

## Report 7: Implementing A Cell Differentiation Modell

In the proceedings, we are implementing a model which simulates the time evolution of the protein and gene expression levels in stem cells and the subsequent differentiation of said cells, which is due to the intercellular interaction via protein diffusion. We will discuss the time evolution and go into detail on how and why differentiation occurs. The model as well as the parameters and the GRNs we simulate are taken from [1].

### 1 The cell differentiation model

We model each cell  $l$  as having two characteristics: The expression levels  $m^l(i, t)$  and  $p^l(i, t)$  of genes and proteins, respectively. Here,  $i$  denotes the gene (or protein) and  $t$  is the time; there are  $k$  genes and  $k$  proteins. The cells change their gene and protein expression over time according to two differential equations. The change in gene expression is governed by

$$\frac{d}{dt}m^l(i, t) = \gamma(F^l(i, t) - m^l(i, t)) \quad (1.1)$$

where

$$F^l(i, t) = f \left( \sum_j J_{ij} p^l(j, t) - \theta_i \right) \quad \text{and} \quad f = \frac{1}{1 + e^{-\beta x}}. \quad (1.2)$$

Here,  $J_{ij} \in \{0, \pm 1\}$  represents the Gene Regulatory Network we are simulating and  $\theta$  contains threshold parameters.  $J$  describes how different proteins increase or decrease each other's expression level in the sense that if  $J_{ij} = \pm 1$ , the presence of protein  $j$  activates / suppresses protein  $i$  (i.e. increases or decreases its expression level). If  $J_{ij} = 0$ , the presence of protein  $j$  does not affect protein  $i$ . The thresholds  $\theta_i$  are included to eliminate any artificial symmetries in the protein expression levels. The factor  $\gamma$  determines the time scale on which protein and gene expression levels change.

$f$  approximates the Heaviside Step Function  $\theta(x)$ : For large values of  $\beta$ ,  $f$  quickly approaches 1 as  $x \rightarrow \infty$  and 0 as  $x \rightarrow -\infty$ . Thus, if  $\sum_j J_{ij} p^l(j, t) > \theta_i$ ,  $F \approx 1$  and the expression of gene  $i$  increases. Otherwise, the gene expression decreases.

The protein expression dynamics are governed by

$$\frac{d}{dt}p^l(i, t) = m^l(i, t) - p^l(i, t) + D_i (P(i, t) - p^l(i, t)) \quad (1.3)$$

where

$$P(i, t) = \frac{1}{N} \sum_l p^l(i, t) \quad (1.4)$$

is the average concentration of protein in the medium between the cells and  $N$  is the number of cells. The first two terms in Equation 1.3 represent the protein synthesis from mRNA and its degradation. The third term represents diffusion through cell membranes;  $D_i$  is the diffusion constant for protein  $i$ .

We also consider the division of cells after specific time intervals; after  $t_{\text{div}}$ , every cell splits in two cells, doubling the total number of cells. The new cells show the protein and gene expressions from the cell they emerged from, we however introduce a little bit of noise in the cell division process. For each division, we take a random value  $\delta \in \{-\sigma, \sigma\}$  and add it to the protein expression of one of the new cells and subtract it from the other one.

#### 1.1 Selecting Networks which allow differentiation

To examine GRNs with respect to their potential for differentiation, we test if, after the simulation, different cell types emerged. We consider two cells to be of different type if their euclidian distance

$$d_{lk} = \sqrt{\sum_i (p^l(i, t) - p^k(i, t))^2} \quad (1.5)$$

in the protein space is above a certain threshold, which we choose to be 0.3.

## 2 Implementing the cell differentiation model

We simulate a system where  $k = 5$ , i.e. there are five genes and five proteins. The other free parameters are chosen to be

$$\begin{aligned}\beta &= 40 \\ \gamma &= 6 \\ \theta &= (-0.01 \quad -0.03 \quad 0.02 \quad 0.01 \quad -0.02) \\ D_j &= 0.4\delta_{ij} \\ \sigma &= 1 \times 10^{-3}.\end{aligned}$$

We choose only one protein to be diffusive, so  $D_j = 0$  for  $i \neq j$ . The GRN  $J_{ij}$  is varied throughout our simulations and will be defined accordingly. We only consider networks with ten paths, i.e. ten connections between proteins which either activate or suppress another protein.

