SIMULATIONS OF ASTROCYTE INDUCTION BY TRANSIENT OVEREXPRESSION

Jonas Petersson

Department of Astronomy and Theoretical Physics, Lund University

Bachelor thesis supervised by Victor Olariu

Abstract

Astrocytes are one of the most common types of cells found in the central nervous system, and they play an important supportive role for neurons and other brain cells. In order to obtain astrocytes for medical or experimental purposes, some protocols enable the production of astrocytes from stem cells, but it is of interest to be able to produce astrocytes from fibroblasts, a cell type found in the skin. By forcing cells to produce specific proteins, it is possible to change a cells specialization, and the goal of this thesis is to model the molecular dynamics that occur during astrocyte formation when starting either from stem or fibroblast cells.

A multi-level model is proposed, with interactions inspired by literature, including genetic and epigenetic regulations. Model simulations are conducted with overexpression of NFIA and Sox9; and overexpression of NFIA in the presence of LIF, which is predicted to result in that the stem and fibroblast cell states will move into the astrocyte state.

Experimental findings validate the model's predictions. The results are also indicative of new experiments and given more experimental data to optimize the parameters, the model will be able to offer recipes for fine-tuning astrocyte production protocols.

Populärvetenskaplig sammanfattning

Alla celler i samma kropp har samma genetiska information lagrad i DNA. Olika delar av den genetiska informationen är lagrade i en otillgänglig form. Vilken del av cellens DNA som är tillgänglig avgör sedan vilken specialisering cellen har. Genom att kontrollera vilken information som används kan man omprogrammera celler, till exempel kan en hudcell bli en hjärncell. En metod för att åstadkomma detta är genom att tvinga cellen producera specifika proteiner som startar en omvandlingsprocess.

Astrocyter är en typ av hjärncell som har en stödjande roll i nervsystemet. Forskning har visat att sjukdomar som Alzheimer's och Parkinson kan vara kopplade till defekta astrocyter. Mediciner för att främja astrocyternas funktioner eller transplantationer av friska astrocyter kan därför hjälpa personer som lider av dessa sjukdomar. Det är en fördel att transplantera astrocyter jämfört med neuroner eftersom astrocyter har en förmåga att migrera till de områden där de behövs. Problemet är att det fortfarande saknas tillräckligt effektiva och säkra metoder för att producera mänskliga astrocyter.

I detta examensarbete utvecklas en datormodell för att simulera processen som sker när stamceller och fibroblaster, en typ av cell som finns i huden, tvingas omvandlas till astrocyter. Modellen lyckas simulera vad som händer när cellen tvingas producera proteinerna NFIA och Sox9, med hjälp av modellen verkar det också gå att optimera de experimentella metoderna som används just nu.

Contents

1 Introduction

The human brain consists mostly of two different types of cells: neurons and glial cells. Glial cells can then be further divided into astrocytes, oligodendrocytes, and microglia. Astrocytes have essential roles in the brain, such as support of the blood-brain barrier, nutrient delivery, extracellular ion balance, glutamate uptake, and scarring processes (Allen and Barres, 2009; Markiewicz and Lukomska, 2006; Matyash and Kettenmann, 2010). It is currently not well known how astrocytes develop, but during recent years it has become of increasing interest to understand. Astrocytes are involved in autism-related genetic disorders such as Rett syndrome and Fragile X syndrome, but also neurodegenerative diseases including multiple sclerosis (MS), Alzheimer's disease, amyotrophic lateral sclerosis (ALS), and Huntington's disease (Molofsky et al., 2012). When mutations in the astrocytes reduce their ability to support surrounding neurons or when the astrocytes start expressing mutant proteins, toxic buildup can cause degeneration of neighbouring neurons (Lobsiger and Cleveland, 2007; Molofsky et al., 2012). Whether the diseases originate from neurons or glial cells, like astrocytes, has become a grey zone that needs to be further studied.

While it is possible to obtain astrocytes from biopsies and conduct *in vivo* studies, less invasive sources are of great value, especially if it is human astrocytes that are needed for the experiments. During recent years several groups have been able to quickly produce functional astrocytes from human embryonic stem cells(hESCs); human induced pluripotent stem cells(hiPSCs); and neural stem cells(NSCs) (Tchieu et al., 2019; Canals et al., 2018; Li et al., 2018). The protocols for pluripotent cells work, but there are drawbacks. When working with hESCs, there are ethical issues since they are obtained from human embryos, and there is a risk that tumours develop in hESC and hiPSC cultures (Ben-David and Benvenisty, 2011).

Experiments have concluded that it is possible to reprogram adult somatic mouse fibroblasts into astrocytes by forcing the expression of the proteins Sox9, NFIA and NFIB (Caiazzo et al., 2015). Direct reprogramming of human fibroblasts to functional astrocytes has not yet been accomplished with human cells, but it should only be a matter of time before an efficient protocol is established.

In order to help to develop a protocol for human fibroblasts and to understand how pluripotent cells and fibroblasts can become astrocytes, computer simulations can aid the process of unveiling the roles of proteins and genes. By looking at experimental data, it is possible to formulate differential equations to fit the data, and also extrapolate new scenarios, suggesting new experiments. Through a dialectic process, it is then possible to develop a deeper understanding of the molecular interactions determining the cells fate.

To explore the differentiation mechanisms leading to astrocytes, for the first time, a multi-level regulation model for astrocyte formation is proposed. Two cases will be simulated: when starting from a multipotent/pluripotent and a somatic state, in this case, fibroblast skin cells. The first level is a deterministic gene regulatory network governed by ordinary differential equations. The second level is epigenetic and attempts to capture the dynamics observed in a region of the DNA that is important for astrocyte development by using a stochastic CpG methylation model. The model presented is based on literature, but

assumptions of interactions have been made, which is suggestive for further experimental exploration.

The model predicts that overexpression of NFIA and Sox9 leads to the formation of astrocytes, both when starting in a fibroblast and pluripotent state, which is confirmed by experimental findings. The model also sheds light on how Sox9 can both help to maintain the pluripotent state and also play a crucial role in the initiation of gliogenesis by its complex formation with NFIA. To improve the efficiency of direct reprogramming of human fibroblasts, the model also suggests to investigate the methylation of the promoter region of the protein GFAP further, and also to investigate the role of N-CoR in fibroblasts.

By readily having access to human astrocytes, researchers will be able to study how drugs and other therapies can restore or enhance the tasks of the astrocytes, for example transplanting healthy astrocytes might be a more feasible alternative to transplanting neurons in patients with neurodegenerative diseases (Lobsiger and Cleveland, 2007). The goal of this thesis is to be the start of an *in silico* investigation, leading to more effective methods of astrocyte production.

2   Background

A study has shown that it is possible to induce astrocytes from human pluripotent stem cells by overexpressing the transcription factors NFIB and Sox9 (Canals et al., 2018). This method was able to yield functional astrocytes in 14 days, which is a considerable improvement compared to older methods taking between 80-180 days (Krencik and Zhang, 2011; Roybon et al., 2013). In another study, it was also shown that NFIA and Sox9 could induce functional human astrocytes in 4-7 weeks (Li et al., 2018). More recently, even NFIA alone was found to be able to induce glial competency in human NSCs and together with leukaemia inhibitory factor, LIF, able to produce functional astrocytes (Tchieu et al., 2019). These three methods together suggest that NFIA and NFIB might have similar roles in the induction process.

There was also a study done on direct conversion from mouse fibroblasts to induced astrocytes by the overexpression of NFIA, NFIB, and Sox9 (Caiazzo et al., 2015). In this study, the authors found that the three transcription factors interact synergistically to induce mature astrocytes. These results are of particular interest since they suggests that it might be possible to conduct similar experiments on human fibroblasts.

Several other factors have also been suggested to play roles in gliogenesis such as Notch signalling, histone methylation, and the protein N-CoR reviewed by Molofsky et al. (Molofsky et al., 2013). N-CoR has been found to play a role in determining whether the cell commits to a neural or glial path by repression of the glial fate (Hermanson et al., 2002; Miller and Gauthier, 2007) and this is why N-CoR will be subject to our modelling study. It has not been experimentally established how N-CoR interacts with the previously mentioned transcription factors, but since NFIA, NFIB, and Sox9 were experimentally shown to accelerate the differentiation process, it is here hypothesized that they suppress N-CoR.

2.1   Relevant Transcription Factors And Proteins

2.1.1   NFIA/NFIB and LIF

One transcription factor that has been identified to be crucial in gliogenesis and astrocytogenesis is NFIA (Deneen et al., 2006). The expression of this transcription factor is maintained in astrocytes while other glial precursors only express it during their development. It has been found that NFIA expression is enough to trigger glial competency. However, the continued expression stops the differentiation into astrocytes. By having the cells in LIF medium and transiently overexpressing NFIA and then letting the NFIA levels drop, it was possible to generate glial fibrillary acidic protein(GFAP) expressing astrocytes quickly. It was also observed that transiently inducing NFIA without LIF does not induce endogenous expression of NFIA but rather after the NFIA levels have dropped, the cell becomes neurogenic again and the glial competency is lost. NFIA regulates DNA methylation and chromatin accessibility as well as lengthening of the G1 cell cycle phase. Continued high levels of NFIA expression also stops the differentiation process into astrocytes since the cell enters a G1 cell cycle arrest (Tchieu et al., 2019).

