Vertebrate species possess diverse phenotypic characteristics, yet they share similar repertoires of coding genes. Evolutionary changes in transcriptomes underlie structural and regulatory differences associated with species-specific characteristics. However, because organ-dependent mRNA expression levels within individual species have been largely conserved during vertebrate evolution, it seems unlikely that changes in gene expression account for the majority of phenotypic diversity among vertebrates. Through the variable use of cis-acting RNA elements in exons and flanking introns that are recognized by trans-acting factors, different pairs of splice sites in primary transcripts can be selected in a cell type–, condition-, or species specific manner. Changes in alternative splicing (AS) may therefore represent a major source of species-specific differences.

The main objective of this project was to perform the first large-scale comparative profiling of tissue-specific AS across vertebrate species and thereby identify biologically informative conserved and condition-specific splicing patterns. In the Blencowe Lab at the University of Toronto (outgoing host), we developed a pipeline for the analysis of RNA-Seq data and profiling of condition-specific AS and, by comparing the transcriptomes of multiple organs from ten vertebrate species spanning ~350 million years of evolution, we observe significant increases in the frequency of "cassette" exon events associated with proximity to the primate lineage. Within 6 million years, the exon-skipping profiles of physiologically equivalent organs diverged such that they are more strongly related to the identity of a species than they are to organ type.

We have also expanded our analysis to intron retention (IR). Of the different classes of AS, IR is the least well understood, although it is known to play important roles in the control of mRNA export, localization and turnover by nonsense-mediated decay. In plants and unicellular eukaryotes, IR is the most common form of AS, whereas in animals it is thought to represent the least prevalent form. We observed that IR is surprisingly frequent in mammals, affecting transcripts from as many as three quarters of multi-exonic genes.

Another goal of the project was to computationally infer combinations of cis-regulatory elements that are predictive of tissue-dependent splicing patterns in vertebrate species and thereby assemble the vertebrate "splicing code(s)". We used a splicing code derived from mouse data to predict, with high classification rates, tissue-dependent AS in vertebrates. We observe that organ-dependent AS patterns appear to be generally controlled by significantly overlapping cis-regulatory features (perhaps comprising an ancestral vertebrate splicing code) and that most vertebrate species-specific splicing patterns are cis-directed. We also found that a highly correlated set of cis-features comprising an "IR code" reliably discriminates retained from constitutively spliced introns.

A third aim of the project was the functional characterization of conserved and tissue/species-specific splice variants. We have identified and characterized species and lineage-classifying "cassette" exon AS events that are predicted to remodel protein-protein interactions involved in gene regulation and other processes. These events likely further contributed to the diversification of splicing and other transcriptomic changes that underlie phenotypic differences among vertebrate species.

We showed that IR acts widely to reduce the levels of transcripts that are less or not required for the physiology of the cell or tissue type in which they are detected. This "transcriptome tuning" function of IR acts through both nonsense mediated mRNA decay and nuclear sequestration and turnover of IR transcripts. We further showed that

IR is linked to a cross-talk mechanism involving localized stalling of RNA polymerase II (Pol II) and reduced availability of spliceosomal components. These results implicate a global checkpoint-type mechanism whereby reduced recruitment of splicing components coupled to Pol II pausing underlies widespread IR-mediated suppression of inappropriately-expressed transcripts.

We analyzed the functional importance of an alternative exon in the PTBP1 gene, whose skipping we had shown to be mammalian-specific. Using human cell lines engineered to uniquely express the different isoforms, we found that exon skipping results in a less potent PTBP1 protein. The relative ratio of PTBP1 isoforms in human cells and tissues impacts the regulation of its downstream targets. Deletion of this exon from a chicken cell line, a species that shows constitutive inclusion of the exon, results in a similar reduction in PTBP1 activity leading to altered regulation of its targets. This investigation reveals that evolution of exon skipping in a splicing regulator has contributed to alter the splicing network of its targets in mammals. Isoform-specific immunoprecipitation and quantitative mass spectrometry assays also suggest that the alternative exon may function by modulating the interaction between PTBP1 and its co-regulator RAVER1.

The fourth and final goal of the project was to determine the minimal set of code elements sufficient to promote constitutive and alternative splicing. To address it, at the the Carmo-Fonseca Lab at the Instituto de Medicina Molecular in Lisbon (return host), genomic integration of a single reporter gene, intron labelling with the MS2 technique and spinning disk confocal microscopy are being combined to understand how different splicing signals affect splicing in real time in the nucleus of living human cells. Using this approach, the Carmo-Fonseca Lab had already shown that replacing the weak polypyrimidine (Py) tract in mouse immunoglobulin Mu pre-mRNA by a U-rich Py decreases the intron lifetime, thus providing direct evidence that splice-site strength influences splicing kinetics (Martin RM et al. 2013, *Cell Reports* 4:1144).

The results on the research done in the outgoing host on “cassette” exon AS events resulted in a high-impact publication (Barbosa-Morais NL et al. 2012, *Science* 338:1587). The results of the research on IR resulted in a manuscript accepted for publication in *Genome Research*. The expertise in the statistical analysis of alternative splicing data acquired during the project has benefitted projects, both in the outgoing and the return hosts, that resulted in high-impact publications (Ward MC et al. 2013, *Molecular Cell* 49:262; Han H et al. 2013, *Nature* 498:241; Pena AC et al, *Molecular Microbiology* 2014 Jun 19, doi:10.1111/mmi.12677).

Fulfilling the knowledge transfer and integration goals of the Marie Curie International Outgoing Fellowship, the work done in this project has paved the way for the Fellow to lead an independent research group at the European return host.