We implemented the simulation in C++ and conduct the data analysis in python; the code can be found in the files `simulation.cpp` and `evaluation.ipynb`. Functions and structs are contained in the file `functions.h`, operator overloads and output functions are contained in `infrastructure.h`. We also included a header file `header.h`.

### 2.1 The Cell struct

The centerpiece of our simulation is the `Cell` struct

```
struct Cell{
    std::vector<double> p,m;

    std::vector<std::vector<double>> p_timeline;
    std::vector<std::vector<double>> m_timeline;

    void update(std::vector<double> Penv,
                std::vector<std::vector<double>> J);

    Cell();
    Cell(const Cell& oldcell,std::vector<double> noise);
};
```

which represents a cell with the current protein expression level `p` and the gene expression level `m`. It also holds all previous expression levels in the containers `p_timeline` and `m_timeline`. The constructor `Cell()` is used to initialize the first cell with random expression levels. The second constructor `Cell(const Cell& oldcell,std::vector<double> noise)` is called during cell division. It uses an existing cell and creates a new one with the same history and the same expression levels `p` and `m`, where `p` is perturbed by `noise`. All current cells are stored in the container `std::vector<Cell> cells`, which is initialized with the value `std::vector<Cell>(1,Cell())`, i.e. with one cell with random expression levels.

The member function `void Cell::update(std::vector<double> Penv, std::vector<std::vector<double>> J)` carries out the time evolution according to Equation 1.1 and Equation 1.3 by using finite time differences `double dt`; it updates `p` and `m` according to

$$\begin{aligned}p &= \text{old\_p} + ((\text{old\_m} - \text{old\_p}) + D * (\text{Penv} - \text{old\_p})) * dt; \\ m &= \text{old\_m} + \gamma * (F(J,\text{old\_p}) - \text{old\_m}) * dt;\end{aligned}$$

where `std::vector<double> Penv` is the protein concentration in the medium and `J` is the medium. `old_p` and `old_m` are the old expression levels of proteins and genes, respectively, taken from the histories of the respective cell. Note that the operators `+`, `-` and `*` have been overloaded to act on the `std::vector` datatype. The function

```
std::vector<double> F(std::vector<std::vector<double>> J,
                      std::vector<double> p);
```

implements the function  $F$  as described in Equation 1.2.

## 2.2 The simulation

The code carrying out the simulation is

```

while(cells.size() <= Nmax and t < tdiv){
    Penv = P(cells);

    // updating mRNA and protein concentrations
    for(auto itr=cells.begin();itr<cells.end();itr++){
        itr->update(Penv,J);
    }

    // time passed
    t++;

    // cell division, if enough time passed
    if(t >= tdiv and cells.size() < Nmax){
        cells = Mitosis(cells);
        t = 0;
    }
}

```

where `std::vector<Cell> Mitosis(std::vector<Cell> cells)` carries out cell division for all cells contained in the argument `cells`. The expression levels of all cells are updated first using the function `void Cell::update`. Afterwards, if enough updates have been performed, cell division is performed using the function `Cell Mitosis(std::vector<Cell> cells)`. The expression level updates are carried out `tdiv` times for every number of cells up until and including `Nmax`. Afterwards, all simulation parameters and the histories of `m` and `p` are saved in a text file.

The Gene Regulatory Network `std::vector<std::vector<double> > J ≡ Jij` is defined in `simulation.cpp`.

After having carried out the simulation, we examine if the cells differentiated, using the distance between cells from Equation 1.5. We are going to demonstrate that the cells often oscillate in the protein space during the differentiation process, so we actually use the average  $\bar{d}_{lk}$ , taken over 50 s. We are also going to demonstrate that, after differentiation, one or more of the proteins are typically fully expressed ( $p \approx 0$  or  $p \approx 1$ ) (see Section 3 for further details). For this reason, the threshold must not be particularly small; we choose a value of 0.3. We thus consider cells  $i$  and  $j$  to be of different type if

$$\overline{d_{ij}} = \frac{1}{\Delta t} \sum_{t=t_0}^{t_0+\Delta t} \sqrt{\sum_k (p^i(k, t) - p^j(k, t))^2} \leq 0.3. \quad (2.1)$$

