Inferring and simulating a gene regulatory network for the sympathoadrenal differentiation from single-cell transcriptomics in human.

Olivier Gandrillon[a,b]

[a]Univ Lyon, ENS de Lyon, Univ Claude Bernard, CNRS UMR 5239, INSERM U1210, Laboratory of Biology and Modelling of the Cell, 46 allée d'Italie Site Jacques Monod, Lyon, F-69007, France

[b]Inria

Abstract

Neuroblastoma is a malignant childhood cancer with significant inter- and intrapatient heterogeneity, arising from abnormal differentiation of neural crest cells into sympathetic neurons. The lack of actionable mutations limits therapeutic options, highlighting the need to better understand the molecular mechanisms driving this differentiation. While RNA velocity has provided some insights, it is limited in modeling regulatory relationships.

To address this, we applied our integrated Gene Regulatory Networks (GRNs) inference and simulation tools using a published single-cell RNAseq dataset from human sympathoadrenal differentiation. Our analysis identified a 97-gene GRN driving the transition from Schwann cell precursors to chromaffin cells and sympathoblasts, highlighting dynamic interactions like self-reinforcing loops and toggle switches.

To showcase the model's ability to predict the impact of perturbations, we conducted *in silico* knockouts (KOs) and overexpressions (OVs) of few selected genes. This analysis revealed that certain perturbations had a notably significant effect on the differentiation process, either acing specifically on one ligneage choice, or preventing cells to exit from a stem compartment. Altogether, these findings demonstrate the utility of our GRN model framework in predicting gene perturbations after inferring the GRN structure.

Introduction

Neuroblastoma is a childhood cancer responsible for around 15% of cancer-related deaths in children aged 0 to 4 years. These tumors typically develop near the adrenal glands and are characterized by significant inter- and intrapatient heterogeneity (1). Treatments are still limited with few actionable mutations and very few drugs have been validated clinically (2), although CAR-T cells targeting GD2 antigen represent a promising prospect (3). Neuroblastoma is an archetypal example of a tumor arising through alteration in a differentiation process, more precisely during the differentiation of migratory neural crest cells into sympathetic neurons (4; 5; 6; 7). There is therefore an urgent need for a better understanding of the mechanisms driving the normal development from neural crest and mesodermal lineages (8), to better apprehend the origin of different neuroblastoma subtypes, their internal heterogeneity and interpatient variation in malignancy.

Initial studies have been performed in mice (9). Although mouse models provide in-depth, experimentally supported insights into developmental processes, inconsistencies between mouse and human development could hinder our understanding of certain diseases. For example, transitions between cell fates during differentiation of migratory neural crest cells into sympathetic neurons appear to occur in a different order in humans versus mice (10). Furthermore, neuroblastomas do not naturally appear in mice, and current genetic models do not recapitulate the full spectrum of natural properties of the disease (11).

It was therefore critical to identify the human-specific aspects of development, which led to a recent surge of studies of the normal development at the single cell level in humans (10; 12; 13). In order to go beyond a static description of the differentiation process, the authors of the aforementioned papers have resorted to the use of RNA velocity analysis to get a more dynamical view (10; 13). RNA velocity is driven by a kinetic model of RNA dynamics and works by distinguishing unspliced and spliced mRNA counts, and fitting gene-specific parameters based on an on-off deterministic transcription model. This allows RNA velocity to predict the direction of future changes in spliced mRNA levels (14; 15; 16). However, while RNA velocity uses a time-dependent gene expression model, in this context time has not explicit biological interpretation (17). Furthermore, each gene is modelled independently and regulatory relationships are ignored (15).

We recently described a model for Gene Regulatory Networks (GRNs) that alleviates most of those drawbacks and gives a detailed access to a fully mechanistic view of gene regulation during a differentiation process (18; 19). In this model, each gene is modelled as a piecewise deterministic Markov process (PDMP) describing bursts in the production of mRNA and the resulting synthesis of proteins (20; 21). Genes are coupled via an interaction function that describes how the proteomic field feedbacks into the frequency of bursts. The coupling of a GRN inference algorithm (CARDAMOM) together with a simulation algorithm (HARISSA) allowed us to demonstrate the potential of this approach for inferring executable models of GRN that can both reproduce the observed experimental data but also makes prediction regarding the time dependent evolution of the studied system in the face of a given perturbation (either knock out (KO) or overexpression (OV) of a given gene).

In the present work, we describe the use of our integrated approach to infer the GRN driving the sympathoadrenal differentiation in human based upon single-cell RNAseq data (10), and to predict the effect of some gene perturbations.

Our analysis identified a 97 genes-based GRN involved in the transition from Schwann cell precursor to chromaffin cells and sympathoblasts. Its study revealed dynamic gene interactions such as self-reinforcing loops and toggle switches. In order to demonstrate the ability of the model to predict the effect of perturbations, we performed both *in silico* KOs and OVs of a few selected genes. This revealed that some perturbation had a quite significant effects on the differentiation process.

Results

1. Organizing the cells

Our first task was to organize the cells according to their differentiation status, since our GRN inference is anchored in the time-dependent evolution of the differentiation process. Cells were collected from human embryos at post-conception weeks 6, 8, 9, 11, 12 and 14. We first assessed whether or not this cell collection time could be a relevant framework for generating time-reconstructed data. As shown in Supplementary Figure 1b, this proved not to be the case, since all cell types were present at all collection times. This is in line with the fact that in the human adrenal gland, the transition from SCP to sympathoadrenal states persists over several weeks (10), which lead to the expectation that cells at all stages of differentiation should be present at all collection times.

We were therefore faced with a different situation from the one we used to benchmark CARDAMOM, where the cells sampling time was in accordance with the actual process time (18).

Consequently, we decided to organize the cells according to their stemness score (Figure 1a). For this we assumed a differentiation sequence starting with Schwann cell precursor (SCP), and giving rise to chromaffin cells and sympathoblasts through a bifurcation. Such a scheme is in line with the conclusions of the original paper from which we obtained the dataset (10) and is shared by some (13; 12; 6) but not all (4; 22) authors working on the sympathoadrenal differentiation in human. This led to the definition of the stemness score (see material and methods), as the ratio of SCP-signature genes expression, evolving from 1 (only the SCP signature genes are expressed) to zero (none of the SCP signature genes are expressed). This allowed us to define a "process time" (Figure 1b) through the use of an overlapping sliding window of 300 cells. This was defined only on the first 1300 cells since the stemness score is almost down to zero at the end of that sequence, and since we wanted to capture in detail the very abrupt decrease in the stemness score. The use of an overlapping window is justified by the fact that differentiation should be considered as a discontinuous non linear process (see e.g. 23), and therefore should allow cells to go back and forth in the gene expression space. This resulted in a 20807 genes x 1800 cells matrix.

2. Gene selection

We then explored the most relevant gene set to be used for GRN inference. Since our inference and simulation scheme CARDAMOM is fundamentally rooted in exploiting the power of distributions (18; 24), we first computed a Kantorovitch distance (25) for all genes for all delta-times. Similarly, since the most delta-entropic genes were shown to represent most relevant genes for the differentiation process (26), we also computed a delta-entropy for all genes for all delta-times.

The union of those two lists (Supplementary Figure 2) provided us with a list of 97 genes (Supplementary table 1) that were used for inferring the GRN. This finally left us with a matrix displaying the expression level of 97 genes in 1800 cells distributed along 6 time points. We verified that the information contained within this matrix was sufficient to correctly capture the differentiation process we were interested in (Supplementary Figure 1c). It is important to note that those genes were chosen purely on their expression pattern, irrespective of their putative biological function (see discussion).

3. GRN inference and analysis

We then applied our CARDAMOM algorithm in order to infer the GRN structure (Figure 2a), resulting in a relatively densely connected network. We first explored the connectivity of the network. We observed a striking difference between the distribution in the number of outgoing and of incoming nodes (Figure 2b). The incoming nodes were distributed according to a normal (i.e. gaussian) distribution centered around a mean of 3, whereas the outgoing nodes displayed a very heavy long tail toward higher values, as well as a large number of null values (i.e. leaves, that is genes only receiving and not sending any signals within the GRN). The long tail was mostly due to 4 genes (CHGA, CHGB, STMN2 and HMGB2).

The intensity of all the outgoing interactions for those 4 genes are shown in Figure 2c. HMGB2 displays a specific pattern with quasi-exclusively inhibitory interactions whereas the three other genes display a relatively similar amount of positive and negative interactions. One should note that all but HMGB2 are part of defining gene signature (see Material and Methods).

One key feature of our inference process is the possibility to decompose the overall GRN into dynamical subparts where each edge appears at the timepoint transition for which it was detected with the strongest intensity by the inference algorithm (Figure 3; 18).

The dynamics starts with a small number of genes being turned on by the stimulus (Figure 3a). In our previous papers, the stimulus was explicitly identified as an exogenous medium change/addition (27; 18). Here it should be understood as a complex influence exerted by the cell environment, whether hormonal of through cell-cell contact, which induces the SCP cell differentiation. It is needed to "set the GRN in motion" and push the cells out of their quasi-steady state.

There is much more happening at the next time point (Figure 3b) including a very large set of genes being repressed by HMGB2. All known SCP-specific genes are repressed at that point which is expected. A CHGA-CHGB self-reinforcing loop starts activating the genes from the chromaffin lineage, while a few sympathoblast-specific genes (STMN2, HAND2) gets activated.

At the next time point, a RTN1-PENK toggle switch is apparent supported by positive loops on both sides (Figure 3c). We can see a lot of genes being repressed by CHGA and CHGB genes. This includes the repression of MPZ and PLP1 SCP-specific genes by CGHB, a chromaffin-specific gene.

The two latest time points show a very reduced amount of interactions (Figure 3d and 4e).

It is clear that the two time points where there is a large amount of interactions corresponds to the times where one observe a quite steep variation in the stemness score (Figure 3b and Figure 1). Altogether this time-decomposition offers a very clear illustration of a signal propagating through the GRN in the form of waves of genes activation and repression (27).

We finally extracted from the full GRN a subset of the genes connected by specific dynamical motifs, that is either cross-positive self-reinforcing loops or mutually repressive toggle switch (Figure 4). Among those only one direct toggle switch can be evidenced, the one linking PENK to RTN1 genes. All other motifs are positive. One can see a group of genes (COL5A2, PTPRZ1, EDRNB, POSTN), including one SCP-specific gene (POSTN) which reinforce together and are repressed concomitantly by chromaffin specific genes (CHGA and CHGB) and a sympathoblast-specific gene (STMN2).

4. GRN simulation

One of the main advantage of our approach is that it produces executable GRN models (28) the time dependent evolution of which can be simulated and the resulting simulated dataset can be confronted to the experimentally observed dataset.

There are many ways through which the result of a simulation can be assessed. Using a UMAP representation allows for a quick evaluation of the quality of the fit. This was first used to calibrate the process-time t1 to t5 into real times (Supplementary Figure 3). We first tested an evenly spaced time frame where each time points was set to be 4 hours apart (Supplementary Figure 3b). We observed that all cells were instantly projected from their t1 position into the final position of the process. We therefore tried different adjustements and obtained the best visual fit using a time sequence extending up to 80 hours (Supplementary Figure 3c), that was used for further simulations.