In mouse NSCs, experiments have shown that LIF enhances NSC self-renewal, and it also promotes GFAP expression (Pitman et al., 2004; Bauer and Patterson, 2006). LIF signalling activates the transcription factor STAT3, and there is a STAT3 binding site in the GFAP promoter region (Bonni et al., 1997). However, an increased GFAP expression in mouse NSC from LIF signalling does not lead to the cell committing to the astroglial lineage (Pitman et al., 2004). However, when human NSCs were grown in LIF medium, transient overexpression of NFIA was required to observe any increase of GFAP expression (Tchieu et al., 2019). The requirement of NFIA indicates that the regulation of GFAP works slightly differently in mice and humans, and in this thesis, it is assumed that LIF does not promote GFAP expression or astroglial commitment without NFIA like observed in human cells.

2.1.2   Sox9

Sox9 regulates the induction of NFIA, and these two transcription factors form a complex that coregulates genes that perform metabolic and migratory roles during both the gliogenesis and astrogenesis. Sox9 has also been associated with the maintenance of neural stem cells (Kang et al., 2012).

2.1.3   N-CoR

The transcription factor N-CoR binds with the DNA-binding protein RBPJ, or CSL, and together they repress GFAP expression (Hermanson et al., 2002) in mice. On the other hand, when RBPJ binds with the Notch intracellular domain, Notch-ICD, the complex activates the GFAP promoter and induces astrogliogenesis (Ge et al., 2002). An experiment has shown that N-CoR is crucial for neural stem cells to remain undifferentiated since cells with the N-CoR gene deleted started to show morphological signs of differentiation and

failed to proliferate, even if treated with Fibroblast growth factor-2, FGF2 (Hermanson et al., 2002).

2.1.4 GFAP

For a long time, GFAP has been used as an astrocytic marker, but there are drawbacks since this is a late-expressed marker which hardly is expressed in several protoplasmic, located in the grey matter, astrocytes (Molofsky and Deneen, 2015). However, since it is one of the most studied markers, this marker and its promoter region will be a cornerstone of the model developed here.

2.2 Direct Cell Reprogramming

In contrast to deriving astrocytes from NSCs, postnatal mouse fibroblasts have also been possible to convert into astrocytes by forced expression of NFIA, NFIB and Sox9 (Caiazzo et al., 2015). During the reprogramming, no expression of OCT4 or Sox2 proteins was found, which indicates that the transition happened without any intermediate pluripotent state (Niwa et al., 2000; Masui et al., 2007). The STAT3-binding site on the GFAP promoter of the mouse embryonic fibroblasts has also been observed to be demethylated, which suggests that the GFAP expression in these cells might not be repressed by DNA methylation (Caiazzo et al., 2015).

2.3 DNA Methylation

In mammalian DNA, there are regions of the genome where there are so-called CpG-islands. Cytosine and guanine connected with a phosphate group are what forms a CpG-site and regions with a high concentration of CpG-sites forms CpG-islands. These CpG-sites can be methylated or demethylated, and the methylation status often changes during the lifetime of mammals. Since around 40% of mammalian promoter regions contain CpG-islands (Fatemi et al., 2005) they are of high interest to include in simulations since methylated CpG-sites prevent any transcription originating from the promoter region. The STAT3-binding site on the GFAP promoter has been found to change from methylated to demethylated during induced astrocytogenesis (Tchieu et al., 2019; Li et al., 2018). It is also observed that overexpression of NFIA demethylates the STAT3-binding site and also that the chromatin accessibility changes to one similar to human pluripotent stem cell, hPSC, derived astrocytes (Tchieu et al., 2019).

3 Results

3.1 The Network

The network developed comes from literature studies, and assumptions of interactions have been made. Mouse NSCs spontaneously develop into astrocyte when the N-CoR gene is

removed (Hermanson et al., 2002) and experiments indicate that DNA methylation does not repress GFAP in mouse fibroblasts (Caiazzo et al., 2015). This led to having an N-CoR node which is promoted by an NSC and a Fibroblast node, the N-CoR node then repress an astrocyte node.

It is here necessary to note that terms pluripotent cells and NSCs are used interchangeably with regards to the network, this is due to a simplification where it is assumed that ESCs, PSCs and NSCs behave the same concerning the other nodes in the network, see figure 1.

Table 1: The main sources of information used to develop the network are specified here.

Transcription Factor	Source	Animal
NFIA/NFIB	(Tchieu et al., 2019)	human
	(Canals et al., 2018)	human
	(Li et al., 2018)	human
	(Caiazzo et al., 2015)	mouse
	(Deneen et al., 2006)	chick & mouse
Sox9	(Li et al., 2018)	human
	(Canals et al., 2018)	human
	(Caiazzo et al., 2015)	human
	(Kang et al., 2012)	chick & mouse
N-CoR	(Hermanson et al., 2002)	mouse
	(Ge et al., 2002)	mouse
LIF	(Tchieu et al., 2019)	human
	(Pitman et al., 2004)	mouse
	(Bauer and Patterson, 2006)	mouse
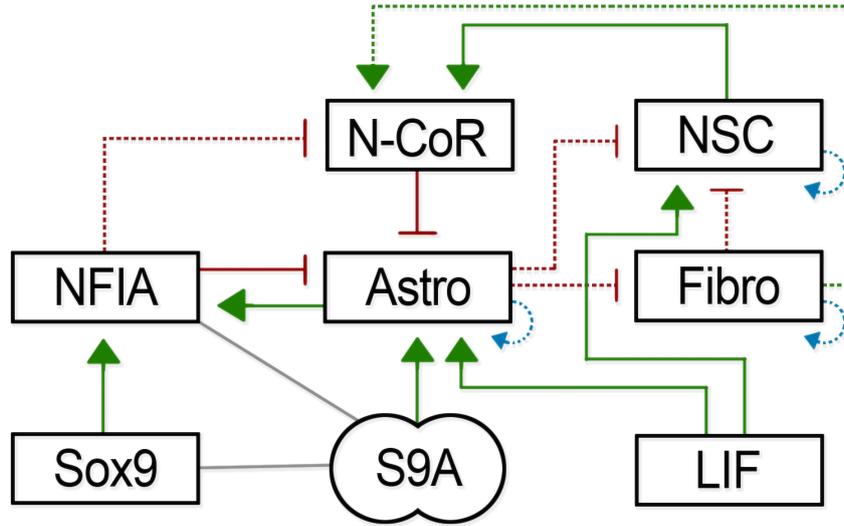
Figure 1: The network used for the simulations. The red lines show repression of transcription, the green lines show promotion of transcription, the blue lines show self-promotion, and the grey lines shows the complex formation of NFIA and Sox9. The dashed lines indicate proposed interactions.

The repression of N-CoR by NFIA is a proposed interaction that makes it possible for NFIA to move the cell into a gliogenic state while at the same time repressing the astrocyte cell fate, though if either LIF or Sox9 is present, NFIA will set the stage for them to differentiate the cell into an astrocyte. Even though the literature suggests that NFIA inhibits astrocyte differentiation through the lengthening of the G1 cell cycle phase (Tchieu et al., 2019), it is simplified as transcriptional repression, seen in figure 1.

This model will not attempt at including details in either the fibroblast state, NSC state or astrocyte state. These three nodes will, therefore, be simplified as self-promoting nodes.
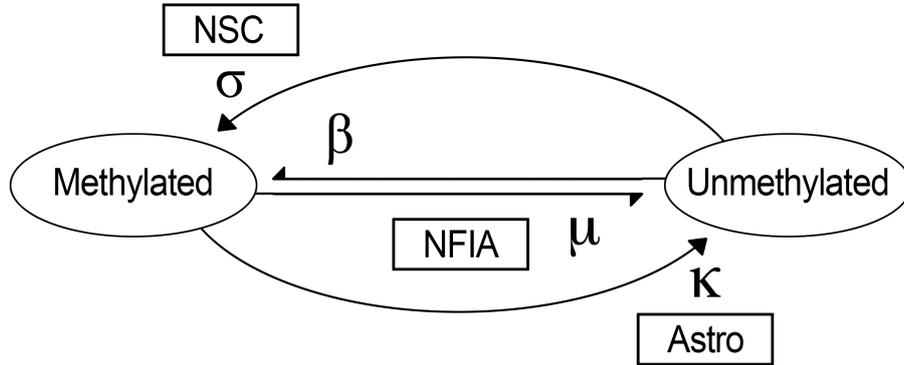
Figure 2: The epigenetic model used to simulate the methylation of the CpG-sites that limits the expression of the astrocyte node. The expression level of the nodes NSC, NFIA and Astro in figure 1 controls the probabilities σ, μ and κ respectively. β and μ are probabilities of switching state. σ and κ are probabilities for two sites in the same state to convert a site in the opposite state, method further described in section 5.4.

The complex formed by Sox9 and NFIA, S9A, will promote the astrocyte node but it will also limit the amount of free NFIA that otherwise would repress N-CoR expression. This dynamic will play an essential role for the CpG-demethylation since the S9A will not be able to increase astrocyte expression until the CpG-sites are sufficiently demethylated and even if high overexpression of Sox9 will promote more NFIA, the amount of free NFIA will be almost none.

