Chapter 6

Concluding Remarks

It is remarkable that a whole complex organism with myriad different cell types and tissues develops from one single zygote. Embryonic stem cells are derived from pluripotent cells within the inner cell mass of the embryo and they have the capacity to differentiate into all tissue types of the organism as well as being able to contribute to chimeric embryos. This creates a promising avenue for the field of regenerative medicine. Understanding the molecular mechanisms of pluripotency and the exit into differentiation is key for designing protocols to grow tissues in culture.

Depending on cell culture condition, mouse embryonic stem cells have different transcriptomes and the population has a different structure. In my thesis, using single cell mRNA sequencing I dissected the heterogeneity of the population of mouse embryonic stem cells cultured in three cell culture conditions. Comparison with previous studies allowed me to generate a comprehensive picture of the gene expression variability. I confirmed that

genes previously suggested to be heterogeneous or fluctuating are indeed doing so. In serum, pluripotency and differentiation genes fluctuate as two modules. In cells with low expression of the pluripotency module, the differentiation module is high and *vice versa*. This corresponds to functional differences between cells, where some of them are more pluripotent and some already express the differentiation programme.

Cells cultured in 2i medium that mediates a ground state of pluripotency are homogeneous for expression of the pluripotency module and do not express markers of differentiation. On the other hand, cell cycle gene expression is heterogeneous in 2i. I was able to use this heterogeneity to assign cells to cell cycle stages: G1/S or G2/M. Using the data presented in this dissertation and other previously published data (Tsang et al., 2015) I observed that there is a relationship between cell cycle heterogeneity and the length of the cell cycle. Cells that cycle quickly have homogeneous expression of cell cycle genes. In cells that cycle with moderate speed, such as those cultured in 2i one can discriminate G1/S form G2/M cells. In slowly cycling cells, such as HSC all phases of the cell cycle can be identified and even G1 can be divided into early and late (Tsang et al., 2015).

I speculate that there are two reasons for low cell cycle noise in fast cycling cells. Firstly, in very quickly cycling cells G1 phase is virtually non-existent causing lower heterogeneity. Secondly, the degradation half-lives of cell cycle related genes are 6-8 hours (Sharova et al., 2009). If the cell cycle is very quick there is not enough time for mRNAs from one phase to degrade when the cell enters the next phase of the cell cycle. This leads to mRNAs from one phase to "bleed" into the following phase. Biologically, this does not necessarily have much effect on cell cycle regulation, because this is achieved at the level of

protein signalling, mostly post-translational phosphorylation by CDKs and protein degradation, mostly of cyclins.

This relationship can be exploited to estimate the speed of the cell cycle in heterogeneous populations. For example, cells from complex tumour tissues can be profiled using single cell mRNA sequencing and subpopulations can be identified. Subsequently, the relative proliferation rates of the subpopulations can be measured using heterogeneity of expression of cell cycle genes in each of the subpopulations. It is quite remarkable that a dynamic feature of a system can be measured from snapshot data such as single cell mRNA sequencing of one time point.

Cells cultured in alternative 2i are similar transcriptomically to cells cultured in 2i, especially for expression of pluripotency genes, suggesting that inhibition of SRC gives rise to a similar phenotype as inhibition of MEK1/2.

Furthermore, I identified a population of previously-reported so-called "2C-like cells" (Macfarlan et al., 2012) in 2i medium and looked at their transcriptomes in relation to transcriptomes of cells from subsequent stages of embryo development. These cells are substantially more similar to cells from the blastocyst than cells from the embryo at the 2 cell stage. The transcriptomes of 2C-like cells are similar to those of the other cells in 2i culture, but in addition to the transcriptomic profile of 2i cells they express some additional genes. 2C-like cells arise probably due to chromatin changes that are forced by signals from the media the cells were cultured in. Derepression of endogenous retroviral elements causes expression of genes that are regulated by MuERV-L in addition to the transcriptomic profile of 2i cells. This is a useful observation, as it allows decoupling regulation of gene

expression by MuERV-L from changes that occur in 2 cell stage embryos, and the study of this process in stead-state culture.

Finally, I discovered several potential regulators of pluripotency and validated that three genes, namely *Ptma*, *Zpf640* and *Zfp710* are regulators of pluripotency. The approach I used can be used for any biological system, for understanding genes that change in transitions or as a result of response to stimulus.

In my work, in addition to gaining biological and mechanistic insights into the pluripotency of mouse embryonic stem cells, I have shown how and what information can be harvested from the single cell transcriptomic data. I measured and understood sources of heterogeneity, found and characterized a rare cell population, assigned cell cycle stage to cells and identified new players important for gene expression networks. These approaches will prove useful for analysis of any type of data in the future.