We further assume that “being of the same type” is transitive, i.e. if  $j$  and  $k$  are of the same type and  $k$  and  $l$  are of the same type, then we also consider  $j$  and  $l$  to be of the same type. The code

```

# the array cell_types will contain an array for each cell type.
# cell_types[i] will contain all cells of the respective type
cell_types = []
threshold = .3
# clustering
for i in range(Nmax):
    matched = False
    # searching for a cell of matching type in the known cell types
    for cell_type in cell_types:
        for j in cell_type:
            if dist[i,j] < threshold:
                # we found a match in one of the existing cell types
                cell_type += [i,]
                matched = True
                break
    if matched: break

```

```

if not matched:
    # we did not find a match; the current cell has a new type
    cell_types += [[i],]

```

then groups the cells into different types. The array `dist` contains the distances in protein space such that  $\text{dist}[i,j] = d_{ij}$ . `dist` is calculated according to Equation 2.1, where we chose  $\Delta t = 50\text{ s}$ .

### 3 Results

We carried out simulations for four different GRNs, which are shown in Figure 3.1 to Figure 3.4. In the graphs on the right, the proteins are shown as nodes and the paths between them as edges. A red edge indicates that the presence of protein  $i$  stimulates protein  $j$ , meaning  $J_{ji} = 1$ . A black edge indicates suppression of the expression level, thus  $J_{ji} = -1$ . When there is no connection between two nodes, we set  $J_{ji} = 0$ .

$$J^{(a)} = \begin{pmatrix} 0 & 0 & 0 & 0 & 0 \\ -1 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 \\ 1 & 0 & -1 & 1 & -1 \\ 1 & 0 & -1 & 1 & -1 \end{pmatrix}$$

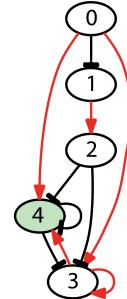


Figure 3.1: GRN (a), as graph and it's matrix representation. The history of it's protein expression levels is shown in Figure A.1.

$$J^{(b)} = \begin{pmatrix} 0 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 \\ 0 & 1 & -1 & 0 & 0 \\ 0 & -1 & 1 & 0 & 1 \\ 0 & 1 & -1 & -1 & 1 \end{pmatrix}$$

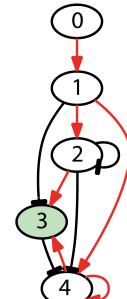


Figure 3.2: GRN (b), as graph and it's matrix representation. The history of it's protein expression levels is shown in Figure A.2.

$$J^{(c)} = \begin{pmatrix} 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & -1 & 0 & 0 \\ 0 & 0 & 0 & -1 & 1 \\ 1 & 0 & -1 & 0 & -1 \\ 0 & 1 & -1 & 1 & -1 \end{pmatrix}$$

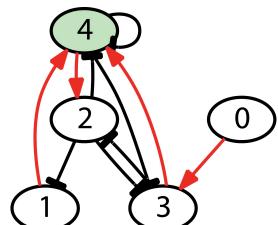


Figure 3.3: GRN (c), as graph and it's matrix representation. The history of it's protein expression levels is shown in Figure A.3.

We save the complete protein and gene expression histories for each simulation and present the results in the following section. All histories are given in Section A; we are going to discuss Figure A.1 and Figure A.3 in detail. These show what we assumed in Section 2.2: After differentiation, one or more of the proteins is fully expressed, justifying our method for finding distinct cell types.

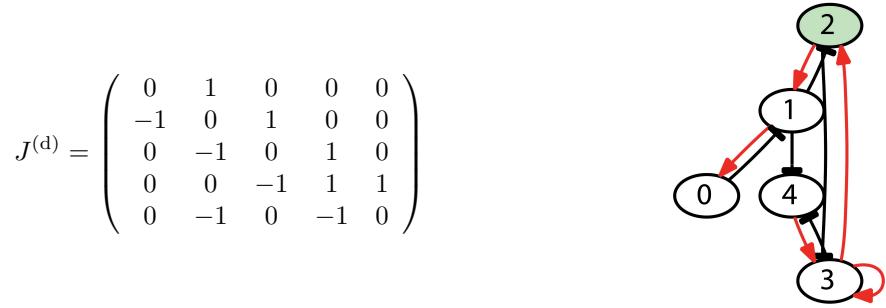


Figure 3.4: GRN (d), as graph and it's matrix representation. The history of it's protein expression levels is shown in Figure A.4.

Figure A.3 shows a behaviour which is common for undifferentiated stem cells: They oscillate in the protein space until they differentiate into a specific state, into which they converge. The differentiated state is an attractor in the protein space; for further details, see Figure 3.5.