We then assessed to which extend the simulation captures the time-dependent evolution of a few genes the expression of which was characteristic for the three cell types we were trying to reproduce (Figure 5). Overall, the gene expression pattern, when observed though its UMAP projection was quite well reproduced by our GRN simulation. Most notably the chromaffin/sympathoblast bifurcation was quite well captured by our model.

Another way for assessing the quality of the fit consists in investigating the distance between the distribution of gene expression values between the experimentally observed mRNA distribution and the

simulated distribution. As shown in supplementary Figure 4, there was a close proximity between the two distributions as assessed by a low Kantorovich distance, for the 4 genes displayed. For a thorough comparison for all genes at all time points, see supplementary Figure 5. Altogether, although some refinement could be performed we considered at that stage that the fit quality of the experimental dataset was sufficient to engage in testing the impact of targeted gene alterations on the global behavior of our GRN.

5. Dynamical predictions

We first decided to assess the impact of genetic perturbations of the 4 genes that occupied a specific place with a high number of outgoing interactions in the GRN, that is CHGA, CHGB, STMN2 and HMGB2 (see Figure 2). We first performed the KOs and verified they were operational (supplementary Figure 4). As expected the distance between the GRN output and the experimental data was largely increased for the knocked out gene. In order to explore more systemic effects, we explored an aggregated value of distance which showed that the simulation of KO GRNs produced outputs that were more distant to the observed values than the simulation of the WT GRN (supplementary Figure 6).

In order to analyze the KO effect at the levels of the cell types, we reasoned that the use of signature genes for assessing cell identity was compromised by the fact that some of the genes that were knocked out were part of the set of signature genes. We therefore decided to take advantage of the clear separation of the three cell types in the 2D UMAP space (supplementary Figure 1c), where cell position in that space results from a more complex set of genes (Figure 6a to c). This allowed us to count the number of cells belonging to each cell type following simulation of the WT or the KO GRNs. It was immediately apparent that neither CHGB nor HMGB2 KOs had any visible effect whereas STMN2 KO lead to a strong reduction in the amount of sympathoblasts and CHGA KO resulted in a strong reduction in the amount of chromaffin cells (Figure 6d).

We then decided to assess the impact of overexpressing the same 4 genes. Interestingly, those results mostly mirrored the KO results: CHGA overexpression resulted in a large increase in the amount of chromaffin cells, whereas STMN2 overexpression lead to a strong increase in the amount of sympathoblasts (Figure 6e). The effect of CHGB overexpression was somewhat more marked than its KO effect, and worked in the same direction than CHGA overexpression phenotype.

We then tested a double perturbation: we simultaneously overexpressed STMN2 and knocked out CHGA. We observed a clear additive effect with a proportion of sympathoblasts of more than 70 % of the total cells using such a double perturbation (Figure 6f).

We finally tested the two SCP-specific genes (POSTN and S100B) present in the dynamical motif subset (see Figure 4) and verified that their overexpression was operational (supplementary Figure 7). We observed that, although POSTN overexpression had a very minor impact, the overexpression of S100B resulted in a very striking phenotype (Figure 7). Whereas in the WT case, the cells progressively exited the SCP state, to reach about 14 % after the 80 days of differentiation, in the case of the S100B overexpression, differentiation was severely impaired, with a strong reduction in the number of cells exiting the SCP state (Figure 7d). This results in over 70 % of cells in the SCP state at the end of the differentiation process. Both sympathoblast and chromaffin cell differentiation were affected (Figure 7b).

Discussion

One of the difficulties we had to solve in the present work is that the sampling time is not equal to the process time, since all differentiation steps appears at all collected time points. As we previously mentioned, "our approach can be applied to any biological process for which time-stamped single-cell transcriptomic data are obtained after applying a given stimulus. When such time-stamped snapshots are not available [which is the case here], the algorithm could in principle take as an input time-reconstructed data (i.e., artificially ordered snapshots). In that case, the quality of the inference will strictly depend on the effectiveness of the time reconstruction algorithm." (18).

We therefore assumed a differentiation sequence starting with Schwann cell precursor (SCP), and giving rise to chromaffin cells and sympathoblasts through a bifurcation. This allowed to organize the cells along a stemness score aimed at capturing the unfolding of the differentiation process. Such a differentiation sequence was demonstrated in the study from which we obtained our dataset (10). Such a sequence is shared by some (13; 12; 6) but not all (4; 22) authors. In the last case, the authors assumed a differentiation sequence starting with sympathoblasts, and giving rise to chromaffin cells and SCP through a bifurcation. The reason for such discrepancies is yet unclear, and sheds a crude light on the complexity of ascertaining a differentiation sequence *in vivo* in humans. One must stress that our method relies upon a given definition of the differentiation sequence and can not be used to untangle the different hypothesis.

Assuming this order in the differentiation sequence, we have proposed a framework through which cells could be organized, genes could be selected, and a 97 genes-based GRN was inferred.

When compared to the GRNs we previously inferred (27; 18), the structure of the adrenal differentiation GRN displays both similarities and differences. Among the differences, one should note that the depth of the GRN (i.e. the largest number of nodes connected in a row) is slightly larger (6) than for the erythroid differentiation sequence (depth of 3) or for the RA-induced ES cell differentiation (depth of 4). This is not imposed by our inference algorithm than can potentially infer GRN of any length. In any case, this depth is limited by the duration of the differentiation sequence, which is in the same range (80 hours) as for the erythroid (72 hours) or the ES (96 hours) differentiation sequences. If the network depth is too large, signal is too damped and delayed to accurately reproduce experimental data. Another difference was the role played by the stimulus that in the case of the sympathoblast differentiation displays a relatively modest effect (4 genes affected directly by the stimulus), whereas it had a much more pronounced effect for the ES (14 genes affected) and the erythroid (29 genes affected) differentiation sequences.

A very clear similarity in all those GRN behavior is that there is a very clear waves-like pattern in the way the signal is propagating through the network. In the case of the sympathoblast differentiation, the most dense interactions time are seen at the t2-t3 and t3-t4 transitions points (Figure 3), in line with the very strong and abrupt change in stemness observed for those time points (Figure 1b).

The analysis of the subset of the genes connected by specific dynamical motifs showed the existence of a group of genes, including two signature SCP-specific gene (POSTN and S100B) which reinforce together and are repressed by chromaffin specific genes and a sympathoblast-specific gene. This is reminiscent of the "team" concept: a team is a group of genes that activate member genes belonging to the same team while inhibiting genes of other teams directly and/or indirectly. It has been shown to be a network design principle that can drive cell-fate canalization in diverse decision-making processes (29) in a robust manner (30). In our case, this could be involved in the maintenance of an undifferentiated SCP phenotype, that will have to be turned down for allowing differentiation to proceed. The effect of overexpressing S100B which maintains cells in an undifferentiated phenotype is in good accordance with this view.

One should stress that the only *bona fide* toggle switch motif, involving the PENK and RTN1 genes, was not involved in the sympathoblast versus chromaffin cells decision-making, since their KO had no major

effect on the amount of both cell types (not shown). Nevertheless the opposite effect of the perturbation affecting STMN2 and CHGA genes (Figure 6) demonstrates the ability of the whole network to behave as a toggle switch.

Our results point to the limitations of making predictions based upon a simple reading of the proposed interactions. For example, both CHGA and CHGB tend to reinforce each other, yet the overexpression of CHGA has a major impact on cell decision whereas the overexpression of CHGB does not displays such an effect. Similarly, the overexpression of POSTN and S100B, both SCP-specific genes, and both repressed by differentiation induced genes do not produce a similar phenotypic effect. We strongly believe that this is a compelling argument in favor of our mechanistic dynamical approach.

This work was aiming at inferring and simulating a GRN operating during the normal sympathoadrenal differentiation in human. The main incentive behind the study of this specific differentiation sequence was to provide insights from human developmental studies relevant in cases of neural crest-related pathologies and cancers. Our work is therefore potentially a first step relevant to the understanding of the developmental origin of neuroblastoma. It has been suggested that some immature sympathetic cells or their immediate progenitors might be a point of tumor origin (5). Nevertheless, the current ambiguity as to what should be considered the actual differentiation sequence (13; 12; 6; 4; 22) complicates the debate. It seems clear that the generation of neuroblastoma cells can not be resumed as a simple differentiation blockade at an immature SCP state. In any case, our analysis advocates for a closer look at the potential role that S100B might play. The S100B is part of the S100 family of calcium-binding proteins, characterized by an "E" helix-loop-"F" helix (EF-hand) structure. It is expressed in a variety of glial cells, specific neuronal cells, and different peripheral cells (31). Ambiguous results have been obtained regarding its possible involvement in neuroblastomas: it has been shown to induce both cell proliferation (32) and cell-cycle arrest (33) on neuroblastoma cells in culture. Furthermore, it has been identified as one of the 6 genes constituting a prediction signature for adverse outcomes in patients with neuroblastoma (34). Clarifying the potential role played by S100B in vivo will therefore require additional studies.

There are of course clear limitations to this approach. Although we describe an informed gene selection procedure, it is dependent upon the use of a given arbitrary threshold in the number of gene selected. The use of a double selection procedure (using both Kantorovich distances as well as entropy) might somehow alleviate that effect, but we can not rule out that selecting more genes would have led to the incorporation of biologically relevant genes. We would also like to emphasize that contrary to most, if not all, GRN inference algorithms our gene selection procedure is completely blind to the function of the selected genes and relies purely on the time-dependent evolution of their mRNA distributions. This notably implies that many genes selected are not transcription factors, a very clear difference with most GRN inference, where regulatory proteins within a GRN are restricted to transcription factors (TF), like in (35; 36; 37). Possible indirect interactions are completely ignored. A trivial example is a gene encoding a protein that induces the nuclear translocation of a constitutive TF. In this case, the regulator gene will indirectly regulate TF target genes, and its effect will be crucial in understanding the GRN behavior. In our case, assuming a time-scale separation, we reasoned that only the genes that will display a change in their expression pattern (i.e. their mRNA distribution at the single-cell level) within the time frame of the experimental sampling can be captured and used for GRN inference, based upon single-cell transcriptomics. This means that most of the interactions that our GRN display are very probably indirect ones, and molecularly mediated by changes that act at scales that are not captured by the sampling process (like very fast protein phosphorylation for example). Edges in our GRN should therefore be interpreted with caution and elucidating their molecular nature would require dedicated efforts using different techniques and sampling strategies.

To sum up, we have proposed an approach for both inferring and simulating GRN based upon single-cell transcriptomics data generated to sample an *in vivo* differentiation process. We have shown that the absence of a time-stamped process in the dataset can be overcome by *ad hoc* measures. There is little

doubt that given the flood of such dataset in the literature, the use of our approach could lead to a better understanding of the mechanisms operating during cell decision-making.

Material and methods

Data preparation

The adrenal.human.seurat.scrublet.rds dataset was dowloaded as of October 2024 from http://pklab.med.harvard.edu/artem/adrenal/data/Seurat/.
This represent single-cell transcriptomic analysis using 10x Chromium on isolated individual cells from dissected adrenal glands with surrounding tissue from human embryos at post-conception weeks 6, 8, 9, 11, 12 and 14. The initial matrix displayed the expression of 25787 genes across 72571 cells. Only cells harboring a cell doublet probability (Scrublet score) < 0.2 were kept.
The cell types were annotated by the authors based upon gene expression pattern.
To elucidate the sympathoadrenal cell fate transitions in humans at higher resolution, we focused on the 3,901 cells of SCP, chromaffin and sympathetic fates, omitting annotated cell cycle genes. As advocated by the authors, we also removed a small potentially contaminating population of cells expressing the STAR gene. This resulted in a final matrix displaying the expression level of 20807 genes in 3759 cells at 6 developmental times points. We verified that the UMAP generated from this dataset was in perfect agreement with the Figure 2a of the original paper (Supplementary Figure 1a). UMAP was computed using the `uwot` package (38). All UMAPs were embedded within the experimental UMAP using the combination of `ret_model` and `umap_transform` functions of the `uwot` package.

