is given in Eq. (5).

$$\begin{cases} \frac{d}{dt} [P] = \alpha_P \frac{1}{1 + ([M]/k_{M,P})^{h_{M,P}}} - \delta_P [P] \\ \frac{d}{dt} [N] = \alpha_N \frac{1}{1 + ([M]/k_{M,N})^{h_{M,N}} + ([P]/k_{P,N})^{h_{P,N}}} - \delta_N [N] \\ \frac{d}{dt} [M] = \alpha_M \frac{\beta_M + \beta_{\text{vAscl1}}}{1 + \beta_M + \beta_{\text{vAscl1}} + ([R]/k_{R,M})^{h_{R,M}}} - \delta_M [M] \\ \frac{d}{dt} [R] = \alpha_R \frac{\beta_R + ([P]/k_{P,R})^{h_{P,R}}}{1 + \beta_R + ([P]/k_{P,R})^{h_{P,R}} + ([M]/k_{M,R})^{h_{M,R}} + \beta_{\text{REST}i}} - \delta_R [R] \\ \frac{d}{dt} [A] = \alpha_A \frac{1}{1 + ([R]/k_{R,A})^{h_{R,A}}} - \delta_A [A] \end{cases} \tag{5}$$

where each node in the GRN is indicated by its initial letter for brevity. For the nPTB → PTB model, we have for the first (PTB) equation:

$$\frac{d}{dt} [P] = \alpha_P \frac{\beta_P + ([N]/k_{N,P})^{h_{N,P}}}{1 + \beta_P + ([M]/k_{M,P})^{h_{M,P}} + ([N]/k_{N,P})^{h_{N,P}}} - \delta_P [P] \tag{6}$$

and an identical formula for the other equations, though with different parameter values after training. Trained parameter values are shown in Supplementary Table 2 for the literature model and nPTB → PTB model. These and all other models presented in this work are given in SBML format in the online supplementary.

As detailed above, these equations can be modified by removing the terms $\beta_{\text{REST}i}$ and β_{vAscl1} as necessary to represent the presence or absence of these external factors. Initial values for our simulations are given in Supplementary Table 3.

Genetic algorithm for parameter fitting. An ODE system is considered a good match for the data if there is some parameter set p that leads to predicted expression curves f_{p,c_0} that closely match observations. Finding such a parameter set is an optimization problem. We obtained the best results using a real-valued genetic algorithm, evolving a vector of real numbers which each represent a system parameter. The algorithm used a population size of 100, uniform crossover, and a mutation operator that modifies a small number of genes by multiplying with log-normally distributed random variables. Specifically, if $m(p, \sigma)$ represents a random distribution that has a probability $(1 - p)$ of being 1 and a probability p of being distributed as $\exp(\mathcal{N}(0, \sigma))$, then mutation was implemented by multiplying each variable with independent random values from the distribution $m(2/n, 1.01) \times m(2/3n, 1.05) \times m(1/5n, 1.5)$, where n is the number of genes. The initial population was composed of individuals with genes chosen uniformly from the unit interval, but more complex models were also run using parameters from simpler models as initial values.

Each real-valued gene represents a system parameter. Hill parameters $h_{T,X}$ were constrained to a maximum value of 4 by transforming the gene value $h'_{T,X}$ to $h_{T,X} = 4 \left(\frac{h'_{T,X}}{1 + h'_{T,X}} \right)$, where $h_{T,X}$ is the parameter value used in the system equations. The other parameters were used directly as they are in the real-valued vector genome. This means they are unbounded (though the mutation operator will never produce non-positive values).

Objective function for parameter fitting. The optimisation procedure is designed to yield ‘good’ parameter sets, according to a pre-defined measure. To quantify this, we used the parameter set to integrate the model equations as specified above in “Model equations”, and compare the outcome to data using a standard sum of squared errors (SSE) with smoothness regularisation. Specifically, the model equations were integrated three times, representing the fibroblast (denote the resulting curve with $f_{p,\text{fibro}}$), REST inhibition ($f_{p,\text{REST}i}$), and conversion ($f_{p,\text{conv}}$) stages of the experiment. Each curve has its own set of initial values and slightly modified parameter set (see Viral inhibition and activation and Supplementary Table 3).