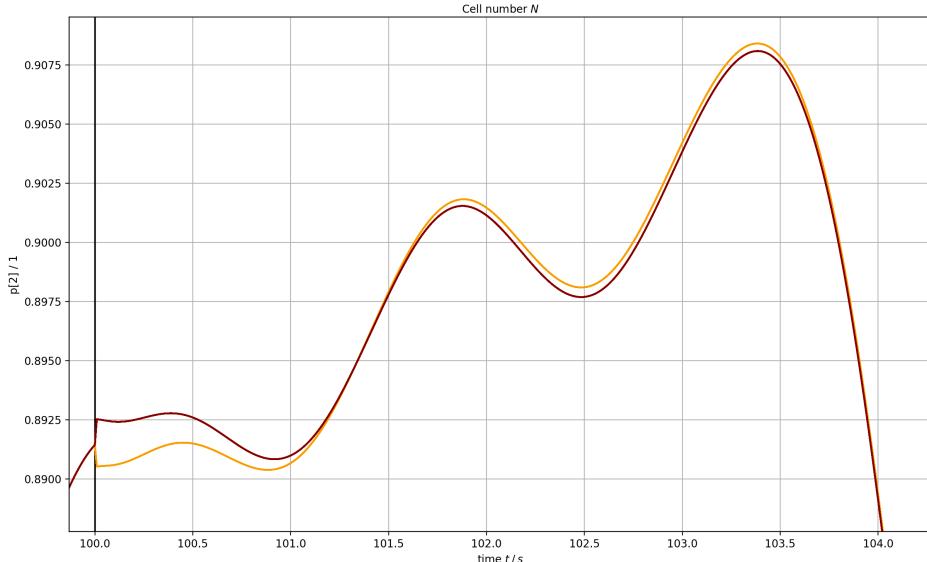


Figure 3.5: The differentiated cell states are attractors in protein space. This is hinted at in Figure A.3 where we see that oscillations at the differentiated states are much smaller and it becomes apparent here: The red curve represents the expression level  $p^1(2, t)$  of protein 2 in the first cell. At  $t = 100\text{s}$ , this (differentiated) cell splits and it's protein expression level is subject to noise (as discussed in Section 1). We thus see two curves after  $t = 100\text{s}$ . They converge towards each other however, meaning the differentiated state is an attractor.

For a graphical overview of the oscillation and differentiation processes, refer to Figure 3.6.

The fact that differentiation only occurs after cell divisions hints at the reason why it occurs: The intercellular interaction through diffusion disturbs the oscillating behaviour of undifferentiated stem cells. One of the cells deviates from the orbit it originally followed and converges towards an attractor in protein space, which is the differentiated state that the respective GRN permits.

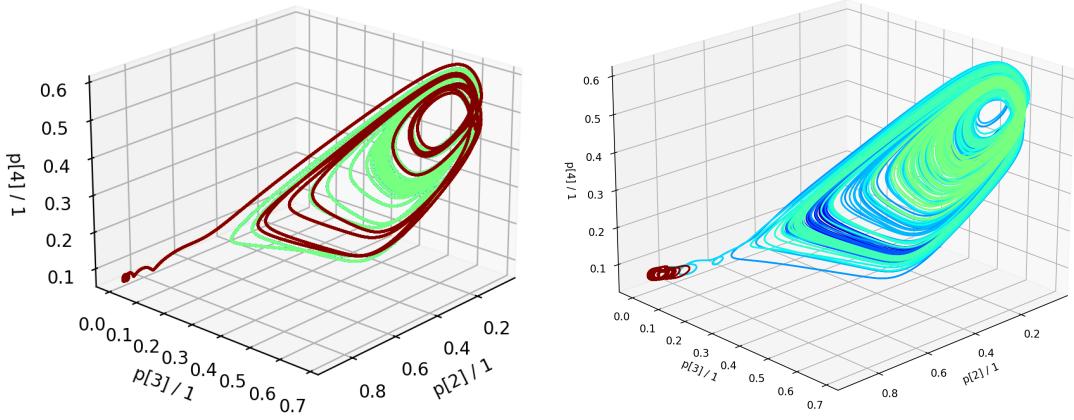


Figure 3.6: The protein expression level time evolution under network  $J^{(c)}$  (Figure 3.3), shown as a three-dimensional subspace of the complete protein space  $[0, 1]^5 \subset \mathbb{R}^5$ . The left plot shows all cells while  $50 \text{ s} \leq t \leq 100 \text{ s}$ , the right plot shows  $250 \text{ s} \leq t \leq 300 \text{ s}$ . Different cells have different colors. In both plots, we see the undifferentiated cells oscillate in a repeating pattern, while one cell differentiates and converges towards a differentiated state. The left plot shows how one cell (red) has already converged while a second cell (light blue) is attracted to the differentiated state.