Stemness score

The stemness score was computed per cell as signature_SCP / signature_SCP + signature_sympatoblast + signature_chromaffin, by summing up the gene expression level for the following signature genes:

- SCP_genes : FOX10, PLP1, FOXD3, FABP7, S100B, NGFR, ERBB3, MBP, MPZ, COL2A1, POSTN, MOXD1 and GAS7
- sympatoblast_genes : STMN2, HAND2, ELAVL4, STMN4, ISL1, PRPH, ELAVL2 and HMX1.
- chromaffin_genes: CHGA, CHGB, INSM1, PENK, PNMT and SLC35D3.

GRN inference and representation

We first cloned the original CARDAMOM git repository at https://github.com/ogandril/cardamom. All computation were made on the IN2P3 computing center (https://cc.in2p3.fr). The scripts used for launching CARDAMOM on a distant server can be found at https://github.com/ogandril/Launching_Cardamom. All interaction values were scaled by a factor 10 and thresholded at a value of 4 to keep only the most relevant interactions.
The resulting inter.csv interaction matrix was open with R (39) and plotted using the `qgraph` package (40). A list of transcription factors was obtained from https://github.com/saezlab/CollecTRI (41). Violin plots were drawn using the `ggbeeswarm` package.

GRN simulation

The simulation was performed using the CARDAMOM build-in version of HARISSA, the software for performing GRN simulations that is based upon an underlying stochastic dynamical model driven by the transcriptional bursting phenomenon (42).
The dynamics of each gene is given by the following ordinary differential equations:

$$\dot{M}_i = s_{0,i}E_i - d_{0,i}M_i$$
$$\dot{P}_i = s_{1,i}M_i - d_{1,i}P_i$$

with M_i being the number of mRNA molecules for gene i, P_i the number of proteins for gene I, E_i the state of the promoter, $s_{0,i}$ the mRNA synthesis rate, $d_{0,i}$ the mRNA degradation rate, $s_{1,i}$ the protein synthesis rate and $d_{1,i}$ the protein degradation rate,
E_i can randomly switch from 0 to 1 (activation) with a protein-dependent rate $k_{on,i}$ (P), and from 1 to 0 (inactivation) with constant rate $k_{off,i}$.
We define the protein-dependent rate function $k_{on,i}$ as in (42):

$$k_{on,i}(P) = \frac{k_{min,i} + k_{max,i} exp\left(\beta_i + \sum_j \theta_{j,i} P_j\right)}{1 + exp\left(\beta_i + \sum_j \theta_{j,i} P_j\right)}$$

such that the promoter activation rate is comprised between $k_{min,i}$ and $k_{max,i}$. The parameter β_i represents the basal activity of gene i, and the parameter $\theta_{j,i}$ encodes the value of the j → i interaction.
Since we simulate a mechanistic model, one needs to provide HARISSA with half-lives for both mRNAs ($d_{0,i}$) and proteins ($d_{1,i}$). Experimentally-determined values for mRNAs and proteins were extracted from (43) and (44), respectively. The genes that were not found in those dataset were attributed the mean value of experimentally-determined half-lives. Degradation rates were obtained through the following formula: degradation_rate = ln(2)/half_life. The half-lives of proteins were limited by a parameter called cell-cycle that was set to 20h.
Color code for a rapid estimation of the fit quality was as follows: given all Kantorovich distances for all genes at all times, we considered that 40 % of the values should be considered as correct (after a careful examination of the marginals, see Supplementary Figure 4) and therefore colored in green. The remaining 60 % was split in two, the first half colored in orange and the last half in red.

Gene perturbation
In order to perform the knock-out of gene i, we imposed a basal activity β_i (see upper) of -1000 and a $k_{min,i}$ value of 0. In order to simulate the overexpression of gene i, we imposed a basal activity β_i of +100 and a $k_{max,i}$ value of 5 (for CHGA, CHGB, STMN2 and HMGB2) or 20 (for POSTN and S100B genes).

Acknoledgments
I am very grateful to Elias Ventre, Ulysse Herbach and Matteo Bouvier for their continuous support and their help in setting up and improving the CARDAMOM/HARISSA framework.
I warmly thank Elias Ventre and Sandrine Gonin-Giraud for their critical reading of the manuscript.
I thank the computational center of IN2P3 (Villeurbanne/France) where computations were performed. I thank the BioSyL Federation and the LabEx Ecofect (ANR-11-LABX-0048) of the University of Lyon for inspiring scientific events.
This work was not financed by the Association Pour la Recherche contre le Cancer (ARC).

Figures

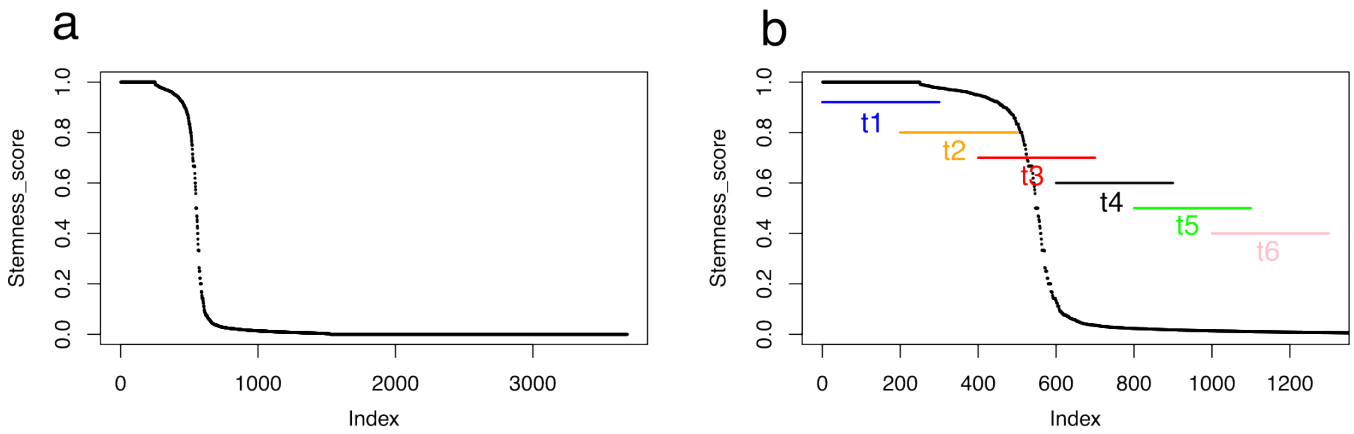


Figure 1: Cell ordering by their stemness score. a: all cells are displayed. b: Focus on the first 1300 cells, where the "process time" are indicated. The process time is defined along six time point with sliding windows of 300 cells overlapping by 100 cells.

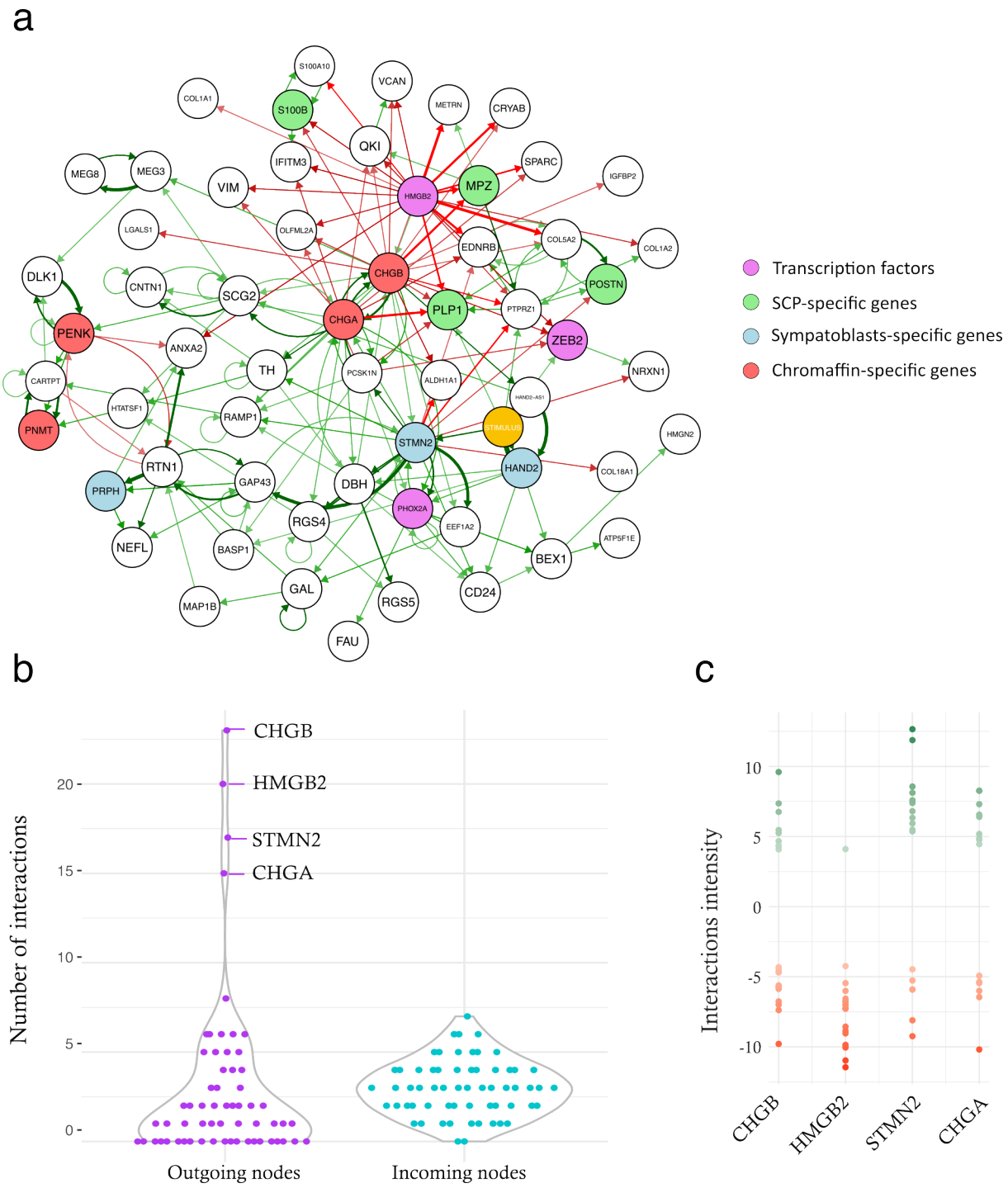


Figure 2: Representation of the inferred GRN structure. a: the full inferred GRN. Activations are displayed in green, inhibitions are displayed in red. The strength of the interaction is reflected by the arrow thickness. On the right is shown the color-code that has been applied to the GRN nodes. The stimulus is highlighted in yellow. b: violin plot representing the number of outgoing or incoming

interactions per gene. c: the actual values of the outgoing interactions are shown for the four genes with the highest number of interactions.