The objective function, which measures the quality of a parameter set p , consisted of several terms. The SSE term measures the difference between the system solution $f_{p,\text{conv}}$ and our observed data points during the first five days of conversion termed X in the following equation. The data values used here are the median values of the triplicate observations in each of the six time points. For the fibroblast stage, no error is included in this term; for the REST inhibition stage, the final value after three days is compared to the observed value; and for the conversion stage, all data points in the first five days are compared to the simulated values.

$$SSE = \sum_X \left(\left(f_{p,\text{REST}i}^X(3 \text{ days}) - [X]_0 \right)^2 + \sum_{t \in \{8 \text{ hours}, 1, 2, 3, 5 \text{ days}\}} \left(f_{p,\text{conv}}^X(t) - [X]_t \right)^2 \right). \quad (7)$$

A pure SSE approach often resulted in highly oscillating fits, which closely matched the data but fluctuated erratically in between. To obtain more realistic solutions, we added a smoothness regularisation term of the form:

$$R_1 = \lambda \sum_X \int_{5 \text{ days}} \left| \frac{d^2}{dt^2} f_{p,\text{conv}}^X(t) \right| dt. \quad (8)$$

Finally, an important property of the system being modelled is that fibroblasts do not spontaneously convert to neurons, i.e. the mRNA concentrations should remain stable when initialized near the observed fibroblast baseline and when not perturbed by the external terms $\beta_{\text{REST}i}$ and β_{vAscl1} . This is enforced with another term in the fitness function, punishing any changes in concentration during a three-day simulation of the fibro stage.

$$R_2 = \mu \sum_X \int_{3 \text{ days}} \left| \frac{d}{dt} f_{p,\text{fibro}}^X(t) \right| dt. \quad (9)$$

We obtained best results with $\lambda = \frac{1}{20}$ and $\mu = \frac{1}{20}$. In the evolutionary algorithm, parameter sets with a higher fitness are considered better, so the fitness is set equal to the negative value $-(SSE + R_1 + R_2)$.

Two nodes in our networks represent a combination of more than one gene for which expression was measured. We fitted the miRs network node to transcription data from miR-124, and the RESTc network node to data from the REST gene. This arbitrary choice did not influence the results due to the high correlation between the concentration values (this was verified by using mRNA concentration levels of miR-9/9* and SCP1 with identical results).

Deterministic simulation. Integration of the ODE system was done using a custom implementation of the fourth-order Runge–Kutta algorithm during training, with a time step size of 0.05 h. Trained models were integrated and plotted using Mathematica. Solutions were verified to be identical using both methods.

Stochastic simulation. Stochastic simulations were performed using a custom implementation of the Gillespie algorithm²⁵. Noise levels were adjusted by modifying the reaction formulas: the stoichiometric coefficients were multiplied by a noise level η , and reaction rates correspondingly divided by η to maintain the same expected number of molecules added or removed per time unit. Our data is a bulk average and therefore does not suggest a realistic value for η . In the figures shown here (Figs. 4b,c, 6a), its value was set to 0.1. However, two adjustments were made. First, PTB and REST have a (relative) concentration range close to 1, whereas the other factors have typical concentration levels closer to 5. For this reason, transcription and decay reactions for these two factors were transformed with an η that was five times lower. Second, these same factors also tend to fit with much higher reaction rates, i.e. both higher α and δ in (2). To obtain more similar noise levels, η for these reactions was adjusted by an additional factor of 1/5, resulting in a total of $\eta = 0.004$ for PTB and REST, and $\eta = 0.1$ for nPTB, miRs and Ascl1.

Overexpression and knockdown were simulated by modifying the appropriate transcription rate coefficient. Specifically, for the overexpression, resp. knockdown, of a factor X , the corresponding transcription rate coefficient α_X was multiplied by 5, resp. 1/5. Overexpression of Ascl1 and knockdown of miRs were instead controlled by multiplying the values $\beta_{\text{REST}i}$ and β_{vAscl1} with 0 or 1, in the same way as when training the models. In the case of PTB knockdown, lack of conversion was also verified when α_{PTB} was multiplied by $\times 1/5000$, and at this rate conversion was observed when the background activation of REST was multiplied by 1/5.

