Cellular differentiation and the regulation of gene expression.

How is gene expression regulated?
How does this relate to cellular differentiation?

Cellular differentiation is one of the 3 main processes needed to form a complex organism from a single fertilized egg cell (Cell division, cell, differentiation and morphogenesis). A complex organism requires many hundreds of different cell types to form structures and carry out specific functions. If all the cells arise from a single fertilised egg cell and all contain the same DNA in their nuclei, how do they become different to each other or "differentiate"? This is what we call cellular differentiation.Cellular differentiation is brought about by differential gene expression: the cells become different because they express different genes.

How are genes turned on and off?

A gene includes more than coding sequences (sequences that are transcribed and translated). It also includes regulatory sequences that determine which cells express that gene and when they turn it on. Regulatory sequences at the start of the coding sequence (Promoters) are needed for the transcriptional machinery to assemble and begin to transcribe the DNA sequence into an RNA message or transcript (mRNA). These are similar in all eukaryotic genes. Other regulatory sequences are gene specific and these determine when and where a gene will be turned on.They can be situated close to the coding sequence (proximal control elements) or at large distances (distal control elements). These are often called enhancers and can be positioned upstream or downstream or within the coding sequence. Transcription initiation is controlled by proteins that interact with DNA (regulatorysequences) and with each other These proteins are called transcription factors and operate by binding to the specific regulatory sequence elements (control elements- enhancers) described above. Cell specific transcription factors influence the efficiency with which the general transcription factors (transcription initiation complex) assemble on the promoter sequence and initiates expression of the gene. Distant control elements, enhancers, may be thousands of nucleotides away from the promoter or even downstream of the gene or within an intron. Factors that turn a gene on in this way are called activators. Eukaryotic genes can also be influenced by repressor proteins that bind to DNAregulatory sequences and tend to destabilize transcription and turn the gene off (can also be called silencers).

Three important points to note about cellular differentiation:

Point 1: Cellular differentiation is usually a result of transcriptional regulation: turning genes on and off.

Point 2: During embryonic development, cells become obviously different instructure and function as they differentiate. But differentiation does not happen suddenly. Differentiation happens progressively as the embryo develops. When differentiated cells appear they already produce the proteins that allow them to carry out their specialised roles in the organism e.g eye lens cells, 80% of their capacity for protein synthesis makes crystallins. However changes will be taking place inside a cell long before it visibly "differentiates". These include a gradual reprogramming of the genes that are expressed. This would show up only at the molecular level. We looked at the example of progressive myoblast differentiation under the control of the cell specific transcription factor MyoD to illustrate progressive differentiation.

Point 3:

The genes that encode transcription factors that control cellular differentiation (e.g. MyoD) are called "Master regulatory genes". These control the expression of sets of target genes (downstream genes), the products of which are needed for the cell to differentiate. Many of the downstream genes may also be regulatory genes controlling the expression of more target genes. This is how a cascade event along a differentiation pathway may be controlled and explains why differentiation is progressive.

Spatiotemporal gene expression patterns

In higher plants, embryogenesis is a process that establishes the body plan required for early vegetative growth after germination. Early embryogenesis includes crucial developmental events for organizing plant architecture; namely, establishment of embryo polarity, pattern formation, and the differentiation of distinct cell types, tissues and organs. Although these determination processes are common among plant species, there are differences in the patterns of cell division and determination processes. The genetic and molecular mechanisms underlying developmental processes during embryogenesis have been investigated in the model eudiocot *Arabidopsis thaliana*. Embryonic studies of *Arabidopsis* also benefit from the abundance of genetic and genomic resources. Forward and reverse genetics studies increase our understanding of the molecular basis of embryo polarity and pattern formation, as well as cell type, tissue and organ differentiation mechanisms. Furthermore, genome-wide expression analyses using microarray and RNA-seq have been used for embryonic studies to estimate the number of genes expressed and predict the gene families associated with developmental and metabolic processes during embryogenesis. Transcriptome analyses combined with laser-capture microdissection during *Arabidopsis* embryogenesis revealed global expression features in different parts of embryos and within embryos of different developmental stages. Although it is important to determine whether the genetic programs governing embryo development during plant embryogenesis are evolutionarily conserved among various plant species, molecular genetic work has mostly been conducted in *Arabidopsis* and is not as well understood in other species.

