Gene regulation in embryonic development

Berg, P.R. van den

SUMMARY

The cells in our body are entrusted with countless different tasks. For instance, some specialized cells form a physical barrier to the outside, some generate the hormone insulin and some produce acid in our stomachs. To achieve these diverse tasks we carry a diverse set of genes, i.e. the functional units of information we inherited from our ancestors. Genes are stored in the DNA molecules in our cells' nucleus, which is collectively called our genome. However, just storing this information is not enough; it needs to be converted into different types of molecules that can perform the cell's tasks. The central dogma of molecular biology describes the flow of genetic information into these other molecules:

1. Genes in the genome can be *replicated*, creating an exact replica of the genome. This happens when cells divide into two cells, that each have to contain the genome.

2. Genes in the genome can be *transcribed*, creating transcripts (RNA molecules). This happens when certain proteins are needed in the cell. All transcripts of the cell are, collectively, called the transcriptome.

3. Genes in the transcriptome can be *translated*, creating proteins. These molecules are responsible for most of the work in the cell. All proteins of the cell are, collectively, called the proteome.

The central dogma of molecular biology is an excellent model to keep in mind when thinking about how genetic information flows from DNA to RNA to protein. However, it fails to describe *which* information flows at a given time and at which *rates*. That is where *gene regulation* comes into play. Gene regulation is a set of biochemical processes the cell employs that ensures a finely tuned proteome, which is crucial for its function. Gene regulation is also one of the common threads among the chapters of this thesis along with mammalian development. Each chapter of this thesis covers a different type of gene regulation (at the different levels of the central dogma, if you will).

In the **Introduction** we introduce the important concept of cell types, and how this subject closely ties into gene regulation and development (another common thread of this thesis). We then give a layman's explanation of several concepts that are needed to understand the remainder of the thesis. We introduce DNA methylation as a form of epigenetic regulation (epigenetics: inheritable alterations of the genome that do not change its DNA sequence). We next explain transcriptional regulation, the control mechanisms that decide which genes are transcribed at which levels, and translational regulation, which controls the rate at which translation take place. And finally we discuss *omics*, a relatively recent collection of techniques that are vital for measuring the above phenomena and are heavily featured in this thesis.

In **Chapter 1** we look at how epigenetics influences cell identity and transcription in mouse embryonic stem cells (mESCs). We investigate a set of super-enhancers (sets of closely-clustered enhancers, which are regions of the genome that do not code for genes but can promote transcription elsewhere). Super-enhancers are often associated with cell identity and therefore highly relevant in developmental systems. Moreover, it was found that in some cell types the DNA of super-enhancers is methylated at varying levels between individual cells. However, the mechanisms causing this heterogeneous methylation were not well understood. We created methylation reporter cell lines for the *Sox2* and miR-290 super-enhancers to uncover these mechanisms. These cell lines have fluorescent proteins of two distinct colors that are turned on and off depending on the methylation state of each of the alleles (the maternal and paternal copies of the super-enhancers). We show that the methylation states at these super-enhancers are not only heterogeneous but also highly dynamic, because they switch from on to off and vice versa over a matter of days. We also show that the methylation state of the super-enhancers influences the transcription of genes *in cis* (meaning the methylation state has a local effect on the same chromosome). Finally, we observe that this dynamic methylation is not an *in vitro* artifact but also occurs in pre-implantation embryos.

In **Chapter 2** we use the transcriptomes of cells as key indicators of their cell type. We performed single-cell transcriptomics on human fetal kidneys at different stages in development. Kidneys consist out of about 1 million nephrons, functional units that operate independently. The development of these nephrons is asynchronous, meaning that we can observe multiple stages of nephron development simultaneously in the fetal kidney. In the transcriptomics dataset we identify 22 different cell types, ranging from precursor cells to fully differentiated cells. Some of these cell types were novel and more nuanced subclassifications of previously known cell types like the *nephron progenitor cells*. We also observe that for the *podocyte* the transcriptome continues to change over the course of development, even though the cell type is already established. This kind of more detailed information on the kidney cell types can potentially open up avenues for the development of cures for kidney diseases in the form of regenerative medicine.

In **Chapter 3** we investigate translation and degradation rates in mESCs. Measuring RNA concentrations is often more convenient than measuring protein concentrations and because of the hierarchical relationship between the two, changes in RNA are often assumed to occur in protein as well. If there is a sudden change in RNA concentration of a gene then the protein concentration should reflect this. How responsive a gene's protein concentration is to its RNA concentration differs per gene and changes dynamically. Therefore, in a system undergoing pervasive changes, e.g. during differentiation, one cannot simply measure RNA concentrations and assume that the corresponding proteins follow a similar pattern. In this chapter we model protein synthesis and degradation rates using RNA and protein data of differentiating mESCs. This allows us to measure the extent of discordance between RNA and protein. More interestingly, the models were useful for identifying when our prior

assumption of constant turnover rates does *not* hold, hinting that these genes may be post-transcriptionally regulated (i.e. regulated at the protein synthesis or degradation steps). Using these models we identify new cases of translational regulation by microRNA (miR), small RNA molecules that block gene-encoding RNAs. Interactions between mRNA and miRs are notoriously hard to identify.

In **Chapter 4** we put the models of the previous chapter to the test. We set out to verify some of the miR-gene interactions we had identified. To probe the interactions we *transfected* the cells with miR mimics and miR inhibitors (to transfect: to introduce nucleic acid molecules or proteins into a cell). To optimize the timings and concentrations of these transfections we first created cell lines that report when these transfections take place by means of fluorescent proteins. We then transfected six different miR mimics, performed transcriptomics and show that these miRs negatively influence the predicted genes' concentrations. These results reinforce the use of the models from Chapter 3 as a means of identifying cases of translational regulation by miR.