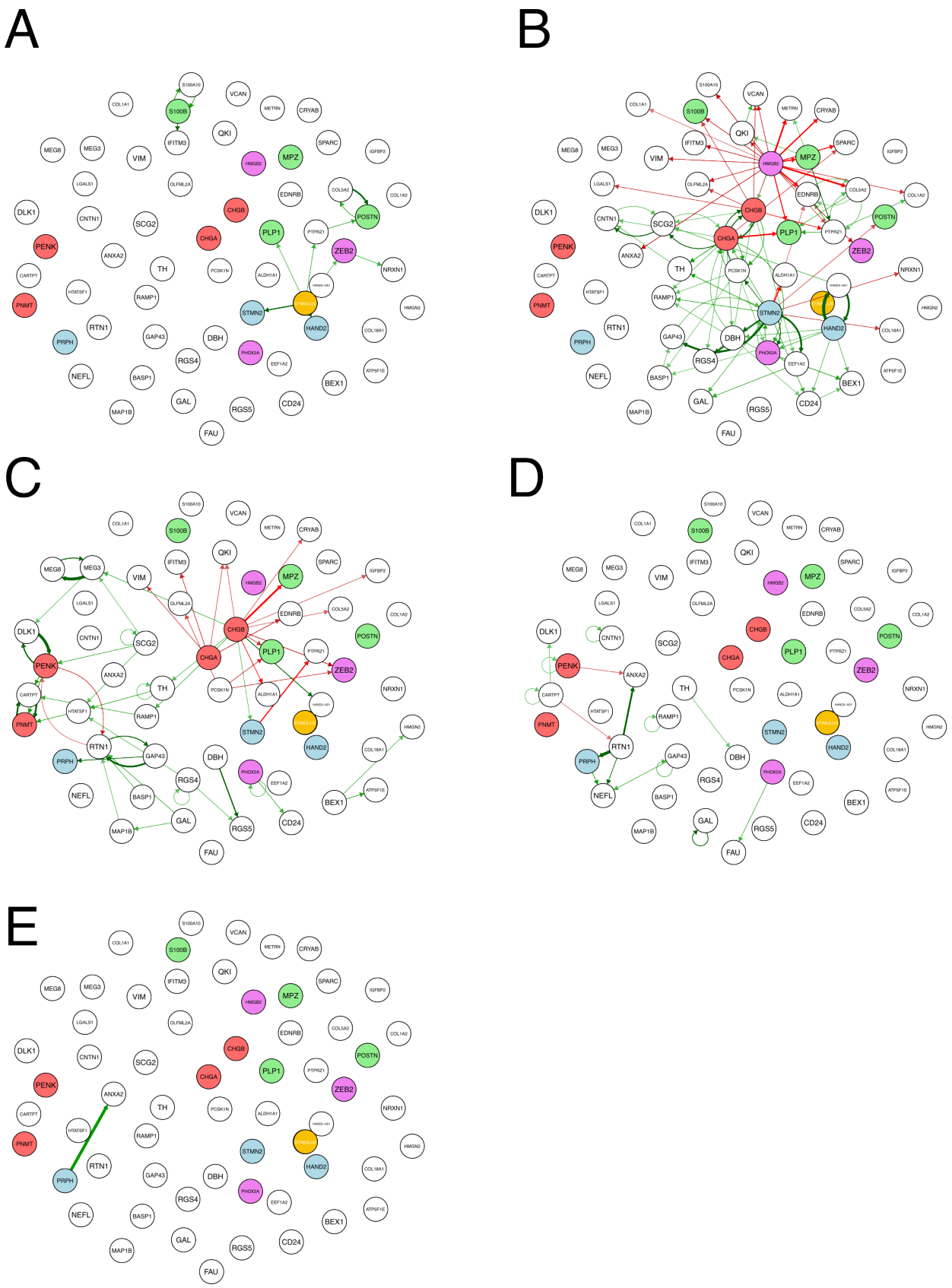


Figure 3: Representation of the time-dependent evolution of the inferred GRN structure. a: t1 → t2; b: t2 → t3; c: t3 → t4; d: t4 → t5; e: t5 → t6. For the time definition, see Figure 1). For the color-code applied

to the GRN nodes, see Figure 2a.

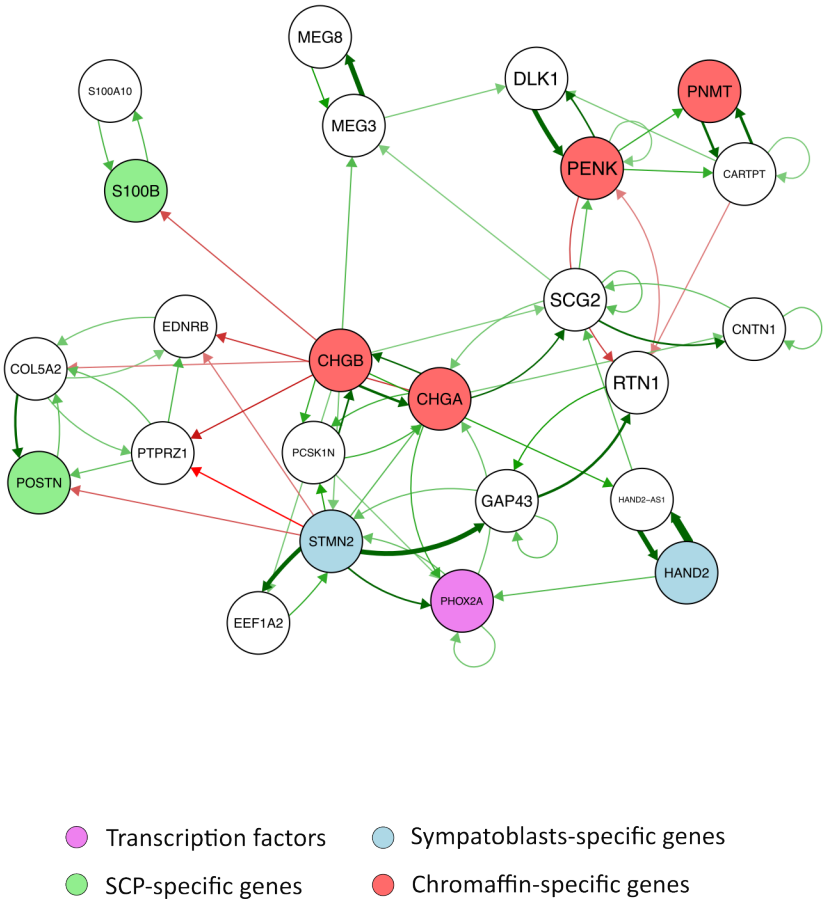


Figure 4: Representation of a subset of the GRN showing genes involved in specific dynamical motifs.

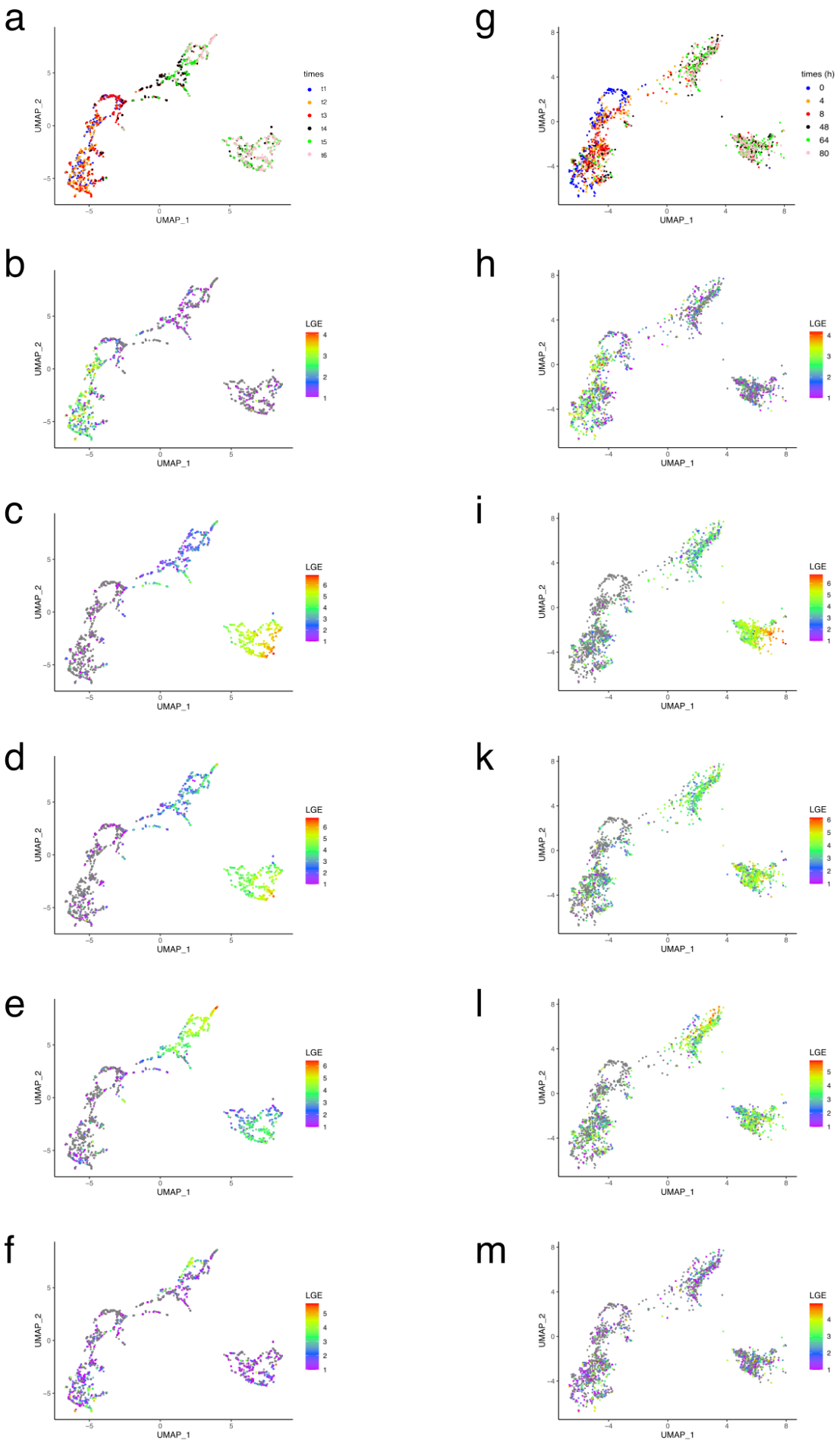

a
b
c
d
e
f
g
h
i
k
l
m
UMAP_1
UMAP_2
times
t1
t2
t3
t4
t5
t6
times (h)
0
4
8
48
64
80
LGE

Figure 5: UMAP representation of the dataset used for GRN inference (a-f) and of the dataset obtained through GRN simulation (g-m). The cells are colored coded according to: (i) time (a and g); (ii) ZEB2 expression level (b and h); (iii) CHGA expression level (c and i); (iv): CHGB expression level (d and k); (v): STMN2 expression level (e and l) and (vi): HMGB2 expression level (f and m). All gene expression are log-transformed (LGE= ln(x+1)).