3.2   From Pluripotent/NSC State

In undifferentiated cells the initial expression level of the Pluripotent node and N-CoR is high, see figure 3, 4. This results in a stable state which can be identified as either a pluripotent or multipotent state. When NFIA is overexpressed, see figure 3a), the cell moves to another state where there is some expression of astrocytic genes which corresponds to a gliogenic state, but that will not result in a terminal commitment to the astroglial lineage. Since NFIA represses the Astro node, the leaking Sox9 expression will not be sufficient to change state, but with the introduction of LIF, this becomes possible, see figure 3b). Without the repression of N-CoR by NFIA, LIF will only stabilize the pluripotent state, but the cooperation of LIF and NFIA results in an induced astrocyte state.
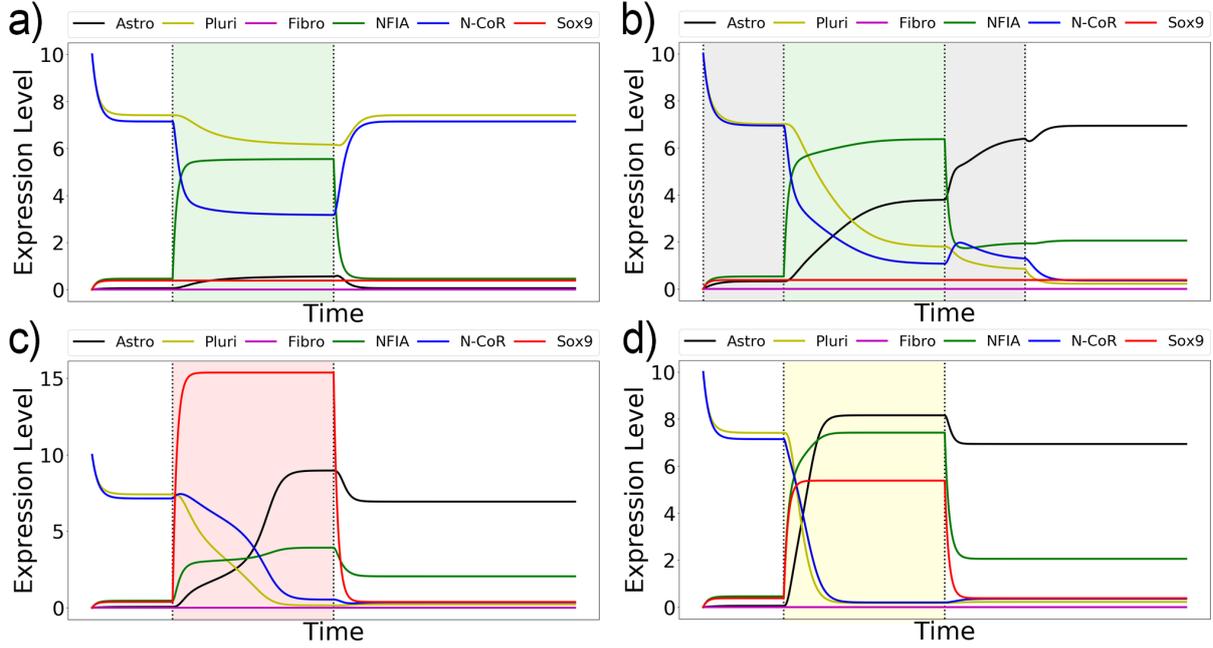
Figure 3: Single level simulation of the network shown in figure 1. Equations 5.7-5.12 are used with the parameters found in table 2. The initial expression level is 0 for all nodes except N-CoR and Pluri which are 10. **a)** NFIA is overexpressed in the green region by $NFIA_{over} = 0.5$, **b)** NFIA is overexpressed in the green region by $NFIA_{over} = 0.5$, and [LIF] = 0.5 in both the grey and green region, **c)** Sox9 is overexpressed in the red region by $Sox9_{over} = 1.5$, **d)** NFIA and Sox9 are overexpressed in the yellow region by $NFIA_{over} = 0.5$ and $Sox9_{over} = 0.5$.

When only Sox9 is overexpressed, see figure 3c), a smaller amount of NFIA is induced which will form a complex with Sox9 and induce the astrocyte node. This leads to the pluripotent node becoming repressed faster than N-CoR which results in slightly different dynamics. When the expression level of Sox9 is high compared to NFIA, almost all of the NFIA will be bound in the complex, and thus NFIA will not be able to repress N-CoR sufficiently.

In the case where both Sox9 and NFIA are overexpressed a fast state change happens, see figure 3d). When the NFIA expression is higher than the Sox9 expression, most of the Sox9 will be bound in the complex, but there will also be free NFIA that can repress the N-CoR.

When the epigenetic level is added to the astrocyte node, see figure 2, the dynamics change, most notably in the case of the Sox9 overexpression, see figure 4c). In the multi-level model it is even more important to have free NFIA since it contributes to an increase of demethylation, see figure 4a), which can further enable a full demethylation switch, see figure 4b),d).
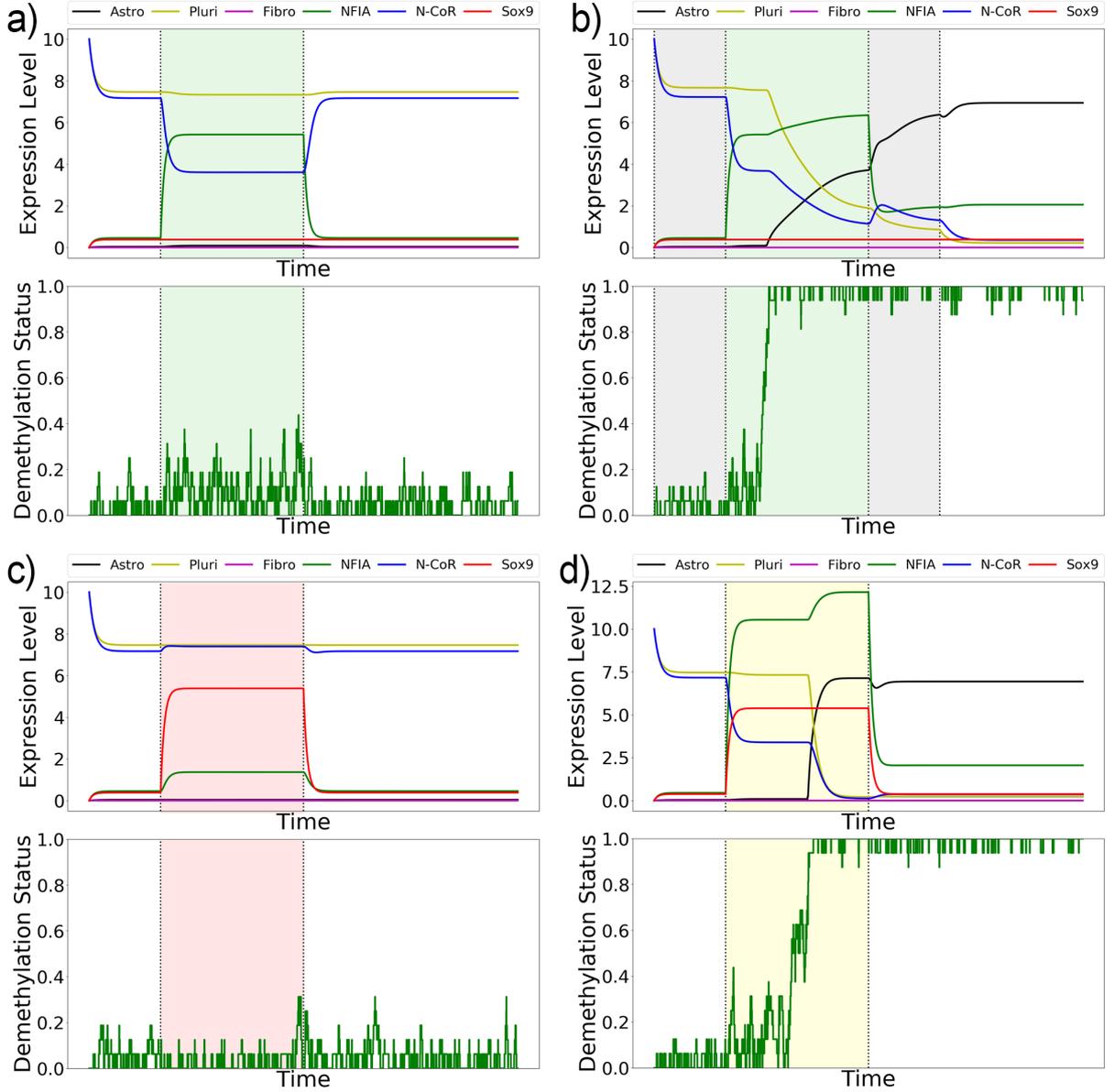
Figure 4: Multi-level simulation of the network shown in figure 1. Equations 5.8-5.12, 5.17 are used with the parameters found in table 2. The initial expression level is 0 for all nodes except N-CoR and Pluri which are 10. Also, the initial number of methylated CpG-sites is 8. The stochastic methylation plots show the proportion of unmethylated CpG-sites for the astrocyte node. a) NFIA is overexpressed in the green region by $NFIA_{over} = 0.5$, b) NFIA is overexpressed in the green region by $NFIA_{over} = 0.5$ and $[LIF] = 0.5$ in both the grey and green region, c) Sox9 is overexpressed in the red region by $Sox9_{over} = 0.5$, d) NFIA and Sox9 are overexpressed in the yellow region by $NFIA_{over} = 1.0$ and $Sox9_{over} = 0.5$.