Overexpression and knockdown simulations. A simulation was considered a neuronal conversion when endogenous expression of *Ascl1* was reached at least 1 (in units corresponding to the detection limit of the qPCR experiment) at some point during the simulation. This was taken to indicate that the system switched to a different state that could be self-sustaining when the externally induced viral expression of *Ascl1* would be removed. An OX/KD scenario was considered to cause neuronal conversion when at least one of fifty performed simulations showed these criteria. These thresholds are arbitrary, but the result that the nPTB \rightarrow PTB model is the best model to predict experimental outcomes is robust.

Measuring relative influence of interactions. In order to break down the total transcription rate of each factor into contributions from its activators and inhibitors, the terms in the numerator and denominator of the transcription rate Eq. (1) were calculated at each time point of a trajectory and plotted, as in Figs. 3d and 6b. Activators and inhibitors are represented by arrows pointing upwards and downwards, respectively, with arrow length showing the relative magnitude of that interaction term.

Specifically, the total length of all activation arrows L_X^{act} of a gene X , at some time point t , equals the fraction of the maximum transcription rate that would be realized if no inhibitors were present. The total length of the inhibition arrows L_X^{inh} is defined so that the difference between the activation and the inhibition arrows equals the realized transcription activity $SA(X)$ (Eq. 1). This results in the following equations:

$$\begin{cases} L_X^{act} = \frac{\Sigma_{act}}{1 + \Sigma_{act}} \\ L_X^{inh} = \frac{\Sigma_{act} \Sigma_{inh}}{(1 + \Sigma_{act})(1 + \Sigma_{act} + \Sigma_{inh})}, \end{cases} \quad (10)$$

where $\Sigma_{act} = \sum_{T \in act} ([T]/k_{T,X})^{h_{T,X}}$ and $\Sigma_{inh} = \sum_{T \in inh} ([T]/k_{T,X})^{h_{T,X}}$ (see Eq. 1). Given these total lengths for activator and inhibitor arrows, individual activators and inhibitors T are given lengths proportional to their current magnitude $([T]/k_{T,X})^{h_{T,X}}$, so that the contribution of each inhibitor or activator to the total transcription rate becomes visible.

Finally, an arrow is added to represent the decay term in the Eq. (2), with a length of $\frac{\partial X}{\partial t} [X]$ (essentially, the decay rate in the same units of α_X as the other arrows). By adding this additional arrow, the difference between the sum of upward and downward pointing arrows shows the rate of change in mRNA concentrations (Eq. 2 divided by α_X).

Code availability

The SBML implementation of the models is available at: <https://github.com/victorahnell/Direct-Neural-Conversion-Models-SBML/blob/main/code.zip>.