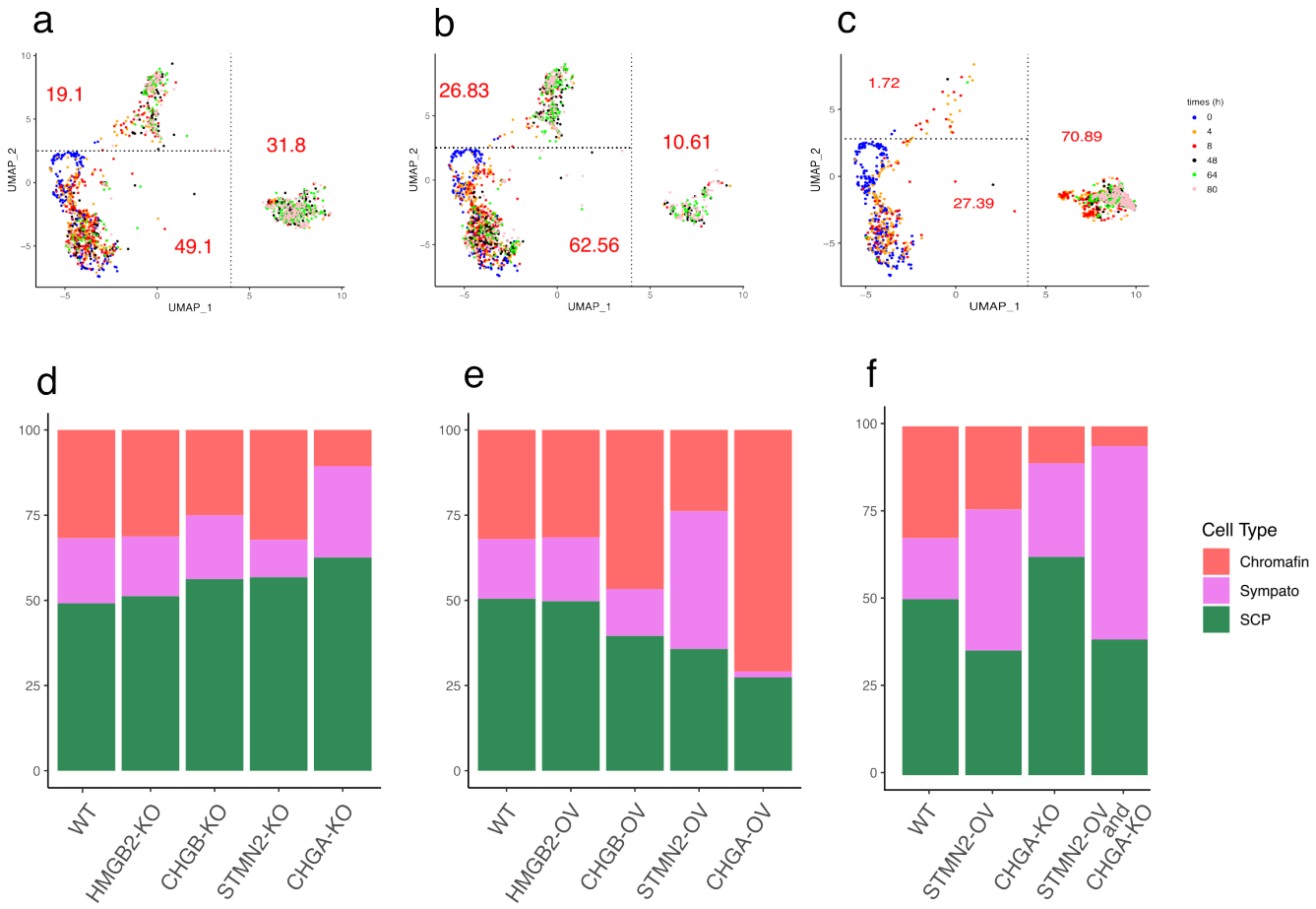


Figure 6: Analysis of the effect of the *in silico* perturbation of 4 genes on the output of the simulation. The UMAP space was divided in three regions, representing SCP (bottom left), sympathoblasts (top left) and chromaffin cells (right part). In a, b and c are shown the percent of cells belonging to each category when simulating the WT GRN (a), the CHGA-KO GRN (b), or the CHGA-OV GRN (c).
In d and e are shown the percent of cells belonging to each category for all 4 genes tested, either through a KO (d) or an overexpression (e). In f is shown either the simple or the double perturbation, associating the KO of the CHGA gene with the overexpression of the STMN2 gene.

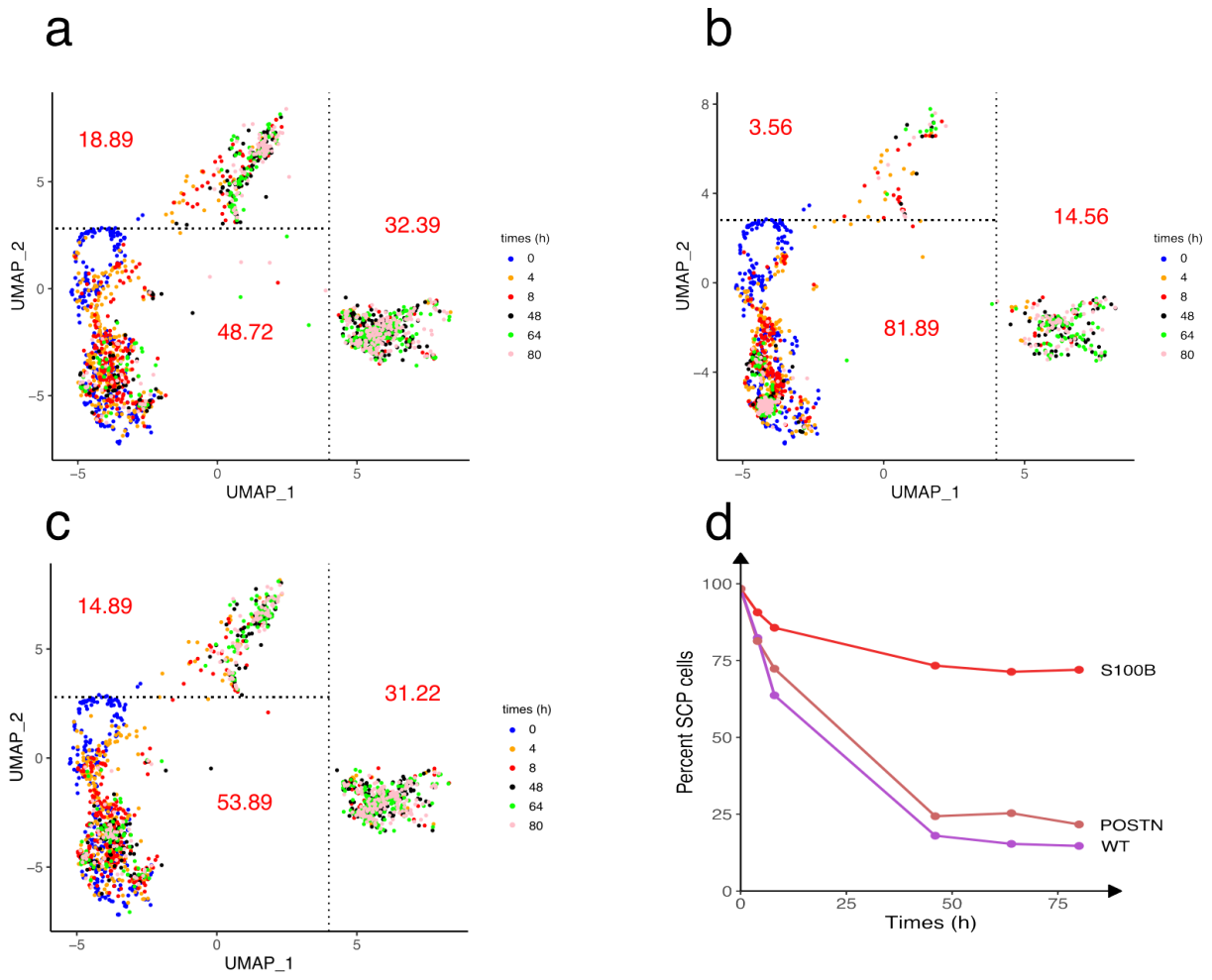
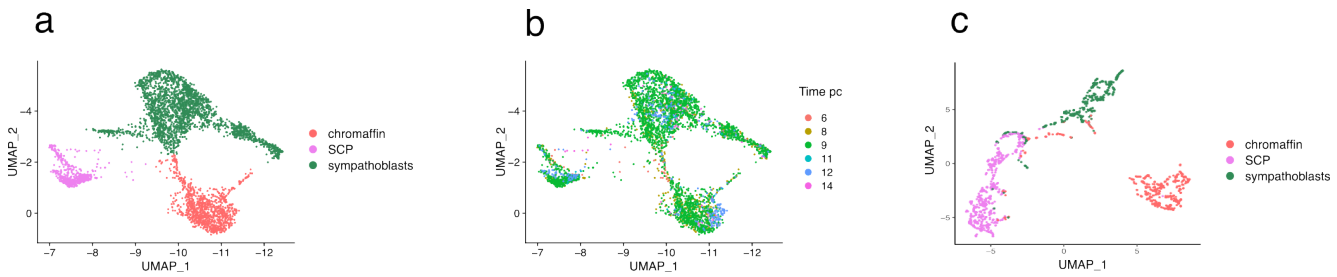
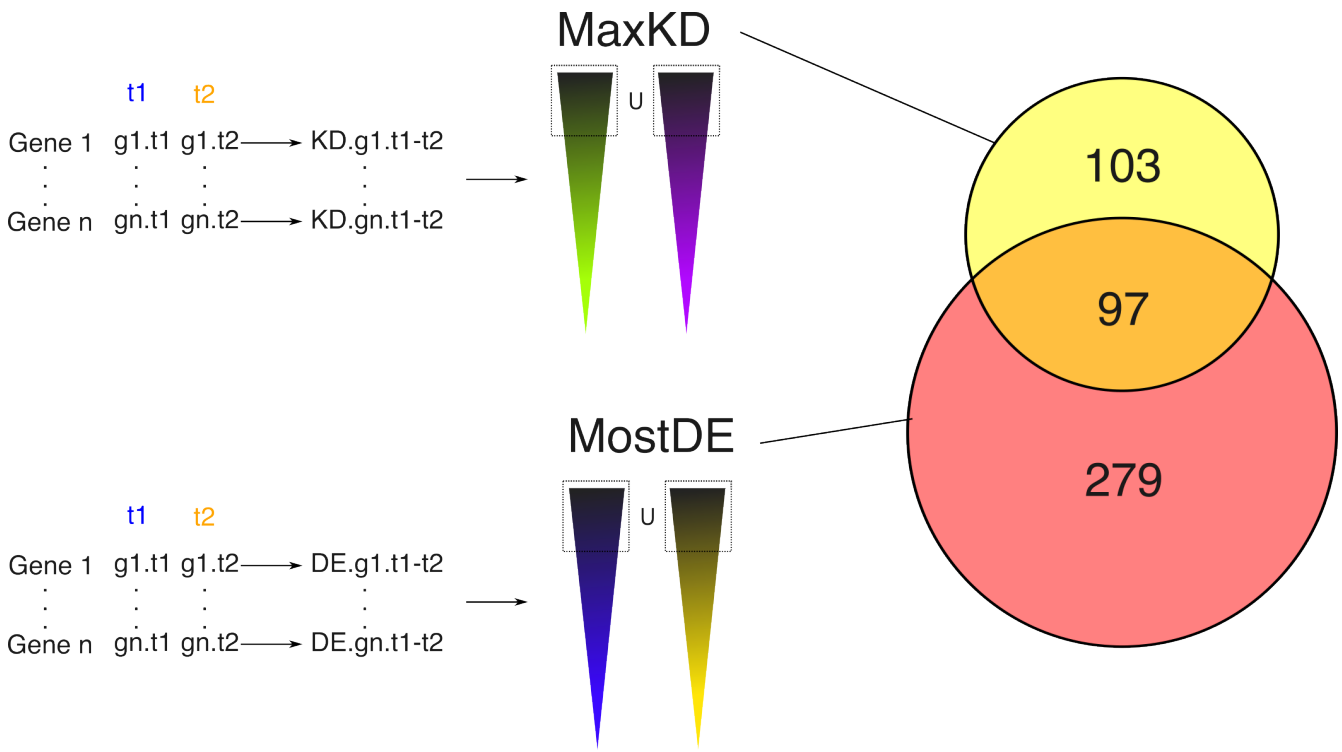