To further explore how much NFIA and Sox9 overexpression and LIF concentration is

needed for the cell to commit to an astrocyte lineage, the corresponding parameter spaces are probed and shown in figure 5. Both the single and multi-level model, figure 5a),b), show a large range of parameters that will lead to astrocyte commitment. It is also clear that in the case of the multi-level model, too much Sox9 will make it impossible for NFIA to demethylate the CpG-sites like in figure 4c).

In the parameter space formed by NFIA overexpression and LIF concentration, see figure 5c),d), a much smaller space is seen that leads to astrocyte commitment. When the LIF concentration becomes too high, there will be very active promotion of the pluripotent node causing the astrocyte node to remain low, even if it is promoted by LIF. When the NFIA expression becomes too high, there is also intense repression of the astrocyte node such that the pluripotent node remains strong, and even if the CpG-sites demethylate during the overexpression, they will methylate back before the astrocyte node can repress the pluripotent node.
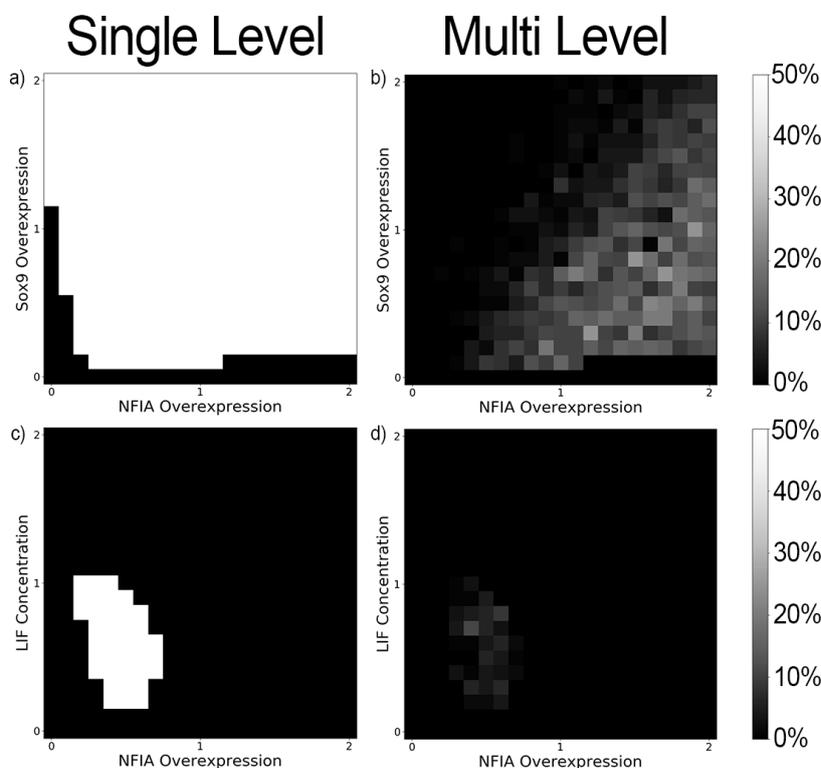
Figure 5: In the case of the single layer plots, a) and c), simulations are carried out in the same manner as in figure 3 by varying [LIF], $NFIA_{over}$ and $Sox9_{over}$ between 0 and 2 in steps of 0.1. The white areas show where the simulation ends in an astrocyte state. In the multi-layer plots, b) and d), simulations are carried out 100 times in the same manner as in figure 4 by varying [LIF], $NFIA_{over}$ and $Sox9_{over}$ between 0 and 2 in steps of 0.1. The percentage of simulations that ends in an Astrocyte state is then plotted. The maximum probability in b) is 22% and in d) 12%.

3.3 From Fibroblast State

When starting in the Fibroblast state, the Fibroblast node and N-CoR are expressed highly, and the CpG-sites are mostly demethylated (Caiazzo et al., 2015). Because of the demethylation status, only the multi-level results, see figure 6, will be shown here since the single level model behaves almost the same. The single-level simulations can be seen in appendix B.

During the Sox9 overexpression in figure 6c), there is not any need to have a large amount of free NFIA since the CpG-sites are already demethylated. Much of the dynamics other than that remain the same as in figure 4. The case where NFIA is overexpressed in a LIF medium, figure 6b), has slightly different implications though. Since the role of LIF is to maintain rather than induce pluripotent states, it is probably not realistic that an increase of pluripotent expression is observed. However, if this artefact of the model is ignored the result is still realistic since LIF does still promote the astrocyte node.
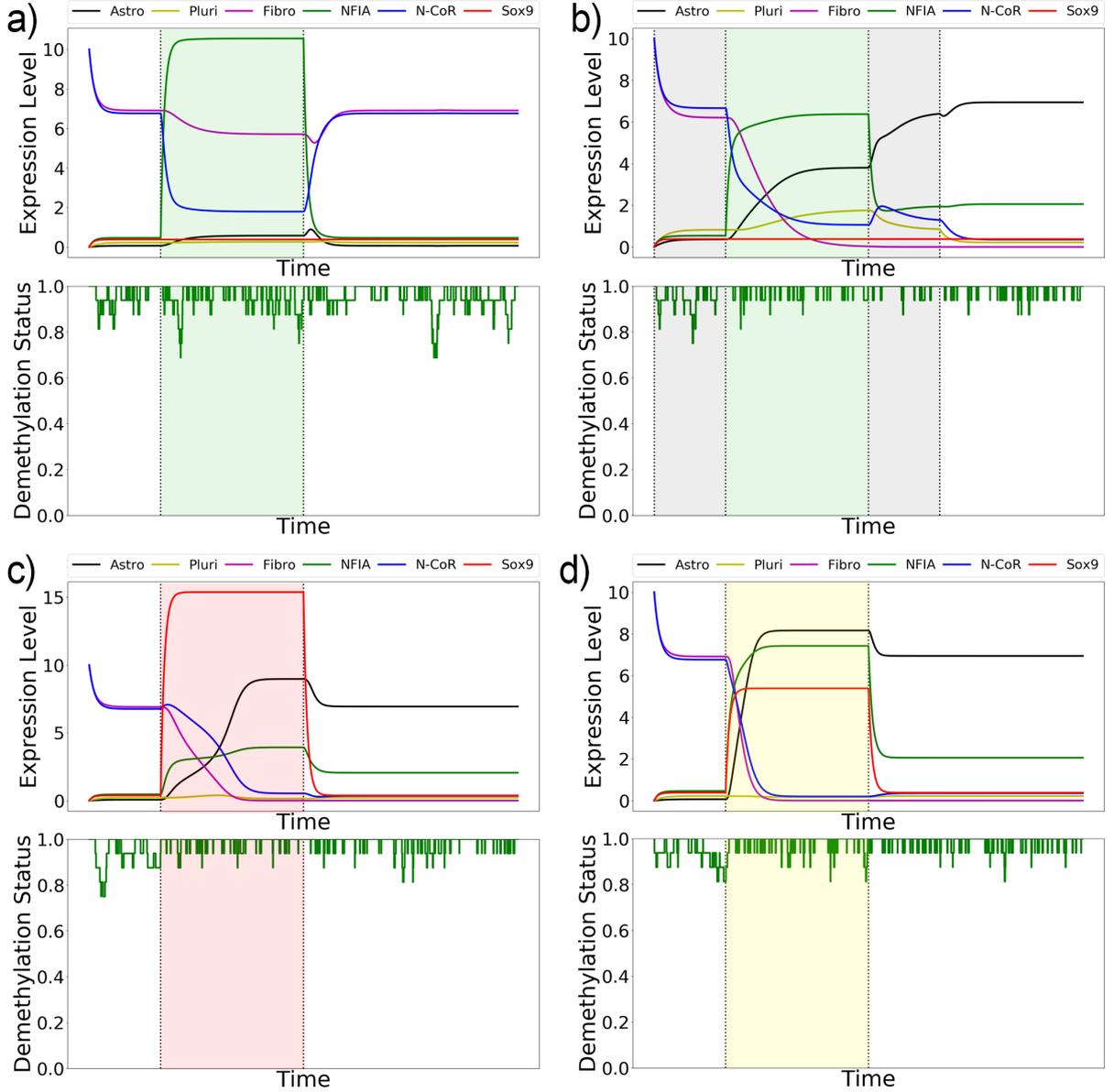
Figure 6: Multi-level simulation of the network shown in figure 1. Equations 5.8-5.12, 5.17 are used with the parameters found in table 2. The initial expression level is 0 for all nodes except N-CoR and Fibro which is 10, also the initial amount of methylated CpG-sites is 0. The stochastic methylation plots shows the proportion of unmethylated CpG-sites for the astrocyte node. **a)** NFIA is overexpressed in the green region by $NFIA_{over} = 1.0$, **b)** NFIA is overexpressed in the green region by $NFIA_{over} = 0.5$ and $[LIF] = 0.5$ in both the grey and green region, **c)** Sox9 is overexpressed in the red region by $Sox9_{over} = 1.5$, d) NFIA and Sox9 are overexpressed in the yellow region by $NFIA_{over} = 0.5$ and $Sox9_{over} = 0.5$.