Received: 8 January 2020; Accepted: 1 January 2021

Published online: 15 January 2021

References

- Huh, C. J. *et al.* Maintenance of age in human neurons generated by microRNA-based neuronal conversion of fibroblasts. *eLife* **5**, e18648 (2016).
- Kim, Y. *et al.* Mitochondrial aging defects emerge in directly reprogrammed human neurons due to their metabolic profile. *Cell Rep.* **23**(9), 2550–2558 (2018).
- Mertens, J. *et al.* Directly reprogrammed human neurons retain aging-associated transcriptomic signatures and reveal age-related nucleocytoplasmic defects. *Cell Stem Cell* **17**(6), 705–718 (2015).
- Tang, Y. *et al.* Direct reprogramming rather than iPSC-based reprogramming maintains aging hallmarks in human motor neurons. *Front. Mol. Neurosci.* **10**, 359 (2017).
- Rackham, O. J. *et al.* A predictive computational framework for direct reprogramming between human cell types. *Nat. Gen.* **48**(3), 331–335 (2016).
- Drouin-Ouellet, J. *et al.* REST suppression mediates neural conversion of adult human fibroblasts via microRNA-dependent and -independent pathways. *EMBO Mol. Med.* **9**(8), 1117–1131 (2017).
- Vierbuchen, T. *et al.* Direct conversion of fibroblasts to functional neurons by defined factors. *Nature* **463**(7284), 1035–1041 (2010).
- Pfisterer, U. *et al.* Direct conversion of human fibroblasts to dopaminergic neurons. *PNAS* **108**(25), 10343–10348 (2011).
- Gopalakrishnan, S., Hor, P. & Ichida, J. K. New approaches for direct conversion of patient fibroblasts into neural cells. *Brain Res.* **1656**, 2–13 (2017).
- Birtele, M. *et al.* Dual modulation of neuron-specific microRNAs and the REST complex promotes functional maturation of human adult induced neurons. *FEBS Lett.* (2019)
- Masserdotti, G. *et al.* Transcriptional mechanisms of proneural factors and REST in regulating neuronal reprogramming of astrocytes. *Cell Stem Cell* **17**(1), 74–88 (2015).
- Xue, Y. *et al.* Direct conversion of fibroblasts to neurons by reprogramming PTB-regulated microRNA circuits. *Cell* **152**(1–2), 82–96 (2013).
- Hu, J. *et al.* PTB/nPTB: Master regulators of neuronal fate in mammals. *Biophys. Rep.* **4**(4), 204–214 (2018).
- Olariu, V. & Peterson, C. Kinetic models of hematopoietic differentiation. *Wiley Interdiscip. Rev. Syst. Biol. Med.* **11**(1), e1424 (2018).
- Herberg, M. & Roeder, I. Computational modelling of embryonic stem-cell fate control. *Development* **142**(13), 2250–2260 (2015).
- Papatsenko, D., Lemischka, I. R. & Wassarman, P. M. Emerging modeling concepts and solutions in stem cell research. *Curr. Top. Dev. Biol.* **116**, 709–721 (2016).
- Kim, J., Ambasudhan, R. & Ding, S. Direct lineage reprogramming to neural cells. *Cur. Opin. Neurobiol.* **22**(5), 778–784 (2012).
- Abernathy, D. G. *et al.* MicroRNAs induce a permissive chromatin environment that enables neuronal subtype-specific reprogramming of adult human fibroblasts. *Cell Stem Cell* **21**(3), 332–348 (2017).
- Xue, Y. *et al.* Sequential regulatory loops as key gatekeepers for neuronal reprogramming in human cells. *Nat. Neurosci.* **19**(6), 807–815 (2016).
- Ackers, G. K., Johnson, A. D. & Shea, M. A. Quantitative model for gene regulation by lambda phage repressor. *PNAS* **79**(4), 1129–1133 (1982).