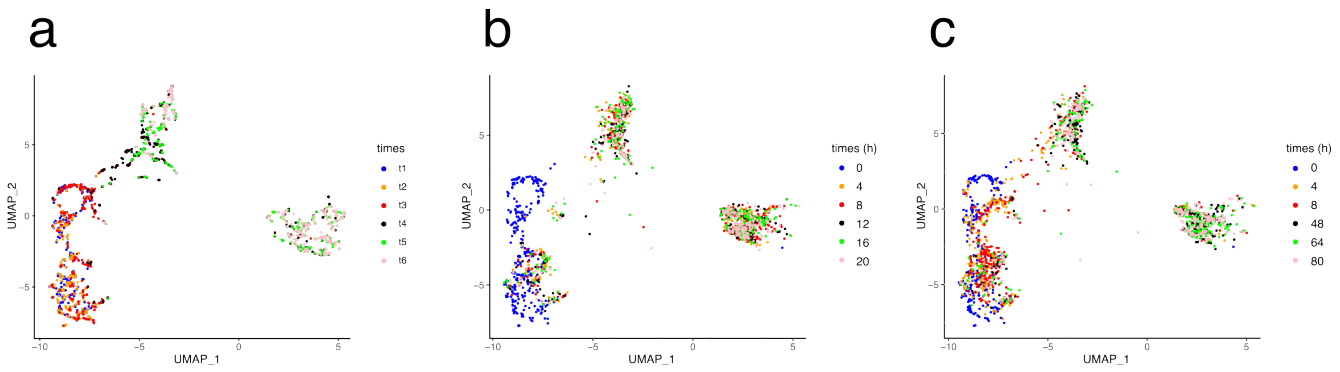
Figure 7: Analysis of the effect of the *in silico* perturbation of 2 genes on the output of the simulation. The UMAP space was divided in three regions, representing SCP (bottom left), sympathoblasts (top left) and chromaffin cells (right part). In a, b and c are shown the percent of cells belonging to each category when simulating the WT GRN (a), the S100B overexpression GRN (b), or the POSTN overexpression GRN (c). In d is shown the percentage of cells in the SCP regions, as a function of time, either in the WT, POSTN overexpression or S100B overexpression.



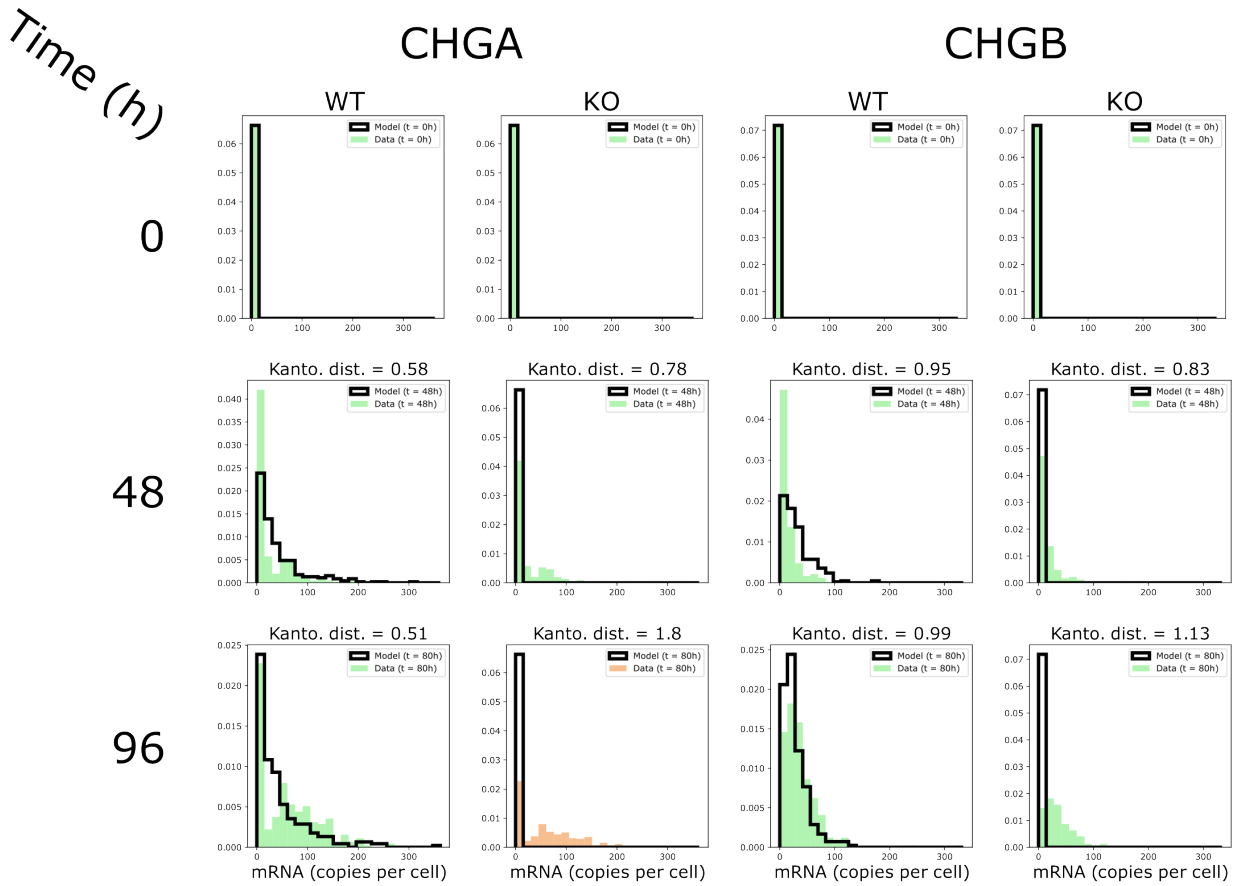
Supplementary Figure 1: UMAP representation of the dataset. a and b: the initial 3759 cells x 20807 genes matrix. c: the final 1800 cell x 97 genes dataset after time sampling (see Figure 1). In a and c cells are color-coded according to their cell fate. In b, cells are color-coded according to their sampling time.

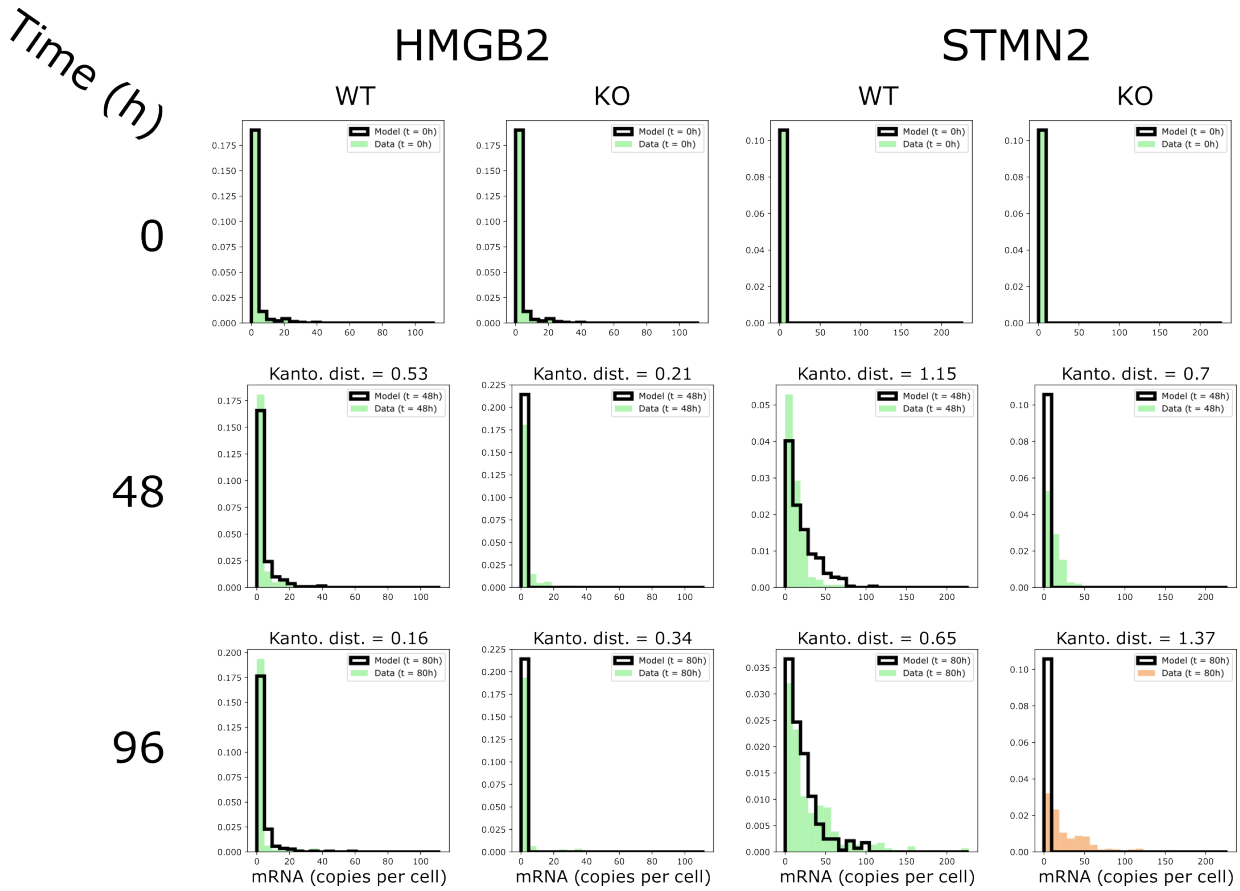


Supplementary Figure 2: Gene selection procedure. We first computed a Kantorovitch distance for all genes between 2 consecutive time points. We then took the union of the 133 genes harboring the largest distance for each delta-time, which left us with a list of 200 genes (some genes might be among the most distant between two or more time points), labelled MaxKD. Very similarly, we computed an entropy value for all genes between 2 consecutive time points. We then took the union of the 50 genes harboring the largest entropy for each delta-time, which left us with a list of 376 genes labelled MostDE. The union of those two lists gave us a list of 97 genes (Supplementary table 1) that were used for inferring the GRN.