Rice embryogenesis is a good candidate for comparative genetic studies, although the rice embryo has a morphologically complex structure and the early embryo is deeply embedded in ovary tissues. Rice embryogenesis exhibits biologically novel features not observed in *Arabidopsis*. First, because the embryo does not show stereotypic cell division patterns, embryonic cell types and tissues cannot typically be traced back to the early cell lineage. Second, the rice embryo shows an evident dorsal-ventral polarity whereby the shoot initiates on the ventral side and the scutellum forms dorsal side. Third, a radicle is endogenously differentiated in the center of the early embryo. Fourth, complex embryo-endosperm interactions affect embryo development and growth throughout embryogenesis. Thus, elucidation of the molecular mechanisms governing embryo development in rice will provide novel insights into embryogenesis of grasses and contribute to the diversity of embryonic studies in higher plants. Functional analyses of genes involved in rice embryogenesis have been mainly performed using forward genetics approaches.

One of the reasons why reverse genetics have not been applied to embryonic studies in rice is the limited number of knockout lines publically available. In addition, knockdown techniques such as RNAi or artificial miRNA are not successful because transgenic plants are typically generated from calli in rice and genes regulating embryo development and fertility are usually not silenced in regenerated plants. However, recent advances in genome editing techniques have improved the feasibility of the study of embryonic lethal null alleles. These techniques allow observation of null phenotypes of essential genes required for embryogenesis in rice. In particular, it has been shown that the clustered regularly-interspaced short palindromic repeats (CRISPR)/Cas9 system is able to efficiently knockout multiple paralogous genes in rice for subsequent characterization. Thus, our understanding of the genetic regulation of embryo development will improve using genome editing-based reverse genetics. For future functional analyses of grass embryogenesis, genome wide expression profiling is necessary. Although expression profiling of various embryonic tissues and stages has been reported in rice and maize predicting the spatial expression pattern and dynamics of gene expression during early embryogenesis remains difficult. Transcriptome profiling of rice embryos using microarray combined with laser microdissection allow us to predict expression patterns and changes in the expression of specific genes during key developmental stages and in distinct embryonic organs and domains

Transcriptional regulation plays an important role in orchestrating a host of biological processes, particularly during development. Advances in microarray and sequencing technologies have allowed biologists to capture genome wide gene expression data; the output of this

transcriptional regulation. This expression data can then be used to identify genes whose expression is correlated with a particular biological process, and to identify transcriptional regulators that coordinate the expression of groups of genes that are important for the same biological process. The identification of such genes and transcriptional regulators is complicated by the complex heterogeneous mixture of cell types and developmental stages that comprise each organ of an organism. Expression patterns that are found only in a subset of cell types within an organ will be diluted and may not be detectable in the collection of expression patterns obtained from RNA isolated from samples of an entire organ. Therefore, techniques have been developed to enrich samples for specific cell types or developmental stages, especially for studies in plants.

In the model plant, *Arabidopsis thaliana*, several features of the root organ reduce its developmental complexity and facilitate analysis. Specifically, most root cell types are found within concentric cylinders moving from the outside of the root to the inside of the root. These cell-type layers display rotational symmetry, thus, simplifying the spatial features of development. This feature has been exploited in the development of a cell type enrichment method. This enrichment method uses green fluorescent protein (GFP)-marked transgenic lines and fluorescently activated cell sorting (FACS) to collect cell type-enriched samples and has allowed for the identification of cell-type-specific expression patterns. Another feature that makes the *Arabidopsis* root a tractable developmental model is that cell types are constrained in files along the root's longitudinal axis and most of these cells are produced from a stem cell population found at the apex of the root. This feature allows a cell's developmental timeline to be represented by its position along the length of the root.

To obtain a developmental time-series expression dataset individual *Arabidopsis* roots are sectioned into 13 pieces, each piece representing a developmental time point. Each of these sections, however, contains a mixture of cell types, and the microarray expression values obtained are therefore the average of the expression levels over multiple cell types present at these specific developmental time points. However, expression from the longitudinal dataset, does contain averaged expression of all cell types, and may be used to infer the missing cell-type data.