21. Keppetipola, N. *et al.* Neuronal regulation of pre-mRNA splicing by polypyrimidine tract binding proteins, PTBP1 and PTBP2. *Crit. Rev. Biochem. Mol. Biol.* **47**(4), 360–378 (2012).
22. Spellman, R., Llorian, M. & Smith, C. W. Crossregulation and functional redundancy between the splicing regulator PTB and its paralogs nPTB and ROD1. *Mol. Cell.* **27**(3), 420–434 (2007).
23. Wollerton, M. C. *et al.* Autoregulation of polypyrimidine tract binding protein by alternative splicing leading to nonsense-mediated decay. *Mol. Cell.* **13**(1), 91–100 (2004).
24. Raj, B. *et al.* Cross-Regulation between an alternative splicing activator and a transcription repressor controls neurogenesis. *Mol. Cell.* **43**(5), 843–850 (2011).
25. Gillespie, D. T. Exact stochastic simulation of coupled chemical reactions. *J. Phys. Chem.* **81**(25), 2340–2361 (1977).
26. Yoo, A. S. *et al.* MicroRNA-mediated conversion of human fibroblasts to neurons. *Nature* **476**(7359), 228–231 (2011).
27. Makeyev, E. V. *et al.* The MicroRNA miR-124 promotes neuronal differentiation by triggering brain-specific alternative pre-mRNA splicing. *Mol. Cell.* **27**(3), 435–448 (2007).
28. Lee, S. W. *et al.* MicroRNAs overcome cell fate barrier by reducing EZH2-controlled REST stability during neuronal conversion of human adult fibroblasts. *Dev. Cell.* **46**(1), 73–84.e7 (2018).
29. Wohl, S. G. & Reh, T. A. miR-124-9/9* potentiates Ascl1-induced reprogramming of cultured Müller glia. *Glia* **64**(5), 742–762 (2016).
30. Drouin-Ouellet, J. *et al.* Direct neuronal reprogramming for disease modeling studies using patient-derived neurons: What have we learned?. *Front. Neurosci.* **11**, 530 (2017).
31. Shrigley, S. *et al.* Simple generation of a high yield culture of induced neurons from human adult skin fibroblasts. *J. Vis. Exp.* **132**, e56904 (2018).
32. Bruce, A. W. *et al.* Genome-wide analysis of repressor element 1 silencing transcription factor/neuron-restrictive silencing factor (REST/NRSF) target genes. *PNAS* **101**(28), 10458–10463 (2004).
33. Schoenherr, C. J. & Anderson, D. J. The neuron-restrictive silencer factor (NRSF): A coordinate repressor of multiple neuron-specific genes. *Science* **267**(5202), 1360–1363 (1995).
34. Gao, Z. *et al.* The master negative regulator REST/NRSF controls adult neurogenesis by restraining the neurogenic program in quiescent stem cells. *J. Neurosci.* **31**(26), 9772–9786 (2011).
35. Ballas, N. *et al.* Regulation of neuronal traits by a novel transcriptional complex. *Neuron* **31**(3), 353–365 (2011).
36. Conaco, C. *et al.* Reciprocal actions of REST and a microRNA promote neuronal identity. *PNAS* **103**(7), 2422–2427 (2006).
37. Wu, J. & Xie, X. Comparative sequence analysis reveals an intricate network among REST, CREB and miRNA in mediating neuronal gene expression. *Genome Biol.* **7**(9), R85 (2006).
38. Li, X. & Jin, P. Roles of small regulatory RNAs in determining neuronal identity. *Nat. Rev. Neurosci.* **11**(5), 329–338 (2010).
39. Li, Q. *et al.* The splicing regulator PTBP2 controls a program of embryonic splicing required for neuronal maturation. *eLife* **3**, e01201 (2014).
40. Castro, D. S. *et al.* A novel function of the proneural factor Ascl1 in progenitor proliferation identified by genome-wide characterization of its targets. *Genes Dev.* **25**(9), 930–945 (2011).
41. Vasconcelos, F. F. & Castro, D. S. Transcriptional control of vertebrate neurogenesis by the proneural factor Ascl1. *Front. Cell Neurosci.* **8**, 412 (2014).
42. Chanda, S. *et al.* Generation of induced neuronal cells by the single reprogramming factor ASCL1. *Stem Cell Rep.* **3**(2), 282–296 (2014).
43. Casarosa, S., Fode, C. & Guillemot, F. Mash1 regulates neurogenesis in the ventral telencephalon. *Development* **125**(3), 525–534 (1999).
44. Chen, Z.-F., Paquette, A. J. & Anderson, D. J. NRSF/REST is required in vivo for repression of multiple neuronal target genes during embryogenesis. *Nat. Genet.* **20**(2), 136–142 (1998).
45. Vuong, J. K. *et al.* PTBP1 and PTBP2 serve both specific and redundant functions in neuronal pre-mRNA splicing. *Cell Rep.* **17**(10), 2766–2775 (2016).
46. Pang, Z. P. *et al.* Induction of human neuronal cells by defined transcription factors. *Nature* **476**(7359), 220–223 (2011).
47. Keppetipola, N. M. *et al.* Multiple determinants of splicing repression activity in the polypyrimidine tract binding proteins, PTBP1 and PTBP2. *RNA* **22**(8), 1172–1180 (2016).

Acknowledgements

JD-O is receiving salary support from the Fonds du Québec en Recherche, Santé (FRQS) and Parkinson Quebec. VO gratefully acknowledges the support of the US National Institutes of Health (USPHS Grant R01HL119102). We thank Joachim Eriksson for preliminary work on building the literature-based network.

Author contributions

A.M., J.D.-O., V.O. designed the study. A.M., J.D.-O., V.O. wrote the manuscript. A.M. built the model, conducted computational analysis. A.M. and V.D. performed parameter optimisation. J.D.-O. and E.M.L. produced and analysed the experimental data. R.A.B. provided cells and material. All authors provided inputs and comments on the manuscript.

Funding

Open Access funding provided by Lund 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/s41598-021-81089-8>.

Correspondence and requests for materials should be addressed to J.D.-O. or V.O.

Reprints and permissions information is available at 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 licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence 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 licence, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2021