The spaces formed by varying NFIA and Sox9 overexpression and LIF concentration

are shown in figure 7. In figure 7a),b) the behaviours of the single and multi-level models are almost identical, indicating that DNA methylation is not a source of repression when starting in a fibroblast state.

When looking at figure 7c),d) there is some discrepancy between the two models, though this is mostly explained by the artefact seen in figure 6b). The expression level of the pluripotency node increases as a response to the high LIF concentration, the expression of pluripotency then makes it much more likely that the CpG-sites methylate and the promotion of the astrocyte node is lost.
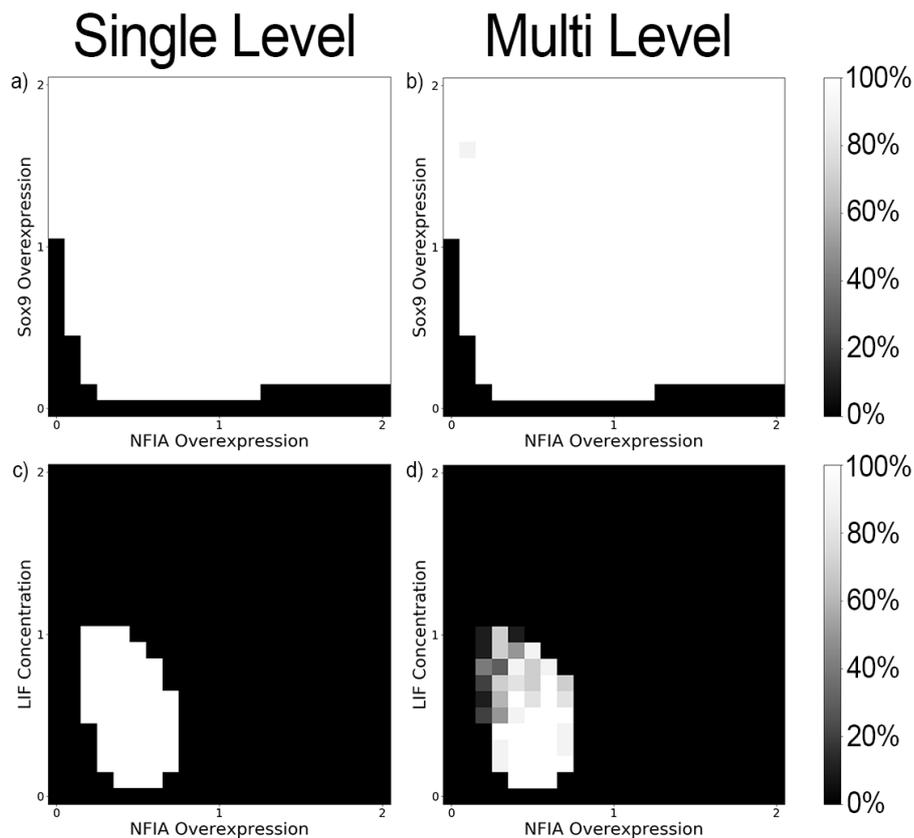
Figure 7: In the case of the single layer plots, a) and c), simulations are carried out in the same manner as in figure 3, but with the initial expression of Fibro being 10 and Pluri as 0, by varying [LIF], $NFIA_{over}$ and $Sox9_{over}$ between 0 and 2 in steps of 0.1. The white areas show where the simulation ends in an astrocyte state. In the multi-layer plots, b) and d), simulations are carried out ten times in the same manner as in figure 6 by varying [LIF], $NFIA_{over}$ and $Sox9_{over}$ between 0 and 2 in steps of 0.1. The percentage of simulations that ends in an Astrocyte state is then plotted.

4 Discussion

We have shown that it is possible to describe the dynamics of astrocyte induction, both from fibroblasts and NSCs, using a multi-level model consisting of a gene regulatory network and a stochastic CpG-site methylation model. While overexpression of NFIA and Sox9 separately struggled to form astrocytes when starting in a pluripotent state, simultaneous overexpression quickly induced an astrocyte state. Also, NFIA overexpression in pluripotent cells led to astrocyte formation while in the presence of LIF.

When starting in a fibroblast state, high overexpression of Sox9 led to astrocytes forming, but NFIA overexpression alone did not suffice. As seen in the case of the pluripotent state, simultaneous overexpression of NFIA and Sox9 quickly induced astrocytes. Our model also predicts that the GFAP promoter remains unmethylated through the process of direct reprogramming, as is also observed experimentally (Caiazzo et al., 2015).

4.1 From Pluripotent/NSC State

The model sheds light on the role of Sox9 during astrocytogenesis. Experimentally, Sox9 was observed to support the NSC state while also playing an important role in the initiation of gliogenesis (Kang et al., 2012). In figure 3c) and 4c) there is a stark difference in the dynamics of the system. When the epigenetic level is added, it becomes important that there is free NFIA that can demethylate the CpG-sites, but when overexpressing Sox9, most of the NFIA is bound in a complex with Sox9. Our model predicts that it is important to pay attention to the relative doses of Sox9 and NFIA, see figure 5b), to increase the chances of obtaining astrocytes from ESCs, iPSCs and NSCs.

When overexpressing NFIA in the presence of LIF the model predicts a lower percentage of formed astrocytes and the range of allowed doses is also smaller compared to when overexpressing NFIA and Sox9, see figures 5b) and d). The model suggests that overexpression of NFIA and Sox9 is a more reliable method for astrocyte production, but more experimental data to optimize the model is needed in order to draw any definitive conclusions.

4.2 From Fibroblast State

Our model predicts that the methylation status of the astrocyte node is not a source of repression since the behaviour of the single-level and multi-level model is the same, see figure 7. It is experimentally established that the GFAP promoter region is unmethylated before and after the direct conversion of mouse fibroblasts to astrocytes (Caiazzo et al., 2015), which validates our model.

The model also predicts that simultaneous expression of NFIA and Sox9 is an effective method of reprogramming the cell, as also is concluded experimentally (Caiazzo et al., 2015). It is noteworthy that the experimental protocol used for mouse fibroblasts, which our model is based on, did not succeed as well with human fibroblasts.

While it seems to be the case that direct conversion of mouse fibroblasts is effective, in the same study by Caiazzo et al., they tried their protocol on human neonatal fibroblasts, and in only 2% of the cells, they found signs of immature astrocytes (Caiazzo et al., 2015). It is, therefore, possible that there is a barrier between human fibroblasts and astrocytes that is not found in mice.

Since there is a higher frequency of methylated CpG-sites in humans compared to mice (Bird, 1980), it is possible that the GFAP promoter region is methylated in human fibroblasts, unlike mouse fibroblasts (Caiazzo et al., 2015). This could be verified experimentally, and it would explain the discrepancy seen between direct reprogramming of mice and human fibroblasts. If the GFAP promoter region is methylated, figure 7b) would be more similar to figure 5b), and if the used protocol expressed too much Sox9, the cells would not go through an epigenetic switch, and the result would be very few or no induced astrocytes.

In this thesis, we propose that N-CoR represses astrogliogenesis in a similar way as REST in neurons (Drouin-Ouellet et al., 2017) and that N-CoR is expressed in fibroblasts. N-CoR has been investigated in pluripotent mouse cells (Hermanson et al., 2002), but it has not been experimentally established what function N-CoR has in fibroblasts during astrocyte induction, neither for mice or human cells. Seeing whether N-CoR is expressed in fibroblasts, and if N-CoR repressed fibroblasts would result in more efficient astrocyte production would confirm our networks predicted mean of N-CoR mediated astrocyte repression in fibroblasts.

Although we tried to overexpress NFIA in the presence of LIF for fibroblasts, it led to the expression of the pluripotent node which is neither desirable nor realistic since LIF is supposed to maintain and not induce pluripotent states. It is possible to resolve this artefact, but it was not further explored here.

Our proposed model proves capable of describing direct reprogramming of astrocytes from mouse fibroblasts. It is also suggestive of new experiments and modification of protocols to avoid saturation of complex bound NFIA, in case that NFIA has a more vital role when reprogramming human fibroblasts.