## A Protein expression level histories

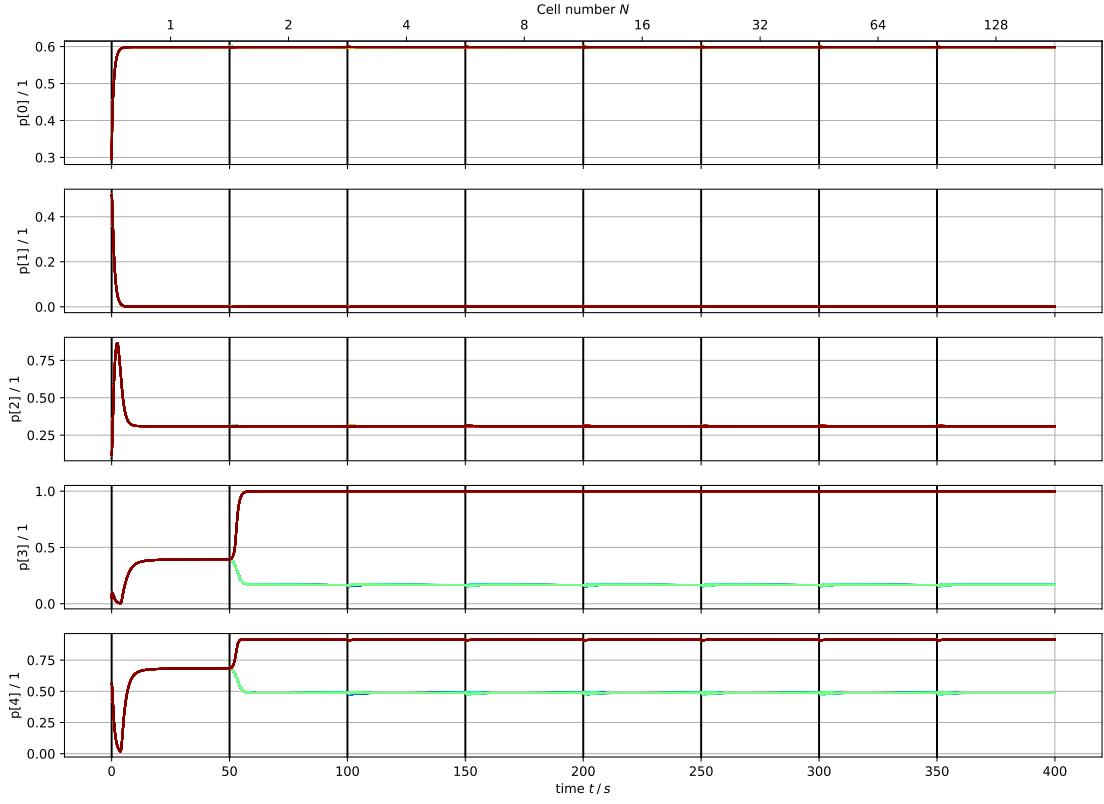


Figure A.1: The protein expression level history for GRN (a), shown in Figure 3.1. We see that, as soon as there are two cells present, the stem cells differentiate and converge into two distinct cell types.