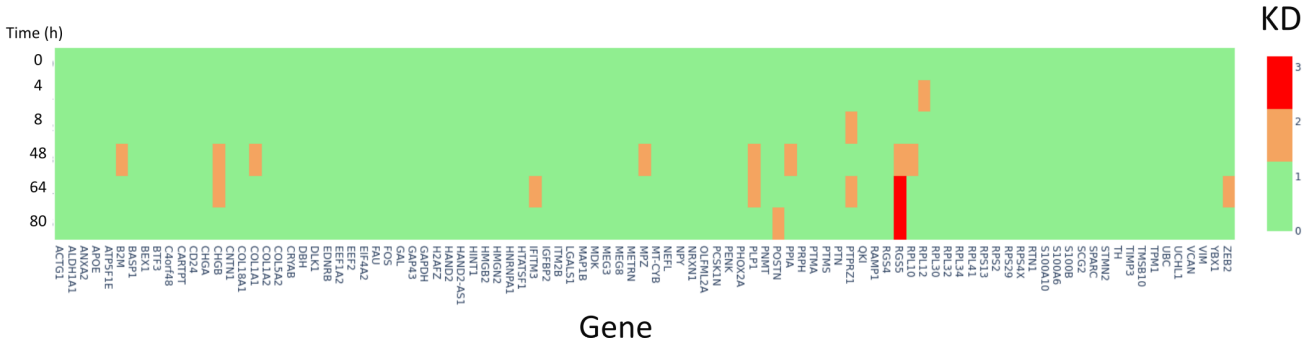


Supplementary Figure 3: Adjusting the process time to real time values. In a, the UMAP is computed on the experimental dataset and color coded according to the process times (see Figure1). b and c: the UMAP is computed on the simulated dataset and color coded according to the model time, in hours.

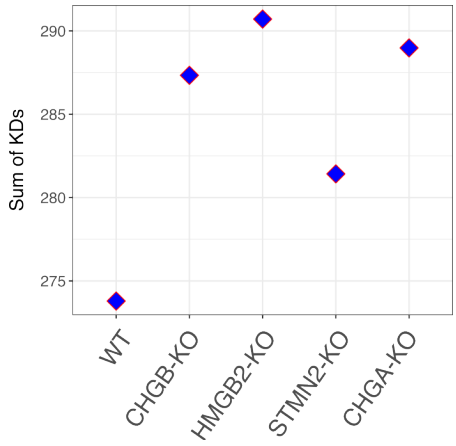

Time (h)
CHGA
CHGB
WT
KO
WT
KO
0
48
96
Model (t = 0h)
Data (t = 0h)
Kanto. dist. = 0.58
Kanto. dist. = 0.78
Kanto. dist. = 0.95
Kanto. dist. = 0.83
Model (t = 48h)
Data (t = 48h)
Kanto. dist. = 0.51
Kanto. dist. = 1.8
Kanto. dist. = 0.99
Kanto. dist. = 1.13
Model (t = 80h)
Data (t = 80h)
mRNA (copies per cell)


Time (h)
HMGB2
STMN2
WT
KO
WT
KO
0
48
96
Model (t = 0h)
Data (t = 0h)
Kanto. dist. = 0.53
Kanto. dist. = 0.21
Kanto. dist. = 1.15
Kanto. dist. = 0.7
Model (t = 48h)
Data (t = 48h)
Kanto. dist. = 0.16
Kanto. dist. = 0.34
Kanto. dist. = 0.65
Kanto. dist. = 1.37
Model (t = 80h)
Data (t = 80h)
mRNA (copies per cell)

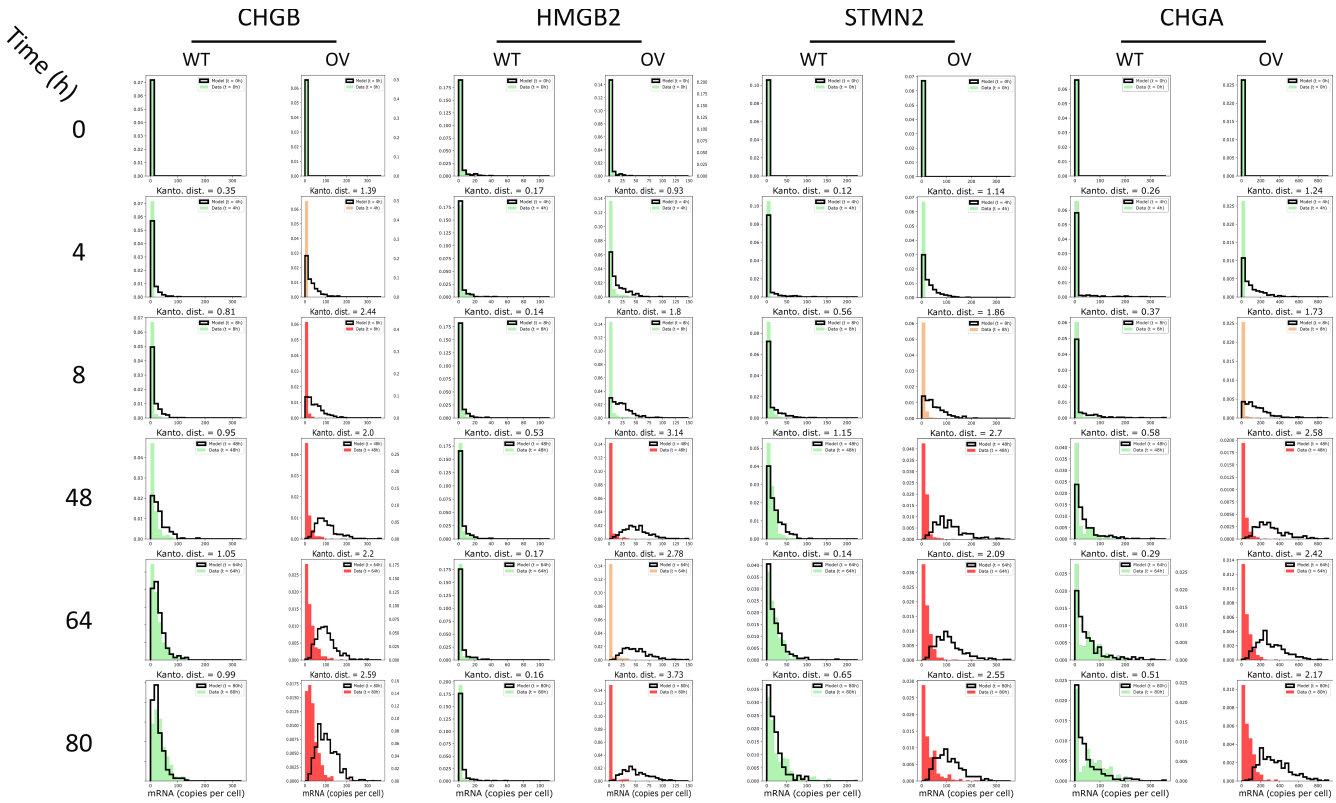
Supplementary Figure 4: Comparison between the experimentally observed mRNA distribution (black-contoured) with the simulated distribution (green- or orange-colored plain bars). Shown are the values for 4 selected genes for three times points, both in the wild-type (WT) or in the gene knock-out (KO) context. On top of each square the Kantorovich distance between the two distribution is shown. The color code gives a rough estimate of the quality of the fit (the inverse of the distance).



Supplementary Figure 5: Fit quality for each gene at each time point as assessed by the Kantorovich distance between the experimentally observed mRNA distribution and the simulated distribution.



Supplementary Figure 6: Overall fit quality by summing Kantorovich distances for all gene at all time points after simulating the wild-type (WT) or after performing the knock-out of selected genes.



Supplementary Figure 7: Comparison between the experimentally observed mRNA distribution (black-contoured) with the simulated distribution (green- or orange or red-colored plain bars). Shown are the values for 4 selected genes for all times points, both in the wild-type (WT) or in the gene overexpression (OV) context. On top of each square the Kantorovich distance between the two distribution is shown. The color code gives a rough estimate of the quality of the fit (the inverse of the distance).

ACTG1	DBH	IGFBP2	PPIA	RTN1
ALDH1A1	DLK1	ITM2B	PRPH	S100A10
ANXA2	EDNRB	LGALS1	PTMA	S100A6
APOE	EEF1A2	MAP1B	PTMS	S100B
ATP5F1E	EEF2	MDK	PTN	SCG2
B2M	EIF4A2	MEG3	PTPRZ1	SPARC
BASP1	FAU	MEG8	QKI	STMN2
BEX1	FOS	METRN	RAMP1	TH
BTF3	GAL	MPZ	RGS4	TIMP3
C4orf48	GAP43	MT-CYB	RGS5	TMSB10
CARTPT	GAPDH	NEFL	RPL10	TPM1
CD24	H2AFZ	NPY	RPL12	UBC
CHGA	HAND2	NRXN1	RPL30	UCHL1
CHGB	HAND2-AS1	OLFML2A	RPL32	VCAN
CNTN1	HINT1	PCSK1N	RPL34	VIM
COL18A1	HMGB2	PENK	RPL41	YBX1
COL1A1	HMGN2	PHOX2A	RPS13	ZEB2
COL1A2	HNRNPA1	PLP1	RPS2	
COL5A2	HTATSF1	PNMT	RPS29	
CRYAB	IFITM3	POSTN	RPS4X	

