Unravelling cell fate decisions through single cell methods and mathematical models

Mircea, M.

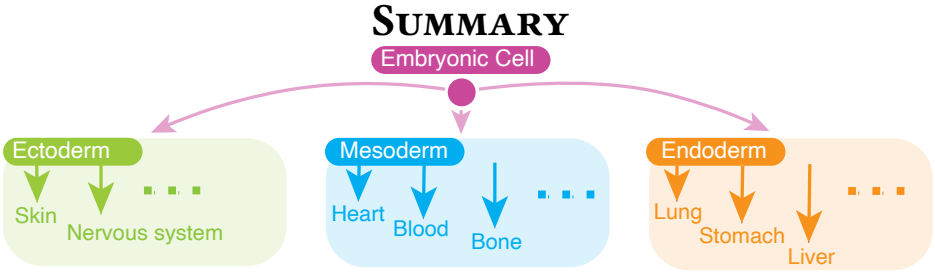

Figure 6.7: The three cell types that occur during gastrulation.

embryonic cells:
Cells of the embryo that have the potential to become any cell of the body.
gastrulation:
The point where three important cell types arise: endoderm, mesoderm and ectoderm.

Human life starts from a single fertilized egg which develops into around 30 trillion cells that make up our body. The egg in the beginning is only a single cell with the potential to divide and change in appearance, function, and composition. These changes are necessary to develop into a complete human, where cells of the heart are very different from cells in the lung or brain. It has been estimated that we have around 200 different cell types in our bodies. Cells at an early stage after fertilization still have the potential to become any cell of the body, and we call them **embryonic cells**. The path from the embryonic cell to a fully specialized cell contains many intermediate stages. We can imagine this path as our way to find a career. As a child, we can still go on to become anything we like, but throughout school and university, we make decisions that further narrow down our field of specialization. With every decision we make, we come one step closer to a specialized field and at the same time it becomes less likely to change career path. For example, one of the first and most crucial decision events in development is **gastrulation** (see Fig. 6.7). At this point, three different cell types emerge from the embryonic cells. This would be similar to making a choice of university studies. This choice will, in most cases, influence the rest of our lives. At this point, the cell can choose between either of the three cell types: endoderm, mesoderm and ectoderm. This choice will narrow down the future career choices: For example, mesoderm will become the heart, blood cells, or bone tissue; Ectoderm will become the nervous system or skin; Endoderm will become the liver, stomach or lungs. In this summary, I want to guide you further through the journey of the embryonic cell.

Information flow in a cell:
DNA → RNA → protein

To really understand how a single cell makes decisions to become a specialized cell type, we have to look into the inside of the cell. We have to figure out where this information is stored and how it is used. Every cell in our body contains the same DNA, which is conserved throughout the cell's life time. It carries information about the building blocks of our bodies. More precisely, it consists of many genes encoded into the DNA. Every gene contains different information, for example about the eye color, hair color or height. It also contains genes that specify cell types. Thus, for each cell type different genes have to become active. In order to do that, the information that the cell currently needs, is passed down from the DNA to the RNA. Every cell in our body has different RNA

molecules. The last information transfer is then, from RNA to protein. It is assumed that the DNA contains around 20.000 genes that code for a protein, and our cells can have as many as 500.000 different proteins. Proteins are the most important components in the cell because they have a wide range of functions. Proteins are like the employees in a company, each one of them has a different job to do and for the company to work everyone must do their job. We can see the inside of an embryonic cell in development like a start-up company. It has to make a lot of changes to its structure and employees until it becomes a well-established firm. With the help of all the employees in the company, the company will in the end make a decision for the next stage. One important kind of protein, in this regard, is called **transcription factor** (TFs). TFs decide which genes will be active and which will remain unused in a cell. They control the protein composition in the cell and can be viewed as the managers in the company. They can hire and fire the people in the company. TFs which can control a high number of genes are called **master TFs** and would therefore be directors of the company. Master TFs in particular often give the cell its identity, meaning that different sets of master TFs are at play in distinct cell types. So, each company has their own group of directors and together they decide on the company's brand. TFs can also control each other and work in parallel. In this way, they form a network called a **gene regulatory network**. In this scenario, you can imagine the management department having to hire or restructure themselves. If several directors are arguing for a position, then the winning group of directors can decide the new company brand.