5 Methods

5.1 Complex Formation

Sox9 and NFIA will form a complex that will be called S9A. An equation for the complex can be formulated:

$$[\text{Sox9}_{free}] + [\text{NFIA}_{free}] \rightleftharpoons [\text{S9A}] \tag{5.1}$$

Since the complex formation is much faster than the DNA synthesis it is assumed for our purpose that the equation is in equilibrium. From this assumption the following equations can be formed:

$$\frac{d[\text{Sox9}_{free}]}{dt} = \frac{d[\text{NFIA}_{free}]}{dt} = 0 \rightarrow [\text{Sox9}_{free}][\text{NFIA}_{free}] - k_d[\text{S9A}] = 0 \tag{5.2}$$

$$k_d = \frac{[\text{Sox9}_{free}][\text{NFIA}_{free}]}{[\text{S9A}]} \tag{5.3}$$

$$[\text{Sox9}] = [\text{Sox9}_{free}] + [\text{S9A}] \tag{5.4}$$

$$[\text{NFIA}] = [\text{NFIA}_{free}] + [\text{S9A}] \tag{5.5}$$

With equation 5.2 - 5.5 the following expression for the complex concentration is obtained:

$$[\text{S9A}] = \frac{k_d + [\text{Sox9}] + [\text{NFIA}]}{2} \\ - \sqrt{\left(\frac{k_d + [\text{Sox9}] + [\text{NFIA}]}{2}\right)^2 - [\text{Sox9}][\text{NFIA}]} \tag{5.6}$$

5.2 Deterministic Model

The differential equations used are based on a thermodynamical approach of modelling transcription (Ackers et al., 1982; Shea and Ackers, 1985). In the simulations the following equations are used:

$$\frac{d[\text{Astro}]}{dt} = \frac{a0 + a1[\text{S9A}]^2 + a2[\text{Astro}] + a3[\text{LIF}]}{1 + a1[\text{S9A}]^2 + a2[\text{Astro}] + a3[\text{LIF}] + a4[\text{N-CoR}]^2 + a5([\text{NFIA}] - [\text{S9A}])} \\ - \gamma[\text{Astro}] \tag{5.7}$$

$$\frac{d[\text{NFIA}]}{dt} = NFIA_{over} + \frac{b0 + b1([\text{Sox9}] - [\text{S9A}]) + b2[\text{Astro}]}{1 + b1([\text{Sox9}] - [\text{S9A}]) + b2[\text{Astro}]} - \gamma[\text{NFIA}] \tag{5.8}$$

$$\frac{d[\text{N-CoR}]}{dt} = \frac{c0 + c1[\text{Fibro}] + c2[\text{NSC}]}{1 + c1[\text{Fibro}] + c2[\text{NSC}] + c3([\text{NFIA}] - [\text{S9A}])} - \gamma[\text{N-CoR}] \tag{5.9}$$

$$\frac{d[\text{Sox9}]}{dt} = Sox9_{over} + d0 - \gamma[\text{Sox9}] \tag{5.10}$$

$$\frac{d[\text{NSC}]}{dt} = \frac{e0 + e1[\text{NSC}] + e2[\text{LIF}]}{1 + e1[\text{NSC}] + e2[\text{LIF}] + e3[\text{Astro}] + e4[\text{Fibro}]} - \gamma[\text{NSC}] \tag{5.11}$$

$$\frac{d[\text{Fibro}]}{dt} = \frac{f0 + f1[\text{Fibro}]}{1 + f1[\text{Fibro}] + f2[\text{Astro}]} - \gamma[\text{Fibro}] \tag{5.12}$$

18

5.3 Parameters

Table 2: The parameters used for equations 5.6, 5.7-5.12, 5.17. The parameters were obtained through a mix of a simple optimization algorithm and manual tuning, described in section 5.6. These parameters are used for all of the simulations carried out, both the deterministic and multi-level model. The step size when solving the equations is h and a the explicit Euler method was used.

h	γ	kd	a0	a1	a2	a3	a4	a5	b0	b1	b2
0.1	0.1	0.6	0.08	0.31	0.62	0.8	0.4	0.55	0.038	0.027	0.03

c0	c1	c2	c3	d0	e0	e1	e2	e3	e4	f0	f1	f2
0	0.38	0.44	0.93	0.038	0.09	0.36	0.5	0.95	0.94	0	0.34	0.8

5.4 CpG-site modelling

To simulate the methylation of the CpG-sites a simplified stochastic model based on a paper by Haerter et al. (Haerter et al., 2013) is used, see figure 2. The model used has also been proven effective to simulate the establishment of pluripotency (Olariu et al., 2016).

- Pick two random sites

  - If those two sites are methylated: pick another random site and methylate that site with probability σ.

  - If those two sites are unmethylated: pick another random site and unmethylate that site with probability κ.

- Pick a random site

  - If the site is methylated: unmethylate it with probability μ.

  - If the site is unmethylated: methylate it with probability β

The updating scheme is applied on every step in time when solving the differential equations.

There is in total 8 CpG-sites for the STAT3 promoter region 1 of the GFAP promoter which the model will use and it is assumed that the transcription is on when 6 or more sites are unmethylated (Tchieu et al., 2019).

The probabilities σ, κ, μ and β are calculated as follows on every step in time:

$$\sigma = 0.6 \frac{[\text{NSC}]}{1 + [\text{NSC}]} \tag{5.13}$$

$$\kappa = 0.1 + 0.9 \frac{[\text{Astro}]}{1 + [\text{Astro}]} \tag{5.14}$$

$$\mu = 0.01 + 0.04 \frac{[\text{NFIA}] - [\text{S9A}]}{1 + [\text{NFIA}] - [\text{S9A}]} \tag{5.15}$$

$$\beta = 0.01 \tag{5.16}$$

5.5 Multi-Layer Model

When the methylation is being simulated, eq. 5.7 is replaced with the following equation where η is defined as follows:

$$\eta(CpG_{unmethylated}) = \begin{cases} 1, & \text{if } CpG_{unmethylated} \geq 6 \\ 0, & \text{otherwise} \end{cases}$$

η will then be used to limit the transcription except for a_0 since it's the leaking transcription:

$$\frac{d[\text{Astro}]}{dt} = \frac{a0 + \eta(a1[\text{S9A}]^2 + a2[\text{Astro}] + a3[\text{LIF}])}{1 + \eta(a1[\text{S9A}]^2 + a2[\text{Astro}] + a3[\text{LIF}]) + a4[\text{N-CoR}]^2 + a5([\text{NFIA}] - [\text{S9A}])} \\ - \gamma[\text{Astro}] \tag{5.17}$$

5.6 Optimization

To find initial parameters for the single-level model a very simple optimization algorithm was deployed:

- Pick a random set of parameters between 0 and 1.

- Compute the cost by:

  - Running simulations with equations 5.6-5.12 for all the cases of overexpression mentioned in table 3 and 4.

  - If the value of the node is within the allowed range the cost is zero. If the value of the node is outside of the allowed range, the absolute error from the closest bound of the allowed range is the cost.

  - The sum of all the simulation costs are returned as the final cost for the set of parameters.

- Randomly pick 1-3 (randomized) of the parameters and add uniform noise between -0.1 and 0.1, but limit the parameters to be between 0 and 1.

- Compute the new cost as in the second step.

- If the new cost is lower than the first, accept the new parameters. If the cost is higher, discard the new parameters and go to step 3 again.

- Stop once the cost is zero.

The parameters obtained qualitatively showed that the wanted dynamics was possible with the proposed single-level model. The parameters, for the single and multi-level model, were then manually tuned with the multi-level equations 5.13-5.16. Since the models were not optimized to an experimental data set, the resulting simulations only show qualitative behaviours and no units of either concentration or time is introduced.

Table 3: Table containing the constraints used to compute the cost during the optimization of the single-level model. The initial expression level for all nodes is 0 except for the N-CoR and pluripotent node which start at 10. The simulations run until t=6000 with stepsize=0.1. (Table continues on next page)

Overexpression	Node	Time	Allowed Range
None	Astro	Before Overexpression	0-0.5
-	-	Endpoint	0-0.5
-	Pluri	Before Overexpression	7-10
-	-	Endpoint	7-10
-	Fibro	Before Overexpression	0-0.5
-	-	Endpoint	0-0.5
-	NFIA	Before Overexpression	0-0.5
-	-	Endpoint	0-0.5
-	Sox9	Before Overexpression	0-0.5
-	-	Endpoint	0-0.5
-	N-CoR	Before Overexpression	5-10
-	-	Endpoint	5-10
$NFIA_{over} = 3$	Astro	Before Overexpression	0-0.5
-	-	Endpoint	0-0.5
-	Pluri	Before Overexpression	7-10
-	-	Endpoint	7-10
-	Fibro	Before Overexpression	0-0.5
-	-	Endpoint	0-0.5
-	NFIA	Before Overexpression	0-0.5
-	-	Endpoint	0-0.5
-	Sox9	Before Overexpression	0-0.5
-	-	Endpoint	0-0.5
-	N-CoR	Before Overexpression	5-10
-	-	Endpoint	5-10
$Sox9_{over} = 3$	Astro	Before Overexpression	0-0.5
-	-	Endpoint	7-10
-	Pluri	Before Overexpression	7-10

