Research Interests – Dr. Eileen Furlong

The process of embryonic development is amazing at many levels. First, in its capacity to utilize the genome in different ways to develop cells and tissues with a huge diversity in function. Second, in its inherent robustness - embryogenesis can generally proceed to give rise to embryos with the same body plan, despite differences in segregating mutations and in the environmental conditions of the embryo, a process Waddington called canalization. Much of this is driven by and buffered within large highly interconnected transcriptional networks. These in turn are regulated through the action of transcription factors that function through enhancer elements. Enhancers are relatively small stretches of DNA scattered throughout the non-coding genome that regulate when and where a gene will be expressed. These 'control elements' are therefore essential drivers of embryogenesis, and actually of almost all biological processes. Mutations in enhancers can lead to devasting developmental defects and are increasingly being found as key driver mutations in a wide-range of human diseases. On the positive side, mutations in enhancers are also a major driver of evolution, leading to speciation differences of different traits.

My long-standing interest is to understand both properties (regulation and robustness), dissecting fundamental principles of genome regulation, including understanding how developmental enhancers function, how they are organised in the three-dimensional nucleus, and how they regulate developmental programs. Over the years, we have studied this at different scales (Fig. 1), mainly using *Drosophila*, which is a very prominent model system to dissect both embryonic development and chromatin/transcriptional regulation – many seminal discoveries in our understanding about both processes were first uncovered in flies.

First, zooming in to elucidate the inherent properties of enhancers themselves: By integrating genetics, genomics and computational modelling, we uncovered an intrinsic complexity and plasticity within developmental enhancers, in addition to mechanisms that explain how some enhancers function. Enhancers are typically bound by multiple transcription factors which give rise to a specific pattern of gene expression. Rather than having 'unique transcription factor binding codes' for a specific expression pattern, which was the prevailing view, we found that different combinations of transcription factors can give rise to the same pattern of expression. How transcription factors bind to an enhancer is also important for activity, including additive or cooperative recruitment (direct or indirect). This has led to several models of enhancer

function. While studying transcription factors essential for heart development, we uncovered an alternative mode of cooperative transcription factor recruitment. This 'Transcription Factor Collective' model explains how some enhancers can still function even in the presence of mutations in some DNA motifs, a property that has also been observed at some neuronal enhancers. By using natural sequence variation, through the integration of Population Genetics and Developmental Biology, we uncovered a number of additional properties that buffer the deleterious effects of genetic variation on gene expression. One surprising finding was the identification of widespread genetic epistasis within developmental enhancers and promoters, demonstrating that large effect variants are attenuated or buffered by other variants within the same element. The more we dig into the functions of enhancers the more plasticity we observe, including elements that have dual functions, including enhancer and silencer activity, or enhancer and promoter activity, or enhancer and boundary activity, depending on their context. This probably speaks to how these inter-related functions evolved – but collectively, these intriguing features highlight the complexity and huge challenge to predict function from sequence alone. With more extensive perturbation data coming, and more sophisticated machine learning models, there should be exciting advances here going forward.

Second, zooming up from individual enhancers to regulatory landscapes within chromatin domains: Each gene typically has may enhancers, especially developmental genes which tend to have very complex expression patterns. A gene's enhancers can be located upstream or downstream from the gene they regulate, and are sometimes even in the introns of other genes. Although some enhancers are located quite close to the gene they regulate (in linear genomic distance), others are at great distances. How such remote enhancers communicate with their promoters is one of the central questions in genome regulation. The prevailing model is that the chromatin is looped (like a string) such that an enhancer and gene promoter are brought into physical proximity to each other. How this looping is regulated during embryogenesis and how an enhancer recognises its correct promoter are two key questions that we are addressing. By combining genetic dissection, genomics and imaging approaches we have uncovered some intriguing and quite surprising properties of enhancer-promoter communication.

Enhancers must be in relatively close proximity to the promoters that they regulate to activate transcription. At some loci this proximity only occurs in the appropriate cell type, and at the appropriate time of embryogenesis when the gene needs to be expressed. Looking at a large number of developmental enhancers at early stages of embryogenesis, we uncovered another mode, where enhancers are already in proximity to their promoters, hours before the gene is

expressed. This intriguing finding indicates that the topology of enhancer-promoter interactions that are required during cell fate decisions are already put in place at very early stages of embryogenesis, where they appear primed for activation at later stages. It also suggests that proximity, at least at this scale, is not the trigger for transcription at these loci. To dissect the functional role of chromatin topological domains and long-range chromatin loops in gene regulation, we are using genomic rearrangements and deletions at different scales, taking advantage of the genetic tools in *Drosophila*. This includes using large-scale highly rearranged chromosomes (balancers) to perturb genome structure and then systematically assessing the impact on chromatin interactions and gene expression. This revealed a very surprising finding – although some enhancer promoter interactions were perturbed, the majority were not. The expression of most genes' did not change upon perturbation of boundaries or fusion of their chromatin domains. This indicates that while some enhancers require chromatin domains to constrain their function and prevent them from mis-appropriately activating another gene, other enhancers do not. Moreover, even after the fusion of TADs that double the size of a domain, bringing many more genes into the same domain, many enhancers ignore them and still regulate their normal target gene's promoter. This opens up many interesting questions about how these elements recognise their correct cognate promoter and why some enhancers (or genes' expression) are affected by chromatin rearrangements while others are not. This is important, as such chromosomal rearrangements are often associated with cancer and developmental defects.