Supplementary table 1: the 97 genes that were used for inferring the GRN

References

1. Gomez RL, Ibragimova S, Ramachandran R, Philpott A, Ali FR: **Tumoral heterogeneity in neuroblastoma**. *Biochim Biophys Acta Rev Cancer* 2022, **1877**(6):188805.
2. Almstedt E, Elgendy R, Hekmati N, Rosen E, Warn C, Olsen TK, Dyberg C, Doroszko M, Larsson I, Sundstrom A, Arsenian Henriksson M, Pahlman S, Bexell D, Vanlandewijck M, Kogner P, Jornsten R, Krona C, Nelander S: **Integrative discovery of treatments for high-risk neuroblastoma**. *Nat Commun* 2020, **11**(1):71.
3. Li CH, Sharma S, Heczey AA, Woods ML, Steffin DHM, Louis CU, Grilley BJ, Thakkar SG, Wu M, Wang T, Rooney CM, Brenner MK, Heslop HE: **Long-term outcomes of GD2-directed CAR-T cell therapy in patients with neuroblastoma**. *Nat Med* 2025.
4. Zeineldin M, Patel AG, Dyer MA: **Neuroblastoma: When differentiation goes awry**. *Neuron* 2022, **110**(18):2916-2928.
5. Ratner N, Brodeur GM, Dale RC, Schor NF: **The "neuro" of neuroblastoma: Neuroblastoma as a neurodevelopmental disorder**. *Ann Neurol* 2016, **80**(1):13-23.
6. Ponzoni M, Bachetti T, Corrias MV, Brignole C, Pastorino F, Calarco E, Bensa V, Giusto E, Ceccherini I, Perri P: **Recent advances in the developmental origin of neuroblastoma: an overview**. *J Exp Clin Cancer Res* 2022, **41**(1):92.
7. Patel AG, Ashenberg O, Collins NB, Segerstolpe Å, Jiang S, Slyper M, Huang X, Caraccio C, Jin H, Sheppard H, Xu K, Chang TC, Orr BA, Shirinifard A, Chapple RH, Shen A, Clay MR, Tatevossian RG, Reilly C, Patel J, Lupo M, Cline C, Dionne D, Porter CBM, Waldman J, Bai Y, Zhu B, Barrera I, Murray E, Vigneau S, Napolitano S, Wakiro I, Wu J, Grimaldi G, Dellostritto L, Helvie K, Rotem A, Lako A, Cullen N, Pfaff KL, Karlström Å, Jané-Valbuena J, Todres E, Thorner A, Geeleher P, Rodig SJ, Zhou X, Stewart E, Johnson BE, Wu G, Chen F, Yu J, Goltsev Y, Nolan GP, Rozenblatt-Rosen O, Regev A, Dyer MA: **A spatial cell atlas of neuroblastoma reveals developmental, epigenetic and spatial axis of tumor heterogeneity**. *bioRxiv* 2024.
8. Del Valle I, Buonocore F, Duncan AJ, Lin L, Barenco M, Parnaik R, Shah S, Hubank M, Gerrelli D, Achermann JC: **A genomic atlas of human adrenal and gonad development**. *Wellcome Open Res* 2017, **2**:25.
9. Furlan A, Dyachuk V, Kastriti ME, Calvo-Enrique L, Abdo H, Hadjab S, Chontorotzea T, Akkuratova N, Usoskin D, Kamenev D, Petersen J, Sunadome K, Memic F, Marklund U, Fried K, Topilko P, Lallemend F, Kharchenko PV, Ernfors P, Adameyko I: **Multipotent peripheral glial cells generate neuroendocrine cells of the adrenal medulla**. *Science* 2017, **357**(6346).
10. Kameneva P, Artemov AV, Kastriti ME, Faure L, Olsen TK, Otte J, Erickson A, Semsch B, Andersson ER, Ratz M, Frisen J, Tischler AS, de Krijger RR, Bouderlique T, Akkuratova N, Vorontsova M, Gusev O, Fried K, Sundstrom E, Mei S, Kogner P, Baryawno N, Kharchenko PV, Adameyko I: **Single-cell transcriptomics of human embryos identifies multiple sympathoblast lineages with potential implications for neuroblastoma origin**. *Nat Genet* 2021, **53**(5):694-706.
11. Rasmuson A, Segerstrom L, Nethander M, Finnman J, Elfman LH, Javanmardi N, Nilsson S, Johnsen JI, Martinsson T, Kogner P: **Tumor development, growth characteristics and spectrum of genetic aberrations in the TH-MYCN mouse model of neuroblastoma**. *PLoS One* 2012, **7**(12):e51297.
12. Kildisiute G, Kholosy WM, Young MD, Roberts K, Elmentaite R, van Hooff SR, Pacyna CN, Khabirova E, Piapi A, Thevanesan C, Bugallo-Blanco E, Burke C, Mamanova L, Keller KM, Langenberg-Ververgaert KPS, Lijnzaad P, Margaritis T, Holstege FCP, Tas ML, Wijnen M, van Noesel MM, Del Valle I, Barone G, van der Linden R, Duncan C, Anderson J, Achermann JC, Haniffa M, Teichmann SA, Rampling D, Sebire NJ, He X, de Krijger RR, Barker RA, Meyer KB, Bayraktar O, Straathof K, Molenaar JJ, Behjati S: **Tumor to normal single-cell mRNA comparisons reveal a pan-neuroblastoma cancer cell**. *Sci Adv* 2021, **7**(6).

13. Jansky S, Sharma AK, Korber V, Quintero A, Toprak UH, Wecht EM, Gartlgruber M, Greco A, Chomsky E, Grunewald TGP, Henrich KO, Tanay A, Herrmann C, Hofer T, Westermann F: **Single-cell transcriptomic analyses provide insights into the developmental origins of neuroblastoma**. *Nat Genet* 2021, **53**(5):683-693.
14. La Manno G, Soldatov R, Zeisel A, Braun E, Hochgerner H, Petukhov V, Lidschreiber K, Kastriti ME, Lonnerberg P, Furlan A, Fan J, Borm LE, Liu Z, van Bruggen D, Guo J, He X, Barker R, Sundstrom E, Castelo-Branco G, Cramer P, Adameyko I, Linnarsson S, Kharchenko PV: **RNA velocity of single cells**. *Nature* 2018, **560**(7719):494-498.
15. Bergen V, Lange M, Peidli S, Wolf FA, Theis FJ: **Generalizing RNA velocity to transient cell states through dynamical modeling**. *bioRxiv* 2019:820936.
16. Gorin G, Fang M, Chari T, L P: **RNA velocity unraveled.** . *PLoS Comput Biol* 2022, **18(9): e1010492.**
17. Fang M, Gorin G, Pachter L: **Trajectory inference from single-cell genomics data with a process time model**. *PLoS Comput Biol* 2025, **21**(1):e1012752.
18. Ventre E, Herbach U, Espinasse T, Benoit G, Gandrillon O: **One model fits all: Combining inference and simulation of gene regulatory networks**. *PLoS Comput Biol* 2023, **19**(3):e1010962.
19. Ventre E: **Reverse engineering of a mechanistic model of gene expression using metastability and temporal dynamics**. *In Silico Biology* 2020, **14**:89-113.
20. Herbach U, Bonnaffoux A, Espinasse T, Gandrillon O: **Inferring gene regulatory networks from single-cell data: a mechanistic approach**. *BMC Systems Biology* 2017, **11**(1):105.
21. Ventre E, Espinasse T, Bréhier C-E, Calvez V, Lepoutre T, Gandrillon O: **Reduction of a stochastic model of gene expression: Lagrangian dynamics gives access to basins of attraction as cell types and metastabilty**. *Journal of Mathematical Biology* 2021, **83**(5):59.
22. Sehgal M, Nayak SP, Sahoo S, Somarelli JA, Jolly MK: **Mutually exclusive teams-like patterns of gene regulation characterize phenotypic heterogeneity along the noradrenergic-mesenchymal axis in neuroblastoma**. *Cancer Biol Ther* 2024, **25**(1):2301802.
23. Moris N, Pina C, Arias AM: **Transition states and cell fate decisions in epigenetic landscapes**. *Nat Rev Genet* 2016, **17**(11):693-703.
24. Mar JC: **The rise of the distributions: why non-normality is important for understanding the transcriptome and beyond**. *Biophys Rev* 2019:89–94.
25. Vershik AM: **Kantorovich metric: initial history and little-known applications**. *Journal of Mathematical Sciences* 2006, **133**(4):1410-1417.
26. Dussiau C, Boussaroque A, Gaillard M, Bravetti C, Zaroili L, Knosp C, Friedrich C, Asquier P, Willems L, Quint L, Bouscary D, Fontenay M, Espinasse T, Plesa A, Sujobert P, Gandrillon O, Kosmider O: **Hematopoietic differentiation is characterized by a transient peak of entropy at a single-cell level**. *BMC Biology* 2022, **20**(1):60.
27. Bonnaffoux A, Herbach U, Richard A, Guillemin A, Gonin-Giraud S, Gros P-A, Gandrillon O: **WASABI: a dynamic iterative framework for gene regulatory network inference**. *BMC Bioinformatics* 2019, **20**(1):220.
28. Fisher J, Henzinger TA: **Executable cell biology**. *Nat Biotechnol* 2007, **25**(11):1239-1249.
29. Hari K, Ullanat V, Balasubramanian A, Gopalan A, Jolly MK: **Landscape of epithelial-mesenchymal plasticity as an emergent property of coordinated teams in regulatory networks**. *eLife* 2022, **11**.
30. Shyam S, Nandan N, Jolly MK, Hari K: **Topological conditions of gene regulatory networks enabling robust binary cell-fate decision-making**. *bioRxiv* 2025:2025.2002.2007.636066.
31. Hernandez-Ortega K, Canul-Euan AA, Solis-Paredes JM, Borboa-Olivares H, Reyes-Munoz E, Estrada-Gutierrez G, Camacho-Arroyo I: **S100B actions on glial and neuronal cells in the developing brain: an overview**. *Front Neurosci* 2024, **18**:1425525.

32. Leclerc E, Fritz G, Weibel M, Heizmann CW, Galichet A: **S100B and S100A6 differentially modulate cell survival by interacting with distinct RAGE (receptor for advanced glycation end products) immunoglobulin domains**. *J Biol Chem* 2007, **282**(43):31317-31331.
33. Bernardini C, Lattanzi W, Businaro R, Leone S, Corvino V, Sorci G, Lauro G, Fumagalli L, Donato FR, Michetti F: **Transcritpional effects of S100B on neuroblastoma cells: perturbation of cholesterol homeostasis and interference on the cell cycle**. *Gene Expr* 2010, **14**(6):345-359.
34. Tan J, Wang C, Jin Y, Xia Y, Gong B, Zhao Q: **Optimal combination of MYCN differential gene and cellular senescence gene predicts adverse outcomes in patients with neuroblastoma**. *Front Immunol* 2023, **14**:1309138.
35. Matsumoto H, Kiryu H, Furusawa C, Ko MS, Ko SB, Gouda N, Hayashi T, Nikaido I: **SCODE: An efficient regulatory network inference algorithm from single-cell RNA-Seq during differentiation**. *Bioinformatics* 2017.
36. Sanchez-Castillo M, Blanco D, Tienda-Luna IM, Carrion MC, Huang Y: **A Bayesian framework for the inference of gene regulatory networks from time and pseudo-time series data**. *Bioinformatics* 2018, **34**(6):964-970.
37. Ocone A, Haghverdi L, Mueller NS, Theis FJ: **Reconstructing gene regulatory dynamics from high-dimensional single-cell snapshot data**. *Bioinformatics* 2015, **31**(12):i89-i96.
38. Melville J: **uwot: The Uniform Manifold Approximation and Projection (UMAP) Method for Dimensionality Reduction**. In.; 2024.
39. **R: A language and environment for statistical computing.** [http://www.R-project.org/.]
40. Epskamp S, Cramer AOJ, Waldorp LJ, Schmittmann VD, Borsboom D: **qgraph: Network Visualizations of Relationships in Psychometric Data**. *Journal of Statistical Software* 2012, **48**:1–18.
41. Muller-Dott S, Tsirvouli E, Vazquez M, Ramirez Flores RO, Badia IMP, Fallegger R, Turei D, Laegreid A, Saez-Rodriguez J: **Expanding the coverage of regulons from high-confidence prior knowledge for accurate estimation of transcription factor activities**. *Nucleic Acids Res* 2023, **51**(20):10934-10949.
42. Herbach U: **Harissa: Stochastic Simulation and Inference of Gene Regulatory Networks Based on Transcriptional Bursting**. In: *Computational Methods in Systems Biology: 2023// 2023; Cham*. Springer Nature Switzerland: 97-105.
43. Blumberg A, Zhao Y, Huang YF, Dukler N, Rice EJ, Chivu AG, Krumholz K, Danko CG, Siepel A: **Characterizing RNA stability genome-wide through combined analysis of PRO-seq and RNA-seq data**. *BMC Biol* 2021, **19**(1):30.
44. Li J, Cai Z, Vaites LP, Shen N, Mitchell DC, Huttlin EL, Paulo JA, Harry BL, Gygi SP: **Proteome-wide mapping of short-lived proteins in human cells**. *Mol Cell* 2021, **81**(22):4722-4735 e4725.