Overexpression	Node	Time	Allowed Range
-	-	Endpoint	0-0.5
-	Fibro	Before Overexpression	0-0.5
-	-	Endpoint	0-0.5
-	NFIA	Before Overexpression	0-0.5
-	-	Endpoint	1-2
-	Sox9	Before Overexpression	0-0.5
-	-	Endpoint	0-0.5
-	N-CoR	Before Overexpression	5-10
-	-	Endpoint	0-0.5
NFIA_{over},$\text{Sox9}_{over} = 3$	Astro	Before Overexpression	0-0.5
-	-	Endpoint	7-10
-	Pluri	Before Overexpression	7-10
-	-	Endpoint	0-0.5
-	Fibro	Before Overexpression	0-0.5
-	-	Endpoint	0-0.5
-	NFIA	Before Overexpression	0-0.5
-	-	Endpoint	1-2
-	Sox9	Before Overexpression	0-0.5
-	-	Endpoint	0-0.5
-	N-CoR	Before Overexpression	5-10
-	-	Endpoint	0-0.5
$\text{NFIA}_{over} = 3$, LIF= 1	Astro	Before Overexpression	0-0.5
-	-	End of Overexpression	0-3
-	-	Endpoint	7-10
-	Pluri	Before Overexpression	7-10
-	-	Endpoint	0-0.5
-	Fibro	Before Overexpression	0-0.5
-	-	Endpoint	0-0.5
-	NFIA	Before Overexpression	0-0.5
-	-	Endpoint	1-2
-	Sox9	Before Overexpression	0-0.5
-	-	Endpoint	0-0.5
-	N-CoR	Before Overexpression	5-10
-	-	Endpoint	0-0.5

Table 4: Table containing the constraints used to compute the cost during the optimization of the single-level model. The initial expression level for all nodes is 0 except for the N-CoR and fibroblast node which start at 10. The simulations run until t=6000 with stepsize=0.1. (Table continues on next page)

Overexpression	Node	Time	Allowed Range
None	Astro	Before Overexpression	0-0.5

Overexpression	Node	Time	Allowed Range
-	-	Endpoint	0-0.5
-	Pluri	Before Overexpression	0-0.5
-	-	Endpoint	0-0.5
-	Fibro	Before Overexpression	7-10
-	-	Endpoint	7-10
-	NFIA	Before Overexpression	0-0.5
-	-	Endpoint	0-0.5
-	Sox9	Before Overexpression	0-0.5
-	-	Endpoint	0-0.5
-	N-CoR	Before Overexpression	5-10
-	-	Endpoint	5-10
$Sox9_{over} = 3$	Astro	Before Overexpression	0-0.5
-	-	Endpoint	7-10
-	Pluri	Before Overexpression	0-0.5
-	-	Endpoint	0-0.5
-	Fibro	Before Overexpression	7-10
-	-	Endpoint	0-0.5
-	NFIA	Before Overexpression	0-0.5
-	-	Endpoint	1-2
-	Sox9	Before Overexpression	0-0.5
-	-	Endpoint	0-0.5
-	N-CoR	Before Overexpression	5-10
-	-	Endpoint	0-0.5
$NFIA_{over}, Sox9_{over} = 3$	Astro	Before Overexpression	0-0.5
-	-	Endpoint	7-10
-	Pluri	Before Overexpression	0-0.5
-	-	Endpoint	0-0.5
-	Fibro	Before Overexpression	7-10
-	-	Endpoint	0-0.5
-	NFIA	Before Overexpression	0-0.5
-	-	Endpoint	1-2
-	Sox9	Before Overexpression	0-0.5
-	-	Endpoint	0-0.5
-	N-CoR	Before Overexpression	5-10
-	-	Endpoint	0-0.5

References

Gary K Ackers, Alexander D Johnson, and Madeline A Shea. Quantitative model for gene regulation by lambda phage repressor. *Proceedings of the national academy of sciences*, 79(4):1129–1133, 1982.

Nicola J Allen and Ben A Barres. Neuroscience: glia—more than just brain glue. *Nature*, 457(7230):675, 2009.

Sylvian Bauer and Paul H Patterson. Leukemia inhibitory factor promotes neural stem cell self-renewal in the adult brain. *Journal of neuroscience*, 26(46):12089–12099, 2006.

Uri Ben-David and Nissim Benvenisty. The tumorigenicity of human embryonic and induced pluripotent stem cells. *Nature reviews cancer*, 11(4):268, 2011.

Adrian P Bird. DNA methylation and the frequency of CpG in animal DNA. *Nucleic acids research*, 8(7):1499–1504, 1980.

Azad Bonni, Yi Sun, Mireya Nadal-Vicens, Ami Bhatt, David A Frank, Irina Rozovsky, Neil Stahl, George D Yancopoulos, and Michael E Greenberg. Regulation of gliogenesis in the central nervous system by the JAK-STAT signaling pathway. *Science*, 278(5337): 477–483, 1997.

Massimiliano Caiazzo, Serena Giannelli, Pierluigi Valente, Gabriele Lignani, Annamaria Carissimo, Alessandro Sessa, Gaia Colasante, Rosa Bartolomeo, Luca Massimino, Stefano Ferroni, et al. Direct conversion of fibroblasts into functional astrocytes by defined transcription factors. *Stem cell reports*, 4(1):25–36, 2015.

Isaac Canals, Aurélie Ginisty, Ella Quist, Raissa Timmerman, Jonas Fritze, Giedre Miskinyte, Emanuela Monni, Marita G Hansen, Isabel Hidalgo, David Bryder, et al. Rapid and efficient induction of functional astrocytes from human pluripotent stem cells. *Nature methods*, 15(9):693, 2018.

Benjamin Deneen, Ritchie Ho, Agnes Lukaszewicz, Christian J Hochstim, Richard M Gronostajski, and David J Anderson. The transcription factor NFIA controls the onset of gliogenesis in the developing spinal cord. *Neuron*, 52(6):953–968, 2006.

Janelle Drouin-Ouellet, Shong Lau, Per Ludvik Brattås, Daniella Rylander Ottosson, Karolina Pircs, Daniela A Grassi, Lucy M Collins, Romina Vuono, Annika Andersson Sjöland, Gunilla Westergren-Thorsson, et al. Rest suppression mediates neural conversion of adult human fibroblasts via microrna-dependent and-independent pathways. *EMBO molecular medicine*, 9(8):1117–1131, 2017.

Mehrnaz Fatemi, Martha M Pao, Shinwu Jeong, Einav Nili Gal-Yam, Gerda Egger, Daniel J Weisenberger, and Peter A Jones. Footprinting of mammalian promoters: use of a CpG DNA methyltransferase revealing nucleosome positions at a single molecule level. *Nucleic acids research*, 33(20):e176–e176, 2005.

Weihong Ge, Keri Martinowich, Xiangbing Wu, Fei He, Alison Miyamoto, Guoping Fan, Gerry Weinmaster, and Yi Eve Sun. Notch signaling promotes astrogliogenesis via direct CSL-mediated glial gene activation. *Journal of neuroscience research*, 69(6):848–860, 2002.

Jan O Haerter, Cecilia Lövkvist, Ian B Dodd, and Kim Sneppen. Collaboration between CpG sites is needed for stable somatic inheritance of DNA methylation states. *Nucleic acids research*, 42(4):2235–2244, 2013.

Ola Hermanson, Kristen Jepsen, and Michael G Rosenfeld. N-cor controls differentiation of neural stem cells into astrocytes. *Nature*, 419(6910):934, 2002.

Peng Kang, Hyun Kyoung Lee, Stacey M Glasgow, Meggie Finley, Tataka Donti, Zachary B Gaber, Brett H Graham, Aaron E Foster, Bennett G Novitch, Richard M Gronostajski, et al. Sox9 and NFIA coordinate a transcriptional regulatory cascade during the initiation of gliogenesis. *Neuron*, 74(1):79–94, 2012.

Robert Krencik and Su-Chun Zhang. Directed differentiation of functional astroglial subtypes from human pluripotent stem cells. *Nature protocols*, 6(11):1710, 2011.

Xiang Li, Yezheng Tao, Robert Bradley, Zhongwei Du, Yunlong Tao, Linghai Kong, Yi Dong, Jeffrey Jones, Yuanwei Yan, Cole RK Harder, et al. Fast generation of functional subtype astrocytes from human pluripotent stem cells. *Stem cell reports*, 11(4): 998–1008, 2018.

Christian S Lobsiger and Don W Cleveland. Glial cells as intrinsic components of non-cell-autonomous neurodegenerative disease. *Nature neuroscience*, 10(11):1355, 2007.

Inga Markiewicz and Barbara Lukomska. The role of astrocytes in the physiology and pathology of the central nervous system. *Acta neurobiol exp (wars)*, 66(4):343–358, 2006.

Shinji Masui, Yuhki Nakatake, Yayoi Toyooka, Daisuke Shimosato, Rika Yagi, Kazue Takahashi, Hitoshi Okochi, Akihiko Okuda, Ryo Matoba, Alexei A Sharov, et al. Pluripotency governed by sox2 via regulation of oct3/4 expression in mouse embryonic stem cells. *Nature cell biology*, 9(6):625, 2007.

Vitali Matyash and Helmut Kettenmann. Heterogeneity in astrocyte morphology and physiology. *Brain research reviews*, 63(1-2):2–10, 2010.

Freda D Miller and Andrée S Gauthier. Timing is everything: making neurons versus glia in the developing cortex. *Neuron*, 54(3):357–369, 2007.

Anna V Molofsky and Benjamin Deneen. Astrocyte development: a guide for the perplexed. *Glia*, 63(8):1320–1329, 2015.