Third, zooming out even more from gene regulatory landscapes (or chromatin domains) to genome-wide developmental networks. Assessing developmental networks has been enabled through the development of genomics approaches. Ever since I was a post-doc, I have been drawn to the unbiased and systematic view that genomics experiments bring. While many of these methods are first developed in tissue culture cells or yeast, I have tried to bring the power of these methods to bear to tackle the complexity of embryonic development. My group optimises, and sometimes develops, genomics methods to understand how multi-cellular embryos develop. Applying these approaches over 15 years ago, revealed some of the first global views of dynamic transcription factor occupancy and enhancer function as an embryo develops, charting changes in transcription factor binding, chromatin state and expression changes as a tissue's development unfolds. Prior to this, concepts were based on a handful of well-studied model loci. These first genome-wide views blew my mind at the time, as they demonstrated the vast scale, complexity and interconnectivity of developmental networks. Combining them with mutants over the ensuing years, has uncovered functional regulatory programmes, in particular during mesoderm specification to different muscle tissues. In our current and future

work we are pushing this at a single cell level. We recently showed that single cell views of open chromatin can identify cell type specific enhancers, predict which stages and tissues they are active in, in addition to being able to identify cell types and follow their developmental trajectories during embryogenesis. Applying this approach in a dense time-course of a tissue's development, provided a very rich view of both the enhancers and transcription factor that are active during each transition. Combining this with similar single cell data from mutant embryos revealed the real power of the system to identify mutant cellular phenotypes *de novo*, while simultaneously revealing the molecular function of the transcription factor. Going forward, we will extend this approach to include more genetic mutants and more diversity in the molecular measurements, with the long-term goal to understand how developmental programmes are regulated and interpreted during embryonic development.

Our research therefore sites at the interface of the field of Transcription/Chromatin Biology and Development Biology, where we strive to uncover general principles in how the genome is regulated and utilized, which has implications to the fields of Developmental Biology, Evolutionary Biology and Human Disease.

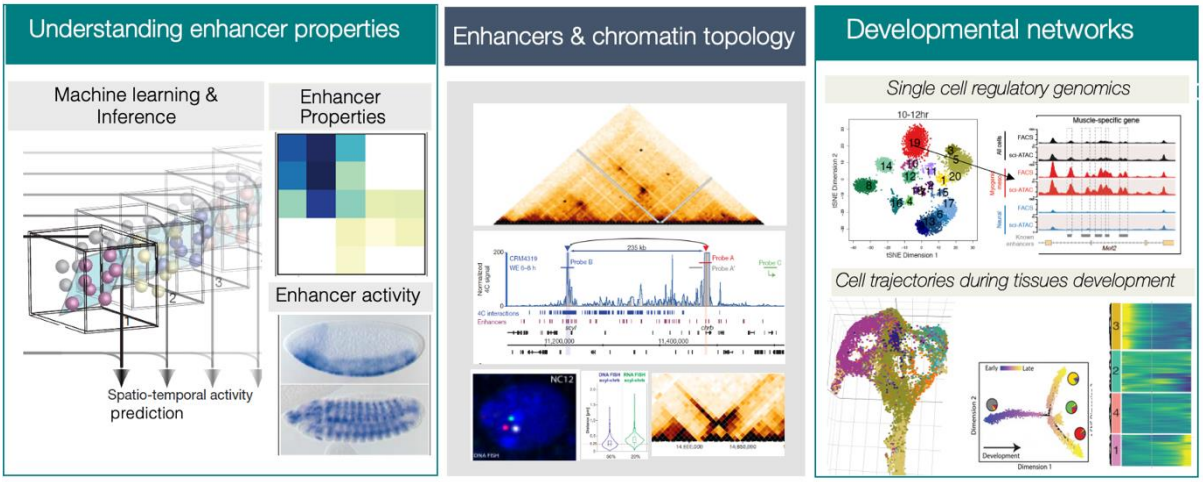

Figure 1:

Left: Integrating approaches from Genomics, Computer Science and Genetics has uncovered a number of surprising features about how enhancers function. Shown is work using a machine learning approach to predict enhancer activity from genome-wide data on transcription factor occupancy.

Middle: Using Genomics, Imaging and Genetics to dissect how enhancers communicate with specific promoters to regulate gene expression and how this changes over embryonic development.

Right: Using single cell regulatory genomics, we can follow developmental trajectories, predict single cell identies and tissue specific enhancers. This is a very powerful approach when combined with mutants, as we can identify mutant phenotypes de novo at both a cellular and molecular resolution, and integrated this information for many mutants to build functional developmental networks.