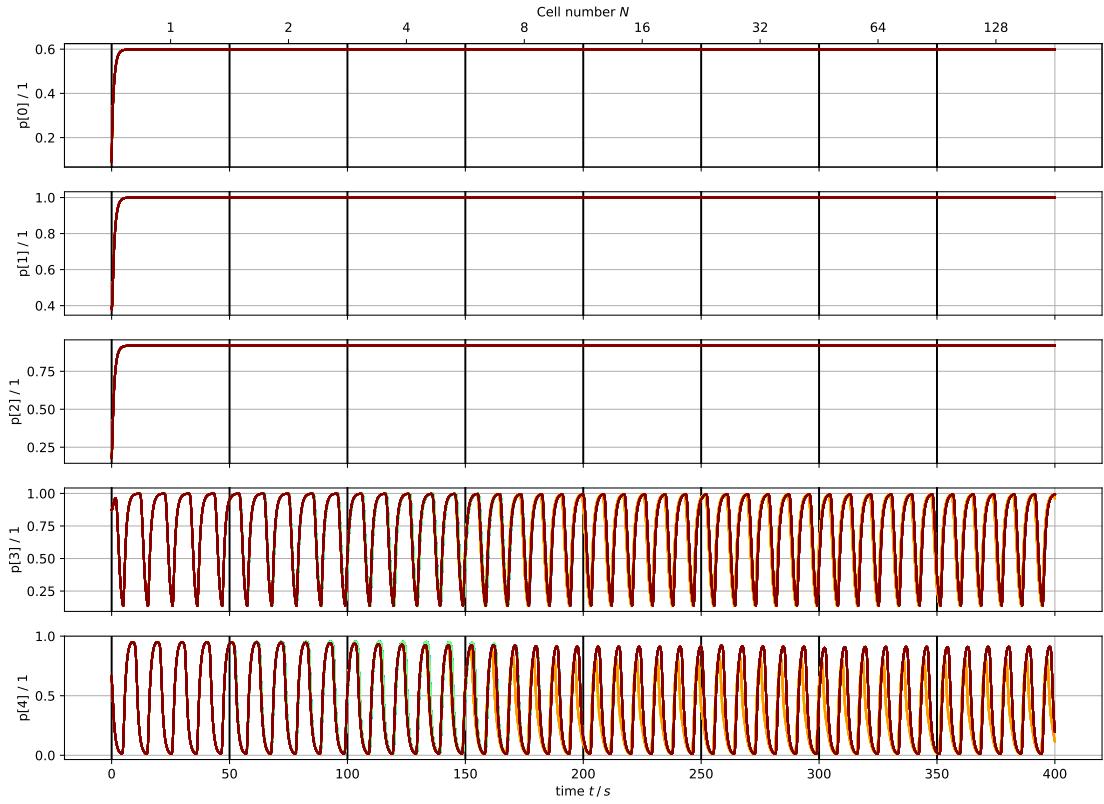


Figure A.2: The protein expression level history for GRN (b), shown in Figure 3.2.

