Dissecting DNA Maintenance Methylation in Mammals: Fidelity, Timing and Instructive Cues

DNA is the carrier of genetic information. The basic message is encoded in a long string with only a four letter alphabet, made from four chemically distinguishable DNA bases. The sequence of letters (bases) is identical for all cells in an organism. However, in multicellular organisms, different cells of different tissues interpret the same "message" differently. The ability to do so depends on a second layer of information on top of the sequence of four bases that includes a modification of one of the bases, by a small hydrophobic group. In more technical language, the modification is methylation of cytosine (C) to 5-methylcytosine (5mC).

When cells divide, the daughter cells need to both inherit a full set of instructions. Cells solve this problem by exploiting a key property of DNA, its composition of two complementary DNA strands, which carry to same information. In cell division, DNA is shared "semi-conservatively", i.e. every daughter cell inherits one original strand and has to newly synthesize another strand, using the instructions from the parental strand. For the first layer, the sequence of bases in the newly made ("nascent") DNA strand is straightforwardly defined by the sequence of bases in the parental strand. Initial synthesis of the new DNA strand has only four standard building blocks (non-modified DNA bases) at its disposal. The second layer of information, often called the "epigenetic" or "epigenomic" layer, is more complicated. Selective modification of C bases to 5mC bases in the nascent DNA strand takes instructions from parental strand, but also other components of the epigenomic machinery, and also proteins acting "in trans".

In this project, we want to focus on the second, epigenomic layer of information in DNA, and more specifically, on the selective conversion of the "standard" base C to modified base 5mC. Most studies to date have focused on equilibrium levels of this modification. In our project, we want to probe the process in a time-resolved manner, and test its accuracy at the single cell level. The key tools for our studies are chemical biology tricks to distinguish modified and unmodified bases, and modern sequencing techniques, that make it possible to read billions of DNA bases in a single experiment. Some of the techniques that we hope to use in this project have come to the market in the last one or two years (in a truly useable form). This creates a window of opportunity to adapt methodologies previously only used with earlier sequencing techniques (such as Illumina reversible terminator sequencing, pyrophosphate sequencing, IonTorrent sequencing, PacBio sequencing) to the new possibilities.

As so often, with technical advances comes the possibility to address new questions. Here, we hope that we will be able to resolve the processes of nascent strand DNA methylation with better temporal resolution, and we hope to be able to more accurately describe its accuracy, both immediately after nascent stand synthesis, and after a time delay to give the epi-genomic machinery time to correct potential errors. We also want to better understand layers of control of DNA methylation, beyond parental DNA strand instruction that is already well understood. Finally, we hope to also better understand how cells avoid replicating methylation information faithfully when this is required for changes of cell fate?

Why does it matter? Apart from the basic interest in a fundamental biological process, there is also some medical interest. Malignancies arise from genetic changes (i.e. changes at the level of the sequence of DNA bases). However, as a malignancy evolves, the cells also show an increasing amount of epigenomic perturbations. Better techniques to monitor how these arise is therefore likely to eventually also have medical relevance.