In order to know which proteins are active inside the cell, they have to be measured. Preferably, we would like to measure all the proteins in a cell to understand exactly what is going on, but that is not yet possible. What is possible, is to measure the RNA molecules in a single cell with a method called **single-cell RNA sequencing** (scRNA-seq). The largest data set produced so far with this method contained 1.3 million brain cells. As proteins are produced from RNA, we can get an estimate of the number of proteins in a cell by measuring the number of RNAs. scRNA-seq has given unprecedented insights into the definition of a cell type and the development of an embryonic cell. Cell types are now defined based on the number of RNAs and many new cell types were discovered in the past decades in this way.

transcription factors:
A protein that can control the activity of some genes.
master TFs:
TFs that can control the activity of many genes.
gene regulatory network:
A network of proteins that activate and suppress each other.
single-cell RNA sequencing:
A method to measure the entire RNA in every single cell.

Chapter 1 of this thesis reviews cell identity in more detail and explains the changes that occur during the development of a cell. It outlines many single-cell techniques currently used to measure DNA, RNA, proteins, and chemical modifications in every single cell. Lastly, it introduces mathematical models for cell development to better understand the mechanisms and causal relationships between genes.

Chapter 2 derives a method named phiclust to evaluate the purity of a cell type in scRNA-seq data. As we can measure all RNA molecules in a cell, it is important to evaluate when two cells are of the same type. Every cell has different RNA molecules due to random fluctuations and processes unrelated to cell identity. Our new method can decide if the differences in RNA molecules between cells are different from random fluctuations. If

the differences are larger, the cells are of different types. We used phiclust on scRNA-seq data of a developing kidney and discovered previously overlooked cell types. In this way, phiclust can help classify and identify the many cell types in our body.

During development, cells are influenced by many factors that can change the cells' decision. Just like for us, when we decide on a career path, it matters what environmental influences we were exposed to. The same happens for cells: Their decision depends on the decisions of other cells in their environment. Naturally, an embryonic cell is inside the womb of the mother. Studying cells in the natural context is called **in vivo**. For ethical reasons, it is impossible to study human development in vivo. Additionally, many processes happen simultaneously in vivo that we can not control and we also do not know their influence. Thus, we decided to look at **in vitro** systems. In vitro refers to studies outside the normal biological context in a controlled lab environment. To study embryonic cells outside the body, they must be extracted and cultured in a dish. Once they are stable in vitro, we call them embryonic stem cells. Embryonic stem cells still maintain their potential to become any cell type of the body but have been extracted from their original environment. Also for ethical reasons, biologists often use **induced pluripotent stem cells** (iPSCs). These cells are taken from mature cells of the body, for example, the skin. Then, they are reverse-engineered to the characteristics of a stem cell and regain the potential to become any cell type of the body. In summary: *embryonic cells in vivo are embryonic stem cell in vitro.*

Chapter 3 studies the influence of the master TF ETV2 in blood vessel cells in vitro. Cells that form the vessels in our body are called **endothelial cells**. In general, their function is to create a barrier between a liquid, for example, blood, and the surrounding tissue. In particular, blood vessel-forming endothelial cells are important for the proper functioning of the heart. To understand the decision process of an endothelial cell, iPSCs were put into a dish and stimulated by specific chemical cues to become endothelial cells. But as described above, there are intermediate stages in the journey of the embryonic cell, which have to be respected in vitro. Using the career path metaphor, this means that first, it has to choose one of three basic career paths that cells have. For an endothelial cell this is mesoderm. In order to become a blood vessel, it has to specialize, like during master studies, on **cardiac** mesoderm (related to the heart). After its master studies, it can now choose the job of an endothelial cell. Surprisingly, we found that in the experiments, two cell types arise: endothelial cells and heart muscle cells which are responsible for the contractions in the heart. This means that after the master studies the cells could still choose between two different jobs. Using scRNA-seq, we showed that the decision depends on the master TF ETV2. If there is less ETV2 the iPSCs will become heart muscle cells and if there is more ETV2 they will become endothelial cells. Here, we can see how important master TFs are in the decision of the cell. As a master TF, ETV2 is able to manage the activity of other genes. So, if enough ETV2 is in the cell, then it will activate genes of the endothelial cell identity and in this way re-brand the cell. After re-branding, ETV2 is no longer needed and will be fired from

in vivo:
A process where cells are in their natural biological context.
in vitro:
A process where cells are outside their biological context in a dish.
induced pluripotent stem cells: Mature cells reverse-engineered to stem cells.
endothelial cells:
Cells that form vessels in the body.
cardiac:
Related to the heart.

the director position. This is an example of a gene regulatory network and how it changes as the cell makes choices.