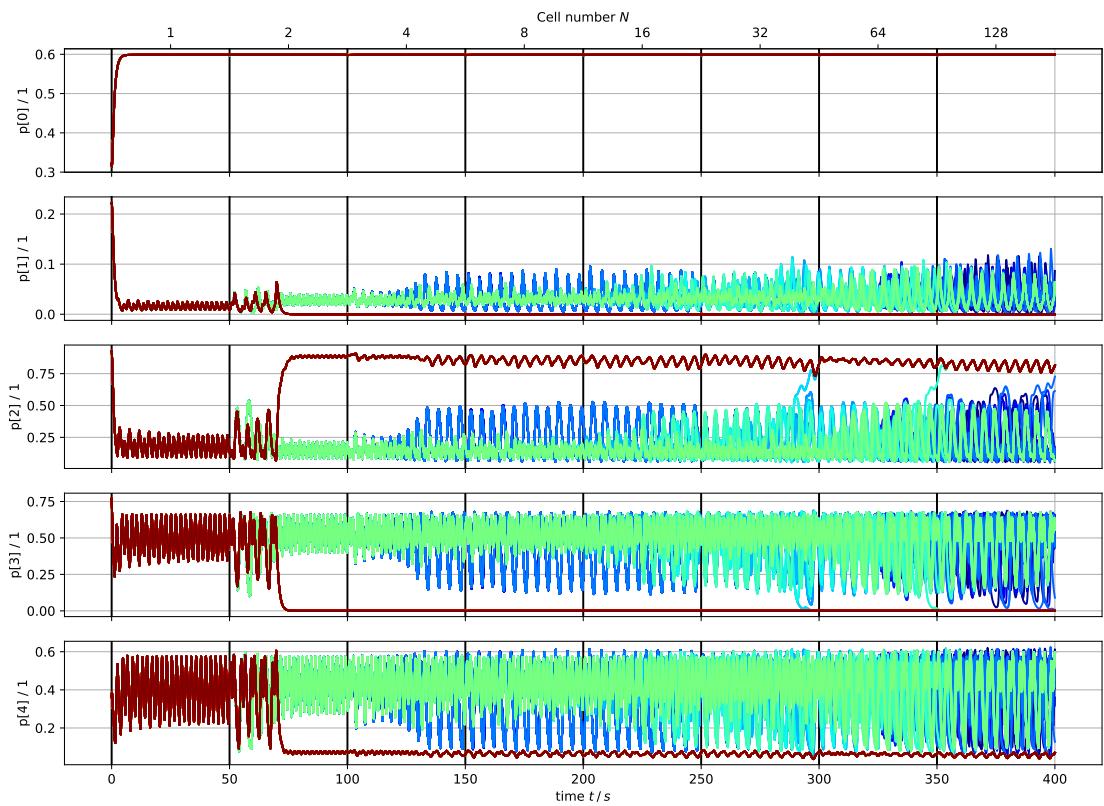


Figure A.3: The proteine expression level history for GRN (c), shown in Figure 3.3. This GRN exhibits oscillations in four of the protein expression levels, which is typical for GRNs which differentiate into different cell types. While two cells are present, we see that one of the cells separates itself from the other one and converges to a non-oscillating state. Other cells differentiate into the same non-oscillating cell-type during the 32- and 64-cell stages.

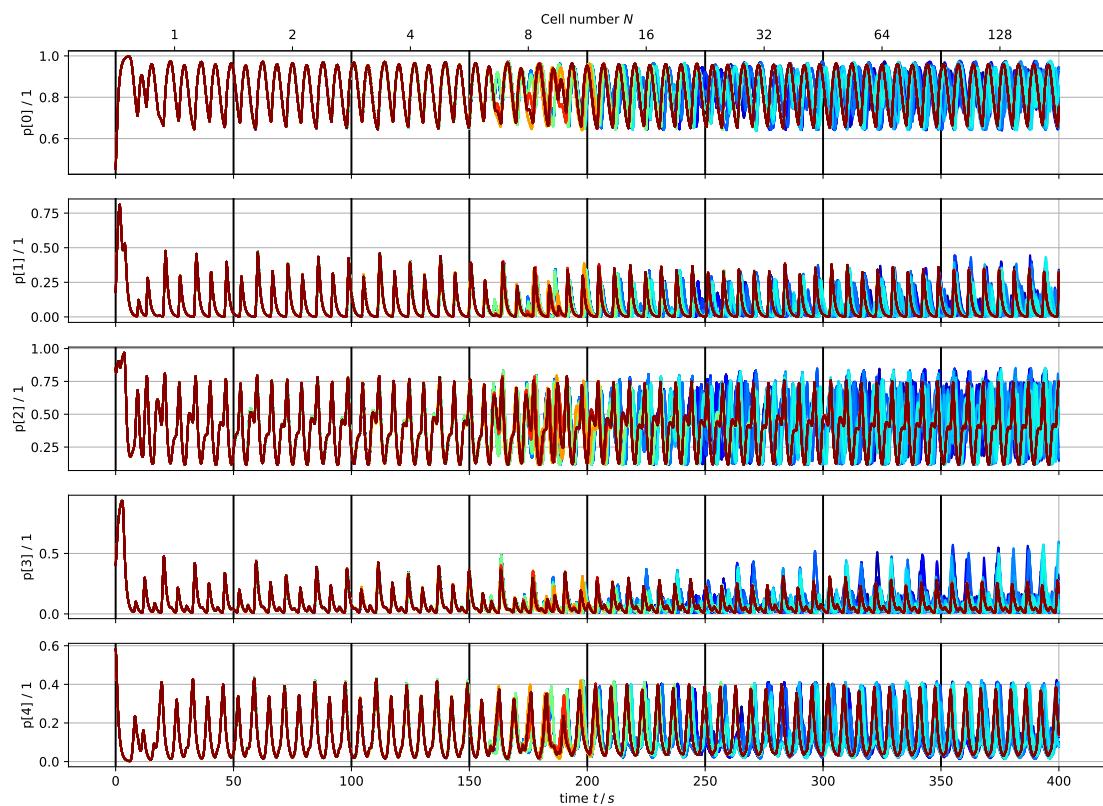


Figure A.4: The proteine expression level history for GRN (d), shown in Figure 3.4.

## References

- [1] Narito Suzuki, Chikara Furusawa, and Kunihiko Kaneko. “Oscillatory Protein Expression Dynamics Endows Stem Cells with Robust Differentiation Potential”. In: *Nature* 470.7334 (2011), pp. 329–334.