Anna V Molofsky, Robert Krenick, Erik Ullian, Hui-hsin Tsai, Benjamin Deneen, William D Richardson, Ben A Barres, and David H Rowitch. Astrocytes and disease: a neurodevelopmental perspective. *Genes & development*, 26(9):891–907, 2012.

Anna V Molofsky, C Hochstim, B Deneen, and D Rowitch. *Comprehensive Developmental Neuroscience: Patterning and Cell Type Specification in the Developing CNS and PNS: Chapter 36. Mechanisms of Astrocyte Development.* Elsevier Inc. Chapters, 2013.

Hitoshi Niwa, Jun-ichi Miyazaki, and Austin G Smith. Quantitative expression of oct-3/4 defines differentiation, dedifferentiation or self-renewal of es cells. *Nature genetics*, 24 (4):372, 2000.

Victor Olariu, Cecilia Lövkvist, and Kim Sneppen. Nanog, oct4 and tet1 interplay in establishing pluripotency. *Scientific reports*, 6:25438, 2016.

M Pitman, B Emery, M Binder, S Wang, H Butzkueven, and TJ Kilpatrick. Lif receptor signaling modulates neural stem cell renewal. *Molecular and cellular neuroscience*, 27 (3):255–266, 2004.

Laurent Roybon, Nuno J Lamas, Alejandro Garcia-Diaz, Eun Ju Yang, Rita Sattler, Vernice Jackson-Lewis, Yoon A Kim, C Alan Kachel, Jeffrey D Rothstein, Serge Przedborski, et al. Human stem cell-derived spinal cord astrocytes with defined mature or reactive phenotypes. *Cell reports*, 4(5):1035–1048, 2013.

Madeline A Shea and Gary K Ackers. The or control system of bacteriophage lambda: A physical-chemical model for gene regulation. *Journal of molecular biology*, 181(2): 211–230, 1985.

Jason Tchieu, Elizabeth L Calder, Sudha R Guttikonda, Eveline M Gutzwiller, Kelly A Aromolaran, Julius A Steinbeck, Peter A Goldstein, and Lorenz Studer. NFIA is a gliogenic switch enabling rapid derivation of functional human astrocytes from pluripotent stem cells. *Nature biotechnology*, page 1, 2019.

A   Stability Analysis
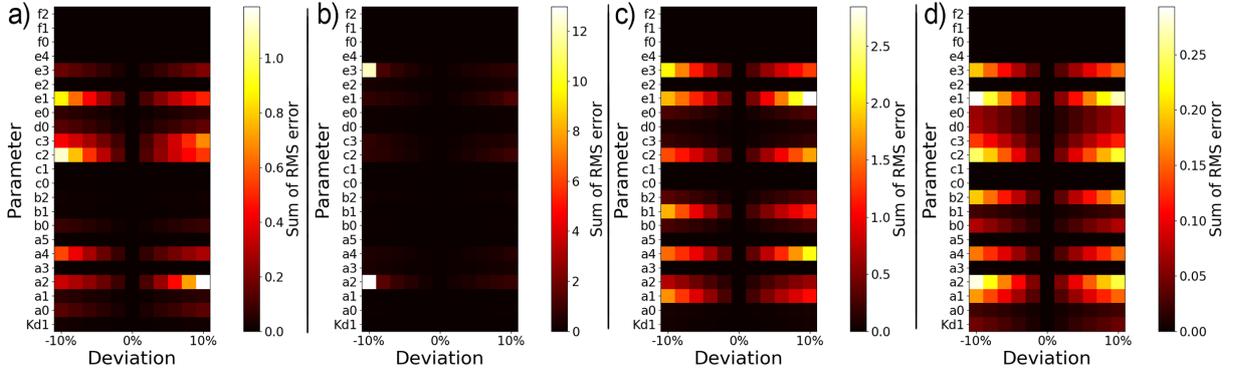
Figure 8: Stability analysis of the parameters used, see table 2. The 0% deviations are the simulations in 3a),b),c),d) respectively. The Z-axis is the sum of the RMS errors of each node when the parameters are varied between -10% and 10% in 11 steps.
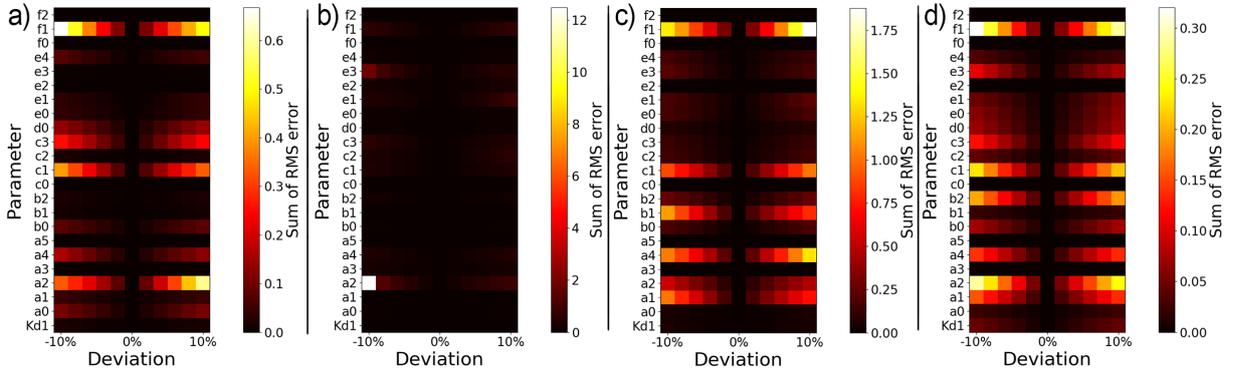
Figure 9: Stability analysis of the parameters used, see table 2. The 0% deviations are the simulations in 10a),b),c),d) respectively. The Z-axis is the sum of the RMS errors of each node when the parameters are varied between -10% and 10% in 11 steps.

B   Single-level Fibroblast Simulations
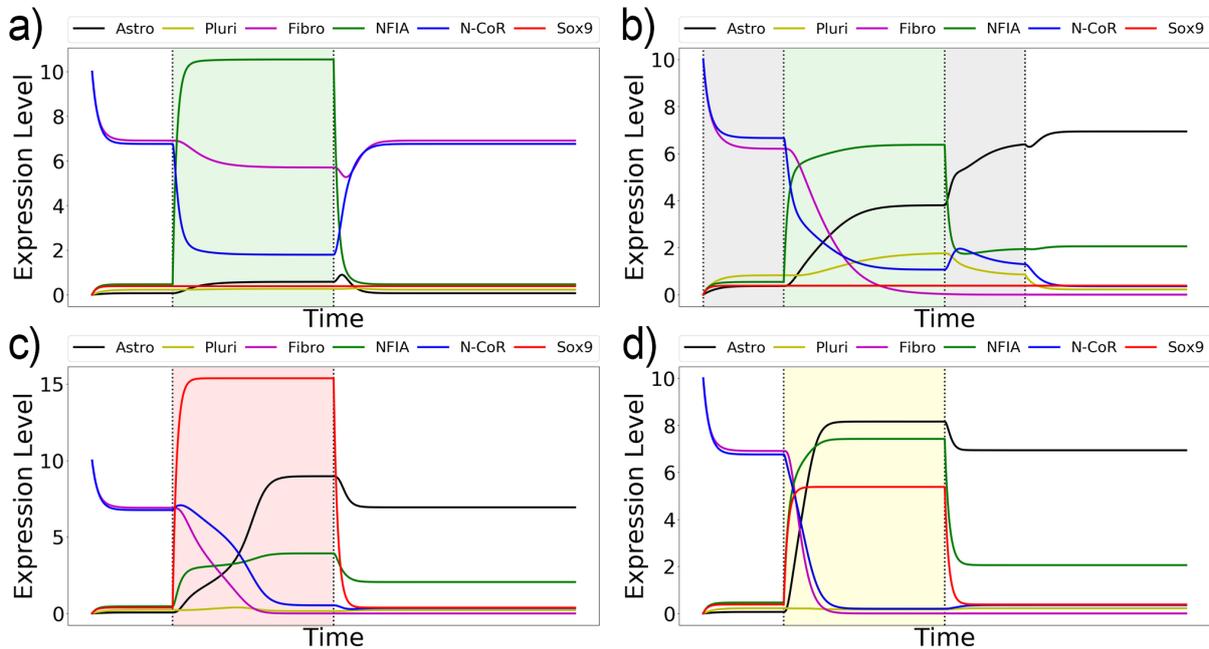
Figure 10: Single level simulation of the network shown in figure 1. Equation 5.7-5.12 are used with the parameters found in table 2. The initial expression level is 0 for all except N-CoR and Fibro which is 10. **a)** NFIA is overexpressed in the green region by $NFIA_{over} = 1.0$, **b)** NFIA is overexpressed in the green region by $NFIA_{over} = 0.5$ and $[LIF] = 0.5$ in both the grey and green region, **c)** Sox9 is overexpressed in the red region by $Sox9_{over} = 1.5$, **d)** NFIA and Sox9 is overexpressed in the yellow region by $NFIA_{over} = 0.5$ and $Sox9_{over} = 0.5$.