As mentioned above, environmental factors can influence a cell's decision. One way in which this happens is by communication with each other. Cells communicate for example via specific proteins named **ligands**. This type of protein has the ability to leave the cell and wander off to the next cell. Once it arrives at another cell it changes the gene regulatory network of the receiving cell. In other words, this messaging protein, can influence the decisions of the managers in another company. In the first half of this thesis, the focus was on changes in the internal compositions of the cells. In the second half, we will extend this by investigating the influence of cellular communication on the cell's composition. We will study cell types combined in vitro that naturally would occur together in vivo to understand how they influence each other.

ligands:
Proteins that are used by cells to communicate.
paraxial:
Cells situated alongside an axis.
extraembryonic endoderm cells:
Cells surrounding the outside of the embryo.

Chapter 4 compares the influence of developmental origin to cellular communication in endothelial cells. This chapter is a continuation of *Chapter 3*, where heart muscle and endothelial cells were derived. In vivo, the human heart comprises multiple cell types, including heart muscle cells, blood vessel cells (endothelial cells), and connective tissue. So, these three cell types were put together in a dish in vitro to create a heart tissue environment. But endothelial cells not only form blood vessels but can for example also form lymph vessels that transport lymph instead of blood in the body. We wanted to understand how endothelial cells decide which type of vessel to become: Are they either influenced by the communication with neighboring cells or influenced by the cell's developmental origin (previous career choices)? For this reason, a second experiment was added where endothelial cells specialized in their master studies on **paraxial** mesoderm instead of cardiac mesoderm. Endothelial cells stemming from paraxial mesoderm usually form lymphatic vessels rather than blood vessels. Then, both endothelial cells, from either background, were put into the same heart tissue environment. With scRNA-seq, we saw that the RNA profiles of both endothelial cells were very similar after integration into the heart tissue environment. This result shows that cellular communication is more important than the developmental background for endothelial cells. It also shows in general that environmental influences have a high impact on the decisions of cells.

In a second example in *Chapter 5*, we show how cellular communication can cause the formation of a neural tube in vitro. We investigated cellular communication in a well-known in vitro model system, namely gastruloids. This system is a model for gastrulation which is the moment of choice between the three university studies of the cell (see Fig. 6.7): endoderm, ectoderm and mesoderm. We used mouse embryonic stem cells to reproduce this important decision event. The embryonic stem cells are able to form all three cell types by themselves in a dish. But we know from in vivo studies that the embryo is surrounded by a layer of cells called **extraembryonic endoderm cells** (XEN). When embryonic stem cells were combined with XEN cells in vitro, the XEN cells formed an outer layer around the stem cells, just like in vivo. Even more striking was that the stem cells started forming structures that looked like a tube. These structures have not been observed without the

addition of XEN cells. Thus, they must be a direct result of the communication between the two types of cells. With scRNA-seq and imaging, we characterized these structures. We found that they evolve from ectoderm, which is responsible for creating the nervous system. These findings indicate the involvement of XEN cells in the formation of neural tube-like structures. Additionally, we observed that also the mouse embryonic stem cells cause changes in the RNA of XEN cells. This experiment shows that communication between cells is already very important in the early stages of a cell's path.

dynamical systems:
Equations that evolve through time.

In the previous chapters, we have characterized elements involved in a cell's decisions with scRNA-seq and imaging. However, to understand the causal relationships in a gene regulatory network we want to formulate mathematical models. Therefore, we set out to use this data to inform mathematical equations about gene interactions during development in *Chapter 6*. These equations describe the time evolution of each gene in the gene regulatory network and are called **dynamical systems**. In a dynamical system, parameters determine the strength of interactions between genes or molecules. We employed physics-informed deep neural networks (PINNs) to determine these parameters with measured data. A neural network is a machine learning technique usually used successfully for pattern recognition. PINNs combine machine learning with physical laws. In this way, we can learn patterns from the data while conforming to predefined mathematical equations. We covered two relevant experimental scenarios: One where cells communicate and measurements of single cells are available. This data can, for example, be generated by imaging proteins through time. In the other scenario, cells do not communicate, and only snapshot data is obtainable. This model could then incorporate, for instance, scRNA-seq measurements. This analysis provides a starting point for acquiring mechanistic insights into the development of an embryonic cell by utilizing different data sets.

I hope this summary has allowed you to get a glimpse of the very complex journey of an embryonic cell. We showed how to define cell identity based on scRNA-seq data, and took a closer look at the influence of master TFs. Then, we added cellular communication to the experiments and found that it crucially influences a cell's developmental path. Lastly, we wanted to combine all this information into a mathematical model that allows us to understand the mechanisms involved in the cell's decisions. There is still much left to study before we can fully understand how a cell makes its decisions.
May the embryonic cell always make the right choice!